UNIVERSITY OF CALIFORNIA SAN DIEGO

In vivo and in vitro strategies for characterizing secondary metabolite biosynthetic pathways from marine bacteria

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

in

Marine Biology

by

Katherine D. Bauman

Committee in charge:

Professor Bradley S. Moore, Chair
Professor Eric E. Allen
Professor William H. Gerwick
Professor Alexis C. Komor
Professor Victor Nizet

2022

The dissertation of Katherine D. Bauman is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To all my teachers over the years, thank you for showing a kid that wanted to do magic that science is really the same thing. To my parents, always, for being my original teachers. And finally, to Mick, for teaching me how to play again.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF DATA

ACKNOWLEDGEMENTS

I would like to thank my advisor and committee chair, Bradley S. Moore, for his unwavering enthusiasm, confidence, and support. I would not be the scientist, or person, I am today without him. Truly not enough can be said about the joyful and collaborative place that he has built in a very special lab down by the beach; his love and support for his students is felt every day. I can only hope to someday inspire as much joy and passion for science and the natural world as he does in his group.

I would also like to acknowledge Dr. Hanna Luhavaya, a phenomenal scientist, teacher, and role model, who I am also lucky enough to call my friend. It was such an honor to be trained by her, and I aim to emulate her rigor, thoughtfulness, and creativity in the lab every day. I can only hope to make her proud.

I'd also like to thank all members of the Moore lab, past and present, that I have had the joy and privilege of working with. I truly believe it is a badge of honor to be a Moore lab member, and it is one I wear with great pride.

Thank you to my parents, Lisa and Reid, and my sisters, Hannah and Sarah, for their unconditional support and love – the Baumans will forever be my favorite team. Finally, the biggest thank you to Hugh and Mick, who I am so lucky to share my life with. Thank you for teaching me how to lean into the mess, embrace imperfections, and find true joy along the way.

Chapter 1 contains material currently being prepared for publication, coauthored with Shende, V.V., and Moore, B.S. The dissertation author was the primary co-investigator and co-author of this chapter with V.V. Shende.

Chapter 2, in full, is a reprint of the material as it appears in *Cell Chemical Biology*, Bauman, K.D., Li, J., Murata, K., Mantovani, S.M., Dahesh, S., Nizet, V., Luhavaya, H., and Moore, B.S., 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in *Nature Chemical Biology,* Bauman, K.D., Shende, V.V., Chen, P.Y.-T., Trivella, D.B.B., Gulder, T.A.M., Vellalath, S., Romo, D., and Moore, B.S., 2022. The dissertation author was the primary researcher and author of this paper.

Chapter 4 contains unpublished material. The dissertation author was the primary researcher and author of this chapter.

VITA

2016 Bachelor of Arts in Chemistry, Middlebury College

2017-2020 National Institutes of Health T32 Marine Biotechnology Trainee

2020-2022 National Institutes of Health F31 Predoctoral Fellow

2022 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Perrault, J. R., Bauman, K. D., Greenan, T. M., Blum, P. C., Henry, M. S. & Walsh, C. J. Maternal transfer and sublethal immune system effects of brevetoxin exposure in nesting loggerhead sea turtles (*Caretta caretta*) from western Florida. *Aquat. Toxicol.* **180**, 131–140 (2016).

Bauman, K. D., Li, J., Murata, K., Mantovani, S. M., Dahesh, S., Nizet, V., Luhavaya, H. & Moore, B. S. Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry. *Cell Chem. Biol.* **26**, 724-736.e7 (2019).

Bauman, K. D., Butler, K. S., Moore, B. S. & Chekan, J. R. Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **38**, 2100–2129 (2021).

Schorn, M. A., Verhoeven, S., Ridder, L., Huber, F., Acharya, D. D., Aksenov, A. A., Aleti, G., Moghaddam, J. A., Aron, A. T., Aziz, S., Bauermeister, A., Bauman, K. D., Baunach, M., Beemelmanns, C., Beman, J. M., Berlanga-Clavero, M. V., Blacutt, A. A., Bode, H. B., Boullie, A., Brejnrod, A., Bugni, T. S., Calteau, A., Cao, L., Carrión, V. J., Castelo-Branco, R., Chanana, S., Chase, A. B., Chevrette, M. G., Costa-Lotufo, L. V., Crawford, J. M., Currie, C. R., Cuypers, B., Dang, T., de Rond, T., Demko, A. M., Dittmann, E., Du, C., Drozd, C., Dujardin, J.-C., Dutton, R. J., Edlund, A., Fewer, D. P., Garg, N., Gauglitz, J. M., Gentry, E. C., Gerwick, L., Glukhov, E., Gross, H., Gugger, M., Guillén Matus, D. G., Helfrich, E. J. N., Hempel, B.-F., Hur, J.-S., Iorio, M., Jensen, P. R., Kang, K. B., Kaysser, L., Kelleher, N. L., Kim, C. S., Kim, K. H., Koester, I., König, G. M., Leao, T., Lee, S. R., Lee, Y.-Y., Li, X., Little, J. C., Maloney, K. N., Männle, D., Martin H., C., McAvoy, A. C., Metcalf, W. W., Mohimani, H., Molina-Santiago, C., Moore, B. S., Mullowney, M. W., Muskat, M., Nothias, L.-F., O'Neill, E. C., Parkinson, E. I., Petras, D., Piel, J., Pierce, E. C., Pires, K., Reher, R., Romero, D., Roper, M. C., Rust, M., Saad, H., Saenz, C., Sanchez, L. M., Sørensen, S. J., Sosio, M., Süssmuth, R. D., Sweeney, D., Tahlan, K., Thomson, R. J., Tobias, N. J., Trindade-Silva, A. E., van Wezel, G. P., Wang, M., Weldon, K. C., Zhang, F., Ziemert, N., Duncan, K. R., Crüsemann, M., rogers, S.,

Dorrestein, P. C., Medema, M. H. & van der Hooft, J. J. J. A community resource for paired genomic and metabolomic data mining. *Nat. Chem. Biol.* **17**, 363–368 (2021).

Bauman, K. D., Shende, V. V., Chen, P. Y.-T., Trivella, D. B. B., Gulder, T. A. M., Vellalath, S., Romo, D. & Moore, B. S. Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead. *Nat. Chem. Biol.* **18**, 538–546 (2022).

ABSTRACT OF THE DISSERTATION

In vivo and in vitro strategies for characterizing secondary metabolite biosynthetic pathways from marine bacteria

by

Katherine D. Bauman

Doctor of Philosophy in Marine Biology

University of California San Diego, 2022

Professor Bradley S. Moore

For much of nature the complexity and diversity of language is found not in verbal cues, but rather in genetically encoded chemical signals. These small molecules, often referred to as “natural products”, have evolved over millennia to serve as a form of language used by organisms across the tree of life. The ecological roles of these compounds are varied, and still largely unknown, but humans have a long history of using these compounds for our own benefit, most famously as medicines.

These metabolites are produced through multi-step enzymatic reactions in which simple metabolic building blocks are assembled to yield a complex molecular structure. In bacteria, the enzymes necessary for these processes are encoded by genes grouped together on contiguous stretches of DNA, forming so-called 'biosynthetic gene clusters,' (BGCs). Connecting natural products to the genes responsible for their production is critical for the discovery of new bioactive molecules, and for expanding our knowledge of biochemical reactions which can inform sustainable, biocatalytic routes to therapeutically valuable chemicals. The overarching goal of this dissertation is to explore the biosynthesis of secondary metabolites from marine bacteria using both in vivo and in vitro strategies.

Both molecular families discussed in this thesis, streptophenazines and salinosporamides, incorporate intermediates from the shikimate pathway. Chapter 1, which serves as an introduction to the thesis, is a review of natural products derived from the shikimate pathway. Chapter 2 deals with the issue of silent BGCs and uses an in vivo, genetic engineering strategy to activate a silent BGC. This work linked the streptophenazine BGC to its molecular products for the first time and developed a series of synthetic biology tools to reliably regulate expression of BGCs. Chapter 3 uses an in vitro, biochemical approach to identify and characterize the enzyme responsible for the novel bicyclization reaction that constructs the bicyclized pharmacophore of salinosporamide A, a marine microbial natural product currently in Phase III clinical trials. Finally, Chapter 4 explores the current status of continued work on the salinosporamide pathway. The work discussed here includes both whole pathway in vivo strategies and

enzyme engineering in vitro approaches towards generating novel salinosporamides with alternative bioactivities.

CHAPTER 1. The biosynthesis of shikimate derived specialized metabolites

1.1 Introduction and context for Chapter 1

In plants and microorganisms the shikimate pathway is the key metabolic process for the biosynthesis of the aromatic amino acids phenylalanine, tyrosine, and tryptophan.[1] Biosynthesis begins when phosphoenolypyruvate (PEP), a product of glycolysis, and erythrose 4-phospate (E4P), from the pentose phosphate pathway, condense to form 3-deoxy-D-arabino-heptulosonate phosphate (DAHP). Six additional steps convert DAHP into chorismic acid, the last common intermediate which serves as the branch point for dedicated aromatic amino acid biosynthesis. Notably, the shikimate pathway is not found in animals, who instead are reliant on aromatic amino acids from their diet. This makes the shikimate pathway an intriguing target for herbicides or antimicrobials, as inhibitory compounds inherently have no human toxicity. Both types of molecules discussed in this dissertation, the streptophenazines (*spz* pathway) in Chapter 2 and the salinosporamides (*sal* pathway) in Chapter 3, are derived from intermediates of the shikimate pathway, which was the original inspiration for the following review article.

Feeding experiments in the 1970s demonstrated that the biosynthesis of phenazines branches from the shikimate pathway with the production of chorismic acid.[2] Dedicated phenazine biosynthesis begins with the conversion of chorismic acid into 2-amino-2-deoxyisochorismic acid (ADIC) by PhzE, an enzyme closely related to anthranilate synthases.[3] A conserved set of biosynthetic genes, including PhzE, are found in all BGCs that produce phenazines and are referred to as the *phz* operon. Likewise, feeding experiments that revealed the cyclohexenyl moiety of salinosporamide originates via a shunt in the shikimate pathway. Subsequent gene disruption experiments[4] and biochemical characterization[5] of a prephenate decarboxylase enzyme, SalX, identified

prephenate as the more immediate branch point for salinosporamide biosynthesis. Both the *spz* and the *sal* BGCs contain pathway-specific DAHP synthases, presumably to ensure ample shikimate intermediates are redirected from primary metabolism to secondary metabolite biosynthesis.

This chapter is currently in preparation for eventual submission to Natural Product Reports, co-authored with Vikram V. Shende and Bradley S. Moore. The purpose of this article is to review natural products derived from intermediates of the shikimate pathway along the seven steps from PEP + E4P to chorismic acid, prior to the dedicated aromatic amino acid biosynthetic steps. However, we have included separate, dedicated sections on natural products derived from prephenate and anthranilate as well, as well as a discussion of the related aminoshikimate pathway. This review is intended to update the 1997 article by Professor Heinz Floss on the same topic. Floss's article was published right as bacterial genome sequencing was emerging, but before the first genomes of any Actinomycetes were published, and the work discussed in his paper primarily uses isotopic labeling studies to link natural products to shikimate origins. The goal of our article is to update Floss's review for the genomic era. We discuss the genomic basis of the compounds identified by Floss, as well as biochemical support for the biosynthesis of these molecules from shikimate pathway origins. Additionally, we have included newer examples of molecules more recently identified as shikimate byproducts. The focus of our review is the genomic and biochemical data that links these small molecules to their biosynthetic pathways.

1.2 References for Chapter 1 Introduction

1. Herrmann, K. The shikimate pathway: early steps in the biosynthesis of aromatic compounds. *Plant Cell* **7**, 907–919 (1995).

2. McDonald, M., Mavrodi, D. V., Thomashow, L. S. & Floss, H. G. Phenazine biosynthesis in *Pseudomonas fluorescens*: branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *J. Am. Chem. Soc.* **123**, 9459–9460 (2001).

3. Blankenfeldt, W. & Parsons, J. F. The structural biology of phenazine biosynthesis. *Curr. Opin. Struct. Biol*. **0**, 26–33 (2014).

4. McGlinchey, R. P. et al. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc*. **130**, 7822 (2008).

5. Mahlstedt, S., Fielding, E. N., Moore, B. S. & Walsh, C. T. Prephenate decarboxylases: a new prephenate-utilizing enzyme family that performs non-aromatizing decarboxylation en route to diverse secondary metabolites. *Biochem*. **49**, 9021–9023 (2010).

1.3 Abstract

The shikimate pathway is the key metabolic process for the biosynthesis of the aromatic amino acids phenylalanine, tyrosine, and tryptophan. Seven metabolic steps convert phosphoenolypyruvate (PEP) and erythrose 4-phosphate (E4P) via shikimic acid into the common intermediate chorismate which in turn serves as the branch point for dedicated aromatic amino acid biosynthesis. The shikimate pathway is not found in animals, and as a consequence bacteria, fungi, algae, and plants uniquely exploit shikimate biosynthetic intermediates in their specialized metabolism. This review highlights the metabolic diversity derived from intermediates of the shikimate pathway along the seven steps from PEP and E4P to chorismate, as well as additional sections on compounds derived from prephenate and anthranilate. Chorismate is by far the most common metabolic branchpoint, leading to the greatest chemical diversity. We discuss the genomic basis and biochemical support leading to shikimate-derived antibiotics, lipids, pigments, cofactors, and other bioactive molecules.

1.4 Introduction

In plants and microorganisms the 20 common proteinogenic amino acids can be organized into six groups based on the shared metabolic intermediates that serve as their biosynthetic origin.[1] One of these biosynthetic families gives rise to the three aromatic amino acids (AAAs): phenylalanine (**1**), tyrosine (**2**), and tryptophan (**3**). The biosynthesis of these amino acids proceeds through a metabolic route called the “shikimate pathway”, named after the toxic Japanese star anise flower, *Illicium anisatum*, called *shikimi* in Japanese, from which the key metabolite shikimate (**4**) was first isolated.[2]

The pathway begins with the intersection of two important primary metabolic pathways when the glycolysis intermediate phosphoenolypyruvate (PEP, **5**) condenses with the pentose phosphate pathway intermediate erythrose 4-phosphate (E4P, **6**), to yield 3-deoxy-D-arabino-heptulosonate phosphate (DAHP, **7**) (**Figure 1.1**).[3] The elimination of phosphate from DAHP catalyzes the intramolecular exchange of the ring heteroatom for a carbon, resulting in a highly oxidized cyclohexyl ring referred to as 3-dehydroquinate (DHQ, **8**). A dehydration reaction produces 3-dehydroshikimic acid (DHS, **9**), the third intermediate in the pathway. Reduction of DHS forms shikimate (**4**), the pathway namesake and a key metabolic intermediate. A shikimate kinase generates shikimate-3-phosphate (**10**), the fifth shikimate pathway intermediate, which is subsequently converted to 5-enolpyruvateshikimate-3-phosphate (EPSP, **11**) by a second condensation with PEP. Finally, the loss of phosphate from this molecule yields chorismate (**12**), the main metabolic branch point and the last common intermediate for aromatic amino acid biosynthesis. At this point, biosynthesis branches and the production of **1** and **2** proceeds through prephenate (**13**), formed by a rare intramolecular pericyclic reaction catalyzed by chorismate mutase, while **3** is biosynthesized via an alternate route by anthranilate (**14**) synthase. In microbes, the shikimate pathway is regulated by AAA inhibition of DAHP synthase, but in higher organisms the complex regulation of this pathway has yet to be fully characterized.[4]

Comparing the shikimate pathway enzymes between plants, fungi, and bacteria reveals remarkable sequence similarity and commonalty in reaction mechanisms across kingdoms.[5] However, the molecular architecture of these pathways varies. In bacteria, the shikimate pathway is encoded by discrete monofunctional enzymes that catalyze

each biosynthetic step, often referred to as *aro* homologs, after the naming system adopted in *E. coli* for their role in aromatic amino acid production. In plants, these seven reactions are performed by six enzymes, including a bifunctional DHQ dehydratase / shikimate dehydrogenase that leads to the formation of shikimate.[6,7] Fungi, however, have evolved a pentafunctional polypeptide, called the AROM complex, that performs the five consecutive reactions from DAHP to EPSP.[8]

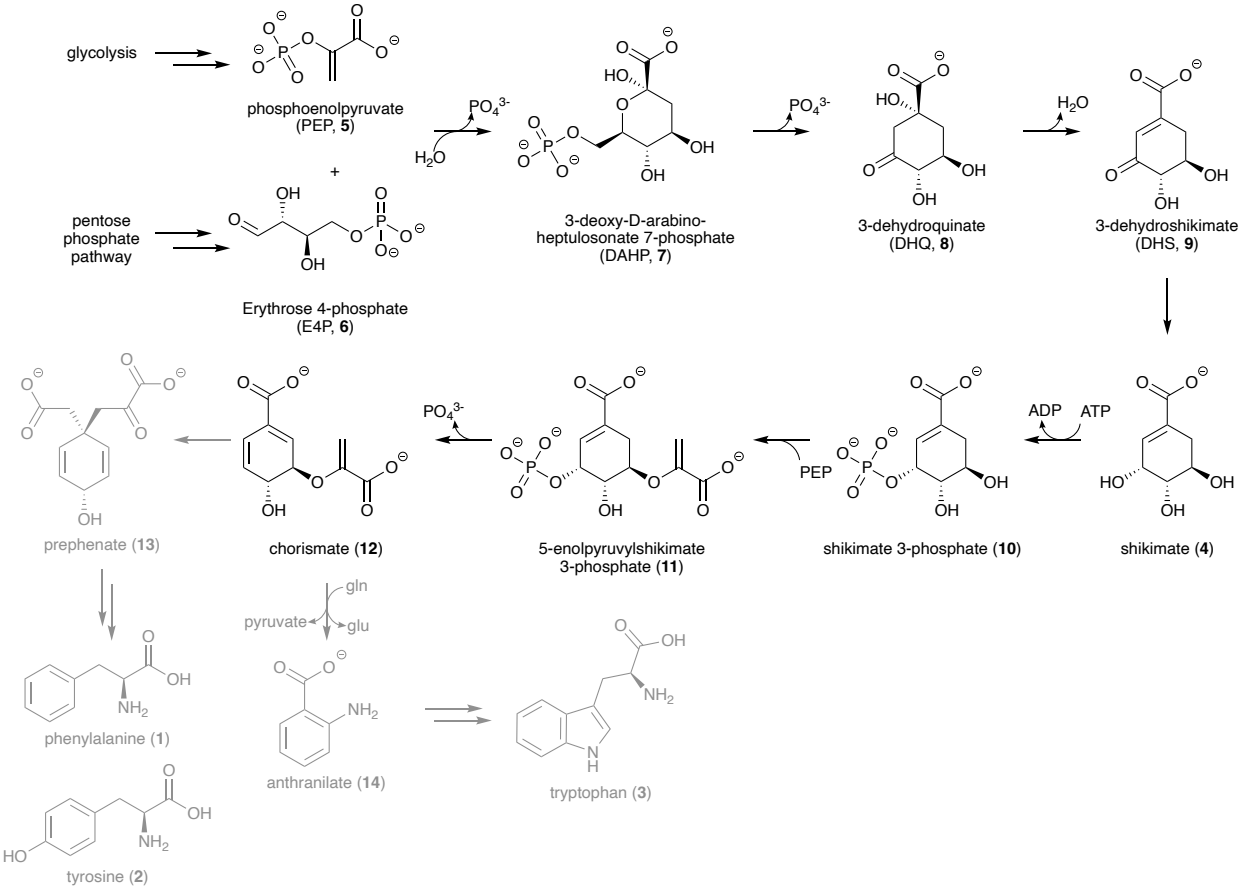


Figure 1.1 The shikimate pathwa

Notably, the shikimate pathway is found across domains of life in bacteria, archaea, fungi, algae, some protozoans, and plants, but never in animals. Instead, animals must retrieve aromatic amino acids from their diet, meaning that phenylalanine and tryptophan are considered “essential” amino acids since tyrosine can be produced in

animals by oxidation of phenylalanine. Because of this metabolic difference, the shikimate pathway has long been recognized as an attractive target for herbicides, antibiotics, and antiparasitics due to the inherent lack of human cross toxicity. The most used herbicide in the world, glyphosate (known as the Monsanto product Roundup), inhibits the sixth step of the shikimate pathway, EPSP synthase.[9] Transgenic Roundup Ready plants carry a form of the EPSP synthase gene that is not sensitive to glyphosate, conferring resistance to the herbicide.[10]

While the endpoints of the shikimate pathway, the amino acids themselves, are commonly incorporated into cofactors, pigments, organic polymers, and natural products, each of the intermediates en route to aromatic amino acids can serve as branchpoints for specialized metabolite biosynthesis as well. These highly oxidized and densely functionalized metabolites are perfectly poised to veer off their primary metabolic pathway and into secondary metabolism instead. Even the multidomained AROM complex is notoriously "leaky," allowing for derivatization of pathway intermediates.[8] This review focuses on these cases where metabolism hijacks intermediates from the shikimate pathway and uses them instead in the biosynthetic pathways of natural products in bacteria, fungi, and plants. We highlight natural products derived from intermediates of the shikimate pathway along the seven steps from PEP and E4P to chorismate, as well as sections on natural products originating from prephenate and anthranilate. Additionally, we have included a section on the aminoshikimate pathway, as many compounds originally thought to be of shikimate origin instead arise from this related pathway. The examples here are not meant to be comprehensive, but rather to

demonstrate the breadth of molecular diversity generated from shikimate intermediates as well as to highlight topics for future discovery.

We intend for this article to update the 1997 review by Heinz Floss on the same topic[11] for the genomic age by focusing on the genomic basis for shikimate-derived natural products and connecting molecular products to their producing genes. We discuss the biochemical and genetic experiments used to link secondary metabolites to their biosynthetic genes and establish their shikimate origin.

1.5 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP)

The condensation of phosphoenolpyruvate (PEP, **5**) and D-erythrose 4-phosphate (E4P, **6**) to 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP, **7**) is commonly recognized as the first committed step in the shikimate biosynthetic pathway (**Figure 1.2**). This reaction is catalyzed by 3-deoxy-D-arabinoheptulosonate 7-phosphate synthase[12] (DAHPS), an aldolase that is dependent on divalent cations.[13] The first high-resolution crystal structure was that of the phenylalanine-regulated DAHPS from *E. coli*,[14] and since then crystal structures of DAHPS from across the kingdoms of life have been solved, many of them in complex with substrates, metal cofactors, and inhibitors which have illuminated the catalytic mechanism of this enzyme and modes of inhibition.

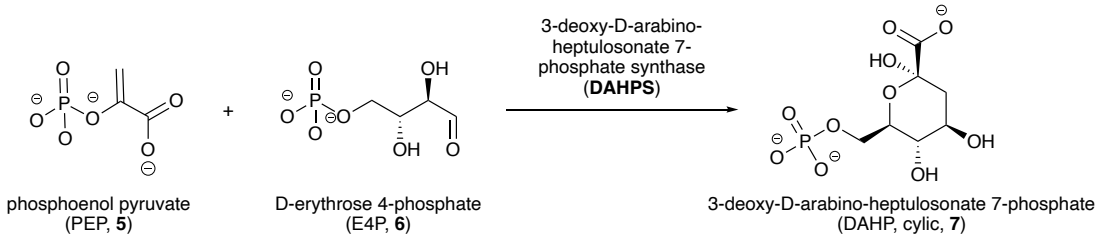


Figure 1.2 Aldol reaction of PEP and E4P catalyzed by DAHP synthase, the first committed step in the shikimate pathway.

In addition to catalyzing this first step in the shikimate pathway, DAHPS is responsible for controlling the flux of carbon into the shikimate pathway and is heavily regulated both via competitive feedback inhibition by downstream products such as amino acids, as well as transcriptional regulation.[15] Notably, in addition to three paralogous DAHPS enzymes ($AroA_I$-type) dedicated to supporting primary metabolic pathways such as amino acid biosynthesis, many bacteria possess an additional copy of DAHPS ($AroA_{II}$-type) which is phylogenetically distinct, is not inhibited by primary metabolic products of the shikimate pathway,[16] and is often embedded in secondary metabolite biosynthetic gene clusters (BGCs).[17,18]

To the best of our knowledge, there are no examples wherein DAHP (**7**) is shunted into a non-shikimate biosynthetic intermediate.

1.6 3-dehydroquinate (DHQ)

3-dehydroquinate (**8**, DHQ) is the second intermediate in the shikimate biosynthetic pathway and the first carbocyclic product of the pathway (**Figure 1.3**). The mechanism for conversion of the pyranose precursor DAHP to carbocyclic 3-DHQ have been informed by high resolution crystal structures of dehydroquinate synthase (DHQS) in complex with metal and nicotinamide cofactors and substrate-based inhibitors, and is proposed to proceed via a complex cascade of oxidation, elimination, reduction, ring opening and intramolecular aldol cyclization.[19–21] Despite the intricate nature of this multistep transformation, 3-DHQS rarely generates shunt products demonstrating the ability for this enzyme to stabilize intermediates and disfavor alternative modes of reactivity throughout the cascade.[21] Structural similarity between the widely distributed

mycosporines mycosporine-like amino acids (MLAAs), and shikimate pathway intermediates garnered hypotheses that 3-DHQ may be an intermediate in their biosynthesis. Initial isotope feeding studies in a variety of mycosporine/MLAA producers demonstrated the likely origin of the cyclohexane core to be derived from the shikimate pathway rather than polyketide origin.[22] To determine the source of the mycosporine/MLAA carbocycle, Walsh and coworkers identified and characterized the complete biosynthetic gene cluster for MLAA, shinorine (**15**) from *Anabaena variabilis*, and demonstrated that the conserved mycosporine precursor, 4-deoxygadusol (**16**), is derived from an alternative seven carbon pyranose, sedoheptulose-7-phosphate (**17**).[23]

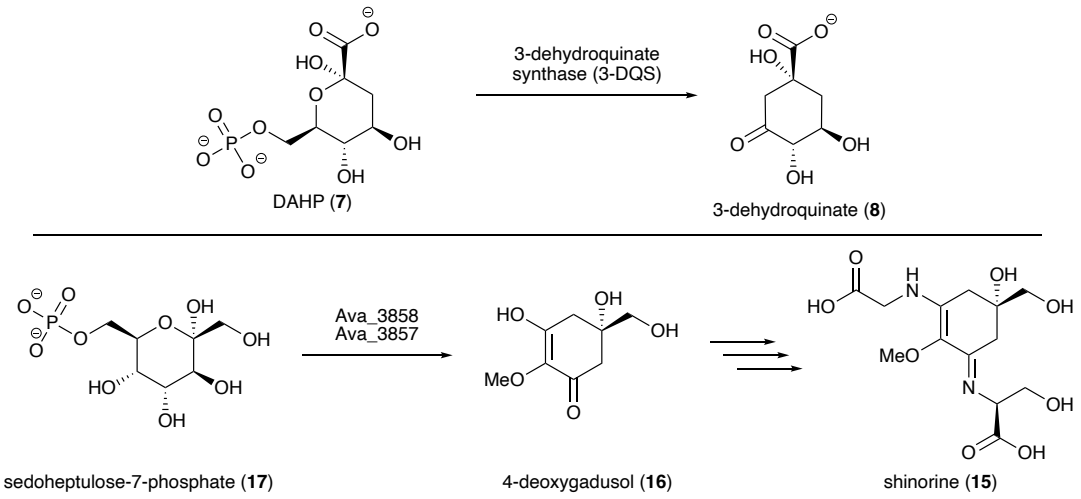


Figure 1.3 Conversion of DAHP to 3-DHQ by DAHPS and sedoheptulose-7-phosphate to 4-deoxygadusol by Ava3858/Ava_3857 from *A. variabilis* en route to mycosporines.

The paucity of microbial specialized metabolites directly derived from DHQ is curious and may reflect an underrepresented group of metabolites and thus a target for genome mining campaigns. This dearth of microbial compounds of DHQ origin is particularly notable given the large family of acyl-quinic acids isolated and characterized from plants.[24] Numbering over 300 examples to date, acyl-quinic acids, also known as

chlorogenic acids (despite the lack of chlorine atoms, **18** and **19**), play fundamental roles in plant maturation and lignin biosynthesis. While their occurrence, isolation, biosynthesis, and bioactivity has been reviewed recently,[24] the general structure of these compounds are quinic acid or a derivative coupled to coumaric acid or an analogue via an ester linkage (**Figure 1.4**).

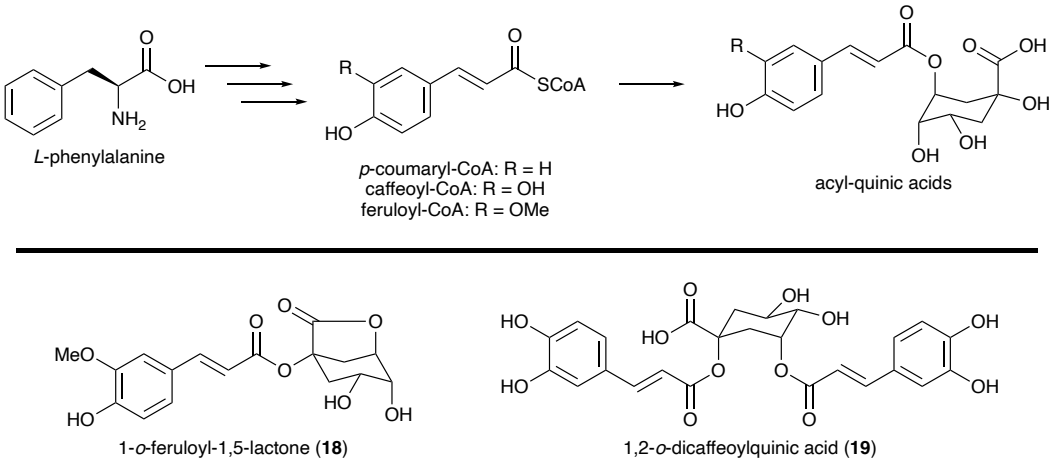


Figure 1.4 General biosynthesis of acyl-quinic acids and representative examples of known acyl-quinic acids from plants.

1.7 3-dehydroshikimate (DHS)

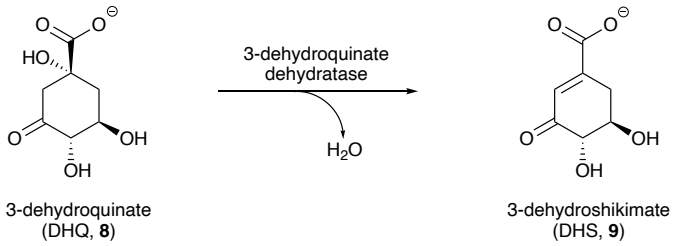
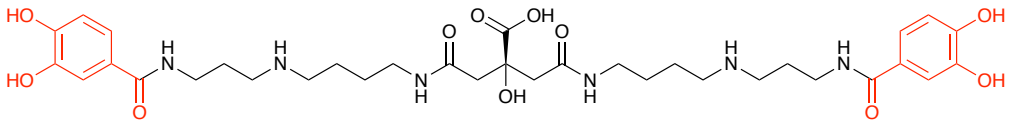


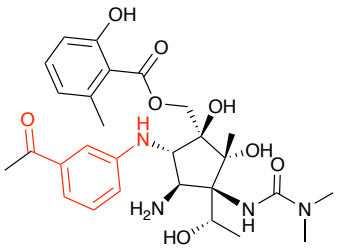
Figure 1.5 The dehydration reaction catalyzed by 3-dehydroquinate dehydratase to generate 3-dehydroshikimate

The dehydration of DHQ by 3-dehydroquinate dehydratase (DHQD) introduces the first double bond into the six membered ring and yields the third intermediate of the shikimate pathway, 3-dehydroshikimate (DHS, **9**) (**Figure 1.5**). Two classes of DHQD

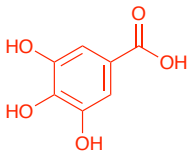
enzymes exist: type I and type II.[25] In one of the few examples of convergent evolution, these proteins share virtually no sequence or structural similarity and they catalyze dehydration via opposite stereochemical routes. Type I DHQD enzymes catalyze dehydration using an active site lysine residue that serves as the amino donor during the formation of a Schiff base intermediate, resulting in a *syn* elimination of water.[26] Type II DHQDs catalyze dehydration without Schiff base formation and proceed via the *anti*-elimination of water.[27] Interestingly, the distribution of type I vs type II DHQDs in bacteria does not appear to follow species phylogeny, although the presence of type II DHQHs in many pathogenic organisms has fueled interest in developing selective inhibitors.[28] In fungi, this reaction is catalyzed by a type I domain as part of the multifunctional protein complex AROM, while in plants, bifunctional enzymes catalyze both DHS and shikimate formation.[29] DHS serves as a starting point for the biosynthesis of several metabolic building blocks in microbes and plants, namely 3,4-dihydrobenzoic acid, 3-aminobenzoic acid, and gallic acid.



petrobactin (**20**)



pactamycin (**22**)



gallic acid (**24**)

Figure 1.6 Natural products derived from 3-dehydroshikimate include petrobactin, pactamycin, and gallic acid. Portions of the molecule arising from DHS are highlighted in red**.**

3,4-dihydrobenzoic acid (3,4-DHBA)

Bacillus anthracis, the causative agent of anthrax, produces a siderophore known as petrobactin that is considered an essential virulence factor for the organism during mammalian infection.[30] Petrobactin (**20**) chelates iron through a 3,4-dihydrobenzoic acid (3,4-DHBA, **21**) biosynthetic unit (**Figure 1.6**). Critically, this feature is particularly important not only for metal binding abilities, but because this moiety enables petrobactin to escape the mammalian immune protein siderocalin that sequesters siderophores as part of an antibacterial iron-depletion defense system.[31,32] Initial isotope feeding experiments suggested a shikimate origin for this chemical feature,[33] and ultimately the actual precursor to 3,4-DHBA formation was shown to be DHS.

Discovery of the petrobactin BGC and sequential deletion of every gene in the cluster revealed that *asbF,* annotated as a hypothetical protein,[34,35] was essential for 3,4-DHBA production. Almost simultaneously, two groups independently revealed that AsbF is a DHS dehydratase, capable of directly converting DHS into 3,4-DHBA via a dehydration reaction (**Figure 1.7**).[36,37] Full enzymatic and structural characterization suggested a requirement for a divalent manganese cation. The proposed AsbF mechanism proceeds via abstraction of a proton from C_4 of DHS by a histidine residue in the active site, followed by dehydration and aromatization. It is notable that in petrobactin biosynthesis it is the off-pathway divergence of a shikimate intermediate that generates the key bioactive chemical feature of the molecule, critical for both its metal chelating abilities and for avoiding detection by the mammalian immune defenses. Interestingly, recent work has demonstrated that production of petrobactin by *Alteromonas* spp. is critical for mediating iron acquisition across ocean environments, further demonstrating

the utility of this shikimate derived moiety, and its widespread distribution across bacteria.[38]

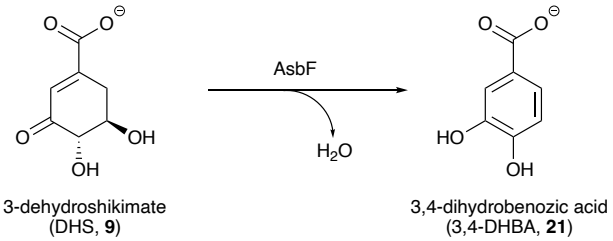


Figure 1.7 Aromatizing dehydration reaction catalyzed by AsbF in the biosynthesis of petrobactin to yield the critical 3,4-DHBA moiety from the shikimate intermediate DHS

3-aminobenzoate (3-ABA)

Beyond dehydration of DHS to 3,4-DHBA in petrobactin biosynthesis, DHS also serves as an unusual building block in the biosynthesis of pactamycin (**22**) through a transamination and double dehydration reaction to yield 3-aminobenzoate (3-ABA, **23**).[39] Pactamycin is a cytotoxic natural product produced by *Streptomyces pactum* containing a distinctive aminocyclopentitol core, decorated with 3-aminoacetophenone (3-AAP), 6-methylsalicyclic acid, and N,N-dimethylurea moieties (**Figure 1.6**).[40] Despite the potent bioactivity of this compound, its broad-spectrum toxicity limited its developed as a therapeutic, leading to interest in developing more selectively bioactive analogs.[41,42] Feeding experiments with isotopically labeled precursors revealed a shikimate origin for the 3-AAP unit based on the labeling patterns observed from incorporation of labeled glucose.[43] Incorporation of labeled 3-ABA demonstrated that this compound is a direct precursor to pactamycin.[39,44] However, the link between **23** and its shikimate origin remained unclear.

The pactamycin BGC was identified independently by two groups[41,45] using a characteristic methyltransferase as a genome mining hook. Gene inactivation of three genes in the putative cluster and heterologous expression of the putative 6-methylsalicyclic acid synthase definitively linked the BGC to pactamycin production. Years later, the conversion of a shikimate pathway intermediate to **23** catalyzed by an unusual aminotransferase, PctV, was biochemically validated.[39] Importantly, PctV was found to selectively accept DHS as its substrate and convert it to **23** (**Figure 1.8**). Not only is PctV able to perform the PLP-dependent transamination reaction on DHS, but it can 1.also perform two subsequent dehydrations to yield 3-ABA. Notably, PctV performs an analogous reaction to the AHBA synthase RifK, discussed at the end of this review in the aminoshikimate section, which catalyzes the dehydration of aminoDHS.

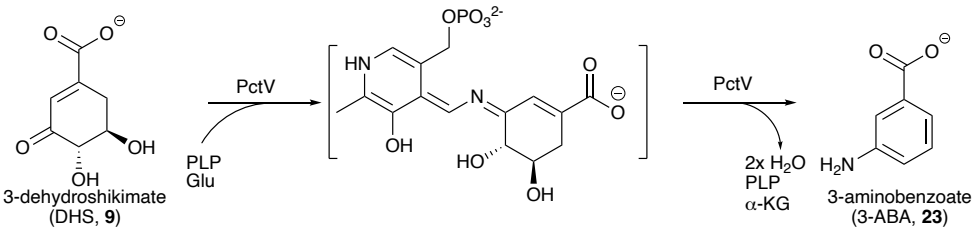


Figure 1.8 PLP-dependent aminotransferase and dehydration reaction catalyzed by PctV in the biosynthesis of pactamycin.

These biochemical experiments firmly established DHS as the shikimate pathway intermediate diverted from primary metabolism into pactamycin biosynthesis. Years later, it was revealed that once PctV generates ABA, this moiety is adenylated and loaded onto a carrier protein for condensation with a malonyl unit.[46] Subsequent glycosylation of this carrier protein tethered polyketide generates the key intermediate en route to the linked aminocyclopentitol and 3-AAP component of pactamycin.

Gallic acid

Outside of bacterial natural products, dehydroshikimate is also the known precursor to the plant metabolite gallic acid (**24**), a key component of plant tannins and recognized on its own for potent antioxidant and anti-inflammatory activity (**Figure 1.6**).[47] Initially, researchers were not sure if gallic acid was derived from an early shikimate precursor or the phenylalanine end product of the pathway, but stable isotope feeding experiments with labeled glucose in the high gallic acid producer fungi *Phycomyces blakesleeanus* demonstrated the direct conversion of DHS to gallic acid.[48] Follow up studies in the plant *Rhus typhina* used natural oxygen isotope abundance to demonstrate the same early shikimate origin DHS.[49] Ultimately, a shikimate dehydrogenase (SDH) enzyme from *Juglans regia* (English walnut) was identified that catalyzed the formation of shikimic acid and gallic acid.[50] The plant SDH was able to complement shikimate acid deficient *E. coli,* restoring shikimate and gallic acid production, and was able to produce gallic acid in vitro from DHS. Additionally, transgenic expression of the *J. regia* SDH enabled gallic acid production in *Nicotiana tabacum.* Gallic acid is thus an unusual case in which both a shikimate pathway intermediate (DHS in this case) and the shikimate pathway enzyme can be appropriated to play a role in secondary metabolite production.

1.8 Shikimate

Shikimate dehydrogenases catalyze the stereoselective reduction of 3-dehydroshikimate (DHS, **9**) to shikimate (**4**) with NADPH as a reducing agent (**Figure 1.9**). The archetypal shikimate dehydrogenase, AroE from *E. coli*, has been the subject of extensive structural and biochemical characterization.[51] Bacterial AroE is a standalone

monomeric enzyme, however shikimate dehydrogenases are discrete domains in eukaryotic multifunctional enzymes that catalyze multiple steps in shikimate biosynthesis. While AroE is selective for both substrates, homologs have demonstrated relaxed substrate specificity for both hydride donor (NADH or PQQ) as well as acceptor (quinate), including *E. coli* paralog, YdiB, which can utilize both NADH and NADPH.[52] To date, the processing of shikimate into reduced and dehydrated cyclohexanes and cyclohexenes are the only known routes to directly incorporate shikimate into natural product scaffolds possible presenting another target for genome mining.

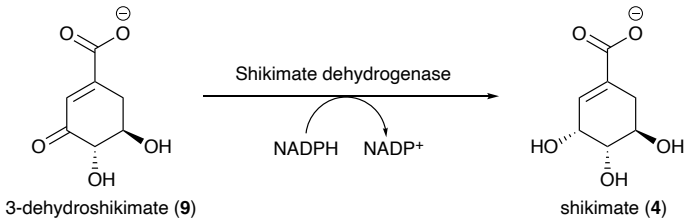


Figure 1.9 Conversion of 3-dehydroshikimate (**9**) to shikimate (**4**) by shikimate dehydrogenase.

Cyclohexane carboxylic acid

Shikimate itself is rarely directly incorporated into natural product scaffolds, with few notable exceptions such as the shikimate-cinnamate esters found in plants as discussed in the 3-DHQ section (**Figure 1.4**). The perhaps most thoroughly characterized pathway that utilizes shikimate as a starting material is the biosynthesis of cyclohexanecarboxyl-CoA (CHC, **25**). Acidophilic and thermophilic *Bacillus acidocaldarius* isolated from a hot acid spring and able to survive pH as low as 2 and temperatures up to 70 ºC, was found to produce ω-cyclohexyl fatty acids (**26**, **27**) as a likely mechanism to survive such conditions (**Figure 1.10**).[53] Foundational work using radioisotope labeled [^{14}C]-shikimate demonstrated the biosynthetic origin of fully reduced

o-cyclohexyl fatty acids to be from shikimate with all seven carbon atoms incorporated into fatty acids. Detailed stable isotope feeding studies with stereoselectively labeled isotopologues of shikimate, genetic manipulations, and in vitro enzymology enabled assignment putative intermediates from shikimate to **25** via a proposed nine-step pathway wherein no intermediates are aromatic. This work has been summarized in depth previously,[11,54] and this section will focus on recently published works that have continued to refine our understanding of the biosynthesis of (CHC) and its incorporation in natural product scaffolds including ansatrienin A (**28**) and asukamycin (**29**) (**Figure 1.10**). Notably both **28** and **29** will be discussed later in this review as well, in the context of other chemical features derived from alternatives to the shikimate pathway.

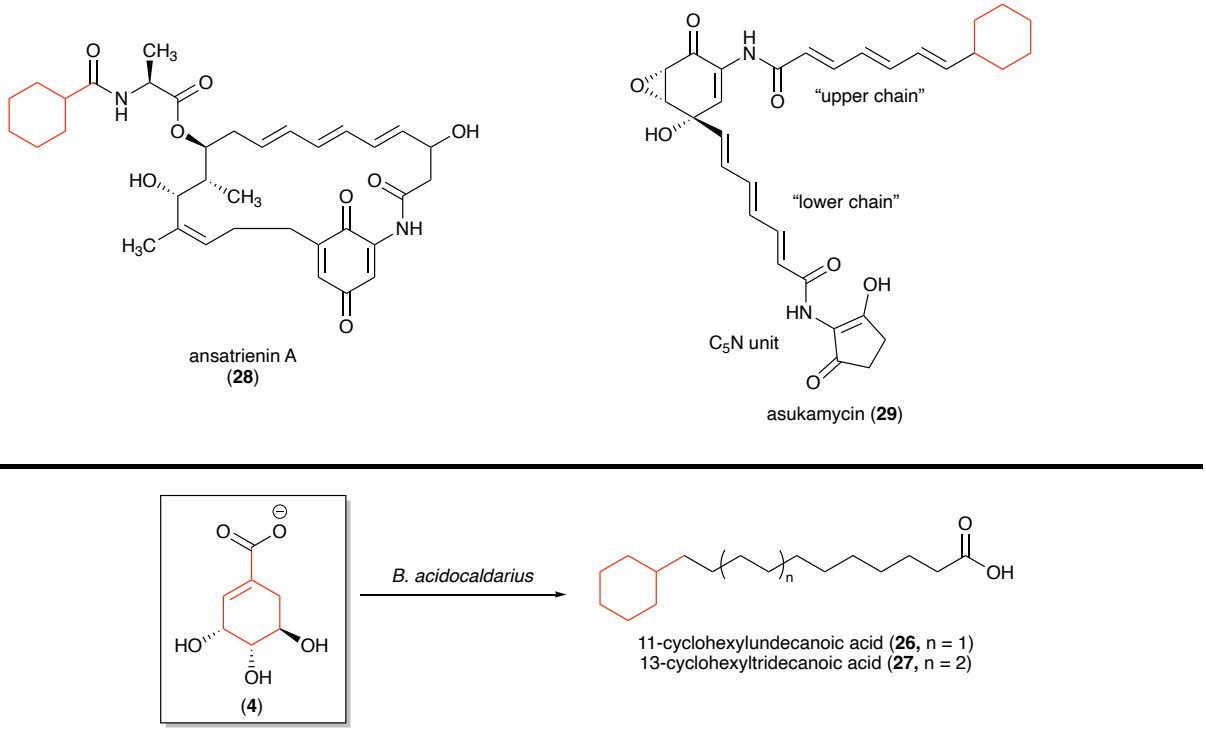


Figure 1.10 Natural products with CHC derived moieties, ansatrienin A and asukamycin, and feeding studies from *B. acidcaldarius* with fate of atoms from shikimate in red.

The first step of the biosynthesis of CHC to be biochemically validated in vitro was the reduction of 1-cyclohexenylcarboxyl-CoA (**32**) to CHC by enzyme ChcA, which represents the final step in CHC biosynthesis (**Figure 1.11**).[55] Examination of the genetic context of *chcA* in the ansatrienin (**28**) BGC from *Streptomyces collinus* revealed putative CHC biosynthetic enzymes upstream and downstream of *chcA*. Heterologous expression of this five gene cassette in *Streptomyces lividans* resulted in production of cyclohexyl fatty acids, demonstrating that these five genes represented the minimal suite of enzymes required for the conversion of shikimate to CHC.[56] To explore these enzymes and the identity of their associated substrates individually, Zhang and coworkers biochemically characterized AsuB1 and AsuB2, homologs to the first three genes in the CHC biosynthetic cluster from asukamycin (**29**) producer *S. nodosus* subsp. asukaensis.[57]

AsuB1 is a didomain protein, with an N-terminal domain homologous to EPSP synthase and C-terminal domain homologous to coumarate-CoA ligase. The full length AsuB1 as well as excised N-terminal domain ($AsuB1_N$) were recalcitrant to functional expression by *E. coli*, however C-terminal domain $AsuB1_C$ was successfully generated and purified. In vitro assays with coenzyme A and shikimate demonstrated $AsuB1_C$ catalyzed ATP-dependent thioester bond formation to yield **30,** thereby directing shikimate into CHC biosynthesis instead of towards amino acid construction (**Figure 1.11**). Assays using proposed alternative substrate shikimate-3-phosphate (**10**) resulted in no product formation, but 16 out 25 tested substrates resulted in the formation of CoA thioesters. This flexibility in acyl substrate displayed by $AsuB1_C$ was leveraged to generate substrates for characterization of downstream enzymes as well. AsuB2 (homologue of AnsL) was proposed to perform a non-redox E_1cB-typ elimination, and

indeed assays with recombinant AsuB2 and 3-hydroxycyclohexanecarboxy-CoA (**31**) resulted in dehydration with 1-cyclohexenecarboxyl-CoA (**32**) as the major product. Finally, a ChcA homolog, AsuB3, was confirmed to catalyze the reduction of **32** to **25**. Through this in-depth biochemical analysis, a revised CHC biosynthetic hypothesis was proposed which excluded shikimate-3-phosphate as a putative intermediate.

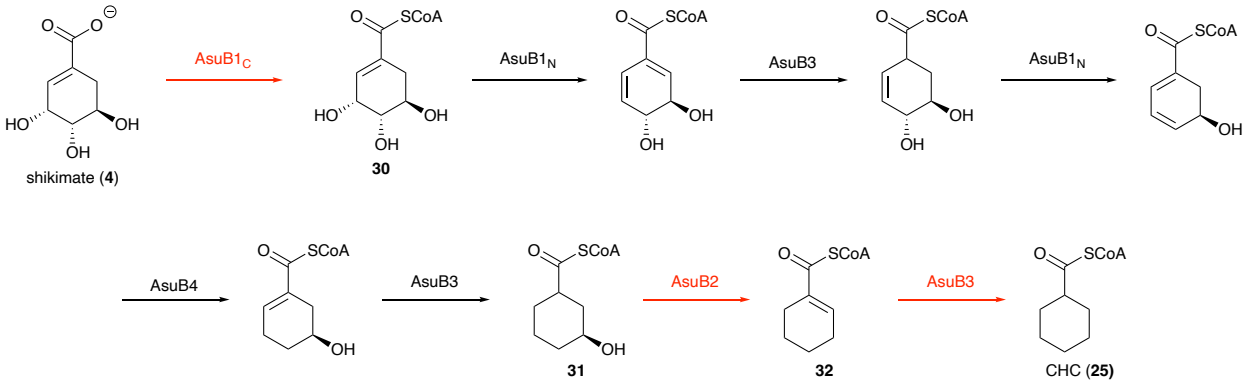


Figure 1.11 Zhang and coworkers' revised biosynthetic hypothesis for biosynthesis of cyclohexane carboxylic acid (biochemically characterized steps from their work in red).

1.9 Shikimate 3-phosphate (S3P)

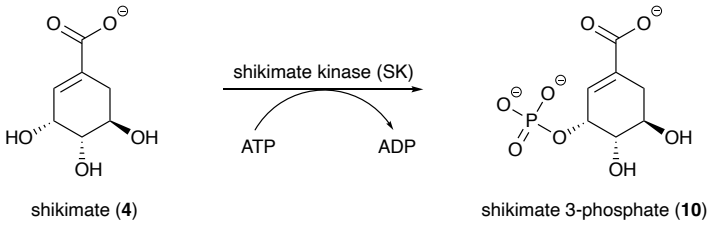


Figure 1.12 The ATP-dependent phosphorylation of shikimate to shikimate 3-phosphate catalyzed by shikimate kinase in the fifth step of the shikimate pathway.

Shikimate kinase, the fifth enzyme in the pathway, catalyzes the ATP-dependent phosphorylation of the 3-hydroxyl group of **4** to yield shikimate-3-phosphate (**10**) (**Figure 1.12**).[3] This enzyme requires a divalent cation (either Mg^{2+} or Mn^{2+}) for activity, and is a member of the nucleoside monophosphate kinase family, containing the characteristic

substrate binding domain, core domain responsible for ATP binding, and lid domains.[58,59] Both substrate and ATP binding induce massive conformational changes, lowering the lid domain and forcing ATP and the substrate into close proximity, resulting in transfer of the phosphoryl group. Two isozymes of shikimate kinase exist in *E. coli,* while the number of isozymes in plants is species specific. Through our research we have not identified any natural products that have been biochemically validated to arise from S3P.

1.10 5-enolpyruvylshikimate-3-phosphate (EPSP)

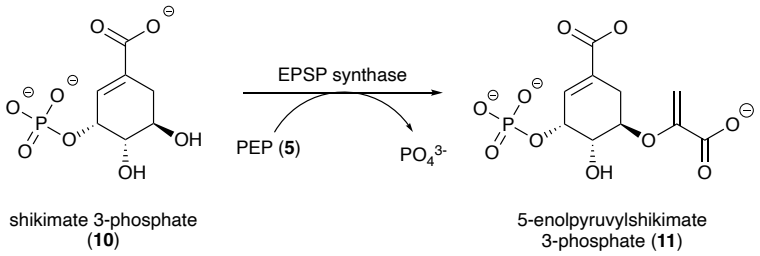


Figure 1.13 The condensation of PEP and shikimate 3-phosphate to 5-enolpyruvylshkimate 3-phosphate is catalyzed by EPSP synthase.

In the penultimate step of the shikimate pathway, a second unit of PEP is condensed with S3P, releasing inorganic phosphate, to generate 5-enolpyruvylshikimate-3-phosphate (EPSP, **11**) (**Figure 1.13**).[3] This reaction is performed by an alkyl transferase-type enzyme referred to as EPSP synthase. This reaction proceeds via C-O bond cleavage of PEP rather than the traditional P-O bond cleavage performed by most PEP-utilizing enzymes.[60] EPSP synthase catalyzes the enol ether transfer from PCP onto a hydroxyl group of S3P, proceeding through a tetrahedral intermediate.

This enzyme is of particular note because it is the target of the herbicide glyphosate, which can occupy the PEP binding site of EPSP synthase, acting as a

competitive inhibitor.[61] Naturally glyphosate tolerant versions of EPSP synthase have been identified in some organisms, which has led to the development of transgenic Roundup Ready crop lines that express a resistant form of this enzyme enabling glyphosate herbicidal treatment of these crops.[10] Because of the success of this system, EPSP synthase inhibitors are of tremendous interest in agricultural research. Despite the focus on EPSP synthase, to the best of our knowledge there are no natural products derived from EPSP. However, this may be a promising area for genome mining as EPSP mimics may serve as potential herbicides through inhibition of EPSPS.

1.11 Chorismate

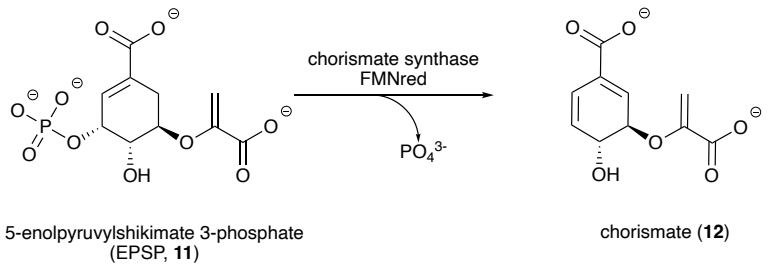


Figure 1.14 The reaction catalyzed by chorismate synthase

The final step in the shikimate pathway is the conversion of 5-enolpyruvylshikimate 3-phosphate (EPSP, **11**) to chorismate (**12**) by chorismate synthase (**Figure 1.14**). This enzyme catalyzes the 1,4-trans elimination of phosphate from EPSP, adding a second double bond to the six membered ring.[62] Chorismate synthase requires reduced flavin mononucleotide (FMN) despite the fact that the reaction involves no net redox change.[63,64] Chorismate synthases can be grouped into two classes: those found in *E. coli* and higher plants are considered monofunctional, and are unable to regenerate reduced FMN, while chorismate synthases from other organisms are considered bifunctional, and possess the ability to use NADPH to reduce FMN themselves.

Once chorismate is formed, this molecule can have a tremendous number of metabolic fates in microorganisms (**Figure 1.15**). Chorismate can be rearranged to yield prephenate or it can be transformed to anthranilate, the two reactions that lead to canonical AAA biosynthesis. However, chorismate can be directed away from amino acid biosynthesis entirely and instead modified by a variety of biosynthetic enzymes for incorporation into secondary metabolites. This section will discuss the plethora of reactions performed on chorismate beyond those towards amino acid biosynthesis, as well as the biosynthesis of the resulting natural products

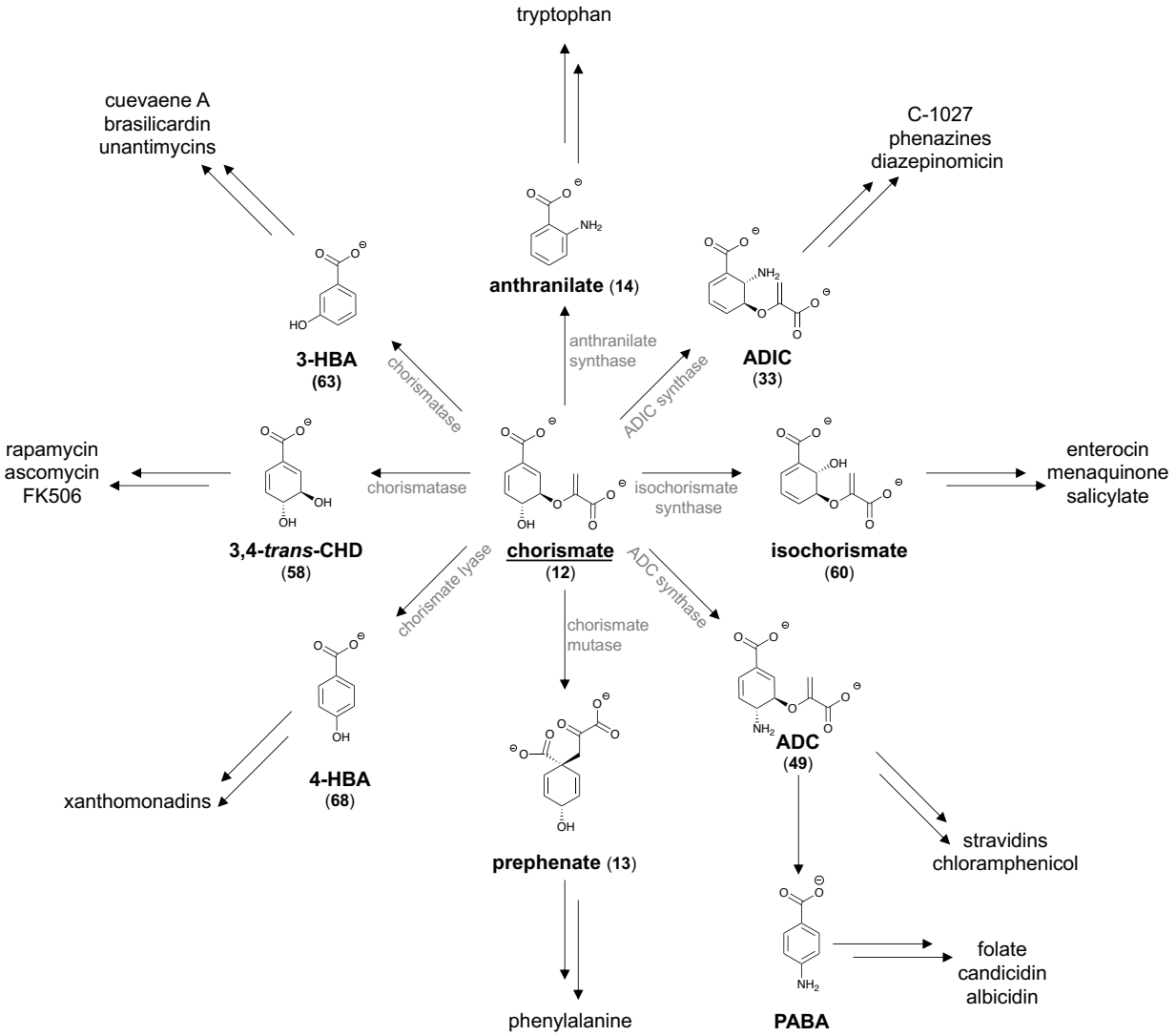


Figure 1.15 Metabolic fates of chorismate

1.11.1 2-amino-2-desoxyisochorismate (ADIC)

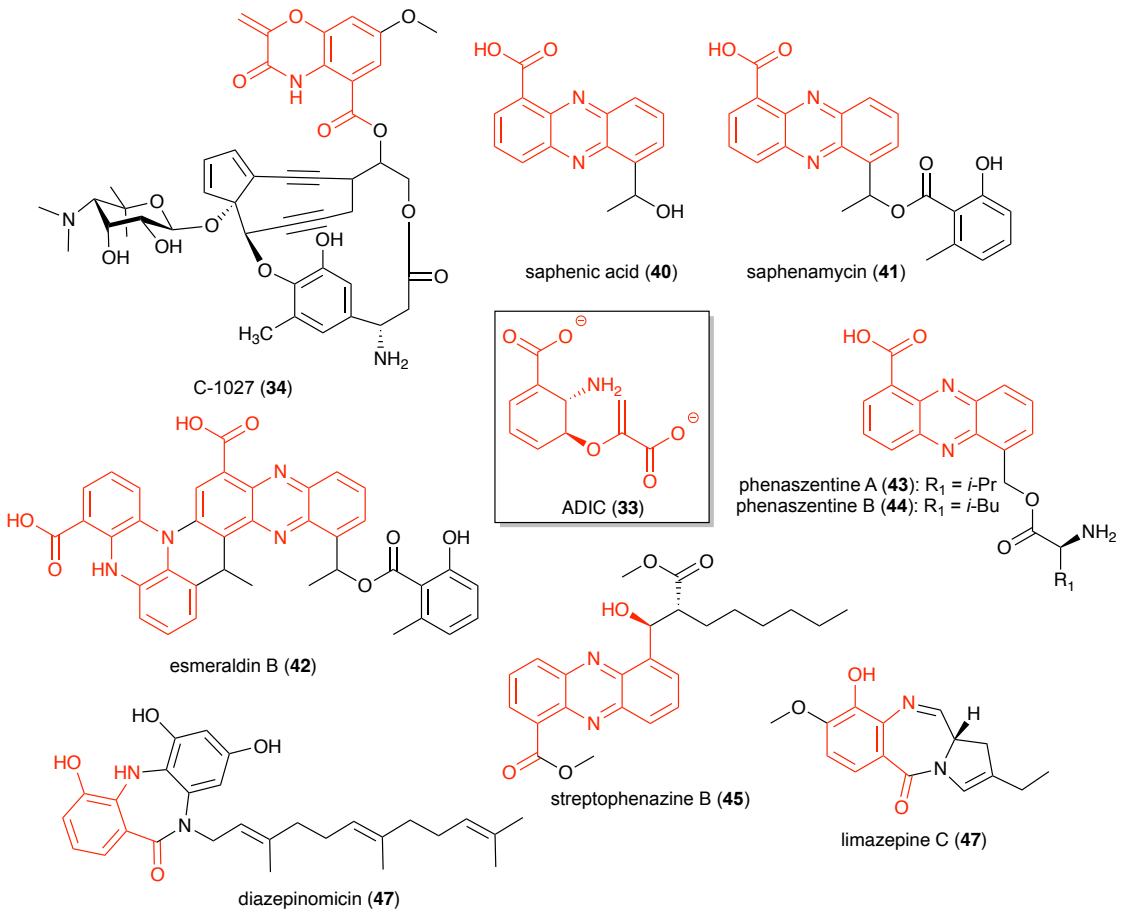


Figure 1.16 Natural products derived from chorismate via ADIC

C-1027

One of the alternative metabolic routes for chorismate beyond its canonical path towards amino acid biosynthesis is the formation of 2-amino-2-desoxyisochorismic acid (ADIC, **33**), an intermediate first identified in phenazine biosynthesis, discussed below, but first biochemically validated in the biosynthesis of the antitumor antibiotic C-1027 (**34**) (**Figure 1.16**).[65] C-1027 is produced by soil dwelling *Streptomyces* and consists of a complex of an enediyne chromophore and an apoprotein responsible for stabilizing and transporting the reactive enediyne small molecule.[66] This molecule is one of the most

cytotoxic structures ever discovered due to its ability to induce double stranded DNA breaks. Because of this potent bioactivity, C-1027 has been evaluated for clinical use, but development has been impeded by poor production from the native organism.[67]

This bioactivity has also initiated tremendous interest in the biosynthesis of this compound. The C-1027 chromophore consists of four distinct biosynthetic units: the reactive enediyne pharmacophore, a β-amino acid, a deoxy aminosugar, and a benzoxazolinate moiety. Importantly, this benzoxazolinate group specifically binds to the apoprotein, and is therefore critical in stabilizing the enediyne core. Initial work on the biosynthesis of this compound focused on the unusual PKS responsible for production of the enediyne, while the benzoxazolinate group was hypothesized to arise from anthranilate due to the presence of two anthranilate synthase homologs, SgcD and SgcG, encoded for in the C-1027 BGC.[66] However, in vitro biochemical experiments revealed that C-1027 biosynthesis does not proceed through anthranilate as previously thought. SgcD lacks the pyruvate lyase activity of anthranilate synthases, and instead catalyzes only the amination reaction to form ADIC. While ADIC is a known transient intermediate of anthranilate synthase, it remains enzyme bound and is not accumulated or released during the reaction. SgcD thus represents a new class of enzyme, a naturally occurring ADIC synthase capable of generating and releasing ADIC without conversion to anthranilate. The next enzyme in the pathway, SgcG, is an ADIC dehydrogenase that generates 3-O-enolpyrvoylanthranilate, further demonstrating that pyruvate is retained during SgcD biosynthesis. These enzymes thus catalyze the first committed steps towards bezoxazolinate biosynthesis, commandeering chorismate from primary metabolism and using it as the substrate for C-1027 biosynthesis. The complete

biochemical characterization of SgcD as a true ADIC synthase validated a new branchpoint for chorismate metabolism and expands the metabolic potential of this pivotal intermediate. While ADIC had been previously identified as an intermediate in phenazine biosynthesis, a bone fide ADIC synthase had never been characterized until this work.

Phenazines

While SgcD from the C-1027 biosynthetic pathway was the first characterized ADIC synthase, ADIC was identified as an intermediate in the biosynthesis of phenazines years earlier. Phenazines are a large class of heterocyclic, redox-active secondary metabolites produced by both Gram-positive and Gram-negative bacteria, but most commonly associated with *Pseudomonas*.[68–70] Most phenazines exhibit broad-spectrum antibiotic activity due to their ability to generate reactive oxygen species, but these compounds play a wide variety of physiological roles as well.[71] Studies have demonstrated the role of these compounds as electron shuttles, ensuring anaerobic survival for their microbial producers under conditions of oxygen limitation and thereby facilitating the formation of biofilms. Because of this, phenazines are considered key virulence factors in *Pseudomonas* infections. These ecological roles, impacts on human health, and the structural diversity of this compound class have made the biosynthesis of phenazines an intriguing question.

Initial feeding experiments in the 1960s showed incorporation of ^{14}C labeled shikimic acid into phenazines, with the symmetrical phenazine core moiety originating from two shikimate molecules.[72,73] Chorismic acid was long suspected as the branchpoint for phenazine biosynthesis,[74,75] but this question was not definitively solved until decades

later, after the genes responsible for phenazine biosynthesis were identified in the late 90s.[76,77] A conserved set of five genes, *phzBDEFG,* make up the *phz* operon and are required for the biosynthesis of phenazine-1-carboxylic acid (PCA) and phenazine-1,6-dicarboxylic acid (PDC), the core phenazines from which the structural diversity of this class is derived (**Figure 1.17**). Additionally, most phenazine BGCs contain a second copy of a DAHP synthase enzyme, encoded by *phzC,* to ensure ample flux through the shikimate pathway, and occasionally a duplicated copy of *phzB* referred to as *phzA.*

The first biochemical analysis of the *phz* operon was completed in 2001, and conclusively identified the conversion of chorismic acid to 2-amino-2-desoxyisochorismic acid (ADIC, **33**) by PhzE as the branchpoint for phenazine biosynthesis (**Figure 1.17**).[78] McDonald et al. expressed the *phz* operon in *E. coli* and incubated the cell-free extracts with chorismic acid, ADIC, or anthranilic acid, which revealed complete conversion to PCA only when ADIC was used as a substrate. Years later, *in vitro* characterization of PhzE and crystallography experiments revealed that this enzyme is a bona fide ADIC synthase that commits chorismic acid to phenazine biosynthesis via the formation of ADIC, requiring $MgCl_2$, and glutamine.[79] PhzE shares structural and sequence homology to anthranilate synthase enzymes and catalyzes a similar SN2 nucleophilic substitution via a glutamine amidotransferase domain. The resulting product, ADIC, is then hydrolyzed to *trans*-2,3-dihydro-3-hydroxyanthranilic acid (DHHA, **35**) by PhzD, an isochorismatase. Subsequent work in 2004[80,81] identified the next enzyme in the pathway, PhzF, as an isomerase that catalyzes the conversion of DHHA to 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid (AOCHC, **36**).[80] Two AOCHC molecules can spontaneously condense to yield the tricyclic phenazine core, but multiple studies have shown this

reaction is greatly facilitated by PhzA/B.[82,83] The resulting intermediate, hexahydrophenazine-1,6-dicarboxylic acid (HHPDC, **37**), can undergo spontaneous decarboxylation en route to phenazine formation. Finally, an additional two electron oxidation is required to achieve aromaticity, which is suspected, although not fully characterized, to be catalyzed by PhzG, to yield the two known phenazine cores, phenazine-1-carboyxlic acid (PCA, **38**), and phenazine-1,6-dicarboyxlic acid (PDC, **39**).[84] These experiments established the biosynthetic route for formation of the phenazine core from chorismate, but there is a tremendous amount of structural diversity in this class of molecules due to the modification and tailoring reactions of that core.

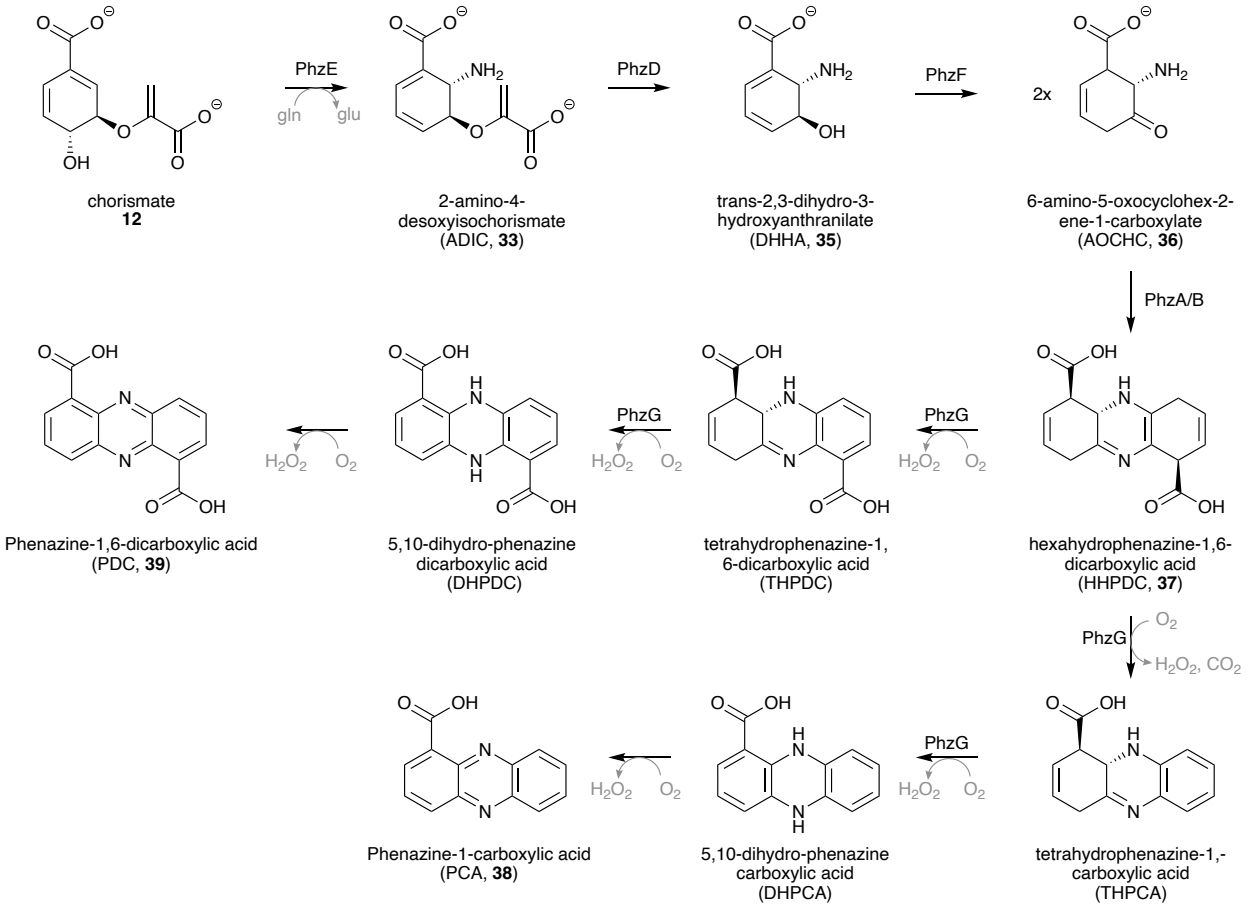


Figure 1.17 Phenazine core biosynthesis

Esmeraldins

The esmeraldins are a family of dimeric, green-pigmented phenazines produced by *Streptomyces antibioticus* Tü 2706.[85,86] They are isolated along with several precursors, including saphenic acid (**40**) and saphenamycin (**41**) (**Figure 1.16**). Esmeraldin B (**42**) has a mild antitumor effect, while **41** exhibits potent antibiotic and mosquito larvicidial activity.[87] The Floss lab was particularly interested in these molecules as much of the early work on phenazine biosynthesis was studied from *Pseudomonas* strains, not *Streptomyces,* and no phenazines with additional carbon scaffolds beyond the phenazine core had been studied biosynthetically. Feeding experiments by the Floss group confirmed the shikimate origin of phenazines from *Streptomyces,* and that the additional carbon appeared to originate from acetate.[88] Both ^{13}C labeled PDC and [^{13}-$^{2}H_3$]-6-acetophenazine-1-carboxylic acid were incorporated into the mature esmeraldins, revealing the start of a proposed biosynthetic pathway. However, this route was only fully established in 2012 with the discovery of the esmeraldin BGC.

This discovery was hampered for a long time because of *S. antibioticus* Tu 2706's tendency to lose the ability to produce esmeraldins, suggesting an instability of the biosynthetic genes. By creating and screening a cosmid library, the Yu group were able to identify a 30 kb region containing the esmeraldin (*esm*) BGC. Ultimately, they revealed that this BGC is harbored on a large plasmid, explaining the instability of the esmeraldin gene and *S. antibioticus* Tu 2706's occasional biosynthetic incompetence. To combat this instability, transposon mutagenesis was used to construct mutants of every gene in the *esm* BGC. Using these single gene mutants and a series of complementation experiments, they were able to establish a biosynthetic route for esmeraldin construction.

The phenazine core PDC is assembled by Esm1-8, which include duplicate copies of three enzymes from the shikimate pathway. PDC can then go on to serve as a starter unit for a one-carbon extension performed by the PKS EsmB4, followed by a ketoreduction to the alcohol catalyzed by EsmC. Finally, EsmD1-3, an atypical PKS, are responsible for esterification of 6-methylsalicylic acid and saphenic acid to yield saphenamycin, demonstrating a novel carrier protein dependent esterification reaction to attach aromatic moieties to a shikimate derived phenazine core. Overall, the work described here directly supports past feeding experiments by demonstrating the link between the shikimic acid pathway and phenazine biosynthesis. Beyond that, this work was the first to characterize any biosynthetic machinery responsible for modification of the phenazine core.

Phenaszentines and streptophenazines

In 2019, two additional papers characterized biosynthetic routes to phenazine production. The Bode lab characterized two phenazine producing BGCs (*xpz*) from *Xenorhabdus szentirmaii,* an entomopathogenic bacterial symbiont of the nematode *Steinernema rarum*. Heterologous expression of the first BGC produced both PCA and PDC that are additionally modified by two monooxygenases to yield oxidatively modified phenazines. The second BGC heterologously produced the previously characterized griseolutetic acid, as well as a hybrid peptide-phenazine compound. A promoter replacement strategy in the native organism resulted in the isolation of three novel phenazine-peptide compounds based on the griseolutetic acid core that the authors named the phenaszentines (**43, 44, Figure 1.16**). The authors also identified a series of

polyketide modified phenazines that appeared to emerge nonenzymatically, due to the lack of PKS genes present in the *xpz* BGCs.

That same year, work in the Moore lab uncovered the gene cluster responsible for the biosynthesis of streptophenazines (**45** is representative streptophenazine B, **Figure 1.16**) for the first time.[89] Streptophenazines are a family of phenazine metabolites with a polyketide derived alkyl chain produced by *Streptomyces.*[90] These molecules display moderate antibiotic activity, and have been isolated from both marine and soil-derived *Streptomyces.* This gene cluster was identified due the presence of the phenazine core operon homologs, as well as a duplicated copy of a DAHP synthase, directly linking the biosynthesis of streptophenazines to the shikimate pathway. Heterologous expression and refactoring off the streptophenazine BGC via promoter replacement resulted in the production of over 100 phenazine metabolites, including a novel series of streptophenazines bearing an N-formylglycine motif. The streptophenazine pathway proceeds via PDC formation followed by polyketide extension, as noted by the presence of PKS biosynthesis genes. The formation of the formylglycine moiety is intriguing, and gene disruption experiments demonstrated that a putative adenylating protein, Spz15, is required for installation of this chemical feature.

Taken together, these works demonstrate the variety of tailoring reactions that can be used to modify the phenazine core. While the diversity of phenazine metabolites come from these biosynthetic modifications, their bioactivity is tied to their redox abilities which originate from their shikimate-derived cores.

Bacterial benzodiazepines

Benzodiazepines are a structurally diverse class of molecules defined by a benzene ring fused to a seven-membered diazepine ring, but differing in substituents off of that core feature. Diazepinomicin (**46**) is a potent antibiotic benzodiazepine isolated from both terrestrial and marine *Micromonospora* strains, containing a tricyclic dibenzodiazepinone ring system that is exceptionally rare in bacterial natural products.[91] This molecule contains two substituted benzene rings, both of which bear structural similarity to shikimate derived natural products. To explore this hypothesis, initial stable isotope feeding experiments identified 3-hydroxyanthranilic acid as a precursor for ring A of the dibenzodiazepinone core.[92] Genome analysis revealed a putative diazepinomicin gene cluster by the presence of a gene cassette involved in isoprenoid biosynthesis colocalized with genes required for shikimate production.

One gene in particular, ORF19, was of particular interest due to its similarity to *phzE* which encodes the key enzyme responsible for stealing the shikimate product chorismate and converting it into ADIC in phenazine biosynthesis.[92] Like PhzE, ORF19 encodes a glutamine amidotransferase domain, and is likewise predicted to produce ADIC from chorismate. Two other genes in the diazepinomicin BGC are predicted to catalyze the downstream conversion of ADIC to 3-hydroxyanthranilate via pyruvate lyase and aromatization reactions. Finally, an adenylate ligase, ORF24, likely adenylates 3-hydroxanthranilate for a condensation reaction with aminohydroxy-1,4-benzoquinon to generate the dibenzodiazepinone core.

However, diazepinomicin biosynthesis was entirely hypothesized bioinformatically and this route to 3-hydroxyanthranilate through ADIC remained biochemically unvalidated

until 2018, when the biosynthetic pathway of the limazepines was uncovered. The limazepines are a family of pyrrolobenzodiazepine molecules with potent antitumor activity produced by soil dwelling *Streptomyces* bacteria.[93] The putative limazepine BGC was identified using antiSMASH, and a gene encoding for a PhzE homolog, Lim6, immediately suggested a biosynthetic route to 3-hdyroxyanthranilate via ADIC, as was suspected for diazepinomicin biosynthesis.[94] In vitro assays with Lim6 indeed confirmed its role as an ADIC synthase, and that downstream enzymes Lim5 and Lim4 are able to convert ADIC to 3-hydroxyanthranilate (**Figure 1.18**). This compound can then enter the *lim* NRPS to generate the pyrrolobenzodiazepine core.

These works demonstrate ADIC's position as an incredibly versatile metabolic building block, capable of incorporation into an enediyne natural product, or of generating a tricyclic phenazine core, or, finally, of transforming into anthranilic acid in bacterial benzodiazepine biosynthesis. However, another route to pyrrolobenzodiazepine biosynthesis found in fungi will be discussed in the anthranilate section of this review.

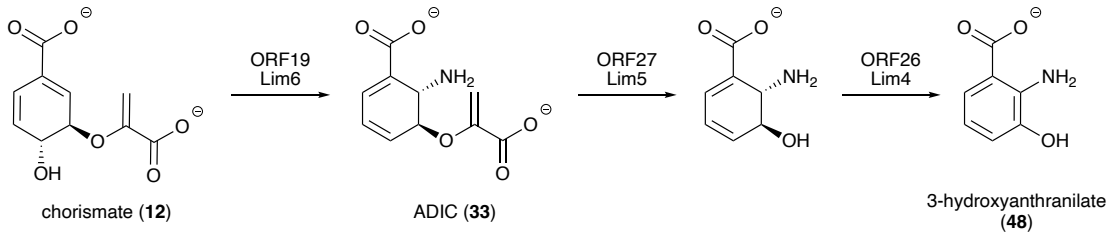


Figure 1.18 Biosynthesis of 3-hydroxyanthranilate from ADIC in the bacterial benzodiazepine natural products diazepinomicin and the limazepines

1.11.2 4-amino-4-deoxychorismate (ADC)

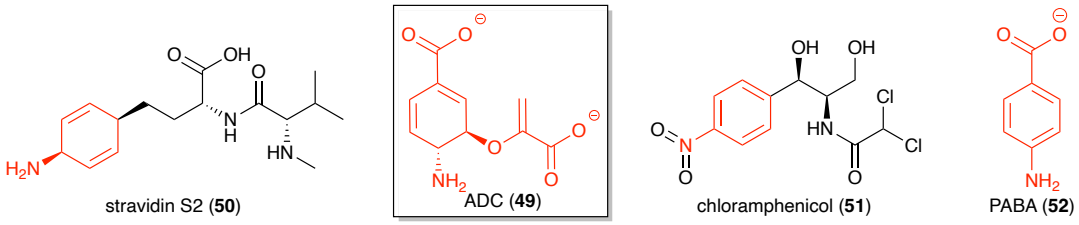


Figure 1.19. Natural products derived from ADC

Stravidins and chloramphenicol

Stravidins are a family of antibiotic dipeptides produced by *Streptomyces avidinii* and are known inhibitors of biotin biosynthesis. These small molecules are isolated along with the biotin binding protein streptavidin, resulting in a potent synergistic cocktail that causes biotin deficiency in microbes.[95,96] Stravidins contain the nonproteinogenic amino acid amiclenomycin (Acm) which binds and inhibits the biotin biosynthesis protein BioA by forming an irreversible adduct with BioA's pyridoxal 5'-phosphate (PLP) cofactor. This activity appears highly specific to *Mycobacteria*. Acm features a characteristic amino-substituted 1,4 cyclohexadiene moiety that is critical for this activity, but the origins of this amino acid, and the biosynthesis of stravidins, remained unknown for decades.

To explore stravidin biosynthesis, the Kelleher lab used their "metabologenomics" pipeline to identify an orphan BGC correlated with stravidin S2 (**50**) (**Figure 1.19**) production across a large *Streptomyces* dataset. Heterologous expression of the putative stravidin (*svn*) operon validated this assignment. Like other BGCs that utilize shikimate pathway intermediates, the *svn* BGC contains an additional copy of a DAHP synthase to ensure ample precursor production and avoid negative feedback regulation by aromatic amino acids, which suggested a shikimate origin for the Acm amino acid. From the gene

cluster organization, the authors proposed a biosynthetic pathway that diverges from the shikimate pathway with chorismate and proceeds through 4-amino-4-deoxychorismate (ADC, **49**), a reaction precedented in the biosynthesis of chloramphenicol (**51**).[97,98]

The conversion of chorismate to ADC was originally characterized in folic acid biosynthesis, where two proteins, PabA and PabB, form a heterodimeric complex to convert chorismate and glutamine to ADC and glutamate.[99,100] ADC can then be subsequently transformed by a lyase to yield *para*-aminobenzoic acid (PABA, **52**), a key intermediate in folate production. In chloramphenicol biosynthesis, ADC is not allowed to proceed to **52**, and is instead converted by a mutase to 4-amino-4-deoxyprephenate (**53**) instead (**Figure 1.20**), followed by aromatic decarboxylation towards mature chloramphenicol synthesis.[97] The stravidin gene cluster contains both an ADC synthase homolog (*svnN*) as well as a ADC mutase homolog (*svnK*), as in chloramphenicol biosynthesis. However, Acm retains the 1,4 cyclohexadiene feature of the prephenate ring, suggesting a novel nonaromatic decarboxylation step (**Figure 1.20**). The enzyme responsible for this reaction is proposed, but not biochemically validated in this work, opening the door for future biosynthetic studies. Ultimately, this work identified the gene cluster responsible for stravidin biosynthesis and linked the production of these molecules to the shikimate pathway. The authors propose a biosynthetic route that follows the shikimate pathway to the production of chorismate, where an ADC synthase hijacks this substrate for stravidin production.

It is important to emphasize here that despite the importance **52** plays in human metabolism because of its incorporation in folate, it is ADC that is the key intermediate between the shikimate pathway (chorismate) and **52**. While **52** can incorporated into

natural product biosynthetic pathways, such as candicidin,[101,102] xanthomonic acid,[103] and albicidin[104], it is not the direct shunt product from the shikimate pathway, and so these PABA containing natural products will not be discussed in detail here

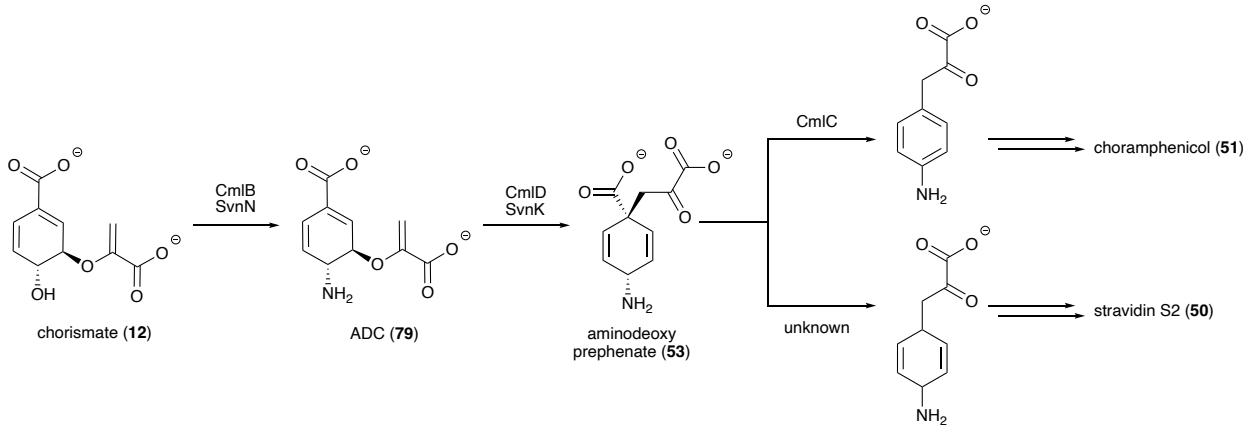


Figure 1.20 Biosynthetic route from chorismate through ADC shared by chloramphenicol and stravidin. At the point of aminodeoxy prephenate these pathways diverge, through either aromatic or non-aromatic decarboxylation reactions.

1.11.3 3,4-*trans*-dihydroxycyclohexa-1,5-dienecarboxylate (*trans*-CHD)

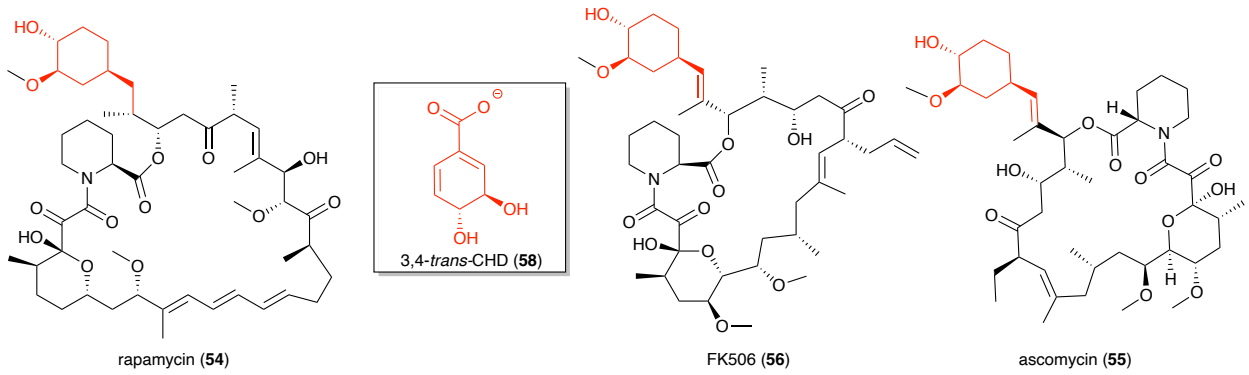


Figure 1.21 Natural products derived from 3,4-trans CHD include the immunosuppressants FK506, rapamycin, and ascomycin (FK520)

Rapamycin, ascomycin, and FK506

The macrocyclic polyketides rapamycin (**54**), ascomycin (FK520, **55**), and FK506 (**56**) are natural products isolated from soil dwelling *Streptomyces* with powerful immunosuppressant activity (**Figure 1.21**).[105] Both rapamycin (Sirolimus) and FK506 (Tacrolimus) are used clinically after organ transplantation to prevent organ rejection by the host. The powerful bioactivity of rapamycin led to the discovery of its physiological target, a protein kinase critical for regulating growth in animals, which was subsequently named mTOR (mammalian target of rapamycin) after its inhibitor.[106] However, the inhibition of mTOR first requires rapamycin to bind the common cellular protein peptidyl prolyl isomerase FK506-binding protein (FKBP12) via a key pipecolate moiety of the molecule.[107] This rapamycin-FKBP12 complex can then interact directly with mTOR and inhibit its activity. FK506 similarly binds to FKBP12, but the complex has a different cellular target, instead inhibiting phosphatase calcineurin instead of mTOR.[108] This powerful bioactivity has made the biosynthesis of these compounds particularly intriguing.

Early feeding experiments clearly demonstrated that ^{13}C-labeled shikimic acid was incorporated into the cyclohexane moiety of both rapamycin[109] and ascomycin.[110] This molecular feature is the presumed started unit of polyketide synthesis for both these molecules. Identification of the mixed PKS/NRPS BGC responsible for rapamycin production in 1995,[111] along with extensive feeding experiments, ultimately identified 4,5-dihydroxycyclohex-1-enecarboxylic acid (DHCHC, **57**), with an unsaturated six membered ring, as the true starter unit for rapamycin biosynthesis.[112] The biosynthesis of **57** was still suspected to branch from the shikimate pathway at the point of **4**, but the actual route between the shikimate pathway and **57** remained unknown.

Gene disruption experiments demonstrated the essential nature of *rapK* in generating the PK/NRP scaffold of **54**, specifically in biosynthesis of the starter unit.[113] Deletion of *rapK* was chemically complemented by the addition of **57**, thereby generating a mutasynthetic route to the production of analogs of **54** and confirming its biosynthetic role in production of **57**.[114] Additionally, genetic complementation of the Δ*rapK* mutant with the homolog *fkbO* from the FK506 BGCs restored rapamycin production, indicating this enzyme has an identical function in all three pathways. Expression of $FkbO_{520}$ in *E. coli* and in vitro biochemical experiments revealed that this enzyme is actually a chorismatase that converts chorismate to 3,4-trans-dihydroxycyclohexa-1,5-dienecarboxylic acid (*trans*-CHD, **58**), catalyzing the first committed step towards **57** biosynthesis.[115] FkbO is therefore the founding member of a new class of chorismate utilizing enzyme that hydrolyze pyruvate from chorismate.

This work identified the true shikimate pathway intermediate that enters rapamycin biosynthesis and identified a new class of chorismatase enzymes, further diversifying the metabolic fates available for chorismate in secondary metabolism. Finally, understanding the biosynthesis of rapamycin and FK506 has enabled the discovery of new bioactive analogs of this family. In a 2019, work by Warp Drive Bio used the lysine cyclodeaminase gene that installs the pipecolate residue essential for FKBP12 binding as a genome mining hook to discover seven BGCs predicted to encode the biosynthesis of novel rapamycin analogs.[116] They ultimately isolated and characterized a new natural product, WDB002, that similarly binds FKBP12 but interacts with a new physiological target, the centrosomal protein CEP250.

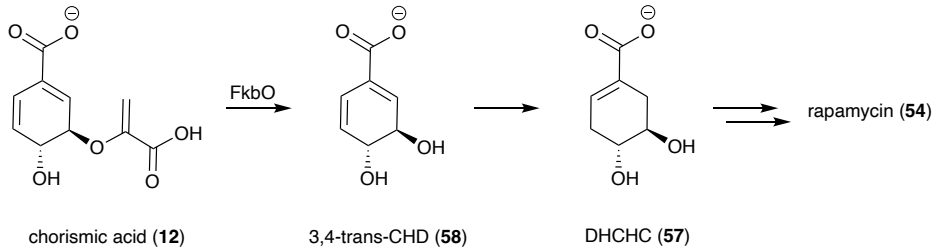


Figure 1.22 Biosynthesis of DHCHC (**57**), the starter unit for rapamycin biosynthesis from chorismate

1.11.4 Isochorismate

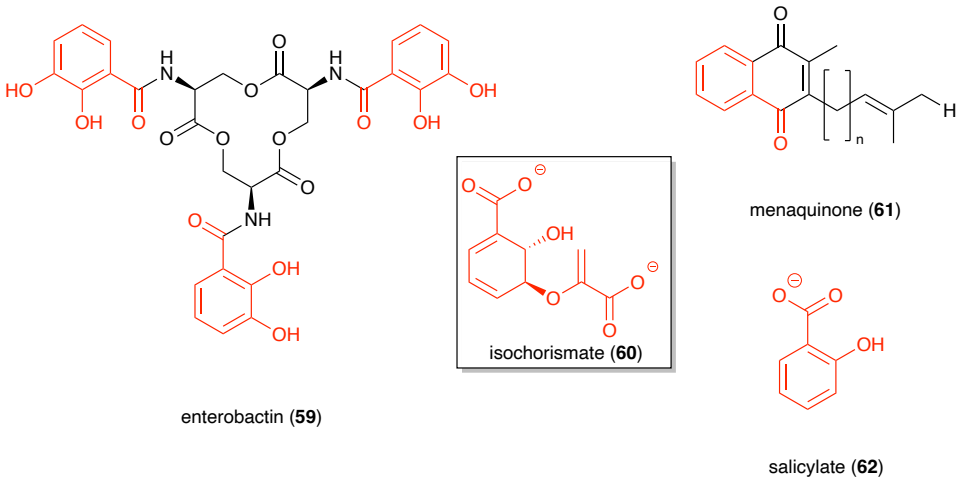


Figure 1.23 Natural products derived from isochorismate

Enterobactin, menaquinone and salicylate

The bacterial siderophore enterobactin (**59**), also known as enterochelin, is one of the most powerful chelators known to man, with an iron binding affinity of 10^{52} M^{-1}. This compound is found widely in Gram-negative bacteria, but is predominantly associated with *E. coli* and *Salmonella* strains. Enterobactin is a cyclic lactone composed of three 2,3-dihydroxybenzoate serine units that chelates iron using a catechol metal binding ligand (**Figure 1.23**). Because of the powerful bioactivity of this molecule, the gene cluster responsible for its production was identified and sequenced from *E. coli* mutants unable

to produce enterobactin.[117,118] Chorismate was identified as an intermediate en route to enterobactin early on, as crude enzymatic preparations showed the conversion of chorismic acid to 2,3-dihdyroxybenzoate (DHB).[119] Isolation of isochorismic acid from crude cell extracts began to establish a biosynthetic route to DHB production.[120] Ultimately, expression and purification of each of the six enzymes, EntA-F, from the gene cluster revealed that EntC initiates enterobactin biosynthesis through the transformation of chorismate into isochorismate (**60**).[121] This divergence from the shikimate pathway proceeds through the stereo- and regio-specific displacement of the 4-hydroxyl group in chorismate by water, yielding isochorismate.[122] Isochorismate then serves as the substrate for EntB, a lyase which catalyzes the loss of pyruvate, and subsequent aromatization by EntA to yield DHB. EntDEF are responsible for the downstream construction of mature enterobactin from three DHB units. The elucidation of enterobactin biosynthesis revealed a new metabolic fate for chorismate in secondary metabolism, and EntC remains archetypical for the identification of isochorismate derived natural products.

Beyond enterobactin, a number of other well-known secondary metabolites diverge from the shikimate pathway at chorismate through isochorismate. Menaquinone (**61**), also known as vitamin K2, is biosynthesized from isochorismate by a specific isochorismate synthase, MenF.[123] Additionally, the anti-inflammatory compound salicylate (**62**), famously recognized as a defensive response in plants and the active ingredient in aspirin, utilizes isochorismate synthase to divert chorismate into secondary metabolite production.[124,125] This route is mirrored in some bacterial producers of salicylate as well.[126]

1.11.5 3-hydroxybenzoate (3-HBA)

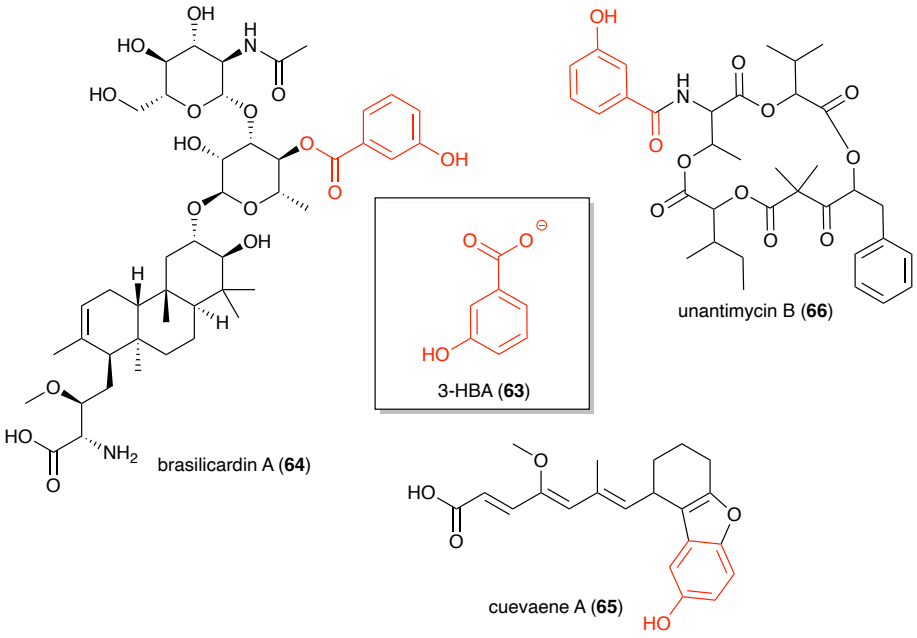


Figure 1.24 Natural products derived from 3-HBA

Brasilicardin, cuevaene A, and unantimycin B

The discovery of FkbO from the FK506 biosynthetic pathway as a new class of chorismate utilizing enzyme also revealed closely related proteins that used chorismate as a substrate, but instead catalyze the production of 3-hydroxybenzoate (3-HBA, **63**) rather than CHD (**Figure 1.24**).[115] One of these proteins, Bra8, is found in the BGC responsible for the production of brasilicardin (**64**), a terpenoid antibiotic with impressive immunosuppressant activity produced by the actinomycete *Nocardia brasiliensis.*[127,128] Heterologous expression of a homologous protein, Hyg5, and in vitro assays confirmed the role of these enzymes as chorismatases / 3-HBA synthases that catalyze both hydrolysis and concomitant dehydration of chorismate.

The discovery of this family of 3-HBA synthases opened the door for genome mining efforts to discover novel natural products with this chemical feature. Genome

analysis of a soil dwelling *Streptomyces* strain revealed an orphan gene cluster containing an Hyg5 homolog.[129] Disruption of this gene, *cuv10,* and comparison of metabolite profiles between the mutant and wild type strains revealed a minor metabolite that, upon isolation, was revealed to be cuevaene A (**65**). Cuevaene A is a previously isolated, tricyclic triene 3-HBA containing natural product produced by *Streptomyces* shown to have moderate antibiotic activity.[130] Expression and purification of Cuv10 from *E. coli* enabled in vitro biochemical experiments with this enzyme, which confirmed it is indeed a 3-HBA synthase. Bioinformatic predictions suggest that 3-HBA serves as the starter unit for six polyketide extensions by Cuv11-14.

A similar pathway using 3-HBA as a starter unit for assembly line biosynthesis was recently identified for the production of unantimycin B (**66**), a *Streptomyces* natural product with potent and selective anticancer activity.[131] A putative chorismatase, *nat-hyg5*, was identified in the gene cluster and inactivated via gene disruption experiments which revealed its essential nature for unantimycin biosynthesis.[132] In vitro characterization of Nat-hyg5 confirmed it is indeed a 3-HBA synthase.

The discovery and characterization of 3-HBA as a starter unit in brasilicardin, cuevaene A, and unantimycin biosynthesis showcases the diverse biosynthetic classes capable of using this shikimate derived metabolite. These pathways encompass terpenoid, PKS, and NRPS biochemistry, but all begin with 3-HBA.

1.11.6 4-hydroxybenzoate (4-HBA)

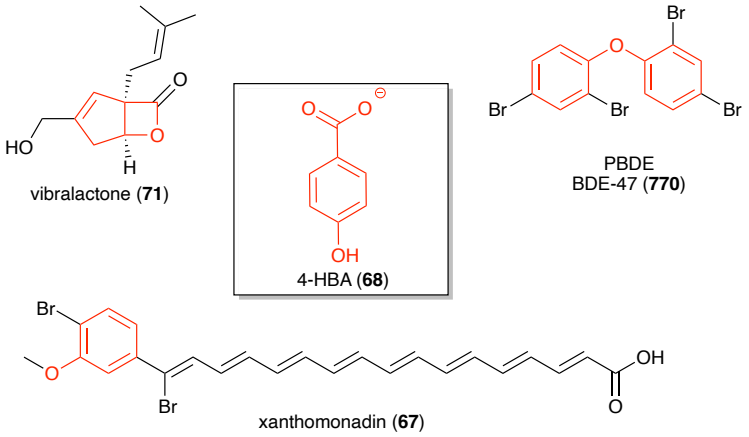


Figure 1.25 Natural products derived from 4-HBA. include the halogenated metabolites xanthomonadin and polybrominated diphenyl ethers including BDE-47. The biosynthesis of vibralactone is suspected to proceed through 4-HBA but this has not been confirmed.

Xanthomonadins and ubiquinone

The xanthomonadins (**67**) are a series of halogenated, aryl-polyene, membrane-bound, yellow pigments produced by the phytopathogenic bacteria *Xanthomonas* spp. (**Figure 1.25**). These compounds are critical for bacterial survival as they protect the pathogen from UV damage and reactive oxygen species.[133] The biosynthesis of these compounds has been shown to be regulated by 3-HBA[134] production, and so a link between an early intermediate of the shikimate pathway and xanthomonadin biosynthesis was well established.[134] Additionally, the discovery of Hyg5 and Bra8 as chorismatases responsible for generating 3-HBA for use in secondary metabolism prompted the examination of the xanthomonadin BGC for a similar chorismate utilizing enzyme.

Gene disruption experiments revealed that a functioning shikimate pathway was required for xanthomonadin production, and identified a gene, *xanB2,* that was involved in 3-HBA synthesis in xanthomonadin producing strains.[135] However, the substrate for

this enzyme remained unknown and any biochemical characterization of XanB2 was lacking. Unexpectedly, in vitro experiments with purified XanB2 and chorismate revealed two products, 3-HBA (**63**) and 4-HBA (**68**). The authors also demonstrated that while 3-HBA is directly incorporated into xanthomonadin biosynthesis, 4-HBA is used instead for an alternative route to ubiquinone (**69**) production. Recent work, however, contradicts this finding and suggests that both 3-HBA and 4-HBA are involved in xanthomonadin biosynthesis.[136] When both 3-HBA and 4-HBA are present, the strain seems to have a preference for 3-HBA. In the absence of 3-HBA, however, 4-HBA can be used towards xanthomonadin production.

Both structurally and mechanistically, XanB2 is thus a unique bifunctional chorismatase capable of producing both 3-HBA and 4-HBA simultaneously. No other chorismatase has been found to have this ability, and examination of the XanB2 active site revealed a deviation from a conserved motif among chorismatases that may be responsible for this broader product range. This reaction is proposed to proceed through an arene oxide intermediate, which in XanB2 may open in an unselective manner, leading to both 3- and 4-HBA.[137]

Polybrominated diphenyl ethers (PBDEs)

Polybrominated diphenyl ethers (**70**) are highly brominated, pervasive marine metabolites produced by marine γ-proteobacteria that bioaccumulate up the food chain (**Figure 1.25**).[138] These natural products resemble man-made PBDEs which have been employed commercially as flame retardants, but have recently been banned due to toxicity concerns owing to their structural similarity to human thyroid hormones. Despite

the abundance of these metabolites and their chemically intriguing structures, no biosynthetic pathway had been established.

To establish a genetic basis for the natural production of PBDEs, the Moore lab used the pyrrole halogenase from the pyoluteorin BGC as a genomic search hook, as halogenation is often a telling and distinctive feature of secondary metabolism. Using this halogenation hook they were able to identify a putative *bmp* biosynthetic gene cluster in the genome of a *Pseudoalteromonas* bacteria and confirm its role in PBDE biosynthesis via heterologous expression in *Escherichia coli*.[139] Previous isotope feeding studies had suggested a shikimate origin for the benzene ring, specifically via a *p*-hydroxybenzoic acid intermediate,[140] and the presence of a chorismate lyase in the putative *bmp* cluster further confirmed that origin. The next enzyme in the pathway, the flavin-dependent decarboxylase-halogenase Bmp5, preferentially used 4-HBA as its substrate and generated 3-bromo-4-hydroxybenzoic acid as a stable intermediate. This work demonstrated that metabolic flux is diverted away from the shikimate pathway at the point of chorismate through a 4-HBA intermediate towards PBDE biosynthesis

A few years later the Moore lab used the *bmp* pathway as a search hook to identify a new *bmp* BGC found in the cyanobacterial sponge symbiont *Hormoscilla spongeliae.*[141] Functional characterization of this BGC revealed it was indeed responsible for production of PBDEs that proceeded through the same shikimate shunt.

Vibralactone

Vibralactone (**71**) is a rare bicyclic fused β-lactone terpenoid natural product produced by the mushroom *Boreostereum vibrans* (**Figure 1.25**).[142] The β-lactone

warhead of this molecule serves as the covalent attachment site for a key serine residue in the human pancreatic lipase, thereby inhibiting fat absorption.[142] ^{13}C labeling studies surprisingly suggested a shikimate origin for the bicyclic skeleton of vibralactone, despite its non-aromatic structure. The labeling patterns suggested that both phenylalanine and an earlier shikimate pathway compound may contribute to this structure, and so the authors hypothesized 4-HBA as a possible intermediate.

Later work demonstrated that VibMO1, an FAD-dependent monooxygenase and aromatic ring hydroxylase, can convert prenyl 4-HBA into preylhydroquinone, confirming a shikimate derived 4-HBA metabolite as a key intermediate in vibralactone production. However, the exact origin of the 4-HBA in vibralactone biosynthesis remains undetermined. It could arise from chorismate lyase activity as seen in the bacterial examples discussed here, or it could come from phenylalanine degradation. However, the genome responsible for vibralactone production has not been made public, and so this question remains unanswered.

1.12 Prephenate

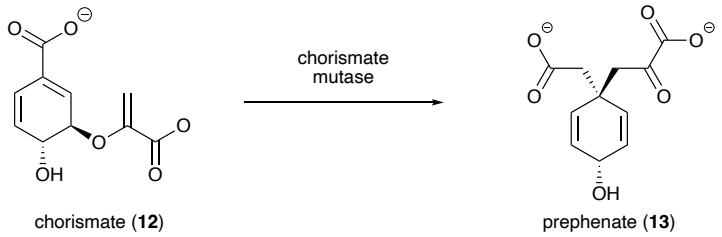


Figure 1.26 Sigmatropic rearrangement of chorismate to prephenate by chorismate mutase.

The conversion of chorismate to prephenate (**13**) by chorismate mutase (CM) is the first committed step towards phenylalanine and tyrosine biosynthesis (**Figure 1.26**).[143] Chorismate mutase performs a 3,3-sigmatropic rearrangement of chorismate, a

particularly notable reaction as it is a proven example of a naturally occurring pericyclic reaction.[144,145] Complex allosteric regulation of chorismate mutase by aromatic amino acids controls the flux of chorismic acid towards either dedicated phenylalanine and tyrosine biosynthesis catalyzed by CM, or tryptophan biosynthesis catalyzed by an anthranilate synthase enzyme. The compounds discussed in this section are those that utilize the entire shikimate pathway for biosynthetic precursor production, and only diverge away from amino acid biosynthesis after the formation of prephenate. The biosynthetic pathways discussed here use prephenate as a substrate, but catalyze a reaction different from those that generate phenylalanine and tyrosine and instead prime the molecule for incorporation into diverse secondary metabolites.

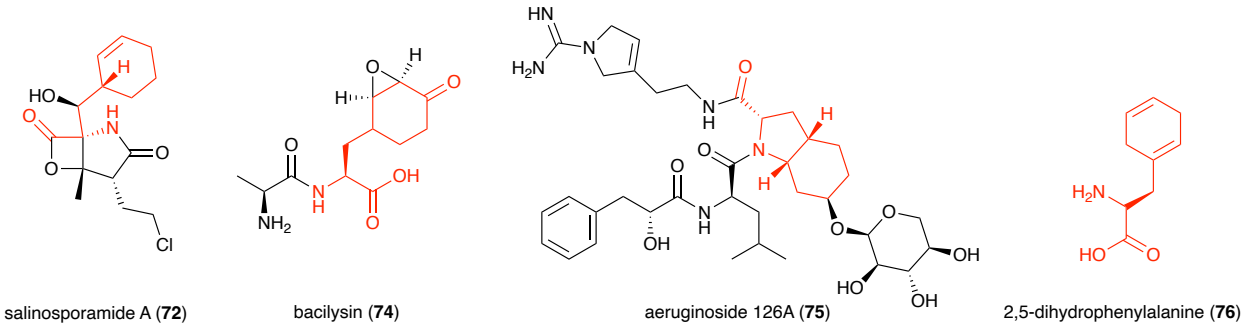


Figure 1.27 Natural products derived from prephenate

Salinosporamide A, bacilysin, and aeruginoside

Salinosporamide A (**72**) is an unusual hybrid polyketide (PK)/non-ribosomal peptide (NRP) natural product originally isolated from the obligate marine actinomycete *Salinispora tropica* (**Figure 1.27**).[146] A potent proteasome inhibitor, the molecule entered oncology clinical trials in 2006, just three years after its initial discovery.[147] Despite the discovery of numerous natural analogs of salinosporamide A, as well as efforts to

generate synthetic and mutasynthetic derivatives, it is the originally discovered natural product itself that has advanced through clinical trials, entering Phase III trials for the treatment of glioblastoma in late 2017.[148] Salinosporamide A selectively and irreversibly binds the β-subunit of the 20S proteasome, thereby inhibiting the cell's ability to control proteolysis.[149] This bioactivity is due to salinosporamide A's γ-lactam-β-lactone core, which appears perfectly designed for proteasome inhibition.[150]

A tremendous amount of work in the Moore lab has focused on understanding the biosynthesis of this molecule[150] beginning with stable isotope feeding experiments that revealed salinosporamide is comprised of three distinct pieces which correlate to the side chains surrounding the biscyclized core.[151] Surprisingly, only one acetate unit was incorporated into the molecule at C3/C14, but assimilation of U-$^{13}C_6$ glucose revealed both the chlorinated four carbon side chain and the cyclohexenyl amino acid moiety originated from tetrose via glycolysis. Asymmetric labeling of the six membered ring suggested a shikimate derived origin from a novel shunt in aromatic amino acid biosynthesis prior to aromatization. Further experiments revealed phenylalanine was not incorporated into salinosporamide A, but shikimate was. Subsequent identification of the salinosporamide BGC[152,153] and gene disruption experiments revealed that deletion of *salX* abolished salinosporamide production and allowed for chemical complementation by alternative amino acids.[154,155]

These labeling and gene disruption experiments directly linked salinosporamide A biosynthesis to the shikimate pathway, but the exact branch point was not definitively established until biochemical work by the Walsh and Moore labs demonstrated the role of SalX as a prephenate decarboxylase. While canonical phenylalanine and tyrosine

biosynthesis proceed through prephenate decarboxylation and aromatization of the cyclohexenyl ring, in vitro characterization of SalX demonstrated that enzyme utilizes prephenate as a substrate and performs a surprising nonaromatizing decarboxylation. SalX converts prephenate to endocyclic dihydro-4-hydroxyphenylpyruvate (H_2HPP, **73**), which can undergo spontaneous conversion to exocyclic H_2HPP, en route to cyclohexenylalanine formation (**Figure 1.28**). However, beyond the formation of H_2HPP, the rest of the steps to cyclohexenylalanine are proposed but remain unvalidated.

The non-aromatizing decarboxylation of prephenate to H_2HPP in salinosporamide biosynthesis was discovered due to the similarity of SalX to a protein found in the bacilysin biosynthetic pathway. Bacilysin (**74**) is a dipeptide antibiotic produced by *Bacillus subtilis* recognized by its signature epoxycyclohexanone amino acid named anticapsin (**Figure 1.27**).[156] Feeding experiments with radioisotope labeled precursors demonstrated that bacilysin was derived from an intermediate of the shikimate pathway prior to phenylalanine,[157] and subsequent gene knockout experiments confirmed prephenate as the likely intermediate.[158] Biochemical validation of the bacilysin gene cluster revealed a prephenate dehydratase homolog, BacA, is capable of performing a decarboxylation and protonation to yield the non-aromatic amino acid **73**.[156] Further in vitro assays have continued to elucidate this pathway, although the exact route to installation of the bioactive epoxide remains unknown.[159,160] Notably a third homologous enzyme, AerD, was identified in the BGC responsible for aeruginoside 126A (**75**) biosynthesis, and in vitro characterization proved its prephenate decarboxylase activity to generate H_2HPP.

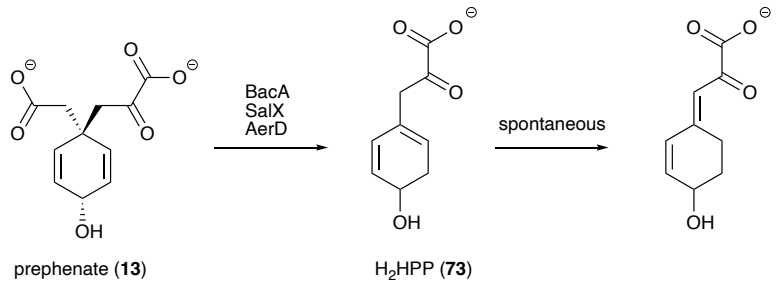


Figure 1.28 BacA, SalX, and AerD catalyzed nonaromatizing decarboxylation of prephenate to yield H2HPP, which can spontaneously isomerize to the exo-cyclic diene.

Dihydrophenylalanine

The novel non-aromatizing decarboxylation of prephenate by BacA, SalX, and AerD, enabled a genome mining approach to search for novel prephenate decarboxylases from other organisms. The Walsh lab chose to target the insect pathogen *Photorhabdus luminescens,* which they hypothesized might utilize a similar prephenate decaryxoylation reaction in the biosynthesis of dihydrophenylalanine (H_2Phe, **76**), a suspected intermediate en route to the antibiotic dihydrostilbene (**Figure 1.27**).[161] Labeled precursor feeding experiments demonstrated that H_2Phe is not derived from a reduction of phenylalanine, due to a lack of labeled Phe incorporation, which suggested an earlier intermediate of the shikimate pathway.[162]

Using this genome mining approach they were able to identify an eight gene cluster that they confirmed through genetic knockouts was responsible for H_2Phe production. In vitro biochemical experiments validated that a homologous prephenate decarboxylase like those found in the earlier bacterial pathways described above, could covert prephenate to non-aromatized H_2HPP. An aminotransferase encoded in the cluster is then able to transaminate H_2HPP, in the endocyclic form, to form H_2 Tyr using Gln as the

amino donor. However, the reactions to convert this compound to H_2Phe remain undiscovered.

1.13 Anthranilate

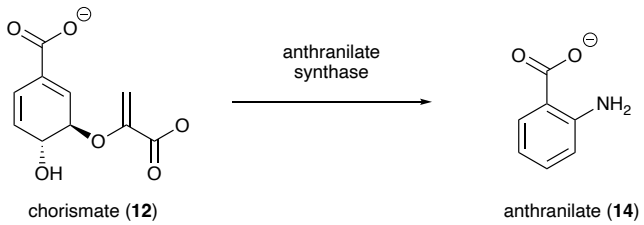


Figure 1.29 Amino transfer and pyruvate lyase reaction of chorismate to prephenate by anthranilate synthase

Chorismate is the central branch point of the shikimate pathway; it is the last shared intermediate in aromatic amino acid biosynthesis. Chorismate mutase, discussed above, directs biosynthesis towards the formation of phenylalanine and tyrosine through prephenate, but indole and tryptophan biosynthesis is initiated by anthranilate synthase which converts chorismic acid to anthranilate (**Figure 1.29**). Anthranilate synthase catalyzes the synthesis of anthranilate in two steps: a glutamine amidotransferase reaction that yields ADIC, followed by the irreversible elimination of pyruvate and conversion to anthranilate.[163,164] While anthranilate is a precursor towards tryptophan biosynthesis, occasionally it can be directly incorporated into secondary metabolites, most commonly in fungal peptides. Secondary metabolites that incorporate anthranilate directly into their biosynthesis pathways will be discussed here.

Fumiquinazoline A, acetylaszonalenin, and asperlicin

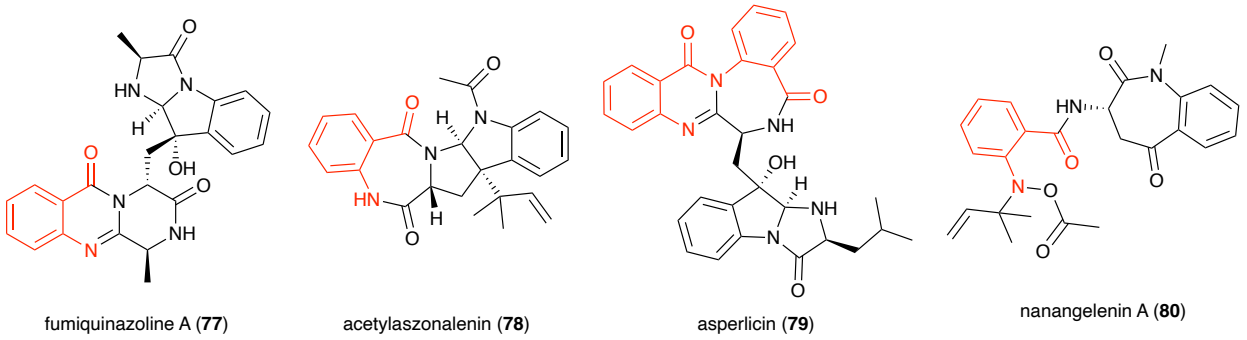


Figure 1.30 Fungal alkaloids derived from anthranilate

Fumiquinazoline (**77**), acetylaszonalenin (**78**), and asperlicin (**79**) are fungal alkaloids isolated from both marine and terrestrial environments with diverse bioactivities including cytotoxicity, toxicity, and cholecystokinin antagonism (**Figure 1.30**). The quinazoline and benzodiazepine rings found in fumiquinazoline, asperlicin and acetylaszonalenin are often referred to as “privileged structures” due to the high numbers of clinically relevant compounds that contain these features.[165,166] Notably, for all three of these structures, anthranilate appears to be a key building block in these privileged ring scaffolds.

Incorporation of anthranilate into fungal natural products is well documented and explored via radioisotope feeding experiments, but no secondary metabolite enzymes that incorporated anthranilate into complex natural product structures had been characterized. However, the gene cluster for acetylaszonalenin had been recently identified[167] and revealed an NPRS (AnaPS) capable of condensing anthranilic acid and tryptophan to yield the benzodiazepinedione core. The Walsh lab bioinformatically identified a module from this NRPS that was capable of selecting, activating, and loading

anthranilate and confirmed this activity via in vitro adenylation assays that demonstrated anthranilate activation.[168] The thiolation half reaction with anthranilate was additionally confirmed. While fungal adenylation domains do not conform to the Stachelhaus code recognized in bacterial systems, Ames and Walsh were able to identify a 10-residue sequence that conferred specificity for anthranilate. The predictive power of this specificity code was further demonstrated through genome mining and heterologous expression of two previously unidentified anthranilate activating NRPS modules from *Aspergillus fumigatus,* which may encode fumiquinazoline biosynthesis, and *Neosartorya fischeri.* In all cases examined in this work, anthranilate serves as the starter unit for assembly via a multimodular NRPS.

A similar anthranilate activating NRPS was recently identified in the BGC responsible for the production of nanangelenin A (**80**), a 1-benzazepine containing natural product produced by the fungi *Aspergillus nanangensis*.[169] While the 1-benzazepine structure is well-represented in clinically approved drugs, this feature is incredibly rare in nature, with just a three examples prior to nanangelenin A. The 1-benzazepine core was predicted to arise from condensation of anthranilate and the nonproteinogenic amino acid L-kynurenine. Based on this prediction, the presence of an anthranilate-activating AnaPS NRPS homolog helped identify the nanangelenin BGC. Co-expression of this NRPS *nanA* with the indolamine-2,3-dioxygenase *nanC* responsible for tryptophan catabolism to produce kynurenine, produced the anthranilate-kynurenine cyclized dipeptide. The formation of this rare 1-benzazepindione was ultimately revealed a regioselective lactamization catalyzed by a terminal condensation domain and aided by the terminal thiolation domain.

Tomaymycin

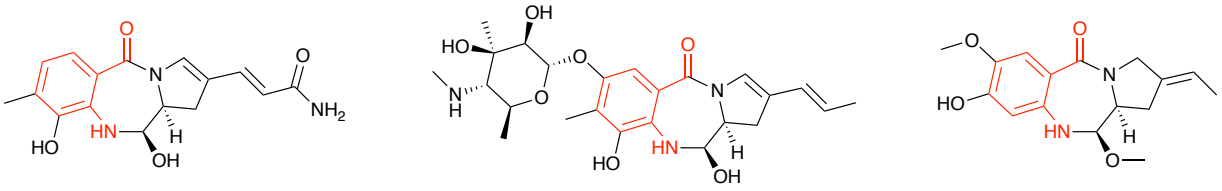


Figure 1.31 Pyrrolobenzodiazepine natural products derived from anthranilate

Pyrrolobenzodiazepines (PBDs) are tricyclic natural products produced by both terrestrial and marine microorganisms that display potent antitumor activity. These compounds bind the minor groove of DNA, forming crosslinks that result in cell death. However, despite this powerful bioactivity, the cardiotoxicity of this class of compounds has precluded their clinical development. The founding member of this class of natural products is anthramycin (**81**), isolated from the thermophilic *Streptomyces refuineus* sbsp. *thermotolerans,*[170] and named for its characteristic anthranilic acid moiety. Additional members of this class include sibiromycin (**82**), and tomaymycin (**83**) (**Figure 1.31**).

The extreme bioactivity of these compounds has led to a tremendous amount of research on the biosynthesis of these compounds. Initial isotope labeled precursor feeding studies demonstrated that 3-hydroxyanthranilic acid derived from tryptophan via the degradative kynurenine pathway is incorporated into the anthranilate moiety of sibiromycin and anthramycin.[171] Subsequent identification of the gene clusters responsible for sibiromycin and anthramycin biosynthesis revealed homologs of the kynurenine tryptophan degradation pathway enzymes, further implicating this route for anthranilate production and incorporation.[172,173] However, feeding experiments with labeled anthranilic acid showed incorporation only in tomaymycin and not in sibiromycin,

which potentially suggested different biosynthetic origins of the anthranilate moiety in these compounds.[174] For pyrrolobenzodiazepines not substituted at the C9 position, like tomaymycin, the authors suggested de novo synthesis of the anthranilate moiety instead of tryptophan degradation. Analysis of the tomaymycin BGC revealed a DAHP synthase, and two anthranilate synthetase homologs, and no genes related to the kynurenine pathway. The authors propose that anthranilic acid can be hydroxylated and methylated before incorporation by the NRPS assembly line proteins TomA and TomB that catalyzed diazepine ring formation. This work demonstrates that the pyrrolobenzodiazepines family of natural products have developed two distinct ways to access the same anthranilate moiety and incorporate it into their biosynthetic pathways. Sibiromcyin and anthramycin utilize the end point of the shikimate pathway, tryptophan, while tomaymycin steals an earlier intermediate, at the point of anthranilate, for secondary metabolite biosynthesis.

Tasikamides

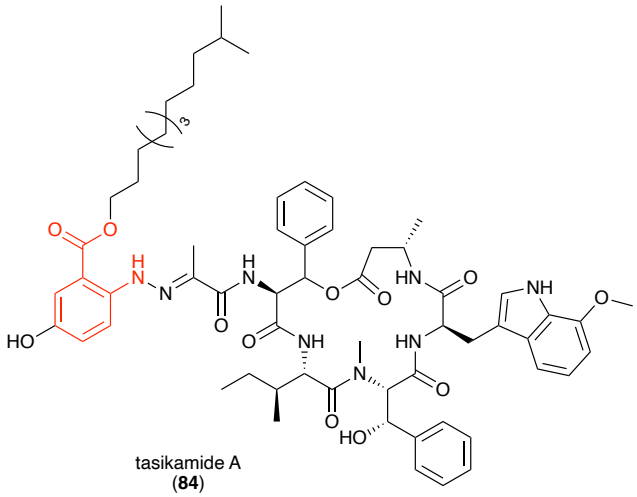
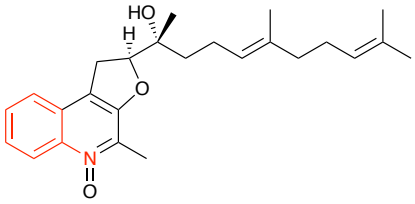


Figure 1.32 Tasikamide A, *Streptomyces* peptide containing an anthranilate derived AHA moiety, highlighted in red.

The tasikamides (tasikamide A, **84**) are a family of cyclic peptides produced by the fresh water sediment-dwelling *Streptomyces tasikensis* P46 (**Figure 1.32**). These compounds are notable for an unusual hydrazone feature rarely seen in natural products that connects an alkyl 5-hydroxylanthranilate (AHA) fragment to a 3-carbon extended cyclic pentapeptide. Whole genome sequencing and subsequent genome mining allowed the authors to identify a five module NRPS that matched the five amino acids present in the tasikamides as well as additional tailoring genes required for tasikamide functionalization. Gene disruption experiments of this cluster indeed confirmed it was required for tasikamide biosynthesis. However, the biosynthesis of the AHA portion of the molecule remained unaccounted for in this BGC. The authors suspected this moiety was derived from either a shikimate pathway shunt product, or via tryptophan degradation. Using these alternative pathways as genome mining hooks, they discovered a 25 kb BGC (*aha*) that contained a putative anthranilate synthase gene, a thioredoxin gene, and a flavin-dependent monooxygenase, all presumably involved in converting chorismate to 5-hydroxyl-anthranilate. Notably, this these genes were found elsewhere on the chromosome, and were not contiguous with the NRPS BGC. This smaller *aha* cluster also contained three genes not required for AHA biosynthesis, but that shared high sequence similarity to three genes from the cremeomycin biosynthetic pathway responsible for nitrite production and diazo formation. In vitro enzymatic assays revealed a similar system in the tasikamide pathway, resulting in the formation of diazo-AHA. Finally, an unusual Japp-Klingemann reaction results in the coupling of the cyclic peptide with diazo-AHA to generate the mature tasikamides. These molecules harbor a wealth of structurally intriguing features, and discovery of their biosynthetic pathway revealed a novel b-amino

acid, two separate BGCs that act in concert, and a new biosynthetic routes to C-N bond formation. Fundamental to this biosynthesis is the anthranilate derived AHA feature. The tasikamide pathway appears to utilize the shikimate pathway until the point of anthranilate, where it hijacks the molecule for production of AHA and downstream biosynthesis.

Aurachins



aurachin A (**85**)

Figure 1.33 Aurachin A, quinoline alkaloid derived from anthranilate

While incorporation of anthranilate is most often associated with fungal NRPS assembly line natural products, this unusual starter unit has also been observed in bacterial systems as well. Aurachin A (**85**) is a member of a family of unusual quinoline alkaloids produced by the myxobacteriym *Stigmatella aurantiaca* Sg a15 known to display mitochondrial respiration inhibitory activity (**Figure 1.33**).[175] Initial feeding experiments with ^{13}C and ^{18}O labeled anthranilic acid resulted in a tremendous increase in aurachin production, demonstrating the role of this compound as an early precursor in the biosynthetic pathway. A putative BGC for aurachin production was identified using a transposon mutagenesis strategy, revealing a five gene cluster encoding a type II PKS, the first identified from a gram negative bacterium.[176]

However, none of these candidate genes appeared capable of activating anthranilate, the presumed first step in aurachin biosynthesis.[177] A BLAST search using the benzoate-CoA ligase, *auaE,* from the aurachin BGC as a hook revealed a homologous gene located almost 8 kb upstream of the gene cluster. Targeted disruption of this gene, *auaEII,* completely abolished aurachin production. Both AuaE and AuaEII were heterologously expressed for in vitro biochemical assays to probe their respective biosynthetic roles. This work demonstrated that AuaEII is an anthranilate CoA ligase, proceeding through an AMP-activated anthranilate intermediate. AuaE is then responsible for transferring anthranilate onto the pathway's carrier protein, AuaB. While AuaE is part of the ANL superfamily of enzymes, it appears only capable of performing only the second transthioesterification half reaction typical of ANL enzymes. The activation of anthranilate in aurachin biosynthesis thus proceeds via an unusual priming mechanism, involving two ANL enzymes located in different genomic regions.

Recently, a series of structurally related prenylated quinoline natural products, the marinoterpins, were isolated by the Fenical lab from a marine actinomcyete.[178] By using the *aua* cluster as a genome mining hook, they were able to identify a putative BGC with a similar benzoate-CoA ligase, MrtE. Notably the putative marinoterpin BGC only contains one of these proteins, instead of the pair required for aurachin biosynthesis. However, the identification of this BGC demonstrates that the activation of anthranilate as a starter unit for bacterial PKS biosynthesis is widespread across bacterial families.

Acridone alkaloids and avenanthramides

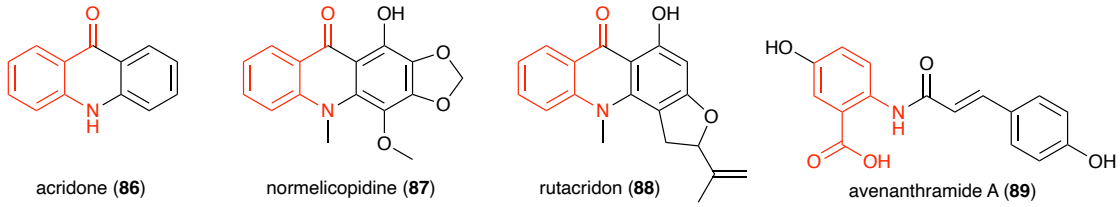


Figure 1.34 Plant natural products derived from anthranilate

While plant alkaloids are typically thought of as amino acid derived, a handful of these compound have been identified that utilize the earlier anthranilate precursor instead. For example, acridone alkaloids are a large family of tricyclic natural products (**86**-**89**), characterized by a shared 9-acridinone core, that are characteristic secondary metabolites of the Rutaceae family of flowering plants, particularly found in citrus (**Figure 1.34**).[179] Like phenazines, a wide variety of modifications to the core have been observed, particularly oxygenated substituents. An early biosynthetic hypothesis proposed these compounds arise from the condensation of anthranilate with three acetate units, perhaps by a PKS. Radioactive precursor feeding experiments in the 1960s indeed confirmed that anthranilate is incorporated into the left half, but not the right, of the molecule.[180] Later cell free extract work demonstrated that N-methylation of anthranilate was essential for the acridone biosynthetic pathway. In vitro biochemical experiments with recombinant anthranilate N-methyltransferase from *Ruta graveolens* demonstrated that this methylation is indeed the first committed step towards acridone biosynthesis.[181] In 2019, acridone biosynthesis was fully reconstituted in *E. coli* for the first time.[182]

Similarly, recent work has revealed anthranilate as a definitive precursor in the biosynthesis of the avenanthramides, antioxidant natural products produced by oat plants

that consist of anthranilic acid coupled to cinnamic acid via an amide bond. Three different types of genes from oat plants have been biochemically characterized including an hydroxycinnamoyl-CoA:hydroxyanthranilate N-hydroxycinnamoyl transferase (HHT) enzyme responsible for the anthranilate condensation reaction.[183,184]

1.14 Aminoshikimate and C_7N units

Many natural products originally thought to arise from early shikimate pathway intermediates have actually been shown to arise from alternative, but analogous, pathways, including the aminoshikimate pathway. The discovery of the aminoshikimate pathway has been the subject of previous reviews[185,186] and as such this section will only briefly discuss key pathways and enzymes known to date involved in the biosynthesis of ASA and 3,5-AHBA in the context of the archetype ansamycins, the rifamycins, the pathway from which many of the enzymes were first characterized (**Figure 1.35**).

The discovery of the aminoshikimate pathway has been the subject of previous reviews[187] and as such this section will only briefly discuss key pathways and enzymes known to date involved in the biosynthesis of ASA and 3,5-AHBA in the context of the archetype ansamycins, the rifamycins, the pathway from which many of the enzymes were first characterized.

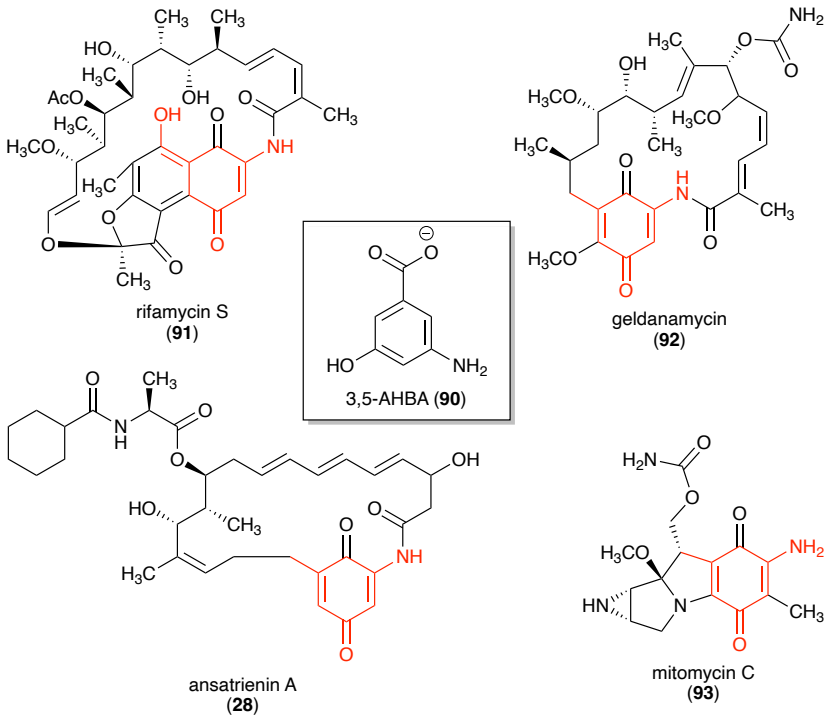


Figure 1.35 3,5-AHBA containing natural products with the AHBA core colored red

Biosynthetic logic suggested 3-amino-5-hydroxybenoic acid (3,5-AHBA, **90**), a known naturally occurring amino acid, may be the source of the naphthalenic core of the ansamycins and which was confirmed in feeding studies[188] with stable isotope labeled 3,5-AHBA as well as via knockout and chemical complementation. However, in contrast with biosynthetic studies of anthranilate, feeding experiments with stable isotope labeled shikimic acid (**4**), quinic acid, and 3-dehydroquinate (**8**) with rifamycin (**91**) and mitomycin (**93**) producer *Amycolatopsis mediterranei* failed to result in isotope enriched metabolites, suggesting that the branch point for the production of the mC_7N core must be prior to DHQ. These results initially lead to hypotheses of crosstalk between observed aminoshikimate and shikimate intermediates and enzymes. Ultimately, however, additional studies with *A. mediterranei* revealed a completely independent

aminoshikimate biosynthetic pathway containing a parallel set of enzymes to those found in shikimate biosynthesis (**Figure 1.36**).

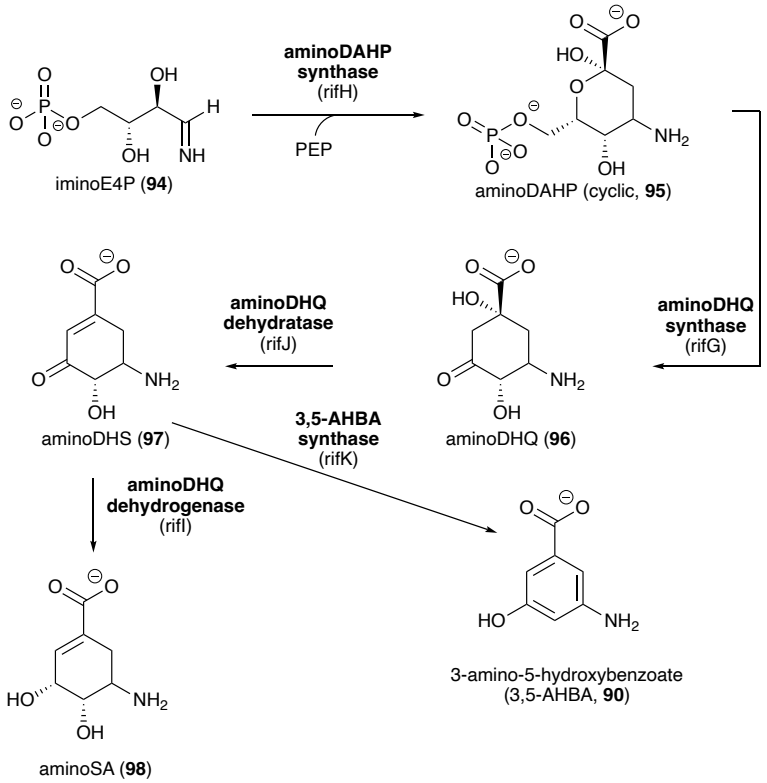


Figure 1.36 The aminoshikimate/3,5-AHBA biosynthetic pathway with representative enzymes from the *A. mediterranei* rifamycin biosynthetic gene cluster.

In short,[189] functional DAHPS homolog, aminoDAHP synthase (rifH) catalyzes the imino-aldol reaction between iminoE4P (**94**, derived from kanosamine) with PEP to give aminoDAHP (**95**). This is followed by two additional steps which mirror those from shikimate biosynthesis with aminoDHQ synthase (rifG) converting **95** to aminoDHQ (**96**), which undergoes dehydration to aminoDHS (**97**) via aminoDHQ dehydratase (rifJ). AminoDHS can then either be trafficked to aminoshikimate (**98**) via aminoDHS dehydrogenase (rifI) again echoing shikimate biosynthesis or can be converted to mC_7N precursor 3,5-AHBA (**90**) via AHBA synthase (rifK) which has no representative

homologue in shikimate assembly. Notably, aminoshikimic acid itself does not serve as an intermediate in the assembly of 3,5-AHBA, however this stereochemically rich scaffold has been proposed to be valuable as a precursor for the construction of antiviral pharmaceuticals such as oseltamivir phosphate (Tamiflu).[190]

Kanglemycins

Elucidation of the core 3,5-AHBA biosynthetic genes not only resolved the key components for the assembly of this scaffold prevalent in a wide variety of natural products, but also enabled their use as "hooks" for genome mining to identify new 3,5-AHBA containing natural products. Despite the potent biological activity of the rifamycins as first in class antibiotics in treatment tuberculosis, there is a concerning increase in the clinical isolates that have demonstrable resistance to approved rifamycins such as rifampicin. Brady and coworkers hypothesized that mutations to bacterial RNA polymerase (RNAP) that give rise to rifamycin resistance in the clinic may also be present in the competitive soil microbiome, and as such evolutionary pressure may have given rise to new biosynthetic gene clusters that produce rifamycin congeners to overcome this resistance.[191] The Brady lab capitalized on the discrete phylogenetic clustering of 3,5-AHBA synthase genes with their corresponding metabolites and leveraged this as a strategy for efficiently identifying rifamycin analogues from complex samples. A degenerate primer approach was used to screen for rifamycin-type 3,5-AHBA synthase genes across "approximately 1500 geographically and ecologically diverse soils" the results of which showed that approximately half of those samples tested possessed rifamycin-type 3,5-AHBA synthase genes. A family of these putative BGCs was prioritized given its diversity of tailoring enzymes as well as an altered predicted selectivity of the

acyltransferase domain of the eighth polyketide synthase module from methylmalonyl-CoA to ethylmalonyl-CoA, and an isolated strain of *Amycolatopsis vancoresmycina* possessing such a BGC was selected for further evaluation as a potential producer of rifamycin analogues. Fermentation of *A. vancoresmycina* resulted in the isolation of known rifamycin congener kanglemycin A (**99**) (originally isolated from *Amycolatopsis mediterranei* var. *kanglensis*, with no characterized BGC associated with its biosynthesis) as well as new kanglemycins, V1 (**100**) which varies from kanglemycin A as the C11 ketone is reduced to an alcohol (stereochemistry undefined) and kanglemycin V2 (**101**) wherein the C11 alcohol forms a methylenedideoxy bridge with C4 hydroquinone phenol (**Figure 1.37**). These new kanglemycins demonstrated improved activity in vivo and in vitro against pathogenic Gram-positive bacteria harboring clinically relevant RNAP mutations associated with rifamycin resistance. Detailed analysis of transcription run off assays suggest that kanglemycins possessed a mechanistically distinct mode of action compared with rifamycin, and high-resolution crystal structures of kanglemycins in complex with RNAP demonstrated the kanglemycins possessed additional points of contact with RNAP which may assist in overcoming loss of contact caused by mutations known to confer rifamycin resistance. While significant effort in genome mining is centered on discovery of structurally novel and uncharacterized classes of natural products, this work demonstrates the utility of using the core 3,5-AHBA biosynthetic genes as hooks to identify natural products of known structural classes with improved efficacy towards pathogenic bacterial with clinically relevant antibacterial resistance.

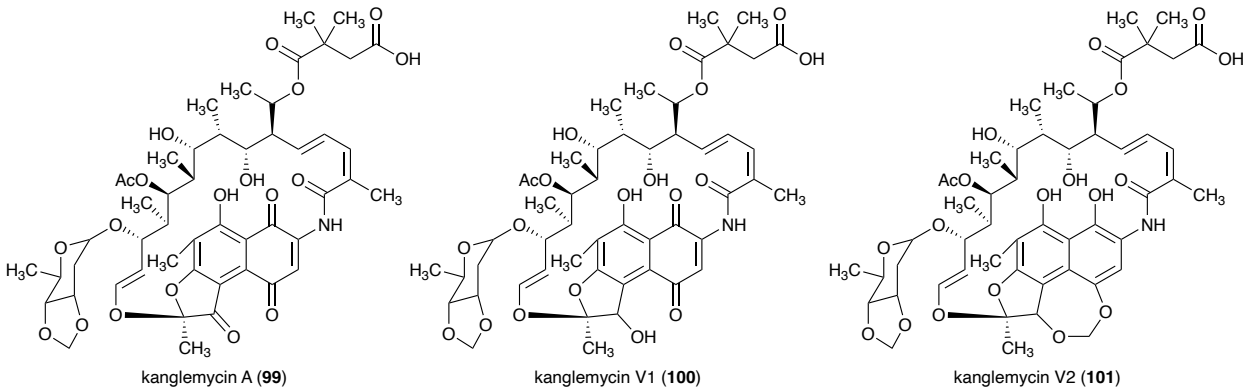


Figure 1.37 Structures of kanglemycins A (**99**), V1 (**100**), and V2 (**101**).

Saliniketals

The saliniketals A (**102**) and B (**103**) were originally isolated from *Salinispora arenicola* by Fenical and coworkers as part of a campaign to isolate inhibitors of ornithine decarboxylase, a putative target for cancer prevention, from marine microbes (**Figure 1.38**).[192] In addition to potent bioactivity, the polyketide structure of the saliniketals posed a unique biosynthetic question as it possessed identical functionalization and stereochemistry as the *ansa* chain of the rifamycins which are also produced by *S. arenicola* (**Figure 1.38**). Given the rifamycins starter unit is aromatic 3,5-AHBA, which is clearly not present in the saliniketals, and the lack of biosynthetic precedent at that time for cleavage of the aromatic C–N bond to generate the primary amide of the saliniketals, the authors proposed the saliniketals were not a shunt product of the rifamycin biosynthetic gene cluster, may be the result of a disparate polyketide synthase.

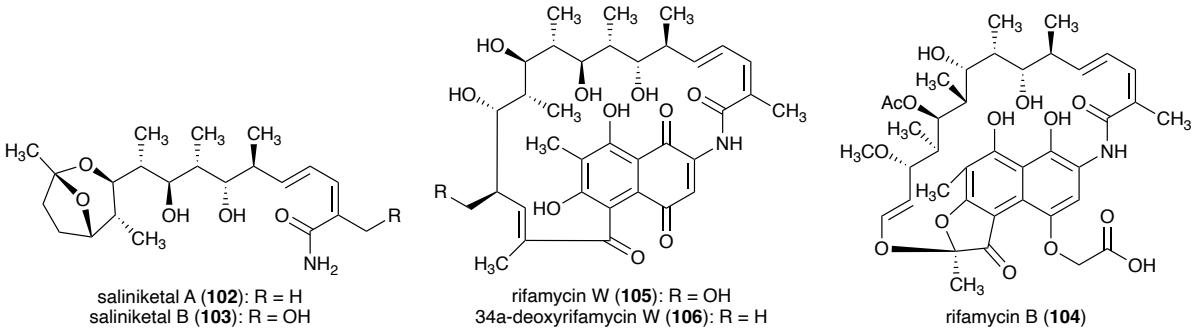


Figure 1.38 Structural similarities between the saliniketals A (**102**) and B(**103**) and rifamycin B (**104**), rifamycin W (**105**) and 34a-deoxyrifamycin W (**106**).

Upon sequencing and annotation of the *S. arenicola* genome, a rifamycin biosynthetic gene cluster was identified (SA-rif) along with an additional 12 PKS-driven BGCs, however none possessed a domain architecture consistent with the assembly of the saliniketals.[193] To demonstrate the shared biosynthetic origin of the saliniketals and rifamycins the 3,5-AHBA synthase gene, *rifK*, from the SA-rif BGC was disrupted which abrogated the production of both metabolites by the *S. arenicola* mutant strain. Chemical complementation of this mutant with [^{15}N]3,5-AHBA resulted in isotopically enriched saliniketals, confirming the origin of the primary amide as the AHBA nitrogen and that cleavage of the aromatic C–N bond in the rifamycins is an on-pathway event in the biosynthesis of the saliniketals. The rifK mutant also enabled feeding of advanced rifamycin biosynthetic intermediates to decipher where in the shared biosynthetic pathways the rifamycins and saliniketals diverge. Upon feeding it was revealed that only intermediates preceding oxidation and polyketide rearrangement (rifamycins W (**105**) and 34a-deoxyrifamycin W(**106**)) were able to rescue saliniketal production. Further disruption of a cytochrome P450 *sare1259* also resulted exclusively in accumulation of these non-rearranged rifamycins, and complete loss of production of mature, rearranged rifamycins

as well as saliniketals implicating *sare1259* as a multifunctional P450 catalyzing multiple reactions involved in late–stage C–C bond cleavage and oxidative rearrangement reactions in the biosynthesis of both the saliniketals and the rifamycins. Notably compounds that had undergone oxidative cleavage of the ansa chain to give the bicyclic ketal architecture of the saliniketals but with the amide in-tact, such as salinisporamicin, had been isolated intimating amide bond cleavage as the ultimate step in saliniketal biosynthesis. At the time, the authors noted that no knock outs of rif tailoring genes were able to support conversion of a salinisporamicin to a saliniketal and that "if this reaction is indeed catalyzed by a dedicated enzyme it may be encoded outside of the *rif* cluster."

Recently, Wright and coworkers discovered a novel and mechanism for resistance against rifamycin and rifamycin derivatives widespread among bacteria wherein the aromatic C–N bond is oxidatively cleaved by flavin monooxygenases dubbed, "Rox" enzymes (**Figure 1.39**).[194] Bioinformatic investigation into rifamycin producers revealed the presence of Rox enzymes outside the bounds of their respective *rif* clusters, and in vitro characterization of these enzymes revealed they are not only capable of deactivating rifamycins via oxidative C–N bond cleavage but are also able to convert 16-demethyl salinisporamycin (**107**) into saliniketals, demonstrating these enzymes play a dual role in self resistance against rifamycins and catalyzing the final step in saliniketal biosynthesis.[195]

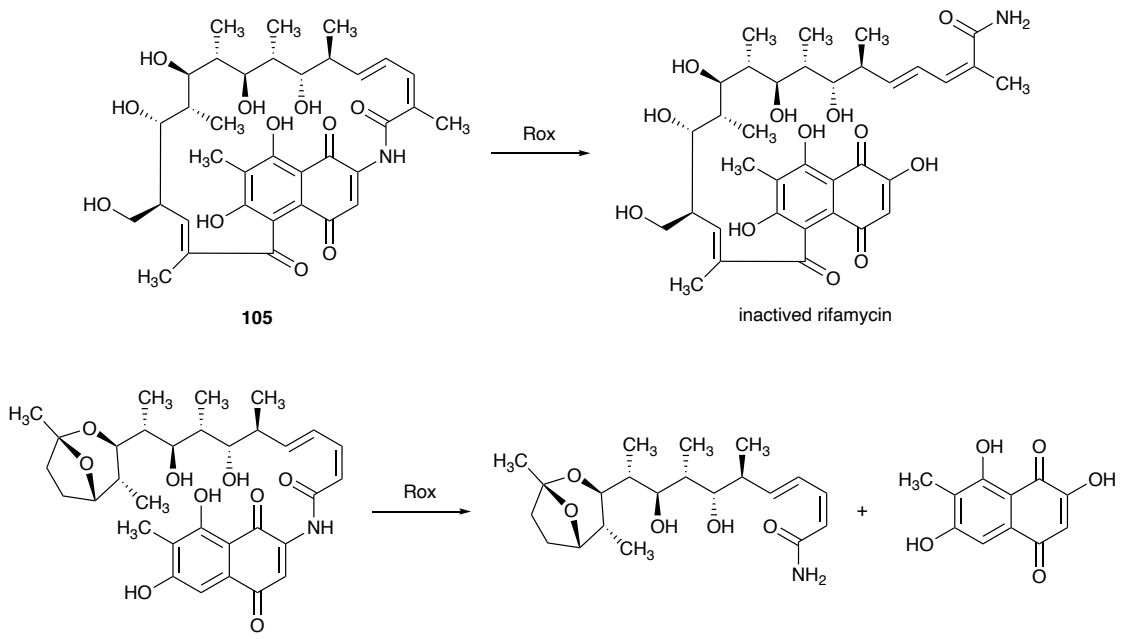


Figure 1.39 Rox enzymes catalyze C–N bond cleavage to generate deactivated rifamycins as well as saliniketals.

3-amino-4-hydroxybenzoic acid (3,4-AHBA)

The manumycin-type antibiotics are a large family of polyketide natural products defined by a central "*m*-C_7N unit" core from which branches a "lower chain" conjugated to a conserved 2-amino-3-hydroxycyclopentenone "C_5N" unit, and an upper chain which varies in branching, double bond position, and ketide extender and starter unit composition (**Figure 1.40**). The structure and classification of manumycin-type polyketides has been described in depth previously,[196] and structurally unique compounds continue to be isolated.[197] The eponymous manumycin A (**109**) was first isolated in 1963 from *Streptomyces parvulus* Tü 64 and the structure elucidated a decade later[198]. The related manumycin-type compound, asukamycin (**29**), was subsequently isolated in 1976[199] from *Streptomyces nodosus* subsp. *asukaensis* as an inhibitor of gram positive *Nocardia asteroides* and its structure was determined in 1979. Manumycin A,

asukamycin and their producing organisms served as invaluable models for the studying the biosynthetic origin of the structurally curious *m*-C_7N and C_5N units.

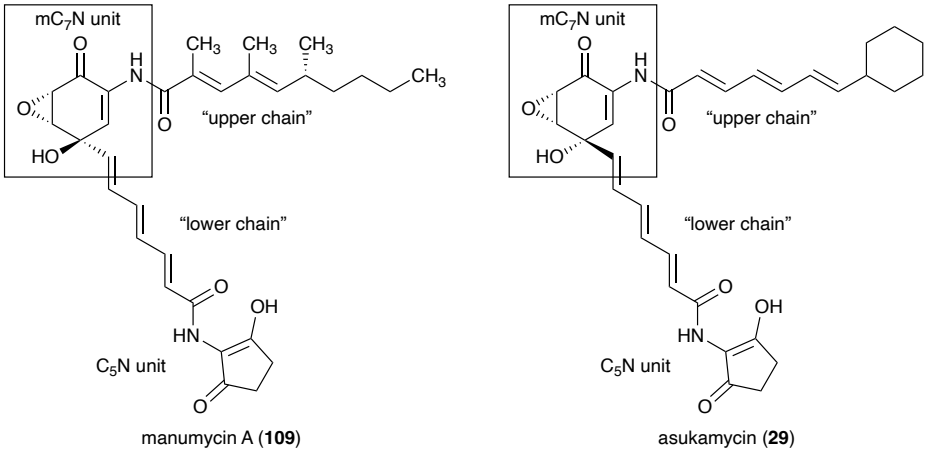


Figure 1.40 Manumycin A and asukamycin and their characteristic structural features.

In isolation reports the authors discuss the structural relation between the *m*-C_7N core of manumycin A and asukamycin and the amino-quinone core of the ansamycins (at the time hypothesized to be derived from the shikimate pathway) and suggested the *m*-C_7N unit in manumycin-type natural products as possibly originating from intermediates in the shikimate biosynthetic pathway as well.[200] Unlike the ansamycins, feeding studies with isotopically labeled $^{13}C_7$-3-amino-5-hydroxybenzoic acid showed no incorporation into either asukamycin nor manumycin, and examination of coupling patterns from asukamycin and manumycin isolated post feeding [U-$^{13}C_3$]glycerol ruled out the possibility of the *m*-C_7N unit originating from shikimate.[201] Incorporation of labeled four carbon dicarboxylic acids, such as succinic acid, into the *m*-C_7N unit pointed to an alternative mode of assembly.[202] This same labelling pattern was observed in feeding studies for 4-hydroxy-3-nitrosobenzamide, a siderophore from *Streptomyces murayamaensis*.[203]

Feeding ^{13}C-labelled 3-amino-4-hydroxybenzoic acid (3,4-AHBA, **110**), to *S. murayamaensis* resulted in labelled 4-hydroxy-3-nitrosobenzamide, and feeding labelled **110** to asukamycin producer *S. nodosus* subsp. *asukaensis* resulted in the production of asukamycin with a ^{13}C-enriched *m*-C_7N unit, demonstrating 3,4-AHBA as a metabolic precursor to the *m*-C_7N core.

Grixazone Biosynthesis

The first biochemical characterization of the enzymes involved in the assembly of 3,4-AHBA was determined through heterologous expression of enzymes from the grixazone biosynthetic gene cluster. Horinouchi and coworkers isolated diffusible yellow pigments, the grixazones A (**111**) and B (**112**), from *Streptomyces griseus* and found they possess the same phenoxazine chromophore as known metabolites michigazone and exfoliazone.[204] Tyrosinase-like binuclear type-III copper proteins had been previously identified to generate phenoxazines in antimycin producer *Streptomyces antibioticus*[205] and this protein class was used as a hook to identify a putative grixazone gene cluster (*gri*).[206] Knockouts of copper protein GriF in *S. griseus* led to accumulation of 3-amino-4-hydroxybenzaldehyde (3,4-AHBAL,) which upon incubation with recombinant GriF and N-acetylcysteine generated **111**, demonstrating 3,4-AHBAL as a direct precursor to the grixazones. The authors also noted that unpublished work, later published, demonstrated 3,4-AHBAL was a product of reduction of 3,4-AHBA by GriCD. Futher subcloning of the *gri* BGC demonstrated to coexpression of griHI in both a non grixazone producing *S. griseus* and *E. coli* led to the production 3,4-AHBA, suggesting that unlike the aminoshikimate pathway which requires a large suite of enzymes production of 3,5-

AHBA, these two enzymes alone may be able to generate the 3,4-AHBA core. Based on feeding studies in the manumycins/asukamycin and the recently discovered alternative route to DHQ (**8**) found in achaea, *Methanocaldococcus jannaschii*, the authors intuited the substrates of GriHI as dihydroxyacetone phosphate (DHAP, **113**) and L-aspartate semialdehyde (ASA, **114**). Among a variety of C3 and C4 building blocks tested, production of 3,4-AHBA was observed exclusively upon incubation of DHAP and ASA with recombinant GriH and GriI. GriI catalyzed aldol addition of DHAP and ASA to give 2-amino-4,5-dihydroxy-6-one-heptanoic acid-7-phosphate (**115**) which was then converted to 3,4-AHBA by Mn^{2+}-dependent, GriH (**Figure 1.41**). This was the first report identifying and characterizing and characterization the enzymatic dyad required for the assembly of the 3,4-AHBA core found in a wide variety of natural products, and retrospective analysis revealed GriH and GriI homologues in their corresponding BGCs and opened the door to using these enzymes as hooks for genome mining.

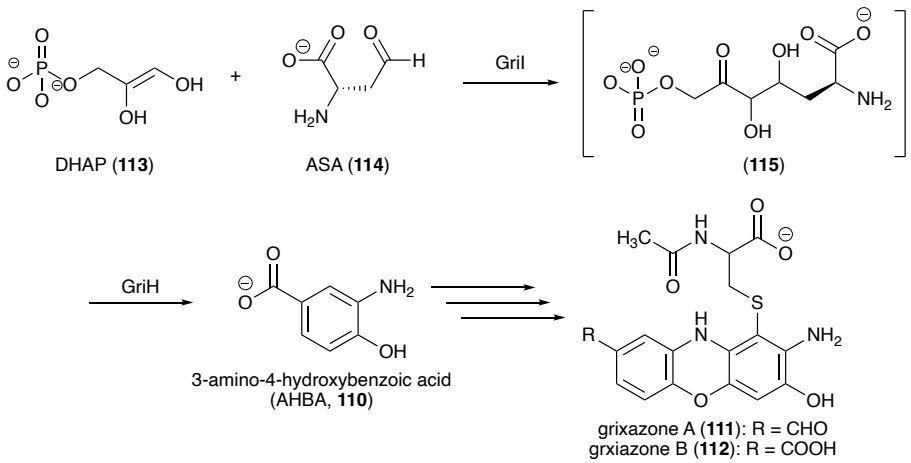


Figure 1.41 Reactions catalyzed by 3,4-AHBA biosynthetic enzymes GriI and GriH in grixazone biosynthesis.

Platencin and platensimycin

Platencin (**116**) and platensimycin (**117**) are a structurally unique class of antibiotic natural products originally isolated from *Streptomyces platensis* that target the bacterial type II fatty acid synthases which catalyze the assembly of essential fatty acids that comprise the cell bacterial cell membranes (**Figure 1.42**). Retrosynthetic analysis of these two compounds suggests they are of mixed origin, and disconnection at their central amide bond reveals a shared 3-amino-2,4-dihydoxybenzoic acid half and an aliphatic half of either terpenoid or polyketide origin. The intricate biosynthesis of the platesimycins and platencins have been the subject of intense study since their discovery in 2006 and 2007 respectively and has been recently summarized.[207] However, prior to the identification of the requisite biosynthetic genes and enzymes responsible for their assembly, the ability to generate significant quantities of these materials as drug leads was in high demand and was the subject of a suite of elegant total syntheses. While not directly used for traditional "genome mining" the use of the recently discovered 3,4-AHBA biosynthetic genes as "hooks" played a significant role in generating a biological platform for production of these important antibiotics.[208]

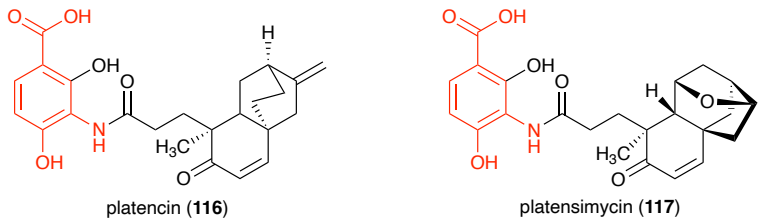


Figure 1.42 Platensin (**116**) and platensimycin (**117**) with AHBA derived cores highlighted in red.

The original strains identified to produce platensimycin (*S. platensis* MA7327) and platencin (*S. platensis* MA7339) produced these metabolites at a meager 2-4 mg/liter and 1 mg/liter respectively. While optimization of fermentative conditions yielded a

platensimycin production up to 56 mg/liter, at the time there was no reports of improvement of platencin production, leading Shen and coworkers to develop a system to genetically manipulate the production of these compounds. Using *S. platensis* MA7327, a producer of both platensimycin and platencin, degenerate PCR primers were used to verify the presence of a 3,4-AHBA synthase gene, which upon sequence revealed a single sequence suggesting *S. platensis* MA7327 had only a single copy of these gene. Screening cosmid libraries for cosmids possessing AHBA synthase by PCR and preliminary sequencing of neighboring genomic regions revealed putative pathway regulator *ptmR1*. This regulator was replaced with an apramycin resistance cassette via conjugation and homologous recombination. Fermentation and analysis of crude extracts from two confirmed exconjugants revealed them to be staggering super producers with one exconjugant producing platensimycin at yields of 323 ± 29 mg/liter and the second platencin at 255 ± 30 mg/liter. This approach demonstrates the incredible power of using biosynthetic logic to rationally target loci for genetic manipulation and strain improvement in the absence of a complete understanding of the biosynthesis of a given molecule.

Cremeomycin

Cremeomycin (**118**) was originally isolated from *Streptomyces cremeus* and patented in 1967 by Upjohn as a photodegradable antibiotic. 28 years later structure elucidation of cremeomycin revealed it to be a member of the rare family of diazo-containing natural products, with the authors proposing the 3,4-AHBA as a likely intermediate in cremeomycin biosynthesis.[209] Balskus and coworkers were able to validate this proposal and were able to identify the presence of a 3,4-AHBA synthase in

S. cremeus using degenerate primers with the amplicon showing high sequence identity to *griH*.[210] This *griH* homolog was used to mine the genome of *S. cremeus* and a suite of enzymes with activities in line with cremeomycin assembly were found in close proximity. This gene cluster (*cre*) was authenticated through heterologous expression in *S. lividans*, and through both stable isotope feeding experiments and *in vitro* characterization revealed the enzymes involved and order of enzymatic steps up to diazotization (**Figure 1.43**). Further characterization by Ohnishi and coworkers revealed L-aspartic acid (**119**) to be a likely source of the second nitrogen in the diazo functionality, as clustered flavin monooxygenase CreE catalyzed the six electron oxidation of L-aspartic acid to nitrosuccinic acid (**120**) which lyase CreD cleaved to fumarate and nitrous acid, a classical reagent used in chemical diazotization.[211] Finally, the full biosynthetic pathway was elucidated upon identification of CreM as the diazatizing enzyme, catalyzing the ATP-dependent convergence of nitrous acid and previously identified penultimate intermediate **121** into cremeomycin.[212]

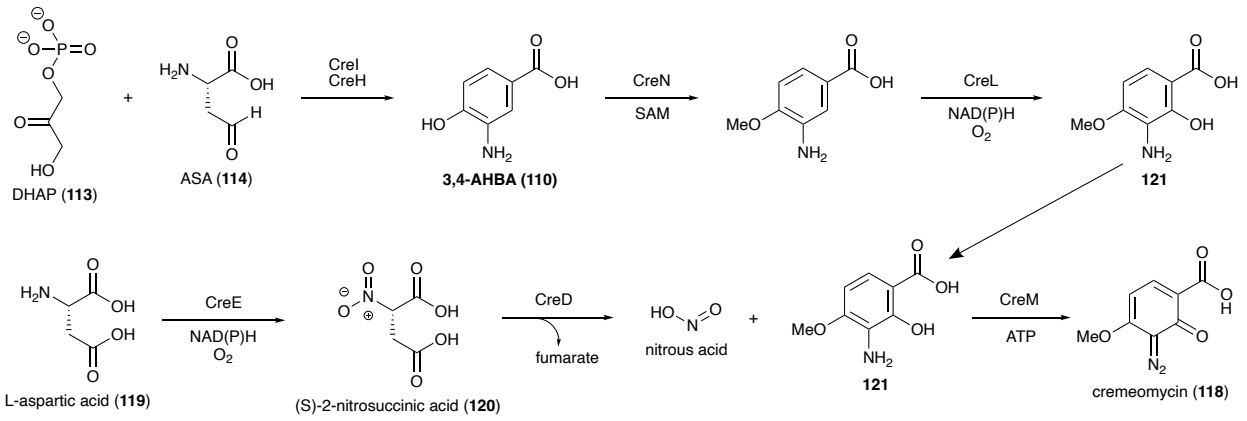


Figure 1.43 Biosynthesis of cremeomycin (**118**)

1.15 Perspectives and outlook

Aromatic amino acid biosynthesis proceeds through the shikimate pathway, seven enzymatic steps that begins with the condensation of PEP and E4P culminates in the production of chorismate, the last common intermediate between phenylalanine, tyrosine, and tryptophan. The biosynthesis of the aromatic amino acids is extraordinarily metabolically expensive,[213] more so than any of the other proteinogenic amino acids,[214] and so therefore are used relatively less frequently in bacterial proteomes. Perhaps because of this high metabolic costs associated with their synthesis, aromatic amino acids are incredibly valuable for secondary metabolism as well. These amino acids are widely incorporated into a variety of nonribosomal peptides, ribosomally synthesized and post-translationally modified peptides, and alkaloids. Often, particularly in the case of neuroactive tryptophan-derived alkaloids, it is these aromatic amino acid moieties themselves that are responsible for the bioactivity of these compounds.

However, as the work highlighted here demonstrates, earlier intermediates from the shikimate pathway can be major drivers of metabolic diversity for secondary metabolism as well. The densely functionalized intermediates between DAHP and chorismate can be diverted away from the shikimate pathway, stolen by biosynthetic enzymes, often through particularly interesting enzymatic reactions to yield a variety of natural products. The tremendous diversity of biosynthetic classes that utilize shikimate derived intermediates is also noticeable here. These compounds end up in peptides, polyketides, alkaloids, terpenes, and hybrid compounds of all of the above.

Notably, for almost all of the biosynthetic pathways described in this article, the shikimate derived portion of the molecule serves as the starter unit upon which additional

biosynthetic enzymes act to construct a mature molecule. This phenomenon perfectly exemplifies the complex interplay between primary and secondary metabolism, and showcases the shikimate pathway's role in both types of metabolism. These shikimate intermediates are derailed from their primary metabolic route and become the base on which on a complex and fully functionalized natural product is built. In fact, these compounds are so valuable in secondary metabolism that some organisms have developed a similar pathway, dubbed "the aminoshikimate pathway" for the biosynthesis of aminovariants of the pathway, all of which end up in secondary metabolism.

Because of how valuable and widely used shikimate intermediates are in secondary metabolism, they are a worthy target for future genome mining efforts. Many of the biosynthetic enzymes that perform the first dedicated biosynthetic step on a shikimate intermediate bear structural or sequence similarity to the shikimate pathway enzyme that typically utilizes that substrate. Perhaps using these enzymes as a biosynthetic search hook might be useful as a genome mining tool that could lead to the discovery of new compounds. Canonically, the genes used as genome mining hooks are those particularly distinctive of secondary metabolism, i.e., PKS or NRPS encoding genes. However, this also means the class of molecule discovered via this approach is predetermined, inherently biasing the search away from chemical novelty. Instead, using an enzyme from primary metabolism as a search hook and loosening the stringency for homologs, could result in the discovery of unknown molecular classes that incorporates a shikimate derived feature.

Despite the diverse pathways and enzymatic reactions discussed here, we were unable to identify any validated biosynthetic pathways that utilize the shikimate

intermediates shikimate-3-phosphate and EPSP. Considering the biosynthetic diversity produced from other intermediates, and the variety of enzymatic reactions at play, it is likely that these intermediates too are capable of entering secondary metabolism. Perhaps using shikimate kinase and EPSP synthase enzymes as genome mining hooks might reveal new pathways with homologs of these enzymes, exclusively used for secondary metabolite production

In preparing this review, we noticed that, not as a rule but as general practice, it seems that bacteria tend to utilize shikimate intermediates from the beginning of the pathway while fungi use intermediates from the end. For example, anthranilate is a relatively common in fungal natural products, but used much less frequently in bacterial systems. We've observed that higher organisms seem to allow the biosynthetic pathway to proceed almost to completion, while bacteria tend to derivative intermediates from earlier in the pathway, and with higher frequency. In summary, the shikimate pathway is a key driver of metabolic diversity in plants, bacteria, and fungi, both from the main pathway itself via aromatic amino acids, and from intermediary branch points.

1.16 References

1. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (W. H. Freeman, 2017).

2. Bochkov, D. V., Sysolyatin, S. V., Kalashnikov, A. I. & Surmacheva, I. A. Shikimic acid: review of its analytical, isolation, and purification techniques from plant and microbial sources. *J. Chem. Biol*. **5**, 5–17 (2011).

3. Herrmann, K. M. & Weaver, L. M. The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 473–503 (1999).

4. Yokoyama, R., de Oliveira, M. V. V., Kleven, B. & Maeda, H. A. The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation. *Plant Cell* **33**, 671–696 (2021).

5. Tohge, T., Watanabe, M., Hoefgen, R. & Fernie, A. R. Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* **4**, 62 (2013).

6. Tzin, V. & Galili, G. The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana. Arabidopsis Book* **8**, e0132 (2010).

7. Maeda, H. & Dudareva, N. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol*. **63**, 73–105 (2012).

8. Arora Verasztó, H., Logotheti, M., Albrecht, R., Leitner, A., Zhu, H. & Hartmann, M. D. Architecture and functional dynamics of the pentafunctional AROM complex. *Nat. Chem. Biol.* **16**, 973–978 (2020).

9. Steinrücken, H. C. & Amrhein, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase*. Biochem. Biophys. Res. Commun.* **94**, 1207–1212 (1980).

10. Funke, T., Han, H., Healy-Fried, M. L., Fischer, M. & Schönbrunn, E. Molecular basis for the herbicide resistance of Roundup Ready crops. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13010–13015 (2006).

11. Floss, H. G. Natural products derived from unusual variants of the shikimate pathway. *Nat. Prod. Rep*. **14**, 433–452 (1997).

12. Srinivasan, P. R. & Sprinson, D. B. 2-Keto-3-deoxy-d-arabo-heptonic acid 7-phosphate synthetase. *J. Biol. Chem*. **234**, 716–722 (1959).

13. Stephens, C. M. & Bauerle, R. Analysis of the metal requirement of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. *J. Biol. Chem.* **266**, 20810–20817 (1991).

14. Shumilin, I. A., Kretsinger, R. H. & Bauerle, R. H. Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli. Structure* **7**, 865–875 (1999).

15. Light, S. H. & Anderson, W. F. The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis: the diverse mechanisms of DAHPS allostery. *Protein Sci.* **22**, 395–404 (2013).

16. Gosset, G., Bonner, C. A. & Jensen, R. A. Microbial origin of plant-type 2-keto-3-deoxy-d-arabino-heptulosonate 7-phosphate synthases, exemplified by the chorismate and tryptophan-regulated enzyme from *Xanthomonas campestris. J. Bacteriol*. **183**, 10 (2001).

17. Choi, Y. S., Johannes, T. W., Simurdiak, M., Shao, Z., Lu, H. & Zhao, H. Cloning and heterologous expression of the spectinabilin biosynthetic gene cluster from *Streptomyces spectabilis. Mol. Biosyst*. **6**, 336–338 (2010).

18. Shawky, R. M., Puk, O., Wietzorrek, A., Pelzer, S., Takano, E., Wohlleben, W. & Stegmann, E. The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains *bbr*, encoding a StrR-like pathway-specific regulator. *Microb Physiol*. **13,** 76–88 (2007).

19. Rotenberg, S. L. & Sprinson, D. B. Mechanism and stereochemistry of 5-dehydroquinate synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1669–1672 (1970).

20. Bender, S. L., Mehdi, S. & Knowles, J. R. Dehydroquinate synthase: the role of divalent metal cations and of nicotinamide adenine dinucleotide in catalysis. *Biochem*. **28**, 7555–7560 (1989).

21. Carpenter, E. P., Hawkins, A. R., Frost, J. W. & Brown, K. A. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. *Nature* **394**, 299–302 (1998).

22. Favre-Bonvin, J., Bernillion, J., Salin, N. & Arpin, N. Biosynthesis of mycosporines: Mycosporine glutaminol in *Trichothecium roseum. Phytochem*. **26**, 2509–2514 (1987).

23. Balskus, E. P. & Walsh, C. T. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* **329**, 1653–1656 (2010).

24. Clifford, M. N., Jaganath, I. B., Ludwig, I. A. & Crozier, A. Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. *Nat. Prod. Rep.* **34**, 1391–1421 (2017).

25. Gourley, D. G., Shrive, A. K., Polikarpov, I., Krell, T., Coggins, J. R., Hawkins, A. R., Isaacs, N. W. & Sawyer, L. The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nat. Struct. Biol*. **6**, 521–525 (1999).

26. Butler, J. R., Alworth, W. L. & Nugent, M. J. Mechanism of dehydroquinase catalyzed dehydration. Formation of a Schiff base intermediate. *J. Am. Chem. Soc*. **96**, 1617–1618 (1974).

27. Roszak, A. W., Robinson, D. A., Krell, T., Hunter, I. S., Fredrickson, M., Abell, C., Coggins, J. R. & Lapthorn, A. J. The structure and mechanism of the type II dehydroquinase from *Streptomyces coelicolor*. *Structure* **10**, 493–503 (2002).

28. Light, S. H., Minasov, G., Shuvalova, L., Duban, M.-E., Caffrey, M., Anderson, W. F. & Lavie, A. Insights into the mechanism of type I dehydroquinate dehydratases from structures of reaction intermediates. *J. Biol. Chem*. **286**, 3531–3539 (2011).

29. Polley, L. D. Purification and characterization of 3-dehydroquinate hydrolase and shikmate oxidoreductase: evidence for a bifunctional enzyme. *Biochim. Biophys. Acta.* **526**, 259–266 (1978).

30. Koppisch, A. T., Browder, C. C., Moe, A. L., Shelley, J. T., Kinkel, B. A., Hersman, L. E., Iyer, S. & Ruggiero, C. E. Petrobactin is the primary siderophore synthesized by *Bacillus anthracis* str. sterne under conditions of iron starvation. *Biometals* **18**, 577–585 (2005).

31. Fischbach, M. A., Lin, H., Liu, D. R. & Walsh, C. T. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat. Chem. Biol.* **2**, 132–138 (2006).

32. Abergel, R. J., Wilson, M. K., Arceneaux, J. E. L., Hoette, T. M., Strong, R. K., Byers, B. R. & Raymond, K. N. Anthrax pathogen evades the mammalian immune system through stealth siderophore production*. Proc. Natl. Acad. Sci. U.S.A.* **103**, 18499–18503 (2006).

33. Koppisch, A. T., Hotta, K., Fox, D. T., Ruggiero, C. E., Kim, C.-Y., Sanchez, T., Iyer, S., Browder, C. C., Unkefer, P. J. & Unkefer, C. J. Biosynthesis of the 3,4-dihydroxybenzoate moieties of petrobactin by *Bacillus anthracis. J. Org. Chem*. **73**, 5759–5765 (2008).

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

35. Lee, J. Y., Janes, B. K., Passalacqua, K. D., Pfleger, B. F., Bergman, N. H., Liu, H., Håkansson, K., Somu, R. V., Aldrich, C. C., Cendrowski, S., Hanna, P. C. & Sherman, D. H. Biosynthetic analysis of the petrobactin siderophore pathway from *Bacillus anthracis. J. Bacteriol*. **189**, 1698–1710 (2007).

36. Fox, D. T., Hotta, K., Kim, C.-Y. & Koppisch, A. T. The missing link in petrobactin biosynthesis: *asbF* encodes a (−)-3-dehydroshikimate dehydratase. *Biochem*. **47**, 12251–12253 (2008).

37. Pfleger, B. F., Kim, Y., Nusca, T. D., Maltseva, N., Lee, J. Y., Rath, C. M., Scaglione, J. B., Janes, B. K., Anderson, E. C., Bergman, N. H., Hanna, P. C., Joachimiak, A. & Sherman, D. H. Structural and functional analysis of AsbF: Origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17133–17138 (2008).

38. Manck, L. E., Park, J., Tully, B. J., Poire, A. M., Bundy, R. M., Dupont, C. L. & Barbeau, K. A. Petrobactin, a siderophore produced by *Alteromonas*, mediates community iron acquisition in the global ocean. *ISME J.* **16**, 358–369 (2022).

39. Hirayama, A., Eguchi, T. & Kudo, F. A single PLP-dependent enzyme PctV catalyzes the transformation of 3-dehydroshikimate into 3-aminobenzoate in the biosynthesis of pactamycin. *ChemBioChem* **14**, 1198–1203 (2013).

40. Wiley, P. F., Jahnke, H. K., MacKellar, F., Kelly, R. B. & Argoudelis, A. D. The structure of pactamycin. *J. Org. Chem.* **35**, 1420–1425 (1970).

41. Ito, T., Roongsawang, N., Shirasaka, N., Lu, W., Flatt, P. M., Kasanah, N., Miranda, C. & Mahmud, T. Deciphering pactamycin biosynthesis and engineered production of new pactamycin analogues. *ChemBioChem* **10**, 2253–2265 (2009).

42. Abugrain, M. E., Lu, W., Li, Y., Serrill, J. D., Brumsted, C. J., Osborn, A. R., Alani, A., Ishmael, J. E., Kelly, J. X. & Mahmud, T. Interrogating the tailoring steps of pactamycin biosynthesis and accessing new pactamycin analogues. *ChemBioChem* **17**, 1585–1588 (2016).

43. Weller, D. D. & Rinehart, K. L. Biosynthesis of the antitumor antibiotic pactamycin. A methionine-derived ethyl group and a C_7N unit. *J. Am. Chem. Soc*. **100**, 6757–6760 (1978).

44. Rinehart, K. L., Potgieter, M., Delaware, D. L. & Seto, H. Direct evidence from multiple carbon-13-labeling and homonuclear decoupling for the labeling pattern by glucose of the m-aminobenzoyl (C_7N) unit of pactamycin. *J. Am. Chem. Soc*. **103**, 2099–2101 (1981).

45. Kudo, F., Kasama, Y., Hirayama, T. & Eguchi, T. Cloning of the pactamycin biosynthetic gene cluster and characterization of a crucial glycosyltransferase prior to a unique cyclopentane ring formation. J. *Antibiot. (Tokyo)* **60**, 492–503 (2007).

46. Eida, A. A., Abugrain, M. E., Brumsted, C. J. & Mahmud, T. Glycosylation of acyl carrier protein-bound polyketides during pactamycin biosynthesis. *Nat. Chem. Biol.* **15**, 795–802 (2019).

47. Kahkeshani, N., Farzaei, F., Fotouhi, M., Alavi, S. S., Bahramsoltani, R., Naseri, R., Momtaz, S., Abbasabadi, Z., Rahimi, R., Farzaei, M. H. & Bishayee, A. Pharmacological effects of gallic acid in health and diseases: A mechanistic review. *Iran J. Basic Med. Sci.* **22**, 225–237 (2019).

48. Werner, I., Bacher, A. & Eisenreich, W. Retrobiosynthetic NMR Studies with 13C-babeled glucose: formation of gallic acid in plants and fungi. *J. Biol. Chem*. **272**, 25474–25482 (1997).

49. Werner, R. A., Rossmann, A., Schwarz, C., Bacher, A., Schmidt, H.-L. & Eisenreich, W. Biosynthesis of gallic acid in *Rhus typhina*: discrimination between alternative pathways from natural oxygen isotope abundance. *Phytochem*. **65**, 2809–2813 (2004).

50. Muir, R. M., Ibáñez, A. M., Uratsu, S. L., Ingham, E. S., Leslie, C. A., McGranahan, G. H., Batra, N., Goyal, S., Joseph, J., Jemmis, E. D. & Dandekar, A. M. Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). *Plant Mol. Biol.* **75**, 555–565 (2011).

51. Michel, G., Roszak, A. W., Sauvé, V., Maclean, J., Matte, A., Coggins, J. R., Cygler, M. & Lapthorn, A. J. Structures of shikimate dehydrogenase AroE and its paralog YdiB. *J. Biol. Chem*. **278**, 19463–19472 (2003).

52. Peek, J. & Christendat, D. The shikimate dehydrogenase family: Functional diversity within a conserved structural and mechanistic framework. *Arch. Biochem. Biophys*. **566**, 85–99 (2015).

53. De Rosa, M., Gambacorta, A. & Bu'lock, J. D. Effects of pH and temperature on the fatty acid composition of *Bacillus acidocaldarius. J. Bacteriol*. **117**, 212–214 (1974).

54. Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep*. **19**, 70–99 (2002).

55. Reynolds, K. A., Wang, P., Fox, K. M., Speedie, M. K., Lam, Y. & Floss, H. G. Purification and characterization of a novel enoyl coenzyme A reductase from *Streptomyces collinus. J. Bacteriol*. **174**, 3850–3854 (1992).

56. Cropp, T. A., Wilson, D. J. & Reynolds, K. A. Identification of a cyclohexylcarbonyl CoA biosynthetic gene cluster and application in the production of doramectin. *Nat. Biotechnol.* **18**, 980–983 (2000).

57. Skyrud, W., Flores, A. D. R. & Zhang, W. Biosynthesis of cyclohexanecarboxyl-CoA highlights a promiscuous shikimoyl-CoA synthetase and a FAD-dependent dehydratase. *ACS Catal*. **10**, 3360–3364 (2020).

58. Gan, J., Gu, Y., Li, Y., Yan, H. & Ji, X. Crystal structure of *Mycobacterium tuberculosis* shikimate kinase in complex with shikimic acid and an ATP analogue. *Biochem*. **45**, 8539–8545 (2006).

59. Hartmann, M. D., Bourenkov, G. P., Oberschall, A., Strizhov, N. & Bartunik, H. D. Mechanism of phosphoryl transfer catalyzed by shikimate kinase from *Mycobacterium tuberculosis. J. Mol. Biol*. **364**, 411–423 (2006).

60. Walsh, C. T., Benson, T. E., Kim, D. H. & Lees, W. J. The versatility of phosphoenolpyruvate and its vinyl ether products in biosynthesis. *Chem. Biol.* **3**, 83–91 (1996).

61. Schönbrunn, E., Eschenburg, S., Shuttleworth, W. A., Schloss, J. V., Amrhein, N., Evans, J. N. S. & Kabsch, W. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1376–1380 (2001).

62. Herrmann, K. The shikimate pathway: early steps in the biosynthesis of aromatic compounds*. Plant Cell* **7**, 907–919 (1995).

63. Maclean, J. & Ali, S. The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure* **11**, 1499–1511 (2003).

64. Kitzing, K., Auweter, S., Amrhein, N. & Macheroux, P. Mechanism of chorismate synthase: role of the two invariant histidine residues in the active site. *J. Biol. Chem*. **279**, 9451–9461 (2004).

65. Van Lanen, S. G., Lin, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 494–499 (2008).

66. Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170–1173 (2002).

67. Yan, X., Hindra, Ge, H., Yang, D., Huang, T., Crnovcic, I., Chang, C.-Y., Fang, S.-M., Annaval, T., Zhu, X., Huang, Y., Zhao, L.-X., Jiang, Y., Duan, Y. & Shen, B. Discovery of alternative producers of the enediyne antitumor antibiotic C-1027 with high titers. *J. Nat. Prod*. **81**, 594–599 (2018).

68. Laursen, J. B. & Nielsen, J. Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. *Chem. Rev*. **104**, 1663–1686 (2004).

69. Pierson, L. S. & Pierson, E. A. Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl. Microbiol. Biotechnol*. **86**, 1659–1670 (2010).

70. Price-Whelan, A., Dietrich, L. E. P. & Newman, D. K. Rethinking 'secondary' metabolism: physiological roles for phenazine antibiotics. *Nat. Chem. Biol.* **2**, 71–78 (2006).

71. Mavrodi, D. V., Blankenfeldt, W. & Thomashow, L. S. Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. *Annu. Rev. Phytopathol*. **44**, 417–445 (2006).

72. Millican, R. C. Biosynthesis of pyocyanine. Incorporation of [14C]shikimic acid. *Biochimi. Biophys. Acta* **57**, 407–409 (1962).

73. Ingledew, W. M. & Campbell, J. J. R. Evaluation of shikimic acid as a precursor of pyocyanine. *Can*. *J. Microbiol*. **15**, 535–541 (1969).

74. Longley, R. P., Halliwell, J. E., Campbell, J. J. R. & Ingledew, W. M. The branchpoint of pyocyanine biosynthesis. *Can. J. Microbiol*. **18**, 1357–1363 (1972).

75. Calhoun, D. H., Carson, M. & Jensen, R. A. The branch point metabolite for pyocyanine biosynthesis in *Pseudomonas aeruginosa. Microbiol.* **72**, 581–583 (1972).

76. Pierson, L. S., Gaffney, T., Lam, S. & Gong, F. Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium *Pseudomonas aureofaciens* 30-84. *FEMS Microbiol. Lett*. **134**, 299–307 (1995).

77. Mavrodi, D. V., Ksenzenko, V. N., Bonsall, R. F., Cook, R. J., Boronin, A. M. & Thomashow, L. S. A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. *J. Bacteriol*. **180**, 2541–2548 (1998).

78. McDonald, M., Mavrodi, D. V., Thomashow, L. S. & Floss, H. G. Phenazine biosynthesis in *Pseudomonas fluorescens*: branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *J. Am. Chem. Soc.* **123**, 9459–9460 (2001).

79. Li, Q.-A., Mavrodi, D. V., Thomashow, L. S., Roessle, M. & Blankenfeldt, W. Ligand binding induces an ammonia channel in 2-amino-2-desoxyisochorismate (ADIC) synthase PhzE. *J. Biol. Chem*. **286**, 18213–18221 (2011).

80. Parsons, J. F., Song, F., Parsons, L., Calabrese, K., Eisenstein, E. & Ladner, J. E. Structure and function of the phenazine biosynthesis protein PhzF from *Pseudomonas fluorescens* 2-79. *Biochem*. **43**, 12427–12435 (2004).

81. Blankenfeldt, W., Kuzin, A. P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L. S. & Mavrodi, D. V. Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens. Proc. Natl. Acad. Sci. U.S.A.* **101**, 16431–16436 (2004).

82. Mentel, M., Ahuja, E. G., Mavrodi, D. V., Breinbauer, R., Thomashow, L. S. & Blankenfeldt, W. Of two make one: the biosynthesis of phenazines. *Chembiochem* **10**, 2295–2304 (2009).

83. Ahuja, E. G., Janning, P., Mentel, M., Graebsch, A., Breinbauer, R., Hiller, W., Costisella, B., Thomashow, L. S., Mavrodi, D. V., & Blankenfeldt, W. PhzA/B catalyzes the formation of the tricycle in phenazine biosynthesis. *J. Am. Chem. Soc.* **130**, 17053–17061 (2008).

84. Parsons, J. F., Calabrese, K., Eisenstein, E. & Ladner, J. E. Structure of the phenazine biosynthesis enzyme PhzG. *Acta Crystallogr. D Biol. Crystallogr*. **60**, 2110–2113 (2004).

85. Keller-Schierlein, W., Geiger, A., Zähner, H. & Brandl, M. Stoffwechselprodukte von Mikroorganismen. Mitteilung. Die Esmeraldine A und B, tief grüne Farbstoffe aus *Streptomyces antibioticus*, Stamm Tü 2706. *Helvetica Chimi. Acta* **71**, 2058–2070 (1988).

86. Rui, Z., Ye, M., Wang, S., Fujikawa, K., Akerele, B., Aung, M., Floss, H. G., Zhang, W., & Yu, T.W. Insights into a divergent phenazine biosynthetic pathway governed by a plasmid-born esmeraldin gene cluster. *Chem. Biol.* **19**, 1116–1125 (2012).

87. Kitahara, M., Nakamura, H., Matsuda, Y., Hamada, M., Naganawa, H., Maeda, K., Umezawa, H., & Iitaka, Y. Saphenamycin, a novel antibiotic from a strain of *Streptomyces. J. Antibiot.* **35**, 1412–1414 (1982).

88. McDonald, M., Wilkinson, B., Van't Land, C. W., Mocek, U., Lee, S. & Floss, H. G. Biosynthesis of phenazine antibiotics in *Streptomyces antibioticus*: stereochemistry of methyl transfer from carbon-2 of acetate. *J. Am. Chem. Soc*. **121**, 5619–5624 (1999).

89. Bauman, K. D., Li, J., Murata, K., Mantovani, S. M., Dahesh, S., Nizet, V., Luhavaya, H. & Moore, B. S. Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry. *Cell Chem. Biol.* **26**, 724-736.e7 (2019).

90. Mitova, M. I., Lang, G., Wiese, J. & Imhoff, J. F. Subinhibitory concentrations of antibiotics induce phenazine production in a marine *Streptomyces* sp. *J. Nat. Prod*. **71**, 824–827 (2008).

91. Charan, R. D., Schlingmann, G., Janso, J., Bernan, V., Feng, X. & Carter, G. T. Diazepinomicin, a new antimicrobial alkaloid from a marine *Micromonospora* sp. *J. Nat. Prod.* **67**, 1431–1433 (2004).

92. McAlpine, J. B., Banskota, A. H., Charan, R. D., Schlingmann, G., Zazopoulos, E., Piraee, M., Janso, J., Bernan, V. S., Aouidate, M., Farnet, C. M., Feng, X., Zhao, Z., & Carter, G. T. Biosynthesis of diazepinomicin/ECO-4601, a *Micromonospora* secondary metabolite with a novel ring system. *J. Nat. Prod*. **71**, 1585–1590 (2008).

93. Fotso, S., Zabriskie, T. M., Proteau, P. J., Flatt, P. M., Santosa, D. A., & Mahmud, T. Limazepines A-F, pyrrolo-1,4-benzodiazepine antibiotics from an Indonesian *Micrococcus* sp.. *J. Nat. Prod.* **72**, 690–695 (2009).

94. Pavlikova, M., Kamenik, Z., Janata, J., Kadlcik, S., Kuzma, M., & Najmanova, L. Novel pathway of 3-hydroxyanthranilic acid formation in limazepine biosynthesis reveals evolutionary relation between phenazines and pyrrolobenzodiazepines. *Sci. Rep.* **8**, 7810 (2018).

95. Baggaley, K. H., Blessington, B., Falshaw, C. P., Ollis, W. D., Chaiet, L. & Wolf, F. J. The constitution of stravidin, a novel microbiological product. *J. Chem. Soc*. **101**–102 (1969).

96. Stapley, E. O., Mata, J. M., Miller, I. M., Demny, T. C. & Woodruff, H. B. Antibiotic msd-235. production by *Streptomyces avidinii* and *Streptomyces lavendulae*. *Antimicrob. Agents Chemother*. **161**, 20–27 (1963).

97. Chang, Z., Sun, Y., He, J. & Vining, L. C. 2001. p-Aminobenzoic acid and chloramphenicol biosynthesis in *Streptomyces venezuelae*: gene sets for a key enzyme, 4-amino-4-deoxychorismate synthase. *Microbiol*. **147**, 2113–2126.

98. He, J., Magarvey, N., Piraee, M., & Vining, L. C. 2001. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiol*. **147**, 2817–2829.

99. Ye, Q. Z., Liu, J., & Walsh, C. T. p-Aminobenzoate synthesis in *Escherichia coli*: purification and characterization of PabB as aminodeoxychorismate synthase and enzyme X as aminodeoxychorismate lyase. *Proc. Natl. Acad. Sci. U.S.A*. **87**, 9391–9395 (1990).

100. Anderson, K. S., Kati, W. M., Ye, Q. Z., Liu, J., Walsh, C. T., Benesi, A. J. & Johnson, K. A. Isolation and structure elucidation of the 4-amino-4-deoxychorismate intermediate in the PABA enzymic pathway. *J. Am. Chem. Soc*. **113**, 3198–3200 (1991).

101. Gil, J. A. & Campelo-Diez, A. B. Candicidin biosynthesis in *Streptomyces griseus*. *Appl. Microbiol. Biotechnol*. **60**, 633–642 (2003).

102. Jørgensen, H., Fjærvik, E., Hakvåg, S., Bruheim, P., Bredholt, H., Klinkenberg, G., Ellingsen, T. E. & Zotchev, S. B. Candicidin biosynthesis gene cluster is widely distributed among *Streptomyces* spp. isolated from the sediments and the neuston layer of the Trondheim Fjord, Norway. *Appl. Environ. Microbiol*. **75**, 3296–3303 (2009).

103. Saleh, H., Petras, D., Mainz, A., Kerwat, D., Nalbantsoy, A., Erzurumlu, Y. & Süssmuth, R. D. Deuterium-labeled precursor feeding reveals a new pABA-containing meroterpenoid from the mango pathogen *Xanthomonas citri* pv. mangiferaeindicae. *J. Nat. Prod*. **79**, 1532–1537 (2016).

104. Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., Vieweg, L., Duplan, S., Marguerettaz, M., Noëll, J., Pieretti, I., Hügelland, M., Kemper, S., Mainz, A., Rott, P., Royer, M. & Süssmuth, R. D. The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. *Nat. Chem. Biol*. **11**, 195–197 (2015).

105. Yoo, Y. J., Kim, H., Park, S. R. & Yoon, Y. J. An overview of rapamycin: from discovery to future perspectives. *J. Ind. Microbiol. Biotechnol*. **44**, 537–553 (2017).

106. Zoncu, R., Efeyan, A. & Sabatini, D. M. mTOR: from growth signal integration to cancer, diabetes and ageing*. Nat. Rev. Mol. Cell Biol*. **12**, 21–35 (2011).

107. Choi, J., Chen, J., Schreiber, S. L. & Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**, 239–242 (1996).

108. Thomson, A. W., Bonham, C. A. & Zeevi, A. Mode of action of tacrolimus (FK506): molecular and cellular mechanisms. *Ther. Drug Monit*. **17**, 584–591 (1995).

109. Paiva, N. L., Roberts, M. F. & Demain, A. L. The cyclohexane moiety of rapamycin is derived from shikimic acid in *Streptomyces hygroscopicus. J. Ind. Microbiol.* **12**, 423–428 (1993).

110. Wallace, K. K., Reynolds, K. A., Koch, K., McArthur, H. A. I., Brown, M. S., Wax, R. G. & Moore, B. S. Biosynthetic studies of ascomycin (FK520): formation of the (1R,3R,4R)-3,4-dihydroxycyclohexanecarboxylic acid-derived moiety. *J. Am. Chem. Soc.* **116**, 11600–11601 (1994).

111. Schwecke, T., Aparicio, J. F., Molnár, I., König, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortés, J. & Lester, J. B. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7839–7843 (1995).

112. Lowden, P. A. S., Wilkinson, B., Böhm, G. A., Handa, S., Floss, H. G., Leadlay, P. F. & Staunton, J. Origin and true nature of the starter unit for the rapamycin polyketide synthase. *Angew. Chem. Int. Ed. Engl*. **40**, 777–779 (2001).

113. Gregory, M. A., Gaisser, S., Lill, R. E., Hong, H., Sheridan, R. M., Wilkinson, B., Petkovic, H., Weston, A. J., Carletti, I., Lee, H.-L., Staunton, J. & Leadlay, P. F. Isolation and characterization of pre-rapamycin, the first macrocyclic intermediate in the biosynthesis of the immunosuppressant rapamycin *by S. hygroscopicus. Angew. Chem. Int. Ed. Engl.* **43**, 2551–2553 (2004).

114. Gregory, M. A., Petkovic, H., Lill, R. E., Moss, S. J., Wilkinson, B., Gaisser, S., Leadlay, P. F. & Sheridan, R. M. Mutasynthesis of rapamycin analogues through the manipulation of a gene governing starter unit biosynthesis. *Angew. Chem. Int. Ed. Engl.* **44**, 4757–4760 (2005).

115. Andexer, J. N., Kendrew, S. G., Nur-e-Alam, M., Lazos, O., Foster, T. A., Zimmermann, A.-S., Warneck, T. D., Suthar, D., Coates, N. J., Koehn, F. E., Skotnicki, J. S., Carter, G. T., Gregory, M. A., Martin, C. J., Moss, S. J., Leadlay, P. F. & Wilkinson, B. Biosynthesis of the immunosuppressants FK506, FK520, and rapamycin involves a previously undescribed family of enzymes acting on chorismate. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4776–4781 (2011).

116. Shigdel, U. K., Lee, S.-J., Sowa, M. E., Bowman, B. R., Robison, K., Zhou, M., Pua, K. H., Stiles, D. T., Blodgett, J. A. V., Udwary, D. W., Rajczewski, A. T., Mann, A. S., Mostafavi, S., Hardy, T., Arya, S., Weng, Z., Stewart, M., Kenyon, K., Morgenstern, J. P., Pan, E., Gray, D. C., Pollock, R. M., Fry, A. M., Klausner, R. D., Townson, S. A. & Verdine, G. L. Genomic discovery of an evolutionarily programmed modality for small-molecule targeting of an intractable protein surface. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17195–17203 (2020).

117. Luke, R. K. & Gibson, F. Location of three genes concerned with the conversion of 2,3-dihydroxybenzoate into enterochelin in *Escherichia coli* K-12. *J. Bacteriol*. **107**, 557–562 (1971).

118. Young, I. G., Langman, L., Luke, R. K. & Gibson, F. Biosynthesis of the iron-transport compound enterochelin: mutants of *Escherichia coli* unable to synthesize 2,3-dihydroxybenzoate. *J Bacteriol.* **106**, 51–57 (1971).

119. Young, I. G., Cox, G. B. & Gibson, F. 2,3-Dihydroxybenzoate as a bacterial growth factor and its route of biosynthesis. *Biochimi. Biophys. Acta - Gen. Subj.* **141**, 319–331 (1967).

120. Young, I. G., Batterham, T. J. & Gibson, F. The isolation, identification and properties of isochorismic acid. An intermediate in the biosynthesis of 2,3-dihydroxybenzoic acid. *Biochimi. Biophys. Acta - Gen. Subj*. **177**, 389–400 (1969).

121. Liu, J., Quinn, N., Berchtold, G. A. & Walsh, C. T. Overexpression, purification and characterization of isochorismate synthase (EntC), the first enzyme involved in the biosynthesis of enterobactin from chorismate. *Biochem*. **29**, 1417–1425 (1990).

122. Sridharan, S., Howard, N., Kerbarh, O., Błaszczyk, M., Abell, C. & Blundell, T. L. Crystal structure of Escherichia coli enterobactin-specific isochorismate synthase (EntC) bound to its reaction product isochorismate: implications for the enzyme mechanism and differential activity of chorismate-utilizing enzymes. *J. Mol. Biol.* **397**, 290–300 (2010).

123. Daruwala, R., Kwon, O., Meganathan, R. & Hudspeth, M. E. A new isochorismate synthase specifically involved in menaquinone (vitamin K2) biosynthesis encoded by the *menF* gene. *FEMS Microbiol. Lett*. **140**, 159–163 (1996).

124. Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).

125. Rekhter, D., Lüdke, D., Ding, Y., Feussner, K., Zienkiewicz, K., Lipka, V., Wiermer, M., Zhang, Y. & Feussner, I. Isochorismate-derived biosynthesis of the plant stress hormone salicylic acid. *Science* **365**, 498–502 (2019).

126. Gaille, C., Reimmann, C. & Haas, D. Isochorismate synthase (PchA), the first and rate-limiting enzyme in salicylate biosynthesis of *Pseudomonas aeruginosa. J. Biol. Chem.* **278**, 16893–16898 (2003).

127. Komaki, H., Nemoto, A., Tanaka, Y., Takagi, H., Yazawa, K., Mikami, Y., Shigemori, H., Kobayashi, J., Ando, A. & Nagata, Y. Brasilicardin A, a new terpenoid antibiotic from pathogenic *Nocardia brasiliensis*: fermentation, isolation and biological activity. *J. Antibiot. (Tokyo)* **52**, 13–19 (1999).

128. Usui, T., Nagumo, Y., Watanabe, A., Kubota, T., Komatsu, K., Kobayashi, J. & Osada, H. Brasilicardin A, a natural immunosuppressant, targets amino acid transport system. *Chem. Biol.* **13**, 1153–1160 (2006).

129. Jiang, Y., Wang, H., Lu, C., Ding, Y., Li, Y. & Shen, Y. Identification and characterization of the cuevaene A biosynthetic gene cluster in *Streptomyces* sp. LZ35. *ChemBioChem* **14**, 1468–1475 (2013).

130. Schlegel, B., Groth, I. & Gräfe, U. Cuevaenes A and B, new polyene carboxylic acids from *Streptomyces* sp. HKI 0180. *J. Antibiot. (Tokyo)* **53**, 415–417 (2000).

131. Lim, C. L., Nogawa, T., Okano, A., Futamura, Y., Kawatani, M., Takahashi, S., Ibrahim, D. & Osada, H. Unantimycin A, a new neoantimycin analog isolated from a microbial metabolite fraction library. *J. Antibiot*. **69,** 456–458 (2016).

132. Shen, Y., Sun, F., Zhang, L., Cheng, Y., Zhu, H., Wang, S.-P., Jiao, W.-H., Leadlay, P. F., Zhou, Y. & Lin, H.-W. Biosynthesis of depsipeptides with a 3-hydroxybenzoate moiety and selective anticancer activities involves a chorismatase. *J. Biol. Chem.* **295**, 5509–5518 (2020).

133. He, Y.-W., Cao, X.-Q. & Poplawsky, A. R. Chemical structure, biological roles, biosynthesis and regulation of the yellow xanthomonadin pigments in the phytopathogenic genus *Xanthomonas. Mol. Plant Microbe Int.* **33**, 705–714 (2020).

134. He, Y.-W., Wu, J., Zhou, L., Yang, F., He, Y.-Q., Jiang, B.-L., Bai, L., Xu, Y., Deng, Z., Tang, J.-L. & Zhang, L.-H. *Xanthomonas campestris* diffusible factor Is 3-hydroxybenzoic acid and is associated with xanthomonadin biosynthesis, cell viability, antioxidant activity, and systemic invasion. *Mol. Plant Microbe Int.* **24**, 948–957 (2011).

135. Zhou, L., Wang, J.-Y., Wang, J., Poplawsky, A., Lin, S., Zhu, B., Chang, C., Zhou, T., Zhang, L.-H. & He, Y.-W. The diffusible factor synthase XanB2 is a bifunctional chorismatase that links the shikimate pathway to ubiquinone and xanthomonadins biosynthetic pathways. *Mol. Microbiol.* **87**, 80–93 (2013).

136. Cao, X.-Q., Ouyang, X.-Y., Chen, B., Song, K., Zhou, L., Jiang, B.-L., Tang, J.-L., Ji, G., Poplawsky, A. R. & He, Y.-W. Genetic interference analysis reveals that both 3-hydroxybenzoic acid and 4-hydroxybenzoic acid are involved in xanthomonadin biosynthesis in the phytopathogen *Xanthomonas campestris* pv. campestris. *Phytopathol*. **110**, 278–286 (2020).

137. Hubrich, F., Juneja, P., Müller, M., Diederichs, K., Welte, W. & Andexer, J. N. Chorismatase mechanisms reveal fundamentally different types of reaction in a single conserved protein fold. *J. Am. Chem. Soc.* **137**, 11032–11037 (2015).

138. Siddiqi, M. A., Laessig, R. H. & Reed, K. D. Polybrominated diphenyl ethers (PBDEs): new pollutants-old diseases*. Clin. Med. Res*. **1**, 281–290 (2003).

139. Agarwal, V., El Gamal, A. A., Yamanaka, K., Poth, D., Kersten, R. D., Schorn, M., Allen, E. E. & Moore, B. S. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* **10**, 640–647 (2014).

140. Hanefeld, U., Floss, H. G. & Laatsch, H. Biosynthesis of the marine antibiotic pentabromopseudilin. Part 1. The benzene ring*. J. Org. Chem*. **59**, 3604–3608 (1994).

141. Agarwal, V., Blanton, J. M., Podell, S., Taton, A., Schorn, M. A., Busch, J., Lin, Z., Schmidt, E. W., Jensen, P. R., Paul, V. J., Biggs, J. S., Golden, J. W., Allen, E. E. & Moore, B. S. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol*. **13**, 537–543 (2017).

142. Liu, D.-Z., Wang, F., Liao, T.-G., Tang, J.-G., Steglich, W., Zhu, H.-J. & Liu, J.-K. Vibralactone: A lipase inhibitor with an unusual fused β-lactone produced by cultures of the basidiomycete *Boreostereum vibrans. Org. Lett*. **8**, 5749–5752 (2006).

143. Cotton, R. G. H. & Gibson, F. The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. *Biochimi. Biophys. Acta - Gen. Subj.* **100**, 76–88 (1965).

144. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A. & Knowles, J. R. Stereochemistry of the rearrangement of chorismate to prephenate: chorismate mutase involves a chair transition state. *J. Am. Chem. Soc.* **106**, 2701–2703 (1984).

145. Khanjin, N. A., Snyder, J. P. & Menger, F. M. Mechanism of chorismate mutase: contribution of conformational restriction to catalysis in the claisen rearrangement. *J. Am. Chem. Soc.* **121**, 11831–11846 (1999).

146. Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. & Fenical, W. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora. Angew. Chem. Int. Ed. Engl.* **42**, 355–357 (2003).

147. Fenical, W., Jensen, P. R., Palladino, M. A., Lam, K. S., Lloyd, G. K. & Potts, B. C. Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* **17**, 2175–2180 (2009).

148. Roth, P., Gorlia, T., Reijneveld, J. C., De Vos, F. Y. F. L., Idbaih, A., Frenel, J.-S., Le Rhun, E., Sepulveda Sánchez, J. M., Perry, J. R., Masucci, L., Freres, P., Hirte, H. W., Seidel, C., Walenkamp, A. M. E., Dhermain, F., Van Den Bent, M. J., O'Callaghan, C. J., Vanlancker, M., Mason, W. P. & Weller, M. A phase III trial of marizomib in combination with temozolomide-based radiochemotherapy versus temozolomide-based radiochemotherapy alone in patients with newly diagnosed glioblastoma. *J. Clin. Oncol*. **39**, 2004–2004 (2021).

149. Groll, M., Huber, R. & Potts, B. C. M. Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of β-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc*. **128**, 5136–5141 (2006).

150. Bauman, K. D., Shende, V. V., Chen, P. Y.-T., Trivella, D. B. B., Gulder, T. A. M., Vellalath, S., Romo, D. & Moore, B. S. Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead. *Nat. Chem. Biol*. **18**, 538–546 (2022).

151. Beer, L. L. & Moore, B. S. Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica. Org. Lett*. **9**, 845–848 (2007).

152. Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica. Proc. Natl. Acad. Sci. U.S.A.* **104***,* 10376–10381 (2007).

153. Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Lechner, A., Kale, A. J., Kobayashi, Y., Reynolds, K. A. & Moore, B. S. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-l-methionine*. Proc. Natl. Acad. Sci. U.S.A.* **106**, 12295–12300 (2009).

154. McGlinchey, R. P., Nett, M., Eustáquio, A. S., Asolkar, R. N., Fenical, W. & Moore, B. S. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **130**, 7822 (2008).

155. Nett, M., Gulder, T. A. M., Kale, A. J., Hughes, C. C. & Moore, B. S. Function-oriented biosynthesis of β-lactone proteasome inhibitors in *Salinispora tropica. J. Med. Chem*. **52**, 6163–6167 (2009).

156. Mahlstedt, S. A. & Walsh, C. T. Investigation of anticapsin biosynthesis reveals a four-enzyme pathway to tetrahydrotyrosine in *Bacillus subtilis. Biochem*. **49**, 912–923 (2010).

157. Roscoe, J. & Abraham, E. P. Experiments relating to the biosynthesis of bacilysin. *Biochem. J*. **99**, 793–800 (1966).

158. Hilton, M. D., Alaeddinoglu, N. G. & Demain, A. L. Synthesis of bacilysin by *Bacillus subtilis* branches from prephenate of the aromatic amino acid pathway. *J. Bacteriol.* **170**, 482–484 (1988).

159. Özcengiz, G. & Öğülür, İ. Biochemistry, genetics and regulation of bacilysin biosynthesis and its significance more than an antibiotic. *New Biotechnol*. **32**, 612–619 (2015).

160. Parker, J. B. & Walsh, C. T. Action and timing of BacC and BacD in the late stages of biosynthesis of the dipeptide antibiotic bacilysin. *Biochem*. **52**, 889–901 (2013).

161. Crawford, J. M., Mahlstedt, S. A., Malcolmson, S. J., Clardy, J. & Walsh, C. T. Dihydrophenylalanine: a prephenate-derived *Photorhabdus luminescens* antibiotic and intermediate in dihydrostilbene biosynthesis. *Chem. Biol*.**18**, 1102–1112 (2011).

162. Scannell, J. P., Pruess, D. L., Demny, T. C., Williams, T. & Stempel, A. L-3-(2, 5-dihydrophenyl) alanine, an antimetabolite of L-phenylalanine produced by a streptomycete. *J. Antibiot*. **23**, 618–619 (1970).

163. Knöchel, T., Ivens, A., Hester, G., Gonzalez, A., Bauerle, R., Wilmanns, M., Kirschner, K. & Jansonius, J. N. The crystal structure of anthranilate synthase from *Sulfolobus solfataricus*: Functional implications. *Proc. Natl. Acad. Sci. U.S.A*. **96**, 9479–9484 (1999).

164. Byrnes, W. M., Goldberg, R. N., Holden, M. J., Mayhew, M. P. & Tewari, Y. B. Thermodynamics of reactions catalyzed by anthranilate synthase. *Biophys. Chem*. **84**, 45–64 (2000).

165. Resende, D. I. S. P., Boonpothong, P., Sousa, E., Kijjoa, A. & Pinto, M. M. M. Chemistry of the fumiquinazolines and structurally related alkaloids. *Nat. Prod. Rep*. **36**, 7–34 (2019).

166. Smith, S. G., Sanchez, R. & Zhou, M.-M. Privileged diazepine compounds and their emergence as bromodomain inhibitors. *Chem. Biol*. **21**, 573–583 (2014).

167. Yin, W.-B., Grundmann, A., Cheng, J. & Li, S.-M. Acetylaszonalenin biosynthesis in *Neosartorya fischeri:* identification of the biosynthetic gene cluster by genomic mining and functional proof of the genes by biochemical investigation. *J. Biol. Chem*. **284**, 100–109 (2009).

168. Ames, B. D. & Walsh, C. T. Anthranilate-activating modules from fungal nonribosomal peptide assembly lines. *Biochem*. **49**, 3351–3365 (2010).

169. Li, H., Gilchrist, C. L. M., Phan, C.-S., Lacey, H. J., Vuong, D., Moggach, S. A., Lacey, E., Piggott, A. M. & Chooi, Y.-H. Biosynthesis of a new benzazepine alkaloid nanangelenin A from *Aspergillus nanangensis* involves an unusual L-kynurenine-incorporating NRPS catalyzing regioselective lactamization. *J. Am. Chem. Soc.* **142**, 7145–7152 (2020).

170. Leimgruber, W., Stefanović, V., Schenker, F., Karr, A. & Berger, J. Isolation and characterization of anthramycin, a new antitumor antibiotic. *J. Am. Chem. Soc.* **87**, 5791–5793 (1965).

171. Hurley, L. H. & Gairola, C. Pyrrolo(1,4)benzodiazepine antitumor antibiotics: biosynthetic studies on the conversion of tryptophan to the anthranilic acid moieties of sibiromycin and tomaymycin. *Antimicrob. Agents Chemother*. **15**, 42–45 (1979).

172. Li, W., Khullar, A., Chou, S., Sacramo, A. & Gerratana, B. Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Appl. Environ. Microbiol.* **75**, 2869–2878 (2009).

173. Hu, Y., Phelan, V., Ntai, I., Farnet, C. M., Zazopoulos, E. & Bachmann, B. O. Benzodiazepine biosynthesis in *Streptomyces refuineus. Chem. Biol.* **14**, 691–701 (2007).

174. Hurley, L. H., Gairola, C. & Das, N. V. Pyrrolo[1,4]benzodiazepine antibiotics. Biosynthesis of the antitumor antibiotic 11-demethyltomaymycin and its biologically inactive metabolite oxotomaymycin by *Streptomyces achromogenes. Biochem.* **15**, 3760–3769 (1976).

175. Höfle, G. & Kunze, B. Biosynthesis of aurachins A−L in *Stigmatella aurantiaca*: a feeding study. *J. Nat. Prod.* **71**, 1843–1849 (2008).

176. Sandmann, A., Dickschat, J., Jenke-Kodama, H., Kunze, B., Dittmann, E. & Müller, R. A type II polyketide synthase from the gram-negative bacterium *Stigmatella aurantiaca* is involved in aurachin alkaloid biosynthesis. *Angew. Chem. Int. Ed. Engl.* **46**, 2712–2716 (2007).

177. Pistorius, D., Li, Y., Mann, S. & Müller, R. Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis. *J. Am. Chem. Soc.* **133**, 12362–12365 (2011).

178. Kim, M. C., Winter, J. M., Asolkar, R. N., Boonlarppradab, C., Cullum, R. & Fenical, W. Marinoterpins A–C: rare linear merosesterterpenoids from marine-derived actinomycete bacteria of the family *Streptomycetaceae. J. Org. Chem.* **86**, 11140–11148 (2021).

179. Skaltsounis, A. L., Mitaku, S. & Tillequin, F. Acridone alkaloids. *The Alkaloids: Chemistry and Biology* vol. **54** 259–377 (Academic Press, 2000).

180. Gröger, D. & Johne, S. Zur Biosynthese einiger alkaloide von *Glycosmis arborea* (Rutaceae). Z. Naturforsh. **23**, 1072–1075 (1968).

181. Rohde, B., Hans, J., Martens, S., Baumert, A., Hunziker, P. & Matern, U. Anthranilate N-methyltransferase, a branch-point enzyme of acridone biosynthesis. *Plant J.* **53**, 541–553 (2008).

182. Choi, G.-S., Choo, H. J., Kim, B.-G. & Ahn, J.-H. Synthesis of acridone derivatives via heterologous expression of a plant type III polyketide synthase in *Escherichia coli. Microb. Cell Factories* **19**, 73 (2020).

183. Li, Z., Chen, Y., Meesapyodsuk, D. & Qiu, X. The biosynthetic pathway of major avenanthramides in oat. *Metabolites* **9**, 163 (2019).

184. Yang, Q. et al. Analysis of the involvement of hydroxyanthranilate hydroxycinnamoyltransferase and caffeoyl-CoA 3-O-methyltransferase in phytoalexin biosynthesis in oat. *Mol. Plant Microbe Interact.* **17**, 81–89 (2004).

185. Kang, Q., Shen, Y. & Bai, L. Biosynthesis of 3,5-AHBA-derived natural products. *Nat. Prod. Rep.* **29**, 243–263 (2012).

186. Escalante, A., Mendoza-Flores, R., Gosset, G. & Bolívar, F. The aminoshikimic acid pathway in bacteria as source of precursors for the synthesis of antibacterial and antiviral compounds. *J.Ind. Microbiol. Biotechnol.* **48**, (2021).

187. Floss, H. G., Yu, T.-W. & Arakawa, K. The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review. *J. Antibiot*. **64**, 35–44 (2011).

188. Yu, T. W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C. G., Leistner, E. & Floss, H. G. Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. *J. Biol. Chem*. **276**, 12546–12555 (2001).

189. Floss, H. G. & Yu, T.-W. Rifamycin – mode of action, resistance, and biosynthesis. *Chem. Rev*. **105**, 621–632 (2005).

190. Adelfo Escalante, A., Carmona, S. B., Diaz Quiroz, D. C. & Bolivar, F. Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. *Res. Rep. Med. Chem*. **35** (2014).

191. Peek, J., Lilic, M., Montiel, D., Milshteyn, A., Woodworth, I., Biggins, J. B., Ternei, M. A., Calle, P. Y., Danziger, M., Warrier, T., Saito, K., Braffman, N., Fay, A., Glickman, M. S., Darst, S. A., Campbell, E. A. & Brady, S. F. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat. Commun.* **9**, 4147 (2018).

192. Williams, P. G., Asolkar, R. N., Kondratyuk, T., Pezzuto, J. M., Jensen, P. R. & Fenical, W. Saliniketals A and B, bicyclic polyketides from the marine actinomycete *Salinispora arenicola. J. Nat. Prod*. **70**, 83–88 (2007).

193. Wilson, M. C., Gulder, T. A. M., Mahmud, T. & Moore, B. S. Shared Biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the sare1259 -encoded cytochrome P450. *J. Am. Chem. Soc.* **132**, 12757–12765 (2010).

194. Koteva, K., Cox, G., Kelso, J. K., Surette, M. D., Zubyk, H. L., Ejim, L., Stogios, P., Savchenko, A., Sørensen, D. & Wright, G. D. Rox, a rifamycin resistance enzyme with an unprecedented mechanism of action. *Cell Chem. Biol.* **25**, 403-412.e5 (2018).

195. Zheng, X.-F., Liu, X.-Q., Peng, S.-Y., Zhou, Q., Xu, B., Yuan, H. & Tang, G.-L. Characterization of the rifamycin-degrading monooxygenase from rifamycin producers implicating its involvement in saliniketal biosynthesis. *Front. Microbiol*. **11**, 971 (2020).

196. Sattler, I., Thiericke, R. & Zeeck, A. The manumycin-group metabolites. *Nat. Prod. Rep.* **15**, 221 (1998).

197. Castro-Falcón, G., Creamer, K. E., Chase, A. B., Kim, M. C., Sweeney, D., Glukhov, E., Fenical, W. & Jensen, P. R. Structure and candidate biosynthetic gene cluster of a manumycin-type metabolite from *Salinispora pacifica. J. Nat. Prod*. **85**, 980–986 (2022).

198. Zeeck, A., Schröder, K., Frobel, K., Grote, R. & Thiericke, R. The structure of manumycin. I. Characterization, structure elucidation and biological activity. *J. Antibiot.* **40**, 1530–1540 (1987).

199. Omura, S., Kitao, C., Tanaka, H., Oiwa, R., Takahashi, Y., Nakagawa, A., Shimada, M. & Iwai, Y. A new antibiotic, asukamycin, produced by *Streptomyces. J. Antibiot. (Tokyo)* **29**, 876–881 (1976).

200. Rui, Z., Petříčková, K., Škanta, F., Pospíšil, S., Yang, Y., Chen, C.-Y., Tsai, S.-F., Floss, H. G., Petříček, M. & Yu, T.-W. Biochemical and genetic insights into asukamycin biosynthesis. *J. Biol. Chem*. **285**, 24915–24924 (2010).

201. Thiericke, R., Zeeck, A., Nakagawa, A., Omura, S., Herrold, R. E., Wu, S. T. S., Beale, J. M. & Floss, H. G. Biosynthesis of the manumycin group antibiotics. *J. Am. Chem. Soc.* **112**, 3979–3987 (1990).

202. Gould, S. J., Melville, C. R. & Cone, M. C. 3-Amino-4-hydroxybenzoic acid is derived from the tricarboxylic acid cycle rather than the shikimic acid pathway. *J. Am. Chem. Soc.* **118**, 9228–9232 (1996).

203. Li, Y., Gould, S. J. & Proteau, P. J. Biosynthesis of 3-amino-4-hydroxybenzoic acid in *Streptomyces murayamaensis*: incorporation of [4-13C]oxalacetate. *Tetrahedron Lett.* **41**, 5181-5185 (2000).

204. Ohnishi, Y., Furusho, Y., Higashi, T., Chun, H.-K., Furihata, K., Sakuda, S. & Horinouchi, S. Structures of Grixazone A and B, a-factor-dependent yellow pigments produced under phosphate depletion by *Streptomyces griseus. J. Antibiot. (Tokyo)* **57**, 218–223 (2004).

205. Hsieh, C. J. & Jones, G. H. Nucleotide sequence, transcriptional analysis, and glucose regulation of the phenoxazinone synthase gene (phsA) from *Streptomyces antibioticus. J. Bacteriol.* **177**, 5740–5747 (1995).

206. Suzuki, H., Ohnishi, Y., Furusho, Y., Sakuda, S. & Horinouchi, S. Novel benzene ring biosynthesis from C3 and C4 primary metabolites by two enzymes. *J. Biol. Chem.* **281**, 36944-36951 (2006).

207. Rudolf, J. D., Dong, L.-B. & Shen, B. Platensimycin and platencin: Inspirations for chemistry, biology, enzymology, and medicine. *Biochem. Pharmacol*. **133**, 139–151 (2017).

208. Smanski, M. J., Peterson, R. M., Rajski, S. R. & Shen, B. Engineered *Streptomyces platensis* strains that overproduce antibiotics platensimycin and platencin. *Antimicrob. Agents Chemother*. **53**, 1299–1304 (2009).

209. McGuire, J. N., Wilson, S. R. & Rinehart, K. L. Cremeomycin, a novel cytotoxic antibiotic from *Streptomyces cremeus*. Structure elucidation and biological activity. *J. Antibiot.(Tokyo).* 48, 516–519 (1995).

210. Waldman, A. J., Pechersky, Y., Wang, P., Wang, J. X. & Balskus, E. P. The cremeomycin biosynthetic gene cluster encodes a pathway for diazo formation. *ChemBioChem* **16**, 2172–2175 (2015).

211. Sugai, Y., Katsuyama, Y. & Ohnishi, Y. A nitrous acid biosynthetic pathway for diazo group formation in bacteria. *Nat. Chem. Biol*. **12**, 73–75 (2016).

212. Waldman, A. J. & Balskus, E. P. Discovery of a diazo-forming enzyme in cremeomycin biosynthesis*. J. Org. Chem*. **83**, 7539–7546 (2018).

213. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A*. **99**, 3695–3700 (2002).

214. Barik, S. The uniqueness of tryptophan in biology: properties, metabolism, interactions and localization in proteins*. Int. J. Mol. Sci.* **21**, 8776 (2020).

1.17 Acknowledgements

Thank you to Professor Heinz Floss for the inspiration and foundational work in shikimate derived natural product biosynthesis. This material is being prepared for submission to *Natural Product Reports,* Bauman, K.D.,* Shende, V.V.,* Moore, B.S., 2022. The dissertation author was the primary co-researcher and co-author of this paper.

CHAPTER 2. Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry

2.1 Introduction and context for Chapter 2

Heterologous expression of biosynthetic gene clusters is a powerful strategy to link natural products to their producing gene clusters and vice versa. Connecting genes to chemistry in this way enables the discovery of both new molecules and new biosynthetic routes to construct structurally complex and bioactive chemical features.[1] Beyond discovery, heterologous expression of BGCs facilitates the genetic manipulation of pathways from genetically intractable organisms, circumvents culturing difficulties to allow access to molecules produced by unculturable and/or symbiotic organisms, and has become an essential strategy for improved production titers of small molecules. Advances in genomic sequencing and synthetic biology tool development have greatly facilitated this process. New capture methods,[2–4] expression vectors,[5] and heterologous hosts[6,7] are constantly adding to the heterologous expression toolbox, enabling easier routes to the cloning, movement, and manipulation of BGCs.

However, despite its benefits, heterologous production of secondary metabolites is not without challenges. Heterologously expressed pathways can have low productivity, and biosynthetic gene clusters may remain silent even within heterologous expression systems.[8] The reasons for this are varied, and could include an incompatibility of regulatory elements between the cloned cluster and the host, an inability to recruit essential building blocks, or failure to export the product.[9] So called “silent” BGCs are a common occurrence, both within native organisms and heterologous hosts. Additional genetic manipulations are often necessary to improve or even achieve production within heterologous hosts.[10,11] Refactoring biosynthetic gene clusters by replacing native genetic elements with well-characterized genetic parts (promoters, ribosome binding

sites, *etc*.) is a popular and often necessary strategy that decouples biosynthesis from its complex native regulation and establishes genetic control over a BGC. Continued development of these strategies is critical for understanding and controlling the biosynthesis of heterologously expressed small molecule natural products.

The work described in this chapter was published in Cell Chemical Biology in 2019 and describes the capture, heterologous expression, and genetic engineering of a phenazine (*phz*) biosynthetic gene cluster from a marine bacterium, *Streptomyces* sp. CNB091. Extensive culturing of this organism resulted in production of a series of compounds referred to as streptophenazines, phenazines containing a polyketide derived alkyl chain previously isolated from other *Streptomyces* strains.[12] Subsequent genome mining revealed a set of genes that encode for the production of phenazines colocalized with genes encoding polyketide and nonribosomal peptide biosynthesis. However, while streptophenazine type compounds had been previously isolated, no gene cluster had ever been linked to their production. Therefore, we took an in vivo, heterologous expression and genetic engineering based approach to link the predicted streptophenazine BGC (*spz)* to its molecular products, regulate the production of these compounds, and answer key questions about their biosynthetic origins.

Importantly, the work presented in this chapter explored this fundamental and persistent problem regarding the inconsistent or absent production of natural products by biosynthetic gene clusters discussed earlier. Upon capture of the *spz* gene cluster via Transformation Associated Recombination cloning and integration into a heterologous host the *spz* cluster proved silent within a heterologous environment. We addressed this issue with a synthetic biology strategy by building a series of plasmid-compatible

promoter cassettes and refactoring the *spz* cluster resulting, in the production of over 100 phenazine compounds. This work also enabled the development of a synthetic biology toolkit in the form of a series of promoter cassettes that can be used to refactor other gene clusters of interest.

This project was my introduction to the Moore lab, and perfectly showcases the breadth of skills required for natural products research, encompassing microbiology, molecular biology, analytical chemistry, and more. I learned how to work with *Streptomyces, Saccharomyces cerevisiae*, and *Escherichia coli,* I performed complex genetic engineering experiments, and I isolated and structurally characterized 15 natural products using HPLC, MS, NMR, and CD techniques. These skills were beneficial throughout my time in the lab, gave me immense confidence in my laboratory skills, and lay the groundwork for my future projects.

2.2 References for Chapter 2 Introduction

1. Zhang, J. J., Yamanaka, K., Tang, X. & Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol*. **621**, 87–110 (2019).
2. Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S. K., Martin, T. A. & Zhao, H. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination. *Nat. Commun*. **12**, 1171 (2021).
3. Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A.F., Zhang, Y. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res*. **46**, e28 (2018).
4. Jiang, W. Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., Zhu, T.F. Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun*. **6**, 8101 (2015).

5. Zhang, J. J., Tang, X., Zhang, M., Nguyen, D. & Moore, B. S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in gram-negative bacteria. *mBio* **8**, e01291-17 (2017).

6. Zhang, J. J., Moore, B. S. & Tang, X. Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **102**, 8437–8446 (2018).

7. Yang, Y., Lin, Y., Li, L., Linhardt, R. J. & Yan, Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metab. Eng*. **29**, 217–226 (2015).

8. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters*. Nat. Rev. Microbiol*. **13**, 509–523 (2015).

9. Gomez-Escribano, J. P. & Bibb, M. J. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 425–431 (2014).

10. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753–1760 (2009).

11. Wilkinson, B. & Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* **3**, 379–386 (2007).

12. Mitova, M. I., Lang, G., Wiese, J. & Imhoff, J. F. Subinhibitory concentrations of antibiotics induce phenazine production in a marine *Streptomyces* sp. *J. Nat. Prod*. **71**, 824–827 (2008).

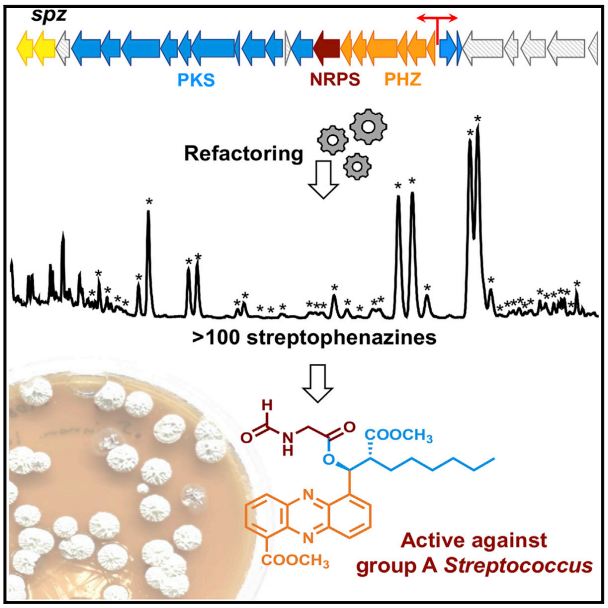
2.3 Reprint of “Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry”

Article

Cell Chemical Biology

Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry

Graphical Abstract



Authors

Katherine D. Bauman, Jie Li, Kazuya Murata, ..., Victor Nizet, Hanna Luhavaya, Bradley S. Moore

Correspondence

hluhavaya@ucsd.edu (H.L.), bsmoore@ucsd.edu (B.S.M.)

In Brief

Bauman et al. designed and utilized promoter cassettes to transcriptionally activate the cryptic streptophenazine biosynthetic pathway from marine *Streptomyces* bacteria. Pathway refactoring resulted in production of over 100 compounds, including streptophenazines with a rare *N*-formylglycine moiety that showed enhanced antibiotic activity.

Highlights

- Streptophenazine biosynthetic pathway (*spz*) was cloned from marine *Streptomyces*
- Refactoring the *spz* pathway led to tremendous chemical diversity (112 compounds)
- Streptophenazines with *N*-formylglycine moiety show potent antibiotic activity
- Discrete adenylation protein is required to produce *N*-formylglycine analogs

Bauman et al., 2019, Cell Chemical Biology *26*, 724–736
May 16, 2019
https://doi.org/10.1016/j.chembiol.2019.02.004

CellPress

Cell Chemical Biology
Article

Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry

Katherine D. Bauman,[1] Jie Li,[1,5] Kazuya Murata,[1,6] Simone M. Mantovani,[1,7] Samira Dahesh,[2] Victor Nizet,[2,3,4] Hanna Luhavaya,[1,*] and Bradley S. Moore[1,4,8,*]
[1]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA
[2]Department of Pediatrics, University of California at San Diego, La Jolla, CA, USA
[3]Collaborative to Halt Antibiotic Resistant Microbes, University of California at San Diego, La Jolla, CA, USA
[4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA
[5]Present Address: Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC, USA
[6]Present address: Faculty of Pharmacy, Kindai University, Osaka, Japan
[7]Present address: Amyris Pharmaceuticals, Emeryville, CA, USA
[8]Lead Contact
*Correspondence: hluhavaya@ucsd.edu (H.L.), bsmoore@ucsd.edu (B.S.M.)
https://doi.org/10.1016/j.chembiol.2019.02.004

SUMMARY

The disconnect between the genomic prediction of secondary metabolite biosynthetic potential and the observed laboratory production profile of microorganisms is well documented. While heterologous expression of biosynthetic gene clusters (BGCs) is often seen as a potential solution to bridge this gap, it is not immune to many challenges including impaired regulation, the inability to recruit essential building blocks, and transcriptional and/or translational silence of the biosynthetic genes. Here we report the discovery, cloning, refactoring, and heterologous expression of a cryptic hybrid phenazine-type BGC (*spz*) from the marine actinomycete *Streptomyces* sp. CNB-091. Overexpression of the engineered *spz* pathway resulted in increased production and chemical diversity of phenazine natural products belonging to the streptophenazine family, including bioactive members containing an unprecedented *N*-formylglycine attachment. An atypical discrete adenylation enzyme in the *spz* cluster is required to introduce the formylglycine moiety and represents a phylogenetically distinct class of adenylation proteins.

INTRODUCTION

Small molecules provide a fundamental means by which microorganisms interact with each other and their environment, facilitating diverse yet critical functions that can include defense, cell signaling, and/or nutrient acquisition (Demain and Fang, 2000). Nature has engineered a chemically diverse array of molecules designed to interact with a variety of biological targets and thereby mediate microbial communication. The potent biological activities that define secondary metabolites have been harnessed by humankind as well, most famously as sources of and inspiration for pharmaceuticals (Koehn and Carter, 2005; Newman and Cragg, 2016). The enzymes necessary for production of these metabolites are genetically encoded, and in bacteria are grouped together on contiguous stretches of DNA, forming biosynthetic gene clusters (BGCs) (Osbourn, 2010). The plummeting cost of high-throughput microbial DNA sequencing and resulting flood of genomic information has revealed that up to 90% of the chemical potential of microorganisms is unexplored as many BGCs are silent, particularly under laboratory conditions, and thus their molecular products have not been identified (Bentley et al., 2002; Omura et al., 2001).

Controlled heterologous expression is a powerful tool to unlock the biosynthetic potential encoded by silent and cryptic BGCs (Gomez-Escribano and Bibb, 2014). This process can be utilized to circumvent issues with culturing and difficulties with genetic manipulation of the native producers, allowing access to the molecules produced by unculturable and symbiotic organisms. However, the heterologous production of secondary metabolites can have low efficiency, and certain BGCs may remain silent even within validated heterologous expression systems (Rutledge and Challis, 2015). Additional genetic manipulations are often necessary to improve or activate production within heterologous hosts. These may include manipulating the pathway-specific regulatory genes (Olano et al., 2008) or using synthetic biology tools to rewrite native genetic elements with optimized components (e.g., promoters, ribosome binding sites [RBSs]) (Freestone et al., 2017; Montiel et al., 2015; Myronovskyi and Luzhetskyy, 2016).

Phenazines are a large class of heterocyclic, nitrogen-containing compounds produced by both Gram-positive and Gram-negative organisms (Laursen and Nielsen, 2004; Pierson and Pierson, 2010; Price-Whelan et al., 2006). Phenazine metabolites are derived from either phenazine-1-carboxylic acid (PCA) (**1**) or phenazine-1,6-dicarboxylic acid (PDC) (**2**), which are biosynthesized by a conserved set of five essential genes, *phz*BDEFG, typically encoded in a single operon (Blankenfeldt and Parsons, 2014; Mavrodi et al., 2010). While all phenazines share a dibenzo annulated pyrazine core structure, they exhibit incredible diversity in

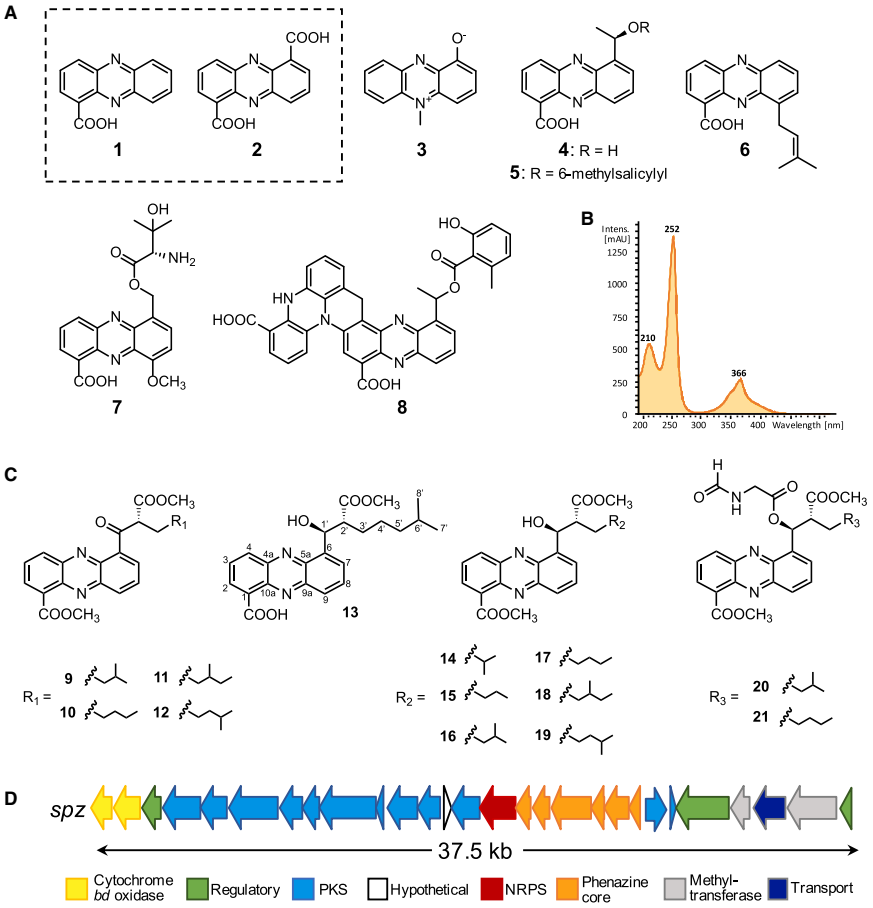


Figure 1. Chemical Structures of Selected Phenazine Metabolites and Gene Organization of the *spz* Biosynthetic Gene Cluster from *Streptomyces* sp. CNB-091

(A) Phenazine compounds isolated from *Streptomyces* and *Pseudomonas* bacteria: phenazine-1-carboxylic acid (PCA) (**1**), phenazine-1,6-dicarboxylic acid (PDC) (**2**), pyocyanin (**3**), saphenic acid (**4**), saphenamycin (**5**), endophenazine A (**6**), pelagiomicin A (**7**), and esmeraldin B (**8**).

(B) Characteristic UV absorption spectrum of streptophenazines (λ_{max} 252 nm and 366 nm).

(C) Chemical structures of streptophenazine metabolites associated with the *spz* BGC from *Streptomyces* sp. CNB-091. Previously reported compounds include streptophenazines A (**16**), B (**17**), C (**13**), D (**14**), F (**19**), and G (**18**). Compounds isolated in this work include oxo-streptophenazines A (**9**), B (**10**), G (**11**), and F (**12**), and streptophenazines P (**15**), Q (**20**), and R (**21**).

(D) Gene arrangement of the *spz* BGC from *Streptomyces* sp. CNB-091.

See also Figures S1, S2, and S5.

modifications of that core that result in potent and often therapeutically relevant bioactivities (Figure 1A). These activities can vary widely, including broad-spectrum antibiotic (saphenic acid, **4**), antitumor (esmeraldin B, **8**), and mosquito larvicidal (saphenamycin, **5**) activities (Geiger et al., 1988; Laursen and Nielsen, 2004; Rui et al., 2012). Phenazines play important ecological roles as well. For instance, pyocyanin (**3**) is an important signaling factor for its microbial producer *Pseudomonas aeruginosa* and a key facilitator of microbial survival in anoxic biofilms (Dietrich et al., 2006, 2008; Ramos et al., 2010). Because of the diverse and important bioactivities of phenazines, their structural similarity to cofactors in primary metabolism, and their abilities to modify gene expression, it has been suggested that phenazine metabolites blur the line between primary and secondary metabolism, and refute the paradigm that so-called secondary metabolites are not essential for survival (Price-Whelan et al., 2006).

Table 1. Open Reading Frames in the *spz* Biosynthetic Gene Cluster and Their Proposed Roles

Proposed Function	Protein	Amino acids	BLAST Closest Homolog	% Cover/Identity	Homolog Accession Number
Cytochrome *bd* ubiquinol oxidase II	Spz1	329	*Streptomyces* sp. ms184	100/98	WP_097875779.1
Cytochrome *bd* ubiquinol oxidase I	Spz2	439	*Streptomyces albovinaceus*	100/96	WP_086670717.1
LysR-type regulator	Spz3	293	*Streptomyces* spp.	100/97	WP_073774781.1
Acyl-CoA synthetase	Spz4	600	*Streptomyces albovinaceus*	100/98	WP_086670715.1
Crotonyl-CoA carboxylase/reductase	Spz5	416	*Streptomyces* spp.	100/99	WP_073774779.1
Type I PKS (KS[a], AT)	Spz6	800	*Streptomyces albovinaceus*	100/97	WP_086670713.1
Ketoreductase (KR)	Spz7	351	*Streptomyces* spp.	100/99	WP_073774774.1
Thioesterase (TE)	Spz8	264	*Streptomyces* sp. ms184	100/97	WP_097876860.1
Type I PKS (KS, ACP)	Spz9	914	*Streptomyces* sp. ms184	100/98	WP_097876858.1
Acyl carrier protein (ACP)	Spz10	109	*Kitasatospora purpeofusca*	95/61	WP_078880679.1
Aldehyde dehydrogenase	Spz11	481	*Streptomyces albovinaceus*	100/97	WP_086670706.1
PDC adenylase	Spz12	365	*Streptomyces* sp. ms184	100/98	WP_097876855.1
Hypothetical protein	Spz13	113	*Streptomyces* sp. TSRI0445	100/96	OKI73654.1
FAD-binding oxidoreductase	Spz14	455	*Streptomyces* spp.	100/98	WP_073776179.1
D-alanine-poly(phosphoribitol) ligase	Spz15	545	*Streptomyces* sp. ms184	100/98	WP_097877165.1
Pyridoxamine 5′-phosphate oxidase (PhzG)	Spz16	220	*Streptomyces* sp. ms184	100/99	WP_097877165.1
2,3-Dihydro-3-hydroxylanthranilate isomerase (PhzF)	Spz17	283	*Streptomyces* spp.	100/98	WP_073774764.1
Anthranilate synthase (PhzE)	Spz18	627	*Streptomyces* sp. ms184	99/98	WP_097877170.1
Isochorismatase (PhzD)	Spz19	207	*Streptomyces* sp. ms184	100/99	WP_097877172.1
Phospho-2-dehydro-3-deoxyheptonate aldolase (DAHP synthetase) (PhzC)	Spz20	391	*Streptomyces* sp. ms184	100/98	WP_097877172.1
Phenazine biosynthesis protein (PhzA/B)	Spz21	166	*Streptomyces* sp. ms184	100/98	WP_097877172.1
3-Oxoacyl-ACP synthase III (FabH)	Spz22	344	*Streptomyces* sp. ms184	100/99	WP_097877172.1
Aryl carrier domain protein	Spz23	97	*Kitasatospora purpeofusca*	92/61	WP_030398483.1
LuxR-type regulator	Spz24	846	*Kitasatospora purpeofusca*	96/55	WP_030398483.1
Class I SAM-dependent methyltransferase	Spz25	289	*Streptomyces* sp. TSRI0445	96/99	WP_079198635.1
MFS transporter	Spz26	516	*Streptomyces* sp. ms184	100/97	WP_097874479.1
HemK family methylase	Spz27	781	*Streptomyces kanasensis*	100/70	WP_058941165.1
TetR/AcrR regulator	Spz28	207	*Streptomyces griseus*	100/98	WP_030809832.1

[a]Lacking catalytic cysteine and histidine residues.

Here we report the identification, capture, engineering, and heterologous expression of a unique cryptic phenazine-type BGC (*spz*) from the marine actinomycete *Streptomyces* sp. CNB-091. Refactoring of the *spz* cluster resulted in tremendous structural diversity of the streptophenazine suite of natural products, PDC-derived phenazines with variable alkyl side chains isolated from *Streptomyces* bacteria (Bunbamrung et al., 2014; Kunz et al., 2014; Liang et al., 2017; Mitova et al., 2008). Our refactoring efforts resulted in the production of previously unseen *N*-formylglycine analogs with antibacterial activity.

RESULTS

Identification of the *spz* BGC

We previously sequenced the 8.2-Mb draft genome of the marine actinomycete *Streptomyces* sp. CNB-091 (GenBank: GCA_000377965.1) and identified 41 metabolically diverse BGCs, revealing the astounding biosynthetic potential of this organism (Ray et al., 2016). Beyond the previously characterized salinamides, a class of anti-inflammatory and antibacterial depsipeptides (Degen et al., 2014; Moore et al., 1999; Trischman et al., 1994), no other natural products have been reported from this strain. To explore the metabolic capacity of this organism, we cultured it in diverse growth conditions and monitored its molecular profile by liquid chromatography-mass spectrometry (LC-MS). We identified a group of orange pigmented molecules that were produced sporadically and at low titers. Comprehensive analysis of UV (Figure 1B) and high-resolution mass spectrometry (HRMS) revealed production of a series of streptophenazine-like compounds with accurate mass and fragmentation identical to previously characterized streptophenazines A (**16**), B (**17**), F (**19**), and G (**18**), first isolated from *Streptomyces* sp. HB202 (Mitova et al., 2008) (Figure 1C).

Inspection of the *Streptomyces* sp. CNB-091 genome revealed a single set of phenazine biosynthetic genes (*spz*16-21) colocalized with genes encoding the biosynthesis of polyketide extender unit(s) and polyketide synthase (PKS) extension cycle(s) (*spz*4-12) as well as nonribosomal peptide synthetase (NRPS) functions (*spz*15), suggestive of a putative streptophenazine BGC (*spz*) (Table 1, Figure 1D) (Weber et al., 2015). While the

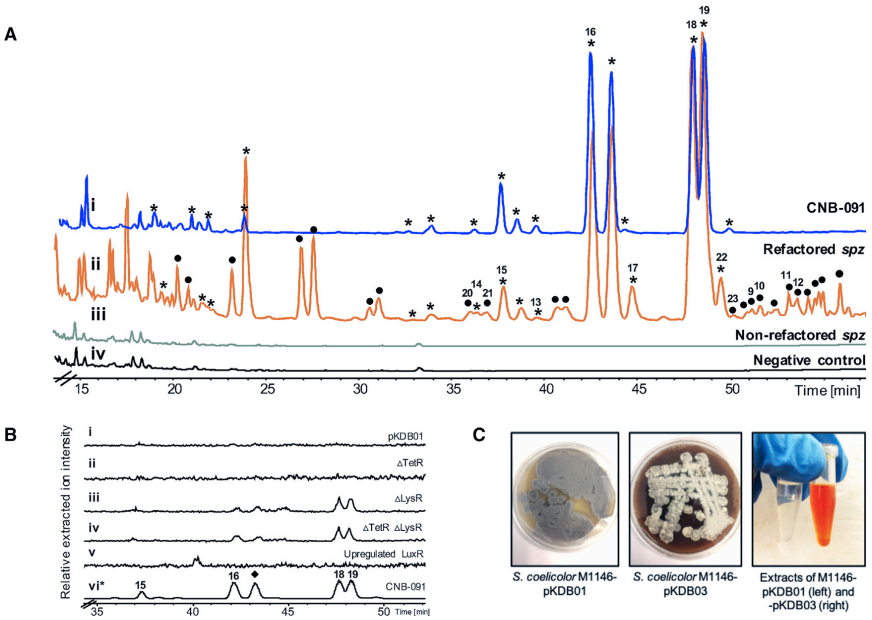


Figure 2. Analysis of the Crude Extracts from the Engineered *spz* Cluster in the Heterologous Host *Streptomyces coelicolor* M1146 and the Wild-Type Producer *Streptomyces* sp. CNB-091

(A) HPLC chromatograms (254 nm) of ethyl acetate extracts of: i, CNB-091, stars indicate peaks with streptophenazine UV profile; ii, M1146-pKDB03 (refactored *spz*), circles indicate peaks with streptophenazine UV profile appearing in refactored strain; iii, M1146-pKDB01 (non-refactored *spz*); iv, M1146-pCAP03 (negative control).

(B) Extracted LC-MS chromatograms (*m/z* 425.2, 439.2, corresponding to streptophenazines A, B, F, and G) of the organic extracts from regulatory gene mutant strains: i, pKDB01 (non-refactored *spz*); ii, pKDB02 (non-refactored *spz*ΔTetR); iii, pKDB01ΔLysR; iv, pKDB02ΔLysR; v, pKDB01-*ermE*p*-LuxR; vi, CNB-091 (*trace is minimized 100× for scale; diamond indicates peak that was not isolated).

(C) Production of streptophenazines on solid media by *S. coelicolor* M1146-pKDB03 and corresponding ethyl acetate extracts of strains with non-refactored (pKDB01) and refactored (pKDB03) *spz* cluster. Bright orange color is due to the abundant production of streptophenazines.

See also Figures S2 and S3.

presence of the PKS genes in the *spz* cluster can be rationalized by the installation of C-6 alkyl chains in streptophenazines (Figure 1C), there have been no previously reported streptophenazine structures that incorporate amino acid(s), indicative of NRPS biochemistry. Therefore, we speculated that the *spz* BGC may encode the biosynthesis of streptophenazine-like molecules with different chemical scaffolds. We thus set out to interrogate the *spz* BGC with the intent of not only exploring its full biosynthetic potential but also to probe issues of BGC expression that might enable the future discovery of additional metabolites from *Streptomyces* sp. CNB-091.

Transformation-Associated Recombination Cloning and Heterologous Expression of the *spz* BGC

To exploit the full biosynthetic potential of the *spz* BGC and facilitate genetic manipulations, we captured the *spz* pathway using transformation-associated recombination (TAR) cloning methodology (Kouprina and Larionov, 2016; Tang et al., 2015; Yamanaka et al., 2014). Initially, a 49.5-kb DNA region covering the proposed *spz* BGC was captured from *Streptomyces* sp. CNB-091 genomic DNA, resulting in an approximately 60.3-kb construct, pSMM. Restriction digest was used to verify correct assembly of the construct isolated from yeast (Figure S1). However, pSMM proved to be unstable upon transformation and manipulation in *Escherichia coli*, limiting its use for future genetic alterations. Inspection of peripheral genes of the DNA insert captured into pSMM revealed almost 12 kb of DNA sequence unlikely associated with streptophenazine biosynthesis. Therefore, the *spz* BGC was recaptured using a PCR-based TAR method (Shao et al., 2009) targeting strictly genes *spz*1-28. Eight approximately 5-kb DNA fragments covering the *spz* pathway were amplified and purified for further TAR cloning into the pCAP03 capture vector (Tang et al., 2015). The resulting construct, pKDB01, remained stable upon transformation into *E. coli* (Figure S1). We then transferred pKDB01 into *E. coli* ET12567 (MacNeil et al., 1992) for further conjugation into the *Streptomyces coelicolor* M1146 heterologous host and chromosomal integration into the ϕC31 attachment site (Gomez-Escribano and Bibb, 2011; Jones et al., 2013). The resulting strain was cultured under a variety of growth conditions,

but we ultimately failed to detect any streptophenazines in its metabolic profile (Figure 2).

While disappointing at first, the silence of the *spz* BGC within *S. coelicolor* M1146 presented an ideal opportunity to explore issues of expression within heterologous host systems and apply genetic tools to activate a silent pathway. The *spz* BGC is an excellent model system to probe these questions for the following reasons: (1) phenazine metabolites are pigmented, facilitating visual detection and aiding purification processes and (2) streptophenazine-like compounds were detected in the native producer, which simplifies the search for these molecules within the heterologous system. To awaken the *spz* cluster in the *S. coelicolor* M1146 host strain, we chose to harness synthetic biology tools (Kim et al., 2015) to establish genetic control over the expression and regulation of the *spz* biosynthetic genes. Ultimately, we predict that these tools can be applied to other BGCs from *Streptomyces* sp. CNB-091, allowing access to the entire biosynthetic potential of this organism.

Genetic Manipulation of the Pathway-Specific *spz* Regulatory Genes

Three putative pathway-specific regulatory genes are found within the *spz* BGC: *spz*28 and *spz*3, encoding TetR and LysR homologs, respectively, which are hypothesized to be repressors, and *spz*24, which encodes a LuxR homolog predicted to be an activator of phenazine biosynthesis. To examine the role of the TetR homolog, we created a construct, pKDB02, with the *spz* pathway lacking the *spz*28 gene (Figure S1). The cosmid was then integrated into the genome of *S. coelicolor* M1146, but the profile of this mutant strain did not reveal any streptophenazine metabolites (Figure 2B). Using λ-Red recombineering, we then replaced the second putative repressor gene, *spz*3, with an apramycin resistance cassette in both plasmids, pKDB01 and the TetR deletion mutant pKDB02 (Figure S2). The resulting strains *S. coelicolor* M1146-pKDB01Δ*spz*3 and the double mutant -pKDB02Δ*spz*3 had identical metabolite profiles, characterized by trace production of four streptophenazine compounds (Figure 2B). However, this production was inconsistent and observed at levels much lower than in the native organism. This observation suggested that the LysR-homolog (Spz3) may serve as a repressor of the *spz* pathway. Finally, we introduced the well-characterized strong constitutive promoter *ermE*p* (Bibb et al., 1985) in front of *spz*24 (Figure S2). The resulting cosmid pKDB01-*ermE*p-spz*24 was then integrated into the genome of *S. coelicolor* M1146. LC-MS analysis revealed that the upregulation of the *spz* LuxR homolog did not activate streptophenazine production (Figure 2B). Ultimately, genetic manipulations of the *spz* regulatory genes failed to achieve reliable production of streptophenazines in the heterologous host.

Refactoring the *spz* Pathway with Strong Constitutive Promoters

We performed RT-PCR to test expression of every gene in the cloned *spz* pathway integrated into the *S. coelicolor* M1146 genome. Results revealed that the *spz* cluster was not transcriptionally active in a heterologous environment (Figure 3A), which strongly suggests a malfunction with the cluster's native promoters. Therefore, we decided to bypass the native regulation of the pathway and refactor the *spz* BGC by introducing characterized, constitutively active, actinomycete-compatible promoters in four locations based on the operonic structure of the cluster (Figure 3B). We created three cassettes (two unidirectional and one bidirectional) that harbor selection markers, promoters, and RBS sequences. Importantly, all cassettes were designed to lack *oriT* that is present in readily available λ-Red cassettes (Gust et al., 2003, 2004), to be compatible with the pCAP series of vectors that already harbor an *oriT* site (Yamanaka et al., 2014).

We designed the first cassette, the *sp44-p21* cassette, to include two promoters, *sp44* and *p21*, facing opposite directions. The *sp44* promoter is a synthetic derivative of the *kasO** promoter from the coelimycin BGC and was shown to be 20-fold stronger than the *ermE** promoter, traditionally used in *Streptomyces* research (Bai et al., 2015). The *sp44* promoter was introduced in front of the phenazine operon to allow robust production of the PDC core for further biosynthetic derivatization. The *p21* promoter is a derivative of *ermEp1*, with 3-fold greater strength than *ermE*p* (Myronovskyi and Luzhetskyy, 2016; Siegl et al., 2013). This cassette also included the apramycin resistance gene *aac(3)IV* sandwiched between two FRT sites to allow excision of the resistance marker after a successful recombination event. Eight additional nucleotides were added between the *sp44* and FRT sequences to ensure an in-frame nonpolar deletion. The second cassette, named *actIp*, was designed to govern transcription of the PKS genes (*spz*4–12). This cassette included the *actI* promoter, RBS, the *actII-ORF4* gene required for activation of the *actI* promoter from the actinorhodin BGC (Fayed et al., 2015; Fernández-Moreno et al., 1991; Rowe et al., 1998), and the *aac(3)IV* gene to leverage apramycin resistance for selection. Finally, the third cassette, *ermE*p*, was used to introduce the *ermE** promoter in front of the cytochrome *bd* oxidase-encoding genes (*spz*1, 2). To avoid cross-resistance, we included the ampicillin resistance gene *bla* to allow for selection in *E. coli*. Gibson assembly was used to stitch together functional elements as discussed in STAR Methods. Cassette sequences were confirmed by restriction digest and Sanger sequencing (Table S1). These cassettes were then used to replace putative native promoter regions within the *spz* cluster in the pKDB01 cosmid in a stepwise manner using λ-Red recombineering (Figure S2B). Insertion of the *sp44-p21* cassette followed by excision of the apramycin resistance marker resulted in construct pKDB03. The latter was used for insertion of the *actIp* cassette to give construct pKDB04, and subsequent insertion of the *ermE*p* cassette yielded pKDB05. Each version of the refactored cosmid was conjugated into *S. coelicolor* M1146 for comparative metabolite profiling.

We performed an analogous RT-PCR experiment, as was done prior to refactoring, on RNA isolated from *S. coelicolor* M1146-pKDB03 carrying the refactored *spz* pathway. Interestingly, insertion of the bidirectional promoter cassette alone was sufficient to activate transcription of all biosynthetic genes within the cluster (Figure 3A).

Streptophenazine Production by the Refactored *spz* BGC

Streptomyces sp. CNB-091 (positive control) and *S. coelicolor* M1146 with integrated pCAP03 (negative control), pKDB01 (non-refactored *spz*), and pKDB03, -04, and -05 (refactored versions of *spz*), were grown for comparative metabolomics. After

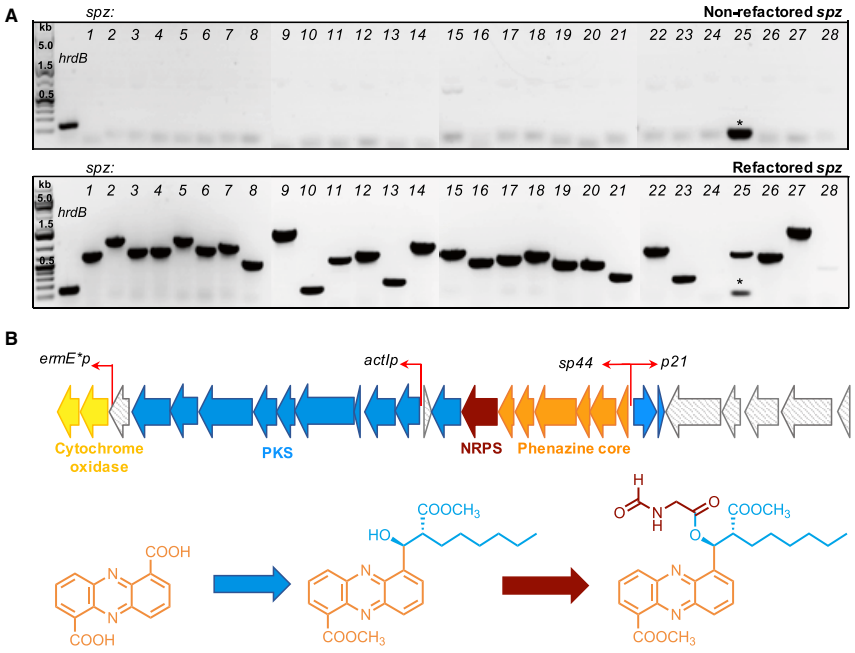


Figure 3. Refactoring the *spz* Biosynthetic Gene Cluster

(A) RT-PCR analysis of expression of biosynthetic genes in the *spz* BGC before and after refactoring. Gene encoding a sigma-like transcription factor of *S. coelicolor*, *hrdB*, was used as a positive control. Strains *S. coelicolor* M1146-pKDB01 (non-refactored *spz*) and *S. coelicolor* M1146-pKDB03 (refactored *spz*) were used for analysis. Expected sizes (bp) of amplified fragments are as follows: *hrdB*, 132, *spz* genes: *spz*1, 679; 2, 1135; 3, 731; 4, 750; 5, 1093; 6, 800; 7, 946; 8, 544; 9, 1134; 10, 180; 11, 587; 12, 689; 13, 268; 14, 963; 15, 768; 16, 532; 17, 611; 18, 722; 19, 515; 20, 540; 21, 328; 22, 727; 23, 242; 24, 1339; 25, 667; 26, 623; 27, 1499; 28, 410. Asterisk indicates low molecular weight band corresponding to primer dimer. Expression of the gene *spz*24 (encodes LuxR-type regulator) was not observed for both versions of the *spz* cluster, refactored and non-refactored.

(B) Schematic illustration of the refactoring strategy indicating locations and directions of the promoters introduced (red arrows).

See also Figure S2 and Table S1.

two days of culturing the production media turned dark orange for the strains only carrying the refactored *spz* BGC, yielding bright orange organic extracts (Figure 2C). Analysis of LC-MS data of these pigmented extracts from *S. coelicolor* M1146-pKDB03, -04, and -05 revealed intense production of at least 38 streptophenazines, as corroborated by HRMS, UV, and tandem MS/MS fragmentation (Figures 2A and S3; Table S2). This chemical bounty included the 16 streptophenazine compounds previously detected in *Streptomyces* sp. CNB-091 (Figure 2A), but with increased production levels, as judged by LC-MS analysis. Importantly, the production profile proved to be consistent and achievable under a variety of growth conditions, which speaks to the advantage of genetic control. No changes in chemical diversity or production titers from *S. coelicolor* M1146-pKDB03, -04, and -05 strains were observed (Figure S3). Therefore, it was decided to continue experiments with *S. coelicolor* M1146-pKDB03, the strain possessing the *spz* pathway with the fewest genetic alterations.

We next used molecular networking to compare *S. coelicolor* M1146-pKDB01, M1146-pKDB03, CNB-091, and the negative control (M1146-pCAP03) to identify analogs that might have been neglected in the initial visual analysis (Figure S4) (Wang et al., 2016). Immediately, the cluster of nodes with molecular ions corresponding to the streptophenazines stood out (Figure 4). Each node in the putative streptophenazine cluster was inspected for association with streptophenazines by manual analysis of the corresponding MS/MS spectra, UV, and accurate mass (Table S2). The network identified 72 nodes corresponding to streptophenazine-type compounds; 67 of the nodes were found within the refactored strain (pink and blue nodes), 34 of which were found in both the refactored strain and in *Streptomyces* sp. CNB-091 (blue nodes). Five nodes were identified only in *Streptomyces* sp. CNB-091 (orange nodes), and one node included in the cluster (gray) was found in the negative control and is likely to be an artifact of the algorithm. Using the molecular network as a guide, manual inspection of the LC-MS data ultimately revealed 112 streptophenazines produced by the refactored *spz* BGC. In most cases, extracting the LC-MS data for the precursor mass found in the network resulted in more peaks than nodes in the network. For example, extracting

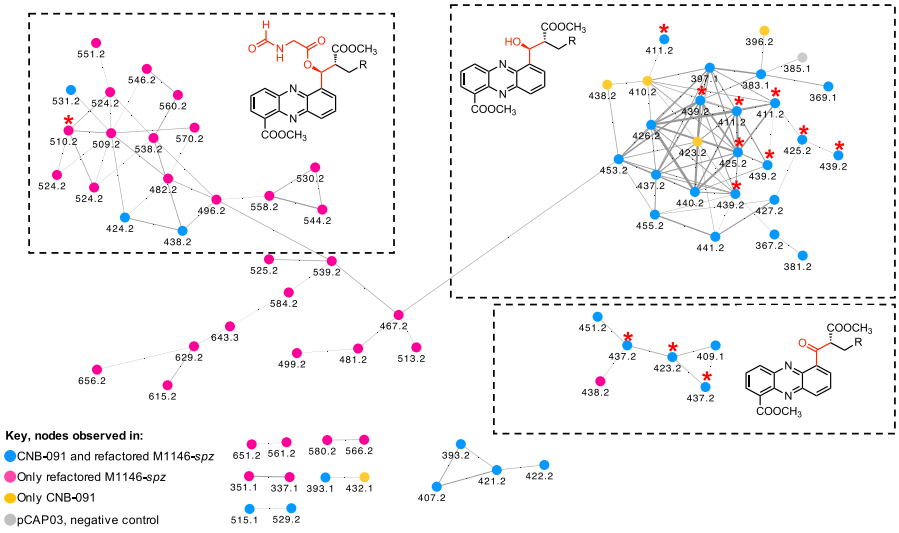


Figure 4. Streptophenazine Cluster of Nodes from the Molecular Network

MS/MS data of extracts of CNB-091, M1146-pKDB01 (non-refactored *spz*), M1146-pKDB03 (refactored *spz*), and pCAP03 (negative control) are included and color coded as per key. All nodes are labeled with the corresponding precursor ion mass. Isolated compounds are marked with red stars. See also Figures S3 and S4; Table S2.

LC-MS data for *m/z* 411.2 revealed eight peaks corresponding to eight different streptophenazines rather than the three nodes seen in the network. Several pairs of nodes (e.g., *m/z* 351.1 and 337.1) appeared separately in the molecular network and did not form clusters with other nodes corresponding to streptophenazines. This finding can be explained by low abundances of these masses in the LC-MS data, which resulted in MS/MS fragment spectra with additional peaks from chemical noise, causing the algorithm to create separate clusters (Caraballo-Rodríguez et al., 2017). Table S2 summarizes all streptophenazine products of the refactored *spz* BGC and contains predicted molecular formulas for compounds where appropriate.

Interestingly, a number of nodes with even mass number were present in the network (e.g., *m/z* 510.2, 524.2, and 538.2), which suggested an additional nitrogen atom in the chemical formula. These molecular ions were present only in the sample from the refactored *spz* pathway. We identified three isomeric compounds with *m/z* 510.2. The early-eluting compound (31.8 min) showed a distinctive loss of 117 Da, while the latter two isomers (36.2 and 37 min) instead had identical losses of 103 Da. These fragments suggested formylalanine and formylglycine attachments, respectively, with different methylation patterns at the C-1 carboxylic group (STAR Methods). Such modifications are reasonable as the *spz* pathway contains the gene *spz*15 that encodes a putative adenylation protein with predicted specificity for alanine.

Isolation and Characterization of Streptophenazines

Production cultures of *S. coelicolor* M1146-pKDB03 were scaled up to 6 L to enable isolation and structural elucidation of observed streptophenazines (STAR Methods). Ethyl acetate extraction of cultures resulted in 8 g of crude mixture, which underwent multiple rounds of flash chromatography (C18 resin and silica resin). Final purification of the targeted compounds was achieved by reversed-phase preparative and/or semipreparative high-performance LC (HPLC). Ultimately, we isolated 15 streptophenazine analogs.

For tables with NMR assignments, spectra, and key correlation spectroscopy (COSY) and heteronuclear multiple bond correlation (HMBC) for all isolated compounds, see Data S1. Analysis of ^{1}H NMR and 2D spectra revealed that isolated compounds **16**, **17**, **13**, **14**, **19**, and **18** correspond to reported streptophenazines A, B, C, D, F, and G, respectively. To confirm stereochemistry, we performed circular dichroism (CD) measurements for compound **18** and compared results with the literature data for streptophenazine G with established 1′*S*,2′*R* absolute configuration (Kunz et al., 2014; Liang et al., 2017; Yang et al., 2012) (Figure S5A). Due to the proximity of the C-1′ stereocenter to the phenazine chromophore, this chiral center will affect the absorption of circularly polarized light and generate a characteristic shift in optical rotation (Cotton effect) evident in a CD spectrum. A strong negative Cotton effect at 251 nm, which mirrored that of previously characterized streptophenazine G, confirmed the *S* conformation of the C-1′ stereocenter of **18**. ^{1}H NMR shifts of compound **18** were identical to those reported in the literature for streptophenazine G, indicating that the absolute configuration of compound **18** was also 1′*S*,2′*R*. Based on the CD result and the fact that stereochemistry of these centers remains unchanged for all reported streptophenazines, we propose the same stereochemistry for all compounds isolated in

this work. To further support the stereochemistry at C-1′, we performed bioinformatics analysis of the ketoreductase (KR) protein Spz7, which we presume catalyzes the stereospecific reduction of the β-keto intermediate to the 1′-hydroxyl group. KR domains of modular PKS enzymes define the configuration at α- and β-carbons, and conserved amino acid motifs that control the stereochemical outcome of this reduction have been defined (Caffrey, 2003; Keatinge-Clay, 2007; Smith and Tsai, 2007). While Spz7 shows relatively low amino acid sequence similarity to modular KR domains (Figure S5B), we identified the LDD amino acid motif characteristic of type B KR domains, which would generate the 1′*S* stereocenter in streptophenazines, consistent with the established stereochemistry.

Compounds isolated in this work included **9**, **10**, **11**, and **12**, which were individually isolated as orange powders. HRMS analysis showed that **9** and **10** share the same molecular formula $C_{24}H_{27}N_2O_5$ (*m/z* 423.1893 $[M + H]^+$ and *m/z* 423.1925 $[M + H]^+$, respectively), while compounds **11** and **12** have molecular formula $C_{25}H_{29}N_2O_5$ (*m/z* 437.2075 $[M + H]^+$ and *m/z* 437.2087 $[M + H]^+$, respectively). These four compounds appeared to be derivatives of streptophenazines A, B, G, and F, respectively, but were missing the C-1′ oxymethine low field shift ($\delta_C \sim 72$) seen in previously characterized streptophenazines. Instead, HMBC correlations revealed a carbon shift indicative of a ketone carbonyl ($\delta_C \sim 198.3$) that showed correlations to both the phenazine core and the alkyl side chain. We suggest the name oxo-streptophenazine A for compound **9**, oxo-streptophenazine B for compound **10**, oxo-streptophenazine G for compound **11**, and oxo-streptophenazine F for compound **12**.

Compound **15** ($C_{23}H_{27}N_2O_5$, *m/z* $[M + H]^+$ 411.1912) shares a 2D structure with streptophenazine D, but terminates in a single methyl group rather than a branched chain (Figure 1D). Paralleling the established nomenclature of streptophenazines, we propose the name streptophenazine P for this compound. Additionally, compounds **22** and **23** had NMR spectra that appeared almost identical to streptophenazines G and F; however, the proton NMR signals associated with C-3′, C-2′, the C-2′ carbomethoxy group (2′-COO**CH$_3$**), and C-1′ were slightly shifted (Data S1), which indicated that compounds **22** and **23** were diastereomers at the C-2′ position of streptophenazines G and F, respectively.

Two of the three detected compounds with *m/z* 510.2 were isolated (**20** and **21**), both with MS/MS fragmentation patterns suggestive of a formylglycine moiety adjoined to the streptophenazine core structure, most plausibly via the hydroxyl group at C-1′. The molecular formula $C_{27}H_{32}N_3O_7$ was established for compounds **20** and **21** (*m/z* $[M + H]^+$ 510.2236 and 510.2234, respectively). NMR spectra of these compounds contained a signal (δ_H 8.24, δ_C 161), suspected to be the formyl group, which showed COSY correlations to two protons (δ_H 4.22, 4.18) with a carbon shift of δ_C 40.25, as well as a COSY correlation to a broad peak δ_H 6.15 with no corresponding carbon shift in the heteronuclear single-quantum coherence (HSQC) shift (amide proton) (Data S1). Two doublets (δ_H 4.22, 4.18) showed key HMBC correlations to two carbonyl groups (δ_C 161 and 168) and COSY correlations to the amide and formyl protons, establishing their position as the glycine α-protons, thus fully confirming the presence of the *N*-formylglycine moiety. Compound **20** was found to terminate the alkyl chain in two branched methyl groups while compound **21** terminated the chain in a single methyl group (Figure 1D). We propose the names streptophenazine Q and streptophenazine R, respectively, for compounds **20** and **21**.

Antimicrobial Activity of Streptophenazines

A series of four streptophenazines, **9**, **13**, **16**, and **20**, were tested for antibiotic activity against two Gram-positive organisms (group A *Streptococcus* [GAS] and methicillin-resistant *Staphylococcus aureus* [MRSA] TCH1516) and two Gram-negative organisms (*Acinetobacter baumannii* 5075 and *Klebsiella pneumoniae* 1100). The four test compounds were selected because they share the same alkyl chain but differ in the C-1′ substituent (keto, hydroxyl, or formylglycine) and methylation state of the C-1 carboxylic group. Results showed that compounds **9**, **13**, and **16** had minimum inhibitory concentration (MICs) higher than their tested concentrations, indicating weak or absent antimicrobial activity (Table S3). However, compound **20**, the *N*-formylglycine analog, showed significantly more potent activity than the hydroxyl or ketone-containing analogs against GAS (MIC 2.5 μg/mL). Compound **20** also inhibited the growth of MRSA and the tested Gram-negative pathogens, while no such activity was observed for the other analogs. This set of structurally similar compounds allows us to infer structure-activity relationships of streptophenazines, and these data suggest that the formylglycine moiety leads to superior bioactivity.

Spz15: A Unique Stand-Alone Adenylation Protein Required for Formylglycine Modification

Arguably the most interesting aspect of the *spz* biosynthetic pathway is the production of the *N*-formylglycine-containing streptophenazines Q and R (**20**, **21**). Although common in ribosomal protein biosynthesis, *N*-formylation is a rare biosynthetic tailoring modification in NRPS-synthesized peptides, famously performed in linear gramicidin A by a specific formylation domain (Kessler et al., 2004; Schoenafinger et al., 2006). Bioinformatics analysis suggests that *spz*15 encodes an adenylation protein, and identification of the GXXGXPKG motif for a phosphate binding loop and AMP binding Pfam domain (PF00501) reinforced this hypothesis (Figure S6) (Perego et al., 1995). However, neither a canonical condensation domain nor a formylation domain is present in the *spz* cluster. The lack of a traditional suite of NRPS domains suggested that perhaps Spz15 is solely responsible for the addition of the *N*-formylglycine moiety. To prove involvement of the Spz15 protein in the biosynthesis of the amino acid-containing compounds, we deleted the *spz*15 gene. The resulting mutant construct (pKDB03Δ*spz*15) was confirmed by restriction digest (Figure S6) and integrated into the genome of the heterologous host. LC-MS analysis of organic extracts of the mutant strain showed no signs of streptophenazines containing the *N*-formylglycine moiety. However, production of all other streptophenazines remained unchanged, ruling out polar effects on downstream genes in the *spz*15 deletion mutant (Figure S6).

Phylogenetic analysis of Spz15 revealed that this enzyme is strikingly distinct from canonical adenylation domains and instead forms a separate evolutionary clade (Figure 5). Analysis of genomes containing Spz15-like proteins revealed that these putative adenylation proteins are all found adjacent to phenazine biosynthesis genes (Figure S7). This arrangement indicates that

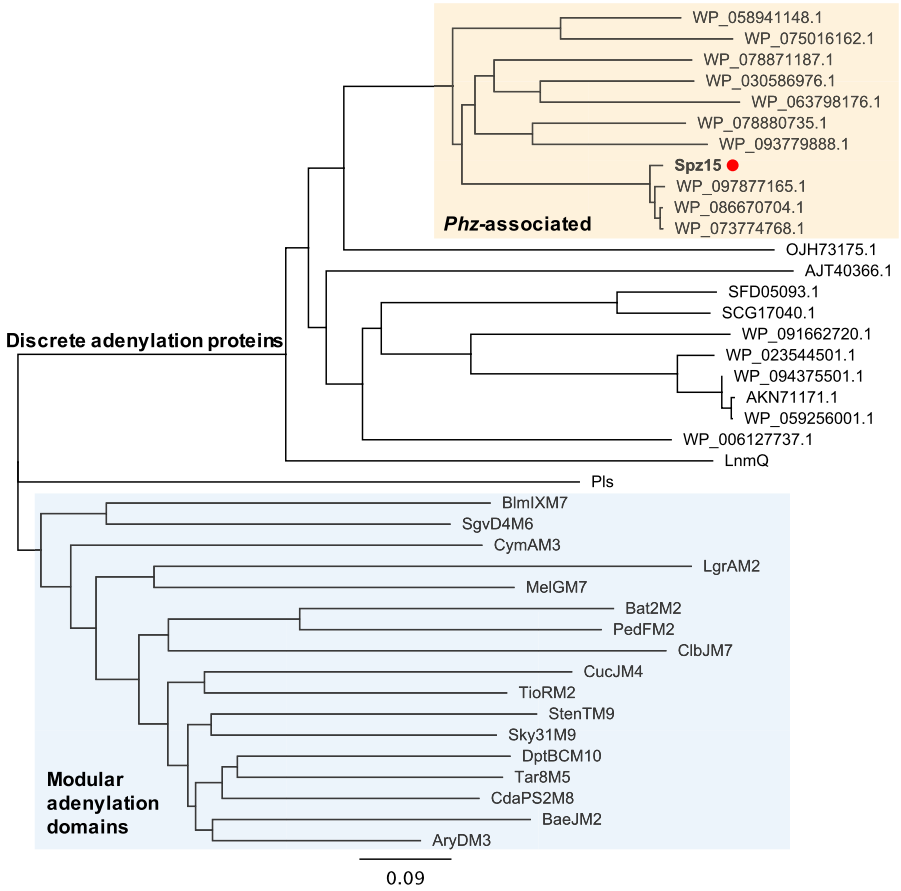


Figure 5. Phylogenetic Analysis of Spz15 Adenylation Protein

Phylogenetic tree of Spz15 with characterized adenylation domains and NCBI blastp homologs of Spz15 was constructed using Geneious software. Spz15 (marked with orange dot) forms a separate clade from canonical modular characterized adenylation domains (highlighted in blue). Proteins that clade with Spz15 (highlighted in orange) are found within putative phenazine-type (*phz*) BGCs. The remaining protein sequences are discrete stand-alone adenylation enzymes. Protein abbreviations are as follows: LnmQ, leinamycin; Pls, ε-poly-L-lysine; NRPS: BlmIXM7, bleomycin module 7; SgvD4M6, griseoviridin module 6; CymAM3, cyclomarin module 3; LgrM2, linear gramicidin module 2; MelGM7, melithiazol module 7; Bat2M2, batumin module 2; PedFM2, pederin module 2; ClbJM7, colibactin module 7; CucJM4, cupriachelin module 4; TioRM2, thiocoraline module 2; StenTM, stenothricin module 9; Sky31M9, skyllamycin module 9; DptBCM10, daptomycin module 10; Tar8M5, taromycin module 5; CdaPS2M8, calcium-dependent antibiotic module 8; BaeJM2, bacillaene module 2; AryDM3, arylomycin module 3. All sequences were downloaded from the NCBI protein database. See also Figures S6 and S7.

*spz*15 may represent a genetic hook into unusual hybrid phenazine-type biosynthetic pathways. Biochemical characterization and investigation of the role of Spz15 in this unusual addition of an *N*-formylglycine moiety to streptophenazine molecules are currently ongoing.

DISCUSSION

The discovery of the unique hybrid PHZ/PKS/NRPS BGC encoded in the genome of marine *Streptomyces* sp. CNB-091 uncovered a pathway that involves unusual biosynthetic transformations and yields a tremendous wealth of chemical products. Fundamentally, the *spz* BGC is arranged as a series of subclusters, each responsible for a distinct aspect of the biosynthesis of the final molecules, evoking the concept of a "cluster of clusters" (Figure 3B). A set of contiguous genes (*spz*16–21) is responsible for the biosynthesis of the PDC core, which serves as the starter unit for a single PKS extension with an unusual long branched-chain malonyl-coenzyme A (CoA) constructed by a crotonyl-CoA carboxylase (CCR)-like mechanism (Figure 6) (Wilson and Moore, 2012). To

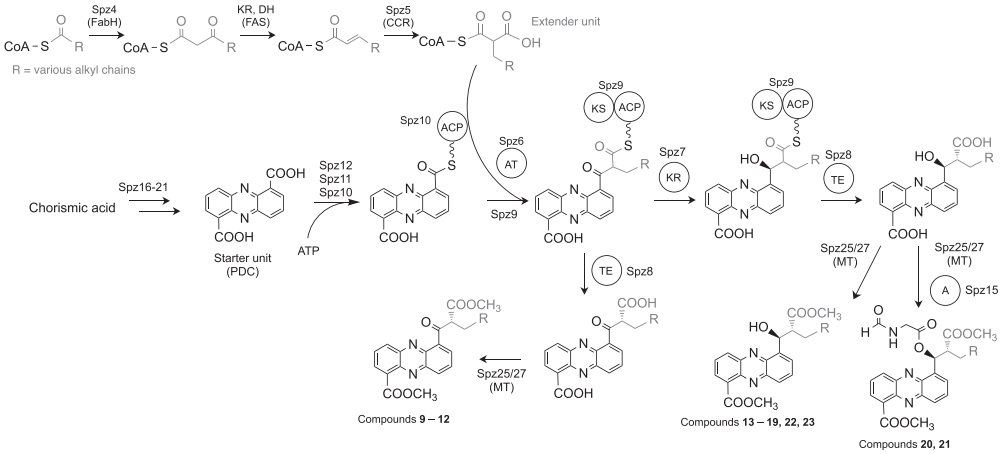


Figure 6. Proposed Biosynthesis of Streptophenazines from *Streptomyces* sp. CNB-091

R denotes alkyl substituents for the side chain. A variety of polyketide extender units arise from diverse substituted malonyl-CoA derivatives from primary metabolism through catalysis by a crotonyl-CoA carboxylase/reductase (Spz5).

date there is only one other example of a PKS extension of a phenazine core substrate, described in the biosynthesis of the esmeraldins and saphenamycin phenazine metabolites (Rui et al., 2012).

A phylogenetically distinct adenylation enzyme (Spz15), transcribed from the same operon as the *spz*16–21 genes, is unambiguously involved in the addition of an *N*-formylglycine moiety to streptophenazines, generating unusual hybrid molecules with potent bioactivities (Figure S6). Finally, two genes at the west boundary of the cluster (*spz*1, 2) encode cytochrome *bd* oxidase enzymes, which may represent a resistance mechanism for the phenazine-producing strain *Streptomyces* sp. CNB-091 (Voggu et al., 2006). The *spz* BGC represents unique biosynthetic machinery that does not fit neatly within any biosynthetic systems described, and therefore opens the door for future mechanistic biochemical studies (Figure 6).

Rewriting native genetic elements in the *spz* gene cluster with optimized promoters and RBS sequences allowed us to circumvent issues associated with the complex network of native regulation and ultimately gain mastery over this unique biosynthetic system. Interestingly, inserting just the first bidirectional promoter cassette resulted in activation of transcription of all *spz* biosynthetic genes (Figure 3). It is possible that production of the PDC is required to trigger expression of the remaining biosynthetic genes (“self-induction”). Certain small molecules can bind to specific DNA regions, resulting in activation of transcription of the downstream genes (Chalabaev et al., 2008; Denison and Kodadek, 1998). Alternatively, the strength of the *sp44* promoter could be sufficient to activate transcription of the genes beyond the phenazine-encoding operon, and there is a precedent of a single promoter driving transcription of multiple genes, spanning over 10 kb (Amagai et al., 2017; Luo et al., 2015).

Ultimately, refactoring the *spz* pathway produced an incredible wealth of chemical diversity (112 streptophenazines) that can serve as a library of compounds, much like a combinatorial chemist might produce. This diversity-oriented biosynthesis is generally considered an evolutionary advantage (Fischbach and Clardy, 2007), as diverse analogs may have different biological targets or act synergistically as an antibiotic cocktail (Wu and Seyedsayamdost, 2017). This may explain the lack of production of the *N*-formylglycine analogs by the wild-type strain; bioactivity testing revealed that these compounds did possess distinct bioactivity, and thus may be produced only under certain environmental and/or physiological stress. Therefore, amino acid-containing streptophenazines are excellent starting points for further chemical derivatization with the goal of ultimately achieving superior antibiotic properties.

The problem regarding inconsistent or absent production of secondary metabolites is a common occurrence, within both native organisms and heterologous hosts, and the disconnect between the genetically encoded biosynthetic potential of an organism and the actual suite of compounds produced is immense. The issue addressed here is not just successful activation of cryptic BGCs, but exploitation of the full range of capabilities housed by these pathways. Through engineering, we were able to achieve production of broad chemical diversity, including compounds not seen in the native organism that contain an unprecedented formylglycine moiety that imparts enhanced antibiotic activity. This example suggests that even transcriptionally active clusters may benefit from refactoring.

SIGNIFICANCE

This work presents the discovery, engineering, and heterologous expression of the hybrid streptophenazine (*spz*) biosynthetic gene cluster from marine *Streptomyces* that unites phenazine, polyketide, and nonribosomal peptide biochemistry in an unprecedented expansion of the biosynthetic and

chemical logic of the phenazine class of metabolites. We developed a set of promoter cassettes compatible with the widely used pCAP series of shuttle vectors to successfully link the cryptic *spz* BGC to its molecular products, including *N*-formylglycine analogs that possess superior antibiotic activity. We identified one gene, *spz*15, which encodes a discrete adenylation protein as a genetic search hook into analogous systems, and propose its involvement in the attachment of the rare *N*-formylglycine moiety. This discovery encourages future work to reveal the mechanism behind this biosynthetic transformation. Complete understanding of the biosynthesis of diverse streptophenazines opens an opportunity for engineering to improve the bioactivities of these compounds. Ultimately, using biology to inform chemistry in this way to reliably regulate expression of BGCs is fundamental in discovering, understanding, and governing the chemistry of our natural world.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - General Methods
  - Bacterial Strains and Growth Conditions
  - Software and Tools Used
  - Preparation of Genomic DNA and PCR Fragments for TAR Cloning
  - TAR Cloning of the *spz* Cluster
  - Heterologous Expression of the *spz* Cluster
  - RNA Extraction and RT-PCR Analysis
  - Promoter Cassette Construction
  - DNA Manipulation
  - Streptophenazine Metabolite Analysis
  - Molecular Networking
  - Purification and Isolation of Streptophenazines
  - Structural Characterization of Streptophenazines
  - Bioactivity Testing of Streptophenazines and MIC Determination
  - Phylogenetic Analysis of Spz15
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.chembiol.2019.02.004.

ACKNOWLEDGMENTS

We are grateful to P.R. Jensen and W. Fenical (Scripps Institution of Oceanography) for providing *Streptomyces* sp. CNB-091, M. Bibb (John Innes Center) for *S. coelicolor* M1146, B.M. Dungan (UC San Diego) for assistance with NMR, F.A. Tezcan and R.G. Alberstein (UC San Diego) for assistance with CD, and L. Ray, S.K. McKinnie, T. de Rond, and P.A. Jordan (Scripps Institution of Oceanography) for helpful discussions. This work was supported by grants from the National Institutes of Health (NIH) through grants R01-GM085770 and R01-AI117712 to B.S.M. and NIH Marine Biotechnology Training Grant Predoctoral Fellowship (T32 GM067550) to K.D.B.

AUTHOR CONTRIBUTIONS

H.L. and B.S.M. designed the research. K.D.B., J.L., K.M., and S.M.M. performed the experiments and analyzed the data. V.N. and S.D. designed and performed the bioactivity testing. K.D.B., H.L., and B.S.M. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 30, 2018
Revised: January 2, 2019
Accepted: January 31, 2019
Published: March 7, 2019

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Flett et al., 1997.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amagai, K., Ikeda, H., Hashimoto, J., Kozone, I., Izumikawa, M., Kudo, F., Eguchi, T., Nakamura, T., Osada, H., Takahashi, S., et al. (2017). Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host. Sci. Rep. *7*, 3382.

Bai, C., Zhang, Y., Zhao, X., Hu, Y., Xiang, S., Miao, J., Lou, C., and Zhang, L. (2015). Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. Proc. Natl. Acad. Sci. U S A *112*, 12181–12186.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature *417*, 141–147.

Bibb, M.J., Janssen, G.R., and Ward, J.M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. Gene *38*, 215–226.

Blankenfeldt, W., and Parsons, J.F. (2014). The structural biology of phenazine biosynthesis. Curr. Opin. Struct. Biol. *29*, 26–33.

Bunbamrung, N., Dramae, A., Srichomthong, K., Supothina, S., and Pittayakhajonwut, P. (2014). Streptophenazines I-L from *Streptomyces* sp. BCC21835. Phytochem. Lett. *10*, 91–94.

Caffrey, P. (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. ChemBioChem *4*, 654–657.

Caraballo-Rodríguez, A.M., Dorrestein, P.C., and Pupo, M.T. (2017). Molecular inter-kingdom interactions of endophytes isolated from *Lychnophora ericoides*. Sci. Rep. *7*, 5373–5387.

Chalabaev, S., Turlin, E., Bay, S., Ganneau, C., Brito-Fravallo, E., Charles, J.-F., Danchin, A., and Biville, F. (2008). Cinnamic acid, an autoinducer of its own biosynthesis, is processed via Hca enzymes in *Photorhabdus luminescens*. Appl. Environ. Microbiol. *74*, 1717–1725.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9–14.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U S A *97*, 6640–6645.

Degen, D., Feng, Y., Zhang, Y., Ebright, K.Y., Ebright, Y.W., Gigliotti, M., Vahedian-Movahed, H., Mandal, S., Talaue, M., Connell, N., et al. (2014). Transcription inhibition by the depsipeptide antibiotic salinamide A. Elife *3*, e02451.

Demain, A.L., and Fang, A. (2000). The natural functions of secondary metabolites. Adv. Biochem. Eng. Biotechnol. *69*, 1–39.

Denison, C., and Kodadek, T. (1998). Small-molecule-based strategies for controlling gene expression. Chem. Biol. *5*, R129–R145.

Dietrich, L.E.P., Price-Whelan, A., Petersen, A., Whiteley, M., and Newman, D.K. (2006). The phenazine pyocyanin is a terminal signaling factor in the quorum sensing network of *Pseudomonas aeruginosa*. Mol. Microbiol. *61*, 1308–1321.

Dietrich, L.E.P., Teal, T.K., Price-Whelan, A., and Newman, D.K. (2008). Redox-active antibiotics control gene expression and community behavior in divergent bacteria. Science *321*, 1203–1206.

Fayed, B., Ashford, D.A., Hashem, A.M., Amin, M.A., Gazayerly, O.N.E., Gregory, M.A., and Smith, M.C.M. (2015). Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in *Streptomyces* spp. and combinatorial biosynthesis. Appl. Environ. Microbiol. *81*, 8402–8413.

Fernández-Moreno, M.A., Caballero, J., Hopwood, D.A., and Malpartida, F. (1991). The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. Cell *66*, 769–780.

Fischbach, M.A., and Clardy, J. (2007). One pathway, many products. Nat. Chem. Biol. *3*, 353–355.

Flett, F., Mersinias, V., and Smith, C.P. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol. Lett. *155*, 223–229.

Freestone, T.S., Ju, K.-S., Wang, B., and Zhao, H. (2017). Discovery of a phosphonoacetic acid derived natural product by pathway refactoring. ACS Synth. Biol. *6*, 217–223.

Geiger, A., Keller-Schierlein, W., Brandl, M., and Zähner, H. (1988). Metabolites of microorganisms. 247 phenazines from *Streptomyces antibioticus*, strain Tü 2706. J. Antibiot. (Tokyo) *41*, 1542–1551.

Gomez-Escribano, J.P., and Bibb, M.J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb. Biotechnol. *4*, 207–215.

Gomez-Escribano, J.P., and Bibb, M.J. (2014). Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J. Ind. Microbiol. Biotechnol. *41*, 425–431.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. U S A *100*, 1541–1546.

Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C.J., and Chater, K.F. (2004). Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. Adv. Appl. Microbiol. *54*, 107–128.

Jones, A.C., Gust, B., Kulik, A., Heide, L., Buttner, M.J., and Bibb, M.J. (2013). Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. PLoS One *8*, e69319.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., and Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Keatinge-Clay, A.T. (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. *14*, 898–908.

Kessler, N., Schuhmann, H., Morneweg, S., Linne, U., and Marahiel, M.A. (2004). The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. J. Biol. Chem. *279*, 7413–7419.

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

Kim, E., Moore, B.S., and Yoon, Y.J. (2015). Reinvigorating natural product combinatorial biosynthesis with synthetic biology. Nat. Chem. Biol. *11*, 649–659.

Koehn, F.E., and Carter, G.T. (2005). The evolving role of natural products in drug discovery. Nat. Rev. Drug Discov. *4*, 206–220.

Kouprina, N., and Larionov, V. (2016). Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma *125*, 621–632.

Kunz, A.L., Labes, A., Wiese, J., Bruhn, T., Bringmann, G., and Imhoff, J.F. (2014). Nature's lab for derivatization: new and revised structures of a variety of streptophenazines produced by a sponge-derived *Streptomyces* strain. Mar. Drugs *12*, 1699–1714.

Laursen, J.B., and Nielsen, J. (2004). Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. Chem. Rev. *104*, 1663–1686.

Liang, Y., Chen, L., Ye, X., Anjum, K., Lian, X.-Y., and Zhang, Z. (2017). New streptophenazines from marine *Streptomyces* sp. 182SMLY. Nat. Prod. Res. *31*, 411–417.

Luo, Y., Zhang, L., Barton, K.W., and Zhao, H. (2015). Systematic identification of a panel of strong constitutive promoters from *Streptomyces albus*. ACS Synth. Biol. *4*, 1001–1010.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene *111*, 61–68.

Mavrodi, D.V., Peever, T.L., Mavrodi, O.V., Parejko, J.A., Raaijmakers, J.M., Lemanceau, P., Mazurier, S., Heide, L., Blankenfeldt, W., Weller, D.M., et al. (2010). Diversity and evolution of the phenazine biosynthesis pathway. Appl. Environ. Microbiol. *76*, 866–879.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., et al. (2015). Minimum information about a biosynthetic gene cluster. Nat. Chem. Biol. *11*, 625–631.

Mitova, M.I., Lang, G., Wiese, J., and Imhoff, J.F. (2008). Subinhibitory concentrations of antibiotics induce phenazine production in a marine *Streptomyces* sp. J. Nat. Prod. *71*, 824–827.

Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z., and Brady, S.F. (2015). Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. Proc. Natl. Acad. Sci. U S A *112*, 8953–8958.

Moore, B.S., Trischman, J.A., Seng, D., Kho, D., Jensen, P.R., and Fenical, W. (1999). Salinamides, antiinflammatory depsipeptides from a marine streptomycete. J. Org. Chem. *64*, 1145–1150.

Myronovskyi, M., and Luzhetskyy, A. (2016). Native and engineered promoters in natural product discovery. Nat. Prod. Rep. *33*, 1006–1019.

Newman, D.J., and Cragg, G.M. (2016). Natural products as sources of new drugs from 1981 to 2014. J. Nat. Prod. *79*, 629–661.

Olano, C., Lombó, F., Méndez, C., and Salas, J.A. (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. Metab. Eng. *10*, 281–292.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., et al. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. U S A *98*, 12215–12220.

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet. *26*, 449–457.

Perego, M., Glaser, P., Minutello, A., Strauch, M.A., Leopold, K., and Fischer, W. (1995). Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis* identification of genes and regulation. J. Biol. Chem. *270*, 15598–15606.

Pierson, L.S., and Pierson, E.A. (2010). Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. Appl. Microbiol. Biotechnol. *86*, 1659–1670.

Price-Whelan, A., Dietrich, L.E.P., and Newman, D.K. (2006). Rethinking "secondary" metabolism: physiological roles for phenazine antibiotics. Nat. Chem. Biol. *2*, 71–78.

Ramos, I., Dietrich, L.E.P., Price-Whelan, A., and Newman, D.K. (2010). Phenazines affect biofilm formation by *Pseudomonas aeruginosa* in similar ways at various scales. Res. Microbiol. *161*, 187–191.

Ray, L., Yamanaka, K., and Moore, B.S. (2016). A peptidyl-transesterifying type I thioesterase in salinamide biosynthesis. Angew. Chem. Int. Ed. *55*, 364–367.

Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J., and Leadlay, P.F. (1998). Construction of new vectors for high-level expression in actinomycetes. Gene *216*, 215–223.

Rui, Z., Ye, M., Wang, S., Fujikawa, K., Akerele, B., Aung, M., Floss, H.G., Zhang, W., and Yu, T.-W. (2012). Insights into a divergent phenazine biosynthetic pathway governed by a plasmid-born esmeraldin gene cluster. Chem. Biol. *19*, 1116–1125.

Rutledge, P.J., and Challis, G.L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol. *13*, 509–523.

Schoenafinger, G., Schracke, N., Linne, U., and Marahiel, M.A. (2006). Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. J. Am. Chem. Soc. *128*, 7406–7407.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. *37*, e16.

Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013). Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab. Eng. *19*, 98–106.

Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol. *1079*, 105–116.

Smith, S., and Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat. Prod. Rep. *24*, 1041–1072.

Tang, X., Li, J., Millán-Aguiñaga, N., Zhang, J.J., O'Neill, E.C., Ugalde, J.A., Jensen, P.R., Mantovani, S.M., and Moore, B.S. (2015). Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem. Biol. *10*, 2841–2849.

Trischman, J.A., Tapiolas, D.M., Jensen, P.R., Dwight, R., Fenical, W., McKee, T.C., Ireland, C.M., Stout, T.J., and Clardy, J. (1994). Salinamides A and B: anti-inflammatory depsipeptides from a marine streptomycete. J. Am. Chem. Soc. *116*, 757–758.

Voggu, L., Schlag, S., Biswas, R., Rosenstein, R., Rausch, C., and Götz, F. (2006). Microevolution of cytochrome *bd* oxidase in *Staphylococci* and its implication in resistance to respiratory toxins released by *Pseudomonas*. J. Bacteriol. *188*, 8079–8086.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat. Biotechnol. *34*, 828–837.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. *43*, W237–W243.

Wilson, M.C., and Moore, B.S. (2012). Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. Nat. Prod. Rep. *29*, 72–86.

Wu, Y., and Seyedsayamdost, M.R. (2017). Synergy and target promiscuity drive structural divergence in bacterial alkylquinolone biosynthesis. Cell Chem. Biol. *24*, 1437–1444.

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl. Acad. Sci. U S A *111*, 1957–1962.

Yang, Z., Jin, X., Guaciaro, M., and Molino, B.F. (2012). Asymmetric synthesis and absolute configuration of streptophenazine G. J. Org. Chem. *77*, 3191–3196.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Streptomyces sp. CNB-091	Trischman et al., 1994	GenBank: GCA_000377965.1
Streptomyces coelicolor M1146	Gomez-Escribano and Bibb, 2011	N/A
Additional strains summarized in Table S4		N/A
Chemicals, Peptides, and Recombinant Proteins		
Adenine hemisulfate salt	Sigma	Cat # A3159
Agar	Thermo Fisher	Cat # BP1423
Apramycin	RPI	Cat # A50020
Ampicillin	CalBioChem	Cat # 171254
Chloramphenicol	Thermo Fisher	Cat # 227920250
5-FOA	Sigma	Cat # F5013
Gibson Assembly Master Mix	NEB	Cat # E2611S
D-Glucose	Sigma	Cat # G5146
Kanamycin	Thermo Fisher	Cat # 11815032
LB broth	BD Difco	Cat # 240230
LB agar	Thermo Fisher	Cat # BP1425
Lysozyme	MP	Cat # 100834
LC-MS solvents	Thermo Fisher	N/A
Magnesium chloride	Macron	Cat # 5958-04
D-Mannitol	TCI	Cat # M0044
Nalidixic acid	MP Biomedicals	Cat # 02190246
NMR solvent ($CDCl_3$ with TMS)	Acros	Cat # A166251000
dNTP set	Thermo Fisher	Cat # R0181
Peptone	BD Bacto	Cat # 211677
PrimeSTAR HS DNA polymerase with GC buffer	Takara	Cat # R044A
Sodium chloride	Thermo Fisher	Cat # S2713
Soya flour	MP	Cat # 960024
Sucrose	JT Baker	Cat # 4072-07
Tryptic soy broth	BD Bacto	Cat # 211825
Tryptone	BD Bacto	Cat # 211705
Yeast extract	BD Bacto	Cat #212750
Yeast nitrogen base w/o amino acids and ammonium sulfate	Sigma	Cat # Y1251
Yeast synthetic drop-out medium w/o tryptophan	Sigma	Cat # Y1876
Zymolase	MP	Cat # 320921
Critical Commercial Assays		
Qiaquick PCR Purification kit	Qiagen	Cat # 28104
Qiaquick Gel Extraction kit	Qiagen	Cat # 28704
PureLink RNA Mini kit	Thermo Fisher	Cat # 12183020
OneStep RT-PCR kit	Qiagen	Cat # 210212
Deposited Data		
sp44-p21	Addgene	Addgene: 120232
actIp	Addgene	Addgene: 120233
*ermE*p*	Addgene	Addgene: 120234
GNPS networks	GNPS	https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9a2d02c583c24422bfaa16e2f8ae9ceb

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LC-MS public datasets	Massive.ucsd.edu	MassIVE: MSV000083081 and MassIVE: MSV000083082
Oligonucleotides		
Summarized in Tables S5 and S6	IDT	N/A
Recombinant DNA		
pCAP03	Tang et al., 2015	N/A
pIJ790	Gust et al., 2003	N/A
Additional plasmids summarized in Table S4	This work	N/A
Software and Algorithms		
antiSMASH	Weber et al., 2015	https://antismash.secondarymetabolites.org
NCBI BLAST	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Clustal Omega	Sievers and Higgins, 2014	https://www.ebi.ac.uk/Tools/msa/clustalo/
Geneious	Kearse et al., 2012	https://www.geneious.com
Global Natural Products Social (GNPS) Molecular Networking	Wang et al., 2016	http://gnps.ucsd.edu
Cytoscape	Shannon et al., 2003	https://cytoscape.org/
HPLC Chem Station software	Agilent	https://www.agilent.com
Compass data analysis software	Bruker Daltonic	https://www.bruker.com
HRMS Mass Hunter software	Agilent	https://www.agilent.com
Topspin 2.1.6 software	Bruker Daltonic	https://www.bruker.com
MestreNova NMR software	Mestre labs	http://mestrelab.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Bradley Moore (bsmoore@ucsd.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Streptomyces sp. CNB-091 was the original streptophenazine producer and *S. coelicolor* M1146 was used for heterologous expression of the streptophenazine biosynthetic gene cluster. *E. coli* DH10B cells were used for plasmid storage and replication, *E. coli* ET12567 and *E. coli* ET12567/pUB307 were used for triparental intergeneric conjugation. *E. coli* BW25113 and *E. coli* DH5α/BT340 (Cherepanov and Wackernagel, 1995; Datsenko and Wanner, 2000) were used for λ-Red recombineering. Culture conditions followed literature procedures and are explained in Method Details.

METHOD DETAILS

General Methods

Primers, plasmids, and strains used in this study are summarized in Tables S4, S5, and S6.

Bacterial Strains and Growth Conditions

Streptomyces strains were grown in TSBY liquid medium (3% tryptic soy broth, 10.3% sucrose, 0.5% yeast extract) for isolation of genomic DNA and on SFM agar plates (2% D-mannitol, 2% soya flour, 2% agar) for conjugation and strain maintenance. Liquid cultures were grown at 30°C with shaking at 220 rpm in a rotary shaker. Solid cultures were grown at 30°C for 10-12 days.

For streptophenazine production, seed medium (TSBY) was inoculated with *S. coelicolor* M1146 spore suspension and cultured for 36 hrs. A 5% inoculum of seed culture were used for fermentation medium (also TSBY medium). Fermentation was carried out at 30°C and 220 rpm in a rotary incubation for 3 days.

E. coli strains were grown in Luria-Bertani (LB) broth or on LB agar at 37°C with appropriate antibiotics for selection (apramycin 50 μg/mL, ampicillin 100 μg/mL, chloramphenicol 25 μg/mL, kanamycin 50 μg/mL, nalidixic acid 25 μg/mL). For conjugation purposes, *E. coli* was grown using 2TY medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl) with appropriate antibiotic selection.

Software and Tools Used

The web-based bioinformatics program antiSMASH was used to analyze the whole genome sequence of *S.* sp. CNB-091 (Weber et al., 2015). The sequence of the *spz* gene cluster was further analyzed using BLAST (Altschul et al., 1990). Protein sequence

alignments were made using Clustal Omega software (Sievers and Higgins, 2014). Geneious software was used for all plasmid maps, promoter cassette design, and phylogenetic tree reconstruction (Kearse et al., 2012). HRMS data were analyzed with MassHunter software (Agilent), low resolution MS data was analyzed with Compass DataAnlalysis software (Bruker Daltonic). NMR data was recorded with Bruker Topspin 2.1.6 software and analyzed with MestreNova software. GNPS (gnps.ucsd.edu) was used for molecular networking analysis (Wang et al., 2016). The spectral networks were imported into Cytoscape 3.6.0 (Shannon et al., 2003).

Preparation of Genomic DNA and PCR Fragments for TAR Cloning

S. sp. CNB-091 gDNA was isolated following standard procedures from *Practical Streptomyces Genetics* (Kieser et al., 2000). Approximately 200 μg of gDNA was digested with SnaBI in an overnight reaction at 30°C. Digested gDNA fragments were precipitated and cleaned by isopropanol precipitation. The resulting gDNA pellet was dissolved in 50 μL 10 mM Tris buffer (pH 7).

Takara PrimeSTAR HS DNA polymerase with GC buffer was used to amplify the eight fragments of the *spz* cluster for TAR cloning using gDNA as a template. PCR reactions were carried out in a BioRad MyCycler with gradient option. Fragments were designed with average 200 bp overlap to neighboring fragments and the first and last fragment with 40 bp overlap to pCAP03 capture vector. For primers see Table S5. PCR conditions were as follows:

PCR Reaction:		Vol (μL)
Milli-Q Water		20.5
Primer 1 (10 μM)		1
Primer 2 (10 μM)		1
Template gDNA		1
2X GC Buffer		25
dNTPs (10 mM)		1
Primestar polymerase		0.5
Total volume		50
PCR Cycling Condition (30 Cycles)		
Initial denaturation	98°C	1 min
Denaturation	98°C	10 s
Annealing	60°C	10 s
Extension	72°C	5 min
Final extension	72°C	10 min
	4°C	∞

PCR fragments were cut from agarose gels and purified using the Qiagen QIAquick Gel Extraction kit according to the manufacturer's instructions.

TAR Cloning of the *spz* Cluster

Initially, gDNA was used as a template to directly capture the 48.5 kb region encompassing the *spz* BGC following general procedures described by (Tang et al., 2015; Yamanaka et al., 2014). The *spz* pathway-specific capture vector was constructed by introducing two PCR-amplified 1 kb homology arms corresponding to flanking regions of the targeted region into pCAP03. *S. cerevisiae* VL6-48N (Table S4) was grown to an OD_{600} of 0.7– to 1.0 in 50 mL of YPD medium (2% D-glucose, 1% yeast extract, 2% peptone) supplemented with adenine hemisulfate salt (100 mg/L) at 30°C with shaking. Cells were harvested and washed with ice cold water and osmotically stabilized in 1M sorbitol at 4°C overnight prior to spheroplast preparation. Preparation of spheroplast cells was carried out with zymolase at a final concentration of 0.1 mg/mL with 30-40 min of incubation. Spheroplast cells were mixed with approximately 0.5 μg of gDNA and 0.5 μg of linearized specific capture vector. Transformed spheroplasts were mixed with 8 mL of synthetic drop-out tryptophan top agar at 55°C and overlaid on synthetic tryptophan drop-out (SD-Trp) top agar containing 5-fluorootic acid (5-FOA) for selection. Plates were incubated at 30°C for 4-5 days before transformants were picked and screened by PCR for positive colonies. Plasmids were extracted and then transferred into *E. coli* DH10B cells by electroporation. Plasmids were then purified from antibiotic resistance clones, and the resulting constructions confirmed by restriction analysis.

PCR-based cloning (Shao et al., 2009) followed the same procedures described above and by Yamanaka et al. (2014) and Tang et al. (2015), mixing 200 ng of each PCR fragment with 200 ng of the linearized capture vector pCAP03.

Heterologous Expression of the *spz* Cluster

pKDB01 and its derivatives were transformed into *E. coli* ET12567 by heat shock at 42°C for 55 s, and then transferred to *S. coelicolor* M1146 by triparental intergeneric conjugation facilitated by *E. coli* ET12567/pUB307 (MacNeil et al., 1992) (Table S4). Exconjugants were grown on soy flour mannitol agar containing $MgCl_2$ (10 mM). After incubation at 30°C for 18 hrs, plates were overlaid with 1 mL

water containing 0.5 mg nalidixic acid followed by 1 mL water containing 1 mg apramycin and then grown until appearance of exconjugants, which were then replated for a second round of selection.

RNA Extraction and RT-PCR Analysis

S. coelicolor M1146-pKDB01 and -pKDB03 were grown following the methods for production of streptophenazines. After 24 hrs growth in production media, 1 mL of culture was sampled, pelleted and washed with RNase free water. RNA was extracted using Ambion's PureLink RNA extraction kit following methods for bacterial cells. Briefly, cells were lysed and homogenized using a 20G needle attached to an RNase-free syringe. RNA was bound to a spin column, washed, and eluted with RNase-free water. Concentrations were analyzed using Invitrogen Qubit Fluorometric Quantitation. For primers see Table S6.

RT-PCR was performed using Qiagen's OneStep RT-PCR kit as follows:

PCR Reaction:		Vol (μL)
RNase-free water		23.5
5x buffer		10
5x Q-solution		10
dNTPs (10 mM each)		2
Primer 1 (30 μM)		1
Primer 2 (30 μM)		1
Enzyme mix		2
Template RNA		0.5
Total volume		50
RT-PCR Cycling Condition (30 Cycles)		
Reverse transcription	50°C	30 min
Initial PCR inactivation	95°C	15 min
Denaturation	94°C	1 min
Annealing	67°C	1 min or 1.5 min
Extension	72°C	1 min
Final extension	72°C	10 min
	4°C	∞

Promoter Cassette Construction

The vector pCAP03 was used as the backbone for cloning and construction of all cassettes. pCAP03 was digested with XhoI and NdeI restriction enzymes and recovered from a 0.7% agarose gel. Apramycin and ampicillin resistance genes were PCR amplified from in house plasmids (pKY01 and pETduet-1, respectively). For the bidirectional promoter cassette, the apramycin resistance gene was amplified with primers that contained flanking FRT sites added synthetically. Promoters regions were PCR amplified from in-house plasmids (not published) with primers containing appropriate homology regions for further Gibson assembly. Target fragments were recovered from a 0.7% agarose gel. Gibson assembly was used to combine the digested and purified PCR products with the vector according to manufacturer's guidelines, followed by transformation into *E. coli* DH10B competent cells using heat shock at 42°C for 55 s. Cells were transferred onto LB plates with appropriate antibiotics for selection. After incubation overnight, transformants were picked and inoculated into 10 mL LB broth containing appropriate antibiotics and incubated at 37°C with shaking at 220 rpm for approximately 16 h. Plasmids were isolated and identity was confirmed by restriction analysis and DNA sequencing. Finally, promoter cassettes were PCR amplified using primers with 50 bp overlap to the adjacent sequence of the targeted DNA within the *spz* cluster (Table S5).

DNA Manipulation

All genetic manipulations (gene deletions and promoter replacements) in pKDB01 and pKDB02 were carried out using the λ-Red recombination-mediated PCR-targeted gene deletion/insertion method, as described by Gust et al. (2003). Electrocompetent *E. coli* BW25113 carrying the pIJ790 plasmid were electroporated (Eppendorf Electroporator 2510) at 2,500 V in a 2-mm cuvette with the pKDB01 or pKDB02 cosmid. Expression of the λ-Red genes was induced with arabinose (10 mM) and the cells were then transformed with the promoter cassette described above and incubated at 37°C. Cosmid DNA was then isolated and tested by restriction digest and Sanger sequencing to confirm correct genetic alteration, prior to conjugation into *S. coelicolor* M1146. For the creation of pKDB03, the pKBD01 cosmid containing the *sp44-p21* promoter cassette was transformed into *E. coli* DH5α cells containing the temperature sensitive FLP recombination plasmid BT340. FLP production and loss of the plasmid were induced at

42°C. Cosmid DNA was isolated and the successful excision of the resistance cassette was verified by restriction digest and Sanger sequencing prior to integration into *S. coelicolor* M1146.

For analytical restriction digests, 1.5 μL 10X CutSmart Buffer (NEB), 500 ng DNA, 1 unit of restriction enzyme, and MQ water to a total volume of 15 μL were incubated for 2 hrs at 37°C and then heat inactivated at 65°C for 20 min. For preparative restriction digests, 50 μL reactions including 5 μL 10X CutSmart Buffer (NEB), 1 μg DNA, 2 units of restriction enzyme, and MQ water were incubated for 5 hrs at 37°C prior to heat inactivation and further use.

Streptophenazine Metabolite Analysis

Production cultures (liquid and solid) were extracted by an equal volume of ethyl acetate. Liquid cultures were adjusted to pH ~4 with formic acid prior to extraction. The organic phase was evaporated and samples were reconstituted in methanol and filtered through a 0.2 μm filter for subsequent LC-MS analysis.

A 10 μL aliquot of methanol extract was injected onto a Phenomenex Luna C18 reversed-phase HPLC column (5 μm, 250 mm × 4.6 mm) and analyzed with a Bruker Amazon Ion Trap MS system coupled to an Agilent 1260 Infinity LC system. A solvent system of acetonitrile and water both containing 0.1% formic acid (v/v) was used. Samples were eluted over a 60 min method with a gradient from 10 to 45% acetonitrile over 15 min, 45 to 65% over the next 30 min, and then to 100% over 10 min. 100% acetonitrile was held for 1 min before concentration was dropped to 10%. Flow rate was 0.75 mL/min.

Eluent was detected using electrospray ionization-mass spectrometry (ESI-MS) monitoring *m/z* 70–2,200 in positive mode with a speed of 32,500 *m/z* /s. HRMS was carried out on an Agilent 6530 Accurate-Mass Q-TOF in positive mode, low resolution data was acquired using an Agilent 1260 Infinity IonTrap in positive mode.

Molecular Networking

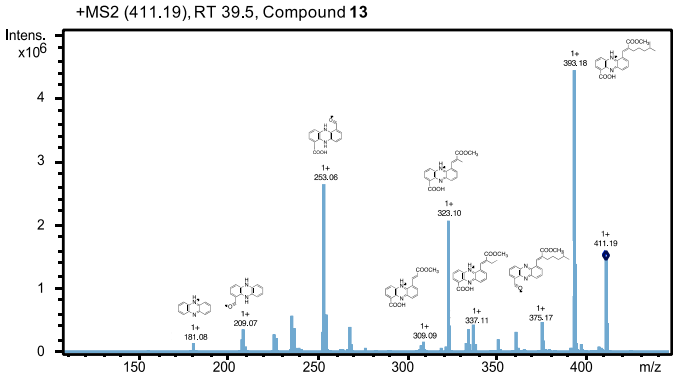
Input data for molecular networking was generated through conversion of the LC-MS/MS raw data to .mzXML data format using ProteoWizard MSConvert software. Data were submitted to the molecular networking workflow at the GNPS online platform (gnps.ucsd.edu).

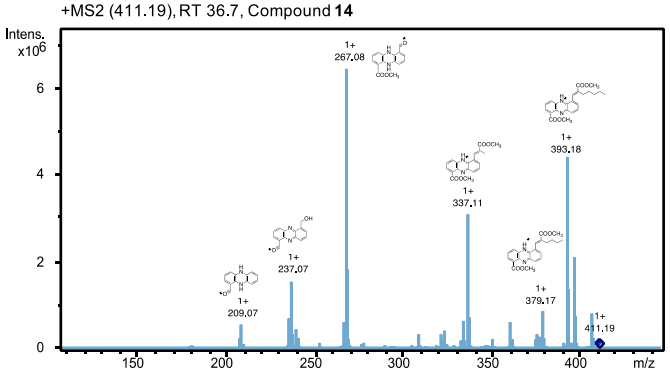
The GNPS molecular network was created using parameters as follows: A molecular network was created using the online workflow at GNPS. The data was clustered with MS-Cluster with a parent mass tolerance of 0.5 Da and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. Further, consensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.5 and more than 2 matched peaks. Edges between two nodes were kept in the network if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. See table below for parameters.

Parameter	Value
Pairs_Min_Cosine	0.5
Analog_Search	0
Tolerance.PM_tolerance	0.5
Tolerance.ion_tolerance	0.5
Min_matched_peaks	2
TopK	10
Cluster_Min_Size	2
Maximum_component_size	100
Min_peak_int	0.0
Filter_Stddev_peak_int	0.0
Run_MScluster	on
Filter_precursor_window	0
Filter_library	1
Window_filter	0
Score_threshold	0.7
Min_matched_peaks_search	6
Max_shift_mass	100.0

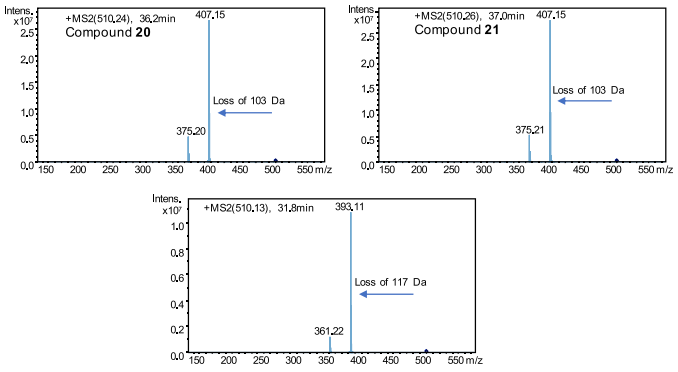
Molecular networks can be viewed at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=24d515f76b4c4d6694a934234435b9bf (Related to Figures 2, 4, S4 and Table S2) and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9a2d02c583c24422bfaa16e2f8ae9ceb (Related to Figures 2 and S3).

HRMS, UV, and MS/MS data for each precursors mass in the network were inspected to confirm relationship to streptophenazines. Example of characteristic MS/MS below:





MS/MS spectra of three *m/z* 510.2 compounds, including the two *N*-formylglycine-containing compounds (**20** and **21**) and the proposed formylalanine-containing compound:



Purification and Isolation of Streptophenazines

Approximately 6 L of 14-day-old culture of *S. coelicolor* M1146-pKDB03 were adjusted to pH ~4 and extracted with an equal volume of ethyl acetate. Organic layers were combined and solvent was evaporated, yielding 8 g of orange solid. The extract was redissolved in MeOH and separated on a bench-packed C18 column, which was washed with increasing MeOH concentrations. Fractions containing streptophenazine metabolites were combined, and the organic solvent was evaporated under vacuum followed by lyophilization. The sample was further purified using a column of silica gel (12 nm, 70 mesh) using a mixture of hexane/ethyl acetate.

Further purification was achieved by preparative HPLC using a Phenomenex Luna C18 column (5 μm, 100 mm x 2 mm), along with an Agilent Technologies system composed of a PrepStar pump, a ProStar 410 autosampler, and a ProStar UV detect (Agilent Technologies, Inc, Santa Clara, USA). Acetonitrile and water, both with 0.1% trifluoroacetic acid (TFA) (v/v), were used as mobile phase. Samples were eluted over 60 min at a flow rate of 15 mL/min. Samples were subsequently subjected to further semi-preparative HPLC purification, using a Phenomenex Luna C18 reversed-phase HPLC column 5 μm, 250 mm × 4.6 mm) using acetonitrile and water, both containing 0.1% TFA, flow rate 2.5 mL/min. Final fractions yielded 1 mg of compounds **13**, **14**, **15**, **22** and **23**, 3 mg compounds **9**–**12**, and over 5 mg of compounds **16**–**21**. All stages of purification were monitored by LC-MS analysis.

Structural Characterization of Streptophenazines

All NMR data were collected using at the UCSD Skaggs School of Pharmacy and Pharmaceutical Sciences NMR Facility on a 600 MHz Bruker NMR spectrometer (Topspin 2.1.6 software, Bruker) consisting of a Magnex 54 mm bore superconducting magnet operating at 14.1 Tesla with a 1.7 mm cryoprobe. Deuterated chloroform containing 1% (v/v) TMS standard was used as a solvent. See Data S1.

Bioactivity Testing of Streptophenazines and MIC Determination

MIC values were determined using broth microdilution in accordance with the Clinical Laboratory Standards Institute (CLSI) guidelines using cation-adjusted Mueller Hinton Broth (MHB) with minor modifications. Briefly, bacteria were grown to mid-log phase (OD_{600nm} = 0.4) at 37°C while shaking except GAS which was grown under static condition. Bacterial cells were then centrifuged, washed, and diluted in PBS to obtain $2x10^6$ cfu/mL, and 10 μL was added to individual wells of a 96-well plate containing 170 μL MHB.

Serial dilutions of compounds **9**, **13**, **16** starting at 500 μg/mL and compound **20** at 400 μg/mL were made in a separate plate, 20 μL of each compound was then added to the test plate. The plates were sealed with parafilm and incubated at 37°C for 24 h. The turbidity of each plate was measured at OD_{600nm} using the EnSpire Alpha plate reader. MIC was defined as the lowest concentration of the drugs that inhibited bacterial growth.

Phylogenetic Analysis of Spz15

The Geneious neighbor-joining tree building method was used for phylogenetic tree reconstruction. Proteins selected for analysis included the closest uncharacterized homologs of Spz15 found using protein BLAST analysis, characterized discrete adenylation proteins, and modular NRPS-associated adenylation domains from characterized BGCs with specificity for alanine or glycine. The sequences of A domains were trimmed according to the annotation by antiSMASH and the MIBiG repository (Medema et al., 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

For bioactivity assays and MIC determination, each sample was repeated three times.

DATA AND SOFTWARE AVAILABILITY

All promoter cassette plasmids are available at Addgene: pCAP03-*sp44/p21* (Addgene: 120232), pCAP03-*actIp* (Addgene: 120233), pCAP03-*ermE*p* (Addgene: 120234). The public datasets used for the GNPS molecular networking are available at massive.ucsd.edu (MassIVE: MSV000083081, MassIVE: MSV000083082). The molecular networks can be accessed at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=24d515f76b4c4d6694a934234435b9bf (for Figures 3 and S4, and Table S2) and: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9a2d02c583c24422bfaa16e2f8ae9ceb (for Figure S3).

Cell Chemical Biology, Volume *26*

Supplemental Information

Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry

Katherine D. Bauman, Jie Li, Kazuya Murata, Simone M. Mantovani, Samira Dahesh, Victor Nizet, Hanna Luhavaya, and Bradley S. Moore

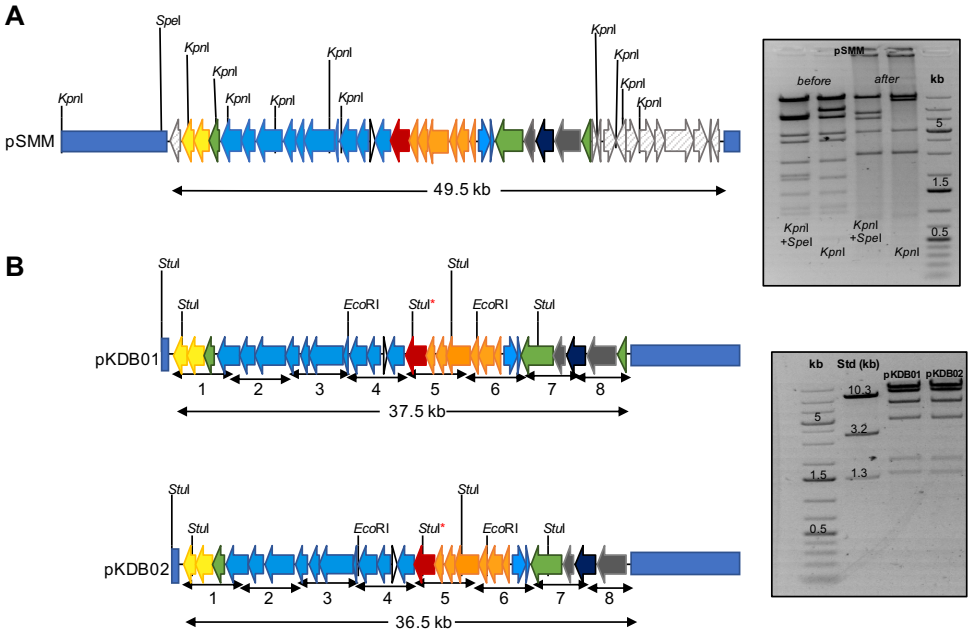


Figure S1. Confirmation of the integrity of the captured *spz* BGC, related to Figures 1 and 2 and Table 1. A. Restriction digest of pSMM before and after transformation in *E. coli* using KpnI and KpnI + SpeI enzymes. KpnI digest, expected sizes (kb) = 23.9, 11.1, 8.2, 4.9, 4.1, 2.7, 1.9, 1.5, 1.1, 1.0, 0.5. KpnI + SpeI digest, expected sizes (kb) = 22.9, 8.9, 8.2, 4.9, 4.1, 2.7, 2.1, 1.9, 1.5, 1.1, 1.0, 0.5. **B.** Restriction digest of pKDB01, expected sizes (kb): 18.0, 13.7, 8.3, 4.9, 2.0, 1.6, and pKDB02, expected sizes (kb): 17.0, 13.3, 8.3, 4.9, 2.0, 1.5, with EcoRI + StuI.* indicates digest is blocked due to overlapping *dcm* methylation. Std – sequenced plasmid digested with EcoRI + StuI, used as a control, expected sizes (kb) = 10.3, 3.2, 1.3. Plasmid backbone represented by blue rectangle.

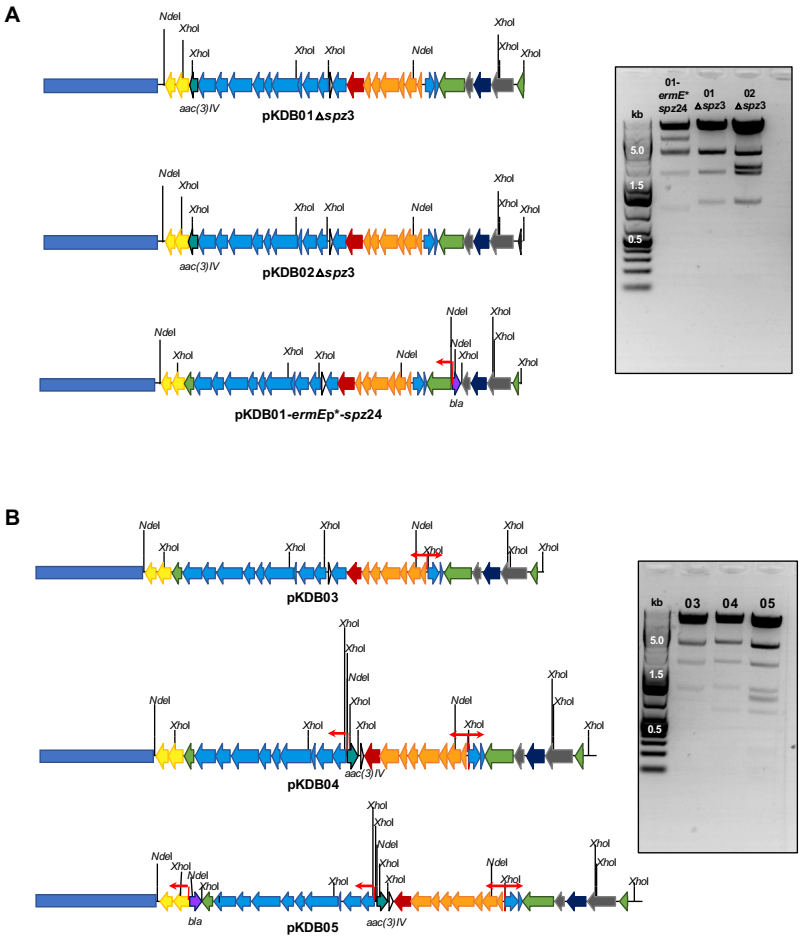


Figure S2. Engineering the *spz* BGC, related to Figures 2 and 3. **A**. Agarose gels of restriction digestions and corresponding maps of regulatory gene mutant plasmids. Enzyme pair XhoI + NdeI used for digestion. pKDB01Δ*spz*3, expected sizes (kb): 10.6, 10.5, 8.7, 8.6, 3.2, 3.1, 1.9, 1.0, 0.1**.** pKDB02Δ*spz*3, expected sizes (kb): 10.6, 10.5, 8.7, 8.6, 3.2, 2.1, 1.9, 1.0, 0.1. pKDB01-*ermE*p-spz*24, expected sizes (kb): 11.7, 10.5, 8.7, 5.2, 3.4, 3.2, 3.1, 1.9, 0.9, 0.2, 0.1. **B.** Restriction digest with XhoI + NdeI of refactored plasmids, agarose gel and corresponding plasmid map showing restriction sites: pKDB03, expected sizes (kb): 11.7, 10.5, 8.7, 7.6, 3.2, 3.1, 1.9, 1.1, 0.1; pKDB04, expected sizes (kb): 11.7, 10.5, 8.6, 7.5, 3.2, 3.1, 1.9, 1.1, 0.7, 0.3, 0.2, 0.1; pKDB05, expected sizes (kb): 10.9, 10.5, 8.7, 7.5, 3.2, 3.1, 1.9, 1.1, 0.9, 0.9 0.7, 0.3, 0.3, 0.1. Plasmid backbone represented by blue rectangle.

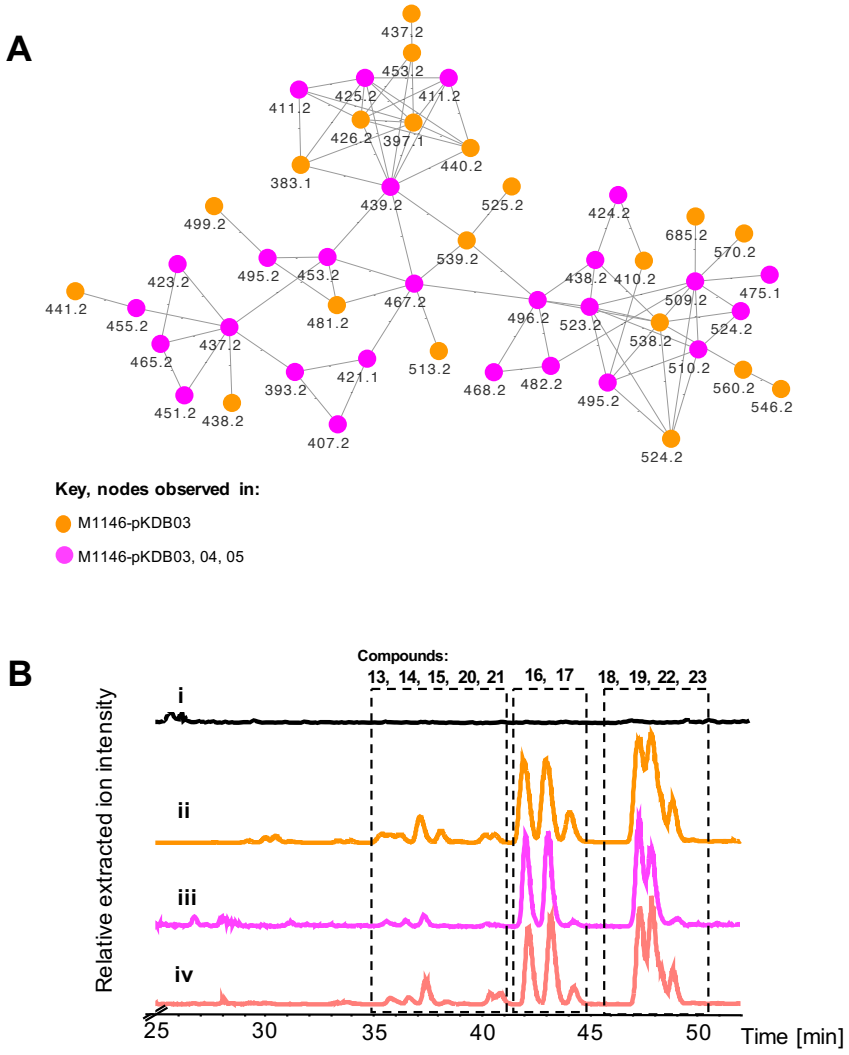


Figure S3. Comparative production of streptophenazines by the refactored *spz* BGC, related to Figures 2, 3, and 4. A. Streptophenazine metabolite cluster of nodes from the molecular network including MS/MS data of extracts of M1146-pKDB03, -pKDB04, and -pKDB05. Nodes are labeled with the corresponding precursor ion mass. **B.** Extracted ion chromatograms (*m/z* 411.2, 425.2, 439.2, 510.2), corresponding to isolated compounds in refactored pathway: (**i**) M1146-pKDB01, (**ii**) M1146-pKDB03, (**iii**) M1146-pKDB04, and (**iv**) M1146-pKDB05.

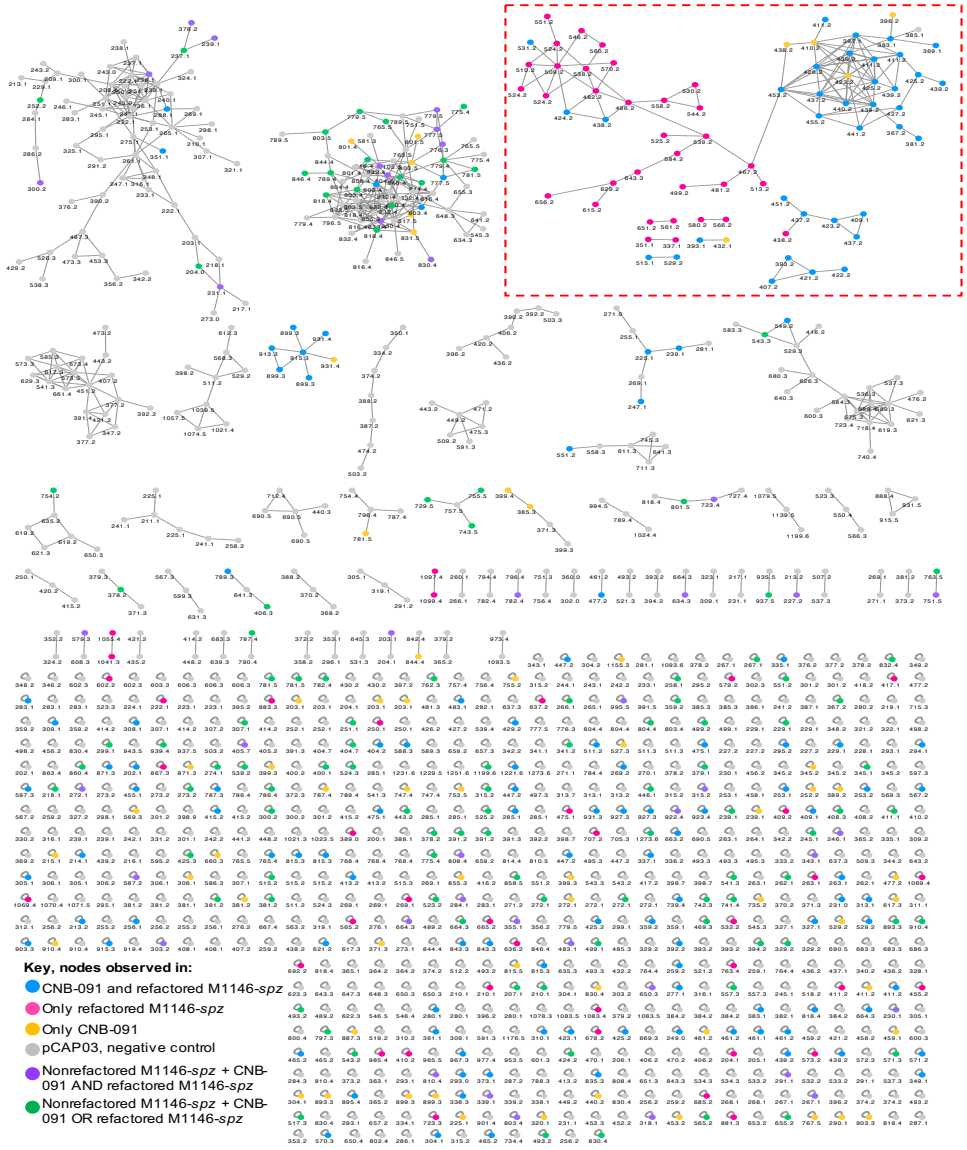
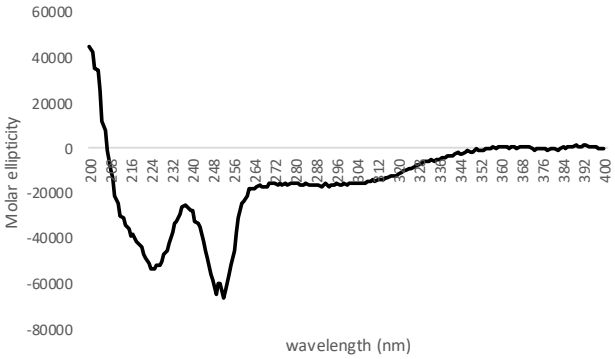


Figure S4. Molecular network showing production of streptophenazines, related to Figure 4. LC-MS/MS data of ethyl acetate extracts from CNB-091, M1146-pCAP03, M1146-pKDB01 (non-refactored *spz* BGC) and M1146-pKDB03 (refactored *spz* BGC) were used for network generation. Nodes corresponding to streptophenazines highlighted in red square.

A



B

```
Spz7_KR      ···–DVYWAHLPDT––VTPVEEILAALDDLVRAGKILHAGLSNFPAWRV–···
Ery1_KR      ···––––LGGIGDDVPLSAVFHAAATLDDGT–VDTLTGERIERASRAKVL···
SlnA1_KR     ···–––LLDRIPEAHPLTGVFHAAGVLDDGM–VGALSAERLDAVLRPKTD···
RifA_KR      ···LEAVLRAIPAEHPLTAVIHTAGVLDDGV–VTELTPDRLATVRRPKVD···
AmphJ_KR     ···–––LLASVPAEHPLTAVVHTAGVLDDGI–FPSLTPDRLDSVMRPKVD···
TlmH_KR      ···–––VLAQIRSRGPIGGVVHAAGLLDDSI–LANMTPEQLHRVLRSKVD···
```

Figure S5. Confirmation of stereochemistry at C-1' position of compound 18, related to Figure 1. **A.** Circular dichroism (CD) spectrum of streptophenazine G (**18**). **B.** Alignment of type I PKS KR domains and Spz7. Amino acid residues (LDD motif) defining Spz7 as type B KR are highlighted. Ery1 = erythromycin, SlnA1 = salinomycin, RifA = rifamycin, AmphJ = amphotericin, TlmH = thiolactomycin.

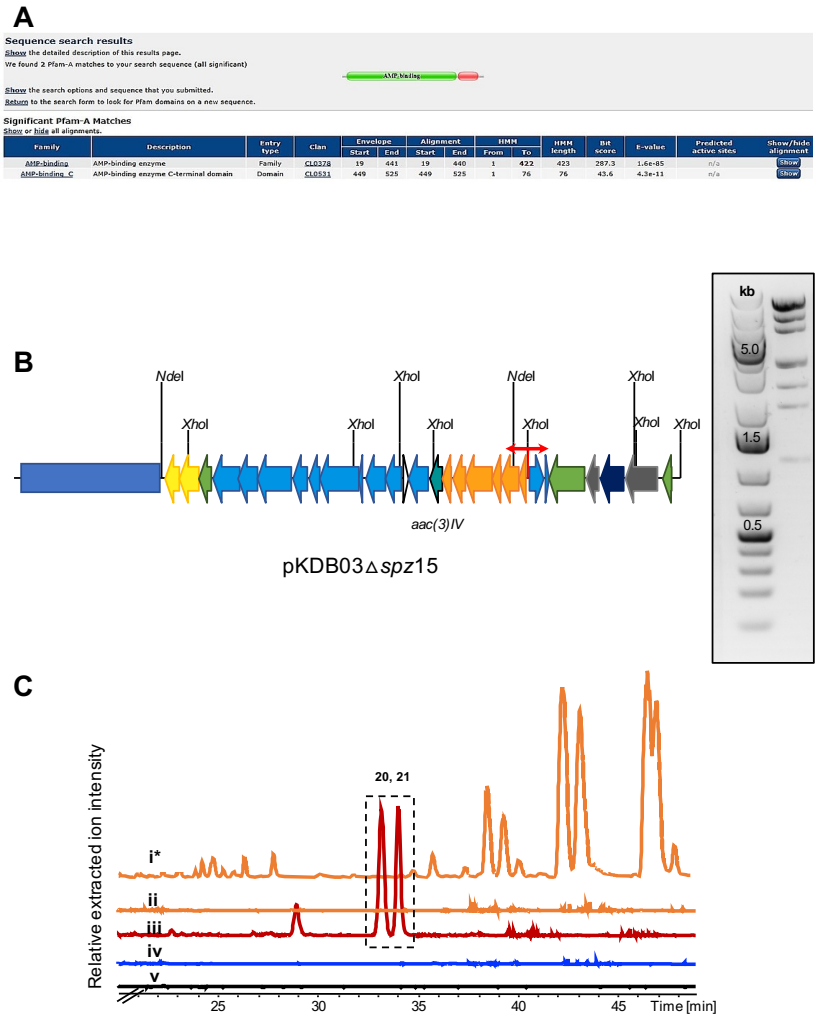


Figure S6. Bioinformatic analysis and gene deletion of *spz*15, related to Figures 5 and 6. A. Protein family (Pfam) analysis of Spz15. Identified CL00378 AMP-binding domain which is characteristic for ANL superfamily of enzymes that includes adenylation domains. **B.** Restriction digest of pKDB03Δ*spz*15 with XhoI and NdeI restriction enzymes. Expected sizes (kb): 11.7, 10.5, 7.5, 5.6, 3.2, 3.1, 2.3, 1.9, 1.1, 0.1. **C.** LC-MS chromatograms: (**i**) Base Peak Chromatogram (BPC) of M1146-pKDB03Δ*spz*15 ******zoomed out 100x,* (**ii**) Extracted Ion Chromatogram (EIC) (*m/z* 510.2, corresponding to compounds **20** and **21**) of M1146-pKDB03Δ*spz*15, (**iii**) M1146-pKDB03 EIC (*m/z* 510.2), (**iv**) CNB-091 EIC (*m/z* 510.2), and (**v**) M1146-pCAP03 EIC (*m/z* 510.2).

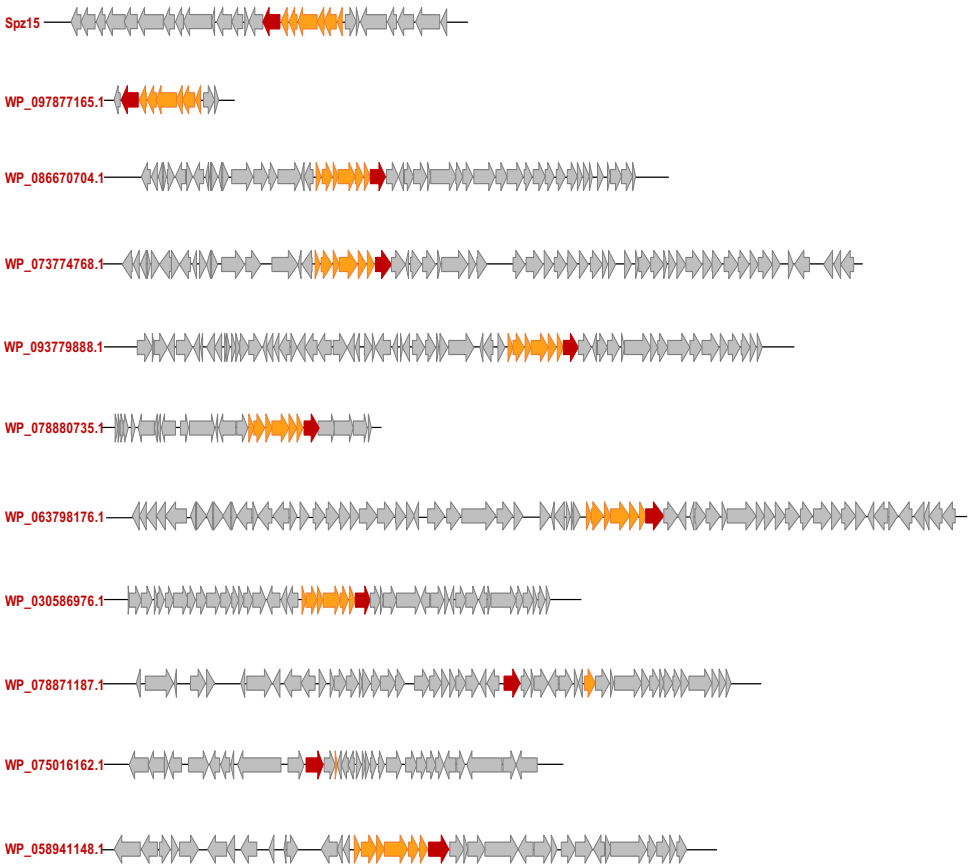


Figure S7. antiSMASH predicted gene neighborhoods of *phz*-associated discrete adenylation proteins, related to Figure 5. Red gene = adenylation enzyme homolog found through blastp, orange genes = phenazine biosynthesis homologs. Spz15 = from *Streptomyces* sp. CNB-091, WP_097877165.1 = from *Streptomyces* sp. ms184, WP_086670704.1 = from *Streptomyces albovinaceus*, WP_073774768.1 = from *Streptomyces* sp. TSRI0445, WP_093779888.1 = from *Streptomyces* sp. yr375, WP_078880735.1 = from *Kitasatospora purpeofusca*, WP_063798176.1 = from *Streptomyces* sp. 150FB, WP_030586976.1 = from *Streptomyces anulatus*, WP_078871187.1 = from *Streptomyces caatingaensis*, WP_075016162.1 = from *Streptomyces rubidus*, WP_058941148.1 = from *Streptomyces kanasensis*.

Table S1, related to Figures 2, 3 and Method Details. Promoter cassette sequences used for refactoring. Bold = restriction site, italic = added for scar, blue = FRT site, red = antibiotic resistance, green = promoter, purple = *actII-ORF4*.

Cassette	Sequence
sp44–p21	**CTCGAG**GGTGAACCGATCTCCTCGTTAGGGTCACACCAGACTTTACAACACCGCACAGCATGT TGTCAAAGCAGAGACGGTTCGAATGTGAACAGCCACTAT**CATATG***TGCAGTTC*GAAGTTCCTAT TCTCTAGAAAGTATAGGAACTTCGGTTCATGTGCAGCTCCATCAGCAAAAGGGGATGATAAGTT TATCACCACCGACTATTTGCAACAGTGCCGTTGATCGTGCTATGATCGACTGATGTCATCAGCG GTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAAC CTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCT CGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACG CTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGT TACACCGGACCTTGGAGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGC CCATCCATTTGCCTTTGCGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCC CCTGCCACCTCACTCGCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTAC TTCTCCTCGGCGTGGGACACGATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAG GTTCCCTATGGGGTGCCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGAT TATCTCGAGAATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAA GAGCCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACAT TGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGG CGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGAAGTTCCTATTCTCTAGA AAGTATAGGAACTTC**AAGCTT**TG**CTCGAG**TGTGCGGGCTCTAACACGTCCTAGTATGGTAGGA TGAGCAATCTAGTCGAGCAACGGAGGTACGGAC**CATATG**
actIp	**CTCGAG**GGTTCATGTGCAGCTCCATCAGCAAAAGGGGATGATAAGTTTATCACCACCGACTAT TTGCAACAGTGCCGTTGATCGTGCTATGATCGACTGATGTCATCAGCGGTGGAGTGCAATGTC GTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCTCAACCTTGGGGTTACCCCC GGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGTCCGGCCCCTCGAAGATGGGCCACT TGGACTGATCGAGGCCCTGCGTGCTGCGCTGGGTCCGGGAGGGACGCTCGTCATGCCCTCG TGGTCAGGTCTGGACGACGAGCCGTTCGATCCTGCCACGTCGCCCGTTACACCGGACCTTGG AGTTGTCTCTGACACATTCTGGCGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTT GCGGCAGCGGGGCCACAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTC GCCTGCAAGCCCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGG GACACGATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTG CCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAGAATGAC CACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAGCCTTCAGAAGGA AGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGACATTGTGGCGACAGCCCT GGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCCGCCAGAGGCGGGATGCGAAGAAT GCGATGCCGCTCGCCAGTCGATTGGCTGA**CATATG**CCACTGCCTCTCGGTAAAATCCAGCAAA AATTAATCAGTGCAGCTCGCTGCACTGATTAATTTTTGATCAATAGGAGATCGCTTGTGACGGC AAGCACATTGAAATCTGTTGAGTAGGCCTGTTATTGTCGCCCCCAGGAGACGGAGAATCTCGA CGGGGGCGCAGATGAGATTCAACTTATTGGGACGTGTCCATGTAATCACCGATGCGGGATGTG TAATTCCGCTTAAATCCTCGAAGGCGACCCAGCTCCTGGTGCTGCTGCTCCTCAGGCGGCACG AGGTGGTGGGATCGGGGGTGCTCATCGAGGAGTTGTGGGCGGACCACCCGCCCCGCAGCGC CATGACGACGCTGCAGACGTACGTGTACCACACCCGCCGGCTGCTGGGGGAGCACCGGGTG ACGAGCGACGACCGGGAATTGGTCCTGACCCAGCCGCCCGGCTACTTCGCCCTGATCGACGA GGACGAACTCGACGTCGCGGTCGCCGAGCGTCTGATCCGCACCGGCGGCCGGCTGCTCGAG GAGAACCGGCTCGAGGAGGCGCTCGCCTCGTTGGACGCGGGACTGGATCTCTGGCGAGGCC CGGCGCTGTCCACCGTACCGTGCGGCCGGGTGCTCGAAAGCAATATCGCGCACCTGGAAGAG CTGCGGCTTTTTGGAATGCAGCTCCGTATCGACGCGAATTGGCGGCTGGGCAGAATAGGGCC GATGATTCCGGAACTCCGGTCCCTGGTAATTTCGCATCCGCTGAACGAGACCCTGCACGCCAA ACTGATGGGCGCGCTCTGTCAGATGGGCAGGCGCGCCGAGGCGCTGGAATCGTATCGGAATC TCCGGCGGATACTGTCCGACGAACTGGGGGGTGGATCCGACGCCGGAAATCCAGCGTATGCAC ATGGAAATTCTCAACGGTGAGAAGGTGCTCGTGTAGCACCGGTCCGTGAACGCGGTGGAGCC CTATGTCTCTTAAGTGTTCCCCTCCCTGCCTCGTGGTCCCTCACGCGCTCAGCTTTGGGCGCC CGGCTCGAGCGGCGGTCGAAGGGAGATGGGGTGCCGCTGGACGCGGCGCCGGTGGATCCG GCATCGAGGGGTCCCGTATCGGCCTTCGAGCCTCCTTCGAGCCACGGGGCCGACGATGACGA CGACCACCGGACGAACGCATC

*ermE*p*	ATGGGGACCTCCTGGGGTGCGTTGGACCGCTGGATCCTACCAACCGGCACGATTGTGCCCAC AACAGCATCGCGGTGCCACGTGTGGACCGCGTCGGTCAGATCCTCCCCGCACCTCTCGCCAG CCGTCAAGATCGACCGCGTGCACCA**CATATG**TTACCAATGCTTAATCAGTGAGGCACCTATCT CAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGAT ACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAG TTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGG TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTA TCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTC TGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC TTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAG CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC TCATACTCTTCCTTTTTCAATCATGATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA CATATTTGAATG**CTCGAG**

Table S3, related to STAR Methods section "Bioactivity testing of streptophenazines and MIC determination." Minimum Inhibitory Concentrations (MICs) (μg/mL) of oxo-streptophenazine A (**9**), streptophenazine C (**13**), streptophenazine A (**16**), and streptophenazine Q (**20**).

Compound **Strain**	**9**	**13**	**16**	**20**
Group A *Streptococcus*	>50	>50	>50	2.5
Acinetobacter baumannii 5075	>50	>50	>50	40
Klebsiella pneumoniae 1100	>50	>50	>50	>40
MRSA TCH1516	>50	>50	>50	40

Table S4, related to STAR methods. Plasmids and strains used in this work.

Plasmids	Description	Sources
pCAP03	TAR cloning and broad-host-range heterologous expression vector; *CEN6-ARS4, oriT, traJ, pUC ori, Kan^r, Apra^r, pADH1, URA3, TRP1*	Tang et al., 2015
pSMM	Derivative of pCAP03 with 48 kb captured *spz* cluster	This work
pKDB01	Derivative of pCAP03 with 37.5 kb captured *spz* cluster	This work
pKDB02	Derivative of pKDB01 without TetR regulatory gene (*spz*28)	This work
pKDB01△*spz*3	Derivative of pKDB01 without *spz*3 (LysR type regulatory gene)	This work
pKDB02△*spz*3	Derivative of pKDB02 without *spz*3 (LysR type regulatory gene)	This work
pKDB01-*ermE*p-spz*24	Derivative of pKDB01 with *ermE** promoter in front of *spz*24 (LuxR-type regulatory gene), *Kan^r*, *Amp^r*	This work
pKDB03	Derivative of pKDB01 with *sp44-p21* cassette	This work
pKDB04	Derivative of pKDB03 with *actIp* cassette	This work
pKDB05	Derivative of pKDB04 with *ermE*p* cassette	This work
pKDB03△*spz*15	Derivative of pKDB03 with *spz*15 (putative adenylation protein-encoding gene) deleted	This work
pCAP03-*ermE*p*	Derivative of pCAP03 containing *ermE*p* cassette between XhoI and NdeI, *Kan^r, Amp^r*	This work
pCAP03-*actIp*	Derivative of pCAP03 containing *actIp* cassette between XhoI and NdeI. *Kan^r, Apra^r*	This work
pCAP03-*sp44-p21*	Derivative of pCAP03 containing *sp44-p21* cassette between XhoI, NdeI. *Kan^r, Apra^r*	This work
pIJ790	*λ*-RED (*gam, bet, exo*), *cat*, *ara*C, rep101^{ts}, oriR101, P araBAD	Gust, 2003
pUB307	Self-transmissible plasmid that mobilizes other plasmids *in trans* for DNA transfer into hosts: RP4, *neo*	Flett, 1997
Strains	**Description**	
Streptomyces		
Streptomyces sp. CNB-091	Native producer of streptophenazines	Trischman et al., 1994
Streptomyces coelicolor M1146	Host strain for heterologous expression derived from *S. coelicolor* M145: Δ*act*, Δ*red*, Δ*cpk*, Δ*cda*.	Gomez-Escribano and Bibb, 2011
S. coelicolor M1146-pCAP03	Heterologous host containing empty pCAP03 as a control	This work
S. coelicolor M1146-pSMM	Heterologous host containing pSMM (*spz* BGC captured in 48 kb DNA fragment)	This work
S. coelicolor M1146-pKDB01	Heterologous host with integrated pKDB01 (37.5 kb captured *spz* cluster)	This work
S. coelicolor M1146-pKDB02	Heterologous host with integrated pKDB02 (△*spz*28)	This work
S. coelicolor M1146-pKDB01△*spz*3	Heterologous host with integrated pKDB01△*spz*3	This work
S. coelicolor M1146-pKDB02△*spz*3	Heterologous host with integrated pKDB02△*spz*3	This work
S. coelicolor M1146-pKDB01-*ermE*p-spz*24	Heterologous host with integrated pKDB01-*ermE*p-spz*24	This work

S. coelicolor M1146-pKDB03	Heterologous host with integrated pKDB03	This work
S. coelicolor M1146-pKDB04	Heterologous host with integrated pKDB04	This work
S. coelicolor M1146-pKDB05	Heterologous host with integrated pKDB05	This work
S. coelicolor M1146-pKDB03△*spz*15	Heterologous host with integrated pKDB03△*spz*15	This work
Escherichia coli		
DH10B	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*ZΔM15, Δ *lac*X74 *rec*A1 *end*A1 *ara*D139 Δ (*ara leu*)7697 *gal*U *gal*K *rps*L *nup*G *λ*–. Storage and maintenance	
BW25113	K-12 derivative: Δ*ara*BAD, Δ*rha*BAD,	Datsenko and Wanner, 2000
BT340	DH5α/pCP20, containing FLP recombinase	Cherepanov and Wackernagel, 1995
ET12567	F- *dam*13::Tn9, *dcm*6, *hsd*M, *hsd*R, *rec*F,143 *zjj*-202::Tn10, *gal*K2, *gal*T22, *ara*-14, *pac*Y1, *xyl*-5, *leu*B6, *thi*-1, *ton*A31, *rps*L136, *his*G4, *tsx*-78, *mtl*-1, *gln*V44. Donor strain for conjugation between *E. coli* and *Streptomyces* in triparental mating	MacNeil et al., 1992
Other		
Group A *Streptococcus*	Clinical isolate used for bioactivity assays	Nizet lab, UCSD
Acetinobacter baumannii 5075	Clinical isolate used for bioactivity assays	Nizet lab, UCSD
Klebsiella pneumoniae 1100	Clinical isolate used for bioactivity assays	Nizet lab, UCSD
MRSA TCH1516	Clinical isolate used for bioactivity assays	Nizet lab, UCSD
S. cerevisiae VL6-48N	MATα *trp*1-Δ1 *ura*3-Δ1 *ade*2-101 *his*3-Δ200 *lys*2 *met*14 cir°, TAR cloning	Kouprina and Larinova, 2016

Table S5, related to STAR methods. Primers used in this work.

Primer name	Sequence (5' to 3')	Purpose
Frag1_F	GAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACAT ATGCATGAGCTGTCTCCTGGTGCTGGTGGGCAGG	TAR cloning *spz* BGC
Frag1_R	CGACCTGCCCGAACTCGACGGGCTGGAACTG	TAR cloning *spz* BGC
Frag2_F	CACGATGCCCAGCAGCAGTCCCATGTCGTGG	TAR cloning *spz* BGC
Frag2_R	GCATCCAGTGCGAGTACAGCCTCGCCGAGCG	TAR cloning *spz* BGC
Frag3_F	CCAGGTACTCCTCCAACTGCTCGCCGGTCCG	TAR cloning *spz* BGC
Frag3_R	CTCATAAGGATGCCTTCTGCGGGTGTGGAGACC	TAR cloning *spz* BGC
Frag4_F	CGATCGCCGCCTGGACCGACTGGAGCAGG	TAR cloning *spz* BGC
Frag4_R	GCTGCCGCTGAACGCCAACGGGAAGGTGGAC	TAR cloning *spz* BGC
Frag5_F	GAGGGCGGTCAGGGCAGGATTCACGGAATGC	TAR cloning *spz* BGC
Frag5_R	GGAGGCAGGCGCGTACTGGACGTTCCTCATCC	TAR cloning *spz* BGC
Frag6_F	CCATCATCTTGAGCTCTTCGTCGACGACCATG	TAR cloning *spz* BGC
Frag6_R	GAACCAGGAGATCGCCGACCGGCTCGTCCT	TAR cloning *spz* BGC
Frag7_F	CTCGGCCTGGAATCACTTCAGATGATGCGCCTG	TAR cloning *spz* BGC
Frag7_R	GTTCCTGATCGCCCAGTTCCTGCAGAGCGTG	TAR cloning *spz* BGC
Frag8_F	GTCACTCCGACGAACATCAGGATGGAGGCCG	TAR cloning *spz* BGC
Frag8_R+TetR	CTCGGTTTGACGCCTCCCATGGTATAAATAGTGGCTC GAGGCACAGGACGGCATCGCCCGCAGCTGAGC	TAR cloning *spz* BGC
Frag8_R-TetR	CTCGGTTTGACGCCTCCCATGGTATAAATAGTGGCTC GAGGGTCCGTCCGGTCCCAGGAGAACCAGCTGATC	TAR cloning *spz* BGC
Δ*spz*3_F	GGCTTGCGGAGAGACCGGGCGGGGATCAGCCGAGG AGGGCGCGGGCGTGTCAGCCAATCGACTGGCGAGC GGCAT	Delete *spz*3 (LysR)
Δ*spz*3_R	GCTCGTCGCACCGTCGCCCCGCTCCTGAACCGGCC GCGGGGAAGGACCACGGTTCATGTGCAGCTCCATCA GCAAAAG	Delete *spz*3 (LysR)
Δ*spz*3_seq_F	GTGAAGGCGACGAACAGGAAGTG	Confirm *spz*3 deletion
Δ*spz*3_seq_R	GATCGAGCCCGCCGACGTG	Confirm *spz*3 deletion
Upreg-*spz*24_F	GTCGGGAGCGTGCGGCACAGTGCGCACGTCGGGCA GGACTGGCACGGTCATATGGGGACCTCCTGGGGTG CGTTGG	Insert *ermE** promoter in front of *spz*24 (LuxR)
Upreg-*spz*24_R	CCCGATCGTGTACGTGCGCGCCGTCGCGCCCTAGG CCCTCCGGGCGGACGCTCGAGCATTCAAATATGTAT CCGCTCATG	Insert *ermE** promoter in front of *spz*24 (LuxR)
Upreg-*spz*24_seq_F	CAGCAGGCTGCTCTTGCCGACAC	Sequencing to confirm insertion
Upreg-*spz*24_seq_R	GATGCGCGGTATGTGGGAGCGC	Sequencing to confirm insertion
***aac(3)IV* +FRT_F**	ATTATACATATGTGCAGTTCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCGGTTCATGTGCAGCTCCATCA GCAAAAG	Amplify *aac(3)IV* gene with FRT sites for cloning between pET28a NdeI and HindIII sites
***aac(3)IV* +FRT_R**	ATTATAAAGCTTGAAGTTCCTATACTTTCTAGAGAATA GGAACTTCTCAGCCAATCGACTGGCGAGCGGCATCG	Amplify *aac(3)IV* gene with FRT sites for cloning between pET28a NdeI and HindIII sites
***sp44*_F**	GACGCCTCCCATGGTATAAATAGTGGCTCGAGGGTG AACCGATCTCCTCGTTAGGGTC	Amplify *sp44* promoter with homology to pCAP03 and *aac(3)IV*
***sp44*_R**	TCTAGAGAATAGGAACTTCGAACTGCACATATGATAG TGGCTGTTCACATTCGAACCGTCTCTG	Amplify *sp44* promoter with homology to pCAP03 and *aac(3)IV*

***p21*_F**	CTATTCTCTAGAAAGTATAGGAACTTCAAGCTTTGCT CGAGTGTGCGGGCTCTAACACGTC	Amplify *p21* promoter with homology to pCAP03 and *aac(3)IV*
***p21*_R**	AGCACGTTCCTTATATGTAGCTTTCGACATATGGTCC GTACCTCCGTTGCTCGACTAGAT	Amplify *p21* promoter with homology to pCAP03 and *aac(3)IV*
***aac(3)IV*+FRT_*sp44*_pCAP03_F**	CAGAGACGGTTCGAATGTGAACAGCCACTATCATAT GTGCAGTTCGAAGTTCCTATTCTCTAGA	Amplify *aac(3)IV* +FRT with homology to *sp*44 promoters
***aac(3)IV*+FRT_*p21*_pCAP03_R**	ACGTGTTAGAGCCCGCACACTCGAGCAAAGCTTGAA GTTCCTATACTTTCTAGAGAATA	Amplify *aac(3)IV* +FRT with homology to *p*21 promoter
***sp44+p21*_cas_insert_F**	TGGTCGGCGTCCTCGTTGCTGAAACTGATGCTCTGG GCGACGTCGGGCATCTCGAGGGTGAACCGATCTCCT CGTTAGG	Amplify *sp44-p21* cassette with homology sequences for targeted insertion into *spz* BGC
***sp44+p21*__cas_insert_R**	CCGGGGACGTGGGTGCCGAGTCCGGCGAGGTAGGT GTTCTCGAACTTCATCATATGGTCCGTACCTCCGTTG CTCGAC	Amplify *sp44-p21* cassette with homology sequences for targeted insertion into *spz* BGC
***sp44+p21*_cas_insert_seq_F**	GTGAACAGGAGATGCCGGGTC	Confirm *sp44-p21* cassette insertion
***sp44+p21*_cas_insert_seq_R**	GCCTCGTACCAGCCTTCCTG	Confirm *sp44-p21* cassette insertion
***actIp*_F**	CCGGTCCGTGAACGCGGTGGAGCCCTATGTCTCTTA AGTGTTCCCCTCCCTGCC	Amplify *actI* promoter with homology to *act*II-ORF4 and pCAP03
***actIp*_R**	CACGTTCCTTATATGTAGCTTTCGAGATGCGTTCGTC CGGTGGTCGTCGTCA	Amplify *actI* promoter with homology to *act*II-ORF4 and pCAP03
***actII*-ORF4_F**	CCGCTCGCCAGTCGATTGGCTGACATATGCCACTGC CTCTCGGTAAAATCC	Amplify *act*II-ORF4 with homology arms for assembly with *aac(3)IV* and *act1p*
***actII*-ORF4_R**	GGCAGGGAGGGGAACACTTAAGAGACATAGGGCTC CACCGCGTTCACGGACCGG	Amplify *act*II-ORF4 with homology arms for assembly with *aac(3)IV* and *act1*p
***aac(3)IV +actIp*_F**	GACGCCTCCCATGGTATAAATAGTGGCTCGAGGGTT CATGTGCAGCTCCATCAGCAAAAG	Amplify *aac(3)IV* with homology to *act*II-ORF4 and pCAP03
***aac(3)IV +actIp*_R**	GGATTTTACCGAGAGGCAGTGGCATATGTCAGCCAA TCGACTGGCGAGCGG	Amplify *aac(3)IV* with homology to *act*II-ORF4 and pCAP03
***actIp*_cas_insert_F**	GGGTCCGCGTCGTACGGAAGGTCAAGAATCTTCGGG TCGGCGGAAGCCACGATGCGTTCGTCCGGTGGTCG TCGTC	Amplify *actIp* cassette with homology sequences for targeted insertion into *spz* BGC
***actIp* _cas_insert_R**	GTGCTCGAATGTCCCATACACCCAAGACGTAGAAGT TCTCTGGAGGAACGACTCGAGGGTTCATGTGCAGCT CCATCAG	Amplify of *actIp* cassette with homology sequences for targeted insertion into *spz* BGC
***actIp*_cas_insert_seq_F**	CTCGAGCCAGTAGCGGGATC	Confirm *actIp* cassette insertion within *spz* BGC
***actIp*_cas_insert_seq_R**	CATTGTGCACGGTCCACCG	Confirm *actIp* cassette insertion within *spz* BGC
***ermE*p*_F**	GTGCCTCACTGATTAAGCATTGGTAACATATGTGGTG CACGCGGTCGATCTTGACGGCTG	Amplify *ermE** promoter with homology arms for assembly with *bla* and pCAP03
***ermE*p*_R**	CAGCACGTTCCTTATATGTAGCTTTCGAATGGGGACC TCCTGGGGTGCGTTGGACC	Amplify *ermE** promoter with homology arms for assembly with *bla* and pCAP03
***bla+ermE*p*_F**	TGACGCCTCCCATGGTATAAATAGTGGCTCGAGCATT CAAATATGTATCCGCTCATGAGA	Amplify *bla* with homology to *ermE*p* and pCAP03 for assembly
***bla+ermE*p*_R**	CAGCCGTCAAGATCGACCGCGTGCACCACATATGTT ACCAATGCTTAATCAGTGAGGCAC	Amplify *bla* with homology to *ermE*p* and pCAP03 for assembly
***ermE*p*_cas _insert_F**	CAGGGTCGCCGTGACGGCGAACTGGAGGCGGGCGA GGTCGACGACGTCCACATGGGGACCTCCTGGGGTG CGTTGG	Amplify *ermE*p* cassette with homology sequences for targeted insertion into *spz* BGC
***ermE*p*_cas_insert_R**	CTTCCACAGCCACGCCCGCGCCCTCCTCGGCTGATC CCCGCCCGGTCTCTCTCGAGCATTCAAATATGTATCC GCTCATG	Amplify *ermE*p* cassette with homology sequences for targeted insertion into *spz* BGC
***ermE*p*_cas_insert_seq_F**	GTGAGGTCGCTCGTGCGCAG	Confirm *ermE*p* cassette insertion within *spz* BGC
***ermE*p*_cas_insert_seq_R**	GTCCTGCCCGAGCAGTATGTCC	Confirm *ermE*p* cassette insertion within *spz* BGC
Δ*spz*15_F	ACGGGTTCCAGGCGGCCGGAGACGACGTCCTCGCG CAGCGCCCCCCGGTCTCAGCCAATCGACTGGCGAG CGGCAT	Delete *spz*15 (adenylation protein)
Δ*spz*15_R	GGCGGGTCGTCCGGCTCCAGCCCTGAGCCCGGGGC ACTCCTCCGCCTCCGGTTCATGTGCAGCTCCATCAG CAAAAG	Delete *spz*15 (adenylation protein)
Δ*spz*15_seq_F	GTCAGGGCAGGATTCACGGAATG	Confirmation of *spz*15 deletion
Δ*spz*15_seq_R	GTACGAACTCCGTTTCGACAGGGTG	Confirmation of *spz*15 deletion

Table S6, related to STAR methods. Primers used for RT-PCR.

Primer name	Sequence (5' to 3')	Primer name	Sequence (5' to 3')
*hrd*B_F	CCTCCGCCTGGTGGTCTCG	*hrd*B_R	AACTTGTAGCCCTTGGTGTAGTCGAAC
*spz1*_F	CGTCACCGAGAGCAGCCACAGCG	*spz1*_R	GTTGCTCCAAGCACTGCGACTGCC
*spz2*_F	CGATCAGGGAGAAGTTCACGACG	*spz2*_R	CTGCACTTCCTGTTCGTCGCCTT
*spz3*_F	GGACATACTGCTCGGGCAGGAC	*spz3*_R	CAGTTGGAGTACTTCCTCGCGG
*spz4*_F	GCCTGATCAGCACCACGTCGG	*spz4*_R	GAACTTCGCGTACGACCTGTGCG
*spz5*_F	CGATCTGCTCGAAGTCGAACAGCC	*spz5*_R	GACACCATGTACGCGTCACTCATCC
*spz6*_F	GTGCAGAGTCTCCAGAACGTCGTGG	*spz6*_R	CACACCGCAGGTCGTGTTCACCTTC
*spz7*_F	GCGAGGCTGTACTCGCACTGGATG	*spz7*_R	GTCTCCGAATACGCGCTGGGCAC
*spz8*_F	GAACGTGGTCACCGGAATGTCCAG	*spz8*_R	GACCGAGCTCGACCTGTACTGCTTC
*spz9*_F	CTCGAAGAAGCTCTCGTCCGGTTCG	*spz9*_R	CTTGAGGACTGGAAGCCGTACGGG
*spz10*_F	GTGGTCGAACAGGTCGATCAGCTC	*spz10*_R	GAGAAGGATTCCCTGACCCTCGTCG
*spz11*_F	GTCTTGCTCCGCATCAGGAATTCC	*spz11*_R	GATCACCTCGACATGATCGTCGAC
*spz12*_F	CTCACGTGGTGCATGACCACCTGG	*spz12*_R	GACATCGTCGGCGTCTTCGAGAG
*spz13*_F	CGATGATCGCCTGGTGGGATCTC	*spz13*_R	GTGGTGGCGACGACGTCGAAC
*spz14*_F	CGAGATGACGTACGTCCACCACGC	*spz14*_R	GCTTCCAGACGGCCTACCGGCATC
*spz15*_F	CACGTCCGCCAGGTGGTAGGTGAG	*spz15*_R	GAGGTGACGGTGTGGTTCTCGGTG
*spz16*_F	CTCCACCCTGTCGAAACGGAGTTC	*spz16*_R	GACCGCCGATATCACCGTCCCGTTTC
*spz17*_F	GCGAGAACATCCTGCTGCGCCAC	*spz17*_R	CTCCTATGTGGTCGTCGACGCCTTC
*spz18*_F	CGGTTGATCTCGACGACACCGAGC	*spz18*_R	CTGGACTCCGCGATCCTCATCCGC
*spz19*_F	CACCAGGAAGGTGCGGATGTCGTG	*spz19*_R	CTGGAATTCCCCGCATCCACCCG
*spz20*_F	CAGGTACTTGTCTCCGACGCCCGAC	*spz20*_R	CTCGACTACGAGCTGCCGATGCTG
*spz21*_F	CTGGTCGAACGGGTTCATGAACTCG	*spz21*_R	GCTGACCCGGCATCTCCTGTTCAC
*spz122*_F	CTGGTCACCACGGAGAAGGCCGTC	*spz22*_R	GATTCCAGATCGAAGCCGATCGGATC
*spz23*_F	CTAGTCGCCGGAGAGGAGCTGCGC	*spz23*_R	GTATCGCGCTCCTGCTCCTTCAGC
*spz24*_F	CAACTCCTCGTAGGCGTGCGACAC	*spz24*_R	CTACCGTCTCTCCCTGCGGCTGGATC
*spz25*_F	GGTGGTGTAGGTGAAGTCCACCCAC	*spz25*_R	GAAGTCGTACGACCAGAGCGGTGTG
*spz26*_F	CGAACATCAGGATGGAGGCCGTCAG	*spz26*_R	GCTGGAGTACTTCTGGTGGGGCTCG
*spz27*_F	CGTCGAGCCGGAGGTCGAAGAAG	*spz27*_R	CAAGGACATCTGGCCGCAGTGGCTG
*spz28*_F	CGATCAGCTGGTTCTCCTGGGACCG	*spz28*_R	GACTAGGGCCCAGGCCAAGGAGCG

ass from m	Determined *m/z*	Retention time	Predicted formula	Calculated *m/z*	ppm	Notes
439.2	439.2229	47.8	$C_{25}H_{31}N_2O_5^+$	439.2227	0.46	Compound **18** (streptophenazine G)
	439.2228	48.3	$C_{25}H_{31}N_2O_5^+$	439.2227	0.23	Compound **19** (streptophenazine F)
	439.2226	48.7	$C_{25}H_{31}N_2O_5^+$	439.2227	-0.23	Compound **22** (diastereomer at C-2' of streptophenazine G)
	439.2228	49.4	$C_{25}H_{31}N_2O_5^+$	439.2227	0.23	Compound **23** (diastereomer at C-2' of streptophenazine G)
425.2	425.207	42.5	$C_{24}H_{29}N_2O_5^+$	425.2071	-0.24	Compound **16** (streptophenazine A)
	425.2078	43.5	$C_{24}H_{29}N_2O_5^+$	425.2071	1.65	
	425.207	44.6	$C_{24}H_{29}N_2O_5^+$	425.2071	-0.24	Compound **17** (streptophenazine B)
	425.2074	33.9	$C_{24}H_{29}N_2O_5^+$	425.2071	0.71	
	425.2063	34.5	$C_{24}H_{29}N_2O_5^+$	425.2071	-1.88	
411.2	411.1914	36.7	$C_{23}H_{27}N_2O_5^+$	411.1914	0	Compound **14** (streptophenazine D)
	411.1919	37.7	$C_{23}H_{27}N_2O_5^+$	411.1914	1.22	Compound **15** (streptophenazine P)
	411.1912	38.7	$C_{23}H_{27}N_2O_5^+$	411.1914	-0.49	*MS/MS indicates methylated at C-1 carboxylic acid (carbomethoxy group)*
	411.1915	39.5	$C_{23}H_{27}N_2O_5^+$	411.1914	0.24	Compound **13** (streptophenazine C)
	411.1913	40.6	$C_{23}H_{27}N_2O_5^+$	411.1914	-0.24	*MS/MS indicates not methylated at C-1 carboxylic acid*
	411.1916	29.9	$C_{23}H_{27}N_2O_5^+$	411.1914	0.49	*MS/MS indicates methylated at C-1 carboxylic acid (carbomethoxy group)*
	411.1917	30.6	$C_{23}H_{27}N_2O_5^+$	411.1914	0.73	*MS/MS indicates not methylated at C-1 carboxylic acid*
	411.1912	31.2	$C_{23}H_{27}N_2O_5^+$	411.1914	-0.49	*MS/MS indicates not methylated at C-1 carboxylic acid*
369.1	369.1441	20.9	$C_{20}H_{21}N_2O_5^+$	369.1445	-1.08	
383.1	383.1604	23.1	$C_{21}H_{23}N_2O_5^+$	383.1601	0.78	*Additional methylene unit from 369.1*
	383.16	23.8	$C_{21}H_{23}N_2O_5^+$	383.1601	-0.26	
397.2	397.1759	26.4	$C_{22}H_{25}N_2O_5^+$	397.1758	0.25	*MS/MS indicates methylated at C-1 carboxylic acid, additional methylene unit from 383.1*
	397.1761	26.9	$C_{22}H_{25}N_2O_5^+$	397.1758	0.76	*MS/MS indicates not methylated at C-1 carboxylic acid*
	397.176	27.6	$C_{22}H_{25}N_2O_5^+$	397.1758	0.5	*MS/MS indicates not methylated at C-1 carboxylic acid*
	397.1758	32.5	$C_{22}H_{25}N_2O_5^+$	397.1758	0	*MS/MS indicates methylated at C-1 carboxylic acid*
	397.1761	32.9	$C_{22}H_{25}N_2O_5^+$	397.1758	0.76	*MS/MS indicates not methylated at C-1 carboxylic acid*
381.2	381.1476	18.5	$C_{21}H_{21}N_2O_5^+$	381.1445	8.13	
367.2	367.2011	50.6	$C_{20}H_{19}N_2O_5^+$	367.2016	-1.36	
427.2	*Manual inspection did not confirm relatedness to phenazine compounds*					
441.2	441.2007	19.4	$C_{24}H_{29}N_2O_5^+$	441.202	-2.95	*HRMS matches known compound streptophenazine H (Mitova* et al., *2002)*
	441.2014	19.8	$C_{24}H_{29}N_2O_5^+$	441.202	-1.36	
423.2	423.1893	51	$C_{24}H_{29}N_2O_5^+$	423.1914	-4.96	Compound **9** (oxo-streptophenazine A)
	423.1925	51.7	$C_{24}H_{29}N_2O_5^+$	423.1914	2.6	Compound **10** (oxo-streptophenazine B)

455.2	455.2176	21.1	$C_{25}H_{31}N_2O_6^+$	455.2177	-0.22	*HRMS matches known compounds I and J from (Kunz* et al., *2014)*
	455.2172	21.8	$C_{25}H_{31}N_2O_6^+$	455.2177	-1.1	
437.2	437.2075	53.6	$C_{25}H_{29}N_2O_5^+$	437.2071	0.91	Compound **11** (oxo-streptophenazine G)
	437.2087	53.9	$C_{25}H_{29}N_2O_5^+$	437.2071	3.66	Compound **12** (oxo-streptophenazine F)
453.2	453.2018	24.2	$C_{29}H_{33}N_2O_5^+$	453.202	-0.44	
	453.2018	50.1	$C_{29}H_{33}N_2O_5^+$	453.202	-0.44	
467.2	467.2156	52.4	$C_{26}H_{31}N_2O_6^+$	467.2182	-5.56	
	467.2167	52.6	$C_{26}H_{31}N_2O_6^+$	467.2182	-3.21	
481.2	481.2314	52.503	$C_{27}H_{33}N_2O_6^+$	481.2333	-3.95	
	481.2339	52.941	$C_{27}H_{33}N_2O_6^+$	481.2333	1.25	*Additional methylene unit from 467.2*
499.2	499.2056	33.9	$C_{26}H_{31}N_2O_8^+$	499.2075	-3.81	
	499.2074	34.8	$C_{26}H_{31}N_2O_8^+$	499.2075	-0.2	
513.2	513.2228	38.4	$C_{27}H_{33}N_2O_8^+$	513.2237	-1.75	*Additional methylene unit from 499.2*
	513.2223	39.1	$C_{27}H_{33}N_2O_8^+$	513.2237	-2.73	
539.2	539.2738	51.3	$C_{30}H_{38}N_2O_7^+$	539.2752	-2.6	
	539.275	51.5	$C_{30}H_{39}N_2O_7^+$	539.2752	-0.37	
584.2	584.2413	40.5				*No chemical formula can be predicted to fit high resolution data*
	584.2425	41.1				
643.3	643.3495	20.2				*No chemical formula can be predicted to fit high resolution data*
629.2	629.2513	53.2	$C_{35}H_{37}N_2O_9^+$	629.2494	3.02	*Additional methylene unit from 615.2*
	629.2505	53.4	$C_{35}H_{37}N_2O_9^+$	629.2494	1.75	
	629.2501	54.1	$C_{35}H_{37}N_2O_9^+$	629.2494	1.11	
	629.2493	54.3	$C_{35}H_{37}N_2O_9^+$	629.2494	-0.16	
615.2	615.2318	51.1	$C_{34}H_{35}N_2O_9^+$	615.2343	-4.06	
	615.2351	51.7	$C_{34}H_{35}N_2O_9^+$	615.2343	1.3	
	615.2352	52.7	$C_{34}H_{35}N_2O_9^+$	615.2343	1.46	
	615.2328	52.9	$C_{34}H_{35}N_2O_9^+$	615.2343	-2.44	
656.2	656.2941	54.6	$C_{37}H_{42}N_3O_8^+$	656.2966	-3.81	
	656.2962	54.9	$C_{37}H_{42}N_3O_8^+$	656.2966	-0.61	
558.2	558.2598	53.4	$C_{32}H_{36}N_3O_6^+$	558.2599	-0.18	*MS/MS suggests glycine attachment at C-1' hydroxy*
	558.2591	53.7	$C_{32}H_{36}N_3O_6^+$	558.2599	-1.43	

	558.258	54.7	$C_{32}H_{36}N_3O_6^+$	558.2599	-3.4	
	558.259	55	$C_{32}H_{36}N_3O_6^+$	558.2599	-1.61	
	558.2598		$C_{32}H_{36}N_3O_6^+$	558.2599	-0.18	
544.2	544.2445	51.5	$C_{31}H_{34}N_3O_6^+$	544.2442	0.55	*One less methylene unit from m/z 558*
	544.2446	52	$C_{31}H_{34}N_3O_6^+$	544.2442	0.73	
	544.2405	53.2	$C_{31}H_{34}N_3O_6^+$	544.2442	-6.8	
	544.2437	53.5	$C_{31}H_{34}N_3O_6^+$	544.2442	-0.92	
	544.2433	55.1	$C_{31}H_{34}N_3O_6^+$	544.2442	-1.65	
	544.2444	55.4	$C_{31}H_{34}N_3O_6^+$	544.2442	0.37	
530.2	530.2274	48.08	$C_{30}H_{32}N_3O_6^+$	530.2286	-2.26	*One less methylene unit from m/z 544*
	530.2284	49.383	$C_{30}H_{32}N_3O_6^+$	530.2286	-0.38	
	530.2266	50.468	$C_{30}H_{32}N_3O_6^+$	530.2286	-3.77	
	530.2257	52.966	$C_{30}H_{32}N_3O_6^+$	530.2286	-5.47	
525.2	525.26	47.2	$C_{29}H_{37}N_2O_7^+$	525.2595	0.95	
	525.2587	48.2	$C_{29}H_{37}N_2O_7^+$	525.2595	-1.52	
	525.261	49.3	$C_{29}H_{37}N_2O_7^+$	525.2595	2.86	
496.2	496.2432	17.9	$C_{27}H_{34}N_3O_6^+$	496.2442	-2.02	*MS/MS suggests formylglycine attachment at C-1' hydroxy*
482.2	482.2281	16.9	$C_{26}H_{32}N_3O_6^+$	482.2286	-1.04	*One less methylene unit from m/z 496.2*
	482.2285	17.1	$C_{26}H_{32}N_3O_6^+$	482.2286	-0.21	
438.2	438.2387	17				*No chemical formula can be predicted to fit high resolution data*
	438.2394	17.1				
538.2	538.2534	40.5	$C_{29}H_{36}N_3O_7^+$	538.2548	-2.6	*MS/MS suggests formylglycine attachment at C-1' hydroxy, additional methylene unit to 52*
	538.2538	41.1	$C_{29}H_{36}N_3O_7^+$	538.2548	-1.86	
560.2	560.2352	40.5	$C_{29}H_{35}N_3O_7Na^+$	560.2367	-2.68	*Sodium adduct of m/z 538.2*
	560.2366	41.1	$C_{29}H_{35}N_3O_7Na^+$	560.2367	-0.18	
570.2	570.2228	32.6	$C_{32}H_{32}N_3O_7^+$	570.2235	-1.23	
	570.2296	33.6	$C_{32}H_{32}N_3O_7^+$	570.2235	10.7	
	570.2258	35.9	$C_{32}H_{32}N_3O_7^+$	570.2235	4.03	
	570.2295	37.2	$C_{32}H_{32}N_3O_7^+$	570.2235	10.52	
424.2	*Mass does not correspond to discrete peak in LCMS chromatogram*					
546.2	546.2183	35.9	$C_{28}H_{33}N_3O_7Na^+$	546.2211	-5.13	*Sodium adduct of m/z 524.2*
	546.2191	36.9	$C_{28}H_{33}N_3O_7Na^+$	546.2211	-3.66	

	546.2201	40.6	$C_{29}H_{33}N_3O_7Na^+$	546.2211	-1.83	
	546.2208	41.5	$C_{29}H_{33}N_3O_7Na^+$	546.2211	-0.55	
524.2	524.2383	35.9	$C_{28}H_{34}N_3O_7^+$	524.2391	-1.53	*Additional methylene from compounds **20** and **21***
	524.2377	36.9	$C_{28}H_{34}N_3O_7^+$	524.2391	-2.67	
	524.2386	40.6	$C_{28}H_{34}N_3O_7^+$	524.2391	-0.95	*Additional methylene unit in side chain, no methylation at C1 carboxylic acid*
	524.2384	41.5	$C_{28}H_{34}N_3O_7^+$	524.2391	-1.34	*Additional methylene unit in side chain, no methylation at C1 carboxylic acid*
510.2	510.225	31.8	$C_{27}H_{32}N_3O_7^+$	510.2235	2.94	*Loss of 117 suggestive of formylalanine at C-1' hydroxy group*
	510.2241	36.2	$C_{27}H_{32}N_3O_7^+$	510.2235	1.18	Compound **20** (streptophenazine Q)
	510.2235	37	$C_{27}H_{32}N_3O_7^+$	510.2235	0	Compoound **21** (streptophenazine R)
509.2	509.2394	17.2				*No chemical formula can be predicted to fit high resolution data*
	509.2397	17.4				
	509.27	17.8				
	509.2678	17.9				
551.2	551.2462	16.3				*No chemical formula can be predicted to fit high resolution data*
531.2	*Mass does not correspond to discrete peak in LCMS chromatogram*					
451.2	*Mass does not correspond to discrete peak in LCMS chromatogram*					
409.1	409.213	55.4	$C_{24}H_{29}N_2O_4^+$	409.2122	1.95	
	409.2118	55.8	$C_{24}H_{29}N_2O_4^+$	409.2122	-0.98	
393.2	393.1221	39.7				*No chemical formula can be predicted to fit high resolution data*
	393.1802	52				
422.2	*Mass does not correspond to discrete peak in LCMS chromatogram*					
421.2	421.2109	56				*MS/MS and UV data do not confirm relatedness to phenazines*
515.1	515.1897	12.1				*MS/MS and UV data do not confirm relatedness to phenazines*
529.1	529.2761	17.2				*MS/MS and UV data do not confirm relatedness to phenazines*
351.1	351.1688	50.9	$C_{21}H_{23}N_2O_3^+$	351.1703	-4.27	
	351.1691	51.7	$C_{21}H_{23}N_2O_3^+$	351.1703	-3.42	
	351.1335	20.5	$C_{20}H_{19}N_2O_4^+$	351.1339	-1.14	
337.1	337.1536	45.2	$C_{19}H_{17}N_2O_4^+$	337.1547	-3.26	
	337.155	46.4	$C_{19}H_{17}N_2O_4^+$	337.1547	0.89	
	337.1179	21.6	$C_{19}H_{17}N_2O_4^+$	337.1183	-1.19	
580.2	*Mass does not correspond to discrete peak in LCMS chromatogram*					

566.2	566.3539	11.7	$C_{33}H_{48}N_3O_5^+$	566.3588	-8.65	
651.2	651.2661	42				*No chemical formula can be predicted to fit high resolution data*
561.2	561.2557	51.2	$C_{32}H_{37}N_2O_7^+$	561.2595	-6.77	
	561.2568	51.5	$C_{32}H_{37}N_2O_7^+$	561.2595	-4.81	

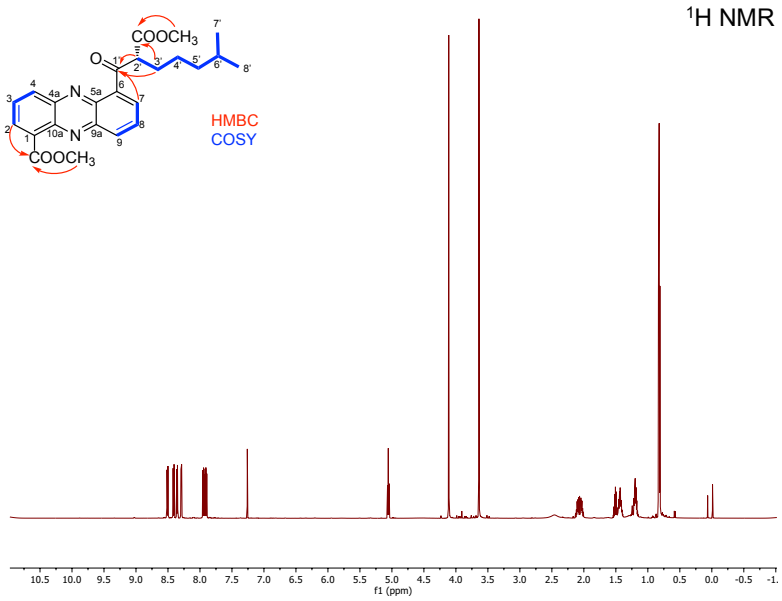
432.1 *Mass does not correspond to discrete peak in LCMS chromatogram*

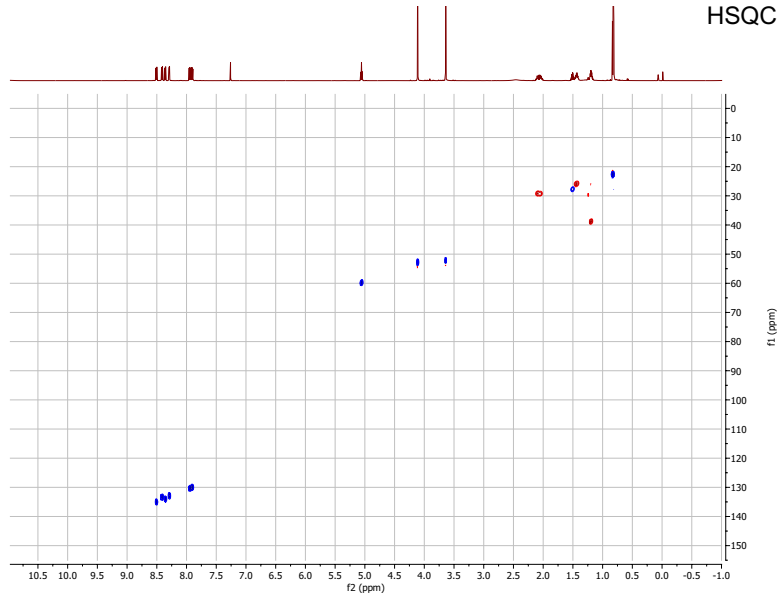
Data S1. NMR analysis and data for isolated streptophenazines, related to **STAR methods** section "Isolation and characterization of streptophenazines" and **Figures 1, 2** and **3**. All spectra recorded in $CDCl_3$ with TMS, 600 MHz. δ in ppm, *J* in Hz. Abbreviations: d = doublet, t = triplet, s = singlet, m = multiplet.

NMR spectroscopic data of compound **9** (oxo-streptophenazine A).

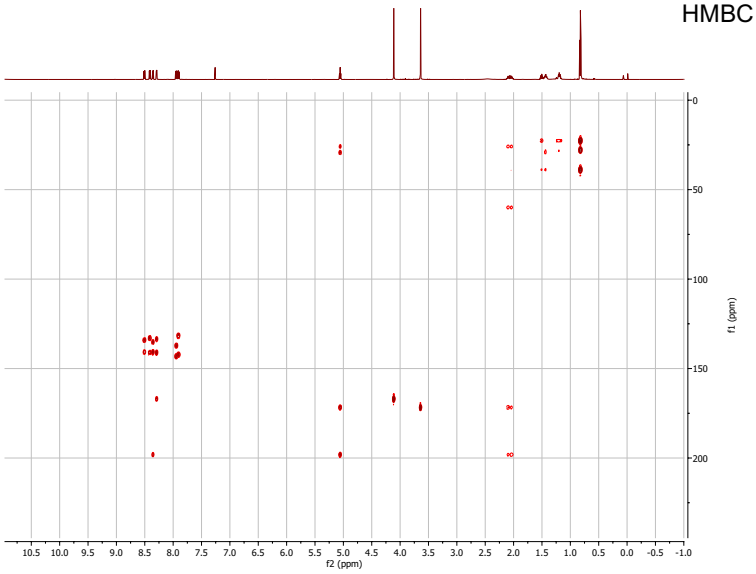
Position	δ_C	δ_H, **mult. (*J*)**	**HMBC**	**COSY**
1	131.5			
2	132.9	8.29, dd (1.45, 6.89)	4, 10a, **1-CO**OCH3	3
3	129.9	7.91, dd (6.87, 8.75)	1, 4a	2, 4
4	133.5	8.41, dd (1.42, 8.69)	2, 10a	3
4a	142.1			
5a	140.9			
6	137.1			
7	133.9	8.36, dd (1.45, 6.99)	1', 9, 5a	8
8	130.1	7.94, dd (6.95, 8.71)	6, 9a	7, 9
9	135.1	8.51, dd (1.45, 8.68)	7, 5a	8
9a	142.9			
10a	141			
1'	198.4			
2'	59.9	5.06, br t (7.02, 7.02)	4', 3', **2'-CO**OCH_3, 1'	3'
3'	29.12	2.09 and 2.05	4', 5', 2', 2'b, 1'	2', 3'
4'	25.82	1.44	3', 5', 2'	5'
5'	38.81	1.2	7', 8', 3', 6', 4'	4', 6'
6'	27.83	1.51, dt (6.68, 6.68, 13.39)	7' and 8', 4', 5'	7', 8'
7'	22.35	0.82, t (6.88, 6.88)	8', 6', 5'	6'
8'	22.35	0.82, t (6.88, 6.88)	7', 6', 5'	6'
1-COOCH_3	167.1			
1-COO**CH_3**	52.8	4.11, s	**1-CO**OCH_3	
2'-COOCH_3	171.5			
2'-COO**CH_3**	52.2	3.64, s	**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **9** (oxo-streptophenazine A).

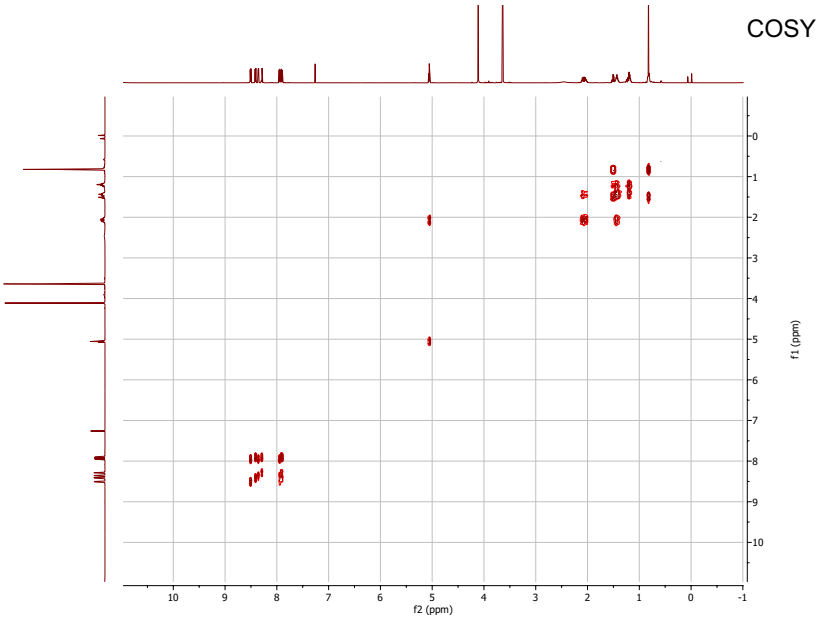




HMBC


0
50
100
150
200
f1 (ppm)
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f2 (ppm)

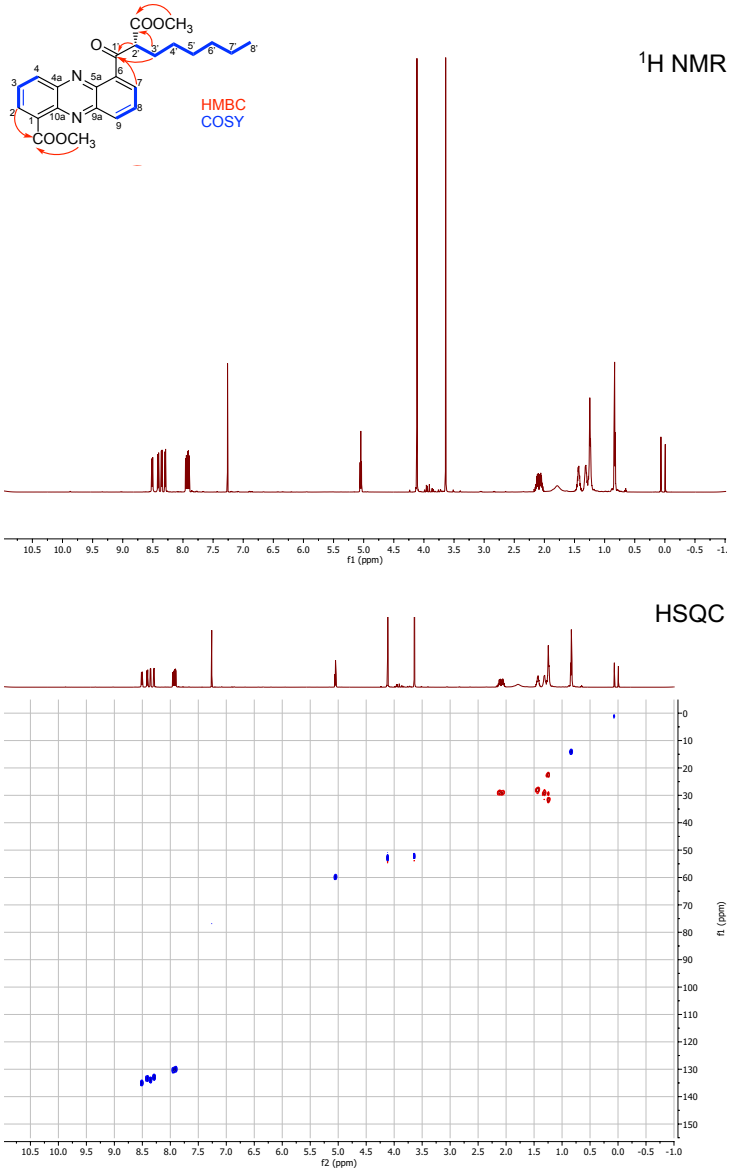
COSY


0
1
2
3
4
5
6
7
8
9
10
f1 (ppm)
10 9 8 7 6 5 4 3 2 1 0 -1
f2 (ppm)

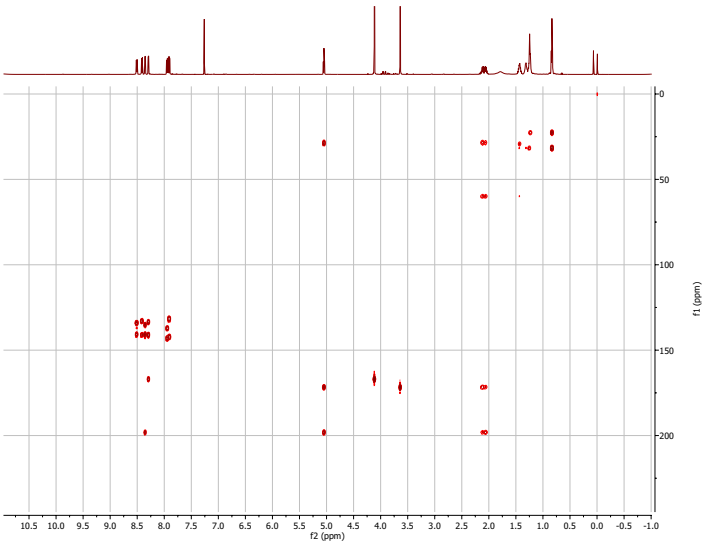
NMR spectroscopic data of compound **10** (oxo-streptophenazine B).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.6			
2	132.9	8.29, dd (1,45, 6.87)	4, 10a, **1-CO**OCH_3	3
3	129.8	7.91, dd (6.97, 8.73)	1, 4a	4, 2
4	133.5	8.41, dd (1.46, 8.74)	2, 10a	3
4a	142.1			
5a	140.7			
6	137.1			
7	133.8	8.36, dd (1.36, 6.88)	1', 9, 5a	8
8	130.2	7.94, dd (6.97, 8.69)	6, 9a	9, 7
9	134.9	8.51, dd (1.46, 8.77)	7, 5a	8
9a	142.9			
10a	141			
1'	198.3			
2'	59.7	5.05, br t (7.03, 7.03)	4', **2'-CO**OCH_3, 1'	3'
3'	29.1	2.11, 2.06	4', 2', **2'-CO**OCH_3, 1'	2', 4'
4'	28.1	1.44	5', 3', 6', 2'	5'
5'	29.3	1.31, dtd (5.47, 7.81, 8.08, 10.35)	3', 6', 7'	4', 5'
6'	31.7	1.24	4', 7', 8'	5', 7'
7'	22.6	1.25	5', 6', 8'	8', 6'
8'	14	0.83, br t (6.93, 6.93)	6', 7'	7'
1-COOCH_3	167.1			
1-COO**CH_3**	52.8	4.12, s	**1-CO**OCH_3	
2'-COOCH_3	171.5			
2'-COO**CH_3**	52.1	3.64, s	**2'-CO**OCH_3	

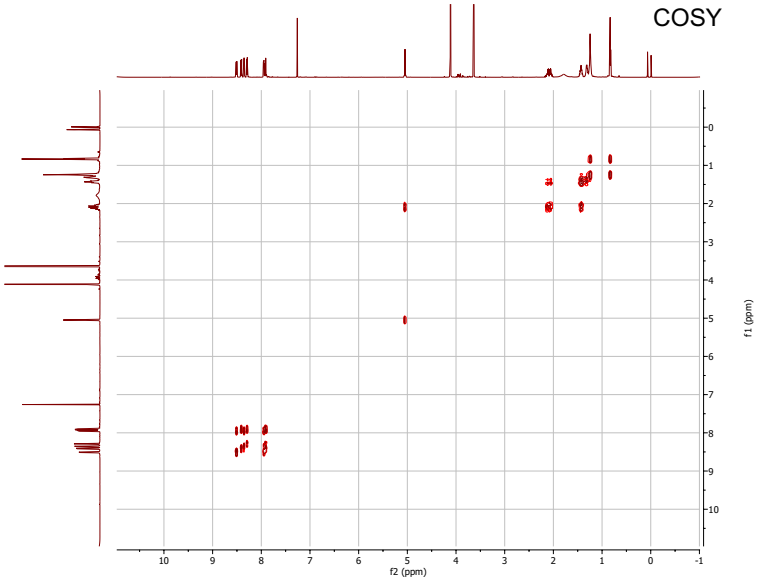
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **10** (oxo-streptophenazine B).



HMBC



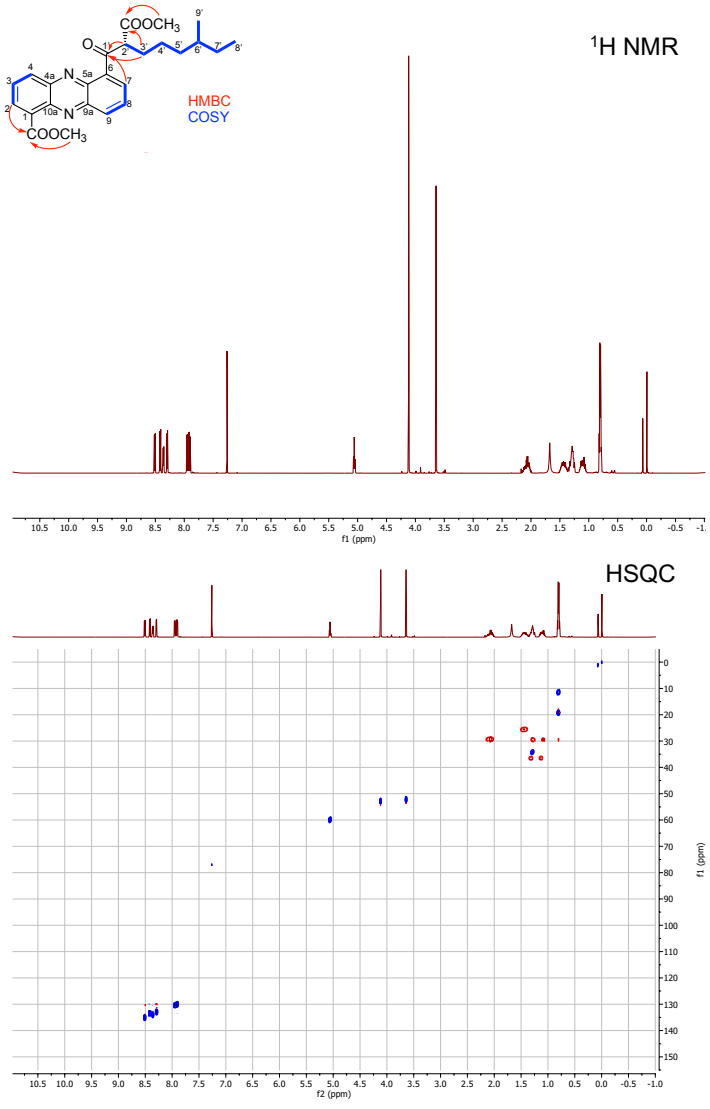
COSY



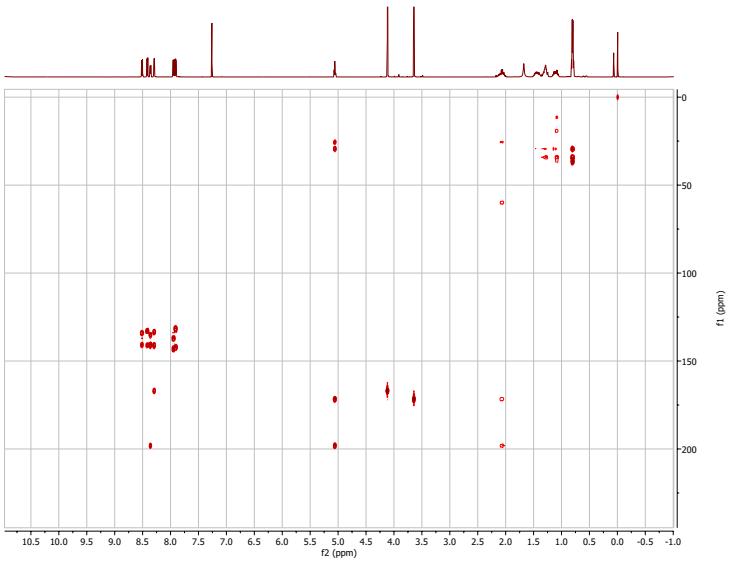
NMR spectroscopic data of compound **11** (oxo-streptophenazine G).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.2			
2	132.89	8.29, dd (1.44, 6.91)	4, 10a, **1-CO**OCH_3	3
3	129.96	7.91, dd (6.87, 8.72)	1, 4a	2, 4
4	133.36	8.41, dd (1.46, 8.71)	2, 10a	3
4a	141.7			
5a	140.96			
6	137.1			
7	133.93	8.36, ddd (1.45, 2.82, 6.98)	9, 5a, 1'	8
8	130.3	7.94, dd (6.97, 8.72)	6, 9a	7, 9
9	135.11	8.51, dd (1.48, 8.67)	7, 5a	8
9a	143.1			
10a	141.1			
1'	198.2			
2'	60.01	5.06, td (2.10, 7.04, 7.04)	4', 3', **2'-CO**OCH_3, 1'	3'
3'	29.31	2.08, 2.05	4', 5', 2', **2'-CO**OCH_3, 1	2', 4'
4'	25.51	1.44, m	6', 5', 2', 3'	3', 5'
5'	36.25	1.13 1.32	8', 9', 4', 3', 6'	4', 6', 9'
6'	34.2	1.29	9', 7', 4'	9', 5', 7'
7'	29.41	1.08, 1.28	5', 6', 9'	5', 6', 8'
8'	11.36	0.81	5', 6', 7', 8'	7'
9'	19.21	0.79	5', 6', 7'	6'
1-COOCH_3	167.3			
1-COO**CH_3**	52.8	4.11, s	**1-CO**OCH_3	-
2'-COOCH_3	171.8			
2'-COO**CH_3**	52.1	3.64, s	**2'-CO**OCH_3	-

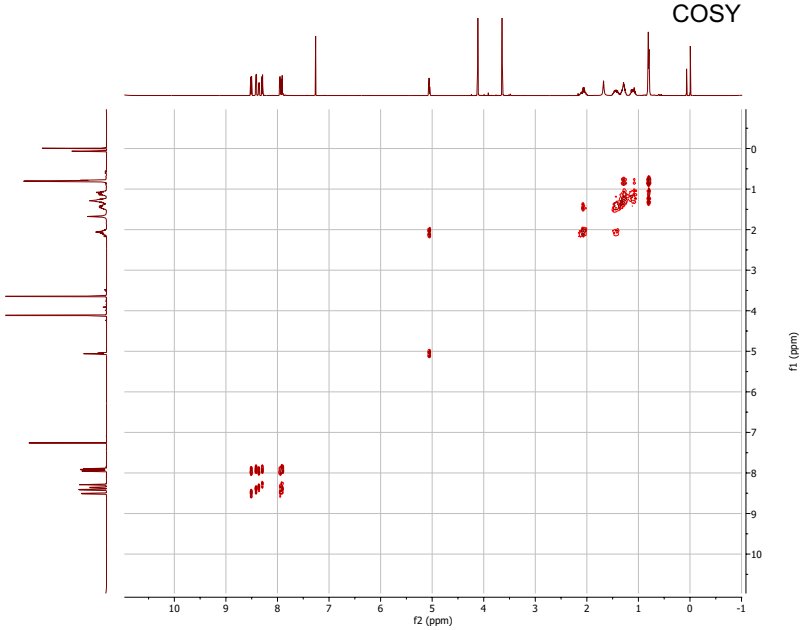
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **11** (oxo-streptophenazine G).



HMBC


0
50
100
150
200
f1 (ppm)
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f2 (ppm)

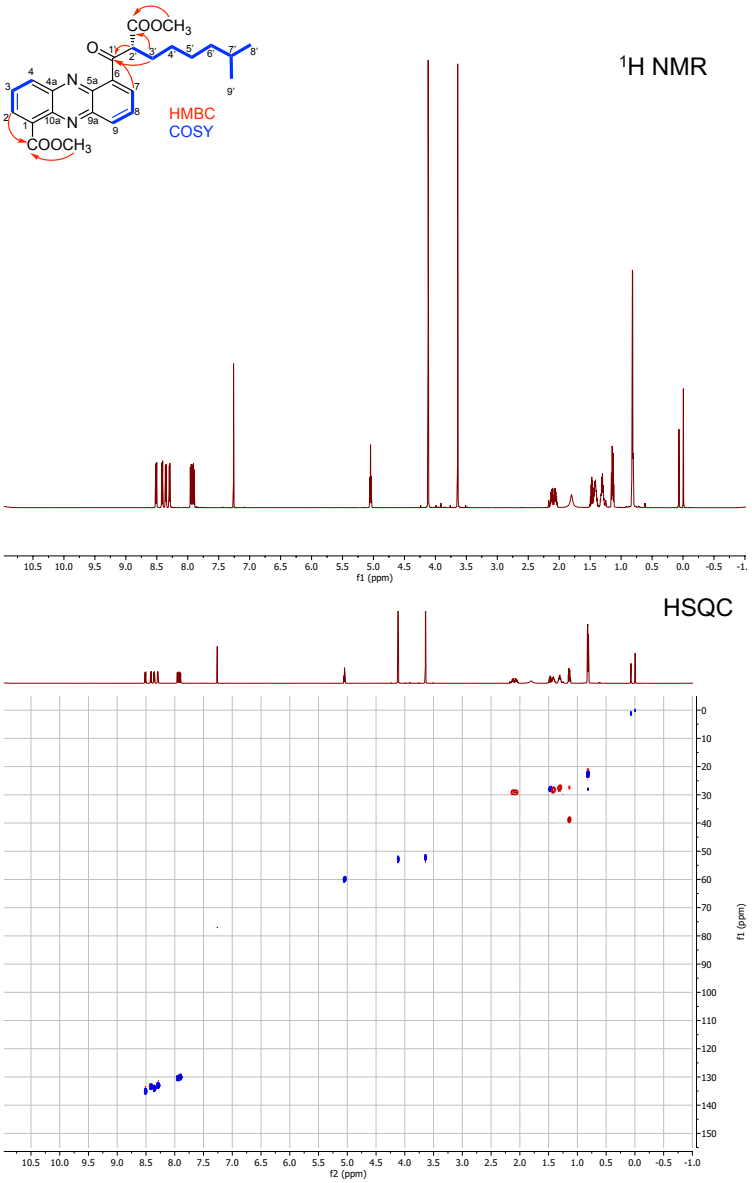
COSY

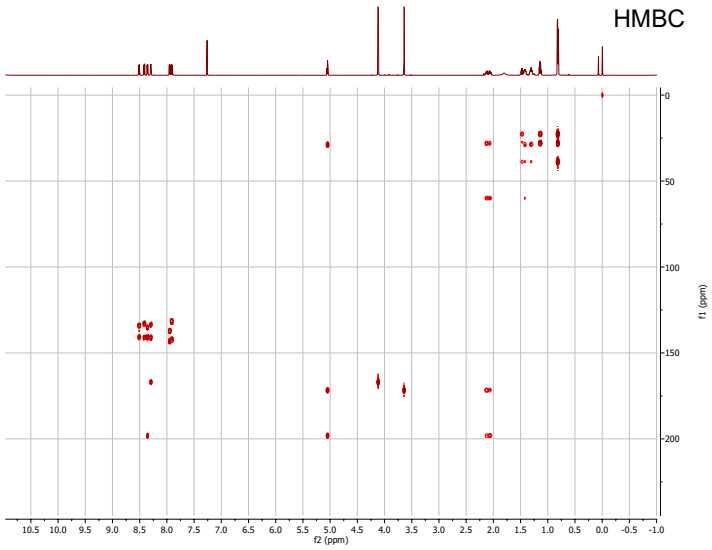

0
1
2
3
4
5
6
7
8
9
10
f1 (ppm)
10 9 8 7 6 5 4 3 2 1 0 -1
f2 (ppm)

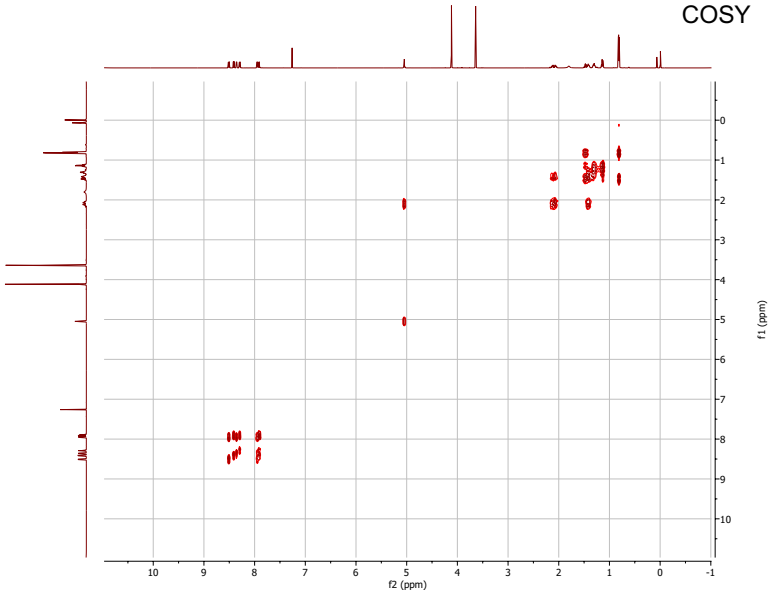
NMR spectroscopic data of compound **12** (oxo-streptophenazine F).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.6			
2	132.9	8.29, dd (1.43, 6.93)	4, 10a, **1-CO**OCH_3	3
3	129.96	7.91, dd (6.89, 8.74)	1, 4a	2, 4
4	133.4	8.41, dd (1.42, 8.77)	2, 10a	3
4a	142.3			
5a	140.9			
6	136.9			
7	133.9	8.36, dd (1.45, 6.95)	9, 5a, 1'	8
8	130.3	7.94, dd (6.94, 8.69)	6, 9a	7, 9
9	135.1	8.51, dd (1.45, 8.67)	7, 5a	8
9a	143			
10a	140.6			
1'	198.2			
2'	59.9	5.05, br t (7.02, 7.02)	4', 3', 1', **2'-CO**OCH_3	3'
3'	29	2.09, m	1', 2', 5', **2'-CO**OCH_3	2', 4'
4'	28.2	1.41	2', 3', 6'	3', 5'
5'	27.4	1.3	6', 7'	4', 6'
6'	38.6	1.14, dt (6.94, 6.94, 8.37)	8', 9'), 4'	5', 7'
7'	27.9	1.48, dt (6.73, 6.73, 13.50)	8', 9', 5', 6'	6', 8'
8'	22.5	0.82, dd (4.02, 6.61)	9', 7', 6'	7'
9'	22.5	0.82, dd (4.02, 6.61)	8', 7', 6'	7'
1-COOCH_3	167.2			
1-COO**CH_3**	52.8	4.12, s	**1-CO**OCH_3	
2'-COOCH_3	171.8			
2'-COO**CH_3**	52.1	3.64, s	**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR and key HMBC and COSY correlations of compound **12** (oxo-streptophenazine F).



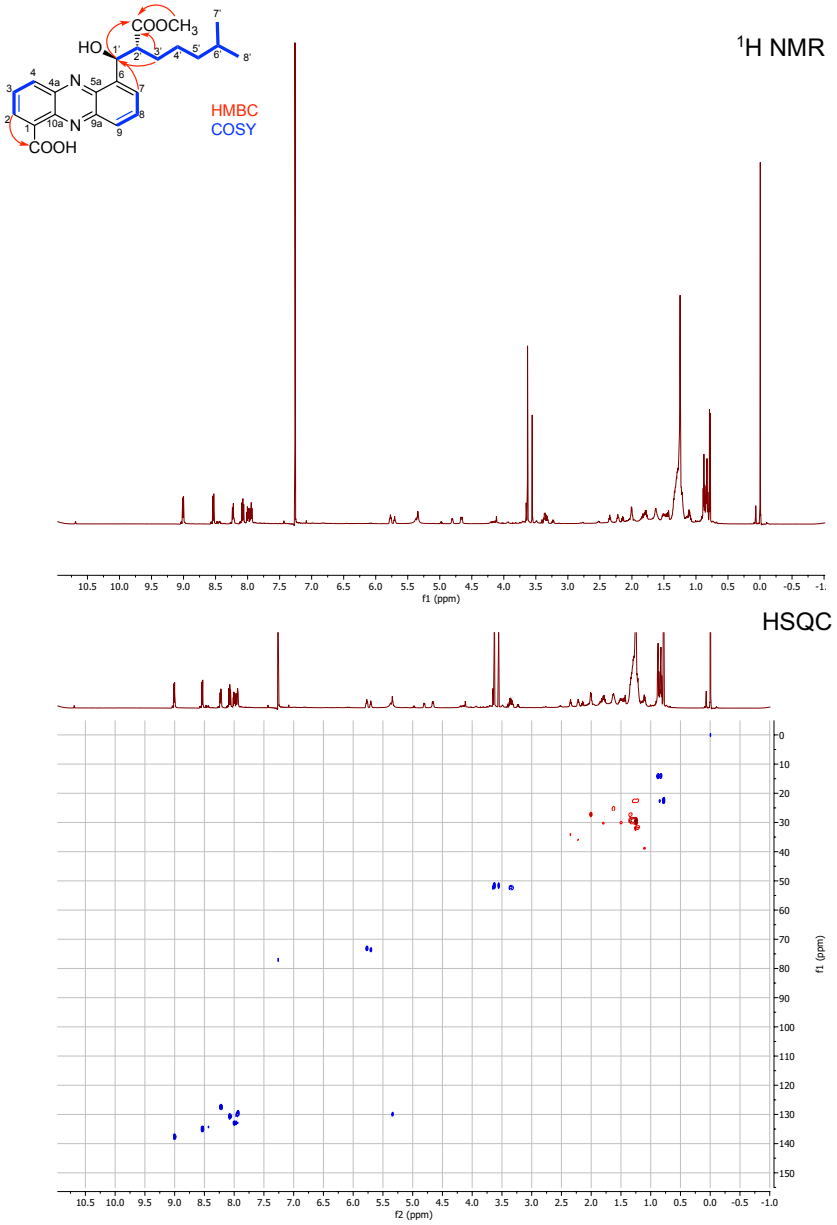

HMBC
f2 (ppm)
f1 (ppm)


COSY
f2 (ppm)
f1 (ppm)

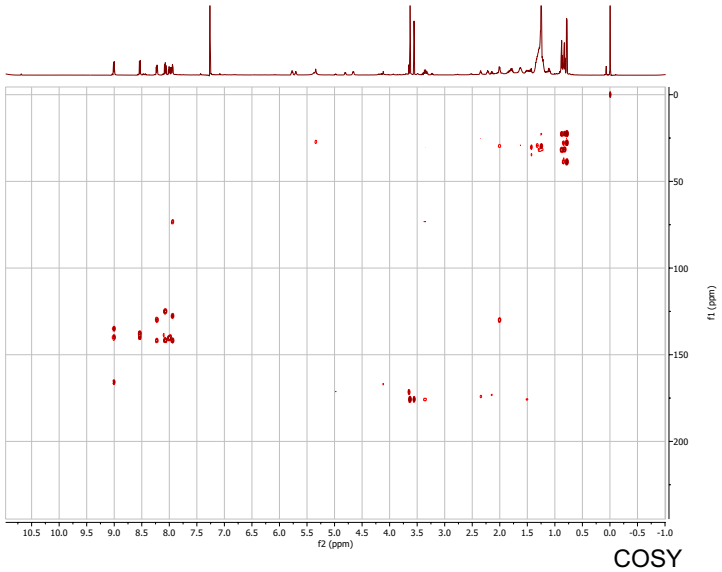
NMR spectroscopic data of compound **13** (streptophenazine C).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	124.93			
2	137.59	9.01, dd (1.43, 7.06)	4, 10a, **1-CO**OH	3,
3	130.69	8.07, dd (7.06, 8.71)	1, 4a	2, 4
4	134.89	8.54, dd (1.44, 8.66)	2, 10a	3
4a	142			
5a	142.2			
6	140.6			
7	129.68	7.95, m	1', 9, 5a	8
8	132.98	8.0, m	6, 9a	7, 9
9	127.58	8.23, ddd (1.53, 5.44, 8.49)	5a, 7	8
9a	140.0			
10a	140.62			
1'	73.38	5.73, dt (7.20, 7.20, 38.29)	3', 2', 7, 5a, **2'-CO**OCH_3	2'
2'	52.25	3.37, m	3', 4', 1', **2'-CO**OCH_3	1', 3'
3'	30.12	1.48 and 1.82	1', 2', 5', **2'-CO**OCH_3	4'
4'	27.1	1.34	2', 6'	3', 5'
5'	38.7	1.1	7'/8', 3'	4', 6'
6'	27.8	1.43	7'/8', 3', 5'	7'/8', 5'
7'	22.37	0.78, d (6.61)	8', 6', 5'	6'
8'	22.37	0.78, d (6.61)	7', 6', 5'	6'
1-COOH	166.85			
2'-COOCH_3	175.98			
2'-COO**CH$_3$**	51.74	3.63 and 3.56	**2'-CO**OCH_3	

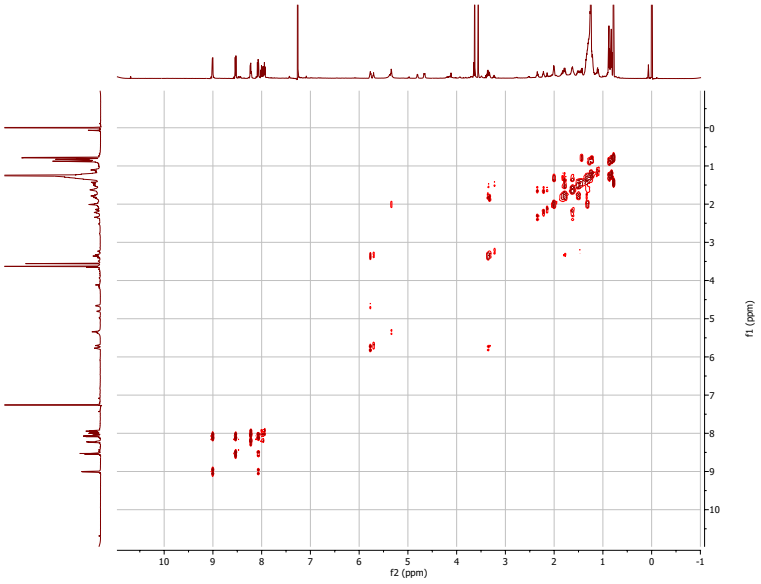
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **13** (streptophenazine C).



HMBC



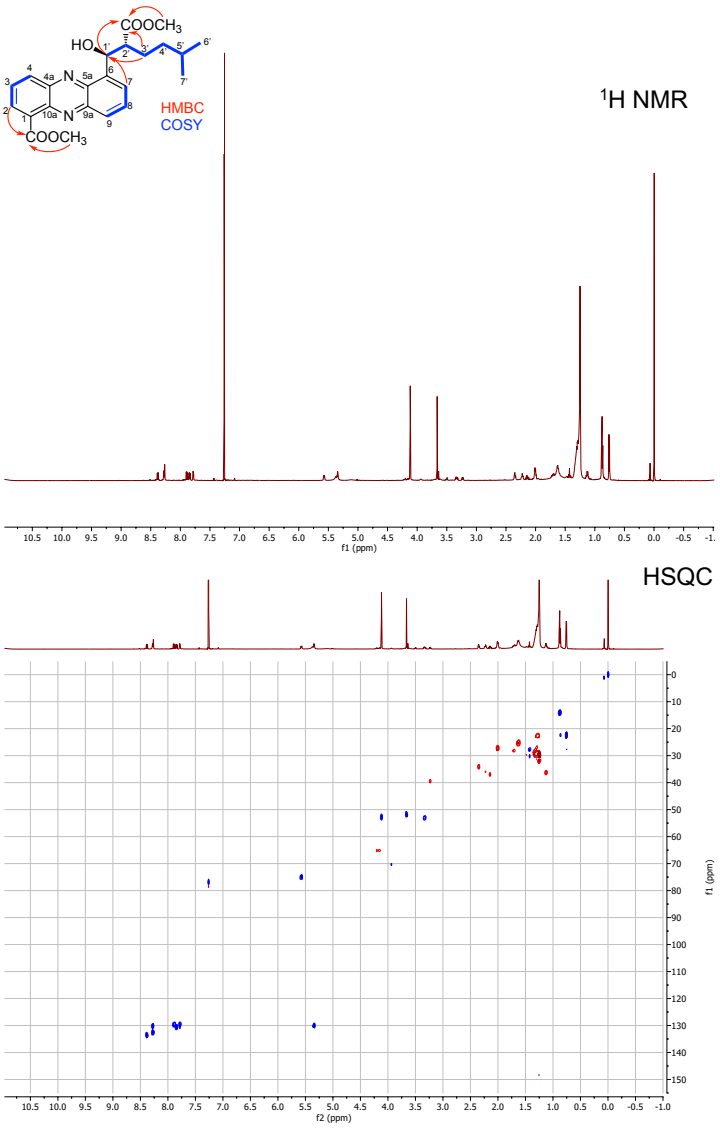
COSY



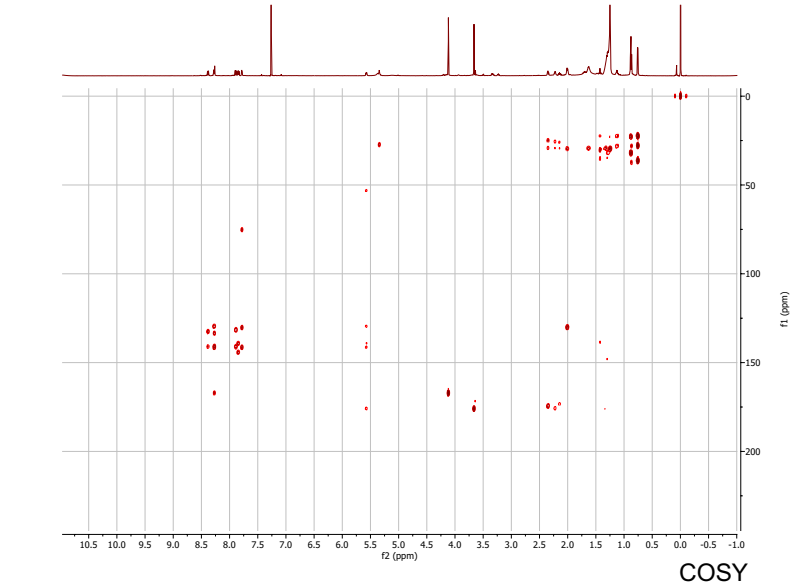
NMR spectroscopic data of compound **14** (streptophenazine D).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.22			
2	132.36	8.26, m	4, 10a, **1-CO**OCH_3	3
3	129.46	7.89, dd (6.90, 8.67)	1, 4a	2, 4
4	133.39	8.38, dd (1.44, 8.70)	2, 10a	3
4a	140.86			
5a	140.93			
6	139.04			
7	129.43	7.78, d (6.72)	1', 5a, 9	8
8	130.48	7.85, dd (6.77, 8.71)	6, 9a	7, 9
9	130.03	8.27, m	7, 5a	8
9a	144.27			
10a	140.68			
1'	75.29	5.57, d (7.53)	3', 2, 5a, 7, **2'-CO**OCH_3	2'
2'	53.02	3.33, ddd (5.03, 7.49, 9.87)	3', 4', 1', **2'-CO**OCH_3	1', 3'
3'	28.91	1.71, dtd (6.34, 9.3, 10.00, 13.43), 1.32	1', 2', 4', 5', **2'-CO**OCH_3	2', 4'
4'	36.22	1.13	2', 3', 5', 6'/7'	3', 5'
5'	27.85	1.42	3', 4', 6'/7'	6'/7', 4'
6'	22.37	0.76, dd (1.66, 6.64)	4', 5', 7'	5'
7'	22.37	0.76	4', 5', 6'	5'
1-COOCH_3	167.1		**1-CO**OCH_3	
1-COO**CH_3**	52.57	4.12, s		
2'-COOCH_3	175.69	-		
2'-COO**CH_3**	51.56	3.66, s	**2'-CO**OCH_3	

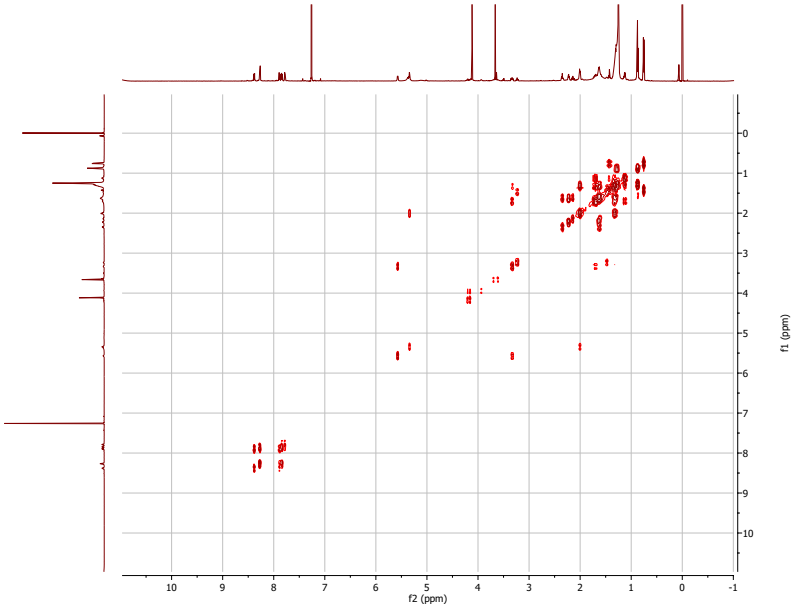
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **14** (streptophenazine D).



HMBC



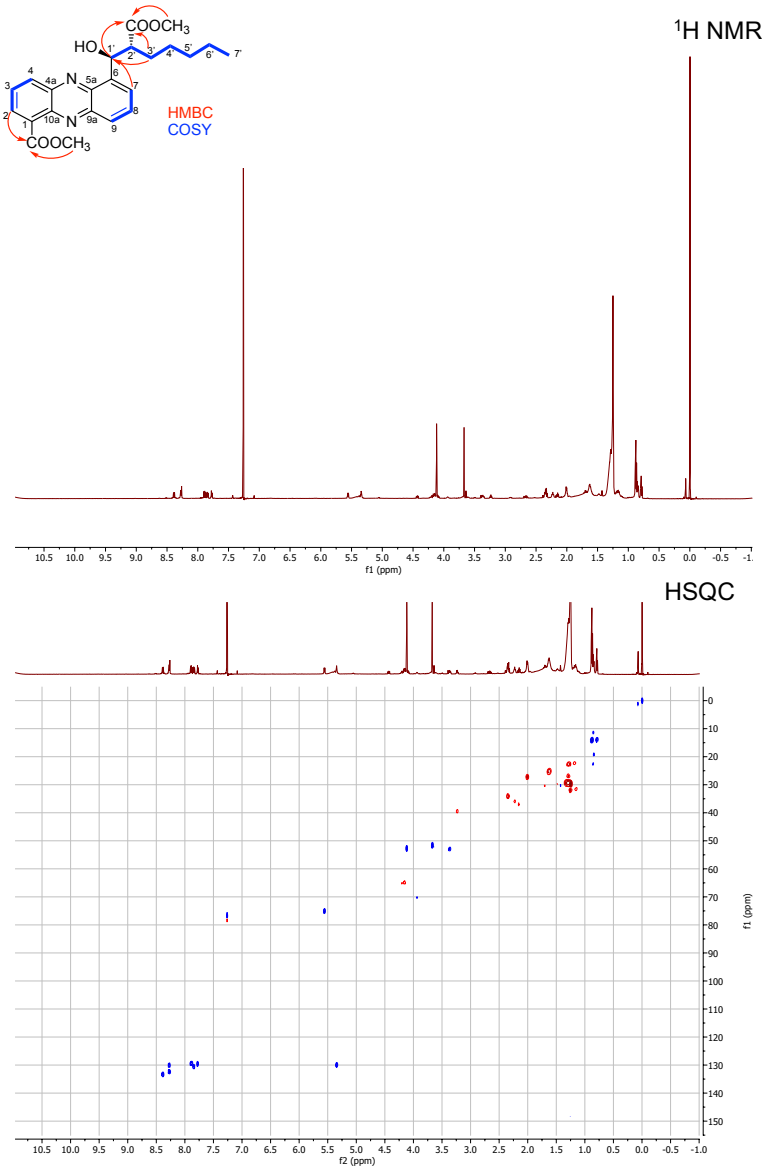
COSY



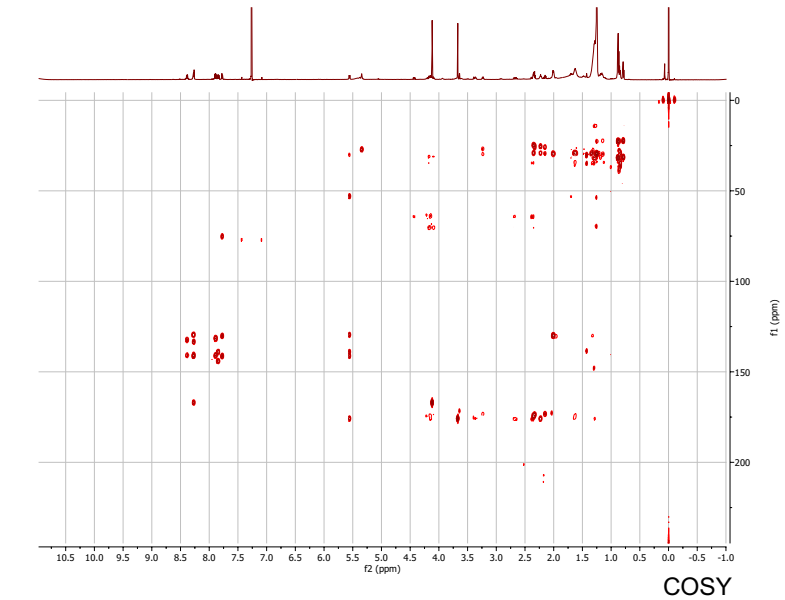
NMR spectroscopic data of compound **15** (streptophenazine P).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	132			
2	132.2	8.27, dd (1.71, 1.76)	4, 10a, **1-CO**OCH_3	3
3	129.1	7.88, dd (6.90, 8.73)	1, 4a	2, 4
4	133.4	8.39, dd (1.42, 8.67)	2, 10a	3
4a	141			
5a	141.4			
6	138.9			
7	129.33	7.77, d (6.76)	1', 5a, 9	8
8	130.4	7.85, dd (6.79, 8.74)	6, 9a	7, 9
9	130.1	8.28, dd (1.40, 3.88)	7, 5a	8
9a	144.08			
10a	141			
1'	75.1	5.56, d (7.63)	2', 3', 7, 5a, 9, **2'-CO**OCH_3	2'
2'	52.9	3.36, ddd (4.41, 7.37, 9.95)	3', 4', 1', **2'-CO**OCH_3	1', 3'
3'	29.4	1.68 and 1.29	1', 2', 5', **2'-CO**OCH_3	2', 4'
4'	29.7	1.3		3', 5'
5'	31.8	1.25	3', 7'	4', 6'
6'	22.37	1.18	5'	5', 7'
7'	14.01	0.79, br t (7.04, 7.04)	6', 5'	6'
1-COOCH_3	167.3			
1-COO**CH_3**	52.7	4.12, s	**1-CO**OCH_3	
2'-COOCH_3	175.6			
2'-COO**CH_3**	51.7	3.67, s	**2'-CO**OCH_3	

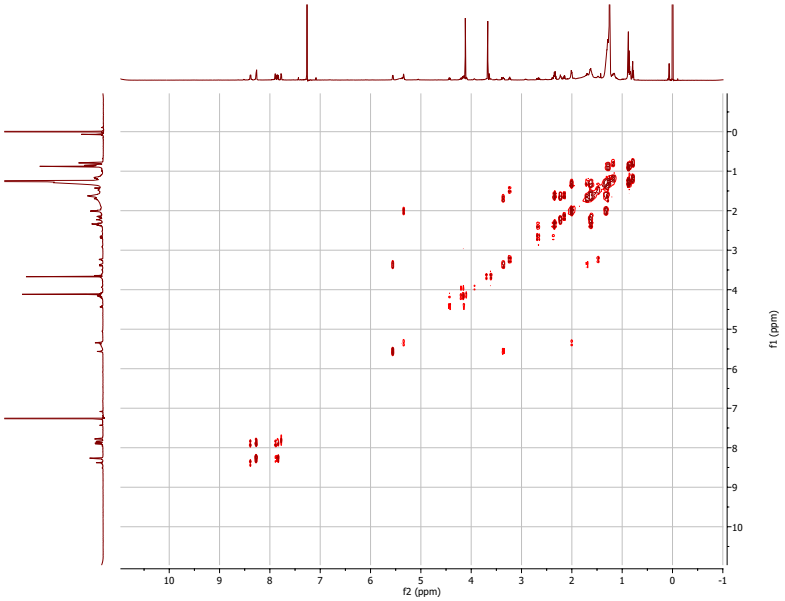
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **15** (streptophenazine P).



HMBC



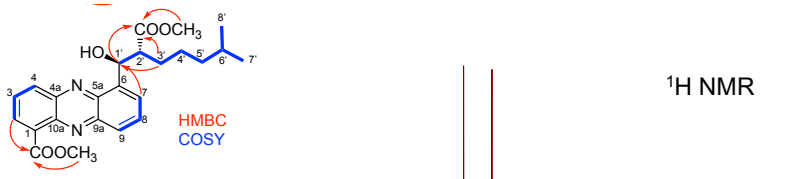
COSY

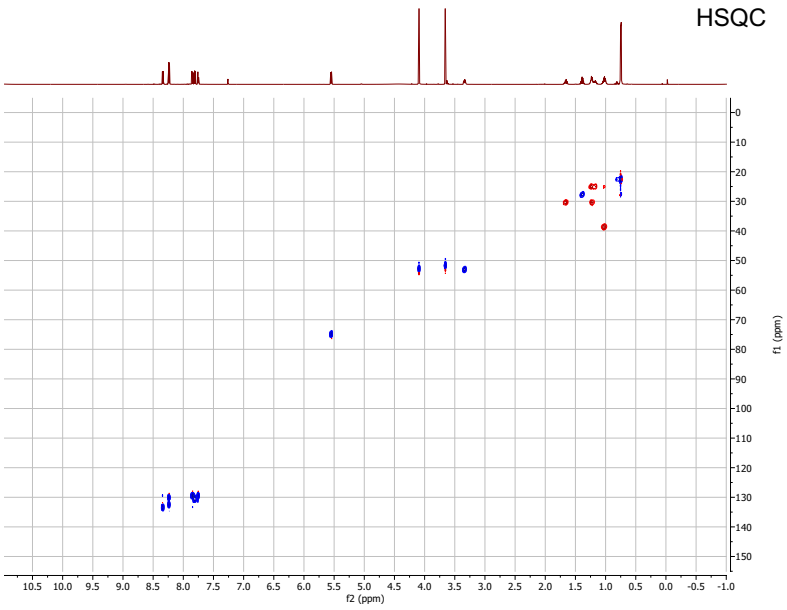


NMR spectroscopic data of compound **16** (streptophenazine A).

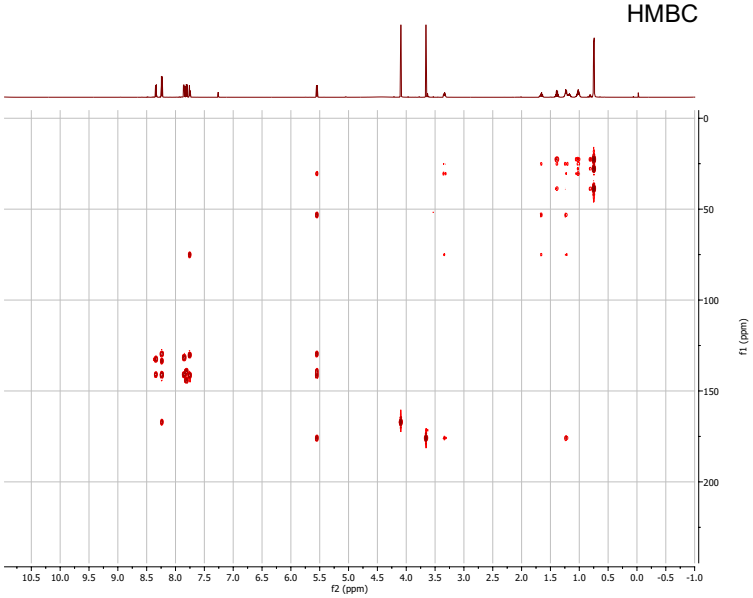
Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.5			
2	132.5	8.24, br t (1.35, 1.35)	4, 10a, **1-CO**OCH_3	3
3	129.4	7.84, dd (6.91, 8.72)	1, 4a	2, 4
4	133.5	8.34, dd (1.46, 8.71)	2, 10a	3
4a	141.1			
5a	140.9			
6	138.9			
7	129.35	7.75, dt (1.07, 1.07, 6.79)	1', 5a, 9	8
8	130.7	7.81, dd (6.80, 8.72)	6, 9a	7, 9
9	129.9	8.23, d (1.45)	7, 5a	8
9a	144			
10a	140.6			
1'	74.99	5.55, d (7.68)	2', 3', 5a, 7, **2'-CO**OCH_3	2'
2'	52.95	3.35, m	1', 3', 4', 6, **2'-CO**OCH_3	1', 3'
3'	30.34	1.24, 1.66	1', 2', 4', 5', **2'-CO**OCH_3	2', 4'
4'	24.9	1.18	2', 6'	5', 3'
5'	38.6	1.01, tddd (2.53, 5.20, 6.91, 9.28, 9.28)	3', 4', 7'/8'	4', 6'
6'	27.8	1.4, dt (6.66, 6.66, 13.31)	5', 7'	7'/8', 5'
7'	22.37	0.75, dd (2.13, 6.66)	7'/8', 6', 5'	6'
8'	22.37	0.75, dd (2.13, 6.66)	7'/8', 6', 5'	6'
1-COOCH_3	166.9			
1-COO**CH_3**	52.3	4.10, s	**1-CO**OCH_3	-
2'-COOCH_3	175.7			
2'-COO**CH_3**	51.68	3.66, s	**2'-CO**OCH_3	-

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **16** (streptophenazine A).

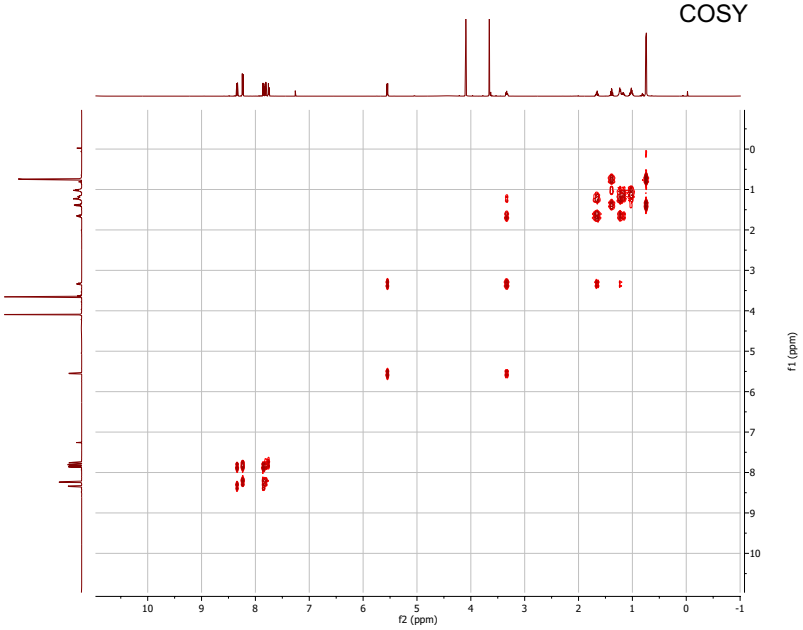




HMBC


0
50
100
150
200
f1 (ppm)
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f2 (ppm)

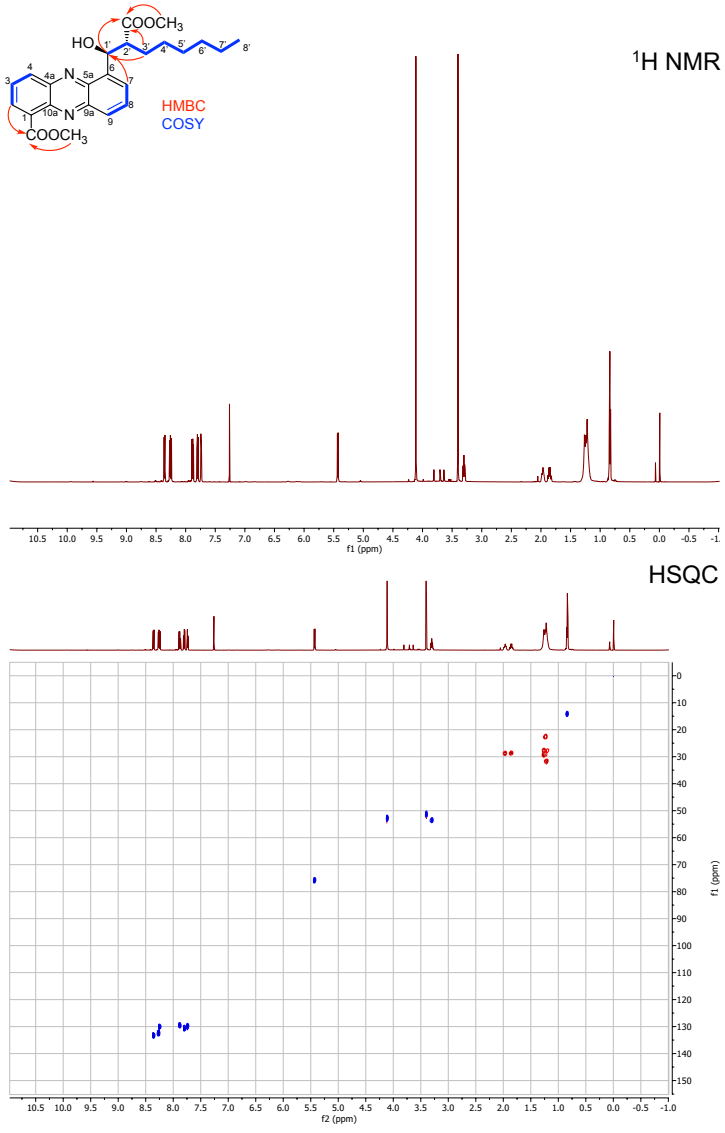
COSY


0
1
2
3
4
5
6
7
8
9
10
f1 (ppm)
10 9 8 7 6 5 4 3 2 1 0 -1
f2 (ppm)

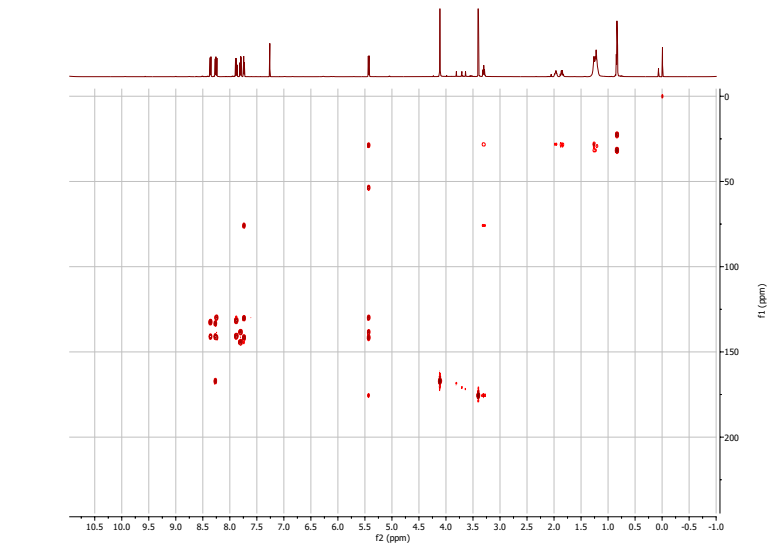
NMR spectroscopic data of compound **17** (streptophenazine B).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.9			
2	132.4	8.26, dd (1.44, 6.96)	4, 10a, **1-CO**OCH_3	3
3	129.5	7.88, dd (6.90, 8.71)	1, 4a	2, 4
4	133.2	8.36, dd (1.45, 8.71)	2, 10a	3
4a	140.5			
5a	141.4			
6	138.2			
7	129.7	7.74, dt (1.10, 1.10, 6.75)	1', 5a, 9	8
8	130.5	7.8, dd (6.80, 8.76)	6, 9a	7, 9
9	130.0	8.24 dd (1.43, 8.73)	7, 5a	8
9a	144.1			
10a	140.7			
1'	75.8	5.43, d (7.75)	2', 3', 5a, 6, 7, **2'-CO**OCH_3	2'
2'	53.4	3.30, m	1', 4', 6, **2'-CO**OCH_3	1', 3'
3'	28.7	1.85, dtd (4.09, 10.54, 10.63, 14.25) and 1.95, tdd (2.39, 5.51, 9.30, 9.30)	1', 2', 5', **2'-CO**OCH_3	2', 4'
4'	28.2	1.15-1.3	2', 6'	3', 5'
5'	28.2	1.15-1.3	3', 7'	6'
6'	31.5	1.15-1.3	4', 8'	5', 7'
7'	22.4	1.15-1.3	5', 6'	6', 8'
8'	14.13	0.84, br t (7.00)	6', 7'	7'
1-COOCH_3	167.2			
1-COO**CH_3**	52.8	4.11, s	**1-CO**OCH_3	-
2'-COOCH_3	175.4			
2'-COO**CH_3**	51.4	3.40, s	**2'-CO**OCH_3	-

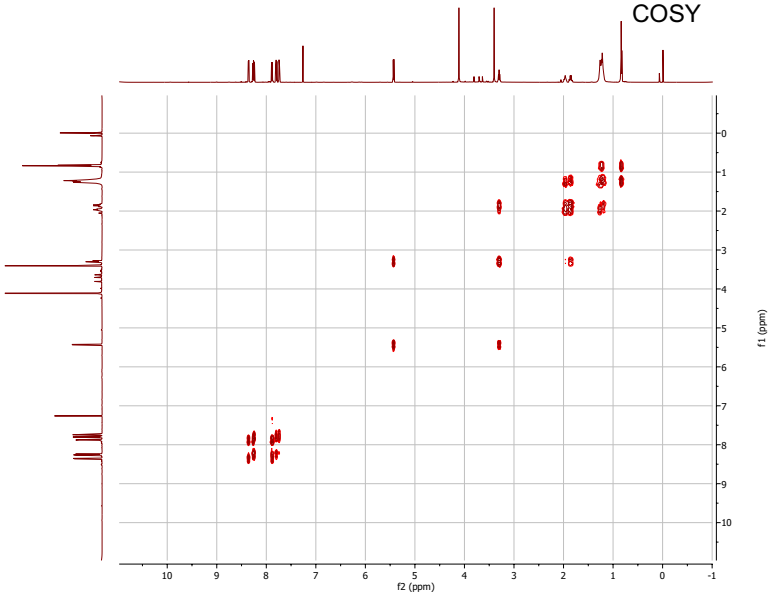
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **17** (streptophenazine B).



HMBC



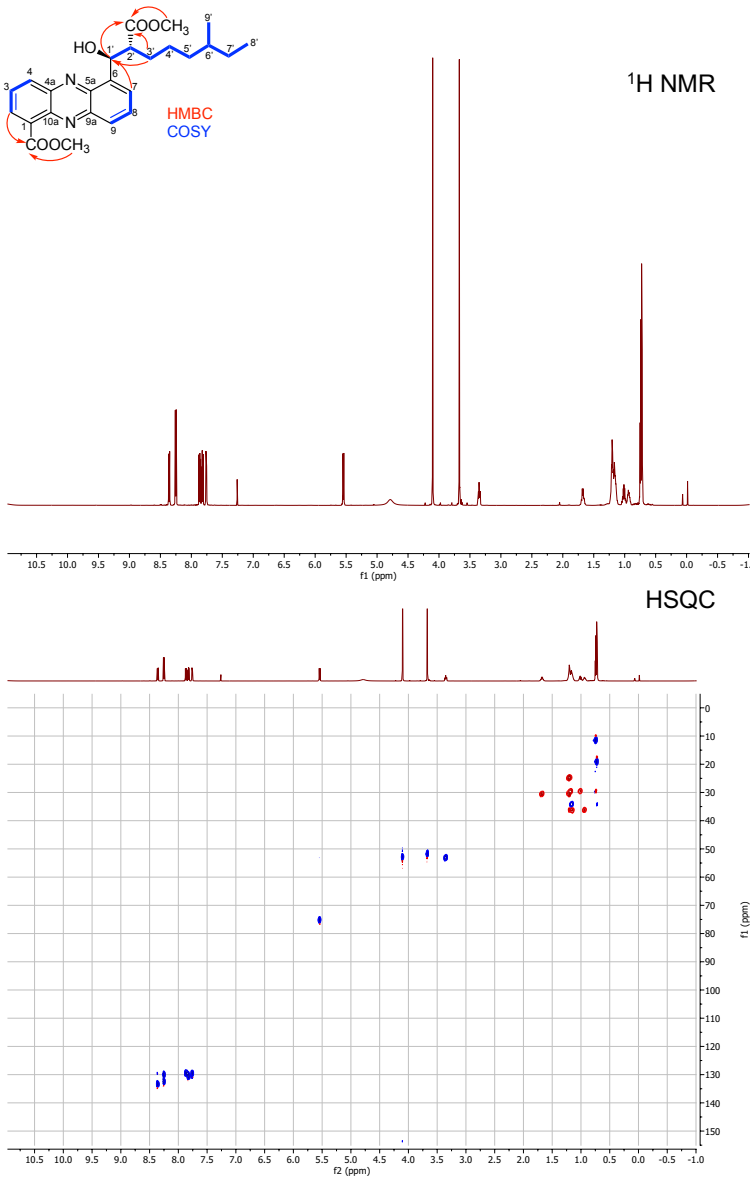
COSY



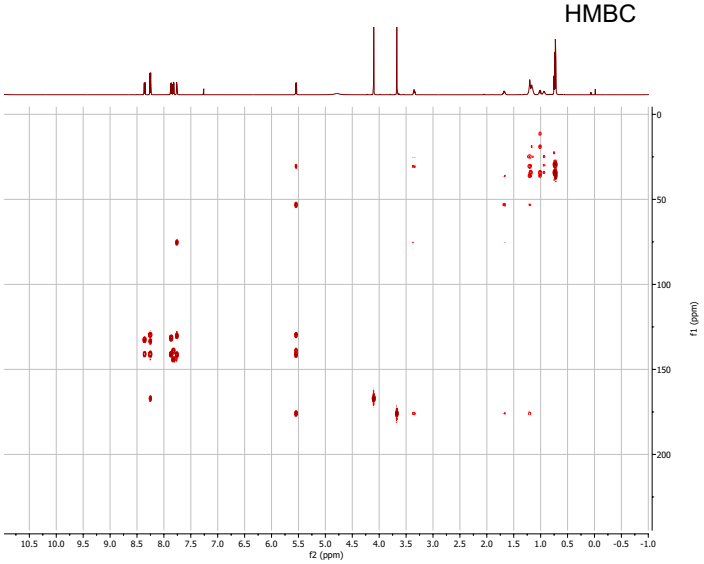
NMR spectroscopic data of compound **18** (streptophenazine G).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.6			
2	132.4	8.26, dd (1.43, 1.43)	4, 10a, **1-CO**OCH_3	3
3	129.5	7.85, dd (6.92, 8.71)	1, 4a	2, 4
4	133.3	8.36, dd (1.43, 8.70)	2, 10a	3
4a	140.7			
5a	141.05			
6	139.2			
7	129.5	7.76, dt (1.11, 1.11, 6.71)	1', 5a, 9	8
8	130.5	7.83, dd (6.77, 8.75)	6, 9a	7, 9
9	129.9	8.25, d (1.43)	7, 5a	8
9a	144.1			
10a	141.1			
1'	75.1	5.54, d (7.82)	2', 3', 5a, 7, **2'-CO**OCH_3	
2'	53.1	3.35, ddd (3.9, 7.61, 9.73)	1', 3', 4', **2'-CO**OCH_3	3', 1'
3'	30.5	1.21 and 1.68	1', 2', 4', 5', **2'-CO**OCH_3	2', 4'
4'	24.8	1.2	2', 3', 6'	3', 5'
5'	36.15	0.94, dtd (3.07, 4.93, 6.22, 10.72) and 1.15	3', 4', 6', 7'	4', 6'
6'	34.2	1.17	4', 8', 9'	5', 7', 9'
7'	29.56	1.02 and 1.19	5', 6', 8', 9'	6', 8'
8'	11.37	0.74, m	5', 6', 7'	6', 7'
9'	19.08	0.72, m	5', 6', 7'	6', 7'
1-COOCH_3	167.2			
1-COO**CH_3**	52.7	4.1, s	**1-CO**OCH_3	
2'-COOCH_3	175.8			
2'-COO**CH_3**	51.6	3.67, s	**2'-CO**OCH_3	

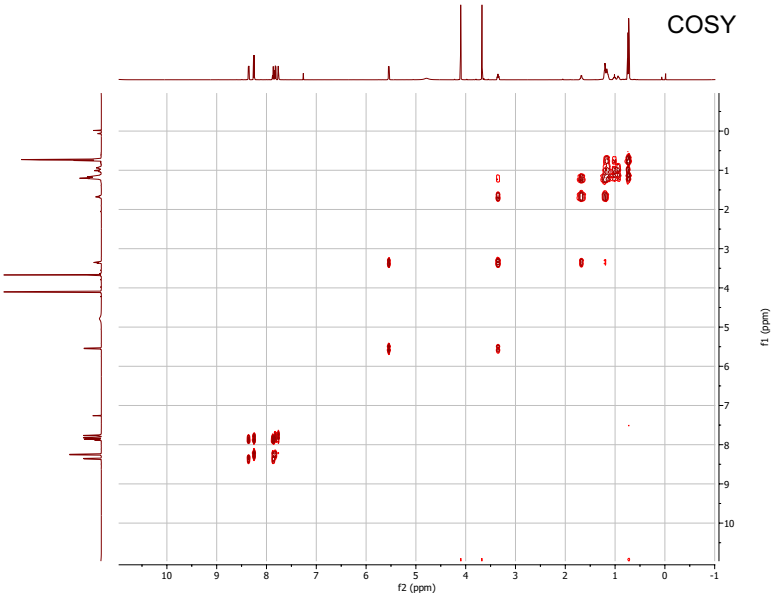
^{1}H, HSQC, HMBC, and COSY NMR spectra key HMBC and COSY correlations of compound **18** (streptophenazine G).



HMBC



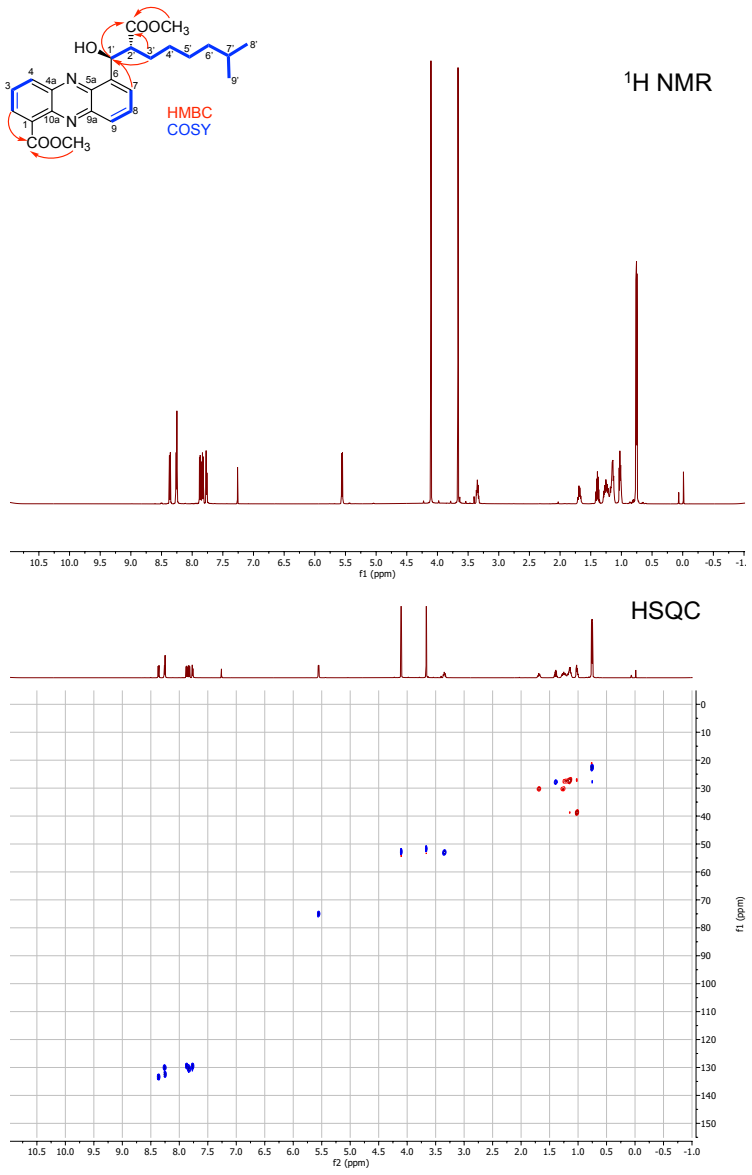
COSY



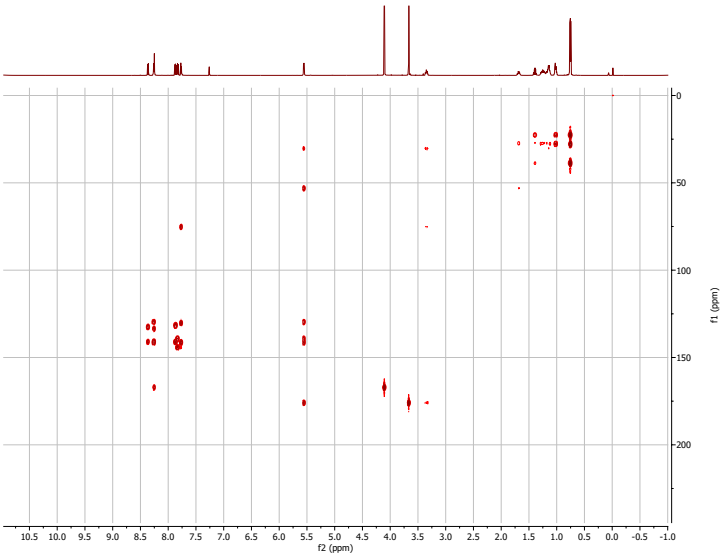
NMR spectroscopic data of compound **19** (streptophenazine F).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.9			
2	132.6	8.26, dd (1.41, 3.51)	4, 10a, **1-CO**OCH_3	3
3	129.4	7.86, dd (6.90, 8.73)	1, 4a	2, 4
4	133.4	8.36, dd (1.44, 8.71)	2, 10a	3
4a	140.9			
5a	141.1			
6	139.1			
7	129.6	7.77, dt (1.04, 1.04, 6.75)	1', 5a, 9	8
8	130.5	7.83, dd (6.7, 8.73)	6, 9a	7, 9
9	130.1	8.25, t (1.55, 1.55)	7, 5a	8
9a	144.2			
10a	140.8			
1'	75.1	5.56, d (7.69)	2', 2', 5a, 7, **2'-CO**OCH_3	2'
2'	53.1	3.35, ddd (4.61, 7.64, 10.16)	1', 3', 4', 6, **2'-CO**OCH_3	1', 3'
3'	30.3	1.68 and 1.27	1', 2', 5', **2'-CO**OCH_3	2', 4'
4'	27.7	1.15, dtt (3.22, 3.22, 5.32, 5.32, 12.64)	2', 3', 5', 6' 7'	3', 5'
5'	27.2	1.18	3', 7'	6'
6'	38.5	1.01	8'/9', 4'	5', 7'
7'	27.8	1.4, dt (6.07, 6.67, 13.28)	8'/9', 5', 6'	6', 8'/9'
8'	22.5	0.76, dd (4.07, 6.70)	8'/9', 7', 6'	7'
9'	22.5	0.76, dd (4.07, 6.70)	8'/9', 7', 6'	7'
1-COOCH_3	167			
1-COO**CH_3**	52.7	4.11, s	**1-CO**OCH_3	
2'-COOCH_3	175.9			
2'-COO**CH_3**	51.5	3.66, s	**2'-CO**OCH_3	

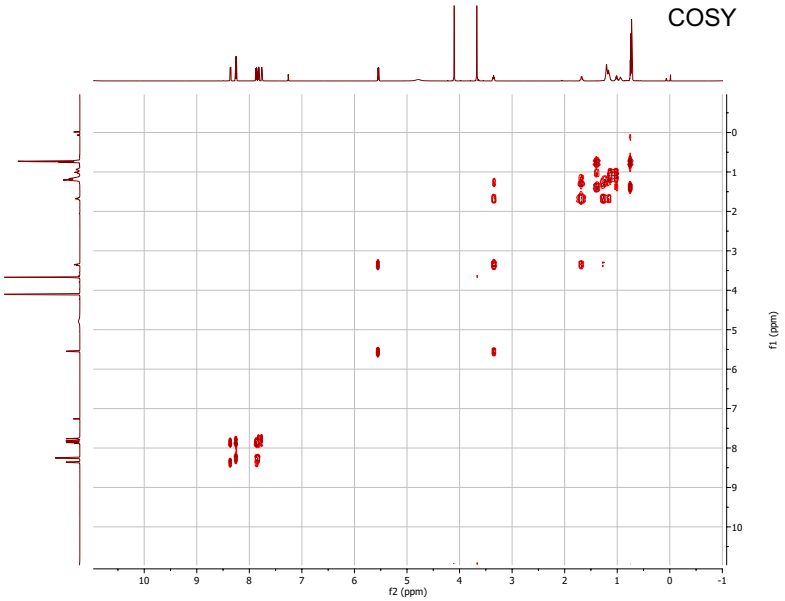
^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **19** (streptophenazine F).



HMBC


0
50
100
150
200
f1 (ppm)
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f2 (ppm)

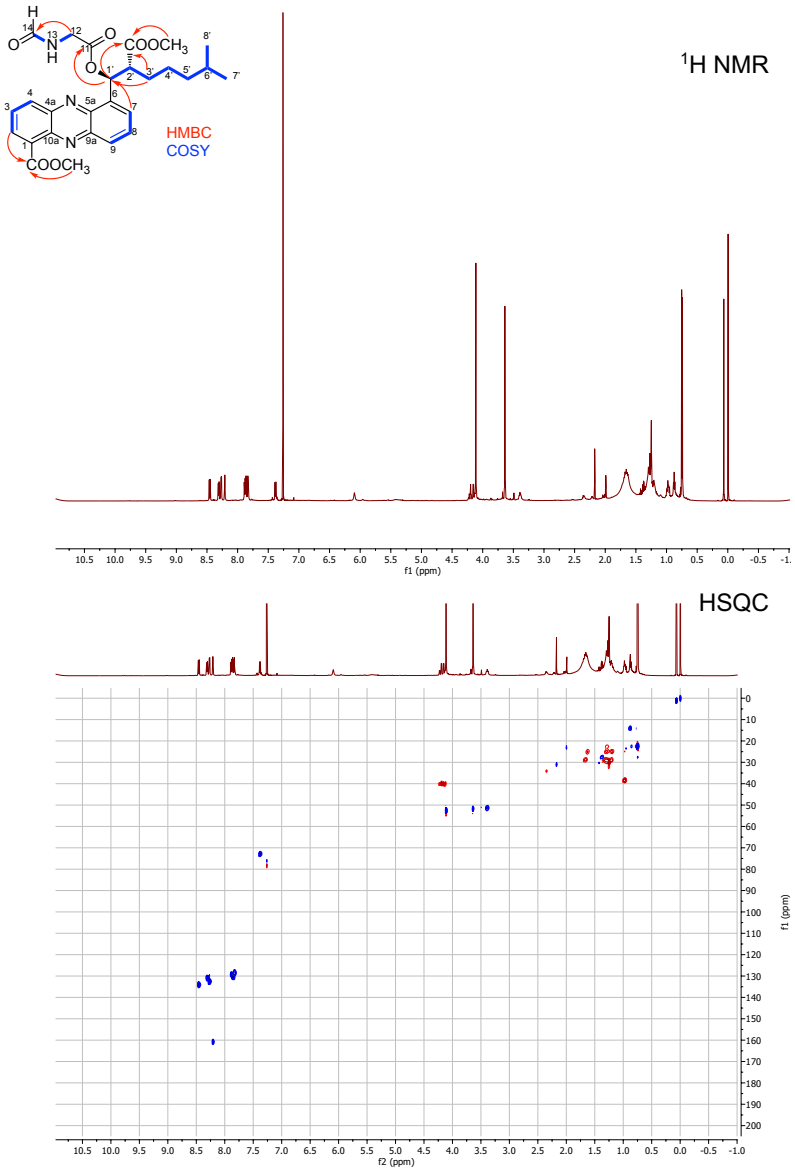
COSY

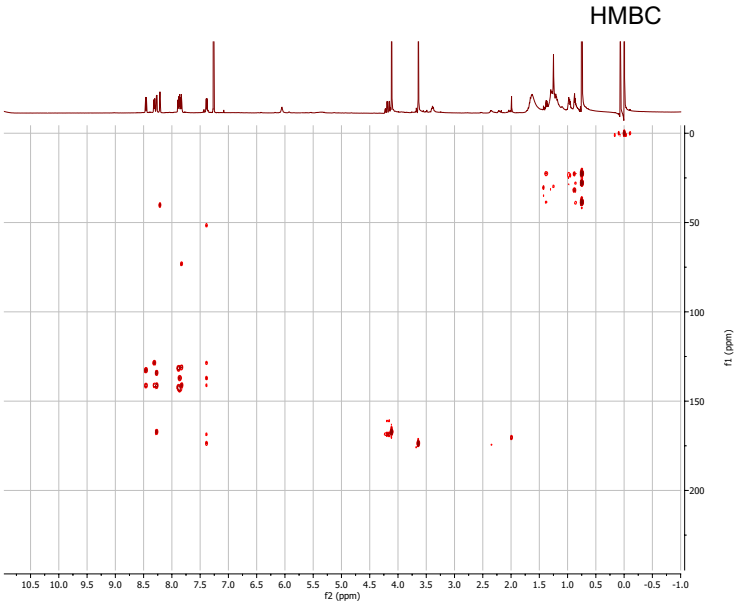

0
1
2
3
4
5
6
7
8
9
10
f1 (ppm)
10 9 8 7 6 5 4 3 2 1 0 -1
f2 (ppm)

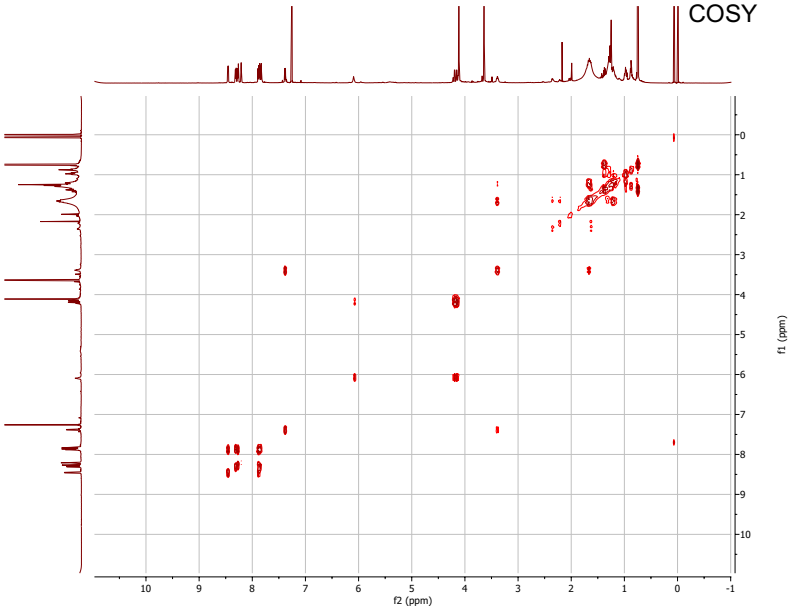
NMR spectroscopic data of compound **20** (streptophenazine Q).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	136.3			
2	133	8.27, dd (1.42, 6.90)	4, 10a, **1-CO**OCH_3	3
3	129.3	7.9, m	1, 4a	2, 4
4	133.7	8.46, dd (1.46 8.72)	2, 10a	3
4a	142.3			
5a	141.3			
6	136.8			
7	129.5	7.9, m	1', 5a, 9	8
8	129.5	7.9, m	6, 9a	7, 9
9	131	8.31, dd (1.53, 8.54)	7, 5a	8
9a	142.1			
10a	141.5			
11	168			
12	40.3	4.19, d (4.98); 4.16, d (5.08)	14	13
13	-	6.05, s	-	12, 14
14	161	8.21, s	12	13
1'	73.1	7.39, d (8.10)	2', 5a, 6, 7, 11, **2'-CO**OCH_3	2'
2'	51.1	3.39, m	1', 3', **2'-CO**OCH_3	1', 3'
3'	28.9	1.67, 1.25	1', 5' **2'-CO**OCH_3	2', 4'
4'	24.9	1.18	2', 5', 6'	3', 5'
5'	38.7	1.01	3', 7'/ 8'	4', 6'
6'	27.7	1.39	5', 4', 7'/ 8'	7', 8', 5'
7'	22.6	0.75, d (6.60)	5', 6', 8'	8', 6'
8'	22.6	0.75, d (6.60)	5', 6', 7'	6', 7'
1-COOCH_3	167.3			
1-COO**CH_3**	52.4	4.11, s	**1-**COO**CH_3**	
2'-COOCH_3	173.4			
2'-COO**CH_3**	51.4	3.64, s	**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **20** (streptophenazine Q).



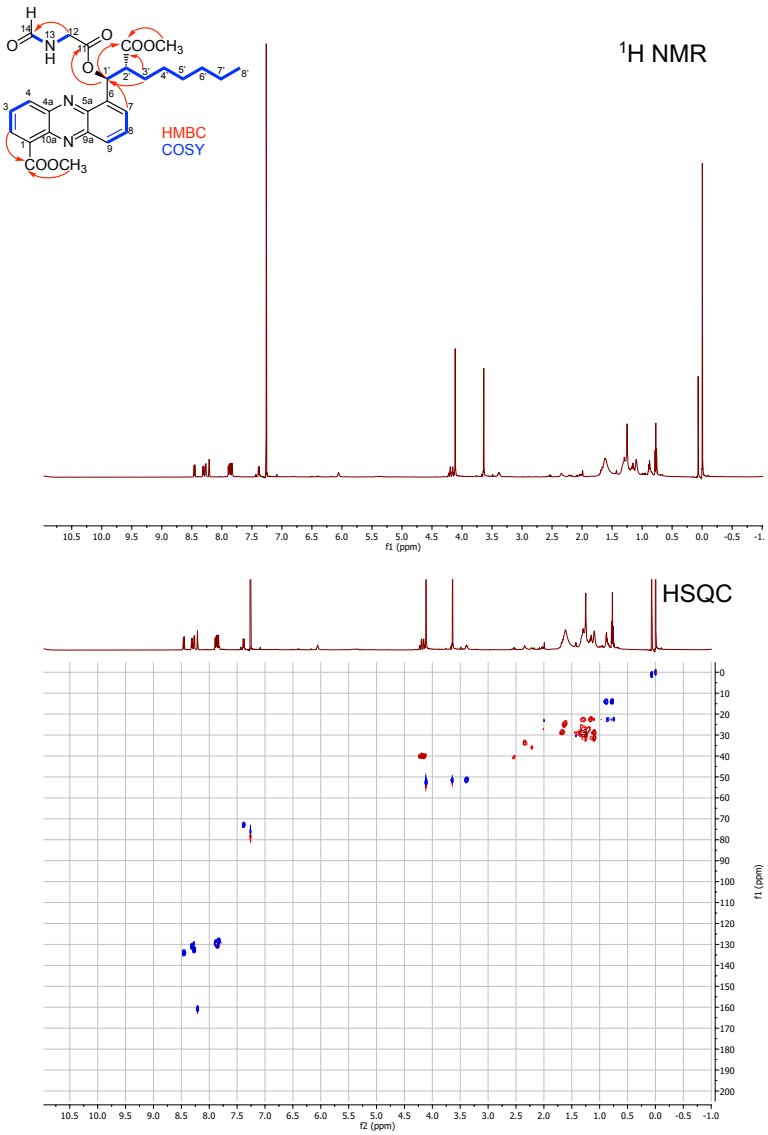

HMBC
f2 (ppm)
f1 (ppm)

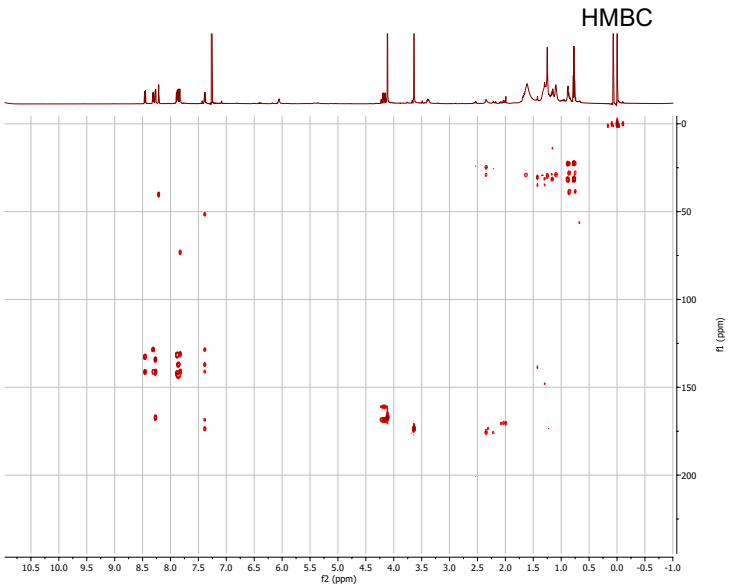

COSY
f2 (ppm)
f1 (ppm)

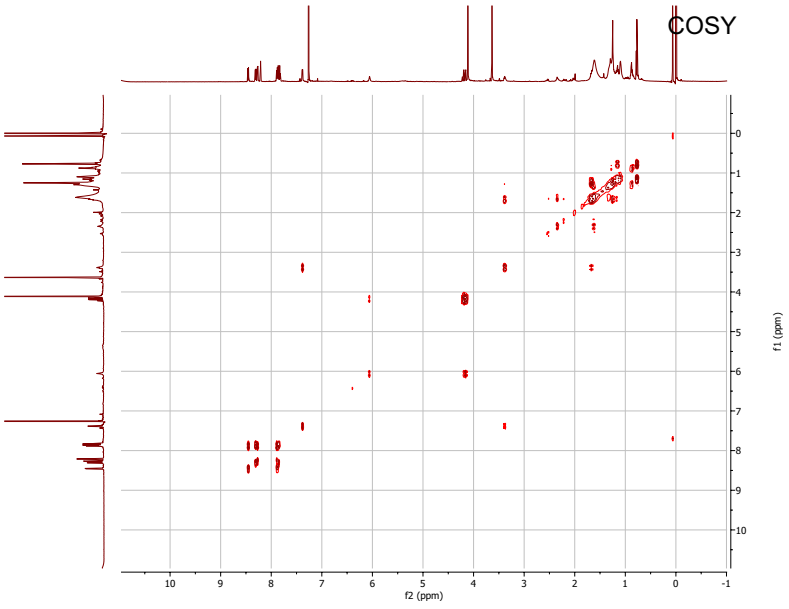
NMR spectroscopic data of compound **21** (streptophenazine R).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	136.6			
2	132.2	8.27, dd (1.46, 6.90)	4, 10a, **1-CO**OCH_3	3
3	129.5	7.9, m	1, 4a	2, 4
4	134.2	8.46, dd (1.44, 8.69)	2, 10a	3
4a	142			
5a	141.5			
6	137			
7	129.5	7.9, m	1', 5a, 9	8
8	129.5	7.9, m	6, 9a	7, 9
9	130.9	8.31, dd (1.54, 8.54)	7, 5a	8
9a	142			
10a	141.3			
11	168			
12	40.2	4.19, d (5.07); 4.16, d (5.13)	14	13
13		6.06, s		12, 14
14	161	8.21, s	12	13
1'	72.8	7.38, d (7.92)	2', 3', 5a, 6, 7, 11, **2'-CO**OCH_3	2'
2'	51.3	3.39, td (4.25, 9.11, 8.75)	1', 3', **2'-CO**OCH_3	1', 3'
3'	28.8	1.65, 1.27	1', 5', **2'-CO**OCH_3	2', 4'
4'	29.8	1.22	2', 5', 6'	3', 5'
5'	29.2	1.29	3', 6', 7'	4', 6'
6'	31.65	1.10	4', 8'	5', 7'
7'	22.5	1.16	5', 6', 8'	6', 8'
8'	14.1	0.77, br t (7.20)	6', 7'	7'
1-COOCH_3	167.1			
1-COO**CH_3**	52.5	4.11, s	**1-CO**OCH_3	
2'-COOCH_3	173.3			
2'-COO**CH_3**	51.6	3.64, s	**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **21** (streptophenazine R).



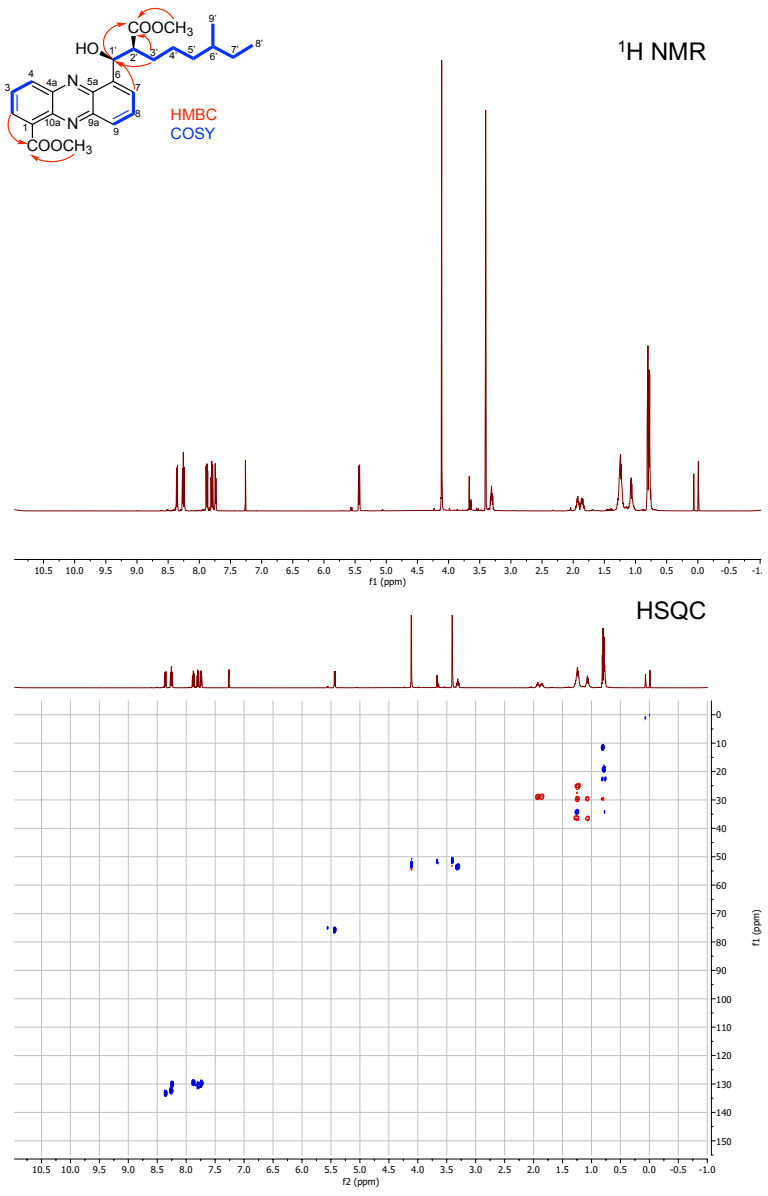

HMBC
f2 (ppm)
f1 (ppm)

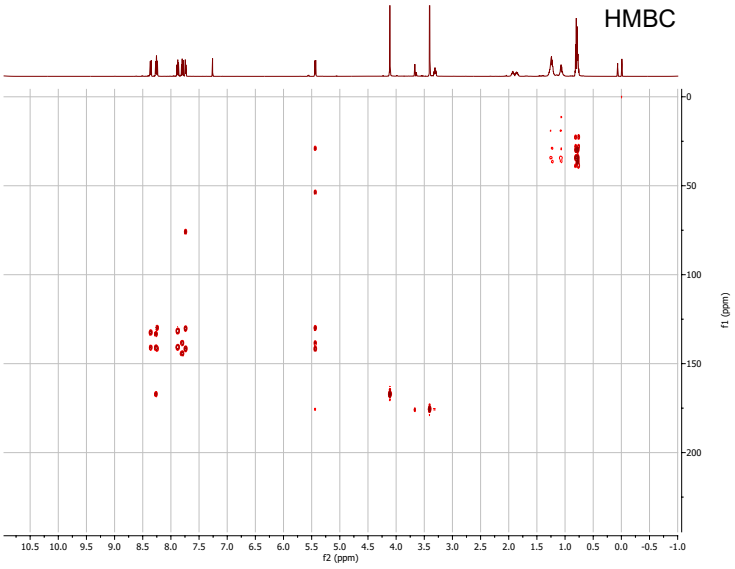

COSY
f2 (ppm)
f1 (ppm)

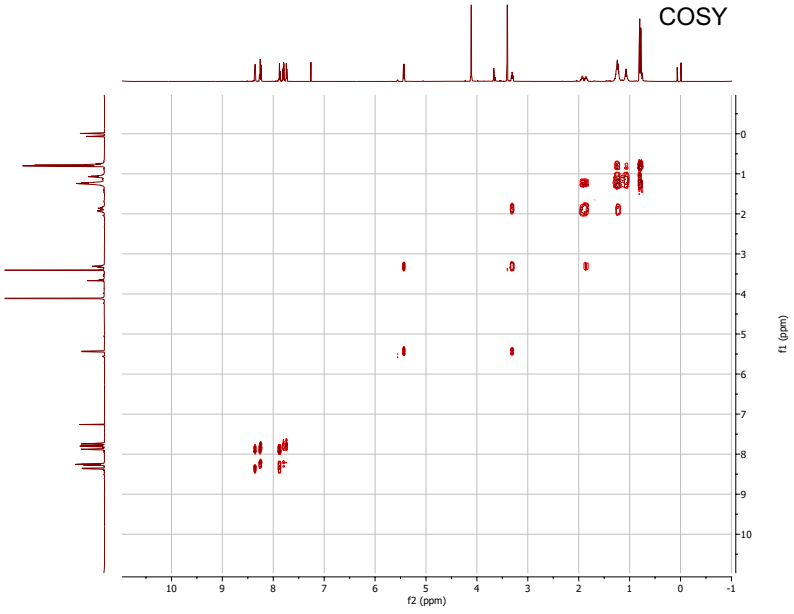
NMR spectroscopic data of compound **22** (diastereomer of streptophenazine G).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.6			
2	132.5	8.26, ddd (1.44, 7.84, 10.14)	4, 10a, **1-CO**OCH_3	3
3	129.5	7.88, dd (6.91, 8.73)	1, 4a	2, 4
4	133.4	8.36, dd (1.44, 8.70)	2, 10a	3
4a	140.6			
5a	141.3			
6	138.2			
7	129.7	7.74, dt (1.09, 1.09, 6.75)	1', 5a, 9	8
8	130.4	7.8, dd (6.79, 8.73)	6, 9a	7, 9
9	130.1	8.25, ddd (1.44, 7.84, 10.14)	7, 5a	8
9a	144.4			
10a	140.8			
1'	76	5.43, d (7.71)	2'. 3', 5a, 6, 7, **2'-CO**OCH_3	2'
2'	53.4	3.31, ddd (3.78, 7.72, 10.64)	1', 3', 6, **2'-CO**OCH_3)	1', 3'
3'	28.9	1.93, 1.85	1', 2', 4', 5', **2'-CO**OCH_3	2', 4'
4'	25.22	1.23	2', 3', 5', 6'	3', 5'
5'	36.21	1.06, 1.26	3', 4', 6', 9'	4', 6'
6'	34.13	1.24	4', 5', 7', 8', 9'	5', 7', 9'
7'	29.61	1.07, 1.22	5', 8', 9'	6', 8'
8'	11.69	0.8	5', 6', 7'	6', 7'
9'	18.97	0.78	5', 6', 7'	6', 7'
1-COOCH_3	167.2			
1-COO**CH_3**	52.8	4.11, s	**1-CO**OCH_3	
2'-COOCH_3	175.3			
2'-COO**CH_3**	51.5	3.4, s	1**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **22** (diastereomer of streptophenazine G).



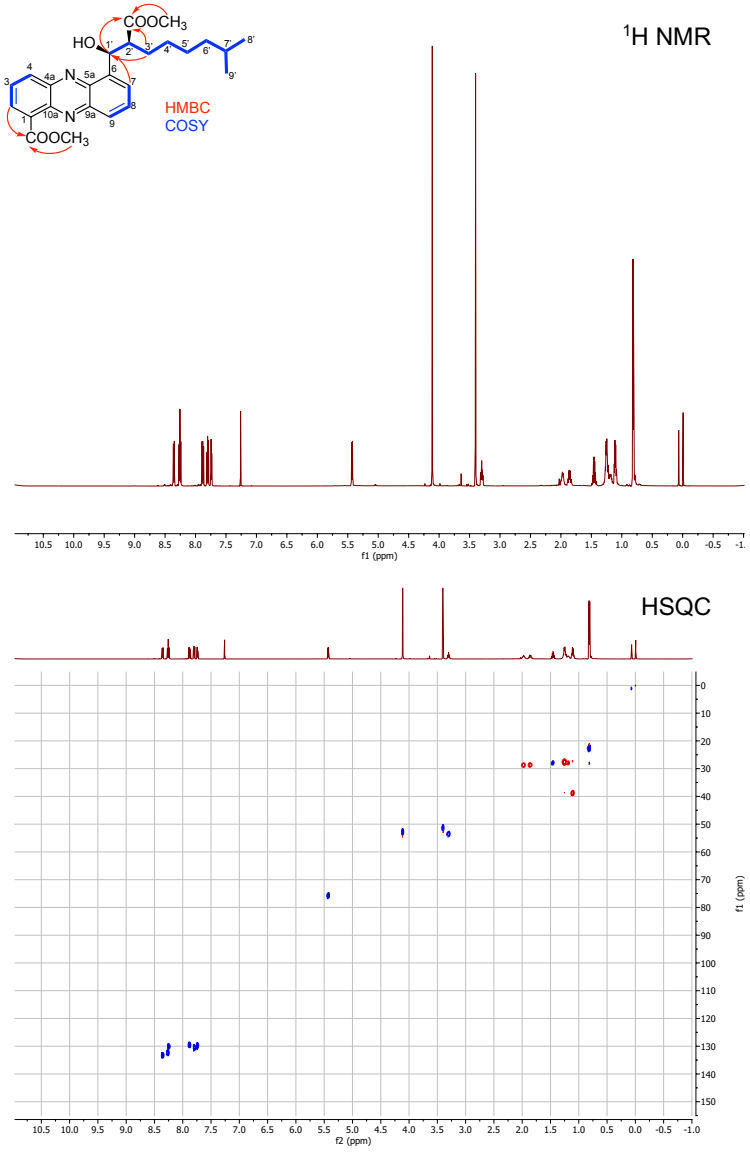

HMBC
f2 (ppm)
f1 (ppm)

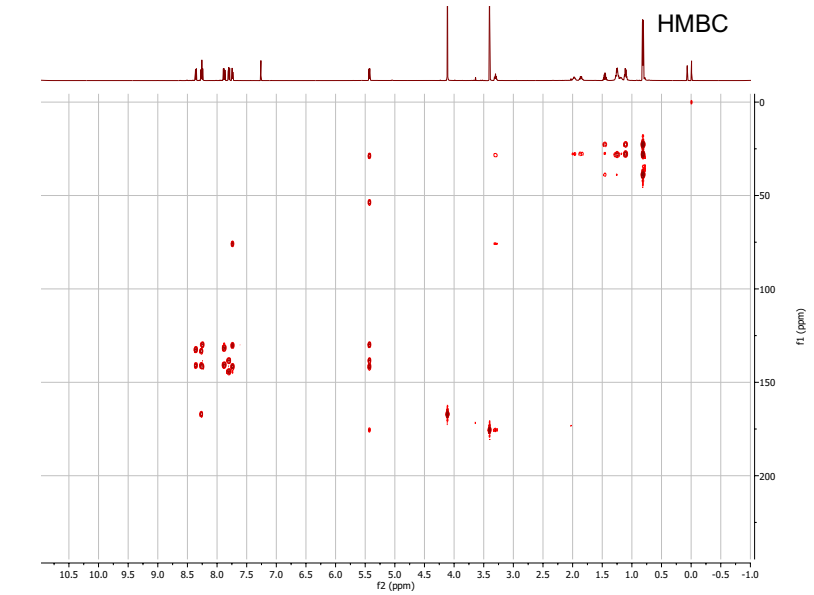

COSY
f2 (ppm)
f1 (ppm)

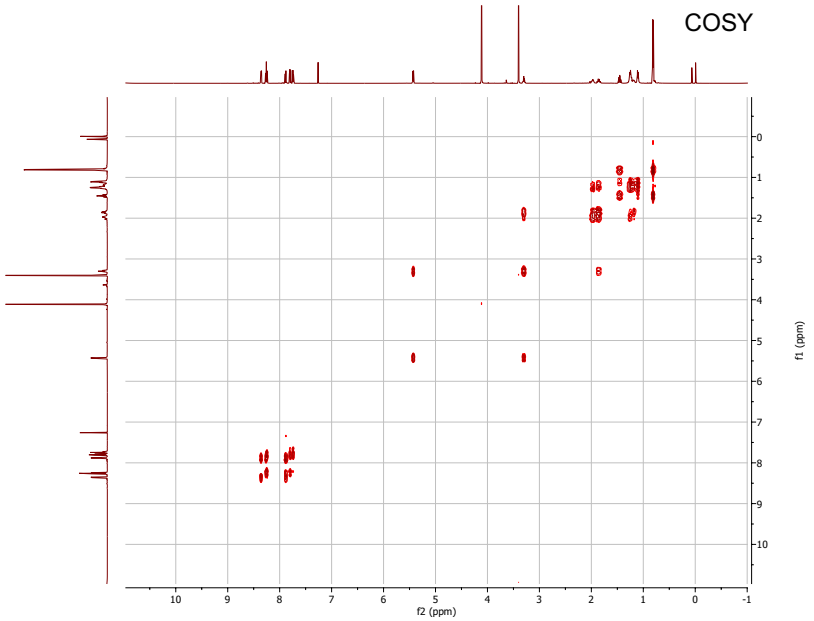
NMR spectroscopic data of compound **23** (diastereomer of streptophenazine F).

Position	δ_C	δ_H, mult. (*J*)	HMBC	COSY
1	131.8			
2	132.35	8.26, m	4, 10a, **1-CO**OCH_3	3
3	129.49	7.88, dd (6.91, 8.75)	1, 4a	2, 4
4	133.15	8.36, dd (1.44, 8.69)	2, 10a	3
4a	140.59			
5a	141.5			
6	138.7			
7	129.63	7.74, dt (1.09, 1.09, 6.91)	1', 5a, 9	8
8	130.53	7.8, dd (6.80, 8.74)	6, 9a	7, 9
9	130.01	8.25, m	7, 5a	8
9a	144.26			
10a	140.96			
1'	75.6	5.43, d (7.76)	2', 3', 5a, 6, 7, **2'-CO**OCH_3	2'
2'	53.41	3.30, ddd (3.69, 7.74, 11.34)	1', 3', **2'-CO**OCH_3	1', 3'
3'	28.5	1.85, 1.95	2', 5'	2', 4'
4'	27.91	1.25	2', 4', 6'	3', 5'
5'	27.65	1.19	3', 4', 6', 7'	4', 6'
6'	38.9	1.11	4', 8'/9'	5', 7'
7'	28.02	1.45, dt (6.62, 6.62, 13.35)	5', 6', 8'/9'	6', 7'/8'
8'	22.62	0.81, dd (2.99, 6.61)	6', 7', 9'	7'
9'	22.62	0.81, dd (2.99, 6.61)	6', 7' ,8'	7'
1-COOCH_3	167.4			
1-COO**CH_3**	52.99	4.11, s	**1-CO**OCH_3	
2'-COOCH_3	175.49			
2'-COO**CH_3**	51.38	3.4, s	**2'-CO**OCH_3	

^{1}H, HSQC, HMBC, and COSY NMR spectra and key HMBC and COSY correlations of compound **23** (diastereomer of streptophenazine F).




HMBC
f2 (ppm)
f1 (ppm)


COSY
f2 (ppm)
f1 (ppm)

2.4 Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in *Cell Chemical Biology*, Bauman, K.D., Li, J., Murata, K., Mantovani, S.M., Dahesh, S., Nizet, V., Luhavaya, H., and Moore, B.S. 2019. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3. Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead

3.1 Introduction and context for Chapter 3

Salinosporamide A is a small molecule natural product produced by the obligate marine actinomycete *Salinispora tropica* and is currently an anti-cancer drug candidate in Phase III clinical trials for the treatment of glioblastoma. The history of this molecule is intimately tied with that of the Scripps Institution of Oceanography. The *Salinispora* genus was isolated by Professors Bill Fenical and Paul Jensen in the early 2000s here at SIO, in the basement of Scholander Hall, and represented the first widespread marine actinomycete ever discovered.[1] Shortly after, a small molecule isolated from *Salinispora* cultures, named salinosporamide A in honor of its producing organism, was found to have potent proteasome inhibitory activity. Just three years later, salinosporamide A entered oncology clinical trials, some of which took place at UCSD's Moores Cancer Center just up the hill from where the molecule was discovered. This powerful bioactivity generated tremendous interest in the genetic underpinnings of salinosporamide. *Salinispora* was one of the first actinomycete genomes to be sequenced, and the first marine actinomycete, an achievement from SIO. Today, salinosporamide A is in Phase III clinical trials for the brain cancer glioblastoma, and biosynthetic studies about the origin of this molecule are still underway in the Moore lab. From discovery to clinical trials, from beach to bedside, salinosporamide A is truly a product of UCSD and SIO, and it was an absolute honor to continue that tradition and history with my own work on this molecule.

My work on salinosporamide biosynthesis focused on the assembly line portion of the pathway. Polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) pathways are often referred to as "assembly lines" because they build molecules through successive, identical reactions performed by highly organized enzymes that each carry

out a distinct step in the process, much like factory workers on an assembly line.[2] While both PKSs and NRPSs assemble molecules in this iterative and organized way, PKSs utilize small acyl units to construct a carbon backbone, while NRPSs use individual amino acids to assemble a polypeptide. In both pathways, enzymatic reactions are performed on a linear intermediate covalently tethered through a 193escry193panthetheine (Ppant) arm to a carrier protein, which serve as molecular shuttles, ferrying intermediates between biosynthetic enzymes. Canonically, PKS biosynthesis is initiated by selection of a coenzyme A-tethered acyl unit by an acyltransferase (AT) domain, or an amino acid by an adenylation (A) domain in NRPS biosynthesis, and transferred to an adjacent carrier protein. Elongation is catalyzed by a ketosynthase (KS) domain that performs a decarboxylative Claisen condensation to generate a carbon-carbon bond, or a condensation (C) domain that forms a peptide bond. Finally, synthesis in both pathways is terminated by a thioesterase domain (TE) that catalyzes release of the mature molecule off the assembly line for further functionalization and tailoring.

The *sal* BGC is dominated by the bimodular PKS SalA and the didomain NRPS SalB, which house the salinosporamide assembly line.[3,4] The first module of the *sal* assembly line loads acetate as the starter unit which initiates salinosporamide biosynthesis. The SalA KS domain then catalyzes diketide formation by condensation with ACP-bound chloroethylmalonate, the first example of a naturally halogenated PKS extender unit, biosynthesized by a cassette of genes found elsewhere in the cluster.[3] the *sal* NRPS selects the nonproteinogenic amino acid cyclohexenylalanine which is suspected to be hydroxylated *in trans* by the cytochrome P450 SalD.[5] Finally, the SalA C domain catalyzes formation of an amide bond and yields the SalB PCP-tethered linear

salinosporamide intermediate. This substrate was proposed to undergo an offloading bicyclization reaction to generate the distinctive salinosporamide bicyclic core. While a tremendous amount of work elucidated the building blocks of the molecule, there was no precedent for the key biosynthetic step(s) to construct the actual pharmacophore itself.

The research described in this chapter was recently published *in Nature Chemical Biology* and describes the discovery, biochemical validation, and structural characterization of SalC as the bicyclase in the salinosporamide biosynthetic pathway, responsible for generating the characteristic and bioactive γ-lactam-β-lactone pharmacophore of the molecule. This project was incredibly collaborative, and spanned generations of Moore lab members. When the salinosporamide biosynthetic gene cluster was first sequenced and revealed in 2007, the original candidate for the cyclizing enzyme was non-obvious. Ultimately, sequential gene disruption experiments of nearly every gene in the *sal* BGC suggested that *salC,* a gene encoding a putative ketosynthase enzyme, may be responsible for this chemistry.

SalC, however, was recalcitrant to heterologous expression in *E. coli*. Accessing the substrate required for the proposed cyclization was also challenging, both from a synthetic perspective and because the system necessitated carrier protein tethered substrate. Finally, working with an unstable β-lactone added an additional challenge to the enzymatic assays. Thus the question regarding the salinosporamide bicyclization step remained unanswered for over a decade. The work described in this chapter is the final answer to this question, and unequivocally establishes SalC as the salinosporamide bicyclase by a series of biochemical experiments, bioactivity-based assays, and structural biology work. This discovery solved a long-standing biosynthetic mystery in the formation

of a clinically relevant compounds, established a novel route to β-lactone formation, and expanded the known biosynthetic reactivities of KS enzymes. Beyond that, understanding the reactivity of SalC grants us access to the most challenging portion of the molecule to install synthetically, opening the door towards future enzyme engineering campaigns and an eventual streamlined chemoenzymatic route to salinosporamide production.

Following my work on the phenazine project, which was entirely in vivo, I was eager to grow my scientific skillset to encompass in vitro biochemical and enzymology experience. This project presented the perfect opportunity to learn those skills and apply them to a biologically and clinically valuable system. Through this project I learned how to express and purify proteins from multiple host systems, I developed enzymatic assays with complex, carrier protein dependent systems and labile products, and I even learned protein crystallography skills and operated the Advanced Light Source beamline myself.

3.2 References for Chapter 3 Introduction

1. Mincer, T. J., Jensen, P. R., Kauffman, C. A. & Fenical, W. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Environ. Microbiol*. **68**, 5005–5011 (2002).

2. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).

3. Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Lechner, A., Kale, A. J., Kobayashi, Y., Reynolds, K. A. & Moore, B. S. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-l-methionine. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12295–12300 (2009).

4. Udwary, D. W., Zeigler, L., Asolkar, R.N., Singan, V. Lapidus, A., Fenical, W., Jensen, P.R., Moore, B.S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica. Proc. Natl. Acad. Sci. U.S.A*. **104**, 10376–10381 (2007).

5. McGlinchey, R. P. Nett, M., Eustáquio, A. S., Asolkar, R. N., Fenical, W. & Moore, B. S. Engineered Biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **130**, 7822 (2008).

3.3 Reprint of “Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead”

nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-022-00993-w

Check for updates

Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead

Katherine D. Bauman[1], Vikram V. Shende[1], Percival Yang-Ting Chen[1,4], Daniela B. B. Trivella[1,5,6], Tobias A. M. Gulder[1,7], Sreekumar Vellalath[2], Daniel Romo[2] and Bradley S. Moore[1,3] ✉

The marine microbial natural product salinosporamide A (marizomib) is a potent proteasome inhibitor currently in clinical trials for the treatment of brain cancer. Salinosporamide A is characterized by a complex and densely functionalized γ-lactam-β-lactone bicyclic warhead, the assembly of which has long remained a biosynthetic mystery. Here, we report an enzymatic route to the salinosporamide core catalyzed by a standalone ketosynthase (KS), SalC. Chemoenzymatic synthesis of carrier protein-tethered substrates, as well as intact proteomics, allowed us to probe the reactivity of SalC and understand its role as an intramolecular aldolase/β-lactone synthase with roles in both transacylation and bond-forming reactions. Additionally, we present the 2.85-Å SalC crystal structure that, combined with site-directed mutagenesis, allowed us to propose a bicyclization reaction mechanism. This work challenges our current understanding of the role of KS enzymes and establishes a basis for future efforts toward streamlined production of a clinically relevant chemotherapeutic.

Salinosporamide A (**1**), also known as marizomib, is a potent 20S proteasome inhibitor originally isolated from the obligate marine actinomycete *Salinispora tropica* and is presently in phase III clinical trials for the treatment of glioblastoma, an aggressive form of brain cancer with a poor prognosis and few therapeutic options (Fig. 1a)[1,2]. Despite the discovery of numerous natural analogs of **1** (ref. [3]), as well as extensive efforts to generate derivatives via chemical synthesis[4] and mutasynthesis[5,6], it is the originally discovered natural product itself that entered clinical trials. Salinosporamide A's compact yet densely functionalized γ-lactam-β-lactone pharmacophore is distinct among proteasome inhibitors, including the Food and Drug Administration (FDA)-approved bortezomib, carfilzomib and ixazomib. The electrophilic β-lactone warhead of **1** serves as the covalent attachment site for the N-terminal catalytic threonine residue found in all three proteasome β-subunits (β1, β2 and β5), resulting in pan-proteasome irreversible inhibition with nanomolar potency (Supplementary Fig. 1)[7]. Unlike other proteasome inhibitors, salinosporamide crosses the blood–brain barrier, leading to its advancement through glioblastoma clinical trials.

Stable isotope feeding[8] and gene inactivation experiments[9] revealed that salinosporamide A is naturally assembled from three distinct metabolic building blocks (Fig. 1a): acetate (orange), cyclohexenylalanine (blue) and chloroethylmalonate (red). These molecules are processed by a hybrid polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) enzymatic assembly line to produce a protein-bound linear intermediate that undergoes cyclization and offloading to yield **1** (Fig. 1a). While the pathway to the unprecedented PKS extender unit chloroethylmalonyl-coenzyme A (chloroethylmalonyl-CoA)[9] and the origins of the unusual non-proteinogenic amino acid cyclohexenylalanine are established[10], the biosynthetic reactions and enzymes responsible for assembly of the γ-lactam-β-lactone pharmacophore are unknown.

Terminal cyclization reactions in microbial biosynthesis are varied and numerous, leading to a diversity of products (Fig. 1b). However, none resemble the suspected bicyclization reaction to **1** that would invoke a two-step intramolecular aldol reaction to install the γ-lactam ring, followed by an offloading β-lactonization reaction. The chemical foundation for this proposed enzymatic reaction was achieved by an elegant biomimetic synthesis of salinosporamide[11,12]. Classically, terminal thioesterase (TE) domains in microbial assembly line megasynth(et)ases are responsible for offloading via macrocyclization[13]. Importantly, the first cyclizing TE capable of β-lactone formation, ObiF, was recently identified in obafluorin (**2**) biosynthesis[14]. While there is a TE encoded within the *sal* BGC, it is hypothesized to be an editing type II TE responsible for removing misprimed molecules from the assembly line[15], and it does not harbor the conserved GXCXG motif from ObiF required for β-lactone synthesis[14]. Microbial biosynthetic pathways have also found alternative routes outside of the TE to catalyze cyclization reactions, including fungal terminal condensation domains (C_T), as seen in fumiquinazoline F (**3**) biosynthesis[16], the hydrolase required for formation of the fused bicyclic ring system in vibralactone (**4**) biosynthesis[17] and the condensation domain responsible for β-lactam formation in the biosynthesis of the antibiotic nocardicin A (**5**)[18]. Recently, a terminal KS domain in the tenuazonic acid (**6**) PKS was shown to catalyze the Dieckmann cyclization of an amino acid β-keto thioester to yield a tetramic acid product[19–21]. While these examples demonstrate a wide variety of non-canonical terminal enzymes that could be involved in a cyclization reaction, none of these previously reported examples resemble a biosynthetic precedent for the proposed bicyclization in salinosporamide A biosynthesis.

Here, we report the identification, biochemical investigation and structural characterization of the KS homolog SalC, which catalyzes the unprecedented tandem aldol-lactonization bicyclization reaction and offloading to assemble the γ-lactam-β-lactone

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA. [2]Department of Chemistry and Biochemistry, Baylor University, Waco, TX, USA. [3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA. [4]Present address: Morphic Therapeutics, Waltham, MA, USA. [5]Present address: Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, Campinas, Brazil. [6]Present address: Institute of Chemistry, University of Campinas, Campinas, Brazil. [7]Present address: Chair of Technical Biochemistry, Technical University of Dresden, Dresden, Germany. ✉e-mail: bsmoore@ucsd.edu

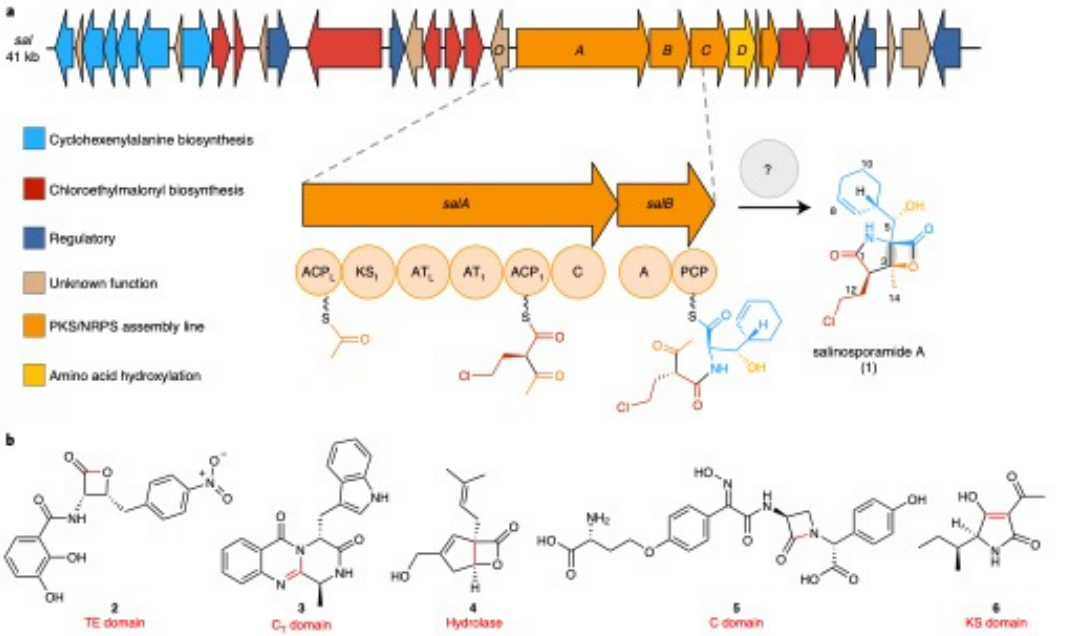


Fig. 1 | Microbial natural products assembled by terminal cyclization reactions. a, The salinosporamide biosynthetic gene cluster (*sal* BGC) and proposed bicyclization of the SalB-peptidyl carrier protein (PCP)-tethered linear intermediate to assemble the salinosporamide A (**1**) pharmacophore. Bonds formed by unknown cyclization enzyme(s) are highlighted in gray. kb = kilobases, ACP = acyl carrier protein, KS = ketosynthase, AT = acyltransferase, C = condensation domain, A = adenylation domain. **b**, Representative additional examples of microbial natural products assembled via terminal cyclization reactions: obafluorin (**2**), fumiquinazoline F (**3**), vibralactone (**4**), nocardicin A (**5**) and tenuazonic acid (**6**). Bonds formed in cyclization reactions and catalyzing enzyme classes are shown in red.

salinosporamide A pharmacophore. This discovery establishes the biosynthetic logic of this brain-penetrant drug candidate and provides a clear roadmap for the generation of new KS-based biocatalysts in medicinal chemistry.

Results

SalC, a KS, is required for salinosporamide biosynthesis. To determine the enzyme responsible for the formation of the bicyclic salinosporamide core, gene inactivation experiments were used to sequentially disrupt candidate genes in the *sal* BGC, as described previously (Fig. 2)[9]. Disruption of *salO*, a putative cyclase, had no effect on salinosporamide production. As expected, disruption of *salA*, encoding the mixed PKS/NRPS assembly line, abolished production of salinosporamides A (**1**) and B (**7**), as did disruption of *salD*, which encodes a P450 enzyme. Instead, the Δ*salD* strain accumulated **8**, the deoxy analog salinosporamide J. A KS, for which no other role in the pathway could be attributed, caught our attention when disruption of its encoding gene, *salC*, completely abolished production of **1** and **7** (Fig. 2). Importantly, *salC* is conserved in all salinosporamide- and cinnabaramide-producing[23] strains (Supplementary Fig. 2). BLAST[23] and NaPDoS[24] analysis of *salC* revealed that its gene product is a standalone KS (Supplementary Figs. 3 and 4). However, SalC lacks the canonical Cys-His-His catalytic triad associated with KSs[25] and instead has an asparagine in place of the first histidine, a variation often associated with 'non-elongating' or 'condensation-incompetent' KSs (Supplementary Fig. 5)[26,27]. Recent work on tenuazonic acid biosynthesis suggests that KSs are capable of catalyzing cyclization reactions, and type III KSs have long been known to catalyze cyclization reactions via aldol condensations on CoA-bound substrates[28], which led to our hypothesis that SalC might function as a bicyclase capable of forming the γ-lactam-β-lactone salinosporamide core.

Characterizing late-stage cyclization and offloading reactions in assembly line biosynthetic pathways presents a unique set of challenges. The substrates required to probe these reactions are fully elongated and functionalized as they prepare for offloading and thus are challenging to access synthetically. To circumvent the need for building these synthetically demanding linear substrates, previous studies of TE domains that perform macrolactonization reactions have capitalized on the reversibility of enzymatic reactions and used cyclized compounds as substrates for the TE to perform the reverse hydrolysis reaction[29]. Therefore, to determine if salinosporamide A is the product of SalC, we chose to first probe the reaction in the reverse direction. To do so, we heterologously prepared SalC in *Streptomyces* (Supplementary Figs. 6 and 8), as all attempts to obtain soluble protein in *Escherichia coli* failed, and incubated the recombinant enzyme with salinosporamide A. In the presence of SalC, we observed a modest decrease in hydrolysis of the β-lactone (compound **9**) compared to the no-enzyme control as well as decreased formation of buffer adduct products in which Tris and glycerol open the β-lactone of **1** (Extended Data Fig. 1). Incubation of **1** with SalC consistently resulted in the formation of unique, non-chlorinated products indicative of the intramolecular formation of a stable tetrahydrofuran (THF) ring (**10**, **11**) by displacement of the chloride, as determined by LC–MS. Notably, this THF ring formation is the same reaction that occurs in the proteasome (Supplementary Fig. 1)[7].

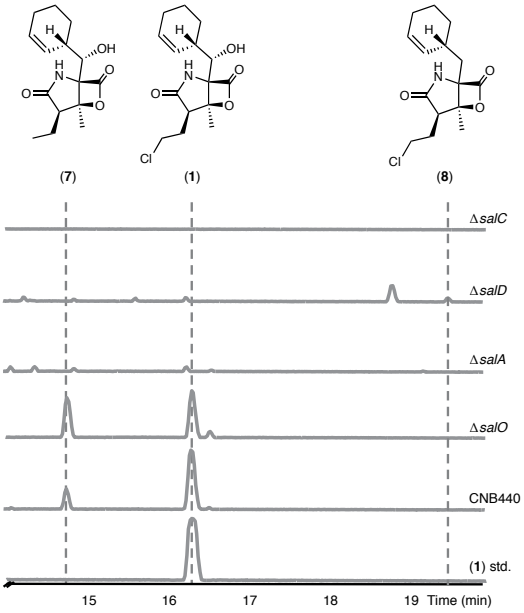


Fig. 2 | SalC is involved in late-stage salinosporamide A biosynthesis. Liquid chromatography–mass spectrometry (LC–MS) chromatograms of salinosporamide A standard (1, std.) and extracts from *Salinispora* cultures, including wild-type *S. tropica* CNB440, *S. tropica* CNB440Δ*salO*, *S. tropica* CNB440Δ*salA*, *S. tropica* CNB440Δ*salD* and *S. tropica* CNB440Δ*salC*; **7**, salinosporamide B; **8**, salinosporamide J.

While **9** could ultimately react to yield **10** and **11**, the presence of SalC clearly increased this reaction rate. The formation of these new products and fewer buffer and hydrolytic cleavage products suggested that, in the presence of SalC, the hydrolysis of **1** occurs in a low-hydration environment, such as within the SalC active site. SalC's ability to hydrolyze the β-lactone of **1** through apparent acylation further implicated SalC in the offloading bicyclization reaction.

SalC is a γ-lactam-β-lactone bicyclase. As our initial experiments suggested a role for SalC in the salinosporamide bicyclization reaction, we sought to develop an assay to directly explore this activity. To do so, we first had to generate an appropriate substrate. Organization of the mixed PKS/NRPS modules in the *sal* pathway indicated that the bimodular PKS (SalA) and the NRPS didomain (SalB) work together to generate the SalB-PCP-tethered β-ketoamide intermediate (Fig. 1a) before the bicyclization reaction. Therefore, we expressed and purified the excised SalB-PCP domain (Supplementary Fig. 7) in its apo form, confirmed by intact proteomics (Supplementary Fig. 8), as well as trypsin digestion and peptide mass fingerprinting (Supplementary Fig. 9), with the goal of modifying the serine residue with a variety of CoA-activated linear precursors.

Salinosporamide A is composed of reactive side chain functional groups that, while essential for the potent bioactivity of this compound, complicate both chemical synthesis and in vitro enzymatic assays. To circumvent possible analytical complications[11,12], we designed a simplified analog of the linear salinosporamide chain elongation intermediate. Once cyclized, this linear precursor would yield 5-deoxy-7,8-dihydro-salinosporamide B, which we refer to as 'simplisporamide'. This substrate was inspired by salinosporamide analogs that have been isolated from the bacterium itself, including the deschloro compound salinosporamide B (**7**), the deoxy version salinosporamide J (**8**), and antiprotealide (**12**), which contains leucine in lieu of the cyclohexenylalanine residue (Extended Data Fig. 2). As all three of these molecules are isolated from *Salinispora* cultures[3,30,31], we hypothesized that a substrate lacking these moieties would be recognized by the putative cyclization enzyme.

Synthesis of the linear pantetheine-activated simplisporamide precursor (**13**) is based on the wealth of literature precedent for the total synthesis of the salinosporamides (Supplementary Note)[11,12,32]. Simplisporamide linear precursors were synthesized as a mixture of diastereomers, and preparative HPLC was used to separate them, as confirmed by NMR. However, following addition to buffer, these pure diastereomers rapidly equilibrated to a ~1:1 mixture (Supplementary Note), which was used for all subsequent experiments and is referred to as **13**.

To generate the PCP-tethered linear substrate, we used a one-pot biocatalytic synthesis beginning with the pantetheine-activated substrate (**13**) and proceeding through the CoA-activated substrate (**14**), as described previously (Fig. 3a, Extended Data Fig. 3 and Supplementary Fig. 10)[33]. Subsequent carrier protein acylation was confirmed by both intact protein LC–MS (Fig. 3b and Supplementary Fig. 11) as well as trypsin digestion of SalB and high-resolution LC–MS analysis of the resulting peptide fragments (Fig. 3c). The MS-based phosphopantetheine (PPant) ejection assay was additionally used to verify the formation of the linear precursor acylated SalB substrate (**15**; Fig. 3c)[34].

Following incubation of SalC (20 μM) with in situ-generated **15**, we observed two products with *m/z* values of 284.19 $[M+H]^+$ at retention times (rts) of 15.2 and 15.6 min (Fig. 3d,e). Coelution with a synthetic standard revealed them to be the hydrolyzed linear β-keto amino acid (**16**). However, the major product of the reaction (*m/z* 266.17 $[M+H]^+$, rt 19.2 min) was indicative of the formation of simplisporamide ((***R***)-**17**; Fig. 3d,e and Supplementary Fig. 12). We noted the appearance of a minor peak with the same molecular ion at 18.9 min ((***S***)-**17**; Fig. 3e). Critically, neither of these compounds were evident in control reactions without SalC. In a time course experiment, we visualized the depletion of these compounds accompanied by formation of a new product (rt 13.7 min) consistent with the mass (*m/z* 284.19 $[M+H]^+$) of the hydrolyzed β-lactone (**18**; Fig. 3e) due to degradative hydrolysis over time, which was also observed with salinosporamide A (Extended Data Fig. 1).

After demonstrating the formation of an enzymatic reaction product indicative of the bicyclic product ((***R***)-**17**), we sought to perform the SalC assay at a preparative scale suitable for purification of the product and NMR characterization. However, the reliance on a single turnover reaction with respect to a carrier protein presented a substantial challenge to scaling up the SalC bicyclization reaction, and, as such, we sought to use diffusible substrates to circumvent the need for stoichiometric SalB-PCP. Given the literature precedent for KSs using CoA mimics as activating groups[35], we tested SalC activity with pantetheine- and CoA-activated linear substrates as well as the products of the CoaA and CoaD reactions en route to the CoA-activated substrate (**19** and **20**, respectively). When SalC was incubated with these diffusible substrates, minimal putative bicyclized product was formed (Extended Data Fig. 3), and the primary product of the reaction appeared to be **16**. Therefore, we proceeded with carrier protein-dependent reactions to scale up the SalC activity assay.

The inherent instability of the simplisporamide β-lactone presented an additional challenge for a larger scale reaction in an aqueous environment[36]. During fermentative production of **1**, degradation via hydrolysis of the β-lactone is mitigated by the addition of the absorbent resin XAD7 to the culture medium, which stabilizes and captures **1** for subsequent purification[37]. Inspired

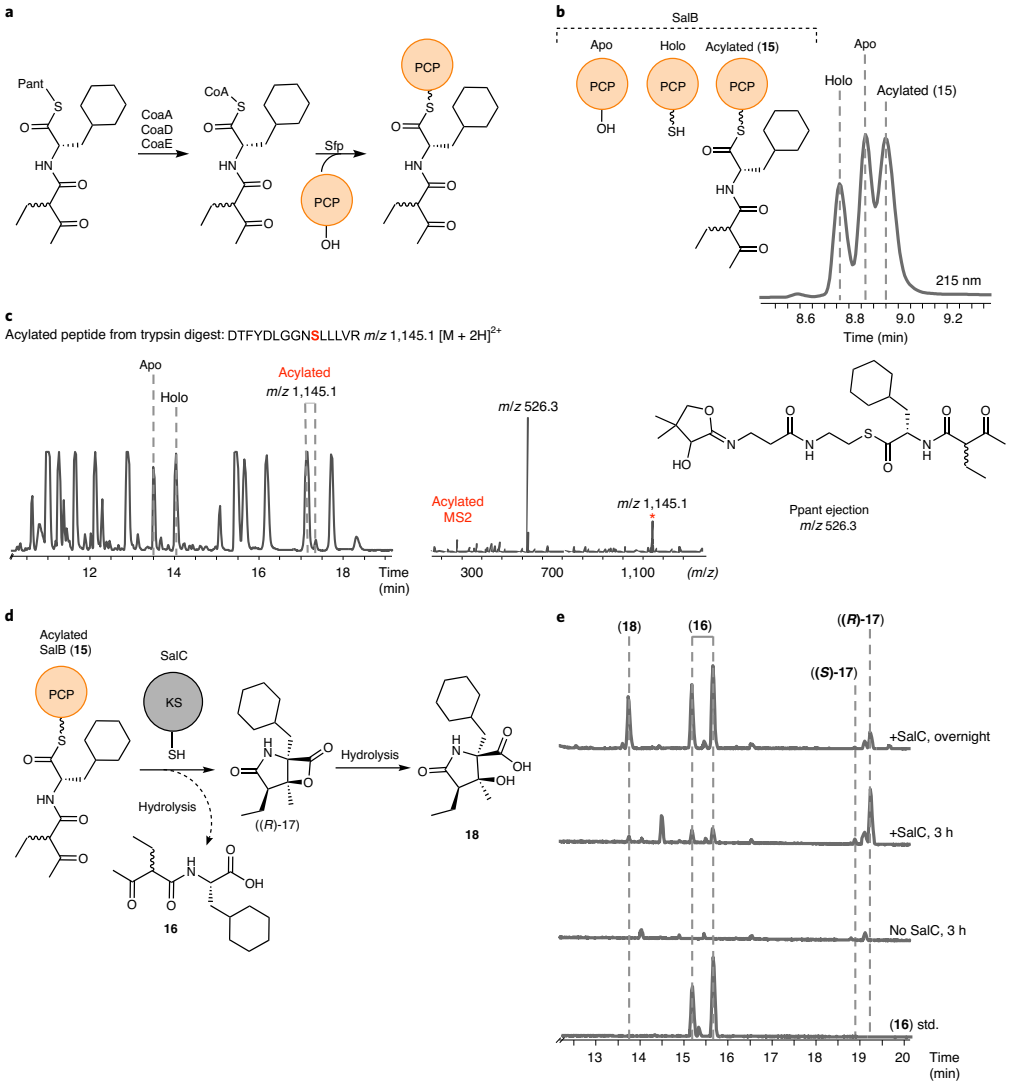


Fig. 3 | SalC is a γ-lactam-β-lactone bicyclase. a, Reaction scheme for the chemoenzymatic synthesis of PCP-tethered linear substrate (**15**) from pantetheine-activated simplisporamide precursor (**13**) via in vitro CoA biosynthesis and Sfp-mediated loading. **b**, Intact protein LC–MS chromatogram of SalB-PCP acylation assay (UV, 215 nm). See Supplementary Fig. 11 for intact proteomics high-resolution MS (HRMS) data. **c**, LC–MS chromatogram of SalB-PCP tryptic digest post acylation assay and MS2 Ppant ejection assay. **d**, Reaction scheme for the SalC activity assay. **e**, LC–MS chromatograms of linear hydrolysis product (**16**) standard, the SalC activity assay with no SalC (3 h) and the SalC activity assay for 3 h and overnight. See Supplementary Fig. 12 for HRMS data for **(*R*)-17** and Supplementary Note for full characterization.

by this strategy, we included an absorbent resin in a large-scale SalC assay to protect the cyclized product as it is offloaded from the protein. Without resin, the product degraded rapidly to yield **18**. The major product ((***R***)-**17**) was captured by the resin (Supplementary Fig. 13), purified and confirmed by NMR (for complete tables of NMR assignments, spectra and key correlations, see Supplementary Note).

The strained β-lactone of **1** is the key pharmacophore of the molecule, serving as the electrophilic trap for proteasome inhibition, and so we aimed to directly link SalC to the formation of this molecular warhead and therefore the bioactivity of the molecule. We performed the SalC activity assay with resin to trap the simplisporamide product and control assays without SalC or **13**. The extracts of these resins (Supplementary Fig. 14) were directly tested

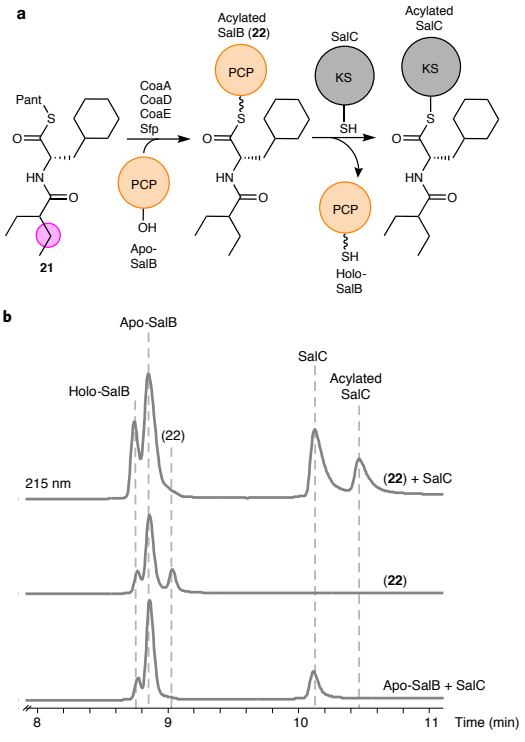


Fig. 4 | Transacylation of SalC by acylated SalB. a, Reaction scheme for the chemoenzymatic synthesis of **22** and subsequent transacylation assay with SalC. **b**, UV chromatograms (215 nm) of intact protein LC–MS of apo-SalB-PCP and SalC (bottom trace; small amount of holo-SalB-PCP present likely due to native *E. coli* phosphopantetheinyl transferase modification), SalB-PCP acylated with the mechanistic probe (**21**) and purified by size-exclusion chromatography to yield **22** (middle trace) and the complete transacylation assay using **22** and SalC (top trace). See Supplementary Figs. 16 and 17 for intact proteomics HRMS data.

for in vitro inhibitory activity against the human 20S proteasome chymotrypsin-like protease activity (β5-subunit; Extended Data Fig. 4). Importantly, this fluorescence-based assay is robust and commercially available and can be performed in a 96-well plate format, which makes it amenable for high-throughput work. The bioassay showed that SalC was required for proteasome inhibition and that the proteasome was not inhibited when SalC or substrate was excluded from the assay. This result further confirmed that SalC is responsible for the formation of the bioactive β-lactone moiety.

SalC is acylated by SalB-PCP. The proposed bicyclization reaction performed by SalC is intramolecular, in contrast to the intermolecular nature of canonical KS biosynthesis. As such, SalC-mediated cyclization could occur on the carrier protein, or the substrate could be transferred onto SalC itself for cyclization to occur (Supplementary Fig. 15a). To distinguish between these two possibilities, we synthesized a mechanistic probe lacking the β-ketone group (**21**) that would be capable of acylating SalB and SalC but would be incompetent for cyclization and offloading. Therefore, this substrate would remain trapped on the protein, potentially allowing for detection via intact protein MS (Fig. 4a). Using **21**, we performed the previously described chemoenzymatic acylation assay with apo-SalB-PCP to yield holo acylated SalB-PCP (**22**). Compound **22** was then purified away from residual CoA proteins and small-molecule intermediates to ensure that all transacylation originated from the carrier protein. Column-purified **22** was incubated with SalC (m/z 66,613 $[M+H]^+$) and subjected to intact protein LC–HRMS analysis, revealing a new peak (m/z 66,864 $[M+H]^+$; Fig. 4b) indicative of chain transfer of the mechanistic probe from the SalB-PCP domain to SalC (Supplementary Figs. 16 and 17). We also performed this transacylation assay with diffusible substrates: the pantetheine-activated linear probe (**21**), the products of the CoaA and CoaD reactions with **21** (**23** and **24**, respectively) and the CoA-activated linear probe (**25**; Extended Data Fig. 5 and Supplementary Figs. 18–23). While SalC was capable of being acylated by diffusible substrates, the largest amount of acylated SalC was observed with carrier protein-activated substrate. Finally, small-molecule LC–MS revealed that once acylated with the mechanistic probe, SalC was able to hydrolyze off **21** over time as well (Supplementary Fig. 15b).

Structure-guided mutagenesis investigated bicyclization. To elucidate the molecular basis for the observed bicyclase activity of this unusual KS, we solved a crystal structure of SalC to 2.85-Å resolution (Protein Data Bank (PDB): 7S2X; Supplementary Tables 3 and 4). Crystallography revealed that the oligomeric state of SalC is a homotetramer (Fig. 5a), which was verified via size-exclusion chromatography (Supplementary Fig. 24)[38]. Overall, SalC possesses high structural homology[39] to *trans*-AT KSs (Extended Data Fig. 6), although they share less than 40% sequence identity. Each monomeric subunit of SalC is composed of two domains, the N-terminal KS domain, which forms tetrameric interfaces, and the C-terminal flanking domain, which does not contact domains from other subunits. The KS domain of SalC resembles a canonical KS with the classic αβαβα thiolase fold[40] (Extended Data Fig. 7); whereas the C-terminal flanking subdomain resembles those seen in *trans*-AT KSs (Extended Data Fig. 6), a ~100-residue structure homologous to the KS-AT adapters in *cis*-AT PKS systems but with an unknown function[41,42].

The active site of SalC lacks the established Cys-His-His catalytic triad found in canonical KS domains[25] and instead contains a Cys180-Asn316-His353 triad similar to non-elongating KSs (Supplementary Figs. 5 and 25). The SalC active site (Fig. 5b) also contains a conserved lysine (Lys348) found in both PKS KSs and the fatty acid synthase β-keto-acyl-ACP synthases. Examination of the active site also revealed the presence of a tyrosine residue (Tyr284) located 3.6 Å from Cys180 and positioned on a loop that protrudes into the active site. Tyr284 is conserved in all SalC homologs from salinosporamide-producing *Salinispora* strains and the cinnabaramide SalC homolog CinC (Extended Data Figs. 7 and 8). Superimposition of SalC with a KS from the bacillaene PKS pathway (bae; Extended Data Fig. 9) bound to its native chain elongation intermediate (PDB ID: 4NA2)[42] revealed that the SalC Tyr284 residue is oriented toward the thioester α-carbonyl of the bae intermediate, suggesting that it may play a key role in salinosporamide bicyclization. To investigate the role these residues play in SalC bicyclization, we performed site-directed mutagenesis experiments of active site residues implicated in catalysis (Supplementary Fig. 26). Resulting SalC variants were subjected to SalB-to-SalC transacylation assays with **22** (Fig. 5c and Supplementary Figs. 27–30) and SalC cyclization activity assays with **15** (Fig. 5d).

As SalC is acylated by the SalB-PCP domain to form a linear acyl-enzyme intermediate before cyclization, it stands to reason that mutating the active site Cys180Ala both abrogated chain transfer and abolished production of the cyclized product (Fig. 5c,d and Supplementary Fig. 27). Furthermore, the Cys180Ser mutation had

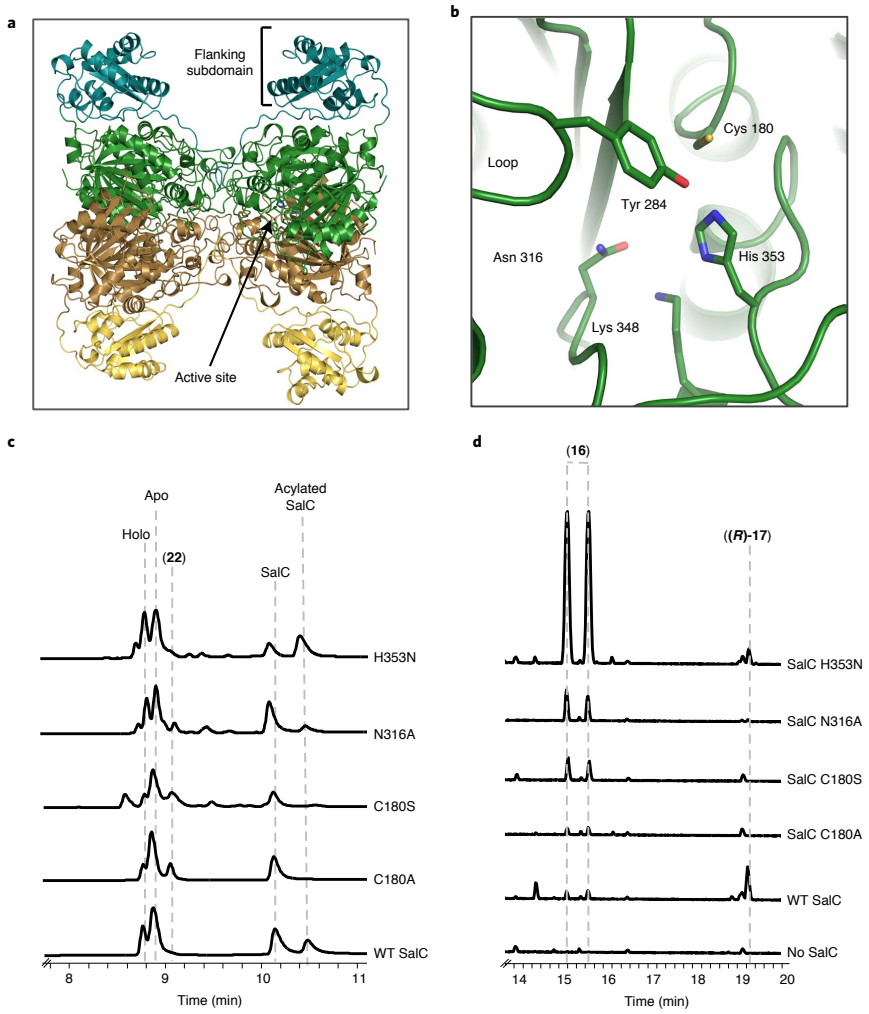


Fig. 5 | SalC structure and active site mutagenesis. a, Overall SalC tetrameric structure with each monomeric domain colored differently (KS domains in green and brown and flanking subdomain in teal and yellow). **b**, SalC active site with proposed catalytic residues labeled. **c**, UV (215 nm) chromatograms from intact protein LC–MS transacylation assays using column-purified **22** and SalC mutants. **d**, LC–MS chromatograms of SalC activity assays (3 h) with wild-type (WT) SalC and SalC active site variants using **15**. See Supplementary Figs. 27–30 for intact proteomics HRMS data.

the same effect (Supplementary Fig. 28)[43]. The Asn316Ala mutant did not retain full transacylation activity in the chain transfer assay and showed reduced production of the bicyclized product to less than 4% of wild-type SalC (Fig. 5c,d and Supplementary Fig. 29). The His353Asn mutation was selected, as type III KSs, known to perform cyclization reactions, typically have an asparagine in place of the final histidine in the catalytic triad. In the case of SalC, this mutant retained roughly 40% of the wild-type SalC cyclization activity, but, notably, it largely appeared to convert SalC from a bicyclase to a hydrolase (Fig. 5c and Supplementary Fig. 30). This mutant produced 43 times the amount of **16** compared to wild-type SalC, which may be due to a potential role of His 353 in deprotonating the thioester α-proton. Without this deprotonation, the intramolecular aldol reaction may not occur, and the stuck linear intermediate is hydrolyzed off SalC as **16**. The Tyr284Phe mutation, unfortunately, resulted in unstable protein and was thus unable to be assessed.

Based on the structure of SalC and results from our mutagenesis, we propose a possible mechanism (Fig. 6 and Extended Data Fig. 10) initiated by deprotonation of Cys 180 by His 353, followed by acylation of the resulting thiolate with the SalB-PCP-tethered substrate (**15**)[25]. Lys 348 may serve as a general base capable of deprotonating His 353, although there are other basic residues near the active site that could perform this deprotonation. Asn 316 may play a stabilizing role in this reaction, as this residue appears critical for transacylation. Hydrogen bonding of the thioamide carbonyl to the Tyr 284 phenol facilitates deprotonation of the thioester α-carbon

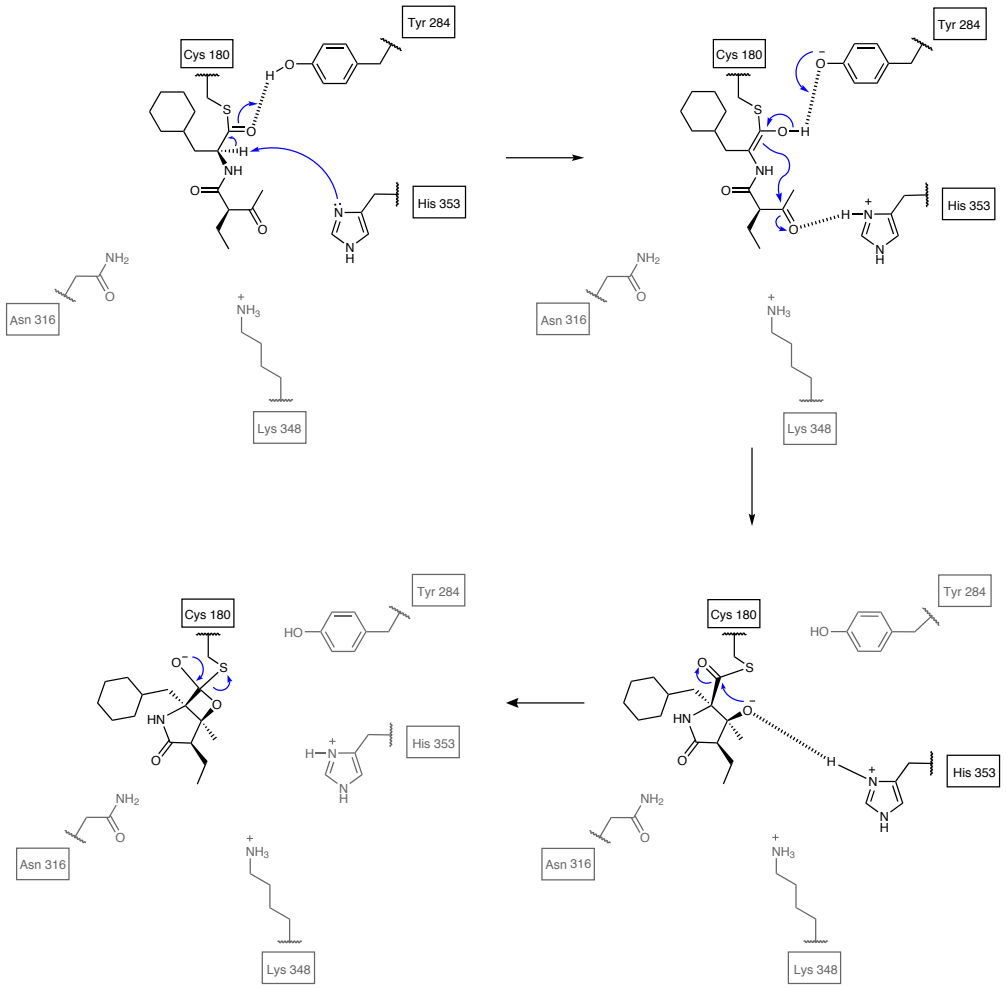


Fig. 6 | Abbreviated key bicyclization steps of the proposed SalC mechanism with 15. Once substrate undergoes transacylation from acylated SalB (**15**) to Cys180 of SalC (not shown), hydrogen bonding by Tyr284 facilitates deprotonation of the thioester α-proton by His353. An intramolecular aldol reaction forms the γ-lactam, and the resulting oxyanion is presumably stabilized by dipole interactions with backbone amides (not shown)[25]. Subsequent β-lactonization through a tetrahedral intermediate leads to release of simplisporamide (**(*R*)-17**) from SalC. See Extended Data Fig. 10 for further details.

by His353, thereby generating an enol intermediate that could react via an intramolecular aldol reaction to form the γ-lactam. The resulting oxyanion is presumably stabilized by dipole interactions with backbone amides, as is hypothesized for KSs[25]. Subsequent β-lactonization through a tetrahedral intermediate releases the bicyclic simplisporamide product from SalC. Finally, Cys180 is reprotonated by His353 to restart the catalytic cycle, as previously proposed for KSs from type I PKS systems[25].

Discussion

The proteasome is a well-established target for cancer therapeutics due to its essential role in intracellular protein degradation, a process that is upregulated in rapidly proliferating cancer cells[44]. Three constitutive proteasome inhibitors are currently FDA approved as cancer therapeutics. Marizomib, however, is fundamentally different from these approved proteasome inhibitors, as it is the only non-peptidic compound in advanced clinical trials and targets all three catalytic subunits of the human 20S proteasome with single-digit nanomolar inhibitory activity against the β5-subunit[7]. This pan-proteasome inhibitory activity allows **1** to retain clinical activity in individuals resistant to other proteasome inhibitors, as resistance is typically due to mutations in a single proteasome subunit[45]. Resistance is a major obstacle toward using proteasome inhibitors in the clinic, which underscores the importance of the continued development of **1** and its analogs. Additionally, salinosporamide A's ability to cross the blood–brain barrier facilitates a new route to treat neurological diseases, which often require invasive surgery for drug delivery to the brain. While salinosporamide

A is still in phase III clinical trials, the potent activity of **1** and its ability to enter the brain could provide hope for individuals with glioblastoma.

The work described here establishes a KS homolog, SalC, acts as a bicyclase and installs the salinosporamide pharmacophore responsible for the clinically relevant bioactivity of **1**. This discovery not only solves a long-standing biosynthetic enigma but also expands the repertoire of known reactivities for the KS family of enzymes. Rather than a decarboxylative Claisen condensation to extend a nascent polyketide chain, the KS SalC plays an unexpected dual role in the *sal* pathway as an intramolecular aldolase/β-lactone synthase, a reaction for which there is no biosynthetic precedent. Furthermore, by using a proteasome inhibition assay to screen for formation of this reactive feature, we were able to directly link the biosynthetic activity of this enzyme to the bioactivity of the molecule.

This ability to use a fluorescence-based assay to validate SalC's biosynthetic ability is especially intriguing as it opens the door for high-throughput screening efforts to generate SalC variants with desirable qualities for biocatalysis. The unusual reactivity of SalC has the potential to be harnessed for chemoenzymatic routes to β-lactone synthesis and complements existing syntheses of **1** that perform bicyclization as the terminal step[12]. Several features of SalC make the enzyme well suited for this.

First, SalC is notable for its ability to perform a complex biosynthetic reaction on a variety of substrates. Typically, the scaffold of assembly line-based natural products is constructed by a PKS and/or NRPS and then functionalized after release from the terminal module via tailoring reactions, such as oxidations or glycosylations, which are essential for the bioactivity of these compounds. Instead, the sal pathway expends metabolic energy upfront to generate complex precursors that are then assembled by SalC, a standalone enzyme that zips together complex pieces in a single step. While TEs are known to be highly selective for their native substrate[46], SalC is able to perform a complex cyclization reaction on a variety of precursors to form the fused bicyclic core that is the constant feature amid the wealth of salinosporamide biosynthetic diversity.

Beyond its flexible substrate preference, SalC also appears capable of resolving a diastereomeric substrate mixture, with a strong preference for forming simplisporamide with the *R* stereocenter at the ethyl side chain. The minor epimer is produced as well (**(*S*)-17**) but in an approximately 1:10 ratio. This preference is borne out in the native bacterium as well; naturally produced salinosporamide F is the C2 epimer of salinosporamide A but is produced in very low titer compared to **1** (ref. [30]). Furthermore, this work also lays the foundation to engineer or design SalC biocatalysts capable of working on small-molecule diffusible substrates (Extended Data Fig. 3).

Finally, SalC is potentially useful as a biocatalyst because it is capable of installing a β-lactone, a chemical moiety known for its reactivity. While the assembly of β-lactam natural products has been studied for decades[18], the study of β-lactone biosynthesis has largely been hindered by the instability of these molecules. Only in the last 5 years have routes to β-lactone biosynthesis been discovered, with just three examples before this work. Furthermore, SalC catalyzes an entirely new reaction type from the previously characterized enzymes, which include adenylate-forming β-lactone synthetases[47,48], thioesterases[14] and α/β-hydrolases[17]. SalC is also different in that it must first perform an aldol reaction to generate the moieties required for β-lactonization; the nucleophilic alcohol is not already installed on the molecule as in the obafluorin and olefinic hydrocarbon pathways. The discovery of SalC as a standalone KS capable of forming a β-lactone suggests the possibility of this type of reactivity in other pathways, particularly where terminal standalone C or KS domains are evident. In fact, NaPDoS analysis reveals that SalC and its homologs, including the cinnabaramide homolog CinC, belong to their own unique clade among known KSs (Supplementary Fig. 4), demonstrating the unique functionality of this enzyme and the potential for discovery of other bicyclized natural products.

Almost 18 years after its initial discovery, and four years after entering phase III clinical trials, a key mystery in the biosynthesis of salinosporamide A has finally been revealed. Using biochemical and structural data, we identified and characterized an unusual KS responsible for performing previously unprecedented biochemistry. The salinosporamide bicycle is the most synthetically challenging portion of the molecule, and, as such, salinosporamide derivatives are prepared by slow and low-yielding fermentative approaches. To this day, salinosporamide A is produced by fermentation for clinical trials, a process that requires the use of specialized reactors to handle the saltwater media necessary for *Salinispora* spp. growth. However, the identification of SalC as a bicyclase capable of generating the γ-lactam-β-lactone pharmacophore completes the minimum genetic architecture required for assembly of the salinosporamide scaffold and has the potential to fundamentally alter how salinosporamides are produced. As our clinical understanding of alternative proteasomes grows[44] and we begin to understand the role these cellular machines play not just in cancer but in autoimmune and parasitic diseases, a hybrid synthetic–biocatalytic blueprint for producing brain-penetrant salinosporamide molecules is now feasible. Understanding the unique reactivity of the SalC bicyclase is the first step toward harnessing that activity for the efficient, straightforward production of new-to-nature bioactive salinosporamides.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-022-00993-w.

Received: 29 August 2021; Accepted: 4 February 2022;
Published online: 21 March 2022

References

1. Roth, P. et al. A phase III trial of marizomib in combination with temozolomide-based radiochemotherapy versus temozolomide-based radiochemotherapy alone in patients with newly diagnosed glioblastoma. *J. Clin. Oncol.* **39**, 2004 (2021).
2. Feling, R. H. et al. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed. Engl.* **42**, 355–357 (2003).
3. Williams, P. G. et al. New cytotoxic salinosporamides from the marine Actinomycete *Salinispora tropica*. *J. Org. Chem.* **70**, 6196–6203 (2005).
4. Macherla, V. R. et al. Structure–activity relationship studies of salinosporamide A (NPI-0052), a novel marine derived proteasome inhibitor. *J. Med. Chem.* **48**, 3684–3687 (2005).
5. McGlinchey, R. P. et al. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **130**, 7822–7823 (2008).
6. Nett, M., Gulder, T. A. M., Kale, A. J., Hughes, C. C. & Moore, B. S. Function-oriented biosynthesis of β-lactone proteasome inhibitors in *Salinispora tropica*. *J. Med. Chem.* **52**, 6163–6167 (2009).
7. Groll, M., Huber, R. & Potts, B. C. M. Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of β-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136–5141 (2006).
8. Beer, L. L. & Moore, B. S. Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica*. *Org. Lett.* **9**, 845–848 (2007).
9. Eustáquio, A. S. et al. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from *S*-adenosyl-L-methionine. *Proc. Natl Acad. Sci. USA* **106**, 12295–12300 (2009).
10. Mahlstedt, S., Fielding, E. N., Moore, B. S. & Walsh, C. T. Prephenate decarboxylases: a new prephenate-utilizing enzyme family that performs non-aromatizing decarboxylation en route to diverse secondary metabolites. *Biochemistry* **49**, 9021–9023 (2010).

11. Ma, G., Nguyen, H. & Romo, D. Concise total synthesis of (±)-salinosporamide A, (±)-cinnabaramide A, and derivatives via a bis-cyclization process: implications for a biosynthetic pathway? *Org. Lett.* **9**, 2143–2146 (2007).
12. Nguyen, H., Ma, G., Gladysheva, T., Fremgen, T. & Romo, D. Bioinspired total synthesis and human proteasome inhibitory activity of (–)-salinosporamide A, (–)-homosalinosporamide A, and derivatives obtained via organonucleophile promoted bis-cyclizations. *J. Org. Chem.* **76**, 2–12 (2011).
13. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
14. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).
15. Gulder, T. A. M. & Moore, B. S. Salinosporamide natural products: potent 20S proteasome inhibitors as promising cancer chemotherapeutics. *Angew. Chem. Int. Ed. Engl.* **49**, 9346–9367 (2010).
16. Gao, X. et al. Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.* **8**, 823–830 (2012).
17. Feng, K.-N. et al. A hydrolase-catalyzed cyclization forms the fused bicyclic β-lactone in vibralactone. *Angew. Chem. Int. Ed. Engl.* **59**, 7209–7213 (2020).
18. Gaudelli, N. M., Long, D. H. & Townsend, C. A. β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520**, 383–387 (2015).
19. Yun, C.-S. et al. Unique features of the ketosynthase domain in a non-ribosomal peptide synthetase–polyketide synthase hybrid enzyme, tenuazonic acid synthetase 1. *J. Biol. Chem.* **295**, 11602–11612 (2020).
20. Yun, C.-S., Motoyama, T. & Osada, H. Biosynthesis of the mycotoxin tenuazonic acid by a fungal NRPS–PKS hybrid enzyme. *Nat. Commun.* **6**, 8758 (2015).
21. Mo, X. & Gulder, T. A. M. Biosynthetic strategies for tetramic acid formation. *Nat. Prod. Rep.* **38**, 1555–1566 (2021).
22. Rachid, S. et al. Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *ChemBioChem* **12**, 922–931 (2011).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Ziemert, N. et al. The Natural Product Domain Seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **7**, e34064 (2012).
25. Robbins, T., Kapilivsky, J., Cane, D. E. & Khosla, C. Roles of conserved active site residues in the ketosynthase domain of an assembly line polyketide synthase. *Biochemistry* **55**, 4476–4484 (2016).
26. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
27. Masschelein, J. et al. A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. *Nat. Chem.* **11**, 906–912 (2019).
28. Satou, R. et al. Structural basis for cyclization specificity of two *Azotobacter* type III polyketide synthases. *J. Biol. Chem.* **288**, 34146–34157 (2013).
29. Boddy, C. N., Schneider, T. L., Hotta, K., Walsh, C. T. & Khosla, C. Epothilone C macrolactonization and hydrolysis are catalyzed by the isolated thioesterase domain of epothilone polyketide synthase. *J. Am. Chem. Soc.* **125**, 3428–3429 (2003).
30. Reed, K. A. et al. Salinosporamides D–J from the marine actinomycete *Salinispora tropica*, bromosalinosporamide, and thioester derivatives are potent inhibitors of the 20S proteasome. *J. Nat. Prod.* **70**, 269–276 (2007).
31. Manam, R. R. et al. Antiprotealide is a natural product. *J. Nat. Prod.* **72**, 295–297 (2009).
32. Ling, T., Macherla, V. R., Manam, R. R., McArthur, K. A. & Potts, B. C. M. Enantioselective total synthesis of (–)-salinosporamide A (NPI-0052). *Org. Lett.* **9**, 2289–2292 (2007).
33. Agarwal, V. et al. Chemoenzymatic synthesis of acyl coenzyme A substrates enables in situ labeling of small molecules and proteins. *Org. Lett.* **17**, 4452–4455 (2015).
34. Dorrestein, P. C. et al. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45**, 12756–12766 (2006).
35. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A. & Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **76**, 195–221 (2007).
36. Denora, N., Potts, B. C. M. & Stella, V. J. A mechanistic and kinetic study of the β-lactone hydrolysis of salinosporamide A (NPI-0052), a novel proteasome inhibitor. *J. Pharm. Sci.* **96**, 2037–2047 (2007).
37. Tsueng, G. & Lam, K. S. Stabilization effect of resin on the production of potent proteasome inhibitor NPI-0052 during submerged fermentation of *Salinispora tropica*. *J. Antibiot.* **60**, 469–472 (2007).
38. Dutta, S. et al. Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
39. Holm, L. Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
40. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931–943 (2007).
41. Lohman, J. R. et al. Structural and evolutionary relationships of 'AT-less' type I polyketide synthase ketosynthases. *Proc. Natl Acad. Sci. USA* **112**, 12693–12698 (2015).
42. Gay, D. C. et al. A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. *Structure* **22**, 444–451 (2014).
43. Davies, C., Heath, R. J., White, S. W. & Rock, C. O. The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* **8**, 185–195 (2000).
44. Cromm, P. M. & Crews, C. M. The proteasome in modern drug discovery: second life of a highly valuable drug target. *ACS Cent. Sci.* **3**, 830–838 (2017).
45. Niewerth, D. et al. Antileukemic activity and mechanism of drug resistance to the marine *Salinispora tropica* proteasome inhibitor salinosporamide A (marizomib). *Mol. Pharmacol.* **86**, 12–19 (2014).
46. Hansen, D. A., Koch, A. A. & Sherman, D. H. Identification of a thioesterase bottleneck in the pikromycin pathway through full-module processing of unnatural pentaketides. *J. Am. Chem. Soc.* **139**, 13450–13455 (2017).
47. Christenson, J. K. et al. β-Lactone synthetase found in the olefin biosynthesis pathway. *Biochemistry* **56**, 348–351 (2017).
48. Robinson, S. L. et al. Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. *J. Biol. Chem.* **295**, 14826–14839 (2020).

Methods

Software and tools. The SalC protein sequence was analyzed via BLAST[23], and all alignments were made using Clustal Omega (1.2.4) software. Geneious Prime 2019.2.3 software was used for plasmid maps and primer design. The NaPDoS platform[24] was used to examine SalC in relation to other KSs, cblaster (1.2.9)[49] was used to examine SalC homologs in salinosporamide-producing strains, and clinker (v0.0.21) was used for visualization[50].

Bacterial strains and growth conditions. *E. coli* strains were grown in LB broth or agar at 37 °C with appropriate antibiotics for selection (apramycin 50 μg ml^{-1}, chloramphenicol 25 μg ml^{-1}, kanamycin 50 μg ml^{-1} and nalidixic acid 25 μg ml^{-1}). For conjugation purposes, *E. coli* was grown using 2TY medium (1.6% (wt/vol) tryptone, 1% (wt/vol) yeast extract and 0.5% (wt/vol) NaCl) with appropriate antibiotic selection. For protein production, TB medium (1.2% (wt/vol) tryptone, 2.4% (wt/vol) yeast extract, 0.4% (vol/vol) glycerol, 2.31% (wt/vol) KH_2PO_4 and 12.54% (wt/vol) K_2HPO) was used. All liquid cultures were shaken at 200 r.p.m.

Streptomyces strains were grown on SFM agar plates (2% (wt/vol) D-mannitol, 2% (wt/vol) soya flour and 2% (wt/vol) agar) for conjugation and strain maintenance at 30 °C or in TSBY liquid medium (3% (wt/vol) tryptic soy broth, 10.3% (wt/vol) sucrose and 0.5% (wt/vol) yeast extract) at 30 °C with shaking at 220 r.p.m.

For protein production in *Streptomyces coelicolor* CH999, a TSBY preculture (100 μl) was used to inoculate the 30-ml 'primary culture' of Super YEME medium (0.3% (wt/vol) yeast extract, 0.5% (wt/vol) peptone, 1% (wt/vol) glucose, 0.3% (wt/vol) malt extract, 34% (wt/vol) sucrose, 0.5% (wt/vol) glycine, 0.235% (vol/vol) $MgCl_2 \cdot 6H_2O$ (2.5 M), 7.5×10^{-3}% (wt/vol) L-proline, 7.5×10^{-3}% (wt/vol) L-arginine, 7.5×10^{-3}% (wt/vol) L-cysteine, 0.01% L-histidine and 1.5×10^{-3}% (wt/vol) uracil, pH 7.2) with 50 mg ml^{-1} apramycin in a 250-ml flask with bottom spring. This culture was grown at 30 °C with shaking at 220 r.p.m. After 4 d, 4 ml of primary culture was used to inoculate 40 ml of Super YEME supplemented with apramycin to generate the 'secondary culture'. This culture was grown at 30 °C for 2 d. Finally, 600 ml of 'expression culture' was inoculated using 30 ml of secondary culture supplemented with apramycin (50 mg ml^{-1}) in 2.8-liter flasks with metal springs. After 2 d, thiostrepton was added to a final concentration of 10 μg ml^{-1} to induce protein expression.

Salinispora strains were grown in sea water-based A1 medium (1% (wt/vol) soluble starch, 0.4% (wt/vol) yeast extract, 0.2% (wt/vol) peptone and 0.1% (wt/vol) $CaCO_3$) at 30 °C with shaking at 220 r.p.m. or A1 agar at 30 °C supplemented with apramycin (50 mg ml^{-1}) and nalidixic acid (25 mg ml^{-1}) for mutant strains. For production of salinosporamide, liquid A1 was supplemented with 2% (vol/vol) KBr (20 g liter^{-1}) and 0.8% (vol/vol) $Fe_2(SO_4)_3$ (8 g liter^{-1}), and XAD7 resin was added to the cultures after 24 h of growth.

DNA isolation and manipulation. Genomic DNA (gDNA) was isolated following standard procedures from *Practical Streptomyces Genetics*[51]. Phusion High-Fidelity DNA Polymerase (NEB) with GC buffer was used to amplify genes from the *sal* cluster using gDNA as a template.

PCR reactions were performed in a BioRad MyCycler with gradient option. PCR products were purified using a Qiagen QIAquick PCR and Gel Cleanup kit according to the manufacturer's instructors. Vectors (pET28, pET28-MBP and pCJW93) were linearized with restriction enzymes for cloning purposes. For cloning, Gibson assembly (NEB HiFi DNA Assembly Master Mix) was used to combine linearized vector and purified PCR products before subsequent transformation into chemically competent *E. coli* DH10B cells. Plasmid DNA was isolated using a QIAPrep Spin Miniprep kit (Qiagen), and cloning was verified by Sanger sequencing (Genewiz).

For protein expression and production of SalC in *Streptomyces*, pCJW93-SalC was transformed into *E. coli* ET12567 (ref. [52]) and conjugated into *S. coelicolor* CH999 via triparental intergeneric mating facilitated by *E. coli* ET12567/pUB307 (ref. [53]) following established procedures[54]. Exconjugates were grown on SFM medium containing 10 mM $MgCl_2$ for exactly 18 h, at which time plates were overlaid with 0.5 mg ml^{-1} naldixic acid followed by 1 mg ml^{-1} apramycin. Plates were incubated at 30 °C until the appearance of exconjugants, which were subsequently replated for further rounds of selection.

S. tropica CNB440 mutants were made by inactivating each gene by replacement with an apramycin resistance cassette (*aac(3)IV*) using λ-Red recombination, as described previously[9].

Salinosporamide identification and purification. Wild-type and knockout *S. tropica* CNB440 strains were grown for 5 d after addition of XAD7 absorbent resin. Resin was collected and extracted with ethyl acetate. The organic phase was evaporated, and samples were reconstituted in acetonitrile and filtered through a 0.2-μm filter for subsequent LC–MS analysis.

Salinosporamides were identified by LC–MS analysis using an Agilent 1260 Infinity LC system coupled to an Agilent 6530 Accurate-Mass Q-TOF. A solvent system of acetonitrile and water both containing 0.1% formic acid (vol/vol) was used. A 10-μl aliquot was injected on a Phenomenex Kinetex C18 reversed-phase HPLC column (5 μm, 250 mm × 4.6 mm) and eluted over a 40-min method with a gradient from 20 to 100% over 25 min; 100% acetonitrile was held for 5 min before the acetonitrile percentage was dropped to 20%. The flow rate was 0.75 ml min^{-1}. Eluent was detected using ESI-MS, monitoring m/z 70–3,200 in positive mode with a speed of 32,500 m/z s^{-1} and 30-kV collision energy. MS data were analyzed using Agilent MassHunter Qualitative Analysis B.05.01 software. Salinosporamide A was purified via methods described previously[5].

SalC expression and purification from *Streptomyces*. *S. coelicolor*[54] CH999-pCJW93-SalC was grown as described above and collected at 4 °C by centrifugation at 16,000*g* for 20 min. The cell pellet was resuspended in collection buffer (100 mM Tris, 300 mM NaCl, 0.8 mM TCEP and 10% glycerol) and lysed on ice with a QSonica sonicator (6-mm tip at 60% amplitude for 15 cycles of 15 s pulse on followed by 45 s pulse off). The lysate was then centrifuged for 60 min at 4 °C and 44,000*g* and subjected to column chromatography.

Protein purification was performed on an ÄKTApurifier instrument (GE Healthcare) with the modules Box-900, UPC-900, R-900 and Frac-900 with all buffers filtered through a nylon membrane 0.2-μm (Merck) before use. Fast protein LC data were analyzed with UNICORN 5.31 (Built 743) software. SalC was initially purified by Ni^{2+} affinity chromatography using 5-ml HisTrap FF (GE Healthcare) columns preequilibrated in buffer A (300 M NaCl, 100 mM Tris, 25 mM imidazole and 0.8 mM TCEP, pH 8.0). Lysate was loaded onto columns at 1.5 ml min^{-1}, after which the column was washed with eight column volumes (40 ml) of buffer A. Protein was eluted by a linear gradient of 0–100% buffer B (300 M NaCl, 100 mM Tris, 250 mM imidazole and 0.8 mM TCEP, pH 8.0) over 30 min at a flow rate of 2.0 ml min^{-1}. Protein-containing fractions were identified by SDS–PAGE in the presence of reducing agents (12% acrylamide) and then combined and concentrated to a volume of ~3 ml using Amicon Ultra centrifugal filters with a 50-kDa molecular weight cutoff (EMD Millipore). Concentrated protein was then loaded onto a HiLoad Superdex 200 prep-grade size-exclusion column (16 cm × 60 cm, GE Healthcare) equilibrated in 50 mM Tris, 150 mM NaCl, 0.8 mM TCEP and 10% glycerol (pH 8.0) buffer and eluted at a constant flow rate of 1.0 ml min^{-1}. Fractions containing the target protein were pooled, concentrated, aliquoted and flash-frozen for storage at −80 °C. Protein concentration was determined by Bradford assay.

Protein expression and purification from *E. coli*. *E. coli* BL21Gold(DE3) transformed with expression plasmids for CoaA, CoaD, CoaE, Sfp and SalB-PCP were inoculated in 10 ml of LB containing kanamycin. After overnight incubation, this 10-ml culture was used to inoculate 1 liter of TB medium in a 2.8-liter baffled flask supplemented with kanamycin. Flasks were incubated at 37 °C and 220 r.p.m. until an optical density at 600 nm (OD_{600}) reached 0.6–0.8, at which time flasks were cooled for 1 h at 18 °C before induction with 1 M isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.2 mM. Flasks were incubated overnight at 18 °C and 220 r.p.m. and collected by centrifugation at 4 °C and 5,000*g* for 10 min. From this point on, all subsequent steps were performed at 4 °C or on ice. The cell pellet was subsequently resuspended in lysis buffer (500 mM NaCl, 20 mM Tris and 10% glycerol, pH 8.0) and lysed by sonication with a Qsonica sonicator (6-mm tip at 40% amplitude for 10 cycles of 15 s pulse on followed by 45 s pulse off). Lysate was then centrifuged at 14,000*g* for 45 min at 4 °C. Protein purification was performed as described in the previous section using buffer A (1 M NaCl, 20 mM Tris and 25 mM imidazole, pH 8.0) and buffer B (1 M NaCl, 20 mM Tris and 250 mM imidazole, pH 8.0).

CoaA, CoaD, CoaE and Sfp proteins were buffer exchanged using PD-10 desalting columns (GE Healthcare) to storage buffer (20 mM Tris and 10% glycerol, pH 8.0) as described previously[55]. Protein concentration was determined by Bradford assay before the samples were aliquoted, flash-frozen and stored at −80 °C.

SalB-PCP was further purified by SEC using a HiLoad Superdex 75 prep-grade gel filtration column (16 cm × 60 cm, GE Healthcare) preequilibrated with storage buffer (20 mM Tris, 200 mM NaCl and 10% glycerol, pH 8.0). Fractions with the target protein were pooled and concentrated with centrifugal filters again. Protein concentration was determined using the Bradford assay and flash-frozen and stored at −80 °C.

Tryptic fingerprinting of SalB protein. To confirm the identity of the purified SalB protein and monitor conversion from apo to holo and acylated holo forms, the protein was digested using a Trypsin Single Proteomics Grade kit (Sigma-Aldrich) and analyzed by high-resolution LC–MS analysis. A 10-μl aliquot was injected onto a Phenomenex Aeris WIDEPORE XB-C18 200-Å LC column (3.6 μM, 250 mm × 4.6 mm) and analyzed with an Agilent 1260 Infinity LC system coupled to an Agilent 6530 Accurate-Mass Q-TOF. A solvent system of acetonitrile and water both containing 0.1% formic acid (vol/vol) was used. Peptide fragments were eluted over a 43-min method with a gradient from 5 to 30% acetonitrile over 4 min, 30 to 65% acetonitrile over the next 15 min and to 100% over 5 min; 100% acetonitrile was held for 6 min before the concentration was dropped to 5%. The flow rate was 0.75 ml min^{-1}. Eluent was detected using ESI-MS, monitoring m/z 70–3,200 in positive mode with a speed of 32,500 m/z s^{-1} and 30-kV collision energy. MS data were analyzed using Agilent MassHunter Qualitative Analysis B.05.01 software. Peptide fragments were compared to predicted trypsin digest fragments using the ExPASy PeptideMass tool.

SalC hydrolysis assay. To probe the SalC reaction in the reverse direction, 50-µl assays containing purified SalC (20 µM), SalB-PCP (50 µM) and 1 mM salinosporamide A were incubated in SalB buffer (20 mM Tris, 200 mM NaCl, 0.8 mM TCEP and 10% glycerol pH 8.0). After 3 h, the assays were quenched with three volumes of acetonitrile and centrifuged at 21,000*g* for 30 min to precipitate protein. Samples were filtered and analyzed by LC–MS with the method for salinosporamide identification described earlier.

Chemical synthesis. The syntheses and accompanying spectral data of the pantetheine-tethered substrates **13** and **21** are described in the Supplementary Notes.

SalC activity assays. Linear pantetheine-activated substrates (**13** or **21**, 250 µM) were incubated with purified CoaA (7 µM), CoaD (10 µM), CoaE (10 µM), Sfp (10 µM) and SalB-PCP (250 µM) in reaction buffer (50 mM phosphate, 100 mM NaCl, 0.8 mM TCEP and 10 mM $MgCl_2$, pH 7.5) in a 50-µl reaction to allow carrier protein loading to occur. The reaction was initiated with the addition of 5 mM ATP. After 3 h, 25 µl of acylated SalB-PCP was removed and used for trypsin fingerprinting or intact MS. SalC (20 µM) was added to the remaining 25 µl for activity assays. After testing different reaction conditions (buffer, pH, concentrations of all proteins and time), these conditions were found to be optimal.

Assays with diffusible substrates omitted required CoA or Sfp proteins. After the addition of SalC, reactions were incubated for an additional 3 h at 30 °C. Reactions were quenched by the addition of three volumes of acetonitrile and centrifuged at 21,000*g* for 30 min to precipitate protein. Finally, samples were analyzed via high-resolution LC–MS using the method for salinosporamide identification described earlier.

Preparative-scale SalC activity assay. For purification of the SalC assay product, reactions were set up as described above at a 50-ml scale in an Erlenmeyer flask. SalB-PCP loading was allowed to proceed for 12 h, at which point approximately 1 g of XAD7 resin and SalC (20 µM) were added to the reaction mixture, and assays were allowed to proceed overnight. In the morning, resin was extracted (3×) with ethyl acetate. Organic extracts were combined and evaporated to dryness. Samples were resuspended in acetonitrile, and the peak corresponding to (***R***)-**17** was purified by preparative HPLC using a Phenomenex Luna C18 column (5 µm, 100 mm, 2-mm inner diameter) along with an Agilent Technologies system composed of a PrepStar pump, a ProStar 410 autosampler and a ProStar UV detect (Agilent Technologies). The samples were eluted by a gradient from 20 to 100% acetonitrile over 40 min at a flow rate of 10 ml min^{-1}. The peak corresponding to (***R***)-**17** was collected and dried by rotary evaporation and lyophilization. Finally, (***R***)-**17** was purified using a small-scale silica column with a dichloromethane/acetonitrile solvent system. All stages of purification were monitored by LC–MS analysis.

Structural characterization of (*R*)-17. All NMR data were collected at the University of California (UC) San Diego Skaggs School of Pharmacy and Pharmaceutical Sciences NMR Facility on a 600-MHz Varian NMR spectrometer (Topspin 2.1.6 software, Bruker) with a 1.7-mm cryoprobe. Deuterated chloroform containing TMS standard was used as a solvent. All spectra for purified (***R***)-**17** can be found in the Supplementary Notes.

Proteasome inhibition assay. SalC activity assays (±SalC and ±**13**) were run on a 1-ml scale with the addition of resin, as described in Preparative-scale SalC activity assay. Resin extracts were resuspended in 50 µl of buffer and used for proteasome inhibition assays and LC–MS analysis.

Proteasome inhibition assays were performed using the QUANTIZYME Assay System (Enzo Life Sciences), designed to measure chymotrypsin-like protease activity (β5-subunit) of the purified 20S human proteasome using the fluorogenic substrate Suc-LLVY-AMC. Inhibitor (0.5 µM epoxomicin, 0.5 µM salinosporamide A or 10 µl of resuspended SalC activity assay extracts) was added to 0.2 µg of proteasome in kit buffer (50 mM Tris-HCl, pH 7.5, 25 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$ and 0.03% SDS) and incubated for 15 min. Substrate was added to a final concentration of 75 µM. Proteasome activity was measured by reading fluorescence of the cleaved substrate at 355 nm (excitation) and 460 nm (emission) using a Spectra Max M2 (Molecular Devices) every 5 min for 30 min. All assays were performed in duplicate.

Chain transfer assays and intact protein mass spectrometry. SalB-PCP was acylated with the pantetheine-activated mechanistic probe **21** using the same methods described earlier. This holo-acyl-SalB-PCP was purified away from residual CoA proteins and small molecules using a Superdex 75 prep-grade gel filtration column (16 cm × 60 cm, GE Healthcare) preequilibrated with buffer (50 mM Tris, 150 mM NaCl and 5% glycerol, pH 8.0). Purified **22** or diffusible substrates (**21**–**25**) were incubated with SalC or a SalC variant (20 µM) for 3 h and subjected to LC–HRMS analysis. All intact protein LC–HRMS was performed by the UC San Diego Molecular Mass Spectrometry Facility on an Agilent 1260 Infinity Binary LC coupled with a 6230 Accurate-Mass TOFMS.

Crystallization of SalC. SalC was crystallized by hanging drop crystallization at room temperature. Briefly, 1.0 µl of 1.7–3.4 mg ml^{-1} SalC in gel filtration buffer was mixed with 1.0 µl of well solution (0.10 M HEPES, pH 8.5, 30% (wt/vol) PEG 3350 and 0.30 M KCl) to make a 2-µl hanging drop in a sealed well with 400 µl of well solution. Transparent plate crystals grew in 24–48 h. The crystals were then transferred to a drop of LV Cryo Oil (MitGen) and flash-cooled in liquid nitrogen.

Data collection and processing. A dataset of SalC was collected at Advanced Light Source (Berkeley) on beamline 8.2.2 using a ADSC Q315R detector at a temperature of 100 K. Reflections could be observed to 2.85 Å. Data were indexed and integrated using XDS[55], with high-resolution cutoff to 2.85 Å, based on $CC_{1/2}$ = ~ 0.7. CC1/2 = Pearson correlation coefficient of the intensity values produced from two half data sets. Data statistics are listed in Supplementary Table 3.

Structure determination and refinement. The structure of SalC was determined to a resolution of 2.85 Å by molecular replacement (MR) using Phaser[56] implemented in CCP4. The KS domain of DEBS1 from *Saccharopolyspora erythraea* (PDB ID: 2HG4)[57], which shared 33.5% identity with SalC, was used to generate a homology model of SalC by I-TASSER[58] as the MR search model. The MR results in one SalC tetramer per asymmetric unit. Resulting log-likelihood gain (LLG), R value (Rval) and translation function Z score (TFZ) were 8,178.1, 44.8 and 66.5, respectively. The MR solution was used as an input for AutoBuild implemented in Phenix[59] for initial model building. After the AutoBuild run, the atomic coordinates and B-factors were iteratively refined in Phenix Refine with model building and manual adjustment of the model in Coot[60]. Water molecules were added manually throughout real space refinements using F_o–F_c electron density contoured to 3.0σ as criteria. Fourfold non-crystallographic symmetry restraints were used throughout refinement. Final cycles of refinements include translation–libration–screw rotation parameterizations with two translation–libration–screw groups per SalC monomer. The division was assigned based on the two domains of SalC. A composite-omit electron density map calculated by Phenix Composite_omit_map was used to verify the model. The refinement statistics are in Supplementary Table 3, and the final model of the SalC structure contains residues listed in Supplementary Table 4. All structure figures were rendered in PyMOL.

SalC mutagenesis. SalC mutagenesis was performed by generating two PCR products using the *salC* forward primer and a mutagenic reverse primer and a mutagenic forward primer and the *salC* reverse primer. Three-piece Gibson assembly was used to assemble these two amplicons into linearized pCJW93. pCJW93-SalC mutants were subsequently conjugated into *S. coelicolor* CH999. Following expression and purification, SalC variants were subjected to SalC activity assays and chain transfer assays as described above.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Strains and plasmids used in this study are described in Supplementary Table 1. All oligonucleotides (Integrated DNA Technology) are shown in Supplementary Table 2. The salinosporamide (*sal*) BGC from *S. tropica* CNB440 is available in the MIBiG database (accession BGC0001041). Other salinosporamide BGCs and the cinnabaramide BGC used for alignments are available in the IMG JGI database. Functional KS and non-elongating KS protein sequences used for alignments can be found in the Supplementary material and through MIBiG. Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the PDB under accession number 7S2X(native SalC). Additionally, the following PBD datasets were used for SalC structural comparison: 2HG4, 4WKY, 2QO3 and 4NA2. Source data are provided with this paper. Other relevant data supporting the findings of this study are available in this published article or its Supplementary files.

References

49. Gilchrist, C. L. M. et al. cblaster: a remote search tool for rapid identification and visualisation of homologous gene clusters. *Bioinformatics Adv.* **1**, vbab016 (2021).
50. Gilchrist, C. L. M. & Chooi, Y.-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
51. Kieser, T. *Practical Streptomyces Genetics* (John Innes Foundation, 2000).
52. MacNeil, D. J. et al. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).
53. Flett, F., Mersinias, V. & Smith, C. P. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting *Streptomycetes*. *FEMS Microbiol. Lett.* **155**, 223–229 (1997).
54. Wilkinson, C. J. et al. Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426 (2002).
55. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

56. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
57. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl Acad. Sci. USA* **103**, 11124–11129 (2006).
58. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

Acknowledgements

We are grateful to P.R. Jensen (Scripps Institution of Oceanography) for the *S. tropica* CNB440 strain, B.M. Dungan (UC San Diego) for assistance with NMR, K.E. Creamer (UC San Diego) for the NaPDoS2.0 analysis and helpful bioinformatics discussions, Y. Su (UC San Diego Molecular Mass Spectrometry Facility) for intact proteomics experiments and J.P. Noel and G. Louie (Salk Institute for Biological Studies) for beamline coordination. Beamline 8.2.2 of the Advanced Light Source, a U.S. DOE Office of Science User Facility under contract number DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30-GM124169-01. This work was supported by NIH grants R01CA127622 (B.S.M.), R01AI047818 (B.S.M.) and F31HD101307 (K.D.B.), R35GM134910 (D.R.), the Robert A. Welch Foundation (grant number AA-1280 (D.R.)) and the São Paulo Research Foundation (FAPESP; grant number 2011/21358-5 (D.B.B.T.)).

Author contributions

K.D.B., V.V.S., T.A.M.G. and B.S.M. designed the study. K.D.B. performed protein expression and purification, enzymatic assays, mutagenesis and data analysis. V.V.S. synthesized and characterized all substrates in this study. P.Y.-T.C., D.B.B.T. and K.D.B. performed all crystallography experiments and subsequent data analysis. T.A.M.G. performed all gene inactivation experiments, preliminary protein expressions and initial substrate synthesis. S.V. and D.R. provided helpful initial synthetic discourse. K.D.B. and B.S.M. wrote the manuscript with input from all coauthors.

Competing interests

The authors declare no competing interests.

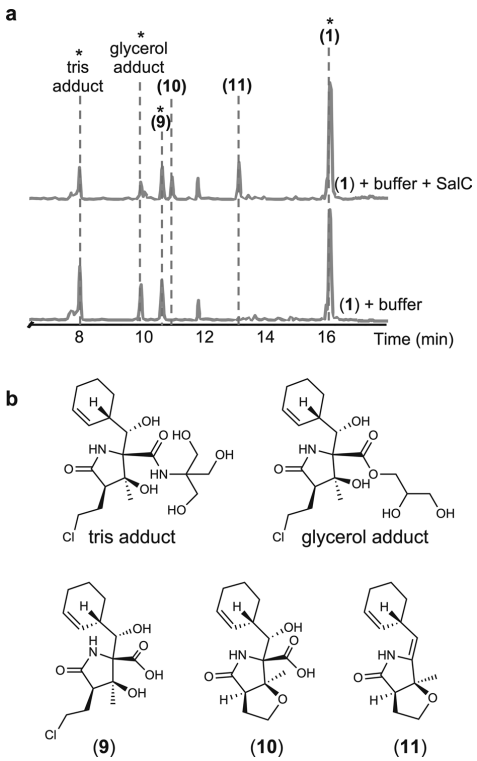
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-022-00993-w.

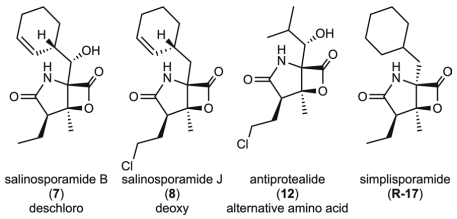
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-022-00993-w.

Correspondence and requests for materials should be addressed to Bradley S. Moore.

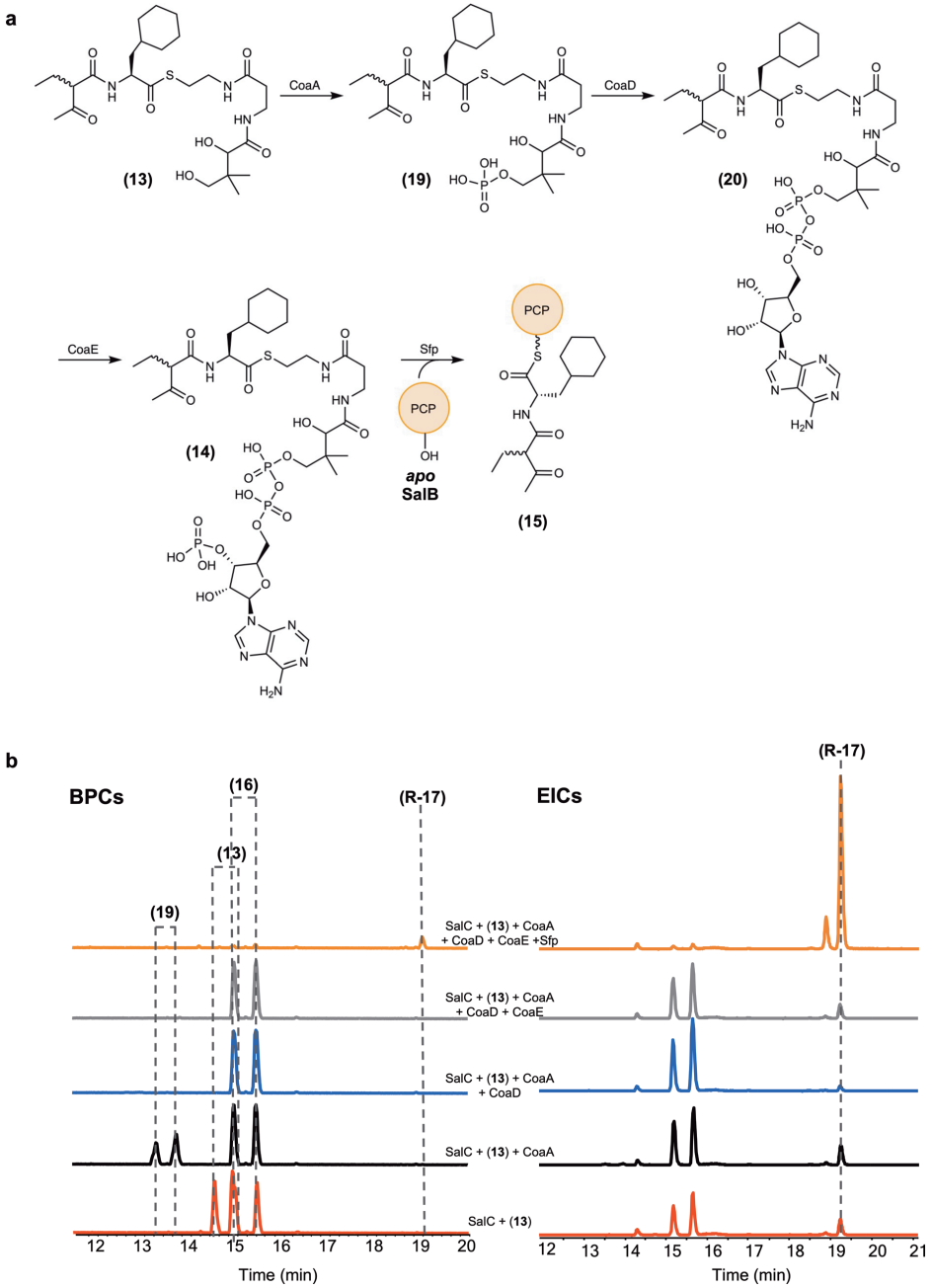
Peer review information *Nature Chemical Biology* thanks Andrew Gulick, Yi Tang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



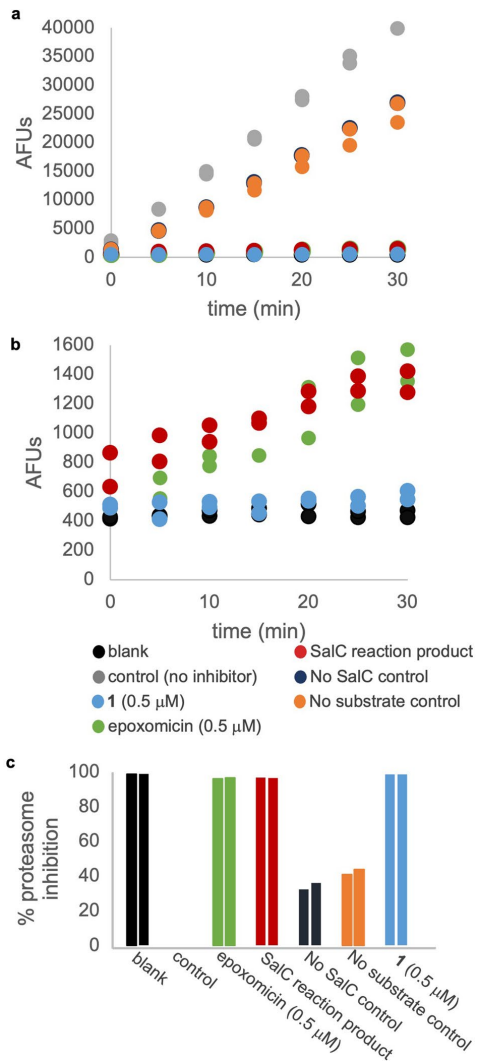
Extended Data Fig. 1 | Salinosporamide A hydrolysis and subsequent THF ring formation is accelerated by the presence of SalC. a, LCMS chromatograms of salinosporamide A (**1**) hydrolysis assay with and without SalC. * indicates compound retains chloride and no THF ring formation has occurred as evidenced by characteristic isotope pattern. **b**, Structures of compounds putatively identified by LCMS in **a**.



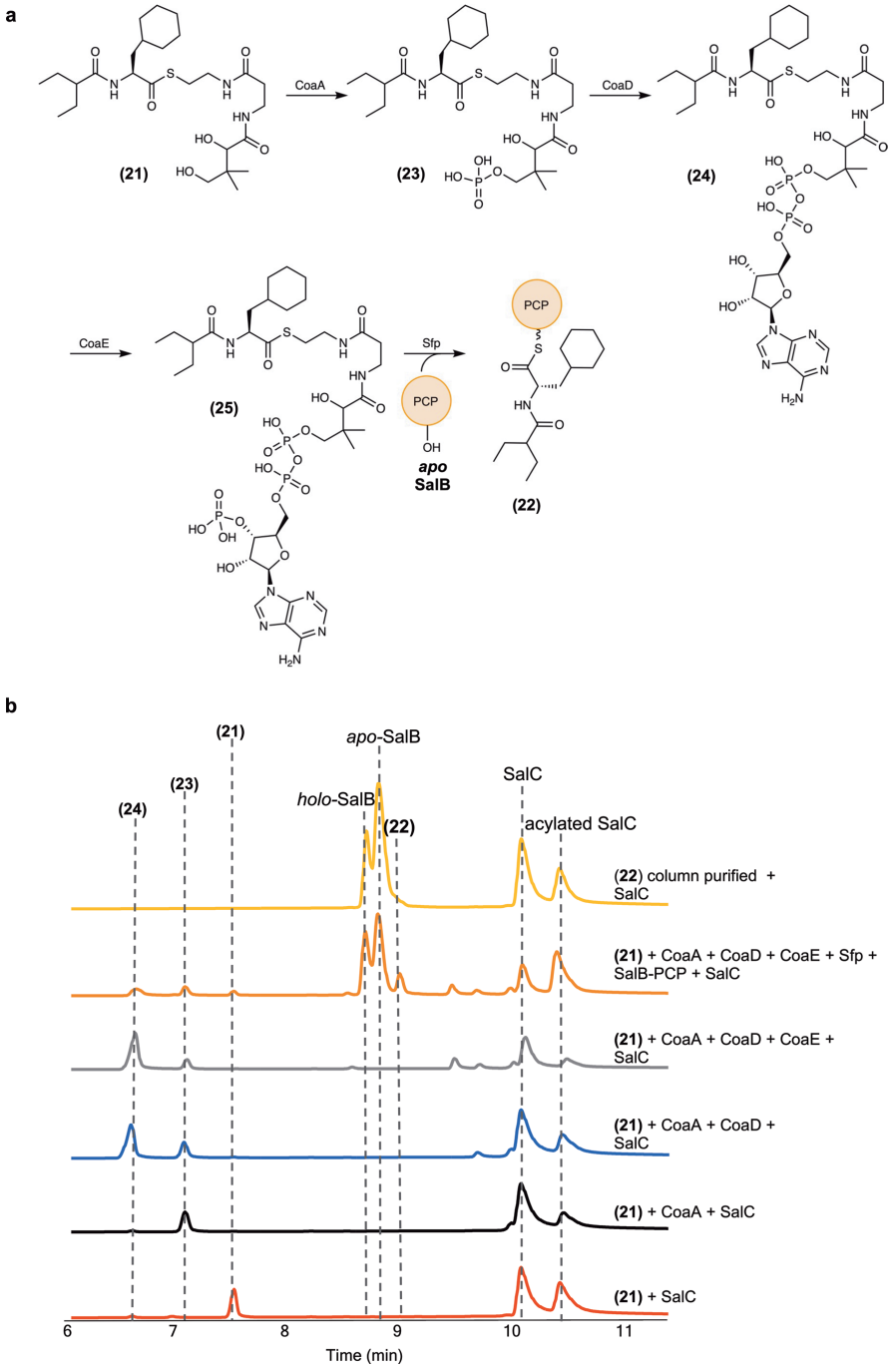
Extended Data Fig. 2 | Naturally produced analogs of salinosporamide A that served as inspiration for simplisporamide.



Extended Data Fig. 3 | SalC activity assay with diffusible substrates. a, Reaction scheme depicting chemoenzymatic synthesis and subsequent SalB-PCP acylation assay to generate **(15). b**, LCMS chromatograms (BPCs and EICs *m/z* 266.18) of SalC activity assay with diffusible substrates shown in part (**a**) including pantetheine-activated (**13**), CoA product-activated (**19**), CoaD product-activated (**20**), CoA activated (**14**), and carrier protein-activated substrate (**15**). Substrates were activated via *in vitro* CoA enzyme biosynthesis. All SalC activity assays contained *apo*-SalB-PCP as well.

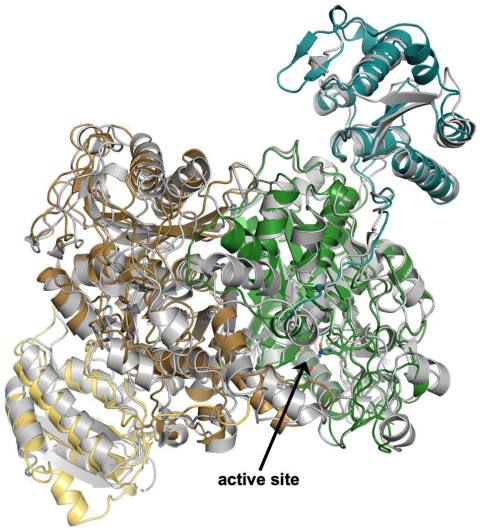


Extended Data Fig. 4 | Proteasome inhibitory activity of SalC assay product using purified human 20S proteasome. a, Proteasome activity determined by reading fluorescence (AFUs) of the cleaved substrate (Suc-LLVY-AMC) at 355 nm (excitation) and 460 nm (emission) every five minutes after substrate was added for 30 min in the presence of various inhibitors. Blank (no proteasome added) = black, control (no inhibitor) = gray, epoxomicin (0.5 μM) = green, SalC reaction product = red, no SalC control = dark blue, no substrate control = orange, salinosporamide A (0.5 μM) = light blue. Samples run in duplicate, all data points shown. b, Magnification of y-axis of plot from a to examine successful inhibition of the 20S proteasome by epoxomicin, SalC reaction product, and salinosporamide A. c, Percent proteasome inhibition at 30 min, relative to control (no inhibitor, 0% inhibition). See Supplementary Fig. 14 for corresponding LCMS traces of extracts used in these assays.

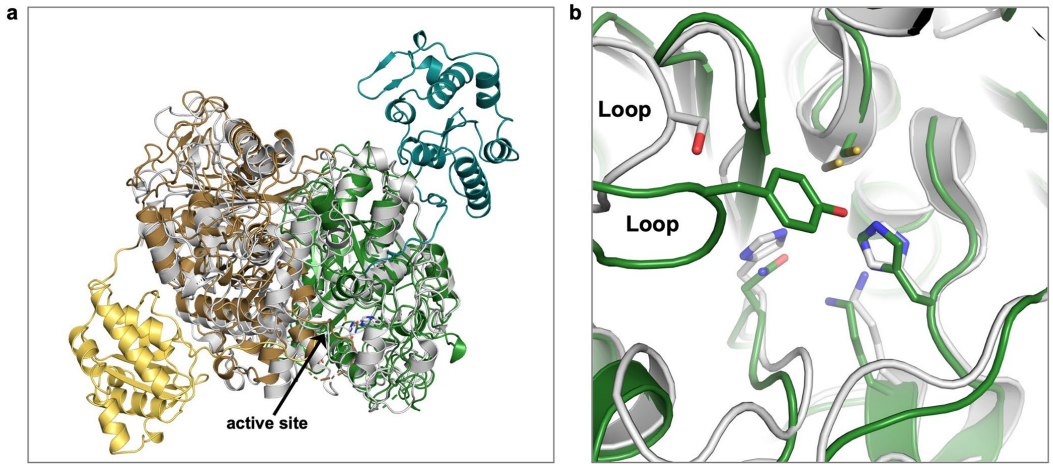


Extended Data Fig. 5 | See next page for caption.

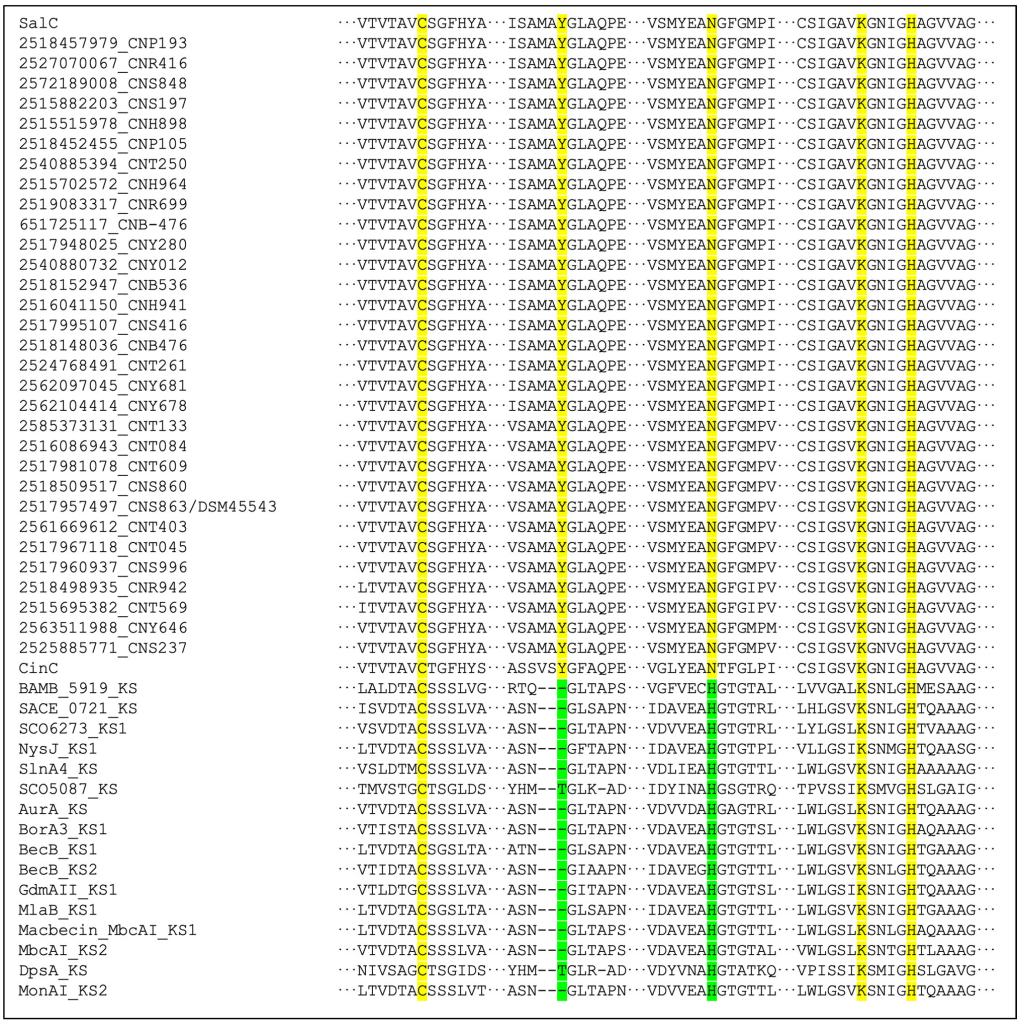
Extended Data Fig. 5 | Acylation of SalC with diffusible substrates. a, Reaction scheme depicting chemoenzymatic synthesis to generate (**22**) **b**, UV chromatograms (215 nm) of intact protein LCMS for transacylation assay with diffusible substrates. Transacylation assay utilized linear mechanistic probe (**21**) activated in different ways (CoA-precursor-activated (**23**, **24**), CoA-activated (**25**), and SalB-PCP-tethered, all generated *in situ*) and SalC. Transacylation assay with column purified **22** shown for comparison.



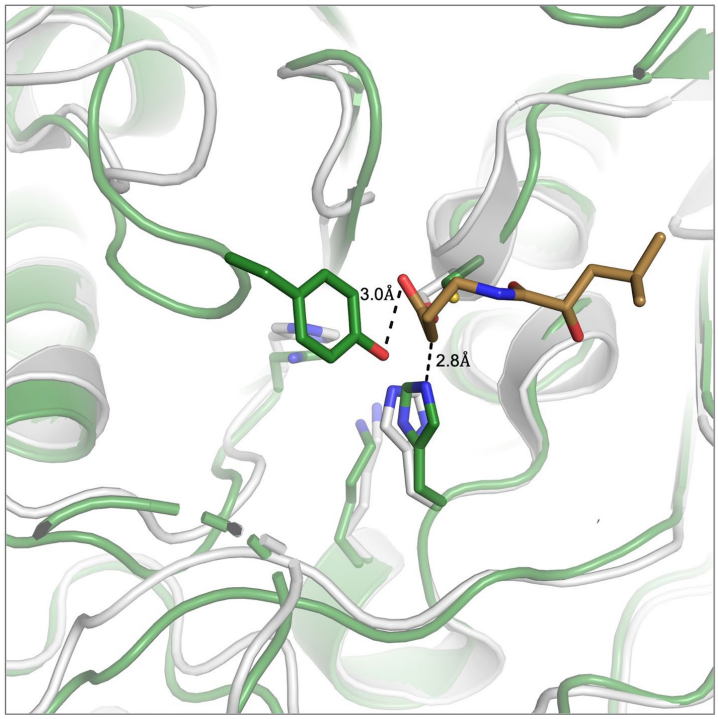
Extended Data Fig. 6 | SalC overlay with a *trans*-AT KS. SalC structure aligned with closest Dali server homolog, the *trans*-AT KS OzmN KS2 (PDB ID: 4WKY) from the hybrid NRPS/PKS oxazolomycin pathway, RMSD 0.842 Å. Overall structure of SalC dimer (colors shown as previous, KS monomers in brown and green, flanking subdomains in yellow and teal) with OzmN KS2 (gray).



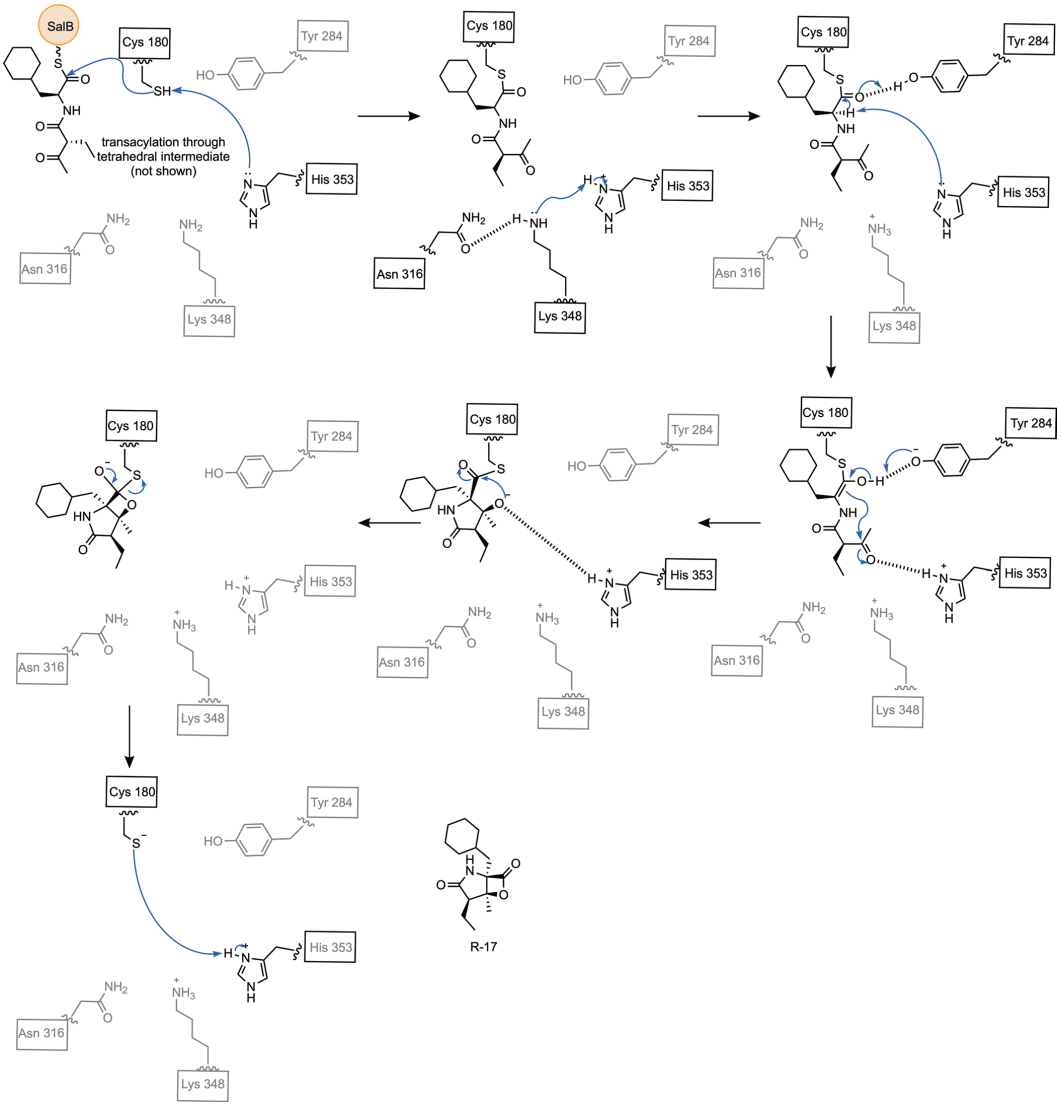
Extended Data Fig. 7 | SalC overlay with functional type I KS. SalC KS aligned with DEBS KS3 (PDB: 2QO3), RMSD 1.225 Å. **a**, Overall structure of SalC dimer (colors shown as previous, KS monomers in brown and green, flanking subdomains in yellow and teal) with DEBS KS3 (gray). **b**, Active site overlay of SalC (green) and DEBS KS3 (gray).



Extended Data Fig. 8 | Tyr284 is conserved in SalC homologs. Condensed alignment showing conservation of Tyr284 in all SalC homologs from *Salinispora* and *Streptomyces cinnabarigriseus* JS360 (CinC) but not in canonical elongating KSs. All SalC homologs from *Salinispora* strains found in JBI IMG database. For functional KS sequences refer to Supplementary Figure 5.



Extended Data Fig. 9 | SalC structure overlaid with bacillaene PKS (*bae*) KS2 bound to its natural intermediate. SalC is shown in green and BaeKS2 is shown in grey, *bae* intermediate in gold. PDB ID: 4NA2, RMSD 1.071 Å.



Extended Data Fig. 10 | Proposed active site mechanism of SalC. Catalysis is initiated by deprotonation of Cys180 followed by transacylation of the SalB-PCP tethered substrate through a tetrahedral intermediate (not shown). Lys348 deprotonates His353, and hydrogen bonding of the thioamide carbonyl to Tyr284 facilitates deprotonation of the thioester α-proton by His353. An intramolecular aldol reaction forms the γ-lactam; the oxyanion is presumably stabilized by dipole interactions with backbone amides, as is hypothesized for KSs[25]. Subsequent β-lactonization through a tetrahedral intermediate leads to release of simplisporamide from SalC. Finally, Cys180 is reprotonated by His353.

nature research

Corresponding author(s): Bradley S. Moore

Last updated by author(s): Sep 6, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	LCMS data was collected using an Agilent 1260 Infinity LC system coupled to an Agilent 6530 Accurate-Mass Q-TOF. NMR data were collected with a 600 MHz Varian NMR spectrometer (Topspin 2.1.6 software, Bruker). A data set of SalC was collected at Advanced Light Source (Berkeley, California, USA) on beamline 8.2.2 using a ADSC Q315R detector at a temperature of 100 K. Reflections could be observed to 2.85 Å. Data were indexed and integrated using XDS,59 with high resolution cut-off to 2.85 Å, based on CC1/2 = ~ 0.7. Data statistics are listed in Supplementary Table 3.
Data analysis	The SalC protein sequence was analyzed via BLAST, and all alignments were made using Clustal Omega (1.2.4) software. Geneious Prime (2019.2.3) software was used for all plasmid maps and primer design. The NaPDoS platform used to examine SalC in relation to other KS domains, cblaster (1.2.9) was used to examine SalC homologs in salinosporamide producing strains, and clinker (v0.0.21) was used for visualization. MS data were analyzed using Agilent MassHunter Qualitative Analysis B.05.01 software and NMR spectra were analyzed using MestraNova (14.2.2) The structure of SalC was determined to 2.85-Å resolution by molecular replacement (MR) using Phaser60 implemented in CCP4. The ketosynthase domain of DEBS1 from Saccharopolyspora erythraea (PDB ID: 2HG4)61, which shared 33.5% identity with SalC, was used to generate a homology model of SalC by I-TASSER62 as the MR search model. The MR results in one SalC tetramer per asymmetric unit (ASU). Resulting LLG, Rval and TFZ were 8178.1, 44.8 and 66.5, respectively. The MR solution was used as an input for AutoBuild implemented in Phenix63 for initial model building. After the AutoBuild run, the atomic coordinates and B-factors were iteratively refined in Phenix Refine with model building and manual adjustment of the model in Coot.64 Water molecules were added manually throughout real space refinements using Fo-Fc electron density contoured to 3.0 σ as criteria. Four-fold non-crystallographic symmetry (NCS) restraints were used throughout refinement. Final cycles of refinements include Translation-Libration-Screw-rotation (TLS) parameterizations with two TLS group per SalC monomer. The division is assigned based on the two domains of SalC. A composite-omit electron density map calculated by Phenix Composite_omit_map was used to verify the model. The refinement statistics are in Supplementary Table 3, and the final model of the SalC

structure contains residues listed in Supplementary Table 4. All structure figures were rendered in PyMOL (version 2.4.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Strains and plasmids used in this study are described in Supplementary Table 1. All oligonucleotides (Integrated DNA Technology) are shown in Supplementary Table 2. The salinosporamide (sal) biosynthetic gene cluster from S. tropica CNB440 is available in the MIBiG database (accession BGC0001041). Other salinosporamide BGCs and the cinnabaramide BGC used for alignments are available in the IMG JGI database. The accession number for the top 50 SalC homologs found via BLASTP are available in Supplementary Fig. 4. Functional KS and non-elongating KS protein sequences used for alignments can be found in the Supplementary material and through MIBiG. Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the Protein Data Bank under accession number 7S2X (native SalC). Additionally, the following PBD datasets were used for SalC structural comparison: 2HG4, 4WKY, 2QO3, 4NA2. Source data for the proteasome inhibition assay (Extended Data Fig. 4) is provided with this paper. Other relevant data supporting the findings of this study are available in this published article or its Supplementary files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation for the proteasome inhibition assay was performed. This assay was meant to provide additional qualitative data supporting the formation for the b-lactone, not to quantify inhibitory activity. Furthermore this assay required a significant amount of protein and substrate and so was run in duplicate.
Data exclusions	No data were excluded from the study.
Replication	All sal knockout strains were cultured independently five times and analyzed for salinosporamide production. SalC hydrolysis assays, activity assays, and chain transfer assays were repeated at least three times on multiple days. All replication attempts were successful. Proteasome inhibition assays were run in duplicate due to the amount of material required for these assays, but all controls (including replicates) were repeated in triplicate on four different days.
Randomization	Randomization was not relevant to our study as for all samples in an experiment the same biological sample or recombinant enzyme was used and the samples underwent identical analysis.
Blinding	Blinding was not relevant to our study because no manual counting, scoring, or sorting was employed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

nature portfolio

https://doi.org/10.1038/s41589-022-00993-w

Supplementary information

Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead

In the format provided by the authors and unedited

Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead

Katherine D. Bauman,[1] Vikram V. Shende,[1] Percival Yang-Ting Chen,[1,2] Daniela B. B. Trivella,[1,3,4] Tobias A. M. Gulder,[1,5] Sreekumar Vellalath,[6] Daniel Romo,[6] Bradley S. Moore[1,7*]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, 92093, USA

[2]Present Address: Morphic Therapeutics, Waltham, MA, 02451, USA

[3]Present Address: Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, Campinas, 13083-970, São Paulo, Brazil

[4]Present Address: Institute of Chemistry, University of Campinas, Campinas, 13083-970, São Paulo, Brazil

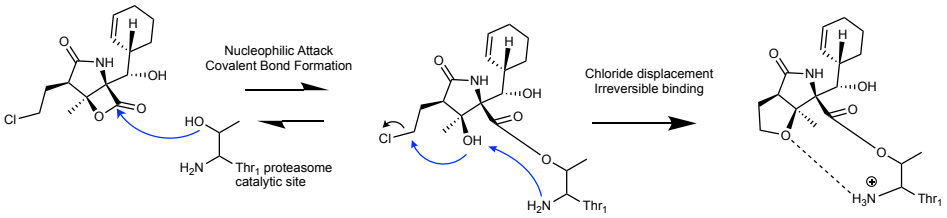
[5]Present Address: Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany

[6]Department of Chemistry and Biochemistry, Baylor University, One Bear Place #97348, Waco, Texas 76798, USA

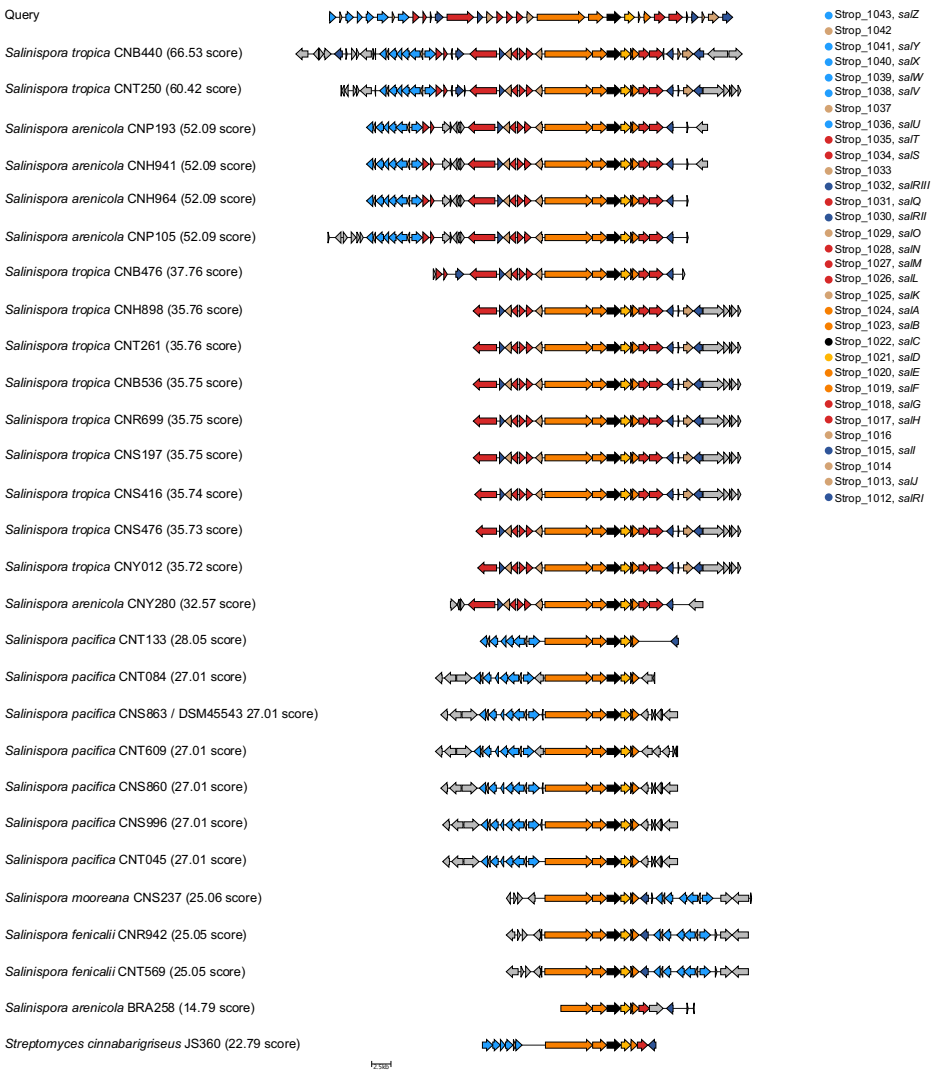
[7]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, 92093, USA

*email: bsmoore@ucsd.edu

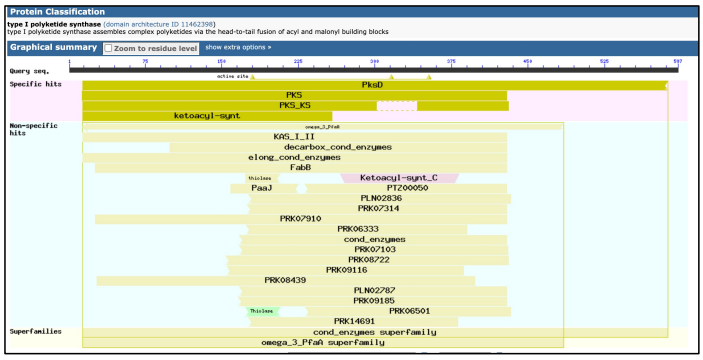
SI Table of Contents



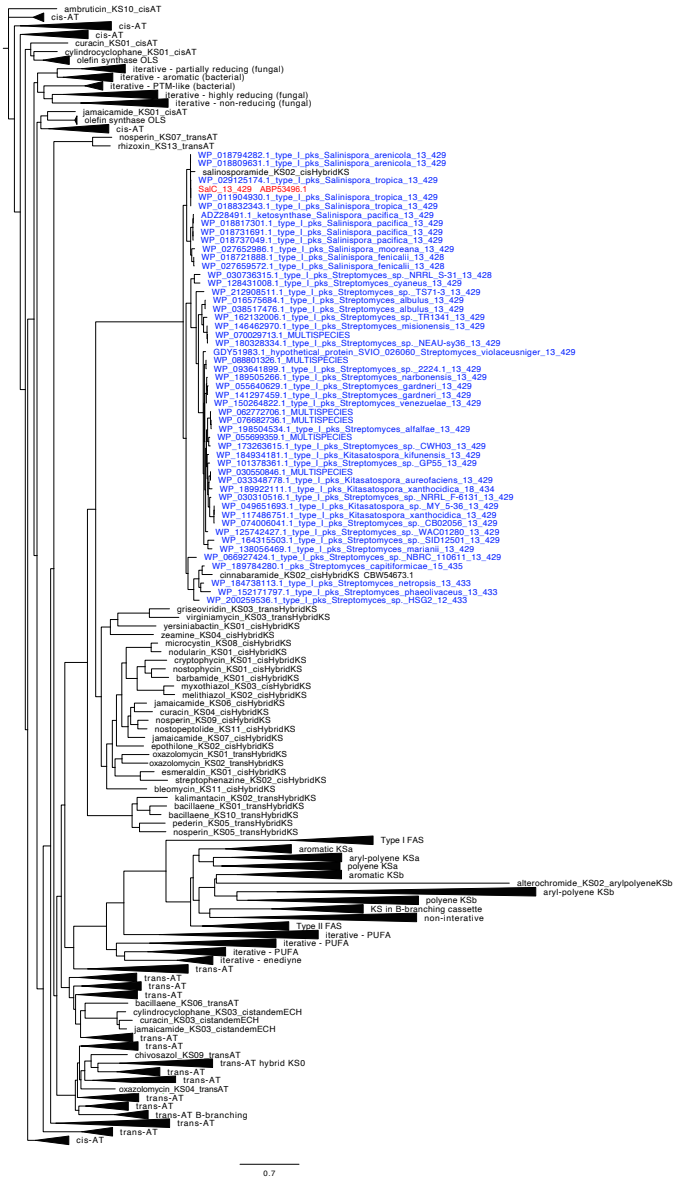
Supplementary Fig. 1: Salinosporamide A mechanism of proteasome inhibition. Nucleophilic attack by the proteasome N-terminal catalytic threonine on the salinosporamide A β-lactone results in formation of a covalent salinosporamide—proteasome adduct. Chloride displacement and formation of the THF ring results in irreversible binding.



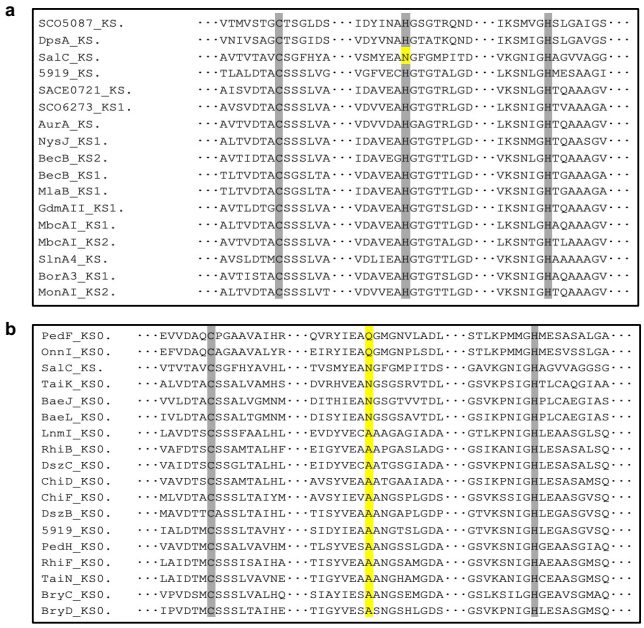
Supplementary Fig. 2: SalC (black) conservation across salinosporamide and cinnabaramide producing strains. Each protein in the *S. tropica* CNB440 BGC was used as a query to search strains found in the NCBI nr protein database using cblaster. Clusters with SalA (required for salinosporamide production) plus at least 2 other unique hits not more than 20000 bp apart was considered a cluster; a hit was considered if it had at least 30% identity, 0.01 e-value, and 50% coverage. Shown here are all hit clusters visualized using clinker from the genus *Salinispora* and the cinnabaramide producer *Streptomyces cinnabarigriseus* JS360 showing the conservation of SalC across salinosporamide producing strains.



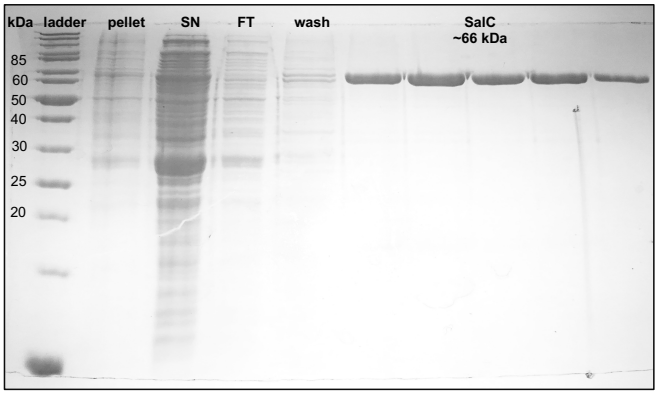
Supplementary Fig. 3: SalC is a standalone ketosynthase. Protein BLAST classification of conserved domains found in SalC.



Supplementary Fig. 4: NaPDoS2 analysis of SalC. SalC ABP53496.1 (red), SalC homolog from cinnabaramide pathway CinC CBW54673.1 (black), plus 50 top BLAST homologs (blue) form their own clade within phylogenetic tree of 414 KS reference sequences. Tree courtesy of Kaitlin E. Creamer and Leesa J. Klau.



Supplementary Fig. 5: Multiple protein sequence alignment of the SalC KS domain active site residues with selected elongating KS and KS^0 domains. a, Elongating KSs: SCO5087 = actinorhodin_SCO5087 KS (aa 47-437), DpsA = doxorubicin DpsA KS (aa 4-394), SalC = salinosporamide KS (aa 13-432), 5919 = enacyloxin KS (aa 32-454), SACE0721 = erythromycin_SACE0721_KS (aa 559-983), SCO6273 = coelimycin_SCO6273 KS1 (aa 35-458), AurA = aureothin AurA KS (aa 118-541), NysJ = nystatin NysJ_KS1 (aa 42-466), BecB KS2 = BE-014106 BecB KS2 (aa 1802-2227), BecB KS1 = BE-014106 BecB KS1 (aa 113-538), MlaB = ML-449_MlaB_KS1 (aa 116-541), GdmAII = geldanamycin GdmAII_KS1 (aa 35-461), MbcAI KS1 = macbecin_MbcAI_KS1 (aa 1121-1543), MbcAI KS2 = macbecin_MbcAI_KS2 (aa 3162-3588), SlnA4 = salinomycin_SlnA4_KS (aa 35-461), BorA3 = borrelidin_BorA3_KS1 (aa 35-461), MonAI = monensin_MonAI_KS2 (aa 1080-1505). **b, KS^{0}s ref #27,** PedF = pederin PedF KS^0 (aa 45-471), OnnI = onnamide I KS^0 (aa 77-504), SalC = salinosporamide KS (aa 13-432), TaiK = thailandamide K KS^0 (aa 1454-1858), BaeJ = bacillaene BaeJ KS^0 (aa 4532-4937), BaeL = bacillaene BaeL KS^0 (aa 4027-4427), LnmI = leinamycin LnmI KS^0 (aa 1931-2344), RhiB = rhizoxin RhiB KS^0 (aa 1041-1450), DszC = disorazole DszC KS^0 (aa 1886-2313), ChiD = chivosazol ChiD KS^0 (aa 2033-2456), ChiF = chivosazol ChiF KS^0 (aa 4825-5249), DszB = disorazole DszB KS^0 (aa 5501-5927), 5919 = enacyloxin KS^0 (aa1819-2242), PedH = pederin PedH KS^0 (aa 4181-4614), RhiF = rhizoxin RhiF KS^0 (aa 1551-1981), TaiN = thailandamide TaiN KS^0 (aa 660-1078), BryC = bryostatin BryC KS^0 (aa 4621-5038), BryD = bryostatin BryD KS^0 (aa 2205-2641).



Supplementary Fig. 6: SalC protein expression and purification. SDS-PAGE (12% acrylamide) analysis of purified SalC from *Streptomyces coelicolor* CH999. EZ-Run™ Rec protein Ladder (Fisher Scientific) was run alongside with the samples. Expected protein size ~66 kDa. SN = supernatant. FT = flowthrough. Protein was purified with the same results five times.

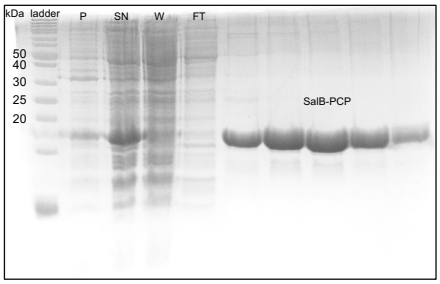
a

```
SalB.                          PLSPVGKLDRTALPA--PPVVDVLPSGDSATTD-TEQALAEIWGALLDRTPQSIGI----      533
fuscachelinPCP_7KW3            -------------------------MVREPATE-AEAALCAVYAEVLGL--DKVGA----       28
eponemycin_epnG_PCP            ------------------------------------RAVAEVWSDLLKV--RAPGA----       18
daptomycin_dptA-PCP1           ---------------------------------EKTLCALFADVLGV--TEVAT----       19
cyclomarin_Sare_4562_PCP3.     ------------------------------------EEILRSVFADVLGI--ERISA----       19
BlmI_PCP-4NEO                  -----------------SNAMSAPRGERTRRRA-LERDIAAIWAETLGR--DSVGP----       36
1DNY_1|Chain                   ------------------MPVTEAQYVAPTNA-VESKLAEIWERVLGV--SGIGI----       34
2N5H_1|Chain                   -------------------------MDGEEVKEKIRRYIMEDLIGPSAKEDEL----       28
5U3H_1|Chain                   -----GTIDYQALKRRHTPEAENPAEADLPQGD-IEKQVAALWQQLLST--GNVTR----      48
2MR8_1|Chain                   ----------------------GAMAKAPESA-TEKVLCALYAEILGV--ERVGV----       30
6LK3_1|Chains                  -----------------QQMGRGSEFHMPSGN-Q---GAAVSVDVLKQLLLDIGIAEQT      48
5JDX_1|Chain                   ------------------------------MLESKLINHIATQF--LDGEKDGL----       22
                                                                :

SalB.                          -------RDTFYDLG-GNSLLLVRLAKRMGQRFHRKVG-VADLF--RFRDIGSLAKWL--      580
fuscachelinPCP_7KW3            -------DADFFALG-GDSVLTLRLVHRARS-AGWEIS-ARHVF--RHPVVADLAAVA--       74
eponemycin_epnG_PCP            -------DQSFFELG-GHSLLAARTLREIAKRLDRDVPPLRMFF--EHPVLQDFADAF--       66
daptomycin_dptA-PCP1           -------DDVFFELG-GHSLNGTRLLARIRTEFGTDLT-LRDLF--AFPTVAGLLPLL--       66
cyclomarin_Sare_4562_PCP3.     -------DDNFFDLG-GHSLLVTRLISRVRGTLGVDVS-MQTFF--NAPTPRQLAGQL--      66
BlmI_PCP-4NEO                  -------HEDFAALG-GNSIHAIKITNRVEELVDAELS-IRVLL--ETRTVAGMTDHV--       83
1DNY_1|Chain                   -------LDNFFQIG-GHSLKAMAVAAQVHREYQVELP-LKVLF--AQPTIKALAQYV--       81
2N5H_1|Chain                   ----DD-QTPLLEWGILNSMNIVKLMVYIRDEMGVSIP-STHITGKYFKDLNAISRTVEQ      82
5U3H_1|Chain                   -------ETDFFQQG-GDSLLATRLTGQLHQ-AGYEAQ-LSDLF--NHPRLADFAATL--      94
2MR8_1|Chain                   -------DDAFHDLG-GSSALAMRLIARIREELGVDLP-IRQLF--SSPTPAGVARAL--      77
6LK3_1|Chains                  LTEIEPGTRLRADLG-LSSVETTDLEIQLRERFGVRIN-LWDKA--DY-TMEQLAVGI--      101
5JDX_1|Chain                   ----DS-QTPLFELNIVDSAAIFDLVDFLRQESKVSIG-MQEIHPANFATVQSMVALVQR      76
                                         .   *                          .

SalB.                          -DDESGKSPEDIEQARRRASTRRSVVRGHSRSPSTRTDPTVKNTPASNGGPHP      632
fuscachelinPCP_7KW3            -QPVTEGS-----------------------------------------------      81
eponemycin_epnG_PCP            -D-----------------------------------------------------      67
daptomycin_dptA-PCP1           -DDN---------------------------------------------------      69
cyclomarin_Sare_4562_PCP3.     -L-----------------------------------------------------      67
BlmI_PCP-4NEO                  -HATLTGERDR--------------------------------------------      93
1DNY_1|Chain                   -ATRSHHHHHH--------------------------------------------      91
2N5H_1|Chain                   LKAESALE-----------------------------------------------      90
5U3H_1|Chain                   -RKTDVPVEQPFVHSPED-------------------------------------     111
2MR8_1|Chain                   -AAKSASWSHPQFEK----------------------------------------      91
6LK3_1|Chains                  -R----------EMPRLE-------------------------------------     108
5JDX_1|Chain                   LKAHPEQGGAALE------------------------------------------      89
```

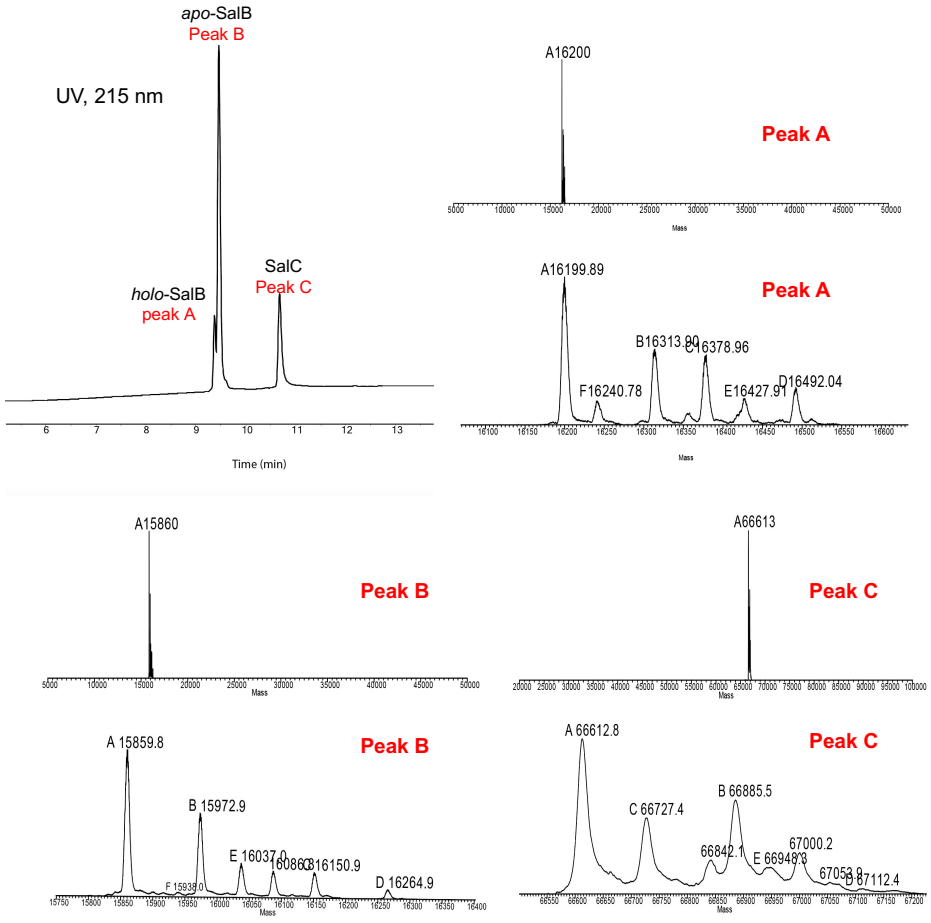
b

Protein name	Amino acid sequence
SalB-PCP domain truncation	MTDTEQALAEIWGALL DRTPQSIGIRDTFYDL GGNSLLLVRLAKRMG QRFHRKVGVADLFRF RDIGSLAKWLDDESGK SPEDIEQARRRASTRR SVVRGHSRSPSTRTD PTVKNTPASNGGPHP

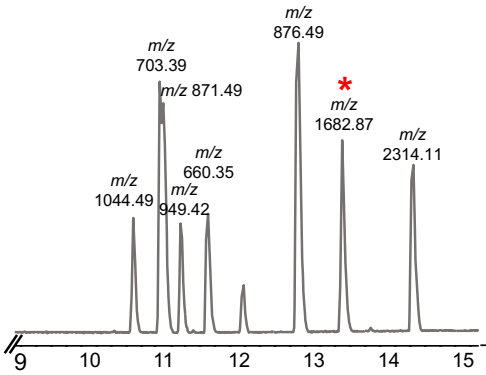
c



Supplementary Fig. 7: SalB-PCP expression and purification. **a,** Truncated alignment of SalB didomain with known PCP domains, particularly those with crystal structures, to determine SalB PCP excision. **b,** Amino acid sequence of SalB PCP domain truncation used for expression and purification. **c,** SDS-PAGE (15% acrylamide) analysis of purified SalB-PCP. EZ-Run™ Rec protein Ladder (Fisher Scientific) was run alongside with the samples, expected protein size ~16 kDa. P = pellet, SN = supernatant, W = wash, FT = flowthrough. Protein was purified with the same results six times.

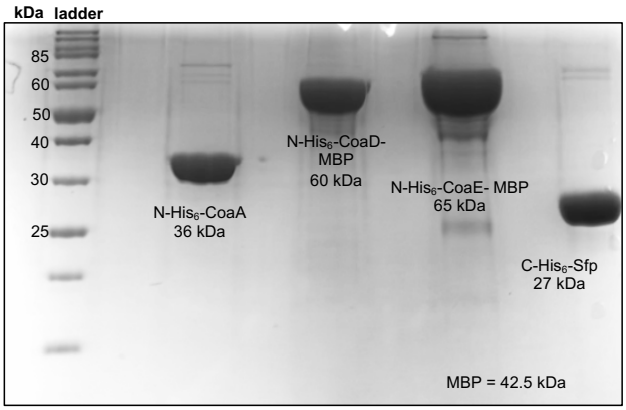


Supplementary Fig. 8: Intact proteomics. Intact protein LCMS chromatograms of purified SalB-PCP (predominantly in *apo* form) and SalC. Related to **Figs. 3** and **4.**

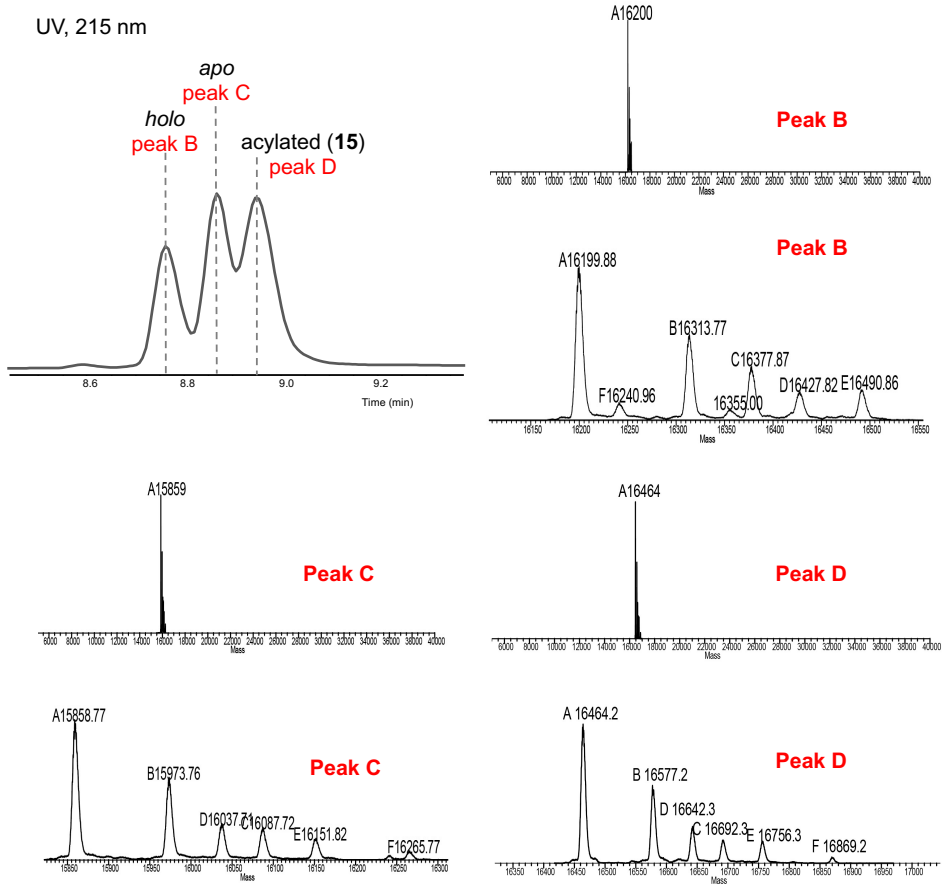


Mass (*m/z*)	Position	Peptide sequence
2314.11	14-34	GSHMTDTEQALAEIWGALLDR
1682.87	43-57	DTFYDLGGN**S**LLLVR
1537.76	1-13	HHHHHHSSGLVPR
1048.48	130-140	NTPASNGGPHP
1044.49	95-103	SPEDIEQAR
949.42	87-94	WLDDESGK
876.49	70-77	VGVADLFR
871.50	35-42	TPQSIGIR
703.39	80-86	DIGSLAK
660.36	124-129	TDPTVK
547.28	119-123	SPSTR

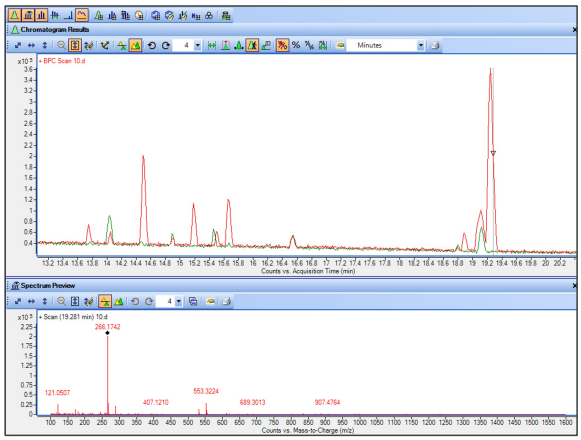
Supplementary Fig. 9: **Tryptic fingerprinting SalB-PCP**. Table depicts predicted peptides determined by *in silico* trypsin digest of the SalB-PCP amino acid sequence. LCMS chromatogram of peptides produced by *in vitro* trypsin digest of SalB-PCP. Bold red S indicates active site serine where phosphopantetheinylation takes place, red asterisk indicates peptide containing active site serine.



Supplementary Fig. 10: CoA and Sfp protein expression and purification. SDS-PAGE (12% acrylamide) analysis of purified proteins. EZ-Run™ Rec protein Ladder (Fisher Scientific) was run alongside with the samples, expected protein sizes are indicated. Proteins were purified with the same results three times.

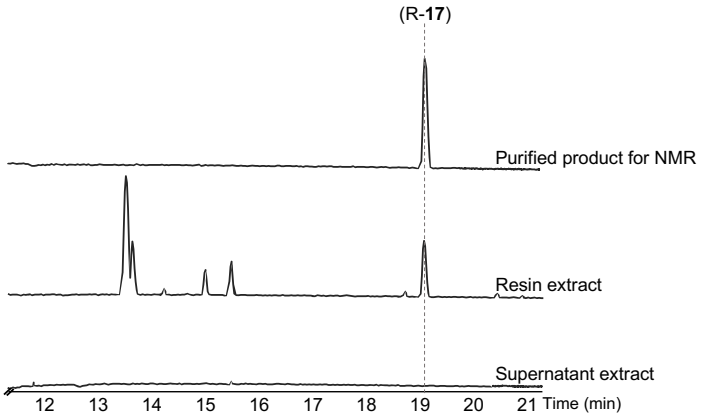


Supplementary Fig. 11: Intact proteomics. Intact SalB-PCP protein LCMS chromatograms for SalB-PCP acylation assay with **13**, refer to **Figure 3b.**

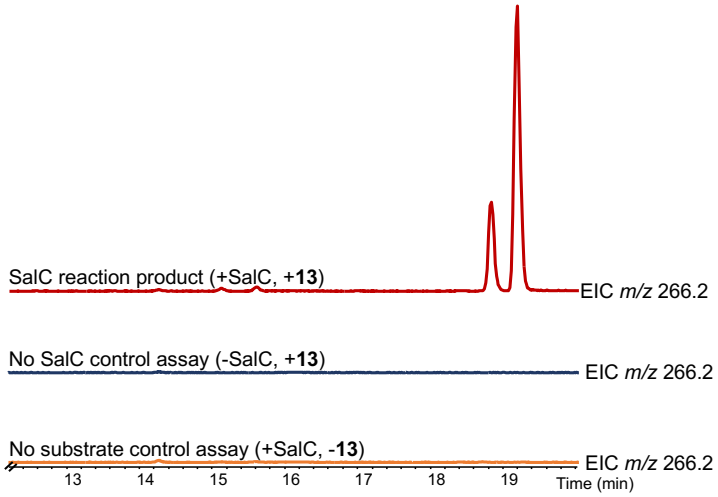


Theoretical (m/z)	Observed (m/z)	Mass error (ppm)
266.1751	266.1742	-3.38

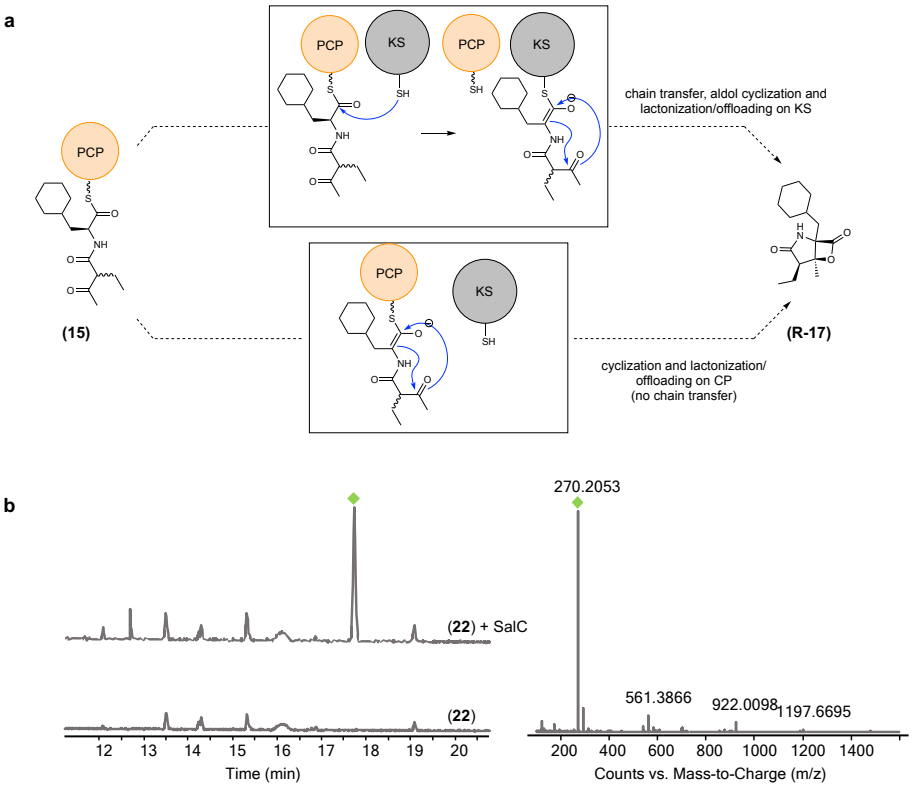
Supplementary Fig. 12: Simplisporamide characterization. High resolution mass spectrometry data for compound **R-17,** refer to **Figure 3e.**



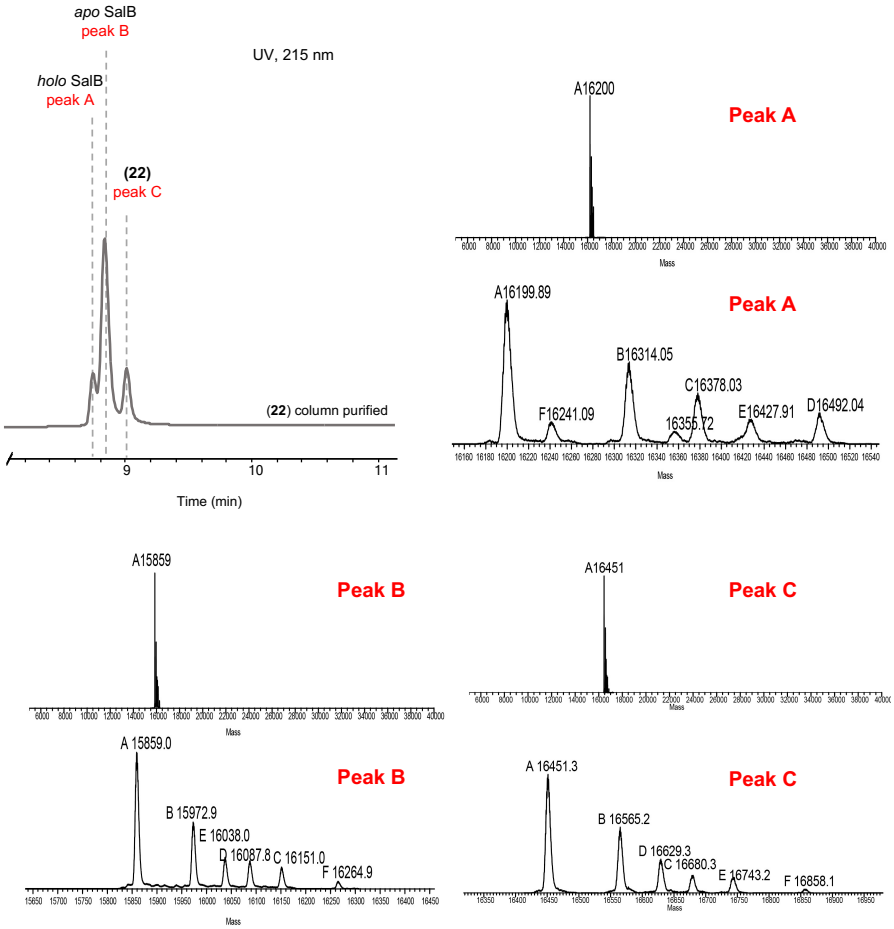
Supplementary Fig. 13: SalC preparative scale assay. LCMS chromatograms of extracts of 50 mL SalC activity assay supernatant, resin, and purified product **(R-17)** ultimately analyzed via NMR.



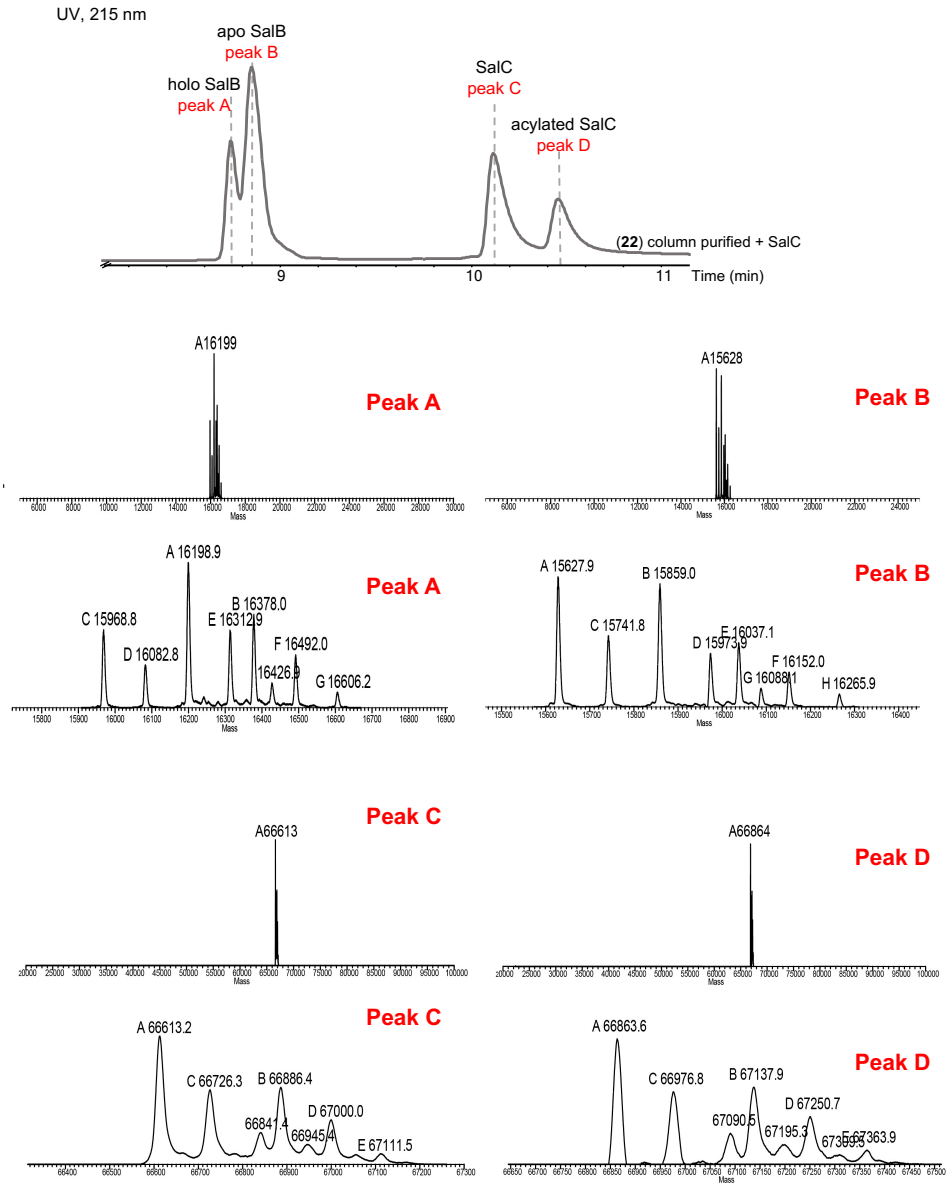
Supplementary Fig. 14: Analysis of SalC activity assays for proteasome inhibition assay. Extracted ion chromatograms (*m/z* 266.2) corresponding to simplisporamide (**R-17**) of resin extracts of SalC activity assays (+/-SalC, +/-**13**) used for proteasome inhibition assays. See **Extended Data Fig. 4**



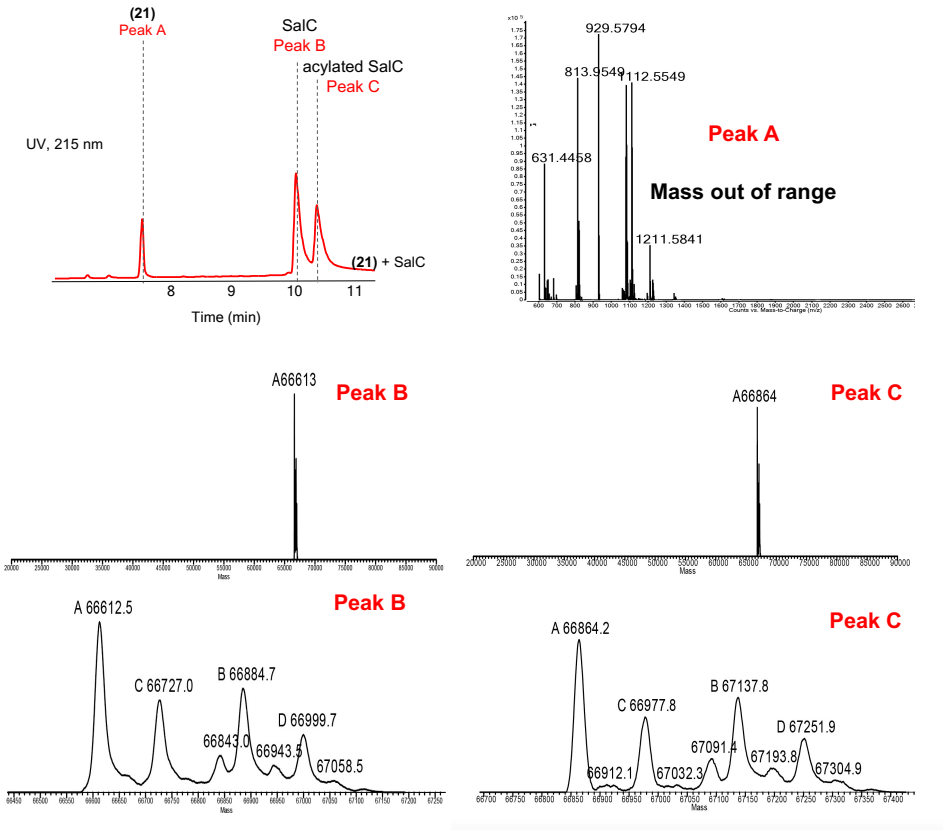
Supplementary Fig. 15: Chain transfer from SalB-PCP to SalC. a, Chain transfer schematic showcasing two options for cyclization via SalC: either linear substrate remains on SalB-PCP for cyclization or is transferred to SalC prior to cyclization. **b,** LCMS chromatograms of **(22)** in buffer and **(22)** + SalC and *m/z* of reaction product (green diamond). When SalC is present the mechanistic probe **(21)** is hydrolyzed off the carrier protein.



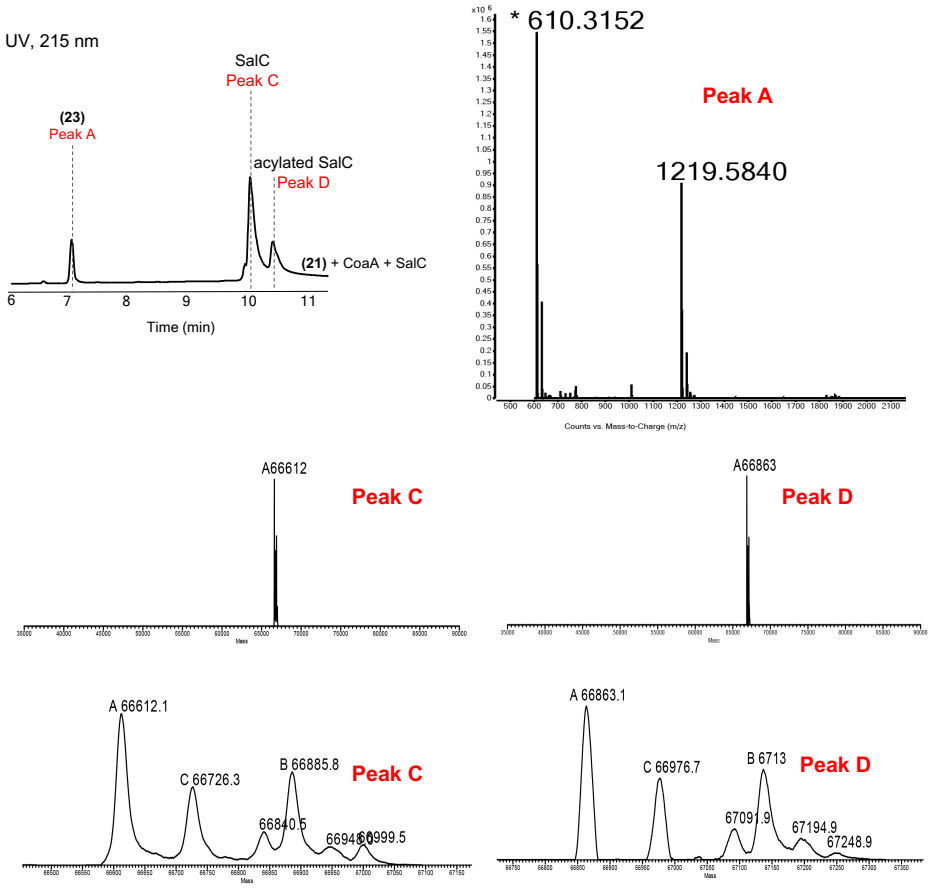
Supplementary Fig. 16: Intact proteomics. Intact protein LCMS chromatograms of column purified **22**, refer to **Figure 4b.**



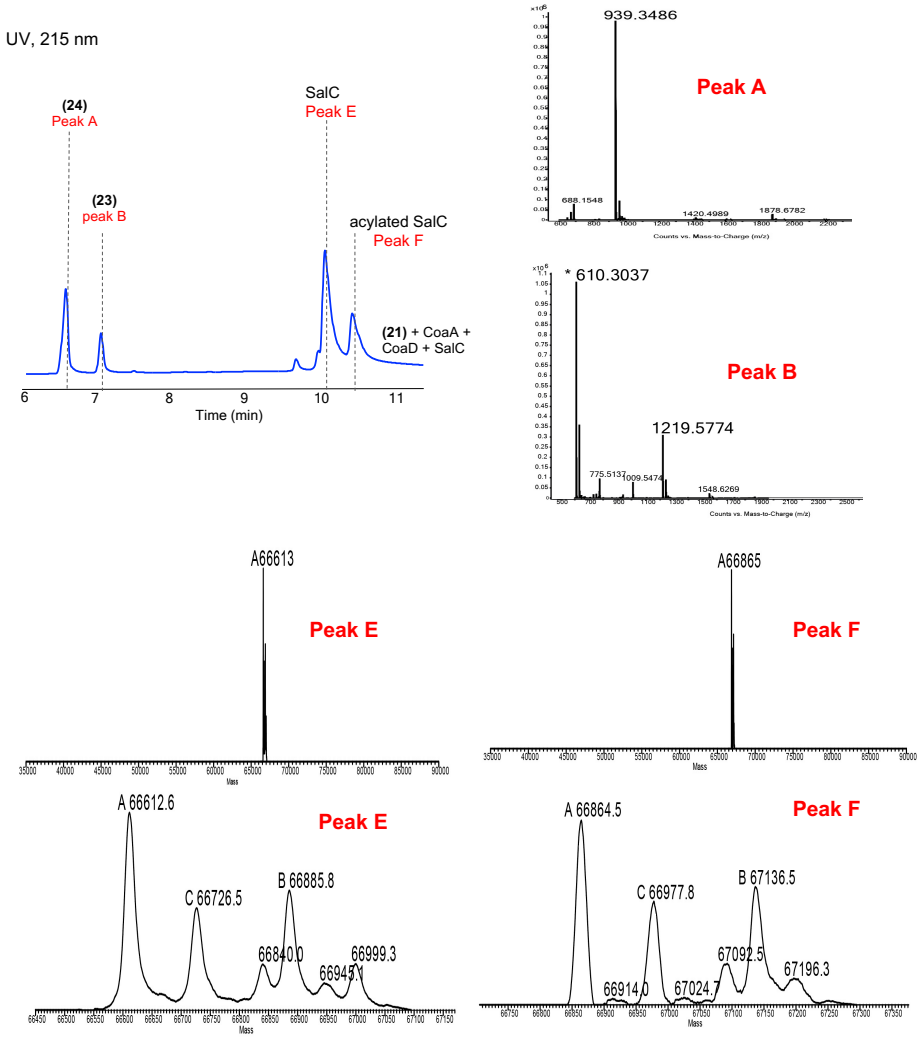
Supplementary Fig. 17: Intact proteomics. Intact protein LCMS chromatograms of column purified **22** + SalC, refer to **Figure 4b.**



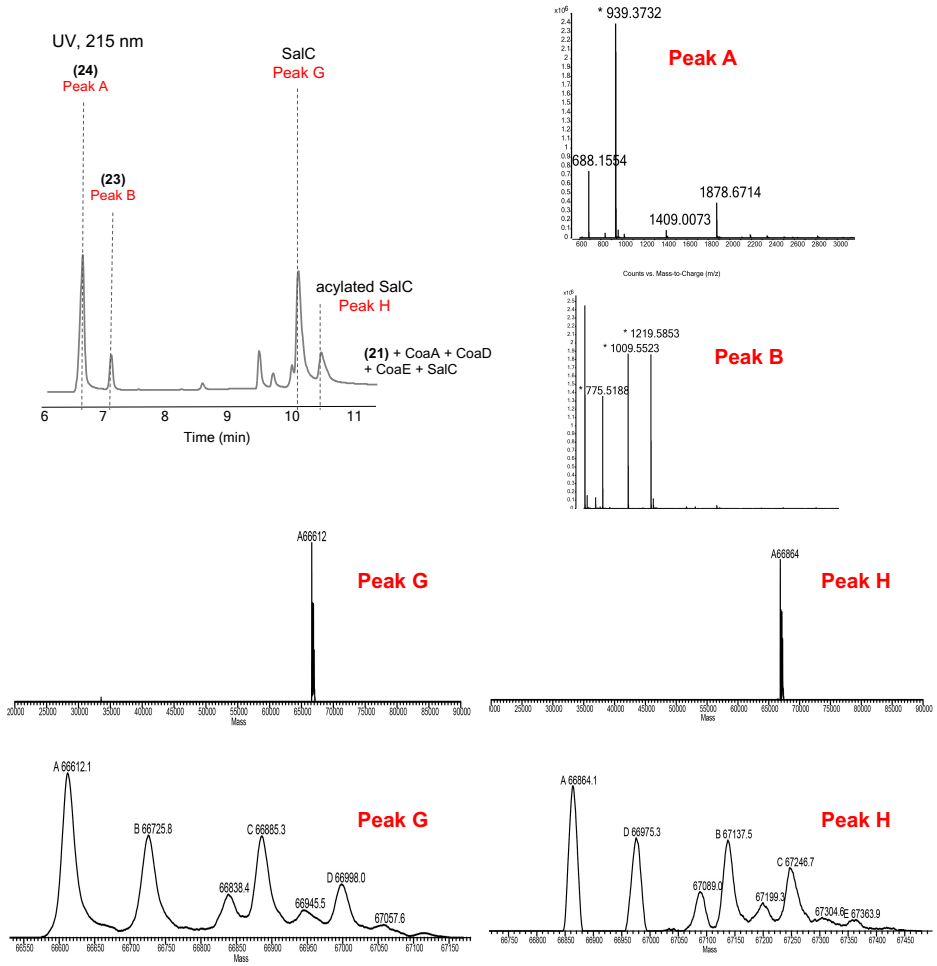
Supplementary Fig. 18: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with diffusible substrate **21**, refer to **Extended Data Fig. 5.**



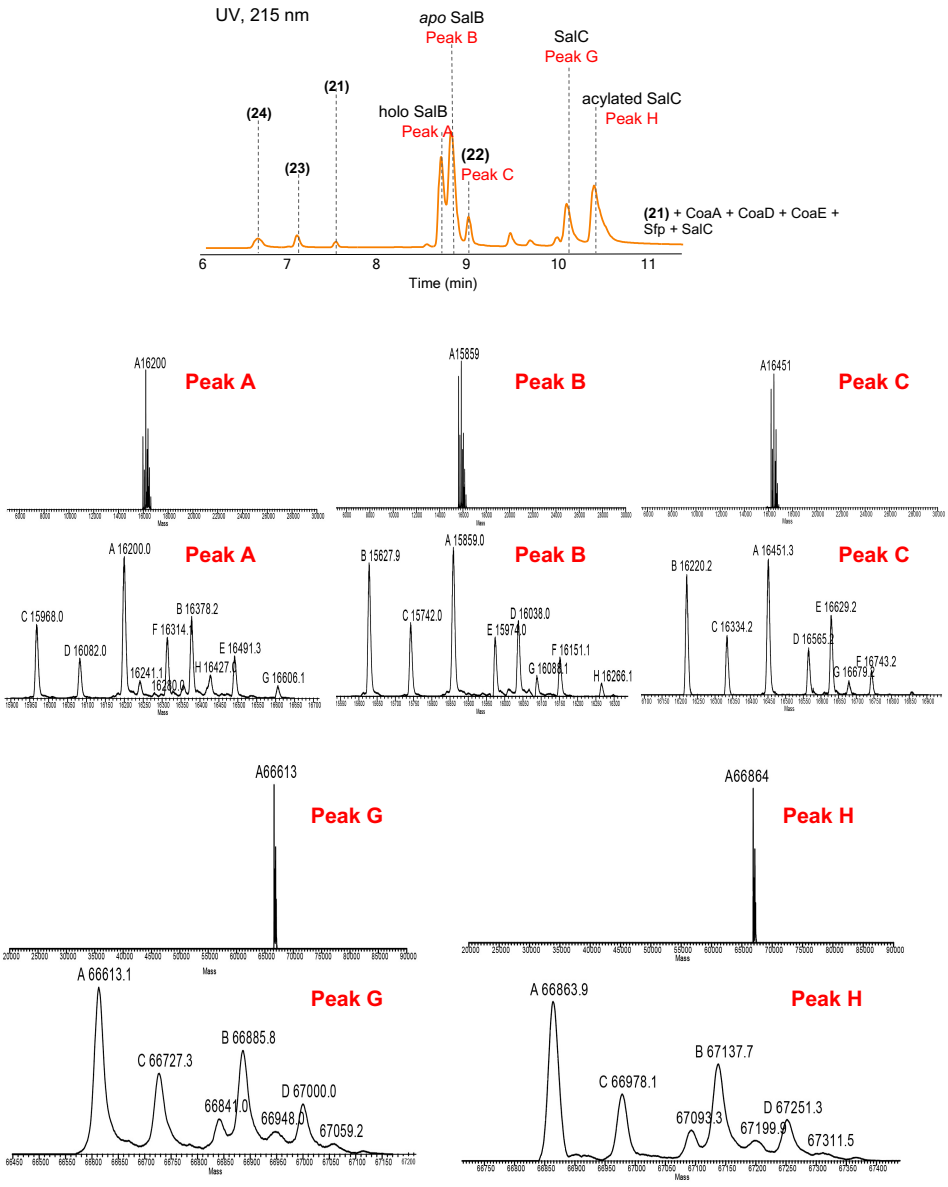
Supplementary Fig. 19: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with diffusible substrate **21** + CoaA (**23** made *in vitro*)**,** refer to **Extended Data Fig. 5.**



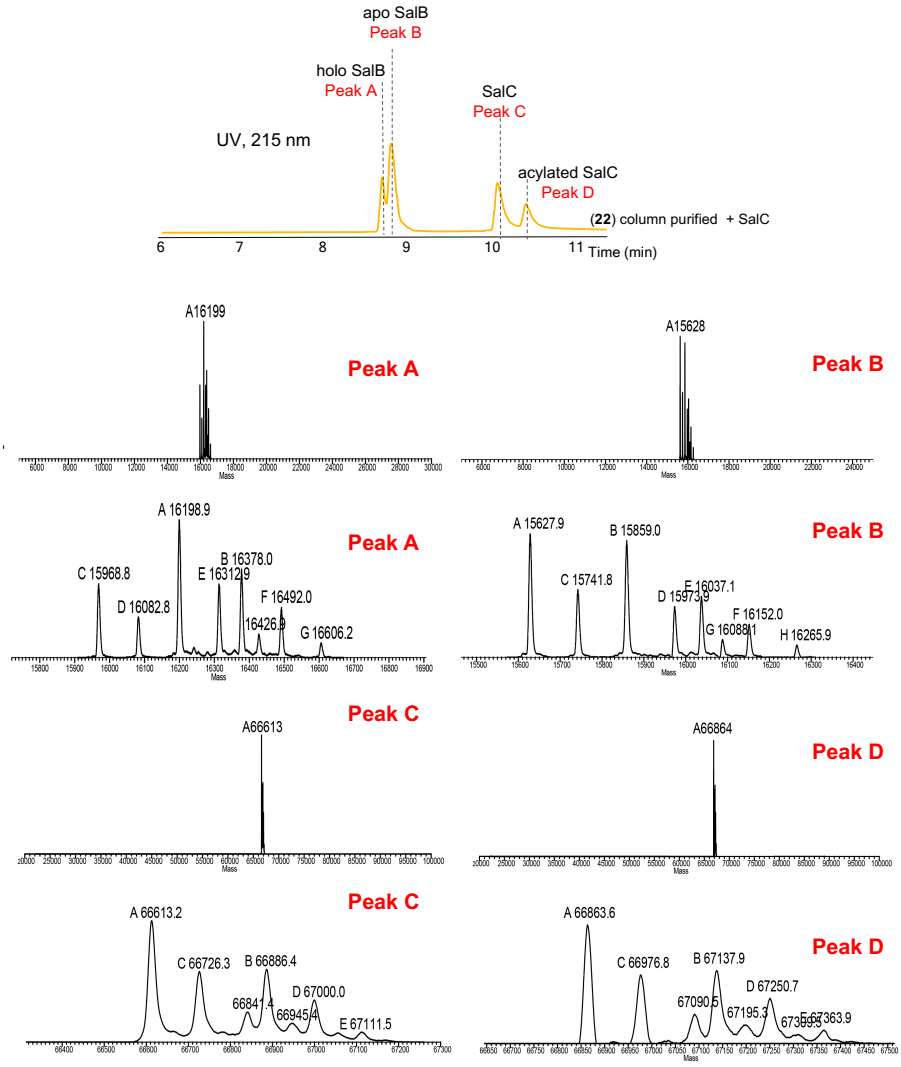
Supplementary Fig. 20: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with diffusible substrate **21** + CoaA + CoaD (**24** made *in vitro*) refer to **Extended Data Fig. 5.**



Supplementary Fig. 21: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with diffusible substrate **21** + CoaA + CoaD + CoaE (**25** made *in vitro*) refer to **Extended Data Fig. 5.**

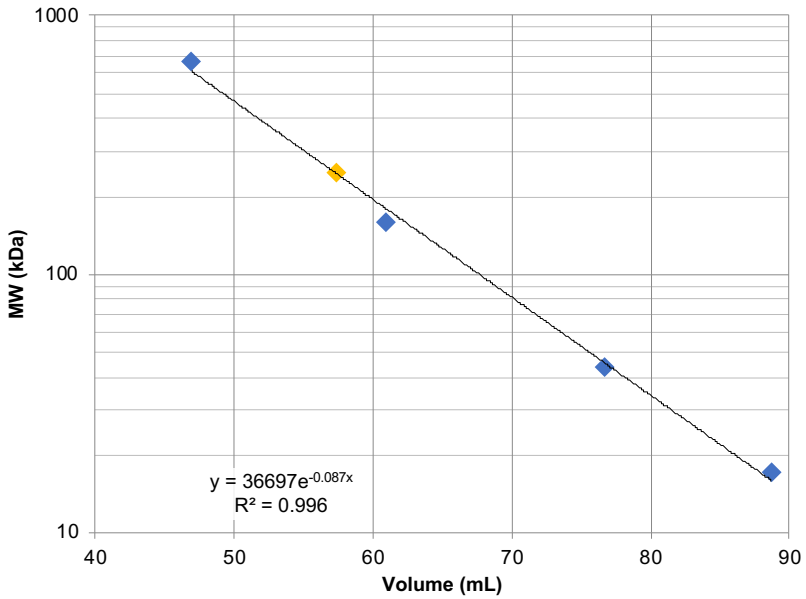


Supplementary Fig. 22: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with **21** + CoaA + CoaD + CoaE +Sfp (**22** made *in vitro*) refer to **Extended Data Fig. 5.**



Supplementary Fig. 23: Intact proteomics. Intact protein LCMS chromatograms for SalC acylation assay with column purified **22,** refer to **Extended Data Fig. 5.**

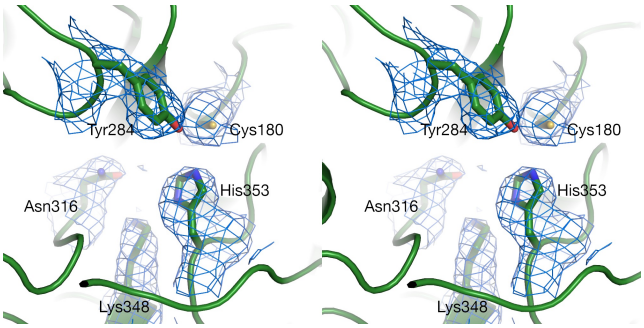
a



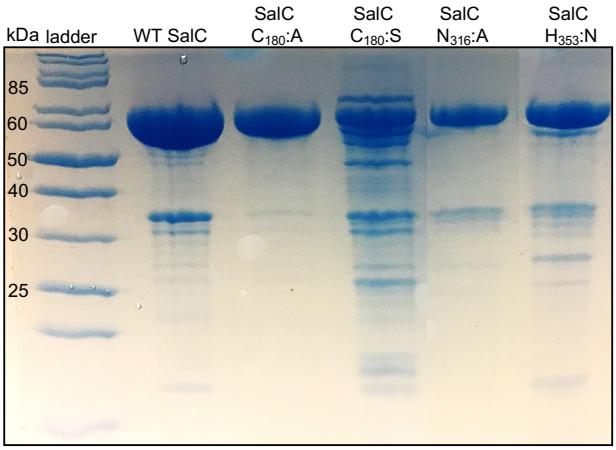
b

		MW (kDa)	Volume (mL)	Differences (mL)
Theoretical	SalC Dimer	132	64.47	-7.07
	SalC Tetramer	264	56.53	0.87
Experimental	SalC Experimental	245	57.40	0.0

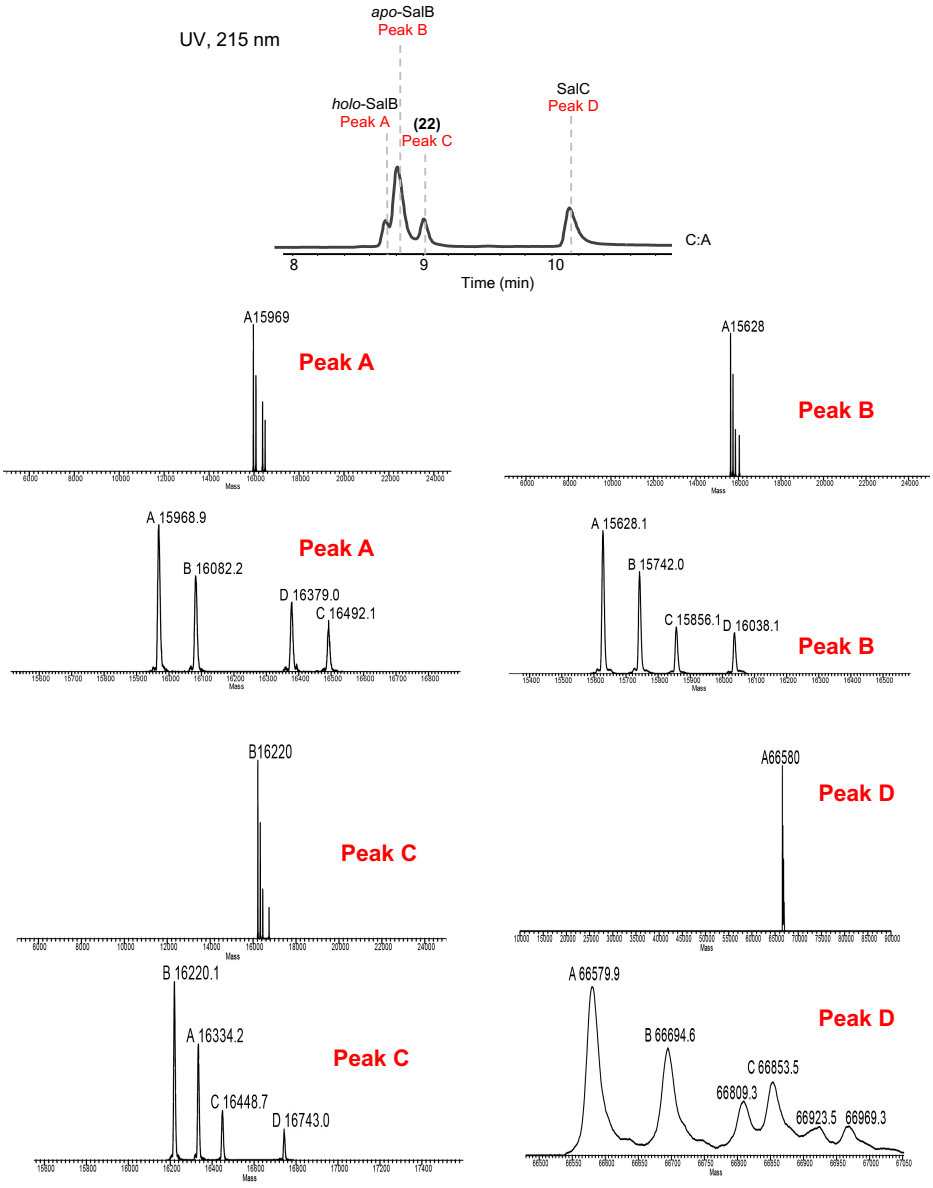
Supplementary Fig. 24: Determining the oligomeric state of SalC. **a,** Size exclusion standard curve with HiLoad Superdex 200 prep grade gel filtration column (16 cm x 60 cm, GE Healthcare) using BioRad Gel Filtration Standard set (blue diamonds; myoglobin = 17 kDa, ovalbumin = 44 kDa, γ-globulin = 158 kDa, thyroglobulin = 670 kDa) and SalC sample (orange diamond). **b,** SalC theoretical and experimental elution volumes.



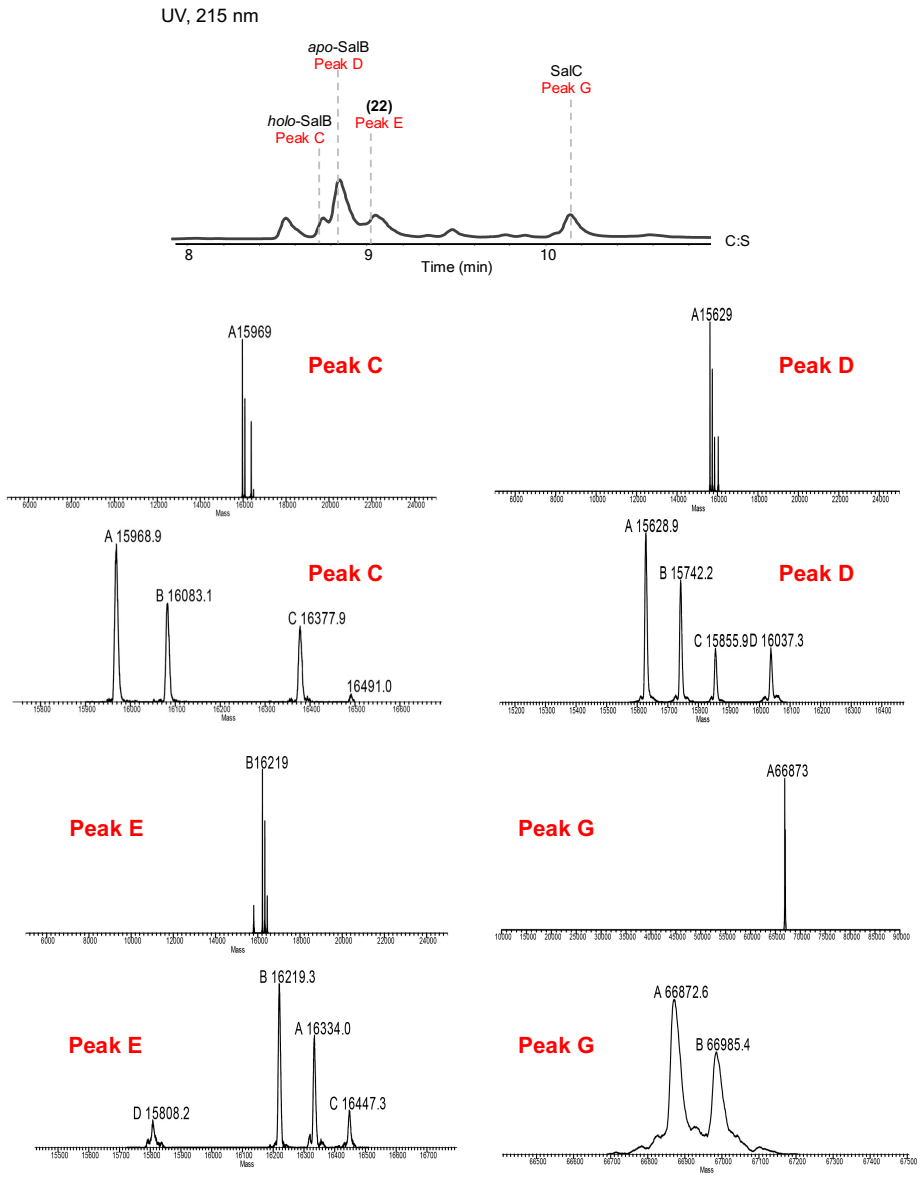
Supplementary Fig. 25: Stereoview showing the active site residues of SalC. The composite-omit electron density map (blue meshes) generated using coefficients of the form 2mFo-DFc for the active site residues (sticks) is contoured at 1.0 σ; protein scaffolds are shown as green ribbons.



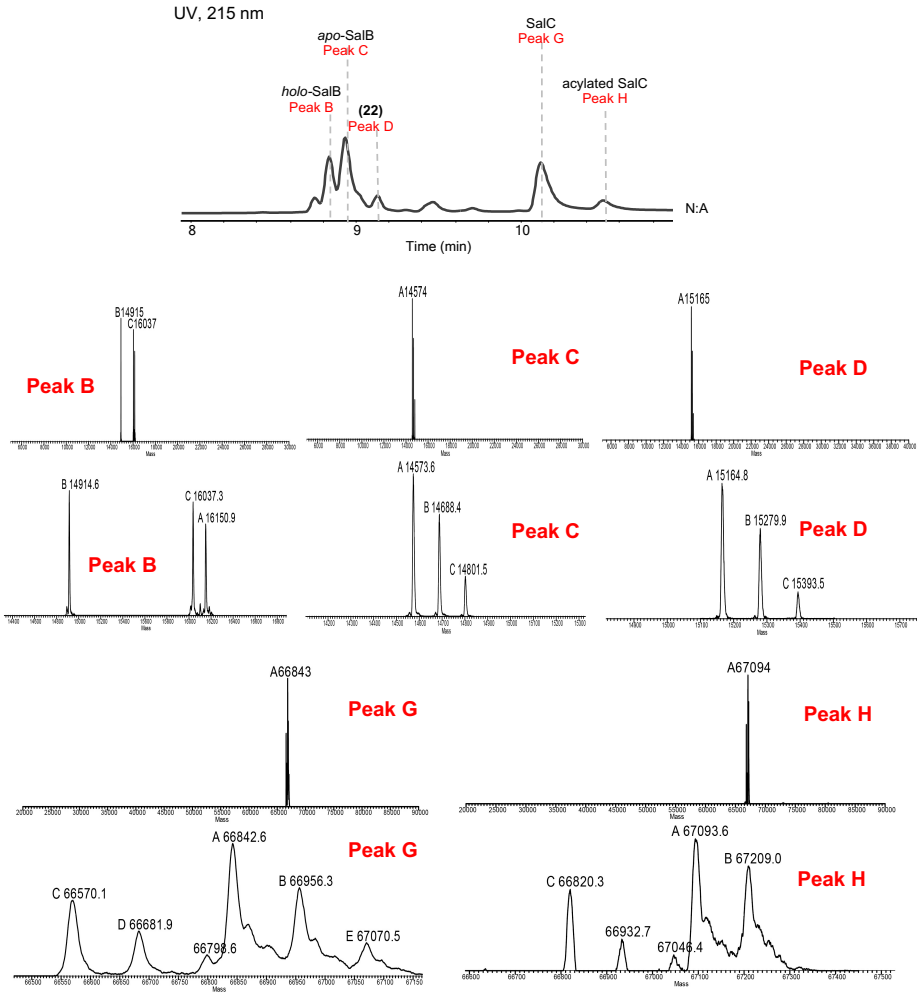
Supplementary Fig. 26: SalC variant protein expression and purification. SDS-PAGE (12% acrylamide) analysis of purified SalC variants from *Streptomyces coelicolor* CH999. EZ-Run™ Rec protein Ladder (Fisher Scientific) was run alongside with the samples. Proteins were purified with the same results two times. See **Fig. 5** and **Supplementary Fig. 27-30**.



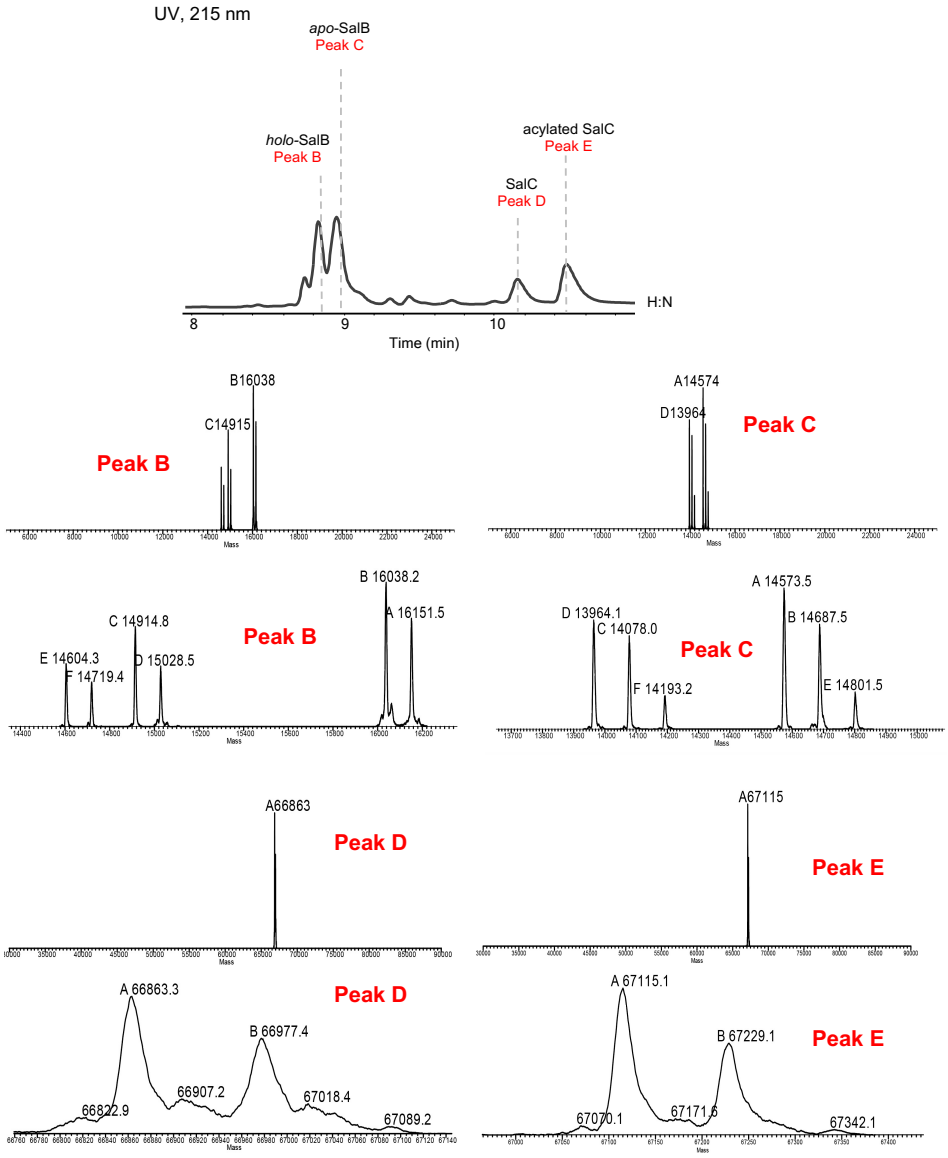
Supplementary Fig. 27: Intact proteomics. Intact protein LCMS data of transacylation assay with column purified **22** + SalC-Cys180Ala, refer to **Figure 5c.**



Supplementary Fig. 28: Intact proteomics. Intact protein LCMS chromatograms of transacylation assay with column purified **22** + SalC-Cys180Ser, refer to **Figure 5c.**



Supplementary Fig. 29: Intact proteomics. Intact protein LCMS chromatograms of transacylation assay with column purified **22** + SalC-Asn316Ala, refer to **Figure 5c.**



Supplementary Fig. 30: Intact proteomics. Intact protein LCMS chromatograms of transacylation assay with column purified **22** + SalC-His353Asn, refer to **Figure 5c.**

Supplementary Tables

Supplementary Table 1. Strains and plasmids used in this work

Strains	Genotype/Description	Source
Salinispora tropica CNB440	Native producing organism of the salinosporamides	Ref #2
S. tropica CNB440 *salC::aac(3)IV*	Salinosporamide producing organism with apramycin resistance gene in place of *salC*	This study
Escherichia coli DH10B	F-,*endA1*, *recA1*, *galE15*, *galK16*, *nupG*, *rpsL*, Δ*lacX74*, Φ80, *lacZ*ΔM15, *araD139*, Δ(*ara*, *leu*)7697, *mcrA*, Δ(*mrr-hsdRMS-mcrBC*), λ- ; Storage and maintenance	Invitrogen
E. coli BL21(DE3)Gold	F-, *ompT*, *gal*, *dcm*, *lon*, *hsdSB*(rB- mB-), λ(DE3 [*lacI* *lac*UV5-T7 gene1, *ind*1, *sam*7, *nin*5]); Host for gene expression and protein purification	Novagen
E. coli ET12567	F-,*dam*-13::Tn9, *dcm-6*, *hsdM*, *hsdR*, *rec*F143, *zjj*-202::Tn10, *galK*2, *gal*T22, *ara*14, *pac*Y1, *xyl*-5, *leu*B6, *thi*-1, *ton*A31, *rps*L136, *his*G4, *tsx*-78, *mtl*-1, glnV44. Donor strain for conjugation	Ref #52
Streptomyces coelicolor CH999	Heterologous host for protein overproduction, Δ*act*, *redE*-, *pro*, *arg*	Ref #54
Plasmids	**Genotype/Description**	**Source**
pET28b(+)	*neo*, P_{T7}, ori^{pBR322}, *lac*I, N- and C-terminal His_6-tag, oriF1	Novagen
pET28b(+)-CoaA	pET28b-derived, *coaA* gene cloned with N-terminal His_6-tag	Ref #33
pET28MBP	pET28b-derived, with additional Maltose Binding Protein tag	Novagen
pET28MBP-CoaD	pET28MBP derived, *coaD* gene cloned with N-terminal His_6-MBP tag	Ref #33
pET28MBP-CoaE	pET28MBP derived, *coaE* gene cloned with N-terminal His_6-MBP tag	Ref #33
pET24a(+)	*neo*, P_{T7}, ori^{pBR322}, *lac*I, C-terminal His_6-tag, N-terminal T7 tag, oriF1	Novagen
pET24a(+)-Sfp	pET24a(+)-derived, *sfp* gene cloned with C-terminal His_6-tag	Ref #33
pET28a(+)	*neo*, P_{T7}, ori^{pBR322}, *lac*I, N- and C-terminal His_6-tag, oriF1	Novagen
pET28a(+)-SalB(PCP)	pET28a-derived, with truncated *salB* PCP domain cloned with N-terminal His_6-tag	This study
pETDUET-1	*bla*, PT7, ori^{pBR322}, *lac*I, oriF1, N-terminal MCS1 His_6-tag	Novagen
pETDUET-1-SalB-SalE	pETDUET-1-derived with *salB* gene in MCS1 with N-terminal His_6-tag, and *salE* gene in MCS2	This study
pCJW93	*aac(3)IV*, *tsr*, *Hcn* replicon, ori^{pIJ101}, *tipAp*, ori^{pUC}, *oriT;* Heterologous protein production in *Streptomyces*	Ref #54
pCJW93-SalC	*salC* gene cloned with N-terminal His_6-tag	This study
pCJW93-SalF	*salF* gene cloned with N-terminal His_6-tag	This study
pUB307	Self-transmissible plasmid, mobilizes other plasmids *in trans* for DNA transfer into hosts: RP4, *neo*	Ref #53
pCJW93-SalC-Cys_{180}:Ala	*salC* gene cloned with N-terminal His_6-tag with Cys_{180}:Ala mutation	This study
pCJW93-SalC-Cys_{180}:Ser	*salC* gene cloned with N-terminal His_6-tag with Cys_{180}:Ser mutation	This study
pCJW93-SalC-Tyr_{284}:Phe	*salC* gene cloned with N-terminal His_6-tag with Tyr_{284}:Phe mutation	This study
pCJW93-SalC-Asn_{316}:Ala	*salC* gene cloned with N-terminal His_6-tag with Asn_{316}:Ala mutation	This study
pCJW93-SalC-His_{353}:Asn	*salC* gene cloned with N-terminal His_6-tag with His_{353}:Asn mutation	This study

Supplementary Table 2. Primers used in this work. Restriction sites shown in bold, vector homology in red, gene homology in black, mutagenesis sites in gray highlight and italic.

Primer name	Sequence (5' → 3')	purpose
CoaA/D/E/Sfp	As described previously in Agarwal *et al*., 2015	
SalB-PCP-pET28-F	CCTGGTGCCGCGCGGCAGC**CATATG**ACCGACACCGAACAAGCGCTTGC	Cloning SalB-PCP
SalB-PCP-pET28-R	GGTGGTGGTGGTG**CTCGAG**TCATGGATGTGGGCCTCCATTTGATG	Cloning SalB-PCP
SalC-pCJW93-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	Cloning SalC
SalC-pCJW93-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGTC	Cloning SalC
C180A-frag1-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	SalC mutagenesis
C180A-frag1-R	GTGAAAGCCCGAG*GCC*ACGGCGGTCACG	SalC mutagenesis
C180A-frag2-F	CGTGACCGCCGT*GGC*CTCGGGCTTTCAC	SalC mutagenesis
C180A-frag2-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGTC	SalC mutagenesis
C180S-frag1-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	SalC mutagenesis
C180S-frag1-R	GAAAGCCCGA*GCT*CACGGCGGTCAC	SalC mutagenesis
C180S-frag2-F	GTGACCGCCGTG*AGC*TCGGGCTTTC	SalC mutagenesis
C180S-frag2-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGTC	SalC mutagenesis
Y284F-frag1-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	SalC mutagenesis
Y284F-frag1-R	GCGAGGCCG*AAC*GCCATGGCACTGATG	SalC mutagenesis
Y284F-frag2-F	CATCAGTGCCATGGCG*TTC*GGCCTCGC	SalC mutagenesis
Y284F-frag2-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGT	SalC mutagenesis
N316A-frag1-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	SalC mutagenesis
N316A-frag1-R	GGCATGCCGAACCC*AGC*GGCCTCGTACATGGA	SalC mutagenesis
N316A-frag2-F	TCCATGTACGAGGCC*GCT*GGGTTCGGCATGCC	SalC mutagenesis
N316A-frag2-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGTC	SalC mutagenesis
H316N-frag1-F	CAGCGGCCTGGTGCCGCGCGGCAGC**CATATG**AGCGAAGCCGTTGAGCG GAACTACG	SalC mutagenesis
H316N-frag1-R	CCACGCCG*TTC*CCGATATTGCC	SalC mutagenesis
H316N-frag2-F	GGCAATATCGGG*AAC*GCCGGCGTGG	SalC mutagenesis
H316N-frag2-R	CTTTAGATCTGGGGAATTCGGATCC**AAGCTT**CACCAGTCGGTGCGCCAGA ACCGTC	SalC mutagenesis

Supplementary Table 3. Data collection and model refinement statistics of SalC

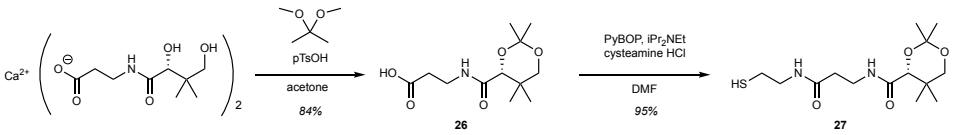
	SalC
Data collection	
Space group	P 1 21 1
Cell dimensions	
a, *b*, *c* (Å)	84.33, 218.29, 87.87
α, β, γ (°)	90, 111.014, 90
Resolution (Å)	45.44-2.85 (3.02-2.85)
R_{sym} or R_{merge}	0.078 (0.404)
$I / \sigma I$	11.32 (2.70)
Completeness (%)	96.8 (95.0)
Redundancy	2.3 (2.4)
Refinement	
Resolution (Å)	44.85-2.85 (2.95-2.85)
No. reflections	66862 (6381)
R_{work} / R_{free}	19.7/23.0
No. atoms	
Protein	17948
Ligand/ion	19
Water	19
B-factors	
Protein	50.02
Ligand/ion	49.2
Water	41.7
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.600

*Values in parentheses are for highest-resolution shell.

Supplementary Table 4. Residues modeled in each chain of SalC (residues 1-597)

	SalC
Chain A	9-597
Chain B	11-597
Chain C	11-597
Chain D	11-597

Supplementary Notes Fig. 1: Synthesis of pantetheine acetonide (**27**).



(*R*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (**26**): Calcium (*R*)-pantothenate (5.0 g, 21 mmol) was suspended in 1M HCl (100 mL) and was stirred until completely dissolved. Resulting solution was saturated with NaCl, extracted with EtOAc (3 × 100 mL), dried over $MgSO_4$, filtered and solvent was removed in vacuo. Resulting colorless oil (4.25 g, 19.4 mmol, 1.0 equiv.) was dissolved in acetone (100 mL, 0.2 M), and cooled to 0 ºC with stirring. *p*-Toluenesulfonic acid (185 mg, 1 mmol, 0.05 equiv.) was added as a solid in a single portion and headspace of reaction was purged with Ar. 2-methoxypropene (5.5 mL, 58.4 mmol, 3.0 equv) was added dropwise and upon completion of addition reaction was stirred for an additional 15 minutes at 0 ºC. Reaction was then allowed to warm to room temperature and stir at r.t. for 1 hour before quenching with saturated $NaHCO_3$ solution (10 mL) and removal of solvent in vacuo to give a yellow solid (4.57 g, 84% yield) which was used without further purification.

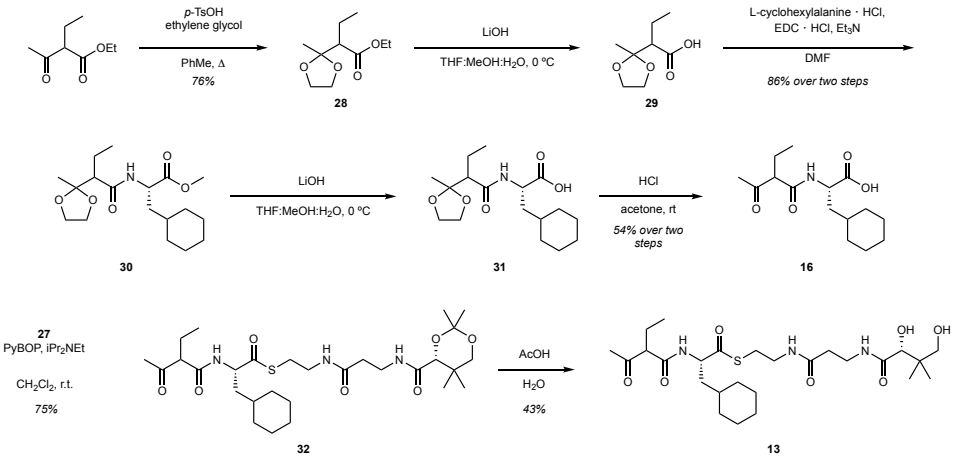
HRMS (ESI) (m/z) calculated for $C_{12}H_{22}NO_5^+$, $[M+H]^+$: 260.1492, observed: 260.1487.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.14 – 6.96 (m, 1H), 4.10 (s, 1H), 3.69 (d, *J* = 11.7 Hz, 1H), 3.66 – 3.54 (m, 1H), 3.54 – 3.44 (m, 1H), 3.28 (d, *J* = 11.7 Hz, 1H), 2.62 (t, *J* = 6.2 Hz, 2H), 1.46 (s, 3H), 1.43 (s, 3H), 1.03 (s, 3H), 0.97 (s, 3H). NMR spectra can be found in **Supplementary Notes Fig. 4** and are in agreement with those previously reported in the literature (main text reference 32).

(*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (**27**): **26** (1.4 g, 5.3 mmol, 1.0 equiv.) was dissolved in anhydrous CH_2Cl_2 (30 mL, 0.2 M) and cysteamine HCl (1.2 g, 10.8 mmol, 2.0 equiv.), PyBOP (4.2 g, 8.0 mmol, 1.5 equiv.), and *i*Pr_2NEt (4.7 mL, 27.0 mmol, 5.0 equiv.) were added sequentially followed by purging reaction headspace with Ar. Reaction was allowed to stir for 12 hours at which point reaction was judged complete by TLC and solvent was removed in vacuo. Crude material was purified by MPLC (50%-100% hexanes:EtOAc) and product containing fractions were pooled to give **27** as an off-white solid (1.6 g, 95% yield).

HRMS (ESI) (m/z) calculated for $C_{14}H_{27}N_2O_4S^+$, $[M+H]^+$: 319.1686, observed: 319. 1677.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.01 (t, *J* = 6.2 Hz, 1H), 6.27 (s, 1H), 4.08 (s, 1H), 3.68 (d, *J* = 11.7 Hz, 1H), 3.63 – 3.52 (m, 2H), 3.50 – 3.37 (m, 2H), 3.28 (d, *J* = 11.7 Hz, 1H), 2.66 (dtd, *J* = 8.7, 6.5, 2.1 Hz, 2H), 2.47 (td, *J* = 6.1, 2.2 Hz, 2H), 1.46 (s, 3H), 1.42 (s, 3H), 1.04 (s, 3H), 0.97 (s, 3H). NMR spectra can be found in **Supplementary Notes Fig. 5** and are in agreement with those previously reported in the literature (main text reference 32).



Supplementary Notes Fig. 2: Synthesis of pantetheine-activated simplisporamide linear precursor (**13**).

Ethyl 2-(2-methyl-1,3-dioxolan-2-yl)butanoate (**28**): Ethyl-2-ethyl acetoacetate (3 mL, 19 mmol, 1 equiv.), was dissolved in anhydrous toluene (15 mL, 1.3 M) followed by addition of *p*-toluenesulfonic acid monohydrate (40 mg, 0.2 mmol, 0.1 equiv.) and ethylene glycol (2.65 mL, 38 mmol, 2 equiv.). The reaction fitted with a jacketed condenser and heated to reflux for 4 hours. After cooling to room temperature the biphasic crude material was diluted in saturated $NaHCO_3$ and extracted (3 × 100 mL) with EtOAc. EtOAc extracts were pooled, dried with $MgSO_4$, filtered and concentrated in vacuo to give **28** as a colorless oil which was used with no additional purification (2.9 g, 76% yield).

GC-MS (EI) analysis of **28** resulted exclusively in degradation.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.17 (dtd, *J* = 7.9, 7.1, 6.6, 2.8 Hz, 2H), 4.05 – 3.88 (m, 4H), 2.55 (dd, *J* = 11.4, 3.6 Hz, 1H), 1.81 – 1.70 (m, 1H), 1.64 (dqd, *J* = 14.7, 7.5, 3.5 Hz, 1H),

1.39 (d, *J* = 1.1 Hz, 3H), 1.27 (t, *J* = 7.1 Hz, 3H), 0.89 (t, *J* = 7.3 Hz, 3H). NMR spectra can be found in **Supplementary Notes Fig. 6**.

Methyl (2*S*)-3-cyclohexyl-2-(2-(2-methyl-1,3-dioxolan-2-yl)butanamido)propanoate (**30**): **28** (2.9 g, 14.3 mmol, 1 equiv.) was dissolved in a 2:1 mixture of THF:MeOH (45 mL, 0.3 M) and chilled to 0 ºC. LiOH (342 mg, 14.3 mmol, 1 equiv.) was added as a solution in deionized water, and after 2 hours reaction was judged as complete by TLC and volatile solvents were removed in vacuo. Residue was diluted with 1M HCl and extracted with CH_2Cl_2 (3 × 50 mL). Combined organic layers were dried over $MgSO_4$, filtered, and concentrated in vacuo to give crude acid **29** as slightly yellow oil (2.74 g) which was used directly with no additional purification.

L-cyclohexylalanine HCl (3.2 g, 14.3 mmol, 1.0 equiv.) was suspended in anhydrous CH_2Cl_2 (50 mL, 0.3 M) with stirring under an Ar atmosphere, and *i*Pr_2NEt (2 mL, 21.5 mmol, 1.5 equiv.) was added and the mixture was allowed to stir for 20 minutes during which time a white precipitate formed. **29** (2.74 g) was added as a solution in anhydrous CH_2Cl_2 followed by PyBOP (8 g, 15.73 mmol, 1.1 equiv.) which was added as a single solid addition followed by a purge of the reaction headspace with Ar. After 12 hours reaction was judged complete by TLC, quenched with 1M HCl and extracted with CH_2Cl_2 (3 × 50 mL). Organic extracts were combined, dried over $MgSO_4$, filtered and concentrated in vacuo. Crude product was purified via MPLC (0-100% hexanes:EtOAc) and product containing fractions were combined and solvent removed in vacuo to give **30** as a white solid and mixture of diastereomers (4.21 g, 86% yield over two steps).

HRMS (ESI) (m/z) calculated for $C_{18}H_{32}NO_5^+$, $[M+H]^+$: 342.2275 observed: 342.2274.

^{1}H NMR (500 MHz, $CDCl_3$) δ 6.49 (d, *J* = 8.0 Hz, 1H), 4.65 (ddd, *J* = 9.6, 8.1, 5.1 Hz, 1H), 4.04 – 3.93 (m, 4H), 3.72 (s, 3H), 2.41 (t, *J* = 7.4 Hz, 1H), 1.83 – 1.76 (m, 1H), 1.76 – 1.60 (m, 9H), 1.56 (m, 1H), 1.34 (s, 3H), 1.29 – 1.11 (m, 3H), 0.94 (m, 4H). NMR spectra can be found in **Supplementary Notes Fig. 7**.

(2*S*)-3-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanoic acid (**16**): **30** (2 g, 5.9 mmol, 1 equiv.) was dissolved in a 2:1 mixture of THF:MeOH (18 mL, 0.3 M), and cooled to 0 ºC. LiOH (140 mg, 5.9 mmol, 1 equiv.) was added as a solution in H_2O, and the reaction was monitored by TLC. After 3 hours, reaction had reached completion and solvent was removed in vacuo. Residue was diluted with 1M $NaHCO_3$, and non-polar impurities were removed via Et_2O (1 × 40 mL) washes. Aqueous layer was then acidified to pH= ~3, and extracted with CH_2Cl_2 (3 × 50 mL). Combined organic layers were dried over $MgSO_4$, filtered and concentrated in vacuo to give crude acid **31** (1.7 g) as an off white solid.

31 (1.7 g) was dissolved in acetone (10 mL, 0.5 M) at followed by 1M HCl (10 mL) and reaction was allowed to stir for 1 hour at room temperature at which point reaction had reached completion as determined by TLC. Solvent was removed in vacuo and residue was purified by MPLC (hexanes:EtOAc, 2:1, isocratic) to give **16** as a white powder (898 mg, 54% yield over two steps).

HRMS (ESI) (m/z) calculated for $C_{15}H_{24}NO_4^-$, $[M-H]^-$: 282.1711 observed: 282.1707.

^{1}H NMR (500 MHz, $CDCl_3$) δ 6.79 (dd, *J* = 13.3, 7.9 Hz, 1H) 4.63 – 4.55 (m, 1H), 3.35 (q, *J* = 6.9 Hz, 1H), 2.25 (s, *J* = 2.4 Hz, 3H), 1.90 (dhept, *J* = 14.5, 7.2 Hz, 2H), 1.79 – 1.53 (m, 8H), 1.39 – 1.27 (m, 1H), 1.27 – 1.09 (m, 3H), 1.03 – 0.82 (m, 4H). NMR spectra can be found in **Supplementary Notes Fig. 8**.

S-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) (2*S*)-3-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanethioate (**32**): **16** (70 mg, 0.25 mmol, 1.2 equiv.) was dissolved in anhydrous CH_2Cl_2 (5 mL, 0.4 M) under inert Ar atmosphere, and **27** (66 mg, 0.21 mmol, 1.0 equiv.) was added as a solution in anhydrous CH_2Cl_2. PyBOP (128 mg, 0.25 mmol, 1.2 equiv.) was added as a single solid addition, followed by a purge of the reaction headspace with Ar and addition of *i*Pr_2NEt (70 mL, 0.41 mmol, 2.0 equiv.). Reaction was allowed to proceed for 18 hours at which time an aliquot was sampled by HPLC-MS and was judged complete with no observable **27** remaining. Solvent was removed *in vacuo* and residue was purified by MPLC (0% - 100%, hexanes: EtOAc). Product containing fractions were pooled and concentrated to give **32** mixture of diastereomers as a yellow oil (108 mg, 75% yield).

HRMS (ESI) (m/z) calculated for $C_{29}H_{50}N_3O_7S^+$, $[M+H]^+$: 584.3364, observed: 584.3351.

NMR spectra is heavily overlapped due to the equilibrating mixture of diastereomers and can be found in **Supplementary Notes Fig. 9**.

S-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(2*S*)-3-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanethioate (**13**): **32** (108 mg, 0.18 mmol) was dissolved in a premixed solution of 2:1 AcOH:H_2O (9 mL, 0.02 M) and solution was stirred at RT. After 5 hours, monitoring by TLC indicated reaction was no longer progressing, and solvent was removed in vacuo. Residue was purified by MPLC (5% MeOH:CH_2Cl_2, isocratic) and product containing fractions were pooled to give **13** as a colorless glass (43 mg, 43% yield).

HRMS (ESI) (m/z) calculated for: $C_{26}H_{46}N_3O_7S^+$, $[M+H]^+$: 544.3051, observed: 544.2040.

NMR spectra is heavily overlapped due to the equilibrating mixture of diastereomers and can be found in **Supplementary Notes Figs. 10** and **11**. NMR spectra of purified diastereomers **13a** and **13b** can be found in **Supplementary Notes Figs. 12-19**. The complicated nature of the spectra makes determining absolute configuration impossible and so **13a** and **13b** are defined arbitrarily.

Preparatory HPLC purification and NMR spectral characterization of **13a** and **13b**: According to previously described conditions, 10 mg (0.02 mmol) of thioester **32** was added a premixed solution of 2:1 AcOH:H_2O (0.9 mL) and solution was stirred at RT for 5 hours, and solvent was removed in vacuo. Crude product was dissolved in minimal 10% MeOH:CH_2Cl_2, filtered through a short plug of silica, and fractions containing **13** were combined and concentrated. Crude **13** was then purified via preparatory HPLC (Phenomenex Luna ® 5mm C18(2) 100 Å LC Column 100 × 21.2 mm) using a gradient of 40-80% CH_3CN:H_2O (0.1% formic acid) over 45 minutes at 10 mL/min with **13a** eluting at 8.7 minutes and **13b** eluting at 9.7 minutes (**a** and **b** assigned arbitrarily, and are not representative of absolute configuration). Solvent from fractions containing **13a** and **13b** was removed in vacuo (rotary evaporation to remove CH_3CN, then frozen as a thin film with liquid N_2 and lyophilized to remove residual water) to yield 1.2 mg of **13a** and 1.1 mg of **13b** (24% combined isolated yield). Upon storage in organic solvent and ambient conditions, these compounds slowly epimerized to mixtures (over the course of 1 day in K_2CO_3 treated $CDCl_3$) as determined by NMR. This epimerization was immediate upon addition of either sample to aqueous buffers (pH≥7.0) as determined by LC-MS. NMR spectra can be found in **Supplementary Notes Figs. 12-19**. This mixture is referred to as **13** and is the substrate used in subsequent enzymatic assays.

13a: HRMS (ESI) (m/z) calculated for: $C_{26}H_{46}N_3O_7S^+$, $[M+H]^+$: 544.3051, observed: 544.3053.

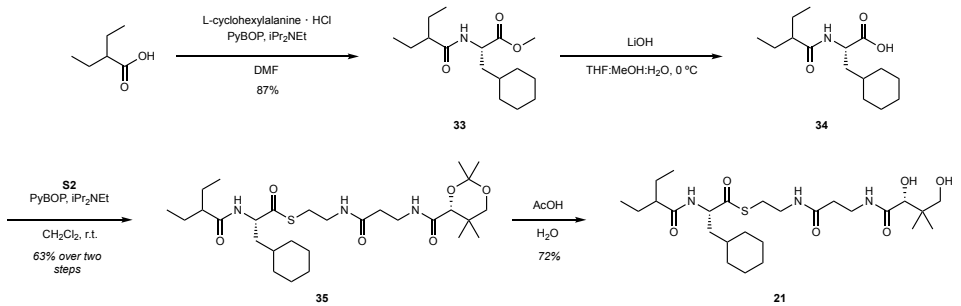
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (d, *J* = 6.7 Hz, 1H), 7.18 (d, *J* = 7.4 Hz, 1H), 6.22 (t, *J* = 5.9 Hz, 1H), 4.57 (ddd, *J* = 11.4, 7.4, 4.6 Hz, 1H), 3.98 (s, 1H), 3.68 – 3.56 (m, 2H), 3.49 (m, 3H), 3.42 (t, *J* = 7.0 Hz, 1H), 3.32 (ddd, *J* = 12.5, 9.1, 5.0 Hz, 1H), 3.15 (m, 1H), 3.09 (ddd, *J* = 12.1, 7.0, 4.4 Hz, 1H), 2.97 (ddd, *J* = 13.8, 7.3, 4.4 Hz, 1H), 2.38 (t, *J* = 5.8 Hz, 2H), 2.30 (s, 3H), 1.94 (m, 2H), 1.82 (m, 1H), 1.76 – 1.49 (m, 10H), 1.39 – 1.10 (m, 3H), 1.04 (s, 3H), 0.97 – 0.91 (m, 6H).

^{13}C NMR (125 MHz, $CDCl_3$) δ 208.02, 201.78, 173.64, 171.66, 169.48, 77.73, 70.91, 61.36, 57.72, 46.27, 39.39, 39.17, 35.70, 34.90, 34.25, 33.66, 31.85, 29.87, 28.42, 26.21, 26.10, 25.88, 24.24, 21.88, 20.28, 11.51.

13b: HRMS (ESI) (m/z) calculated for: $C_{26}H_{46}N_3O_7S^+$, $[M+H]^+$: 544.3051, observed: 544.3049 .

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33 (d, *J* = 6.3 Hz, 1H), 7.01 (d, *J* = 7.4 Hz, 1H), 6.30 (t, *J* = 5.8 Hz, 1H), 4.60 (ddd, *J* = 11.7, 7.3, 4.6 Hz, 1H), 4.01 (s, 1H), 3.61 (m, 1H), 3.57 – 3.45 (m, 4H), 3.45 – 3.32 (m, 2H), 3.09 – 2.98 (m, 2H), 2.40 (t, *J* = 5.9 Hz, 2H), 2.28 (s, 3H), 1.95 (p, *J* = 7.4 Hz, 2H), 1.79 – 1.57 (m, 10H), 1.37 – 1.08 (m, 3H), 1.04 (s, 3H), 0.99 (t, *J* = 7.4 Hz, 4H), 0.95 (s, 4H).

^{13}C NMR (125 MHz, $CDCl_3$) δ 207.66, 201.69, 173.75, 171.78, 169.34, 77.72, 70.92, 62.26, 57.77, 44.09, 39.54, 39.27, 35.85, 35.16, 34.23, 33.73, 32.02, 29.86, 28.46, 26.32, 26.19, 25.97, 24.14, 21.93, 20.49, 12.02.



Supplementary Notes Fig. 3: Synthesis of pantetheine-activated chain transfer probe **21**.

(*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanoate (**33**): In an oven-dried round bottom flask L-cyclohexylalanine HCl (1.0 g, 4.5 mmol, 1.0 equiv.) was suspended in anhydrous DMF with stirring followed by addition of *i*Pr_2NEt (1 mL, 5.7 mmol, 1.3 equiv.) and upon addition. After 20 minutes, 2-ethylbutyric acid (600 uL, 5.4 mmol, 1.2 equiv.) and PyBOP (2.5 g, 4.7 mmol, 1.05 eq) were added sequentially, headspace was purged with Ar, and the reaction was allowed to stir for 12 hours at room temperature. Reaction was quenched by the addition of 1 M HCl followed by extraction with Et_2O (3 × 25 mL). Combined organic fractions were washed with saturated $NaHCO_3$, dried over $MgSO_4$, filtered and solvent was removed in vacuo. Crude residue was purified by MPLC (0-100% hexanes:EtOAc), product containing fractions were combined and concentrated to give **33** as a stiff foam (1.1 g, 87% yield).

HRMS (ESI) (m/z) calculated for $C_{16}H_{30}NO_3^+$, $[M+H]^+$: 284.2220 , observed: 284.2229.

^{1}H NMR (500 MHz, $CDCl_3$) δ 5.77 (d, *J* = 8.5 Hz, 1H), 4.72 (td, *J* = 9.1, 5.0 Hz, 1H), 3.72 (s, 3H), 1.92 (tt, *J* = 9.5, 5.1 Hz, 1H), 1.81 (d, *J* = 13.0 Hz, 1H), 1.74 – 1.56 (m, 8H), 1.56 – 1.42 (m, 4H), 1.32 – 1.10 (m, 3H), 0.90 (q, *J* = 7.2 Hz, 7H). NMR spectra can be found in **Supplementary Notes Fig. 20**.

S-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl)(*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanethioate (**35**): **33** (925 mg, 3.3 mmol, 1 equiv.) was dissolved in a 2:1 mixture of THF:MeOH (15 mL, 0.2 M) and chilled to 0 ºC. LiOH (80 mg, 3.3 mmol, 1 equiv.) was added as a solution in deionized water, and after 2 hours reaction was judged as complete by TLC and volatile solvents were removed in vacuo. Residue was diluted with 1M $NaHCO_3$, non-polar impurities were removed via Et_2O (1 × 50 mL) wash, aqueous layer was acidified to pH= ~3, and extracted with CH_2Cl_2 (3 × 50 mL). Combined organic layers were dried over $MgSO_4$, filtered, and concentrated in

vacuo to give crude acid **34** as a white powder (733 mg) which was used directly with no additional purification.

34 was then dissolved in anhydrous CH_2Cl_2 under an atmosphere of argon and **27** (1.05 g, 3.3 mmol, 1 equiv.) was added as a solution in CH_2Cl_2 followed by *i*Pr_2NEt (0.7 mL, 3.6 mmol, 1.1 equiv.) and PyBOP (1.7 g, 3.3 mmol, 1.0 equiv.) which was added as a single solid addition followed by a purge of the reaction headspace. Reaction was no longer progressing after 12 hours and was quenched with 1M HCl and extracted with CH_2Cl_2. Organic extracts were combined, dried over $MgSO_4$, filtered and concentrated in vacuo. Crude product was purified via MPLC (0-100% EtOAc:hexanes) and product containing fractions were combined and solvent removed in vacuo to give **35** as a white solid (1.3 g, 63% yield over two steps).

HRMS (ESI) (m/z) calculated for $C_{29}H_{52}N_3O_6S^+$, $[M+H]^+$: 570.3571, observed: 570.3498.

^{1}H NMR (500 MHz, CD_3OD) δ 4.58 (dd, *J* = 11.2, 4.3 Hz, 1H), 3.81 (dd, *J* = 93.6, 11.7 Hz, 1H), 3.42 (m, 2H), 3.36 – 3.21 (m, 2H), 2.95 (t, *J* = 6.8 Hz, 2H), 2.36 (t, *J* = 6.8 Hz, 2H), 2.21 – 2.07 (m, 1H), 2.05 – 1.92 (m, 2H), 1.79 – 1.43 (m, 12H), 1.41 (d, *J* = 4.0 Hz, 6H), 1.39 – 1.10 (m, 3H), 1.03 – 0.81 (m, 14H).

NMR spectra can be found in **Supplementary Notes Figs. 21**.

S-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanethioate (**21**): **35** (100 mg, 0.175 mmol, 1.0 equiv.) was dissolved in a premixed solution of 2:1 AcOH:H_2O (9 mL, 0.02 M) and solution was stirred at RT. After 5 hours, HPLC-MS analysis determine reaction was no longer progressing, and solvent was removed in vacuo. Residue was purified by MPLC (5% MeOH:CH_2Cl_2, isocratic) and product containing fractions were pooled to give **21** as a white solid glass (66 mg, 72% yield).

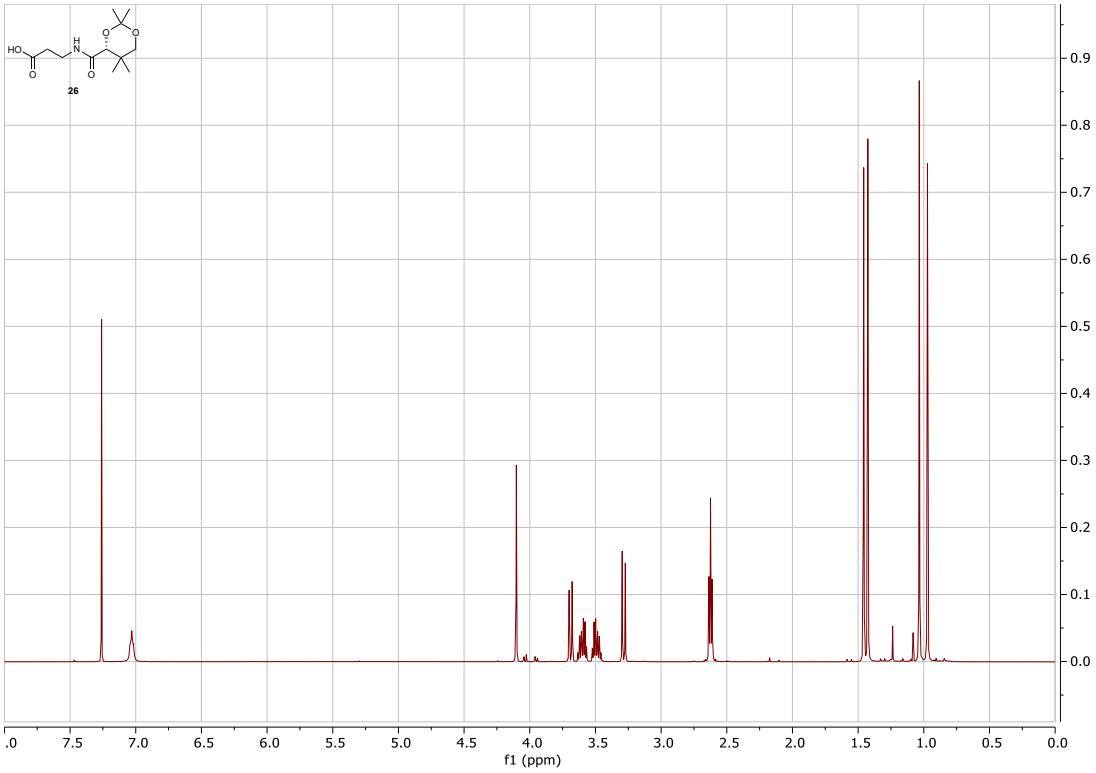
HRMS (ESI) (m/z) calculated for $C_{26}H_{48}N_3O_6S^+$, $[M+H]^+$: 530.3258, observed: 530.3264.

^{1}H NMR (500 MHz, CD_3OD) δ 4.58 (dd, *J* = 11.2, 4.3 Hz, 1H), 3.86 (s, 1H), 3.50 – 3.33 (m, 4H), 3.31 – 3.25 (m, 2H), 2.96 (t, *J* = 6.7 Hz, 2H), 2.37 (t, *J* = 6.7 Hz, 2H), 2.14 (m, 1H), 1.77 – 1.41 (m, 11H), 1.37 (m, 1H), 1.31 – 1.08 (m, 2H), 0.99 (m, 1H), 0.94 (t, *J* = 7.4 Hz, 3H), 0.91 – 0.85 (m, 9H), 0.83 (m, 1H).

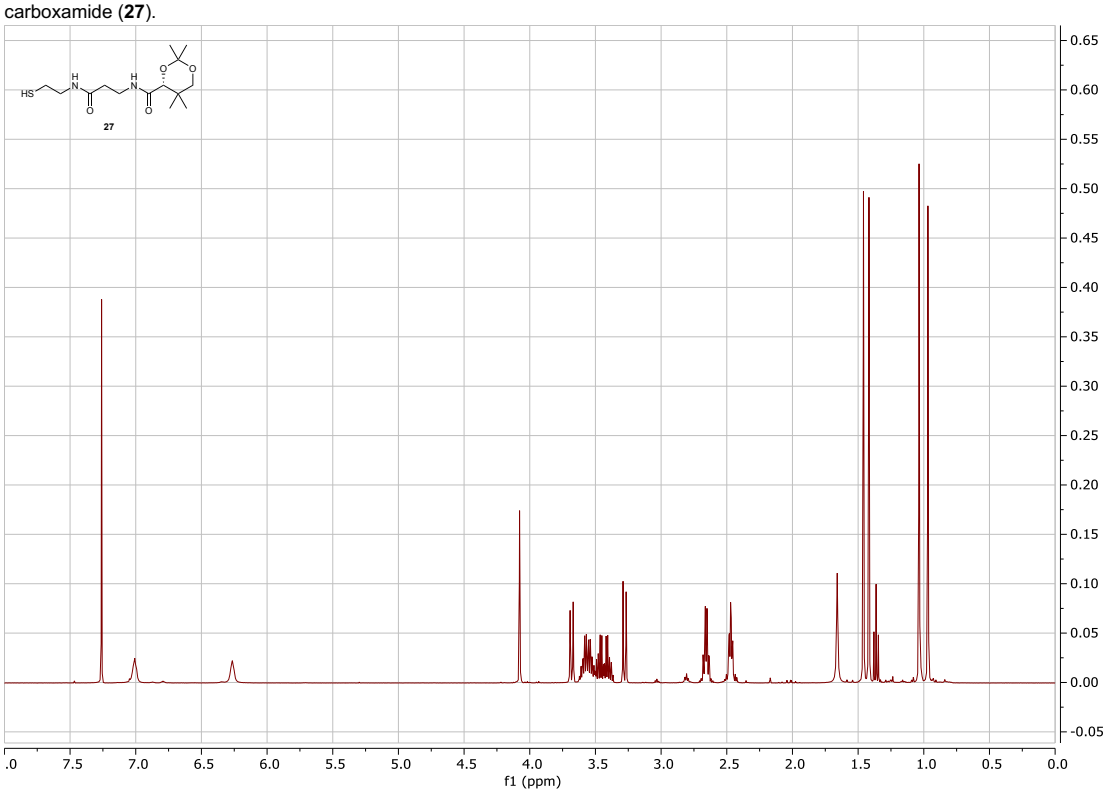
^{13}C NMR (125 MHz, CD_3OD) δ 201.46, 178.00, 174.71, 172.55, 75.93, 69.02, 57.05, 50.28, 39.05, 38.78, 38.62, 35.08, 35.00, 33.97, 33.70, 31.32, 29.38, 27.55, 26.17, 26.13, 25.77, 25.47, 20.02, 19.59, 11.46, 11.19.

NMR spectra can be found in **Supplementary Notes Figs. 22** and **23**.

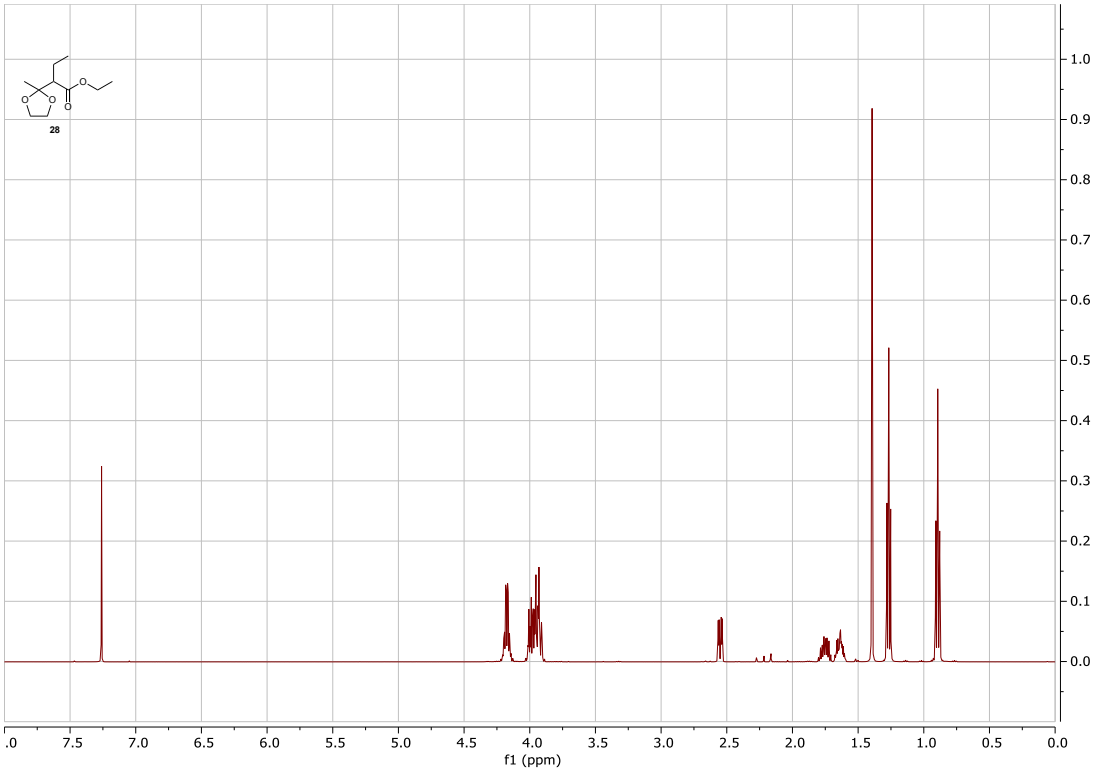
Supplementary Notes Fig. 4: ^{1}H NMR spectra (*R*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (**26**).



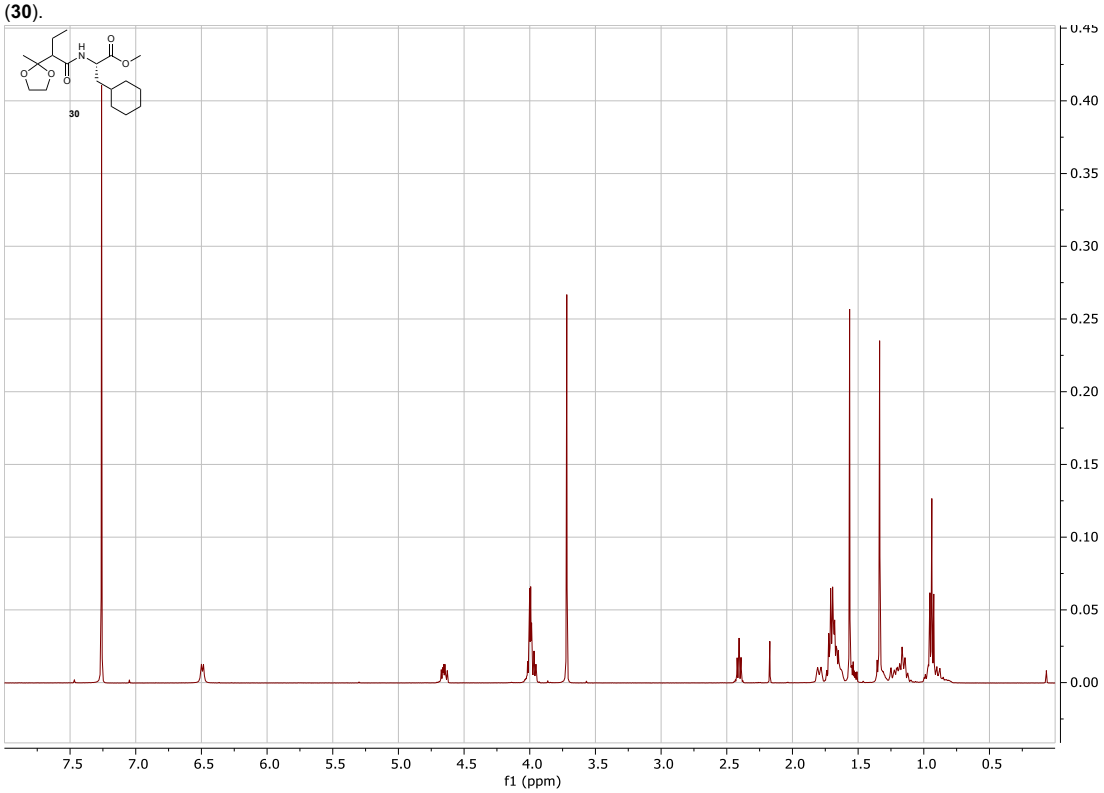
Supplementary Notes Fig. 5: ^{1}H NMR spectra (*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (**27**).



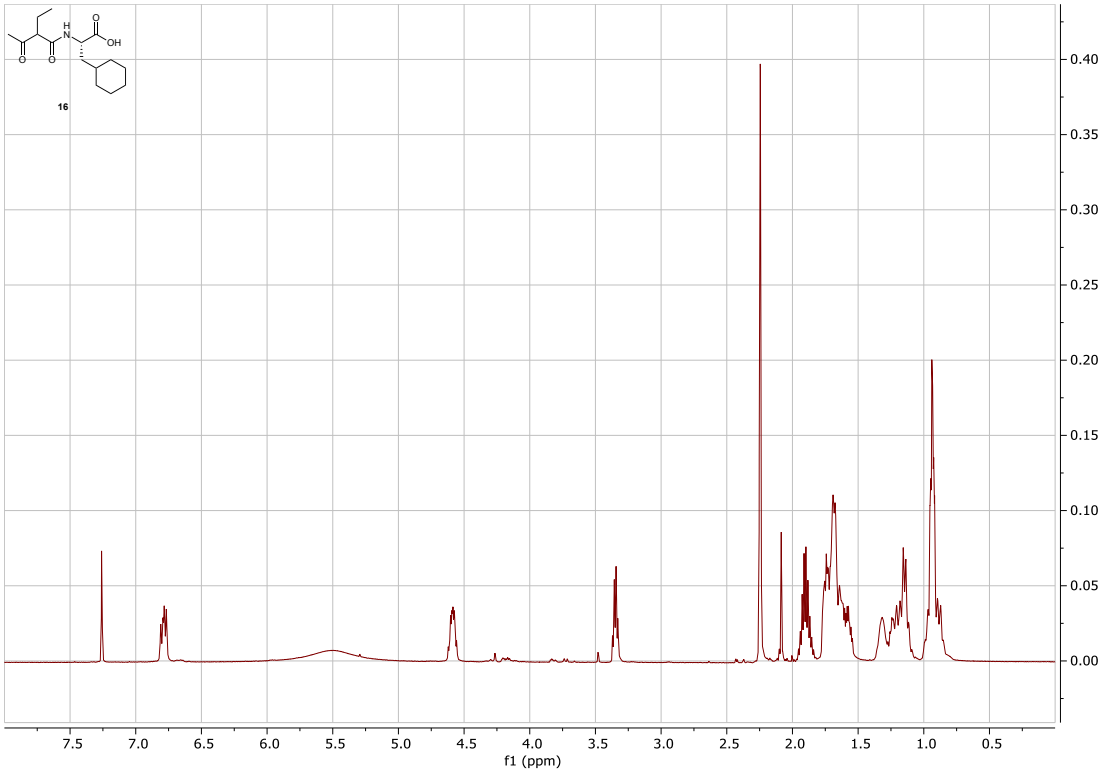
Supplementary Notes Fig. 6: ^{1}H NMR spectra ethyl 2-(2-methyl-1,3-dioxolan-2-yl)butanoate (**28**).



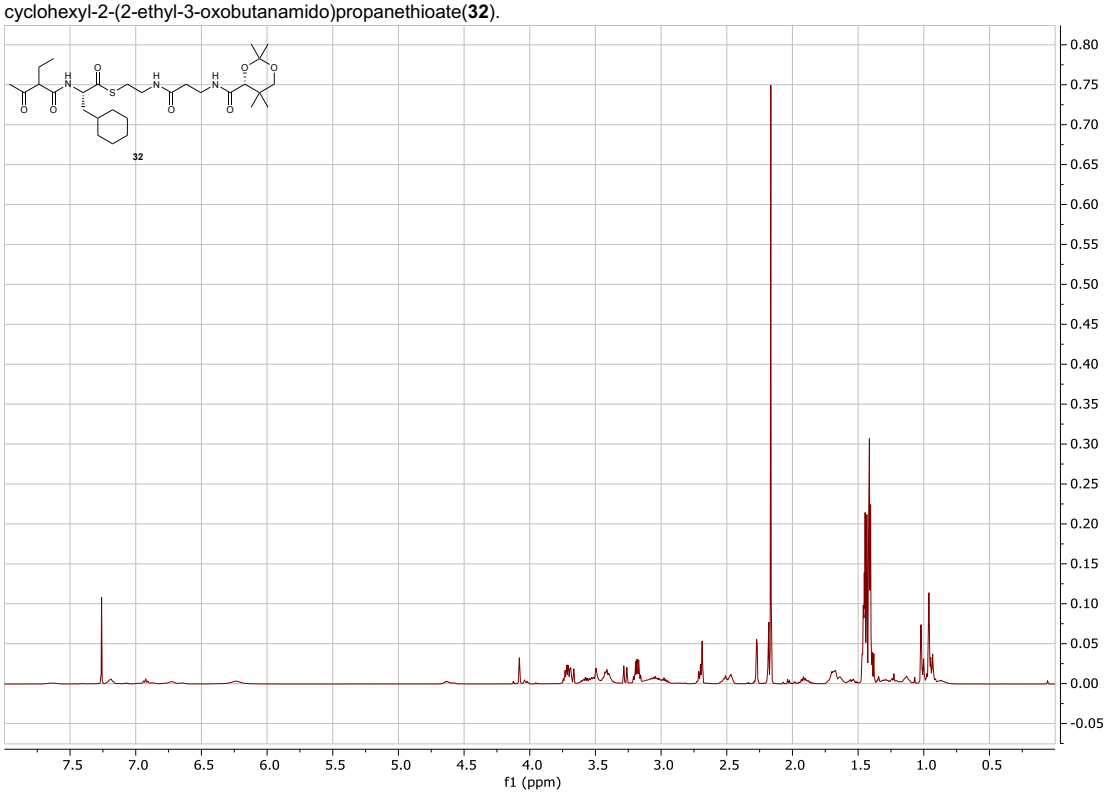
Supplementary Notes Fig. 7: ^{1}H NMR spectra methyl (2*S*)-3-cyclohexyl-2-(2-(2-methyl-1,3-dioxolan-2-yl)butanamido)propanoate (**30**).



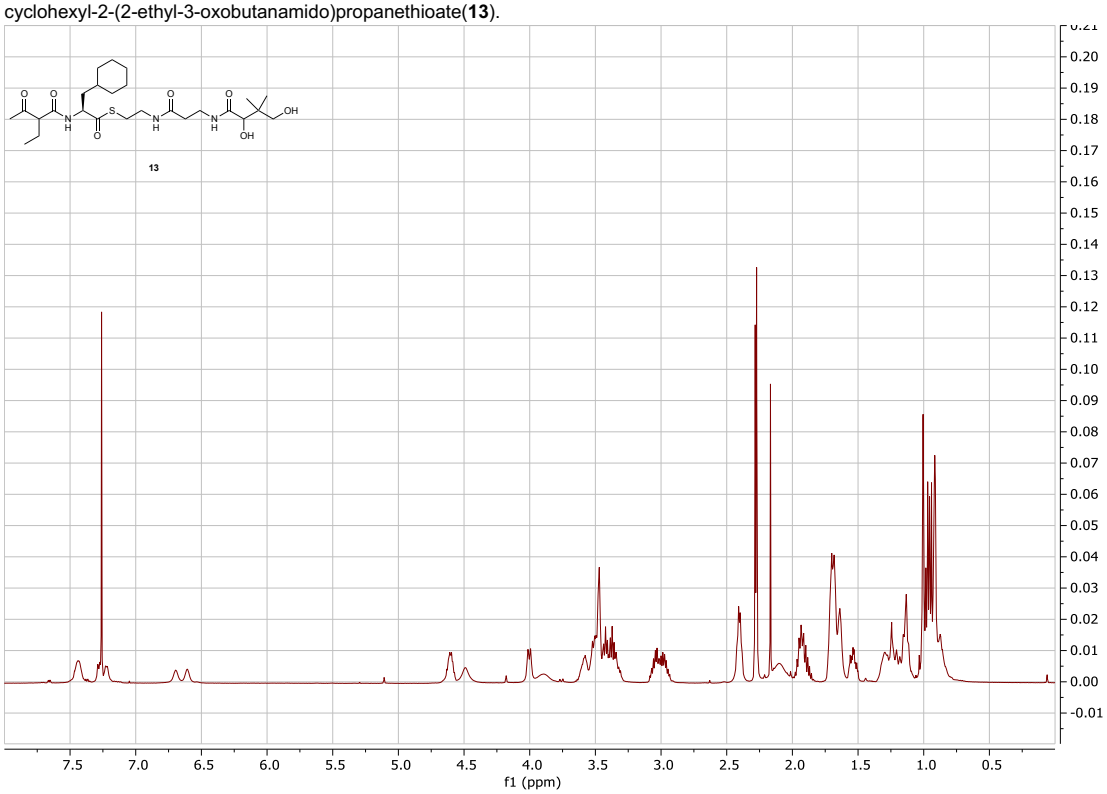
Supplementary Notes Fig. 8: ^{1}H NMR spectra (2*S*)-3-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanoic acid (**16**).



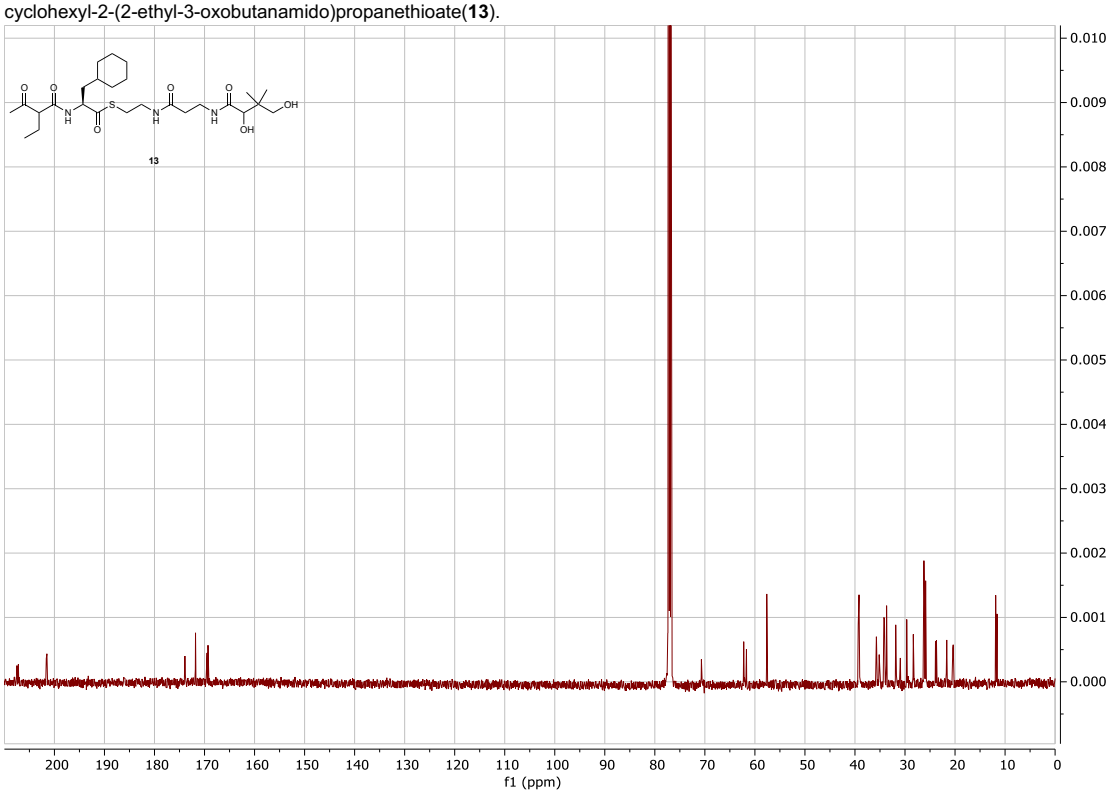
Supplementary Notes Fig. 9: ^{1}H NMR spectra *S*-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl)(2*S*)-3-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanethioate(**32**).



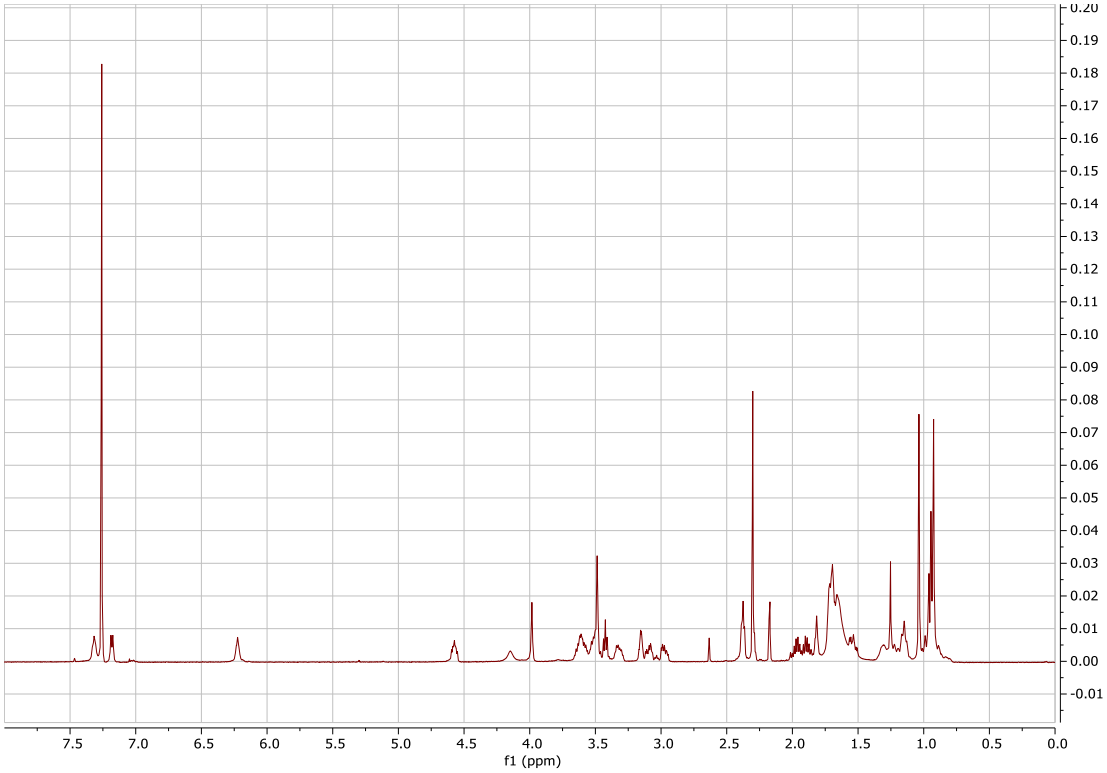
Supplementary Notes Fig. 10: ^{1}H NMR spectra *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(2*S*)-3 cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanethioate(**13**).



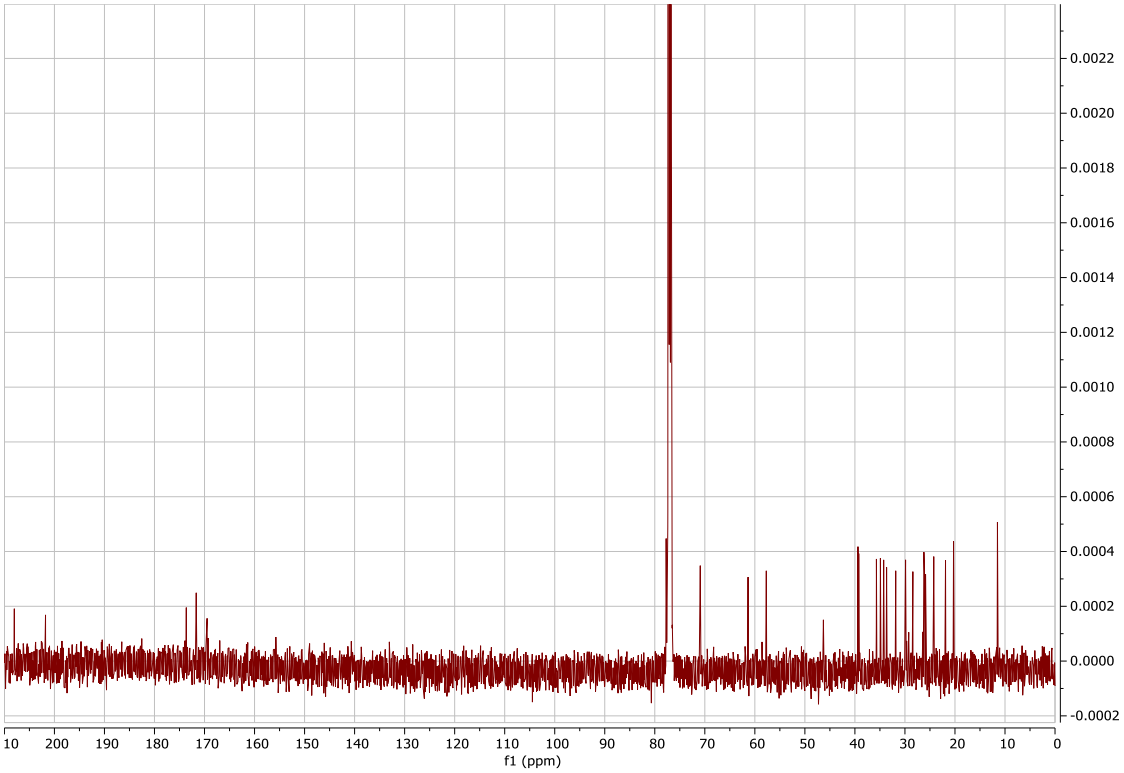
Supplementary Notes Fig. 11: ^{13}C NMR spectra *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(2*S*)-2-cyclohexyl-2-(2-ethyl-3-oxobutanamido)propanethioate(**13**).



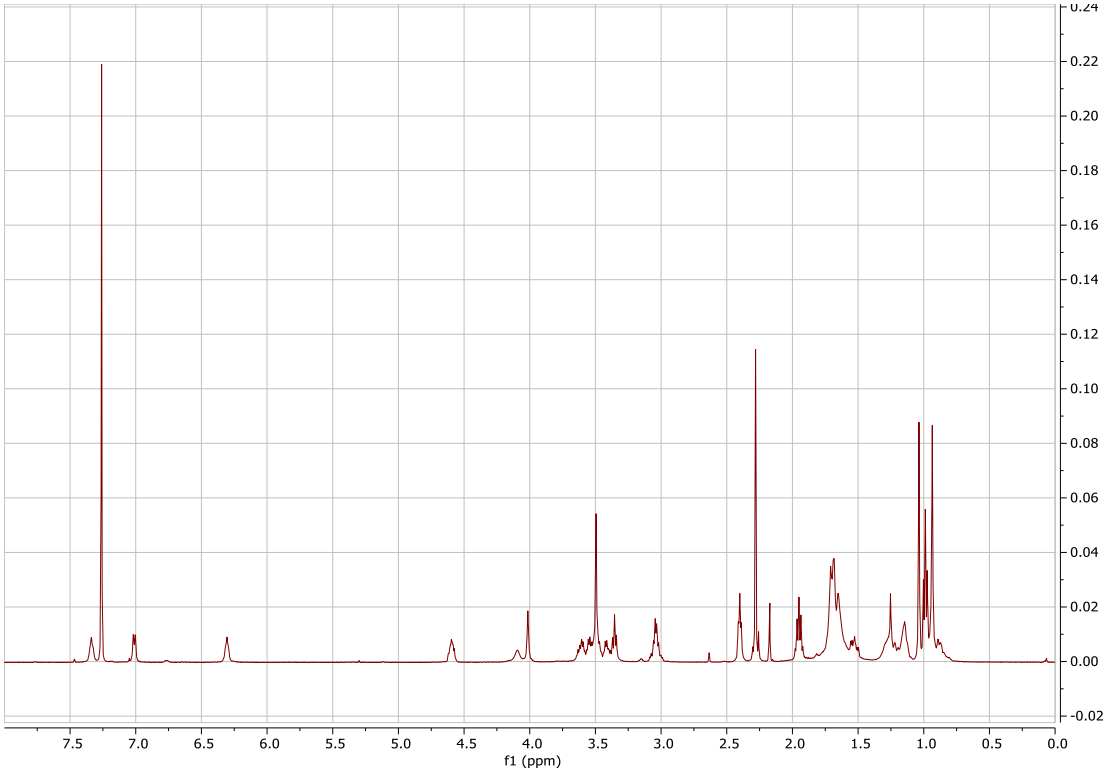
Supplementary Notes Fig. 12: ^{1}H NMR spectra for preparatory HPLC purified **13a**.



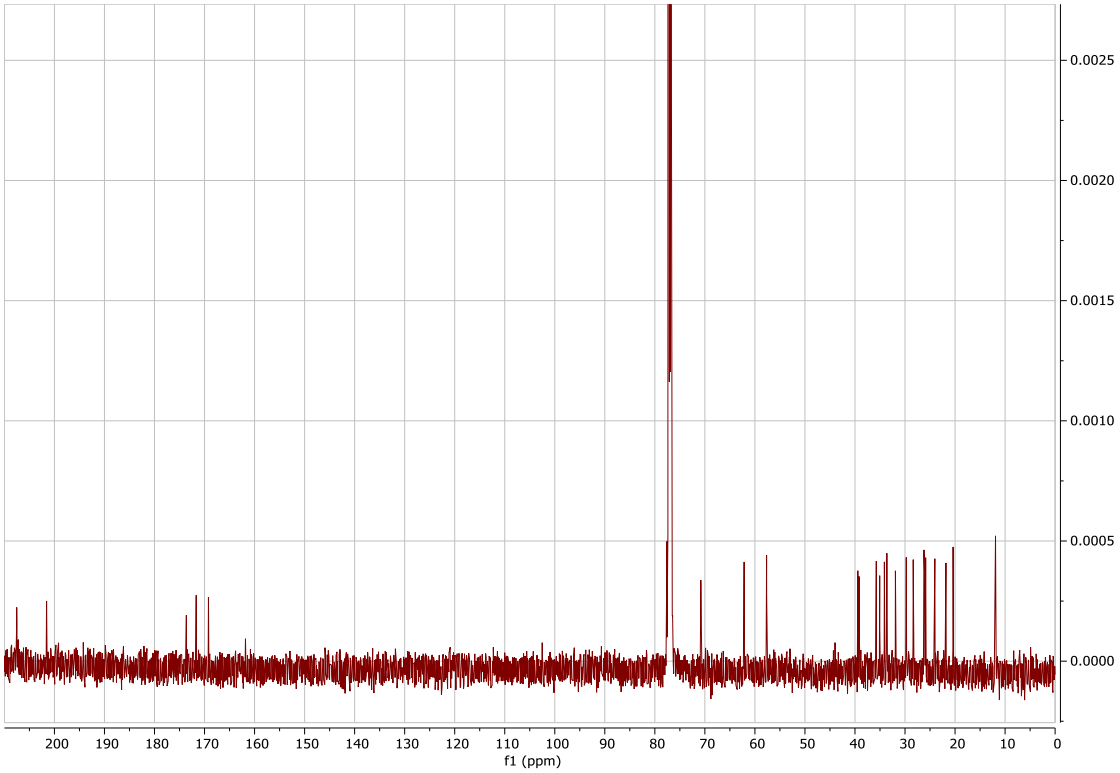
Supplementary Notes Fig. 13: ^{13}C NMR spectra for preparatory HPLC purified **13a**.



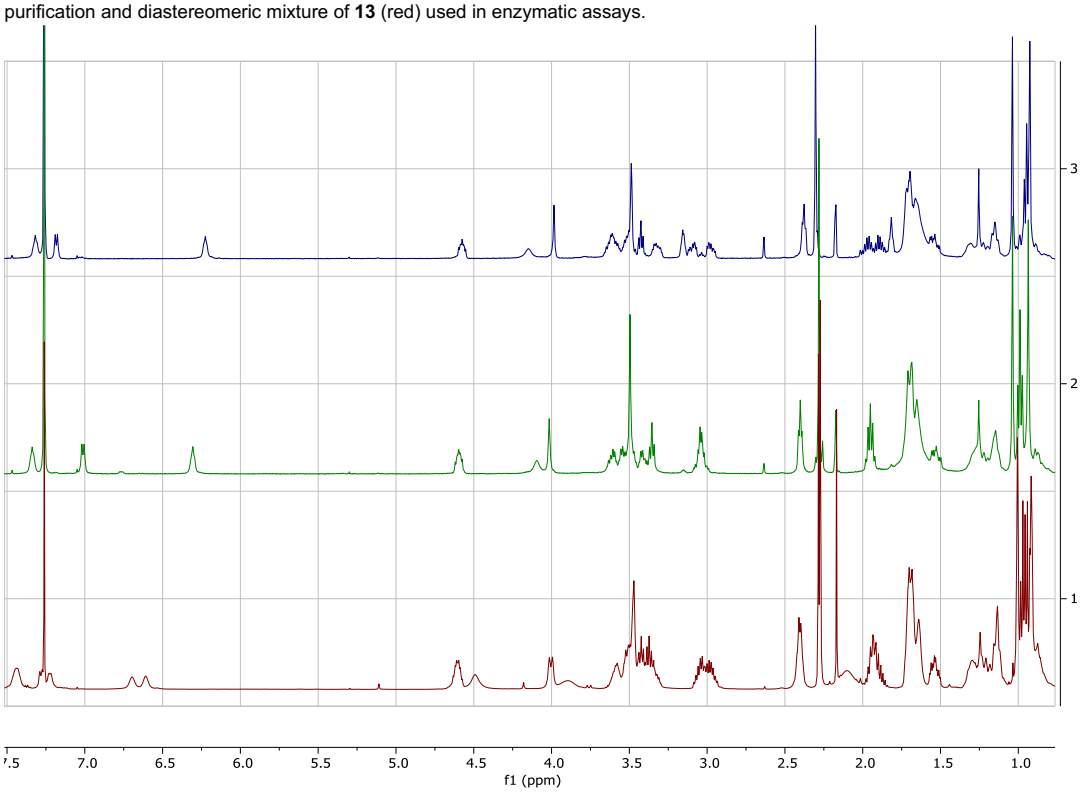
Supplementary Notes Fig. 14: ^{1}H NMR for preparatory HPLC purified spectra **13b**.



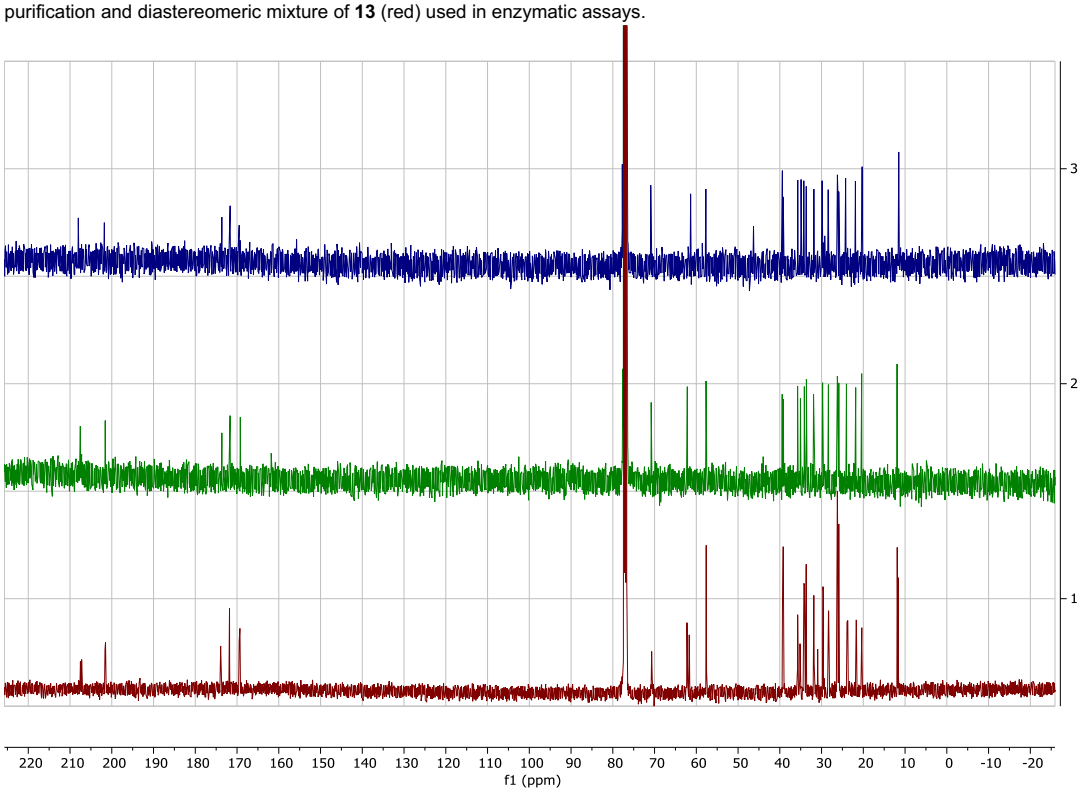
Supplementary Notes Fig. 15: ^{13}C NMR spectra for preparatory HPLC purified **13b**.



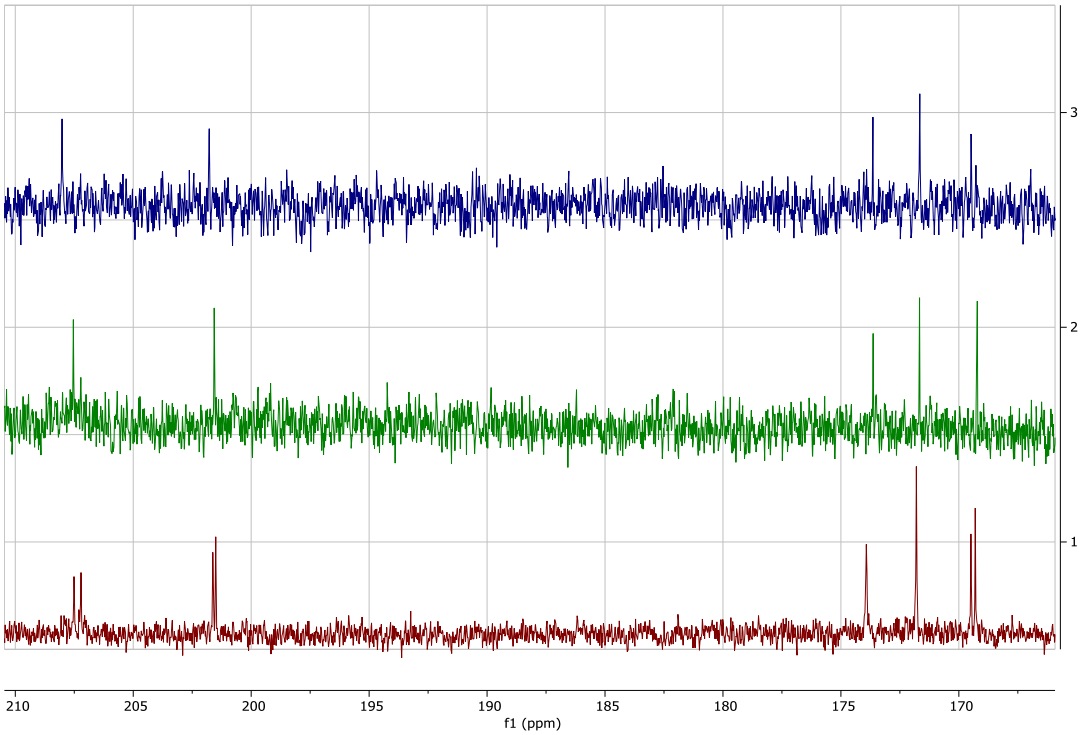
Supplementary Notes Fig. 16: Stacked ^{1}H NMR spectra of isolated diastereomers **13a** (blue), **13b** (green) from preparative HPLC purification and diastereomeric mixture of **13** (red) used in enzymatic assays.



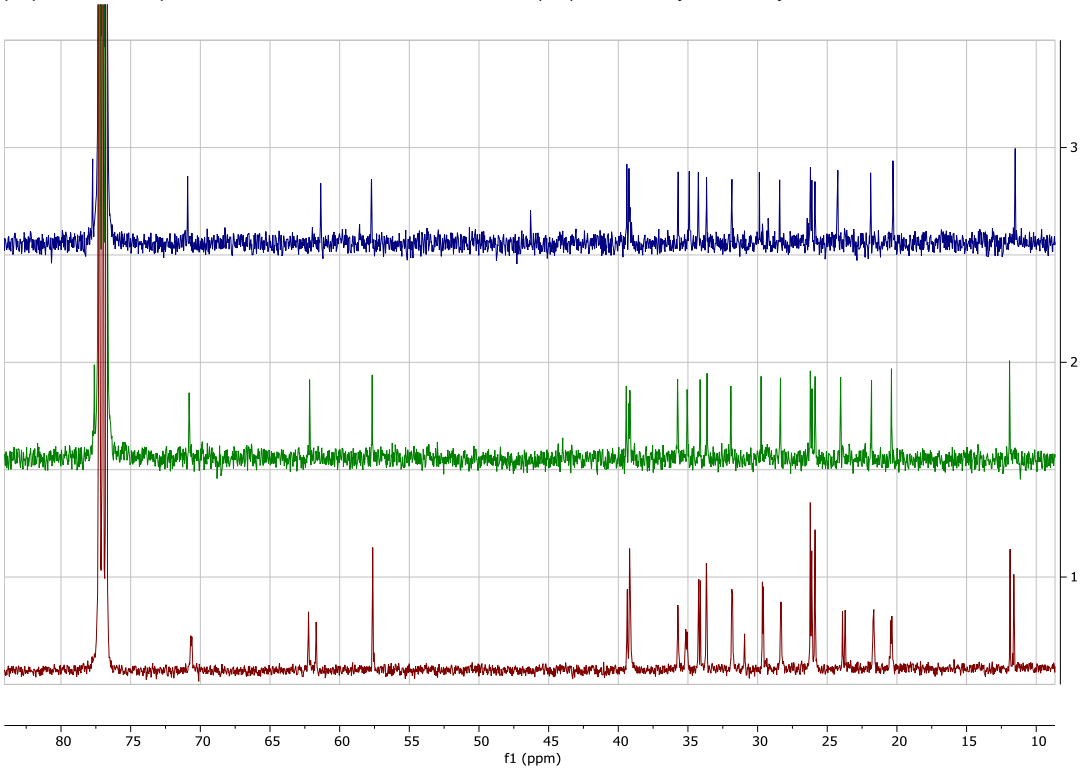
Supplementary Notes Fig. 17: Stacked ^{13}C NMR spectra of isolated diastereomers **13a** (blue), **13b** (green) from preparative HPLC purification and diastereomeric mixture of **13** (red) used in enzymatic assays.



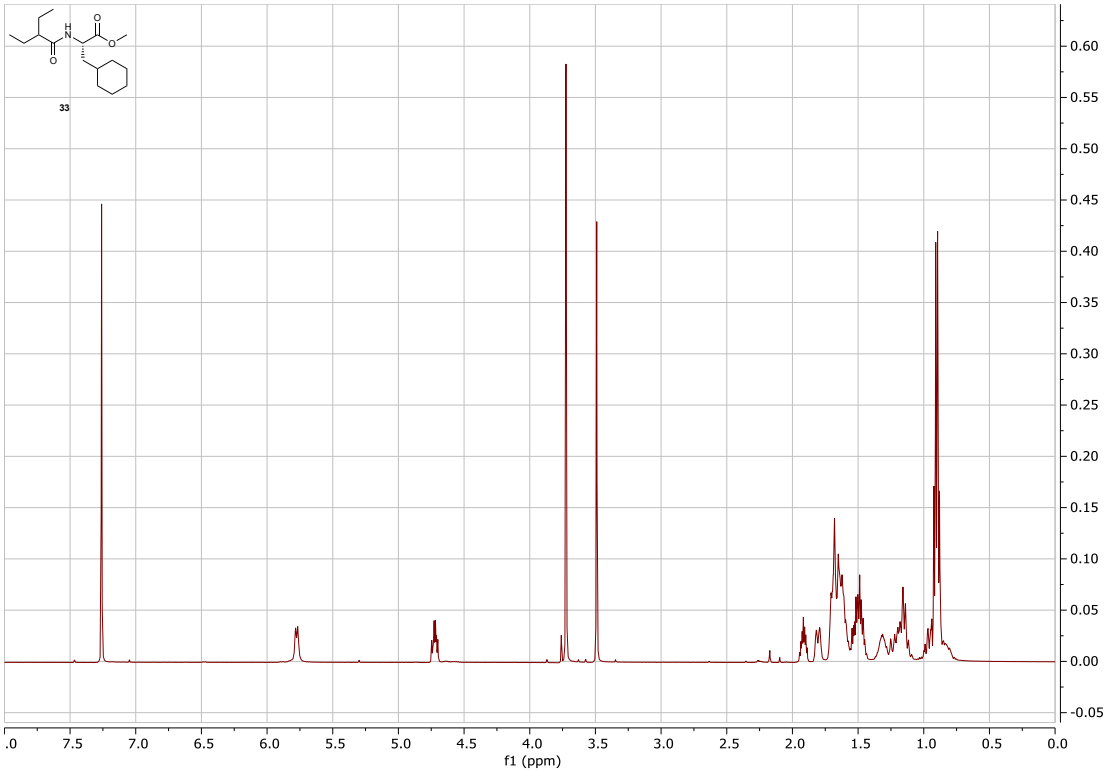
Supplementary Notes Fig. 18: Stacked ^{13}C NMR spectra from 210-170 ppm of isolated diastereomers **13a** (blue), **13b** (green) from preparative HPLC purification and diastereomeric mixture of **13** (red) used in enzymatic assays.



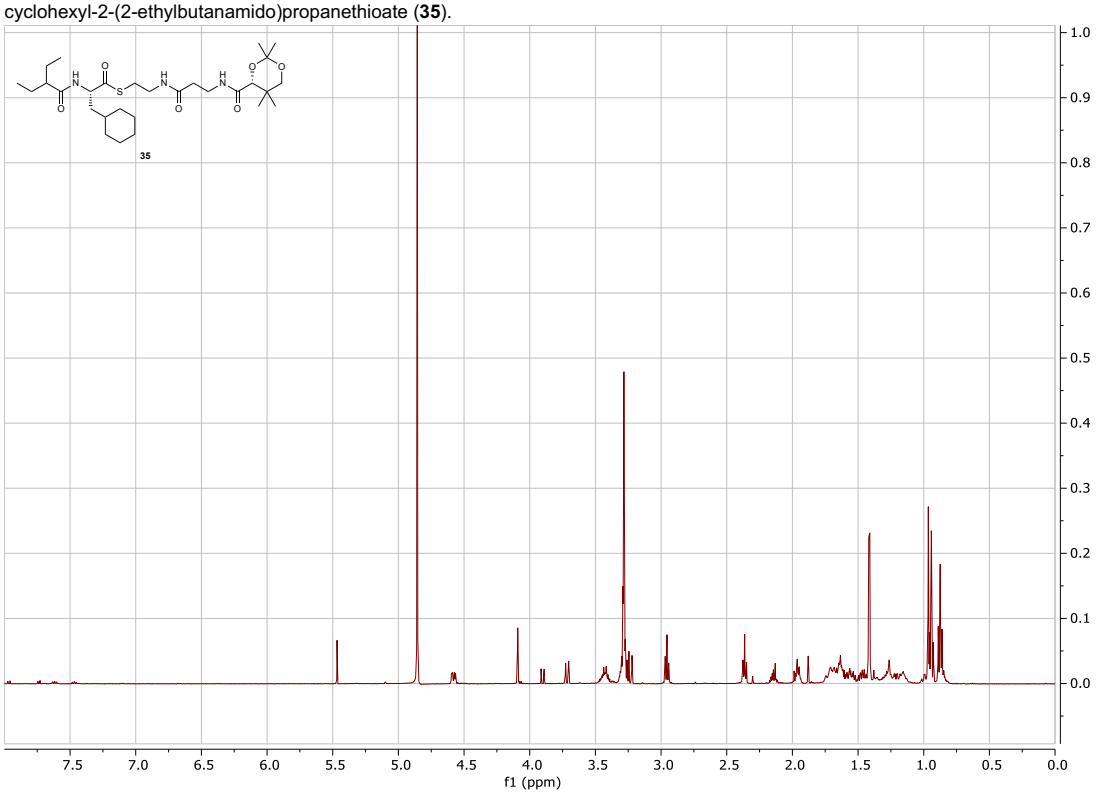
Supplementary Notes Fig. 19: Stacked ^{13}C NMR spectra from 90-10 ppm of isolated diastereomers **13a** (blue), **13b** (green) fron preparative HPLC purification and diastereomeric mixture of **13** (red) used in enzymatic assays.



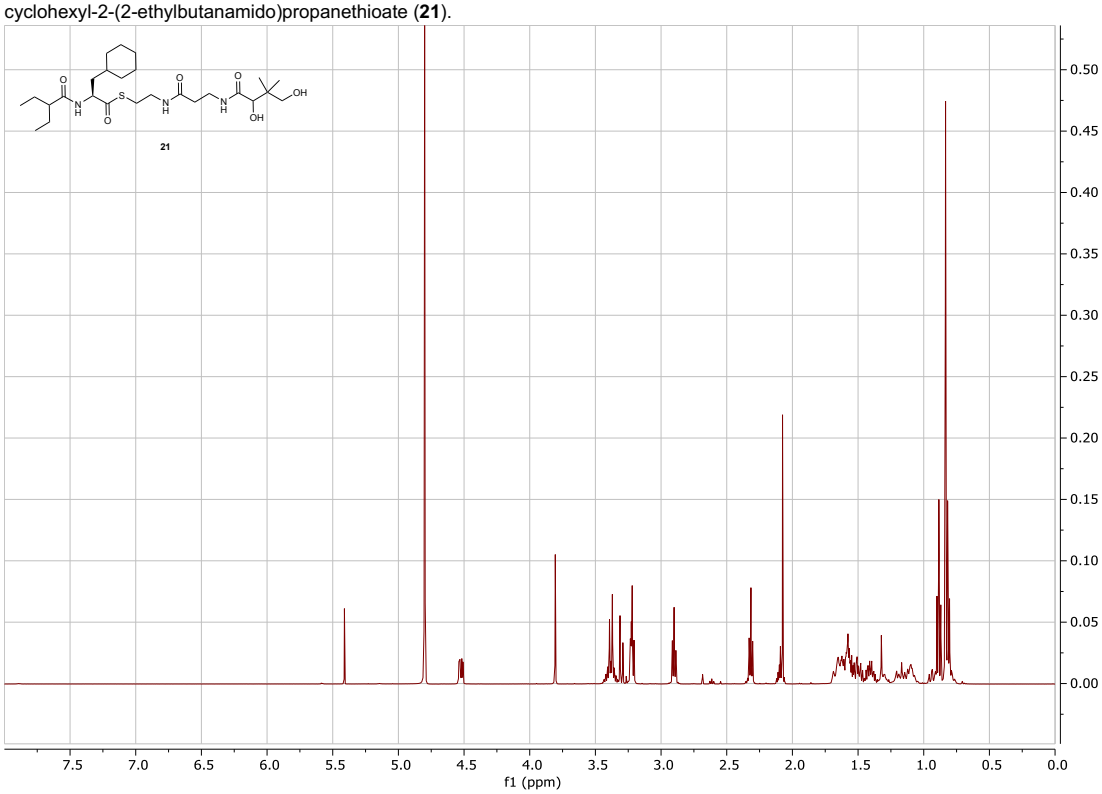
Supplementary Notes Fig. 20: ^{1}H NMR spectra methyl (*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanoate (**33**).



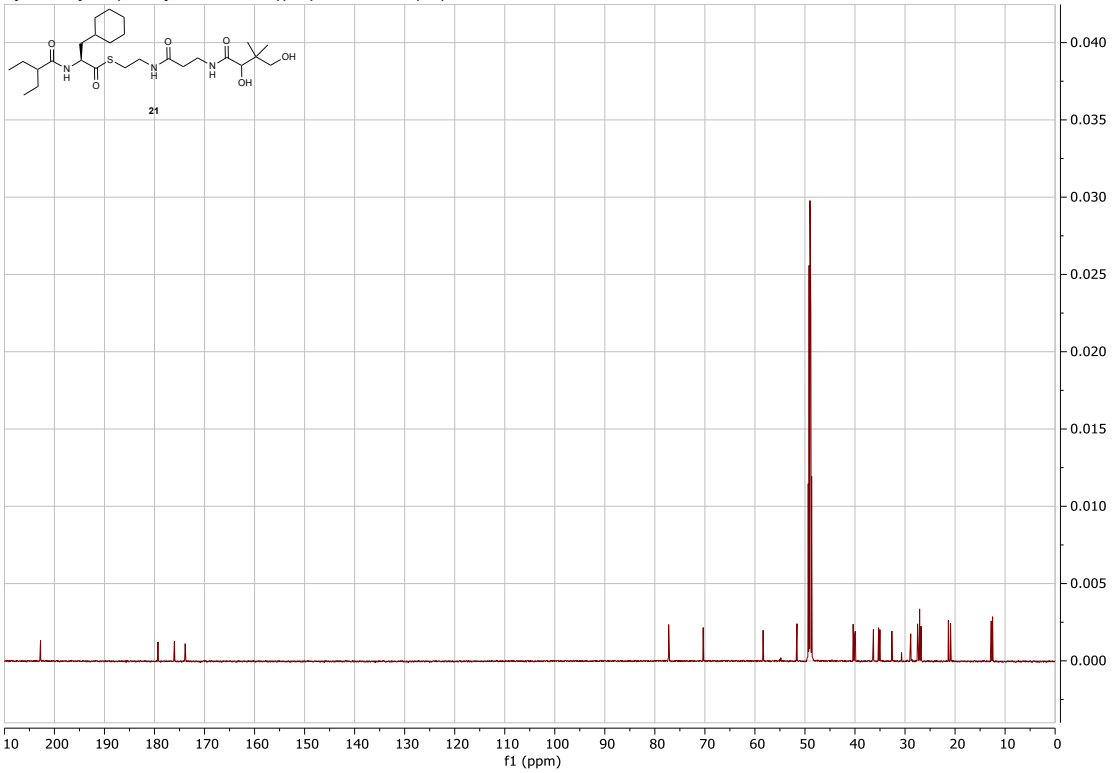
Supplementary Notes Fig. 21: ^{1}H NMR spectra *S*-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl)(*S*)-2
cyclohexyl-2-(2-ethylbutanamido)propanethioate (**35**).



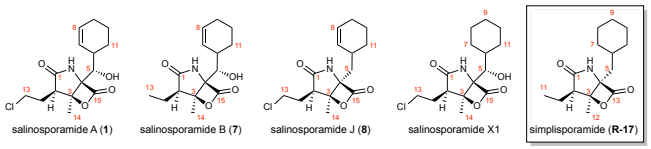
Supplementary Notes Fig. 22: ^{1}H NMR spectra *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanethioate (**21**).



Supplementary Notes Fig. 23: ^{13}C NMR spectra *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)(*S*)-3-cyclohexyl-2-(2-ethylbutanamido)propanethioate (**21**).



Supplementary Notes Fig. 24: Numbering scheme for known salinosporamides A (**1**), B (**7**), J (**8**) and X1, and simplisporamide (**R-17**).

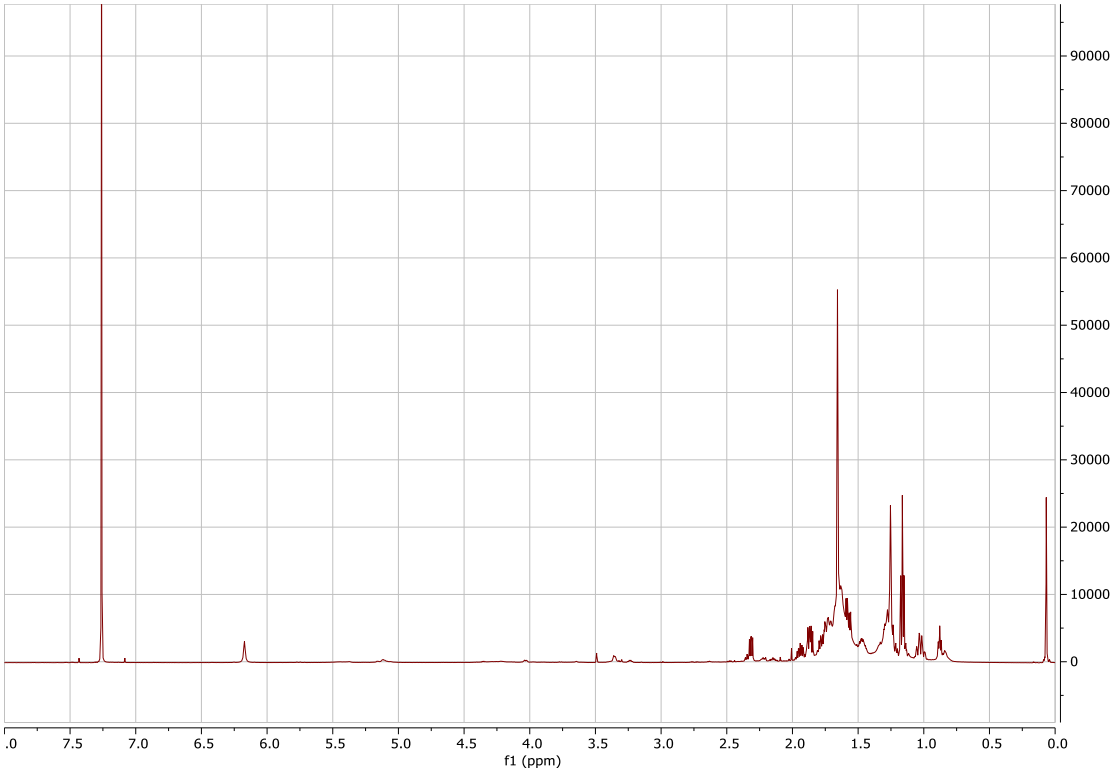


Supplementary Notes Table 1: Simplisporamide (**R-17**) NMR assignment table.

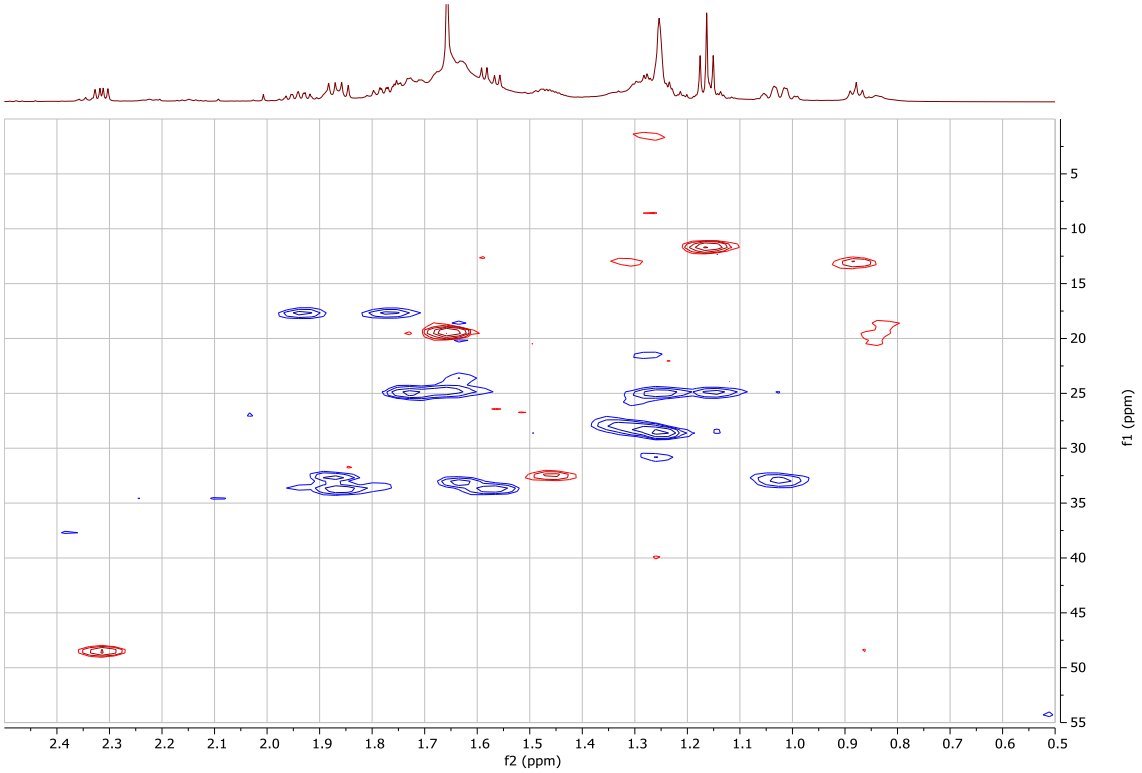
C #	δ_H, (multiplicity, *J* in Hz)	δ_C
1	-----	176.94
2	2.31, (dd, 9.2, 5.5)	48.52
3	-----	84.70
4	-----	73.46
5	1.57 (dd, 14.8, 6.2), 1.86 (m)	33.72
6	1.46 (m)	32.47
7	1.02 (m)	32.85
8	1.87 (m), 1.63 (m)	32.90
9	1.25 (m)	28.50
10	1.77 (m), 1.94 (m)	17.69
11	1.17 (t, 7.5)	11.67
12	1.66 (s)	19.44
13	-----	168.20

The structure of our simplisporamide (**17a**) was solved *de novo*, and our spectra is in agreement with of those of salinosproamide B (**7**, main text reference #2),salinosporamide J (**8**, main text reference #30), and salinosporamide X1 (main text reference #4).

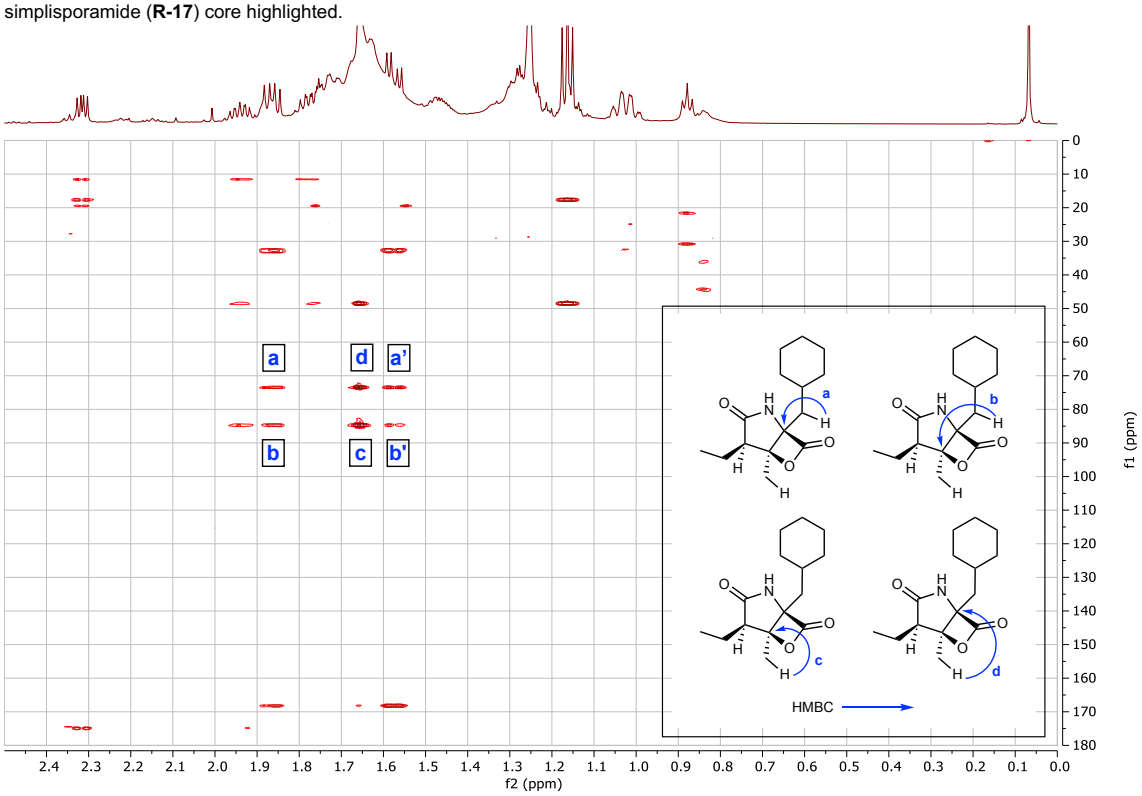
Supplementary Notes Fig. 25: ^{1}H NMR spectra simplisporamide (**R-17**).



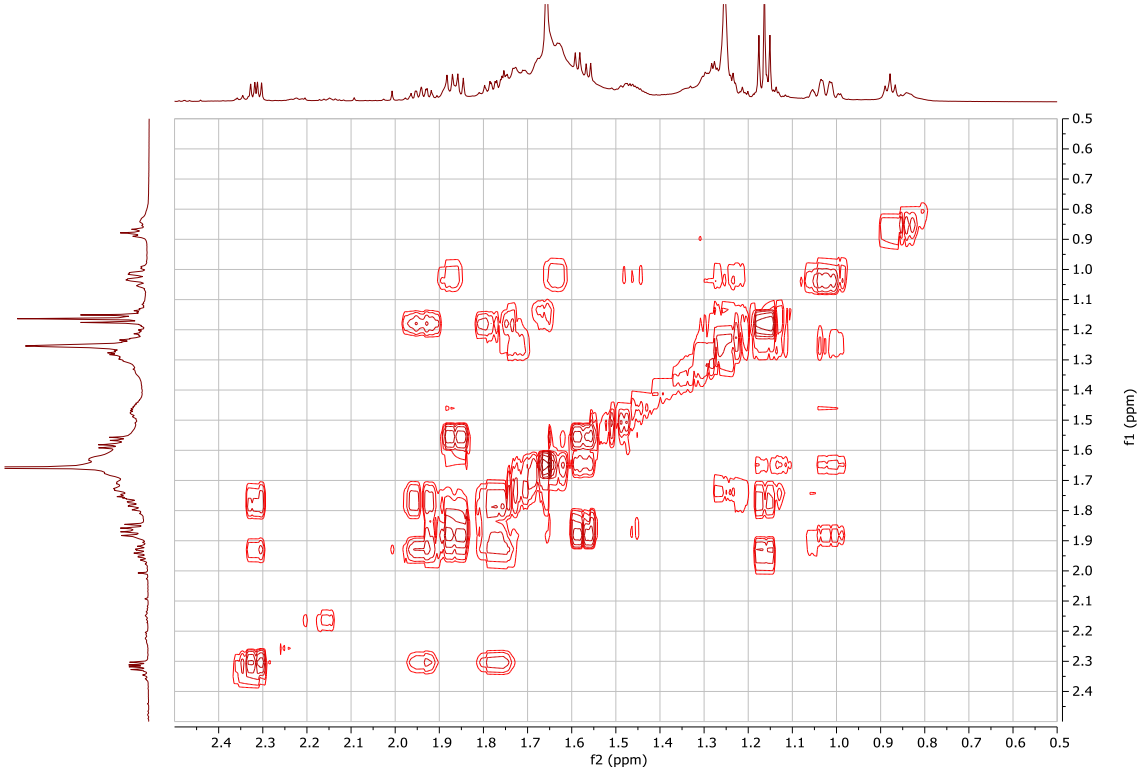
Supplementary Notes Fig. 26: HSQC spectra simplisporamide (**R-17**).



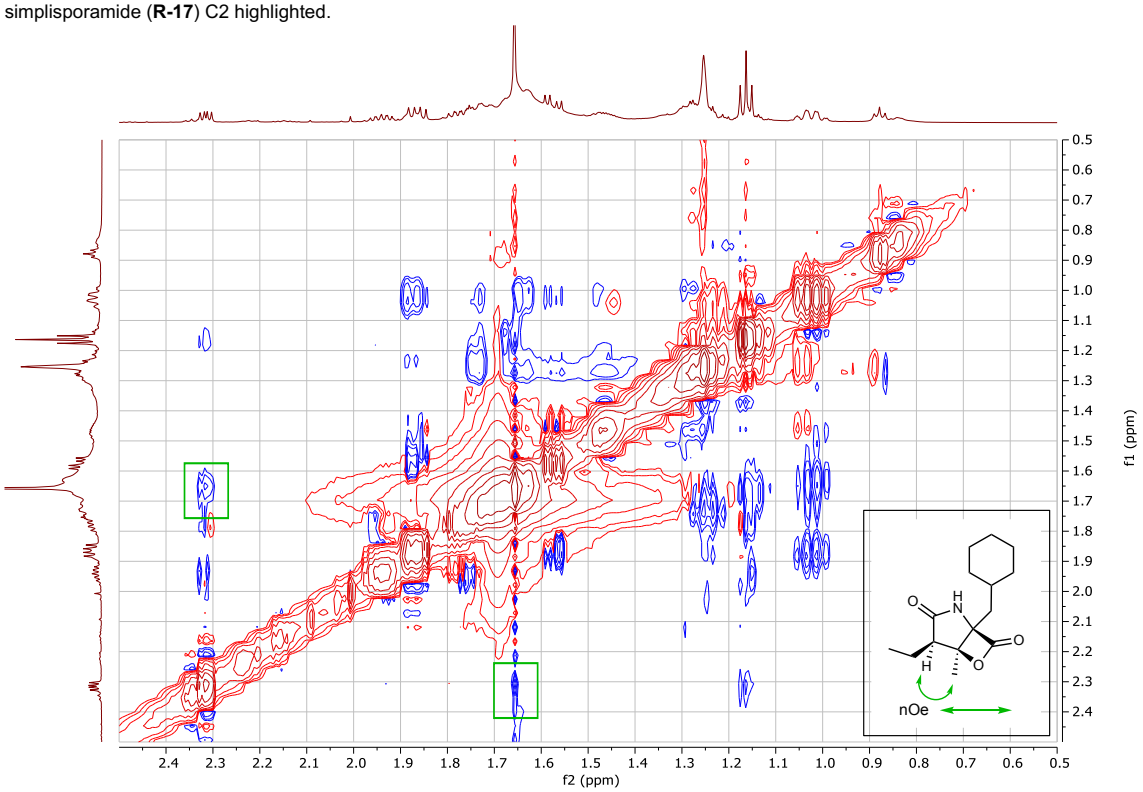
Supplementary Notes Fig. 27: HMBC spectra simplisporamide (**R-17**) with select HMBC correlations used to assign bicyclic simplisporamide (**R-17**) core highlighted.



Supplementary Notes Fig. 28: COSY spectra simplisporamide (**R-17**).



Supplementary Notes Fig. 29: NOESY spectra simplisporamide (**R-17**) with select nOe used to resolve configuration of simplisporamide (**R-17**) C2 highlighted.



3.4 Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in *Nature Chemical Biology,* Bauman, K.D., Shende, V.V., Chen, P.Y.-T., Trivella, D.B.B., Gulder, T.A.M., Vellalath, S., Romo, D., and Moore, B.S.. The dissertation author was the primary researcher and author of this paper.

CHAPTER 4. Ongoing investigations of the salinosporamide pathway

4.1 Introduction and context for Chapter 4

The marine natural product salinosporamide A is a potent proteasome inhibitor with the unique ability to cross the blood-brain-barrier. Because of this ability, salinosporamide A is presently in phase III clinical trials for the treatment of glioblastoma, a deadly brain cancer with few therapeutic options.[1] Our recent goal is to repurpose this molecule into new drug leads that maintain their brain permeability but selectively target alternative proteasomes. While the constitutive proteasome is a validated target for cancer therapy,[2] recent work has demonstrated that inhibition of the immunoproteasome is successful in the treatment of autoimmune and inflammatory diseases,[3–5] and that proteasomes of pathogenic organisms are intriguing targets for infectious diseases.[6] These proteasomes expand the druggable space for a variety of diseases of the brain ranging from Huntington's Disease to fungal meningitis. This highlights the need for continued development of brain-penetrant proteasome inhibitors with selectivity for these targets.[7] While salinosporamide A is powerfully bioactive against the human constitutive proteasome, SAR studies have revealed salinosporamide analogs that lack this activity. This suggests that these molecules may be selective inhibitors of other proteasomes, and thus may serve as potential therapeutics for other diseases.

However, this work has been stalled by lack of access to material as salinosporamide analogs are incredibly difficult to access synthetically and are produced in excruciatingly low titers by fermentation. However, our recent discovery and characterization of the salinosporamide bicyclase SalC gives us a biochemical route to install the most synthetically challenging portion of the molecule, which means a chemoenzymatic route to salinosporamide production is within reach. My work in the

Moore lab has been focused on two alternative approaches to access salinosporamide diversity: an in vitro biocatalytic strategy for the chemoenzymatic preparation of novel salinosporamides, and an in vivo strategy to develop a library of designer strains engineered for the production of bespoke salinosporamides. This chapter is intended as a progress report on the current status of these ongoing projects and a roadmap for future lab members interested in this work.

4.2 References for Chapter 4 introduction

1. Fenical, W., Jensen, P. R., Palladino, M. A., Lam, K. S., Lloyd, G. K. & Potts, B. C. Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* **17**, 2175–2180 (2009).

2. Adams, J., Palombella, V. J., Sausville, E. A., Johnson, J., Destree, A., Lazarus, D. D., Maas, J., Pien, C. S., Prakash, S. & Elliott, P. J. Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res.* **59**, 2615–2622 (1999).

3. Basler, M., Dajee, M., Moll, C., Groettrup, M. & Kirk, C. J. Prevention of experimental colitis by a selective inhibitor of the immunoproteasome. *J. Immunol.* **185**, 634–641 (2010).

4. Muchamuel, T., Basler, M., Aujay, M. A., Suzuki, E., Kalim, K. W., Lauer, C., Sylvain, C., Ring, E. R., Shields, J., Jiang, J., Shwonek, P., Parlati, F., Demo, S. D., Bennett, M. K., Kirk, C. J. & Groettrup, M. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat. Med.* **15**, 781–787 (2009).

5. Jenkins, T. W., Downey-Kopyscinski, S. L., Fields, J. L., Rahme, G. J., Colley, W. C., Israel, M. A., Maksimenko, A. V., Fiering, S. N. & Kisselev, A. F. Activity of immunoproteasome inhibitor ONX-0914 in acute lymphoblastic leukemia expressing MLL–AF4 fusion protein. *Sci. Rep.* **11**, 10883 (2021).

6. Bibo-Verdugo, B., Jiang, Z., Caffrey, C. R. & O'Donoghue, A. J. Targeting proteasomes in infectious organisms to combat disease. *FEBS J.* **284**, 1503–1517 (2017).

7. Groll, M., Huber, R. & Potts, B. C. M. Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of β-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136–5141 (2006).

4.3 Introduction

The ubiquitin-proteasome pathway (UPP) is the major route for protein disposal in eukaryotes, responsible for the degradation of over 80% of cellular proteins.[1] The degradation process begins when three enzymes work in concert to sequentially attach ubiquitin to a protein, creating a polyubiquitin tag that marks that protein for degradation by the proteasome. The proteasome (**Figure 4.1a**) is a complex proteolytic molecular machine that serves as a sort of cellular trashcan, degrading proteins into a mixture of short peptides. These peptides can be further broken down into individual amino acids by various cellular enzymes, or can be bound by major histocompatibility complex (MHC) class I molecules as epitopes, thus playing a critical role in the body's immune system.[2,3]

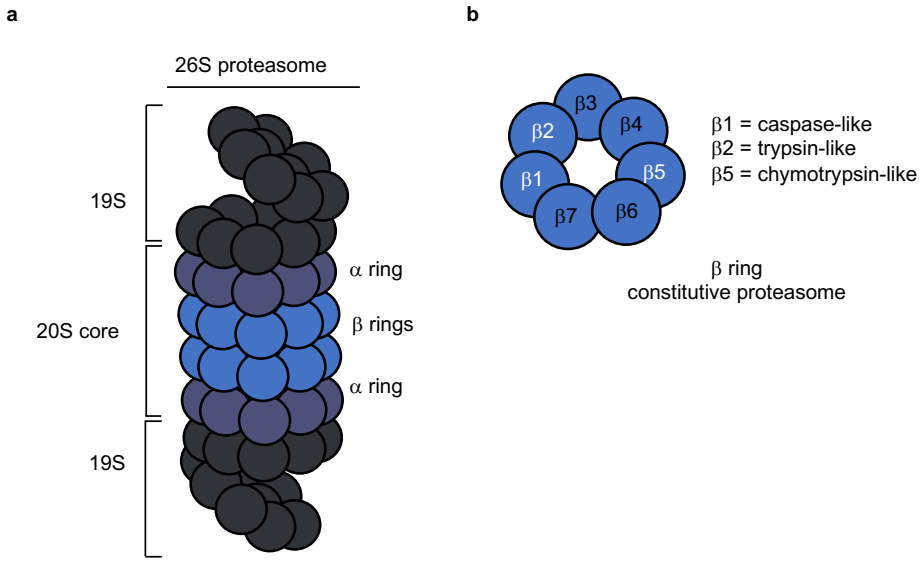


Figure 4.1 The 26S proteasome. a. The 26S proteasome is comprised of the 20S core particle, made of α and β rings, capped by two 19S regulatory particles. b. β-subunits are organized in rings of seven. The three catalytic subunits, β1, β2, and β5 are highlighted in white.

In humans, the 26S constitutive proteasome is a highly abundant protein, found in all tissue types and comprised of a 20S core capped by two 19S regulatory particles (**Figure 4.1a**). The 20S core is a large (~700 kDa) and well-organized barrel-shaped protein consisting of four rings made up of seven subunits each. The two outer rings (α rings) are identical and non-catalytic, while the inner β rings contain catalytically active subunits that exhibit caspase-like (β1), trypsin-like (β2), and chymotrypsin-like (β5) proteolytic activities (**Figure 4.1b**). Ubiquinated proteins are recognized by the 19S particles and fed into the hollow barrel of the 20S for proteolysis.

Due to the essential nature of protein degradation, the proteasome is critical for maintaining cellular homeostasis and has therefore been identified as an important target for cancer therapeutics. There are currently three FDA-approved proteasome inhibitors that are used for the treatment of blood cancers: bortezomib (**1**), ixazomib (**2**), and carfilzomib (**3**), (**Figure 4.2**).[1] All three compounds are small peptides with electrophilic warheads (boronate or epoxyketone moieties) that covalently modify the proteasome via the N-terminal active site threonine (Thr1) and have various side chains that interact with the S pockets of the proteasome. Of particular interest in the Moore lab is the proteasome inhibitor salinosporamide A, also known as Marizomib (**4**), that was discovered at SIO and has long been a subject of biosynthetic investigation in our lab.[4,5] Marizomib is fundamentally structurally different from the approved inhibitors; it is the only nonpeptidic inhibitor in advanced clinical trials and utilizes a β-lactone electrophilic moiety as its warhead. Importantly, **4** targets all three catalytic subunits of the proteasome (β1, β2, and β5), and has the lowest IC50 value among the approved inhibitors.[6] Marizomib is also

capable of crossing the blood-brain-barrier, resulting in its advancement through clinical trials for glioblastoma, a brain cancer with a poor prognosis and few therapeutic options.

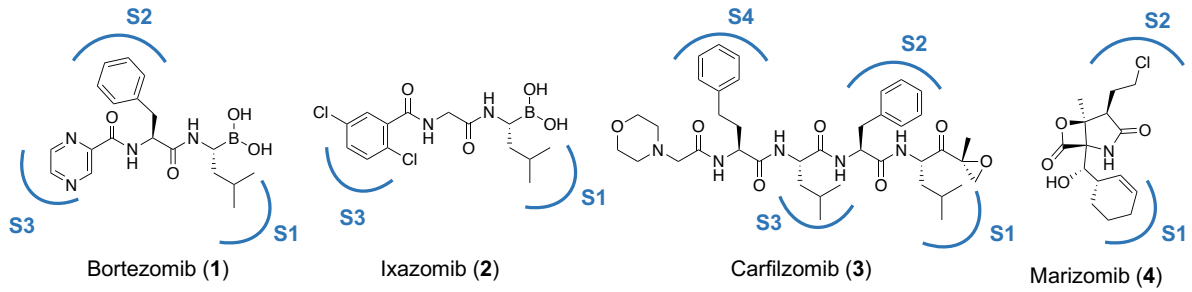


Figure 4.2 Proteasome inhibitors. FDA improved inhibitors, bortezomib (first in class), ixazomib, and carfilzomib. Marizomib is currently in phase III clinical trials. Interactions with the S pockets of the proteasome are depicted in blue.

While the 26S proteasome is essential for cellular homeostasis and is now recognized as an important therapeutic target, humans actually have two additional proteasomes: the immunoproteasome (iCP), an inducible form of the proteasome differentially expressed in response to cytokines throughout the body and constitutively expressed in immune cells, and the thymoproteasome (tCP), found solely in the thymus, a lymphatic organ that is responsible for the production of the body's T-cells. These two alternative proteasomes are structurally distinct from the 26S constitutive proteasome, which results in different catalytic activities and cleavage specificities. The iCP incorporates three lower molecular weight proteolytic β-subunits (β1i, β2i, and β5i) (**Figure 4.3**), and preferentially hydrolyzes proteins after nonpolar amino acids. This results in peptides with hydrophobic C-termini that are ideal for anchoring to MCH molecules, thereby shaping the immune system to a greater degree than the peptides produced by the constitutive proteasome.[1,7] The tCP incorporates β1i, β2i, and β5t

subunits, which has reduced chymotrypsin activity, and the resulting peptides induce positive selection of $CD8^+$ T cells.[8]

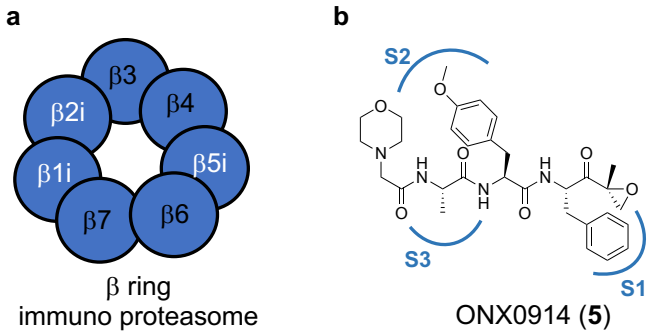


Figure 4.3 The immunoproteasome. a. β-subunit of the immunoproteasome with incorporation of low molecular weight immunoproteasome specific subunits β1i, β2i, and β5i. b. Immunoproteasome specific inhibitor ONX0914. Interactions with S pockets of the immunoproteasome are shown in blue.

Because of the differential expression of the immunoproteasome and its connection to the immune system, inhibition of the iCP has shown therapeutic success in the treatment of autoimmune and inflammatory diseases.[9–11] Notably, selective inhibition of the iCP avoids toxicities seen during treatment with constitutive proteasome inhibitors that can cause on-target inhibition of proteasomes in non-lymphoid tissue (i.e. the brain, GI tract, etc).[11–13] In 2016, the first selective immunoproteasome inhibitor, an epoxyketone bearing peptide (KRZ-616, an analog of ONX0914 (**5**), **Figure 4.3**), entered clinical trials for the treatment of lupus, an autoimmune disease with no cure.

Key structural differences between the immuno and constitutive proteasomes explain the selectivity of inhibitors.[7] The S1 pocket of the iCP β1 subunit (β1i) is smaller and more hydrophobic than observed in the corresponding constitutive proteasome subunit (β1c), therefore preferentially accommodating amino acids with branched,

nonpolar side chains. This results in the observed differential proteolytic activity of the iCP, where peptide bond hydrolysis preferentially occurs after small, hydrophobic, residues which are generated for presentation on MHC-I molecules. However, the key structural differences that explain the enhanced affinity of iCP selective inhibitors are found in the β5 subunit. β5i showcases a shallower S2 pocket and a smaller, more hydrophilic S3 pocket. Most importantly, however, were the conformational changes observed by residue Met45, which results in a spacious S1 pocket in β5i and a smaller pocket in β5c. These differences in the S1 pocket, and in particular the orientation of Met45, are responsible for OXN0914's 15-fold selectivity for β5i over β5c. The large, hydrophobic phenyl moiety in the P1 position of OXN0914 exploits the more spacious S1 pocket of β5i to gain selectivity for the immunoproteasome. These structural differences, supported by inhibitory activity data, show that generating iCP selective inhibitors is an exciting and therapeutically relevant avenue for drug discovery efforts.

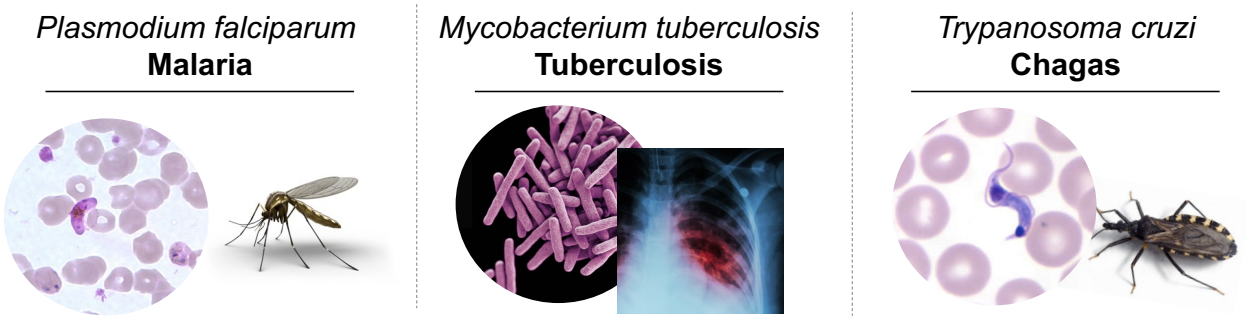


Figure 4.4 Infectious disease targets for proteasome inhibition.

Beyond the constitutive proteasome as a target for cancer and the immunoproteasome as a target for autoimmune and inflammatory diseases, proteasomes found in other organisms represent intriguing therapeutic targets for a host of other diseases (**Figure 4.4**).[14] Notably, bacteria in the *Mycobacterium* genus have proteasomes, including the pathogenic *Mycobacterium tuberculosis* (Mtb) and *Mycobacterium leprae,* the causative agents of tuberculosis and leprosy, respectively. A series of oxathiazole-2-one compounds have been identified as highly selective mycobacterium proteasome (MtbCP) inhibitors with 1,000 fold selectivity for the MtbCP over the human constitutive proteasome.[15] The proteasome of the protozoan parasite *Plasmodium falciparum* has likewise been an exciting new target in the fight against malaria especially as resistance to frontline drugs like artemisinin has become firmly established.[16,17] Structural evidence has revealed an unusually open β2 active site in the *P. falciparum* proteasome, and β2 selective inhibitors have shown selective and potent inhibition of the parasite's proteasome.[17] Furthermore, the proteasome may serve as an important target for additional neglected tropical diseases caused by protozoan parasites, including trypanosomiasis (sleeping sickness), Chagas disease, and leishmaniasis.[18,19] These so-called 'alternative' proteasomes expand the druggable space for a variety of diseases ranging from autoimmune disorders to malaria and highlights the need for continued development of proteasome inhibitors with selectivity for these targets. As seen with OXN0914, small changes in a peptide can yield tremendous selectivity for one proteasome over another, which suggests that building libraries and testing analogs of validated proteasome inhibitors may be a good starting point towards this selectivity.

Previous work by the Moore lab and others have generated large libraries of salinosporamide analogs via mutasynthesis, total synthesis, and extensive culturing of the native producer *Salinispora tropica*.[20–26] These compounds incorporate different amino acids, opposite stereochemistries, alternative side chains, and varying degrees of halogenation and oxidation (**Figure 4.5**). While these analogs were beneficial for SAR studies, none of these molecules showed constitutive proteasome inhibitory activity on par with **4**, and thus it was the originally isolated natural product itself that has advanced through clinical trials. However, it is precisely this poor activity against the constitutive proteasome that suggests these other salinosporamide analogs may make incredibly selective inhibitors of alternative proteasomes. These analogs retain the key pharmacophore for proteasome inhibition, the γ-lactam-β-lactone, and simply vary in side chains that interact with the S pockets of the proteasome. This bioactivity testing is something we are eager to pursue in collaboration with Professor Anthony O'Donoghue here at UCSD, and Professor Michael Groll at the Technical University of Munich.

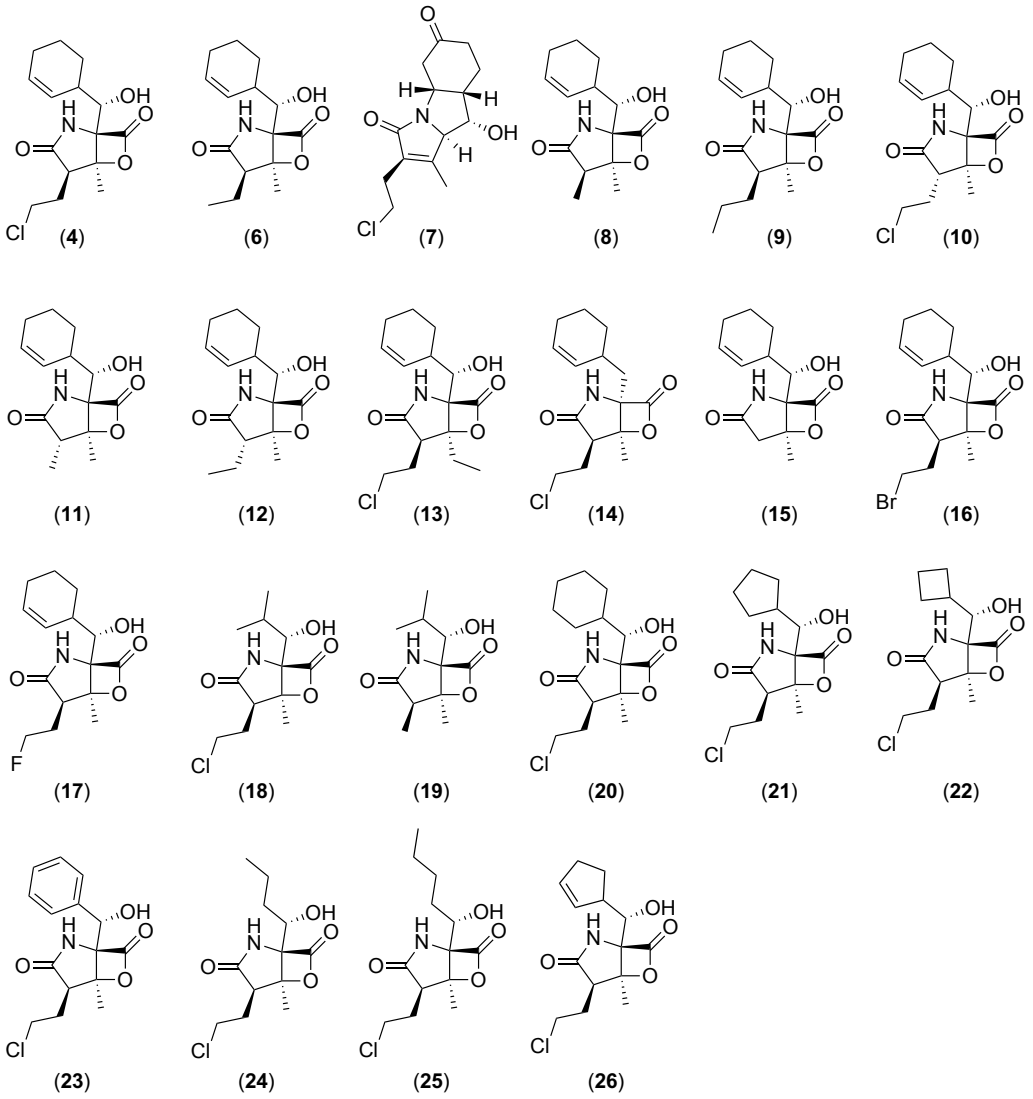


Figure 4.5 Salinosporamide analogs produced via fermentation and mutasynthesis. **4** = salinosporamide A, **6** = salinosporamide B, **7** = salinosporamide C, **8** = salinosporamide D, **9** = salinosporamide E, **10** = salinosporamide F, **11** = salinosporamide G, **12** = salinosporamide H, **13** = salinosporamide I, **14** = salinosporamide J, **15** = salinosporamide K, **16** = bromosalinosporamide, **17** = fluorosalinosporamide, **18** = antiprotealide, **19** = omuralide, **20** = salinosporamide X1, **21** = salinosporamide X2, **22** = salinosporamide X3, **23** = salinosporamide X4, **24** = salinosporamide X5, **25** = salinosporamide X6, **26** = salinosporamide X7.

However, these salinosporamide analogs are incredibly difficult to access synthetically, and are produced in excruciatingly low titers by fermentation. For example, of particular interest, is the compound known as salinosporamide X4 (**23**), which has a phenyl group in place of salinosporamide A's cyclohexenyl ring.[21] Importantly, salinosporamide X4 shows 400-fold decreased activity against the constitutive proteasome than salinosporamide A**.** However, the iCP has a more open, spacious S1 pocket that is differentially inhibited by bulkier, more hydrophobic residues.[1] The phenyl ring of **5** sits in the S1 pocket and explains the specificity of that molecule for the iCP, which suggests that **23** may show this same specificity. However, this molecule is currently only accessible via mutasynthesis, resulting in titers less than 0.1 mg/L. Additionally, beyond the compounds produced naturally, we are interested in generating new-to-nature salinosporamides for bioactivity testing. We are particularly interested in creating analogs with alternative side chains in place of the C3 methyl group, as only one molecule with a substitution in this position has ever been isolated. This compound, salinosporamide I (**13**), bears an ethyl group in place of the C3 methyl, and has 1000 fold decreased activity compared to salinosporamide A against the constitutive proteasome, which suggests it and similar analogs may make extremely selective inhibitors of other proteasomes and leave the constitutive proteasome untouched.

Therefore, my current goal in the Moore laboratory is to develop alternative routes to access additional salinosporamide diversity. Here, I will discuss two complementary approaches to access salinosporamide diversity: an in vitro strategy using SalC as a biocatalyst for the chemoenzymatic preparation of novel salinosporamides, and an in vivo genetic engineering based strategy for the development of a library of heterologous host

strains that produce bespoke salinosporamides. These strategies will ultimately allow us to access salinosporamide diversity for testing against alternative proteasomes in collaboration with the Groll and O'Donoghue labs.

4.4 Methods

Bacterial strains

All *E. coli* strains were grown on LB agar or in LB broth with shaking at 37 °C. Antibiotics were used for selection as appropriate: kanamycin 50 μg/mL, apramycin 50 μg/mL, tetracycline 8 μg/mL, ampicillin 100 μg/mL. For protein production, TB media (1.2% w/v tryptone, 2.4% w/v yeast extract, 0.4% (v/v) glycerol, 2.31% w/v KH_2PO_4, 12.54% w/v K_2HPO) was used.

Attempted expression of SalC in *E. coli*

Codon optimized SalC from four strains (*Salinispora tropica* CNB440, *Streptomyces cinnabarigriseus* JS360, *Salinispora pacifica* CNS237 and *Salinispora pacifica* CNT133) were ordered from Twist Biosciences cloned into pET28a vectors. *E. coli* BL21(DE3) competent cells were transformed with expression plasmids. Transformants were inoculated into 10 mL LB with kanamycin and incubated overnight at 37 °C with 220 rpm shaking. The following morning, the entire 10 mL overnight culture was used to inoculate 1 L of TB in a 2.8 L baffled flask with kanamycin. This culture was incubated at 37 °C with shaking at 220 rpm until OD600 reached ~0.8 when the culture was subsequently cooled to 18 °C and induced with isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were incubated at 18 °C overnight, and

harvested by centrifugation the following morning. All subsequent steps were performed at 4 °C or on ice. The cell pellet was resuspended in lysis buffer (100 mM Tris, 300 mM NaCl, 0.8 mM TCEP, 10% glycerol) and sonicated on ice with a Qsonica sonicator (6 mm tip at 40% amplitude). Lysate was then clarified by centrifugation at 14,000 x *g* for 45 min at 4 °C. Successful protein expression was evaluated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis in the presence of reducing agents (SDS-PAGE, 12% acrylamide). Both the pellet and the supernatant were evaluated for expression of SalC.

SalB protein expression

The SalB didomain protein (A-PCP) was cloned into multiple cloning site 1 of pET-DUET-1 (N-terminal His tag), along with the MbtH-like protein SalE which was cloned into multiple cloning site 2. This expression plasmid was transformed into *E. coli* BL21(DE3) *ybdZ::acc(IV)* and expressed the same way as described above, with ampicillin for selection.

Protein purification was performed on an ÄTKApurifier FPLC (GE Healthcare Life Sciences) at 4 °C. Proteins were purified by Ni^{2+}-affinity chromatography using 5 mL HisTrap FF (GE Healthcare) columns pre-equilibrated in buffer A (1 M NaCl, 20 mM Tris, 25 mM imidazole, pH 8.0). Lysate was then loaded onto the column at 1.5 mL/min after which the column was washed with 8 column volumes (40 mL) of buffer A. Protein was eluted by a linear gradient of 0 – 100% buffer B (1 M NaCl, 20 mM Tris, 250 imidazole, pH 8.0) over 30 min at a flow rate of 2.0 mL/min. Protein-containing fractions were identified by SDS-PAGE gel (12% acrylamide) and then combined, concentrated and

buffered exchanged (50 mM Tris, 500 mM NaCl, 10% glycerol, pH 8.0) using Amicon Ultra centrifugal filters with a 50 kDa molecular weight cut-off (EMD Millipore).

While SalE was coexpressed with SalB, I also cloned it into pET28a with an N-terminal His tag and purified it by itself using the same methods described above.

SalB activity assays

SalB (A-PCP) was converted from *apo* to *holo* SalB (A-PCP) in a 50 μL reaction containing 20 μM SalB, 10 mM $MgCl_2$, 0.8 mM TCEP, 1 mM CoASH, and 1 μM Sfp. Conversion was monitored via trypsin digestion using Proteomics Grade trypsin singles (Sigma-Aldrich) following manufacturer's directions. Resulting peptide fragments were analyzed with an Agilent 1260 Infinity LC system coupled to an Agilent 6530 Accurate-Mass Q-TOF using a Phenomenex Aeris WIDEPORE XB-C18 200 Å, LC column (3.6 μM, 250 mm x 4.6 mm).

Adenylation domain assays were performed in 50 μL reactions including 20 μM SalB (A-PCP), 10 mM $MgCl_2$, 1 mM CoASH, 10 μM Sfp, 20 μM His-tagged SalE, 1 mM cyclohexylalanine, and 1 mM ATP. The assay was incubated at room temperature for 2 hours and then delivered to the Molecular Mass Spectrometry Facility in the UCSD Chemistry Department for intact proteomics experiments. A control reaction containing just 20 μM SalB(A-PCP) was also analyzed.

Cas12a assisted CAPTURE method

Genomic DNA from *Salinispora tropica* CNB-440 was extracted following methods described by Enghiad et al.[27] Briefly, *Salinispora* cultures were harvested prior to

sporulation and resuspended in harvest buffer (50 mM Tris, 25 mM EDTA, pH 8.0) with lysozyme and RNAse A and incubated for 2 hrs at 37 °C. Proteinase K was added to the mixture and incubated for an additional 1 hr, followed by addition of 1.2 mL of 10% SDS and an additional incubation for 2 hrs at 50 °C. The sample was then extracted with phenol:chloroform:isoamyl alcohol, followed by pure chloroform. gDNA was precipitated with isopropanol, washed with 70% ethanol, and rehydrated in 10 mM Tris, pH 8.0.

*Fn*Cas12a was purified from *E. coli* BL21(DE3) following standard protocols. Cells were grown in TB media with kanamycin and induced with 1mM (final concentration) IPTG and 0.1% (w/b) L-rhamnose when OD_{600} reached 1.2-1.5. Cells were harvested the following morning, resuspended in 20 mM Tris, 1 M NaCl, 20 mM imidazole, pH 8.0 buffer, and lysed by sonication. Protein was purified by Ni^{2+}-affinity chromatography using a 5 mL HisTrap FF (GE Healthcare) column on the FPLC. Protein was eluted by a linear gradient of 20 – 500 mM imidazole. Fractions containing protein were identified by 12% SDS-PAGE, pooled, and concentrated using 50 kDa Amicon Ultra Spin Filters.

The BGC capture experiment was completed over three days. On the first day, DNA receivers were PCR amplified using Q5 polymerase, DpnI digested, gel purified, and stored at -20 °C for future use. ssDNA oligonucleotides for gRNAs (two per BGC, one upstream and one downstream of BGC) were hybridized to generate dsDNA by incubating forward and reverse templates together for 5 min at 98 °C, followed by slow temperature decrease of 0.1 °C/sec until 10 °C was reached. Resulting dsDNA was the template for RNA transcription overnight using the HiScribe T7 Quick Yield RNA Synthesis Kit (NEB). Finally, DH10B *E. coli* were transformed with the circularization plasmid pBE14 and incubated overnight at 30 °C.

The second day, gRNAs were purified using the Zymo RNA Clean and Concentrator Kit, and DH10B-pBE14 were inoculated into liquid cultures overnight at 30 °C. On the third day, gDNA was digested with Cas12a in a reaction containing 15 μg gDNA, 2100 ng each gRNA, 60 pmol *Fn*Cas12a, and 30 μL NEBuffer3.1. Digested gDNA was purified using the same method as described previously, except with an ice cold ethanol DNA precipitation. Digested gDNA was then assembled into a linear DNA molecule with the two DNA receivers purified previously following the Zhao lab T4 exo+fill method. The resulting assembly mixture was dialyzed against ddH_2O using membrane filters prior to transfromation. Meanwhile, DH10B-pBE14 competent cells are prepared at room temperature, following induction with L-arabinose, and transformed with the DNA assembly mixture. Transformants were screened by colony PCR and positive clones were confirmed by restriction digest.

4.5 Results and current status

In vitro strategies for accessing salinosporamide diversity

My recent work, discussed in Chapter 3 of this dissertation, established that an unusual ketosynthase (KS), SalC, is responsible for the bicyclization reaction that generates the salinosporamide γ-lactam-β-lactone pharmacophore.[28] This work expanded the known types of chemical transformations performed by KSs, revealed a new route to β-lactone formation, and opened the door towards using SalC as a biocatalyst in the chemoenzymatic (bio)synthesis of the salinosporamides. Using SalC as a biocatalyst is tempting for a number of reasons. First, SalC already appears capable of relatively promiscuous cyclization activity. Unlike many canonical assembly line systems,

in which a simple scaffold is assembled and then functionalized post release, the salinosporamide pathway generates three functionalized pieces and then fuses them together in a single step performed by SalC. This is evidenced in the wealth of salinosporamide diversity that can be accessed by mutasynthesis or that is simply produced during fermentation. SalC is a constant in the construction of these compounds, capable of cyclizing a diverse set of linear substrates to yield the distinctive core structure. This broad substrate scope suggests that SalC may be a useful biocatalyst for generating salinosporamide diversity. Furthermore, SalC appears capable of resolving a diastereomeric substrate mixture, a key advantage of enzymatic catalysis over chemical catalysis. Finally, the successful use of a proteasome inhibition assay as a means to screen for β-lactone formation by SalC is additional evidence that a high-throughput engineering campaign for optimized SalC activity is feasible.

However, while seemingly promising, using SalC as a biocatalyst remains problematic. Most concerningly, SalC has only been successfully expressed in *Streptomyces.* While this does result in soluble protein expression (~8 mg/L), *Streptomyces coelicolor* is slow growing and requires DNA transfer by conjugation. Any protein used for biocatalysis will require engineering for optimization, especially if that protein is being asked to perform a chemical reaction on a nonnative substrate, which is what we would ask of SalC. The speed of *Streptomyces* protein expression is prohibitive to any sort of engineering – a month between mutant generation is entirely unsustainable. Therefore, a major goal has been to achieve soluble expression of SalC in *E. coli*. Previous postdocs have attempted expression in Rosetta(DE3), ArcticExpress, Origami, and Shuffle *E. coli* expression strains, as well as co-expression of SalC with chaperone

plasmids. I have additionally attempted expression of SalC with a maltose-binding protein (MBP) tag, and expression of SalC in a *Corynebacterium glutamicum* host,[29] all without ever seeing any expression of SalC. More recently I have attempted expression of four *E. coli* codon optimized SalC homologs: SalC from *S. tropica* CNB440, CinC from the cinnabaramide producer *Streptomyces cinnabarigriseus* JS360, and two SalC homologs from salinosporamide K producing strains *S. pacifica* CNS237 and *S. pacifica* CNT133 (**Figure 4.6**). Gratifyingly, these homologs appeared to express in *E. coli* BL21(DE3), although the protein remained in the insoluble pellet fraction. However this was the first time I had observed expression of SalC in *E. coli* at all. This solubility issue may now be solved by using additional tags, or expression in a different cell line. These are experiments currently underway in the laboratory.

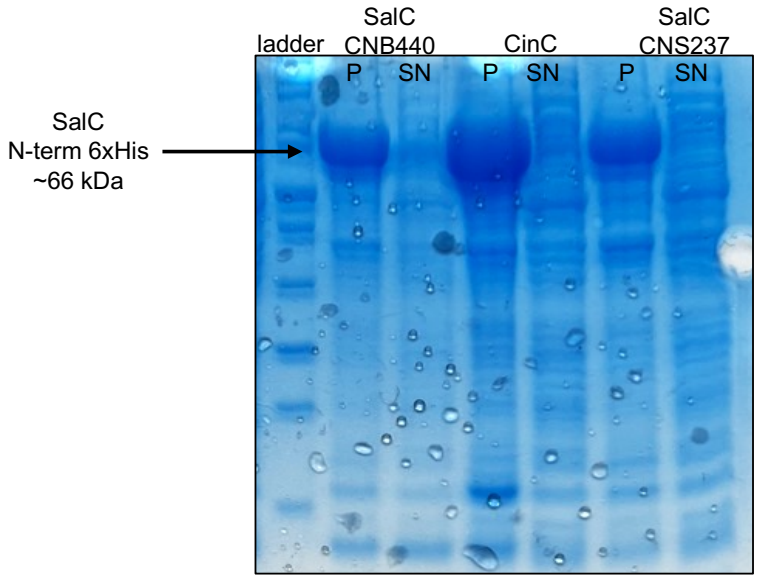


Figure 4.6 SalC protein expression in *E. coli* BL21(DE3). E. coli codon optimized SalC from *S. tropica* CNB440, *E. coli* codon optimized CinC from *Streptomyces cinnabarigriseus* JS360, SalC from salinosporamide K producer *S. pacifica* CNS237. Pellet and supernatant from lysed *E. coli* shown here.

As evidenced in Chapter 3, the major bottleneck to using SalC in a biocatalytic role is its reliance on carrier protein tethered substrate. It takes five individual proteins just to generate the substrate for SalC and results in a reaction that is single turnover with respect to the carrier protein; once the linear substrate is transferred to the active site of SalC, the resulting *holo*-PCP cannot be reused. While other KSs have been shown to accept diffusible substrates with CoA mimics or even thiophenol as activating groups,[30] SalC has a strong preference for a substrate bound to a carrier protein. Therefore, the first major goal of an engineering campaign for SalC is to engineer the protein away from PCP dependence. Critically, I have observed weak SalC bicyclization activity when using small molecule diffusible substrates (**26**-**29**) (**Figure 4.7**) which suggests a starting point for engineering SalC to preferentially accept diffusible substrates.

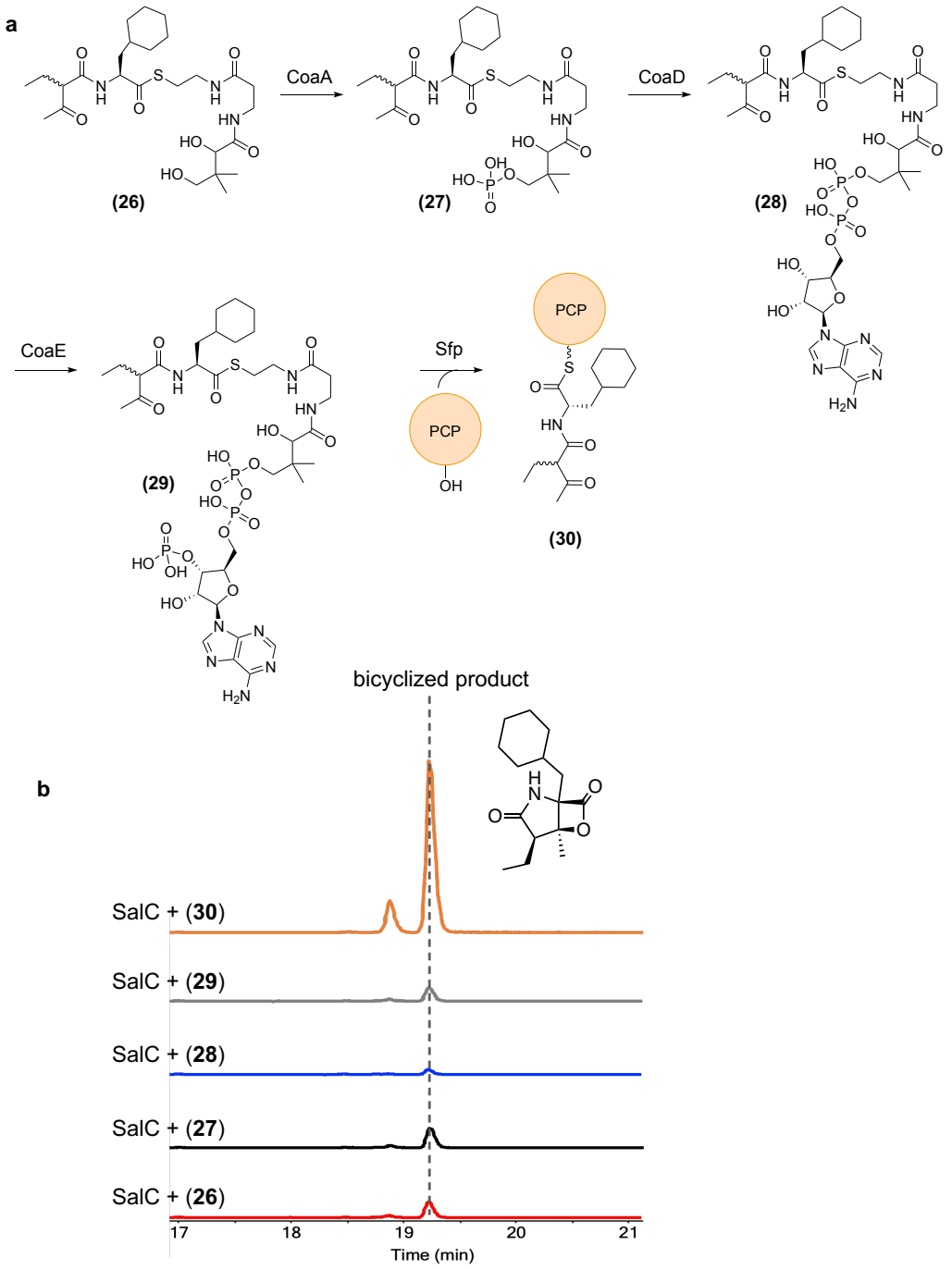


Figure 4.7 SalC activity assay with diffusible substrates. a. Reaction scheme depicting chemoenzymatic synthesis and subsequent SalB-PCP acylation assay to generate SalC substrate. b. LCMS chromatograms (EIC *m/z* 266.18) of SalC activity assay with diffusible substrates shown in part (**a**) including pantetheine-activated (**26**), CoaA product-activated (**27**), CoaD product-activated (**28**), CoA activated (**29**), and carrier protein-tethered substrate (**30**).

Once soluble protein expression in *E. coli* is achieved, both structure-guided and random-mutagenesis methods will be used for a directed evolution campaign to engineer SalC away from carrier protein dependence for ultimate use as a biocatalyst in the production of novel salinosporamides. To identify residues in SalC that are important for carrier protein interaction we are pursuing crystallography experiments to obtain a co-crystal structure of SalB and SalC. A phosphopanthethine probe with a reactive electrophile can be used to irreversibly label the KS active site cysteine, forming a cross-link between the KS and PCP domains, and allow for visualization of their interaction.[31,32] Using both targeted mutagenesis and error-prone PCR, a library of SalC variants will be generated and mutants will be subjected to bicyclization activity assays with diffusible substrates prepared synthetically as demonstrated in my recent publication.[28] Successful bicyclization will be validated using the high-throughput compatible proteasome inhibition assay that screens for formation of the reactive β-lactone moiety via a fluorescent readout. Successive rounds of mutagenesis will be performed until a mutant capable of efficient, PCP-independent bicyclization is identified.

This PCP-independent SalC mutant will then be used to generate a library of salinosporamide analogs via chemoenzymatic (bio)synthesis, with particular attention to molecules bearing alternative amino acids and side chains, as discussed earlier. These compounds can then be screened for proteasome inhibition against a variety of proteasome types. Importantly, this work has the potential to be a breakthrough for production of other therapeutically important assembly line molecules. The products of PKS and NRPS pathways are well recognized for their bioactivity,[33] but their biosynthesis requires carrier proteins, which is a massive barrier towards their chemoenzymatic

production. The work described here has the potential to be foundational in the development of carrier protein-free biosynthesis of assembly line antibiotics.

In vitro reconstitution of the *sal* pathway

As I have worked towards soluble production of SalC in *E. coli*, I have also turned my attention towards expression of the other assembly line enzymes in the salinosporamide pathway. While the in vitro reconstitution of the entire *sal* assembly line (*salA-D*) would not be a particularly efficient route to salinosporamide production, it is incredibly exciting from a structural biology perspective. The modular architecture of assembly line pathways appears well suited for engineering, where individual domains or even entire modules could be swapped for a 'plug-and-play' approach to generate tremendous molecular diversity from a single pathway. However, despite this intrinsic modularity, assembly line engineering has proved difficult, plagued by low yields and non-functional pathways due to problematic protein-protein interactions.[34,35] Understanding these dynamic interactions is therefore critical for improving future engineering efforts, and so studying structural biology of PKS and NRPS pathways is an important goal in the field of natural products biosynthesis.

Individual domains and even di-domains from NRPS or PKS pathways have been structurally characterized and well-studied for over a decade,[36–38] but full modules have proved more elusive. A major breakthrough came in 2014 when the first full length PKS module was visualized by cryo-electron microscopy (cryoEM), revealing an unexpectedly different architecture compared to the homologous dimeric mammalian fatty acid synthase.[39] For NRPS pathways, a landmark structure of the termination module from

surfactin biosynthesis was solved in 2008 by x-ray crystallography,[40] and the first holo-NRPS structures were visualized around the same time as the PKS structure.[41] More recent structural biology work has focused on the dynamic nature of these pathways, with atomic scale resolution used to study distinct conformational changes of these modules during chain elongation.[42–44] However, despite these advances, no structure of a full assembly pathway has ever been visualized. Understanding the structure, movement, and interactions of an entire assembly line will be critical for redesigning and engineering these pathways in the future.

The salinosporamide pathway presents an intriguing system to visualize an entire intact assembly line pathway for a number of reasons. First, the salinosporamide assembly line is relatively small, encompassing a loading module, two extender modules, and the unusual KS SalC, followed by a tailoring cytochrome P450 encoded by *salD*. The assembly line biosynthetic proposal for salinosporamide A construction is shown in **Figure 4.8.** Many assembly line pathways can have over a dozen modules required to create a mature product, and so it is rare that the salinosporamide A pathway can construct a bioactive molecule with just three modules. Notably, the *sal* assembly line is a hybrid pathway, incorporating both PKS and NRPS domains that interact with each other. While PKS assembly lines are dimeric, NRPS pathways are monomeric, and the 3D interaction between these differently shaped proteins has never been visualized. Finally, salinosporamide A is a biologically and clinically relevant molecule, further underscoring the importance of fully visualizing and characterizing this pathway. Because of this, we have targeted the salinosporamide pathway for full reconstitution and eventual visualization by cryo-EM and/or X-ray crystallography in collaboration with a structural

biology lab. Towards this end I have been working towards the expression and purification of the proteins that make up the *sal* assembly line.

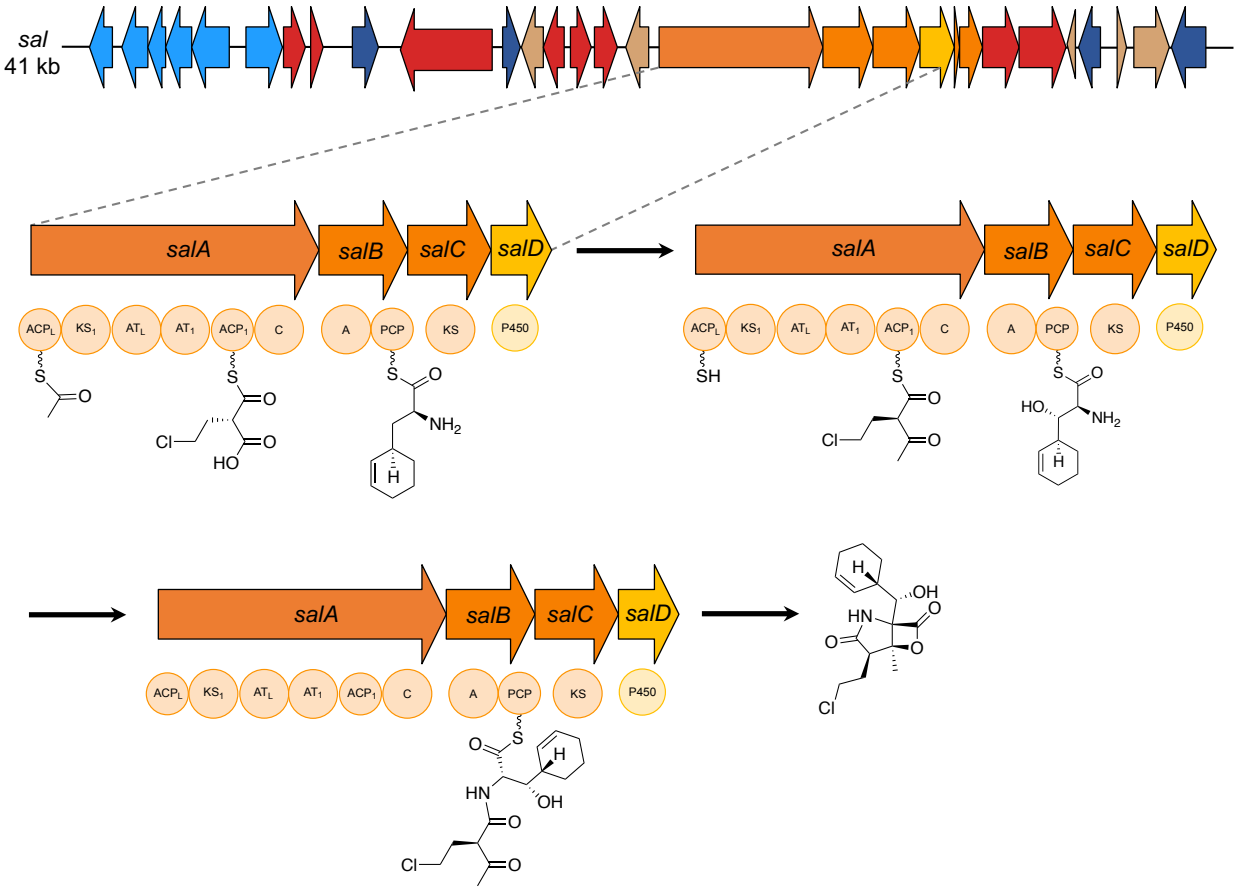


Figure 4.8 The salinosporamide assembly line. Biosynthetic proposal for the assembly of salinosporamide A by the multimodular PKS *salA,* the didomain *salB,* the bicyclizing KS *salC,* and the P450 *salD.* These four proteins encompass the *sal* assembly line.

SalA

The bimodular PKS SalA is the largest protein (~220 kDa) in the *sal* pathway, dominating the BGC, and is characterized by an unusual organization of its six domains. The protein harbors a loading ACP (ACP_L) domain, a KS domain (KS_1), two contiguous acyltransferase domains for the selection of the starter unit (AT_L) and the extender unit

(AT_1), a second ACP (ACP_1), and finally terminates in the NRPS condensation domain (C_1). At 220 kDa, SalA comprises over half the molecular weight of the *sal* assembly line.

I have successfully cloned SalA into pET28a(+) between the XhoI and NdeI sites, resulting in an N-terminal His tag, and verified the sequence integrity of the insert by Sanger sequencing (**Figure 4.9**). I am planning to attempt protein expression in BL21(DE3) cells, as well as BAP1 *E. coli* cells, a strain specifically designed for complex natural product biosynthesis and encodes a chromosomal copy of the promiscuous phosphopantetheinyl transferase *sfp.* Recently, the Keatinge-Clay lab was able to purify a ~230 kDa and a ~140 kDa polypeptide to reconstitute the entire venemycin PKS in vitro using a strain of *E. coli* originating from BAP1.[45] These proteins were His-tagged and purified using standard Ni-NTA columns, followed by size exclusion chromatography. I plan to attempt a similar procedure for the purification of SalA.

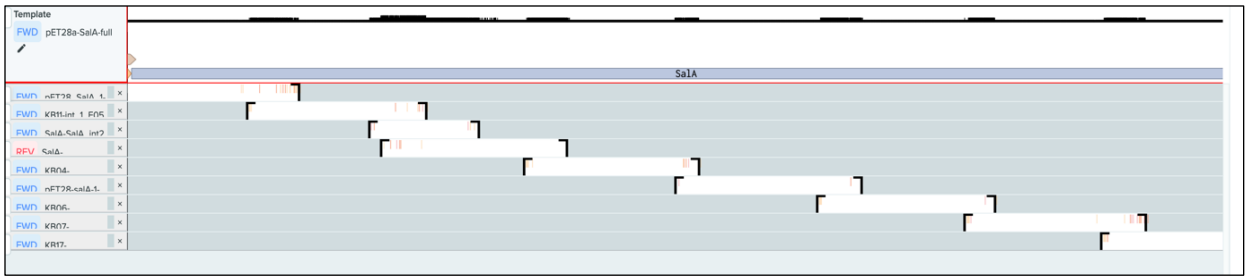


Figure 4.9 Sequence alignment of *salA* and Sanger sequencing data. *salA* was cloned into pET28a and sequence integrity was validated by primer walking and Sanger sequencing.

SalB

SalB is a didomain protein, comprising an adenylation domain and a PCP domain around 67.5 kDa in size. In my previous work with SalC I created a truncated version of this protein, excising just the carrier protein for acyl transfer of the linear substrate onto

SalC. I attempted multiple truncations to purify soluble carrier protein, and was ultimately able to express just the carrier protein domain in *E. coli.* Native SalB, however, also contains an adenylation domain upstream of this PCP. Adenylation domains catalyze amino acid activation in NRPS biosynthesis via ATP-dependent adenylation of an amino acid followed by transfer onto the downstream carrier protein through a thioester linkage (**Figure 4.10**).

Purification of the SalB didomain, however, proved more difficult. Previous Moore lab members attempted purification of the entire SalB (A-PCP) protein in the same set of strains and conditions used for attempted SalC expression, also with no success. However, I was able to achieve soluble expression of the SalB didomain (12 mg/L) by co-expressing SalB in a pETDUET vector with SalE, an MbtH-like protein from the *sal* BGC. I expressed this plasmid in *E. coli* BL21(DE3) *ybdZ::acc(IV)* gifted by the Thomas lab,[46] an expression strain with the native *E. coli* MbtH-like gene *ybdZ* disrupted by an apramycin resistance cassette. The exact role of MbtH-like proteins remains unknown, but they have been shown to be important for the solubility of NRPS proteins and can have an effect amino acid activation by A domains.[46,47]

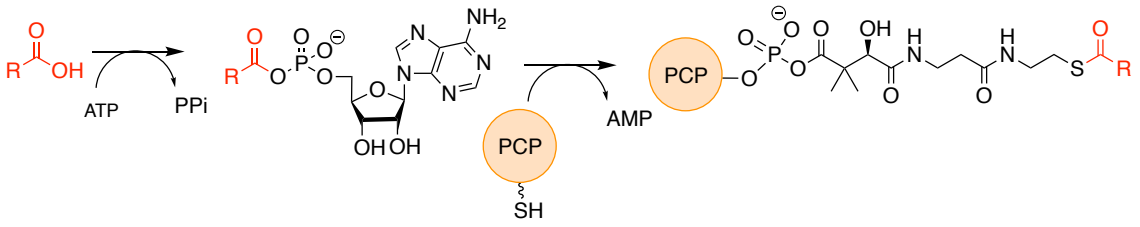


Figure 4.10 Enzymatic reaction catalyzed by adenylation domain. First, an amino acid is adenylated using ATP and then loaded onto the downstream peptidyl carrier protein (PCP) via a thioester linkage.

The identity of the purified SalB(A-PCP) didomain protein was confirmed by trypsin digestion and peptide mass fingerprinting (**Figure 4.11**). I was able to observe the same peaks originating from the PCP described in our recent publication, corresponding to the *apo* peptide fragment and a very small amount of *holo* peptide due to the activity of native *E. coli* PPTases. To generate more *holo* carrier protein, I incubated the SalB(A-PCP) protein with Sfp and coenzyme A, and upon trypsin digestion could observe a major increase in the peptide fragment corresponding to *holo*-SalB, indicating that the carrier protein can be enzymatically converted from *apo* to *holo* (**Figure 4.11**).

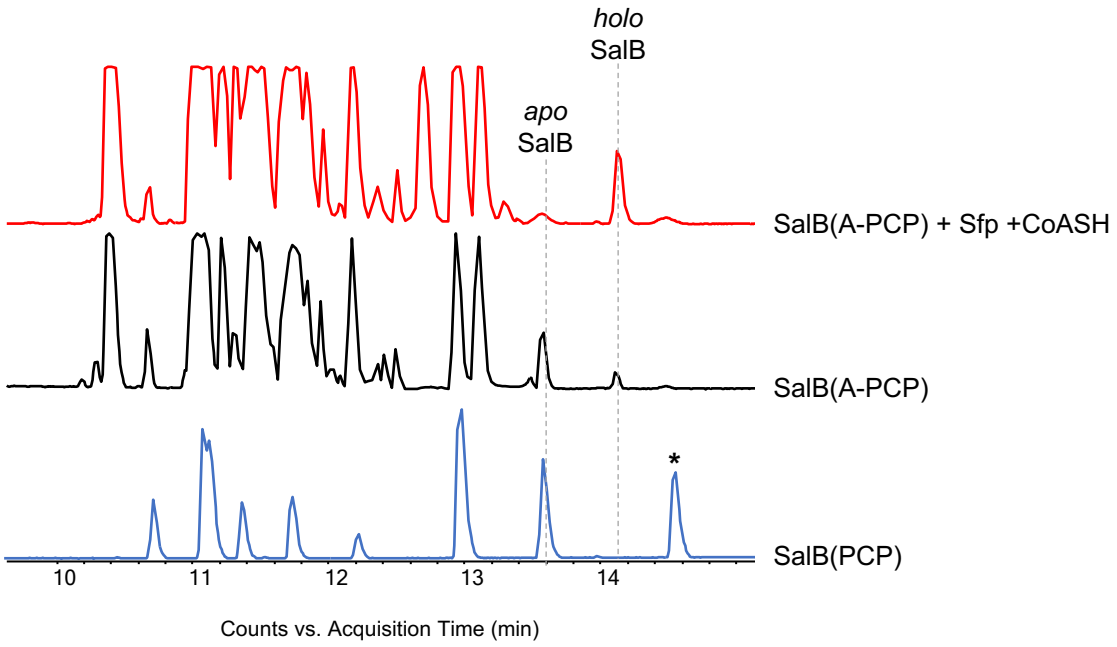


Figure 4.11 Peptide mass fingerprinting of trypsin digested SalB. Bottom trace (blue) is *apo* SalB(PCP)(excised PCP domain) from our recent publication. * indicates peak for N-terminal trypsin fragment of excised PCP domain due to His-tag that will not be present in the di-domain protein. Middle trace (black) is *apo* didomain SalB(A-PCP). Small amount of holo protein likely due to native *E. coli* PPTases. Top trace (red) is *holo* didomain SalB(A-PCP) after in vitro reaction with CoA and Sfp.

Once I confirmed formation of the *holo*-PCP, I attempted adenylation assays with cyclohexylalanine to determine if the SalB A-domain could select and adenylate this amino acid and transfer it to the adjacent PCP (**Figure 4.12**). While this amino acid is not used by the native organism, it can be incorporated into the mutasynthetic product salinosporamide X1 (**20**), which suggested that SalB is capable of adenylating this nonnative substrate. The His-tagged SalB didomain (20 μM, coexpressed with SalE) was incubated with 10 mM $MgCl_2$, 0.8 mM TCEP, 1 mM CoASH, 10 μM Sfp, 20 μM His-tagged SalE (purified separately), 1 mM cyclohexylalanine, and 1 mM ATP for 2 hours. The assay was then subjected to intact proteomics at the Molecular Mass Spectrometry Facility (MMSF) in the UCSD Chemistry Department. As a control, the SalB didomain (coexpressed with SalE) was also analyzed by intact proteomics.

LC-HRMS analysis of the SalB control revealed both *apo* (*m/z* 68500) and *holo* SalB (*m/z* 68840) species, although the protein was predominantly found, as expected, in its *apo* form. I also observed peaks indicative of a SalB-SalE complex (*m/z* 78765), and coexpressed SalE (*m/z* 8241). Analysis of the A-domain assay revealed the same peaks described above for *holo* SalB, the SalB-SalE complex, and coexpressed SalE. However, I observed no *apo* SalB, but did observe two new peaks, one indicative of His-tagged SalE (*m/z* 10404), which was purified separately and added to this assay, and a new peak indicative of cyclohexylalanine loading (*m/z* 68993). These data indicate that SalB is in fact active, and capable of selecting, adenylating, and transferring amino acids to the downstream PCP domain.

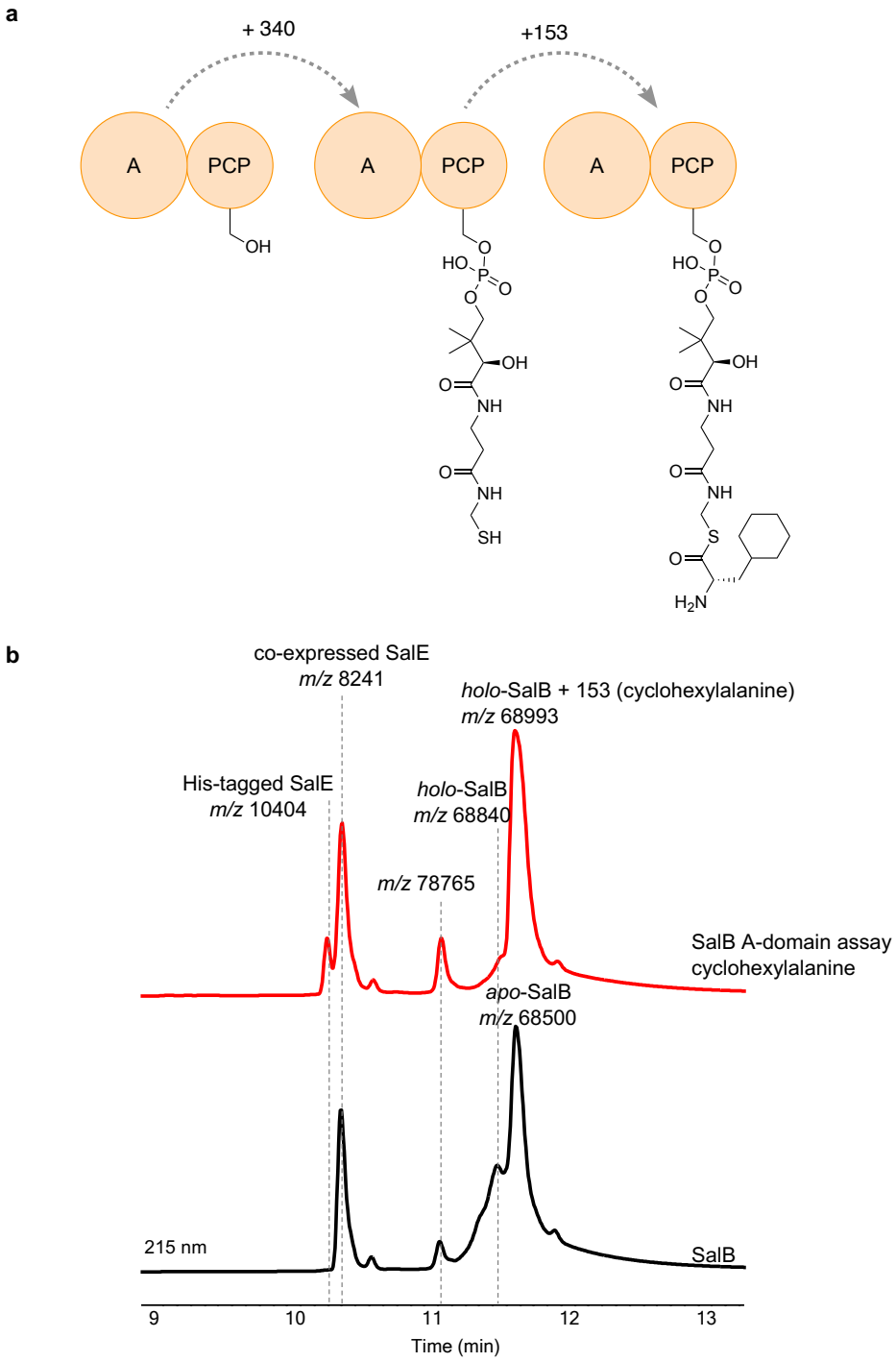


Figure 4.12 SalB adenylation domain assay. a**.** Cartoon schematic indicating *apo, holo*, and amino acid loaded *holo* SalB. b. UV chromatograms (215 nm) of intact protein LC-HRMS of SalB(A-PCP) didomain co-expressed with MbtH-like protein SalE (black), and the SalB(A-PCP) adenylation (A) domain assay with additional SalE and cyclohexylalanine (red).

SalC

Attempted expression of SalC (~65 kDa) in *E. coli* has been discussed extensively above. However, successful expression and purification of SalC from *Streptomyces coelicolor* CH999 has been achieved with a yield of ~8 mg/L. Therefore, even if soluble expression in *E. coli* is not achieved, in vitro reconstitution of the pathway will not be hindered by SalC as we have purified protein in hand.

SalD

While not technically a core assembly line protein, the hydroxylation of salinosporamide A is important for its bioactivity and likely occurs while the linear molecule is tethered to the carrier protein SalB. Therefore, this tailoring cytochrome P450 SalD (~47 kDa) is currently under active investigation by other members of the Moore lab. The protein has been successfully expressed and purified from *E. coli,* and shown to retain its heme cofactor. The timing and mechanism for amino acid hydroxylation is ongoing.

In vivo strategies for accessing salinosporamide diversity

While an in vitro strategy for salinosporamide production is appealing for the generation of new analogs and subsequent bioactivity testing, I have additionally decided to pursue a complementary, strain engineering based approach to generating salinosporamide diversity. The Moore lab has a long history of capturing biosynthetic gene clusters for bioengineering purposes, and of developing methodology for the heterologous expression of these BGCs.[48–50] However, despite this history, the salinosporamide BGC has never been captured. Now that we are interested in accessing

salinosporamide analogs, a strain engineering approach is an attractive strategy, and therefore my goal was to capture, engineer, and heterologously express the *sal* BGC.

Heterologous expression of biosynthetic gene clusters has long been used to improve the production of natural products, particularly those produced by genetically intractable organisms. While a number of different methods have been developed for this process, heterologous expression always requires the cloning or 'capture' of a biosynthetic gene cluster in an expression vector, and then expression within a heterologous host system.[51] Traditionally, for BGCs originating from GC-rich actinomycetes, *Streptomyces* are the heterologous host of choice, and a series of genome minimized *Streptomyces* hosts have been engineered specifically for the production of natural products.[52] Heterologous expression of a BGC alone may be enough to improve titer, but the real advantage of capturing a BGC lies in the genetic tractability of a plasmid-expressed cluster which can be easily engineered in *E. coli*. Using a well-developed toolkit of *E. coli* synthetic biology tools, a BGC can be decoupled from its complex native regulation and refactored with strong promoters, regulatory genes can be overexpressed or deleted, and core biosynthetic genes can be duplicated or disrupted, depending on the desired product of the strain. These types of engineering experiments may be impossible in wild type organisms, but are straightforward in *E. coli,* exemplifying a key benefit of a plasmid expressed BGC.

Previously in my PhD I captured the streptophenazine BGC using Transformation Associated Recombination (TAR) cloning methodology (Chapter 2). This method relies on the natural recombination ability of *Saccharomyces cerevisiae* to enable cloning of large segments of DNA (i.e. biosynthetic gene clusters) directly from genomic DNA.[53]

TAR cloning has been heavily developed and optimized in the Moore lab for the purpose of cloning large BGCs through the creation of the pCAP vectors, a series of shuttle vectors designed for the movement of BGCs between yeast, *E.coli*, and Actinobacteria.[48,54] However, this method is not without its challenges. Working with yeast is technically demanding, TAR cloning has low efficiency, and this method relies on homologous recombination which can present problems for GC-rich, highly repetitive BGCs from Actinomycetes. Therefore, I was interested in trying a new method for capturing the salinosporamide BGC.

In 2021, the Zhao lab at the University of Illinois, Urbana-Champaign published a Cas12a-assisted site-specific recombination based cloning method for biosynthetic gene clusters, which they claimed had an efficiency > 93%. This method utilizes FnCas12a directed by BGC-specific guide RNAs to cut genomic DNA in a targeted way and release a BGC of interest. A hallmark of the Cas12a enzyme is that it leaves behind sticky ends, which is advantageous for downstream cloning efforts. An in vitro DNA assembly step between the Cas12a-digested BGC and two "DNA receivers" with homology arms to the BGC generates a linear DNA fragment flanked by loxP sites. This linear piece of DNA is subsequently transformed into *E. coli* for an in vivo Cre-*lox* recombination-based circularization step that generates a plasmid carrying the BGC of interest capable of both *E. coli* and *Streptomyces* expression. The methodology differences between TAR cloning and Cas12a-assisted cloning are highlighted in **Figure 4.13,** but the Cas12a-assisted method is advantageous for a number of reasons. First, it doesn't rely on restriction enzyme sites, which may or may not be found near the BGC of interest. This method is also performed entirely *in vitro,* until the final *E. coli* based circularization step, eliminating

any reliance on yeast and simplifying the techniques required. Furthermore, this method does not rely on homologous recombination, which can be problematic for highly homologous, high GC assembly line pathways from Actinomycetes. Finally, once gRNAs are designed and Cas12a is purified, the method can be performed quickly in 3-4 days, including screening.

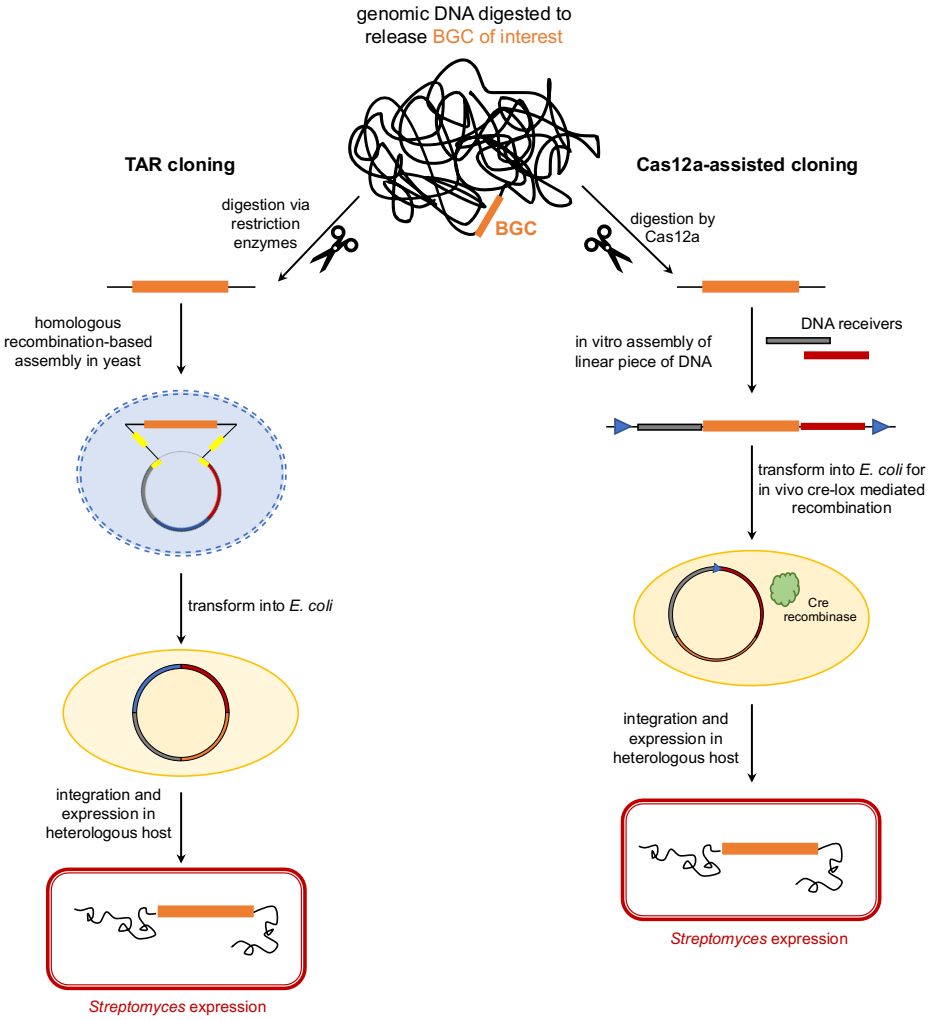


Figure 4.13 Capture methodology. Comparison of TAR cloning and Cas12a-assisted cloning methodology for capturing BGCs.

In practice, this method proved much simpler and more efficient than TAR cloning. I designed gRNA pairs that cut up- and downstream of the *sal* BGC, and screened 20

colonies per gRNA pair. Colony PCR analysis for the presence of the *salB* gene revealed an efficiency of ~20% (**Figure 4.14a**). Plasmids were purified from positive colonies and assembly was confirmed via restriction digests with three enzyme pairs (**Figure 4.14b**). This verified that the salinosporamide BGC was successfully captured.

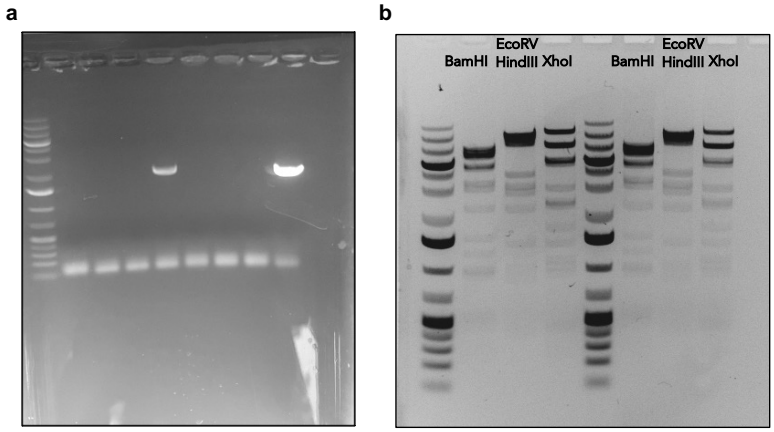


Figure 4.14 Cas12a-assisted cloning of the sal BGC. a. Colony PCR of the salB gene from Cas12a-assisted cloning attempt indicating two positive clones. b. Restriction digest confirming capture of the sal BGC in the two positive clones from a.

Now that I have captured the *sal* BGC, I plan to express the cluster and validate the production of salinosporamides in either a *Streptomyces*[52] or *Salinispora*[55] heterologous host. Wild type *Streptomyces* have been found to produce a related series of molecules, the cinnabaramides,[56] which suggests that this may be a successful host. However, as discussed in Chapter 2 of this dissertation, heterologous expression of BGCs does not always result in successful expression of the pathway and production of the molecules of interest. In this case, it is possible to express the *sal* BGC in the engineered strain *Salinispora tropica* CNB4401, where we know the molecules will be produced.[55] This strain was constructed by replacing *salA-C* with the ϕC31 phage attachment site *attB*, rendering the strain incapable of producing salinosporamides, but

allowing for genomic integration of BGCs from plasmids bearing an *attP* site, which our *sal* plasmid includes. Using a *Salinispora* host is not ideal because it still requires saltwater media and the organism is relatively difficult to work with compared to *Streptomyces,* but a captured, plasmid-based BGC is still advantageous as it allows for engineering and the creation of designer salinosporamide-producing strains.

Once production of salinosporamide A can be verified, the plasmid-borne BGC can be engineered for production of salinosporamide analogs. Of particular interest is swapping the adenylation domain (A domain) specificity of *salB* to incorporate different amino acids, thereby generating salinosporamide analogs with differential binding for the S1 pocket of the proteasome. A domains are highly specific for the amino acid they recognize, and this specificity is determined by a distinct series of residues in the substrate binding pocket, often referred to as the "Stachelhaus code" that enables the in silico prediction of an A domain's substrate.[57,58] Once the salinosporamide BGC is plasmid-based, the Stachelhaus code of the *salB* A domain can be altered to incorporate phenylalanine instead of cyclohexenylalanine, thereby creating a designer strain that specifically makes salinosporamide X4 (**23**). Methodology developed in the Moore lab for the rapid engineering of cloned BGCs will be employed for this goal.[59] We aim to engineer a library of designer salinosporamide producing strains, each with preferential incorporation of a different amino acid. The resulting inhibitors produced by these strains will be used directly for high-throughput bioactivity screening efforts against alternative proteasomes.

4.6 Discussion

When the *Salinispora tropica* CNB440 genome was sequenced in 2007, notably the first whole genome-sequenced marine actinomycete, it opened the door towards the biosynthetic investigation of the organism's famed natural product, salinosporamide A. Soon after, nearly every gene in the *sal* cluster was targeted for inactivation, thereby establishing a proposed biosynthetic pathway for mature salinosporamide production.[22] Subsequent in vitro biosynthetic investigations focused on the unusual SAM-dependent halogenation reaction[60] and the formation of the halogenated PKS extender unit chloroethylmalonyl.[61] A tremendous amount of attention was paid to fermenting and synthesizing analogs of this compound, but none of the resulting molecules were able to best the bioactivity of the originally isolated natural product itself. By the mid 2010s, with salinosporamide A comfortably advancing through clinical trials, much of the scientific attention around this molecule had dissipated, despite unanswered biosynthetic questions. However, our recent discovery of the enzyme responsible for the construction of salinosporamide's distinctive γ-lactam-β-lactone core has reinvigorated interest in the salinosporamide pathway and helped established a system for further biosynthetic investigations.

A major area of research sparked by our recent publication is the development of the salinosporamide bicyclase SalC as a biocatalyst. This enzyme is incredibly exciting as a potential biocatalyst due to its ability to perform a complex reaction, installing two stereocenters and a reactive β-lactone moiety, on a variety of complex substrates. However, the biocatalytic potential of this enzyme is hindered by its reliance on carrier protein tethered substrate. Therefore our goal is a directed evolution campaign to

engineer SalC away from carrier protein dependence. This work would not only ensure an efficient and scalable route to salinosporamide diversity, but could be universally applied to fundamentally transform biocatalysis using assembly line enzymes.

This chapter additionally describes work towards the ultimate reconstitution of the entire salinosporamide assembly line. This salinosporamide assembly line is a small, hybrid PKS-NRPS, clinically important system, and potentially ideal for structural characterization. Three of the four *sal* assembly line proteins have been expressed, making complete reconstitution a tangible goal. Finally, my recent capture of the salinosporamide biosynthetic gene cluster opens the door for future strain engineering efforts that could be used to generate additional new-to-nature salinosporamides.

Beyond the work described here, additional components of the salinosporamide pathway remain unexplored. The proposed P450 catalyzed hydroxylation of the amino acid remains experimentally uninvestigated, and is an intriguing target for further study of carrier dependent P450s. Furthermore, while the route to the unusual amino acid cyclohexenylalanine has been proposed,[5] just one step in this biosynthesis has been experimentally validated,[62] leaving the rest of the pathway still open to investigation. The ongoing work described in this chapter aims to lay the groundwork for future investigations of the salinosporamide biosynthetic pathway that I am eager to see other Moore lab members continue in the future.

4.7 References

1. Cromm, P. M. & Crews, C. M. The proteasome in modern drug discovery: second life of a highly valuable drug target. *ACS Cent. Sci.* **3**, 830–838 (2017).

2. Adams, J. The proteasome: structure, function, and role in the cell. *Cancer Treat. Rev.* **29**, 3–9 (2003).

3. Tanaka, K. The proteasome: Overview of structure and functions. *Proc. Jpn Acad. Ser. B. Phys. Biol. Sci.* **85**, 12–36 (2009).

4. Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. & Fenical, W. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed. Engl.* **42**, 355–357 (2003).

5. Gulder, T. A. M. & Moore, B. S. Salinosporamide Natural Products: Potent 20S Proteasome Inhibitors as Promising Cancer Chemotherapeutics. *Angew. Chem. Int. Ed. Engl.* **49**, 9346–9367 (2010).

6. Fenical, W., Jensen, P. R., Palladino, M. A., Lam, K. S., Lloyd, G. K. & Potts, B. C. Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* **17**, 2175 (2009).

7. Huber, E. M., Basler, M., Schwab, R., Heinemeyer, W., Kirk, C. J., Groettrup, M. & Groll, M. Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* **148**, 727–738 (2012).

8. Sasaki, K., Takada, K., Ohte, Y., Kondo, H., Sorimachi, H., Tanaka, K., Takahama, Y. & Murata, S. Thymoproteasomes produce unique peptide motifs for positive selection of CD8+ T cells. *Nat. Commun.* **6**, 7484 (2015).

9. Basler, M., Dajee, M., Moll, C., Groettrup, M. & Kirk, C. J. Prevention of experimental colitis by a selective inhibitor of the immunoproteasome. *J. Immunol.* **185**, 634–641 (2010).

10. Muchamuel, T., Basler, M., Aujay, M. A., Suzuki, E., Kalim, K. W., Lauer, C., Sylvain, C., Ring, E. R., Shields, J., Jiang, J., Shwonek, P., Parlati, F., Demo, S. D., Bennett, M. K., Kirk, C. J. & Groettrup, M. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat. Med.* **15**, 781–787 (2009).

11. Jenkins, T. W., Downey-Kopyscinski, S. L., Fields, J. L., Rahme, G. J., Colley, W. C., Israel, M. A., Maksimenko, A. V., Fiering, S. N. & Kisselev, A. F. Activity of immunoproteasome inhibitor ONX-0914 in acute lymphoblastic leukemia expressing MLL–AF4 fusion protein. *Sci. Rep.* **11**, 10883 (2021).

12. Stansborough, R. L. & Gibson, R. J. Proteasome inhibitor-induced gastrointestinal toxicity. *Curr. Opin Support. Palliat. Care* **11**, 133–137 (2017).

13. Badros, A., Goloubeva, O., Dalal, J. S., Can, I., Thompson, J., Rapoport, A. P., Heyman, M., Akpek, G. & Fenton, R. G. Neurotoxicity of bortezomib therapy in multiple myeloma: A single-center experience and review of the literature. *Cancer* **110**, 1042–1049 (2007).

14. Zhang, H. & Lin, G. Microbial proteasomes as drug targets. *PLoS Pathog.* **17**, e1010058 (2021).

15. Lin, G., Li, D., de Carvalho, L. P. S., Deng, H., Tao, H., Vogt, G., Wu, K., Schneider, J., Chidawanyika, T., Warren, J. D., Li, H. & Nathan, C. Inhibitors selective for mycobacterial versus human proteasomes. *Nature* **461**, 621–626 (2009).

16. Li, H., Ponder, E. L., Verdoes, M., Asbjornsdottir, K. H., Deu, E., Edgington, L. E., Lee, J. T., Kirk, C. J., Demo, S. D., Williamson, K. C. & Bogyo, M. Validation of the proteasome as a therapeutic target in *Plasmodium* using an epoxyketone inhibitor with parasite-specific toxicity. *Chem. Biol.* **19**, 1535–1545 (2012).

17. Li, H., O'Donoghue, A. J., Linden, W. A. van der, Xie, S. C., Yoo, E., Foe, I. T., Tilley, L., Craik, C. S., Fonseca, P. C. A. da & Bogyo, M. Structure and function based design of *Plasmodium*-selective proteasome inhibitors. *Nature* **530**, 233 (2016).

18. Bibo-Verdugo, B., Wang, S. C., Almaliti, J., Ta, A. P., Jiang, Z., Wong, D. A., Lietz, C. B., Suzuki, B. M., El-Sakkary, N., Hook, V., Salvesen, G. S., Gerwick, W. H., Caffrey, C. R. & O'Donoghue, A. J. The proteasome as a drug target in the metazoan pathogen, *Schistosoma mansoni*. *ACS Infect. Dis.* **5**, 1802–1812 (2019).

19. Khare, S., Nagle, A. S., Biggart, A., Lai, Y. H., Liang, F., Davis, L. C., Barnes, S. W., Mathison, C. J. N., Myburgh, E., Gao, M.-Y., Gillespie, J. R., Liu, X., Tan, J. L., Stinson, M., Rivera, I. C., Ballard, J., Yeh, V., Groessl, T., Federe, G., Koh, H. X. Y., Venable, J. D., Bursulaya, B., Shapiro, M., Mishra, P. K., Spraggon, G., Brock, A., Mottram, J. C., Buckner, F. S., Rao, S. P. S., Wen, B. G., Walker, J. R., Tuntland, T., Molteni, V., Glynne, R. J. & Supek, F. Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature* **537**, 229–233 (2016).

20. Reed, K. A., Manam, R. R., Mitchell, S. S., Xu, J., Teisan, S., Chao, T.-H., Deyanat-Yazdi, G., Neuteboom, S. T. C., Lam, K. S. & Potts, B. C. M. Salinosporamides D-J from the marine actinomycete *Salinispora tropica*, bromosalinosporamide, and thioester derivatives are potent inhibitors of the 20S proteasome. *J. Nat. Prod.* **70**, 269–276 (2007).

21. Nett, M., Gulder, T. A. M., Kale, A. J., Hughes, C. C. & Moore, B. S. Function-oriented biosynthesis of β-lactone proteasome inhibitors in *Salinispora tropica*. *J. Med. Chem.* **52**, 6163–6167 (2009).

22. McGlinchey, R. P., Nett, M., Eustáquio, A. S., Asolkar, R. N., Fenical, W. & Moore, B. S. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **130**, 7822 (2008).

23. Reddy, L. R., Fournier, J.-F., Subba Reddy, B. V. & Corey, E. J. An efficient, stereocontrolled synthesis of a potent omuralide−salinosporin hybrid for selective proteasome inhibition. *J. Am. Chem. Soc.* **127**, 8974–8976 (2005).

24. Eustáquio, A. S. & Moore, B. S. Mutasynthesis of fluorosalinosporamide, a potent and reversible inhibitor of the proteasome. *Angew. Chem. Int. Ed. Engl.* **120**, 4000–4002 (2008).

25. Macherla, V. R., Mitchell, S. S., Manam, R. R., Reed, K. A., Chao, T.-H., Nicholson, B., Deyanat-Yazdi, G., Mai, B., Jensen, P. R., Fenical, W. F., Neuteboom, S. T. C., Lam, K. S., Palladino, M. A. & Potts, B. C. M. Structure-activity relationship studies of salinosporamide A (NPI-0052), a novel marine derived proteasome inhibitor. *J. Med. Chem.* **48**, 3684–3687 (2005).

26. Manam, R. R., McArthur, K. A., Chao, T.-H., Weiss, J., Ali, J. A., Palombella, V. J., Groll, M., Lloyd, G. K., Palladino, M. A., Neuteboom, S. T. C., Macherla, V. R. & Potts, B. C. M. Leaving groups prolong the duration of 20S proteasome inhibition and enhance the potency of salinosporamides. *J. Med. Chem.* **51**, 6711–6724 (2008).

27. Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S. K., Martin, T. A. & Zhao, H. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination. *Nat. Commun.* **12**, 1171 (2021).

28. Bauman, K. D., Shende, V. V., Chen, P. Y.-T., Trivella, D. B. B., Gulder, T. A. M., Vellalath, S., Romo, D. & Moore, B. S. Enzymatic assembly of the salinosporamide γ-lactam-β-lactone anticancer warhead. *Nat. Chem. Biol.* **18**, 538–546 (2022).

29. Kortmann, M., Kuhl, V., Klaffl, S. & Bott, M. A chromosomally encoded T7 RNA polymerase-dependent gene expression system for *Corynebacterium glutamicum*: construction and comparative evaluation at the single-cell level. *Microb. Biotech.***8**, 253–265 (2015).

30. Hansen, D. A., Koch, A. A. & Sherman, D. H. Substrate controlled divergence in polyketide synthase catalysis. *J. Am. Chem. Soc.* **137**, 3735–3738 (2015).

31. Worthington, A. S., Rivera, H., Torpey, J. W., Alexander, M. D. & Burkart, M. D. Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS Chem. Biol.* **1**, 687–691 (2006).

32. Gulick, A. M. & Aldrich, C. C. Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Nat. Prod. Rep.* **35**, 1156–1184 (2018).

33. Walsh, C. T. The chemical versatility of natural-product assembly lines. *Acc. Chem. Res.* **41**, 4–10 (2008).

34. Bozhüyük, K. A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51**, 88–96 (2019).

35. Calcott, M. J. & Ackerley, D. F. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol. Lett.* **36**, 2407–2416 (2014).

36. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Mol. Biol.* **9**, 522–526 (2002).

37. May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12120–12125 (2002).

38. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11124–11129 (2006).

39. Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).

40. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321** (659-63) (2008).

41. Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Leigh Allen, C., Skiniotis, G., Aldrich, C. C. & Gulick, A. M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).

42. Cogan, D. P., Zhang, K., Li, X., Li, S., Pintilie, G. D., Roh, S.-H., Craik, C. S., Chiu, W. & Khosla, C. Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science* **374**, 729–734 (2021).

43. Bagde, S. R., Mathews, I. I., Fromme, J. C. & Kim, C.-Y. Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science* **374**, 723–729 (2021).

44. Wang, J., Li, D., Chen, L., Cao, W., Kong, L., Zhang, W., Croll, T., Deng, Z., Liang, J. & Wang, Z. Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. *Nat. Commun.* **13**, 592 (2022).

45. Miyazawa, T., Hirsch, M., Zhang, Z. & Keatinge-Clay, A. T. An in vitro platform for engineering and harnessing modular polyketide synthases. *Nat. Commun.* **11**, 80 (2020).

46. Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W. & Thomas, M. G. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochem.* **49**, 8815–8817 (2010).

47. Schomer, R. A., Park, H., Barkei, J. J. & Thomas, M. G. Alanine scanning of YbdZ, an MbtH-like protein, reveals essential residues for functional interactions with its nonribosomal peptide synthetase partner EntF. *Biochem.* **57**, 4125–4134 (2018).

48. Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., Dorrestein, P. C. & Moore, B. S. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1957–1962 (2014).

49. Zhang, J. J., Tang, X., Zhang, M., Nguyen, D. & Moore, B. S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in gram-negative bacteria. *mBio* **8**, e01291-17 (2017).

50. Bauman, K. D., Li, J., Murata, K., Mantovani, S. M., Dahesh, S., Nizet, V., Luhavaya, H. & Moore, B. S. Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry. *Cell Chem. Biol.* **26**, 724-736.e7 (2019).

51. Zhang, J. J., Tang, X. & Moore, B. S. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* **36**, 1313–1332 (2019).

52. Gomez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207–215 (2011).

53. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protocols* **3**, 371–377 (2008).

54. Zhang, J. J., Yamanaka, K., Tang, X. & Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* **621**, 87–110 (2019).

55. Zhang, J. J., Moore, B. S. & Tang, X. Engineering *Salinispora tropica* for heterologous expression of natural product biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **102**, 8437–8446 (2018).

56. Rachid, S., Huo, L., Herrmann, J., Stadler, M., Köpcke, B., Bitzer, J. & Müller, R. Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *ChemBioChem* **12**, 922–931 (2011).

57. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).

58. Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C. & Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).

59. Zhang, J. J. & Moore, B. S. Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synth. Biol.* **9**, 1917–1922 (2020).

60. Eustáquio, A. S., Pojer, F., Noel, J. P. & Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **4**, 69–74 (2008).

61. Eustáquio, A. S., McGlinchey, R. P., Liu, Y., Hazzard, C., Beer, L. L., Florova, G., Alhamadsheh, M. M., Lechner, A., Kale, A. J., Kobayashi, Y., Reynolds, K. A. & Moore, B. S. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-l-methionine. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12295–12300 (2009).

62. Mahlstedt, S., Fielding, E. N., Moore, B. S. & Walsh, C. T. Prephenate decarboxylases: a new prephenate-utilizing enzyme family that performs non-aromatizing decarboxylation en route to diverse secondary metabolites. *Biochem.* **49**, 9021–9023 (2010).

4.8 Acknowledgements

I would like to thank Professor Michael G. Thomas for the *E. coli* BL21(DE3) *ybdZ::aac*(3)IV strain used for expression of SalB, and Professor Huimin Zhao for the Cas12a CAPTURE plasmids. This dissertation authors was the primary investigator and author of this chapter.

CHAPTER 5. Perspectives and future outlook

Secondary metabolites are the language of the microbial world. It is the unique mixture of compounds synthesized by an organism that defines how that organism interacts with its environment, ultimately determining its ecological niche. In this way, chemistry mediates biology and shapes our natural world. These small molecule natural products can have profound impacts on human life as well, and are used regularly in diverse areas of our lives, including as pesticides, pigments, and, of course, as pharmaceuticals. Because of this, there is tremendous value in understanding how these compounds are produced naturally. Only by first understanding the biosynthesis of a compound can we use its biosynthetic genes to discover new compounds, or employ its biosynthetic enzymes in bio-catalytic manufacturing routes, or engineer the entire pathway for improved production of a high-value chemical. It is only through understanding the biosynthesis of a compound, therefore, that we gain true mastery of that system.

The work described in my dissertation used a combination of in vivo genetics based approaches and in vitro biochemical experiments to explore the biosynthesis of two bioactive families of natural products, the streptophenazines and the salinosporamides, both produced by marine bacteria. Here, I summarize the key findings of these research projects with special attention to the broader impact of the work and directions for future research.

In Chapter 2, I used a genetic engineering strategy to address one of the most fundamental and persistent problems in natural product research – that of inadequate or absent expression of biosynthetic pathways. Recent estimates suggest that only 3% of

secondary metabolites encoded in bacterial genomes have been discovered, indicating that the vast majority of biosynthetic gene clusters are silent, or not expressed, under traditional laboratory culture conditions. I was faced with this issue directly when studying a predicted phenazine BGC from a marine *Streptomyces* bacterium. To definitively link this gene cluster to its molecular products and aid in downstream genetic engineering experiments, I captured this BGC and attempted expression in a *Streptomyces* heterologous host. However, I saw no production of any phenazine-type metabolites, and RT-PCR experiments ultimately revealed that this cluster was not expressed in a heterologous environment.

I used a synthetic biology approach to address this question by refactoring the gene cluster, replacing the defective native regulatory system with engineered genetic parts. This ensured consistent, reliable, and high-yielding production of over 100 compounds that belonged to the streptophenazine family, including a series of structurally novel members. Furthermore, the genetic engineering strategy used in this work enabled some understanding of the streptophenazine biosynthetic pathway. Insertion of a promoter in front of the operon responsible for biosynthesis of the phenazine core activated expression of every other gene in the cluster. This allowed us to propose a biosynthetic mechanism where phenazine dicarboxylic acid serves as the starter unit for polyketide extension to generate the mature streptophenazines, and suggests that transcription of this cluster may be regulated by production of this starter unit.

Looking towards the future, there is much left to be explored in regards to the streptophenazine pathway. The biosynthesis of the formylglycine containing compounds is particularly interesting as formylation is a rare feature in natural products and is usually

accomplished by a dedicated formylation domain, which is not found in the *spz* BGC, nor elsewhere in the genome. While gene disruption experiments demonstrated that a putative adenylating protein is required for production of these compounds, unpublished in vitro experiments showed no activity with this enzyme in adenylating formylglycine. However, there are handful of other interesting targets in the cluster that may be involved in this chemistry. Of particular interest is Spz27*,* a 781 amino acid protein that appears to harbor both a methyltransferase and an oxidoreductase domain, which could be a route to formylation of an amino acid.

In Chapter 3 of this dissertation I used an vitro approach to fully elucidate a specific enzymatic step in the assembly of the anticancer clinical candidate salinosporamide A. This work revealed SalC as the salinosporamide bicyclase, capable of generating the characteristic γ-lactam-β-lactone core that imparts salinosporamide with its anticancer activity. By sequence and structural similarity, SalC resembles a ketosynthase (KS), but demonstrates a fundamentally different reactivity than canonical KSs, and exhibits a new route to PKS-NRPS cyclization and chain release. The in vitro strategy used in this chapter allowed us to explore this biosynthetic step in tremendous detail; structural characterization of SalC through x-ray crystallography enabled site-directed mutagenesis experiments that have informed a proposed mechanism for this unusual reaction.

The cyclization step in salinosporamide biosynthesis had remained a mystery for over a decade, largely due to the experimental challenges associated with this question. The salinosporamide pathway is an assembly line system, necessitating carrier protein tethered substrates, and the product is a fragile β-lactone, complicating many traditional experiments. Additionally, many of the *sal* enzymes were not readily purified from *E. coli*

and required alternative expression systems or heterologous hosts. Truly what I am most proud of with this project is the sheer grit and determination it required. Many of the experiments were physically demanding, done on a large scale, or required extensive planning, development, and attention to detail.

The work described in Chapter 3 established a working system for the continued exploration of salinosporamide biosynthesis, particularly the assembly line portion of the pathway, which is discussed in Chapter 4 of this dissertation. Chapter 4 combines strategies from Chapters 2 and 3, using both in vivo and in vitro approaches to further investigate salinosporamide assembly line biosynthesis. Contrary to many assumptions, the biosynthesis of this molecule is not fully characterized, and questions surrounding its construction still remain. In this chapter I describe my work towards complete in vitro reconstitution of the salinosporamide pathway, which I see as a particularly exciting goal for future structural biology work. Additionally, this chapter discusses work towards engineering SalC away from carrier protein dependence, which could ultimately enable a streamlined, chemoenzymatic approach to salinosporamide production. With the new Illumina high-throughput screening laboratory, the Moore lab will be increasingly focused on protein engineering, and SalC is an intriguing target for towards this goal. Finally, I have recently captured the salinosporamide biosynthetic gene cluster for the first time, which enables future genetic engineering efforts towards construction of designer salinosporamide producing strains.

The late 20th century has been heralded as the so-called "Third Industrial Revolution," a revolution based in silicon and focused on digital innovation. The technological and computing advancements developed during those years have

irrevocably shaped our world; the invention of the internet and the personal computer have enabled organization on a global scale, facilitating the near constant connection and communication to which we have all become accustomed. These technological advances have transitioned us from an analog to a digital world, and, for better or for worse, are now fundamentally intertwined with modern life.

With the dawn of the 21st century, however, the digital revolution of the 80s and 90s has given way to a bio-based revolution, driven by recent biological discoveries and bio-technological advancements. Innovations in sequencing technologies, epitomized by the long read sequencing from Oxford Nanopore or the recent reveal of the $100 human genome by Ultima Genomics, an astounding feat given the multimillion dollar price tag it took to sequence the human genome in 2006, have been foundational to this era of bio-revolution. Similarly, the cost of DNA synthesis has dropped orders of magnitude over the last 10 years, enabling affordable access to genetic material. The discovery of the bacterial CRISPR-Cas based immune system has led to the most precise and accurate genome editing tool ever developed, now used in clinical trials to treat genetic diseases. As these biological advancements have accelerated, we've already begun to see them shape our world. This was perhaps most clear in 2020, when our ability to quickly sequence and then synthesize the genetic code of the novel SARS-CoV-2 virus led to the unprecedented development of an mRNA vaccine in less than a year – a feat previously deemed impossible. With this enhanced ability to read, write, and edit genetic code at scale, synthetic biology is poised to take on some of the biggest challenges we face as a species. Climate change, antibiotic resistance, global pandemics, and food shortages are all areas where synthetic biology could provide a solution.

These biotechnological advances have also fundamentally altered the field of natural products – the work described in this dissertation relied heavily on genome sequencing, gene synthesis, and genetic engineering, and would not have been possible even 10 years ago. Our ability to engineer, edit, and manipulate biosynthetic pathways gives us a mastery of these systems never previously possible. However, I also believe that natural products research has an outsized role to play in shaping synthetic biology as well. Understanding the enzymatic processes that make bioactive molecules creates a genetic roadmap to discover new compounds, perhaps with enhanced bioactivity or other advantageous properties. Studying the biosynthesis of secondary metabolites opens the door for sustainable, biocatalytic routes to making valuable chemicals. During my graduate career I had the opportunity to intern at Synlogic, a biotech company bringing live microbial therapeutics to the clinical for the treatment of genetic metabolic diseases. They were constantly searching for new enzymes with specific properties to build into their genetic circuits, which is an area where natural product research can have a tremendous impact. As we continue to discover new enzymes from biosynthetic pathways capable of performing new and difficult chemistries, we grow the synthetic biology toolbox for the sustainable production of high-value chemicals.

Throughout my PhD I have become increasingly enamored with Nature as a chemist; the incredibly precise molecular machines used to produce intricate chemical structures controlled by complex regulatory networks cannot be rivaled by anything we produce in a lab. I am continually astounded by how much remains to discovered in the natural world, how little we know even about seemingly “well-studied” organisms. One can only imagine the additional secrets harbored in bacterial genomes around the world,

and how we can use them to help engineer biology. I'd like to imagine a world where someday little bacterial chemists make sustainable plastic substitutes in giant fermenters, algae help sequester carbon, and an engineered microbiome helps keep us safe from disease. I became a scientist because I wanted to build something, and I believe that synthetic biology is the key to building a sustainable future. The interdisciplinary skills I have learned throughout my PhD have enabled me to reach across disciplines, to not be siloed or scared into a single area of research, and to turn to new fields and techniques for the answers to my questions. With the skills developed through the work described here, I'm eager to take my place as a scientist helping shape the biology based world of the future.