Decoding the Bio-Assembly of Hapalindole-Type Alkaloids from Cyanobacteria for Drug Discovery

by

Shasha Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2017

Doctoral Committee:

Professor David H. Sherman, Chair
Professor George A. Garcia
Professor John Montgomery
Professor Janet L. Smith

Shasha Li
shashali@umich.edu
ORCID iD: 0000-0001-7468-0377

Dedication

To My Parents,
Who bestowed me life, raised me with love, educated me with wisdom,
And support me pursuing my dream from thousands of miles away.

To My Darling,
Who doubles my joys, divides my sorrows, calms me down, and cheers me up,
Our hearts are tied together no matter how far apart.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. David H. Sherman for his guidance and support throughout my graduate education. I was always told that a “right” advisor is one of the most important elements for an enjoyable and fruitful Ph.D. life. I feel extremely fortunate and grateful to have Dr. Sherman as my Ph.D. research advisor, whose guidance is far-sighted, inspiring, patient and comprehensive. He also offered me the flexibility to choose projects and valuable opportunity for collaboration.

I also want to acknowledge my committee members, Dr. George Garcia, Dr. John Montgomery, Dr. Janet Smith, and my former committee members, Dr. Carol Fierke and Dr. Ronald Woodard. Their continued guidance, constructive suggestions, and warmly encouragement helped me a lot in accomplishing my dissertation research. I also want thank Dr. Shaomeng Wang for his valued guidance during my rotation.

A harmonious research environment is sometimes more important than the research project. I feel extremely lucky to have all my labmates in the past five years. Especially, I want to thank Dr. Andrew Lowell, Dr. Sean Newmister, Dr. Fengan Yu, Dr. Ashootosh Tripathi, Dr. Sung Ryeol Park, Dr. Yogan Khatri, Dr. Karoline Chiou, Dr. Douglas Hansen, Dr. Pamela Schultz, Dr. Michael Schofield, Jennifer Schmidt, Sam Slocum, Callie Chappell, and those I failed to list here, for their fruitful collaboration, inspiring discussion, and warmly encouragement.

I am grateful for the funding from NSF-CCI center for Selective C-H Functionalizatio (CHE-1205646), and Rackham Predoctoral Fellowship.

I am also sincerely grateful to all my friends and colleagues, especially Mr. Su for his constructive suggestions which greatly helped me in planning my life, and Lulu for her best friendship.

Last, but most importantly, I want to thank the love and support from my family, my father Zhengyuan Li, my mother Xiaoqun Wu, my brother Xunjie Li, who are the source of my motivation and courage to optimistically face every possibility in life. Most luckily, my soulmate Chen Feng, even though we both joked how we delayed each other's graduation, but without his companionship and encouragement, my life would be boring and painful.

Preface

This thesis contains six chapters covering my dissertational studies to understand the biosynthesis of hapalindole-type alkaloids isolated from cyanobacteria. Chapter 1 is an overall introduction of hapalindole metabolites. Chapter 2 is adapted from the paper *Journal of the American Chemical Society*, **2015**, *137*, 15366-15369, describing the gene cluster annotation and the initial discovery of the novel Stig cyclases. Chapter 3 is adapted from paper *Nature Chemical Biology*, **2017**, *13*, 467-470, characterizing the formation of hapalindoles and fischerindoles controlled by the Stig cyclases. Chapter 4 is a continuous work in elucidating the enzymatic functions of all annotated Stig cyclase and analyzing its mechanism in controlling the regio- and stereoselective ring formation, this work is in preparation for submission. Chapter 5 is focused on structural study of Stig cyclase HpiC1 followed by computational QM/MD analysis to understand the Cope rearrangement mechanism, this work is collaborated with Sean N. Newmister and have been submitted for publication. Chapter 6 describes my perspectives about future study on the current hapalindole project, together with some preliminary data

Table of Contents

List of Tables

List of Figures

Abstract

Biosynthetic exploration of natural products provides a promising opportunity to produce novel bioactive molecules for drug development. Hapalindole-type alkaloids are a large group of secondary metabolites isolated from terrestrial and freshwater Hapalosiphonaceae cyanobacteria. Their diverse pharmacological activity and intriguing polycyclic skeletons have compelled us to pursue a detailed mechanistic understanding of their biosynthesis, hoping to overcome the limitations of chemical synthesis and natural product isolation in developing these metabolites as drug leads.

My dissertation research focus on elucidating the molecular basis of hapalindole and fischerindole core ring formation, which are the basic structures to generate more complex ambiguine and welwitindolinone subgroups through late-stage tailoring. All reported hapalindoles and fischerindoles can be categorized into six stereochemical patterns based on the C10, C11, C12, and C15 chiral centers. This stereo- and regiochemically diverse polycyclic ring formation represents a fascinating platform to investigate variations during initial cyclization and subsequent tailoring reactions.

In this report, we describe a thorough study of the novel Stig cyclases responsible for catalyzing hapalindole and fischerindole formation from a common intermediate C3-geranylated indole isonitrile, whose coupled Cope rearrangement/6-exo-trig cyclization/electrophilic aromatic substitution together elaborates the four chiral centers and three types of ring systems. We have characterized the enzymatic activity of more than twenty annotated Stig cyclases generating stereo-diverse hapalindoles and fischerindoles. Our crystal structure

study, computational modeling and mutational analysis together reveal that the Stig cyclases may function as Ca^{2+}-induced higher-order heteromeric oligomers to collaboratively control the stereochemistry. This work also enable manipulation to diversify these remarkable alkaloids.

Chapter 1
Introduction

1.1 Hapalindole-type alkaloids and drug discovery

According to the CDC, pathogenic microbes cause more than two million illnesses and 23,000 deaths annually in United States, and, the proliferation of antibiotic resistance is raising the specter of a potential public health crisis.[1-2] Since 1962, following the golden-era of antibiotic discovery, the pharmaceutical industry has largely focused on developing next-generation analogs of existing antibiotics, providing a temporary solution for recurring problems.[3] However, once resistance emerges to a particular drug class, it often spreads rapidly and the new analogs become ineffective for treating bacterial infections. Current antibiotics are losing their effectiveness for treating common infections and must be replaced with new classes of drugs. Natural products remain an important resource for antimicrobial drug discovery based on a strong track record that includes β-lactams, macrolides, tetracyclines, and most recently, daptomycin.[4-7] Many antibiotics introduced into the clinic over the past 50 years were discovered by screening extracts from cultivable soil microorganisms.[4] However, due to the advent of combinatorial chemistry, and the perceived slow pace and high cost of natural product discovery, most large pharmaceutical companies terminated their programs in this area over twenty years ago.[8] In addition, due to the generally low return on investment for new antibiotics, the pharmaceutical industry lost its enthusiasm for developing these agents. Small biotechnology companies and academic drug discovery programs are now assuming much of this effort.[9] The startling increase in drug-resistant pathogens and newly identified infectious

agents has dramatically altered the biomedical urgency of finding new structural families of antibiotics targeting new biochemical and biological targets.[3]

Over the past thirty years, terrestrial, aquatic and marine cyanobacteria have been investigated for their ability to generate structurally unique natural product molecules with important biological activities. Historically, cyanobacterial natural product screening programs have been focused on anticancer activity, but this has been expanded recently to include other targets, including infectious agents.[10-12] The discovery of Hapalindole-type alkaloids with antiinfective activity was one of the outcomes of expanded screening campaigns.[13]

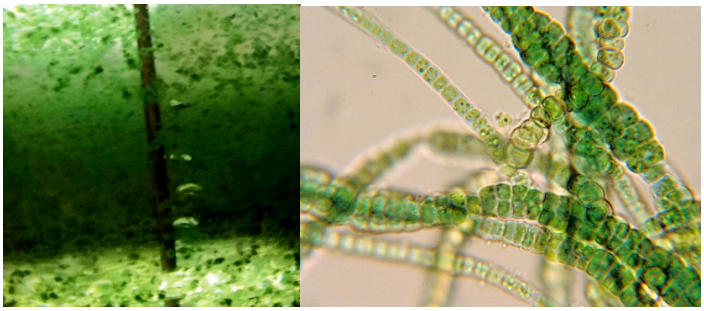


Figure 1-1. Laboratory cultivation of cyanobacteria and cell-level image of the family Hapalosiphonaceae which producing the hapalindole-type alkaloids.

1.2 Occurrence and isolation

In 1984, Moore and co-workers published the first isolation work of hapalindole family from the branched filamentous blue-green algae *Hapalosiphon fontinalis* V-3-1. The motivation was to isolate the active ingredients responsible for the antialgal activity toward *Anabaena oscillarioides*, which was observed from its lipophilic extract.[13] This work yielded two similar tetracyclic indole alkaloids, hapalindole A (**Figure 1-3**, **10**) and B (**11**), naming from the genus *Hapalosiphon* and structural type indole alkaloid. Following this work, another eighteen new compounds, hapalindoles C-Q and T-V, have been isolated from the same strain (**Figure 1-3**). The investigation toward other cyanophytes uncovered more similar tri-, tetra- and pentacyclic indole alkaloids, all given names to indicate the genus or species from which they were isolated. Since

then, at least 82 hapalindole-type natural products have been isolated from over 18 documented cyanobacterial order Stigonematales strains (**Table 1-1**)[14]

Table 1-1. A summary of documented cyanobacterial strains which produce hapalindole-type alkaloids.

No.	Cyanophyte	Strain No.	Ref
[1]	*Hapalosiphon fontinalis* (Agardh) Bornet	ATCC 39694 [V-3-1]	13, 15-17
[2]	*Fischerella muscicola* (Thuret) Gomont	UTEX 1829	18-19
[3]	*Hapalosiphon welwitschii* W. & G.S. West	UH IC-52-3	20
	Hapalosiphon welwitschii	UTEX 1830	20
[4]	*Westiella intricate* Borzi	UH HT-29-1	20
[5]	*Hapalosiphon laingii* L. Hoffmann	89-785/4	21
[6]	*Fischerella* sp.	ATCC 43239	22
[7]	*Westiellopsis* sp.	SAG 20.93	19
[8]	*Fischerella* sp.	SAG 46.79	23
[9]	*Fischerella ambigua* (Nageli) Gomont	UTEX 1903	24-27
[10]	*Hapalosiphon hibernicus* W. & G.S. West	UH BZ-3-1	24
[11]	*Westiellopsis prolifica* Janet	UH EN-3-1;	24
[12]	*Hapalosiphon delicatulus* W. & G.S. West	UH IC-13-1	28
[13]	*Fischerella* sp.	TAU IL-199-3-1	29
[14]	*Fischerella muscicola* (Thuret) Gomont	HG-39-5 (UTEX 1301)	30
[15]	*Fischerella* major Gomont	HX-7-4	30
[16]	*Fischerella*	ATCC 53558	31-32
[17]	*Fischerella*	52-1	33

The polycyclic pattern of each core divides these natural products into four subgroups, including hapalindoles (tri-/tetracyclic), ambiguines (tetra-/pentacyclic), fischerindoles, and welwitindolinones (**Figure 1-2**). Although each subgroup has a distinct ring system, all molecules are united by several common features, most notably a *cis*-indole isonitrile and a geranyl-pyrophosphate monoterpene unit, a highly functionalized D-ring, most of them have a regio- and stereoselective chlorine substituent, and varying levels of oxidation.

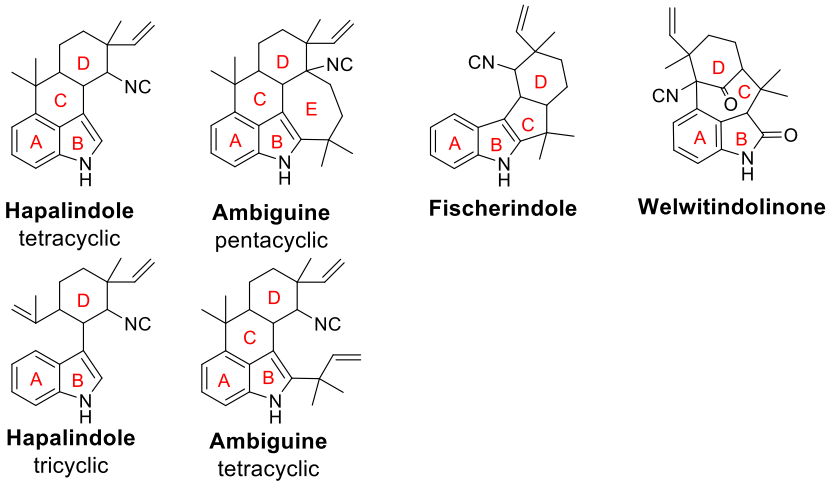


Figure 1-2. The common cores of hapalindole-type alkaloids.

1.3 Structural diversity

1.3.1 Hapalindole subgroup

Hapalindoles comprise the biggest A and B represent the first members of hapalindole subgroup.[13] Being the biggest subgroup with at least 31 members, the core ring skeleton of hapalindoles has the most complicated stereochemistry, which are categorized into six stereochemical patterns based on the chirality of the C10, C11, C12, and C15 stereocenters. In order to display the structural relationship, all hapalindoles are ordered according to their polycyclic skeleton (tri-/tetracycle), core stereochemistry, and various functional groups (**Figure 1-3**).

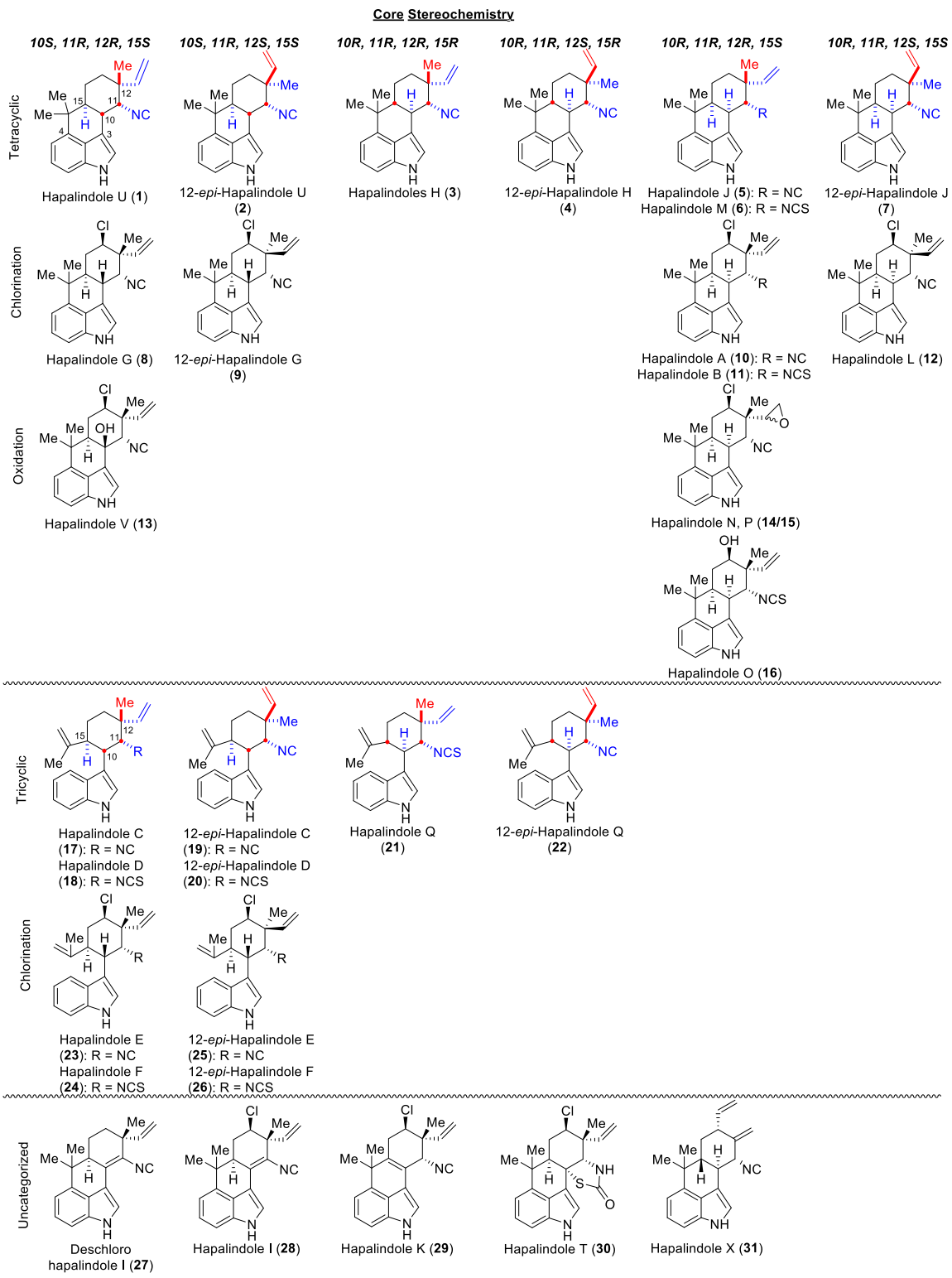


Figure 1-3. Isolated hapalindoles categorized according to the core stereochemistry and functional groups.

1.3.2 Fischerindole subgroup

The fischerindole subgroup is structurally differing from hapalindoles by indole C-2/3 ring fusion rather than C-3/4. The first member of fischerindole subgroup is fischerindole L (**35**) from *Fischerella muscicola* UTEX 1829 isolated in 1992, naming after the *Fischerella* genus and identical stereochemistry as hapalindole L (**12**, C10*R*, 11*R*, 12*S*, 15*S*). This metabolite was discovered through a bioassay-guided isolation looking for antifungal components toward *Aspergillus oryzae*, *Penicillium notatum*, *Saccharomyces cerevisiae*, and *Trichophyton mentagrophytes*.[18] Since then, another eight similar metabolites were isolated[18-20, 23, 30] including a new type of core stereochemistry (**32**-**34**, C10*S*, 11*R*, 12*S*, 15*S*). The structure of deschloro 12-*epi*-fischerindole W nitrile (**40**) was recently revised by Li et al.[34] The "12-*epi*" nomenclature of fischerindoles was intended to display the different stereochemistry at C-12 position from previously named hapalindoles.

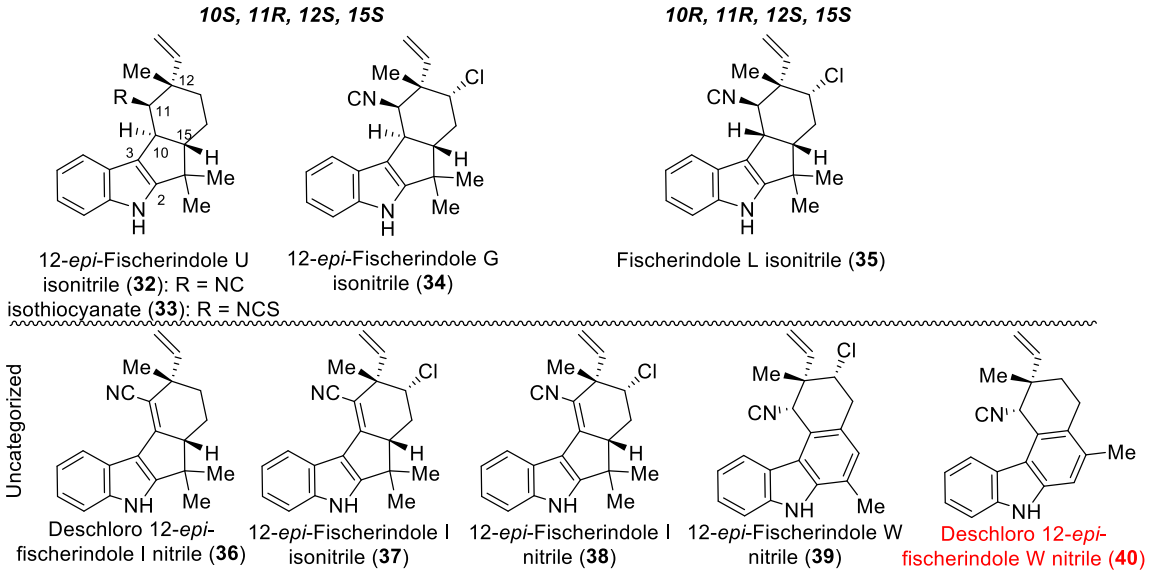


Figure 1-4. Isolated fischerindoles.

1.3.3 Ambiguine subgroup

The ambiguines possess the core skeleton of tetracyclic hapalindoles, but with an indole C-2 reverse prenyl group or further fused six/seven-membered

ring. The first ambiguine molecule was reported by Smitka and Moore in 1992. In a fungicidal activity-directed research, three species belong to the same Stigonemataceae family shown potential inhibition of fungi growth, viz. *Fischerella ambigua* UTEX 1903, *Hapalosiphon hibernicus* UH BZ-3-1, and *Westiellopsis prolifica* UH EN-3-1.[24] In total 6 novel ambiguines (A-F) were identified from strain *Fischerella ambigua*, ambiguine A and E were also isolated from *Hapalosiphon hibernicus*, while ambiguine D and E were found in *Westiellopsis prolifica*. So far, 20 members of ambiguine subgroup have been characterized by sharing the same stereochemistry with hapalindole U (**1**) at C10, C11, C12, C15, except the newly identified 12-*epi*-ambiguine B (**60**) from strain *Fischerella* 52-1.[33] This compound has not only a different configuration at C12 position, but also shown as nitrile instead of the common isonitrile group. The most intriguing structural feature of ambiguines is the heavily functionalized E-ring, including the ring formation through C-C bond formation and varying level of oxidation.

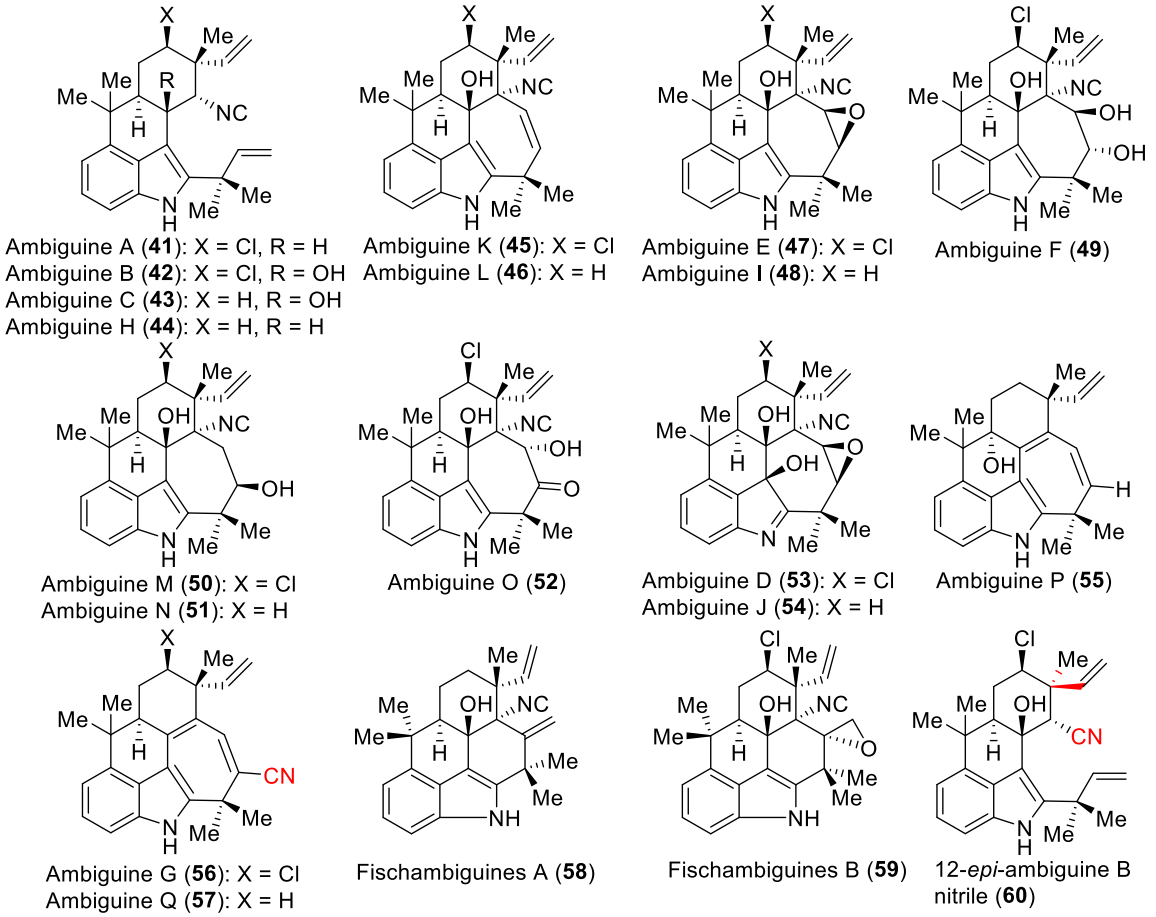


Figure 1-5. Isolated ambiguines.

1.3.4 Welwitindolinone subgroup

In 1994, the last subgroup, welwitindolinones were isolated from *hapalosiphon welwitschii*, reporting to be responsible for some of the multiple-drug-resistance reversing activity.[20] Now, in total ten members were isolated from four cyano strains (**Table 1-1**, [3], [4], [14], [15]). Welwitindolinones are notably different from the other three subgroups by its bicyclo[4.3.1]decane ring system (welwitindolinone A has a unique spirocyclic cyclobutene C-ring), which is proposed to be derived from fischerindoles through oxidation and following ring rearrangement (**Figure 1-9**). Besides, all but one (**70**) of these alkaloids contain a chloride at C-13 position, which contrasts with other subgroups.

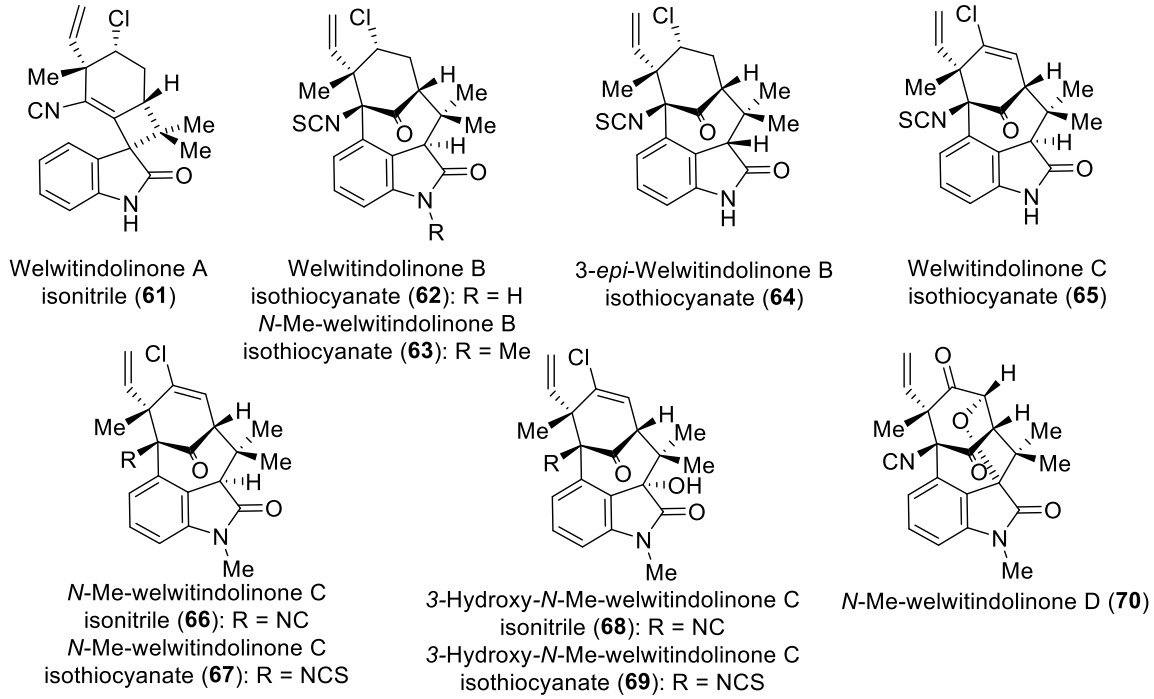


Figure 1-6. Isolated welwitindolinones.

1.3.5 Other unclassified compounds

Finally, there are another twelve hapalindole-related alkaloids, isolated along with previously discussed metabolites, that do not fit directly into the four subgroups. After analyzing their structures, however, it is reasonable to propose that compounds **74**-**77** are oxidized form of tetracyclic hapalindoles on the indole moiety, which would derive into compounds **78**-**83** through ring-opening. The two hapalindolinones (**72**, **73**) are also hypothesized to be derivatives from fischerindoles (**Figure 6-11**).

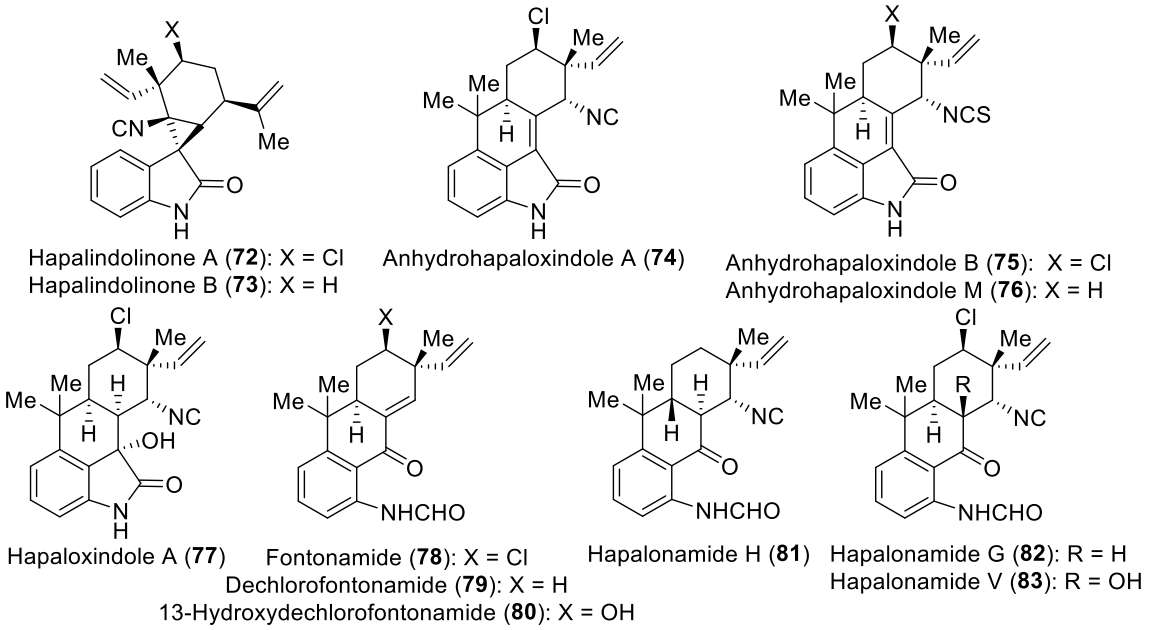


Figure 1-7. Unclassified hapalindole-related alkaloids.

1.4 Pharmaceutical significance

The hapalindole-type alkaloids display a diverse biological activity profiles, including antibacterial, antimycotic, insecticidal, and anticancer properties. This chemotype is worthy of further exploration especially as promising antiinfective lead compounds with activities against >20 gram-positive and gram-negative bacterial and fungal strains, including several declared by the CDC as serious antibiotic resistance threats (*Candida albicans*, *Mycobacterium tuberculosis*, *Staphylococcus aureus*).[14] Most importantly, hapalindoles exhibit antibacterial activity against *Staphylococcus aureus*, *Bacillus anthracis*, and *Mycobacterium tuberculosis*, and fungicidal activity against *Candida albicans*, *Trichophyton mentagrophytes*, and *Aspergillus fumigatus*. Several examples in particular include Hapalindole T (**30**) that inhibits the growth of *M. tuberculosis* with activity comparable to rifampicin, and is more effective than streptomycin.[35] Ambiguine I (**48**) inhibits *C. albicans* (MIC 0.39 μg/mL) with higher activity than amphotericin B (MIC 1.56 μg/mL).[29] Fischambiguine B (**59**) displays potent activity against *C. albicans* (MIC = 2 μM) without any detectable cytotoxicity against Vero cells (IC_{50}>128 μM).[27] The molecular mechanism of action for these antimicrobial

agents is mostly unclear and constitutes a further compelling justification for studying this chemotype.[36-37] Although we will focus our work on their potential as antiinfectives, some other activities for these indole alkaloids have been previously described.[38] For example, Welwitindolinone C (**66**) inhibits drug-resistant breast carcinoma with IC_{50} 0.12 uM;[38-39] and Hapalindolinone A (**72**) inhibits arginine vasopressin binding to kidney tissue (V_2 receptor) and V_2 vasopressin-stimulated adenylate cyclase, which could aid in treating congestive heart failure, hypertension, edema, and hyponatremia.

Due to the slow growth rate of Stigonematales cyanobacteria, and low isolation yield (0.01~0.05% of dry cell weight), large-scale fermentation of these compounds has not been feasible. Interest in their high level of functional group density and stereochemical complexity has driven efforts toward total synthesis of select molecules within this class, which have recently been reported (**Figure 1-8**). However, access to significant amounts of these compounds for antibiotic screening programs has been stymied (only half of the compounds have been screened for bioactivity with detailed IC_{50} or MIC values), as well as allied efforts to identify relevant biochemical targets to understand mechanism of action. All these facts motivated us to initiate biosynthetic studies, in purposes to develop Chemoenzymatic approach to access these unique bioactive molecules and analogs for drug discovery.

1.5 Total synthesis

The first total synthesis about this family was finished by Natsume et al. in 1989, a few years after the initial isolation.[40] The synthesized compounds were hapalindole J (**5**) and M (**6**) in a racemic form. Followed this work, in total 24 compounds have been successfully synthesized, including 11 hapalindoles, 4 fischerindoles, 7 welwitindolinones, 1 ambiguine H (**44**), and 1 hapalonamide H (**81**).[34, 40-69]

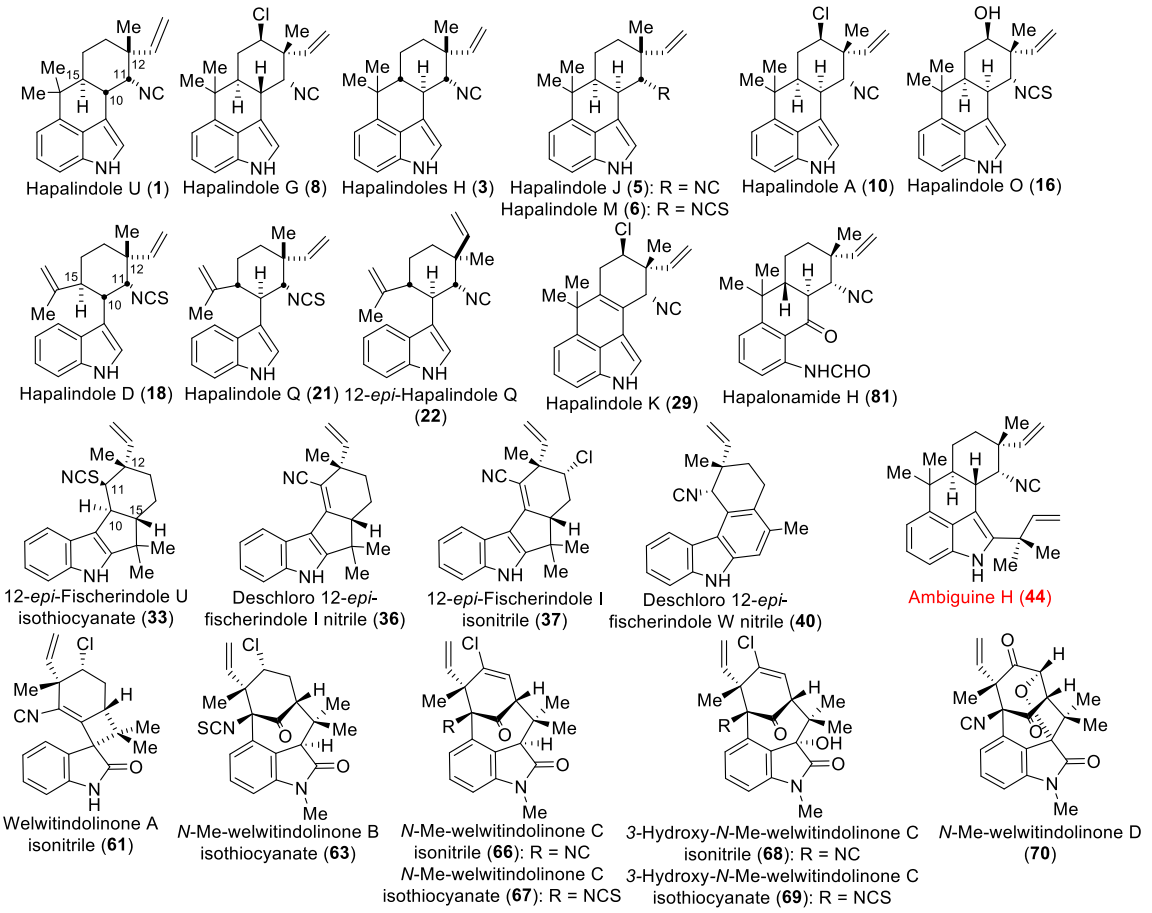


Figure 1-8. A summary of successfully synthesized hapalindole-type alkaloids.

1.6 Biosynthesis study

The initial hypothesis of hapalindole biosynthesis was formulated by Moore (**Figure 1-9**).[13, 18, 20] Although each subgroup has a distinct ring system, all hapalindole compounds share the common features of an indole isonitrile core (**1**) and a geranyl monoterpene unit, with many also containing a chlorine substituent. Inspired from the acid-catalyzed reactions about hapalindoles, especially the transformation of tricyclic hapalindoles to fischerindoles after treated with strong acid[70], Moore and co-workers proposed that tricyclic hapalindole is the first intermediate assembled by indole isonitrile and geranyl pyrophosphate (GPP) via a chloronium ion or a proton-catalyzed enzymatic cyclization. This intermediate can be further derived to tetracyclic hapalindoles or fischerindoles through enzyme-controlled, acid-catalyzed ring fusion at indole C-

4 or C-2, or elaborated into welwitindolinones through oxidation and acid-involved cyclization. The more structurally complex ambiguines are developed from tetracyclic hapalindoles via indole C-2 prenylation with dimethylallyl pyrophosphate (DMAPP), additional ring-closure of the tetracyclic ambiguines would afford the pentacyclic members. The formation of welwitindolinones in this proposal were modified by Baran.[56] Instead of generating from tricyclic hapalindoles, he suggested an oxidation of the indole C-2/3 bond of fischerindole followed by a rearrangement to produce welwitindolinones.

In 2007, a single-step cyclization was proposed by Carmeli (**Figure 1-9**, shown in the dash box).[29] Based on the observation that *Fischerella* sp. only produces tetracyclic hapalindoles and pentacyclic ambiguines, they suggested that the tetracyclic compounds would be synthesized in one step, induced by chloronium or proton ion.

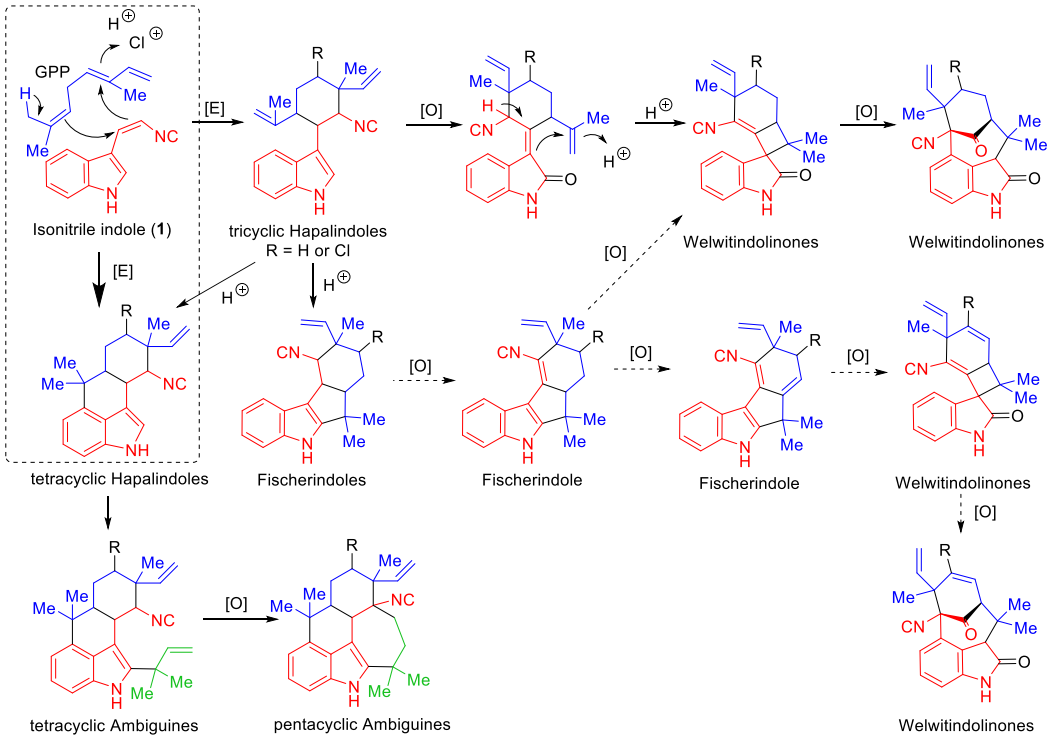


Figure 1-9. Current proposed biosynthetic formation of the polycyclic skeleton for the four hapalindole subgroups. Pipeline connected in solid arrows is proposed by Moore and co-workers; dash arrows are modification made by Baran and co-workers; pathway in dash box is created by Carmeli and co-workers as one-step cyclization.

1.7 Specific Aims

The ultimate goal of this biosynthetic study toward hapalindole-type natural products is to develop chemoenzymatic approach to synthesize native hapalindoles and its derivatives for drug discovery. The short-term goal was elucidating the biosynthetic pathway of hapalindoles, including the core-ring skeleton formation of hapalindoles and fischerindoles, and late-stage tailoring reactions. This thesis is mainly focused on regio- and stereospecific inter-intramolecular ring formation leading to various core systems. Our initial bioinformatics analysis revealed that all hapalindole-type alkaloids are derived from indole isonitrile and geranyl pyrophosphate, followed by C/D core-ring formation. Thus, by employing heterologous gene expression and in vitro enzymatic assays, we propose to: a) identify key enzymes involved in regio- and stereospecific assembling of the individual hapalindole and fischerindole cores; b) pursue molecular analysis of the key enzymes by applying site-directed mutagenesis and X-ray crystallography to probe the catalytic mechanism.

1.8 Reference

1. Antibiotic Resistance Threats in the United States, 2013. Prevention, C. f. D. C. a., Ed. CDC: 2013.
2. Ventola, C. L., The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics* **2015,** *40* (4), 277-283.
3. Coates, A. R.; Halls, G.; Hu, Y., Novel classes of antibiotics or more of the same? *Br. J. Pharmacol.* **2011,** *163* (1), 184-194.
4. Lewis, K., Antibiotics: Recover the lost art of drug discovery. *Nature* **2012,** *485* (7399), 439-440.
5. Newman, D. J.; Cragg, G. M., Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012,** *75* (3), 311-335.
6. Morrison, K. C.; Hergenrother, P. J., Natural products as starting points for the synthesis of complex and diverse compounds. *Natural Product Reports* **2014,** *31* (1), 6-14.
7. Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schaberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K., A new antibiotic kills pathogens without detectable resistance. *Nature* **2015,** *517* (7535), 455-+.
8. Henkel, T.; Brunne, R. M.; Muller, H.; Reichel, F., Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem. Int. Ed.* **1999,** *38* (5), 643-647.
9. Frye, S.; Crosby, M.; Edwards, T.; Juliano, R., US academic drug discovery. *Nat. Rev. Drug Discov.* **2011,** *10* (6), 409-410.
10. Singh, R. K.; Tiwari, S. P.; Rai, A. K.; Mohapatra, T. M., Cyanobacteria: an emerging source for drug discovery. *J. Antibiotics* **2011,** *64* (6), 401-412.
11. Patterson, G. M. L.; Baldwin, C. L.; Bolis, C. M.; Caplan, F. R.; Karuso, H.; Larsen, L. K.; Levine, I. A.; Moore, R. E.; Nelson, C. S.; Tschappat, K. D.; Tuang, G. D.; Furusawa, E.; Furusawa, S.; Norton, T. R.; Raybourne, R. B., Antineoplastic Activity of Cultured Blue-Green-Algae (Cyanophyta). *J Phycol* **1991,** *27* (4), 530-536.
12. Gerwick, W. H.; Proteau, P. J.; Nagle, D. G.; Hamel, E.; Blokhin, A.; Slate, D. L., Structure of Curacin-a, a Novel Antimitotic, Antiproliferative, and Brine Shrimp Toxic Natural Product from the Marine Cyanobacterium Lyngbya-Majuscula. *J. Org. Chem.* **1994,** *59* (6), 1243-1245.
13. Moore, R. E.; Cheuk, C.; Patterson, G. M. L., Hapalindoles: new alkaloids from the blue-green alga Hapalosiphon fontinalis. *J. Am. Chem. Soc.* **1984,** *106* (21), 6456-6457.
14. Bhat, V.; Dave, A.; MacKay, J. A.; Rawal, V. H., Chapter two - the chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. In *The Alkaloids: Chemistry and Biology*, Hans-Joachim, K., Ed. Academic Press: San Diego, 2014; Vol. 73, pp 65-160.
15. Moore, R. E.; Cheuk, C.; Yang, X. Q. G.; Patterson, G. M. L.; Bonjouklian, R.; Smitka, T. A.; Mynderse, J. S.; Foster, R. S.; Jones, N. D.,

Hapalindoles, antibacterial and antimycotic alkaloids from the cyanophyte *Hapalosiphon fontinalis*. *J. Org. Chem.* **1987,** *52* (6), 1036-1043.
16. Moore, R. E.; Yang, X. Q. G.; Patterson, G. M. L., Fontonamide and anhydrohapaloxindole A, two new alkaloids from the blue-green alga Hapalosiphon fontinalis. *J. Org. Chem.* **1987,** *52* (17), 3773-3777.
17. Moore, R. E.; Yang, X.-q. G.; Patterson, G. M. L.; Bonjouklian, R.; Smitka, T. A., Hapalonamides and other oxidized hapalindoles from Hapalosiphon fontinalis. *Phytochem.* **1989,** *28* (5), 1565-1567.
18. Park, A.; Moore, R. E.; Patterson, G. M. L., Fischerindole L, a new isonitrile from the terrestrial blue-green alga *Fischerella muscicola*. *Tetrahedron Lett.* **1992,** *33* (23), 3257-3260.
19. Kim, H.; Lantvit, D.; Hwang, C. H.; Kroll, D. J.; Swanson, S. M.; Franzblau, S. G.; Orjala, J., Indole alkaloids from two cultured cyanobacteria, *Westiellopsis* sp. and *Fischerella muscicola*. *Bioorg. Med. Chem.* **2012,** *20* (17), 5290-5295.
20. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, unusual alkaloids from the blue-green algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to fischerindoles and hapalinodoles. *J. Am. Chem. Soc.* **1994,** *116* (22), 9935-9942.
21. Klein, D.; Daloze, D.; Braekman, J. C.; Hoffmann, L.; Demoulin, V., New Hapalindoles from the Cyanophyte Hapalosiphon laingii. *J. Nat. Prod.* **1995,** *58* (11), 1781-1785.
22. Becher, P. G.; Keller, S.; Jung, G.; Sussmuth, R. D.; Juttner, F., Insecticidal activity of 12-epi-hapalindole J isonitrile. *Phytochem.* **2007,** *68* (19), 2493-2497.
23. Kim, H.; Krunic, A.; Lantvit, D.; Shen, Q.; Kroll, D. J.; Swanson, S. M.; Orjala, J., Nitrile-containing fischerindoles from the cultured cyanobacterium *Fischerella* sp. *Tetrahedron* **2012,** *68* (15), 3205-3209.
24. Smitka, T. A.; Bonjouklian, R.; Doolin, L.; Jones, N. D.; Deeter, J. B.; Yoshida, W. Y.; Prinsep, M. R.; Moore, R. E.; Patterson, G. M. L., Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae. *J. Org. Chem.* **1992,** *57* (3), 857-861.
25. Lin, Y.; Schiavo, S.; Orjala, J.; Vouros, P.; Kautz, R., Microscale LC-MS-NMR platform applied to the identification of active cyanobacterial metabolites. *Anal. Chem.* **2008,** *80* (21), 8045-8054.
26. Mo, S.; Krunic, A.; Chlipala, G.; Orjala, J., Antimicrobial ambiguine isonitriles from the cyanobacterium Fischerella ambigua. *J. Nat. Prod.* **2009,** *72* (5), 894-899.
27. Mo, S.; Krunic, A.; Santarsiero, B. D.; Franzblau, S. G.; Orjala, J., Hapalindole-related alkaloids from the cultured cyanobacterium Fischerella ambigua. *Phytochem.* **2010,** *71* (17-18), 2116-2123.
28. Huber, U.; Moore, R. E.; Patterson, G. M. L., Isolation of a nitrile-containing indole alkaloid from the terrestrial blue-green alga hapalosiphon delicatulus. *J. Nat. Prod.* **1998,** *61* (10), 1304-1306.

29. Raveh, A.; Carmeli, S., Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* **2007,** *70* (2), 196-201.
30. Jimenez, J. I.; Huber, U.; Moore, R. E.; Patterson, G. M., Oxidized welwitindolinones from terrestrial fischerella spp. *J Nat Prod* **1999,** *62* (4), 569-572.
31. Schwartz, R. E.; Hirsch, C. F.; Springer, J. P.; Pettibone, D. J.; Zink, D. L., Unusual cyclopropane-containing hapalindolinones from a cultured cyanobacterium. *J. Org. Chem.* **1987,** *52* (16), 3704-3706.
32. Schwartz, R. E.; Hirsch, C. F.; Sesin, D. F.; Flor, J. E.; Chartrain, M.; Fromtling, R. E.; Harris, G. H.; Salvatore, M. J.; Liesch, J. M.; Yudin, K., Pharmaceuticals from Cultured Algae. *J. Ind. Microbiol.* **1990,** *5* (2-3), 113-123.
33. Walton, K.; Gantar, M.; Gibbs, P.; Schmale, M.; Berry, J., Indole Alkaloids from Fischerella Inhibit Vertebrate Development in the Zebrafish (Danio rerio) Embryo Model. *Toxins* **2014,** *6* (12), 3568-3581.
34. Lu, Z.; Yang, M.; Chen, P.; Xiong, X.; Li, A., Total Synthesis of Hapalindole-Type Natural Products. *Angew. Chem. Int. Ed.* **2014,** *53* (50), 13840-13844.
35. Asthana, R.; Srivastava, A.; Singh, A.; Deepali; Singh, S.; Nath, G.; Srivastava, R.; Srivastava, B., Identification of an antimicrobial entity from the cyanobacterium Fischerella sp. isolated from bark of Azadirachta indica (Neem) tree. *J. Appl. Phycol.* **2006,** *18* (1), 33-39.
36. Doan, N. T.; Stewart, P. R.; Smith, G. D., Inhibition of bacterial RNA polymerase by the cyanobacterial metabolites 12-epi-hapalindole E isonitrile and calothrixin A. *FEMS Microbiol. Lett.* **2001,** *196* (2), 135-139.
37. Thanh Doan, N.; Rickards, R.; Rothschild, J.; Smith, G., Allelopathic actions of the alkaloid 12-epi-hapalindole E isonitrile and calothrixin A from cyanobacteria of the genera Fischerella and Calothrix. *J. Appl. Phycol.* **2000,** *12* (3-5), 409-416.
38. Zhang, X.; Smith, C. D., Microtubule effects of welwistatin, a cyanobacterial indolinone that circumvents multiple drug resistance. *Mole. Pharmacol.* **1996,** *49* (2), 288-294.
39. Smith, C. D.; Zilfou, J. T.; Stratmann, K.; Patterson, G. M.; Moore, R. E., Welwitindolinone analogues that reverse P-glycoprotein-mediated multiple drug resistance. *Mol. Pharmacol.* **1995,** *47* (2), 241-247.
40. Muratake, H.; Natsume, M., Total synthesis of marine alkaloids (±)-hapalindoles J and M. *Tetrahedron Lett.* **1989,** *30* (14), 1815-1818.
41. Reyes, J. R.; Xu, J.; Kobayashi, K.; Bhat, V.; Rawal, V. H., Total Synthesis of (-)-*N*-Methylwelwitindolinone B Isothiocyanate. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (33), 9962-9966.
42. Liu, Y.; Cheng, L.-J.; Yue, H.-T.; Che, W.; Xie, J.-H.; Zhou, Q.-L., Divergent enantioselective synthesis of hapalindole-type alkaloids using catalytic asymmetric hydrogenation of a ketone to construct the chiral core structure. *Chem. Sci.* **2016,** *7* (7), 4725-4729.

43. Maimone, T. J.; Ishihara, Y.; Baran, P. S., Scalable Total Syntheses of (-)-Hapalindole U and (+)-Ambiguine H. *Tetrahedron* **2015,** *71* (22), 3652-3665.
44. Komine, K.; Nomura, Y.; Ishihara, J.; Hatakeyama, S., Total Synthesis of (-)-*N*-Methylwelwitindolinone C Isothiocyanate Based on a Pd-Catalyzed Tandem Enolate Coupling Strategy. *Org. Lett.* **2015,** *17* (15), 3918-3921.
45. Weires, N. A.; Styduhar, E. D.; Baker, E. L.; Garg, N. K., Total Synthesis of (−)-N-Methylwelwitindolinone B Isothiocyanate via a Chlorinative Oxabicycle Ring-Opening Strategy. *J. Am. Chem. Soc.* **2014,** *136* (42), 14710-14713.
46. Styduhar, E. D.; Huters, A. D.; Weires, N. A.; Garg, N. K., Enantiospecific Total Synthesis of N-Methylwelwitindolinone D Isonitrile. *angew. Chem. Int. Ed.* **2013,** *52* (47), 12422-12425.
47. Rafferty, R. J.; Williams, R. M., Formal Synthesis of Hapalindole O and Synthetic Efforts Towards Hapalindole K and Ambiguine A. *Heterocycles* **2012,** *86* (1), 219-231.
48. Rafferty, R. J.; Williams, R. M., Total synthesis of hapalindoles J and U. *J. Org. Chem.* **2012,** *77* (1), 519-524.
49. Quasdorf, K. W.; Huters, A. D.; Lodewyk, M. W.; Tantillo, D. J.; Garg, N. K., Total Synthesis of Oxidized Welwitindolinones and (−)-N-Methylwelwitindolinone C Isonitrile. *J. Am. Chem. Soc.* **2012,** *134* (3), 1396-1399.
50. Fu, T.-h.; McElroy, W. T.; Shamszad, M.; Martin, S. F., Formal Syntheses of Naturally Occurring Welwitindolinones. *Org. Lett.* **2012,** *14* (15), 3834-3837.
51. Allan, K. M.; Kobayashi, K.; Rawal, V. H., A Unified Route to the Welwitindolinone Alkaloids: Total Syntheses of (−)-N-Methylwelwitindolinone C Isothiocyanate, (−)-N-Methylwelwitindolinone C Isonitrile, and (−)-3-Hydroxy-N-methylwelwitindolinone C Isothiocyanate. *J. Am. Chem. Soc.* **2012,** *134* (3), 1392-1395.
52. Huters, A. D.; Quasdorf, K. W.; Styduhar, E. D.; Garg, N. K., Total Synthesis of (−)-N-Methylwelwitindolinone C Isothiocyanate. *J. Am. Chem. Soc.* **2011,** *133* (40), 15797-15799.
53. Chandra, A.; Johnston, J. N., Total Synthesis of the Chlorine-Containing Hapalindoles K, A, and G. *Angew. Chem. Int. Ed.* **2011,** *50* (33), 7641-7644.
54. Bhat, V.; Rawal, V. H., Stereocontrolled synthesis of 20,21-dihydro N-methylwelwitindolinone B isothiocyanate. *Chem. Commun.* **2011,** *47* (34), 9705-9707.
55. Bhat, V.; Allan, K. M.; Rawal, V. H., Total Synthesis of N-Methylwelwitindolinone D Isonitrile. *J. Am. Chem. Soc.* **2011,** *133* (15), 5798-5801.
56. Richter, J. M.; Ishihara, Y.; Masuda, T.; Whitefield, B. W.; Llamas, T.; Pohjakallio, A.; Baran, P. S., Enantiospecific total synthesis of the hapalindoles, fischerindoles, and welwitindolinones via a redox economic approach. *J. Am. Chem. Soc.* **2008,** *130* (52), 17938-17954.

57. Reisman, S. E.; Ready, J. M.; Weiss, M. M.; Hasuoka, A.; Hirata, M.; Tamaki, K.; Ovaska, T. V.; Smith, C. J.; Wood, J. L., Evolution of a Synthetic Strategy: Total Synthesis of (±)-Welwitindolinone A Isonitrile. *J. Am. Chem. Soc.* **2008,** *130* (6), 2087-2100.
58. Baran, P. S.; Maimone, T. J.; Richter, J. M., Total synthesis of marine natural products without using protecting groups. *Nature* **2007,** *446* (7134), 404-408.
59. Reisman, S. E.; Ready, J. M.; Hasuoka, A.; Smith, C. J.; Wood, J. L., Total Synthesis of (±)-Welwitindolinone A Isonitrile. *J. Am. Chem. Soc.* **2006,** *128* (5), 1448-1449.
60. Baran, P. S.; Richter, J. M., Enantioselective total syntheses of welwitindolinone A and fischerindoles I and G. *J. Am. Chem. Soc.* **2005,** *127* (44), 15394-15396.
61. Baran, P. S.; Richter, J. M., Direct Coupling of Indoles with Carbonyl Compounds: Short, Enantioselective, Gram-Scale Synthetic Entry into the Hapalindole and Fischerindole Alkaloid Families. *J. Am. Chem. Soc.* **2004,** *126* (24), 7450-7451.
62. Kinsman, A. C.; Kerr, M. A., The Total Synthesis of (+)-Hapalindole Q by an Organomediated Diels−Alder Reaction. *J. Am. Chem. Soc.* **2003,** *125* (46), 14120-14125.
63. Kinsman, A. C.; Kerr, M. A., Total Synthesis of (±)-Hapalindole Q. *Org. Lett.* **2001,** *3* (20), 3189-3191.
64. Sakagami, M.; Muratake, H.; Natsume, M., Preparation of Alkyl-Substituted Indoles in the Benzene Portion. Part 12. Enantiospecific Synthesis of Hapalindole O. *Chem. Pharm. Bull.* **1994,** *42* (7), 1393-1398.
65. Fukuyama, T.; Chen, X., Stereocontrolled Synthesis of (-)-Hapalindole G. *J. Am. Chem. Soc.* **1994,** *116* (7), 3125-3126.
66. Vaillancourt, V.; Albizati, K. F., Synthesis and absolute configuration of (+)-hapalindole Q. *J. Am. Chem. Soc.* **1993,** *115* (9), 3499-3502.
67. Muratake, H.; Natsume, M., Synthetic studies of marine alkaloids hapalindoles. Part I Total synthesis of (±)-hapalindoles J and M. *Tetrahedron* **1990,** *46* (18), 6331-6342.
68. Muratake, H.; Natsume, M., Synthetic studies of marine alkaloids hapalindoles. Part 2. Lithium aluminum hydride reduction of the electron-rich carbon-carbon double bond conjugated with the indole nucleus. *Tetrahedron* **1990,** *46* (18), 6343-6350.
69. Muratake, H.; Kumagami, H.; Natsume, M., Synthetic studies of marine alkaloids hapalindoles. Part 3 Total synthesis of (±)-hapalindoles H and U. *Tetrahedron* **1990,** *46* (18), 6351-6360.
70. Bonjouklian, R.; Moore, R. E.; Patterson, G. M. L., Acid-Catalyzed Reactions of Hapalindoles. *J. Org. Chem.* **1988,** *53* (25), 5866-5870.

Chapter 2
Hapalindole/ambiguine biogenesis is mediated by a Cope re-arrangement, C-C bond-forming cascade

2.1 Abstract

In this chapter, we describe the *fam* gene cluster from the cyanobacterium *Fischerella ambigua* UTEX 1903 encoding hapalindole and ambiguine biosynthesis along with the characterization of two aromatic prenyltransferases, FamD1 and FamD2, and a previously undescribed cyclase, FamC1. These studies demonstrate that FamD2 and FamC1 act in concert to form the tetracyclic core ring system of the hapalindoles from cis-indole isonitrile and geranyl pyrophosphate through a presumed biosynthetic Cope rearrangement and subsequent *6-exo-trig* cyclization/electrophilic aromatic substitution reaction.

2.2 Introduction

Hapalindoles are a large group of structurally fascinating indole alkaloids from the cyanobacterial order Stigonematales with diverse biological activity profiles,[1] including antibacterial,[2] antimycotic,[3-4] insecticidal,[5-6] and anticancer properties.[7] This family of metabolites is comprised of at least 81 members isolated from over 18 documented cyanobacterial strains,[1, 8] and is classified into four subgroups. These subgroups include the hapalindoles (tri-/tetracyclic), ambiguines, fischerindoles, and welwitindolinones, as distinguished by their variant polycyclic ring systems (**Figure 2-1**A). Because of their noteworthy biological properties, as well as daunting structural features, several groups have initiated research programs devoted to the total synthesis of these molecules.[1] However, the challenge of obtaining the highly functionalized ring systems in a

regio- and stereospecific manner has impeded these efforts with some notable exceptions (**Figure 1-8**). In contrast to this large body of work, limited research concerning the biogenesis[9-12] of these structurally diverse terpenoids motivated us to initiate biosynthetic studies.

Although each subgroup has a distinct ring system, all hapalindole compounds share the common features of an indole isonitrile core (**1**) and a geranyl monoterpene unit (**Figure 2-1**B), with many also containing a chlorine substituent. While the biosynthesis of these shared subunits has been recently identified by Liu and co-workers,[9, 11] the key fusion step to generate the regio- and stereospecific ring systems from **1** and geranyl pyrophosphate (GPP) remained unknown. Current hypotheses include a proton-promoted two-step polyolefin cyclization to first furnish tricyclic hapalindoles and subsequently tetracyclic hapalindoles and fischerindoles via electrophilic aromatic substitution,[13-16] or a one-step enzymatic cyclization to directly form tetracyclic hapalindoles and fischerindoles (**Figure 2-1**B).[9, 17] In this work, we disclose a surprising intermediate generated by the aromatic prenyltransferases and the allied cyclase responsible for forming the tetracyclic core structure of this family, which provides key insights into a new mechanistic proposal.

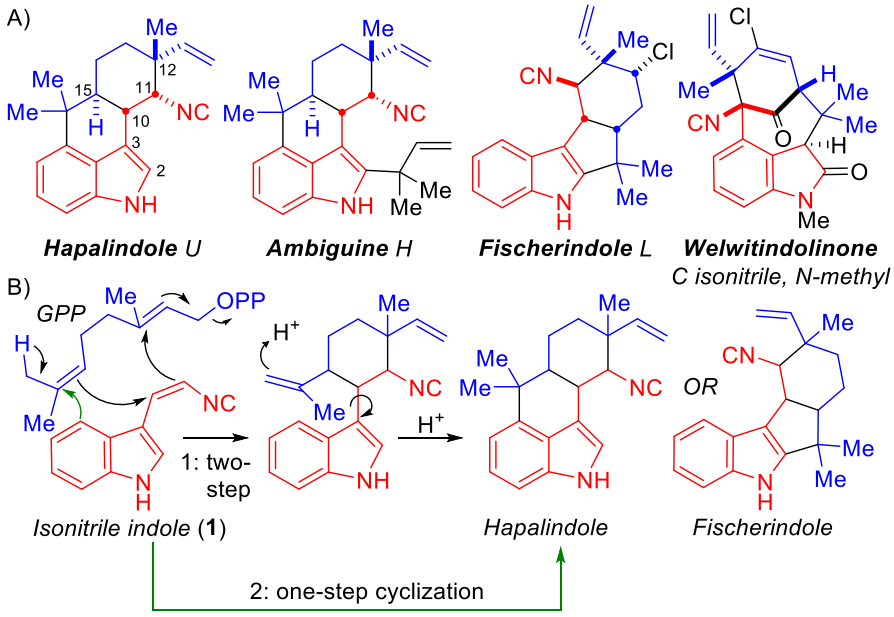


Figure 2-1. Representatives of hapalindole-type alkaloids and proposed mechanism. A) Examples of the four hapalindole alkaloid subgroups with the indole core (red) and monoterpenoid group (blue) highlighted; B) Current proposed hypotheses for biosynthetic formation of the tri- and tetracyclic cores. 1) Moore proposed a proton-promoted two-step cyclization;[14] 2) Carmeli proposed a concerted enzymatic cyclization.[17]

2.3 Research and Discussion

We initiated our investigation using the hapalindole/ambiguine producing strain *Fischerella ambigua* UTEX 1903[3-4, 18] as a model. The whole genome was independently sequenced and mined for the ambiguine biosynthetic gene cluster using prenyltransferases as an *in silico* probe, revealing a 41 kb region encoding 32 proteins that was putatively annotated as the ambiguine biosynthetic gene cluster (*fam*) (**Figure 2-2**A). This sequence is 100% identical to the recently published *amb* gene cluster[9] (**Figure 2-2**B), and now includes the complete protein functional annotation for FamC1-4 based on our findings detailed below.

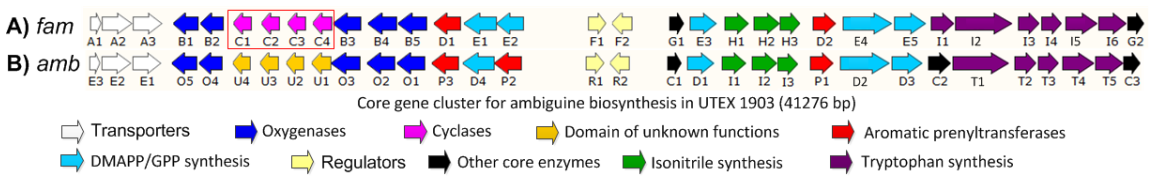


Figure 2-2. Comparison of the gene cluster from *Fischerella ambigua* UTEX 1903. A) *fam* gene cluster identified in this study. B) *amb* gene cluster assigned previously.[9]

2.3.1 Aromatic prenyltransferases FamD1, FamD2

FamD1 and FamD2 align with ABBA aromatic prenyltransferases that catalyze prenyl transfer to aromatic groups.[19-21] Thus, the presence of a monoterpene unit and a dimethylallyl group in the ambiguine natural products indicate that FamD1 and FamD2 may play critical roles in fusing **1** with GPP to generate tri- or tetracyclic hapalindoles (**Figure 2-1**B), and prenylating tetracyclic hapalindoles with dimethylallyl pyrophosphate (DMAPP) to form the ambiguines (**Figure 2-1**A). Hence, we cloned, overexpressed, and purified FamD1 and FamD2 in *E. coli* as 6x His-tagged proteins to assess their biochemical function *in vitro*. Contrary to previous reports,[9, 12] we observed that FamD2 prenylates **1** with GPP to give two distinct products that we have isolated and identified as novel geranylated C-2 and C-3 isocyano-indoles **2** and **3** (**Figure 2-3**A, **SI Table 2-2** and **SI Table 2-3**). Their production is pH dependent: a slightly acidic pH (6.0) gives C-2 prenylation (**2**) as the major product, and a high pH (10.5) gives C-3 prenylation (**3**) as the major product (**Figure 2-3**C). Compound **3** rearranges to **2** in the presence of FamD2 at pH 6.0 and 7.8 but not pH 10.5, while **2** did not convert to **3** at any pH tested (**SI Figure 2-1**). FamD2 can also accept DMAPP and react with **1** in a pH dependent manner to give dimethylallylated isocyano-indoles **4** and presumably **5** (**Figure 2-3**B,**SI Table 2-4**), the latter structure being determined by comparing its ^{1}H/COSY NMR and MS to **4** and **3**. We also examined the activity of FamD1 toward substrate **1**. Surprisingly, FamD1 was able to prenylate **1** with GPP and DMAPP in an analogous manner to FamD2, albeit with much lower efficiency. Compared to DMAPP, FamD1 shows diminished conversion with GPP and produced only **2** (**Figure 2-3**D-ii). FamD1 also preferentially accepted DMAPP when incubated with **1**:DMAPP:GPP (1:1:1) (**SI Figure 2-2**), indicating that DMAPP is the natural substrate for FamD1.

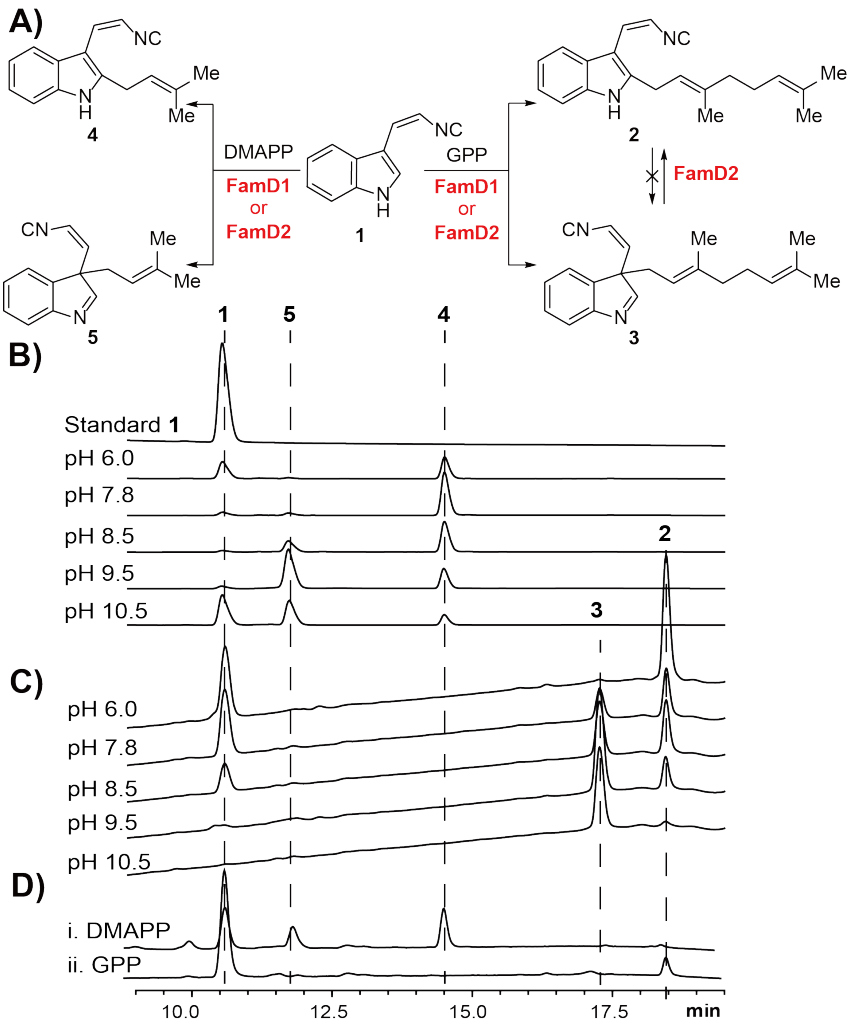


Figure 2-3. *In vitro* assay of aromatic prenyltransferases FamD1 and FamD2. A) The four products generated by FamD1 and FamD2 in reaction with GPP or DMAPP. B) pH effects of **1** and DMAPP treated with FamD2. C) pH effects of **1** and GPP treated with FamD2. D) An assessment of FamD1 reactivity showing production of i) **4** and **5** from **1** and DMAPP; and ii) **2** from **1** and GPP.

We also investigated whether FamD1 or FamD2 could tailor tetracyclic hapalindole U to ambiguine H (**Figure 2-4**) and confirmed that only FamD1 was capable of this transformation. Neither enzyme prenylated hapalindole U with GPP.

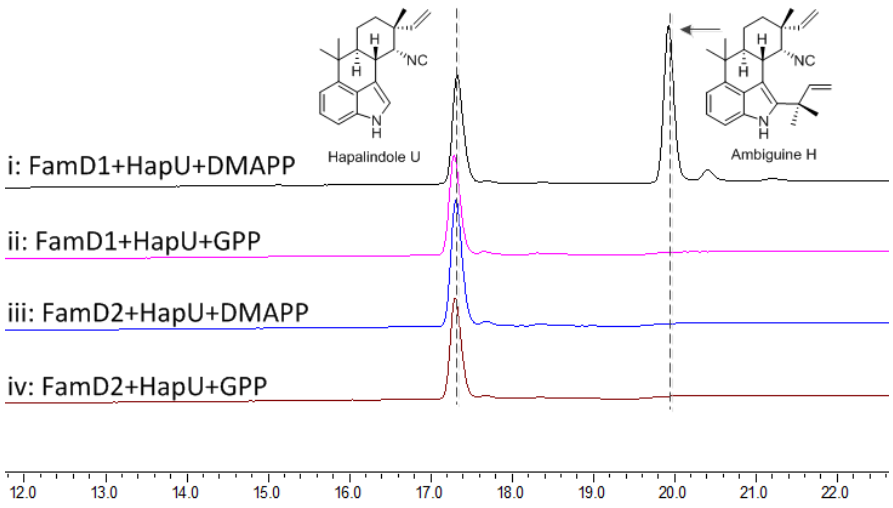


Figure 2-4. *In vitro* assay with Hapalindole U and FamD1/FamD2. Hap U was incubated with DMAPP/GPP and FamD1/FamD2, and only FamD1 can prenylate hapalindole U with DMAPP to generate Ambiguine H which was confirmed by NMR.

These results confirm that FamD1 is responsible for processing tetracyclic hapalindoles into ambiguines and that FamD2 prenylates **1** with GPP, although not to the tri- or tetracyclic hapalindoles. Instead, FamD2 catalyzes the first step of a ring-forming cascade, generating an intermediate that we reasoned subsequently serves as substrate for an additional enzyme responsible for the final ring formations. Thus, as the ultimate goal for this investigation, we set out to determine which gene encodes the hapalindole Fam cyclase.

2.3.2 Cell-free lysate

After the characterization of FamD1 and FamD2, the remaining genes that could play a role in the cyclization included five oxygenases (FamB1-5) and four genes previously annotated as unknown proteins (FamC1-4). Additionally, it was possible that the gene encoding the cyclase was located outside of the defined gene cluster. To identify the gene(s) responsible for cyclization, we employed traditional lysate fractionation techniques to isolate and identify the enzyme responsible.

F. ambigua was cultured and filtered to obtain sufficient cell mass, which was disrupted and extracted to give a crude cell-free lysate.[22] We tested the

lysate activity by reacting it with **1** and GPP, and two new products (**6** and **7**) were generated (**Figure 2-5** i). We examined the possible relationship between **2**/**3** and **6**/**7** (**Figure 2-5** iv-ix) by treating **2** and **3** with cell-free lysate, observing **2** is unreactive (**Figure 2-5** iv-vi) while **3** converts into a mixture of **6** and **7** (**Figure 2-5** vii, viii). Surprisingly, the negative control where the cell-free lysate was boiled for 5 min (**Figure 2-5** ix) showed the formation of **6** only. Boiling the lysate for an extended period (15 min) showed a decreased conversion to **6** (**SI Figure 2-3**). These results suggested that **2** is a shunt metabolite, while **3** is an intermediate between starting materials **1** and GPP on the biosynthetic pathway to **6** and **7**, and a thermostable protein in the boiled supernatant cyclizes **3** to **6.** The protein responsible for forming **7** from **3** is present in the cell-free lysate but is thermolabile. A large scale reaction enabled us to identify **6** as a previously unreported tetracycle, 12-*epi*-hapalindole U (**SI Table 2-5**); the structure of **7** was latterly identified as hapalindole H (see Chapter 3).

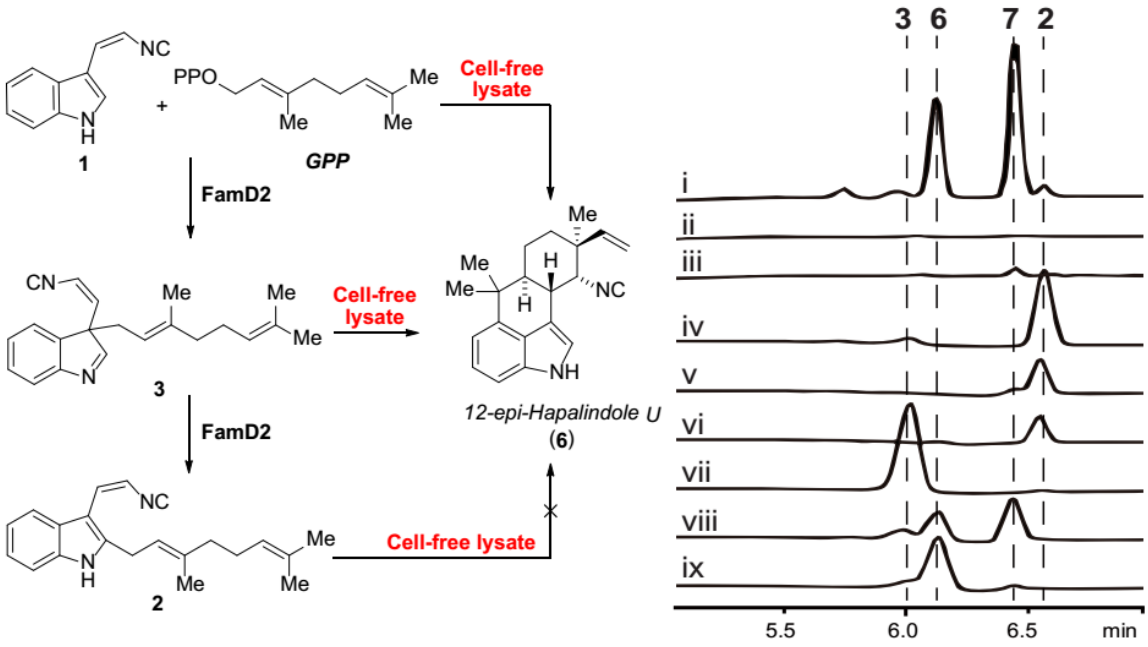


Figure 2-5. *In vitro* assays with cell-free lysate. i) **1**+GPP+cell-free lysate; ii) **1**+GPP+boiled cell-free lysate; iii) blank with only cell-free lysate; iv) standard of **2**; v) **2+**cell-free lysate; vi) **2**+boiled cell-free lysate; vii) standard of **3**; viii) **3+**cell-free lysate; ix) **3**+boiled cell-free lysate.

2.3.3 Stig cyclase FamC1

We examined the composition of the cell-free lysate and boiled supernatant by protein gel electrophoresis (**SI Figure 2-4**) and subsequently purified the active cyclase directly from the supernatant after finding that it contained relatively few proteins. Thus, the lysate was boiled for 5 min prior to fractionation on an anion exchange column, and the resulting fractions were tested for activity with **3** (**Figure 2-6**A). The HPLC traces revealed that some fractions retained the ability to effectively cyclize **3** to **6** (**Figure 2-6**B).

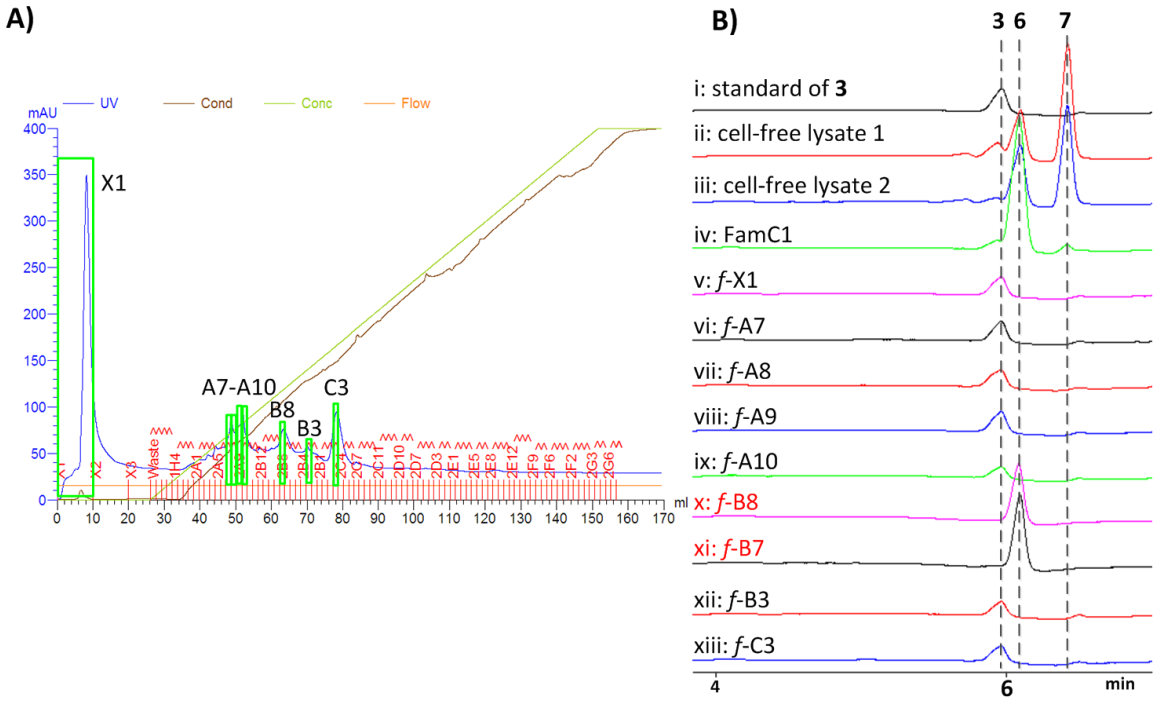


Figure 2-6. Protein fractionation with boiled cell-free lysate. A) FPLC spectrum of the ion exchange column purification. The fractions in green squares were tested for *in vitro* enzymatic assay. B) HPLC traces from the *in vitro* assay. v-xiii are fractions from purification and incubated with molecule **3**. *f*-B8 (x) and *f*-B7 (xi) showed the ability to catalyze the transformation of **3** into **6**.

Protein gel analysis of the active fraction showed one major band with a second minor component (**Figure 2-7**A). These bands were excised and analyzed at the University of Michigan Proteomics & Peptide Synthesis Core, with the active protein being conclusively identified as FamC1; while the minor component was identified as an allophycocyanin protein (**Figure 2-7**B and C).

BLAST analysis of the three additional previously unknown proteins in the *fam* gene cluster against FamC1 showed high sequence identities of 63%, 63%, and 73%, respectively, thus leading us to propose that the encoding genes represent a new class of cyclases (*fam*C1-4) (**SI Figure 2-5**).

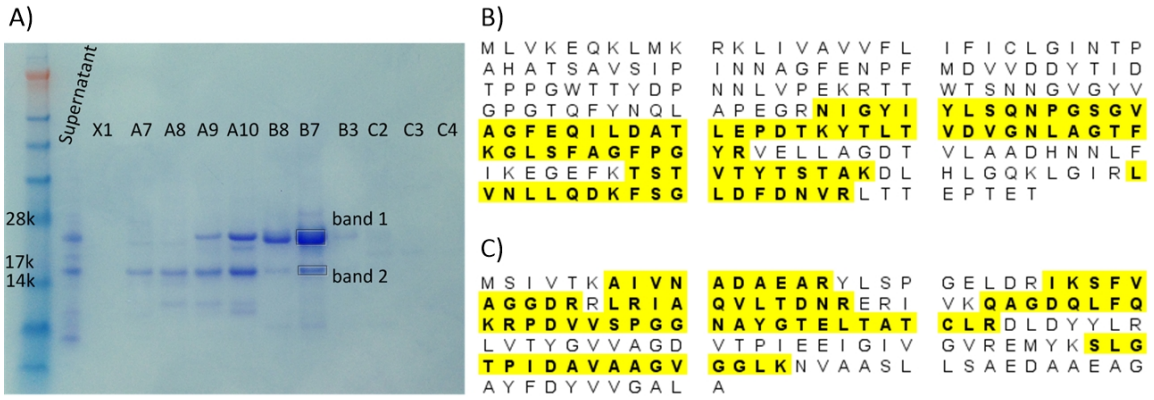


Figure 2-7. Proteomics analysis. A) SDS-PAGE gel image for ion exchange purification fractions. *f*-B8 and B7 are the active fractions, and band 1 and 2 were excised and analyzed at the University of Michigan Proteomics & Peptide Synthesis Core. B) Protein identification of band 1. The highlighted residues were identified as contiguous peptide segments encoded exclusively by the *fam*C1 gene of *F. ambigua* UTEX 1903. C) Protein identification of band 2. The highlighted residues were identified as contiguous peptide segments encoded in the allophycocyanin protein in *F. ambigua*. Allophycocyanin is a protein belongs to the light-harvesting phycobiliprotein family.

2.3.4 Proposed mechanism

The results above indicate that an unexpected biosynthetic mechanism is mediating formation of the core common to all hapalindoles, ambiguines, fischerindoles and welwitindolinones, in contrast to previous proposals (**Figure 2-1**B).[9, 13-17] Our results suggest the mechanism begins with the FamD2 catalyzed prenylation of GPP to the C3 position of **1** (**Figure 2-8**), generating a quaternary product (**3**) that can re-aromatize to form C2 prenylated shunt metabolite (**2**) in the presence of FamD2 at neutral or low pH (**SI Figure 2-1**).[23-25] To continue along the biosynthetic pathway, prenylated indole **3** must undergo a Cope rearrangement[23-24, 26] catalyzed by FamC1 to achieve alkene intermediate (**8**). The stereochemistry of the C11 and C12 centers would be set during this [3,3]-sigmatropic rearrangement, with the 12-*epi*-hapalindoles being directed through a chair-like transition state by FamC1, and the normal hapalindoles likely

the result of another cyclase. FamC1 subsequently catalyzes the formation of the tetracyclic hapalindole core (**6**), setting the stereochemistry of the C10 and C15 positions in either a concerted (**8** to **6**) or stepwise manner (**8** to **9** to **6**).

The modular mechanism where rearomatization first generates tricyclic hapalindoles (**9**) from **8** and electrophilic aromatic substitution forms **6** as a separate step is particularly intriguing as it retains the possibility for **9** to reenter the catalytic cycle and diverge to either hapalindoles (**6**) or fischerindoles (**10**) as controlled by the directing effect of pathway-specific cyclase enzymes. Further study is necessary to determine if **9** is indeed an active intermediate on the pathway to **6** or **10**, as was previously proposed[14] and effectively utilized in total syntheses of the tetracyclic hapalindole and ambiguine natural products, or a shunt metabolite that is unable to undergo further cyclization.

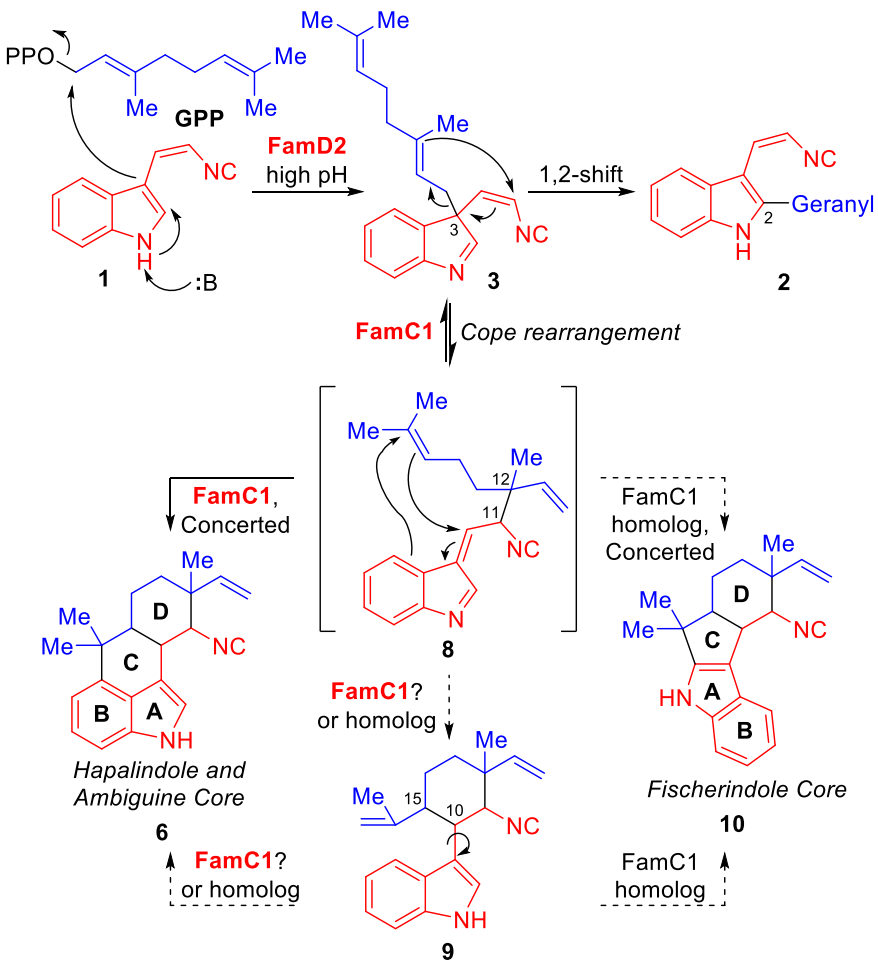


Figure 2-8. Proposed mechanism for the origin of hapalindole and fischerindole core ring systems.

2.4 Conclusion

We identified the functions of aromatic prenyltransferases FamD1 and FamD2, discovering previously unreported abilities to prenylate indole isonitrile **1** with either GPP or DMAPP to give C-2 prenylation (**2** and **4**) and C-3 prenylation (**3** and **5**). We also confirmed previous reports[10] that FamD1 prenylates tetracyclic hapalindoles with DMAPP into ambiguines while FamD2 does not. The hypothesis that **3** could act as precursor for tri- or tetracyclic hapalindoles led to our identification of the FamC1 cyclase from the cell-free lysate based on its ability to convert **3** into tetracyclic 12-*epi*-hapalindole U (**6**). Thus, we propose that the core ring formation is the combined result of a prenyltransferase (FamD2) generating intermediate **3** and a cyclase (FamC1) completing the ring

fusion to generate tetracyclic hapalindoles, which represents the first biocatalytic demonstration of such a process.

The studies described herein support a novel mechanism involving the union of indole isonitrile **1** and GPP to generate key biosynthetic intermediate **3**, and a subsequent Cope rearrangement and ring-cyclization cascade to generate the tetracyclic hapalindoles. It is especially significant that the current mechanistic hypothesis constitutes perhaps one of the only biosynthetic Cope rearrangements. The rarity of pericyclic reactions in biosynthetic transformations has been the subject of much recent debate,[27] and the current mechanistic hypothesis is consistent with all of the available experimental data. The underlying molecular basis for the biocatalytic conversion of **3** to **9** via the proposed [3,3]-sigmatropic rearrangement and subsequent 6-*exo*-trig ring closure/electrophilic aromatic substitution constitutes a provocative starting point for further studies in this intriguing arena. Moreover, the discovery of 12-*epi*-hapalindole U (**6**) indicates a larger metabolite scope produced by UTEX 1903 than previously recognized and suggests that our work characterizing the other *fam*C genes in this biosynthetic pathway will facilitate access to the previously isolated hapalindole natural products and additional structural diversity within this class of secondary metabolites.

2.5 Experimental section

2.5.1 Materials and Methods

All NMR spectra were acquired on a Varian 600 MHz spectrometer. Proton and carbon signals are reported in parts per million (δ) using residual solvent signals as an internal standard. Deuterated solvents ($CDCl_3$ and C_6D_6) were purchased as individual ampules of 99.6% minimum purity. The LCMS analysis was performed on a Shimadzu 2010 EV APCI spectrometer. High-resolution APCIMS spectra were obtained from an Agilent 6520 Q-TOF mass spectrometer equipped with an Agilent 1290 HPLC system at the University of Michigan core facility in the Department of Chemistry. Solvents were MS or HPLC grade (EMD or Fisher).

Escherichia coli strain DH5α (Invitrogen) was used for plasmid manipulation, BL21(DE3) was used for protein expression. Plasmid pET28a (Novagen) was used for cloning and expression. KOD Xtreme Hot Start DNA polymerase (EMD Millipore) was used for polymerase chain reactions. PureLink Quick Plasmid Miniprep Kit (Invitrogen) was used to prepare plasmid DNA. Primers were purchased from Integrated DNA Technologies. All cloned plasmids were sequenced by Sanger sequencing at the University of Michigan DNA Sequencing Core. For protein expression and purification, isopropyl *β*-D-thiogalactopyranoside (IPTG) was used to induce expression; Benzonase® Nuclease (EMD 70746-3 10KUN; ca. 250 units/μL, ≥90%) and lysozyme (Fisher BioReagents™, Egg White (White Cryst. Powder)) were used for protein purification on Ni-NTA affinity column (Invitrogen).

2.5.2 Genomic DNA

<u>*Cyanobacteria culturing*</u>. The cyanobacteria strain *Fischerella ambigua* UTEX 1903 was obtained from UTEX Culture Collection of Algae at the University of Texas at Austin. The cells were cultured in 10 L glass bottles containing Blue-Green 11 (BG-11) medium.[28] The cultures were grown at room temperature with 16/8 h light/dark cycles illuminated with 100 μmol photons/M^2/s for 6-8 weeks. The cells were collected on nitrocellulose membrane, and then

used directly for genomic DNA isolation, cell-free lysate extraction, or stored at -80 ℃ for later use.

Genomic DNA isolation. Pelleted cells (8 g) from UTEX1903 were re-suspended in 20 mL of Buffer A (Tris 100 mM, pH 8, EDTA 50 mM, NaCl 100 mM) and subjected to three freeze-thawing cycles (alternating freezing in liquid nitrogen and thawing at 37 ℃ in a water bath) to damage the cell walls and render the cells more susceptible for enzymatic lysis. 1% Sarkosyl detergent was added to a final concentration of 0.1% and the cells were incubated for 45 min at 4 ℃ on nutator. The cells were centrifuged (8,000 x *g*, 4 °C, 10 min) resulting in a light blue supernatant over a blue pellet. The supernatant was aspirated away and the pellet was re-suspended in 20 mL of Buffer B (Tris 50 mM, pH 8, EDTA 5 mM, NaCl 50 mM). Centrifugation was repeated and the pellet was resuspended in Buffer B (15 mL). The mixture was sonicated for 30 s total time with 10 s pulses followed by a 50 s rest, and then centrifuged (8,000 x *g*, 4 °C, 10 min). The pellet was suspended in Buffer B (16 mL) with lysozyme (0.5 mg/mL) and incubated at 37 ℃ for 30 min. The sample was subsequently degraded by the addition of proteinase K (final conc. 50 µg/mL) and SDS (2% final concentration), and incubated on nutator overnight at 4 ℃. The sample was centrifuged (14,000 x *g*, 4 °C, 10 min) and the supernatant was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, Fisher Biotech). It was shaken vigorously until an emulsion formed and the tube was centrifuged (14,000 x *g*, 22 °C, 10 min). The aqueous phase was transferred to a fresh tube and repeated with the phenol/chloroform/isoamyl alcohol extraction until no protein was visible at the interface of the phases. The resulting aqueous phase was washed with one volume of chloroform and centrifuged (14,000 x *g*, 22 °C, 1 min). The washed aqueous phase was mixed with 2/3 volume of 5 M NaCl, and then one volume of isopropanol was added, resulting in visible clumps of DNA. The DNA clumps were collected in an Eppendorf tube, resuspended in 1 mL of 70% ethanol and then pelleted by centrifugation (14,000 x *g*, 22 °C, 1 min). The supernatant was aspirated away and the DNA pellet was air-dried. The DNA was finally re-dissolved in 200 µL of ddH_2O.

Genome sequencing and assembly. The genome of UTEX 1903 was 108 Cycle Paired-End sequenced using the Illumina Genome Analyzer IIx at University of Michigan DNA Sequencing Core. De novo assembly of genomes was performed using Velvet version 1.2.10.

2.5.3 Protein preparation

Gene cloning. FamD1 and FamD2 genes were cloned from UTEX 1903 genomic DNA, and ligated into vector pET28a. The construct was transformed into chemically competent DH5α *E. coli* strain and incubated with LB agar plate with 50 μg/mL Kanamycin. Colonies were picked and inoculated overnight in LB media to isolate plasmid for DNA sequencing. **SI Table 2-1**

Protein expression. Positive plasmids from gene cloning were transformed into electro-competent BL21(DE3) *E. coli* strain. A single colony was picked and inoculated overnight in LB medium (10 mL, 50 μg/mL kanamycin) at 37 ℃, 200 rpm. This overnight culture was added to a 2.8 L Fernbach flask containing pre-warmed LB medium (1 L, 50 μg/mL kanamycin). The cells were incubated (37 ℃, 200 rpm) to an optical density of 0.6, then cooled down to 16 ℃. IPTG was added to a final concentration of 0.2 mM to induce protein expression. After incubation (16 °C, 200 rpm, 16 h), the cells were pelleted (6,000 x *g*, 4 °C, 15 min).

Protein purification. The following procedures were performed at 4 ℃ in a temperature controlled room. 10 g of cell pellets were re-suspended in buffer A (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol), containing 1 mM of PMSF, 0.5 mg/mL of lysozyme and 1 μL of benzonase. The mixture was stirred for 30 min and sonicated on ice for 100 s total time using 10 s pulses followed by a 50 s pause. The sample was centrifuged (60,000 x g, 4 °C, 35 min) to remove the cellular debris and imidazole was added to the supernatant to a final concentration of 10 mM. This clarified lysate was loaded onto Ni-NTA agarose prewashed with lysis buffer and gravity eluted. The column was then washed with buffer B (10 mM HEPES, 300 mM NaCl, 0.2 mM TCEP, 10% glycerol, 20 mM imidazole) to remove unbound proteins, followed with buffer C (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol, 300 mM imidazole) for target

protein. The fractions with the His-tag protein were combined and dialyzed using a PD10-desalting column (GE Healthcare) using buffer A. The purified protein was analyzed by SDS-PAGE gel for purity, measured by Nanodrop for concentration, and flash-frozen in liquid nitrogen for storage at -80 ℃. Rapid purification of FamD2 was essential for optimal activity and a loss of activity was observed with longer purifications.

2.5.4 Cell-free lysate extraction and fractionation

Extraction. Full protein extraction followed an adaption of the protocol by Ivleva and Golden.[22] A fresh UTEX 1903 cyanobacterial cell pellet (10 g) was rinsed with buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol; 20 mL x 2) to remove the BG-11 medium. The resulting pellet was resuspended in 15 mL of buffer containing 0.5 mg/mL of lysozyme, 1 mM PMSF, and 2 µL of Benzonase, stirred at 4 ℃ for 2 h, and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The mixture was centrifuged (60,000 x g, 4 °C, 35 min) to remove cellular debris. The supernatant was concentrated using a 10 kDa MWCO centrifugal tube and centrifuged (4,500 x g, 4 °C) until the desired concentration was reached. This sample was either flash-frozen for storage at -80 ℃, or used directly.

Fractionation. The concentrated lysate was boiled in water bath for 5 min and centrifuged (17,000 x g, 30 min) to pellet the denatured proteins. The supernatant was filtered through a 0.45 µm syringe filter and loaded onto an anion ion exchange column (Mono Q HR 16/10, GE Healthcare). Two buffers were used for fractionation: buffer A (20 mM Tris pH 8.2, 20 mM NaCl, 0.2 TCEP), buffer B (20 mM Tris pH 8.2, 1000 mM NaCl, 0.2 TCEP).

2.5.5 In vitro enzymatic assay

FamD1/D2 analytical assay. Indole isonitrile (**1**) was synthesized as described below. Authentic hapalindole U was provided by Professor Phil Baran. GPP and DMAPP were purchased from companies Isoprenoids and Sigma Aldrich. The general activity assays were conducted as a 50 µL-scale reaction containing 5 µM of FamD1 or FamD2, 1 mM of **1**, 1 mM of GPP or DMAPP, 5

mM of $MgCl_2$, and 50 mM of buffer (Citrate pH 6.0, Tris pH 7.8/8.5, Glycine pH 9.5/10.5). The reaction was incubated at 37 ℃ for 4 h, then quenched and extracted with 1x ethyl acetate thrice. The organic layers were combined, dried and re-dissolved in 100 µL acetonitrile for LCMS or HPLC analysis. The same condition was used to test the activity of FamD1 and FamD2 on hapalindole U by replacing **1** with hapalindole U. For the conversion assay between **2** and **3**, the reactions were conducted with 5 µM of FamD2 or boiled FamD2, **2** or **3** as substrate, 5 mM of $MgCl_2$, and different pH buffers (6/7.8/10.5) as shown in **SI Figure 2-1**.

Cell-free lysate assay. Similar condition was used to examine the activity of cell-free lysate except replacing the purified protein with 1 mg/mL cell-free lysate. In the conversion assay between **2**/**3** and **6**/**7**, substrates **2** and **3** were isolated from the above reactions with FamD2 (different buffer, pH 6.0 and 10.5, respectively), dissolved in DMSO and incubated with cell-free lysate. The boiled lysate was prepared by boiling the cell-free lysate for 5 min, centrifuging (17,000 x g, rt, 10 min) to remove precipitated proteins, and the resulting supernatant was used in the reaction.

Scale-up for structure characterization. For the structure analysis of enzymatic products (**2**/**3**/**4**/**5**/**ambiguine H**), the reactions were scaled up to 5 mL and incubated under same conditions. The extracted products were purified by HPLC (Column: XBridge^{TM} Prep C18 5 µM, 10 x 250 mm), using a 50-100% gradient of acetonitrile in water over 28 min. The purified compounds were concentrated, dissolved in C_6D_6 or $CDCl_3$, and analyzed using a Varian 600 MHz NMR.

2.5.6 Chemical synthesis of (Z)-2-(3-Indolyl)vinyl isocyanide (1)

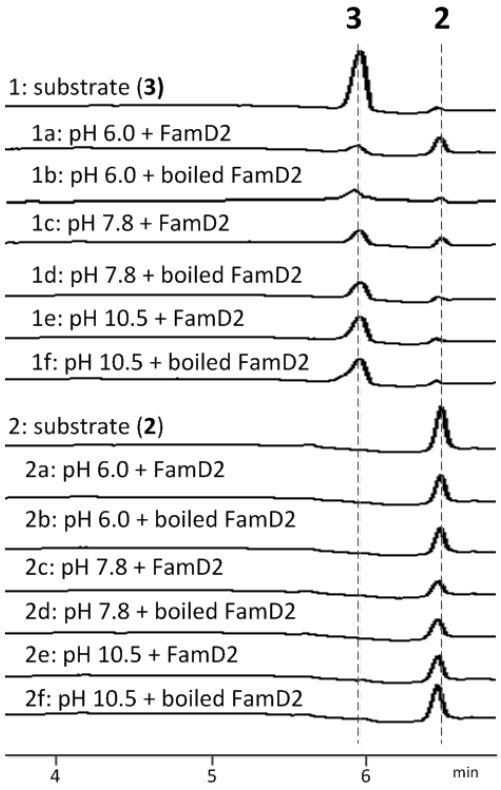
To a stirring solution of LiHMDS (2.12 g, 12.7 mmol) in THF (17 mL) at -78 °C (dry ice/acetone) was added diethyl (isocyanomethyl) phosphonate (1.00 mL, 6.24 mmol). The mixture was stirred at -78 °C for 0.75 h and then treated dropwise with indole-3-carboxaldehyde (0.820 g, 5.65 mmol) dissolved in THF (30 mL + 5 mL rinse). The resulting mixture was wrapped in foil and stirred at 2-3

°C (cryocool) for 22 h. The resulting blue solution was quenched by the addition of AcOH (0.37 mL, 6.5 mmol) in THF (5 mL) and concentrated. The dark red residue was diluted with EtOAc (40 mL), sonicated for 10 min, washed with 1 M aqueous potassium phosphate buffer (40 mL, pH 7), washed with water (40 mL), dried with $MgSO_4$, and concentrated to a dark red oil. The residue was dissolved in Et_2O and purified by flash chromatography (40% Et_2O/pentane, SiO_2) to give **1** as an off-white crystalline solid (0.024 g, 2.5%) that became faintly red upon standing. Spectral data were in accord with those reported.[29,30] ^{1}H NMR (599 MHz; $CDCl_3$): δ 8.59 (1H, bs), 8.14 (1H, d, *J* = 4.3 Hz), 7.68 (1H, dd, *J* = 12.0, 1.7 Hz), 7.44 (1H, dt, *J* = 12.0, 1.7 Hz), 7.29 (1H, td, *J* = 11.3, 1.9 Hz), 7.23 (1H, td, *J* = 11.2, 1.7 Hz), 6.80 (1H, m), 5.75 (1H, d, *J* = 13.3 Hz). ^{13}C NMR (151 MHz; $CDCl_3$): δ 168.8, 135.2, 127.0, 126.6, 124.3, 123.4, 121.1, 118.2, 111.7, 110.3, 104.6 (t).

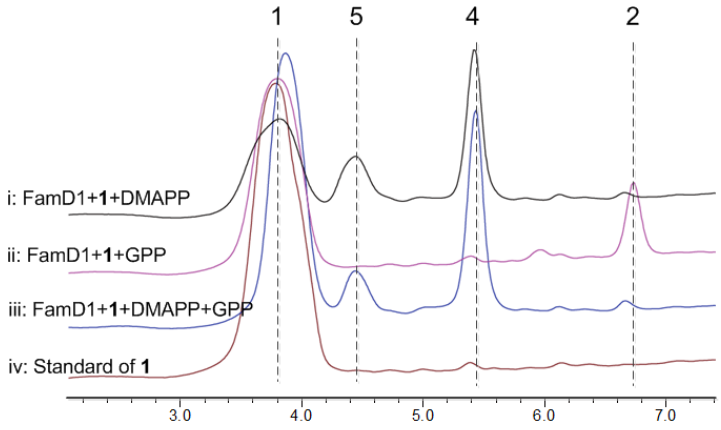
2.6 Supplementary information

SI Table 2-1. Primers used for cloning.

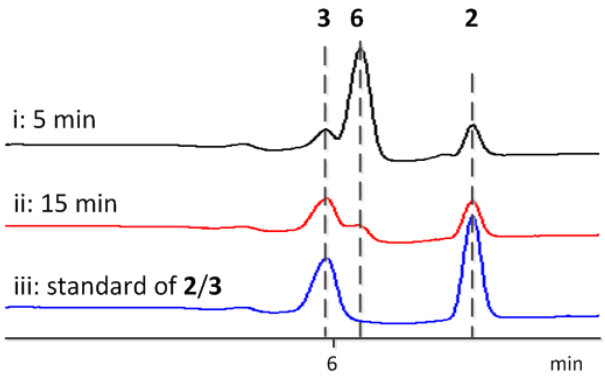
Gene	Forward primer	Reverse primer
famD1	AAAAAAGCTAGCATGACCATTGTTAACCGTATACG	AAAAAACTCGAGTTATGACATGACAATACTATTAAGTTTAT
famD2	AAAAAACATATGAACGATGTTAACCGTATAC	AAAAAAAAGCTTAAGCAAGAGCAAAATCAGCAG



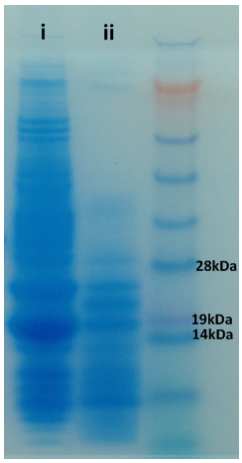
SI Figure 2-1. HPLC traces of the conversion assay between **2** and **3** catalyzed by FamD2. 1) Conversion assay with **3** as substrate reacting with FamD2: **3** can convert to **2** at pH 6.0 and pH 7.8 in the presence of FamD2, but no conversion is observed at pH 10.5. Trace amount of **2** results from the initial isolation of **3** and are present in all samples. 2) Conversion assay with substrate **2** and FamD2: **2** does not convert to **3** at any pH.



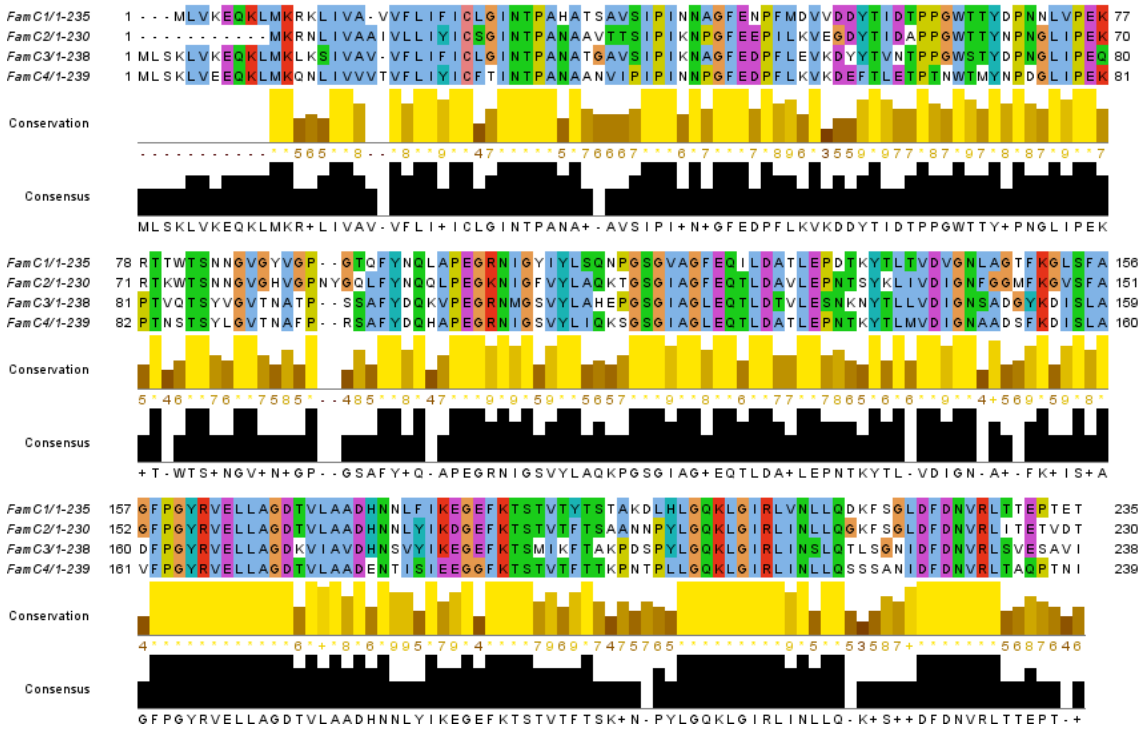
SI Figure 2-2. FamD1 shows a preference for DMAPP as substrate when incubated with **1**:DMAPP:GPP (1:1:1), trace iii.



SI Figure 2-3. The cell-free lysate was boiled for different time and incubated with a mixture of **2** and **3**. i) cell-free lysate was boiled for 5 min; ii) cell-free lysate was boiled for 15 min; iii) standard of starting material used in the reaction.

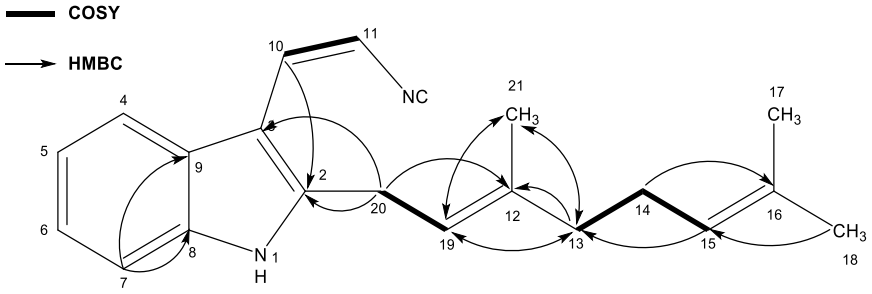


SI Figure 2-4. SDS-PAGE gel image of i) cell-free lysate and ii) supernatant of cell-free lysate after boiling for 5 min.



SI Figure 2-5. Sequence alignment of the three domains previously annotated as unknown function in *fam* gene cluster show strong sequence identities with the cyclase FamC1 and thus categorized into the cyclase class FamC1-4.

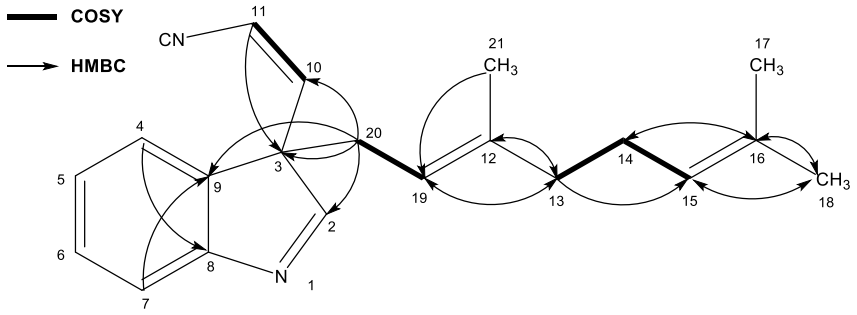
SI Table 2-2. Key correlations for **2** and complete NMR spectroscopic data table in $CDCl_3$ and C_6D_6. HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2008.



Position	δ_{13C} ($CDCl_3$)	δ_{1H}, multi (*J*) (C_6D_6)	COSY (C_6D_6)	HMBC ($CDCl_3$)
1		7.29, bs		
2	138.6			
3	106.2			
4	109.5	6.98, d (8.4)	5	6
5	119.6	7.18, t (7.8)		
6	119.7	7.23, t (7.2)	7	
7	118.1	7.74, d (7.8)		8,9
8	135.1			
9	122.3			
10	125.2	6.19, d (7.2)	11	2
11	105.7	5.21, d (8.4)	10	3, 10
12	140.2			
13	39.5	2.00, t (7.2)	14	12, 14, 19, 21
14	26.1	2.09, q (7.2)	13, 15	13, 15, 16
15	122.4	5.12, t (6.6)	14	17
16	132.3			
17	25.5	1.52, s		15, 16, 18
18	18.6	1.67, s		15, 16, 17
19	116.9	5.14, t (7.2)	20	13, 21
20	27.4	3.35, d (6.6)	19	2, 3, 12, 19
21	16.5	1.51, s		12, 13, 19

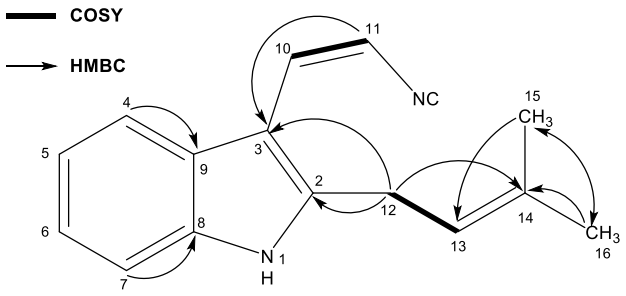
NMR splitting coefficients are reported a s = singlet, d = doublet, t = triplet, m = multiplet, b = broad

SI Table 2-3. Key correlations for **3** and complete NMR spectroscopic data table in C_6D_6. HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2004.



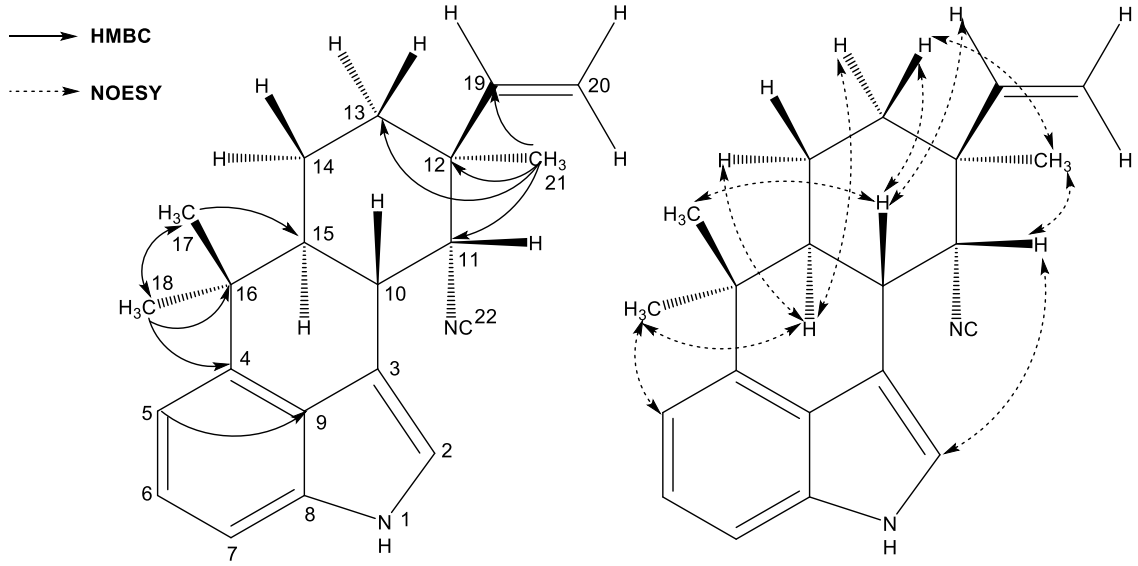
Position	δ_{13C}	δ_{1H}, multi (*J*)	COSY	HMBC
1				
2	171.6	8.2, s		
3	62.0			
4	121.9	7.05, dd, (7.3, 0.7)	5	6, 8
5	126.2	7.02, td, (7.4, 1.1)	4, 6	4, 9
6	128.7	7.10, td, (7.4, 1.4)	5, 7	7, 8
7	121.9	7.74, d, (7.7)	6	5, 9
8	156.3			
9	141.1			
10	131.3	4.93, d, (8.9)	11	
11	112.2	4.78, d, (8.9)	10	3
12	139.8			
13	39.8	1.87, t, (7.5)	14	12, 14, 15, 19, 21
14	26.6	1.98, q, (7.4)	13, 15	12, 13, 15, 16
15	124.2	5.06, t, (6.9)	14	17, 18
16	131.4			
17	17.6	1.52, s		15, 16, 18
18	25.7	1.68, s		15, 16, 17
19	117.5	4.88, t, (7.5)	20	
20	33.8	2.37, dd, (7.4, 14.2)	19	2, 3, 9, 10, 12, 19,
		2.33, dd, (7.6, 14.2)	19	
21	16.2	1.33, s		12, 13, 19

SI Table 2-4. Key correlations for **4** and complete NMR spectroscopic data table in C_6D_6. HRMS: Calcd for $C_{16}H_{16}N_2$ $[M+H]^+$ 237.1386, found 237.1385.



Position	δ_{13C}	δ_{1H}, multi (*J*)	COSY	HMBC
1				
2	138.3			
3	106.7			
4	110.8	6.90, d, (7.9)	5	6, 9
5	122.1	7.18, td, (7.7, 1.1)	4	7
6	120.5	7.23, td, (7.4, 1.1)	7	4
7	119.9	7.74, d, (7.9)	6	5, 8
8	135.6			
9	127.8			
10	126.1	6.18, d, (8.7)	11	
11	108.4	5.21, d, (8.9)	10	3
12	27.2	3.30, d, (7.1)	13	2, 3, 13, 14
13	119.7	5.06, tm, (7.0)	12	
14	135.3			
15	25.5	1.57, s		13, 14, 16
16	17.6	1.47, s		13, 14, 15

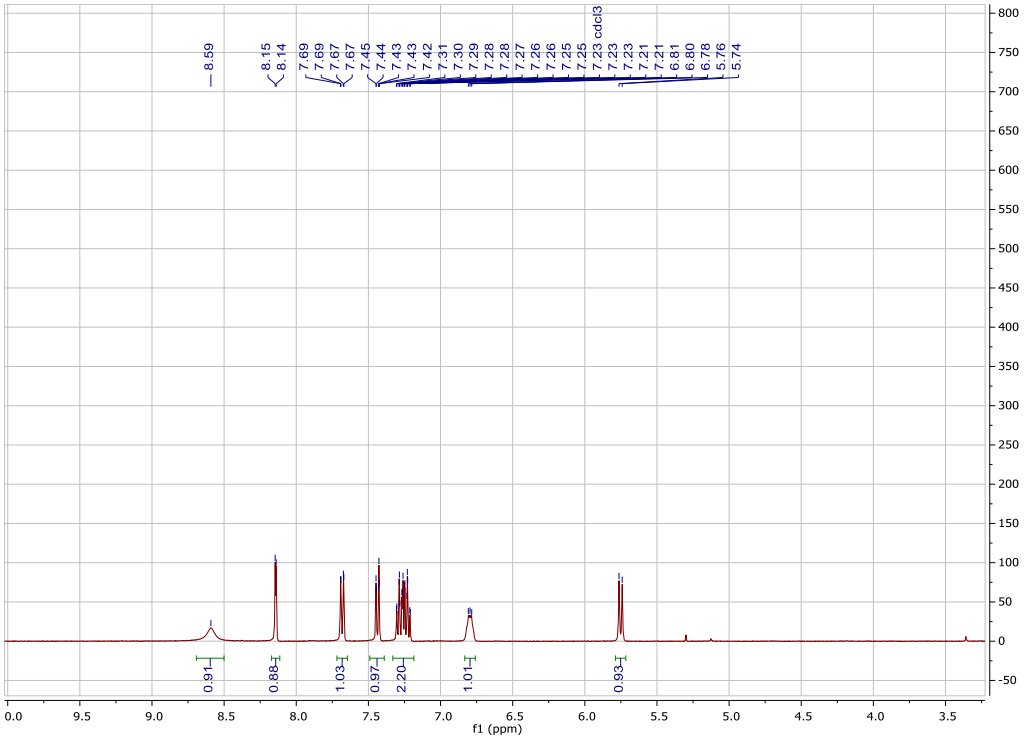
SI Table 2-5. Key correlations for 12-*epi*-Hapalindole U (**6**) and complete NMR spectroscopic data table in C_6D_6. HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2008.

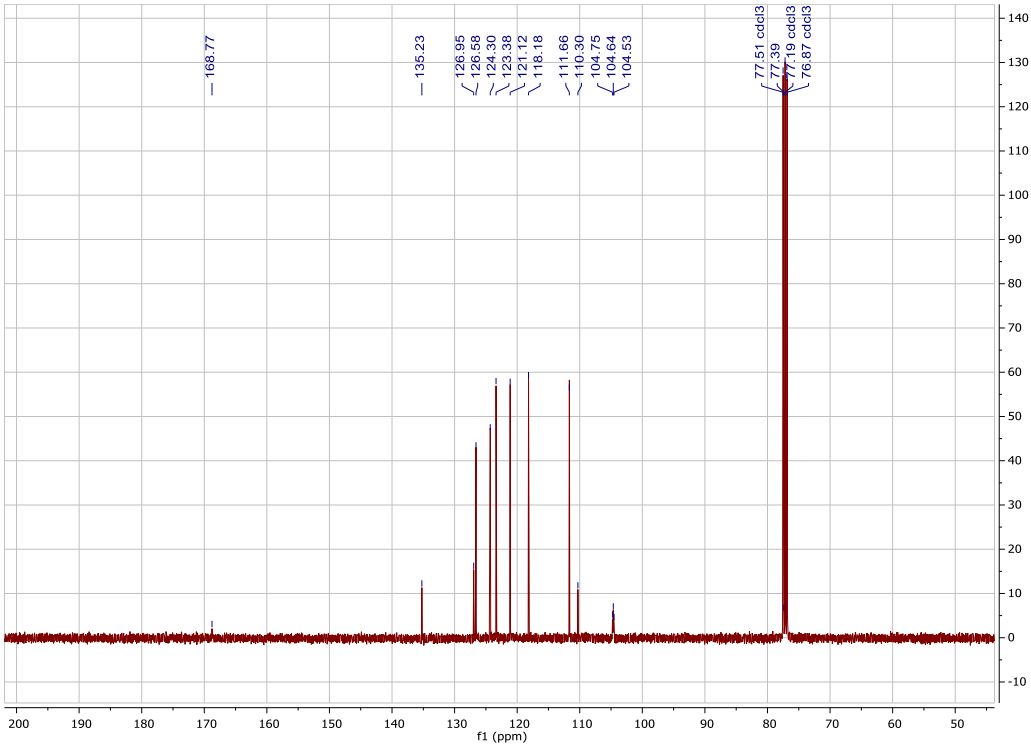


Position	δ_{13C}	δ_{1H}, multi (*J*)	COSY	HMBC	NOESY
1		6.67, bs			
2	116.4	6.38, t, (1.9)			11
3	113.1				
4	141.1				
5	113.4	7.08, d, 7.3)	6	7, 9	18
6	123.4	7.29, t, (7.7)	5, 7		
7	108.7	6.97, d, (8.2)	6		
8	134.6				
9	126.0				
10	34.7	3.03, d, (11.3)	11, 15		14ax, 17, 19
11	63.0	3.94, s	10		2, 21
12	39.8				
13	31.0	(H_{ax}) 1.69, td, (13.0, 3.9) (H_{eq}) 1.48, m	14		15, 21 20trans, 21
14	21.6	(H_{ax}) 1.37, td, (12.9, 3.8) (H_{eq}) 1.48, m	13, 15		10, 17, 19, 20trans
15	43.0	2.01, tc, (12.0, 3.5)	10, 14		13ax, 14eq, 18
16	37.2				
17	25.0	1.01, s		4, 15, 16, 18	10
18	24.2	1.32, s		4, 15, 16, 17	5
19	142.7	5.43, dd, (17.6, 11.0)			10, 13eq, 14ax, 11, 21
20	114.4	(cis) 4.93, dd, (11.0, 0.6) (trans) 4.88, dd, (17.6, 0.6)		12	13eq, 21
21	28.0	1.19, s		11, 12, 13, 19	
22	160.7				

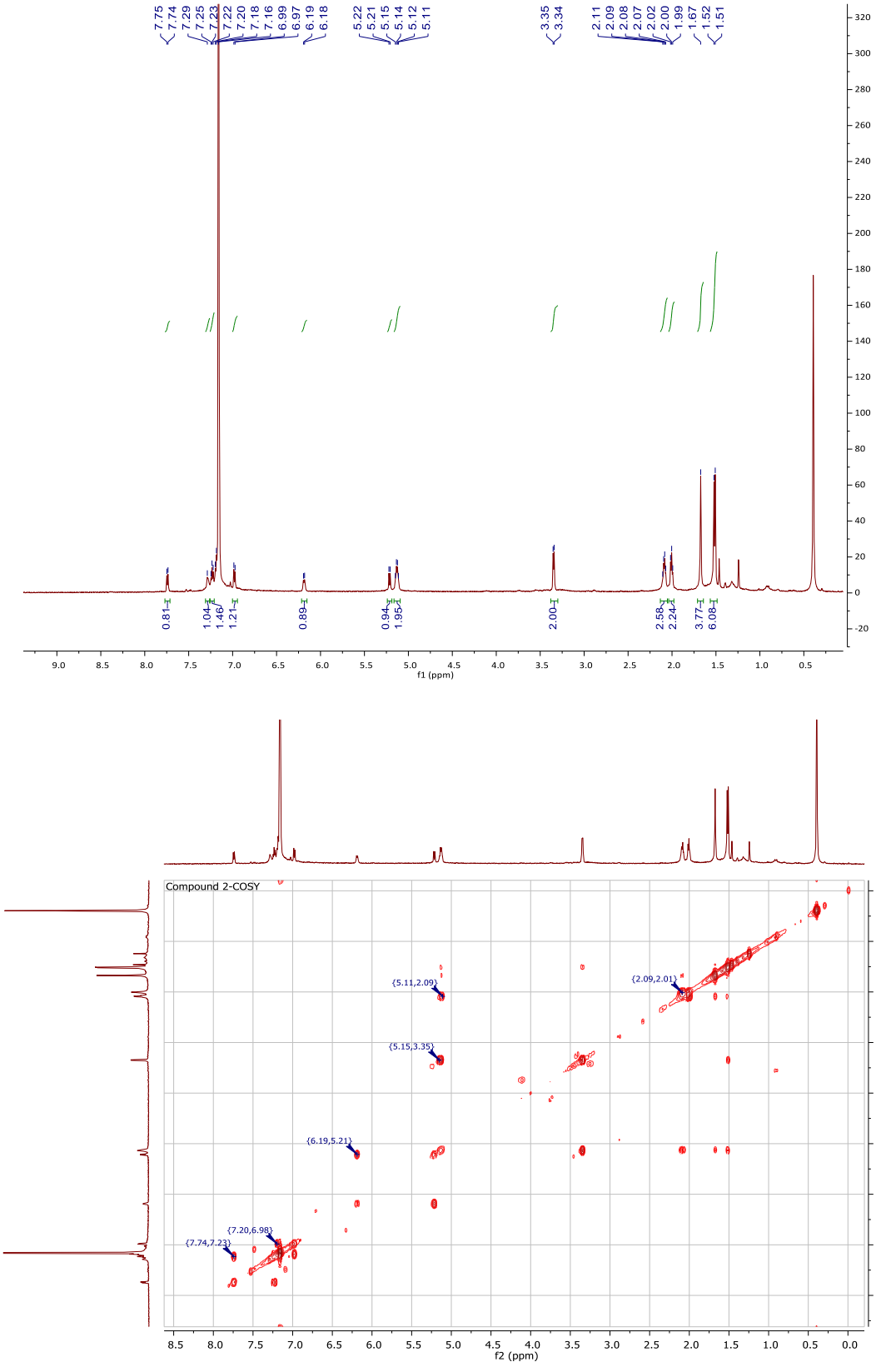
2.7 Spectra section

^{1}H and ^{13}C NMR spectra of indole isonitrile (**2**) in $CDCl_3$ at 600 MHz and 125 MHz, respectively.

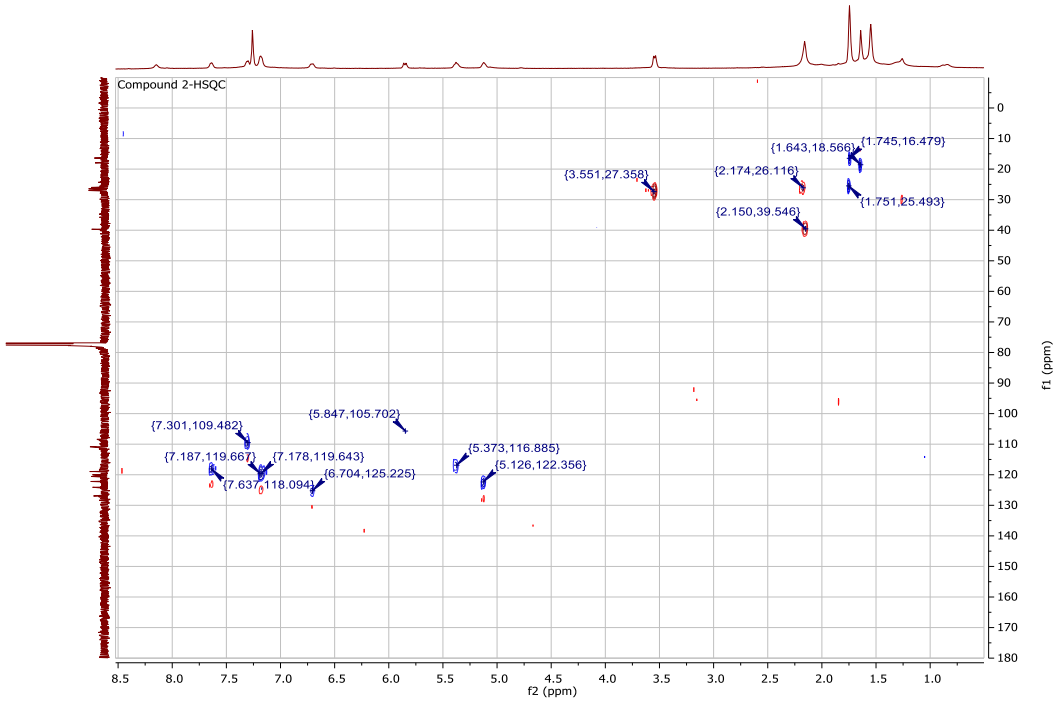


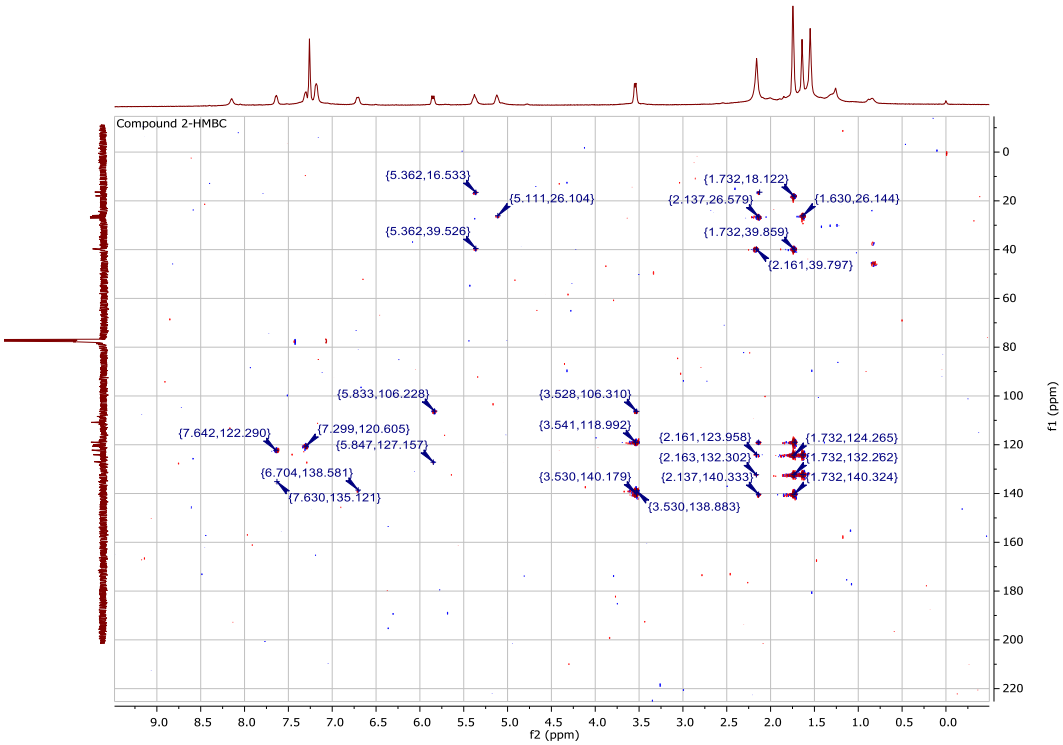


^{1}H and COSY NMR spectra of Compound **2** in C_6D_6 at 600 MHz

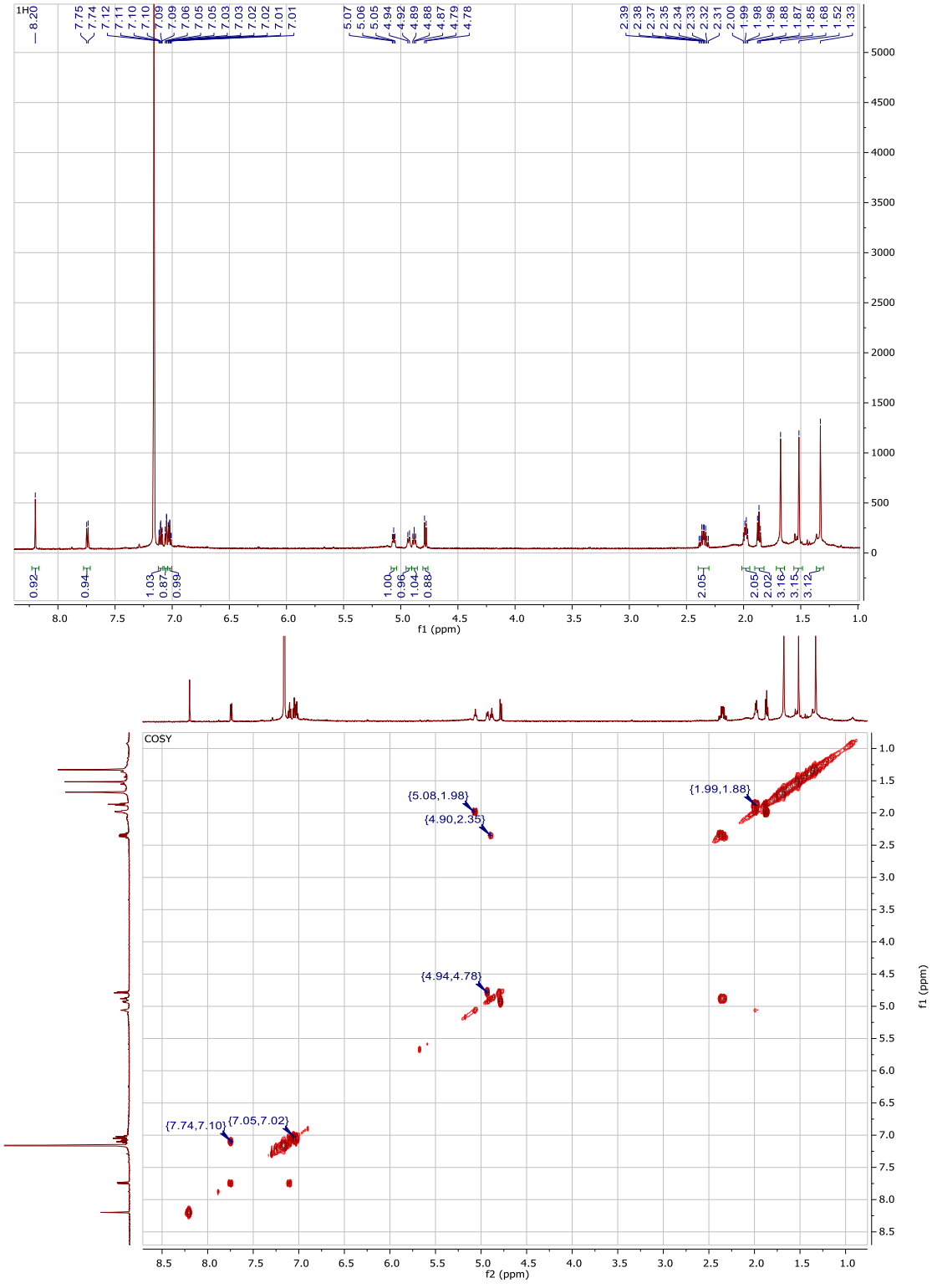


HSQC and HMBC NMR spectra of Compound **2** in $CDCl_3$ at 600 MHz

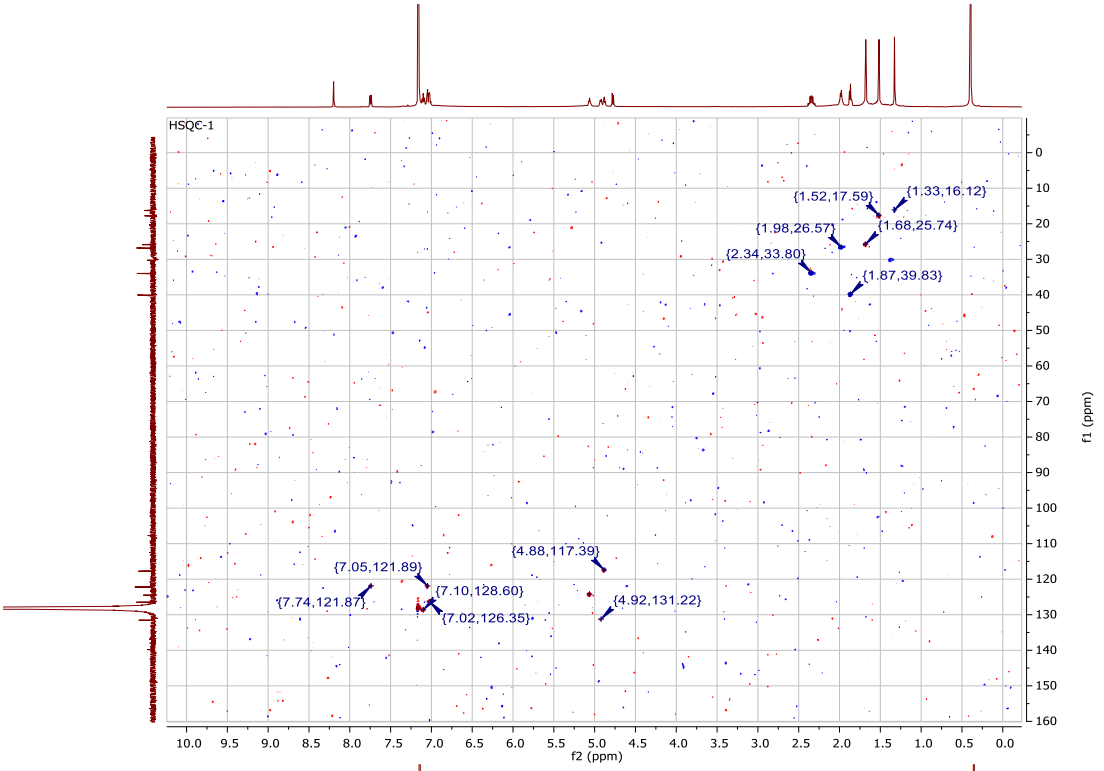


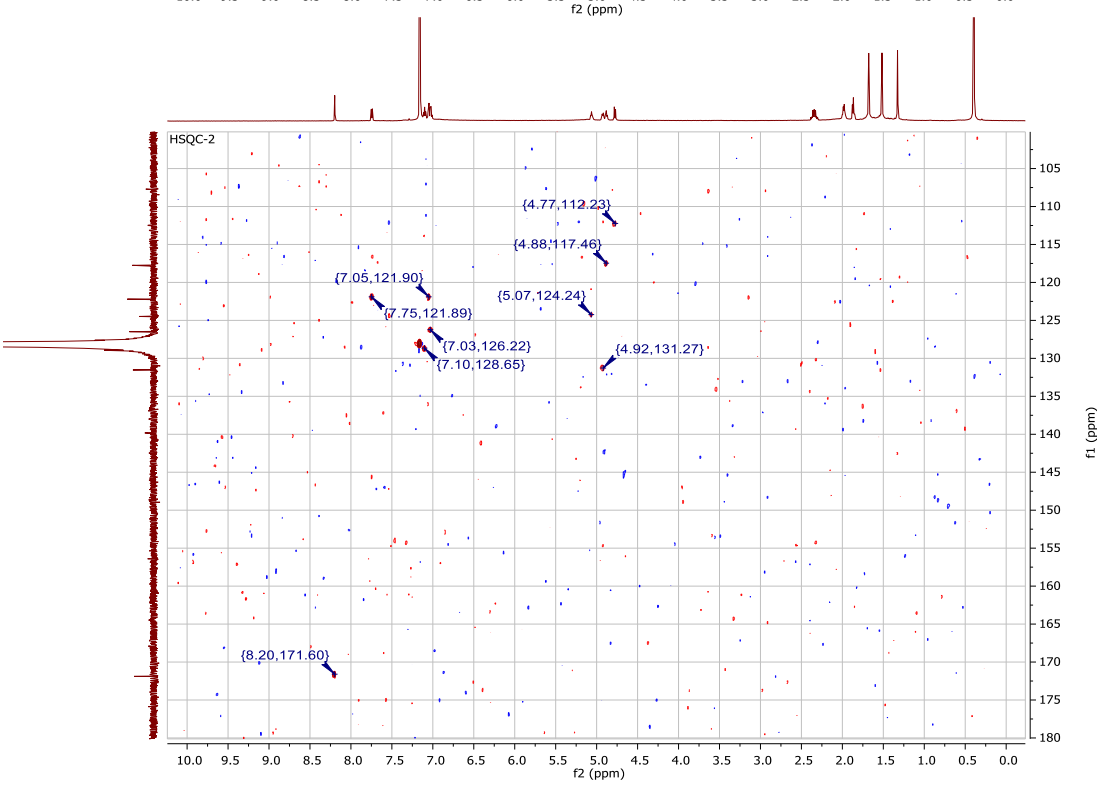


^{1}H and COSY NMR spectra of Compound **3** in C_6D_6 at 600 MHz

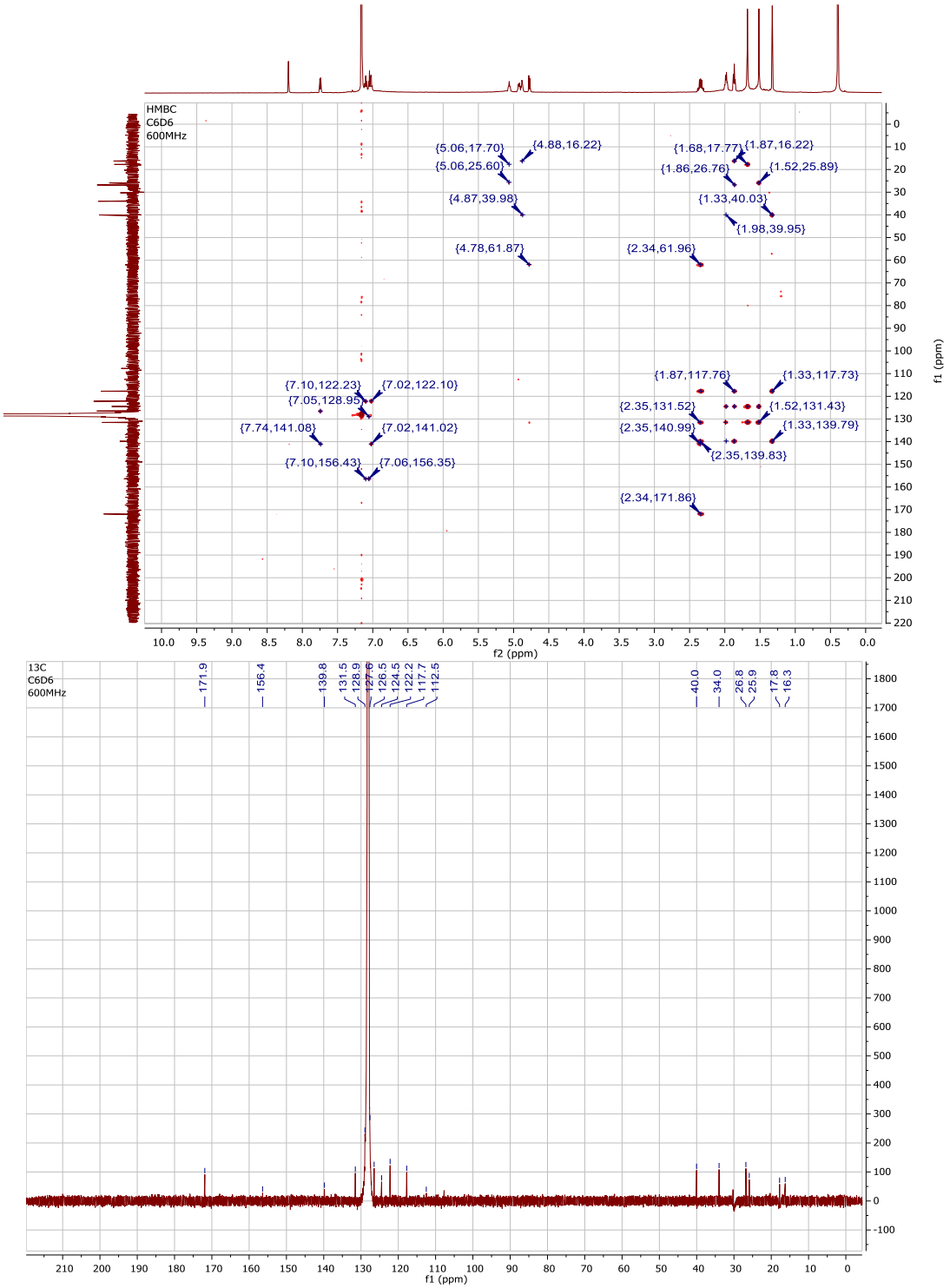


HSQC NMR spectra of Compound **3** in C_6D_6 at 600 MHz and 125 MHz, respectively.

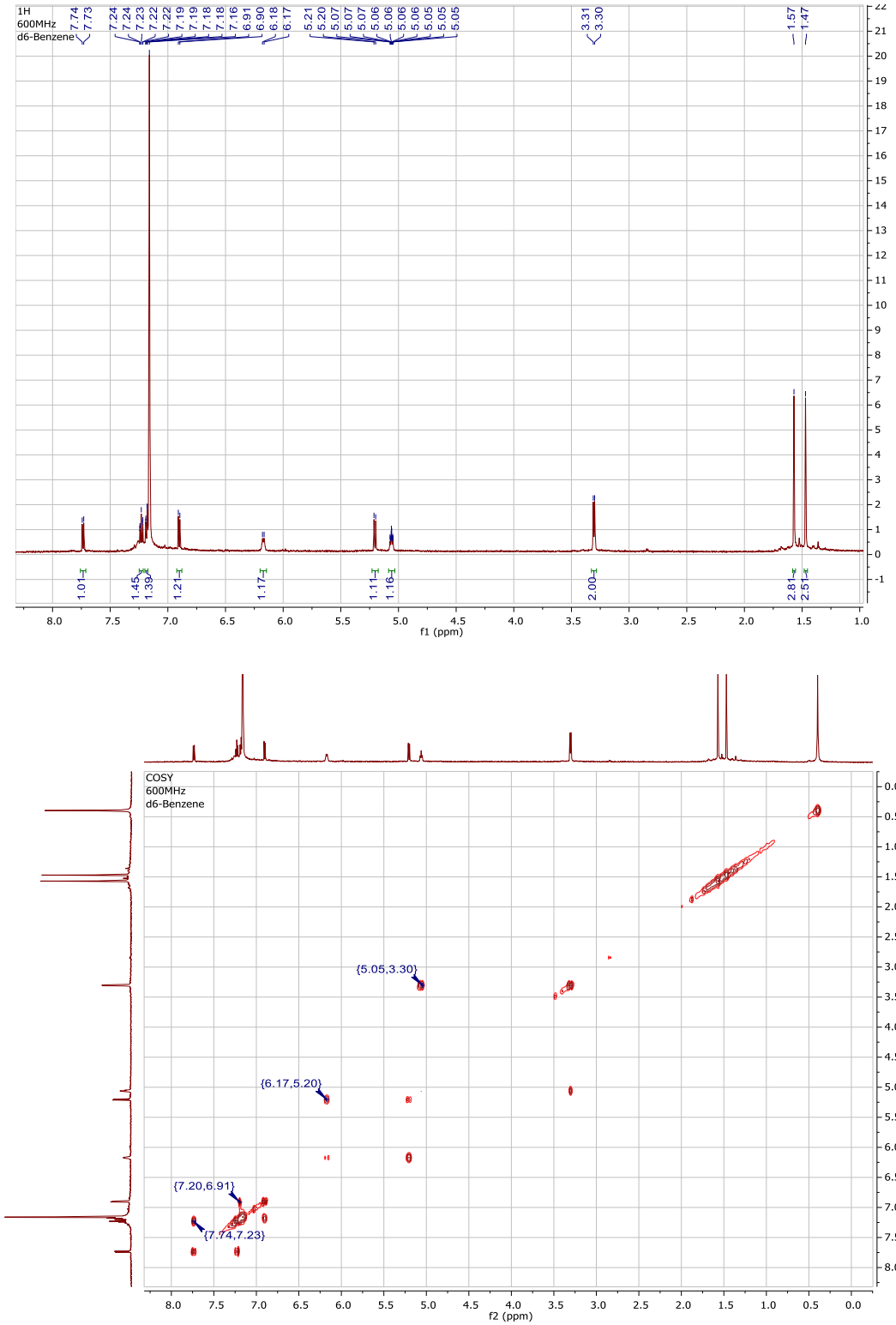




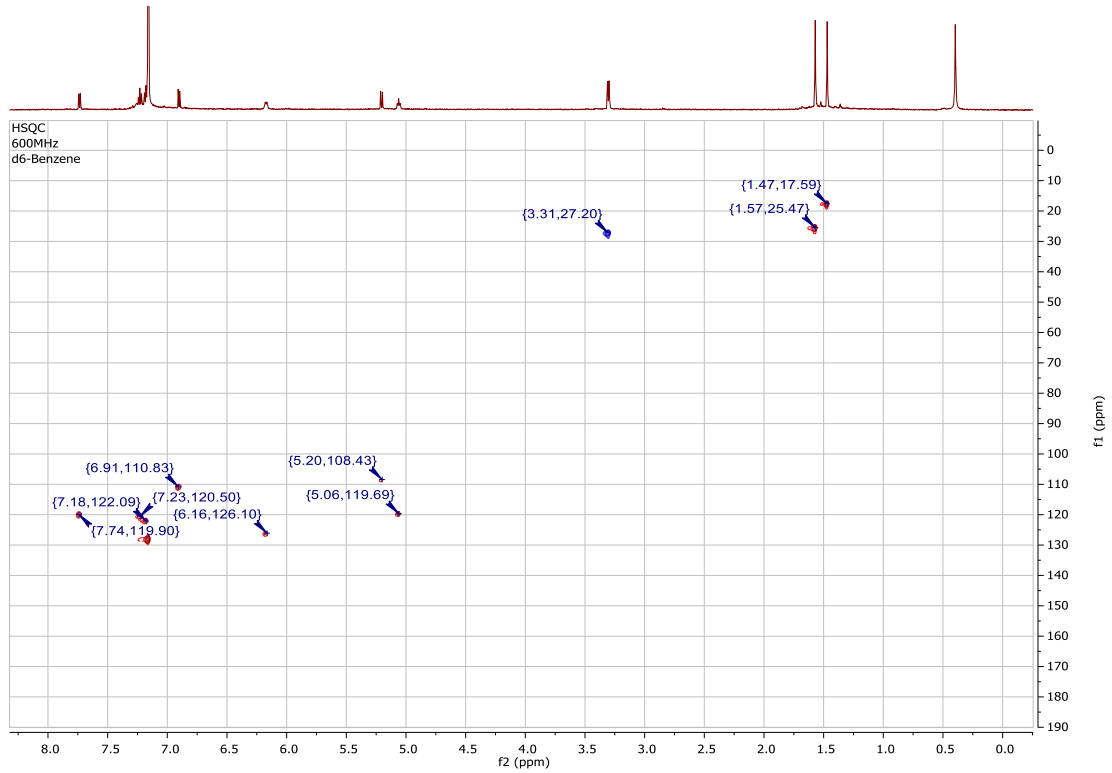
HMBC and ^{13}C NMR spectra of Compound **3** in C_6D_6 at 600 MHz and 150 MHz, respectively.

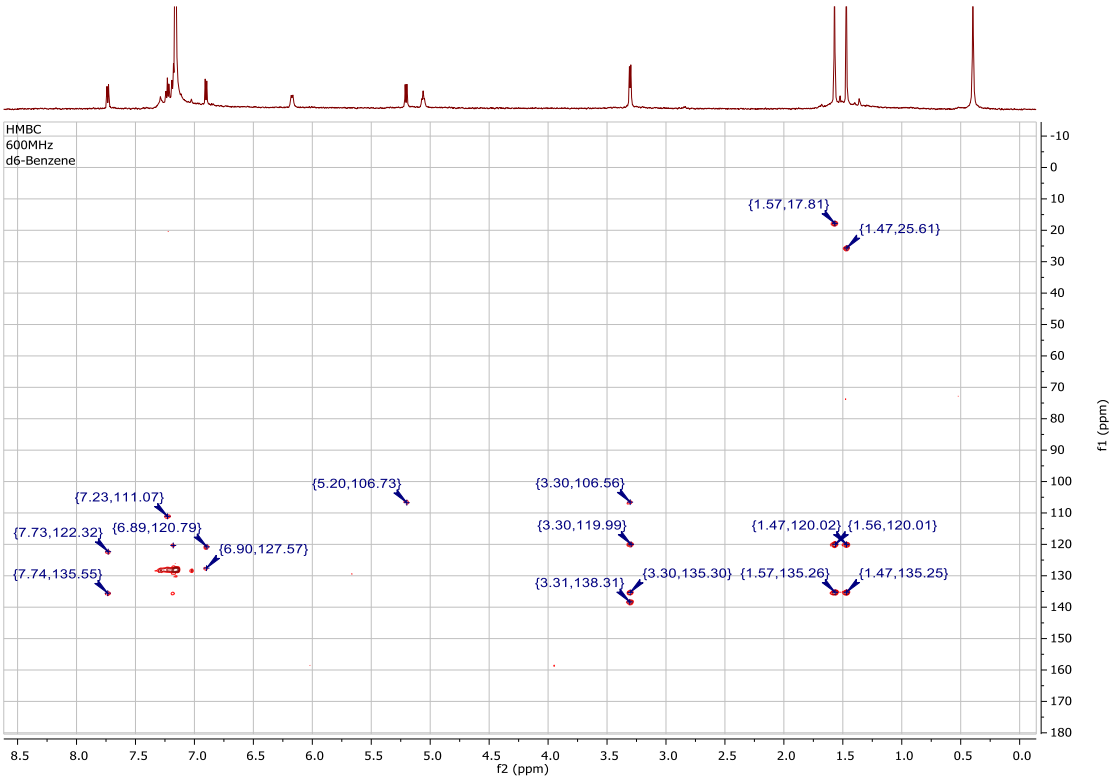


^{1}H and COSY NMR spectra of Compound **4** in C_6D_6 at 600 MHz

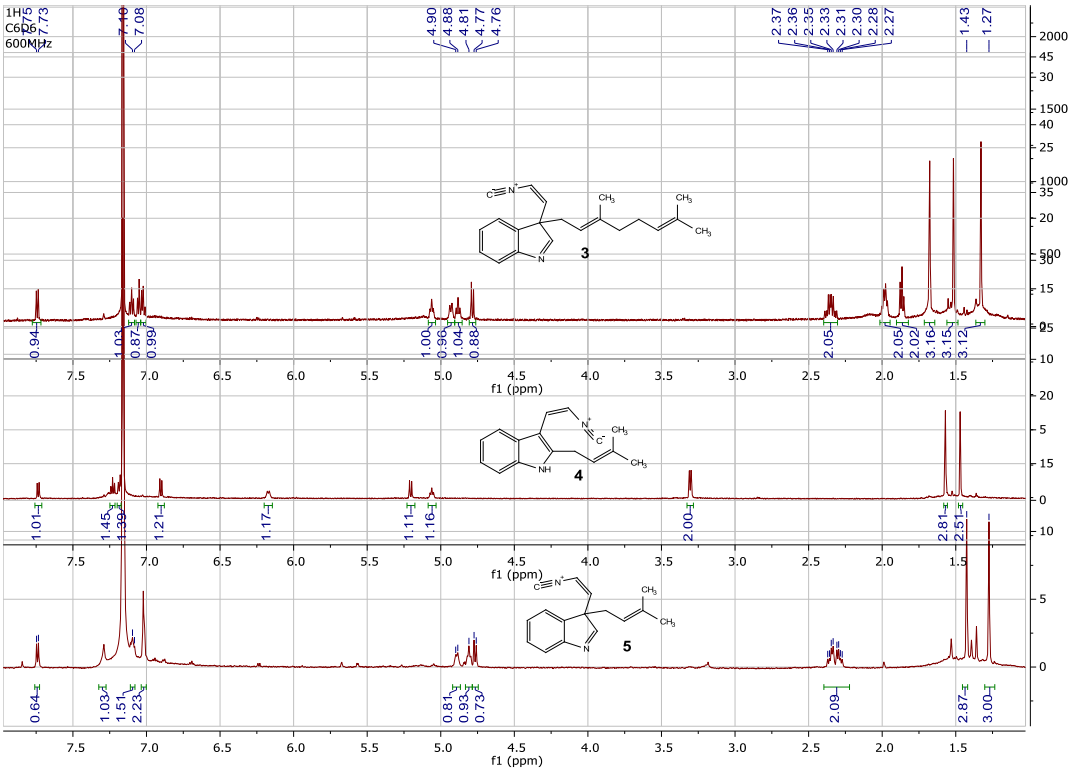


HSQC and HMBC NMR spectra of Compound **4** in C_6D_6 at 600 MHz

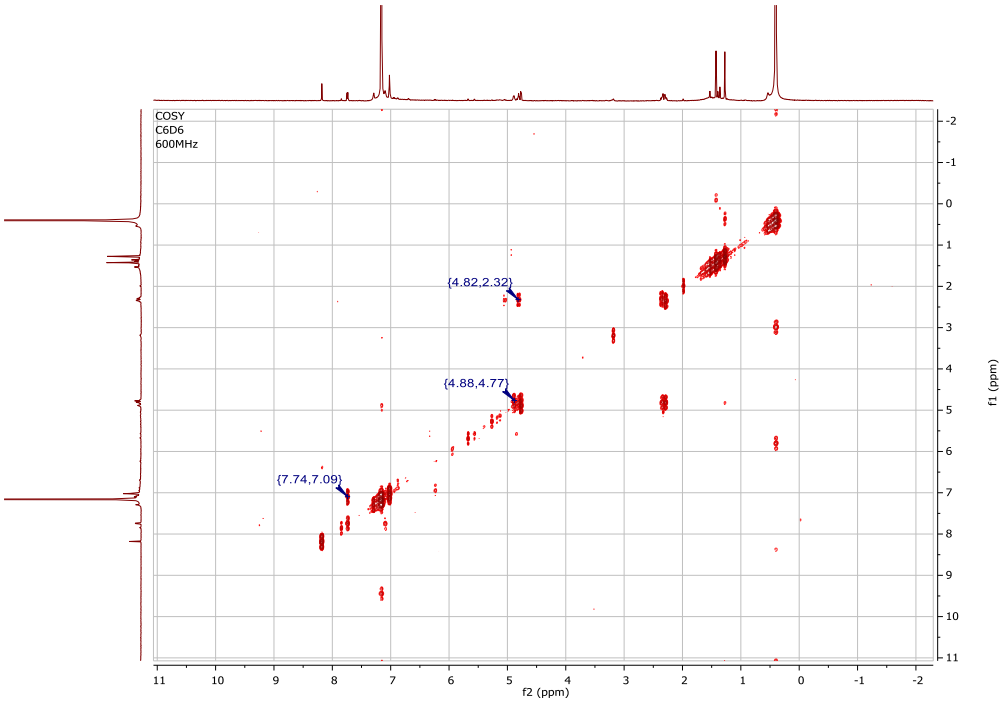




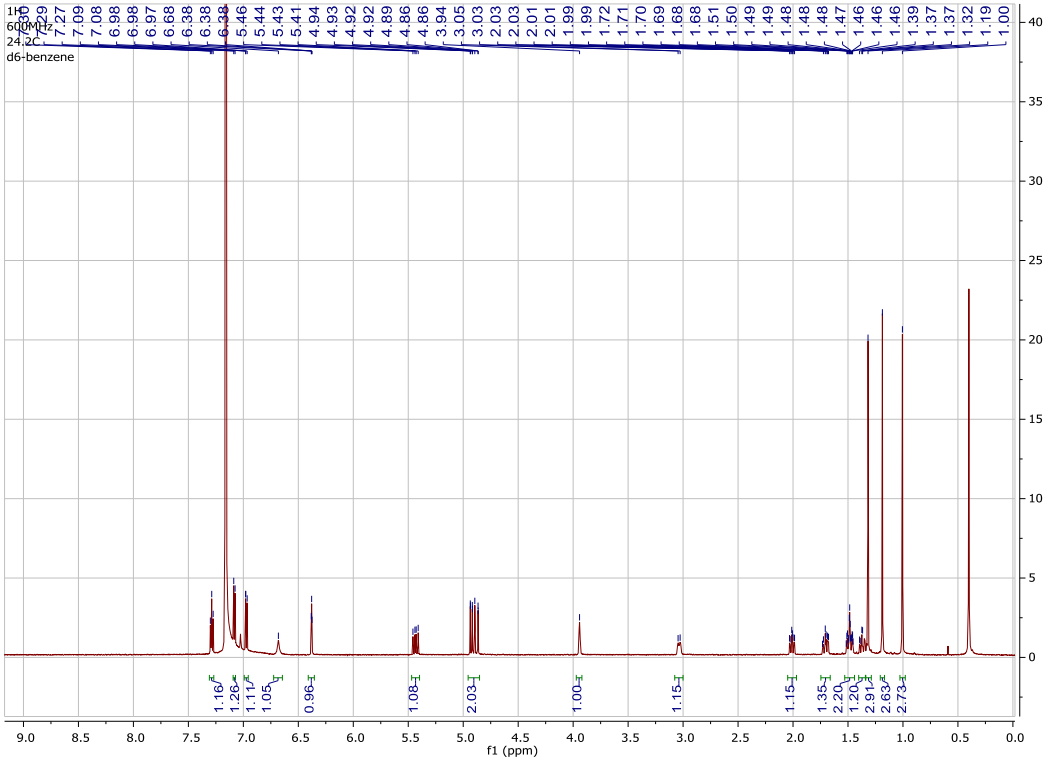
1H NMR spectrum of Compound 5 in C6D6 at 600 MHz and comparison with compounds 3 and 4. HRMS: Calcd for C16H16N2 [M+H]+ 237.1386, found 237.1381.

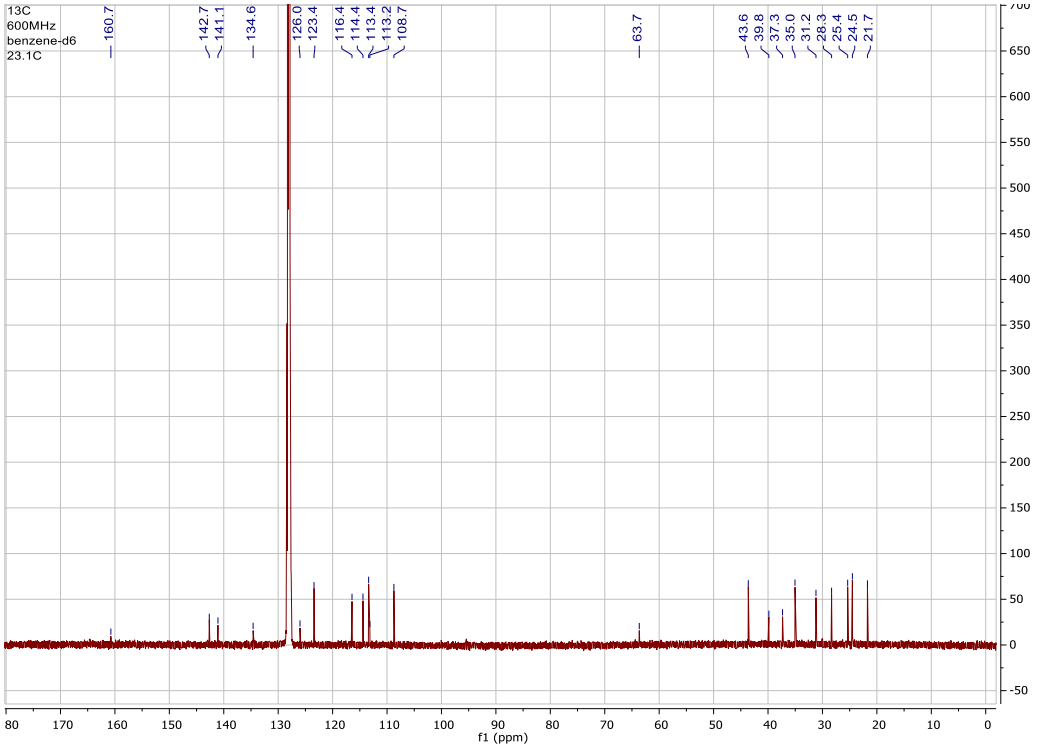


COSY NMR spectra of Compound **5** in C_6D_6 at 600 MHz

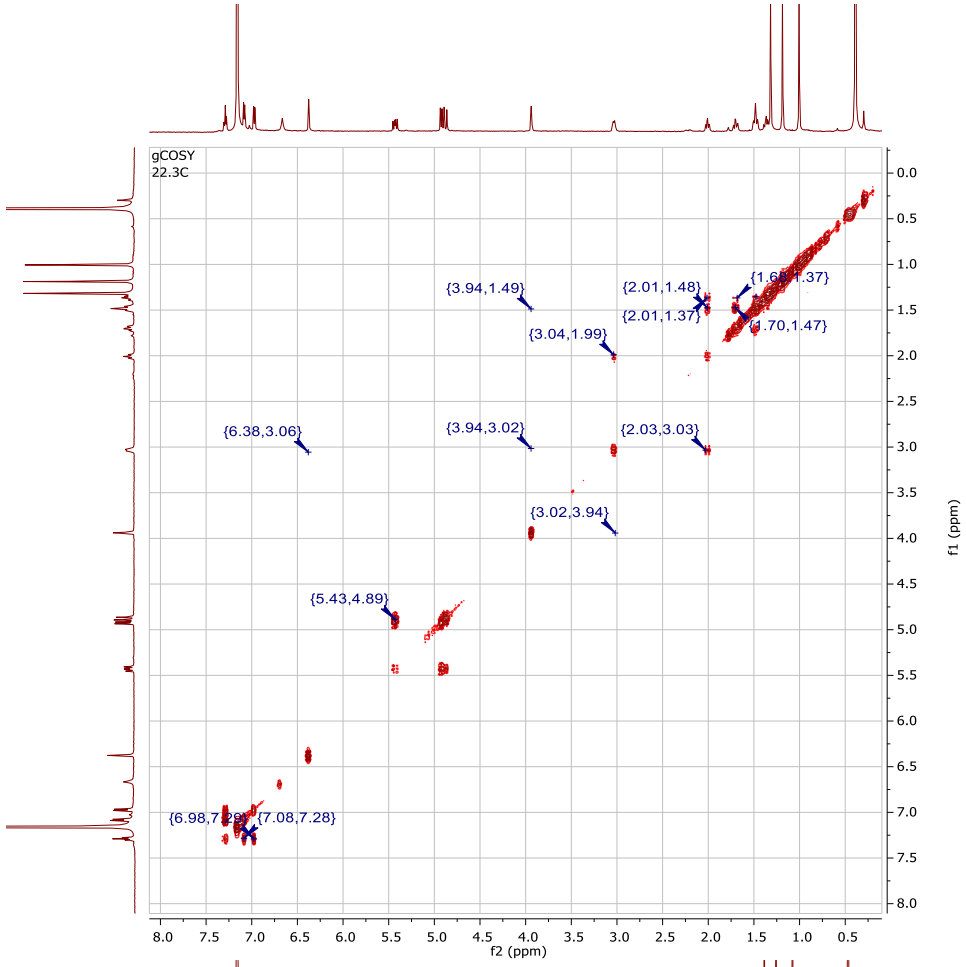


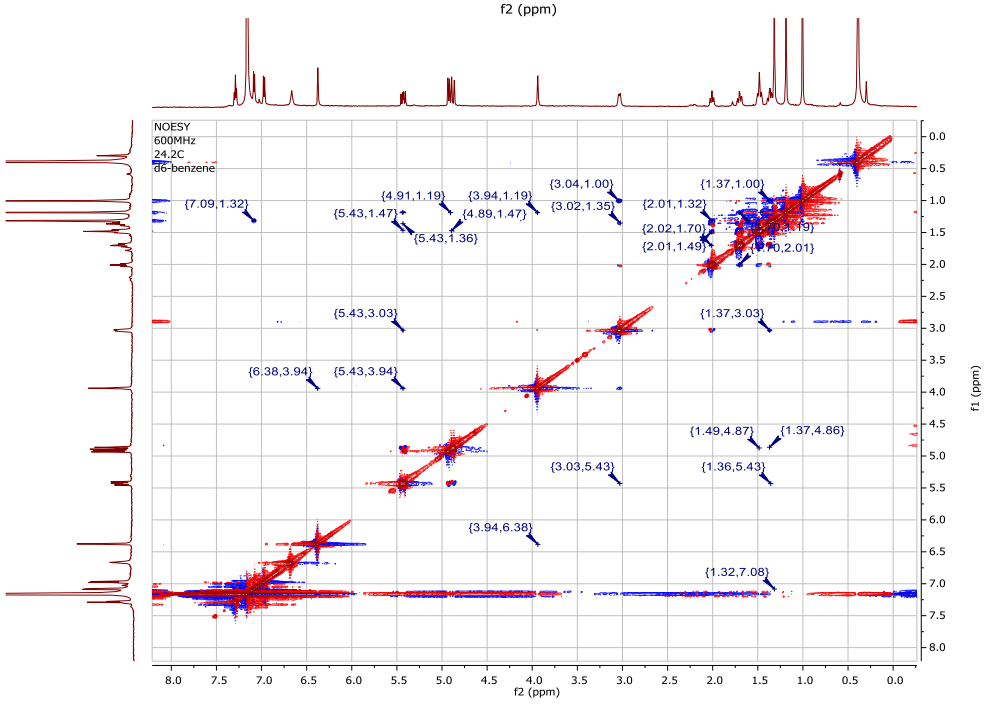
^{1}H and ^{13}C NMR spectra of Compound **6** in C_6D_6 at 600 MHz and 150 MHz, respectively.



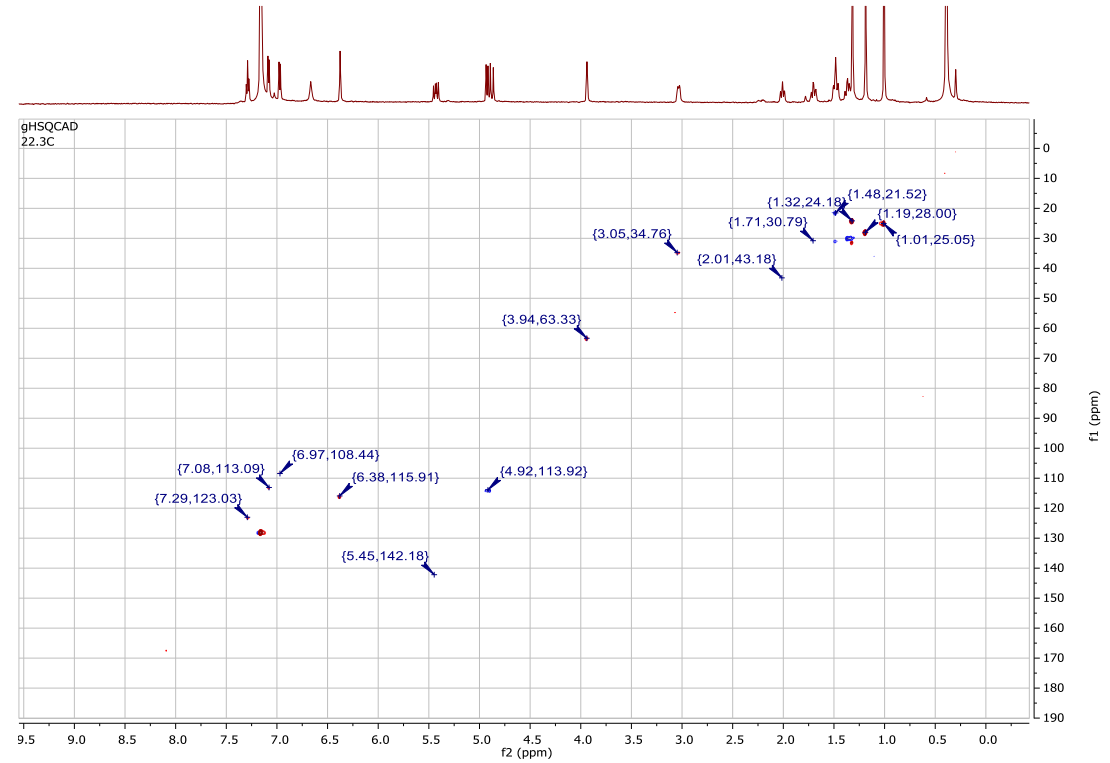


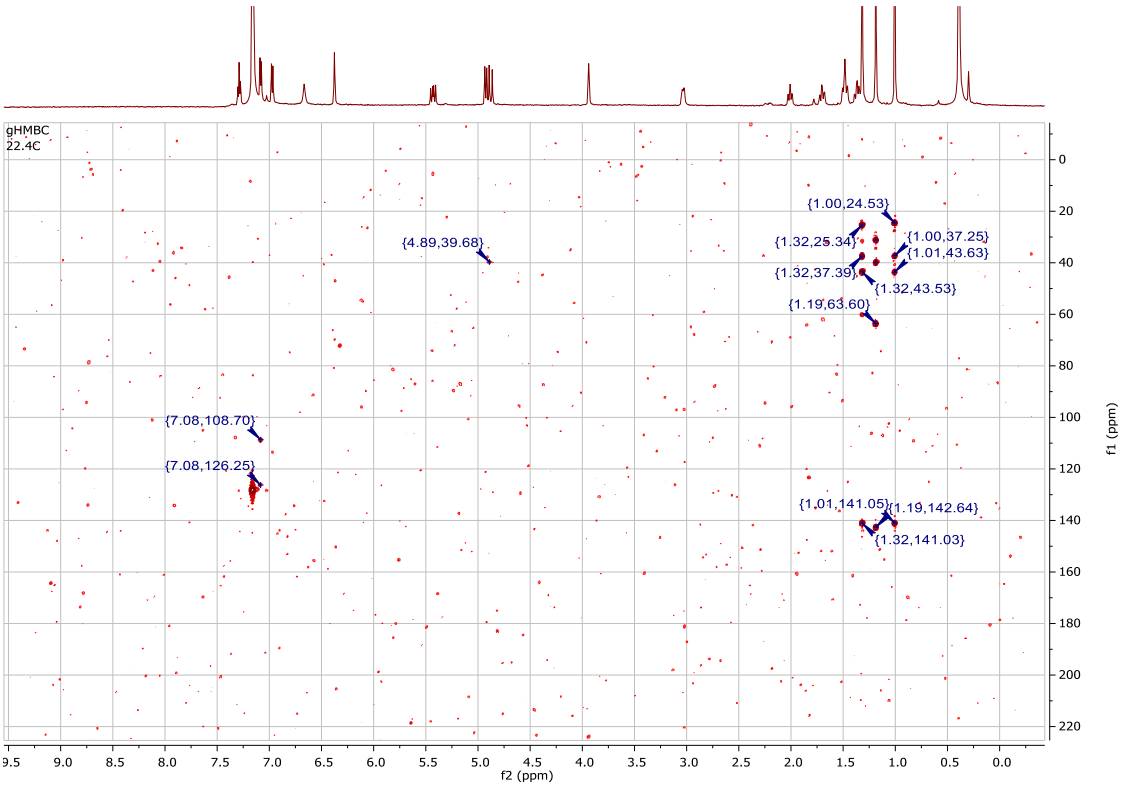
COSY and NOESY NMR spectra of Compound **6** in C_6D_6 at 600 MHz



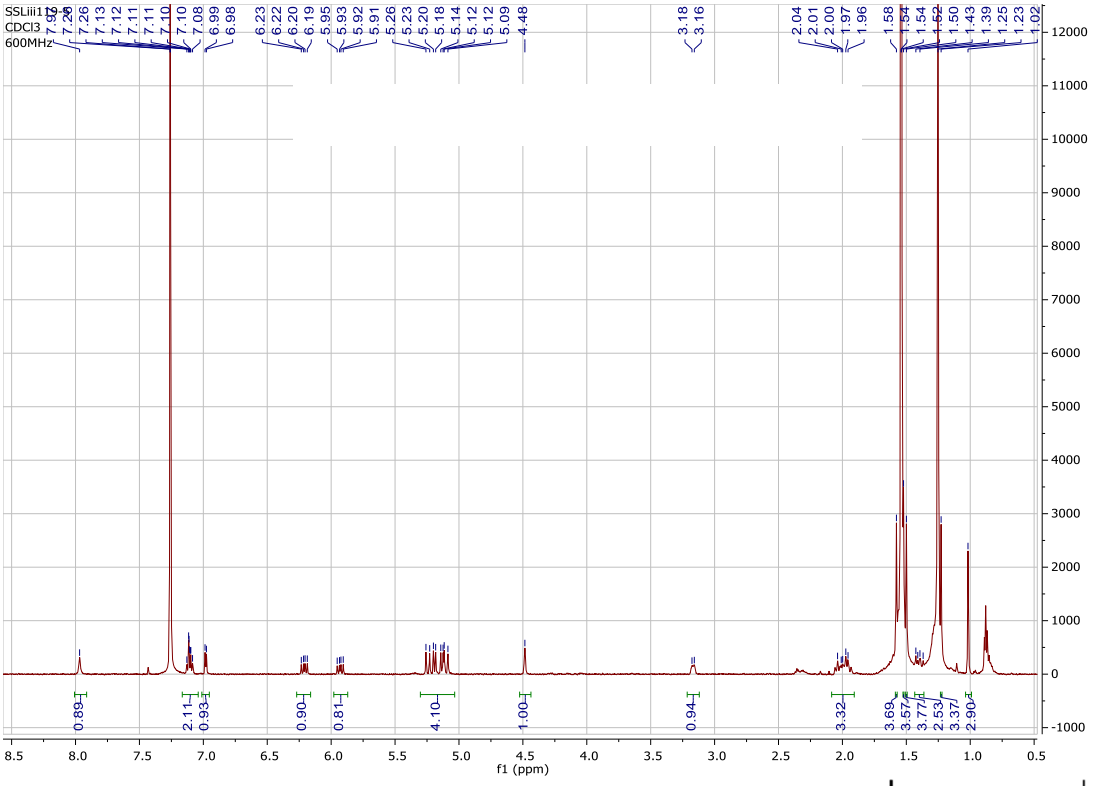


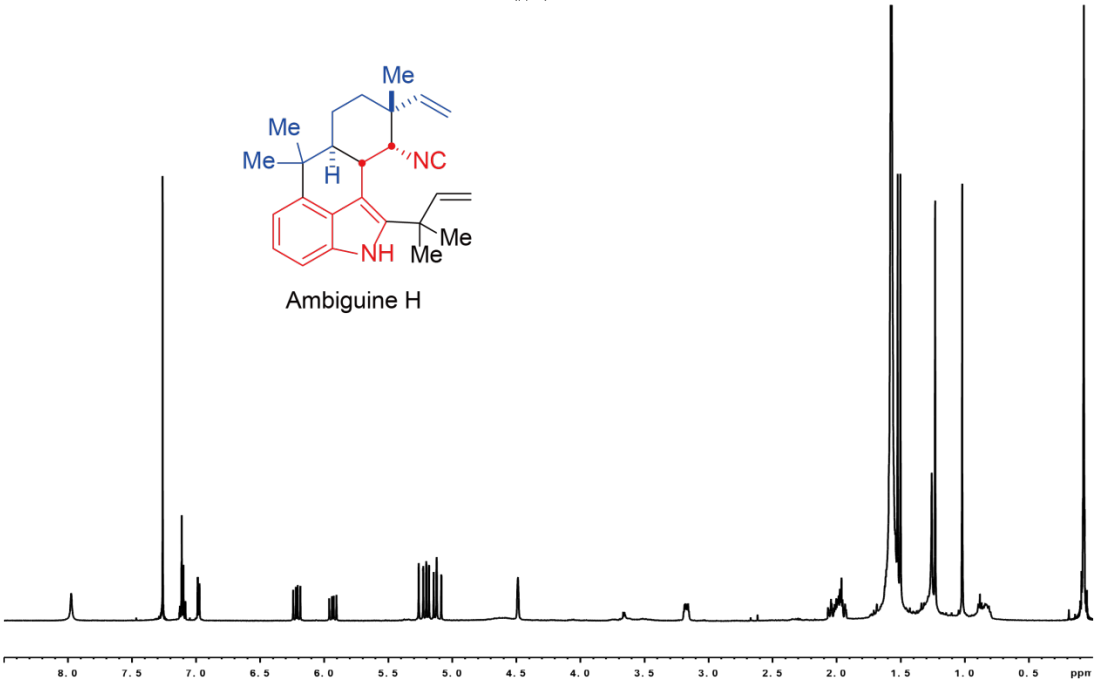
HSQC and HMBC NMR spectra of Compound **6** in C_6D_6 at 600 MHz





^{1}H NMR spectrum of Ambiguine H in $CDCl_3$ at 600 MHz (A) in comparison with published spectrum (B)[17] HRMS: Calcd for C26H32N2 [M+H]+ 373.2638, found 373.2638.





2.8 Reference

1. Bhat, V.; Dave, A.; MacKay, J. A.; Rawal, V. H., Chapter two - the chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. In *The Alkaloids: Chemistry and Biology*, Hans-Joachim, K., Ed. Academic Press: San Diego, 2014; Vol. 73, pp 65-160.
2. Asthana, R.; Srivastava, A.; Singh, A.; Deepali; Singh, S.; Nath, G.; Srivastava, R.; Srivastava, B., Identification of an antimicrobial entity from the cyanobacterium Fischerella sp. isolated from bark of Azadirachta indica (Neem) tree. *J. Appl. Phycol.* **2006,** *18* (1), 33-39.
3. Mo, S.; Krunic, A.; Santarsiero, B. D.; Franzblau, S. G.; Orjala, J., Hapalindole-related alkaloids from the cultured cyanobacterium Fischerella ambigua. *Phytochem.* **2010,** *71* (17-18), 2116-2123.
4. Mo, S.; Krunic, A.; Chlipala, G.; Orjala, J., Antimicrobial ambiguine isonitriles from the cyanobacterium Fischerella ambigua. *J. Nat. Prod.* **2009,** *72* (5), 894-899.
5. Becher, P. G.; Keller, S.; Jung, G.; Sussmuth, R. D.; Juttner, F., Insecticidal activity of 12-epi-hapalindole J isonitrile. *Phytochem.* **2007,** *68* (19), 2493-2497.
6. Cagide, E.; Becher, P. G.; Louzao, M. C.; Espiña, B.; Vieytes, M. R.; Jüttner, F.; Botana, L. M., Hapalindoles from the Cyanobacterium Fischerella: Potential Sodium Channel Modulators. *Chem. Res. Toxicol.* **2014,** *27* (10), 1696-1706.
7. Zhang, X.; Smith, C. D., Microtubule effects of welwistatin, a cyanobacterial indolinone that circumvents multiple drug resistance. *Mole. Pharmacol.* **1996,** *49* (2), 288-294.
8. Walton, K.; Gantar, M.; Gibbs, P.; Schmale, M.; Berry, J., Indole Alkaloids from Fischerella Inhibit Vertebrate Development in the Zebrafish (Danio rerio) Embryo Model. *Toxins* **2014,** *6* (12), 3568-3581.
9. Hillwig, M. L.; Zhu, Q.; Liu, X., Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium Fischerella ambigua. *ACS Chem. Biol.* **2014,** *9* (2), 372-377.
10. Hillwig, M. L.; Fuhrman, H. A.; Ittiamornkul, K.; Sevco, T. J.; Kwak, D. H.; Liu, X., Identification and Characterization of a Welwitindolinone Alkaloid Biosynthetic Gene Cluster in the Stigonematalean Cyanobacterium Hapalosiphon welwitschii. *ChemBioChem* **2014,** *15* (5), 665-669.
11. Hillwig, M. L.; Liu, X., A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014,** *10* (11), 921-923.
12. Micallef, M. L.; Sharma, D.; Bunn, B. M.; Gerwick, L.; Viswanathan, R.; Moffitt, M. C., Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol.* **2014,** *14*, 213-230.
13. Park, A.; Moore, R. E.; Patterson, G. M. L., Fischerindole L, a new isonitrile from the terrestrial blue-green alga *Fischerella muscicola*. *Tetrahedron Lett.* **1992,** *33* (23), 3257-3260.
14. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, unusual

alkaloids from the blue-green algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to fischerindoles and hapalinodoles. *J. Am. Chem. Soc.* **1994,** *116* (22), 9935-9942.
15. Richter, J. M.; Ishihara, Y.; Masuda, T.; Whitefield, B. W.; Llamas, T.; Pohjakallio, A.; Baran, P. S., Enantiospecific total synthesis of the hapalindoles, fischerindoles, and welwitindolinones via a redox economic approach. *J. Am. Chem. Soc.* **2008,** *130* (52), 17938-17954.
16. Maimone, T. J.; Ishihara, Y.; Baran, P. S., Scalable Total Syntheses of (-)-Hapalindole U and (+)-Ambiguine H. *Tetrahedron* **2015,** *71* (22), 3652-3665.
17. Raveh, A.; Carmeli, S., Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* **2007,** *70* (2), 196-201.
18. Smitka, T. A.; Bonjouklian, R.; Doolin, L.; Jones, N. D.; Deeter, J. B.; Yoshida, W. Y.; Prinsep, M. R.; Moore, R. E.; Patterson, G. M. L., Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae. *J. Org. Chem.* **1992,** *57* (3), 857-861.
19. Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L., Evolutionary relationships of microbial aromatic prenyltransferases. *PLoS One* **2011,** *6* (11), e27336.
20. Steffan, N.; Grundmann, A.; Yin, W. B.; Kremer, A.; Li, S. M., Indole prenyltransferases from fungi: a new enzyme group with high potential for the production of prenylated indole derivatives. *Curr. Med. Chem.* **2009,** *16* (2), 218-231.
21. Ding, Y.; Williams, R. M.; Sherman, D. H., Molecular analysis of a 4-dimethylallyltryptophan synthase from Malbranchea aurantiaca. *J. Biol. Chem.* **2008,** *283* (23), 16068-16076.
22. Ivleva, N. B.; Golden, S. S., Protein extraction, fractionation, and purification from cyanobacteria. *Methods Mol. Biol.* **2007,** *362*, 365-373.
23. Tanner, M. E., Mechanistic studies on the indole prenyltransferases. *Nat. Prod. Rep.* **2015,** *32* (1), 88-101.
24. Mahmoodi, N.; Tanner, M. E., Potential rearrangements in the reaction catalyzed by the indole prenyltransferase FtmPT1. *ChemBioChem* **2013,** *14* (15), 2029-2037.
25. Caballero, E.; Avendano, C.; Menendez, J. C., Brief total synthesis of the cell cycle inhibitor tryprostatin B and related preparation of its alanine analogue. *J. Org. Chem.* **2003,** *68* (18), 6944-6951.
26. Luk, L. Y.; Qian, Q.; Tanner, M. E., A cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase? *J. Am. Chem. Soc.* **2011,** *133* (32), 12342-12345.
27. Mahmoodi, N.; Tanner, M. E., Potential Rearrangements in the Reaction Catalyzed by the Indole Prenyltransferase FtmPT1. *Chembiochem* **2013,** *14* (15), 2029-2037.
28. Andersen, R. A., Algal culturing techniques. **2005**.

29. Hoppe, I.; Schöllkopf, U., Synthesis and Biological Activities of the Antibiotic B 371 and its Analogs. *Liebigs Annalen der Chemie* **1984,** *1984* (3), 600-607.
30. Evans, J. R.; Napier, E. J.; Yates, P., Isolation of a new antibiotic from a species of Pseudomonas. *J Antibiot (Tokyo)* **1976,** *29* (8), 850-852.

Note:

This work has been published as “Hapalindole/Ambiguine Biogenesis is Mediated by a Cope Rearrangement, C−C Bond-Forming Cascade.” Shasha Li, Andrew N. Lowell, Fengan Yu, Avi Raveh, Sean A. Newmister, Nathan Bair, Jeffrey M. Schaub, Robert M. Williams, and David H. Sherman. *J. Am. Chem. Soc.* **2015**, *137*, 15366-15369.

Author contributions:

S.L. and D.H.S. designed the research. S.L. performed all experiments. A.N.L. synthesized the indole isonitrile. S.L., A.N.L., S.A.N., D.H.S., A.R., N.B., and J.M.S. conducted data analysis and interpretation. F.Y. performed bioinformatics analyses. S.L., A.N.L., D.H.S., and R.M.W. contributed to manuscript preparation.

Chapter 3
Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids

3.1 Abstract

The formation of C–C bonds in an enantioselective fashion to create complex polycyclic scaffolds in the hapalindole- and fischerindole-type alkaloids from Stigonematales cyanobacteria represents a compelling and urgent challenge in adapting microbial biosynthesis as a catalytic platform in drug development. Here we determine the biochemical basis for tri- and tetracyclic core formation in these secondary metabolites, involving a new class of cyclases that catalyze a complex cyclization cascade, which was identified and briefly described in Chapter 2.

3.2 Introduction

Carbon–carbon bond construction is the backbone of organic chemistry and encompasses some of the most challenging reactions in synthetic methodology[1]. Enzymes that control enantioselectivity during C–C bond formation are of great interest because of their potential to catalyze difficult transformations under mild, environmentally sustainable conditions[2]. While biocatalysts have been investigated extensively during the past few decades, those responsible for complex cyclization reactions remain underexplored[3-5], despite some notable progress in terpenoid and polyketide systems[6-7] and several examples of Diels–Alderases[8-10]. Here, we report a new class of Stigonematales (Stig) cyclases that catalyze a highly stereoselective intramolecular ring formation, creating four new chiral centers via a Cope

rearrangement and C–C bond–forming cascade, the mechanism of which was recently described[11].

Hapalindoles and the related ambiguines, fischerindoles, and welwitindolinones are bioactive indole alkaloids from cyanobacteria that have attracted substantial attention because of their unique pharmacological profiles and complex chiral structures[12-13]. The most notable structural feature of this family is their stereo- and regiochemically diverse polycyclic ring system (**Figure 3-1**). We recently disclosed a new pathway intermediate (**3**; **Figure 3-1a**) that completely revised the prevailing hypotheses about the formation of this complex heterocyclic core[11]. Instead of a cationic catalysis that directly forms hapalindoles (**Figure 3-1b**, **4**-**7**) from *cis*-indole isonitrile (**1**) and geranylpyrophosphate (GPP, **2**), the action of an aromatic prenyltransferase (FamD2) appends **2** onto the C3 position of **1**, and the resulting product (**3**) is subsequently processed into a hapalindole (**4**) via a Cope rearrangement and intramolecular ring cyclization catalyzed by FamC1.[11] However, these initial results prompted two intriguing questions: what explains the regioselective C-ring divergence between hapalindoles and fischerindoles; and what sets the chirality of the C10, C11, C12, and C15 stereocenters?

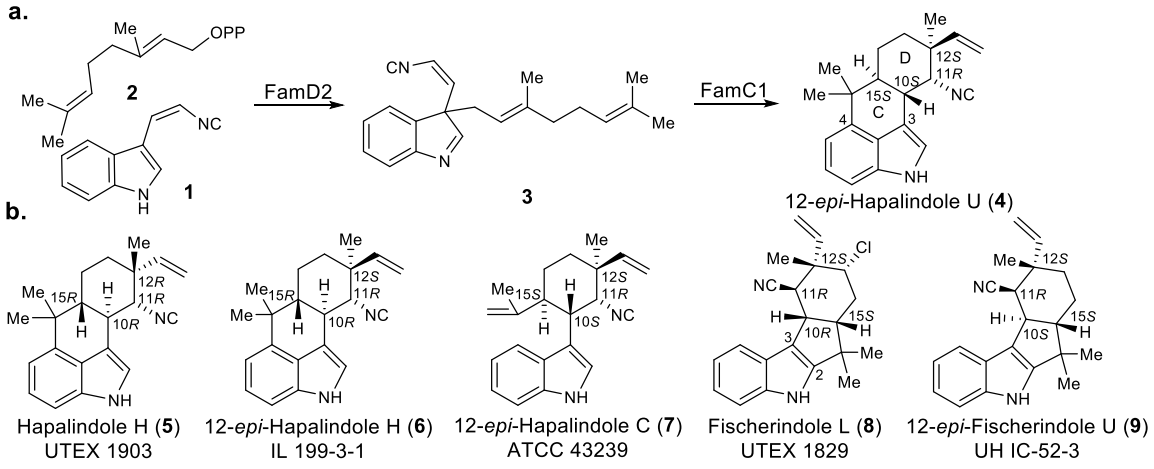


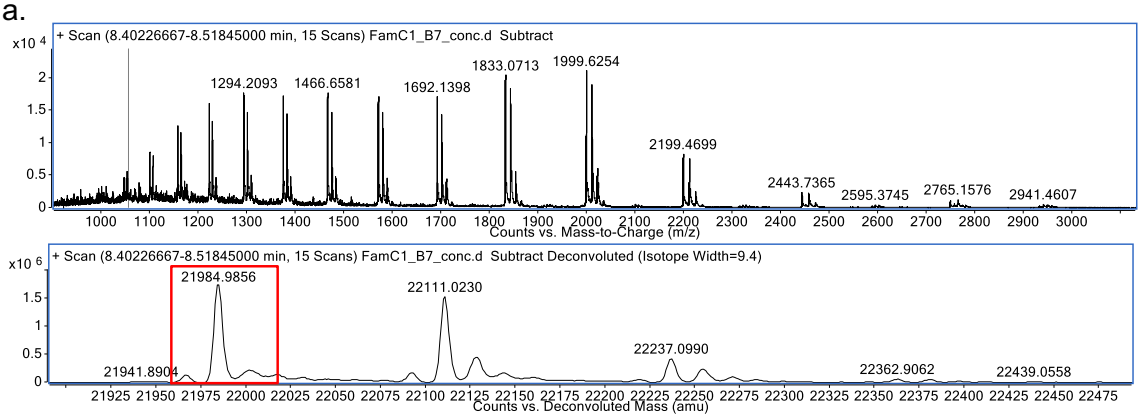
Figure 3-1. Select hapalindole -type alkaloids. a) Biosynthetic pathway leading to 12-epi-hapalindole U;[11] b) representative indole alkaloids from related cyanobacterial metabolic pathways illustrating the variant polycyclic core systems.[14-18]

3.3 Research and Discussion

3.3.1 Expression of Stig cyclases

We originally identified Stig cyclase FamC1 through cell-free lysate fractionation and demonstrated its direct role in formation of 12-*epi*-hapalindole U (**4**)[11]. However, *in vitro* assays with purified protein were hampered by the apparent inability to heterologously express FamC1 in *Escherichia coli*. Through protein sequence analysis, an N-terminal 23 amino acid region was predicted to be a transmembrane segment, which we reasoned led to the failed FamC1 expression. Mass spectrometry analysis of the isolated functional enzyme showed a mass that exactly matched the protein lacking the predicted transmembrane sequence (**Figure 3-2**). This result indicated that a post-translational modification is required to form the mature cyclase. Thus, by cloning the corresponding truncated gene, we successfully overexpressed FamC1 in *E. coli* as a soluble protein and confirmed its ability to convert intermediate **3** into **4** *in vitro* (**Figure 3-1**a).

a.



b.

MW1: 24734.8 MW2: 21984.4

Predicted transmembrane sequence

```
MKRKLIVAVV FLIFICLGIN TPAHATSAVS IPINNAGFEN PFMDVVDDYT IDTPPGWTTY  60
DPNNLVPEKR TTWTSNNGVG YVGPGTQFYN QLAPEGRNIG YIYLSQNPGS GVAGFEQILD  120
ATLEPDTKYT LTVDVGNLAG TFKGLSFAGF PGYRVELLAG DTVLAADHNN LFIKEGEFKT  180
STVTYTSTAK DLHLGQKLGI RLVNLLQDKF SGLDFDNVRL TTEPTET                227
```

Figure 3-2. Mass spectrometry analysis of functional FamC1 isolated from UTEX 1903 cell-free lysate. a) The mass of the protein was 21984 Da (lower spectrum) after deconvolution (upper spectrum); b) MW1 is the calculated mass of intact FamC1, MW2

is the protein mass start from amino acid T, which matched exactly with results from a, indicating the mature FamC1 sequence is truncated at that position. Transmembrane region prediction was performed at TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and GenScript-OptimumGene™.

3.3.2 FamC1 and homologs FilC1, HpiC1

12-*epi*-Hapalindole U is an unexpected new metabolite bearing a pattern of stereocenters (C10*S*, 11*R*, 12*S*, and 15*S*) that had not been reported from UTEX 1903 cyanobacterium, suggesting that our knowledge of the full spectrum of natural products generated by this strain is likely incomplete. This discovery inspired us to search for this metabolite in other biosynthetic pathways based on previously characterized compounds and the presence of FamC1 homologs. Thus, genome sequencing and bioinformatic analysis of hapalindole-producing cyanobacteria led us to identify several gene clusters (**Figure 3-3**), including the *fil* gene cluster from *Fischerella* sp. IL-199-3-1[ref.16] and the *hpi* gene cluster from *Fischerella* sp. ATCC 43239[ref.14]. Protein sequence analysis of the nine annotated Stig cyclases (FilC1–FilC4 and HpiC1–HpiC5) against FamC1 resulted in identification of two FamC1 homologs, FilC1 (94% identity) from IL-199-3-1 and HpiC1 (85% identity) from ATCC 43239 (**SI Figure 3-1**).

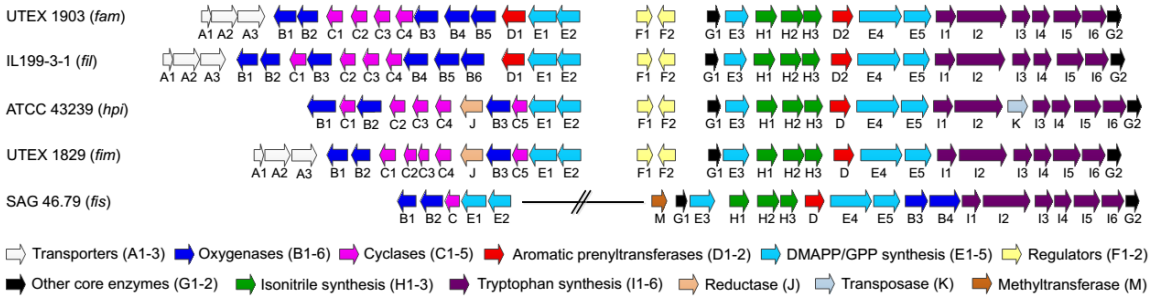


Figure 3-3. Gene clusters identified from hapalindole producing strains. fam gene cluster from ambiguine producing strain Fischerella ambigua UTEX 1903;[11] fil gene cluster from ambiguine producing strain Fischerella sp. IL-199-3-1 (this study); hpi gene cluster from hapalindole producing strain Fischerella sp. ATCC 43239[14] (nomenclature has been revised for consistency); fim gene cluster from fischerindole producing strain Fischerella muscicola UTEX 1829 (this study); fis gene cluster from fischerindole producing strain Fischerella sp. SAG 46.79 (this study).

To verify our hypothesis that FilC1 and HpiC1 catalyze formation of the same product as FamC1, we overexpressed and purified these two proteins from

E. coli using the strategy of transmembrane-segment truncation to achieve *in vitro* activity. Both of these FamC1 homologs exhibited the ability to cyclize **3** to **4** (**Figure 3-4**; see **SI Figure 3-2** for the structure of **10**), indicating that proteins of this subclass are responsible for catalyzing formation of a hapalindole-type tetracyclic core containing the C10*S*, 11*R*, 12*S*, and 15*S* stereocenter motif.

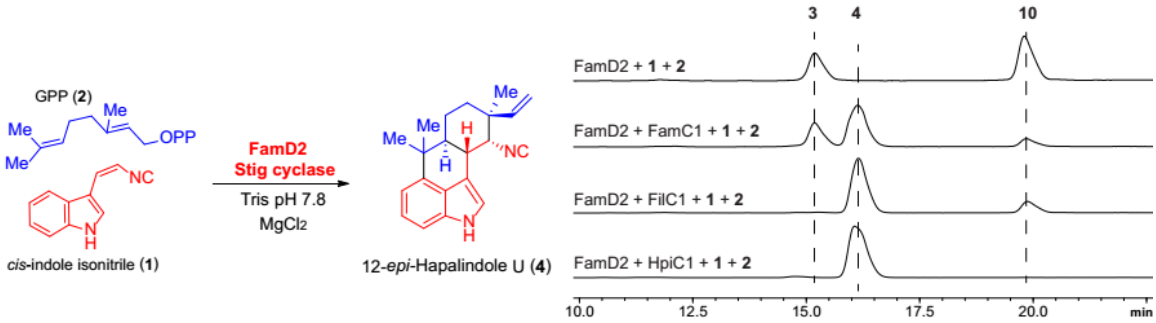


Figure 3-4. In vitro assay of recombinant Stig cyclases FamC1, FilC1 and HpiC1 by producing 12-*epi*-hapalindole U.Assays were conducted in a one-pot enzymatic tandem reaction by incubating substrates indole isonitrile (**1**) and GPP (**2**) with FamD2 and the Stig cyclases.

3.3.3 Fischerindole producers FimC5, FisC

One of the most intriguing questions regarding construction of the broader family of hapalindole core structures is the basis for regiospecific C-ring formation in hapalindoles (**4**, C3–C4) and fischerindoles (**8**, C3–C2). Our studies on 12-*epi*-hapalindole U suggested a novel mechanism involving a concerted FamC1-catalyzed cyclization to afford the tetracyclic hapalindoles. However, it remained unclear whether a similar type of concerted mechanism was employed to generate the fischerindole core and if tricyclic hapalindoles could serve as active intermediates toward formation of the tetracyclic cores. Thus, we expanded our analysis to include two fischerindole-producing strains, *Fischerella muscicola* UTEX 1829[ref.18] and *Fischerella* sp. SAG 46.79 (**Figure 3-3**)[15]. The genomic DNA of UTEX 1829 was sequenced and mined for the fischerindole biosynthetic gene cluster (*fim*), revealing a 39-kb region encoding 31 predicted proteins, including five hypothetical Stig cyclases (FimC1–FimC5). The *fis* gene cluster from strain SAG 46.79 is split into a 22-kb region encoding 18 proteins

and a 5.3-kb region encoding five proteins, with only one open reading frame annotated as a cyclase (FisC).

Our ability to predict the activity of FilC1 and HpiC1 through comparative sequence analysis suggested that the same strategy could be applied to identify the Stig cyclases responsible for fischerindole formation. Thus, all putative cyclase amino acid sequences in our study were compared and classified into subgroups according to their identities. FimC1 to FimC4 showed relatively high identities (>90%) with FamC1 to FamC4, respectively, and were classified into the same subgroups. FimC5 and FisC were 90% identical and displayed relatively low similarity to any FamC subgroup. These results led us to hypothesize that FimC5 and FisC are responsible for fischerindole formation and may generate the same stereochemical pattern at the four chiral centers. Thus, the two heterologously expressed and purified proteins were separately incubated with chemoenzymatically generated **3** for activity analysis, and two new compounds were observed from both reactions (**Figure 3-5**). NMR analysis revealed these two proteins catalyzed formation of the same molecules: the major product was 12-*epi*-fischerindole U (**9**, C10*S*, 11*R*, 12*S*, and 15*S*) and the minor one 12-*epi*-hapalindole C (**7**), a tricyclic hapalindole possessing the same stereochemistry as **9**. **7** also failed to convert to **9** upon incubation with FimC5 or FisC, indicating that it is a shunt metabolite and cannot re-enter the catalytic cycle to form fischerindoles.

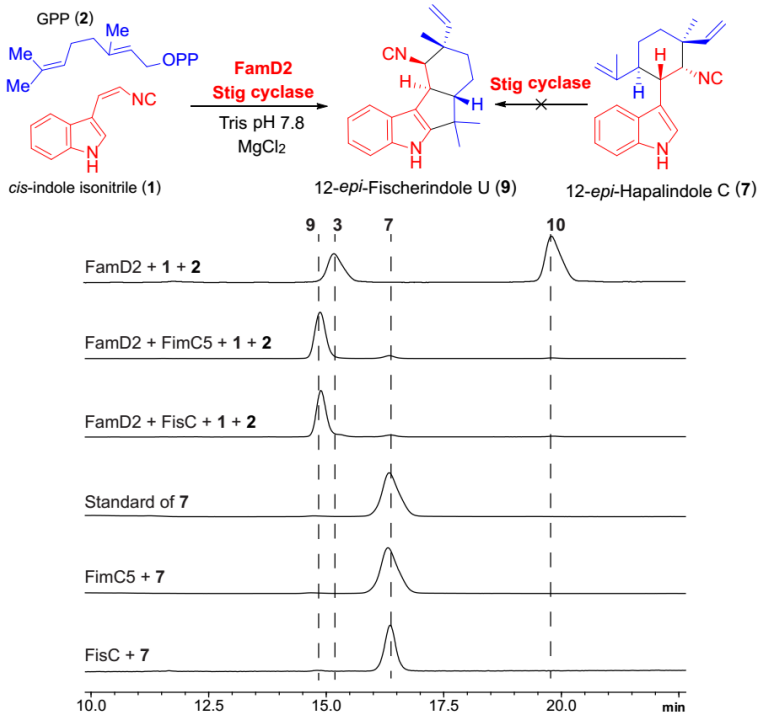


Figure 3-5. In vitro assay of recombinant Stig cyclases FimC5 and FisC by producing 12-*epi*-fischerindole U (**9**) and 12-*epi*-hapalindole C (**7**). Assays were conducted in a one-pot enzymatic tandem reaction by incubating substrates indole isonitrile (**1**) and GPP (**2**) with FamD2 and the Stig cyclases.

3.3.4 Hapalindole H producer FamC2-FamC3 heterodimer

Our studies conclusively demonstrated that **3** is the common intermediate to all hapalindole- and fischerindole-type core ring systems. However, the functional basis of stereodiversification at positions C10, C11, C12, and C15 by the suite of homologous Stig cyclases remained unclear. We hypothesized that each cyclase subgroup is responsible for a specific stereochemistry observed in the core ring system. In support of this reasoning, we were able to isolate hapalindole H (**5**, C10*R*, 11*R*, 12*R,* and 15*R*) as an additional product from the cell-free lysate reaction of **1** with **2** (**Figure 3-6**), a compound with a different stereochemical configuration from the previously isolated **4**[refs.11, 17]. Thus, we sought to use the FamC cyclases as a model to explore the basis for stereochemical control in these indole alkaloids.

To probe the hypothesis that the alternative cyclases control the relative stereochemical configurations observed amongst the variant hapalindole

metabolites, the three additional FamC proteins (FamC2–FamC4) were overexpressed in *E. coli* and purified. Surprisingly, *in vitro* assays with FamC2 and FamC3 individually failed to generate any product (**Figure 3-6**). However, testing of the proteins in pairs revealed that co-incubating FamC2 and FamC3 in a 1:1 ratio resulted in efficient formation of hapalindole H. We also observed that mixing the two proteins helped solubilize FamC2, which precipitated partially when purified as a homogeneous protein. Because size-exclusion chromatography showed that the other FamC homologs were active as homodimers, it is reasonable to suggest that the association between FamC2 and FamC3 may involve heterodimer formation (**SI Figure 3-3**)[19-21]. This hypothesis was tested using a Ni-NTA pulldown assay in which untagged FamC2 co-purified with His-tagged FamC3. After further purification using size-exclusion chromatography, this complex was shown to generate **5**. Thus, we propose that the hapalindole biosynthetic machinery may engage various homo- and heterodimeric forms of the cyclase monomers to control the stereo- and regiochemical outcomes of the indole alkaloid metabolites. This hypothesis could address the apparent conflict that different stereochemical patterns in the ambiguine and fischerindole metabolites are produced between UTEX 1903 and UTEX 1829 despite the high sequence identity between the FamC1–FamC4 and FimC1–FimC4 Stig cyclases (**SI Figure 3-1** and **4**), a factor we now attribute to expanded cyclase scope through differential dimer formation.

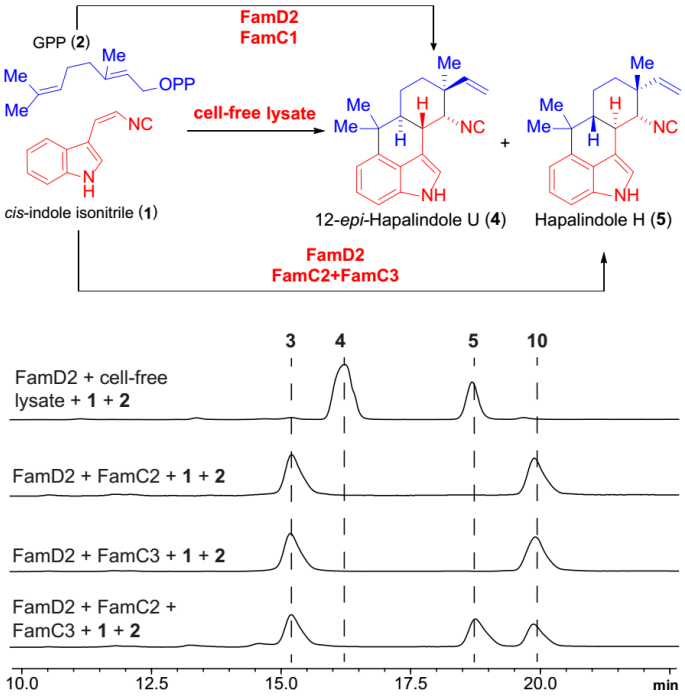


Figure 3-6. In vitro assay of recombinant Stig cyclases FamC2-FamC3 by producing hapalindole H (**5**). Assays were conducted in a one-pot enzymatic tandem reaction by incubating substrates indole isonitrile (**1**) and GPP (**2**) with FamD2 and the Stig cyclases.proteins, cell-free lysate is from UTEX 1903 strain.[11] (see **SI Figure 3-2** for the structure of **10**).

3.4 Conclusion

In summary, *in vitro* reconstitution of the functional activities of this new class of indole alkaloid cyclases provides conclusive evidence that they are responsible for ring formation and stereochemical control in the biosynthesis of hapalindoles and fischerindoles. We identified and characterized the function of several Stig cyclases from five hapalindole-producing cyanobacterial strains, including three novel gene clusters. This work demonstrates that the FamC1–FilC1–HpiC1 class catalyzes formation of 12-*epi*-hapalindole U (**4**) and FimC1–FisC directly produces 12-*epi*-fischerindole U (**9**), while hapalindole H (**5**) is assembled by a heterodimeric combination of FamC2 and FamC3. Moreover, tricyclic hapalindole **7** was not transformed into fischerindole **9**, suggesting that tricyclic compounds are shunt products and not intermediates on the pathway to tetracycle formation. These data provide a plausible basis for enzymatic regio- and stereoselective ring formation, which is controlled by the related Stig-type

cyclases acting on central bicyclic precursor **3**. Importantly, the protein–protein interactions between FamC2 and FamC3 suggest that some functional cyclases are comprised of more than one protein to create the variant stereo- and regiochemical motifs observed within this large class of indole alkaloids. Further efforts involving protein structural studies to probe directly the mechanism of these remarkable biocatalysts are underway.

3.5 Experimental section

3.5.1 General materials and methods

All NMR spectra were acquired on Varian 400, 600 and 700 MHz spectrometers. Proton and carbon signals are reported in parts per million (δ) using residual solvent signals as an internal standard. The LC–MS analysis was performed on a Shimadzu 2010 EV APCI spectrometer equipped with an Agilent Extend C18 5 μm 4.6 × 150 mm column, using a mobile phase gradient of 70–90% acetonitrile in water over 22 min and was monitored by UV absorption at 220 nm. Preparative-scale HPLC was performed using an Agilent Extend C18 10 μm 10 × 250 mm column, using a mobile phase gradient of 70–90% acetonitrile in water over 28 min. High-resolution APCIMS spectra and protein mass spectrometry sequence analysis were obtained from an Agilent 6520 Q-TOF mass spectrometer equipped with an Agilent 1290 HPLC system at the University of Michigan core facility in the Department of Chemistry, with MS grade solvents. Optical rotations were measured in CH_2Cl_2 at 25 °C at the sodium D line.

Escherichia coli strain DH5α (Invitrogen) was used for plasmid manipulation, BL21(DE3/pRARE) was used for protein expression. KOD Xtreme Hot Start DNA polymerase (EMD Millipore) was used for polymerase chain reactions. Restriction endonucleases (NheI, XhoI, NdeI and HindIII) and T4 DNA ligase were purchased from New England BioLabs. Primers were purchased from Integrated DNA Technologies. PureLink Quick Plasmid Miniprep Kit (Invitrogen) was used to prepare plasmid DNA. All cloned plasmids were confirmed by Sanger sequencing at the University of Michigan DNA Sequencing Core. Isopropyl-D-thiogalactopyranoside (IPTG; GoldBio) was used to induce expression; benzonase and lysozyme used in purification were purchased from Sigma-Aldrich; phenylmethane sulfonyl fluoride (PMSF) was dissolved in isopropanol and used as serine protease inhibitor during protein purification. Ni-NTA agarose from Invitrogen was used to purify 6× His-tag proteins. LB broth and agar (EMD Millipore) were used for all *E. coli* culturing.

3.5.2 Cyanobacterial culturing, genomic DNA extraction and sequencing

Cyanobacteria strains *Fischerella ambigua* UTEX 1903 and *Fischerella muscicola* UTEX 1829 were purchased from the UTEX Culture Collection of Algae at the University of Texas at Austin. *Fischerella* sp. IL-199-3-1 was obtained from S. Carmeli and A. Raveh (Tel Aviv University), *Fischerella* sp. ATCC 43239 was purchased from the ATCC microbial collection, and *Fischerella* sp. SAG 46.79 was obtained from J. Orjala (University of Illinois at Chicago). The cyanobacterial culturing, genomic DNA isolation, and whole genome sequencing of these strains followed the established protocols[11].

3.5.3 Protein preparation

Protein construct. Vector pET28a (NdeI and HindIII sites) was used to construct the FamD2 expression system. Vector pET28a (NheI and XhoI sites) was used for the cloning of all seven cyclases. The *fam*D2 gene was amplified from UTEX 1903 genomic DNA, *fil*C1 from IL-199-3-1 genomic DNA, and *fis*C from SAG 46.79 genomic DNA. For the other five cyclase genes, *fam*C1/*fam*C2/*fam*C3 (UTEX 1903), *hpi*C5 (ATCC 43239), and *fim*C5 (UTEX 1829), the DNA sequences used for cloning were purchased from Integrated DNA Technologies (IDT) after rare codon optimized because of low expression levels in *E. coli* using the native DNA sequence. All primers used are listed in **SI Table 3-1**.

Protein expression and purification. Positive plasmids from gene cloning were transformed into electro-competent BL21 (DE3/pRARE) *E. coli* cells. A single colony was picked and inoculated into LB medium (10 ml, 50 μg/ml kanamycin), shaken overnight at 37 °C (200 r.p.m.), and used to inoculate a 2.8-L Fernbach flask containing prewarmed LB medium (1 L, 50 μg/ml kanamycin). After incubation (37 °C, 200 r.p.m.) to an optical density of 0.6, the culture was cooled to 16 °C. IPTG was added to a final concentration of 0.2 mM. After overnight incubation (16 °C, 200 r.p.m.), the cells were centrifuged (6,000 × *g* at 4 °C for 15 min). The following procedures were performed at 4 °C in a temperature controlled room. The cell pellets were resuspended in 20 ml of lysis

buffer (10 mM HEPES, 50 mM NaCl, 20 mM imidazole, 0.2 mM TCEP, 10% glycerol), containing 1 mM PMSF, 0.5 mg/ml of lysozyme, and 1 μl of benzonase. The mixture was incubated for 30 min and sonicated on ice for 2 min using 10 s pulses followed by a 50 s pause. The sample was centrifuged (60,000 × *g* at 4 °C for 35 min) to remove cellular debris, and the supernatant was loaded onto Ni-NTA agarose prewashed with lysis buffer and loaded by gravity. The column was washed with buffer (10 mM HEPES, 300 mM NaCl, 0.2 mM TCEP, 10% glycerol, 20 mM imidazole) to remove unbound proteins, followed by 10 ml of elution buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol, 300 mM imidazole) to elute the desired protein. The protein-containing fractions were combined and desalted using a PD10-desalting column (GE Healthcare) pre-equilibrated with storage buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol). The purified protein was analyzed by SDS–PAGE gel for homogeneity, assessed by Nanodrop (Company) for concentration using calculated molar extinction coefficients, flash-frozen in liquid nitrogen, and stored at –80 °C.

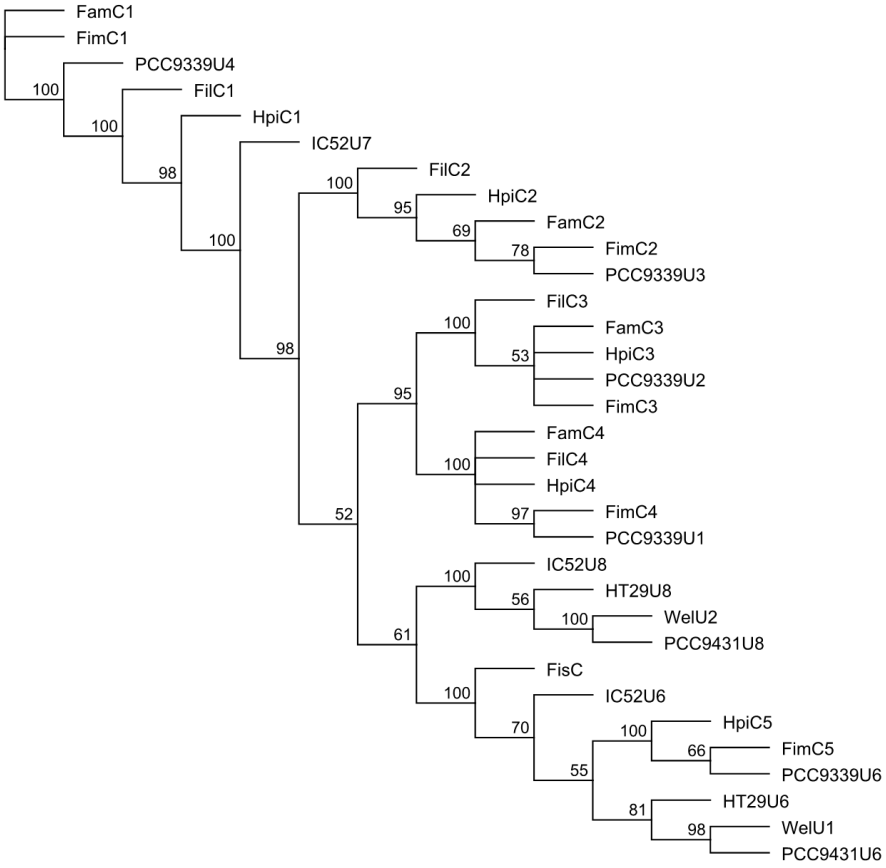
3.5.4 In vitro enzymatic assays

The synthesis of indole isonitrile (**1**) was described previously[11], GPP (**2**) was purchased from supplier http://Isoprenoids.com/ (purity > 95%). The activity assays for cyclases (FamC1/FilC1/HpiC1/FimC5/FisC/FamC2/FamC3) were initially conducted by incubating them with pure intermediate **3**, which was produced by incubating FamD2 with substrates **1** and **2** at pH 10.5. However, compound **3** was not stable in extended storage. Thus, the reactions were set up in one-pot by incubating substrates **1** and **2** with enzyme cascades (FamD2 and the selected cyclase(s)), which also resulted in the generation of byproduct **10** (**SI Figure 3-2**)[11]. The general protocol for enzymatic reactions is as follows. A 50 μl reaction containing 5 μM of FamD2, 10 μM of cyclase(s) or 1 mg/ml of cell-free lysate[11], 1 mM of **1**, 1 mM of **2**, 5 mM of $MgCl_2$, and 50 mM of Tris buffer (pH 7.8) was incubated at 37 °C. The reaction was quenched after 5 h and extracted three times with an equal volume of ethyl acetate. The organic layers were combined,

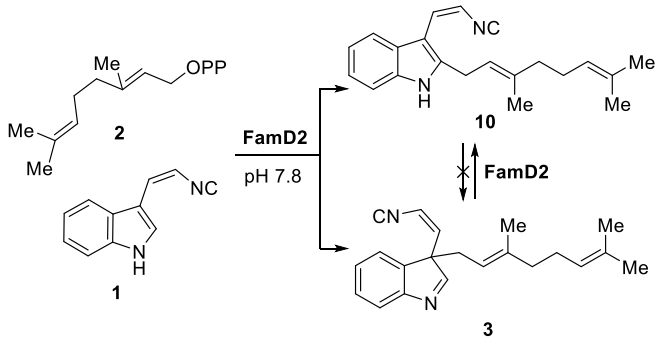
dried, and re-dissolved in 100 μl of acetonitrile for LC–MS analysis. The same conditions were used to test the activity of FamC2 and FamC3 in combination by mixing 10 μM of each protein with FamD2. Fresh FamC2 was used for *in vitro* assay because of precipitation during the freeze-thawing process. For the conversion assay between 12-*epi*-hapalindole C (**7**) and 12-*epi*-fischerindole U (**9**), compound **7** was purified by HPLC from the reaction with FimC5, and dissolved in DMSO. 0.5 mM of **7** was incubated at 37 °C with 15 μM FimC5 or FisC in 50 mM Tris buffer (pH 7.8). The reactions were quenched and analyzed as described above.

For structure analysis of enzymatic products (**4**, **5**, **7**, and **9**), the reactions were scaled up to 5 ml and incubated under identical conditions. The extracted products were purified by HPLC (see section “General materials and methods”), then concentrated, dissolved in C6D6, CDCl3, or CD2Cl2, and analyzed by NMR (**SI Figure 3-5**; **SI Table 3-6** and **7**, and **Spectra section**). The spectra of known compounds **5**, **7**, and **9** were identical to those previously reported[22-23].

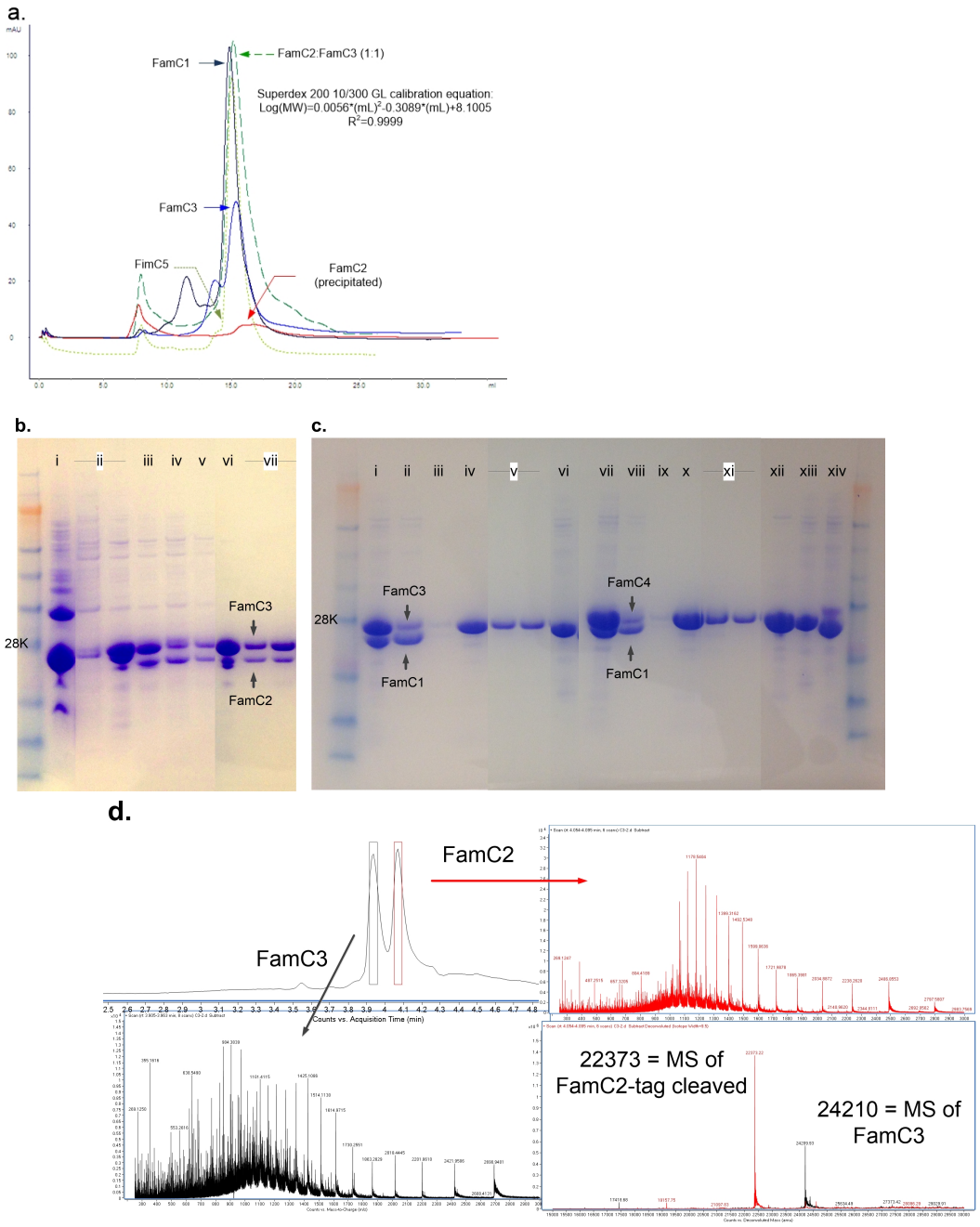
3.6 Supplementary information

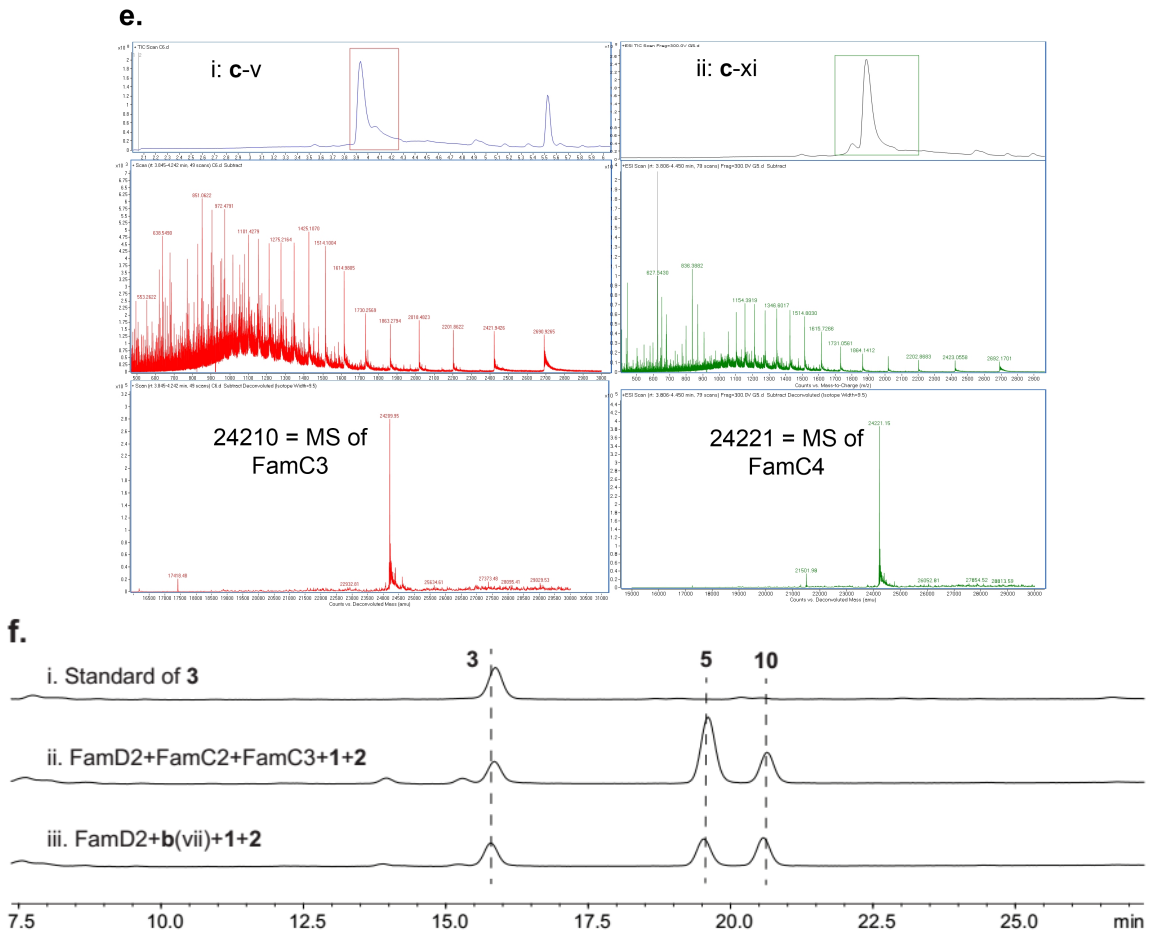


SI Figure 3-1. Phylogenetic analysis and identity table of Stig cyclases (Supplementary Data Set 1) annotated from this study and public database information: FamC1 - FamC4, FilC1 - FilC4, HpiC1 - HpiC5, FimC1 – FimC5, and FisC (this study); PCC9339U1 – PCC9339U4, and PCC9339U6 from *Fischerella* sp. PCC 9339 (JGI IMG/ER: 2516653082);[14] PCC9431 U6 and PCC9431 U8 from *Fischerella* sp. PCC 9431 (JGI IMG/ER: 2512875027);[14] IC52U6 – IC52U8 from *Hapalosiphon welwitschii* UH strain IC-52-3;[14] HT29U6 and HT28U8 from *Westiella intricata* UH strain HT-29-1.[14] Phylogenetic trees were constructed using the Geneious Tree Builder program and neighbour-joining method. Numbers at each branch point are the bootstrap values for percentages of 100 replicate trees.



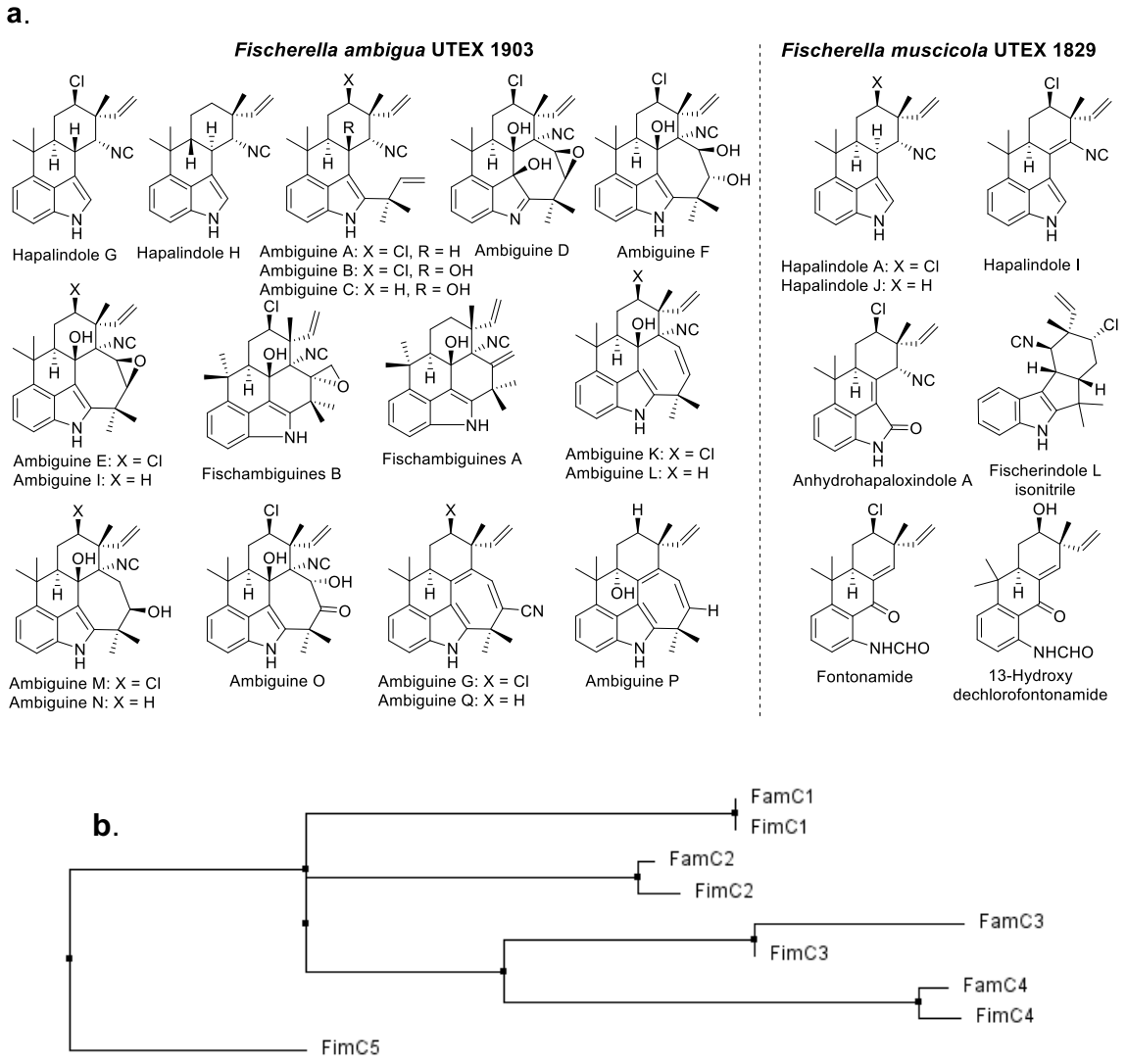
SI Figure 3-2. Compound 10 was produced along with 3 by FamD2 with indole isonitrile (1) and GPP (2) at pH 7.8.[11]


a.
FamC1
FamC2:FamC3 (1:1)
Superdex 200 10/300 GL calibration equation:
Log(MW)=0.0056*(mL)2-0.3089*(mL)+8.1005
R2=0.9999
FamC3
FimC5
FamC2
(precipitated)
b.
i ii iii iv v vi vii
28K
FamC3
FamC2
c.
i ii iii iv v vi vii viii ix x xi xii xiii xiv
28K
FamC3
FamC1
FamC4
FamC1
d.
FamC2
FamC3
Counts vs. Acquisition Time (min)
22373 = MS of
FamC2-tag cleaved
24210 = MS of
FamC3



SI Figure 3-3. Investigation of homodimer and heterodimer formation among Stig cyclases. a). Size exclusion chromatography assays revealed the existence of homodimeric and heterodimeric forms of the proteins. The experiments were conducted using Superdex 200 10/300 GL column (GE Healthcare Life Sciences). The injected mass for samples FamC1 and FimC5 were 1.5 mg, and 0.8 mg for FamC3, which were consistent with their absorptions. For sample FamC2, 0.8 mg of protein was thawed in ice, but precipitated, which resulted in a very low absorption with the filtered supernatant. For the mixture sample FamC2:FamC3, 0.8 mg of FamC2 and 0.8 mg of FamC3 were thawed together and incubated on ice for 1 h before assay; no visible precipitation was observed and its absorption matched FamC1 and FimC5. According to the calibration equation, the molecular weights of FamC1, FimC5, FamC3, and the mixture of FamC2:FamC3 are approximately 50 kDa, indicating the likely existence of homodimers or heterodimers. b). Protein-protein interaction of FamC2 and FamC3 was analyzed using a pull-down assay. The protein construct of FamC2 is His-tag-cleavable using TEV-protease, while FamC3 His-tag is not cleavable. The experiment was conducted by mixing 10 g of FamC2 *E. coli* cell pellet with 3 g of FamC3 *E. coli* cell pellet, and co-purifying the protein extract by Ni-NTA chromatography to obtain a FamC2-FamC3 mixture. TEV-protease was added to cleave the FamC2 His-tag at 4 ℃, overnight. The resulting sample was further purified by a second Ni-NTA column, followed by size exclusion chromatography. SDS-PAGE analysis showed that FamC2 is eluted together with FamC3 after the TEV-cleavage, indicating the existence of stable protein-protein

interaction. i. A mixture of FamC2 and FamC3 pellet; ii. Elution fractions from the first Ni-NTA purification; iii. Proteins after incubation with TEV-protease; iv. Flow-through from the second Ni-NTA purification; v. Wash fraction from the second Ni-NTA purification; vi. Eluted fraction from the second Ni-NTA purification; vii. Fractions from the size exclusion assay containing both FamC3 and His-tag-cleaved FamC2. c). Controls were designed to verify the specific interaction between FamC2 and FamC3, including the combination of FamC1 – FamC3 (i-vi), and FamC1 – FamC4 (vii-xiv). 2.5 mg of purified FamC1 (His-tag cleavable) was mixed with either 2.5 mg of FamC3 (His-tag not cleavable) or same amount of FamC4 (tag not cleavable). The FamC1 His-tag was cleaved by adding TEV-protease into each combination, then applying the same procedure described in b to analyze the possible interactions. The SDS-PAGE results showed that there is no interaction between FamC1 and FamC3, or FamC1 and FamC4, which further demonstrated the specific interaction between FamC2 and FamC3. i. FamC1-FamC3 mixture after incubated with TEV-protease; ii. Flow-through from the Ni-NTA purification, showing that FamC1 passed through the column due to a lack of His-tag binding and no stable protein-protein interaction; iii. Wash fraction from the Ni-NTA purification; iv. Eluted fraction containing only FamC3; v. Fractions from size exclusion column containing only FamC3; vi. Standard of FamC3; vii. FamC1-FamC4 mixture after incubation with TEV-protease; viii. Flow-through from the Ni-NTA purification showing that FamC1 passed through the column due to a lack of His-tag binding and no stable protein-protein interaction; ix. Wash fraction from the Ni-NTA purification; x. Eluted fraction containing only FamC4; xi. Fractions from size exclusion column, containing only FamC4; xii. Standard of FamC4; xiii. Standard of FamC1; xiv. Standard of FamC1-tag cleaved. d) Fraction (b-vii) from size exclusion purification of FamC2-FamC3 combination was examined by Q-TOF mass spectrometry, which showed the existence of both FamC2 (His-tag cleaved) and FamC3. e). Fraction c-v from FamC1-FamC3 combination and fraction c-xi from FamC1-FamC4 combination was examined by Q-TOF mass spectrometry. FamC1 was not detected from either of the fractions, indicating that there is no stable protein-protein interaction between FamC1 and FamC3 or FamC1 and FamC4. f). *In vitro* assay was conducted for fraction b (vii) to verify the existence of a FamC2-FamC3 heterodimer. i. Standard of substrate 3; ii. Positive control with pure FamC2 and FamC3 to produce 5 (hapalindole H); iii. Fraction b (vii) also produced 5, indicating the existence of a FamC2-FamC3 dimer.



SI Figure 3-4. Isolated hapalindole alkaloids. a) UTEX 1903 is a hapalindole and ambiguine producing strain, while UTEX 1829 is a hapalindole and fischerindole producing strain; b) Phylogenetic analysis of the Stig cyclases FamC1-4 (UTEX 1903) and FimC1-5 (UTEX 1829) shows that FamC1-4 and FimC1-4 belong to same subgroups, respectively, and FimC5 is divergent in sequence (and function). Phylogenetic trees were constructed using Jalview program and method “neighbour joining using % identity”.

SI Table 3-1. Primers used for cloning.

Genes	Forward primer	Reverse primer
famD2	AAAAAACATATGAACGATGTTAACCGTATAC	AAAAAAAAGCTTAAGCAAGAGCAAAATCAGCAG
famC1	AAAGCTAGCACCAGCGCTGTGAGCATC	CTCGAGTTAGGTCTCGGTCGGTTCG
famC2	AAAGCTAGCGCGGTGACCACCAGCATC	AAAAAACTCGAGTTAGGTATCAACGGTTTC
famC3	AAAGCTAGCACCGGTGCGGTTAGCATC	AAAAAACTCGAGTTAAATCACCGCGCTTTC
hpiC1	AAAGCTAGCACCAGCGTGGTTAGCATC	CTCGAGTTAGGTTTGCGCCGGTTC
fimC5	AAAGCTAGCGCGAGCGCGGTTTGCATCC	AAAAAACTCGAGTTAGGTTTCCGCCGGCTCC
filC1	AAGGCGCTAGCACATCTGTCGTTTCCATTCCTATCAACAATG	ATCCTCGAGTTAGGCCTCAGTTGGTTCTGC
fisC	AAGGCGCTAGCGCTGTTTCTATTCCTATCAAAAATGCAGG	ATCCTCGAGTTAAGTTTCAGCTGGTTCCGCA

SI Table 3-2. Function annotation of ORFs in the *fam* biosynthetic gene cluster. The Genbank accession number for the *fam* gene cluster is KX451322.

Protein	Size (aa)	Proposed function	Protein homolog (identity/positives %/%)	Accession no.
FamA1	151	hypothetical protein	hypothetical protein [Fischerella sp. PCC 9339] (100/100)	WP_017308642.1
FamA2	397	DevB family ABC transporter membrane fusion protein	DevB family ABC transporter membrane fusion protein [Microcoleus sp. PCC 7113] (60/81)	YP_007124464.1
FamA3	388	ABC transporter	ABC transporter [Coleofasciculus chthonoplastes] (56/75)	WP_006102398.1
FamB1	290	2OG-Fe(II) oxygenase	2OG-Fe(II) oxygenase [Myxococcus hansupus] (36/53)	WP_002634351.1
FamB2	285	Rieske oxygenase	PrnD [Burkholderia vietnamiensis] (26/42)	WP_021160485.1
FamC1	227	Stig cyclase	hypothetical protein [Fischerella sp. PCC 9339] (100/100)	WP_017308636.1
FamC2	230	Stig cyclase	hypothetical protein [Fischerella sp. PCC 9339] (99/100)	WP_017308633.1
FamC3	227	Stig cyclase	hypothetical protein [Fischerella sp. PCC 9339] (67/79)	WP_017308633.1
FamC4	228	Stig cyclase	hypothetical protein [Fischerella sp. PCC 9339] (97/98)	WP_017308633.1
FamB3	363	Rieske oxygenase	PrnD [Myxococcus fulvus] (41/59)	AF161185_3
FamB4	360	Rieske oxygenase	PrnD [Myxococcus fulvus] (44/62)	AF161185_3
FamB5	359	Rieske oxygenase	PrnD [Myxococcus fulvus] (43/63)	AF161185_3
FamD1	322	aromatic prenyltransferases	Orf2 [Clostridium clariflavum DSM 19732] (32/52)	YP_005045476.1
FamE1	412	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	4-hydroxy-3-methylbut-2-enyl diphosphate reductase [Fischerella sp. PCC 9339] (96/98)	WP_017308629.1
FamE2	346	polyprenyl synthetase	polyprenyl synthetase [Cyanothece sp. PCC 7425] (70/86)	YP_002481719.1
FamF1	232	LuxR family transcriptional regulator	LuxR family transcriptional regulator [Saccharomonospora xinjiangensis] (33/53)	WP_006236595.1
FamF2	244	LuxR family transcriptional regulator	LuxR family transcriptional regulator [Crinalium epipsammum PCC 9333] (62/74)	YP_007141135.1
FamG1	192	dephospho-CoA kinase-like protein	dephospho-CoA kinase-like protein [Odoribacter splanchnicus DSM 20712] (45/66)	YP_004252647.1
FamE3	406	1-deoxy-D-xylulose 5-phosphate reductoisomerase	1-deoxy-D-xylulose 5-phosphate reductoisomerase [Moorea producens 3L] (81/90)	ZP_08426551.1
FamH1	319	isonitrile synthase	PvcA protein [Cystobacter fuscus] (48/69)	WP_002632709.1
FamH2	330	isonitrile synthase	PvcA protein [Cystobacter fuscus] (53/68)	WP_002632709.1

FamH3	273	isonitrile synthase	PvcB protein [Cystobacter fuscus] (53/69)	WP_002632708.1
FamD2	309	aromatic prenyltransferase	Orf2 [Clostridium clariflavum DSM 19732] (34/51)	YP_005045476.1
FamE4	647	1-deoxy-D-xylulose-5-phosphate synthase	1-deoxy-D-xylulose-5-phosphate synthase [Fischerella sp. PCC 9339] (99/99)	WP_017308617.1
FamE5	408	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [Fischerella sp. PCC 9339] (100/100)	WP_017308616.1
FamI1	301	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase [Chroococcidiopsis thermalis PCC 7203] (76/85)	YP_007092480.1
FamI2	734	anthranilate synthase subunit I	anthranilate synthase subunit I [Mastigocladopsis repens] (72/82)	WP_017317604.1
FamI3	283	indole-3-glycerol-phosphate synthase	indole-3-glycerol-phosphate synthase [Mastigocladopsis repens] (68/83)	WP_017317605.1
FamI4	270	tryptophan synthase subunit alpha	tryptophan synthase subunit alpha [Rivularia sp. PCC 7116] (69/83)	YP_007057474.1
FamI5	416	tryptophan synthase subunit beta	tryptophan synthase subunit beta [Fischerella sp. PCC 9339] (98/99)	WP_017308611.1
FamI6	365	anthranilate phosphoribosyl transferase	anthranilate phosphoribosyltransferase [Rivularia sp. PCC 7116] (67/80)	YP_007057471.1
FamG2	214	phosphoglycerate mutase	phosphoglycerate mutase [Propionibacterium acnes] (39/56)	WP_002548864.1

SI Table 3-3. Function annotation of ORFs in the *fil* biosynthetic gene cluster. The Genbank accession number for the *fil* gene cluster is KY026488.

Protein	Size (aa)	Proposed function	Protein homolog (identity/positives %/%)
FilA1	151	hypothetical protein	FamA1[b] (100/100)
FilA2	397	DevB family ABC transporter membrane fusion protein	FamA2[b] (99/100)
FilA3	388	ABC transporter	FamA3[b] (99/100)
FilB1	316	2OG-Fe(II) oxygenase	FamB1[b] (99/99)
FilB2	285	Rieske oxygenase	FamB2[b] (100/100)
FilC1	227	Stig cyclase	FamC1[b] (94/96)
FilB3	362	Rieske oxygenase	FamB3[b] (68/80)
FilC2	230	Stig cyclase	FamC2[b] (98/98)
FilC3	233	Stig cyclase	FamC3[b] (96/96)
FilC4	228	Stig cyclase	FamC4[b] (98/99)
FilB4	363	Rieske oxygenase	FamB3[b] (98/99)
FilB5	360	Rieske oxygenase	FamB3[b] (73/84)
FilB6	359	Rieske oxygenase	FamB3[b] (79/85)
FilD1	343	aromatic prenyltransferases	FamD1[b] (100/100)
FilE1	412	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	FamE1[b] (100/100)
FilE2	346	polyprenyl synthetase	FamE2[b] (100/100)
FilF1	238	LuxR family transcriptional regulator	FamF1[b] (100/100)
FilF2	244	LuxR family transcriptional regulator	FamF2[b] (100/100)
FilG1	192	dephospho-CoA kinase-like protein	FamG1[b] (99/100)
FilE3	406	1-deoxy-D-xylulose 5-phosphate reductoisomerase	FamE3[b] (99/99)
FilH1	319	isonitrile synthase	FamH1[b] (98/99)
FilH2	331	isonitrile synthase	FamH2[b] (98/98)
FilH3	273	isonitrile synthase	FamH3[b] (99/100)
FilD2	309	aromatic prenyltransferase	FamD2[b] (99/100)
FilE4	647	1-deoxy-D-xylulose-5-phosphate synthase	FamE4[b] (99/99)
FilE5	408	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	FamE5[b] (100/100)
FilI1	301	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	FamI1[b] (100/100)
FilI2	734	anthranilate synthase subunit I	FamI2[b] (100/100)
FilI3	283	indole-3-glycerol-phosphate synthase	FamI3[b] (100/100)
FilI4	270	tryptophan synthase subunit alpha	FamI4[b] (100/100)
FilI5	415	tryptophan synthase subunit beta	FamI5[b] (99/99)
FilI6	365	anthranilate phosphoribosyltransferase	FamI6[b] (99/100)
FilG2	214	phosphoglycerate mutase	FamG2[b] (98/98)

[b] Fischerella ambigua UTEX 1903

SI Table 3-4. Function annotation of ORFs in the *fim* biosynthetic gene cluster. The Genbank accession number for the *fim* gene cluster is KY026487.

Protein	Size (aa)	Proposed function	Protein homolog (identity/positives %/%)
FimA1	151	hypothetical protein	FamA1[b] (100/100)
FimA2	397	DevB family ABC transporter membrane fusion protein	FamA2[b] (99/100)
FimA3	388	ABC transporter	FamA3[b] (98/99)
FimB1	290	2OG-Fe(II) oxygenase	FamB1[b] (98/99)
FimB2	285	Rieske oxygenase	FamB2[b] (100/100)
FimC1	227	Stig cyclase	FamC1[b] (100/100)
FimC2	227	Stig cyclase	FamC2[b] (100/100)
FimC3	143	Stig cyclase	FamC3[b] (100/100 within the first 138aa)
FimC4	228	Stig cyclase	FamC4[b] (97/98)
FimJ	336	reductase	dihydrodipicolinate reductase [Tepidanaerobacter acetatoxydans] (50/70)[a]
FimB3	358	Rieske oxygenase	FamB3[b] (71/81)
FimC5	228	Stig cyclase	FamC2[b] (70/80)
FimE1	412	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	FamE1[b] (96/98)
FimE2	346	polyprenyl synthetase	FamE2[b] (100/100)
FimF1	232	LuxR family transcriptional regulator	FamF1[b] (100/100)
FimF2	244	LuxR family transcriptional regulator	FamF2[b] (100/100)
FimG1	192	dephospho-CoA kinase-like protein	FamG1[b] (100/100)
FimE3	406	1-deoxy-D-xylulose 5-phosphate reductoisomerase	FamE3[b] (100/100)
FimH1	319	isonitrile synthase	FamH1[b] (100/100)
FimH2	331	isonitrile synthase	FamH2[b] (99/99)
FimH3	272	isonitrile synthase	FamH3[b] (99/100)
FimD	309	aromatic prenyltransferase	FamD2[b] (100/100)
FimE4	647	1-deoxy-D-xylulose-5-phosphate synthase	FamE4[b] (99/99)
FimE5	408	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	FamE5[b] (100/100)
FimI1	301	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	FamI1[b] (99/99)
FimI2	734	anthranilate synthase subunit I	FamI2[b] (96/97)
FimI3	283	indole-3-glycerol-phosphate synthase	FamI3[b] (94/97)
FimI4	270	tryptophan synthase subunit alpha	FamI4[b] (97/98)
FimI5	415	tryptophan synthase subunit beta	FamI5[b] (98/99)
FimI6	365	anthranilate phosphoribosyltransferase	FamI6[b] (99/99)
FimG2	214	phosphoglycerate mutase	FamG2[b] (97/99)

[a] Accession no. WP_013777407.1 [b] Fischerella ambigua UTEX 1903

SI Table 3-5. Function annotation of ORFs in the *fis* biosynthetic gene cluster. The Genbank accession number for the *fis* gene cluster is KY026489.

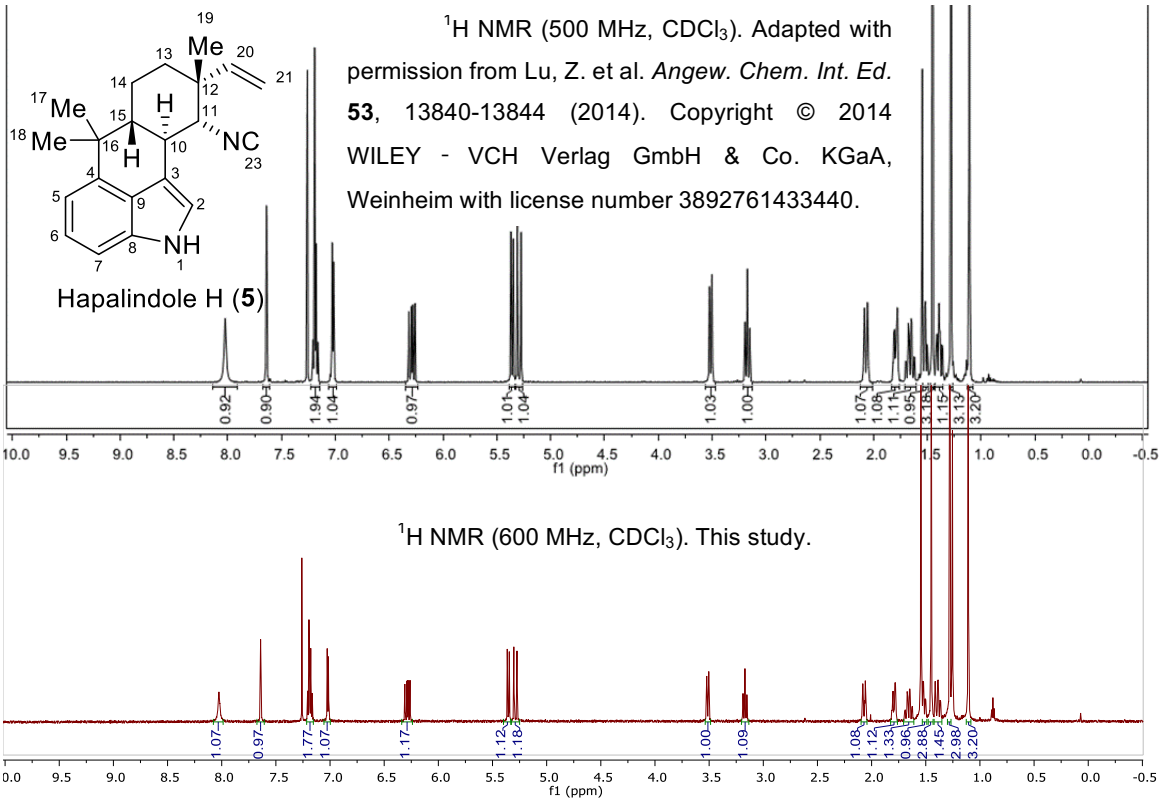
Protein	Size (aa)	Proposed function	Protein homolog (identity/positives %/%)
FisB1	290	2OG-Fe(II) oxygenase	FamB1[b] (81/90)
FisB2	356	Rieske oxygenase	FamB3[b] (68/78)
FisC	228	Stig cyclase	FamC2[b] (70/79)
FisE1	400	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	FamE1[b] (92/96)
FisE2	353	polyprenyl synthetase	FamE2[b] (88/93)
FisM	254	methyltransferase	SAM-dependent methyltransferase [Rhodococcus] (29/52)[a]
FisG1	192	dephospho-CoA kinase-like protein	FamG1[b] (87/92)
FisE3	406	1-deoxy-D-xylulose 5-phosphate reductoisomerase	FamE3[b] (95/97)
FisH1	319	isonitrile synthase	FamH1[b] (89/96)
FisH2	348	isonitrile synthase	FamH2[b] (96/98)
FisH3	273	isonitrile synthase	FamH3[b] (92/97)
FisD	308	aromatic prenyltransferase	FamD2[b] (96/97)
FisE4	648	1-deoxy-D-xylulose-5-phosphate synthase	FamE4[b] (96/98)
FisE5	408	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	FamE5[b] (99/99)
FisB3	358	Rieske oxygenase	FamB3[b] (69/80)
FisB4	486	monooxygenase	cytochrome P450 [Fischerella muscicola] (100/100)[c]
FisI1	301	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	FamI1[b] (98/99)
FisI2	734	anthranilate synthase subunit I	FamI2[b] (94/96)
FisI3	283	indole-3-glycerol-phosphate synthase	FamI3[b] (90/96)
FisI4	274	tryptophan synthase subunit alpha	FamI4[b] (92/94)
FisI5	415	tryptophan synthase subunit beta	FamI5[b] (96/97)
FisI6	365	anthranilate phosphoribosyltransferase	FamI6[b] (96/97)
FisG2	214	phosphoglycerate mutase	FamG2[b] (96/97)

[a] Accession no. WP_005260767.1; [b] Fischerella ambigua UTEX 1903; [c] Accession no. WP_033378141.1.

SI Table 3-6. The complete NMR spectroscopic data table for hapalindole H (**5**) in benzene-d6, and the ^{1}H NMR spectrum overlay with literature data[23] in $CDCl_3$. HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2007. $[\alpha]_D$ +52.9 (CH_2Cl_2, *c* 0.05).

Position	δ_{13C}	δ_{1H} (multi, *J* value)	COSY	HMBC	NOESY
1		6.76 (bs)	2		2
2	118.5	7.62 (dd, *J* = 1.4, 2.2 Hz)	1, 10	3, 8, 9	1
3	113.6				
4	140.9				
5	112.7	7.05 (d, *J* = 7.2 Hz)	6	6, 9, 16	17, 18
6	122.8	7.26 (t, *J* = 7.8 Hz)	5,7	4, 5, 8	
7	108.3	6.96 (d, *J* = 8.0 Hz)	6	5, 9	
8	133.9				
9	125.8				
10	36.5	3.14-3.06 (m)	2, 15	3, 11, 12, 15	20
11	67.6	3.12-3.10 (m)		10, 20, 19	
12	40.7				
13	36.0	(Heq) 1.62 (td, *J* = 3.4, 13.8 Hz) (Hax) 0.81 (dt, *J* = 5.6, 12.7 Hz)	14	12, 14, 20, 19	
14	20.9	1.35-1.30 (m)	13, 15		
15	49.6	1.13-1.07 (m)	10, 14		
16	37.4				
17	24.7	0.96 (s)		4, 15, 16, 17	
18	24.5	1.29 (s)		4, 15, 16, 18	
19	27.3	1.09 (s)		11, 12, 13, 20	13ax
20	138.9	6.18 (dd, *J* = 11.1, 17.6 Hz)	21	11, 12, 13, 19	10, 13eq, 14, 19
21	115.5	(cis) 5.14 (dd, *J* = 11.1, 1.3 Hz) (trans) 5.09 (dd, *J* = 17.6, 1.4 Hz)	20	12, 20	13eq, 14ax, 19
23	161.2				

NMR splitting coefficients are reported a s = singlet, d = doublet, t = triplet, m = multiplet, b = broad



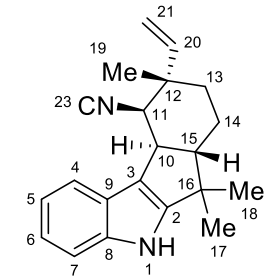
^{1}H NMR (500 MHz, $CDCl_3$). Adapted with permission from Lu, Z. et al. *Angew. Chem. Int. Ed.* **53**, 13840-13844 (2014). Copyright © 2014 WILEY - VCH Verlag GmbH & Co. KGaA, Weinheim with license number 3892761433440.

^{1}H NMR (600 MHz, $CDCl_3$). This study.

SI Table 3-7. The complete NMR spectroscopic data table for 12-*epi*-fischerindole U (**9**) in benzene-d6 and the ^{1}H NMR spectrum overlay with the literature data[22] in CD_2Cl_2. HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2012. $[\alpha]_D$ +59.8 (CH_2Cl_2, *c* 0.1).

Position	δ_{13C}	δ_{1H} (multi, J value)	COSY	HMBC	NOESY
1	N	6.54			7, 17
2	151.8				
3	115				
4	118.8	7.54 (m)	5, 6, 7	3, 5, 8, 9	11
5	121.1	7.24 (m, 2H)	4, 7	4, 7, 8, 9	
6	120.4				
7	112.2	7.11 (m)	4, 5, 6	6, 8, 9	1
8	140.2				
9	124.6				
10	42.55	2.97 (dtd, J = 10.6, 3.5, 3.3 Hz)	11, 15	3	11, 14a, 18, 20, 21
11	62.6 (t, J = 5.6 Hz)	4.13 (d, J = 1.9 Hz)	10	10, 12, 13, 15, 20, 23	4, 10, 19, 21, 19
12	41.1				
13	31.66	1.62-1.55 (m)	14a, 14b	11, 12, 14, 15, 20, 19	15, 21, 19
14a	20.4	1.45-1.33 (m)	15, 13a/b, 14b	10, 13, 15	10, 18, 20
14b		1.30-1.23 (m)	13a/b, 14a, 15		15, 18
15	55.1	2.41 (ddd, J = 13.2, 10.5, 3.3 Hz)	10, 14a, 14b		13, 14b, 17
16	40.1				
17	24.87	1.05 (3H)	18	2, 15, 16, 18	1 (NH), 15, 18
18	20.58	0.78 (3H)	17	2, 15, 16, 17	10, 14a, 14b, 17
19	27.96	1.14 (3H)		11, 12, 13, 20	11, 13, 20, 21
20	143	5.34 (dd, J = 17.8, 10.7 Hz)	21a, 21b	11, 12, 13, 19	10, 11, 13, 14a, 19
21a	114.1	4.88 (dd, J = 10.7, 0.8 Hz)	20	12, 20	10, 11, 13, 14a, 19
21b		4.87 (dd, J = 17.8, 0.8 Hz)			
23	161.2				

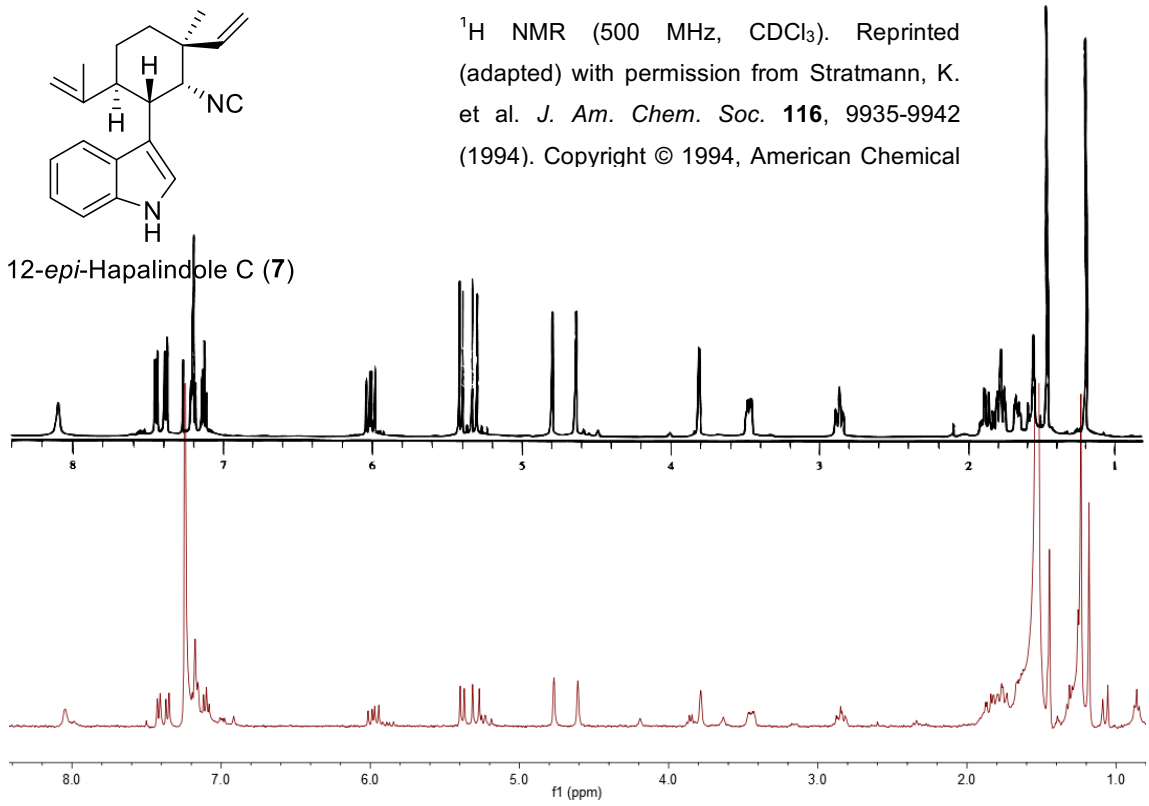
NMR splitting coefficients are reported a s = singlet, d = doublet, t = triplet, m = multiplet, b = broad



12-*epi*-Fischerindole U (**9**)

^{1}H NMR (500 MHz, CD_2Cl_2). Reprinted (adapted) with permission from Stratmann, K. et al. *J. Am. Chem. Soc.* **116**, 9935-9942 (1994). Copyright © 1994, American Chemical

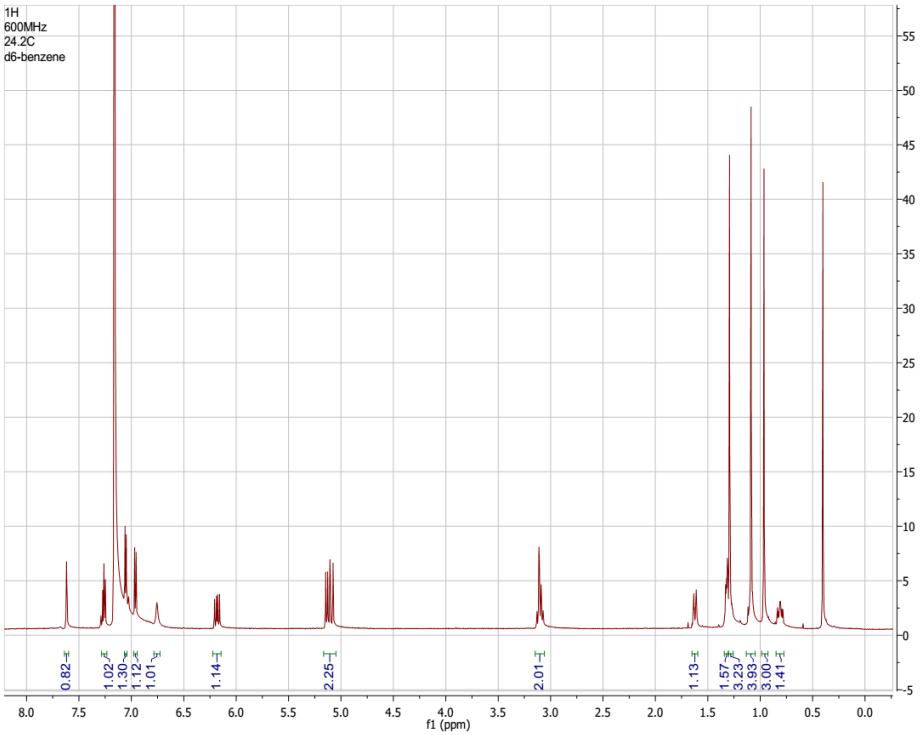
^{1}H NMR (400 MHz, CD_2Cl_2). This

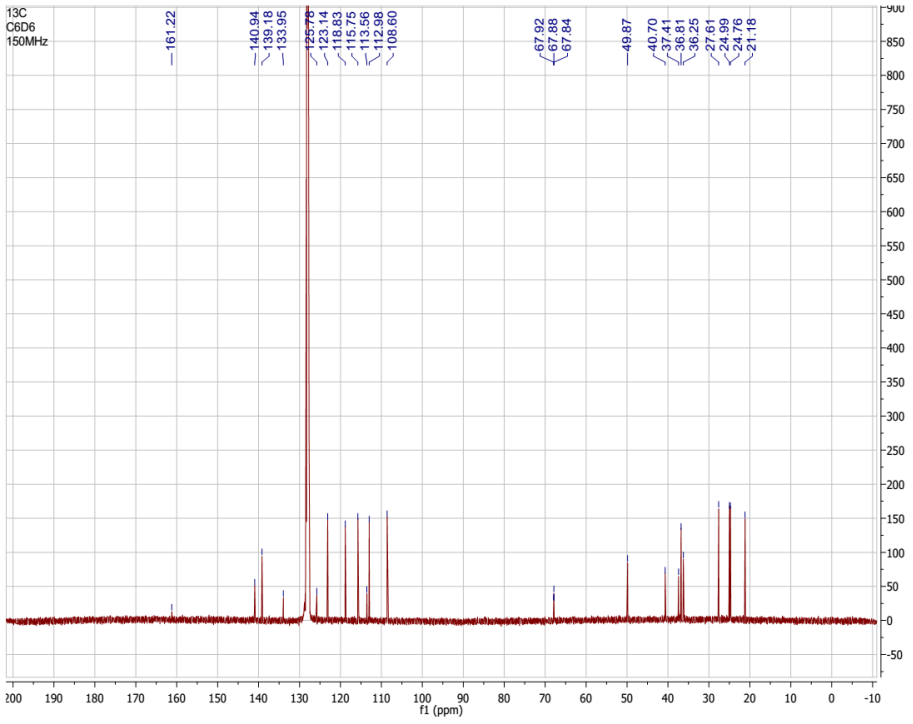


SI Figure 3-5. ^{1}H NMR spectrum of 12-*epi*-hapalindole C (**7**) in CD_2Cl_2 at 400 MHz compared to the published spectrum.[22] HRMS: Calcd for $C_{21}H_{24}N_2$ $[M+H]^+$ 305.2012, found 305.2008.

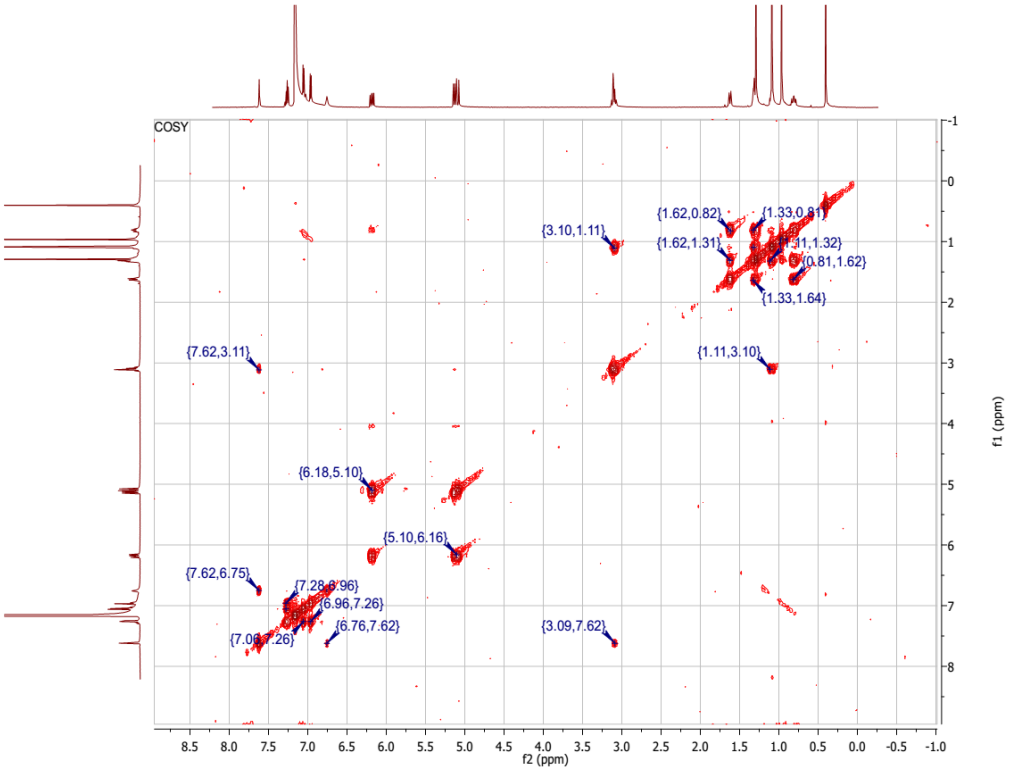
3.7 Spectra section

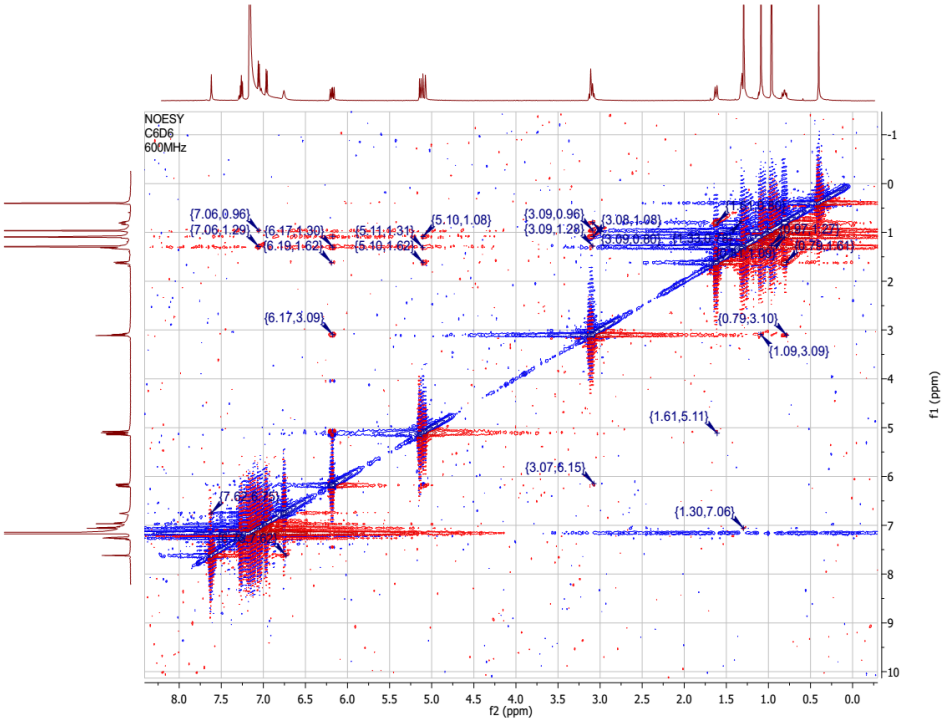
^{1}H and ^{13}C NMR spectra of hapalindole H (**5**) in benzene-d6 at 600/150 MHz.



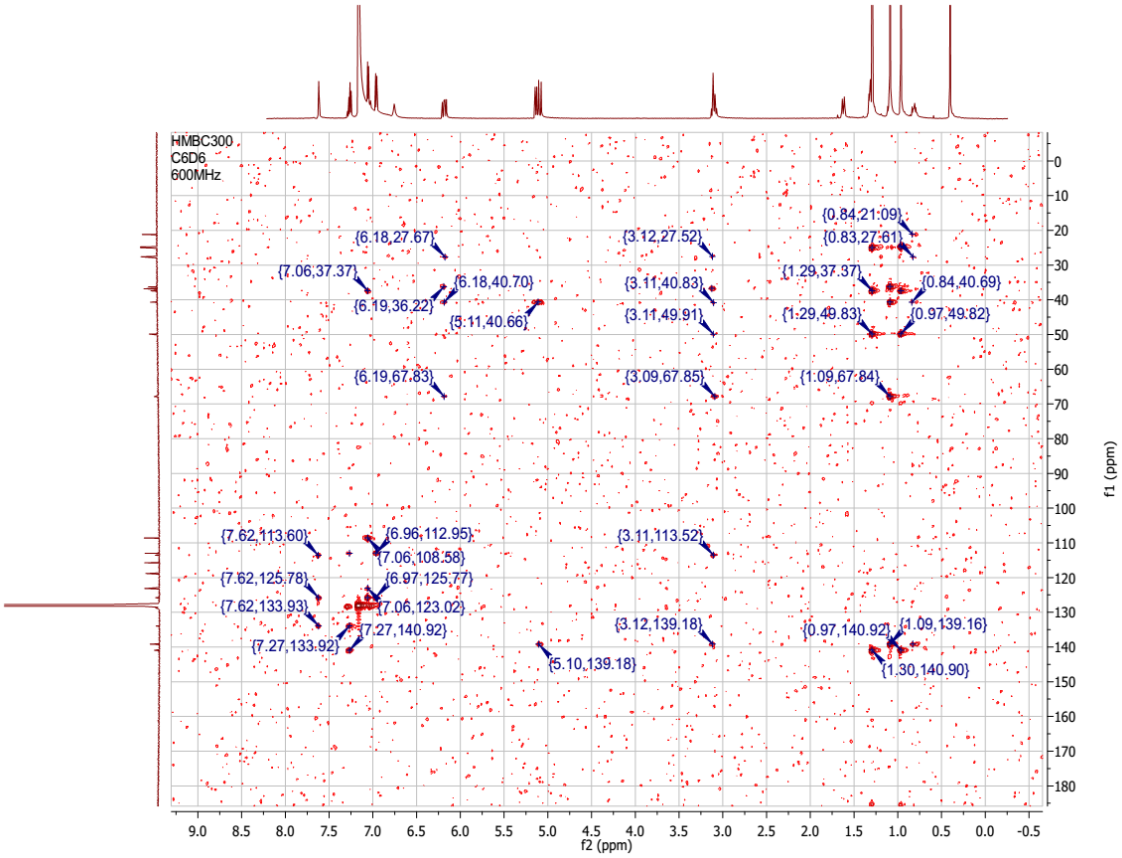


COSY and NOESY NMR spectra of hapalindole H (**5**) in benzene-d6 at 600 MHz.

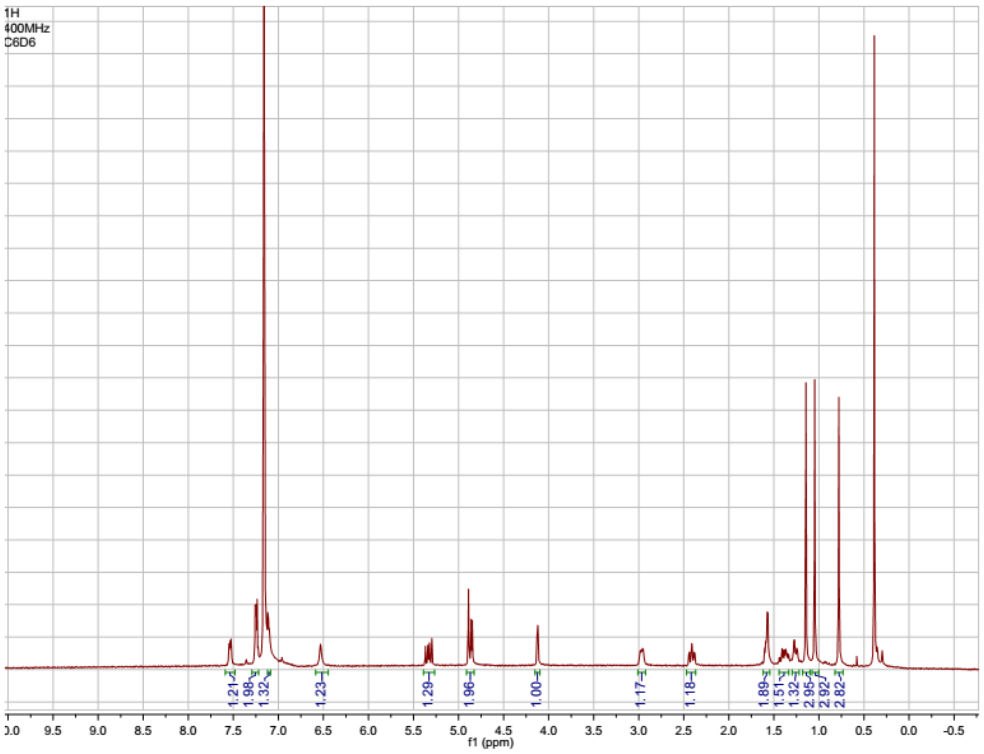


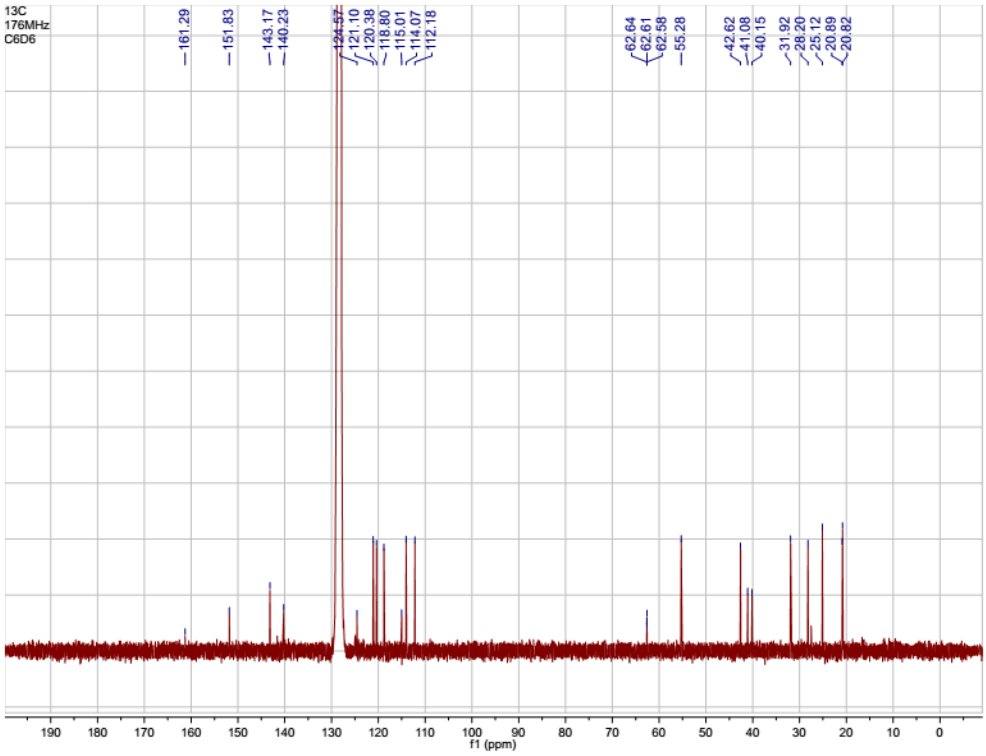


HMBC NMR spectrum of hapalindole H (**5**) in benzene-d6 at 600 MHz.

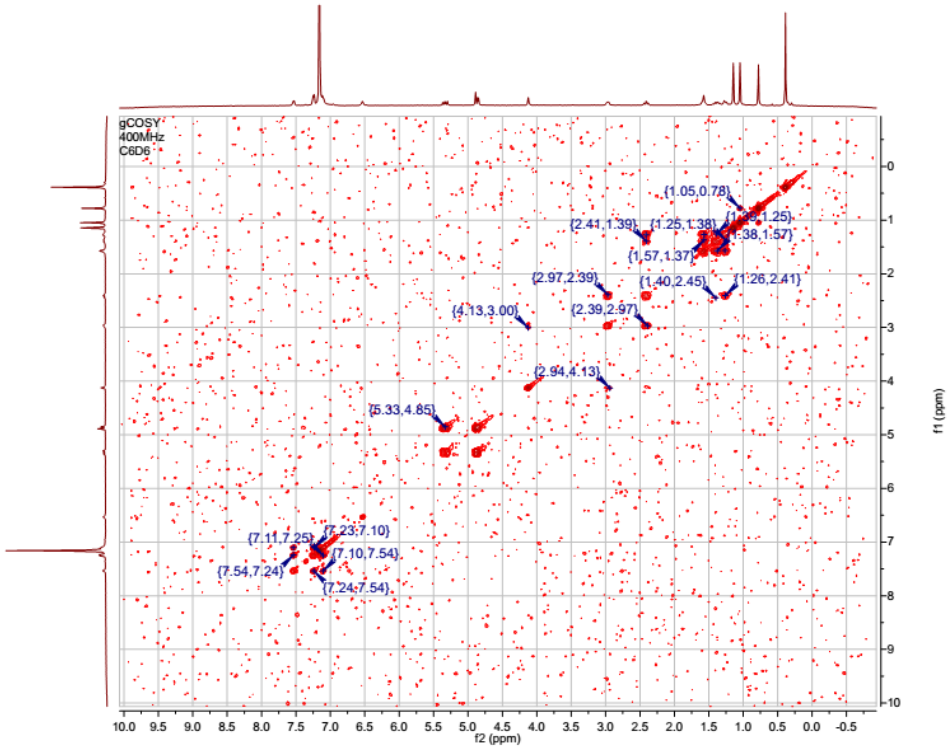


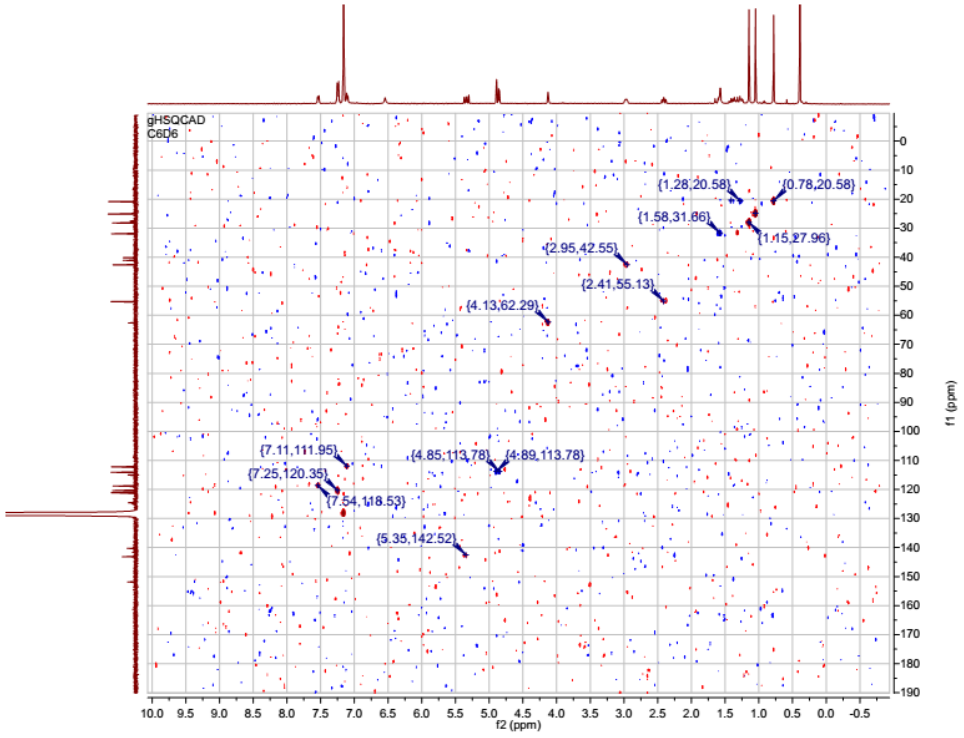
^{1}H and ^{13}C NMR spectra of 12-*epi*-fischerindole U (**9**) in benzene-d6 at 400/100 MHz.



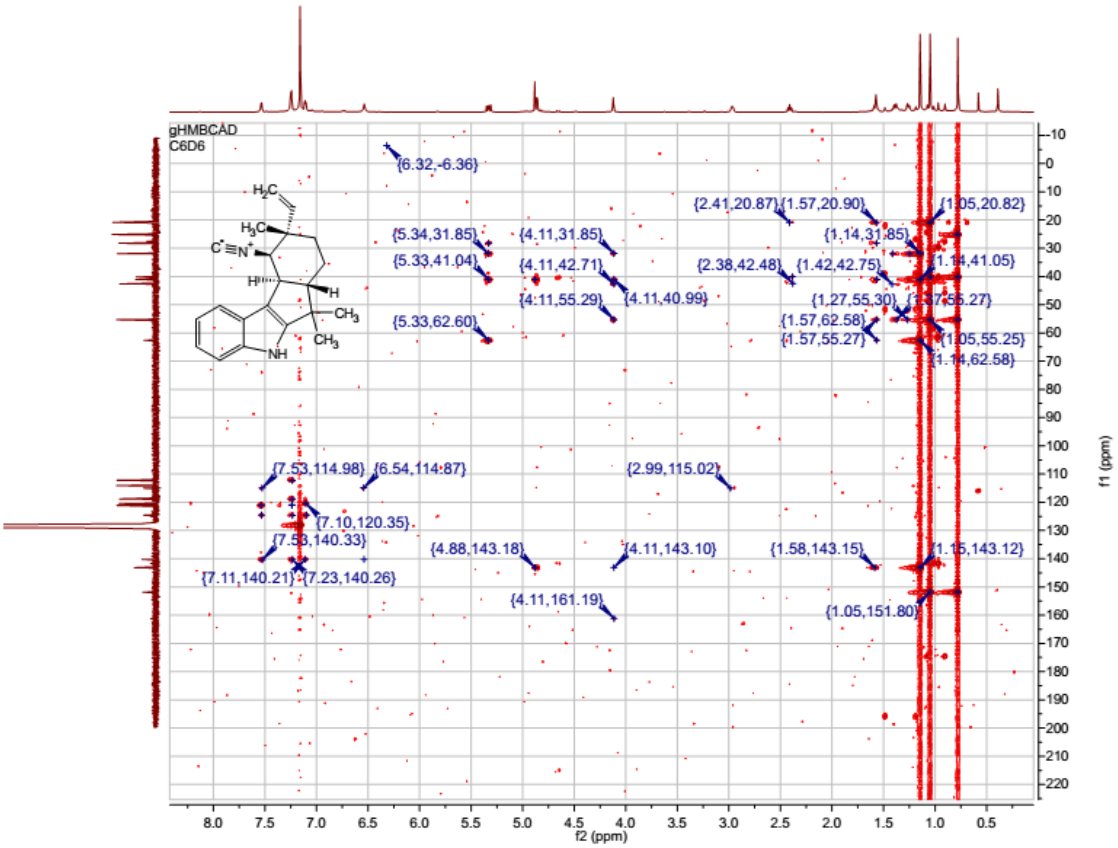


COSY and HSQC NMR spectrum of 12-*epi*-fischerindole U (**9**) in benzene-d6 at 400 MHz.





HMBC NMR spectrum of 12-*epi*-fischerindole U (**9**) in benzene-d6 at 700 MHz.



3.8 References

1. Li, C. J., Organic reactions in aqueous media with a focus on carbon-carbon bond formations: a decade update. *Chem. Rev.* **2005,** *105* (8), 3095-3165.
2. Koeller, K. M.; Wong, C. H., Enzymes for chemical synthesis. *Nature* **2001,** *409* (6817), 232-240.
3. Tsunematsu, Y.; Ishikawa, N.; Wakana, D.; Goda, Y.; Noguchi, H.; Moriya, H.; Hotta, K.; Watanabe, K., Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.* **2013,** *9* (12), 818-825.
4. Sanchez, C.; Mendez, C.; Salas, J. A., Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. *Nat. Prod. Rep.* **2006,** *23* (6), 1007-1045.
5. Jakubczyk, D.; Cheng, J. Z.; O'Connor, S. E., Biosynthesis of the ergot alkaloids. *Nat. Prod. Rep.* **2014,** *31* (10), 1328-1338.
6. Fesko, K.; Gruber-Khadjawi, M., Biocatalytic Methods for C-C Bond Formation. *Chemcatchem* **2013,** *5* (6), 1248-1272.
7. Miao, Y.; Rahimi, M.; Geertsema, E. M.; Poelarends, G. J., Recent developments in enzyme promiscuity for carbon-carbon bond-forming reactions. *Curr. Opin. Chem. Biol.* **2015,** *25*, 115-123.
8. Klas, K.; Tsukamoto, S.; Sherman, D. H.; Williams, R. M., Natural Diels-Alderases: Elusive and Irresistable. *J. Org. Chem.* **2015,** *80* (23), 11672-11685.
9. Fage, C. D.; Isiorho, E. A.; Liu, Y. N.; Wagner, D. T.; Liu, H. W.; Keatinge-Clay, A. T., The structure of SpnF, a standalone enzyme that catalyzes [4+2] cycloaddition. *Nat. Chem. Biol.* **2015,** *11* (4), 256-258.
10. Auclair, K.; Sutherland, A.; Kennedy, J.; Witter, D. J.; Van den Heever, J. P.; Hutchinson, C. R.; Vederas, J. C., Lovastatin nonaketide synthase catalyzes an intramolecular Diels-Alder reaction of a substrate analogue. *J. Am. Chem. Soc.* **2000,** *122* (46), 11519-11520.
11. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., Hapalindole/Ambiguine biogenesis is mediated by a Cope rearrangement, C–C bond-forming cascade. *J. Am. Chem. Soc.* **2015,** *137* (49), 15366-15369.
12. Bhat, V.; Dave, A.; MacKay, J. A.; Rawal, V. H., Chapter two - the chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. In *The Alkaloids: Chemistry and Biology*, Hans-Joachim, K., Ed. Academic Press: San Diego, 2014; Vol. 73, pp 65-160.
13. Baran, P. S.; Maimone, T. J.; Richter, J. M., Total synthesis of marine natural products without using protecting groups. *Nature* **2007,** *446* (7134), 404-408.
14. Micallef, M. L.; Sharma, D.; Bunn, B. M.; Gerwick, L.; Viswanathan, R.; Moffitt, M. C., Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol.* **2014,** *14*, 213-230.

15. Kim, H.; Krunic, A.; Lantvit, D.; Shen, Q.; Kroll, D. J.; Swanson, S. M.; Orjala, J., Nitrile-containing fischerindoles from the cultured cyanobacterium *Fischerella* sp. *Tetrahedron* **2012,** *68* (15), 3205-3209.
16. Raveh, A.; Carmeli, S., Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* **2007,** *70* (2), 196-201.
17. Smitka, T. A.; Bonjouklian, R.; Doolin, L.; Jones, N. D.; Deeter, J. B.; Yoshida, W. Y.; Prinsep, M. R.; Moore, R. E.; Patterson, G. M. L., Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae. *J. Org. Chem.* **1992,** *57* (3), 857-861.
18. Park, A.; Moore, R. E.; Patterson, G. M. L., Fischerindole L, a new isonitrile from the terrestrial blue-green alga *Fischerella muscicola*. *Tetrahedron Lett.* **1992,** *33* (23), 3257-3260.
19. Wang, X.; Feng, J.; Xue, Y.; Guan, Z.; Zhang, D.; Liu, Z.; Gong, Z.; Wang, Q.; Huang, J.; Tang, C.; Zou, T.; Yin, P., Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **2016,** *534* (7608), 575-578.
20. Newmister, S. A.; Chan, C. H.; Escalante-Semerena, J. C.; Rayment, I., Structural insights into the function of the nicotinate mononucleotide:phenol/p-cresol phosphoribosyltransferase (ArsAB) enzyme from *Sporomusa ovata*. *Biochemistry* **2012,** *51* (43), 8571-8582.
21. Pyriochou, A.; Papapetropoulos, A., Soluble guanylyl cyclase: more secrets revealed. *Cell. Signal.* **2005,** *17* (4), 407-413.
22. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, unusual alkaloids from the blue-green algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to fischerindoles and hapalinodoles. *J. Am. Chem. Soc.* **1994,** *116* (22), 9935-9942.
23. Lu, Z.; Yang, M.; Chen, P.; Xiong, X.; Li, A., Total Synthesis of Hapalindole-Type Natural Products. *Angew. Chem. Int. Ed.* **2014,** *53* (50), 13840-13844.

Notes:

This work has been published as “Decoding Cyclase-Dependent Assembly of Hapalindole and Fischerindole Alkaloids.” Shasha Li, Andrew N Lowell, Sean A Newmister, Fengan Yu, Robert M Williams & David H Sherman. *Nat. Chem. Biol.* **2017**, *13*, 467-470.

Author contributions:

S.L. and D.H.S. designed the research. S.L. performed all experiments. A.N.L. synthesized the indole isonitrile. S.L., S.A.N., A.N.L. and D.H.S. conducted data analysis and interpretation. F.Y. performed bioinformatics analyses. S.L., A.N.L., D.H.S., and R.M.W. contributed to manuscript preparation.

Chapter 4
Control of stereoselectivity in diverse hapalindole/fischerindole metabolites is mediated by cofactor induced combinatorial pairing of Stig cyclases

4.1 Abstract

In Chapter 2 and 3, we disclosed a novel class of Stig cyclases responsible for the hapalindoles and fischerindole formation. A mechanism was proposed for this reaction comprising by a Cope rearrangement, 6-exo-trig cyclization and electrophilic aromatic substitution. Based on the protein-protein interaction study of FamC2 and FamC3, we also proposed that these Stig cyclases are functional as homodimer or heterodimer. In this chapter, I will describe our continuous work toward all annotated Stig cyclases. Through in vitro analysis, mutagenesis and information obtain from crystal structure study (will be described in Chapter 5), this work revealed a more complex mechanism of the Stig cyclases in conducting the polycyclic core formation, that the stereochemistry is controlled by Ca^{2+} promoted combinatorial pairing.

4.2 Introduction

Hapalindole-type alkaloids are a large group of secondary metabolites isolated from terrestrial and freshwater Hapalosiphonaceae[1] cyanobacteria. Their diverse pharmacological activity and intriguing polycyclic skeletons have compelled us to pursue a detailed mechanistic understanding of their biosynthesis.[2-5] Our hypothesis is that biogenesis of the hapalindole or fischerindole core structures with variant stereochemical patterns at four chiral

centers is followed by late-stage tailoring reactions to generate more complex ambiguine and welwitindolinone subgroups (**SI Figure 4-1**).[6] All reported hapalindoles and fischerindoles can be categorized into six stereochemical patterns based on the C10, C11, C12, and C15 chiral centers (**Figure 4-1**). This stereo- and regiochemically diverse polycyclic ring formation represents a fascinating platform to investigate variations during initial cyclizations and subsequent tailoring reactions. We recently reported a novel class of Stig cyclases responsible for catalyzing hapalindole and fischerindole formation from a common intermediate C3-geranylated indole isonitrile (**1**), whose coupled Cope rearrangement/ring-forming cascade together elaborates the four chiral centers and three types of ring systems.[4-5] Considerable progress has been made recently to identify the core assembly reactions, and to solve the first structure of a Stig cyclase.[7] However, further studies of these remarkable biocatalysts are required to fully understand the basis for stereochemical control during assembly of diverse hapalindole and fischerindole metabolites.

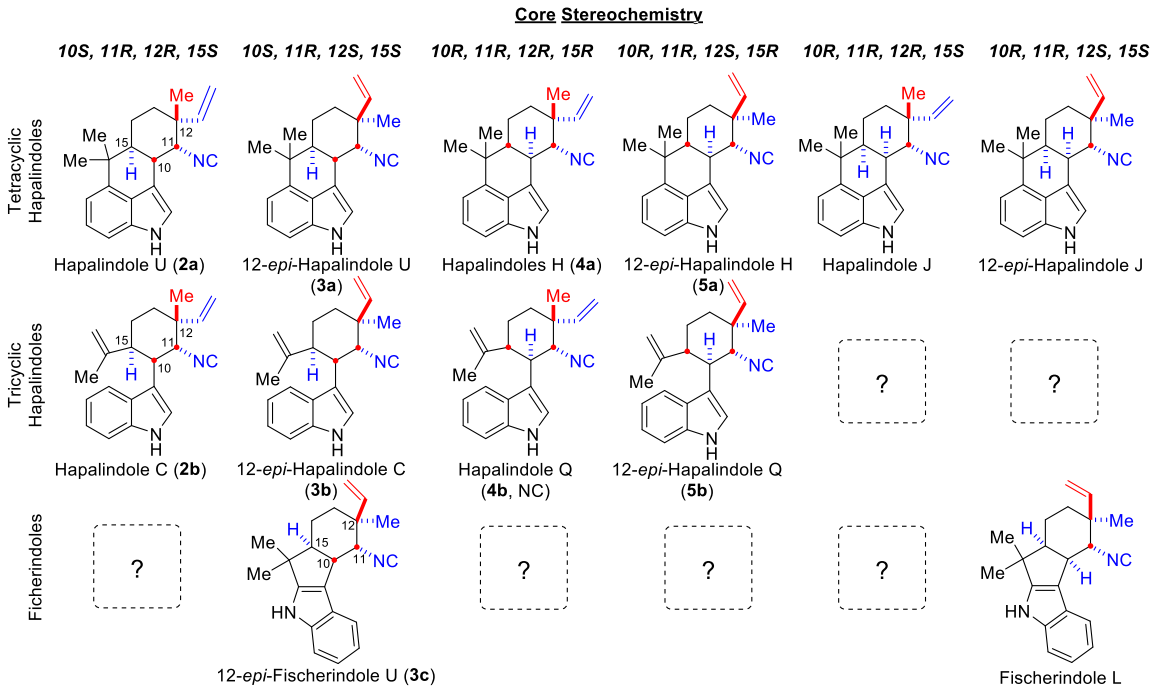


Figure 4-1. Stereochemical classification of isolated hapalindoles and fischerindoles based on the C10, C11, C12, and C15 stereocenters. Dash-line boxes with a question mark represent predicted structures that remain unidentified as natural products.

In a previous report,[5] we characterized several Stig cyclases, including FamC1 and its homologs FilC1/HpiC1 that generate 12-*epi*-hapalindole U (**3a**), FimC5/FisC that generate 12-*epi*-fischerindole U (**3c**), and the heterodimeric combination FamC2-FamC3 that produce hapalindole H (**4a**). These results provided a preliminary explanation for the stereoselectivity and regioselective divergence for C-ring coupling between hapalindoles and fischerindoles. We applied a phylogenetic analysis to predict the pattern of catalytic activity of Stig cyclases based on the hypothesis that homologous proteins with high % levels of sequence identity will generate the same or highly similar hapalindole-type metabolites. In addition, we demonstrated that Stig cyclases function as homodimeric (*e.g.* FamC1) or heterodimeric (*e.g.* FamC2-FamC3) complexes to generate distinct, stereochemically defined products. Finally, the cyclase reaction mechanism is composed of three steps including, 1) a Cope-rearrangement to introduce the C11 and C12 chiral centers, 2) a 6-*exo*-trig cyclization to establish

the C10 and C15 stereocenters, and 3) an electrophilic aromatic substitution (C-ring formation) or deprotonation (on D-ring) to generate tri- or tetracyclic hapalindoles or fischerindoles (**Figure 4-2**).

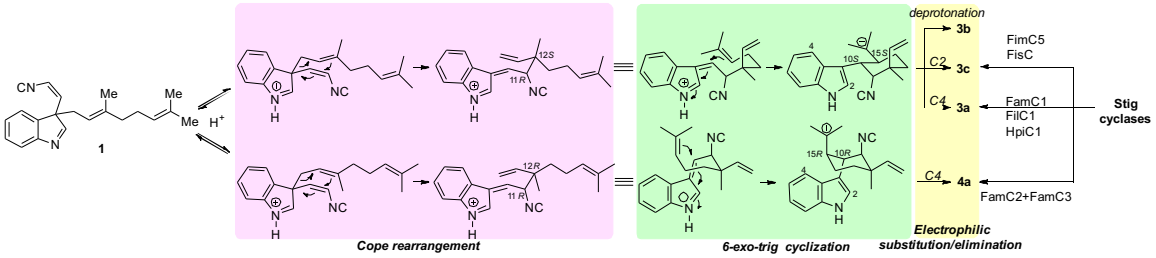


Figure 4-2. Proposed catalytical mechanism of Stig cyclases in forming tri- and tetracyclic hapalindoles and fischerindoles from the common intermediate **1** and summarization of identified Stig cyclases.

In the current study to decode the formation of alternative stereochemical patterns during hapalindole and fischerindole core construction, we reveal that Stig cyclases can display differential activities based on Ca^{2+} dependence and formation of higher order protein complexes.

4.3 Results and Discussion

New information arose as we continued to address the hypothesis that similar Stig proteins generate identical hapalindole or fischerindole core products. By comparing the annotated gene clusters and previously reported metabolites from individual cyanobacterial strains, we noticed circumstances where divergent stereochemical patterns of hapalindoles were found within related Stig cyclase subgroups (**Figure 4-3**), or cases where the same patterns of stereochemistry at the four chiral centers were derived from distinct cyclase subgroups (**Figure 4-4**). This analysis suggested that a more nuanced spectrum of molecular determinants is responsible for the pattern of chirality at C10, C11, C12, and C15, which motivated us to expand our studies of these fascinating biocatalysts.

4.3.1 12-*epi*-Hapalindole H producer FilC2-FilC3

To further dissect the basis for stereocontrol in Stig cyclase enzymes, we selected two ambiguine producing strains, *Fischerella ambigua* UTEX 1903 (*fam*)

and *Fischerella sp.* IL 199-3-1 (*fil*) (**SI Figure 4-2**).[5] Both *fam* and *fil* gene clusters contain four Stig cyclases (C1-C4) sharing high sequence similarities. FilC2 and FilC3 are 97.8% and 98.2% identical to FamC2 and FamC3, respectively, and are expected to produce hapalindole H according to our initial hypothesis. However, the isolated metabolites of IL 199-3-1 contains 12-*epi*-hapalindole H (**5a**, C10*R*, 11*R*, 12*S*, and 15*R*) instead of hapalindole H (**4a,** C10*R*, 11*R*, 12*R*, and 15*R*), which is isolated from the Fam system (**Figure 4-3**A and B, **SI Figure 4-3**).[8-10] Based on this metabolite profile, we decided to address whether FilC2-FilC3 behave differently from FamC2-FamC3 and produce **5a** instead of **4a** despite such high sequence identities. Thus, we heterologously expressed FilC2 and FilC3 in *E. coli* and analyzed the products of the *in vitro* reaction.

In this experimental system FilC2 and FilC3 were almost completely inactive both as a pair and as individual proteins, in contrast to FamC2 and FamC3 producing **4a** in a heterodimeric form.[5] Thus, we began screening for cofactors that might be required to promote the reaction after observing activity by introducing *Fischerella sp.* IL 199-3-1 cell-free lysate into the reaction mixture. We found that BG-11 culture medium contains the necessary cofactor; with high levels of calcium chloride and magnesium chloride promoting the reaction, and Ca^{2+} exhibiting higher efficiency than Mg^{2+} (**Figure 4-3**D, **SI Figure 4-5**). To characterize the structure of the FilC2 and FilC3 derived metabolites, we conducted a large-scale reaction using central intermediate **1** with 5 mM of $CaCl_2$ and isolated the compounds for NMR analysis. Under these conditions, the major product was identified as 12-*epi*-hapalindole H **5a** (**SI Figure 4-6** and **SI Table 4-2**), which is consistent with its isolation from IL 199-3-1, but surprising as these proteins (FilC2/FilC3) are 98% identical compared to the Fam (FamC2/FamC3) system that were shown previously to generate hapalindole H **4a**. Minor products were also produced, two of which we identified as 12-*epi*-hapalindole C (**3b**) and 12-*epi*-hapalindole Q (**5b**, **SI Figure 4-7**).

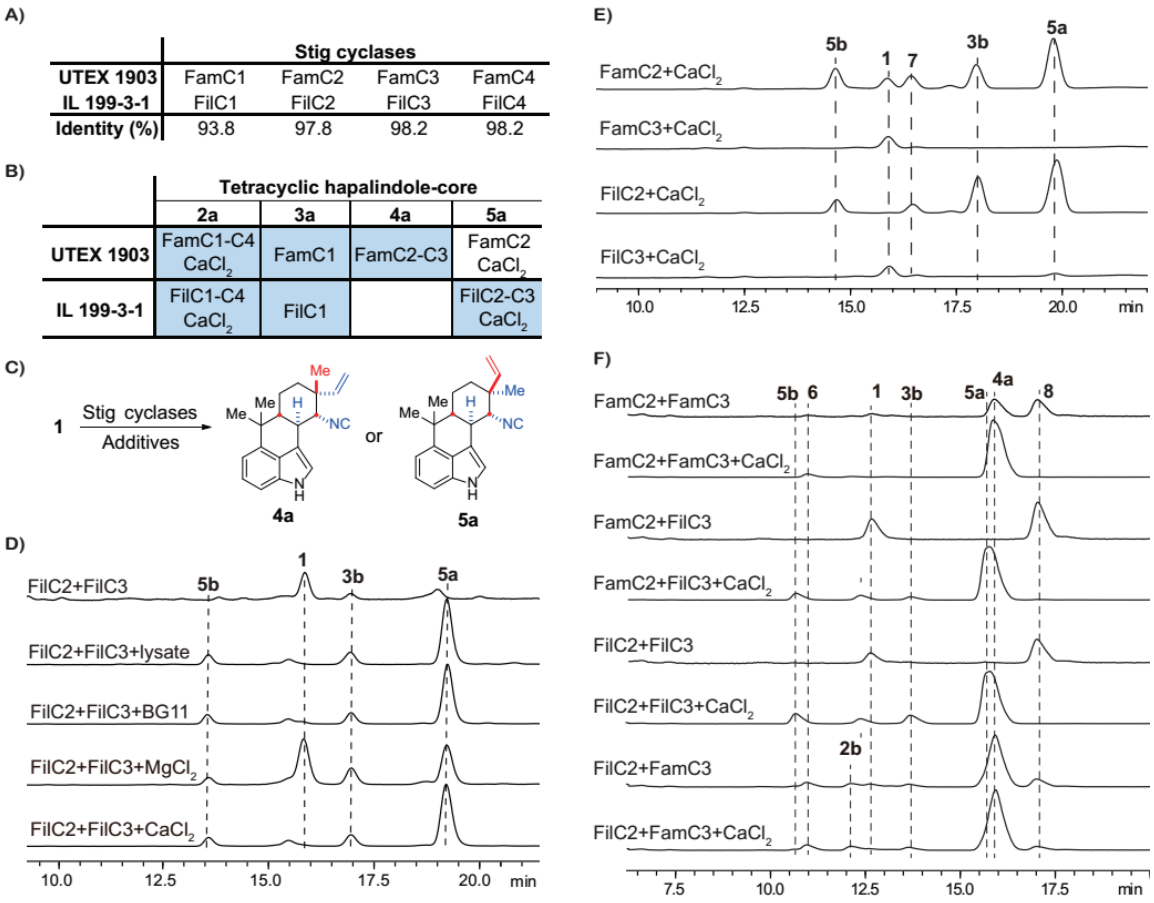


Figure 4-3. Information related to FilC2 and FilC3. A) Protein identities of the Stig cyclases from *Fischerella ambigua* UTEX 1903 and *Fischerella sp*. IL 199-3-1 strains. B) Stereochemical patterns of tetracyclic hapalindoles isolated from UTEX 1903 and IL 199-3-1 labeled in blue, and a summary of enzymatic products from their Stig cyclases. C) Main products generated in the in vitro assay with common intermediate **1** and Stig cyclases. D) HPLC traces of FilC2 and FilC3 with additives. E) Effect of $CaCl_2$ in promoting the activities of Stig cyclases. F) In vitro analysis of eterologous-pairing of Stig cyclases (structure of compound **8** see **SI Figure 4-1**).

The FilC2/FilC3 pair is the first example we investigated that requires supplemental Ca^{2+} to be functional and results in formation of different metabolites compared to its homologs that possess high sequence identity. To assess the effect of Ca^{2+} on other Stig cyclase reactions, we compared the products of FamC2-FamC3 and other previously characterized Stig cyclases[5] after introduction of excess Ca^{2+}, and found that these metabolite profiles remained unchanged. Furthermore, all reactions could be suppressed by adding EDTA. As demonstrated functionally by Liu et al[11]., assigned a direct role from

the x-ray structure of HpiC1 cyclase[7], and further assessed herein, these data suggested that Ca^{2+} is a critical structural component for these Stig cyclases. However, as addressed experimentally below, this co-factor appears unlikely to interact directly with the substrate to influence product profiles, but instead through promoting the protein-protein interaction.

We next tested the individual Stig cyclases with excess $Ca^{2+}$and found that this metal ion can activate FilC2 alone to produce the same metabolites observed when it is paired with FilC3, albeit at slightly lower levels. On the other hand, even with excess Ca^{2+}, FilC3 and FamC3 remained catalytically inactive in the absence of FilC2 and FamC2, respectively. Surprisingly, FamC2 alone generated 12-*epi*-hapalindole H (**5a**) instead of hapalindole H (**4a**), which is the product of heterodimeric FamC2/FamC3 cyclase (**Figure 4-3**E, **SI Figure 4-8**). This result suggests that FamC3 may actively control the stereochemistry of hapalindoles, in addition to serving as a solubilizing structural component for the heterodimeric FamC2/FamC3 complex. To test this hypothesis, we decided to investigate the heterologous pairing of the FamC2/FamC3 and FilC2/FilC3 Stig cyclases (**Figure 4-3**F). We found that pairing of FamC2-FilC3 resulted in production of 12-*epi*-hapalindole H **5a**, which is the product formed by pairing of FilC2-FilC3. Similarly, pairing of FilC2-FamC3 in vitro with intermediate **1** resulted in formation of hapalindole H **4a**, the same product identified from FamC2-FamC3, which is further characterized by ^{1}H NMR to be a 4:1 mixture of **4a** and **5a** (**SI Figure 4-9**). Two of the three minor products were confirmed to be hapalindole C (**2b**, **SI Figure 4-10**) and 12-*epi*-hapalindole C (**3b**).

4.3.2 Hapalindole U producer WepC1-WepC2

This product reconfiguration was also observed with proteins from the *wep* gene cluster, from ambiguine-producing strain *Westiellopsis prolifica* SAG 16.93. Two types of tetracyclic hapalindoles were isolated from this strain using traditional extraction methods, including hapalindole U and hapalindole H (**SI Figure 4-3**). Interestingly, only two Stig cyclases were identified from bioinformatics analysis of the *wep* gene cluster, including WepC1 and WepC2 (**SI**

Figure 4-4). WepC1 was initially classified as a FamC1 homolog based on its 89% sequence identity, while WepC2 is a FamC4 homolog with 93% similarity. Despite *Westiellopsis prolifica* SAG 16.93 being identified as a hapalindole H producer, there are no FamC2/FamC3 Stig cyclase homologs identified from *wep* or the broader genome based on extensive mining (**Figure 4-4**A and B). In vitro assays demonstrated that WepC1 behaves similarly to FamC1 by producing 12-*epi*-hapalindole U **3a** as the major metabolite in the presence of calcium. WepC1 is also able to produce low levels of tricyclic 12-*epi*-hapalindole C **3b** (10%). A major difference is the dependence of WepC1 on supplemental Ca^{2+} to be functional. By contrast, FamC1 does not require supplemental calcium for the cyclase reaction to proceed. In our initial screen for cyclase activity, WepC2 was inactive (similarly to FamC4), both with and without excess calcium. Thus, after analyzing both cyclases individually, neither of them generated the hapalindoles originally isolated from the *Westiellopsis prolifica* SAG 16.93 strain. Inspired by our previous finding that demonstrated the activity of heteromeric cyclase pairs (e.g. FamC2/FamC3), we combined WepC1 and WepC2 in 1:1 ratio with 5 mM of $CaCl_2$. In this experimental configuration, the product profile was completely altered. Instead of **3a**, the WepC1/WepC2 heteromeric combination produced three new metabolites (**Figure 4-4**D), characterized through NMR analysis to be hapalindole U (**2a**, C10*S*, 11*R*, 12*R*, and 15*S*), tricyclic hapalindole C **2b** (5:1 ratio), and trace level of hapalindole H **4a**.

Hapalindole U is a notable metabolite in this family by serving as the common precursor to all ambiguines except 12-*epi*-ambiguine B. Thus, every ambiguine-producing strain is expected to encode a Stig cyclase that catalyzes hapalindole U formation, including strain *Fischerella ambigua* UTEX 1903. Recently, Liu et al[11] described the formation of hapalindole U by combining FamC1-10xFamC4 with 20 mM of $CaCl_2$, which we confirmed using FamC1/FamC4 at a 1:1 ratio (5 mM $CaCl_2$). Beyond this pair, we tested the same conditions on the phylogenically homologous FilC1/FilC4 and FimC1/FimC4 and also observed hapalindole U formation (**Figure 4-4**D, **SI Figure 4-11**).

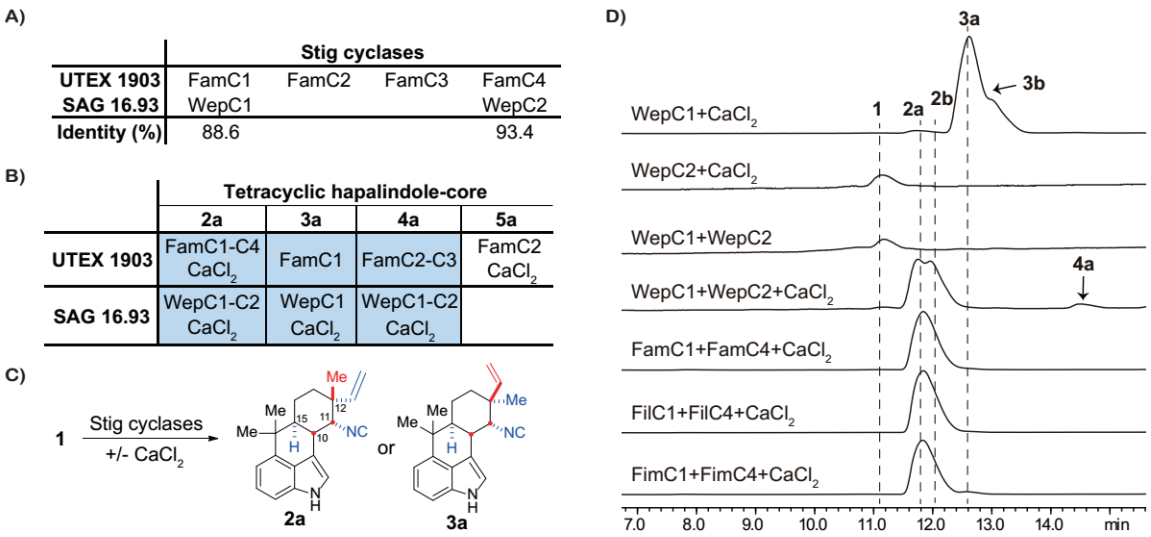


Figure 4-4. Information related to WepC1 and WepC2. A) Protein identities of the Stig cyclases from *Fischerella ambigua* UTEX 1903 and *Westiellopsis prolifica* SAG 16.93 strains. B) Stereochemical patterns of tetracyclic hapalindoles isolated from UTEX 1903 and SAG 16.93 labeled in blue, and a summarization of enzymatic products from their Stig cyclases. C) Main products generated in the in vitro assay with common intermediate **1** and Stig cyclases. D) HPLC traces of WepC1 and WepC2 and the homologs.

4.3.3 Hypothesis of oligomerization

In these two cases (FamC2/FamC3, FamC1/FamC4), we observed a distinct alteration in stereochemistry by adding another cyclase. Our initial hypothesis was that FamC3, FamC4, and their homologs were providing structural support for their cognate partners through protein-protein interactions as a heterodimeric assembly, and were not able to function as cyclases as single entities. However, we reasoned that another possibility is that the active sites of both cyclases are engaged to complete the reaction of a cognate cyclase partner. To evaluate this hypothesis, we further considered insights from the HpiC1 crystal structure[7] to guide mutations in the FamC cyclases. Thus, we decided to test the activity of the hypothesized dimeric assembly when one partner has been catalytically deactivated by site-directed mutagenesis at a key active site residue.

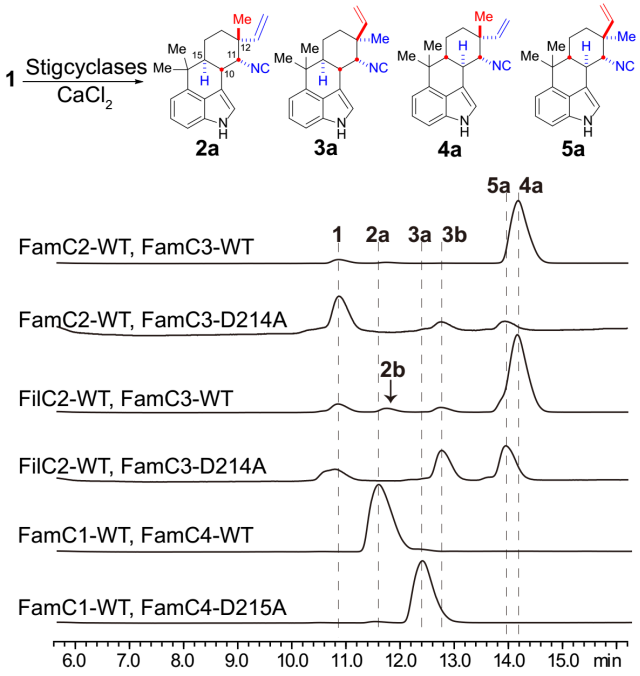


Figure 4-5. In vitro assay of FamC1, C2, C3, C4 mutants with substrate **1** and calcium chloride.

In the structural study of HpiC1,[7] we probed the critical role (source of acid catalysis) of residue D214, which is 100% conserved across all identified Stig cyclases. By mutating this residue, HpiC1 lost its activity, but is expected to retain its structural conformation. Accordingly, this conserved residue was mutated in four Stig cyclases, FamC1-D214A, FamC2-D217A, FamC3-D214A and FamC4-D215A. All the mutants were confirmed to be inactive either as an individual cyclase or as heterodimeric pairs. When FamC2-D217A was combined with FamC3-wild-type or FamC3-D214A, no products were obtained. Similar results were observed for the FamC1-D214A mutant combined with FamC4-WT or FamC4-D215A (**SI Figure 4-12**). These results indicated that FamC2 and FamC1 mediate an indispensable part of the catalytic reaction to form hapalindole alkaloids and are incapable of functioning in the absence of D214. We then analyzed the activity of the FamC3-D214A mutant by incubating with wild-type FamC2 or FilC2 (**Figure 4-5**). We found that the catalytically inactive FamC3 mutant was no-longer capable of altering the product profiles; generating **5a** in both cases, which is the expected result when an individual FamC2 or FilC2

cyclase is employed. With respect to FamC1/FamC4, the reaction with FamC1-WT and FamC4-D215A was also unable to generate **2a**, although **3a** continued to be produced (**Figure 4-5**). These data suggest that the active site of FamC3 and FamC4 is interacting with its partner active site in FamC2 and FamC1, respectively, to cooperatively control stereochemistry of the alkaloid metabolites.

There is an interesting pattern that both 12-*epi*-hapalindole U **3a** and 12-*epi*-hapalindole H **5a** are not produced from *Fischerella ambigua* UTEX 1903, but the reconfigured compounds hapalindole U **2a** and hapalindole H **4a** are isolated as natural products. Based on our previous in vitro data that showed functional heterodimer formation between FamC2 and FamC3, it is reasonable to assume that some Stig cyclases function by cooperating with others in the intracellular environment. But how do they interact with each other to control generation of specific stereo- and regiochemical variants? Currently, we have established the homodimeric conformation of HpiC1 through protein crystallization,[7] and have biochemical evidence to support the formation of Stig cyclase heterodimers, such as FamC2/FamC3.[5] Based on the HpiC1 crystal structure and the high homology of all Stig cyclases, we can deduce that if FamC2/FamC3 is functional as a heterodimer, the two active sites are too far away for a viable interaction.[7] But if they form a higher order oligomer with at least two dimers, the observed packing of dimers in the HpiC1 crystal structure suggests that the two active sites from FamC2 and FamC3 will be sufficiently close to create a joint active site.[7] As shown in **Figure 4-6**, we employed the crystal packing of HpiC1 to build an interaction model of FamC1-FamC4 complex, which is composed of one FamC1 homodimer and one FamC4 homodimer, instead of two FamC1/FamC4 heterodimer. This hypothesis stems our protein-protein interaction assay, showing unlike FamC2/FamC3, FamC1 and FamC4 did not exhibit noticeable interaction.[5] Further investment is necessary to validate this hypothesis, however, this may also help to explain the role of calcium in facilitating reactivity. The need for excess calcium remains unclear, as the FamC2/FamC3 heterodimer and FamC1 homodimer do not require supplemental co-factor to function, but the FamC1-FamC4 hetero-oligomer and WepC1 homodimer require

excess levels of this metal ion. Taken together, these data show that while Ca^{2+} facilitates specific stereochemical outcomes in some Stig cyclases, it is not sufficient to explain the complete product profile. If the role of Ca^{2+} is to assist formation of higher-order Stig cyclase protein complexes, then we can rationalize the inconsistency as a consequence of varied binding affinity among these enzymes.

Another intriguing observation about the reconfiguration is that either **3a** (*10S, 11R, **12S**, 15S*) to **2a** (*10S, 11R, **12R**, 15S*), or **5a** (*10R, 11R, **12S**, 15R*) to **4a** (*10R, 11R, **12R**, 15R*), the alteration was only located at C-12 chiral center, which we reasoned is established during the Cope rearrangement step (**Figure 4-2** and **Figure 4-6**). The C-10 and C-15 chiral centers installed in the 6-*exo*-trig cyclization remain intact, along with the C-ring regiochemistry determined during electrophilic aromatic substitution. One more example to delineate this observation is WepC1-WepC2, with a product profile change from **3a**+**3b** to **2a**+**2b**, exhibiting a stereochemical change at C-12 while retaining the same polycyclic core. These data suggest that the Cope rearrangement is conducted in the Stig cyclases comprised of a combined active site between cognate pairs. The 6-*exo*-trig cyclization and subsequent electrophilic aromatic substitution (or lack of this final step to generate the tricyclic hapalindoles) are controlled by the independently active FamC1/FamC2 monomers and their homologs, which may be considered as the primary subunit. More structural information is needed to demonstrate the Cope rearrangement hypothesis, but we can reinforce our assumption about the electrophilic substitution/elimination with the HawC3 work by identifying some critical residues that mediate C-ring formation.

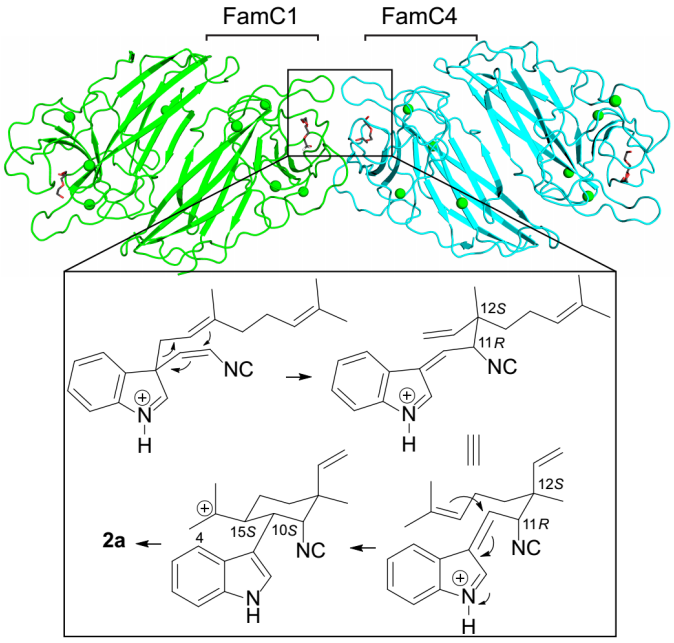


Figure 4-6. Proposed oligomeric model of FamC1/FamC4 complex to explain the interaction of the active sites of FamC1 and FamC4, which pairing controlling the stereochemistry of product hapalindole U (**2a**). The protein structure is adapted from HpiC1 cyclase.

4.3.4 Mutagenesis of HpiC1

As discussed above, in addition to stereoselectivity, the other intriguing structural feature involves regioselectivity during the C-ring formation, which results in tri- and tetracyclic hapalindoles and fischerindoles. It is natural to expect that protein catalyst is a perfect biosynthetic machinery by selectively producing a single metabolite. Through our study, we frequently observed minor tricyclic hapalindoles produced along with stereoidentical tetracyclic hapalindoles or fischerindoles, such as **3b** from HpiC1/WepC1 together with **3a**, and **3b** from FimC5 together with **3c**. Since **3b** is not the precursor to **3a** or **3c**,[5] this phenomenon suggest that the energetic barrier to access the three polycyclic cores is not difficult to overcome, or occurs after the rate determining step. We have recently identified a Stig cyclase HawC3 from strain *Hapalosiphon welwitschii* UH IC-52-3 (**SI Figure 4-2**)[12-14] by producing 12-*epi*-hapalindole C **3b** as the major metabolite (which is named as WelU3 in Liu et al report[14]). By investigating together with FamC1 and FimC5, whose major products are **3a** and

3c, respectively, we hope to gain insight to the factors that control C-ring formation.

	1	2	3	Native product
HpiC1	T72W73	Y101	F138	**3a/3b**
FamC1	T72W73	Y101	L138	**3a**
HawC3	K72L73	F102	F139	**3b**
FimC5	R73I74	F102	S139	**3c**

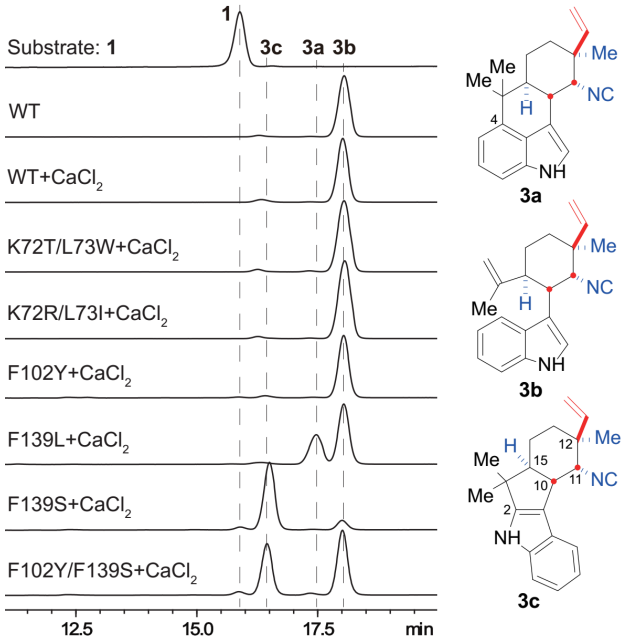


Figure 4-7. In vitro assay with HawC3 wide-type and mutants.

Our structural study of HpiC1 probed some key residues in controlling the activity which led us to reconfigure the normal product of this cyclase from **3a** to **3c**, and FimC5 product from **3c** to **3b**.[7] In this study, we aimed to change the product (**3b**) of HawC3 to both **3a** and **3c**. The strategy is to mine these key residues in FamC1/HawC3/FimC5 through protein alignment and then mutate them in HawC3 to match the ones in FamC1 and FimC5. As shown in **Figure 4-7**, mutating residues K72 and L73 to TW as in FamC1, or RI as in FimC5 did not change the regioselectivity of HawC3, nor did a F102Y mutation of HawC3 mimicking the FamC1 change product profile. However, the mutation of residue F139 proved to be the key in reconfiguring the C-ring profile of the metabolites. F139L slightly shift the tricyclic product **3b** to **3a**, with a ratio 1:5 according to

NMR data (**SI Figure 4-13**); while F139S dramatically altered the product to mainly **3c**. Subsequent effort to thoroughly understand the mechanism is in process, but these data provide new insights regarding active site amino acids that control the electrophilic aromatic substitution/elimination reaction.

4.4 Conclusion

In this study, we explored the activity of FilC2-FilC3 to produce an unexpected molecule 12-*epi*-hapalindole H, which demonstrated the role of supplemental calcium's critical structural role for these Stig cyclases and more importantly, the discovery of FamC3's capability in reconfiguring the product profile. This product reformation was further observed in WepC1-WepC2 by generating the common ambiguine precursor hapalindole U, as well as in FamC1-FamC4 and their homologs. To explain the stereoselectivity, we proposed that Stig cyclases are forming functional, higher-order oligomers, and every monomer is actively contributing the stereo- and regiospecific ring formation but with one monomer as the primary catalytic subunit. Furthermore, we conducted mutagenesis studies on FamCs and HawC3, which suggests that the Cope-rearrangement step is mediated by a combined active site of the heteromeric complexes, which shifted to the primary monomer to accomplish the C-ring formation. This work also showed the ability to change the type of polycyclic ring skeleton by controlling the C-ring formation. Further structural studies are required to verify the hetero-oligomeric complex formation and characterize the formation of other stereochemical patterns (such as hapalindole J), with which, we will be able to manipulate the synthesis of stereo- and regiospecific hapalindole formation.

4.5 Experimental section

4.5.1 General materials and methods.

All NMR spectra were acquired on Varian 400, 600 and 700 MHz spectrometers. Proton and carbon signals are reported in parts per million (d) using residual solvent signals as an internal standard. The LC–MS analysis was performed on a Shimadzu 2010 EV APCI spectrometer. Preparative-scale HPLC was performed using an Agilent Extend C18 10 mm 10 × 250 mm column, using a mobile phase gradient of 70–90% acetonitrile in water over 28 min. High-resolution APCIMS spectra and protein mass spectrometry sequence analysis were obtained from an Agilent 6520 Q-TOF mass spectrometer equipped with an Agilent 1290 HPLC system at the University of Michigan core facility in the Department of Chemistry, with MS grade solvents.

Escherichia coli strain DH5a (Invitrogen) was used for plasmid manipulation, BL21(DE3/pRARE) was used for protein expression. KOD Xtreme Hot Start DNA polymerase (EMD Millipore) was used for polymerase chain reactions. Restriction endonucleases (NheI, XhoI, and BamHI) and T4 DNA ligase were purchased from New England BioLabs. Primers were purchased from Integrated DNA Technologies. PureLink Quick Plasmid Miniprep Kit (Invitrogen) was used to prepare plasmid DNA. All cloned plasmids were confirmed by Sanger sequencing at the University of Michigan DNA Sequencing Core. Isopropyl-D-thiogalactopyranoside (IPTG; GoldBio) was used to induce expression; benzonase and lysozyme used in purification were purchased from Sigma-Aldrich; phenylmethane sulfonyl fluoride (PMSF) was dissolved in isopropanol and used as serine protease inhibitor during protein purification. Ni-NTA agarose from Invitrogen was used to purify 6× His-tag proteins. Terrific broth, lysogeny broth and agar (EMD Millipore) were used for all *E. coli* culturing.

4.5.2 Protein preparation.

Vector pET28 (Novagen) was used to build the expression construct for all cyclases with their N-terminal leading peptides truncated.[5] Except *famC4* and *hawC3* which cloned from codon optimized synthetic genes (IDT gBlocks), the

rest were cloned from genomic DNA: *filC2/filC3/filC4* from IL-199-3-1 genomic DNA, *wepC1/wepC2* from SAG 16.93 genomic DNA, *fimC1/fimC4* from UTEX 1829 genomic DNA. *Westiellopsis prolifica* SAG 16.93 was obtained from Professor Jimmy Orjala (University of Illinois at Chicago), its genomic DNA preparation together with other genomes were described previously.[4] Site-directed mutagenesis of FamC1, FamC2, FamC3, FamC4 and HawC3 was performed using a single primer method based on QuikChange mutagenesis (Agilent Genomics). All primers used are listed in **SI Table 4-1**.

The expression and purification procedure was described in previous work,[4-5] the protein sequences showed in **SI Figure 4-4**.**Error! Reference source not found.**

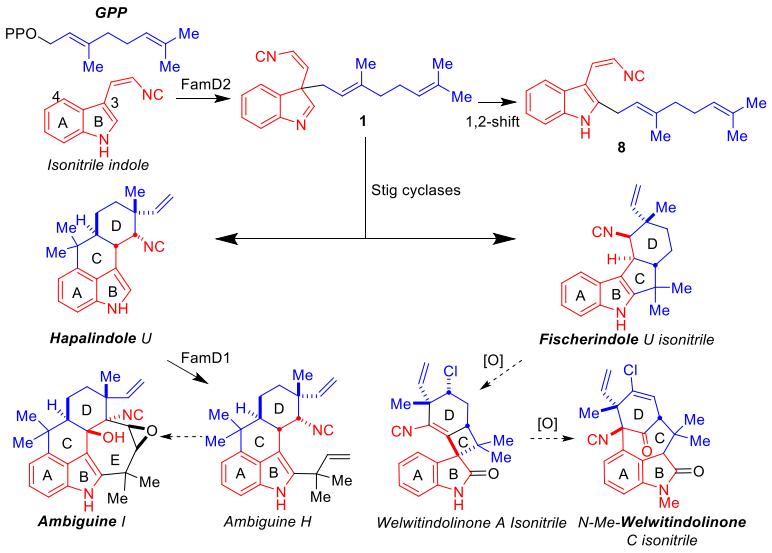
4.5.3 *In vitro* enzymatic assays.

All activity assays were conducted in a 50 µL scale and incubated at 37 °C for 3 h. These reactions were extracted with 1x ethyl acetate twice, the organic layers were dried under N_2 gas and re-dissolved in 50 µL of acetonitrile for LC-MS or HPLC analysis. The reaction condition of each figure was described in detail along with each SI figures. For structural characterization of enzymatic products, the reactions were scaled up to 5 ml and incubated under identical conditions. The extracted products were purified by HPLC equipped with an Agilent Extend C18 10 mm 10 × 250 mm column, and a mobile phase gradient of 70–90% acetonitrile in water over 28 min. The products were dissolved in C6D6 or $CDCl_3$ for NMR analysis.

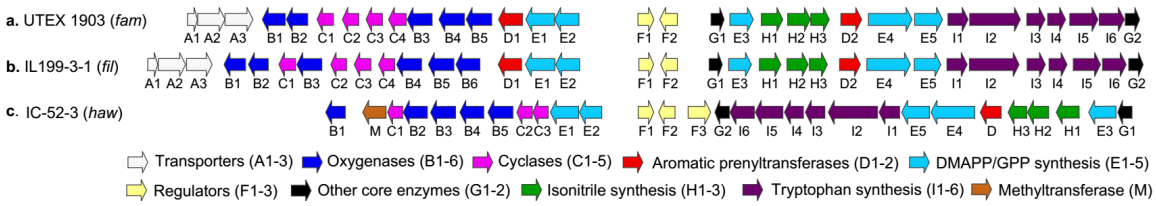
4.6 Supplementary information

SI Table 4-1. Primers used in this study.

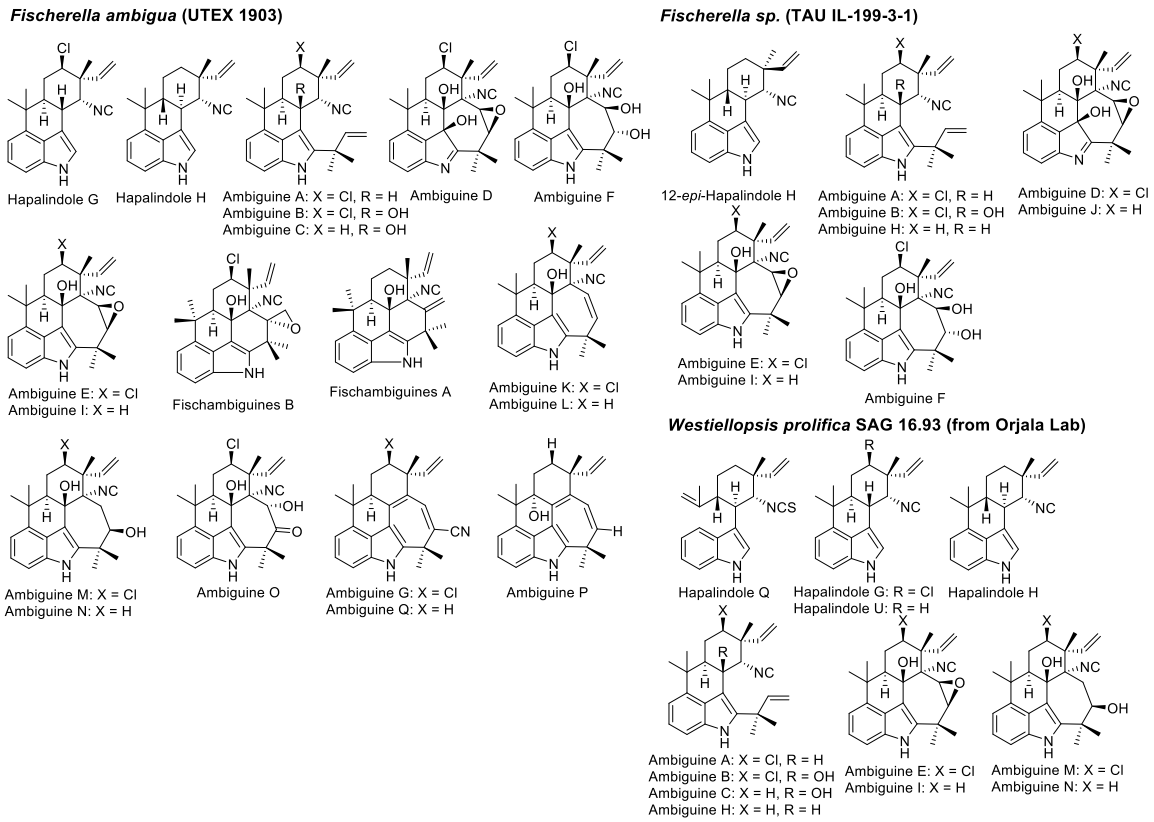
Genes	Forward primer	Reverse primer
filC2	TGCTGCTAGCGCAGTTACT ACTTCCATTC	GTAACTCGAGTTAGGTATCAA CGGTTTCTG
filC3	TGCTGGATCCACAGGTGCT GTTTCTATTCC	ATACCTCGAGCTAAATTACAG CCGATTCAAC
wepC1	CTTCTGAGAATCTCTACTTC CAAGGCGCTAGCAAAGGTG CTGTTTCTATTCCGATAAAC	GATCTCAGTGGTGGTGGTGGT GGTGCTCGAGTTAGGTCTCAG TTGGTTCTGCCGTT
wepC2	CTTCTGAGAATCTCTACTTC CAAGGCGCTAGCGCAAATG TGATTTCAATTCC	GATCTCAGTGGTGGTGGTGGT GGTGCTCGAGTTAAATATTGG TAGGTTGTGC
famC4	AAAGCTAGCGCGAACGTTA TCCCGATTC	AAAAAACTCGAGTTAGATGTT GGTCGGTTGC
fimC1	AAAAGCTAGCACATCTGCT GTTTCCATTC	GTCCTCGAGTTAAGTCTCAGT GGGTTCTGTG
fimC4	TGCTGCTAGCGCAAATGTG ATTCAAATTCC	GTGACTCGAGTTAAATATTGG TAGGTTGTGCAG
filC4	TGCTGCTAGCGCAAATTTG ATTCCAATTCC	GTGACTCGAGTTAAATATTGG TAGGTTGTGC
hawC3	AAAGCTAGCACCACCGCGA TCAGCATTCC	AAAAAACTCGAGTTAGATGTT CGCCGGCTC
famC1-D214A	GCTGCAAGATAAGTTTAGCGGCCTGgcgTTTGATAATGTGCGC CTGACCACCG	
famC2-D217A	GCAGGGTAAATTCAGCGGCCTGgcgTTTGATAACGTGCGTCT GATCACC	
famC3-D214A	CTGCAAACCCTGAGCGGCAACATCgcgTTCGATAACGTGCGT CTGAGCGTTG	
famC4-D215A	CTGCAAAGCAGCAGCGCGAACATTgcgTTTGATAACGTGCGTC TGACCGCG	
hawC3-K72T/L73W	ACGGTCTGGTGCCGGAAAAGCGTACCacctggACCAGCAACAA CGGTGTGGGTCATG	
hawC3-K72R/L73I	ACGGTCTGGTGCCGGAAAAGCGTACCcgtatcACCAGCAACAA CGGTGTGGGTCATG	
hawC3-F102Y	GGAGGGCCGTAACATCGGTtatGTTTATCTGAGCCAGAAACC	
hawC3-F139L	CTGACCGTGGATGTTGGTAACctcGGTGGCGGTTTTCAAAAC	
hawC3-F139S	CTGACCGTGGATGTTGGTAACtccGGTGGCGGTTTTCAAAAC	



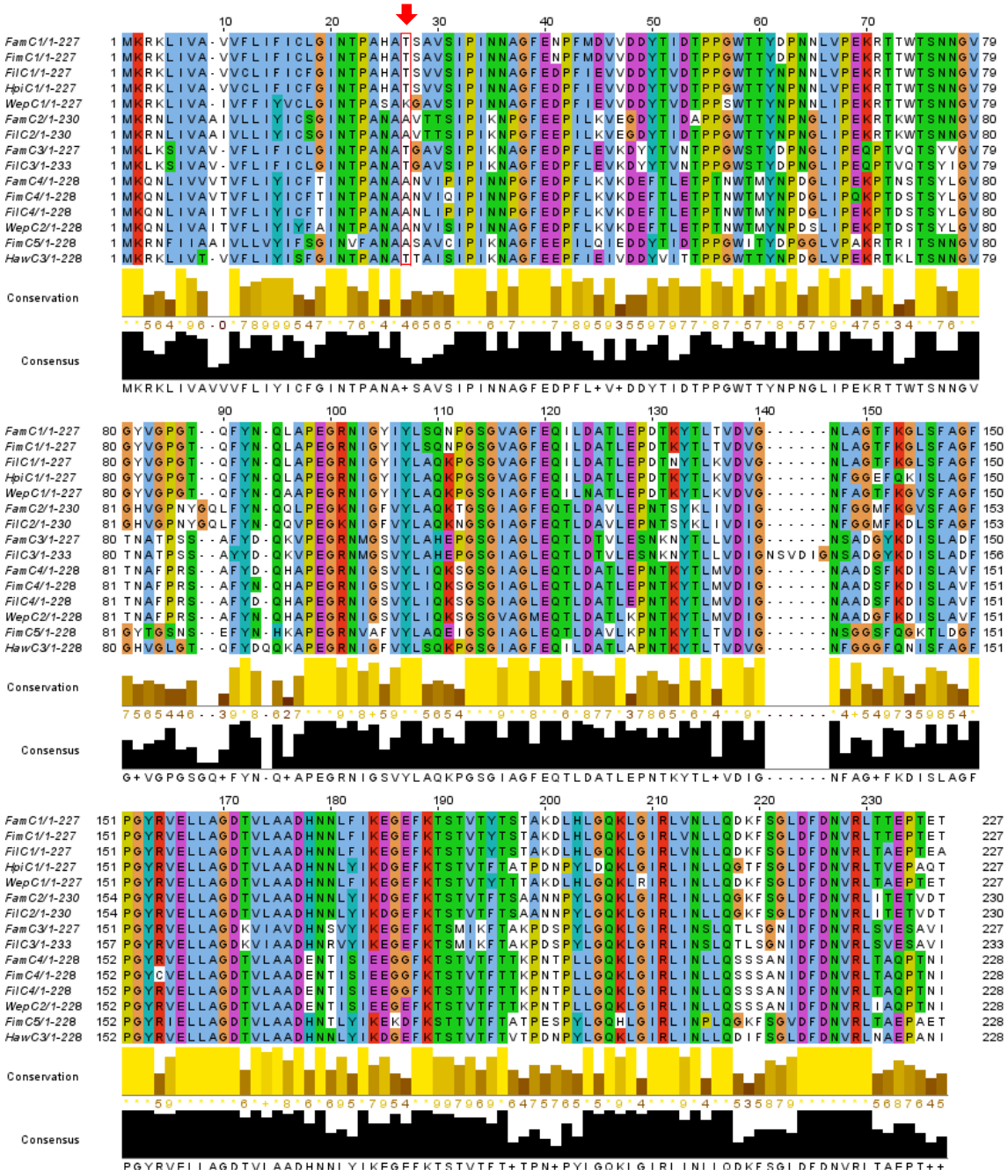
SI Figure 4-1. Proposed biogenesis of the four subgroups, hapalindoles, fischerindoles, ambiguines and welwitindolinones from the common units, GPP and isonitrile indole.



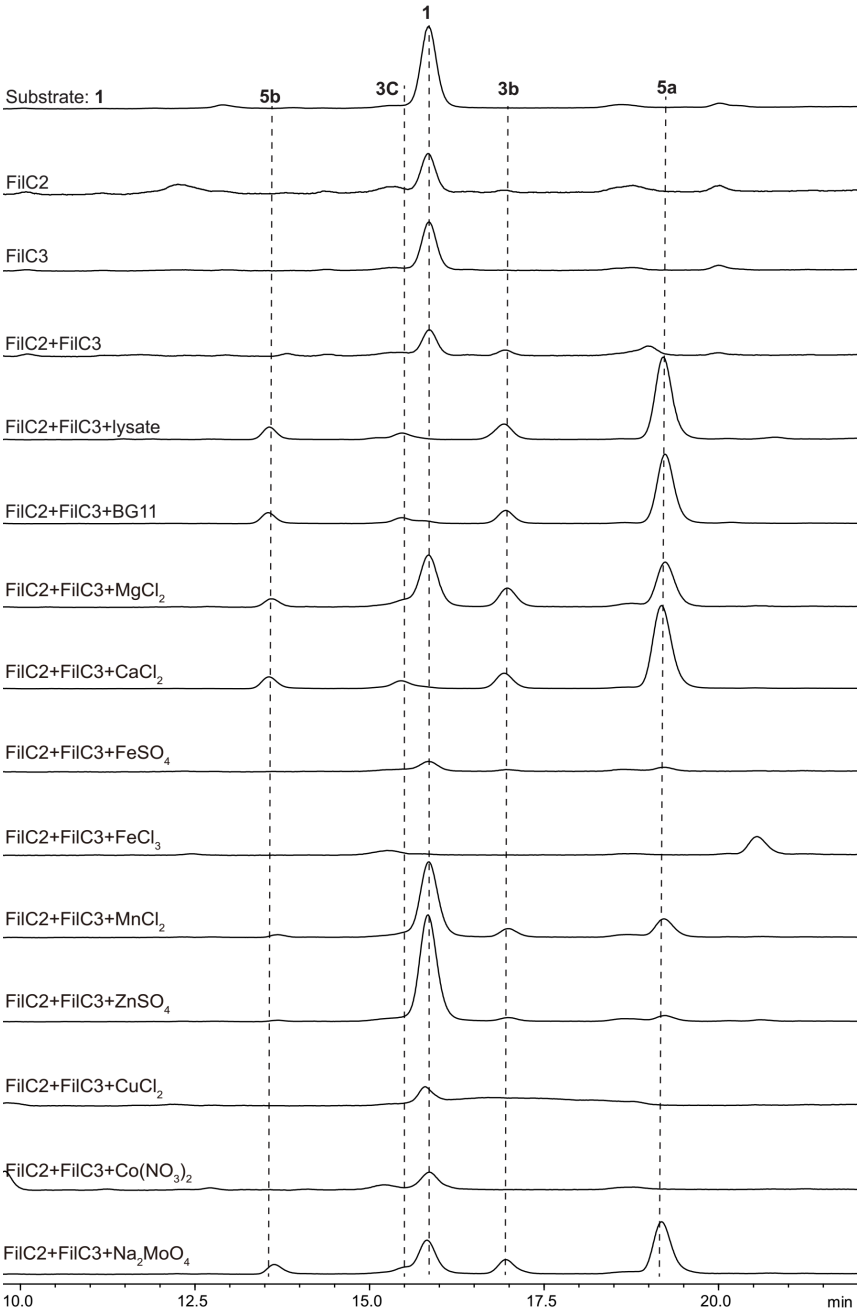
SI Figure 4-2. Gene clusters identified from Hapalindole-producing strains. a) *fam* gene cluster from Ambiguine-producing strain *Fischerella ambigua* UTEX 1903;[4] b) *fil* gene cluster from Ambiguine-producing strain *Fischerella* sp. IL-199-3-1;[5] c) *haw* gene cluster from Welwitindolinone-producing strain *Hapalosiphon welwitschii* W. & G.S. West UH IC-52-3[13].



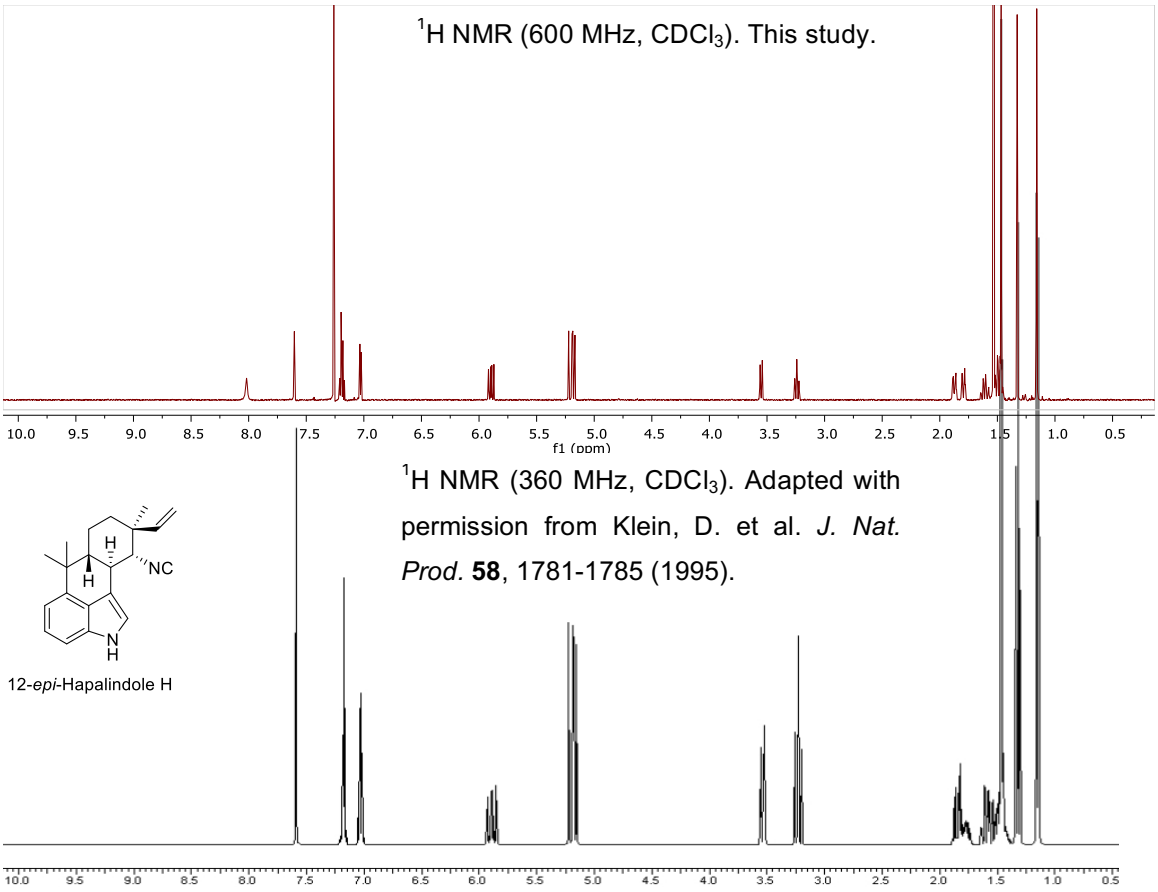
SI Figure 4-3. Isolated hapalindoles from strain UTEX 1903, TAU IL-199-3-1 and SAG 16.93.



SI Figure 4-4. Protein sequence alignment of Stig cyclases analyzed in this report. All heterologously expressed Stig cyclases including their mutants were started from the red-arrow pointed position with the N-terminal lead sequence cleaved.

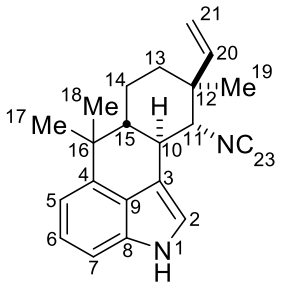


SI Figure 4-5. Complete HPLC traces of in vitro assays to identify the key cofactor for activating FilC2 and FilC3. The assay was conducted in 50 µL scale containing 50 mM Tris buffer (pH 7.8), 1 mM of **1**, 5 µM of each Stig cyclase, 5 mM of the respective metal additive or 10 µL for lysate and BG11 additives. The reaction was incubated at 37 °C and quenched as described in the general assay condition for LC-MS analysis with method: Extend C18 5 mm 4.6 × 150 mm column, and a mobile phase gradient of 70–90% acetonitrile in water over 22 min with flowrate 0.4 mL/min. Some especial components are: Lysate, cell-free lysate of IL 199-3-1[4]; BG-11, the metal-rich media commonly used in cyanobacterial cultivation; the metal-ion screened below are components from the BG-11 media.

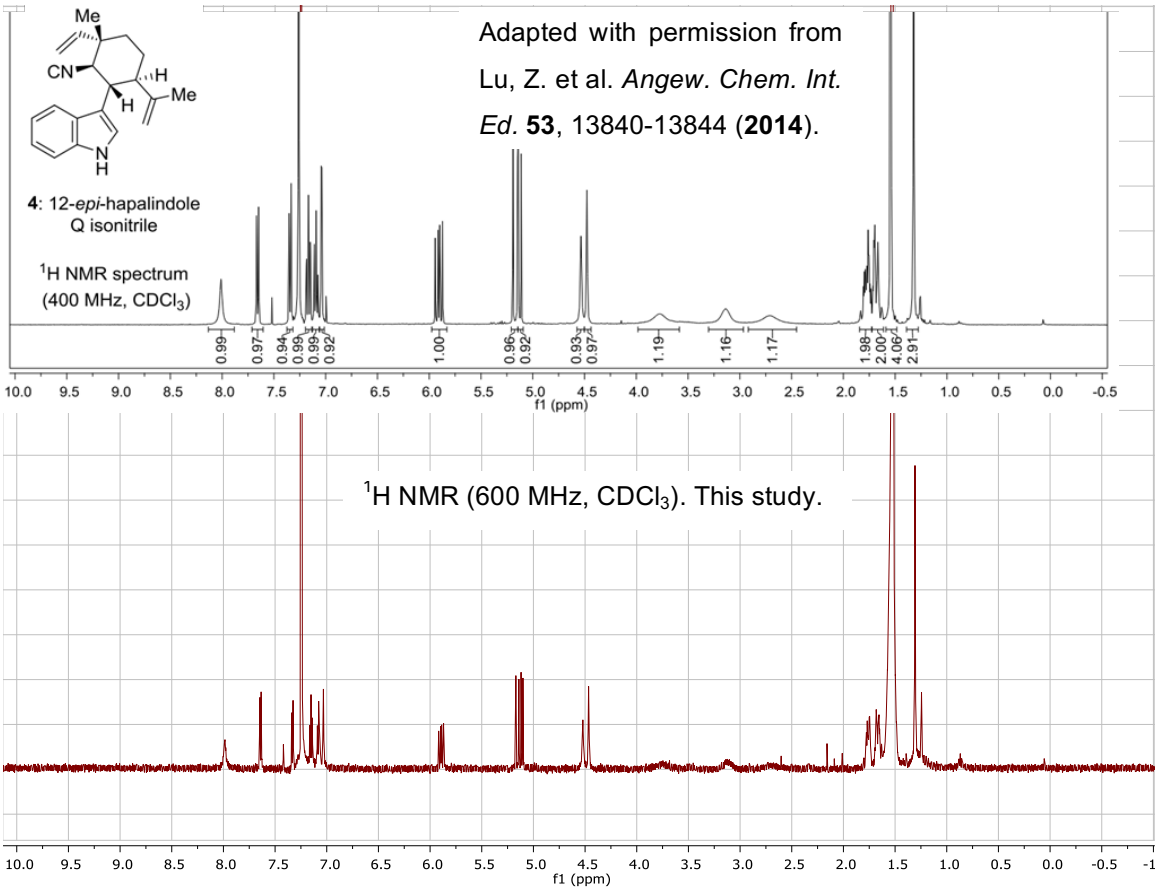


SI Figure 4-6. ^{1}H NMR spectrum overlay of 12-*epi*-hapalindole H (**5a**) with literature data in $CDCl_3$.

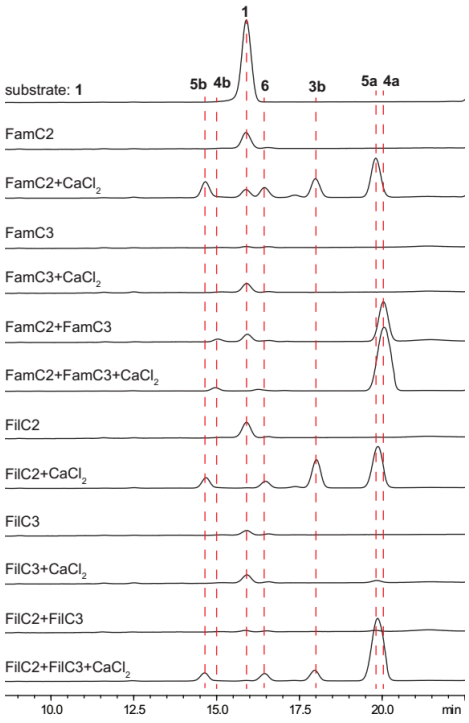
SI Table 4-2. The complete NMR spectroscopic data table for 12-*epi*-hapalindole H (**5a**) in benzene-d6.



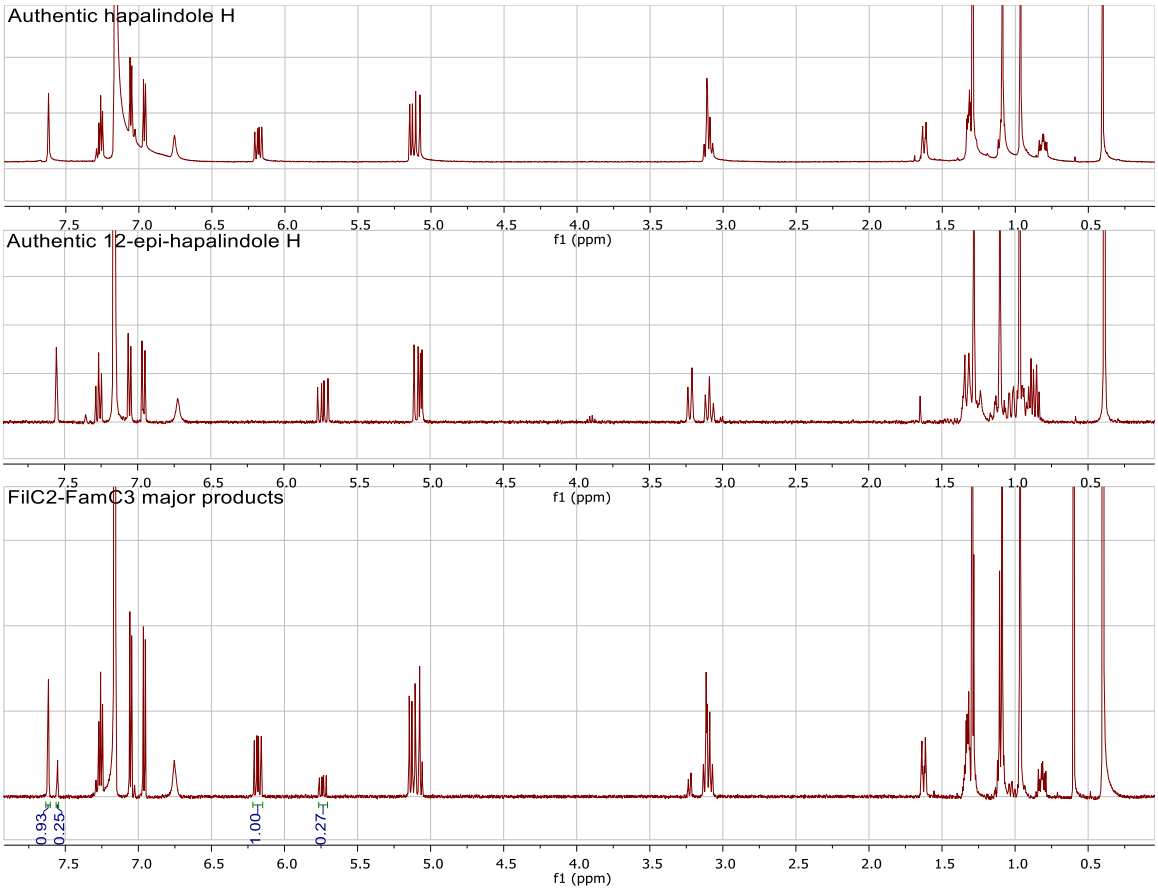
Position	δ_{13C}	δ_{1H} (multi, *J* value)	COSY	HMBC
1	N	6.74 (bs)		
2	118.7	7.56 (s)		3, 8, 9, 10
3	113.2			
4	140.9			
5	112.8	7.06 (d, *J* = 7.1 Hz)	6	6, 7, 9, 16
6	123.2	7.27 (t, *J* = 7.3 Hz)	5,7	4, 5, 7, 8
7	108.6	6.96 (d, *J* = 8.0 Hz)	6	5, 6, 8, 9
8	134.0			
9	125.8			
10	36.3	3.09 (t, *J* = 11.1 Hz)	11, 15	2, 3, 9, 11, 12, 14, 15, 16
11	65.5	3.22 (d, *J* = 11.1 Hz)	10	3, 10, 12, 13, 19, 20, 23
12	40.2			
13	35.9	(Heq) 1.35-1.30 (1H)	14	11, 12, 14, 15, 19
		(Hax) 0.96-0.91 (1H)		
14	20.6	1.35-1.30 (1H), 1.15-1.08 (1H)	13, 15	12, 13, 15
15	49.6	1.04-0.98 (m)	10, 14	
16	37.4			
17	24.9	0.97 (s)		4, 15, 16, 17
18	24.8	1.28 (s)		4, 15, 16, 18
19	17.0	1.10 (s)		11, 12, 13, 20
20	146.0	5.73 (dd, *J* = 10.9, 17.5 Hz)	21	11, 12, 13, 19
21	113.4	(cis) 5.07 (d, *J* = 10.7 Hz)	20	11, 12, 19, 20
		(trans) 5.09 (d, *J* = 17.3 Hz)		
23	160.8			



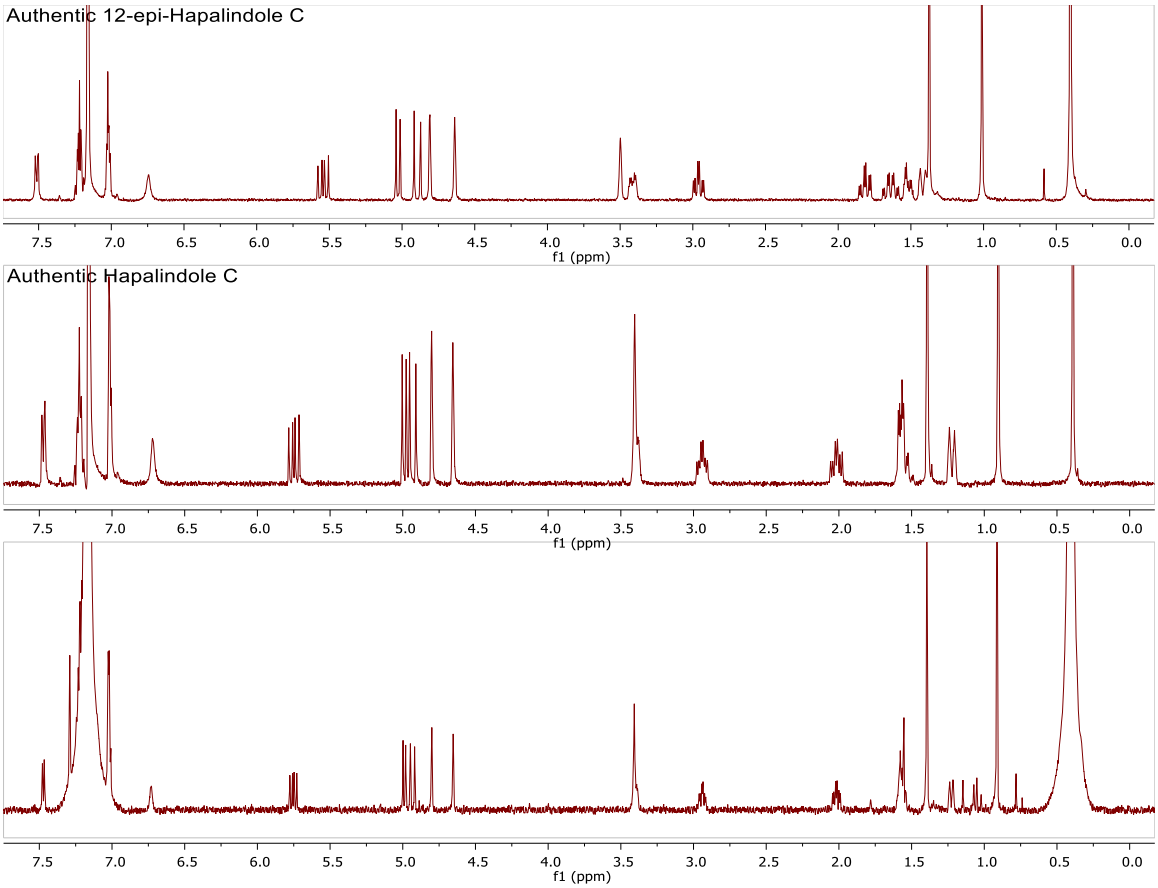
SI Figure 4-7. ^{1}H NMR spectrum of 12-*epi*-hapalindole Q (**5b**) produced by FilC2-FilC3 combination and overlay with literature data in $CDCl_3$.



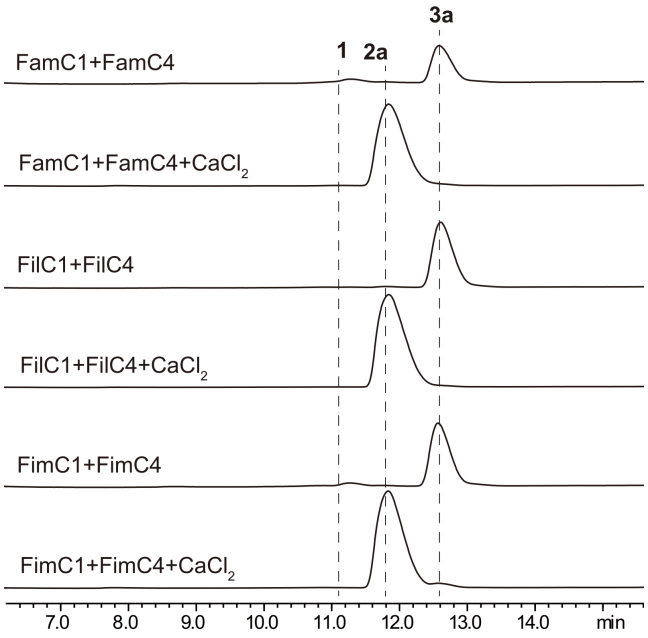
SI Figure 4-8. In vitro assay of FamC2, FamC3, FilC2, and FilC3 to examine the effect of Ca^{2+}. The assay was conducted in 50 µL scale containing 50 mM Tris buffer (pH 7.8), 1 mM of **1**, 5 µM of each Stig cyclase, with or without 5 mM of $CaCl_2$, and incubated at 37 °C. The reaction was quenched as described in the general assay condition and analyzed on HPLC with method: Extend C18 5 mm 4.6 × 150 mm column, and a mobile phase gradient of 50–72% acetonitrile in water over 23 min with flowrate 1.5 mL/min.



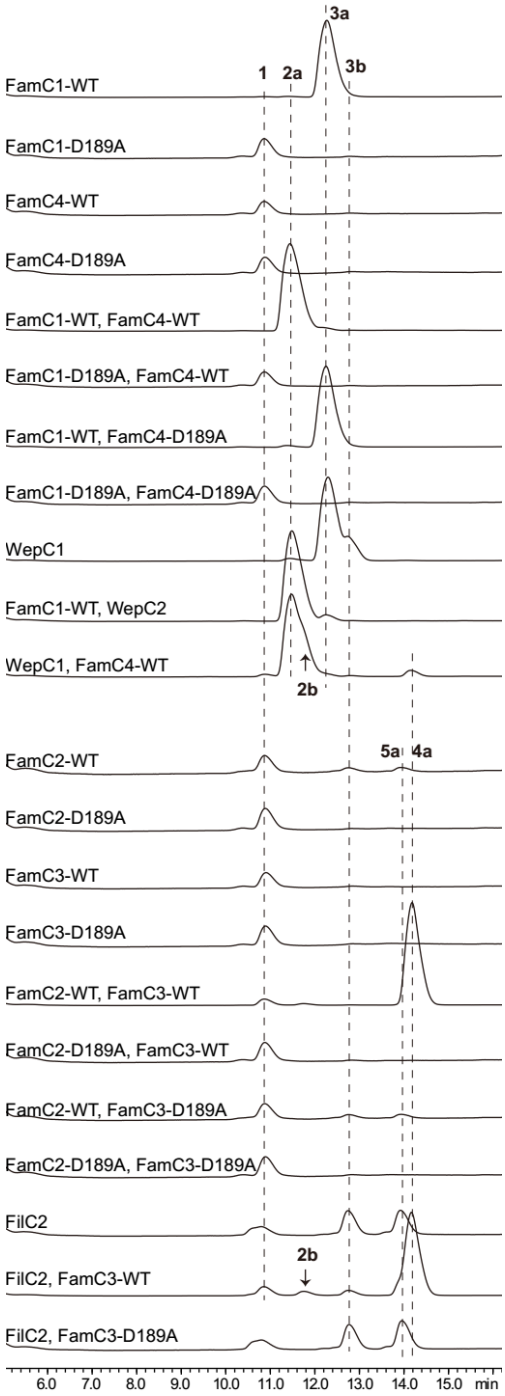
SI Figure 4-9. ^{1}H NMR spectrum of FilC2-FamC3 major product, which is a mixture of hapalindole H and 12-*epi*-hapalindole H with a ratio 4:1.



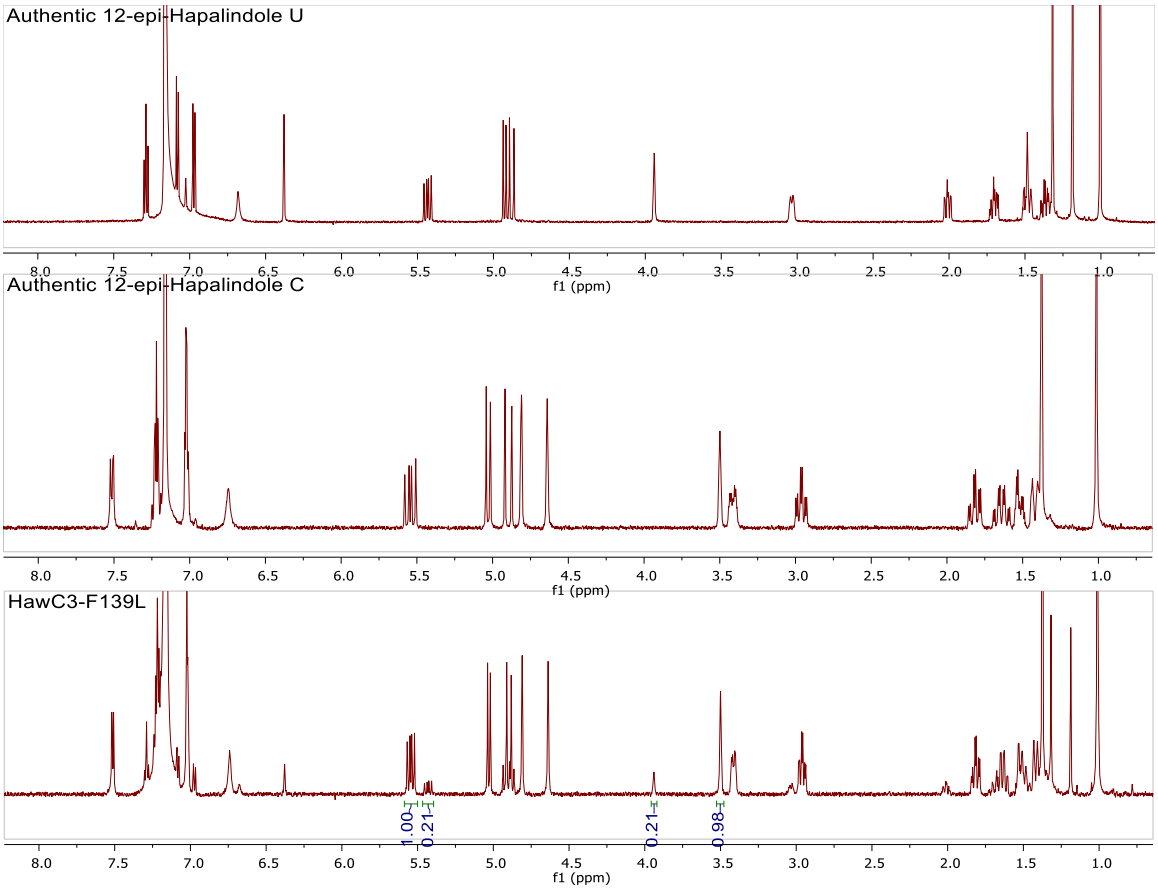
SI Figure 4-10. ^{1}H NMR spectrum of one of the FilC2-FamC3 minor product, by comparing to authentic hapalindole C and 12-*epi*-hapalindole C, this molecule was identified as hapalindole C.



SI Figure 4-11. Complete HPLC traces of in vitro assay for WepC1-WepC2 homologs. The assay was conducted in 50 µL scale containing 50 mM MES buffer (pH 6.5), 1 mM of **1**, 5 µM of each Stig cyclase, with or without 5 mM of $CaCl_2$, and incubated at 37 °C. The reaction was quenched as described in the general assay condition and analyzed on HPLC with method: XBridge Shield RP18 3.5 µm, 3.0x150mm column, and a mobile phase gradient of 50–72% acetonitrile in water over 16 min with flowrate 0.8 mL/min.



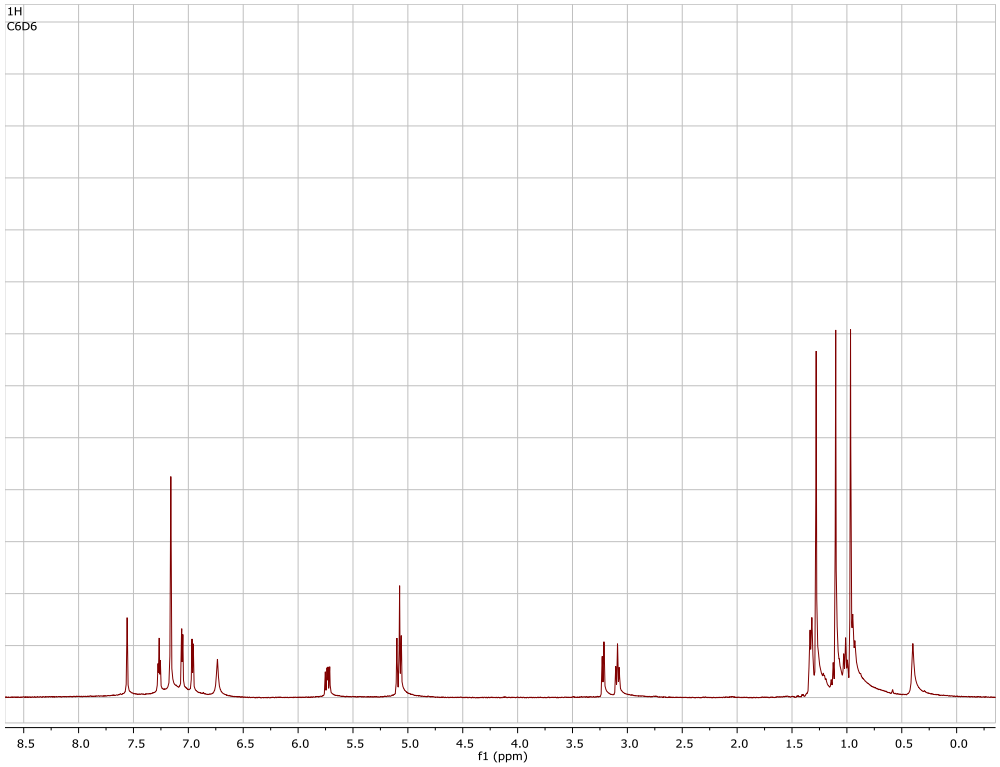
SI Figure 4-12. Complete HPLC traces of in vitro assay for all FamCs mutants. The assay was conducted as a one-pot reaction in 50 µL scale containing 50 mM MES buffer (pH 6.5), 1 mM of isonitrile indole, 1.5 mM of geranyl pyrophosphate (GPP), 5 µM of FamD2, 5 mM of $MgCl_2$, 10 µM of each Stig cyclase, with or without 5 mM of $CaCl_2$, and incubated at 37 °C. The reaction was quenched as described in the general assay condition and analyzed on HPLC with method: XBridge Shield RP18 3.5 µm, 3.0x150mm column, and a mobile phase gradient of 50–72% acetonitrile in water over 16 min with flowrate 0.8 mL/min.

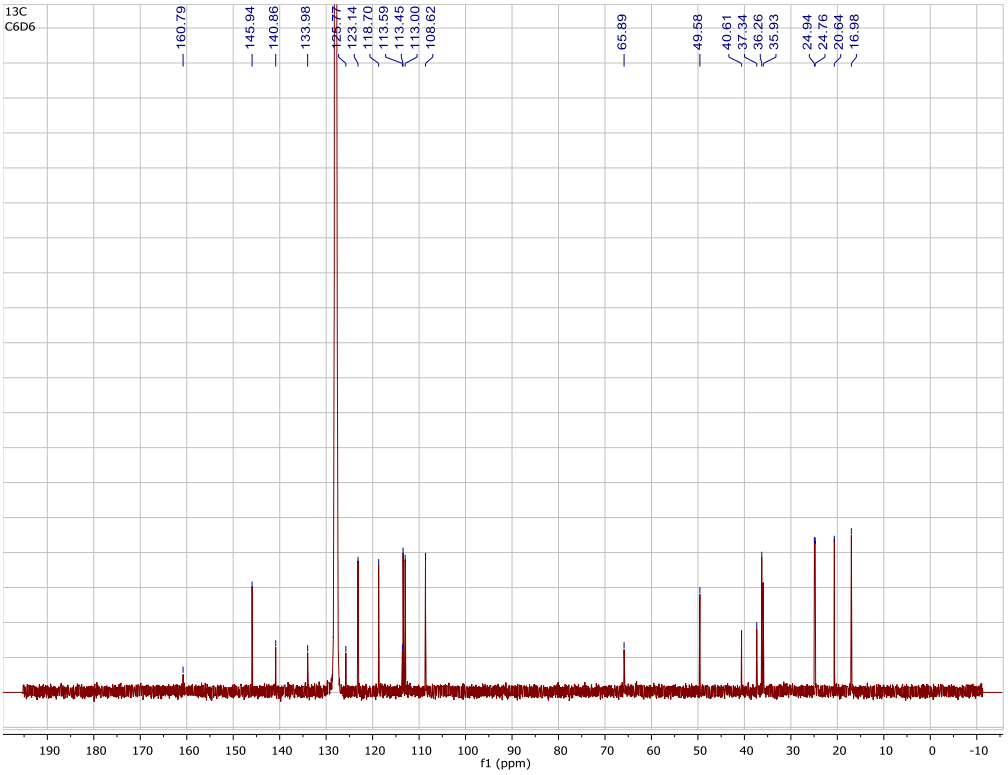


SI Figure 4-13. ^{1}H NMR spectrum of HawC3-F139L product, which is a mixture of 12-*epi*-hapalindole U and 12-*epi*-hapalindole C with a ratio 1:5.

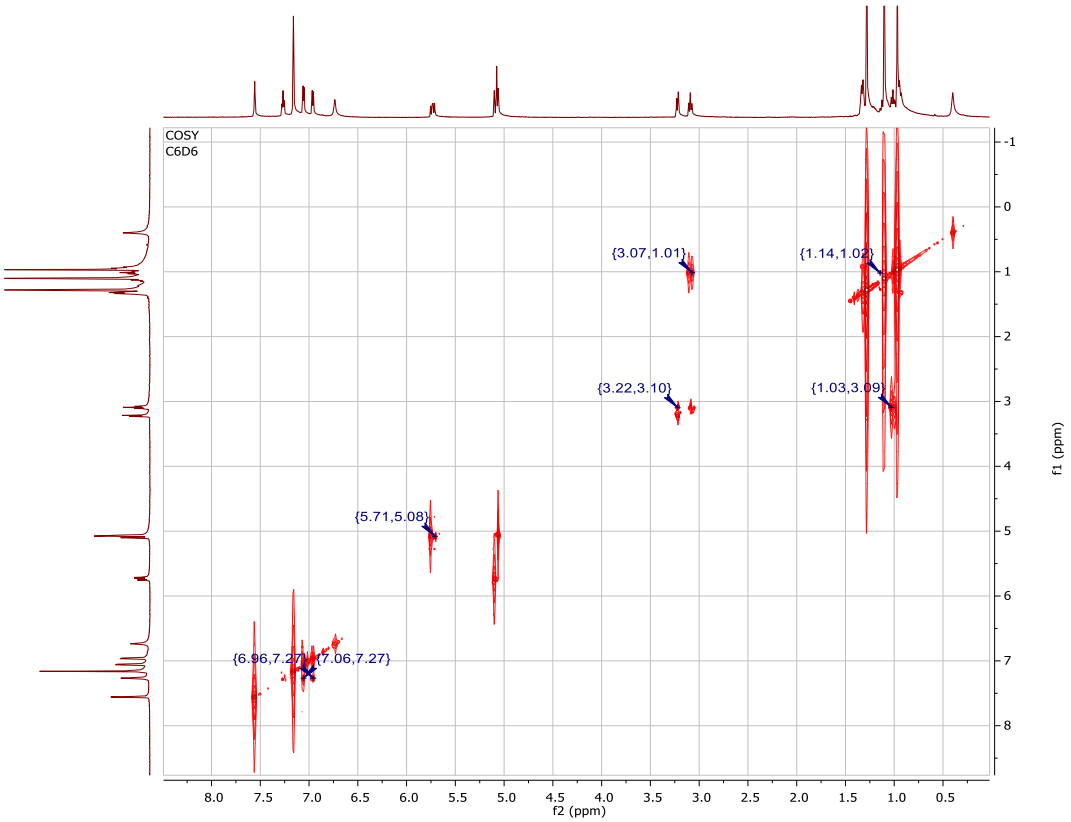
4.7 Spectra section

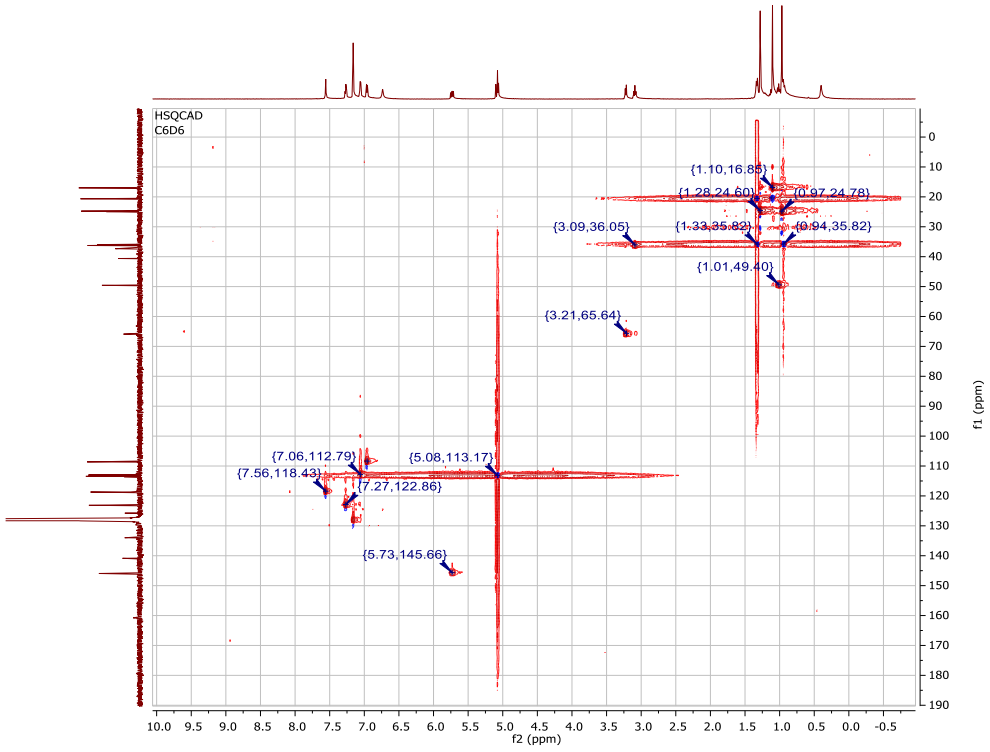
^{1}H and ^{13}C NMR spectra of 12-*epi*-hapalindole H (**5a**) in benzene-d6 at 700/176 MHz.



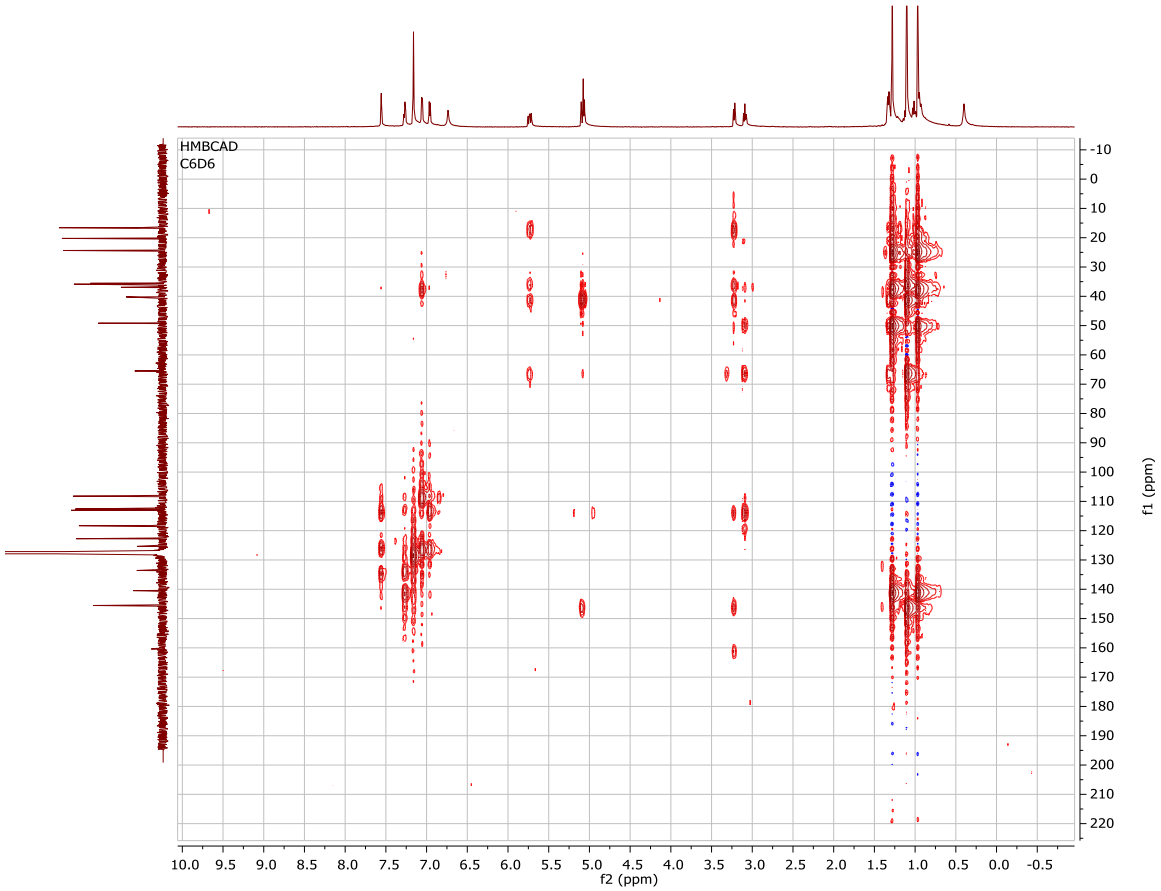


COSY and HSQC NMR spectra of 12-*epi*-hapalindole H (**5a**) in benzene-d6 at 700 MHz.





HMBC NMR spectra of 12-*epi*-hapalindole H (**5a**) in benzene-d6 at 700 MHz.



4.8 Reference

1. Komarek, J.; Kastovsky, J.; Mares, J.; Johansen, J. R., Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* **2014,** *86* (4), 295-335.
2. Moore, R. E.; Cheuk, C.; Patterson, G. M. L., Hapalindoles: new alkaloids from the blue-green alga Hapalosiphon fontinalis. *J. Am. Chem. Soc.* **1984,** *106* (21), 6456-6457.
3. Bhat, V.; Dave, A.; MacKay, J. A.; Rawal, V. H., Chapter two - the chemistry of hapalindoles, fischerindoles, ambiguines, and welwitindolinones. In *The Alkaloids: Chemistry and Biology*, Hans-Joachim, K., Ed. Academic Press: San Diego, 2014; Vol. 73, pp 65-160.
4. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., Hapalindole/Ambiguine biogenesis is mediated by a Cope rearrangement, C–C bond-forming cascade. *J. Am. Chem. Soc.* **2015,** *137* (49), 15366-15369.
5. Li, S.; Lowell, A. N.; Newmister, S. A.; Yu, F.; Williams, R. M.; Sherman, D. H., Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids. *Nat. Chem. Biol.* **2017,** *13* (5), 467-469.
6. Richter, J. M.; Ishihara, Y.; Masuda, T.; Whitefield, B. W.; Llamas, T.; Pohjakallio, A.; Baran, P. S., Enantiospecific total synthesis of the hapalindoles, fischerindoles, and welwitindolinones via a redox economic approach. *J. Am. Chem. Soc.* **2008,** *130* (52), 17938-17954.
7. Newmister, S. A.; Li, S.; Garcia-Borràs, M.; Sanders, J. N.; Yang, S.; Lowell, A. N.; Yu, F.; Smith, J. L.; Williams, R. M.; Houk, K. N.; Sherman, D. H., Structural basis of the Cope rearrangement and C−C bond-forming cascade in hapalindole/fischerindole biogenesis. **2017**.
8. Raveh, A.; Carmeli, S., Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* **2007,** *70* (2), 196-201.
9. Smitka, T. A.; Bonjouklian, R.; Doolin, L.; Jones, N. D.; Deeter, J. B.; Yoshida, W. Y.; Prinsep, M. R.; Moore, R. E.; Patterson, G. M. L., Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae. *J. Org. Chem.* **1992,** *57* (3), 857-861.
10. Mo, S.; Krunic, A.; Chlipala, G.; Orjala, J., Antimicrobial ambiguine isonitriles from the cyanobacterium Fischerella ambigua. *J. Nat. Prod.* **2009,** *72* (5), 894-899.
11. Zhu, Q.; Liu, X., Discovery of a Calcium-Dependent Enzymatic Cascade for the Selective Assembly of Hapalindole-Type Alkaloids: On the Biosynthetic Origin of Hapalindole U. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (31), 9062-9066.
12. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, unusual alkaloids from the blue-green algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to fischerindoles and hapalinodoles. *J. Am. Chem. Soc.* **1994,** *116* (22), 9935-9942.

13. Micallef, M. L.; Sharma, D.; Bunn, B. M.; Gerwick, L.; Viswanathan, R.; Moffitt, M. C., Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol.* **2014,** *14*, 213-230.
14. Zhu, Q.; Liu, X., Molecular and genetic basis for early stage structural diversifications in hapalindole-type alkaloid biogenesis. *Chem. Commun.* **2017,** *53* (19), 2826-2829.

Notes:

This work is in preparation for submission as “Control of Stereoselectivity in Diverse Hapalindole/Fischerindole Metabolites is Mediated by Cofactor Induced Combinatorial Pairing of Stig Cyclases.” Shasha Li, Andrew N. Lowell, Sean A. Newmister, Fengan Yu, Robert M. Williams & David H. Sherman.

Author contributions:

S.L., S.A.N. and D.H.S. designed the research. S.L. performed biological experiments. A.N.L. synthesized the indole isonitrile and geranl pyrophosphate. S.L. and S.A.N. conducted the mutagenesis, S.L., A.N.L., S.A.N. and D.H.S. conducted data analysis and interpretation. F.Y. performed bioinformatics analyses. S.L., A.N.L., S.A.N., D.H.S., and R.M.W. contributed to manuscript preparation.

Chapter 5
Structural basis of the Cope rearrangement and C−C bond-forming cascade in hapalindole/fischerindole biogenesis

5.1 Abstract

In Chapter 2-4, we thoroughly studied the novel Stig cyclases in forming hapalindoles and fischerindoles by characterizing the enzymatic activity and product profile of each cyclase. The reaction includes a rare Cope-rearrangement, 6-*exo*-*trig* cyclization, and electrophilic aromatic substitution. In order to understand the molecular basis of these alluringly Stig cyclases in controlling this complex reaction, we studied the protein structure. Here we report the structure of HpiC1, a Stig cyclase that catalyzes the formation of 12-*epi*-hapalindole U in vitro. The 1.5 Å structure reveals a dimeric assembly with two calcium ions per monomer and the active sites located at the distal ends of the protein dimer. Mutational analysis and computational methods uncovered key residues for an acid catalyzed [3,3]-sigmatropic rearrangement and specific determinants that control the position of terminal electrophilic aromatic substitution leading to a switch from hapalindole to fischerindole alkaloids.

5.2 Introduction

The hapalindole family of alkaloids are a large and structurally diverse class of natural products from cyanobacteria of the order Hapalosiphonaceae [1]. These metabolites are active against a broad range of biological targets, which include antibacterial, antifungal, insecticidal, and antimitotic activities[2-7]. Each member is classified as a hapalindole, ambiguine, fischerindole, or welwitindolinone based on its core ring system (**SI Figure 5-1**), and they have

been the subject of various total syntheses due to their challenging structural complexity and unique biological properties[1]. Until recently comparatively little was known regarding the biogenesis of these alkaloids, and particularly the construction of the tetracyclic core ring system.

Initial reports demonstrated that hapalindoles are derived from cis-indole isonitrile and geranyl pyrophosphate (GPP)[8-10], but the biogenesis of the polycyclic ring systems remained elusive. We recently identified an unexpected biosynthetic intermediate **1** that undergoes a Cope rearrangement followed by a cyclization cascade to generate 12-*epi*-hapalindole U (**12H**) (**Figure 5-1**)[11]. The Cope rearrangement is a 3,3-sigmatropic rearrangement that proceeds through a cyclic transition state (**Figure 5-1**b)[12]. Although this pericyclic reaction is prevalent in organic synthesis[13], it has rarely been identified as a biosynthetic transformation[14-16], and no structural information has been previously reported on enzymes that catalyze this reaction in secondary metabolism. The biosynthesis of **12H** was proposed to proceed through a three-part reaction mechanism: (1) Cope rearrangement of **1** to generate intermediate **3**, which sets the stereochemistry at positions C11 and C12; (2) 6-*exo-trig* cyclization of **3** to intermediate **4**, which sets the stereochemistry at positions C10 and C15; and (3) electrophilic aromatic substitution of **4** to give **12H** upon deprotonation (**Figure 5-1***b*). This discovery was expanded to include several Stig cyclases and revealed that the variant configurations observed in this class of alkaloids are generated from a central biosynthetic intermediate **1**, transformed to products in a regio- and stereospecific fashion (**Figure 5-1***b*)[17-19]. Thus, biogenesis of hapalindole-type metabolites includes a fascinating mechanistic puzzle regarding how homologous Stig cyclases maintain stereochemical and regiochemical control at each of these biosynthetic steps in the formation of varied hapalindoles and fischerindoles.

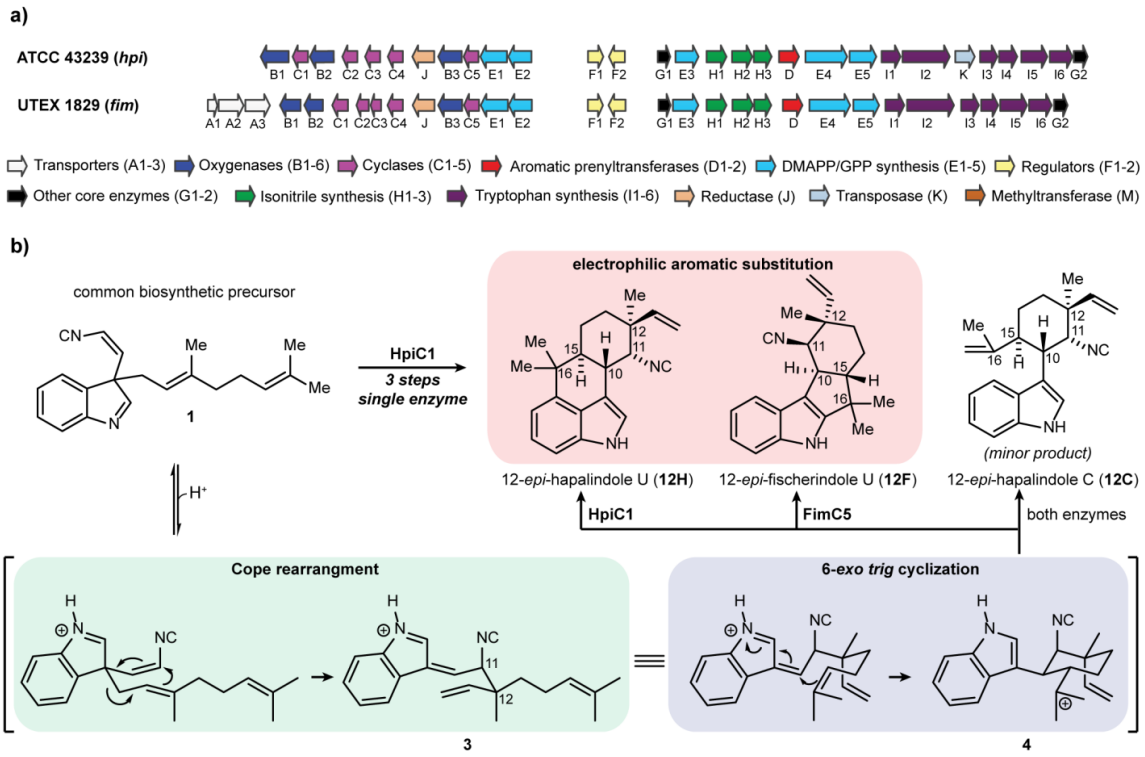


Figure 5-1. Biogenesis of hapalindole alkaloids. (**a**) *hpi* and *fim* gene clusters from *Stigonematales* cyanobacterial strains *Fischerella* sp. ATCC 43239 and *Fischerella muscicola* UTEX 1829. (**b**) The hapalindole and fischerindole core ring systems arise from the common biosynthetic intermediate **1**. Stig cyclases catalyze a Cope rearrangement followed by cyclization to generate tetracyclic products and trace levels of tri-cyclic shunt products. HpiC1 catalyzes formation of **12H**, while FimC5 catalyzes formation of **12F**, with identical stereochemistry at C11 and C12, but different C-ring regiochemistry.

In this study, we describe the molecular basis for the Stig cyclase ability to control three reactions. Originally annotated as unknown proteins[8], no previous information was available regarding the structure of this new type of biosynthetic enzyme. We describe herein the first crystal structure of a Stig cyclase, HpiC1, and show through a mutational analysis that localizes its active site (HpiC1, and other members of these unique enzymes) and demonstrates the ability to reconfigure its metabolite profile. Density functional calculations on the mechanism and modes of hydrogen bonding catalysis, and molecular dynamics simulations on the wild-type and mutant enzymes provide detailed information on the enzymatic processes that control product formation. These data provide compelling insights into the mechanism of Cope rearrangement, 6-*exo*-*trig*

cyclization and electrophilic aromatic substitution for this broad class of natural products.

5.3 Results

5.3.1 Dimeric structure and Ca^{2+} binding

Following recent work that reported the function and selectivity of Stig cyclase proteins[11, 17-19], we sought a crystal structure of HpiC1 in an effort to understand the structural basis of the complex cyclization cascade of its substrate **1**, and to understand how the Stig cyclases catalyze formation of variant alkaloid products from this common biosynthetic intermediate. Ultimately, we obtained HpiC1 crystals in four different forms, under conditions that differed in Ca^{2+} concentration. The 1.7 Å structure in Form 1 was solved by selenomethionyl SAD phasing from a HpiC1 W73M/K132M double mutant, as the wild-type protein lacks Met. The SeMet W73M/K132M structure was used as a search model to solve structures in the other three crystal forms by molecular replacement. The overall fold of the HpiC1 polypeptide is a flattened β-jelly roll fold (**Figure 5-2***a*) composed of two antiparallel β-sheets. In all crystal forms, the antiparallel pairing of strands β6 in two monomers creates a continuous β-sheet across an extensive dimer interface, which buries 2060 $Å^2$ of total surface area (PISA, **Figure 5-2***b*)[20], and encompasses approximately 20% of the total surface area of each monomer. This dimer interface is consistent with size exclusion chromatography analysis in which HpiC1 and other Stig cyclases migrated as apparent dimers[17]. HpiC1 shares highest structural similarity (2.3 Å rmsd, DALI[21]) with the carbohydrate-binding module (CBM) from xylanase in the thermostable bacterium *Rhodothermus marinus* (PDBid: 2Y64, **Figure 5-2***c*)[22]. The proteins have highly similar tertiary structure and topology with the most significant differences occurring at their N-termini and in the loop regions between shared β-strands.

HpiC1 has two integral Ca^{2+} ions, each with octahedral coordination geometry. Ca^{2+} site 1 is ligated by the carbonyl oxygens from F138, L147, G149, the side chain oxygens of N137 and D175, and a water molecule; and Ca^{2+} site 2

by the G37, E95 and N98 carbonyl oxygens and the E14, E95, and D216 side chains (**SI Figure 5-2.** Two hexacoordinate Ca2+ ions are present in each HpiC1 monomer.

(**a**) The first site is closer to the active site, and is ligated by carbonyl oxygen atoms from F138, L147, G149, and aspartate/asparagine side chain atoms from N137 and D175. A water molecule is the sixth ligand. (**b**) A second Ca2+ is ligated by carbonyl oxygen atoms from G37, E95, N98, and carboxylate/amide side chain atoms from E39, E95, and D216.). These sites were a key starting point to assess the structural and catalytic role that Ca^{2+} plays in the Stig cyclase enzymes. In vitro assays conducted in the presence of 5 mM EDTA showed no activity, which confirmed that Ca^{2+} is required for catalytic function in HpiC1[18]. From sequence comparisons, we expect the integral calcium-binding sites and the core dimeric assembly to be maintained in the other Stig cyclase enzymes (**SI Figure 5-3**).

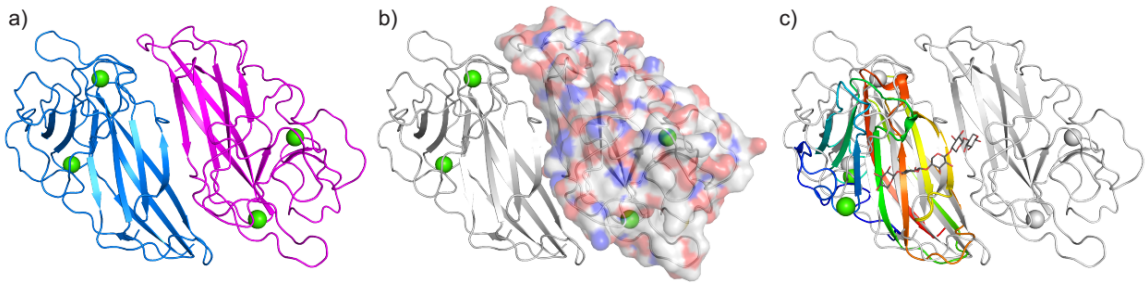


Figure 5-2. HpiC1 structure overview at 1.5 Å. (**a**) Cartoon representation of HpiC1 homodimer; the subunits are colored blue and magenta, green spheres indicate bound calcium ions. (**b**) Surface representation colored by atom of a single protomer shows 2060 $Å^2$ buried surface area between the subunits. (**c**) Superposition of xylanase CBM homolog (PDB id: 2Y64, rainbow); CBM is monomeric despite sharing the same fold as HpiC1.

5.3.2 Active site identification

We next sought to determine the location of the enzyme active site. The high resolution, substrate-free HpiC1 structure did not immediately suggest a location for substrate binding and catalysis. Initial localization of the active site was revealed by abnormal electron density in a pocket located at the distal end of each subunit between Ca^{2+} sites 1 and 2. Polyethylene glycol from the

crystallization was modeled into this pocket (**SI Figure 5-4**), which was composed of numerous aromatic amino acids (**Figure 5-3**a). Given the hydrophobicity of substrate **1**, we explored whether this hydrophobic pocket could be the enzyme active site. This region was probed using Autodock VINA with **12H** as the ligand[23]. The major product was chosen based on its defined stereochemistry and rigid scaffold. The top three docking solutions had affinities ranging between -9.7 and -9.4 kcal/mol (**Figure 5-3**b, **SI Figure 5-5**), which suggested to a first approximation that this site possesses an appropriate size and shape to accommodate the hapalindole core.

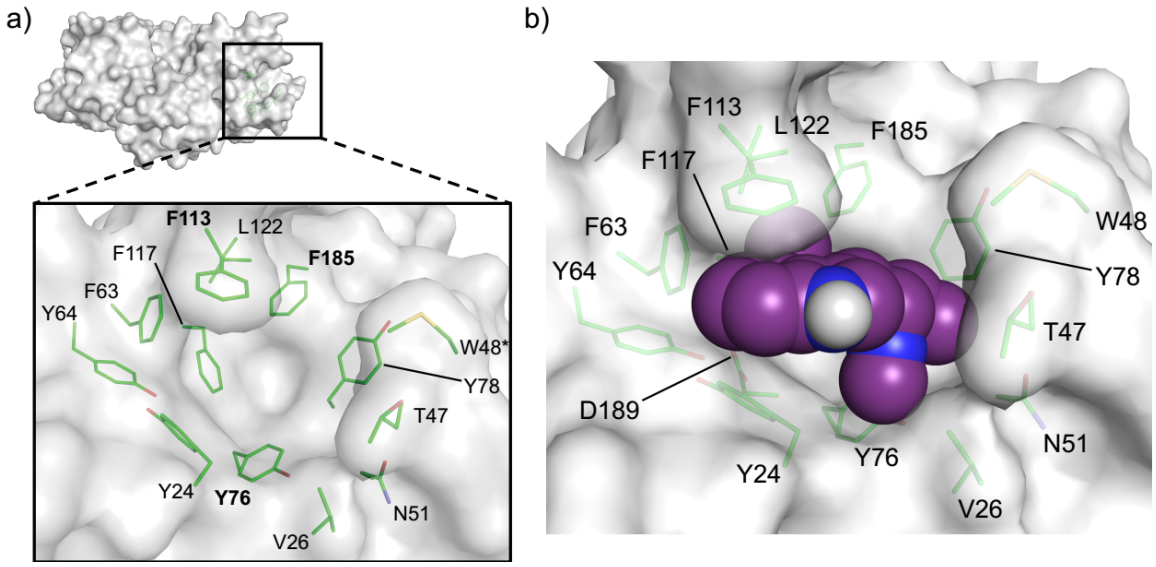


Figure 5-3. Active site of SeMet HpiC1 W73M/K132M. a) Surface representation of the SeMet HpiC1 active site. Key residues are shown as green sticks. Met73 is substituted for the native tryptophan residue. This mutant protein retained wild-type activity. b) Autodock VINA was used to examine if the binding pocket was appropriately sized for 12-*epi*-hapalindole U.

The initial docking solutions compelled us to further interrogate this region of the protein as the putative active site. First, we probed the role of HpiC1 D214 toward catalyzing formation of **12H**. This residue is 100% conserved in all of the currently identified Stig cyclases and is exceptional as the amino acid in the hydrophobic pocket most likely to participate in acid/base chemistry, which can be inferred to promote [3,3]-sigmatropic rearrangments[24]. The HpiC1 structure shows that D214 carboxylate lacks a counter-ion and is hydrogen bonded to the

Y89 hydroxyl. Substitution of D214 to alanine abolished activity in HpiC1 (**Figure 5-6**a), indicating that it plays a critical role in the catalytic cascade to generate cyclized products (addressed computationally below), and provides compelling evidence that this hydrophobic pocket is the enzyme active site.

The enzyme conformation about D214 varies among the individual crystal forms (**Figure 5-4**a). In Form 1, D214 is shielded by F138, whereas this residue is shifted in Forms 2, 3, and 4 to expose D214 to the binding pocket. The loop containing F138 is also shifted ~3 Å from its position in Form 1, although coordination of the calcium ion by the carbonyl oxygen of F138 is maintained. The atomic temperature factors in this region also indicate a higher degree of mobility at this site compared to other sections of the protein (**Figure 5-4**b). It is apparent that conformational flexibility would be required to perform multiple catalytic steps in a single binding domain, and also to accommodate the broad range of regio- and stereochemical configurations present in this class of polycyclic alkaloids derived from the initial bicyclic species **1**[1].

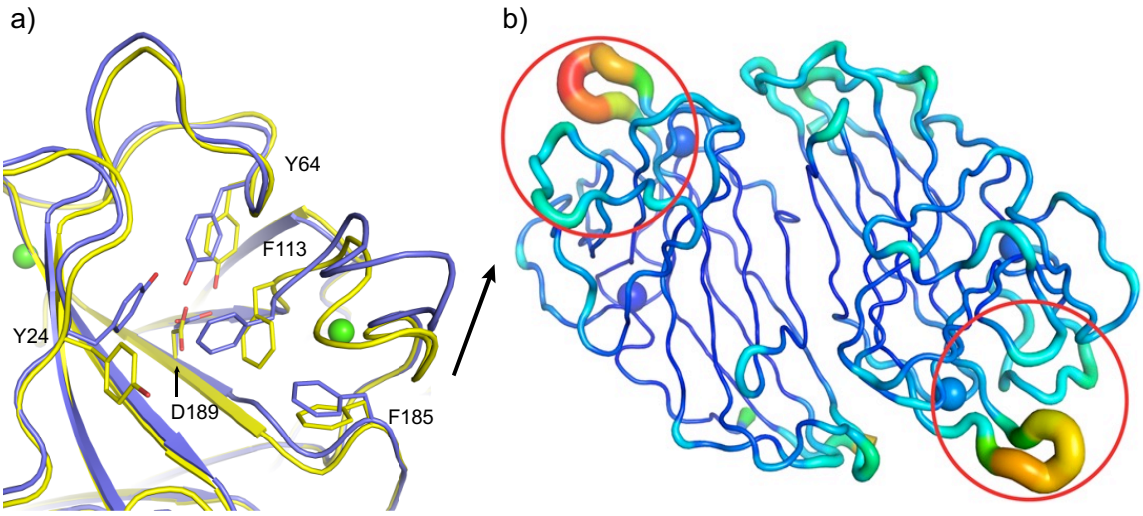


Figure 5-4. Active site of HpiC1. a) Comparison of SeMet HpiC1 W73M/K132M (*blue*, Form 1) and native HpiC1 (*yellow*, Forms 2 and 3) shows coupled side chain and main chain differences in the active site. D214 is shielded by F138 in SeMet HpiC1, whereas this residue is shifted in the native protein to expose D214 to the binding pocket. D214 is 100% conserved in known Stig cyclases. The loop containing F138 is also shifted ~3 Å from its position in the SeMet protein, although coordination of the structural calcium ion by the carbonyl oxygen of F138 is maintained. b) B-factor putty representation of HpiC1

in crystal form 1 suggests that the enzyme active site is a comparatively mobile region of the protein (PyMOL)[25].

Based on the different product profiles of the various Stig cyclases, we examined the conservation of key residues in this hydrophobic pocket (**Figure 5-5**). Comparison of the Stig cyclases revealed that primary sequence clustering correlates to similar product profiles. We reasoned that information regarding key determinants of cyclase reactivity was contained within these localized sequences, and this active site was explored by mutagenesis. A comparative analysis between HpiC1 (catalyzes **12H** production from **1**) and FimC5 (catalyzes **12F** production from **1**) was pursued based on their function as homodimers in vitro, and their corresponding ability to differentially produce hapalindole or fischerindole core ring systems, respectively[17].

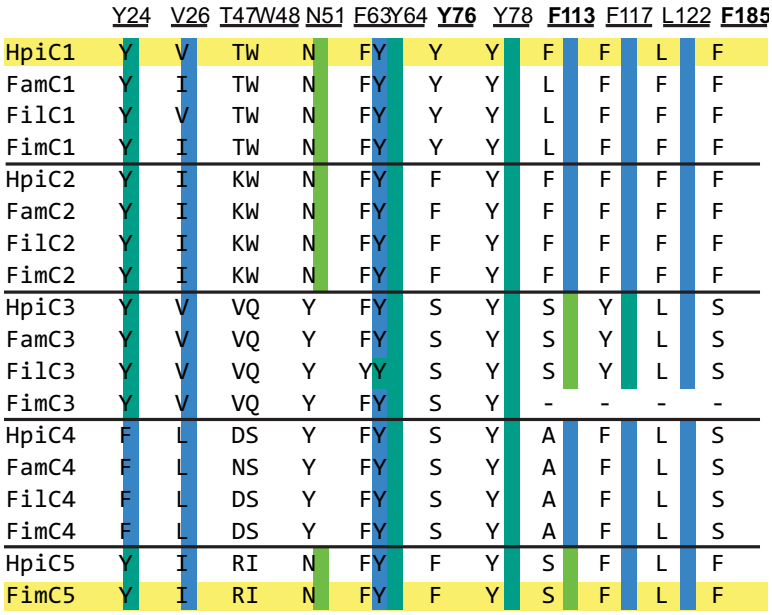


Figure 5-5. Key active site residues shown in an alignment with other Stig cyclases. Residues are colored by conservation and side chain composition (ClustalX).

Remarkably, substitution of F138 in HpiC1 to the corresponding serine from FimC5 led to the generation of a mixture of its major product, hapalindole **12H**, and the FimC5 major product, fischerindole **12F** (**Figure 5-6**a). The product ratio in F138S is approximately 1:1 (**12H**:**12F**) and is further shifted to 2:1 (**12F**:**12H**) in the Y101F/F138S double mutant. The HpiC1 Y101F single mutant

had comparable activity to wild-type protein, indicating that this effect is driven primarily by F138. A corresponding mutation at this key position in FimC5, S139F, did not lead to formation of **12H**; instead the product profile was shifted towards the production of **12C**, a minor tricyclic shunt product of the native HpiC1 and FimC5 reactions (**Figure 5-6**b)[17, 26]. These data (and computational studies described below) indicate that both HpiC1 F138 and FimC5 S139 play a key role in directing terminal electrophilic aromatic substitution.

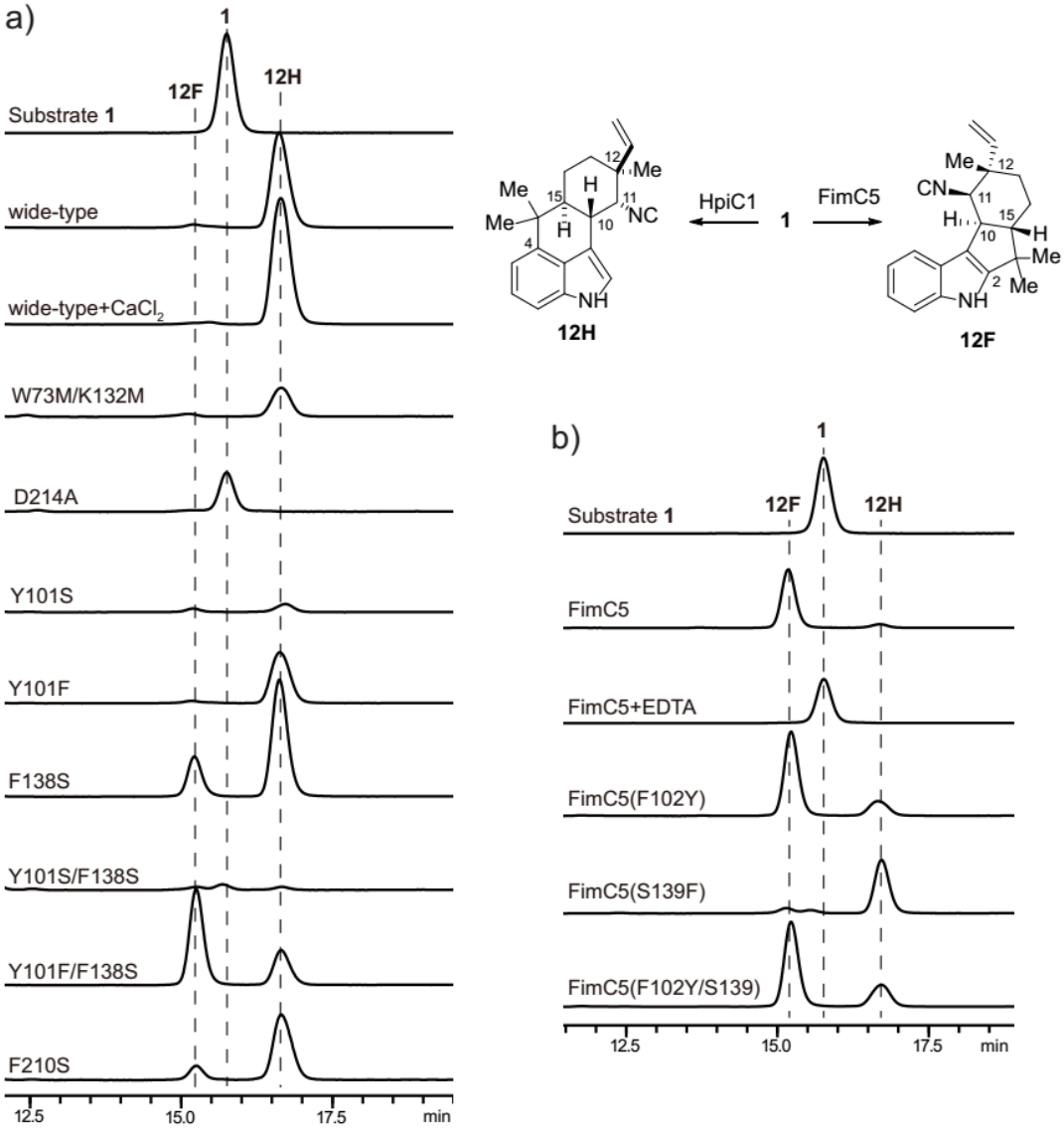


Figure 5-6. HPLC traces of in vitro assay. a) Characterization of HpiC1 mutants using **1** as substrate. Substitution of the catalytic acid D214 to alanine abolished activity. A single mutation F138S altered the native product profile of HpiC1 to produce **12F**. The product became predominantly **12F** in the HpiC1 Y101F/F138S double mutant. b) In

vitro reactions with FimC5. FimC5 is inactivated by 5 mM EDTA. The S139F variant produced predominantly the tricyclic 12-*epi*-hapalindole C. Identity of reaction products was confirmed by ^{1}H-NMR.

Additional mutations guided by sequence alignments were introduced into HpiC1 to identify key residues in the enzyme active site (**Figure 5-5**). Of particular interest was the Y101S/F138S HpiC1 variant that corresponds to the FamC3/HpiC3/FilC3 homologs. This mutant showed no activity with **1** in vitro (**Figure 5-6**a), which is consistent with the fact that no products were observed with the homodimeric forms of FamC3/HpiC3/FilC3[17]. The HpiC1 Y101S mutation also showed reduced activity, though to a lesser extent than in combination with F138S. Intriguingly, FamC3 was shown to associate with FamC2 as a heterodimer, and catalyze the formation of hapalindole H[17].

5.3.3 MD simulations

In order to understand the impact of these various mutations, high resolution structures were determined for HpiC1 variants Y101F, Y101S, F138S, and Y101F/F138S. While our efforts to observe bound ligands by either soaking or co-crystallization only afforded complexes with DMSO and Tris buffer (**SI Figure 5-4**), these structures significantly aided our efforts to interrogate the mechanism of cyclization in HpiC1 using computational methods. We first applied Molecular Dynamics (MD) simulations to the substrate-free structures to gain insights into the structure and dynamics of the active site and its impact on catalysis. Starting from the apo HpiC1 dimeric structure, analysis of the MD trajectories revealed large fluctuations of the loop containing F138 (N137-F150) (**SI Figure 5-6**), in agreement with the different conformations found for this loop in the crystal structures (**Figure 5-4**a). The flexibility of the loop, which is in part controlled and reduced by a calcium ion, induces conformational changes in the F138 active site residue. These data show that F138 primarily explores two conformations along the MD trajectory.

MD simulations show that D214, which is essential for enzyme activity, stays preferentially in a conformation in which the D214 side chain points

towards the inner cavity of the active site, while F138 acts as a wall on the side of the active site pocket (**Figure 5-7**a, **SI Figure 5-7**). This is due in part to the Y89 hydroxyl H-bond with the D214 carboxylic acid group (**SI Figure 5-7**, **SI Figure 5-8**). Importantly, the predicted pKa for D214, estimated from different snapshots obtained along the 500ns of MD simulation (**SI Figure 5-9**a), is 6.5 - 7.0, indicating that it can be protonated in an acid-base equilibrium to act as a protonating species during catalysis. Although the arrangement of the active site pre-organized for catalysis is the preferred one, an alternative conformation of F138 is sampled during the 500 ns trajectory. In this alternative conformation, the F138 side chain is displaced towards the center of the active site, and D214 becomes buried and inaccessible for interaction with the substrate, generating an inactive conformation (**SI Figure 5-7** and **8**). This alternative conformation is very similar to the active site arrangement observed in the crystal form 1 (**Figure 5-4**).

We next considered the absence of activity in the HpiC1 Y101S/F138S double mutant. MD simulations for Y101S/F138S in its apo state showed that the introduced serine Y101S interacts closely with the catalytic D214, in addition to the original Y89 hydroxyl H-bond (**SI Figure 5-10**, **SI Figure 5-11**). These two H-bonds between D214 and Y101S and Y89 favor stabilization of the negatively charged D214 carboxylate group, which will not be protonated in this more polar environment also favored by the neighboring F138S. This is confirmed by the decrease of the predicted pKa value of D214 (pKa ≈ 5.5 - 6) (**SI Figure 5-9***b*). Moreover, D214 becomes more buried in the active site cavity, preventing its interaction with the substrate (**SI Figure 5-11**).

As described above, F138S and Y101F/F138S mutants change the product profile in HpiC1 leading to increased formation of fischerindole **12F** (**Figure 5-6**a). We performed 500 ns MD simulations on both F138S and Y101F/F138S mutants, finding important changes in the active site pocket shape. The HpiC1 F138S mutation creates more space around the catalytic D214 residue (**Figure 5-7**b) compared to wild-type enzyme containing the original F138 residue. This mutation also releases the interaction between the two phenyl rings

of F138 and F210, with the F210 side chain becoming more flexible in HpiC1 F138S and Y101F/F138S mutants (**Figure 5-7**a-b, **SI Figure 5-12**). This active site reshaping is responsible for the change in the reaction outcome, as discussed below.

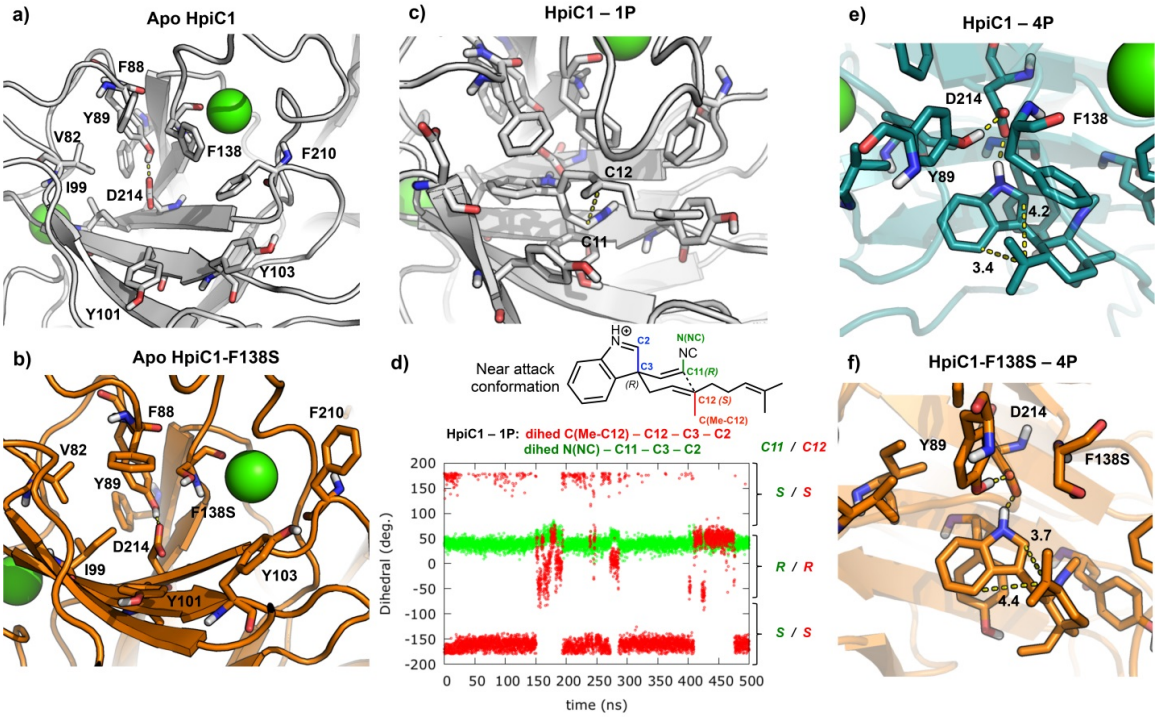


Figure 5-7. MD simulations of the active site. Representative snapshots of: (**a**) apo HpiC1; (**b**) F138S; (**c**) substrate **1**(*R*)**P** bound into HpiC1; (**d**) Dihedral angles explored during the 500 ns of MD simulation for substrate **1**(*R*)**P** bound into HpiC1. Dihedral-1 (N(NC) – C11 – C3 – C2) and dihedral-2 (C(Me-C12) – C12 – C3 – C2) define the relative orientation of substituents at C11 and C12 positions, respectively, during the MD simulation. The right axis indicates the final stereochemistry of C11 and C12 expected after the Cope rearrangement coming from the given near attack conformation (NAC) of **1**(*R*)**P**, as shown in (**d**) and **SI Figure 5-18**. **1**(*R*)**P** mainly explores one conformation during the MD trajectory leading to (*R*) configuration at C11 and (*S*) configuration at C12. Representative snapshots obtained from 500 ns MD simulations of the active site for intermediate **4P** bound to (**e**) HpiC1 (**SI Figure 5-19**, **SI Figure 5-20**); and (**f**) F138S (**SI Figure 5-21**). Distances (in Å) show that the conformation adopted by intermediate **4P** in the HpiC1 wild-type enzyme moves C16 closer to C4 to form the hapalindole product, while the F138S mutant enables the exploration of a conformation of intermediate **4P** in which C16 gets closer to C2 to allow fischerindole product formation.

Next, we employed density functional theory (DFT) calculations to explore the possible reaction mechanism for Stig cyclases, in particular the [3,3]-sigmatropic rearrangement (Cope rearrangement), which is the first step in the

HpiC1 catalyzed three-part reaction cascade starting from **1**[11]. The instability of precursor **1** has precluded determining its chiral configuration at the indole C3 position. Thus, we computed the Cope rearrangement and cyclization cascade mechanism (**Figure 5-8**) starting from both (*R*)-**1**, the enantiomer predicted to be accepted by HpiC1 (as discussed below), and from the (*S*)-**1** enantiomer (illustrated in **SI Figure 5-13**). The biosynthesis of hapalindole **12H** by HpiC1 and fischerindole **12F** by FimC5 (and related enzymes) is thought to occur in three mechanistic steps: (1) Cope rearrangement; (2) 6-*exo-trig* cyclization; and (3) electrophilic aromatic substitution (**Figure 5-8**).

The first proposed step in the biosynthesis of **12H** and **12F** is the Cope rearrangement of **1** to intermediate **3**. Although a typical Cope rearrangement has four possible transition states, two chairs and two boats (**SI Figure 5-14**, **SI Figure 5-15**), only one chair-like transition state can account for the known stereochemistry at positions C11 and C12 in the final products and, therefore, only this transition state was further considered. Given that the otherwise hydrophobic active site of HpiC1 contains the essential D214 residue, we explored the impact of this aspartic acid residue on accelerating the Cope rearrangement. The conversion of neutral starting material **1**(*R*)**N** to intermediate **3N** is concerted, proceeding through a single chair-like transition state **2N**, which lies 20.6 kcal/mol above the near attack conformation (NAC) of the starting material. This chair-like transition structure (**SI Figure 5-16**) has dissociative character, with breaking and forming partial single bonds of 2.58 Å and 2.53 Å, respectively. The partial negative charge on the isocyanidovinylindoline fragment of –0.43 *e* can be stabilized by a hydrogen bonding donor at the indole nitrogen, which has a partial negative charge of –0.49 *e*. Adding an acetic acid molecule to mimic possible hydrogen-bonding between D214 and the indole nitrogen lowers this barrier by 2.2 kcal/mol (40-fold rate enhancement), with the conversion of **1**(*R*)**A** to **3A** proceeding through a single chair-like transition state **2A** with a free energy barrier of 18.4 kcal/mol. These computations suggest that D214 could facilitate the Cope rearrangement of **1** to **3** by hydrogen-bonding to the indole nitrogen. Fully protonating the indole nitrogen, which represents the maximum

limit of potential acid catalysis by D214, results in a change of mechanism. The conversion from **1**(*R*)**P** to **3P** is stepwise and dissociative, with **2P** being an intermediate rather than a transition state, and has a much lower overall free energy barrier of 6.8 kcal/mol. Intermediate **2P** is stabilized by full conjugation between the indole and isonitrile groups, and by an allylic cation to produce this much lower overall barrier (**Figure 5-8**).

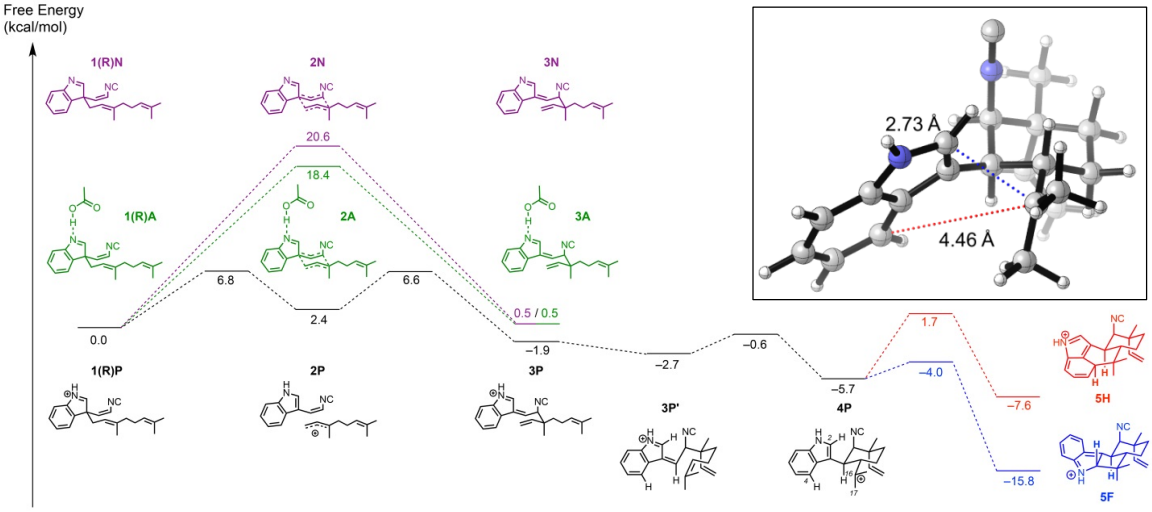


Figure 5-8. Quantum mechanics simulation. Cope rearrangement, 6-*exo-trig* cyclization, and electrophilic aromatic substitution cascade starting from the (*R*)-enantiomer of substrate **1** in a near attack conformation and leading to **12H** precursor **5H** and **12F** precursor **5F**. The energetics of the Cope rearrangement are computed with the neutral indole (pathway **N**), the *N*-protonated indole (pathway **P**), and the indole forming a hydrogen bond with acetic acid (pathway **A**). Inset: Optimized geometry of key intermediate **4P** that undergoes regioselective electrophilic aromatic substitution to form **12H** or **12F**.

The second proposed step in the biosynthesis of **12H** and **12F** is the 6-*exo-trig* cyclization of intermediate **3** to intermediate **4**, which sets the stereochemistry at positions C10 and C15 in the final products. In gas-phase DFT optimizations, it was possible to locate a transition state for this cyclization only when the indole nitrogen was protonated. Without protonation, the zwitterionic character of the possible transition state leads to bond formation between negatively-charged position C3 of the indole and positively-charged position C16 to generate a cyclobutane ring. In contrast, the protonated species undergoes facile cyclization from intermediate **3P** to intermediate **4P** through a

low-lying transition state. This suggests that protonation is crucial and that D214 may catalyze cyclization in this way, and possibly the preceding Cope rearrangement as well.

Continuing along the protonated pathway, the third step is electrophilic aromatic substitution of intermediate **4P**, with two different transition states leading to the two major products. Electrophilic aromatic substitution at position C4 of the indole yields intermediate **5H**, which gives **12H** upon deprotonation, while electrophilic aromatic substitution at position C2 of the indole yields intermediate **5F**, which gives **12F** upon deprotonation. Rapid deprotonation at the methyl position (C16) instead of electrophilic aromatic substitution can lead to formation of tricyclic **12C** (**Figure 5-1**), which is generated as a trace product by HpiC1[26]. QM calculations show that formation of **12F**, which has the lower-energy transition state, should be intrinsically favored. Thus, the regioselectivity of electrophilic aromatic substitution to generate **12H** appears to be controlled by the HpiC1 active site as opposed to inherent energetics of the system.

To understand enzymatic control of the initial Cope rearrangement, we conducted 500 ns MD trajectories on the wild-type HpiC1 with both **1**(*R*) and **1**(*S*) substrate enantiomers as their protonated states, **1**(*R*)**P** and **1**(*S*)**P,** respectively, bound into the enzyme active site (**SI Figure 5-17**). Both enantiomers retain the key H-bond interaction between the protonated indole NH and the D214 residue for the entirety of the MD trajectories. However, the C11-C12 distance, which corresponds to the new C-C bond formed during the Cope rearrangement, is shorter (~3.5Å) for the **1**(*R*)**P** substrate than for the **1**(*S*)**P** substrate (>4.0Å). In addition, substrate **1**(*R*)**P** is stabilized by the enzyme active site in a near attack conformation (NAC) that leads to the correct stereochemistry at positions 11 and 12, (*R*) and (*S*), respectively (**Figure 5-7***c-d*). In contrast, substrate **1**(*S*)**P** never explores a NAC that could lead to the correct stereochemistry at positions 11 and 12 observed in the **12H** natural product (**SI Figure 5-17**, **SI Figure 5-18**). Based on these observations, the (*R*) enantiomer **1**(*R*)**P** is the most plausible natural substrate for the HpiC1 enzyme.

To understand how the HpiC1 active site could control the formation of the hapalindole/fischerindole products, and how select mutations can alter the product distribution, we performed additional MD simulations with intermediates prior to electrophilic aromatic substitution bound into the active site of the HpiC1 enzyme and associated mutants. We analyzed the impact of the active site in dictating the regioselectivity of the reaction (to form hapalindole or fischerindole cores) by considering the two intermediate precursors **4P** and **8P**, characterized during our QM modeling, that derive from the **1**(*R*)**P** and **1**(*S*)**P** starting materials, respectively. Since either intermediates **4P** or **8P** can generate both the final **12H** and **12F** products, we analyzed the binding of both intermediates in the HpiC1 active site. MD simulations with the two docked intermediates show that intermediate **4P**, derived from the **1**(*R*)**P** substrate, is more effectively bound into the active site, maintaining the H-bond between the protonated NH-indole and D214 more consistently than the **8P** intermediate, which rapidly dissociates during the simulation (**SI Figure 5-19**). These results reinforce the idea that the **1**(*R*)**P** substrate is the most plausible natural substrate for the HpiC1 enzyme. Moreover, MD simulations showed that when the **4P** intermediate is bound into the active site, it adopts a conformation in which the distance between C4 and C16 (~3.5Å) that corresponds to hapalindole product formation is shorter than the distance between C2 and C16 (~4.0Å) that corresponds to fischerindole product formation (**Figure 5-7**e, **SI Figure 5-19**). Although the intrinsic fischerindole-forming TS is lower in energy than the hapalindole-forming TS, the enzyme active site promotes the formation of the intrinsically less-favored hapalindole product. Finally, late during the MD simulation (around 200 ns) a conformational change of the F138 side chain occurs, which reverses this trend to finally disrupt the interaction of **4P** and D214 (**SI Figure 5-19**, **SI Figure 5-20**), highlighting the key role of F138 residue in controlling the enzyme activity and also the site-selectivity of the reaction.

The critical role of F138 has been investigated through further MD simulations on the F138S and Y101F/F138S mutants with **4P** intermediate bound. In both trajectories, the distance between C2 and C16 leading to

fischerindole product formation is shortened compared to the wild-type enzyme, becoming similar and slightly closer than the distance between C4 and C16 leading to hapalindole product formation, consistent with the experimentally observed product ratios (**Figure 5-7**f, **SI Figure 5-21**). The absence of the bulky F138 residue near the C2 position allows the intermediate to adopt a slightly different conformation when it is interacting with the D214 residue (Fig. 5*f*, **SI Figure 5-22**), enabling **12F** formation without completely suppressing generation of **12H**.

5.4 Discussion and Conclusion

The structure of HpiC1 has provided the first high resolution insights into a fascinating mechanistic puzzle in which the cyanobacterial indole alkaloid Stig cyclases are able to generate extensive stereochemical and regiochemical diversity through the central biosynthetic precursor **1**. The surprising function of the Stig cyclases[11] could not be inferred from bioinformatic analysis and annotation, and similarly, structural and mechanistic insights from homology-based tools were unavailable for these remarkable biocatalysts. HpiC1 is not homologous with any characterized terpene cyclase, but instead is most similar to bacterial carbohydrate binding modules (CBM). CBMs function primarily to bring various hydrolytic enzymes into contact with their carbohydrate substrates. Target sugar molecules bind the CBMs in an extended cleft at the protein surface, which is mediated through several amino acids that are not conserved in HpiC1 (**Figure 5-2**c)[22]. This indicates a divergent functionality in HpiC1 that is based on a shared protein scaffold. Nevertheless, Stig cyclases and CBMs share several properties including a common fold, thermostability (**SI Figure 5-23**)[27], and structural calcium ions, which play an important role in CBM stabilization, substrate recognition, and oligomerization[27-28].

The molecular basis for the Ca^{2+} dependence of HpiC1 was demonstrated through two well-ordered binding sites near the enzyme active site. A paradoxical aspect of this calcium requirement is the observation that low millimolar concentrations of $CaCl_2$ caused HpiC1 to precipitate. This effect was reversible

by stoichiometric addition of EDTA, indicative of calcium-dependent higher order oligomerization of HpiC1. However, a requirement for millimolar concentrations of Ca^{2+} in the reaction buffer to achieve activity has been reported for some cyclases[18]. There are also fascinating instances of heteromeric association of some Stig cyclases leading to variant stereochemical outcomes of the products, compared to the products of their homomeric counterparts[17-18]. Taken together these data are indicative of an important, yet complex structural role for Ca^{2+} in the Stig cyclase catalysis that may involve higher order oligomerization. Because HpiC1 could not be crystallized without supplemental $CaCl_2$, we examined the crystal lattice contacts for any evidence of additional calcium binding sites and found two sites in which calcium could facilitate formation of higher-order complexes. We observed one fully occupied interfacial calcium ion far from the active site in crystal form 1 (**SI Figure 5-24**), and a second Ca^{2+} at half occupancy in unrelated crystal forms 1 and 2 (**SI Figure 5-25**). This Ca^{2+} is adjacent to the active sites indicating that a bridging calcium ion could influence the activity of cyclase oligomers. A functional role of higher order oligomerization in HpiC1 remains unclear as the addition of supplemental (1 – 20 mM) Ca^{2+} enhances the enzymatic activity of HpiC1, but is not required for turnover (**Figure 5-6**a). Further structural investigation will be required to understand the assembly of heteromeric Stig cyclase complexes, and the relevance of Ca^{2+} in those cases.

We have established the location of the HpiC1 enzyme active site, and using mutational analysis identified critical residues for catalysis and demonstrated a key relationship between amino acid sequence and product outcome. Most importantly, we identified D214 as the source of an active site acid that is required for catalysis, and is consistent with numerous reports of acid catalyzed Cope rearrangements[24]. Notably, the hydrophobic environment around D214 is essential to maintain a suitable population of the protonated species, as the Y101S/F138S mutation significantly reduced enzyme activity. We also identified a key regiochemical switch at the F138 position that gave rise to production of the fischerindole core in HpiC1. We also confirmed the general importance of F138 in the HpiC1 homolog FimC5, where a corresponding

mutation also affected the product distribution with respect to regiochemistry (**Figure 5-6**b). Together these findings will facilitate our efforts to anticipate the product profiles in both uncharacterized cyclases and new ones that are uncovered as additional strains and gene clusters are discovered through sequencing of relevant microbial genomes.

We also explored the HpiC1 active site dynamics using a combination of DFT quantum mechanical calculations of mechanism and modes of catalysis, and MD simulations on the protein. We studied the origins of the three-part catalytic mechanism, and how HpiC1 controls the regiochemistry of product formation by favoring a particular conformation of substrate **1** and the reaction intermediate **4P**. We have also examined the role of key mutations in HpiC1 that switch the native product outcome from hapalindole **12H** to fischerindole **12F**. Together, these results address several of the catalytic steps in the formation of **12H** from **1**(*R*). Further analysis will be required to establish the mechanistic basis by which the variant Stig cyclases achieve differentiation at the 6-*exo*-*trig* cyclization step, where the stereocenters at C10 and C15 are set (e.g. Hapalindole U, H, and J series) (**SI Figure 5-1**). Structural studies on additional Stig cyclases, mutational analysis across key active site residues, and computational modeling of the reaction intermediates together will enable prediction of product profiles, and engineering of new selectivities to diversify further this remarkable family of natural products.

5.5 Experimental section

5.5.1 Protein preparation and in vitro assay

Cloning and mutagenesis of HpiC1 and FimC5. HpiC1 and FimC5 were cloned into pET28 (Novagen) from codon optimized synthetic genes (IDT gBlocks®) with their N-terminal leader peptides truncated[17]. Site-directed mutagenesis of HpiC1 and FimC5 was performed using a single primer method based on "Quikchange" mutagenesis (Agilent Genomics). Mutagenic primer sequences are listed in Supplementary Table 1. All mutations were verified by DNA sequencing at the University of Michigan DNA Sequencing Core.

Expression of HpiC1 and FimC5 proteins. HpiC1 and FimC5 and their corresponding active site mutants were overexpressed in *Escherichia coli* strain BL21 (DE3). Cultures from a single colony were used to inoculate 1.5 L terrific broth (TB) supplemented with 50 μg/mL kanamycin. Expression was induced with 0.7 mM isopropyl-β-D-thiogalactopyranoside when cultures reached OD_{600} ~ 1.0. After 20 h induction at 18 °C the cells were harvested by centrifugation and stored at -80 °C.

Expression of HpiC1 W73M/K132M selenomethionine derivative. An initial challenge involved the lack of native methionine residues in HpiC1. Therefore, a series of mutants containing methionine substitutions were screened for crystallization. Sites for substitution were selected based on positions containing a native methionine in sequence comparisons with other Stig cyclases. Selenomethionine (SeMet) HpiC1 W73M/K132M was produced by metabolic inhibition[29]. Briefly, freshly transformed BL21 (DE3) cells harboring the hpiC1 gene on pET28 were used to incolulate 3 L M9 minimal medium supplemented with 50 μg/mL kanamycin. An amino acid cocktail containing L-selenomethionine was added when the cells reached OD_{600} = 1.0. The cells were cooled to 18 °C and shaken for 30 min prior to induction with 0.7 mM isopropyl-β-D-thiogalactopyranoside. After 20 h induction the cells were harvested by centrifugation and stored at -80 °C.

Purification of recombinant proteins. All proteins were purified as described previously[17]. Briefly, 10 g *E. coli* wet cell mass containing the recombinant cyclase was resuspended in 75 mL lysis buffer (10 mM HEPES pH 7.6, 50 mM NaCl, 10% glycerol). Cells were lysed by the addition of lysozyme (0.5 mg/mL) and sonication and clarified by centrifugation at 60,000 x*g* for 25 min. Clarified lysate was loaded by gravity onto 8 mL NiNTA Superflow resin (Qiagen). The column was washed with 100 mL lysis buffer containing 20 mM imidazole and 50 mL lysis buffer containing 40 mM imidazole. The proteins were eluted with elution buffer (250 mM imidazole, pH 7.9 and 10% glycerol). Fractions containing the purified cyclase were concentrated using Amicon Ultra 15 centrifugal filters and desalted using PD-10 columns (GE Healthcare) equilibrated with storage buffer (10 mM HEPES pH 7.6, 50 mM NaCl). The purified cyclases were drop-frozen in 30 µL aliquots directly into liquid N_2 and stored at -80 °C.

In vitro cyclase assays. In vitro assays were performed with HpiC1, FimC5 and their corresponding active site mutants as described previously[17]. Briefly, FamD2, GPP, Indole isonitrile, cyclase, $CaCl_2$. Products were analyzed using LC/MS (Shimadzu) using C18 (Agilent) HPLC column.

Scaleup, purification, and NMR of F138S product (12-epi-fischerindole U). The semi-prep scale reaction was performed as described previously[17].

5.5.2 Process of crystallization

Crystallization of SeMet HpiC1 W73M/K132M. Single, diffraction quality crystals of the HpiC1 W73M/K132M selenomethionine derivative were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 11 mg/mL SeMet HpiC1 in storage buffer with 1 µL of a well solution containing 23% PEG 3350, 200 mM $CaCl_2$, 5% trehalose. Sitting droplets were nucleated after 18 h from an earlier spontaneous crystallization event using a cat whisker. Single, rod-shaped crystals grew to approximate dimensions of 50 x 50 x 250 µm after 14 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 23% PEG 3350, 200 mM $CaCl_2$, 9.1% trehalose

was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. SeMet HpiC1 crystallized in Form 1, space group $P2_12_12_1$ with unit cell dimensions of *a* = 44.9 Å, b = 81.1 Å, *c* = 131.7 Å, and two chains in the asymmetric unit.

Crystallization of Native HpiC1 ($P4_2$). Single, diffraction quality crystals of HpiC1 native were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 20 mg/mL HpiC1 in storage buffer with 1 µL of a well solution containing 22% PEG 4000, 200 mM $CaCl_2$, 100 mM Tris pH 8.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, rod-shaped crystals grew to approximate dimensions of 50 x 50 x 150 µm after 7 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 200 mM $CaCl_2$, 100 mM Tris pH 8.5, 15% ethylene glycol was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. In these conditions, HpiC1 native crystallized in Form 2, space group $P4_2$ with unit cell dimensions of *a* = 71.3 Å, *b* = 71.3 Å, *c* = 80.6 Å, and two chains in the asymmetric unit.

Crystallization of Native HpiC1 (C2). Single, diffraction quality crystals of HpiC1 native were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 20 mg/mL HpiC1 in storage buffer and 5% DMSO with 1 µL of a well solution containing 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate dimensions of 200 x 200 x 100 µm after 7 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 15% ethylene glycol, 5% DMSO was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. In these conditions, HpiC1 native crystallized in Form 3, space group *C*2 with unit

cell dimensions of *a* = 113.8 Å, *b* = 49.5 Å, *c* = 53.1 Å, α = 90°, β = 110.5°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 Y101F. Single, diffraction quality crystals of HpiC1 Y101F were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 15 mg/mL protein in storage buffer with 1 µL of a well solution containing 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate dimensions of 250 x 250 x 270 µm after 7 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 22% PEG 4000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 15% ethylene glycol was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 Y101F crystallized in Form 3, space group *C*2 with unit cell dimensions of *a* = 113.8 Å, *b* = 49.8 Å, *c* = 53.4 Å, α = 90°, β = 110.4°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 Y101S. Single, diffraction quality crystals of HpiC1 Y101S were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 15 mg/mL protein in storage buffer with 1 µL of a well solution containing 20% MEPEG 5000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, diamond-shaped crystals grew to approximate dimensions of 250 x 250 x 270 µm after 7 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 20% MEPEG 5000, 150 mM $CaCl_2$, 100 mM Tris pH 8.5, 15% ethylene glycol, 5% DMSO was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 Y101S crystallized in Form 3, space group *C*2 with unit cell dimensions of *a* = 113.9 Å, *b* = 49.6 Å, *c* = 53.4 Å, α = 90°, β = 110.3°, γ = 90° and one chain in the asymmetric unit.

Crystallization of HpiC1 F138S and Y101F/F138S. Single, diffraction quality crystals of HpiC1 F138S and Y101F/F138S were grown in Intelli-Plate 96-2 shallow well plates (Hampton research) at 20 °C by mixing 1 µL of 15 mg/mL protein in storage buffer, 20 mM $CaCl_2$, 5% DMSO with 1 µL of a well solution containing 20% MEPEG 5000, 100 mM BisTris pH 6.5, 5% ethylene glycol. Sitting droplets were nucleated after 4 h from an earlier spontaneous crystallization event using a cat whisker. Single, plate-shaped crystals grew to approximate dimensions of 50 x 50 x 300 µm after 7 days. 8 µL of a cryoprotecting solution containing 10 mM HEPES pH 7.6, 50 mM NaCl, 20% MEPEG 5000, 20 mM $CaCl_2$, 100 mM BisTris pH 6.5, 15% ethylene glycol, 5% DMSO was added directly to the sitting drops and the crystals were harvested using nylon loops and vitrified by rapid plunging into liquid nitrogen. HpiC1 F138S and HpiC1 Y101F/F138S crystallized in Form 4, space group $P2_1$ with unit cell dimensions of a = 62.0 Å, b = 47.9 Å, c = 174.2 Å, α = 90°, β = 97.2°, γ = 90° and four chains in the asymmetric unit.

Data collection and processing. X-ray data were collected at 100 K on beamline 23ID-B at the General Medical Sciences and Cancer Institutes Structural Biology Facility at the Advanced Photon Source in Argonne, IL, USA. Diffraction data were integrated and scaled using XDS[30]. Data collection statistics are given in Supplementary Table 2.

Experimental phasing (SAD) and molecular replacement, model building and refinement. The structure of SeMet HpiC1 W73M/K132M was solved using single wavelength anomalous diffraction (SAD). Phasing and initial model building were performed using Phenix Autosol[31]. This resulted in an initial model that could be extended by alternating cycles of manual building in *Coot*[32] and least-squares refinement with Refmac[33]. The structures for HpiC1 native and Y101F, Y101S, F138S, Y101F/F138S were solved by molecular replacement using Phaser-MR[34] with the structure of the HpiC1 selenomethionine derivative as a search model. Final models were generated by alternating cycles of manual

building in *Coot*[32] and refinement in Refmac[33] and Phenix[31], and were validated using MolProbity[35].

*Docking 12-epi-hapalindole U with Autodock VINA.*12-*epi*-hapalindole U was docked into the SeMet HpiC1 model using Autodock VINA[23]. Default parameters for Autodock VINA were used with the exception of exhaustiveness, which was set to 100.

Molecular graphics. All figures depicting the protein structure were generated with PyMOL[25].

5.5.3 QM /MD analysis

Quantum mechanical calculations. Conformational searches of the hapalindole and fischerindole products were performed using the Schrödinger MacroModel[36] software package, and the lowest energy conformation was used for all reported quantum mechanical calculations. All quantum mechanical calculations were performed using the Gaussian 09[37] software package. Structures were optimized in the gas phase at the B3LYP[38-39]/6-31G(d) level of theory; frequency calculations were used to confirm the presence of local minima (no imaginary frequencies) and transition states (one imaginary frequency) and to calculate free energies at 298 K. To obtain more accurate energetics, single-point energy calculations were performed on the optimized structures at the B3LYP/6-311++G(d,p) level of theory using Grimme's D3(BJ) dispersion correction[40-41] and the IEFPCM[42-44] solvent model for diethyl ether ($\varepsilon = 4$). The use of the dielectric constant $\varepsilon=4$ has proven to be a good model to estimate the dielectric permittivity in the enzyme active site, accounting for electronic polarization and small backbone fluctuations[45-46].

Molecular Dynamics simulations. Molecular Dynamics simulations were performed using the GPU code (*pmemd*)[47] of the AMBER 16 package[48]. Parameters for intermediates and substrates were generated within the *antechamber* module using the general AMBER force field (*gaff*)[49], with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by

the RESP model[50]. The charges were calculated according to the Merz–Singh–Kollman scheme[51-52] using the Gaussian 09 package[36]. Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P[53] water molecules using the *leap* module, resulting in the addition of around 15,000 solvent molecules. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber14 force field (*ff14sb*)[54]. A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the solute by a harmonic potential with a force constant of 500 $kcal \cdot mol^{-1} \cdot Å^{-2}$ and the second stage minimizes all the atoms in the simulation cell. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was kept fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method[55]. An 8 Å cutoff was applied to Lennard–Jones and electrostatic interactions. Harmonic restraints of 10 $kcal \cdot mol^{-1}$ were applied to the solute and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated without restrains for 2 ns with a 2 fs time step at a constant pressure of 1 atm and temperature of 300 K. After the systems were equilibrated in the NPT ensemble, subsequent MD simulations were performed for an additional 500 ns under an NVT ensemble and periodic-boundary conditions.

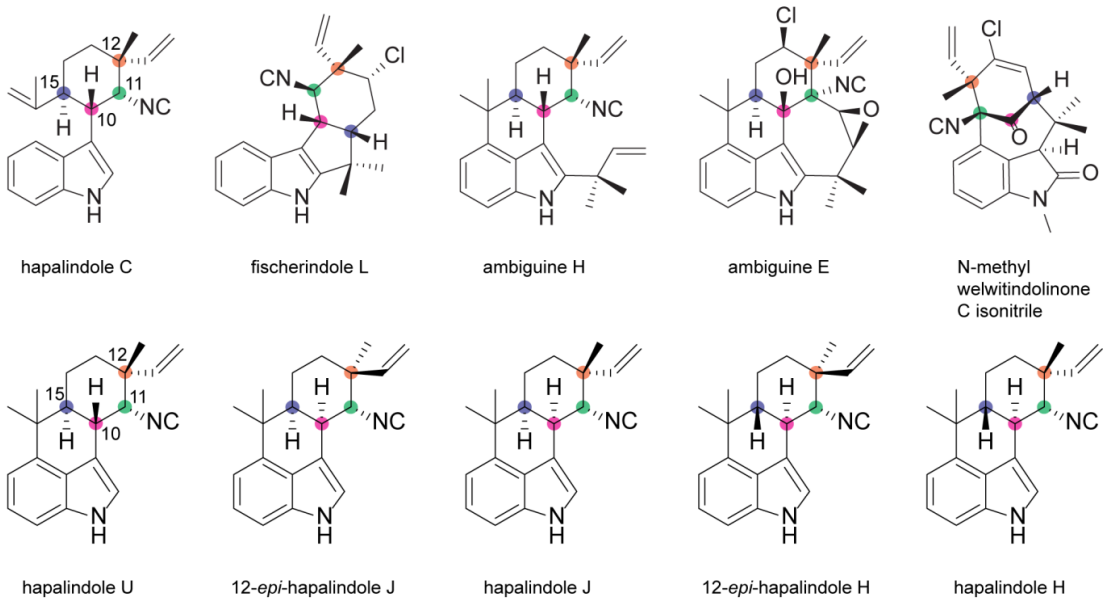
5.6 Supplementary information

SI Table 5-1. Mutagenic primer sequences (5'→3').

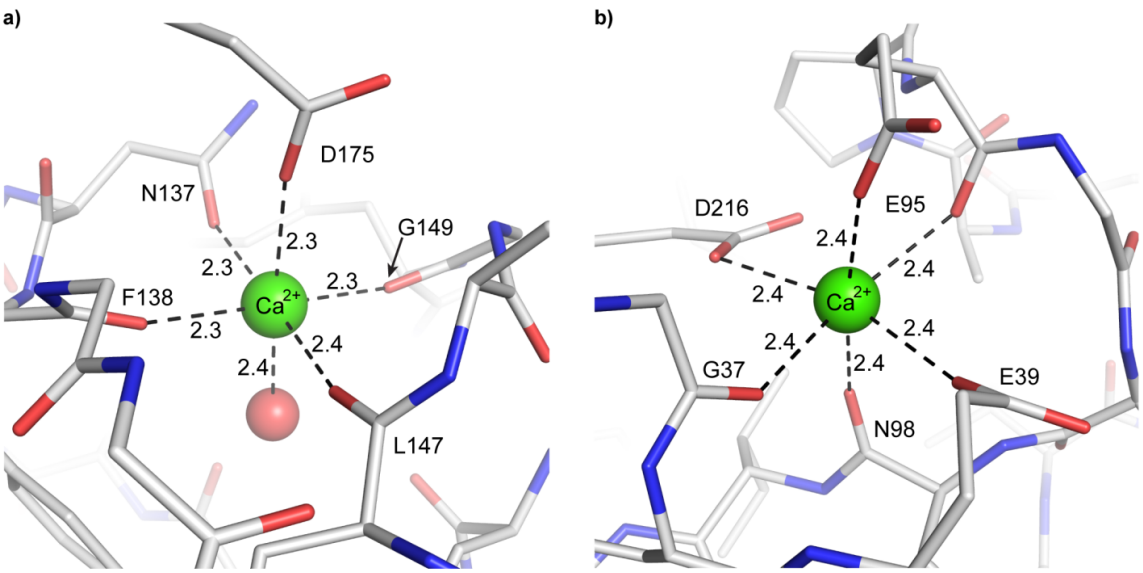
Gene	Primer
SN_hpiC1_W73M	CTGGTTCCGGAGAAGCGTACCACCatgACCAGCAACAA CGGTGTGGGCTAC
SN_hpiC1_K132M	GGAACCGGATACCAAGTACACCCTGatgGTGGACGTTG GTAACTTCGGTGGC
SN_hpiC1_Y101S	CGCCGGAAGGTCGTAACATCGGCtctATTTATCTGGCGC AGAAACCGGG
SN_hpiC1_Y101F	CGCCGGAAGGTCGTAACATCGGCttcATTTATCTGGCGC AGAAACCGGG
SN_hpiC1_F138S	CCCTGAAAGTGGACGTTGGTAACtctGGTGGCGAGTTTC AGAAAATTAGCC
SN_hpiC1_F210S	CTGATTAACCTGCTGCAAGGCACCtctAGCGGCCTGGAC TTTGATAACGTG
SN_hpiC1_D214A	CAAGGCACCTTCAGCGGCCTGgcgTTTGATAACGTGCG TCTGACCGTTG
SN_fimC5_F102Y	CCGGAAGGCCGTAACGTGGCGtatGTTTACCTGGCGCA GGAGATCGG
SN_fimC5_S139L	CAAATATACCCTGACCGTTGATATTGGTAACctgGGTGG CAGCTTCCAGGGCAAAAC

SI Table 5-2. Data collection and refinement statistics.

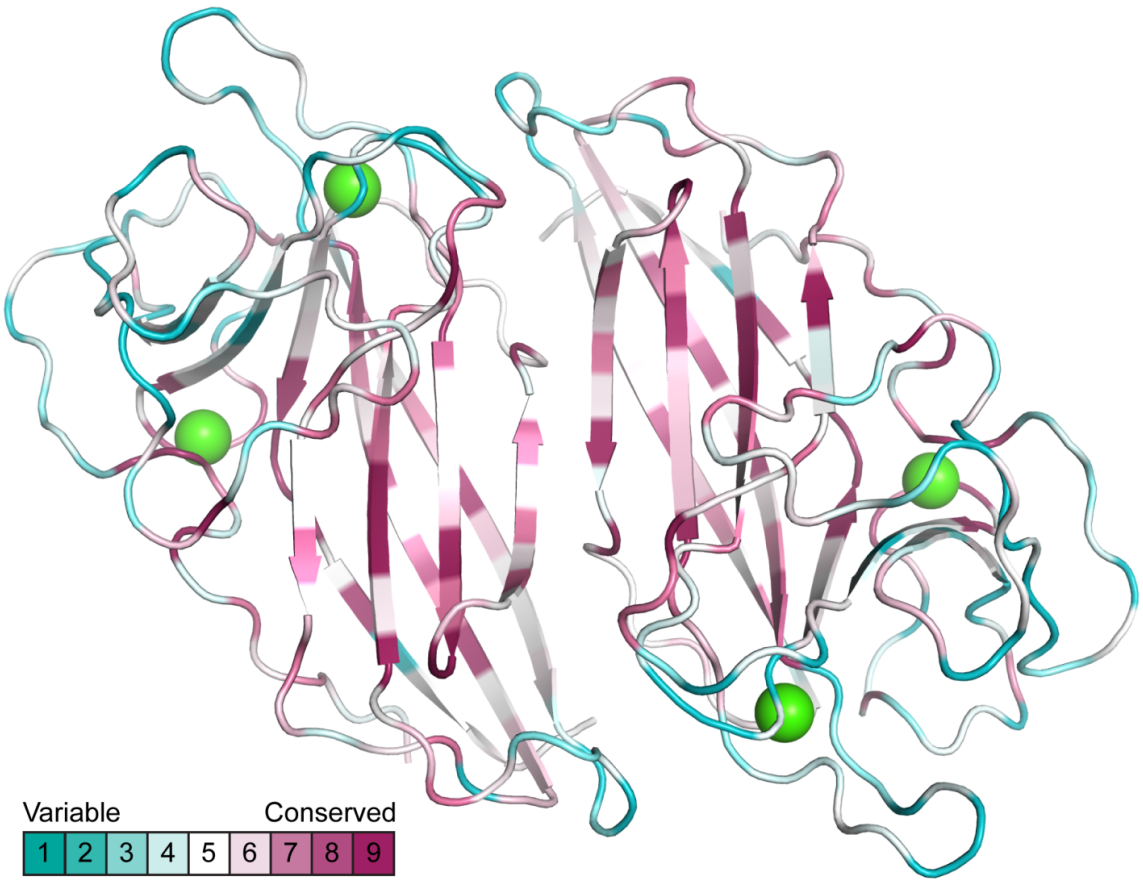
	Form 1 HpiC1 SeMet	Form 2 HpiC1 native	Form 3 HpiC1 native	Form 3 HpiC1 Y101F	Form 3 HpiC1 Y101S	Form 4 HpiC1 F138S	Form 4 HpiC1 Y101F/F138S
PDB	5WPP	6AL6	5WPR	5WPS	5WPU	6AL7	6AL8
Ca^{2+} (mM)	200	200	<150	<150	<150	20	20
Wavelength (Å)	0.979	1.033	1.033	1.033	1.033	1.033	1.033
d_{min}	39.3 - 1.7 (1.76 - 1.70)	42.76 - 2.09 (2.16 - 2.09)	45.11 - 1.49 (1.54 - 1.49)	36.28 - 1.39 (1.44 - 1.39)	45.21 - 1.39 (1.44 - 1.39)	46.2 - 1.69 (1.75 - 1.69)	46.22-1.64 (1.70-1.64)
Space group	$P2_12_12_1$	$P4_2$	C2	C2	C2	$P2_1$	$P2_1$
Unit cell (Å)	a=44.9 b=81.1 c=131.7	a=71.3 b=71.3 c=80.6	a=113.8 b=49.5 c=53.1 β=110.5°	a=113.8 b=49.8 c=53.4 β=110.4°	a=113.9 b=49.6 c=53.4 β=110.3°	a=62.0 b=47.9 c=174.2 β=97.2°	a=62.0 b=48.0 c=174.8 β=97.0°
Total reflections	359,857 (36,743)	323,567 (32,999)	300,958 (24,314)	337,601 (14,570)	334,789 (14,285)	782,992 (76,532)	812,940 (49,680)
Unique reflections	53,600 (5260)	24,033 (2401)	45,278 (4440)	51,294 (2965)	52,971 (3352)	112,827 (10,916)	122,677 (9983)
Multiplicity	6.7 (7.0)	13.5 (13.7)	6.6 (5.5)	6.6 (4.9)	6.3 (4.3)	6.9 (7.0)	6.6 (5.0)
Completeness (%)	99.5 (99.0)	99.8 (98.9)	99.7 (97.6)	90.5 (51.3)	93.6 (58.6)	97.8 (95.4)	97.7 (79.9)
Mean I/sigma(I)	16.87 (2.04)	18.94 (1.62)	19.41 (1.30)	20.44 (0.66)	18.24 (0.92)	16.37 (1.11)	13.42 (0.97)
Wilson B-factor	24.0	44.4	21.4	20.6	18.8	32.4	29.5
R-merge	0.069 (0.815)	0.104 (1.287)	0.050 (0.984)	0.050 (1.496)	0.055 (0.9158)	0.054 (1.653)	0.065 (1.020)
R-meas	0.075 (0.880)	0.108 (1.336)	0.054 (1.088)	0.052 (1.672)	0.060 (1.044)	0.060 (1.784)	0.071 (1.142)
CC1/2	0.999 (0.887)	0.999 (0.632)	0.999 (0.659)	0.999 (0.537)	0.999 (0.844)	0.999 (0.737)	0.998 (0.808)
R-work	0.181	0.210	0.164	0.137	0.149	0.218	0.218
R-free	0.215	0.255	0.184	0.167	0.182	0.252	0.245
Number of atoms	3483	3066	1794	1892	1861	6370	6618
protein	3024	2953	1547	1556	1535	5957	6102
ligand	33	6	14	18	10	9	9
Solvent	426	107	233	318	316	404	507
RMS(bonds)(Å)	0.005	0.008	0.005	0.005	0.005	0.009	0.009
RMS(angles)(°)	0.81	0.99	0.83	0.82	0.80	0.99	0.93
Average B-factor	29.1	46.4	25.7	26.6	24.5	53.3	50.7
Ramachandran favored (%)	99.24	97.91	99.49	100.00	100.00	97.78	97.06
Allowed (%)	0.76	1.83	0.51	0.00	0.00	2.22	2.81
Outliers (%)	0.00	0.26	0.00	0.00	0.00	0.00	0.13



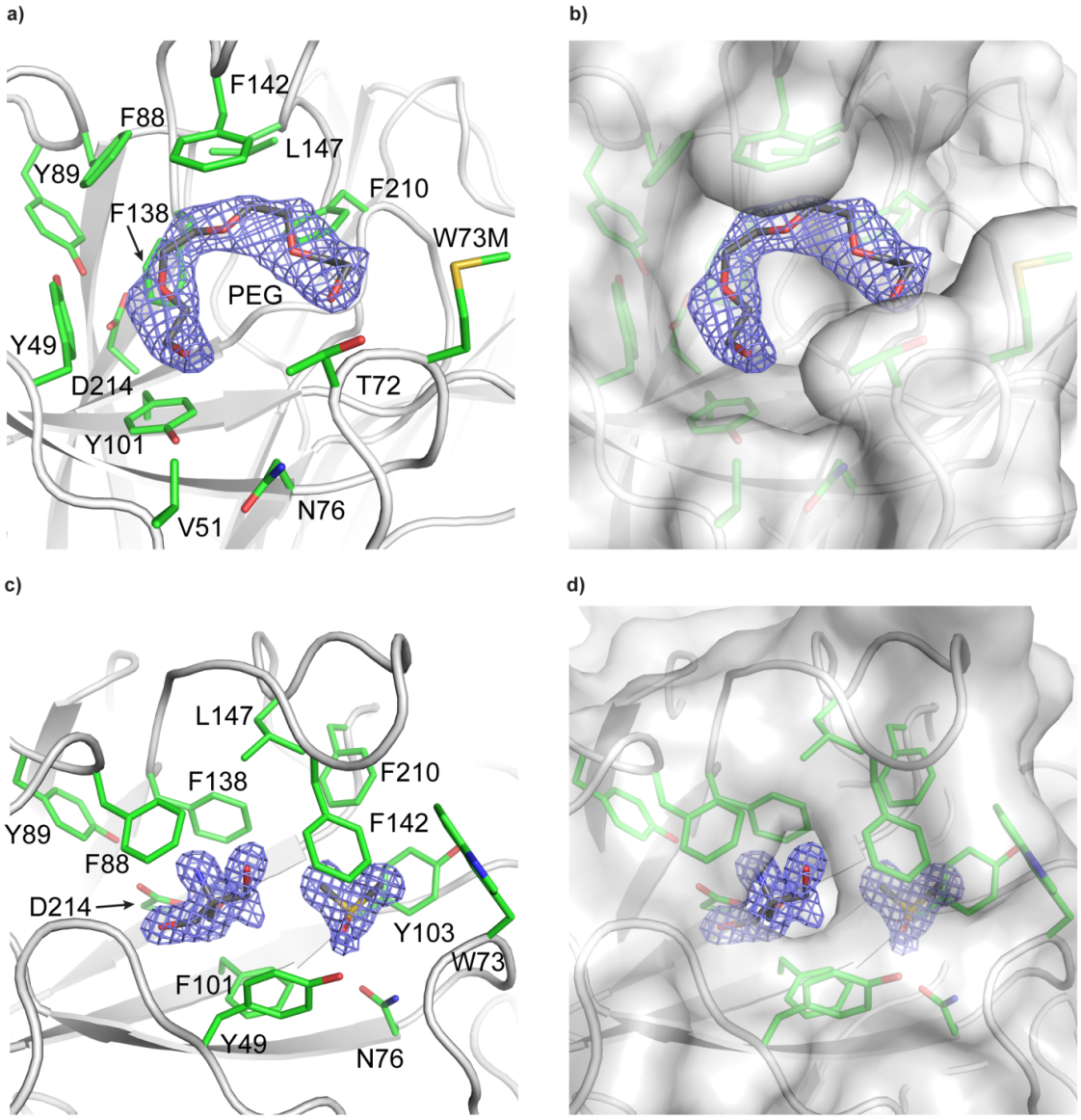
SI Figure 5-1. Representative hapalindole-type alkaloids. Tricyclic, tetracyclic, and pentacyclic species are shown. Stereochemical differences among the hapalindole U, J, and H series are highlighted.



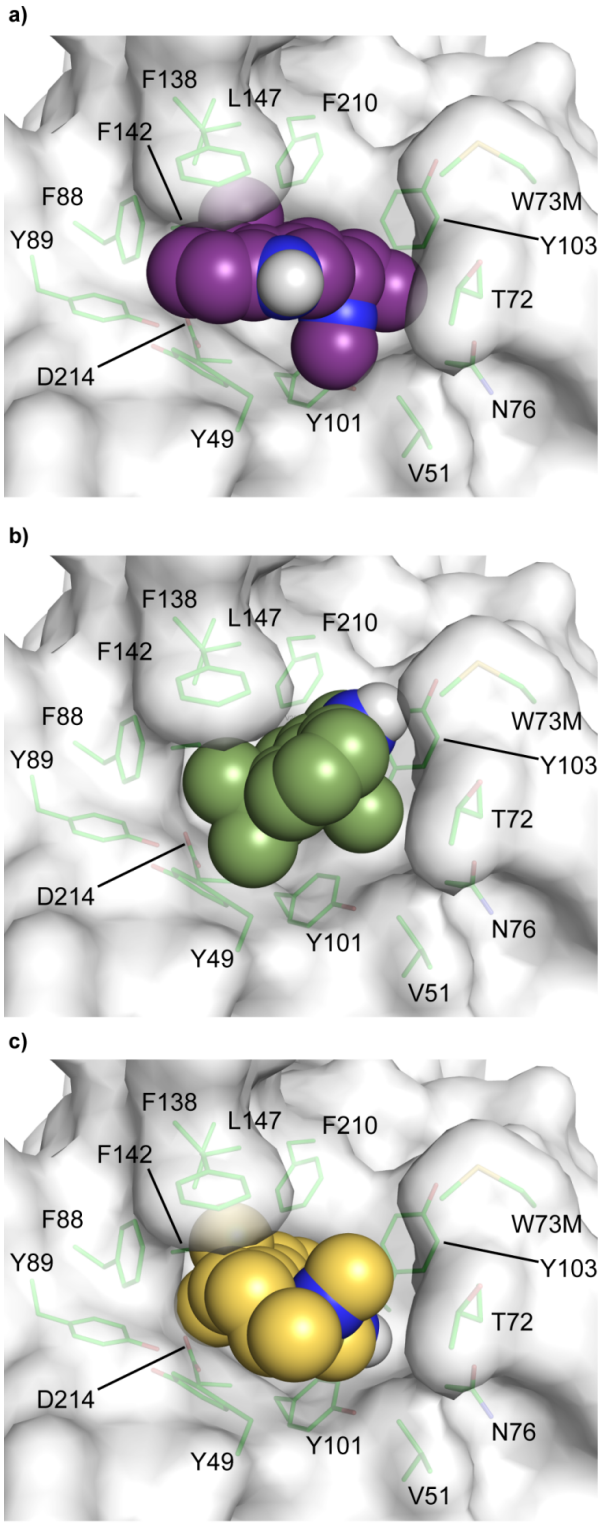
SI Figure 5-2. Two hexacoordinate Ca^{2+} ions are present in each HpiC1 monomer. (**a**) The first site is closer to the active site, and is ligated by carbonyl oxygen atoms from F138, L147, G149, and aspartate/asparagine side chain atoms from N137 and D175. A water molecule is the sixth ligand. (**b**) A second Ca^{2+} is ligated by carbonyl oxygen atoms from G37, E95, N98, and carboxylate/amide side chain atoms from E39, E95, and D216.



SI Figure 5-3. ConSurf Analysis of HpiC1[56]. ConSurf sequence conservation projected onto the HpiC1 structure.High conservation is most evident in the secondary structure elements including the dimer interface, while the active site region shows higher variability. Amino acid positions for which conservation could not be determined (20 out of 200 residues) are colored white. The Uniprot-90 database was used to generate the alignment.

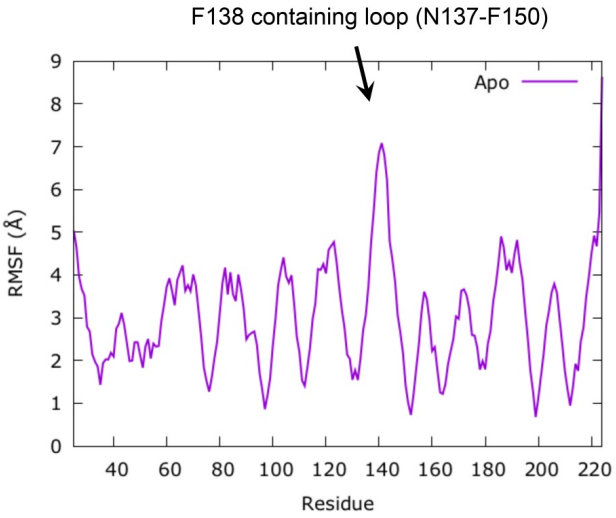


SI Figure 5-4. Polyethylene glycol (4 units) modeled into difference density in the SeMet HpiC1 W73M/K132M active site. The 1.7 Å omit electron density (Fo-Fc, 3.0σ contour) was calculated from a model where polyethylene glycol was omitted from the phase calculation and refinement. The active site is shown in cartoon representation (**a**) and surface representation (**b**) with key residues highlighted as green sticks. DMSO and Tris were modeled into difference density in the HpiC1 Y101F active site. The 1.4 Å omit electron density (Fo-Fc, 3.0σ contour) was calculated from a model where DMSO and Tris were omitted from the phase calculation and refinement. The active site is shown in cartoon representation (**c**) and surface representation (**d**) with key residues highlighted as green sticks.

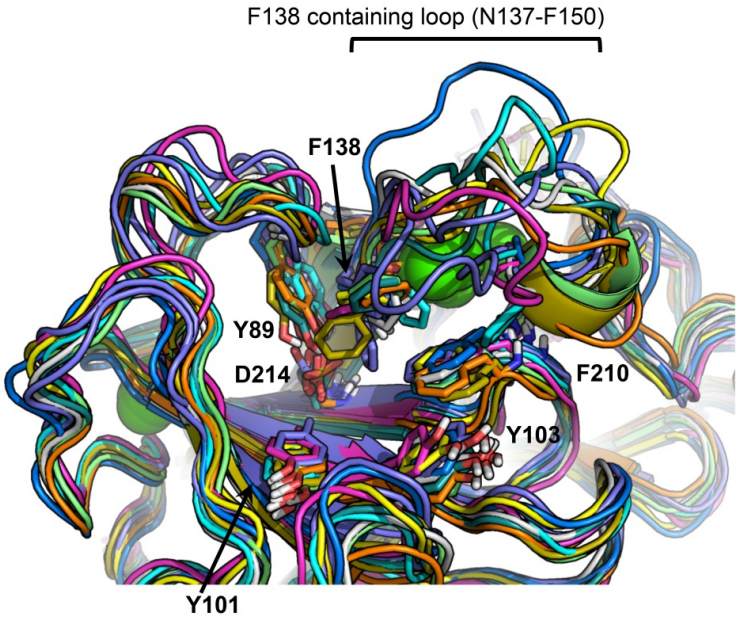


SI Figure 5-5. Autodock VINA[23] was used to examine if the binding pocket was appropriately sized for 12-epi-hapalindole U. Several docked structures of similar energy were generated (**a-c**): -9.7 kcal/mol, -9.4 kcal/mol, -9.2 kcal/mol.

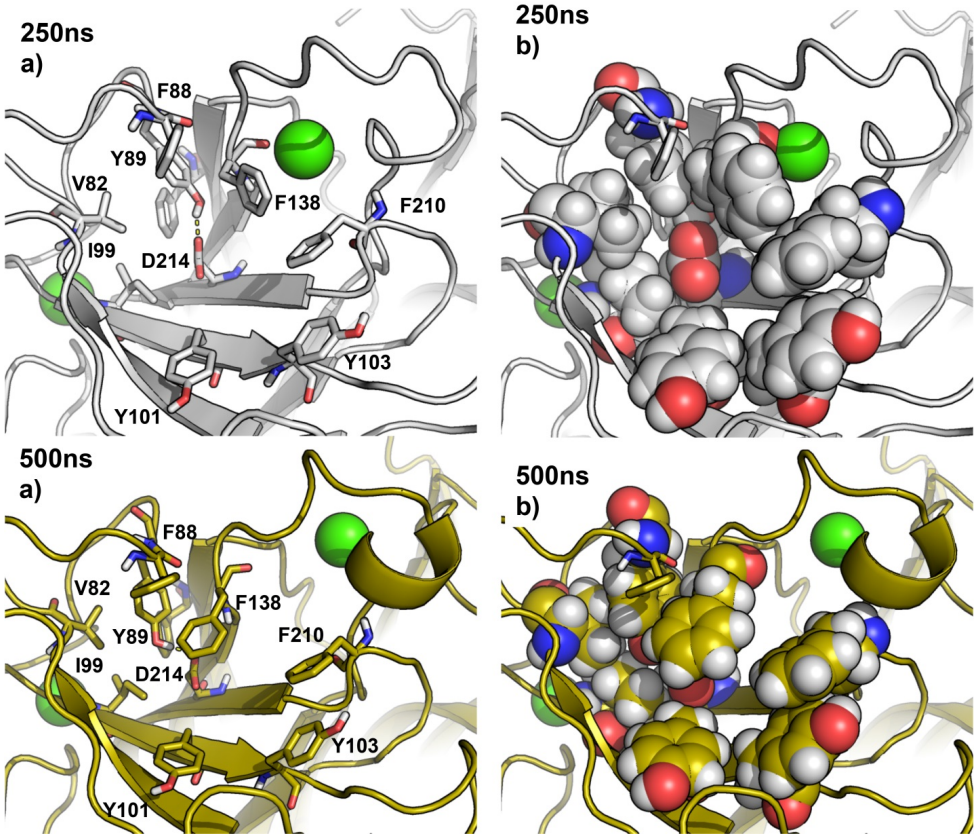
a)



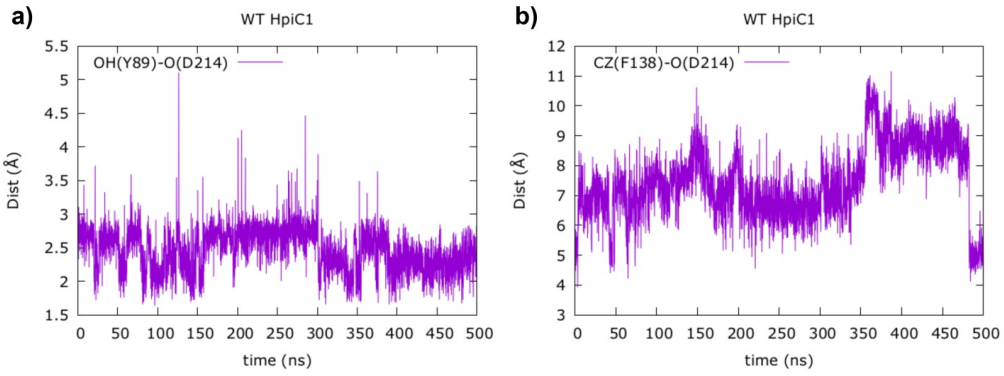
b)



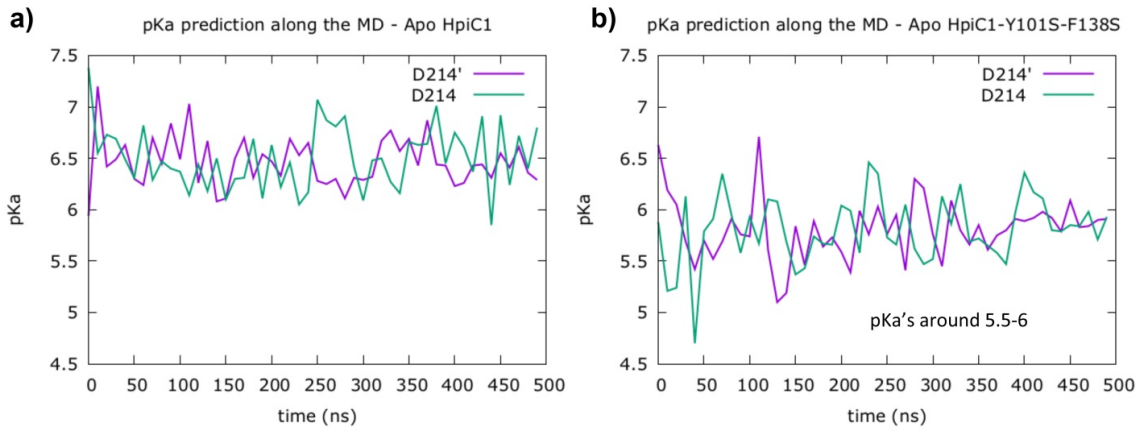
SI Figure 5-6. (**a**) RMSF measured over 500 ns MD simulation for the apo HpiC1 enzyme. (**b**) Overlay of representative snapshots obtained from this MD simulation for apo wild-type.



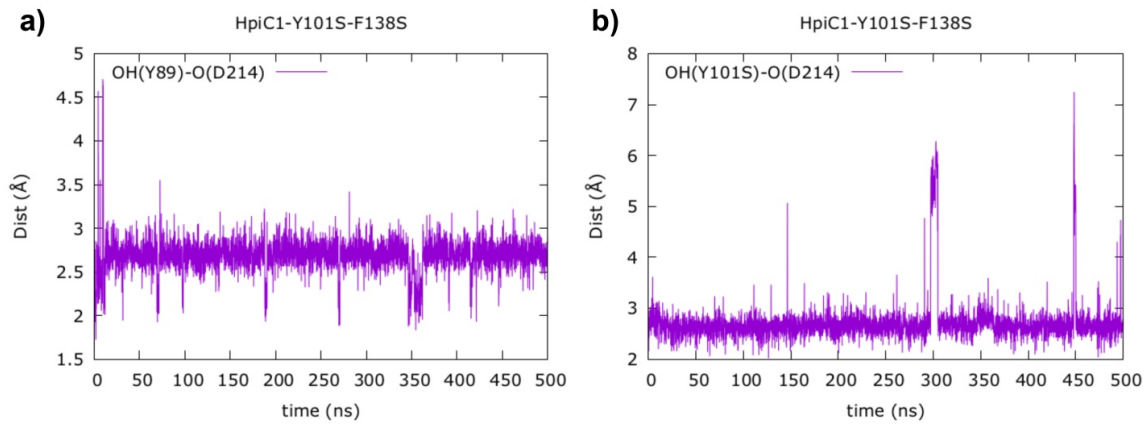
SI Figure 5-7. Representative snapshots (at 250 ns and 500 ns) of the active site arrangement and the two conformations of F138 side chain observed during the 500 ns of MD trajectory for the apo wild-type HpiC1 enzyme. The most important active site residues are represented (**a**) as sticks, and (**b**) in space-filling representation.



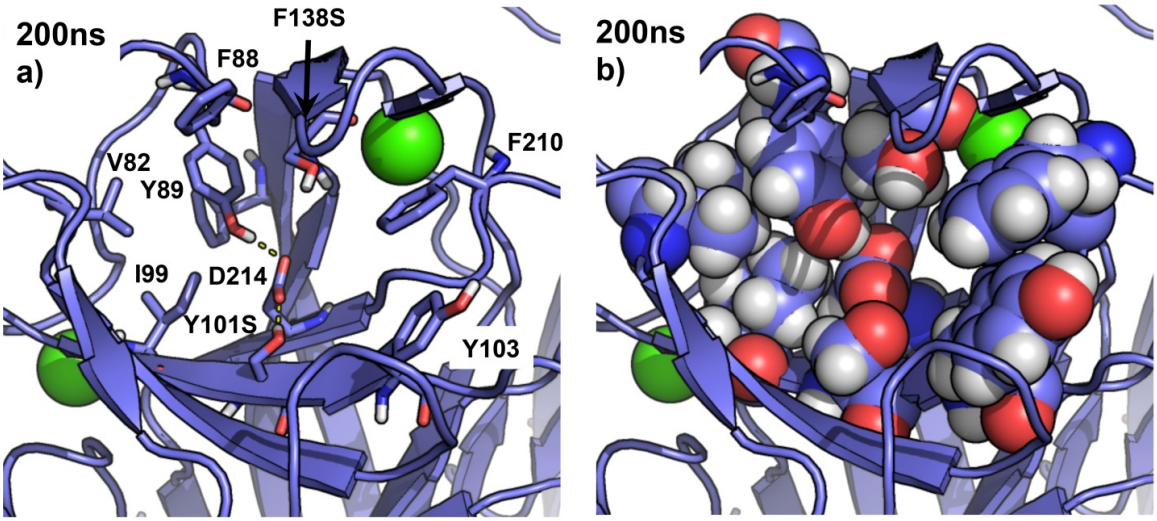
SI Figure 5-8. Distances measured along the 500 ns MD trajectoryfor the apo wild-type HpiC1 enzyme between (**a**) the center of mass of the two oxygen atoms of the D214 carboxylate group and the HO(Y89); and (**b**) Cz from the F138 side chain and the center of mass of the two oxygen atoms of the D214 carboxylate group.



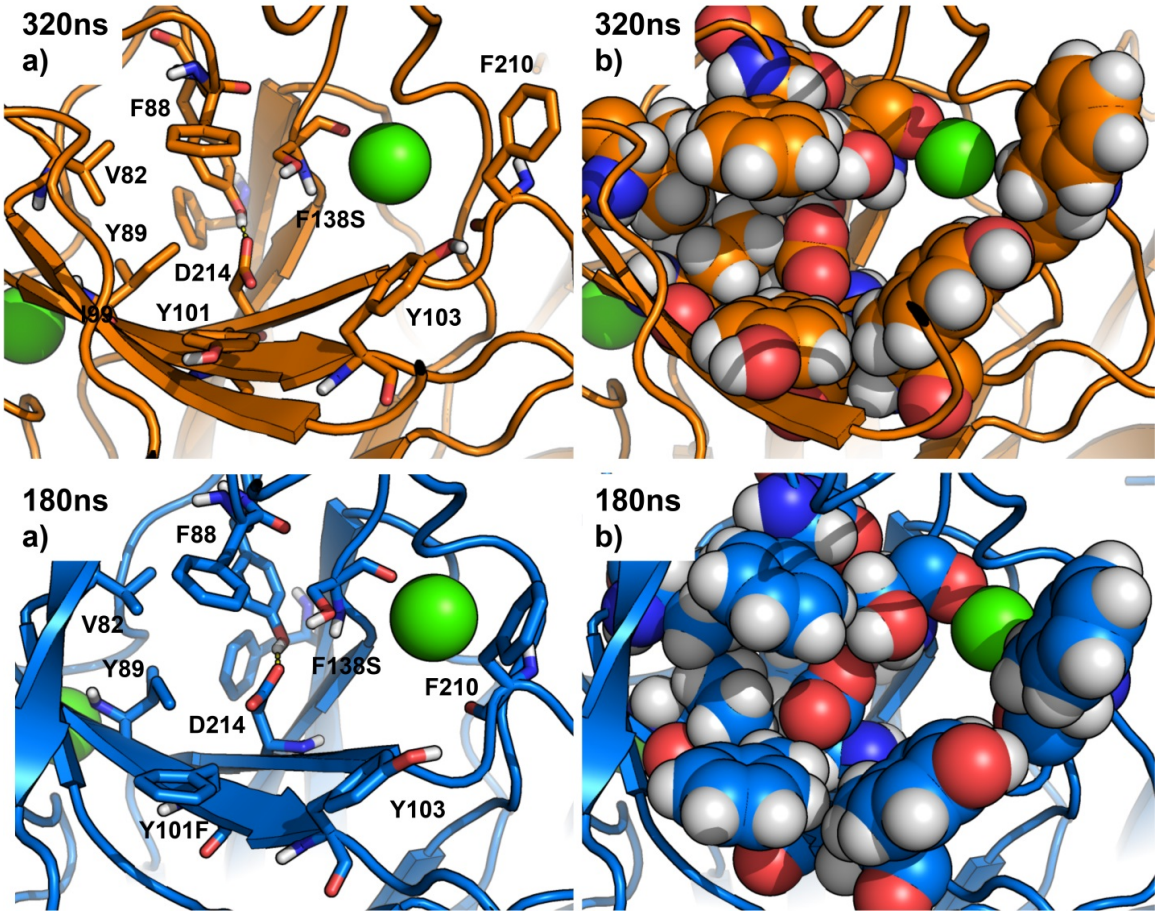
SI Figure 5-9. pK_a predictions for catalytic D214 and D214' in the apo state of the dimeric (**a**) wild-type HpiC1 enzyme and the (**b**) Y101S/F138S double mutant, for different snapshots obtained during the 500 ns MD trajectories. pKa estimations are obtained from the Propka3.1 program[57-58].



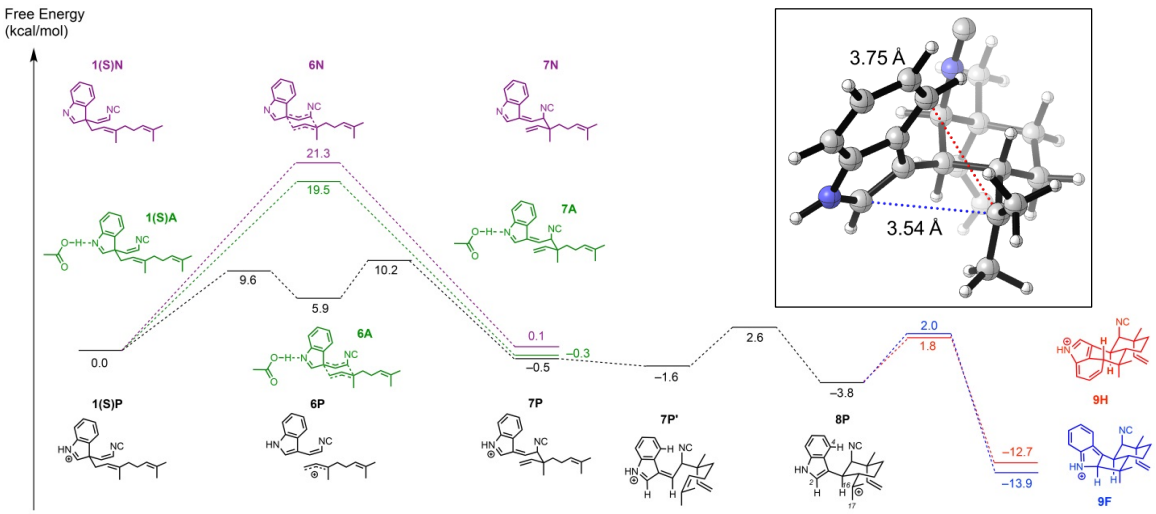
SI Figure 5-10. Distances measured along the 500 ns MD trajectory for the apo HpiC1 Y101S/F138S double mutant between the center of mass of the two oxygen atoms of the D214 carboxylate group and the (**a**) HO(Y89); and (**b**) HO(Y101S) side chains.



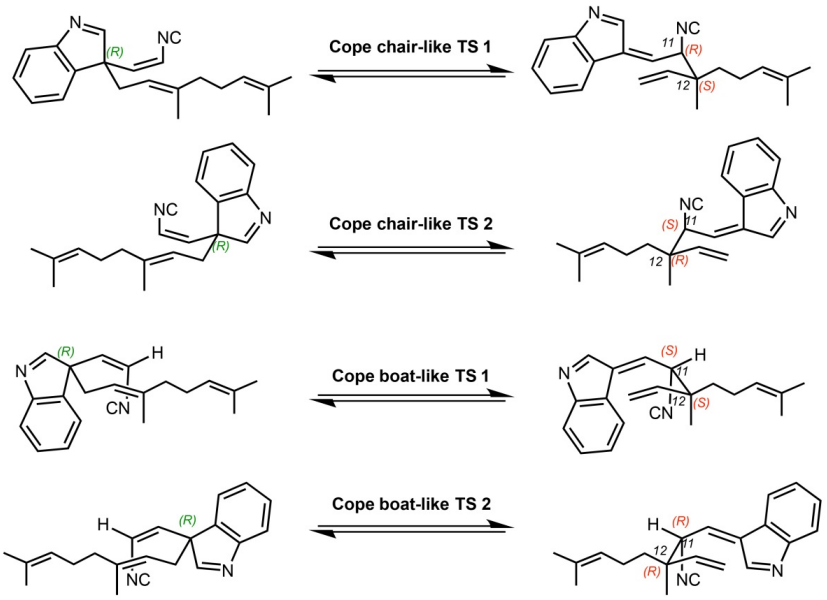
SI Figure 5-11. Representative snapshot (at 200 ns) of the active site arrangement in the MD trajectory of the apo HpiC1 F138S/Y101S double mutant. The most important active site residues are represented (**a**) as sticks, and (**b**) in space-filling representation.



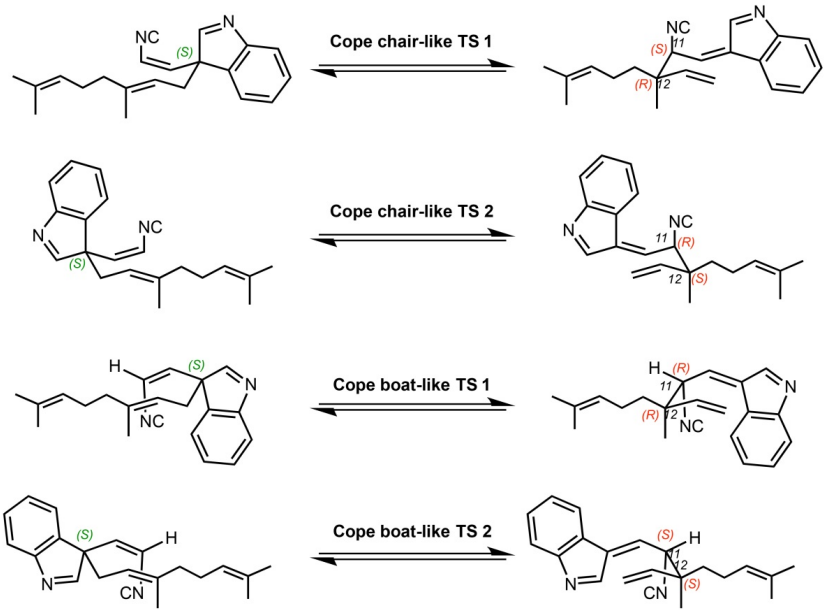
SI Figure 5-12. Representative snapshots of the active site arrangement observed during the 500 ns of MD trajectories for the HpiC1 F138S (top in orange, at 320 ns) and Y101F/F138S (bottom in blue, at 180 ns) mutants. The most important active site residues are represented (**a**) as sticks, and (**b**) in a space-filling representation.



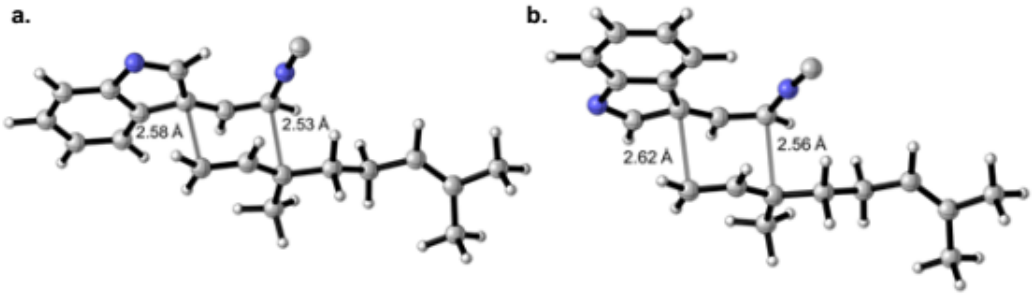
SI Figure 5-13. Quantum mechanics simulation of (*S*)-enantiomer of starting material **1.** Cope rearrangement, cyclization, and electrophilic aromatic substitution cascade starting from the (*S*)-enantiomer of starting material **1**(*S*) in a near-attack conformation and leading to 12-*epi*-hapalindole U precursor **9H** and 12-*epi*-fischerindole U precursor **9F**. The energetics of the Cope rearrangement are computed with the neutral indole (pathway **N**), the *N*-protonated indole (pathway **P**), and the indole forming a hydrogen bond with acetic acid (pathway **A**). Inset: Optimized geometry of key intermediate **8P** which undergoes regioselective electrophilic aromatic substitution to form 12-*epi*-hapalindole U or 12-*epi*-fischerindole U.



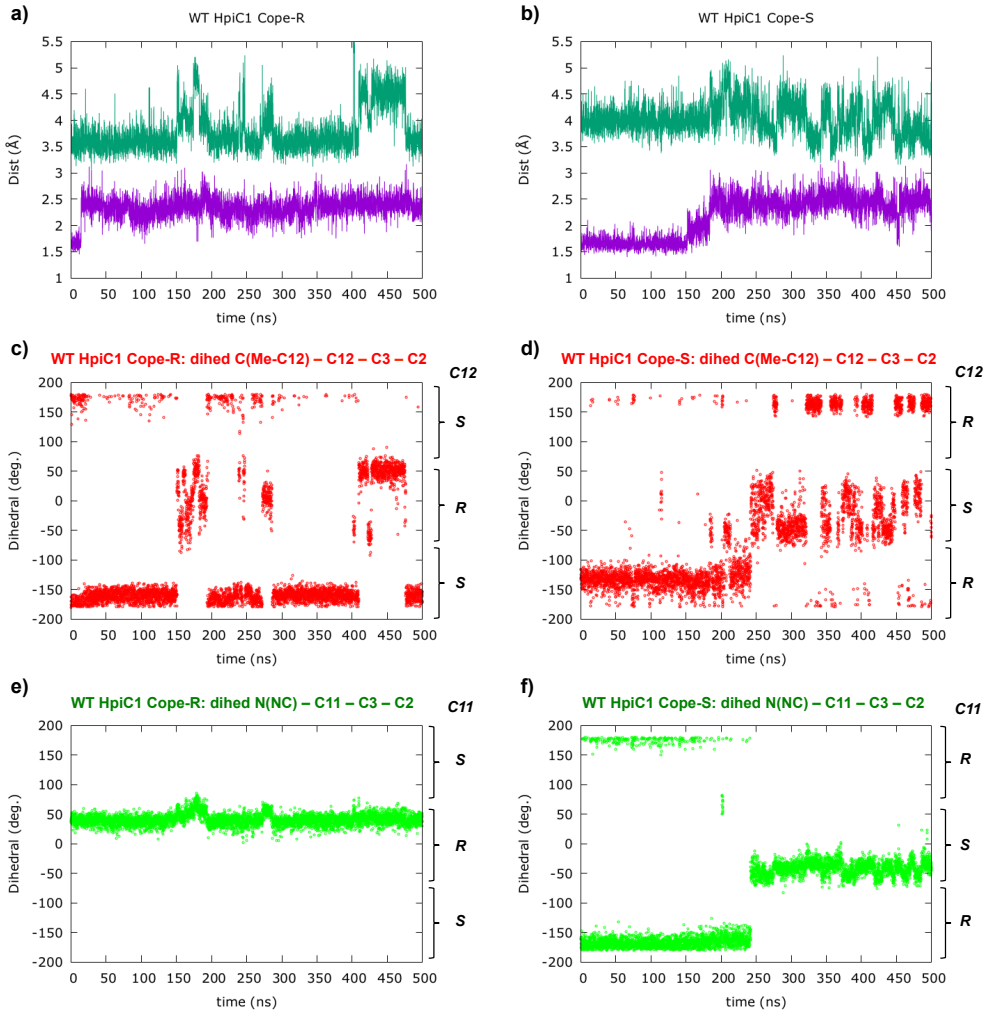
SI Figure 5-14. The four possible chair-like and boat-like transition states for the Cope rearrangement starting from the (*R*)-enantiomer of the starting material. Only chair-like TS 1 yields a product consistent with the known stereochemistry of 12-*epi*-hapalindole U and 12-*epi*-fischerindole U at positions 11 and 12, so only this possibility was investigated computationally (**Figure 5-8**).



SI Figure 5-15. The four possible chair-like and boat-like transition states for the Cope rearrangement starting from the (*S*)-enantiomer of the starting material. Only chair-like TS 2 yields a product consistent with the known stereochemistry of 12-*epi*-hapalindole U and 12-*epi*-fischerindole U at positions 11 and 12, so only this possibility was investigated computationally (**SI Figure 5-13**).

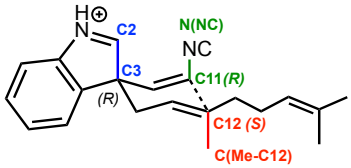


SI Figure 5-16. Optimized Cope rearrangement transition states (**a**) **2N** coming from the **1**(*R*)**N** starting material and (**b**) **6N** coming from the **1**(*S*)**N** starting material. Both transition states are chair-like with dissociative character, featuring breaking and forming partial single bonds with long lengths ranging from 2.53 Å to 2.62 Å.



MD simulations predict that (*R*) substrate **1**(*R*)**P** is stabilized by the enzyme active site in a near attack conformation (NAC) that leads to the C11-(*R*) and C12-(*S*) configurations in the Cope rearrangement product, which corresponds to the stereochemistry of 12-*epi*-hapalindole U. On the other hand, (*S*) substrate **1**(*S*)**P** does not explore any conformation that could lead to the C11-(*R*) and C12-(*S*) configurations during the entire 500 ns MD trajectory.

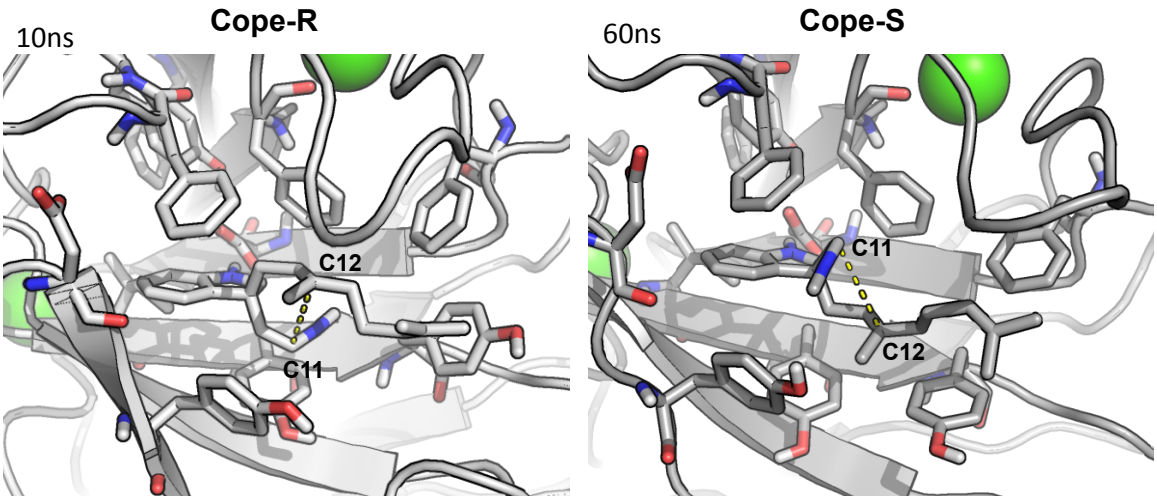
Dihedral angles explored:



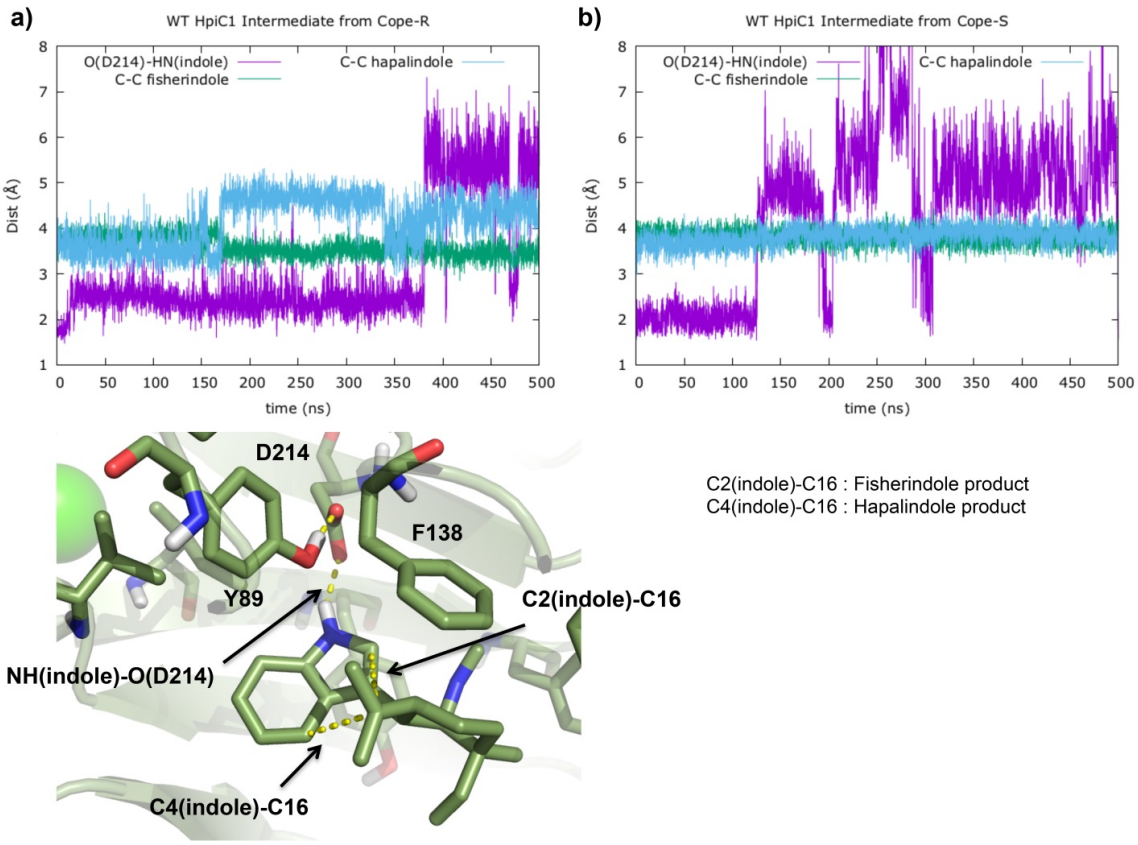
SI Figure 5-17. MD simulations. (**a**) and (**b**) show the distances between HN(indole) - O(D214) (*purple*), and C11-C12 (*green*) measured along the 500 ns MD trajectories for wild-type HpiC1 with (*R*) substrate **1**(*R*)**P** (**a**, left) and (*S*) substrate **1**(*S*)**P** (**b**, right) bound. During the simulations, the indole ring is protonated on nitrogen while D214 side chain is in the deprotonated carboxylate form. The **1**(*R*)**P** substrate mantains a shorter C11-C12 distance than **1**(*S*)**P**, which is the new C-C bond formed during the Cope rearrangement.

(c) and (**d**) show the dihedral angle defined by C(Me-C12) – C12 – C3 – C2 explored during the 500 ns MD simulation for both substrate **1**(*R*)**P** (**c**, left) and **1**(*S*)**P** (**d**, right) bound complexes. This dihedral angle describes the relative orientation of the alkyl chain substituent with respect to the indole ring. On the right of each plot, the stereochemistry at **C12** in the product of the Cope rearrangement that comes from this particular near attack conformation (NAC) is reported.

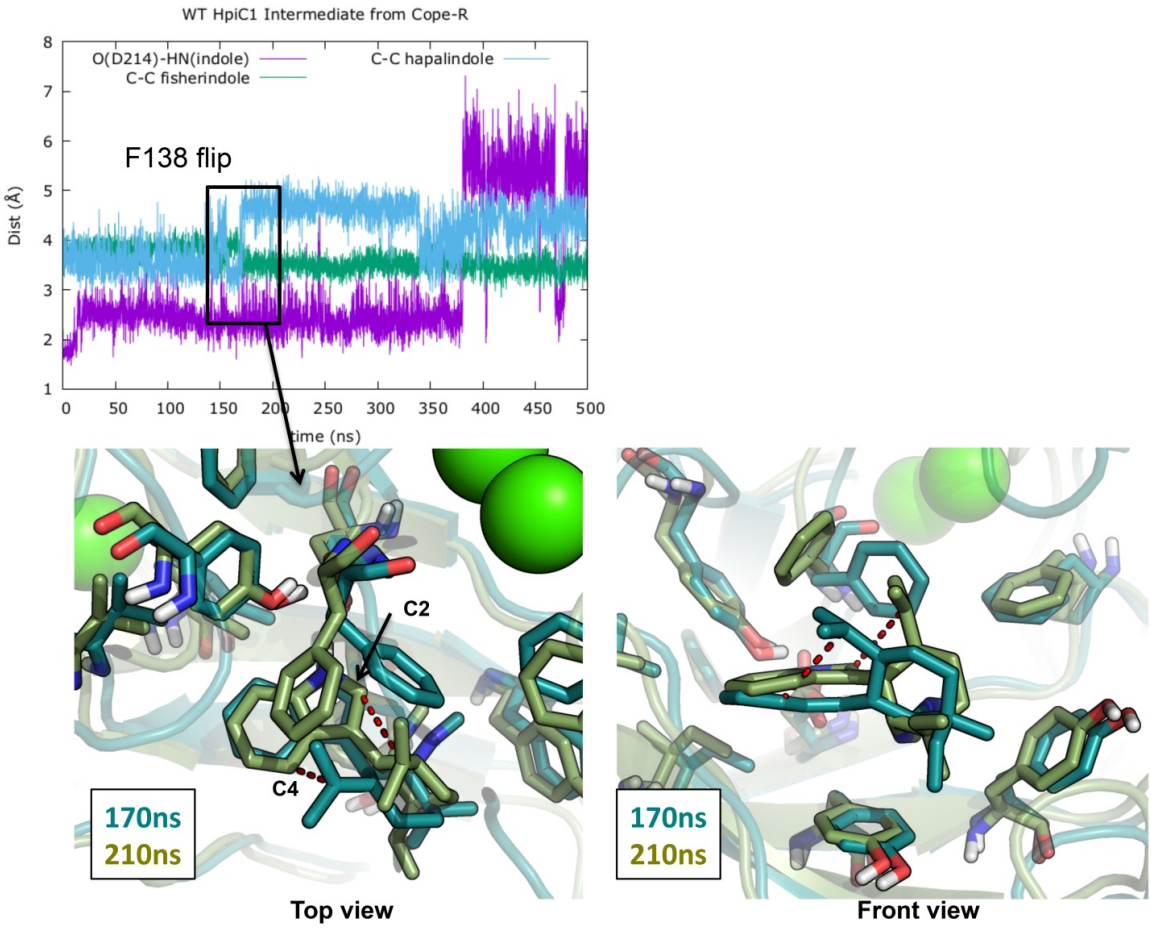
(e) and (**f**) show the dihedral angle defined by N(NC) – C11 – C3 – C2 explored during the 500 ns MD simulation for both substrate **1**(*R*)**P** (**e**, left) and **1**(*S*)**P** (**f**, right) bound complexes. This dihedral angle describes the relative orientation of the NC-containing substituent with respect to the indole ring. On the right of each plot, the stereochemistry at **C11** in the product of the Cope rearrangement that comes from this particular near attack conformation (NAC) is reported.



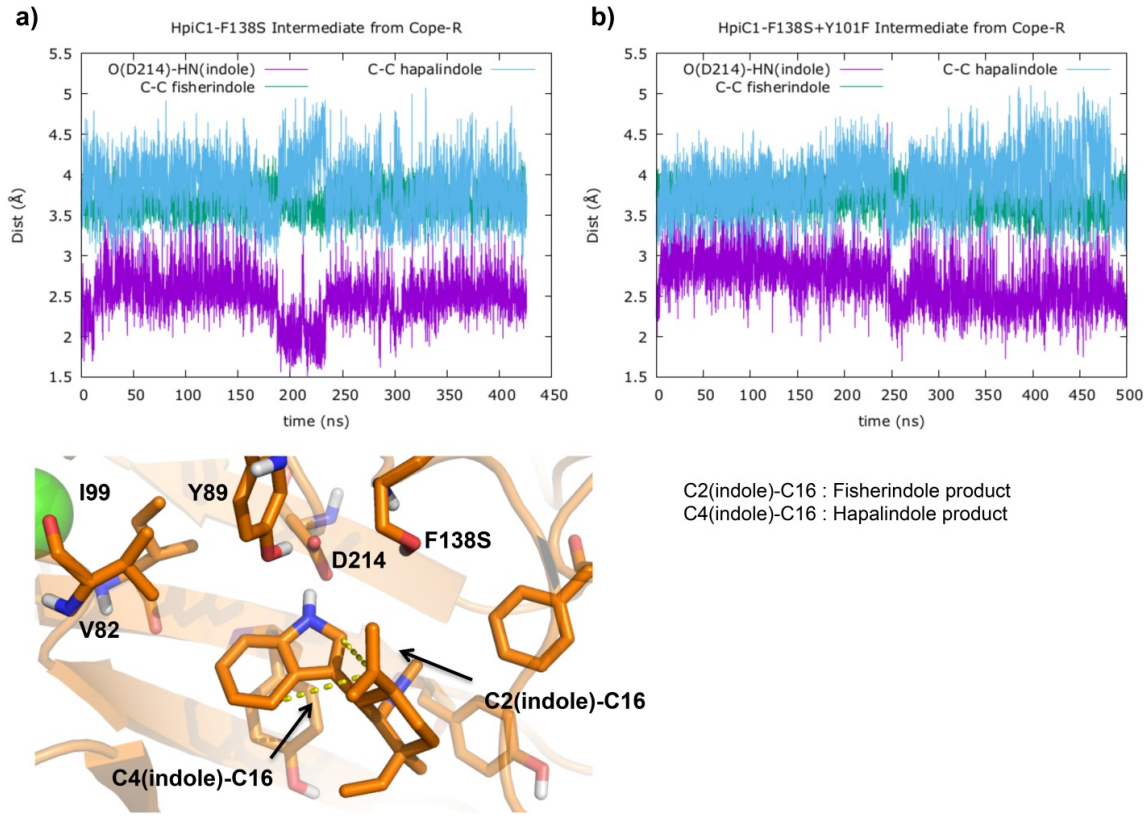
SI Figure 5-18. Representative snapshots of the active site arrangement observed during the 500 ns MD trajectories for substrate **1**(*R*)**P** (Cope-R, left) and **1**(*S*)**P** (Cope-S, right) bound into wild-type HpiC1 enzyme. MD simulations predict that substrate **1**(*R*)**P** is stabilized by the enzyme active site in a near attack conformation that leads to the C11-*R* and C12-*S* configurations (left), as described in **SI Figure 5-14**. This observation highlights the importance of the active site shape and arrangement in controlling the stereochemistry of the Cope rearrangement.



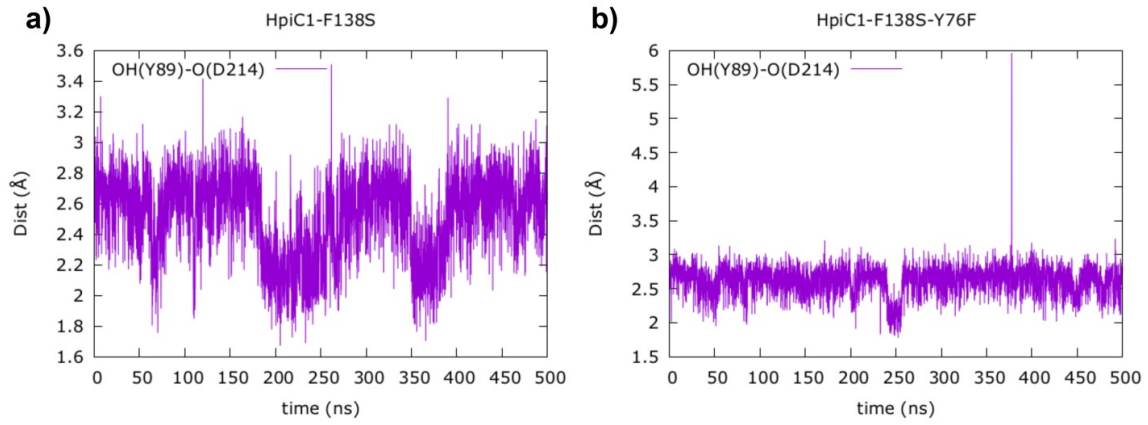
SI Figure 5-19. Key distances measured along the 500 ns MD trajectory for intermediate **4P** bound into wild-type HpiC1 enzyme. Two possible intermediates are considered: (**a**) **4P** coming from Cope-R pathway; and (**b**) **8P** coming from Cope-S pathway. **8P** rapid dissociates from the active site, as illustrated by the disruption of the O(D214)-HN(indole) H-bond.



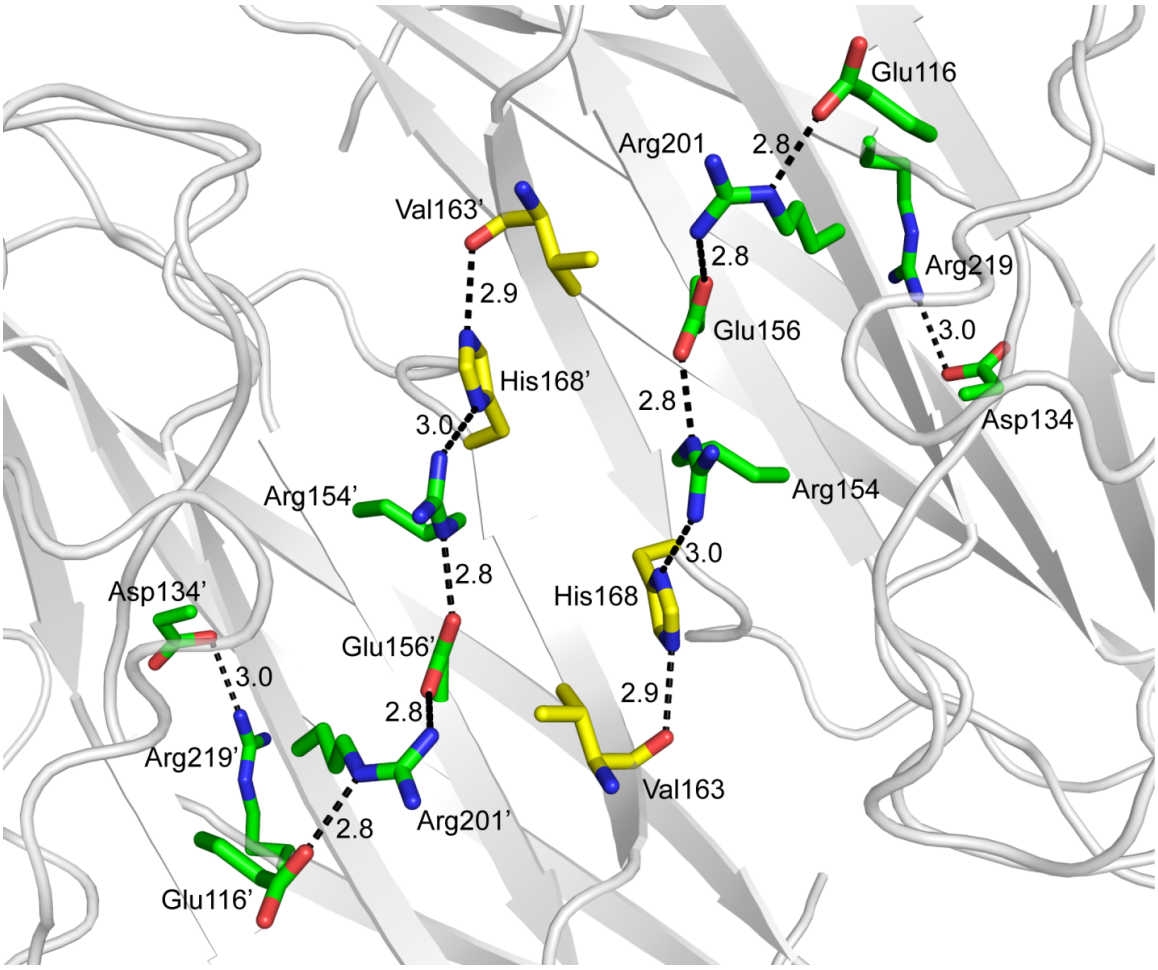
SI Figure 5-20. Representative snapshots of the active site arrangement observed during the 500 ns MD trajectories for intermediate **4P** (from Cope-R pathway) bound into wild-type HpiC1 enzyme. Preference for C-C hapalindole formation is observed (snapshot at 170ns) for the preferred orientation of F138 side chain, and C-C fischerindole formation could be possible due to a flip of the F138 side chain after ~180ns of MD trajectory. This observation highlights the importance of F138 residue in dictating the regiochemistry of the process.



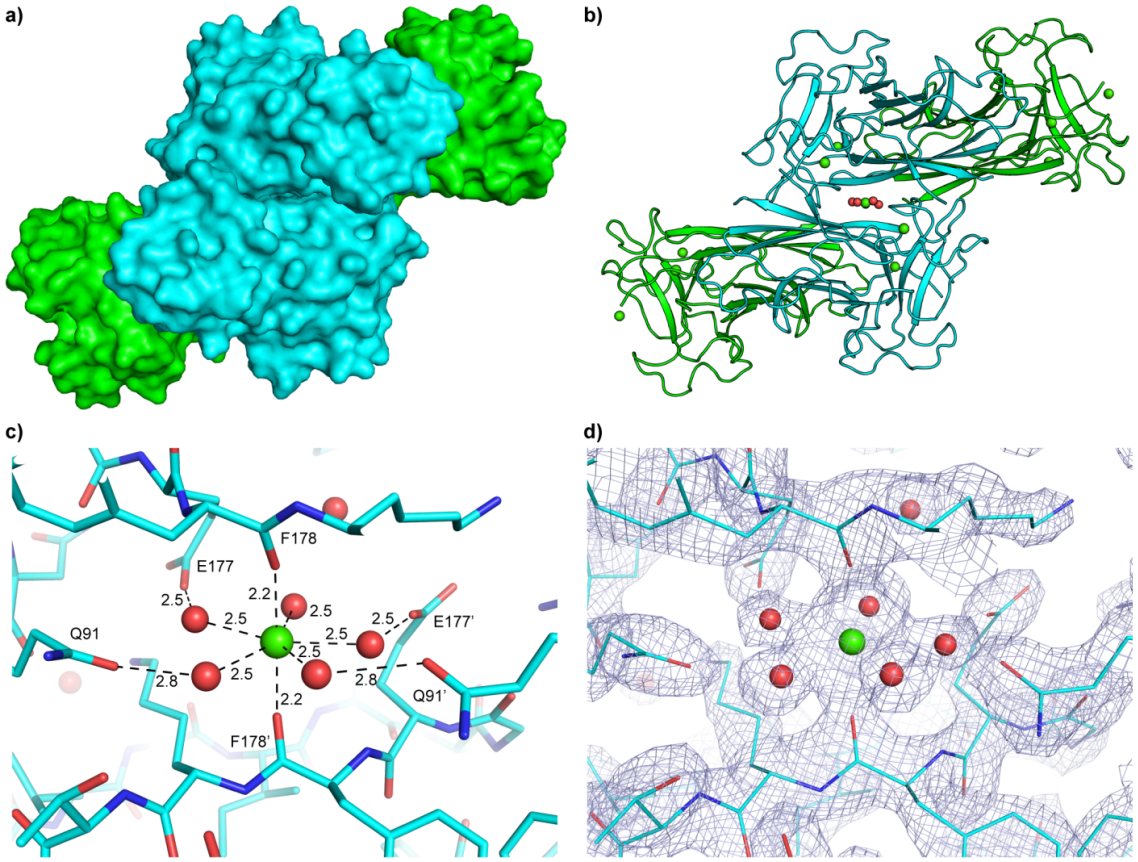
SI Figure 5-21. Key distances measured along the 500 ns MD trajectory for intermediate **4P** (from Cope-R pathway) bound into (**a**) F138S mutant and (**b**) Y101F/F138S double mutant. Preference for C-C fischerindole formation (C4-C16 distance ~3.5Å) is observed over hapalindole (C4-C16 distance ~4.0Å).



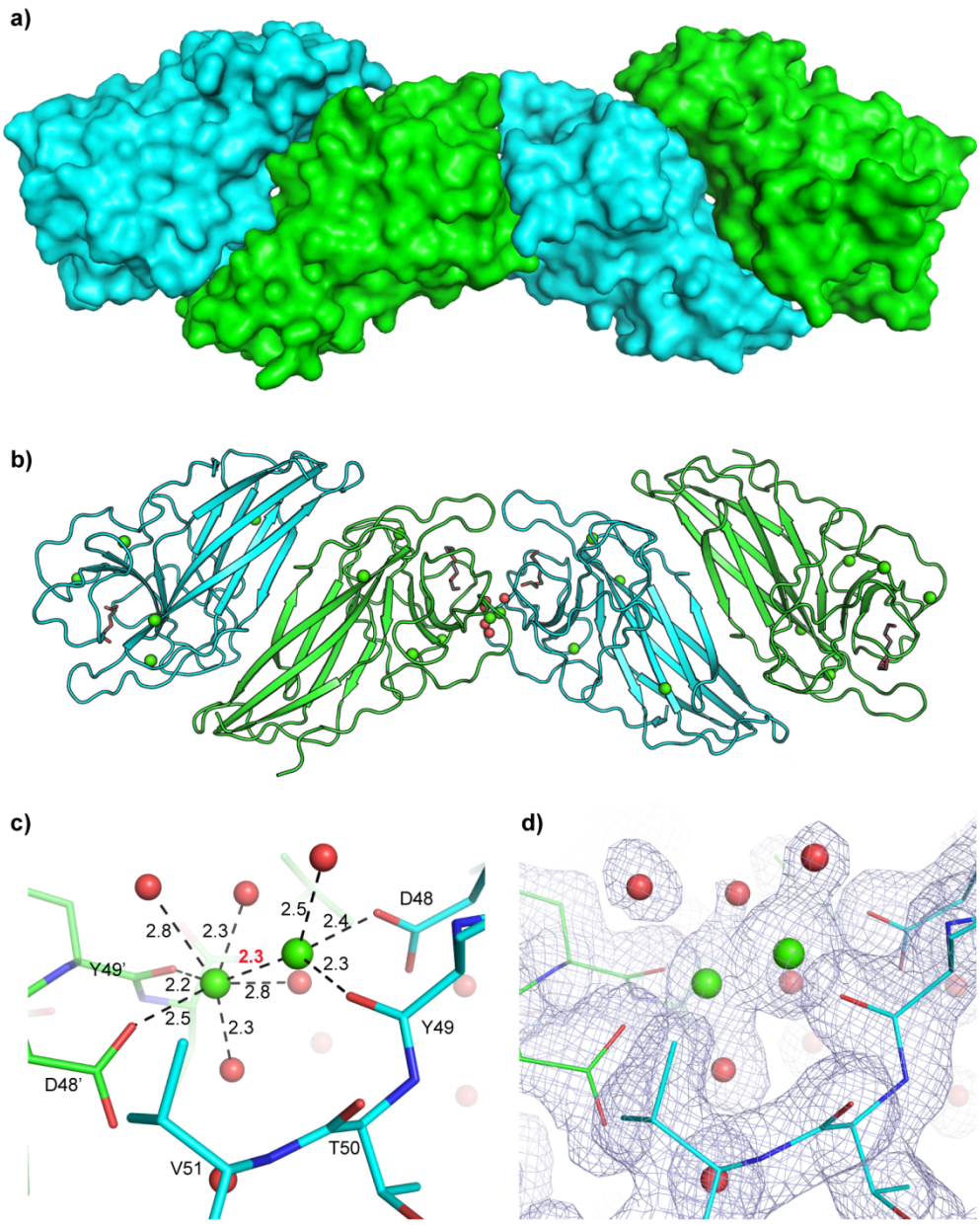
SI Figure 5-22. Distances measured along the 500 ns MD trajectory for the apo (**a**) F138S mutant and (**b**) Y101S/F138S double mutant between the center of mass of the two oxygen atoms of the D214 carboxylate group and the HO(Y89) side chain.



SI Figure 5-23. As with the CBMs[27], HpiC1 and other Stig cyclases are thermostable. This characteristic was highly beneficial for facile purification of an HpiC1 homolog, FamC1, from the *Fischerella ambigua* cell-free lysate[11]. As observed with other thermophilic proteins[59] an extended salt bridge array in HpiC1 is readily apparent, and spans the dimeric interface in coordination with hydrogen bonds. These residues are conserved in other Stig cyclases and we expect that many will share this thermostable property. Charged amino acids are shown as green sticks and hydrogen bonded amino acids are shown as yellow sticks.



SI Figure 5-24. Bridging calcium ion at crystal packing interface in HpiC1 crystal form 2. (**a**) Surface representation showing two dimers, each with green and cyan monomers. The total buried surface area for this assembly is 1700 Å^2 (PISA)[20]. (**b**) Cartoon representation: calcium ions are shown as green spheres; ligating waters are shown as red spheres. (**c**) Stick representation of the bridging calcium and ligating waters. (**d**) 2Fo-Fc electron density map at 2.1 Å resolution contoured at 1.5σ.



SI Figure 5-25. Second bridging calcium site. (**a**) Surface representation of HpiC1 in crystal form 1 showing adjacent dimers, each with green and cyan monomers. The total buried surface area for this assembly is 2160 Å^2 (PISA)[20] (**b**) cartoon representation of the HpiC1 packing interaction. PEG molecules bound in the active site are shown as sticks. All calcium ions are shown as green spheres, ligating waters at the bridging calcium ions are shown as red spheres. (**c**) Stick representation of the bridging calcium ions and ligating waters. Due to the close positioning between the two cations, each calcium ion was modeled at half occupancy. (**d**) 2Fo-Fc electron density map at 1.7 Å resolution contoured at 1.0σ.

5.7 References

1. Bhat, V.; Dave, A.; MacKay, J. A.; Rawal, V. H., The Chemistry of Hapalindoles, Fischerindoles, Ambiguines, and Welwitindolinones. *Alkaloids Chem. Biol.* **2014,** *73*, 65-160.
2. Asthana, R. K.; Srivastava, A.; Singh, A. P.; Deepali; Singh, S. P.; Nath, G.; Srivastava, R.; Srivastava, B. S., Identification of an antimicrobial entity from the cyanobacterium *Fischerella* sp isolated from bark of *Azadirachta indica* (Neem) tree. *J. Appl. Phycol.* **2006,** *18* (1), 33-39.
3. Becher, P. G.; Keller, S.; Jung, G.; Sussmuth, R. D.; Juttner, F., Insecticidal activity of 12-epi-hapalindole J isonitrile. *Phytochemistry* **2007,** *68* (19), 2493-2497.
4. Cagide, E.; Becher, P. G.; Louzao, M. C.; Espina, B.; Vieytes, M. R.; Juttner, F.; Botana, L. M., Hapalindoles from the cyanobacterium *Fischerella*: potential sodium channel modulators. *Chem. Res. Toxicol.* **2014,** *27* (10), 1696-1706.
5. Mo, S.; Krunic, A.; Chlipala, G.; Orjala, J., Antimicrobial ambiguine isonitriles from the cyanobacterium *Fischerella ambigua*. *J. Nat. Prod.* **2009,** *72* (5), 894-899.
6. Mo, S.; Krunic, A.; Santarsiero, B. D.; Franzblau, S. G.; Orjala, J., Hapalindole-related alkaloids from the cultured cyanobacterium *Fischerella ambigua*. *Phytochemistry* **2010,** *71* (17-18), 2116-2123.
7. Zhang, X. Q.; Smith, C. D., Microtubule effects of welwistatin, a cyanobacterial indolinone that circumvents multiple drug resistance. *Mol. Pharmacol.* **1996,** *49* (2), 288-294.
8. Hillwig, M. L.; Zhu, Q.; Liu, X., Biosynthesis of ambiguine indole alkaloids in cyanobacterium *Fischerella ambigua*. *ACS Chem. Biol.* **2014,** *9* (2), 372-377.
9. Raveh, A.; Carmeli, S., Antimicrobial ambiguines from the cyanobacterium *Fischerella* sp. collected in Israel. *J. Nat. Prod.* **2007,** *70* (2), 196-201.
10. Stratmann, K.; Moore, R. E.; Bonjouklian, R.; Deeter, J. B.; Patterson, G. M. L.; Shaffer, S.; Smith, C. D.; Smitka, T. A., Welwitindolinones, Unusual Alkaloids from the Blue-Green Algae *Hapalosiphon welwitschii* and *Westiella intricata*. Relationship to Fischerindoles and Hapalinodoles. *J. Am. Chem. Soc.* **1994,** *116* (22), 9935-9942.
11. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., Hapalindole/Ambiguine Biogenesis Is Mediated by a Cope Rearrangement, C-C Bond-Forming Cascade. *J. Am. Chem. Soc.* **2015,** *137* (49), 15366-15369.
12. Cope, A. C.; Hardy, E. M., The Introduction of Substituted Vinyl Groups. V. A Rearrangement Involving the Migration of an Allyl Group in a Three-Carbon System1. *J. Am. Chem. Soc.* **1940,** *62* (2), 441-444.
13. Ilardi, E. A.; Stivala, C. E.; Zakarian, A., [3,3]-Sigmatropic rearrangements: recent applications in the total synthesis of natural products. *Chem. Soc. Rev.* **2009,** *38* (11), 3133-3148.

14. DeClue, M. S.; Baldridge, K. K.; Kunzler, D. E.; Kast, P.; Hilvert, D., Isochorismate pyruvate lyase: a pericyclic reaction mechanism? *J. Am. Chem. Soc.* **2005,** *127* (43), 15002-15003.
15. Luk, L. Y.; Qian, Q.; Tanner, M. E., A cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase? *J. Am. Chem. Soc.* **2011,** *133* (32), 12342-12345.
16. Tanner, M. E., Mechanistic studies on the indole prenyltransferases. *Nat. Prod. Rep.* **2015,** *32* (1), 88-101.
17. Li, S.; Lowell, A. N.; Newmister, S. A.; Yu, F.; Williams, R. M.; Sherman, D. H., Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids. *Nat. Chem. Biol.* **2017,** *13* (5), 467-469.
18. Zhu, Q.; Liu, X., Discovery of a Calcium-Dependent Enzymatic Cascade for the Selective Assembly of Hapalindole-Type Alkaloids: On the Biosynthetic Origin of Hapalindole U. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (31), 9062-9066.
19. Zhu, Q.; Liu, X., Molecular and genetic basis for early stage structural diversifications in hapalindole-type alkaloid biogenesis. *Chem. Commun. (Camb.)* **2017,** *53* (19), 2826-2829.
20. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-797.
21. Holm, L.; Rosenstrom, P., Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **2010,** *38* (Web Server issue), W545-549.
22. von Schantz, L.; Hakansson, M.; Logan, D. T.; Walse, B.; Osterlin, J.; Nordberg-Karlsson, E.; Ohlin, M., Structural basis for carbohydrate-binding specificity--a comparative assessment of two engineered carbohydrate-binding modules. *Glycobiology* **2012,** *22* (7), 948-961.
23. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
24. Lutz, R. P., Catalysis of the Cope and Claisen Rearrangements. *Chem. Rev.* **1984,** *84* (3), 205-247.
25. Xue, Y.; Zhao, L.; Liu, H.-w.; Sherman, D. H., A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity. *Proceedings of the National Academy of Sciences* **1998,** *95* (21), 12111-12116.
26. Prof. Xinyu Liu, personal communication.
27. Abou-Hachem, M.; Karlsson, E. N.; Simpson, P. J.; Linse, S.; Sellers, P.; Williamson, M. P.; Jamieson, S. J.; Gilbert, H. J.; Bolam, D. N.; Holst, O., Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. *Biochemistry* **2002,** *41* (18), 5720-5729.
28. Montanier, C. Y.; Correia, M. A.; Flint, J. E.; Zhu, Y.; Basle, A.; McKee, L. S.; Prates, J. A.; Polizzi, S. J.; Coutinho, P. M.; Lewis, R. J.; Henrissat, B.; Fontes, C. M.; Gilbert, H. J., A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides

through calcium-mediated oligomerization. *J. Biol. Chem.* **2011,** *286* (25), 22499-22509.
29. Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J., Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **1993,** *229* (1), 105-124.
30. Kabsch, W., Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **2010,** *66* (Pt 2), 125-132.
31. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-221.
32. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **2010,** *66* (Pt 4), 486-501.
33. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **2011,** *67* (Pt 4), 355-367.
34. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.
35. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **2010,** *66* (Pt 1), 12-21.
36. Bonnett, Shilah A.; Rath, Christopher M.; Shareef, A.-R.; Joels, Joanna R.; Chemler, Joseph A.; Håkansson, K.; Reynolds, K.; Sherman, David H., Acyl-CoA Subunit Selectivity in the Pikromycin Polyketide Synthase PikAIV: Steady-State Kinetics and Active-Site Occupancy Analysis by FTICR-MS. *Chemistry & Biology* **2011,** *18* (9), 1075-1081.
37. Zhang, W.; Fortman, J. L.; Carlson, J. C.; Yan, J.; Liu, Y.; Bai, F.; Guan, W.; Jia, J.; Matainaho, T.; Sherman, D. H.; Li, S., Characterization of the Bafilomycin Biosynthetic Gene Cluster from Streptomyces lohii. *ChemBioChem* **2013,** *14* (3), 301-306.
38. Becke, A. D., Density ‐ functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98* (7), 5648-5652.
39. Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B: Condens. Matter* **1988,** *37* (2), 785-789.
40. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010,** *132* (15), 154104.

41. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32* (7), 1456-1465.
42. Miertuš, S.; Scrocco, E.; Tomasi, J., Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981,** *55* (1), 117-129.
43. Miertus, S.; Tomasi, J., Approximate Evaluations of the Electrostatic Free-Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* **1982,** *65* (2), 239-245.
44. Pascual-ahuir, J. L.; Silla, E.; Tuñon, I., GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J. Comput. Chem.* **1994,** *15* (10), 1127-1138.
45. Li, L.; Li, C.; Zhang, Z.; Alexov, E., On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013,** *9* (4), 2126-2136.
46. Schutz, C. N.; Warshel, A., What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **2001,** *44* (4), 400-417.
47. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9* (9), 3878-3888.
48. Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A.; California, U. o., AMBER 2016. **2016**.
49. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
50. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993,** *97* (40), 10269-10280.
51. Besler, B. H.; Merz, K. M.; Kollman, P. A., Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990,** *11* (4), 431-439.
52. Singh, U. C.; Kollman, P. A., An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984,** *5* (2), 129-145.
53. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
54. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and

Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
55. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
56. Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N., ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005,** *33* (Web Server issue), W299-302.
57. Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7* (2), 525-537.
58. Sondergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011,** *7* (7), 2284-2295.
59. Kumar, S.; Tsai, C. J.; Nussinov, R., Factors enhancing protein thermostability. *Protein Eng.* **2000,** *13* (3), 179-191.

Notes:

This work has been submitted for publication as “Structural Basis of the Cope Rearrangement and C−C Bond-Forming Cascade in Hapalindole/Fischerindole Biogenesis.” Sean A. Newmister, Shasha Li, Marc Garcia-Borràs, Jacob N. Sanders, Song Yang, Andrew N. Lowell, Fengan Yu, Janet L. Smith, Robert M. Williams, K. N. Houk & David H. Sherman. **2017**

Author contributions:

S.A.N. conducted protein crystallography and preparation. S.L. cloned and assayed the enzymes. M.G.-B. conducted MD simulations. J.N.S. conducted DFT calculations. A.N.L synthesized substrates. F.Y. performed bioinformatics analyses. S.A.N., S.L., M.G.-B., J.N.S, S.Y., J.L.S, K.N.H and D.H.S designed research and conducted data analysis and interpretation. S.A.N., S.L., M.G.-B., J.N.S, J.L.S, K.N.H and D.H.S wrote the manuscript with contributions from all authors.

Chapter 6
Summary and Future

6.1 Abstract

This thesis was mainly focused on the Stig cyclases in forming hapalindoles and fischerindoles which has been detailly discussed in Chapter 2-5. In this chapter, I will briefly discuss the future directions of the Stig cyclases research, and present the preliminary data toward the oxygenases which proposed to control the late-stage tailoring reactions.

6.2 Summary of Chapter 2-5

Heretofore, in total twelve hapalindole-related gene clusters have been disclosed from our lab and other research labs,[1-5] nine of them (**Figure 6-1**, a-i) have been reported to produce hapalindole alkaloids, while the other three (j-l) were identified from publicly available genomes.[5] Besides, the very first strain for hapalindole isolation[6], *Hapalosiphon fontinalis* Bornet ATCC 39694 [V-3-1], was also sequenced in our lab and the gene cluster annotation is on-going. This new gene cluster will provide more biological material in elucidating the molecular basis of the Stig cyclases and oxygenases.

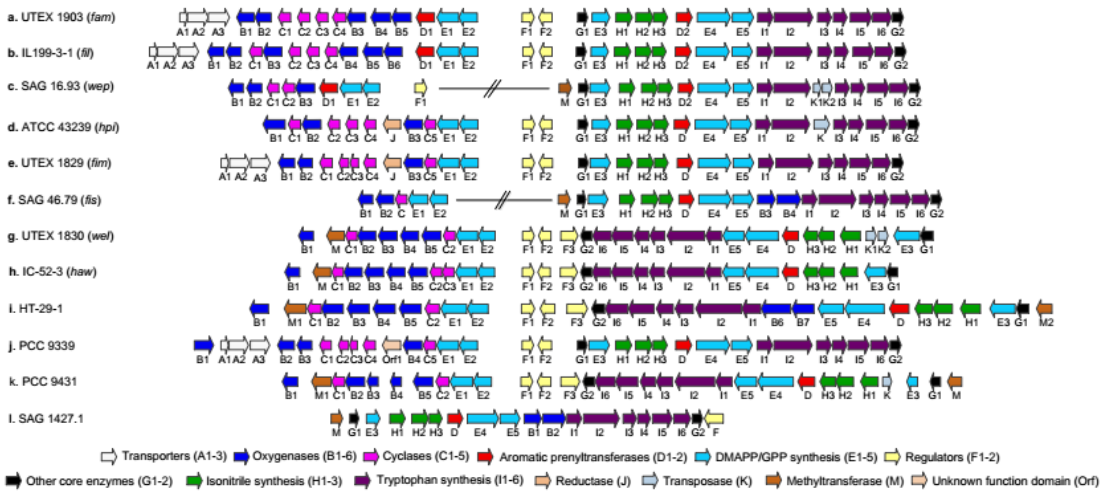
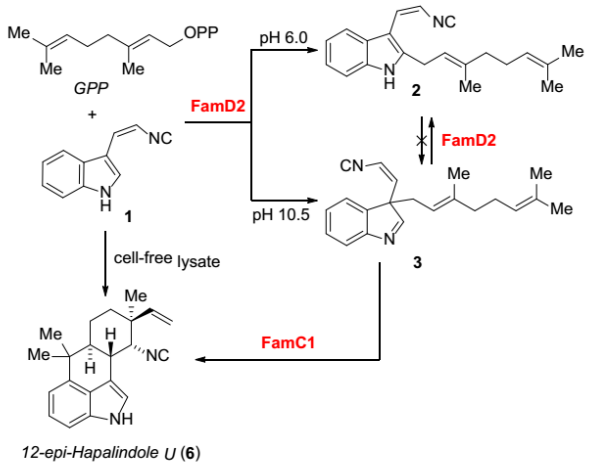


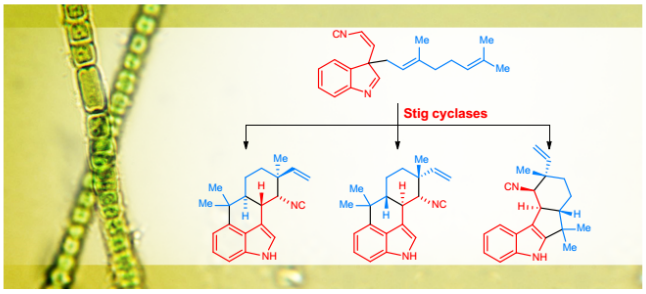
Figure 6-1. A summary of reported hapalindole-related gene clusters. a) *fam* gene cluster from ambiguine-producing strain *Fischerella ambigua* UTEX 1903; b) *fil* gene cluster from ambiguine-producing strain *Fischerella* sp. IL-199-3-1; c) *wep* gene cluster from ambiguine-producing strain *Westiellopsis prolifica* SAG 16.93; d) *hpi* gene cluster from hapalindole-producing strain *Fischerella* sp. ATCC 43239; e) *fim* gene cluster from fischerindole-producing strain *Fischerella muscicola* UTEX 1829; f) *fis* gene cluster from fischerindole-producing strain *Fischerella* sp. SAG 46.79; g) *wel* gene cluster from welwitindolinone-producing strain *Hapalosiphon welwitschii* UTEX 1830; h) *haw* gene cluster from welwitindolinone-producing strain *Hapalosiphon welwitschii* W. & G.S. West UH IC-52-3; i) welwitindolinone-producing strain Westiella intricate UH strain HT-29-1; j) *Fischerella* sp. PCC 9339 (JGI IMG/ER: 2516653082); k) *Fischerella* sp. PCC 9431 (JGI IMG/ER: 2512875027); l) *Fischerella muscicola* SAG 1427-1 (JGI IMG/ER: 2548876995).

As summarized in **Figure 6-2**, my research was focused on elucidating the molecular basis of hapalindole/fischerindole formation, starting from the initial mining of Stig cyclase, to the enzymatic activity identification of each Stig cyclase in assembling the individual hapalindole and fischerindole cores, and finally the mechanism in controlling the complex ring-formation. More work is required to fully understand these intriguing enzymes, which will be described in **Section 6.4**.

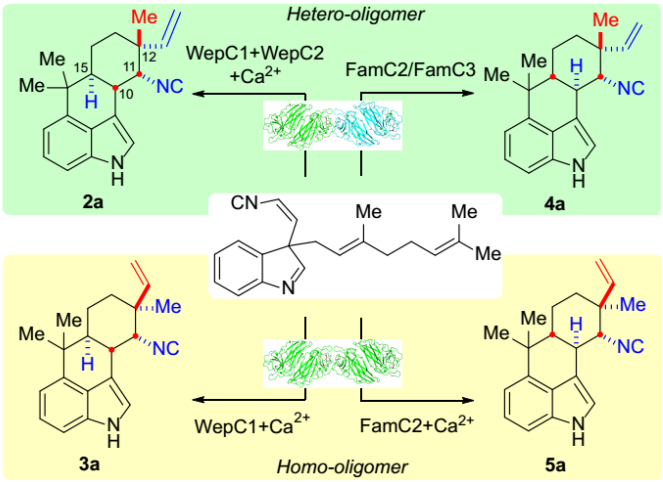
Chapter 2:
1. PT: FamD1/FamD2
2. Common intermediate **3**
3. Stig cyclase FamC1



Chapter 3:
1. Regioselectivity between hapalindoles and fischerindoles
2. Stereoselectivity within hapalindoles
3. Heterodimer (FamC2-FamC3)



Chapter 4:
1. Ca^{2+} dependence
2. Stereochemical reconfiguration through protein partnering
3. Oligomerization
4. Core-ring alteration through mutagenesis



Chapter 5:
1. Protein structure as dimer
2. MD/QM study
3. Critical residues in controlling the core-ring formation

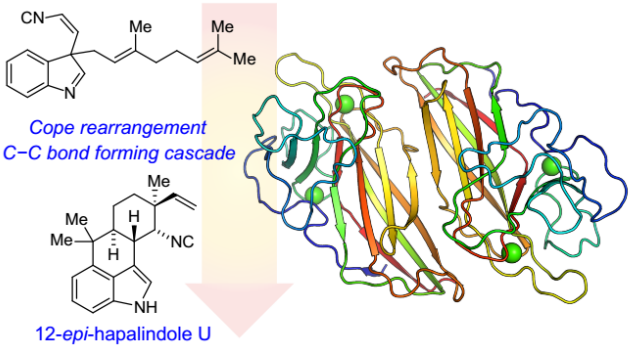


Figure 6-2. A summary of research data described in Chapter 2-5, the most important information was listed on the left.

6.3 Aromatic prenyltransferases (PTs)

As the only two PTs described in this thesis, FamD1 is forming ambiguines through C-2 reverse prenylation on hapalindoles and FamD2 is catalyzing the common intermediate formation through normal prenylation (C-3 geranylated indole isonitrile, **Figure 6-2**, compound **3**). We observed that FamD1 and homologs only exist in ambiguine-producing strain (FamD1-FilD1, 100% identical; FamD1-WepD1, 95% identical), while FamD2 and homologs are presented in every hapalindole-related gene cluster sharing very high sequential identities (≥96%) (**Figure 6-3**). According to our study with FamD2/FilD2/WepD2/FimD/FisD/WelD, these FamD2 homologs are exchangeable in the enzymatic assay to produce the same compounds. In this section, I will describe the additional study on protein structure and substrate scope, mainly focused on FamD1.

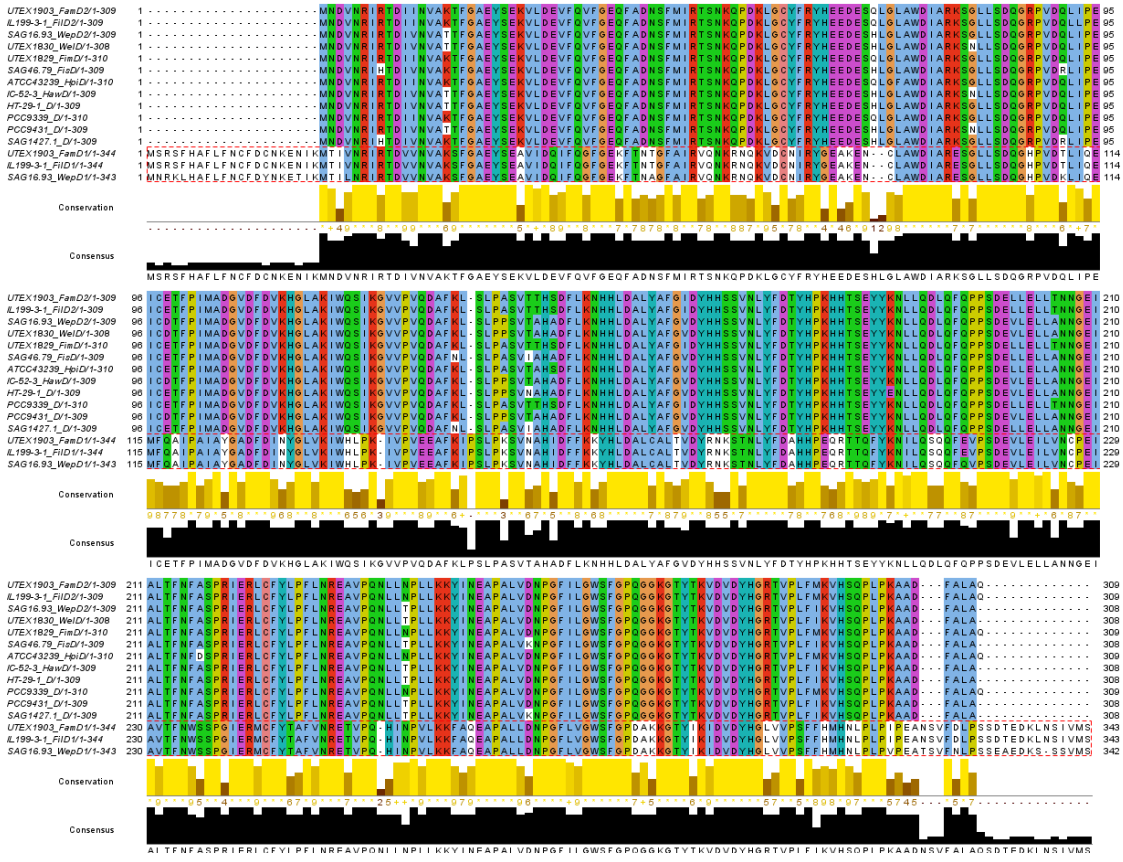


Figure 6-3. Sequence alignment of FamD2, FamD1 and their homologs. The bottom three sequences in red-dash box are FamD1 and homologs.

6.3.1 Protein structure of FamD1

With a 59% sequence identity, FamD2 catalyzes indole C-3 normal prenylation, while FamD1 conducts C-2 reverse prenylation (Section **2.3.1**). Moreover, except 12-*epi*-ambiguine B, all identified natural ambiguines are derived from hapalindole U, suggesting a stereochemical preference of FamD1 in choosing substrates. In order to understand the molecular basis on these regio- and stereoselective prenylation, we conducted protein structure study through X-ray crystallography with FamD1, FamD2 and homolog FisD (SAG 46.79). The study of FamD2 and FisD was hampered by the failure to get good crystals, but we were able to crystallize FamD1. The structure of FamD1 was solved with 1.9 Å diffraction to be a TIM barrel fold (**Figure 6-4**), a common protein scaffold for aromatic prenyltransferases. However, a pyrophosphate unit (PPi) was bound tightly in the predicted active site, and cannot be replaced by adding substrates indole isonitrile or hapalindoles. To remove this insert, we tried different crystallization condition and more robust protein purification, but none of them worked. With the speculation that this PPi may come from *E. coli* during protein expression, we tried to detach it by refolding the protein. The fresh inclusion body of expressed FamD1 was treated with a high concentration of denaturant, such as urea and guanidinium chloride to obtain unfolded protein. However, we failed to recover soluble FamD1 after trying several refolding techniques, such as dialysis, dilution and on-column refolding.[7-9]

Further exploration to obtain the real apo- and holo-structure of FamD1 would be valuable to understand the substrate preference on hapalindole U. Together with FamD2 structure information, we will be able to elucidate the selectivity between normal and reverse prenylation, and help to verify the observation that FamD2 can promote 1-2 shift to get the shunt C-2 prenylated metabolite.

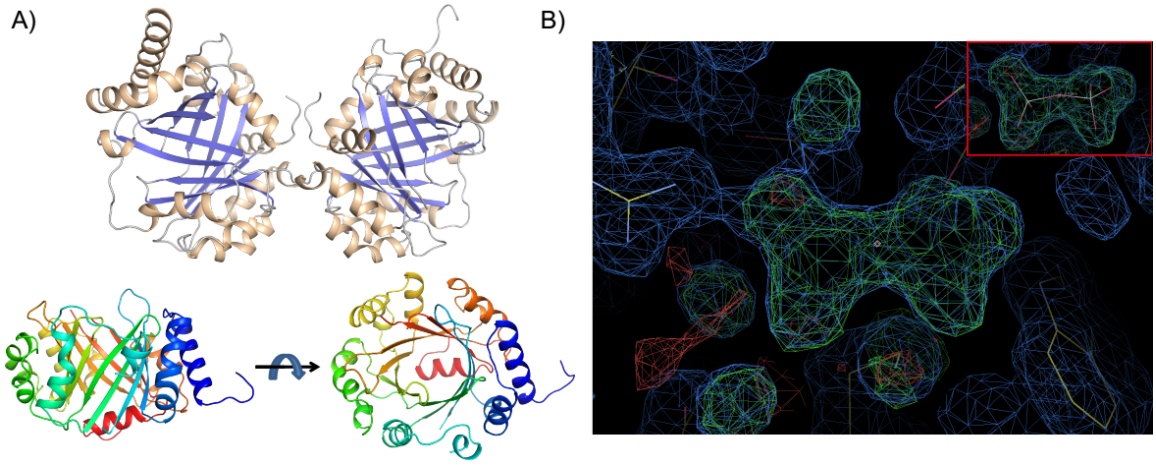


Figure 6-4. A) Protein structure of FamD1. B) The predicted active site of FamD1, in where a pyrophosphate unit was present.

6.3.2 Substrate scope of FamD1/FamD2

One of the intriguing structural features of hapalindole-type alkaloids is the C-10/11/12/15 stereogenic centers, which contribute to the diversity across this family by forming at least six stereochemical patterns. UTEX 1903 has produced at least two types of stereospecific hapalindoles (hapalindole U and H), but its ambiguines are specifically developed from hapalindole U. Besides, 20 ambiguines have been reported so far, and 19 of them are derived from hapalindole U. This strong stereochemical preference in ambiguine formation may suggest an ecological advantage of this specific stereochemistry among the diverse metabolite pool, and FamD1 may has a very restricted substrate flexibility. To examine the flexibility of FamD1, we fed it with several stereo-varied hapalindoles, including hapalindole U, H, 12-*epi*-hapalindole U, and 11-*epi*-hapalindole C, D and hapalindole Q. FamD1 was able to accept most of the substrates, except hapalindole Q and 11-*epi*-hapalindole D, which have a NCS group instead of NC (**Figure 6-5**a). We scaled up the reaction with 11-*epi*-hapalindole C for product characterization. However, instead of the expected indole C-2 prenylated product, two compounds were formed, including a “reverse” N-1-prenylated as the minor product and a “normal” indole C-7-prenylated as the major one (**Figure 6-5**b). We proposed that a reverse prenylation is first occurred at N-1 position, the resulted molecule then undergoes

a Cope rearrangement to convert into C-7 normal prenylated major product. This unexpected prenylation raises a question about how FamD1 controls regioselectivity of prenylation among all the indole substrates. Further studies including scale-up for structural characterization on other substrates and protein crystallization are necessary to answer this.

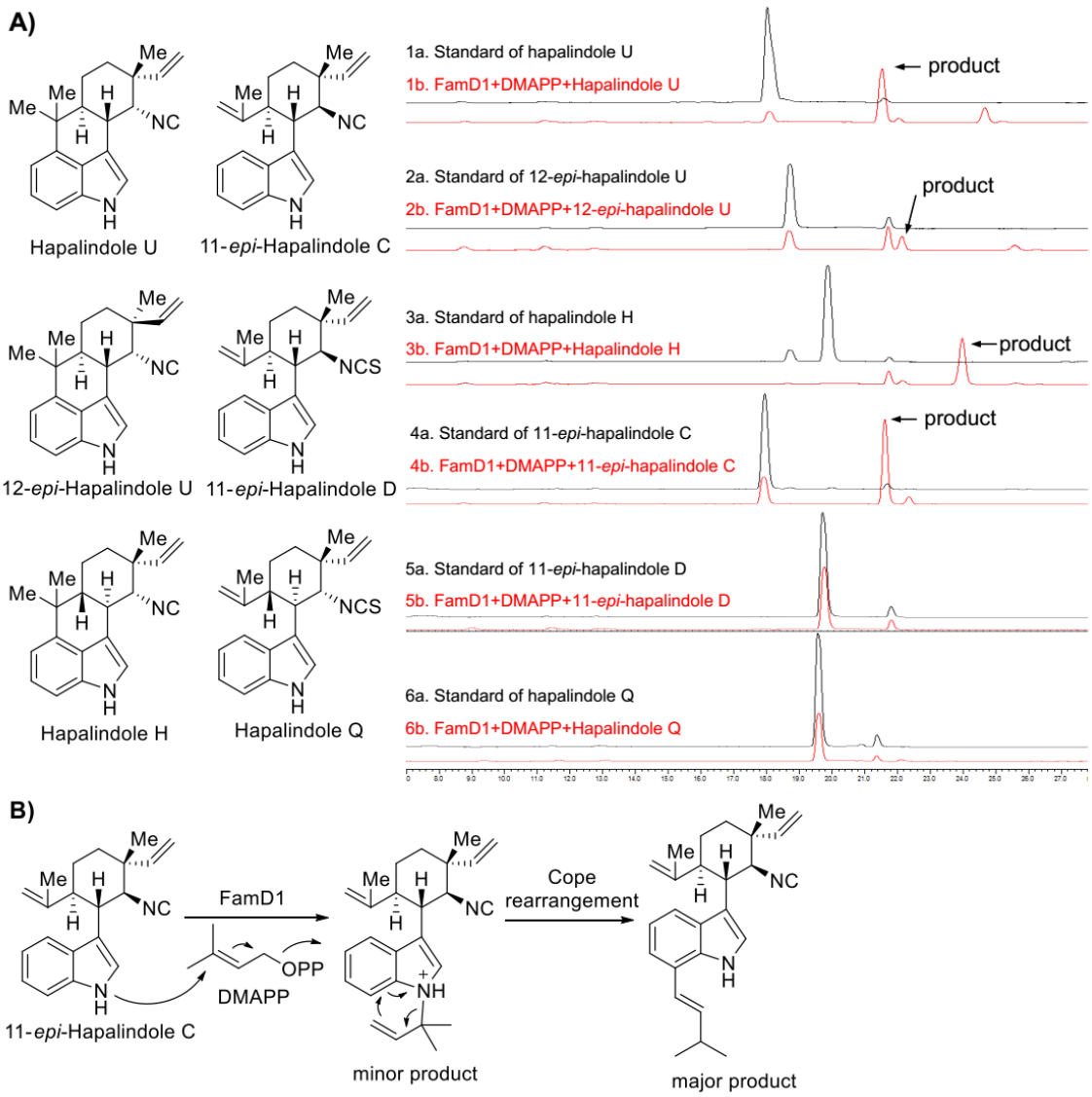


Figure 6-5. Substrate flexibility assay of FamD1. A) HPLC traces of *in vitro* reactions; B) Products from 11-*epi*-hapalindole C and proposed Cope rearrangement.

One approach to expand the chemical diversity of hapalindole-type alkaloids is to employ synthetic *cis*-indole isonitrile analogs and heterocyclic

cores in the enzymatic reaction with FamD2 and Stig cyclases. To examine the substrate flexibility of FamD2, we have proposed some indole analogs for chemoenzymatic assay. As shown in **Figure 6-6**, all compounds will be treated with diethyl cyanomethylphosphonate to build the *cis*-2-isocyanovinyl group, which will be tested in vitro with FamD2. This work will provide substrates for the downstream enzymatic conversions with Stig cyclases and late-stage tailoring enzymes, and finally generate diverse hapalindole/fischerindole derivatives.

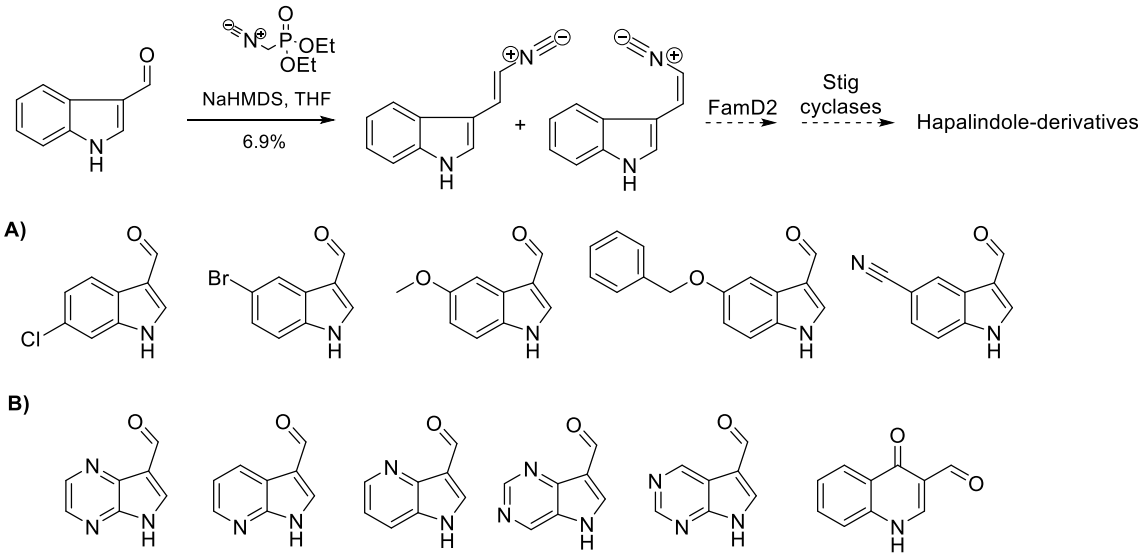


Figure 6-6. Preparation of FamD2 unnatural substrates. A) Indoles with various substitutions which have been converted to the *cis*-2-isocyanovinyl indole analogs. B) Proposed heterocyclic analogs for substrate exploration.

6.4 Stig cyclases

6.4.1 Enzymatic activity

Our study toward the Stig cyclases is mainly focused on gene clusters a-h, viz cyclases FamC1-C4, FilC1-C4, WepC1-C2, HpiC1-C5, FimC1-C5, FisC, WelC1-C2, and HawC1-C3 (**Figure 6-1**). All Stig cyclases have been heterologously expressed in *E. coli* and analyzed for their in vitro enzymatic activities, except HawC2 and HawC3. Among these 24 recombinant proteins, the function of HpiC4, FimC3 and WelC1 are still unknown after testing various in vitro conditions, for example, changing pH/temperature, adding metal-ion/cell-

free lysate, partnering other Stig cyclases. We believe that these proteins carry unique functions but need more study to find the correct in vitro condition. One circumstance supporting this assumption is that both *hpi* and *fim* have hapalindole/fischerindole with unidentified bio-transformer, such as hapalindole J and fischerindole L from *fim* gene cluster, 12-*epi*-hapalindole J from *hpi* gene cluster (**Figure 6-7**). This indicates that the uncharacterized Stig cyclases should play some roles in forming these hapalindole/fischerindole, which requires a more thorough research.

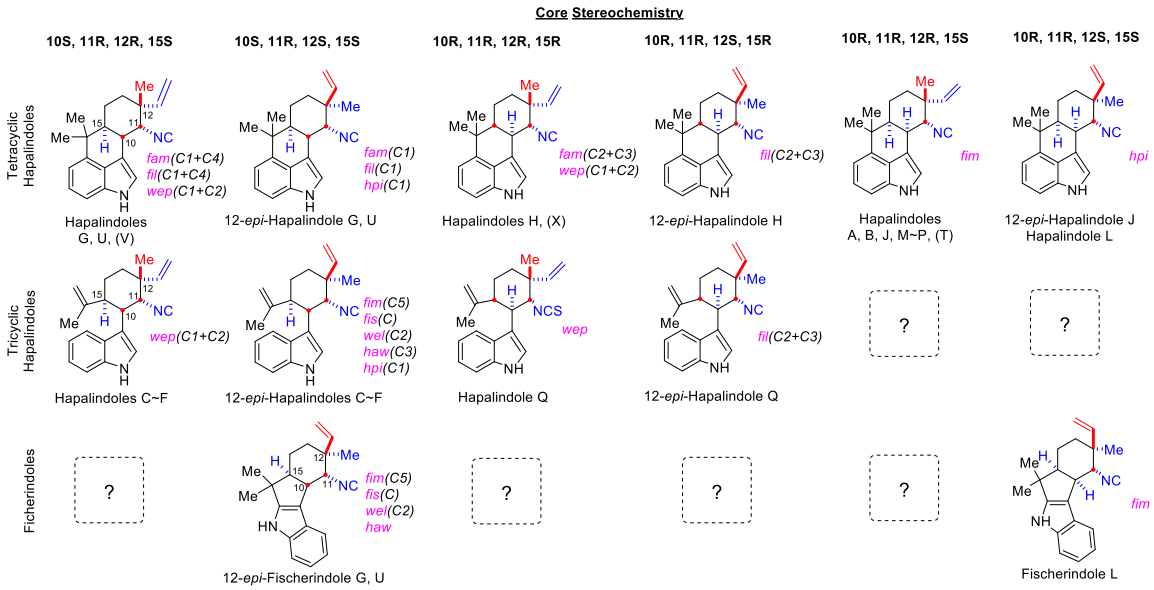


Figure 6-7. Stereochemical classification of isolated hapalindoles and fischerindoles and the gene cluster information where the alkaloids were isolated from. Genes in parentheses were characterized to generate the hapalindole scaffold on its left-side. The structures with only gene cluster name instead of specific gene name are metabolites with known isolation resource, but unknown biosynthetic model.

6.4.2 Oligomer and heteromeric configuration

Besides of the biosynthetic function, another important topic is the oligomerization proposed in Chapter 4 and 5. In order to interpret the mechanism of Stig cyclases in reconfiguring the product stereochemistry through Ca^{2+} promoted pairing, we proposed that the Stig cyclases are functioning by forming oligomer to build a conjugated active site to cooperatively control the stereochemistry of the hapalindole metabolites. Right now, we only observed

oligomer packing form from HpiC1 crystal structure, but crystal structures of other heteromeric complexes (such as FamC1-FamC4, and FamC2-FamC3) are critical to verify this hypothesis. Our previous effort to crystallize FamC2-FamC3 was impeded by the instability of FamC2 to maintain equivalent ratio with FamC3 in high concentration, but after characterizing the function of FamC1-FamC4 and their homologs as heteromeric complexes, this work will be much easier since these proteins have good solubility and stability. Besides of FamC1-FamC4 pair, there are other options for crystallization, viz FilC1-FilC4, FimC1-FimC4, WepC1-WepC2.

6.5 Oxygenases

6.5.1 Bioinformatics

In addition to the core-ring formation conducted by the prenyltransferases and Stig cyclases, these alkaloids are modified by multiple late-stage tailoring reactions, including hydroxylation, epoxidation, sulfuration, and chlorination. Thereinto, the chlorine functional group is one of the most common substituent with regio- and stereospecificity (C13*R*), which was previously proposed as the catalytic co-factor in C/D-ring formation. Recently, Liu *et al* have characterized WelB1 and FamB1 (WelO5 and AmbO5 in their papers) as nonheme 2-oxoglutarate-iron-dependent halogenases to conduct this chlorination.[10-13] Since every hapalindole-producing strain contains chlorinated-metabolites, we are able to identify a halogenase from each gene cluster (**Figure 6-8**, G1), except *hpi*, whose HpiB1 was annotated as a FAD dependent oxidoreductase and HpiB2-B3 as Rieske oxygenases.

Besides of the halogenases (G1), all other annotated oxygenases were categorized into oxygenase-like cyclases (G2) and Rieske oxygenases (G3) based on protein sequence. In addition, there are other uncharacterized enzymes, such as FimJ/HpiJ annotated as dihydrodipicolinate reductase, and FisB4 as P450 (**Figure 6-8**). It is hypothesized that these both uncharacterized enzymes will control most of the tailoring reactions, the E-ring formation in ambiguines, and C-ring modification in welwitindolinones.

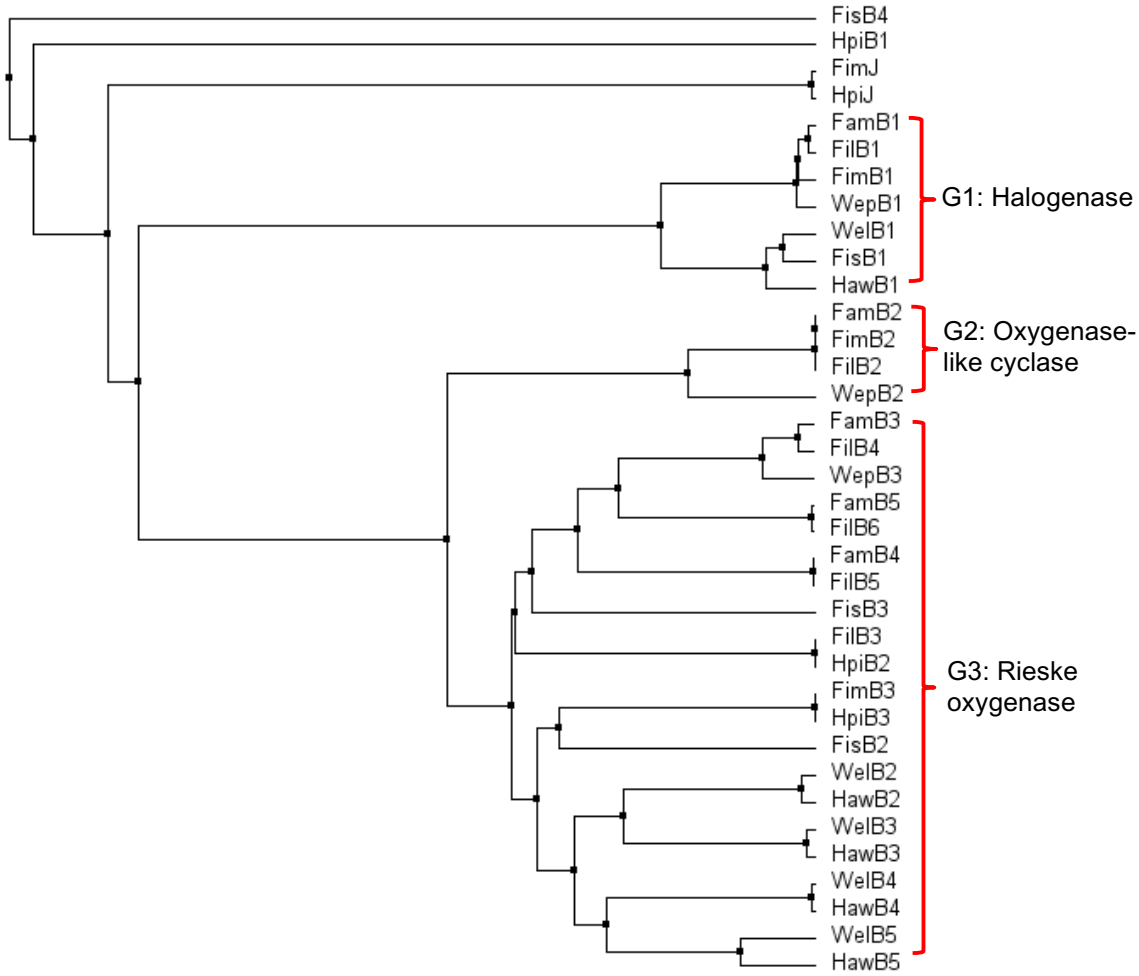


Figure 6-8. Phylogenetic analysis of annotated oxygenases from gene clusters *fam, fil, wep, fim, fis, hpi, wel, haw*.

6.5.2 Ambiguine E-ring formation

The ambiguines are the only chemotype within this family which possess a pentacyclic skeleton, involving a very interesting 7-*endo*-trig or 6-exo-trig cyclization of the C-11 isonitrile onto the unactivated alkenyl residue of the C-2-dimethylallyl group (**Figure 6-9**). According to our bioinformatics study, we found FamB2 and its homologs WepB2/FimB2/FilB2 are sequentially different from other oxygenases. More importantly, this group of proteins only presents in ambiguine-producing gene clusters, such as UTEX 1903, IL 199-3-1 and SAG 16.93. Thus, we proposed that FamB2 homologs (**Figure 6-8**, G2) are the enzymes to cyclize the E-ring through C-C bond formation. These proteins were

initially annotated as PrnD-like[14] aminopyrrolnitrin Rieske oxygenases, however, in-depth bioinformatics study revealed that they do not contain the conserved Rieske domain (**Figure 6-10**, **SI Figure 6-1**). Thus, instead of treating them as Rieske oxygenases, we renamed this group as oxygenase-like cyclase.

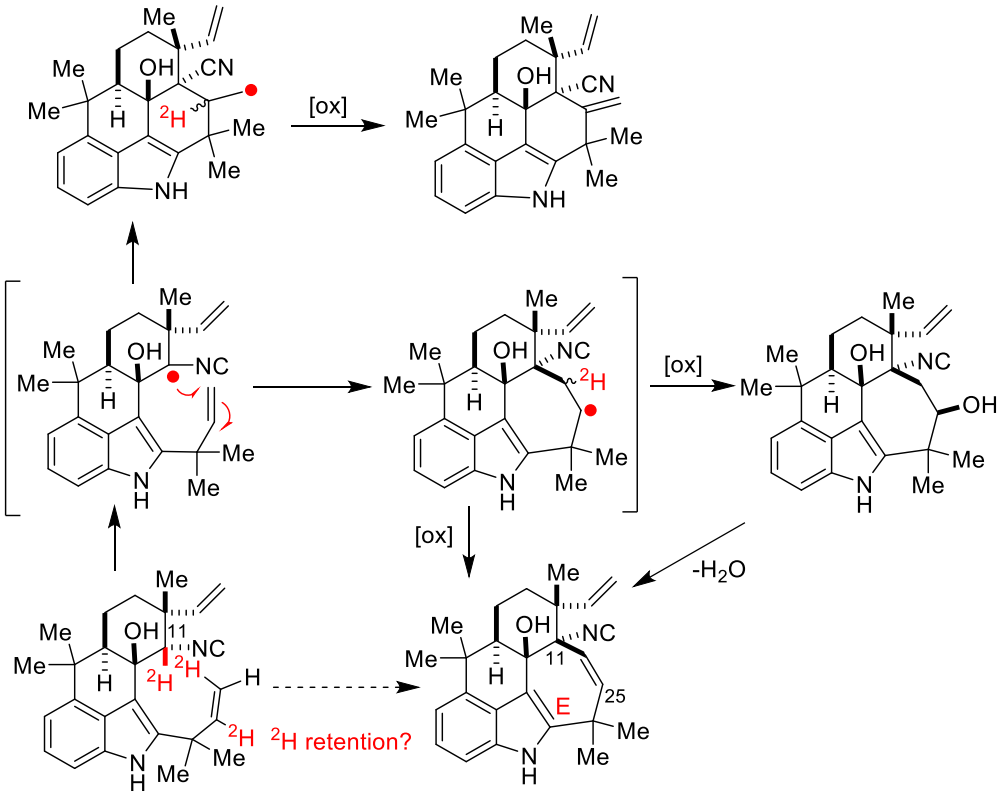


Figure 6-9. Proposed mechanism for ambiguine E-ring formation through oxidation.

We have already prepared FamB2/WepB2 as recombinant protein, and tested their in vitro enzymatic activities by incubating with substrates tetracyclic ambiguines. The tetracyclic ambiguines were chemoenzymatically synthesized by mixing prenyltransferase (FamD2) and Stig cyclase (FamC1) with initial substrate indole isonitrile and GPP, then the generated hapalindole core will be further prenylated with a dimethylallyl function group by prenyltransferase FamD1 to get the final tetracyclic ambiguines. These three steps can also be combined as one-pot reaction in high yield. In our initial exploration of in vitro conditions, we treated these proteins as co-factor independent enzymes by only mixing the

protein with substrate ambiguine and buffer, but no new compound was generated yet. Further study is on-going to understand the enzymatic function.

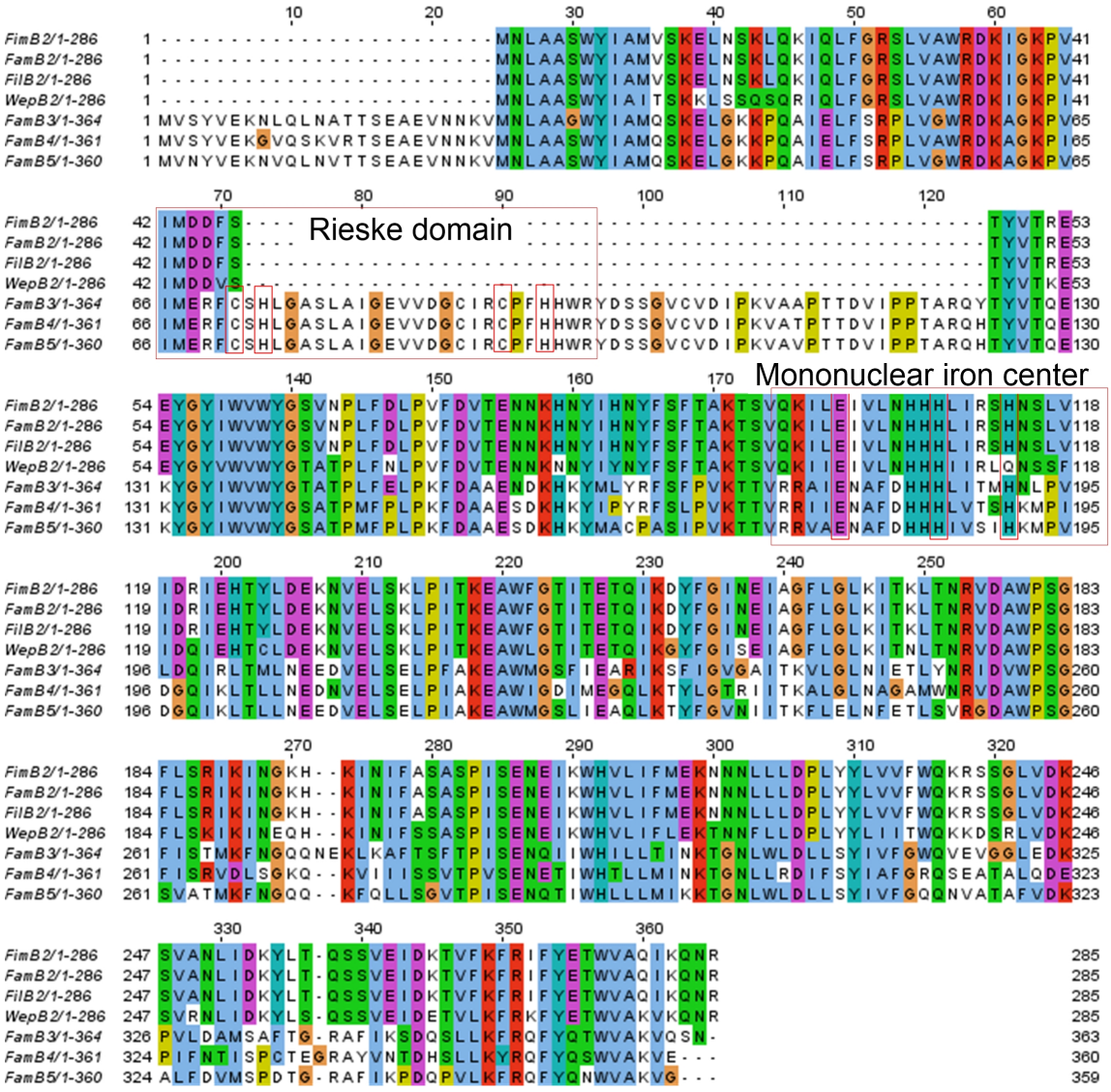


Figure 6-10. Protein sequence alignment of FamB2/FimB2/FilB2/WepB2. Compared with the real Rieske oxygenases FamB3/FamB4/FamB5, these proteins do not contain Rieske domain, and WepB2 does not even contain the mononuclear iron center. FimB2 is 100% identical to FamB2, even though it comes from a fischerindole-producing strain (UTEX 1829), but many genes from cluster *fim* and *fam* are identical.

6.5.3 Late-stage tailoring transformations

The late-stage tailoring includes two major parts, the first is forming new polycyclic patterns such as welwitindolinones, and the second is generating

regio- and stereoselective functional groups. According to our hypothesis, welwitindolinones are derived from fischerindoles through oxidation and rearrangement to form the four-membered C-ring, followed by another oxidation and ring-expansion to get the bridged-ring system (**Figure 6-11**A). Thus, the *wel* gene cluster from welwitindolinone-producing strain *Hapalosiphon welwitschii* UTEX 1830 will be the main system to study. Through comprehensive bioinformatics comparison, the candidates are focused on the four Rieske oxygenases (WelB2-5). The genes will be cloned, overexpressed in *E. coli*. The activity of the proteins will be examined in vitro by loading substrates fischerindoles, which will be provided through biosynthesis. Besides of welwitindolinone core pattern, another highly interested polycycle is 12-epi-fischerindole W only isolated from *Fischerella sp.* SAG 46.79, which possessing a benzoic C-ring (**Figure 6-11**B). We propose that this molecule is derived from 12-*epi*-fischerindole I through ring-expansion and aromatization. Annotation study showed that this *fis* gene cluster has a P450 (FisB4), which does not present in other gene clusters. Thus, this protein is the most potential candidate to work with.

The second portion of late-stage tailoring is the rich stereochemical and functional group arrays including Cl, OH, -O-, NCS, NC displayed in rigid polycyclic cores. Some reactions involve C-H functionalization on unactivated-alkyl carbon, such as C-13 halogenation, and C-15 hydrogenation. These functional groups are not only chemically important, but more attractively show the ability for structural modification to optimize the pharmaceutical potency in drug development.

So far, we have built the expression constructs of FamB3/FamB4/FamB5/FisB4/WepB3, but unfortunately, we did not yet find a good expression system for soluble protein. This will be one of the major direction in the future study.

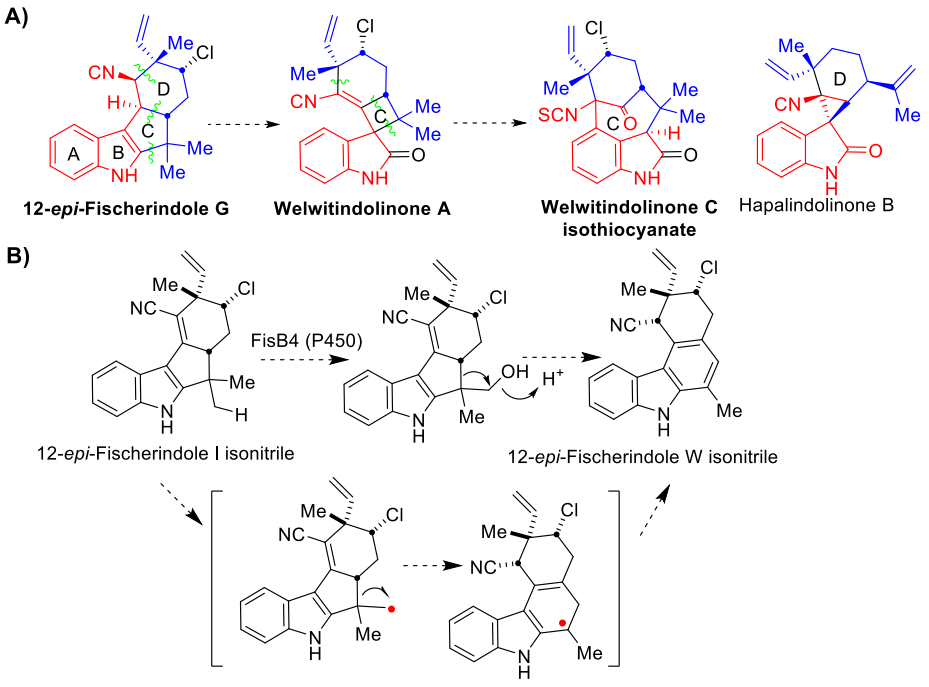
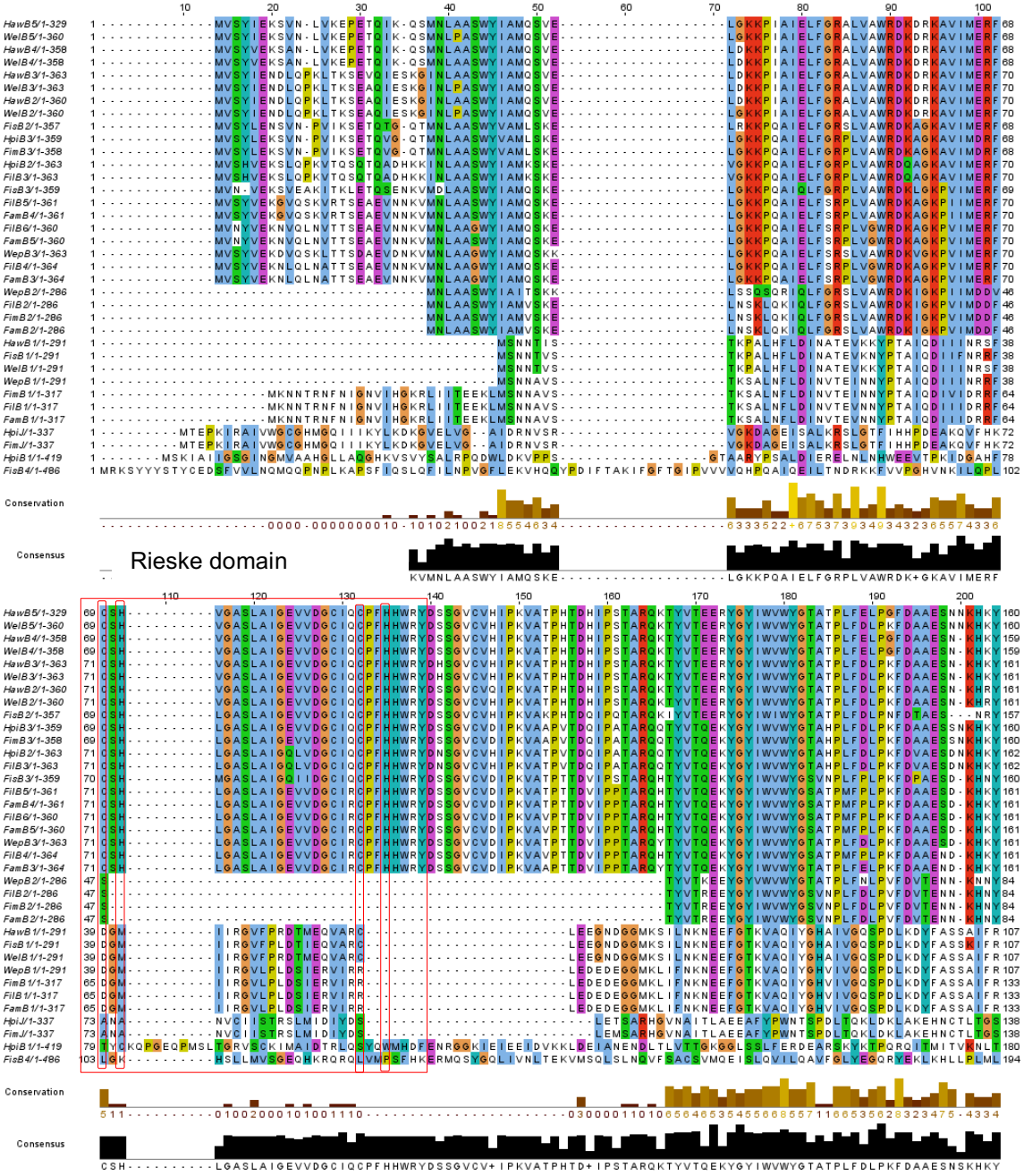
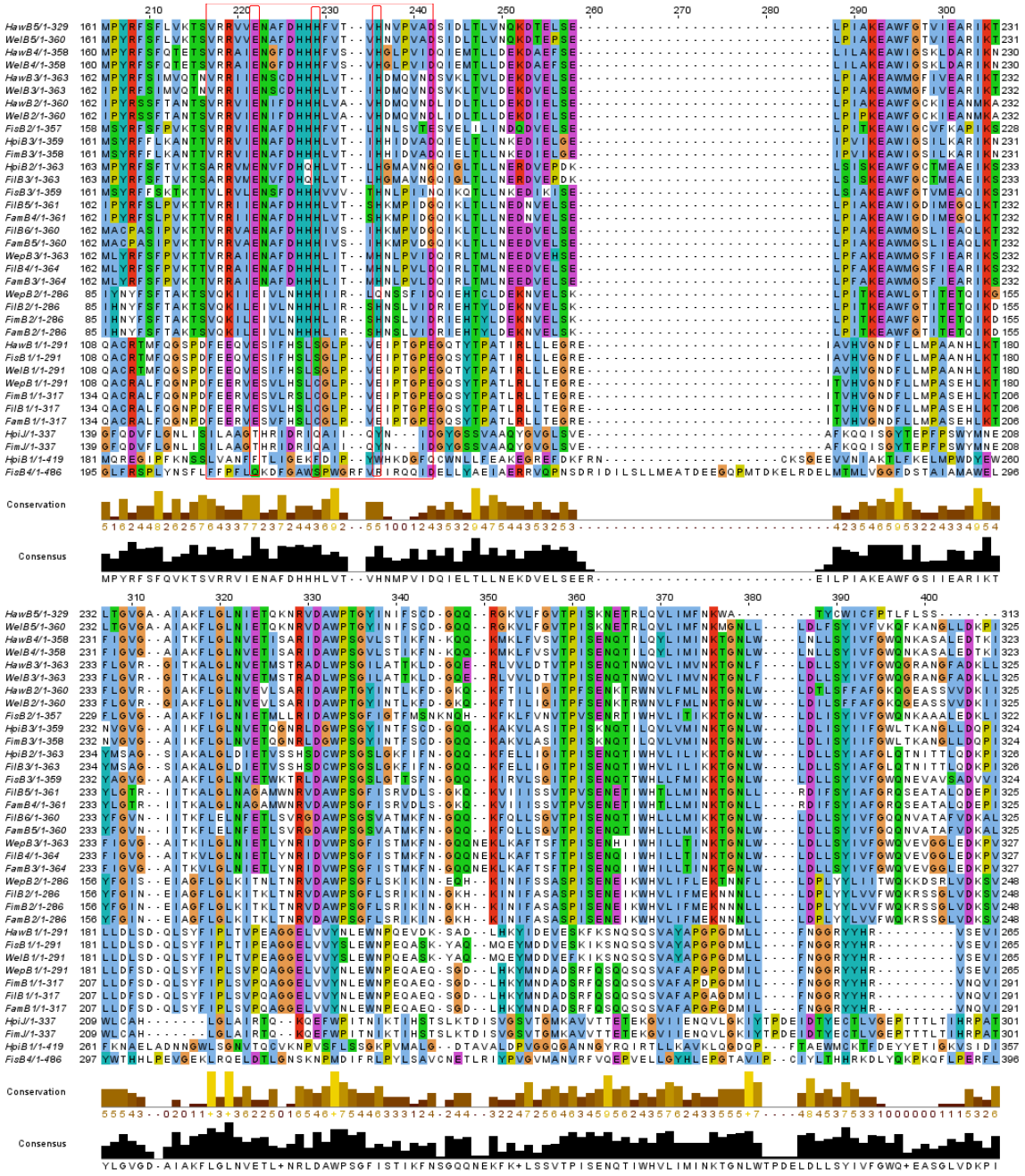


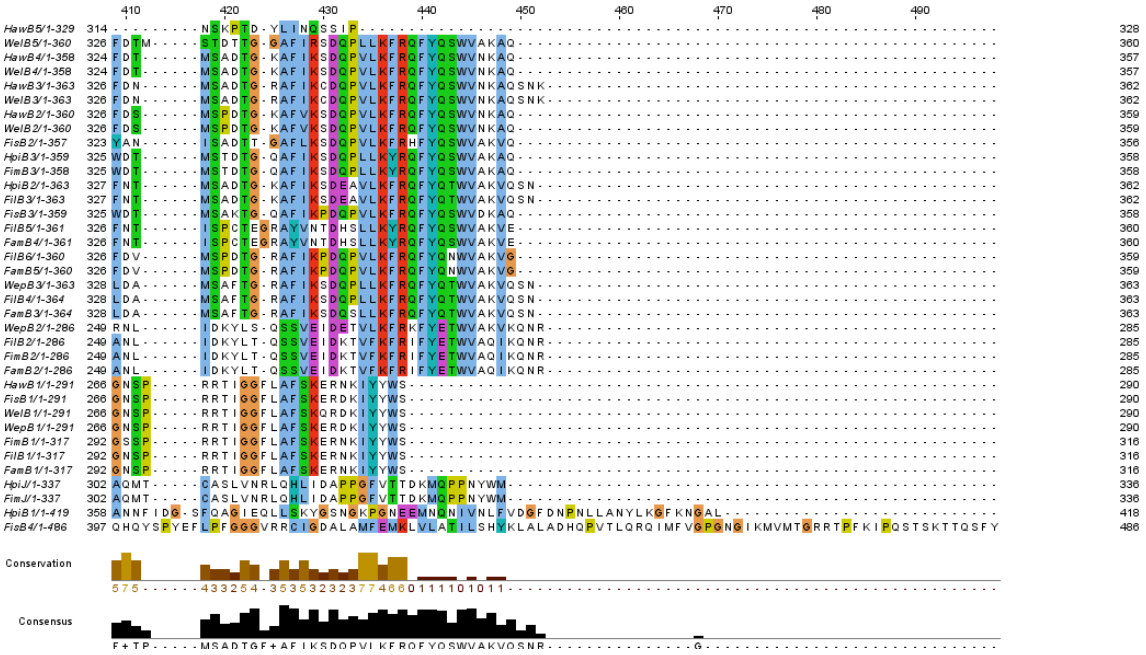
Figure 6-11. Proposed ring rearrangement. A) Transformation from fischerindole to welwitindolinone through oxidation and ring-rearrangement. B) Oxidative C-ring expansion of fischerindole.

6.6 Supplementary information



Mononuclear iron center





SI Figure 6-1. Protein sequence alignment of all oxygenases from gene clusters *fam, fil, wep, hpi, fim, fis, wel, haw*. Two conserved domains includ the Rieske domain (CxH, CxxH) and mononuclear iron domain ($Ex_nHx_nHx_n$).

6.7 Reference

1. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H., Hapalindole/Ambiguine biogenesis is mediated by a Cope rearrangement, C–C bond-forming cascade. *J. Am. Chem. Soc.* **2015,** *137* (49), 15366-15369.
2. Li, S.; Lowell, A. N.; Newmister, S. A.; Yu, F.; Williams, R. M.; Sherman, D. H., Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids. *Nat. Chem. Biol.* **2017,** *13* (5), 467-469.
3. Hillwig, M. L.; Zhu, Q.; Liu, X., Biosynthesis of ambiguine indole alkaloids in cyanobacterium Fischerella ambigua. *ACS Chem. Biol.* **2014,** *9* (2), 372-377.
4. Hillwig, M. L.; Fuhrman, H. A.; Ittiamornkul, K.; Sevco, T. J.; Kwak, D. H.; Liu, X., Identification and characterization of a welwitindolinone alkaloid biosynthetic gene cluster in the stigonematalean Cyanobacterium Hapalosiphon welwitschii. *Chembiochem* **2014,** *15* (5), 665-669.
5. Micallef, M. L.; Sharma, D.; Bunn, B. M.; Gerwick, L.; Viswanathan, R.; Moffitt, M. C., Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol.* **2014,** *14*, 213-230.
6. Moore, R. E.; Cheuk, C.; Patterson, G. M. L., Hapalindoles: new alkaloids from the blue-green alga Hapalosiphon fontinalis. *J. Am. Chem. Soc.* **1984,** *106* (21), 6456-6457.
7. Yang, Z.; Zhang, L.; Zhang, Y.; Zhang, T.; Feng, Y.; Lu, X.; Lan, W.; Wang, J.; Wu, H.; Cao, C.; Wang, X., Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PLoS One* **2011,** *6* (7), e22981.
8. Burgess, R. R., Refolding solubilized inclusion body proteins. *Methods Enzymol.* **2009,** *463*, 259-282.
9. Oganesyan, N.; Kim, S. H.; Kim, R., On-column protein refolding for crystallization. *J. Struct. Funct. Genomics* **2005,** *6* (2-3), 177-182.
10. Zhu, Q.; Hillwig, M. L.; Doi, Y.; Liu, X., Aliphatic Halogenase Enables Late-Stage C-H Functionalization: Selective Synthesis of a Brominated Fischerindole Alkaloid with Enhanced Antibacterial Activity. *Chembiochem* **2016,** *17* (6), 466-470.
11. Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K., Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **2016,** *12* (8), 636-640.
12. Hillwig, M. L.; Zhu, Q.; Ittiamornkul, K.; Liu, X., Discovery of a Promiscuous Non-Heme Iron Halogenase in Ambiguine Alkaloid Biogenesis: Implication for an Evolvable Enzyme Family for Late-Stage Halogenation of Aliphatic Carbons in Small Molecules. *Angew. Chem. Int. Ed.* **2016,** *55* (19), 5780-5784.
13. Hillwig, M. L.; Liu, X., A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014,** *10* (11), 921-923.

14. Lee, J.; Simurdiak, M.; Zhao, H., Reconstitution and Characterization of Aminopyrrolnitrin Oxygenase, a Rieske N-Oxygenase That Catalyzes Unusual Arylamine Oxidation. *J. Biol. Chem.* **2005,** *280* (44), 36719-36727.