Harnessing Cascade Biocatalysis for the Chemoenzymatic Synthesis of Unnatural Hapalindole-Type Metabolites and Further Exploration into their Biosynthesis

by

Robert M. Hohlman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2022

Doctoral Committee:

Professor David H. Sherman, Chair
Professor Roland Kersten
Professor Alison Narayan
Professor Andy White

Robert M. Hohlman

rhohlman@med.umich.edu

ORCID ID: 0000-0002-0361-1611

Acknowledgements

I would be lying if I said I did a Ph.D. all on my own. There are many individuals along the way who offer support when things are going poorly and celebrate with you when things are going well.

I first have to thank my family for all their love and support. They may not fully understand what I did for the past five years but they were there every step of the way reminding me that I could do this when things were looking down. Second, I have to thank my friends, both the ones I made in graduate school and the ones from past experiences. Having them around to get a drink with, play trivia, go to concerts/sporting events, etc. helps keep you sane when your whole world revolves around lab work and science.

Next are the people who I was around day in and day out in the lab. Being able to discuss sports, music, life, etc. helps pass the time when you're waiting for that key experiment to finish up. I was blessed to have such fantastic co-workers in the Sherman lab that I could go to with any question. Picking people's brains has been one of my favorite things to do in graduate school. The amount of information you can learn is incredible.

I have been blessed to have two phenomenal scientific mentors throughout my career so far. At Calvin, I was brought into the world of medicinal chemistry by Dr. Michael Barbachyn and decided I wanted to pursue a research career instead of pursue a Pharm.D degree. At Michigan, my advisor, Dr. David Sherman, fostered a collaborative environment and provided me with a project that fit perfectly with my skill set. My one-on-one meetings with David not only involved talking about science but about many other aspects of life. David's insights and the way he approaches scientific research are things I will never forget.

Finally, I have to thank my fiancée Kaylie Bullock. Who would have thought my life would have changed for the best when Kaylie and I went on our first date 4 years ago? Kaylie and I have balanced me being in graduate school and her being in medical school. There have been challenging moments for both of us but having each other for support has only made our relationship grow stronger. I couldn't have finished this without all her love and support.

Table of Contents

List of Figures

List of Tables

List of Schemes

Abstract

In recent years, the rise in resistance by bacteria, viruses, tumor cells, etc. to already approved therapeutics has led to a greater demand for drug production. To meet this demand, a revival of interest in natural products has opened the door for a once dormant pipeline of biologically active compounds. Natural products are secondary metabolites produced by organisms that aid in their survival (but are not critical for the organism's survival itself). Over the past 40 years, natural products have provided the actual compound (or inspiration) for greater than 50% of all FDA approved drugs. While a diverse library of natural products already exists to be explored as potential therapeutics; many limitations arise. Natural products are generally complex structures that are difficult to produce synthetically in an acceptable amount of steps and yields. Isolation methods from producing strains are also not great enough to meet the demand of drug production. It has long been proposed that the enzymatic machinery from the biosynthetic gene clusters that produce these compounds could be harnessed to not only produce the parent compounds but also derivatives. Today, sequencing technology and protein engineering has made that proposal a reality and opened the door even further for natural product exploration.

This thesis presents results focused on the exploration of the enzymes responsible for the biosynthesis of the hapalindole-type metabolites. Hapalindole-type metabolites are a diverse group of indole alkaloids defined by their polycyclic ring system, various stereoisomers, unique functional groups and promising biological activities. Recently, the biosynthetic gene cluster responsible for their biosynthesis has been elucidated. This revealed a wide range of enzymes that included prenyltransferases, oxygenases, and cyclases. The cyclases, in particular, have drawn interest due to their ability to catalyze a unique three step cyclization cascade: 1) a Cope rearrangement, 2) 6-*exo-trig* cyclization, and 3) terminal electrophilic aromatic substitution (EAS) from a common indole C-3 geranylated (3-GC) intermediate. Utilizing cascade biocatalysis with a prenyltransferase and various cyclases, a chemoenzymatic route to produce unnatural hapalindole-type metabolites has been devised on both milligram and *in vitro* scales in

reaction vessels and cell-free protein synthesis reactions. Further medicinal testing and semisynthetic modifications to this library could provide a drug candidate to combat increasing drug resistance.

Another member of the hapalindole-type metabolites, the ambiguines, showcases similar biological activity but a greater amount of structural modifications. One of the most interesting is the addition of a fifth (E) ring. To date, it is unknown how this fifth ring is formed but it is proposed to arise from reactions catalyzed by Rieske-type oxygenases. To tackle this question, two routes were explored; one focusing on the Rieske-type oxygenases themselves and another exploring a chemoenzymatic synthesis of a pentacyclic ambiguine. While the Rieske-type oxygenases proved challenging to elucidate, the chemoenzymatic route has shown promise. To date, an optimized, efficient biocatalytic method for generating an unnatural ambiguine derivative, 12-*epi*-ambiguine H nitrile, has been developed. Efficient biosynthesis of the unnatural derivative is crucial to further synthetic chemistry or biocatalysis efforts to generate the E-ring. In addition, during early phases of this work, we uncovered an unexpected new metabolite of a one pot reaction previously unknown to the Stig Cyclases. This finding strengthens the Cope rearrangement hypothesis and provides a potential synthon towards total synthesis efforts of a diverse range of compounds. These works showcase the potential natural product enzymes have to produce complex metabolites and provide a new route to further develop the diverse library of already known natural products.

Chapter 1
Introduction

1.1 Hapalindole-type metabolites & drug discovery

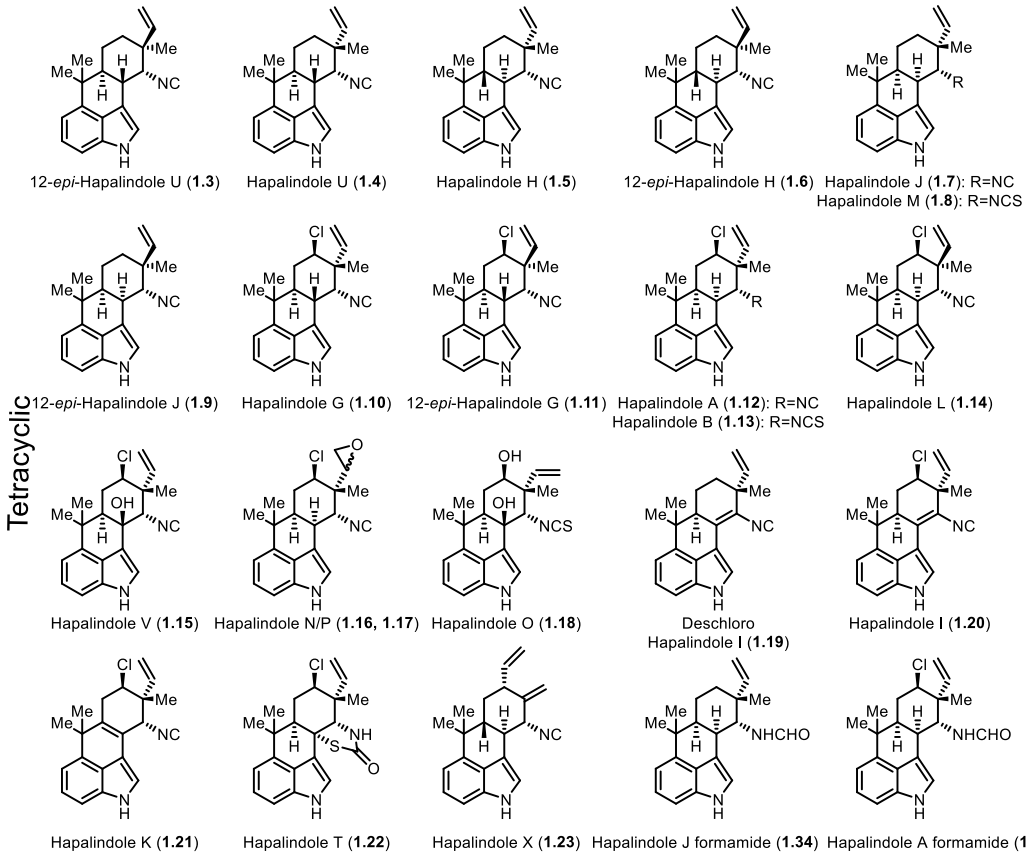
1.1.1 Introduction

Terrestrial, marine and aquatic cyanobacteria thrive in many different habitats worldwide. Due to their resilience in even the most extreme environments, they have been explored extensively for specialized metabolites that may aid in their survival.[1] This has led to the discovery of biologically active metabolites such as cryptophycin-1, dolastatin, noscomin and pahayokolide A.[2] A derivative of dolastatin (monomethyl auristatin E) serves as a cytotoxic payload in several FDA-approved antibody drug conjugates.[3]

In 1984, Richard Moore and colleagues isolated the first hapalindoles from the cyanobacterial strain *Hapalosiphon fontinalis*. The authors were searching for a compound responsible for the antimicrobial effects against *Anabaena oscillarioides*. Their work uncovered hapalindoles A **1.12** and B **1.13**.[4] Following initial isolation, hapalindoles C-Q (**1.5**, **1.7**, **1.8**, **1.10**, **1.14**, **1.16-1.18**, **1.20**, **1.24-1.25**, **1.28**, **1.30-1.31**) and T-V (**1.4**, **1.15**, **1.22**) (Figure 1.1) were isolated from the same strain.[5] Investigation of other cyanobacterial strains have uncovered a variety of tri-, tetra- and pentacyclic hapalindole–type alkaloids. These compounds have been shown to have a wide scope of biological activities, which include antibacterial,[1,6–9] antimycotic,[1,4,10] insecticidal,[11,12] antitumoral,[13–15] and immunomodulatory.[16] To date, over 80 hapalindole-type metabolites have been isolated from 18 distinct cyanobacterial strains, all arising from the Stigonematales order.[1]

The polycyclic compounds are divided into four subgroups, the hapalindoles (tri-/tetracyclic), fischerindoles (tetracyclic), ambiguines (tetra-/pentacyclic) and welwitindolinones. The core fused ring system arises from a *cis*-indole isonitrile moiety and a geranyl monoterpene unit, and contains similar structural features such as a functionalized D-ring, which include late-stage C-H oxidation and halogenation (Figure 1.2).

In 2009, Gault and Marler published a handbook on cyanobacteria in which Moffitt and Burns contributed a chapter with the first comprehensive review on the hapalindoles.[17] Bhat and co-workers subsequently published an extensive review on the hapalindole-type metabolites.[1] This review highlighted each class within the family of molecules, their biological properties, the proposed biosynthesis, and significant advances in their total syntheses. Since 2014, remarkable progress has been made on their biosynthetic assembly and significant advances in total syntheses of the pentacyclic ambiguines have also been reported.



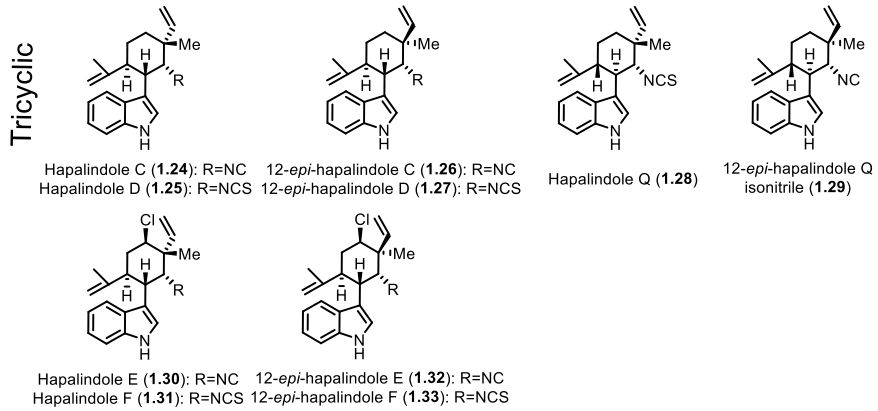


Figure 1.1: Currently reported hapalindole compounds. Grouped into tetracyclic and tricyclic categories.

1.2 Structural diversity

1.2.1 Hapalindoles

The hapalindoles are the most prevalent group of compounds comprising this class of indole alkaloids, and account for at least 31 molecules. The majority of the hapalindoles possess a polycyclic system containing four fused rings with the key connection between the C-4 and C-16 positions (Figure 1.2). Stereocenters at the C-10, C-11, C-12 and C-15 positions endow these structures with a spectrum of stereoisomers. Adding further to this diversity are a group of tricyclic hapalindoles that lack the C-C bond between the C-4 and C-16 positions. Lastly, the presence of a halogen or hydroxyl group in select compounds brings more variety to an already distinct family of indole alkaloids (Figure 1.1). Most recently, two new compounds, hapalindole A formamide **1.34** and hapalindole J formamide **1.35** were isolated from the branching cyanobacterium *Hapalosiphon* sp. CBT1235.[16]

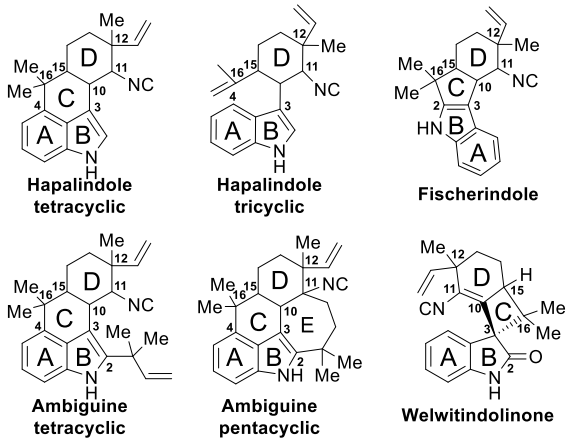


Figure 1.2: Common hapalindole-type indole alkaloid core. Letter labels for each ring, key carbon stereocenters and fused ring points are shown.

1.2.2 Fischerindoles

First isolated from *Fischerella muscicola* in 1992 by Moore and colleagues, the fischerindoles represented a structurally distinct, but related group of compounds.[18] These metabolites contain many of the same characteristics that defined the hapalindoles; a polycyclic system containing four rings, stereocenters at the C-10, C-11, C-12 and C-15 positions and the presence of a halogen or oxygenation in some of the compounds. The main difference is the key connection is now between the C-2 and C-16 positions as opposed to the C-4 and C-16 positions. Isolated in 2012, the Orjala group discovered fischerindoles that contain a nitrile moiety instead of the hallmark isonitrile (Figure 1.3, **1.41**, **1.43**, **1.44**). A nitrile group had previously been observed only in the ambiguines (Figure 1.4, **1.51**, **1.61**).[19]

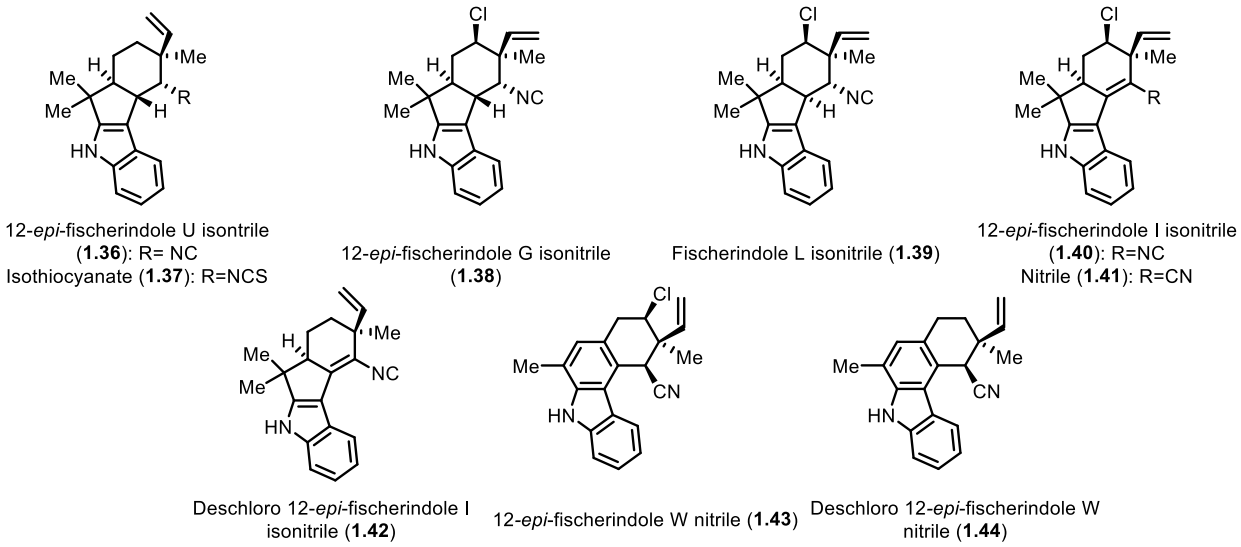


Figure 1.3: Currently reported fischerindole compounds.

1.2.3 Ambiguines

Further work by the Moore group in 1992 uncovered ambiguines A-F (**1.45-1.50**) from *Fischerella ambigua*, revealing pentacyclic hapalindole-type indole alkaloids for the first time (Figure 1.4).[10] The seven membered E-ring encompasses a connection between the C-2 and C-11 positions. The search for additional ambiguines expanded when the Orjala group identified metabolites that contain a 6-membered E-ring (Fischambiguines A **1.62** and B **1.63**).[20] Because ambiguines A **1.45**, B **1.46**, C **1.47**, and H **1.52** are comprised of a reverse prenyl group at the C-2 position, the E-ring is proposed to arise from further 7-*endo-trig* (ambiguine) or 6-*exo-trig* (fischambiguine) cyclization via a radical mechanism catalyzed by a putative Rieske-type protein. Further halogenation and oxidation of both the E-ring and core hapalindole scaffold give rise to further structural diversity observed in this family of molecules.

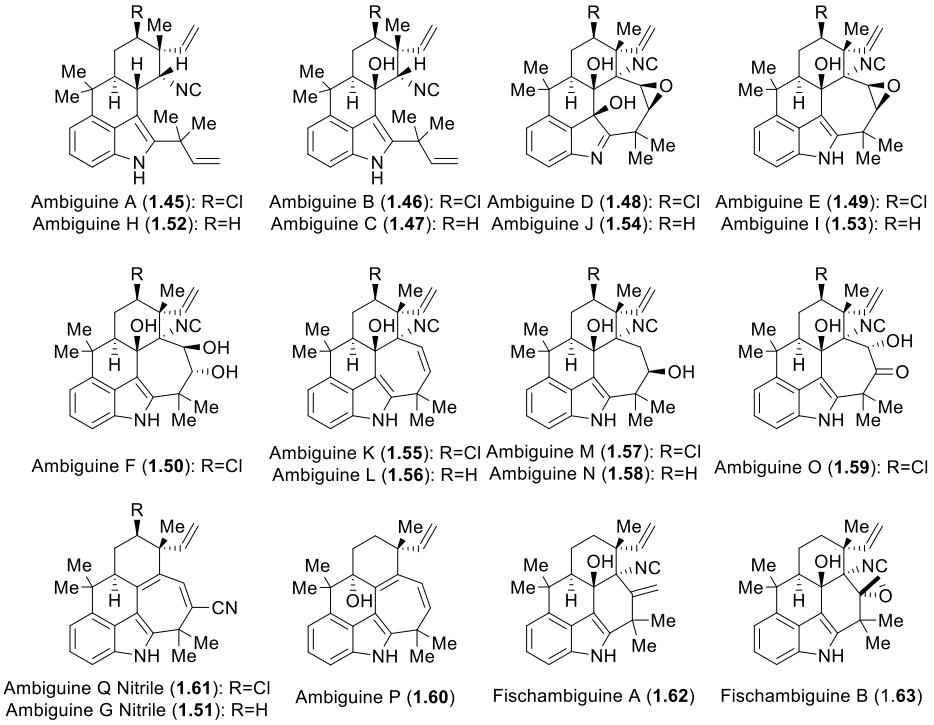


Figure 1.4: Currently reported ambiguine compounds.

1.2.4 Welwitindolinones

Moore's group reported that the lipophilic extract of *Hapalosiphon welwitschii* reversed P-glycoprotein-mediated multi-drug resistance in human ovarian adenocarcinoma cells.[11] Upon isolation, the compound responsible for this action was identified as *N*-methylwelwitindolinone C **1.69**, a similar, but structurally unique compound related to the hapalindoles. From this strain, six structurally related congeners were also discovered (**1.64**, **1.65**, **1.66**, **1.67**, **1.68**, **1.70**).[11] The welwitindolinones represent a unique structural class in this family of indole alkaloids. Their defining feature is the bicyclo[4.3.1]decane ring system except for welwitindolinone A isonitrile **1.64**, which contains a spirooxindole moiety. Every compound is also oxidized at the C-2 position, which raises additional questions about the biogenesis of these highly functionalized molecules (Figure 1.5).

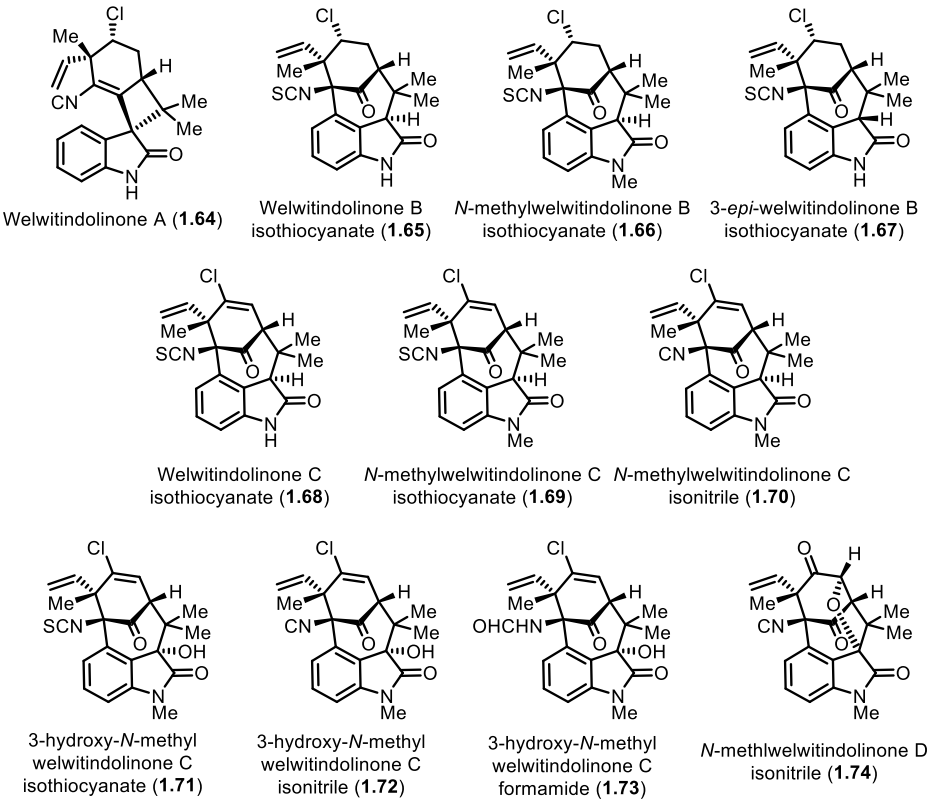


Figure 1.5: Currently reported welwitindolinone compounds.

1.2.5 Other related hapalindole-type indole alkaloids

There have been several hapalindole-type indole alkaloids isolated from various extracts that do not fit directly within the families discussed above (Figure 1.6). Isolated from another *Fischerella* strain, the hapalindolinones (**1.75-1.76**) contain an oxidized C-2 position (oxo-indole) and a highly strained spiro-cyclopropane moiety.[21] Moore and colleagues also reported the isolation of several other oxidized hapalindoles from *Hapalosiphon fontinalis* (**1.77-1.82**, **1.84-1.86**).[22,23] More recently, the Orjala group isolated 13-hydroxydechlorofontonamide **1.83** from *Fischerella muscicola*, which alongside the fontonamides and the hapalonamides, lacks the indole B-ring.[14]

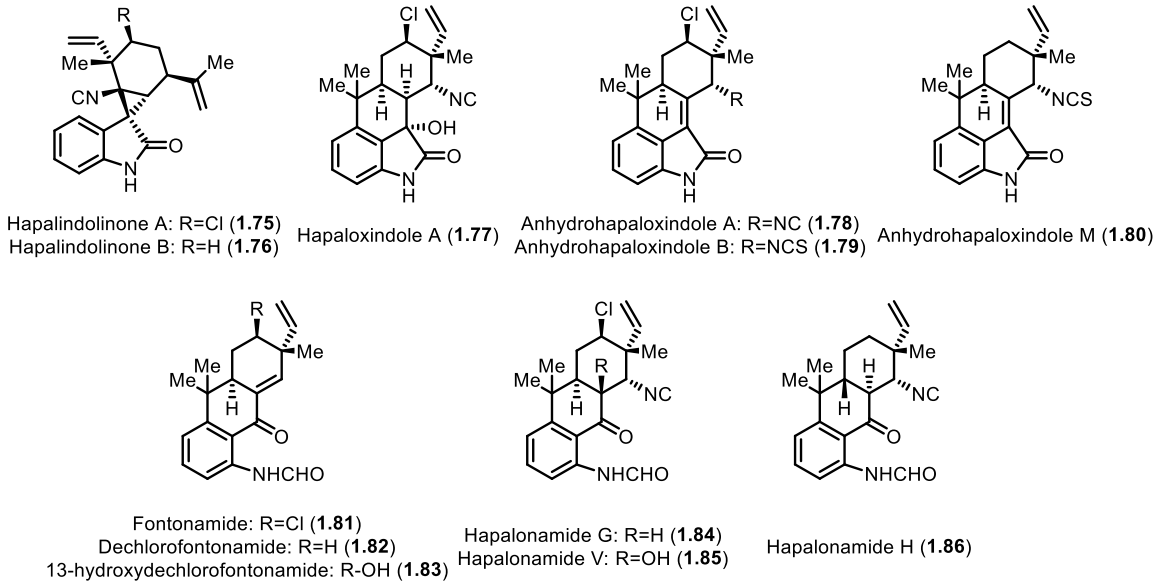


Figure 1.6: Currently reported hapalindole-type indole alkaloids.

1.3 Biological activity

1.3.1 Antimicrobial

Many isolated hapalindole-type molecules have displayed antibiotic activity against a wide variety of both Gram-positive and Gram-negative bacteria. Hapalindoles A-H (**1.5**, **1.10**, **1.12-1.13**, **1.24-1.25**, **1.30-1.31**) were reported to have MIC values ranging from 0.06 to 64 µg/mL against Gram-positive pathogens such as *Staphylococcus aureus* and *Streptococcus pneumoniae*.[1] Hapalindoles J-Q (**1.7-1.8**, **1.14**, **1.16-1.18**, **1.21**, **1.28**) and T-V (**1.4**, **1.15**, **1.22**) were also shown to be active against numerous multi-drug resistant (MDR) bacteria.[7] Hapalindole T **1.22**, isolated from a *Fischerella* strain growing on the bark of the Neem tree, showed very low concentration values for inhibition of MDR bacteria including 0.25 µg/mL against *S. aureus* and 2.0 µg/mL against *Pseudomonas aeruginosa*.[6]

In 2007, ambiguines H **1.52** and I **1.53** were shown to have MIC values of 0.625 µg/mL and 0.078 µg/mL, respectively against *Staphylococcus albus* and 1.25 µg/mL and 0.312 µg/mL against *Bacillus subtilis*.[7] These values were similar, if not better than the clinical antibiotic streptomycin. Orjala and co-authors reported numerous ambiguines with low micromolar MIC values against *Mycobacterium tuberculosis, Bacillus anthracis* and *S. aureus.*[8] These values were similar to approved antibiotics such as rifampin, ciprofloxacin and gentamicin.

It is believed that the presence of halogens, hydroxyl groups, and epoxides give this family of indole alkaloids their biological activity against numerous pathogens.[7] However, the

target of these molecules remained unknown until 2001 when Smith and colleagues showed inhibition of the bacterial RNA polymerase with 12-*epi*-hapalindole E isonitrile **1.32** from a *Fischerella* strain. The authors calculated an IC_{50} value of 500 μM for the RNA polymerase and the compound was shown to prevent elongation of the growing nucleic acid chain. However, the authors note that initial assays only required an IC_{50} of 3 μM to inhibit growth of *B. subtilis* suggesting that there is likely more than one biological target for the hapalindole-type alkaloids.[9] As such, the result reported should be regarded as preliminary and requires further rigorous studies.

1.3.2 Antimycotic

Initial discovery of the hapalindoles was driven by the antimicrobial activity observed against *Anabaena oscillarioides.*[4] Since then, numerous hapalindole-type alkaloids have also shown antimycotic activity against a wide variety of fungi including *Aspergillus*, *Penicillium*, *Fusarium*, and *Candida* strains.[7] Ambiguines A-F (**1.45-1.50**) exhibited low μg/mL values against *C. albicans* and *Trichophyton mentagrophytes*.[10] While these results are encouraging, further medicinal chemistry efforts will be necessary to improve potency compared to already approved fungicides.

1.3.3 Insecticidal

While not all hapalindole-type alkaloids have been screened for insecticidal activity, some have shown promise against specific agricultural pests. Becher and colleagues noted that at 26 μM 12-*epi*-hapalindole J isonitrile **1.9** killed 100% of growing *Chironomus riparius* larvae over 48 hours. Similar activity was observed with 12-*epi*-hapalindole C **1.26**, hapalindole L **1.14** and **1.32**.[12] While the Moore group noted insecticidal activity for *N*-methylwelwitindolinone C isothiocyanate **1.69**, more recent attention has focused on the compound's antitumoral activities and P-glycoprotein transporter inhibition.[13,15]

1.3.4 Antitumoral and other activity

Another promising area for hapalindole-type metabolites is in the realm of antitumoral therapeutics. Hapalindoles A **1.12**, C **1.24**, H **1.5**, I **1.20** and J **1.7** along with fischerindole L **1.39** showed low micromolar inhibition of many human carcinoma cell lines including colon, breast,

large cell lung and glioblastoma.[14] As noted earlier, Moore's group uncovered the P-glycoprotein inhibition of **1.69** (IC_{50}=3.03 μM), *N*-methylwelwitindolinone C isonitrile **1.70** (IC_{50}=0.12 μM) and welwitindolinone C isothiocyanate **1.68** (IC_{50}=0.13 μM) in breast cancer cell lines.[13] P-glycoprotein mutations are a common resistance mechanism observed in cancer cells and its inhibition allows for more effective chemotherapy treatments. Roughly one year later, Zhang and Smith revealed another mechanism of action for **1.68** against ovarian cancer cell lines. The authors showed that the welwitindolinone analog caused disorganization of the microtubules at the reported IC_{50} value of 0.13 μM and by 3.3 μM showed complete loss of the microtubules.[15]

In other therapeutic areas, **1.12** showed inhibition of T-cell proliferation. As reported recently by Grundemann and colleagues, the IC_{50} value of **1.12** (1.56 μM) shows promise as a new drug lead to treat auto-immune disorders. Interestingly, the authors also uncovered hapalindoles D **1.25**, M **1.8** and hapalindoles A and J formamide (**1.34**, **1.35**) from *Hapalosiphon* sp. CBT1235. None of these compound showed the same potency compared to **1.12**, which indicated that a chlorine at the C-13 and isonitrile functionality at C-11 position were critical for activity (Figure 1.1).[16] Another potential biological target was uncovered in 2014 by Botana et al. Using isolated **1.7**, **1.14**, **1.26**, and **1.32**; these investigators reported the hapalindoles to be sodium channel modulators. Neuroblastoma cell lines were treated with veratridine and a bis-oxonol fluorescent dye to activate sodium channels and monitor the membrane potential changes induced by hapalindole compounds. Addition of the hapalindoles reduced the intensity of the dye, showing a decrease in membrane potential. Using linear regression, the IC_{50} values for almost all the compounds were determined (4.8 μM for **1.26**, 6.7 μM for **1.32** and 10.6 μM for **1.14**) suggesting that the hapalindoles may be a starting point for development of more potent neuroactive agents.[24]

1.4 Overview of total synthesis efforts

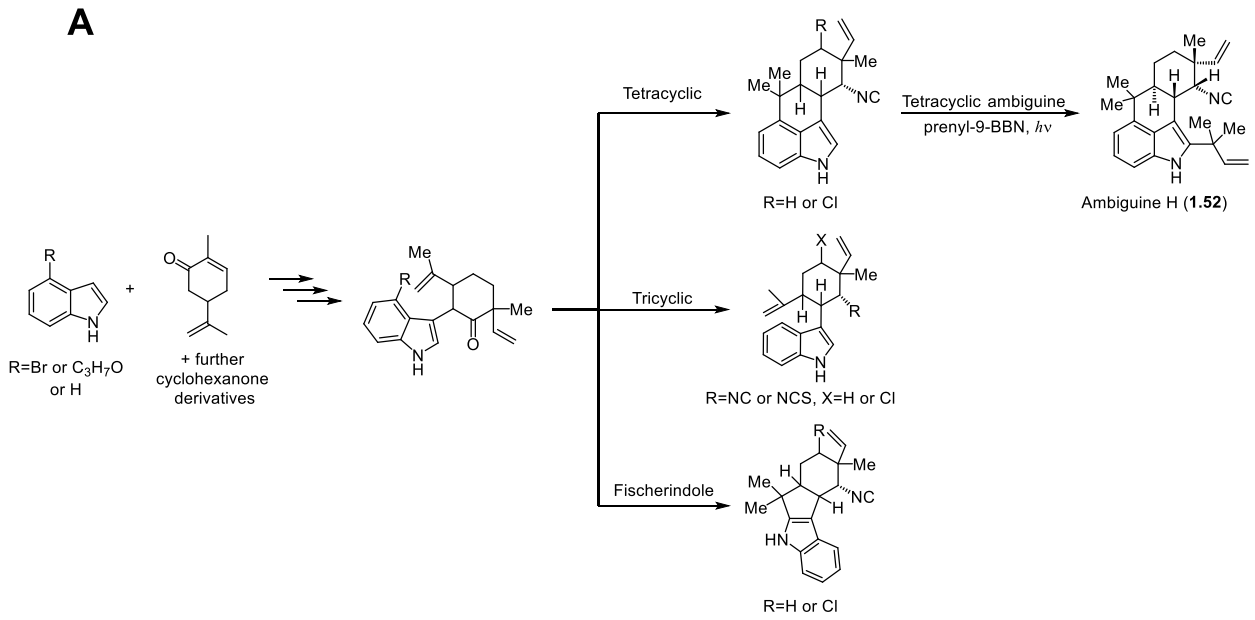
1.4.1 Introduction

To date, 27 hapalindole-type metabolites have been synthesized, which include efforts to develop novel routes or improve earlier ones.[25–60] The overview below highlights the general routes used in the total syntheses of each class of compounds, which sets the stage for contrasting the widely variant strategies employed between total synthesis and biosynthesis

(discussed in the next section) of these molecules. Rawal's excellent 2014 review contains in-depth explanations of the total synthesis efforts, which will not be covered here.[1]

1.4.2 General routes to tri- and tetracyclic hapalindoles, ambiguines, fischerindoles and pentacyclic ambiguines

Routes toward the synthesis of tri- and tetracyclic hapalindoles and ambiguines along with fischerindoles rely on functionalized indole derivatives coupled with various cyclohexanone derivatives to produce a tricyclic-like intermediate. Final tailoring reactions complete the construction of tricyclic hapalindoles, while Friedel-Crafts type reactions are used to complete the tetracyclic framework of hapalindoles and fischerindoles (Figure 1.7 A).[26,28,29,33–35,37,48–53,57–59] Currently, ambiguine H remains the only synthesized tetracyclic ambiguine with prenyl-9-BBN and light used to induce a rearrangement to install the reverse prenyl group at the C-2 position (Figure 1.7 A).[29,33,38,58] Two groups have synthesized pentacyclic ambiguine P **1.60**, and both routes start by cross-coupling indole and either (*S*) or (*R*)-carvone. Different strategies are then employed to install the ambiguine E-ring followed by further transformations to complete the molecule (Figure 1.7 B).[54,55] Most recently, Hu and Rawal reported the total synthesis of the first chlorine containing pentacyclic ambiguine, ambiguine G Nitrile **1.51** with a [4+3] cycloaddition being the key reaction to construct the E-ring.[56]



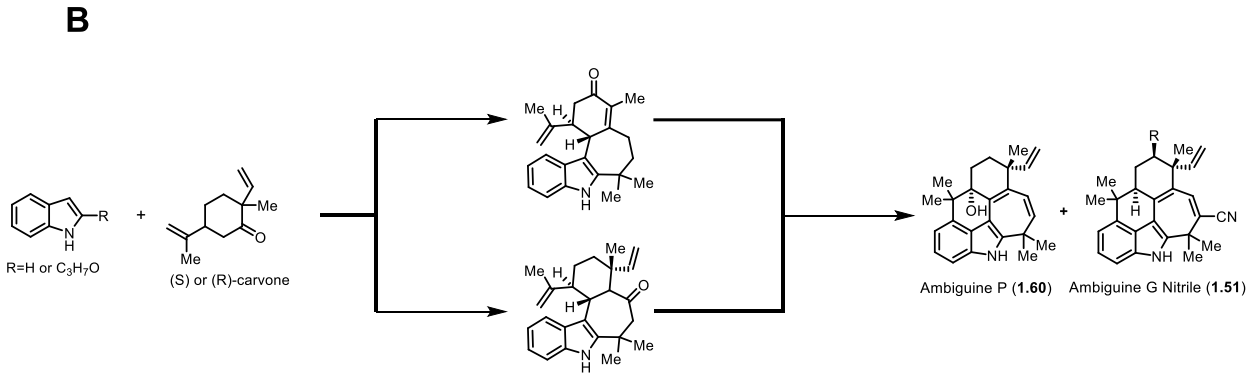


Figure 1.7: Hapalindole, fischerindole and ambiguine general synthetic strategies. (**A**): General synthesis route used to produce tri- and tetracyclic hapalindoles along with fischerindoles. A functionalized indole derivative is coupled with a cyclohexanone derivative to produce a tricyclic-like intermediate. Final tailoring reactions complete the routes to tricyclic hapalindoles, and tetracyclic hapalindoles and fischerindoles, (**B**): General synthesis route used to produce pentacyclic ambiguines. Ambiguine P **1.60** and ambiguine G nitrile **1.51** have been synthesized using these methods.

1.4.3 General routes to welwitindolinones

Various routes have been explored to construct the unique bicyclo[4.3.1]decane ring system or the spirooxindole functionality of the welwitindolinones. Baran et al. uncovered an acid catalyzed rearrangement reaction from 12-*epi*-fischerindole I isonitrile **1.40**, which gives rise to welwitindolinone A **1.64** (Figure 1.8 A).[57,58] Reisman and Wood explored another route to **1.64** and the spirooxindole moiety. Using a synthesized hydroxyl-enone derivative and a diazo Grignard reagent, the spirooxindole can be effectively produced leading to the total synthesis of **1.64** (Figure 1.8 B).[36,60] For the bicyclo[4.3.1]decane ring system, Rawal and Garg developed efficient routes starting from functionalized indole and cyclohexanone derivatives. Rawal utilized a Friedel-Crafts type reaction to complete the ring system, while Garg has relied upon a unique benzyne intermediate to produce the framework. For both groups, further chemistry led to the total synthesis of various welwitindolinone compounds (Figure 1.8 C).[25,39–47]

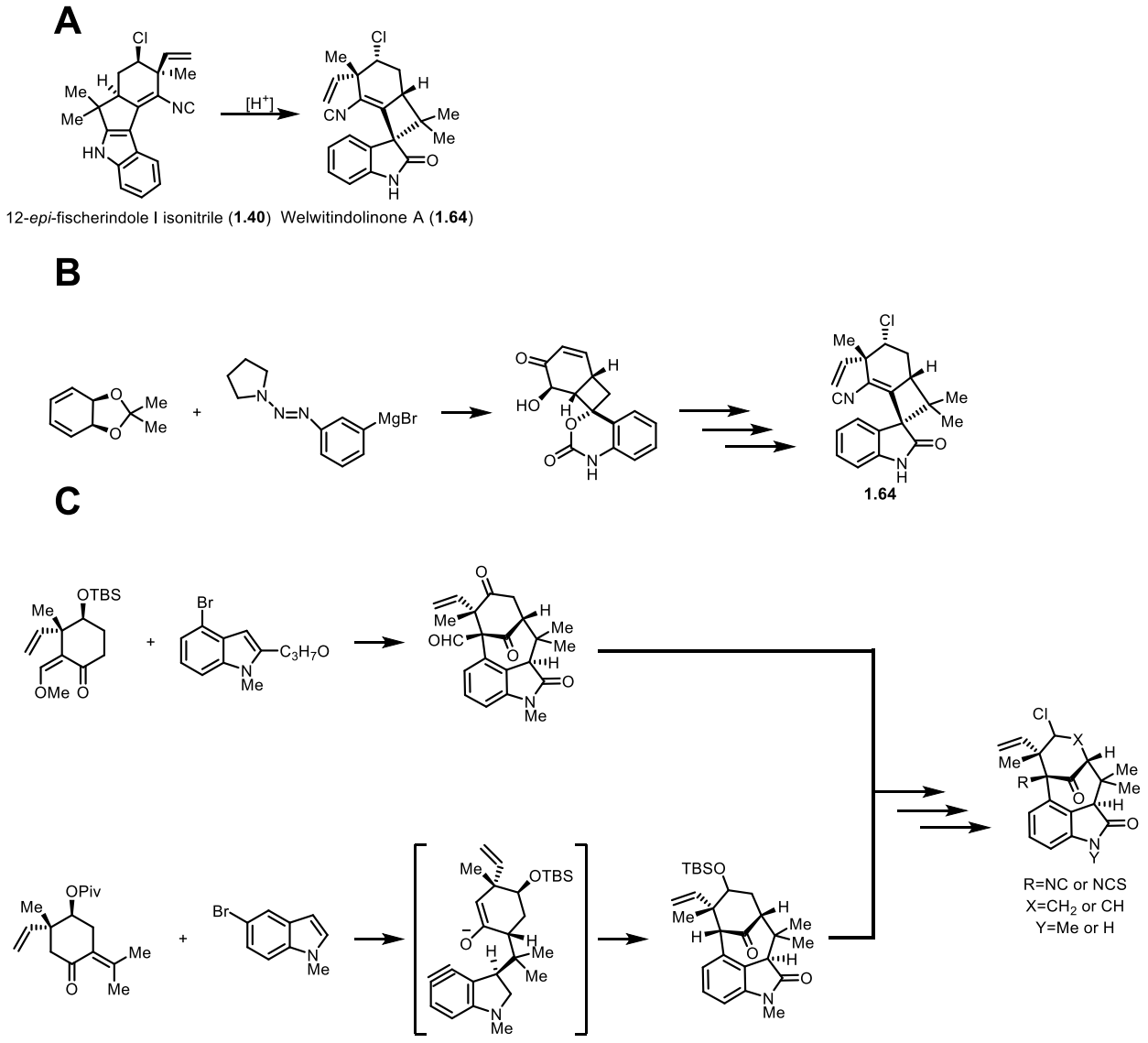


Figure 1.8: Welwitindolinone general synthetic strategies. (**A**) Acid-catalyzed rearrangement route employed by Baran et al. to produce **1.64**.[57,58] (**B**) The route used by Reisman et al. in the total synthesis of **1.64**. A hydroxyl-enone derivative is reacted with a diazo Grignard reagent to aid in producing the spirooxindole moiety.[36,60] (**C**) General synthesis routes used by Rawal and Garg to produce various welwitindolinone compounds. Both routes use functionalized cyclohexanone and indole derivatives with key intermediates highlighted.[25, 39-47]

1.4.4 Summary

The general synthetic approaches described above highlight the creative strategies employed by a number of groups to access the hapalindole-type natural products. Alternative routes have also been utilized to produce the compounds described above. The isolation of numerous natural products and total synthesis efforts helped motivate initiation of biosynthetic studies to determine how the compounds are produced by the Stigonematales order of cyanobacteria. Following the identification of a series of hapalindole-type biosynthetic gene

clusters through genome sequencing,[61–70] a period of intense exploration has led to surprising and fascinating conclusions, described below.

1.5 Hapalindole biosynthesis

1.5.1 Early biosynthetic proposals

Concurrent with their discovery, Moore and colleagues developed a hypothesis regarding hapalindole biogenesis. By examining each compound, they surmised an indole isonitrile core **1.1** and geranyl monoterpene subunit, and reasoned that every hapalindole-type compound was derived from an initial tricyclic hapalindole intermediate. A chloronium ion or proton-catalyzed enzymatic reaction was postulated to produce the tricyclic intermediate. Further enzyme catalyzed acidic condensation reactions were proposed to complete the tetracyclic hapalindole and fischerindole core. For the welwitindolinones, Moore and colleagues also proposed that an early stage oxidation of the indole ring played a role in an acid-catalyzed cyclization reaction to produce the unusual core.[11] Isolation of ambiguine H **1.52** led to the suggestions that the pentacyclic ambiguines are produced by a reverse prenylation with dimethylallyl pyrophosphate (DMAPP) followed by an enzyme catalyzed cyclization reaction.

In 2006, Raveh and Carmeli proposed a new biosynthetic mechanism for assembly of the hapalindole core. This was motivated by their inability to isolate any tricyclic hapalindole compound from a *Fischerella* strain, and they hypothesized a condensed chloronium ion or proton catalyzed cyclization reaction to produce the hapalindole system.[7] In 2008, Baran revised the hypothesis for welwitindolinone formation suggesting that welwitindolinone A **1.64** arises from an oxidative ring contraction from 12-*epi*-fischerindole I **1.40** followed by epoxide formation and rearrangement to produce welwitindolinone B isonitrile (Figure 1.9).[37]

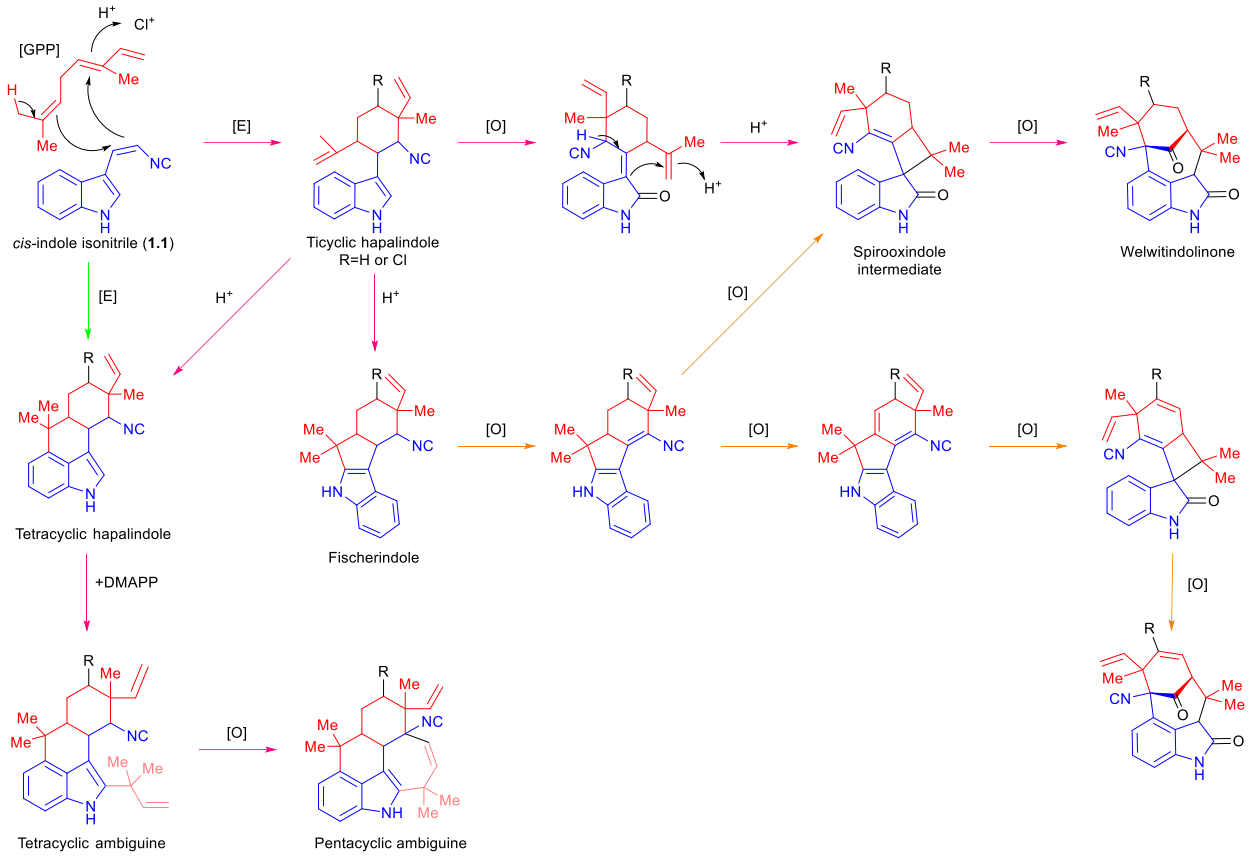


Figure 1.9: Summary of previously proposed biosynthetic mechanisms. The *cis*-indole isonitrile core (**1.1**) is highlighted in blue and the geranyl monoterpene unit (GPP) is highlighted in red. Magenta arrows highlight the mechanism hypothesized by Moore et. al.[11] Green arrows highlight the mechanism hypothesized by Carmeli et. al.[7] Orange arrows highlight the mechanism hypothesized by Baran et. al.[37]

1.5.2 Biosynthetic gene cluster analysis and revised proposal for hapalindole biogenesis

Recent advances in bioinformatics have enabled more in depth analysis of the biosynthetic gene clusters (BGCs) of the hapalindole-type producing cyanobacterial strains. In 2014, the Liu group published the first annotated BGC from *Fischerella ambigua* UTEX 1903 (*amb*).[61] In their work with the *amb* gene cluster, they were able to use heterologous expression to characterize several proteins from the BGC. AmbI1-3 were shown to be homologues of the isonitrile synthases IsnA and IsnB. These proteins have been previously shown to produce the *trans*-indole isonitrile derivative.[71,72] However, AmbI1-3 were shown to produce the *cis*-indole isonitrile **1.1**, a proposed subunit for hapalindole metabolites. This unique isonitrile functionality is believed to be derived from a β-keto imine formation between L-tryptophan and ribulose-5-phosphate followed by two retro-aldol type condensations.[71–73] They also annotated three genes that encode prenyltransferases, AmbP1-3. AmbP2 was uncovered to catalyze formation of

geranyl pyrophosphate (GPP) from isopentenyl pyrophosphate (IPP) and DMAPP. AmbP3 was found to prenylate hapalindole G **1.10** with DMAPP to afford ambiguine A **1.45**, thus confirming Moore's original hypothesis that the tetracyclic ambiguines are the result of a reverse prenylation with DMAPP. A small substrate screen was performed with various other indole derivatives, but none of them were found to be prenylated by AmbP3 suggesting that the terpenoid component in the cyclization of the hapalindole core was needed for further prenylation. The authors did not examine AmbP1, but hypothesized that it plays a role in geranylating **1.1** with GPP to give an intermediate before the cyclized core is completed. The pathway was found to contain five non-heme iron dependent oxygenases, including four Rieske-type proteins (AmbO1-5). These proteins were hypothesized to mediate halogenation and further oxidative modifications to the hapalindole and ambiguine cores (Figure 1.10).[61] Later, AmbO5 was characterized to halogenate various hapalindole, fischerindole and ambiguine metabolites.[74]

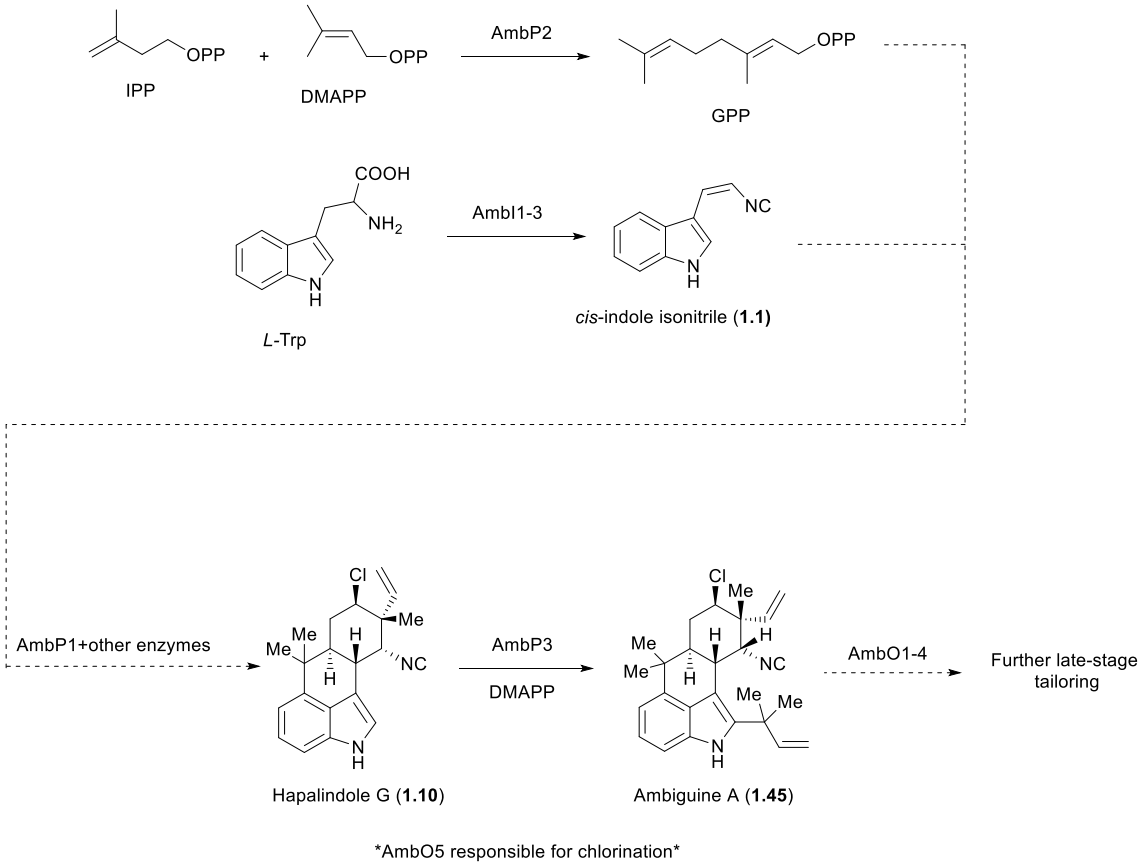


Figure 1.10: Summary of Hillwig et al.[61,69] discoveries. All derived from annotation of the *amb* BGC from *Fischerella ambigua* UTEX 1903. Solid arrows represent enzymes whose function was confirmed. Dotted arrows represent hypothesized enzyme functions.

Soon thereafter, Liu and co-workers published the BGC of a welwitindolinone producing strain, *Hapalosiphon welwitschii* UTEX B1830 (*wel*).[62] They discovered many similarities to the *amb* gene cluster, suggesting horizontal gene transfer across the Stigonematales order of cyanobacteria. In order to confirm that the annotated homologues performed the same function, heterologous expression was again employed to access the proteins. WelP2 (homolog of AmbP2) was found to function as a geranyl diphosphate synthase. WelI1-3 were responsible for biosynthesis of **1.1**, consistent with the function of AmbI1-3. However, there were eight proteins annotated that did not correlate with any protein from the *amb* gene cluster. One of the proteins, WelM, was determined to be a SAM-dependent methyltransferase responsible for the N-methylation of the welwitindolinone core. Enzyme kinetic studies revealed that welwitindolinone C isothiocyanate **1.68** is likely the native substrate for WelM. Substrate scope studies of WelM revealed that the enzyme is highly selective for an oxindole core and likely is the final biosynthetic enzyme in the formation of *N*-methylwelwitindolinone C isothiocyanate **1.69**. Five non-heme iron dependent oxygenases (WelO1-5) shared moderate (61-79%) identity sequences with their *amb* counterparts; nevertheless, they were all hypothesized to play a role in late-stage tailoring (Hillwig and Liu later characterized WelO5 as the halogenase responsible for chlorinating **1.36** and **1.26**[69]).[62]

Building from this initial gene cluster analysis, Micallef et al. performed an extensive bioinformatic analysis from the genome sequences of numerous Stigonematales cyanobacterial strains.[63] Seven strains were analyzed, including *Fischerella* and *Hapalosiphon* strains, in addition to *Westiella intricata* UH strain HT-29-1. The goal of the study was to compare numerous hapalindole/fischerindole BGCs and examine similarities and differences. The majority of the annotated gene clusters showed overall similarities in architecture, with notable differences. Welwitindolinone gene clusters were all highly conserved but were distinct from hapalindole and ambiguine BGCs in the sequence identity of their late-stage tailoring enzymes (i.e. *N*-methyltransferases, oxygenases). In all strains, the enzymes that catalyze formation of **1.1**, GPP, and prenyltransferases were highly conserved. Interestingly, the isonitrile synthases from *Westiella intricata* UH strain HT-29-1 were found to produce both **1.1** and its *trans* isomer uncovering the first isonitrile synthases from a hapalindole-type producing strain to generate the *trans* isomer. Based on phylogenetic analysis, the authors hypothesized that gene cluster transmission is vertical (rather than horizontal as suggested by Liu) because their analysis

identified a monophyletic clade between all the strains examined in the study.[63] This extensive study set the ground work for further elucidation of the biosynthetic pathway responsible for hapalindole-type metabolite formation.

In 2013, the Sherman lab initiated a study to examine the prenyltransferases and proteins with "unknown" annotations responsible for hapalindole core formation. Li et al. examined the BGC of *Fischerella ambigua* UTEX 1903 (which they elected to rename *fam*).[64] Their efforts revealed four novel cyclase-type enzymes (FamC1-C4) alongside previously annotated proteins. Upon more in depth exploration, they discovered that FamD2 (identical to AmbP1) catalyzes indole C-3 geranylation in a pH dependent manner. At relatively acidic pH, a 1-2 shift of the geranyl group at C-3 to the indole C-2 position results in formation of a terminal shunt metabolite. This result was later confirmed by Liu and co-workers upon examination of the magnesium dependence of AmbP1.[70] Li et al. also explored the function of FamD1 (identical to AmbP3). They showed that while FamD1 can catalyze prenylation of **1.1** with both GPP and DMAPP; its higher selectivity for DMAPP leads to formation of ambiguine H **1.52** from hapalindole U **1.4.** Analysis of a cell free extract from *F. ambigua* 1903 in the presence of **1.1** and GPP resulted in formation of a new metabolite 12-*epi*-hapalindole U **1.3.** Proteomic analysis of the soluble protein fraction revealed the identity of the thermostable cyclase FamC1. Cloning and expression of *famC1* devoid of the leader peptide provided ready access to soluble, functional recombinant protein, which in the presence of FamD2, **1.1** and GPP, catalyzed formation of **1.3**. These results led to a new hypothesis for the biosynthesis of hapalindole-type natural products. The process begins when **1.1** is geranylated at the C-3 position to give **1.2,** and in the presence of a Stig cyclase, undergoes a remarkable three-step reaction including 1) [3,3]-sigmatropic (Cope) rearrangement, 2) 6-*exo*-*trig* cyclization and 3) terminal electrophilic aromatic substitution (EAS) reaction to generate either the hapalindole or fischerindole core (Figure 1.11). Further late stage prenylation, cyclization and oxidation reactions are presumed to be responsible for the formation of the ambiguines and welwitindolinones.[64] This work revealed one of the first biosynthetic Cope rearrangements,[75] which represents the initial reaction in the three-step cascade catalyzed by FamC1 and related Stig cyclases.

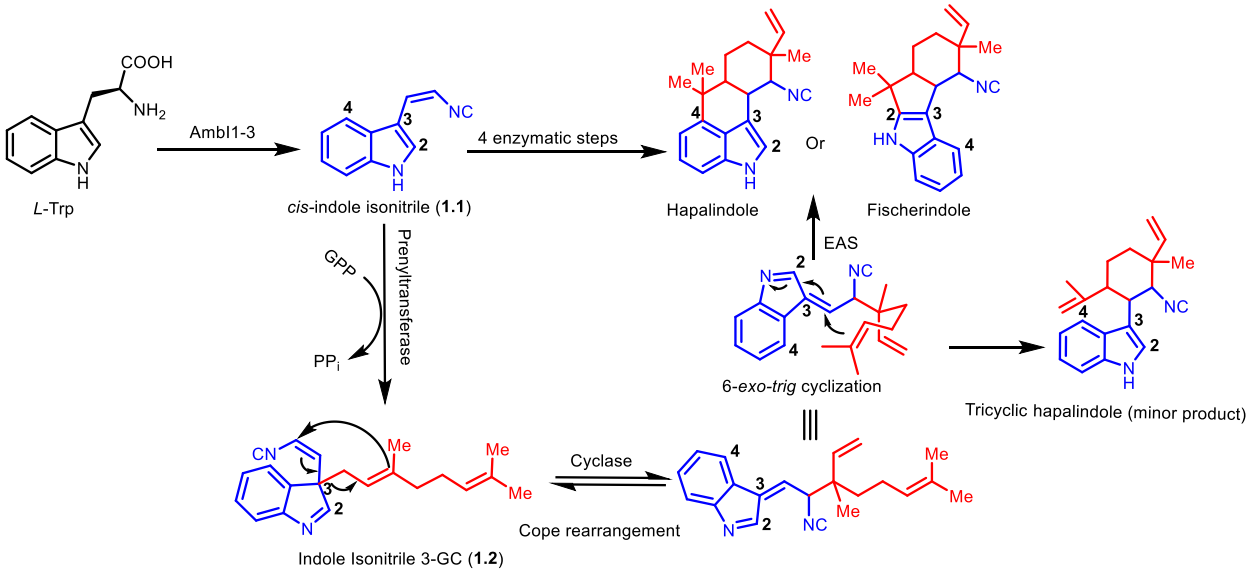


Figure 1.11: Biosynthetic mechanism for assembly of the hapalindole and fischerindole cores. Based on independent studies of Sherman[64] and Liu.[61] *Cis*-indole isonitrile core (**1.1**) is highlighted in blue. GPP is highlighted in red.

1.5.3 Further exploration of Stig cyclases

Upon the discovery of FamC1 as the central hapalindole core forming biocatalyst, the Sherman and Liu groups independently continued to explore the newly named Stig cyclases (named after the Stigonematales cyanobacterial strains they have been isolated from) further. In early 2017, Zhu and Liu reported the characterization of all WelU cyclases from *Hapalosiphon welwitschii* UTEX B1830 and IC-52-3.[65] They discovered that WelU1 assembled 12-*epi*-fischerindole U **1.36** and WelU3 assembled the tricyclic 12-*epi*-hapalindole C **1.26**. They also noted that the cyclases were more efficient at lower pHs, but made the reaction more challenging as acidic pHs induce the 1,2-shift of **1.2**.[65]

Concurrently and independently, the Sherman lab examined the *famC* gene cluster and discovered that a combination of different cyclases leads to variant stereochemical outcomes at the molecule's four chiral centers (C-10, C-11, C-12, C-15). A reaction mixture containing FamC2-FamC3 showed production of hapalindole H **1.5**.[66] Li suggested that these varying product profiles resulted from different interactions in these heterodimeric combinations, which was shown by protein pull down experiments supporting a direct association between FamC2 and FamC3.[66] Examination of close homologs HpiC1 (from *Fischerella* sp. ATCC43239) and FilC1 (from *Fischerella* sp. IL-199-3-1) showed production of **1.3**, as expected. Fischerindole production was explored in cyclases from *Fischerella muscicola* UTEX1829 (FimC5) and

Fischerella sp. SAG 46.79 (FisC). Both enzymes produced **1.36** even though they were about 90% identical in sequence to FamC1 (although FimC1-FimC4, from the same BGC as FimC5, was found to produce **1.4**).[66] With HpiC1 and FimC5, the tricyclic hapalindole **1.26** was observed as a minor product. This metabolite was not converted to the tetracyclic form when provided exogenously to the cyclases, supporting the hypothesis that the C-3 geranylated intermediate is the substrate for the cyclases and the tricyclic hapalindoles are shunt products, not intermediates, in the pathway as originally hypothesized by Moore. [66]

Later in 2017, Zhu and Liu published another follow up study examining the cyclases AmbU1-4.[67] Already knowing that AmbU4 (identical to FamC1) produces **1.3**, they elected to look at AmbU1-3. Through heterologous expression, all four proteins were incubated together, which lead to the production of **1.3**, **1.4**, and **1.6**, a result which was consistent with previous studies.[66] However, Zhu and Liu made an important observation by adjusting the stoichiometric ratios of select Stig cyclases, and were able to shift product profiles of AmbU2-AmbU3 (FamC2/C3) and AmbU1-AmbU4 (FamC1/C4). Upon addition of supplementary Ca^{2+} to AmbU1-AmbU4, they were able to shift the product profile exclusively to **1.4**. Likewise, the addition of Ca^{2+} to AmbU2-AmbU3 shifted the product profile to assemble **1.5** over **1.4**. They also noted that AmbU2 (FamC2) by itself could assemble **1.5**, but less efficiently compared to the combination. To confirm the Ca^{2+} dependence, the authors dialyzed the proteins (including WelU1) and showed that EDTA abolished activity, which was restored upon addition of Ca^{2+}.[67]

After this initial series of publications that revealed a fundamentally new type of biosynthetic cascade for indole alkaloid assembly, attention was drawn to crystallographic and mechanistic insights for the Stig cyclases (see below). However, in early 2020, the Sherman lab published a new study on this unique class of enzymes. Li et al. decided to further explore cyclases from the *fil* gene cluster concurrently with a newly annotated gene cluster from *Westiellopsis prolifica* SAG 16.93 (*wep*).[68] FilC1-FilC3 was shown to catalyze formation of **1.3**, with **1.4** as a minor product. FilC1-FilC4 heterodimer led to production of **1.4**, and FilC2-FilC3 catalyzed assembly of a new metabolite, 12-*epi*-hapalindole H **1.6**. FilC2 could also generate **1.6** as a homodimer, with lower efficiency than the heterodimeric combination. FilC2-FilC3 was also reported to assemble **1.26** and 12-*epi*-hapalindole Q **1.29** as minor products. This broader product profile from FilC2-FilC3 was shown to be cofactor dependent as the addition of supplementary Ca^{2+} shifted the product profile to produce **1.6** as the major product. As described

in Liu's study[67], all Stig cyclase enzymatic activity was abolished by dialyzing the proteins with EDTA. Further analysis of the Wep pathway derived cyclases showed similar results. WepC1 produced **1.3** and heterodimeric combination WepC1-WepC2 produced three metabolites including, **1.4** and hapalindole C **1.24** as major products alongside **1.5** as a minor product. Li et al. were able to shift product profiles by incubating the Stig cyclase homologues from different gene clusters together. Thus, FamC2-FilC3 showed production of **1.6** unlike FamC2-FamC3, which produces **1.5**. Mutational analysis of active site amino acids along the presumed dimeric interface (identified by Newmister et al.[76]) (Figure 1.12) led to alterations in the product profiles of specific hetero-oligomeric cyclase combinations. These results suggest that key residues in a composite active site play a role in affecting the stereochemical outcome of the cyclization cascade.[68] Further studies on heterodimeric and oligomeric combinations of Stig cyclases are expected from studies on the growing list of annotated Stigonematales cyanobacterial genome derived BGCs (Figure 1.13). Currently characterized Stig cyclases and their respective products are shown in Figure 1.14.

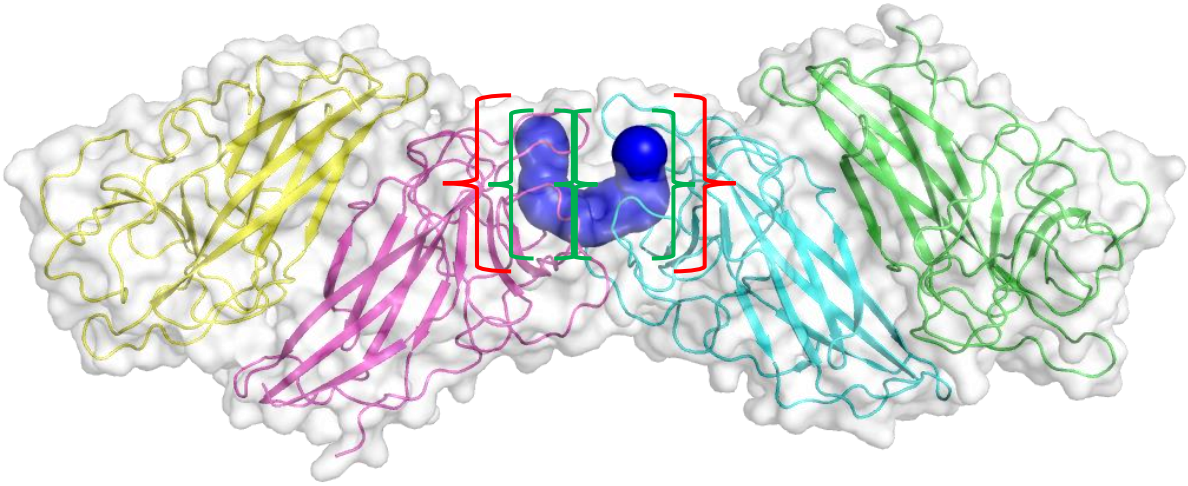


Figure 1.12: Crystal structure of dimeric interface of HpiC1. Reported in Newmister et al.[76] and supported by functional data in Li et al.[68] Dimeric interface is highlighted by the blue cavity. Red brackets encompass the combined active site while green brackets encompass the active site from the individual Stig cyclase monomers. Blue cavity calculated using CAVER software.[87]

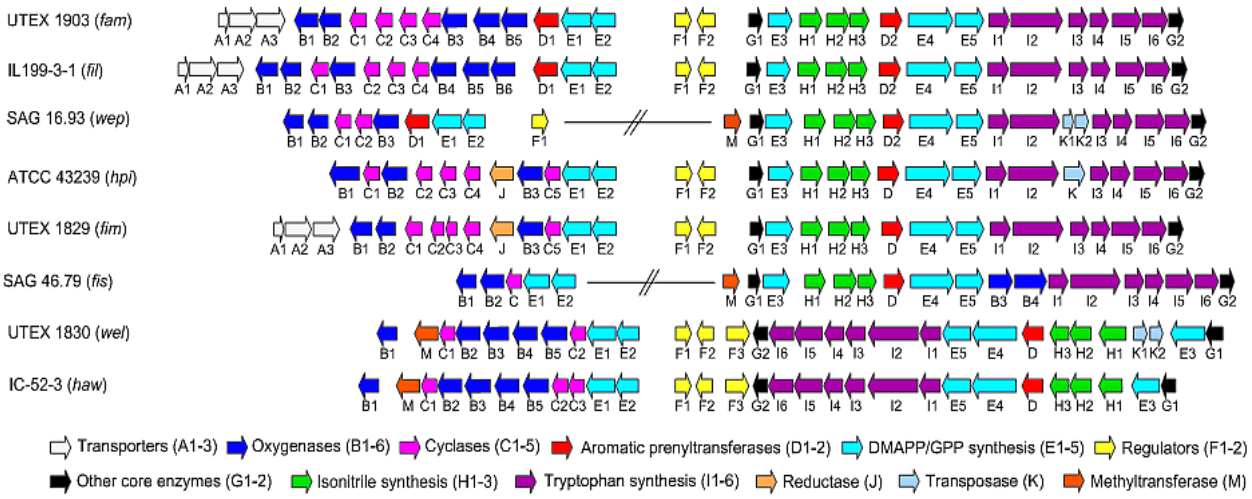


Figure 1.13: Currently annotated hapalindole/fischerindole encoding BGCs. All from Stigonematales cyanobacteria.

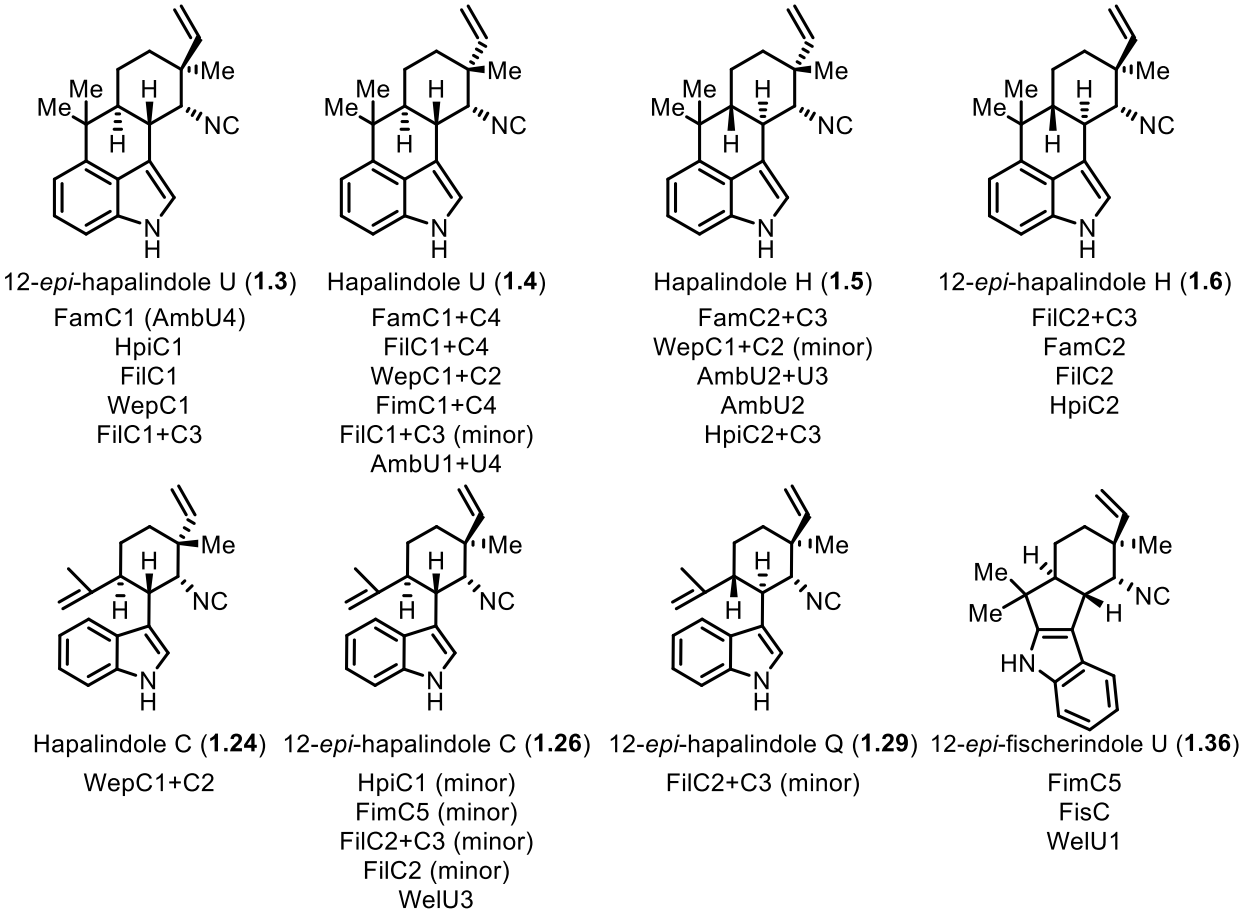


Figure 1.14: Currently characterized Stig cyclases and the compounds derived from their biocatalytic reactions.[64-68]

1.5.4 Stig cyclase and prenyltransferase protein structure and mechanistic insights

Following the initial functional characterization of various Stig cyclases in 2017, there was a shift towards X-ray crystal structure studies to identify the active site and elucidate the mechanistic basis of the three-step cyclization cascade. In 2018, Newmister et al. reported the first high resolution structure of a Stig cyclase,[76] and subsequent mutational analysis of HpiC1 revealed some remarkable insights. The HpiC1 hapalindole-forming cyclase exists in a homodimeric state with four Ca^{2+} coordination sites, which confirmed and provide molecular details behind the metal cofactor dependence noted in previous studies.[66,67] Analysis of a likely active site region identified by sequence comparisons uncovered numerous aromatic residues, and the presence of a conserved aspartic acid residue suggested that the cyclization cascade may be acid catalyzed (Figure 1.15). Mutating the active site aspartic acid residue to an alanine abrogated Stig cyclase (HpiC1) activity, and confirmed its importance in the cyclization cascade. Comparing the proposed active site of HpiC1 to FimC5 (catalyzes fischerindole formation)

facilitated the identification of key amino acid residues responsible for selectivity of the indole C-4 versus C-2 position in the terminal EAS step. Thus, Y101 and F138 were mutated to their corresponding residues in FimC5 (F101 and S138) to interrogate their impact on hapalindole vs fischerindole production. The Y101F mutant did not show a shift in product profile; however the F138S mutant catalyzed production of both **1.36** and **1.3** (1:2 ratio), suggesting that these active site amino acids play a more important role in selectivity of the terminal cyclization step. This profile was shifted even further with the double mutant (Y101F/F138S) form of HpiC1, which produced **1.36** as the major product in (2:1 ratio) over **1.3**.[76] Molecular dynamics (MD) and density-functional theory (DFT) calculations were performed to understand further the basis for hapalindole vs fischerindole formation. MD simulations showed that the aspartate residue retains its conformation toward the inner cavity of the active site while F138 acts as a wall on the side of the pocket. However in the F138S mutant, there is additional space and the active site becomes more flexible offering an explanation for the change in selectivity of the C-2 vs C-4 EAS. DFT calculations showed that in the proposed chair transition state for the Cope rearrangement, the aspartate residue can lower the energy barrier from 20.6 kcal/mol to 18.4 kcal/mol (a 40-fold rate enhancement) suggesting that the aspartate residue catalyzes a proton transfer with the indolenine nitrogen, inducing the acid catalyzed Cope rearrangement. This was further confirmed from computations showing that the 6-*exo-trig* cyclization can only occur when the indole nitrogen is protonated, while suppressing reaction at the C-3 and C-16 positions to form a cyclobutane ring.[76] Lastly, the intrinsic energy difference that favors the terminal EAS occurring at the C-2 position versus the C-4 position is about 8 kcal/mol (fischerindole production). This suggests that HpiC1 and likely all hapalindole producing cyclases overcome the intrinsic energy barrier by controlling the terminal EAS at C-4 instead of the intrinsic lower energy at C-2.[76] A crystal structure was also obtained with HpiC1 homodimers bearing the dimeric interface comprising the individual active sites. This interface was functionally confirmed by Li et al. through mutagenesis studies that confirmed the dependence of both Asp residues at the interface of a dimeric Stig cyclase (Figure 1.12).[68] These remarkable findings gave the first in-picture view of how the Stig cyclase class of enzymes catalyze a complex cyclization cascade.

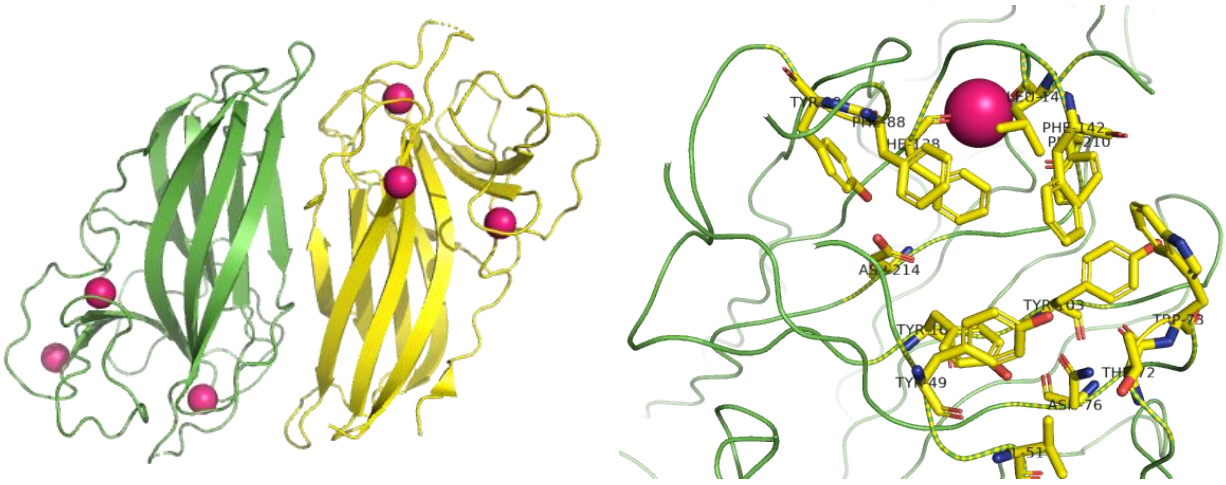


Figure 1.15: Crystal structure and active site of HpiC1. **Left**: Crystal structure of HpiC1 obtained by Newmister et al.[76] Cyclase exists in dimeric state with each individual subunit colored. Ca^{2+} ions highlighted in magenta. **Right**: Structure of active site with important residues highlighted in yellow and labeled. Ca^{2+} ion highlighted in magenta.

Chen et al. followed this study by publishing X-ray crystal structures of FamC1, FilC1 and FisC.[77] The majority of their work confirmed insights gleaned from HpiC1, while also revealing that the cyclases exhibit significant conformational changes around the active site. They also showed that hydrophobic residues in FamC1 and FilC1 push the terminal EAS towards the C-4 position while the polar residues found in FisC and FimC5 allow for the energetically favored cyclization at the C-2 position.[77]

Other groups elected to explore the prenyltransferases and reported structural studies of both FamD1 (AmbP3) and FamD2 (AmbP1). Wang et al. disclosed the first X-ray crystal structure of FamD1,[78] which is similar to all other ABBA prenyltransferases, in that the protein contains a central β-barrel of ten anti-parallel strands surrounded by ten α-helices. Obtaining a crystal structure with DMAPP bound to the enzyme, the authors observed the pyrophosphate group interacting with a group of basic residues on the top of the active site in the center of the β-barrel. They also uncovered two tyrosine residues that may help facilitate pyrophosphate release. Furthermore, a crystal structure with both **1.1** and DMAPP bound was obtained. They observed **1.1** interacting with a group of hydrophobic residues in the lower half of the active site, opposite of DMAPP. From this complex structure, the authors proposed a mechanism wherein DMAPP is dephosphorylated and reacts via an SN_1 type reaction with **1.1**. Following proton abstraction, the formation of the C-3 and C-2 prenylated products originally identified by Li et al. are produced.[64]

Soon after publication of the Wang study, the Liu and Abe groups collaborated to determine the first crystal structure of FamD2 (AmbP1),[79] which was shown to contain the same ABBA prenyltransferase structure as FamD1 (ten anti-parallel β-barrel strands surrounded by α-helices). They uncovered a unique Mg^{2+} binding site that was not observed in FamD1. However, this proved to not be the Mg^{2+} responsible for FamD2's selectivity as another Mg^{2+} binding site was uncovered in the ternary complex of the protein, **1.1**, and GSPP (a non-hydrolyzable GPP analog) at pH 8. As in the case of FamD1, the top half of the active site is rich in polar residues while the bottom half is replete in hydrophobic residues, which helped orient GSPP. Interestingly, in structures without Mg^{2+} bound, **1.1** turned by nearly 100 degrees showing the importance of Mg^{2+} for catalysis. The presence of Mg^{2+} actually shifts the C-2 of **1.1** 2Å further away from C-1 of GPP helping to favor geranylation at the C-3 position in the predicted[65,67,76] (*R*) configuration.[79] This was the first structural work on a prenyltransferase where a metal co-factor was shown to play a role in enzymatic selectivity.

These two groups also produced a structure of FamD1 (AmbP3) bound to both **1.4** and hapalindole A **1.12**.[80] The authors observed the expected reverse prenylation at the C-2 position with **1.4** but observed normal prenylation at the C-2 position with **1.12**. Upon obtaining separate crystal structures with **1.4** and **1.12** in complex with dimethylallyl S-thiolodiphosphate (DMSPP), they observed DMSPP bound to the polar region, near the top of the β-barrel cavity. Compounds **1.4** and **1.12** were bound to the more hydrophobic region of the pocket and **1.4** formed a hydrogen bond with Glu207; with **1.12** this key hydrogen bond was not observed. Both **1.4** and **1.12** were found to have completely altered orientations suggesting that the terpenoid component of the molecule is more important for substrate recognition by the prenyltransferase. For compound **1.4**, the distance from indole C-2 to the C-3 of DMSPP was shorter than with compound **1.12**, supporting the original observation of reverse versus normal prenylation (Figures 1.16 & 1.17). [80]Along with the other studies discussed in this section, we now have an in-depth picture regarding how the Stig cyclases and allied prenyltransferases catalyze formation of hapalindole-type metabolites.

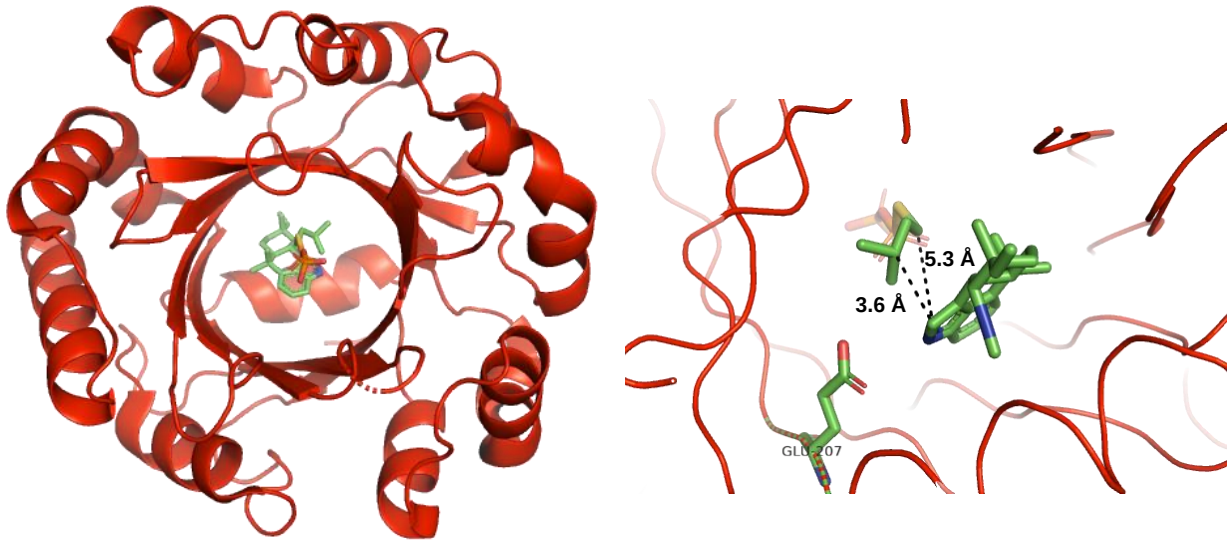


Figure 1.16: Crystal structure and active site of FamD1 with hapalindole U **1.4**. **Left**: Crystal structure of FamD1 with **1.4** obtained by Wong et al.[80] Looking down the β-barrel **1.4** and DMSPP are visible. **Right**: Structure of FamD1 active site. Glu207 is highlighted forming key hydrogen bond with **1.4.** This hydrogen bond places **1.4** into a configuration favoring reverse prenylation over normal prenylation at indole C-2. Distances between carbons are also displayed.

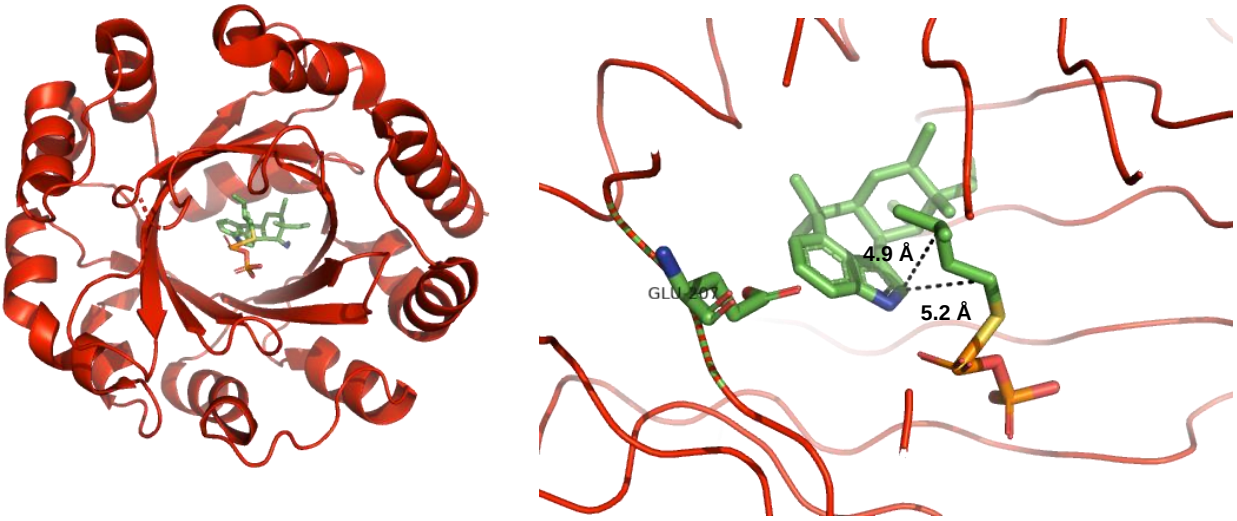


Figure 1.17: Crystal structure and active site of FamD1 with hapalindole A **1.12** **Left**: Crystal structure of FamD1 with **1.12** obtained by Wong et al.[80] Looking down the β-barrel **1.12** and DMSPP are visible. **Right**: Structure of FamD1 active site. Glu207 is highlighted without forming a key hydrogen bond with **1.12.** The loss of this hydrogen bond places **1.12** into a configuration favoring normal prenylation over reverse prenylation at indole C-2. Distances between carbons are also

1.5.5 Further analysis of WelO5 halogenase and homologs

As noted previously, Hillwig and Liu were able to characterize WelO5 as a Fe(II)/α-ketoglutarate dependent halogenase responsible for chlorination of **1.36** and **1.26**.[69] Because halogens are important bioisosteres and late-stage halogenation continues to be a challenge for medicinal chemists, halogenase enzymes have attracted much interest to solve these

problems.[81,82] Due to the higher selectivity of WelO5 compared to its counterpart AmbO5, Liu et al. continued their work by showing that WelO5 could stereoselectively halogenate **1.36** with both chlorine and bromine, although the enzyme showed a 10:1 preference for Cl^- over Br^-.[83]

With these results in hand, Liu (in collaboration with Boal) obtained a crystal structure of WelO5.[84] Like most Fe(II)/α-ketoglutarate dependent proteins, it contained an eight-stranded β-sandwich topology; however, it contained a glycine where normally a glutamate or aspartate would be found, thus removing a carboxylate moiety that enables direct coordination between the substrate and chloride ion. The active site is largely hydrophobic but key hydrogen bonds are made between the indole N-H and the isonitrile functionality. This leaves C-13 of **1.36** a short distance from the Fe(II) center but farther from the chloride ion, thus indicating that hydroxylation would occur instead of chlorination (Figure 1.18). From this crystal structure, site directed mutagenesis was performed to convert the noted glycine residue into an aspartate. This led to the formation of C-13 hydroxylated **1.36**. With these results in hand, the investigators hypothesized that **1.36** would have to shift positions to allow for selective chlorination. Closer observation of the active site revealed a key serine residue (Ser189) that could form a hydrogen bond with the iron-oxo intermediates, thus allowing **1.36** to shift positions and suppress hydroxylation. Mutagenesis of Ser189 to an alanine led to 1:1 production of chlorinated and hydroxylated metabolites.[84] Building off this work, Zhang and co-workers used quantum mechanical (QM) and molecular mechanical (MM) calculations to confirm both the mechanism for chlorination and the importance of the serine residue.[85] This crystallographic study and further follow up analysis revealed the unique features of WelO5 as a halogenating biocatalyst and how various residues work together to abrogate hydroxylation.

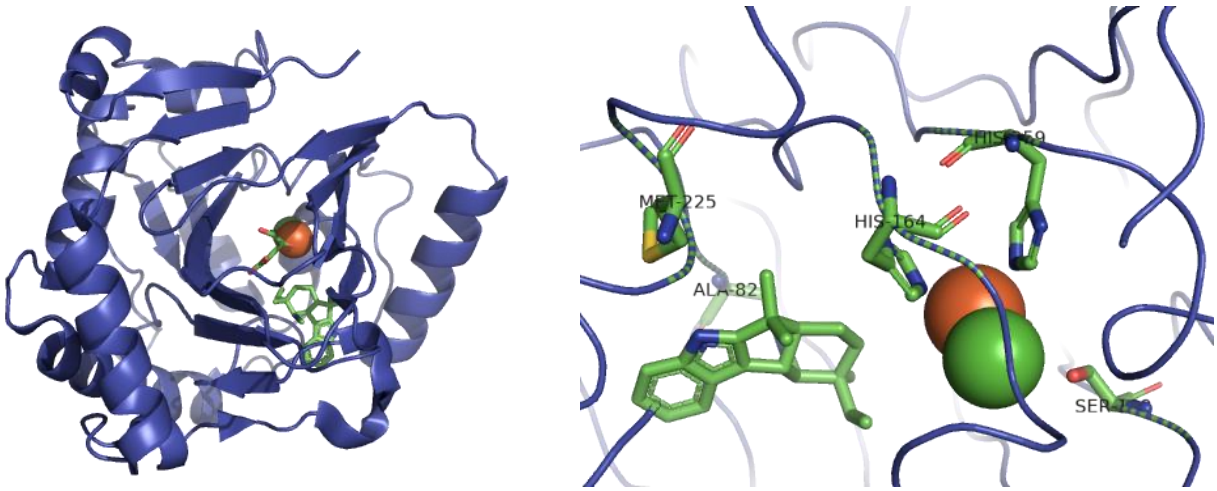


Figure 1.18: Crystal structure and active site of WelO5 with 12-*epi*-fisherindole U **1.36**. **Left**: Crystal structure of WelO5 with **1.36** obtained by Mitchell et al.[84] **Right**: Structure of active site with key residues highlighted**.** His164 and His259 coordinate to Fe(II) (orange sphere). Ala82 and Met225 hydrogen bond to **1.36** for proper configuration. Ser189 coordinates the chloride ion (green sphere).

Continuing their work, Liu and Zhu characterized the homologous halogenase from *Hapalosiphon welwitschii* IC-52-3 (also known as the *haw* BGC), which they referred to as WelO5*.[86] WelO5* was found to be more selective for **1.26** while still showing some promiscuity for **1.36**. A comparative analysis of the two halogenases found an 11 amino acid residue difference at the C-terminus. This C-terminus difference is believed to account for selectivity as a similar sequence was observed in AmbO5.[86]

1.6 Conclusion and thesis specific aims

Cyanobacteria generate an amazing array of structurally complex natural products, which include the hapalindole-type metabolites. With their unique polycyclic ring system, functional groups and stereochemical differences, these indole alkaloids continue to attract attention. The diverse biological activities of these metabolites enhance their importance as we continue to search for novel small molecules to combat infectious agents, cancer and other human diseases. While there have been numerous total syntheses, we continue to search for new approaches to explore and diversify these natural products for on-going drug development efforts. Because nature has developed elegant ways to synthesize natural products, numerous groups have taken the challenge to dissect the biosynthetic construction and tailoring of these compounds. This work has uncovered a collection of fascinating enzymes that catalyze distinctive reactions. Isonitrile synthases, aromatic prenyltransferases, the newly discovered Stig cyclases, carrier protein independent halogenases and various Rieske oxygenases all work in unison to produce these remarkable indole alkaloids.

Currently, we have only scratched the surface of the potential versatility these unique biocatalysts provide. While the hapalindole-type metabolites possess a wide variety of biological activities, total synthesis and fermentation efforts do not provide the amount of material, nor the ability to diversify, that is needed for true drug development efforts. This thesis contains two specific aims that seek to address this issue using chemoenzymatic methods.

The first aim looks at examining the Stig cyclases and aromatic prenyltransferases to produce unnatural hapalindole and fischerindole metabolites. This was undertaken in two different studies. The first looks at using microscale *in vitro* transcription/translation systems to

produce both native and unnatural derivatives without the need of heterologous host for protein expression and purification (**Chapter 2**). The second method looks at larger scale reactions with purified protein designed to both isolate and test derivatives. These studies also allow for further in depth study of the enzymes and what energetic barriers they can, and cannot, overcome (**Chapter 3**).

The second aim of the thesis addresses questions related to the pentacyclic ambiguines. The fifth (E) ring of these fascinating compounds has provided a synthetic challenge to overcome. To date, it is still unknown which enzymes catalyze the formation of this ring *in vivo*. This question was tackled using two different methods. The first was the elucidation of proposed Rieske-type proteins believed to be responsible for formation of the E-ring (**Chapter 4**). The second is looking at a chemoenzymatic route to its production. This involves early stage enzymatic construction of the tetracyclic core followed by C-H functionalization chemistry to close the E-ring. During the course of these studies, a novel shunt metabolite was discovered giving further validation to the proposed biosynthesis of the hapalindole and fischerindole core involving a Cope rearrangement (**Chapter 5**).

1.7 References

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*; Knölker, H.-J., Ed.; Chemistry and Biology; Academic Press, 2014; Vol. 73, pp 65–160. https://doi.org/10.1016/B978-0-12-411565-1.00002-0.
2. Singh, R. K.; Tiwari, S. P.; Rai, A. K.; Mohapatra, T. M. Cyanobacteria: An Emerging Source for Drug Discovery. *The Journal of Antibiotics* **2011**, *64* (6), 401–412. https://doi.org/10.1038/ja.2011.21.
3. Polakis, P. Antibody Drug Conjugates for Cancer Therapy. *Pharmacol Rev* **2016**, *68* (1), 3–19. https://doi.org/10.1124/pr.114.009373.
4. 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. https://doi.org/10.1021/ja00333a079.
5. 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.; Swartzendruber, J. K.; Deeter, J. B. Hapalindoles, Antibacterial and Antimycotic Alkaloids from the Cyanophyte *Hapalosiphon Fontinalis*. *J. Org. Chem.* **1987**, *52* (6), 1036–1043. https://doi.org/10.1021/jo00382a012.
6. 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. https://doi.org/10.1007/s10811-005-9011-9.
7. Raveh, A.; Carmeli, S. Antimicrobial Ambiguines from the Cyanobacterium *Fischerella* Sp. Collected in Israel. *J. Nat. Prod.* **2007**, *70* (2), 196–201. https://doi.org/10.1021/np060495r.
8. Mo, S.; Krunic, A.; Chlipala, G.; Orjala, J. Antimicrobial Ambiguine Isonitriles from the Cyanobacterium *Fischerella Ambigua*. *J. Nat. Prod.* **2009**, *72* (5), 894–899. https://doi.org/10.1021/np800751j.
9. 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. https://doi.org/10.1111/j.1574-6968.2001.tb10554.x.
10. 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. https://doi.org/10.1021/jo00029a014.
11. 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. https://doi.org/10.1021/ja00101a015.
12. Becher, P. G.; Keller, S.; Jung, G.; Süssmuth, R. D.; Jüttner, F. Insecticidal Activity of 12-*Epi*-Hapalindole J Isonitrile. *Phytochemistry* **2007**, *68* (19), 2493–2497. https://doi.org/10.1016/j.phytochem.2007.06.024.

13. 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.
14. 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. https://doi.org/10.1016/j.bmc.2012.06.030.
15. Zhang, X.; Smith, C. D. Microtubule Effects of Welwistatin, a Cyanobacterial Indolinone That Circumvents Multiple Drug Resistance. *Mol. Pharmacol.* **1996**, *49* (2), 288–294.
16. Chilczuk, T.; Steinborn, C.; Breinlinger, S.; Zimmermann-Klemd, A. M.; Huber, R.; Enke, H.; Enke, D.; Niedermeyer, T. H. J.; Gründemann, C. Hapalindoles from the Cyanobacterium *Hapalosiphon* Sp. Inhibit T Cell Proliferation. *Planta Med.* **2019**. https://doi.org/10.1055/a-1045-5178.
17. *Handbook on Cyanobacteria: Biochemistry, Biotechnology and Applications*; Gault, P. M., Marler, H. J., Eds.; Bacteriology research developments series; Nova Science Publishers: New York, 2009.
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. https://doi.org/10.1016/S0040-4039(00)92061-6.
19. 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. https://doi.org/10.1016/j.tet.2012.02.048.
20. 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. https://doi.org/10.1016/j.phytochem.2010.09.004.
21. 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. https://doi.org/10.1021/jo00392a045.
22. 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. https://doi.org/10.1021/jo00226a009.
23. Moore, R. E.; Yang, X. G.; Patterson, G. M. L.; Bonjouklian, R.; Smitka, T. A. Hapalonamides and Other Oxidized Hapalindoles from *Hapalosiphon Fontinalis*. *Phytochemistry* **1989**, *28* (5), 1565–1567. https://doi.org/10.1016/S0031-9422(00)97798-7.
24. 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. https://doi.org/10.1021/tx500188a.
25. Bhat, V.; Allan, K. M.; Rawal, V. H. Total Synthesis of *N* -Methylwelwitindolinone D Isonitrile. *J. Am. Chem. Soc.* **2011**, *133* (15), 5798–5801. https://doi.org/10.1021/ja201834u.
26. 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. https://doi.org/10.1002/anie.201100957.

27. Connon, R.; Guiry, P. J. Recent Advances in the Development of One-Pot/Multistep Syntheses of 3,4-Annulated Indoles. *Tetrahedron Lett.* **2020**, 151696. https://doi.org/10.1016/j.tetlet.2020.151696.
28. Dethe, D. H.; Das, S.; Kumar, V. B.; Mir, N. A. Enantiospecific Total Syntheses of (+)-Hapalindole H and (-)-12-*Epi*-Hapalindole U. *Eur. J. Chem.* **2018**, *24* (36), 8980–8984. https://doi.org/10.1002/chem.201800970.
29. Gademann, K.; Bonazzi, S. Total Synthesis of Complex Cyanobacterial Alkaloids without Using Protecting Groups. *Angew. Chem. Int. Ed.* **2007**, *46* (30), 5656–5658. https://doi.org/10.1002/anie.200701881.
30. Kinsman, A. C.; Kerr, M. A. Total Synthesis of (±)-Hapalindole Q. *Org. Lett.* **2001**, *3* (20), 3189–3191. https://doi.org/10.1021/ol0165138.
31. 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. https://doi.org/10.1021/ja036191y.
32. 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. https://doi.org/10.1002/anie.201406626.
33. Maimone, T. J.; Ishihara, Y.; Baran, P. S. Scalable Total Syntheses of (−)-Hapalindole U and (+)-Ambiguine H. *Tetrahedron* **2015**, *71* (22), 3652–3665. https://doi.org/10.1016/j.tet.2014.11.010.
34. Rafferty, R. J.; Williams, R. M. Total Synthesis of Hapalindoles J and U. *J. Org. Chem.* **2012**, *77* (1), 519–524. https://doi.org/10.1021/jo202139k.
35. 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. https://doi.org/10.3987/COM-12-S(N)3.
36. 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. https://doi.org/10.1021/ja057640s.
37. 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. https://doi.org/10.1021/ja806981k.
38. Sahu, S.; Das, B.; Maji, M. S. Stereodivergent Total Synthesis of Hapalindoles, Fischerindoles, Hapalonamide H, and Ambiguine H Alkaloids by Developing a Biomimetic, Redox-Neutral, Cascade Prins-Type Cyclization. *Org. Lett.* **2018**, *20* (20), 6485–6489. https://doi.org/10.1021/acs.orglett.8b02804.
39. 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. https://doi.org/10.1021/ja5087672.
40. 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. https://doi.org/10.1002/anie.201307464.
41. 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. https://doi.org/10.1021/ja210837b.
42. Huters, A. D.; Styduhar, E. D.; Garg, N. K. Total Syntheses of the Elusive Welwitindolinones with Bicyclo[4.3.1] Cores. *Angew. Chem. Int. Ed.* **2012**, *51* (16), 3758–3765. https://doi.org/10.1002/anie.201107567.
43. MacKay, J. A.; Bishop, R. L.; Rawal, V. H. Rapid Synthesis of the *N*-Methylwelwitindolinone Skeleton. *Org. Lett.* **2005**, *7* (16), 3421–3424. https://doi.org/10.1021/ol051043t.
44. Bhat, V.; Rawal, V. H. Stereocontrolled Synthesis of 20,21-Dihydro *N*-Methylwelwitindolinone B Isothiocyanate. *Chem. Commun.* **2011**, *47* (34), 9705–9707. https://doi.org/10.1039/C1CC13498A.
45. Bhat, V.; MacKay, J. A.; Rawal, V. H. Lessons Learned While Traversing the Welwitindolinone Alkaloids Obstacle Course. *Tetrahedron* **2011**, *67* (52), 10097–10104. https://doi.org/10.1016/j.tet.2011.09.088.
46. Bhat, V.; MacKay, J. A.; Rawal, V. H. Directed Oxidative Cyclizations to C2- or C4-Positions of Indole: Efficient Construction of the Bicyclo[4.3.1]Decane Core of Welwitindolinones. *Org. Lett.* **2011**, *13* (12), 3214–3217. https://doi.org/10.1021/ol201122f.
47. 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. https://doi.org/10.1021/ja210793x.
48. Muratake, H.; Natsume, M. Total Synthesis of Marine Alkaloids (±)-Hapalindoles J and M. *Tetrahedron Lett.* **1989**, *30* (14), 1815–1818. https://doi.org/10.1016/S0040-4039(00)99587-X.
49. 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. https://doi.org/10.1016/S0040-4020(01)96006-5.
50. 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. https://doi.org/10.1016/S0040-4020(01)96007-7.
51. Vaillancourt, V.; Albizati, K. F. Synthesis and Absolute Configuration of (+)-Hapalindole Q. *J. Am. Chem. Soc.* **1993**, *115* (9), 3499–3502. https://doi.org/10.1021/ja00062a013.
52. Fukuyama, T.; Chen, X. Stereocontrolled Synthesis of (-)-Hapalindole G. *J. Am. Chem. Soc.* **1994**, *116* (7), 3125–3126. https://doi.org/10.1021/ja00086a053.
53. 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.
54. Johnson, R. E.; Ree, H.; Hartmann, M.; Lang, L.; Sawano, S.; Sarpong, R. Total Synthesis of Pentacyclic (−)-Ambiguine P Using Sequential Indole Functionalizations. *J. Am. Chem. Soc.* **2019**, *141* (6), 2233–2237. https://doi.org/10.1021/jacs.8b13388.
55. Xu, J.; Rawal, V. H. Total Synthesis of (−)-Ambiguine P. *J. Am. Chem. Soc.* **2019**, *141* (12), 4820–4823. https://doi.org/10.1021/jacs.9b01739.

56. Hu, L.; Rawal, V.H. Total Synthesis of the Chlorinated Pentacyclic Indole Alkaloid (+)-Ambiguine G. *J. Am. Chem. Soc.* **2021**, *143* (29), 10872-10875. https://doi.org/10.1021/jacs.1c05762.
57. 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. https://doi.org/10.1021/ja056171r.
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. https://doi.org/10.1038/nature05569.
59. 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. https://doi.org/10.1039/C6SC00686H.
60. Ready, J. M.; Reisman, S. E.; Hirata, M.; Weiss, M. M.; Tamaki, K.; Ovaska, T. V.; Wood, J. L. A Mild and Efficient Synthesis of Oxindoles: Progress Towards the Synthesis of Welwitindolinone A Isonitrile. *Angew. Chem. Int. Ed.* **2004**, *43* (10), 1270–1272. https://doi.org/10.1002/anie.200353282.
61. Hillwig, M. L.; Zhu, Q.; Liu, X. Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium *Fischerella Ambigua*. *ACS Chem. Biol.* **2014**, *9* (2), 372–377. https://doi.org/10.1021/cb400681n.
62. 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. https://doi.org/10.1002/cbic.201300794.
63. 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* (1), 213. https://doi.org/10.1186/s12866-014-0213-7.
64. 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. https://doi.org/10.1021/jacs.5b10136.
65. 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. https://doi.org/10.1039/C7CC00782E.
66. 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. https://doi.org/10.1038/nchembio.2327.
67. 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.* **2017**, *56* (31), 9062–9066. https://doi.org/10.1002/anie.201703932.
68. Li, S.; Newmister, S. A.; Lowell, A. N.; Zi, J.; Chappell, C. R.; Yu, F.; Hohlman, R. M.; Orjala, J.; Williams, R. M.; Sherman, D. H. Control of Stereoselectivity in Diverse Hapalindole Metabolites Is Mediated by Cofactor-Induced Combinatorial Pairing of Stig

Cyclases. *Angew. Chem. Int. Ed.* **2020**, *59* (21), 8166–8172. https://doi.org/10.1002/anie.201913686.

69. Hillwig, M. L.; Liu, X. A New Family of Iron-Dependent Halogenases Acts on Freestanding Substrates. *Nat. Chem. Biol.* **2014**, *10* (11), 921–923. https://doi.org/10.1038/nchembio.1625.
70. Liu, X.; Hillwig, M. L.; Koharudin, L. M. I.; Gronenborn, A. M. Unified Biogenesis of Ambiguine, Fischerindole, Hapalindole and Welwitindolinone: Identification of a Monogeranylated Indolenine as a Cryptic Common Biosynthetic Intermediate by an Unusual Magnesium-Dependent Aromatic Prenyltransferase. *Chem. Commun.* **2016**, *52* (8), 1737–1740. https://doi.org/10.1039/C5CC10060G.
71. Brady, S. F.; Clardy, J. Systematic Investigation of the *Escherichia Coli* Metabolome for the Biosynthetic Origin of an Isocyanide Carbon Atom. *Angew. Chem. Int. Ed.* **2005**, *44* (43), 7045–7048. https://doi.org/10.1002/anie.200501942.
72. Brady, S. F.; Clardy, J. Cloning and Heterologous Expression of Isocyanide Biosynthetic Genes from Environmental DNA. *Angew. Chem. Int. Ed.* **2005**, *44* (43), 7063–7065. https://doi.org/10.1002/anie.200501941.
73. Chang, W.; Sanyal, D.; Huang, J.-L.; Ittiamornkul, K.; Zhu, Q.; Liu, X. In Vitro Stepwise Reconstitution of Amino Acid Derived Vinyl Isocyanide Biosynthesis: Detection of an Elusive Intermediate. *Org. Lett.* **2017**, *19* (5), 1208–1211. https://doi.org/10.1021/acs.orglett.7b00258.
74. 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. https://doi.org/10.1002/anie.201601447.
75. Mahmoodi, N.; Qian, Q.; Luk, L.; Tanner, M. Rearrangements in the Mechanisms of the Indole Alkaloid Prenyltransferases. *Pure Appl. Chem.* **2013**, *85* (10), 1935–1948. https://doi.org/10.1351/pac-con-13-02-02.
76. 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 Cyclization in Hapalindole Biogenesis. *Nat. Chem. Biol.* **2018**, *14* (4), 345–351. https://doi.org/10.1038/s41589-018-0003-x.
77. Chen, C.-C.; Hu, X.; Tang, X.; Yang, Y.; Ko, T.-P.; Gao, J.; Zheng, Y.; Huang, J.-W.; Yu, Z.; Li, L.; Han, S.; Cai, N.; Zhang, Y.; Liu, W.; Guo, R.-T. The Crystal Structure of a Class of Cyclases That Catalyze the Cope Rearrangement. *Angew. Chem. Int. Ed.* **2018**, *57* (46), 15060–15064. https://doi.org/10.1002/anie.201808231.
78. Wang, J.; Chen, C.-C.; Yang, Y.; Liu, W.; Ko, T.-P.; Shang, N.; Hu, X.; Xie, Y.; Huang, J.-W.; Zhang, Y.; Guo, R.-T. Structural Insight into a Novel Indole Prenyltransferase in Hapalindole-Type Alkaloid Biosynthesis. *Biochem. Biophys. Res. Commun.* **2018**, *495* (2), 1782–1788. https://doi.org/10.1016/j.bbrc.2017.12.039.
79. Awakawa, T.; Mori, T.; Nakashima, Y.; Zhai, R.; Wong, C. P.; Hillwig, M. L.; Liu, X.; Abe, I. Molecular Insight into the Mg2+-Dependent Allosteric Control of Indole Prenylation by Aromatic Prenyltransferase AmbP1. *Angew. Chem. Int. Ed.* **2018**, *57* (23), 6810–6813. https://doi.org/10.1002/anie.201800855.
80. Wong, C. P.; Awakawa, T.; Nakashima, Y.; Mori, T.; Zhu, Q.; Liu, X.; Abe, I. Two Distinct Substrate Binding Modes for the Normal and Reverse Prenylation of

Hapalindoles by the Prenyltransferase AmbP3. *Angew. Chem. Int. Ed.* **2018**, *57* (2), 560–563. https://doi.org/10.1002/anie.201710682.

81. Hernandes, M. Z.; Cavalcanti, S. M. T.; Moreira, D. R. M.; de Azevedo Junior, W. F.; Leite, A. C. L. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **2010**, *11* (3), 303–314. https://doi.org/10.2174/138945010790711996.
82. Fraley, A. E.; Sherman, D. H. Halogenase Engineering and Its Utility in Medicinal Chemistry. *Bioorg. Med. Chem. Lett.* **2018**, *28* (11), 1992–1999. https://doi.org/10.1016/j.bmcl.2018.04.066.
83. 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. https://doi.org/10.1002/cbic.201500674.
84. 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. https://doi.org/10.1038/nchembio.2112.
85. Zhang, X.; Wang, Z.; Gao, J.; Liu, W. Chlorination versus Hydroxylation Selectivity Mediated by the Non-Heme Iron Halogenase WelO5. *Phys. Chem. Chem. Phys.* **2020**, *22* (16), 8699–8712. https://doi.org/10.1039/D0CP00791A.
86. Zhu, Q.; Liu, X. Characterization of Non-Heme Iron Aliphatic Halogenase WelO5* from Hapalosiphon Welwitschii IC-52-3: Identification of a Minimal Protein Sequence Motif That Confers Enzymatic Chlorination Specificity in the Biosynthesis of Welwitindolelinones. *Beilstein J. Org. Chem.* **2017**, *13* (1), 1168–1173. https://doi.org/10.3762/bjoc.13.115.
87. Chovancová,E., Pavelka, A., Beneš, P., Strnad, O., Brezovský, J., Kozlíková, B., Gora, A., Šustr, V., Klvaňa, M., Medek, P., Biedermannová, L., Sochor, J., Damborský, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. PLoS Computational Biology **2012**, *8* (10), e1002708. https:// 10.1371/journal.pcbi.1002708.

Note:

Majority of this introduction has been published as “Recent advances in hapalindole-type cyanobacterial alkaloids: biosynthesis, synthesis, and biological activity.” Robert M. Hohlman and David H. Sherman. *Nat. Prod. Rep.* **2021**, *38*, 1567-1588.

Chapter 2

Multi-component microscale biosynthesis of unnatural cyanobacterial indole alkaloids

2.1 Abstract

Genome sequencing and bioinformatics tools have facilitated the identification and expression of an increasing number of cryptic biosynthetic gene clusters (BGCs). However, functional analysis of all components of a metabolic pathway to precisely determine biocatalytic properties remains time-consuming and labor intensive. One way to speed this process involves microscale cell-free protein synthesis (CFPS) for direct gene to biochemical function analysis, which has rarely been applied to study multi-component enzymatic systems in specialized metabolism. In this chapter and work, we established an *in vitro* transcription/translation (TT)-assay to assess the assembly of both native hapalindole-type natural products and unnatural derivatives.

2.2 Introduction

Natural products (NPs) continue to be invaluable molecules for developing front-line drugs against infectious diseases, parasites and cancers.[1,2] Current advances in bioinformatics tools for mining genomic and metagenomic data have facilitated our ability to rapidly decode biosynthetic pathways to identify and exploit novel structural diversity. Heterologous expression and refactoring of NP biosynthetic gene clusters (BGCs) is also unlocking cryptic pathways and expanding our ability to discover novel molecules of pharmaceutical interest.[3-8] It is also evident that evolutionary processes give rise to subtle DNA sequence variations[9] especially in the complex multi-component enzymatic systems including bacterial polyketide synthases (PKSs), non-ribosomal polyketide synthases (NRPSs),[9] the ribosomally synthesized and post-translationally modified peptides (RiPPs)[10,11] and cyanobacterial indole alkaloid systems[12-17] which lead to diverse product profiles. A remaining challenge is to rapidly predict from genome

sequencing data the likely change in catalytic function and metabolite profile alterations mediated by closely related biosynthetic enzymes. In addition, defining the functional consequences continues to require rigorous, time-consuming biochemical experiments.

It is often impractical to express and purify all components of a biosynthetic pathway and corresponding families of related proteins to decipher biochemical and functional properties. Thus, cell-free protein synthesis (CFPS) has emerged[18,19] with direct synthetic biology applications including production of metalloenzymes,[20-22] biosynthesis of mevalonate[23] and limonine,[24] assembly of bioplastics,[25] NP biosynthesis from NRPS,[26] characterizing and prototyping genetic networks,[27,28] and most recently for the study of modular construction of protein glycosylation pathways.[29] In the current study, we decided to explore Stig cyclases using microscale CFPS as 1) an alternative approach for screening hapalindole-related biosynthetic enzyme function, and 2) as a means to rapidly assess production of unnatural hapalindoles using synthetic unnatural substrates.

In this chapter and work, we sought to expand the application of CFPS for cyanobacterial NPs representing a multi-complex enzyme system (Figure 2.1). Two cyanobacterial derived indole alkaloid BGCs were selected, including *Fischerella ambigua* UTEX 1903[30] (hapalindole producer) and *Fischerella* sp. SAG 46.79[14] (fischerindole producer). As a first step, we demonstrated that CFPS is able to express plasmid-borne FamD2 to synthesize 3GC intermediate **2.3** using native substrates **2.1** and **2.2**. **2.3** is subsequently cyclized in a three-step reaction by the combination of strain specific Stig cyclases FamC1-C4 and FisC to afford the corresponding fused ring system either at the 3,4- (hapalindole) or 2,3-position (fischerindole) of

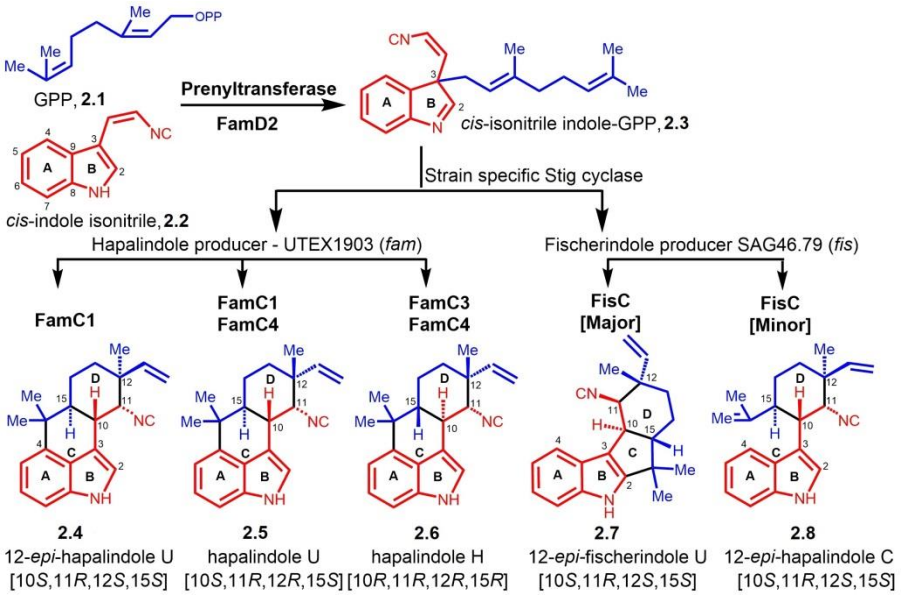


Figure 2.1: Biosynthesis of hapalindole and fischerindole metabolites.

the indole ring, respectively (Figure 2.1). This data was supported by proteomics identification of the CFPS derived prenyltransferase (FamD2) and cyclase (FamC/FisC) enzymes. The *in vitro* transcription-translation assay (TT-assay) was further employed to screen a series of unnatural halogenated *cis*-indole isonitrile derivatives to generate new hapalindole and fischerindole metabolites using a second plasmid bearing FamC1 or FisC, respectively. The new indole alkaloid molecules were confirmed by MS and NMR following direct chemoenzymatic scale-up (Figure 2.2). This approach represents the first example of *in vitro* TT applied to a multi-component pathway to generate a fully elaborated NP. With the rapidly growing natural product BGC database, we envision multiple applications for rapid functional analysis of biosynthetic enzymes, and substrate profiling to create and assess natural product structural diversity.

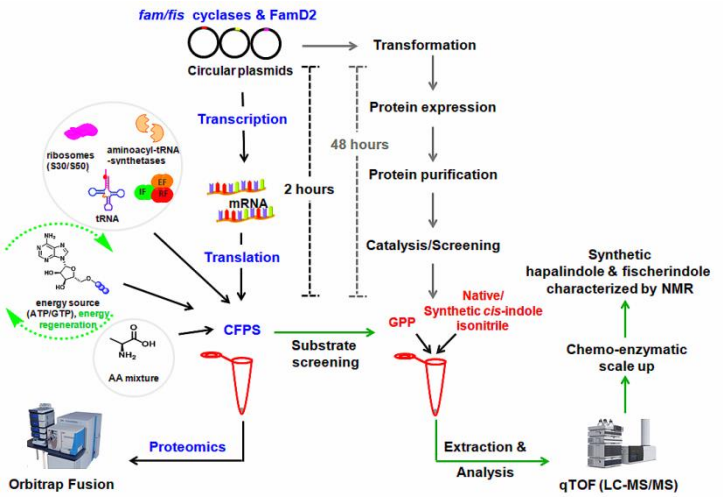


Figure 2.2: Overview of cell-free protein synthesis (CFPS). Using exogenous plasmids for the synthesis of prenyltransferase and Stig cyclases, identification of expressed protein by proteomics analysis, and detection and analysis of secondary metabolites. The timeline of CFPS is compared with the traditional protein expression, purification and catalysis.

2.3 Results and discussion

2.3.1 Development of an *in vitro* TT-assay for the biosynthesis of native hapalindole metabolites

To establish an *in vitro* transcription-translation platform for hapalindole biosynthesis in *Fischerella ambigua* UTEX 1903, individual plasmids containing genes *famD2, famC1, famC2, famC3* and *famC4* were cloned into the pET28a or pET28H8T systems as described.[14,30] Geranylated *cis*-indole isonitrile **2.3** is the key intermediate for subsequent formation of the tetracyclic core ring system of the hapalindoles.[30] Thus, our *in vitro* TT-assay was performed using a plasmid bearing the aromatic prenyltransferase FamD2 in the presence of **2.1** and **2.2**, with transcription and translation mediated through the T7 promoter (PureExpress NEB-I system). Proteomics analysis of the *in vitro* TT-assay showed 25 unique FamD2 peptides

covering 72.49% of the total amino acids sequence, providing strong evidence for efficient *in vitro* expression of FamD2 (Supplemental information Table S2.1). LC-MS analysis of the extracted reaction products ($[M+H]^+$= 305) possessed identical retention time (t_R) and MS-pattern compared to authentic **2.3** (Figure 2.3 B). Although enzymatic prenylation of **2.2** with GPP showed the production of geranylated C-2 and C-3 isocyano-indoles under pH dependent conditions favoring **2.3** at basic pH;[30,32] LC-MS analysis of the TT-assay reaction showed no detectable formation of the C-2 form as the reaction was performed in the presence of Tris-buffer at pH 7.8.

Table 2.1: Proteomics analysis of FamD2 and FamC1 expression from CFPS system. Plasmids bearing individual *famD2* and *famC1* genes were analyzed for expression of expected protein. The sequence highlighted in yellow represents the coverage of the synthesized protein.

FamD2			**FamC1**		
Total peptide	20		Total peptide	11	
Unique peptide	20		Unique peptide	11	
Coverage (%)	67.96		Coverage (%)	62.11	
Annotated Peptide Sequence to cover FamD2 sequence	**Sequence Positions**	**PSM FamD2**	**Annotated Peptide Sequence to cover FisC sequence**	**Sequence Position**	**PSM FamC1**
[R].IRTDIINVAK.[T]	FamD2 [7-16]	10	[K].RTTWTSNNGVGYVGPGTQFYNQLAPEGR.[N]	FamC1 [70-97]	2
[R].TDIINVAK.[T]	FamD2 [9-16]	6	[R].TTWTSNNGVGYVGPGTQFYNQLAPEGR.[N]	FamC1 [70-97]	8
[K].TFGAEYSEK.[V]	FamD2 [17-25]	3	[R].NIGYIYLSQNPGSGVAGFEQILDATLEPDTK.[Y]	FamC1 [98-128]	35
[K].VLDEVFQVFGEQFADNSFMIR.[T]	FamD2 [26-46]	19	[K].YTLTVDVGNLAGTFK.[G]	FamC1 [129-143]	87
[R].TSNKQPDK.[L]	FamD2 [47-54]	1	[K].GLSFAGFPGYR.[V]	FamC1 [144-154]	19
[K].LGCYFR.[Y]	FamD2 [55-60]	3	[R].VELLAGDTVLAADHNNLFIK.[E]	FamC1 [155-174]	306
[R].YHEEDESQLGLAWDIAR.[K]	FamD2 [61-77]	4	[K].TSTVTYTSTAK.[D]	FamC1 [180-190]	11
[R].KSGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [78-112]	6	[K].DLHLGQK.[L]	FamC1 [191-197]	20
[K].SGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [79-112]	8	[R].LVNLLQDK.[F]	FamC1 [202-209]	2
[K].IWQSIK.[G]	FamD2 [118-123]	3	[K].FSGLDFDNVR.[L]	FamC1 [210-219]	15
[K].GVVPVQDAFK.[L]	FamD2 [124-133]	5	[R].LVNLLQDKFSGLDFDNVR.[L]	FamC1 [202-219]	2
[K].LSLPASVTTHSDFLK.[N]	FamD2 [134-148]	7			
[R].LCFYLPFLNR.[E]	FamD2 [224-233]	2			
[R].EAVPQNLLNPLLK.[K]	FamD2 [234-246]	1			
[R].EAVPQNLLNPLLKK.[Y]	FamD2 [234-247]	2			
[K].KYINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2 [247-273]	4			
[K].YINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2 [248-273]	2			
[K].VDVDYHGR.[T]	FamD2[279-286]	1			
[R].TVPLFMK.[V]	FamD2 [287-293]	1			
[K].VHSQPLPK.[A]	FamD2 [294-301]	2			
FamD2 synthesized by NEBI			**FamC1 synthesized by NEBI**		
MNDVNRIRTDIINVAKTFGAEYSEKVLDEVFQVFGEQFADNSFMIRTSNKQPDKLGCYFRYHEEDESQLGLAWDIARKSGLLSDQGRPVDQLIPEICETFPIMADGVDFDVKHGLAKIWQSIKGVVPVQDAFKLSLPASVTTHSDFLKNHHLDALYAFGIDYHHSSVNLYFDTYHPKHHTSEYYKNLLQDLQFQPPSDELLELLTNNGEIALTFNFASPRIERLCFYLPFLNREAVPQNLLNPLLKKYINEAPALVDNPGFILGWSFGPQGGKGTYTKVDVDYHGRTVPLFMKV			MKRKLIVAVVFLIFICLGINTPAHATSAVSIPINNAGFENPFMDVVDDYTIDTPPGWTTYDPNNLVPEKRTTWTSNNGVGYVGPGTQFYNQLAPEGRNIGYIYLSQNPGSGVAGFEQILDATLEPDTKYTLTVDVGNLAGTFKGLSFAGFPGYRVELLAGDTVLAADHNNLFIKEGEFKTSTVTYTSTAKDLHLGQKLGIRLVNLLQDKFSGLDFDNVRLTTEPTET		

HSQPLPKAADFALAQ , 309 residues	, 227 residues

As a next step to test the cyclization of **2.3** using Stig cyclases (FamC1-C4), we tested equal concentrations of FamC1 and FamD2 plasmids for the *in vitro* TT-assay incubated in presence of **2.1** and **2.2**. Proteomics analysis of the reaction revealed 20 and 11 unique peptides covering 67.96% and 62.11% of the total protein sequences for FamD2 and FamC1, respectively (Table 2.1). LC-MS analysis of the TT-assay demonstrated the formation of a metabolite with *m/z* of 305, and its t_R and MS-pattern were identical to the standard 12-*epi*-hapalindole U **2.4** (Figure 2.3 C)**.**

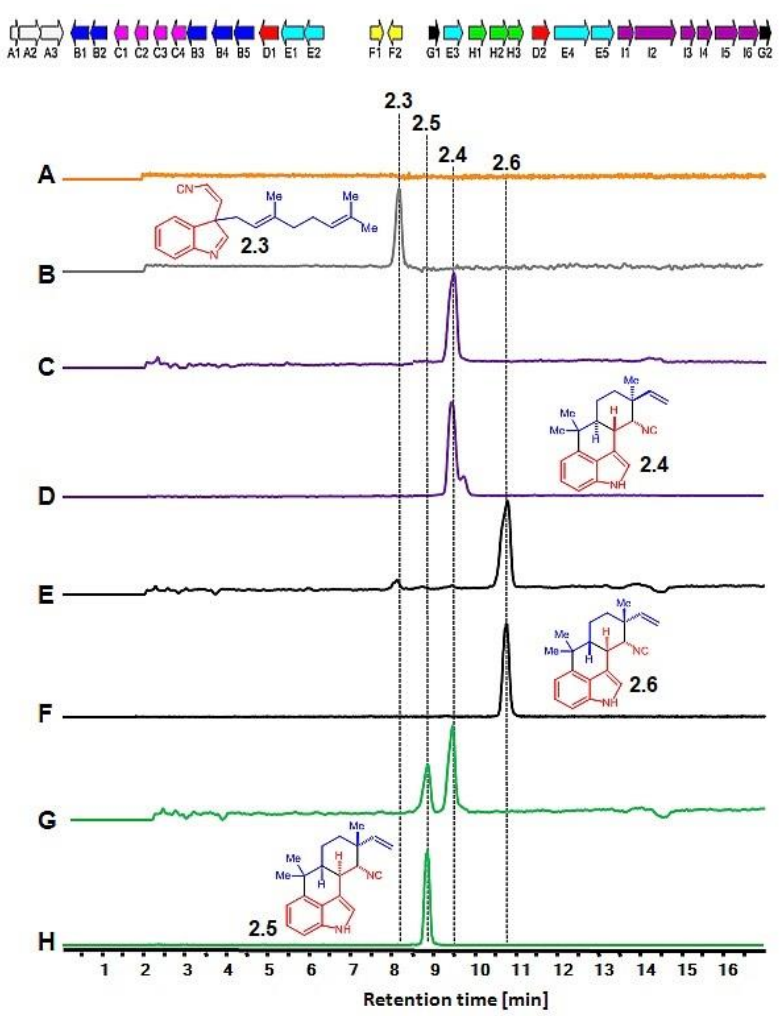


Figure 2.3: *In vitro* TT-assay and production of native hapalindoles. Top: Biosynthetic gene cluster of *Fischerella ambigua* UTEX 1903 (*fam*). The genes for transporters (white, A1-3), oxygenases (blue, B1-6), cyclases (pink, C1-5), aromatic prenyltransferases (red, D1-2), GPP synthases (cyan, E1-5), regulators (yellow, F1-2), *cis*-indole isonitrile synthase (green, H1-3) and other core enzymes (black, G1-2) are shown. Bottom: Extracted ion chromatograms (EICs) are shown in all panels. Detection of *in vitro* synthesized hapalindole products showing EIC (*m/z* of 305) with no plasmid (A), FamD2 alone (B) for **2.3**, FamD2 plus FamC1 (C) for **2.4,** FamD2 plus FamC2 and FamC3 (E) for **2.6**, and FamD2 plus FamC1 and FamC4 (G) for **2.5** and **2.4** production. The traces D, F, and H are the authentic standards of **2.4, 2.6** and **2.5**, respectively.

While performing CFPS of FamD2, FamC1 and FamC4 plasmids containing exogenous **2.1** and **2.2** in the presence of supplementary calcium (20 mM), the synthesized heterodimeric cyclases showed production of hapalindole U **2.5** and predominantly **2.4**, which were confirmed by LC-MS and correlated with authentic standards (Figure 2.3 G). The heterodimerization of FamC1 and FamC4 has been shown previously to produce hapalindole U **2.5**.[14,33] However, the CFPS reaction with *famC1*

and *famC4* plasmids under the same conditions showed the reverse product profile with **2.4** as the predominate product. We reasoned that the stoichiometry of FamC1 is relatively high compared to the calcium-dependent FamC4 resulting in increased production of **2.4** relative to **2.5** in the TT-assay. In order to study further the effect of alternative cyclases on stereocontrol of hapalindole molecules, we tested FamC2 and FamC3 combined with FamD2 using supplementary calcium (20 mM). This combination resulted in formation of hapalindole H **2.6** (Figure 2.3 E), which was confirmed by an authentic standard and corroborates our earlier findings using chemoenzymatic synthesis.[14] In this multi-plasmid format, we found that lower calcium ion resulted in failed hapalindole production, which further confirms the dependence of Stig cyclase heterodimerization and catalysis on this metal co-factor

2.3.2 Development of an *in vitro* TT-assay for the biosynthesis of native fischerindole metabolites

Following conclusive demonstration of hapalindole-type core ring system production by CFPS, we sought to employ the cyclase FisC from *Fischerella* sp. SAG 46.79 to generate the fischerindole core (Figure 2.1). Although the biosynthetic gene cluster of this strain contains only the one cyclase and a single prenyltransferase (FisD, a homolog of FamD2), it is known to produce 12-*epi*-fischerindole U **2.7**, having the same core stereochemistry (C10*S*, 11*R*, 12*S*, 15*S*) of **2.4**, bearing the indole C-2/3 ring connection rather than C-3/4.[14,31] In this case, the *in vitro* TT-assay was designed using two plasmids encoding individually FamD2 and FisC, and supplying exogenous **2.1** and **2.2**. Proteomics analysis of the *in vitro* TT-assay product showed 22 and 11 unique peptides covering 70.55% and 78.07% of FamD2 and FisC amino acids sequences, respectively (Table 2.2), demonstrating *in vitro* production of the prenyltransferase and cyclase enzymes.

The LC-MS analysis of the reaction showed production of 12-*epi*-fisherindole U **2.7**, which matched an authentic standard (Figure 2.4). This result also correlated with our previous results derived from chemoenzymatic synthesis of **2.7.**[14]

Table 2.2: Proteomics analysis of FamD2 and FisC expression from CFPS system. Plasmids bearing individual *famD2* and *fisC* genes were analyzed for expression of expected protein. The sequence highlighted in yellow represents the coverage of the synthesized protein.

FamD2		**FisC**	
Total peptide	22	Total peptide	11
Unique peptide	22	Unique peptide	11
Coverage (%)	70.55	Coverage (%)	78.07

Annotated Peptide Sequence	Sequence Positions	PSM FamD2	Annotated Peptide Sequence	Sequence Position	PSM FamC1
[R].IRTDIINVAK.[T]	FamD2 [7-16]	8	[K].NAGFEEPSLTVEDYYTIDTPPGWITYDPNGLVPAK.[R]	FisC [36-70]	8
[R].TDIINVAK.[T]	FamD2 [9-16]	6	[K].NAGFEEPSLTVEDYYTIDTPPGWITYDPNGLVPAKR.[T]	FisC [36-71]	1
[K].TFGAEYSEK.[V]	FamD2 [17-25]	10	[R].ITSNNGVGYTGPNSAYYNHK.[A]	FisC [74-93]	110
[K].VLDEVFQVFGEQFADNSFMIR.[T]	FamD2 [26-46]	5	[R].ITSNNGVGYTGPNSAYYNHKAPEGR.[N]	FisC [74-98]	6
[R].TSNKQPDK.[L]	FamD2 [47-54]	5	[R].NVAYVYLAQEIGSGIAGLEQTLDAVLKPNTK.[Y]	FisC [99-129]	1
[R].TSNKQPDKLGCYFR.[Y]	FamD2 [47-60]	1	[K].YTLTVDIGNSGGSFQGFPLDGFPGYR.[V]	FisC [130-155]	2
[K].LGCYFR.[Y]	FamD2 [55-60]	5	[R].VELLAGDTVLAADQNNLYIK.[E]	FisC [156-175]	107
[R].YHEEDESQLGLAWDIAR.[K]	FamD2 [61-77]	30	[K].TTTVTFIATPESPYLGQHLGIR.[L]	FisC [181-202]	32
[R].KSGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [78-112]	4	[R].LINPLQGK.[F]	FisC [203-210]	27
[K].SGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [79-112]	2	[R].LINPLQGKFSGVDFDNVR.[L]	FisC [203-220]	1
[K].IWQSIK.[G]	FamD2 [118-123]	5	K].FSGVDFDNVR.[L]	FisC [211-220]	20
[K].GVVPVQDAFK.[L]	FamD2 [124-133]	8			
[K].LSLPASVTTHSDFLK.[N]	FamD2 [134-148]	8			
[K].HHTSEYYK.[N]	FamD2 [178-185]	1			
[R].LCFYLPFLNR.[E]	FamD2 [224-233]	1			
[R].EAVPQNLLNPLLK.[K]	FamD2 [234-246]	15			
[R].EAVPQNLLNPLLKK.[Y]	FamD2 [234-247]	4			
[K].KYINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2 [247-273]	1			
[K].YINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2[248-273]	2			
[K].VDVDYHGR.[T]	FamD2 [279-286]	1			
[R].TVPLFMK.[V]	FamD2 [287-293]	10			
[K].VHSQPLPK.[A]	FamD2 [294-301]	4			

FamD2
MNDVNRIRTDIINVAKTFGAEYSEKVLDEVFQVFGEQFADNSFMIRTSN
KQPDKLGCYFRYHEEDESQLGLAWDIARKSGLLSDQGRPVDQLIPEICET
FPIMADGVDFDVKHGLAKIWQSIKGVVPVQDAFKLSLPASVTTHSDFLK
NHHLDALYAFGIDYHHSSVNLYFDTYHPKHHTSEYYKNLLQDLQFQPPS
DELLELLTNNGEIALTFNFASPRIERLCFYLPFLNREAVPQNLLNPLLKKY
INEAPALVDNPGFILGWSFGPQGGKGTYTKVDVDYHGRTVPLFMKVHS
QPLPKAADFALAQ, 309 residues

FisC
MKRNFIIAAIVLLVYICFGISISANAASAVSIPIKNAGFE
EPSLTVEDYYTIDTPPGWITYDPNGLVPAKRTRITSNN
GVGYTGPNSAYYNHKAPEGRNVAYVYLAQEIGSGIA
GLEQTLDAVLKPNTKYTLTVDIGNSGGSFQGFPLDGFP
GYRVELLAGDTVLAADQNNLYIKEKDFKTTTVTFIAT
PESPYLGQHLGIRLINPLQGKFSGVDFDNVRTAEPAET,
228 residues

2.3.3 Generation of unnatural hapalindoles and fischerindoles

Once we were able to employ CFPS to produce prenyltransferase and cyclase enzymes for assembly of **2.3** for hapalindole and fischerindole assembly, we sought to test this system as a tool for rapid screening of non-native substrates. Toward this end, we synthesized a series unnatural substrates including 5- and 6-halogenated *cis*-indole isonitrile derivatives, specifically 5-fluoro-**2.9**, 6-fluoro-**2.10**, 6-chloro-**2.11**, 5-bromo-**2.12**, 6-bromo-**2.13**, and 5-iodo-*cis*-indole isonitrile **2.14** (Figure 2.5). For these experiments, we employed the *in vitro* TT-assay to assess the unnatural *cis*-indole isonitriles using plasmids encoding FamC1 and FisC individually with

the goal of generating the corresponding unnatural hapalindoles linked at C-3/4, or C-2/3, respectively. The metabolite profile for each reaction was initially determined using mass spectrometry, and halogenated products were further characterized by chemoenzymatic scale-up methods for full NMR structure elucidation.

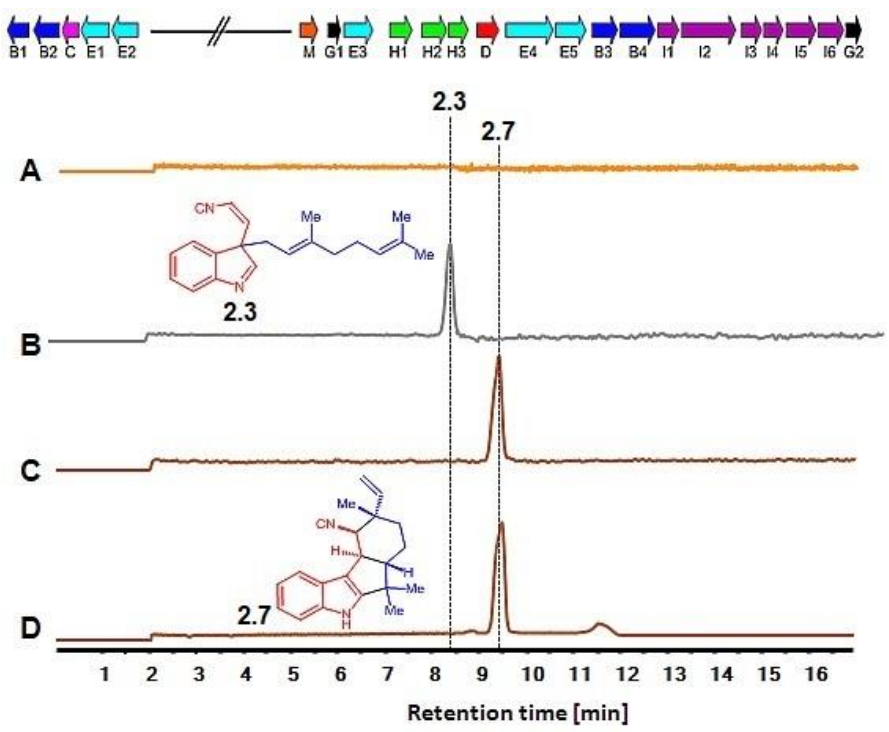


Figure 2.4: *In vitro* TT-assay and production of native fischerindole. Top: Biosynthetic gene cluster of *Fischerella sp* SAG 46.79 (fis). The genes for oxygenases (blue, B1-4), cyclase (pink, C), aromatic prenyltransferases (red, D), GPP synthases (cyan, E1-5), methyl transferase (red, M), tryptophan synthase (magenta, I1-6), isonitrile synthase (green, H1-3) and other core enzymes (black, G1-2) are shown. Bottom: Extracted ion chromatograms (EICs) are shown in all panels. Detection of in vitro synthesized fischerindole products showing EIC (m/z of 305) with no plasmid (A), FamD2 alone (B) for **2.3**, and FamD2 plus FisC (C) for **2.7** production. The trace D is the authentic standards of **2.7.**

LC-MS analysis of the *in vitro* TT-assay comprised of plasmids encoding FamD2 and FamC1 in the presence of substrate **2.1** and the fluorinated *cis*-indole isonitrile **2.9** or **2.10** revealed a cyclized product (*m/z* of 323.18) (Figure 2.6 A-B). The new metabolite was further validated by employing heterologously purified FamD2 and FamC1 enzymes for *in vitro* chemoenzymatic reactions of unnatural *cis*-indole isonitrile **2.9** or **2.10** containing **2.1**, and the molecule was isolated with an identical t_R and MS-pattern compared to CFPS-derived product. Similarly, the *in vitro* TT-assay performed with the exogenous FamD2 and FisC with the substrate **2.1** and **2.9** or **2.10**, also showed the cyclized product (*m/z* of 323.19) identical with the molecule derived from standard chemoenzymatic synthesis (Figure 2.6 C-D). However, screening of the other halogenated *cis*-indole isonitrile **2.11, 2.12, 2.13** and **2.14** for both FamC1 and FisC failed to show cyclized products. This suggested the inability of FamC1 or FisC to cyclize more sterically demanding halogenated (F<Cl<Br<I) *cis*-indole isonitrile substrates, despite forming the corresponding C3-geranylated *cis*-indole isonitrile derivatives. The structure-function study of

aromatic prenyltransferase and its biocatalytic potential to synthesize hapalindole family related molecules has recently been reported.[17] However this study using CFPS represents the first reported effort to examine the substrate scope of Stig cyclases using unnatural *cis*-indole isonitrile derivatives.

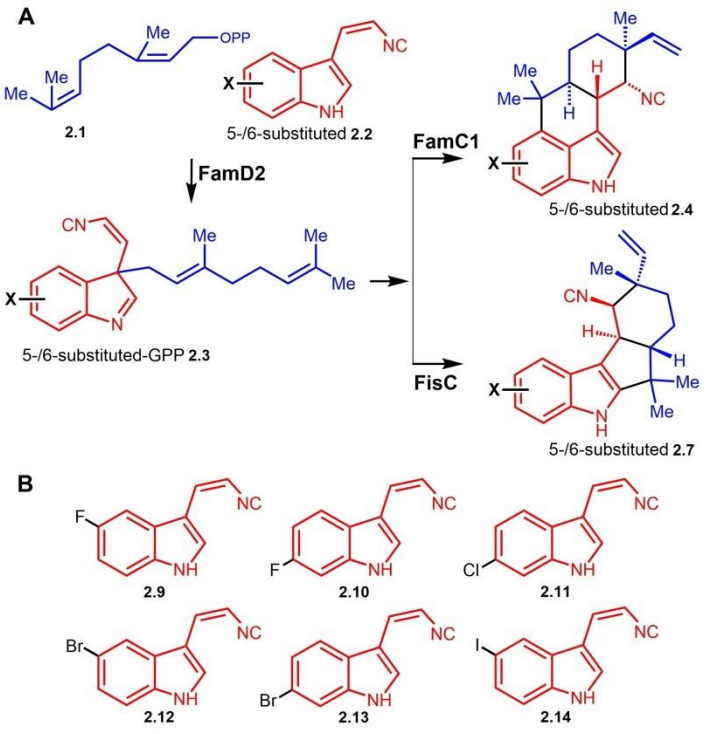


Figure 2.5: Screening of unnatural *cis*-indole isonitrile derivatives. **A**: Scheme showing the screening of unnatural 5- or 6-halogenated *cis*-indole isonitrile by CFPS. **B**: Structure of the halogenated *cis*-indole isonitrile derivatives. 5-fluoro- **2.9**, 6-fluoro- **2.10**, 6-chloro- **2.11**, 5-bromo- **2.12**, 6-bromo- **2.13** and 5-iodo **2.14** were used in this study.

In each experiment where the *in vitro* TT-assay showed a new metabolite, we proceeded with chemoenzymatic scale-up for structure elucidation. Thus, reaction of purified FamD2 and FamC1 proteins in the presence of GPP **2.1** and unnatural *cis*-indole isonitrile **2.9** and **2.10** provided new hapalindoles, 5-fluoro- **2.15** and 6-fluoro-12-*epi*-hapalindole U **2.16**, respectively (Table S2.2 & S2.3). However, the scaled-up product of **2.1** with **2.9** and **2.10** for FisC was generated using homologous protein FimC5 from the fischerindole producing strain *Fischerella muscicola* UTEX 1829[14] because of its identical catalytic function[15] and higher enzymatic activity. The corresponding novel identified products were indole C-2/3 ring fused unnatural fischerindole, 5-fluoro- **2.17** and 6-fluoro-12-*epi*-fischerindole U **2.18** (Table S2.4 & S2.5). Stereochemistry of **2.15, 2.16, 2.17,** and **2.18** were assigned by NMR association with authentic **2.4** and **2.8**.

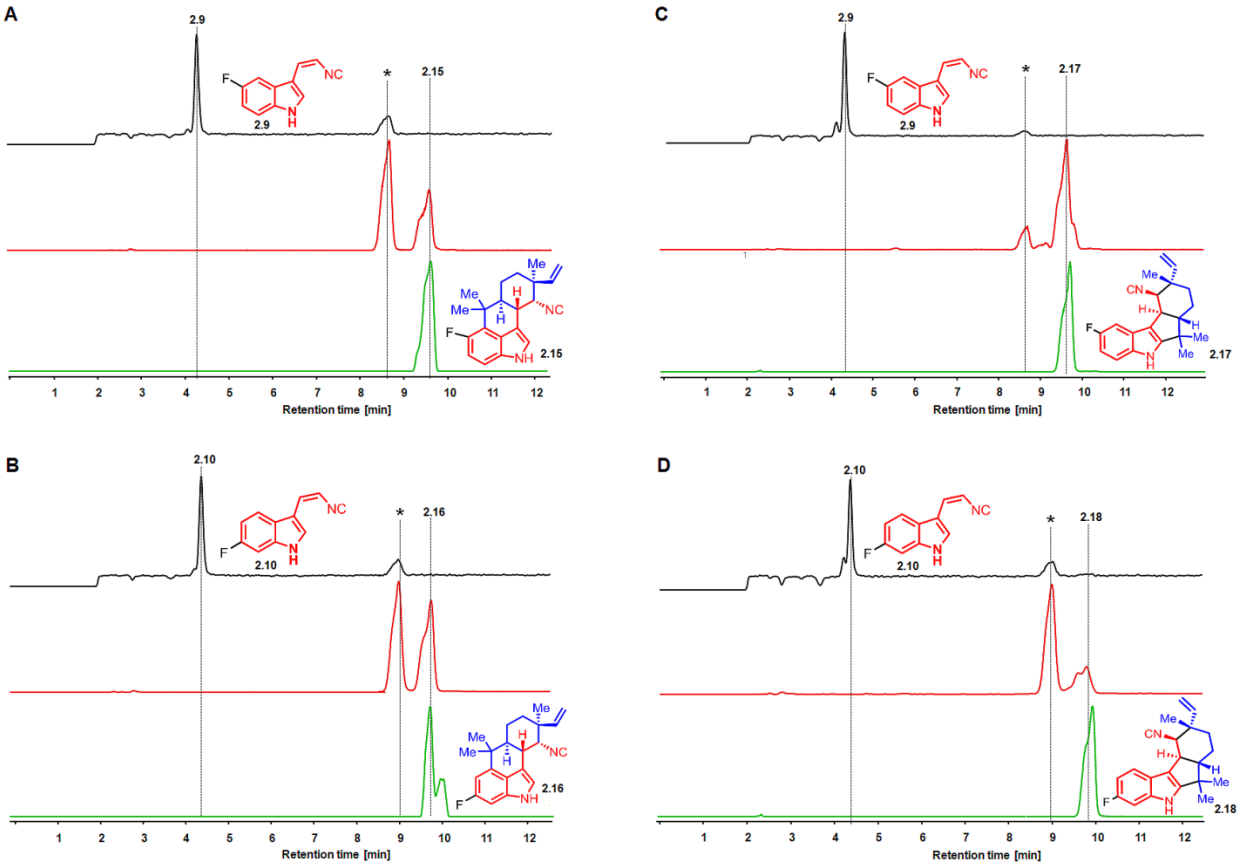


Figure 2.6: Screening of unnatural fluorinated *cis*-indole isonitrile substrates. 5-fluoro **2.9** (Panel A and C) and 6-fluoro **2.10** (Panel B and D). Substrates **2.9** and **2.10** are converted into 5-fluoro-12-*epi*-hapalindole U **2.15** and 6-fluoro-12-*epi*-hapalindole U **2.16** using FamD2 and FamC1 plasmids (Panel A and B) with **1.1**. Whereas the combination of exogenous plasmids FamD2 and FisC showed the production of 5-fluoro-12-*epi*-fischerindole U **2.17** and 6-fluoro-12-*epi*-fischerindole U **2.18** (Panel C and D). Extracted ion chromatogram (EIC) for the substrate **2.9** or **2.10** (black line, *m/z* of 187.02) and TT-assay (brown line, *m/z* of 323.18) are shown. EICs of authentic product-standard **2.15**, **2.16, 2.17** and **2.18** (green line, *m/z* of 323.18) obtained by corresponding chemoenzymatic reaction are shown. The '*' represents 5F-/6F **2.3** intermediates.

2.4 Conclusion

In this work, we sought to establish an *in vitro* TT-assay platform as a biosynthetic prototype for the synthesis and screening of natural and unnatural hapalindole-type molecule. By using detailed proteomics analysis, we have demonstrated that the assay comprised of multiple plasmids showed production of a prenyltransferase and selected Stig cyclases from cyanobacterial indole alkaloid pathways. The assay was further applied to screen chemically synthesized halogenated *cis*-indole isonitrile substrates for assembly of unnatural hapalindole-type molecules. The TT-assay guided screening for metabolite production was scaled-up using chemoenzymatic synthesis, and the structural identities of the novel unnatural hapalindole-type molecules were determined. We believe this approach represents an effective, high throughput

strategy to determine the functional role of homologous genes available in the large pool of diverse natural product BGCs.

2.5 Experimentals

2.5.1 General

All chemicals were purchased from Sigma-Aldrich, ACROS, and Combi-Blocks. Plasmid pET28a bought from Invitrogen was used for cloning and expression of full length FamD2, FamC3 and FamC4. Plasmid pET28H8T (pET28b variant generated in house containing an N-terminal 8xHis-tag followed by a TEV cleavage sequence upstream of the multiple cloning site) was used for cloning and expression of N-terminally truncated FamC1, full length FamC2, FimC5 and FisC. The protein expression host *Escherichia coli* strain BL21(DE3) and Ni-NTA agarose to purify His-tag proteins were also purchased from Invitrogen. The liquid chromatography–mass spectrometry (LC–MS) was performed on an Agilent G6545A quadrupole-time-of-flight (Q-TOF) or Agilent 6230 time-of-flight (TOF) using an XBridge Shield RP18 3.5 μm, 3.0 mm × 150 mm column from Waters. The chromatographic method for all substrate was from 60-100% acetonitrile water gradients over 27 min. Preparative-scale HPLC was performed on a Shimadzu 20-AT equipped with an LUNA C18 250 x 10 mm column, using a mobile phase gradient of 60-100% acetonitrile in water over 60 min. All NMR spectra were acquired on a Varian 400 and 600 MHz and Bruker 800 MHz spectrometers. Proton and carbon signals are reported in part per million (δ) using residual solvent signals as an internal standard. Multiplicities are abbreviated as following: singlet (s), doublet (d), triplet (t), quartet (q), doublet-doublet (dd), triplet-doublet (td), doublet-doublet-doublet (ddd), triplet-doublet-doublet (tdd), and multiplet (m). *Chemical abbreviations:* Ethyl Acetate (EtOAc), Tetrahydrofuran (THF), Potassium bis(trimethylsilyl)amide (KHMDS), Acetic Acid (AcOH), Sodium Sulfate (Na_2SO_4), Diethyl Ether (Et_2O), Acetonitrile (CH_3CN), Magnesium Chloride ($MgCl_2$), Calcium Chloride ($CaCl_2$), Sodium Chloride (NaCl).

2.5.2 *In vitro* TT-assay, proteomics analysis and substrate conversion

***In vitro* TT-assay**: The microscale *in vitro* synthesis of natural and unnatural hapalindole and fischerindole molecules was performed using a PURExpress® *In vitro* Protein Synthesis kit (New England Biolabs). This system is based on the PURE system originally developed by the

Ueda group[34] and later commercialized as PURESYSTEM® by Biocomber (Tokyo, Japan).[35] The assay comprised of synthetic substrates **2.1** and **2.2** or halogenated 5-/6-substituted **2.2** (**2.9**-**2.14**) and exogenous prenyltransferase (FamD2) and *fam* (FamC1-C4) or *fis* (FisC) Stig cyclase plasmids to imitate the hapalindole/fischerindole biosynthetic gene cluster.

The reaction consists of exogenous plasmid(s) constructed in pET28a or pET28H8T that allows transcription of a gene under the T7 promoter into mRNA and subsequently translated into a protein in the same reaction. Our typical TT-assay comprised of 500 ng freshly extracted plasmid (dissolved in nuclease free H_2O, added <2 µL), 0.6 µL of murine RNase inhibitor (40 U/µL, New England Biolabs), commercially available solution A (12 µL) and solution B (8 µL) (New England Biolabs). Solution A is the premix of tRNAs, rNTPS and amino acids mixture, while solution B is comprised of ribosomes and protein components such as T7 RNA polymerase, translation factors, aminoacyl-tRNA synthetases and energy regeneration enzymes. In our assay, 1.2 µL of DTmix chaperone (provided by New England Biolabs) was also added. The noncommercial DTmix solution of 1 µL comprised of 2.4 µM DnaK, 0.48 µM DnaJ, 1.4 µM GrpE and 1.2 µM of Trigger Factor. The *in vitro* TT-assay was performed in a PCR tube at an end volume of 30 µL adjusted with 50 mM Tris HCl buffer, pH 7.8. The assay was incubated at 25°C for 3 h with agitation at 150 rpm for the transcription and translation of exogenous plasmid(s). Depending on the target for biosynthetic assembly of natural or unnatural hapalindole and fischerindole compounds, the corresponding exogenous plasmid(s) were used. The control TT-assay devoid of plasmid was used as a negative control in which only the substrate **2.2 (**or 5-/6-substituted **2.2**) could be detected in the LC-MS, whereas the assay comprised of FamD2 plasmid alone produce **2.3 (**or 5-/6-substituted-GPP **2.3**). The presence of FamD2 plus FamC1 plasmids produce **2.4**, FamD2 plus FamC2 and FamC3 produce **2.6**, and FamD2 plus FamC1 and FamC4 showed the production of **2.4** and **2.5**. The combination of FamD2 and FisC plasmids result in production of **2.7.** Use of unnatural 5-/6-fluorinated **2.2**, FamD2 in combination with FamC1 or FisC showed the production of 5-/6-substituted **2.15**, **2.16**, **2.17** and **2.18**. The sample was subjected to proteomics analysis and substrate conversion as described below. Assays were conducted with proteins known to be soluble from previous *E. coli* production. Proteins that may be insoluble in *E. coli* or other hosts were not assessed in the current study.

Proteomics analysis by LC-Tandem MS*:* Proteomics analysis of the *in vitro* TT-assay samples were performed in the Proteomics Resource Facility at the University of Michigan using an LC-MS based approach. Briefly, following reduction (10 mM DTT, 30 min) and alkylation (65 mM 2-chloroacetamide, 30 min), proteins were digested overnight with 500 ng of sequencing grade modified trypsin (Promega). The resulting peptides were resolved on a nano-capillary reverse-phase column (Acclaim PepMap C18, 2 microns, 50 cm, Thermo Fisher Scientific) using 0.1% formic acid/acetonitrile gradient at 300 nL/min (2-25% acetonitrile in 105 min; 25-40% acetonitrile in 20 min followed by a 90% acetonitrile wash for 10 min and a further 30 min re-equilibration with 2% acetonitrile) and directly introduced into Q Exactive HF mass spectrometer (Thermo Fisher Scientific, San Jose CA).

MS1 scans were acquired at 60 K resolution. Data-dependent high-energy C-trap dissociation MS/MS spectra were acquired with top speed option (3 sec) following each MS1 scan (relative CE ~28%). Proteins were identified by searching the data against *E. coli* database (UniProtKB, 4331 entries, downloaded on 11/07/2018) appended with FamD2, FamC1 and FisC protein sequences using Proteome Discoverer (v2.1, Thermo Fisher Scientific). Search parameters included MS1 mass tolerance of 10 ppm and fragment tolerance of 0.1 Da; two missed cleavages were allowed; carbamidimethylation of cysteine was considered fixed modification and oxidation of methionine, deamidation of asparagine and glutamine were considered as potential modifications. FDR was determined using Percolator and proteins/peptides with a FDR of ≤1% were retained.

Substrate conversion and analysis of product*:* Following a 3 h incubation period of *in vitro* TT-assay components with exogenous plasmid, the substrates GPP **2.1** and native/unnatural *cis*-indole isonitrile **2.2** were added and incubated further for 2 h, and extracted and analyzed as described below. Typically, for the TT-assay consisting of the exogenous plasmid FamD2 alone or the combination of FamD2 plus FamC1 or FamD2 plus FisC, 10 mM each of GPP **2.1** and *cis*-indole isonitrile **2.2,** $MgCl_2$ (5 mM), $CaCl_2$ (5 mM) and the energy mix (glucose-6-phosphate (5 µM), G-6-P-dehydrogenase (0.5 U), NADPH (1 mM) and glucose (1 mM) were used. For the assay consisting of three plasmids- FamD2 plus FamC1 and FamC4 or FamD2 plus FamC2 and FamC3- higher concentration of the $CaCl_2$ (20 mM) was used.

In order to analyze the product, the reaction was quenched by adding equal volume of ethyl acetate and extracted twice. The organic layer was combined, dried in a nitrogen stream,

and reconstituted in 20 μL of acetonitrile for LC–MS analysis. The analytical scale LC–MS run was performed on an Agilent G6545A quadrupole-time-of-flight (Q-TOF) or Agilent 6230 time-of-flight (TOF) mass spectrometer equipped with a dual AJS ESI source and an Agilent 1290 Infinity series diode array detector, auto sampler, and binary pump. An XBridge Shield RP18 3.5 μm, 3.0 mm × 150 mm from Waters was used for all separations. The chromatographic method for all substrate gradients was from 60% to 100% of acetonitrile in water over 27 min. A volume of 1 μL injections was made for each sample.

2.5.3 Protein expression and purification

The expression and purification of FamD2, FamC1 and FimC5 was performed as described.[14,31] Briefly, a single BL21(DE3) colony of FamD2/FamC1/FimC5 transformant was inoculated in LB medium containing 50 μg/mL kanamycin and grown overnight at 37°C shaking at 200 rpm. The main culture (1 L) was prepared at the dilution of 1:100 in 2.8 L of Fernbach flask containing LB medium and the same concentration of antibiotic. The cells were grown (37°C, 200 rpm) to an optical density (A_{600} nm) of 0.6. The culture flasks were chilled in ice for 30 min, induced with IPTG (0.2 mM), and incubated (16°C, 200 rpm) further for 16 h. The cells were harvested (5000 rpm, 4°C, 15 min), flash frozen in liquid nitrogen, and stored at −80°C until purification.

For the protein purification, the cell pellets were resuspended in chilled lysis buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol), containing 0.5 mg/mL of lysozyme and 1 mL of 2 mg/mL DNAse. The mixture was stirred for 30 min and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The cellular debris was removed by centrifugation (35000xg, 4°C, 35 min). Imidazole (10 mM) was added in the clarified lysate and loaded onto Ni-NTA agarose column equilibrated with lysis buffer. The column was washed with five column volume of wash buffer (10 mM HEPES, 300 mM NaCl, 0.2 mM TCEP, 10% glycerol, 20 mM imidazole) and the His-tagged protein was eluted with elution buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol, 300 mM imidazole). The fractions were pooled and dialyzed using a PD10-desalting column (GE Healthcare) using lysis buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol). The purified protein was analyzed by SDS-PAGE for purity, measured by Nanodrop for concentration, and flash-frozen in liquid nitrogen to store at −80°C.

2.5.4 Scale-up chemoenzymatic reactions

For the structure analysis of the enzymatic products, reactions were scaled-up to 10 mL containing FamD2 (5 μM), cyclase (20 μM), substrate (1 mM), GPP (1 mM), $MgCl_2$ (5 mM) and 7.5 mM $CaCl_2$ in 50 mM Tris-HCl, pH 7.0 buffer. The reaction was incubated at 37°C until completion by HPLC, quenched and extracted with EtOAc, dried with Na_2SO_4 and concentrated. The extracted products were purified by Prep HPLC. The purified compounds were concentrated, dissolved in C_6D_6 and analyzed using a Varian 600 MHz NMR and Bruker 800 MHz NMR. All compounds were obtained as a white solid.

2.5.5 Chemical synthesis of *cis*-indole isonitrile derivatives

All derivatives were prepared using method previously described.[14] Briefly, in a 50 mL two-neck round-bottom flask purged with nitrogen at −78°C (dry ice/acetone), diethyl (isocyanomethyl) phosphonate (0.20 mL, 1.248 mmol) was diluted with THF (5 mL). KHMDS (1 M THF, 2.60 mL, 2.60 mmol) was added dropwise, and the reaction was stirred at −78°C for 30 min. To a separate 4 mL vial, indole-3-carboxaldehyde derivative (1.13 mmol) was dissolved in THF (5 mL), and the resulting solution was added dropwise to the KHMDS solution at −78°C. The resulting mixture was stirred at 0°C (cryocool) overnight. The resulting solution was quenched by the addition of AcOH (0.15 mL, 2.6 mmol) and concentrated. The resulting residue was diluted with EtOAc (20 mL), washed with 1 M aqueous potassium phosphate buffer (20 mL, pH 7), washed with brine, dried with Na_2SO_4, and concentrated to an oil. The residue was dissolved in Et_2O and purified by flash chromatography (24%–100% pentane/Et_2O, SiO_2) to afford the titled compound as reported below. Yields and spectral data reported below.

(Z)-3-(2-isocyanovinyl)-1H-indole (*cis*-indole isonitrile) (2.2): Red Solid, 21mg, 11%

^{1}H NMR (600 MHz, Acetone-d_6) δ 5.37 (dd, J = 11.8, 1.1 Hz, 1H), 7.14 – 7.31 (m, 2H), 7.50 – 7.60 (m, 1H), 7.67 (d, J = 11.7 Hz, 1H), 7.77 – 7.86 (m, 1H), 8.37 (s, 1H), 11.02 (s, 1H).

(Z)-5-fluoro-3-(2-isocyanovinyl)-1H-indole (2.9): Tan Solid, 19mg, 9%

^{1}H NMR (600 MHz, Acetone-d_6) δ 5.95 (d, J = 8.9 Hz, 1H), 6.94 (p, J = 4.8 Hz, 1H), 7.03 (td, J = 9.1, 2.5 Hz, 1H), 7.52 (td, J = 9.9, 3.5 Hz, 2H), 8.23 (s, 1H), 10.97 (s, 1H).

^{13}C NMR (151 MHz, dmso) δ 168.74, 158.87, 157.32, 132.31, 128.84, 127.59, 127.52, 124.97, 113.67, 113.61, 111.19, 111.02, 109.61, 109.58, 103.90, 103.74.

(Z)-6-fluoro-3-(2-isocyanovinyl)-1H-indole (2.10): Tan Solid, 14mg, 7%

^{1}H NMR (600 MHz, Acetone-d_6) δ 5.97 (d, J = 8.9 Hz, 1H), 6.93 – 7.01 (m, 2H), 7.27 (dd, J = 9.7, 2.3 Hz, 1H), 7.77 (dd, J = 8.7, 5.2 Hz, 1H), 8.17 (s, 1H), 10.93 (s, 1H).

(Z)-6-chloro-3-(2-isocyanovinyl)-1H-indole (2.11): Yellow solid, 36mg, 31%

^{1}H NMR (599 MHz, Acetone-d_6) δ 5.96 (d, J = 8.9 Hz, 1H), 6.94 (dt, J = 9.5, 4.9 Hz, 1H), 7.16 (dd, J = 8.5, 1.9 Hz, 1H), 7.57 (d, J = 1.8 Hz, 1H), 7.75 (d, J = 8.5 Hz, 1H), 8.17 – 8.20 (m, 1H), 11.02 (s, 1H). ^{13}C NMR (151 MHz, acetone) δ 170.89, 137.02, 128.93, 128.64, 128.47, 126.68, 124.71, 121.72, 120.25, 112.65, 110.61.

(Z)-5-bromo-3-(2-isocyanovinyl)-1H-indole (2.12): Red Solid, 51mg, 37%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.96 (d, J = 8.9 Hz, 1H), 6.97 (dt, J = 9.2, 4.7 Hz, 1H), 7.33 (dd, J = 8.6, 1.9 Hz, 1H), 7.49 (d, J = 8.6 Hz, 1H), 7.95 (d, J = 1.8 Hz, 1H), 8.20 (d, J = 2.2 Hz, 1H), 11.08 (s, 1H). ^{13}C NMR (101 MHz, acetone) δ 170.77, 135.27, 129.69, 128.97, 128.80, 126.21, 124.63, 121.61, 114.62, 114.27, 110.11

(Z)-6-bromo-3-(2-isocyanovinyl)-1H-indole (2.13): Red Solid, 10mg, 4%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.98 (d, J = 8.9 Hz, 1H), 6.85 – 7.00 (m, 1H), 7.29 (dd, J = 8.5, 1.8 Hz, 1H), 7.67 – 7.76 (m, 2H), 8.18 (d, J = 1.9 Hz, 1H), 10.99 (s, 1H). ^{13}C NMR (151 MHz, acetone) δ 169.98, 136.50, 127.61, 126.03, 123.70, 123.38, 119.69, 115.57, 114.74, 109.70, 104.77.

(Z)-5-iodo-3-(2-isocyanovinyl)-1H-indole (2.14): Red Solid, 10mg, 6%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.98 (d, J = 8.9 Hz, 1H), 7.00 (d, J = 9.0 Hz, 1H), 7.40 (d, J = 8.5 Hz, 1H), 7.51 (dd, J = 8.6, 1.7 Hz, 1H), 8.16 (d, J = 3.0 Hz, 2H), 11.03 (s, 1H). ^{13}C NMR (151 MHz, acetone) δ 169.93, 152.01, 134.82, 130.90, 129.54, 127.63, 127.06, 123.71, 114.15, 108.91, 83.50.

2.5.6 Chemical synthesis of geranyl diphosphate, tri ammonium

The synthesis was done as described previously.[36] Briefly, in a 10 mL round bottom flask purged with nitrogen, tris (tetrabutylammonium) hydrogen pyrophosphate (1.0 g, 1.04 mmol) was dissolved in CH_3CN (1.0 mL). Geranyl chloride (0.09 mL, 0.475 mmol) was added and the reaction mixture was stirred at room temperature for 2 h.

Dowex 50WX8 resin preparation*:* Dowex 50WX8 resin (20 g, hydrogen form) was washed with half saturated aqueous ammonium chloride (5 x 50 mL) and water (5 x 50 mL) until the pH

of the supernatant equaled 5. The slurry was rinsed twice with ion exchange buffer (2% isopropanol in 25 mM aqueous ammonium bicarbonate) and loaded into a flash column and equilibrated with ion exchange buffer.

Purification*:* The reaction mixture was concentrated to afford an orange residue which was diluted with ion exchange buffer. The crude mixture was chromatographed with two column volumes of ion exchange buffer (75 mL). The fractions were combined and concentrated by rotary evaporation, flash frozen and lyophilized for 2 days. The resulting white powder was diluted with 1 M ammonium bicarbonate (4 mL) and 50% isopropanol/CH_3CN (10 mL), vortexed for 30 seconds and centrifuged (2,000 rpm, rt, 5 min). The organic layer was extracted and the residual 0.5 mL of yellow liquid was diluted with 50% isopropanol/CH_3CN and the dilution/vortex/centrifugation process was repeated twice. The combined organic layers were concentrated to afford a while solid (360 mg). The white solid was taken up in 50% isopropanol:25% CH_3CN:25%.1M aqueous ammonium bicarbonate and chromatographed with cellulose. The resulting fractions were combined and lyophilized affording the titled compound as a white powder (108 mg, 62.2%).

^{1}H NMR (400 MHz, D_2O/ND_4OD) δ 1.92 (d, *J* = 1.3 Hz, 3H), 1.98 (s, 3H), 2.01 (d, *J* = 1.3 Hz, 3H), 2.39 (d, *J* = 6.5 Hz, 2H), 2.41 – 2.49 (m, 2H), 5.45 – 5.53 (m, 1H), 5.74 (dt, *J* = 6.1, 3.9 Hz, 1H).

^{13}C NMR (151 MHz, D_2O/ND_4OD) δ 142.46, 133.67, 124.40, 120.46, 62.61, 39.10, 25.92, 25.16, 17.26, 15.89. ^{31}P NMR (162 MHz, D_2O/ND_4OD) δ -9.93 (d, *J* = 21.6 Hz), -6.08 (d, *J* = 21.6 Hz).

2.6 References

1. Katz, L., and Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016** *43* (2-3), 155-176. https://doi.org/10.1007/s10295-015-1723-5.
2. Patridge, E., Gareiss, P., Kinch, M. S., and Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* **2016**, *21* (2), 204-207. https://doi.org/10.1016/j.drudis.2015.01.009.
3. Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158* (2), 412-421. https://doi.org/10.1016/j.cell.2014.06.034.
4. Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963-968. http://www.nature.com/doifinder/10.1038/nchembio.1659.
5. Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., and Medema, M. H. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36-W41. https://doi.org/10.1093/nar/gkx319.
6. Skinnider, M. A., Merwin, N. J., Johnston, C. W., and Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **2017**, *45* (W1), W49-W54. https://doi.org/10.1093/nar/gkx320.
7. Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Dusterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kotter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nutzmann, H. W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gomez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y. Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J., Ehling-Schulz, M., Entian, K. D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Hofte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H. J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Mendez, C., Metsa-Ketela, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Muller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J. L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Sussmuth, R. D.,

Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., and Glockner, F. O. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625-631.

8. Noda-Garcia, L., Liebermeister, W., and Tawfik, D. S. Metabolite-Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annu. Rev. Biochem.* **2018**, *87*, 187-216. https://doi.org/10.1146/annurev-biochem-062917-012023.
9. Fischbach, M. A., and Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106* (8), 3468-3496. https://doi.org/10.1021/cr0503097.
10. Hudson, G. A., and Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Curr. Opin. Microbiol.* **2018**, *45*, 61-69. https://doi.org/10.1016/j.mib.2018.02.010.
11. Li, Y., and Rebuffat, S. The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. *J. Biol. Chem.* **2019,** *295* (1), 34-54. https://doi.org/10.1074/jbc.REV119.006545.
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* (1), 213. https://doi.org/10.1186/s12866-014-0213-7.
13. 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. https://doi.org/10.1039/C7CC00782E.
14. 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. https://doi.org/10.1038/nchembio.2327.
15. Wong, C. P.; Awakawa, T.; Nakashima, Y.; Mori, T.; Zhu, Q.; Liu, X.; Abe, I. Two Distinct Substrate Binding Modes for the Normal and Reverse Prenylation of Hapalindoles by the Prenyltransferase AmbP3. *Angew. Chem. Int. Ed.* **2018**, *57* (2), 560–563. https://doi.org/10.1002/anie.201710682.
16. Wang, J.; Chen, C.-C.; Yang, Y.; Liu, W.; Ko, T.-P.; Shang, N.; Hu, X.; Xie, Y.; Huang, J.-W.; Zhang, Y.; Guo, R.-T. Structural Insight into a Novel Indole Prenyltransferase in Hapalindole-Type Alkaloid Biosynthesis. *Biochem. Biophys. Res. Commun.* **2018**, *495* (2), 1782–1788. https://doi.org/10.1016/j.bbrc.2017.12.039.
17. Awakawa, T.; Mori, T.; Nakashima, Y.; Zhai, R.; Wong, C. P.; Hillwig, M. L.; Liu, X.; Abe, I. Molecular Insight into the Mg2+-Dependent Allosteric Control of Indole Prenylation by Aromatic Prenyltransferase AmbP1. *Angew. Chem. Int. Ed.* **2018**, *57* (23), 6810–6813. https://doi.org/10.1002/anie.201800855.
18. Perez, J. G., Stark, J. C., and Jewett, M. C. Cell-Free Synthetic Biology: Engineering Beyond the Cell. *Cold Spring Harb Perspect Biol.* **2016**, *8*. 10.1101/cshperspect.a02385.
19. Silverman, A. D., Karim, A. S., and Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* **2020**, *21*, 151-170. https://doi.org/10.1038/s41576-019-0186-3.
20. Li, J., Lawton, T. J., Kostecki, J. S., Nisthal, A., Fang, J., Mayo, S. L., Rosenzweig, A. C., and Jewett, M. C. Cell-free protein synthesis enables high yielding synthesis of an

active multicopper oxidase.*Biotechnol. J.* **2016**, *11* (2), 212-218. https://doi.org/10.1002/biot.201500030.

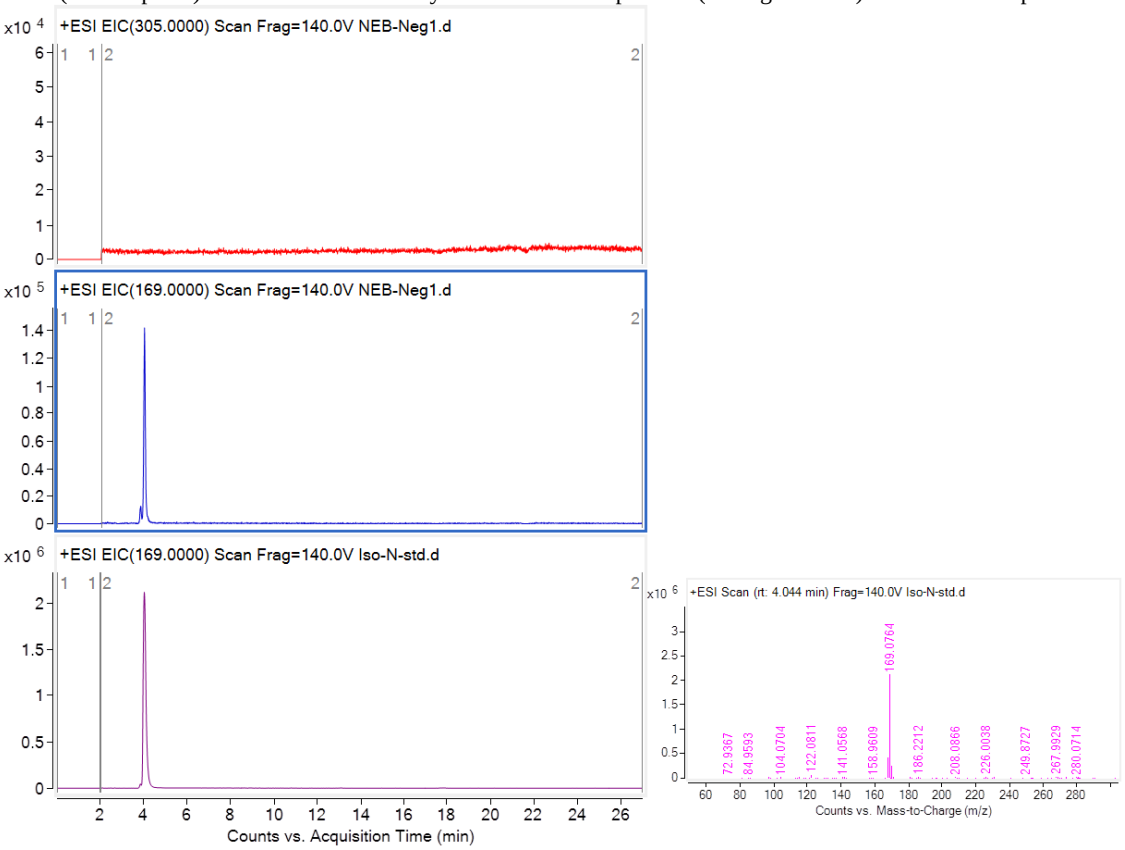
21. Boyer, M. E., Stapleton, J. A., Kuchenreuther, J. M., Wang, C. W., and Swartz, J. R. Cell-free synthesis and maturation of [FeFe] hydrogenases. *Biotechnol. Bioeng.* **2008**, *99* (1), 59-67. https://doi.org/10.1002/bit.21511.
22. Kwon, Y. C., Oh, I. S., Lee, N., Lee, K. H., Yoon, Y. J., Lee, E. Y., Kim, B. G., and Kim, D. M. Integrating cell-free biosyntheses of heme prosthetic group and apoenzyme for the synthesis of functional P450 monooxygenase. *Biotechnol. Bioeng.* **2013**, *110* (4), 1193-1200. https://doi.org/10.1002/bit.24785.
23. Dudley, Q. M., Anderson, K. C., and Jewett, M. C. Cell-Free Mixing of Escherichia coli Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis. *ACS Synth. Biol.* **2016**, *5* (12), 1578-1588. https://doi.org/10.1021/acssynbio.6b00154.
24. Dudley, Q. M., Nash, C. J., and Jewett, M. C. Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates. *Synth. Biol.* **2019**, *4* (1), 14. https://doi.org/10.1093/synbio/ysz003.
25. Kelwick, R., Ricci, L., Chee, S. M., Bell, D., Webb, A. J., and Freemont, P. S. Cell-free prototyping strategies for enhancing the sustainable production of polyhydroxyalkanoates bioplastics. *Synth.Biol.* **2018**, *3* (1), 16. https://doi.org/10.1093/synbio/ysy016.
26. Goering, A. W., Li, J., McClure, R. A., Thomson, R. J., Jewett, M. C., and Kelleher, N. L. In Vitro Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free Protein Synthesis. *ACS Synth. Biol.* **2017**, *6* (1), 39-44. https://doi.org/10.1021/acssynbio.6b00160.
27. Takahashi, M. K., Chappell, J., Hayes, C. A., Sun, Z. Z., Kim, J., Singhal, V., Spring, K. J., Al-Khabouri, S., Fall, C. P., Noireaux, V., Murray, R. M., and Lucks, J. B. Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synth. Biol.* **2015**, *4* (5), 503-515. https://doi.org/10.1021/sb400206c.
28. Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **2016**, *5* (4), 344-355. https://doi.org/10.1021/acssynbio.5b00296.
29. Kightlinger, W., Duncker, K. E., Ramesh, A., Thames, A. H., Natarajan, A., Stark, J. C., Yang, A., Lin, L., Mrksich, M., DeLisa, M. P., and Jewett, M. C. A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.* **2019**, *10* (1), 019-12024. https://doi.org/10.1038/s41467-019-12024-9.
30. 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. https://doi.org/10.1021/jacs.5b10136.
31. 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 Cyclization in Hapalindole Biogenesis. *Nat. Chem. Biol.* **2018**, *14* (4), 345–351. https://doi.org/10.1038/s41589-018-0003-x.
32. Hillwig, M. L.; Zhu, Q.; Liu, X. Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium *Fischerella Ambigua*. *ACS Chem. Biol.* **2014**, *9* (2), 372–377. https://doi.org/10.1021/cb400681n.

33. 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.* **2017**, *56* (31), 9062–9066. https://doi.org/10.1002/anie.201703932.
34. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19* (8), 751-755. https://doi.org/10.1038/90802.
35. Shimizu, Y., and Ueda, T. PURE technology, *Methods Mol. Biol.* **2010**, *607*, 11-21. https://doi.org/10.1007/978-1-60327-331-2_2.
36. Davisson, V. J., Woodside, A. B., Neal, T. R., Stremler, K. E., Muehlbacher, M., and Poulter, C. D. Phosphorylation of isoprenoid alcohols. *J. Org. Chem.* **1986**, *51* (25), 4768-4779. https://doi.org/10.1021/jo00375a005.

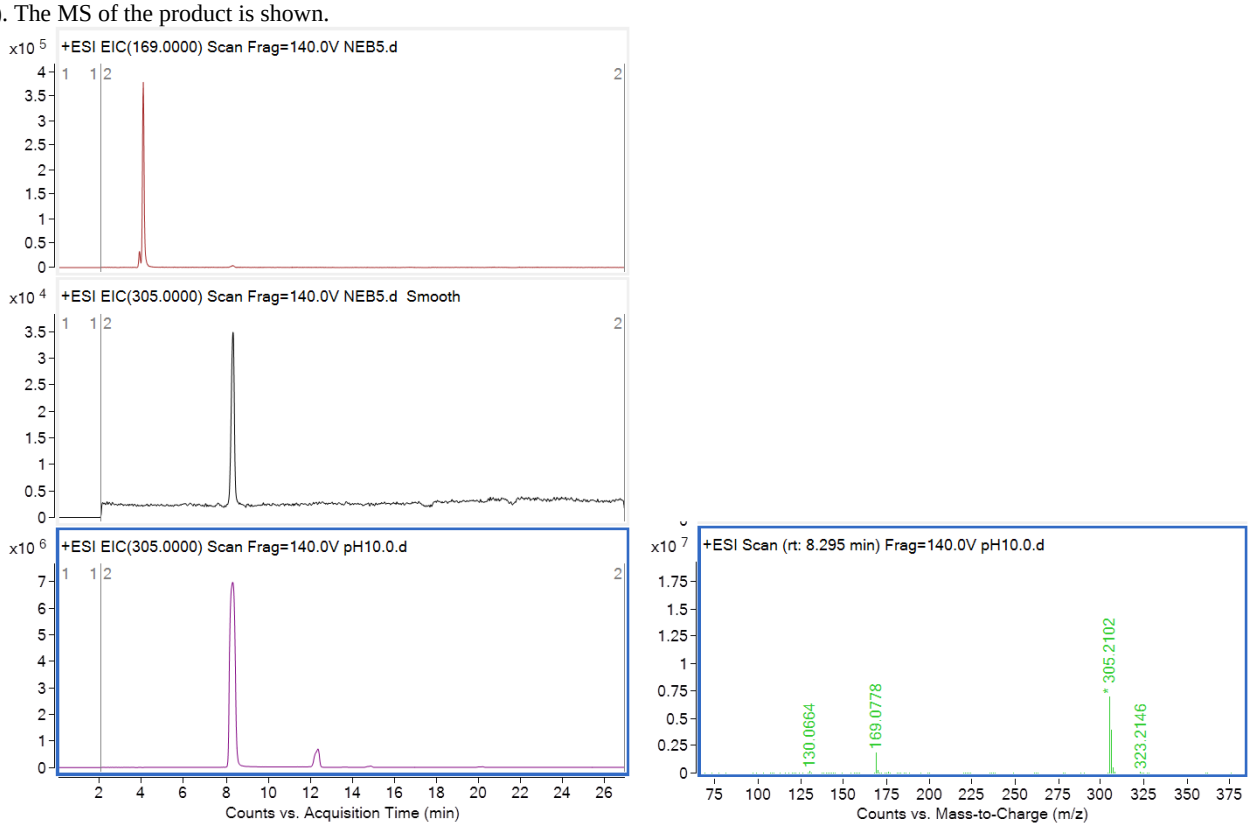
Note: This work has been published as "Multi-component microscale biosynthesis of unnatural cyanobacterial indole alkaloids." Yogan Khatri, Robert M. Hohlman, Johnny Mendoza, Shasha Li, Andrew N. Lowell, Haruichi Asahara, and David H. Sherman. *ACS Syn. Biol* **2020**, *9* (6), 1349-1360.

2.7 Supplemental information

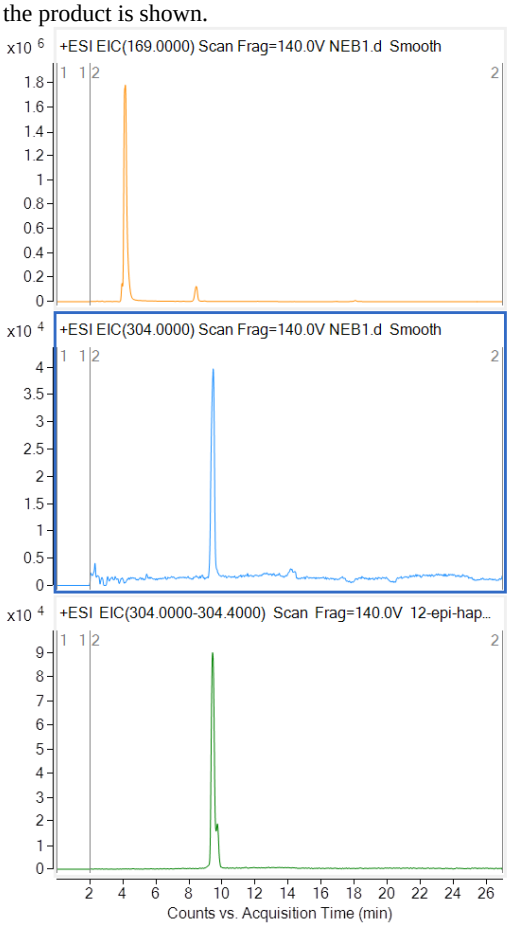
Supplemental Figure S2.1: Comparative MS data of Figure 2.3 and Figure 2.4
A. EIC of hapalindole product (*m/z* 305) (top panel) and *cis*-indole isonitrile **2.2** (*m/z* 169) (middle panel) compared with the standard (bottom panel) for the *in vitro* TT-assay in the absence of plasmid (for Figure 2.3A). The MS of the product is shown.

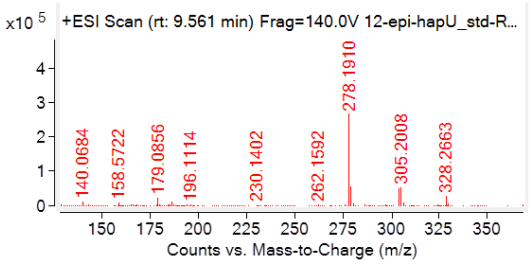


B. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel) and *cis*-indole isonitrile-GPP intermediate **2.3** (*m/z* 305) (middle panel) compared with the standard **2.3** (bottom panel) for the *in vitro* TT-assay in the presence of FamD2 plasmid alone (for Figure 2.3 B). The MS of the product is shown.

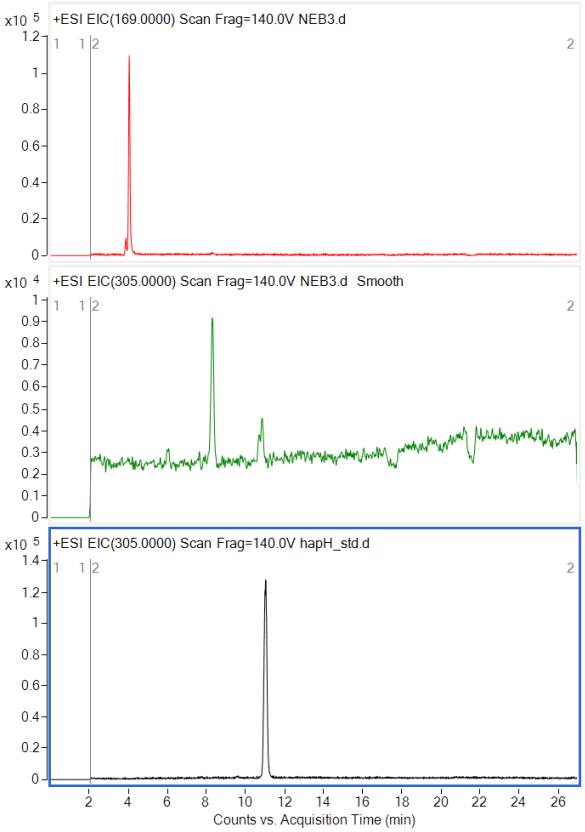


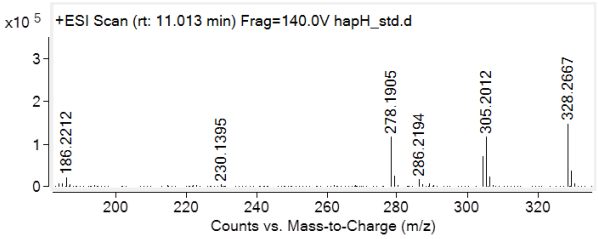
C. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel) and 12-*epi*-hapalindole U **2.4** (*m/z* 305) (middle panel) compared with the standard **2.4** (bottom panel) for the *in vitro* TT-assay in the presence of FamD2 and FamC1 plasmids (for Figure 2.3 C). The MS of the product is shown.



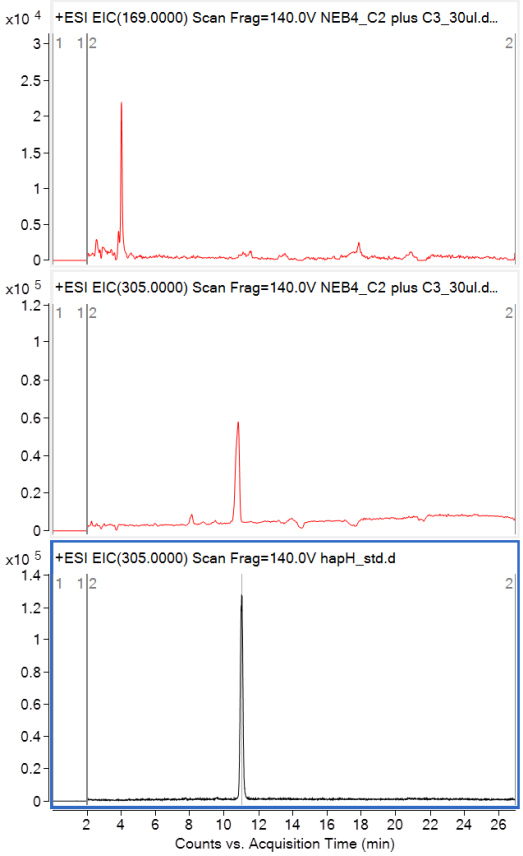


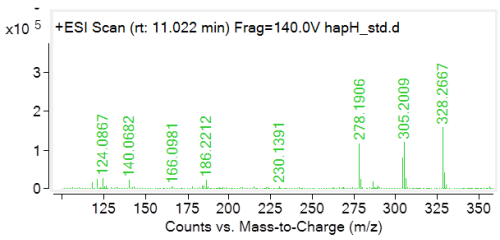
D. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel) and hapalindole H **2.6** (*m/z* 305) (middle panel) compared with the standard **2.6** (bottom panel) for the *in vitro* TT-assay in the presence of FamD2 and FamC2 and FamC3 plasmids in the presence of lower 5 mM calcium. The MS of the product is shown.



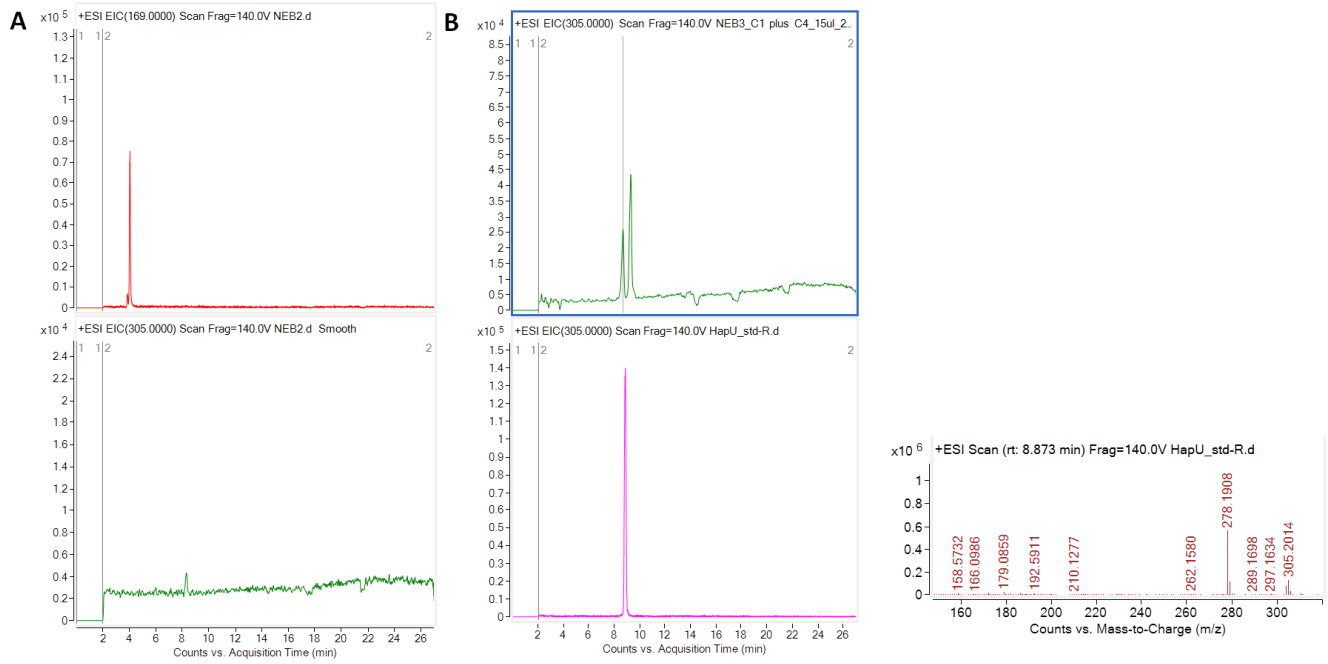


E. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel) and hapalindole H **2.6** (*m/z* 305) (middle panel) compared with the standard **2.6** (bottom panel) for the *in vitro* TT-assay in the presence of FamD2 and FamC2 and FamC3 plasmids in the presence of higher 20 mM calcium (for Figure 2.3 E). The MS of the product is shown.

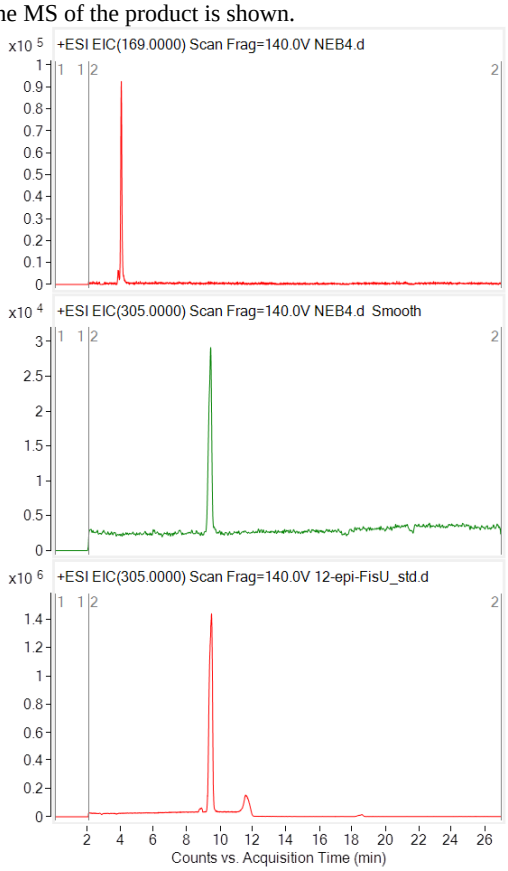




F. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel A) and hapalindole U **2.5** (*m/z* 305) (bottom panel **A** and top panel **B**) compared with the standard **2.5** (bottom panel B) for the *in vitro* TT-assay in the presence of FamD2 and FamC1 and FamC4 plasmids in the presence of lower 5 mM (**A**) and higher 20 mM (**B**) calcium concentration. Note that lower concentration of calcium did not show distinct product (for Figure 2.3 G). The MS of the product is shown.



G. EIC of *cis*-indole isonitrile **2.2** (*m/z* 169) (top panel A) and 12-epi-fischerindole U **2.8** (*m/z* 305) (bottom panel) compared with the standard **2.8** (bottom panel) for the *in vitro* TT-assay in the presence of FamD2 and FisC plasmids (for Figure 2.4 C). The MS of the product is shown.



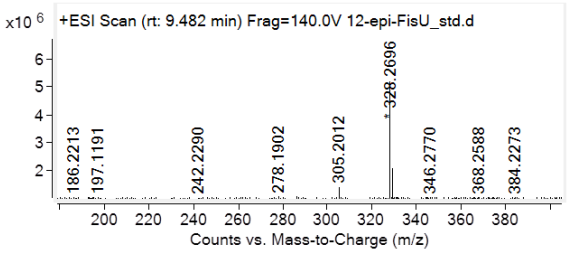
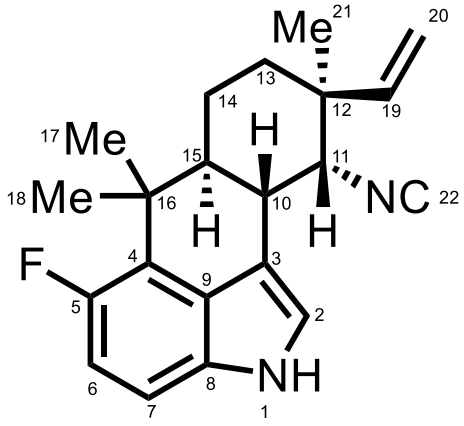


Table S2.1: Proteomics analysis of FamD2 expression in the TT-assay. Exogenous plasmids containing *famD2*, and the substrates **2.1** and **2.2** during *cis*-indole isonitrile-GPP intermediate **2.3** synthesis were added in the reaction. The sequence highlighted in yellow showed the coverage of the synthesized protein.

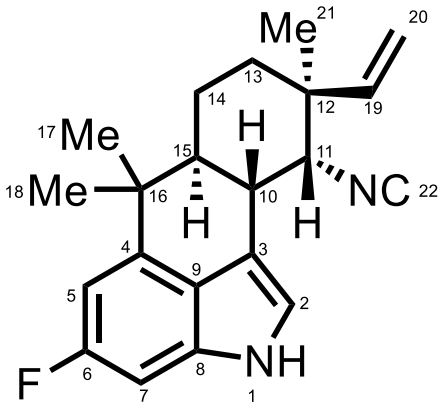
FamD2		
Total peptide	25	
Unique peptide	25	
Total PSM	906	
Coverage (%)	72.49	
Annotated Peptide Sequence	Sequence Positions	PSM FamD2
[-].MNDVNR.[I]	FamD2[1-6]	1
[R].IRTDIINVAK.[T]	FamD2 [7-16]	30
[R].TDIINVAK.[T]	FamD2 [9-16]	11
[K].TFGAEYSEK.[V]	FamD2 [17-25]	14
[K].VLDEVFQVFGEQFADNSFMIR.[T]	FamD2 [26-46]	155
[R].TSNKQPDK.[L]	FamD2 [47-54]	10
[R].TSNKQPDKLGCYFR.[Y]	FamD2 [47-60]	2
[K].QPDKLGCYFR.[Y]	FamD2 [51-60]	7
[K].LGCYFR.[Y]	FamD2 [55-60]	16
[R].YHEEDESQLGLAWDIAR.[K]	FamD2 [61-77]	335
[R].KSGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [78-112]	45
[K].SGLLSDQGRPVDQLIPEICETFPIMADGVDFDVK.[H]	FamD2 [79-112]	28
[K].IWQSIK.[G]	FamD2 [118-123]	12
[K].GVVPVQDAFK.[L]	FamD2 [124-133]	22
[K].LSLPASVTTHSDFLK.[N]	FamD2 [134-148]	42
[K].HHTSEYYK.[N]	FamD2 [178-185]	1
[R].LCFYLPFLNR.[E]	FamD2 [224-233]	5
[R].EAVPQNLLNPLLK.[K]	FamD2 [234-246]	42
[R].EAVPQNLLNPLLKK.[Y]	FamD2 [234-247]	14
[K].KYINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2 [247-273]	10
[K].YINEAPALVDNPGFILGWSFGPQGGK.[G]	FamD2 [248-273]	16
[K].VDVDYHGR.[T]	FamD2 [279-286]	3
[R].TVPLFMK.[V]	FamD2 [287-293]	64
[R].TVPLFMKVHSQPLPK.[A]	FamD2 [287-301]	1
[K].VHSQPLPK.[A]	FamD2 [294-301]	6
FamD2 MNDVNRIRTDIINVAKTFGAEYSEKVLDEVFQVFGEQFADNSFMIRTSNKQPDKLGCYFRYHEEDESQLGLAWDIARKSGLLSDQGRPVDQLIPEICETFPIMADGVDFDVKHGLAKIWQSIKGVVPVQDAFKLSLPASVTTHSDFLKNHHLDALYAFGIDYHHSSVNLYFDTYHPKHHTSEYYKNLLQDLQFQPPSDELLELLTNNGEIALTFNFASPRIERLCFYLPFLNREAVPQNLLNPLLKKYINEAPALVDNPGFILGWSFGPQGGKGTYTKVDVDYHGRTVPLFMKVHSQPLPKAADFALAQ, 309 residues		

Table S2.2: 5-fluoro-12-*epi*-hapalindole U (**2.15**) characterization
$[\alpha]_D^{25}$=+7.7 (c=0.07, CH_2Cl_2)
Calc. $[M+H]^+$ 323.19 Obsv. 323.1922



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.55, bs**	**2**	
2	**118.03**	**6.35, s**	**1**	**3, 4, 8, 9**
3	**113.75**			
4	**127.60**			
5	**161.65, d, (236.4)**			
6	**112.75, d, (29.2)**	**7.01, dd, (12.9, 4.4)**	**6**	**8, 9**
7	**109.71, d, (10.1)**	**6.69, dd, (3.0, 8.7)**	**7**	**4**
8	**130.96**			
9	**124.13**			
10	**34.58**	**2.90, d, (11.5)**	**15**	
11	**63.35**	**3.90, s**	**10**	
12	**39.73**			
13	**31.27**	**1.66, d, (3.7) & 1.44, m**	**14**	**15**
14	**21.17**	**1.48, d, (3.7) & 1.27, m**	**13,15**	**13, 21**
15	**43.98**	**1.98, td, (12.0, 3.4)**	**10,14**	
16	**37.27**			
17	**23.07**	**1.11, s**	**18**	**9, 15, 16, 18**
18	**25.05**	**1.63, s**	**17**	**9, 15, 16, 17**
19	**142.17**	**5.39, dd, (17.7, 11.0)**	**20**	**12**
20	**114.47**	**(cis) 4.91, d, (11.0) (trans), 4.84, d, (17.7)**	**19**	**12, 19**
21	**28.35**	**1.18, s**		**11, 12, 13, 19**
22	**160.84**			

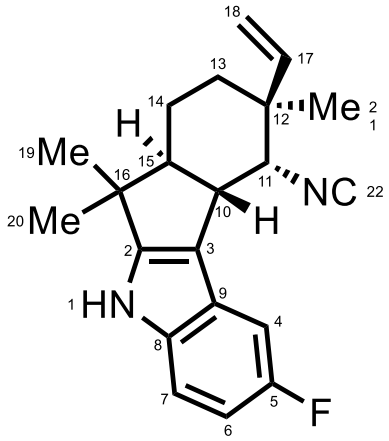
Table S2.3: 6-fluoro-12-*epi*-hapalindole U (**2.16**) characterization
$[\alpha]_D^{25}$=+12.9 (c=0.04, CH_2Cl_2)
Calc. $[M+H]^+$ 323.19 Obsv. 323.1920



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.51, bs**		
2	**116.62, d, (3.2)**	**6.29, s**		**3, 8, 9**
3	**113.12**			
4	**142.37, d, (8.7)**			
5	**102.96, d, (25.9)**	**6.93*, dd, (11.0, 1.8)**		**6,7,9**
6	**162.10, d, (236.5)**			
7	**95.36, d, (27.4)**	**6.69*, d, (9.8)**		**5,6,9**
8	**134.02**			
9	**122.36**			
10	**34.88**	**2.95, d, (11.5)**	**15**	**3,15**
11	**63.45**	**3.88, s**	**10**	**10,12,13,15**
12	**39.81**			
13	**31.08**	**1.66, m & 1.43, m**	**14**	**11,15**
14	**21.66**	**1.41, d, (3.7) & 1.26, m**	**13,15**	**19**
15	**43.57**	**1.93, td, (11.9, 3.5)**	**10,14**	
16	**37.44**			
17	**24.26**	**1.14, s**	**18**	**4, 15, 16, 18**
18	**24.96**	**0.89, s**	**17**	**4, 15, 16, 17, 19**
19	**142.55**	**5.40, dd, (17.7, 11.0)**	**20**	**11, 12, 13**
20	**114.49**	**(cis) 4.92, d, (11.0) (trans), 4.85, d, (17.7)**	**19**	**12,19**
21	**28.27**	**1.17, s**		**11, 12, 13, 19**
22	**160.88**			

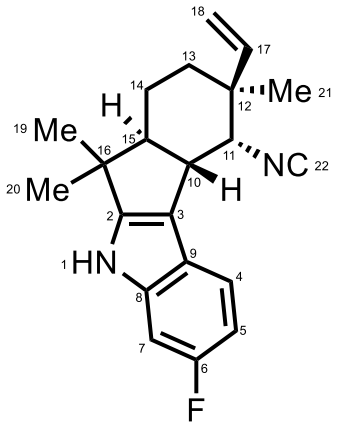
*-Protons observe same correlations, could be switched

Table S2.4: 5-fluoro-12-*epi*-fischerindole U (**2.17**) characterization
$[\alpha]_D^{25}$=+12.0 (c=0.15, CH_2Cl_2)
Calc. $[M+H]^+$ 323.19 Obsv. 323.1945



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.43, bs**		**2,3,8,9**
2	**153.90**			
3	**115.16**			
4	**103.93 (23.1)**	**7.27, dd (9.4,2.5)**	**6**	**3,5,6,8**
5	**159.42 (233.5)**			
6	**109.07 (26.3)**	**6.99, td, (9.0, 2.6)**	**7**	**4,5,8**
7	**112.55 (10.1)**	**6.81, dd, (8.8, 4.4)**	**6**	**4,5,9**
8	**136.53**			
9	**124.63 (10.1)**			
10	**42.37**	**2.88, m**	**11,15**	**3,9,14,15**
11	**62.21**	**3.95, d (3.0)**	**11**	**3,12,13,15,17,22**
12	**41.04**			
13	**31.82**	**1.55, m**	**14**	**11,12,14,15,17**
14	**20.81**	**1.34, ddd (13.1, 6.8, 4.0) & 1.23, m**	**13,15**	**12,13,15**
15	**55.21**	**2.36, ddd (13.3, 10.6, 3.3)**	**10, 14**	**10,14,16,19**
16	**40.18**			
17	**143.02**	**5.31, dd (17.6, 11.0)**	**18**	**11,13,12,21**
18	**114.16**	**(cis) 4.87, d (11.1) (trans) 4.84, d (17.7)**	**17**	**12,17,21**
19	**24.93**	**1.01, s**		**2,15,16,20**
20	**20.78**	**0.74, s**		**2,15,16,19**
21	**28.15**	**1.13, s**		**11,1213,17**
22	**161.39**			

Table S2.5: 6-fluoro-12-*epi*-fischerindole U (**2.18**) characterization
$[\alpha]_D^{25}$=+10.8 (c=0.16, CH_2Cl_2)
Calc. $[M+H]^+$ 323.19 Obsv. 323.1948



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		6.38, bs		
2	152.16			
3	114.88			
4	119.18, d, (10.1)	7.20, dd, (8.6, 5.3)	5	2
5	108.67, d, (24.2)	6.99, td, (9.1, 2.4)	4	3
6	161.97, d, (239)			
7	98.80, d, (25.9)	6.85, dd, (9.8, 2.3)		
8	136.10			
9	121.08			
10	42.51	2.89, m	11, 15	15
11	62.44	4.03, s	10	
12	41.03			
13	31.87	1.56, m	14	
14	20.82	1.35, m & 1.24, m	13	
15	55.21	2.36, ddd, (13.4, 10.6, 3.2)	10, 14	
16	40.25			
17	143.08	5.34, dd, (17.5, 11.0)	18	13
18	114.16	Cis 4.89, d, (11.0) Trans 4.87, d, (17.5)	17	12,17
19	25.07	1.02, s	20	2,15,16,20
20	20.70	0.74, s	19	2,15,16,19
21	28.17	1.15, s		11,12,13,17
22	161.17			

Table S2.6: Amino acid sequences of cyclases

FamC1
MKRKLIVAVVFLIFICLGINTPAHATSAVSIPINNAGFENPFMDVVDDYTIDTPPGWTTYDPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLSQNPGSGVAGFEQILDATLEPDTKYTLTVDVGNLAGTFKGLSFAGFPGYRVELLA
GDTVLAADHNNLFIKEGEFKTSTVTYTSTAKDLHLGQKLGIRLVNLLQDKFSGLDFDNVRLTTEPTET

FamC2
MKRNLIVAAIVLLIYICSGINTPANAAVTTSIPIKNPGFEEPILKVEGDYTIDAPPGWTTYNPNGLIPEKRTKWTSNNGVGH
VGPNYGQLFYNQQLPEGKNIGFVYLAQKTGSGIAGFEQTLDAVLEPNTSYKLIVDIGNFGGMFKGVSFAGFPGYRVELL
AGDTVLAADHNNLYIKDGEFKTSTVTFTSAANNPYLGQKLGIRLINLLQGKFSGLDFDNVRLITETVDT

FamC3
MKLKSIVAVVFLIFICLGINTPANATGAVSIPIKNAGFEDPFLEVKDYYTVNTPPGWSTYDPNGLIPEQPTVQTSYVGVTN
ATPSSAFYDQKVPEGRNMGSVYLAHEPGSGIAGLEQTLDTVLESNKNYTLLVDIGNSADGYKDISLADFPGYRVELLAG
DKVIAVDHNSVYIKEGEFKTSMIKFTAKPDSPYLGQKLGIRLINSLQTLSGNIDFDNVRLSVESAVI

FisC
MKRNFIIAAIVLLVYICFGISISANAASAVSIPIKNAGFEEPSLTVEDYYTIDTPPGWITYDPNGLVPAKRTRITSNNGVGYT
GPNSAYYNHKAPEGRNVAYVYLAQEIGSGIAGLEQTLDAVLKPNTKYTLTVDIGNSGGSFQGFPLDGFPGYRVELLAGD
TVLAADQNNLYIKEKDFKTTTVTFIATPESPYLGQHLGIRLINPLQGKFSGVDFDNVRLTAEPAET

FimC5
MKRNFIIAAIVLLVYIFSGINVFANAASAVCIPIKNAGFEEPILQIEDDYTIDTPPGWITYDPGGLVPAKRTRITSNNGVGYT
GSNSEFYNHKAPEGRNVAFVYLAQEIGSGIAGLEQTLDAVLKPNTKYTLTVDIGNSGGSFQGKTLDGFPGYRIELLAGDT
VLAADHNTLYIKEKDFKSTTVTFTATPESPYLGQHLGIRLINPLQGKFSGVDFDNVRLTAEPAET

Chapter 3
Structural diversification of hapalindole and fischerindole natural products via cascade biocatalysis

3.1 Abstract

Hapalindoles and related compounds (ambiguines, fischerindoles, welwitindolinones) are a diverse class of indole alkaloid natural products. They are typically isolated from the Stigonemataceae order of cyanobacteria and possess a broad scope of biological activities. Recently the biosynthetic pathway for assembly of these metabolites has been elucidated. In order to generate the core ring system, *L*-tryptophan is converted into the *cis*-indole isonitrile subunit before being prenylated with geranyl pyrophosphate at the C-3 position. A class of cyclases (Stig) catalyzes a three-step process including a Cope rearrangement, 6-*exo*-*trig* cyclization and electrophilic aromatic substitution to create a polycyclic core. Formation of the initial alkaloid is followed by diverse late-stage tailoring reactions mediated by additional biosynthetic enzymes to give rise to the wide array of structural variations observed in this compound class. In this chapter, we demonstrate the versatility and utility of the Fam prenyltransferase and Stig cyclases toward core structural diversification of this family of indole alkaloids. Through synthesis of *cis*-indole isonitrile subunit derivatives, and aided by protein engineering and computational analysis, we have employed cascade biocatalysis to generate a range of derivatives, and gained insights into the basis for substrate flexibility in this system.

3.2 Introduction

Hapalindoles are a large family of indole alkaloids that have been isolated from the cyanobacterial order Stigonematales.[1] Along with their related compounds, fischerindoles, ambiguines and welwitindolinones, there are at least 81 members isolated from over 18 cyanobacterial strains.[2] They have also been shown to have antimicrobial,[3-8] antimycotic,[7,9] anticancer[10-12] and immunomodulatory activity.[13] Hapalindole/fischerindole metabolites possess three distinguishing features: a polycyclic ring system, diverse stereochemical variations and the

late-stage introduction of a range of functional groups such as additional rings, halogens, hydroxyls, isothiocyanate, and others. Due to these unique structural features and diverse biological activities, a number of total syntheses have been devoted to this family of indole alkaloids.[14-47] However, these efforts have been hindered due to the highly functionalized ring system and variant stereo-chemical patterns. Recently, work by several groups has explored the biogenesis of these metabolites,[48-67] which has uncovered new prenyltransferases, cyclases, halogenases and other unique biosynthetic enzymes.

While each hapalindole subgroup has a characteristic ring connectivity and stereochemistry, current evidence indicates that they are all derived from a *cis*-indole isonitrile core combined with a geranyl monoterpene subunit. Through biosynthetic analysis, the mechanism for assembly of the hapalindole core was recently elucidated. *L*-Tryptophan is converted to *cis*-indole isonitrile precursor **3.1** by a three enzyme cascade[48] followed by geranylation at the C-3 position.[53] In the presence of a Stig cyclase(s), this 3-geranyl *cis*-indole isonitrile (3-GC) intermediate **3.2** undergoes a Cope rearrangement, 6-*exo*-*trig* cyclization and a terminal electrophilic aromatic substitution (EAS) at either the indole C-2 or C-4 position to afford the fischerindole or hapalindole core, respectively (Figure 3.1 A).[53] Subsequent late stage tailoring of these molecules provide further access to the ambiguines and welwitindolinones.

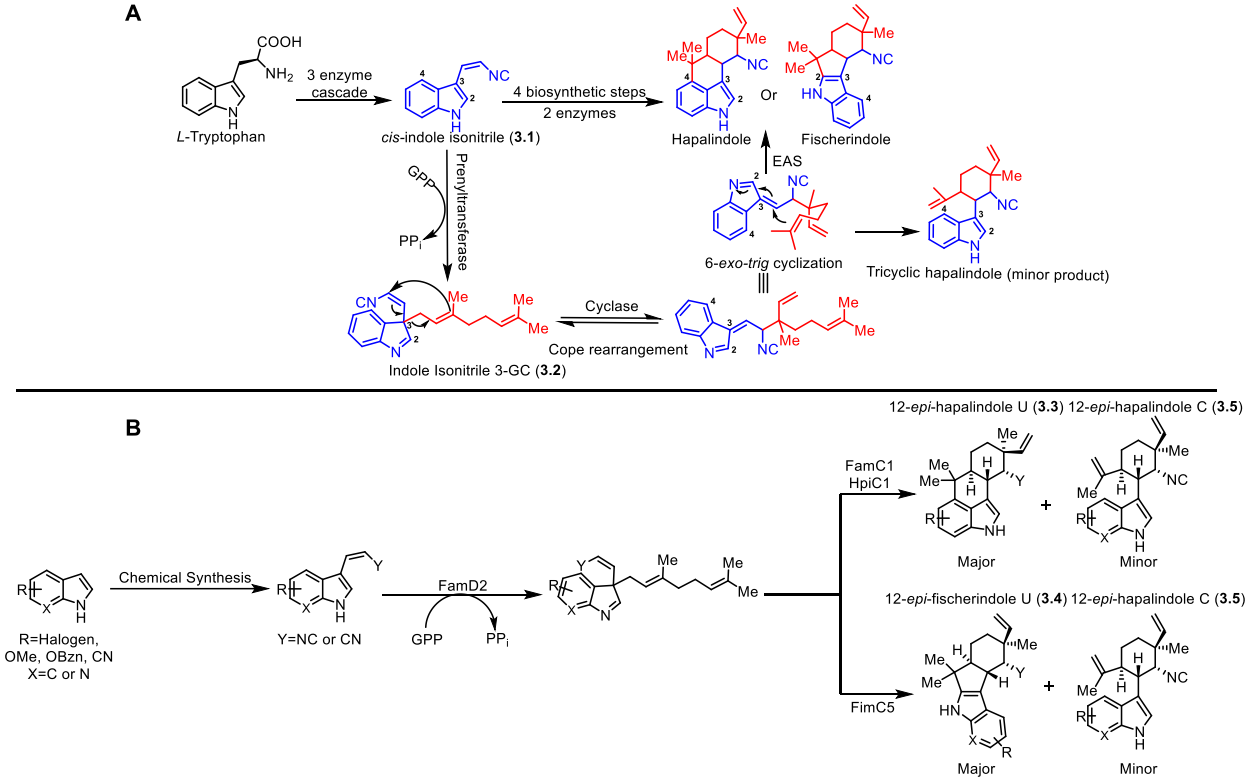


Figure 3.1: Overview of presented work. **A**: Proposed biosynthesis of hapalindole type molecules. *L*-Tryptophan is converted into the *cis*-indole isonitrile core (**3.1**) (in blue) via a 3 enzyme cascade followed by geranylation (in red) at C-3 position to afford the 3-GC intermediate (**3.2**). In the presence of the cyclase(s), **3.2** undergoes a Cope rearrangement, 6-*exo*-*trig* cyclization and electrophilic aromatic substitution (EAS) at the C-4 or C-2 position to afford the core hapalindole or fischerindole structure, respectively. **B**: The approach described herein includes chemical synthesis of unnatural *cis*-indole derivatives to show that the geranyltransferase (FamD2) and Stig cyclases can produce a range of unnatural 12-*epi*-hapalindole U (**3.3**), 12-*epi*-fischerindole U (**3.4**) and 12-*epi*-hapalindole C (**3.5**) compounds.

Recently, cyanobacterial-derived Stig cyclases have attracted interest for their ability to catalyze a multi-step core-forming cyclization cascade. Previous work has shown that the Stig cyclases exist in a dimeric state that may involve higher order oligomeric complexes to catalyze cyclization, including a terminal C-H functionalization reaction.[54,63] Depending on the components of this oligomeric complex, different regio- and stereochemical outcomes have been observed.[53,54,56,66,67] Structural and mutagenesis studies have revealed key residues responsible for core hapalindole and fischerindole formation alongside computational analysis that has examined the cyclase-mediated Cope rearrangement and terminal EAS.[55] Although gram scale total syntheses of the hapalindoles has been achieved on a few select metabolites,[28,33] further diversification is necessary for drug lead exploration. The Stig cyclases show potential as novel biocatalytic tools for developing unnatural hapalindole and fischerindole metabolites. Recently, we showed that fluorinated unnatural fischerindole and hapalindole derivatives could be produced using a microscale in vitro transcription/translation system with genes encoding prenyltransferase and Stig cyclase proteins.[68] To validate and explore this approach further, we demonstrate this chapter that key Stig cyclases can be employed as biocatalytic tools to produce numerous unnatural hapalindole and fischerindole compounds while also providing further mechanistic insights into the unusual cyclization cascade (Figure 3.1 A and 3.1 B).

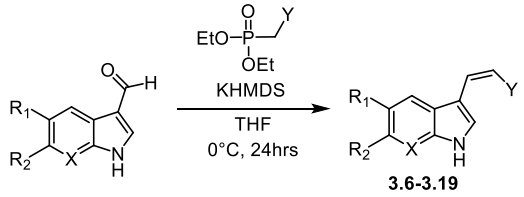
3.3 Results and discussion

3.3.1 Substrate synthesis

To pursue our analysis, chemical synthesis of unnatural *cis*-indole isonitrile derivatives was initiated. We synthesized these molecules via a Horner-Wadsworth-Emmons (HWE) olefination reaction from carboxyaldehyde indole derivatives based on the ability to produce unnatural derivatives in a facile manner (Table 3.1). While the HWE reaction is known to stereochemically favor *trans*-(*E*) product formation, we found that performing the reaction at 0°C yields a 50:50 ratio of the *cis*-(*Z*) and *trans* isomers. Using normal phase chromatography, these isomers are readily separated to obtain the desired Z-isomer product. The majority of these

derivatives were designed to have functional group modification at the indole C-5 and C-6 positions, which included halogens (**3.10-3.16**), methoxy groups (**3.8, 3.9**), benzyl groups (**3.6**) and cyano (**3.7**) functionality. Two of the isomers (**3.17**, **3.18**) were modified at the C-7 position of the indole ring by replacing with a nitrogen atom. For derivatives **3.16** and **3.18**, the carboxyaldehyde indole precursor was commercially unavailable. Thus, we obtained the indole precursor and added the carboxyaldehyde functionality via a Vilsmeier-Haack reaction that offered **3.20** and **3.21** (Experimentals), allowing us to obtain derivatives **3.16** and **3.18**.

Table 3.1: *Cis*-indole isonitrile derivatives generated in this study. Isolated % yield of the desired Z-isomer is reported.



3.6-3.19

Compound	Substituent	% Yield of Z-isomer
3.6	R_1=OBzn,R_2=H,X=C,Y=NC	11
3.7	R_1=CN,R_2=H,X=C,Y=NC	22
3.8	R_1=OMe,R_2=H,X=C,Y=NC	16
3.9	R_1=H,R_2=OMe,X=C,Y=NC	5
3.10	R_1=Cl,R_2=H,X=C,Y=NC	11
3.11	R_1=H,R_2=Cl,X=C,Y=NC	16
3.12	R_1=F,R_2=H,X=C,Y=NC	18
3.13	R_1=H,R_2=F,X=C,Y=NC	23
3.14	R_1=Br,R_2=H,X=C,Y=NC	18
3.15	R_1=H,R_2=Br,X=C,Y=NC	23
3.16	R_1=I,R_2=H,X=C,Y=NC	18
3.17	R_1=H,R_2=H,X=N,Y=NC	13
3.18	R_1=F,R_2=H,X=N,Y=NC	18
3.19	R_1=H,R_2=H,X=C,Y=CN	16

3.3.2 FamD2 prenyltransferase

With a suite of *cis*-indole isonitrile derivatives in hand, we assessed the flexibility of FamD2 (also known as AmbP1), the prenyltransferase from *Fischerella ambigua* UTEX 1903[48,53] for the first key transformation of **3.1**. Through initial analytical reactions, we observed geranylation of 13 of the 14 substrates with only the 5-cyano derivative **3.7** failing to be geranylated. This result was rationalized based on previous structural studies of FamD2 that revealed conformational flexibility in the active site in the presence of a Mg^{2+} cofactor.[61] In the case of **3.7**, we hypothesized that either steric hindrance or unfavorable electronics played a role in the lack of C-3 geranylation. Because substrate **3.16** (containing a large 5-iodo group) was able to be geranylated, we believed that steric hindrance was less likely to be a factor. Upon examination of the crystal structure of FamD2,[61] we hypothesized that the highly polar 5-cyano functionality is interacting with a specific residue or the backbone of FamD2, disrupting the derivative from a favorable position for geranylation. Because the 3-GC intermediate has been shown to undergo a 1,2-shift,[53] we sought a different method to analyze the efficiency of the

geranylation step. Using an HPLC assay, total turnover numbers (TTN) were calculated over a one hour reaction period at pH 10 to achieve maximum enzyme efficiency. Further cyclization reactions were conducted at pH 7 or 7.8 as cyclase activity is greater under more acidic conditions.[55,62] It is possible that FamD2 turnover is attenuated at a more neutral pH, but this was not tested. At pH 10, TTN values ranged from 710 to 1902, which highlights the versatility and efficiency of FamD2 as a biocatalyst over a range of unnatural substrates (Table 3.2 & Supplemental information Table S3.3-S3.16).

Table 3.2: TTN values for geranylation reaction with FamD2.[a]

Compound	TTN Value
3.1	1495
3.6	1615
3.7	N/A[b]
3.8	1481
3.9	1902
3.10	947
3.11	1547
3.12	1625
3.13	803
3.14	755
3.15	1174
3.16	710
3.17	1281
3.18	1366
3.19	1454

[a]Assay conditions: 1 μM FamD2, 2 mM substrate, 1.5 mM GPP, 5 mM $MgCl_2$, 50 mM glycine (pH 10.0), 100 μL, 37°C, 200 rpm, 1 hr.
[b]Compound **3.7** showed no conversion in assay

3.3.3 FamC1 cyclase

We next investigated the FamC1 cyclase from the *Fischerella ambigua* UTEX 1903 *fam* biosynthetic gene cluster. The homodimeric form of FamC1 produces 12-*epi*-hapalindole U[53] and was previously shown to accept fluorinated *cis*-indole isonitrile derivatives **3.12** and **3.13** to generate new hapalindole compounds **3.22** and **3.23**.[68] We sought to analyze and confirm its scope beyond fluorinated derivatives. Initially, few additional substrates showed conversion to cyclized product. Lowering the pH and concentration of FamD2 did not increase cyclase activity. As Ca^{2+} has been shown to be an important cyclase cofactor, we supplemented the reaction mixture to test its effect on FamC1 activity.[54-56,66] Upon the addition of 5 mM calcium chloride, we observed increased production of cyclized products in 8 of the 14 unnatural substrates. Based on HPLC analysis, we estimated the conversion values ranged from 10% to >99%. Scale up efforts of 2 mg of each substrate enabled structural characterization of these compounds. We confirmed that in all but one case, the corresponding 12-*epi*-hapaplindole U derivative was produced. By contrast, FamC1 catalyzed formation of tricyclic 12-*epi*-hapalindole C derivative **3.31** from azaindole compound **3.17**. Tricyclic hapalindoles are normally a minor product observed in enzymatic reactions from previously studied Stig cyclases (Table 3.3). We initially reasoned that the unnatural electronegative N-7 of the azaindole inhibited the terminal EAS

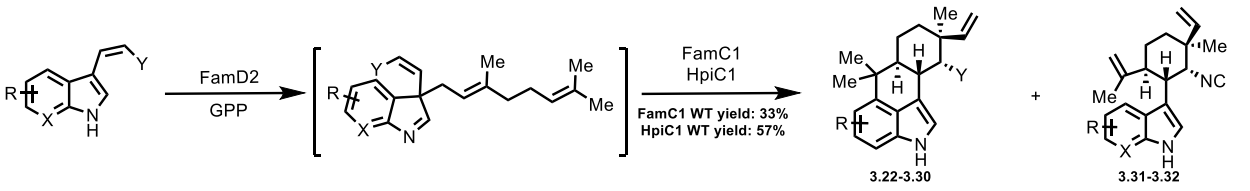
reaction at the C-4 position. However, C-4 tetracycle formation was observed with C-5-fluoro derivative **3.12**, which suggests the selectivity may be guided by skeletal variation of the N-7 position instead of the electronics of the indole ring.

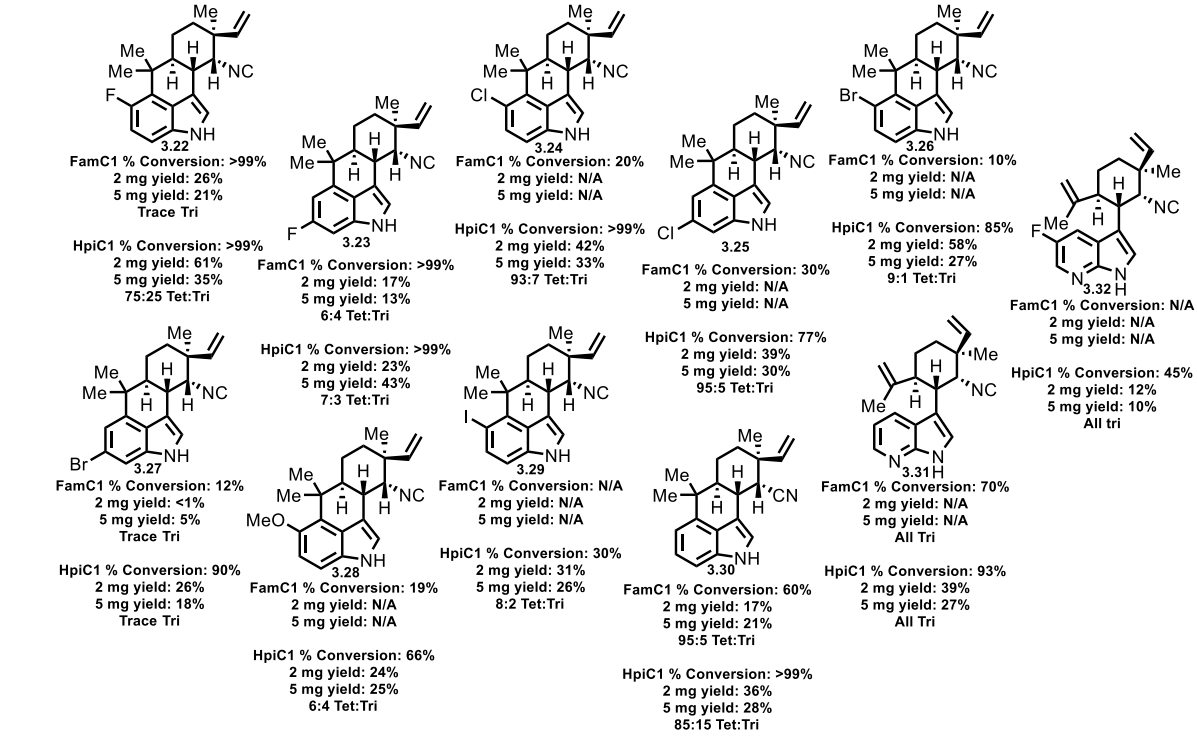
3.3.4 HpiC1 cyclase

While FamC1 demonstrated the ability to convert 8 of 14 substrates, we sought a cyclase with greater flexibility. We next examined HpiC1 from *Fischerella* sp. ATCC 43239, which shares 84% sequence homology with FamC1. Complexes consisting of only HpiC1 have been shown to produce 12-*epi*-hapalindole U and trace levels of 12-epi-hapalindole C.[54,55] Sequence comparison revealed only three active site sequence variations compared to FamC1.[55] Analytical reactions showed an increased scope as 10 of the 14 substrates were converted to hapalindole products with conversion values ranging from 30 to >99%. As with FamC1, the majority of the compounds were confirmed to be the corresponding 12-*epi*-hapalindole U derivatives. Similar to reactions with FamC1, we observed production of tricyclic 12-*epi*-hapalindole C derivatives **3.31** and **3.32** from *cis*-indole isonitrile substrates **3.17** and **3.18**, respectively (Table 3.3).

We hypothesized that the increased scope and reaction efficiency of HpiC1 could be attributed to the amino acid differences in the active site,[55] resulting in less steric hindrance to accommodate larger substituents. We investigated the three amino acid differences in the active site between HpiC1 and FamC1 using site directed mutagenesis to assess their role in substrate scope, which are Val51, Phe138 and Leu147 in HpiC1. The HpiC1_V51I mutant revealed no changes in the substrate scope. Although single mutants (HpiC1_F138L and HpiC1_L147F) failed to show noticeable changes in scope, the % conversion of substrates **3.8**, **3.14**, and **3.16** were reduced (Supporting Information Table S3.36). The attenuated substrate scope of the double mutant (HpiC1_F138L_L147F) closely matched WT FamC1 (Supplemental information Table S3.36). This mutagenesis study revealed that two of the three active site residue differences (138 and 147) were responsible for the change in substrate scope and conversion values between the two cyclases. Given that FamC1 exhibits low % conversions for **14** and **15**, and completely fails to convert **16**, we hypothesize that the two mutations present in HpiC1 (Phe138 and Leu147) reduce steric hindrance in the area of the active site that interacts with the C-5 and C-6 substituents of the indole, resulting in higher conversion efficiencies.

Table 3.3: Structures of 12-*epi*-hapalindole U and 12-*epi*-hapalindole C derivatives produced by FamC1 and HpiC1.[c]





[c]Percent conversions, isolated yield values and tetracyclic:tricyclic ratio (ratio estimated by NMR and/or HPLC) are shown below each derivative. N/A=Derivative was either not produced by FamC1 or was not screened further in this study due to enhanced versatility of HpiC1. HPLC conversion values determined after 4 hrs in 100 µL reactions. Isolated yield values from overnight reactions.

3.3.5 FimC5 cyclase

Following analysis of two hapalindole producing cyclases, we explored FimC5, a fischerindole producing cyclase derived from *Fischerella muscicola* UTEX 1829. This strain produces 12-*epi*-fischerindole U[54] and was used in our previous study to generate two fluorinated derivatives **3.33** and **3.37**.[68] Analytical reactions showed production of cyclized products for 8 of the 14 substrates tested with estimated conversion values ranging from 20% to >99%. The majority of the substrates were characterized as the corresponding 12-*epi*-fischerindole U derivatives. With substrate **3.17,** FimC5 catalyzed formation of a 50:50 mixture of two products: the previously observed tricyclic derivative **3.31** and new tetracyclic derivative

3.36. The structural assignment of **3.36** was also confirmed using X-ray crystallography (Figure 3.2). In the case of derivative **3.18**, we only observed production of the tricyclic derivative **32** (Table 3.4), with the C-5 fluorine appearing to limit reactivity at the C-2 position of the indole ring.

Previous analysis of FimC5 indicated that two active site amino acid residues (Phe101 and Ser138) play a key role in selectivity towards the indole C-2 or C-4 position for the terminal EAS.[54,55] Through site directed mutagenesis, we decided to explore this further with unnatural 3-GC derivatives. Three mutants of HpiC1 were produced to match the same residues identified in the FimC5 active site (Y101F, F138S and Y101F_F138S). Our results provide further support that these residues are critical for selectivity of the terminal reaction. For the Y101F and F138S mutants, we observed co-production of fischerindole and hapalindole products. However, upon screening the HpiC1 double mutant, we observed almost complete conversion to fischerindole metabolites. The F138S mutation also aided in uncovering the substrate selectivity differences between HpiC1 and FimC5 as product formation was either inhibited or completely abolished in select substrates. Exceptions to these results include compounds **3.13** and **3.17**, which displayed a 50:50 ratio of products for all mutants and compound **3.18**, which led to formation of the tricyclic hapalindole **3.32** for all HpiC1 mutants tested (Supplemental information Table S3.37).

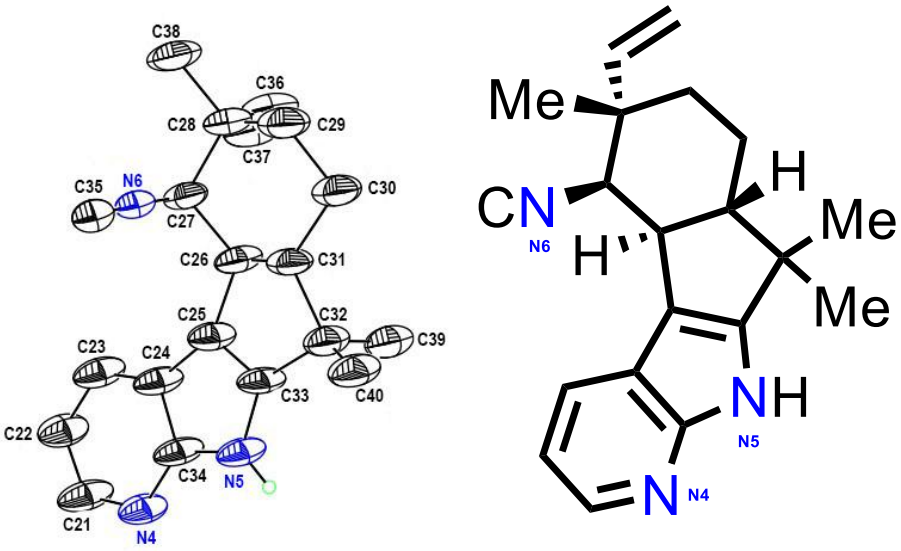
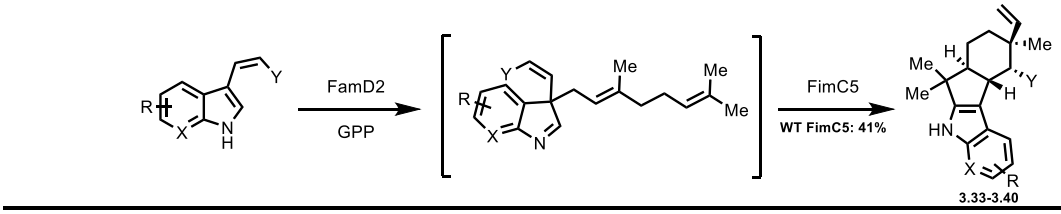
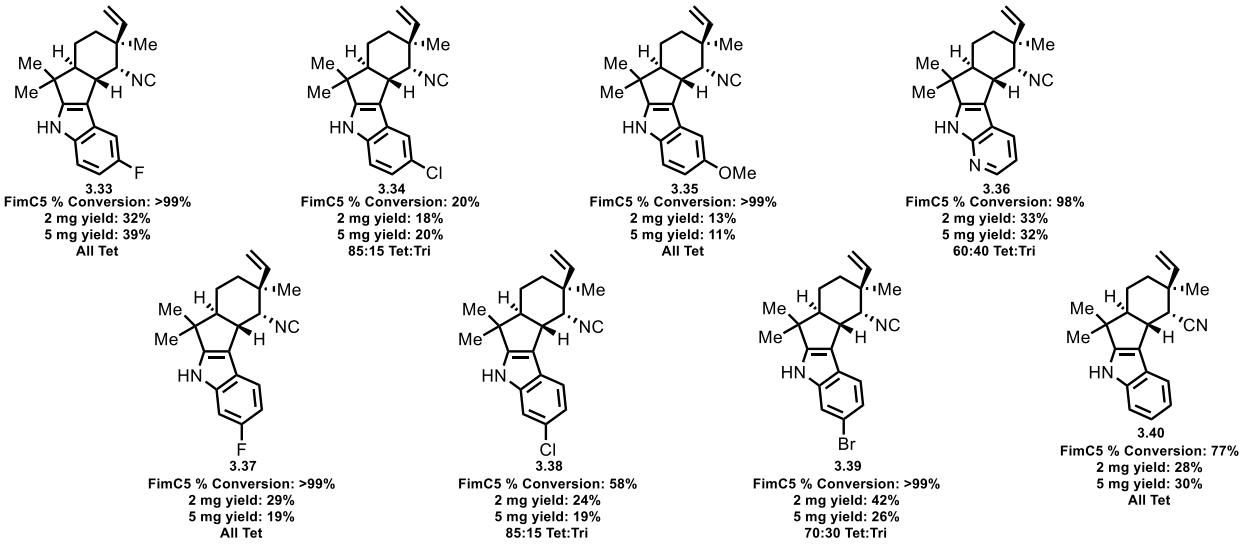


Figure 3.2: X-ray crystal structure of **3.36**. NMR assigned structure has been added to highlight the structures similarity. Nitrogen atoms have been highlighted in blue with same numbers to orient the reader.

Table 3.4: Structures of 12-*epi*-fischerindole U derivatives produced by FimC5.[d]





[d]Percent conversions, isolated yields and tetracyclic:tricyclic ratio (ratio estimated by NMR and/or HPLC) are shown below each derivative. HPLC conversion values determined after 4 hrs in 100 µL reactions. Isolated yield values from overnight reactions.

3.3.6 Discovery and isolation of nitrile containing compounds 3.30 and 3.40

Hapalindole type-indole alkaloids are noted for containing a rare isonitrile (-NC) moiety although nitrile (-CN) containing fischerindole and ambiguine molecules have been isolated previously.[1,5,69] Derivation of the nitrile functionality from isonitrile rearrangement has been suggested previously, and we decided to examine this hypothesis by screening the Stig cyclases using *cis*-indole *nitrile* derivative **3.19**, which was accepted by the prenyltransferase FamD2 at a TTN value comparable to the native isonitrile compound (Table 3.2). Cyclized product from wild-type HpiC1 was not initially observed with **3.19**, however, generation of two new products from each of the three HpiC1→FimC5 mutant cyclases in variant ratios was observed. Scale up and structure characterization of products derived from the HpiC1 Y101F mutant led to identification of the 12-*epi*-hapalindole U nitrile derivative **3.30** (we also observed small amounts of the corresponding tricyclic hapalindole derivative). Wild-type HpiC1 and FamC1 were then rescreened with optimized conditions (Experimentals) resulting in production of **3.30**

for both enzymes. To our knowledge, these are the first nitrile containing hapalindole molecules reported. With this result in hand, we returned to wild type FimC5 cyclase and, under the same optimized conditions, retested substrate **3.19**. We observed production of nitrile containing 12-*epi*-fischerindole U derivative **3.40** that was inadvertently overlooked in our initial analysis. Thus, our results demonstrated that the FamC1, HpiC1 and FimC5 cyclases have the ability to catalyze cyclization of the isonitrile and nitrile containing indole subunits. However, the nitrile substrate appeared to have lower conversion and isolated yield values (Tables 3.3 and 3.4) compared to the native isonitrile substrate, suggesting that the nitrile moiety affects turnover from the 3-GC intermediate to the terminal tetracyclic product. Regardless, this result supports that the nitrile functionality may come from early modifications of the intermediates but the source remains unknown.

3.3.7 Computational analysis for reactivity of substrates 3.17 and 3.18

To investigate why substrates **3.17** and **3.18** form tricyclic hapalindole compounds **3.31** and **3.32** with FamC1/HpiC1, respectively, while tetracyclic fischerindole **3.36** is formed with FimC5, we performed quantum mechanical density functional theory computations (Figure 3.3). For substrates **3.17** and **3.18**, as well as the parent indole intermediate **3.2**, we began with the tricyclic cationic intermediates **T** and considered two possible electrophilic aromatic substitutions: (1) reaction at C-4 to form the tetracyclic hapalindole scaffolds **H** and (2) reaction at C-2 to form the tetracyclic fischerindole scaffolds **F**. As shown in Figure 3.3, formation of hapalindole scaffolds is favored for the parent indole but disfavored for the azaindoles. The difference can be understood by consideration of resonance structures that stabilize the hapalindole scaffold. While conversion of indole cation **3.2-T** into **3.2-H** is exergonic by 1.9 kcal/mol and has a low free energy barrier of 7.4 kcal/mol, the conversion of azaindole cation **3.17-T** into **3.17-H** is endergonic by 10.0 kcal/mol and has a higher barrier of 14.7 kcal/mol. In particular, of the four resonance structures in addition to the iminium structure, the resonance structure shown places positive charge on the electronegative nitrogen, resulting in less stabilization than with the parent indole derivative where this positive charge is on a carbon. The conversion of fluorinated azaindole cation **3.18-T** into **3.18-H** is only slightly more favorable due to near-cancellation of resonance and inductive effects of F; this reaction is endergonic by 9.8 kcal/mol and has a barrier of 14.0 kcal/mol. Given the higher barriers to tetracyclic

hapalindole formation for **3.17** and **3.18**, it is not surprising that FamC1 and HpiC1 fail to catalyze this reaction; instead, deprotonation of cations **3.17-T** and **3.18-T** yield tricyclic compounds **3.31** and **3.32**, respectively. Unlike tetracyclic hapalindole formation, conversion of azaindoles into tetracyclic fischerindole scaffolds does not exhibit differential resonance effects compared to the parent indole because the azaindole nitrogen and fluorine atoms are not on positively charged positions in any of the four additional resonance structures that stabilize each tetracyclic fischerindole scaffold. Instead, tetracyclic fischerindole formation is influenced by smaller inductive effects. Thus, conversion of indole **3.2-T** into **3.2-F** is exergonic by 10.1 kcal/mol, and conversion of azaindole **3.17-T** into **3.17-F** (with an additional inductively withdrawing nitrogen atom) is exergonic by only 8.5 kcal/mol. Conversion of fluorinated azaindole **3.18-T** into **3.18-F**, which contains inductively withdrawing nitrogen and fluorine atoms is exergonic by only 6.5 kcal/mol. The free energy barriers, which are 1.7, 1.9, and 2.9 kcal/mol, respectively, follow the same trend in which addition of inductively withdrawing atoms raises the transition state energy. However, because these inductive effects are smaller than resonance effects, FimC5 can efficiently catalyze the conversion of **3.17** into **3.36**. Although it is important to note that these reactions take place in an enzyme active site, these computations addressing the innate reactivity of azaindoles reveal that tetracyclic hapalindole formation has a substantially higher free energy barrier than tetracyclic fischerindole formation. Accordingly, it is not surprising that FimC5 catalyzes tetracyclic fischerindole formation while FamC1 and HpiC1 form tricyclic products rather than tetracyclic hapalindoles. Evidently, the enzyme active sites do not overcome the innate difference in ease of tetracyclic product formation.

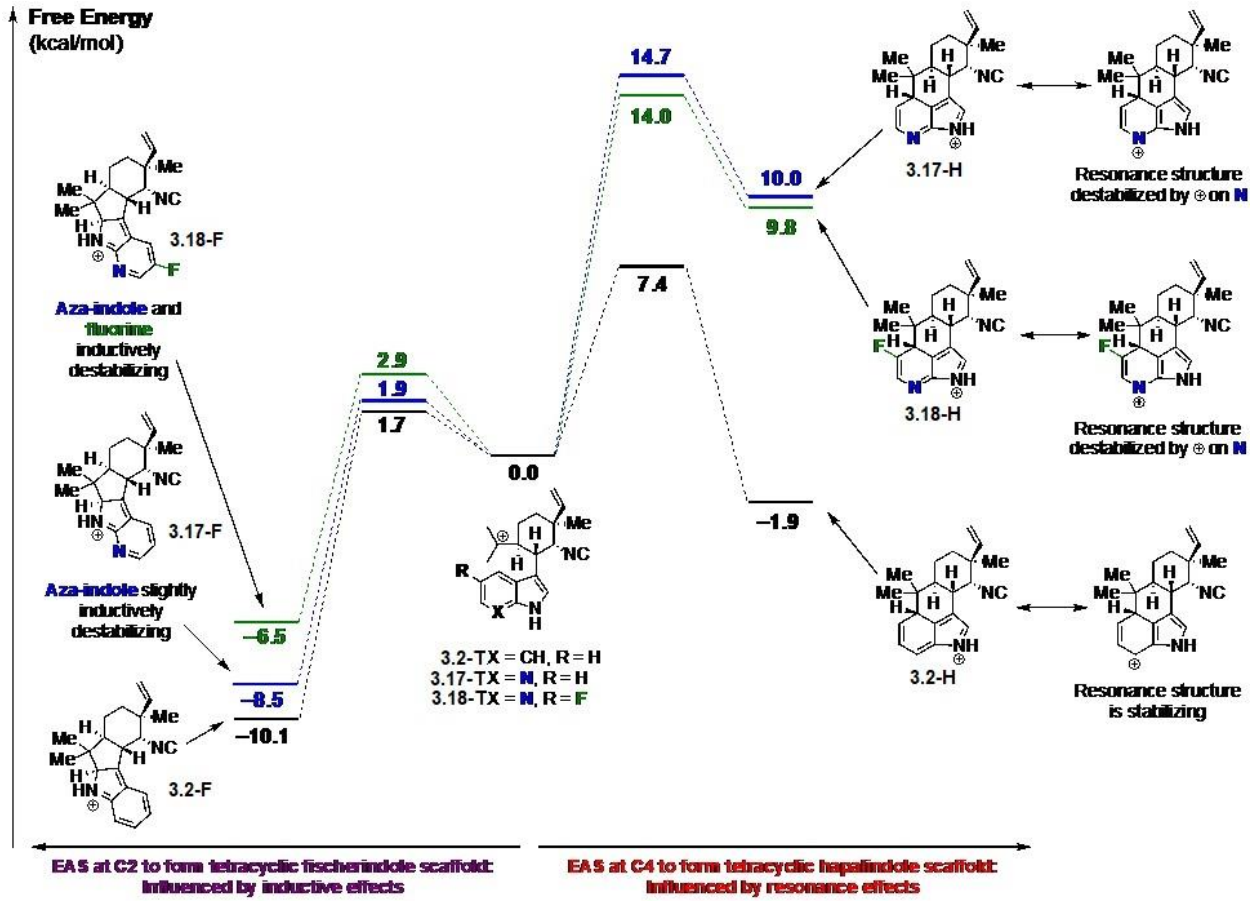


Figure 3.3: Quantum mechanical density functional theory computations. Energetics of tetracyclic hapalindole formation and tetracyclic fischerindole formation are compared. Starting points are from the cationic intermediates derived from **3.2**, **3.17**, and **3.18**. Tetracyclic fischerindole formation (left side) is only influenced by inductive effects, while tetracyclic hapalindole formation (right side) is also influenced by resonance effects.

3.3.8 Analysis of Stig cyclase biocatalytic ability

In this work, we explored the biocatalytic versatility of select Stig cyclases by assessing their substrate scope and ability to generate new derivatives. We analyzed HpiC1, select FamC1 and FimC5 reactions at both a 2 mg and 5 mg scale to assess how scalability may affect overall yield, which ranged from 10-60% (Table 3.3 & 3.4) for both reactions. The isolated yields of certain substrates (**3.22**, **3.26** for HpiC1 and **3.39** for FimC5) were coincident with, or exceeded, the levels of the native substrate, suggesting certain substituents enhance cyclase reactivity. In the majority of cases though, the isolated yields appeared to be significantly lower than the percent conversion values observed in the analytical scale reactions. We believe that through the filtration, workup and purification process, a significant amount of material was lost and optimizing isolation methods may be necessary to address this issue. At the 5 mg scale with

HpiC1, we observed that the tricyclic minor product was generated in a higher ratio compared to the 2 mg or analytical reactions. While these two compounds are separable, further scalability represents an objective for protein engineering to maximize tetracycle formation. However, FamC1 did show lower production of the tricyclic product, but appears to be less efficient and flexible compared to HpiC1.

3.4 Conclusion

Through the work showcased in this chapter, we have shown the biocatalytic potential of core biosynthetic enzymes to produce unnatural hapalindole and fischerindole derivatives. The three cyclases investigated for this work, FamC1, HpiC1, and FimC5 along with select mutants and FamD2 geranyltransferase have demonstrated their ability to accept numerous unnatural *cis*-indole isonitrile (and nitrile) derivatives. These new compounds are poised for investigation of biological activity directly or for additional semi-synthetic or biocatalytic modifications. While HpiC1 appears to be the most versatile Stig cyclase, there are numerous additional homologs and heteromeric combinations that can be screened to further probe the full scope of Stig cyclase biocatalytic versatility. This work further supports the growing opportunity to employ natural product biosynthetic enzymes for assembly of complex, bioactive small molecules and as a complement to synthetic chemistry approaches.

3.5 Experimentals

3.5.1 General

All NMR spectra were acquired on a Varian 400 and 600 MHz and Bruker 800 MHz spectrometers. Proton and carbon signals are reported in part per million (δ) using residual solvent signals as an internal standard. Analytical HPLC analysis was performed on a Shimadzu 2010 EV APCI spectrometer equipped with an LUNA C18 250 x 4.6 mm column, using a mobile phase gradient of 70-100% acetonitrile in water over 16 min. at 40°C and was monitored by UV absorption at 280 nm. LC-MS analysis was performed on a Agilent Infinity II TOF using an XBridge C18 2.1 x 150 mm column, using a mobile phase gradient of 70-100% acetonitrile in water over 12 min. Preparative-scale HPLC was performed on a Shimadzu 20-AT equipped with an LUNA C18 250 x 10 mm column for 2 mg reactions and an LUNA C8 250 x 21 mm column

for 5 mg reactions, using a mobile phase gradient of 50 or 60-100% acetonitrile in water over 60 min. Optical rotations were obtained using a Jasco P2000 polarimeter at 25°C.

Escherichia coli strain BL21(DE3) was used for protein expression. Plasmid pET28H8T[8g] was used for cloning and expression of N-truncated FamC1 and FimC5. Plasmid pET28a was used for cloning and expression of FamD2, HpiC1 and HpiC1 mutants. Isopropyl β-D-thiogalactopyranoside (IPTG) was used to induce expression; DNAse and lysozyme were purchased form Sigma-Aldrich. Ni-NTA agarose from Invitrogen was used to purify His-tag proteins.

All chemicals were purchased from Sigma-Aldrich, ACROS, and Combi-Blocks. Multiplicities are abbreviated as following: singlet (s), doublet (d), triplet (t), quartet (q), doublet-doublet (dd), triplet-doublet (td), doublet-doublet-doublet (ddd), triplet-doublet-doublet (tdd), and multiplet (m). Chemical abbreviations: Ethyl Acetate (EtOAc), Dichloromethane (DCM), Tetrahydrofuran (THF), Potassium bis(trimethylsilyl)amide (KHMDS), Acetic Acid (AcOH), Sodium Sulfate (Na_2SO_4), Diethyl Ether (Et_2O), Phosphorous Oxychloride ($POCl_3$), Sodium Hydroxide (NaOH), Acetonitrile (CH_3CN), Magnesium Chloride ($MgCl_2$), Calcium Chloride ($CaCl_2$), Sodium Chloride (NaCl),

3.5.2 Protein expression and purification

The expression and purification of proteins was performed as described.[54,68] Briefly, a single BL21(DE3) colony was inoculated in LB medium containing 50 μg/mL kanamycin and grown overnight at 37 °C shaking at 200 rpm. The main culture (1 L) was inoculated at the dilution of 1:100 in 2.8 L of Fernbach flask containing TB medium and the same concentration of antibiotic. The cells were grown (37 °C, 200 rpm) to an optical density (A_{600} nm) of 1.0. The culture flasks were chilled in ice, induced with IPTG (0.2 mM), and were further incubated (18 °C, 200 rpm) for 16 h. The cells were harvested (5000 rpm, 4 °C, 15 min), flash frozen, and stored at −80 °C until purification. The cell pellets were resuspended at 4 °C in the lysis buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol), containing 0.5 mg/mL of lysozyme, 1 mM PMSF and 1 mL of 2 mg/mL DNAse. The mixture was stirred for 30 min and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The cellular

debris was removed by centrifugation (65,000 x g, 4 °C, 35 min). The clarified lysate was loaded onto Ni-NTA agarose column equilibrated with lysis buffer. The column was washed with two column volume of wash buffer (10 mM HEPES, 300 mM NaCl, 0.2 mM TCEP, 10% glycerol, 20 mM imidazole) and the His-tagged protein was eluted with elution buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol, 300 mM imidazole). The fractions were pooled and dialyzed overnight or by using a PD10-desalting column (GE Healthcare) using storage buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol). The purified protein was analyzed by SDS-PAGE gel for purity, measured by Nanodrop using a calculated molar extinction coefficient for concentration, and flash-frozen in liquid nitrogen to store at −80 °C.

3.5.3 Analytical, TTN and scale up enzymatic reactions

Cis-indole isonitrile derivatives were synthesized as described below. For the initial assays, a 50µL reaction containing 10 µM FamD2, 15 µM cyclase, 1 mM substrate, 1 mM GPP, 5 mM $MgCl_2$, 50 mM of Tris pH 7.8 buffer and 5 mM $CaCl_2$, was incubated at 37°C for 4 hrs. The reaction was quenched with 3x volume of EtOAc twice. The organic layers were combined, dried and re-dissolved in 100 µL acetonitrile for LCMS and HPLC analysis. HPLC conversion values were determined by the area under the curve of the residual starting material and newly formed product. For later assays, a 100µL reaction containing 5µM FamD2, 20µM cyclase, 1 mM substrate, 1.5 mM GPP, 5 mM $MgCl_2$, 50 mM of Tris pH 7.0 buffer and 7.5 mM $CaCl_2$ was incubated, quenched and analyzed as previous. For FamD2 TTN assays, a 100 µL reaction containing 1 µM FamD2, 2 mM substrate, 1.5 mM GPP, 5 mM $MgCl_2$, and 50 mM Glycine pH 10.0 buffer was incubated at 37°C for 1 hr. The reaction was quenched and analyzed as previous. TTN values were determined by standard curve analysis for the starting material (Supporting Information Tables S3.3-S3.15). For the structure analysis and isolated yield values of the enzymatic products, the reactions were scaled up to 2 mg and 5 mg starting material (10 and 25 mL respectively) and incubated at 37°C overnight or until HPLC showed consumption of starting material. Products were extracted with EtOAc and purified by preparative HPLC as described in the general methods. All products were obtained as a white solid. The purified compounds were concentrated, dissolved in C_6D_6 and analyzed using a Varian 600 MHz NMR and Bruker 800 MHz NMR.

3.5.4 Chemical synthesis of *cis*-indole isonitrile derivatives

All derivatives were prepared using method previously described.[53,68] Briefly, to a 50 mL two-neck round-bottom flask purged with nitrogen at −78 °C (dry ice/acetone), diethyl (isocyanomethyl) phosphonate (0.37 mL, 2.26 mmol) (diethyl cyanomethyl phosphonate was used for production of **3.19**) was diluted with THF (5 mL). KHMDS (1 M THF, 2.60 mL, 2.60 mmol) was added dropwise, and the reaction was stirred at −78 °C for 15 min. To a separate 4 mL vial, indole-3-carboxaldehyde derivative (1.13 mmol) was dissolved in THF (5 mL), and the resulting solution was added dropwise to the KHMDS solution at −78 °C. The resulting mixture was stirred at 0 °C (cryocool) overnight or until TLC showed consumption of starting material. The resulting solution was quenched by the addition of AcOH (0.15 mL, 2.6 mmol) and concentrated. The resulting residue was diluted with EtOAc (20 mL), washed with 1 M aqueous potassium phosphate buffer (20 mL, pH 7), washed with brine, dried with Na_2SO_4, and concentrated to a residue. The residue was dissolved in EtOAc and purified by flash chromatography (24%–100% pentane/Et_2O, SiO_2) to afford the titled compound as reported below. Yields and spectral data reported below.

(Z)-5-(benzyloxy)-3-(2-isocyanovinyl)-1H-indole (3.6): Blue solid, 17mg, 11%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.17 (s, 2H), 5.90 (d, *J* = 8.9 Hz, 1H), 6.90 – 6.99 (m, 2H), 7.32 (t, *J* = 7.3 Hz, 1H), 7.37 – 7.47 (m, 4H), 7.51 (d, *J* = 7.5 Hz, 2H), 8.14 (d, *J* = 2.7 Hz, 1H), 10.80 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 169.52, 154.15, 137.99, 130.84, 128.32, 127.70, 127.58, 127.55, 127.32, 124.44, 113.46, 112.60, 109.49, 101.38, 70.17.

(Z)-3-(2-isocyanovinyl)-1H-indole-5-carbonitrile (3.7): Tan Solid, 25mg, 22%

^{1}H NMR (400 MHz, Acetone-d_6) δ 6.06 (d, *J* = 8.9 Hz, 1H), 7.06 (dt, *J* = 9.2, 4.7 Hz, 1H), 7.53 (dd, *J* = 8.5, 1.6 Hz, 1H), 7.71 (dd, *J* = 8.4, 0.8 Hz, 1H), 8.24 – 8.29 (m, 1H), 8.32 (s, 1H), 11.38 (s, 1H).

^{13}C NMR (101 MHz, acetone) δ 171.15, 138.20, 129.86, 127.72, 126.16, 124.72, 124.10, 120.80, 114.08, 111.07, 104.41

(Z)-3-(2-isocyanovinyl)-5-methoxy-1H-indole (3.8): Red Solid, 36mg, 16%

^{1}H NMR (400 MHz, Acetone-d6) δ 3.84 (s, 3H), 5.89 (d, J = 8.8 Hz, 1H), 6.87 (dd, J = 8.8, 2.4 Hz, 1H), 6.97 (d, J = 6.0 Hz, 1H), 7.30 (d, J = 2.4 Hz, 1H), 7.41 (d, J = 8.7 Hz, 1H), 8.07 – 8.15 (m, 1H), 10.77 (s, 1H).

^{13}C NMR (151 MHz, dmso) δ 169.82, 154.74, 130.64, 127.65, 127.49, 125.45, 124.09, 113.16, 113.00, 109.24, 100.54, 55.83.

(Z)-3-(2-isocyanovinyl)-6-methoxy-1H-indole (3.9): Tan Solid, 5mg, 5%

^{1}H NMR (599 MHz, Acetone-d_6) δ 3.82 (s, 3H), 5.91 (d, *J* = 8.8 Hz, 1H), 6.82 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.93 (dt, *J* = 9.8, 4.8 Hz, 1H), 7.04 (d, *J* = 2.3 Hz, 1H), 7.64 (d, *J* = 8.7 Hz, 1H), 8.03 (d, *J* = 2.3 Hz, 1H), 10.67 (s, 1H).

^{13}C NMR (151 MHz, dmso) δ 169.84, 156.73, 136.44, 125.96, 125.32, 124.46, 123.68, 119.32, 110.83, 109.34, 95.19, 55.67.

(Z)-5-chloro-3-(2-isocyanovinyl)-1H-indole (3.10): Tan Solid, 20mg, 11%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.98 (d, *J* = 8.9 Hz, 1H), 6.93 – 7.06 (m, 1H), 7.22 (dd, *J* = 8.6, 2.0 Hz, 1H), 7.55 (d, *J* = 8.6 Hz, 1H), 7.82 (d, *J* = 2.0 Hz, 1H), 8.22 (d, *J* = 2.0 Hz, 1H), 11.04 (s, 1H).

(Z)-6-chloro-3-(2-isocyanovinyl)-1H-indole (3.11): Yellow solid, 36mg, 16%

^{1}H NMR (599 MHz, Acetone-d_6) δ 5.96 (d, *J* = 8.9 Hz, 1H), 6.94 (dt, *J* = 9.5, 4.9 Hz, 1H), 7.16 (dd, *J* = 8.5, 1.9 Hz, 1H), 7.57 (d, *J* = 1.8 Hz, 1H), 7.75 (d, *J* = 8.5 Hz, 1H), 8.17 – 8.20 (m, 1H), 11.02 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 170.89, 137.02, 128.93, 128.64, 128.47, 126.68, 124.71, 121.72, 120.25, 112.65, 110.61.

(Z)-5-fluoro-3-(2-isocyanovinyl)-1H-indole (3.12): Tan Solid, 36.7mg, 18%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.95 (d, *J* = 8.9 Hz, 1H), 6.94 (p, *J* = 4.8 Hz, 1H), 7.03 (td, *J* = 9.1, 2.5 Hz, 1H), 7.52 (td, *J* = 9.9, 3.5 Hz, 2H), 8.23 (s, 1H), 10.97 (s, 1H).

(Z)-6-fluoro-3-(2-isocyanovinyl)-1H-indole (3.13): Tan Solid, 48mg, 23%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.97 (d, *J* = 8.9 Hz, 1H), 6.93 – 7.01 (m, 2H), 7.27 (dd, *J* = 9.7, 2.3 Hz, 1H), 7.77 (dd, *J* = 8.7, 5.2 Hz, 1H), 8.17 (s, 1H), 10.93 (s, 1H).

(Z)-5-bromo-3-(2-isocyanovinyl)-1H-indole (3.14): Red Solid, 51mg, 18%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.96 (d, *J* = 8.9 Hz, 1H), 6.97 (dt, *J* = 9.2, 4.7 Hz, 1H), 7.33 (dd, *J* = 8.6, 1.9 Hz, 1H), 7.49 (d, *J* = 8.6 Hz, 1H), 7.95 (d, *J* = 1.8 Hz, 1H), 8.20 (d, *J* = 2.2 Hz, 1H), 11.08 (s, 1H).

^{13}C NMR (101 MHz, acetone) δ 170.77, 135.27, 129.69, 128.97, 128.80, 126.21, 124.63, 121.61, 114.62, 114.27, 110.11

(Z)-6-bromo-3-(2-isocyanovinyl)-1H-indole (3.15): Red Solid, 63mg, 23%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.98 (d, *J* = 8.9 Hz, 1H), 6.85 – 7.00 (m, 1H), 7.29 (dd, *J* = 8.5, 1.8 Hz, 1H), 7.67 – 7.76 (m, 2H), 8.18 (d, *J* = 1.9 Hz, 1H), 10.99 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 169.98, 136.50, 127.61, 126.03, 123.70, 123.38, 119.69, 115.57, 114.74, 109.70, 104.77.

(Z)-5-iodo-3-(2-isocyanovinyl)-1H-indole (3.16): Red Solid, 60mg, 18%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.98 (d, *J* = 8.9 Hz, 1H), 7.00 (d, *J* = 9.0 Hz, 1H), 7.40 (d, *J* = 8.5 Hz, 1H), 7.51 (dd, *J* = 8.6, 1.7 Hz, 1H), 8.16 (d, *J* = 3.0 Hz, 2H), 11.03 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 169.93, 152.01, 134.82, 130.90, 129.54, 127.63, 127.06, 123.71, 114.15, 108.91, 83.50.

(Z)-3-(2-isocyanovinyl)-1H-pyrrolo[2,3-b]pyridine (3.17): White Solid, 26mg, 13%

^{1}H NMR (599 MHz, Acetone-d_6) δ 6.01 (d, *J* = 8.9 Hz, 1H), 6.97 (dt, *J* = 9.8, 5.1 Hz, 1H), 7.20 (dd, *J* = 7.9, 4.7 Hz, 1H), 8.20 (dd, *J* = 7.9, 1.6 Hz, 1H), 8.26 (s, 1H), 8.35 (dd, *J* = 4.6, 1.6 Hz, 1H), 11.38 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 170.07, 148.22, 144.26, 127.05, 126.63, 123.64, 118.94, 116.61, 108.38, 105.03.

(Z)-5-fluoro-3-(2-isocyanovinyl)-1H-pyrrolo[2,3-b]pyridine (3.18): White Solid, 38mg, 18%

^{1}H NMR (400 MHz, Acetone-d_6) δ 6.04 (d, *J* = 8.9 Hz, 1H), 6.95 (dt, *J* = 9.5, 5.0 Hz, 1H), 8.04 (dd, *J* = 9.3, 2.7 Hz, 1H), 8.26 (t, *J* = 2.2 Hz, 1H), 8.33 (d, *J* = 2.1 Hz, 1H), 11.44 (s, 1H).

^{13}C NMR (151 MHz, acetone) δ 170.20, 156.91, 155.31, 144.88, 132.67, 132.48, 129.23, 123.38, 119.29, 119.25, 112.50, 112.35, 108.61.

(Z)-3-(1H-indol-3-yl)acrylonitrile (3.19): Yellow solid, 30mg, 16%

^{1}H NMR (400 MHz, Acetone-d_6) δ 5.37 (dd, *J* = 11.8, 1.1 Hz, 1H), 7.14 – 7.31 (m, 2H), 7.50 – 7.60 (m, 1H), 7.67 (d, *J* = 11.7 Hz, 1H), 7.77 – 7.86 (m, 1H), 8.37 (s, 1H), 11.02 (s, 1H).

3.5.5 Chemical synthesis of indole-3-carboxaldehyde derivatives

All derivatives were prepared using the same method. Briefly, in a 25mL round-bottom flask purged with nitrogen at 0 °C (ice-water bath), $POCl_3$ (1.38 mL, 14.7 mmol) was stirred in dry DMF (4 mL) for 20 minutes. To a separate 4 mL vial, the reactant indole compound (2.94 mmol) was dissolved in dry DMF (4 mL) and added to the $POCl_3$ solution at 0 °C. The reaction was slowly brought to room temperature and allowed to stir for 1 h or until TLC showed consumption of starting material. The reaction mixture was cooled to 0°C and quenched with

ice-water and 1M NaOH (5 mL each). The reaction mixture was allowed to stir at room temperature for 1 h or until TLC showed consumption of intermediate. The resulting solution was extracted with EtOAc (2x10 mL), washed with brine and dried with Na_2SO_4 and concentrated to a residue. The residue was dissolved in EtOAc and purified by flash chromatography (16%-100% Hexanes/EtOAc, SiO_2) to afford the titled compound as reported below. Yields and spectral data reported below.

5-iodo-1H-indole-3-carbaldehyde (3.20): Off-white solid, 259mg, 46%

^{1}H NMR (400 MHz, Acetone-d_6) δ 7.42 (dd, J = 8.5, 0.6 Hz, 1H), 7.58 (dd, J = 8.6, 1.8 Hz, 1H), 8.22 (s, 1H), 8.56 – 8.66 (m, 1H), 10.01 (s, 1H).

5-fluoro-1H-pyrrolo[2,3-b]pyridine-3-carbaldehyde (3.21): White Solid, 88mg, 18%

^{1}H NMR (400 MHz, Acetone-d_6) δ 8.20 (dd, J = 8.8, 2.8 Hz, 1H), 8.31 (dd, J = 2.8, 1.8 Hz, 1H), 8.48 (s, 1H), 10.01 (s, 1H), 11.69 (s, 1H).

3.5.6 Chemical synthesis of geranyl diphosphate, tri ammonium

Geranyl diphosphate was synthesized as described previously.[70] To a 10 mL round bottom flask purged with nitrogen, tris (tetrabutylammonium) hydrogen pyrophosphate (1.0 g, 1.04 mmol) was dissolved in CH_3CN (1.0 mL). Geranyl chloride (0.09 mL, 0.475 mmol) was added and the reaction mixture was stirred at room temperature for 2 h.

Dowex 50WX8 resin preparation: Dowex 50WX8 resin (20 g, hydrogen form) was washed with half saturated aqueous ammonium chloride (5x50 mL) and water (5x50 mL) until the pH of the supernatant equaled 5. The slurry was rinsed twice with ion exchange buffer (2% isopropanol in 25 mM aqueous ammonium bicarbonate) and loaded into a flash column and equilibrated with ion exchange buffer.

Purification: The reaction mixture was concentrated to afford an orange residue which was diluted with ion exchange buffer. The crude mixture was chromatographed with two column volumes of ion exchange buffer (75 mL). The fractions were combined and concentrated by rotary evaporation, flash frozen and lyophilized for 2 d. The resulting white powder was diluted with 0.1 M ammonium bicarbonate (4 mL) and 50% isopropanol/CH_3CN (10 mL), vortexed for 30 seconds and centrifuged (2,000 rpm, rt, 5 min). The organic layer was extracted and the residual 0.5 mL of yellow liquid was diluted with 50% isopropanol/CH_3CN and the dilution/vortex/centrifugation process was repeated twice. The combined organic layers were

concentrated to afford a while solid. The white solid was taken up in 50% isopropanol:25% CH_3CN:25% 0.1 M aqueous ammonium bicarbonate and chromatographed with cellulose. The resulting fractions were combined and lyophilized affording the title compound as a white powder (138 mg, 80.3%).

^{1}H NMR (400 MHz, D_2O/ND_4OD) δ 1.92 (d, *J* = 1.3 Hz, 3H), 1.98 (s, 3H), 2.01 (d, *J* = 1.3 Hz, 3H), 2.39 (d, *J* = 6.5 Hz, 2H), 2.41 – 2.49 (m, 2H), 5.45 – 5.53 (m, 1H), 5.74 (dt, *J* = 6.1, 3.9 Hz, 1H).

^{13}C NMR (151 MHz, D_2O/ND_4OD) δ 142.46, 133.67, 124.40, 120.46, 62.61, 39.10, 25.92, 25.16, 17.26, 15.89.

^{31}P NMR (162 MHz, D_2O/ND_4OD) δ -9.93 (d, *J* = 21.6 Hz), -6.08 (d, *J* = 21.6 Hz).

3.5.7 Structure determination

Colorless plates of **3.36** were grown from a MeOH/Et_2O solution of the compound at 4 °C. A crystal of dimensions 0.06 x 0.04 x 0.02 mm was mounted on a Rigaku AFC10K Saturn 944+ CCD-based X-ray diffractometer equipped with a low temperature device and Micromax-007HF Cu-target micro-focus rotating anode (λ = 1.54187 A) operated at 1.2 kW power (40 kV, 30 mA). The X-ray intensities were measured at 85(1) K with the detector placed at a distance 42.00 mm from the crystal. A total of 2028 images were collected with an oscillation width of 1.0° in ω. The exposure times were 5 sec. for the low angle images, 45 sec. for high angle. Rigaku d*trek images were exported to CrysAlisPro for processing and corrected for absorption. The integration of the data yielded a total of 61153 reflections to a maximum 2θ value of 139.83° of which 7384 were independent and 5502 were greater than 2σ(I). The final cell constants (Table S36) were based on the xyz centroids of 8462 reflections above 10σ(I). Analysis of the data showed negligible decay during data collection. The structure was solved and refined with the Bruker SHELXTL (version 2018/3) software package, using the space group P2(1)2(1)2(1) with Z = 4 for the formula $C_{44}H_{56}N_6O$. All non-hydrogen atoms were refined anisotropically with the hydrogen atoms placed in a combination of idealized and refined positions. Full matrix least-squares refinement based on F^2 converged at R1 = 0.0885 and wR2 = 0.2349 [based on I > 2σ(I)], R1 = 0.1102 and wR2 = 0.2582 for all data.

3.6 References

1. Walton, K.; Berry, J.P. Indole Alkaloids of the Stigonematales (Cyanophyta): Chemical Diversity, Biosynthesis and Biological Activity. *Mar. Drugs* **2016**, *14* (4), 73. https://www.mdpi.com/1660-3397/14/4/73.
2. 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;* Academic Press: New York, **2014**; Vol. 73, p 65-160. http://www.sciencedirect.com/science/article/pii/B9780124115651000020.
3. Shunyan, M.; Krunic, A.; Chlipala, G.; Orjala, J. Antimicrobial Ambiguine Isonitriles from the Cyanobacterium *Fischerella ambigua*. *J. Nat. Prod.* **2009**, *72* (5), 894-899. https://pubs.acs.org/doi/10.1021/np800751j.
4. Shunyan, M.; Krunic, A.; Santarsiero, B.D.; Orjala, J. Hapalindole-related alkaloids from the cultured cyanobacterium *Fischerella ambigua*. *Phytochemistry* **2010**, *71* (17-18), 2116-2123. https://linkinghub.elsevier.com/retrieve/pii/S0031942210003456.
5. Kim, H.; Lantivit, 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. http://www.sciencedirect.com/science/article/pii/S0968089612005081.
6. Raveh, A.; Carmeli, S. Antimicrobial Ambiguines from the Cyanobacterium *Fischerella* sp. Collected in Israel. *J. Nat. Prod.* **2007**, *70* (2), 196-201. https://pubs.acs.org/doi/10.1021/np060495r.
7. 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.; Swartzendruber, J.K.; Deeter, J.B. Hapalindoles, antibacterial and antimycotic alkaloids from the cyanophyte *Hapalosiphon fontinalis*. *J. Org. Chem.* **1987**, *52* (6), 1036-1043. http://pubs.acs.org/doi/abs/10.1021/jo00382a012.
8. 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. App. Phycol.* **2006**, *18* (1), 33-39. http://link.springer.com/10.1007/s10811-005-9011-9.
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. https://doi.org/10.1021/jo00029a014.
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. https://doi.org/10.1021/ja00101a015.
11. Smith, C.D.; Zilfou, J.T.; Stratmann, K., Patterson, G.M.L.; Moore, R.E. Welwitindolinone analogues that reverse P-glycoprotein-mediated multiple drug resistance. *Mol. Pharmacol.* **1995**, *47* (2), 241-247. http://molpharm.aspetjournals.org/content/47/2/241.

12. Zhang, X.; Smith, C.D. Microtubule effects of welwistatin, a cyanobacterial indolinone that circumvents multiple drug resistance. *Mol. Pharmacol.* **1996**, *49* (2), 288-294. http://molpharm.aspetjournals.org/content/49/2/288.
13. Chilczuk, T.; Steinborn, C.; Breinlinger, S.; Zimmermann-Klemd, A.M.; Huber, R.; Enke, H.; Enke, D.; Niedermeyer, T.H.J.; Grundemann, C. Hapalindoles from the Cyanobacterium *Hapalosiphon* sp. Inhibit T Cell Proliferation. *Planta Med.* **2020**, *86* (2), 96-103. http://www.thieme-connect.de/DOI/DOI?10.1055/a-1045-5178.
14. Dethe, D.H.; Das, S.; Kumar, V.B.; Mir, N.A. Enantiospecific Total Syntheses of (+)-Hapalindole H and (−)-12-*epi*-Hapalindole U. *Eur. J. Chem.* **2018**, *24* (36), 8980-8984. https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.201800970.
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. https://pubs.acs.org/doi/10.1021/ja806981k.
16. 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. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010148/.
17. Connon, R.; Guiry, P.J. Recent advances in the development of one-pot/multistep syntheses of 3,4-annulated indoles. *Tetrahedron Lett.* **2020**, *61* (14), 151696. https://www.sciencedirect.com/science/article/pii/S0040403920301192.
18. Maimone, T.J.; Ishihara, Y.; Baran, P.S. Scalable total syntheses of (−)-hapalindole U and (+)-ambiguine H. *Tetrahedron* **2015**, *71* (22), 3652-3665. http://www.sciencedirect.com/science/article/pii/S0040402014015683.
19. Fukuyama, T.; Chen, X. Stereocontrolled Synthesis of (-)-Hapalindole G. *J. Am. Chem. Soc.* **1994**, *116* (7), 3125-3126. http://pubs.acs.org/doi/abs/10.1021/ja00086a053.
20. Sahu, S.; Das, B.; Maji, M.S. Stereodivergent Total Synthesis of Hapalindoles, Fischerindoles, Hapalonamide H, and Ambiguine H Alkaloids by Developing a Biomimetic, Redox-Neutral, Cascade Prins-Type Cyclization. *Org. Lett.* **2018**, *20* (20), 6485-6489. https://pubs.acs.org/doi/10.1021/acs.orglett.8b02804.
21. Vaillancourt, V.; Albizati, K.F. Synthesis and absolute configuration of (+)-hapalindole Q. *J. Am. Chem. Soc.* **1993**, *115* (9), 3499-3502. http://pubs.acs.org/doi/abs/10.1021/ja00062a013.
22. 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. https://pubs.acs.org/doi/10.1021/ja036191y.
23. Kinsman, A.C.; Kerr, M.A. Total Synthesis of (±)-Hapalindole Q. *Org. Lett.* **2001**, *3* (20), 3189-3191. https://pubs.acs.org/doi/10.1021/ol0165138.
24. Bhat, V.; Allan, K.M.; Rawal, V.H. Total Synthesis of *N* -Methylwelwitindolinone D Isonitrile. *J. Am. Chem. Soc.* **2011**, *133* (15), 5798-5801. https://pubs.acs.org/doi/10.1021/ja201834u.
25. Gademann, K.; Bonazzi, S. Total Synthesis of Complex Cyanobacterial Alkaloids without Using Protecting Groups. *Angew. Chem. Int. Ed.* **2007**, *46* (30), 5656-5658. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200701881.
26. 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. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201406626.

27. Rafferty, R.J.; Williams, R.M. Total Synthesis of Hapalindoles J and U. *J. Org. Chem.* **2012**, *77* (1), 519-524. http://pubs.acs.org/doi/10.1021/jo202139k.
28. 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. https://www.nature.com/articles/nature05569.
29. Johnson, R.E.; Ree, H.; Hartmann, M.; Lang, L.; Sawano, S.; Sarpong, R. Total Synthesis of Pentacyclic (−)-Ambiguine P Using Sequential Indole Functionalizations. *J. Am. Chem. Soc.* **2019**, *141* (6), 2233-2237. https://pubs.acs.org/doi/10.1021/jacs.8b13388.
30. Xu, J.; Rawal, V.H. Total Synthesis of (−)-Ambiguine P. *J. Am. Chem. Soc.* **2019**, *141* (12), 4820-4823. https://doi.org/10.1021/jacs.9b01739.
31. 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. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201100957.
32. 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. https://doi.org/10.1021/ja057640s.
33. 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. https://pubs.acs.org/doi/10.1021/ja047874w.
34. Muratake, H.; Natsume, M. Total synthesis of marine alkaloids (±)-hapalindoles J and M. *Tetrahedron Lett.* **1989**, *30* (14), 1815-1818. http://www.sciencedirect.com/science/article/pii/S004040390099587X
35. 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. http://www.sciencedirect.com/science/article/pii/S0040402001960053
36. 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. http://www.sciencedirect.com/science/article/pii/S0040402001960053
37. Bhat, V.; Allan, K.M.; Rawal, V.H. Total Synthesis of *N*-Methylwelwitindolinone D Isonitrile. *J. Am. Chem. Soc.* **2011**, *133* (15), 5798-5801. https://pubs.acs.org/doi/10.1021/ja201834u.
38. 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. https://doi.org/10.1021/ja5087672.
39. 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. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201307464.
40. 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. https://doi.org/10.1021/ja210837b.

41. Huters, A. D.; Styduhar, E. D.; Garg, N. K. Total Syntheses of the Elusive Welwitindolinones with Bicyclo[4.3.1] Cores. *Angew. Chem. Int. Ed.* **2012**, *51* (16), 3758–3765. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201107567.
42. MacKay, J. A.; Bishop, R. L.; Rawal, V. H. Rapid Synthesis of the *N*-Methylwelwitindolinone Skeleton. *Org. Lett.* **2005**, *7* (16), 3421–3424. https://doi.org/10.1021/ol051043t.
43. Bhat, V.; Rawal, V. H. Stereocontrolled Synthesis of 20,21-Dihydro *N*-Methylwelwitindolinone B Isothiocyanate. *Chem. Commun.* **2011**, *47* (34), 9705–9707. https://pubs.rsc.org/en/content/articlelanding/2011/cc/c1cc13498a.
44. Bhat, V.; MacKay, J. A.; Rawal, V. H. Lessons Learned While Traversing the Welwitindolinone Alkaloids Obstacle Course. *Tetrahedron* **2011**, *67* (52), 10097–10104. http://www.sciencedirect.com/science/article/pii/S004040201101458X.
45. Bhat, V.; MacKay, J. A.; Rawal, V. H. Directed Oxidative Cyclizations to C2- or C4-Positions of Indole: Efficient Construction of the Bicyclo[4.3.1]Decane Core of Welwitindolinones. *Org. Lett.* **2011**, *13* (12), 3214–3217. https://doi.org/10.1021/ol201122f.
46. 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. https://doi.org/10.1021/ja210793x.
47. Hu, L.; Rawal, V.H. Total Synthesis of the Chlorinated Pentacyclic Indole Alkaloid (+)-Ambiguine G. J. Am. Chem. Soc. **2021**, *143* (29), 10872-10875. https://doi.org/10.1021/jacs.1c05762.
48. Hillwig, M.L.; Zhu, Q.; Liu, X. Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium *Fischerella ambigua*. *ACS Chem. Biol.* **2014**, *9* (2), 372-377. http://pubs.acs.org/doi/10.1021/cb400681n.
49. Chang, W.C.; Sanyal, D.; Huang, J.L.; Ittiamornkul, K.; Zhu, Q.; Liu, X. In Vitro Stepwise Reconstitution of Amino Acid Derived Vinyl Isocyanide Biosynthesis: Detection of an Elusive Intermediate. *Org. Lett.* **2017**, *19* (5), 1208-1211. http://pubs.acs.org/doi/10.1021/acs.orglett.7b00258.
50. Ittiamornkul, K.; Zhu, Q.; Gkotsi, D.S.; Smith, D.M.R.; Hillwig, M.L; Nightingale, N.; Goss, R.J.M.; Liu, X. Promiscuous indolyl vinyl isonitrile synthases in the biogenesis and diversification of hapalindole-type alkaloids. *Chem. Sci.* **2015**, *6* (12), 6836-6840. https://pubs.rsc.org/en/content/articlelanding/2015/sc/c5sc02919h.
51. Liu, X.; Hillwig, M.L.; Koharudin, L.M.I.; Gronenborn, A.M. Unified biogenesis of ambiguine, fischerindole, hapalindole and welwitindolinone: identification of a monogeranylated indolenine as a cryptic common biosynthetic intermediate by an unusual magnesium-dependent aromatic prenyltransferase. *Chem. Comm.* **2016**, *52* (8), 1737-1740. https://pubs.rsc.org/en/content/articlelanding/2016/cc/c5cc10060g.
52. 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. https://onlinelibrary.wiley.com/doi/abs/10.1002/cbic.201300794.

53. 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. https://doi.org/10.1021/jacs.5b10136.
54. 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. https://www.nature.com/articles/nchembio.2327.
55. Newmister, S.A.; Li, S.; Garcia-Borras, 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 cyclization in hapalindole biogenesis. *Nat. Chem. Biol.* **2018**, *14* (4), 345-351. https://www.nature.com/articles/s41589-018-0003-x.
56. Li, S.; Newmister, S.A.; Lowell, A.N.; Zi, J.; Chappell, C.R.; Yu, F.; Hohlman, R.M.; Orjala, J.; Williams, R.M.; Sherman, D.H. Control of Stereoselectivity in Diverse Hapalindole Metabolites is Mediated by Cofactor-Induced Combinatorial Pairing of Stig Cyclases. *Angew. Chem. Int. Ed.* **2020**, *59* (21), 8166-8172. 10.1002/anie.201913686.
57. 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 Microbiology* **2014**, *14*, 213. http://europepmc.org/article/med/25198896.
58. Yu, C.P.; Tang, Y.; Cha, L.; Milikisiyants, S.; Smirnova, T.I.; Smirnov, A.I.; Guo, Y.; Chang, W.C. Elucidating the Reaction Pathway of Decarboxylation-Assisted Olefination Catalyzed by a Mononuclear Non-Heme Iron Enzyme. *J. Am. Chem. Soc.* **2018**, *140* (45), 15190-15193. https://doi.org/10.1021/jacs.8b10077.
59. Bornemann, V.; Patterson, G.M.L.; Moore, R.E. Isonitrile biosynthesis in the cyanophyte *Hapalosiphon fontinalis*. *J. Am. Chem. Soc.* **1988**, *110* (7), 2339-2340. https://pubs.acs.org/doi/abs/10.1021/ja00215a075.
60. Harris, N.C.; Born, D.A.; Cai, W.; Huang, Y.; Martin, J.; Khalaf, R.; Drennan, C.L.; Zhang, W. Isonitrile Formation by a Non-Heme Iron(II)-Dependent Oxidase/Decarboxylase. *Angew. Chem. Int. Ed.* **2018**, *130* (31), 9855-9858. https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.201804307.
61. Awakawa, T.; Mori, T.; Nakashima, Y.; Zhai, R.; Wong, C.P.; Hillwig, M.L.; Liu, X.; Abe, I. Molecular Insight into the Mg2+-Dependent Allosteric Control of Indole Prenylation by Aromatic Prenyltransferase AmbP1. *Angew. Chem. Int. Ed.* **2018**, *57* (23), 6810-6813. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201800855.
62. Wang, J.; Chen, C.C.; Yang, Y.; Liu, W.; Ko, T.P.; Shang, N.; Hu, X.; Xie, Y.; Huang, J.W.; Zhang, Y.; Guo, R.T. Structural insight into a novel indole prenyltransferase in hapalindole-type alkaloid biosynthesis. *Biochem. Biophys. Res. Commun.* **2018**, *495* (2), 1782-1788. http://www.sciencedirect.com/science/article/pii/S0006291X17324312.
63. Chen, C.C.; Hu, X.; Tang, X.; Yang, Y.; Ko, T.P.; Gao, J.; Zheng, Y.; Huang, J.W.; Yu, Z.; Li, L.; Han, S.; Cai, N.; Zhang, Y.; Liu, W.; Guo, R.T. The Crystal Structure of a Class of Cyclases that Catalyze the Cope Rearrangement. *Angew. Chem. Int. Ed.* **2018**, *57* (46), 15060-15064. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201808231.
64. Wong, C.P.; Awakawa, T.; Nakashima, Y.; Mori, T.; Zhu, Q.; Liu, X.; Abe, I. Two Distinct Substrate Binding Modes for the Normal and Reverse Prenylation of Hapalindoles by the Prenyltransferase AmbP3. *Angew. Chem. Int. Ed.* **2018**, *57* (2), 560-563. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201710682.

65. Hillwig, M.L.; Liu, X. A new family of iron-dependent halogenases act on freestanding substrates. *Nat. Chem. Bio.* **2014**, *10* (11), 921-923. https://www.nature.com/articles/nchembio.1625.
66. 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.* **2017**, *56* (31), 9062-9066. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201703932.
67. Zhu, Q.; Liu, X. Molecular and genetic basis for early stage structural diversifications in hapalindole-type alkaloid biogenesis. *Chem. Comm.* **2017**, *53* (19), 2826-2829. https://pubs.rsc.org/en/content/articlelanding/2017/cc/c7cc00782e.
68. Khatri, Y.; Hohlman, R.M.; Mendoza, J.; Li, S.; Lowell, A.N.; Asahara, H.; Sherman, D.H. Multicomponent Microscale Biosynthesis of Unnatural Cyanobacterial Indole Alkaloids. *ACS Syn. Biol.* **2020**, *9* (6), 1349-1360. https://doi.org/10.1021/acssynbio.0c00038.
69. 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. https://pubs.acs.org/doi/10.1021/np9801561.
70. Davisson, V.J.; Woodside, A.B.; Neal, T.R.; Stremler, K.E.; Muehlbacher, M.; Poulter, C.D. Phosphorylation of isoprenoid alcohols. *J. Org. Chem.* **1986**, *51* (25), 4768-4779. https://doi.org/10.1021/jo00375a005.

Note: This work has been published as "Structural diversification of hapalindole and fischerindole natural products via cascade biocatalysis." Robert M. Hohlman, Sean A. Newmister, Jacob N. Sanders, Yogan Khatri, Shasha Li, Nikki R. Keramati, Andrew N. Lowell, K. N. Houk and David H. Sherman. *ACS Catal.* **2021**, *11* (8), 4670-4681.

3.7 Supplemental information

Table S3.1: Protein amino acid sequence

FamD2
MNDVNRIRTDIINVAKTFGAEYSEKVLDEVFQVFGEQFADNSFMIRTSNKQPDKLGCYFRYHEEDESQLGLAWDIARKS
GLLSDQGRPVDQLIPEICETFPIMADGVDFDVKHGLAKIWQSIKGVVPVQDAFKLSLPASVTTHSDFLKNHHLDALYAF
GIDYHHSSVNLYFDTYHPKHHTSEYYKNLLQDLQFQPPSDELLELLTNNGEIALTFNFASPRIERLCFYLPFLNREAVPQN
LLNPLLKKYINEAPALVDNPGFILGWSFGPQGGKGTYTKVDVDYHGRTVPLFMKVHSQPLPKAADFALAQ

FamC1
MKRKLIVAVVFLIFICLGINTPAHATSAVSIPINNAGFENPFMDVVDDYTIDTPPGWTTYDPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLSQNPGSGVAGFEQILDATLEPDTKYTLTVDVGNLAGTFKGLSFAGFPGYRVELLA
GDTVLAADHNNLFIKEGEFKTSTVTYTSTAKDLHLGQKLGIRLVNLLQDKFSGLDFDNVRLTTEPTET

FimC5
MKRNFIIAAIVLLVYIFSGINVFANAASAVCIPIKNAGFEEPILQIEDDYTIDTPPGWITYDPGGLVPAKRTRITSNNGVGYT
GSNSEFYNHKAPEGRNVAFVYLAQEIGSGIAGLEQTLDAVLKPNTKYTLTVDIGNSGGSFQGKTLDGFPGYRIELLAGDT
VLAADHNTLYIKEKDFKSTTVTFTATPESPYLGQHLGIRLINPLQGKFSGVDFDNVRLTAEPAET

HpiC1
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNFGGEFQKISLAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

HpiC1_Y101F
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGFIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNFGGEFQKISLAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

HpiC1_F138S
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNSGGEFQKISLAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

HpiC1_Y101FF138S
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGFIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNSGGEFQKISLAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

HpiC1_L147F
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNFGGEFQKISFAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

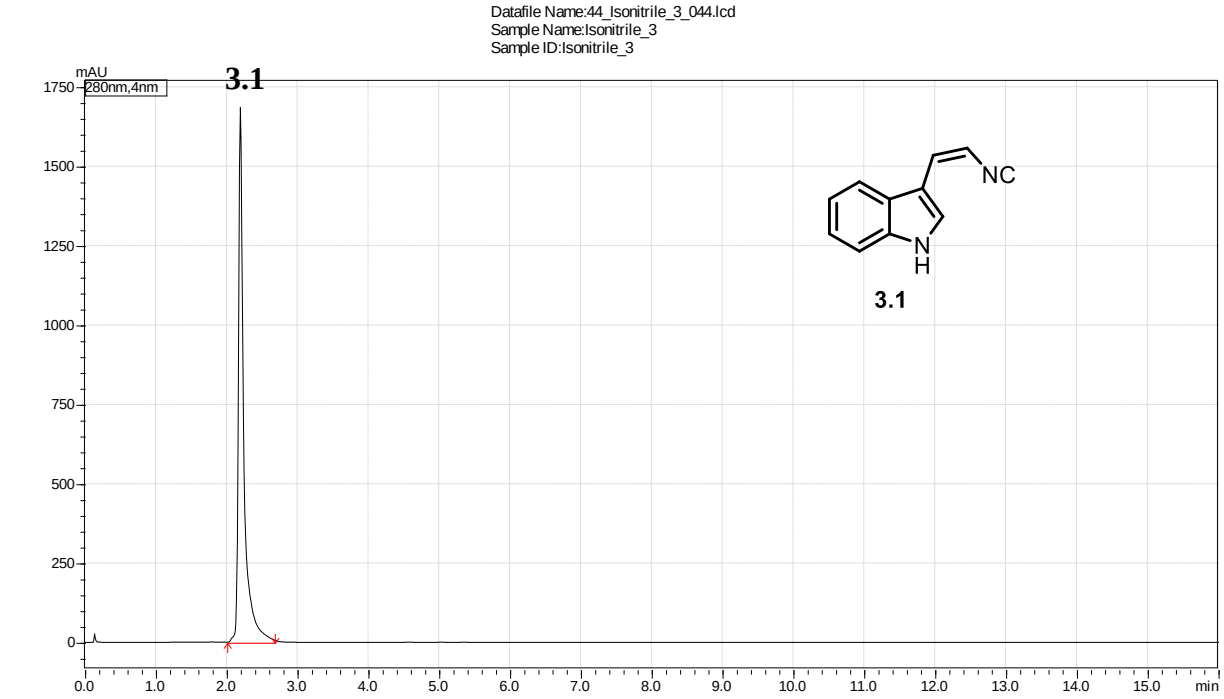
HpiC1_F138L
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNLGGEFQKISLAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

HpiC1_V51I
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTIDTPPGWTTYNPNNLVPEKRTTWTSNNGVGY
VGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNFGGEFQKISLAGFPGYRVELLAG
DTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

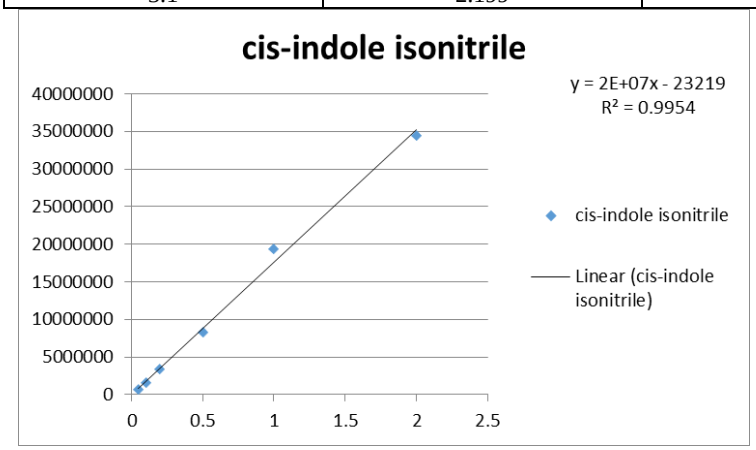
HpiC1_L147FF138L
MGSSHHHHHHSSGLVPRGSHMASTSVVSIPINNAGFEDPFIEVVDDYTVDTPPGWTTYNPNNLVPEKRTTWTSNNGVG
YVGPGTQFYNQLAPEGRNIGYIYLAQKPGSGVAGFEQILDATLEPDTKYTLKVDVGNLGGEFQKISFAGFPGYRVELLA
GDTVLAADHNNLYIKDGEFKTSTVTFTATPDNPYLDQKLGIRLINLLQGTFSGLDFDNVRLTVEPAQT

Table S3.2: Mutagenic primers. (5'→3')	
SN_HpiC1_F138S	CCCTGAAAGTGGACGTTGGTAACtctGGTGGCGAGTTTCAGAAAATTAGCC
SN_HpiC1_Y101F	CGCCGGAAGGTCGTAACATCGGCttcATTTATCTGGCGCAGAACCGGG
RMH_HpiC1_F138L	CACCCTGAAAGTGGACGTTGGTAACctgGGTGGCGAGTTTCAG
RMH_HpiC1_L147F	GGTGGCGAGTTTCAGAAAATTAGCtttGCGGGTTTTCCGGGCTACCG
RMH_HpiC1_V51I	CGAAGTGGTTGACGATTACACCattGATACCCCGCCGGGTTGGAC

Tables S3.3-S3.15: FamD2 TTN calculations (Compound #'s given)
All TTN values were calculated by using (#mol substrate consumed/#mol of enzyme). Each starting concentration was 2 mM substrate and 1 µM FamD2 (1:2000 ratio) for 1 hour
Note: Molarity absorptivity of 3-GC and 2-GC intermediates is much lower than starting material at 280 nm.
Table S3.3: FamD2 TTN calculations for **3.1**

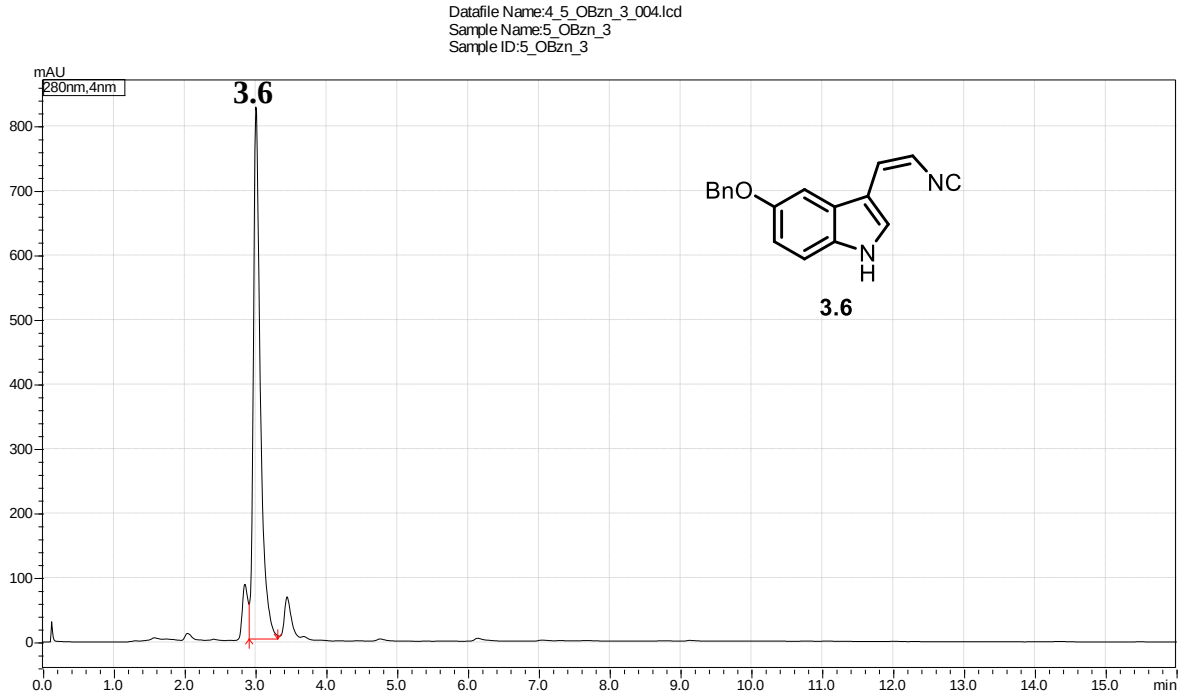


Peak#	Ret. Time (min)	Area	Height
3.1	2.199	8894557	1685057

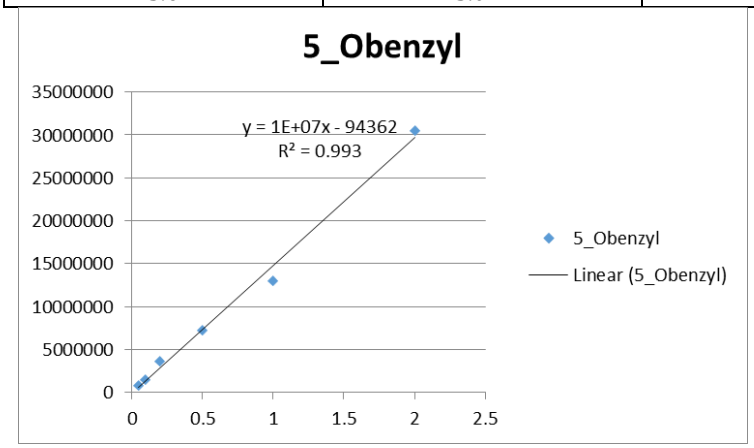


Concentration (mM)	Peak Area
0.05	653893
0.1	1554188
0.2	3331471
0.5	8326797
1	19411897
2	34535180
Eq: Area=17650000C-23219 TTN=(2-(8894557+23219)/17650000))*1000 TTN=1495	

Table S3.4: FamD2 TTN calculations for **3.6**

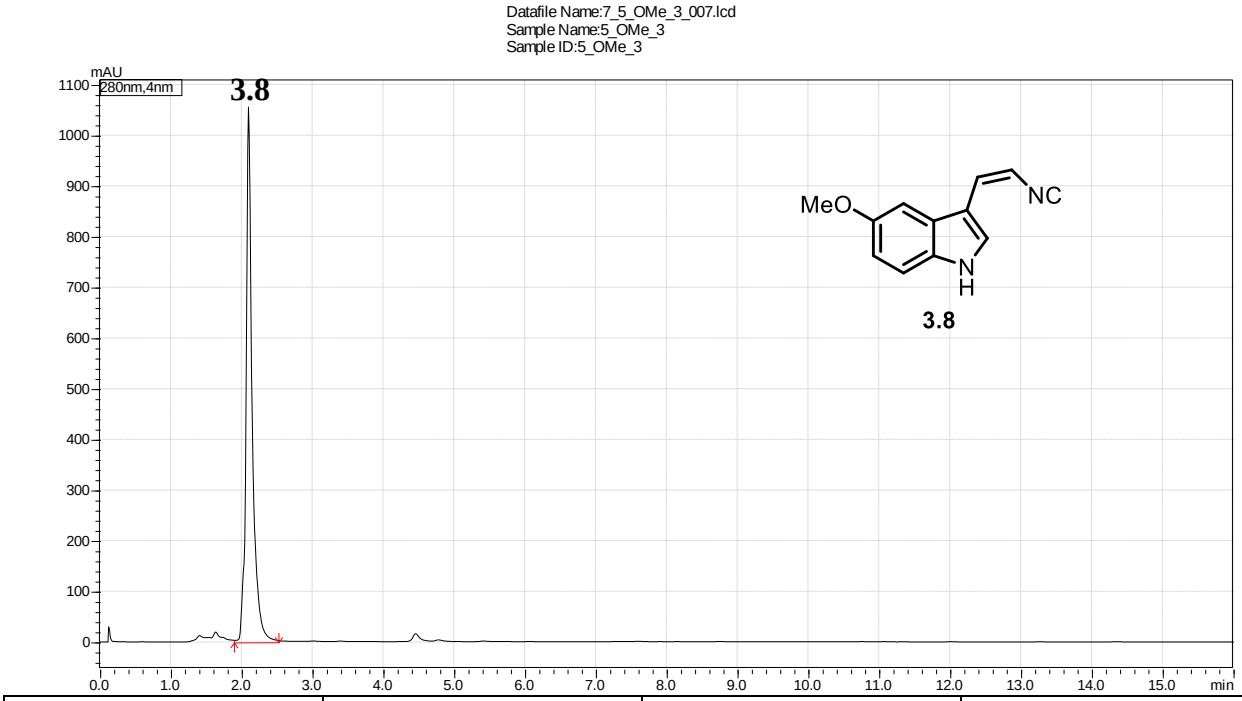


Peak#	Ret. Time (min)	Area	Height
3.6	3.012	5643864	823270

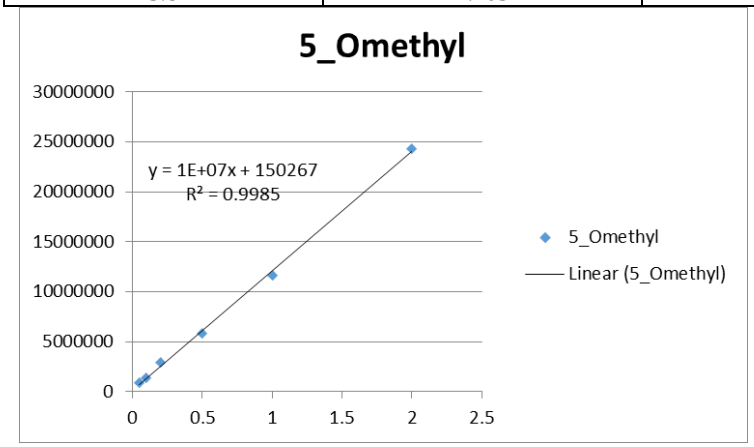


Concentration (mM)	Peak Area
0.05	845526
0.1	1552167
0.2	3612819
0.5	7210308
1	13033420
2	30538409
Eq: Area=14898000C-94361 TTN=(2-(5643864+94361)/14898000))*1000 TTN=1615	

Table S3.5: FamD2 TTN calculations for **3.8**

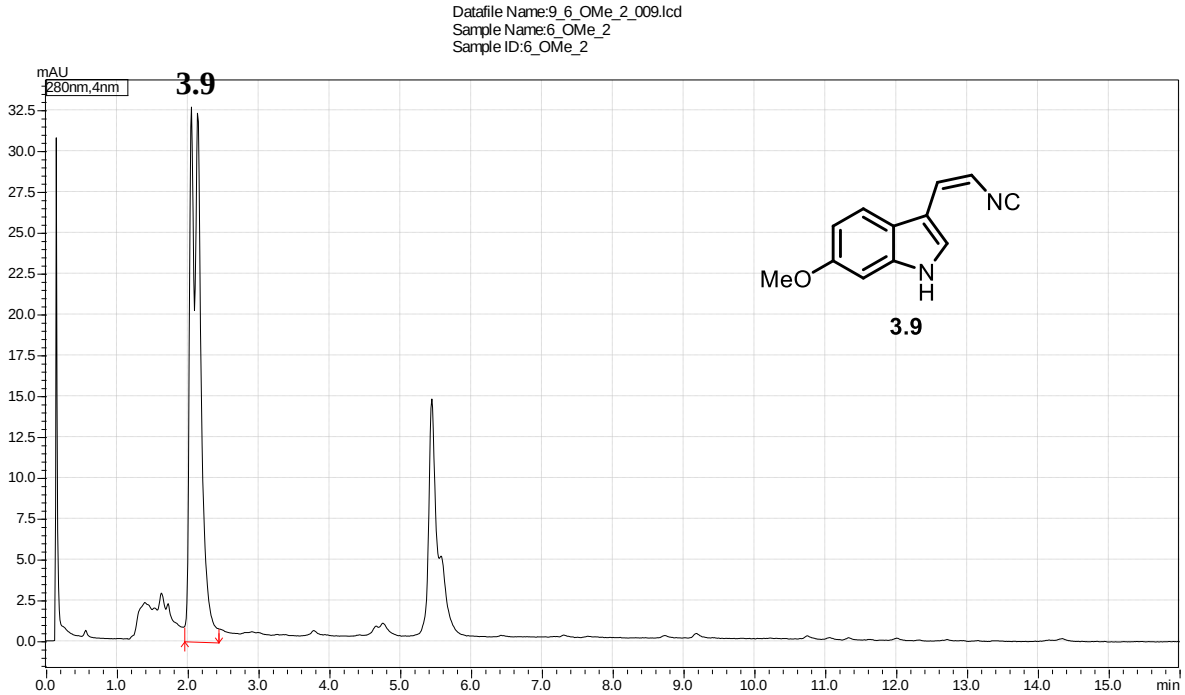


Peak#	Ret. Time (min)	Area	Height
3.8	2.103	6345503	1055592

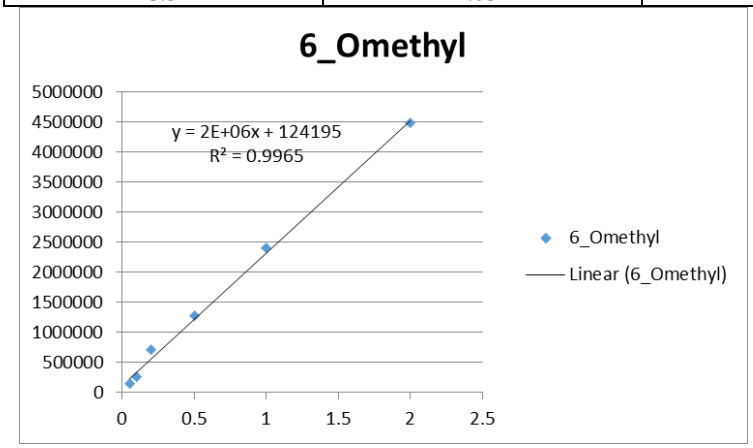


Concentration (mM)	Peak Area
0.05	880751
0.1	1350825
0.2	2934652
0.5	5795652
1	11602799
2	24324966
Eq: Area=11945000C+150267 TTN=(2-(6345503-150267)/11945000))*1000 TTN=1481	

Table S3.6: FamD2 TTN calculations for **3.9**

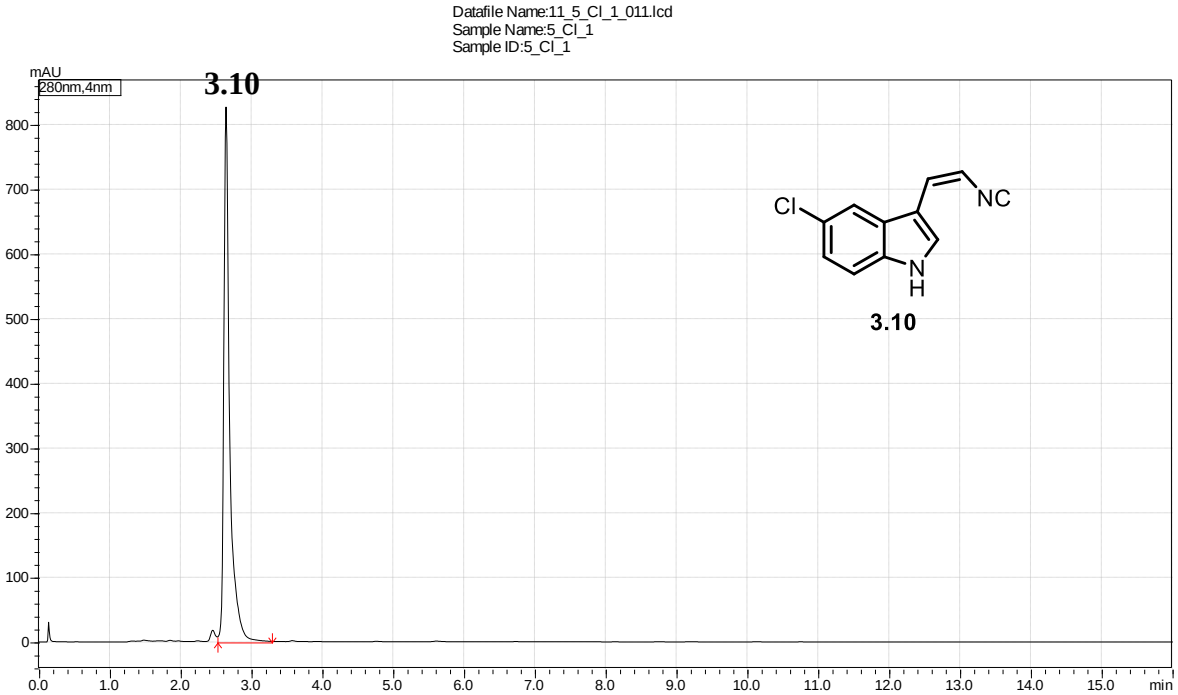


Peak#	Ret. Time (min)	Area	Height
3.9	2.057	339537	32684

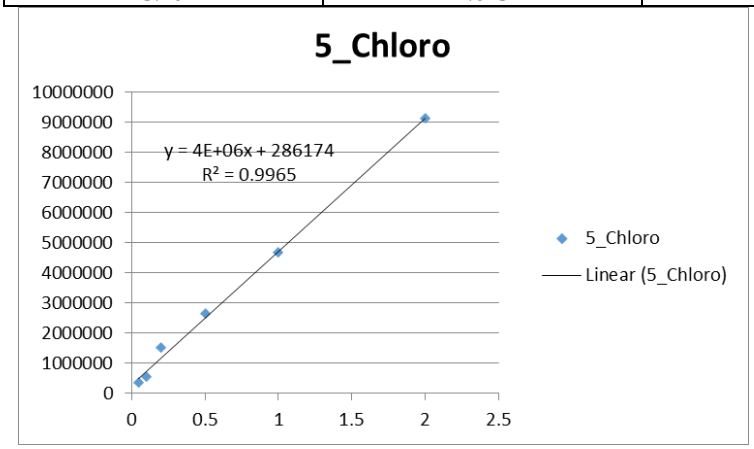


Concentration (mM)	Peak Area
0.05	126960
0.1	250899
0.2	702230
0.5	1269780
1	2384291
2	4465861
Eq: Area=2196100C+124195 TTN=(2-(339537-124196)/2196100))*1000 TTN=1902	

Table S3.7: FamD2 TTN calculations for **3.10**

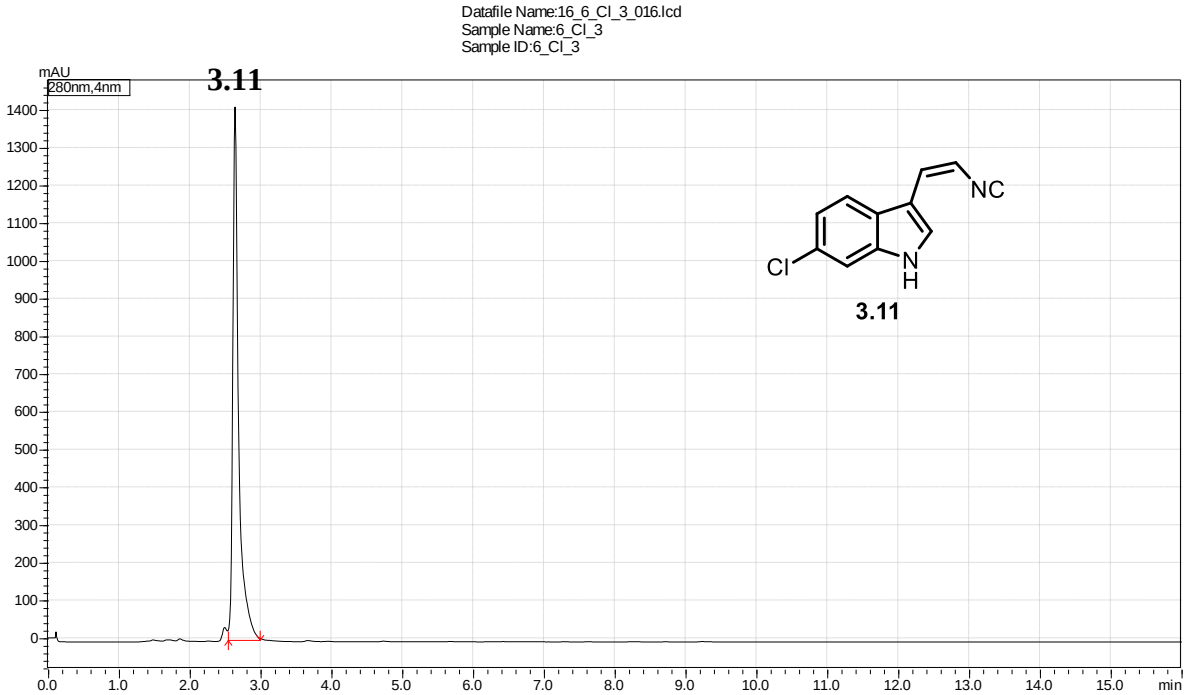


Peak#	Ret. Time (min)	Area	Height
3.10	2.645	4768339	827062

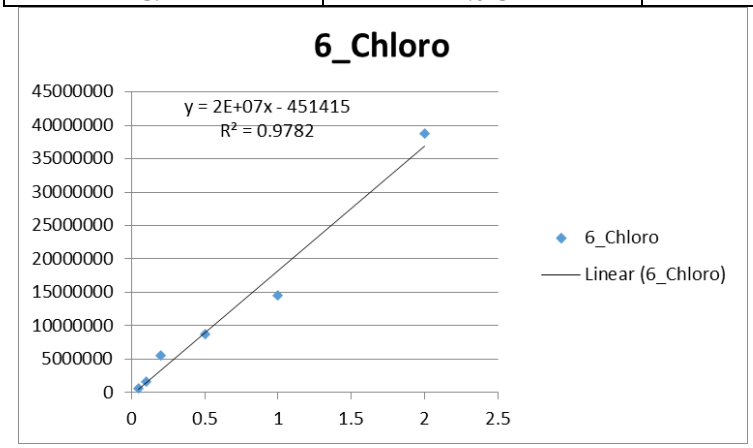


Concentration (mM)	Peak Area
0.05	338062
0.1	521558
0.2	1488020
0.5	2633172
1	4663983
2	9110272
Eq: Area=4254600C+286174 TTN=(2-(4768339-286175)/4254600))*1000 TTN=947	

Table S3.8: FamD2 TTN calculations for **3.11**

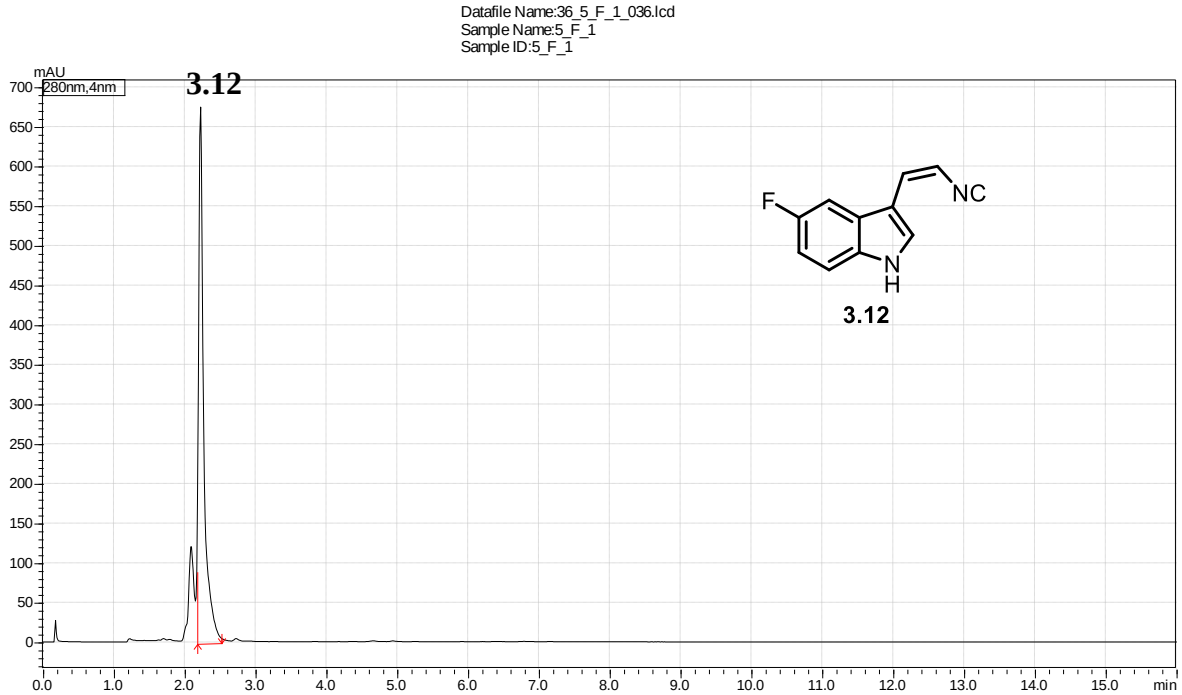


Peak#	Ret. Time (min)	Area	Height
3.11	2.645	8018316	1411263

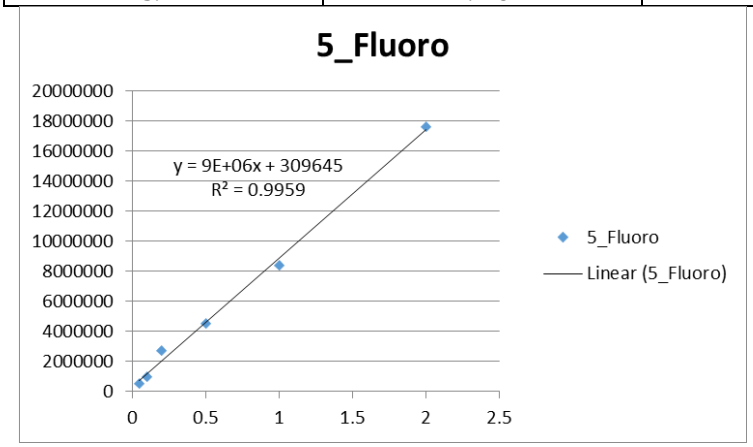


Concentration (mM)	Peak Area
0.05	550712
0.1	1523935
0.2	5421020
0.5	8628390
1	14447101
2	38655405
Eq: Area=18684000C-451415 TTN=(2-(8018316+451415)/18684000))*1000 TTN=1547	

Table S3.9: FamD2 TTN calculations for **3.12**

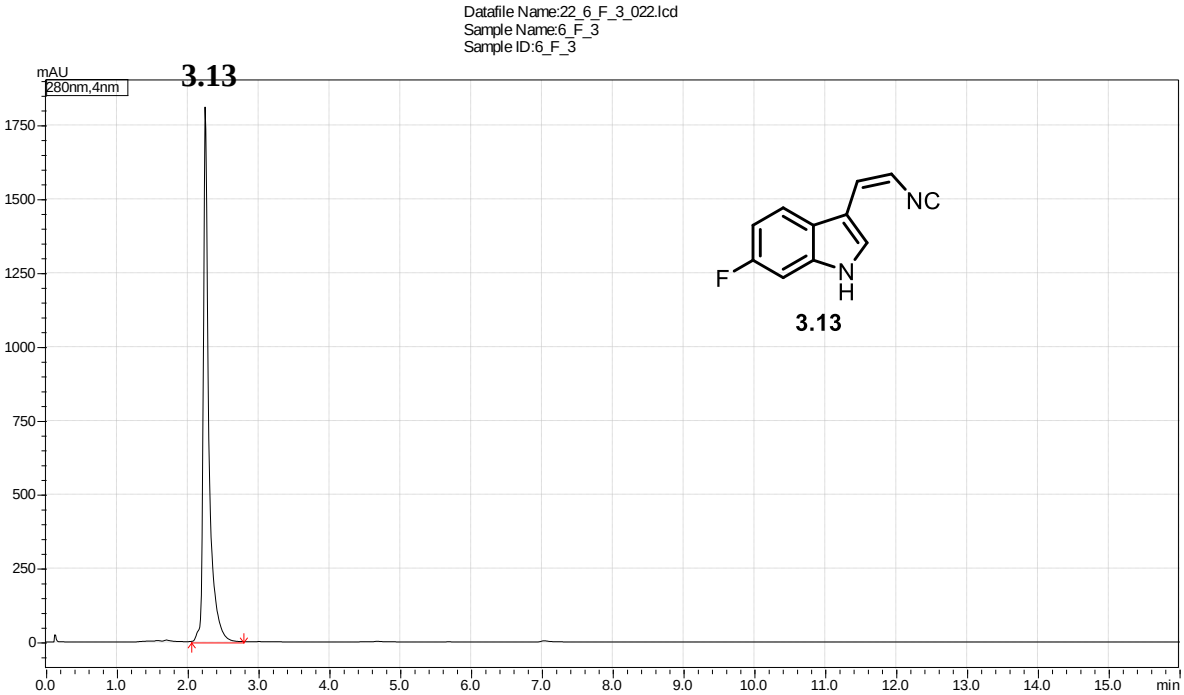


Peak#	Ret. Time (min)	Area	Height
3.12	2.228	3450687	676146

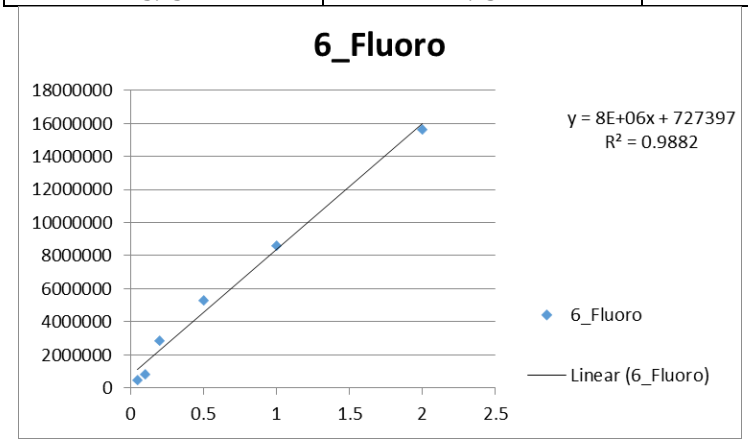


Concentration (mM)	Peak Area
0.05	524363
0.1	998106
0.2	2736403
0.5	4522171
1	8384625
2	17593782
Eq: Area=8372900C+309645 TTN=(2-(3450687-309645)/8372900))*1000 TTN=1625	

Table S3.10: FamD2 TTN calculations for **3.13**

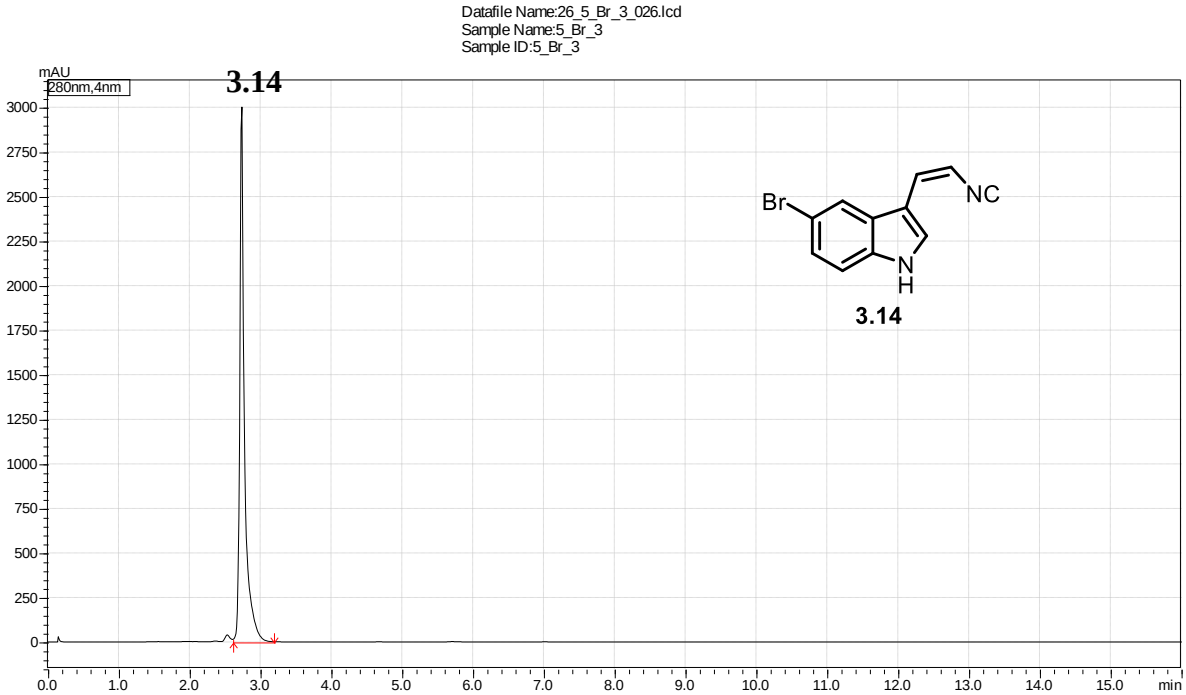


Peak#	Ret. Time (min)	Area	Height
3.13	2.254	9844920	1810594

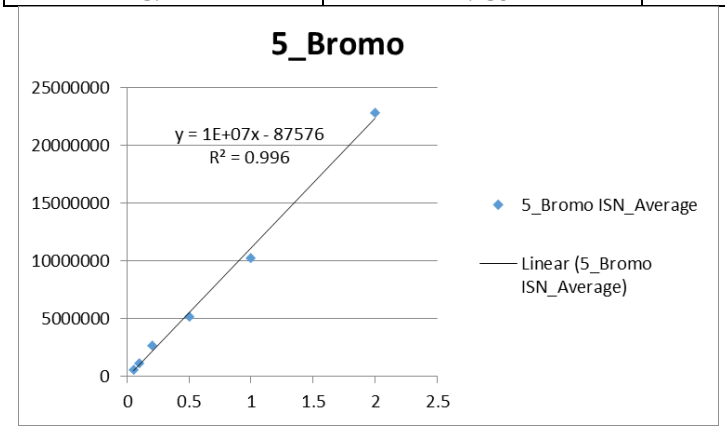


Concentration (mM)	Peak Area
0.05	463672
0.1	833895
0.2	2879462
0.5	5292420
1	8583214
2	15642615
Eq: Area=7618400C+727397 TTN=(2-(98944920-727397)/7618400))*1000 TTN=803	

Table S3.11: FamD2 TTN calculations for **3.14**

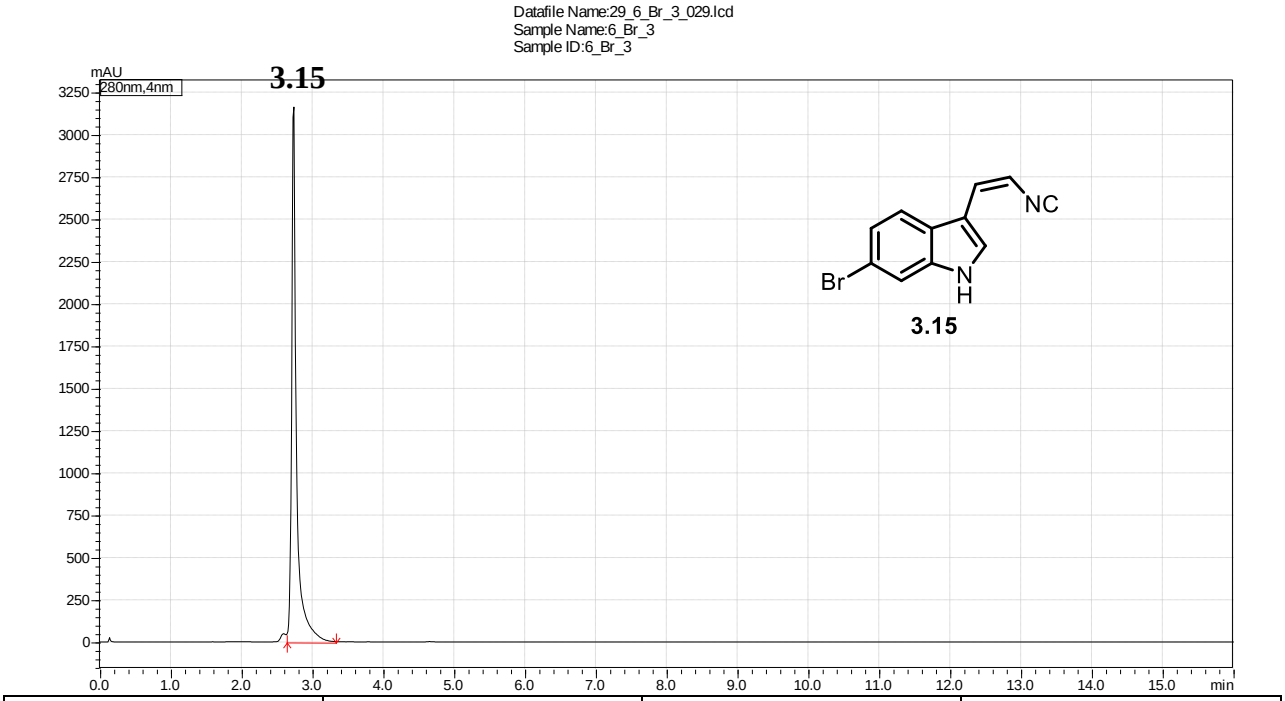


Peak#	Ret. Time (min)	Area	Height
3.14	2.738	13892693	3002168

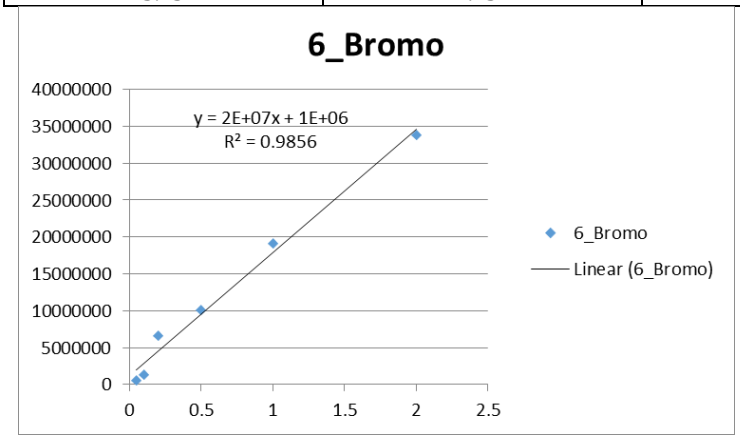


Concentration (mM)	Peak Area
0.05	574234
0.1	1172302
0.2	2680616
0.5	5195618
1	10238688
2	22839230
Eq: Area=11228000C-97576 TTN=(2-(13892693+97576)/11228000))*1000 TTN=755	

Table S3.12: FamD2 TTN calculations for **3.15**

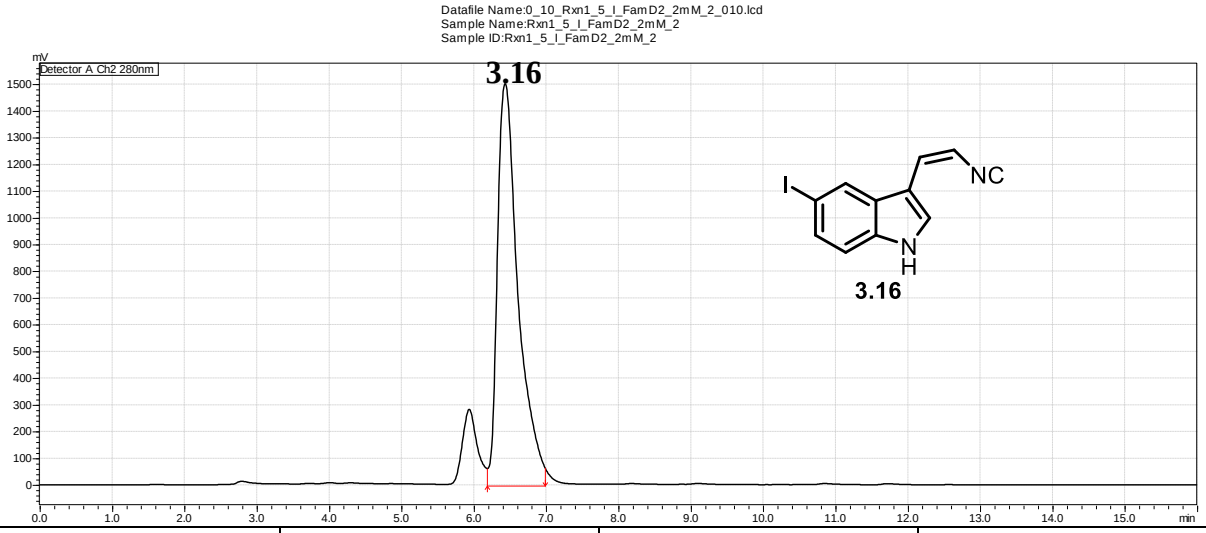


Peak#	Ret. Time (min)	Area	Height
3.15	2.737	14992583	3162897

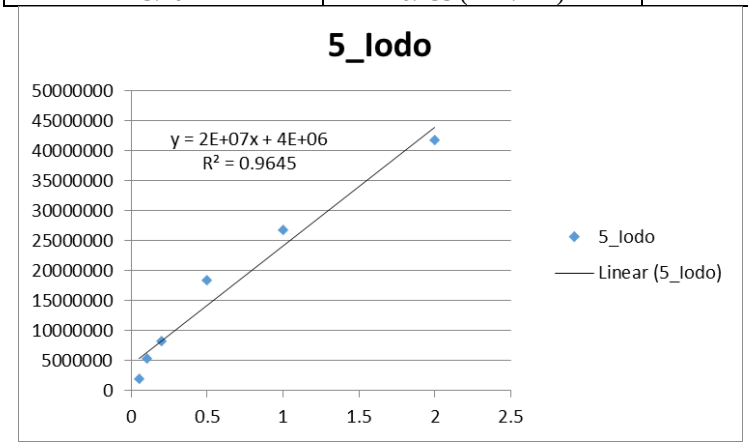


Concentration (mM)	Peak Area
0.05	523053
0.1	1281699
0.2	6598523
0.5	10125820
1	19156883
2	33784053
Eq: Area=16737000C+1171800 TTN=(2-(14992583-1171800)/16737000))*1000 TTN=1174	

Table S3.13: FamD2 TTN calculations for **3.16**

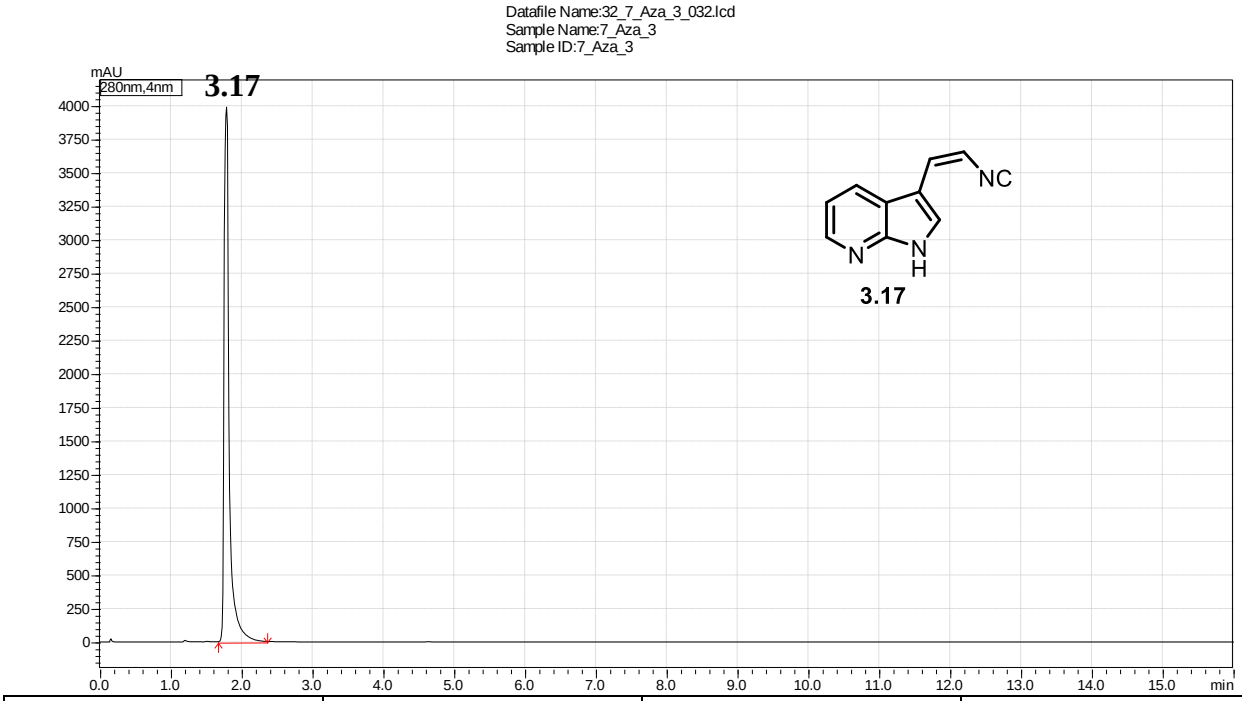


Peak#	Ret. Time (min)	Area	Height
3.16	6.435 (1mL/min)	29908556	1519361

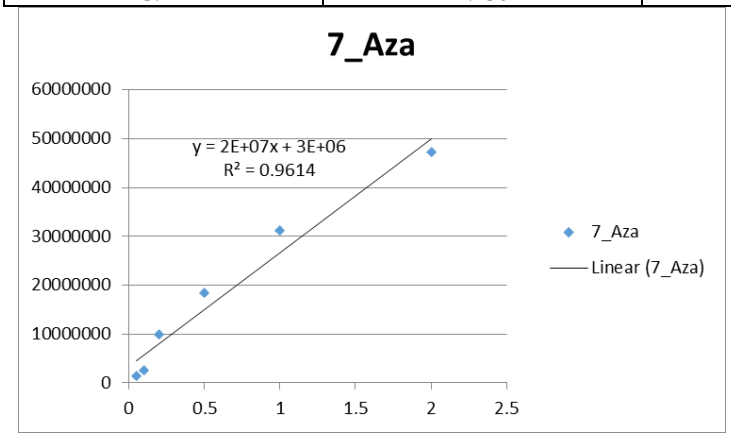


Concentration (mM)	Peak Area
0.05	2021327
0.1	5413201
0.2	8326473
0.5	18357085
1	26803087
2	41734933
Eq: Area=19734000C+4446700 TTN=(2-(29908566-4446700)/19734000))*1000 TTN=710	

Table S3.14: FamD2 TTN calculations for **3.17**

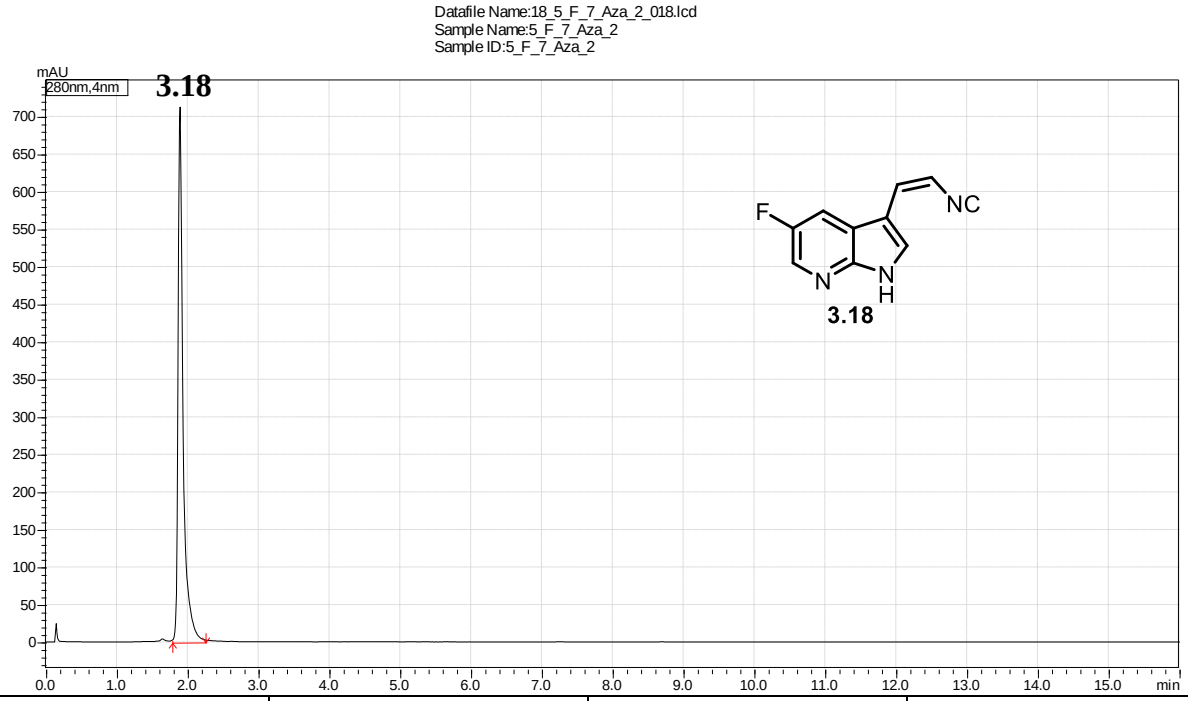


Peak#	Ret. Time (min)	Area	Height
3.17	1.790	20093276	3989928

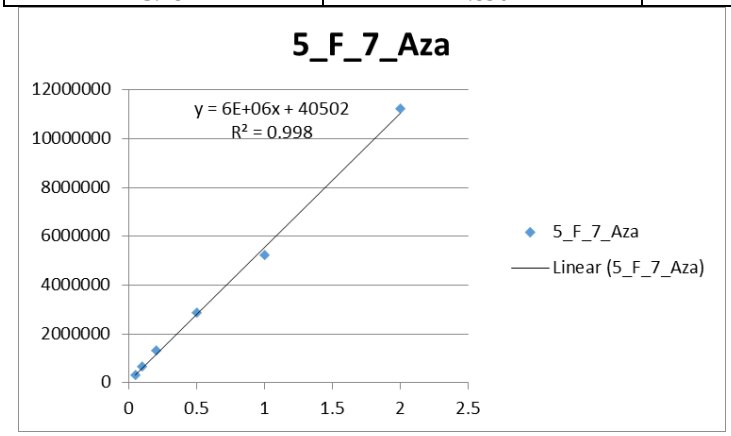


Concentration (mM)	Peak Area
0.05	1258424
0.1	2439309
0.2	9730293
0.5	18168987
1	31076307
2	47068189
Eq: Area=23351000C+3306500 TTN=(2-(20093276-3306500)/23351000))*1000 TTN=1281	

Table S3.15: FamD2 TTN calculations for **3.18**

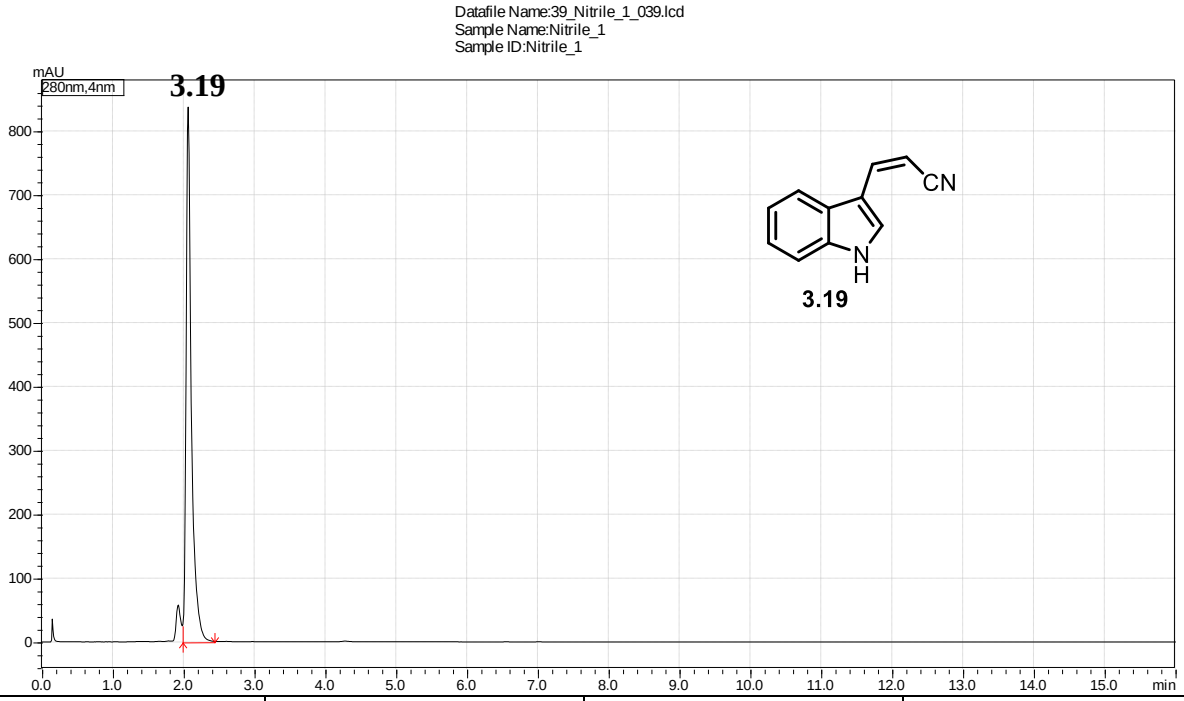


Peak#	Ret. Time (min)	Area	Height
3.18	1.896	3530558	712821

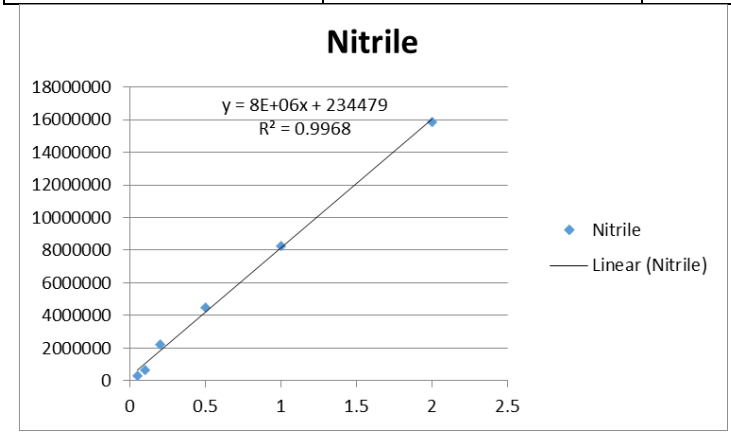


Concentration (mM)	Peak Area
0.05	283081
0.1	645116
0.2	1286867
0.5	2826221
1	5191694
2	11199120
Eq: Area=5503700C+40502 TTN=(2-(3530558-40502)/5503700))*1000 TTN=1366	

Table S3.16: FamD2 TTN calculations for **3.19**



Peak#	Ret. Time (min)	Area	Height
3.19	2.068	4373769	837043



Concentration (mM)	Peak Area
0.05	326968
0.1	628233
0.2	2231327
0.5	4492460
1	8254223
2	15871448
Eq: Area=7579000C+234479 TTN=(2-(4373769-234479)/7579000))*1000 TTN=1454	

Characterization Tables (Tables S3.17-S3.35)

12-*epi*-hapalindole U derivatives

Table S3.17: 5-fluoro-12-*epi*-hapalindole U (**3.22**)
Full characterization data[68]

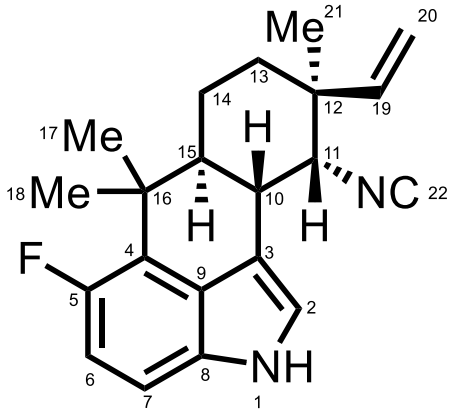


Table S3.18: 6-fluoro-12-*epi*-hapalindole U (**3.23**)
Full characterization data[68]

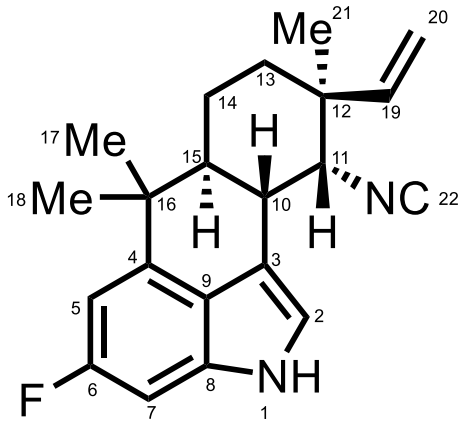
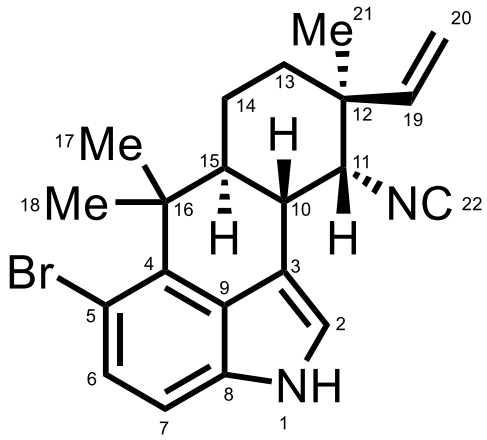
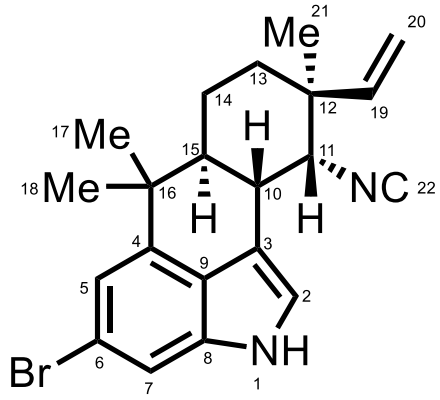


Table S3.19: 5-bromo-12-*epi*-hapalindole U (**3.26**)
$[\alpha]_D^{25}$=+10.5 (c=0.15, CH_2Cl_2)
Calc. $[M+H]^+$ 383.1117 Obsv. 383.1117



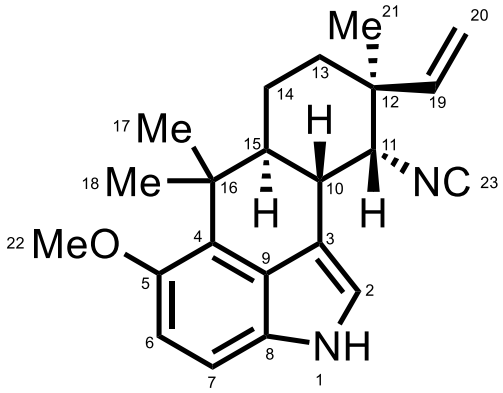
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		6.53, bs		
2	117.00	6.26, t, (2.0)	10	3,8,9
3	113.74			
4	137.84			
5	110.12			
6	130.05	7.52, d, (8.5)	7	4,5,8
7	110.62	6.59, d, (8.5)	6	5,9
8	133.68			
9	127.61			
10	34.00	2.83, d, (11.8)	11,15	3,15
11	63.18	3.88, s	10,14	3,10,12,14,15,16,19,22
12	39.66			
13	22.10	1.48, dd, (13.1, 3,6) & 1.29, td, (13.0, 3.5)	14	15,17,21
14	31.49	1.65, td, (13.9, 3.9) & 1.42, d, (14.2)	13	13,16,19
15	44.64	1.96, td, (12.0, 3.3)	10,14	10,16,17
16	39.51			
17	21.09	1.22, s		4,15,16,18
18	24.50	1.77, s		4,15,16,17
19	142.51	5.37, dd, (17.7, 11.1)	20	11,12,14
20	114.50	Cis 4.90, d, (11.0) Trans 4.83, d, (17.6)	19	12,19
21	28.21	1.17, s		11,12,15,19
22	160.92			

Table S3.20: 6-bromo-12-*epi*-hapalindole U (**3.27**)
$[\alpha]_D^{25}$=+6.1 (c=0.08, CH_2Cl_2)
Calc. $[M+H]^+$ 383.1117 Obsv. 383.1121



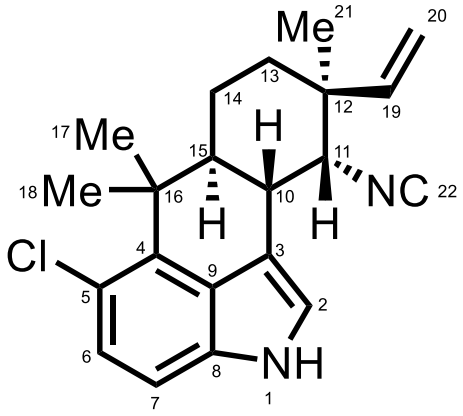
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		6.46, bs		
2	117.06	6.24, t, (1.9)		3,8,9
3	113.22			
4	142.74			
5	117.33	7.31, d, (1.9)		6,7,9,16
6	117.42			
7	111.94	7.16 (in solvent)		
8	135.10			
9	124.75			
10	34.81	2.92, d, (10.3)	11,15	3,15
11	63.43	3.85, d, (3.1)	10	3,10,12,13,15,19,22
12	39.78			
13	31.07	1.64, td, (13.8,3.9) & 1.42, dt, (14.3,3.7)	14	
14	21.58	1.39, dd, (13.0,3.5) & 1.26, qd, (13.0,3.7)	13,15	
15	43.49	1.88, td, (12.0,3.5)	10,14	10,14,16,18
16	37.41			
17	24.07	1.10, s		4,15,16,18
18	24.94	0.85, s		4,15,16,17
19	142.49	5.39, dd, (17.7,11.0)	20	11,12,13
20	114.51	(cis) 4.91, d, (11.0) (trans) 4.84, d, (17.7)	19	12,19,21
21	28.23	1.16, s		11,12,13,19
22	160.89			

Table S3.21: 5-methoxy-12-*epi*-hapalindole U (**3.28**)
$[\alpha]_D^{25}$=+7.4 (c=0.06, CH_2Cl_2)
Calc. $[M+H]^+$ 335.2118 Obsv. 335.2127



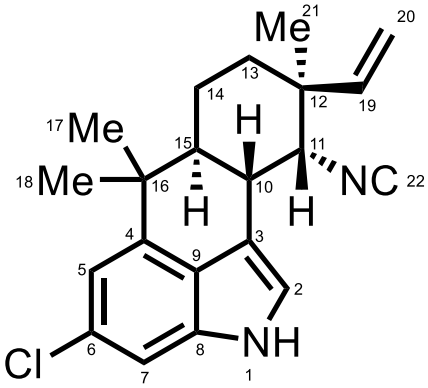
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.57, bs**	**2**	
2	**117.40**	**6.40, t, (1.90)**	**1**	**3,4,9**
3	**113.53**			
4	**130.57**			
5	**150.60**			
6	**111.74**	**6.88, m**	**7**	**5,8,9**
7	**109.12**	**6.88, m**	**6**	**5,9**
8	**126.52**			
9	**127.46**			
10	**34.57**	**2.96, d, (11.6)**	**11, 15**	**3,15**
11	**63.47**	**3.96, s**	**10, 13, 15**	**3,10,12,15,13,19,21,23**
12	**39.77**			
13	**31.54**	**1.72 & 1.49, m**	**14**	**11,14,15,17**
14	**21.61**	**1.63 & 1.40, m**	**13**	**13,18**
15	**44.59**	**2.06, td, (12.0, 3.4)**	**10, 11, 14**	**10,16,17**
16	**37.84**			
17	**22.53**	**1.25, s**	**18**	**4,15,16,18**
18	**25.09**	**1.75, s**	**17**	**4,15,16,17**
19	**142.74**	**5.43, dd, (17.7, 10.9)**	**20**	**11,12,13,21,**
20	**114.36**	**4.92 (cis), d, (10.9)** **4.87 (trans), d, (17.7)**	**19**	**12,19,21**
21	**28.32**	**1.20, s**		**11,12,13,14,19**
22	**57.12**	**3.59, s**		**5**
23	**160.72**			

Table S3.22: 5-chloro-12- *epi*-hapalindole U (**3.24**)
$[\alpha]_D^{25}$=+3.2 (c=0.08, CH_2Cl_2)
Calc. $[M+H]^+$ 339.1623 Obsv. 339.1633



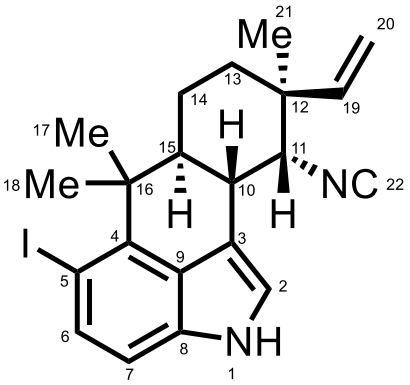
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.52, bs**		
2	**117.27**	**6.29, d, (2.0)**		**3,8,9**
3	**113.76**			
4	**136.16**			
5	**121.85**			
6	**126.70**	**7.30, d, (8.4)**	**7**	**4,8,5**
7	**110.24**	**6.66, d, (8.4)**	**6**	**5,9**
8	**133.26**			
9	**127.69**			
10	**34.13**	**2.85, d, (11.7)**	**15**	**3**
11	**63.22**	**3.89, s**	**10,13**	**10,13,15,22**
12	**39.67**			
13	**31.48**	**1.66, td, (13.8,3.8) & 1.43, m**	**14**	
14	**21.87**	**1.50, dt, (13.3,3.5) & 1.30, m**	**10,13**	
15	**44.86**	**1.97, td, (11.9,3.2)**	**10,14**	
16	**39.10**			
17	**21.23**	**1.20, s**		**4,15,16,18**
18	**24.52**	**1.74, s**		**4,15,16,17**
19	**142.49**	**5.38, dd, (17.7,11.0)**	**20**	**11,12,13,**
20	**114.48**	**Cis 4.90, d, (11.0) Trans 4.83, d, (17.7)**	**19**	**12,19**
21	**28.21**	**1.18, s**		**11,12,13,19**
22	**160.90**			

Table S3.23: 6-chloro-12-*epi*-hapalindole U (**3.25**)
$[\alpha]_D^{25}$=+2.6 (c=0.15, CH_2Cl_2)
Calc. $[M+H]^+$ 339.1623 Obsv. 339.1631



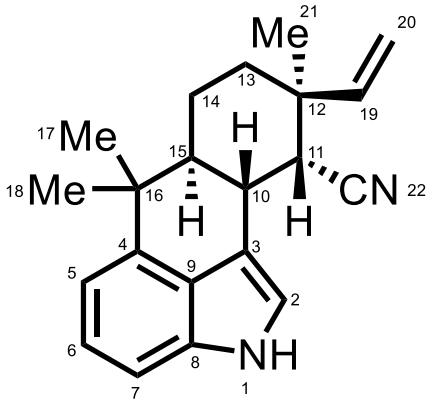
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.48, bs**		
2	**116.70**	**6.26, s**	**1,10**	**3,4,9**
3	**112.76**			
4	**141.93**			
5	**108.58**	**7.02, s**	**7**	**4,7**
6	**124.07**			
7	**114.38**	**7.22, s**	**5**	**5**
8	**129.09**			
9	**134.19**			
10	**34.40**	**2.92, d, (11.5)**	**11, 15**	
11	**63.07**	**3.85, s**	**14,15**	**15**
12	**39.36**			
13	**21.17**	**1.39 & 1.25, m**	**14**	
14	**30.63**	**1.64, t, (13.0) & 1.41, m**	**13,15**	**13,19**
15	**43.09**	**1.89, t, (12.0)**	**10,11,14**	
16	**37.00**			
17	**23.67**	**1.12, s**		**4,15,16,18**
18	**24.52**	**0.86, s**		**4,15,16,17**
19	**142.07**	**5.38, dd, (17.7, 11.0)**	**20**	**11,12,14**
20	**114.10**	**Cis 4.92, d, (11.0) Trans 4.84, d, (17.7)**	**19**	**12,19**
21	**27.83**	**1.16, s**		**11,12,13,19**
22	**160.44**			

Table S3.24: 5-iodo-12-*epi*-hapalindole U (**3.29**)
$[\alpha]_D^{25}$=+13.7 (c=0.10, CH_2Cl_2)
Calc. $[M+H]^+$431.0979 Obsv. 431.0990



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.54, bs**	**2**	
2	**116.47**	**6.21, s**	**1,10**	**3,8,9**
3	**113.46**			
4	**140.95**			
5	**81.03**			
6	**137.30**	**7.88, d, (8.5)**	**7**	**4,8**
7	**111.10**	**6.45, d, (8.4)**	**6**	**5,9**
8	**134.38**			
9	**127.61**			
10	**33.82**	**2.83, d, (11.8)**	**11,15**	**3,15**
11	**63.18**	**3.87, s**	**10,14**	**10,12,13,15,22**
12	**39.64**			
13	**31.50**	**1.63, td, (13.7, 3.6) & 1.39, m**	**14**	**12,14,19**
14	**22.43**	**1.44, m & 1.30, m**	**11,13,15**	**11,12,15**
15	**45.25**	**1.93, td, (12.0, 3.4)**	**10,14**	**10,16,17**
16	**39.37**			
17	**21.08**	**1.22, s**	**18**	**4,15,16,18**
18	**24.44**	**1.77, s**	**17**	**4,15,16,17**
19	**142.50**	**5.37, dd, (17.7, 11.0)**	**20**	**11,12,13**
20	**114.50**	**(cis) 4.89, d, (11.0) (trans) 4.82, d, (17.7)**	**19**	**12,19**
21	**28.20**	**1.17, s**		**11,12,13,19**
22	**161.00**			

Table S3.25: 12-*epi*-hapalindole U nitrile (**3.30**)
$[\alpha]_D^{25}$=+3.4 (c=0.13, CH_2Cl_2)
Calc. $[M+H]^+$ 305.2012 Obsv. 305.2019



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.66, bs**		
2	**116.64**	**6.38, d, (1.9)**		**3,8,9**
3	**114.20**			
4	**141.10**			
5	**113.48**	**7.08, d, (7.3)**	**6**	**7,9**
6	**123.54**	**7.29, t, (7.7)**	**5,7**	**4,8**
7	**108.76**	**6.97, d, (8.1)**	**6**	**5,9**
8	**134.60**			
9	**125.62**			
10	**33.34**	**3.04, ddd, (11.5, 4.4, 1.6)**	**11,15**	**3,13,15,22**
11	**44.16**	**3.09, dd, (4.3, 2.0)**	**10**	**3,10,12,13,15,22**
12	**38.98**			
13	**33.46**	**1.69, m & 1.53, m**	**14**	**15,19**
14	**22.01**	**1.53, m & 1.33, m**	**10,13**	**15**
15	**46.15**	**1.97, m**	**10,14**	**13,14,16,17**
16	**37.48**			
17	**25.31**	**1.00, s**		**4,15,16,18**
18	**24.43**	**1.30, s**		**4,15,16,17**
19	**143.71**	**5.44, dd, (17.7, 11.0)**	**20**	**10,11,12,13**
20	**114.08**	**Cis 4.93, d, (11.0) Trans 4.85, d, (17.7)**	**19**	**12,19**
21	**29.68**	**1.23, s**		**11,12,13,19**
22	**119.28**			

12-*epi*-fischerindole U derivatives

Table S3.26: 5-fluoro-12-*epi*-fischerindole U (**3.33**)
Full characterization data[68]

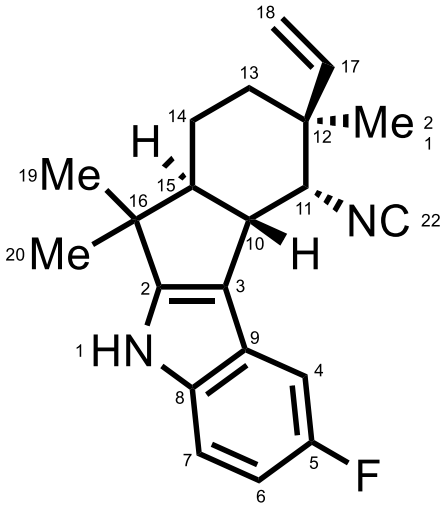


Table S3.27: 6-fluoro-12-*epi*-fischerindole U (**3.37**)
Full characterization data[68]

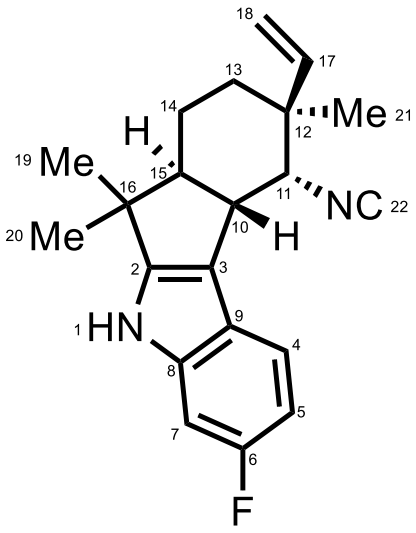
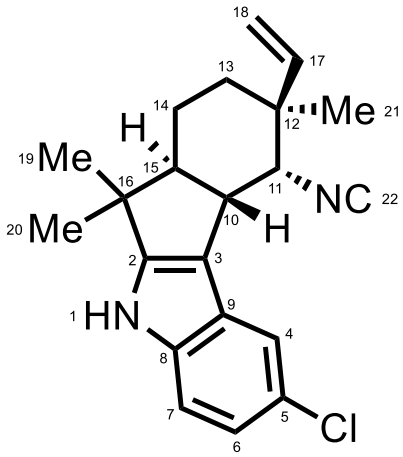
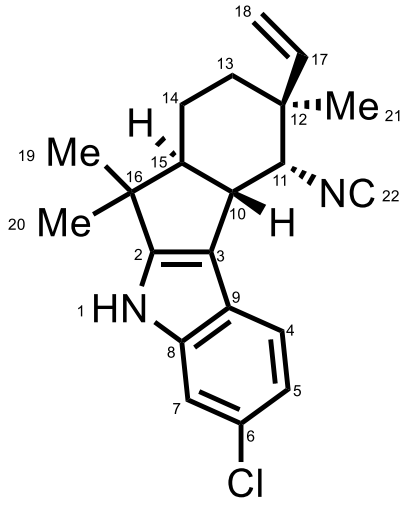


Table S3.28: 5-chloro-12-*epi*-fischerindole U (**3.34**)
$[\alpha]_D^{25}$=+9.8 (c=0.08, CH_2Cl_2)
Calc. $[M+H]^+$ 339.1623 Obsv. 339.1621



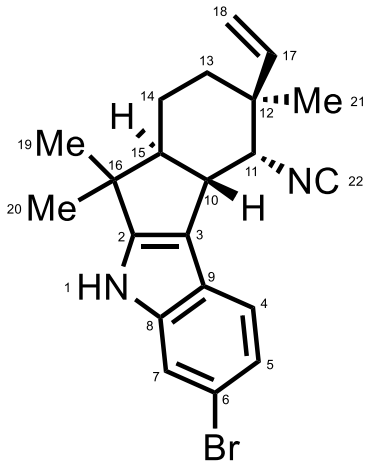
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		6.48, bs		
2	153.58			
3	114.76			
4	118.24	7.68, d, (2.0)		3,5,6,8
5	125.40			
6	121.31	7.23, dd, (8.6,2.0)	6	4,8,9
7	113.07	6.81, d, (8.6)	7	4,5
8	138.40			
9	126.18			
10	42.37	2.86, dd,(10.8,4.3)	11	3,15
11	62.22	3.89, d,(2.9)	10,15	12,13,15
12	41.06			
13	31.81	1.54, m	15	11,12,14,15,17
14	20.82	1.33, m & 1.22, m	13,15	
15	55.16	2.35, ddd, (13.3,10.7,3.1)	11,14	16,19,20
16	40.16			
17	143.02	5.31, dd,(17.6,11.0)	18	11,12,13
18	114.19	(cis) 4.88, d, (10.9) (trans) 4.83, d, (17.6)	17	12,17
19	24.90	1.00, s		2,16,15,20
20	20.73	0.72, s		2,16,15,19
21	28.13	1.12, s		11,12,13,17
22	161.37			

Table S3.29: 6-chloro-12-*epi*-fischerindole U (**3.38**)
$[\alpha]_D^{25}$=+20.3 (c=0.05, CH_2Cl_2)
Calc. $[M+H]^+$ 339.1623 Obsv. 339.1634



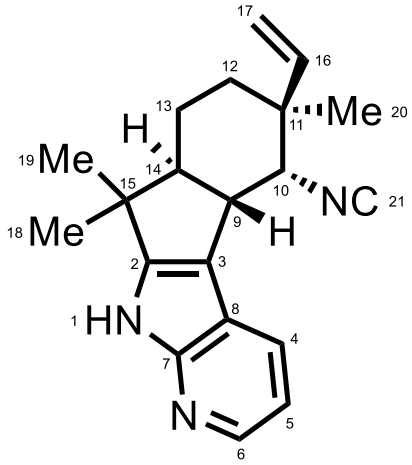
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.39, bs**		
2	**152.70**			
3	**115.00**			
4	**112.25**	**7.16, in solvent peak**	**5**	**3**
5	**120.96**	**7.23, dd, (8.4, 1.9)**	**4**	**4,9**
6	**126.93**			
7	**119.44**	**7.17, s**		
8	**140.37**			
9	**122.93**			
10	**42.39**	**2.86, m**	**11,15**	
11	**62.34**	**4.00, d, (3.0)**	**10**	**12,13,15**
12	**41.02**			
13	**31.84**	**1.55, dd, (9.0, 3.8)**	**14**	**12,15**
14	**20.64**	**1.32, m & 1.22, m**	**10,13**	**13,15**
15	**55.30**	**2.34, ddd, (13.5, 10.7, 3.2)**	**10,14**	
16	**40.22**			
17	**143.03**	**5.32, dd, (17.6, 11.0)**	**18**	**11,12,13**
18	**114.20**	**4.87, m**	**17**	**12,17**
19	**24.92**	**0.99, s**		**2,15,16,20**
20	**20.79**	**0.72, s**		**2,15,16,19**
21	**28.15**	**1.15, s**		**11,12,13,17**
22	**161.38**			

Table S3.30: 6-bromo-12-*epi*-fischerindole U (**3.39**)
$[\alpha]_D^{25}$=+14.8 (c=0.07, CH_2Cl_2)
Calc. $[M+H]^+$ 383.1117 Obsv. 383.1133



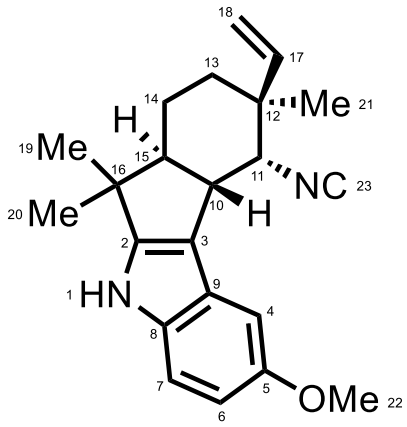
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.41, bs**		
2	**152.65**			
3	**115.03**			
4	**119.82**	**7.12, d, (8.4)**	**5**	**6,8**
5	**123.55**	**7.35, dd, (8.3,1.8)**	**4**	**3,7,9**
6	**114.49**			
7	**115.22**	**7.29, d, (1.8)**		**5,6**
8	**140.78**			
9	**123.18**			
10	**42.38**	**2.86, dd, (11.1,4.1)**	**11,15**	**3**
11	**62.33**	**4.00, d, (3.0)**	**10**	**10,13,15,22**
12	**41.02**			
13	**31.83**	**1.55, m**		**11,12,15,17**
14	**20.61**	**1.22, m & 1.34, m**	**15**	
15	**55.31**	**2.33, ddd, (13.4,10.6,3.1)**	**10,14**	**16,19,20**
16	**40.20**			
17	**143.01**	**5.33, dd, (17.6,11.0)**	**18**	**11,12,13**
18	**114.20**	**(cis) 4.88, d, (11.6) (trans) 4.86, d, (18.1)**	**17**	**12,17**
19	**24.89**	**0.99, s**		**2,15,16,20**
20	**20.79**	**0.72, s**		**2,15,16,19**
21	**28.15**	**1.14, s**		**11,12,13,17**
22	**161.35**			

Table S3.31: 7-aza-12-*epi*-fischerindole U (**3.36**)
$[\alpha]_D^{25}$=+17.2 (c=0.17, CH_2Cl_2)
Calc. $[M+H]^+$ 306.1965 Obsv. 306.1972



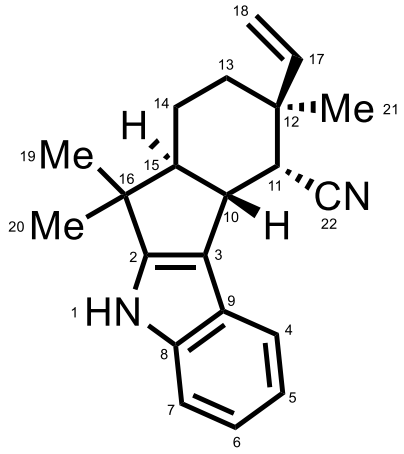
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		10.42, bs		
2	153.16			
3	112.46			
4	126.02	7.55, dd, (7.7, 1.6)	5	3,6,7,8
5	116.25	6.89, dd, (7.7, 4.7)	4,6	6,8
6	141.04	8.35, dd, (4.7,1.6)	5	4,5,7
7	152.56			
8	117.63			
9	42.91	2.96, dd, (10.5,3.6)	10,14	3,14
10	62.47	4.06, d, (3.0)	9	9,12,14,16,21
11	41.03			
12	31.89	1.56, dd, (11.3,4.0)	13	14,16
13	20.75	1.35, dq, (17.8,6.3) & 1.24, dt, (12.7,3.7)	12,14	14
14	54.82	2.37, ddd, (13.4, 10.7,3.1)	9,13	13,15
15	40.63			
16	143.07	5.37, dd, (17.6,11.0)	17	10,11,12
17	114.22	(cis) 4.90, d, (11.2) (trans) 4.88, d, (17.9)	16	11,16,20
18	20.62	0.80, s		2,14,15,19
19	24.94	1.14, s		2,14,15,18
20	28.15	1.16, s		10,11,12,16
21	161.50			

Table S3.32: 5-methoxy-12-*epi*-fischerindole U (**3.35**)
$[\alpha]_D^{25}$=+36.6 (c=0.03, CH_2Cl_2)
Calc. $[M+H]^+$ 335.2118 Obsv. 335.2131



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.47, bs**		
2	**152.81**			
3	**114.78**			
4	**101.45**	**7.20, d, (2.4)**		**3,6,9**
5	**155.28**			
6	**110.76**	**7.05, dd, (8.8,2.4)**		**4,9**
7	**112.69**	**6.99, d, (8.8)**		**4,8,5**
8	**125.00**			
9	**135.34**			
10	**42.64**	**3.01, dd, (11.0,4.2)**	**11,15**	**3,15**
11	**62.66**	**4.18, d, (3.0)**	**10**	**10,12,13,15**
12	**41.07**			
13	**31.86**	**1.59, m**	**14**	**12,15,17**
14	**20.91**	**1.40, m & 1.27, dt, (12.3,3.6)**	**13**	
15	**55.25**	**2.42, ddd, (13.4,10.7,3.5)**	**10,14**	**16,19**
16	**40.11**			
17	**143.12**	**5.34, dd, (17.5,11.0)**	**18**	**12,13**
18	**114.10**	**(cis) 4.89, d, (1.5) (trans) 4.87, d, (8.4)**	**17**	**12,17**
19	**25.07**	**1.06, s**		**2,15,16,20**
20	**20.86**	**0.81, s**		**2,15,16,19**
21	**28.12**	**1.12, s**		**11,12,13,17**
22	**55.59**	**3.58, s**		**5**
23	**161.34**			

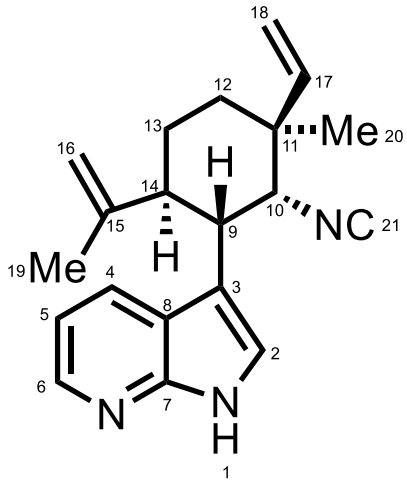
Table S3.33: 12-*epi*-fischerindole U nitrile (**3.40**)
$[\alpha]_D^{25}$=+66.5 (c=0.03, CH_2Cl_2)
Calc. $[M+H]^+$ 305.2012 Obsv. 305.2026



Position	^{13}C shift (ppm),	^{1}H shift (ppm), multi (J)	COSY	HMBC	NOESY
1		**6.54, bs**			
2	**151.63**				
3	**115.55**				
4	**118.77**	**7.52, dd, (6.2,2.9)**	**5**	**5,8**	
5	**120.40**	**7.24, dd, (6.0,3.1)**		**4,7,8**	
6	**121.13**	**7.24, dd, (6.0,3.1)**		**4,7,8**	
7	**112.19**	**7.11, m**	**6**	**6,9**	
8	**140.23**				
9	**124.61**				
10	**40.93**	**2.98, dd, (10.6,4.0)**	**11,15**	**3,14,15,22**	**14a,20**
11	**43.04**	**3.25, d, (3.9)**	**10,13**	**12,13,15,17,22**	**21**
12	**40.40**				
13	**33.76**	**1.61, m**	**14**	**12,14,15,17**	**14,21**
14	**21.03**	**1.38 (a), m & 1.28 (b), m**	**13,15**	**12,21**	
15	**58.49**	**2.39, ddd, (13.3,10.5,3.1)**	**10,14**	**10,14,19**	**14b,19**
16	**40.32**				
17	**113.94**	**5.34, dd, (17.6,10.9)**	**18**	**12,18**	**18**
18	**144.19**	**Cis 4.88, d, (10.9) Trans 4.85, d, (17.6)**	**17**	**11,12,13**	**17**
19	**25.03**	**1.03, s**		**2,15,16,20**	**14b,20,21**
20	**20.72**	**0.76, s**		**2,15,16,19**	**14a,19**
21	**29.68**	**1.19, s**		**11,12,13,14,17**	
22	**118.91**				

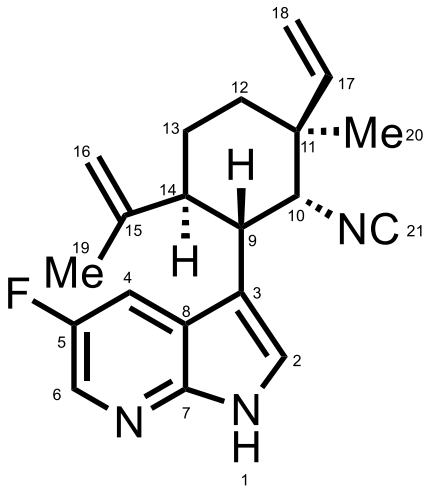
12-*epi*-hapalindole C Derivatives

Table S3.34: 7-aza-12-*epi*-hapalindole C (**3.31**)
$[\alpha]_D^{25}$=+2.5 (c=0.27, CH_2Cl_2)
Calc. $[M+H]^+$ 306.1965 Obsv. 306.1975



Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**7.72, bs**		
2	**124.20**	**6.88, s**		**3,7,8**
3	**112.65**			
4	**125.65**	**7.55, d, (7.9)**	**5**	**3,6,7,8**
5	**115.82**	**6.82, t (11.9, 6.9)**	**4,6**	**4,6,8**
6	**143.71**	**8.42, d, (3.9)**	**5**	**4,5,7**
7	**148.86**			
8	**119.24**			
9	**36.58**	**3.23, d, (12.2)**		**2,3,8,10,14**
10	**66.36**	**3.35, s**		**3,9,11,12,14,21**
11	**40.42**			
12	**30.81**	**1.77, td, (13.9, 3.5) & 1.37, m**	**13,20**	**11,13,14,17**
13	**28.69**	**1.56, m**	**12**	**9,12,14,15**
14	**43.79**	**2.87, td, (12.8, 3.5)**	**13**	**3,9,10,13,15,19**
15	**147.72**			
16	**112.68**	**4.72, s & 4.60, s**	**19**	**14,15,19**
17	**142.84**	**5.48, dd, (17.7, 10.9)**	**18**	**10,11,12**
18	**114.73**	**(cis) 5.01, d, (11.1) (trans) 4.84, d, (17.7)**	**17**	**11,17**
19	**18.82**	**1.32, s**	**16**	**14, 15,16**
20	**28.43**	**1.02, s**		**11,12,13,17**
21	**161.25**			

Table S3.35: 5-fluoro-7-aza-12-*epi*-hapalindole C (**3.32**)
$[\alpha]_D^{25}$=+17.3 (c=0.05, CH_2Cl_2)
Calc. $[M+H]^+$ 324.1871 Obsv. 324.1867



Position	^{13}C shift (ppm) (Fluorine coupling, J)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**7.64, bs**	**2**	
2	**126.46**	**6.89, d, (2.6)**	**1**	**3,7,8**
3	**112.74**			
4	**111.38, d, (21)**	**7.31, dd, (9.1, 3.0)**	**6**	**3,5,7**
5	**156.21, d, (242)**			
6	**132.19, d, (29)**	**8.29, t, (2.1)**	**4**	**4,7**
7	**145.27**			
8	**119.07**			
9	**36.50**	**3.08, dq, (10.9, 3.4)**	**10,14**	**2,3,8,14**
10	**66.16**	**3.27, s**	**9,12**	**9,11,12,14**
11	**40.29**			
12	**30.72**	**1.73, td, (13.7, 4.5) & 1.37, m**	**10,13,20**	**11,17,20**
13	**28.54**	**1.52, m & 1.46, m**	**12,14,20**	**12,14,17**
14	**43.64**	**2.80, td, (12.0, 4.2)**	**9,13,20**	**3,9,13,15,16,19**
15	**147.05**			
16	**112.77**	**4.67, s & 4.58, s**	**19**	**15,19**
17	**142.54**	**5.39, dd, (17.7, 11.0)**	**18**	**10,11,12**
18	**114.91**	**(cis) 4.95, d, (11.0) (trans) 4.78, d, (17.7)**	**17**	**11,17**
19	**18.79**	**1.28, s**	**16**	**14,15,16**
20	**28.40**	**1.00, s**	**10,13,14**	**10,11,12,17**
21	**161.34**			

Table S3.36: HpiC1→FamC1 mutagenesis percent conversion

Substrate	WT HpiC1	WT FamC1	V51I	F138L	L147F	F138LL147F
3.12	>99%	>99%	70%	92%	33%	56%
3.13	>99%	>99%	76%	31%	19%	>99%
3.10	>99%	20%	60%	26%	>99%	42%
3.11	77%	30%	13%	39%	81%	73%
3.14	85%	10%	67%	31%	18%	13%
3.15	90%	12%	11%	52%	27%	47%
3.8	14% tri, 52% tet	<1% tri, 19% tet	4% tri, 63% tet	<1% tri, 36% tet	<1% tri, 47% tet	<1% tri, 17% tet
3.9	N/A	N/A	N/A	N/A	N/A	N/A
3.16	30%	N/A	30%	8%	10%	N/A
3.19	>99%	60%	30%	40%	54%	35%
3.6	N/A	N/A	N/A	N/A	N/A	N/A
3.17	93%	70%	8%	9%	5%	5%
3.18	45%	N/A	3%	2%	<1%	N/A

Table S3.37: HpiC1→FimC5 mutagenesis percent conversions and hapalindole/fischerindole product ratio

Substrate	WT HpiC1	WT FimC5	Y101F	F138S	Y101FF138S
3.12	>99% 3.22	>99% 3.33	60% 3.22	9% 3.33, 29% 3.22	37% 3.33, 22% 3.22
3.13	>99% 3.23	>99% 3.37	57% 3.23	14% 3.37, 18% 3.23	27% 3.37, 5% 3.23
3.10	>99% 3.24	20% 3.34	36% 3.24	N/A	1% 3.34
3.11	77% 3.25	58% 3.38	59% 3.25	10% 3.38, 59% 3.25	12% 3.38, 10% 3.25
3.14	85% 3.26	N/A	18% 3.26	N/A	N/A
3.15	90% 3.27	>99% 3.39	34% 3.27	22% 3.39, 4% 3.27	50% 3.39, 20% 3.27
3.8	14% tri, 52% 3.28	>99% 3.35	2% tri, 18% 3.28	20% 3.35	30% 3.35
3.9	N/A	N/A	N/A	N/A	N/A
3.16	30% 3.29	N/A	17% 3.29	N/A	N/A
3.19	>99% 3.30	77% 3.40	18% 3.30	6% 3.40, 14% 3.30	10% 3.40, 3% 3.30
3.6	N/A	N/A	N/A	N/A	N/A
3.17	93% 3.31	59% 3.36 , 39% 3.31	13% 3.31	10% 3.36, 10% 3.31	13% 3.36, 6% 3.31
3.18	45% 3.32	14% 3.32	73% 3.32	24% 3.32	24% 3.32

Chapter 4

Ambiguine E-ring synthesis Part I: Rieske-type oxygenase exploration

4.1 Abstract

In this chapter and the following chapter, experiments exploring the synthesis of the fifth (E) ring in pentacyclic ambiguines will be discussed. The ambiguines are a related family of compounds to the hapalindoles and fischerindoles. While they all contain the hapalindole core, the ambiguines possess further tailoring including the addition of a fifth (E) ring. To date, the biosynthetic formation of this ring remains uncharacterized. This chapter will explore experiments conducted to elucidate the biosynthesis of the E-ring through examining Rieske-type oxygenases found in the ambiguine biosynthetic gene cluster. While a majority of these experiments were unsuccessful, they continue the foundation of experiments already conducted and showcase future experiments that still need to be performed.

4.2 Introduction

The ambiguine metabolites are a closely related class of compounds within the hapalindole family. Originally isolated in 1992[1], the ambiguines were shown to possess some more unique structural characteristics than the hapalindoles. The ambiguines show varying levels of oxidation on their core structure but the most prominent feature was the addition of a 7-membered fifth (E) ring. Due to isolation of tetracyclic ambiguines (containing a reverse prenylation with dimethylallyl at the indole C-2 position) it was hypothesized that this E-ring was derived from the dimethylallyl group.[1] However, it was unknown at the time what enzymes were responsible for this carbocyclization reaction. The intrigue in the biosynthetic mechanism only grew when Orjala's group isolated the fischambiguines containing a 6-membered E-ring (Figure 4.1).[2] In order to address this question and the question on hapalindole biosynthesis in general, the biosynthetic gene clusters (BGCs) of various ambiguine produce cyanobacterial strains were sequenced and annotated (Figure 4.1).[3-7] These annotations revealed a suite of

Rieske-type proteins believed to be responsible for the cyclization reaction through a radical based mechanism (Figure 4.2).[8]

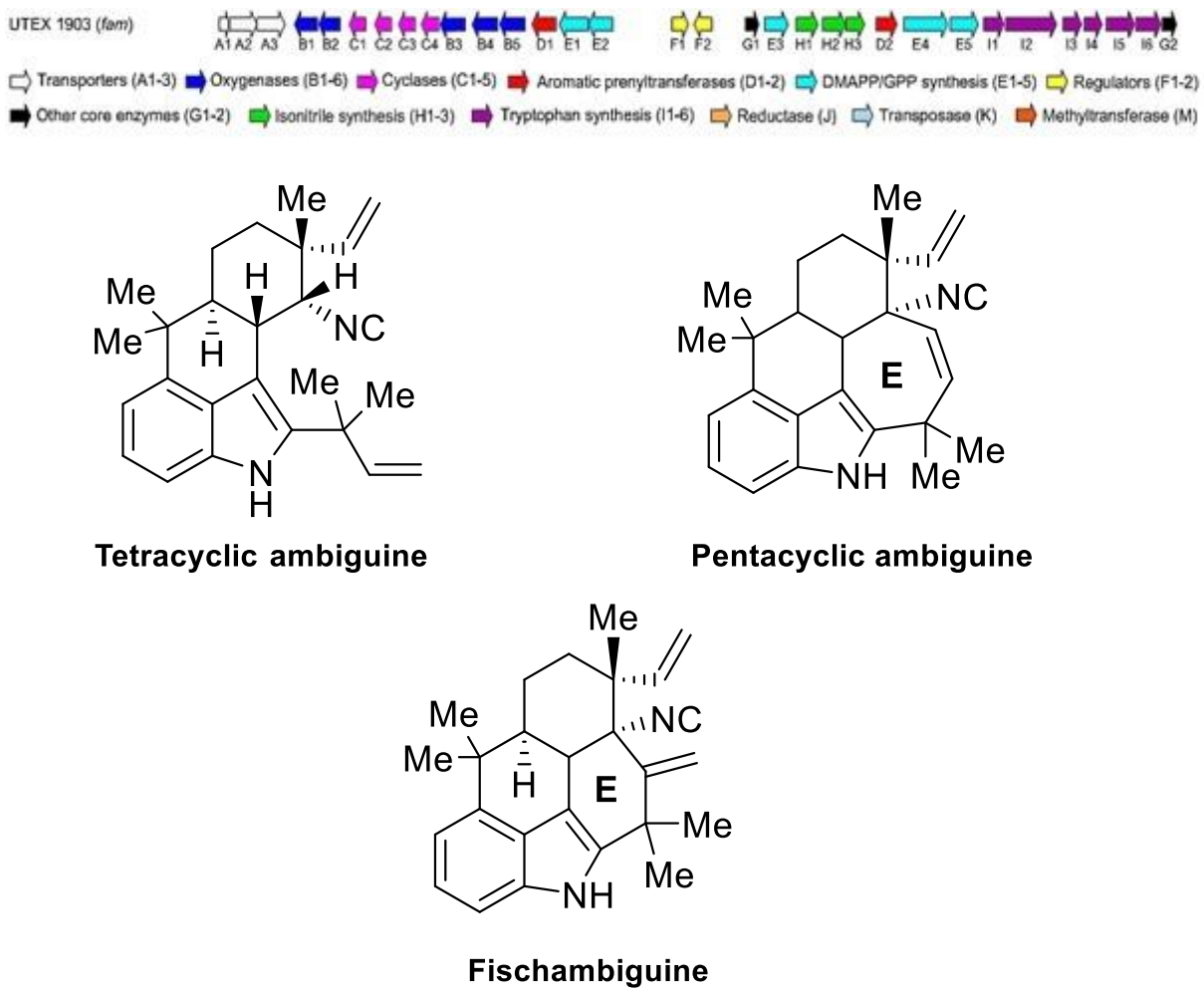


Figure 4.1: Representative BGC and ambiguine compounds. **Top**: The BGC from *F. ambigua* UTEX 1903 highlighting the different annotated enzymes. The blue arrows represent annotated Rieske-type oxygenases. **Bottom**: Representative examples of the three different types of ambiguines that have been observed.

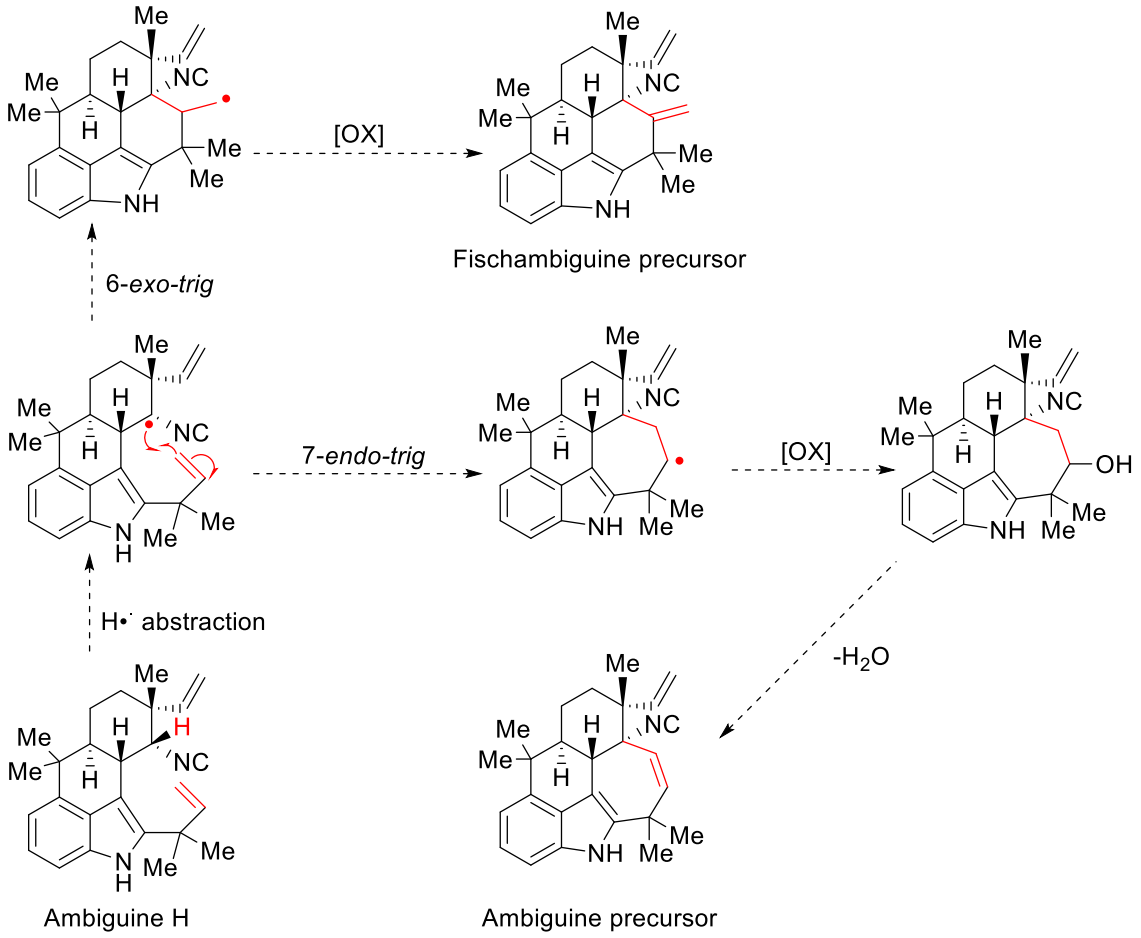


Figure 4.2: Proposed E-ring synthesis mechanism. Key carbons and bonds are highlighted in red.

Rieske non-heme iron oxygenases were first isolated in 1964 by John Rieske and co-workers; however, their function wasn't elucidated until 1968 in isolated fractions from *Pseudomonas putida*. The proteins were responsible for the *cis*-dihydroxylation in the metabolism of certain arenes. Molecular oxygen was identified as the source for the hydroxyl groups and NAD(P)H was identified as a reducing agent for the reaction. This differed from known P450 mechanisms at the time.[9] Rieske proteins, unlike P450s, contain no porphyrin system for binding iron. Instead, Rieske proteins contain iron-sulfur clusters. These can range from [2Fe-2S] up to [4Fe-4S] however [2Fe-2S] is the most commonly observed complex. The two irons are normally coordinated by two cysteine residues and two histidine residues; however, complexes containing four cysteine residues have been observed in plants (Figure 4.3).[9] All Rieske systems contain at least two protein components, a reductase and the oxygenase. Some systems also contain a ferredoxin protein to facilitate more efficient election transportation

between the oxygenase and reductase.[9] A specific amino acid sequence (known as the Rieske domain) is conserved across all Rieske-type proteins. Rieske proteins have been shown to perform a wide variety of reactions including oxidation, *O*-demethylation and carbocyclization.[10]

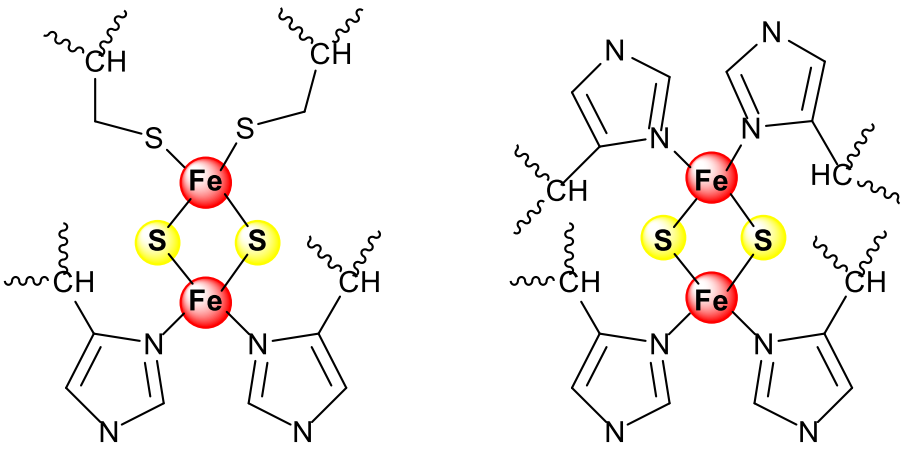


Figure 4.3: Examples of [2Fe-2S] clusters found in Rieske proteins. The first shows the most common coordination of the irons with two histidine and two cysteine residues. The second shows the coordination of the irons with four histidine residues.

In 2014, Hillwig and Liu were the first to elucidate the function of one these annotated "oxygenases". They showed that WelO5 (from *Hapalosiphon welwitschii*) was able to chlorinate 12-*epi*-hapalindole C (**4.1**) and 12-*epi*-fischerindole U (**4.2**). While WelO5 was actually an Fe(II)/α-ketoglutarate dependent protein containing no Rieske domain, it was an important discovery as this was one of the first protein independent halogenases to be characterized.[11] A couple years later, they showed the same result with AmbO5 (also known as FamB1) from an ambiguine producing strain *Fischerella ambigua*. AmbO5 was shown to more promiscuous in its substrate scope by being able to chlorinate a variety of hapalindole and ambiguine substrates, most notably hapalindole U (**4.3**) and ambiguine H (**4.4**) (Figure 4.4).[12] However, Liu and co-workers went back to WelO5 to show case its promiscuity in halogenating substrates with both chlorine and bromine.[13] A crystal structure of WelO5 has been obtained[14] and further mutagenesis work has been undertaken to halogenate non-native substrates[15,16] but these will not be discussed in further detail here.

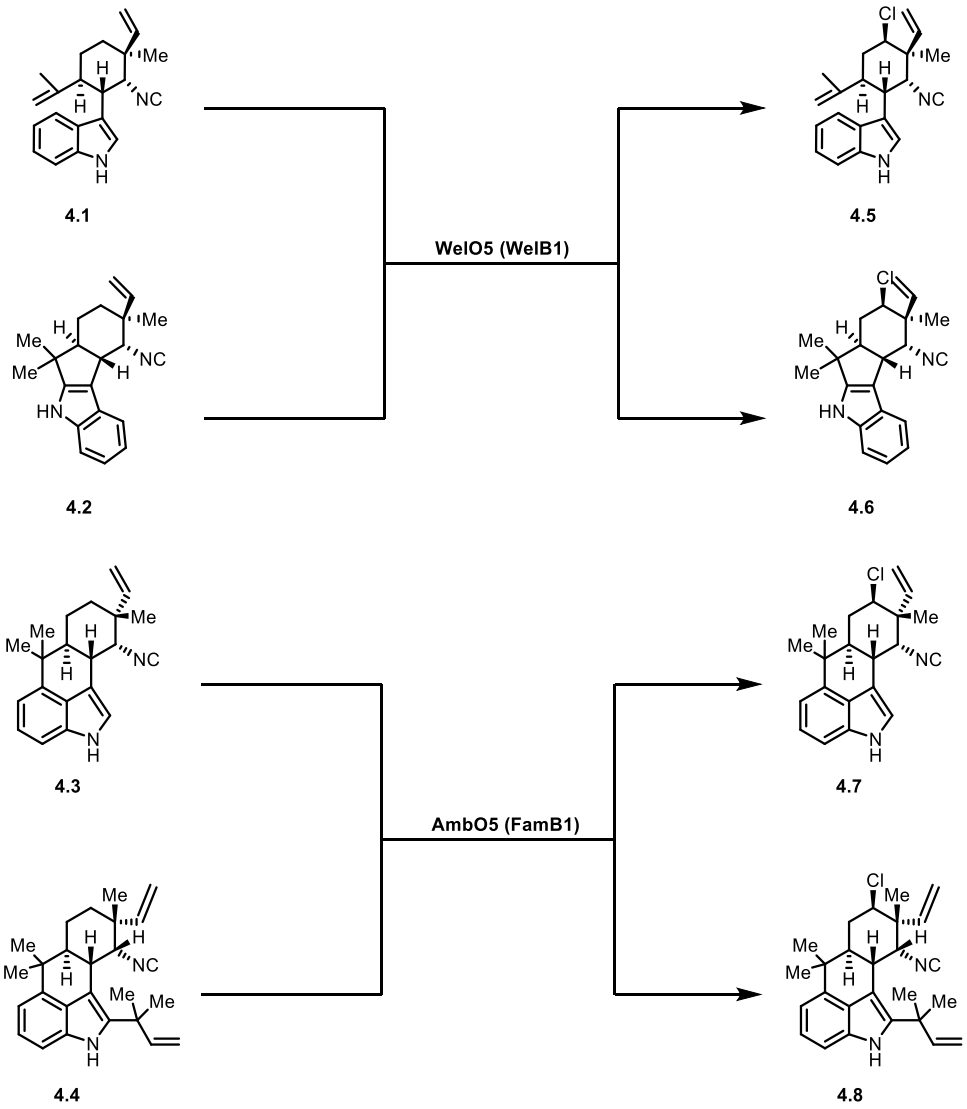


Figure 4.4: Characterized Fe(II)/α-ketoglutarate halogenases. WelO5 (WelB1) has been shown to chlorinate 12-*epi*-hapalindole C (**4.1**) and 12-*epi*-fischerindole U (**4.2**) resulting in 12-*epi*-hapalindole E (**4.5**) and 12-*epi*-fischerindole G (**4.6**) respectively. AmbO5 (FamB1) has been shown to chlorinate hapalindole U (**4.3**) and ambiguine H (**4.4**) resulting in hapalindole G (**4.7**) and ambiguine A (**4.8**) respectively.

In this chapter, work regarding the expression, isolation and attempted characterization of Rieske-type proteins from select cyanobacterial BGCs will be discussed. While this work was not successful in elucidating a function for one of the annotated proteins, it lays the groundwork for future experiments.

4.3 Results and discussion

4.3.1 Cell free lysate and previous purified proteins

Initial experiments involved examining the cell free lysate of *F. ambigua* UTEX 1903. This strain has had numerous ambiguines isolated from it[1] and it was the BGC where the Stig cyclases FamC1/C2/C3/C4 had been characterized from.[4,5] To start, similar experiments to the ones conducted by Li and co-workers were undertaken.[4] These involved growth of the cyanos, cell lysis and then screening of the cell free lysate with additional substrates and cofactors. Due to the hypothesis that the E-ring is derived from a cyclization involving a tetracyclic ambiguine, ambiguine H was the first substrate elected to be screened. Various cofactors such as α-ketoglutarate, NADP+, ferredoxin/ferredoxin reductase and exogenous iron were added to the reaction mixture. Initial HPLC results showed the formation of a new product. This new product was observed in every reaction that contained α-ketoglutarate. This result was unsurprising given that AmbO5 and WelO5 were both characterized as Fe(II)/α-ketoglutarate dependent halogenases.[11,12]

However, the annotated Rieske proteins were not believed to require α-ketoglutarate for activity, only exogenous iron. Upon closer examination of the BGC, the X2 proteins (FamB2, WelB2, WepB2, etc.) are not classified as Rieske proteins but instead classified as mononuclear non-heme iron proteins which may still require α-ketoglutarate. A close homolog of FamB2, WepB2 (from *Westiellopsis prolifica*) was previously purified. Reactions containing WepB2 and α-ketoglutarate produced the same results as the cell free lysate reactions hinting towards the activity coming from FamB2. However, the observed mass of $[M+H]^+$ 391.2748 was not the expected mass for a cyclization, oxidation or halogenation reaction. Upon scale up for characterization, it was determined that the new product was actually ambiguine H formamide (**S4.1**) (Supplemental information figure S4.1), a product resulting from the hydrolysis of the isonitrile in an acdic environment. Reactions were re-run in a higher concentration of pH 7.0 buffer with no reactivity observed. Further reactions with hapalindole substrates (hapalindole U, H and 12-*epi*-hapalindole U) revealed no reactivity as well. To validate all results, a protein gel was run containing various cell lysates with very little protein expression (Supplemental information Figure S4.2).

4.3.2 FamB2 expression and screening

While the cell free lysate reactions did not produce a positive result, we were encouraged about the idea of purified protein providing the solution. FamB2 had been previously

heterologously expressed and purified in our group. This batch of protein was screened with various substrates and cofactors; however, no positive results were observed. This was believed to be because the batch of FamB2 was old and no longer functionable. Expression and purification of new FamB2 was undertaken using conditions previously reported in our group. However, expression levels of the new FamB2 remained low (Supplemental information Figure S4.3). For confirmation, the new FamB2 was screened but no positive results were obtained. In order to address this low expression, growth conditions were modified but no improvement was observed. We elected to obtain a synthetic plasmid from Twist Biosciences with the amino acid codons being optimized for *E. coli*. Expression and purification of this new codon optimized plasmid led to much greater expression levels (Supplemental information Figure S4.4). This new batch of protein was screened with various substrates under various conditions but no positive results were observed (Experimentals S4.5.4).

4.3.3 FamB3/B4 expression, screening and PrnD condition exploration

With FamB2 producing no positive results, attention was turned towards the annotated Rieske proteins of the BGC, FamB3 & B4. As was the case with FamB2, initial expression of cloned *famB3* and *famB4* genes were very low (Supplemental information Figures S4.5 & S4.7). Use of different cell lines and growth conditions could not improve expression levels. We elected to order *E. coli* codon optimized plasmids from Twist Biosciences, just as we had with *famB2*. The expression levels of these new codon optimized plasmids were not much greater than previously observed but we moved ahead with the newly purified protein batches (Supplemental information Figures S4.6, S4.8 & S4.9). Various reaction conditions and substrates did not produce any positive results.

With no positive results obtained, a BLAST search was conducted looking for any characterized homologous proteins. While majority of the results were uncharacterized, one protein, PrnD, provided some homology (55% by sequence) to FamB3 and FamB4. PrnD is a Rieske oxygenase from *Pseudomonas fluorescens* that catalyzes the

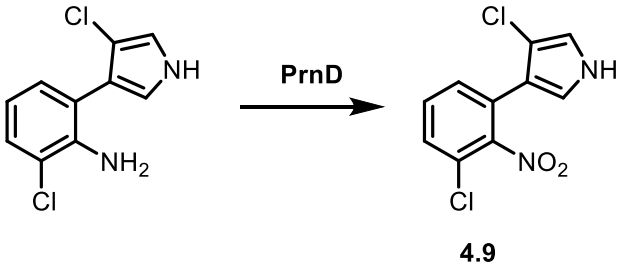


Figure 4.5: Function of PrnD.[17] The aryl amine of the pyrrolnitrin precursor is converted into an aryl nitro group to give the final compound **4.9**.

conversion of an aryl amine to an aryl nitro group in the natural product pyrrolnitrin (**4.9**) (Figure 4.5).[17] The two main findings from this work were that the protein's function had to be reconstituted anaerobically and that it utilized a Flavin recycling system. Up to this point, we had only used a spinach ferredoxin/ferredoxin reductase electron recycling system. Also, the practice of reconstituting Rieske proteins function anaerobically is common but seems counterintuitive given that the proteins need molecular oxygen to function.[18] Regardless, we forged ahead reconstituting FamB2, B3 and B4 all anaerobically (O_2 levels<40 ppm) using iron(II) sulfate and sodium sulfide.[17] Initial anaerobic reconstitution of FamB2 appeared successful based on color change. FamB3 and B4 did not appear to reconstitute as efficiently but all reconstituted proteins were moved forward for screening. However, reactions with the new protein and Flavin recycling system showed no positive results with various substrates. While we observed ample FamB2 expression, FamB3 and B4 remain difficult to obtain adequate levels of expression. Thus, it is unknown if the lack of function with the two Rieske proteins is due to the very low expression levels.

4.4 Conclusion

In this chapter, experiments related to Rieske-type proteins FamB2, B3 and B4 were discussed. While results appeared promising early on, these proteins continue to be difficult to elucidate. Various experiments under a variety of conditions with a variety of substrates were conducted with no positive results observed (Experimentals S4.5.4). The greatest challenge to elucidating these proteins functions continues to be optimizing their expression and reconstruction. Further dives into Rieske protein literature may uncover a hint about how to crack the code of these proteins. However, efforts were turned towards a chemoenzymatic synthesis route to the E-ring which is discussed further in the next chapter.

4.5 Experimentals

4.5.1 General

Analytical HPLC analysis was performed on a Shimadzu 2010 EV APCI spectrometer equipped with an LUNA C18 250 x 4.6 mm column, using a mobile phase gradient of either 60-100% acetonitrile in water over 16 min or 55-100% acetonitrile in water over 16 min at 40°C and was monitored by UV absorption at 280 nm. LC-MS analysis was performed on a Agilent

Infinity II QTOF using an Phenomenex C4 Aeris Widepore 3.6 µ 50 x 2.10 mm column, using a mobile phase gradient of 10-100% acetonitrile in water over 8 min.

Escherichia coli strain BL21(DE3) and BL21-AI pRARE was used for protein expression. Plasmid pET28H8T[4] was used for cloning and expression of N-truncated FamB2, B3 and B4. Plasmid pET28b *E. coli* codon optimized plasmids with an N-terminal TEV cleavage site were obtained from Twist Biosciences. Isopropyl β-D-thiogalactopyranoside (IPTG) was used to induce expression; DNAse and lysozyme were purchased form Sigma-Aldrich. Ni-NTA agarose from Invitrogen was used to purify His-tag proteins.

All chemicals were purchased from Sigma-Aldrich, ACROS, and Combi-Blocks. Chemical abbreviations: Ethyl Acetate (EtOAc), Acetonitrile (CH_3CN), Magnesium Chloride ($MgCl_2$), Calcium Chloride ($CaCl_2$), Iron(II) Sulfate heptahydrate ($FeSO_4 \bullet 7H_2O$), Iron(II) Citrate ($FeC_6H_6O_7$), Sodium Sulfide (Na_2S), Glucose-6-phosphate (G6P), Glucose-6-phosphate dehydrogenase (G6PDH), α-ketoglutarate (α-KG), Nicotinamide adenine dinucleotide phosphate (NADP+), Reduced nicotinamide adenine dinucleotide phosphate (NADPH), ferredoxin (Fdx), ferredoxin reductase (Fdr), Flavin mononucleotide (FMN), Phenylmethylsulfonyl fluoride (PMSF), Sodium Chloride (NaCl)

4.5.2 Protein expression and purification

The expression and purification of proteins was performed as described. Briefly, a single BL21(DE3) or BL21-AI pRARE colony was inoculated in LB medium containing 50 µg/mL kanamycin and grown overnight at 37 °C shaking at 200 rpm. The main culture (1 L) was inoculated at the dilution of 1:100 in 2.8 L of Fernbach flask containing TB medium and the same concentration of antibiotic. Exogenous $FeSO_4 \bullet 7H_2O$ (0.4 mg/mL) and $FeC_6H_6O_7$ (0.2 mg/mL) were added to the growth media. For BL21-AI pRARE growths, exogenous *L*-(+)-arabinose (2 g/L) was added as well. The cells were grown (37 °C, 180 rpm) to an optical density (A_{600} nm) of 1.0. The culture flasks were chilled in ice, induced with IPTG (1 mM), and were further incubated (16-20°C, 180 rpm) for 16 h. The cells were harvested (6500 x g, 4 °C, 30 min), flash frozen, and stored at −80 °C until purification. The cell pellets were resuspended at 4 °C in the lysis buffer (50 mM Tris, 500 mM NaCl, 0.2 mM TCEP, 20 mM Imidazole, 10%

Glycerol, pH 7.4), containing 0.5 mg/mL of lysozyme, 1 mM PMSF and 1 mL of 2 mg/mL DNAse. The mixture was stirred for 30 min and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The cellular debris was removed by centrifugation (65,000 x g, 4 °C, 30 min). The clarified lysate was loaded onto Ni-NTA agarose column equilibrated with lysis buffer. The column was washed with two column volume of wash buffer (50 mM Tris, 500 mM NaCl, 0.2 mM TCEP, 20 mM Imidazole, 10% Glycerol, pH 7.4) and the His-tagged protein was eluted with elution buffer (50 mM Tris, 500 mM NaCl, 0.2 mM TCEP, 300 mM Imidazole, 10% Glycerol, pH 7.4). The fractions were pooled and dialyzed overnight or by using a PD10-desalting column (GE Healthcare) using storage buffer (50 mM Tris, 50 mM NaCl, 0.2 mM TCEP, 10% Glycerol, pH 7.4). The purified protein was analyzed by SDS-PAGE gel for purity, measured by Nanodrop using a calculated molar extinction coefficient for concentration, and flash-frozen in liquid nitrogen to store at −80 °C.

4.5.3 Anaerobic reconstitution of iron-sulfur clusters

Anaerobic reconstitution was performed as described.[15] The reconstitution was carried out in an anaerobic chamber (COY Laboratory Products INC. Anaerobic Chamber) with O_2 levels<40ppm. All solution were sparged with N_2 for 30 min before being added to chamber. The purified protein was diluted to 10-20 µM with sparged storage buffer (50 mM Tris, 50 mM NaCl, 0.2 mM TCEP, 10% Glycerol, pH 7.4). $FeSO_4 \bullet 7H_2O$ and Na_2S were added to the solution at a final concentration of 1.0 mM. The solution was gently mixed every 15 minutes for 2 h. The solution was diluted 8-fold with sparged storage buffer, equilibrated in air and desalted to remove excess Fe^{2+} and S^{2-}. The reconstituted proteins were concentrated, flash frozen in liquid nitrogen and stored at -80 °C. The iron-sulfur cluster incorporation was monitored by color change to a yellow-brown.

4.5.4 Enzymatic reactions

Enzymatic reactions were run under a variety of conditions. Substrates tested include ambiguine H, 12-*epi*-ambiguine H, hapalindole U, hapalindole H, 12-*epi*-hapalindole U, 12-*epi*-hapalindole C and 12-*epi*-fischerindole U. Each reaction condition will be listed separately:

Cell free lysate reactions: A variety 100 or 150 µL reactions were ran containing various mixtures of the following: 500 µM substrate, 250 mM α-KG, 1 mM NADP+, 40 µM Fdx, 6 µM

Fdr, 100 mM $FeSO_4 \bullet 7H_2O$, 5 mM G6P, 1 U/mL G6PDH, 1 mM $MgCl_2$, 7.5 mM $CaCl_2$ and the remaining volume was diluted with cell free lysate. All Reactions were ran at 30 °C and 160 rpm for 3 hrs. Quenched with 3x volume EtOAc, dried under N_2, dissolved in 100 µL CH_3CN and prepped for HPLC analysis.

FamB2/B3/B4 reactions: A variety of 150 µL reactions were ran containing various mixtures of the following: 1 mM substrate, 20 µM FamB2/B3/B4, 5 mM α-KG, (in reactions not reconstituted: 5 mM $FeSO_4 \bullet 7H_2O$, 5 mM Na_2S), 1 mM NADP+, 5 mM G6P, 1 U/mL G6PDH, 40 µM Fdx, 6 µM Fdr, 50 mM Tris buffer pH 7.0. In reactions with Flavin recycling system, 100 µM NADPH, 3 µM FMN and 80 µM HpaC (Flavin reductase). All reactions were run at 30 °C and 160 rpm for 3 hrs. Quenched with 3x volume EtOAc, dried under N_2, dissolved in 100 µL CH_3CN and prepped for HPLC analysis.

4.5.5 Cyanobacteria culturing and lysis

Cyanobacteria culturing: *Fischerella ambigua* UTEX 1903 was grown as previously described.[4] The cells were cultured in two 250 mL Erlenmeyer flasks containing Blue-Green 11 (BG-11) medium (2% v/v). 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, pelleted and lysed as described below.

Cyanobacteria lysis: Cells were lysed as previously described.[4] Pelleted cells were re-suspended at 4 °C in lysis buffer (100 mM Tris pH 8.0, 50 mM EDTA, 100 mM NaCl) containing 2 mg/mL DNAse and 0.5 mg/mL lysozyme. The mixture was stirred for 30 min and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The cellular debris was removed by centrifugation (65,000 x g, 4 °C, 30 min). The supernatant was flash frozen with liquid N_2 and stored at -80 °C.

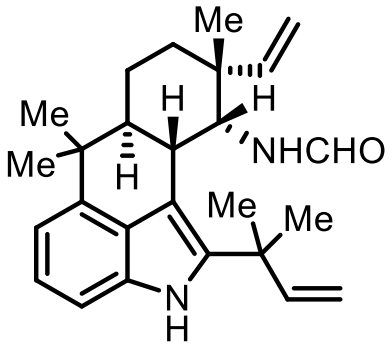
4.6 References

1. 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. https://doi.org/10.1021/jo00029a014.
2. 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. https://doi.org/10.1016/j.phytochem.2010.09.004.
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. https://doi.org/10.1021/cb400681n.
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. https://doi.org/10.1021/jacs.5b10136.
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. https://doi.org/10.1038/nchembio.2327
6. Li, S.; Newmister, S. A.; Lowell, A. N.; Zi, J.; Chappell, C. R.; Yu, F.; Hohlman, R. M.; Orjala, J.; Williams, R. M.; Sherman, D. H. Control of Stereoselectivity in Diverse Hapalindole Metabolites Is Mediated by Cofactor-Induced Combinatorial Pairing of Stig Cyclases. *Angew. Chem. Int. Ed.* **2020**, *59* (21), 8166–8172. https://doi.org/10.1002/anie.201913686.
7. 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* (1), 213. https://doi.org/10.1186/s12866-014-0213-7.
8. Hohlman, R.M.; Sherman, D.H. Recent advances in hapalindole-type cyanobacterial alkaloids: biosynthesis, synthesis and biological activity. *Nat. Prod. Rep.* **2021**, *38* (9), 1567-1588. https://pubs.rsc.org/en/content/articlelanding/2021/np/d1np00007a.
9. Ferraro, D.J.; Gakhar, L.; Ramaswamy, S. Rieske business: Structure-function of Rieske non-heme oxygenases. *Biochem. Biophys.Res. Commun*. **2005**, *338* (1), 175-190. http://www.sciencedirect.com/science/article/pii/S0006291X05019480.
10. Barry, S.M.; Challis, G.L. Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS Catal.* **2013**, *3* (10), 2362-2370. https://doi.org/10.1021/cs400087p.
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. https://doi.org/10.1038/nchembio.1625.
12. 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. https://doi.org/10.1002/cbic.201500674.

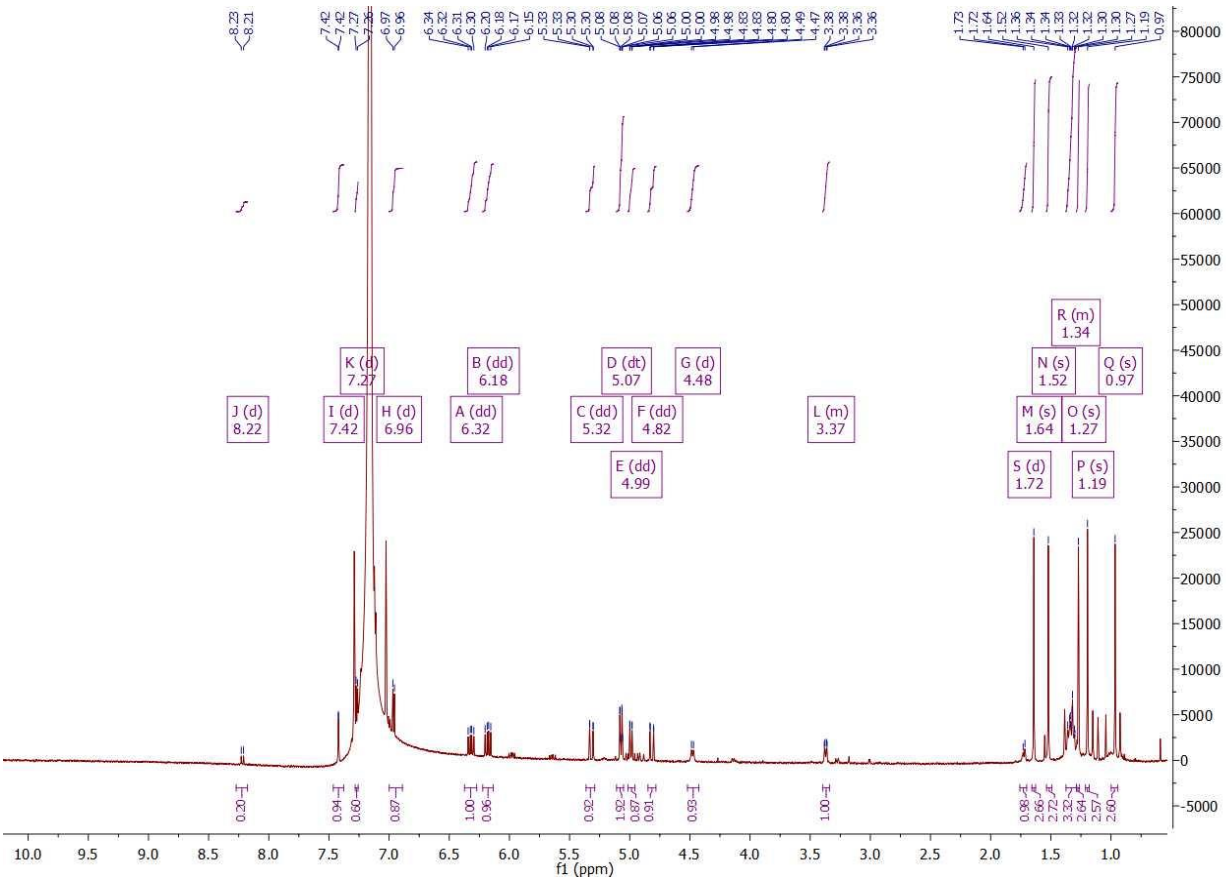
13. Hillwig, M. L.; Zhu, Q.; Ittiamornkul, K.; Liu, X. Discovery of a Promiscuous Non-Heme Iron Halogenase in Ambiguine Alkaloid Biogenesis: Implication for an Evolvable Enzyme Family for Late-Stage Halogenation of Aliphatic Carbons in Small Molecules. *Angew. Chem. Int. Ed.* **2016**, *55* (19), 5780–5784. https://doi.org/10.1002/anie.201601447.
14. 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. https://doi.org/10.1038/nchembio.2112.
15. Duewel, S.; Schmermund, L.; Faber, T.; Harms, K.; Srinivasan, V.; Meggers, E.; Hoebenreich, S. Directed Evolution of an Fe(II)-Dependent Halogenase for Asymmetric C(sp3)-H Chlorination. *ACS Catal.* **2020**, *10* (2), 1272-1277. https://doi.org/10.1021/acscatal.9b04691.
16. Hayashi, T.; Ligibel, M.; Sager, E.; Voss, M.; Hunziker, J.; Schroer, K.; Snajdrova, R.; Buller, R. Evolved Aliphatic Halogenases Enable Regiocomplementary C-H Functionalization of a Pharmaceutically Relevant Compound. *Angew. Chem. Int. Ed.* **2019**, *58* (51), 18535-18539. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201907245.
17. Lee, J.; Simurdiak, M.; Zaho, H. Reconstitution and Characterization of Aminopyrrolnitrin Oxygenase, a Rieske *N*-Oxygenase That Catalyzes Unusual Arylamine Oxidation* *J. Biol. Chem.* **2005**, *280* (44), 36719-36728. https://www.sciencedirect.com/science/article/pii/S002192582059306X.
18. Liu, X. Chapter Twelve-*In Vitro* Analysis of Cyanobacterial Nonheme Iron-Dependent Aliphatic Halogenases WelO5 and AmbO5. *Methods in Enzymology*. **2018**, 389-404. https://www.sciencedirect.com/science/article/pii/S007668791830082X.

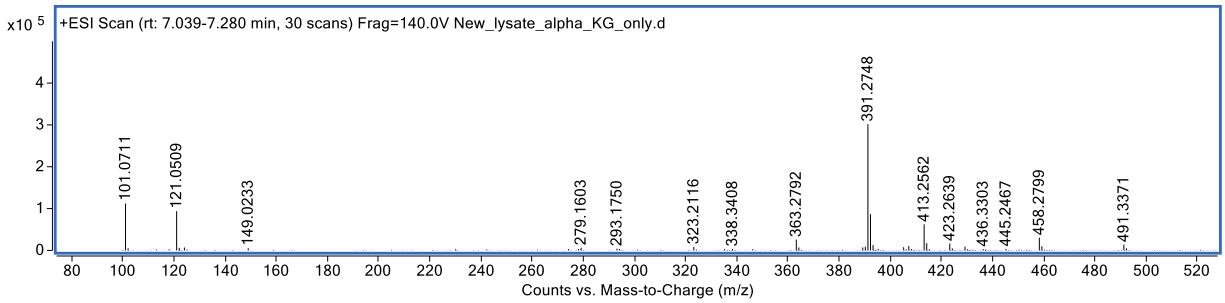
4.7 Supplemental information

Figure S4.1: Structure, H^1 NMR and HRMS spectra of ambiguine H formamide **(S4.1)**



MS: Calc. $[M+H]^+$ 391.2744 Obsv: 391.2748

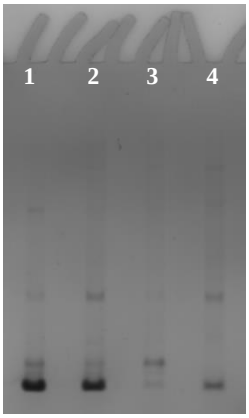




Figures S4.2-4.9: Protein gel images

Key: **L**=ladder, **CP**=Cell Pellet, **FT**=Flow Through, **R1**=Rinse 1, **R2**=Rinse 2, **R3**=Rinse 3 (if applicable), **E**=Elution Fraction

Figure S4.2: Cyano cell lysate protein gel. (no protein ladder)



1: *F. ambigua* growth 1

2: *F. ambigua* growth 2

3: *H. welwitschii*

4: *F. SAG 46.79*

Figure S4.3: FamB2 protein gel, non-codon optimized

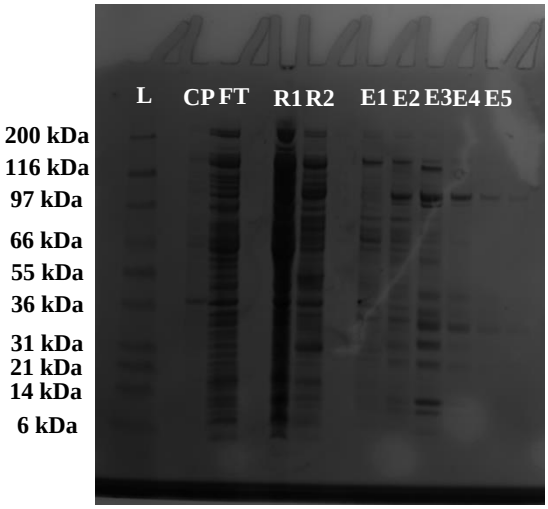


Figure S4.4: FamB2 protein gel, codon optimized

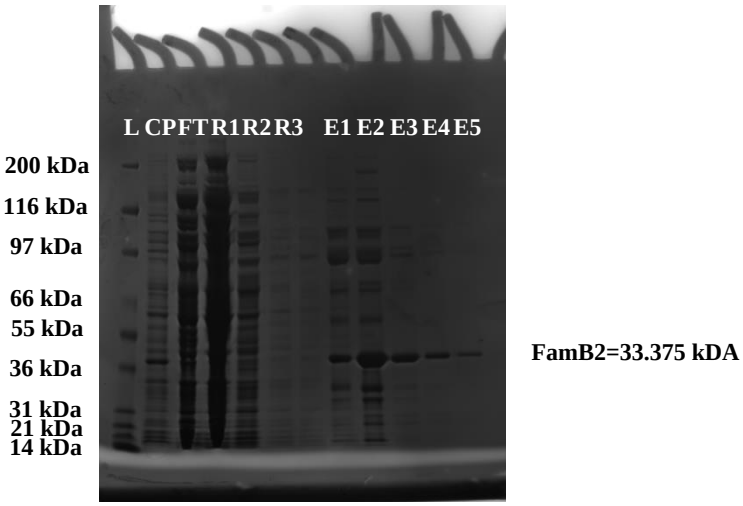


Figure S4.5: FamB3 protein gel, non-codon optimized

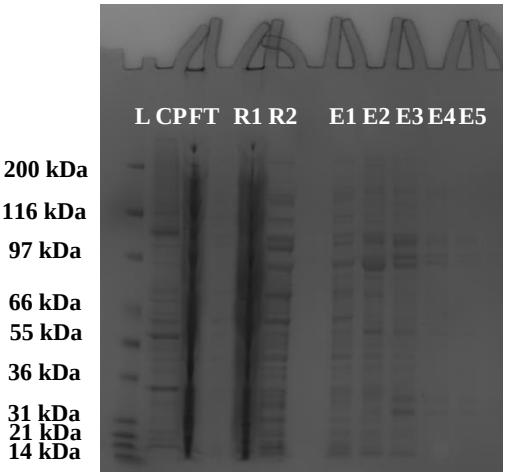


Figure S4.6: FamB3 protein gel, codon optimized

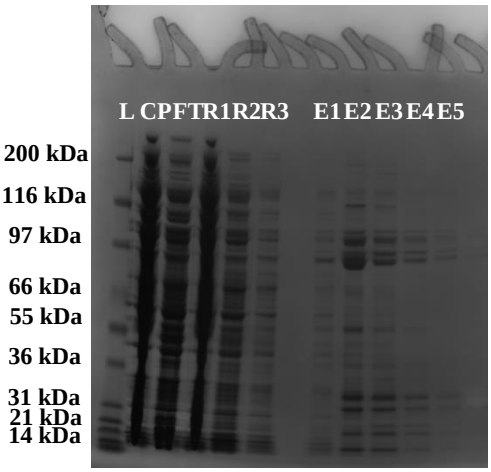


Figure S4.7: FamB4 protein gel, non-codon optimized

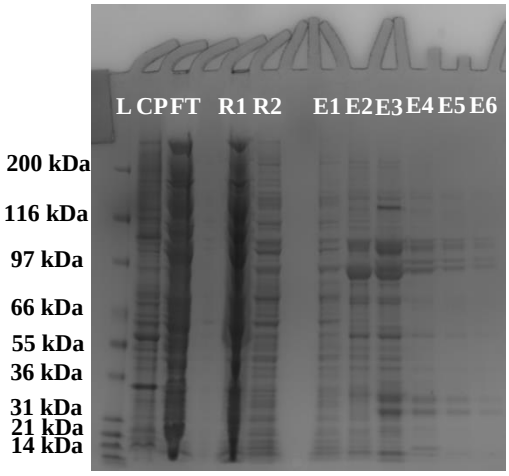


Figure S4.8: FamB4 protein gel, codon optimized. (inefficient ladder stain)

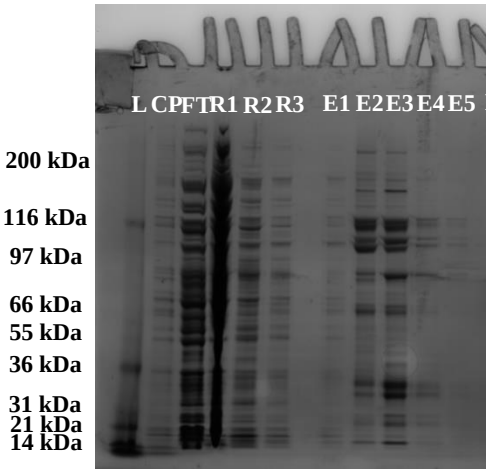
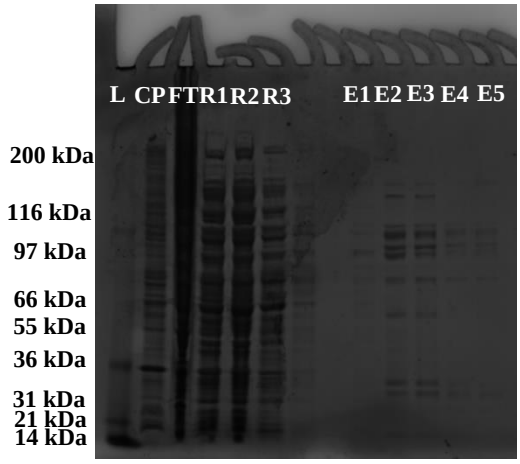


Figure S4.9: FamB3/B4 coexpression gel. (inefficient ladder stain)



Chapter 5

Ambiguine E-ring synthesis Part II: Chemoenzymatic synthesis and 11-DMAC discovery

5.1 Abstract

In a continuation of work from chapter 4, the second part of the ambiguine E-ring work will be presented. Recently, a new array of C-H functionalization chemistry has made previous reactions thought too difficult a reality. Proposing a C-H functionalization, or C-H activation, route to the synthesis of the ambiguine E-ring from a tetracyclic ambiguine now seems feasible. In order to accomplish this goal, the efforts to develop an efficient biocatalytic route for the production of an unnatural ambiguine derivative, 12-*epi*-ambiguine H nitrile, will be discussed. This substrate was provided to develop a C-H functionalization route to the production of the ambiguine E-ring and further synthesis efforts. During the biocatalytic route optimization, a unique shunt metabolite (11-DMAC) was uncovered. This metabolite further strengthens the Cope rearrangement hypothesis and warrants further exploration into its biosynthesis and uses.

5.2 Introduction

With the biosynthetic elucidation of the Rieske-type oxygenases being difficult, we elected to turn our attention towards another route for ambiguine E-ring synthesis. To date only the Sarpong group and Rawal group have been able to synthesize pentacyclic ambiguines.[1-3] Both ambiguine P (**5.1**) and ambiguine G nitrile (**5.2**) were synthesized from starting materials of indole and carvone derivatives (Figure 5.1). The Sarpong group utilized a cobalt catalyzed Nicholas reaction to form the E-ring of **5.1**.[1] In both of their total syntheses, the Rawal group utilized a [4+3] Diels-Alder type reaction to form the E-ring in **5.1** and **5.2**.[2-3]

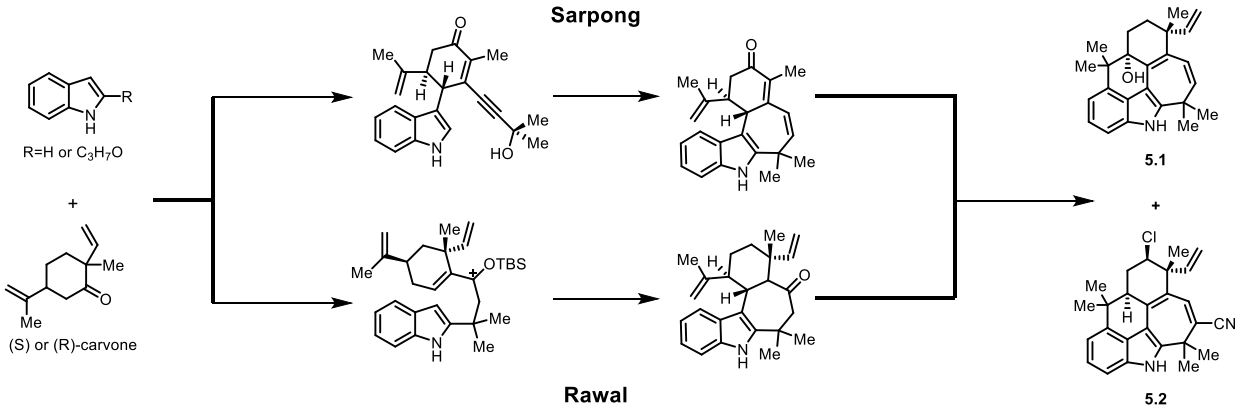
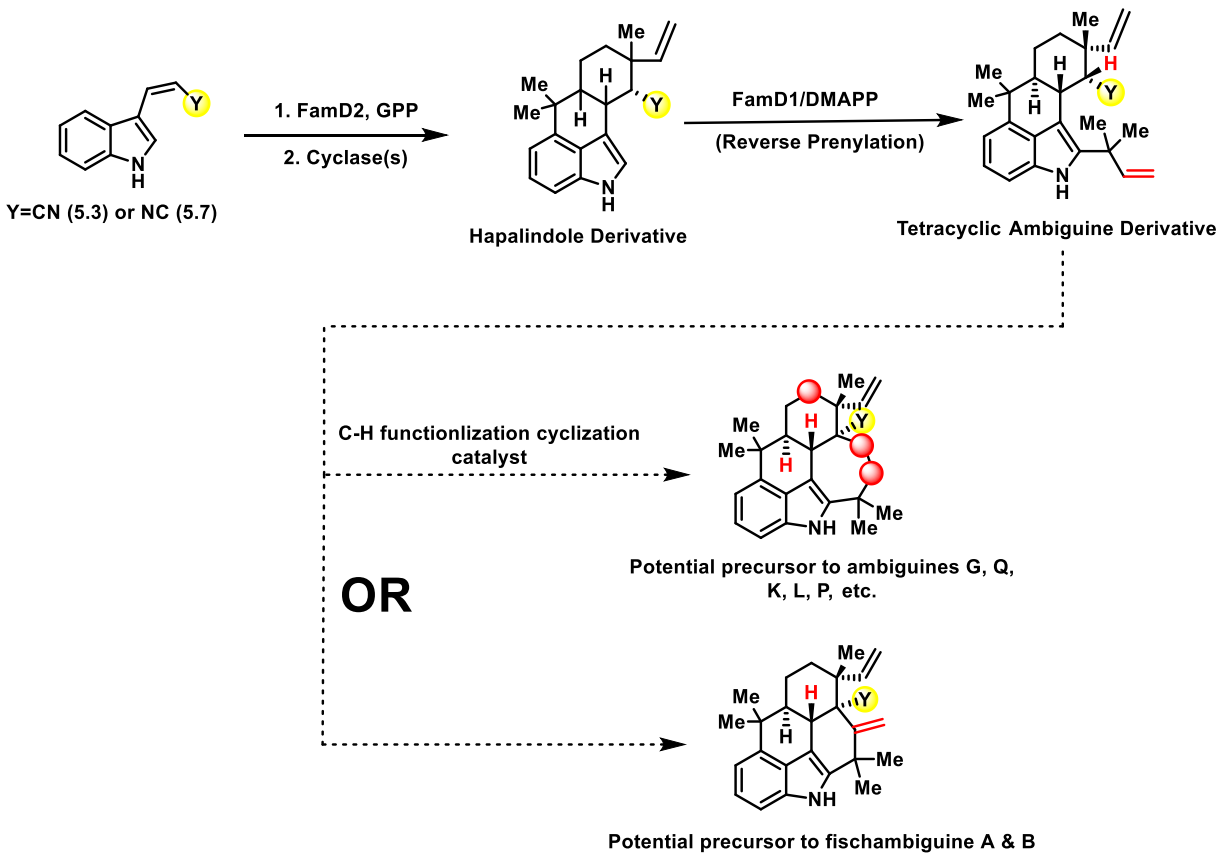


Figure 5.1: General methods used by Sarpong and Rawal for ambiguine E-ring synthesis. General starting materials and key intermediates are shown leading to the synthesis of ambiguine P **(5.1**) and ambiguine G nitrile (**5.2**).

While both strategies succeed in completing the ring, E-ring formation is believed to arise from a tetracyclic ambiguine. With great strides made recently in the field of C-H functionalization chemistry[4-6], we believed that a semisynthetic route could be produced from a tetracyclic ambiguine. As a member of the Center for Selective C-H Functionalization (CCHF), we brought this idea up for a molecule of the month (MOM) project. The resulting discussions were very informative and showcased to us that this route could be viable. To that end, we proposed a general route where the tetracyclic ambiguine compound is produced fully enzymatically (besides early synthesis of the *cis*-indole starting material). Then C-H functionalization or activation chemistry is used to close the E-ring and further synthetic chemistry allows for a total synthesis of an already known pentacyclic ambiguine (Scheme 5.1). This work would push the scale of enzymatic reactions up to levels not yet utilized and along the way, would lead to the discovery of a unique metabolite never before seen in the hapalindole pathway.



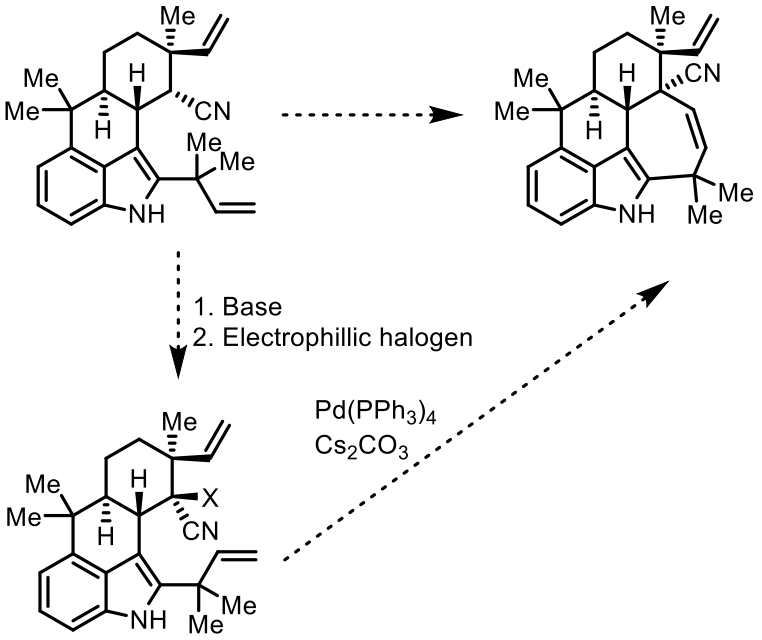
Scheme 5.1: Proposed route to ambiguine E-ring using C-H functionalization. Early stage enzymatic work leads to the formation of a tetracyclic ambiguine derivative followed by potential C-H functionalization/activation chemistry. Key areas for further chemistry are highlighted in red.

5.3 Results and discussion for chemoenzymatic synthesis

5.3.1 12-*epi*-ambiguine H nitrile formation and initial synthetic results

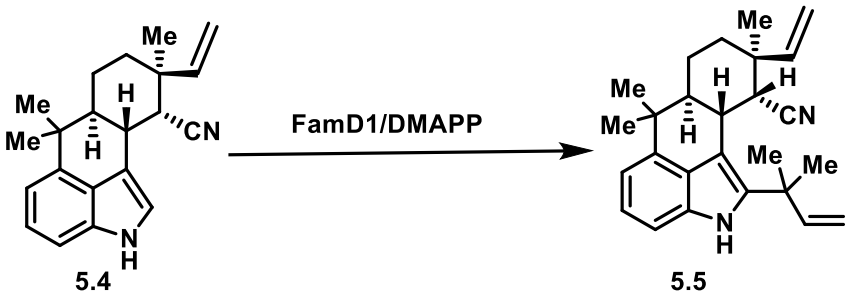
After discussions within the CCHF, we elected to continue our long standing collaboration with Bob Williams's group at Colorado State University. The Williams lab has had a history working on the total synthesis of hapalindole compounds[7-8] and we believed they would be the best group to work with moving forward; one of Bob's students, John, was willing to tackle this challenge. During initial discussions with John, the idea of a Heck coupling reaction was proposed to close the E-ring. However, it was believed that the reaction conditions for aliphatic halogenation at the C-11 position would easily hydrolyze the isonitrile. In order to address this concern, we proposed the idea of using a nitrile containing tetracyclic ambiguine in the reaction (Scheme 5.2). It was unknown if the nitrile would survive the reaction conditions but it was believed it had a stronger chance than the isonitrile.

While it was known that the *cis*-indole nitrile substrate (**5.3**) could be accepted for geranylation and cyclization into a tetracyclic hapalindole compound[9], it was unknown if this compound could be further reverse prenylated by FamD1. Initial analytical reactions with 12-*epi*-hapalindole U nitrile (**5.4**), purified FamD1 and dimethylallyl pyrophosphate (DMAPP) resulted in the formation of a new peak in the HPLC (Supplemental information Figure S5.1). Further scale up and characterization work revealed this new peak be the desired tetracyclic ambiguine compound, 12-*epi*-ambiguine H nitrile (**5.5**) (Scheme 5.3 and Supplemental information Table S5.1). While some ambiguine compounds do contain a nitrile moiety, all of the ambiguines contain hapalindole U stereochemistry (not 12-*epi*). However, it was easier to produce greater quantities of **5.4** and we elected to proceed using the 12-*epi* stereochemistry.



Scheme 5.2: Proposed Heck coupling to ambiguine E-ring. It was hypothesized that using the nitrile containing compound (in yellow) would be more stable to reaction conditions.

Initial methods to produce acceptable amounts of **5.5** proved challenging. While reactions using FamD2 and HpiC1 to produce **5.4** on 5-7 mg scale gave acceptable yields (30-35%), the reverse prenylation reaction with FamD1 was both slow and low yielding (10-15%). FamD1's expression is not substantial and a large amount of protein had to be used in each reaction in



Scheme 5.3: Enzymatic reverse prenylation reaction of unnatural substrates. 12-*epi*-hapalindole U nitrile (**5.4**) could be converted into 12-*epi*-ambiguine H nitrile (**5.5**) using the reverse prenyltransferase FamD1.

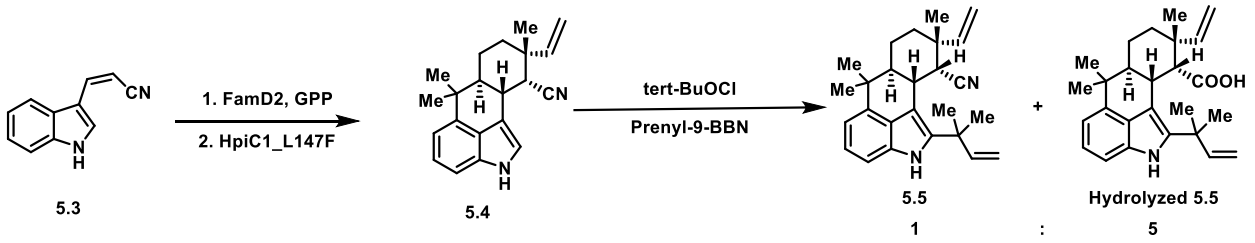
order to get greater than 50% conversion values. After numerous runs with purified protein, roughly 16 mg of **5.5** was sent to John for initial synthetic experiments.

As with the enzymatic reactions, the initial synthetic experiments also proved to be challenging. Indole N-H protection was needed for further halogenation chemistry but the steric strain of the compound limited the use of certain protecting groups. Large groups such as BOC and FMOC could not be used. After some experiments, only methylation appeared to be a viable protecting route moving forward. Most of the material at this point had been used up in protecting group optimization and further discussions were warranted.

5.3.2 HpiC1_L147F mutant and further synthetic attempts

Based on the initial results, the main issue hindering the project was the ability to produce acceptable amounts (greater than 20 mg) of **5.5** in a fast time frame. While FamD1 turnover remained an issue, it was reasoned that finding a method to produce more of **5.4** would be able to overcome the limitations of the later reverse prenylation. With a library of cyclase mutants from previous projects, a large scale screen was conducted to see if any of them could enhance production of **5.4.** After various screens and optimization efforts, the HpiC1_L147F mutant gave a large improvement in yield on a 5-7 mg scale (50-70%). While this result was promising, FamD1 turnover continued to be an issue and the increase in yield of **5.4** from the mutant only gave a moderate increase of **5.5** at the end of the reaction pathway.

At this point, the issue holding the project back was now solely focused on FamD1 and its turnover to produce **5.5**. Protein engineering of FamD1 was discussed but it would be too labor and time intensive. Danishefsky's method for reverse prenylation[10] had been used previously by Baran's group in their total synthesis of ambiguine H.[11] Baran reported that this method caused a unique reaction between the indole and isonitrile moieties. However, it was reasoned that the nitrile would not react in the same way. Fresh **5.4** was prepared for this synthesis route. After some experiments, the route proved to be successful in producing **5.5** however, hydrolysis of the nitrile to the carboxylic acid was observed in a 5:1 ratio of hydrolyzed:nonhydrolzyed (Scheme 5.4). Thus, **5.5** had to be produced enzymatically and a viable route needed to be exploited.



Scheme 5.4: Synthetic reverse prenylation results. Production of **5.4** proceeded in good yields with HpiC1_L147F mutant. Reverse prenylation using Danishefky's method[10] yielded a 5:1 ratio of hydrolysis of the nitrile to the corresponding carboxylic acid.

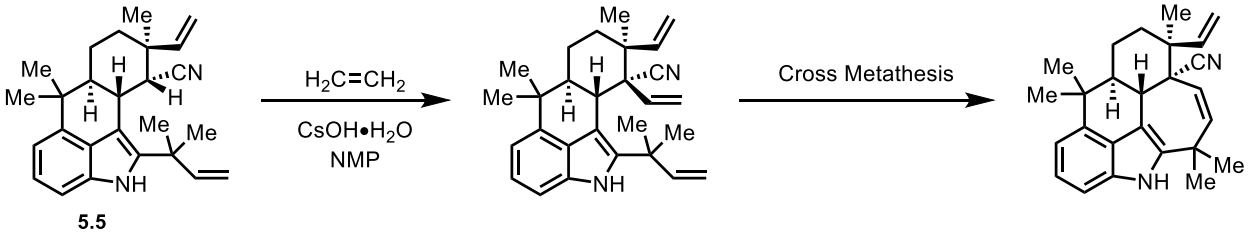
5.3.3 Enzymatic one pot synthesis optimization and current synthesis plans

During resultant discussions, the idea of a one pot enzymatic synthesis route was explored. Here, all the proteins (FamD2, HpiC1_L147F and FamD1) would be mixed in the same pot with cofactors, geranyl pyrophosphate (GPP) and DMAPP. This would cut work up to one purification step and ideally lead to an increase of **5.5**. However, it was known that FamD1 and FamD2 were somewhat promiscuous with prenyl groups[12] and it was feared that these prenylation reactions would hinder cyclization to **5.4**. The analytical reactions gave encouraging results. While percent conversion was not great (10-15%) of **5.5**, feared side reactions occurred at a minimum (except for a noteable exception which will be discussed later in this chapter). Scale up reactions to 5-10 mg scale afforded a slight increase in overall yield of the **5.5** (20-25%) but the side reaction to the new metabolite was hindering efforts (Supplemental information Figure S5.2). In order to counter act this reaction, a one pot stepwise route was developed. Here, the formation of **5.4** is allowed to go to completion before the addition of FamD1 and DMAPP. This reaction is then allowed to run until greater than 80% conversion by HPLC is observed. While this allowed for a more rapid production of **5.5**, it still required a lot of purified protein and a more efficient way to using raw materials was needed.

In order to meet this need, a one pot stepwise reaction was optimized using clarified (or purified) cell lysate. These reactions gave substantial increase in percent conversion values and after optimizing the work up method, the route proved to be the best plan moving forward (Supplemental information Figure S5.3**)**. To date, the reaction has been run three times on a 25 mg scale to give yields of 40, 49 and 61% of **5.5**. In just three reactions, greater than 80 mg of **5.5** was produced for synthetic efforts.

With the enzymatic route fully optimized, the synthetic efforts could be undertaken more extensively. While the nitrile could survive the aliphatic halogenation, the route proved to be

unsuccessful. Moving forward, a new route has been proposed where a vinyl group is added to the C-11 position using Cesium hydroxide.[13] This would lead to an Olefin cross metathesis reaction for the formation of the E-ring (Scheme 5.5).



Scheme 5.5: Future synthetic strategy. **5.5** will undergo a vinylation reaction using cesium hydroxide[13] followed by a olefin cross metathesis.

5.4 Results and discussion for discovery and characterization of 11-DMAC

5.4.1 Initial discovery and characterization

As discussed previously, during the one pot optimization for the production of **5.5**, a new peak was observed in roughly the same quantity as **5.5**. This new peak did not match any previous standards or known retention times of any previous hapalindole enzymatic reactions (Supplemental information Figure S5.2**)**. After isolation and characterization, this new compound was revealed to be a previously unknown compound in which the reverse prenyl group was connected in the C-11 position of **5.3** (Figure 5.2). It was hypothesized that this compound arises from a Cope rearrangement after the normal prenylation of **5.3** at the C-3 position. This new compound has been titled 11-DMAC (DiMethylAllylCope) nitrile (**5.6**). While a rearrangement similar to this has been observed in the enzyme dimethylallyltryptophan

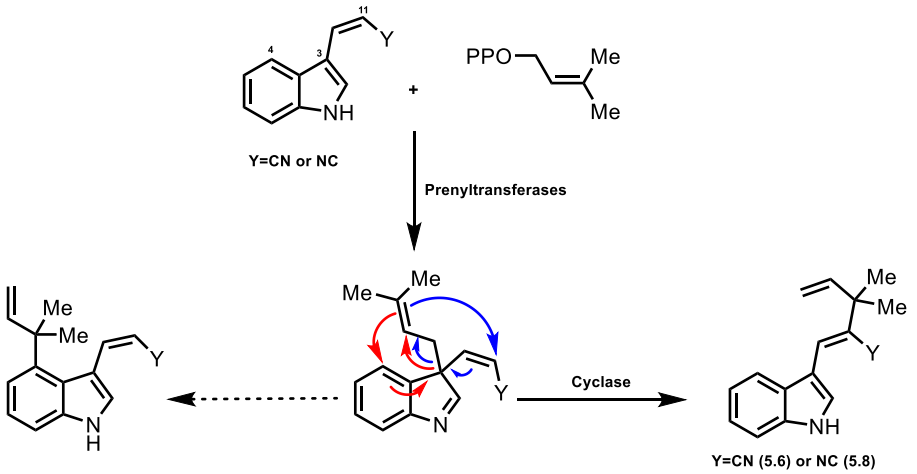


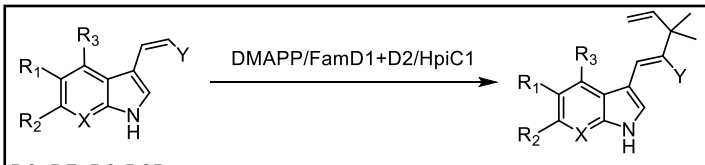
Figure 5.2: General formation of 11-DMAC compounds. *Cis*-indole nitrile (**5.3**) or -indole isonitrile (**5.7**) can be prenylated at the C-3 position with DMAPP with FamD2 and/or FamD2 prenyltransferases. A Cope rearrangement (in blue) induced by the cyclases leads to 11-DMAC nitrile (**5.6**) or isonitrile (**5.8**). A reaction similar to the Cope rearrangement catalyzed by DMAT[14] (in red) has been hypothesized but not observed.

synthase (DMAT), the reaction involves the indole C-4 position for the rearrangement (Figure 5.2).[14]

5.4.2 Follow up studies and derivatization of 11-DMAC

The discovery of **5.6** led to some follow up questions relating to if we would observe the same result with the isonitrile and if the Cope rearrangement is enzyme catalyzed. The same reaction involving FamD2, HpiC1 and DMAPP was run with *cis*-indole isonitrile (**5.7**) as the starting material. A peak with a similar retention time to **5.6** was observed and upon scale up and characterization, the compound was identified as 11-DMAC isonitrile (**5.8**) (Figure 5.2 and Supplemental information Figure S5.5). With this result in hand, we looked next to determine if the Cope rearrangement was cyclase catalyzed. Our initial observations showed slight spontaneous formation of **5.8** but reactions containing cyclase appeared to show greater conversion to the compound (Supplemental information Figure S5.4**)**. Kinetic assays were run looking at the rate of formation of **5.8** in reactions with and without cyclase. These results showed that **5.8** formation was 92-128x faster in reactions containing cyclase versus reactions without cyclase

Table 5.1: Analytical % conversions to 11-DMAC derivatives



5.3, 5.7, 5.9-5.25

Compound	Substituent	Analytical % conversion
5.3	R_1=H,R_2=H,R_3=H,X=C,Y=CN	34
5.7	R_1=H,R_2=H,R_3=H,X=C,Y=NC	26
5.9	R_1=OBzn,R_2=H,R_3=H,X=C,Y=NC	<1
5.10	R_1=CN,R_2=H,R_3=H,X=C,Y=NC	N/A*
5.11	R_1=OMe,R_2=H,R_3=H,X=C,Y=NC	16
5.12	R_1=H,R_2=OMe,R_3=H,X=C,Y=NC	14
5.13	R_1=F,R_2=H,R_3=H,X=C,Y=NC	23
5.14	R_1=H,R_2=F,R_3=H,X=C,Y=NC	14
5.15	R_1=Cl,R_2=H,R_3=H,X=C,Y=NC	>99
5.16	R_1=H,R_2=Cl,R_3=H,X=C,Y=NC	40
5.17	R_1=Br,R_2=H,R_3=H,X=C,Y=NC	<1
5.18	R_1=H,R_2=Br,R_3=H,X=C,Y=NC	8
5.19	R_1=I,R_2=H,R_3=H,X=C,Y=NC	<1
5.20	R_1=H,R_2=H,R_3=H,X=N,Y=NC	3
5.21	R_1=F,R_2=H,R_3=H,X=N,Y=NC	<1
5.22	R_1=H,R_2=F,R_3=H,X=N,Y=NC	<1
5.23	R_1=H,R_2=H,R_3=F,X=C,Y=NC	9
5.24	R_1=H,R_2=H,R_3=Cl,X=C,Y=NC	<1
5.25	R_1=H,R_2=H,R_3=Br,X=C,Y=NC	<1

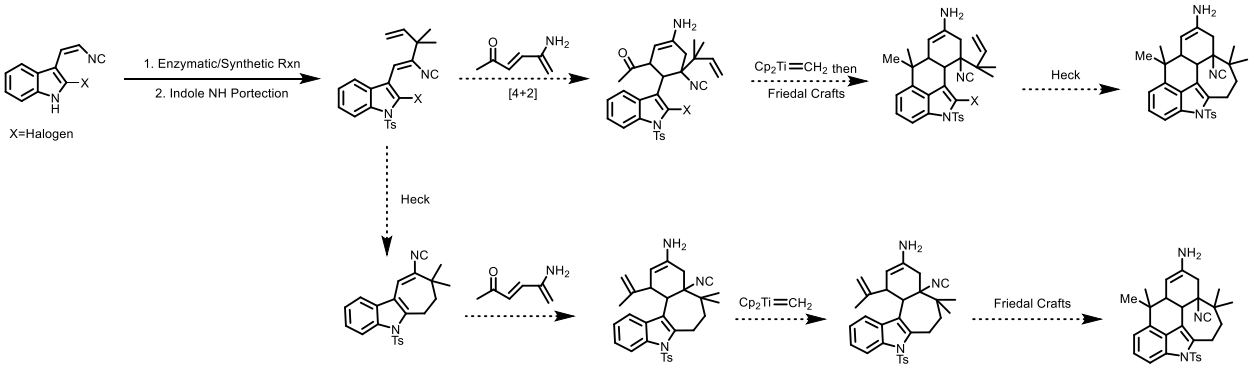
Assay conditions: 1 mM substrate, 20 µM HpiC1, 10 µM FamD2, 5 µM FamD1, 5 mM $MgCl_2$, 7.5 mM $CaCl_2$, 50 mM Tris pH 7.0
N/A*: Compound did not show ability to be prenylated by FamD1 or FamD2

(Supplemental information Table S5.2**)**. Not only did these results support our initial observations but also further strengthened the hypothesis of the Cope rearrangement being catalyzed by the cyclase in the biosynthesis of hapalindole-type metabolites.

With these questions answered, derivatization studies with HpiC1 and HpiC1_L147F were next to be explored. However, optimization of the reaction conditions had to first be conducted. These experiments revealed the addition of a small amount of FamD1 and a reduced amount of FamD2 actually increased the reaction rate. This is believed to occur because FamD1 is better able to recognize DMAPP whereas FamD2 recognizes GPP for prenylation. With the reaction fully optimized, derivatization studies were undertaken. At this point, the library of isonitrile derivatives (**5.9-5.25**) had expanded to include C-4 halogenated derivatives (**5.23-5.25**) and a 6-fluoro-7-Azaindole (**5.22**) derivative. 11-DMAC derivative formation appeared to occur in nine of the derivatives (Table 5.1 and Supplemental information Figures S5.6-S5.14) but the conversion values are fairly low (<15%) for majority of the substrates excluding the chlorine containing substrates which show elevated conversion values relative to the other compounds. This could be due to the chlorine being the perfect size substituent to push the compound into a more favorable position for prenylation in the FamD1 and FamD2 active sites. To date, only the 6-chloro (**5.16, S5.1**) and 6-bromo (**5.18, S5.2**) (Supplemental information Tables S5.6 and S5.7) derivatives have been characterized so further characterization work needs to be undertaken to determine if the other seven derivatives are indeed 11-DMAC derivatives.

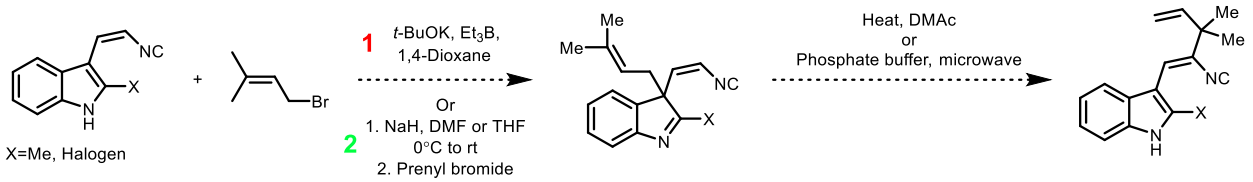
5.4.3 C-2 11-DMAC derivative total synthesis and proposed synthetic utility

While producing unnatural of derivatives of 11-DMAC may give us insight into the



Scheme 5.6: Proposed synthetic strategy of pentacyclic ambiguine using a C-2 halogenated 11-DMAC derivative as starting material.

promiscuity of the enzymes, other uses for 11-DMAC need to be explored. We hypothesized that a C-2 derivative could be used to form the E-ring of a pentacyclic ambiguine. From here, further reactions such as a Diels-Alder and Friedel-Crafts could be used to complete the remaining ring systems (Scheme 5.6). Based on our previous results, we concluded that enzymatic reactions would not give us the desired amount of material for further synthetic reactions. To address this limitation, we proposed a biomimetic total synthesis of a C-2 11-DMAC derivative. This would involve a C-3 normal prenylation followed by an induced Cope rearrangement (Scheme 5.7). Literature searches revealed that route was possible using a heat or microwave induced Cope rearrangement after the normal prenylation at the indole C-3 position.[15-16] We elected to start with a C-2 methyl derivative (**5.26**) as a test system with the plan to move to a C-2 halogenated derivative in the future. Our initial attempt proved successful but follow up attempts were unsuccessful. Investigation of the reaction progress revealed that the chelation between the indole nitrogen and the boron chelating agent proved too difficult to break. To solve this issue, we moved forward with a different method not involving a chelating agent.[17] This new reaction was first tested on the *trans*-indole isonitrile derivative (**5.27**). The desired C-3 prenylated material was obtained (Experimentals **5.28**) and the reaction conditions will next be moved to the *cis* derivative. Once enough material is obtained, the Cope rearrangement will be explored. Based on electronics of the system, it is unknown if the rearrangement will occur at the C-11 position as desired or the C-4 position as observed with DMAT (Figure 5.2).[14] This result could lead to further questions and experiments but for now the main focus is on getting to our desired C-2 11-DMAC compound.



Scheme 5.7: Proposed synthetic strategy for a C-2 11-DMAC derivative. The first conditions screened[15] (**1**) on the C-2 methyl derivatives (**5.26** and **5.**27) was unsuccessful due to the strong chelation between the indole nitrogen and borane. The second conditions screened[17] (**2**) have been successful with **5.27** and will next be screened with **5.26**. The next part of the reaction will involve inducing the Cope rearrangement by either microwave or high heat.[16]

5.5 Conclusion

With the Rieske oxygenases proving difficult to elucidate, a chemoenzymatic synthesis route to a pentacyclic ambiguine was explored. Initial discussions within the CCHF showed that this challenge was viable. Further exploration and collaboration has led an optimized one pot

enzymatic reaction for the production of 12-*epi*-ambiguine H nitrile (**5.5**) on 25 mg scale. This new route has allowed for further synthetic studies. While these synthetic studies have yet to yield the same positive results, it is not unreasonable to believe this route will succeed. During this one pot optimization, a novel metabolite was discovered. Both 11-DMAC nitrile (**5.6**) and 11-DMAC isonitrile (**5.8**) helped further strengthen the hypothesis that the hapalindole-type metabolite biosynthesis begins with a cyclase catalyzed Cope rearrangement. Initial enzymatic derivatization studies have showed some promising results and a viable route for a total synthesis of a C-2 11-DMAC derivative appears to be close to being obtained. These results show that while the Rieske proteins may still be the holy grail for ambiguine E-ring formation, there are other viable routes and discoveries related to this critical question.

5.6 Experimentals

5.6.1 General

All NMR spectra were acquired on a Varian 400 and Bruker 600 MHz spectrometers. Proton and carbon signals are reported in part per million (δ) using residual solvent signals as an internal standard. Analytical HPLC analysis was performed on a Shimadzu 2010 EV APCI spectrometer equipped with an LUNA C18 250 x 4.6 mm column, using a mobile phase gradient of 70-100% acetonitrile in water over 16 min and was monitored by UV absorption at 280 nm. LC-MS analysis was performed on a Agilent Infinity II TOF using an XBridge C18 2.1 x 150 mm column, using a mobile phase gradient of 60-100% or 70-100% acetonitrile in water over 12 min. Preparative-scale HPLC was performed on a Shimadzu 20-AT equipped with an LUNA C18 250 x 10 mm column for 2 mg reactions and an LUNA C8 250 x 21 mm column for 5+mg reactions, using a mobile phase gradient of 60-100% acetonitrile in water over 60 min.

Escherichia coli strain BL21(DE3) was used for protein expression. Plasmid pET28a was used for cloning and expression of FamD1, FamD2, HpiC1 and HpiC1_L147F. Isopropyl β-D-thiogalactopyranoside (IPTG) was used to induce expression; DNAse and lysozyme were purchased form Sigma-Aldrich. Ni-NTA agarose from Invitrogen was used to purify His-tag proteins.

All chemicals were purchased from Sigma-Aldrich, ACROS, Combi-Blocks, Enamine and Oakwood. Multiplicities are abbreviated as following: singlet (s), doublet (d), triplet (t), quartet (q), doublet-doublet (dd), triplet-doublet (td), doublet-doublet-doublet (ddd), triplet-doublet-doublet (tdd), and multiplet (m). Chemical abbreviations: Ethyl Acetate (EtOAc), Dichloromethane (DCM), Tetrahydrofuran (THF), Potassium bis(trimethylsilyl)amide (KHMDS), Acetic Acid (AcOH), Sodium Sulfate (Na_2SO_4), Diethyl Ether (Et_2O), Sodium Hydroxide (NaOH), Acetonitrile (CH_3CN), Magnesium Chloride ($MgCl_2$), Calcium Chloride ($CaCl_2$), Sodium Chloride (NaCl), Methanol (MeOH), Water (H_2O), Sodium Hydride (NaH)

5.6.2 Protein expression and purification

The expression and purification of proteins was performed as described.[12] Briefly, a single BL21(DE3) colony was inoculated in LB medium containing 50 μg/mL kanamycin and grown overnight at 37 °C shaking at 200 rpm. The main culture (1 L) was inoculated at the dilution of 1:100 in 2.8 L of Fernbach flask containing TB medium and the same concentration of antibiotic. The cells were grown (37 °C, 200 rpm) to an optical density (A_{600} nm) of 1.0. The culture flasks were chilled in ice, induced with IPTG (0.4 mM), and were further incubated (18 °C, 200 rpm) for 16 h. The cells were harvested (5000 rpm, 4 °C, 15 min), flash frozen, and stored at −80 °C until purification. The cell pellets were resuspended at 4 °C in the lysis buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol), containing 0.5 mg/mL of lysozyme, 1 mM PMSF and 1 mL of 2 mg/mL DNAse. The mixture was stirred for 30 min and sonicated on ice for 120 s total time using 10 s pulses followed by a 50 s pause. The cellular debris was removed by centrifugation (65,000 x g, 4 °C, 35 min). The clarified lysate was loaded onto Ni-NTA agarose column equilibrated with lysis buffer. The column was washed with two column volume of wash buffer (10 mM HEPES, 300 mM NaCl, 0.2 mM TCEP, 10% glycerol, 20 mM imidazole) and the His-tagged protein was eluted with elution buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol, 300 mM imidazole). The fractions were pooled and dialyzed overnight or by using a PD10-desalting column (GE Healthcare) using storage buffer (10 mM HEPES, 50 mM NaCl, 0.2 mM TCEP, 10% glycerol). The purified protein was analyzed by SDS-PAGE gel for purity, measured by Nanodrop using a calculated molar extinction coefficient for concentration, and flash-frozen in liquid nitrogen to store at −80 °C.

5.6.3 Purified protein assay and scale up reactions for 5.5 formation

12-*epi*-haplindole U nitrile (**5.4**) was previously produced enzymatically.[9] For the analytical scale reactions, a 100µL reaction containing 1 mM **5.4**, 15 µM FamD1, 1.5 mM DMAPP, 5 mM $MgCl_2$, 50 mM of Glycine pH 10.0 buffer was incubated at 37°C for 2.5 hrs. The reaction was quenched with 3x volume of EtOAc. The organic layers were combined, dried and re-dissolved in 100 µL acetonitrile for LCMS and HPLC analysis. HPLC conversion values were determined by the area under the curve of the residual starting material and newly formed product. For the structure analysis and isolated yield values of the enzymatic products, the reactions were scaled up to 5 mg starting material (25 mL) and incubated at 37°C overnight or until HPLC showed consumption of starting material. Products were extracted with EtOAc and purified by preparative HPLC as described in the general section. The purified compound was concentrated, dissolved in C_6D_6 and analyzed using a Bruker 600 MHz NMR. **5.5** was obtained as a white solid. Characterization can be found at Supplemental information Table S5.3.

5.6.4 Purified protein assay and scale up for one pot non stepwise reaction for 5.5 formation

Cis-indole nitrile (**5.3**) was synthesized as described in section 5.6.7. For the analytical scale reactions, a 100µL reaction containing 1 mM **5.3**, 20 µM HpiC1, 5 µM FamD2, 15 µM FamD1, 1.5 mM GPP, 1.5 mM DMAPP, 5 mM $MgCl_2$, 7.5 mM $CaCl_2$, 50 mM Tris Buffer pH 7.0 buffer was incubated at 37°C for 2.5 hrs. The reaction was quenched with 3x volume of EtOAc. The organic layers were combined, dried and re-dissolved in 100 µL acetonitrile for LCMS and HPLC analysis. HPLC conversion values were determined by the area under the curve of the residual starting material and newly formed product. For the scale up reactions, the reactions were scaled up to 5-10 mg starting material (25-50 mL total volume) and incubated at 37°C overnight or until HPLC showed consumption of starting material. Products were extracted with EtOAc and purified by preparative HPLC as described in the general section.

5.6.5 One pot stepwise purified cell lysate reaction

25 g of cell pellet of HpiC1_L147F and 12.5 g of cell pellet of FamD2 were prepared in a similar fashion to the methods described in the protein expression and purification section. The cell pellets were mixed in 6.25 mL MES pH 6.5 buffer, 0.5 mg/mL lysozyme, 2 mg/mL DNAse,

10% glycerol and H_2O. Cells were sonicated and pelleted as previously described. Once the cell debris was pelleted, the lysate was passed through a 0.45µ filter. The purified lysate was added to a 250 mL round bottom containing 25 mg of **5.3** dissolved in 2% DMSO:H_2O (1 mM final reaction concentration), 5 eq of GPP (271 mg), 8 mM $MgCl_2$ and 10 mM $CaCl_2$ stirring at 37 °C. After the reaction showed >80% conversion by HPLC to **5.4**, 20 g of cell pellet of FamD1 was prepared in the same manner as previous enzymes. The purified lysate and 6 eq of DMAPP (262 mg) was added to the reaction for a final reaction volume of ~125 mL. The reaction was monitored by HPLC until the reaction showed >80% completion. The reaction was quenched with 3x volume with either acetone or MeOH to crash out all biologics. The reaction mixture was split between 50 mL conical vials and centrifuged (3000 x g, rt, 15 min). The resulting supernatant was combined and had the organic removed by rotary evaporation. The remaining water/buffer layer was extracted with EtOAc (3x), washed with brine, dried with Na_2SO_4, filtered and concentrated. The crude mixture was dissolved in 5 mL acetone and purified by preparative HPLC as described in the general section. **5.5** was obtained as an off-white solid on larger reaction scale.

5.6.6 11-DMAC analytical assays, kinetic assays and scale up reactions

Cis-indole derivatives were synthesized as described in section 5.6.7. For the analytical scale reactions, a 100µL reaction containing 1 mM substrate, 20 µM HpiC1, 10 µM FamD2, 5 µM FamD1, 3 mM DMAPP, 8 mM $MgCl_2$, 7.5 mM $CaCl_2$, 50 mM Tris Buffer pH 7.0 buffer was incubated at 37°C for 4 hrs. Kinetic assays were ran under the same conditions except select reactions did not contain cyclase. The reaction was quenched with 3x volume of EtOAc. The organic layers were combined, dried and re-dissolved in 100 µL acetonitrile for LCMS and HPLC analysis. HPLC conversion values were determined by the area under the curve of the residual starting material and newly formed product. For the scale up reactions, the reactions were scaled up to 5 mg starting material (25 mL total volume) and incubated at 37°C overnight or until HPLC showed consumption of starting material. Products were extracted with EtOAc and purified by preparative HPLC as described in the general section. All compounds were obtained as a white solid. The purified compounds were concentrated, dissolved in C_6D_6 and analyzed using a Bruker 600 MHz NMR. Characterization can be found at Supplemental information Tables S5.4-S5.7.

5.6.7 Chemical synthesis of *cis*-indole isonitrile derivatives

All derivatives were prepared using method previously described.[9] Briefly, to a 50 mL two-neck round-bottom flask purged with nitrogen at −78 °C (dry ice/acetone), diethyl (isocyanomethyl) phosphonate (0.37 mL, 2.26 mmol) (diethyl cyanomethyl phosphonate was used for production of **5.3**) was diluted with THF (5 mL). KHMDS (1 M THF, 2.60 mL, 2.60 mmol) was added dropwise, and the reaction was stirred at −78 °C for 15 min. To a separate 4 mL vial, indole-3-carboxaldehyde derivative (1.13 mmol) was dissolved in THF (5 mL), and the resulting solution was added dropwise to the KHMDS solution at −78 °C. The resulting mixture was stirred at 0 °C (cryocool) overnight or until TLC showed consumption of starting material. The resulting solution was quenched by the addition of AcOH (0.15 mL, 2.6 mmol) and concentrated. The resulting residue was diluted with EtOAc (20 mL), washed with 1 M aqueous potassium phosphate buffer (20 mL, pH 7), washed with brine, dried with Na_2SO_4, and concentrated to a residue. The residue was dissolved in EtOAc and purified by flash chromatography (24%–100% pentane/Et_2O, SiO_2) to afford the titled compound as reported below. Yields and spectral data reported below. Note: Majority of compounds are reported in literature already.[9] Only the new 4-halogenated derivatives (**5.23-5.25**) and 6-F-7-Aza derivative (**5.22**) and both C-2 methyl *cis* & *trans*-indole isonitrile (**5.26, 5.27**) spectral data and yield are reported.

(Z)-4-fluoro-3-(2-isocyanovinyl)-1H-indole (5.23): Tan solid, 36.7 mg, 16%

^{1}H NMR (400 MHz, Acetone-d_6) δ 11.12 (s, 1H), 8.21 (s, 1H), 7.37 (d, *J* = 8.2 Hz, 1H), 7.18 (td, *J* = 8.0, 5.2 Hz, 1H), 7.09 (s, 1H), 6.87 (dd, *J* = 11.7, 7.8 Hz, 1H), 5.97 (d, *J* = 9.0 Hz, 1H).

(Z)-4-chloro-3-(2-isocyanovinyl)-1H-indole (5.24): White-red solid, 56.6 mg, 23%

^{1}H NMR (400 MHz, Acetone-d_6) δ 11.16 (s, 1H), 8.30 (d, *J* = 2.6 Hz, 1H), 7.61 (s, 1H), 7.53 (dd, *J* = 7.4, 1.6 Hz, 1H), 7.23 – 7.12 (m, 2H), 5.98 (d, *J* = 9.1 Hz, 1H).

(Z)-4-bromo-3-(2-isocyanovinyl)-1H-indole (5.25): Red solid, 42.6 mg, 15%

^{1}H NMR (400 MHz, Acetone-d_6) δ 11.16 (s, 1H), 8.30 (s, 1H), 7.74 (s, 1H), 7.58 (dd, *J* = 8.1, 0.9 Hz, 1H), 7.36 (dd, *J* = 7.6, 0.9 Hz, 1H), 7.12 (t, *J* = 7.9 Hz, 1H), 6.01 (d, *J* = 9.2 Hz, 1H).

(Z)-6-fluoro-3-(2-isocyanovinyl)-1H-pyrrolo[2,3-b]pyridine (5.22): Red solid, 4.0 mg, 9%

^{1}H NMR (400 MHz, Acetone-d_6) δ 11.36 (s, 1H), 8.36 (t, *J* = 8.1 Hz, 1H), 8.19 (s, 1H), 6.97 (s, 1H), 6.91 (d, *J* = 8.4 Hz, 1H), 6.05 (d, *J* = 8.9 Hz, 1H).

(Z)-3-(2-isocyanovinyl)-2-methyl-1H-indole (5.26): Red-brown solid, 40.7 mg, 16%

^{1}H NMR (400 MHz, Acetone-d_6) δ 10.85 (s, 1H), 7.64 (d, *J* = 7.5 Hz, 1H), 7.41 – 7.32 (m, 1H), 7.16 – 7.02 (m, 2H), 6.83 (s, 1H), 6.03 (d, *J* = 8.9 Hz, 1H), 2.51 – 2.43 (m, 3H).

(E)-3-(2-isocyanovinyl)-2-methyl-1H-indole (5.27): Red-brown residue, 45.4 mg, 20%

^{1}H NMR (400 MHz, Acetone-d_6) δ 10.61 (s, 1H), 7.76 – 7.69 (m, 1H), 7.41 – 7.33 (m, 1H), 7.26 (d, *J* = 14.2 Hz, 1H), 7.11 (dtd, *J* = 8.9, 7.4, 6.9, 3.9 Hz, 2H), 6.52 (d, *J* = 14.2 Hz, 1H), 2.54 (s, 3H).

5.6.8 Chemical synthesis of dimethylallyl diphosphate, tri ammonium

Dimethyallyl diphosphate was synthesized in a similar method to geranyl diphosphate.[9] To a 10 mL round bottom flask purged with nitrogen, tris (tetrabutylammonium) hydrogen pyrophosphate (1.71 g, 1.04 mmol) was dissolved in CH_3CN (1.0 mL). Prenyl bromide (0.1 mL, 0.475 mmol) was added and the reaction mixture was stirred at room temperature for 2 h.

Dowex 50WX8 resin preparation: Dowex 50WX8 resin (30 g, hydrogen form) was washed with half saturated aqueous ammonium chloride (5x50 mL) and water (5x50 mL) until the pH of the supernatant equaled 5. The slurry was rinsed twice with ion exchange buffer (2% isopropanol in 25 mM aqueous ammonium bicarbonate) and loaded into a flash column and equilibrated with ion exchange buffer.

Purification: The reaction mixture was concentrated to afford an orange residue which was diluted with ion exchange buffer. The crude mixture was chromatographed with two column volumes of ion exchange buffer (75 mL). The fractions were combined and concentrated by rotary evaporation, flash frozen and lyophilized for 2 d. The resulting white powder was obtained as the final product.

^{1}H NMR (400 MHz, Deuterium Oxide) δ 5.70 (td, *J* = 7.1, 6.6, 3.5 Hz, 1H), 4.69 (t, *J* = 6.8 Hz, 2H), 2.02 (s, 3H), 1.97 (s, 3H).

^{31}P NMR (242 MHz, Deuterium Oxide) δ -6.06 (d, *J* = 22.9 Hz), -10.06 (dt, *J* = 22.3, 6.1 Hz).

5.6.9 C-2 indole isonitrile C-3 prenylation synthesis

Both *cis* and *trans* compounds were produced using method described previously.[17] Briefly, to a 25 mL two neck round bottom flask purged with nitrogen at 0 °C (ice water bath), the desired starting material (10-20 mg, 0.0548-0.1097 mmol) was dissolved in dry THF (1.5-2.0

mL). NaH (60% dispersion oil, 9.20 mg-18.42, 0.274-0.5485 mmol) was added slowly. After 1 h, prenyl bromide (25.3-50.7 µL, 0.2192-0.4388 mmol) was added. The reaction was brought to room temp slowly and monitored by TLC until consumption of starting material. The resulting reaction was cooled back to 0 °C and quenched with ice-water. The resulting mixture was extracted with EtOAc (3x), washed with brine, dried with Na_2SO_4 and concentrated to a residue. The residue was dissolved in EtOAc and purified by flash chromatography (5%–40% Hexanes/EtOAc, SiO_2) to afford the titled compound as reported below.

(E)-3-(2-isocyanovinyl)-2-methyl-3-(3-methylbut-2-en-1-yl)-3H-indole (**5.28**): Red-brown solid, 5.3 mg, 23%

^{1}H NMR (599 MHz, Chloroform-*d*) δ 7.55 (dt, *J* = 7.8, 0.9 Hz, 1H), 7.37 (ddd, *J* = 7.8, 5.5, 3.3 Hz, 1H), 7.23 – 7.21 (m, 2H), 6.20 (d, *J* = 14.1 Hz, 1H), 5.62 (d, *J* = 14.2 Hz, 1H), 4.54 (tp, *J* = 6.3, 1.4 Hz, 1H), 2.69 – 2.63 (m, 1H), 2.50 – 2.44 (m, 1H), 2.27 (s, 3H), 1.55 (d, *J* = 1.5 Hz, 3H), 1.49 (d, *J* = 1.3 Hz, 3H).

HRMS: $[M+H]^+$ Calc: 251.1543 Obsv: 251.1545

5.7 References

1. Johnson, R. E.; Ree, H.; Hartmann, M.; Lang, L.; Sawano, S.; Sarpong, R. Total Synthesis of Pentacyclic (−)-Ambiguine P Using Sequential Indole Functionalizations. *J. Am. Chem. Soc.* **2019**, *141* (6), 2233–2237. https://doi.org/10.1021/jacs.8b13388.
2. Xu, J.; Rawal, V. H. Total Synthesis of (−)-Ambiguine P. *J. Am. Chem. Soc.* **2019**, *141* (12), 4820–4823. https://doi.org/10.1021/jacs.9b01739.
3. Hu, L.; Rawal, V.H. Total Synthesis of the Chlorinated Pentacyclic Indole Alkaloid (+)-Ambiguine G. *J. Am. Chem. Soc.* **2021**, *143* (29), 10872-10875. https://doi.org/10.1021/jacs.1c05762.
4. Roudesly, F.; Oble, J.; Poli, G. Metal-catalyzed CH activation/functionalization: The fundamentals. *J. Mol. Catal. A Chem.* **2017**. *426*. 275-296. https://www.sciencedirect.com/science/article/pii/S1381116916302424.
5. Davies, H.M.L.; Morton, D. Recent Advances in C-H Functionalization. *J. Org. Chem.* **2016**, *81* (2), 343-350. https://doi.org/10.1021/acs.joc.5b02818.
6. Davies, H.M.L.; Morton, D. Collective Approach to Advancing C-H Functionalization. *ACS Cent. Sci.* **2017**, *3* (9), 936-943. https://doi.org/10.1021/acscentsci.7b00329.
7. Rafferty, R. J.; Williams, R. M. Total Synthesis of Hapalindoles J and U. *J. Org. Chem.* **2012**, *77* (1), 519–524. https://doi.org/10.1021/jo202139k.
8. 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. https://doi.org/10.3987/COM-12-S(N)3.
9. Hohlman, R.M.; Newmister, S.A.; Sanders, J.N.; Khatri, Y.; Li, S.; Keramati, N.R.; Lowell, A.N.; Houk, K.N.; Sherman, D.H. Structural diversification of hapalindole and fischerindole natural products via cascade biocatalysis. *ACS Catal.* **2021**, *11* (8), 4670-4681. https://doi.org/10.1021/acscatal.0c05656.
10. Schkeryantz, J.M.; Woo, J.C.G.; Siliphaivanh, P.; Depew, K.M.; Danishefsky, S.J. Total Synthesis of Gypsetin, Deoxybrevianamide E, Brevianamide E, and Tryprostatin B: Novel Constructions of 2,3-Disubstituted Indoles. *J. Am. Chem. Soc.* **1999**, *121* (51), 11964-11975. https://pubs.acs.org/doi/abs/10.1021/ja9925249.
11. 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. https://www.nature.com/articles/nature05569.
12. 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. https://doi.org/10.1021/jacs.5b10136.
13. Tzalis, D.; Knochel, P. Cesium Hydroxide: A Superior Base for the Catalytic Alkynylation of Aldehydes and Ketones and Catalytic Alkenylation of Nitriles. *Angew. Chem. Int. Ed.* **1999**, *38* (10), 1463-1465. https://onlinelibrary.wiley.com/doi/abs/10.1002.
14. Luk, L.Y.P.; Qian, Q.; Tanner, M.E. A Cope Rearrangement in the Reaction Catalyzed by Dimethylallyltryptophan Synthase?. **2011**, *133* (32), 12342-12345. https://pubs.acs.org/doi/10.1021/.

15. Lin, A.; Yang, J.; Hashim, M. *N*-Indolytriethylborate: A Useful Reagent for Synthesis of C3-Quaternary Indolenines. *Org. Lett.* **2013**, *15* (8), 1950-1953. https://doi.org/10.1021/ol4005992.
16. Thandavamurthy, K.; Sharma, D.; Porwal, S.K.; Ray, D.; Viswanathan, R. Regioselective Cope Rearrangement and Prenyl Transfers on Indole Scaffold Mimicking Fungal and Bacterial Dimethylallyltryptophan Synthases. *J. Org. Chem.* **2014**, *79* (21), 10049-10067. https://doi.org/10.1021/jo501651z.
17. Adla, S.K.; Sasse, F.; Kelter, G.; Fiebig, H. H.; Lindel, T. Doubly prenylated tryptamines: cytotoxicity, antimicrobial activity and cyclisation to the marine natural product flustramine A. *Org. Biomol. Chem.* **2013**, *11* (36), 6119-6130. https://pubs.rsc.org/en/content/articlelanding/2013/ob/c3ob40896e.

5.8 Supplemental information

Table S5.1: Protein sequences and mutagenic primers[9]

FamD1:
MSRSFHAFLFNCFDCNKENIKMTIVNRIRTDVVNVAKSFGAEYSEAVIDQIFQGFGEKFTNTGFAIRVQN
KRNQKVDCNIRYGEAKENCLAWDIARESGLLSDQGHPVDTLIQEMFQAIPAIAYGADFDINYGLVKIWHL
PKIVPVEEAFKIPSLPKSVNAHIDFFKKYHLDALCALTVDYRNKSTNLYFDAHHPEQRTTQFYKNILQSQ
QFEVPSDEVLEILVNCPEIAVTFNWSSPGIERMCFYTAFVNRETVPQHINPVLKKFAQEAPALLDNPGFL
VGWSFGPDAKKGTYIKIDVDYHGLVVPSFFHMHNLPLPIPEANSVFDLPSSDTEDKLNSIVMS

Figures S5.1-5.5: Analytical HPLC traces

Figure S5.1: 12-*epi*-hapalindole U nitrile (**5.4**) to 12-*epi*-ambiguine H nitrile (**5.5**) reaction

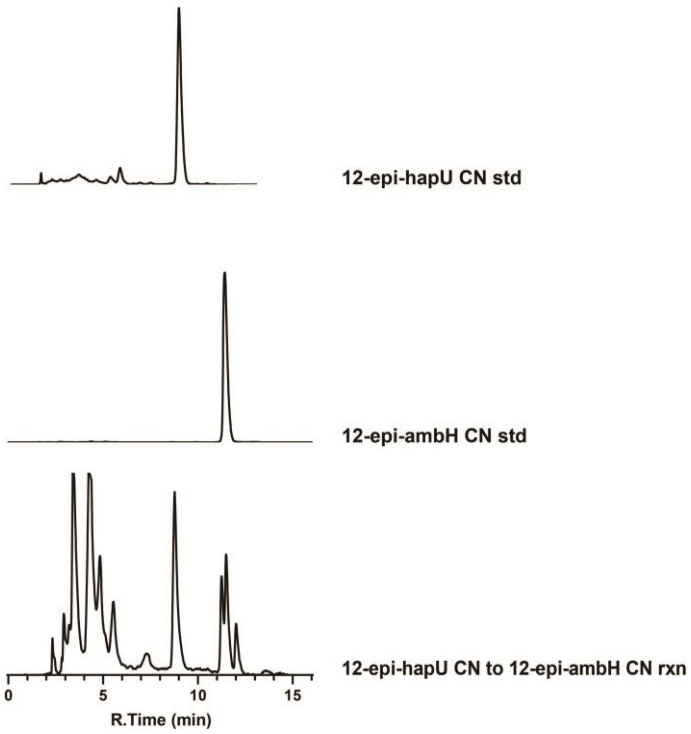


Figure S5.2: *Cis*-indole nitrile (**5.3**) one pot nonstepwise to 12-*epi*-ambiguine H nitrile (**5.5**) and 11-DMAC nitrile (**5.6**)

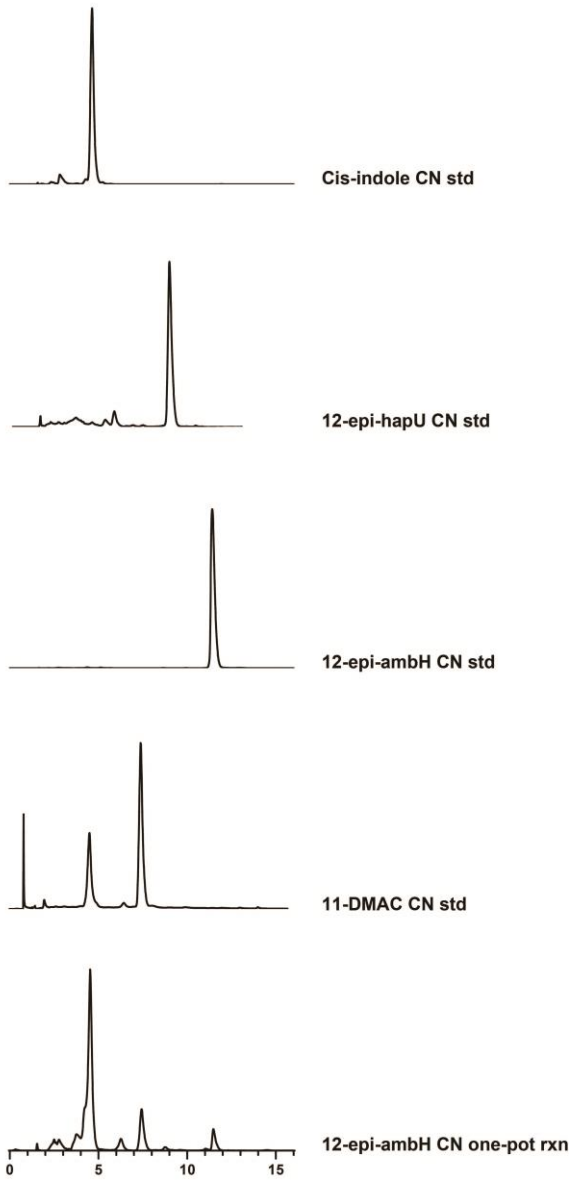


Figure S5.3: One pot clarified cell lysate reaction time course for production of 12-*epi*-ambiguine H nitrile (**5.5**)

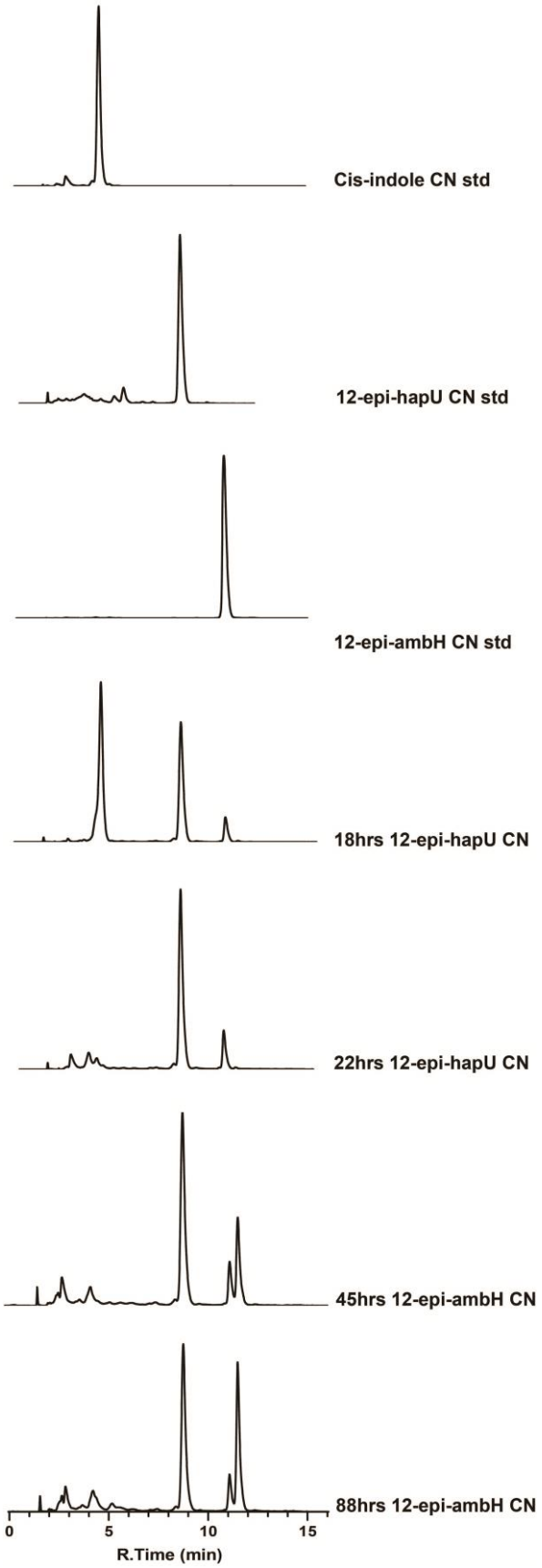

Cis-indole CN std
12-epi-hapU CN std
12-epi-ambH CN std
18hrs 12-epi-hapU CN
22hrs 12-epi-hapU CN
45hrs 12-epi-ambH CN
88hrs 12-epi-ambH CN
0
5
10
15
R.Time (min)

Figure S5.4: 11-DMAC nitrile (**5.6**) formation with cyclase and without cyclase

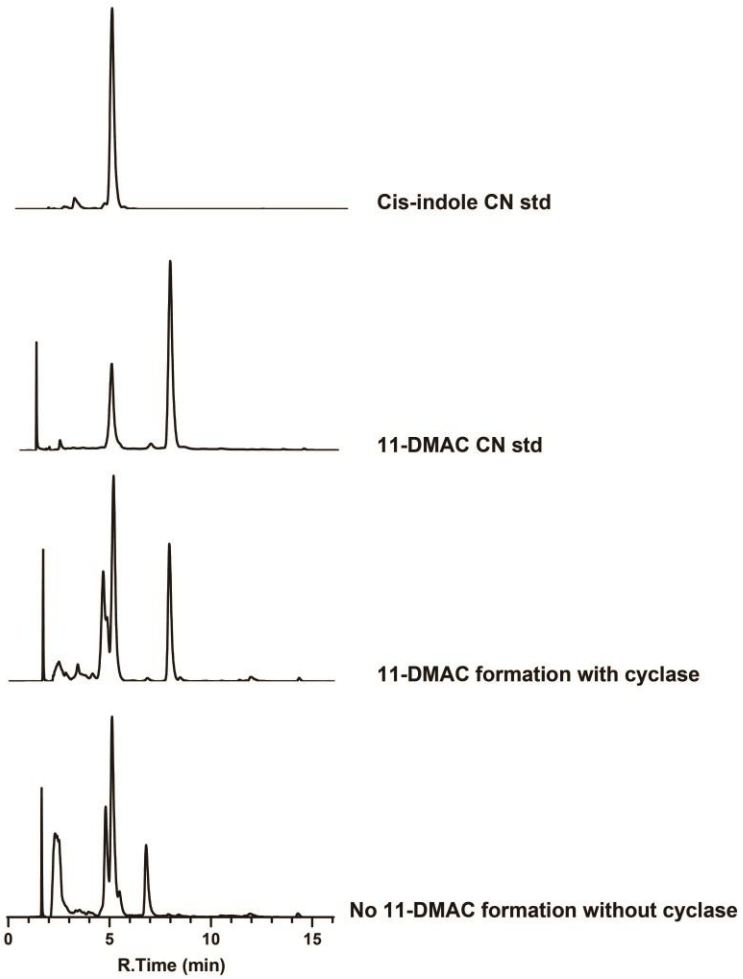
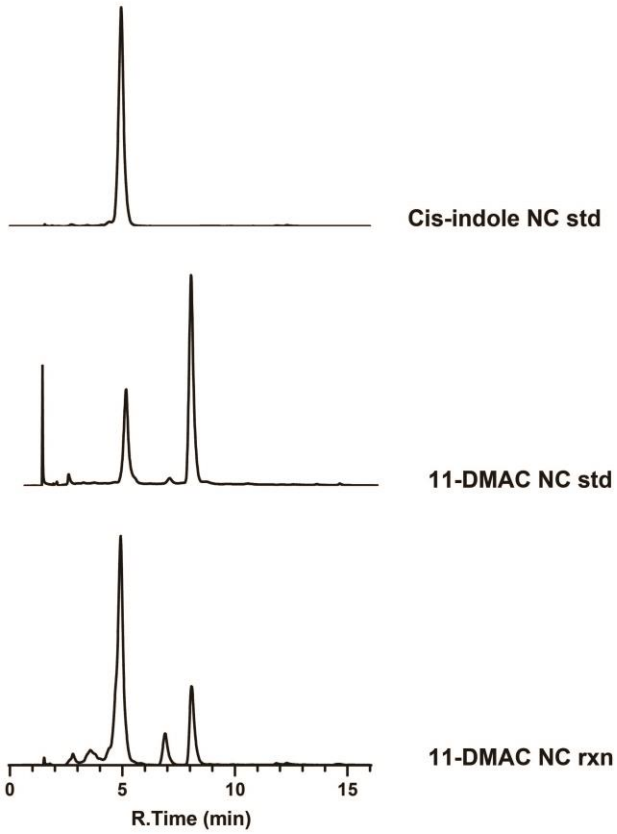


Figure S5.5: 11-DMAC isonitrile (**5.8**) formation



Figures S5.6-S5.14: 11-DMAC isonitrile derivative HPLC traces with either confirmed or hypothesized formation

Figure S5.6: 5-methoxy (**5.11**) 11-DMAC reaction. (*****=C-2 prenylated shunt metabolite, &=11-DMAC hypothesized derivative)

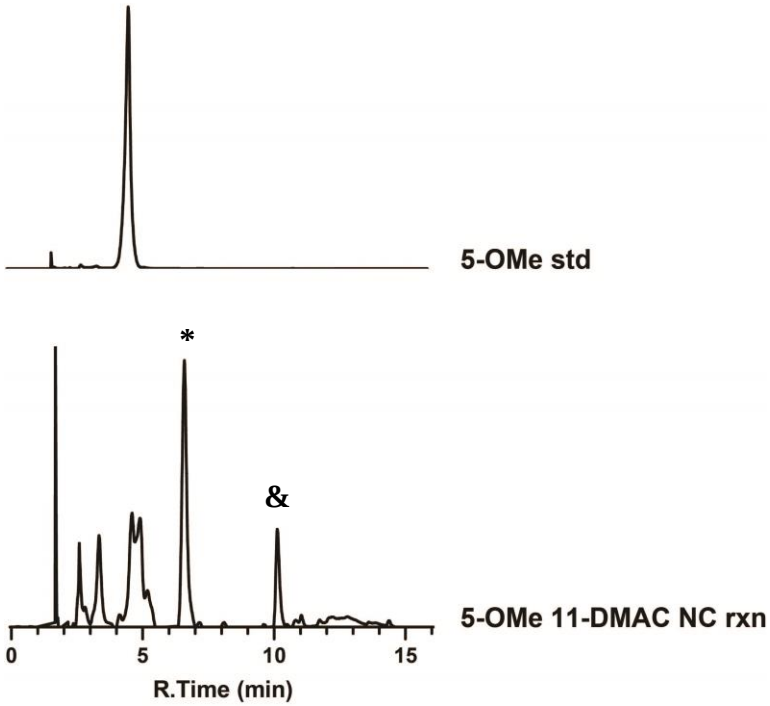


Figure S5.7: 6-methoxy (**5.12**) 11-DMAC reaction. (*****=C-2 prenylated shunt metabolite, &=11-DMAC hypothesized derivative)

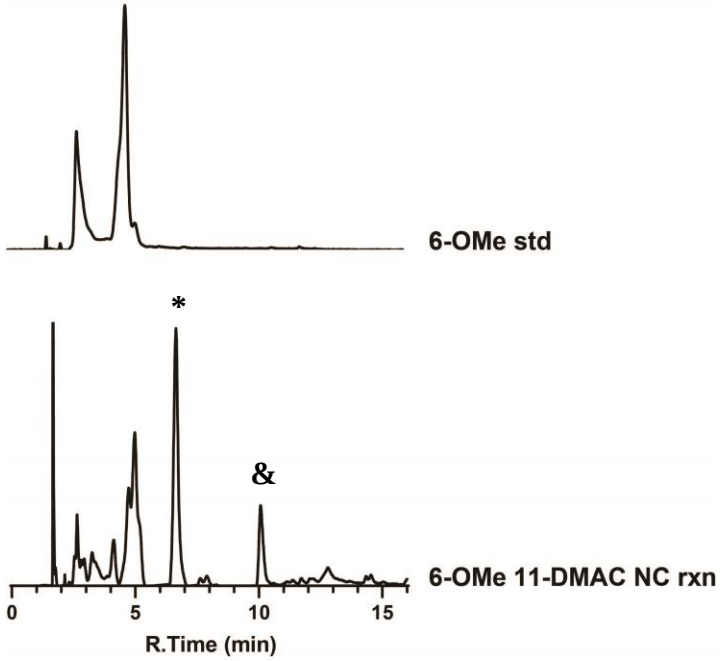


Figure S5.8: 4-fluoro (**5.23**) 11-DMAC reaction. (&=11-DMAC hypothesized derivative)

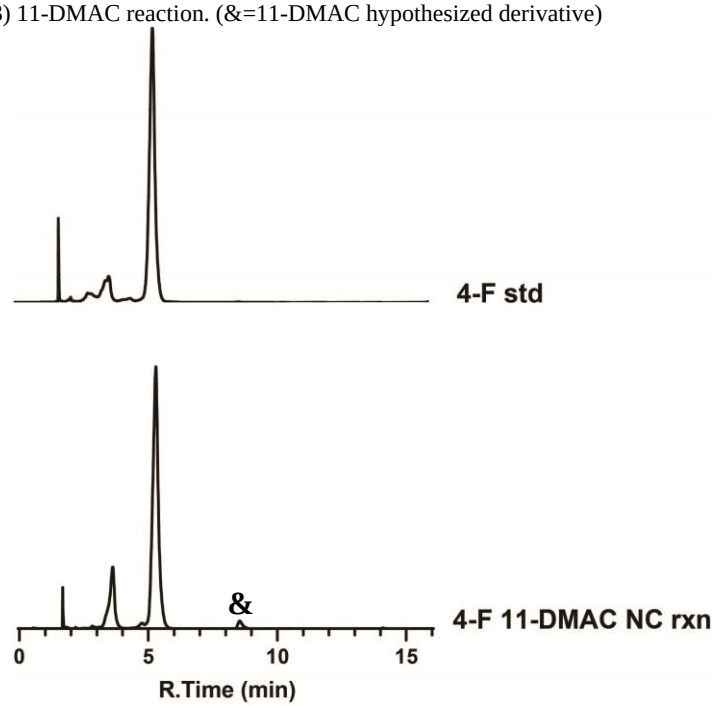


Figure S5.9: 5-fluoro (**5.13**) 11-DMAC reaction. (&=11-DMAC hypothesized derivative)

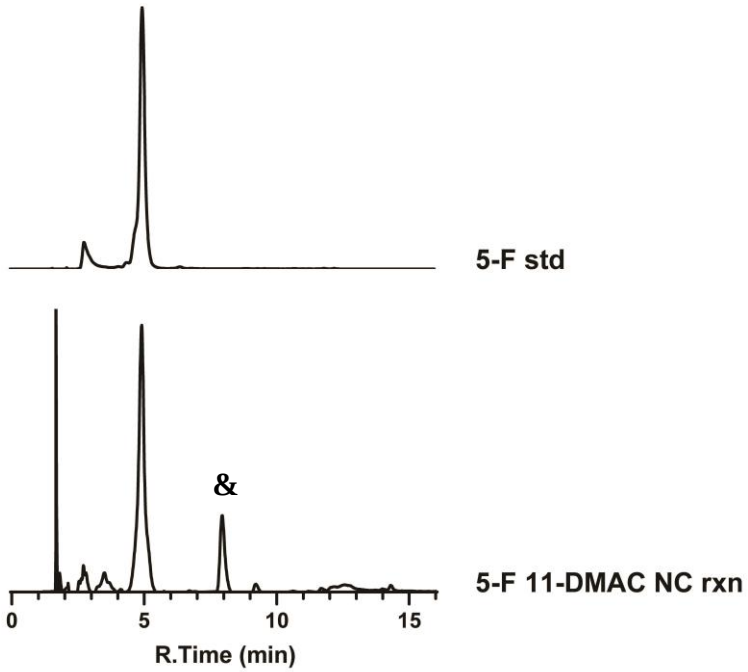


Figure S5.10: 6-fluoro (**5.14**) 11-DMAC reaction. (&=11-DMAC hypothesized derivative)

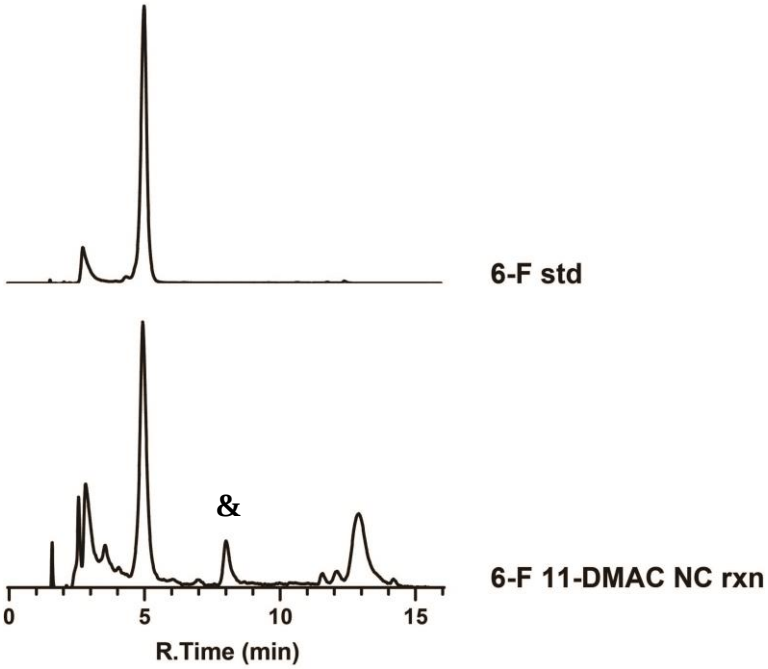


Figure S5.11: 5-chloro (**5.15**) 11-DMAC reaction. (&=11-DMAC hypothesized derivative)

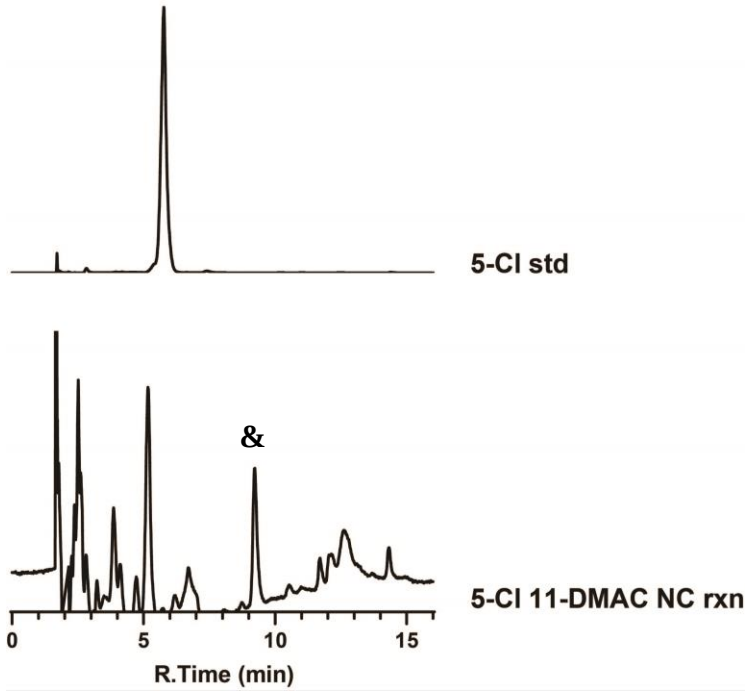


Figure S5.12: 6-chloro (**5.16**) 11-DMAC reaction. (&=11-DMAC confirmed derivative)

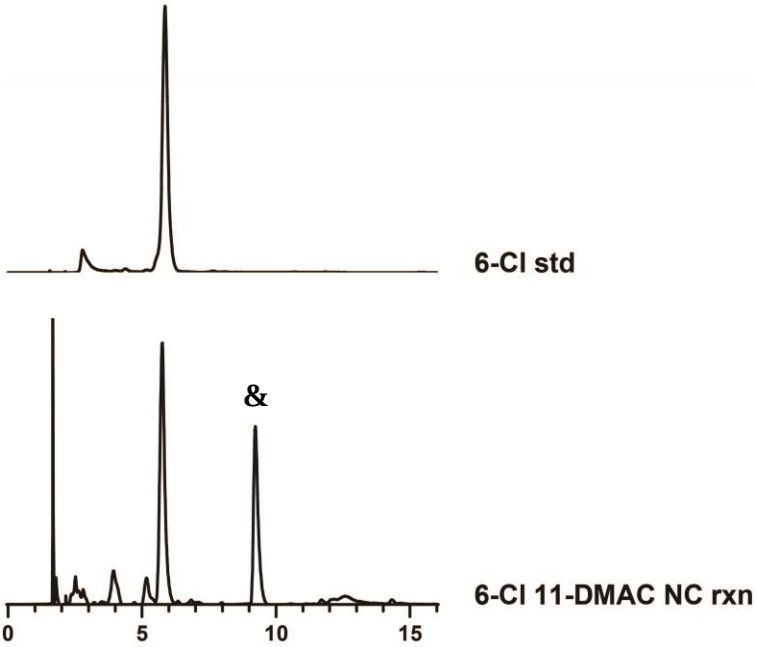


Figure S5.13: 6-bromo (**5.18**) 11-DMAC reaction. (&=11-DMAC confirmed derivative)

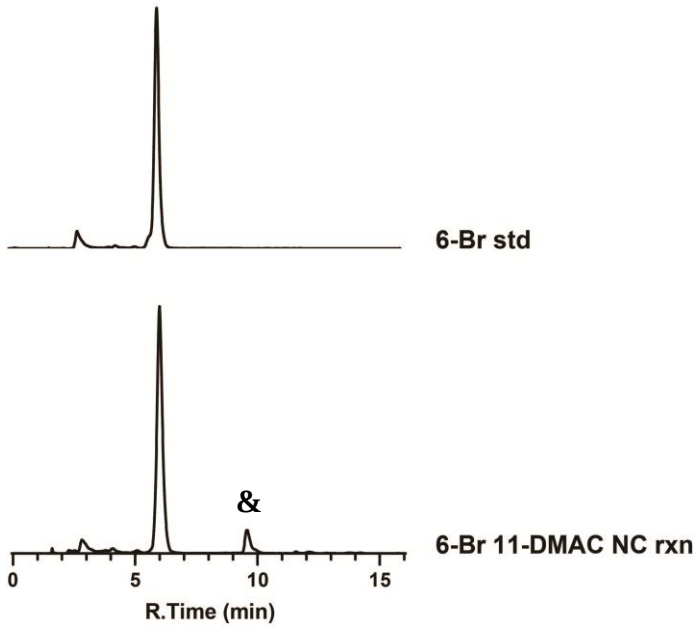


Figure S5.14: 7-azaindole (**5.20**) 11-DMAC reaction. (*****=C-2 prenylated shunt metabolite, &=11-DMAC hypothesized derivative)

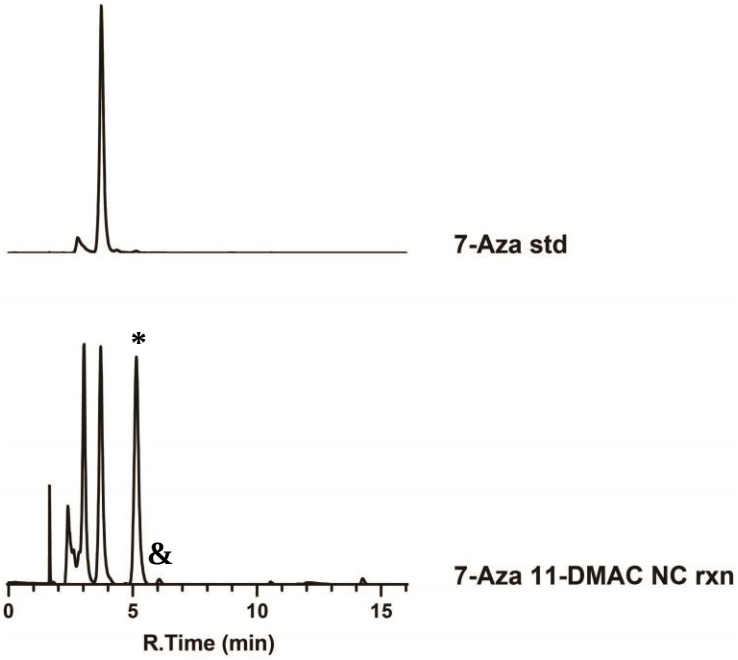
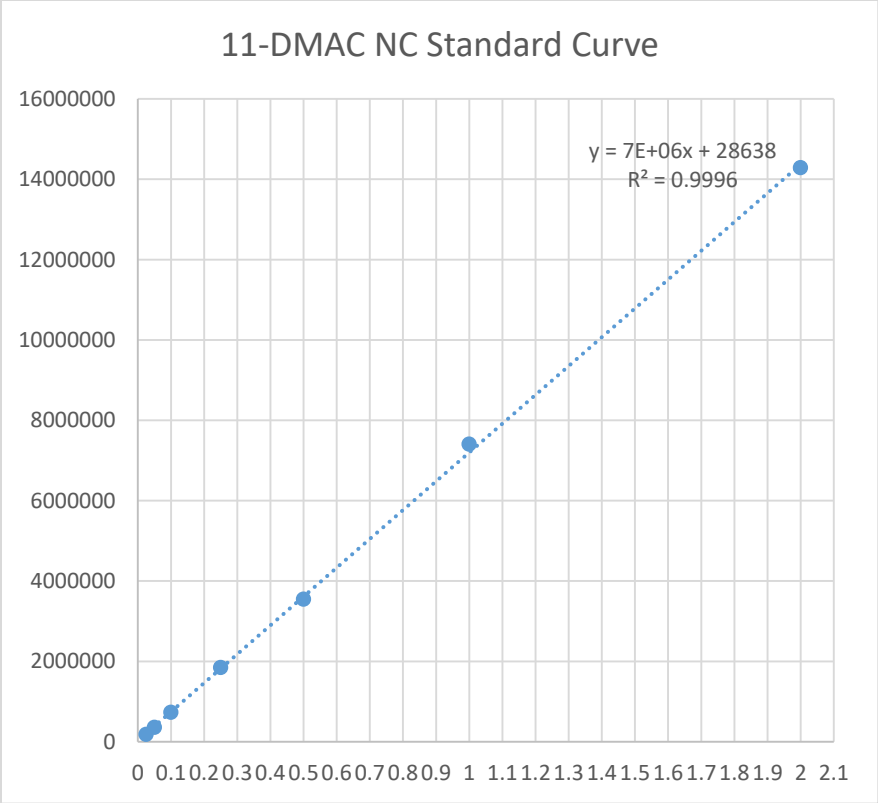


Table S5.2: 11-DMAC NC kinetics tables

Concentration (mM)	Ret	Area	Height
0.025	8.077	180040	13433
0.05	8.086	356789	26407
0.1	8.074	730835	54192
0.25	8.055	1845612	136432
0.5	8.055	3544316	266798
1	8.041	7405890	559120
2	8.034	14285134	1095933

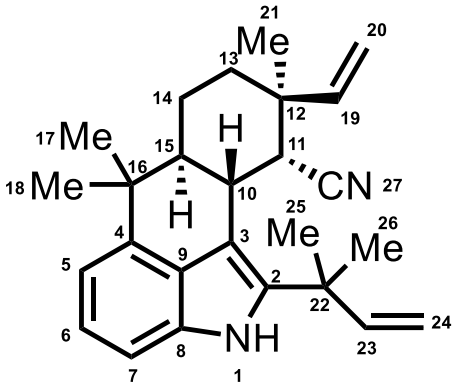


Equation: Peak area=(7171500(mmols produced))+28658

Catalyst ratio=average mmols produced with cyclase/average mmol produced without cyclase

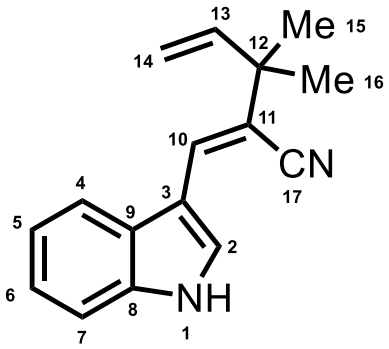
WT Rxns	11-DMAC peak area	Mmols produced	Avg	Catalyst ratio
1	1468766	0.200809872	0.201516	127.9476465
2	1317357	0.179697274		
3	1635357	0.224039462		
L147F Rxns				
1	910805	0.123007321	0.145214	92.20032463
2	1047032	0.142002928		
3	1252345	0.170631946		
No Cyclase				
1	46720	0.00251858	0.001575	N/A
2	40985	0.001718887		
3	32154	0.000487485		

Table S5.3:12-*epi*-ambiguine H nitrile (**5.5**) characterization



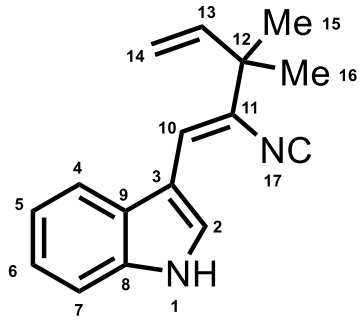
Position	^{13}C shift (ppm), C_6D_6	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		7.26, bs		
2	137.32			
3	108.35			
4	141.03			
5	113.14	7.13, d, (7.3)		3,6,7
6	122.73	7.30, t, (15.5, 7.6)	7	4,5,9
7	108.28	7.02, d, (8.3)	6	5,8
8	127.61			
9	132.81			
10	35.16	3.09, dd, (10.8, 3.5)	11,15	3,11,15,27
11	47.42	3.82, dd, (3.6, 1.6)	10,14	10,12,13,15
12	39.97			
13	32.04	1.70, m	14	12,14
14	22.54	1.68, dt, (12.7,3.1) & 1.45, m	13,15	12,13
15	47.11	2.25, ddd, (12.9,10.9,2.6)	10,14	10,16
16	36.77			
17	23.77	1.39, s		4,15,16,18
18	24.06	0.98, s		4,15,16,17
19	143.39	5.79, dd, (17.6, 10.9)	20	11,12,13
20	114.42	5.03, m	19	12,19,21
21	29.79	1.27, s		11,12,13,19
22	38.79			
23	147.08	6.08, dd, (17.5, 10.6)	24	2,22,25
24	112.74	5.03, m	23	22,23
25	28.44	1.31, s		2,22,23,26
26	28.10	1.10, s		2,22,23,25
27	119.97			

Table S5.4: 11-DMAC nitrile (**5.6**) characterization



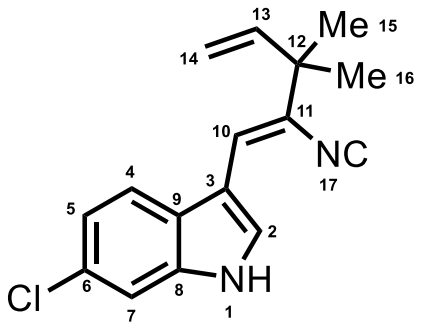
Position	^{13}C shift (ppm), C_6D_6	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		6.83, bs		
2	125.50	8.20, d, (2.8)	1	3,8,9
3	111.50			
4	117.20	7.60, dd, (7.1,1.7)	5	6,8
5	121.04	7.14	4	7
6	123.16	7.18	7	4
7	111.85	6.93, d, (7.2)	6	5,9
8	135.65			
9	127.61			
10	131.03	7.24, s		2,12,17
11	114.10			
12	41.15			
13	144.67	5.75, dd, (17.4, 10.6)	14	12,15,16
14	113.32	5.05, d, (17.4) & 5.01, d, (10.6)	13	12,13
15	26.66	1.22, s		11,12,13,15
16	26.66	1.22, s		11,12,13,16
17	120.19			

Table S5.5: 11-DMAC isonitrile (**5.8**) characterization



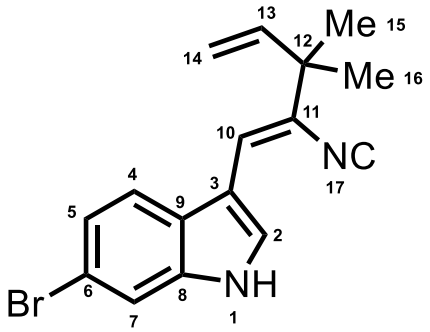
Position	^{13}C shift (ppm), C_6D_6	^{1}H shift (ppm), multi (J) C_6D_6	COSY	HMBC
1		6.84, bs	2	
2	125.63	8.00, s	1	3,8,9
3	110.44			
4	118.29	7.62, d, (7.7)	5	3,5,8
5	123.08	7.19, m	4	4,8
6	120.76	7.16 (in C_6D_6)	7	7
7	111.68	6.95, d, (7.3)	6	
8	135.43			
9	127.69			
10	117.45	6.70, s		2,11,12
11	127.61			
12	41.75			
13	143.79	5.75, ddd, (17.4,10.6,1.8)	14	11,12,15,16
14	113.53	(cis) 5.01, dd, (10.7,1.9) (trans) 5.06, dd, (17.4,1.9)	13	12,13
15	25.47	1.18, s		11,12.13,16
16	25.47	1.18, s		11,12,13,15
17	172.32			

Table S5.6: 6-chloro 11-DMAC (**S5.1**) characterization



Position	^{13}C shift (ppm), C_6D_6	^{1}H shift (ppm), multi (J) C_6D_6	COSY	HMBC
1		6.64, bs	2	
2	126.17	7.87, d, (2.8)	1	3,8,9
3	110.47			
4	119.23	7.27, d, (8.5)		3,8
5	121.40	7.14, d		
6	129.04			
7	111.77	6.96, d, (1.8)		5,6,8,9
8	135.64			
9	126.09			
10	116.80	6.50, s		9
11	127.61			
12	41.60			
13	143.57	5.72, dd, (17.4,10.6)	14	12,15,16
14	113.68	(cis) 5.01, dd, (10.6,0.9) (trans) 5.04, dd, (17.4,0.9)	13	12,13
15	25.42	1.16, s		11,12,13,16
16	25.42	1.16, s		11,12,13,15
17	172.63			

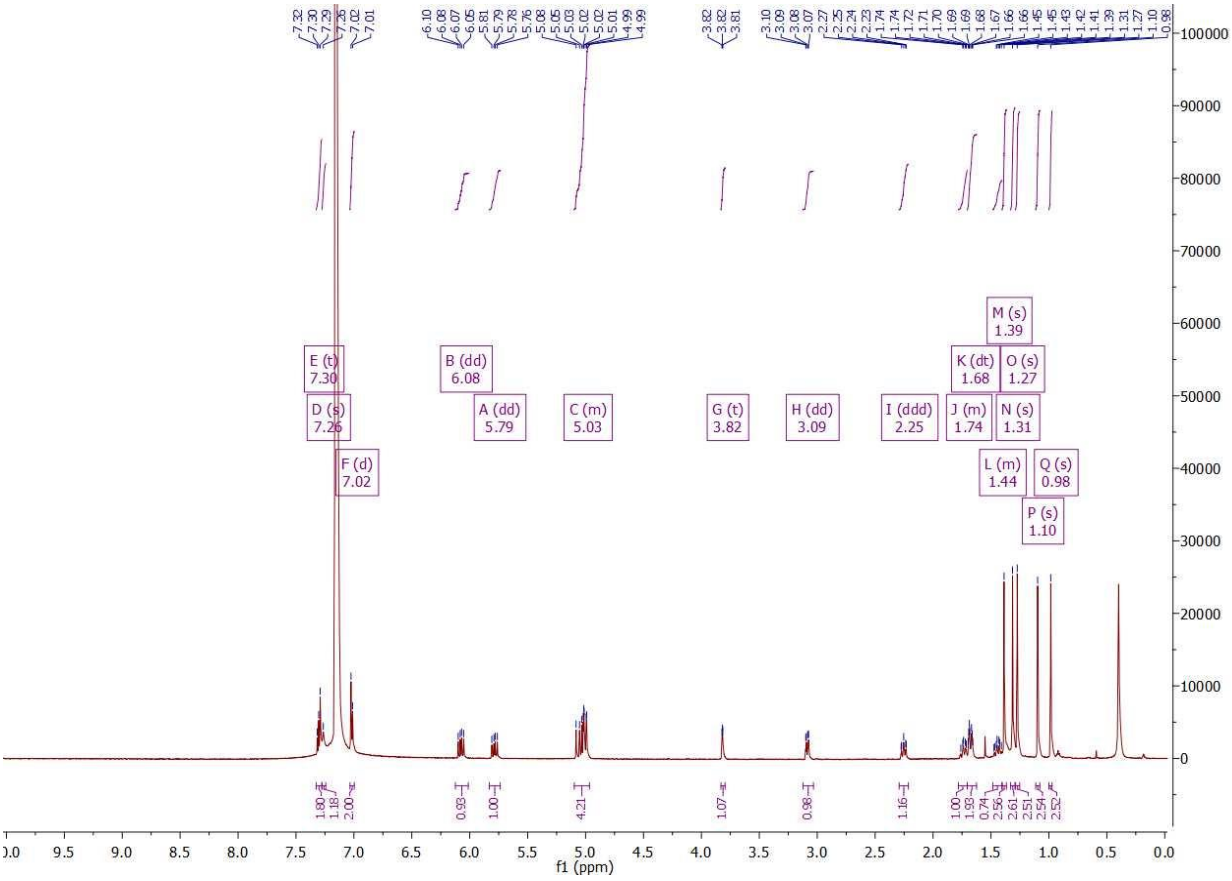
Table S5.7: 6-bromo 11-DMAC (**S5.2**) characterization

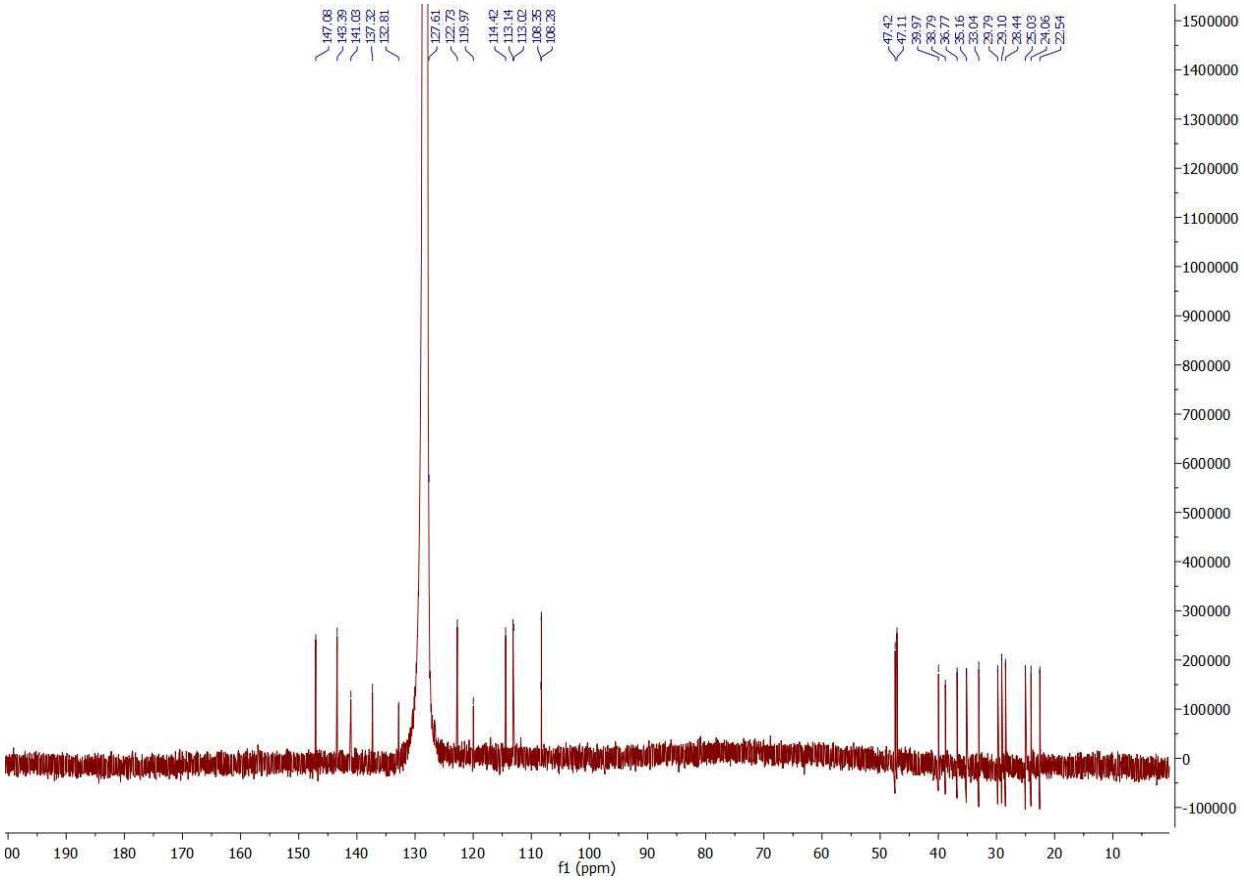


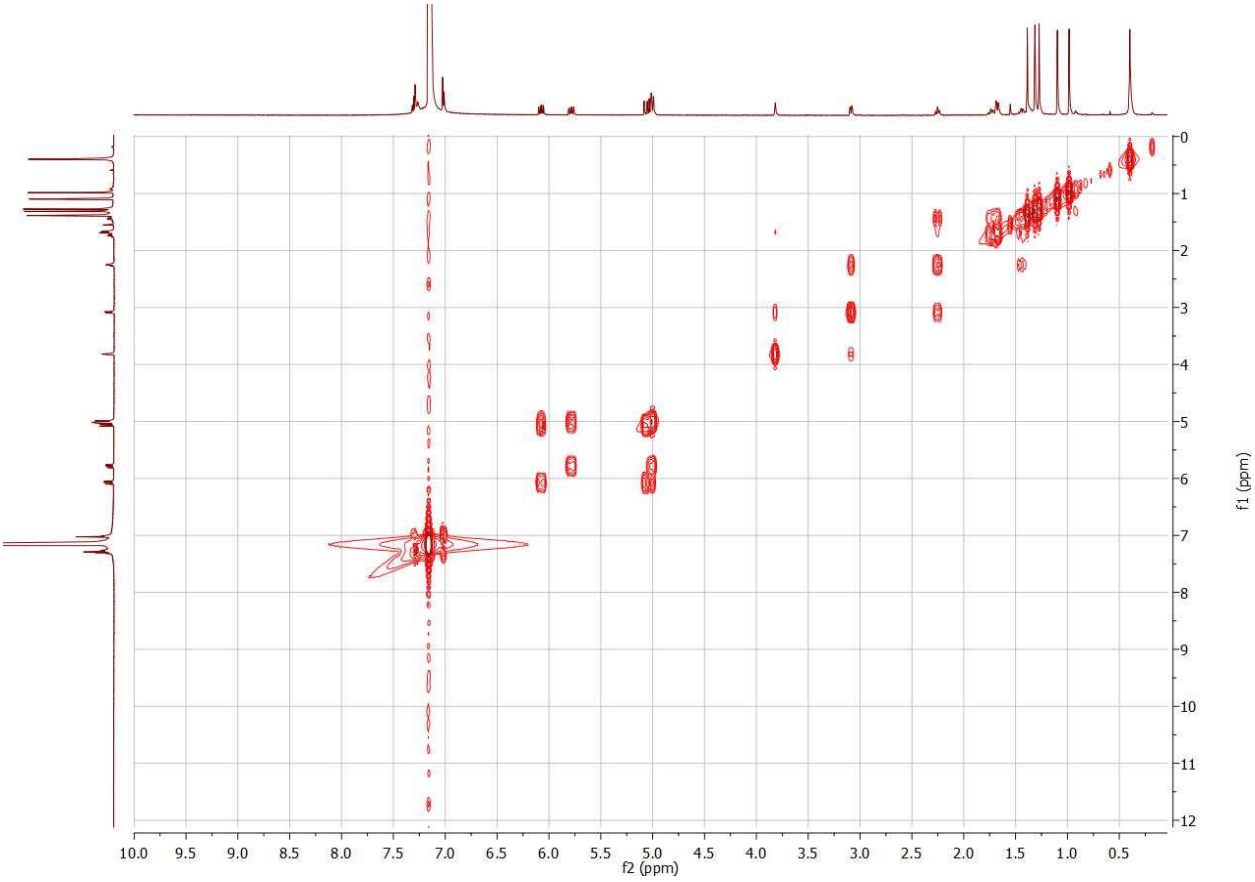
Position	^{13}C shift (ppm), C_6D_6	^{1}H shift (ppm), multi (J) C_6D_6	COSY	HMBC
1		**6.62, bs**	**2**	
2	**126.09**	**7.85, d, (2.8)**	**1**	**3,8,9**
3	**110.46**			
4	**123.97**	**7.28, d**		**7**
5	**119.56**	**7.21, d, (8.3)**		**6,8**
6	**116.66**			
7	**114.77**	**7.11, s**		**4,6,9**
8	**136.04**			
9	**126.38**			
10	**116.73**	**6.49, s**		**9,12**
11	**127.61**			
12	**41.77**			
13	**143.55**	**5.72, dd, (17.4,10.6)**	**14**	**12,15,16**
14	**113.69**	**(cis) 5.01, d, (10.8) (trans) 5.04, d, (17.4)**	**13**	**12,13**
15	**25.42**	**1.16, s**		**11,12,13,16**
16	**25.42**	**1.16, s**		**11,12,13,15**
17	**172.65**			

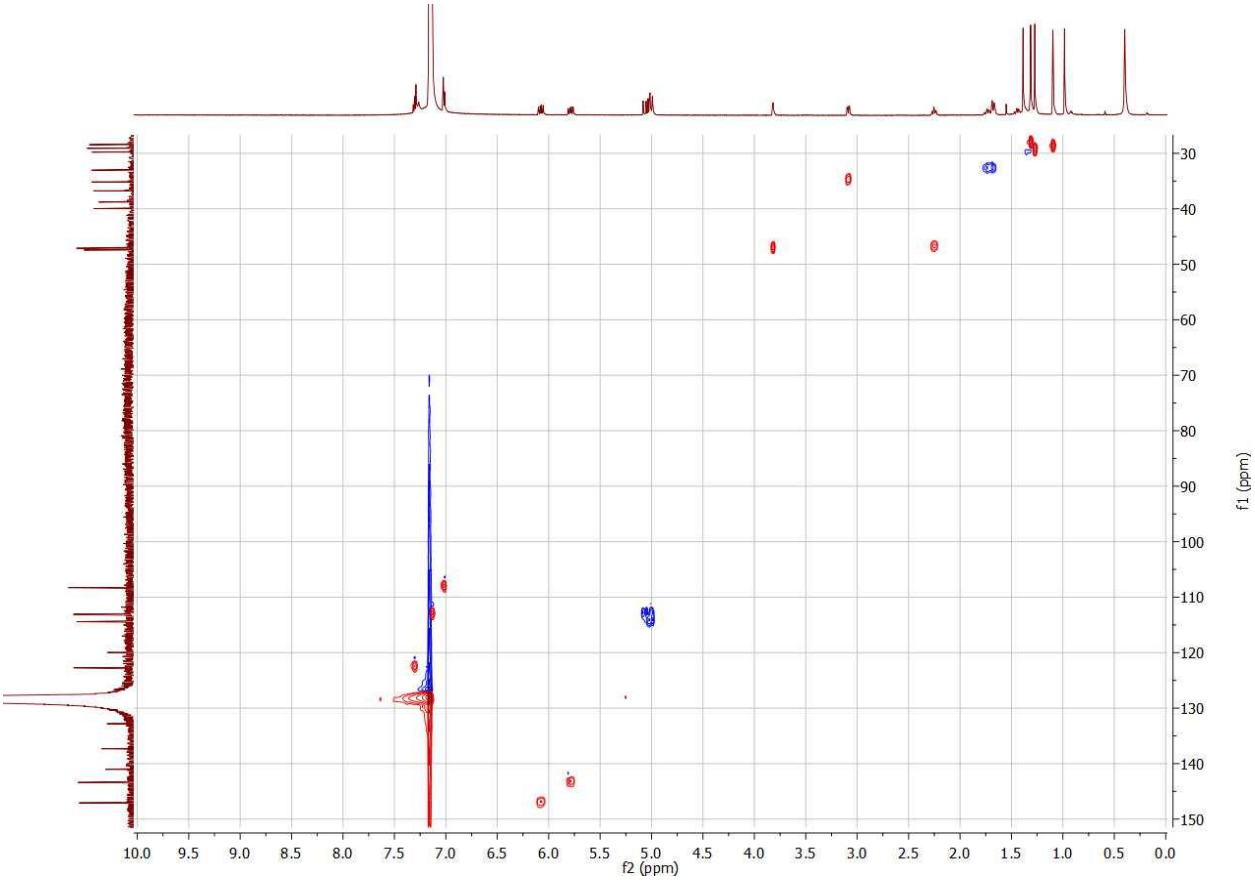
Figures S5.15-5.19: NMR and HRMS traces

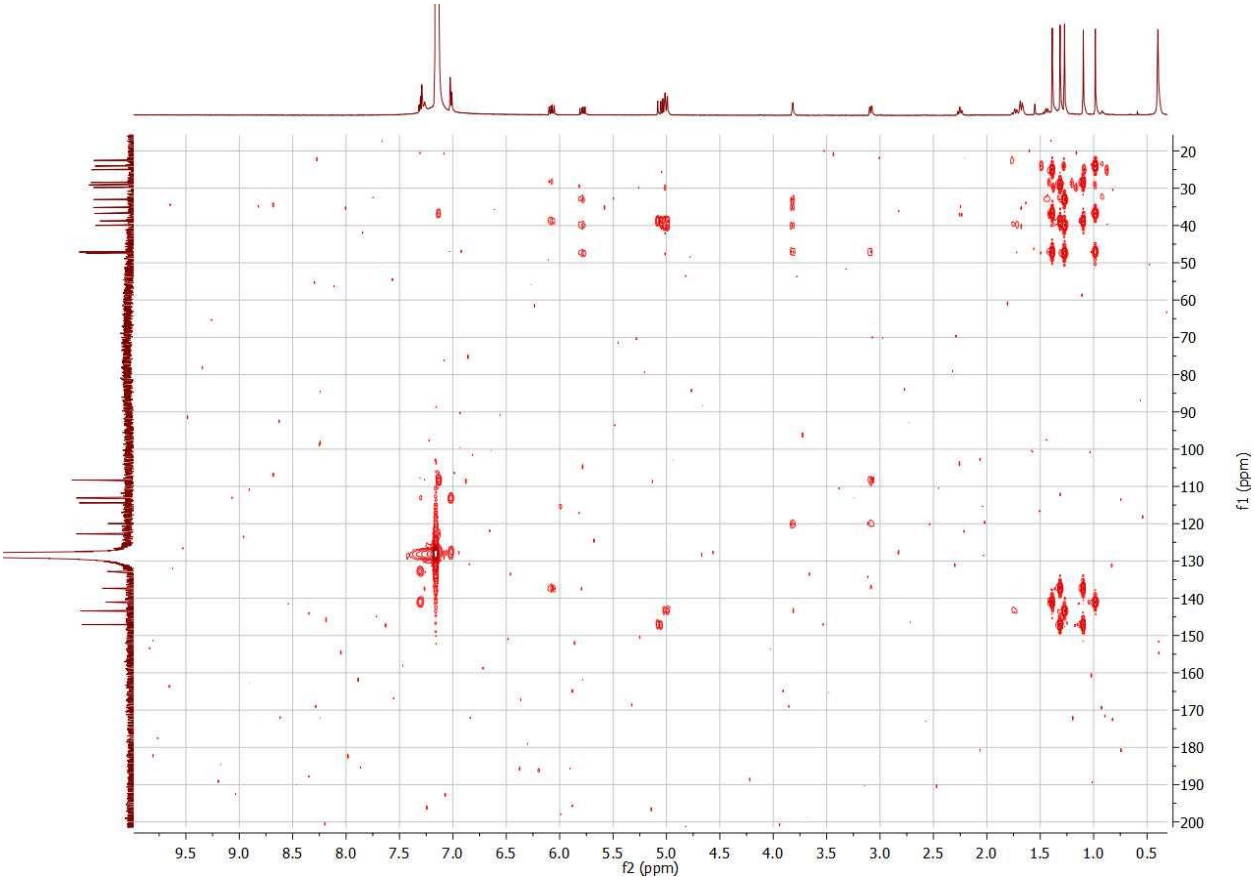
Figure S5.15:^{1}H, ^{13}C, HSQC, HMBC NMR spectra and HRMS of 12-*epi*-ambiguine H nitrile **(5.5)** in C_6D_6 at 600 MHZ and 151 MHZ respectively.

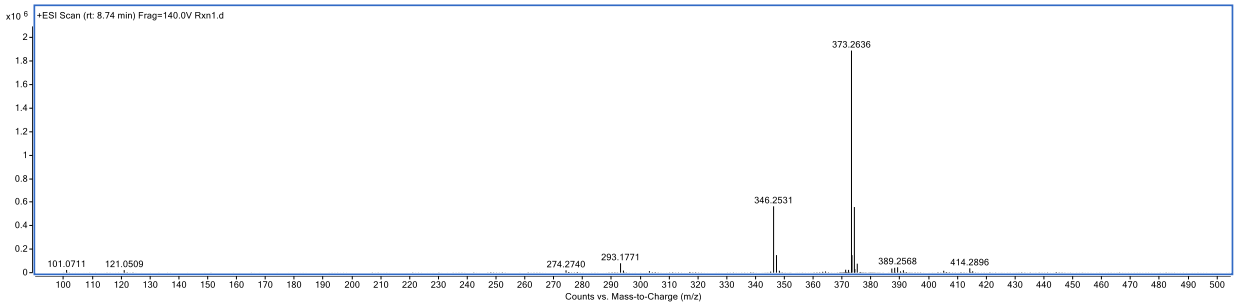



147.08
143.39
141.09
137.32
132.81
127.61
122.73
119.97
114.42
113.14
113.02
108.35
108.28
47.42
47.11
39.97
38.79
36.77
35.16
33.04
29.79
29.10
28.44
25.03
21.06
22.54
1500000
1400000
1300000
1200000
1100000
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
-100000
00
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)


10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
10
11
12
f1 (ppm)

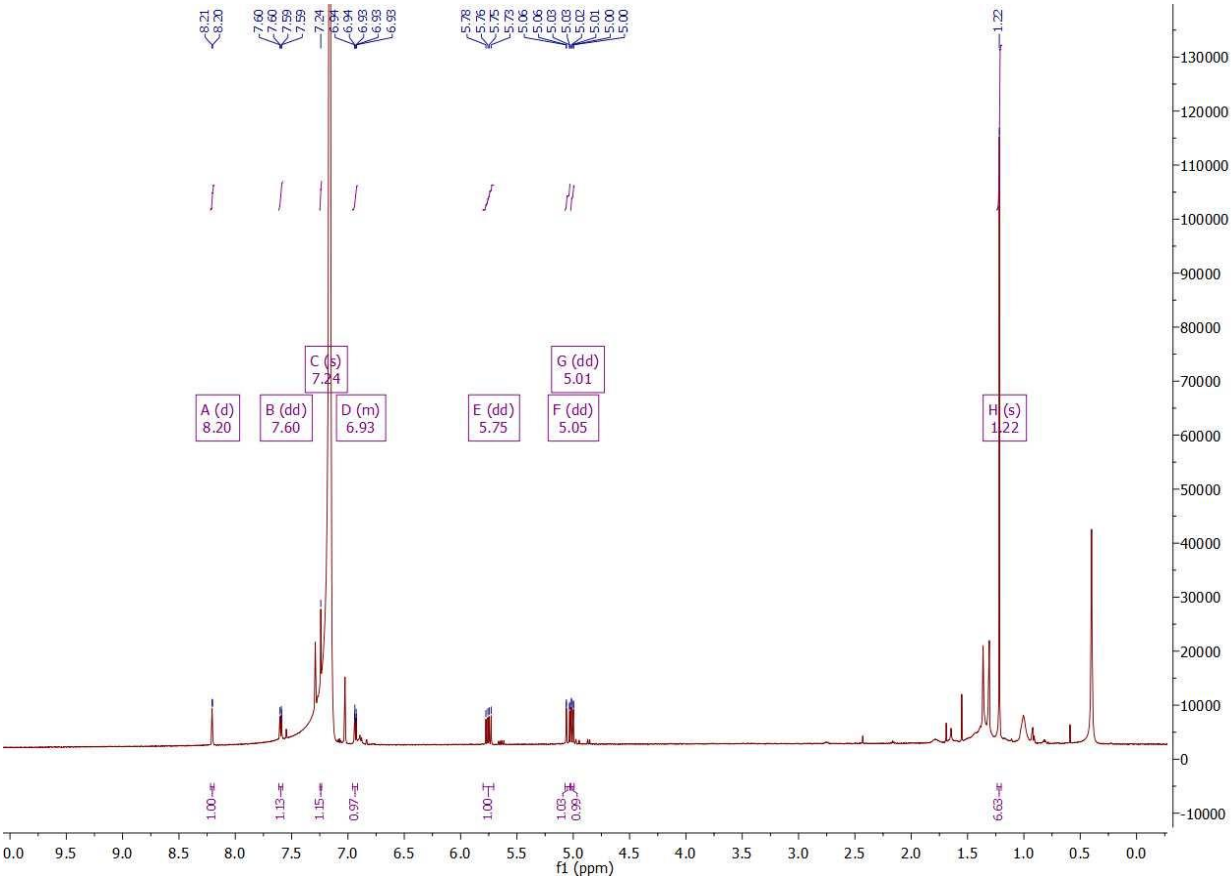

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
30
40
50
60
70
80
90
100
110
120
130
140
150
f1 (ppm)

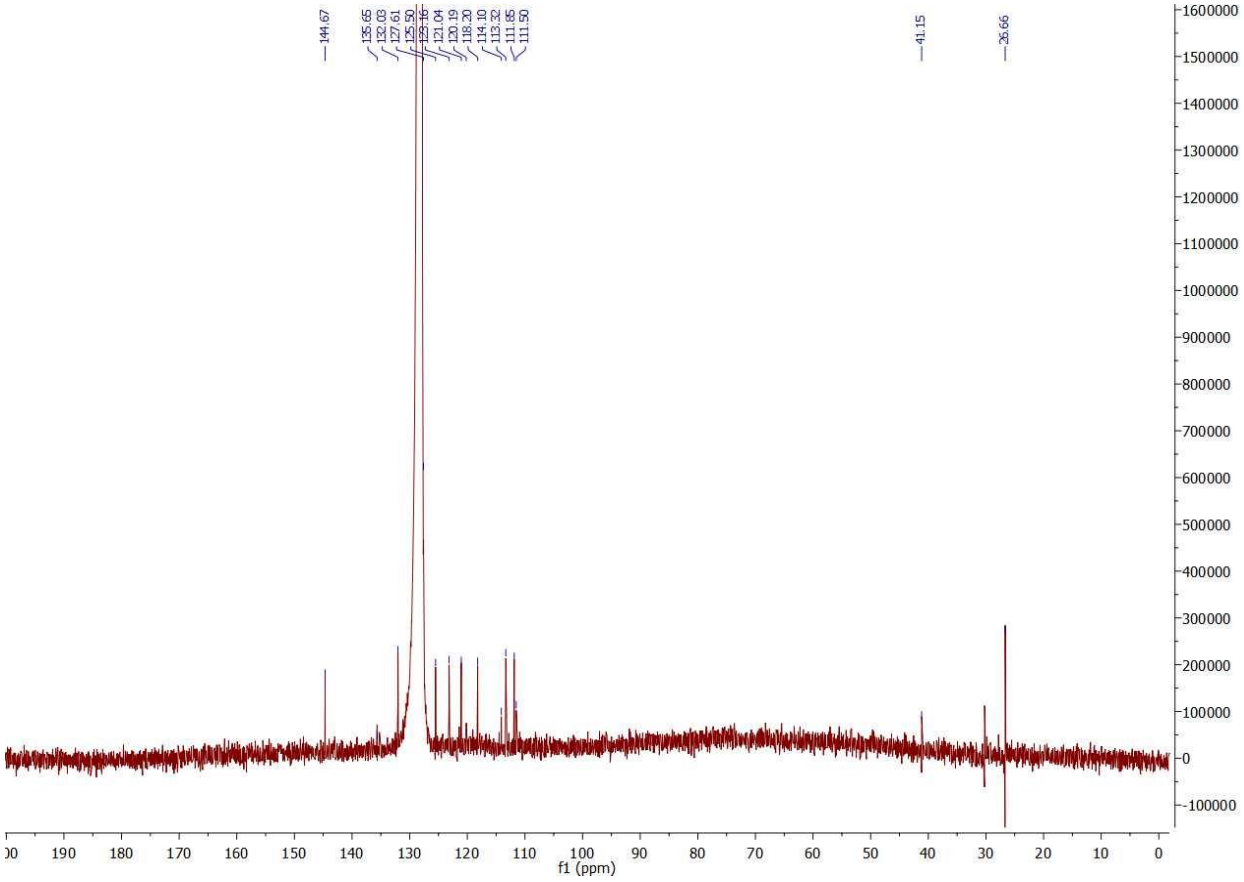


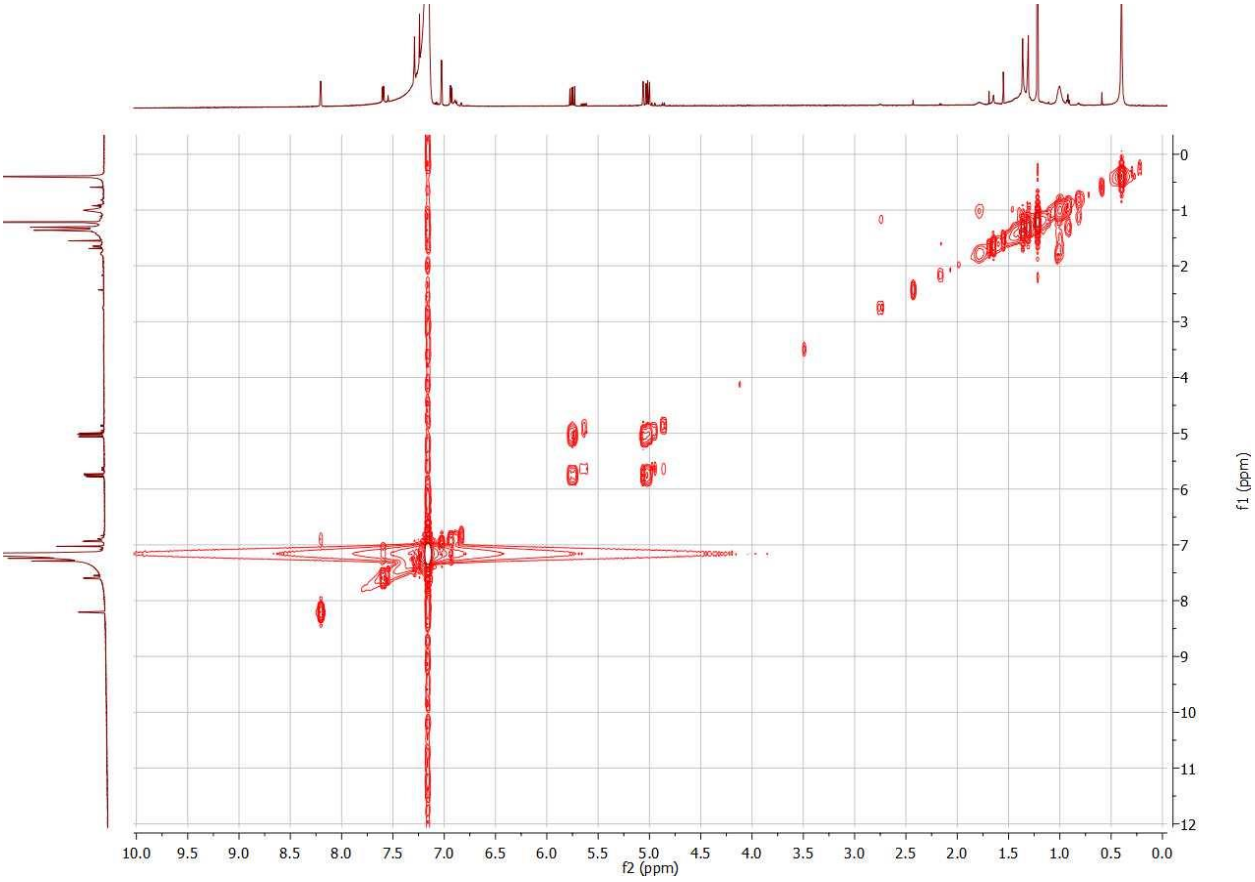


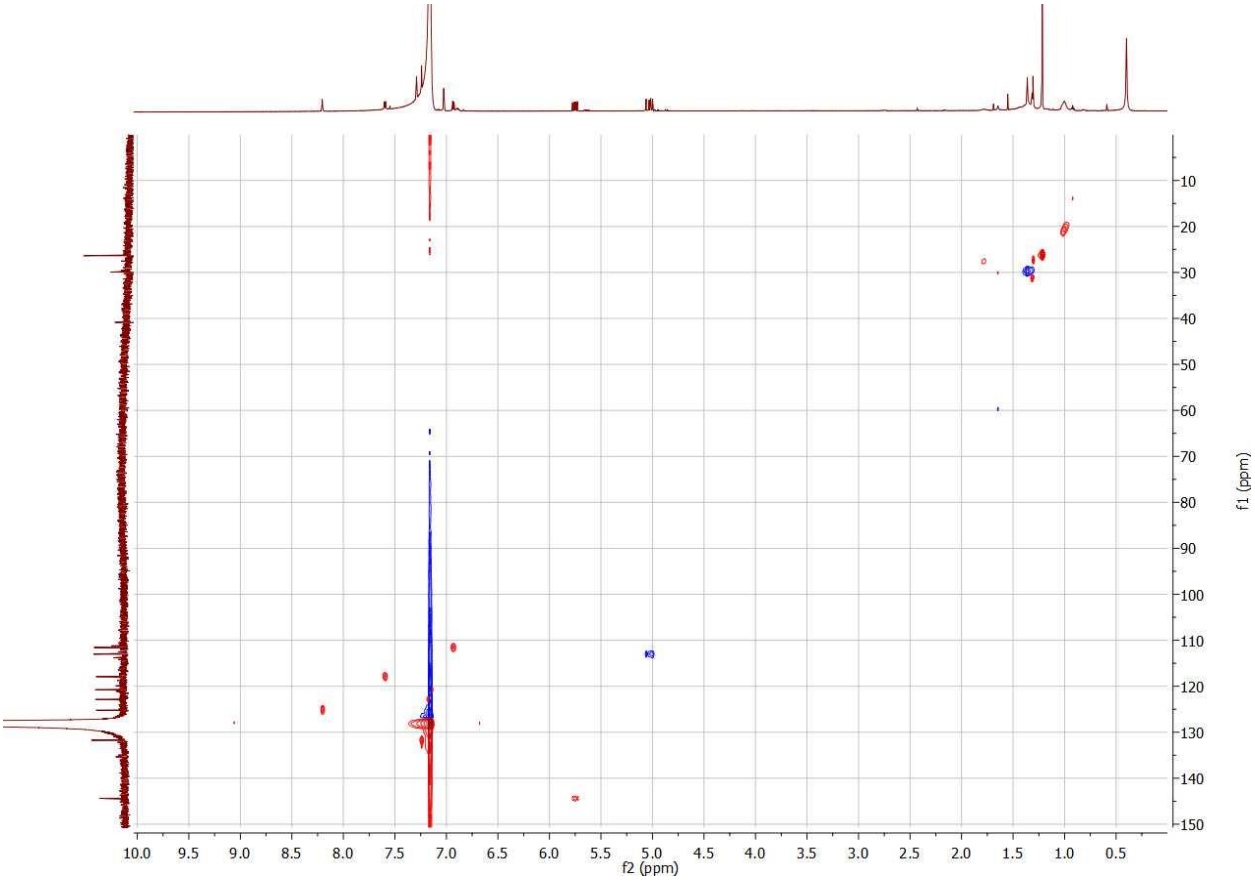
$[M+H]^+$ Calc: 373.2638 Obsv: 373.2636

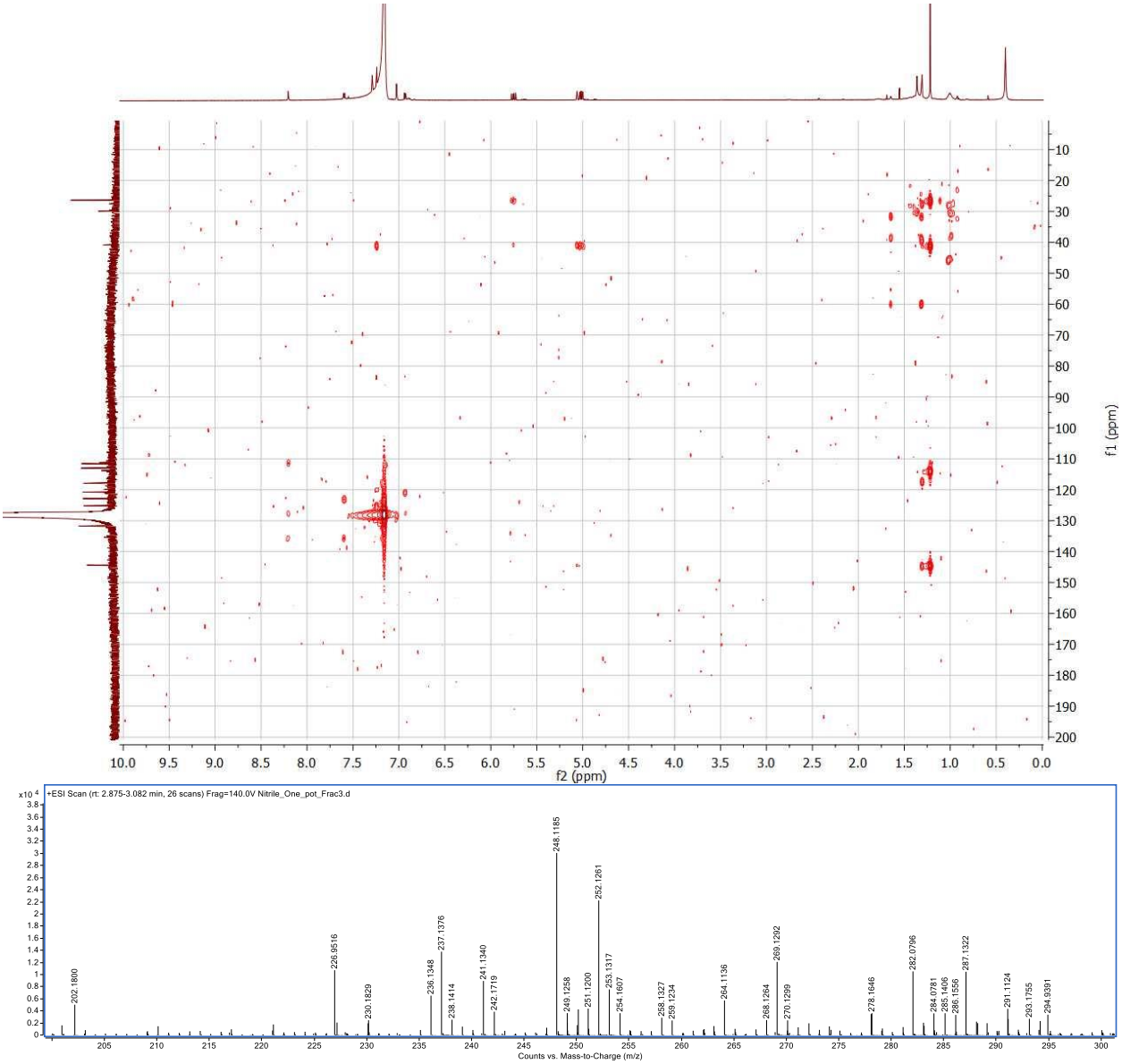
Figure S5.16: ^{1}H, ^{13}C, COSY, HSQC, HMBC NMR spectra and HRMS of 11-DMAC nitrile (**5.6**) in C_6D_6 at 600 MHZ and 151 MHZ respectively.


8.21
8.20
7.60
7.60
7.59
7.59
7.24
6.94
6.93
6.93
6.93
5.78
5.76
5.75
5.73
5.06
5.03
5.02
5.01
5.00
5.00
1.22
C (s)
7.24
G (dd)
5.01
A (d)
8.20
B (dd)
7.60
D (m)
6.93
E (dd)
5.75
F (dd)
5.05
H (s)
1.22
1.00
1.13
1.15
0.97
1.00
1.03
0.99
6.63
f1 (ppm)


144.67
135.65
132.03
127.61
125.50
123.16
121.04
120.19
118.20
114.10
113.32
111.85
111.50
41.15
26.66
1600000
1500000
1400000
1300000
1200000
1100000
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
-100000
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

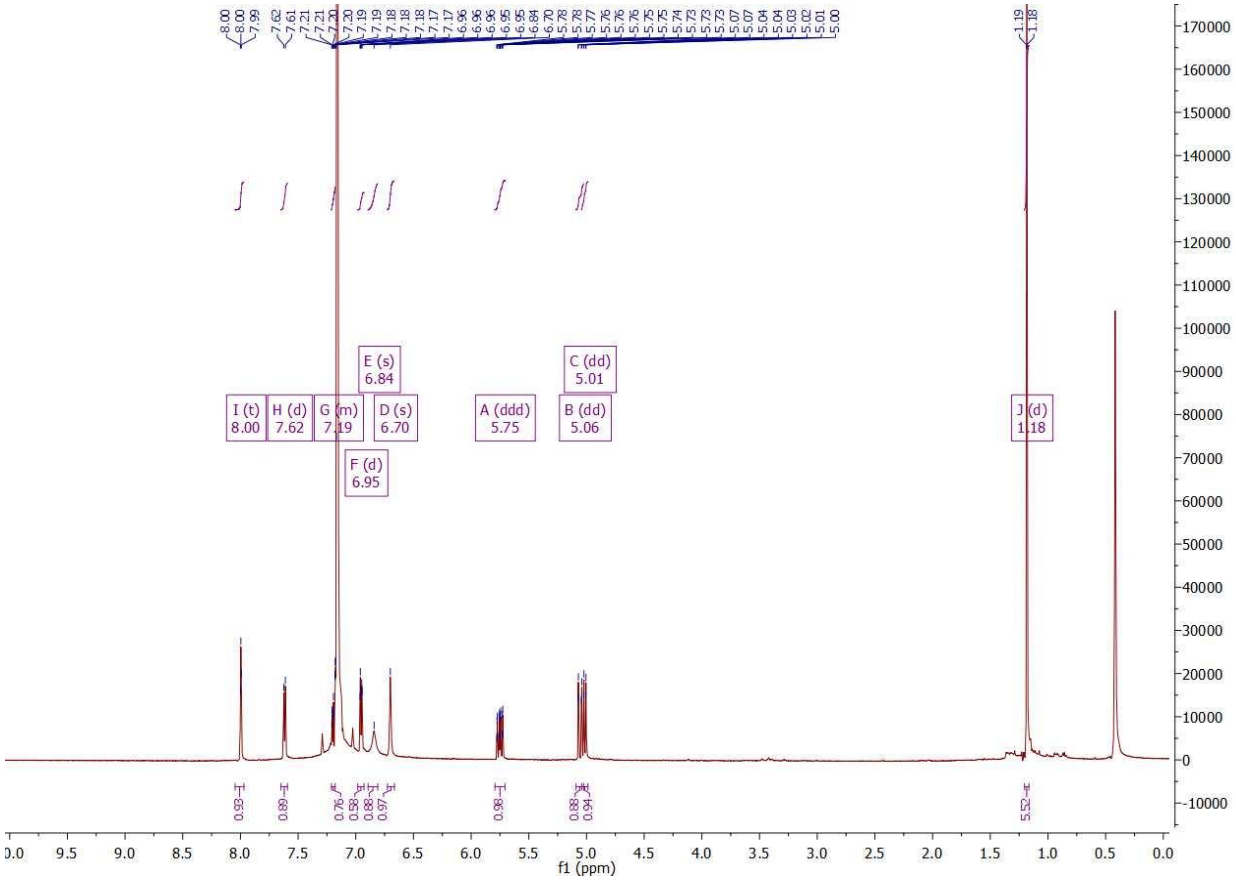

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
10
11
12
f1 (ppm)

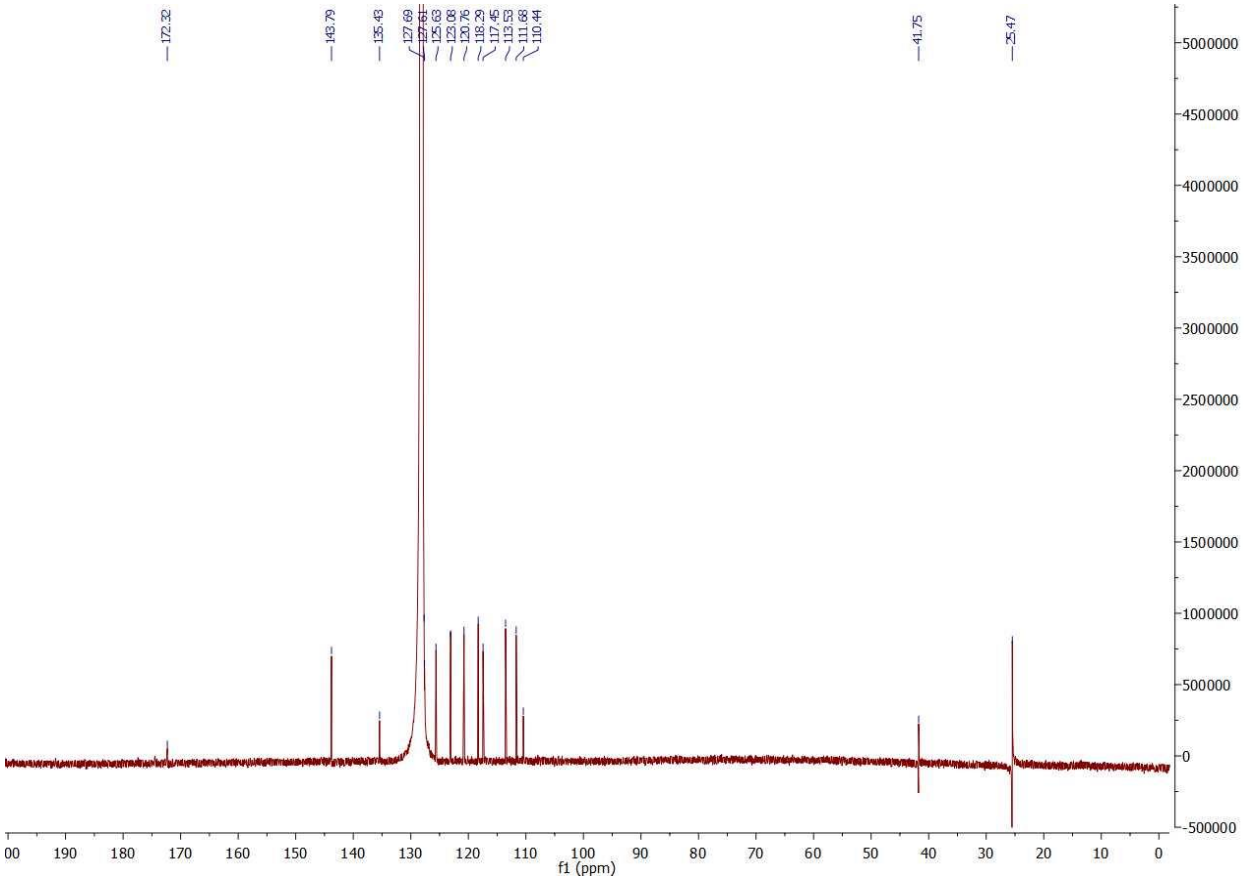

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f2 (ppm)
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
f1 (ppm)

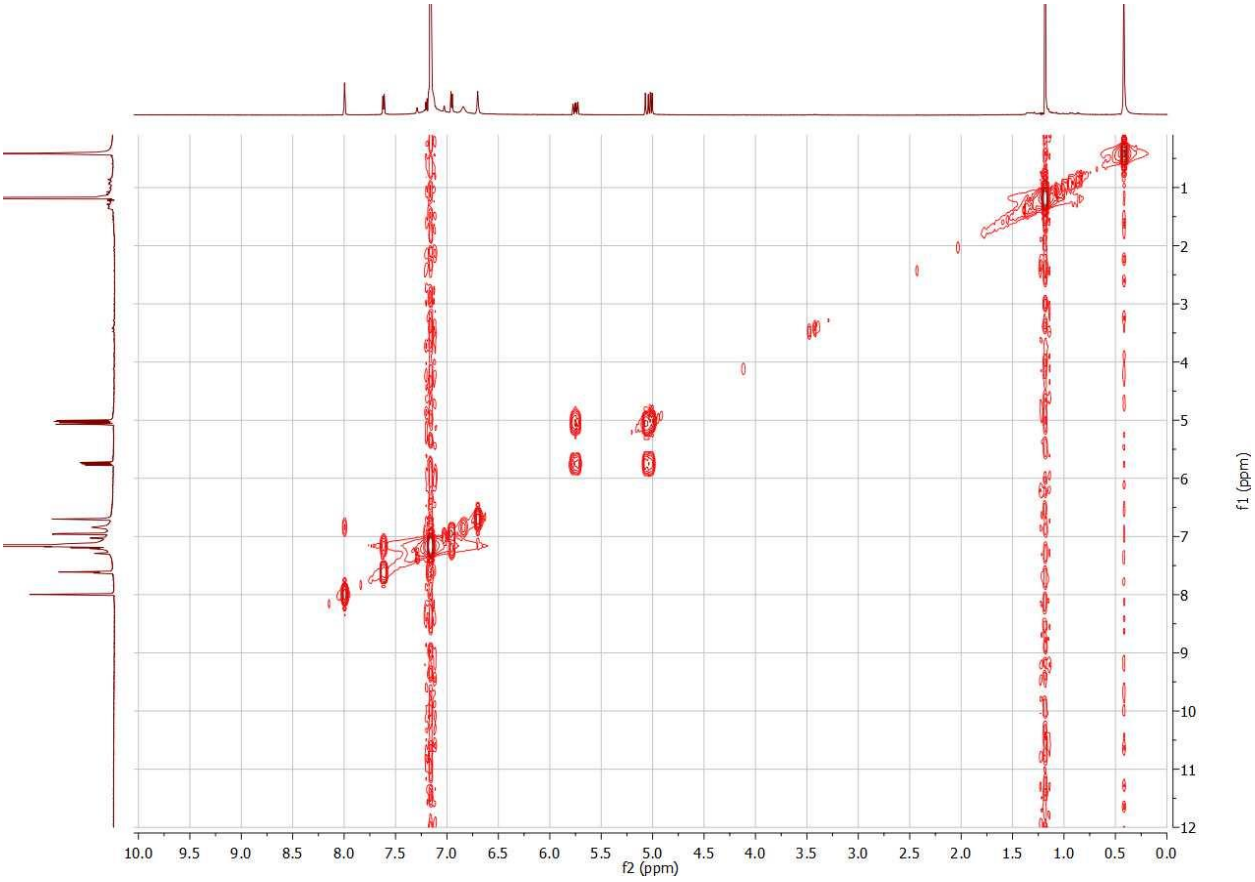


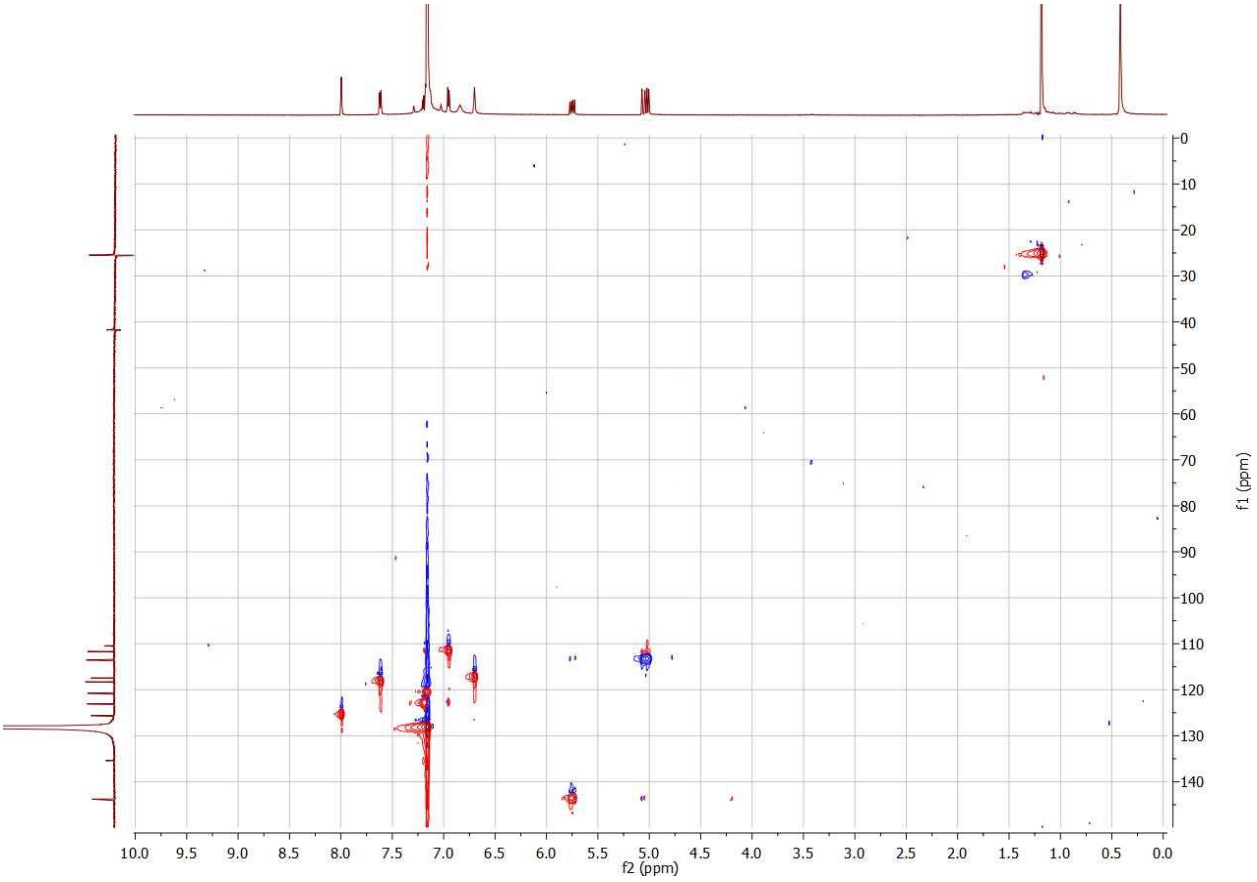
$[M+H]^+$ Calc: 237.1386 Obsv: 237.1376

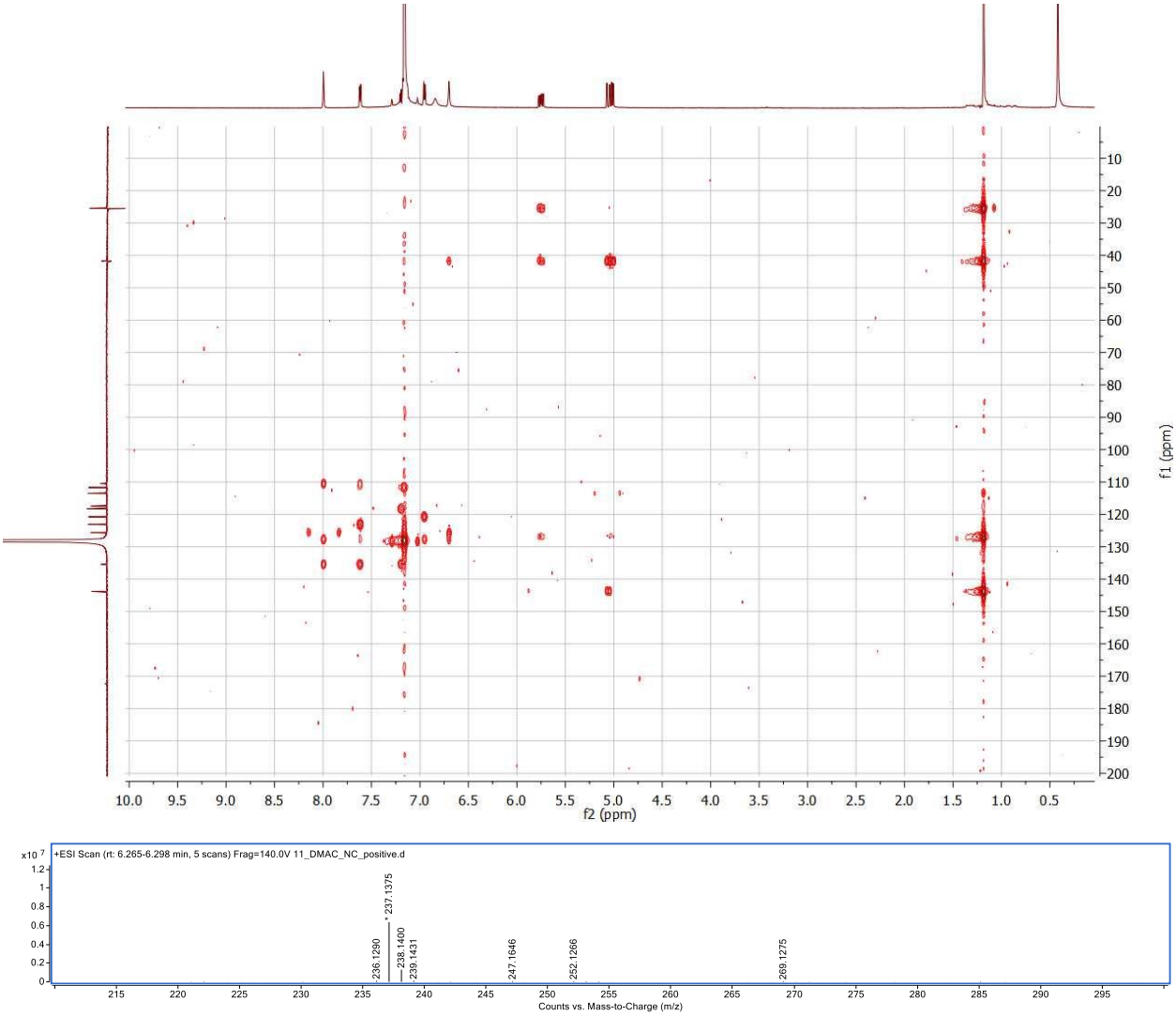
Figure S5.17: ^{1}H, ^{13}C, COSY, HSQC, HMBC NMR spectra and HRMS of 11-DMAC isonitrile (**5.8**) in C_6D_6 at 600 MHZ and 151 MHZ respectively.


I (t)
8.00
H (d)
7.62
G (m)
7.19
F (d)
6.95
E (s)
6.84
D (s)
6.70
A (ddd)
5.75
C (dd)
5.01
B (dd)
5.06
J (d)
1.18
f1 (ppm)


172.32
143.79
135.43
127.69
127.61
125.63
123.08
120.76
118.29
117.45
113.53
111.68
110.44
41.75
25.47
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
-500000
00 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

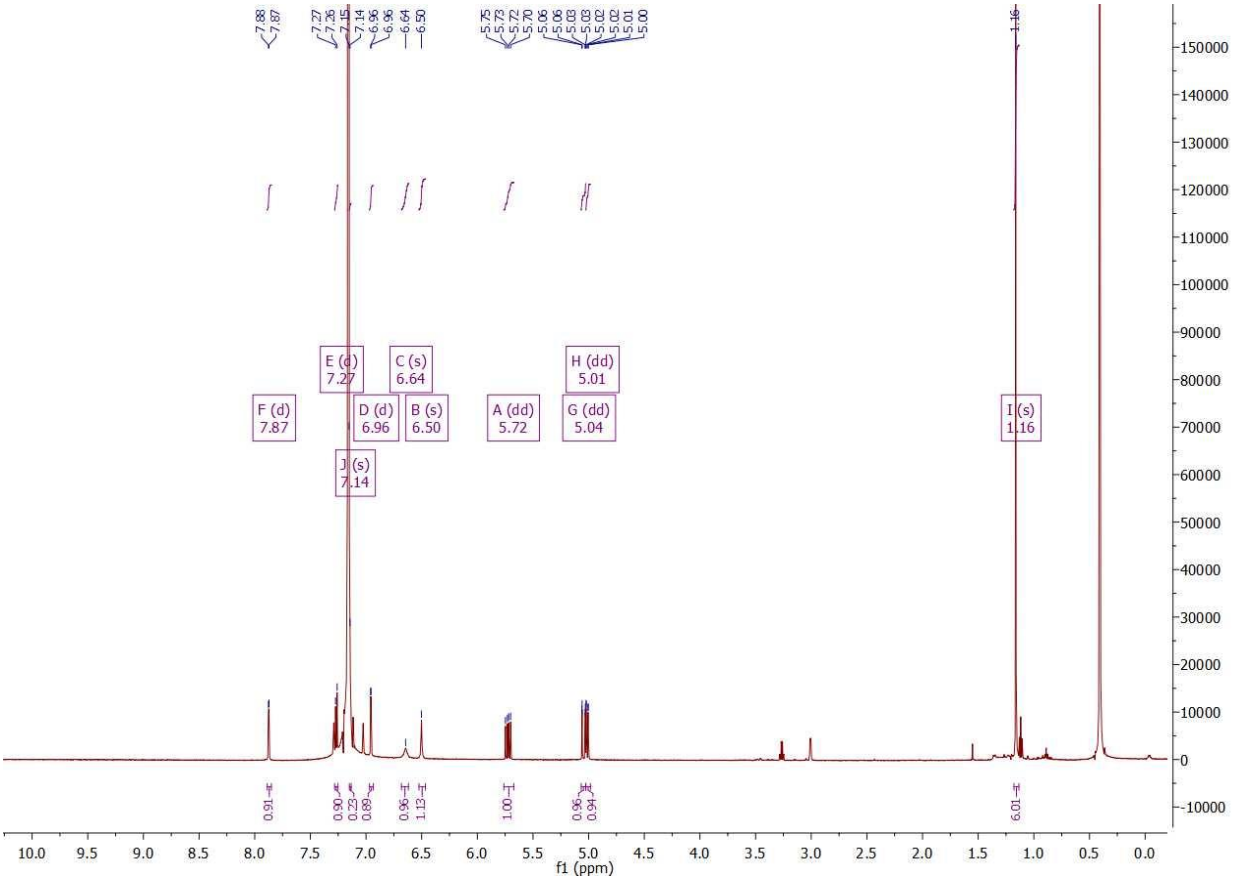

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
1
2
3
4
5
6
7
8
9
10
11
12
f1 (ppm)

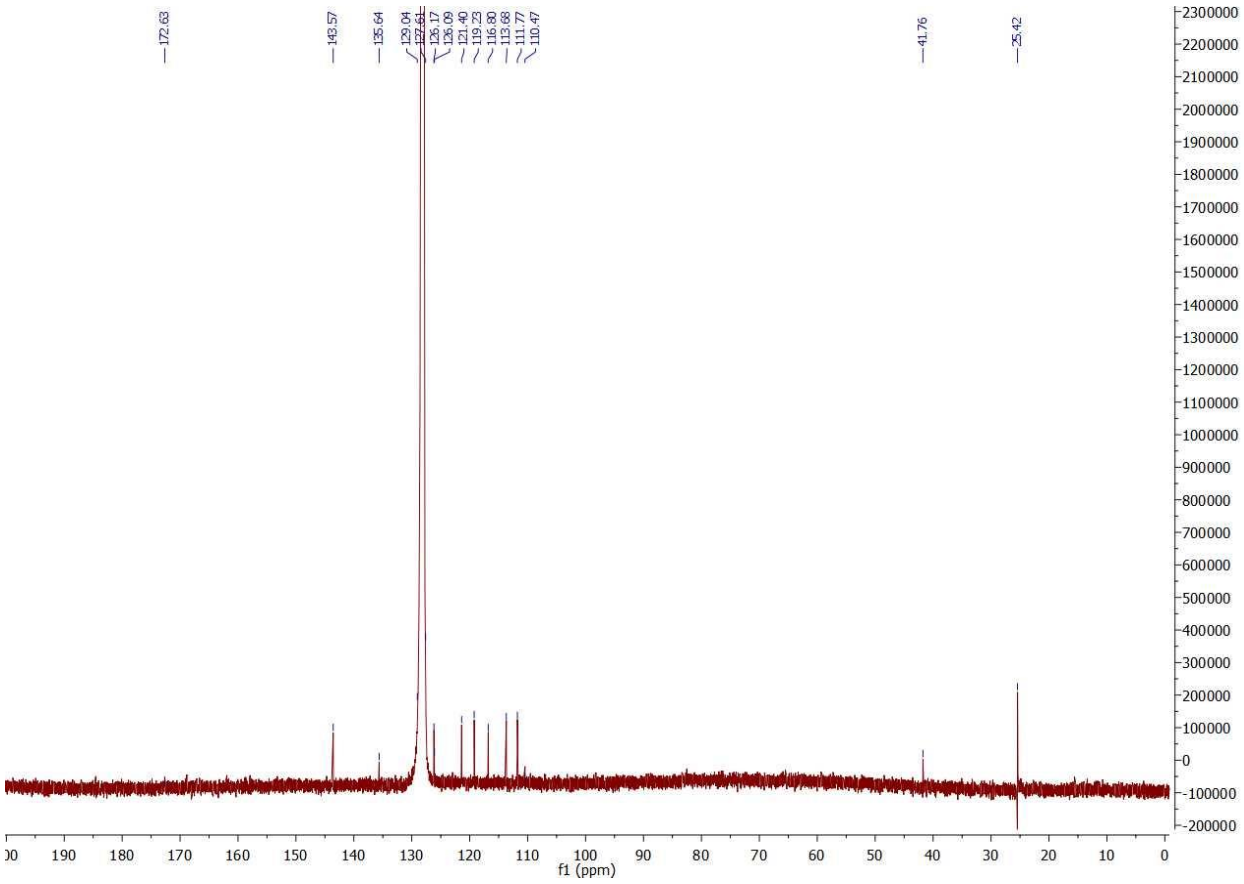

f1 (ppm)
f2 (ppm)

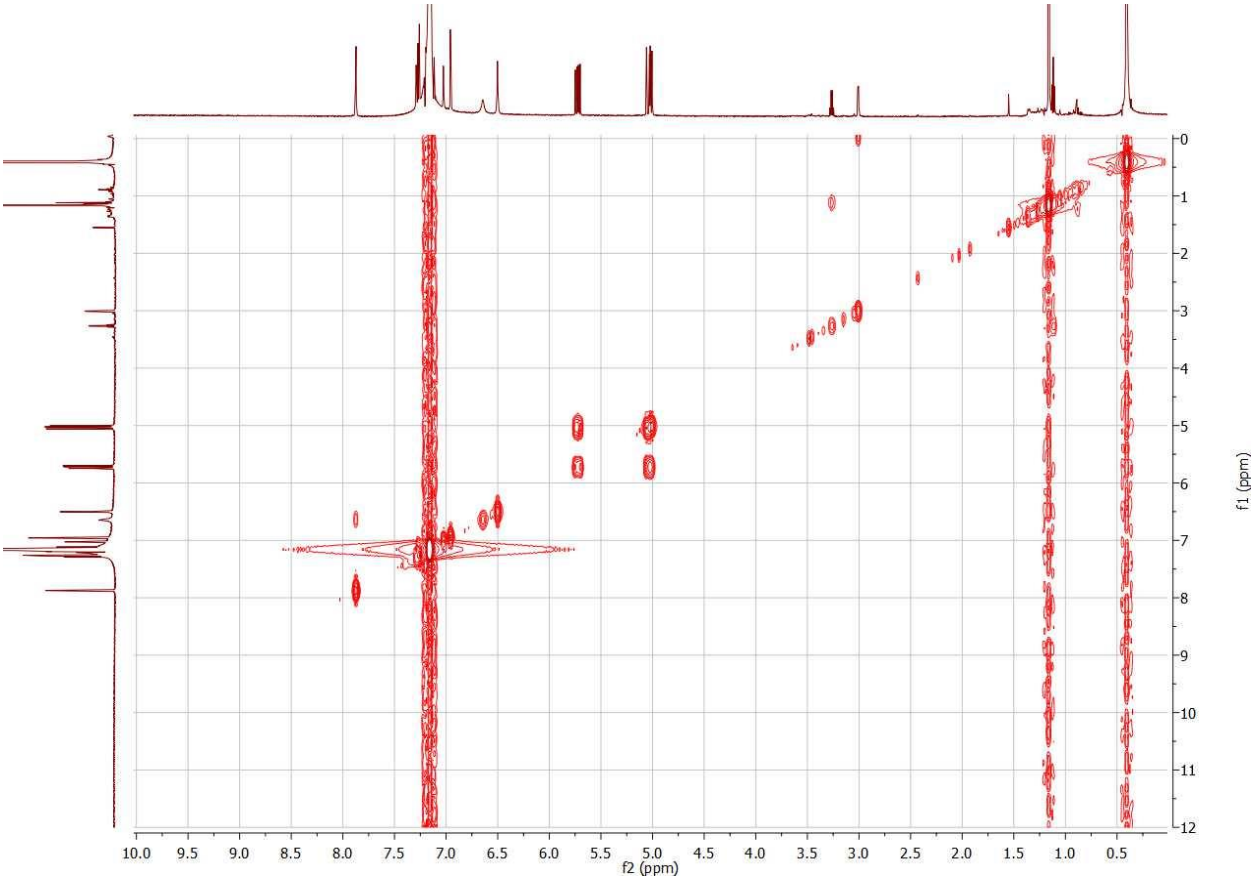


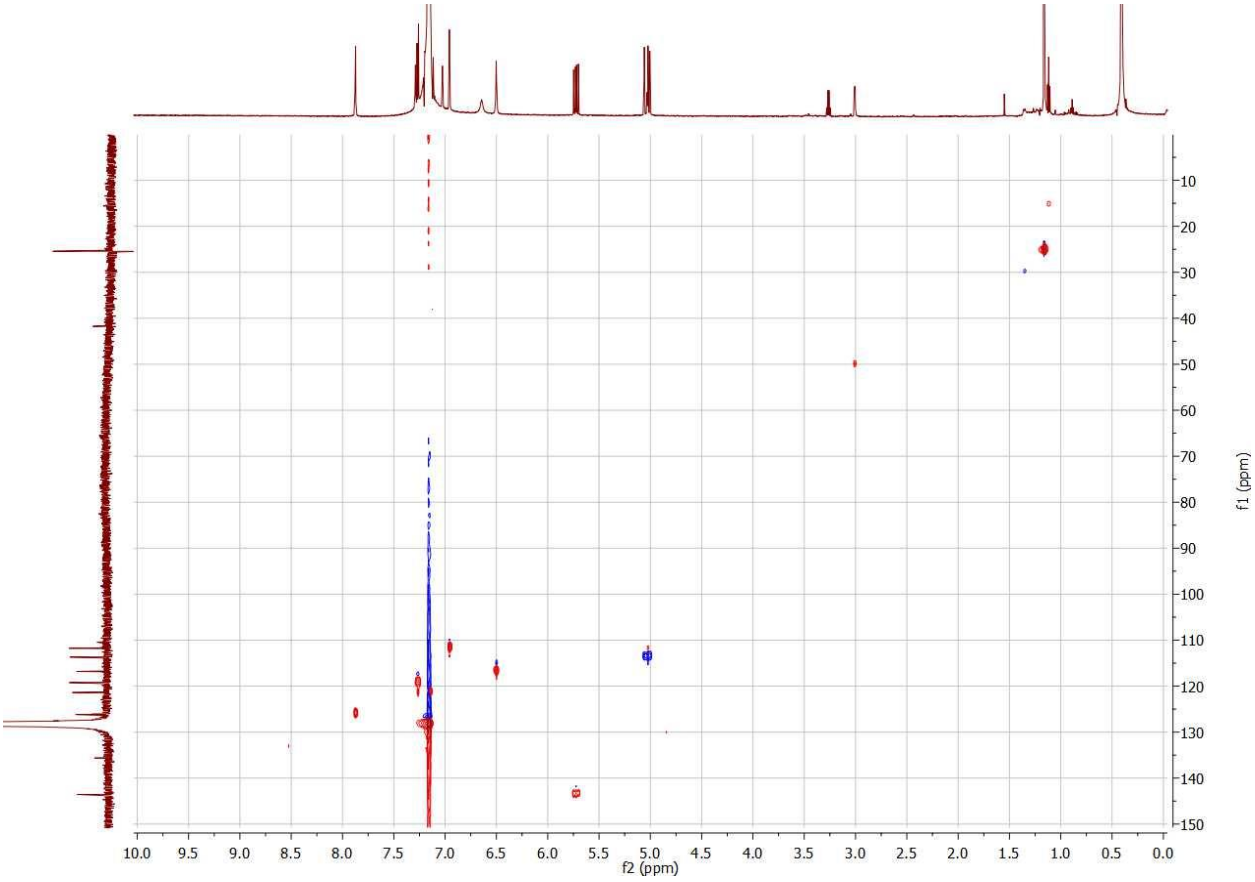
$[M+H]^+$ Calc: 237.1386 Obsv: 237.1375

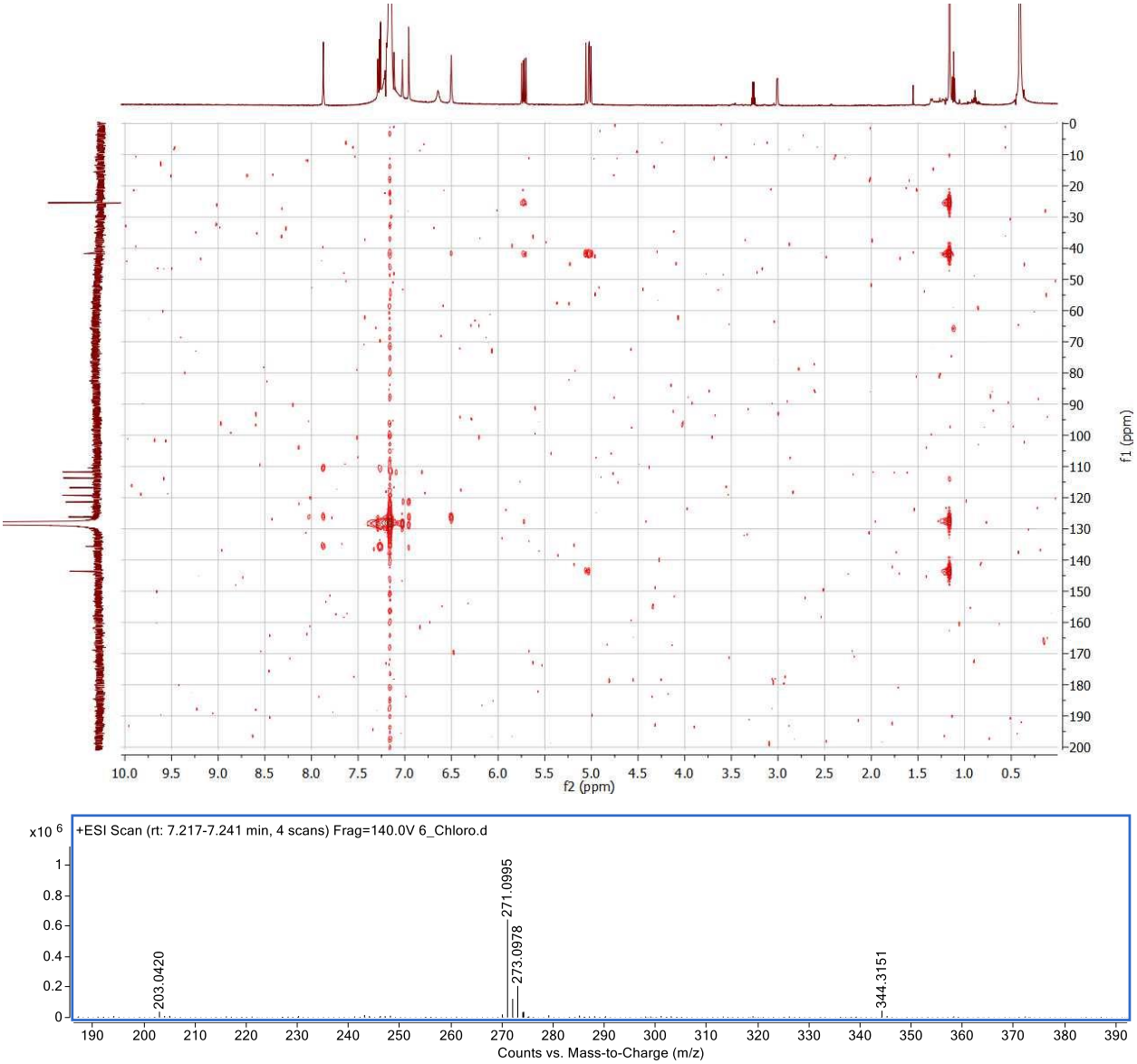
Figure S5.18: ^{1}H, ^{13}C, COSY, HSQC, HMBC NMR spectra and HRMS of 6-chloro 11-DMAC isonitrile (**S5.1**) in C_6D_6 at 600 MHZ and 151 MHZ respectively.


7.88
7.87
7.27
7.26
7.15
7.14
6.96
6.96
6.64
6.50
5.75
5.73
5.72
5.70
5.06
5.03
5.03
5.02
5.02
5.01
5.00
1.16
E (d)
7.27
C (s)
6.64
H (dd)
5.01
F (d)
7.87
D (d)
6.96
B (s)
6.50
A (dd)
5.72
G (dd)
5.04
I (s)
1.16
J (s)
7.14
0.91
0.90
0.23
0.89
0.96
1.13
1.00
0.96
0.94
6.01
f1 (ppm)


172.63
143.57
135.64
129.04
127.61
126.17
126.09
121.40
119.23
116.80
113.68
111.77
110.47
41.76
25.42
f1 (ppm)

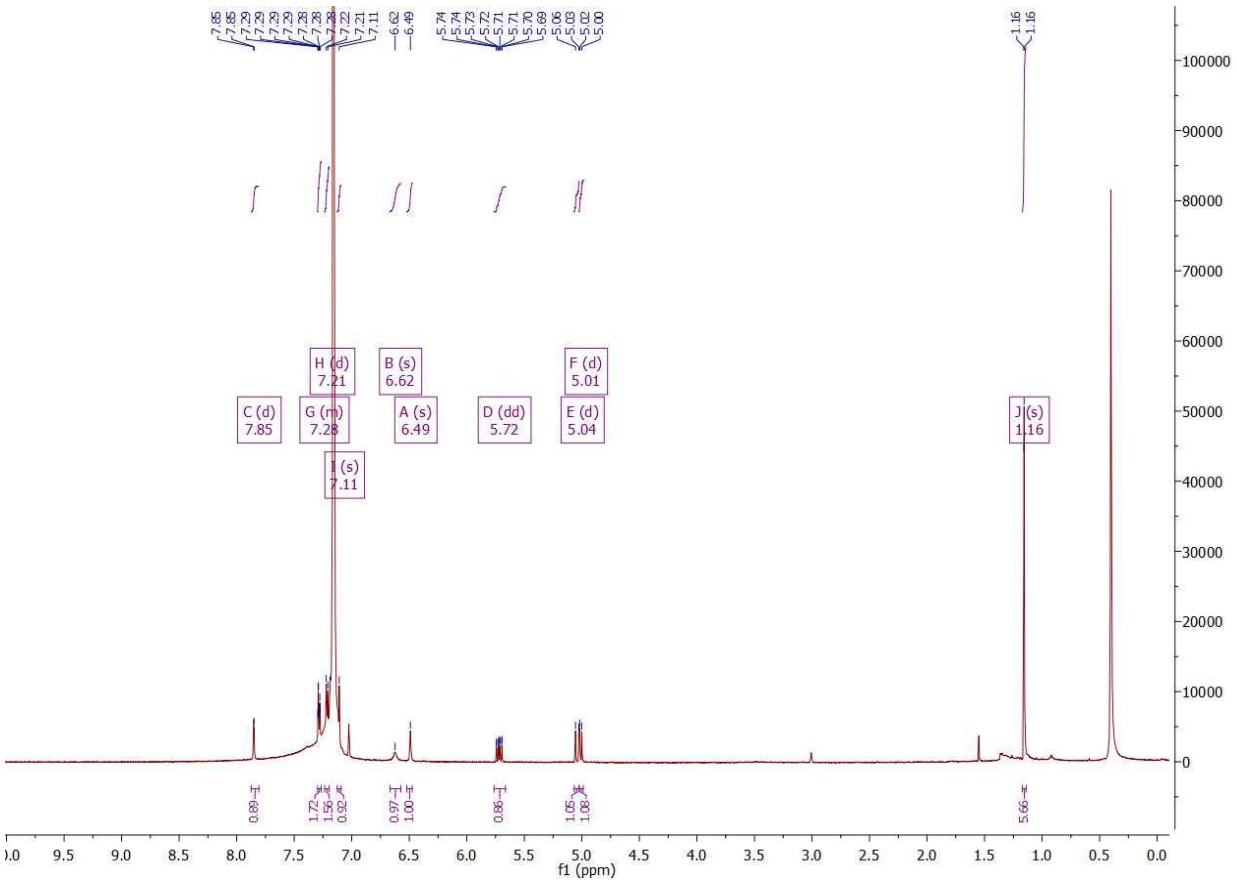

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
10
11
12
f1 (ppm)

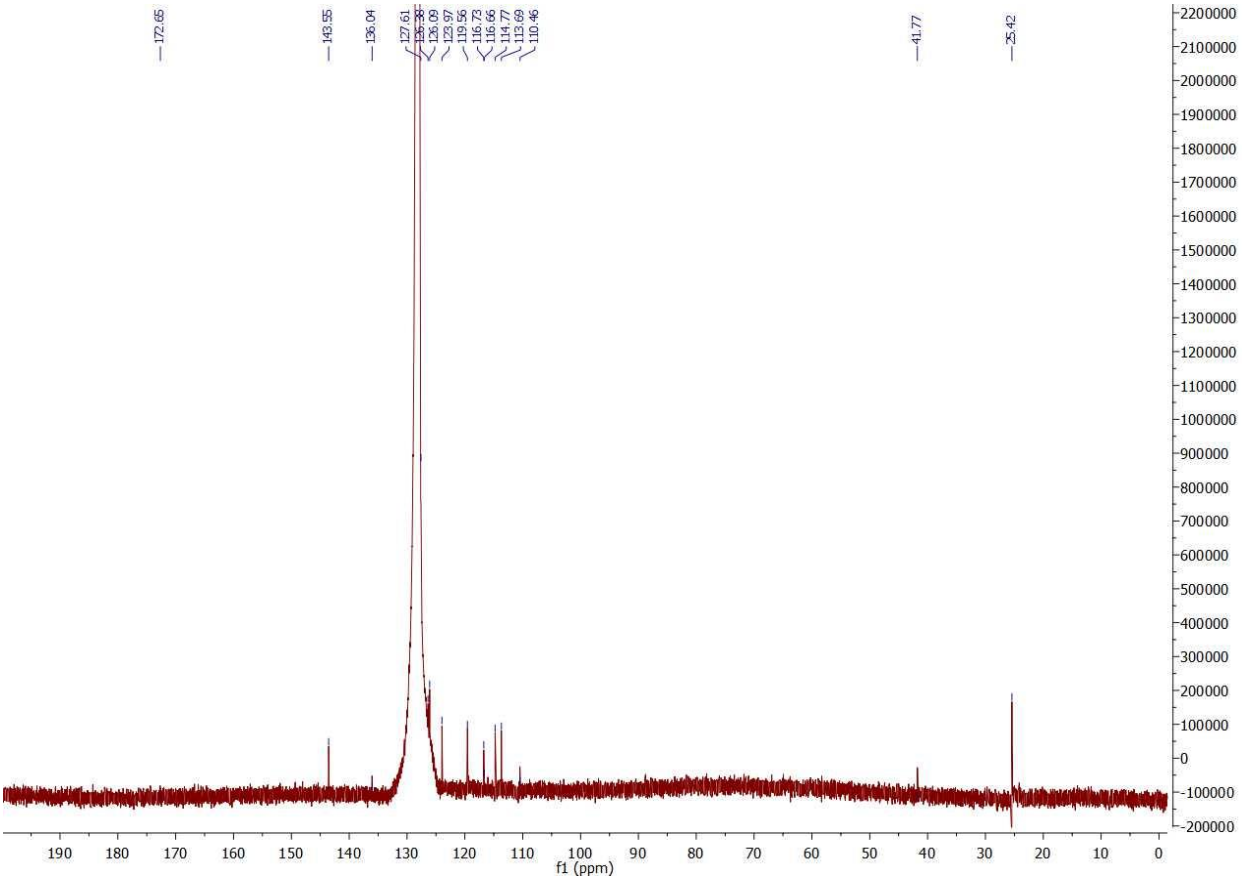

10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
f1 (ppm)

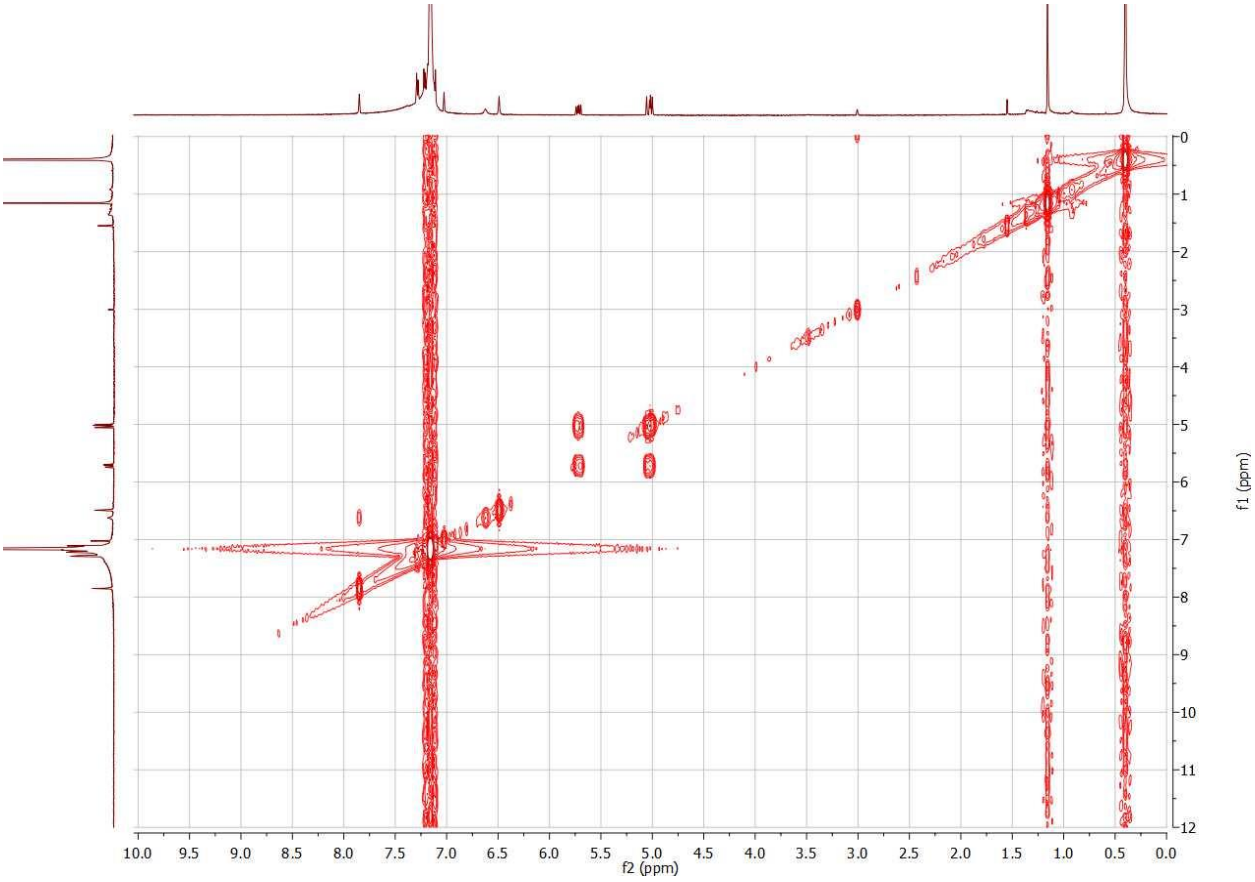


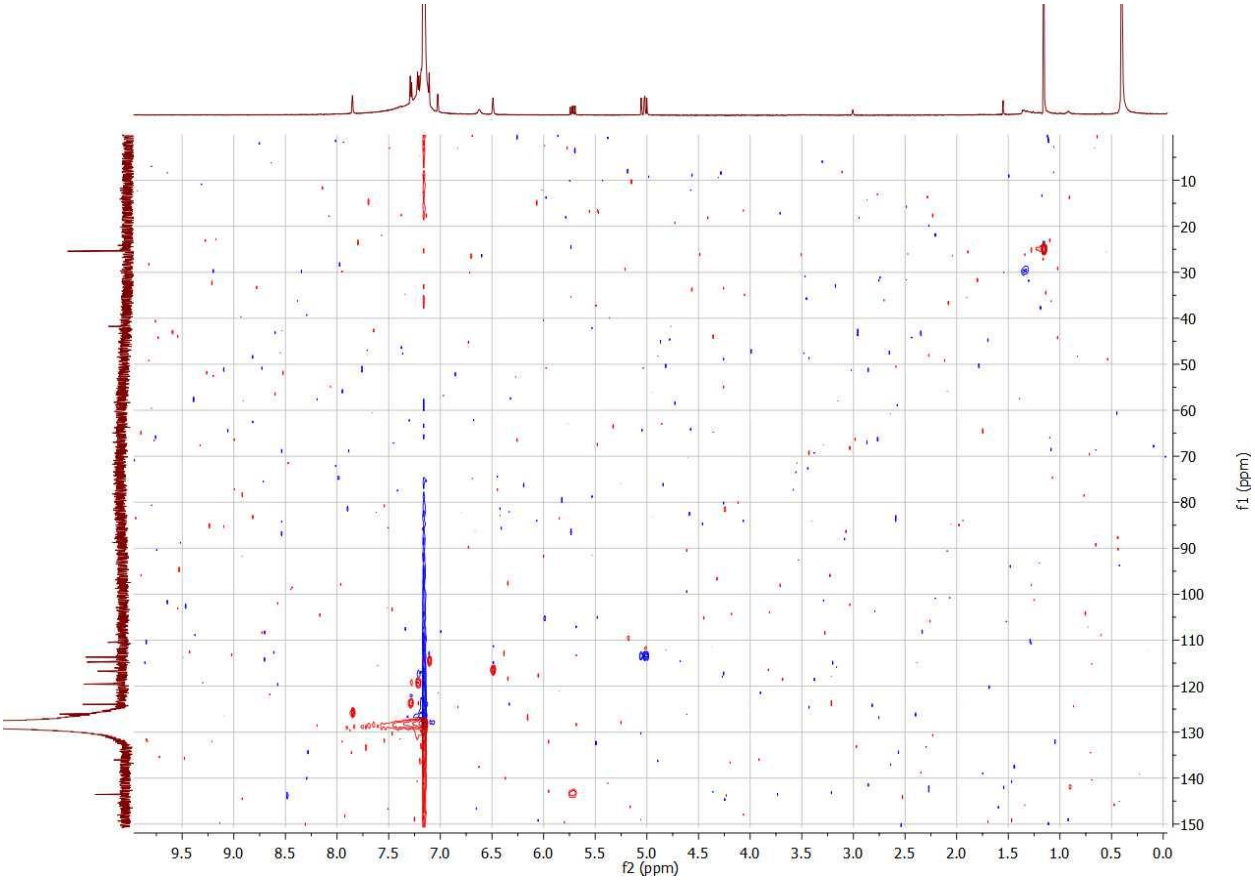
$[M+H]^+$ Calc: 271.0997 Obsv: 271.0995

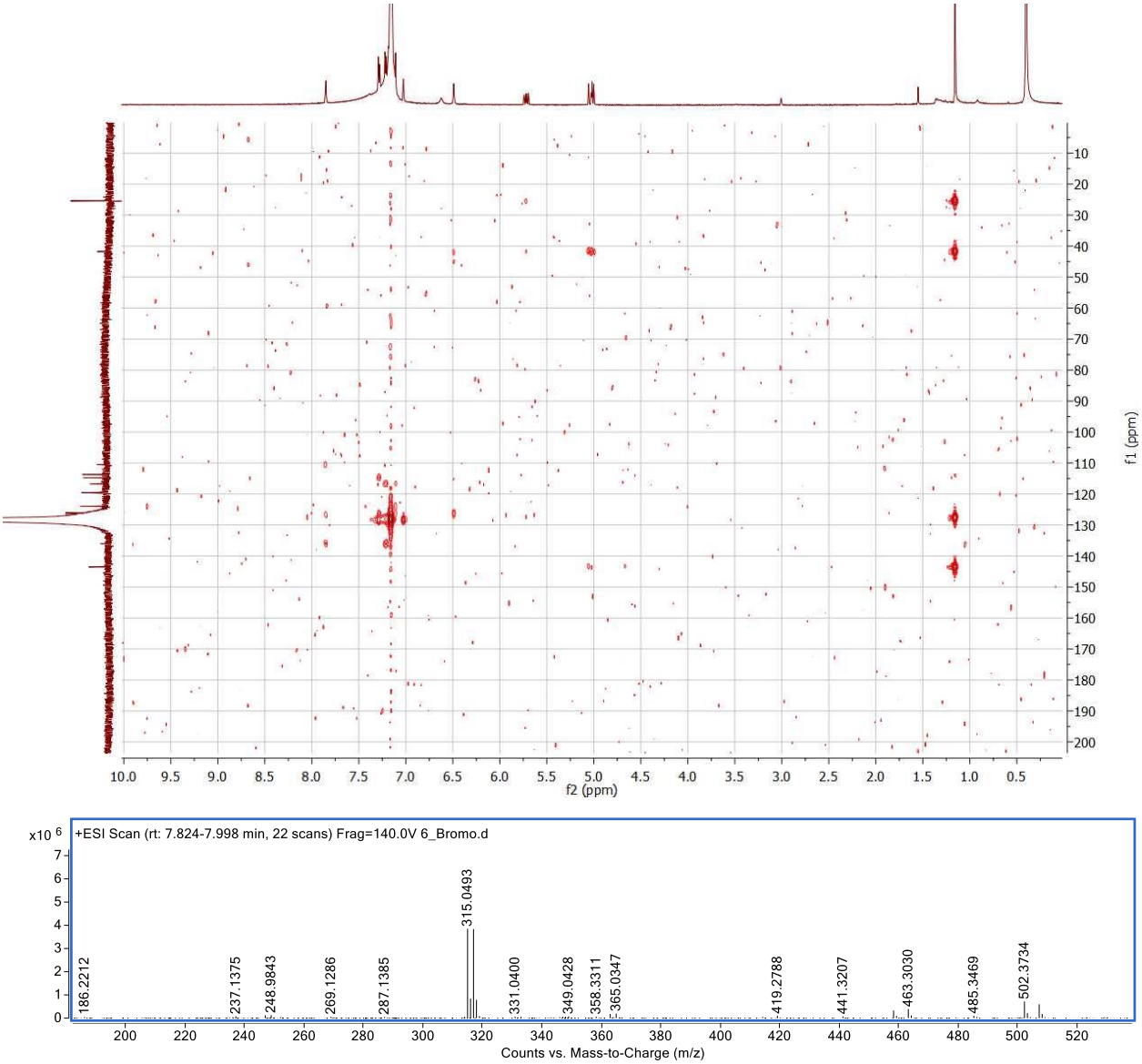
Figure S5.19: ^{1}H, ^{13}C, COSY, HSQC, HMBC NMR spectra and HRMS of 6-bromo 11-DMAC isonitrile (**S5.2**) in C_6D_6 at 600 MHZ and 151 MHZ respectively.


H (d)
7.21
B (s)
6.62
F (d)
5.01
C (d)
7.85
G (m)
7.28
A (s)
6.49
D (dd)
5.72
E (d)
5.04
J (s)
1.16
I (s)
7.11
f1 (ppm)


172.65
143.55
136.04
127.61
126.09
123.97
119.56
116.73
116.66
114.77
113.69
110.46
41.77
25.42
f1 (ppm)


10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
10
11
12
f1 (ppm)


f2 (ppm)
f1 (ppm)



$[M+H]^+$ Calc: 315.0491 Obsv: 315.0493

Chapter 6
Summary and future directions

6.1 Abstract

In this thesis, biocatalytic diversification and elucidation projects relating to the hapalindole-type metabolites have been presented. These works, presented in chapters 2-5, showcased the interactions between organic chemistry, biocatalysis, chemical biology and computational chemistry. In this concluding chapter, I will briefly summarize each chapter before looking at future directions these projects can go. Preliminary data relating to some of these future projects will be presented as well.

6.2 General summaries of chapters 2-5

6.2.1 Chapter 2

In chapter 2, experiments about using Cell-Free Protein Synthesis (CFPS) assays for the production of hapalindole-type metabolites in a multi-enzymatic system were presented. CFPS assays provide a faster route to elucidating protein functions when compared to traditional expression methods. However, there had never been a proof-of-concept showing that the assay could work in a multi-enzymatic system. In that work, it was shown that FamD2, FamC1 and FisC could all be expressed in these CFPS assays. FamD2 and the cyclases (FamC1 or FisC) could be expressed together which was shown through proteomics analysis and production of native hapalindole compounds. The assay components were then screened on six unnatural *cis*-indole isonitrile derivatives with production of fluorinated hapalindole and fischerindole compounds observed. Thus, it was shown that not only could CFPS assays be applied to multi-enzymatic systems for production of their native compounds but also for the production of unnatural derivatives (Figure 6.1).[1]

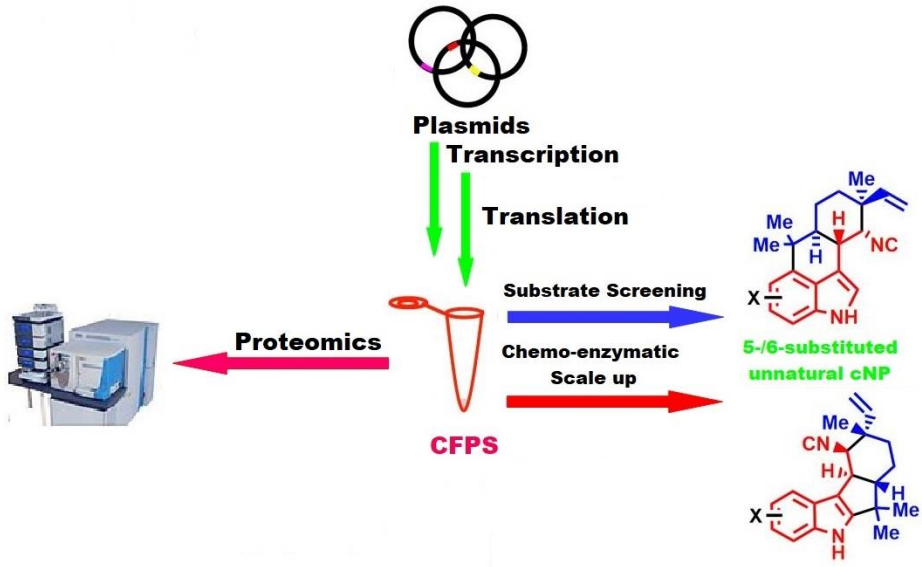


Figure 6.1: General summary of chapter 2. Highlights CFPS methods for production of native and unnatural hapalindole-type metabolites. Note: Figure is an adaption of figure used by Khatri et al.[1]

6.2.2 Chapter 3

Building off of chapter 2, chapter 3 explored unnatural diversification of hapalindole and fischerindole metabolites further. Hapalindole-type metabolites possess a wide variety of biological activities but they are difficult to diversify and produce in large enough quantities. Utilizing the biosynthetic machinery that produces these compounds in cyanos, a suite of unnatural hapalindole and fischerindole compounds were produced using FamD2, FamC1, HpiC1 and FimC5. Further protein engineering work was able to showcase differences between cyclases in substrate scope and regioselectivity. Finally, computational chemistry explored the reaction barriers in select substrates that the cyclases could and could not overcome.[2] These results highlight the interface between organic chemistry and biocatalysis in the potential for producing derivatives of already known natural products (Figure 6.2).

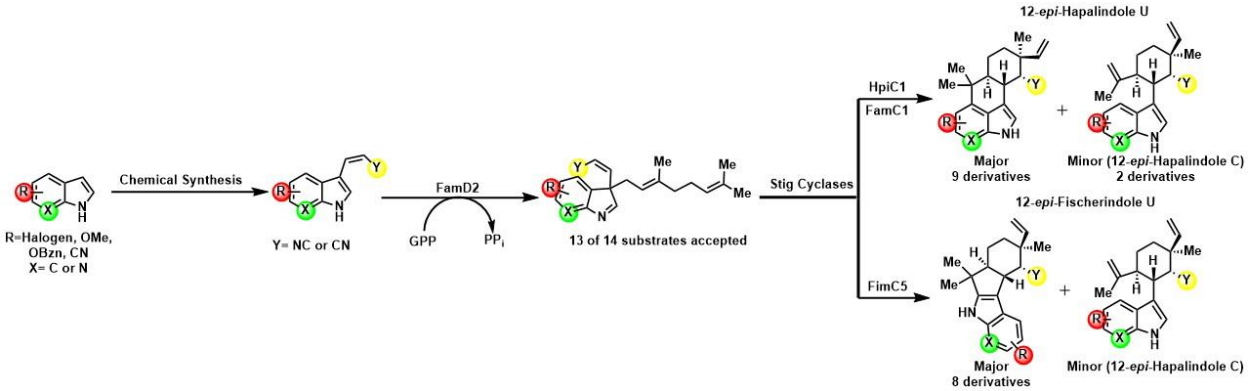


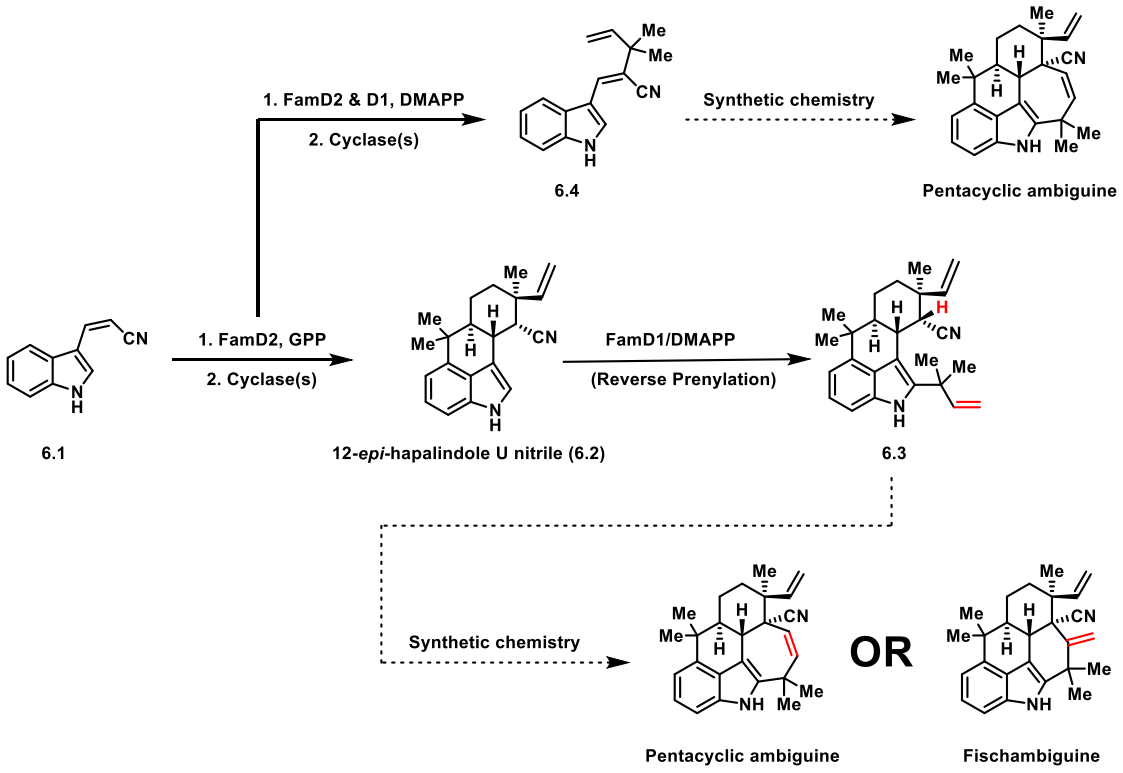
Figure 6.2: General summary of chapter 3. Highlights use of Fam prenyltransferase and Stig cyclases to produce unnatural 12-*epi*-hapalindole U, 12-*epi*-hapalindole C and 12-*epi*-fischerindole U derivatives.

6.2.3 Chapter 4

In chapter 4, Rieske-type oxygenases were explored to elucidate their function for the production of the ambiguine E-ring. The formation of the E-ring in the ambiguines remains one of the largest unsolved questions relating to the hapalindole-type metabolites. The biosynthetic gene clusters (BGCs) of ambiguine producing cyanos points to annotated iron-sulfur Rieske proteins as the most likely to catalyze this carbocyclization reaction. Even though numerous experiments under different conditions and different starting materials were conducted, the function of any of these proteins could not be elucidated. This is a question still worth looking at but may require some outside of the box thinking to succeed.

6.2.4 Chapter 5

Chapter 5 continued to look at ambiguine E-ring formation but through a chemoenzymatic approach. Early stage enzymatic work would be used to produce a tetracyclic ambiguine followed by a late stage C-H functionalization/activation reaction to complete the ambiguine E-ring. During this work, an efficient enzymatic route was developed to produce 84 milligrams of an unnatural tetracyclic ambiguine derivative, 12-*epi*-ambiguine H nitrile (**6.3**) (Scheme 6.1). Synthetic methods are still being explored to complete the E-ring synthesis. Serendipitously, during the optimization work, a novel metabolite (11-DMAC nitrile) (**6.4**) was uncovered. This compound further strengthened the cyclase catalyzed Cope rearrangement hypothesis for the formation of hapalindole-type metabolites. Initial enzymatic derivatization efforts for this compound were moderately successful as its potential uses were still being explored. At the same time, a total synthesis effort of an indole C-2 derivative is was undertaken in order to supplement the enzymatic production. Ultimately, this compound appears poised to be used in the production of a pentacyclic ambiguine (Scheme 6.1).



Scheme 6.1: General summary of chapter 5. Highlights one pot enzymatic optimization route for the production of **6.3** and future synthetic chemistry efforts to produce a pentacyclic ambiguine or fischambiguine. 11-DMAC nitrile (**6.4**) discovery and future uses are also highlighted. Bonds involved in the cyclization reaction are highlighted in red.

6.3 Future experiments and initial results

6.3.1 Other characterized cyclases/cyclase combinations for unnatural production

While initial exploration with FamC1, HpiC1, FimC5 and certain engineered cyclases for production of unnatural hapalindole and fischerindole compounds was a promising start, there are numerous other characterized cyclase combinations to be explored.[2-4] These include both homoligomeric and heteroligomeric combinations. Even cyclases that produce the same compounds are worth exploring i.e.: HpiC1's scope and conversion rates were greater than FamC1's even though they both produce 12-*epi*-hapalindole U natively.[2] These could potentially be advanced even further with the characterized halogenases in the hapalindole biosynthetic gene clusters.[5-6]

To explore this possibility, the *cis*-indole isonitrile (& nitrile) library were screened with FamC1-C4 and FamC2-C3. With FamC1-C4 combination, production of the hypothesized hapalindole U derivative was not seen in many derivatives. Only the *cis*-indole nitrile (**6.1**), 5-&-6 methoxy isonitrile (**6.5-6.6**) and 5-&-6 fluorinated isonitrile (**6.7-6.8**) showed production of what appeared to a hapalindole U derivative (Figure 6.3). Scaling up for characterization,

reactions with **6.1** were fully characterized to be hapalindole U nitrile (**6.9**) (Supplemental information Table S6.2) and a proton NMR was obtained for **6.8** which appeared to show the formation of a hapalindole U derivative but further characterization is required. FamC2-C3 gave a similar result but with a suite of different derivatives for hapalindole H formation. **6.1** and **6.5** appeared to show production of a hapalindole H derivative along with 5-obenzyl isonitrile (**6.10**), 5-chloro isonitrile (**6.11**), 6-bromo isonitrile (**6.12**) and 7-azaindole isonitrile (**6.13**) (Figure 6.3). To date, none of those derivatives have been scaled up for characterization, leaving a large area of unexplored work to cover.

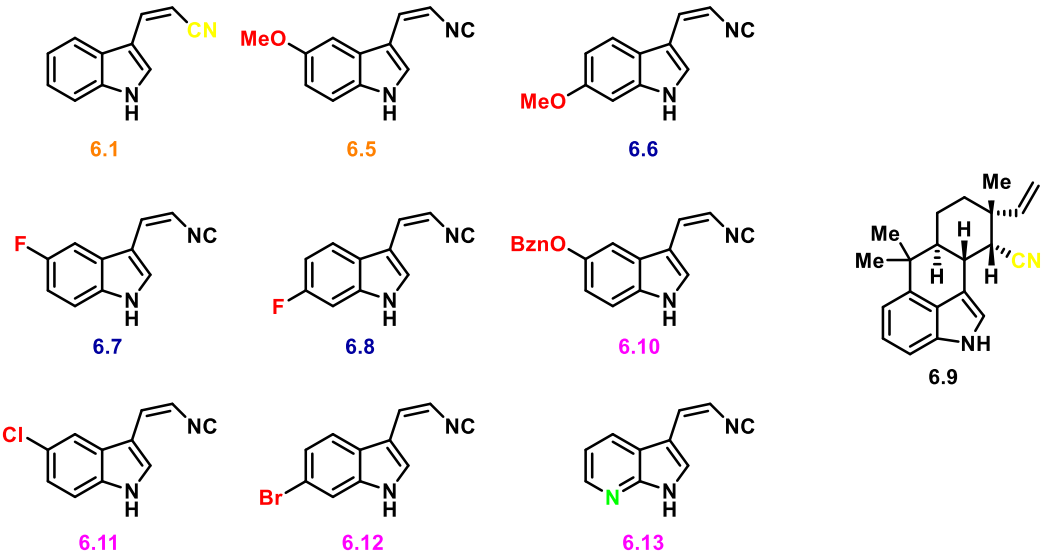


Figure 6.3: Early FamC1/C4 and FamC2/C3 screen results. Structural differences are highlighted in yellow, red and green. Compound numbers in orange are compounds that showed some production of hapalindole compounds with both enzymatic screens. Blue indicates production from only FamC1/C4. Magenta indicates production from only FamC2/C3. Hapalindole U nitrile (**6.9**), on the right, is the only unnatural compound to be characterized so far.

6.3.2 Biological testing of hapalindole/fischerindole library

To date, a majority of the new hapalindole and fischerindole library has not undergone extensive biological screening. The hapalindole-type metabolites possess a wide range of biological activites[7] so it would be reasonable to assume that one of the new derivatives would be bioactive against some/many targets. Large antimicrobial and antifungal screens would be the first place to begin before moving on to antitumoral screens.

For some initial work on this matter, all of the compounds in the library[2] were screened in an AlphaLISA HIV-1 protease assay developed by Chaoping Chen's group.[8] Unfortunately;

none of the compounds in the library came back active against this target. Due to RNA polymerase being a known target of the hapalindole-type metabolites[9], select compounds (12-*epi*-hapalindole U (**6.14**), 5-&-6-fluoro-12-*epi*-hapalindole U (**6.15-6.16**), 12-*epi*-fischerindole U (**6.17**), 5-&-6-fluoro-12-*epi*-fischerindole U (**6.18-6.19**), and 7-aza-12-*epi*-hapalindole C (**6.20**))[2] were screened at 10 µM in duplicate against RNA polymerase. While **6.18-6.20** showed some inhibition of RNA polymerase, the percent inhibitions were not close to a known RNA polymerase inhibitor, rifampin, which was screened at 1 µM as a control (Figure 6.4 and Supplemental information Table S6.1). This result was unsurprising as the original reported inhibition value of RNA polymerase by 12-*epi*-hapalindole E (**6.21**) was an IC_{50} value of 500 µM.[9] This result further supports the author's hypothesis that RNA polymerase is not the only target of the hapalindole-type metabolites. Future experiments will need to shed light on this unknown area.

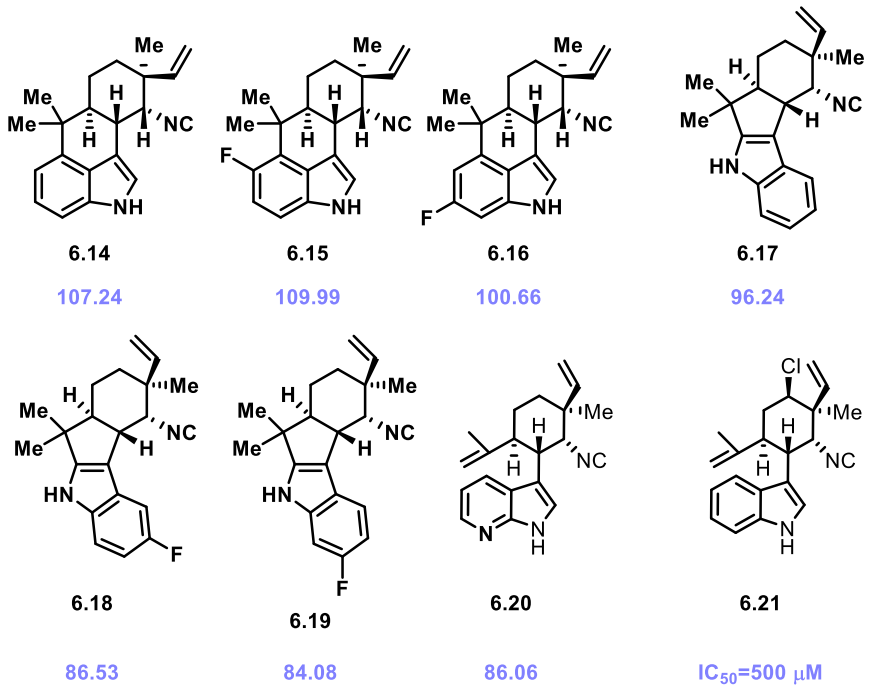


Figure 6.4: Compounds tested in RNA polymerase assay. 12-*epi*-hapalindole E (**6.21**) has been shown as reference for structure and IC_{50} value.[9] Percent inhibition values at 10 µM (n=2) are shown in light blue below compound numbers. Values at or above 100% are assumed no inhibition.

6.3.3 Future of 11-DMAC

As discussed near the end of chapter 5, there is still extensive work to be performed in regards to 11-DMAC and derivatives. With the unnatural screen complete, the remaining hypothesized compounds will need to scaled up for characterization. Synthesis of a C-2

derivative will need to be completed and potentially moved forward in the total synthesis of a pentacyclic ambiguine. The total synthesis may also give us clues into what drives the rearrangement to the C-11 position over the indole C-4 position. This could elicit further computational studies to explain this selectivity preference. Finally, because 11-DMAC is not a native compound to the prenyltransferases or cyclases, protein engineering efforts should be undertaken to engineer the cyclases and prenyltransferases to accept this substrates and increase their turnover. An engineered cyclase with a selectivity preference for the C-3 prenylated intermediate over the C-3 geranylated intermediate would aid in gaining enough material for future efforts regarding this compound.

6.3.4 Planned Rieske experiments

The Rieske proteins continue to be a challenge to elucidate and characterize. For majority of the Rieske proteins already screened, proper levels of expression continue to be a large hindrance in their characterization. However, almost all of the annotated BGCs contain annotated Rieske-type proteins. There are numerous homologs to test and see if better levels of expression can be obtained. There is also another area relating to a ring expansion in select fischerindole compounds (Figure 6.5). These proteins could be another area to tackle to determine which one is responsible for this unique reaction. If heterologous expression and *in vitro* reactions continue to fail, *in vivo* experiments would be another route to try. It has already been shown that hapalindole expression levels can be increased in engineered cyano strains[10] so an engineered cyano strain expressing all the genes from the gene cluster (or just the Rieske proteins) could help elucidate their function.

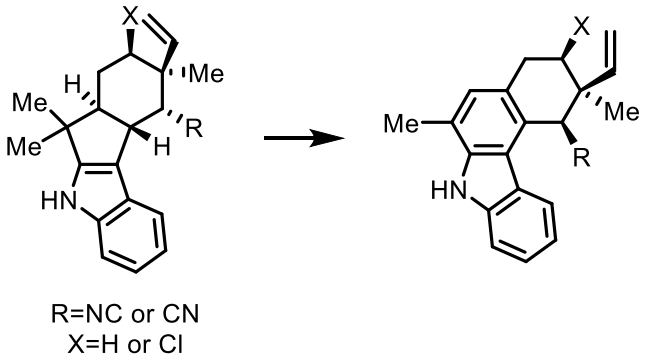


Figure 6.5: Ring expansion seen in select fischerindoles.

6.3.5 Further cyclase exploration and elucidation

The Stig cyclases remain one of the more unique classes of proteins recently uncovered. Based on the large stereochemical differences seen in the hapalindole-type metabolites, it is clear that there are other cyclases needed to be characterized. Proper elucidation and characterization

work could uncover the cyclases, or combination of cyclases, that lead to other stereoisomers. Due to different ratios and different levels of cofactors affecting Stig cyclases activity, all conditions need to be screened with newly expressed proteins. X-ray crystallography work will also be important in determining how certain higher order oligomeric complexes come together to giver different stereo- and regiochemical outcomes. These structures can help guide future experiments in terms of cyclase ratios and cofactor ratios. This elucidation work has not been tackled by anyone in recent years but is still a pivotal part of the story of the hapalindole-type metabolites.

6.4 References

1. Khatri, Y.; Hohlman, R.M.; Mendoza, J.; Li, S.; Lowell, A.N.; Asahara, H.; Sherman. D.H. Multicomponent Microscale Biosynthesis of Unnatural Cyanobacterial Indole Alkaloids. *ACS Synth. Biol.* **2020**, *9* (6), 1359-1360. https://doi.org/10.1021/acssynbio.0c00038.
2. Hohlman, R.M.; Newmister, S.A.; Sanders, J.N.; Khatri, Y.; Li, S.; Keramati, N.R.; Lowell, A.N.; Houk, K.N.; Sherman, D.H. Structural diversification of hapalindole and fischerindole natural products via cascade biocatalysis. *ACS Catal*. **2021**, *11* (8), 4670-4681. https://doi.org/10.1021/acscatal.0c05656.
3. 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. https://www.nature.com/articles/nchembio.2327.
4. Li, S.; Newmister, S.A.; Lowell, A.N.; Zi, J.; Chappell, C.R.; Yu, F.; Hohlman, R.M.; Orjala, J.; Williams, R.M.; Sherman, D.H. Control of Stereoselectivity in Diverse Hapalindole Metabolites is Mediated by Cofactor-Induced Combinatorial Pairing of Stig Cyclase. *Agnew. Chem. Int. Ed.* **2020**, *59* (21), 8166-8172. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201913686.
5. Hillwig, M.L.; Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014**, *10* (11), 921-923. https://www.nature.com/articles/nchembio.1625.
6. 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. *Agnew. Chem. Int. Ed.* **2016**, *55* (19), 5780-5784. https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201601447.
7. Hohlman, R.M.; Sherman. D.H. Recent advances in hapalindole-type cyanobacterial alkaloids: biosynthesis, synthesis and biological activity. *Nat. Prod. Rep.* **2021**, *38*, 1567-1588. https://pubs.rsc.org/en/content/articlelanding/2021/np/d1np00007a.
8. Huang, L.; Li, L.; Tien, C.F.; LaBarbera, D.V.; Chen, C. Targeting HIV-1 Protease Autoprocessing for High-throughput Drug Discovery and Drug Resistance Assessment. *Sci. Rep.* **2019**, *9* (1), 301. https://www.nature.com/articles/s41598-018-36730-4.
9. 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. https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2001.tb10554.x.
10. Knoot, C.J.; Khatri, Y.; Hohlman, R.M.; Sherman, D.H.; Pakrasi, H.B. Engineered Production of Hapalindole Alkaloids in the Cyanobacterium *Synechococcus* sp. UTEX 2973. *ACS Synth. Biol.* **2019**, *8* (8), 1941-1951. https://doi.org/10.1021/acssynbio.9b00229.

6.5 Supplemental information

Table S6.1: RNA polymerase (RNAP) inhibition data. (RMP=rifampin control)

Compound #	Percent Activity	Rel SD	SD
6.14 (1)	107.24	0.09	9.98
6.17 (2)	109.99	0.09	9.97
6.15 (3)	100.66	0.09	9.22
6.16 (4)	96.24	0.10	9.90
6.18 (5)	86.53	0.21	17.91
6.19 (6)	84.08	0.20	16.80
6.20 (7)	86.06	0.09	7.89
DMSO	100.00	0.13	12.81
RMP (1 µM)	11.60	0.15	1.69

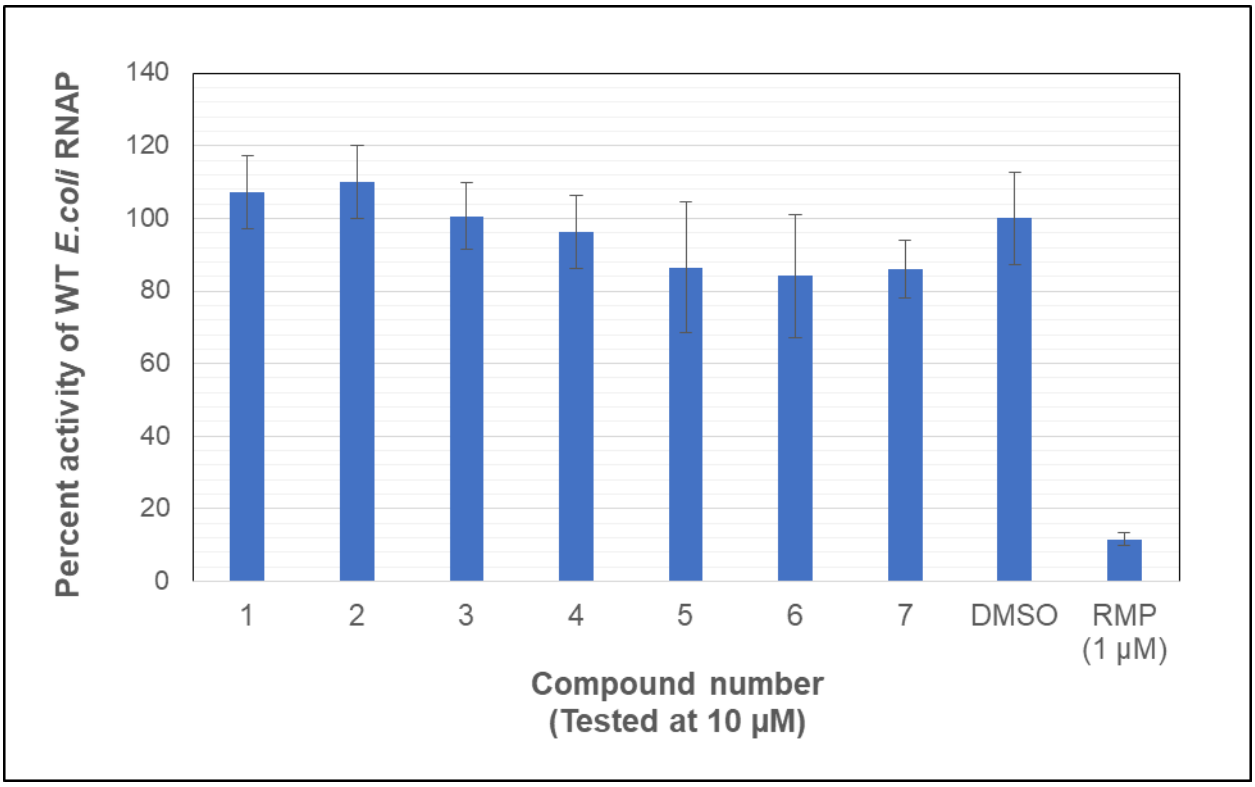
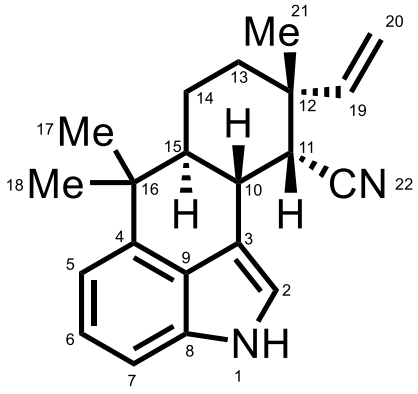
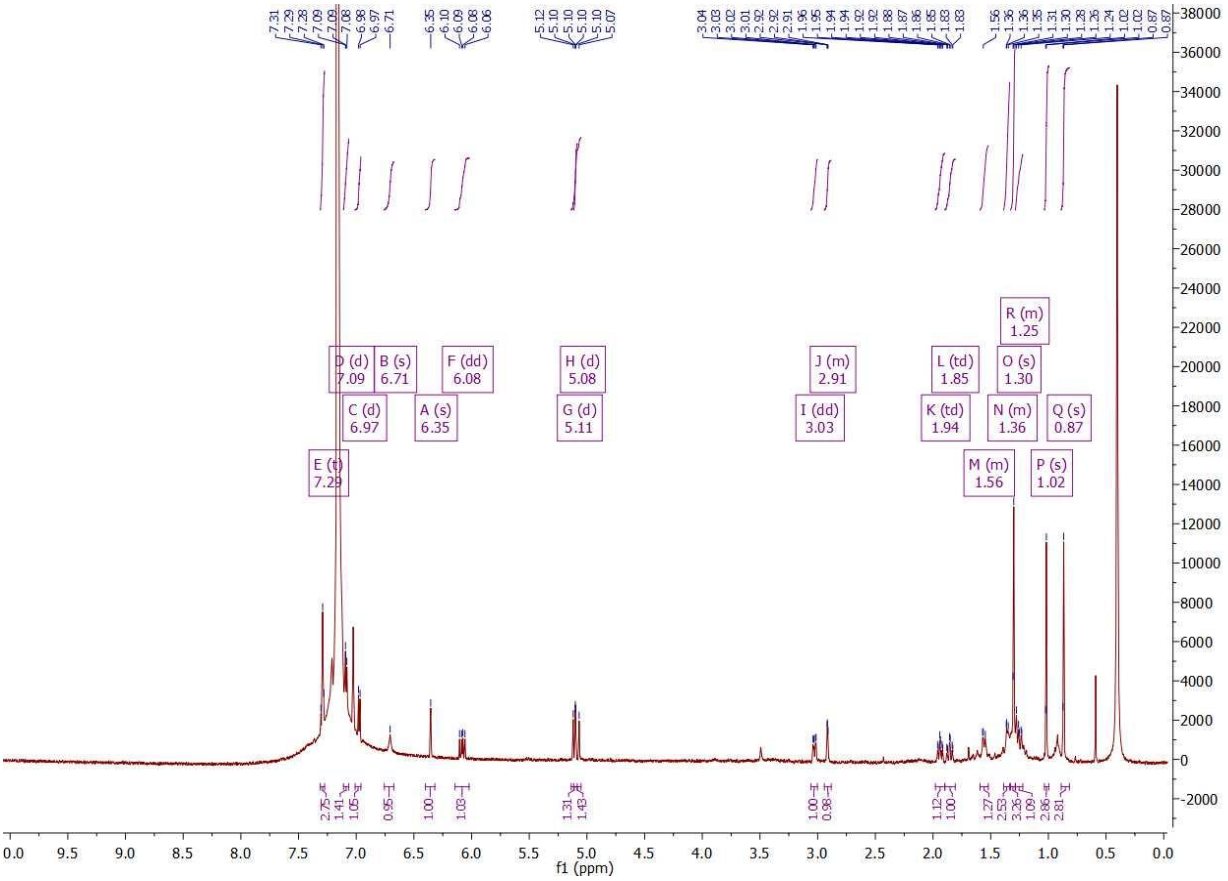


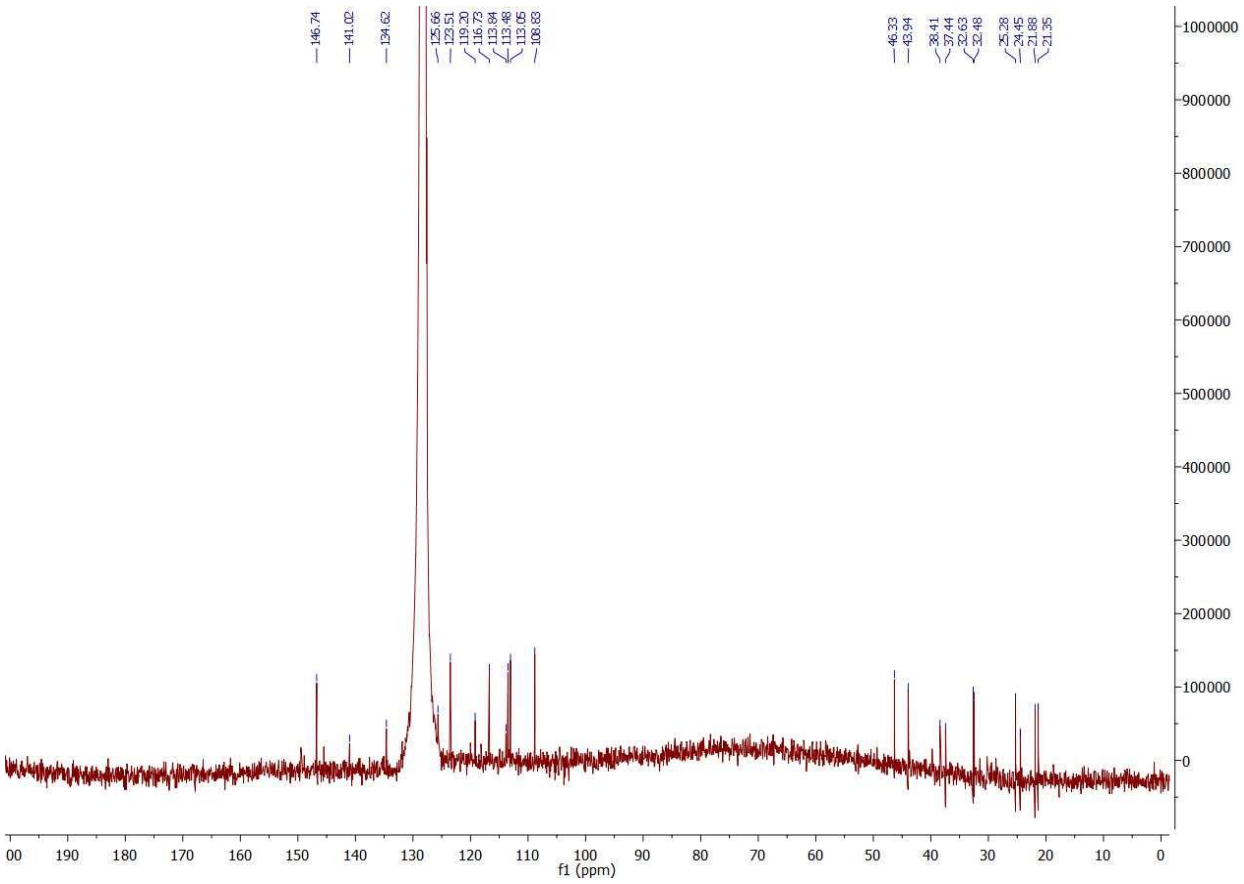
Table S6.2: Hapalindole U nitrile (**6.9**) characterization

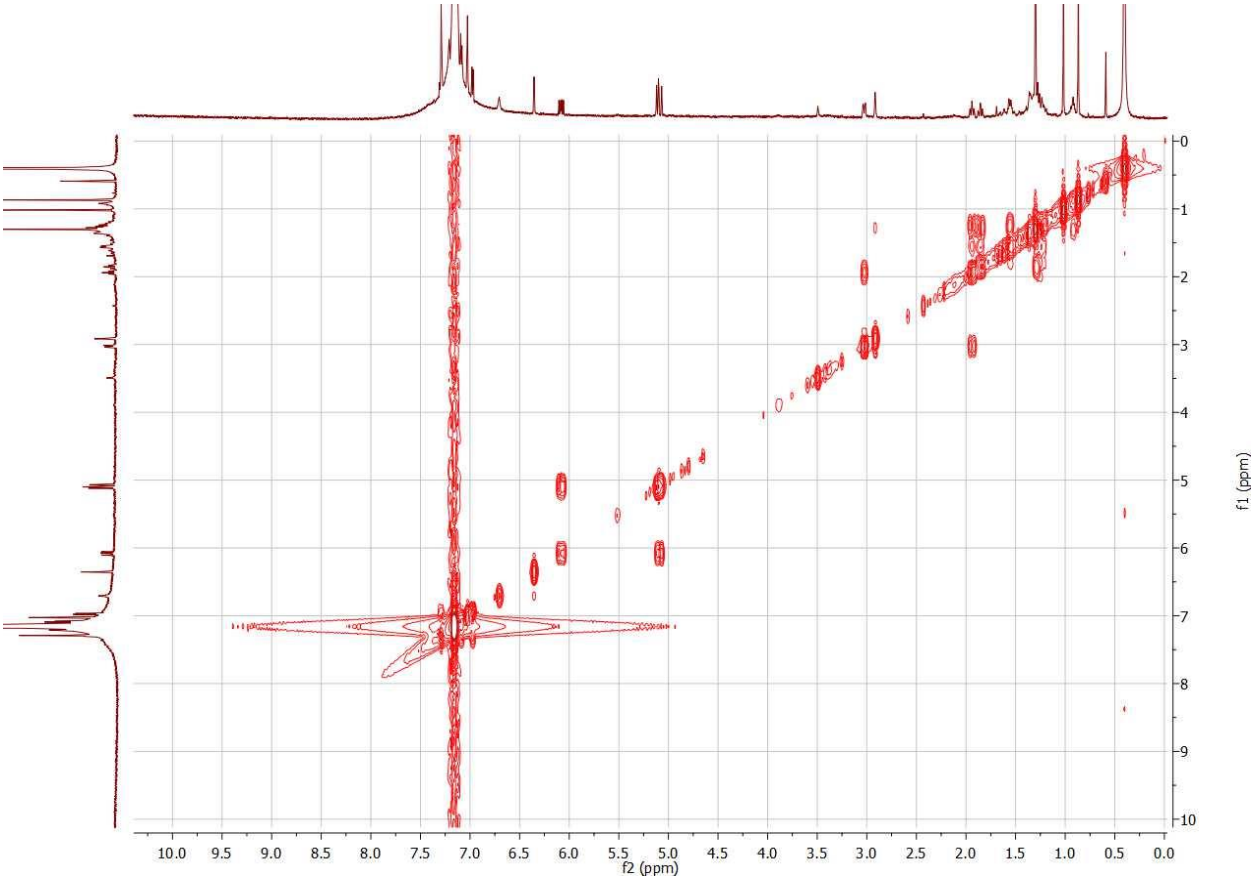


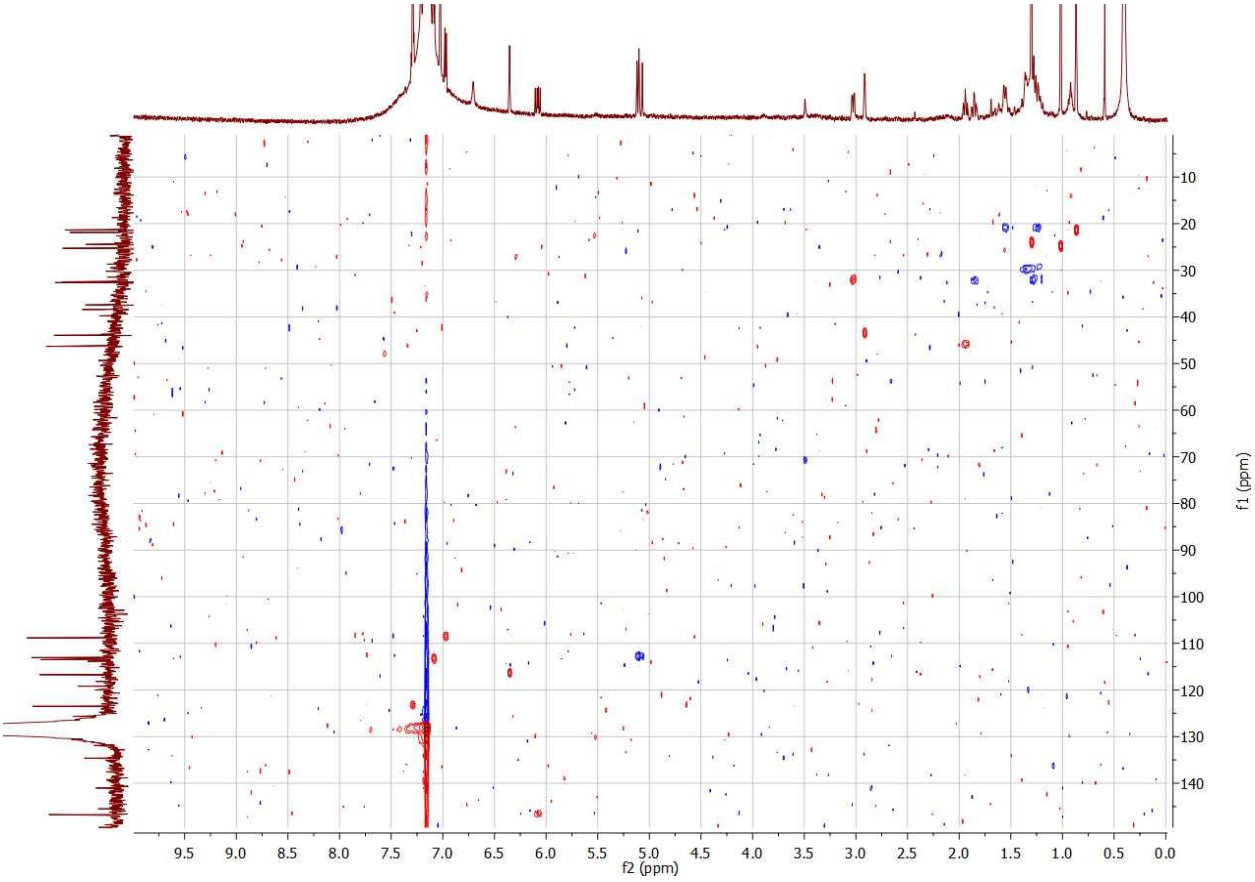
Position	^{13}C shift (ppm)	^{1}H shift (ppm), multi (J)	COSY	HMBC
1		**6.71, bs**	**2**	
2	**116.73**	**6.35, s**	**1**	**3,8,9**
3	**113.84**			
4	**141.02**			
5	**113.48**	**7.09, d, (7.3)**	**6**	**7,9**
6	**123.51**	**7.29, t, (8.1, 6.5)**	**5,7**	**4,8**
7	**108.83**	**6.97, d, (8.1)**	**6**	**5,9**
8	**134.62**			
9	**125.66**			
10	**32.48**	**3.03, dd, (11.6, 4.2)**	**11,15**	
11	**43.94**	**2.91, m**	**10**	
12	**38.41**			
13	**32.63**	**1.85, td, (13.4, 3.5) & 1.25, m**	**14**	
14	**21.35**	**1.56, m & 1.25, m**	**13,15**	
15	**46.33**	**1.94, td, (11.8, 3.5)**	**10,14**	
16	**37.44**			
17	**24.45**	**1.30, s**		**4,15,16,18**
18	**25.28**	**1.02, s**		**4,15,16,17**
19	**146.74**	**6.08, dd, (17.5, 10.8)**	**20**	
20	**113.05**	**Cis 5.11, d, (11.0) Trans 5.08, d, (18.2)**	**19**	**12**
21	**21.87**	**0.87, s**		**11,12,13,19**
22		**119.20**		

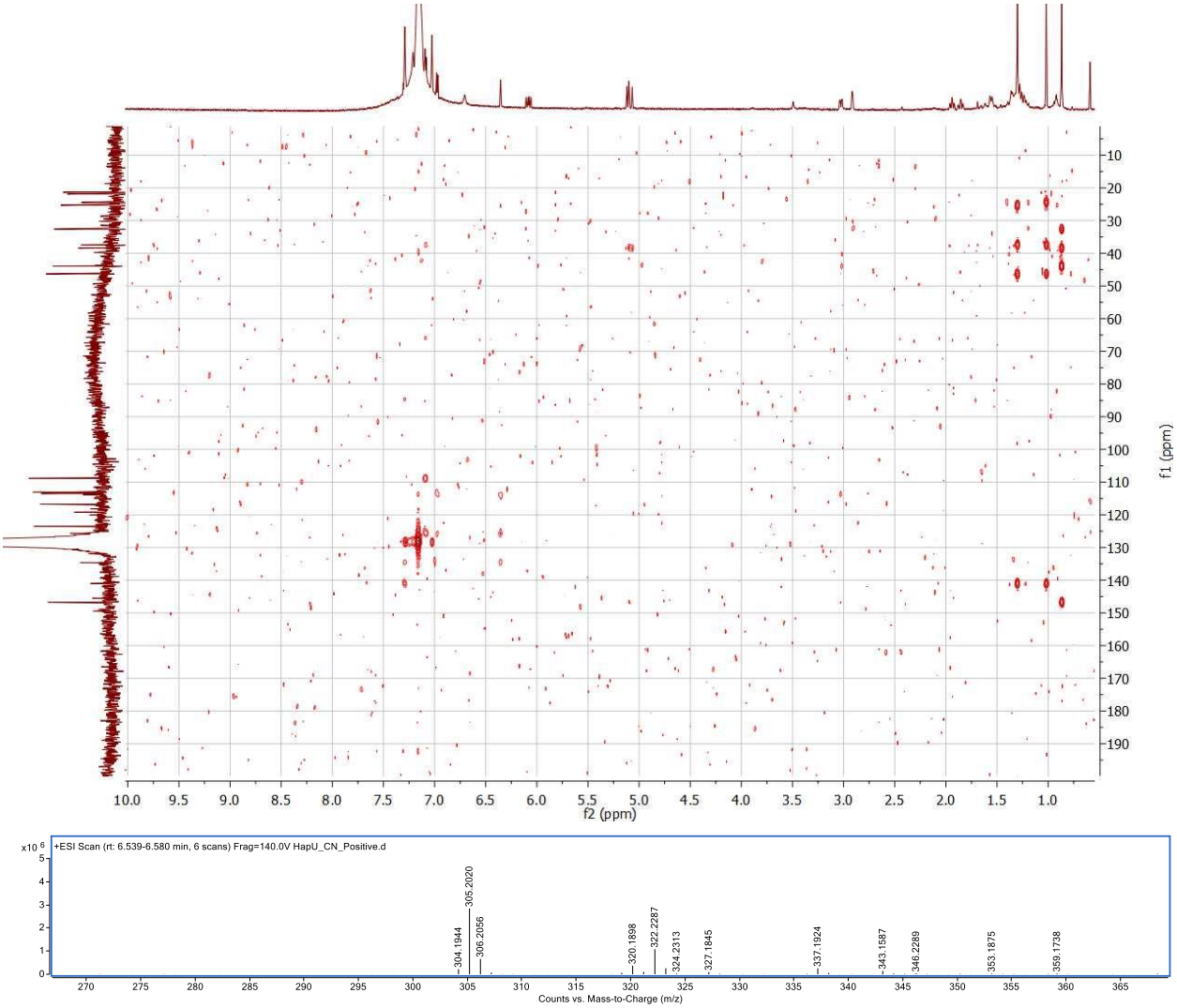
Figure S6.1: ^{1}H, ^{13}C, COSY, HSQC, HMBC NMR spectra and HRMS of hapalindole U nitrile (**6.9**) in C_6D_6 at 600 MHZ and 151 MHZ respectively.




146.74
141.02
134.62
125.66
123.51
119.20
116.73
113.84
113.48
113.05
108.83
46.33
43.94
38.41
37.44
32.63
32.48
25.28
24.45
21.88
21.35
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
00 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)


10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
10
f1 (ppm)


f1 (ppm)
f2 (ppm)



$[M+H]^+$ Calc: 305.2012 Obsv: 305.2020