UCLA
UCLA Electronic Theses and Dissertations

Title

Insights in the Biosynthesis of Griseofulvin and Echinocandin B, two Antifungal Compounds from Ascomycetes

Permalink

Author

Cacho, Ralph Adrian

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Insights in the Biosynthesis of Griseofulvin and Echinocandin B, two Antifungal Compounds

from Ascomycetes

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

Philosophy in Chemical Engineering

by

Ralph Adrian Cacho

2015

ABSTRACT OF DISSERTATION

Insights in the Biosynthesis of Griseofulvin and Echinocandin B, two Antifungal Compounds

from Ascomycetes

by

Ralph Adrian Cacho

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2015

Professor Yi Tang, Chair

Natural products, since the discovery of the first antibiotic penicillin in 1928, have been the source of and inspiration for drugs. Due the emergence of drug-resistant pathogens and discovery of drug targets in cancer biology, the need for the discovery of new bioactive natural compounds and synthesis of analogs thereof remains present. Fortuitously, the development of next-generation sequencing and tools for the heterologous expression of the biosynthetic genes for natural products has accelerated the discovery of new natural products as well as the elucidation of the respective biosynthetic pathways. This dissertation describes the use of these two emerging technologies in revealing insights in the biosynthesis of two antifungal natural compounds synthesized by filamentous fungi: **griseofulvin**, an antifungal drug used in treatment against dermatophytes and is in the World Health Organization List of Essential Medicine and **echinocandin B**, the parent compound of the frontline anti-Candidiasis drug anidulafungin.

This dissertation will first present the work in the investigation the biosynthesis of echinocandin B. Invasive candidiasis caused by opportunistic pathogenic strains of genus

Candida, accounts for 17% of ICU-related infections and through understanding of the biosynthesis of echinocandin B will aid in the mutasynthesis of analogs with better efficacy than the parent compound. In this work, we characterized the key steps in the biosynthesis of the natural product. Deletion of the biosynthetic genes also allowed for the mutasynthesis of echinocandin B analogs with more potent antifungal activity and showed the potential for re-engineering of the pathway to produce "unnatural" natural compounds.

Moreover, this work also describes the full elucidation and reconstitution of the enzymatic reactions leading to the biosynthesis of griseofulvin using a combination of gene knockout and biochemical characterization of the griseofulvin biosynthetic enzymes. Among the important biosynthetic enzymes characterized in this study is the non-reducing polyketide synthase GsfA which was implicated in the biosynthesis of norlichexanthone and the cytochrome P450 GsfF which catalyzes the formation of the spirobicyclic grisan ring in griseofulvin. The culmination of this study is the demonstration of the total *in vitro* biosynthesis of the compound using purified griseofulvin biosynthetic enzymes.

The dissertation of Ralph Adrian Cacho is approved.

Neil K. Garg

Tatiana Segura

Yi Tang, Committee Chair

University of California, Los Angeles

2015

TABLE OF CONTENTS

LIST OF FIGURES FOR SECTIONS 1-5

LIST OF TABLES FOR SECTIONS 1-5

LIST OF SCHEMES FOR SECTIONS 1-5

APPENDIX TABLE OF CONTENTS:

SUPPLEMENTARY INFORMATION FOR SECTION 2

Supplementary Tables

Supplementary Figures 2

SUPPLEMENTARY INFORMATION FOR SECTION 3

Supplementary Tables

Supplementary Figures

SUPPLEMENTARY INFORMATION FOR SECTION 4

Supplementary Tables.

Supplementary Figures

ACKNOWLEDGEMENTS

There a number of people that I am greatly indebted to that have made the work in this dissertation possible. These people have served as my mentors and my inspiration and I am not who I am today without their guidance and mentorship. Foremost among them is my graduate advisor Prof. Yi Tang for being a caring and patient mentor and for giving me this wonderful opportunity to study the biosynthesis of these fascinating natural products. I greatly appreciate his faith in my capabilities as a scientist and his trust in me has continually motivated me throughout my whole career. I would also like to thank Prof. Neil K. Garg, Prof. James C. Liao and Prof. Tatiana Segura for their advice and encouragement. I would also like to thank the NIH Chemistry-Biology Interface Training Grant, the UCLA Graduate Division and the Dissertation Year Fellowship Award for supporting me during my studies.

Secondly, I want to thank the numerous collaborators and colleagues who helped me in my projects and mentored me throughout my academic career. In particular I would like to thank Dr. Yit-Heng Chooi for his wonderful mentorship during the initial stages of my research career and for teaching me the wonders of working with ascomycetes and Prof. Christopher T. Walsh and Dr. Wei Jiang for mentoring me in the field of enzymology and teaching me creative ways to design experiments. I would also like to express my gratitude to Dr. Grace Chiou, Dr. Hui Zhou, Dr. Wei Xu, Prof. Jaclyn Winter, Dr. Hsiao-Ching Lin, Dr. Xue Gao, Dr. Travis McMahon, Jessica Grandner, Prof. Ken Houk, Justin Thuss and Prof. John C. Vederas for their patience and their excellent contributions to my research and career. I would also like to thank my labmates Prof. Wenjun Zhang, Dr. Suzanne Ma, Dr. Lauren Pickens, Prof. Zhen Gu, Dr. Xinkai Xie, Dr. Yanran Li, Dr. Kangjian Qiao, Dr. Anuradha Biswas, Dr. Peng Wang, Dr. Muxun Zhao, Dr. Angelica Zabala, Dr. Li Li, Dr. Wenbing Yin, Dr. Youcai Hu, Dr. Woncheol Kim, Dr. Mancheng Tang, Dr. Yi Zou, Dr. Xuming Mao, Dr. Shushan Gao, Dr. Xia Yu, Dr. Nidhi Tibrewal, Dr.

<div align="center">

VITA

</div>

2010	B.S. Biochemistry University of California Los Angeles Los Angeles, CA
2011-2014	NIH Chemistry-Biology Interface Training Grant University of California Los Angeles Los Angeles, CA
2010-2015	Research Assistant Department of Chemical and Biomolecular Engineering University of California Los Angeles
2013-2014	Teaching Fellow Department of Chemistry and Biochemistry University of California, Los Angeles Los Angeles, CA
2014-2015	Dissertation Year Fellowship University of California Los Angeles Los Angeles, CA

<div align="center">

PUBLICATION AND PRESENTATIONS

</div>

1. Zabala, A. O.*, **Cacho, R.A.***, and Tang, Y. "Protein Engineering towards Natural Product Synthesis and Diversification."*J. Ind.Microb.Biotechnol*. **2012**, 39, 227-241. (Invited Review)
   a. *- Equal contribution*

2. **Cacho, R.A.**, Jiang, W., Chooi, Y-H., Walsh, C.T., and Tang, Y. "Identification and Characterization of the Echinocandin B Biosynthetic Gene Cluster from *Emericella rugulosa* NRRL 11440" *J. Am. Chem. Soc.* **2012** 134, 16781–16790

3. **Cacho, R. A**., and Tang, Y. "Bringing Protein Engineering and Natural Product Biosynthesis Together." *Chem. Biol* .**2013**, 20, 3-5 (Commentary)

4. Jiang, W.*, **Cacho, R. A.***, Chiou, G., Garg. N. K., Tang, Y., and Walsh, C. T., "EcdGHK Are Three Tailoring Iron Oxygenases for Amino Acid Building Blocks of the Echinocandin Scaffold." *J. Am. Chem. Soc.* **2013**, 135, 4457-4466
   - Equal contribution

5. Yin, W. B., Chooi, Y., Smith, A. R., **Cacho, R. A.,** Hu, Y., White, T. C., and Tang, Y. "Discovery of Cryptic Polyketide Metabolites from Dermatophytes using Heterologous Expression in *Aspergillus nidulans.*" *ACS Syn. Biol.* **2013**, 2, 629–634

6. **Cacho, R. A.**, Chooi, Y-H, Zhou, H., and Tang, Y. "Complexity Generation in Fungal Polyketide Biosynthesis: a Spirocycle-Forming P450 in the Concise Pathway to the Antifungal Drug Griseofulvin." *ACS Chem. Biol.* **2013**. 8, 2322–2330

7. Chooi, Y-H., Hong, Y.J., **Cacho, R.A.**, Tantillo, D.J., and Tang, Y. "A Cytochrome P450 Serves as an Unexpected Terpene Cyclase during Fungal Meroterpenoid Biosynthesis*" J. Am. Chem. Soc*. **2013.** 135, 16805–16808

8. **Cacho, R.A.**, Chooi, Y-H. and Tang, Y. "Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi ". *Frontiers in Microbiology.***2015.** 5:774 (Review)

9. Zou, Y., Zhan, Z., Li, D., Tang, M., **Cacho, R.A.**, Watanabe, K., Tang, Y. "Tandem Prenyltransferases Catalyze Isoprenoid Elongation and Complexity Generation in Biosynthesis of Quinolone Alkaloids" *J. Am. Chem. Soc.*.**2015**.(*in press*)

Cacho, R.A., Chooi, Y.H., Zhou, H., and Tang, Y. "Investigation of the early steps in the biosynthesis of Griseofulvin". ACS National Conference, March 27-31, 2011, Anaheim, CA

Cacho, R.A., Chooi, Y.H., Zhou, H., and Tang, Y. "Investigation of the early steps in the biosynthesis of Griseofulvin". Annual UCLA Molecular Biology Research Conference and Retreat. October 21-23, 2011, Lake Arrowhead, CA

Cacho, R.A., Jiang, W., Chooi, Y-H., Walsh, C.T., Tang,Y., "Identification and characterization of the Echinocandin B biosynthetic gene cluster from *Emericella rugulosa* NRRL 11440". Annual UCLA Molecular Biology Research Conference and Retreat. October 19-21, 2012, Lake Arrowhead, CA

Cacho, R.A., Jiang, W., Chiou, G., Chooi, Y.H., Garg, N.K., Walsh, C.T., and Tang, Y. "Genetic and enzymatic characterization of Echinocandin biosynthesis". Annual UCLA Molecular Biology Research Conference and Retreat. January 25-26, 2014, Los Angeles, CA

Cacho, R.A., Jiang, W., Chiou, G., Chooi, Y.H., Garg, N.K., Walsh, C.T., and Tang, Y. "Elucidation of the biosynthesis of the antifungal lipopeptide Echinocandin B". ACS National Conference, March 16, 2014, Dallas, TX

Section 1: Biosynthesis of fungal nonribosomal peptides and aromatic polyketides

Natural product or secondary metabolites are small molecular weight (<10 kDa), and often bioactive, molecules produced by bacteria, plants or fungi that are not part of the primary metabolism of the producer organism.[1] Natural product or secondary metabolites (SMs) have been an important source of drugs or inspiration for molecular scaffolds for drug leads. A recent review of the drugs currently on market estimated that up to 34% of these compounds are either natural products or are derived from natural products.[2] The kingdom fungi is one of the main clades of life that produce bioactive SMs that are crucial to human health.[1] Foremost of the beneficial fungal SMs are the antibiotic penicillin (**1**),[3] the cholesterol-lowering drug lovastatin (**26**),[4] the immunosuppressive drug cyclosporine A (**13**)[5] and the antifungal compounds griseofulvin (**18**)[6] and echinocandin B.[7] Three of the compounds mentioned, **1**, **13** and echinocandin B, belong to a family of compounds known as nonribosomal peptides. On the other hand, **18** and **26** belong to the polyketide class of compounds. This section will discuss these two classes of SMs, highlighting the studies that contribute to the functional elucidation of the cornerstone enzymes in the biosyntheses of these two classes of compounds; the nonribosomal peptide synthetases (NRPSs) and the polyketide synthases (PKSs). Emphasis is also given to the enzymology of non-reducing polyketide synthases (NR-PKSs), the cornerstone enzymes in the biosynthesis of aromatic polyketides such as **18**, the biosynthesis of which is the focus of the study in Section 4.

1.1 Enzymology and Structural Biology of Nonribosomal Peptide Synthetases

One of the keys to understanding how Nature is able to synthesize a diverse array of nonribosomal peptides is through understanding the basic enzymology of nonribosomal peptide synthetases, the proteins involved in the assembly of the monomeric building blocks of the final scaffold.[8-10] The NRPS product, in layman terms, serves as a scaffold upon which tailoring enzymes can introduce further complexities into the structures. Hence, it crucial to elucidate the mechanism of how NRPSs work in hopes of finding ways to manipulate these assembly-line enzymes to synthesized "designer" nonribosomal peptides.
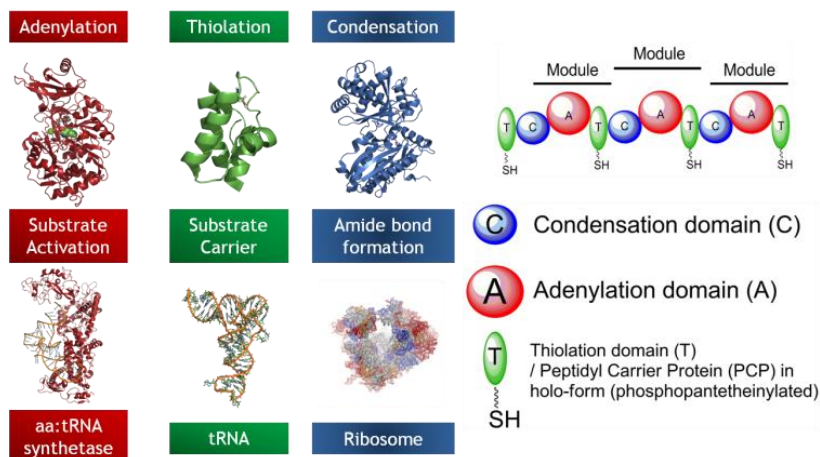
Figure 1.1: Comparison of the functions of nonribosomal peptide synthetase domains with their ribosomal counterparts

Nonribosomal peptide synthetases are multimodular complexes arranged in functional units called modules. Typically, a module will incorporate one monomer- in most instances this monomer is an amino acid- into the scaffold of the product; hence one can envision the length of the NRPS product by knowing the number of modules found in the NRPS. Each module, except for the initiation module, is consists of three domains: the adenylation (A) domain, the thiolation (T) or peptidyl carrier protein (PCP) domain and the condensation (C) domain.[8-10] These three

domains act analogously to the enzymes involved in the ribosome-dependent biosynthesis of proteins. The adenylation domain selects and activates the amino acid building blocks for the NRPS and hence performs an equivalent function of the amino-acyl:tRNA synthetase. The thiolation domain or PCP is analogous in function to the transfer RNA (tRNA); both are involved in acting as substrate carrier. Finally, the condensation domain and the ribosome act in a similar role in nonribosomal and ribosomal peptide synthesis, respectively, where both catalyze the formation of amide bonds between the substrates.[8] Most of the understanding in the enzymology of the adenylation, thiolation and condensation domain are garnered from the structural studies of these domains and will be summarized in the subsequent sections.
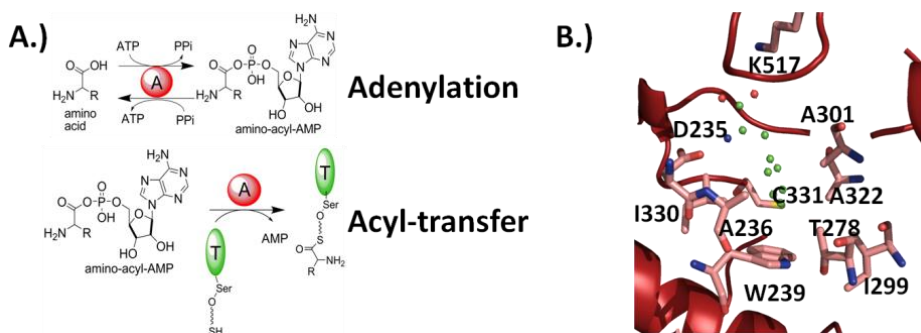
Figure 1.2: A.) Two half reaction of adenylation domain **B.)** The residues in the A domain active site involved in substrate specificity known as the 10 aa code.

1.1.1 Adenylation domain. The NRPS adenylation domain catalyzes the two-step activation of the amino acid substrate via, as the name implies, adenylation of the carboxyl group of the substrate. In the first step of the A domain reaction, the A domain catalyzes the addition of an adenosine monosphosphate group to the amino acid forming an aminoacyl-adenosyl-monophosphate (aminoacyl-AMP). Following formation of the aminoacyl-AMP, the adenylation domain catalyzes the second half of the reaction; the transfer of the aminoacyl group to the phosphantetheine prosthetic group of the thiolation domain. Earlier biochemical studies of the A

domain revealed that A domains have remarkable selectivity towards their cognate substrates and hence acts as a "gatekeeper" that selects the correct substrate for the NRPS module.[11]

The first breakthrough in understanding the mechanism of substrate recognition for the A domains were wrought from the crystal structure of PheA, a phenylalanine-activating A domain form the gramicidin S synthetase NRPS complex (PDB code: **1AMU**).[12] The crystal structure of PheA shows that the enzyme is organized into a larger N-terminal subdomain consisted of a five-layered $\alpha_1\beta_1\alpha_2\beta_2\alpha_3$ sandwich and a smaller C-terminal subdomain. Based on the substrate-bound crystal structure of PheA, the active site of the enzyme is situated between the two subdomains and nestled on top of the middle helix of the N-terminal subdomain.

While all of the residues involved in the interaction with ATP are in the C-terminal subdomain, almost all of these residues involved in interaction with the amino acid substrate, in exception with Lys517, are situated in the $\beta_1\alpha_2$ fold of the N-terminal subdomain. Shown in Figure 1-2B are the eight other residues forming the critical interactions that mediate substrate selection of A domains. Alignment of A domain protein sequences revealed that these eight amino acids are conserved among A domains that utilized the same or very similar amino acid substrates. Hence these residues, along with Asp235 and Lys517, which are currently known as the 10 aa code, were used as predictive tools for A-domain selectivity.[13] Modification of these 10 aa residues led to the engineering of A domain variants with selectivity towards alternative substrates.[14]

1.2.1 Thiolation domain. The thiolation (T) domain or peptidyl carrier protein (PCP) is a small (8-10 kDa) protein that acts as a substrate carrier for the growing peptidyl intermediate of the NRPS.[15, 16] This highly dynamic domain shuttles the phosphopantheteine bound substrate

from the adenylation domain to the condensation domain, forming significant contact with both domains during the catalytic cycle. It is composed of a three- or four-antiparallel helix bundle, depending on the catalytic stage of the NRPS module.[17] Connecting helices 2 and 3 and helices 3 and 4 are loops that are four and two residues in length. The loop connecting helices 1 and 2 (loop 1-2) is significantly longer (18 residues) and forms contacts with all helices except for helix 3. The conserved IGGHSL motif is found at the end of loop 1-2 and the N-terminus of helix 2. After ribosomal synthesis of the NRPS, a phosphopantetheine transferase installs the prosthetic phosphopantetheine (ppant) arm via a phosphate ester linkage with the sidechain of the serine residue in the conserved IGGHSL motif. The phosphopantetheinylated PCP is then competent for receiving and forming a thioesters linkage with the aminoacyl substrate. NMR solution structure of the dissected PCP revealed that helix 3 unravels to form to a loop when the PCP is interacting with a thioesterase domain, a dramatic conformational change described as the H-state by Koglin et al.[16, 18] During the H-state, helix 2 rotates to an almost orthogonal angle in relation to helix 1 causing the ppant arm to move up to 30-35 Å. This "swivel" movement of helix 2 was proposed to be the mechanism by which the Ppant arm delivers the substrates into the active sites of the catalytic domains. [18]
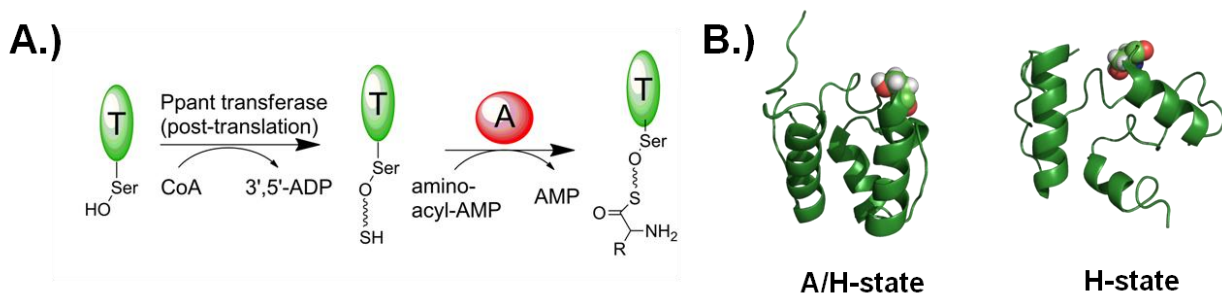
Figure 1.3:A.) Role of thiolation (T) domain **B.)** Conformational change in the T domain from A/H state to H-state

1.1.3 Condensation domain. The NRPS condensation (C) domain, as the name implies, catalyzes the formation of the amide linkages between the donor (upstream)-PCP-bound-aminoacyl thioester acting as a nucleophile and the acceptor-PCP-bound aminoacyl thioesters as the electrophile. Crystal structures of three bacterial C domains, either as standalone protein or as part of multidomain constructs, have shed light into understanding the reaction mechanism of this enzyme. [19-21]

The standalone C domain VibH, involved in the biosynthesis of the siderophore vibriobactin from *Vibrae cholerae*, is the first NRPS C domain that was structurally characterized.[19, 22] The crystal structure of VibH revealed that NRPS C domains are monomeric pseudodimers. The N-terminal and C-terminal subdomains that make up the pseudodimers are each consisted of a distorted αβα sandwich. The interface between the two subdomains is a V-shaped canyon that span ~40 Å, which is approximately the twice the length of a Ppant group.
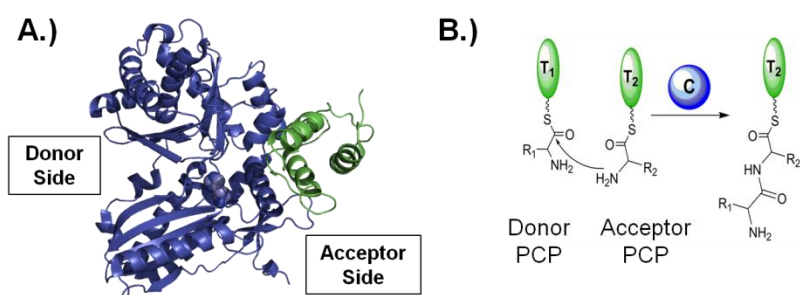
Figure 1.4: A.) Structure of condensation domain from terminal module of surfactin synthetase showing the donor and acceptor sides. The acceptor PCP is also shown in green. **B.)** Reaction catalyzed by the C domain.

Roughly in the middle of the 40 Å solvent channel is the conserved HHXXXDG motif; it was initially proposed that the second histidine of the conserved motif acts as a general base for the deprotonation of the α-ammonium group prior to the nucleophilic attack of the α-nitrogen to

the carboxyl group of the acceptor substrate.[23, 24] However, retention of activity of the alanine mutant of the second histidine of VibH and the evidence that the histidine sidechain is protonated in the structure of the TycC PCP-C didomain structure, suggests that it might instead provide electrostatic stabilization of the putative zwitterionic tetrahedral intermediate of reaction.[19-21]

1.1.4 Tailoring and terminating domains: Other domains typically found in NRPS assembly lines are the epimerase, cyclization, *N*-methyltransferase, thioesterase and reductase domains.[8, 15] The epimerase domain catalyzes the conversion of the PCP-bound l-aminoacyl thioesterase to d-aminoacyl thioesterase via deprotonation-reprotonation of the α-carbon. The cyclization domain, typically found in bacterial NRPSs, catalyzes formation of oxazoline or thiazoline groups from serine and cysteine, respectively. N-methyltransferase domains catalyze addition of a methyl group from *S*-adenosyl-methionine to the amide nitrogen of the PCP-bound substrates. Thioesterase domains catalyze either the cyclization of the NRPS product to form a macrocylic peptide or hydrolysis resulting in a linear peptide. Reductase domains are typically found at the terminal module of NRPS and catalyze reduction of the C-terminus to an aldehyde or alcohol.

1.2 The Myriad Forms of Fungal Nonribosomal Peptides

While the NRPS is the cornerstone enzyme in the synthesis of the NRP, the post-NRPS tailoring reactions introduces a cornucopia of modifications that give rise to a myriad of complex structures ranging from the linear peptaibols to the multicyclic scaffolds of indole-containing NRP-derived alkaloids. This section gives a brief overview of the host of different fungal NRPs and the bioactivity and biosynthesis thereof.

1.2.1 β-Lactams are one of the most beneficial, and also well-known, class of fungal nonribosomal peptides. This class of compounds, such as the quintessential β-lactam drugs

penicillin G (**1**) from *Penicillium chrysogenum* and cephalosporin (**2**) synthesized by several *Acremonium* species, are potent suicide inhibitors of DD-transpeptidase, an enzyme critical in formation of the peptidoglycan linkages in bacterial cell walls.[3, 25] Biosynthesis of both **1** and **2** from fungi begin from the synthesis of the linear tripeptide L-δ-aminoadipyl-L-cysteiny-D-valine (ACV) by the NRPS ACV synthetase.[26-28] Following the annulations reactions of ACV to form the β-lactam and thiazolidine rings of isopenicillin N,[29] the biosynthesis of fungal β-lactams diverges into two different pathways: isopenicillin N can undergo transamidation to form **1** or it can undergo epimerization of amino group in the adipate moiety followed by ring expansion to form **2**.[26, 27, 30]

Scheme 1.1: Examples of fungal nonribosomal peptides

Penicillin G (**1**)

Cephalosporin (**2**)

Gliotoxin (**3**)

(-)-notoamide A (**4**)

(-)-paraherquamide A (**5**)

malbrancheamide (**6**)

acetylaszonalenin (**7**)

penigequinolone (**8**)

ardeemin (**9**)

asperlicin E (**10**)

tryptoquialanine (**11**)

ergotamine (**12**)
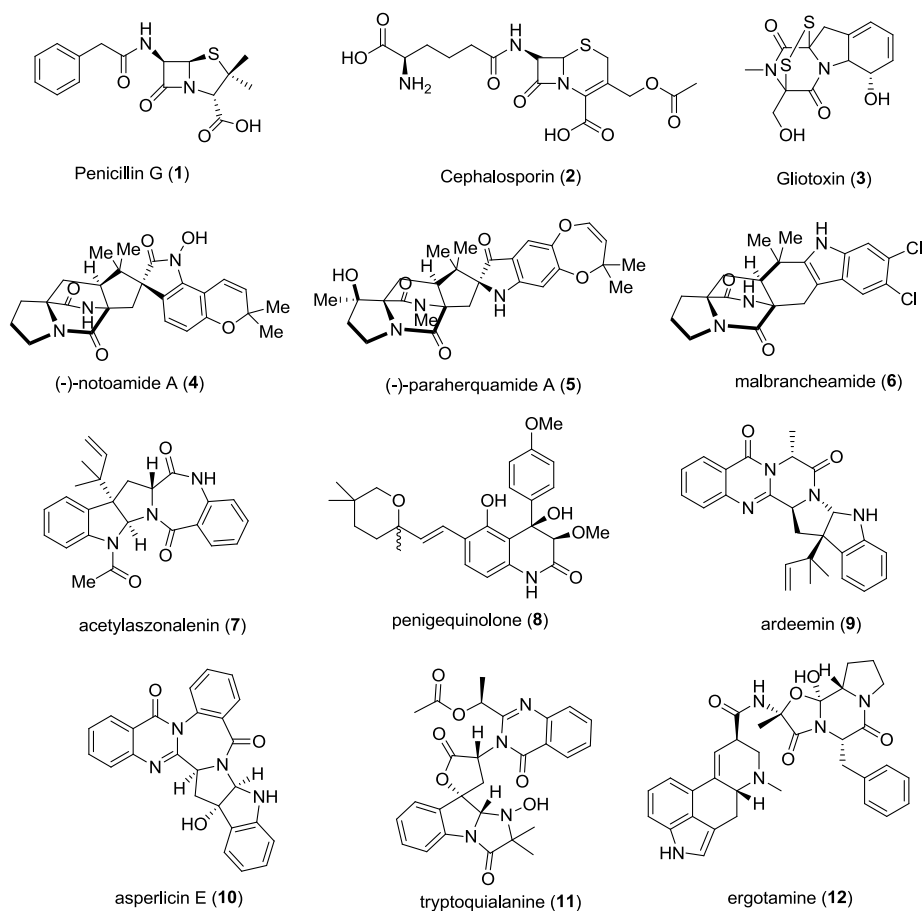
1.2.2 Diketopiperazines (DKPs) are another class of fungal NRPs that can undergo a series of post-NRPS modification to introduce complexity into final structure. This class of compounds is typically made up of two amino acids that annulate to form the core DKP structure. Two subclasses of DKP-derived fungal NRPs will be discussed in this text: the epipolythio-dioxopiperazines (ETP) and the fungal bicyclo[2.2.2]diazaoctane indole alkaloids. Other examples of DKPs not in either categories are the cell-growth inhibitor tryprostatin A and the tryptophanyl-histidine DKPs roquefortine C and meleagrin. A. A more extensive treatment of these compounds as well as other indole-containing alkaloids is discussed in the review by Xu et al.[31]

ETPs are a class of secondary metabolites with a diverse host of biological activity such as an antiviral, antimicrobial and immunosuppressant.[32] The immunosuppressive ability of this class of compounds implicates their role in the pathogenicity of the producer fungi. Much of what is known in the biosynthesis of ETPs is garnered from the studies in the biogenesis of gliotoxin (**3**), the prototypical member of this class of compound. Biochemical characterization of the NRPS GliP confirmed its role in the biosynthesis of **3**.[33] Furthermore, the studies in the recent decade have shed light into the post-NRPS tailoring steps, in particular, the formation of the distinctive transannular disulfide bridge in **3**.[34-36]

Bicyclo[2.2.2]diazaoctane indole alkaloids are another class of fungal DKPs also displaying broad biological activity such as cytotoxic compound (-)-notoamide A(**4**) from *Aspergillus* sp. MF297-2, the anthelmintic paraherquamide (**5**)from *P. fellutanum* ATCC20841 and the calmodulin-inhibitor malbrancheamide (**6**) from *Malbranchea aurantiaca* RRC1813.[37] Common among this type of compound is the reverse-prenylation at C2 of the indole ring.[38] The formation of the bicyclo[2.2.2]diazaoctane ring system was proposed to be through an

intramolecular Diels-Alders reaction.[39] The spiro-ring system in **4** and **6** was proposed to have originated through epoxidation of the C2 and C3 of the indole ring followed by pinacol rearrangement.[37]

1.2.3 Anthranilate-containing Dipeptide and Tripeptide NRPs. Other classes of fungal NRPs are derived from the seven-membered benzodiazepinedione such as acetylaszonalenin (**7**) from *Neosatorya fischeri[40]* and penigequinolone (**8**) from *Penicillium thymicola*.[41] As with the other examples highlighted in this chapter, post-NRPS tailoring enables the emergence of two very diverse scaffolds in the two compounds. In the biosynthesis of **7**, the reverse prenylation of C3 of the pendant indole group promotes attack of the tryptophan α-nitrogen to the incipient imine carbon in the indole ring.[40] Meanwhile **8** is proposed to undergo oxidative ring rearrangement to form the intermediate 14-methoxyviridicatin followed by hydroxylation and prenylation of the anthranilic acid portion of the compound, to afford the final compound.[41]

Other examples of anthranilate-containing NRPs are from the tripeptide-derived metabolites ardeemin (**9**), asperlicin E (**10**) and tryptoquialanine (**11**). Readers are recommended to refer to the reviews by Walsh for a more comprehensive discussion of the biosyntheses of these compounds.[42-44] This chapter will discuss the biosyntheses of the three aforementioned tripeptide-derived NRPs to illustrate the strategies by which Nature utilized a seemingly trivial building block anthranilate to form a myriad of complex multicyclic chemical structures.

The multidrug resistance pump inhibitor **9** is a prenylated indole NRPS from *Aspergillus fischeri*. Post-NRPS modification of post-NRPS intermediate ardeemin FQ is reminiscent with that of **7** wherein the reverse-prenylation of the C3 of the pendant indole group of ardeemin FQ sets up subsequent annulations to afford the hexacyclic **9**.[45] Similarly, a single post-NRPS

enzymatic processing triggers the formation of the heptacyclic ring system of **10** from *A. alliaceus*. Epoxidation of the C2 and C3 of the sidechain indole of the unadorned NRPS product asperlicin C triggers the cyclization cascade to form the heptacyclic **10**.[46]

In comparison to **9** and **10**, **11**, a tremorgenic quinazoline-containing indole alkaloid from *P. aethiopicum*, undergoes a gauntlet of post-NRPS modification to afford the final compound from the initial intermediate fumiquinazoline F.[47] After the formation of fumiquinazoline F, the pendant indole group undergoes epoxidation at C2 and C3 and undergoes addition of α-amino isobutyric acid to form 15-dimethyl-2-*epi*-fumiquinazoline A. Thereafter TqaG and TqaI mediate the oxidative ring-of 15-dimethyl-2-*epi*-fumiquinazoline A and the subsequent rearrangement reaction to form *N*-deoxytryptoquialanone. Followed by N-hydroxylation by TqaE, ketoreduction by TqaC and an acetyltransfer step by TqaD, the final compound is afforded from the pathway.

1.2.4 Ergopeptides are a class of four-amino acid fungal nonribosomal peptides. Due to the similarity of the structures of this family of compounds to the neurotransmitters noradrenaline, dopamine and serotonin, ergopeptides can serve as receptor agonists or antagonists by mimicking the binding of these neurotransmitters to their targets.[48] Ergotamine (**12**) is a quintessential example of this class of compounds and it is made up D-lysergic acid, L-alanine, L-phenylalanine and L-proline. The unusual building block D-lysergic acid, the distinguishing feature of this family of compounds, originates from dimethylallyl-N-L-tryptophan (dimethylallyl-L-abrine) and undergoes a series of oxidative reactions to yield D-lysergic acid.[48, 49] Thereafter, through the collaborative action of the combined NRPS assembly line of LPS B and LPS A2, the DKP-containing intermediate ergotamam is synthesized.[50]

Oxidation of L-alanine portion of ergotamam by EasH triggers the cyclolization of the compound to afford ergotamine.[51]

1.2.5 Peptaibols are antibiotic linear short-chain ($\geq$20 amino acids) peptides with a characteristic alcohol group in the C-terminus arising from the reduction of the carboxyl group.[52] Another distinguishing feature of this class of compounds is the bevy of nonproteinogenic amino acids that are incorporated into the final scaffold such as α-aminoisobutyric acid, isovaline and hydroxyproline. Recently, the gene for the peptaibol synthetase in *Trichoderma virens* Gv29-8, a fungus that synthesizes a variety of 11- and 14-amino acid peptaibols, was genetically characterized.[53] It was demonstrated that the peptaibol synthetase gene *tex2* is responsible for the biosynthesis of all the 11- and 14-amino acid peptaibols from this strain of *T. virens,* a rare example of an NRPS that synthesizes a product with alternative chain length.

1.2.6 Fungal macrocyclic peptides: While there is much debate on what constitutes a macrocyclic peptide, for the purpose of this chapter, it is defined as a peptide consists of four or more monomeric building blocks that form a ring with 12 or more atoms.[54] Typically, the terminal carboxyl group of the peptide forms a covalent bond with a sidechain alcohol or amine or the N-terminus.[55] Fungal macrocyclic peptides fall into two broad varieties, those that solely have amide linkages within the macrocycle and depsipeptides, macrocycles that have one or more ester bond connecting the building blocks of the scaffold. While the low bioavailability of macrocyclic peptides due to their large molecular weight and low calculated lipophilicity (cLogP) has made it difficult to develop this class of compounds as drugs or drug leads, due to the ability of this class of compounds to inhibit non-classical drug targets like protein-protein interaction, medicinal chemists are taking a second look in the untapped potential of these compounds in the past decades. [56]

This discussion of fungal macrocyclic peptides will be remiss if cyclosporine A (**13**), the poster child of the success of this class of compounds as drugs, is not mentioned. **13** is a potent immunosuppressive agent widely used in organ transplant to prevent rejection of the donor tissues via inhibition of T-cell activation.[57] The compound, made by the soil-dwelling fungus *Tolypocladium inflatum*, is a cyclic undecapeptide containing two nonproteinogenic amino acids, D-alanine (aa$_1$) and the highly unusual butenyl-methyl-L-threonine (aa$_4$). Sequencing of the genome of *T. inflatum* and annotation of the putative biosynthetic gene cluster has revealed new insights in the biosynthesis of the compound.[58] With the discovery of the complete gene cluster, the future is bright in elucidation of the biosynthesis of this critical compound and as well as in the engineering of the pathway to create cyclosporine-like analogs with modified or modulated activities.

Scheme 1.2: Examples of fungal macrocyclic peptides

cyclosporine A (**13**)

echinocandin B (**14**)

emericellamide B (**15**)

apicidin (**16**)

AM-toxin I (**17**)
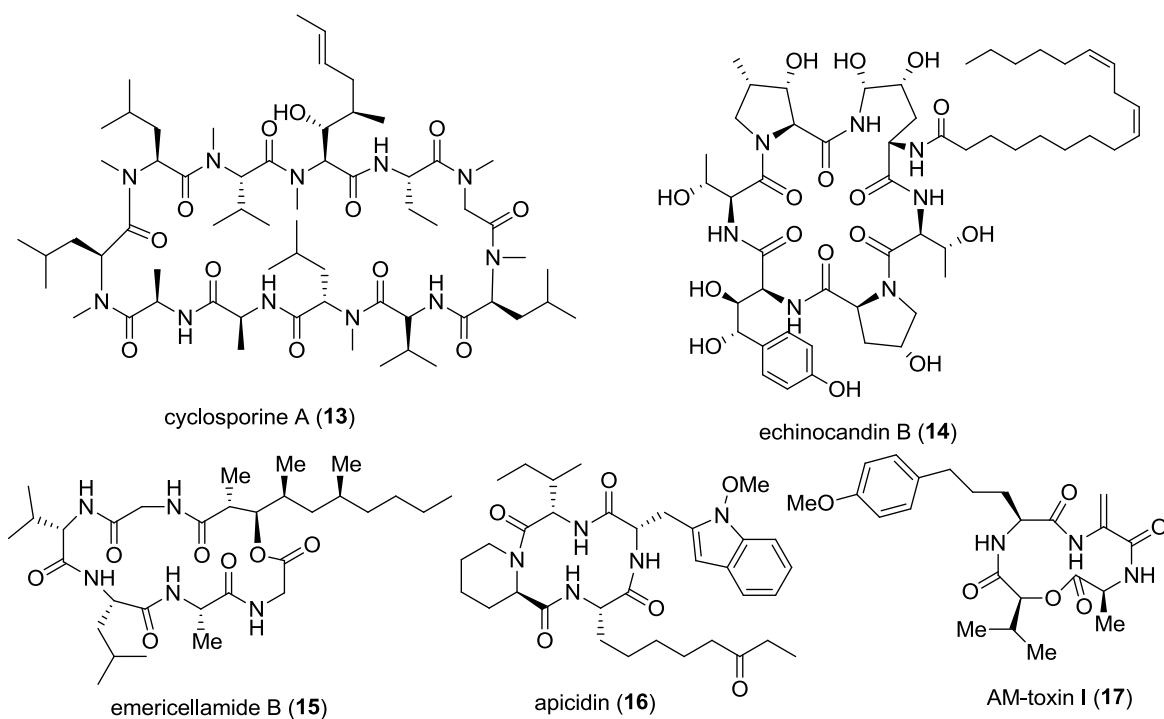
Another class of fungal cyclic macropeptides that were utilized as drug or drug leads are the echinocandins, cyclic peptides that have a distinctive fatty acid chain attached to the macrocycles via an amide bond. A prototypical member of this class of compound is echinocandin B (**14**) from *Emericella rugulosa.* These compounds are known for their potent activity against *Candida albicans* and while the hemolytic activity of natural echinocandins prevented the use of these compounds as antifungal drugs, semisynthetic derivatives of natural echinocandins such as anidulafungin, caspofungin and micafungin have retained the inhibitory activity of their parent compound while minimizing their adverse effect and are currently used as frontline treatment against invasive candidiasis. Further discussion of the biosynthesis of these compounds is presented in chapter 2 and 3 of this manuscript.

Other examples of fungal macrocyclic peptides that received renewed attention through the discovery of their respective biosynthetic genes are emericellamide B (**15**), apicidin (**16**) and AM-toxin I (**17**). The biosynthetic genes antibacterial hexalipodepsipeptide **15** was recently genetically characterized in *A. nidulans* A4.[59] **16**, another example of fungal macrocylic peptide with interesting biological activity, shows broad-spectrum antiprotozoal activity through inhibition of histone deacetylases (HDAC). The biosynthetic gene cluster for **16**, a cyclic tetrapeptide in *Fusarium semitectum* strain KCTC16676 was also discovered fairly recently.[60] Meanwhile AM-toxins are a family of cyclic tetradepsipeptides produced by the plant pathogen *Alternaria alternata,* the causative organism of apple blotch disease.[61, 62] The AM-toxin scaffold also contains a plethora of non-proteinogenic monomers such as variants of L-α-amino-phenyl-valeric acid, dehydroalanine and L-hydroxyisovaleric acid.

As demonstrated by this section, fungal NRPSs come in manifold of sizes and shapes. However, a number of these compounds, in particular those highlighted in this review, play an

important role in biology whether beneficial as drugs and drug leads or detrimental such as mycotoxins and host-specific plant toxins. Understanding the biosyntheses of these compounds serves as the first step in harnessing the endless potential of these compounds for the overall improvement of human condition.

1.3 Fungal Polyketides and Fungal Iterative Polyketide Synthase:
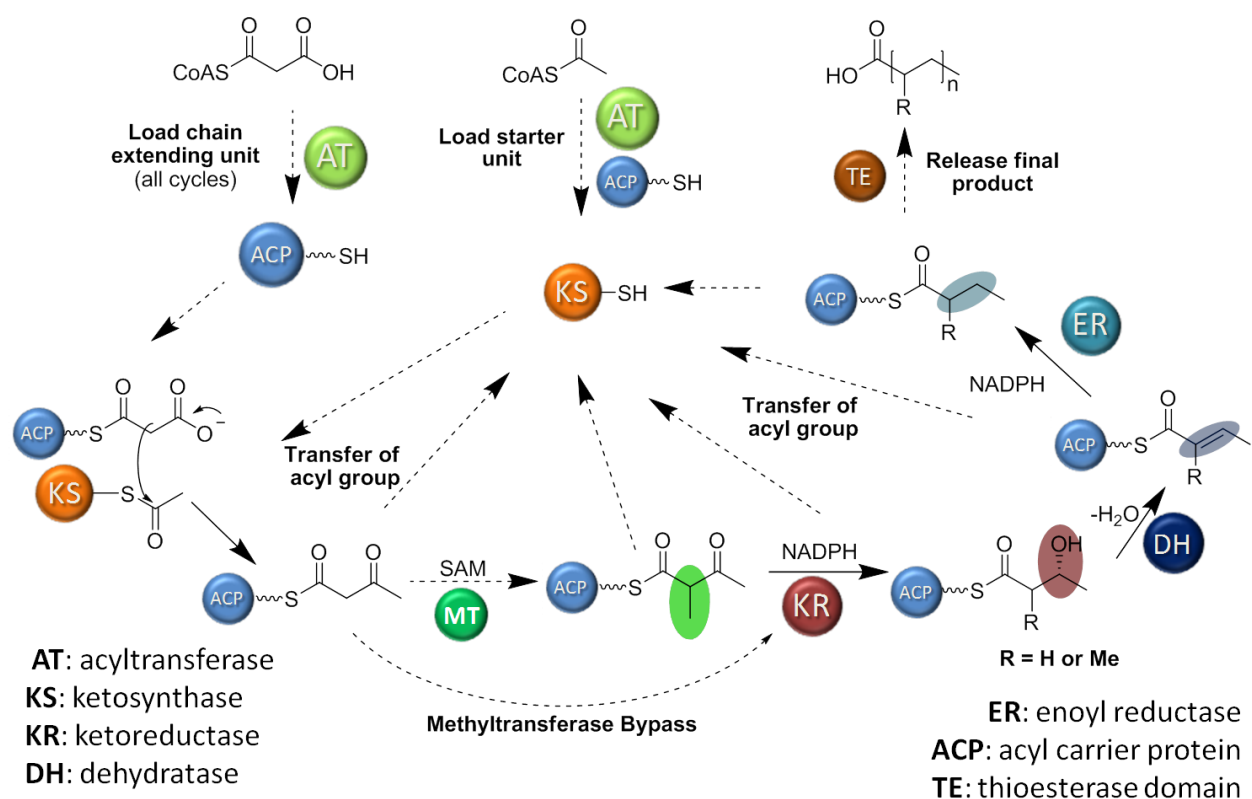
Figure 1.5: Function of the different polyketide synthase and fatty acid synthase domains.

Polyketides are a group of structurally diverse natural products that share a common biosynthetic origin from the polymerization of short chain carboxylic acids-typically acetate and propionate, via decarboxylative Claisen condensation and synthesized by wide variety of organisms of the plants, bacteria and fungi.[63-65] Like nonribosomal peptides, polyketides play

an important role in human lives, not only as source of drug or drug leads but also as toxins and virulence factors for pathogens.

The multidomain enzyme complexes polyketide synthases (PKS), like the nonribosomal peptide synthetases (NRPSs), are the cornerstone protein in the biosynthesis of polyketides. Polyketide synthases function similar to fatty acid synthase, and as such, it is crucial to understand the enzymology of the fatty acid synthase before any further discussions of the polyketide synthases is presented (Figure 1.5).[63, 65-68]

The fatty acid synthase is a multidomain enzyme complex that catalyzes the formation of long-chain fatty acids; most often times palmitic acid (C16 fatty acid), using malonyl-CoA as extender unit and NADPH as reducing equivalent. Fatty acid biosynthesis is initiated by the loading of the starter unit acetate from acetyl-CoA to the phosphopantetheine prosthetic group of acyl-carrier protein (ACP) in a reaction catalyzed by the malonyl/acetyl-CoA:ACP transacylase (MAT) domain. The acyl carrier protein then transfers the Ppant-bound acetate group to the thiol of the active site cysteine of the ketosynthase (KS) domain. Thus, the ACP performs an analogous function to the PCP of the NRPSs.[63, 66]

The extension cycle then begins by the transfer of the malonate group from malonyl-CoA to the free thiol of the Ppant arm of the ACP in a reaction catalyzed once again by the MAT domain.[69, 70] The ACP then delivers the Ppant-bound malonate to the active site of the KS domain where the KS catalyzes the deprotonation of the α-carbon of the malonate group; the presence of the additional carboxylate group in malonate allows the pKa of this carbon to be lower in comparison to acetate, hence the use of malonyl-CoA instead of acetyl-CoA as extender units. After the malonate is deprotonated, the resulting carbanion performs an attack on carboxyl

group of the KS-bound starter unit, followed by the decarboxylation of the addition product to yield the Ppant-bound acetoacetyl thioester.[71]

Thereafter, the ACP delivers the substrate to the active site of the ketoreductase (KR) domain where the β-keto group of the acetoacetate is reduced to an alcohol using NADPH as a formal hydride donor to form D-β-hydroxybutyryl-*S*-ACP thioester.[72] The ACP then transfers the incipient β-hydroxy-acyl-thioester to the dehydratase (DH) domain where a histidine-aspartate/glutamate dyad catalyzes the removal of water from C2 and C3 of the substrate to form *trans*-Δ^2-butenoyl-*S*-ACP.[73, 74] Finally, the double bond of the resulting enoyl-ACP is reduced to form butyryl-*S*-ACP by the action of the enoyl reductase (ER) domain. Once again, NADPH serves as the formal hydride donor in the reaction.[63]

Following production of the butyryl-*S*-ACP, the ACP transfers the full saturated acyl product to the thiol group of the active site cysteine of the ketosynthase domain to begin another round of extension cycle catalyzed by the collaborative function of the MAT, ACP and the KS using another equivalent of malonyl-CoA as an extender unit. Following the Claisen condensation, the reductive steps catalyzed by the KR, DH and ER occurs. These extension-reduction cycles repeat until the fatty acid reaches its final length. Thereafter, the thioesterase (TE) domain catalyzes the hydrolysis of the palmitate from ACP.[63]

One of the important distinctions of polyketide biosynthesis is that one or more of the reductive steps found in fatty acid biosynthesis are skipped giving rise to a manifold of different product. Moreover, the length of the polyketide product can be varied as well, thus introducing a second mechanism for generation of complexity in the product. In addition, unlike the fatty acid synthases, polyketide synthases, especially the fungal type I systems, have an inline

methyltransferase domains that can introduce a methyl group into the α-carbon during the extension cycle.[65, 75]

Polyketide synthases are classified based on their similarity to fatty acid synthases and fall into three types: the type I PKSs which are similar to the mammalian fatty acid synthases wherein the catalytic domains are juxtapose in a single polypeptide; the type II PKSs which share the same architecture as bacterial fatty acid synthases wherein the catalytic domains are dissociated proteins and the type III polyketide synthases which are ACP-less PKSs and are responsible for the synthesis of chalcones and polyphenols in plants and bacteria.[63, 65, 75] The type II and type III will only be briefly mentioned in this chapter and the readers are referred to the reviews that discuss in detail the two PKS systems.[63, 76-78]

The type I PKS are further subdivided into two classes, the modular type I PKSs which are found in bacterial organisms and the iterative type I PKS which are mostly found in fungi, except for the enediyne PKS which are from an actinomycete. Modular type I PKSs work in an assembly-line-like fashion reminiscent of the NRPS wherein each domain only acts once in the synthesis of one molecule of the final compound. Thus, one can envision what the product of the PKS will be based on the number of modules found in the PKS. Moreover, the identity of the product can be deduced based on what reducing domains are missing or inactivated on the PKS modules. In addition, it was previously shown that the acyl transferase (AT) domains of modular type I PKS, unlike the MAT domain of fatty acid synthases, can utilize methylmalonyl-CoA or ethylmalonyl-CoA as extender units.[63, 65]

Iterative type I PKS, on the other hand, works in the similar way to the fatty acid synthases where in both cases a single set of domains is responsible for the biosynthesis of the

entire PKS product. Fungal iterative type I PKSs are further classified into three major classes based on the degree of β-keto reduction performed by the PKS; namely the non-reducing (NR), the partially-reducing (PR) and the highly-reducing (HR)-PKSs. The enzymology and classification of fungal iterative type I PKSs have been reviewed extensively. [64, 75, 79] As a general rule, aromatic polyketide compounds are synthesized by NR-PKSs, while HR-PKSs produce aliphatic compounds. PR-PKSs, on the other hand, has been shown to produce compounds lacking a phenolic hydroxyl group at the aromatic ring on the position where a β-keto group has been reduced to alcohol on the polyketide chain before cyclization, such as 6-methylsalicyclic acid [80, 81] and (*R*)-mellein. Griseofulvin (**18**), the subject of study in Section 4, is an example of an aromatic polyketide arising from a product of NR-PKS. Thus, we will primarily discuss the studies done on NR-PKSs in the last two decades to illustrate the progress in uncovering the mechanisms of these amazing enzymatic machineries and show the utility of heterologous expression systems and enzymatic studies in bringing forth this progress.

1.4 Diversity of Fungal Polyketides

Aside from the antifungal compound **18**, other examples of aromatic fungal polyketide arising from NR-PKS products are the mycotoxin aflatoxin (**19**) and its intermediate norsolorinic acid (**20**) and the tetracycline-like compound viridicatumtoxin (**21**).[64] Studies involving PksA, NR-PKS involved in the biosynthesis of **20**, have been particularly insightful in understanding NR-PKS enzymology and will be discussed further in later sections.[79, 82-85]

A method by which fungal organisms can diversify the product profile of NR-PKS, aside from tailoring reactions, is by using the NR-PKS in tandem with a HR-PKS. Examples of this strategy is found in the resorcylic acid lactone (RAL) hypomycetin (**22**),[86-88] the

dihydroxyphenylacetic acid lactone (DAL) 10,11-dehydrocurvularin (**23**),[89, 90] the azaphilone

azanigerone (**24**)[91]from *Aspergillus niger* and asperfuranone (**24**).[92] Studies of the

biosynthesis of **22** and other RALs are especially enlightening in understanding some of the

programming rules of HR-PKSs.[86, 88, 93]

Scheme 1.3: Examples of fungal polyketide

griseofulvin (**18**)

aflatoxin (**19**)

norsolorinic acid (**20**)

viridicatumtoxin (**21**)

hypomycetin (**22**)

10,11-dehydrocurvularin (**23**)

azanigerone (**24**)

asperfuranone (**25**)

lovastatin (**26**)

compactin (**27**)

equisetin (**28**)

tenellin (**29**)
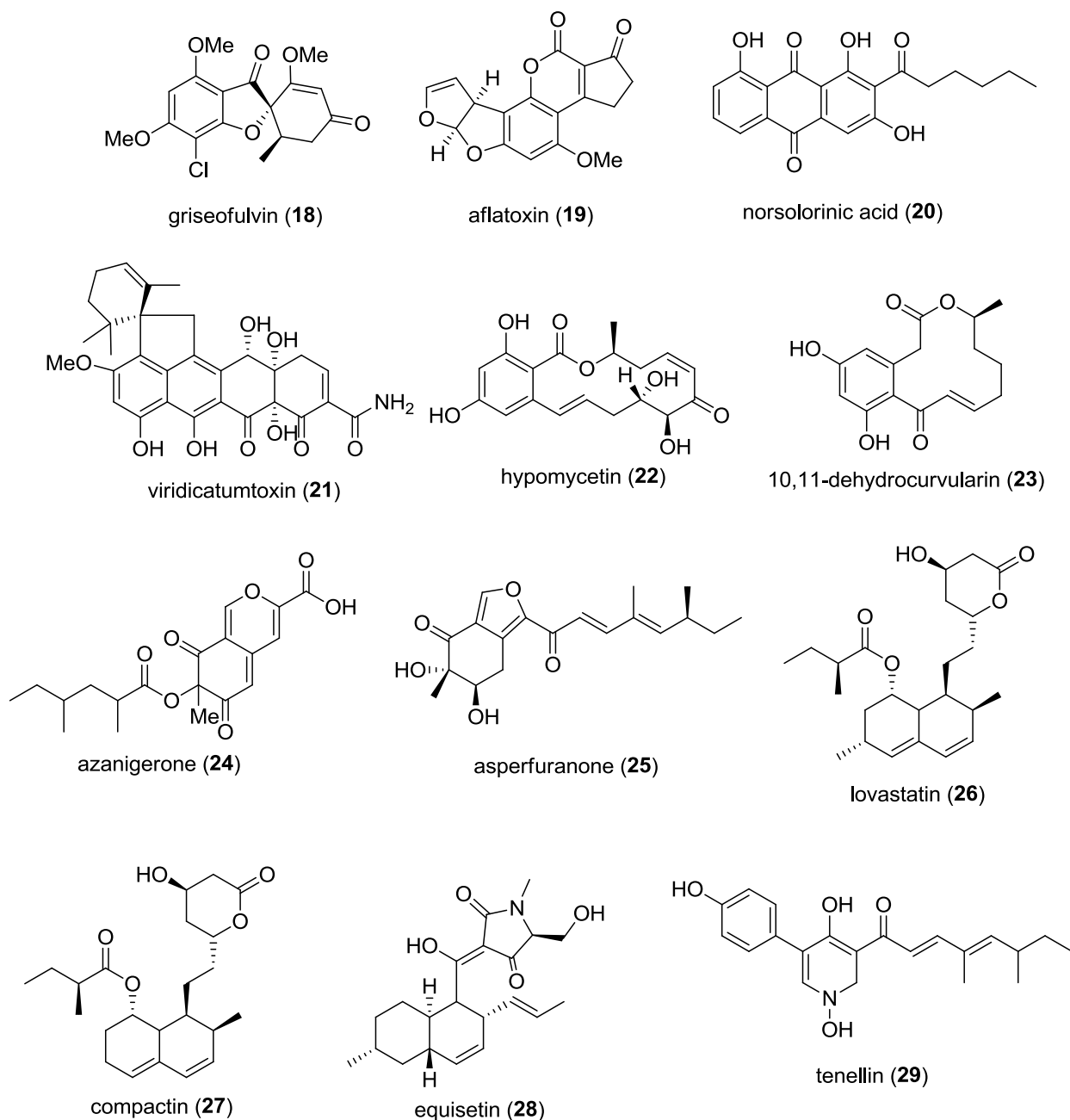
Similarly, fungal polyketide can also arise as a tandem product of two HR-PKSs such as in the HMG-CoA reductase inhibitors lovastatin(**26**) and compactin (**27**).[94-96] Both compounds also feature a decalin ring moiety which arose from an intramolecular Diels-Alder reaction of the HR-PKS intermediate. The HR-PKS involved in the biosynthesis of **26**, the lovastatin nonaketide synthase LovB, is the first example of a fungal HR-PKS that was reconstituted *in vitro*.[94] The study gave important insights into the biosynthetic programming of HR-PKS as well as opens the door towards production of lovastatin intermediates in *Saccharomyces cerevisiae*.[94, 97]

Finally, another example of product diversity arising from PKS is found in PKS-NRPS systems, where a HR-PKS is fused to a single module of NRPS. Examples of fungal PKS-NRPS product include the decalin-containing compound equisetin (**28**) from *Fusarium heterosporum* and tenellin (**29**) from *Beauveria bassiana*.[98, 99] Biochemical studies of the tenellin synthase TenS contributed to the understanding of the mechanism of fungal hybrid PKS-NRPS systems.[100, 101]

1.5 Enzymology of Iterative Nonreducing Polyketide Synthase.

Like typical PKS, iterative contain the minimal PKS domains KS, AT and ACP. However, as their name indicates, NR-PKSs lack the triad of the reductive domains KR, DH and ER.[64, 79] On the other hand, NR-PKSs contain two additional domains not found in HR-PKSs and fatty acid synthase; the starter unit:ACP transacylase (SAT) domain and the product (PT) domain.[82-84] Moreover, the thioesterase (TE) domain of NR-PKSs can also act as a Claisen-like-cyclase (CLC) domain as well.[85, 102] These three domains are critical towards imparting structural diversity found in products of NR-PKSs.[64, 79]

Most of the understanding of the enzymology of NR-PKSs was brought forth from the deconstructive studies of PksA, the norsolorinic acid synthase involved in the biosynthesis of **19** in *Aspergillus flavus*. Discussion of the studies primarily of these two systems will be presented here in relation to their respective contribution towards understanding the SAT, PT and the TE/CLC domains.
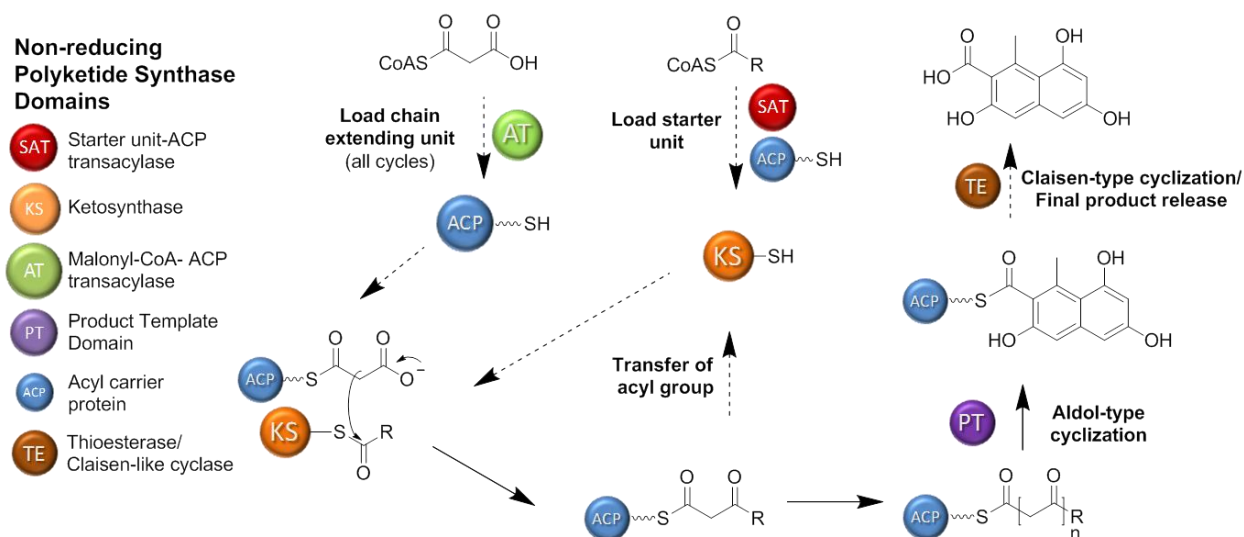
Figure 1.6: Function of the different iterative non-reducing polyketide synthase domains

1.5.1 Starter unit:ACP transacylase. One of the mechanisms that NR-PKSs utilize to provide product diversity is to change the starter unit carboxylic acid during the initiation stage of the PKS reaction. Previously characterized NR-PKSs shows that SATs can utilize simple starter units such as acetyl-CoA as starter unit as in the case for GsfA in griseofulvin biosynthesis and PKS4 in bivakerin biosynthesis.[103-105] Moreover, SATs can be utilized as a mechanism to incorporate a product of an HR-PKS as a starter unit for a downstream NR-PKS as in the case for the resorcylic acid lactones (RALs) and the azaphilones.[86, 91] Aside from the fatty acid and polyketide-derived starter unit, the structure of viridicatumtoxin suggests that a malonamate starter unit is used for priming the VrtA NR-PKS.[103]

Previous genetic and feeding studies have implicated that norsolorinic acid synthase utilizes a hexanoate starter unit to initiate the polyketide elongation.[106] In silico analysis of PksA using sequence alignment and the Udwary-Merski algorithm revealed the presence of previously uncharacterized conserved domain in NR-PKSs that contains structural similarity to MATs despite sharing minimal sequence similarity to previously characterized enzymes.[82, 107] Moreover, the previously uncharacterized domain, dubbed the starter unit:ACP transacylase (SAT) domain from PksA contains a conserved GXCXG motif and the conserved cysteine is proposed to be the active site residue involved in the transthioesterification reaction between the CoA thioesters and the ACP.[82] However, later characterization of other NR-PKSs, in particular those involved in the biosynthesis of the RALs revealed SATs can have a conserved GXSXG motif instead of a GXCXG motif, indicating that an oxyester intermediate can substitute for the thioesters intermediate found in the PksA SAT.[86, 108]

Subsequently, the transthioesterification reaction catalyzed by the PksA SAT was reconstituted using dissected SAT and ACP from PksA, demonstrating that SAT is indeed capable of transferring the hexanoate from hexanoyl-CoA to the ACP.[82] Moreover, testing of SAT activity with other acyl-CoA substrates revealed that the PksA SAT shows fidelity towards hexanoyl-CoA over noncognate substrates with lengths deviating from C6, although a more thorough study of the PksA SAT demonstrated that it can also recognize acyl-CoA substrates that have same chain length as hexanoyl-CoA such as 5-methyl-hexanoyl-*S*-NAC.[82, 109] Later studies involving swapping of the SAT of the asperfuranone NR-PKS (AfoE) SAT with the PksA SAT resulted in a chimeric AfoE:: PksA-SAT producing a hexanoyl-primed 1,4-napthoquinone derivative.[110]

1.5.2 Product template domain. As a consequence of the lack of the three reductive domains KR, DH and ER, the NR-PKS synthesizes an extremely reactive β-polyketone which can undergo aberrant cyclization via Knoevagel condensation as well as formation of polyketide lactone.[65, 75] Whereas bacterial aromatic PKSs utilizes small proteins called cyclases to guide and control the stereoselectivity of the cyclization reaction of the nascent reactive full-length polyketide product, fungal NR-PKS utilizes a different enzymatic mechanism to accomplish this feat.[111] Seminal studies by the Townsend group helped in identifying the catalytic domain responsible for the stereoselective cyclization of the NR-PKS product.[83, 84]

Once again, the Udwary-Merski algorithm was utilized to identify from the PksA the previously uncharacterized product template (PT) domain.[84] Assays of different combinations of the dissected PksA domains revealed that the PT is necessary towards the synthesis of the correct product norsolorinic acid anthrone which contains the correct C4-C9 and C2-C11 cyclization. Furthermore, structural studies of the PT domain revealed the mechanism behind the stereoselective control of cyclization.[83] Crystal structure of the dimeric PT domain revealed that it exhibits a modified double-hotdog-fold similar to what is observed with the DH domain, suggesting an ancestral relationship between DHs and PTs. The PT structure with the bound palmitate and substrate analog showed the presence of a Ppant binding pocket, and a deep two-part chamber with a hexyl-binding pocket and a cyclization chamber. It was proposed that a histidine-aspartate dyad deprotonates the C4 of the full-length PksA product and initiates the C4-C9 first ring cyclization. The first ring cyclization then re-positions the C2 proton for abstraction by the histidine to trigger the C2-C11 second ring cyclization.

Moreover, bioinformatic study and biochemical characterization of NR-PKS PT domain by Li et al helped further the understanding of the mechanism behind the PT domain

cyclization.[112] Phylogenetic analysis of PT domains from this study resulted in the classification of NR-PKS PTs into five major groups that roughly correlate with cyclization mode as well as the length of the polyketide product; three of the groups (I-III) consist of PTs that mediate C2-C7 type cyclization while group IV and group V catalyzes C4-C9 and C6-C11 first-ring cyclization, respectively. The ability of group V PTs to promote C6-C11 cyclization was also demonstrated in the study using two dissected PTs from AptA and VrtA and the minimal *G. fujikuroi* PKS4 construct PKS_WJ.

1.5.3 Thioesterase (TE)/ Claisen-like cyclase domain and other NR-PKS releasing domains.

While the function of the TE domain in fatty acid synthase is limited to the hydrolytic release of the final product, the TE domains of NR-PKSs can play multiple roles in shaping the structure of the PKS product. Foremost of this secondary function of the TE is the macrolactonization of the polyketide as is what is observed in RALs. Moreover, NR-PKS C-terminal TE domains were also implicated in final ring cyclization via Claisen condensation as first demonstrated in the NR-PKS YWA1; leading to the renaming of TE to the Claisen-like cyclase (CLC) domain.[102] More recent work on the TE domain revealed that that it plays a role in removal of stalled products from the ACP and possibly plays a role in maintaining product fidelity for the NR-PKS.[113] The crystal structure of the dissected PksA TE showed that NR-PKS TEs display a α/β-hydrolase fold and Ser-His-Asp catalytic triad found in typical serine hydrolases.[85]

While the TE/CLC domain is the better known method for NR-PKS product release, it is far from the paradigm for releasing mechanism for NR-PKS. One alternative mechanism for release is through a C-terminal reductive domain as observed in the biosynthesis of 3-methylorcinaldehyde and azanigerones.[91, 114] More recently, NR-PKSs with C-

methyltransferase domains were shown to also contain esterase/lipase-like terminal domains that have evolved separately from the α/β-hydrolase-like TEs.[115, 116]

In addition, more NR-PKS are being characterized without an in-line releasing domain as in the case of AptA which utilizes a separate enzyme to catalyze the final ring cyclization and product release. The enzyme, annotated as AptB, shows similarity to metal-dependent- β-lactamase-type hydrolases.[117] Similar β-lactamase-type hydrolase-like enzymes were also shown to be involved in the release of atrochrysone carboxylic acid and in conjunction with an oxidase, the Claisen-like cyclization of the fungal naphthacenedione TAN-1612.[118]

Section 2: Discovery and Characterization of Echinocandin B Biosynthetic Genes in *Emericella rugulosa* NRRL 11440

2.1 Introduction

Invasive candidiasis caused by opportunistic pathogenic strains of genus *Candida,* accounts for 17% of ICU-related infections, third highest after *Staphylococcus aureus* and *Pseudomonas* spp.-related infections.[119] Moreover, there has been a steady increase in the incidence of invasive candidiasis correlating with the increased use of immunosuppressants, broad-spectrum antibiotics, intravenous catheters and prosthetics and invasive clinical procedures.[120, 121] Echinocandins, a family of lipohexapeptides that prevent fungal wall synthesis through noncompetitive inhibition of 1,3-β-glucan synthase, rapidly rose to the top ranks of antifungal agents due to their activity against a wide range of *Candida* spp., in particular azole-resistant strains, and are significantly less toxic compared to Amphotericin B.[122, 123]

As the first echinocandin discovered, Echinocandin B **1**, was isolated from *Aspergillus nidulans* var. *echinulatus[124]* and *A. nidulans* var. *roseus* NRRL 11440.[7] While **1** and family members subsequently isolated, such as pneumocandin A_0 **4,**[125] aculeacin,[126] cryptocandin,[127] and mulundocandin[128], show anti-*Candida* activity, the hemolytic properties of natural echinocandins prevented their use as therapeutics. Derivatization of **1** and **4**, especially in the fatty acid moiety, led to the development of cilofungin **2**,[129, 130] anidulafungin **3** (Eraxis, Pfizer),[131] caspofungin (Cancidas, Merck and Co) **5**,[132] and micafungin **6** (Mycamine, Astellas Pharma)[133] (Scheme 2.1) that have less hemolytic properties while retaining the bioactivity of their parent compound. For example, **3**, a

semisynthetic derivative of **1** containing a substituted terphenyl acyl chain, was approved by the FDA in 2006.[131, 134]

Scheme 2.1: Echinocandin B (**1**) and its semisynthetic derivatives cilofungin (**2**) and anidulafungin (**3**). Also shown are the related natural compound Pneumocandin A$_0$ (**4**) and its semisynthetic derivatives caspofungin (**5**) and micafungin (**6**).
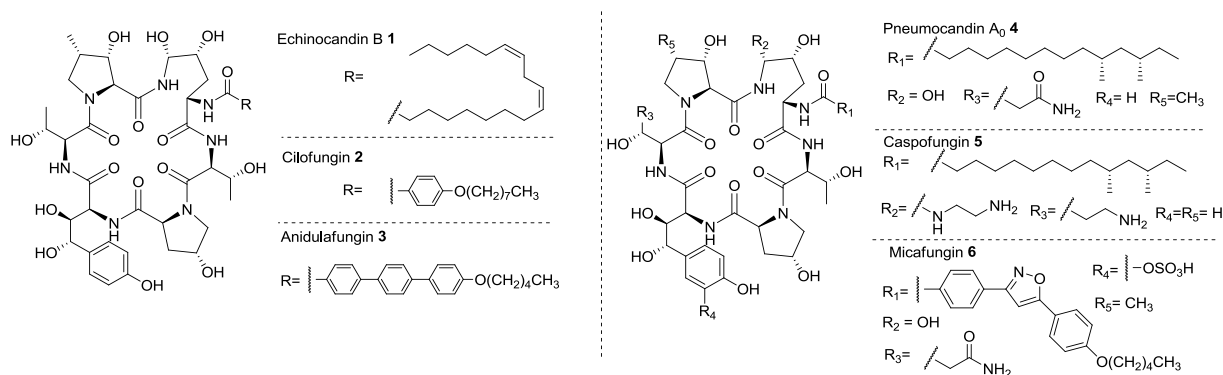
In addition to their antifungal activity, echinocandins reflect interesting biosynthetic features in their structures. Aside from the long chain fatty acyl amide, the presence of nonproteinogenic amino acids *4R*, *5R*-dihydroxyl-L-ornithine, *3S*-hydroxyl,*4S*-methyl-L-proline, *4R*-hydroxyl-L-proline and *3S, 4S*-dihydroxyl-L-homotyrosine suggests echinocandins are synthesized by nonribosomal peptide synthetases (NRPSs). An intriguing feature of echinocandins is the presence of multiple alcohol and diol groups within the scaffold as a result of the incorporation of these unnatural residues; most distinct of these hydroxyl groups are at Cδ of L-ornithine which creates a hydrolytically labile hemiaminal in the macrocycle[135, 136] and the vicinal diol found in 3*S*,4*S*-dihydroxy-L-homotyrosine. It is noted that enzymatic Cδ-hydroxylation of L-ornithine is a particularly novel and challenging reaction due to the presence of the neighboring amine. Because of this unstable linkage, the total synthesis of **1** has not been

demonstrated so far. Only simpler versions without the hydroxyl groups in the L-ornithine position, such as echinocandin D, have been synthesized and used in SAR studies.[137-139]

Despite the medical importance and intriguing structural motifs in echinocandin, the genetic and molecular basis for the biosynthesis of this family has remained unknown to date. The enzymology (sequence, specificity, type of oxygenase) behind this multitude of hydroxylation steps is also unresolved. *A. nidulans* var. *roseus* NRRL 11440 (ATCC 58397) is an industrially important strain that produces **1**, which is a precursor for the semisynthetic **3** (Scheme 2.1). A polyphasic characterization showed that this strain should belong to the *Emericella rugulosa* species and henceforth, we use that nomenclature here.[140] In this section, we report the discovery of the gene cluster of **1** in this strain through gene deletion of a multimodular NRPS and biochemical characterization of the enzymes involved in the lipo-initiation process. Through gene deletion and chemical complementation, we have also uncovered the separate gene cluster responsible for the biosynthesis of L-homotyrosine in the peptide scaffold of **1.**

2.2 Materials and Methods

2.2.1 General Methods and Material. *E. rugulosa* (*A. nidulans* var. *roseus*) strain NRRL 11440 was obtained from Agricultural Research Services Culture Collection (Peoria, Il). Authentic standard for **1** was purchased from Santa Cruz Biotechnology, Inc (Santa Cruz, CA). Primers used in this study were ordered from Integrated DNA Technologies and are listed in Table 2.2. Sequencing of heterologous expression constructs and knockout cassettes was performed by Laragen, Inc (Culver City, CA). RNA for cDNA amplification was isolated using RiboPure-

Yeast Kit from Ambion. First strand cDNA synthesis was performed using SuperScript III- First

Strand Synthesis SuperMix (Invitrogen Corp.).

	Unused Reads	All Contigs	Contigs > 100bp	Contigs > 1000bp
No. of contigs	25	433	433	373
Min length	88	139	139	1009
Median length		154,740	1547,40	21,740
Mean length	188	74,420	74,420	86,322
N50 length		235,313	235,313	235,313
No. of contigs >N50		38	38	38
Length sum	4719	32,224,016	32,224,016	32,198,232

Table 2.1: Genome Assembly Statistics for *Emericella rugulosa*.

2.2.2 Illumina Hiseq2000 Sequencing and Bioinformatic Analysis. The genomic DNA for

sequencing was isolated as described elsewhere from stationary liquid cultures.[141] Shotgun

sequencing was performed at Ambry Genetics (Aliso, Viejo, CA) using Illumina Hiseq 2000

with a read size of 157 bases resulting in a total of ~17 Gbp reads. The Illumina sequencing

reads were assembled using a hierarchical assembly method using the assembly programs

SOAPdeNOVO[142] and Geneious™ and performed by the UCLA Hoffman Cluster. First, the

~17 Gbp total reads were assembled using SOAPdeNOVO using k-mer size of 87 bp. The first-

tier contigs were then assembled using Geneious to generate 433 second tier contigs with a N50

of 235,313 and total length of ~32 Mb.

The second-tier contigs were converted to BLAST database format for local BLAST

search using stand-alone BLAST software (v. 2.2.18). Gene predictions were performed using

the FGENESH program (Softberry) and manually checked by comparing with homologous

gene/proteins in the GenBank database. Functional domains in the translated protein sequences were predicted using Conserved Domain Search (NCBI) or InterproScan (EBI). The nucleotide sequences for the *ecd* and *hty* gene clusters are deposited to Genbank database with accession numbers JX421684 and JX421685, respectively.

2.2.3 Fungal Transformation and Gene Disruption. Polyethylene glycol-mediated transformation of *E. rugulosa* NRRL 11440 was performed as done previously[143] with the following modifications: the spores from two plates were grown in 250 mL GMM with 10 mM ammonium tartrate as sole nitrogen source for 16 hrs at 250 rpm, 28°C and 1 g of the germlings from the culture was used for digestion using 3 g of Vinotaste Pro enzyme mixture (Novozyme). The glufosinate resistance gene *bar* with an upstream *trpC* promoter was amplified from pBARGPE1 plasmid obtained from the Fungal Genetic Stock Center (Kansas City, MO). Glufosinate used for the selection of transformants was prepared by extracting twice with equal volume of 1-butanol from commercial herbicide Finale(Bayer), which contains 11.33% (w/v) glufosinate-ammonium.[144] Construction of the knockout cassette was performed by fusion PCR as described elsewhere.[143] Fungal genomic DNA from the transformants was isolated using ZR-Fungal/Bacterial DNA Miniprep Kit (Zymo). Primers used for PCR screening are listed in Table 2.2.

2.2.4 Extraction and Characterization of 1. The echinocandin extraction method is based on US Patent #4,288,549.[7] Briefly, the strain was grown in Medium 2 [2.5% (w/v) glucose (Sigma), 1% peptone (BD Biosciences), 1% (w/v) starch (Sigma), 1%(w/v) molasses, 0.4% (w/v) N-Z Amine A (Sheffield Biosciences), 0.2% (w/v) calcium carbonate (Sigma)] in 28°C for 7 days in 10 mL cultures for screening of transformants and 2 L for large scale compound extraction. An equal volume of methanol was added to the whole fermentation broth and the

mixture was shaken at 16°C for 1 hour. The mixture was then filtered and the pH adjusted to 4.0.

The filtrate was extracted twice with equal volumes of chloroform. The concentrated extract was

purified using Sephadex LH-20 resin using 1:1 methanol and chloroform solvent system. The

fractions containing **1**, determined by LCMS, was further purified using HPLC with C-18

column and a 40-95%, 20 minute water:acetonitrile gradient. Purified **1** was further characterized

via Agilent 6520 high-resolution Q-TOF/LC-MS and its 1D-^{1}H NMR, 2D-COSY, HSQC and

HMBC spectra were recorded using a Varian 600 MHz NMR spectrometer.

2.2.5 Anti-*Candida* Assay. $\Delta ecdA$ and $\Delta htyA$ mutants were grown in 10 mL of liquid Medium 2

at static conditions for 7 days at 28°C. Discs of 10 mm in diameter were cut from the fungal mat

and transferred to yeast extract-peptone-dextrose (YPD) agar plates that were previously

inoculated with *Candida albicans* ATCC 90234. The co-cultures were grown at 28°C overnight.

For chemical complementation of $\Delta htyA$, L-homotyrosine (0.1 mg/mL) was fed to the static

liquid cultures at day-4 before transferring the mycelial discs from individual clones to the YPD

plate pre-inoculated with *C. albicans* at day-7.

2.2.6 Cloning of EcdA-M1. The C-terminal boundary of EcdA-T_1 was predicted through

alignment of gramicidin synthetase PCP (accession: 1DNY) and fungal NRPSs with EcdA by

ClustalW. EcdA-M_1 was amplified from *E.rugulosa* cDNA sample using the primer pair *Nde*I-

EcdA-T_0 and *EcoR*I-EcdA-T_1 (See Table 2.2). The amplicon was gel-purified, digested with

*Nde*I and *EcoR*I and ligated into pET28a expression vector to create pET28a-EcdA-M1.

QuikChange I Site-Directed Mutagenesis Kit (Agilent Technologies) was used to clone the

EcdA-M1 variants using pET28A-EcdA-M1 as template and ecdA-S47A-F and ecdA-S47A-R as

primer pair to create S47A variant and ecdA-S1127A-F and ecdA-S1127A-R primer pair to

create S1127A variant.

Table 2.2: List of primers used for Section 2.

Primer Name	Primer Sequence (5' -3')	Notes
ecdA-KO-P1	aggggggtcataacaaccgct	For mutant screening and KO cassete cloning
ecdA-KO-P2	gcctgatgcccaggcaacctc	For knockout cassette cloning
ecdA-KO-P3	cgtccggtcctgcccgtcaccgagatttaggctgtgggctgatgcgtggt	
ecdA-KO-P4	tgctccttcaatatcatcttctgtcggctggctcgatcatgtcacagta	
ecdA-KO-P5	ggccacaatgttgcctacttgccac	
ecdA-KO-P6	gtccagtcaatgtgagtccgca	For mutant screening and KO-cassete cloning
htyA-KO-P1	cgtaatacgcatgttggcagataaatagtt	For mutant screening and KO-cassete cloning
htyA-KO-P2	ccgtcctccactagctggaat	For knockout cassette cloning
htyA-KO-P3	cgtccggtcctgcccgtcaccgagatttagagctagagtctccgcttccgaa	
htyA-KO-P4	aaagtgctccttcaatatcatcttctgtcgccgcgtttcagcatcccggtt	
htyA-KO-P5	tctcccaggtgggttggcgga	
htyA-KO-P6	cacaacgcgaggttcccacgaaca	For mutant screening and KO-cassete cloning
bar-f	aagtaaccatgagcccagaacga	For mutant screening
bar-r-stp	ctaaatctcggtgacgggca	For mutant screening and KO cassette cloning
PtrpC-F	cgacagaagatgatattgaa	For KO cassette cloning
*Nde*I-EcdA-T$_0$	aaaaacatatgcacagaaccaacgagatggaa	For cloning of EcdA-M1
*EcoR*I-EcdA-T$_1$	tttttgaattcgggtccgcctcctgcat	
ecdA-S47A-F	ggaggcgacgcgatggctacc	For cloning of EcdA-M1-S47A variant
ecdA-S47A-R	ggtagccatcgcgtcgcctcc	
ecdA-S1127A-F	gggtgatgcaatcgccgcgat	For cloning of EcdA-M1-S1127A variant
ecdA-S1127A-R	atcgcggcgattgcatcaccc	
ecdI-*Nde*I-F	aaaaaacatatggtcttcacttcccagca	For cloning EcdI
ecdI-*EcoR*I-R	tttttgaattcttaaatcaactccttcaact	

2.2.7 Cloning of EcdI. EcdI cDNA was amplified from *E. rugulosa* cDNA sample using primers EcdI-*Nde*I-F and EcdI-*EcoR*I-R. The amplicon was ligated into PCR-blunt vector and was transformed into TOP10 cells. The plasmid was sequenced to confirm correct splicing of the transcript. The plasmid bearing the correct sequence was digested with *Nde*I and *EcoR*I and the insert was cloned into pET28a vector to create pET28a-EcdI.

2.2.8 Heterologous expression of EcdA-M1 and EcdI. pET28a-EcdA-M1 or pET28a-EcdI was transformed into BL21 (DE3) and the cells were grown in 500 mL LB at 37°C and 250 rpm. When the OD600 reading reached 0.4, the cultures were cooled to 16°C and protein expression was induced by addition of 60 µM IPTG. After 16 hr of shaking at 16°C, the cells were pelleted and resuspended in Buffer A (50 mM Tris-HCl, pH 7.9, 5 mM NaCl, 1 mM DTT) with 20 mM imidazole. The cells were lysed via sonication and centrifuged at 4°C at 15000 rpm. Nickel-NTA resin was added to the supernatant and was gently stirred at 4°C for 2 hours. The protein/resin mixture was loaded into a gravity flow column and the His-tagged proteins were purified with increasing concentration of imidazole in Buffer A. Additional anion exchange column purification for EcdI in was performed using a 80 minute gradient of 0-100% Buffer A and Buffer B (50 mM Tris-HCl, 1 M NaCl, 2mM DTT) using a MonoQ 10/100L Anion exchange column (GE Healthcare Life Sciences). The fractions containing the desired protein were concentrated using ultrafiltration column (100 kDa cutoff, for EcdA-M1 and 30 kDa cutoff for EcdI). Protein concentration was determined via by UV absorbance at λ=280 nm.

2.2.9 ATP-[^{32}P]PPi Exchange Assays. A typical reaction mixture (500 µL) contained 1.0 µM EcdA, 2 mM amino acid substrate (unless specified), 5 mM ATP, 10 mM MgCl$_2$, 5 mM Na[^{32}P]-pyrophosphate (PPi) (~1.8 x 10^6 cpm/mL), and 50 mM Tris-HCl (pH 8). Mixtures were incubated at ambient temperature for regular time intervals (e.g., 5 min), and 150 µL aliquots

were removed and quenched with 500 µL of a charcoal suspension (100 mM NaPPi, 350 mM HClO$_4$, and 16 g/L charcoal). The mixtures were vortexed and centrifuged at 13000 rpm for 3 min. Pellets were washed twice with 500 µL of wash solution (100 mM NaPPi and 350 mM HClO$_4$). Each pellet was resuspended in 500 µL wash solution and added to 10 mL Ultima Gold scintillation fluid. Charcoal-bound radioactivity was measured using a Beckman LS 6500 scintillation counter.

2.2.10 Loading of [^{14}C]-substrate onto NRPS. The assay was carried out in two steps. First, a 50 µL reaction containing 10 µM EcdA-M1, 20 µM Sfp, 1mM CoA, 10 mM MgCl$_2$, 1 mM TCEP, and 50 mM HEPES (pH 7.0) was incubated at ambient temperature for 30 min to convert *apo* EcdA-M1 to its holo-form. Afterwards, 8 µM EcdI, 5 mM ATP, and ~ 40 µM [^{14}C]-substrate (~ 4.4 x 10^6 cpm/mL) were added and incubated for 30 min. The reaction was quenched by 600 µL acetonitrile for the assays with [^{14}C]-acyl substrate or 600 µL 10% trichloroacetic acid for [^{14}C]-L-ornithine with addition of 100µL of 1 mg/mL BSA. The mixture was vortexed and centrifuged at 13,000 rpm for 3 min. The pellet was then washed twice with 600 µL acetonitrile, dissolved in 250 µL formic acid, added into 10 mL Ultima Gold scintillation fluid and subjected to a Beckman LS 6500 scintillation counter.

2.3 Results

2.3.1 Whole Genome Sequencing and Analysis of NRPS Gene Clusters of *Emericella rugulosa* NRRL 11440. Whole genome shotgun sequencing of *E. rugulosa* NRRL 11440 was performed using Illumina HiSeq2000 to generate ~17 Gbp of sequence. Assembly of the sequence reads generated 433 contigs with N50 length of 235,313 bases (See Table 2.1).

Table 2.3: Putative NRPSs in *Emericella rugulosa.*

Protein annotation	NRPS domains	Closest homolog in NCBI database through BlastP search (Identity, Similarity)	E values, Score
ErNRPS284	C-A-T-C-A-T-C-A-T-E-C-A-T-C-A-T-C_T	EHK21807.1* (47%, 63%), *Trichoderma virens* Gv29-8	0.0, 1.095×10^4
ErNRPS57	A-T-C-A-T-E-C-A-T-C-A-T-C-A-T-C-A-T-R	AN7884.2 (92%,96%), *A.nidulans* A4	0.0, 1.379×10^4
ErNRPS123 (*ecdA*)	T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C_T	EGR45389* (33%, 52%), *Trichoderma reesei* QM6a	0.0, 3.404×10^4
ErNRPS74	A-T-E-C-A-C-A-T-C-A-T-E-C-T-C-T	AN0016 (92%, 96%) , *A.nidulans* A4	0.0, 1.127×10^4
ErNRPS239	A-T-E-C-A-C-A-T-C-A-T-E-C-T-C-T-C-T	AN1242.2 (91%,94%), *A.nidulans* A4	0.0, 1.340×10^4
ErNRPS99	T-E-C-A-T-E-C-A-T-C-A-T-C-A-T-C-A-T-C_T	AN2545, *easA* (91%, 94%), *A.nidulans* A4	0.0, 1.303×10^4
ErNRPS32	A-T-C-A-T-C-A-T-C-T-C-T-C_T	AN0607 (91%, 95%) (sidC), *A.nidulans* A4	0.0, 9032
ErNRPS184	A	AN8433.2 (91%,94%), *A.nidulans* A4	0.0, 1477
ErNRPS30	A	ATEG09019 (76%, 86%), *A.terreus*	5×10^{-70}, 265
ErNRPS153a	A-T-Te	AN8513 (92%, 94%) TdiA (terrequinone NRPS) *A.nidulans* A4	0.0, 1842
ErNRPS153b	A-T-C-A-T-C_T	EED13064.1, GliP-like,(48%, 68%), *Talaromyces stipitatus* ATCC 10500	0.0, 1896
ErNRPS 92	T-C-A-T-C_T	EGE04746 (38%, 55%)*, *Trichophyton equinum* CBS 127.97	0.0, 1134
ErNRPS84	A-T-Te	NFIA_045590(69%, 80%), *N.fischeri* NRRL 181	0.0, 1479
ErNRPS24a	A-T-Te	AN8105.2 (90%, 94%), *A.nidulans* A4	0.0, 1957

Table 2.3: NRPS genes in *Emericella rugulosa* (continued).

Gene annotation	NRPS domains	Closest homolog in NCBI database through BlastP search	E values, Score
ErNRPS24b	A-T	AN9216 (87%,90%), *A.nidulans* A4	0.0, 1003
ErNRPS105a	A-T-Te	AN5610.2 (99%, 99%), *A.nidulans* A4	0.0, 2861
ErNRPS105b	A-T	AN5626.2 (98%, 99%), *A.nidulans* A4	0.0, 1352
ErNRPS146	A-T-Te	AFUA_8G01640 (62%, 75%), *A.fumigatus*	0.0, 1296
ErNRPS93	A-T-Te	NFIA_005280 (76%, 87%), *N.fischeri* 181	0.0, 1614
ErNRPS20	A-T	AFLA_066840 (40%, 58%), *A.flavus* NRRL3357	7×10^{-11},348
ErNRPS100a	C-A	AN2621 (68%, 76%) AcvA synthetase, *A.nidulans* A4	3×10^{-145}, 750
ErNRPS100b	A	AN2621 (63%, 69%)* AcvA synthetase *A.nidulans* A4	2×10^{-100}, 774
ErNRPS291	A-T	AN5990.2 (98%, 99%), *A.nidulans* A4	0.0, 1128
ErNRPS247	A-T	AAX09993.1 (52%, 68%)*, *Cochliobolus heterostrophus*	0.0, 1015
ErNRPS87	A-C	EFY94744.1 (37%, 55%), *Metarhizium anisopliae* ARSEF 23	6×10^{-63}, 229
ErNRPS43	A-T-R	AN5318 (98%, 99%), *A.nidulans* A4	0.0, 2536

Definitions:

A-adenylation domain

T- Thiolation/ Peptidyl Carrier Protein Domain

C – Condensation domain

R – Reductase domain

Te- thioesterase domain

C_T- Terminal Condensation domain

Table 2.4: The Echinocandin B Biosynthetic Gene Cluster.

Gene	Length (aa)	Conserved Domain/Function	Nearest BLAST hit (Identity, Similarity)
ecdB	545	fungal transcription factor	*A. fumigatus* AFUA_048990 (72%, 80%)
ecdC	556	transporter (MFS)	*N. fischeri* NFIA_042010 (70%, 79%),
ecdD	541	transporter (MFS)	*N. fischeri* NFIA_042010 (87%, 93%)
ecdE	703	glycosyl hydrolase	*N. fischeri* NFIA_042050 (78%, 88%)
ecdF	508	glycosidase	*P. purpurogenum* BAA12320 (71%, 84%)
ecdA	7260	NRPS (T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C)	*T. reesei* EGR45389(33%, 52%)
ecdG	338	non-heme iron, α-ketoglutarate dependent dioxygenase	*C. militaris* CCM_03049 (31%, 50%),
ecdH	503	cytochrome P450 heme-iron-dependent oxygenase	*N. haematococca* 100691 (28%, 48%),
ecdI	559	fatty-acyl- AMP ligase	*A. nidulans* A4, AN3490 (52%, 67%),
ecdJ	668	hypothetical protein	*A. capsulatus* 05345 (34%, 47%),
ecdK	332	non-heme iron, α-ketoglutarate dependent dioxygenase	*T. reesei* TRIREDRAFT_58580 (43%, 63%)
ecdL	1479	multidrug transporter (ABC)	*M. anisoplae* MAA_01638 (40%, 59%)

The total length of the 433 contigs amounts to 32,224,016 bases, which is slightly larger than the genome size of the previously sequenced *A. nidulans* A4 strain of ~31 Mb.[145]

Using the first adenylation (A) domain of the enzyme TqaA (accession number: ADY16697) from the tryptoquialanine pathway[47] as BLAST query, we were able to find 26

putative NRPS genes in *E. rugulosa* (Table 2.3). Four out of the twenty-six NRPSs, which we denoted as ErNRPSs, are NRPSs with five or more modules. Two of these four ErNRPSs, annotated as ErNRPS99 and ErNRPS57, are also found in the model fungi *A. nidulans* A4 that does not produce **1**. ErNRPS99 is likely a homolog of EasA found in the biosynthesis of emericellamide[59] due to the high shared sequence identity between the two. ErNRPS57, on the other hand, contains a terminal reductive domain (R) similar to what is found in peptaibol synthetases.[146] Thus, this leaves ErNRPS284 and ErNRPS123 as echinocandin synthetase candidates. ErNRPS284 (Table 2.3) contains five modules, which are insufficient for catalyzing the formation of the hexalipopeptide scaffold of **1** based on the colinearity hypothesis.[147] On the other hand, the six-module ErNRPS123 (799 kDa), which is annotated as *ecdA*, has the correct number of modules necessary for the assembly of **1**. However, performing A domain selectivity prediction using NRPS predictor[148] offered little confirmation that this is the correct echinocandin NRPS, with only the third A domain prediction (L-proline) matching the corresponding amino acid in **1** (L-proline or *4R*-hydroxyl-L-proline) (Table A1). Further bioinformatics analysis of the domain architecture of EcdA, as well as the genes within the *ecd* cluster, gave indications that this is most likely the correct gene cluster that is consistent with some of the expected biosynthetic transformations required for the assembly of **1**. First, EcdA has a terminal condensation domain (C_T) that has been shown to catalyze the cyclization of NRPS products in fungi[149], in agreement with the anticipated macrocyclization of the hexapeptide. Furthermore, in proximity to *ecdA* is *ecdI* (Figure 2.1A and Table 2.4), an acyl-AMP ligase homolog gene, giving a plausible route for lipo-initiation of the NRPS. In addition, other genes adjacent to *ecdA* encode putative non-heme iron, α-ketoglutarate-dependent oxygenases (*ecdG* and *ecdK*), and a cytochrome P450 heme-iron-dependent oxygenase
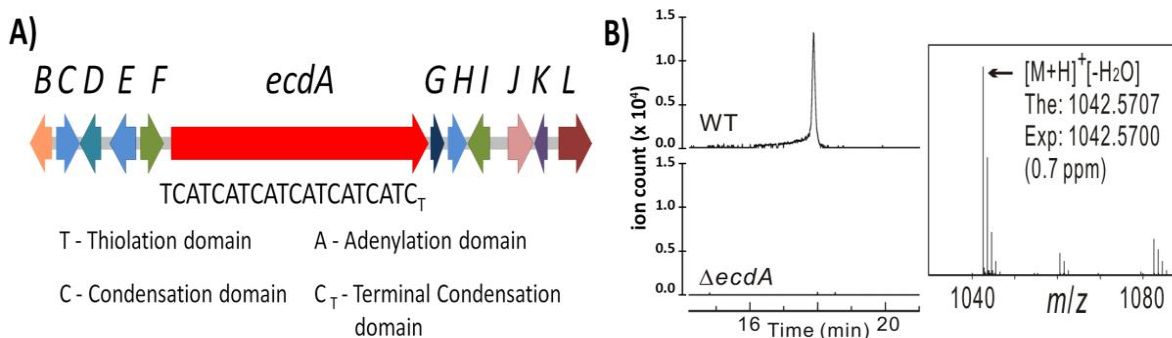
Figure 2.1: Identification of the *ecd* gene cluster. A) Organization of the *ecd* gene cluster. *ecdA* encodes for the six-module nonribosomal peptide synthetase. *ecdI* encodes for the linoleyl-AMP ligase responsible for lipo-initiation of EcdA. Also found in the cluster are *ecdG*, *ecdH*, and *ecdK* encoding for proposed hydroxylases; B) Metabolic profile of Δ*ecdA* mutant shows the loss of production of **1**.

(*ecdH*) (Table 2.4), indicating that the nascent NRPS product may undergo several oxygenation steps, as anticipated for the biosynthesis of **1**. Other genes flanking the putative *ecd* cluster encode a fungal transcription factor gene (*ecdB*), three transporter protein genes (*ecdC, ecdD* and *ecdL*), a glycosyl hydrolase gene (*ecdE*), a glycosidase gene (*ecdF)* and a gene encoding for a protein with no predicted conserved domain (*ecdJ*). Interestingly, genes encoding the biosynthesis of unnatural amino acids such as l-homotyrosine are not present in the vicinity of *ecd* gene cluster, but are found to reside elsewhere on the genome (see below).

2.3.2 Verification of the role of *ecdA* in the biosynthesis of 1. To confirm the production of **1**, *E. rugulosa* was grown in Medium 2 and the metabolite extracted from the fermentation broth in the same manner as previously described.[7] High resolution LCMS trace of the extracted metabolites shows a peak at 17.6 min with *m/z* of 1042.5700 [M+H-H$_2$O]$^+$ corresponding to the theoretical mass-to charge ratio of **1** (*m/z* =1042.5707 [C$_{52}$H$_{81}$N$_7$O$_{16}$+H-H$_2$O]$^+$) (Figure 2.1B). Purification and subsequent characterization by NMR (Table A3 and Figure A10-A13) and comparison with authentic standard (Figure A1) verified that the compound is indeed **1.**
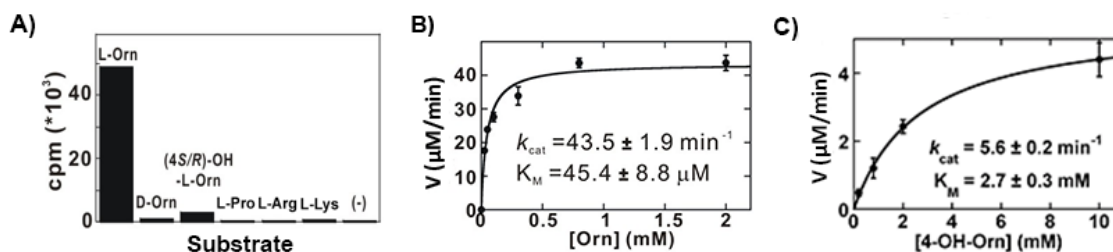
Figure 2.2: ATP-[^{32}P]-PPi exchange assay of EcdA-M1 A) The adenylation (A) domain of EcdA-M1 shows preference towards L-ornithine   B) Michaelis-Menten plot for the A domain of EcdA-M1 with L-ornithine as substrate. C) Michaelis-Menten plot for the A domain of EcdA-M1 with 4(R/S)-OH-L-ornithine as substrate.

To verify whether the *ecd* cluster is responsible for the production of **1**, we developed a gene deletion method for *E. rugulosa* based on previous methods for *A. nidulans* A4,[143] using the glufosinate resistance gene *bar* as selection marker (See Materials and Methods).[150] A gene deletion cassette containing portions of *ecdA* internally disrupted by the *bar* gene driven by the *trpC* promoter was introduced into *E. rugulosa* protoplasts via PEG-mediated transformation and the resulting glufosinate-resistant strains were selected. A bioassay-guided knockout screening was developed in which individual fungal colonies are spotted on plates pre-inoculated with *C. albicans*. Approximately 100 mutants were screened by both loss of anti-*Candida* activity, and PCR-based screening using a *bar* gene primer and a primer found outside the knockout cassette. One mutant (Δ*ecdA* I-16) was isolated which lost the ability to inhibit the growth of *C. albicans* (Figure A2) and also showed the correct PCR-amplified product (Figure A3). LCMS analysis of the extracted metabolites after 7 days of growth showed no production of **1** (Figure 2.1B), suggesting that EcdA is the NRPS responsible for the biosynthesis of **1**.
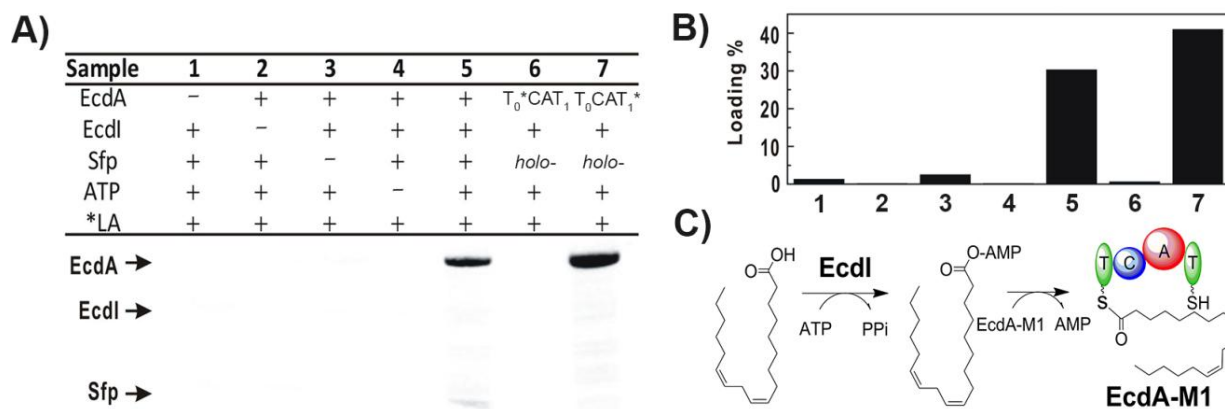
Figure 2.3. Loading Assay of EcdA-M1. A) EcdA-M1, converted to *holo* form *in vitro* by Sfp, is loaded with ^{14}C-linoleic acid only in the presence of the fatty-acyl-AMP ligase EcdI and ATP as shown in the autoradiogram (lane 5). Loading of EcdA-M1 variants with ^{14}C-linoleic acid in the same condition as sample 5 B) Quantification of percentage of EcdA-M1 loaded with radioactive linoleic acid. Samples 1-7 carried out in identical conditions as in A. C) Linoleic acid is activated by EcdI to form linoleyl-AMP which is subsequently transferred to EcdA-M1.

2.3.3 Characterization of the Adenylation domain (A) of the First Module of EcdA.

Guided by the results of the gene knockout of *ecdA*, we proceeded to determine the amino acid specificity of the first A domain of EcdA by cloning the initiation module of EcdA, which is fused to a thiolation domain (T_0) that is likely the site of lipid attachment (EcdA-M1, T_0CAT_1). The 130 kDa protein was expressed in *Escherichia coli* BL21(DE3) cells to a final titer of ~20 mg/L and purity of >95% (Figure A4). Amino acid-dependent ATP-[^{32}P]-PPi exchange assay with 2 µM of EcdA-M1 at ambient temperature and incubation time of 30 minutes showed that A_1 could activate L-ornithine (L-Orn) and ($4R/S$)-4-hydroxyl-L-ornithine (4-OH-L-Orn). D-ornithine and other amino acids with basic side chains such as L-lysine and L-arginine were not activated (Figure 2.2A). Full kinetic analysis shows that L-Orn (k_{cat}= 43.5 ± 1.9 min^{-1}, K_M= 45.4 ± 8.8 µM) is about a 500-fold better substrate than 4-OH-L-Orn (k_{cat}= 5.6 ± 0.2 min^{-1}, K_M= 2.7 ± 0.3 mM) as judged by k_{cat}/K_M ratios (Figure 2.2B and 2.2C), suggesting that oxidation at Cγ of L-

ornithine most likely occurs after loading of L-Orn to T_1. The activation of L-Orn by the first module of EcdA further supports the link between the NRPS and the biosynthesis of **1.**

2.3.4 EcdI catalyzed Lipo-initiation of EcdA. The addition of the lipid chain in the biosynthesis of lipopeptides, also known as lipo-initiation[151, 152], is a critical step not only because it connects the fatty acid pool and NRPS biosynthesis, but also due to the importance of the lipid to the antimicrobial activities.[131] Several mechanisms exist in activating and transferring the lipid chain to the NRPS.[152] These include fusion of a fatty acid synthetase (FAS)-like module to the N-terminus of a NRPS (mycosubtilin)[153]; and transfer of the lipid chain to a dissociated T domain (daptomycin)[154, 155] or coenzyme A (surfactin)[151] by a fatty acyl ligase, followed by condensation with an aminoacyl adenylate catalyzed by the first C domain. Since EcdA only contains six A domains, one for each of the residues in the hexapeptide scaffold, we reasoned that EcdI, containing an AMP-binding domain might be responsible for the formation of the activated form of linoleic acid and its transfer to the T_0 domain of EcdA. To test this hypothesis, we cloned EcdI Into an *E. coli* expression vector and expressed the *N*-terminal His-tagged protein in BL21(DE3) with a titer of ~30mg/L and purity of ~95% (Figure A4). Using the purified phosphopantetheinyl transferase Sfp from *Bacillus subtilis,*[156] we converted *apo* EcdA-M1 into its *holo* form *in vitro*. After co-incubation of *holo* EcdA-M1 with EcdI, ATP and [^{14}C]-linoleic acid, the assay mixture was analyzed using SDS-PAGE and autoradiography. The autoradiogram shows a strong radiolabeling of the 130 kDa band, indicating that the ^{14}C-linoleic acid is covalently bound to EcdA-M1 (Figure 2.3A). In contrast, in the absence of EcdI, ATP or Sfp, nearly no labeling of EcdA-M1 can be detected, confirming the proposed mechanism (Figure 2.3C). Quantification of the radiolabel suggests that ~30% of

EcdA-M1 is loaded with the labeled substrate (Figure 2.3B), consistent with fractional stoichiometries seen in other labeling studies.[157, 158]

To pinpoint which of the two thiolation domains in EcdA initiation module (T_0CAT_1) is acylated by the linoleoyl starter unit in the presence of EcdI, single mutations to the active site serines in the two thiolation domains were made. The S47A mutant (T_0*CAT_1) prevents phosphopantetheinylation of the initiation T domain, but leaves the T_1 domain available for conversion into the holo form that should be capable of undergoing covalent loading with [^{14}C]-L-ornithine. Conversely, the corresponding S1127A mutant (T_0CAT_1*) will not be phosphopantetheinylated at T_1 but should be able to load [^{14}C]-linoleate. Both the S47A and the S1127A mutants were expressed from *E.coli* BAP1,[159] which coexpresses Sfp for thiolation domain phosphopantetheinylation (thus no Sfp pre-incubation is required) (Figure A5). As shown in Figure A5, [^{14}C]-linoleate is loaded onto the T_0CAT_1* but not the T_0*CAT_1 variant of the EcdA initiation module, as assessed by both autoradiography and by radioactive counting of protein precipitated via addition of trichloroacetic acid. The complementary result is seen for [^{14}C]-L-ornithine covalent loading in which T_0*CAT is labeled but T_0CAT_1* is not (Figure A6). The relatively low amount of L-Orn radioactivity may reflect lability of the thioester due to the attack by the Nδ of the ornithine side chain to the activated carbonyl carbon followed by the subsequent release of the cyclic δ-lactam, a known propensity of ornithine thioesters.[160]

In previously characterized lipopeptide NRPSs, the incorporation of fatty acids into the assembly line requires the formation of either acyl-CoA[151] or acyl-AMP[154] prior to loading onto the NRPS. To explore the mechanism of linoleic acid activation by EcdI, we removed excess CoA after conversion of *apo* EcdA-M1 to its *holo* form, followed by addition of 0, 0.5 or 5.0 mM CoA to the assay. The reaction rates for the acyl loading step, as determined by the

fraction of EcdA-M1 loaded with labeled substrate at different time points, are essentially the same (Figure A7), ruling out the CoA-dependent mechanism and indicating that the acyl-AMP is directly transferred to EcdA-T_0 by EcdI (Figure 2.3C).

In order to probe the substrate specificity of EcdI, we co-incubated *holo*-EcdA-M1 with alternative acyl substrates. Palmitic acid (C-16) showed similar degree of loading to EcdA-M1 compared to linoleic acid (27% vs. 30 % for linoleic acid). Decreasing the length of the acyl chain to C-10 ($[^{14}C]$-decanoic acid), on the other hand, dramatically reduced the loading of EcdA to ~8% at time points where the loading of linoleic acid reaches saturation. Nevertheless, this range of fatty acyl chain lengths that can be transferred to EcdA-M1 indicates there is room for incorporation of alternative lipid starter molecules into the structure of **1.** Incubation of EcdI and EcdA-M1 with non-fatty-acid substrate such as benzoic acid, an aryl carboxylate, did not show covalent loading onto T_0 (Figure A8).
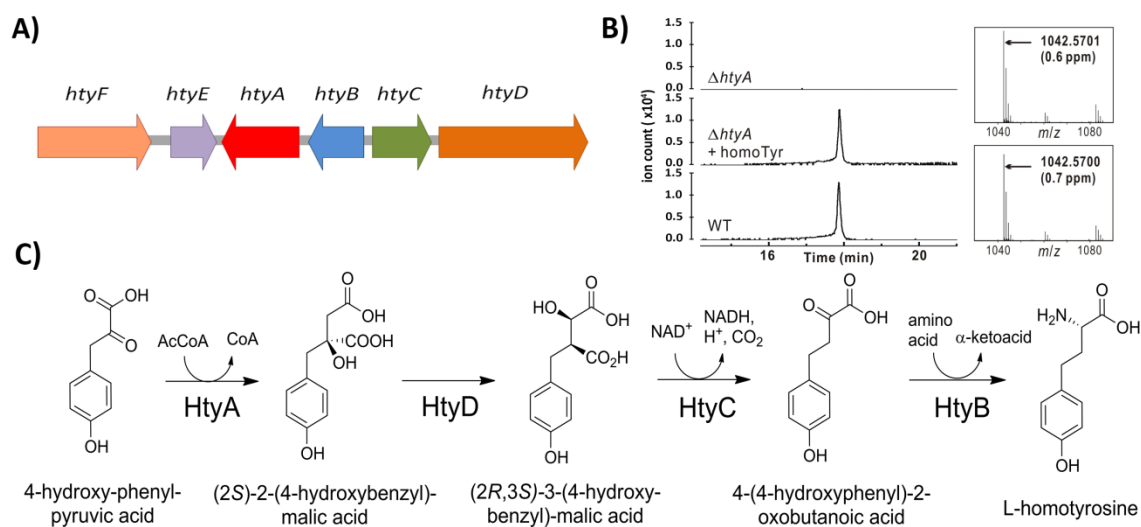
Figure2.4: Biosynthesis of L-homotyrosine. A) Organization of *hty* gene cluster B) Disruption of *htyA* led to the loss of production of **1**, which was restored upon addition of L-homotyrosine to the culture. C) Putative biosynthesis of L-homotyrosine.

2.3.5 Discovery of the L-homotyrosine biosynthetic gene cluster. One of the characteristic features of the echinocandin family is the presence of nonproteinogenic *3S, 4S*-dihydroxyl-L-homotyrosine in the 4th amino acid position of the cyclic hexapeptide. This residue is derived from the dihydroxylation of L-homotyrosine either prior to, or after incorporation into the peptide scaffold. It was previously shown that L-homotyrosine in the scaffold of **4** is derived from the condensation of 4-hydroxyphenyl-pyruvate and acetate to form 2-(4- hydroxybenzyl)-malate.[161] Likewise, the biosynthesis of L-homophenylalanine in watercress plants is proposed to follow a pathway analogous to leucine biosynthesis,[162] beginning with the condensation of acetyl-CoA and phenylpyruvate to form benzylmalic acid. In order to find putative 2-(4-hydroxybenzyl)-malate synthase (HBMS) that may be involved in L-homotyrosine biogenesis, we searched the *E. rugulosa* genome for genes encoding the functionally analogous isopropyl-malate synthase (IPMS), the enzyme that catalyzes the condensation of α-ketovalerate with acetyl-CoA in leucine biosynthesis. We used the IPMS gene from *Mycobacterium tuberculosis*

Table 2.5: Genes found within the *hty* cluster

Gene	Length (aa)	Conserved Domain/Function	Nearest BLAST hit (Identity, Similarity)
htyF	683	heme-dependent P450 oxygenase	*T. stipitatus* TSTA_09270 (45%, 64%)
htyE	329	non-heme iron, α-ketoglutarate dependent dioxygenase	*A. terreus* ATEG09098 (50%, 66%),
htyA	584	isopropyl malate synthase	*A. alternata* BAI44742 (51%, 68%)
htyB	379	transaminase	*A. alternata* BAI44740 (55%, 71%),
htyC	366	isopropyl malate dehydrogenase	*A. alternata* BAI44741 (63%, 73%)
htyD	877	aconitase	*A. alternata* BAI44743 (57%, 68%),

(accession number: MT3813) as BLAST query. Two significant hits were found, ErIPMS48 and

ErIPMS66, both of which are found outside the contig containing the *ecd* gene cluster.

ErIPMS48 shares 99% protein sequence identity to the predicted *A. nidulans* A4 housekeeping

IPMS (AN0804) as well as >85% identity to predicted IPMS in other ascomycetes, suggesting

that this gene encodes for the actual IPMS involved in leucine biosynthesis in *E. rugulosa*. On

the other hand, the ErIPMS66 protein sequence has a lower similarity to AN0804 (43% protein

identity), while such a second IPMS homolog is not present in *A. nidulans* A4. Downstream of

ErIPMS66 (designated as *htyA*) are genes putatively encoding a transaminase (*htyB*), a 3-

isopropyl-malate dehydrogenase homolog (*htyC*), and an isopropyl-malate isomerase homolog

(*htyD*), all possibly involved in biosynthesis of L-homotyrosine(Table 2.4 and Figure 2.4A).

Moreover, the presence of immediately upstream genes encoding for a predicted non-heme iron,

α-ketoglutarate-dependent oxygenase (*htyE*) and cytochrome P450 oxygenase (*htyF*) are

consistent with the requirement of Cγ and Cδ hydroxylation of L-homotyrosine in **1**. Thus, we

reasoned that the gene cluster, here designated as *hty*, may be responsible for the biosynthesis of

L-homotyrosine in *E. rugulosa*.

To confirm this hypothesis, we genetically disrupted the *htyA* gene in the same manner as

for construction of Δ*ecdA*. Screening of ~100 colonies yielded 3 PCR-positive mutants (Figure

A9). All three Δ*htyA* mutants lost the ability to inhibit the growth of *C. albicans* under screening

conditions (Figure A2), accompanied by the loss of production of **1** (Figure 2.4B). To chemically

complement the Δ*htyA* mutant, 0.1 mg/mL of L-homotyrosine were supplemented to the growth

media. As expected, adding free homotyrosine restored the ability of the mutant to inhibit

Candida (Figure A2), as well as the production of **1** (Figure 2.4B) to wild type levels. As a

negative control, feeding of L-homotyrosine to Δ*ecdA* I-16 mutant did not restore the production

of **1** (Figure A2). Therefore, based on whole genome sequencing, we were able to identify a separately located gene cluster that is responsible for the biosynthesis of an unnatural amino acid building block for the *ecd* pathway.

Based on the putative functions of of HtyA-D (Table 2.4), the biosynthesis of L-homotyrosine is predicted to be as follows (Figure 2.4C): 4-hydroxy-phenylpyruvate undergoes an aldol-type condensation by HtyA with the C-2 of acetyl-CoA followed by the release of CoA to form 2-(4-hydroxybenzyl)-malate. This is followed by isomerization of 2-(4- hydroxybenzyl)-malate to 3-(4-hydroxybenzyl)-malate by HtyD. Thereafter, 3-(4-hydroxybenzyl)-malate undergoes decarboxylation and oxidation to form 2-oxo-4-(4-hydroxybenzyl)-butanoic acid, coupled to reduction of NAD^+ to NADH by HtyC. The product then undergoes transamination catalyzed by HtyB to form L-homotyrosine. The closest homologs of HtyA-D is found in a four gene cassette from *Alternaria alternata* that is predicted to catalyze the formation of 2-amino-4-phenyl-valeric acid (APVA) in AM-toxin.[163] Interestingly, this suggests that the HtyA-D homologs in the AM-toxin gene cluster must perform two cycles of the α-ketoacid elongation to afford APVA.

2.4 Discussion

Echinocandins are a family of antifungal cyclic lipopeptides from ascomycetes. Through the sequencing of the genome of *E. rugulosa* NRRL 11440 and subsequent bioinformatics analysis of NRPS genes, we have identified *ecdA*, encoding a 799 kDa six module NRPS, which is confirmed by gene deletion to be required for production of **1.** The *ecd* gene cluster is flanked by two microsyntenic blocks belonging to two different chromosomes in *A. nidulans* A4. Upstream of the *ecd* cluster are a group of genes that are syntenic to genes found in Chromosome

VII of *A. nidulans* A4 while the genes found downstream of the cluster is syntenic to genes

found in *A. nidulans* A4 Chromosome V (Figure 2.5). This suggests that a chromosomal

translocation occurred when *E. rugulosa* NRRL 11440 and *A. nidulans* A4 diverged from their

common ancestor.
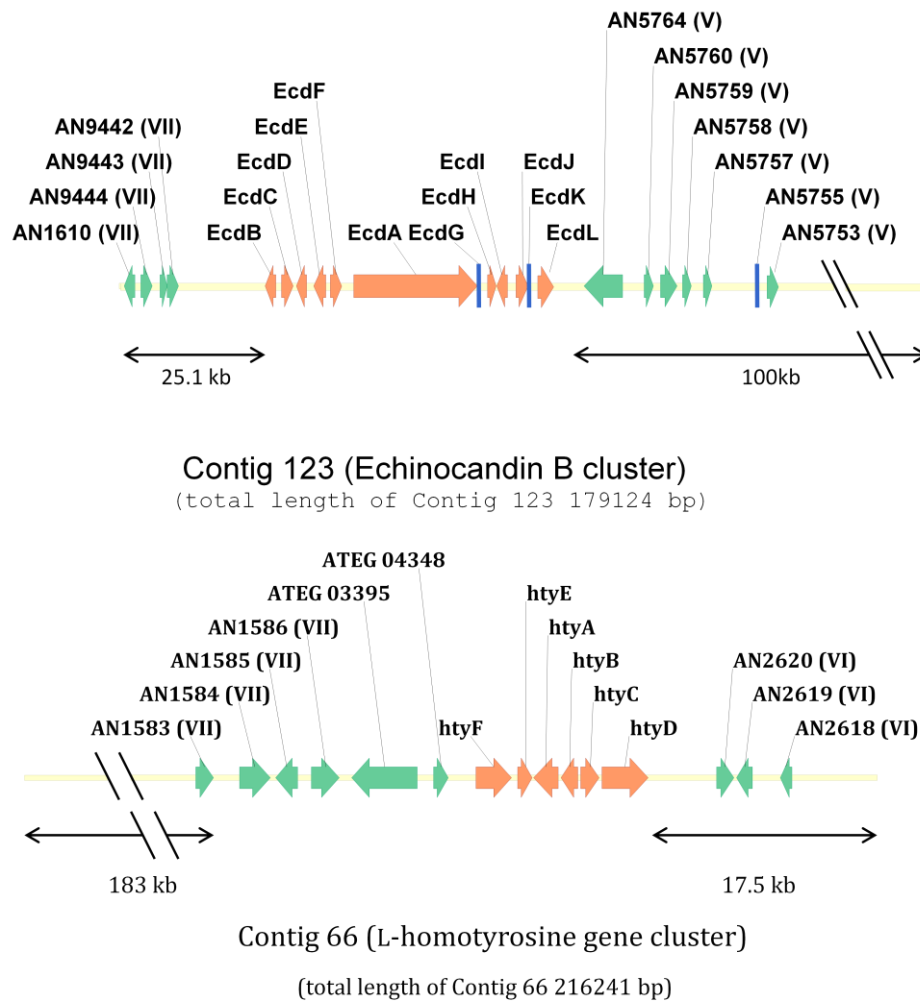
Figure 2.5: The *ecd* and *hty* cluster are found on two separate contigs (123 and 66). Based on the available genome sequence information, we cannot determine if the two gene clusters are located on the same or separate chromosomes. The shorter flanking regions of *ecd* and *hty* clusters are ~25 kb and ~17.5 kb respectively Thus, if the two gene clusters are co-located on a single chromosome on *E. rugulosa* the minimal distance between them will be ~42.5 kb. The presence of microsyntenic blocks flanking the *ecd* and *hty* cluster corresponds to different chromosomes of *A . nidulans* A4 suggests that there are chromosomal translocation events during the divergence of the two species, thus we cannot map the chromosomal location of the two cluster using *A. nidulans* as reference genome.

49

Another interesting result revealed by the study is the separation of the *ecd* gene cluster and the L-homotyrosine biosynthetic genes. By adding the distance to the nearest end of their respective contigs, the minimum distance between the two clusters is ~42.5 kb, assuming that both clusters are located in the same chromosome (Figure 2.5). Moreover, similar to the phenomenon seen in the *ecd* gene cluster, the *hty* gene cluster is flanked by microsyntenic blocks from two different *A. nidulans* A4 chromosomes; upstream of *hty* cluster are genes that are syntenic to genes from Chromosome VII of *A. nidulans* A4 while downstream are genes that are syntenic to genes found in the *A. nidulans* A4 Chromosome VI (Figure 2.5). This chromosomal translocation in the *E. rugulosa* genome, in comparison to *A. nidulans* A4 genome, also prevents us from mapping the *hty* and *ecd* gene cluster in relation to the *A. nidulans* A4 chromosomes. While separation of the biosynthetic genes for **1** is unusual, it is not unprecedented. Examples of the separation of secondary metabolite biosynthetic genes in fungi are found in the dothistromin pathway in *Dothistroma septosporum*[164], the convergence of the orsellinic acid and anthrone biosynthetic pathways in *A. nidulans* to form spiroanthrones,[165] the austinol and dehydroaustinol pathways in *A.nidulans*,[116] and the putative tryptoquivaline pathway in *A. clavatus*, which is a feature identified after comparison to the closely related tryptoquialanine pathway in *Penicillium aethiopicum.[47]*

Due to the size of the full-length EcdA, which is a formidable challenge for *in vitro* evaluation, we dissected the EcdA initiation module to establish the predicted specificity for L-ornithine activation by A_1. The ~130 kDa, four domain (T_0CAT_1) EcdA initiation module (EcdA-M1) was expressed from *E. coli*. Determination of selectivity of A_1 by amino acid-dependent ATP-[^{32}P]-PPi exchange assay established L-ornithine as the most likely substrate. Heterologous expression of EcdA-M1 and EcdI, an acyl-AMP ligase homolog, allowed us to

investigate the mechanism of lipo-initiation of EcdA. Co-incubation of EcdA-M1 with EcdI, ATP and [^{14}C]-linoleic acid indicated loading of linoleic acid to the initiation T_0 domain of EcdA. While much is known regarding the lipo-initiation strategies of the bacterial NRPS's,[151-155] this study presents the first *in vitro* characterization of the lipo-initiation of fungal NRPS. In comparison to the bacterial NRPS systems, EcdI acts analogously to DptE in daptomycin biosynthesis:[154] each adenylates the fatty acid substrate which is subsequently transferred to the initiation T_0 domain. The difference between the two systems is that the acceptor thiolation domain is standalone to the NRPS DptA in daptomycin biosynthesis but is fused at the N-terminal of EcdA for the biosynthesis of **1**. Hence it is expected that the smaller EcdI requires interaction with a thiolation domain that is part of a multimodular megasynthetases. This appears to be a commonly employed strategy among fungal PKSs and NRPSs, in which smaller enzyme partners are recruited to interact with the thiolation and acyl carrier protein domains.[95, 166] Currently, semisynthesis of echinocandins used for medical use such as **3** and **6** require the biological deacylation of their parent compound via feeding to separate cultures of *Actinoplanes* spp. followed by chemical reacylation using protective chemistry.[129, 131] Thus, deciphering the lipo-initiation strategy may also enable the engineered biosynthesis of approved derivatives of **1** containing alternative lipid groups.

The presence of acyl-AMP ligase homolog gene *ecdI* within the *ecd* gene cluster led us to investigate whether other organisms, in particular filamentous fungi, also have clustering of genes encoding an acyl-AMP ligase and an NRPS with an N-terminal T_0 domain. A search in the NCBI database revealed five genes for fatty-acyl-AMP ligase that share identity $\geq$39% with EcdI and are clustered with genes of NRPS with an initiation T_0 domain (Table S6). Moreover, four of the five EcdI homolog genes are also clustered with highly-reducing PKS genes in addition to

NRPS genes, further hinting that these gene clusters may encode for lipopeptide biosynthetic enzymes. Furthermore, *easD* from *A. nidulans,* the only characterized out of the five fatty acyl-CoA ligase genes, is shown to be involved in the biosynthesis of emericellamide. The emericellamide synthetase EasA, however, requires an additional acyltransferase EasC for lipo-initiation.[59] Such an acyltransferase is notably missing in the *ecd* gene cluster and is now proven not to be required for the linoleic acid loading in our *in vitro* assays.

Based on the results of the lipo-initiation by EcdI and determination of the selectivity of A_1, we can propose one possible pathway for the biosynthesis of **1** (Scheme 2.2): linoleoyl-AMP, produced by EcdI, is transferred to the initiation T_0 of EcdA. The linoleoyl-*S*-phosphopantetheinyl-T_0 is sequentially extended with L-ornithine, L-threonine, L-proline, L-

Scheme 2.2: Putative biosynthetic pathway of Echinocandin B (1)
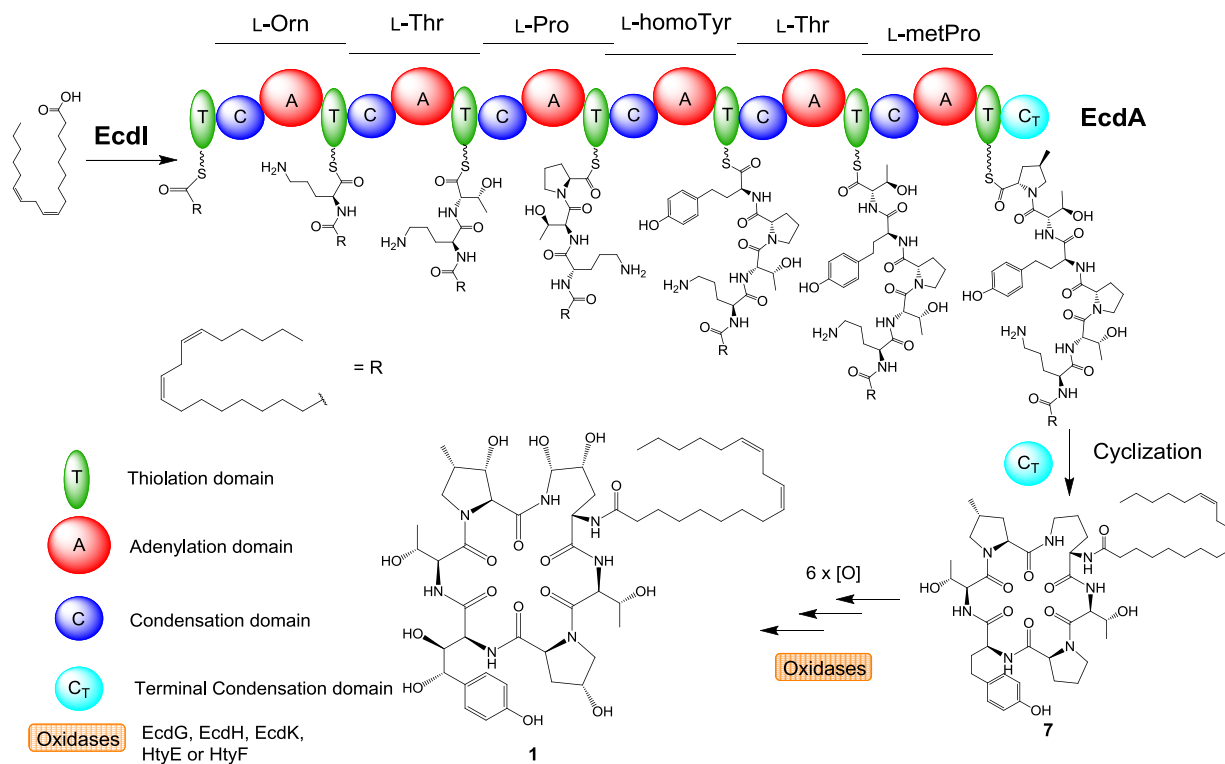
homotyrosine, L-threonine and *4S*-methyl-L-proline to form the linear hexapeptide. Thereafter, the terminal condensation (C$_T$) performs macrocyclization of the NRPS product[149] and the cyclic scaffold **7** is released from EcdA. In this pathway, in which all the hydroxylation reactions are proposed to occur following completion of the cyclic peptide, the unhydroxylated precursor **7** will undergo six rounds of hydroxylation. In congruence to modification of the residues found in **1**, five hydroxylase genes (*ecdG*, *ecdH*, *ecdK*, *htyE* and *htyF*) are embedded within the *ecd* and *hty* clusters. At this point, it is not possible to assign the hydroxylases based on sequence alone, as all are proposed to act on *sp*^{3} hybridized carbon atoms. It was previously shown that L-proline hydroxylation to 4*R*-hydroxyl-L-proline in protein scaffolds is catalyzed by an non-heme iron, α-ketoglutarate dependent oxygenase.[167] Thus it is likely that the hydroxylation of L-proline in **1** might be catalyzed by any of EcdG, EcdK or HtyE. However, the possibility that a P450 oxidase such as EcdH or HtyF can catalyze the reaction cannot be excluded. On the other hand, the formation of vicinal diols to give the 4*R*,5*R*-dihydroxyl-L-ornithine and 3*S*,4*S*-dihydroxyl-L-homotyrosine residue are more novel compared to that of the modified proline, and may each require two hydroxylases to separately install the two hydroxyl groups. Due to the lability of the resulting hemiaminal,[135] we anticipate that Cδ hydroxyl group in L-ornithine must be installed after peptide macrocycle formation, and is likely the last step in the oxidative tailoring cascade. An equally likely pathway to **1** is that some of the amino acids are hydroxylated prior to incorporation into the hexapeptide. The most plausible candidate for this scenario is *4R*-hydroxyl-L-proline, which is a commonly observed unnatural amino acid in different organisms[168-170]. The exact timing and substrate of this plethora of hydroxylation enzymes will be determined in subsequent efforts through a combination of A domain activation assays, genetic knockouts of the candidate oxygenases and *in vitro* biochemical investigation.

In addition to L-homotyrosine and L-ornithine, **1** contains the nonproteinogenic amino acid *3S*-hydroxyl-*4S*-methyl-L-proline, which is presumably derived from *4S*-methyl-L-proline. Previous studies of **4**,[161] nostopeptolide,[171] and nostocyclopeptide[172] biosyntheses showed that that 4-methyl-L-proline originates from Cδ oxidation and subsequent cyclization of L-leucine. Oxidation of L-leucine to 5-hydroxyl-L-leucine was recently identified to be catalyzed by a non-heme iron, α-ketoglutarate dependent oxygenase in *Nostoc punctiforme*.[173] Thus, it is probable that one of the non-heme, α-ketoglutarate dependent oxygenase such as EcdG, EcdK and HtyE can perform this reaction. However, both the *ecd* and *hty* clusters do not harbor the genes encoding enzymes involved in the reactions downstream to oxidation of L-leucine. A genome-wide search for genes for pyrroline-5-carboxylate (P5C) reductase homolog, proposed to catalyze the final step of 4-methyl-L-proline biosynthesis,[171] revealed the presence of four candidate genes in *E. rugulosa*. However, an ortholog for each of the four candidate genes are also present in *A. nidulans* A4 (Table A3) so at present it is not yet clear which, if any, is involved in synthesis of the 4S-methyl-L-proline for **1**.

In conclusion, we report the discovery of the biosynthetic gene cluster of **1**, the first such cluster for a member of the medically relevant fungal lipopeptide family of compounds. This study has also uncovered the genetic basis for the biosynthesis of the nonproteinogenic L-homotyrosine. The mechanism and timing of the hydroxylation of the residues L-homotyrosine, L-proline, *4S*-methyl-L-proline and L-ornithine, are intriguing features of the biosynthesis of **1** and are currently under investigation in our groups.

Section 3: Characterization of the Mononuclear Iron-dependent Oxygenases EcdG, H and K in the Biosynthesis of Echinocandin B

3.1 Motivation

The echinocandins are a small group of fungal lipopeptides that act to kill other fungi by targeting the 1,3-glucan synthase that builds the 1,3-glucans that are key structural components of fungal cell walls.[174, 175] Three semisynthetic versions (with the natural fatty acyl chains replaced by synthetic hydrophobic groups) of the naturally occurring pneumocandin or echinocandin scaffolds have been commercialized and approved for human clinical use as fungicidal and fungistatic agents for the treatment of candidiasis and aspergillosis, respectively.[132, 176]

For naturally occurring echinocandin B (**1**, Figure 3.1) the lipid is the di-olefinic linoleic acid. It has an amide linkage to a modified L-ornithine residue that is part of a cyclic hexapeptide framework (Figure 3.1). Four of the six residues in the hexapeptide framework are nonproteinogenic: only Thr_2 and Thr_5 are proteinogenic amino acids. The $4R,5R$-dihydroxy-L-Orn_1, $4R$-OH-Pro_3, $3S,4S$-dihydroxy-L-homotyrosine$_4$, and $3S$-OH-$4S$-methyl-L-proline$_6$ (3-OH-Me-Pro) residues are typically not found in proteins (although collagen has hydroxyproline residues introduced by posttranslational hydroxylation). Echinocandin C (**2**) has the C_3-OH but lacks the C_4-OH on the homoTyr residue, while echinocandin D (**3**) is a trideoxy form that has the 3-OH-homotyrosine$_4$ (3-OH-homoTyr$_4$) but a nonhydroxylated Orn_1. [177]

The unusual building block inventory suggested that echinocandins and the closely related pneumocandin are constructed by fungal nonribosomal peptide synthetase

(NRPS) assembly lines. This expectation has been recently validated by the identification of a six module NRPS biosynthetic gene *ecdA* in *A. nidulans var roseus* (recently renamed *Emericella rugulosa*),[178] whose disruption abolishes production of **1**.[179] Among the coding sequences in the vicinity of the *ecdA* gene are *ecdG, ecdH, ecdK,*
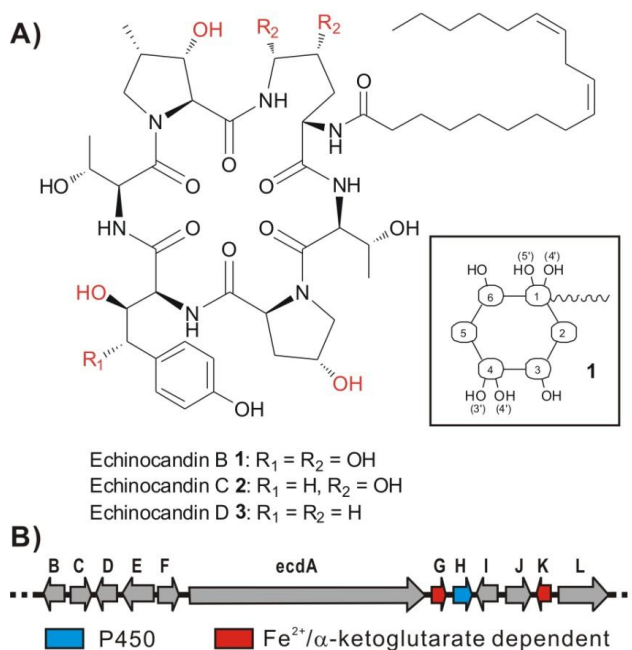
Figure 3.1: A) Structure of Echinocandin B (**1**), C (**2**), and D (**3**). A schematic representation of **1** used in Figures 3.5 and 3.6 is shown in the inset. B) Proposed nonheme mononuclear iron oxygenase and P450 type heme genes in *ecd* gene clusters.

two of which are predicted to be α-ketoglutarate-dependent mononuclear, nonheme iron oxygenases genes (*ecdG*, *ecdK*), with the third predicted to be a P450 type of hemeprotein monooxygenase gene (*ecdH*) (Figure 3.1). Given that all six of the echinocandin B amino acid residues have at least one side chain hydroxyl group (Orn$_1$ and homoTyr$_4$ each have two such –OH groups), it is likely that these oxygenases may fulfill tailoring functions at some stage to create the hydrophilic surface of the mature echinocandin B antifungal agent. In addition, it is likely that the Me-Pro framework at

56

residue 6 arises from the proteinogenic amino acid building block L-leucine by hydroxylation at C_5,[180] suggesting another, albeit cryptic, step for hydroxylase action in echinocandin producers.

To address the nature of the involvement of *ecdG*, *ecdH*, and *ecdK* in the maturation of the lipopeptide scaffold of echinocandin B, we report the gene disruptions in each of these three genes and analysis by mass spectrometry and NMR of the des-hydroxy forms of echinocandin that accumulate in the media in such fermentations. Studies with purified EcdG and EcdK confirm regiospecific oxygenase action on L-homoTyr and L-Leu, respectively. The results allow assessment of the specificity of each of the three oxygenases and the consequences for biological activity (antifungal action) as the nascent echinocandin scaffold is progressively matured.

3.2 Background:

3.2.1 Overall structure and active site residues of α-ketoglutarate, non-heme-iron-dependent dioxygenases. We have previously reported the presence of three α-ketoglutarate, non-heme-iron-dependent dioxygenase genes *ecdG, ecdK* and *htyE* in the combined *ecd* and *hty* clusters (see Section 2) and proposed their involvement in the biosynthesis of **1**.[179] α-Ketoglutarate (α-KG), non-heme-iron-dependent dioxygenase is a superfamily of protein found in all three domains of life that share a similar protein fold despite the low sequence similarity within the superfamily.[181-183] Members of the superfamily share a distorted jelly-roll fold (shown in blue in Fig 3.2A) motif in the protein structure that serves as the catalytic cavity for the enzyme. Within this jelly-roll fold are the two conserved histidines (His) and one conserved aspartate or glutamate

(Asp/Glu) that served as the ligands for the reactive iron center, thereby anchoring the iron atom in the active site.[181, 182, 184] Sequence alignment of EcdG, EcdK and HtyE with other more closely related members of (α-KG), non-heme-iron-dependent dioxygenase superfamily verified the presence of the three conserved residues that forms one of the faces in the octahedral iron complex (Fig 3.2B).
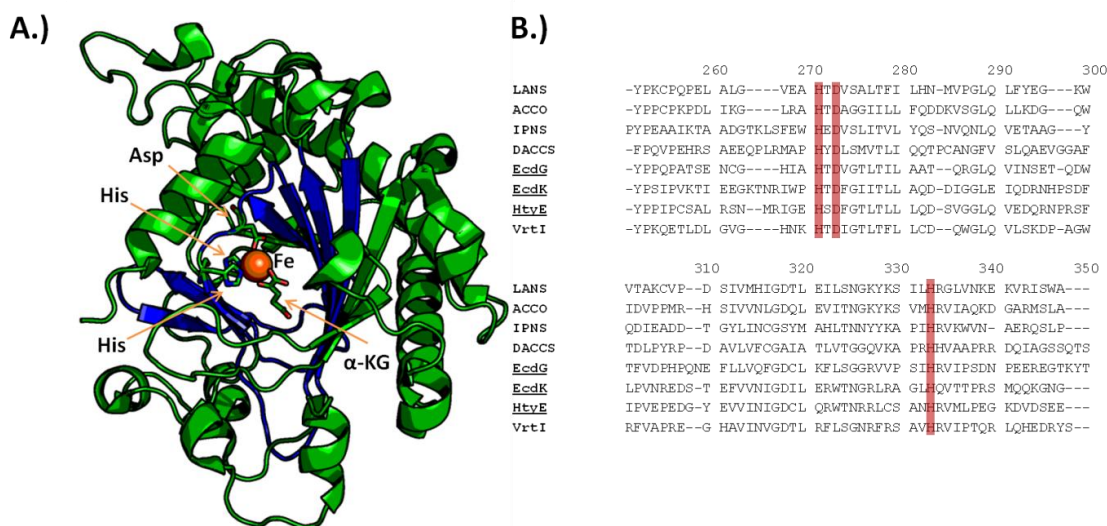
Figure 3.2: A.) Protein structure of leucoanthoncyanin synthase (**LANS**) from *Arabidopsis thaliana* (**PDB code: 1GP4**),[185] a member of α-KG, non-heme-iron-dependent dioxygenase family. Shown in blue is the distorted jelly-roll motif conserved within the superfamily. Inside the jell-roll fold is the 2-His-Asp triad coordinated to iron. Also shown is α-KG, which acts as a cosubstrate for the reaction. **B.)** Alignment of protein sequences of characterized α-KG, non-heme-iron-dependent dioxygenases LANS,1-aminocyclopropane-1-carboxylate oxidase 1 (**ACCO**) from *Petunia hybrida*, isopenicillin N synthase (**IPNS**) from *Aspergillus nidulans*, deacetocephalosporin C synthase (**DACCS**) from *Streptomyces clavuligerus* with **EcdG**, **EcdK** and **HtyE**. Highlighted in red are the conserved residues comprising the 2-His-Asp facial triad.

Whereas the porphyrin ring and the proximal cysteine or histidine occupy five of the six possible coordination site for heme-iron-dependent oxygenases, α-KG, non-heme–

iron-dependent dioxygenases have three vacant sites opposite of the 2His-Asp/Glu "facial triad" in the octahedral complex.[186, 187] One of the vacant coordination serves as the binding site for molecular oxygen during the catalytic cycle, as in the case of the final coordination site for heme-iron-dependent oxygenases. In prototypical non-heme–iron-dependent dioxygenases, the two remaining equatorial coordination sites of the octahedral complex serves the binding site for the cosubstrate α-ketoglutarate, which coordinates with the iron atom in a bidentate fashion.

However, members of the non-heme, mononuclear-iron–dependent dioxygenase superfamily have evolved to recognize alternative bidentate ligands other than α-ketoglutarate such as other 2-ketoacids as well as catechols as in the case of the extradiol catechol dioxygenase.[188, 189] An even more unorthodox member of the superfamily is isopenicillin N synthase (IPNS), which catalyzes the cyclization of the linear peptide amino-adipyl-cysteinyl-valine (ACV) to the bicyclic isopenicillin N. (IPNS) does not utilize a bidentate ligand but instead directly coordinates the thiol (SH) group of ACV.[182, 190]

Aside from changing the exogenous bidentate ligands, other members of the non-heme–iron-dependent dioxygenase superfamily have evolved to have alternative number of endogenous ligands coordinating the reactive iron center allowing alternative modes of oxidative chemistry for these enzymes.[186, 188, 189] For instance, the halogenase SyrB involved in the syringomycin biosynthesis lacks the Asp/Glu residue in the "facial triad" and instead replaces it with an exogenous chloride anion.[187, 191, 192] This allows the chloride anion to participate in the catalytic cycle, where it forms a C-Cl bond with the radical resulting from the hydrogen abstraction step (see mechanism below). On the other

59

side of the coin is the decreased number of available coordination sites found in tetrahyropterin-dependent oxygenases, where the glutamate in the facial triad coordinates to the iron atom in a bidentate fashion and utilizes tetrahydropterin as the reducing equivalent during the catalytic cycle.[189, 193]

3.2.2 Mechanism of α-ketoglutarate(α-KG)- non-heme-iron-dependent dioxygenase.

Scheme 3.1 shows the reaction mechanism of proline hydroxylase during the post-translational modification of procollagen to collagen, a quintessential example of hydroxylation reaction catalyzed by α-KG-non-heme dependent oxygenase.[188, 194] The catalytic cycle begins in the resting state of the enzyme, where the iron (II) center is coordinated by the 2-His-Asp/Glu triad that forms one face of the octahedral complex

Scheme 3.1: Mechanism of α-ketoglutarate(α-KG)- non-heme-iron-dependent dioxygenase
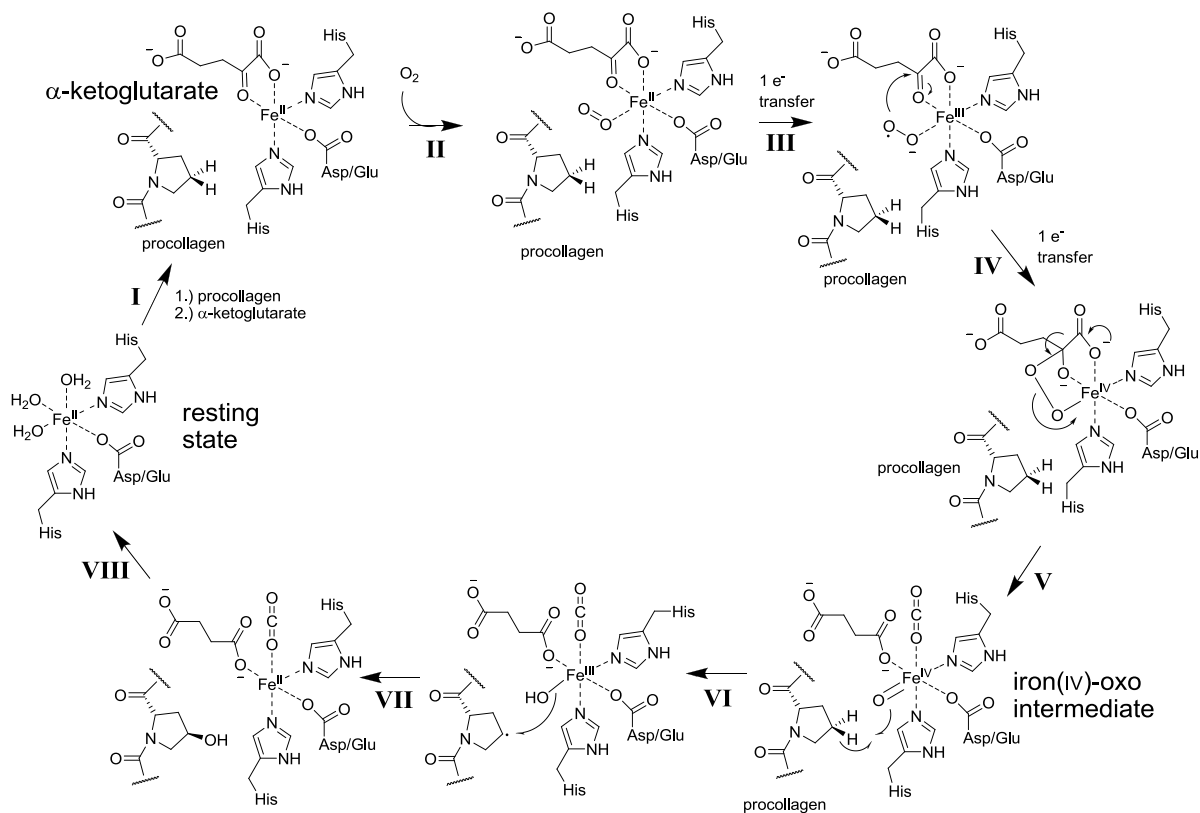
and three water molecules. Release of the three water molecules in the octahedral complex then allows binding of the substrate procollagen and co–substrate 2-oxoglutarate (α-KG), the latter coordinating with the iron atom in a bidentate fashion (**I**).[184, 194] Replacement of the neutral water ligand with the anionic α-KG causes the Fe(II)/Fe(III) redox potential to decrease thereby facilitating the oxidation of the iron center by dioxygen (**II**).[195] Following electron transfer from iron (II) to molecular oxygen (**III**), the emergent Fe(III)-[O_2]$^{\bullet-}$ species attacks α-KG and , following a second electron transfer step (**IV**), results in the oxidative cleavage of peroxide intermediate to carbon dioxide and succinate (**V**).[182, 188] During the oxidative cleavage step, the highly reactive oxoiron(IV) species is formed.[187] This high-valent oxo-iron species was previously detected via stopped-flow (SF) absorption and freeze-quench (FQ) Mössbauer spectroscopies and was proven to be the species responsible for the hydrogen abstraction in the substrate (**VI**).[196-199] This hydrogen abstraction mechanism was proven experimentally by the high observed deuterium kinetic isotope effect (KIE) on the decay of the oxoiron (IV) monitored by Mössbauer experiments.[200] Following formation of the radical on the proline substrate, the intermediate undergoes an oxygen rebound (**VII**) to form the hydroxylated product and restoration of the redox state of iron to Fe(II). Replacement of the exogenous ligands with water brings the enzyme back to the resting state.

3.2.3 Overall structure and conserved residues for CYPs. The two remaining oxidoreductases in the combined *ecd* and *hty*, annotated as EcdH and HtyF belong to the cytochrome P450 family. While there is low shared sequence identity (10-30%) among cytochrome P450, members of the family nevertheless share an overall fold as
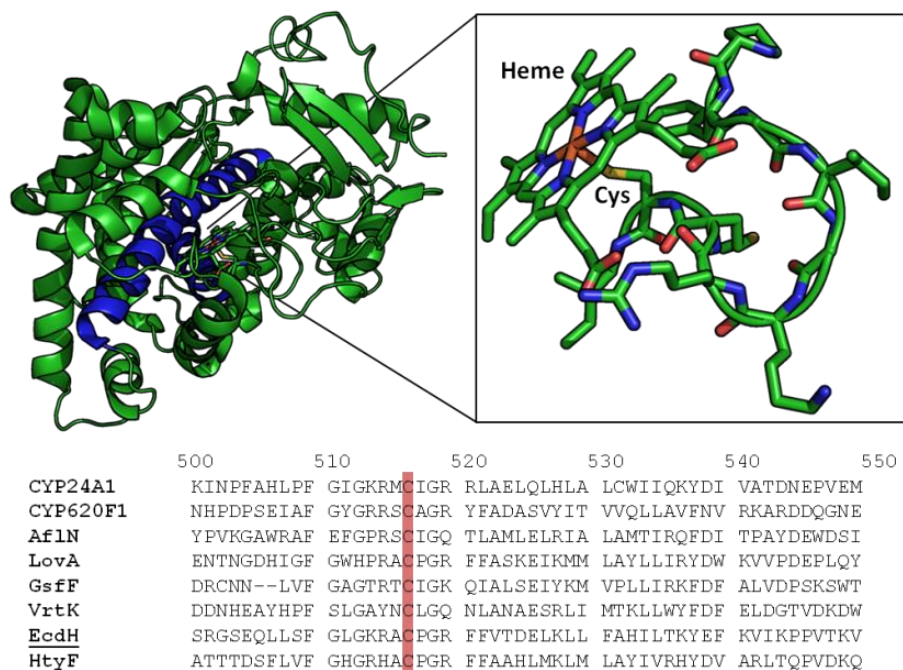
Figure 3.3: Structure of mitochondrial 1, 25-dihydroxyvitamin D_3 24-hydroxylase (**CYP24A1**, PDB code:**3K9Y**) showing the characteristic helix-rich fold typical of P450s Helix I (foreground) and helix L (background) are shown in blue. Shown in the inset is the heme prosthetic group and the proximal thiol ligand from the cysteine residue in the β-bulge. This cysteine residue is conserved among the P450sas shown for **EcdH** and **HtyF** from *E. rugulosa*, **GsfF** and **VrtK** from *P. aethiopicum*, **LovA** from *A. terreus* and **AflN** from *A. parasiticus.*

exemplified by a number of P450 structures from the three dominions of life obtained through x-ray crystallography. [201, 202] Characteristic in the family is the helix-rich tertiary structure with the helix I and L (shown in blue in Fig 3.3) nearest to the heme prosthetic group and therefore the most conserved in sequence among the cytochrome P450 family. [201, 203] Surrounded by the helices is the β-bulge (shown in inset) where the conserved cysteine residue is situated. [203] The sidechain thiolate of this residue serves as the proximal axial ligand for the iron atom of the heme group and modulates the reactivity of the heme iron[204] as shown by the mutation of the proximal cysteine to

serine in the rabbit liver microsomal cytochrome P450 2B4. [205] The C436S mutant of 2B4 has no detectable monooxygenase activity but can catalyze formation of H_2O_2 using NADPH as electron donor. Further spectroscopic characterization of the 2B4 variant revealed that the alcohol of the proximal serine remains in neutral (ROH) form when O_2 occupies the final iron coordination site, in contrast to the sidechain of cysteine which is in thiolate form [206]. It was proposed that the thiolate ligand contributes electron density to the heme causing the scission of the O-O bond to become more favorable, a supposition supported by the formation of Compound 0 (*vide infra*) and H_2O_2 in the C436S mutant of 2B4.[205-207] Interestingly,a Cys to Ser mutation in the proximal ligand of the $P450_{BM3}$ variant increased the efficiency of carbene transfer to olefin substrates by the engineered P450.[208] Recent evidence also suggests that the thiolate proximal ligand also elevates the pK_A and lowers the redox potential of the heme iron and prevents deleterious auto-oxidation of tyrosine and tryptophan residues within the active site.[209]

Moreover, the other residues found in the β-bulge, especially ones in close proximity to the conserved cysteine, also contribute to fine-tuning the reactivity of the heme iron. In particular, the hydrogen bonding interaction of the α-amide of the cysteine backbone with the backbone carbonyl in the N-terminus of the β-bulge creates a suitable environment for maintaining the higher than expected redox potential for the heme iron.[203]

In addition to core cytochrome P450 fold, eukaryotic CYPs also contains an N-terminal membrane anchoring domain.[203] Unlike bacterial CYPs, a vast majority of eukaryotic cytochrome P450s are membrane-bound, in particular in the endoplasmic

reticulum membrane. The membrane anchoring domain is composed of one or two helices that anchor the enzyme in the endoplasmic reticulum membrane. The 20 amino acid transmembrane helix (TMH), which corresponds to 3 nm long structure, spans across the bilipid membrane and is connected to the catalytic portion of the P450 by a ten-amino acid long linker composed of basic amino acids. Recent elucidation of the structure of the full-length *Saccharomyces cerevisiae* CYP51A1 (PDB code: 4KOF) showed evidence that in addition to the TMH, an additional N-terminal amphipathic helix was found in the membrane anchoring domain. The helix is situated in the distal leaflet of the bilipid membrane. The hydrophobic portion of the amphipathic helix faces the interior of the membrane while the hydrophilic portion interacts with the polar head groups of the distal leaflet of the membrane.

3.2.4 Mechanism of Cytochrome P450-type oxidase. Shown in Scheme 3.2 are the reaction mechanisms for the hydroxylation catalyzed by cytochrome P450.[210, 211] In the enzyme resting state, the, all six coordination sites for heme iron is occupied; four by nitrogen atoms from the porphyrin ring, one by the proximal cysteine thiol and one in the axial coordination site by an exogenous water ligand. Upon binding of the substrate R-H (**I**), axial water is removed from the coordination complex, causing the iron atom to move towards the proximal cysteine thiol and out of plane of the porphyrin ring. This conformational change in the heme complex causes a weakening of the ligand field and switching of heme to a high-spin complex. At this state, the first electron transfer step (**II**), becomes more favorable since the high-spin heme iron complex contains more unpaired electrons (five) compared to the low-spin complex (one). Following the first electron transfer step (**II**), molecular oxygen binds to the heme iron via the axial ligand

site (**III**).Oxygen binding is followed by a second electron transfer step (**IV**) and protonation of the ferric-peroxy anion to form **Compound 0** (**V**).

Scheme 3.2: Mechanism of hydroxylation by cytochrome P450
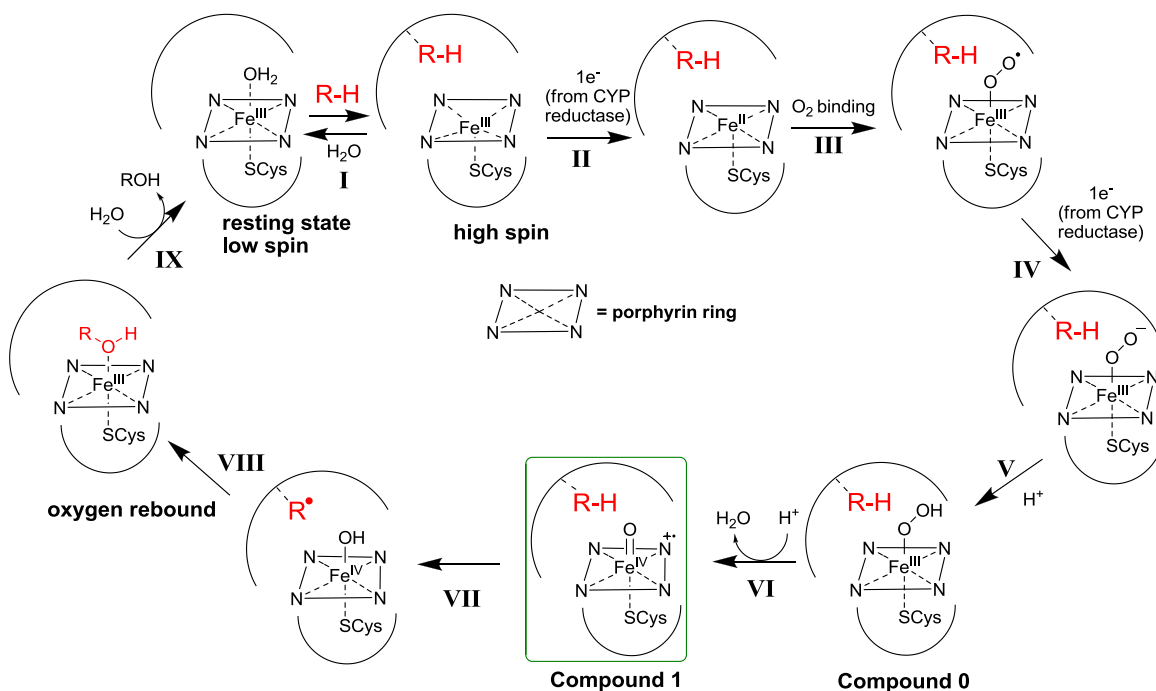
A second protonation step occurs on the nascent Compound 0 leading to the scission of the O-O bond and release of water (**VI**); resulting from this scission of the O-O bond in Compound 0 is the formation of **Compound 1**. As in the case of α-KG, non-heme-iron dependent dioxygenases, this high valent iron (IV)-oxo intermediate serves as the reactive species competent to perform the hydrogen abstraction step in the substrate R-H, which was proven experimentally fairly recently.[212, 213] The same study also was able to characterize by Mössbauer, electron paramagnetic resonance (EPR) and ultraviolet-visible wavelength (UV-Vis) spectroscopies this elusive but crucial

intermediate in heme-iron-dependent enzymes by stopped-flow reaction of purified cytochrome P450 with *m*-chloroperoxybenzoic acid (*m*-CPBA).

While the cytochrome P450 Compound I is reminiscent of the iron(IV)-oxo intermediate of α-KG, non-heme-iron dependent dioxygenases and performs an analogous role in the catalytic cycles of the respective classes of enzymes; distinct in Compound I is the relatively more direct participation of the porphyrin ligand in the formation of the reactive intermediate. Whereas the oxidation state of the 2-His-Asp/Glu triad remains constant throughout the catalytic cycle of α-KG, non-heme-iron dependent dioxygenases, the porphyrin ligand aids in the formation of Compound I by donating an electron to iron. The resulting radical ligand is stabilized due to the highly conjugated structure of the porphyrin ring.

Following the hydrogen abstraction step (**VII**), the radical intermediate undergoes oxygen rebound to form the final hydroxylated product concurrent with the regeneration of the original oxidation state of the heme iron (Fe(III)). Replacement of the product with a water molecule in the axial coordination site (**IX**) brings the enzyme back to the resting state.

3.2.5 Structure and Function of Cytochrome P450 Reductase (CPR). The catalytic cycle of cytochrome P450 requires two electron transfer steps (i) to enable conversion of the Fe(III) heme complex to Fe(II) (**II**) and (ii) to convert the iron(III)-superoxide complex to the peroxy-iron(III) intermediate (**IV**). While there are currently ten classes of cytochrome P450 classified based on the nature of their reductive partner, majority of the characterized CYPs are members of two classes (class I and class II) and hence make up

the paradigm of the CYP redox systems [214, 215]. Class I P450 systems are comprised of majority of known bacterial cytochrome P450 systems as well as eukaryotic mitochondrial P450 systems and will be only briefly discussed here. Class II P450 systems, the class in which EcdH and HtyF are a member of, are the most common system found in eukaryotes and are capable of performing diverse array of chemical transformation within the cell.
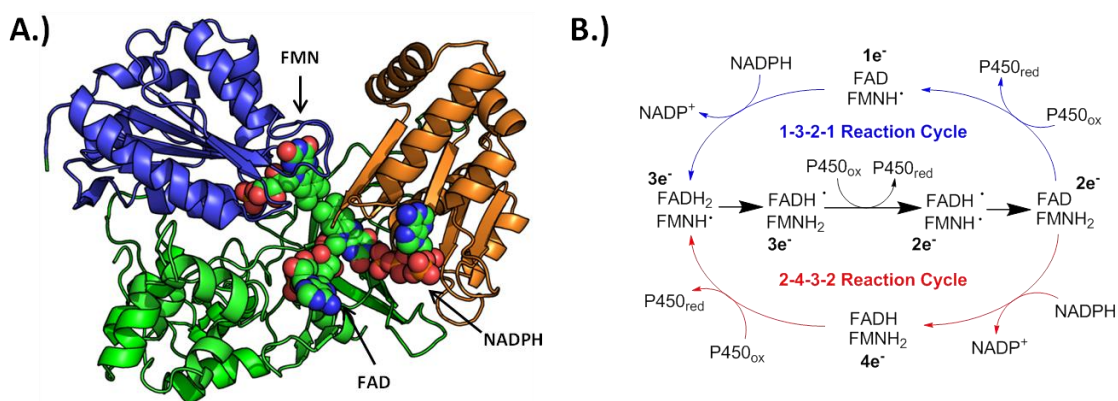
Figure 3.4 A.) Structure of *Rattus norvegicus* cytochrome P450 reductase (CPR) (**PDB code: 1AMO**) showing its three cofactors and the respective cofactor binding domains: flavin mononucleotide phosphate (FMN), flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide phosphate (NADPH). **B.)** The two models for the electron transfer steps in the catalytic cycle of the CYPs: the 1-3-2-1 reaction cycle (upper half, in blue) and the 2-4-3-2 reaction cycle (lower half, in red). Both cycles utilize the middle pathway (in black).

Class I cytochrome P450 systems are composed of three dissociated proteins: flvain-dependent ferrodoxin:NAP$^+$ reductase (FNR), ferredoxin (Fdx), and the cytochrome P450 [214]. The FNR takes a hydride from NADPH as a two-electron source and transfers the electrons in a step-wise manner to ferrodoxin, a small protein that contains an iron-sulfur cluster that serves as an electron storage complex. The ferrodoxin

in turn transfers the electrons to the heme prosthetic group of CYP during the catalytic cycle.

Class II cytochrome P450 systems, in contrast to Class I systems, are two-component systems made up of the diflavin-dependent cytochrome P450 reductase (CPR) and the cytochrome P450[214]. Based on DNA sequence homology as well as the fold of the individual domains of the *Rattus norvegicus* (Norway rat) CPR (PDB code: **1AMO**), the CPR evolved via the fusion of two ancestral proteins with homology to flavodoxin and ferrodoxin-NADP$^+$ reductase (FNR)[215, 216]. The N-terminal flavodoxin-like domain (in blue in Fig 3.4) serves as the binding site for the flavin mononucleotide (FMN) cofactor. On the other hand, the FNR-like domain serves as the binding site for the flavin adenine dinucleotide (FAD) cofactor (binding subdomain is shown in green in Fig 3.4) and binding site for NADPH (in orange in Fig 3.4). In addition to these two domains, the CPR also contains a hinge region joining the flavodoxin-like and the FNR-like domains and a 60-amino acid N-terminal membrane binding domain which anchors the protein in endoplasmic reticulum membrane.[215] Aside from the membrane binding domain, the CPR does not form significant interaction with the membrane, hence allowing the soluble expression of the *R. norvergicus* CPR for X-ray crystallography.

As mentioned previously, the CPR contains the two flavin cofactors FMN and FAD that are crucial to the role of the enzyme transferring electrons from NADPH to the heme iron. During the reductive half reaction of flavin, the isoalloxazine ring of the oxidized cofactor (Fl$_{ox}$) can take an equivalent of a hydride from NADPH and a proton from the environment to form the reduced form (Fl$_{red}$) (Scheme 3.3). In addition to these two manifolds, flavin can also undergo a single-electron oxidation from Fl$_{red}$ to the

semiquinone form (Fl_{sq}). Likewise, the oxidize form can accept an electron from a one-electron donor (*e.g.* Fe(II)) to form Fl_{sq}. This versatile redox chemistry of flavin allows it to take electrons from an obligate two-electron donor NADPH and transfer it in a step-wise manner to an obligate one-electron acceptor (Fe(III)).[45]

Scheme 3.3: The three redox states of flavin
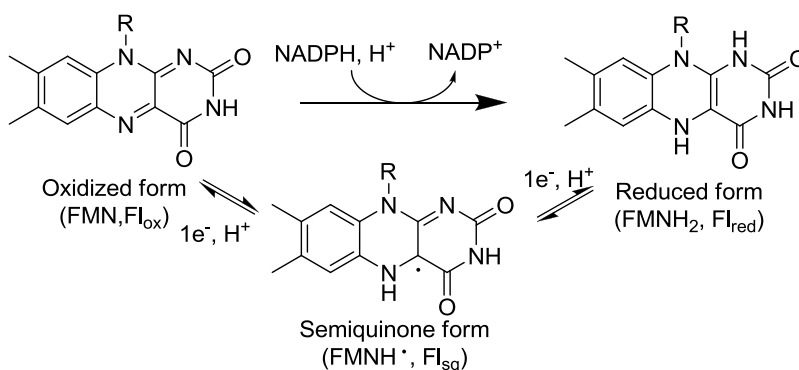
Based on the measured redox potentials of FAD and FMN in CPR and of NADPH in solution, it was proposed that the electron transfer is from NADPH to FAD to FMN and then finally to Fe(III).[215] Two electron transfer pathways were proposed (Fig 3.4B), the 1-3-2-1 cycle, where the second electron transfer step happens directly after the first and the 2-4-3-2 cycle where a hydride transfer step from NADPH occurs between the electron transfer steps to the P450. While the 1-3-2-1 cycle is believed to be the dominant pathway for P450 reduction, the 2-4-3-2 cycle is believed to be possible depending on the $NADPH/NADP^+$ ratio.

Having discussed the mechanism and enzymology of the mononuclear iron-dependent oxygenase, this section will present the functional elucidation of the enzymes EcdG, H and K within the biosynthesis of echinocandin B.

3.3 Materials and Methods

3.3.1 Gene knock-out in *Emericella rugulosa.* Procedure for the gene knockout for *E. rugulosa* is based on a protocol established for *A. nidulans* A4 with modification.[12] The linear knockout cassette was constructed as done in previous studies [12] except that the glufosinate resistance gene driven by the *trpC* promoter was used as selection marker.

Transformation of *E. rugulosa* was done as follows: *E. rugulosa* conidiospores were inoculated to 250 mL of glucose minimal media with 10 mM ammonium tartrate as nitrogen source (GMMT)[8] at a concentration of 10^7 conidiospores/mL. The culture was shaken at 28 °C for 16 hrs and 250 rpm. Thereafter, 2 g of germlings was digested at 28 °C while shaking at 100 rpm with 3 g of Vinotaste Enzyme Mixture (Novozyme) in 8 mL GMMT supplemented with 8 mL of 1.2 M KCl and 50 mM citric acid (pH 5.5). After 3 hrs, the digestion mixture was overlaid on top of an equal volume of 1.2 M sucrose and centrifuged at 1800 g and 4 °C. The protoplasts trapped on top of the sucrose cushion were isolated, washed with 0.6 M KCl three times and resuspended in 500 μL 0.6 M KCl and 50 mM $CaCl_2$ solution. Ten micrograms of the linear knockout cassette was added to 100 μL of resuspended protoplasts and the mixture was briefly vortexed for 3 seconds. 50 μL of PEG solution (0.6 M KCl, 50 mM $CaCl_2$, 75 mM polyethylene glycol with average molecular weight of 3250 g/mol, 10 mM Tris-HCl, pH 7.5) was added to the protoplast and the resulting mixture was incubated at 4 °C for 20 minutes. Thereafter, 1 mL of PEG solution was added to the protoplast mixture and incubated at room temperature for an additional 20 minutes and subsequently plated to GMMT plates supplemented with 1.2 M sorbitol and 10 mg/mL glufosinate.

Table 3.1: Primers used for Section 3

Primer Name	Primer Sequence (5'-> 3')	Notes
bar-f	aagtaaccatgagcccagaacga	For mutant screening
bar-st-R	ctaaatctcggtgacgggca	For knockout cassette cloning and mutant screening
PtrpC-F	cgacagaagatgatattgaa	For knockout cassette cloning
ecdG-KO-P1	atctccaacggattgcatctcgca	For knockout cassette cloning and mutant screening
ecdG-KO-P2	cgacaactgacttccaagcct	For knockout cassette cloning
ecdG-KO-P3	ggtcctgcccgtcaccgagatttagcgcagcagcccatcgta	
ecdG-KO-P4	gctccttcaatatcatcttctgtcgattcacagggttataccga	
ecdG-KO-P5	ataacgagctgctcttgagga	
ecdG-KO-P6	cttggctacttgacccgtctt	For knockout cassette cloning and mutant screening
ecdH-KO-P1	ggagtcgaaataatcatggtgccata	For knockout cassette cloning and mutant screening
ecdH-KO-P2	ccagtcagacaatgttccgaagagca	For knockout cassette cloning
ecdH-KO-P3	gtcctgcccgtcaccgagatttagcacccacaagcccatcacgcaca	
ecdH-KO-P4	aaagtgctccttcaatatcatcttctgtcgttataagacgggtcaagta	
ecdH-KO-P5	tggtgcagccaactggcaagca	
ecdH-KO-P6	ccgttgacattgcattgcactgca	For knockout cassette cloning and mutant screening
ecdK-KO-P1	tgctgtatgttccgaatgcga	For knockout cassette cloning and mutant screening
ecdK-KO-P2	ctttctcgccgacagtaagct	For knockout cassette cloning
ecdK-KO-P3	ggtcctgcccgtcaccgagatttagctcgacttcagagcctga	
ecdK-KO-P4	tgctccttcaatatcatcttctgtcgtgctctggcgtatcagca	

Primer Name	Primer Sequence (5'-> 3')	Notes
ecdK-KO-P5	cgactctcgctcttcggcga	
ecdK-KO-P6	cctttacacgcctccgtca	For knockout cassette cloning and mutant screening

*ecd*G-F	ggtattgagggtcgcatggcatctatctcatacgat	For cloning of EcdG
*ecd*G-R	agaggagagttagagccttcaagcagcactaaccccgt	expression construct
ecdK-F	ggtattgagggtcgcatgtctgttctaactctcga	For cloning of EcdK
ecdK-R	agaggagagtctagtacacaatccctgtcctca	expression construct

Genomic DNA from the resulting transformants was isolated using ZR Fungal/Bacterial DNA Miniprep kit (Zymo) and was used as template for PCR screening. Primers for the PCR screening are given in Table 3.1.

3.3.2 Culturing of *Emericella rugulosa* and mutants in large scale.The media and process used for *Emericella rugulosa* growth are adapted from US Patent 4,288,549.[10] Wild-type or verified mutant *E. rugulosa* cultures were grown in solid GMMT plates (or solid GMMT plates supplemented with 10 mg/mL glufosinate in the case of the mutants) at 28 °C for 4 days. The harvested conidiospores from the plates were inoculated to 8 L of medium 2 ([2.5% (w/v) glucose (Sigma), 1% peptone (BD Biosciences), 1% (w/v) starch (Sigma), 1%(w/v) molasses, 0.4% (w/v) N-Z Amine A (Sheffield Biosciences), 0.2% (w/v) calcium carbonate (Sigma)] and was shaken at 250 rpm and 28 °C for 7 days.

3.3.3 Preparative Scale Isolation of Echinocandins for NMR analysis. The extraction process for the echinocandins was adapted from US Patent 4,288,549.[10] The whole cell culture and broth was mixed with an equal volume of methanol and stirred thoroughly for one hour at 16 °C. It was then filtered through a layer of Miracloth and adjusted to pH 4.0 by the addition of concentrated HCl. The acidified filtrate was extracted twice with equal volumes of chloroform. The chloroform extracts were combined and concentrated *in vacuo* to 5-10 mL with a dark yellow color. The solution was centrifuged and the supernatant was loaded onto a Sephadex LH20 size-exclusion column (400x40 mm). The

compounds were eluted by using 1:1 chloroform/methanol (v/v) with a rate of 1.5 mL/min. Fractions with desired compounds were combined and dried *in vacuo*. The mixture was re-suspended in methanol, centrifuged, and the supernatant was injected onto a Phenomenex Luna prep-HPLC C18 column (250 x 21.1 mm, 10 micron, part # 00G-4253-P0-AX) using a solvent gradient of 40 to100% B over 30 min (solvent A, 0.1% TFA/H2O; solvent B, 0.1% TFA/MeCN). Elution of echinocandins was monitored at 280 nm. The fractions with desired compounds were lyophilized for further analysis.

3.3.4 Culturing, feeding and extraction of *Emericella rugulosa* with deuterated amino acids for MS analysis. 10 mL *Emericella rugulosa* culture, either in medium 2 described above or GMMT, was inoculated and incubated at 28 °C without shaking. Deuterated amino acids (D_6-L-ornithine, D_7-L-proline, and D_7,^{15}N-L-tyrosine) were purchased from Cambridge Isotope Laboratories (Andover, MA). Orn and Pro for feeding studies were prepared as neutral aqueous solutions with a concentration of 60 mg /mL and sterile filtered. Tyrosine solution with a concentration of 36 mg / mL was prepared in a similar way but at pH 13 due to tyrosine's poor solubility at neutral pH. 250 μL of compound solutions were added to 10 mL *Emericella rugulosa* culture at Day 4 after inoculation, and harvested at Day 7. For compound extraction, an equal volume of methanol was added to the culture broth, mixed well, incubated at room temperature for 15 min and filtered. An equal volume of chloroform was then added to the mixture. The organic layer was separated, dried with anhydrous Na_2SO_4, and filtered. The crude extract was dried *in vacuo* and stored at -20 °C for further use.

3.3.5 LC-MS methods for natural compound profiles. All LCMS analysis was carried out on an Agilent 1200 Series HPLC coupled to an Agilent 6520 QTOF spectrometer.

The crude material was re-dissolved in 200 µL methanol and filtered through a 0.45 µM membrane. 10 µL were injected for each analytical run. Separation was achieved on a 50 x 2.0 mm Phenomenex Gemini C18 column (part # 00B-4435-B0) using a solvent gradient of 20 to 100% B over 30 min (solvent A: 0.1% formic acid in H_2O; solvent B: 0.1% formic acid in acetonitrile). Temperature was 35.0 °C.

3.3.6 Cloning of EcdG and EcdK. RNA from 7-day old wild-type *E. rugulosa* liquid static culture was isolated using Ribopure-Yeast kit (Ambion).First strand cDNA synthesis was performed with SuperScript III-First Strand Synthesis SuperMix using the RNA from wild-type *E. rugulosa* as template and oligo-dT as reverse-transcriptase primer. EcdG and EcdK cDNA were amplified using primers listed in Table 3.1 and cloned into p*ET*30-Xa/LIC vector. The resulting constructs were transformed into *E.coli* Xl1 Blue cells (Stratagene) and sent for sequencing to verify correct splicing of the cDNA.

3.3.7 Reconstitution of EcdG/EcdK and hydroxylation assays. Anaerobic reconstitution of EcdG/EcdK with Fe(II), α-ketoglutarate (α-KG), and DTT was conducted as for other nonheme Fe(II)-dependent halogenases.[217] Then the hydroxylation assay was initiated by adding 100 µM or otherwise indicated concentration of reconstituted EcdG or EcdK to a 300 µL solution containing 50 mM HEPES (pH 7.0), 2 mM α-KG, and 1 mM homoTyr (EcdG) or Leu or 5-hydroxyl-Leu (EcdK). The mixture was incubated at ambient temperature and 50 µL aliquot was quenched by addition of 50 µL acetonitrile at various time points. The solution was centrifuged at 13k rpm for 5 min and the supernatant was subjected for further use.

3.3.8 Fmoc derivatization and LC/MS analysis. To 100 μL of quenched hydroxylation assay mixture described above was added 50 μL of 200 mM sodium borate (pH 8.0) and 20 μL of 10 mM 9-fluorenylmethyl chloroformate (Fmoc-Cl). After incubation at room temperature for 5 min, 20 μl of 100 mM 1-adamantanamine was added to stop the reaction. 10 μL of the mixture were injected onto a 50 x 2.0 mm Phenomenex Gemini C18 column (part # 00B-4435-B0) using a solvent gradient of 20 to 100% B over 20 min (solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid in acetonitrile).

3.3.9 Isolation of 3-OH-homoTyrosine. 1.2 mL of the quenched EcdG hydroxylation assay mixture described above was dried *in vacuo* and re-dissolved in 200 μL ddH$_2$O. 50 μL was injected for each run to a 100 *2.1 mm Thermo Hypercarb column (5 micron, part # 35005-102130). Solvent A was composed of 20 mM perfluoropentanoic acid in water; solvent B was composed of acetonitrile without any additive. Separation was achieved by using the LC gradient 0% B for 5 min followed by 0-70% B over 30 min.

The process was monitored at 275 nm. Solutions from two major absorption peaks were collected and subjected to MS to confirm that hydroxyl-homoTyr was eluted prior to homoTyr. Hydroxyl-homoTyr from 10 runs was collected, combined, lyophilized and subjected to NMR.

3.3.10 Ortho-aminobenzaldehyde derivatization and LC/MS analysis. The hydroxylation assay containing 50 mM HEPES (pH 7.0), 2 mM α-KG, 1 mM Leu or 5-hydroxyl-Leu, and 0.2 mM reconstituted EcdK proceeded for 15 min at ambient temperature. The solution was filtered through a Microcon YM-3 (Amicon) microcentrifuge filter to remove the enzyme. The flow-through was collected, to which

75

solid ortho-aminobenzaldehyde (*o*-AB) was directly added at 0.5 mg per 100 µL, mixed well and incubated at ambient temperature for 1 hour. Afterwards, the insoluble *o*-AB was removed by filtration and the solution was subjected to LC/MS or HPLC (see below). 10 µL were injected onto a 50 x 2.0 mm Phenomenex Gemini C18 column (part # 00B-4435-B0) using a solvent gradient of 1 to 100% B over 20 min (solvent A: 0.1% formic acid in water; solvent B: 0.1% formic acid in acetonitrile).

3.3.11 Isolation of *o*-AB-derivatized 3-methyl-Δ^1-pyrroline-5-carboxylic acid (MeP5C) by HPLC. 100 µL of *o*-AB derivatized solution described above was injected for each run to a 25 cm *4.6 mm Supelco Discovery C18 HPLC column (5 micron, part # 504971). Separation was achieved by using a very mild solvent gradient of 5-8% B over 20 min (solvent A: 0.1% TFA in water; solvent B: 0.1% TFA in acetonitrile). The process was monitored at 440 nm. Solutions from the major absorption peak (the 2nd) were collected. Material from ~20 runs was combined, lyophilized and subjected to NMR.

3.3.12 Synthesis of 4*R*-methyl-proline and 4-hydroxyl-L-leucine. Following the protocol disclosed by Rapoport et al, [218] 4*R*-methyl proline was synthesized starting from commercially available materials. 5*R*-hydroxyl-L-leucine was synthesized by the known method described by Moore et al. [219]

3.3.13 Antifungal susceptibility test by broth microdilution method. The potency of Echinocandin B **1** and its analogs against *Candida albicans* was measured following a published protocol using 96-well plate.[220] MIC_{90} was designated as the lowest concentration that produced an increase of less than 10% in OD over that of the adjacent sterile control.

3.4 RESULTS
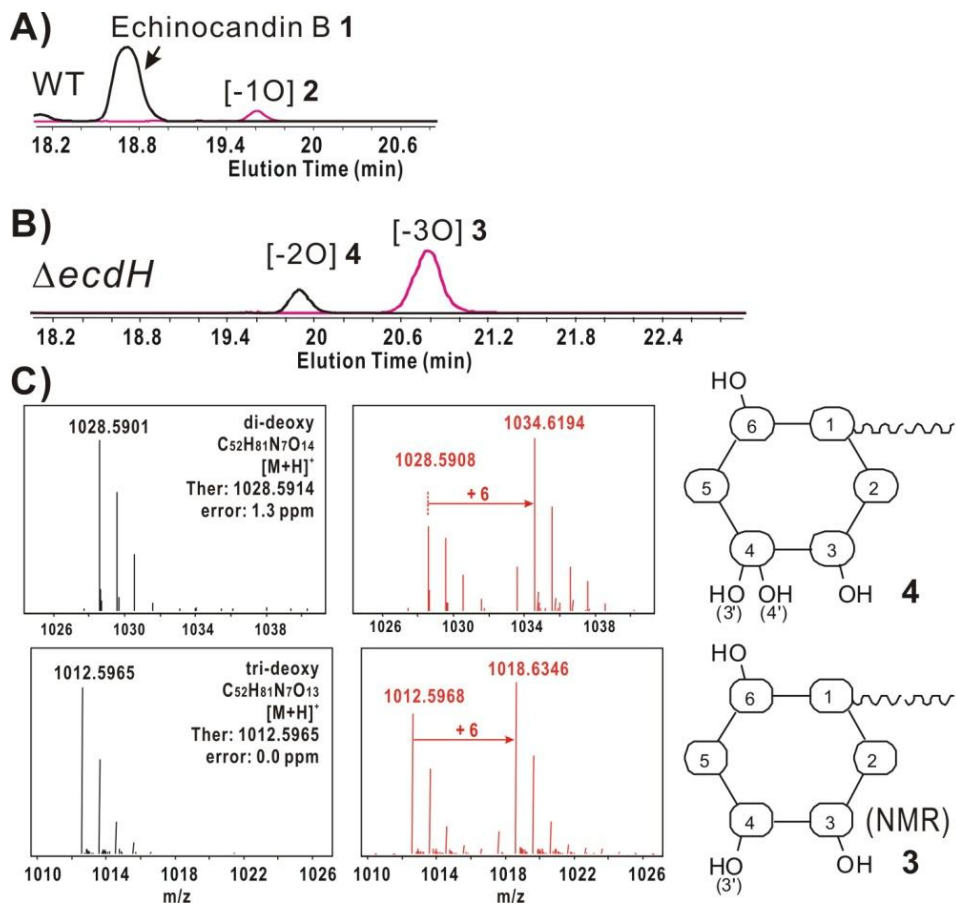
Figure 3.5: A) Extracted ion chromatogram of Echinocandin B (**1**) and C (**2**) from *Emericella rugulosa* profile. B) Extracted ion chromatogram of Echinocandin mutants from Δ*ecdH* profile. C) MS of parent ions from Echinocandin mutants produced by Δ*ecdH* (**3** & **4**), either from regular media (black) or fed with (D$_6$)-L-ornithine (red). The proposed structure of each compound is on the right.

3.4.1 Echinocandin Production from wild type *E. rugulosa*. In preparation for analysis of the echinocandin forms produced in gene deletion mutants, we first examined the forms of echinocandin generated by the native *E. rugulosa* producer under echinocandin production conditions.[221] In addition to the fully hydroxylated **1**, (*m/z* calcd for $C_{52}H_{81}N_7O_{16}$ [M+H]$^+$[-H$_2$O] 1042.5707, found 1042.5700), there is a monodeoxy form **2**

(= echinocandin C) noted previously by earlier workers [177] that corresponds to the lack of a hydroxyl group at the 4-position of the homoTyr residue (m/z calcd for $C_{52}H_{81}N_7O_{15}$ [M+H]$^+$[-H$_2$O] 1026.5758, found 1026.5757) (Figure 3.5A). Presumably, conversion of homoTyr to 3,4-dihydroxy-homoTyr, either as the free monomer or as part of a hexapeptide intermediate, is incomplete and either occurs late in maturation or does not prevent hydroxylation at other sites during maturation. The presence of **2** along with **1** suggests hydroxylation at the 3-position of homoTyr$_4$ can occur before hydroxylation at C_4.

3.4.2 Echinocandin Product Profile from a *ΔecdH Emericella rugulosa* strain:

Construction and characterization of the *ΔecdH* strain is shown in Fig S1 and is representative of the approaches used also for the *ΔecdG* and *ΔecdK* genes. Deletion mutants in each of the *ecdGHK* genes were screened for insertion of *bar* knockout cassettes followed by PCR validation of *bar* gene insertion into the desired *ecd* gene. [143, 150] Mutant cultures were grown under echinocandin production conditions,[221] compounds were extracted and evaluated by MS and NMR.

The *ΔecdH* strain loses production of **1** and **2** while yielding dideoxy (m/z calcd for $C_{52}H_{81}N_7O_{14}$ [M+H]$^+$ 1028.5914; found 1028.5901) and trideoxy forms (m/z calcd for $C_{52}H_{81}N_7O_{13}$ [M+H]$^+$ 1012.5965; found 1012.5965) of **1** (Figure 3.5B). The trideoxy compound is the major component with yields ~10 mg from 8 L culture. NMR characterization reveals the loss of hydroxyl group at 4-homoTyr$_4$ and both of the 4- and 5-OH of Orn$_1$ (Table A5). This is identical to echinocandin D, compound **3,** which is one of the naturally occurring minor echinocandins (Table A6).[138, 177]

The yield of dideoxy compound **4** from *ΔecdH* was too low for NMR assignment. We thus carried out feeding studies with deuterated amino acid monomers to evaluate the hydroxylation state at certain residues. 10 mg of 3,3-4,4-5,5-perdeutero-(D_6)-L-ornithine was used in 10 mL culture to assess the hydroxylation of the Orn_1 residue. **1** and **2** extracted from the WT strain exhibit an M+4 parent ion (Figure A15), consistent with loss of a deuterium at each of C_4 and C_5 as 4-OH, 5-OH-L-Orn is incorporated into the antibiotic framework. Variants **3** and **4** isolated from *ΔecdH* strain fed with (D_6)-L-Orn, on the other hand, both had prominent peaks at M+6, indicating that none of the Orn side chain deuterium atoms had been lost (Figure 3.5C). Therefore, compound **4** must retain all the hydroxylations except C_4 and C_5 at Orn_1.

We also tested feeding with perdeuteroproline (D_7) or perdeuteroTyr (D_7) as a parallel control. While no incorporation was observed in rich medium, in minimal medium, echinocandins were still produced and a low but detectable level of (D_6)-L-proline incorporation (loss of one deuterium at C_4 during conversion to 4-OH-Pro) was obtained in **1** from the native producer (Figure A16). Compound **3** from the *ΔecdH* strain exhibited the same M+6 parent ion (Figure A16), indicating no structural difference between them on the Pro_4 residue. On the other hand, there was still no detectable incorporation of perdeutero-Tyr. Proline is not expected to be a precursor to 3*S*-OH-4*S*-Me-L-Pro_6 since precedent suggests a route from leucine,[219] and accordingly no incorporation of deuterium from (D_6)-L- proline is found at that residue.

Thus, one can unambiguously assign the hemeprotein P450 EcdH as the Orn oxygenase acting to hydroxylate two adjacent carbons. The *ecdH* gene knockout studies are silent on two aspects of the timing of C_4 vs C_5 hydroxylation of Orn_1: (a) whether the

free amino acid undergoes hydroxylation or only the Orn residue after incorporation into the nascent peptide is oxygenated; and (b) whether C_4 or C_5 is oxygenated first. Efforts to isolate soluble EcdH from *E. coli* expression were unsuccessful (overproduction yielded insoluble inactive protein). Because EcdH may be membrane-associated, expression in yeast microsomes was conducted. Neither free L-Orn nor the trideoxy form **3** of echinocandin was detectably oxygenated on incubation with those microsomal preparations, so no conclusion can be drawn about point (a).

However, given that C_5 hydroxylation of Orn_1 in mature **1** is a key feature of the hemiaminal linkage, one expects this would not have happened before Orn was incorporated into the echinocandin peptide backbone. Otherwise a free C_5-OH-L-Orn tetrahedral addition compound would unravel to the aldehyde and in that oxidation state will not be able to form the amide bond with 3-OH-Me-Pro_6. The relative timing of C_5 vs C_4 hydroxylation can be intuited in the *ΔecdG* studies described below, from the trideoxy echinocandin **7** NMR where the C_5 of Orn has been hydroxylated but C_4 has not. In aggregate, the results suggest EcdH acts on Orn only after incorporation into the nascent echinocandin chain and acts first at C_5, then at C_4. There are precedents for a single P450 enzyme acting iteratively on its substrate. Notably, the P450 side chain cleavage enzyme in the adrenal steroid hormone pathway converts cholesterol to 22*R*-OH-cholesterol, then to the 20*R*, 22*R*-diol, before a third P450$_{scc}$ oxygenative step cleaves between the diol to yield pregnenolone and isocaproic aldehyde. [222]

3.4.3 Echinocandin Product Profile from a *ΔecdG Emericella rugulosa* strain.

The *ΔecdG* strain (Figure A17) gave a more complex echinocandin metabolite profile, producing a mixture of mono- (**5**), di- (**6**), tri- (**7**) and tetra-deoxy (**8**) compounds,

80

with **6** and **7** as the major components (Figure 3.6A). Both of them were successfully isolated (4.0 mg **6** and 5.5 mg **7** from 8 L culture) and characterized by NMR. The dideoxy form **6** (m/z calcd for $C_{52}H_{81}N_7O_{14}$ [M+H]$^+$ -[H$_2$O] 1010.5809; found 1010.5819) is missing both hydroxyl groups on 3- and 4-homoTyr$_4$ (Table A5). The trideoxy **7** (m/z calcd for $C_{52}H_{81}N_7O_{13}$ [M+H]$^+$ -[H$_2$O] 994.5859; found 994.5826), in addition to missing two -OH groups on homoTyr$_4$, also is lacking an -OH at C$_4$ of Orn$_1$ (Table A5). Consistent with NMR analysis, feeding with (D$_6$)-L-Orn to the $\Delta ecdG$ strain led to M+4 and M+5 shifts in compound **6** and **7**, respectively (Figure 3.6B).

The yield of **8** (m/z calcd for $C_{52}H_{81}N_7O_{12}$ [M+H]$^+$ 996.6016; found 996.6035) was too low for NMR analysis, however, its structure can be deduced from feeding results. It exhibits an M+6 peak upon (D$_6$)-L-Orn addition, indicating both hydroxyl groups on Orn$_1$ are absent. Together with the confirmed structure of **6** and **7**, it is reasonable to propose that Orn$_1$ and homoTyr$_4$ side chains in the tetradeoxy compound **8** are completely unmodified. Consistent with this scheme, **8** shows an M+6 pattern upon (D$_7$)-L-proline feeding (Figure A16), indicating hydroxylation on the proline moiety. As expected, **7** presents the same M+6 pattern on proline as well (the incorporation level in **5** and **6** is too low to be identified).
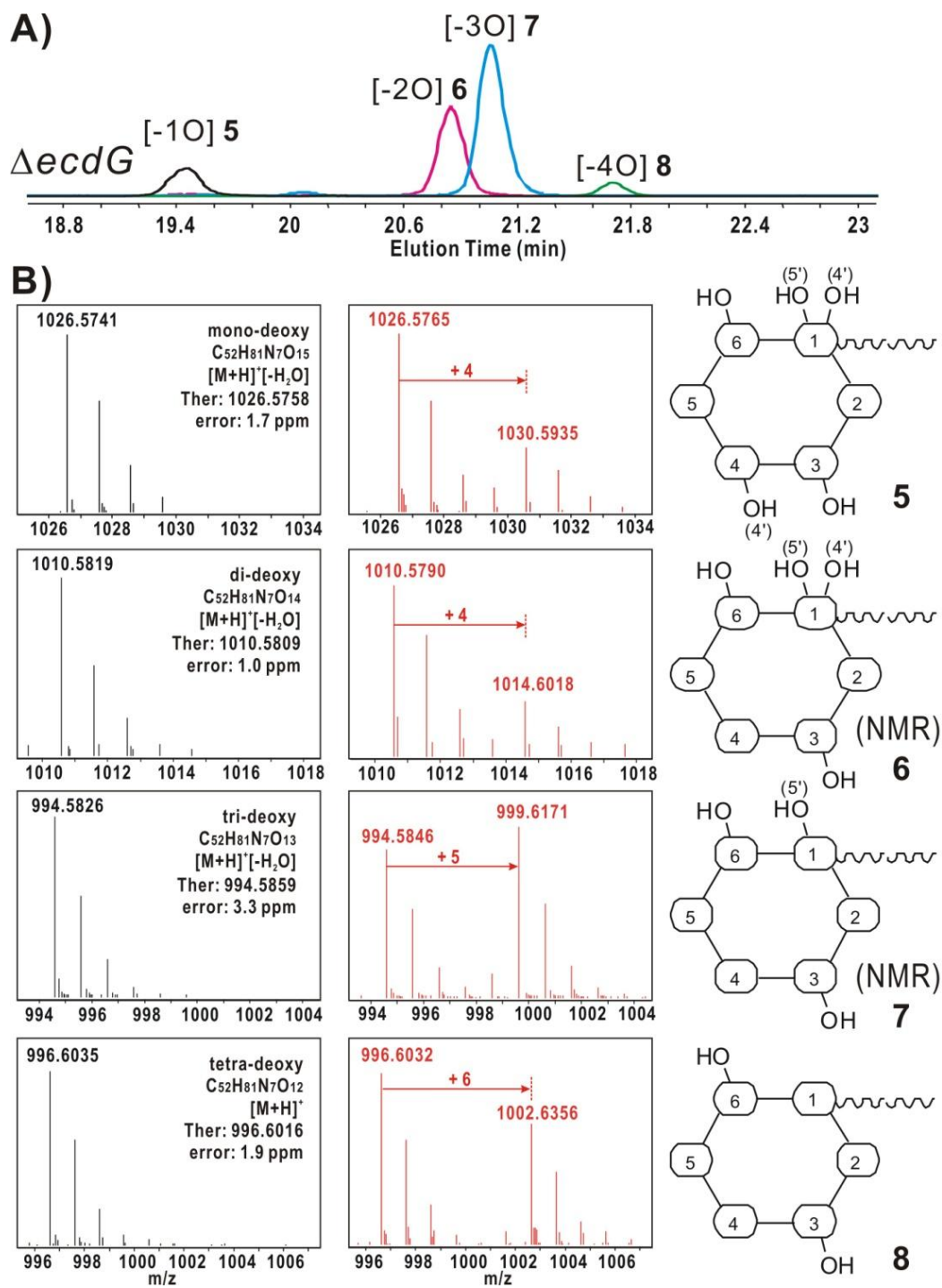
Figure 3.6: A) Extracted ion chromatogram of Echinocandin mutants from ΔecdG profile. B) MS of parent ions from Echinocandin mutants produced by ΔecdG (**5-8**), either from regular media (black) or fed with (D_6)-L-ornithine (red). The proposed structure of each compound is on the right.

3.4.4 Purified EcdG is a Homotyrosine 3-Hydroxylase. The mononuclear nonheme iron enzyme EcdG could be expressed and purified in soluble form from *E. coli* (Figure 3.7A) with a yield of ~30 mg/L. Assay with the amino acid building block L-homoTyr (Figure 3.7B) followed by Fmoc derivatization revealed that EcdG acts to introduce one hydroxyl group on free L-homoTyr (Fmoc-homoTyr: m/z calcd for $C_{25}H_{23}NO_5$ $[M+Na]^+$ 440.1468, found 440.1460; Fmoc-hydroxy-homoTyr: m/z calcd for $C_{25}H_{23}NO_6$ $[M+Na]^+$ 456.1418, found 456.1411). As has been seen with other nonheme iron oxygenases that generate high valent oxoiron intermediates,[191, 192] EcdG inactivates itself after a small number of turnovers. Variation of the amount of pure EcdG relative to excess substrate indicates a proportional amount of product formation (Figure 3.7C) and indicates that about 3-3.5 turnovers occur per inactivation event. The estimated initial rate of hydroxylation is higher than 5 catalytic events per minute.

Although only 3-3.5 oxygenation events per EcdG enzyme molecule occur before enzyme suicide, it was possible to scale up the monohydroxy homoTyr product and separate it from unreacted L-homoTyr by Hypercarb column using 20 mM perfluoropentanoic acid / acetonitrile as mobile phase (Figure 3.7D, homoTyr: m/z calcd for $C_{10}H_{13}NO_3$ $[M+H]^+$ 196.0968, found 196.0975; hydroxy-homoTyr: m/z calcd for $C_{10}H_{13}NO_4$ $[M+H]^+$ 212.0917, found 212.0925). Proton NMR and gCOSY clearly indicated the enzymatic product is a 3-OH-homoTyr (Figure A18 and A19) and correlates with the chemical shifts reported for synthetic (2*S*, 3*R*)-OH-homoTyr.[223] We have not determined the absolute stereochemical configuration at the C3-OH of the EcdG *in vitro* product but anticipate it will be 3*R*.
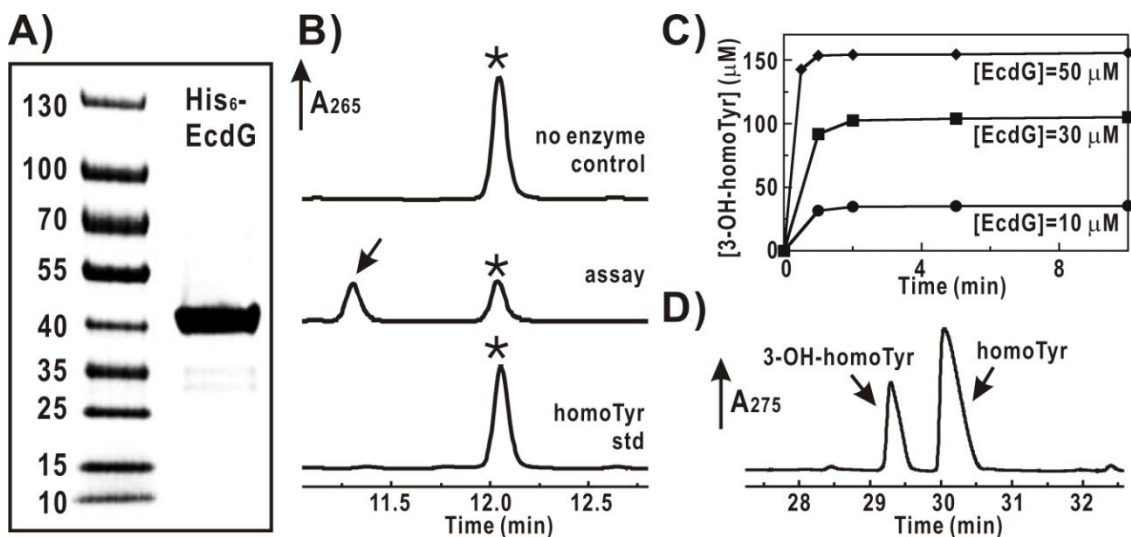
Figure 3.7: A) SDS-PAGE gel showing purified His$_6$-EcdG enzyme (expected molecular weight: 40.0 kDa). B) HPLC-UV (259 nm) traces indicate formation of hydroxy-homoTyr (arrow, middle trace) relative to negative control without enzyme (top) and homoTyrosine standard (star, bottom trace) as Fmoc-derivatives. C) Time course of formation of hydroxy-homoTyr at different enzyme concentrations. D) HPLC-UV (275 nm) trace showing isolation of hydroxyl-homoTyr from EcdG assay solutions for further NMR analysis.

Pure EcdG showed no activity with the deshydroxy echinocandin scaffold isolated from the *ΔecdG* mutant strain (**6** & **7**), confirming that it does not act after the NRPS assembly line has produced the hexapeptide framework. Moreover, EcdG will not take the 3-OH-homoTyr product and convert it on to the 3,4-dihydroxy amino acid (data not shown), making it unlikely to be the hydroxylase for introducing the 4-OH group in addition to the hydroxylation at C3. Therefore, the prediction is that the adenylation domain in module 4 of the six module EcdA would activate 3-OH-homoTyr (or 3,4-dihydroxy-homoTyr but less likely, see discussion below) for incorporation into the growing echinocandin chain.

3.4.5 Deletion of *ecdK* abolishes production of echinocandins Given the above results from *ecdG* and *ecdH* deletion strains, we anticipated that the remaining putative oxygenase, the nonheme mononuclear iron protein EcdK, could act either on Pro_3, 4R-Me-Pro_6, or both, since these are the remaining two hydroxylated nonproteinogenic residues in **1**. It was also possible that EcdK could be involved in the biosynthesis of 4-Me-Pro, which in cyanobacteria is known to involve initial hydroxylation at C_5 of L-leucine. [219] When metabolites of *ΔecdK* mutants were analyzed by LCMS, indeed, production of **1** was abolished. Feeding experiments with homoTyr or 4*R*-OH-Pro did not restore production of **1**. On the other hand, either 3*S*-OH-4*S*-Me-Pro or 4*R*-Me-Pro did restore robust echinocandin production (Figure 3.8A). These results clearly implicate EcdK in a step upstream of 4*R*-Me-Pro formation. The first step in Me-Pro construction is proposed in other systems to be hydroxylation of Leu to 5-OH-Leu (Figure 3.8B).[219] Further oxidation (either by a second hydroxylation at C_5 or via action of an alcohol oxidase) to γ-methyl-glutamic acid-γ-semialdehyde (γ-Me-Glu-γ-semialdehyde) would set up nonenzymatic cyclization to 3-methyl-Δ^1-pyrroline-5-carboxylic acid (MeP5C). Subsequent reduction of the 1-pyrroline by a nicotinamide-utilizing reductase would give 4-Me-Pro (Figure 3.8B).

3.4.6 EcdK oxygenates C_5 of L-Leucine iteratively. In order to verify the results observed in the knockout of ecdK, we sought to reconstitute *in vitro* the oxidation of L-Leu to γ-Me-Glu-γ-semialdehyde by EcdK. EcdK could be overproduced in and purified from *E. coli* as a soluble N-His$_6$-tagged enzyme (Figure 3.8C) with a yield of 33 mg/L. Assay of pure EcdK with L-leucine followed by analysis of products revealed O_2-dependent conversion to two new products. One was identified as a hydroxy-Leu by

LC/MS after Fmoc derivativation (Figure 3.8D, trace 2 & 3) (Fmoc-OH-Leu: m/z calcd for $C_{21}H_{23}NO_5$ [M+Na]$^+$ 392.1468, found 392.1465). A synthetic standard of Fmoc-5-OH-Leu, which has the same retention time, confirms the assignment of the enzymatic product (Figure 3.8D, trace 1). The second proposed product, MeP5C, could be captured by ortho-aminobenzaldehyde (o-AB)[224] to yield the characteristic A_{440} nm chromophore of the dihydroquinazolinium derivative in a pair of products (major to minor = 6/1, Figure 3.8D, trace 1 & 2) with identical mass (MeP5C: m/z calcd for $C_{13}H_{15}N_2O_2$ [M]$^+$ 231.1128, found 231.1121) and UV spectrum (Figure A21). This is consistent with a pair of diastereomers created at C_1 on capture of the C_1 imine (Figure A22). Scale up of the EcdK reaction was challenging, given that EcdK underwent suicidal inactivation during turnover (Figure 3.8F). This is characteristic of a subset of the nonheme iron oxygenases,as exemplified also by EcdG (Figure 3.7C), and most probably reflects adventitious, uncontrolled reaction of an Fe^{IV}=O high valent intermediate in the oxygenation process. From the three traces in Figure 3.8F, there are about 5-6 turnovers/inactivation event for EcdK. Even with the low turnover/inactivation ratio it was possible to use enough pure EcdK to generate sufficient dihydroquinazolinium product adduct. It was purified by HPLC (Figure A21), collected for NMR analysis of the major isomer, and the structure was validated (Figure A24-A27). In turn, this supports the scheme in Figure 3.8B that EcdK acts tandemly to oxygenate C_5 of L-Leu. The *gem* diol product is the carbonyl hydration product of γ-Me-Glu-γ-semialdehyde, in turn in equilibrium with the cyclic imine that is trappable with o-AB.
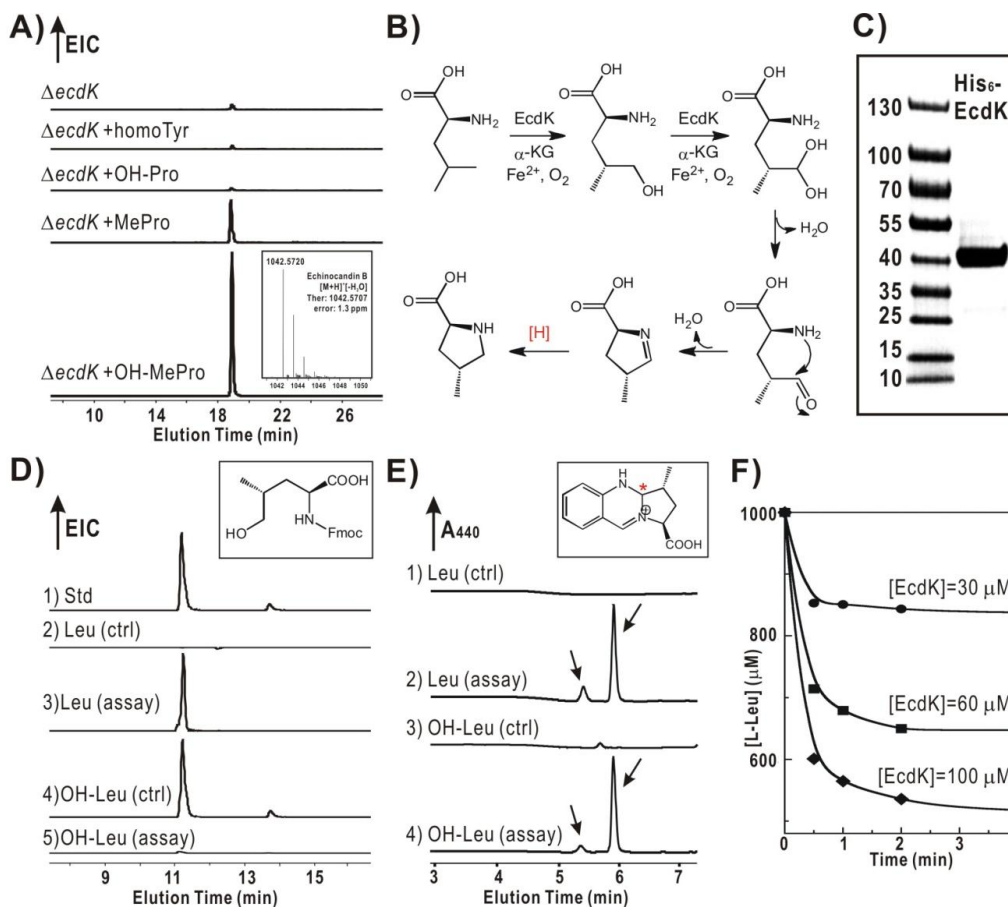
Figure 3.8: **A**) Extracted ion chromatogram of Echinocandin B (**1**) from Δ*ecdK* mutant. The culture media is supplemented with amino acid as labeled (OH-Pro = 4*R*-OH-L-Pro; Me-Pro = 4*R*-Me-L-Pro; OH-Me-Pro=3*S*-OH-4*S*-Me-L-Pro). **B**) Proposed biosynthesis pathway for 4*R*-methyl-L-proline from L-Leu by tandem hydroxylation at C_5. **C**) SDS-PAGE gel showing purified His$_6$-EcdK enzyme (expected molecular weight: 39.7 kDa). **D**) Extracted ion chromatogram of Fmoc-derivatized 5-OH-Leu (structure shown in inset). Top to bottom: 1) synthetic standard of Fmoc-5-OH-Leu; 2) assay with Leu, no enzyme; 3) assay with Leu, with enzyme; 4) assay with 5-OH-Leu, no enzyme; 5) assay with 5-OH-Leu, with enzyme, reflecting comlete consumption of 5-OH-Leu. **E**) HPLC-UV (440 nm) traces showing the formation of MeP5C after *o*-AB derivatization (derivatized product structure shown in inset). Arrows indicate a pair a diastereomers with a ratio of ~ 1:6. Top to bottom: 1) assay with Leu, no enzyme; 2) assay with Leu, with enzyme; 3) assay with 5-OH-Leu, no enzyme; 4) assay with 5-OH-Leu, with enzyme. (**F**) Time course of utilization of L-leucine at different EcdK concentrations. Residual Leu is quantified in its Fmoc-derivatized form by UV/Vis absorption.

87

Table 3.2: MIC_{90} measurement of compound echinocandin B (**1**) and its deshydroxy analogs against *Candida albicans* strain ATCC 90234.

Compound	MIC (µg/mL)
1 (wt)	0.6
3 ($\Delta ecdH$)	0.2
5 ($\Delta ecdG$)	3.1
6 ($\Delta ecdG$)	0.8
7 ($\Delta ecdG$)	1.6

To further corroborate the view that EcdK can act at C_5 of L-Leu iteratively, 5-OH-Leu was synthesized from an intermediate in Me-Pro synthesis [14]. The 5-OH-Leu served as a substrate for EcdK yielding the same imine as seen from L-Leu (Figure 3.8D, trace 4 & 5; Figure 3.8E, trace 3 & 4). The cyclic imine, the penultimate compound in biosynthesis of the cyclic amino acid Me-Pro, presumably then undergoes reduction (e.g. by an NADH-dependent oxidoreductase) to afford Me-Pro. Recently, a comparable mononuclear iron-dependent L-Leu 5-hydroxylase has been characterized from *Nostoc punctiforme*.[225] Apparently, L-Leu 5-hydroxylase from *N. punctiforme* stops after the first hydroxylation, in contrast to EcdK which performs both oxidation steps.

3.4.7 Anticandidal activities of echinocandin variants. Earlier reports noted that **2** and **3** retained similar activity against *Candida* as **1**, albeit without detailed data.[177, 226] Studies on the semisynthetic cilofungin, which has the same hexapeptide core as echinocandin but different fatty acid side chain, indicated a sixteen-fold drop in potency from the fully oxygenated hexapeptide scaffold to the tetra-deshydroxy type

scaffold.[226] It is known that the nature of the hydrophobic acyl chain in the semisynthetic variants is an important determinant of their antifungal activity and toxicity.[130, 131, 227] Given that we had access to **1** from wild type and **3** from the *ΔecdH* strain, we examined their antifungal activity directly side by side along with other deshydroxy analogs against *Candida albicans* strain ATCC 90234 with the results shown in Table 3.2. Under these assay conditions **3** is three fold more potent, while the compounds lacking the 3-hydroxy on homotyrosine (**5**-**7**) are 1.3-5.1 fold less active. Zambias et al noted from work on synthetic derivatives of cilofungin that the hydroxy groups on Pro_3 and Me-OH-Pro_6 , along with the homoTyr residue, were the most important hydroxylation sites for anticandidal activity.[226] Our results are consistent with observations on cilofungin but we do not yet see production of echinocandin variants lacking hydroxylation on Pro_3 or Me-Pro_6.

3.5 Discussion

The echinocandins undergo a cascade of hydroxylation reactions during the maturation process of the peptide scaffold. Six hydroxyl groups are distributed over the four nonproteinogenic building blocks in **1**. Two additional hydroxylations, at C_5 of L-Leu have been presumed to initiate the pathway to 4*R*-Me-Pro (as a precursor to the 3*S*-OH-4*S*-Me-Pro_6 residue). Thus, at least eight hydroxylations are required for generation of the mature echinocandin B scaffold, albeit the two hydroxyls introduced at C_5 of L-Leu are cryptic, *i.e.* not persisting in the mature echinocandin scaffold.

Among the questions we had at the outset were which of the putative three EcdGHK oxygenases, if any, act iteratively on Orn_1, and homoTyr$_4$ and which may act

89

on Pro_3 and/or $4R$-Me-Pro_6 or the L-Leu precursor to Me-Pro. The timing of side chain oxygenations in the antifungal scaffold assembly was also unknown; to wit, it is not apparent whether the hydroxylases act upon free amino acid, amino acyl- or peptidyl-S-NRPS stage, or on a deshydroxy echinocandin framework. In addition, there was ambiguity about whether hydroxylations must go in a prescribed order and whether independent routes of oxygenative maturation can go in parallel.
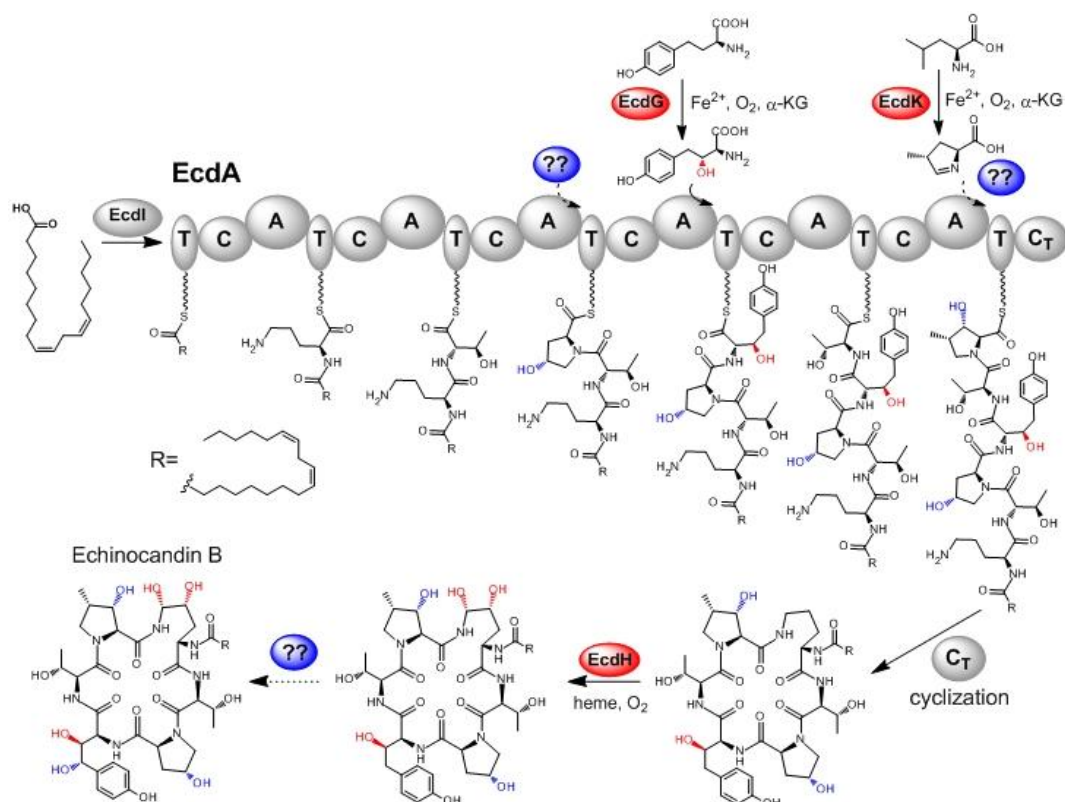
Figure 3.9: Proposed timing of post-translational hydroxylations for biosynthesis of echinocandin B (**1**). Confirmed steps and modifications are highlighted in red while proposed ones are in blue.

This study identifies five oxygenation events by these three iron enzymes. The nonheme iron EcdG works on free L-homotyrosine to generate 3-OH-L-homoTyr. This step occurs at an early stage during echinocandin biosynthesis before the building block

90

is loaded onto the hexamodular NRPS EcdA. On the other hand, the hydroxylase for C_4 of homotyrosine remains unidentified. Knockout of *ecdG* or *ecdH* significantly decreases the hydroxylation efficiency at this position in the final echinocandin scaffold, although neither enzyme directly hydroxylates it. We thus propose that the homoTyr-4-hydroxylation reaction occurs after EcdH, probably as the last step during echinocandin B biosynthesis (Figure 3.9). That placement implies this particular hydroxylase takes the partially deshydroxy macrocycle as a substrate and can produce minor compounds **4** and **5**, with a preference for generating the fully hydroxylated echinocandin B scaffold.

The hemeprotein EcdH, which iteratively works at C_5 and then C_4 of Orn_1, probably functions at a late stage after the release of the macrocyclic peptide from the NRPS EcdA, since a free C_5-OH-L-Orn would unravel to the aldehyde and not be competent to be joined to 3*S*-OH-Me-Pro_6. EcdH is sensitive to the conformational change of the hexapeptide macrocycle when particular-OH groups are missing. When *ecdG* is deleted, EcdH becomes less efficient and produces three variants at Orn_1 (compound **5-8**). The EcdH-mediated hydroxylation at C_5 of Orn_1 creates the hydrolytically labile hemiaminal linkage: a subsequent elimination of water to the cyclic imine would be on pathway to hydrolytic linearization and deactivation of the hexapeptide macrocycle. The post NRPS generation of that hemiaminal linkage is in contrast to known routes to imine linkages in other nonribosomal cyclic peptides such as nostocyclopeptide[228] and koranimine.[229] In those cases the NRPS assembly lines have C-terminal reductase domains that release the linear peptide aldehydes as nascent products. Imine formation with the N-terminal-NH_2 is the cyclization step. In contrast

EcdA releases the cyclic hexapeptide apparently via a fungal macrocyclizing C_T domain and the hemiaminal (precursor to imine) is a post assembly line maturation step.

EcdK, in analogy to EcdG, works as a nonheme mononuclear iron oxygenase at the level of a free amino acid. In this case L-Leu is substrate and undergoes tandem hydroxylation at C_5 in the first and second step in the pathway to 4R-Me-proline assembly. Thus, both EcdH and EcdK do iterative hydroxylations, one on the same carbon (EcdK), the other on adjacent carbons (EcdH) of their substrates. The proposed timing of EcdG, EcdK, and EcdH in echinocandin scaffold maturation is noted in Figure 3.9.

Three echinocandin oxygenation events are still to be characterized. To date we have not identified the hydroxylase(s) for Pro_3 (oxygenation at C_4) and 4R-Me-Pro_6 (oxygenation at C_3) and at what stage they occur. Given the absence of any deshydroxy forms of echinocandins at either of these two residues, hydroxylations on the macrocycle after release by the nonribosomal peptide synthetase EcdA is disfavored. Until the Pro/Me-Pro hydroxylases are found, we cannot tell if the free amino acids or peptidyl-S-NRPS species are substrates. Precedent does suggest conversion of Pro to 4R-OH-Pro can occur at the level of the free amino acid [230]. It may be that is also a reasonable expectation for the transformation of 4R-Me-Pro to 3S-OH-4S-Me-Pro. As noted earlier, the homotyrosine-4- hydroxylase is also not yet identified. We have noted that two genes annotated as *htyE* and *htyF* are adjacent to the *htyA-D* genes required for conversion of Tyr to homoTyr. [179] Whether either HtyE or HtyF are any of the missing three oxygenation catalysts for echinocandin B maturation is a subject for further study.

Section 4: Complexity generation in fungal polyketide biosynthesis: a spirocycle-forming P450 in the concise pathway to the antifungal drug griseofulvin

4.1 Introduction:

Griseofulvin (**1**), a polyketide produced by ascomycetes such as *Penicillium aethiopicum*,[231] is an antifungal discovered in 1940s.[232] It was the first oral antifungal drug [233-235] and has been in use for many years in medical and veterinary applications.[236] While **1** is largely superseded by newer antifungal drugs such as azoles and echinocandins, **1** has a niche in treatment of dermatophytes such as *tinea capitis* (ringworm of the scalp) and *tinea pedis* (athlete's foot). **1** recently gained medical attention due to its ability to disrupt mitotic spindle[237] and act as potential inhibitor of centrosomal clustering in tumor cells. [238]

While the utility of **1** as an antifungal drug as well as its newly discovered bioactivity show its medical relevance, **1** occupies a place in the pantheon of natural products due to its historical significance in establishing the field of study of polyketide biosynthesis.[239] Arthur Birch demonstrated the acetate origin of the carbon skeleton of **1** through the feeding of [1-^{14}C]-acetate to the griseofulvin-producing strain *Penicillium griseofulvum* Dierckx and determined the labeling pattern in **1** using degradation experiments.[240, 241] These results, together with studies on the biosynthesis of 6-methylsalicylic acid,[242] laid the foundation in proving the polyketide hypothesis.[63] Results from radioactive tracer studies using [^{14}C]-acetate was confirmed by NMR studies using singly and doubly ^{13}C-labeled acetate (Figure 4.1).[243] In addition, the

formation of the spirobicyclic grisan ring of **1** was one of the examples used by Barton

and Cohen to illustrate the prevalence of phenol oxidative coupling reactions in
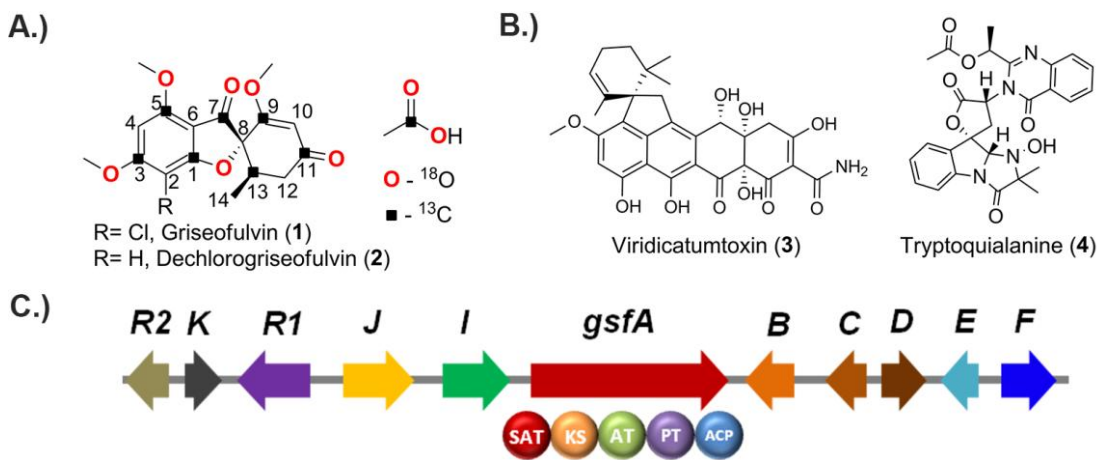
Figure 4.1: Griseofulvin (**1**) and other secondary metabolites produced by *Penicillium aethiopicum.* **A.**)
Structure of **1** and dechlorogriseofulvin (**2**) shown together with acetate origin of carbon and oxygen atoms.
B.) Viridicatumtoxin (**3**) and tryptoquialanine (**4**) are the other chemotaxonomic marker in *P. aethiopicum*;
and **C.**) Arrangement of genes in the biosynthetic gene cluster of **1** in *P. aethiopicum.*

Nature.[244]   Subsequent studies using feeding of the radiolabeled benzophenone

intermediates griseopheone B (**10**) and C (**11**) implicated their role in the biosynthesis of

1 and supported the hypothesis that **1** undergoes phenol oxidative coupling to form the

grisan ring moiety.[245-247]

Although labeling and feeding studies opened the path to elucidating the timing of

the tailoring reactions, the genes and the enzymes in the biosynthetic pathway of **1** have

not been discovered until recently when we sequenced the genome of *P.*

aethiopicum.[103] In addition to **1** and the related metabolite dechlorogriseofulvin (**2**), *P.*

aethiopicum also produces viridatumtoxin (**3**) and tryptoquialanine (**4**), two other

secondary metabolites with spirobicyclic moiety (Figure 4.1B).[47, 103, 231] Bioinformatic mining of the *P. aethiopicum* genome led to the discovery of the putative *gsf* cluster (Figure 4.1C), which contains *gsfA*, a gene for a non-reducing polyketide (NR-PKS) and genes for tailoring enzymes that appear consistent with the modifications required to form **1**. Gene deletion of *gsfA* and flavin-dependent halogenase gene *gsfI* confirmed the association between *gsf* cluster and biosynthesis of **1**.[103] However, the biochemical properties of the NR-PKS, which synthesizes a polyketide backbone that is not typical among the characterized aromatic polyketides; and the enzymatic steps leading to the maturation of the griseofulvin scaffold structure, highlighted by the spirocyclic grisan ring, have not been studied. In this work, we present the complete genetic and biochemical characterization of the *gsf* pathway. These insights led to the in vitro synthesis of **1** from malonyl-CoA using purified *gsf* enzymes and cofactors.

4.2 Materials and Methods

4.2.1 Strains and Culture Conditions. *P. aethiopicum* IBT 5753 was obtained from the IBT culture collection (Kgs. Lyngby, Denmark) and maintained on YMEG-agar (4g/L yeast extract, 10g/L malt extract, 4g/L dextrose, 16g/L agar) or glucose minimal media with 10 mM ammonium tartrate as sole nitrogen source (GMM-NH$_4^+$)[248] at 28 °C. *Saccharomyces cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2- Δ1 trp1 pep4::HIS3 prb1Δ1.6R can1GAL*) was used as the yeast expression strain. *Escherichia coli* BL21 (DE3) (Novagen) was used as the *E.coli* expression strain.

4.2.2 General molecular biology experiments. General molecular cloning techniques were done as described elsewhere.[249] PCR was performed using Phusion® DNA

Table   4.1:   Primers   used   for   Section   4

Primer name	Sequence (5'-> 3')	Notes
gsfB-KO-P1	ctatgaagacccttccacctggcta	Screening of mutants and cloning of KO cassette
gsfB-KO-P2	acgcactcgttgctatgtcaaaca	cloning of KO cassette
gsfB-KO-P3	cctgcccgtcaccgagatttagaacgagcaacctgtcaaggatg	
gsfB-KO-P4	cttcaatatcatcttctgtcgacgtcgcctgctaagggaatctgacggg caa	
gsfB-KO-P5	cttgagcgcctcctgagagtt	
gsfB-KO-P6	tccgacgacaggttccatcactct	Screening of mutants and cloning of KO cassette
gsfC-KO-P1	catccttgacaggttgctcgtt	Screening of mutants and cloning of KO cassette
gsfC-KO-P2	cgacacgttcttgccctctcgaa	cloning of KO cassette
gsfC-KO-P3	ctgcccgtcaccgagatttagtaccttactcgtcagccagtct	
gsfC-KO-P4	cttcaatatcatcttctgtcgac-acaactctcaggaggcgctc	
gsfC-KO-P5	tccataccccccgacacctcc	
gsfC-KO-P6	tcgaccagtctctcggcgta	Screening of mutants and cloning of KO cassette
gsfD-KO-P1	acgacttgccatcggcacca	Screening of mutants and cloning of KO cassette
gsfD-KO-P2	ctttctgggcatgttaggcgaa	cloning of KO cassette
gsfD-KO-P3	ctgcccgtcaccgagatttagaggctggatcggtattgagc	
gsfD-KO-P4	cttcaatatcatcttctgtcgacg-gactaagaatatcacgaggt	
gsfD-KO-P5	gattacgccaagccataaggca	
gsfD-KO-P6	gacacagtctgcactatgtcgaata	Screening of mutants and cloning of KO cassette
gsfE-KO-P1	gcaagttctagtacccgcgc	Screening of mutants and cloning of KO cassette
gsfE-KO-P2	gtctgagtcggtacgctcg	cloning of KO cassette
gsfE-KO-P3	tgcccgtcaccgagatttaggaagccagcactgagcactgc	
gsfE-KO-P4	tcaatatcatcttctgtcgaccgacaagtttagacagctgggg	
gsfE-KO-P5	gccttgagattcttggtctggg	
gsfE-KO-P6	gcacctgggcaaattgaatgg	Screening of mutants and cloning of KO cassette

Table 4.1 (continued): Primers used for Section 4

Primer name	Sequence (5'-> 3')	Notes
gsfF-KO-P1	ggaccttaacccgactaagaatatc	Screening of mutants and cloning of KO cassette
gsfF-KO-P2	aaggtccagctgatctcatgaatgtg	cloning of KO cassette
gsfF-KO-P3	ctgcccgtcaccgagatttagcataggtcaaccatagtcgggtg	
gsfF-KO-P4	cttcaatatcatcttctgtcgac gtgctgacatcgtcacagattgctc	
gsfF-KO-P5	Ataggatggcatcgcgtataagg	
gsfF-KO-P6	cttagcttcaggctctacgcg	Screening of mutants and cloning of KO cassette
gsfH-KO-P1	gtatggtccctcgggtgc	Screening of mutants and cloning of KO cassette
gsfH-KO-P2	ggatacattttcttcgtcggc	Cloning of KO cassette
gsfH-KO-P3	tgcccgtcaccgagatttaggtggactagacatcctgatcc	
gsfH-KO-P4	tcaatatcatcttctgtcgacaccaaggtagtgtcgaatcc	
gsfH-KO-P5	gtggcccaggaaattggg	
gsfH-KO-P6	gcttggatcgggctcg	Screening of mutants and cloning of KO cassette
gsfK-KO-P1	gtggtggcggttctcag	Screening of mutants and cloning of KO cassette
gsfK-KO-P2	gacctcaggcaaggagac	Cloning of KO cassette
gsfK-KO-P3	tgcccgtcaccgagatttaggctcgatgaggaactgtggcag	
gsfK-KO-P4	tcaatatcatcttctgtcgaccgtctaccacaacctaccgc	
gsfK-KO-P5	ggaaacccacccagtcttgc	
gsfK-KO-P6	ggccgagcgaaatgacgg	Screening of mutants and cloning of KO cassette
gsfA-*Nhe*I-F	aaaatggctagcatgacttccgctaaggtctta	Cloning of YepLac195-gsfA
gsfA-*Afl*II-R	gcctcttaagaaccacgcttccgactcc	
gsfA-*Afl*II-F	aattcttaagaggcttgaggatgcagaagcc	
gsfA- *EcoR*V-R	gatatcacgcacctcgatgccaaggtctt	
gsfB- F	aaattcatatggcgtccaatacaagtcggt	Cloning of pET28-gsfB
gsfB- R	caatggatccacacatgccgaaatcgatgtttctga	
gsfC- F	ggtattgagggtcgcatgactcttgaccaaattagtcggatac	Cloning of pET30-gsfC
gsfC_R	agaggagagttagagccttatgacgtcgctaagctggc	

Table 4.1 (continued): Primers used for Section 4

Primer name	Sequence (5'-> 3')	Notes
gsfD-F	agctacatatgtctacccctgagcaatggatcca	Cloning of pET24-gsfD
gsfD- R	aaatttggatccctccctctttaaccttgcct	
gsfE- F	tttacatatgccaaaaacagctttcatcactggc	Cloning of pET28-gsfE
gsfE- R	aatgaattcatttaggtatcaaccccagctgtc	
gsfF- F	accctcactaaagggcggccgcatgactgttttgtttattct	Cloning of pESC-Leu-AtCPR/GsfF
*gsfF*_R	cttatcgtcgtcatccttgtaatctagacctcggactgtaactttaacc	
gsfI-F	aaaaagctagcgcgattcctcaatcttgtac	Cloning of pET24-gsfI
gsfI-R	ataatgaatcccatttggagatcccct	

Polymerase (New England Biolabs). DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs). DNA sequences of the gene constructs were confirmed by Sanger sequencing by Laragen, Inc (Culver City, CA). *E. coli* TOP10 (Invitrogen) and XL1-Blue (Stratagene) were used for cloning, following standard recombinant DNA techniques.[249] RNA extraction was performed using a RiboPure Yeast Kit (Ambion) and ImProm-II[TM] Reverse Transcription System for RT-PCR (Promega) was used to synthesize complementary DNA (cDNA) from total RNA. Transformation of *S. cerevisiae* BJ5464 transformation was done using the *SC* EasyComp Transformation kit (Invitrogen).

4.2.3 General protocol for protein expression in *E.coli*. Expression plasmid p*ET*28-*gsfB*, p*ET*30-*gsfC,* p*ET*24-*gsfD,* p*ET*24-*gsfI* or p*ET*28-*gsfE* was transformed into electrocompetent *E.coli* BL21 (DE3) and the cells were grown in 500 mL LB at 37°C and 250 rpm. When the OD_{600} reading reached 0.5, the cultures were cooled to 16 ^{o}C and protein expression was induced by addition of 60 µM IPTG and grown for additional 16 hours prior to His-tag-fusion protein purification (*vide infra*).

4.2.4 Heterologous expression of GsfA in *S.cerevisiae*. The two fragments of the cDNA for *gsfA*, was amplified using primer pairs GsfA-*Nhe*I-F and GsfA-*Afl*II-R and GsfA-*Afl*II-F and GsfA-*EcoR*V-R and cut with the appropriate restriction enzymes. The two fragments were simultaneously ligated into YepLac195 vector (ura3 selection marker), linearized by digestion with *Nhe*I and *Sma*I, such that the *gsfA* cDNA is flanked upstream by ADH2 promoter and downstream by an in-frame C-terminal His-tag and ADH2 terminator. The resulting YepLac195-GsfA plasmid was transformed for propagation in *E. coli* XL1and verified by sequencing. BJ5464-NpgA cells harboring YepLac195-GsfA were grown in YPD (10 g/L yeast extract, 20 g/L peptone) supplemented with 1% dextrose and incubated at 28 °C with shaking for 72 hours prior to His-tag fusion protein purification (*vide infra*).

4.2.5 General His-tag fusion protein purification. *S.cerevisiae* BJ5464-NpgA or *E.coli* BL21 cells expressing the desired protein were pelleted and resuspended in Buffer A (50 mM Tris-HCl pH 7.9, 500 mM NaCl) with 20 mM imidazole prior to lysing by sonification. The cell lysate was subjected to centrifugation at 27000g for 30 min for *E.coli* or 40000g for 1 hour for *S.cerevisiae*. Nickel-NTA-agarose resin was then added to the clarified lysate and the mixture was gently stirred at 4 °C overnight. The desired protein was then purified using gravity-flow column chromatography with increasing concentrations of imidazole (20-250 mM) in Buffer A. Purified protein was concentrated and buffer was exchanged into Buffer B (50 mM Tris-HCl, 2 mM EDTA, 50 mM NaCl, pH 8.0) using an Amicon Ultra-15 Centrifugal Filter Unit and stored in 10% glycerol.

4.2.6 *In vitro* assay of GsfA. For the *in vitro* synthesis of **5**, 10 μM GsfA was incubated with 2 mM malonyl-CoA and 15 μM tailoring enzyme (Hpm TE, GsfB, GsfC or GsfD) in 100 mM phosphate buffered saline pH 7.4 in a 100 μL reaction. The reaction was incubated overnight and

99

extracted twice with ethyl acetate. For the reactions using malonyl-CoA generated *in situ* by MatB, the assay was done essentially the same as above except malonyl-CoA was substituted with 20 μM MatB, 5mM CoA, 100 mM malonate, 5 mM $MgCl_2$, 5mM DTT and 20 mM ATP. The organic phase was dried and dissolved in 20 μL methanol and subjected to LCMS analysis. LC-MS was conducted with a Shimadzu 2010 EV Liquid Chromatograpy Mass Spectrometer using both positive and negative electrospray ionization monitoring the *m/z* range 200-800 and photodiode array monitoring the 290 nm wavelength. Samples were separated on a Phenomenex Luna 5μ 100 x 2 mm C18 reverse-phase column using a flow rate of 0.1 mL/min on a linear gradient of 5-95% solvent B in 30 min followed by isocratic 95% solvent B for another 15 min (solvent A: water with 0.1% (v/v) formic acid, solvent B: acetonitrile with 0.1% (v/v) formic acid).

4.2.7 Purification and Characterization of Norlichexanthone (5) from *S.cerevisiae*. *S. cerevisiae* strain BJ5464-NpgA harboring YepLac195-GsfA plasmid was inoculated to 4 mL Yeast Synthetic Drop-Out medium without uracil. The cells were grown for 36 hours with constant shaking at 28 °C. The seed culture was inoculated 1 L YPD (10 g yeast extract, 20 g peptone and 950 mL Milli-Q water) supplemented with 1% dextrose and grown for an additional 72 hours at 28 °C with constant shaking. The cells were harvested by centrifugation (3750 rpm, 10 minutes, 4 °C), extracted twice with equal volume of ethyl acetate and was evaporated to dryness. The extract was redissolved, loaded into a Sephadex LH-20 column (300 mm x 30 mm) and eluted using 1:1 ratio of chloroform and methanol. The fractions containing the desired compound, as ascertained using thin-layer chromatography and LCMS, were pooled and dried. The dried extract was resuspended in minimal volume of methanol, centrifuged and injected to the HPLC equipped with a Phenomenex Luna 5μ 250 x 10mm C18 reverse-phase column at a

flow rate 2.5 mL/min and using a linear gradient of 40-80% solvent B over 30 min (Solvent A: water with 0.1% trifluoroacetic acid (TFA), solvent B: acetonitrile with 0.1% TFA). NMR characterization (1D ^{1}H and ^{13}C, 2D HSQC and HMBC) of the purified compound was then performed on a Bruker AV500 NMR (500 MHz) equipped with 5mm dual cryoprobe at the UCLA Molecular Instrumentation Center.

4.2.8 Fungal Transformation and Gene Deletion in *P. aethiopicum*. Polyethylene glycol (PEG)-mediated transformation was done essentially as described previously.[103] Briefly, conidia from *P.aethiopicum* was inoculated to 250 mL liquid GMM-NH$_4$$^+$ for 13 hours at 28 °C, 250 rpm for germination. The harvested germlings were then digested with 3 mg/mL lysing enzyme (Sigma-Aldrich) and 2 mg/mL Yatalase (Takara Bio) to obtain the protoplasts, which were then transformed with the linear knockout cassette. The linear knockout cassettes containing the glufosinate resistance gene *bar* was constructed as described elsewhere[103, 250] and using primers listed in Table 3.1. After PEG-mediated transformation, the protoplasts were inoculated into GMM-NH$_4$$^+$ media supplemented with 1.2 M sorbitol, agar and 10mg/mL of glufosinate as selective agent. Genomic DNA from the transformants was isolated using Carlson lysis buffer and chloroform extraction followed by precipitation of DNA in the aqueous phase by addition of equal volume of isopropanol. PCR screening was performed as described previously[103] using primers listed in Table 3.1.

4.2.9 Chemical analysis and compound isolation from *P. aethiopicum* gene deletion mutants. For small-scale secondary metabolic profile analysis, the *P.aethiopicum* wild-type and transformants were grown in 10-20 mL YMEG liquid medium for 7 days at 28 °C without shaking. The cultures were extracted with equal volume of ethyl acetate with 1% acetic acid, evaporated to dryness and redissolved in methanol for LCMS analysis in the same manner as the

in vitro characterization of GsfA (*vide supra*). For preparative scale compound isolation, gene deletion strain of *P.aethiopicum* was grown in 3 L YMEG in the same manner as the small-scale cultures. The culture was extracted twice with equal volume of ethyl acetate and was evaporated to dryness and subjected to Sephadex and HPLC purification and NMR characterization in the same manner as **5** (*vide supra*).

4.2.10 Biotransformation of 11 to 14. *S. cerevisiae* BJ5464 transformed with *pESC-gsfF/AtCPR* was inoculated in 5 mL synthetic leucine dropout synthetic media (SDM, -Leu) and was grown overnight at 28 °C, 250 rpm. The overnight culture was inoculated to 500 mL synthetic leucine dropout media with galactose (SGMM, -Leu) for induction. The cells were grown at 28 °C, 250 rpm for 36 hours after induction and were pelleted by centrifugation at 4 °C. The cell pellet was resuspended in 50 mL SGMM, -Leu and **11** was added to a final concentration of 1 μM. The culture was grown at 28 °C, 250 rpm for 24 hours after feeding. Thereafter 1 mL of whole cell culture was taken, extracted twice with ethyl-acetate and dried to completeness before being subjected to LCMS analysis as described above.

4.2.11 Yeast microsome isolation and *in vitro* assay of GsfF. Yeast microsome extraction was adapted from the method described by Ralston *et al* and Barriuso *et al*.[251, 252] The overnight culture of BJ5464-NpgA harboring p*ESC-gsfF/AtCPR*, grown in the same manner as was done for the biotransformation of **11** to **14**, was inoculated to 20 mL SDM, -Leu and shaken for an additional 24 hrs prior to transfer to 500 mL synthetic leucine dropout media with galactose (SGMM, -Leu) for induction. The cells were grown at 28 °C, 250 rpm for 24 hours after induction and were pelleted by centrifugation at 4 °C. The cell pellet was then resuspended in 100 mL TES buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA and 0.6 mM sorbitol). The cells were pelleted again, resuspended in 100 mL TES buffer with 10 mM β-mercaptoethanol and

incubated at 25 °C for 10 minutes. The cells were then pelleted and resuspended in 2.5 mL extraction buffer (TES buffer supplemented with 1% bovine serum albumin and 2 mM β-mercaptoethanol and 1mM phenylmethylsulfonyl fluoride (Sigma-Aldrich)). Zirconia/silica beads (0.5 mm in diameter, Biospec Products) were added until skimming the surface of the cell suspension. Cell walls were disrupted manually by hand-shaking in a cold room for 10 min at 30 s intervals separated by 30 s intervals on ice. Cell extracts were transferred to a 50 mL centrifuge tube, the Zirconia/silica beads were washed three times with 5 mL of extraction buffer, and the washes were pooled with the original cell extracts. Finally, microsomes were obtained by differential centrifugation at 10,000g for 10 min at 4°C to remove cellular debris followed by centrifugation at 100,000g for 70 min at 4°C. The microsomal pellets were weighed prior to resuspension in 1.5 mL of TEG-M buffer (50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 20% glycerol, and 1.5 mM 2-mercaptoethanol) and stored frozen at -80 °C. The GsfF in vitro assay was performed by addition of microsomal protein to a final concentration of 0.42 mg/mL to 50 µM griseophenone B and 2mM NADPH in a 100 µL reaction. The reaction was incubated at 28 °C overnight and was extracted with ethyl acetate. The organic phase was dried and redissolved in 10 µL methanol prior to injection to LC-MS and analyzed as the same manner as above.

4.2.12 *In vitro* assay of GsfI. The assay for the chlorinase was done essentially the same as described in Zhou *et al*.[87] Briefly, 50 µM of GsfI was incubated with 200 µM Griseophenone C (**10**) in 100 mM sodium phosphate buffer (pH 7.4) and 50 mM NaCl. In order to regenerate the reduced flavin in GsfI, the flavin reductase SsuE (15 µM), FAD (5 µM) and NADPH (2 mM) were added to the reaction mix. After overnight incubation, the reaction mix was extracted twice with equal volume ethyl acetate, dried *in vacuo* and subjected to LCMS analysis as described above for **5** (*vide supra*).

4.2.13 *In vitro* assay of GsfE. For in vitro synthesis of **1** from **18**, 10 µM of GsfE was incubated with 100 µM and 1 mM NADPH in 100 mM Tris-HCl (pH 7.5) buffer and 100 µL reaction volume. After 5 hour incubation in room temperature, the reaction mix was extracted twice with equal volume of ethyl acetate, dried *in vacuo* and analyzed by LCMS in the same manner as **5** (*vide supra*).

4.2.14 Total *in vitro* biosynthesis of 1 and 2. For the total *in vitro* synthesis of **2**, 4 µM of GsfA, 2 µM each of the methyltransferases (GsfB, GsfC and GsfD), 2 µL of microsomal protein containing GsfF and *At*CPR and 2 µM of GsfE were incubated with 2 mM of malonyl-CoA, 100 µM of *S*-adenosyl methionine and 2 mM of NADPH in 100 µL reaction mix buffered with 100 mM sodium phosphate (pH 7.4) and 50 mM NaCl. After overnight incubation, the reaction mix was extracted twice with equal volume ethyl acetate, dried and subjected to the same LCMS analysis as **5** (*vide supra*). The total *in vitro* biosynthesis of **1** was done in the same manner as **2**, except GsfA-C were incubated with 10 µM of GsfI and 2 µM SsuE for 3 hours prior to addition of GsfF, GsfE and GsfD. The total yield of **1** was measured by comparing the area under the chromatogram peak of **1** against a standard curve of known amount of **1** injected and analyzed by LCMS.

4.3 Results

4.3.1 GsfA is a norlichexanthone synthase. Our initial genetic studies with *P. aethiopicum* showed that *gsfA,* which encodes a thioesterase (TE)-less NR-PKS, is required to produce the heptaketide backbone of **1**.[103] To biochemically investigate the role of GsfA, *gsfA* cDNA was cloned into the YEpLac195 yeast-*E. coli* shuttle vector driven by the ADH2 promoter. The resulting YepLac195-GsfA construct was transformed into *S. cerevisiae* BJ5464-NpgA strain,

which contain a chromosomal copy of *npgA*, the phosphopantetheine transferase gene from *Aspergillus nidulans* which post-translationally installs the phosphopantetheine arm to the acyl carrier domains (ACP) domain of PKS.[94] Intact holo-GsfA (185 kDa) with a C-terminal His-tag was purified to near homogeneity to a final titer of 2 mg/L of yeast culture.
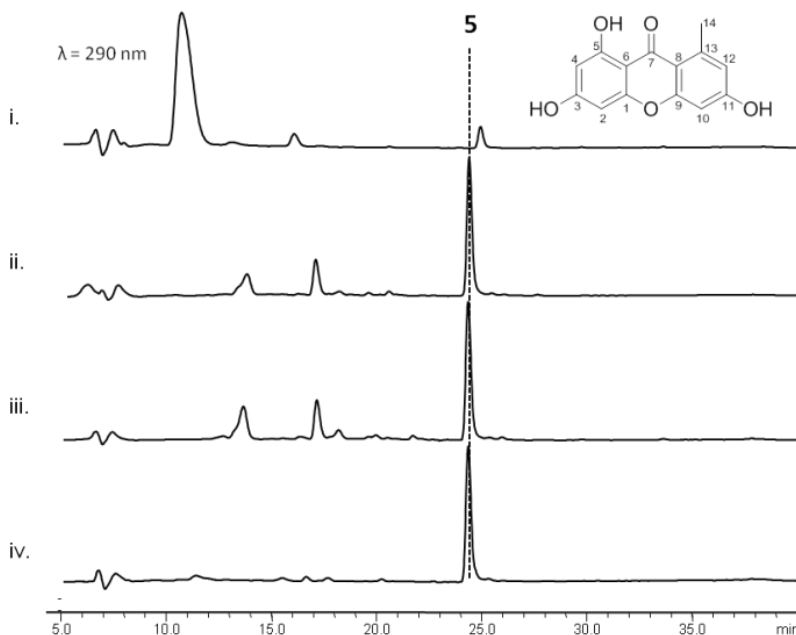
Figure 4.2: GsfA catalyzes formation of **5** from malonyl-CoA as confirmed by HPLC analysis. i.) No GsfA control ii.) GsfA and the excised TE domain of Hpm3 incubated with *in situ* generated malonyl-CoA. iii.) GsfA incubated with *in situ* generated malonyl-CoA using MatB, malonate, CoA and ATP. iv.) GsfA incubated with malonyl-CoA only.

To assay the activity of recombinant GsfA, we incubated the enzyme with either malonyl-CoA directly or with the *in situ* malonyl-CoA generation system (Figure 4.2). Furthermore, since GsfA lacks a TE domain that may release the mature polyketide product and no standalone TE is found in the gene cluster, we incubated GsfA with a promiscuous TE domain from the hypothemycin NR-PKS Hpm3.[86] Surprisingly, GsfA alone was sufficient for the production and release of a predominant product **5** with *m/z*=259, [M+H]$^+$ (Figure 4.2).

105

Addition of the Hpm3 TE domain does not enhance the turnover of the product **5**. To purify sufficient amount of **5** for structural characterization, we cultured the BJ5464-NpgA strain expressing GsfA, which was able to produce **5** at a final titer of 10 mg/L. Compound purification followed NMR characterization identified **5** as norlichexanthone (Figure 4.2), which bears the expected heptaketide backbone.

Table 4.2: Griseophenones that were isolated from the *gsf* knockout strains.
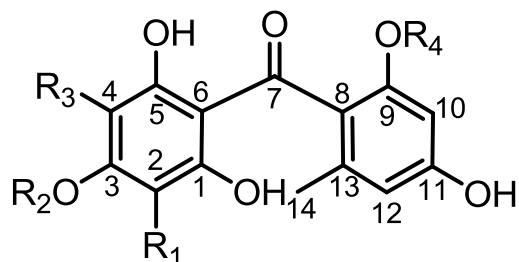
Compound Name	R_1	R_2	R_3	R_4
griseophenone D (**6**)	H	Me	H	H
griseophenone E (**7**)	H	H	H	Me
griseophenone F (**8**)	Cl	H	H	Me
griseophenone C (**10**)	H	Me	H	Me
griseophenone B (**11**)	Cl	Me	H	Me
griseophenone G (**12**)	Cl	Me	Cl	Me
griseophenone H (**17**)	Cl	Me	H	H

Based on the structure of **5**, it appears that the nascent linear product of GsfA undergoes a C8-C13 cyclization via aldol-condensation to form the orcinol ring, presumably catalyzed by the product template domain of GsfA[83, 84] and a Claisen condensation between C1 and C6 to form the phloroglucinol ring.
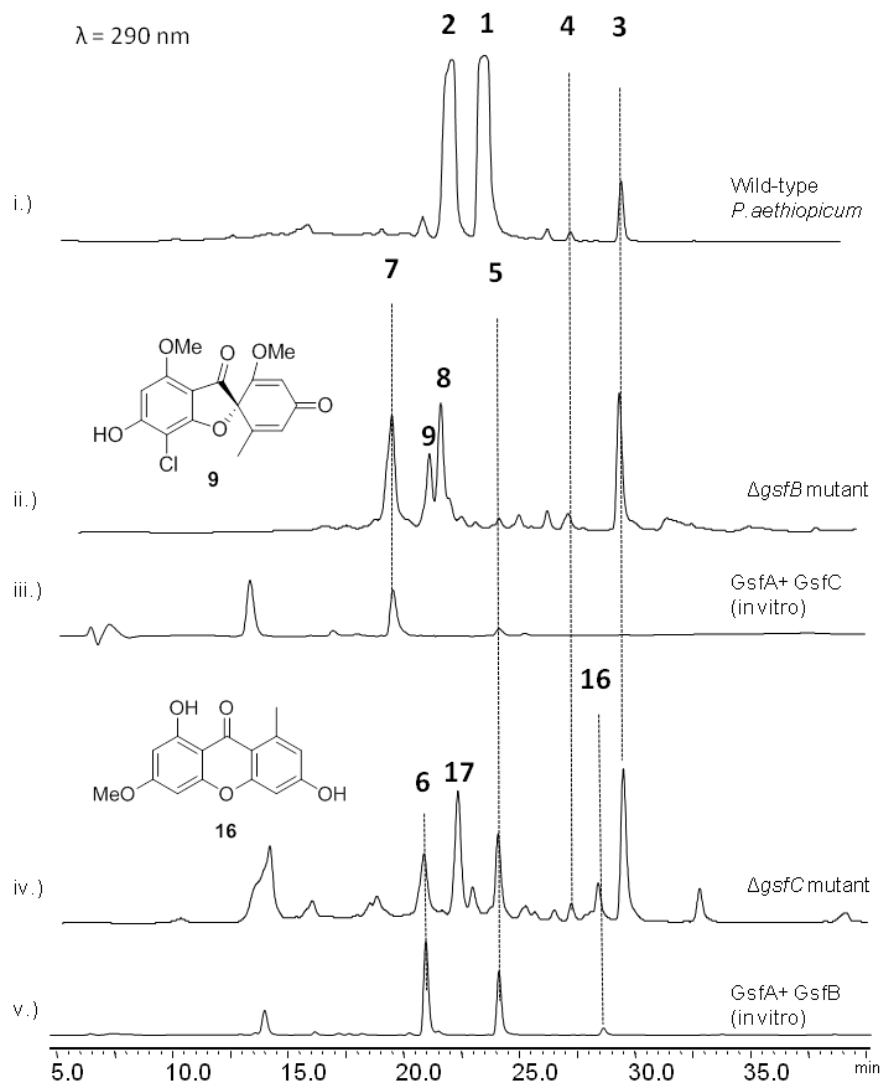
Figure 4.3: The *O*-methyltransferases GsfB and GsfC tailor the nascent GsfA product. i.) Secondary metabolic profile of *Penicillium aethiopicum* showing the production of 1-4. ii.) Deletion of *gsfB* led to the production of **5, 7, 8,** and **9**. iii.) Coupled *in vitro* assay of GsfA and GsfC led to the production of **7** thereby confirming 9-OH regioselectivity of GsfC iv.) Deletion of *gsfC* led to the production of **5, 6, 16** and **17**. v.) Coupled *in vitro* assay of GsfA and GsfB led to the production of **5, 6** and **16** thereby confirming the 3-OH regioselectivity of GsfB.

4.3.2 Methylation of GsfA product by GsfB and GsfC hinders xanthone formation.

Although the xanthone **5** is detected as the only product from the in vitro assay, it is most likely that the true heptaketide product of GsfA is the benzophenone **5a** (Table 3.2), which can undergo

a spontaneous dehydration to form **5**. Formation of the grisan ring found in **1** must proceed through the intermediate **5a**, thereby making **5** a likely off-pathway shunt product. Indeed, feeding of **5** to the Δ*gsfA* mutant did not restore the production of **1** (Figure. A28), thereby ruling out the possibility of **5** undergoing rehydration to open the xanthone ring and form **5a** during biosynthesis of **1.** Furthermore, this suggests that a mechanism in which downstream tailoring reaction that modify **5a** can prevent the undesirable dehydration reaction. We hypothesize that the methylation of the phenols in **5a**, especially at 9-OH may hinder dehydration and suppress xanthone formation. To investigate this hypothesis, genetic inactivation and *in vitro* reconstitution of the three *O*-methyltransferases (MTs) GsfB, GsfC and GsfD were performed.

Scheme 4.1: Non-phenone shunt metabolites from Δ*gsfB* mutant

Griseoxanthone C (**16**)

Normethyl-
dehydrogriseofulvin B (**9**)
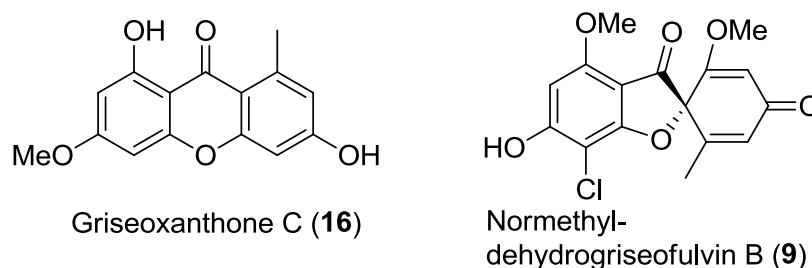
Deletion of *gsfB* led to the production of **5** as a minor product, **7** (obs. *m/z*= 291, [M+H]$^+$), **8** (obs *m/z*= 325, [M+H]$^+$), and **9** (obs. *m/z*= 337, [M+H]$^+$) (Figure 4.3, ii). Elucidation of the structures of the compounds using 1D ^{1}H and ^{13}C and 2D HSQC and HMBC NMR revealed the identity of **7** as griseophenone E, **8** as griseophenone F and **9** as desmethyl-dehydrogriseofulvin B. **7**-**9** share the characteristic lack of methylation at 3-OH, suggesting that GsfB is the 3-OH MT. The isolation of **9** also suggests that downstream steps, including the formation of the grisan structure, are not dependent on the 3-OH methylation step. On the other hand, the Δ*gsfC* mutant showed production of **5** as a more prominent product in comparison to the metabolic profile of the Δ*gsfB,* along with **6** (obs. *m/z* = 291, [M+H]$^+$), **16** (obs. *m/z* = 273,

[M+H]$^+$) and **17**(obs. *m/z=* 325, [M+H]$^+$), which were identified as griseophenone D, griseoxanthone C and griseophenone H respectively (Figure 4.3, iv). In contrast to the metabolites found in *ΔgsfB* mutant, **6**, **16** and **17** are methylated at 3-OH, but lack methylation at 9-OH, thereby confirming that GsfC is the 9-OH MT. Isolation of these benzophenone intermediates in the *ΔgsfB* and *ΔgsfC* mutants confirms our hypothesis that methylations at 3-OH and more so at 9-OH, can suppress the formation of the off-pathway shunt product **5**.

In order to verify the roles of the MTs, in vitro assay of GsfA together with each of the recombinant MT was performed. The coupled reaction with GsfA generates the labile benzophenone **5a** in situ that serve as substrates for the MTs. Assay containing GsfA and GsfB produced predominantly xanthones **5** and **16**, as well as **6**, demonstrating limited suppression of dehydration after methylation of 3-OH (Figure 4.3, v). Presumably, **6** undergoes dehydration to form **16**. Alternatively, the nascent polyketide product of GsfA undergoes dehydration to **5** followed by methylation by GsfB to form **16**. The coupled assay of GsfA with GsfC, on the other hand, produced predominantly **7** as expected, with a trace amount of **5,** further confirming the 9-OH methylation step as a means to keep the GsfA product **5a** on pathway during biosynthesis of **1** (Figure 4.3, iii). Taken together, these studies with MTs confirm that the GsfA product undergoes methylation by both GsfB and GsfC before forming the grisan ring, which is in agreement of the feeding studies done by Rhodes *et al.*[245, 246]

On the other hand, knockout of *gsfD* yielded the production of grisan-containing compounds desmethyl-dechlorogriseofulvin (**13**) (obs. *m/z* = 305, [M+H]$^+$), desmethyl-dehydrogriseofulvin A (**14**) (obs. *m/z* = 337, [M+H]$^+$), and desmethyl-griseofulvin (**15**) (obs. *m/z* = 339, [M+H]$^+$) (Figure 4.4, iii). These metabolites represent late-stage shunt products (**13** and **15**) and intermediate **14** during the formation of **1.** The structures of **13**-**15** show the lack of

methylation at the 5-OH position, implying that Gsf regioselectively targets 5-OH. Based on these results, we propose the role of GsfD is to methylate of **14** to produce dehydrogriseofulvin (**18**).
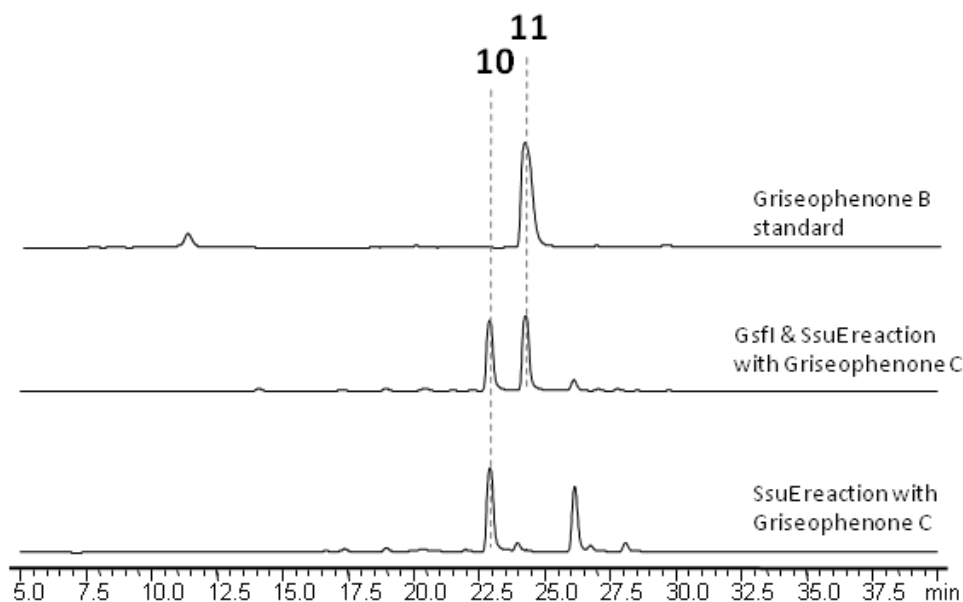
Figure 4.4: *In vitro* assay of GsfI using griseophenone C (**10**) as substrate. GsfI, heterologously expressed and purified from *E.coli* BL21 cells was incubated together with the NADPH-dependent flavin reductase SsuE and griseophenone C in 100 mM sodium phosphate buffer (pH 7.4) and 50 mM NaCl showing the conversion of **10** to griseophenone B

4.3.3 Biochemical confirmation of the role of GsfI. Previously, loss of **1** and production of only **2** were observed when the gene *gsfI* encoding the flavin-dependent halogenase was deleted in *P. aethiopicum*. To biochemically reconstitute the halogenation reaction, we cloned, expressed and purified GsfI and performed *in vitro* assay using **10**, which is the proposed intermediate from the previous labeling studies,[246] as substrate. In order to regenerate the reduced flavin cofactor of GsfI, the NADPH-dependent flavin reductase SsuE was added to the reaction. As

expected, assay of the non-chlorinated **10** with GsfI led to the production of chlorinated phenone **11** at ~ 50% conversion (Figure 4.4).
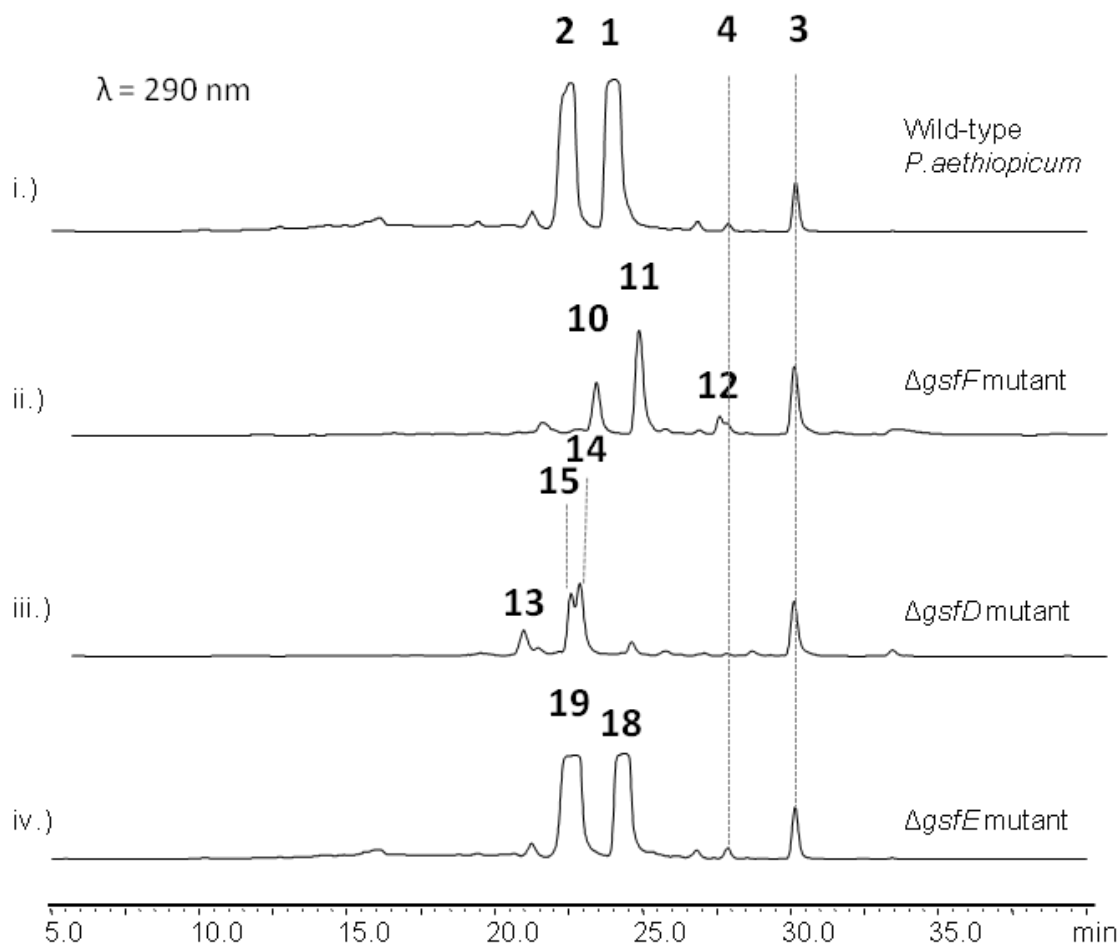
Figure 4.5: Tailoring reactions in biosynthesis of 1. i.) Wild-type *P.aethiopicum* showing the production of **1-4**. ii.) Knockout of cytochrome P450 gene *gsfF* led to the production of the dimethylated phenones **10**, **11** & **12** containing different degrees of chlorination. iii.) Knockout of the *O*-methyltransferase *gsfD* led to the production of **13**, **14** and **15**. iv.) Knockout of the short-chain dehydrogenase /reductase gene *gsfE* led to the production of **18**, and **19**.

4.3.4 GsfF performs phenolic coupling to afford the grisan scaffold. Oxidation of the phenone ring to form the grisan ring was first proposed by Barton and Cohen.[244] This biosynthetic step dramatically transforms the phenone scaffold of griseophenone B (**11**) into the

111

spirobicyclic grisan compound **14**, thereby dearomatizing the orcinol ring of **11** and creating the cyclohexadienone in **14**. An analogous reaction during biosynthesis of geodin in *A. terreus* was proposed to be catalyzed by copper-centered enzyme dehydrogeodin oxidase.[253] However, such a copper laccase enzyme is not encoded in the *gsf* cluster, thus we proposed that the cytochrome P450 (CYP) GsfF may instead catalyze the phenolic coupling reaction of **11** to afford **14**.[103] To probe the role of GsfF, we deleted *gsfF* in *P. aethiopicum* and characterized the resulting metabolites via LC-MS and NMR.

The Δ*gsfF* strain produced unchlorinated griseophenone C (**10**) (obs. $m/z = 305$, $[M+H]^+$), monochlorinated **11** (obs. $m/z = 339$, $[M+H]^+$) and dichlorinated griseophenone G (**12**) (obs. $m/z = 373$, $[M+H]^+$) (Figure 4.5, ii). Accumulation of **10** and **11** as major intermediates therefore supports the coupling role of GsfF in the pathway. The degrees of chlorination of the metabolites were further ascertained based on the presence of the expected 3:1 isotopic distribution of the monochlorinated **11** and the 1.5:1 isotopic distribution of the dichlorinated **12** (Appendix Tables A11-A13). The C2-regioselectivity of the first chlorination reaction was determined in the loss of the symmetry along the C3-C6 axis of **11** in comparison to **10**, as evident in comparative NMR analysis. The symmetry along C3-C6 axis as shown in the NMR spectra is restored upon further chlorination of **11** to afford **12,** thus confirming the regioselectivity of the second chlorination at C4.

To functionally confirm the role of GsfF, the CYP was expressed in *S. cerevisiae* BJ5464-NpgA strain together with the *A. terreus* CYP oxidoreductase (CPR) using Gal1 and Gal10 bidirectional promoter,[254] which allows the temporal coexpression of the CYP and the CPR that is required for turnover of the CYP. Following induction of GsfF and CPR expression, **11** was supplemented to the yeast culture at a final concentration of 1 μM. The organic

metabolites was extracted twenty-four hours later with ethyl acetate and analyzed by LCMS. *S. cerevisiae* cells expressing both CPR and GsfF were able to completely convert **11** to a product that is identical (UV, mass and retention time) to **14** isolated from *ΔgsfD* mutant (Figure 4.6, ii). In contrast, yeast cells expressing only CPR did not show any conversion of **11**(Figure 4.6, i).
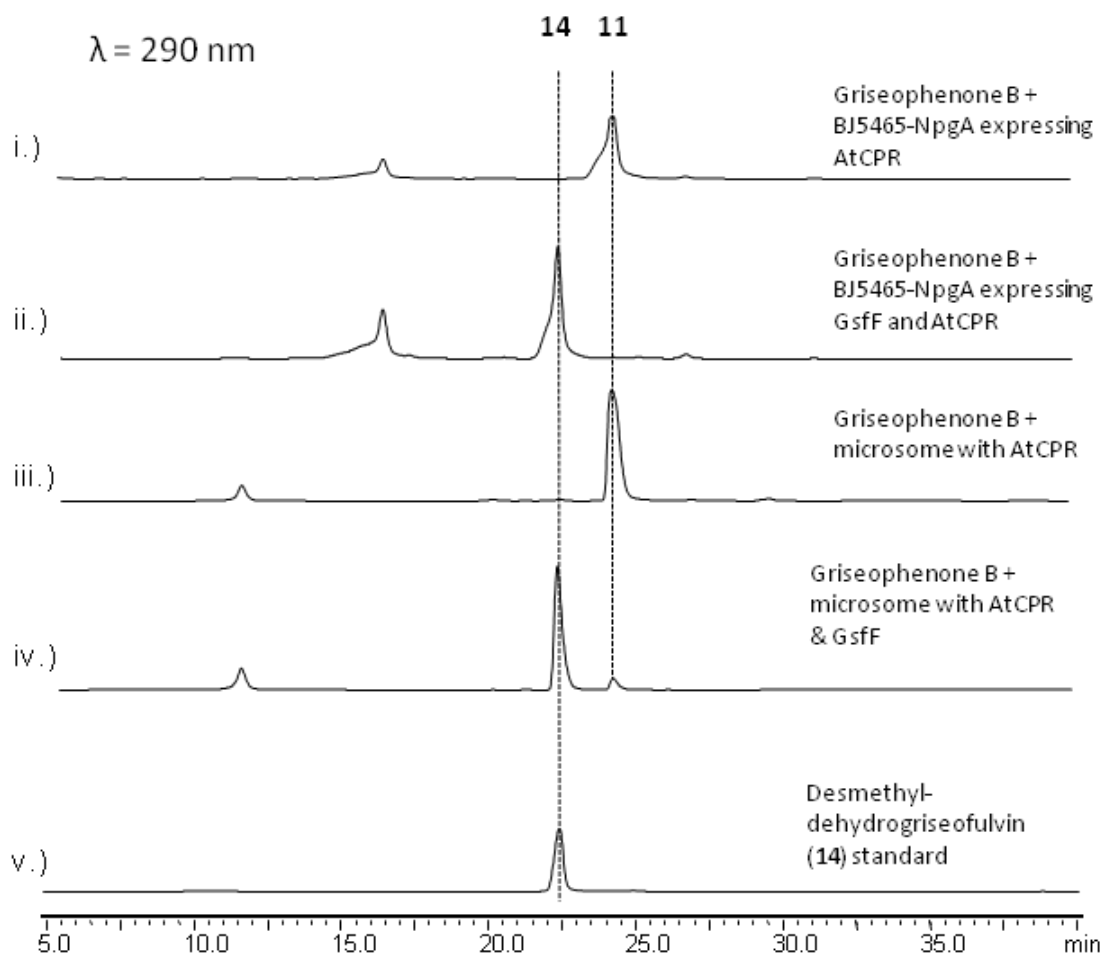
Figure 4.6: The cytochrome P450 GsfF catalyzes the coupling of orcinol and phloroglucinol rings in griseophenone B (**11**) to form desmethyl-dehydrogriseofulvin A(**14**). i.) BJ5464-NpgA expressing CPR only did not convert **11** to **14**. ii.) BJ5464-NpgA expressing both CPR and GsfF completely converted **11** to **14** after 24 hours. iii.) Yeast microsomes containing CPR do not convert **11** in vitro. iv.) Yeast microsomes containing both CPR and GsfF converted **11** to **14** in the presence of NADPH. v.) Standard of **14** purified from Δ*gsfD* deletion strain.
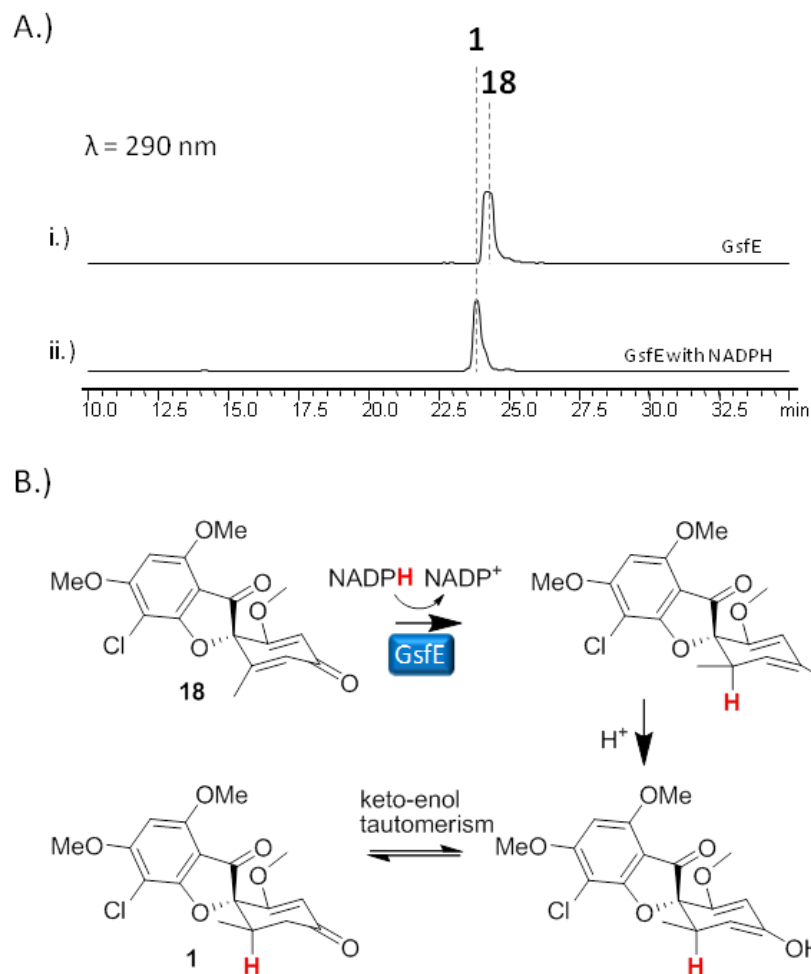
Figure 4.7: Verifying the activity of GsfE. A.) Incubation of **18** with 10 µM GsfE and 1 mM NADPH led to the production of **1**. B.) Putative mechanism of GsfE showing the stereospecific 1,4-(Michael) addition of the hydride from NADPH to **18,** followed by protonation of the resulting cyclohexadienolate to form **1**.

To further prove the conversion of **11** to **14** is catalyzed by GsfF, microsomal fraction from yeast cells coexpressing CPR and GsfF was isolated and *in vitro* assay was performed in the presence of NADPH. Microsomes containing both GsfF and CPR facilitated 90% conversion of **11** to **14** after overnight incubation (Figure 4.6, iv). Time course kinetics of GsfF revealed conversion of **11** to **14** at a turnover rate of 0.437 µM min^{-1}·mg^{-1} of microsomal protein (Appendix figure A29). In contrast, identically prepared microsomes containing only CPR did

not shown any conversion (Figure 4.6, iii). Therefore, we unequivocally confirmed the role of GsfF in the oxidative formation of the spirocyclic grisan.

4.3.5 GsfE performs stereospecific reduction of enol 18 to afford the final product 1. Having functionally characterized the PKS GsfA, the MTs GsfB-D, the chlorinase GsfI and the P450 GsfF, the only uncharacterized biosynthetic step in the pathway is reduction of the cyclohexadienone ring in **18** or dechloro-dehydrogriseofulvin (**19**) to afford **1** or **2,** respectively. Of the remaining uncharacterized genes in the cluster, GsfE and GsfK; both

Scheme 4.2: Late-shunt metabolites and intermediates from the biosynthesis of **1** showing the distinct presence of the grisan scaffold.

R= H, Normethyl-
    dechlorogriseofulvin (**13**)

R= Cl, Normethylgriseofulvin (**15**)

R_1= Cl, R_2 = H, Normethyl-
    dehydrogriseofulvin A (**14**)

R_1= Cl, R_2=CH_3,Dehydrogriseofulvin (**18**)
R_1= H, R_2=CH_3, Dechloro-
    dehydrogriseofulvin (**19**)
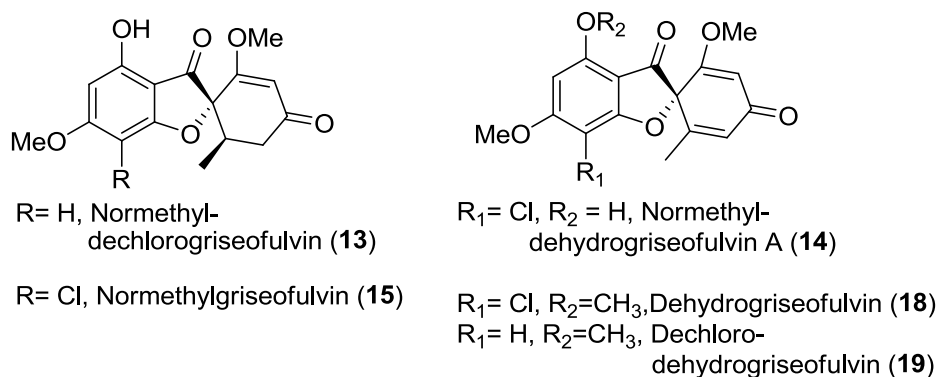
contain a conserved binding site for a nicotinamide cofactor and one of these may therefore be capable of performing the final enoylreduction. Thus, single gene deletion mutants of *gsfE* and *gsfK* were constructed followed by metabolite analysis. All of the Δ*gsfK* mutants isolated retained the ability to produce **1** and **2**. On the other hand, knockout of *gsfE* led to the production of **18** and **19** (Figure 4.3, iv), both of which were structurally verified by extensive NMR characterization (Table A20-A21). Interestingly, while initial bioinformatic analysis of GsfE suggested that it is related to nucleoside-sugar epimerases based on sequence alignment, further

analysis of the GsfE protein sequence via structural homology prediction revealed that it shares

an overall folding, as well as conserved catalytic lysine and tyrosine as the progesterone-5β-

reductase (POR) from the cardenolide biosynthetic pathway in *Digitalis lanata*.[255] POR

catalyzes the stereospecific reduction of progesterone to form 5β-pregnane-3,20-dione, a reaction

that is analogous to the reduction of the cyclohexadienone ring of **18** to form **1**. To verify the

genetic and bioinformatic results, we reconstituted the reaction *in vitro* using His-tag fused-GsfE

expressed and purified from *E. coli* BL21(DE3),and **18** or **19** as substrates. NADPH was added

as the reducing cofactor of GsfE. As shown in Figure 4.7A and Figure A31, GsfE fully converted

18 or **19** to **1** or **2**, respectively, thereby confirming its role as the last enzyme in the pathway.
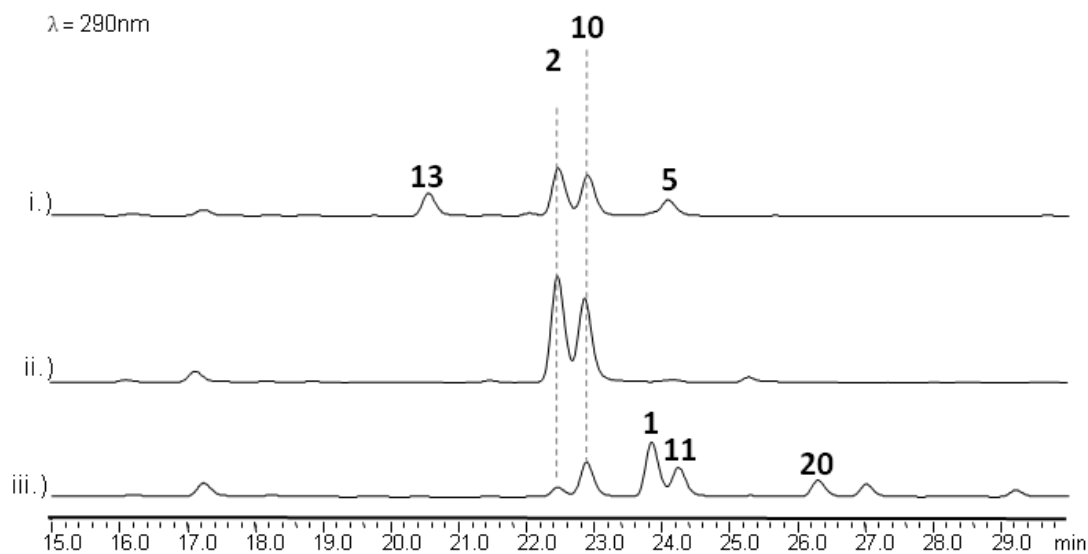
Figure 4.8: In v itro total biosynthesis of **1** and **2**. i.) GsfA-F, SAM and malonyl-CoA added in one step. ii.) GsfA-C, SAM and malonyl-CoA incubated together for 3 hours prior to addition of NADPH, GsfF ,GsfD and GsfE. iii.) GsfA-C, GsfI, SsuE, SAM, malonyl-CoA and NADPH incubated together for 3 hours prior to addition of GsfF, GsfD and GsfE.

4.3.6 Total *in vitro* biosynthesis of 1 and 2. Upon identifying all enzymatic components of the

biosynthetic pathway, we attempted in vitro synthesis of **1** and **2** using purified Gsf enzymes (see

Methods). Incubation of GsfA-F with malonyl-CoA, SAM and NADPH in PBS resulted in the expected production of **2** as a major product, along with **5**, **10** and **13**. Addition of GsfD-F after three-hour preincubation of GsfA-C with malonyl-CoA and SAM led to the production of **2** and **10** only. Finally, addition of the chlorinase GsfI and SsuE together with GsfA-F led to the production of the desired product **1** as a major product along with **2**, **10**, **11** and griseophenone A (**20**) (Figure 4.8, iii). The final yield of of **1** with respect to malonyl-CoA was ~1%. The presence of **10** and **11** in all assays indicate the microsomal GsfF-catalyzed oxidative coupling is the rate limiting step in the multi-enzyme reactions.

4.4 Discussion

Scheme 4.3: Proposed Biosynthesis of Griseofulvin (1)
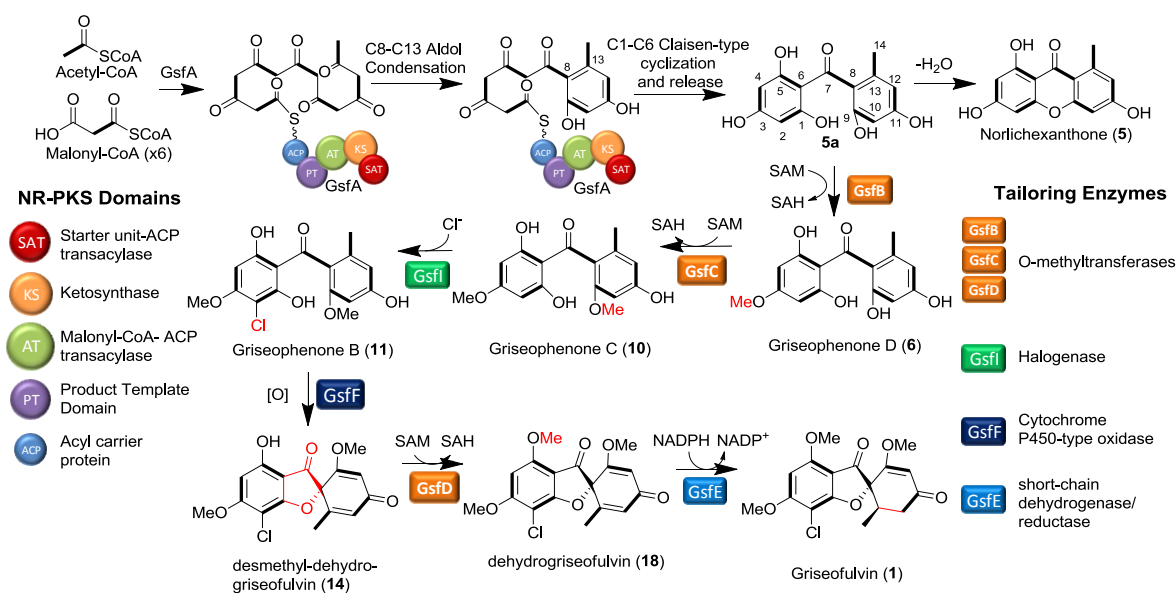
In this study, we performed comprehensive single-gene deletion of the *gsf* genes and functional characterization of the corresponding enzymes to elucidate the biosynthetic pathway of **1**. The functional roles of the enzymes the pathway, except GsfK which is not involved in the pathway, were elucidated. Following biosynthesis of the polyketide backbone by GsfA, six

enzymes are required to transform the reactive benzophenone product **5a** into the grisan containing **1**. In addition, several grisan analogs of **1** were isolated from the different gene deletion mutants (Schemes 4.1 and 4.2), demonstrating the overall flexibility of the pathway and its potential for biosynthetic engineering.

Based on the results from the genetic and enzymatic data, as well as the prior biosynthetic studies using labeled substrates, we can establish the biosynthetic pathway for **1** (Scheme 4.3). The biosynthesis is initiated by the formation of the heptaketide by GsfA using one acetyl-CoA starter unit and six malonyl-CoA extender units. Thereafter, the product template (PT) domain of GsfA catalyzes the unorthodox C8-C13 aldol-type cyclization followed by C1-C6 Claisen-type cyclization to form **5a**, wherein dehydration of **5a** can readily take place in the absence of downstream steps to form the xanthone shunt product **5**. Meanwhile, on pathway processing of **5a** continues through the methylation at 3-OH and 9-OH by GsfB and GsfC, respectively, to form **10**. Although feeding studies by Harris *et al* suggest that 3-OH methylation has priority over 9-OH methylation,[247] results from the knockout of *gsfB* and *gsfC* indicate that either can perform the first methylation reaction on the GsfA product **5a**. Following the two methylation steps, **10** undergoes chlorination by the flavin-dependent halogenase GsfI to form **11.** Following chlorination, **11** undergoes phenolic ring coupling by the cytochrome P450 GsfF to form the grisan compound **14**. Subsequently, **14** is subjected to two additional tailoring steps: methylation at 5-OH and reduction of the cyclohexadienone motif to afford **1**. Based on the presence of **14** in the Δ*gsfD* mutant, indicating the inefficient reduction by GsfE of the 5-OH-desmethyl substrates, it can be proposed that methylation by GsfD to yield **18** takes place before GsfE-catalyzed enoylreduction, which is in agreement with previous feeding studies on the biosynthesis of **1**.[245-247, 256] The chlorination step by GsfI is apparently incomplete in *P. aethiopicum* and *in*

vitro (Figure 4.8). Enzymes downstream of GsfI, including GsfF, GsfD and GsfE, can all act on the respective dechlorinated substrates to yield the final product **2**.

The NR-PKS GsfA produced the heptaketide **5** *in vitro* using malonyl-CoA as substrate. A typical NR-PKS contains starter-unit-ACP transacylase (SAT),[82] ketosynthase (KS), malonyl-CoA-ACP transacylase (AT), PT,[83, 84] ACP and TE domains.[85, 102] The cyclization regioselectivity of the nascent polyketide chain is determined by the functions of the PT and the TE domains.[83-85] The PT domain, in particular, mediates the regioselective aldol-type cyclization through steric interactions with the newly-formed polyketide.[83, 84] Using phylogenetic analysis and domain shuffling experiment, we previously classified ~100 PT domains into distinct clades based on ring size and first aldol condensation regioselectivity.[112] Interestingly, the GsfA PT does not classify into any of the clades, suggesting it should have unique cyclization selectivity. Indeed, based on previous labeling studies and the biochemical studies using purified GsfA, we can conclude that GsfA PT must catalyze the unusual C8-C13 aldol condensation as the first cyclization step in the formation of **5a**.

The enzymatic basis of the Claisen-like condensation (C1-C6) to afford the phloroglucinol ring, however, is unclear. GsfA belongs to a TE-less NR-PKS, of which no terminal TE domain is found in the NR-PKS. For TE-less NR-PKSs, a standalone β-lactamase-like enzyme is typically present in the gene cluster to release the product either through hydrolysis or Claisen-like condensation.[118] No such releasing enzyme is present in the *gsf* cluster, and our reconstitution studies show that GsfA alone is able to turnover products. Therefore the activities to catalyze the C1-C6 cyclization must be within the GsfA NR-PKS. It is reasonable to propose that because of the smaller size of the polyketide, the PT domain, when distorting the polyketide backbone for the aldol cyclization, may also promote the C1-C6

cyclization in its active site. Hence, formation of **5a** represents an intriguing example in which a NR-PKS performs a Claisen-like cyclization without using a dedicated, in-line or dissociated TE.

While the two cyclization events in the biosynthesis of **5a** (and **5**) are unusual, it is far from unique. To wit, compounds bearing the norlichexanthone scaffold have been isolated from lichens.[257] Thus one can reasonably speculate that a GsfA homolog may be found in the fungal symbionts in the lichens that produce compounds related to **5**. Interestingly, these norlichexanthone-derived compounds found in lichens undergo similar *O*-methylation and chlorination as the xanthone and phenone shunt products and intermediates characterized here, albeit with different regioselectivities.[257] These myriad of post-tailoring steps generate additional levels of complexity on the structurally simplistic polyketide product such as **5a**.

Based on the relative abundance of xanthone in comparison to phenone in the *gsfB* and *gsfC* deletion mutants, as well as the amount of **5** in the coupled GsfA and GsfB/C assays, methylation of 9-OH by GsfC serves an additional role in slowing down the dehydration of the phenone into the xanthone. The resulting hemiacetal formed after the nucleophilic attack of 1-OH to C9 is less prone to dehydration compared to the corresponding *gem* diol that forms when 9-OH is not methylated. Thus, by methylating the 9-OH position, *P. aethiopicum* can keep the biosynthetic pathway on track and minimize derailment as a result of xanthone formation. After the GsfA product undergoes the first cascade of tailoring steps by the actions of GsfB, C and I (Scheme 2), the cytochrome P450 oxygenase GsfF performs the coupling of the phloroglucinol and orcinol rings in **11** to form **14.** In this reaction, the orcinol ring is contorted resulting in the spirobicyclic grisan scaffold found in **14**. In the same way that the 9-OH methylation is the bifurcation point that determines the benzophenone or xanthone fate of the GsfA product, the ring coupling reaction serves as a switch for the drastic metamorphosis of the di-aromatic **11** into

the tricyclic **14**. Based on the feeding studies in the biosynthesis of **1** using [1,1-^{18}O]-labeled

acetate, all of the oxygen atoms in **1** originated from acetate.[256] Based on known reactions

catalyzed by cytochrome P450, one can propose two different mechanisms for GsfF reaction

(Scheme 3): A.) a di-radical mechanism in which the orcinol and phloroglucinol rings

sequentially lose a hydrogen atom and B.) a mechanism involving the formation of an arene

oxide intermediate (**11b**). In the first mechanism, the reaction is initiated by the formation of

Compound I, the active form of the heme cofactor of GsfF. Compound I then abstracts a

hydrogen atom from**11** to form the resonance-stabilized phenoxy radical on the orcinol ring.

Scheme 4.4: Proposed Mechanism for GsfF A.) Di-radical ring coupling mechanism of GsfF. B.) Alternative
mechanism for the ring coupling involving an arene oxide intermediate**.**
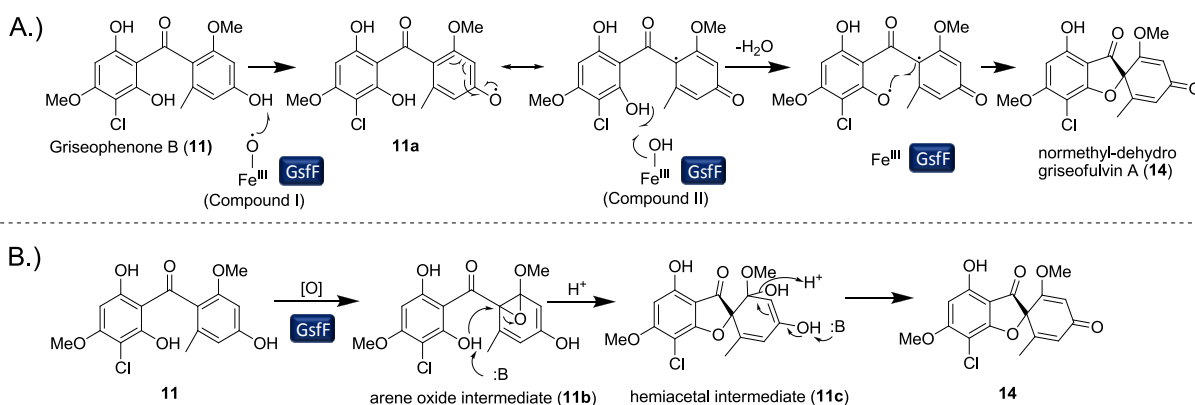
Thereafter, Compound II form of GsfF abstracts the second hydrogen atom from **11a** to form the

phenoxy radical on the phloroglucinol ring with water as a by-product of the reaction. The

incipient phloroglucinol radical then couples with the orcinol radical in a stereospecific fashion

to form **14**. This di-radical mechanism, which was first envisioned by Barton and Cohen,[244]

has also been proposed for the mechanisms of the coupling between the isoquinoline and

phenolic rings in reticuline by salutaridine synthase during morphine biosynthesis,[258] the

coupling of the two indole rings of chromopyrrolic acid by StaP to form indolocarbazole in the

biosynthesis of staurosporine,[259] and the dimerization of coniferyl alcohol monomers in the biosynthesis of lignin.[260] Incidentally, both salutaridine synthase and StaP are cytochrome P450-type oxygenase as well.[258, 259] Moreover, this proposed mechanism is in agreement with the [1,1-^{18}O]-acetate labeling study since all oxygen atoms are retained during the transformation of **11** to **14**.[256] Alternatively, the oxidation of **11** by GsfF can go through an arene oxide intermediate (**11b**), which will then be subjected to nucleophilic attack by the hydroxyl of the phloroglucinol ring to form **11c**. The hemiacetal intermediate then undergoes dehydration to afford **14**. During the dehydration of the hemiacetal intermediate, the oxygen atom that originated from molecular oxygen is removed. Thus, this mechanism agrees as well with the labeling studies.[256] The analogous mechanism was proposed, together with a di-radical mechanism, in the ring coupling reaction by the cytochrome P450 OxyB in the biosynthesis of the vancomycin family of antibiotics.[261, 262]

GsfE, converting the penultimate **18** to **1,** shares an overall tertiary structure to progesterone 5β-reductase (POR)*,* the key enzyme in the biosynthesis of plant cardenolides such as digitoxin. POR was recently characterized to belong to a novel family of short-chain dehydrogenase/reductase and the authors suggested members of this family may not be exclusively found in plants.[255] GsfE, which catalyzes a similar 1,4-type hydride addition to a cyclic enone moiety, gives additional proof that enzymes belonging to this family are indeed short-chain dehydrogenases/reductases rather than nucleoside-sugar epimerases as they are commonly annotated in databases. An enzymatic reaction mechanism for GsfE, as shown in Figure 4.7B, can be proposed: A formal hydride from NADPH performs a pro-*R*, 1,4 (Michael)-type addition to the cyclohexadienone ring of **18**. The resulting Michael addition product undergoes keto-enol tautomerizaton to afford **1**.

This study, through a combination of genetic and enzymatic methodologies, have characterized and reconstituted all of the reactions leading to **1**, culminating in the total *in vitro* biosynthesis of the complex natural product. The classical methodologies in elucidation of biochemical pathways such as feeding of radiolabeled precursors and isolation of intermediates from wild-type producers provide extremely valuable knowledge to the current field of natural product biosynthesis.[263] This is demonstrated by the work done by Rhodes *et al* on **1,** which deciphered key intermediates and formulated a sound hypothesis of the timing of the tailoring cascades without knowledge of genes in organism.[245-247, 256] Nevertheless, contemporary methodologies such as gene deletion of candidate biosynthetic genes identified through bioinformatic analysis and *in vitro* assay of heterologously expressed biosynthetic enzymes have opened new avenues in elucidating biosynthetic pathways as demonstrated by the discovery of the interesting reactions within this pathway, such as the unorthodox C1-C6 and C8-C13 cyclization of **5a** by GsfA; and P450-mediated ring coupling reaction of the orcinol and phloroglucinol rings of **11** to form **14**. . In conclusion, we performed comprehensive single-gene deletion of the *gsf* genes and functional characterization of the corresponding enzymes to elucidate the biosynthetic pathway of **1**. Several grisan analogues of **1** were isolated from the different gene deletion mutants (Scheme 3.1), hinting at the overall flexibility of the pathway and its potential for biosynthetic engineering. Moreover, this study has uncovered the novel and interesting reactions within this pathway, such as the unorthodox C1-C6 and C8-C13 cyclization of **5a** by GsfA, and P450-mediated ring coupling reaction of the orcinol and phloroglucinol rings of **11** to form **14**. These two reactions warrant further mechanistic investigations that may possibly lead to a better understanding of the biosynthesis of natural products with similar features.

Section 5: Conclusion

The work presented in this dissertation provided critical insights towards understanding the biosynthesis of two important fungal secondary metabolites echinocandin B and griseofulvin. Echinocandin B is an antifungal nonribosomal peptide that is used a precursor for the semisynthesis of the frontline drug for treatment of candidiasis Anidulafungin. On the other hand, griseofulvin is an aromatic polyketide currently being used as treatment against fungal dermatophytes. Using a combination of genetic and enzymatic characterization, we have assigned the functions of the genes involved in key steps towards the biosynthesis of both compounds.

Firstly, we have sequenced and assembled the genome of the organism *Emericella rugulosa*, the producer organism of echinocandin B. Bioinformatic analysis of the genome of the organism revealed the presence of twenty-six NRPS genes in the organism. By comparing the NRPS genes of *E. rugulosa* with the closely related strain *A. nidulans*, which does not produce echinocandin B, we have discovered the gene for echinocandin synthetase *ecdA*. Knockout of the *ecdA* gene and biochemical characterization of the first adenylation domain of EcdA confirmed the function of EcdA in assembling the echinocandin antibiotic scaffold.

This work also describes the biochemical characterization of the fatty-acyl-AMP ligase EcdI and its role in the priming of the EcdA assembly-line with linoleic acid. Based on the *in vitro* reconstitution of the enzymatic reaction, EcdI activates linoleic acid as linoleyl-AMP prior to its transfer to the T_0 domain of EcdA. This work also investigated the fidelity of EcdI to recognize linoleic acid over noncognate fatty acid substrate. Using the relationship of EcdA and EcdI, we have also discovered putative lipopeptide biosynthetic gene clusters in other fungi.

This work also describes the discovery of the *hty* gene cluster, a separate from the *ecd* gene cluster involved in the biosynthesis of the non-proteinogenic amino acid L-homotyrosine. Knockout of the isopropylmalate synthase homolog gene *htyA* and subsequent rescue of the production of Echinocandin B in the Δ*htyA* verified the role of the cluster in the biosynthesis of L-homotyrosine.

Section 3 describes the work in characterization of the three mononuclear iron-dependent oxidases in the *ecd* gene cluster. First, we have assigned the function of the cytochrome P450 EcdH in the hydroxylation of C4 and C5 of L-ornithine using a combination of NMR characterization of the shunt metabolites from Δ*ecdH* mutant and feeding of perdeuterated L-ornithine to the Δ*ecdH* mutant. EcdH is proposed to install the hydroxyl groups in L-ornithine after the echinocandin scaffold is assembled by EcdA. Moreover, the role of EcdG in the hydroxylation of C3 of L-homotyrosine was also characterized and reconstituted by this study. The bioactivity against *Candida albicans* of the shunt metabolites isolated from the two *E. rugulosa* mutants were also characterized in this study. Finally section 3 describes the role of EcdK in conversion of L-leucine to the nonproteinogenic amino acid L-methylproline.

This work also presents (section 4) the complete characterization of the enzymatic steps towards the biosynthesis of griseofulvin. This study has reconstituted the formation of norlichexanthone by the non-reducing polyketide synthase (NR-PKS) GsfA. The study showed that GsfA is capable of catalyzing Claisen-type cyclization and product release without the aid of a thioesterase domain. The roles of methyltransferases GsfB and GsfC in methylation of the GsfA product and prevention of the closure of the xanthone ring was also verified. In vitro reconstitution of the GsfI reaction also established the timing of the chlorination of griseofulvin to be prior to the formation of the grisan ring.

This study also characterized the function of the cytochrome P450 GsfF in the biosynthesis of griseofulvin, in particular, its role in the formation of the grisan ring moiety in the final compound. Reconstitution of the GsfF reaction confirmed its role in the conversion of Griseophenone B to the grisan compound desmethyldehydrogriseofulvin A. The final tailoring reactions in the biosynthesis of griseofulvin by the *O*-methyltransferase GsfD and the NADPH-dependent reductase GsfE was also characterized in this study. Total *in vitro* biosynthesis of the griseofulvin was achieved using the enzymes GsfA-F and GsfI as a culmination of section 4.

Section 6: Appendix

6.1 Supplemental Information for Section 2

Table A1: A domain selectivity of EcdA A domain as predicted by NRPSpredictor1 and NRPS predictor2 algorithms.

	amino acid in **1**	10 aa code	Bacterial		Fungal	
			NRPS1	NRPS2	NRPS1	NRPS2
A1	L-ornithine	DVMELSSITK	hpg	dab	hpg	hpg
A2	L-threonine	DAQTAVAIHK	phe	phe	phe	ala
A3	L-proline	DVSSATTVCK	no prediction	trp	no prediction	pro
A4	L-homotyrosine	DGEAVGCVFK	val,leu	ile	val	leu
A5	L-threonine	DAQTIVAIHK	phe	phe	phe	Ala
A6	4-methyl-L-proline	DNTMITAMSK	no prediction	asn	no prediction	tyr

Hpg = hydroxyphenylglycine, dab= di-aminobutyric acid, phe = L-phenylalanine, ala= L-alanine, pro = L-proline, trp = L-tryptophan, val = L-valine, leu = L-leucine, ile = L-isoleucine, asn = L-asparagine, tyr = L-tyrosine

Table A2[:] ^{1}H NMR and 2D-NMR (HSQC, HMBC, gCOSY) data of six amino acid residues in Echinocandin B **1** (600 MHz) in CD_3OD. The parameters in red are from pneumocandin $A_{0[135]}$ and parameters in blue are from ^{13}C-NMR spectrum of **1** from literature as comparison[264]. COSY and HMBC correlation signals that are used to identify **1** are shown in the structure.

4R-OH, 5R-OH-L-ornithine$_1$

C-1	174.1	174.6	174.0			
C-2	51.3	51.4	52.8	2-H	4.39	4.43
C-3	34.8	34.8	36.73	3-Ha	2.07	2.00
				3-Hb	1.93	2.00
C-4	70.6	70.6	70.87	4-H	3.94	4.00
C-5	74.2	74.0	74.3	5-H	5.25	5.26

L-threonine₂

C-1	172.6	172.7	172.2		
C-2	58.4	58.4	58.6	2-H	4.95
C-3	68.1	68.2	68.3	3-H	4.51
C-4	19.5	19.7	20.1	4-H	1.20

Table S4: L-threonine₅

C-1		172.7	172.2		
C-2	56.7	58.4	56.6	2-H	4.84
C-3	69.5	68.2	69.42	3-H	4.15
C-4	19.5	19.7	19.5	4-H	1.21

4-hydroxy-L-proline₃

C-1	173.3	173.4	173.2		
C-2	62.3	62.5	62.3	2-H	4.58
C-3	38.3	38.5	38.78	3-Ha	2.42
				3-Hb	2.04
C-4	71.1	71.3	70.87	4-H	4.50
C-5	56.8	57.1	56.6	5-Ha	3.96
				5-Hb	3.79

3,4-dihydroxy-L-homotyrosine₄

C1		172.5	169.7			
C-2	56.1	56.4	56.6	2-H	4.32	4.31
C-3	76.7	76.9	76.8	3-H	4.24	4.27
C-4	75.4	75.8	75.51	4-Ha	4.30	4.28
C-1'	132.9	133.0	132.5			
C-2'/C-6'	129.5	129.6	129.5	2'-H/6'-H	7.14	7.13
C-3'/C-5'	116.0	116.2	116.1	3'-H/5'-H	6.76	6.75
C-4'	158.3	158.5	158.2			

3-Hydroxy-4-methyl-L-proline[6]

C-1		172.7	172.5			
C-2	69.3	70.2	69.42	2-H	4.34	4.36
C-3	75.4	75.9	75.51	3-H	4.19	4.14
C-4	38.8	39.1	38.78	4-H	2.52	2.48
C-5	52.7	53.0	55.9	5-Ha	3.86	4.09
				5-Hb	3.38	3.34
4-CH₃	11.1	11.1	11.3	4-CH₃	1.05	1.04
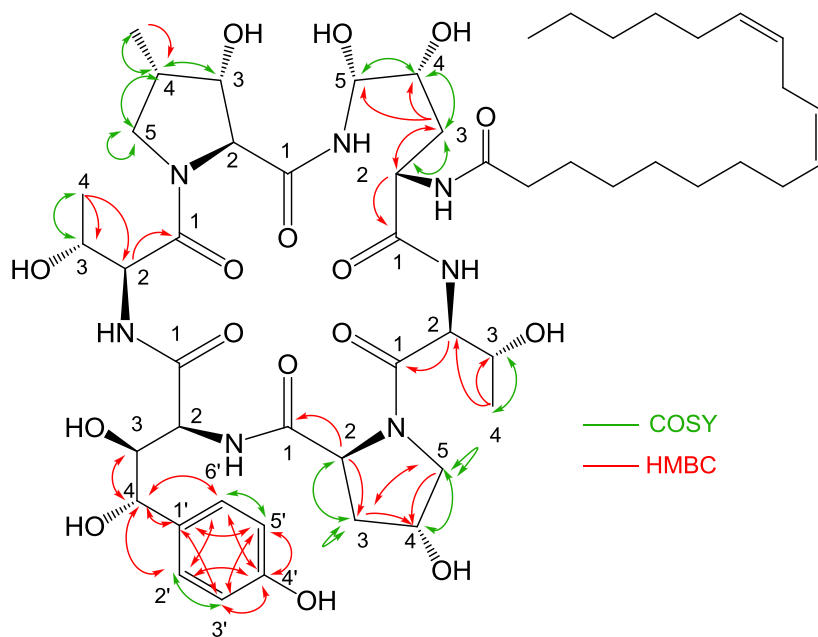
Table A3: Pyrroline-5-reductase genes in *E.rugulosa.*

Annotation	Closest Homolog	Identity/Similarity
ErP5CR65	AN9279, *A.nidulans* A4	94%, 96%
ErP5CR54	AN7387, *A.nidulans* A4	92%, 93%
ErP5CR11	AN4355, *A.nidulans* A4	93%, 96%
ErP5CR150	AN6025, *A.nidulans* A4	92%, 95%

Table A4: EcdI homologs that are clustered together with NRPS with initiation T domains.

Identity and similarity percentages of EcdI homologs in comparison to EcdI sequence are given.

EcdI blast hits	NRPS in the region (domain organization)	PKS in region (domain organization)
PMAA_001090 (45%, 63%) *Penicillium marneffei*	PMAA_001110 (TCATCATCATCT)	PMAA_001080 (KS-AT-DH-Met-ER-KR-ACP)
AN2549 (46%, 63%) *A.nidulans* A4	AN2545 (easA) (TECATECATCATCAT)	EasB (KS-AT-DH-Met-ER-KR-ACP)
TSTA_098290 (45%, 63%) *Talaromyces stipitatus* ATCC 10500	TSTA_098310 (TCATCATCATCT)	TSTA_098280 (KS-AT-DH-Met-ER-KR-ACP)
FOXB_07957 (43%, 53%) *Fusarium oxysporum* Fo5176	FOXB_07954 (TCATCATCATCATC)	None
EGC42542 (41%, 51%) *Ajellomyces capsalatus* H88	EGC42544 (TCATCATCATCTCAATC)	EGC42541 (KS-AT-DH-Met-ER-KR-ACP)
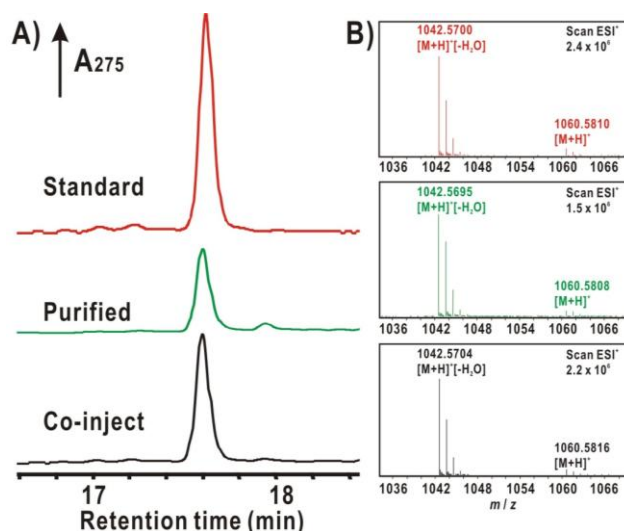
Figure A1: Chromatographic traces at 275 nm (A) and mass spectra (B) of authentic Echinocandin B standard (top, red), Echinocandin B purified from fermentation broth of *E.rugulosa* NRRL 11440 culture (middle, green), and co-injection of the two (bottom, black). The theoretical mass-to charge ratio of Echinocandin B **1** is 1060.5813 for $[C_{52}H_{81}N_7O_{16}+H]^+$ and 1042.5707 for $[C_{52}H_{81}N_7O_{16}+H-H_2O]^+$.
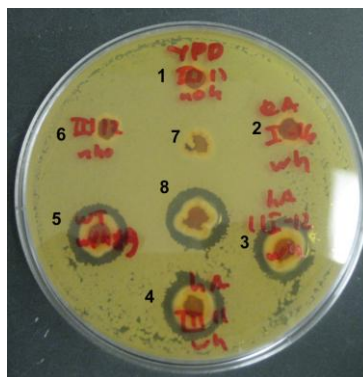
Figure A2: Anti-*Candida* assay performed on Δ*htyA*-III-11 without(1) and with(4) L-homotyrosine feeding , Δ*htyA*-III-12 without(6) and with(3) L-homotyrosine, on Δ*ecdA*-I-16 with(2) and without(7) L-homotyrosine and wild-type *E.rugulosa* with(colony 5) and without(8) L-homotyrosine. Anti-*Candida* bioassay was performed by feeding L-homotyrosine (0.1 mg/mL) to the static liquid cultures of Δ*htyA* and Δ*ecdA* mutants at day-4 before transferring the mycelial discs from individual clones to the YPD plate pre-inoculated with *C. albicans* at day-7
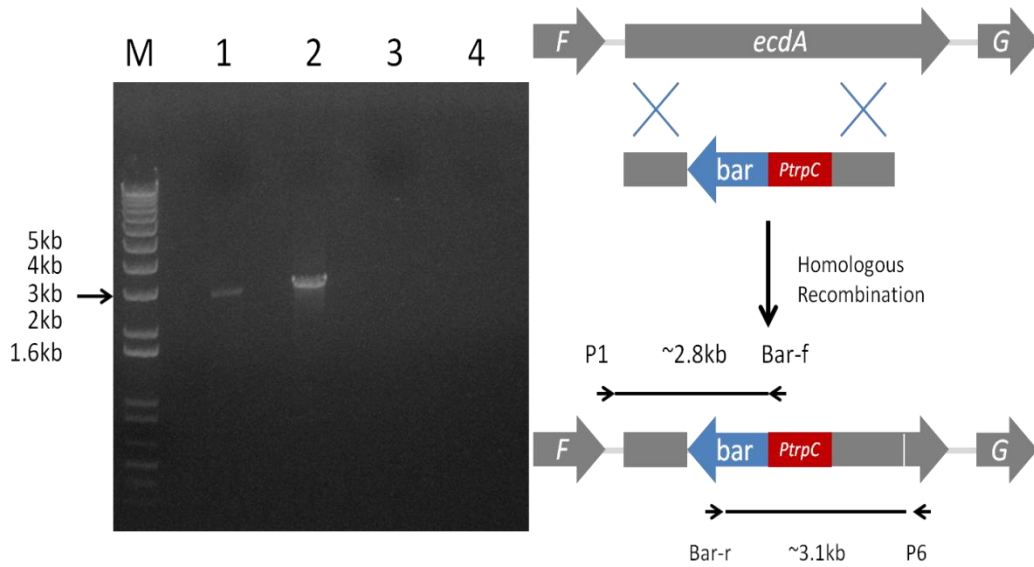
Figure A3: PCR screening of Δ*ecdA*-I-16. (M)1 kb plus ladder (Invitrogen). Amplification of ΔecdA-I-16 genomic DNA (gDNA) using primer pair P1 and bar-f showing expected ~2.8kb amplicon (lane 1) and using primer pair ecdA-KO-P6 and bar-r showing expected ~3.1 kb amplicon (lane 2). Amplification of wild-type *E.rugulosa* gDNA using primer pair ecdA-KO-P1 and bar-f (lane 3) and using primer pair ecdA-KO-P6 and bar-r (lane 4)
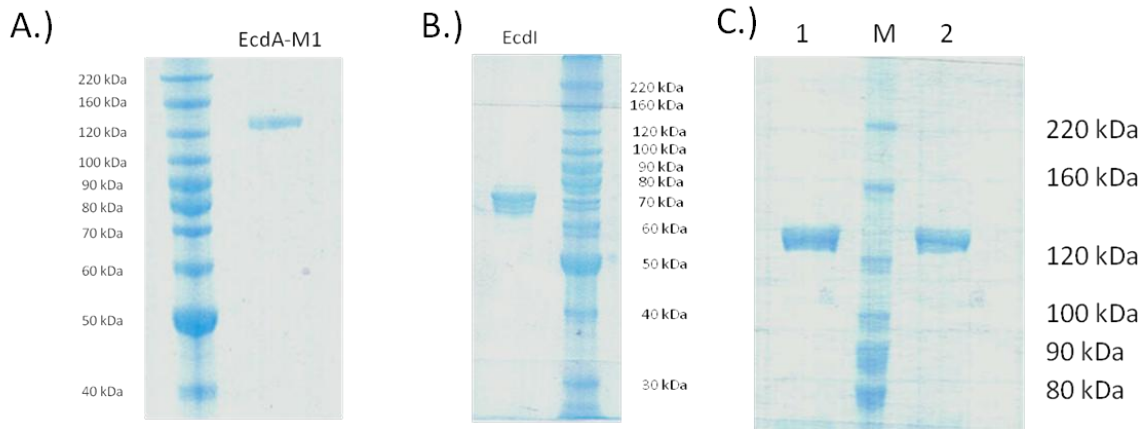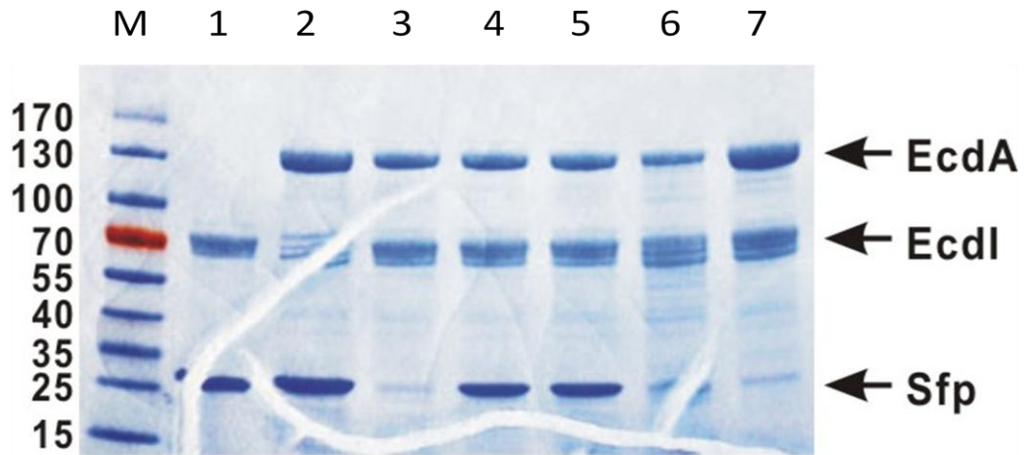

Figure A4:A.) SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1 using a Nickel-NTA resin showing the expected size of 130 kDa. **B.)** SDS-PAGE gel of purified N-terminal-His tagged EcdI using Nickel-NTA resin showing expected size of 70 kDa. **C.)** SDS-PAGE gel of purified N-terminal-His tagged EcdA-M1 variants using a Nickel-NTA resin showing the expected size of 130 kDa. **Lane 1**: S47A variant of EcdA-M1 (T_0*CAT_1). **Lane 2**: S1127A variant of EcdA-M1 (T_0CAT_1)

A)

| Sample	1	2	3	4	5	6	7
EcdA	−	+	+	+	+	$T_0{*}CAT_1$	$T_0CAT_1{*}$
EcdI	+	−	+	+	+	+	+
Sfp	+	+	−	+	+	holo-	holo-
ATP	+	+	+	−	+	+	+
*LA	+	+	+	+	+	+	+

B)

C)

.
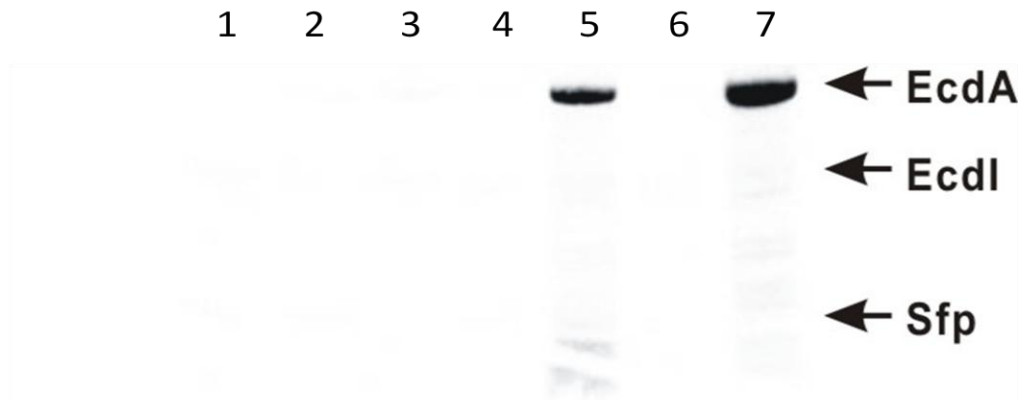
Figure A5: A) The components of the assay mixture used in the radiography B) SDS-PAGE gel from EcdA-M1 ^{14}C-linoleic acid loading assay showing the protein present in the assay mixture and C)its matching radiogram used for Figure 3A of main text. EcdA variants S47A and S1127A are in *holo-* form due to expression in *E.coli* BAP1.

Sample	1	2	3	4
EcdA	+	+	T_0*CAT_1	T_0CAT_1*
Sfp	+	+	*holo-*	*holo-*
ATP	−	+	+	+
*L-Orn	+	+	+	+
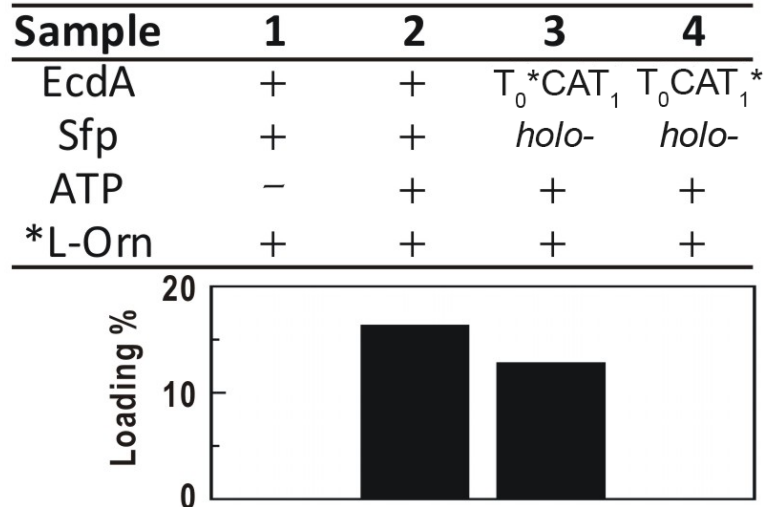
Figure A6: Loading assay of EcdA-M1 with ^{14}C-L-ornithine showing incorporation of [14C]-L-ornithine to wild-type EcdA-M1 and T_0*CAT_1 (S47A) but not to T_0CAT_1* (S1127A) as expected. EcdA variants S47A and S1127A are in *holo-* form due to expression in *E.coli* BAP1.
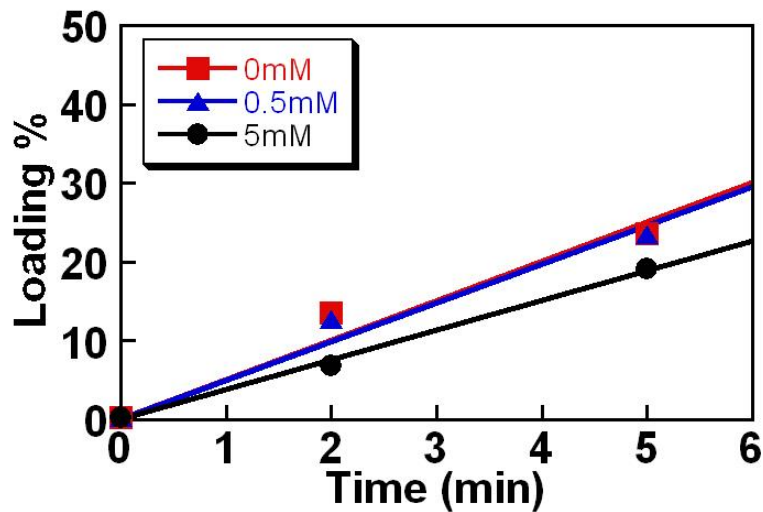

Figure A7: No apparent effect of [CoA] on the reaction rate of loading of linoleic acid onto EcdA-M1 by EcdI. After the conversion of *apo* to *holo* form, the residual CoA is removed by buffer exchange through a ultracentrifugal filter (Molecular weight cutoff of 10 kDa). 0, 0.5, or 5 mM CoA was added afterwards and the loading percentages were measured at 2 and 5 min.
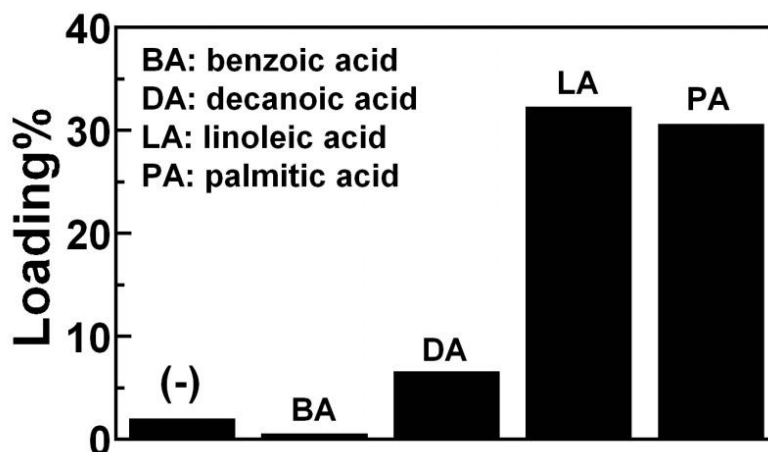
Figure A7: Loading of alternative substrates onto EcdA-M1 by EcdI. The assay procedure was described in Materials and Methods. The reactions were run at ambient temperature for 30 min and subject to quantification. The negative control sample contains linoleic acid but no addition of ATP.
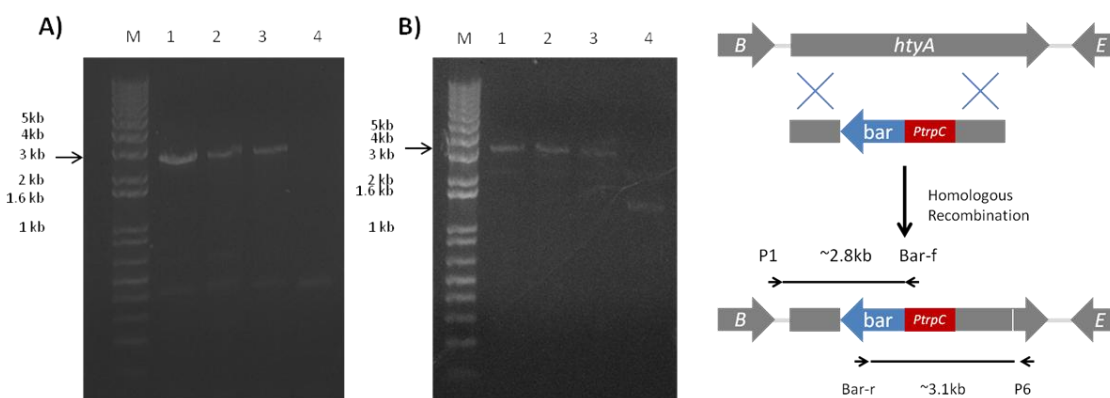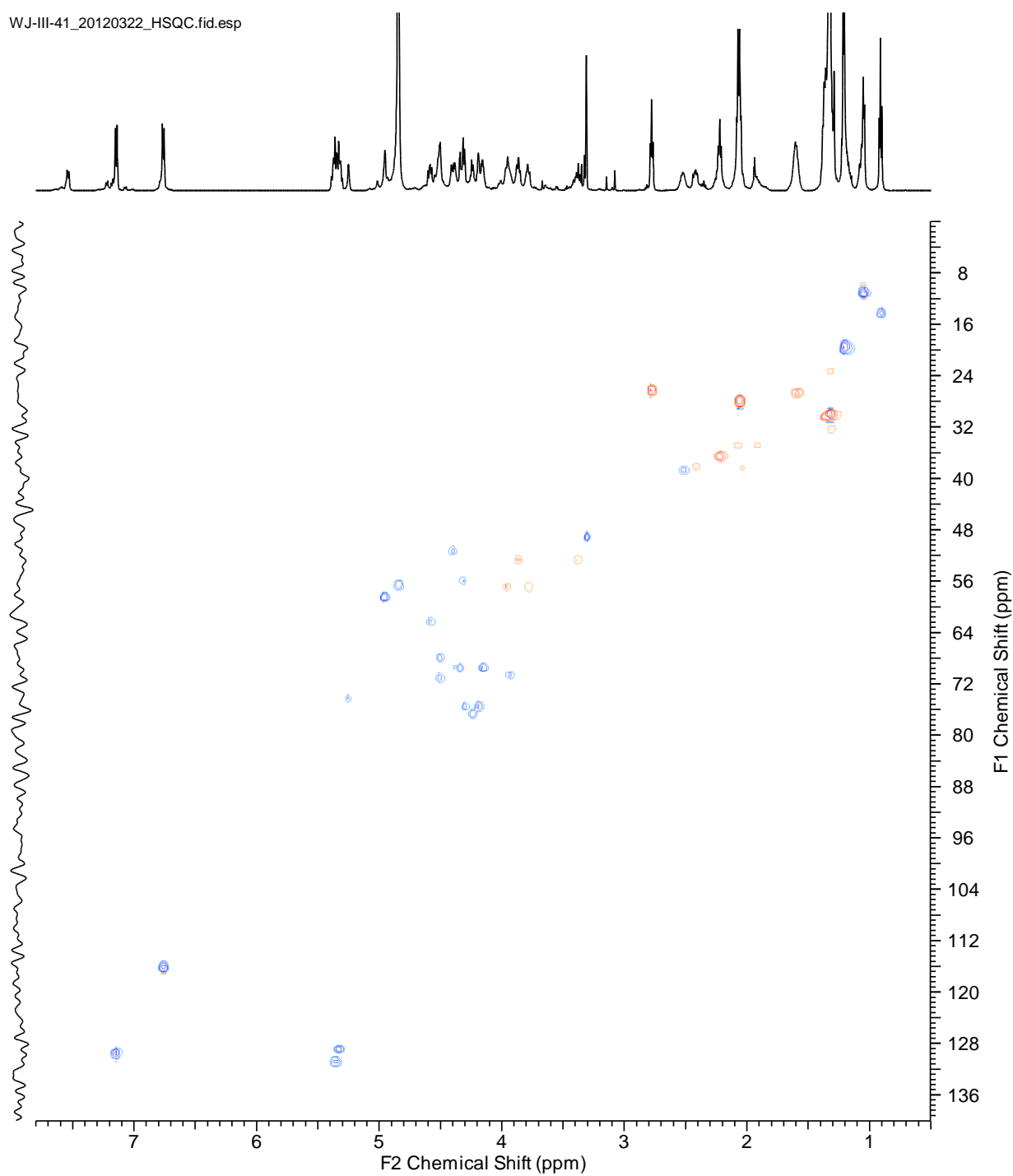

Figure A8: PCR screening of Δ*htyA* mutants **A.)** Amplification of gDNA from Δ*htyA* III-7 (lane 1), III-11(lane 2), III-12(lane 3),and wild-type (lane 4) using primer pairs htyA-KO-P1 bar-f. Δ*htyA* mutants show the expected ~2.8 kb amplicon **B.)** Amplification of gDNA from of (lane 1) Δ*htyA* III-7, (lane 2) III-11, (lane 3) III-12, and (lane 4)wild-type using primer pairs htyA-KO-P6 bar-r. Δ*htyA* mutants show the expected ~3.1 kb amplicon
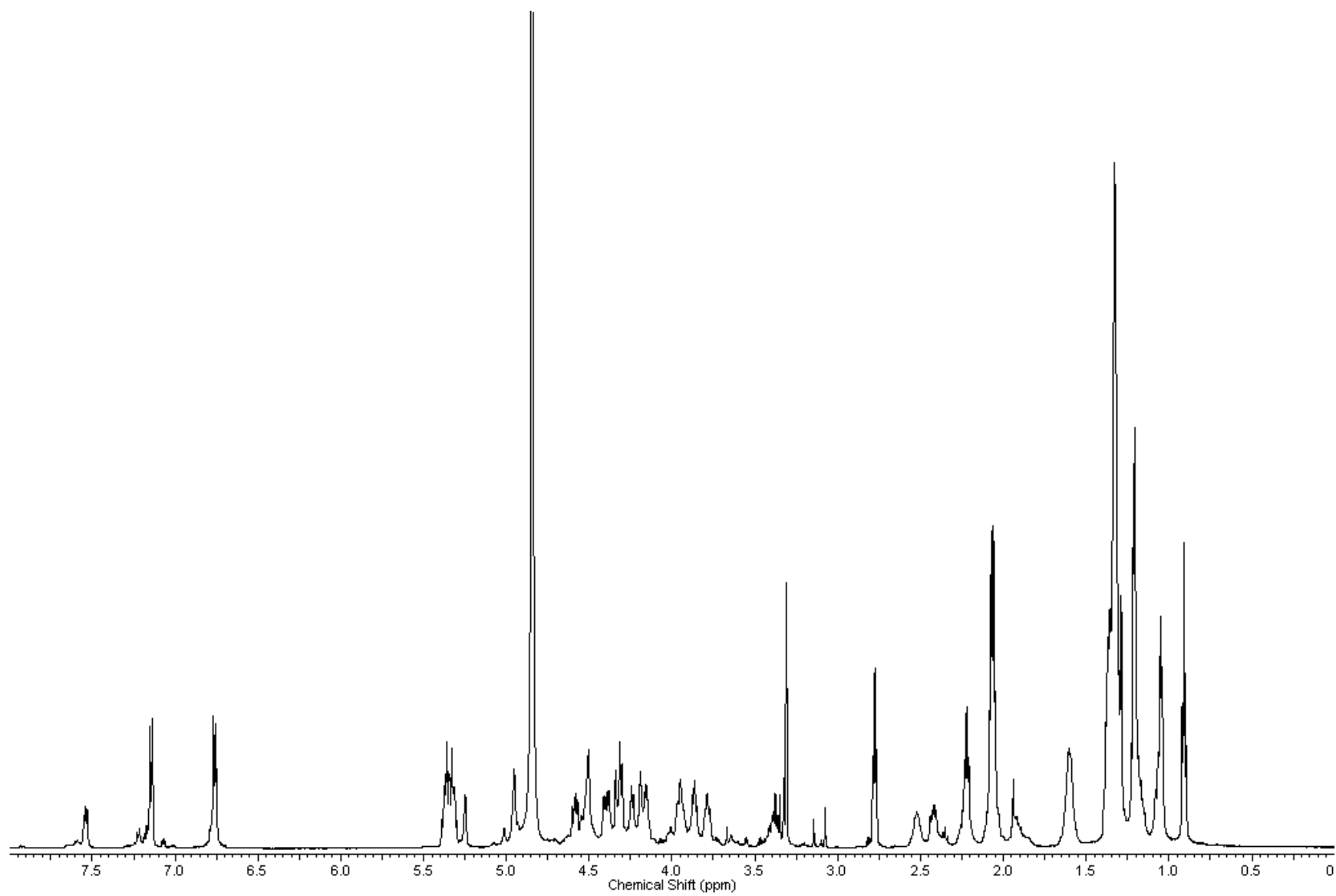
Figure A9: ^{1}H spectrum of **1** (600 MHz) in CD$_3$OD

Figure A11: HSQC Spectrum of **1** (600 MHz) in CD$_3$OD

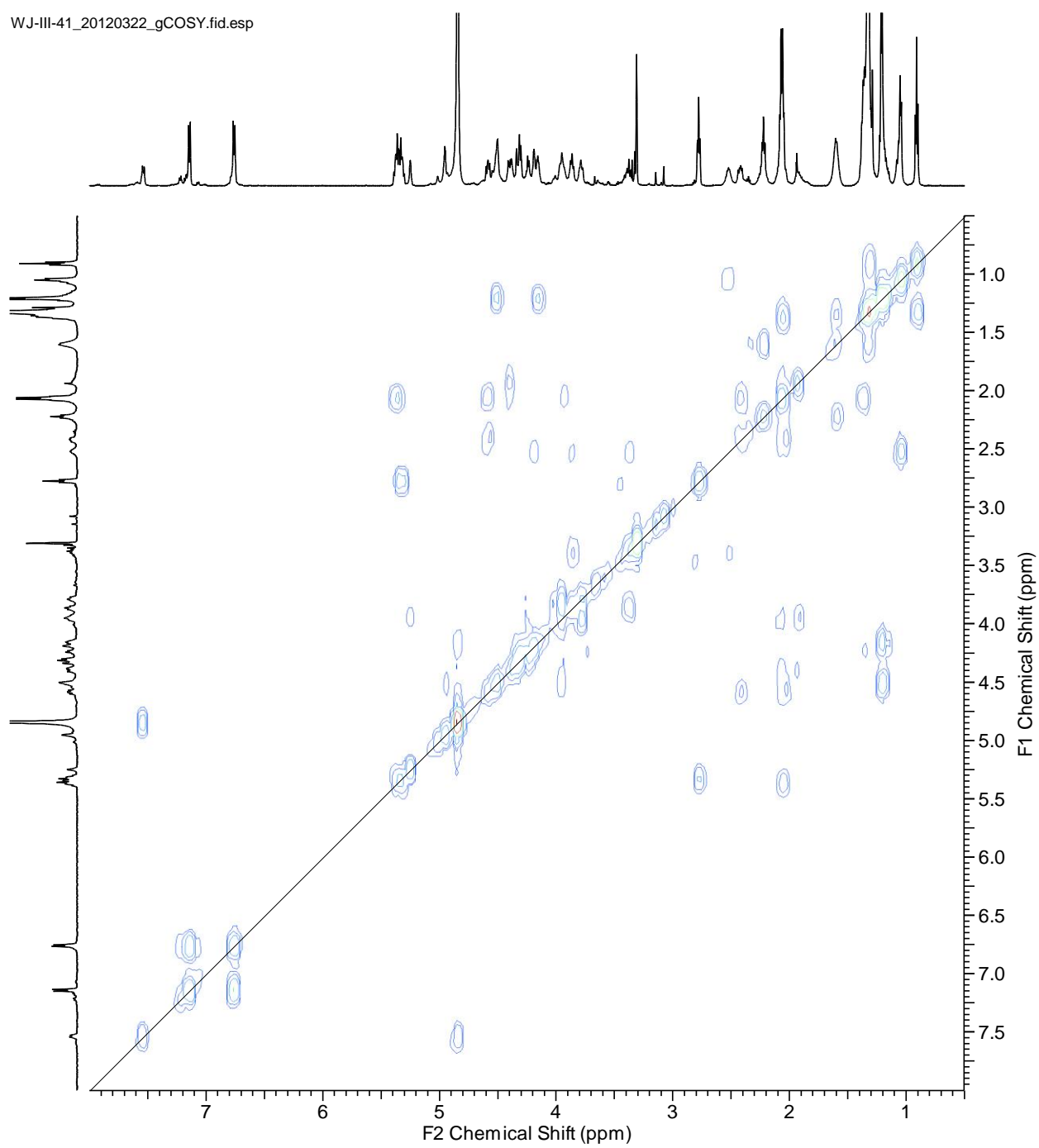WJ-III-41_20120322_gCOSY.fid.esp

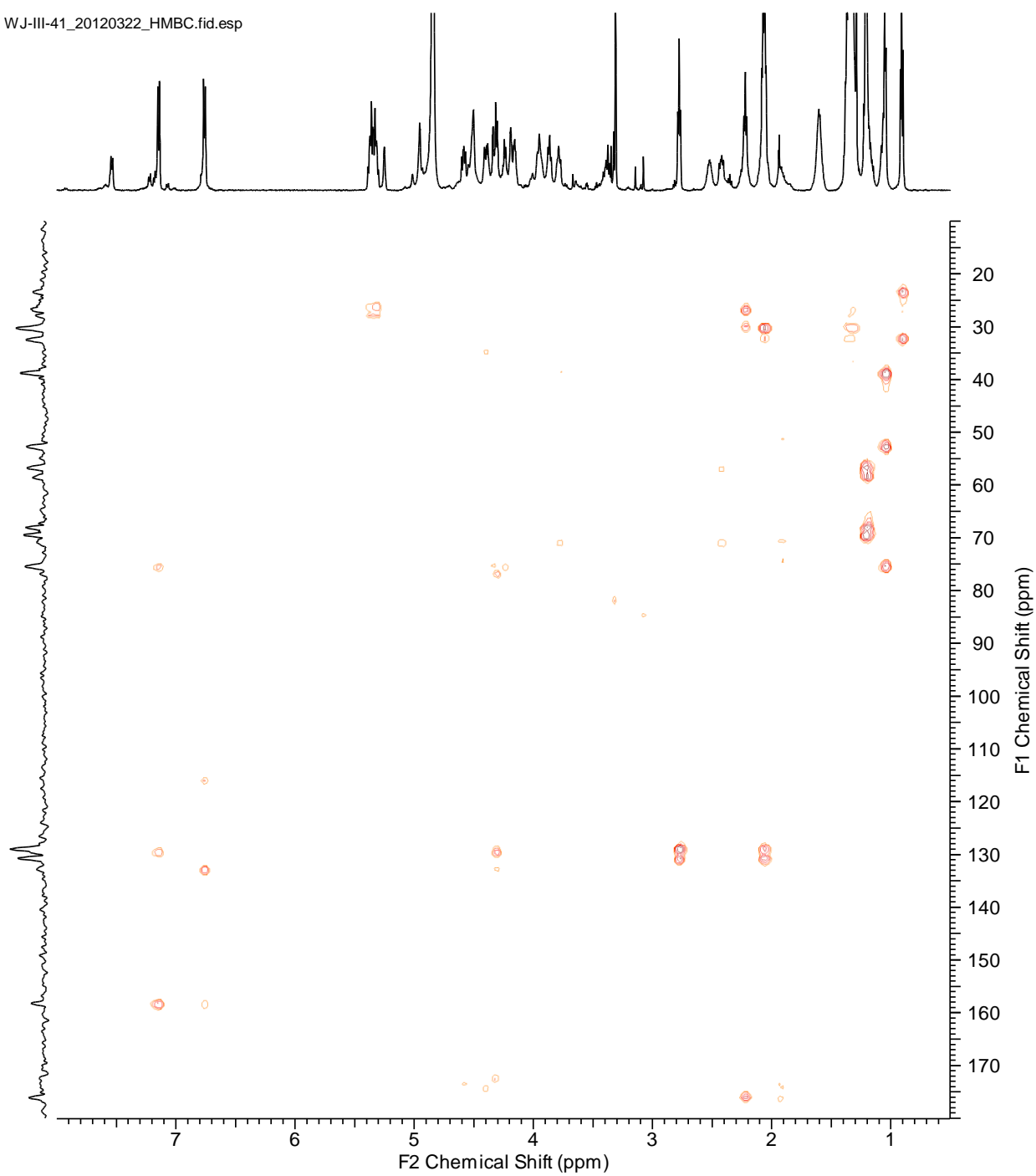WJ-III-41_20120322_HMBC.fid.esp

6.2 Supplemental Information for Section 3

Table A5: ^{13}C and ^{1}H NMR data for Echinocandin B **1** (from WT), D **3** (fromΔ*ecdH* variant), **6** and **7** (from Δ*ecdG* variant) in CD$_3$OD (600 MHz). The signals most useful for indicating structural changes are highlighted.
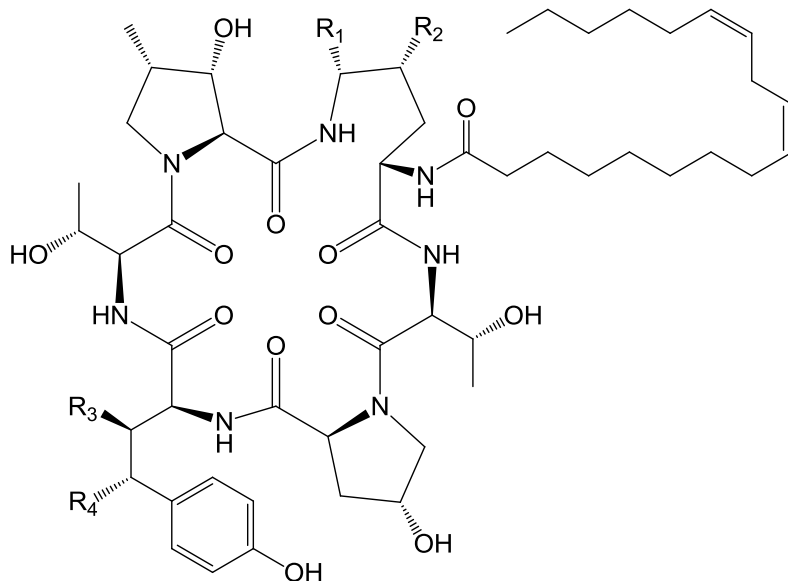

	WT 1	Δ*ecdH* (tri) 3	Δ*ecdG* (di) 6	Δ*ecdG* (tri) 7		WT 1	Δ*ecdH* (tri) 3	Δ*ecdG* (di) 6	Δ*ecdG* (tri) 7
4,5-dihydroxyornithine (DiOHOrn)									
C-2	51.3	52.8	51.5	52.1	2-H	4.39	4.39	4.41	4.37
C-3	34.8	27.7	34.8	26.8	3-Ha	2.07	1.55	2.07	1.65
					3-Hb	1.93	2.12	1.93	1.65
C-4	70.6	24.6	70.6	30.8	4-Ha	3.94	1.71	3.90	2.00
					4-Hb	--	1.71	--	1.73
C-5	74.2	37.8	74.4	79.3	5-Ha	5.25	2.97	5.25	5.10
					5-Hb	--	3.48	--	--
Threonine A (ThrA)									
C-2	58.4	58.2	58.5	58.5	2-H	4.95	4.91	4.92	4.96
C-3	68.1	67.9	68.5	68.8	3-H	4.51	4.48	4.45	4.46
C-4	19.5	19.5	19.2	19.2	4-H	1.20	1.23	1.23	1.22
Threonine B (ThrB)									
C-2	56.7	57.0	55.9	55.7	2-H	4.84	4.91	4.85	4.89
C-3	69.5	69.1	69.0	69.1	3-H	4.15	4.26	4.17	4.14
C-4	19.5	20.1	20.3	20.3	4-H	1.21	1.19	1.18	1.19

	WT **1**	Δ*ecdH* (tri) **3**	Δ*ecdG* (di) **6**	Δ*ecdG* (tri) **7**		WT **1**	Δ*ecdH* (tri) **3**	Δ*ecdG* (di) **6**	Δ*ecdG* (tri) **7**
4-hydroxy-proline (OHPro)									
C-2	62.3	62.2	61.8	61.9	2-H	4.58	4.65	4.55	4.53
C-3	38.3	38.8	38.4	38.1	3-Ha	2.42	2.45	2.33	2.33
					3-Hb	2.04	2.08	1.77	1.75
C-4	71.1	71.1	70.9	71.0	4-H	4.50	4.58	4.48	4.47
C-5	56.8	57.0	56.8	56.7	5-Ha	3.96	4.00	3.76	3.76
					5-Hb	3.79	3.81	3.74	3.72
3,4-dihydroxyhomotyrosine (diOHTyr)									
C-2	56.1	57.6	54.7	54.5	2-H	4.32	4.44	4.38	4.48
C-3	76.7	74.0	34.9	35.5	3-Ha	4.24	4.41	2.20	2.20
					3-Hb	--	--	2.07	1.99
C-4	75.4	40.8	32.6	32.4	4-Ha	4.30	2.65	2.59	2.59
					4-Hb	--	2.57	2.59	2.59
C-1'	132.9	129.5	132.8	132.7					
C-2'/C-6'	129.5	131.2	130.2	130.2	2'-H/6'-H	7.14	7.01	7.01	7.01
C-3'/C-5'	116.0	116.0	116.0	116.0	3'-H/5'-H	6.76	6.71	6.70	6.70
C-4'	158.3	156.9	156.4	156.5					
3-hydroxy-4-methylproline (OHMePro)									
C-2	69.3	69.7	69.3	69.3	2-H	4.34	4.32	4.37	4.28
C-3	75.4	75.8	75.2	76.0	3-H	4.19	4.14	4.21	4.17
C-4	38.8	38.8	38.8	38.8	4-H	2.52	2.48	2.55	2.56
C-5	52.7	52.7	52.6	52.7	5-Ha	3.86	3.85	3.88	3.88
					5-Hb	3.38	3.38	3.40	3.42
4-CH3	11.1	11.1	11.1	11.1	4-CH3	1.05	1.05	1.08	1.08

Table A6: ^{1}H NMR data of trideoxy compound from Δ*ecdH* (black) and reported Echinocandin

D **3** (red) in CD$_3$OD are consistent.

|	Δ*ecdH* (tri)	Echinocandin D **3**		Δ*ecdH* (tri)	Echinocandin D **3**
4,5-dihydroxyornithine (DiOHOrn)			4-hydroxy-proline (OHPro)		
2-H	4.39	4.40-4.36	2-H	4.65	4.64
3-Ha	1.55	1.62-1.51	3-Ha	2.45	2.49-2.43
3-Hb	2.12	2.15-2.04	3-Hb	2.08	2.15-2.04
4-Ha	1.71	1.72-1.66	4-H	4.58	4.56
4-Hb	1.71	1.72-1.66	5-Ha	4.00	4.00
5-Ha	2.97	2.99-2.94	5-Hb	3.81	3.80
5-Hb	3.48	3.49-3.43	3,4-dihydroxyhomotyrosine (diOHTyr)		
Threonine A (ThrA)			2-H	4.44	4.41
2-H	4.91	4.89	3-Ha	4.41	4.40-4.36
3-H	4.48	4.48-4.44	4-Ha	2.65	2.64
4-H	1.23	1.21	4-Hb	2.57	2.55
Threonine B (ThrB)			2'-H/6'-H	7.01	7.00
2-H	4.91	4.88	3'-H/5'-H	6.71	6.69
3-H	4.26	4.26-4.22	3-hydroxy-4-methylproline (OHMePro)		
4-H	1.19	1.18	2-H	4.32	4.30
			3-H	4.14	4.14
			4-H	2.48	2.49-2.43
			5-Ha	3.85	3.83
			5-Hb	3.38	3.38
			4-CH3	1.05	1.04
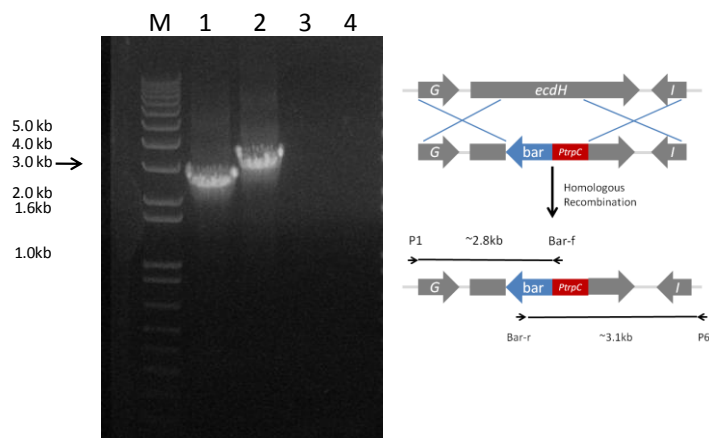
Figure A14: PCR screening of Δ*ecdH:* (M)1 kb plus ladder (Invitrogen). Amplification of Δ*ecdH* genomic DNA (gDNA) using prime pair *ecdH*-KO-P1 and bar-f showing expected ~2.8kb amplicon (lane 1). Amplification of Δ*ecdH* gDNA using primer pair *ecdH*-KO-P6 and bar-r showing expected ~3.1 kb amplicon (lane 2). Amplification of wild-type *E.rugulosa* gDNA using primer pair *ecdH*-KO-P1 and bar-f (lane 3) Amplification of wild-type *E.rugulosa* gDNA using primer pair *ecdH*-KO-P6 and bar-r (lane 4).
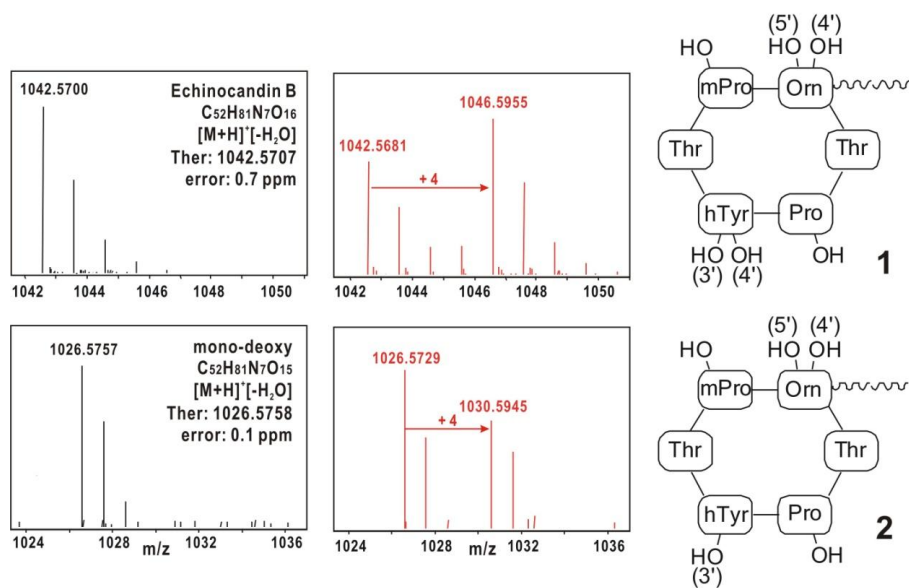

Figure A15: MS of parent ions from Echinocandin B **1** and C **2** produced by *Emericella rugulosa*, either from regular media (black) or fed with D_6-L-ornithine (red). The schematic structure of each compound is on the right.

143
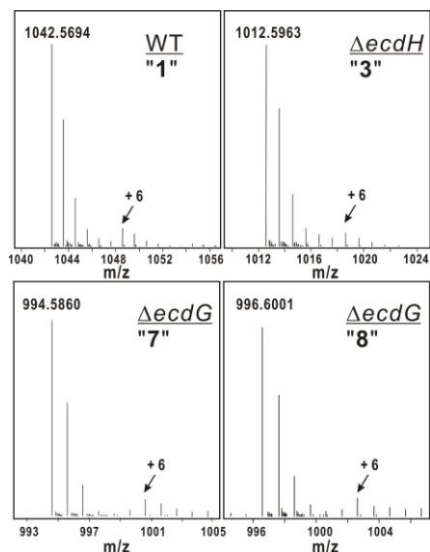
Figure A16: MS of parent ions from Echinocandin **1**, **2**, **7** and **8** produced by *Emericella rugulosa* wildtype or gene knockouts (as labeled) grown on minimal media with supplement of D_7-L-proline.
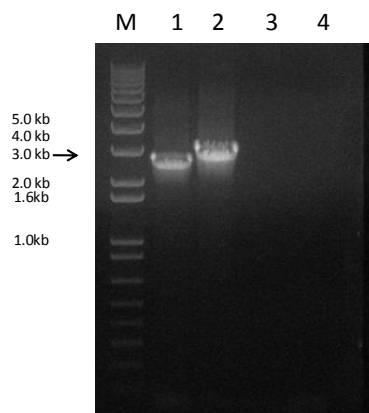
Figure A17: PCR screening of Δ*ecdG*.1 kb plus ladder (Invitrogen)(M). Amplification of Δ*ecdG* genomic DNA (gDNA) using prime pair *ecdG*-KO-P1 and bar-f showing expected ~2.8kb amplicon (lane 1). Amplification of Δ*ecdG* gDNA using primer pair *ecdH*-KO-P6 and bar-r showing expected ~3.1 kb amplicon (lane 2). Amplification of wild-type *E.rugulosa* gDNA using primer pair *ecdG*-KO-P1 and bar-f (lane 3) Amplification of wild-type *E.rugulosa* gDNA using primer pair *ecdG*-KO-P6 and bar-r (lane 4).
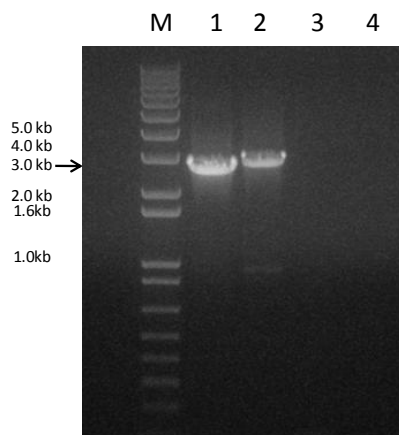
144
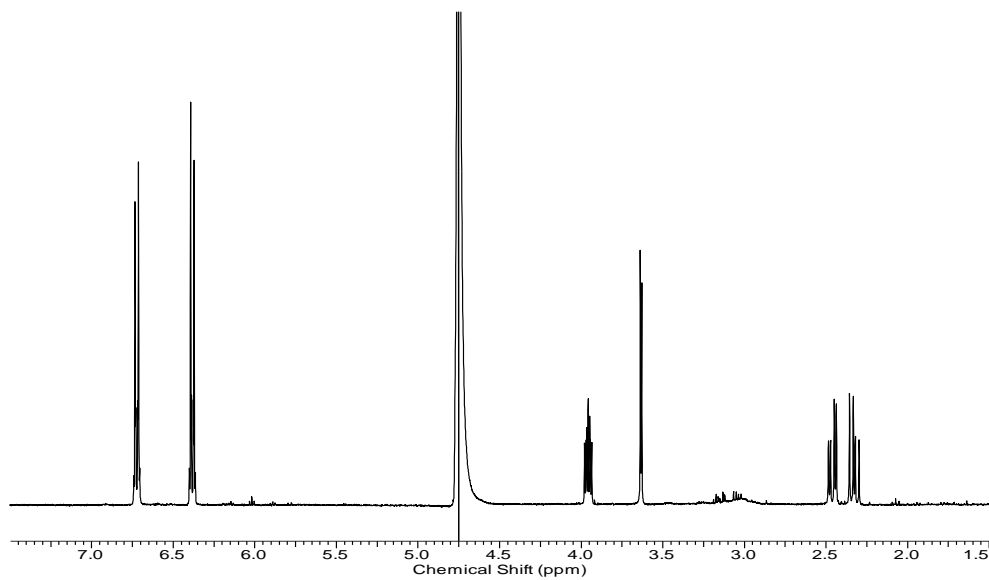
Figure A18: ^{1}H spectrum of 3-hydroxyl-homoTyr (400 MHz) in 1N DCl: δppm 6.68 - 6.76 (m, 2 H), 6.36 - 6.41 (m, 2 H), 3.96 (ddd, J=9.1, 5.2, 3.7 Hz, 1 H), 3.63 (d, J=3.9 Hz, 1 H), 2.46 (dd, J=14.1, 5.1 Hz, 1 H), 2.33 (dd, J=14.1, 9.4 Hz, 1 H)
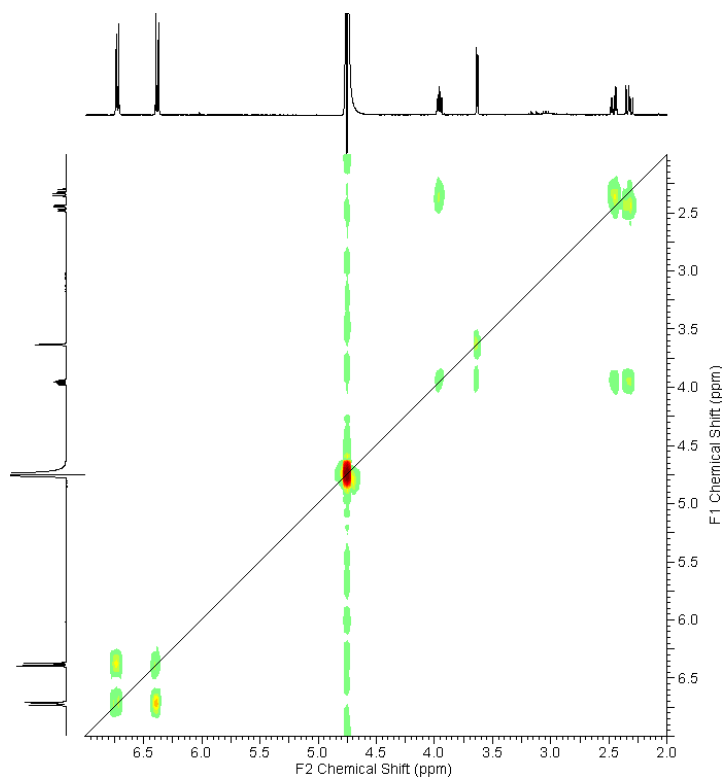

**Figure A19**: gCOSY spectrum of 3-hydroxyl-homoTyr (400 MHz) in 1N DCl.

145

Figure A20: PCR screening of Δ*ecdK*. 1 kb plus ladder (Invitrogen)(M). Amplification of Δ*ecdK* genomic DNA

(gDNA) using prime pair *ecdK*-KO-P1 and bar-f showing expected ~2.8kb amplicon (lane 1). Amplification of

Δ*ecdK* gDNA using primer pair *ecdK*-KO-P6 and bar-r showing expected ~3.1 kb amplicon (lane 2) and using

primer pair *ecdG*-KO-P1 and bar-f (lane 3) Amplification of wild-type *E.rugulosa* gDNA using primer pair *ecdG*-

KO-P6 and bar-r (lane 4)
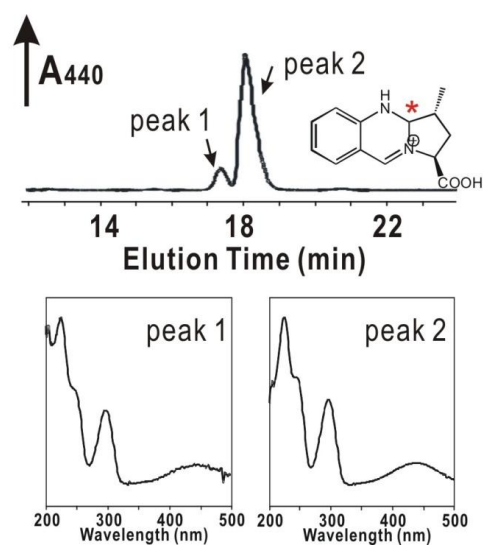
Figure A21: HPLC-UV (440 nm) trace showing two peaks of *o*-aminobenzaldehyde derivatized MeP5C

(major/minor = 6:1). They have nearly identical mass and UV-Vis spectrum, indicating a pair of diastereomers.

(define Me5PC)
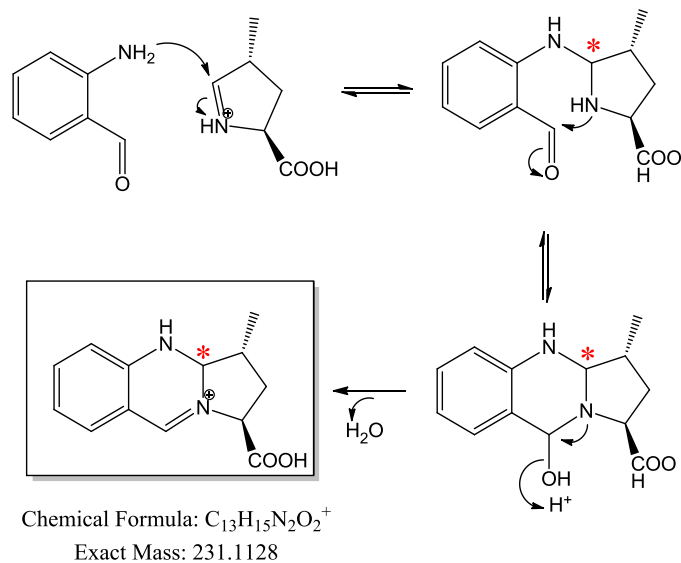
Figure A22: Proposed mechanism for *o*-aminobenzaldehyde derivatization of MeP5C.
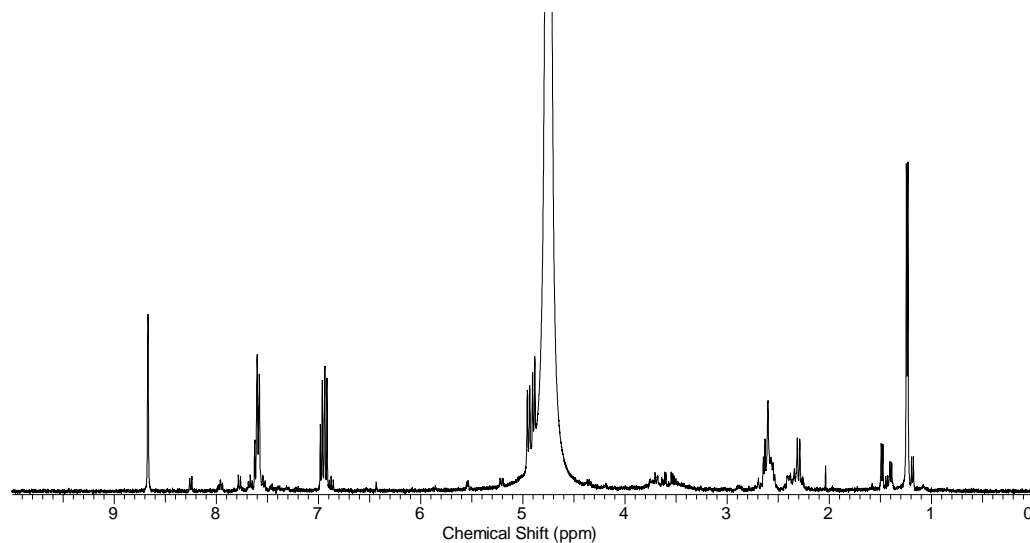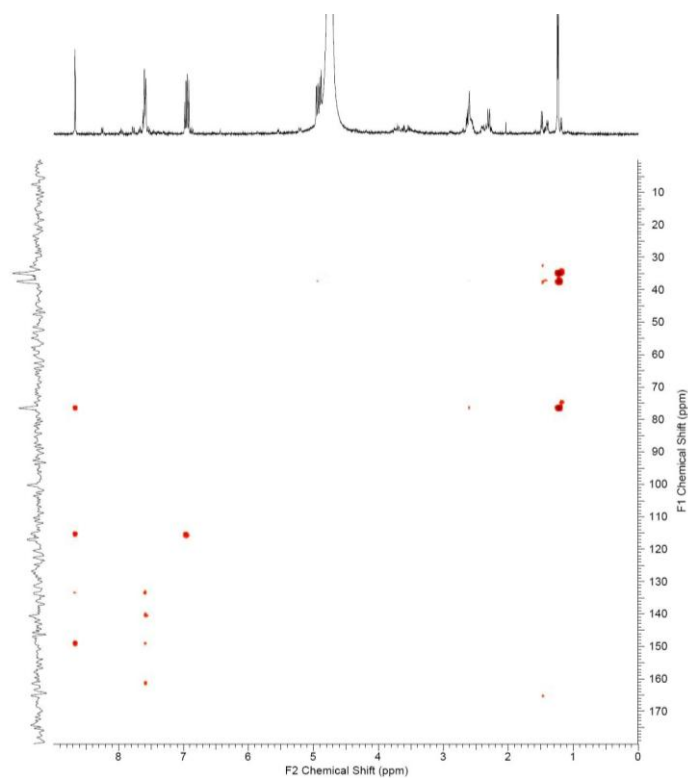

Figure A23: ^{1}H NMR spectrum of *o*-aminobenzaldehyde derivatized MeP5C (400 MHz) in D_2O: □ppm 8.67 (s, 1 H), 7.56 - 7.64 (m, 2 H), 6.95 (dd, *J*=18.0, 8.0 Hz, 2 H), 4.94 (d, *J*=9.4 Hz, 1 H), 4.89 (d, *J*=8.6 Hz, 1 H), 2.52 - 2.67 (m, 3 H), 2.30 (d, *J*=9.8 Hz, 1 H), 1.24 (d, *J*=5.9 Hz, 3 H)
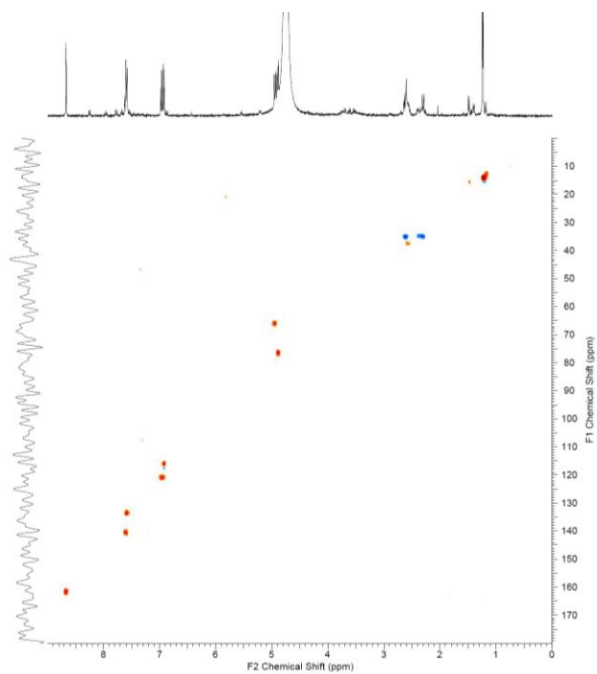
Figure A24: HSQC spectrum of *o*-aminobenzaldehyde derivatized MeP5C (400 MHz) in D$_2$O.
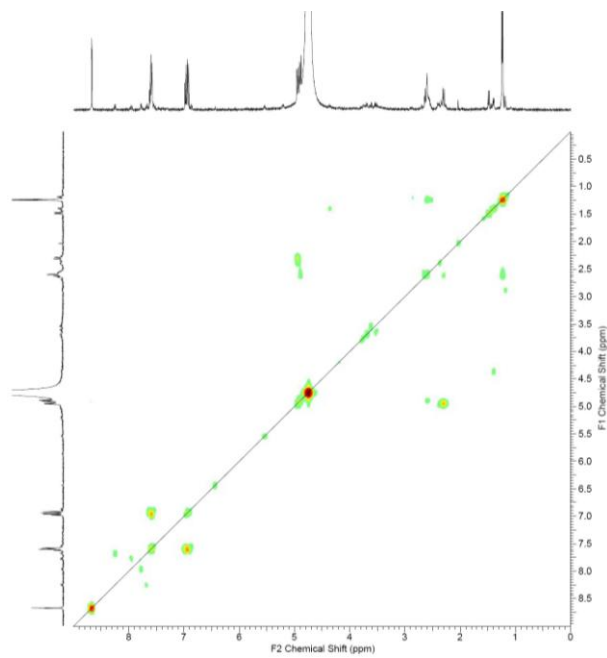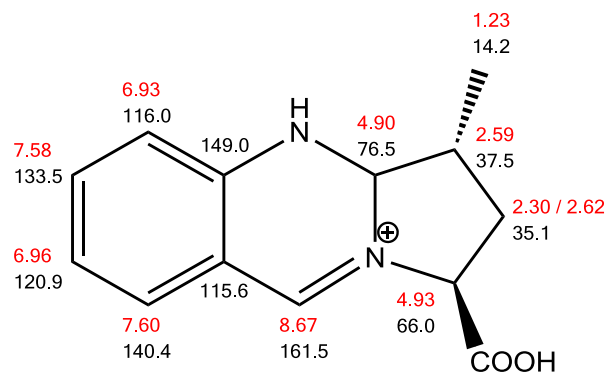

Figure A25: gCOSY spectrum of *o*-aminobenzaldehyde derivatized MeP5C (400 MHz) in D$_2$O.

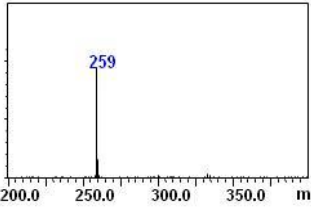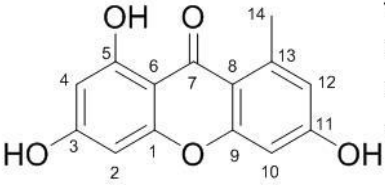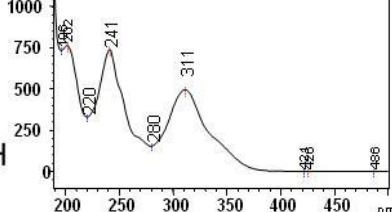**Figure A26:** HMBC spectrum of *o*-aminobenzaldehyde derivatized MeP5C (400 MHz) in D$_2$O.

Figure A27: ^{1}H and ^{13}C parameters of *o*-aminobenzaldehyde derivatized MeP5C extracted from ^{1}D and ^{2}D-NMR (Figure S10-S13).

6.3 Supplemental Information for Section 4

Table A7: Mass and UV-Vis spectra and NMR data of **5** in $(CD_3)_2CO$ on 500 MHz Bruker

NMR spectrometer

Norlichexanthone (**5**) (M.W. 258.2 g/mol)			
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	158.026		H2
C2	93.946	6.321, 1H, t, 2 Hz	H4
C3	164.889		H4
C4	98.686	6.189, , 1H, t, 2 Hz	H2, O5-H
C5	164.563		H4, H14, O5-H
C6	104.183		H4, H14, O5-H
C7	182.984		H2
C8	112.558		H12, H14
C9	160.214		H10
C10	101.967	6.721, 1H, s	H12, H14
C11	160.231		H10, H12
C12	116.889	6.708, 1H, s	H10, H14
C13	144.368		H10, H14
C14	23.426	2.776, 3H, s	H12
O5-H		13.452, 1H, s	

Table A8: Mass and UV-Vis spectra of **6**.

Griseophenone D (**6**)(M.W. 290.3)

290.8 [M+H]⁺

247.1 255.9 272.9 284.9 294.9 302.8 312.8 324.8 331.9 339.9 351.0

201

296

249

464 481 486 532 555 581
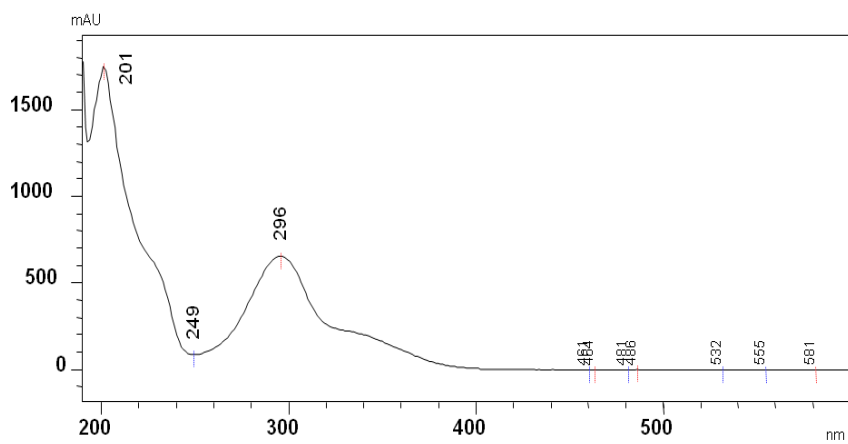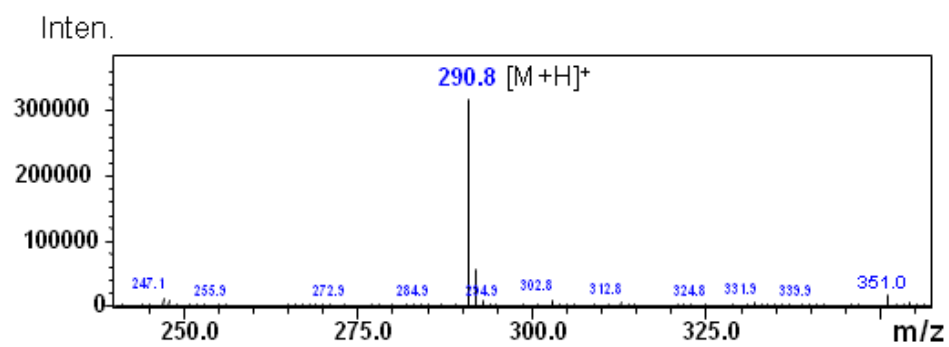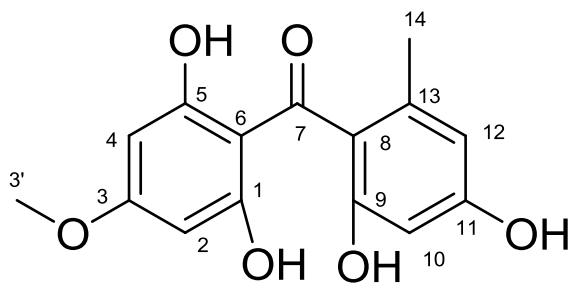
Table A9: Mass and UV-Vis spectra and NMR data of **7** in $(CD_3)_2CO$ on 500 MHz

Bruker NMR spectrometer

Griseophenone E (**7**)(M.W. 290.3 g/mol)			
Assignment	^{13}C ppm	^1H ppm, integration, mult, J_{HH} (Hz)	HMBC signals
C1, C5	166.54		H2, H4
C2, C4	95.78	5.86, 2H, s	H2, H4
C3	165.66		H2, H4
C6	106.88		H2, H4
C7	199.92		
C8	125.41		H14, H10, H12
C9	158.32		H9', H10
C9'	55.96	3.64, 3H,s	
C10	97.45	6.31,1H, d, 1.5 Hz	
C11	159.36		H10
C12	109.62	6.29, 1H, d, 1.5Hz	H14, H10, H12
C13	136.40		H14
C14	19.17	2.08, 3H,s	H12

Table A10: Mass and UV-Vis spectra and NMR data of **8** in $(CD_3)_2CO$ on 500 MHz

Bruker NMR spectrometer

<table>
<tr><td colspan="4" align="center">Griseophenone F (**8**)(M.W. 324.7 g/mol)</td></tr>
<tr><td colspan="4"></td></tr>
<tr><th>Assignment</th><th>^{13}C ppm</th><th>^{1}H ppm, integration, mult, J_{HH} (Hz)</th><th>HMBC signals</th></tr>
<tr><td>C1</td><td>161.38</td><td></td><td></td></tr>
<tr><td>C2</td><td>99.95</td><td></td><td>H4</td></tr>
<tr><td>C3</td><td>161.57</td><td></td><td>H4</td></tr>
<tr><td>C4</td><td>95.81</td><td>6.08 , 1H, s</td><td></td></tr>
<tr><td>C5</td><td>162.21</td><td></td><td></td></tr>
<tr><td>C6</td><td>107.27</td><td></td><td>H4</td></tr>
<tr><td>C7</td><td>200.51</td><td></td><td>H4</td></tr>
<tr><td>C8</td><td>124.97</td><td></td><td>H14, H12, H10</td></tr>
<tr><td>C9</td><td>158.46</td><td></td><td>H9', H10</td></tr>
<tr><td>C9'</td><td>55.98</td><td>3.64 , 1H, s</td><td></td></tr>
<tr><td>C10</td><td>97.63</td><td>6.32 , 1H, d, 1.8 Hz</td><td></td></tr>
<tr><td>C11</td><td>159.55</td><td></td><td>H12</td></tr>
<tr><td>C12</td><td>109.66</td><td>6.31, 1H, d, 1.8 Hz</td><td>H14, H10</td></tr>
<tr><td>C13</td><td>136.61</td><td></td><td>H14</td></tr>
<tr><td>C14</td><td>19.18</td><td>2.09, 3H, s</td><td>H12</td></tr>
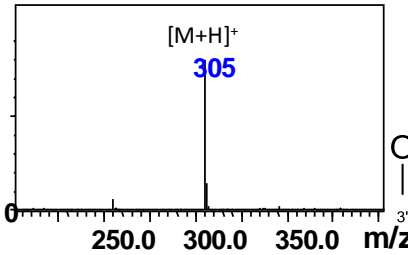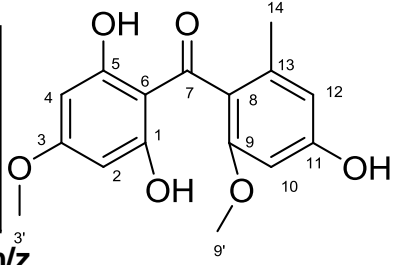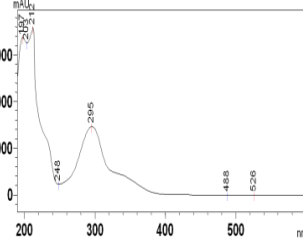<tr><td></td><td></td><td></td><td></td></tr>
</table>

153

Table A11: Mass and UV-Vis spectra and NMR data of **9** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer

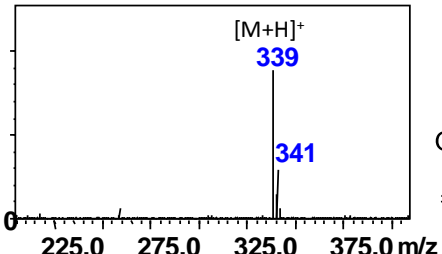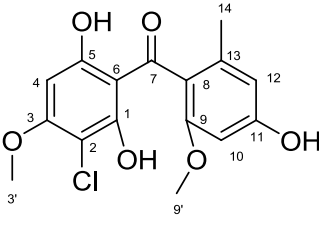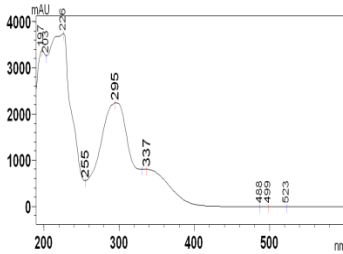|	Desmethyl-dehydrogriseofulvin B (**9**)(M.W. 336.7 g/mol)		
Assignment	^{13}C ppm	^{1}H ppm, integration, mult, J_{HH} (Hz)	HMBC signals
C1	170.3114		
C2	95.4402		H4
C3	165.0561		H4
C4	94.6581	6.795 ,1H, s	
C5	157.9998		H5', H4
C5'	55.6593	3.868, 3H, s	
C6	102.9		H4
C7	187.4717		H4
C8	88.4008		H14, H10, H12
C9	167.8811		H9', H10
C9'	56.0094	3.707,3H, s	
C10	103.1345	5.681 ,1H, d, 1.1 Hz	
C11	185.3387		H10
C12	128.6774	6.149,1H, t, 1.1 Hz	H10, H14
C13	147.2017		H14
C14	15.4892	1.773 ,3H, d, 1.1 Hz	H12

Table A12: Mass and UV-Vis spectra and NMR data of **10** in (CD$_3$)$_2$CO on 500 MHz

Bruker NMR spectrometer

<table>
<tr><td colspan="4" align="center">Griseophenone C (**10**) (M.W. 304.1 g/mol)</td></tr>
<tr><td colspan="4"></td></tr>
<tr><th>Assignment</th><th>^{13}C ppm</th><th>^{1}H ppm, integration, mult, J$_{HH}$ (Hz)</th><th>HMBC signals</th></tr>
<tr><td>C1, C5</td><td>165.3484</td><td></td><td>H(2,4)</td></tr>
<tr><td>C4, C2</td><td>94.18</td><td>5.93, 2H, s</td><td>H(2,4)</td></tr>
<tr><td>C3</td><td>167.5571</td><td></td><td>H3', H(2, 4)</td></tr>
<tr><td>C3'</td><td>55.94</td><td>3.81, 3H, s</td><td></td></tr>
<tr><td>C6</td><td>107.5</td><td></td><td>H(2,4)</td></tr>
<tr><td>C7</td><td>200.56</td><td></td><td></td></tr>
<tr><td>C8</td><td>125.31</td><td></td><td>H12, H14</td></tr>
<tr><td>C9</td><td>158.36</td><td></td><td>H9'</td></tr>
<tr><td>C9'</td><td>55.8272</td><td>3.64, 3H, s</td><td></td></tr>
<tr><td>C10</td><td>97.3642</td><td>6.33, 1H, d, 1.8 Hz</td><td>H12</td></tr>
<tr><td>C11</td><td>159.41</td><td></td><td>H10, H12</td></tr>
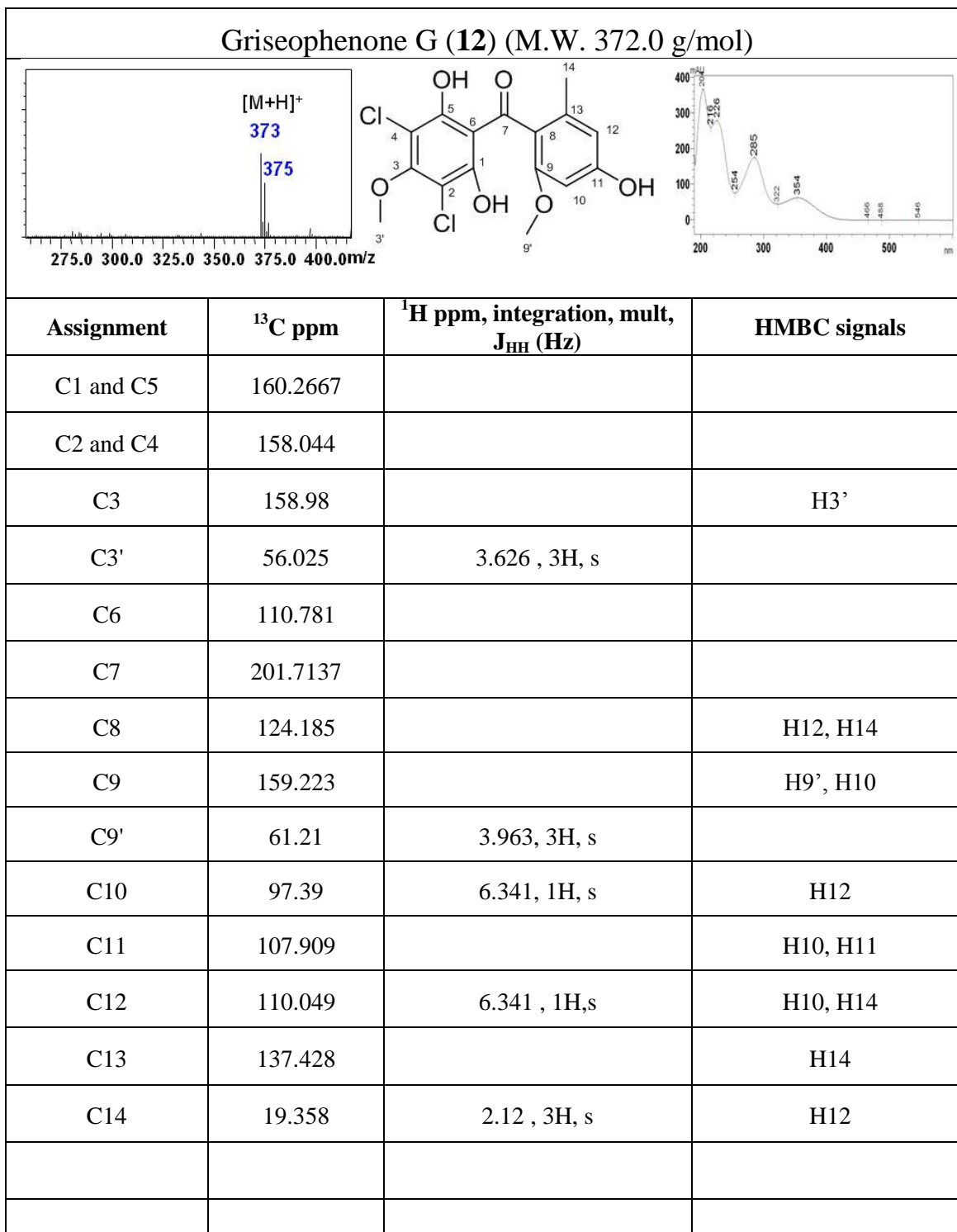<tr><td>C12</td><td>109.59</td><td>6.31, 1H, d, 1.8 Hz</td><td>H10,H14</td></tr>
<tr><td>C13</td><td>136.41</td><td></td><td>H14</td></tr>
<tr><td>C14</td><td>19.1588</td><td>2.08, 3H, s</td><td>H9'</td></tr>
</table>

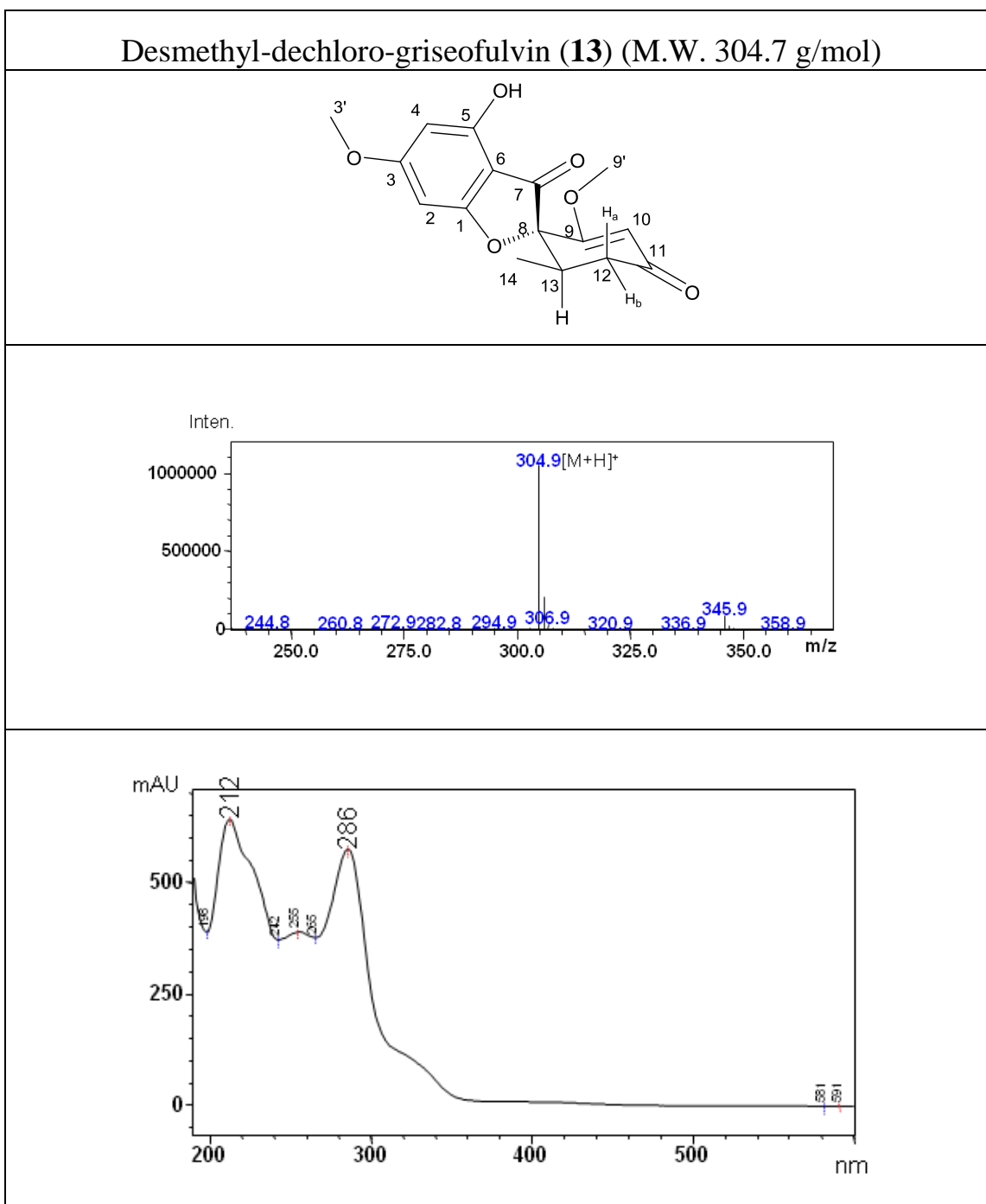**Table A13**: Mass and UV-Vis spectra and NMR data of **11** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer.

Griseophenone B (**11**) (M.W. 338.1 g/mol)			
Assignment	^{13}C ppm	^1H ppm, integration, mult, J_{HH} (Hz)	HMBC signals
C1	159.1		
C2	99.72		H4
C3	161.52		H3',H4
C3'	55.93	3.905, 3H, s	
C4	91.79	6.164, 1H, s	
C5	162.33		
C6	106.8		H4
C7	200.05		H4
C8	123.96		H10,H12, H14
C9	157.64		H9'
C9'	55.11	3.628, 3H, s	
C10	96.48	6.323, 1H, d, 1.8 Hz	H9', H12
C11	158.79		H10, H12
C12	108.84	6.315, 1H, d 1.8 Hz	H10, H14
C13	135.83		H14
C14	18.3164	2.08, 3H, s	H12

156

Table A14: Mass and UV-Vis spectra and NMR data of **12** in $(CD_3)_2CO$ on 500 MHz

Bruker NMR spectrometer.

Griseophenone G (**12**) (M.W. 372.0 g/mol)			
Assignment	**^{13}C ppm**	**^1H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1 and C5	160.2667		
C2 and C4	158.044		
C3	158.98		H3'
C3'	56.025	3.626 , 3H, s	
C6	110.781		
C7	201.7137		
C8	124.185		H12, H14
C9	159.223		H9', H10
C9'	61.21	3.963, 3H, s	
C10	97.39	6.341, 1H, s	H12
C11	107.909		H10, H11
C12	110.049	6.341 , 1H,s	H10, H14
C13	137.428		H14
C14	19.358	2.12 , 3H, s	H12

Desmethyl-dechloro-griseofulvin (**13**) (M.W. 304.7 g/mol)

304.9[M+H]+

345.9

244.8   260.8 272.9282.8 294.9 306.9   320.9   336.9   358.9
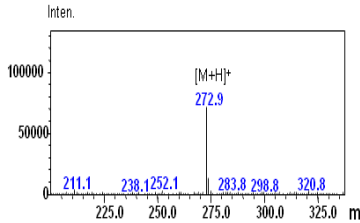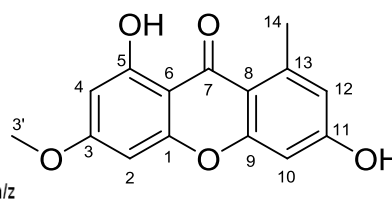

212

286

Table A16: Mass and UV-Vis spectra and NMR data of **14** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer.

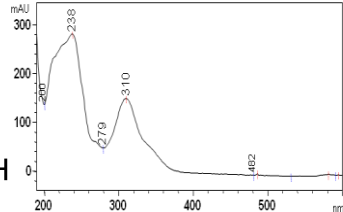| Desmethyl-dehydro-griseofulvin A (**14**) (M.W. 336.7 g/mol)			
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	169.42		
C2	104.52		H4
C3	165.69		H3', H4
C3'	57.69	4.01, 3H, s	
C4	95.59	6.42, 1H, s	
C5	158.46		H4
C6	96.15		H4
C7	190.24		H4
C8	89.31		H10, H12, H14
C9	168.62		H9', H10
C9'	57.09	3.72, 3H, s	
C10	104.24	5.7, 1H, d, 1.05 Hz	
C11	186.2		H10
C12	129.82	6.17, 1H, t, 1.3 Hz	H10, H14
C13	147.87		H14
C14	16.53	1.79, 3H, d, 1.2 Hz	H12

Table A17: Mass and UV-Vis spectra and NMR data of **15** in (CD$_3$)$_2$CO on 500 MHz Bruker NMR spectrometer.

Desmethyl-griseofulvin (**15**) (M.W. 338.7 g/mol)			
Assignment	^{13}C ppm	^{1}H ppm, integration, mult, J$_{HH}$ (Hz)	HMBC signals
C1	168.327		
C2	95.204		H4
C3	164.704		H3', H4
C3'	56.763	3.99, 3H, s	
C4	94.197	6.393, 1H, s	
C5	156.209		H4
C6	104.181		H4
C7	192.884		H4
C8	90.499		H10,H14
C9	170.4		H9', H10
C9'	56.404	3.69, 3H, s	
C10	104.443	5.554, 3H, s	
C11	194.957		H10, H12a, H12b,
C12	39.769	H12a-2.35, 1H, dd, 3.6 Hz, 14 Hz	H10, H13
		H12b,-2.80, 1H, dd(overlap)	
C13	36.235	2.80, 1H, m (overlap)	H12, H14
C14	13.573	0.92, 3H, d, 6.2 Hz	H13

Table A18: Mass and UV-Vis spectra and NMR data of **16** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer.

Griseoxanthone C (**16**) (M.W. 272.3 g/mol)			
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	157.87		H2
C2	92.48	6.42, 1H, d, 2 Hz	H4
C3	166.98		H4, H3'
C3'	56.27	3.92, 3H, s	
C4	97.50	6.260, , 1H, d, 2 Hz	H2, O5-H
C5	163.80		H4, H14, O5-H
C6	104.33		H4, H14, O5-H
C7	183.08		H2
C8	112.50		H12, H14
C9	160.30		H10
C10	101.50	6.786, 1H, s	H12, H14
C11	164.58		H10, H12
C12	117.05	6.736, 1H, s	H10, H14
C13	144.39		H10, H14
C14	23.42	2.770, 3H, d, 1.5 Hz	H12
O5-H		13.489, 1H, s	

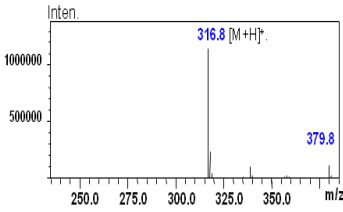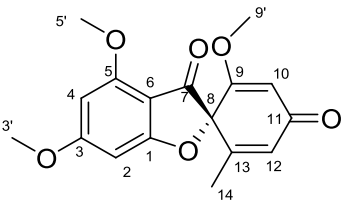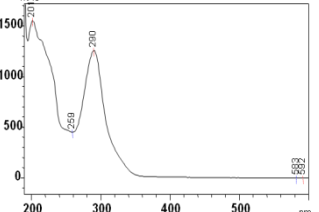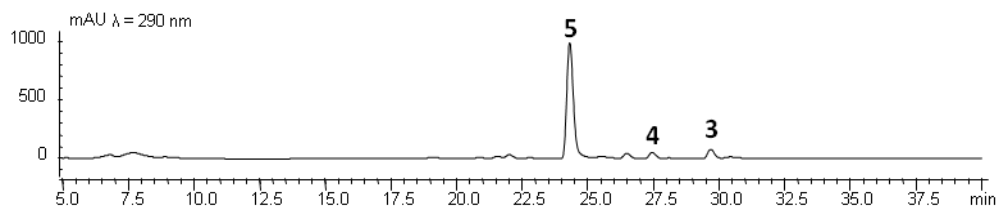**Table A19**: Mass and UV-Vis spectra and NMR data of **17** in $(CD_3)_2CO$ on 500 MHz

Bruker NMR spectrometer.

Griseophenone H (**17**) (M.W. 324.7 g/mol)			
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	157.09		
C2	100.6		H4
C3	162.07		H3', H4
C3'	56.8	3.91 ,3H, s	
C4	92.83	6.21,1H, s	
C5	162.63		H4
C6	108.49		H4
C7	200.73		
C8	122.55		H12, H14
C9	159.29		H10
C10	101.17	6.26, 1H, d, 1.8 Hz	
C11	160.15		H10, H12
C12	109.84	6.25, 1H, d, 1.8 Hz	H10,H14
C13	138.13		H14
C14	19.72	2.1,3H, s	H12

Table A20: Mass and UV-Vis spectra and NMR data of **18** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer.

Dehydro-griseofulvin (**18**) (M.W. 350.8 g/mol)			
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	170.303		
C2	97.4693		H4
C3	165.9245		H3', H4
C3'	57.9741	4.125,3H,s	
C4	91.8759	6.789 ,1H, s	
C5	159.4749		H5', H4
C5'	56.9964	4.000 ,3H,s	
C6	104.2393		H4
C7	186.9645		H4
C8	89.3667		H14, H10, H12
C9	168.5715		H9', H10
C9'	57.0895	3.714 ,3H,s	
C10	104.9241	5.700,1H, d, 1.3 Hz	
C11	186.2087		H10
C12	129.8836	6.169, 1H, t, 1.4 Hz	H10, H14
C13	147.7481		H14
C14	16.4739	1.772,3H, d, 1.37 Hz	H12

Table A21: Mass and UV-Vis spectra and NMR data of **19** in $(CD_3)_2CO$ on 500 MHz Bruker NMR spectrometer.
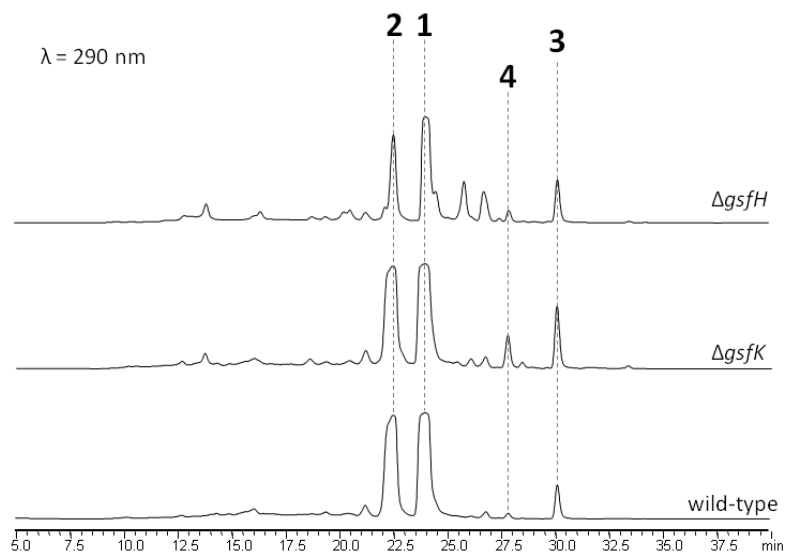

Dehydro—dechlorogriseofulvin (**19**)(M.W. 316.3 g/mol)		
Assignment	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**

| **Assignment**	**^{13}C ppm**	**^{1}H ppm, integration, mult, J_{HH} (Hz)**	**HMBC signals**
C1	177.1366		H2
C2	94.2726	6.266, 1H, d,1.7 Hz	H4
C3	169.0983		H2,H3', H4
C3'	56.9068	3.9773, 3H,s	
C4	90.1736	6.4462 ,1H, d, 1.7 Hz	H2
C5	160.5407		H5', H4
C5'	56.5022	3.909, 3H,s	
C6	104.0223		H4
C7	189.0155		H4
C8	88.9061		H14, H10, H12
C9	171.6062		H9', H10
C9'	56.8816	3.699, 3H,s	
C10	104.0091	5.664 ,1H, d, 1.3 Hz	H12
C11	186.3979		H10
C12	129.8836	6.130, 1H, t, 1.4 Hz	H10, H14
C13	148.4656		H14
C14	16.4971	1.741, 3H, d, 1.5 Hz	H12

Figure A28: Feeding of norlichexanthone (**5**) to Δ*gsfA* mutant. Chromatogram from the LCMS analysis from the extracts of Δ*gsfA* mutant fed with 0.1 mg/mL of **5**, showing that feeding of **5** did not restore the production of griseofulvin(**1**) or dechlorogriseofulvin (**2**). Also shown in the trace are viridicatumtoxin (**3**) and tryptoquialanine (**4**).
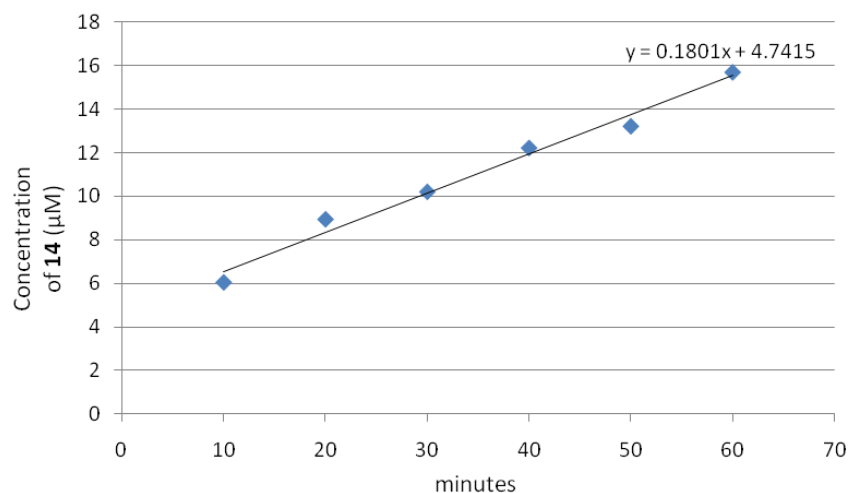

Figure A29: Time course-*in vitro* assay of GsfF from 0-60 minutes using 50 μM griseophenone B (**10**) and 1mM NADPH as substrates and 0.412 mg/mL microsomal protein in 100 mM Tris-HCl (pH 7.5) showing a turnover rate of 0.437 μM/ min·mg of micromal protein. The concentration of protein from the microsomes was determined by comparison with bovine serum albumin (BSA) standard curve. The amount of product desmethyl-dehydrogriseofulvin (**14**) was quantified by comparison with a standard curve using purified **14** from Δ*gsfD* fermentation.

Figure A30: Metabolic profile of Δ*gsfH* and Δ*gsfK* mutants showing the production of griseofulvin (**1**) and dechlorogriseofuvlin (**2**).
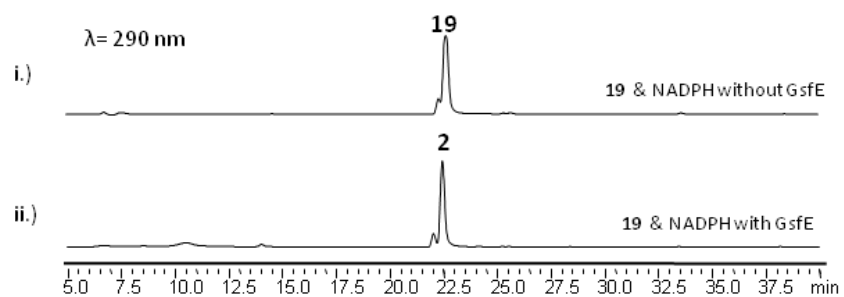
Figure A31: *In vitro* assay of GsfE using dechloro-dehydrogriseofulvin (**19**) as substrate.) i.) Chromatogram (λ = 290 nm) from extract of **19** incubated overnight with GsfE only in 100 mM Tris-HCl (pH 7.5). ii.) .) Chromatogram (λ = 290 nm) from extract of **19** incubated overnight with GsfE and NADPH showing conversion to **2**.
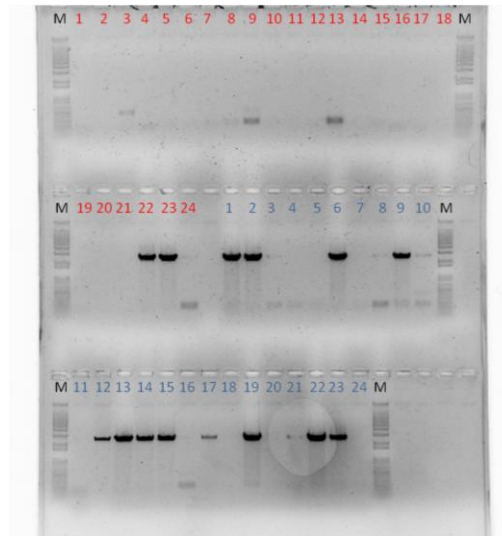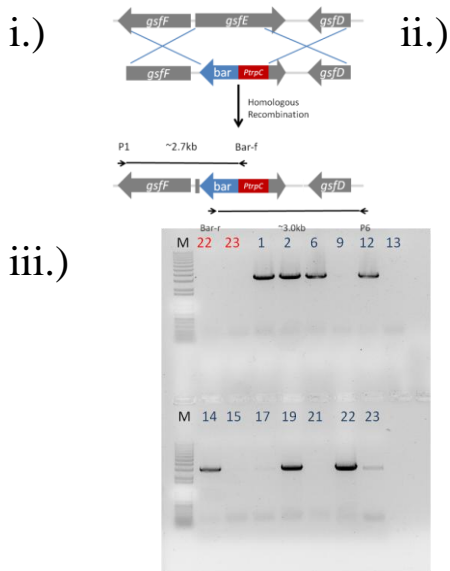
Figure A32: Δ*gsfE* mutant screening shown as example of PCR screening method. i.) Linear knockout cassette showing the positions of the primer binding site after successful disruption of the targeted gene. ii.) Amplification of genomic DNA from 48 Δ*gsfE* transformants (plate I, 1-24 in red and plate II, 1-24 in blue) using primer pair Δ*gsfE*-KO-P1 and bar-f showing expected ~2.7 kb amplicon. iii.) Amplification of genomic DNA from Δ*gsfE* transformants that passed the screening from (ii.) using primer pairs Δ*gsfE*-KO-P6 and bar-r.
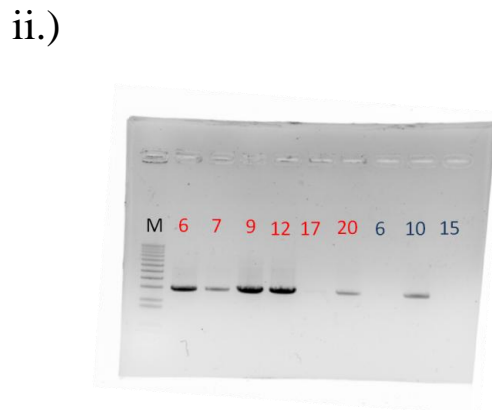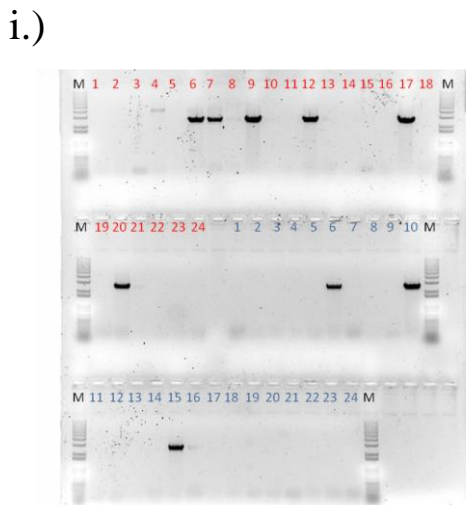
Figure A33: Δ*gsfH* mutant screening. i.) Amplification of genomic DNA from 48 Δ*gsfH* transformants (plate I, 1-24 in red and plate II, 1-24 in blue) using primer pair Δ*gsfH*-KO-P1 and bar-f showing expected ~2.7 kb amplicon. ii.) Amplification of genomic DNA from Δ*gsfH* transformants that passed the screening from (i.) using primer pairs Δ*gsfH*-KO-P6 and bar-r.
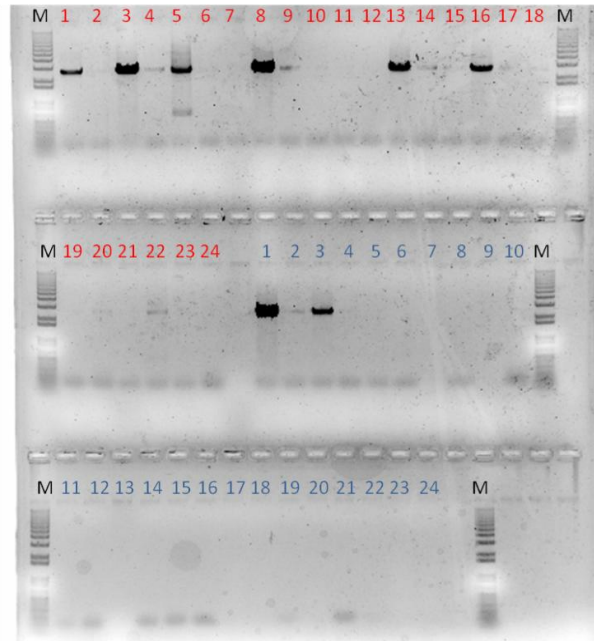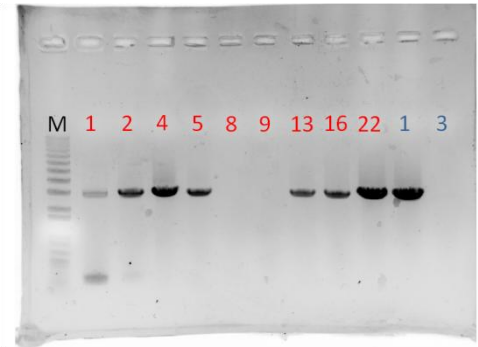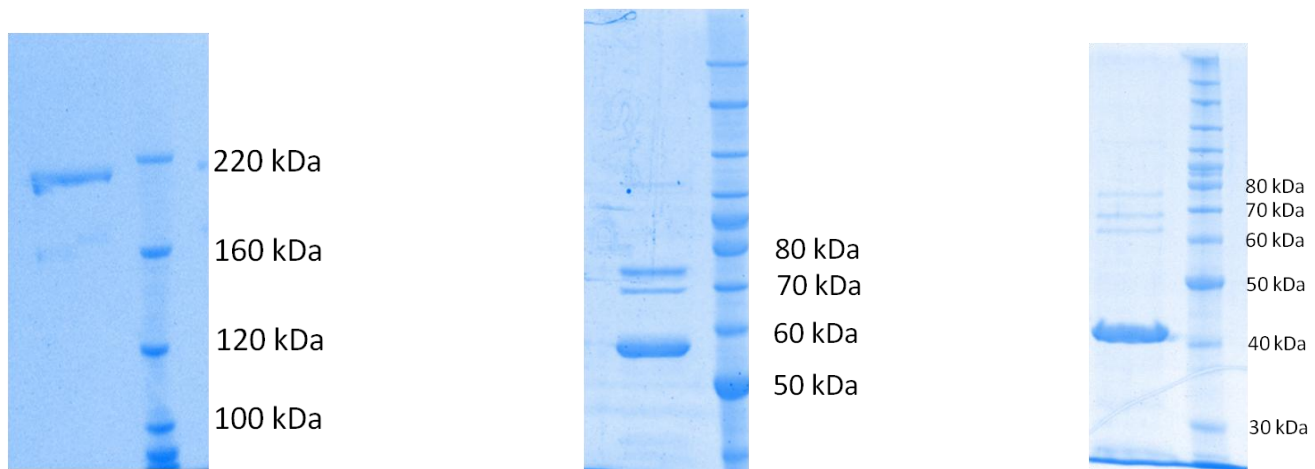
Figure A34: Δ*gsfK* mutant screening. i.) Amplification of genomic DNA from 48 Δ*gsfK* transformants (plate I, 1-24 in red and plate II, 1-24 in blue) using primer pair Δ*gsfK*-KO-P1 and bar-f showing expected ~2.7 kb amplicon. ii.) Amplification of genomic DNA from Δ*gsfK* transformants that passed the screening from (i.) using primer pairs Δ*gsfK*-KO-P6 and bar-r.

Figure A35: SDS-PAGE gels of the PKS GsfA (left, expected size 185 kDa), halogenase GsfI (middle, expected size 58 kDa) and the dehydrogriseofulvin

reductase GsfE (right, expected size 41 kDa)
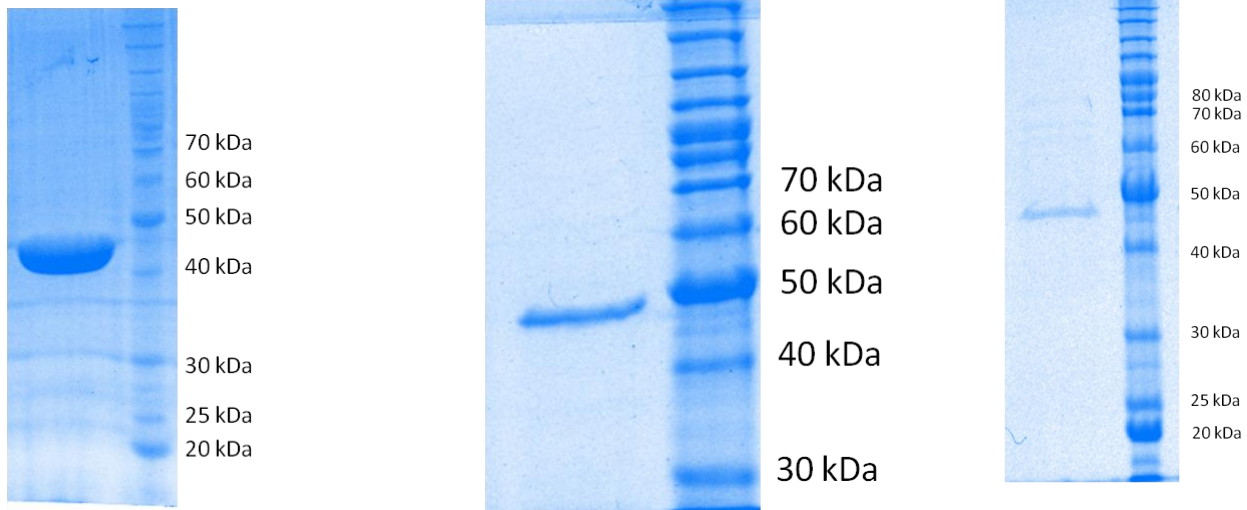
Figure A36: SDS-PAGE gel of GsfB (left, expected size 47 kDa) , GsfC (middle, 49 kDa) and GsfD (right, 47 kDa)
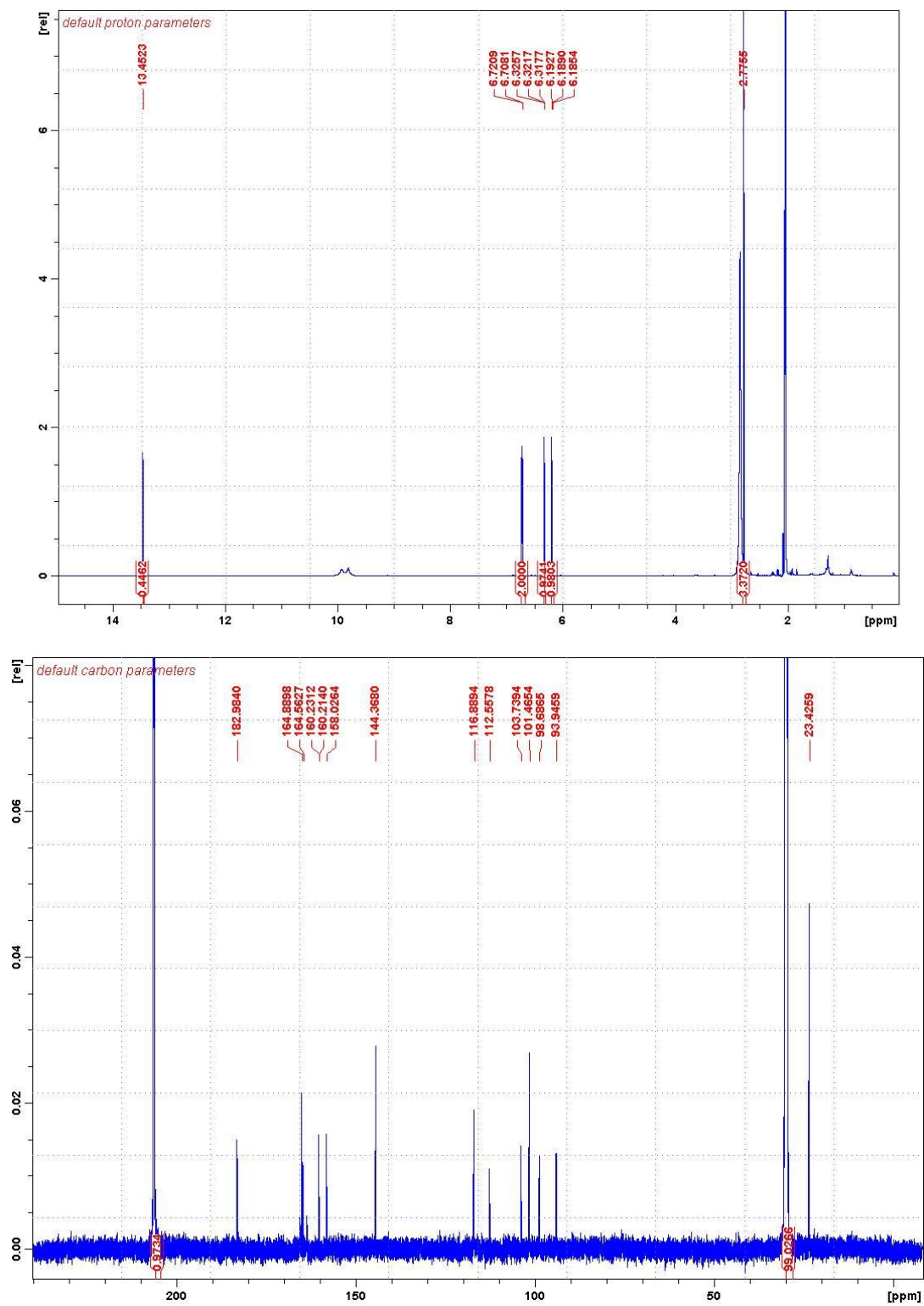
Figure A37: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **5** in $(CD_3)_2CO$ (500 MHz).

Figure A38: 2D HSQC (top) and HMBC (bottom) of **5** in $(CD_3)_2CO$ (500 MHz).

Figure A39: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **7** in (CD$_3$)$_2$CO (500 MHz).
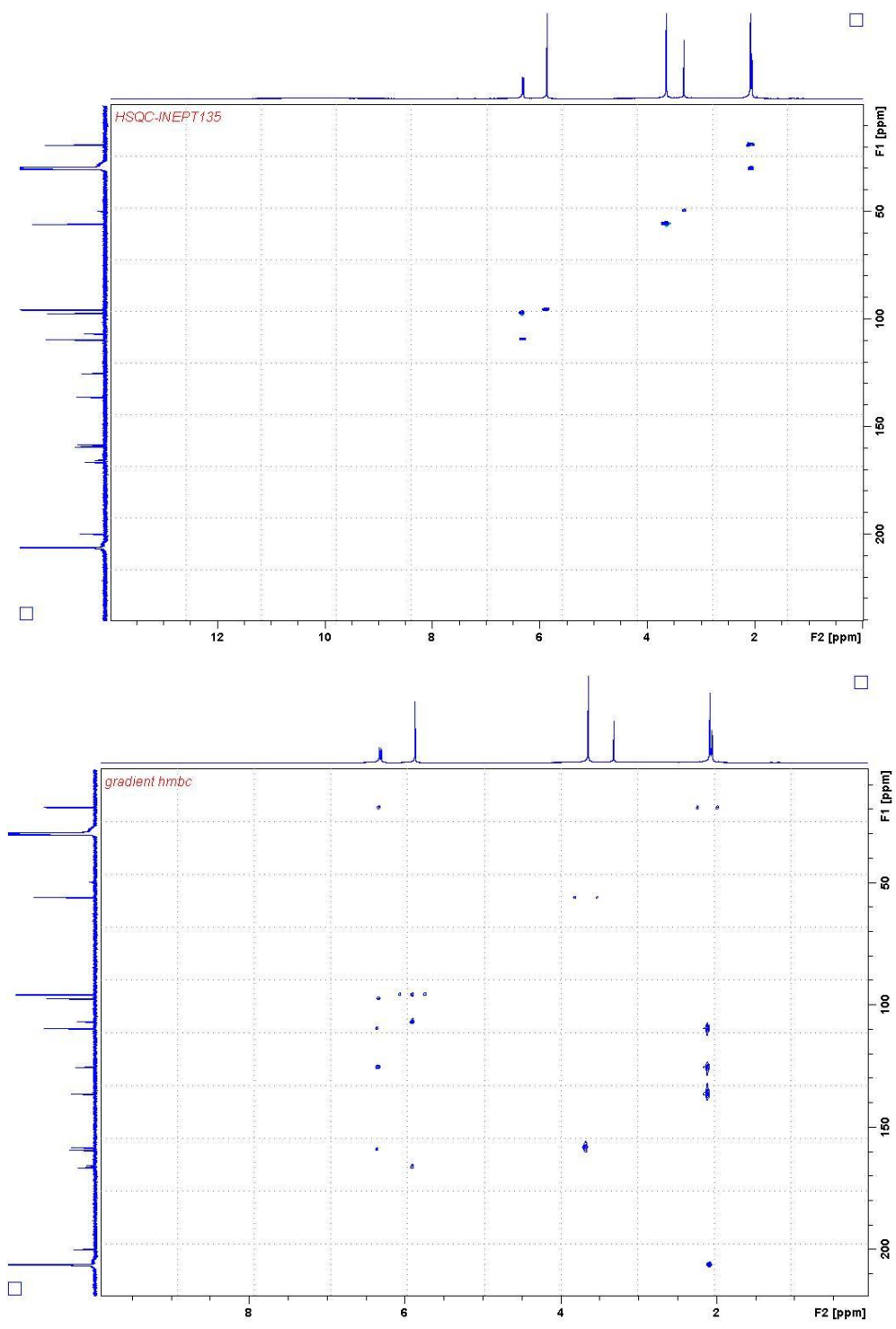
Figure A40: 2D HSQC (top) and HMBC (bottom) of **7** in $(CD_3)_2CO$ (500 MHz).

Figure A41: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **8** in $(CD_3)_2CO$ (500 MHz).

Figure A42: 2D HSQC (top) and HMBC (bottom) of **8** in (CD$_3$)$_2$CO (500 MHz).

Figure A43: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **9** in $(CD_3)_2CO$ (500 MHz).
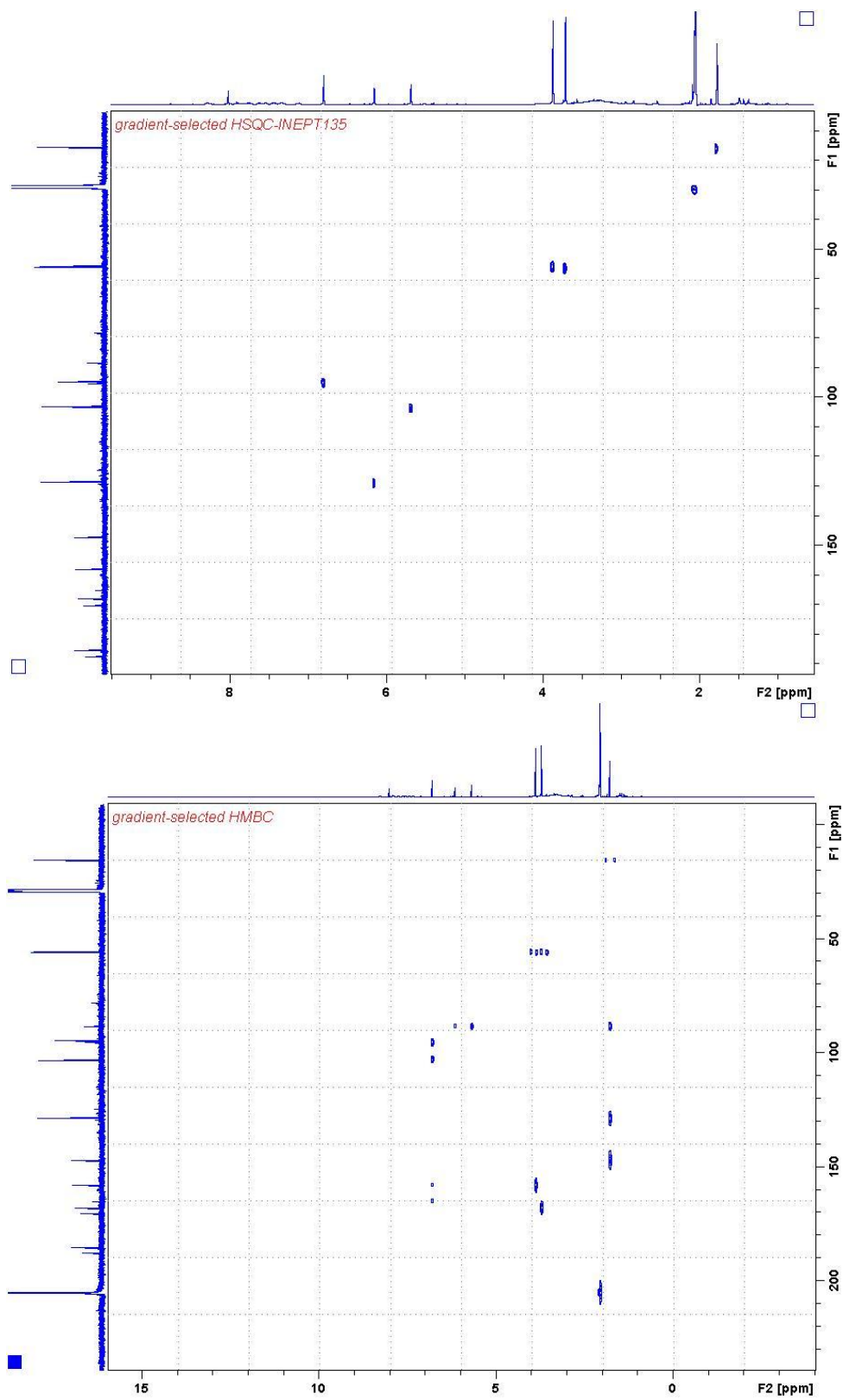
Figure A44: 2D HSQC (top) and HMBC (bottom) of **9** in $(CD_3)_2CO$ (500 MHz).
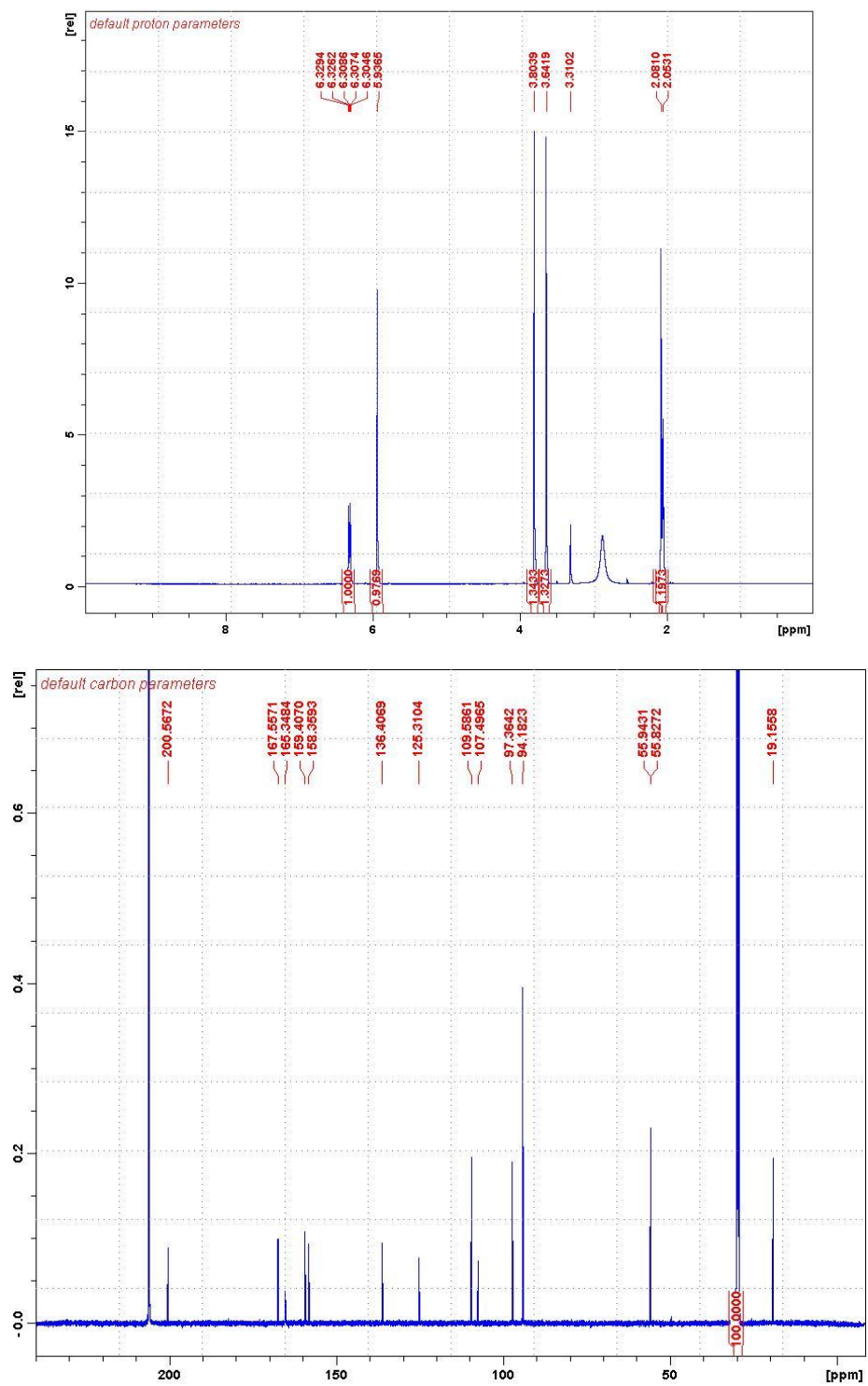
Figure A45: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **10** in $(CD_3)_2CO$ (500 MHz).
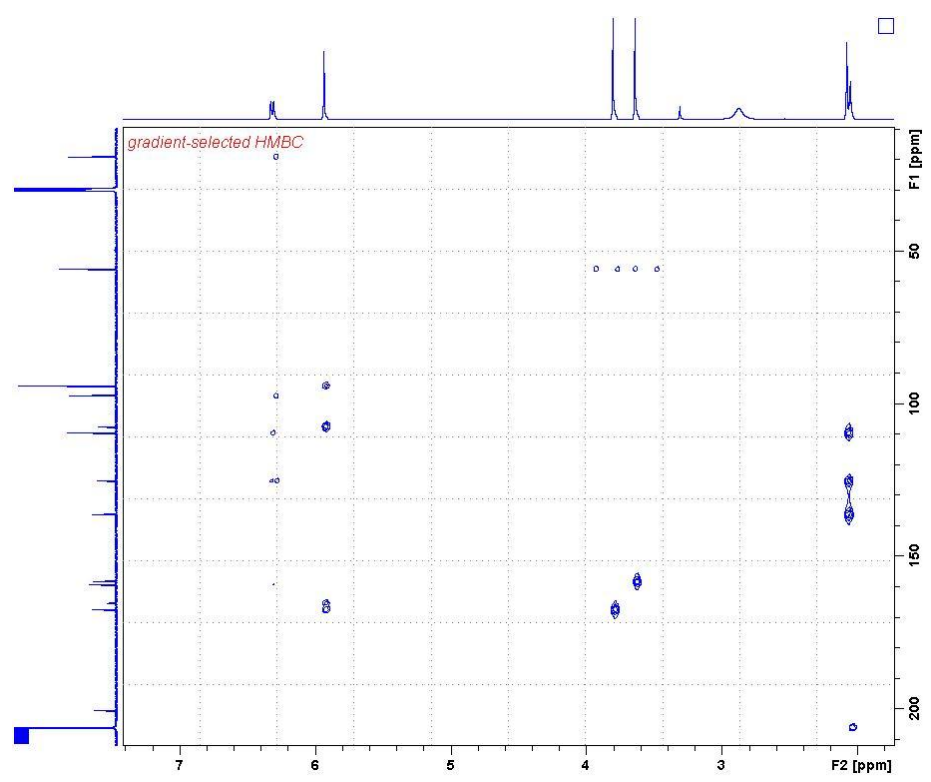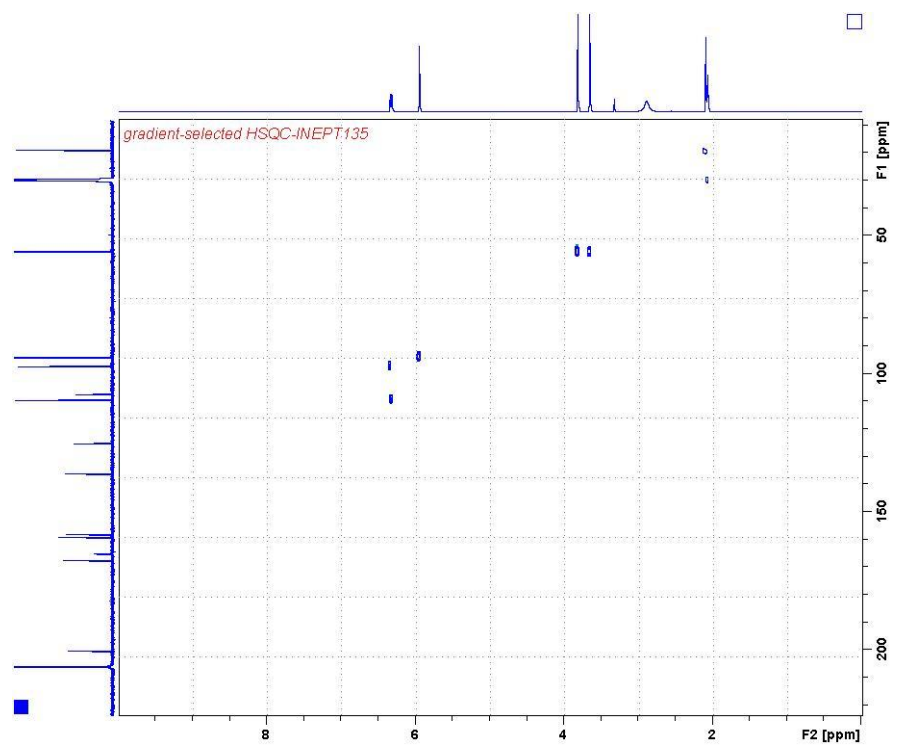
Figure A46: 2D HSQC (top) and HMBC (bottom) of **10** in $(CD_3)_2CO$ (500 MHz).
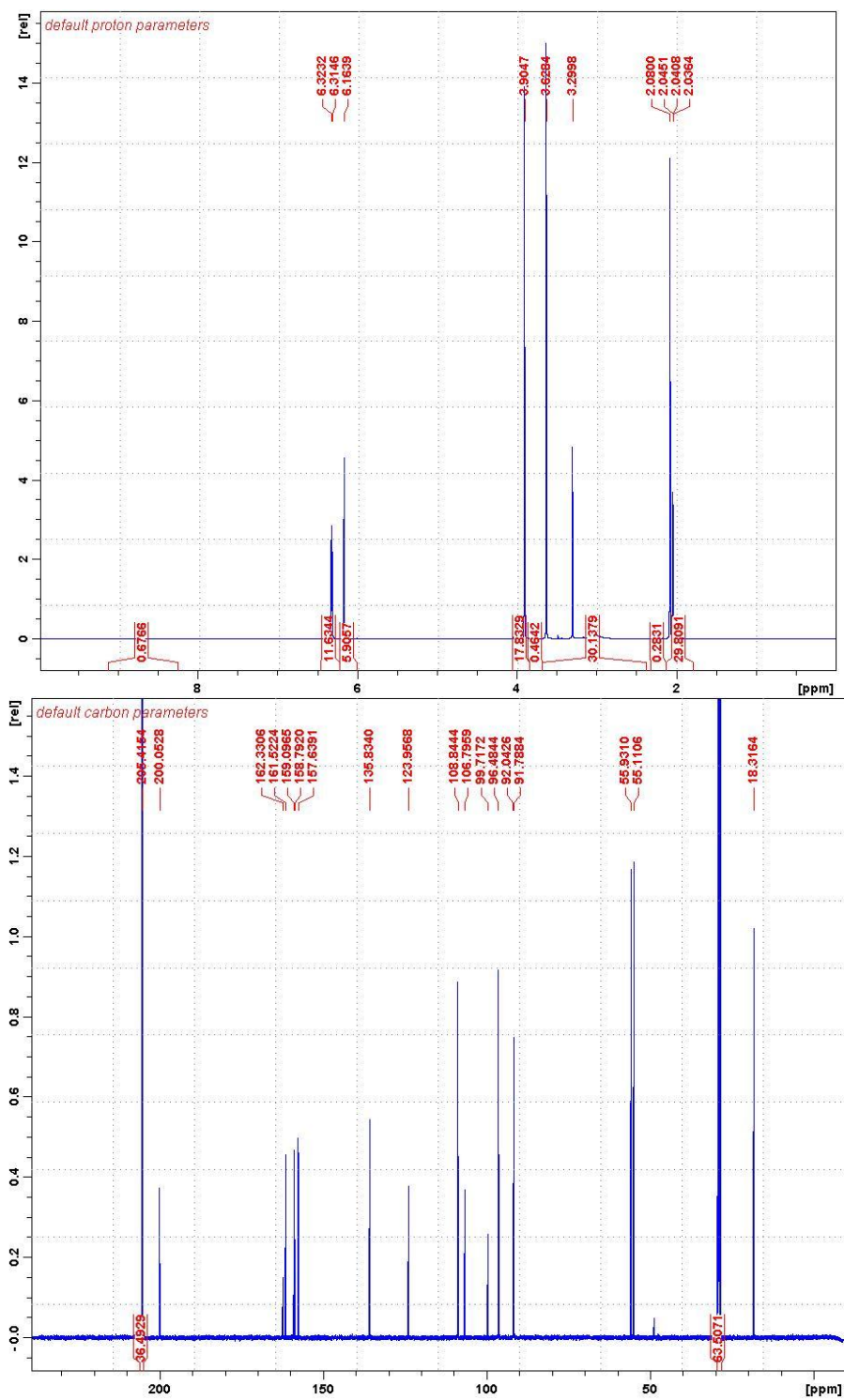
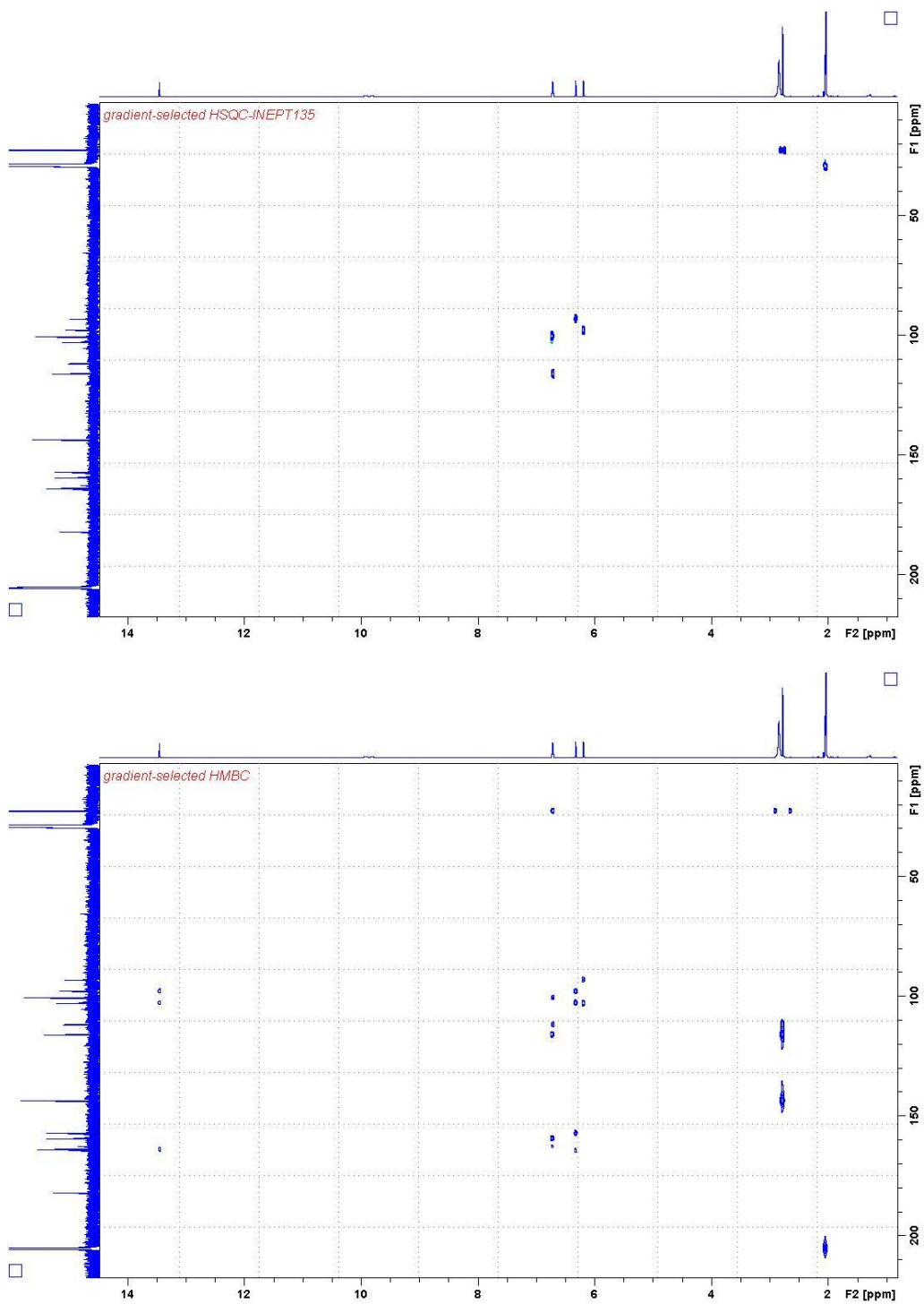**Figure A47:** 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **11** in $(CD_3)_2CO$ (500 MHz).
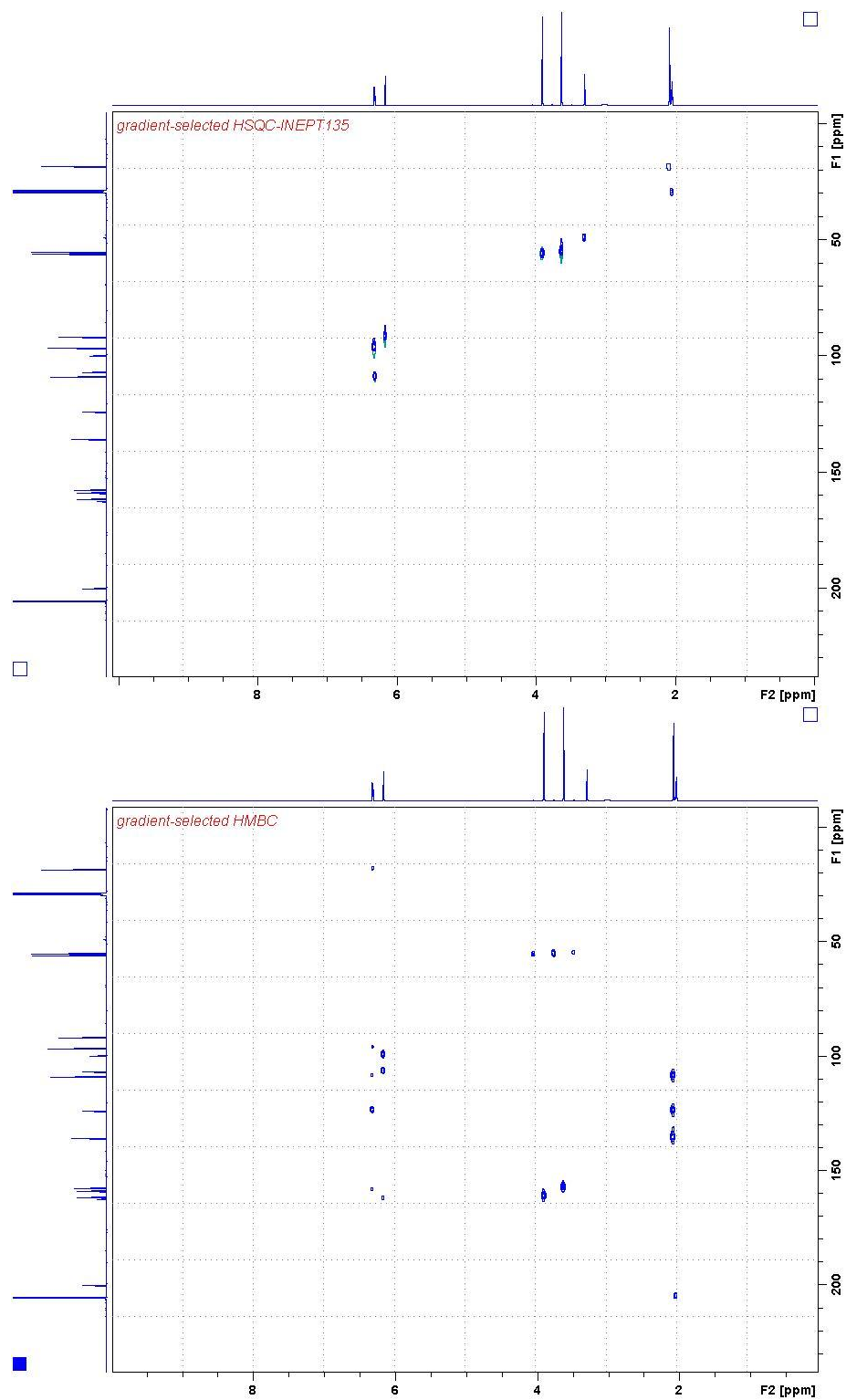
Figure A48: 2D HSQC (top) and HMBC (bottom) of **11** in $(CD_3)_2CO$ (500 MHz).
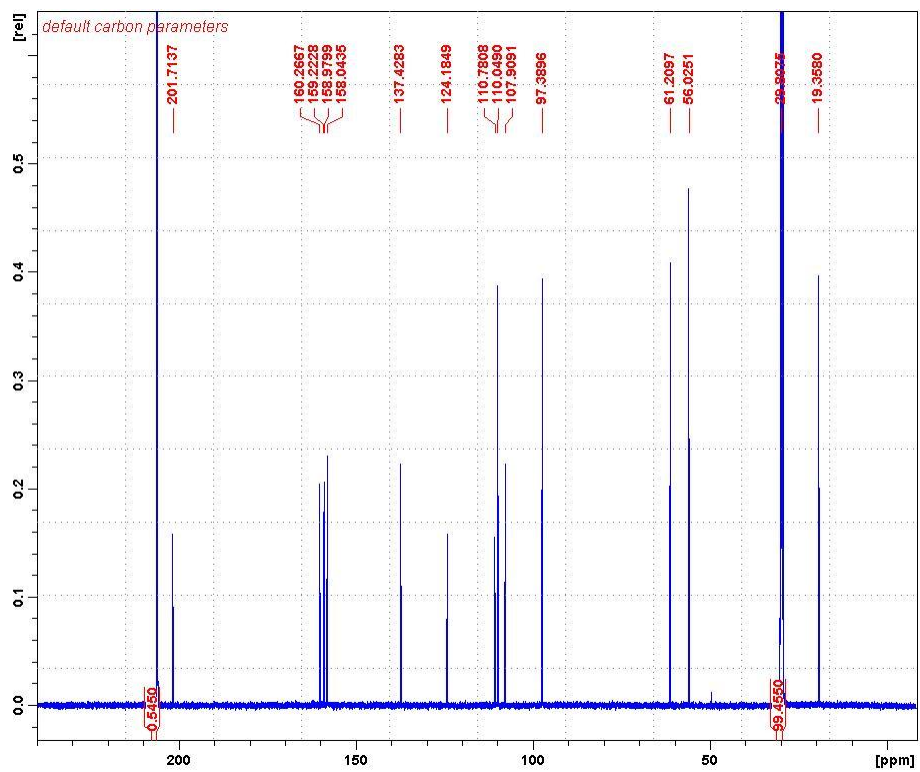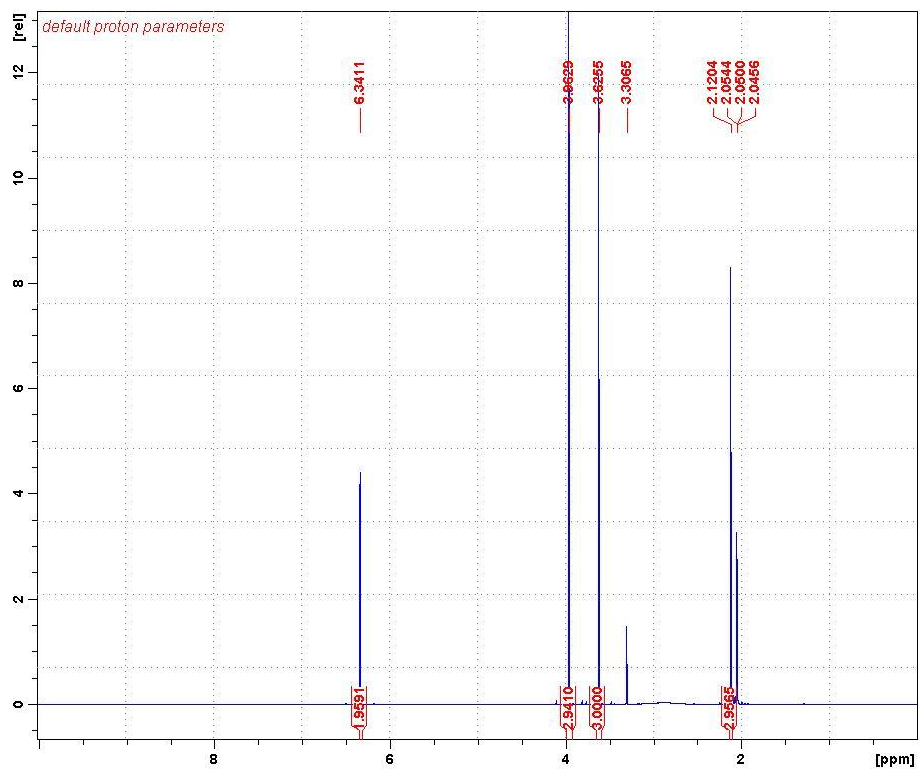
Figure A49: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **12** in $(CD_3)_2CO$ (500 MHz).
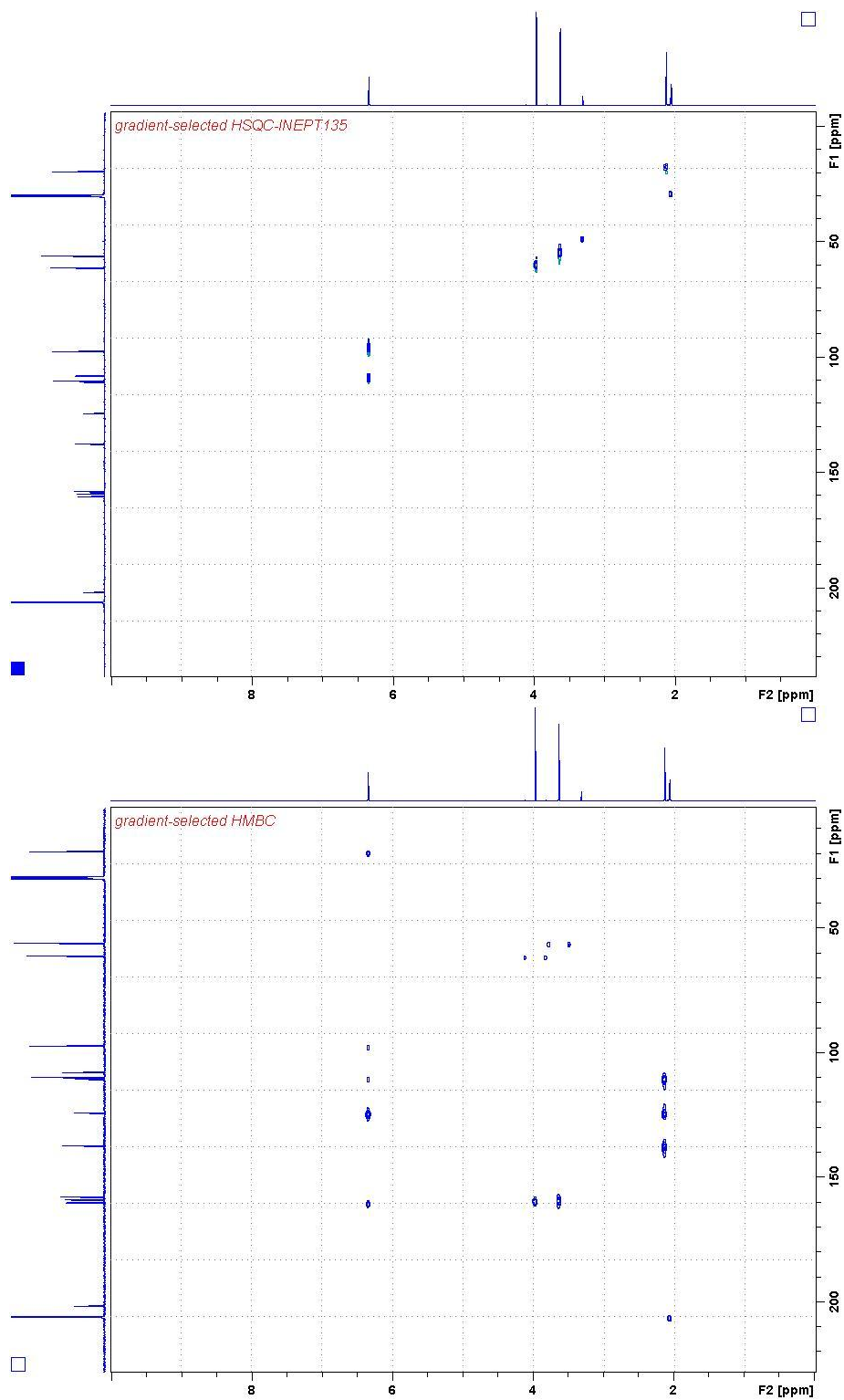
Figure A50: 2D HSQC (top) and HMBC (bottom) of **12** in (CD$_3$)$_2$CO (500 MHz).
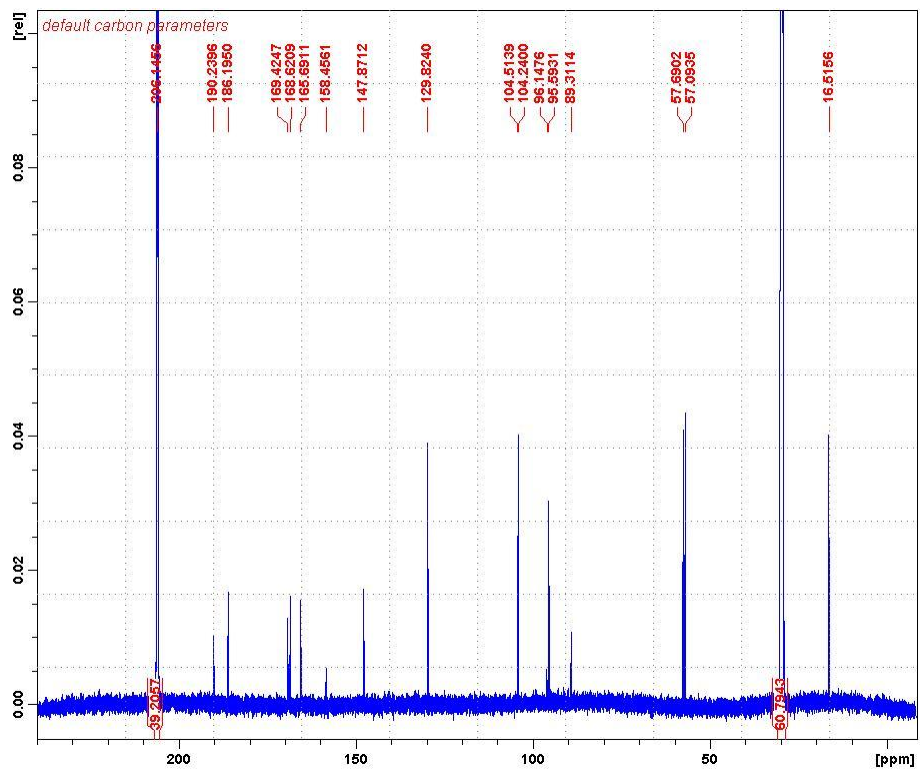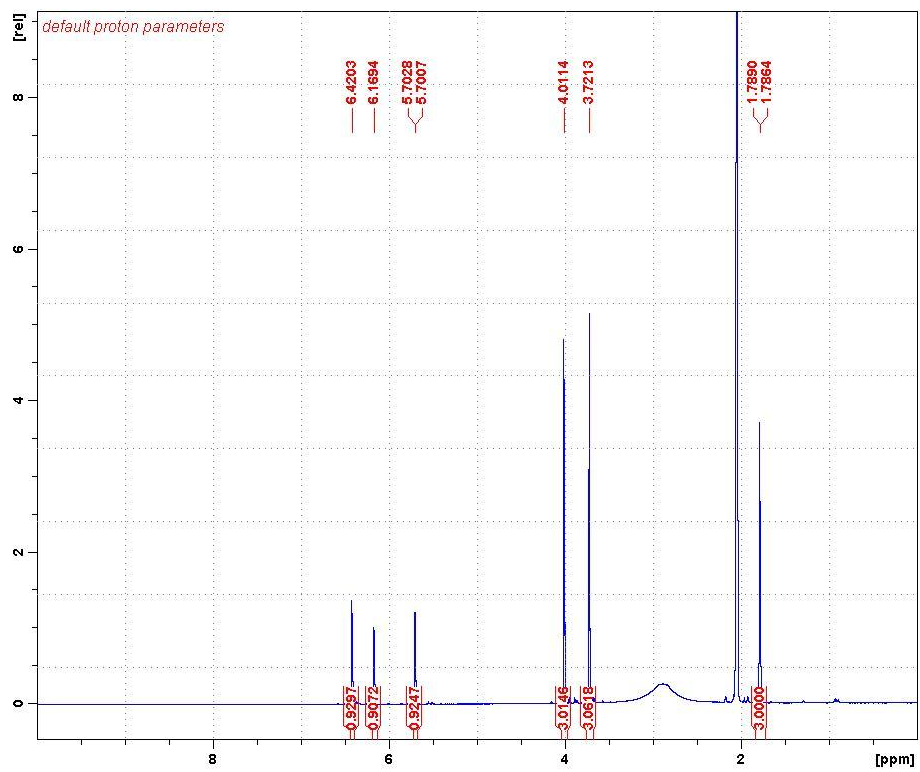
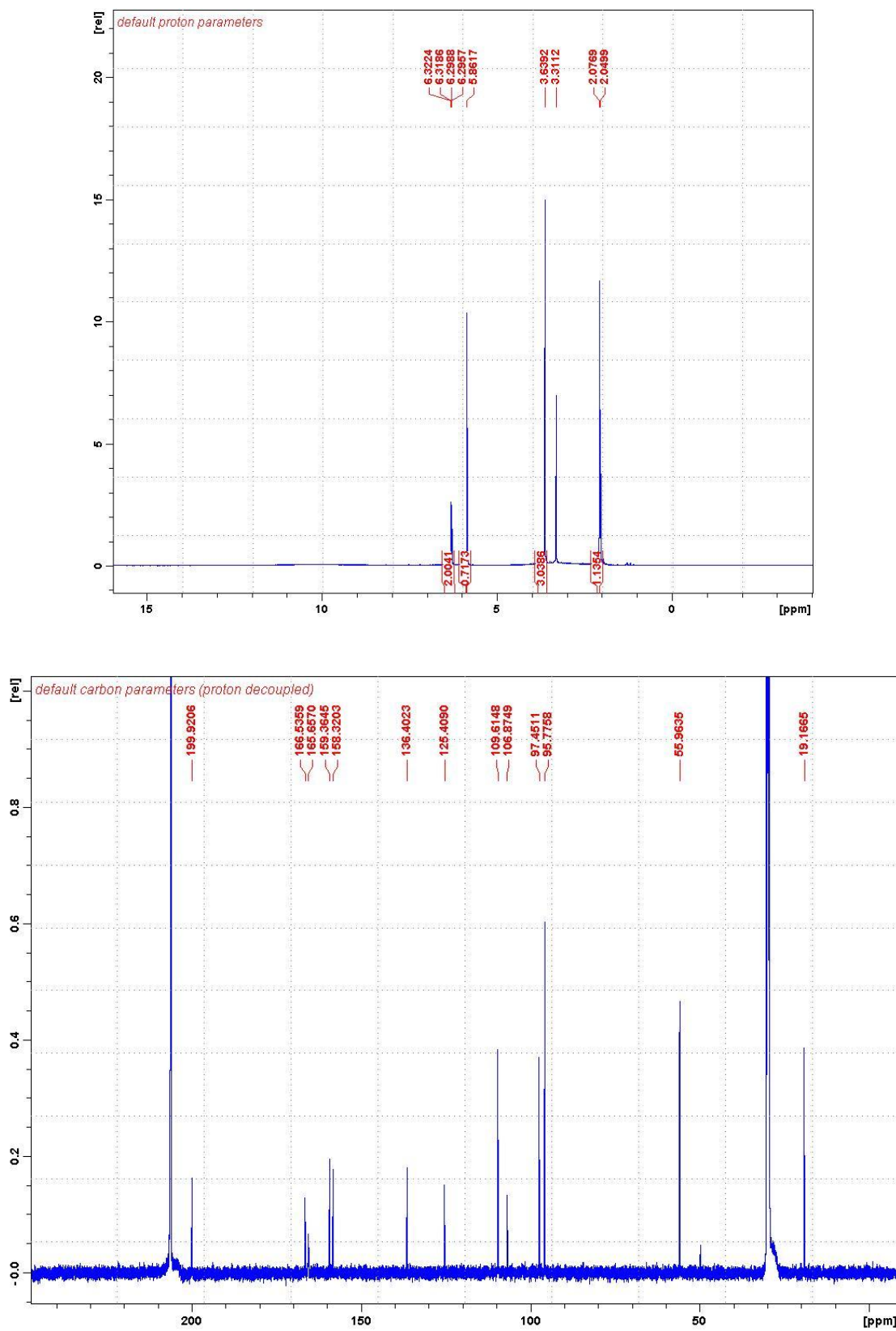**Figure A51**: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **14** in $(CD_3)_2CO$ (500 MHz).
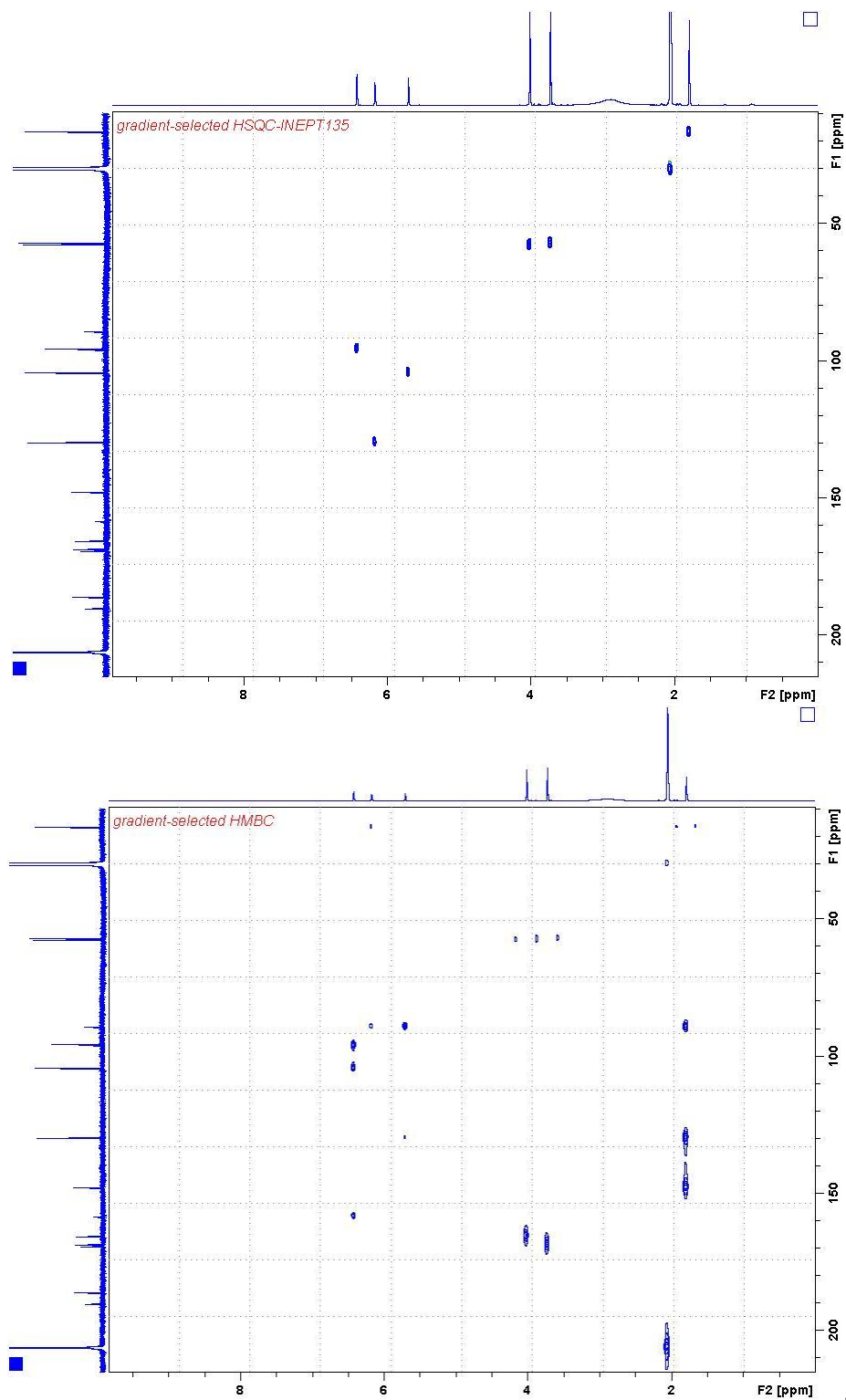
Figure A52: 2D HSQC (top) and HMBC (bottom) of **14** in $(CD_3)_2CO$ (500 MHz).
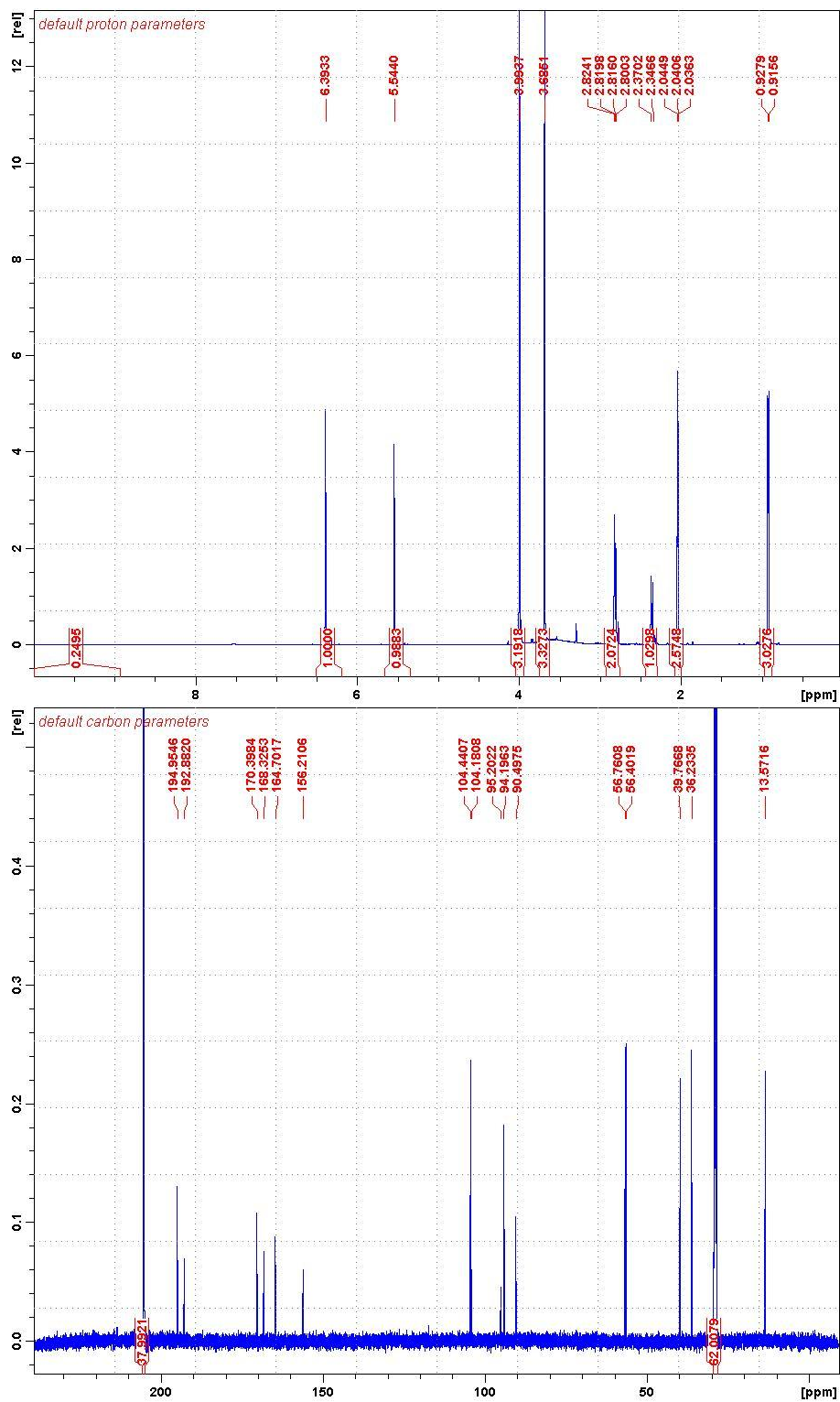
Figure A53: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **15** in $(CD_3)_2CO$ (500 MHz).
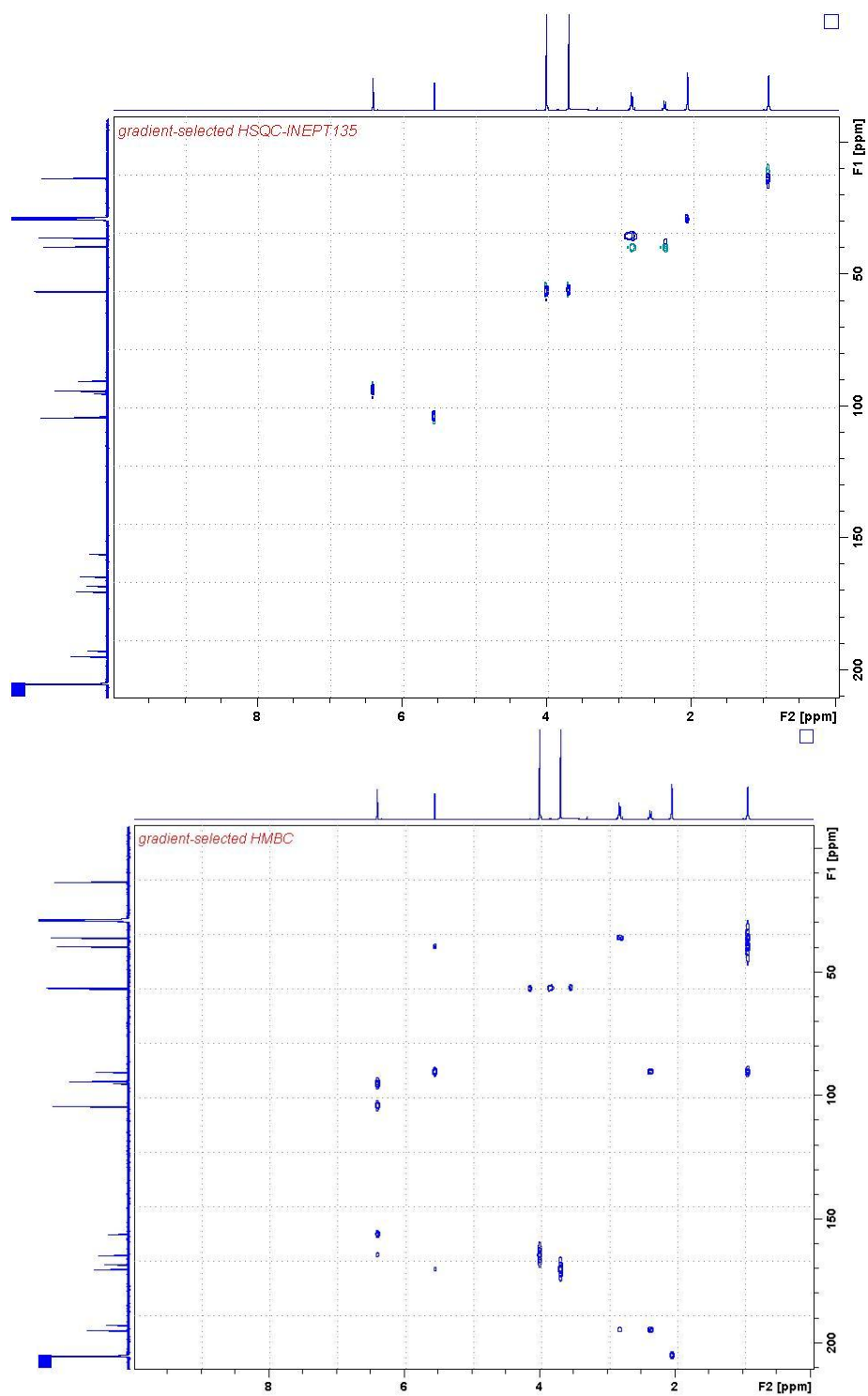
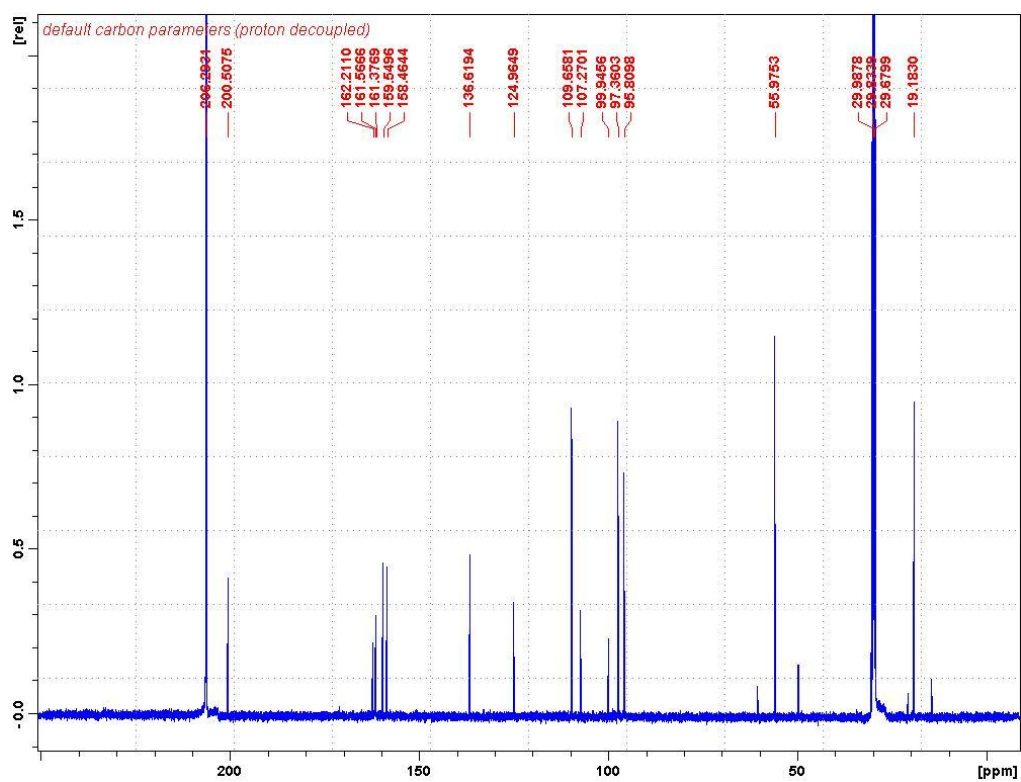**Figure A54**: 2D HSQC (top) and HMBC (bottom) of **15** in (CD$_3$)$_2$CO (500 MHz).
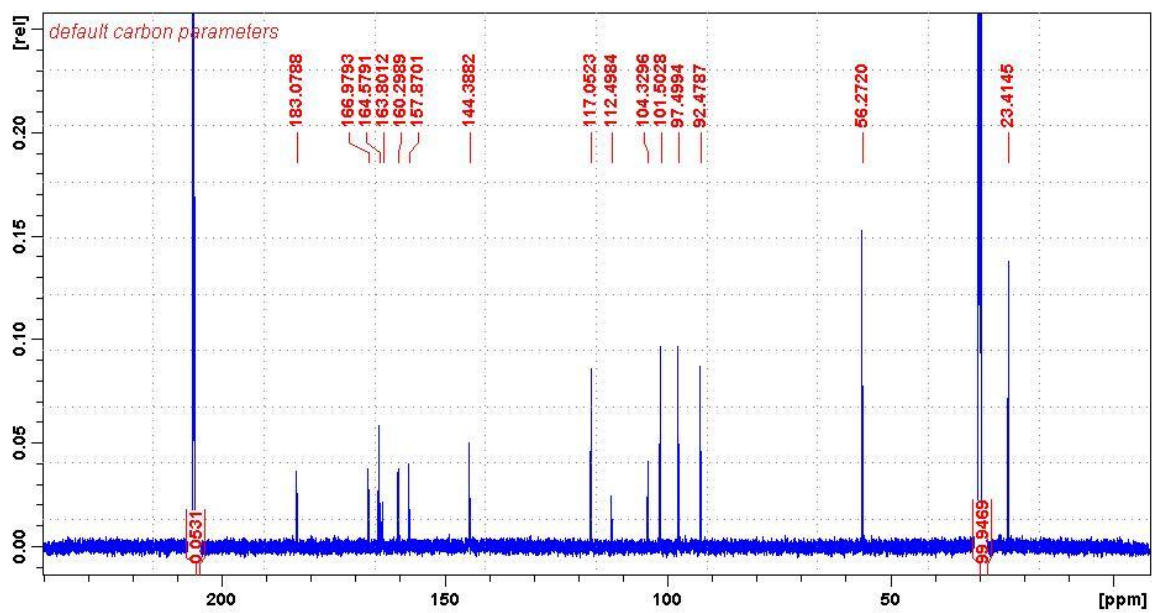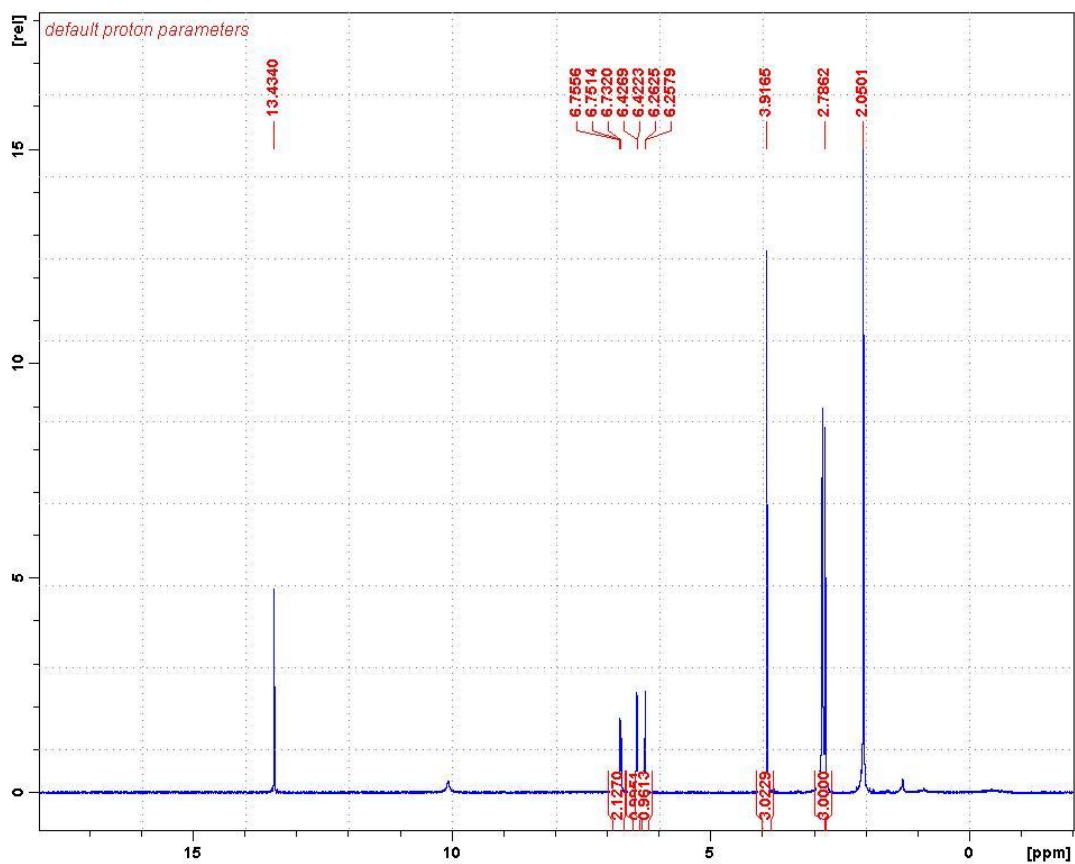
Figure A55: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **16** in $(CD_3)_2CO$ (500 MHz).
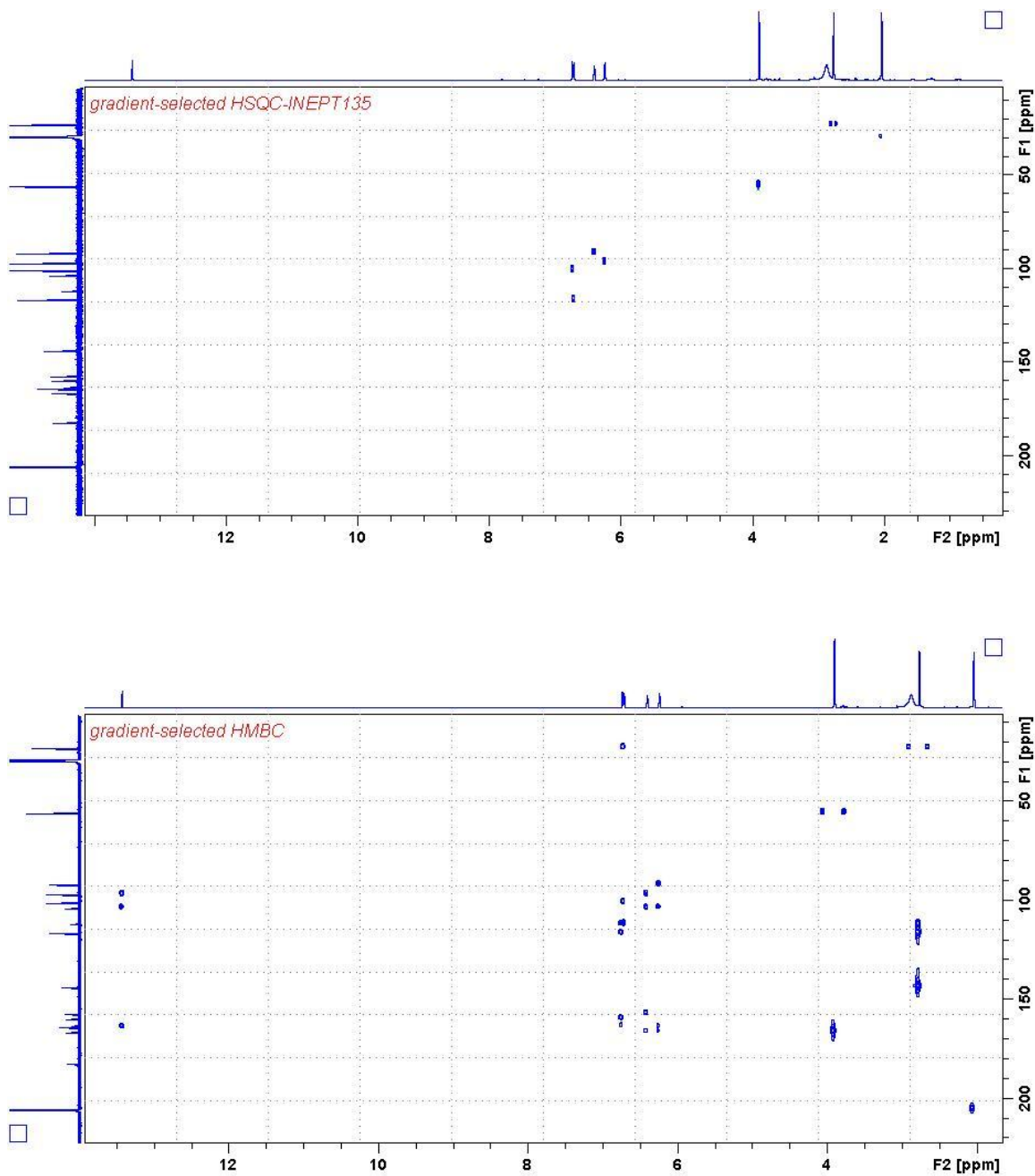
Figure A56: 2D HSQC (top) and HMBC (bottom) of **16** in $(CD_3)_2CO$ (500 MHz).
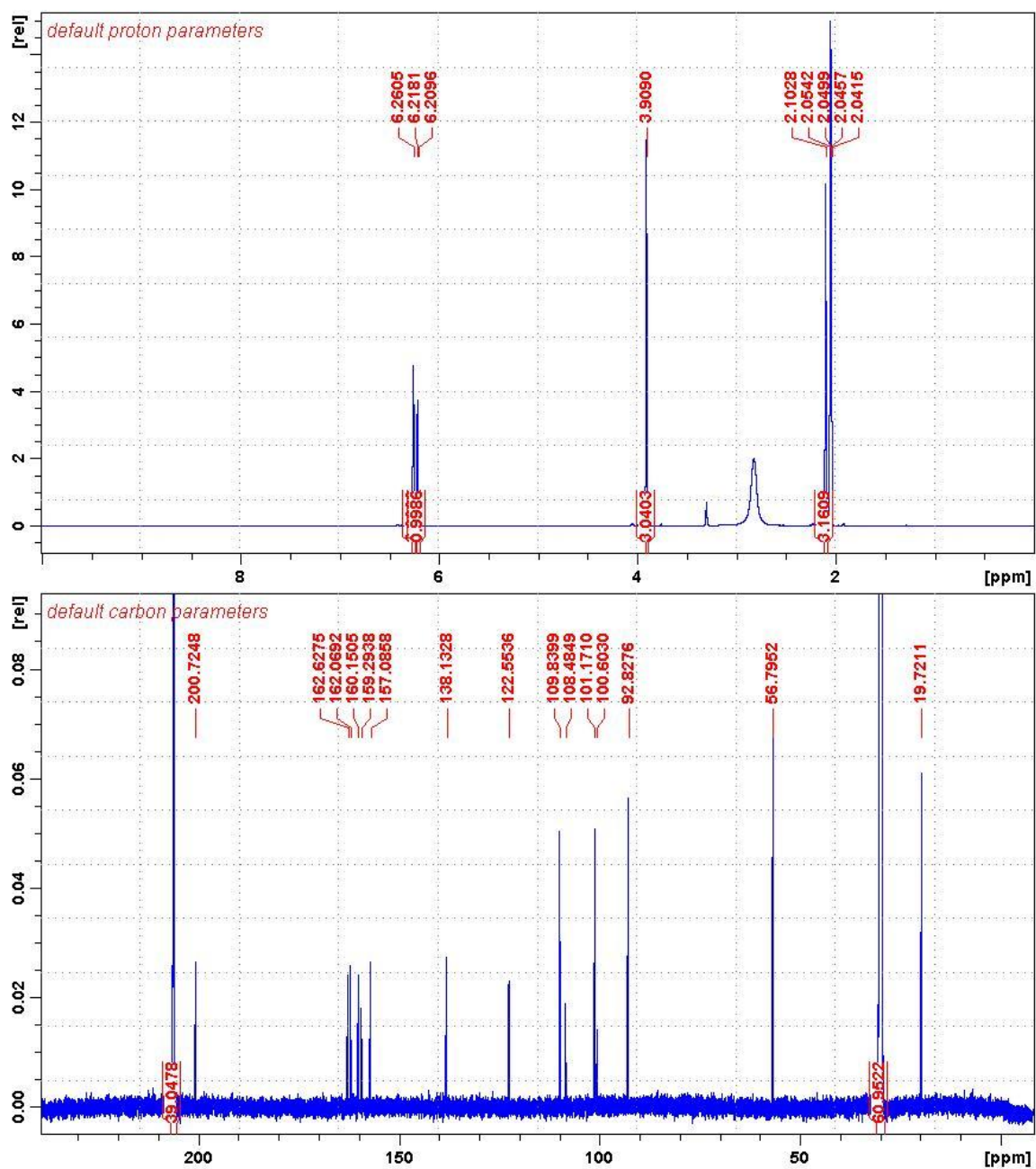
Figure A57: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **17** in $(CD_3)_2CO$ (500 MHz).
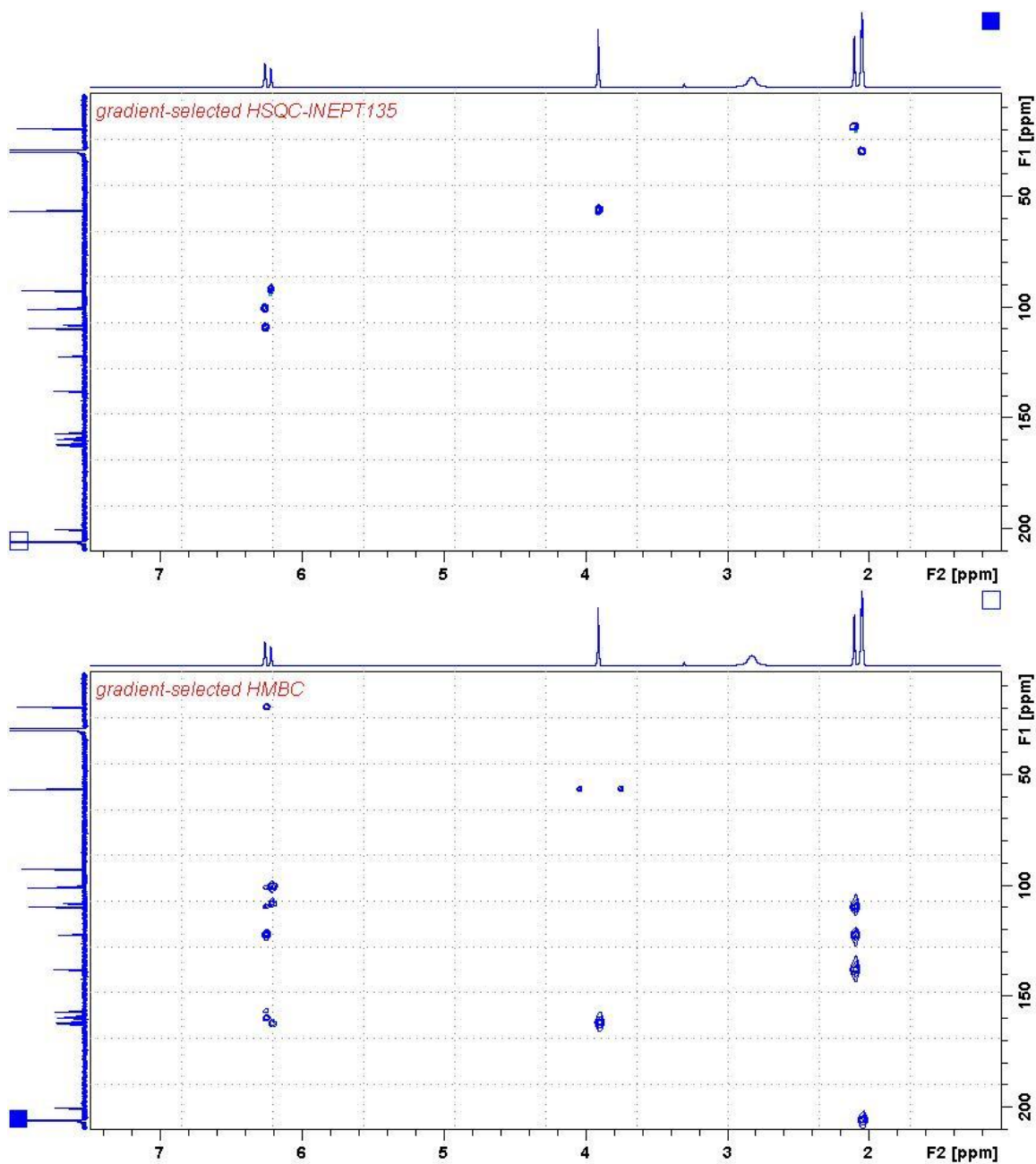
Figure A58: 2D HSQC (top) and HMBC (bottom) of **17** in $(CD_3)_2CO$ (500 MHz).
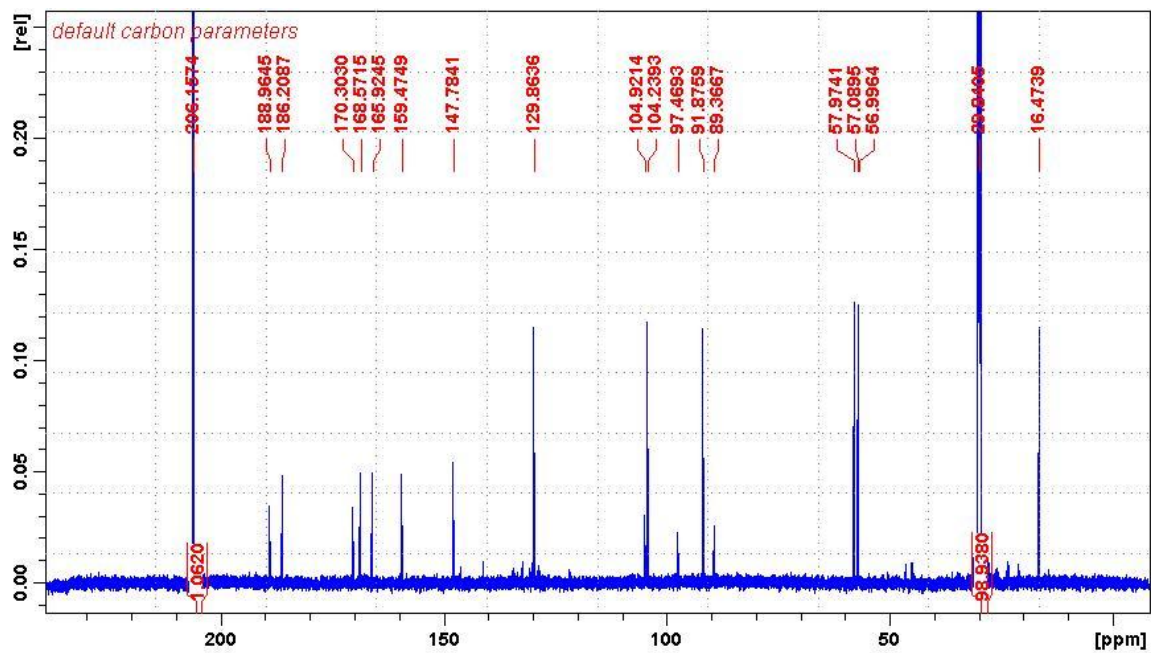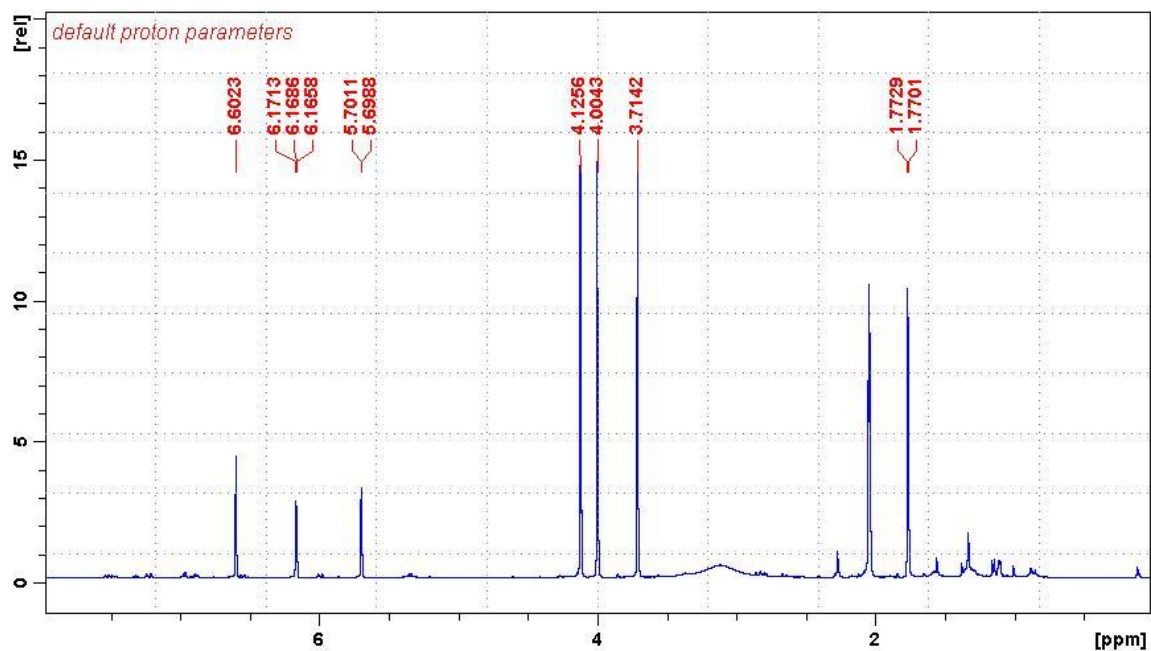
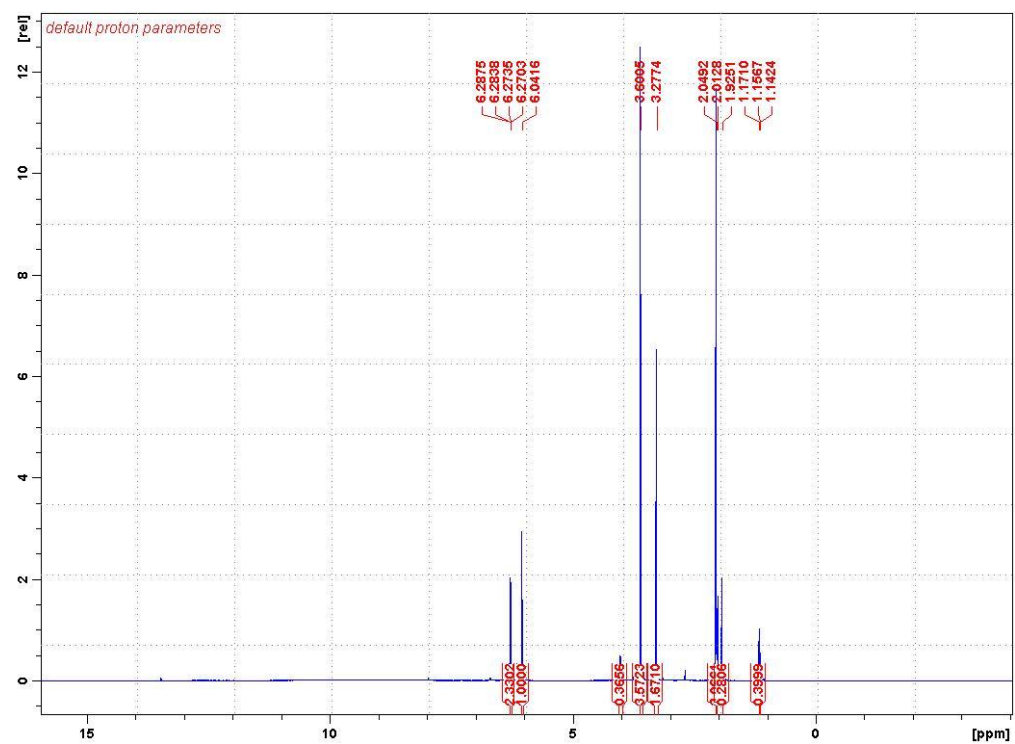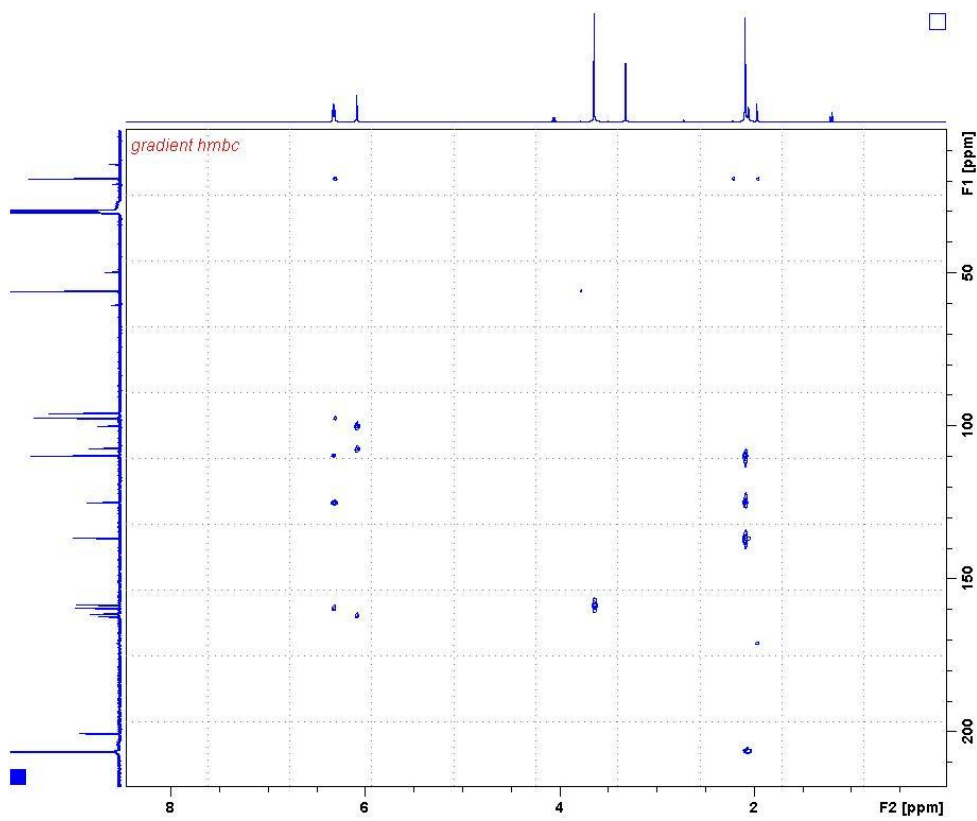**Figure A59:** 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **18** in $(CD_3)_2CO$ (500 MHz).
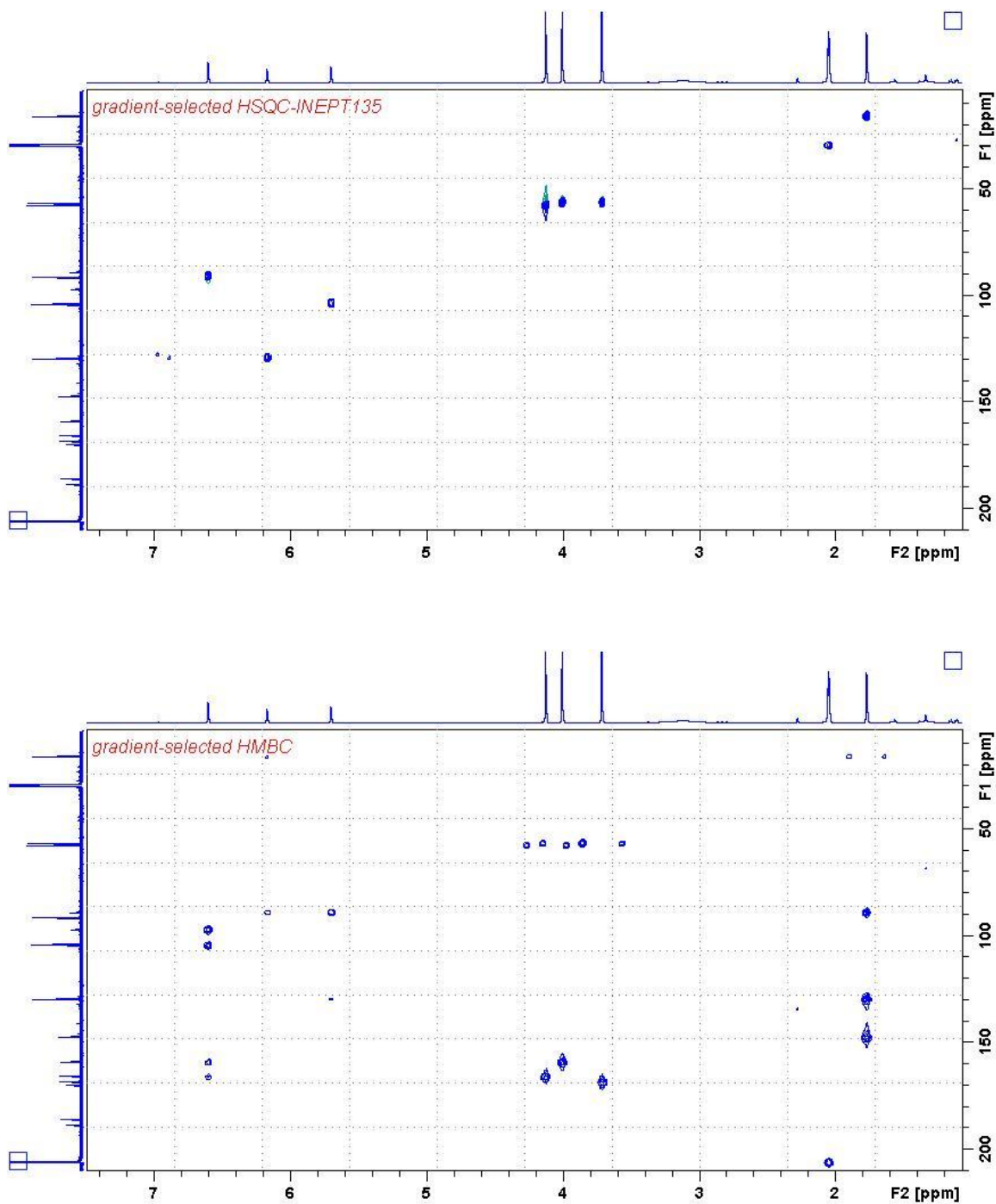
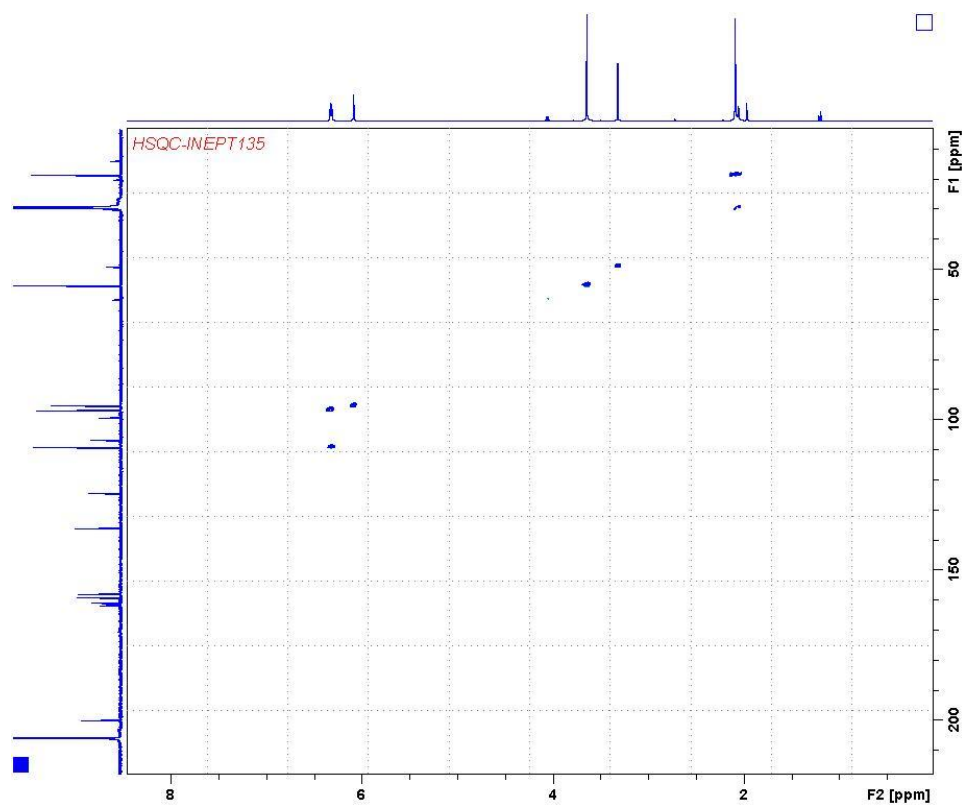**Figure A60:** 2D HSQC (top) and HMBC (bottom) of **18** in (CD$_3$)$_2$CO (500 MHz).
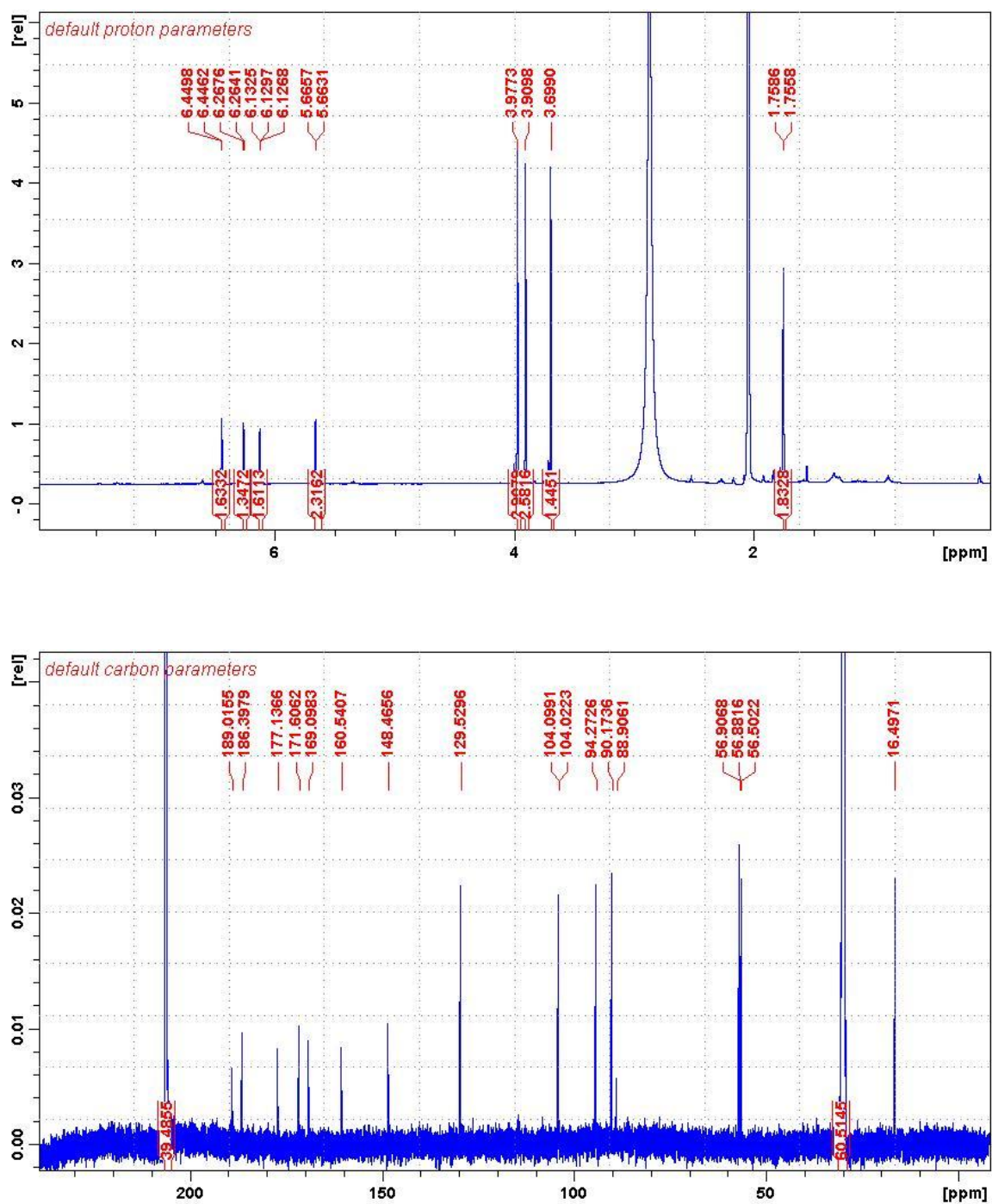
Figure A61: 1D ^{1}H (top) and ^{13}C NMR (bottom)spectrum of **19** in $(CD_3)_2CO$ (500 MHz).
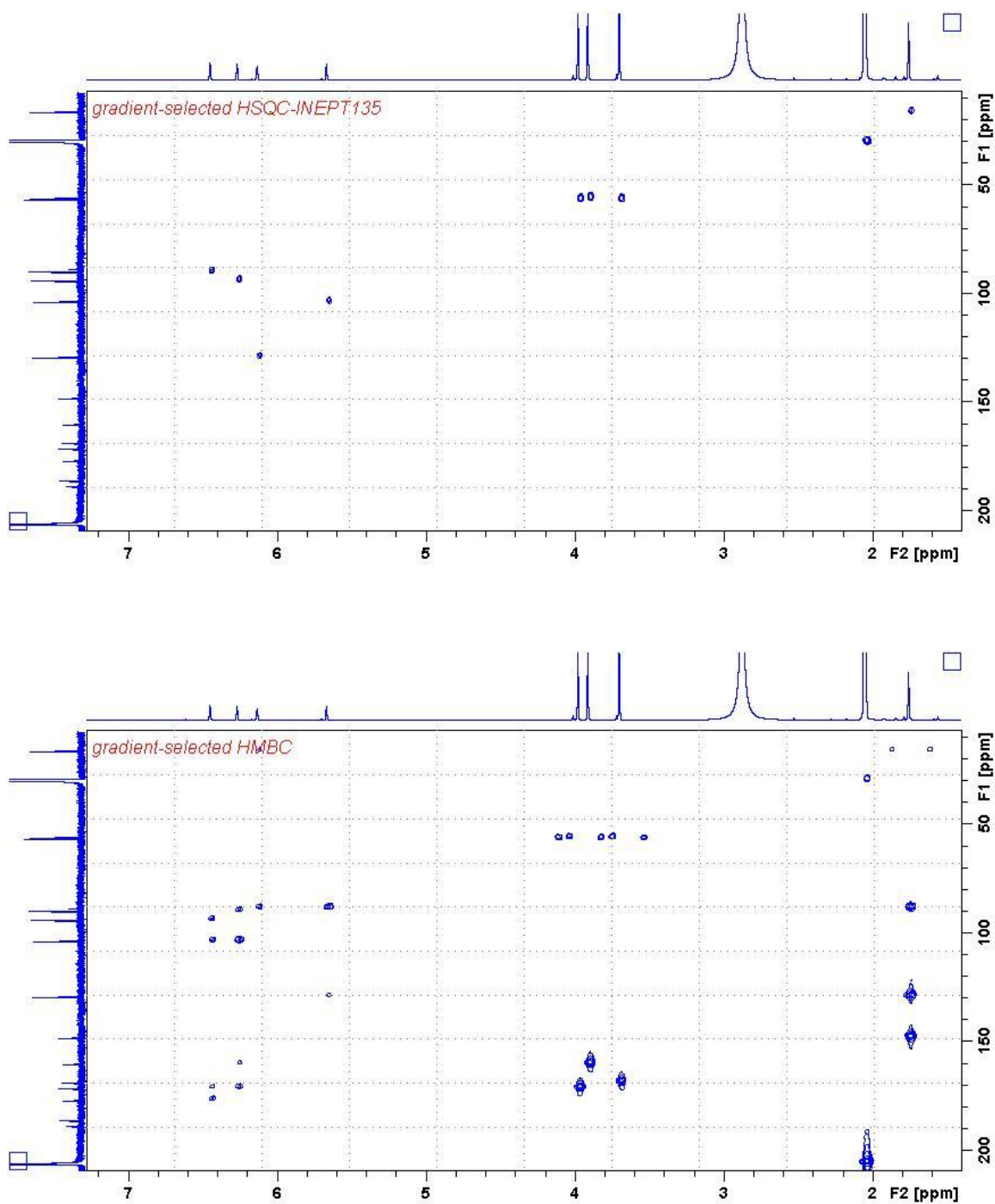
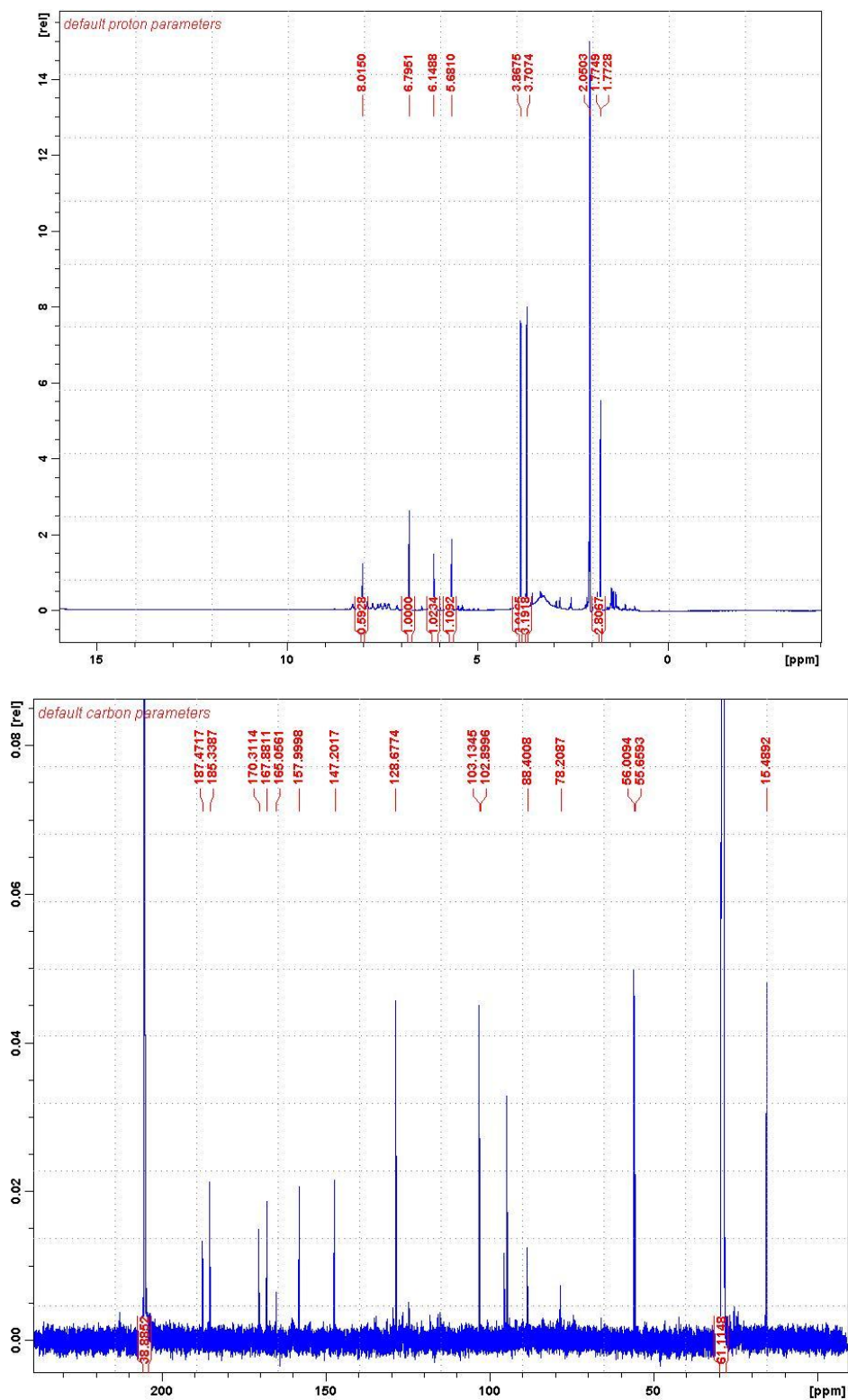**Figure A62:** 2D HSQC (top) and HMBC (bottom) of **19** in $(CD_3)_2CO$ (500 MHz).

SECTION 7: BIBLIOGRAPHY

1.   Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism - from biochemistry to genomics.* Nat Rev Micro, 2005. **3**(12): p. 937-947.

2.   Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the 30 Years from 1981 to 2010.* J Nat Prod, 2012. **75**(3): p. 311-335.

3.   Kardos, N. and A. Demain, *Penicillin: the medicine with the greatest impact on therapeutic outcomes.* App Microbiol Biotech, 2011. **92**(4): p. 677-687.

4.   Endo, A., *A historical perspective on the discovery of statins.* Proceedings of the Japan Academy, Series B, 2010. **86**(5): p. 484-493.

5.   Britton, S. and R. Palacios, *Cyclosporin A – usefulness, risks and mechanism of action.* Immunol Rev, 1982. **65**(1): p. 5-22.

6.   Huber, F.M., *Griseofulvin*, in *Mechanism of Action*, D. Gottlieb and P. Shaw, Editors. 1967, Springer Berlin Heidelberg. p. 181-189.

7.   Boeck, L. and R. Kastner, *Methods of producing the A-30912 antibiotics*, U.S. Patent, Editor. 1981, Eli Lilly and Company: United States.

8.   Schwarzer, D., R. Finking, and M.A. Marahiel, *Nonribosomal peptides: from genes to products.* Nat Prod Rep, 2003. **20**(3): p. 275-287.

9.   Finking, R. and M.A. Marahiel, *Biosynthesis of nonribosomal peptides.* Annu Rev Microbiol, 2004. **58**: p. 453-88.

10.   Mootz, H.D., D. Schwarzer, and M.A. Marahiel, *Ways of assembling complex natural products on modular nonribosomal peptide synthetases.* Chembiochem, 2002. **3**(6): p. 490-504.

11.   Gocht, M. and M.A. Marahiel, *Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis.* J Bacteriol, 1994. **176**(9): p. 2654-62.

12.   Conti, E., et al., *Structural basis for the activation of phenylalanine in the non‑ribosomal biosynthesis of gramicidin S.* EMBO J, 1997. **16**(14): p. 4174-4183.

13.   Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chem Biol, 1999. **6**(8): p. 493-505.

14. Eppelmann, K., T. Stachelhaus, and M.A. Marahiel, *Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics.* Biochemistry, 2002. **41**(30): p. 9718-9726.

15. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.

16. Koglin, A. and C.T. Walsh, *Structural insights into nonribosomal peptide enzymatic assembly lines.* Nat Prod Rep, 2009. **26**(8): p. 987-1000.

17. Weber, T., et al., *Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases.* Structure, 2000. **8**(4): p. 407-18.

18. Koglin, A., et al., *Conformational switches modulate protein interactions in peptide antibiotic synthetases.* Science, 2006. **312**(5771): p. 273-6.

19. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nat Struct Biol, 2002. **9**(7): p. 522-6.

20. Samel, S.A., et al., *Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase.* Structure, 2007. **15**(7): p. 781-92.

21. Tanovic, A., et al., *Crystal structure of the termination module of a nonribosomal peptide synthetase.* Science, 2008. **321**(5889): p. 659-63.

22. Keating, T.A., C.G. Marshall, and C.T. Walsh, *Vibriobactin biosynthesis in Vibrio cholerae: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains.* Biochemistry, 2000. **39**(50): p. 15513-21.

23. Stachelhaus, T., et al., *Peptide Bond Formation in Nonribosomal Peptide Biosynthesis: Catalytic role of the condensation domain.* J Biol Chem, 1998. **273**(35): p. 22773-22781.

24. Bergendahl, V., U. Linne, and M.A. Marahiel, *Mutational analysis of the C-domain in nonribosomal peptide synthesis.* Eur J Biochem, 2002. **269**(2): p. 620-629.

25. Walsh, C.T. and T.A. Wencewicz, *Prospects for new antibiotics: a molecule-centered perspective.* J Antibiot, 2014. **67**(1): p. 7-22.

26. Brakhage, A.A., *Molecular regulation of β-Lactam biosynthesis in filamentous fungi.* Microbiol Mol Biol Rev, 1998. **62**(3): p. 547-585.

27. Brakhage, A.A., et al., *Aspects on evolution of fungal β-lactam biosynthesis gene clusters and recruitment of trans-acting factors.* Phytochemistry, 2009. **70**(15–16): p. 1801-1811.

28. Byford, M.F., et al., *The mechanism of ACV synthetase.* Chem Rev, 1997. **97**(7): p. 2631-2650.

29. Roach, P.L., et al., *Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes.* Nature, 1995. **375**(6533): p. 700-4.

30. Lloyd, M.D., et al., *Controlling the Substrate Selectivity of Deacetoxycephalosporin/deacetylcephalosporin C Synthase.* J Biol Chem, 2004. **279**(15): p. 15420-15426.

31. Xu, W., D.J. Gavia, and Y. Tang, *Biosynthesis of fungal indole alkaloids.* Nat Prod Rep, 2014. **31**(10): p. 1474-87.

32. Waring, P. and J. Beaver, *Gliotoxin and related epipolythiodioxopiperazines.* Gen Pharmac, 1996. **27**(8): p. 1311-1316.

33. Balibar, C.J. and C.T. Walsh, *GliP, a multimodular nonribosomal peptide synthetase in Aspergillus fumigatus, makes the diketopiperazine scaffold of gliotoxin.* Biochemistry, 2006. **45**(50): p. 15029-15038.

34. Davis, C., et al., *The role of glutathione S-transferase GliG in gliotoxin biosynthesis in Aspergillus fumigatus.* Chem Biol, 2011. **18**(4): p. 542-552.

35. Scharf, D.H., et al., *A dedicated glutathione S-transferase mediates carbon–sulfur bond formation in gliotoxin biosynthesis.* J Am Chem Soc, 2011. **133**(32): p. 12322-12325.

36. Scharf, D.H., et al., *Epidithiodiketopiperazine Biosynthesis: A Four-Enzyme Cascade Converts Glutathione Conjugates into Transannular Disulfide Bridges.* Angew Chem Int Ed, 2013. **52**(42): p. 11092-11095.

37. Li, S., et al., *Comparative analysis of the biosynthetic systems for fungal bicyclo[2.2.2]diazaoctane indole alkaloids: the (+)/(-)-notoamide, paraherquamide and malbrancheamide pathways.* MedChemComm, 2012. **3**(8): p. 987-996.

38. Ding, Y., et al., *Genome-based characterization of two prenylation steps in the assembly of the stephacidin and notoamide anticancer agents in a marine-derived Aspergillus sp.* J Am Chem Soc, 2010. **132**(36): p. 12733-12740.

39.   Sunderhaus, J.D., et al., *Synthesis and bioconversions of notoamide T: a biosynthetic precursor to stephacidin A and notoamide B.* Org Lett, 2013. **15**(1): p. 22-25.

40.   Yin, W.-B., et al., *Acetylaszonalenin biosynthesis in Neosartorya fischeri.* J Biol Chem, 2009. **284**(1): p. 100-109.

41.   Zou, Y., et al., *Tandem Prenyltransferases Catalyze Isoprenoid Elongation and Complexity Generation in Biosynthesis of Quinolone Alkaloids.* J Am Chem Soc, 2015. **137**(15): p. 4980-4983.

42.   Walsh, C.T., et al., *Short pathways to complexity generation: fungal peptidyl alkaloid multicyclic scaffolds from anthranilate building blocks.* ACS Chem Biol, 2013. **8**(7): p. 1366-1382.

43.   Walsh, C.T., S.W. Haynes, and B.D. Ames, *Aminobenzoates as building blocks for natural product assembly lines.* Nat Prod Rep, 2012. **29**(1): p. 37-59.

44.   Walsh, C.T., *Biological matching of chemical reactivity: pairing indole nucleophilicity with electrophilic isoprenoids.* ACS Chem Biol, 2014. **9**(12): p. 2718-2728.

45.   Haynes, S.W., et al., *Complexity generation in fungal peptidyl alkaloid biosynthesis: a two-enzyme pathway to the hexacyclic MDR export pump inhibitor ardeemin.* ACS Chem Biol, 2013. **8**(4): p. 741-748.

46.   Haynes, S.W., et al., *Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two-enzyme pathway generates heptacyclic scaffold complexity in Asperlicin E.* J Am Chem Soc, 2012. **134**(42): p. 17444-17447.

47.   Gao, X., et al., *Fungal indole alkaloid biosynthesis: Genetic and biochemical investigation of the tryptoquialanine pathway in Penicillium aethiopicum.* J Am Chem Soc, 2011. **133**(8): p. 2729-2741.

48.   Wallwey, C. and S.-M. Li, *Ergot alkaloids: structure diversity, biosynthetic gene clusters and functional proof of biosynthetic genes.* Nat Prod Rep, 2011. **28**(3): p. 496-510.

49.   Haarmann, T., et al., *Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways.* ChemBioChem, 2006. **7**(4): p. 645-652.

50.   Ortel, I. and U. Keller, *Combinatorial Assembly of Simple and Complex d-Lysergic Acid Alkaloid Peptide Classes in the Ergot Fungus Claviceps purpurea.* J Biol Chem, 2009. **284**(11): p. 6650-6660.

51. Havemann, J., et al., *Cyclolization of D-lysergic acid alkaloid peptides.* Chem Biol, 2014. **21**(1): p. 146-155.

52. Daniel, J.F.d.S. and E. Rodrigues Filho, *Peptaibols of Trichoderma.* Nat Prod Rep, 2007. **24**(5): p. 1128-1141.

53. Mukherjee, P.K., et al., *Two classes of new peptaibols are synthesized by a single non-ribosomal peptide synthetase of Trichoderma virens.* J Biol Chem, 2011. **286**(6): p. 4544-4554.

54. Newman, D.J. and G.M. Cragg, *CHAPTER 1 Bioactive Macrocycles from Nature*, in *Macrocycles in Drug Discovery*. 2015, The Royal Society of Chemistry. p. 1-36.

55. Keating, T.A. and C.T. Walsh, *Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis.* Curr Opin Chem Biology, 1999. **3**(5): p. 598-606.

56. Giordanetto, F. and J. Kihlberg, *Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties?* J Med Chem, 2014. **57**(2): p. 278-295.

57. Huai, Q., et al., *Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12037-42.

58. Bushley, K.E., et al., *The genome of Tolypocladium inflatum: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster.* PLoS Genet, 2013. **9**(6): p. e1003496.

59. Chiang, Y.M., et al., *Molecular genetic mining of the Aspergillus secondary metabolome: discovery of the emericellamide biosynthetic pathway.* Chem Biol, 2008. **15**(6): p. 527-32.

60. Jin, J.-M., et al., *Functional characterization and manipulation of the apicidin biosynthetic pathway in Fusarium semitectum.* Mol Microbiol, 2010. **76**(2): p. 456-466.

61. Johnson, R.D., et al., *Cloning and characterization of a cyclic peptide synthetase gene from Alternaria alternata apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity.* Mol Plant-Microb Int, 2000. **13**(7): p. 742-753.

62. Harimoto, Y., et al., *Expression profiles of genes encoded by the supernumerary chromosome controlling AM-Toxin biosynthesis and pathogenicity in the apple pathotype of Alternaria alternata.* Mol Plant-Microbe Int, 2007. **20**(12): p. 1463-1476.

63.   Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.

64.   Chooi, Y.H. and Y. Tang, *Navigating the fungal polyketide chemical space: from genes to molecules.* J Org Chem, 2012. **77**(22): p. 9933-53.

65.   Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed, 2009. **48**(26): p. 4688-4716.

66.   Tsai, S.C. and B.D. Ames, *Chapter 2 Structural enzymology of polyketide synthases*, in *Methods Enzymol*, A.H. David, Editor. 2009, Academic Press. p. 17-47.

67.   Maier, T., S. Jenni, and N. Ban, *Architecture of mammalian fatty acid synthase at 4.5 Å resolution.* Science, 2006. **311**(5765): p. 1258-1262.

68.   Maier, T., M. Leibundgut, and N. Ban, *The crystal structure of a mammalian fatty acid synthase.* Science, 2008. **321**(5894): p. 1315-1322.

69.   Keatinge-Clay, A.T., et al., *Catalysis, specificity, and ACP docking site of Streptomyces coelicolor malonyl-CoA:ACP transacylase.* Structure, 2003. **11**(2): p. 147-154.

70.   Serre, L., et al., *The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-Å Resolution.* J Biol Chem, 1995. **270**(22): p. 12961-12964.

71.   Tang, Y., et al., *Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase.* Chem Biol, 2007. **14**(8): p. 931-943.

72.   Keatinge-Clay, A.T. and R.M. Stroud, *The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases.* Structure, 2006. **14**(4): p. 737-748.

73.   Gay, D., et al., *Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase.* Biochemistry, 2013. **52**(49): p. 8916-8928.

74.   Leesong, M., et al., *Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site.* Structure, 1996. **4**(3): p. 253-264.

75.   Cox, R.J., *Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets.* Org Biomol Chem, 2007. **5**(13): p. 2010-26.

76. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Nat Prod Rep, 2007. **24**(1): p. 162-190.

77. Hashimoto, M., T. Nonaka, and I. Fujii, *Fungal type III polyketide synthases.* Nat Prod Rep, 2014. **31**(10): p. 1306-1317.

78. Katsuyama, Y. and Y. Ohnishi, *Chapter Sixteen - Type III polyketide synthases in microorganisms*, in *Methods Enzymol*, A.H. David, Editor. 2012, Academic Press. p. 359-377.

79. Crawford, J.M. and C.A. Townsend, *New insights into the formation of fungal aromatic polyketides.* Nat Rev Microbiol, 2010. **8**(12): p. 879-89.

80. Beck, J., et al., *The multifunctional 6-methylsalicylic acid synthase gene of Penicillium patulum. Its gene structure relative to that of other polyketide synthases.* Eur J Biochem, 1990. **192**(2): p. 487-98.

81. Fujii, I., et al., *Cloning of the polyketide synthase gene atX from Aspergillus terreus and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression.* Mol. Gen. Genet., 1996. **253**(1-2): p. 1-10.

82. Crawford, J.M., et al., *Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase.* Proc Natl Acad Sci USA, 2006. **103**(45): p. 16728-16733.

83. Crawford, J.M., et al., *Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization.* Nature, 2009. **461**(7267): p. 1139-1143.

84. Crawford, J.M., et al., *Deconstruction of iterative multidomain polyketide synthase function.* Science, 2008. **320**(5873): p. 243-246.

85. Korman, T.P., et al., *Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis.* Proc Natl Acad Sci USA, 2010. **107**(14): p. 6246-6251.

86. Zhou, H., et al., *Enzymatic synthesis of resorcylic acid lactones by cooperation of fungal iterative polyketide synthases involved in hypothemycin biosynthesis.* J Am Chem Soc, 2010. **132**(13): p. 4530-4531.

87. Zhou, H., et al., *Insights into radicicol biosynthesis via heterologous synthesis of intermediates and analogs.* J Biol Chem, 2010. **285**(53): p. 41412-41421.

88. Gao, Z., et al., *Investigation of fungal iterative polyketide synthase functions using partially assembled intermediates.* J Am Chem Soc, 2013. **135**(5): p. 1735-1738.

89.   Xu, Y., et al., *Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat shock response-modulating anticancer fungal polyketide from Aspergillus terreus.* Appl Environ Microbiol, 2013. **79**(6): p. 2038-2047.

90.   Xu, Y., et al., *Rational reprogramming of fungal polyketide first-ring cyclization.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5398-5403.

91.   Zabala, Angelica O., et al., *Characterization of a silent azaphilone gene cluster from Aspergillus niger ATCC 1015 reveals a hydroxylation-mediated pyran-ring formation.* Chem Biol, 2012. **19**(8): p. 1049-1059.

92.   Chiang, Y.-M., et al., *A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in Aspergillus nidulans.* J Am Chem Soc, 2009. **131**(8): p. 2965-2970.

93.   Zhou, H., et al., *A fungal ketoreductase domain that displays substrate-dependent stereospecificity.* Nat Chem Biol, 2012. **8**(4): p. 331-333.

94.   Ma, S.M., et al., *Complete reconstitution of a highly reducing iterative polyketide synthase.* Science, 2009. **326**(5952): p. 589-592.

95.   Xie, X., et al., *Acyltransferase mediated polyketide release from a fungal megasynthase.* J Am Chem Soc, 2009. **131**(24): p. 8388-8389.

96.   Abe, Y., et al., *Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in Penicillium citrinum.* Mol Genet Genomics, 2002. **267**(5): p. 636-646.

97.   Xu, W., et al., *LovG: the thioesterase required for dihydromonacolin L release and lovastatin nonaketide synthase turnover in lovastatin biosynthesis.* Angew Chem Int Ed, 2013. **52**(25): p. 6472-6475.

98.   Kakule, T.B., et al., *Two related pyrrolidinedione synthetase loci in Fusarium heterosporum ATCC 74349 produce divergent metabolites.* ACS Chem Biol, 2013. **8**(7): p. 1549-1557.

99.   Eley, K.L., et al., *Biosynthesis of the 2-pyridone tenellin in the insect pathogenic fungus Beauveria bassiana.* ChemBioChem, 2007. **8**(3): p. 289-297.

100.   Fisch, K.M., et al., *Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite.* J Am Chem Soc, 2011. **133**(41): p. 16635-16641.

101.   Halo, L.M., et al., *Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires coexpression with an enoyl reductase.* ChemBioChem, 2008. **9**(4): p. 585-594.

102.   Fujii, I., et al., *Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of Aspergillus nidulans.* Chem Biol, 2001. **8**(2): p. 189-197.

103.   Chooi, Y.-H., R. Cacho, and Y. Tang, *Identification of the viridicatumtoxin and griseofulvin gene clusters from Penicillium aethiopicum.* Chem Biol, 2010. **17**(5): p. 483-494.

104.   Crawford, J.M., et al., *Synthetic strategy of nonreducing iterative polyketide synthases and the origin of the classical "starter-unit effect".* ChemBioChem, 2008. **9**(7): p. 1019-1023.

105.   Ma, S.M., et al., *Enzymatic synthesis of aromatic polyketides using PKS4 from Gibberella fujikuroi.* J Am Chem Soc, 2007. **129**(35): p. 10642-10643.

106.   Townsend, C.A., S.B. Christensen, and K. Trautwein, *Hexanoate as a starter unit in polyketide biosynthesis.* J Am Chem Soc, 1984. **106**(13): p. 3868-3869.

107.   Udwary, D.W., M. Merski, and C.A. Townsend, *A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase.* J Mol Biol, 2002. **323**(3): p. 585-598.

108.   Zhou, H., et al., *A polyketide macrolactone synthase from the filamentous fungus Gibberella zeae.* Proc Natl Acad Sci U S A, 2008. **105**(17): p. 6249-6254.

109.   Huitt-Roehl, C.R., et al., *Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA.* ACS Chem Biol, 2015.

110.   Liu, T., et al., *Engineering of an "Unnatural" Natural Product by Swapping Polyketide Synthase Domains in Aspergillus nidulans.* J Am Chem Soc, 2011. **133**(34): p. 13314-13316.

111.   Zhou, H., Y. Li, and Y. Tang, *Cyclization of aromatic polyketides from bacteria and fungi.* Nat Prod Rep, 2010. **27**(6): p. 839-868.

112.   Li, Y., W. Xu, and Y. Tang, *Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains.* J Biol Chem, 2010. **285**(30): p. 22764-22773.

113.   Vagstad, A.L., et al., *Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity.* J Am Chem Soc, 2012. **134**(15): p. 6865-6877.

114. Bailey, A.M., et al., *Characterisation of 3-methylorcinaldehyde synthase (MOS) in Acremonium strictum: first observation of a reductive release mechanism during polyketide biosynthesis.* Chem Comm, 2007(39): p. 4053-4055.

115. Regueira, T.B., et al., *Molecular basis for mycophenolic acid biosynthesis in Penicillium brevicompactum.* Appl Environ Microbiol, 2011. **77**(9): p. 3035-3043.

116. Lo, H.-C., et al., *Two separate gene clusters encode the biosynthetic pathway for the meroterpenoids austinol and dehydroaustinol in Aspergillus nidulans.* J Am Chem Soc, 2012. **134**(10): p. 4709-4720.

117. Szewczyk, E., et al., *Identification and characterization of the asperthecin gene cluster of Aspergillus nidulans.* Appl Environ Microbiol, 2008. **74**(24): p. 7607-7612.

118. Li, Y., et al., *Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an α-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase.* J Am Chem Soc, 2011. **133**(39): p. 15773-15785.

119. Vincent, J.L., et al., *International study of the prevalence and outcomes of infection in intensive care units.* JAMA, 2009. **302**(21): p. 2323-9.

120. Arendrup, M.C., *Epidemiology of invasive candidiasis.* Curr Opin Crit Care, 2010. **16**(5): p. 445-52.

121. Mayr, A., M. Aigner, and C. Lass-Florl, *Caspofungin: when and how? The microbiologist's view.* Mycoses, 2012. **55**(1): p. 27-35.

122. Andes, D.R., et al., *Impact of treatment strategy on outcomes in patients with candidemia and other forms of invasive candidiasis: a patient-level quantitative review of randomized trials.* Clin Infect Dis, 2012. **54**(8): p. 1110-22.

123. Kett, D.H., et al., *Anidulafungin compared with fluconazole in severely ill patients with candidemia and other forms of invasive candidiasis: support for the 2009 IDSA treatment guidelines for candidiasis.* Crit Care, 2011. **15**(5): p. R253.

124. Benz, F., et al., *Stoffwechselprodukte von Mikroorganismen 143. Mitteilung. Echinocandin B, ein neuartiges Polypeptid-Antibioticum aus Aspergillus nidulans var. echinulatus: Isolierung und Bausteine.* Helv Chim Acta, 1974. **57**(8): p. 2459-2477.

125. Schwartz, R.E., et al., *Pneumocandins from Zalerion arboricola. I. Discovery and isolation.* J Antibiot (Tokyo), 1992. **45**(12): p. 1853-66.

126.	Mizuno, K., et al., *Studies on aculeacin. I. Isolation and characterization of aculeacin A.* J Antibiot (Tokyo), 1977. **30**(4): p. 297-302.

127.	Strobel, G.A., et al., *Cryptocandin, a potent antimycotic from the endophytic fungus Cryptosporiopsis cf. quercina.* Microbiology, 1999. **145 (Pt 8**): p. 1919-26.

128.	Roy, K., et al., *Mulundocandin, a new lipopeptide antibiotic. I. Taxonomy, fermentation, isolation and characterization.* J Antibiot (Tokyo), 1987. **40**(3): p. 275-80.

129.	Boeck, L.D., et al., *Deacylation of echinocandin B by Actinoplanes utahensis.* J Antibiot (Tokyo), 1989. **42**(3): p. 382-8.

130.	Debono, M., et al., *Synthesis of new analogs of echinocandin B by enzymatic deacylation and chemical reacylation of the echinocandin B peptide: synthesis of the antifungal agent cilofungin (LY121019).* J Antibiot (Tokyo), 1989. **42**(3): p. 389-97.

131.	Debono, M., et al., *Semisynthetic chemical modification of the antifungal lipopeptide echinocandin B (ECB): structure-activity studies of the lipophilic and geometric parameters of polyarylated acyl analogs of ECB.* J Med Chem, 1995. **38**(17): p. 3271-81.

132.	Bouffard, F.A., et al., *Synthesis and antifungal activity of novel cationic pneumocandin B_o derivatives.* J Med Chem, 1994. **37**(2): p. 222-5.

133.	Tomishima, M., et al., *FK463, a novel water-soluble echinocandin lipopeptide: synthesis and antifungal activity.* J Antibiot (Tokyo), 1999. **52**(7): p. 674-6.

134.	Joseph, J.M., R. Kim, and A.C. Reboli, *Anidulafungin: a drug evaluation of a new echinocandin.* Expert Opin Pharmacother, 2008. **9**(13): p. 2339-48.

135.	Hensens, O.D., et al., *Pneumocandins from Zalerion arboricola. III. Structure elucidation.* J Antibiot (Tokyo), 1992. **45**(12): p. 1875-85.

136.	Leonard, W.R., Jr., et al., *Synthesis of the antifungal beta-1,3-glucan synthase inhibitor CANCIDAS (caspofungin acetate) from pneumocandin B0.* J Org Chem, 2007. **72**(7): p. 2335-43.

137.	Kurokawa, N. and Y. Ohfune, *Total synthesis of echinocandins. II. Total synthesis of echinocandin D via efficient peptide coupling reactions.* J Am Chem Soc, 1986. **108**(19): p. 6043-6045.

138. Evans, D.A. and A.E. Weber, *Synthesis of the cyclic hexapeptide echinocandin D. New approaches to the asymmetric synthesis of b-hydroxy-a-amino acids.* J Am Chem Soc, 1987. **109**(23): p. 7151-7157.

139. Kurokawa, N. and Y. Ohfune, *Synthetic studies on antifungal cyclic peptides, echinocandins. Stereoselective total synthesis of echinocandin D via a novel peptide coupling.* Tetrahedron, 1993. **49**(28): p. 6195-6222.

140. Toth, V., et al., *Polyphasic characterization of "Aspergillus nidulans var. roseus" ATCC 58397.* Folia Microbiol (Praha), 2011. **56**(5): p. 381-8.

141. Carlson, J.E., et al., *Segregation of random amplified DNA markers in F1 progeny of conifers.* Theor Appl Genet, 1991. **83**(2): p. 194-200.

142. Li, R., et al., *De novo assembly of human genomes with massively parallel short read sequencing.* Genome Res, 2010. **20**(2): p. 265-72.

143. Szewczyk, E., et al., *Fusion PCR and gene targeting in Aspergillus nidulans.* Nat Protoc, 2006. **1**(6): p. 3111-20.

144. Hays, S. and E. Selker, *Making the selectable marker bar tighter and more economical.* Fungal Genet Newl, 2000. **47**: p. 107.

145. Galagan, J.E., et al., *Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae.* Nature, 2005. **438**(7071): p. 1105-15.

146. Wiest, A., et al., *Identification of peptaibols from Trichoderma virens and cloning of a peptaibol synthetase.* J Biol Chem, 2002. **277**(23): p. 20862-8.

147. Marahiel, M.A., T. Stachelhaus, and H.D. Mootz, *Modular peptide synthetases involved in nonribosomal peptide synthesis.* Chem Rev, 1997. **97**(7): p. 2651-2674.

148. Röttig, M., et al., *NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity.* Nucleic Acids Res, 2011. **39**(suppl 2): p. W362-W367.

149. Gao, X., et al., *Cyclization of fungal nonribosomal peptides catalyzed by a terminal condensation-like domain.* Nat Chem Biol, 2012. **8**(10): p. 823-830.

150. Nayak, T., et al., *A versatile and efficient gene-targeting system for Aspergillus nidulans.* Genetics, 2006. **172**(3): p. 1557-1566.

151. Kraas, F.I., et al., *Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation.* Chem Biol, 2010. **17**(8): p. 872-80.

152. Chooi, Y.H. and Y. Tang, *Adding the lipo to lipopeptides: do more with less.* Chem Biol, 2010. **17**(8): p. 791-3.

153. Hansen, D.B., et al., *The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity.* J Am Chem Soc, 2007. **129**(20): p. 6366-7.

154. Wittmann, M., et al., *Role of DptE and DptF in the lipidation reaction of daptomycin.* FEBS J, 2008. **275**(21): p. 5343-54.

155. Baltz, R.H., V. Miao, and S.K. Wrigley, *Natural products to drugs: daptomycin and related lipopeptide antibiotics.* Nat Prod Rep, 2005. **22**(6): p. 717-41.

156. Mootz, H.D., R. Finking, and M.A. Marahiel, *4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis.* J Biol Chem, 2001. **276**(40): p. 37289-98.

157. Stachelhaus, T., A. Huser, and M.A. Marahiel, *Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases.* Chem Biol, 1996. **3**(11): p. 913-21.

158. Stachelhaus, T., et al., *Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain.* J Biol Chem, 1998. **273**(35): p. 22773-81.

159. Pfeifer, B.A., et al., *Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli.* Science, 2001. **291**(5509): p. 1790-1792.

160. Gadow, A., et al., *Gramicidin S synthetase. Stability of reactive thioester intermediates and formation of 3-amino-2-piperidone.* Eur J Biochem, 1983. **132**(2): p. 229-234.

161. Adefarati, A.A., et al., *Biosynthesis of L-671,329, an echinocandin-type antibiotic produced by Zalerion arboricola: origins of some of the unusual amino acids and the dimethylmyristic acid side chain.* J Am Chem Soc, 1991. **113**(9): p. 3542-3545.

162. Underhill, E.W., *Biosynthesis of mustard oil glucosides: 3-benzylmalic acid, a precursor of 2-amino-4-phenylbutyric acid and of gluconasturtiin.* Can J Biochem, 1968. **46**(5): p. 401-5.

163. Harimoto, Y., et al., *Expression profiles of genes encoded by the supernumerary chromosome controlling AM-toxin biosynthesis and pathogenicity in the apple pathotype of Alternaria alternata.* Mol Plant Microbe Interact, 2007. **20**(12): p. 1463-76.

164. Zhang, S., et al., *A fragmented aflatoxin-like gene cluster in the forest pathogen Dothistroma septosporum.* Fungal Genet Biol, 2007. **44**(12): p. 1342-54.

165. Scherlach, K., et al., *Two induced fungal polyketide pathways converge into antiproliferative spiroanthrones.* ChemBioChem, 2011. **12**(12): p. 1836-1839.

166. Ames, B.D., et al., *Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis.* Proc Natl Acad Sci U S A, 2012. **109**(28): p. 11144-11149.

167. Hutton, J.J., Jr., A. Kaplan, and S. Udenfriend, *Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase.* Arch Biochem Biophys, 1967. **121**(2): p. 384-91.

168. Petersen, L., et al., *Novel proline hydroxylase activities in the pneumocandin-producing fungus Glarea lozoyensis responsible for the formation of trans 3- and trans 4-hydroxyproline.* Appl Microbiol Biotechnol, 2003. **62**(2): p. 263-267.

169. Lawrence, C.C., et al., *Purification and initial characterization of proline 4-hydroxylase from Streptomyces griseoviridus P8648: a 2-oxoacid, ferrous-dependent dioxygenase involved in etamycin biosynthesis.* Biochem J, 1996. **313 (Pt 1**): p. 185-91.

170. Hausinger, R.P., *Fe(II)/α-ketoglutarate-dependent hydroxylases and related enzymes.* Crit Rev Biochem Mol Biol, 2004. **39**(1): p. 21-68.

171. Luesch, H., et al., *Biosynthesis of 4-methylproline in cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities.* J Org Chem, 2003. **68**(1): p. 83-91.

172. Becker, J.E., R.E. Moore, and B.S. Moore, *Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization.* Gene, 2004. **325**: p. 35-42.

173. Hibi, M., et al., *L-Leucine 5-hydroxylase of Nostoc punctiforme is a novel type of Fe(II)/α-ketoglutarate-dependent dioxygenase that is useful as a biocatalyst.* Appl Microbiol Biotechnol, 2012: p. (in press).doi: 10.1007/s00253-012-4136-7.

174. Benz, F., et al., *Stoffwechselprodukte von Mikroorganismen 143. Mitteilung. Echinocandin B, ein neuartiges Polypeptid-Antibioticum aus Aspergillus nidulans var. echinulatus: Isolierung und Bausteine.* Helv. Chim. Acta, 1974. **57**(8): p. 2459-2477.

175. Sawistowska-Schröder, E.T., D. Kerridge, and H. Perry, *Echinocandin inhibition of 1,3-β-D-glucan synthase from Candida albicans.* FEBS Lett., 1984. **173**(1): p. 134-138.

176.   Tomoshima, M., et al., *FK463, a novel water-soluble echinocandin lipopeptide: synthesis and antifungal activity.* J Antibiot (Tokyo), 1999. **52**: p. 674-676.

177.   Traber, R., et al., *Cyclopeptid-antibiotika aus Aspergillus-arten. struktur der Echinocandine C und D.* Helv Chim Acta, 1979. **62**(4): p. 1252-1267.

178.   Tóth, V., et al., *Polyphasic characterization of "Aspergillus nidulans var. roseus" ATCC 58397.* Folia Microbiol., 2011. **56**(5): p. 381-388.

179.   Cacho, R.A., et al., *Identification and characterization of the Echinocandin B biosynthetic gene cluster from Emericella rugulosa NRRL 11440.* J Am Chem Soc, 2012. **134**(40): p. 16781-16790.

180.   Adefarati, A.A., et al., *Biosynthesis of L-671,329, an echinocandin-type antibiotic produced by Zalerion arboricola: origins of some of the unusual amino acids and the dimethylmyristic acid side chain.* J. Am. Chem. Soc., 1991. **113**(9): p. 3542-3545.

181.   Aik, W., et al., *Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases.* Curr Opin Struct Biol, 2012. **22**(6): p. 691-700.

182.   Schofield, C.J. and Z. Zhang, *Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes.* Curr. Opin. Struct. Biol., 1999. **9**(6): p. 722-731.

183.   Clifton, I.J., et al., *Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins.* J. Inorg. Biochem., 2006. **100**(4): p. 644-669.

184.   Elkins, J.M., et al., *X-ray crystal structure of Escherichia coli taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates.* Biochemistry, 2002. **41**(16): p. 5185-5192.

185.   Wilmouth, R.C., et al., *Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana.* Structure, 2002. **10**(1): p. 93-103.

186.   Koehntop, K., J. Emerson, and L. Que, Jr., *The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes.* J Biol Inorg Chem, 2005. **10**(2): p. 87-93.

187.   Krebs, C., et al., *Non-heme Fe(IV)–oxo intermediates.* Acc Chem Res, 2007. **40**(7): p. 484-492.

188.   Bugg, T.D.H., *Dioxygenase enzymes: catalytic mechanisms and chemical models.* Tetrahedron, 2003. **59**(36): p. 7075-7101.

189. Kovaleva, E.G. and J.D. Lipscomb, *Versatility of biological non-heme Fe(II) centers in oxygen activation reactions.* Nat Chem Biol, 2008. **4**(3): p. 186-193.

190. Roach, P.L., et al., *Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes.* Nature, 1995. **375**(6533): p. 700-704.

191. Galonić, D.P., et al., *Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3.* Nat Chem Biol, 2007. **3**(2): p. 113-116.

192. Vaillancourt, F.H., J. Yin, and C.T. Walsh, *SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O_2-dependent halogenase.* Proc. Natl. Acad. Sci. U. S. A., 2005. **102**(29): p. 10111-10116.

193. Fitzpatrick, P.F., *Tetrahydropterin-dependent amino acid hydroxylases.* Ann. Rev. Biochem., 1999. **68**(1): p. 355-381.

194. Hausinger, R.P., *Fe(II)/α-ketoglutarate-dependent hydroxylases and related enzymes.* Crit. Rev. Biochem. Mol. Biol., 2004. **39**(1): p. 21-68.

195. Hegg, E.L. and L.Q. Jr, *The 2-his-1-carboxylate facial triad — an emerging structural motif in mononuclear non-heme iron(II) enzymes.* Eur J Biochem, 1997. **250**(3): p. 625-629.

196. Price, J.C., et al., *The first direct characterization of a high-valent iron intermediate in the reaction of an α-ketoglutarate-dependent dioxygenase: a high-spin Fe(IV) complex in taurine/α-ketoglutarate dioxygenase (TauD) from Escherichia coli.* Biochemistry, 2003. **42**(24): p. 7497-7508.

197. Krebs, C., et al., *Rapid freeze-quench ^{57}Fe Mössbauer spectroscopy: monitoring changes of an iron-containing active site during a biochemical reaction.* Inorg Chem, 2005. **44**(4): p. 742-757.

198. Proshlyakov, D.A., et al., *Direct detection of oxygen intermediates in the non-heme Fe enzyme taurine/α-ketoglutarate dioxygenase.* J. Am. Chem. Soc., 2004. **126**(4): p. 1022-1023.

199. Riggs-Gelasco, P.J., et al., *EXAFS spectroscopic evidence for an FeO Unit in the Fe(IV) intermediate observed during oxygen activation by taurine:α-ketoglutarate dioxygenase.* J. Am. Chem. Soc., 2004. **126**(26): p. 8108-8109.

200. Price, J.C., et al., *Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:α-ketoglutarate dioxygenase (TauD).* J. Am. Chem. Soc., 2003. **125**(43): p. 13008-13009.

201. Hasemann, C.A., et al., *Structure and function of cytochromes P450:a comparative analysis of three crystal structures.* Structure, 1995. **3**(1): p. 41-62.

202. Peterson, J.A. and S.E. Graham, *A close family resemblance: the importance of structure in understanding cytochromes P450.* Structure, 1998. **6**(9): p. 1079-1085.

203. Poulos, T.L. and E.F. Johnson, *Structures of cytochrome P450 enzymes*, in *Cytochrome P450 Structure Mechanism and Biochemistry*, P. Ortiz de Montellano, Editor. 2015, Springer International Publishing. p. 3-32.

204. Green, M.T., *CH bond activation in heme proteins: the role of thiolate ligation in cytochrome P450.* Curr. Opin. Chem. Biol., 2009. **13**(1): p. 84-88.

205. Vatsis, K.P., H.-M. Peng, and M.J. Coon, *Replacement of active-site cysteine-436 by serine converts cytochrome P450 2B4 into an NADPH oxidase with negligible monooxygenase activity.* J. Inorg. Biochem., 2002. **91**(4): p. 542-553.

206. Perera, R., et al., *Molecular basis for the inability of an oxygen atom donor ligand to replace the natural sulfur donor heme axial ligand in cytochrome P450 catalysis. Spectroscopic characterization of the Cys436Ser CYP2B4 mutant.* Arch. Biochem. Biophys., 2011. **507**(1): p. 119-125.

207. Sono, M., et al., *Heme-containing oxygenases.* Chem Rev, 1996. **96**(7): p. 2841-2888.

208. Coelho, P.S., et al., *A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo.* Nat. Chem. Biol., 2013. **9**(8): p. 485-487.

209. Yosca, T.H., et al., *Iron(IV)hydroxide pKa and the role of thiolate ligation in C–H Bond activation by cytochrome P450.* Science, 2013. **342**(6160): p. 825-829.

210. Denisov, I.G., et al., *Structure and chemistry of cytochrome P450.* Chem. Rev., 2005. **105**(6): p. 2253-2278.

211. Denisov, I.G. and S.G. Sligar, *Activation of molecular oxygen in cytochrome P450*, in *Cytochrome P450 Structure, Mechanism and Biochemistry*, P. Ortiz de Montellano, Editor. 2015, Springer International Publishing. p. 69-109.

212. Krest, C.M., et al., *Reactive intermediates in cytochrome P450 catalysis.* J Biol Chem, 2013. **288**(24): p. 17074-17081.

213. Rittle, J. and M.T. Green, *Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics.* Science, 2010. **330**(6006): p. 933-937.

214.	Hannemann, F., et al., *Cytochrome P450 systems—biological variations of electron transport chains.* Biochim. Biophys. Acta., 2007. **1770**(3): p. 330-344.

215.	Waskel, L. and J.-J.P. Kim, *Electron transfer partners of cytochrome P450*, in *Cytochrome P450 Structure, Mechanism and Biochemistry*, P. Ortiz de Montellano, Editor. 2015, Springer International Publishing. p. 33-68.

216.	Wang, M., et al., *Three-dimensional structure of NADPH–cytochrome P450 reductase: Prototype for FMN- and FAD-containing enzymes.* Proc. Natl. Acad. Sci. USA, 1997. **94**(16): p. 8411-8416.

217.	Vaillancourt, F.H., et al., *Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis.* Nature, 2005. **436**(7054): p. 1191-1194.

218.	Koskinen, A.M.P. and H. Rapoport, *Synthesis of 4-substituted prolines as conformationally constrained amino acid analogs.* J. Org. Chem, 1989. **54**(8): p. 1859-1866.

219.	Luesch, H., et al., *Biosynthesis of 4-Methylproline in Cyanobacteria: Cloning of nosE and nosF Genes and Biochemical Characterization of the Encoded Dehydrogenase and Reductase Activities.* J. Org. Chem, 2003. **68**(1): p. 83-91.

220.	Schwalbe, R., L. Steele-Moore, and A.C. Goodwin, eds. *Antimicrobial susceptibility testing protocols*. 2007, CRC Press. 432.

221.	Boeck, L.D. and R.E. Kastner, *A-30912 antibiotics*. 1981, Eli Lilly and Co., USA .

222.	Strushkevich, N., et al., *Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10139-10143.

223.	Kurokawa, N. and Y. Ohfune, *Total synthesis of echinocandins. I. Stereocontrolled syntheses of the constituent amino acids.* J Am Chem Soc, 1986. **108**(19): p. 6041-6043.

224.	Mezl, V.A. and W.E. Knox, *Properties and analysis of a stable derivative of pyrroline-5-carboxylic acid for use in metabolic studies.* Anal. Biochem., 1976. **74**(2): p. 430-440.

225.	Hibi, M., et al., *L-Leucine 5-hydroxylase of Nostoc punctiforme is a novel type of Fe(II)/α-ketoglutarate-dependent dioxygenase that is useful as a biocatalyst.* Appl. Microbiol. Biotechnol., 2012: p. 1-6.

226. Zambias, R.A., et al., *Preparation and structure-activity relationships of simplified analogs of the antifungal agent cilofungin: a total synthesis approach.* J Med Chem, 1992. **35**(15): p. 2843-2855.

227. Boeck, L.D., D.S. Fukuda, and B.J. Abbott, *Deacylation of Echinocandin B by Actinoplanes utahensis.* J. Antibiot. (Tokyo), 1989. **42**: p. 382-388.

228. Becker, J.E., R.E. Moore, and B.S. Moore, *Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization.* Gene, 2004. **325**(0): p. 35-42.

229. Evans, B.S., et al., *Proteomics-based discovery of koranimine, a cyclic imine natural product.* J. Am. Chem. Soc., 2011. **133**(19): p. 7316-7319.

230. Lawrence, C.C., et al., *Purification and initial characterization of proline 4-hydroxylase from Streptomyces griseoviridus P8648: a 2-oxoacid, ferrous-dependent dioxygenase involved in etamycin biosynthesis.* Biochem. J., 1996. **313**(1): p. 185-191.

231. Frisvad, J.C. and R.A. Samson, *Polyphasic taxonomy of Penicillium subgenus Penicillium. A guide to identification of food and air-borne terverticillate Penicillia and their mycotoxins.* Studies in Mycology, 2004. **49**(1): p. 1-174.

232. Oxford, A.E., H. Raistrick, and P. Simonart, *Studies in the biochemistry of micro-organisms: Griseofulvin, $C_{17}H_{17}O_6Cl$, a metabolic product of Penicillium griseofulvum Dierckx.* Biochem. J., 1939. **33**(2): p. 240-248.

233. Gentles, J.C., *Experimental ringworm in guinea pigs: oral treatment with griseofulvin.* Nature, 1958. **182**(4633): p. 476-477.

234. Pillsbury, D.M., *Griseofulvin theraphy in dermatophytic infections.* Trans Am Clin Climatol Assoc, 1960. **71**: p. 52-57.

235. Gull, K. and A.P.J. Trinci, *Griseofulvin inhibits fungal mitosis.* Nature, 1973. **244**(5414): p. 292-294.

236. Finkelstein, E., B. Amichai, and M.H. Grunwald, *Griseofulvin and its uses.* Int J Antimicrob Agents, 1996. **6**(4): p. 189-194.

237. Panda, D., et al., *Kinetic suppression of microtubule dynamic instability by griseofulvin: Implications for its possible use in the treatment of cancer.* Proc Natl Acad Sci USA, 2005. **102**(28): p. 9878-9883.

238. Rebacz, B., et al., *Identification of griseofulvin as an inhibitor of centrosomal clustering in a phenotype-based screen.* Cancer Res, 2007. **67**(13): p. 6342-6350.

239.   Hutchinson, C.R., *The use of isotopic hydrogen and nuclear magnetic resonance spectroscopic techniques for the analysis of biosynthetic pathways.* J Nat Prod, 1982. **45**(1): p. 27-37.

240.   Birch, A.J., et al., *66. Studies in relation to biosynthesis. Part XIII. Griseofulvin.* J Chem Soc, 1958. **0**(0): p. 360-365.

241.   Birch, A. and F. Donovan, *Studies in relation to biosynthesis. I. Some possible routes to derivatives of orcinol and phloroglucinol.* Aus J Chem, 1953. **6**(4): p. 360-368.

242.   Birch, A., R. Massy-Westropp, and C. Moye, *Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in Penicillium griseofulvum Dierckx.* Aus J Chem, 1955. **8**(4): p. 539-544.

243.   Simpson, T.J. and J.S.E. Holker, *^{13}C-NMR studies on griseofulvin biosynthesis and acetate metabolism in Penicillium patulum.* Phytochemistry, 1977. **16**(2): p. 229-233.

244.   Barton, D.H.R. and T. Cohen, *In Festschrift Prof. Dr. Arthur Stoll*. 1957: Basel, Birkha□user.

245.   Rhodes, A., et al., *Biosynthesis of griseofulvin: the methylated benzophenone intermediates.* Biochem J, 1961. **81**: p. 28-37.

246.   Rhodes, A., G.A. Somerfield, and M.P. McGonagle, *Biosynthesis of griseofulvin. Observations on the incorporation of [^{14}C]Griseophenone C and [^{36}Cl]Griseophenones B and A.* Biochem J, 1963. **88**: p. 349-57.

247.   Harris, C.M., J.S. Roberson, and T.M. Harris, *Biosynthesis of griseofulvin.* J Am Chem Soc, 1976. **98**(17): p. 5380-6.

248.   Cove, D.J., *The induction and repression of nitrate reductase in the fungus Aspergillus nidulans.* Biochim et Biophys Acta, 1966. **113**(1): p. 51-56.

249.   Sambrook, J. and D.W. Russell, *Molecular cloning: a laboratory manual*. 2001: Cold Spring Harbor Laboratory.

250.   Szewczyk, E., et al., *Fusion PCR and gene targeting in Aspergillus nidulans.* Nat Protocols, 2007. **1**(6): p. 3111-3120.

251.   Barriuso, J., et al., *Double oxidation of the cyclic nonaketide dihydromonacolin L to monacolin J by a single cytochrome P450 monooxygenase, LovA.* J Am Chem Soc, 2011. **133**(21): p. 8078-81.

252. Ralston, L., et al., *Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from tobacco (Nicotiana tabacum).* Arch Biochem Biophys, 2001. **393**(2): p. 222-235.

253. Huang, K.-x., et al., *Molecular cloning and heterologous expression of the gene encoding dihydrogeodin oxidase, a multicopper blue enzyme from Aspergillus terreus.* J Biol Chem, 1995. **270**(37): p. 21495-21502.

254. Partow, S., et al., *Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae.* Yeast, 2010. **27**(11): p. 955-964.

255. Thorn, A., et al., *The crystal structure of progesterone 5β-reductase from Digitalis lanata defines a novel class of short chain dehydrogenases/reductases.* J Biol Chem, 2008. **283**(25): p. 17260-17269.

256. Lane, M.P., T.T. Nakashima, and J.C. Vederas, *Biosynthetic source of oxygens in griseofulvin. Spin-echo resolution of oxygen-18 isotope shifts in carbon-13 NMR spectroscopy.* J Am Chem Soc, 1982. **104**(3): p. 913-915.

257. Peres, V. and T.J. Nagem, *Trioxygenated naturally occurring xanthones.* Phytochemistry, 1997. **44**(2): p. 191-214.

258. Gesell, A., et al., *CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy.* J Biol Chem, 2009. **284**(36): p. 24432-24442.

259. Makino, M., et al., *Crystal structures and catalytic mechanism of cytochrome P450 StaP that produces the indolocarbazole skeleton.* Proc Natl Acad Sci USA, 2007. **104**(28): p. 11591-11596.

260. Vanholme, R., et al., *Lignin biosynthesis and structure.* Plant Physiol, 2010. **153**(3): p. 895-905.

261. Zerbe, K., et al., *An oxidative phenol coupling reaction catalyzed by OxyB, a cytochrome P450 from the vancomycin-producing microorganism.* Angew Chem Int Ed, 2004. **43**(48): p. 6709-6713.

262. Holding, A.N. and J.B. Spencer *Investigation into the mechanism of phenolic couplings during the biosynthesis of glycopeptide antibiotics.* ChemBioChem, 2008. **9**(14): p. 2209-2214.

263. Simpson, T.J., *Genetic and biosynthetic studies of the fungal prenylated xanthone shamixanthone and related metabolites in Aspergillus spp. revisited.* ChemBioChem, 2012. **13**(11): p. 1680-1688.

264. Benz, F., et al., *Stoffwechselprodukte von Mikroorganismen 143. Mitteilung. Echinocandin B, ein neuartiges Polypeptid-Antibioticum aus Aspergillus nidulans var. echinulatus: Isolierung und Bausteine.* Helvetica Chimica Acta, 1974. **57**(8): p. 2459-2477.