Characterization and Analysis of Biosynthetic Systems from *Nostoc* sp. ATCC 53789 and Selected Fungal Natural Product Pathways

by

Yousong Ding

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in The University of Michigan
2010

Doctoral Committee:

Professor David. H. Sherman, Chair
Professor David P. Ballou
Professor Shaomeng Wang
Associate Professor Anna K. Mapp
Assistant Professor Jason E. Gestwicki

To My Wife, Han, and Our Parents

Acknowledgements

The first and foremost person I would like to thank is my advisor, Dr. David Sherman. David generously allowed me to work on many interesting and challenging projects and gave me the freedom to pursue my interests in past four and half years. His continuous support, warm encouragement, and valuable guidance were critical in my productive doctoral studies and will be important in my future career development as well. His passion and love for science and fine personality not only made my work at this lab to be fun and enjoyable but also teach me how to face challenges in my life. He is the best model for me to be a scientist, teacher, and mentor.

I would like to express my sincere gratitude to Dr. David Ballou, Dr. Jason Gestwicki, Dr. Anna Mapp, and Dr. Shaomeng Wang for serving on my committee and for their scientific and insightful suggestions on my projects. I am especially grateful to both Jason and Shaomeng for their loads of help in my future career development.

I want to extend my gratitude to Dr. Robert M. Williams and his group in Chemistry Department of Colorado State University for an exciting collaboration and Dr. Paul Hollenberg in Biological Chemistry Department of University of Michigan for his generosity to share chemicals in CrpE studies.

I thank all present and past members of the Sherman group: Dr. Zachary Beck, Dr. Jeffery Kittendorf, Dr. Wolfgang Seufert, Dr. Douglas Burr, Dr. Liangcai Guo, Dr. Shengying Li, Dr. Fengan Yu, Dr. Pamela Schultz, Dr. Patricia Cruz Lopez, Kyle Bolduc, Chris Rath, Tonia Buchholz, Karoline Chiou, Tyler Nusca, Rafay Shareef, Jon Mortison, Shamilya Williams, and all I’ve failed to list here. I got immense help and support from them in past several years. Especially, Zack introduced cryptophycin and verrucarin projects to me and gave me invaluable suggestions to conduct and design experiments.

Funding to support my research mainly came from NIH. I also would like to thank the additional supports from Department of Medicinal Chemistry, Elizabeth Broomfield Foundation scholarship, Eli Lilly graduate fellowship, and Rackham Predoctoral Fellowship.

I would like to thank all of my friends for providing support and friendship that I needed. I can not list all names here but all of them made my life in Ann Arbor to be happy, healthy, and unforgotten.

Finally, I especially thank my best friend, supporter, soul-mate, and wife, Han. Han gives me all of her unconditioned support and love. She always has faith in me and my intellect, which even makes me blush sometimes. With her, both of my academic and social lives are meaningful, colorful, and enjoyable. Also, I truly thank all of my family members, especially my parents and parents-in-law, for their uncounted supports. My father loved and supported me with all he had in his whole life. My mom and parents-in-law came to Ann Arbor and gave me so many enjoyable moments and valuable supports.

Preface

This dissertation contains five chapters covering three major projects in my doctoral studies to understand natural product biosynthesis from different perspectives. Chapter 1 is an introduction of my studies to develop drugs from natural products and is partially adapted from one book chapter written by me and David in *Comprehensive Natural Products Chemistry Edition II*, which is in press. Chapter 2 describes my studies to generate cryptophycin analogs through chemoenzymatic approaches and to characterize novel biosynthetic enzymes. It is mainly adapted from two of our papers published in *Journal of the American Chemical Society* (**2008**, 130: 5492-5498) and *ACS Chemical Biology* (**2006**, 1: 766-779), and twos manuscripts to be submitted. Chapter 3 focuses on characterization of fungal alkaloid biosynthesis with different approaches and is adapted from four of our papers (*Organic Letters* **2008**, 10: 4863-4866; *Journal of Natural Product* **2008**, 71: 1574-1578; *Journal of Biological Chemistry* **2008**, 283: 16068-16076; and *Journal of Organic Chemistry* **2008**, 73: 3116-3119), and one manuscript to be submitted. Chapter 4 represents our latest understanding about trichothecene macrolide biosynthesis in a marine fungal strain and is adapted from one manuscript in preparation. Chapter 6 describes my own points of view about future work of these three projects and summarizes my entire doctoral research in this thesis. Due to the limitation of space, several other projects I conducted in my studies are notincluded here.

Table of Contents

List of Figures

Figure

List of Tables

Table

List of Abbreviations

PK	Polyketide
NRP	Nonribosomal peptide
PKS	Polyketide synthase
NRPS	Nonribosomal peptide synthetase
AT	Acyltransferase
KS	Ketoreductase
T	Thiolation
ACP	Acyl carrier protein
PCP	Peptidyl carrier protein
C	Condensation
DH	Dehydration
ER	Enoylreductase
MT	Methyltransferase
E	Epimerase
TE	Thioesterase
6DEB	6-Deoxyerythronolide
PrpE	Propionyl-CoA synthase
PCC	Propionyl-CoA carboxylase
Sfp	Phosphopantetheinyl transferase
MSRA	Methicillin-resistant Staphylococcus aureus
2KIC	2-Ketoisocaproic acid
2HIC	2-Hydroxyl-isocaproic acid
CrpD-M2	CrpD second module
AKGB	2-Keto-γ-(methylthio) butyrate
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
MBP	Maltose binding protein
Cr-1	Cryptophycin 1
CaM	Calmodulin
PDE1	Phosphodiesterase 1
DMAPP	Dimethylallyl pyrophosphate
DMATS	Dimethylallyl tryptophan synthase
SIM	Selective ion monitoring
PT	Prenyltransferase
FPP	Farnesyl pyrophosphate
DON	Deoxynivalenol
DAS	Diacetoxyscirpenol
TAS	3,4,15-Triacetoxyscirpenol
ORF	Open reading frame
GST	Glutathione *S*-transferase

Abstract

Characterization and Analysis of Biosynthetic Systems from *Nostoc* sp. ATCC 53789 and Selected Fungal Natural Product Pathways

by

Yousong Ding

Chair: David H. Sherman

Complex secondary metabolites display diverse biological activities and together with their derivatives have provided over two-thirds of new pharmaceutical agents introduced during the past two decades. However, limitations in isolation and in rapid structural determination continue to be inherent hurdles for using natural products as leads in drug discovery and design. My dissertation research focused on selected biosynthetic pathways with the hope to overcome some of these limitations. Three projects are described in this dissertation thesis. The first project demonstrates my efforts to generate natural product analogs using the biocatalysts, a strategy that provides significant advantages in catalytic specificity, efficiency, and impacts on the environment. Several natural and synthetic anticancer agent analogs were produced with a single P450 epoxidase and an excised thioesterase involved in the production of cryptophycin in *Nostoc* sp. Moreover, the homotropic and heterotropic cooperativity of the bacterial P450 epoxidase toward its substrates was characterized in details. This enzyme may serve as a more operable model to study the same features in several human P450s involving

in xenobiotcs metabolism. The second project describes how unique prenylated indole alkaloids are biosynthesized in various fungal genera. These biosynthetic pathways were extensively investigated by isolation and characterization of several key biosynthetic intermediates from *Penicillium*, *Aspergillus*, and *Malbranchea* sp. Subsequently, these pathways were examined at the first time through the elucidation of the biosynthetic gene cluster for stephacidin/notoamide from a marine *Aspergillus* strain and biochemical characterization of two critical aromatic prenyltransferases catalyzing two committed steps. Finally, one trichothecene macrolide gene cluster was cloned from a marine *Myrothecium verrucaria* strain and validated with biochemical characterization of a sesquiterpene synthase and a multifunctional P450, representing the latest understanding of the biosynthesis of structurally complex mycotoxins. With the identification and characterization of natural product gene clusters, more new fungal secondary metabolite analogs may be generated through metabolic engineering and heterologous production.

Chapter 1

Introduction

A significant number of natural products contain pharmacological activities that are beneficial to human health. There are many examples of natural product compounds that are in clinical use, including antibacterial penicillins, cephalosporins, immunosuppressive cyclosporine A, and the cholesterol-lowering HMG-CoA reductase inhibitors best known as the "statins" [1]. These natural products, which have achieved their functions over the course of millions of years of evolution, offer chemical scaffolds for development of new analogs with improved or altered functions. New bioactive analogs that contain novel structural elements may be generated by both semisynthesis and total synthesis efforts [2, 3]. In fact, more than two-thirds of newly introduced drugs worldwide were natural products or natural product derivatives in the past two decades [4].

The number of natural products that function in biological systems is large but represents only a small fraction of the total possible number of small carbon-based compounds, indicating the importance of stereochemistry and functional groups in natural product functions [5]. Modern synthetic chemistry has encountered difficulty in preparation of complex, high molecular weight natural products containing a great number of reactive groups and stereocenters for the generation of drug leads in the pharmaceutical industry. Incorporating the use of biocatalysts during natural product synthesis represents a promising strategy for the production of compounds that are desperately needed for pharmaceutical development [6, 7]. For example, all anti-cholesterol statins have a common 3,5-dihydroxyacid side chain, which contains two chiral hydroxyl groups. In the current manufacturing process of atorvastatin side chain, (*R*)-4-cyano-3-hydroxybutyrate, a key side chain precursor, was produced by a two-enzyme system [8]. In this process, ethyl 4-chloroacetoacetate was first reduced by a ketoreductase to form (*R*)-

4-chloro-3-hydroxybutyrate, followed by a halohydrin dehydrogenase- catalyzed cyanation to produce the desired product. Moreover, enzymes residing in living cells have been extensively used as biocatalysts in the food and beverage industry, and isolated enzymes also play critical roles in performing chemical transformations on organic compounds in many areas [9-12]. Often, these remarkable catalysts are able to perform a wide array of reactions on structurally diverse compounds. Furthermore, enzymes can also selectively catalyze reactions with chiral (enantio-) and positional (regio-) selectivities [12]. With these advantages, enzymes are applied in organic synthesis to avoid tedious protection and deprotection steps commonly required for enantio- and regioselective synthesis. The inherent selectivity of enzymes generates few by-products, making it an environmentally friendly alternative to chemical catalysts. Enzymes used in organic synthesis include acyl transferases (e. g., lipases, esterases, peptidases, amidases, and acylases), carbohydrate processing enzymes (e. g., glycosidases, glycosyl-transferases), hydrolytic enzymes (e. g., epoxidases, nitrilases, nitrile hydratases), reductases, oxidases and oxygenases, aldolases, and oxynitrilases [9].

Polyketides (PKs), nonribosomal peptides (NRPs), and PK/NRP or NRP/PK hybrids represent three large subclasses of highly diverse natural products with various bioactivities [13]. These natural products are produced by large megaenzymes, polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs). Type I PKSs consist of multiple modules, with each module minimally containing three core domains: acyltransferase (AT) domain, ketosynthase (KS) domain, and thiolation (T) domain [also called acyl carrier protein (ACP) domain]. Typically, one type I PKS module catalyzes a single elongation cycle for PK production (**Figure 1-1**). During elongation, the AT domain serves as the gatekeeper for specificity, responsible for selecting the appropriate CoA extender unit (e. g., malonyl-CoA, methylmalonyl-CoA) and transferring the extender unit to the sulfhydryl terminus of the phosphopantetheinyl arm on the T domain [14]. The KS domain catalyzes the decarboxylation of acyl-*S*-T to generate a carbanion that reacts with the PK intermediate linked to T domain generated in the previous elongation cycle. The resulting α-ketoacyl-*S*-T becomes the substrate for the next cycle of elongation catalyzed by the subsequent module. In addition to type I PKSs, there are two other PKS

classes, type II PKSs and type III PKSs [15, 16]. Unlike the type I class, type II PKSs consist of discrete enzymes that are organized as a multi-component system [16]. The type III PKSs are distinguished from the others by lack of an AT and T domain. Type III PKS systems typically use CoA substrates (i. e. malonyl-CoA), but there is precedent for their ability to accept acyl-*S*-T substrates [15, 17]. Similar to the type I PKSs, NRPSs are comprised of multifunctional enzymes that are arranged into modules. Each NRPS module contains three core domains: adenylation (A), condensation (C), and thiolation (T) [also called peptidyl carrier protein (PCP) domain] [18] (**Figure 1-1**). The A domain is responsible for selecting and activating the natural or modified amino acid monomer. The activated amino acid monomer is covalently attached via a thioester bond to the cysteamine group of a phosphopantetheinyl arm in the holo-T domain. The condensation (C) domain catalyzes formation of the peptide bond between the amino acid monomer and the peptidyl intermediate tethered to a T domain in an adjacent module. Similar to type I PKS modules, each NRPS module performs a single elongation step of the growing peptidyl chain. In both NRPSs and PKSs, there are several additional domains that contribute to

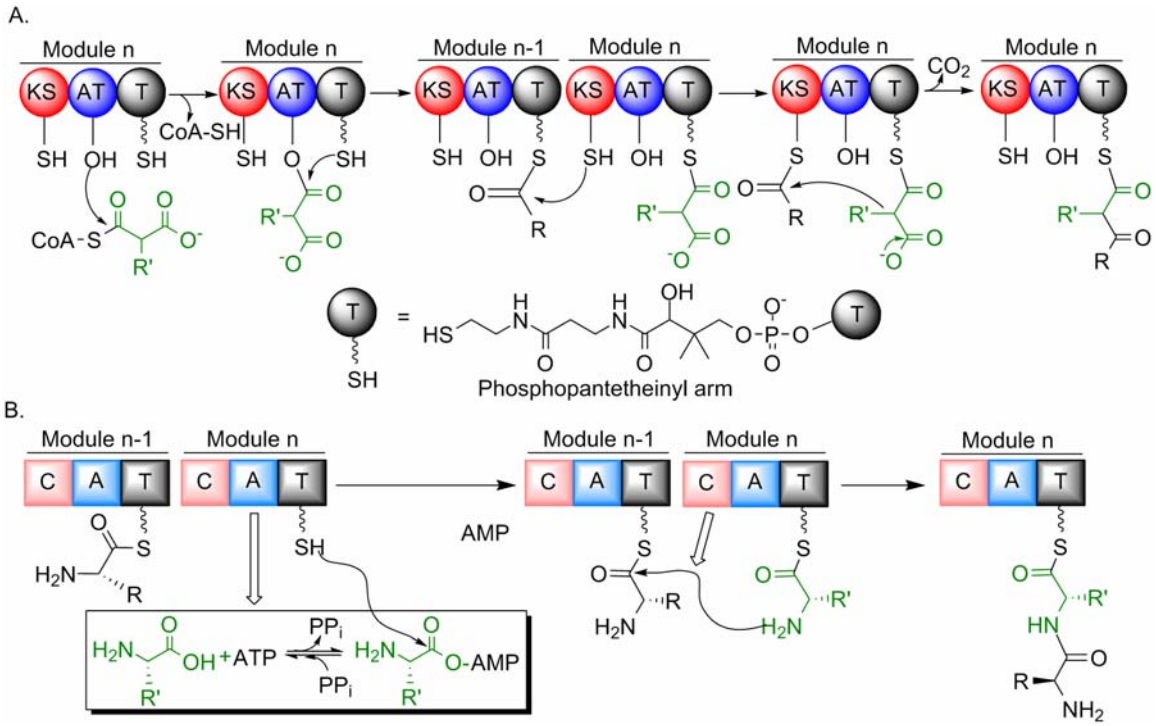


Figure 1-1. Schematic representations of one elongation cycle catalyzed by the minimal module of PKS (A) and NRPS (B). A common phosphopantetheinyl arm is found in the holo-T domain of both PKS and NRPS as shown in **A**.

natural product structural diversity. Ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER), and methyltransferase (MT) domains are commonly found in PKS modules while *N*-MT and epimerase (E) domains are generally embedded within NRPS modules. These additional domains contribute significantly to the diversity and bioactivity of PKs and NRPs. Thioesterase (TE) domains, typically found at the C-terminus of the final elongation module in both PKSs and NRPSs are responsible for terminating biosynthesis. In most cases, TE domains catalyze intra-molecular macrocyclization or hydrolysis of the thioester bond between the final T domain and the PK or NRP intermediate [19]. The structures of the nascent PK and NRP products are often further modified through oxidation, glycosylation, acylation, alkylation, and halogenation reactions catalyzed by tailoring enzymes in natural product biosynthetic pathways [20, 21].

Recent advances in recombinant DNA technology, genetics, and DNA synthetic methodology have further boosted the development of natural products for clinical practice through heterologous production and metabolic engineering [22-24]. For example, the antibiotic erythromycin has been heterologously produced in *E. coli*, presenting the first example of the production of complex PKs in heterologous microbes (**Figure 1-2**) [25]. Although *E. coli* produces propionyl-CoA, the starting unit for 6-deoxyerythronolide B (6DEB) biosynthesis, this substrate is not formed in high-cell density conditions, requiring metabolic engineering of host strain to fulfill specific requirements. Pfeifer *et al.* circumvented this problem by introducing one propionyl-CoA synthetase (PrpE) gene, propionyl-CoA carboxylase (PCC) genes, and one phosphopantetheinyl transferase (Sfp) gene to *E. coli* while disrupting its *prp* operon and one methylmalonyl-CoA decarboxylase (*ygfG*) gene [25]. These engineering efforts avoided squandering propionyl-CoA and (2*S*)-methylmalonyl-CoA, another building block for 6DEB biosynthesis, by unexpected metabolism in *E. coli*. Moreover, the engineered *E. coli* strain produced both CoAs after feeding propionic acid. After further optimization, the titer of 6dEB was improved to 1.1 g per liter of fermentation culture [26]. More importantly, replacing the loading module for 6DEB biosynthesis with that for rifamycin biosynthesis made *E. coli* to produce a 6DEB analog carrying an aromatic moiety, further suggesting feasibility of metabolic engineering in generation of novel natural product analogs for pre-clinical and

clinical studies [25]. Furthermore, one mature erythromycin analog, erythromycin C, was prepared at 0.4 mg/L of culture after co-expressing two biosynthetic gene clusters responsible for the production of deoxysugars, TDP-L-mycarose and TDP-D-desosamine in the above engineered *E. coli* strain [27]. These results indicated that heterologous microbes with well-developed genetic engineering tools have the ability to produce

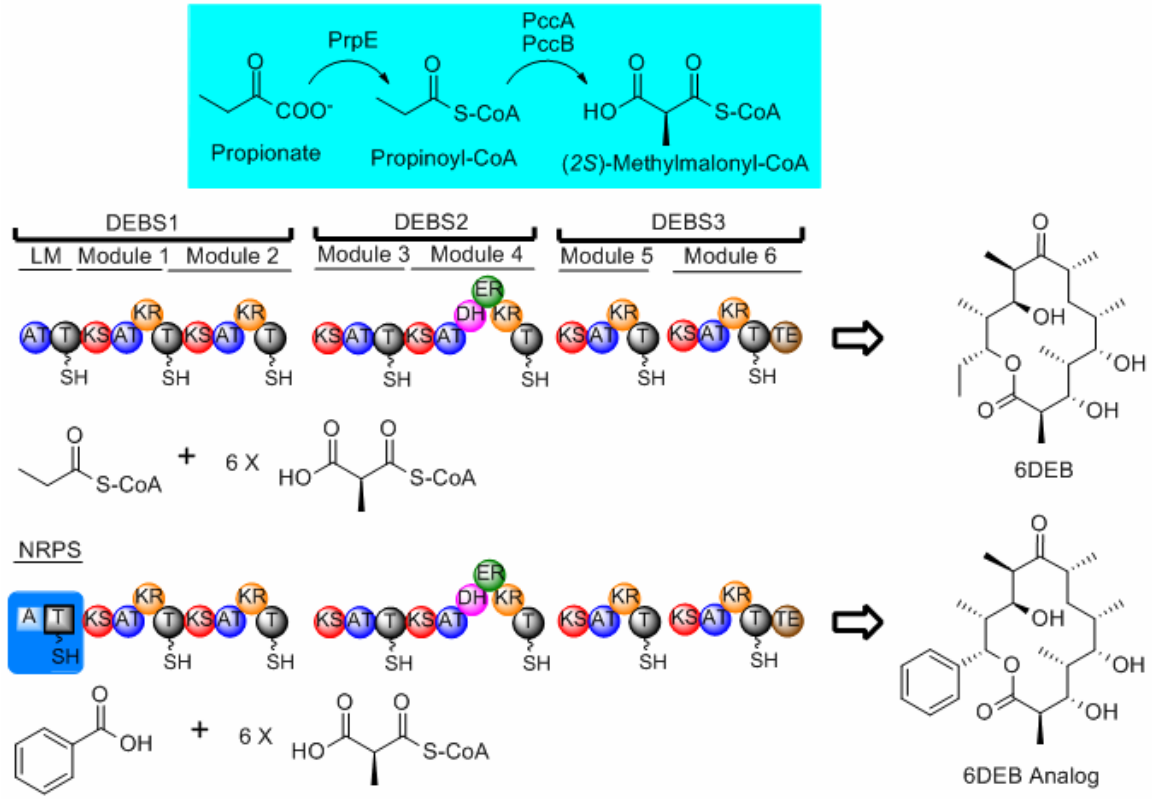


Figure 1-2. Metabolic engineering and heterologous production of 6DEB in *E. coli*. The efficiency and feasibility of metabolic engineering and heterologous production of pharmaceutically valuable natural products and analogs thereof is exemplified by the production of 6DEB and its analog in *E. coli*. 6DEB analog was produced after replacing DEBS1 loading module with one NRPS module with specificity to benzoic acid.

pharmaceutically valuable both natural and novel natural product analogs through DNA recombination and metabolic engineering.

Complex secondary metabolites display diverse biological activities, and along with their derivatives, have provided over two-thirds of new pharmaceutical agents introduced during the past two decades. However, drug-resistant bacteria like methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus*

faecalis continue to emerge as a threat to human health [28-30]. Along with medical needs to treat a spectrum of diseases, an increasing number of natural products have been isolated and screened for potential bioactive secondary metabolites [31]. Adversely, limitations in isolation and rapid structural determination continue to be inherent hurdles for natural products as leads in drug discovery and design. Chemoenzymatic approaches along with natural product heterologous production and metabolic engineering have proven to be effective and feasible for producing natural product analogs regardless of the above challenges. On the other hand, identification and characterization of natural product biosynthetic pathways are prerequisite and indispensable to these strategies. My Ph.D. dissertation research under the mentorship of Professor David H. Sherman was designed to broadly investigate selected natural product biosynthetic pathways and to extensively employ biocatalysts discovered from the characterized pathways to identify compounds with enhanced anticancer activity. The specific aims of my project included (1) development of chemoenzymatic routes to cryptophycin analogues, (2) isolation of key biosynthetic intermediates to dissect fungal alkaloid biosynthetic pathways; (3) genome mining and detailed characterization of notoamide biosynthetic gene cluster; (4) cloning and characterization of a novel macrocyclic trichothecene gene cluster. The details of each specific aim will be thoroughly covered in following chapters of this thesis.

Reference:

1. Keller, N. P.; Turner, G.; Bennett, J. W., *Nat Rev Microbiol* **2005,** 3, (12), 937-47.
2. Nicolaou, K. C.; Vourloumis, D.; Winssinger, N.; Baran, P. S., *Angew Chem Int Ed Engl* **2000,** 39, (1), 44-122.
3. Wohlleben, W.; Pelzer, S., *Chem Biol* **2002,** 9, (11), 1163-4.
4. Newman, D. J.; Cragg, G. M.; Snader, K. M., *J Nat Prod* **2003,** 66, (7), 1022-37.
5. Dobson, C. M., *Nature* **2004,** 432, (7019), 824-8.
6. Hojati, Z.; Milne, C.; Harvey, B.; Gordon, L.; Borg, M.; Flett, F.; Wilkinson, B.; Sidebottom, P. J.; Rudd, B. A.; Hayes, M. A.; Smith, C. P.; Micklefield, J., *Chem Biol* **2002,** 9, (11), 1175-87.
7. Kopp, F.; Marahiel, M. A., *Curr Opin Biotechnol* **2007,** 18, (6), 513-20.
8. Muller, M., *Angew Chem Int Ed Engl* **2005,** 44, (3), 362-5.
9. Davis, B. G.; Boyer, V., *Nat Prod Rep* **2001,** 18, (6), 618-40.
10. Gupta, M. N., *Eur J Biochem* **1992,** 203, (1-2), 25-32.
11. Roland, W., *Journal of Chemical Technology & Biotechnology* **2007,** 82, (12), 1055-1062.
12. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B., *Nature* **2001,** 409, (6817), 258-68.
13. Fischbach, M. A.; Walsh, C. T., *Chem Rev* **2006,** 106, (8), 3468-96.
14. Walsh, C. T., *Science* **2004,** 303, (5665), 1805-10.
15. Austin, M. B.; Noel, J. P., *Nat Prod Rep* **2003,** 20, (1), 79-110.
16. Hopwood, D. A., *Chem Rev* **1997,** 97, (7), 2465-2498.
17. Gruschow, S.; Buchholz, T. J.; Seufert, W.; Dordick, J. S.; Sherman, D. H., *Chembiochem* **2007,** 8, (8), 863-8.
18. Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D., *Chem Rev* **1997,** 97, (7), 2651-2674.
19. Kopp, F.; Marahiel, M. A., *Nat Prod Rep* **2007,** 24, (4), 735-49.
20. Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M., *Curr Opin Chem Biol* **2001,** 5, (5), 525-34.
21. Murphy, C. D., *Nat Prod Rep* **2006,** 23, (2), 147-52.
22. Chemler, J. A.; Koffas, M. A., *Curr Opin Biotechnol* **2008,** 19, (6), 597-605.
23. Wenzel, S. C.; Muller, R., *Curr Opin Biotechnol* **2005,** 16, (6), 594-606.
24. Zhang, H.; Wang, Y.; Pfeifer, B. A., *Mol Pharm* **2008,** 5, (2), 212-25.
25. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., *Science* **2001,** 291, (5509), 1790-2.
26. Lau, J.; Tran, C.; Licari, P.; Galazzo, J., *J Biotechnol* **2004,** 110, (1), 95-103.
27. Peiru, S.; Menzella, H. G.; Rodriguez, E.; Carney, J.; Gramajo, H., *Appl Environ Microbiol* **2005,** 71, (5), 2539-47.
28. Gin, A. S.; Zhanel, G. G., *Ann Pharmacother* **1996,** 30, (6), 615-24.
29. Walsh, C., *Nat Rev Microbiol* **2003,** 1, (1), 65-70.
30. Spurgeon, D., *Bmj* **2007,** 335, (7627), 961.
31. Pelaez, F., Biological activities of fungal metabolites. In *Handbook of Industrial Mycology*, An, Z., Ed. Marcel Dekker: New York, 2005; pp 49-922.

Chapter 2

Chemoenzymatic Synthesis of Anticancer Agent Cryptophycin Analogs with Biocatalysts from *Nostoc* sp.

2.1. Summary

This chapter contains three successive stories toward developing chemoenzymatic routes to synthesize cryptophycin analogs and biochemically characterizing the substrate cooperativities in cryptophycin P450 epoxidase. The first study focused on the mechanistic understanding of the incorporation of L-leucic acid into the cryptophycin structure as its Unit D. Non-amino acid substrate incorporation into NRPs has been reported previously but the complete characterization of this process was lacking and only preliminarily addressed in this study. The substrate selectivity of the A domain of the CrpD second module (CrpD-M2) was predicted to be specific toward non-amino acid chemicals in bioinformatics analysis, and this was subsequently confirmed using a radioactive ATP-PPi assay. Both 2-ketoisocaproic acid (2KIC) and 2-hydroxyisocaproic acid (2HIC) were activated by the CrpD-M2 A domain and loaded onto its T domain. Substrate loading was further confirmed with high-resolution fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). After 2KIC loading, the unique α-ketoreductase (KR) embedded into CrpD-M2 reduced 2KIC into 2HIC with the consumption of either NADPH or NADH. Working in progress is to characterize the formation of a *seco*-cryptophycin biosynthetic intermediate with chemically synthesized UnitABC-NAC ester using the CrpD-M2 C domain. The full understanding of L-leucic acid incorporation will establish one chemoenzymetic route relying on one multifunctional enzyme, which has the potential to generate cryptophycin analogs carrying unnatural Unit D moieties. The second study addressed the most challenging issue in chemical synthesis of cryptophycin analogs. The most potent natural cryptophycin analogues retain a β-epoxide at the C2'-C3' position of the molecule.

Chemical installation of this functional group faces problems of selectivity, specificity, and undesirable byproducts. A P450 epoxidase encoded by c*rpE* was shown to install this key functional group in a regio- and stereospecific manner. A detailed characterization of the CrpE epoxidase using an engineered maltose binding protein (MBP)-CrpE fusion investigated the substrate tolerance of the CrpE polypeptide, and a series of structurally related cryptophycin analogues were generated by chemoenzymatic synthesis. The k_{cat}/K_m values of the enzyme for these substrates were determined to provide further insights into how the P450 epoxidase catalytic efficiency was affected by substrate structural variation. Finally, binding analysis revealed cooperativity of MBP-CrpE toward natural and unnatural desepoxy cryptophycin substrates. The last story in this chapter continued my efforts to understand the unusual cooperativity feature of this bacterial P450. I further characterized its homotropic cooperativity from different aspects using Job's titration to calculate ligand:enzyme molar ratio, substrate titration under high concentration salt to exclude enzyme oligomerization contributing to observed cooperativity, and linear fatty acids titration to estimate enzyme active pocket size. Furthermore, steroid testosterone was found to be a both homotropic and heterotropic ligand of CrpE, suggesting CrpE as a good model system for human P450 cooperativity studies. Most importantly, testosterone was an enhancer for CrpE catalytic reaction, making it as a good tool for generating anticancer cryptophycin analogs in a chemoenzymatic approach.

2.2. Introduction

Cryptophycins are isolated from the cyanobacterial symbiont *Nostoc* sp. ATCC 53789, and *Nostoc* sp. GSV 224 [1, 2]. They are potent anticancer agents due to their ability to stimulate cellular microtubule instability, inhibit microtubule assembly, and induce tubulin self-association, resulting in a G2/M phase transition block in the cell cycle [3, 4]. The cryptophycin mode of action and cellular target resembles both the v*inca* alkaloids and taxol [5, 6]. Significantly, the cryptophycins are not active substrates of P-glycoprotein and/or multiple drug resistance-associated protein, making them viable chemotherapeutic alternatives to the v*inca* alkaloids and taxol for treatment of *vinca* alkaloid- and taxol-resistant cancers [7, 8]. The clinical potential, validated mode of action, and synthetically

challenging structure of cryptophycins, as well as the lack of large-scale fermentation methods for their production, have stimulated the development of synthetic methods to provide suitable amounts of material and new analogs with improved physiochemical properties for clinical studies.

The cryptophycin gene cluster from *Nostoc* sp. ATCC 53789 and *Nostoc* sp. GSV 224 was recently identified in a 40-kb of region DNA, offering potential chemoenzymatic tools for cryptophycin production (**Figure 2-1**) [9]. The gene cluster is comprised of two type I PKS genes, *crpA* and *crpB*, two NRPS genes, *crpC* and *crpD*, and four tailoring enzyme genes including a P450 epoxidase gene (*crpE*), a putative 2-ketoglutarate iron-dependent hydroxylase gene (*crpF*), a decarboxylase gene (*crpG*), and a flavin-dependent halogenase gene (*crpH*). CrpA and CrpB each contain two elongation modules that are hypothesized to generate the δ-hydroxy phenyloctenoic acid polyketide moiety (Unit A)

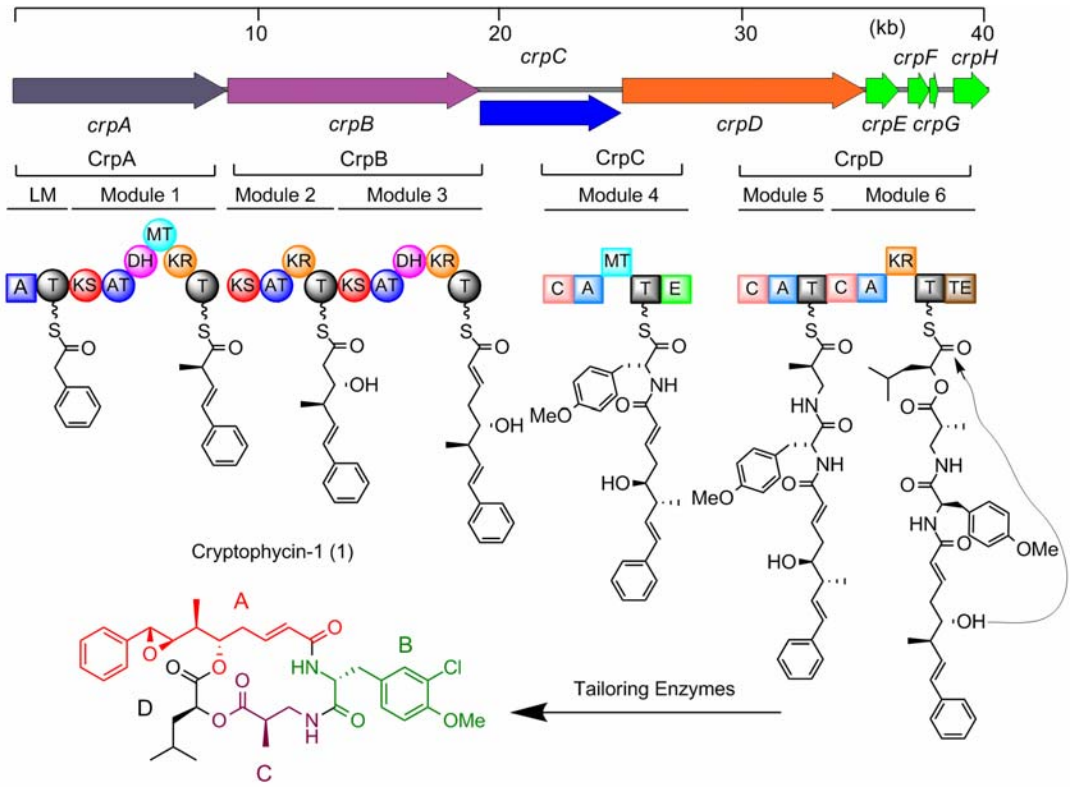


Figure 2-1. **Schematic organization of cryptophycin gene cluster (*crp*) and its biosynthetic pathway with CrpA, CrpB, CrpC and CrpD.** The structure of predominant natural cryptophycin, Cr-1, is shown and dissected into four Units (A-D).

in cryptophycin 1 (Cr-1) (**1**) from one molecule of phenylacetate or its derivative as a starter unit, and three molecules of malonyl-CoA as extender units. CrpC is a monomodular NRPS containing a single elongation module that activates and epimerizes L-tyrosine to D-tyrosine and subsequently condenses the activated D-tyrosine (Unit B) with the PK acyl intermediate tethered to the CrpB T domain. The activated amino acid may also be methylated by a single MT domain in CrpC. CrpD is a bimodular NRPS enzyme. Its first module activates methyl-β-alanine (Unit C) that is converted from L-aspartic acid by CrpG, a β-methylaspartate-α-decarboxylase [10]. The second module of CrpD assembles an activated 2KIC (Unit D) to produce the final PK/NRP hybrid intermediate. The immediate source of 2KIC remains unknown, but may result from transformation of L-leucine. Interestingly, this NRPS module contains one KR domain, which may convert 2KIC into 2HIC. The final domain of CrpD is a thioesterase that catalyzes release of the linear intermediate from the final T domain of CrpD and generates a cyclic 16-membered ring natural product. This cyclic depsipeptide is further structurally diversified by tailoring enzymes whose genes are encoded within the cryptophycin biosynthetic gene cluster. More than 25 naturally occurring cryptophycin analogues are produced by this biosynthetic machinery in *Nostoc* sp. ATCC 53789 [9], with the major isolate being Cr-1 (**Figure 2-1**).

The large number of cryptophycins produced by *Nostoc* sp. ATCC 53789 is indicative of the flexibility of the cryptophycin biosynthetic system, including PKSs, NRPSs, and tailoring enzymes. The versatility of this assembly line was first assessed using a precursor directed biosynthesis approach [9]. In total, 22 unnatural amino acids and halogen sources were introduced to the *Nostoc* culture and 44 unnatural cryptophycin analogs including cryptophycin 52, the synthetic lead molecule that later advanced to clinical trials, were isolated and identified. This result highlighted the exciting possibility that the cryptophycin biosynthetic machinery could be employed to generate and identify new bioactive cryptophycin analogs as anti-cancer leads [9].

The extended growth period of *Nostoc* sp. ATCC 53789 and GSV224 limits their ability to produce cryptophycin analogs in large-scale fermentation. Alternatively, total

synthetic strategies have been developed to provide adequate supplies for clinical evaluation [11]. Several effective synthetic approaches have been employed for the generation of cryptophycins, including sufficient quantities of cryptophycin 52 required for clinical trials. Besides, a number of cryptophycin analogs with unnatural Unit A, B, and C moieties have been produced by chemical methods and their activities were examined to identify novel chemical with improved bioactivities and reduced toxicity. However, proper macrocyclization and efficient epoxidation are still the two most challenging issues for cryptophycin chemical synthesis, and chemical reagents to address these issues add considerable time and expense to achieve the final desired products. Cryptophycin TE has the potential to solve the problem of proper macrocyclization in both an environmentally friendly and economical manner. In a recently described approach, the Crp TE was excised and heterologously overexpressed as a recombinant enzyme [12]. Four NAC activated *seco*-cryptophycins were chemically synthesized and utilized to interrogate the *in vitro* activity and substrate specificity of Crp TE (**Figure 2-2**). Naturally occurred cryptophycin 4 (**2**), cryptophycin 24 (**3**), and an unnatural C6 *gem*-dimethyl analog, cryptophycin-B (**4**), were produced as the major products in the chemoenzymatic reactions, suggesting the termination enzyme is robust in its ability to produce both natural and unnatural products and tolerant to variants at the cryptophycin β-alanine moiety. TE specificity was further investigated using a substrate lacking the phenyl group in its PK A-subunit moiety. Interestingly, Crp TE was also capable of generating the cyclic product (**5**) but with significantly increased levels of hydrolysis (cyclization:hydrolysis = 1: 8.3) [12].

Solid-phase peptide synthesis is widely used in preparation of NRP substrates for NRPS TE domain studies. Recently, this approach was applied to PK/NRP natural product studies [13]. Several *seco*-cryptophycin analogs were synthesized on safety-catch PEGA resin and subjected to chemoenzymatic reactions with Crp TE (**Figure 2-2**). Crp TE also recognized these substrates and catalyzed macrocyclization to produce cryptophycin 29 (**6**) and the unnatural des-epoxy cryptophycin 24 (**7**) [13]. Since the ester bond between methyl-β-alanine and α-ketoisocaproate moieties in cryptophycins is labile to hydrolysis, Crp TE was probed for its tolerance regarding replacement of this ester

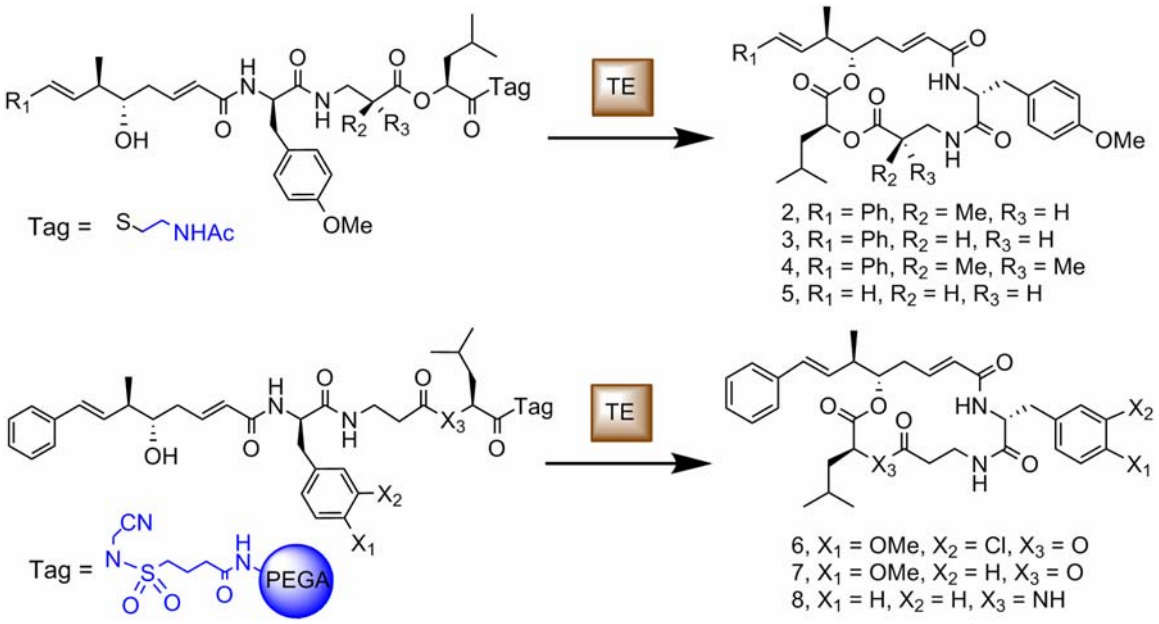


Figure 2-2. Chemoenzymatic synthesis of cryptophycin analogs using the excised Crp TE domain. Recombinant TE from this prototypical PKS/NRPS system tolerates structural modifications on its native *seco*-cryptophycin substrates. Two different tags are used to mimic the T domain to which the PK/NRP intermediate is tethered *in vivo*.

with an amide linkage. Generation of a novel cyclic compound bearing the amide (**8**) in the Crp TE reaction demonstrates the unusual versatility of this enzyme. Thus, Crp TE is an attractive tool to overcome the macrocyclization challenge in chemical synthesis, and to generate an array of new analogs in sufficient quantity for bioactivity analysis.

The most significant challenge in cryptophycin chemical preparation is the late-stage regio- and stereospecific installation of the epoxide moiety due to the labile and highly reactive nature of this functionality [11]. Initial efforts with mCPBA or dimethyl dioxirane (DMD) resulted in conversion of 25% of the starting material into the unfavorable α-epoxy-cryptophycin. The potency of this isomer is at least 100 times less than the β-epoxy-product [14, 15]. Although several other stereospecific epoxidation methods have been developed, there remains significant need for improvement in yield and selectivity [16, 17]. In this chapter, I will describe my studies to develop the cytochrome P450 CrpE as an effective biocatalyst to fulfill this demand. In addition, efforts to investigate less explored Unit D biosynthesis will also expand the current understanding

of NRP biosynthesis and provide one more chemoenzymatic route to generate numerous novel cryptophycin analogs.

2.3. Results and Discussion

2.3.1. Chemoenzymatic Synthesis of Cryptophycin with a Single Multifunctional Cryptophycin Biosynthetic Enzyme

To activate and install CrpD-M2 substrates into the cryptophycin structure, the DNA fragment of this megaenzyme consisted of condensation (C), adenylation (A), ketoreductase (KR), and thiolation (T) domains was amplified by PCR, cloned into the *Bam*HI and *Xho*I sites of pET28a, and heterologously overexpressed in *E. coli* BAP1 strain with the chromosomally integrated *sfp* phosphopantetheinyl transferase gene [18]. The N-terminally His-tagged protein was purified with Ni-NTA resin to reach about 80 % purity (**Figure S2-1A**). Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) was used to verify that the protein was the full length, and phosphopantetheinated on the T domain active site (**Figure S2-1B**).

Subsequently, the well-proven radioactive ATP-PPi exchange assay with ten acids as enzyme substrates was used to determine the substrate specificity of the CrpD-M2 A domain by measuring reversible acyl-AMP formation (**Figure S2-1C**). CrpD-M2 was found to preferentially activate 2KIC over L-leucine, which was consistent with results of CrpD-M2 substrate bioinformatic prediction and feeding experiment (**Figure 2-3**, **Table S2-1**) [9]. Moreover, CrpD-M2 showed the similar selection to natural substrate 2-oxovalerate while its specificity to two other natural Unit D fragments, 3-methyl-2-oxovalerate and 3-methly-2-oxobutyrate, was decreased about 50 % and 90 %, respectively, compared to 2KIC. This result along with the weak activation of unnatural substrates 2-oxobutyrate and phenylpyruvate strongly suggested the importance of side chains for substrate recognition by CrpD-M2. More importantly, the unnatural substrate, 2-keto-γ-(methylthio) butyrate (AKGB), was effectively activated, suggesting that CrpD-M2 might have the potential to produce cryptophycin analogs with enhanced hydrophilicity. Interestingly, L-2HIC was the best substrate in the assay. Thus, CrpD-M2 A domain is the first example with similar selectivity toward both 2-keto and 2-hydroxy

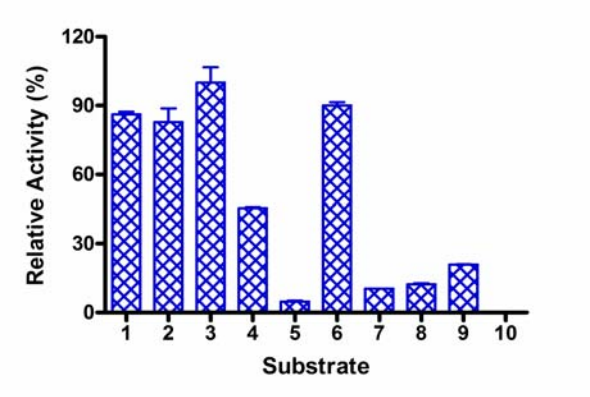


Figure 2-3. CrpD-M2 A domain substrate selectivity. Totally, ten chemicals were used in ATP-PPi exchanging assay. They were (1) 2KIC, (2) AKGB, (3) L-2HIC, (4) 3-methyl-2-oxovalerate, (5) L-leucine, (6) 2-oxovalerate, (7) 3-methly-2-oxobutyrate, (8) 2-oxobutyrate, (9) 4-methyl-valerate, and (10) phenylpyruvate. The structures of these substrates were shown in **Figure S2-1C**.

acids, although 2-keto acid activating A domains have been described in bacillaene, barbamide, cereulide, and valinomycin NRPS [19-21]. The weak activation of 4-methyl-valerate by CrpD-M2 firmly suggested the importance of functional group at its C-2 for enzyme recognition.

Both 2KIC and L-2HIC were activated by CrpD-M2 in an *in vitro* assay, however, only 2KIC is a direct precursor of leucine in amino acid biosynthetic pathway and may serve as the physiological substrate of this NRPS module. Following 2KIC activation and loading onto CrpD-M2 T domain, 2HIC may be produced in one α-keto reduction reaction catalyzed by one unusual KR domain embedded in CrpD-M2. Bioinformatic analysis indicated that this KR was grouped with other NRPS KR domains while was separated from any type of PKS β-KR domain (**Figure S2-2**).

FTICR-MS was utilized to monitor substrate loading directly on the T-domain [22-24]. Interestingly, both L-2HIC and D-2HIC were loaded on the T domain, indicating an unusual degree of substrate flexibility (**Table 2-1**). However, as only L-2HIC cryptophycin analogs have been isolated and characterized, one of the following hypotheses is likely true: 2KIC is stereoselectivly reduced to L-2HIC, only L-2HIC is productively processed to *seco*-cryptophycin, and/or D-2HIC is not physiologically

available. A further degree of substrate flexibility, with possible applications for combinatorial biosynthesis, was displayed by the loading of 2KIC and the unnatural substrate AKGB, in agreement with the ATP-PPi results (**Table 2-1**).

Table 2-1. FTICR-MS analysis of substrate loading to CrpD-M2 and ketoreduction by its domain

#	Reaction	MW_{obs}	$MW_{cal.}$	Max Intensity	Retention Time
1	L-2HIC	4,116.21	4,116.12	2.1E+05	82
2	D-2HIC	4,116.18	4,116.12	4.3E+05	82
3	AKGB	4,151.28	4,151.16	2.8E+04	79
4	KIC	4,114.15	4,114.12	8.8E+04	81
5	KIC+NADPH	4,116.14	4,116.12	1.5E+05	82
6	KIC+NADH	4,116.20	4,116.12	7.6E+05	82

A mass shift was evident after loading of 2KIC and addition of either NADH or NADPH (**Table 2-1**). The mass shift of 2 Da, consistent with 2KIC to 2HIC, indicated the occurrence of α-ketoreduction reaction. Both NADH and NADPH appeared to operate with qualitatively similar efficiency as hydride donors.

Next, the ability of CrpD-M2 to form the full-length cryptophycin was investigated with newly synthesized UnitABC-NAC ester (**Figure S2-3**). The synthetic scheme followed one previously established route and the final product was confirmed with NMR and high resolution mass spectrometry. The UnitABC-NAC ester was then mixed with CrpD-M2 and L-2HIC. Despite numerous attempts, T-domain tethered *seco*-cryptophycin could not be identified by LC FTICR-MS. Possible reasons include: product instability (although other advanced NRPS intermediates can be observed[25-27] using similar methodologies), low steady-state concentrations of enzyme bound *seco*-cryptophycin, or the cleavage of C domain from CrpD-M2 during storage. Now, the CrpD-M2 C domain is being overexpressed as a single domain to investigate these possibilities.

The presumed *seco*-cryptophycin intermediate may be macrocylized from CrpD-M2 T domain by addition of the CrpD thioesterase domain *in trans*. Furthermore, this

putative substrate will be analyzed with HPLC, FTICR-MS and MS/MS. Finally, the feasibility of the above cryptophycin chemoenzymatic route will be tested with one unnatural substrate. In consistence with ATP-PPi exchange assay, AKGB was effectively loaded onto CrpD-M2 T domain, which was then reduced to 2-hydroxy species with the consumption of NAD(P)H (**Table 2-1**). Feeding of the UnitABC-NAC ester may initiate the formation of unnatural *seco*-cryptophycin. I expect that CrpD-M2 has the potential to generate novel cryptophycin analogs with altered physicochemical properties that may be beneficial to clinical application.

Previously, understanding of non-amino acid moieties incorporation into many bioactive nonribosomal peptides came from two separate systems. The biochemical studies of NRPS systems in bacillaene, cereulide, and valinomycin biosyntheses are the only examples so far to investigate 2-keto acid activation, loading, and reduction while the subsequent formation of ester bonds in natural product biosynthesis was indirectly deduced from studies of Fum14p consisted of T and C domains for fumonisin biosynthesis and of SgcC5, a standalone C domain for C-1027 biosynthesis [19, 21, 28, 29]. Herein, one complete scheme will be presented by the characterization of CrpD-M2 in cryptophycin biosynthesis. CrpD-M2 may also have the potential to generate novel cryptophycin analogs with improved therapeutic index and reduced toxicity, as exemplification of the putative production of both natural and unnatural cryptophycin analogs.

2.3.2. Analysis of the Cryptophycin P450 Epoxidase Reveals Substrate Tolerance and Cooperativity

Overexpression of *crpE* in *E. coli*. The 1.4-kb c*rpE* gene was identified at the 3'-end of cryptophycin biosynthetic gene cluster[9] and cloned into the pSJ8 expression vector to produce CrpE tagged with maltose-binding protein (MBP) at its amino-terminus. The MBP-CrpE fusion protein, when co-expressed with select chaperones, resulted in significant levels of soluble protein that displayed the expected red-brown color.[30] A protein of molecular weight ~57 kD (presumed to be GroEL) was co-purified with MBP-

CrpE (**Figure S2-4**) and all attempts to separate this contaminant from the fusion protein by addition of ATP and urea[31, 32] were unsuccessful (**Figure S2-4**). Thus, the studies described in this report were performed on the MBP-CrpE fusion protein in the presence of this bound polypeptide.

Reconstruction of P450 MBP-CrpE Reaction System. CrpE was hypothesized to be an epoxidase based on bioinformatic analysis. Comparison of its sequence to four bacterial P450 epoxidases from secondary metabolic pathways (EpoK, PimD, MycG, and OleP) revealed that all share two conserved motifs (**Figure S2-5**).[33-38] A Ser or Thr residue in motif-I is likely responsible for O_2 binding and cleavage, and an invariant Cys residue in motif-II presumably serves as the fifth ligand for the heme iron. The presence of these conserved sequence motifs suggested that CrpE is a P450 enzyme. Subsequently, UV-visible CO difference spectrum analysis of MBP-CrpE gave direct evidence for this conclusion (**Figure S2-4**). Although TEV protease digestion provided native CrpE, this protein lacked suitable stability (estimated half time of 2 hours) for biochemical studies, and therefore the studies reported in this paper utilized the MBP fusion protein.

The native *Nostoc* sp. electron donor partners for CrpE are unknown. In our studies, spinach ferredoxin (Fer) and ferredoxin $NADP^+$ reductase (FNR) were used to provide electrons for MBP-CrpE catalytic turnover (**Figure 2-4**). An NADPH regenerating system (e.g. glucose-6-phosphate and glucose-6-phosphate dehydrogenase[33]) had only a minor effect on MBP-CrpE activity. Thus, the reconstituted MBP-CrpE reaction included Fer, FNR, MBP-CrpE, NADPH, and O_2. Maximal enzyme activity occurred in 50 mM sodium phosphate buffer, pH 6.4 at 25 °C.

Analysis of MBP-CrpE Substrate Tolerance. Compounds used to assess the substrate specificity of MBP-CrpE (Cr-438, Cr-424, Cr-442, Cr-430, and Cr-537, see **Figure 2-5**) were synthesized chemoenzymatically (**Figure S2-6**) from the corresponding *N*-acetylcysteamine (SNAC) *seco*-cryptophycin. The structure of each compound was confirmed by ^{1}H NMR, ^{13}C NMR and MS analysis. In the final step, CrpD TE was used to catalyze macrocyclization of *seco*-cryptophycins[12] to provide Cr-3, Cr-4, Cr-17, Cr-43,

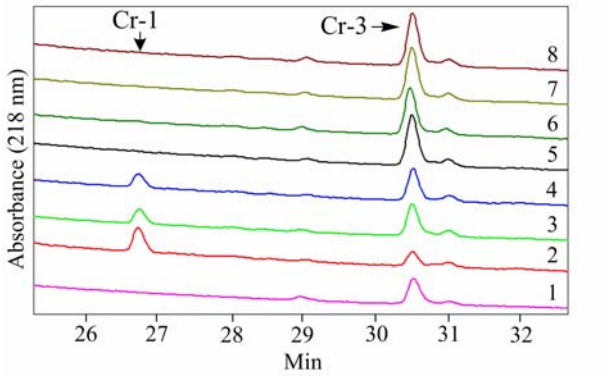
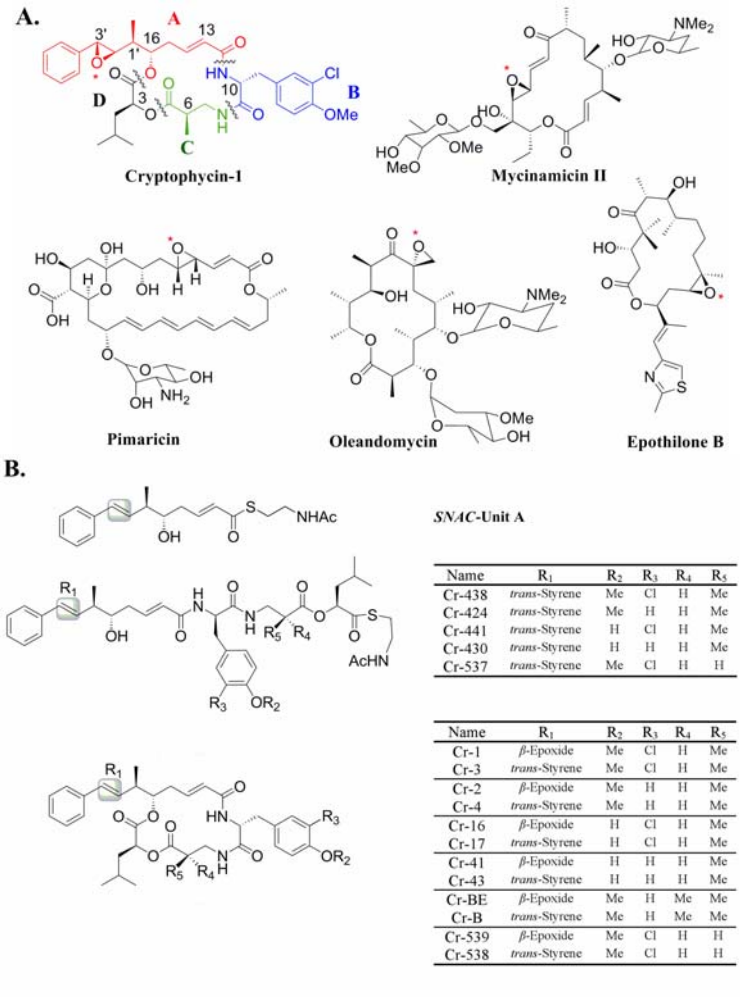


Figure 2-4. Investigation of MBP-CrpE reaction system components. The standard enzyme reaction with Cr-3 as substrate contained MBP-CrpE, spinach ferredoxin (Fer), spinach ferredoxin $NADP^+$ reductase (FNR), glucose-6-phosphate, and glucose-6-phosphate dehydrogenase (G6PDH) in 100 µl of storage buffer, and was analyzed by HPLC-UV. Lane 1, standard Cr-3; lane 2, standard enzyme reaction; lane 3, standard reaction with CrpE replacing MBP-CrpE; lane 4, standard reaction omitting G6PDH; lane 5, standard reaction with *E. coli* flavodoxin (Fld) and NADPH-flavodoxin reductase (Fpr) instead of Fer and FNR; lane 6, standard reaction with rat $NADP^+$ P450 reductase instead of Fer and FNR; lane 7, standard reaction omitting Fer; lane 8 the standard reaction omitting FNR. Only Fer and FNR were accepted by MBP-CrpE to produce Cr-1 from Cr-3. The presence of an NADPH regenerating system was not necessary for P450 activity. CrpE was capable of producing Cr-1 from Cr-3, although a decrease in overall production was observed. Thus, subsequent studies utilized MBP-CrpE as the enzyme source.

and Cr-538, respectively. Two additional members in the library included *SNAC*-Unit A as a linear CrpE substrate, and Cr-B as a cyclic substrate (**Figure 2-5**).

MBP-CrpE substrate selectivity was investigated with the above twelve-member library that was designed to probe the affect of structural variation in Unit B and Unit C on P450 catalysis and binding. None of the linear compounds were converted to product, while all cyclic cryptophycin analogs were epoxidized by MBP-CrpE (**Figure 2-6**). Products of Cr-1, Cr-2, and Cr-16 in the reactions gave m/z values of 655.3, 621.4, and 641.3, respectively, consistent with their calculated molecular weights (**Figure 2-6**). Cr-41, (the product of MBP-CrpE conversion of Cr-43) represents a new cryptophycin analog, not observed previously in nature. MBP-CrpE was also able to utilize unnatural cyclic cryptophycins as substrates to provide Cr-BE and Cr-539, respectively, as demonstrated by MS analysis (**Figure 2-6**).



Name	R_1	R_2	R_3	R_4	R_5
Cr-438	*trans*-Styrene	Me	Cl	H	Me
Cr-424	*trans*-Styrene	Me	H	H	Me
Cr-441	*trans*-Styrene	H	Cl	H	Me
Cr-430	*trans*-Styrene	H	H	H	Me
Cr-537	*trans*-Styrene	Me	Cl	H	H

Name	R_1	R_2	R_3	R_4	R_5
Cr-1	*β*-Epoxide	Me	Cl	H	Me
Cr-3	*trans*-Styrene	Me	Cl	H	Me
Cr-2	*β*-Epoxide	Me	H	H	Me
Cr-4	*trans*-Styrene	Me	H	H	Me
Cr-16	*β*-Epoxide	H	Cl	H	Me
Cr-17	*trans*-Styrene	H	Cl	H	Me
Cr-41	*β*-Epoxide	H	H	H	Me
Cr-43	*trans*-Styrene	H	H	H	Me
Cr-BE	*β*-Epoxide	Me	H	Me	Me
Cr-B	*trans*-Styrene	Me	H	Me	Me
Cr-539	*β*-Epoxide	Me	Cl	H	H
Cr-538	*trans*-Styrene	Me	Cl	H	H

Figure 2-5. Chemical structure of several epoxy natural products and MBP-CrpE substrates. A: cryptophycin-1 (Cr-1) and other examples of natural products containing an epoxide moiety installed by P450s. B: six *seco*- and six cyclic cryptophycin substrates used in MBP-CrpE studies. The six observed reaction products are also shown.

The stereochemistry and position of the β-epoxide in Cr-2 produced from Cr-4 has been confirmed previously by ^{1}H NMR analysis (**Table S2-2**).[9] Additionally, the two unnatural substrates, Cr-B and Cr-538, were treated with mCPBA, a reagent known to provide non-selective epoxidization,[39] in order to generate both the α- and β-epoxide stereoisomers, for comparison with the CrpE catalyzed reaction product (**Figure S2-7**). The β-epoxide products from mCPBA reactions had the same retention times as those from enzyme reactions in HPLC analysis, and thus confirming that in each case, CrpE

generated a single epoxidized cryptophycin product contained the desired β-epoxide configuration.

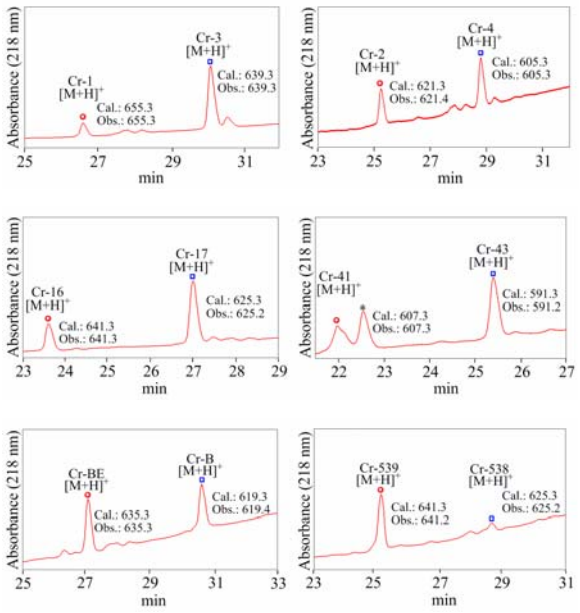


Figure 2-6. HPLC-UV and MS analyses of MBP-CrpE reactions with Cr-3, Cr-4, Cr-17, Cr-43, Cr-B, and Cr-538 as substrates. All substrates and products yielded the expected $[M+H]^+$ values. The Blue squares and the red circles represent substrates and products, respectively. A single contaminant (labeled with dark star) was found in all reactions and is attributed to DMSO used to dissolve all substrates.

Kinetic Analysis of MBP-CrpE with Different Substrates. Further studies of MBP-CrpE were conducted to establish the kinetic parameters for epoxidation. In these experiments, all plots of v (pmol·min^{-1}) *vs* [S] (μM) were linear with 100 μM as the highest substrate concentration. Thus, the expected K_m values are much larger than final cyclic substrate concentrations ([S]) in the assays. In this study, the Michaelis-Menten equation was converted into $v = k_{cat}\cdot[E]\cdot[S]/K_m$ giving slopes in the plots equal to $k_{cat}/K_m\cdot[E]$. The data (**Table 2-2**) show that MBP-CrpE most efficiently epoxidized Cr-538, whose k_{cat}/K_m value is 0.272 μM^{-1}·min^{-1}, while Cr-4 had the lowest k_{cat}/K_m value at 0.029 μM^{-1}·min^{-1}. Among the group of natural substrates, Cr-3 was most efficiently epoxidized by CrpE, consistent with Cr-1 as the predominant cryptophycin isolated from *Nostoc* sp. ATCC 53789. Structural modifications at Unit B and Unit C revealed

significant affects on MBP-CrpE catalytic efficiency, although substrate solubility ultimately limited kinetic analysis.

Table 2-2. Substrates, slopes, R, and k_{cat}/K_m values in kinetic analysis of MBP-CrpE

Substrate	Slope	R	k_{cat}/K_m
Cr-3	0.396	0.9999	0.099
Cr-4	0.116	0.9981	0.029
Cr-17	0.121	0.9965	0.031
Cr-43	0.320	0.9973	0.080
Cr-B	0.242	0.9999	0.061
Cr-538	1.086	0.9956	0.272

All experiments were performed in duplicate. The final MBP-CrpE concentration was 0.2 μM in all experiment. The volume of all reaction here was 100 μl. The units of Slope and k_{cat}/K_m are pmol·μM^{-1}·min^{-1} and μM^{-1}·min^{-1}, respectively.

Binding Analysis of MBP-CrpE. The characteristics of MBP-CrpE were assessed further by analyzing spectral changes caused by substrate binding to the enzyme. Each cyclic depsipeptide substrate generated a type I binding spectrum with a peak at ~388 nm and a trough at ~422 nm (**Figure 2-7**). The initial plot of ΔA *vs.* S for most substrates could not be fit to the $\Delta A = \Delta A_{max}S/(K_d+S)$ equation. Instead, an equation $\Delta A = \Delta A_{max}S^n/(K_d{}^n+S^n)$ was used to generate sigmoidal curves. Here n is the Hill coefficient, a measure that normally indicates substrate cooperativity. By this equation, ΔA_{max}, K_d, and the n value of each cyclic substrate were generated (**Table 2-3**). All substrates had n values larger than 1, indicating cooperative binding to MBP-CrpE. Interestingly, the calculated Hill coefficient values varied according to specific structural modifications on cryptophycin Unit B and Unit C, suggesting the importance of these subunits in binding of substrates in the enzyme active site pocket. As shown in Eadie-Hofstee plots, all macrocyclic substrates except Cr-43 displayed cooperativity curves (**Figure S2-8**).

The K_d values of all CrpE substrates evaluated in this study were very similar. Cr-17 had the tightest binding with the K_d value of 1.06 ± 0.03 μM, while Cr-3 had the highest K_d value of 1.72 ± 0.03 μM, indicating that the interaction between substrates and the enzyme are substantial. This result also revealed that the structural variations among

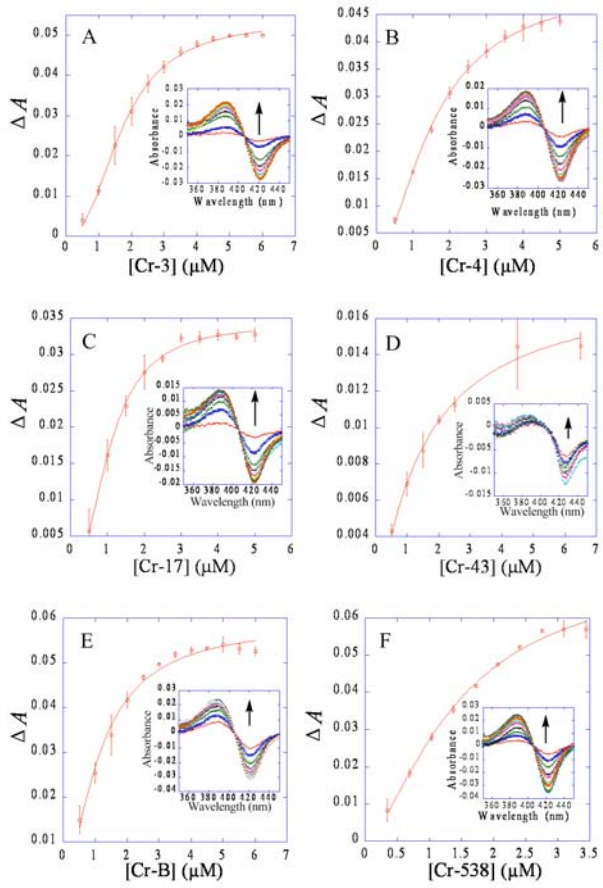


Figure 2-7. **Spectral titration of 0.3 μM MBP-CrpE.** Cr-3 (0.5-6 μM) (A), Cr-4 (0.5-5 μM) (B), Cr-17 (0.5-5 μM) (C), Cr-43 (0.5-6 μM) (D), Cr-B (0.5-6 μM) (E), and Cr-538 (0.35-3.5 μM)(F). The difference spectra resulting from different substrates are shown as insets (A to F). The direction of spectral shift upon substrate addition is shown (arrows). Absorbance changes were determined by subtracting A388 with A422 and were fitted to the equation $\Delta A = \Delta A_{max} S^n/(K_d^n+S^n)$.

these substrates do not significantly affect substrate binding affinity. Maximal absorbance changes caused by different substrate binding varied from 0.0179 ± 0.0015 $\mu M^{-1}cm^{-1}$ of Cr-43 to 0.0744 ± 0.0047 $\mu M^{-1}cm^{-1}$ of Cr-538. Five *seco*-cryptophycin analogues and the linear *SNAC*-Unit A were also assessed in binding analysis (data not shown). Up to 100μM of each linear substrate was added to the protein solution, but none induced a spectral change, indicating the importance of the cryptophycin macrocyclic core for

substrate binding to MBP-CrpE. As a control, GroEL was substituted in place of MBP-CrpE (**Figure S2-9**), and when titrated with Cr-538, no meaningful spectral change was observed.

Table 2-3. Binding analysis of MBP-CrpE with cyclic substrates

Substrates	ΔA_{max} ($\mu M^{-1} \cdot cm^{-1}$)	K_d (μM)	n
Cr-3	0.0536 ± 0.0007	1.72 ± 0.03	2.37 ± 0.10
Cr-4	0.0511 ± 0.0014	1.57 ± 0.07	1.66 ± 0.09
Cr-17	0.0343 ± 0.0004	1.06 ± 0.03	2.20 ± 0.12
Cr-43	0.0179 ± 0.0015	1.50 ± 0.27	1.11 ± 0.16
Cr-B	0.0590 ± 0.0021	1.13 ± 0.07	1.56 ± 0.16
Cr-538	0.0744 ± 0.0047	1.43 ± 0.14	1.53 ± 0.13

All experiments were performed in triplicate.

2.3.3. Characterization of Homotropic and Heterotropic Cooperativity in Cryptophycin Biosynthetic P450 Epoxidase

Stoichiometry of Cr-3 binding to CrpE. In previous studies, CrpE showed homotropic cooperativity to five natural and unnatural cyclic cryptophycin analogs in spectral titration [40]. The Hill coefficient (n) was greater than 2 for Cr-3 in previous assay, presumably indicating at least 3:1 of Cr-3 to CrpE. Although P450 x-ray studies might provide direct evidence to explain substrate cooperativity, the presence of co-purified contaminant, putative chaperone protein GroEL, with CrpE enzyme significantly hampered x-ray studies of this biosynthetic P450 but also triggered efforts to probe deeply into its cooperativity with other methods [41-46]. Job's titration was chosen for its unbiased measurement of ligand to protein molar ratio [47]. Equal concentrations of Cr-3 and CrpE enzyme were used in the assay and the data for Cr-3 binding were fit to produce a bell-shaped graph with its vertex at around 0.25 (**Figure 2-8**). The value of molar fraction (defined as [CrpE]/([CrpE] + [Cr-3])) literally indicated an approximate 1:3 enzyme/substrate molar ratio. This result is consistent with the observation in our previous report and strongly supports CrpE substrate homotropic cooperativity [40].

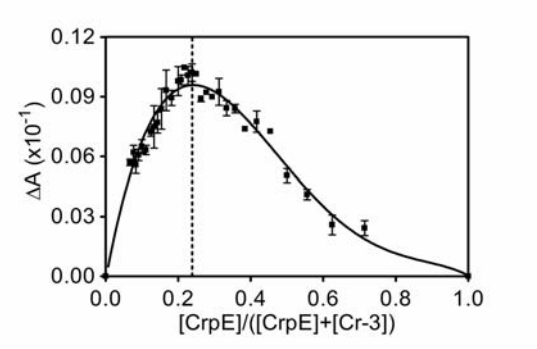


Figure 2-8. Stoichiometry of Cr-3 binding to CrpE. The Job's plot at a total protein and substrate concentration of 2 μM showed a peak with the mole fraction of about 0.25. This value indicated an approximate 3:1 ratio for the Cr-3 binding to CrpE.

Protein Oligomerization upon Substrate Binding. To exclude the possibility of CrpE oligomerization, which might be reflected as substrate cooperativity in substrate titration assay, the concentrations of KCl were varied in the substrate binding analysis (**Table 2-4**). The binding cooperativity of Cr-4 substantially remained the same with up to 300 mM KCl, whose ionic intensity is unambiguously high enough to abolish most protein-protein interaction induced by ligand binding. Moreover, both ΔA_{max} and K_d values kept nearly unchanged in the presence of elevated ionic strength, possibly providing sights into non-ionic interaction between Cr-4 and CrpE. Accordingly, this experiment further confirmed CrpE substrate cooperativity in another respect.

Table 2-4. Salt effect on cryptophycin-4 binding to CrpE.

KCl (mM)	ΔA_{max} ($\mu M^{-1} \cdot cm^{-1}$)	K_d (μM)	n
0	0.0511 ± 0.0014	1.57 ± 0.07	1.66 ± 0.09
50	0.0575 ± 0.0020	1.02 ± 0.01	1.60 ± 0.13
100	0.0618 ± 0.0069	1.37 ± 0.26	1.39 ± 0.24
200	0.0565 ± 0.0068	1.37 ± 0.29	1.36 ± 0.25
300	0.0457 ± 0.0022	1.34 ± 0.11	1.46 ± 0.11

All experiments were performed in duplicate.

Fatty Acid Binding Studies. Linear fatty acids with different lengths have been used to measure the effective spaces in P450 active site pockets [34, 48]. In this study, all four linear saturated fatty acids (C_{12} –C_{18}) induced a "type I" binding spectrum (**Figure S2-10**,

Table 2-5). This type of spectral shift suggests the increasing of high-spin state of heme ion along with the elevated concentration of the fatty acids. The poor binding of stearic acid (C_{18}) indicated that longer chain fatty acid can not displace the distal water ligand of CrpE heme in its substrate binding pocket to produce the high spin ferric state. This result essentially estimated the substrate binding cavity size in CrpE. Interestingly, P450 EpoK in epothilone biosynthesis also accepted linear fatty acids with length from C_{10} to C_{18} in its binding pocket, possibly suggesting the similar effective sizes in two biosynthetic P450s both accommodating 16-membered macrocyclic substrates [34]. However, no cooperativity was observed in EpoK.

Table 2-5. Binding of linear fatty acids to CrpE.

	C12	C14	C16	C18
ΔA_{max} ($\mu M^{-1} \cdot cm^{-1}$)	0.061±0.002	0.067±0.002	0.077±0.002	-
K_d (μM)	152±10	79±5	119±6	-

All experiments were performed in duplicate. Symbol "-" represents that the constants were too big to be determined.

Testosterone Binding to CrpE. Steroid testosterone is a positive cooperative substrate for human CYP3A4 and also bacterial P450EryF mutants [49, 50]. The relative large binding pocket in CrpE estimated by fatty acid binding suggested that this biosynthetic P450 from cyanobacterium could accommodate this chemical. As shown in **Figure 2-9A**, a blue-shift of Soret absorbance band was indeed observed in CrpE spectrum upon testosterone binding. The collected ΔA values generated by adding 0-900 μM testosterone were fitted to the Hill equation (**Figure 2-9B**). The binding affinity of testosterone to CrpE (K_d = 548.4 ± 42.2 μM) was about two-fold and ten times lower than its affinities to P450EryF (K_d = 278 ± 35 μM) and CYP3A4 (K_d = 56 ± 3 μM), respectively. Furthermore, plot of ΔA vs [Testosterone] gave the ΔA_{max} value of 0.23 ± 0.02 μM^{-1} and a Hill coefficient number of 1.97 ± 0.13. Thus, testosterone is a positive cooperative ligand toward cyanobacterial biosynthetic P450 CrpE, making this enzyme the second bacterial P450 featured with homotropic cooperativity toward steroid. Moreover, this result provided further sights into the substrate binding pocket size of this remarkable enzyme.

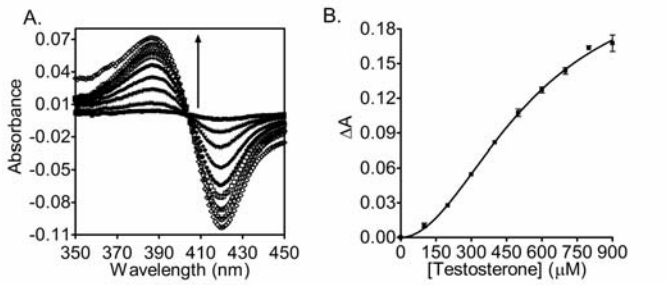


Figure 2-9. Spectral titration of 0.3 μM CrpE with 0-900 μM of testosterone. A: The difference spectra induced by different concentrations of testosterone. The direction of spectral shift upon substrate addition is shown (arrows). B: Plot of absorbance change vs ligand concentration. Absorbance changes were determined by subtracting A_{388} with A_{418} and were fitted to the equation $\Delta A = \Delta A_{max} S^n/(K_d^n+S^n)$.

Interestingly, no hydroxylated or epoxidized testosterones were produced in the CrpE reaction as determined by HPLC analysis, indicating substrate non-catalytic effective binding (**Figure S2-11**).

Heterotropic Cooperativity in CrpE. The similar physical properties of cryptophycin analogs possibly indicate the same binding mode and position of substrate in CrpE. In this study, Cr-3 and Cr-538 were used as two competitive ligands in binding assays to validate this hypothesis. Indeed, the presence of 1 μM Cr-538 in the binding solution significantly reduced both ΔA_{max} and n values of Cr-3 (ΔA_{max} = 0.068 ± 0.005 $\mu M^{-1} \cdot cm^{-1}$; n = 1.31 ± 0.16) (**Figure 2-10**, **Table 2-6**). Also, Cr-538 slightly changed Cr-3 binding affinity (K_d = 1.55 ± 0.22 μM). Similarly, 1 μM Cr-3 caused the same effects on Cr-538 binding (ΔA_{max} = 0.058 ± 0.002 $\mu M^{-1} \cdot cm^{-1}$; n = 1.47 ± 0.10; K_d = 1.13 ± 0.06 μM), which indicated the same site for CrpE to interact with two ligands. Although epothilone D doesn't cooperatively bind to P450EryF, both homotropic and heterotropic cooperativites were observed in this bacterial P450 with foreign substrates and effectors including flavones and steroids [51]. In this study, CrpE was also examined with testosterone for its effects on natural substrate binding. A series of testosterone (0, 50, 100, 300, and 500 μM) were included in the binding solution and both Cr-3 and Cr-538 were gradually added (**Figure 2-10**). Both ΔA_{max} and n values of Cr-3 and Cr-538 were decreased step-by-step with the elevated concentrations of testosterone in the solution, while their K_d values were mainly constant (**Table 2-6**). At 500 μM testosterone, ΔA_{max} values of Cr-3 and Cr-538 were 0.050 ± 0.003 $\mu M^{-1} \cdot cm^{-1}$ and 0.040 ± 0.004 $\mu M^{-1} \cdot cm^{-1}$, respectively. In contrast,

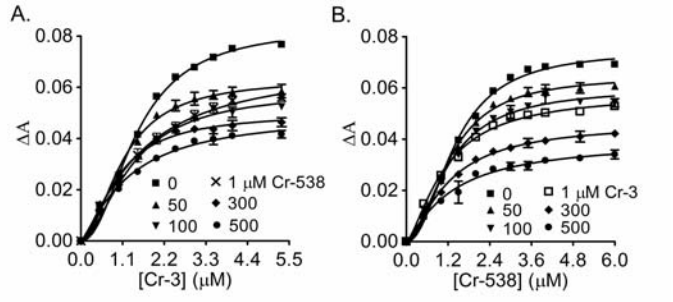


Figure 2-10. Effect of testosterone on substrate binding to CrpE. A: Cr-3 (0-5.5 μM) induced absorbance changes in the presence of testosterone (0, 50, 100, 300, 500 μM) and 1 μM Cr-538. B: Cr-538 (0-6.0 μM) induced absorbance changes in the presence of testosterone (0, 50, 100, 300, 500 μM) and 1 μM Cr-3. The curves represent the results of fitting data sets to the Hill equation.

Table 2-6. Testosterone effects on cryptophycin-3 and cryptophycin-538 binding to CrpE

Testosterone	Cr-3			Cr-538		
(μM)	ΔA_{max}	n	K_d (μM)	ΔA_{max}	n	K_d (μM)
0	0.083±0.002	2.10±0.13	1.46±0.05	0.075±0.001	2.10±0.11	1.39±0.04
50	0.063±0.002	1.89±0.17	1.11±0.06	0.065±0.001	2.00±0.14	1.25±0.05
100	0.059±0.002	1.72±0.13	1.33±0.07	0.060±0.002	1.80±0.18	1.22±0.07
300	0.050±0.002	1.64±0.18	0.98 ±0.06	0.046±0.002	1.46±0.12	1.21 ±0.08
500	0.050±0.003	1.26±0.15	1.24±0.16	0.040±0.004	1.21±0.23	1.31±0.26

All experiments were performed in duplicate.

these values of Cr-3 and Cr-538 were 0.083 ± 0.002 $\mu M^{-1} \cdot cm^{-1}$ and 0.075 ± 0.001 $\mu M^{-1} \cdot cm^{-1}$, respectively, in the absence of testosterone. Similarly, the Hill coefficient values of Cr-3 and Cr-538 were decreased from 2.10 ± 0.13 and 2.10 ± 0.11 in the absence of testosterone to 1.26 ± 0.15 and 1.21 ± 0.23 with 500 μM testosterone, respectively. Thus, the hetereotropic cooperativity in CrpE was revealed with steroid testosterone and natural cryptophycin substrates, representing the first example in any bacterial biosynthetic P450 with its natural substrate as one of ligands.

Testosterone Effects on CrpE Catalysis. Many chemicals including drugs on the market may alter human P450 catalytic activities by acting as activators or inhibitors [52]. Previously, P450eryF A245F was the only bacterial P450 investigated with various effectors including testosterone, which also served as its substrate [51]. In this study,

testosterone was also examined for its effects on CrpE kinetic parameter. From **Table 2-7**, it was shown that the increased concentrations of testosterone triggered significant enhancement of CrpE catalytic efficiency to epoxidize both Cr-3 and Cr-538. In the presence of 500 μM testosterone, the k_{cat}/K_m values of CrpE toward Cr-3 and Cr-538

Table 2-7. Testosterone effects on epoxidizing Cr-3 and Cr-538 by CrpE

Testosterone	Cr-3		Cr-538	
(μM)	Slope	k_{cat}/K_m	Slope	k_{cat}/K_m
0	0.105	0.525	0.538	2.690
50	0.276	1.380	0.671	3.355
100	0.435	2.175	0.833	4.165
300	0.698	3.490	1.523	7.615
500	0.876	4.380	1.657	8.285

All experiments were performed in duplicate. The final CrpE concentration was 0.2 μM in all experiment. The volume of all reactions was 100 μl. The units of Slope and k_{cat}/K_m are pmol·μM^{-1}·min^{-1} and pmol·min^{-1}, respectively.

were 4.380 pmol·min^{-1} and 8.285 pmol·min^{-1}, respectively, which were enhanced respectively by 8.3 and 3.1 times compared with those in the absence of steroid. Subsequently, testosterone effect on overall epoxidation conversion of Cr-538 by CrpE was explored. The amount of Cr-539 in the reactions was increased along with testosterone elevated concentration (**Figure 2-11**). In the presence of 900 μM testosterone, Cr-539 concentration was increased about 6.9 times in CrpE reaction compared to its productivity in the enzyme reaction lack of testosterone, suggesting

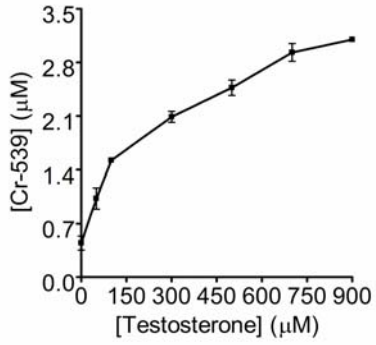


Figure 2-11. **Testosterone effect on the overall epoxidation of Cr-538 by CrpE.** The amount of Cr-539 was increased with the elevated concentration of testosterone (0, 50, 100, 300, 500, 700, 900 μM).

testosterone potential application in efficiently generating anticaner cryptophycin analogs. Interestingly, testosterone did not have any significant effect on reactions of two other bacterial P450s, PikC in pikromycin biosynthesis and MycG in mycinamicin production (**Figure S2-12**) [53, 54]. In these two P450 reactions, product levels had no considerable changes regardless of the presence of 500 μM testosterone. Spectral titration analysis revealed that up to 750 μM testosterone didn't induce any type I spectral shifts in both PikC and MycG solution (data not shown). This result possibly indicated bacterial P450 activity enhancement induced by testosterone might be stemmed from its specific type I binding to CrpE.

2.4. Materials and Methods

General Chemicals, DNA Sub-cloning, and Bacterial Strains. Molecular biology reagents and enzymes were supplied by New England Biolabs with the exception of for *Pfu* DNA polymerase (Stratagene), dNTPs (Takara), T4 DNA Ligase (Invitrogen). Chemicals were purchased from Sigma-Aldrich, Fisher Scientific and BaChem. *Escherichia coli* XL-1 Blue (Stratagene) was used for cloning and plasmid harvesting while *E. coli* BL21 (DE3) (Invitrogen) was used for protein overexpression. All *E. coli* strains were grown in Luria-Bertani broth. Preparation and manipulation of plasmid DNA from *E. coli* was accomplished following manufacture protocols from Qiagen. DNA sequencing was performed at the University of Michigan DNA Sequencing Core. A Beckman Coulter HPLC, fitted with an XBridge C18 column (5μm, 4.6 x 250 mm), coupled with a System Gold I68 Detector was used for HPLC-UV analysis. A SHIMADZU LCMS-2010EV system was used for LC-MS analysis in the studies.

Heterologous Overexpression and Purification of Proteins. A). CrpD-M2 The *crpD-M2* gene was amplified by PCR using cosmid pDHS500 as template. A typical 50 μl reaction mixture contained 5 ng pDHS500, 2 μM forward primer (5'- CAA <u>*GGA TCC*</u> TTA CGT ACT ACT AAT AGC GCA-3') and 2 μM reverse primer (5'- ATG <u>*CTC GAG*</u> TAG TTG TTG AAT TGG TAC TAA TGG-3') (*Bam*HI and *Xho*I sites are underlined), 200 μM dNTP, 5 μl 10 X PCR buffer. Conditions for *crpD-M2* amplifications included

an initial 5 min. denaturation step (94°C) and cycling conditions of 94°C (30 sec), 55°C (30 sec) and 72°C (7 min 45 sec) for 30 cycles followed by a final 72°C extension step (10 min). The amplicon was purified and digested with corresponding enzymes to clone into pET28a. The construct was then transformed into *E. coli* BAP1 competent cell for protein overexpression, which was induced with 0.1 mM IPTG at 15 °C with constant shaking (200 rpm) for 16 hours. The cells were pelleted by centrifugation and resuspended in 20 mL/L cell pellet of lysis buffer (200 mM NaCl, 50 mM Tris-Cl, pH 7.5, 3 mM β-mercaptoethanol, 10 % glycerol, and 20 mM immidazole). The cell suspension was sonicated and the lysate collected following centrifugation (36,000 x g for 32 min). The resulting supernatant was collected, incubated with pre-equilibrated Ni-NTA resin at 4 °C for 2 hours with agitation. The resin was washed (40 column volumes of lysis buffer and lysis buffer with 30 mM immidazole) and CrpD-M2 was eluted with lysis buffer containing 200 mM immidazole. The protein was desalted with a PD-10 column and its purity was assessed with 4-12 % SDS-PAGE. The concentration of the purified protein was determined by its predicted extinction coefficient (259000 M^{-1} cm^{-1} at 280 nm).

B). MBP-CrpE. The *crpE* CrpE gene was amplified by PCR using cosmid pDHS500 as template. A typical 50 µl reaction mixture contained 5 ng pDHS500, 2 µM forward primer (5'-TGC GGA TCC ATG ATT AAT ACT GCT AAA TCC-3') and 2 µM reverse primer (5'- ACG CGA ATT CTT ACA ATA CAA CCA TTT TTA ATC C-3') (*Bam*HI and *Eco*RI sites are underlined), 200 µM dNTP, 5 µl 10 X PCR buffer. Conditions for *crpE* amplifications included an initial 5 min. denaturation step (94°C) and cycling conditions of 94°C (30 sec), 58.5°C (30 sec) and 72°C (1 min 45 sec) for 35 cycles followed by a final 72°C extension step (7 min). The *crpE* amplicon was phosphorylated using T4 Kinase and ligated into pSMart-HCKan to produce pDing1. The insert within pDing1 was sequenced and shown to be free of PCR error. Insert in pDing1 was digested with *Bam*HI and *Eco*RI and ligated into linear pSJ8 (*Eco*RI/*Bam*HI cut). The resulting plasmid, pDing2, and chaperone expressing plasmid pGRO7 were used to co-transform *E. coli* BL21(DE3). Transformants were selected by ampcilin (amp) (50 µg /ml) and chloramphenicol (cml) (25 µg /ml). A 5 mL overnight culture was diluted in 1 L LB supplemented with amp (50 µg /ml), cml (25 µg /ml), 0.25 mM $Fe(NH_4)_2(SO_4)_2$, 1 mM thiamine, and 0.25 mM of 5-aminolevulinic acid that was added 30 minutes prior to

induction (OD_{600} ~ 0.6). The culture was then induced with 0.1 mM IPTG and 1 g /L L-arabinose. The culture was grown at 15 °C with constant shaking (200 rpm) for 20 hours. The cells were pelleted by centrifugation and resuspended in 80 mL of PBS buffer (140 mM NaCl, 2.7 mM potassium chloride, 10 mM sodium hydrogen phosphate, and 1.8 mM potassium dihydrogen phosphate, pH 7.4, 3 mM β-mercaptoethanol, 10 % glycerol). The cell suspension was sonicated and the lysate collected following centrifugation (40,000 x g for 45 min). The resulting supernatant was collected, incubated with pre-equilibrated amylase agarose resin at 4 °C for three hours with agitation. The amylase agarose resin was washed (100 column volumes of lysis buffer) and MBP-His-CrpE eluted with lysis buffer containing 8 mM maltose. Maltose was removed from the protein sample with a PD-10 column, and the MBP-His-CrpE fusion protein was then treated with His-TEV protease at 4 °C overnight to remove the MBP-His portion. The CrpE protein was separated from the MBP-His polypeptide and His-TEV by passing the mixture through a Ni-agarose column. The concentration of the purified protein was determined by its predicted extinction coefficient (81820 M^{-1} cm^{-1} at 280 nm). The active CrpE concentration was determined using previously described methods[55].

Identification of CrpD-M2 and Determination of Phosphopantetheinylation. Full length expression of CrpDm2 was determined by peptide map finger printing. Briefly, CrpDm2 was reduced and digested with trypsin (Pierce TPCK modified). The sample was desalted with Handee Microspin columns (Pierce) packed with 20 µl of 300 Å polymeric C4 resin (Vydac). Samples were loaded onto the columns and washed with 30 column volumes of 0.1% formic acid prior to elution with 10 column volumes of 50% acetonitrile plus 0.1% formic acid. Peptides were then introduced into the FTICRMS at 70ul/hour with direct infusion. Peaks were identified with the thrash algorithm as implemented in MIDAS data analysis workstation (National High Magnetic Field Laboratory). Peaks were matched against a theoretical digest of CrpDm2 (Protein Prospector) with a tolerance of +/-20ppm for the uncelebrated mass spectrum. Due to the extremely large size of the protein (221 kDa) the bottom-up approach was identified as more convenient. Note that low sequence is not surprising due to lack of LC separation,

however, sequence coverage is obtained from residue 3-1,922 out of 1,964 suggesting that the purified protein is full-length.

Determination of CrpD-M2 active site loading. The CrpD-M2 T domain active site peptide (QLVEIFQEVLNLPSIGIHDNFFSLGGH**S**LLAVR) was then loaded with the substrates: L-2HIC, D-2HIC, AKGB, and 2KIC. Peaks were initially identified through LC FITCR-MS using accurate mass. For online confirmation of peak identity the same LC gradient (described below) was run using a Thermo LTQ linear ion trap (LIT) MS. Product peaks were subjected to MS2 and MS3 for further confirmation. Notably, the phosphopantetheine ejection assay performed well on the low resolution instrument, as previously described [56]. The KR reduction reaction was monitored with the same procedure.

ATP-[^{32}P]PP$_i$ Exchange Assays for Aminoacyl-AMP Formation. To determine substrate specificity of CrpD-M2 A domains, ATP-[^{32}P]PP$_i$ reactions (100 μL) containing Tris-HCl (pH 7.5) (75 mM), $MgCl_2$ (10 mM), TCEP (5 mM), ATP (5 mM), NaPP$_i$ (1 mM), α-amino acid, 2-keto acids, 2-hydroxy acid substrates (5 mM), [^{32}P]PP$_i$ (2 Ci/mM) to give at least 400,000 cpm, and 1 μM CrpD-M2 were performed at room temperature. Reactions were initiated by addition of protein. At 5 min, the reactions were quenched into 500 μL of 100 mM NaPP$_i$, 5 % perchloric acid (70 %), and 1.6 % activated charcoal. The charcoal was collected by centrifugation, washed twice with 500 μL of 100 mM NaPP$_i$ and 5% (v/v) perchloric acid, then resuspended in 500 μL of water, added to scintillation fluid, and counted by a Beckman Coulter LS 6500 liquid scintillation counter.

Chemical Synthesis and Characterization:

(*S*)-*S*-2-Acetamidoethyl-2-(*tert*-butoxycarbonyloxy)-4-methylpentanethioate (2). To a stirring solution of Boc-protected (*S*)-leucic acid (1) (295 mg, 1.27 mmol) in CH_2Cl_2 (8 mL) was added EDC (268 mg, 1.40 mmol), DMAP (16 mg, 0.13 mmol) and *N*-acetylcysteamine (149 μL, 1.40 mmol). The reaction was stirred for 18 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (2% MeOH/CH_2Cl_2) afforded the title compound 2 (387

mg, 91%) as a colorless oil. TLC R_f = 0.32 (5% MeOH/CH_2Cl_2, vanillin stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 5.98 (br t, 1H), 4.99 (dd, *J* = 9.9, 3.8, 1H), 3.30–3.49 (m, 2H), 2.97–3.10 (m, 2H), 1.93 (s, 3H), 1.69–1.82 (m, 2H), 1.53–1.61 (m, 1H), 1.48 (s, 9H), 0.92 (d, *J* = 6.6, 3H), 0.91 (d, *J* = 6.6, 3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 200.4, 170.5, 152.9, 83.6, 79.8, 40.8, 39.4, 27.8, 27.8, 24.5, 23.2, 23.1, 21.5; MS (ESI+) *m*/*z* 334.0 $[M + H]^+$ ($C_{15}H_{28}NO_5S$ requires 334.2).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-(*tert*-butoxycarbonyl-amino)-2-methylpropanoate (4a). Compound 2 (387 mg, 1.16 mmol) was treated with anhydrous 4 M HCl in dioxane (5 mL) at room temperature for 40 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (20 mL) and (*R*)-3-*tert*-butyloxycarbonylamino-2-methylpropanoic acid (3a) (236 mg, 1.16 mmol), EDC (222 mg, 1.16 mmol) and DMAP (14 mg, 0.12 mmol) was added. The reaction was stirred for 18 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (2% MeOH/CH_2Cl_2) afforded the title compound 4a (285 mg, 59%) as a light yellow solid. TLC R_f = 0.23 (5% MeOH/CH_2Cl_2, $KMnO_4$ stain); ^{1}H-NMR ($CDCl_3$, 500 MHz) δ 5.83 (br t, 1H), 5.22–5.25 (m, 1H), 5.00 (br t, 1H), 3.35–3.44 (m, 3H), 3.23–3.29 (m, 1H), 3.05 (t, *J* = 6.3, 2H), 2.75–2.83 (m, 1H), 1.96 (s, 3H), 1.61–1.80 (m, 3H), 1.43 (s, 9H), 1.25 (d, *J* = 7.2, 3H), 0.95 (d, *J* = 6.4, 3H), 0.93 (d, *J* = 6.4, 3H); MS (ESI+) *m*/*z* 441.2 $[M + Na]^+$ ($C_{19}H_{34}N_2NaO_6S$ requires 441.2).

((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-(*tert*-butoxycarbonyl-amino)-propanoate (4b). Compound 2 (170 mg, 0.51 mmol) was treated with anhydrous 4 M HCl in dioxane (5 mL) at room temperature for 30 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (8 mL) and Boc-*β*-Ala-OH (3b) (104 mg, 0.55 mmol), EDC (105 mg, 0.55 mmol) and DMAP (7 mg, 0.06 mmol) was added. The reaction was stirred for 18 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (10% MeOH/CH_2Cl_2) afforded the title compound 4b (163 mg, 78%) as a colorless oil. TLC R_f = 0.20 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H-NMR ($CDCl_3$, 400 MHz) δ 5.96 (br t, 1H), 5.21–5.24 (m, 1H), 5.09 (br t, 1H), 3.37–3.46 (m, 4H), 2.97–3.08 (m, 2H), 2.56–2.70 (m, 2H), 1.95 (s, 3H), 1.57–1.78 (m, 3H), 1.42 (s, 9H), 0.93 (d, *J* = 6.4, 3H), 0.91 (d, *J* = 6.5,

3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 199.7, 171.9, 170.6, 155.9, 77.7, 40.9, 40.8, 39.4, 36.3, 34.7, 28.5, 28.1, 27.9, 24.7, 23.1, 21.7; MS (ESI+) *m*/*z* 427.2 $[M + Na]^+$ ($C_{18}H_{32}N_2NaO_6S$ requires 427.2).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-(*tert*-butoxy–carbonylamino)-3-(3-chloro-4-methoxyphenyl)propanamido)-2-methylpropanoate (6a). Compound 4a (46 mg, 0.11 mmol) was treated with anhydrous 4 M HCl in dioxane (2 mL) at room temperature for 30 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (5 mL) and Boc-3-Cl-D-Tyr(Me)-OH (5a) (40 mg, 0.12 mmol), TBTU (39 mg, 0.12 mmol) and DIEA (63 μL, 0.36 mmol) was added. The reaction was stirred for 3 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (3% MeOH/CH_2Cl_2) afforded the title compound 6a (67 mg, 96%) as a colorless oil. TLC R_f = 0.14 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.24 (br s, 1H), 7.04–7.08 (m, 2H), 6.83 (d, *J* = 8.4, 1H), 6.56 (br t, 1H), 5.33 (d, *J* = 7.6, 1H), 5.23–5.27 (m, 1H), 4.37 (dt, *J* = 8.2, 7.0, 1H), 3.86 (s, 3H), 3.63–3.70 (m, 1H), 3.45–3.52 (m, 1H), 2.98–3.22 (m, 5H), 2.91 (dd, *J* = 13.8, 7.3, 1H), 2.75–2.81 (m, 1H), 1.94 (s, 3H), 1.63–1.82 (m, 3H), 1.37 (s, 9H), 1.19 (d, *J* = 6.8, 3H), 0.96 (d, *J* = 6.3, 3H), 0.92 (d, *J* = 6.3, 3H); MS (ESI+) *m*/*z* 652.0 $[M + Na]^+$ ($C_{29}H_{44}ClN_3NaO_8S$ requires 652.2).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-(*tert*-butoxy–carbonylamino)-3-(3-chloro-4-hydroxyphenyl)propanamido)-2-methylpropanoate (6b). Compound 4a (42 mg, 0.10 mmol) was treated with anhydrous 4 M HCl in dioxane (2 mL) at room temperature for 30 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (5 mL) and Boc-3-Cl-D-Tyr-OH (5b) (35 mg, 0.11 mmol), TBTU (35 mg, 0.11 mmol) and DIEA (57 μL, 0.33 mmol) was added. The reaction was stirred for 3 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (4% MeOH/CH_2Cl_2) afforded the title compound 6b (30 mg, 49%) as a colorless oil. TLC R_f = 0.64 (10% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.18 (s, 1H), 7.00 (d, *J* = 8.0, 1H), 6.93 (br s, 1H), 6.91 (d, *J* = 8.4, 1H), 6.55 (br t, 1H), 5.32 (d, *J* = 7.8, 1H), 5.25 (dd, *J* = 10.0, 3.1, 1H), 4.34 (dt, *J* = 8.2, 7.0, 1H), 3.62–3.68 (m, 1H), 3.44–3.52 (m, 1H), 2.90–3.26 (m, 6H), 2.72–2.79 (m, 1H), 1.95 (s, 3H), 1.62–1.79 (m, 3H), 1.38 (s, 9H),

1.18 (d, J = 7.0, 3H), 0.95 (d, J = 6.1, 3H), 0.92 (d, J = 6.3, 3H); MS (ESI+) m/z 638.0 $[M + Na]^+$ ($C_{28}H_{42}ClN_3NaO_8S$ requires 638.2).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-(*tert*-butoxy–carbonylamino)-3-(4-methoxyphenyl)propanamido)-2-methylpropanoate (6c). Compound 4a (45 mg, 0.11 mmol) was treated with anhydrous 4 M HCl in dioxane (2 mL) at room temperature for 20 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (5 mL) and Boc-D-Tyr(Me)-OH (5c) (35 mg, 0.12 mmol), TBTU (39 mg, 0.12 mmol) and DIEA (63 μL, 0.36 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (3% MeOH/CH_2Cl_2) afforded the title compound 6c (58 mg, 90%) as a white solid. TLC R_f = 0.32 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.07–7.12 (m, 2H), 6.78–6.84 (m, 3H), 6.52 (br t, 1H), 5.20–5.26 (m, 2H), 4.32–4.38 (m, 1H), 3.77 (s, 3H), 3.58–3.64 (m, 1H), 3.42–3.49 (m, 1H), 3.16–3.28 (m, 2H), 2.92–3.10 (m, 4H), 2.71–2.78 (m, 1H), 1.94 (s, 3H), 1.62–1.78 (m, 3H), 1.37 (s, 9H), 1.18 (d, J = 6.6, 3H), 0.95 (d, J = 6.2, 3H), 0.92 (d, J = 6.5, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 200.4, 173.7, 171.5, 170.9, 158.7, 155.4, 130.5, 128.9, 114.1, 80.1, 56.1, 55.4, 40.8, 40.4, 39.3, 39.2, 38.8, 29.8, 28.4, 28.3, 25.0, 23.2, 23.2, 21.6, 14.7; MS (ESI+) m/z 618.2 $[M + Na]^+$ ($C_{29}H_{45}N_3NaO_8S$ requires 618.3).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-(*tert*-butoxy–carbonylamino)-3-(4-hydroxyphenyl)propanamido)-2-methylpropanoate (6d). Compound 4a (40 mg, 0.10 mmol) was treated with anhydrous 4 M HCl in dioxane (2 mL) at room temperature for 30 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (5 mL) and Boc-D-Tyr-OH (5d) (31 mg, 0.11 mmol), TBTU (35 mg, 0.11 mmol) and DIEA (57 μL, 0.33 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (3% MeOH/CH_2Cl_2) afforded the title compound 6d (51 mg, 92%) as a white solid. TLC R_f = 0.21 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.62 (br s, 1H), 7.01–7.04 (m, 2H), 6.76 (d, J = 8.4, 2H), 6.63 (br s, 1H), 6.50 (br s, 1H), 5.22–5.27 (m, 2H), 4.27–4.32 (m, 1H), 3.53–3.60 (m, 1H), 3.41–3.47 (m, 1H), 3.19–3.32 (m, 2H), 2.85–3.07 (m, 4H), 2.67–2.74 (m, 1H), 1.96

(s, 3H), 1.57–1.77 (m, 3H), 1.39 (s, 9H), 1.16 (d, J = 6.6, 3H), 0.93 (d, J = 6.3, 3H), 0.90 (d, J = 6.2, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 200.3, 173.8, 171.7, 171.5, 155.8, 155.4, 130.5, 127.9, 115.7, 80.1, 56.4, 41.7, 40.8, 40.4, 39.4, 38.8, 29.8, 28.4, 28.1, 24.8, 23.2, 23.2, 21.6, 14.7; MS (ESI+) m/z 604.1 $[M + Na]^+$ ($C_{28}H_{43}N_3NaO_8S$ requires 604.3).

((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-(*tert*-butoxy–carbonylamino)-3-(3-chloro-4-methoxyphenyl)propanamido)-propanoate (6e). Compound 4b (160 mg, 0.40 mmol) was treated with anhydrous 4 M HCl in dioxane (5 mL) at room temperature for 30 min then the reaction was concentrated *in-vacuo*. The residue was dissolved in CH_2Cl_2 (8 mL) and Boc-3-Cl-D-Tyr(Me)-OH (5a) (148 mg, 0.45 mmol), TBTU (144 mg, 0.45 mmol) and DIEA (226 μL, 1.30 mmol) was added. The reaction was stirred for 90 min at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash chromatography (7% MeOH/CH_2Cl_2) afforded the title compound 6e (235 mg, 95%) as a white foam. TLC R_f = 0.67 (10% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.26 (d, J = 1.9, 1H), 6.98–7.06 (m, 2H), 6.82 (d, J = 8.4, 1H), 6.43 (br t, 1H), 5.30 (d, J = 7.6, 1H), 5.21–5.23 (m, 1H), 4.33–4.37 (m, 1H), 3.85 (s, 3H), 3.48–3.56 (m, 3H), 3.22–3.27 (m, 1H), 2.95–3.12 (m, 4H), 2.58 (d, J = 6.8, 2H), 1.94 (s, 3H), 1.58–1.77 (m, 3H), 1.36 (s, 9H), 0.93 (d, J = 6.3, 3H), 0.90 (d, J = 6.3, 3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 200.4, 171.3, 171.2, 171.0, 155.4, 154.0, 131.2, 130.2, 128.7, 122.3, 112.1, 77.7, 56.2, 55.8, 40.7, 39.1, 38.1, 35.0, 34.3, 28.4, 28.3, 28.2, 24.7, 23.2, 23.1, 21.5; MS (ESI+) m/z 638.2 $[M + Na]^+$ ($C_{28}H_{42}ClN_3NaO_8S$ requires 638.2).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-((2*E*,5*S*,6*R*,7*E*)-5-hydroxy-6-methyl-8-phenylocta-2,7-dienamido)-3-(3-chloro-4-methoxyphenyl)propan-amido)-2-methylpropanoate (Cr-438). Compound 6a (67 mg, 0.11 mmol) and the *tert*-butyl ester of unit A (7) (32 mg, 0.11 mmol) were stirred in CH_2Cl_2 (4 mL) and TFA (1 mL) for 1 h at room temperature. The solvents were evaporated *in-vacuo* and residual TFA was removed by co-evaporation with toluene. The residue was dissolved in CH_2Cl_2 (5 mL) and TBTU (35 mg, 0.11 mmol) and DIEA (57 μL, 0.33 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash-chromatography (3% MeOH/CH_2Cl_2) afforded the title compound Cr-438 (52 mg, 65%)

as a white foam. TLC R_f = 0.57 (10% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.36 (d, *J* = 7.4, 2H), 7.30 (t, *J* = 7.6, 2H), 7.19–7.25 (m, 2H), 7.05–7.07 (m, 2H), 6.95 (br t, 1H), 6.79–6.83 (m, 2H), 6.59 (d, *J* = 8.0, 1H), 6.44 (d, *J* = 16.0, 1H), 6.15 (dd, *J* = 16.0, 8.5, 1H), 5.91 (d, *J* = 15.4, 1H), 5.22 (dd, *J* = 9.9, 3.4, 1H), 4.79 (dt, *J* = 7.6, 7.1, 1H), 3.85 (s, 3H), 3.60–3.67 (m, 2H), 3.44–3.53 (m, 1H), 3.12–3.26 (m, 2H), 3.09 (t, *J* = 6.6, 2H), 2.97–3.02 (m, 2H), 2.72–2.78 (m, 1H), 2.26–2.45 (m, 3H), 2.12 (br s, 1H), 1.96 (s, 3H), 1.59–1.77 (m, 3H), 1.15 (d, *J* = 6.6, 3H), 1.13 (d, *J* = 6.3, 3H), 0.94 (d, *J* = 6.1, 3H), 0.90 (d, *J* = 6.1, 3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 200.6, 173.5, 171.2, 171.0, 165.6, 154.0, 142.3, 137.2, 131.8, 131.3, 131.2, 130.0, 128.7, 128.7, 127.5, 126.3, 125.6, 122.3, 112.2, 77.6, 73.9, 56.2, 54.6, 43.4, 42.0, 40.7, 40.5, 39.2, 37.9, 37.5, 28.3, 24.9, 23.2, 23.2, 21.6, 17.0, 14.6; MS (ESI+) *m*/*z* 758.0 $[M + H]^+$ ($C_{39}H_{53}ClN_3O_8S$ requires 758.3).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-((2*E*,5*S*,6*R*,7*E*)-5-hydroxy-6-methyl-8-phenylocta-2,7-dienamido)-3-(3-chloro-4-hydroxyphenyl)propan-amido)-2-methylpropanoate (Cr-441). Compound 6b (48 mg, 0.08 mmol) and the *tert*-butyl ester of unit A (7) (24 mg, 0.08 mmol) were stirred in CH_2Cl_2 (4 mL) and TFA (1 mL) for 1 h at room temperature. The solvents were evaporated *in-vacuo* and residual TFA was removed by co-evaporation with toluene. The residue was dissolved in CH_2Cl_2 (5 mL) and TBTU (25 mg, 0.08 mmol) and DIEA (41 μL, 0.23 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash-chromatography (3%, 5% MeOH/CH_2Cl_2) afforded the title compound Cr-441 (26 mg, 45%) as a white solid. TLC R_f = 0.59 (10% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.36 (d, *J* = 7.4, 2H), 7.30 (t, *J* = 7.5, 2H), 7.18–7.23 (m, 2H), 6.97–7.03 (m, 2H), 6.80–6.92 (m, 3H), 6.64 (br s, 1H), 6.44 (d, *J* = 15.8, 1H), 6.15 (dd, *J* = 15.8, 8.5, 1H), 5.91 (d, *J* = 15.4, 1H), 5.22 (dd, *J* = 9.9, 3.4, 1H), 4.76 (dt, *J* = 8.0, 6.9, 1H), 3.56–3.66 (m, 2H), 3.43–3.50 (m, 1H), 3.18–3.28 (m, 2H), 3.07 (t, *J* = 6.7, 2H), 2.97 (d, *J* = 6.8, 2H), 2.70–2.77 (m, 1H), 2.26–2.45 (m, 4H), 1.96 (s, 3H), 1.60–1.77 (m, 3H), 1.14 (d, *J* = 7.2, 3H), 1.13 (d, *J* = 6.6, 3H), 0.94 (d, *J* = 6.0, 3H), 0.90 (d, *J* = 6.1, 3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 200.6, 173.6, 171.5, 171.1, 165.6, 150.9, 142.5, 137.2, 131.8, 131.3, 130.2, 129.7, 129.2, 128.7, 127.5, 126.3, 125.6, 120.1, 116.6, 77.6, 73.9, 54.8, 43.5,

41.9, 40.7, 40.4, 39.3, 38.1, 37.4, 28.2, 24.9, 23.2, 23.1, 21.6, 17.0, 14.6; MS (ESI+) *m*/*z* 744.0 $[M + H]^+$ ($C_{38}H_{51}ClN_3O_8S$ requires 744.3).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-((2*E*,5*S*,6*R*,7*E*)-5-hydroxy-6-methyl-8-phenylocta-2,7-dienamido)-3-(4-methoxyphenyl)propanamido)-2-methylpropanoate (Cr-424). Compound 6c (58 mg, 0.10 mmol) and the *tert*-butyl ester of unit A (7) (29 mg, 0.10 mmol) were stirred in CH_2Cl_2 (4 mL) and TFA (1 mL) for 1 h at room temperature. The solvents were evaporated *in-vacuo* and residual TFA was removed by co-evaporation with toluene. The residue was dissolved in CH_2Cl_2 (5 mL) and TBTU (32 mg, 0.10 mmol) and DIEA (52 μL, 0.30 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash-chromatography (3% MeOH/CH_2Cl_2) afforded the title compound Cr-424 (46 mg, 66%) as colorless oil. TLC R_f = 0.23 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.36 (d, *J* = 7.4, 2H), 7.30 (t, *J* = 7.4, 2H), 7.21 (t, *J* = 6.9, 1H), 7.11 (d, *J* = 8.3, 2H), 6.96 (br t, 1H), 6.90–6.95 (m, 1H), 6.80–6.85 (m, 1H), 6.79 (d, *J* = 8.1, 2H), 6.48 (br t, 1H), 6.44 (d, *J* = 16.0, 1H), 6.15 (dd, *J* = 16.0, 8.6, 1H), 5.90 (d, *J* = 15.2, 1H), 5.20–5.23 (m, 1H), 4.76 (app q, *J* = 7.4, 1H), 3.76 (s, 3H), 3.57–3.65 (m, 2H), 3.44–3.51 (m, 1H), 3.13–3.22 (m, 2H), 3.08 (t, *J* = 6.4, 2H), 3.02 (d, *J* = 6.6, 2H), 2.72–2.78 (m, 1H), 2.26–2.47 (m, 3H), 2.15 (br s, 1H), 1.96 (s, 3H), 1.60–1.75 (m, 3H), 1.14 (d, *J* = 6.7, 3H), 1.13 (d, *J* = 6.3, 3H), 0.94 (d, *J* = 6.0, 3H), 0.90 (d, *J* = 6.0, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 200.3, 173.6, 171.1, 171.1, 165.5, 158.7, 142.2, 137.2, 131.8, 131.3, 130.4, 128.7, 128.7, 127.5, 126.3, 125.7, 114.1, 77.5, 73.9, 55.3, 54.8, 43.4, 41.9, 40.7, 40.4, 39.2, 38.0, 37.5, 28.4, 24.9, 23.2, 23.2, 21.6, 17.0, 14.7; MS (ESI+) *m*/*z* 724.3 $[M + H]^+$ ($C_{39}H_{54}N_3O_8S$ requires 724.4).

(*R*)-((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-((2*E*,5*S*,6*R*,7*E*)-5-hydroxy-6-methyl-8-phenylocta-2,7-dienamido)-3-(4-hydroxyphenyl)propanamido)-2-methylpropanoate (Cr-430). Compound 6d (58 mg, 0.09 mmol) and the *tert*-butyl ester of unit A (7) (27 mg, 0.09 mmol) were stirred in CH_2Cl_2 (4 mL) and TFA (1 mL) for 1 h at room temperature. The solvents were evaporated *in-vacuo* and residual TFA was removed by co-evaporation with toluene. The residue was dissolved in CH_2Cl_2 (5 mL) and TBTU (28 mg, 0.09 mmol) and DIEA (46

μL, 0.26 mmol) was added. The reaction was stirred for 2 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash-chromatography (3%, 10% MeOH/CH_2Cl_2) afforded the title compound Cr-430 (38 mg, 61%) as a white solid. TLC R_f = 0.17 (5% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.34 (d, *J* = 7.2, 2H), 7.28 (t, *J* = 7.4, 2H), 7.20 (t, *J* = 7.2, 1H), 7.00 (d, *J* = 8.4, 2H), 6.94 (t, *J* = 5.6, 1H), 6.79–6.89 (m, 3H), 6.72 (d, *J* = 8.4, 2H), 6.42 (d, *J* = 16.0, 1H), 6.15 (dd, *J* = 16.0, 8.5, 1H), 5.90 (d, *J* = 15.4, 1H), 5.20 (dd, *J* = 9.8, 3.5, 1H), 4.70 (dt, *J* = 7.6, 7.2, 1H), 3.59–3.65 (m, 1H), 3.37–3.52 (m, 2H), 3.21–3.30 (m, 2H), 3.02 (t, *J* = 6.6, 2H), 2.92–2.98 (m, 2H), 2.66–2.75 (m, 1H), 2.24–2.42 (m, 4H), 1.93 (s, 3H), 1.57–1.77 (m, 3H), 1.12 (d, *J* = 6.4, 6H), 0.92 (d, *J* = 6.1, 3H), 0.88 (d, *J* = 6.2, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 200.2, 173.7, 171.7, 171.7, 165.9, 155.8, 142.5, 137.3, 131.5, 131.4, 130.4, 128.7, 127.7, 127.4, 126.3, 125.6, 115.8, 77.6, 74.0, 55.2, 43.4, 41.8, 40.7, 40.2, 39.3, 38.2, 37.5, 28.1, 24.8, 23.2, 23.1, 21.6, 17.0, 14.6; MS (ESI+) *m*/*z* 710.2 $[M + H]^+$ ($C_{38}H_{52}N_3O_8S$ requires 710.4).

((*S*)-1-(2-Acetamidoethylthio)-4-methyl-1-oxopentan-2-yl) 3-((*R*)-2-((2*E*,5*S*,6*R*,7*E*)-5-hydroxy-6-methyl-8-phenylocta-2,7-dienamido)-3-(3-chloro-4-methoxyphenyl)propan-amido)-propanoate (Cr-537). Compound 6e (50 mg, 0.08 mmol) and the *tert*-butyl ester of unit A (7) (24 mg, 0.08 mmol) were stirred in CH_2Cl_2 (4 mL) and TFA (1 mL) for 1 h at room temperature. The solvents were evaporated *in-vacuo* and residual TFA was removed by co-evaporation with toluene. The residue was dissolved in CH_2Cl_2 (6 mL) and TBTU (26 mg, 0.08 mmol) and DIEA (42 μL, 0.24 mmol) was added. The reaction was stirred for 1 h at room temperature then washed with water, dried with Na_2SO_4, filtered and concentrated *in-vacuo*. Flash-chromatography (3% MeOH/CH_2Cl_2) afforded the title compound Cr-537 (40 mg, 67%) as a white solid. TLC R_f = 0.48 (10% MeOH/CH_2Cl_2, PMA stain); ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.36 (d, *J* = 7.2, 2H), 7.29 (t, *J* = 7.2, 2H), 7.21–7.25 (m, 2H), 7.12 (t, *J* = 5.6, 1H), 7.06 (dd, *J* = 8.4, 2.1, 1H), 6.75–6.87 (m, 3H), 6.66 (d, *J* = 7.8, 1H), 6.43 (d, *J* = 15.8, 1H), 6.15 (dd, *J* = 15.8, 8.5, 1H), 5.90 (d, *J* = 15.2, 1H), 5.22 (dd, *J* = 9.8, 3.7, 1H), 4.78 (dt, *J* = 7.8, 7.0, 1H), 3.85 (s, 3H), 3.61–3.66 (m, 1H), 3.46–3.57 (m, 3H), 3.21–3.29 (m, 1H), 2.95–3.11 (m, 4H), 2.28–2.58 (m, 6H), 1.96 (s, 3H), 1.56–1.75 (m, 3H), 1.13 (d, *J* = 6.6, 3H), 0.94 (d, *J* = 6.4, 3H), 0.90 (d, *J* = 6.4, 3H); ^{13}C NMR ($CDCl_3$, 101 MHz) δ 200.4, 171.3, 171.3,

171.0, 165.7, 154.0, 142.4, 137.2, 131.7, 131.3, 131.2, 130.0, 128.7, 128.7, 127.5, 126.3, 125.6, 122.3, 112.1, 77.8, 73.9, 56.2, 54.6, 43.4, 40.7, 39.2, 37.6, 37.5, 35.2, 34.3, 28.3, 24.7, 23.2, 23.2, 21.5, 17.0; MS (ESI+) *m/z* 744.3 $[M + H]^+$ ($C_{38}H_{51}ClN_3O_8S$ requires 744.3).

mCPBA Reactions with Cr-B and Cr-538 as Substrates. 0.5 mg (~ 0.8 µmol) of Cr-B and Cr-538 were dissolved in DCM. About 1.0 µmol of mCPBA in DCM was added to 2 mL reaction mixture dropwise with stirring in water-ice bath. After 1 h, the reaction was naturally warmed to room temperature and continued for 16 h. The reaction mixture was dried under nitrogen gas and redissolved in MeOH for HPLC analysis, whose conditions were as the same as enzymatic reaction assays.

Enzyme Activity Assay. A). CrpD-M2 To load and reduce substrates, the enzyme reaction mixture contained 100 mM Tris-Cl, pH 7.5, 10 mM $MgCl_2$, 5 mM ATP, 1 mM TCEP, 2.5 mM substrate (2KIC, 2HIC, or AKGB), about 5 µM holo-CrpD-M2 in a total volume of 50 µl. The reaction mixture was incubated at room temperature for 30 minutes for loading reaction. The KR catalyzed reduction reaction was subsequently initiated with addition of 2 mM NADPH or NADH and incubated at room temperature for 30 minutes. UnitABC-NAC ester (100 µM) was added to start the condensation reaction catalyzed by CrpD-M2 C domain. The reaction was performed at room temperature for 1 h and then 10 µM CrpD-TE was added to the above mixture. The reaction mixture was further incubated for 1-2 hours at room temperature and was then terminated by adding two volumes of MeOH. Products were detected and collected with analytical reverse-phase HPLC (XBridge C18 column, 1 mL/min, 30-100% acetonitrile/water + 0.1% TFA, 30 min, 218 nm). The product was then re-dissolved and subjected to MS analysis.

B). MBP-CrpE MBP-CrpE reactions contained 100 µg/ml ferredoxin from spinach, 0.2 unit/ ml ferredoxin-$NADP^+$ reductase from spinach, 1.4 mM NADPH, 10 mM glucose-6-phosphate, 8 units/ml glucose-6-phosphate dehydrogenase in 100 µl storage buffer (50 mM sodium phosphate solution, pH 6.4). The alternative redox partners tested were *E. coli* NADPH-flavodoxin reductase and flavodoxin or rat NADPH cytochrome P450 reductase. A DMSO stock solution of Cr-3 (2 mM) was added to the reaction mixture to

reach a final concentration of 5 μM, such that the final concentration of DMSO in the reaction was 5%. The reaction mixture was pre-warmed at room temperature for 2 min. The reactions were initiated by addition of MBP-CrpE to a final concentration of 0.2 μM. Reaction mixtures were incubated at room temperature for 2-4 hours. In order to investigate optimal conditions for the MBP-CrpE reaction, reactions were performed at 4 °C, 15 °C, 25 °C, 30 °C, 37 °C, and 42 °C. In addition, optimal pH of the reaction was determined by monitoring product formation as a function of reaction pH (pH of 5, 6, 6.4, 7, 8, and 9). From these analyses, optimal reaction conditions were determined to be 50 mM sodium phosphate buffer, pH 6.4 at 25 °C. Cr-4, Cr-17, Cr-43, Cr-B, Cr-538, *SNAC*-Unit A, and *seco*-cryptophycin Cr-438, Cr-424, Cr-442, Cr-430, and Cr-537 (each at a final concentration of 5 μM) were used investigate MBP-CrpE substrate selectivity. Following incubation at 25 °C for 2-4 hours, reaction mixtures were extracted twice with equal volume of ethyl acetate. The combined ethyl acetate extracts were dried and subsequently concentrated *in vacuo*. The concentrated compounds were dissolved into DMSO:water (1:1) and separated by analytical reverse-phase HPLC (XBridge C18 column, 1 mL/min, 30-100% acetonitrile/water + 0.1% TFA, 40 min, 218 nm).

Kinetic Analysis of MBP-CrpE Reactions. MBP-CrpE reactions contained 100 μg/ml ferredoxin, 0.2 unit/ml ferredoxin-$NADP^+$ reductase, and 1.4 mM of NADPH in 100 μl storage buffer. Substrate concentrations ranged from 2-100 μM for Cr-3, Cr-4, Cr-17, Cr-43, Cr-B 2-60 μM for Cr-538. The reactions contained 5 % DMSO (v/v). The reaction mixture was pre-warmed at room temperature for 2 min, after which MBP-CrpE was added to a final concentration of 0.2 μM to initiate each reaction. The reactions were incubated at room temperature for 10 min and quenched with addition of two volumes of methanol. After centrifugation at 13,200 rpm for 5 minutes, the 300-μl solutions were subjected to HPLC analysis. The areas of the product peaks were calculated and converted to moles by fitting into Cr-4 standard curve. The kinetic data was analyzed with KaleidaGraph software and fit to the linear function. All experiments were performed in duplicate.

Binding Analysis of MBP-CrpE. Binding affinities of ligands to MBP-CrpE were determined using 0.3 μM of enzyme in a total volume of 1.0 ml of 50 mM sodium phosphate solution, pH 6.4. A Carey 300 Bio UV-Visible spectrophotometer was used to record the spectrum from 350 nm to 450 nm. Cryptophycin analogs and testosterone were dissolved in DMSO while fatty acids stock solutions were prepared in ethanol. The same amount of organic solvents was added to the reference cuvette. KCl (0-300 mM) was included in the binding buffer to examine binding constant changes under high ionic strength. To determine K_d, ΔA_{max}, and the Hill coefficient (n), titration data sets were fitted to the Hill equation of $\Delta A = \Delta A_{max} S^n/(K_d^n+S^n)$. Here, ΔA is the absorbent shift, ΔA_{max} is the maximal absorbent shift, K_d is the apparent dissociation constant for ligand-enzyme complex, S is the ligand concentration, and n is a measure of cooperativity. All experiments were conducted in duplicate. In the control experiment, 1 μM GroEL solution was used with Cr-538 as the ligand.

Job's Titration. This analysis followed a protocol in a previous report [41]. Briefly, a Carey 300 Bio UV-Visible spectrophotometer was used to record the spectrum from 350 nm to 450 nm. Both CrpE and Cr-3 were kept in 100 mM sodium phosphate solution, pH 6.4 (binding buffer), and had the same concentration as 2 μM in the assay. Experiments were carried out at 25 °C. An aliquot of 100 μl CrpE solution was initially transferred into plastic 1-cm cuvette. The Cr-3 solution was gradually added to enzyme solution until the volume of the mixture reached 1580 μl and spectral changes were record. The total concentration of reactants was kept constant to be 2 μM. As the reference, an equal volume of binding buffer was gradually added to the enzyme solution. Spectral changes were calculated by subtracting $A_{365\sim370}$ with $A_{410\text{-}415}$ and the data were analyzed with Prism 4 (GraphPad). All experiments were conducted in duplicate.

Testosterone Effects on CrpE Activity Assay. CrpE reaction contained 200 μg/ml Fer, 0.2 unit/ml FNR, and 2.5 mM of freshly prepared NADPH. When testosterone (250 μM) was the enzyme substrate, the reactions were incubated at room temperature for 2 hours and putative products were analyzed by analytical reverse-phase HPLC (XBridge C18 column, 1 mL/min, 254 nm). Solvent B (acetonitrile) was kept at 10 % for 5 min and then

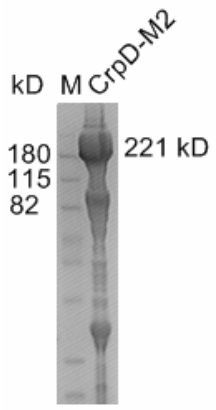
increased to 24 % over next 5 min. Subsequently, solvent B was increased to 62 % in 7 min, at which it was kept for 20 min. The solvent A was water. Testosterone was eluted at 21.7 min in HPLC trace. To investigate its effect on enzyme kinetics, testosterone (0, 50, 100, 300, 500 μM) was added to the enzyme solutions containing Cr-3 or Cr-538. The kinetic analyses were followed the previously described protocol [40]. The kinetic data were analyzed with Prism 4 and fitted to the linear function. All experiments were performed in duplicate.

Testosterone Effect on Overall Catalytic Activity of Selected Bacterial P450s. Testosterone was further applied to investigate its effect on the overall catalytic activity of CrpE. A series of testosterone (0, 50, 100, 300, 500, 700, 900 μM) were added to the above CrpE reaction mixtures with Cr-538 as substrate, which were incubated at room temperature for 2 hours. The reactions were terminated with 2 volumes of methanol and centrifuged at top speed of centrifuge for 5 min before subjecting 250 μl to HPLC analysis as above. The amount of product, Cr-539, was calculated with the same method in the previous report [40]. All experiments were performed in duplicated.

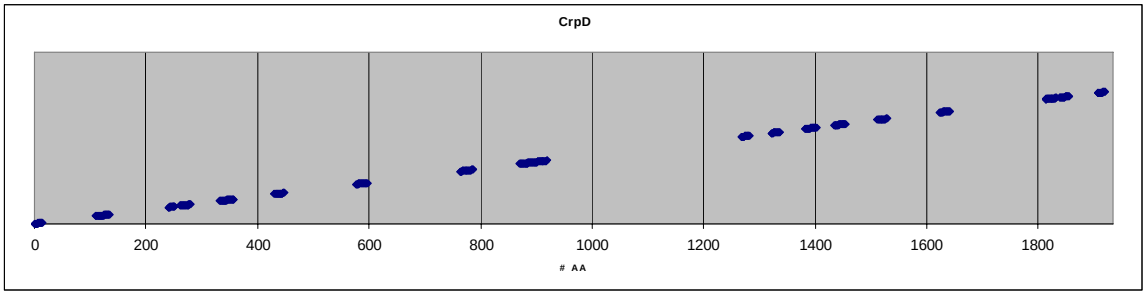
PikC and MycG, two P450 hydroxylases in macrocyclic natural products pikromycin and mycinamicin biosyntheses, respectively, were selected to investigate testosterone effect on their catalytic conversion [53, 54]. Their 100-μl reaction mixtures contained 0.5 mM testosterone, 100 μg/ml Fer, 0.1 unit/ml FNR, 2.5 mM of freshly prepared NADPH, 1 μM P450, and 0.5 mM narbomycin for PikC or 0.5 mM mycinamicin-IV for MycG in 50 mM NaH_2PO_4, pH 7.3 with 1 mM EDTA, 0.2 mM dithioerythritol, and 10% glycerol. The reactions were incubated at 30 °C for 40 min and the products were extracted with an equal volume of chloroform twice. After chloroform evaporation, the residues were dissolved in 300 μl methanol. For HPLC analysis, 250 μl solutions were subjected with previously described methods [53, 54]. All experiments were performed in duplicated.

2.5. Supplementary Figures and Tables

A.



B.



C.

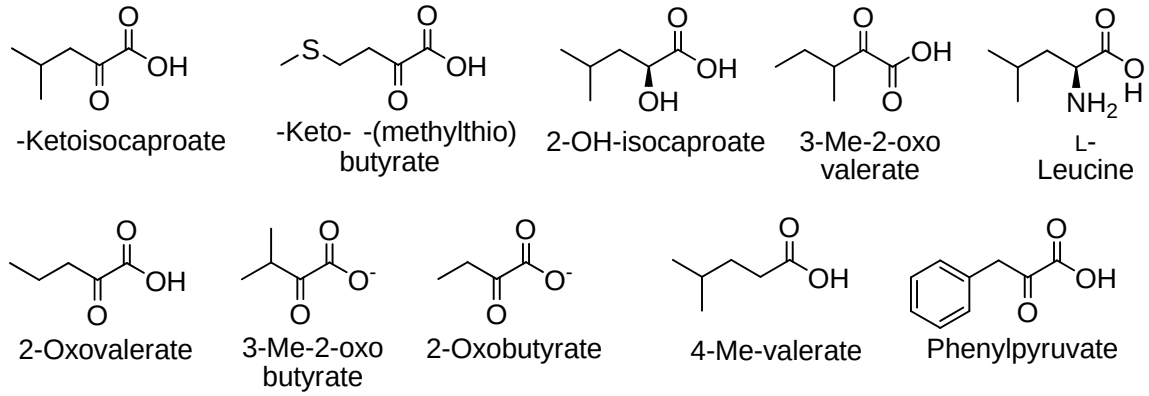


Figure S2-1. A: 4-12 % SDS-PAGE analysis of N-terminally His-tagged CrpD-M2 after Ni-NTA resin. The enzyme purity was roughly about 80 %. **B:** FT-ICR-MS analysis of trypsin-digested CrpD-M2 oligopeptides. **C:** Ten chemicals for examination of CrpD-M2 A domain substrate flexibility.

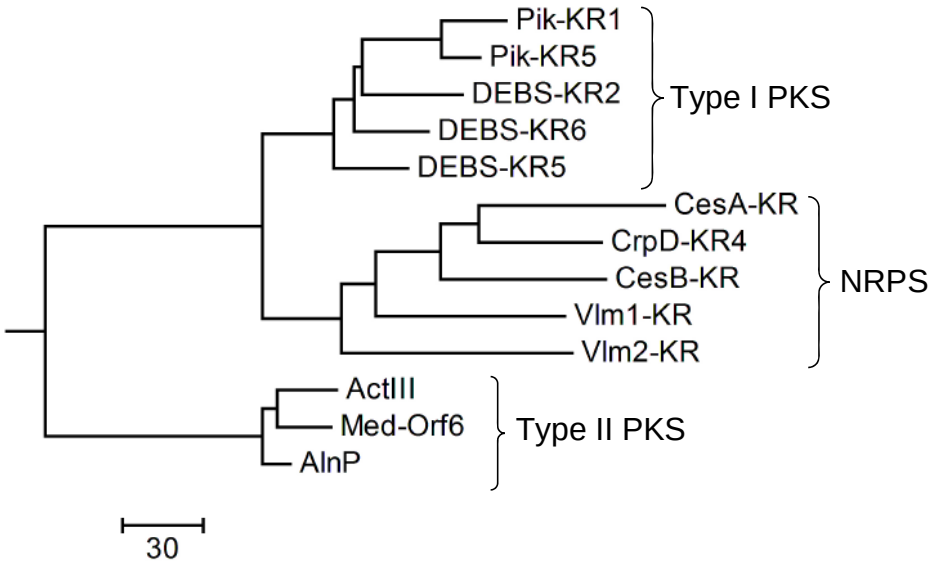


Figure S2-2. Phylogenetic analysis of CrpD-M2 KR domain. The CrpD-M2 KR domain was grouped with other NRPS KR domains, which was separated from others from both type I and type II PKS KR domains. The tree scale was shown as the neighboring joining identity percentage.

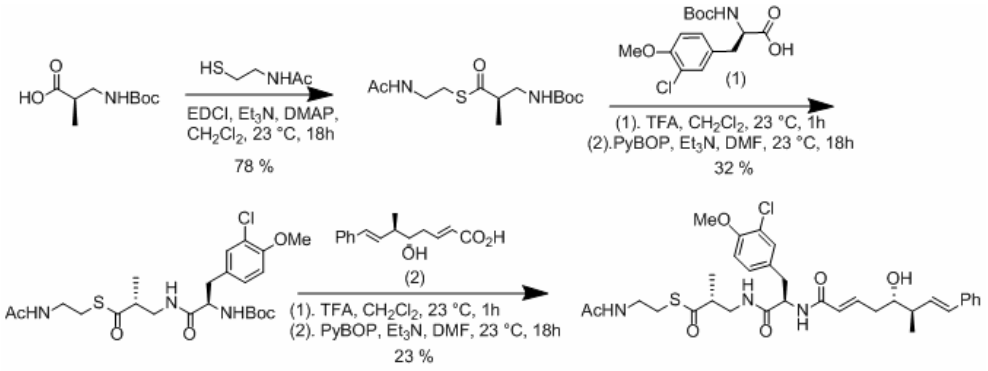


Figure S2-3. A synthetic scheme for cryptophycin Unit ABC-NAC ester.

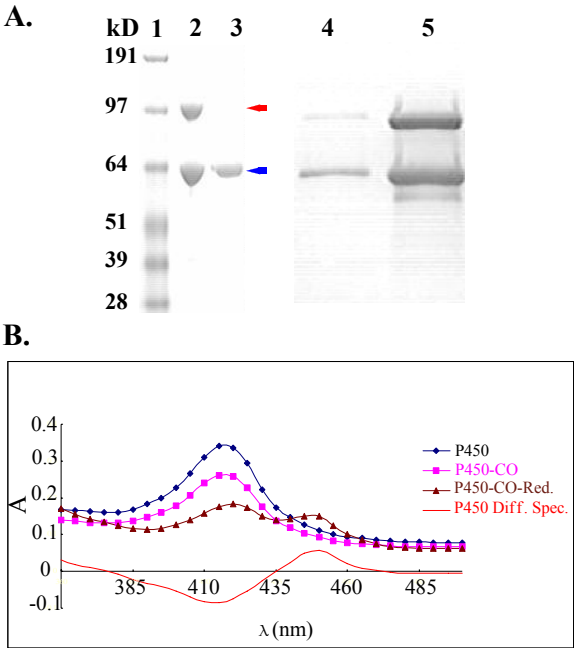


Figure S2-4. A: SDS-PAGE analysis of MBP-CrpE. Lane 1, molecular marker; lane 2, purified MBP-CrpE; lane 3, authentic GroEL; lane 4, MBP-CrpE treated with ATP; lane 5, MBP-CrpE treated with ATP and denatured bacterial lysate. The red arrow indicates the MBP-CrpE enzyme while the blue arrow shows the position of contaminant. B: Spectral analysis of purified MBP-CrpE. The blue trace represents the absorbance spectrum of purified MBP-CrpE in storage solution, which has a peak at 420 nm. The pink trace denotes the spectrum of CO-saturated MBP-CrpE solution. The maroon trace represents the spectrum of the CO-saturated MBP-CrpE solution reduced by $Na_2S_2O_4$. The spectrum of reduced CO-difference spectrum is shown by the red trace, which has a peak at 450 nm.

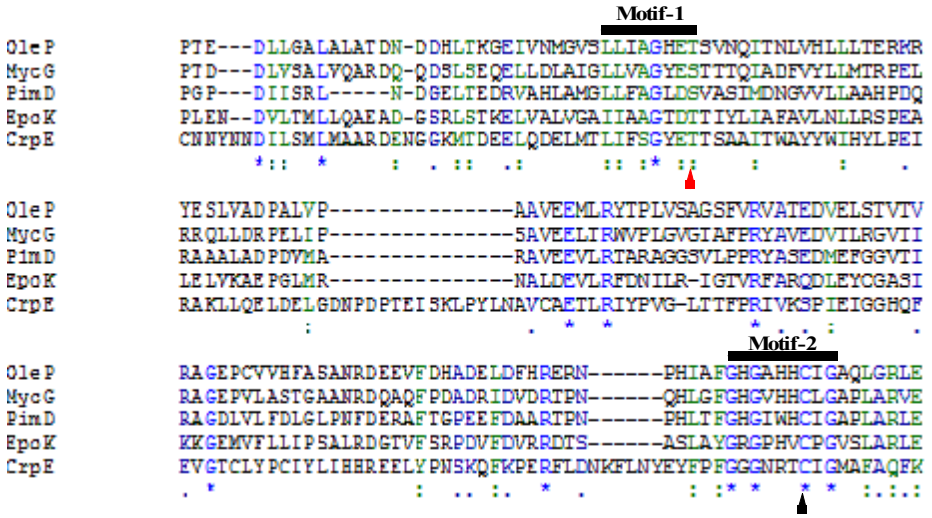
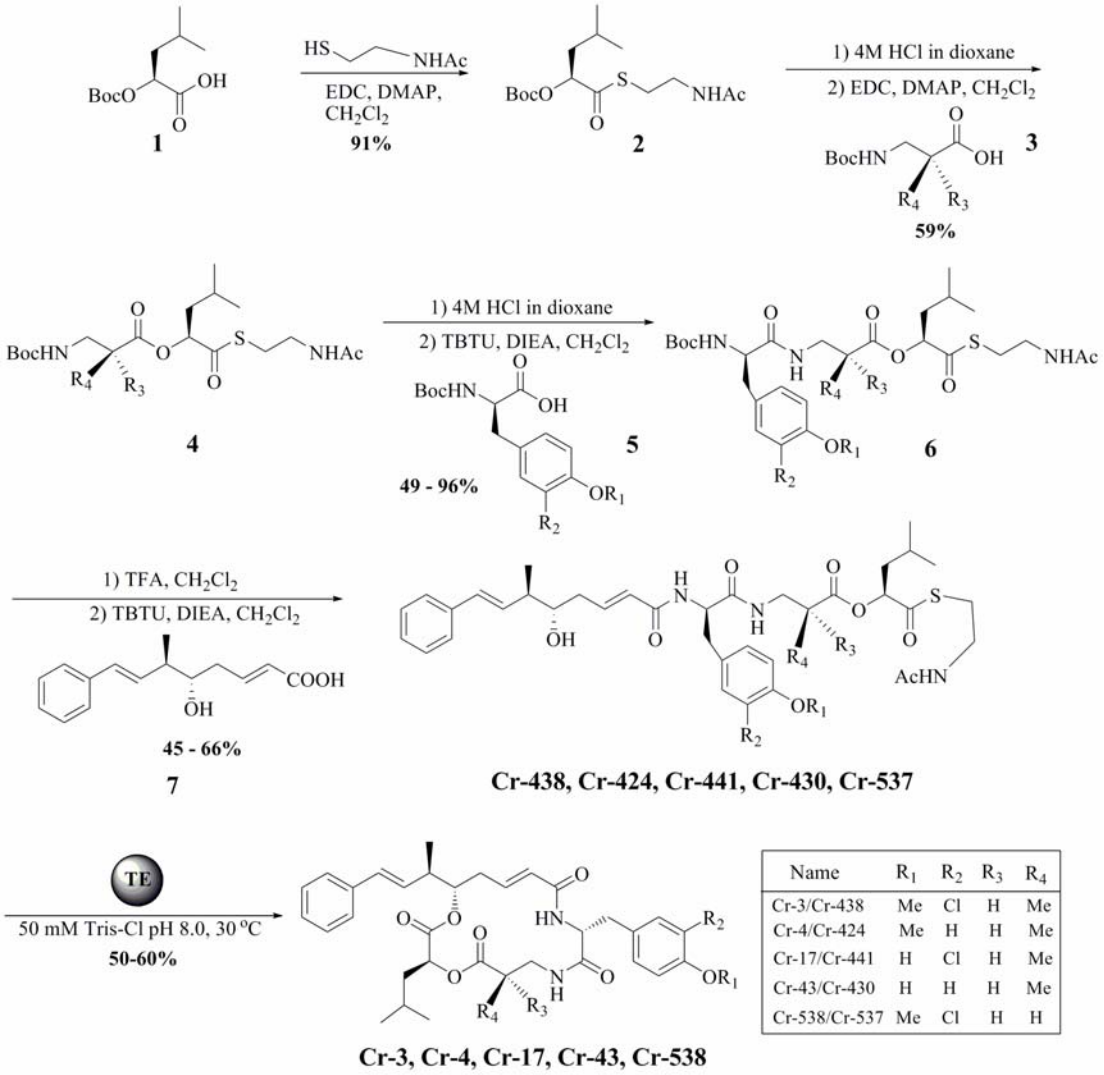


Figure S2-5. Multiple sequence alignment of CrpE with other P450s involved in natural product biosynthesis. CrpE shares two conserved motifs with other P450 epoxidases. The T/S residue in motif-1 is possibly responsible for O_2 binding and cleavage while the invariant C in motif-2 is likely responsible for heme binding.



Name	R_1	R_2	R_3	R_4
Cr-3/Cr-438	Me	Cl	H	Me
Cr-4/Cr-424	Me	H	H	Me
Cr-17/Cr-441	H	Cl	H	Me
Cr-43/Cr-430	H	H	H	Me
Cr-538/Cr-537	Me	Cl	H	H

Figure S2-6. Chemoenzymatic approach used to generate MBP-CrpE substrates.

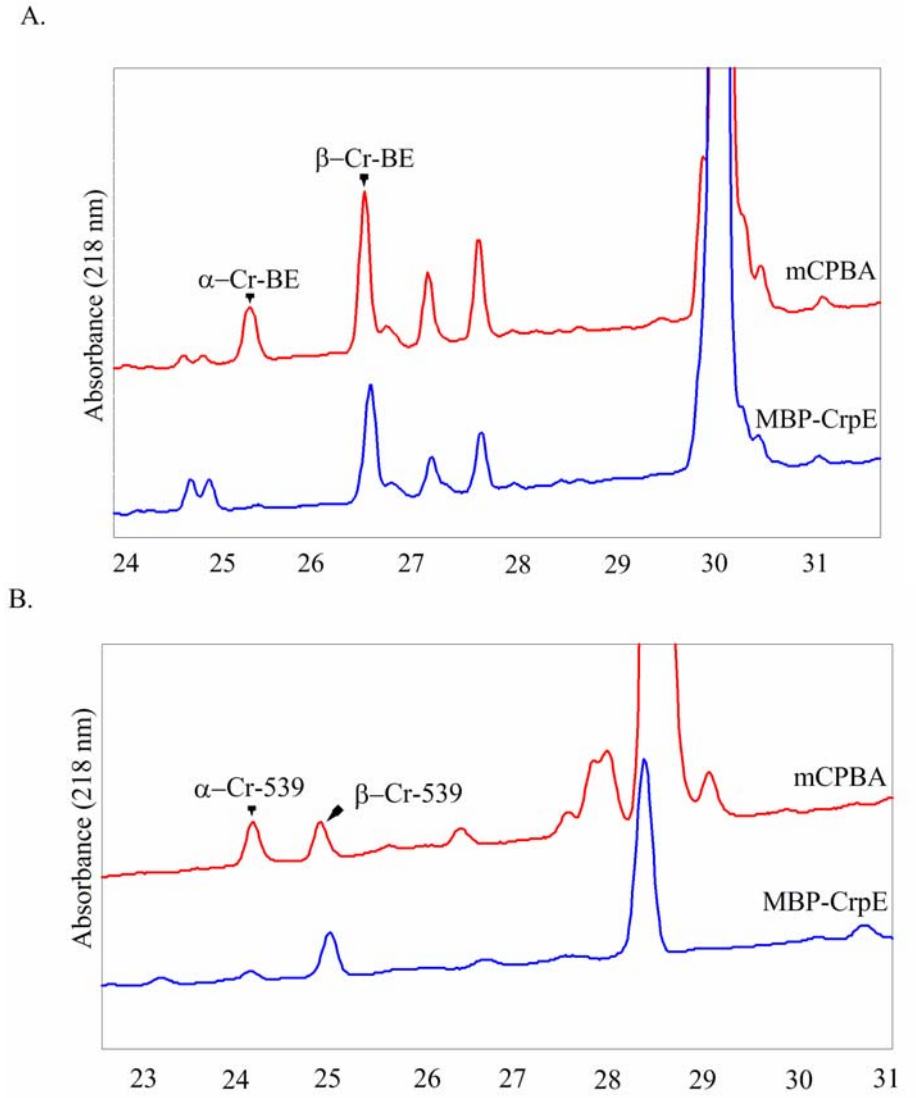


Figure S2-7. HPLC analysis of mCPBA reactions with Cr-B (A) and Cr-538 (B) as substrates. Both the α- and β-epoxide containing products were produced in chemical reactions utilizing mCPBA, whereas only the β-epoxide containing products were generated in enzymatic reactions using MBP-CrpE as catalyst.

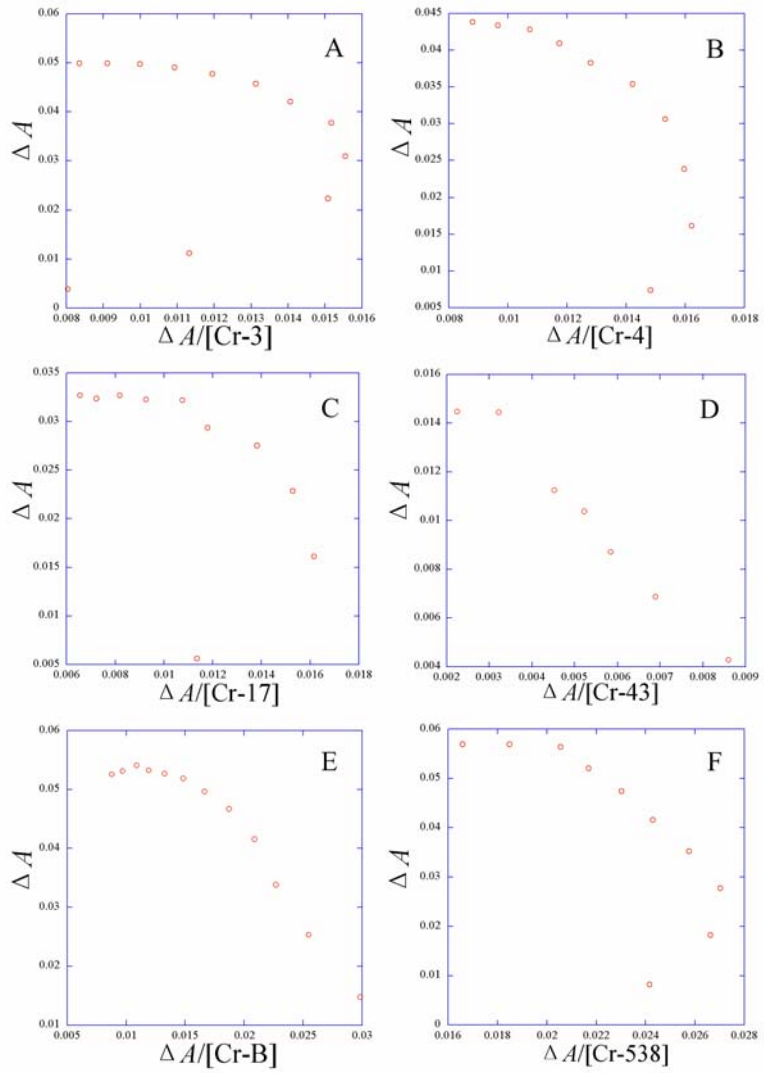


Figure S2-8. Putative cooperativity in substrate binding to MBP-CrpE. The data were fit to the Eadie-Hofstee equation.

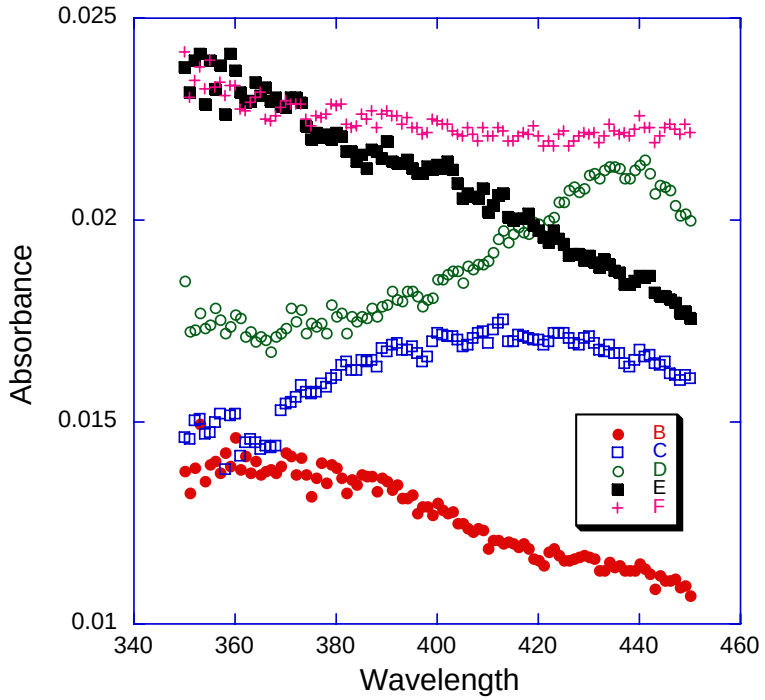


Figure S2-9. Absorbance change caused by addition of different concentrations of Cr-538. B, C, D, E, and F represent 1 μM, 2 μM, 3 μM, 4 μM, and 5 μM, respectively. 1 μM GroEL was used as a control in the assay. No meaningful absorbance change was caused by the addition of Cr-538 to GroEL.

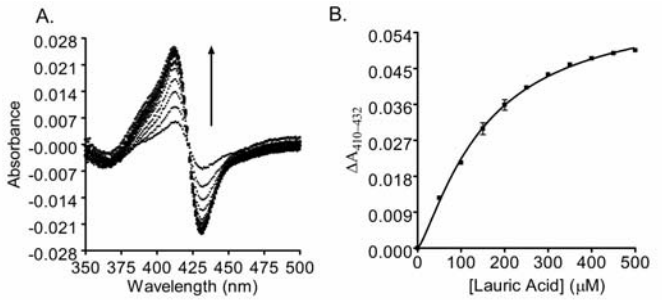


Figure S2-10. Lauric acid binding to CrpE. A: The difference spectra induced by different concentrations of lauric acid (0-500 μM). The direction of spectral shift upon substrate addition is shown (arrows). B: Plot of absorbance change vs ligand concentration. Absorbance changes were determined by subtracting A_{410} with A_{432} and were fitted to a hyperbolic binding equation. Similar to lauric acid, C_{14}, C_{16}, and C_{18} linear acids also induced the similar spectral shifts and the binding constants were summarized in Table 2.

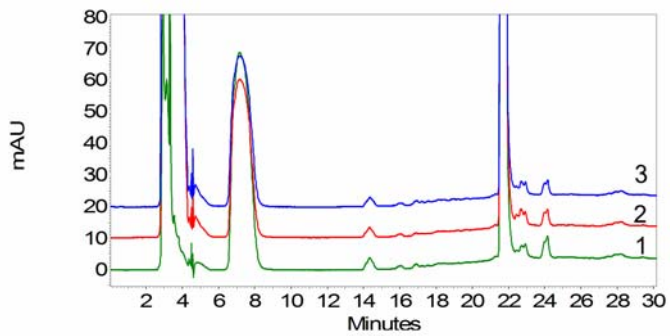


Figure S2-11. HPLC analysis of CrpE reaction with testosterone as substrate. Lane 1, authentic testosterone; lane 2, negative control reaction with CrpE boiled at 100 °C for 10 min as the enzyme source; lane 3, complete enzyme reaction. Testosterone was eluted at 21.7 min but no additional peak was found in the complete enzyme reaction compared to negative control, indicating that testosterone is not the enzyme catalytic substrate.

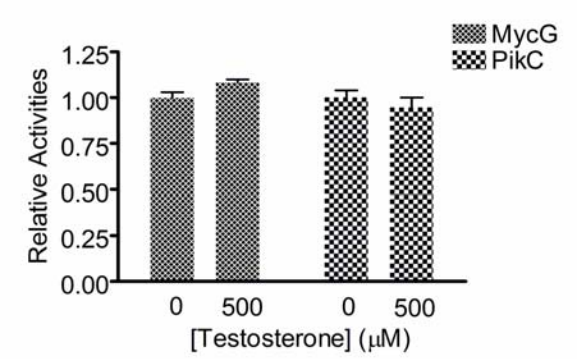


Figure S2-12. Testosterone effect on reactions of bacterial P450s, MycG and PikC. In contrast to CrpE reactions, the presence of testosterone didn't cause obvious effect on these two bacterial biosynthetic P450 catalytic abilities.

Table S2-1: Specificity codes for A domains from CrpD-M2, one enniatin NRPS, two cereulide NRPSs, two valinomycin NRPSs, one bacillaene NRPS, and two barbamide NRPSs.

A Domain	235	236	239	278	299	301	322	330	331	517	Specificity
CrpDM2-A	V	A	I	F	L	G	S	S	G	K	HIC
ENSYN-A1	G	A	L	H	V	V	G	I	C	K	D-HIV
CseA-A1	V	G	V	W	V	G	T	S	G	K	2-KIC
CseB-A1	V	G	F	W	V	A	V	S	D	K	2-KIV
Vlm1-A1	A	A	L	W	I	A	V	S	G	K	2-KIV
Vlm2-A1	V	V	I	W	I	A	E	N	M	K	Pyruvate
PksJ-A1	V	G	W	T	T	A	A	I	C	K	2-KIC
BarE-A	V	G	I	L	V	G	G	T	S	K	Trichloro-2-KIC
BarD	**D**	A	I	L	L	G	G	A	A	K	L-Leucine

Table S2-2. 1H NMR data for Unit A of Cr-1, Cr-38, and enzyme product with Cr-4 as substrate in $CDCl_3$*

Position	Enzyme Product Unit A	Cr-1 Unit A[57]	Cr-38 Unit A[58]
2	5.73 (d; 15.4)	5.74 (dt; 15.5, 0.9)	5.80 (d; 15.4)
3	6.73 (ddd; 15.1, 10.1, 4.8)	6.68 (ddd; 15.5, 9.6, 5.2)	6.68 (ddd; 15.4, 9.9, 5.4)
4 (*proS*)	2.40 (ddd; 14.5, 10.8, 10.5)	2.45 (ddd; 14.2, 11.1, 9.6)	2.65 (dt; 14.5, 9.9, 11.2)
4 (*proR*)	2.58 (dm; 14.5)	2.55 (brdd; 14.2, 5.2)	2.55 (brdd; 14.5, 5.4)
5	5.21 (ddd; 9.5, 4.8, 1.6)	5.16 (ddd; 11.1, 4.9, 1.9)	5.12 (ddd; 11.2, 5.0, 1.6)
6	1.73 (m)	1.80 (m)	1.76 (pd, 7-7.8, 5.0)
6-Me	1.17 (d; 6.9)	1.14 (d; 7.1)	1.03 (d; 7.0)
7	2.94 (dd; 7.5, 1.8)	2.92 (dd; 7.5, 2.0)	2.88 (dd; 7.8, 2.0)
8	3.70 (d; 1.8)	3.69 (d; 2.0)	3.58 (d; 2.0)
10/14	7.24-7.28 (m)	7.25 (m)	7.22 (m)
11/12/13	7.35-7.42 (m)	7.34-7.39 (m)	7.28-7.36 (m)

*: (1) δ_H (multiple; *J* in Hz);
(2) The chemical shifts of H-7 and H-8 indicates the stereochemistry of epoxide in these products. Both Cr-1 and enzyme product contain β-epoxide while Cr-38 contains α-epoxide.

2.6. References

1. Golakoti, T.; Yoshida, W. Y.; Chaganty, S.; Moore, R. E., *J Nat Prod* **2001,** 64, (1), 54-9.
2. Schwartz, R. E.; Hirsch, C. F.; Sesin, D. F.; Flor, J. E.; Chartrain, M.; Fromtling, R. E.; Harris, G. H.; Salvatore, M. J.; Liesch, J. M.; Yudin, K., *Journal of Industrial Microbiology* **1990,** 5, (2-3), 113-123.
3. Mitra, A.; Sept, D., *Biochemistry* **2004,** 43, (44), 13955-62.
4. Hamel, E., *Biopolymers* **2002,** 66, (3), 142-60.
5. Herbst, R. S.; Khuri, F. R., *Cancer Treat Rev* **2003,** 29, (5), 407-15.
6. Gupta, S.; Bhattacharyya, B., *Mol Cell Biochem* **2003,** 253, (1-2), 41-7.
7. Borst, P.; Evers, R.; Kool, M.; Wijnholds, J., *J Natl Cancer Inst* **2000,** 92, (16), 1295-302.
8. Gottesman, M. M.; Pastan, I., *Annu Rev Biochem* **1993,** 62, 385-427.
9. Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H., *ACS Chem Biol* **2006,** 1, (12), 766-79.
10. Beck, Z. Q.; Burr, D. A.; Sherman, D. H., *Chembiochem* **2007,** 8, (12), 1373-1375.
11. Eggen, M.; Georg, G. I., *Med Res Rev* **2002,** 22, (2), 85-101.
12. Beck, Z. Q.; Aldrich, C. C.; Magarvey, N. A.; Georg, G. I.; Sherman, D. H., *Biochemistry* **2005,** 44, (41), 13457-66.
13. Seufert, W.; Beck, Z. Q.; Sherman, D. H., *Angew Chem Int Ed Engl* **2007,** 46, (48), 9298-300.
14. Liang, J.; Hoard, D. W.; Van Khau, V.; Martinelli, M. J.; Moher, E. D.; Moore, R. E.; Tius, M. A., *Journal of Organic Chemistry* **1999,** 64, (5), 1459-1463.
15. Barrow, R. A.; Hemscheidt, T.; Liang, J.; Paik, S.; Moore, R. E.; Tius, M. A., *J Am Chem Soc* **1995,** 117, (9), 2479-2490.
16. Gardinier, K. M.; Leahy, J. W., *J Org Chem* **1997,** 62, (21), 7098-7099.
17. Chang, H. T.; Sharpless, K. B., *J Org Chem* **1996,** 61, (18), 6456-6457.
18. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., *Science* **2001,** 291, (5509), 1790-2.
19. Calderone, C. T.; Bumpus, S. B.; Kelleher, N. L.; Walsh, C. T.; Magarvey, N. A., *Proc Natl Acad Sci U S A* **2008,** 105, (35), 12809-14.
20. Chang, Z.; Flatt, P.; Gerwick, W. H.; Nguyen, V. A.; Willis, C. L.; Sherman, D. H., *Gene* **2002,** 296, (1-2), 235-47.
21. Magarvey, N. A.; Ehling-Schulz, M.; Walsh, C. T., *J Am Chem Soc* **2006,** 128, (33), 10698-9.
22. Dorrestein, P. C.; Kelleher, N. L., *Natural product reports* **2006,** 23, (6), 893.
23. Gu, L.; Geders, T. W.; Wang, B.; Gerwick, W. H.; Hakansson, K.; Smith, J. L.; Sherman, D. H., *Science* **2007,** 318, (5852), 970.
24. Gu, L.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Håkansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H., *Nature* **2009,** 459, (7247), 731.
25. Leslie, M. H.; Hicks, *ACS chemical biology* **2006,** 1, (2), 93.
26. Kelleher, N. L.; Hicks, L. M., *Current Opinion in Chemical Biology* **2005,** 9, (5), 424-430.

27. McLoughlin, S. M.; Kelleher, N. L., *J Am Chem Soc* **2005,** 127, (43), 14984.
28. Lin, S.; Van Lanen, S. G.; Shen, B., *Proc Natl Acad Sci U S A* **2009,** 106, (11), 4183-8.
29. Zaleta-Rivera, K.; Xu, C.; Yu, F.; Butchko, R. A.; Proctor, R. H.; Hidalgo-Lara, M. E.; Raza, A.; Dussault, P. H.; Du, L., *Biochemistry* **2006,** 45, (8), 2561-9.
30. Nishihara, K.; Kanemori, M.; Yanagi, H.; Yura, T., *Appl Environ Microbiol* **2000,** 66, (3), 884-9.
31. Rye, H. S.; Burston, S. G.; Fenton, W. A.; Beechem, J. M.; Xu, Z.; Sigler, P. B.; Horwich, A. L., *Nature* **1997,** 388, (6644), 792-8.
32. Viitanen, P. V.; Gatenby, A. A.; Lorimer, G. H., *Protein Sci* **1992,** 1, (3), 363-9.
33. Mendes, M. V.; Anton, N.; Martin, J. F.; Aparicio, J. F., *Biochem J* **2005,** 386, (Pt 1), 57-62.
34. Ogura, H.; Nishida, C. R.; Hoch, U. R.; Perera, R.; Dawson, J. H.; Ortiz de Montellano, P. R., *Biochemistry* **2004,** 43, (46), 14712-21.
35. Anzai, Y.; Saito, N.; Tanaka, M.; Kinoshita, K.; Koyama, Y.; Kato, F., *FEMS Microbiol Lett* **2003,** 218, (1), 135-41.
36. Shah, S.; Xue, Q.; Tang, L.; Carney, J. R.; Betlach, M.; McDaniel, R., *J Antibiot (Tokyo)* **2000,** 53, (5), 502-8.
37. Rodriguez, A. M.; Olano, C.; Mendez, C.; Hutchinson, C. R.; Salas, J. A., *FEMS Microbiol Lett* **1995,** 127, (1-2), 117-20.
38. Andersen, J. F.; Tatsuta, K.; Gunji, H.; Ishiyama, T.; Hutchinson, C. R., *Biochemistry* **1993,** 32, (8), 1905-13.
39. Ghosh, A. K.; Swanson, L., *J Org Chem* **2003,** 68, (25), 9823-6.
40. Ding, Y.; Seufert, W. H.; Beck, Z. Q.; Sherman, D. H., *J Am Chem Soc* **2008,** 130, (16), 5492-8.
41. Ouellet, H.; Podust, L. M.; de Montellano, P. R., *J Biol Chem* **2008,** 283, (8), 5069-80.
42. Zhao, Y.; White, M. A.; Muralidhara, B. K.; Sun, L.; Halpert, J. R.; Stout, C. D., *J Biol Chem* **2006,** 281, (9), 5973-81.
43. Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M., *J Biol Chem* **2006,** 281, (11), 7614-22.
44. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., *Science* **2004,** 305, (5684), 683-6.
45. Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D.; Jhoti, H., *Nature* **2003,** 424, (6947), 464-8.
46. Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D., *Proc Natl Acad Sci U S A* **2003,** 100, (23), 13196-201.
47. Job, P., *Ann. Chim.* **1928,** 9, 113-203.
48. Koo, L. S.; Tschirret-Guth, R. A.; Straub, W. E.; Moenne-Loccoz, P.; Loehr, T. M.; Ortiz de Montellano, P. R., *J Biol Chem* **2000,** 275, (19), 14112-23.
49. Khan, K. K.; He, Y. A.; He, Y. Q.; Halpert, J. R., *Chem Res Toxicol* **2002,** 15, (6), 843-53.
50. Harlow, G. R.; Halpert, J. R., *Proc Natl Acad Sci U S A* **1998,** 95, (12), 6636-41.
51. Khan, K. K.; Liu, H.; Halpert, J. R., *Drug Metab Dispos* **2003,** 31, (4), 356-9.
52. Zhou, S. F., *Curr Drug Metab* **2008,** 9, (4), 310-22.

53. Li, S.; Ouellet, H.; Sherman, D. H.; Podust, L. M., *J Biol Chem* **2009,** 284, (9), 5723-30.
54. Anzai, Y.; Li, S.; Chaulagain, M. R.; Kinoshita, K.; Kato, F.; Montgomery, J.; Sherman, D. H., *Chem Biol* **2008,** 15, (9), 950-9.
55. Omura, T.; Sato, R., *J Biol Chem* **1964,** 239, 2379-85.
56. Meluzzi, D.; Zheng, W. H.; Hensler, M.; Nizet, V.; Dorrestein, P. C., *Bioorganic & medicinal chemistry letters* **2008,** 18, (10), 3107.
57. Golakoti, T.; Ogino, J.; Heltzel, C. E.; Le Husebo, T.; Jensen, C. M.; Larsen, L. K.; Patterson, G. M. L.; Moore, R. E.; Mooberry, S. L.; Corbett, T. H.; Valeriote, F. A., *J. Am. Chem. Soc.* **1995,** 117, (49), 12030-12049.
58. Chaganty, S.; Golakoti, T.; Heltzel, C.; Moore, R. E.; Yoshida, W. Y., *J Nat Prod* **2004,** 67, (8), 1403-6.

Notes:

Yousong Ding and David H. Sherman designed all experiments. Yousong Ding performed the experiments.

Dr. Wolfgang H. Seufert and Kyle Bolduc chemically prepared the substrates for CrpD-M2 and CrpE studies. Chris Rath performed all LC-FTICR-MS analyses.

Chapter 3

Chemical and Biochemical Characterization of Unique Prenylated Indole Alkaloid Biosynthesis

3.1. Summary

Four consecutive studies are included in this chapter to investigate the biosynthetic pathways of one group of unique prenylated fungal indole alkaloids containing a characteristic bicyclo[2.2.2]diazaoctane core. The first study focused on identification of the secondary metabolites VM55599 and pre-paraherquamide by LC-MS^n analysis as natural metabolites in cultures of *Penicillium fellutanum* and *Aspergillus japonicus*. The identification of both metabolites, which have a diastereomeric relationship, provides indirect support for a previously proposed unified biogenetic hypothesis. The second study continued my efforts to identify and characterize key biosynthetic intermediates in the prenylated fungal indole alkaloid biosynthesis. An advanced metabolite, pre-malbrancheamide, involved in the biosynthesis of malbrancheamide, was synthesized as a double ^{13}C-labeled form and was incorporated into the indole alkaloid malbrancheamide B by *Malbranchea aurantiaca*. In addition, pre-malbrancheamide has been detected as a natural metabolite in cultures of *M. aurantiaca*. Besides the chemical characterization, my efforts to biochemically investigate the fungal alkaloid biosynthesis led to the last two studies. One prenyltransferase gene was isolated from genomic DNA of *M. aurentiaca*. Enzyme specificity was investigated with a series of amino acid substrates, revealing its function as a 4-dimethylallyltryptophan synthase. MaPT activity was not dependent on a divalent cation co-factor, although it was reversibly inactivated by 5 mM EDTA. Analysis of kinetic parameters showed reduced enzyme efficiency upon simple modification of L-Trp. Moreover, D-Trp had 0.5% relative activity and functioned as a competitive inhibitor with K_i = 40.41 μM. Finally, T105, D179, K189 and K261 in MaPT were serially

mutated and the resulting lesions displayed low or complete loss of activity. Finally, the whole genome sequencing of marine notoamide-producing *Aspergillus* MF-297-2 generated 480-Mb genome information. One putative stephacidin/notoamide biosynthetic gene cluster was identified by genome mining. Two putative prenyltransferase genes, two P450 genes, and one bi-module NRPS gene were predicted from the gene cluster. The cDNAs of two prenyltransferase genes were prepared, cloned, and overexpressed to produce two soluble polypeptides. Biochemical characterization of the second prenyltransferase (NotD) indicated this enzyme is a 6-deoxybrevianamide E synthase, one long-seeking enzyme in fungal alkaloid studies, and catalyzes a reverse prenylation reaction. Subsequent biochemical characterization included enzyme activity metal dependence, kinetic analysis, and substrate specificity. E108, R122, and W424 in NotD were serially mutated and the resulting lesions displayed low or nearly complete loss of activity. Moreover, the first prenyltransterase (NotA) exhibited considerable activities to keto-premalbrancheamide, premalbrancheamide, and pre-paraherquamide. These results indicated a putative biosynthetic pathway for stephacidin and notoamide.

3.2. Introduction

Fungi are phylogenetically diverse microorganisms that produce thousands of low-molecular-weight natural products. A significant number of these metabolites are in clinical use as human therapeutics, including antibacterial penicillins, cephalosporins, immunosuppressive cyclosporine A, and cholesterol-lowering HMG-CoA reductase inhibitor "statins" [1]. However, drug-resistant bacteria like methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus faecalis* continue to emerge as a threat to human health [2-4]. Along with medical needs to treat a spectrum of diseases, an increasing number of natural products have been isolated from fungal sources and screened for potential bioactive secondary metabolites [5].

Numerous secondary metabolic indole alkaloids and related nitrogenous substances are derived from tryptophan and isoprenic building blocks. The rich nucleophilic chemistry of the indole ring has been extensively exploited by Nature to create an array of structurally intriguing natural substances such as various ergot

alkaloids. A number of structurally interesting natural alkaloids have been isolated from various fungi that contain the unique bicyclo [2.2.2] ring system constituted mainly from tryptophan, proline or its derivatives, and dimethylallyl pyrophosphate (DMAPP) (**Figure 3-1**) [6]. These fungal metabolites include but not limit to brevianamides, paraherquamides, asperparalines, notoamides, stephacidins, and malbrancheamides [6-10]. Interestingly, these alkaloids contain various bioactivities such as allergens, calmodulin (CaM) antagonist, antinematodal agents, and anticancer agents, although they possibly stem from one common biosynthetic precursor. The brevianamides might be the first members of this family of indole alkaloids and are constructed from tryptophan, proline, and one isoprene unit (**Figure 3-1**). Brevianamide A was originally isolated from *Penicillium brevicompactum* [8]. The paraherquamides are comprised of two isoprene units, tryptophan and variously substituted proline derivatives, making them to be more complex in structure. Their potent anthelmintic and antinematodal activities stand them out as the next generation of veterinary medicine candidates to treat intestinal parasites [11]. Although the parent and most potent member, paraherquamide A, was isolated from cultures of *Penicillium* species, other members of paraherquamides have also been isolated from various both *Penicillium* and *Aspergillus* species. Similar to paraherquamides, the asperparalines also have paralytic effects on silkworms [12]. Structurally, these fungal metabolites contain a *spiro*-succinimide ring system, which is replaced as a *spiro*-oxindole in the brevianamides and paraherquamides. Asperparaline A was isolated from *Aspergillus japonicus* JV-23. The malbrancheamides have been recently isolated from *Malbranchea aurantiaca* collected from bat guano [9]. Malbrancheamide is the first alkaloid in this class of prenylated indole alkaloids to contain a halogenated indole ring, and it is further characterized by the lack of a tertiary amide in the bicyclo[2.2..2]diazaoctane core. In addition to these notable structural features, this fungal alkaloid has been shown to be a CaM antagonist that inhibits the activity of CaM-dependent phosphodiesterase (PDE1) in a concentration dependent manner. Recently, marine-derived fungi have proven to be rich sources of structurally novel and biologically active secondary metabolites, which are emerging as a significant resource for new chemicals in drug discovery [13]. Several structurally related alkaloids have been isolated from marine environments and shown to be anticancer agents.

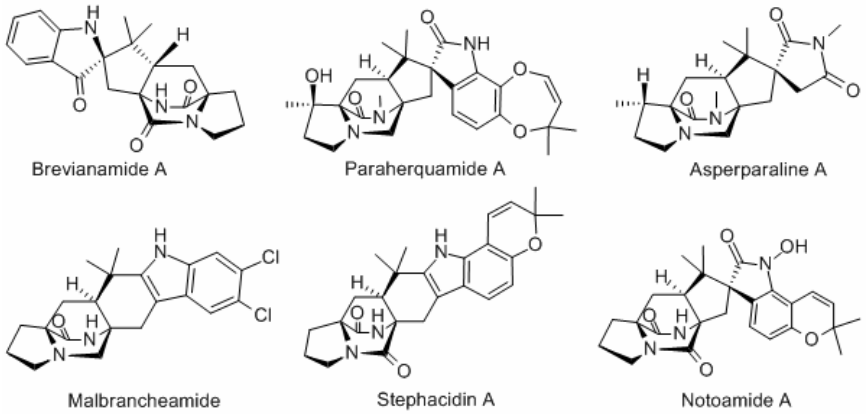


Figure 3-1. Structures of several prenylated fungal alkaloids. These fungal metabolites are isolated from various fungal strains but share a common bicyclo[2.2.2]diazaoctane core.

Stephacidin A was isolated from afungal strain *Aspergillus ochraceus* WC76466 and exhibited potent in vitro cytotoxicity against various human tumor cell lines through a novel mechanism of action different than those of p53, mdr, bcl2, tubulin, or topoisomerase II mediating pathways [10]. Notoamides were isolated from a culture of marine-derived fungus, *Aspergillus* MF297-2, which was found in the common mussel, *Mytilus edulis* [14]. Many of notoamides showed moderate anticancer abilities with a series of human tumor cell lines.

The interesting structures and promising bioactivities of these fungal alkaloids have attracted considerable efforts to investigate their biosynthetic scheme. In 1974, Baldas *et al.* observed the incorporation of radioactive Trp, Pro, DMAPP, and brevianamide F into brevianamide A, making them to propose one brevianamide A biosynthetic pathway (**Figure 3-2A**) [15]. Later, the Williams group also conducted a series of feeding experiments to explore the biosynthetic origin and pathway of paraherquamide [16-18]. These experiments strongly indicated that this family of prenylated indole alkaloids shares a common biosynthetic pathway, and their structural differences come from the decoration of pathway-specific tailoring enzymes. However, it is still unknown how Nature creates the characteristic bicyclo[2.2.2]diazaoctane core of these natural products, although an intramolecular hetero-Diels-Alder reaction of a 5-hydroxypyrazin-2(1H)-one has been long proposed (**Figure 3-2A**). In 2006, one dimodular NRPS gene was mined

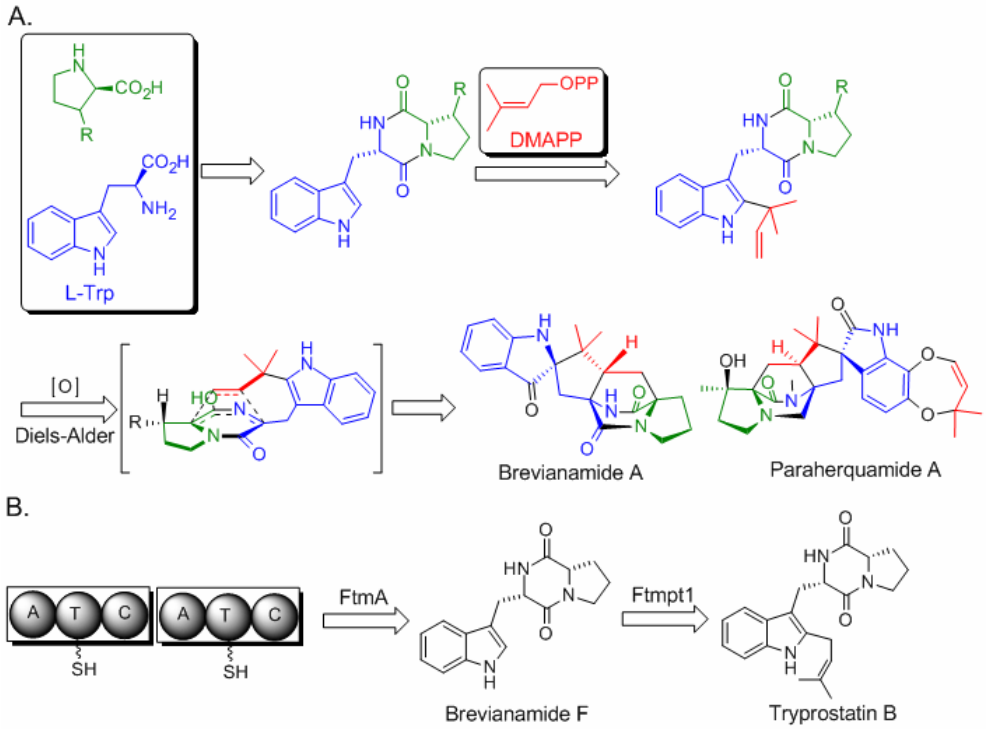


Figure 3-2. Putative biosynthetic pathways of brevianamide A, paraherquamide A, and ergot alkaloids. Feeding experiments and genetic and biochemical characterization of FtmA and FtmB (also termed as FtmPT1) led to the understanding of these pathways.

from *Aspergillus fumigatus* genome and heterologous expression of this gene led to the accumulation of brevianamide F, the first genetic evidence about the biosynthesis of fumitrmorgins and this family of prenylated indole alkaloids (**Figure 3-2B**) [19]. Furthermore, FtmB (previously named as FtmPT1) was derived from the *A. fumigatus* genome, and it catalyzed a regular C-2 prenylation of brevianamide F to produce tryprostatin B, providing additional evidence towards the understanding of ergot alkaloid biosynthesis [20]. However, one reverse prenyltransferse to produce deoxybrevianamide E from brevianamide F is desperately required to confirm the biosynthetic scheme of prenylated indole alkaloids with bicyclo[2.2.2]diazaoctane core. In this chapter, I will describe my studies to isolate key biosynthetic intermediates in paraherquamide, asperparaline, and malbrancheamide biosyntheses. Moreover, one regular prenyltransferase was isolated from *M. aurantica* and characterized to be a dimethylallyl tryptophan synthase (DMATS). Finally, one putative notoamide gene cluster was identified through genome mining of the *Aspergillus* MF297-2 genome database, and the long-sought deoxybrevianamide E synthase was characterized in detail.

3.3. Results and Discussion

3.3.1. Isolation of VM55599 and Pre-paraherquamide from *A. japonicus* and *P. fellutanum*. Biosynthetic Implications

In my studies, *P. fellutanum* and *A. japonicus* JV-23 were cultured under the conditions described in the experiment section and were first interrogated for viable production of paraherquamide A, asperparaline A, and their derivatives, respectively. Authentic paraherquamide A was first analyzed by LC-MS^n and its MS, MS^2, MS^3, and MS^4 spectra were informative to identify this alkaloid, establishing the successful application of LC-MS^n analysis in this study (**Figure S3-1**). Next, paraherquamide A in the *P. fellutanum* isolation was eluted at 14.47 min in selective ion monitoring (SIM) chromatograph and exhibited an ion at *m/z* 494.36 (calculated $[M+H]^+$: 494.26) in MS analysis (**Figure 3-3C** and **Figure 3-4A**). Further MS^2, MS^3, and MS^4 analyses produced the identical spectra to those of authentic paraherquamide A (**Figure 3-4A** and **Figure S3-1**). Along with paraherquamide A, paraherquamides B-G have previously been isolated from *P. fellutanum* (previously named *P. charlesii*) [6]. In this study, one compound at *m/z* 464.29 was detected in the fungal isolation and had the retention time of 14.95 min (**Figure 3-3C** and **Figure 3-4B**). This compound was identified as paraherquamide B (calculated $[M+H]^+$: 464.25) by comparing its retention time and MS^2 spectrum to those of an authentic specimen (**Figure 3-3** and **Figure S3-1**).

In extracts from *A. japonicus* JV-23 cultures, asperparaline A had the retention time of 7.58 min and exhibited an ion at *m/z* 360.29 (calculated $[M+H]^+$: 360.22) (**Figure 3-3D** and **Figure 3-4C**). This metabolite was further analyzed with MS^2 analysis. In previous reports, *Aspergillus* species IMI 337664 and *A. sclerotiorum* represented the first organisms outside of *Penicillium sp.* to produce paraherquamide congeners [21, 22]. In this study, I also investigated paraherquamide production in *A. japonicus* JV-23. By comparing their retention times and MS^n spectra to those of authentic compounds, both paraherquamide A (14.45 min) and paraherquamide B (14.88 min) were identified in this *Aspergillus* sp. isolation, further indicating that both the *Penicillium sp.* and the

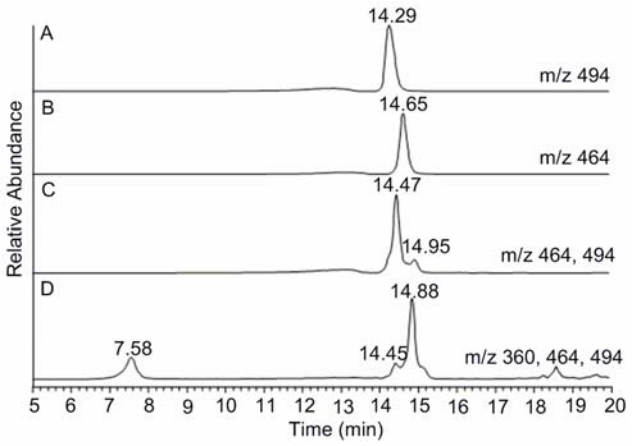


Figure 3-3. Selective ion monitoring (SIM) chromatographs corresponding to the authentic paraherquamide A (*m/z* 494) (A), authentic paraherquamide B (*m/z* 464) (B), isolation from *P. fellutanum* culture (*m/z* 464 and 494) (C), and isolation from *A. japonicus* JV-23 (*m/z* 360, 464, and 494) (D).

Aspergillus sp. are able to produce these anthelmintic alkaloid metabolites (**Figure 3-3,** and **Figure S3-2**). Moreover, this result strongly suggested one common biosynthetic pathway shared by both asperparalines and paraherquamides in this fungus.

The *A. japonicus* JV-23 strain produced more paraherquamide B than paraherquamide A in PDB medium under the above growth conditions. *A. japonicus* JV-23 is the first reported *Aspergillus* sp. to produce paraherquamide A itself to the best of our knowledge, although Everett and co-workers isolated the paraherquamide congeners VM54159, SB203105 and SB200437 from *Aspergillus* strain IMI 337664 [21]. In extracts from *P. fellutanum*, two metabolites with *m/z* of 350 were separated and identified by LC-MS/MS analysis (**Figure 3-5**). The first peak had the retention time of 12.15 min while the second metabolite was eluted at 12.64 min. These peaks were initially proposed to be VM55599 and pre-paraherquamide, considering their theoretical molecular weights ($C_{22}H_{27}N_3O$, 349.22) and the previous isolation of VM55599 as the minor metabolite from *Penicillium* sp. IMI337664 [23].

To further identify these metabolites, synthetic and authentic (racemic) samples of VM55599 and pre-paraherquamide were used to secure standard MS and MS^2 fragmentation data (**Figure S3-3**). Interestingly, these two compounds, which are

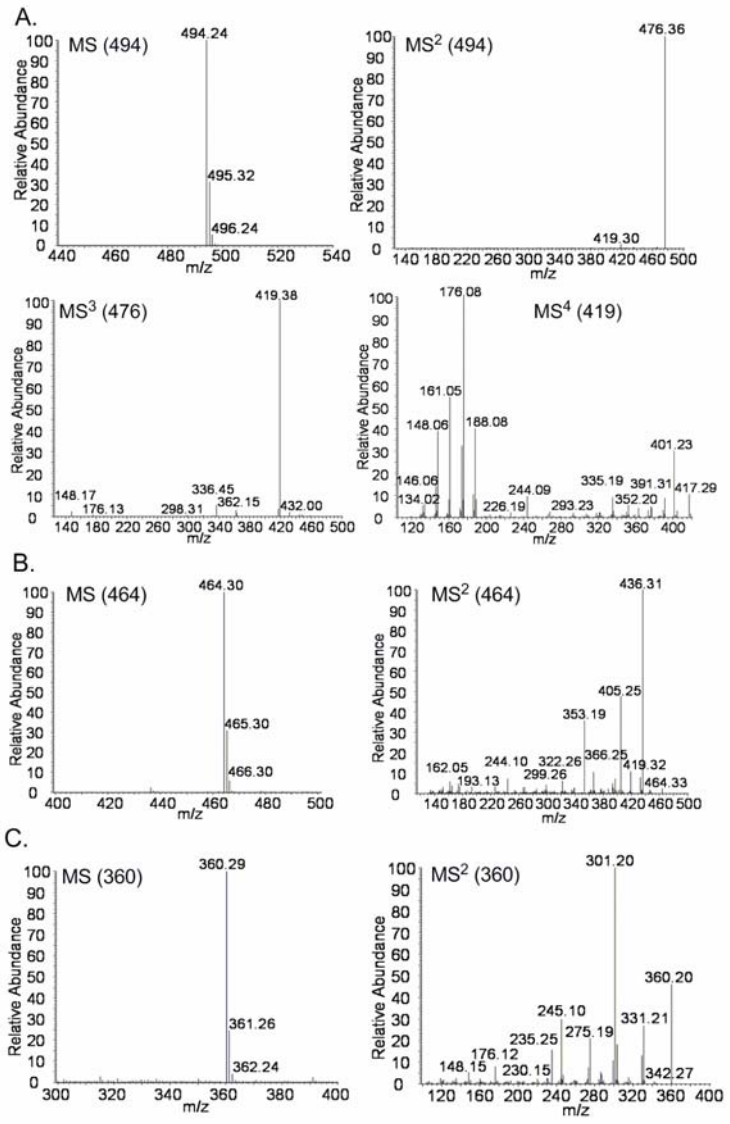


Figure 3-4. MS^n spectra of paraherquamide A (1) (A) and paraherquamide B (B) from the isolation of *P. fellutanum* cultures, and asperparaline A (2) (C) from the extracts of *A. japonicus* JV-23. The integral *m/z* values of ions for each MS^n analysis are included in corresponding graphs.

diastereomers, exhibited different fragmentation patterns in their MS^2 spectra. The ratio of the peak at *m/z* 322.26 to the peak at *m/z* 305.25 in VM55599 MS^2 spectrum was larger than 1 while this ratio was significantly smaller than 1 in the pre- paraherquamide MS^2 spectrum with the fragment at *m/z* 305.24 as the most intense peak, which likely serves as the distinctive feature of these two compounds' MS^2 spectra (**Figure S3-3**). CO was lost from the peptide bond of parent ions to produce the first fragment at *m/z*

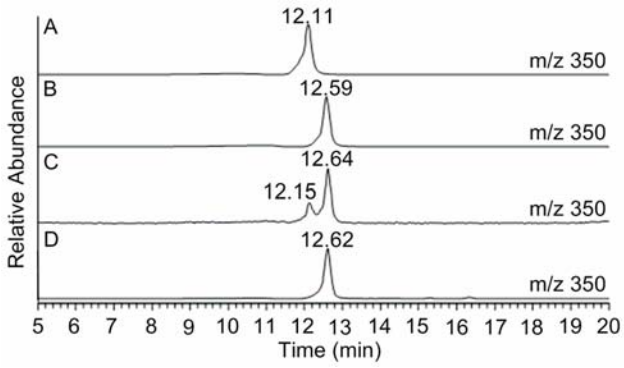


Figure 3-5. Selective ion monitoring chromatographs corresponding to the LC-MS^n analysis of authentic VM55599 (*m/z* 350) (A), authentic pre-paraherquamide (*m/z* 350) (B), isolation from *P. fellutanum* cultures (*m/z* 350) (C), and isolation from *A. japonicus* JV-23 (*m/z* 350) (D).

322.26, which was further fragmented to give a signal at *m/z* 305.25 by losing NH_3 (**Figure 3-6** and **Figure S3-3**). The same fragmentation pathway was observed for pre-paraherquamide. The fragmentation discrepancy observed in VM55599 and pre-paraherquamide MS^2 spectra were apparently affected by the single relative stereochemical difference at C14 (paraherquamide numbering) between these two compounds. Comparing their MS^2 spectra with those of authentic compounds, VM55599 (12.15 min) and pre-paraherquamide (12.64 min) were identified in the extract from *P. fellutanum* cultures (**Figure 3-5A-C** and **Figure 3-6A-B**). Pre-paraherquamide was thus observed as one natural metabolite, further strongly validating the putative pathway in **Figure 3-7**.

In extracts from *A. japonicus* JV-23 cultures, only one metabolite at *m/z* 350 was identified by LC-MS/MS analysis (**Figure 3-5D**). When compared to the MS and MS^2 spectra of authentic standards, the metabolite with the retention time of 12.62 min was validated to be pre-paraherquamide (**Figure 3-6C**). This represents the first identification of this putative precursor in an asperparaline-producing organism. Pre-paraherquamide has been proposed as the key, common biosynthetic precursor to both the paraherquamides and the asperparalines [6]. The identification of this substance in the paraherquamide- and asperparaline-producing *A. japonicus* JV-23 strain further supports

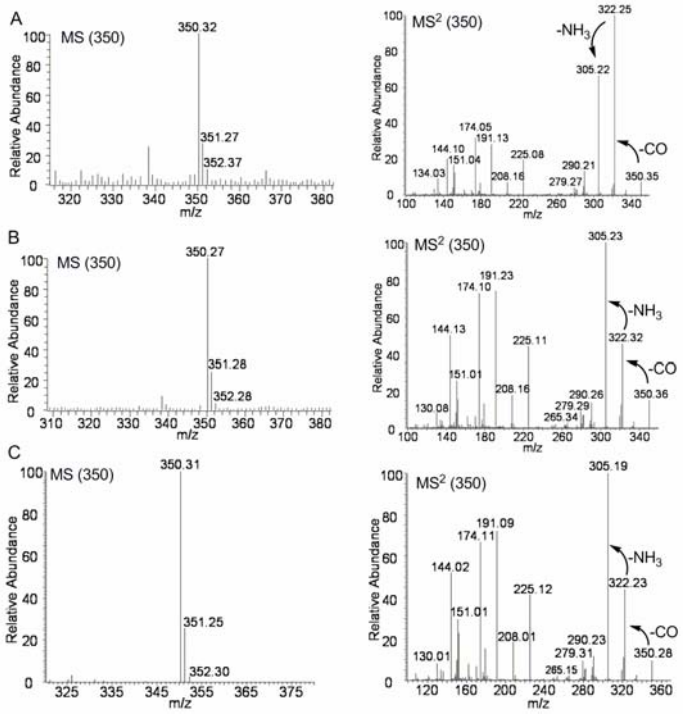


Figure 3-6. MS and MS^2 spectra of VM55599 (A) and pre-paraherquamide from the isolation from *P. fellutanum* culture (B), and pre-paraherquamide from the extracts of *A. japonicus* JV-23 cultures (C). The integral *m/z* values of ions for each analysis are included in corresponding graphs.

the unified biogenetic hypothesis detailed in (**Figure 3-7**). Curiously, VM55599 was not detected as a natural metabolite from the *A. japonicus* JV-23 cultures.

The identification of pre-paraherquamide as a natural trace metabolite suggests that structurally related common precursors may be involved in the biosynthesis of the family of prenylated indole alkaloids and that the structural diversity of these alkaloids are likely introduced by downstream tailoring enzymes following the construction of the bicyclo[2.2.2]diazaoctane core. LC-MS/MS analysis has several important advantages with respect to sensitivity and selectivity in the detection of trace natural metabolites. In natural products identification, NMR techniques are widely used to directly provide structural information when sufficient amounts of purified substances are available. However, in the case of biosynthetic intermediates, there is often a paucity of material that is thus insufficient for NMR structural studies. This is a manifestation of biosynthetic

intermediates being largely consumed by downstream tailoring enzymes, and thus these substrates do not accumulate. LC-MS/MS is an effective and powerful alternative in these cases and has been successfully used to identify many natural products in crude extracts by comparison with the respective reference compounds.[17] Herein, I was able to successfully deploy LC-MS/MS analysis to identify the presence of paraherquamide A, paraherquamide B, asperparaline A, VM55599, and pre-paraherquamide in crude fungal extracts using this technique.

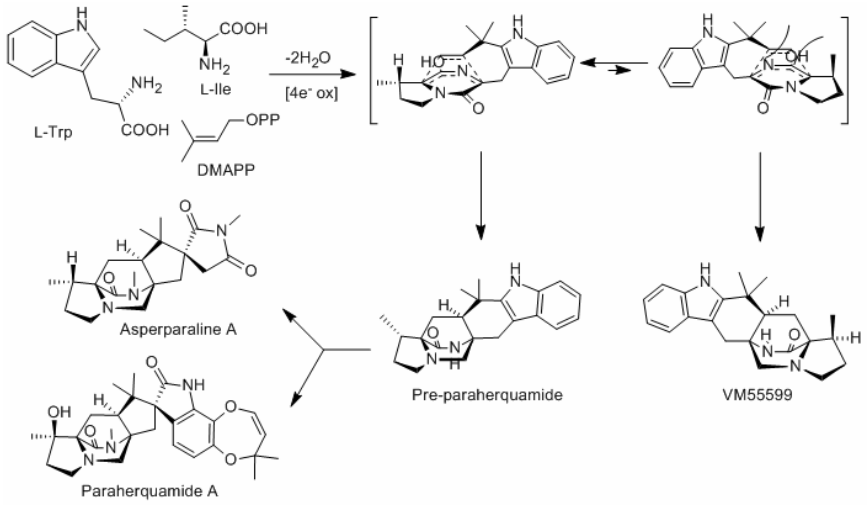


Figure 3-7. Proposed unified biogenesis of paraherquamides and asperparalines.

In conclusion, I have demonstrated for the first time, that pre-paraherquamide is a natural, albeit at low levels, secondary metabolite of the paraherquamide-producing organism *P. fellutanum* and the paraherquamide- and asperparaline-producing organism *A. japonicus* JV-23. This report constitutes the first confirmatory evidence for the natural existence of pre-paraherquamide and provides additional support for the unified biogenesis we have proffered (**Figure 3-7**). VM55599 is also produced by *P. fellutanum* and is consistent with the initial identification of this substance from the related paraherquamide-producing organism *Penicillium* sp. IMI332995 described by Everett and co-workers [23]. As the identification of new paraherquamide-producing fungi are discovered in various environments around the world, I speculate here, that VM55599 can be expected to be detected as a co-metabolite. It should be further noted that Miller and co-workers recently described the detection of VM55599 as a metabolite in several

strains of *Penicillium paneum* on the basis of mass spectral data [24]. It is entirely possible that these workers might have instead detected pre-paraherquamide (or both), which has the same mass as VM55599. As in the previously established case of *P. fellutanum*, VM55599 is a shunt (dead-end) metabolite as it possesses the incorrect absolute (and relative) stereochemistry to be processed further to a paraherquamide-like structure. Likewise, in *A. japonicus,* the major pathway metabolite pre-paraherquamide is largely consumed by the down-stream biosynthetic machinery responsible for the substantial oxidative elaboration of this biosynthetic intermediate into the asperparalines and paraherquamides.

3.3.2. Pre-malbrancheamide: Synthesis, Isotopic Labeling, Biosynthetic Incorporation, and Detection in Cultures of *Malbranchea aurantiaca*

Malbrancheamide and malbrancheamide B have been isolated from *M. aurantiaca* [9, 25]. In the current study, both natural products were also found in the fungal extract by LC-MS analysis (**Figure 3-8**). The retention times of malbrancheamide and malbrancheamide B were 35.3 min and 31.0 min, respectively, in the LC spectrum. Both compounds exhibited the expected chlorine isotope patterns and *m/z* values. In the MS/MS analysis, the loss of CO_2 (C-14 atom) and then NH_3 from the molecular ion of malbrancheamide generated fragments at *m/z* of 376.25 and 359.24, respectively (**Figure 3-9**). A similar fragmentation pattern was observed in the malbrancheamide B MS/MS spectrum, further confirming the production of this natural product in *M. aurantiaca* [25]. Interestingly, one compound in the fungal extract had both the same *m/z* value (336.31) and the retention time (24.6 min) as authentic pre-malbrancheamide (**Figure 3-8**). Moreover, this isolated compound had a similar MS/MS fragmentation pattern compared to malbrancheamide and malbrancheamide B, indicative of the structural homology of these three compounds (**Figure 3-9**). Furthermore, the identical MS/MS spectra of this compound and synthetic, authentic compound confirmed the presence of pre-malbrancheamide in the fungal extract (**Figure 3-9**, **Figure S3-4**).

In order to investigate the role of pre-malbrancheamide in malbrancheamide biosynthesis, doubly ^{13}C-labeled pre-malbrancheamide was synthesized. As shown in

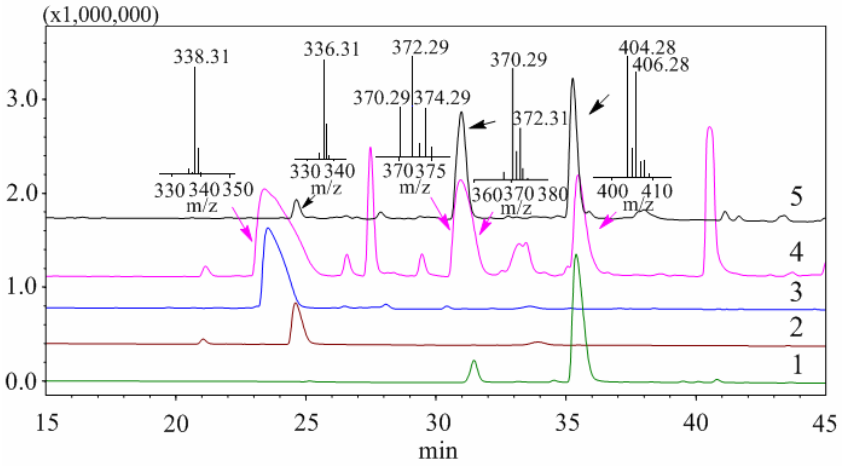


Figure 3-8. LC-MS analysis of extracts from *M. aurantiaca* liquid culture. Lane 1, authentic malbrancheamide and malbrancheamide B; lane 2, authentic pre-malbrancheamide; lane 3, doubly labeled pre-malbrancheamide; lane 4, *M. aurantiaca* fungal extract in feeding experiment with ^{13}C-labeled premalbrancheamide; lane 5, *M. aurantiaca* fungal extract. All natural products exhibited expected *m/z* values in MS analysis. Pre-malbrancheamide was identified in *M. aurantiaca* fungal extract by comparing to authentic compound. In the feeding experiment, both ^{13}C labeled malbrancheamide B and native compound were isolated while no ^{13}C labeled malbrancheamide was detected by LC-MS.

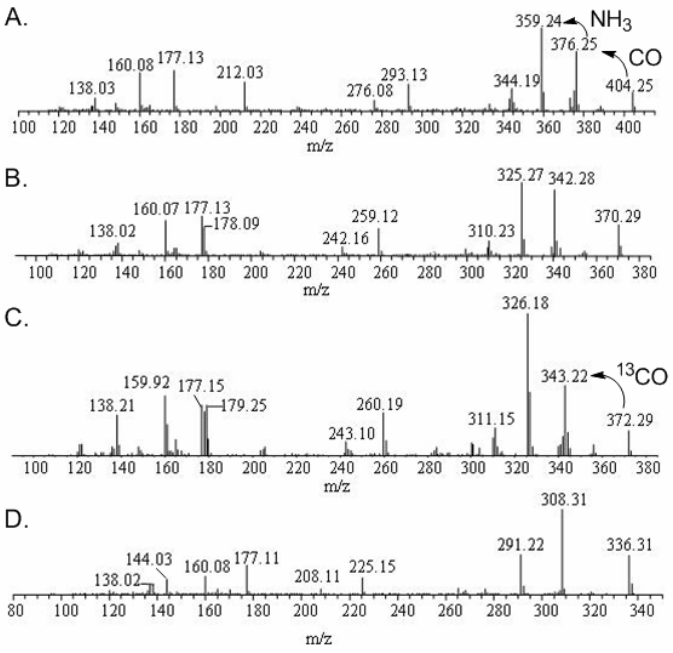


Figure 3-9. MS/MS spectra of malbrancheamide (A), malbrancheamide B (B), doubly ^{13}C-labeled malbrancheamide B (C), and pre-malbrancheamide (D) from the fungal extract.

Figure 3-10, amino acid coupling of the ^{13}C-labeled reverse prenylated tryptophan derivative and ^{13}C-labeled *cis*-3-hydroxyproline ethyl ester (as the TFA salt) in the presence of HATU provided a mixture of diastereomers. Treatment of the resulting peptide with TFA resulted in deprotection of the carbamate, and subsequent heating of the resultant primary amine with 2-hydroxypyridine in toluene gave the dioxopiperazine. Dehydration of this intermediate under Mitsunobu conditions led to the enamide, which smoothly underwent a hetero-Diels-Alder reaction when treated with aqueous KOH in MeOH affording cycloadducts. Treatment of the major cycloadduct with excess DIBAL-H led to selective reduction of the tertiary amide in the presence of a secondary amide to provide double ^{13}C-labeled pre-malbrancheamide in excellent yield.

Double ^{13}C-labeled pre-malbrancheamide was added to cultures of *M. aurantiaca* in a precursor incorporation experiment. As a putative precursor of pre-malbrancheamide, the major cycloadduct was also included in the analysis. Fungal extracts from these precursor incorporation studies were analyzed by LC-MS and ^{13}C enrichment was revealed by MS/MS analysis. Double ^{13}C-labeled pre-malbrancheamide was clearly

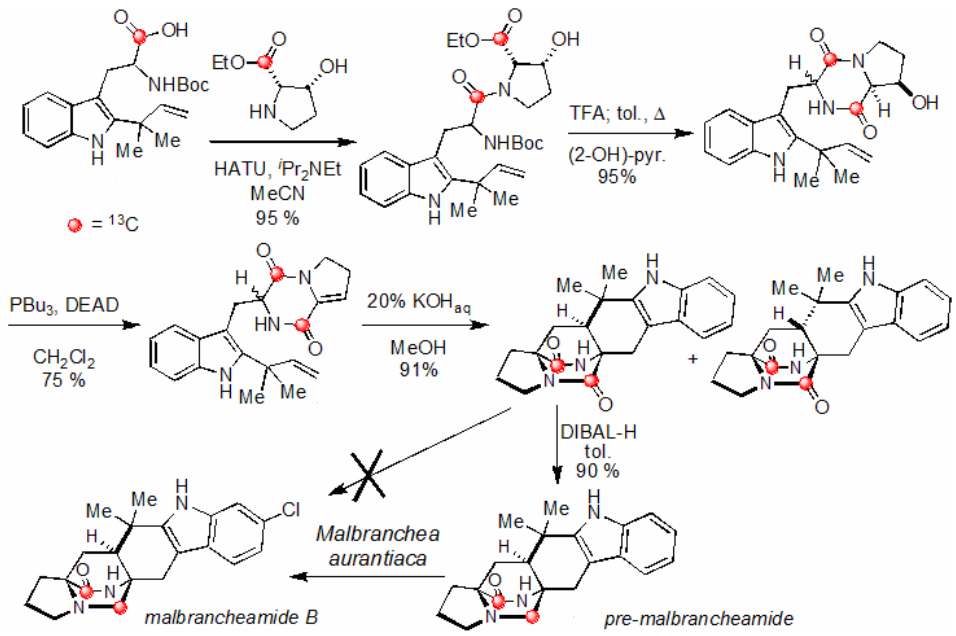


Figure 3-10. Doubly ^{13}C-labeled pre-malbrancheamide and product incorporation.

incorporated intact into malbrancheamide B, whose parent ion had an *m/z* value of 372.29 (**Figure 3-8**). Its retention time was the same as that of the native malbrancheamide B. In the MS/MS spectrum of doubly ^{13}C-labeled malbrancheamide B, the fragment at *m/z* of 343.22 was produced by the loss of ^{13}CO containing a ^{13}C atom at its C-14 position (**Figure 3-9C**). A similar fragmentation pattern was observed in the MS/MS spectrum of compound **17** (**Figure S3-4**). The *m/z* difference (=1) of many fragments in MS/MS spectra of labeled malbrancheamide B and its natural compound is due to ^{13}C atom incorporation in the fragments. From analysis of the electrospray mass spectrum, incorporation was determined to be 5.5% for the intact doubly labeled material. Furthermore, C-5 and C-14 of the isolated malbrancheamide B had significant chemical shifts in the ^{13}C NMR spectrum, compared to compound un-labeled malbrancheamide B.

Interestingly, ^{13}C-labeling of malbrancheamide itself was not detected by LC/MS-MS analysis. Only double ^{13}C-labeled malbrancheamide B was produced in this feeding experiment (**Figure 3-8**). This might be due to the kinetics of the second chlorination reaction being considerably slower than the first. Efforts are currently underway to prepare doubly ^{13}C-labeled malbrancheamide B in sufficient quantities for analogous feeding studies to investigate that malbrancheamide arises from a subsequent C6-chlorination of malbrancheamide B. Curiously, feeding of doubly ^{13}C-labeled dioxopiperazine to *M. aurantiaca* did not label either malbrancheamide or malbrancheamide B, which again raises some important questions regarding timing of reduction of the tryptophan carbonyl residue.

3.3.3. Molecular Analysis of a 4-Dimethylallyltryptophan Synthase from *Malbranchea aurantiaca*

Identification and purification of MaPT. Several PT amino acid sequences were aligned to identify conserved regions for PT gene degenerate primer design in *M. aurantiaca* RRC1813 (**Figure 3-11A**). A 750-bp fragment was isolated and predicted to be a dimethylallyltryptophan synthase (DMATS) based on sequence comparisons (data not shown). TAIL-PCR was subsequently applied to isolate the 1371-bp full length MaPT genomic DNA [26]. A 63-bp intron was predicted in the MaPT genomic DNA (BCM

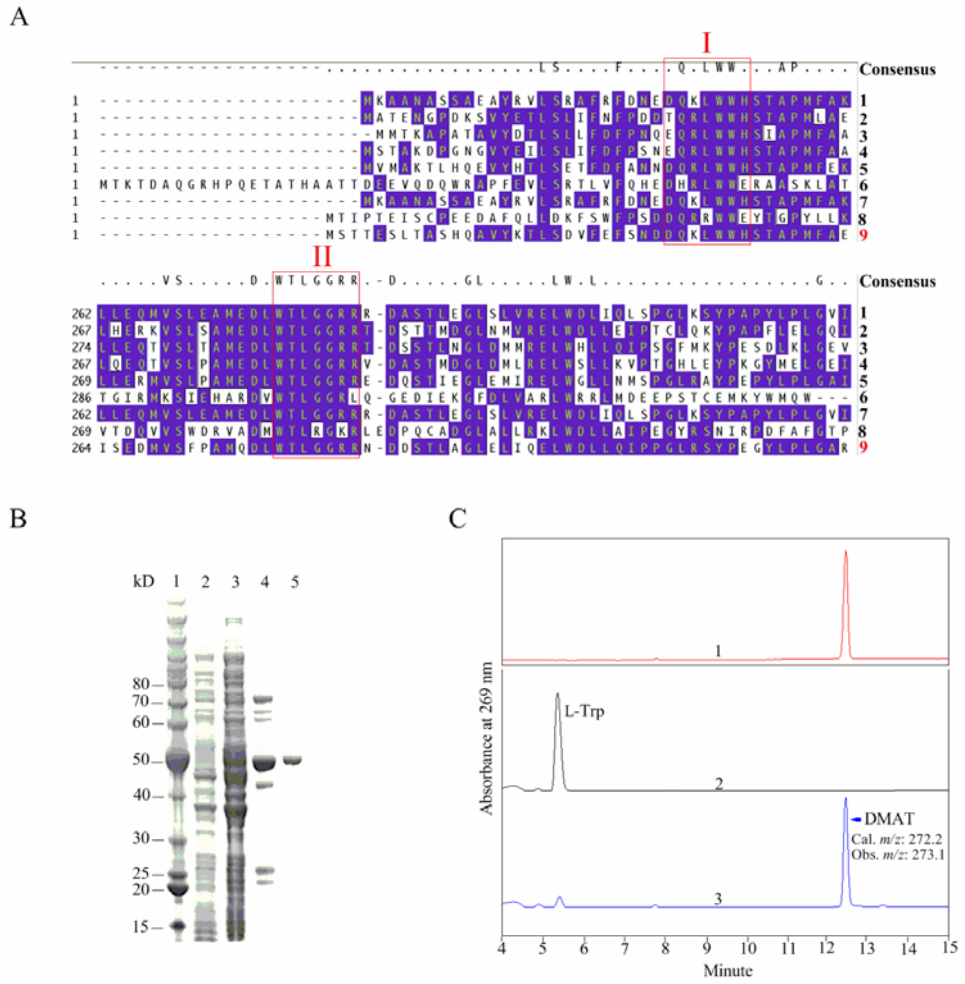


Figure 3-11. Preparation of prenyltransferase (MaPT) from *M. aurantiaca* RRC1813 and HPLC analysis of enzyme reactions. A: alignment analysis of several members of the new class of PTs. Two motifs were identified for designing degenerate primers. Lane 1 to 5 represents DMATS from *A. fumigatus*, *B. obtecta*, *C. fusiformi*, *C. purpurea*, and *N. coenophialum*, respectively. Lane 6 to 9 are FgaPT1, FgaPT2, FtmPT2, and MaPT, respectively. B: 4-12% SDS-PAGE analysis of purified MaPT. Lane 1, molecular marker; lane 2, IPTG-uninduced total protein fraction; lane 3, IPTG-induced total protein fraction; lane 4, enzyme purified with Ni-NTA agarose; lane 5, enzyme purified with gel chromatography. The purified enzyme showed expected molecular weight of ~51 kDa. C: HPLC analysis of enzyme reactions. Lane 1, authentic dimethylallyltryptophan (DMAT); lane 2, negative control with boiled enzyme used in the reaction; lane 3, MaPT enzyme reaction mixture. The product was analyzed by MS and showed an expected *m/z* value of mono-prenylated L-Trp.

Gene Finder) and removed to produce the corresponding MaPT cDNA. The PT encoded from this cDNA was aligned with DMATSs from *A. fumigatus* (EAL94103), *Balansia obtecta* (Q6X1E1), *Claviceps fusiformis* (Q12594), *C. purpurea* (Q9C141), and *Neotyphodium coenophialum* (Q6X2E1), FgaPT1 (EAL94098), FgaPT2 (AAX08549),

and FtmPT2 (**Figure 3-11A**) [20]. It shares two conserved regions for designing degenerate primers with other PTs and has the highest sequence identity with FgaPT2 (64 %). The lack of a conserved (N/D)DXXD motif in this PT from *M. aurantiaca* RRC1813 indicated that this enzyme is a member of a recently identified new class of PTs.

The MaPT cDNA was cloned into pET28a for protein expression in *E. coli* BL21(DE3). Purified polypeptide was obtained by Ni-NTA agarose and further by gel chromatography with a Superdex 200 column to generate a product with apparent homogeneity. The observed molecular weight of enzyme in 4-12 % SDS-PAGE analysis was about 51 kDa, consistent with a theoretical weight of 52 kDa for His_6-MaPT (**Figure 3-11B**). Using gel-filtration chromatography, the native molecular weight of the enzyme was determined to be 61.6 kDa, suggesting a monomeric species (**Figure S3-5**).

Enzyme activity and substrate selectivity. L-Trp and a dipeptide, brevianamide F, were first tested as enzyme substrates. Brevianamide F was utilized by the enzyme with very poor conversion (**Table 3-1**), whereas L-Trp was converted effectively into product, demonstrating it to be a more acceptable enzyme substrate (**Table 3-1**, **Figure 3-11C**). The *m/z* of the product corresponds to the theoretical weight of DMAT, suggesting the enzyme to be a DMATS. This result further confirms that DMATS is an enzyme common in many fungal strains such as *Aspergillus* sp., *Balansia* sp., *Claviceps* sp., *Neotyphodium* sp., and the newly isolated *Malbranchea* sp. MaPT was specific for DMAPP as no product was detected by LC-MS when L-Trp and brevianamide F were tested with IPP, GPP, and FPP (data not shown).

Although both L-Trp and brevianamide F were accepted by MaPT, its native substrate was not predictable by comparing its sequence to other known enzymes in this class. To assess further the substrate flexibility, we interrogated 14 additional compounds as substrates with nine converted into products observed by LC-MS analysis (**Table 3-1**). The enzyme activity was less when L-Trp was modified, and this is exemplified by 1-Me-L-Trp and the dipeptide Ala-Trp with relative activities of 3.2 ± 0.35 % and 0.6 ± 0.06 %, respectively. Additionally, amino acid chirality was critical for substrate recognition with

Table 3-1. Investigation of MaPT substrate selectivity

Substrate	Relative Activity (%)	Mr of Product	$[M+1]^+$ of Product
L-Trp	100 ± 18.39	272.2	273.1
D-Trp	0.5 ± 0.04	272.2	273.1
1-Methyl- L-Trp	3.2 ± 0.35	286.2	287.4
5-Br-DL-Trp	0.3 ± 0.02	350.1	350.9
5-F- DL-Trp	ND[1]	ND[1]	ND[1]
5-Me- DL-Trp	16.1 ± 0.01	286.2	287.0
5-MeO- DL-Trp	0.7 ± 0.02	302.2	303.1
5-OH- L-Trp	17.0 ± 0.37	288.2	289.1
L-Abrine	24.9 ± 1.45	286.2	287.1
6-Me- DL-Trp	27.0 ± 2.66	286.2	287.1
Ala-Trp	0.6 ± 0.06	343.2	344.2
N-(2,4-Dinitrophenyl)- L-Trp	ND[1]	ND[1]	ND[1]
L-Phe	ND[1]	ND[1]	ND[1]
L-Tyr	ND[1]	ND[1]	ND[1]
L-His	ND[1]	ND[1]	ND[1]
Brevianamide F[2]	0.5 ± 0.03	353.2	354.1

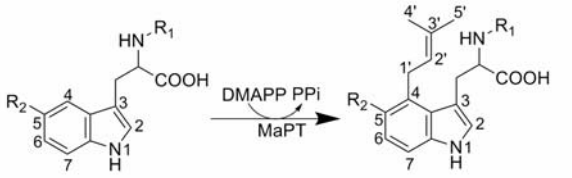
1: ND represents no product detected; 2: brevianamide F was labeled with two ^{13}C atoms.

0.5 ± 0.04 % relative activity of MaPT toward D-Trp. Finally, the enzyme is strictly specific for L-Trp and did not utilize other aromatic amino acid analogues including L-Phe, L-His, or L-Tyr.

Five substrates used in this study contained side chains with variant electronegativity at C-5 of the L-Trp indole ring. 5-OH-L-Trp was the best substrate with 17.0 ± 0.37 % relative activity, followed by 5-Me-DL-Trp (16.1 ± 0.01 %), 5-MeO-DL-Trp (0.7 ± 0.02 %), and 5-Br-DL-Trp (0.3 ± 0.02 %), while 5-F-DL-Trp was not utilized by the enzyme. Thus, electronegativity of the side chain substituent on the L-Trp indole ring was critical for enzyme catalytic efficiency. Interestingly, FgaPT2 exhibits the same trend towards these five substrates, suggesting it is common to other DMATSs [27].

Identification of MaPT enzyme reaction products. Products formed by reaction of MaPT with L-Trp, L-abrine, 5-OH-L-Trp, and 5-Me-DL-Trp in large-scale reactions were purified by semi-preparative HPLC and analyzed by ^1H and ^{13}C NMR and HRMS

(**Figure 3-12**). ^{1}H NMR analysis of the MaPT reaction with L-Trp revealed that a dimethylallyl group was appended at C-4 of the indole ring (**Table S3-1**). HRMS analysis of this product showed an ion with the *m/z* of 273.1610, which corresponds to the theoretical mass $[M+H]^{+}$ (273.1603) of prenylated tryptophan. Further analysis of products with 5-OH-L-Trp, 5-Me-DL- Trp, and L-abrine as enzyme substrates confirmed C-4 of indole as the site for covalent linkage of the dimethylallyl group (**Table S3-1**). ^{13}C NMR analysis provided additional evidence to confirm the structures of 4-prenylated L-abrine, 5-OH-L-Trp, and 5-Me-DL-Trp. These three products also showed expected *m/z* values in the HRMS analysis.



	R_1	R_2
L-Trp	H	H
L-Abrine	Me	H
5-OH-L-Trp	H	OH
5-Me-DL-Trp	Me	H

	R_1	R_2
DMAT	H	H
DMAA	Me	H
DMA5HT	H	OH
DMA5MT	Me	H

Figure 3-12. Four identified MaPT products. Four different enzyme substrates were converted into corresponding products in large scale enzyme reactions. The products were isolated by semi-prep HPLC and subjected for ^{1}H and ^{13}C NMR analyses and high resolution MS analysis (HRMS).

Metal dependence of MaPT. The optimal temperature and pH value for enzyme reactions were 30 °C and 7.4, respectively (**Figure S3-6**). The product in the reaction including L-Trp was produced to be correlated linearly with up to 0.43 μg protein per 100 μl assay, and up to 12 min under the above optimal conditions. The enzyme activity did not require any additive divalent cation while addition of 5 mM of Mg^{2+} or Ca^{2+} slightly enhanced MaPT activity by about 130% (**Figure S3-7**). Five divalent metal ions (Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, and Zn^{2+}) at 5 mM inhibited the enzyme activity from 50% to 80%, and 5 mM of Cu^{2+} or Sn^{2+} significantly reduced enzyme activity to less than 10%. EDTA was found to hinder enzyme activity in a concentration dependent manner (**Figure S3-7**).

More than 4 mM EDTA totally abrogated enzyme activity after incubating it with the enzyme at 30 °C for 5 min; this was not restored by addition of 10 mM Mg^{2+}.

Cation dependence in prenyltransferases has been well documented, and represents a key feature of the mechanistic biochemistry in these enzymes. Therefore, divalent cation content in MaPT was investigated with ICP-MS. Although Ni^{2+} concentration (1.62 μM in 17 μM MaPT solution) was increased about 26 times, all other metal ion content values were not changed, and found to be <0.05 mol per mol of MaPT. The increased Ni^{2+} content was presumed to be derived from Ni-NTA agarose used in protein purification. Therefore, purified MaPT contained none of the typical metal ion co-factors, and EDTA presumably exhibited its inhibition through a mechanism that is independent of divalent cation sequestration. Enzyme inhibition by metal ion chelating agents was further investigated with EGTA, EDDA, DHPTA, and EDTA-Mg^{2+} complex (**Figure 3-13A**). Dialysis of the enzyme solution containing 5 mM EDTA against reaction buffer regenerated active enzyme whose efficiency was similar to the positive control (**Figure 3-13B**), demonstrating EDTA as a reversible inhibitor of MaPT. Interestingly, 5 mM EDTA was also reported to reduce CloQ prenyltransferase activity to 25 % [28], and inhibits horseradish peroxidase, human liver arginase, and lignin peroxidase H2 in a metal ion independent manner [29-31].

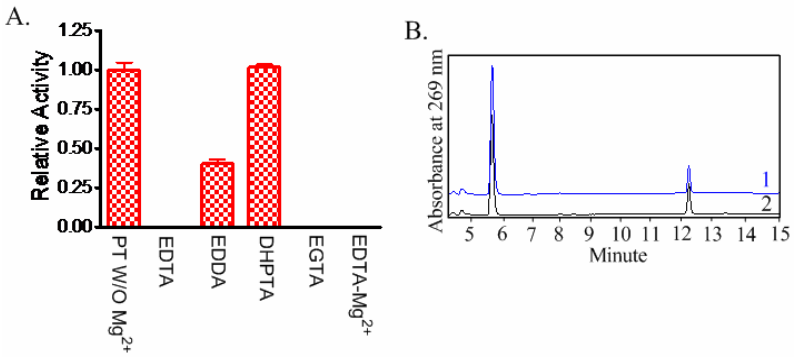


Figure 3-13. Reversible inhibition of MaPT activity by EDTA. A: inhibition of MaPT by EDTA analogues. Each chemical (5 mM) was incubated with the enzyme at 30 °C for 5 min before initiating the reaction. B: Regaining enzyme activity by dialysis. EDTA (5 mM) was incubated with MaPT at 30 °C for 5 min, and then dialyzed against gel chromatography running buffer at 4 °C overnight. Line 1 (blue), reaction with 1 mM L-Trp and dialyzed MaPT treated with EDTA; line 2 (black), reaction with 1 mM L-Trp and dialyzed MaPT untreated with EDTA. Mg^{2+} was not added to either reaction.

Kinetic analysis of MaPT. Several DMATS from *Claviceps* sp. and PTs from *A. fumigatus* have been biochemically characterized with radiolabeled substrates or directly by HPLC-UV [20, 32-39] analysis. In this study, HPLC-UV was used to investigate MaPT kinetics with L-Trp, 5-OH-L-Trp, and L-abrine coupled with DMAPP as enzyme substrates. The enzyme reaction followed Michaelis-Menton kinetics when L-Trp concentration was varied and DMAPP concentration held constant at 0.2 mM (**Table 3-2**). The K_m and V_{max} values for L-Trp were 48.88 ± 1.86 μM and 3.41 ± 0.04 μM/min, respectively, giving a turnover value of 80.42 ± 0.94 min^{-1}. Interestingly, DMAPP showed self-inhibition when its concentration was varied. The K_m and K_s values for DMAPP were 13.97 ± 1.86 μM and 126.37 ± 27.30 μM, respectively. Its V_{max} value (5.63 ± 0.45 μM/min) was higher than that with L-Trp (measured at 0.2 mM DMAPP in L-Trp kinetic analysis). Both 5-OH-L-Trp and L-abrine had higher K_m values and lower V_{max} values than those of L-Trp. 5-OH-L-Trp K_m and V_{max} values were 303.08 ± 25.73 μM and 1.43 ± 0.04 μM/min, respectively. No self-inhibition was observed for up to 200 μM DMAPP in the presence of 5-OH-L-Trp. L-Abrine was converted into product with lower catalytic efficiency and showed self-inhibition in the kinetic assays. Its K_m and K_s values were 105.99 ± 21.14 μM and 964.15 ± 260.85 μM, respectively. The V_{max} value of L-abrine (2.70 ± 0.30 μM/min) was lower than that of L-Trp. K_m and V_{max} values of L-Trp were comparable to other PTs including FgaPT2 and DMATS from *C. purpurea* [35]. MaPT exhibited catalytic efficiencies about 5-fold higher toward both 5-OH-L-Trp and L-abrine compared to FgaPT2 [27].

In the next assay, both L-Phe and D-Trp were included in the enzyme reaction to investigate their affect on enzyme kinetics. The kinetic parameters of L-Trp and DMAPP were not changed by addition of 100 μM L-Phe. However, L-Trp kinetic parameters were reduced by addition of D-Trp, while DMAPP kinetic data were not changed (**Figure 3-14**). For example, $K_{m,\,obs}$ and $V_{max,\,obs}$ values for L-Trp with 50 μM D-Trp was 107.79 ± 5.07 μM and 3.56 ± 0.07 μM/min, respectively, while $K_{m,\,obs}$ and $V_{max,\,obs}$ values for DMAPP were 12.14 ± 3.36 μM and 4.87 ± 0.707 μM/min, respectively (**Table 3-2**). The K_i for D-Trp was determined to be 40.41 μM after plotting $K_{m,\,obs}$ values and

Table 3-2. Kinetic parameters of MaPT with three substrates and with D-Trp as inhibitor.

Substrate	K_m (μM)	V_{max} (μM/min)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1} \cdot min^{-1}$)
L-Trp	48.88±1.86	3.41±0.04	80.42±0.94	1.65±0.04
5-OH-L-Trp	303.08±25.73	1.43±0.04	33.73±0.94	0.11±0.09
L-Abrine	105.99±21.14	2.70±0.30	63.68±7.08	0.60±0.23
L-Trp*	107.79±5.07	3.56±0.07	83.96±1.65	0.78±0.05
L-Trp†	13.97 ±1.86	5.63 ±0.45	132.78±10.61	9.50±0.16
5-OH-L-Trp†	41.20 ±3.48	1.40 ±0.04	33.02±0.94	0.80±0.09
L-Abrine†	19.00 ±3.34	1.19 ±0.06	28.07±1.42	1.48±0.18
L-Trp*†	12.14 ±3.36	4.87 ±0.70	114.86±16.51	9.46±0.31

***:** kinetic data were obtained in the presence of 50 μM D-Trp. †: DMAPP kinetic parameters.

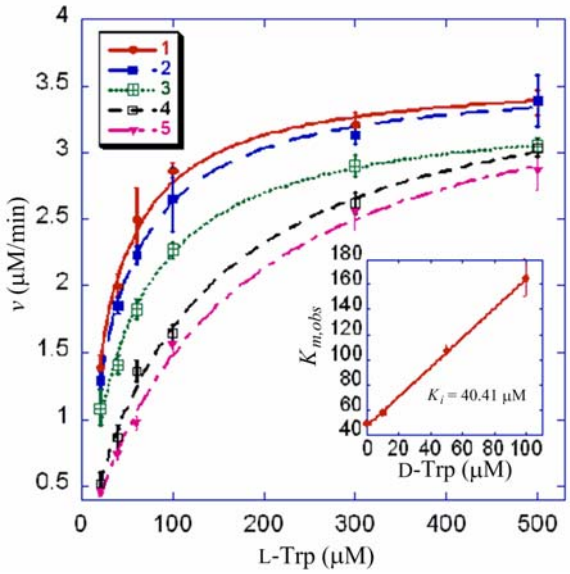


Figure 3-14. MaPT reaction with L-Trp as substrate was inhibited by varying concentrations of D-Trp. Line 1, the Michaelis-Menton curve resulting from reactions in the presence of 100 μM L-Phe; line 2 to 5, the Michaelis-Menton curve generated by reactions in the presence of 0, 10, 50, and 100 μM D-Trp, respectively. The right inset graph is the plot of $K_{m, obs}$ vs. [D-Trp] to provide the K_i value.

corresponding D-Trp concentrations by the equation of $K_{m, obs} = K_m [1 + [\text{D-Trp}]/K_i]$. The data revealed that D-Trp acts as a competitive inhibitor in the MaPT reaction.

Mutagenesis of MaPT. The MaPT amino acid sequence was aligned to Orf2, CloQ, and NovQ sequences by ESPript (**Figure S3-8**). MaPT K261 (corresponding to N173 in Orf2)

was identified as a possible key residue involved in DMAPP binding. This residue was mutated to Glu and Leu (**Figure S3-9**). MaPT K261E was highly soluble and readily purified, while its K261L form was only partially purified due to significant insolubility. No product was generated in either reaction, reflecting the potential key role of K261 in enzyme catalysis (**Figure 3-15**). MaPT K189 (the counterpart of K119 in Orf2) was also presumed to be involved in diphosphate binding. Both MaPT K189E and K189L retained only 1.1% and 0.9% of wild type enzyme activity, respectively (**Figure S3-9**, **Figure 3-15**). Thus, this residue is also a key candidate for being involved in enzyme catalysis. Crystal structure analysis revealed that S51 in Orf2 is involved in substrate binding through a hydrogen-bond. A similar residue, T105, is found in MaPT while two basic residues are present in CloQ, NovQ, and Fnq26. These positively charged residues have been suggested to act as surrogates of Mg^{2+} for α-phosphate binding. The key residues for Mg^{2+} binding in Orf2 are D62 and D110. Although D62 is not a conserved residue, D110 is seemingly conserved among these PTs, casting some uncertainty on the role limited to metal binding. In this study, MaPT T105 was mutated to Asp and Val while D179 (corresponding to D110 in Orf2) was mutated to Asn, Glu, and Val (**Figure S3-9**). MaPT

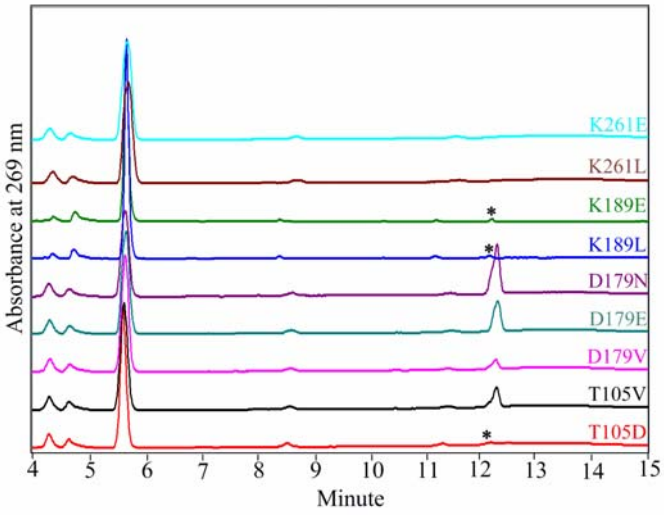


Figure 3-15. HPLC analysis of reaction products using mutant forms of MaPT and L-Trp as substrate. All reactions were run with the same reaction components and conditions. Addition of 0.86 µg MaPT mutants except K261L initiated the reactions, which were incubated at 30 °C for 30 min before termination. Instead, 11.7 µg partially purified MaPT K261L was used in the reaction. Peaks with low levels of product were labeled with black asterisks.

T105V retained about 11% of wild type enzyme activity while the T105D mutant form was almost completely inactive (1.4% of wild type MaPT) (**Figure 3-15**). Considering S51 in Orf2 also points toward the GPP binding pocket (5-7 Å distance), this result suggests that T105, along with other hydrophobic residues, possibly defines a binding pocket for DMAPP. MaPT D179N, D179E, and D179V retained 40%, 25%, and 7% of wild type enzyme activity, respectively.

3.3.4. Detailed Characterization of One Deoxybrevianamide E Synthase in Stephacidin/Notoamide Biosynthetic Pathway

The biochemical understanding of prenylated indole alkaloids remains limited because no biosynthetic gene cluster from any alkaloid producing fungal strain is available. My initial efforts to clone out asperparaline and malbrancheamide gene clusters relied on pairs of degenerate primers designed from bacterial NRPS A domains, but resulted in limited successes possibly because of the inconsistence of A domain specific codes between bacterial and fungal NRPS systems. The advance and low cost of the fungal whole genome sequencing represents a new strategy to attack this problem. Notoamide- and stephacidin-producing marine *Aspergillus* sp. was chosen for the whole genome sequencing target because of its natural product anticancer activities and significantly unique structure features. State-of-the-art 454FLX technology in the University of Michigan DNA Sequencing Core facility generated a 480-Mb fungal genomic database. With assists from bioinformatics specialists, a partial fungal genome has been assembled and the genome size was estimated to be about 36 Mb, similar to all published *Aspergillus* genome sizes [40]. This project not only represents the first effort to investigate marine *Aspergillus* sp. to develop an understanding fungal genetic evolution and adaptation to different environmental niches, but also provided the possibility to study the biosynthesis of prenylated indole alkaloids.

Although there are structural differences among brevianamide A, notoamides, and stephacidins, they all may share the same biosynthetic precursor, brevianamide F [15]. Recently, *FtmA* in the *Aspergillus fumigatus* genome was identified and characterized as the brevianamide F synthase in ergot alkaloid production [19]. In my studies, *FtmA* was

used to screen the *Aspergillus* MF297-2 preliminary genome, and one putative dimodular NRPS gene (*NotC*) was isolated (**Figure 3-16**). This 6723-bp intronless gene encodes a polypeptide product with 47 % identity to FtmA in *A. fumigatus* (**Table 3-3**). Besides this gene, 18 other genes were also identified in a 42457-bp chromosomal region. At the left side of the gene cluster, the products of *Orf1-3* might be involved in a primary metabolic pathway. At its right side, the predicted functions of *Orf4-9* suggested that they might not be directly related to notoamide/stephacidin biosynthesis. Instead, ten genes *(NotA-J*), including *NotC*, were predicted to be essential. NotA and NotD are two predicted DMATSs, which possibly catalyze two required prenylation steps in fungal alkaloid biosynthesis. The 1350-bp *NotA* contained a 69-bp intron and its encoded product showed highest identity (50 %) to FtmH (also termed as FtmPT2) in *A. fumigatus* [41]. Similarly, NotD whose gene was 1431 bp with a 72-bp intron has the highest identity (40 %) to one putative DMATS (EER24759) in *Coccidioides posadasii*. NotB was predicted as a flavin-dependent oxidoreductase but its role in notoamide/stephacidin biosynthesis remains unknown. NotE and NotF are two P450s, which together might catalyze the formation of the terminal 6-membered ring attached to indole ring. The function of NotG was predicted as an FAD-dependent oxidoreductase while NotH was a hypothetical enzyme whose function remains unknown. The pathway-specific resistant machinery is NotI, which was predicted as an efflux pump. NotJ was predicted as a C6 zinc finger domain and possibly acts as a pathway-specific transcriptional activator.

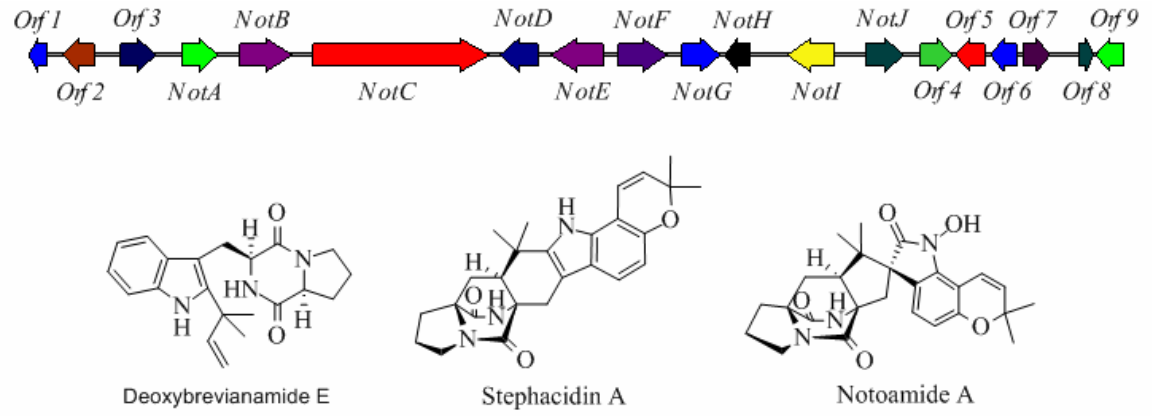


Figure 3-16. One putative gene cluster for stephacidin/notoamide biosynthesis in *Aspergillus* MF297-2. Besides notoamide A and stephacidin A, one putative biosynthetic intermediate, deoxybrevianamide E, was also isolated from marine fungal culture [14].

Table 3-3. Features of the *not* gene products

Protein	Size bp/aa	Exon	Function	Relative (identity/similarity [%])	Accession number
Orf1	731/224	1-114, 173-731	partial polysaccharide synthase	Capsule polysaccharide biosynthesis protein from *Aspergillus fumigatus* (43/63)	XP_748327
Orf2	1199/339	1293-1555, 1643-1971, 2023-2179, 2234-2491	negative regulator	NmrA family protein from *Ajellomyces capsulatus* (45/65)	EEH03447
Orf3	1344/401	3486-4029, 4141-4487, 4568-4829	FAD binding domain protein	FAD binding domain protein from *A. clavatus* (44/63)	XP_001268514
NotA	1350/427	5819-6996, 7016-7168	prenyl-transferase	FtmH from *A. fumigatus* (50/66)	BAH24002
NotB	2025/621	8012-8294, 8389-8927, 8996-10036	oxidoreductase	oxidoreductase from *Microsporum canis* (40/59)	EEQ33235
NotC	6723/2241	10787-17509	NRPS	FtmA from *A. fumigatus* (47/67)	XP_747187
NotD	1431/453	17924-18053, 18126-19354	prenyl-transferase	tryptophan dimethylallyltransferase from *Coccidioides posadasii* (40/62)	EER24759
NotE	1901/544	19899-20086, 20171-20272, 20347-20635, 20689-20811, 20878-21799	P450	cytochrome P450 from *A. fumigatus* (62/75)	XP_747185
NotF	-	-	P450	cytochrome P450 from *A. fumigatus* (47/65)	EDP49182
NotG	1423/434	24804-24963, 25022-25990, 26004-26226	FAD binding domain protein	FAD binding domain protein from *A. clavatus* (44/63)	XP_001268514
NotH	1113/371	26391-27503	unknown	hypothetical protein from *Salinispora arenicola* (52/65)	YP_001537335
NotI	1851/564	28772-29142, 29197-29570, 29621-30390, 30446-30622	efflux pump	MFS transporter from *Neosartorya fischeri* (87/93)	XP_001265322
NotJ	1455/484	31790-33244	transcriptional activator	C6 zinc finger domain protein from *N. fischeri* (53/62)	XP_001265321
Orf4	1266/422	33817-35082	unknown	hypothetical protein from *Talaromyces stipitatus* (76/83)	XP_002482929
Orf5	1126/340	35194-35246, 35300-35897, 35950-36319	dehydrogenase	alcohol dehydrogenase from *Penicillium marneffei* (60/76)	XP_002147947
Orf6	993/331	36522-37514	short-chain dehydrogenases /reductase	hypothetical protein from *Nectria haematococca* (66/80)	EEU36425
Orf7	1020/322	37772-37932, 37987-38791	unknown	metallo-β-lactamase domain protein from *T. stipitatus* (80/88)	XP_002482927
Orf8	569/152	39873-40062, 40122-40318, 30372-40441	unknown	hypothetical protein from *T. stipitatus* (88/94)	XP_002482928
Orf9	1517/461	40516-41142, 41214-41729, 41793-42032	transcriptional co-activator	hypothetical protein from *P. marneffei* (45/61)	XP_002144868

To verify functions of NotA and NotD and their roles in notoamide/stephacidin biosynthesis, their cDNAs were prepared by removing introns with overlapping PCR (**Table S3-2**). The predicted cDNAs were sequenced to exclude any error introduced by PCR and cloned into the *Nde*I and *Not*I sites of pET28a for heterologous overexpression in *E. coli* BL21 CodonPlus-(DE3)-RIPL strain. The N-terminally His-tagged proteins were purified with Ni-NTA resin to reach about 90 % purity (**Figure S3-10A**). The native status of NotD (53.6 kD) was determined as the oligomer with observed molecular

weight of 292 kD, suggesting the formation of a pentamer or hexamer (**Figure S3-10B**). Protein concentrations were determined with their corresponding extinction coefficients at 280 nm (97,950 and 91,853 M^{-1} cmP^{-1} for NotA and NotD, respectively).

Next, the functions of NotA and NotD were examined with L-Trp, doubly ^{13}C-labeled brevianamide F, deoxybrevianamide E, and doubly ^{13}C-labelled keto-premalbrancheamide as substrates (**Figure S3-11**). No prenylated product was detected with LC-MS when L-Trp, doubly ^{13}C-labeled brevianamide F, and deoxybrevianamide E were used in NotA reactions, while one small peak appeared in the LC trace of the enzyme reaction with doubly ^{13}C-labeled keto-premalbrancheamide as its substrate (**Figure 3-17**). The peak content exhibited the m/z value at 420.10. This result indicated the single prenylation reaction catalyzed by NotA and its putative position in notoamid/stephacidin biosynthetic scheme. The relatively low conversion possibly suggested that keto-premalbrancheamide was not the physiological substrate of NotA. Subsequently, NotA substrate specificity was preliminarily investigated with several other chemicals including, premalbrancheamide, preparaherquamide, and its diastereomer

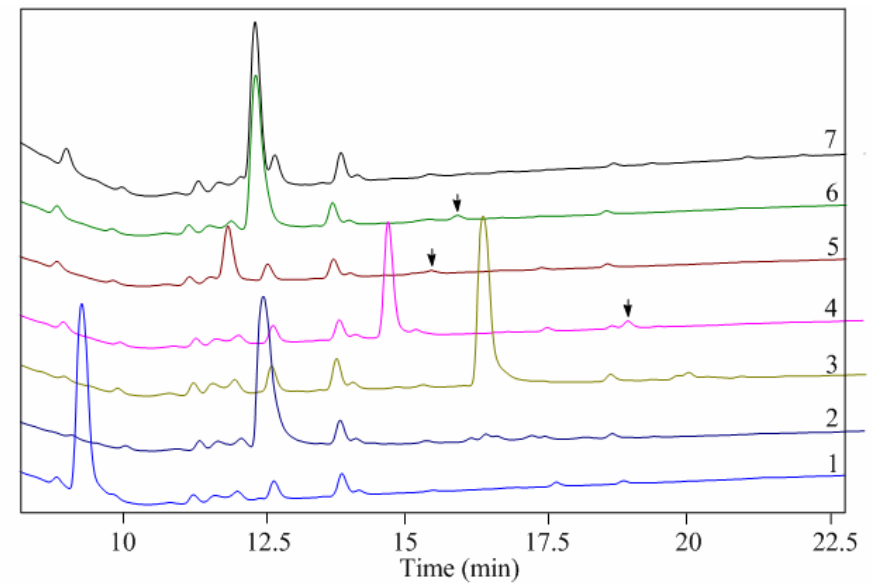


Figure 3-17. LC-MS analysis of NotA reactions with several different substrates. 1: L-Trp, 2: doubly ^{13}C-labeled brevianamide F, 3: deoxybrevianamide E, 4: doubly ^{13}C-labelled keto-premalbrancheamide, 5: premalbrancheamide, 6: preparaherquamide, 7: VM55599. All peaks labeled with black arrows showed the expected *m/z* values of singly prenylated products.

VM55599. Both premalbrancheamide and preparaherquamide were utilized by NotA, albeit at the low level, while VM55599 was not an enzyme substrate. The *m/z* values of 404.15 and 418.00 indicated the formation of singly prenylated premalbrancheamide and preparaherquamide, respectively. Moreover, this study possibly suggested that VM55599 was the dead end metabolite in paraherquamide/asperparaline biosynthetic pathway.

NotD showed the relatively high activity toward doubly ^{13}C-labeled brevianamide F (**Figure 3-18**). The product peak retained the same retention time (17.38 min) as deoxybrevianamide E and the peak content showed the m/z value at 354.22, 2-Da shift from that of authentic deoxybrevianamide E in MS analysis. In MS^2 analysis, the NotD product showed the same fragmentation pattern as that of deoxybrevianamide E and the m/z differences (1 or 2 Da) of some fragments in two MS^2 spectra came from the two labeled ^{13}C atoms in brevianamide F (**Figure S3-12**). Subsequently, NotD substrate selectivity was investigated with more substrates (**Figure S3-11**). L-Trp, deoxybrevianamide E, and doubly ^{13}C-labelled keto-premalbrancheamide were used to probe the timing of NotD reaction in notoamide/stephacidin biosynthesis. None of them was utilized by NotD as determined in LC-MS analysis, indicating that NotD is the long-sought deoxybrevianamide E synthase and catalyzes the committed step to produce the key common biosynthetic intermediate shared by the family of unique prenylated indole alkaloids. Furthermore, cyclo-(L-Phe-L-Pro), cyclo-(L-Trp-L-Trp), and cyclo-(L-Trp-L-Tyr) were used as the NotD substrates to determine enzyme substrate selectivity. Singly or doubly prenylated chemicals were not detected in enzyme reaction mixtures by LC-MS. Thus, both amino acid residues in brevianamide F are restrictively critical for substrate binding to NotD. However, amino acid residues in NotD contributing to substrate recognition remain unknown and X-ray studies of NotD will be necessary to provide certain information to understand this process and facilitate future enzyme engineering.

Interestingly, NotD exhibited similar catalytic activity toward brevianamide F over a broad range of temperature (4 to 42 °C) and in buffers with pH values from 6.0 to 9.0 (**Figure S3-13**). The formation of product in the enzyme reaction was linear with up

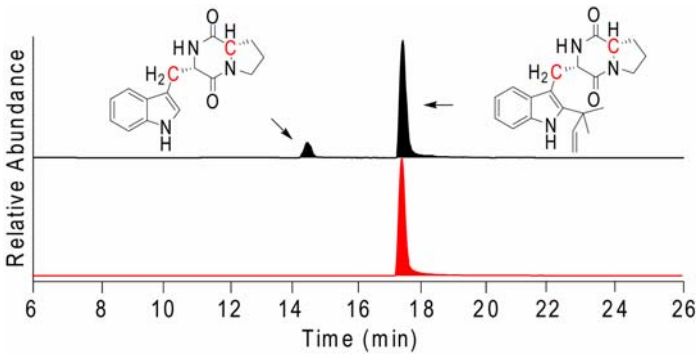


Figure 3-18. LC spectrum of NotD reaction with doubly ^{13}C-labeled brevianamide F as substrate. The enzyme product had the same retention time (17.38 min) as authentic deoxybrevianamide E.

to 2 µg protein per 100 µl assay for up to 15 min under the above optimal conditions. The enzyme activity was not dependent on any additive divalent cation while addition of 5 mM Mg^{2+}, Ca^{2+} or Mn^{2+} slightly enhanced NotD activity by about 106% (**Figure 3-19**). Two divalent metal ions (Co^{2+} and Ni^{2+}) at 5 mM inhibited the enzyme activity by 5% to 20% and 5 mM Cu^{2+}, Zn^{2+}, Fe^{2+}, or Sn^{2+} significantly reduced enzyme activity to 12% to 35% (**Figure 3-19**). Unlike to MaPT, 5 mM EDTA caused only minor effect on NotD activity (remaining 95%), indicating that the NotD reaction pocket might be less exposed to solvent.

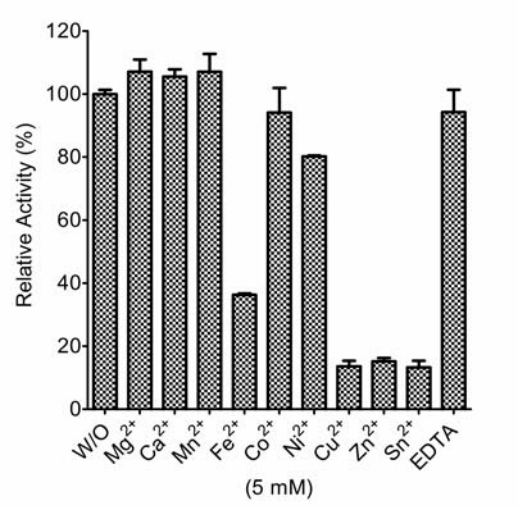


Figure 3-19. Metal dependence of NotD. Addition of 5 mM divalent metal ions or EDTA caused various effects on enzyme activity. Experiments were conducted in duplicate.

In this study, HPLC-UV was used to investigate NotD kinetics with doubly ^{13}C-labeled brevianamide F coupled with DMAPP as enzyme substrates. The enzyme reaction followed Michaelis-Menton kinetics when doubly ^{13}C-labeled brevianamide F concentration was varied (1 - 90 μM) and DMAPP concentration held constant at 0.1 mM (**Figure S3-14**). The K_m and V_{max} values for doubly ^{13}C-labeled brevianamide F were 4.33±0.43 μM and 0.89±0.02 μM/min, respectively, giving a maximal turnover value of 19.1±0.4 min^{-1}. Similarly, the DMAPP concentration was varied (2 - 60 μM) while the other substrate concentration was kept as 0.1 mM. Its K_m and V_{max} values for DMAPP were 1.31±0.22 μM and 1.18±0.03 μM/min, respectively. Its maximal turnover value (25.3±0.6 min^{-1}) and enzyme catalytic efficiency value (19.31 $\mu M^{-1} \bullet min^{-1}$) were slightly higher than those of doubly ^{13}C-labeled brevianamide F (19.1±0.4 min^{-1}, 4.41 $\mu M^{-1} \bullet min^{-1}$, respectively). Compared to FtmB (with brevianamide F as its substrate, K_m = 55 μM, k_{cat}/K_m = 6.08 $\mu M^{-1} \bullet min^{-1}$), NotD showed more restricted substrate selectivity and greater substrate binding affinity but a similar level of enzyme catalytic efficiency [42].

Recently, the FgaPT2 crystal structure was solved to provide sights for understanding the reaction mechanism of this group of metal-independent aromatic prenyltransferases [43]. In my studies, three key amino acid residues located in enzyme reaction pocket were mutated to test their importance in this enzyme. The NotD amino acid sequence was first aligned to MaPT (ABZ80611), FtmB (AAX56314), FtmH (BAH24002), FgaPT1 (EAL94098), FgaPT2 (AAX08549), DMATSs from *A. fumigatus* (EAL94103), *Balansia obtecta* (Q6X1E1), and *Claviceps fusiformis* (Q12594) by ESPript (**Figure S3-15**). Both K212 and K282 in NotD (corresponding to K187 and K189 in FgaPT2) were identified as two possible key residues involved in DMAPP binding and their roles in FgaPT2 reaction have been validated by mutagenesis and structural studies [43, 44]. In FgaPT2 crystal structure, R100 was suggested as the key residue to interact with the pyrophosphate group in DMAPP, and subsequent mutagenesis studies confirmed this role [43]. R108 in NotD is the corresponding residue, and it was mutated into both His and Gly to test for its function in reverse prenyltransferase (**Figure S3-10**). Both NotD lesions resulted in less than 2% of wild type catalytic activity (**Figure 3-20**). Another highly conserved residue in NotD suspected to be involved in substrate binding is E108,

corresponding to E89 in FgaPT2. This residue in FgaPT2 forms one H-bond with N-H of L-Trp indole ring [43]. It was expected that E108 will have a similar function in NotD since brevianamide F also contain the L-Trp structural moiety. Both NotD E108D and E108G mutants had at least 92 % less activity than wild type. The lost function of NotD E108D indicated that the proper length of acidic residue side chain is critical for H-bond formation in this group of aromatic prenyltransferases and NotD may lack structural plasticity to accommodate aromatic substrate analogs. Indeed, two E108 mutants showed no activity toward cyclo-(L-Phe-L-Pro) with a phenyl ring instead of indole ring (data not shown). The FgaPT2 structure also revealed the presence of one aromatic network consisting of five Tyr residues to shield the reactive DMAPP carbocation from solvent [43]. All of these aromatic residues are conserved in NotD, except that Y413 is replaced with W424 in NotD (**Figure S3-15**). This tryptophan residue was mutated into Tyr and Gly in my study. Although the W424G lesion lost at least 98 % of catalytic activity, W424Y was capable to produce deoxybrevianamide E with about 25 % efficiency in comparison to wild type enzyme. This observation validated the importance of the aromatic network in the aromatic prenyltransferase reaction and also suggested that W424 was not the sole determining factor for regulating regular or reverse prenylation in NotD.

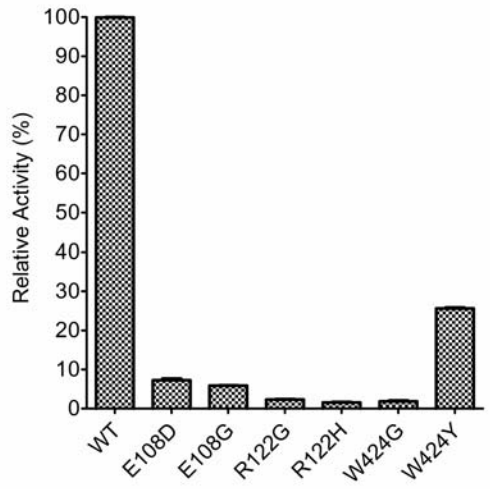


Figure 3-20. Catalytic activities of wild type NotD and its mutants. In enzyme reactions, 2 µg of enzyme was used in a 100-µl reaction mixture containing 0.1 mM doubly ^{13}C-labeled brevianamide F and 0.1 mM DMAPP. The reactions were conduced at room temperature for 1.5 h and performed in duplicate.

Overall, my studies in this section isolated and biochemically confirmed one putative gene cluster for notoamide/stephacidin biosynthesis. Bioinformatic analysis suggested that NotC is a brevianamide F synthase, similar to FtmA in ergot alkaloid biosynthesis. Subsequently, the brevianamide F product will be singly prenylated in a NotD-catalyzed reaction (**Figure 3-21**). One oxidoreductase in the gene cluster will then oxidize doxybrevianamide E by 2 electrons. Although the subsequent heteroatom Diels-Alder cyclization in prenylated fungal alkaloid biosynthesis has been proposed for a long time, such an enzyme remains unknown here. Since NotA showed limited catalytic efficiency toward keto-premalbrancheamide, it is proposed that a P450-catalyzed hydroxylation on the indole ring may occur prior to this regular prenylation reaction. The subsequent ring closing might be catalyzed by the other P450 to produce stephacidin. Recently, a notoamide E feeding experiment indicated that this biosynthetic intermediate might not be the direct precursor of notoamide A [45]. Thus, stephacidin may be the

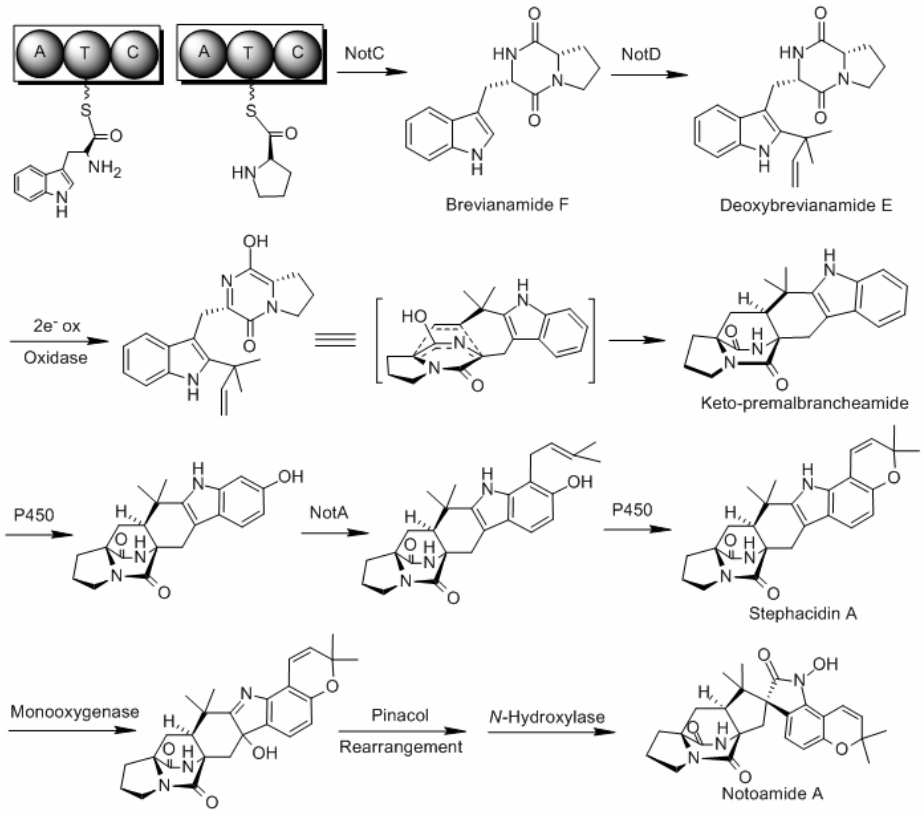


Figure 3-21. One putative biosynthetic route for stephacidin and notoamide.

intermediate in notoamide biosynthetic pathway. I propose that stephacidin will be further hydroxylated by one monooxygenase, and it will then undergo a pinacol rearrangement. It is unknown whether such a pinacol rearrangement will automatically occur under fungal physiological conditions or will be catalyzed by one enzyme. Notoamide B will be produced in a pinacol rearrangement reaction and will be subjected to an *N*-hydroxylation reaction to produce the final product notoamide A. It is believed that this prototypical biosynthetic pathway will lead to a better understanding of how other members in the family of prenylated fungal alkaloids are biosynthesized, and may offer opportunities to generate numerous natural product analogs that are beneficial to drug discovery and development.

3.4. Materials and Methods

General Chemicals, DNA Sub-cloning, and Bacterial Strains. Chemicals were purchased from Sigma-Aldrich, Fisher Scientific and BaChem while acetonitrile used in LC-MS^n analysis was LC/MS grade. A MilliQ H_2O purification system generated water for LC-MS^n analysis. Trifluoroacetic acid (99%, reagent plus) and formic acid (>98%, ACS reagent) were also purchased from Sigma Aldrich. DMAPP was purchased Isoprenoids. Authentic paraherquamide A, VM55599, and pre-paraherquamide were synthesized following previously published procedures. Paraherquamide B (as the unnatural enantiomer) was obtained by total synthesis. Authentic malbrancheamide was kindly provided by Professor Rachel Mata of Universidad Nacional Autónoma de México while Dr. Anthony E. Glenn of the USDA for provided *M. aurantiaca* RRC1813 strain (originally obtained from Dr. María del Carmen González). *P. fellutanum* (ATCC20841) was purchased from American Type Culture Collection (Manassas, VA) while *A. japonicus* JV-23 was provided by Dr. Hideo Hayashi of Osaka Prefecture University. Standard methods for DNA isolation and manipulation were performed as described by Sambrook *et al* [46]. Genomic DNA from *M. aurantiaca* RRC1813 was isolated with MasterPure Yeast DNA Purification kit (Epicentre Biotechnologies) as described in the manual. The MaPT GenBank Accession Numbers are EU420001 for its genomic DNA and EU420002 for its putative cDNA. Molecular biology reagents and

enzymes were supplied by New England Biolabs with the exception of for *Pfu* DNA polymerase (Stratagene), dNTPs (Takara), T4 DNA Ligase (Invitrogen). *Escherichia coli* XL-1 Blue was used for cloning and plasmid harvesting while *E. coli* BL21 (DE3) or *E. coli* BL21 CodonPlus-(DE3)-RIPL was used for protein overexpression. All *E. coli* strains were grown in Luria-Bertani broth. DNA sequencing was performed at the University of Michigan DNA Sequencing Core. A Beckman Coulter HPLC, fitted with an XBridge C18 column (5μm, 4.6 x 250 mm), coupled with a System Gold I68 Detector was used for HPLC-UV analysis. A SHIMADZU LCMS-2010EV system was used for LC-MS analysis in the studies.

Cultures of Fungal Strains. *P. fellutanum* and *A. japonicus* JV-23 were initially grown in solid medium (20 g malt extract, 20 g glucose, 1 g peptone, and 20 g agar in 1 L deionized water) and solid (20 g potato-dextrose-broth and 20 g agar in 1 L deionized water), respectively, at 25 °C in dark for 2 weeks. Fungal mycelium and spores were then transferred into 300 ml sterile corn steep liquor medium (22g corn steep liquor and 40 g glucose per liter deionized water) for *P. fellutanum* or 300 ml sterile potato-dextrose-broth (PDB) (24 g potato-dextrose-broth per liter deionized water) for *A. japonicus* JV-23, in 2 L Erlenmeyer flasks. Both fungal strains were then grown at 25 °C in dark for 4 weeks. *Malbranchea aurantiaca* RRC1813 was cultured in potato dextrose broth (PDB) at 28 °C, 200 rpm under dark for 14 days.

Sample Extraction. The cultures with *P. fellutanum* and *A. japonicus* JV-23 mycelia were adjusted to pH 10-12 by 10 M KOH. The cultures were then extracted with the equal volume of ethyl acetate twice. The combined organic layer from each culture was washed with water, dried over anhydrous magnesium sulfate, and evaporated to dryness. The residues were re-dissolved in methanol prior to LC-MS^n analysis. Malbrancheamides and pre-malbrancheamide were isolated by extracting the fungal culture with equal volume of ethyl acetate three times following *in vacuo* concentration. The dried residue was stored at -20 °C until ready to use.

Feeding ^{13}C-Labeled Chemicals to *M. aurantiaca*. The feeding experiment followed a modified protocol previously described [18]. Briefly, the fungal strain was initially grown in PDB for 5 days. Then the mycelium from 500 ml culture was collected and rinsed with 100 ml of 50 mM Tris-Cl, pH 7.2. Doubly ^{13}C-labeled chemicals were dissolved in methanol (1.25 mg/ml) and then 2.5 mg of each compound was added to 50 ml of trace element solution (35 mM $NaNO_3$, 5.7 mM K_2HPO_4, 4.2 mM $MgSO_4$, 1.3 mM KCl, 36 μM $FeSO_4$•$7H_2O$, 25 μM $MnSO_4$•H_2O, 7 μM $ZnSO_4$•$7H_2O$, 1.5 μM $CuCl_2$•$2H_2O$) with 0.1% Tween-20. The mixture was sonicated for about 15 min to aid in the formation of micelle, in which the mycelium from 500 ml culture was resuspended. The fungal strain was further grown at 28 °C, 200 rpm under dark for 7 more days. The secondary metabolites were extracted with 100 ml of ethyl acetate twice. The aqueous-organic mixtures were stirred for about 30 min to facilitate the extraction. Fungal mycelium was then removed with Waterman filter paper. The organic layers were treated as above. ^{13}C-labled malbrancheamide B was purified by semi-preparative HPLC. An Alltech Econosil C18 10μ (250 mm x 10 mm) was used with the flow rate of 3 ml/min in an isocratic 50 % methanol in 50 % water with 0.1 % TFA over 25 min. The compound was collected with detection wavelength 233 nm. The isolated compound was lyophilized and subjected for LC-MS analysis.

LC-MS^n Analyses. LC-MS^n analyses were performed by using a ThermoFinnigan LTQ linear ion-trap instrument equipped with electrospray source and Surveyor HPLC system at room temperature. Separations were carried out with a Waters XBridge^{TM} C18 (3.5 μm, 2.1x150mm) column at a flow rate of 210 μl/min with solvent A (water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid). Solvent B was kept at 15% in solvent A for 2.5 min and then was gradually increased to 40% over 12.5 min, 80% over 2 min, and then was kept at 80% for 6 min to elute fungal metabolites. The column was further re-equilibrated with 15% solvent B for 25 min. For mass spectrometry, the capillary temperature was set to 275°C with the source voltage at 3.6 kV, the source current at 3.5 μA, the capillary voltage at 30 V, and the tube lens at 119 V. Sheath gas flow was set to 28 psi and auxiliary gas flow was to be 5 arbitrary units. The normalized collision energy for ion fragmentation was 30%. The injection volume was 5-10 μl and

spectra were recorded in the positive ion mode. Selective ion monitoring (SIM) chromatographs were achieved with the selected *m/z* values.

Structural Analysis of Naturally Isolated Malbrancheamide B. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.41 (s, 1 H), 7.32 (d, *J* = 8.4 Hz, 1 H), 7.27 (d, *J* = 1.7 Hz, 1 H), 6.95 (dd, *J* = 8.4, 1.7 Hz, 1 H), 3.36 (s, 1 H), 3.27 (d, *J* = 10.0 Hz, 1 H), 2.95 (m, 1 H), 2.76 (s, 2 H), 2.43 (m, 1 H), 2.13 (d, *J* = 9.9 Hz, 1 H), 2.10-1.70 (comp, 6 H), 1.32 (s, 3 H), 1.26 (s, 3 H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 173.1, 142.6, 136.8, 125.3, 125.2, 118.7, 118.5, 110.3, 103.7, 64.1, 58.5, 55.3, 53.9, 47.0, 34.0, 31.1, 30.0, 28.7, 26.6, 23.7, 22.5; IR (neat) 3297, 1652, 1457 cm^{-1}; HRMS (TOF+) calcd for $C_{21}H_{25}N_3OCl$ (M+H) 370.1680, found 370.1670.

MaPT Analytical Methods. A Beckman Coulter HPLC, fitted with an XBridge^{TM} C18 column (5μm, 4.6 x 250 mm), coupled with a System Gold I68 Detector was routinely used for HPLC-UV analysis. The detection wavelength was set at 269 nm. A linear gradient of 20 – 70 % acetonitrile in 0.1 % TFA over 20 min was used for product detection. One 10-minute re-equilibration program with 20% acetonitrile in 0.1 % TFA followed each run. A SHIMADZU LCMS-2010EV system was used for LC-MS analysis with a linear gradient of 20 – 80 % acetonitrile in 0.1 % FA over 15 min. An XBridge^{TM} C18 (3.5 μm, 2.1x150mm) column was used with a flow rate of 208 μl/min. Varian Inova 500 MHz or Bruker DRX 500 MHz instruments were used to record NMR spectra. VG (Micromass) 70-250-S Magnetic Sector Mass Spectrometer in ESI^{+} mode was used to record high resolution mass spectra.

Isolation of MaPT Gene. Degenerate primers were designed from conserved sequences of the new class of PTs. The pair of PT-FW and PT-RV primers was used to screen putative PTs from isolated *M. aurantiaca* RRC1813 genomic DNA by PCR (**Table S3-3**). Fragments of about 750 bp were subcloned into pGEM-Teasy vector (Promega) and sequenced. BLASTX in NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) was used to identify sequenced PT fragments. TAIL-PCR was applied to amplify fragments extended from both ends of PT gene to produce full length genomic DNA [26]. The long specific

primers were designed on the known sequence, while two arbitrary primers were first used by Liu and coworkers [26]. All primers used in TAIL-PCR are listed in Supplement Table S1. BCM Gene Finder (http://www.bioscience.org/urllists/genefind.htm) was used to predict putative introns in the full-length PT gene. A 63-bp intron was predicted and removed with primers of InterFW and InterRV along with PTFW(NdeI) and PTRV(BamHI) to produce the MaPT cDNA. Similarly, *Aspergillus* MF297-2 genomic DNA was isolated and used as the template to generate both *NotA* and *NotD* cDNAs with the above overlapping PCR technology.

Gene Cloning, Expression and Protein Purification. The purified cDNAs were ligated to pET28a for transformation of *E. coli* DH5α competent cells. The constructs were isolated and submitted for sequencing to exclude mutations introduced during PCR amplification and gene manipulation. The constructs was transferred into *E. coli* BL21 (DE3) competent cells or BL21 CodonPlus-(DE3)-RIPL for protein expression. Cells harboring the construct were cultured in LB medium containing 25 μg/ml of kanamycin and grown at 37 °C in a shaker at 200 rpm until OD_{600} reached 0.6. The cultures were then induced by IPTG with a final concentration of 0.2 mM. The cultures were further grown at 16 °C in a shaker at 180 rpm for 16 hours. The cells were harvested by centrifugation (7250 *g*, 12 min, and 4 °C) and cell pellets were stored in -80 °C or directly used for protein purification. For MaPT, cell pellet was resuspended in 50 ml lysis buffer [1 X PBS, 3 mM β-mercaptoethanol (BME), 10 % glycerol, 20 mM imidazole, pH 7.4], which was sonicated to release soluble proteins. The supernatant after centrifugation at 35,000 x *g* at 4 °C for 36 min was incubated with 5 ml pre-equilibrated Ni-NTA agarose resin (Qiagen) at 4 °C for 2 h. The resin was then washed with 150 ml lysis buffer and recombinant His_6-MaPT was eluted with 10 ml 200 mM imidazole in 1 X PBS (pH 7.4), 3 mM BME, and 10 % glycerol. The protein was further purified with Superdex 200 column (Amersham Biosciences) with gel chromatography running buffer (1 X PBS, 1 mM DTT, 10 % glycerol, pH 7.4). Protein absorbance at 280 nm was measured for calculating its concentration with an extinction coefficient of 88490. The protein was then aliquoted and stored at -80 °C. The final yield of His_6-MaPT was 2.5 mg/L. For NotA and NotD purification, the cell pellets were resuspended in 40 ml

binding buffer (50 mM Tris-Cl, pH 7.7, 200 mM NaCl, 3 mM β-mercaptoethanol, 10 % glycerol) with 20 mM imidazole. The resuspended cell solutions were sonicated to release the soluble proteins. The soluble supernatants were collected and incubated with 2.5 ml pre-equilibrated Ni-NTA agarose resin (Qiagen) after centrifugation at 36 000 *g* at 4 °C for 32 min. After 2-hour incubation at 4 °C with gentle shaking, the resin was washed with 50 ml binding buffer with 20 mM imidazole, then 50 ml binding buffer with 30 mM imidazole, and finally 80 ml binding buffer with 40 mM imidazole. The recombinant His_6-NotA and NotD were gradually eluted with elution buffer (50 mM Tris-Cl, pH 7.7, 50 mM NaCl, 3 mM β-mercaptoethanol, 10 % glycerol). Subsequent desalting was attained by buffer exchange into elution buffer with a PD-10 column. The protein was then aliquoted and stored at -80 °C.

Preparation of Enzyme Mutants. MaPT, NotA, and NotD gene mutants were prepared following the protocols from Stratagene QuikChange® II Site-Directed Mutagenesis Kit. The mutagenesis primers were shown in **Table S3-2** and **S3-3**. The same procedures for preparation of wild type proteins were followed to produce their mutants except that only Ni-NTA agarose resin was used for MaPT mutant purification.

Protein Analysis. The subunits of prenyltransferase were analyzed by 4-12 % SDS-PAGE and protein bands were stained with Commassie brilliant blue R-250. The molecular weight of MaPT with His_6-tag at its N-terminus was determined by Superdex 75 column (Amersham Biosciences) with gel chromatography running buffer. The column was calibrated with bovine serum albumin (66 kD), chicken egg ovalbumin (43 kD), bovine trypsinogen (25 kD), and ribonuclease A (13. 7 kD). The molecular weight of MaPT was calculated as 61.6 kD. For NotD, Superdex S-200 column (Amersham Biosciences) was used with running buffer (50 mM Tris-Cl, pH 7.5, 200 mM NaCl, 10 % glycerol, with or without 1 mM DTE).

Enzyme Assays. **A). MaPT** The 100-µl reaction mixture routinely contained 5 mM $MgCl_2$, 0.25 mM L-Trp, 0.2 mM DMAPP, and 1 µl of 0.86 µg/µl MaPT in reaction buffer (1 X PBS, pH 7.4, 10 % glycerol, and 3 mM BME). Enzyme reactions were initiated by

adding MaPT after pre-warming the other components at 30 °C for 1 min and then kept at 30 °C for 30 min. Enzyme boiled at 100 °C for 10 min was used as a negative control. In the mutagenesis studies, MaPT mutants were used as the enzyme source. The reaction was stopped with 10 µl of 1.5 M TCA. The mixture was agitated well and centrifuged at 13,000 g for 5 min. The enzyme products were detected by HPLC. In order to investigate substrate specificity, L-Trp was substituted with a series of derivatives. IPP, GPP, or FPP also substituted DMAPP in the reactions. **B). NotA and NotD**. The 100-µl reaction mixture contained 1 µg of NotA or NotD, 5 mM $MgCl_2$, 0.1 mM brevianamide F, 0.1 mM DMAPP in the reaction buffer (50 mM Tris-Cl, pH 7.5, 10 % glycerol, and 3 mM BME). The reaction was initiated by adding enzyme after pre-warming the other components at room temperature for 1 min. After mixing well and briefly centrifuging, the reactions were further incubated at room temperature for 45 min and stopped with 15 µl of 1.5 M TCA. The mixtures was mixed well and centrifuged at 13000 g for 5-10 min. The mixture was mixed well and centrifuged at 13000 g for 5 min. To detect product, an aliquot of 100-µl solution was subjected to HPLC with detection at 222 nm. Solvent B (acetonitrile in 0.1 % TFA) was increased from 20 % to 30 % for 5 min and then increased to 80 % in next 20 min. One 5-minute re-equilibration program with initial B concentration was followed after linear gradient. All experiments were performed in duplicate. The area under product peak was measured with Beckman Software and the content in the peak was converted into nano-moles with deoxybrevianamide E standard curve.

Metal Dependence of Prenyltransferases. 5 mM of divalent metal ions (Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, and Sn^{2+}) were added to the routine reaction mixture. 5 mM EDTA, EGTA, ethylenediamine-*N,N'*-diacetic acid (EDDA), and 1,3-diamino-2-propanol-*N,N,N',N'*-tetraacetic acid (DHPTA) were added to reaction buffer containing 0.86 µg MaPT and the mixture was incubated at 30 °C for 5 min. Then 0.25 mM L-Trp and 0.2 mM DMAPP were added to bring the volume to 100 µl and initiate the reaction, which was further incubated at 30 °C for 30 min before terminating with 10 µl of 1.5 M TCA. The products were detected by HPLC, and experiments were performed in duplicate. For NotD, the reaction mixtures were incubated at room temperature.

Optimal Conditions for Enzyme Reaction. A). MaPT. For investigation of optimal temperature, reactions were conducted under 4 °C, 15 °C, 23 °C, 30 °C, 37 °C, and 42 °C in the reaction buffer described above. For investigation of optimal reaction buffer pH value, 50 mM Tris-Cl or sodium phosphate with various pH values (5, 6, 7, 7.4, 8, 9, and 10) were used in the reactions at 30 °C. The same procedure for routine enzyme assay was followed in this section. **B). NotD.** For investigation of optimal temperature, reactions were conducted under 4 °C, 20 °C, 26 °C, 30 °C, 37 °C, 42 °C, and 55 °C in the reaction buffer described above. For investigation of optimal reaction buffer pH value, 50 mM Tris-Cl or sodium phosphate with various pH values (4, 5, 6, 7, 7.5, 8, 9, and 10) were used in the reactions at room temperature. The same procedure for routine enzyme assay was followed in this section.

Kinetics Analysis. A). MaPT. In order to find suitable MaPT concentration linearly correlated with product generation, 0.1, 0.25, 0.5, 1, 2, 4, and 8 µl of 0.86 µg/µl enzyme was added to reaction mixtures. The reaction was incubated at 30 °C for 6 min before terminating with 10 µl of 1.5 M TCA. In order to determine the time course of enzyme reactions, 50 µl of reaction mixture was removed and mixed with 5 µl of 1.5 M TCA at 1 min, 3 min, 5 min, 8 min, 12 min, 16 min, 20 min, 25 min, 30 min, and 35 min. The enzyme product was detected with the same method described above.

Reaction mixtures in the enzyme kinetic analysis contained 0.22 µg MaPT, and 5 mM $MgCl_2$. To determine kinetic parameters of L-Trp, 5-OH-L-Trp, and L-abrine, 0.2 mM DMAPP was included in the reaction and 2-600 µM L-Trp, 5-1500 µM 5-OH-L-Trp, or 5-1000 µM L-abrine was added. To determine kinetic parameters of DMAPP, 0.25 mM L-Trp, 2 mM 5-OH-L-Trp, or 1 mM L-abrine was included in the reaction while DMAPP final concentration was varied (1-100 µM, 5-200 µM, 5-160 µM, respectively). The reaction mixtures without L-Trp, 5-OH-L-Trp, L-abrine, or DMAPP were pre-warmed at 30 °C for 1 min. The reactions were initiated by mixing them with the single omitted component and were further incubated at 30 °C for 5 min and 6 min for L-Trp and 5-OH-L-Trp or L-abrine, respectively, before terminating with 10 µl of 1.5 M TCA. The

mixtures were mixed well and centrifuged at 13,000 x *g* for 5 min. A 100 µl solution was subjected to HPLC for product detection. All experiments were performed in duplicate. The area under the product peak was measured and the peak content amount was calculated using a dimethylallyltryptophan (DMAT) standard curve. The initial rate was calculated with equation of product (nmol)/100 µl x 110 µl/100 µl. The data were fit to the Michaelis-Menton equation or the substrate inhibition equation ($v = V_{max}[S]/(K_m + [S](1+[S]/K_s))$. For the competitive kinetic analysis, 0-100 µM D-Trp was included in the reactions containing varied concentration of L-Trp or DMAPP. 100 µM L-Phe was also used as the putative inhibitor in the assay.

B). NotD. The NotD concentration and reaction time were investigated to examine the linear correlation with product generation. The 100-µl reaction mixture contained 0.25 µg of NotD and 5 mM $MgCl_2$ in the reaction buffer (50 mM Tris-Cl, pH 7.5, 10 % glycerol, and 3 mM BME). For investigation of brevianamide F kinetic parameters, 0.1 mM DMAPP was included in the reaction and brevianamide F final concentration was varied (1, 2, 3, 5, 70, 10, 30, 60, and 90 µM). For determination of DMAPP kinetic parameters, 0.1 mM brevianamide F was included in the reaction while DMAPP final concentration was varied (2, 5, 7, 10, 30, and 60 µM). The reactions without brevianamide F or DMAPP were pre-warmed at room temperature for 1 min and initiated with mixing with the omitted component. After mixing well and briefly centrifuging, the reactions were further incubated at room temperature for 5 min and stopped with 15 µl of 1.5 M TCA. The mixture was mixed well and centrifuged at 13000 g for 5 min. To detect product, an aliquot of 100-µl solution was subjected to HPLC with detection at 220 nm. All experiments were performed in duplicate. The area under product peak was measured with Beckman Software and the content in the peak was converted into nano-moles with deoxybrevianamide E standard curve. The initial rate was calculated with product (nmol)/100 µl * 115 µl / 100 µl. The data were fit to Micheal-Menton equation.

Characterization of MaPT Reaction Products. 2 ml reaction mixtures containing 2 mg of L-Trp, 5-OH-L-Trp, 5-Me-DL-Trp, or L-abrine and 1.2 mg of DMAPP were incubated

at 30 °C overnight and stopped by adding 200 µl of 1.5 M TCA. The mixtures were mixed well and centrifuged at 13,000 x *g* for 5 min. The products were separated by a semi-preparative XBridge^{TM} C18 column with the same program except using a flow rate of 3 ml/min. The isolated products were lyophilized and subjected for ^{1}H and ^{13}C NMR analyses and high resolution MS analysis (HRMS). DMAT: ^{1}H NMR (D_2O, 500 MHz) δ 7.40 (d, *J*= 8.0 Hz, 1H), 7.31 (s, 1H), 7.20 (t, *J*= 8.0, 7.5 Hz, 1H), 6.99 (d, *J*= 7.0 Hz, 1H), 5.37 (s, 1H), 4.11 (dd, *J*= 10.5, 4.5 Hz, 1H); 3.77 (d, *J*= 6.5 Hz, 1H), 3.74 (dd, *J*= 15.5, 4.5 Hz, 1H), 3.27 (dd, *J*= 15.5, 10.5 Hz, 1H), 1.75 (s, 3H), 1.78 (s, 3H). HRMS (ESI^{+}): observed *m/z* = 273.1610; calculated *m/z* = 273.1603. Dimethylallyl 5-OH-L-Trp (DMA5HT): ^{1}H NMR (CD_3OD, 500 MHz) δ 7.09 (d, *J*=3.0 Hz, 1H), 7.07 (s, 1H), 6.74 (d, *J*=8.5 Hz, 1H), 5.12 (s, 1H), 4.92 (s, 1H), 4.15 (dd, *J*= 11.5, 3.5 Hz 1H), 3.81-3.84 (m, 1H), 3.77 (d, *J*= 15.1, 3.6 Hz, 1H), 3.58 (dd, *J*= 21.1, 16.3 Hz, 1H), 3.08 (dd, *J*= 15.4, 11.4 Hz, 1H), 1.81 (s, 3H), 1.66 (s, 3H). ^{13}C NMR (CD_3OD, 500 MHz) δ 171.9, 149.2, 134.5, 132.9, 126.9, 126.3, 123.1, 119.5, 113.8, 110.9, 108.4, 55.4, 30.2, 26.3, 25.9, 18.6. HRMS (ESI^{+}): observed *m/z* = 289.1552; calculated *m/z* = 289.1552. Dimethylallyl 5-Me-DL-Trp (DMA5MT): ^{1}H NMR (CD_3OD, 500 MHz) δ 7.14 (d, *J*=8.5 Hz, 1H), 7.12 (s, 1H), 6.96 (d, *J*=8.5 Hz, 1H), 5.10 (s, 1H), 4.16 (dd, *J*= 11, 4 Hz, 1H); 3.79-3.82 (m, 1H), 3.77 (d, *J*= 4 Hz, 1H), 3.67 (d, *J*= 18 Hz, 1H), 3.12 (dd, *J*= 15, 11 Hz, 1H), 2.31 (s, 3H), 1.83 (s, 3H), 1.71 (s, 3H). ^{13}C NMR (CD_3OD, 500MHz) δ 171.5, 138.0, 133.5, 132.4, 126.7, 126.6, 126.5, 125.9, 125.0, 109.8, 108.0, 55.1, 30.6, 29.8, 25.3, 19.1, 18.5. HRMS (ESI^{+}): observed *m/z* = 287.1762; calculated *m/z* = 287.1760. Dimethylallyl L-abrine (DMAA): ^{1}H NMR (CD_3OD, 500 MHz) δ 7.23 (d, *J*=8 Hz, 1H), 7.19 (s, 1H), 7.04 (t, *J*=15.5, 7.4 Hz, 1H), 6.82 (d, *J*=6.9 Hz, 1H), 5.30 (s, 1H), 4.20 (dd, *J*= 9.5, 5 Hz, 1H); 3.40-3.77 (m, 1H), 3.71 (d, *J*= 4.6 Hz, 1H), 3.34-3.36 (m, 1H), 3.25-3.27 (m, 1H), 2.65 (s, 3H), 1.78 (s, 3H), 1.75 (s, 3H). ^{13}C NMR (CD_3OD, 500 MHz) δ 171.3, 142.3, 139.5, 135.1, 133.6, 125.8, 125.4, 123.4, 121.2, 111.0, 108.4, 63.8, 33.5, 33.1, 29.6, 26.0, 18.4. HRMS (ESI^{+}): observed *m/z* = 287.1766; calculated *m/z* = 287.1660.

LC-MS Analysis of NotD Products. LC-MS analysis was performed by using a ThermoFinnigan LTQ linear ion-trap instrument equipped with electrospray source and Surveyor HPLC system at room temperature. Separations were carried out with a Waters

XBridge^{TM} C18 (3.5 μm, 2.1x150mm) column at a flow rate of 200 μl/min with solvent A (water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid). Solvent B was kept at 2% in solvent A for 4 min and then was gradually increased to 90% over 15 min and then was kept at 90% for 2 min to elute fungal metabolites. The column was further re-equilibrated with 2% solvent B for 10 min. For mass spectrometry, the capillary temperature was set to 275°C with the source voltage at 3.5 kV, the source current at 3.5 μA, the capillary voltage at 30 V, and the tube lens at 119 V. Sheath gas flow was set to 28 psi and auxiliary gas flow was to be 5 arbitrary units. The normalized collision energy for ion fragmentation was 20 %. The injection volume was 19 μl and spectra were recorded in the positive ion mode. Selective ion monitoring (SIM) chromatographs were achieved with the selected *m/z* values.

Analysis of Metal Content of MaPT. Twelve divalent metal ions (Mg^{2+}, Ca^{2+}, V^{2+}, Cr^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Pb^{2+}) in MaPT were detected by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) (Finnigan) in the Department of Geological Science, University of Michigan. The gel chromatography running buffer was used as the blank sample. All detected metal ion content measurements except Ni^{2+} were < 0.05 mol per mol of MaPT.

3.5. Supplementary Figures and Tables

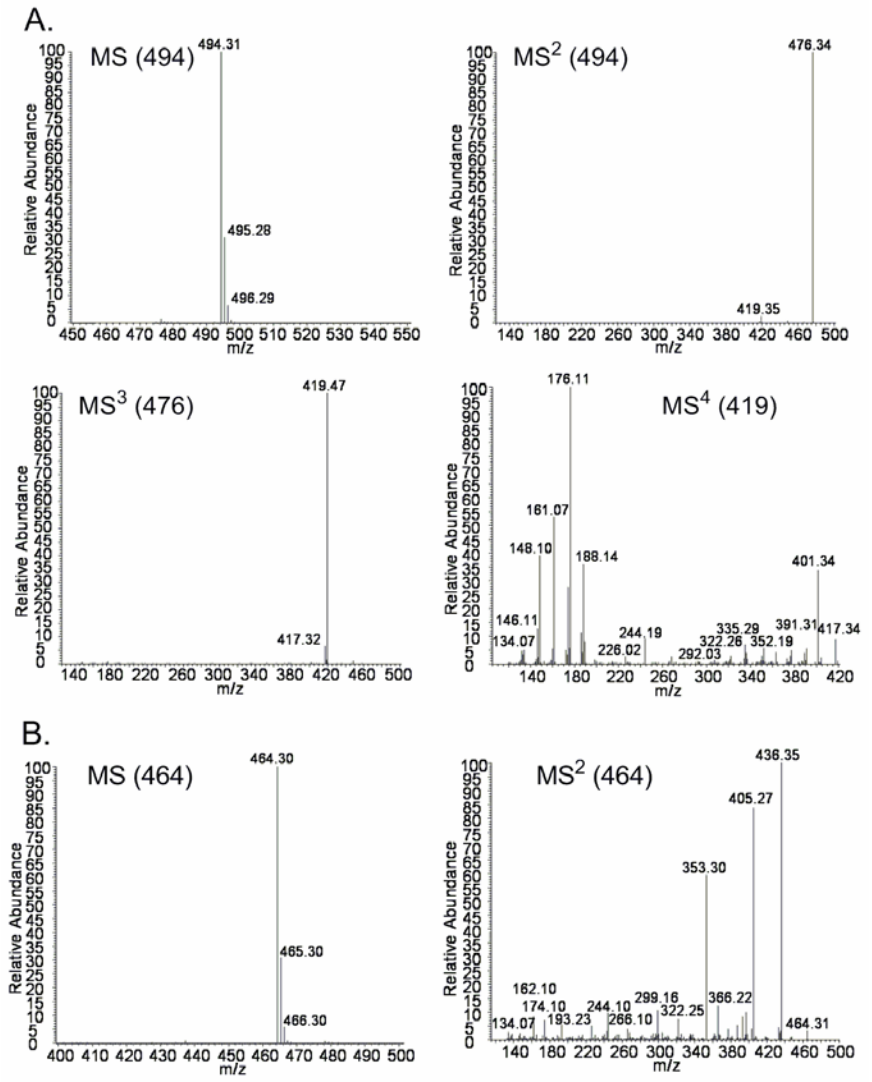


Figure S3-1. MS^n spectra of authentic paraherquamide A (**A**) and authentic paraherquamide B (**B**). These spectra were used to identify both compounds in extracts from isolations of *P. fellutanum* and *A. japonicus* JV-23.

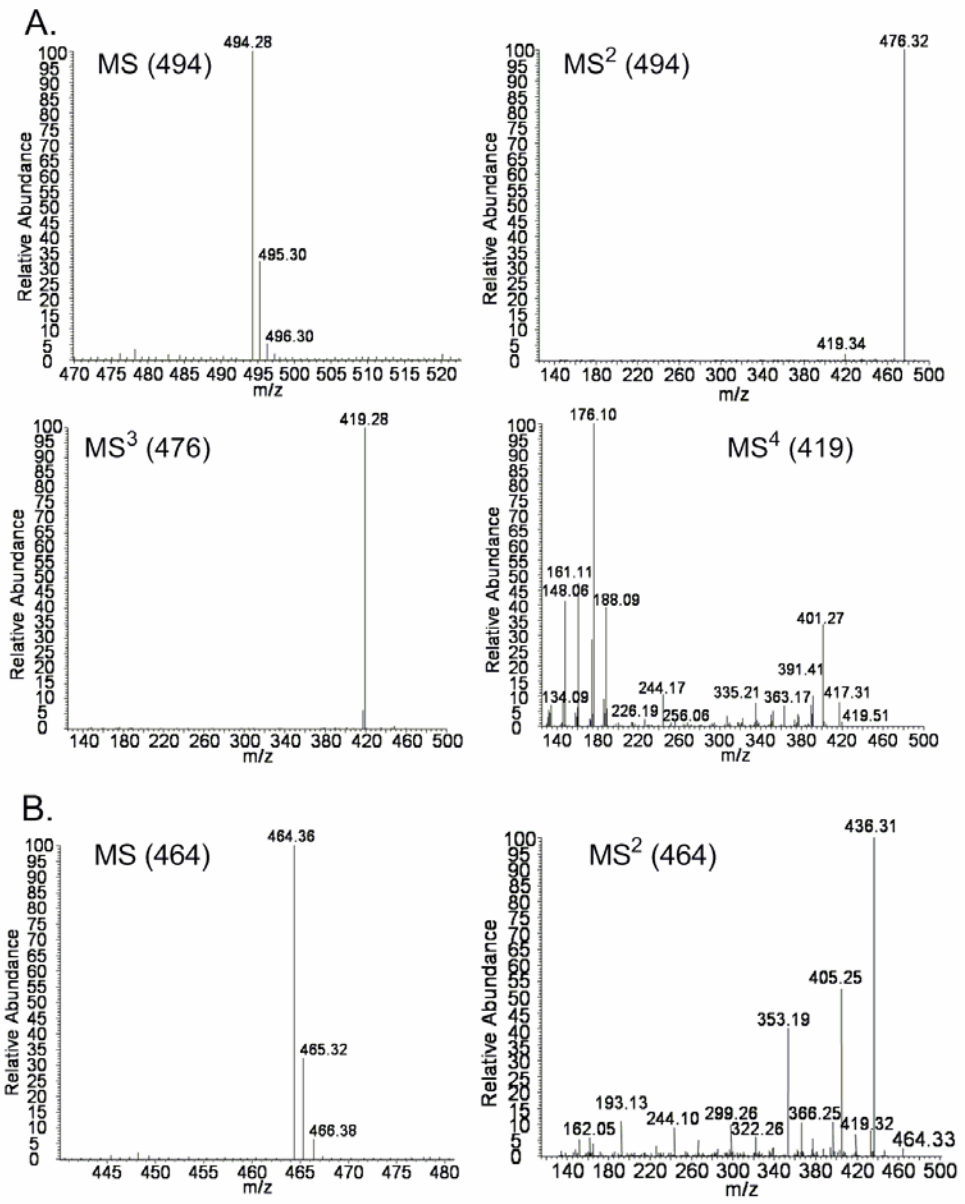


Figure S3-2. MS^n spectra of two metabolites at 14.45 min (**A**) and at 14.88 min (**B**) from the extract from *A. japonicus* JV-23. These metabolites were identified as paraherquamide A (14.45 min) and paraherquamide B (14.88 min) by comparing their MS^n spectra to those of authentic compounds.

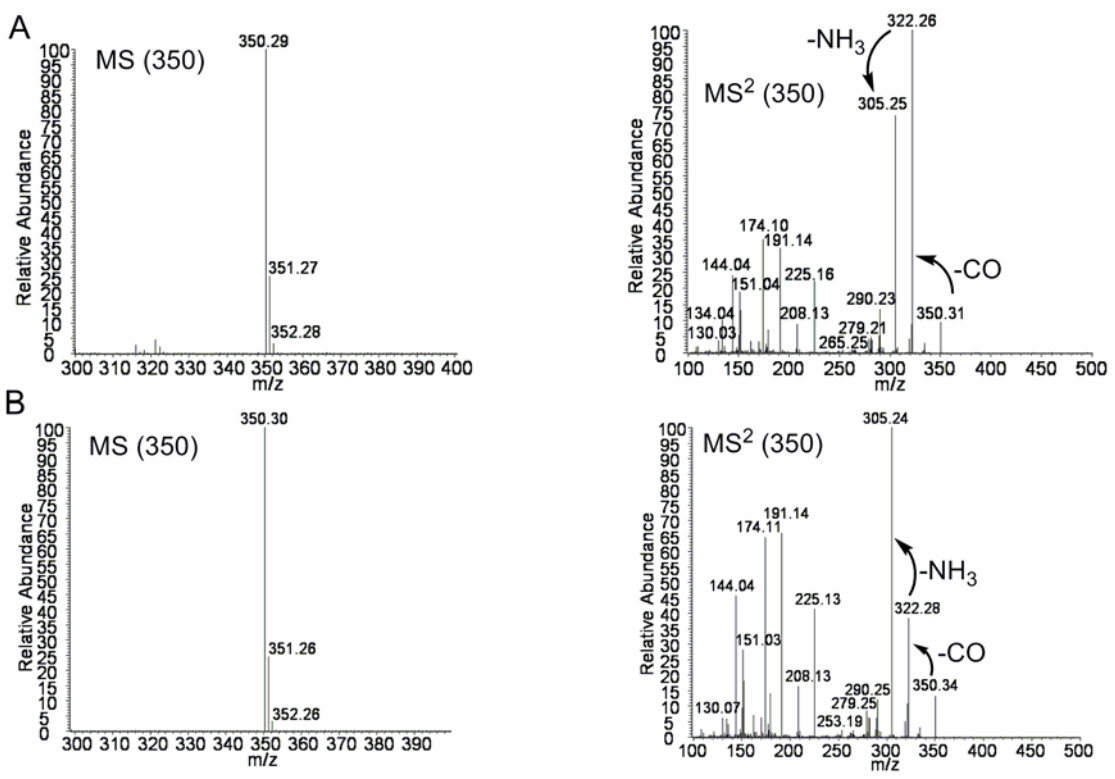


Figure S3-3. MS and MS^2 spectra of authentic VM55599 (**A**) and authentic pre-paraherquamide (**B**). These spectra were used to identify VM55599 and pre-paraherquamide in extracts from isolations of *P. fellutanum* and *A. japonicus* JV-23.

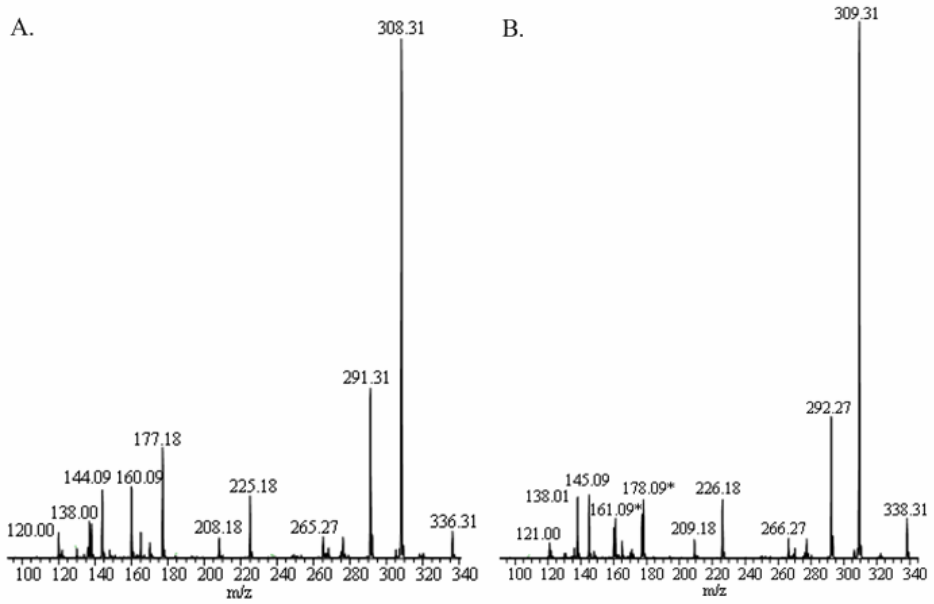


Figure S3-4. MS/MS spectrum of authentic pre-malbrancheamide (A) and doubly ^{13}C-labeled pre-malbrancheamide (B). In the MS/MS spectrum of doubly ^{13}C-labeled pre-malbrancheamide, fragments at *m/z* of 177.19 and 178.09 were observed. Similarly, fragments at *m/z* of 160.01 and 161.09 were also identified.

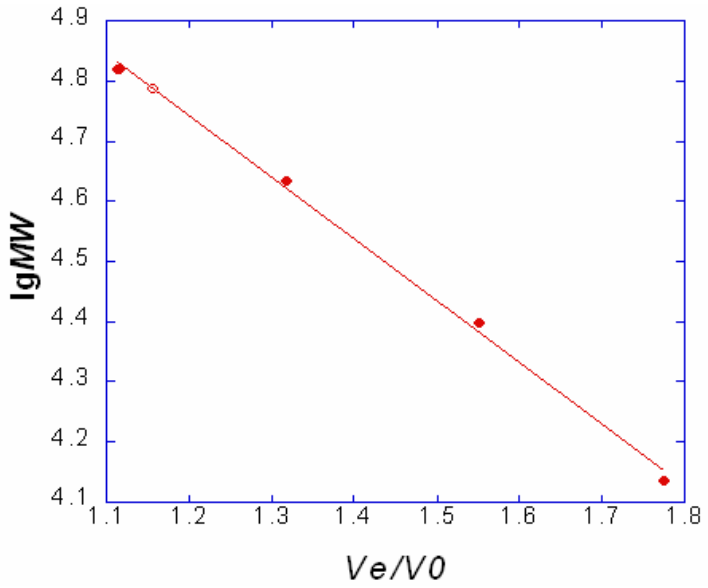


Figure S3-5. Determination of molecular weight of native His_6-MaPT by gel filtration. The Superdex 75 column was calibrated with bovine serum albumin (66 kD), chicken egg ovalbumin (43 kD), bovine trypsinogen (25 kD), and ribonuclease A (13. 7 kD). The molecular weight of native MaPT was calculated as 61.6 kD.

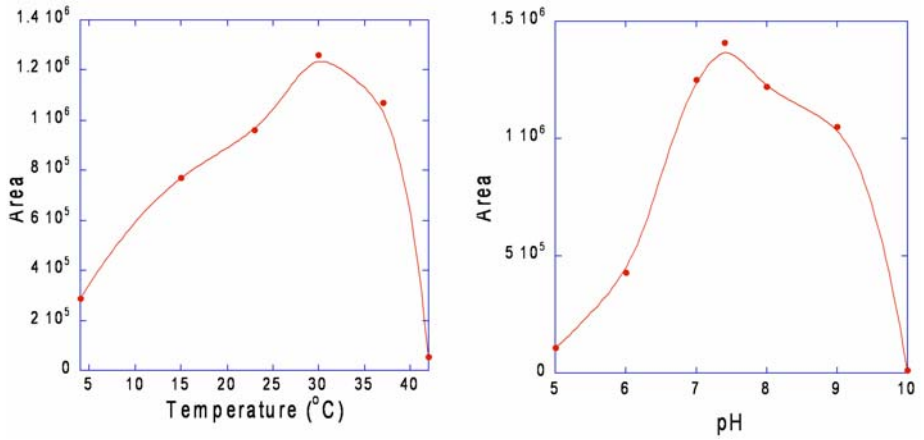


Figure S3-6. Determination of optimal reaction conditions for MaPT. The optimal temperature and pH value for reaction with L-Trp as substrate were 30 °C and 7.4, respectively.

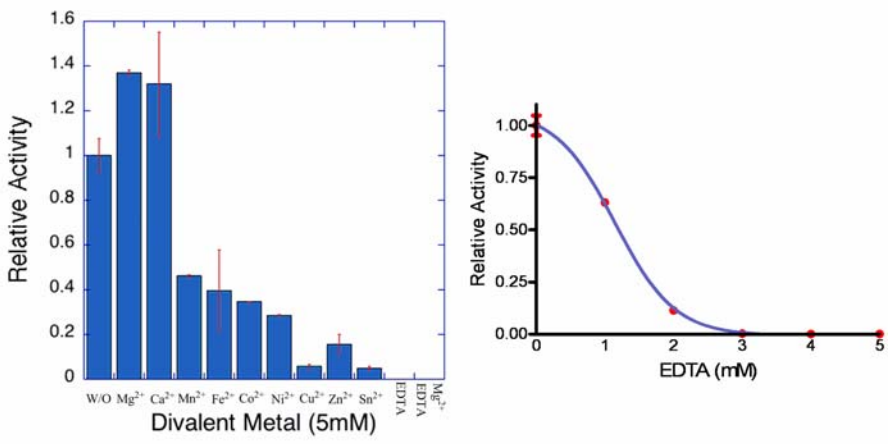


Figure S3-7. Investigation of MaPT metal ion dependence under optimal reaction conditions. Addition of 5 mM EDTA caused the total loss of enzyme activity which can not be restored by adding 10 mM Mg^{2+}. Experiments were conducted in duplicate. The inhibition by EDTA was concentration dependent.

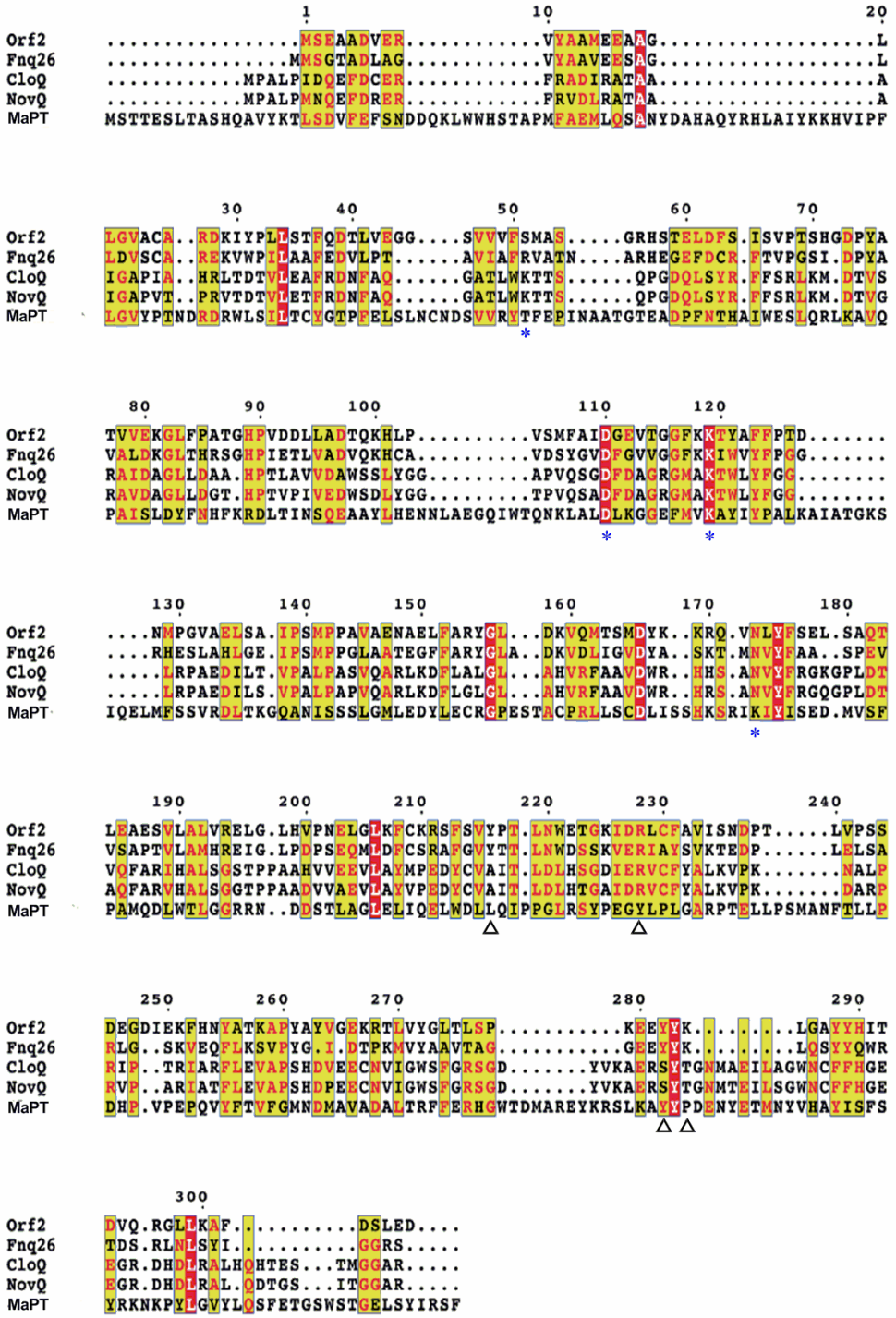


Figure S3-8. Sequence analysis of MaPT with CloQ (AAN65239), NovQ (AF170880), Fnq26 (CAL34101) and Orf2 (BAE00106). The alignment was performed with ESPript (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) using the known Orf2 secondary structure (1ZCW). The residues labeled with blue asterisks were mutated to investigate their potential roles in substrate binding and enzyme catalysis. Residues labeled with triangles are additional amino acids corresponding to those involved in binding of diphosphate, especially α-phosphate, in Orf2. The sequence labeling corresponds to Orf2.

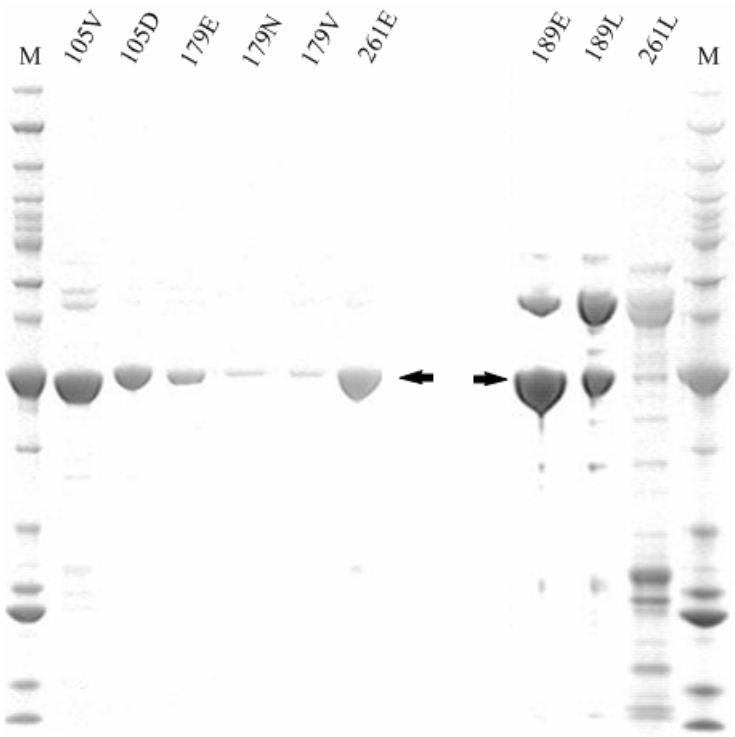


Figure S3-9. 4-12 % SDS-PAGE analysis of MaPT site-directed mutants. All mutants were purified with Ni-NTA agarose resin.

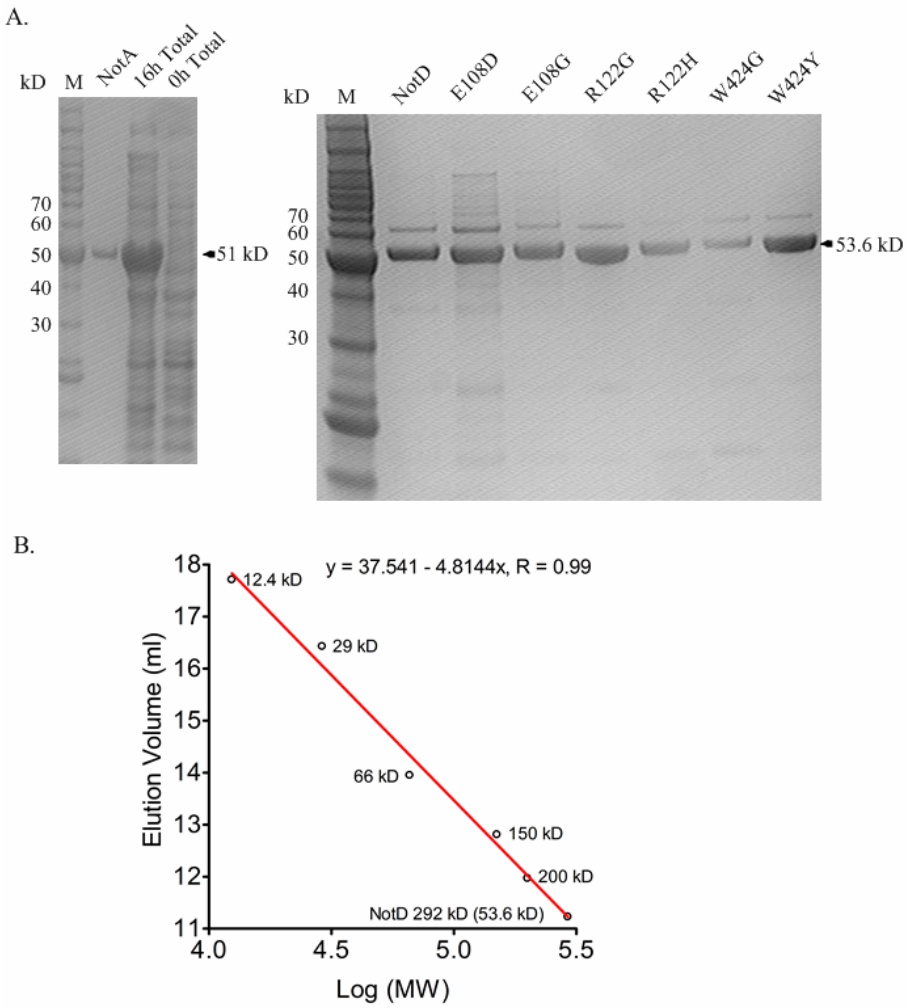


Figure S3-10. A. 4-12 % SDS-PAGE analysis of NotA and NotD with its mutants. B. Determining molecular weight of native His_6-NotD by gel filtration. The molecular weight of native NotD was calculated as 292 kD but its monomeric theoretical molecular weight is 53.6 kD.

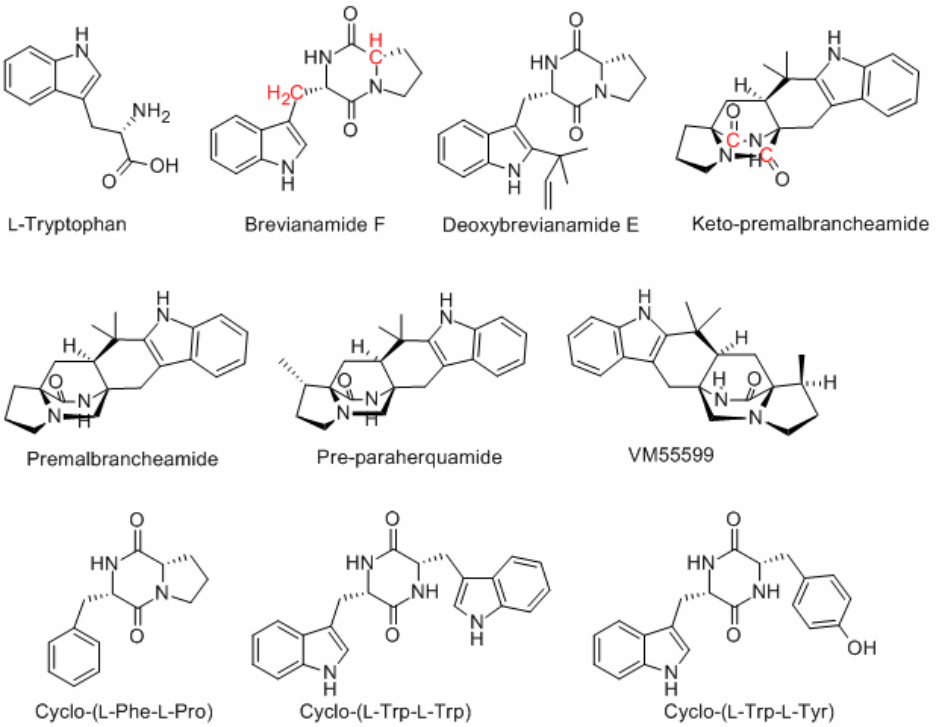


Figure S3-11. Selected substrates to test NotA and NotD activities. Brevianamide F and keto-premalbrancheamide were doubly labeled with ^{13}C atoms shown in red.

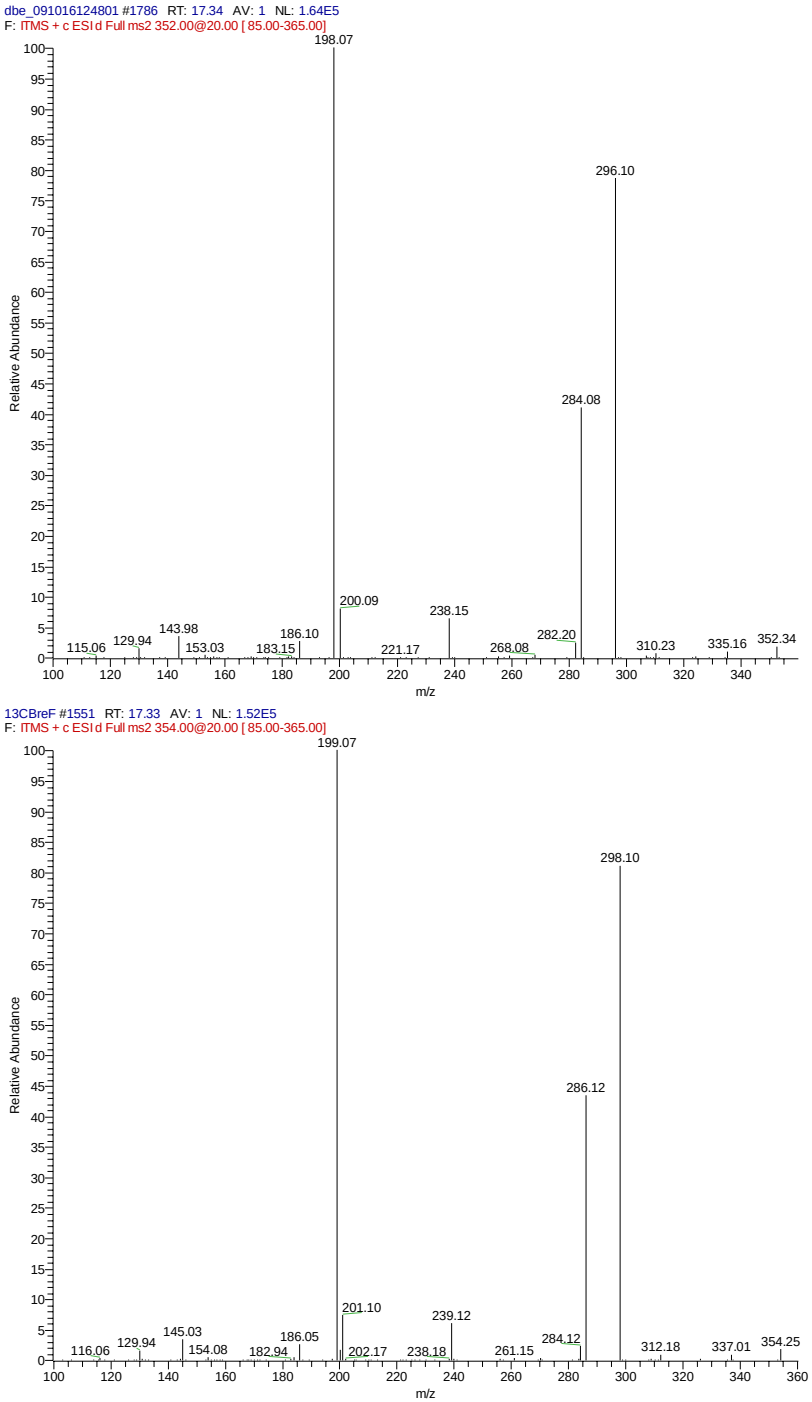


Figure S3-12. MS^2 analysis of NotD enzyme product (top) and authentic deoxybrevianamide E (bottom).

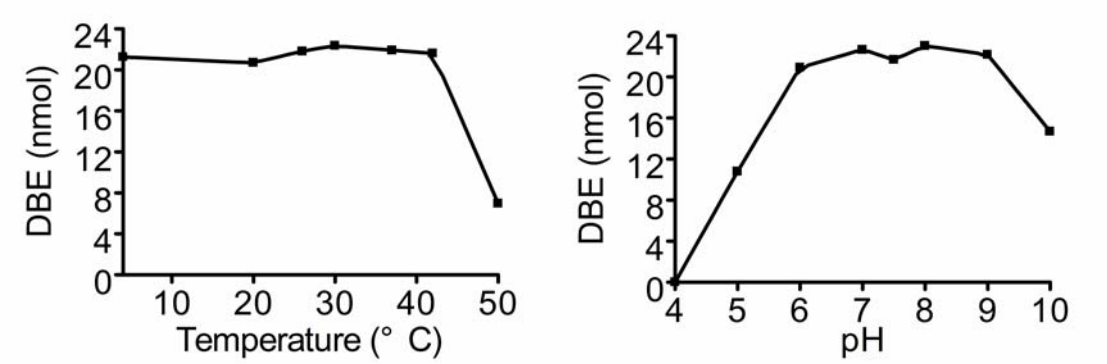


Figure S3-13. Determination of optimal conditions for NotD reaction.

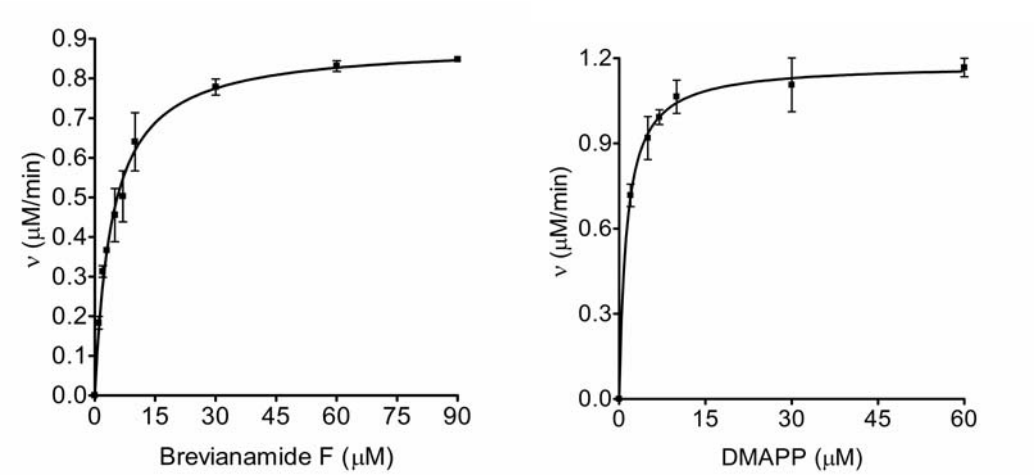


Figure S3-14. Kinetic analysis of NotD. The 100-μl reaction mixture contained 0.25 μg of NotD and 5 mM $MgCl_2$ in the buffer (50 mM Tris-Cl, pH 7.5, 10 % glycerol, and 3 mM BME).

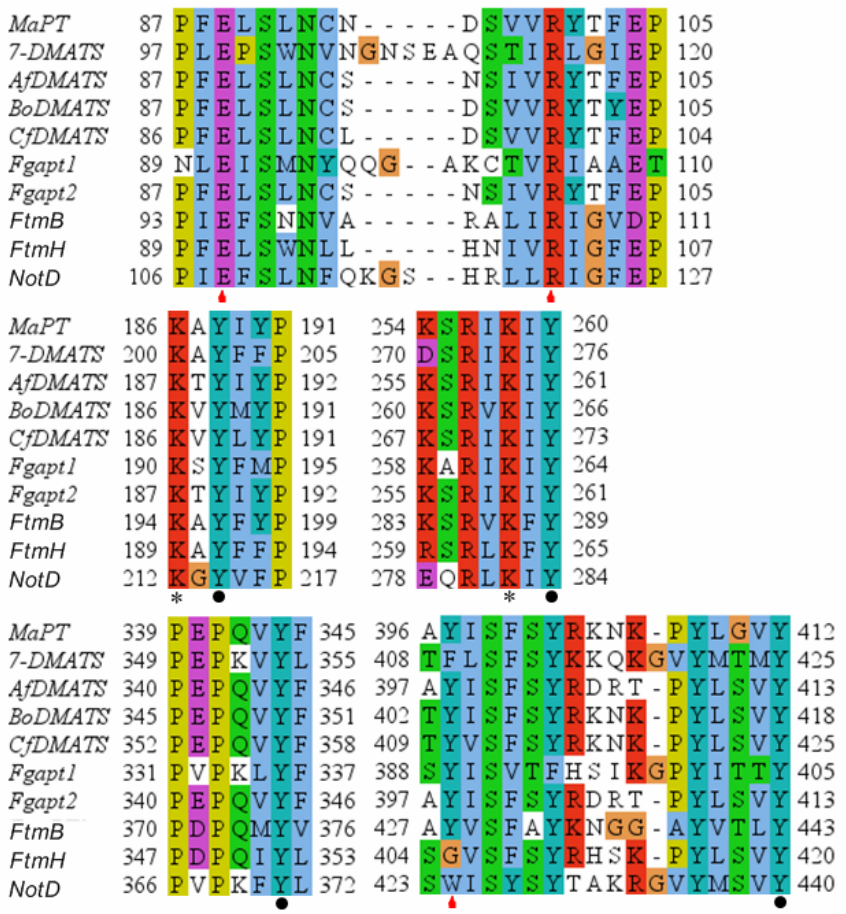


Figure S3-15. Alignment analysis of NotD. NotD shares key amino acid residues with other prenyltransferases for DMAPP and indole ring interaction. Two lysine residues labeled with stars were responsible for DMAPP interaction. Four tyrosine residues labeled with black cycles may shield the reactive carboncation from solvent. Three residues mutated in my studies were labeled with red arrows.

Table S3-1 ^{1}H NMR analysis of four MaPT products

Proton	DMAT	DMAT	DMA5HT	DMA5MT	DMAA
H-2	7.30 s	7.31 s	7.07 s	7.12 s	7.19 s
H-4	--	--	--	--	--
H-5	6.99 d (7.5)	6.99 d (7.0)	4. 92 s (5-OH)	2.31 s (5-CH_3)	6.82 dd (6.9)
H-6	7.20 dd (8.1, 7.5)	7.20 t (8.0, 7.5)	6.74 d (8.5)	6.96 d (8.5)	7.04 dd (15.5, 7.4)
H-7	7.41 d (8.1)	7.40 d (8.0)	7.09 d (3.0)	7.14 d (8.5)	7.23 d (8)
H-α	3.89 dd (10.4, 4.5)	4.11 dd (10.5, 4.5)	4.15 dd (11.5, 3.5)	4.16 dd (11, 3.5)	4.20 dd (9.5, 5)
H-β	3.25 dd (15.6, 10.4) 3.71 dd (15.6, 4.5)	3.27 dd (15.5, 10.5) 3.74 dd (15.5, 4.5)	3.08 dd (15.4, 11.4) 3.77 dd (15.1, 3.6)	3.12 dd (15, 11) 3.77 d (4)	3.25-3.27 m 3.71 d (4.6)
H_2-1’	3.79 m	3.77 d (6.5)	3.81-3.84 m 3.58 dd (21.1, 16.3)	3.79-3.82 m 3.67 d (18)	3.40-3.77 m 3.34-3.36 m
H-2’	5.37 m	5.37 m	5.12 s	5.10 s	5.30 s
H_3-4’	1.76 s	1.75 s	1.66 s	1.71 s	1.75 s
H_3-5’	1.79 s	1.78 s	1.81 s	1.83 s	1.78 s
NH-CH_3	--	--	--	--	2.65 s

*: DMAT was dissolved in D_2O and other three products were dissolved in CD_3OD.

Table S3-2 Primers for *NotA* and *NotD* intron removal, gene expression, and mutant preparation

Name	Sequence (5'→3')	Function
FW1(*Nde*I)	AAGCGCATATGGCCATTGAAGAAAAGTC	*NotA* expression
RV1(*Not*I)	ATAATGCGGCCGCTCAGAAAGGAGAGTAGTAC	*NotA* expression
InterF1	GAACGTCCAGTCGTACTATCCCAATCTGGATCTCAA	*NotA* Intron removal
InterR1	TTGAGATCCAGATTGGGATAGTACGACTGGACGTTC	*NotA* Intron removal
FW(*Nde*I)	ATCGGCATATGACGGCCCCAGAGCTC	*NotD* expression
RV(*Not*I)	ATTATGCGGCCGCTCAATCTTCTTCCCACAG	*NotD* expression
InterF	GATACCTTGCAGCAGCTGTACCCCGACCAGGACATTTC	*NotD* Intron removal
InterR	GAAATGTCCTGGTCGGGGTACAGCTGCTGCAAGGTATC	*NotD* Intron removal
E108GF	GTAGCGGACTGCCGATCGGATTCAGTCTCA ACTTCCAG	mutagenesis
E108GR	CTGGAAGTTGAGACTGAATCCGATCGGCAGTCCGCTAC	mutagenesis
E108DF	GTAGCGGACTGCCGATCGATTTCAGTCTCAACTTCCAG	mutagenesis
E108DR	CTGGAAGTTGAGACTGAAATCGATCGGCAGTCCGCTAC	mutagenesis
R122GF	GGCTCGCACCGACTGTTGGGCATCGGGTTTGAACCCG	mutagenesis
R122GR	CGGGTTCAAACCCGATGCCCAACAGTCGGTGCGAGCC	mutagenesis
R122HF	GGCTCGCACCGACTGTTGCACATCGGGTTTGAACCCG	mutagenesis
R122HR	CGGGTTCAAACCCGATGTGCAACAGTCGGTGCGAGCC	mutagenesis
F216YF	CTCGTCAAGGGCTATGTATACCCGTATCTGAAAGCAAAGGC	mutagenesis
F216YR	GCCTTTGCTTTCAGATACGGGTATACATAGCCCTTGACGAG	mutagenesis
F216VF	CTCGTCAAGGGCTATGTAGTCCCGTATCTGAAAGCAAAG GC	mutagenesis
F216VR	GCCTTTGCTTTCAGATACGGGACTACATAGCCCTTGACGAG	mutagenesis
Y371FF	GTTCCCCGTGCCCAAGTTCTTCCTCCCCGTGCATGGC	mutagenesis
Y371FR	GCCATGCACGGGGAGGAAGAACTTGGGCACGGGGAAC	mutagenesis
Y371VF	GTTCCCCGTGCCCAAGTTCGTCCTCCCCGTGCATGGC	mutagenesis
Y371VR	GCCATGCACGGGGAGGACGAACTTGGGCACGGGGAAC	mutagenesis
W424YF	GACAACCCGTTTGCAATCTTATATATCGTATTCCTATACTGC	mutagenesis
W424YR	GCAGTATAGGAATACGATATATAAGATTGCAAACGGGTTGTC	mutagenesis
W424GF	GACAACCCGTTTGCAATCTGGGATATCGTATTCCTATACTGC	mutagenesis
W424GR	GCAGTATAGGAATACGATATCCCAGATTGCAAACGGGTTGTC	mutagenesis

Table S3-3 Primers for MaPT gene identification, intron removal, gene expression, and mutant preparation

Name	Sequence (5'→3')	Function
PT-FW	GANGAYCASCGNCTNTGGTGG	gDNA screening
PT-RV	NCCNCCCAGNGTCCANASRTC	gDNA screening
3SP1	AGGGGCAAGCAAATATATCCTCTTC	Specific primer
3SP2	ATTATCTCGAATGCCGTGGTCCC	Specific primer
3SP3	TGATACGAGACTTATGGGAGGAAATCA	Specific primer
5SP1	AGCCAACGATCTCGATCGTTCGTA	Specific primer
5SP2	CATGTTTCTTGTAGATAGCAAGGTGTCTG	Specific primer
5SP3	GTCATAGTTGGCGGATTGGAGC	Specific primer
AD1	CAWCGICNGAIASGAA	Arbitrary primer
AD2	TCSTICGNACITWGGA	Arbitrary primer
PTFW (*Nde*I)	CCACATATGTCGACTACAGAATCCTTGACC	Gene expression
PTRV (*Bam*HI)	GCGCGGATCCTTAAAATGATCTAATATATGAC	Gene expression
InterFW	GCTTGAAAGCATACTACCCGGATGAAAATTATG	Intron removal
InterRV	CATAATTTTCATCCGGGTAGTATGCTTTCAAGC	Intron removal
T105VF	CTCAGTAGTCAGATATGTGTTCGAGCCCATAAATGC	mutagenesis
T105VR	GCATTTATGGGCTCGAACACATATCTGACTACTGAG	mutagenesis
T105DF	GCAACGACTCAGTAGTCAGATATGACTTCGAGCCCATAAATGCTG	mutagenesis
T105DR	CAGCATTTATGGGCTCGAAGTCATATCTGACTACTGAGTCGTTGC	mutagenesis
D179EF	CAGAACAAGTTAGCTTTGGAATTAAAGGGCGGGGAATTC	mutagenesis
D179ER	GAATTCCCCCCCCTTTAATTCCAAAGCTAACTTGTTCTG	mutagenesis
D179NF	CAGAACAAGTTAGCTTTGAATTTAAAGGGCGGGGAATTC	mutagenesis
D179NR	GAATTCCCCCCCCTTTAAATTCAAAGCTAACTTGTTCTG	mutagenesis
D179VF	CAGAACAAGTTAGCTTTGGTTTTAAAGGGCGGGGAATTC	mutagenesis
D179VR	GAATTCCCCCCCCTTTAAAACCAAAGCTAACTTGTTCTG	mutagenesis
K189EF	GGGGGAATTCATGGTTGAGGCTTATATCTACCCTGCTTTAAAAGC	mutagenesis
K189ER	GCTTTTAAAGCAGGGTAGATATAAGCCTCAACCATGAATTCCCCC	mutagenesis
K189LF	GGGGGAATTCATGGTTCTGGCTTATATCTACCCTGCTTTAAAAGC	mutagenesis
K189LR	GCTTTTAAAGCAGGGTAGATATAAGCCAGAACCATGAATTCCCCC	mutagenesis
K261EF	CTCCCATAAGTCTCGTATCGAGATCTACATTTCGGAAG	mutagenesis
K261ER	CTTCCGAAATGTAGATCTCGATACGAGACTTATGGGAG	mutagenesis
K261LF	CTCCCATAAGTCTCGTATCCTGATCTACATTTCGGAAG	mutagenesis
K261LR	CTTCCGAAATGTAGATCAGGATACGAGACTTATGGGAG	mutagenesis

3.6. Reference

1. Keller, N. P.; Turner, G.; Bennett, J. W., *Nat Rev Microbiol* **2005,** 3, (12), 937-47.
2. Gin, A. S.; Zhanel, G. G., *Ann Pharmacother* **1996,** 30, (6), 615-24.
3. Walsh, C., *Nat Rev Microbiol* **2003,** 1, (1), 65-70.
4. Spurgeon, D., *Bmj* **2007,** 335, (7627), 961.
5. Pelaez, F., Biological activities of fungal metabolites. In *Handbook of Industrial Mycology*, An, Z., Ed. Marcel Dekker: New York, 2005; pp 49-922.
6. Williams, R. M.; Cox, R. J., *Acc Chem Res* **2003,** 36, (2), 127-39.
7. Artman, G. D., 3rd; Grubbs, A. W.; Williams, R. M., *J Am Chem Soc* **2007,** 129, (19), 6336-42.
8. Birch, A. J.; Wright, J. J., *Tetrahedron* **1970,** 26, (10), 2329-44.
9. Martinez-Luis, S.; Gonzalez, M. C.; Ulloa, M.; Mata, R., *Phytochemistry* **2005,** 66, (9), 1012-6.
10. Qian-Cutrone, J.; Huang, S.; Shu, Y. Z.; Vyas, D.; Fairchild, C.; Menendez, A.; Krampitz, K.; Dalterio, R.; Klohr, S. E.; Gao, Q., *J Am Chem Soc* **2002,** 124, (49), 14556-7.
11. Shoop, W. L.; Haines, H. W.; Eary, C. H.; Michael, B. F., *Am J Vet Res* **1992,** 53, (11), 2032-4.
12. Hayashi, H.; Nishimoto, Y.; Akiyama, K.; Nozaki, H., *Biosci Biotechnol Biochem* **2000,** 64, (1), 111-5.
13. Blunt, J. W.; Copp, B. R.; Hu, W. P.; Munro, M. H.; Northcote, P. T.; Prinsep, M. R., *Nat Prod Rep* **2008,** 25, (1), 35-94.
14. Kato, H.; Yoshida, T.; Tokue, T.; Nojiri, Y.; Hirota, H.; Ohta, T.; Williams, R. M.; Tsukamoto, S., *Angew Chem Int Ed Engl* **2007,** 46, (13), 2254-6.
15. John Baldas, A. J. B. a. R. A. R., *J. Chem. Soc., Perkin Trans. 1* **1974**, 50 - 52.
16. Domingo, L. R.; Zaragoza, R. J.; Williams, R. M., *J Org Chem* **2003,** 68, (7), 2895-902.
17. Stocking, E. M.; Martinez, R. A.; Silks, L. A.; Sanz-Cervera, J. F.; Williams, R. M., *J Am Chem Soc* **2001,** 123, (14), 3391-2.
18. Stocking, E. M.; Sanz-Cervera, J. F.; Williams, R. M., *Angew Chem Int Ed Engl* **2001,** 40, (7), 1296-1298.
19. Maiya, S.; Grundmann, A.; Li, S. M.; Turner, G., *Chembiochem* **2006,** 7, (7), 1062-9.
20. Grundmann, A.; Li, S. M., *Microbiology* **2005,** 151, (Pt 7), 2199-207.
21. Banks, R. M.; Blanchflower, S. E.; Everett, J. R.; Manger, B. R.; Reading, C., *J Antibiot (Tokyo)* **1997,** 50, (10), 840-6.
22. Whyte, A. C.; Gloer, J. B.; Wicklow, D. T.; Dowdw, P. F., *J Nat Prod* **1996,** 59, (11), 1093-5.
23. Blanchflower, S. E.; Banks, R. M.; Everett, J. R.; Reading, C., *J Antibiot (Tokyo)* **1993,** 46, (9), 1355-63.
24. Nielsen, K. F.; Sumarah, M. W.; Frisvad, J. C.; Miller, J. D., *J Agric Food Chem* **2006,** 54, (10), 3756-63.
25. Figueroa, M.; Gonzalez Mdel, C.; Mata, R., *Nat Prod Res* **2008,** 22, (8), 709-14.
26. Liu, Y. G.; Whittier, R. F., *Genomics* **1995,** 25, (3), 674-81.
27. Steffan, N.; Unsold, I. A.; Li, S. M., *Chembiochem* **2007,** 8, (11), 1298-307.

28. Pojer, F.; Wemakor, E.; Kammerer, B.; Chen, H.; Walsh, C. T.; Li, S. M.; Heide, L., *Proc Natl Acad Sci U S A* **2003,** 100, (5), 2316-21.
29. Shah, M. M.; Grover, T. A.; Barr, D. P.; Aust, S. D., *J Biol Chem* **1992,** 267, (30), 21564-9.
30. Bhattacharyya, D. K.; Adak, S.; Bandyopadhyay, U.; Banerjee, R. K., *Biochem J* **1994,** 298 (Pt 2), 281-8.
31. Carvajal, N.; Orellana, M. S.; Borquez, J.; Uribe, E.; Lopez, V.; Salas, M., *J Inorg Biochem* **2004,** 98, (8), 1465-9.
32. Cress, W. A.; Chayet, L. T.; Rilling, H. C., *J Biol Chem* **1981,** 256, (21), 10917-23.
33. Gebler, J. C.; Poulter, C. D., *Arch Biochem Biophys* **1992,** 296, (1), 308-13.
34. Tsai, H. F.; Wang, H.; Gebler, J. C.; Poulter, C. D.; Schardl, C. L., *Biochem Biophys Res Commun* **1995,** 216, (1), 119-25.
35. Unsold, I. A.; Li, S. M., *Microbiology* **2005,** 151, (Pt 5), 1499-505.
36. Unsold, I. A.; Li, S. M., *Chembiochem* **2006,** 7, (1), 158-64.
37. Haagen, Y.; Unsold, I.; Westrich, L.; Gust, B.; Richard, S. B.; Noel, J. P.; Heide, L., *FEBS Lett* **2007,** 581, (16), 2889-93.
38. Kremer, A.; Westrich, L.; Li, S. M., *Microbiology* **2007,** 153, (Pt 10), 3409-16.
39. Yin, W. B.; Ruan, H. L.; Westrich, L.; Grundmann, A.; Li, S. M., *Chembiochem* **2007,** 8, (10), 1154-61.
40. Galagan, J. E. et al. *Nature* **2005,** 438, (7071), 1105-15.
41. Grundmann, A.; Kuznetsova, T.; Afiyatullov, S.; Li, S. M., *Chembiochem* **2008,** 9, (13), 2059-63.
42. Zou, H.; Zheng, X.; Li, S. M., *J Nat Prod* **2009,** 72, (1), 44-52.
43. Metzger, U.; Schall, C.; Zocher, G.; Unsold, I.; Stec, E.; Li, S. M.; Heide, L.; Stehle, T., *Proc Natl Acad Sci U S A* **2009,** 106, (34), 14309-14.
44. Stec, E.; Steffan, N.; Kremer, A.; Zou, H.; Zheng, X.; Li, S. M., *Chembiochem* **2008,** 9, (13), 2055-8.
45. Tsukamoto, S.; Kato, H.; Greshock, T. J.; Hirota, H.; Ohta, T.; Williams, R. M., *J Am Chem Soc* **2009,** 131, (11), 3834-5.
46. Sambrook, J.; Russel, D. W., *Molecular Cloning - A Laboratory Manual*. 3rd ed.; Cold Spring Harbor Laboratory Press: New York, 2001.

Notes:

Yousong Ding and David H. Sherman designed all experiments and Yousong Ding performed the experiments.

Substrates in feeding experiments as well as the substrates to test NotA and Not D activities were obtained from Dr. Robert M. Williams group.

Chapter 4

Biochemical Characterization of the Biosynthesis of Anticancer Trichothecene Macrolides in Marine *Myrothecium verrucaria*

4.1. Summary

This chapter focuses on the first studies of the gene cluster responsible for the biosynthesis of trichothecene macrolides in one marine fungal strain, *Myrothecium verrucaria*. Fungal genomic libraries were constructed and screened with specific probes. One gene cluster was identified in a 102-kb DNA fragment. In total, 36 genes were identified in the gene cluster, and functions of these genes were predicted. VerH containing three C_2H_2 zinc finger domains was predicted as a transcriptional activator. Bioinformatic analysis found multiple VerH binding sites in promoter regions of the verrucarin biosynthetic genes. VerN was predicted as a sesquiterpene synthase and the homolog of Tri5 in acyclic trichothecene biosynthesis. Removal of one intron resulted in the generation of its cDNA for heterologous overexpression in *E. coli*. The purified enzyme was shown to convert farnesyl pyrophosphate (FPP) into trichodiene, validating its role in natural product biosynthesis. The cDNA of one P450, *VerE*, was produced after removing four introns, and was overexpressed in a yeast strain. The isolated yeast microsome fraction was utilized to convert trichodiene into two putative metabolites as determined in GC-MS analysis. Finally, a putative biosynthetic pathway to produce verrucarin and roridin was proposed, based on predicted functions of genes in the cluster and biochemical characterization of VerN and VerE.

4.2. Introduction

Fungi are ubiquitous in *Nature* and are rich sources of numerous secondary

metabolites. Many fungal natural products, such as antibacterial penicillins, cephalosporins, immunosuppressive cyclosporine A, and cholesterol-lowering HMG-CoA reductase inhibitor "statins", have been clinically used to fight against many diseases [1]. Conversely, many mycotoxins, which commonly contaminate food supply, upon consumption, threaten the lives of animals and human beings and cause enormous economical loss. Although mycotoxins are less well known than mushroom poisons, they indeed cause a higher incidence of disease, because they contaminate about 25 % of the worldwide agricultural crop. As early as 430 B.C., ergot alkaloids commonly produced by *Claviceps* sp., *Aspergillus* sp., and *Penicillium* sp., were correlated with a human disease called ergotism, or St. Anthony's Fire (**Figure 4-1**) [1]. The most well-known mycotoxins are the aflatoxins named after the producer, *Aspergillus flavus*. The four major aflatoxins — aflatoxin B_1, B_2, G_1 and G_2 — were identified on the basis of their blue or green fluorescence under ultraviolet light and aflatoxin B_1 is one of the most toxic and carcinogenic compounds ever discovered (**Figure 4-1**) [2]. *A. flavus* is widely distributed and is highly infective to most major crop plants. These factors make it difficult to prevent aflatoxin contamination of foods and feeds. For example, a severe outbreak was reported in Kenya and had at least a 39% incidence of death (317 cases with 125 deaths) resulting from acute hepatotoxicity in 2004 [3]. Another group of mycotoxins, fumonisins, is mainly produced by the maize pathogens, *Fusarium verticillioides* (formerly *F. moniliforme*) and *F. proliferatum* (**Figure 4-1**). The predominant fumonisin, fumonisin B_1, is the agent responsible for leukoencephal-omalacia in horses and pulmonary edema in swine and may be associated with esophageal cancer in human [4]. Due to the health concerns related to fumonisin consumption, the Food and Drug Administration (FDA) has established recommended

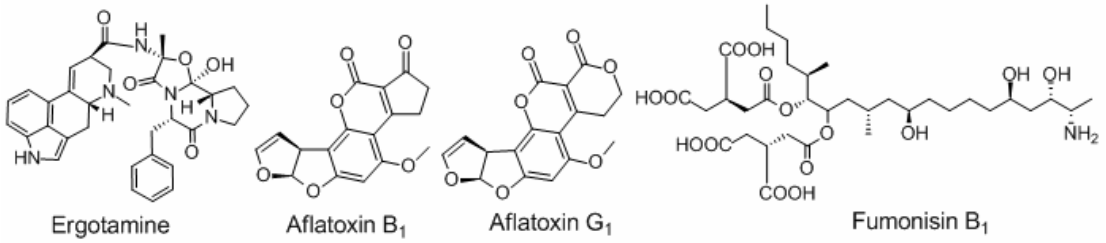


Figure 4-1. Chemical structures of selected fungal mycotoxins.

guidelines for the levels in food and feedstuff. The fumonisin content was recommended to be lower than 2 mg / kg in degermed dry milled corn products such as flaking grits, corn grits, corn meal, and corn flour, and 5 mg / kg in maize fed to horses [5].

Trichothecenes are one group of mycotoxins produced by a number of unrelated fungal genera including *Fusarium*, *Mycothecium*, *Trichoderma*, *Trichothecium*, *Stachybotrys*, *Verticimonosporium*, and *Cephalosporium* sp. (**Figure 4-2**) [6]. Currently, more than 200 trichothecenes have been isolated, and a few cause a range of acute and chronic symptoms in animals and human beings after consumption of contaminated crops or feeds [7]. For example, thousands of people in Russia died after consuming grain contaminated with *Fusarium* species during World War II. One trichothecene, T-2 toxin, is believed to correlate with alimentary toxic aleukia outbreaks since the 19th century and with outbreaks of hemorrhagic syndrome in farm animals in North America and Europe [8]. The ease for trichothecene-producing fungal strains growing on crop plants suggests the necessity to detect its level in food supplies. Accordingly, about $ 637 million are annually lost due to deoxynivalenol (DON) contaminating human crop foods.

Structurally, all trichothecenes are sesquiterpenes with the common features of one tricyclic nucleus, one double bond between C-9 and C-10, and one epoxide between C-12 and C-13. According to the different side chains attached to the trichothecene skeleton, these mycotoxins are divided into four types (A-D) (**Figure 4-2**) [6]. T-2 toxin and diacetoxyscirpenol (DAS, also named as anguidine) are two examples in type A, while DON belongs to type B because of the presence of keto group at its C-8. Type C includes several non-*Fusarium* trichothecenes that contain an additional epoxide between their C-7 and C-8. In contrast to all in the above three types, trichothecenes in type D contain a macrocyclic ring between C-4 and C-15. These macrocyclic mycotoxins, such as satratoxins, verrucarins, and roridins, are more toxic than their acyclic counterparts and have attracted increasing interests to control their production.

The primary mechanism of trichothecene toxicity is to inhibit protein synthesis [9]. They interact with 60S subunits of the mammalian ribosome in a non-competitive manner,

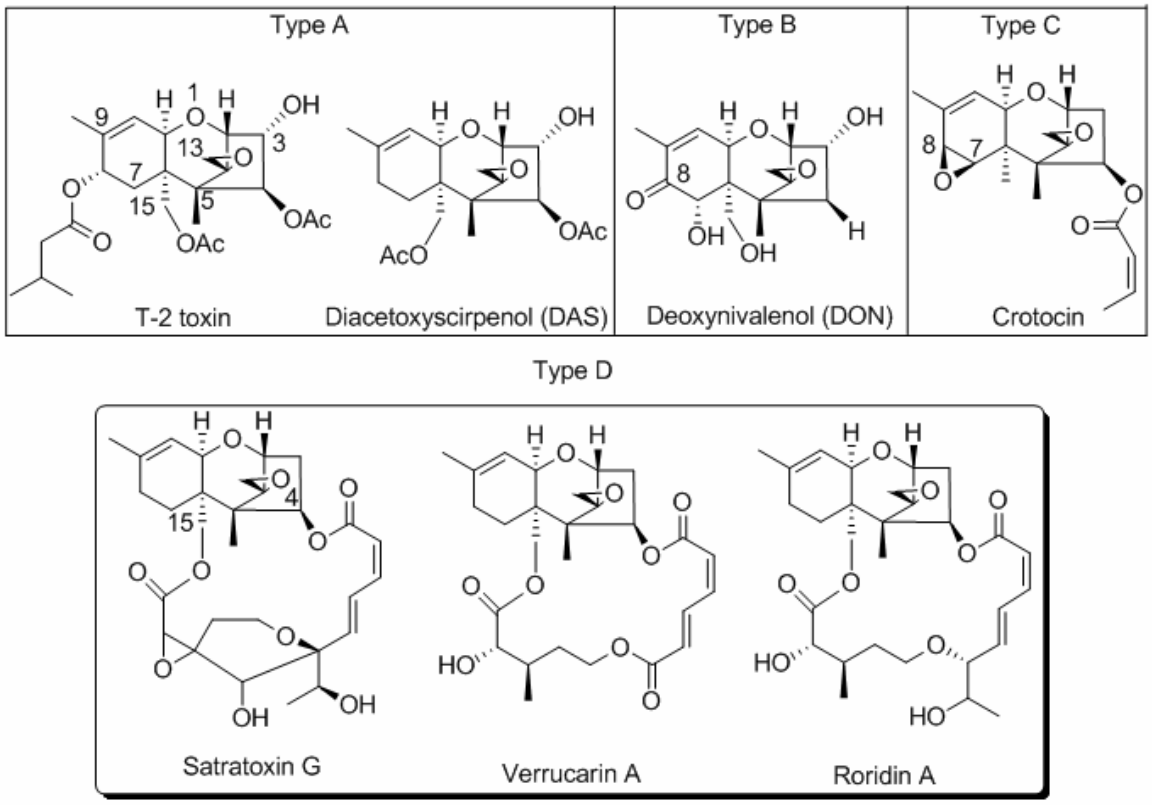


Figure 4-2. Chemical structures of selected trichothecenes. All of trichothecene mycotoxins are filled into four types (A-D) based on their side chains.

and the different trichothecenes specifically inhibit initiation, elongation, or termination steps in protein synthesis [6, 10]. Moreover, structural differences of their side chains cause different degrees of toxicity toward eukaryotic cells [6]. The most potent protein synthesis inhibitors are T-2 toxin, verrucarin A, and roridin A from a group of nineteen 12,13-epoxytrichothecene mycotoxins that were tested for their relative capabilities to inhibit protein synthesis in Vero cells and rat spleen lymphocytes [11]. The other cellular effects of trichothecenes on eukaryotic cells include inhibition of DNA and RNA synthesis, inhibition of mitosis, and cell membrane effects [12]. Recently, some trichothecenes were found to induce apoptosis, a programmed cell death response in various cell lines [13]. Their apoptosis induction ability might require both translational arrest and mitogen-activated protein kinase activity. However, how trichothecenes activate the eukaryotic signal transduction cascades and downstream gene products remains obscure. The overall effects of trichothecenes on eukaryotic cells led to develop them as the anticancer agents [14-16]. Anguidine (DAS) showed good potency and excellent selectivity toward leukemia

cells [16]. However, in the late 1970s, five distinct series of Phase II studies involving 484 patients produced unfavorable results with serious side effects and a lack of substantive anticancer activity [17].

The intriguing structure, potent toxicity, and wide distribution of trichothecences has fostered considerable interests in understanding the biosynthetic pathway, genes, enzymes, regulation, and evolution, for this group of important mycotoxins. The biosynthesis of acyclic trichothecenes (type A and B) produced by *Fusarium* sp. have been extensively studies by one research group of USDA in past 25 years (**Figure 4-3**) [8]. Twelve genes were found in the Tri5 gene cluster, and at least three essential genes were found to be outside this chromosomal region, but all were required for T-2 toxin biosynthesis in *F. sporotrichioides* [18]. Briefly, one sesquiterpene synthase (Tri5) produces trichodiene, the key biosynthetic intermediate, from farnesyl pyrophosphate (FPP) [19]. This intermediate is then converted into isotrichodermol by a single P450, Tri4 [20, 21].

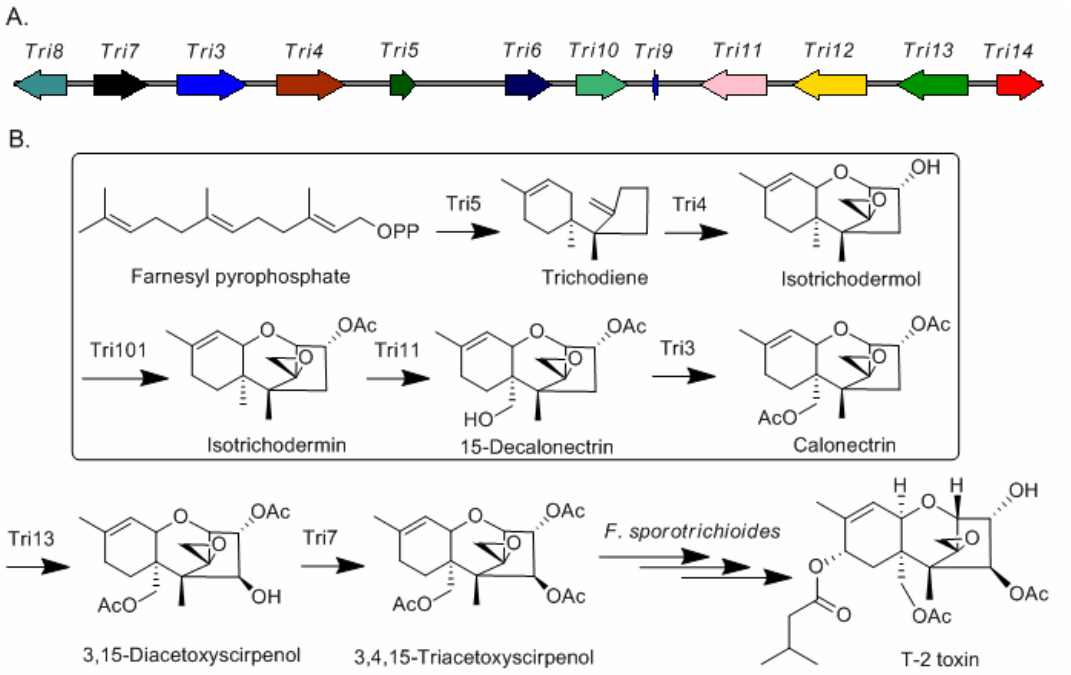


Figure 4-3. A. One gene cluster for T-2 toxin biosynthesis in *F. sporotrichioides*. This gene cluster contains 12 essential genes. **B. One biosynthetic pathway for type A and B trichothecenes produced by *Fusarium* sp**. The biosynthetic steps in the box are shared by both types of trichothecenes, while T-2 toxin production in *F. sporotrichioides* requires three pathway specific enzymes.

This remarkable enzyme is proposed to catalyze at least two hydroxylations and one epoxidation on trichodiene. Trichothecene 3-*O*-acetyltransferase (Tri101) catalyzes the next step in trichothecene biosynthesis [22]. Its product, isotrichodermin, is then hydroxylated by another P450 (Tri11) to generate 15-decalonectrin [23]. Calonectrin is produced in the next step from acetylation of the OH group at C-15 by the second *O*-acetyltransferase (Tri3) [24]. All of above enzymes are shared in the biosyntheses of all type A and B trichothecenes produced by *Fusarium* sp. In *F. sporotrichioides* and *F. graminearum*, DAS is produced after one hydroxylation at C-4 by Tri13, and one subsequent C-4-*O*-acetylation of DAS by Tri7 produces 3,4,15-triacetoxyscirpenol (TAS) [25]. In *F. sporotrichioides*, three unique enzymes are required to convert TAS into T-2 toxin, while five pathway-specific enzymes are expected in *F. graminearum* to produce nivalenol, another type B trichothecene [8]. The biosynthetic pathway of DON in *F. graminearum* branches from the T-2 toxin pathway after the production of calonectrin. The studies of trichothecene serve as the primary model for fungal secondary metabolite. The findings not only help to develop practical control strategies but also offer opportunities for generating new sesquiterpenes through metabolic engineering.

The biosynthesis of type D trichothecenes has not been studied extensively, in contrast to the biosynthesis of the acyclic trichothecenes in *Fusarium*. Trichothecene macrolides are mainly produced by plant fungal pathogens such as *Myrothecium* and *Stachybotrys*. *Stachybotrys* species are commonly found in indoor environments, and recently, there has been increased public attention to mycotoxins, including the satratoxins produced by this fungal species, because of their correlations with many acute and chronic diseases (**Figure 4-2**) [26-28]. Moreover, *M. verrucaria* and *M. roridum* are commonly found on materials such as paper, textiles, and living plants. Fungal infections and verrucarins and roridins produced by these fungal strains may cause adverse effects on human living environments and crops. An understanding of their biosynthetic systems is critical for developing the control of their production. However, only three genes (*Tri4*, *Tri5*, and *Tri6*) were previously identified from *M. roridum* and shown to be involved in the biosynthesis of acyclic trichothecene moiety of roridins [29]. In 2003, the Crews group

at the University of California-Santa Cruz isolated the marine fungus, *M. verrucaria*, in Hawaii. A group of type D trichothecenes, including verrucarins and roridins, were purified and characterized from this strain [30]. This chapter describes my studies to isolate the complete gene cluster for verrucarin and roridin biosynthesis from the marine fungal species and the characterization of two enzymes involved in their biosynthetic pathwaies.

4.3. Results and Discussion

4.3.1. Isolation of Trichothecene Macrolide Gene Cluster from the Marine Fungus *Myrothecium verrucaria*

Three genes similar to *Tri4*, *Tri5*, and *Tri6*, and one fungal iterative PKS gene were expected to be present in the verrucarin gene cluster. Thus, the *Tri4*, *Tri5*, and PKS *KS* domain were used as probes to screen *M. verrucaria* genome library. The fungal genomic DNA was isolated and nebulized into 30-40 kb fragments for the construction of the library. The protocols in the Epicentre CopyControl™ Fosmid Library Production Kit were followed to guide the processes. Initially, about 10,000 colonies were screened with one pair of degenerate primers designed on the basis of PKS *KS* domain, and two pairs of primers on the basis of conserved regions of *Tri4* and *Tri5* from *M. roridium*. Initially, 10 positive hits were found with probes of *KS* and *Tri5* (**Table S1**). Fosmid N1 was chosen for shotgun sequencing. In the 40.5-kb chromosomal fragment, 11 genes were identified. Seven exhibited high similarities to the characteristic trichothecene biosynthetic genes, *Tri5*, *Tri6*, *Tri7*, *Tri9*, *Tri10*, *Tri11*, and *Tri12*. Moreover, one fungal iterative PKS gene was identified in this region, strongly indicating the isolation of the correct gene cluster for verrucarin biosynthesis. Extensive efforts using the probes for the identified CoA ligase gene and *Tri5* allowed us to identify fosmid C10, adding a new 14.6-kb sequence and revealing four more genes. However, the absence of the essential *Tri4* caused us to prepare and screen the second fungal genomic fosmid library. About 1,500 colonies were further screened with the primers from both ends of the above gene cluster, and one positive hit covering each end was chosen for the subsequent shotgun sequencing (**Figure 4-4**, **Table S1**). In total, one 101.6-kb chromosomal region was sequenced and its overall GC content is 49 %.

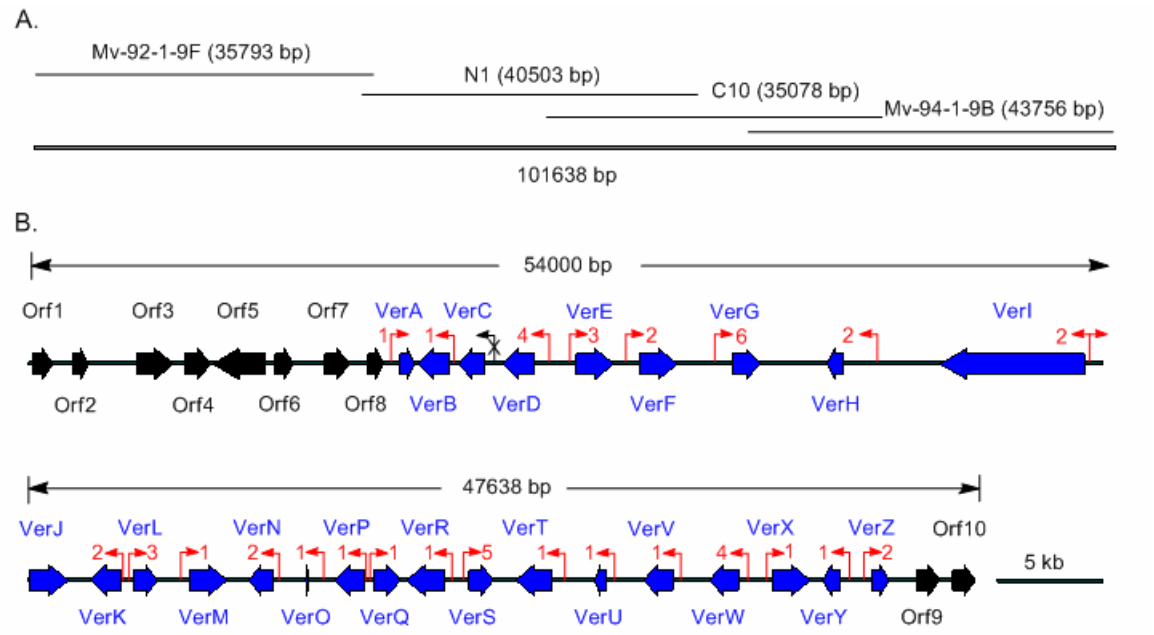


Figure 4-4 A. One sequenced chromosomal region covered by four fosmids. B. Thirty-six open reading frames were predicted in the sequenced region. The boundaries of trichothecene macrolide gene cluser were defined by the presence of VerH binding motifs in the gene promoter region.

In total, 36 open reading frames (ORF) were identified in the sequenced region. Preliminary functional assignments of individual ORFs were made by comparison of the deduced gene products with proteins of known functions in the NCBI database (**Table 4-1**). At the left side of the sequenced fragment, Orf1, Orf2, Orf3 were predicted to be one putative fungal phospholipase A2 isoform 1, one putative fungal endoglucanase V, and one galactose oxidase precursor, respectively, indicating they may not be involved in verrucarin biosynthesis. Orf4 encoded a putative salicylate 1-monooxygenase while Orf5 may be a C6 transcription factor. Subsequently, Orf6 and Orf8 were predicted as a putative short-chain dehydrogenase /reductase and a putative NAD-dependent epimerase /dehydratase, respectively, but the function of Orf7 is unknown. The involvement of these gene products in verrucarin biosynthesis was predicted by the analysis of their promoter regions. VerH shows 66 % identity to Tri6, a C_2H_2 trichothecene-specific transcriptional activator [31]. Tri6 binds to the conserved DNA motif of YNAGGCC in the promoter regions of all trichothecene biosynthetic genes to directly regulate their expression [31]. It is expected that all verrucarin biosynthetic genes have at least one copy of the YNAGGCC motif in their promoter regions. Bioinformatic analysis indicated that

Table 4-1. Features of gene products in the sequenced chromosomal region

Code	Position	Size	Features / putative functions
Orf1	209-1225	338 (28/44)	Partial *M. musculus* group XII-1 phospholipase A2 isoform 1 (P_075685.2)
Orf2	2213-2295 2347-2595 2661-3021	230	*N. crassa* endoglucanase V (XP_957107.1)
Orf3	5410-7287	625 (61/72)	*P. tritici-repentis* galactose oxidase precursor (EDU43973.1)
Orf4	7874-9157	427 (28/46)	*M. sp.* salicylate 1-monooxygenase (ABR69922.1)
Orf5	9255-10713 11060-11734 11799-11872	735 (33/50)	Putative *N. fischeri* C6 transcription factor (EAW20173.1)
Orf6	12407-13306	299 (41/59)	*A. fumigatus* short-chain dehydrogenase /reductase (XP_748579.1)
Orf7	14881-15421 15570-16138	369 (36/54)	*G. zeae* hypothetical protein (XP_390502.1); unknown protein
Orf8	17054-17203 17275-17823	232 (46/66)	*A. fumigatus* conserved hypothetical protein (XP_752469.1); putative NAD-dependent epimerase /dehydratase
VerA	18667-19395	242 (66/80)	*G. zeae* hypothetical protein (XP_384723.1); glutathione *S*-transferase
VerB	19556-20901 20966-21116	498 (66/79)	*A. fumigatus* MFS multidrug transporter (XP_748638.1); self-resistant machinery
VerC	21614-22927	437 (49/69)	*A. niger* hypothetical protein (XP_001395378.1); C6 finger domain protein
VerD	23814-25196 25250-25399	510 (36/53)	*S. sclerotiorum* hypothetical protein (XP_001589913.1); P450
VerE	27525-27768 27825-28684 28747-28857 28950-29116 29187-29391	528 (81/89)	*M. roridum* cytochrome P450 (AAC49958.1); P450 similar to Tri4
VerF	30693-31490 31584-32105 32287-32411 32469-32601	525 (44/64)	*G. zeae* trichothecene 15-O-acetyltransferase (BAC22114.1); acyltransferase similar to Tri3
VerG	35361-35748 35881-36749	418 (52/67)	*G. zeae* regulatory protein (AAM48887.1); regulatory protein similar to Tri10
VerH	40098-40906	264 (66/73)	*M. roridum* trichothecene biosynthesis transcription factor (AAC49959.1); transcription factor similar to Tri6
VerI	45774-52679 52739-53020	2396 (32/51)	*A. clavatus* polyketide synthase (XP_001273842.1); fungal iterative type I polyketide synthase
VerJ	54048-54998 55077-55555 55576-55633 55721-55842 55906-56038	580 (30/50)	*F. sporotrichioides* 15-O-acetyltransferase (AAK33072.1); acyltransferase similar to Tri3
VerK	57123-57464 57525-58670	495 (52/70)	*A. clavatus* benzoate 4-monooxygenase (XP_001273721.1); putative P450
VerL	59319-59507 59572-60484	367 (82/92)	*S. chartarum* TRI14 (AAG47844.1); unknown protein similar to Tri14
VerM	62090-62702 62781-63491 63550-64013	595 (57/75)	*F. pseudograminearum* trichothecene efflux pump (AAM48858.1); self-resistant machinery similar to Tri12
VerN	65094-65782 65842-66310	385 (83/89)	*M. roridum* trichodiene synthase (AAC49957.1); sesquiterpene synthase similar to Tri5
VerO	67949-68077	42 (47/72)	*F. sporotrichioides* TRI9; unknown protien similar to Tri9
VerP	69419-70888	489 (34/56)	N. uncinatum LolU-1 (AAV68707.1); putative DNA binding protein
VerQ	71352-72723	455 (66/82)	*G. zeae* hypothetical protein (XP_384510.1); putative acyl transferase
VerR	73008-74110 74218-74897	593 (59/73)	*A. fumigatus* long-chain-fatty-acid-CoA ligase (XP_753087.1); acyl-CoA ligase

VerS	76175-76530 76581-77391	388 (33/51)	*M. grisea* hypothetical protein (XP_367888.1); protein similar to AtmA
VerT	78523-78941 79016-79610 79672-79845 79907-80023 80076-80306	511 (39/61)	*G. fujikuroi* cytochrome P450 monooxygenase (CAA75565.1); P450
VerU	82487-82725 82791-83034	160 (35/59)	*G. zeae* hypothetical protein (XP_387387.1); unknown protein
VerV	84920-86407	495 (44/62)	*A. clavatus* flavin-binding monooxygenase (XP_001275023); flavin-binding monooxygenase
VerW	88228-88780 88837-88961 89012-89157 89213-89308 89452-89522 89594-89709	368 (67/82)	*N. crassa* norsolorinic acid reductase (XP_961973.2); ketone reductase
VerX	91436-91685 91747-91836 91900-92013 92073-92545 92603-92769 92836-92963 93015-93318	511 (33/54)	*P. tritici-repentis* ent-kaurene oxidase (XP_001942213.1); putative P450
VerY	93950-94819	289 (38/62)	*A. fumigatus* short-chain dehydrogenase/reductase (XP_747739.2); short-chain dehydrogenase/reductase
VerZ	96428-97270	280 (38/56)	*S. cellulosum* α/β fold hydrolase /acyltransferase (YP_001612520.1); hydrolase/acyltransferase
Orf9	98705-99838	369 (63/75)	*G. zeae* hypothetical protein (XP_390717.1); unknown protein
Orf10	100491-101638	316 (45/60)	Partial *A. fumigatus* ankyrin repeat protein (EDP48589.1)

such a motif was absent in the promoter regions of *Orf1-8*, indicating that they might not be the verrucarin biosynthetic genes. Similarly, *Orf9-10* at the right end of the sequenced region lacked this conserved motif in their promoter regions and may define the right boundary of verrucarin gene cluster.

Twenty-six ORFs (*VerA-Z*) were tentatively identified in the middle of the sequenced regions (**Figure 4-4**). All of these ORFs except *VerC* contained at least one copy of the VerH binding motif in their promoter regions (**Table S4-2**). Compared to the T-2 toxin biosynthetic gene cluster, ten genes in verrucarin gene cluster showed homologous (**Figure S4-1**, **Table 4-1**). Some of these gene products (VerE, VerF, VerK, VerJ, and VerN) were possibly involved in the biosynthesis and diversification of verrucarol, while others, including VerG, VerH, VerL, VerM, and VerO, may be transcriptional regulators, resistance machinery, or unknown proteins. Based on the predicted functions of these enzymes, one putative biosynthetic pathway of verrucarol was proposed (**Figure 4-5**). The pathway was starts with the production of trichodiene

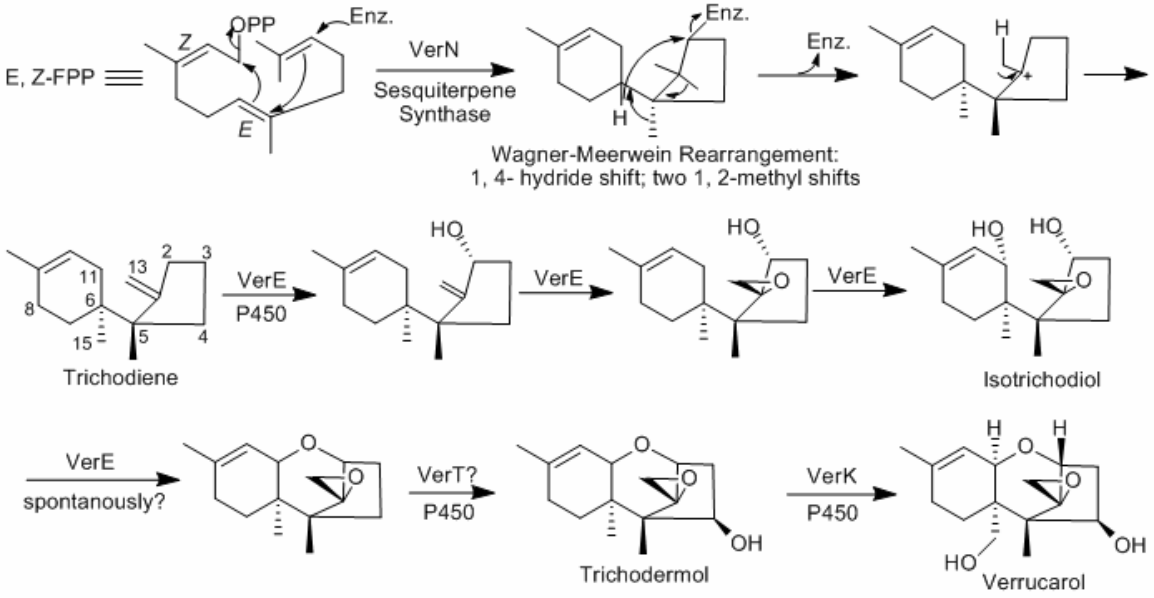


Figure 4-5. The putative pathway for verrucarol product. The putative gene product catalyzing each step was labeled.

from FPP catalyzed by VerN. Subsequently, VerE homologous to Tri4 transformed trichodiene into isotrichodiol after two hydroxylations and one epoxidation. The subsequent ring closure might occur spontaneously under physiological conditions or be promoted by VerE in fungal cells. Trichodermol is produced in a C-4 hydroxylation reaction. This step is catalyzed by another P450, possibly VerT. VerK, another P450 that is similar to Tri11, may catalyze the C-15 hydroxylation reaction to generate verrucarol. The identification of genes for verrucarol production indicated that the isolated gene cluster was responsible for one trichothecene biosynthesis.

Besides verrucarol, verrucarins and roridins also contain polyketide and verrucarinate moieties. VerI was predicted as a fungal iterative PKS and may be involved in the biosynthesis of the polyketide moiety (**Figure 4-6**). Interestingly, this moiety contains six carbons in verrucarins but eight carbon atoms in roridins, indicating VerI may be capable of producing various lengths of polyketides. The second unique feature of this polyketide moiety is the presence of one *cis*-double bond between C-9' and C-10'. The alkenes in the polyketides are generated by the actions of a ketoreductase-dehydratase (KR-DH) didomain, and are generally in the *trans* form [32, 33]. The KR domain catalyzes a 3-keto reduction while the DH domain facilitates the subsequent $\Delta^{2,3}$

elimination reaction. In the formation of a *trans* double bond, a D-3-hydroxyacyl intermediate should be produced while a L-3-hydroxyacyl product indicates the *cis* double bond formation [32, 33]. Remarkably, both *cis* and *trans* double bond are present in the polyketide moiety of verrucarins and roridins, although VerI contains putative KR and DH domains. Thus, it is highly possible that the one DH-independent enzymatic reaction is involved the formation of *cis*-double bond between C-9' and C-10'. Recently, one glutathione *S*-transferase (GST) was proposed to catalyze the *cis-trans* isomerization in the biosynthesis of fungal polyketide hypothemycin [34]. VerA was predicted to be a GST enzyme and may play the same role in trichothecene macrolide biosynthesis (**Figure 4-6**). In verrucarin, one carboxylic group was expected at one end of its polyketide moiety. A P450 involved in biotin biosynthesis was shown to produce the carboxylic diacid after oxidative cleavage of a fatty acid in *Bacillus subtilis* [35]. The substrate in this remarkable P450 reaction was a fatty acid tethered to one fatty acid synthase thiolation (T) domain. In the biosynthesis of trichothecene macrolide verrucarin, such a P450, possibly VerD, was proposed to introduce a carboxylic acid at C-6' position. Subsequently, VerR, a putative long-chain CoA ligase, will activate this group with CoA. In roridin biosynthesis, one epoxide was produced from the double bond between C-6' and

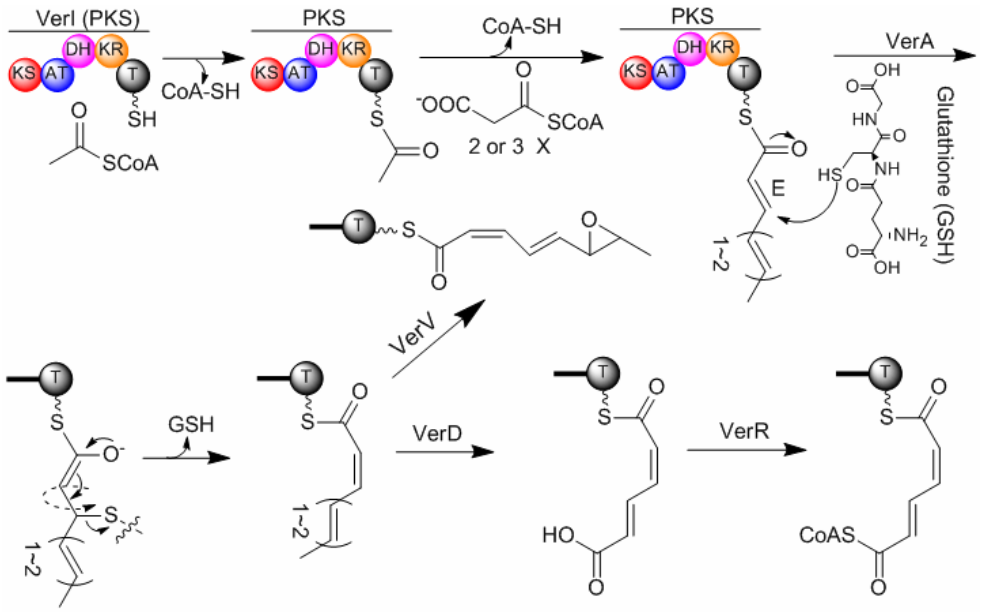


Figure 4-6. One putative pathway for the biosynthesis and modification of the polyketide moiety in trichothecene macrolide.

C-13' (**Figure 4-6**). Possibly, VerV, a putative flavin-binding monooxygenase, will catalyze this reaction. The investigation of VerI and VerA activities is in progress to probe into fungal iterative PKS biosynthesis and *cis-trans* isomerization in fungal polyketides.

Unlike the biosyntheses of the verrucarol and the polyketide moieties in trichothecene macolides, the production of verrucarinate was only reported in this type of trichothecenes and remained obscure. However, the isolation of verrucarin J, verrucarin B, and 2'-dehydroverrucarin A suggests one putative pathway from mevalonate into verrucarinate (**Figure 4-7**). It is proposed that Orf8, one presumable dehydratase, may catalyze the removal of one molecule of water from mevalonate to produce *cis*-anhydromevalonate. The biochemical characterization of this enzyme is in progress to test this hypothesis. In the next step, this newly produced *cis*-double bond was converted into an epoxide possibly under the catalysis of the P450, VerX. How this epoxide ring is opened to produce a ketone functiona group is unclear, but the ketone was reduced to a hydroxyl group by one putative ketoreductase, VerW. It is believed that VerR will activate the terminal carboxylic group of verrucarinate with a CoA.

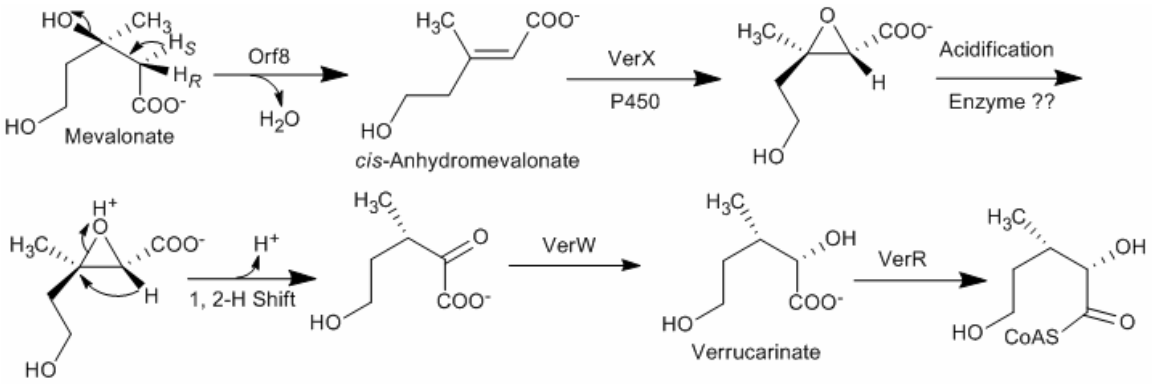


Figure 4-7. One putative pathway for the biosynthesis and modification of verrucarinate moiety in trichothecene macrolide.

The last group of enzymes in trichothecene macrolide biosynthesis was responsible for the assembly of the three above moieties. Both VerF and VerJ share the highly conserved motif (HxxxDG) with Tri3 in *F. sporotrichioides* and *F. graminearum*

(**Figure S4-2A**) [36-38]. Since Tri3 utilizes one acetyl-CoA to acylate C-15-OH of acyclic trichothecenes, acyl-CoA was one of substrates in both VerF and VerJ reactions (**Figure 4-8**). However, it is unclear which enzyme acylates the C-15-OH of verrucarol. The third acyltransferase in the trichothecene macrolide gene cluster is VerQ. This enzyme showed high similarity to Rv1347c in mycobactin biosynthesis (**Figure S4-2B**). The Rv1347c enzyme effectively utilizes both fatty acid bound to the fatty acid synthase T domain and acyl-CoA as its substrate. The conserved active residues, H353 and E391, are common for both VerQ and Rv1347c, indicating that VerQ may cleave the polyketide intermediate from VerI in macrocyclic trichothecene biosynthesis (**Figure 4-8**). It is proposed that the C-4-OH of verrucarol will be positioned in VerQ to attach the polyketide intermediate tethered to VerI T domain. Thus, the last step in trichothecene macrolide biosynthesis is to attack the terminal CoA of the polyketide intermediate by the terminal OH of verrucarinate, which might be catalyzed by either VerF or VerJ. In the roridin biosynthesis, this step is to open the terminal epoxide in polyketide intermediate by the terminal OH of verrucarinate, and one ether bond instead one ester bond is generated. Overall, trichothecene macrolide biosynthesis involves enzymes from sesquiterpene, polyketide, and unique verrucarinate biosynthetic pathways. This work offers numerous possibilities to control mycotoxin production and generate natural product analogs with potential anticancer activities. It also provides basic information to understand natural product production with different origins.

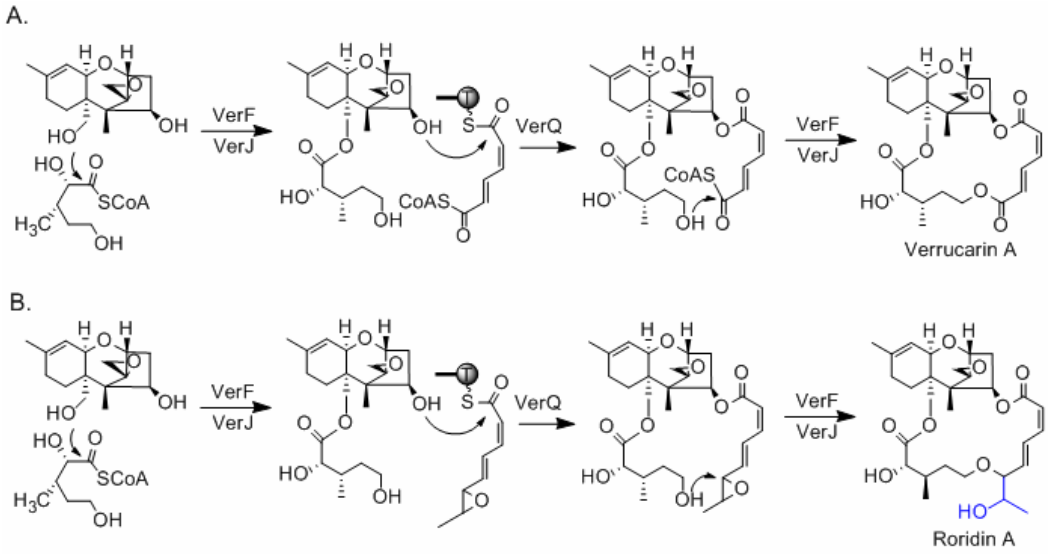


Figure 4-8. One putative pathway to assemble three moieties of trichothecene macrolides. A was proposed for verrucarin A production while B was for roridin A.

4.3.2. Detailed Characterization of One Sesquiterpene Synthase and One Multifunctional P450 in the Verrucarol Biosynthetic Pathway

The above proposed biosynthetic pathways for trichothecene macrolides were preliminarily validated by the biochemical characterization of VerN and VerE in my studies. Bioinformatic analysis indicated that VerN shared three conserved motifs with Tri5s from *F. sporotrichioides* and *M. roridum* (**Figure 4-9**). The first aspartate-rich motif and two divalent metal ions are responsible for FPP binding in sesquiterpene synthase [39]. The third divalent metal ion was positioned to enzyme FPP binding pocket with the assists from the conserved N, S, and E residues in the second motif. The last basic motif was shown to trigger the ionization of the substrate. Thus, VerN was predicted to be a sesquiterpene synthase.

Figure 4-9. Bioinformatic analysis of VerN.

To validate the role of VerN in verrucarol biosynthesis, a 61-bp intron in this 1217-bp gene was removed with the use of overlapping PCR technology. The resultant cDNA was sequenced and cloned into pET28a for overexpression in *E. coli*. Single Ni-NTA resin purification gave VerN in about 80 % purity (**Figure S4-3**). The activity of purified VerN was examined with FPP as the substrate. Two broad peaks were identified in GC-MS analysis (**Figure 4-10**). Each of these products had the molecular ion value at 204. In fact, from the single substrate FPP, various sesquiterpene synthases produce more than 300 known monocyclic, bicyclic, or tricyclic hydrocarbon or alcohol products of widely varied structure and stereochemistry [40]. Moreover, multiple products were also reported in the Tri5 reactions, strongly indicating the feasibility and flexibility of sesquiterpene synthase in natural product biosynthesis [41]. Compared to the retention time

in the GC trace and MS spectrum of authentic trichodiene, the peak from VerN reaction was identified as the trichodiene with retention time of 9.18 min (**Figure 4-10**, **Figure S4-4**). The production of one key biosynthetic intermediate by VerN indicated that the trichothecene biosynthetic gene cluster was isolated from *M. verrucarin*.

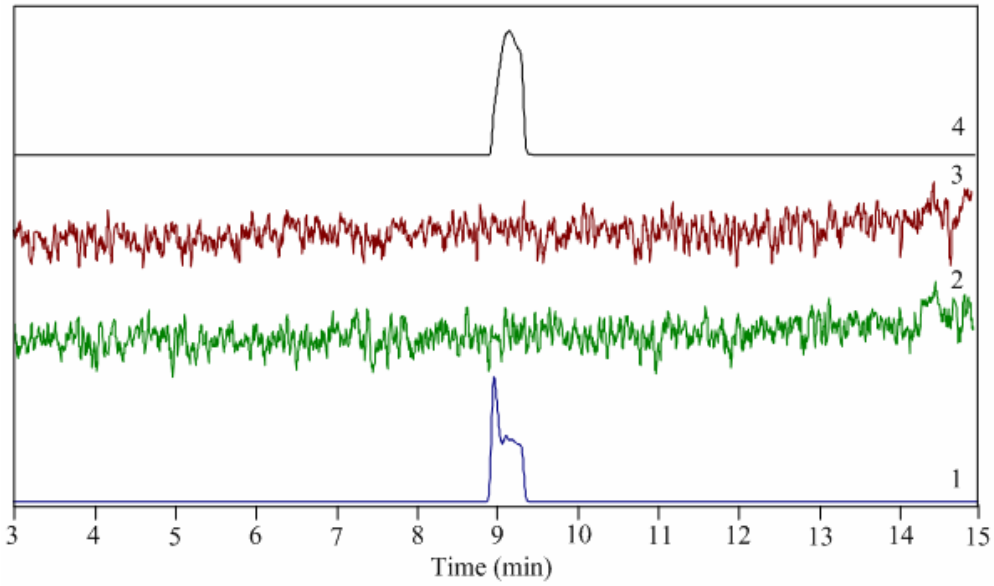


Figure 4-10. GC-MS analysis of VerN reactions. Lane 1, VerN full reaction; lane 2, VerN reaction without Mg^{2+}; lane 3, enzyme reaction with boiled VerN as enzyme source; lane 4, authentic trichodiene.

In the isolated gene cluster, five P450s were predicted and possibly required for trichothecene macrolide biosynthesis. All of these remarkable biocatalysts contained three conserved motifs (**Figure S4-5**). The first motif is responsible for protonating the reactive iron-oxygen intermediate, while the Cys residue in the third motif serves as the fifth ligand of the heme iron. Interestingly, all of these fungal P450s contain a 30-AA transmembrane segment at their N-termini. VerE was chosen for further investigation in this study. This enzyme may convert trichodiene into isotrichodiol by two hydroxylations and one epoxide in an unprecedented manner. Four introns were removed from its gDNA to produce its corresponding cDNA by overlapping PCR technology. Initial efforts to produce soluble and active full-length and truncated VerE in *E. coli* failed, possibly indicating the codon bias between *E. coli* and the marine fungal strain. Thus, the *VerE* cDNA was cloned into the pYES2/C NT vector for its overexpression in yeast

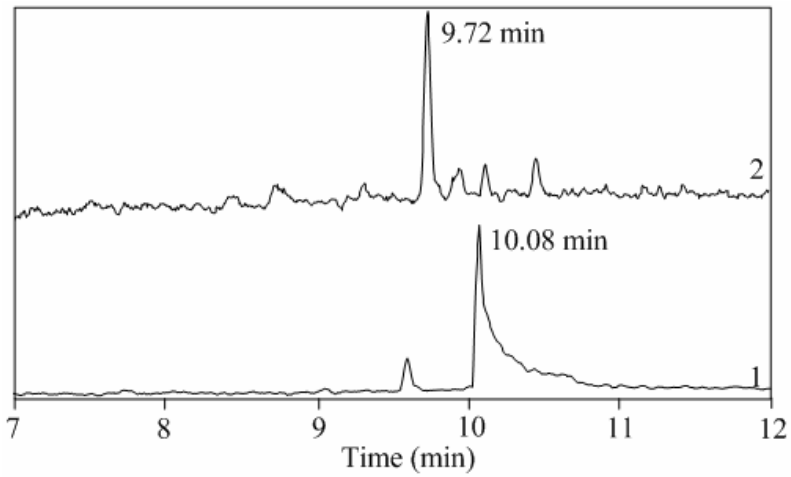


Figure 4-11. GC-MS analysis of VerE reaction. In the coupled reactions including VerN and VerE from the yeast microsomes, FPP was used as the substrate. In SIM mode, 12,13-epoxy-9,10-trichoene-2α-ol (*m/z* = 236.2) was identified as the peak with retention time of 10.08 min in lane 1 while 2α-hydroxytrichodiene (*m/z* = 220.2) was found with retention time of 9.72 min in lane 2.

Saccharomyces cerevisiae. The resultant yeast microsomes containing VerE were isolated and used as the source of enzymes to investigate the P450 reactions (**Figure 4-11**). In the preliminary assay, two peaks, one for 2α-hydroxytrichodiene (9.72 min) and the other for 12,13-epoxy-9,10-trichoene-2α-ol (10.08 min), were identified in SIM mode. These peaks were not observed in the single enzyme (VerN or VerE) reaction, and they were characterized by the expected *m/z* values (**Figure S4-6**). Further validation of these putative products is in progress. Overall, the preliminary analysis of VerE indicated that this P450 was able to catalyze one hydroxylation and one epoxidation reaction on trichodiene, as would be expected for it to be involved in the early stage of the biosynthetic pathway of trichothecene macrolide.

4.4. Materials and Methods

General Chemicals, DNA Sub-cloning, and Bacterial Strains. Chemicals were purchased from Sigma-Aldrich and Fisher Scientific. FPP was purchased Isoprenoids. Authentic trichodiene, isotrichodiol, and trichothecene were kindly provided by Dr. Susan McCormick at the USDA (Peoria, IL). *Myrocethium verrucaria* was isolated from Hawaii *Spongia* sp. and was provided by Professor Philip Crews in University of

California-Santa Cruz. The yeast overexpression system was from Invitrogen. Standard methods for DNA isolation and manipulation were performed as described by Sambrook *et al* [42]. Genomic DNA from *M.verrucaria* was isolated with a MasterPure Yeast DNA Purification kit (Epicentre Biotechnologies) as described in its manual. Molecular biology reagents and enzymes were supplied by New England Biolabs, with the exception of for *Pfu* DNA polymerase (Stratagene), dNTPs (Takara), and T4 DNA Ligase (Invitrogen). *Escherichia coli* XL-1 Blue was used for cloning and plasmid harvesting while *E. coli* BL21 (DE3) or *E. coli* BL21 CodonPlus-(DE3)-RIPL was used for protein overexpression. All *E. coli* strains were grown in Luria-Bertani broth. DNA sequencing was performed at the University of Michigan DNA Sequencing Core.

Construction of Fungal Genomic DNA Library. *M. verrucaria* was grown in 50 ml of 1.5 % malt extract broth in artificial sea water at room temperature without shaking. Fungal mycelium was collected by filtering through Waterman filter paper #4 and washed with 100 ml 50 mM Tris-Cl, pH 7.5. Mycelia were dried with paper and ground into powder with mortar and pestle in liquid nitrogen. Mycelia powder was transferred to one sterile 1.5-ml Eppendorf tube, and fungal genomic DNA was isolated with an Epicentre Masterpure Yeast DNA Purification Kit. The manufacturer protocol was followed. The size of the isolated fungal gDNA was checked with pulse field gel electrophoresis and was estimated to be about 40 kb. An Epicentre CopyControl™ Fosmid Library Production Kit was used to construct the fungal gDNA library. The protocols in the kit were used with minor modifications, but the end-repaired DNA was not purified from the gel. Instead, DNA was directly precipitated, dried, and re-dissolved for library construction. A genomic DNA library was then screened by PCR with one pair of degenerate primers designed on the basis of KS domain, one pair of primers on *Tri4* from *M. roridium*, one pair of primers on *Tri5* from *M. roridium*, one pair on the identified *CoA ligase* region, one pair on the 5'-end of assembled sequence of fosmid N1 and C10, and one pair on the 3'-end of assembled sequence of fosmid N1 and C10. In total, four fosmids were isolated and subjected to shotgun sequencing. The sequencing results were analyzed and assembled to reveal a 101638-bp chromosomal region by the use of SeqMan in Lasergene 6 package. BLASTX in NCBI

(http://www.ncbi.nlm.nih.gov/BLAST/) was used to predict the functions of each open reading frame (ORF). Subsequently, BCM Gene Finder (http://www.bioscience.org/urllists/genefind.htm) was used to predict putative introns in each ORF.

Gene Cloning, Expression and Protein Purification. Primers in Table **S4-1** were designed to remove intron(s) from the full-length *VerN* and *VerE* and to produce te corresponding cDNAs by overlapping PCR. The purified *VerN* cDNA was ligated to pET28a for transformation of *E. coli* DH5α competent cells. The constructs were isolated and submitted for sequencing to exclude errors introduced during PCR amplification and gene manipulation. The construct was transformed into *E. coli* BL21 (DE3) competent cells for protein expression. Cells harboring the construct were cultured in LB media containing 25 μg/ml of kanamycin and grown at 37 °C in a shaker at 200 rpm until OD_{600} reached 0.6. The culture was then induced by IPTG with a final concentration of 0.2 mM and was further grown at 16 °C in a shaker at 180 rpm for 16 hours. The cells were harvested by centrifugation (7250 *g*, 12 min, and 4 °C) and cell pellets were stored at -80 °C or used directly for protein purification. The cell pellet was re-suspended in 40 ml binding buffer (50 mM Tris-Cl, pH 7.5, 200 mM NaCl, 3 mM β-mercaptoethanol, 10 % glycerol) with 10 mM imidazole. The resuspended cell solutions were sonicated to release the soluble proteins. The soluble supernatants were collected and incubated with 2.5 ml pre-equilibrated Ni-NTA agarose resin (Qiagen) after centrifugation at 36 000 *g* at 4 °C for 32 min. After a 2-hour incubation at 4 °C with gentle shaking, the resin was washed with 50 ml binding buffer containing 10 mM imidazole, and then with 50 ml binding buffer containing 20 mM imidazole. Recombinant His_6-VerN gradually eluted with the elution buffer (50 mM Tris-Cl, pH 7.5, 50 mM NaCl, 3 mM β-mercaptoethanol, 10 % glycerol, and 200 mM imidazole). The protein was further purified with FPLC. The protein fractions were then combined and the concentration was measured by its specific absorbance at 280 nm (ε = 77900 M^{-1} cm^{-1}). The protein was aliquoted and stored at -80 °C for use.

Yeast Transformation and Protein Overexpression. VerE was obtained as a microsomal preparation from yeast that had been transformed with its cDNA [43]. Briefly, the pYES2/C NT expression vector carrying *VerE* cDNA was introduced in *Saccharomyces cerevisiae* INVSc1 from Invitrogen. A single transformed yeast colony was selected and grown at 30°C for approximately 24 h in 15 ml SC selective medium (without uracil) containing 2% glucose as the carbon source. Upon depletion of the glucose in the medium, expression of *VerE* cDNA was initiated by the addition of galactose to a final concentration of 2% in 50 ml SC selective medium. The cultures were maintained at 30°C for an additional 16 h before collecting the cells by centrifugation at 7000 g for 10 min. The pelleted cells were then washed with 100 mL of TES buffer (50mM Tris–HCl, pH, 7.5, 1mM EDTA, 0.6M sorbitol), centrifuged as above, resuspended in 100 mL of TES-M (TES supplemented with 10 mM BME), and allowed to incubate at room temperature for 10 min. The cells were collected by centrifugation, and the pellet was resuspended in 1 mL of extraction buffer (1% bovine serum albumin (fraction V), 2 mM BME, and 1mM PMSF in TES buffer). Glass beads (0.2-0.4 mm in diameter) (cleaned by rinsing in ethanol, dilute acid, and copious amounts of water) were added until they occupied approximately 90 % of the resuspended cell volume and the cells were disrupted by vigorously shaking the mixture in a cold room for 10 min at 30-s intervals separated by 30-s intervals on ice. The cell extract was decanted into a centrifuge tube and combined with three washes of the glass beads, each with 0.5-1 mL of extraction buffer. Microsomes were prepared by differential centrifugation at 10,000 g for 10 min at 4°C to remove cellular debris followed by centrifugation at 107,000 g for 70 min at 4°C. The final microsomal pellets were resuspended in 0.5-1.0 mL of TEG-M buffer (50mM Tris–HCl, pH 7.5, 1 mM EDTA, 20% glycerol, and 1.5 mM BME) and stored frozen at -80 °C until use.

Enzyme Assays. **A). VerN** The enzyme reaction mixture contained 10 mM Tris-Cl, pH 7.5, 5 mM $MgCl_2$, 15 % glycerol, 5 mM BME, 250 μM FPP, and 0.5 μM VerN in one 200-μl total volume. The reaction mixture was overlaid with HPLC-grade n-pentane (10% overall volume) in a glass test tube at 30°C for 30 minutes. The reaction mixture was then extracted with HPLC-grade n-pentane (10 %) twice and the extracts were

combined and purified on a 3 cm 230-400 mesh silica gel column. The extract was further concentrated on an ice-water mixture under reduced pressure. One 10-µl aliquot was injected in a splitless mode to GC-MS to analyze the enzyme product. The GC-MS analysis was conducted with a Finnigan Trace GC/MS machine. Finnigan Trace GC 2000 was installed with a DB-1 capillary column (25 meters, 0.2 mm I.D., 0.22 µm film thickness) in a split injector. The split ratio is 1:100. The injection temperature was 120 °C, which was held for 2 min. The initial ramp with a rate of 6 °C/min was utilized to increase the temperature to 210 °C, which was held for 1 additional min. The MS used the electron impact mode with detection voltage of 300 V.

B). VerE. Purified microsome was used in VerE reactions with trichodiene as the substrate. Otherwise, VerE activity was examined in the coupled reaction containing VerN and FPP. The reaction mixture contained 50 mM Tris-Cl, pH 7.5, 10 % glycerol, 3 mM BME, 5 mM $MgCl_2$, 250 µM FPP, 1 µM VerN, 0.5 mM NADPH, and 30 µl of enzyme microsome in a 200-µl total volume. The reaction mixture was overlaid with HPLC-grade n-pentane (10% overall volume) in a glass test tube at 30°C for 2 hours, and reaction mixture was then extracted with HPLC-grade n-pentane (10 %) twice. The combined extracts were concentrated on an ice-water mixture under reduced pressure. One 10-µl aliquot was injected to GC-MS to analyze the enzyme product. The injection temperature was 120 °C, which was held for 2 min. The initial ramp with a rate of 15 °C/min was utilized to increase the temperature to 210 °C, which was held for 1 additional min. Subsequently, the final temperature of 260 °C was reached with the second ramp of 10 °C/min and was held for 3 min. The MS used the electron impact mode with detection voltage of 300 V

4.5. Supplementary Figures and Tables

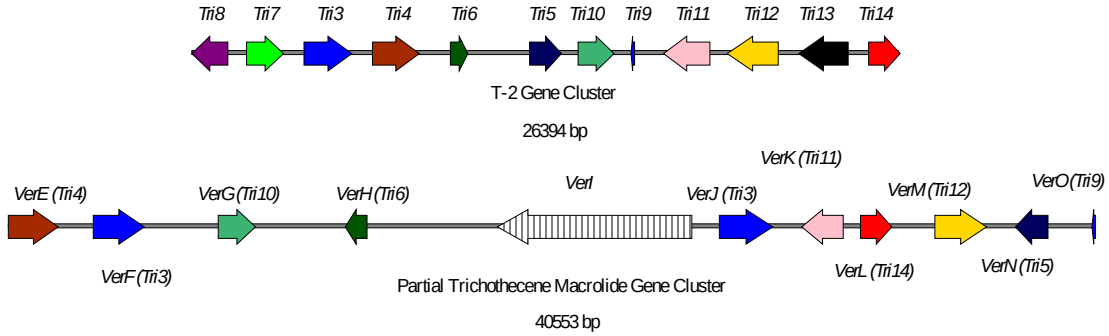


Figure S4-1. Several genes shared in the T-2 toxin gene cluster and the trichothecene macrolide gene cluster.

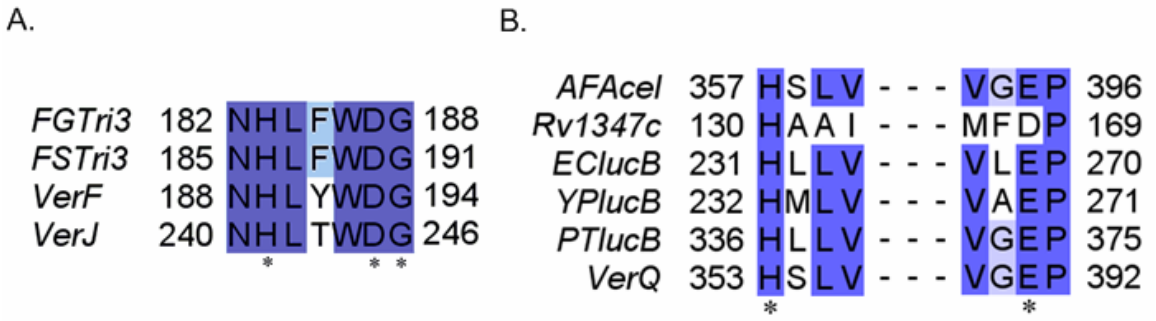


Figure S4-2. A. One conserved motif shared by Tri3 and VerF and VerJ. B. One conserved motif shared by VerQ and Rv1347c and other unique acyltransferases.

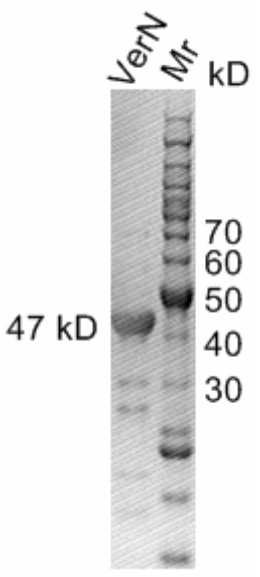


Figure S4-3. SDS-PAGE analysis of purified VerN.

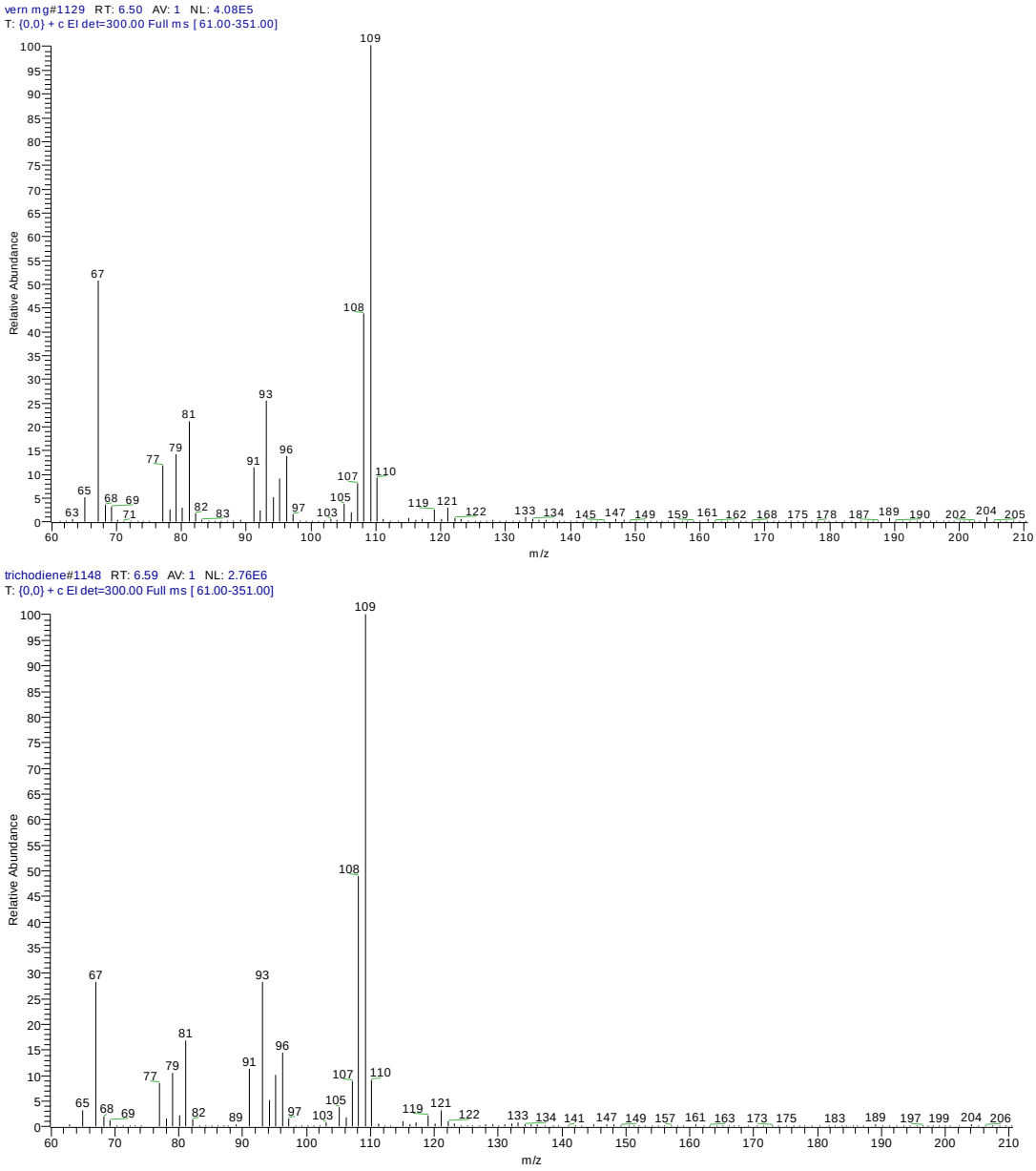


Figure S4-4. MS spectra of VerN product (top) at 9.18 min and authentic trichodiene (bottom).

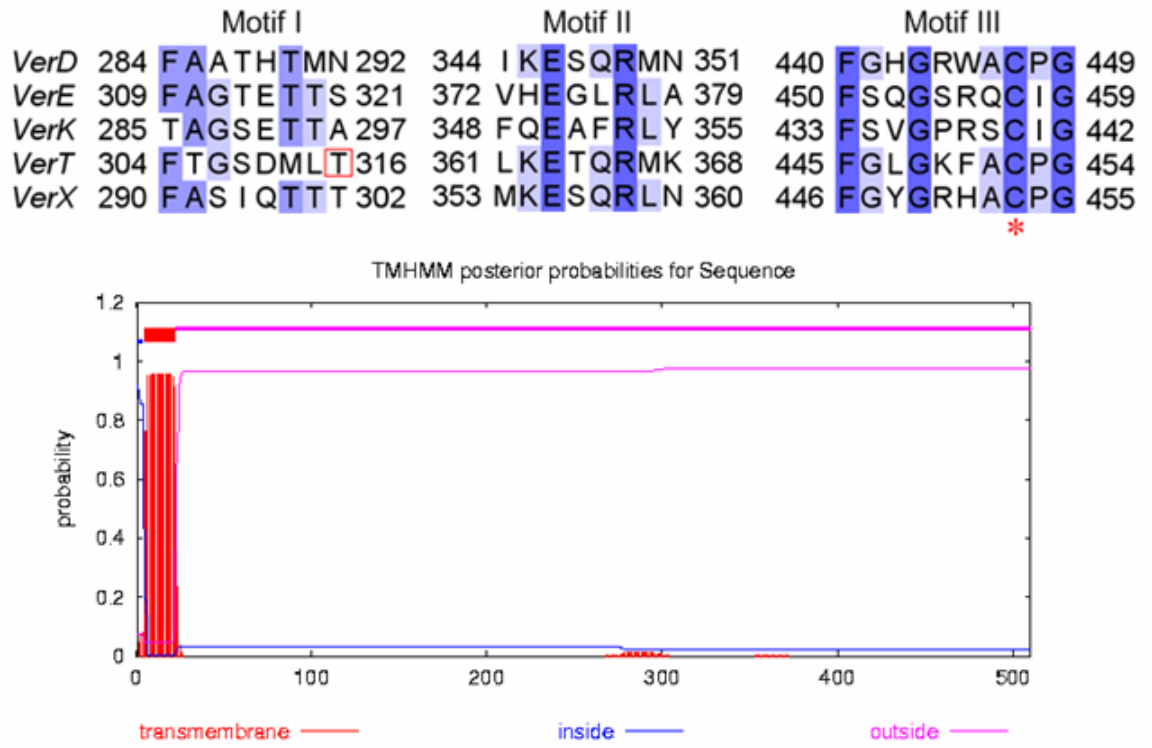


Figure S4-5. Three conserved motifs are shared by five P450s involved in trichothecene macrolide biosynthesis. Moreover, all of these enzymes contained a transmembrane segment at their N-termini.

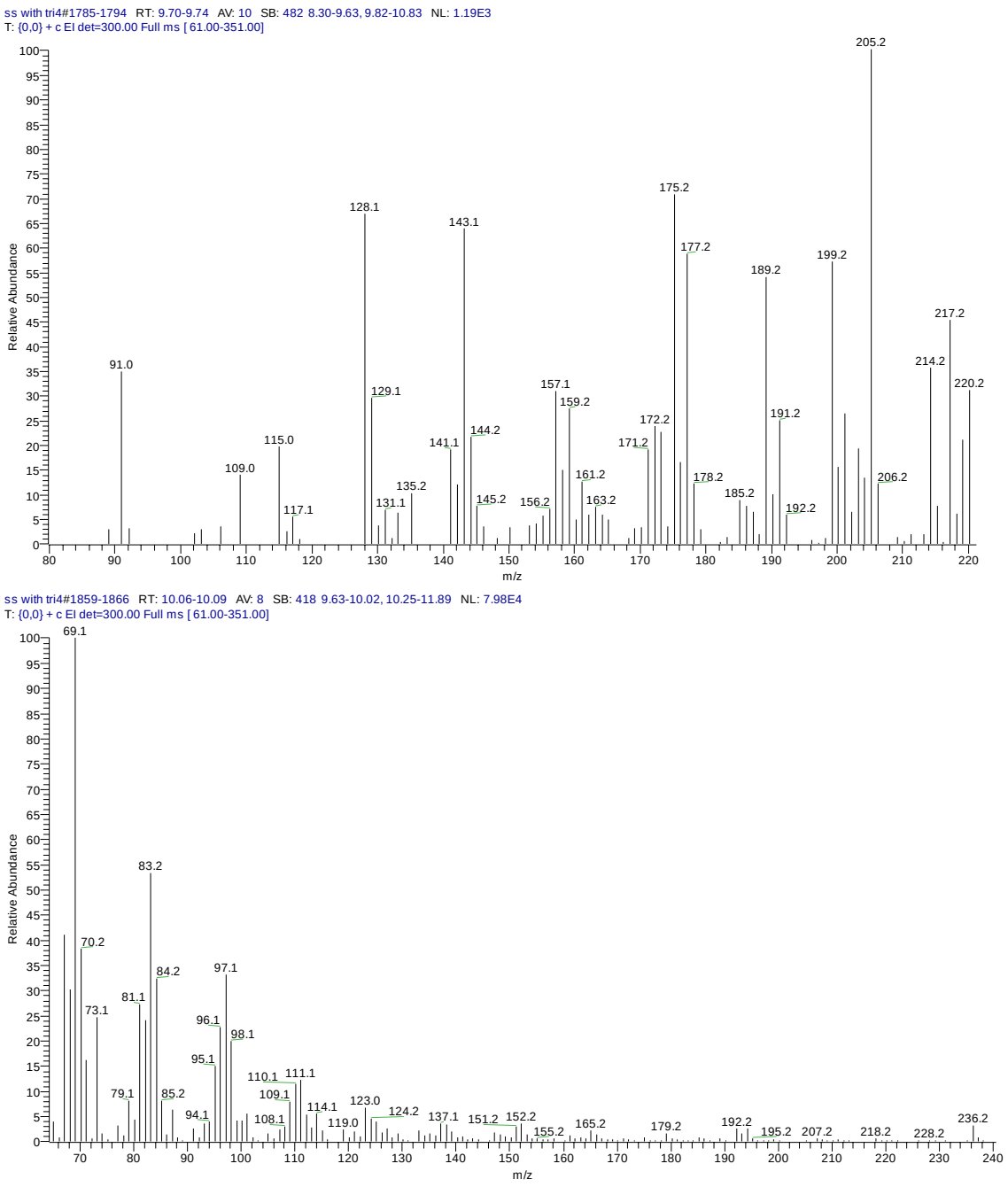


Figure S4-6. SIM spectra of 2α-hydroxytrichodiene (top) and 12,13-epoxy-9,10-trichoene-2α-ol (bottom)

Table S4-1 Primers for gene cluster isolation, intron removal, and gene expression

Name	Sequence (5'→3')	Function
KS3	TTYGAYGCIGCITTYTTYAA	*KS* domain
KS4C	RTGRTTIGGCATIGTIATICC	*KS* domain
Mrtri4F	TTCTACCATGAAGTCATCCG	*Mrtri4* primer
Mrtri4R	ATCCATCGCTCAGGCTTGAA	*Mrtri4* primer
Mrtri5F	CTTGTCAACGAGCACTTCCCCAACG	*Mrtri5* primer
Mrtri5R	GCCAGGTGATGTAGCCGTGCATAAA	*Mrtri5* primer
N1CLF	TGGCAGGTTACCAACTACAG	*CoA ligase* primer
N1CLR	CAAGGGAGCGATTCTTTC	*CoA ligase* primer
DY5F	CATAGCAGATCACACCGGTTATTC	Mv-92-1-9F isolation
DY5R	CGTTGGTTATCGTGCTGTGAATC	Mv-92-1-9F isolation
DY3F	CGTTTGGAGAGTGTGTGCAGT	Mv-94-1-9B isolation
DY3R	GACACCTACTGACCTTCCCAGT	Mv-94-1-9B isolation
VerNF(*Nde*I)	AACCCATATGGATACCTTCCCCACC	VerN expression
VerNR(*Hind*III)	CCGAAAGCTTTTACACCAGAACCTCATG	VerN expression
VerNIF	ACGCTTGACTTCGAGGGCTGCTGGATT	VerN intron removal
VerNIR	GCCCTCGAAGTCAAGCGTGCTGCGAAT	VerN intron removal
VerEF(*Eco*RI)	GCGGAATTCATGCCCTCCTTCTCTGAT	VerE yeast expression
VerER(*Xho*I)	GCGCTCGAGTTAAATCAACTGCTTGGC	VerE yeast expression
VerEI1F	AGATGCACGAGAAATATGCACAAGGCCCCATTGTCC	VerE intron1 removal
VerEI1R	GGACAATGGGGCCTTGTGCATATTTCTCGTGCATCT	VerE intron1 removal
VerEI2F	GAAAGTCTTCCTTACTTGAGCGGTGTCGTCCATGAG	VerE intron2 removal
VerEI2R	CTCATGGACGACACCGCTCAAGTAAGGAAGACTTTC	VerE intron2 removal
VerEI3F	TACACTATTCCCGCTGGAACCCCATTGAGCATGTCC	VerE intron3 removal
VerEI3R	GGACATGCTCAATGGGGTTCCAGCGGGAATAGTGTA	VerE intron3 removal
VerEI4F	ACAGTGCATTGGTATCAACATGTCTTTTGCCGAAAT	VerE intron4 removal
VerEI4R	ATTTCGGCAAAAGACATGTTGATACCAATGCACTGT	VerE intron4 removal

Table S4-2 Putative VerH binding sites in the promoter region of each trichothecene macrolide biosynthetic gene

Gene Name	Binding Site (5' → 3')	Position	Gene Name	Binding Site (5' → 3')	Position
VerA	CCAGGCC	18537-18543	*VerM*	TCAGGCCT	61450-61457
VerB	GG***G***CTCG	21468-21474	*VerN*	AGGCCTAG	66812-66819
VerC	-	-		GGCCTTA	67700-67706
VerD	GGCCTCA	26485-26491	*VerO*	AGGCCTGA	68557-68564
	AGGCCTGA	26634-26641	*VerP*	GGCCTTA	71148-71154
	AGGCCTGA	26681-26688	*VerQ*	TAAGGCC	71066-71072
	GGCCTGA	27234-27240	*VerR*	AGGCCTTG	75686-75693
VerE	TAAGGCCT	25583-25590	*VerS*	TTAGGCC	75149-75155
	TAAGGCC	26338-26344		TTAGGCC	75171-75177
	CGAGGCC	26995-27001		TAAGGCCT	75359-75366
VerF	CTAGGCC	29394-29400		TCAGGCCT	75684-75691
	TCAGGCCT	30410-30419		TTAGGCC	76005-76011
VerG	CTAGGCC	33626-33632	*VerT*	GGCCTTA	80560-80566
	TTAGGCC	33658-33664	*VerU*	AGGCCTTA	83768-83775
	TCAGGCC	33954-33960	*VerV*	GGCCTGA	86841-86847
	TAAGGCC	34304-34310	*VerW*	GGCCTTA	89983-89989
	TCAGGCC	34950-34956		GGCCTGA	90484-90490
	CTAGGCC	35011-35017		GGCCTAA	91145-91151
VerH	GGCCTTA	43956-43962		AGGCCTAA	91195-91202
	AGGCCTCA	43968-43975	*VerX*	TCAGGCC	90886-90892
VerI+J	TAAGGCCTGA	53524-53533	*VerY*	AGGCCTAG	96007-96014
	TACGGCCTCA	53760-53769	*VerZ*	TAAGGCC	95091-95097
VerK	GGCCTGA	58983-58989		TAAGGCCT	96005-96012
	GGCCTGA	59157-59163			
VerL	TGAGGCCT	58836-58843			
	CTAGGCC	58898-58904			
	CAAGGCC	58952-58958			

4.6. Reference

1. Keller, N. P.; Turner, G.; Bennett, J. W., *Nat Rev Microbiol* **2005,** 3, (12), 937-47.
2. Squire, R. A., *Science* **1981,** 214, (4523), 877-80.
3. Azziz-Baumgartner, E.; Lindblade, K.; Gieseker, K.; Rogers, H. S.; Kieszak, S.; Njapau, H.; Schleicher, R.; McCoy, L. F.; Misore, A.; DeCock, K.; Rubin, C.; Slutsker, L., *Environ Health Perspect* **2005,** 113, (12), 1779-83.
4. Stockmann-Juvala, H.; Savolainen, K., *Hum Exp Toxicol* **2008,** 27, (11), 799-809.
5. Bush, B. J.; Carson, M. L.; Cubeta, M. A.; Hagler, W. M.; Payne, G. A., *Phytopathology* **2004,** 94, (1), 88-93.
6. Ueno, Y., *Crit Rev Toxicol* **1985,** 14, (2), 99-132.
7. Grove, J. F., *Fortschr Chem Org Naturst* **2007,** 88, 63-130.
8. Desjardins, A. E., *J Agric Food Chem* **2009,** 57, (11), 4478-84.
9. Liao, L. L.; Grollman, A. P.; Horwitz, S. B., *Biochim Biophys Acta* **1976,** 454, (2), 273-84.
10. Rocha, O.; Ansari, K.; Doohan, F. M., *Food Addit Contam* **2005,** 22, (4), 369-78.
11. Thompson, W. L.; Wannemacher, R. W., Jr., *Toxicon* **1986,** 24, (10), 985-94.
12. Miller, J. D.; Ewen, M. A., *Nat Toxins* **1997,** 5, (6), 234-7.
13. Shifrin, V. I.; Anderson, P., *J Biol Chem* **1999,** 274, (20), 13985-92.
14. Jarvis, B. B.; Midiwo, J. O.; Mazzola, E. P., *J Med Chem* **1984,** 27, (2), 239-44.
15. Jarvis, B. B.; Stahly, G. P.; Pavanasasivam, G.; Mazzola, E. P., *J Med Chem* **1980,** 23, (9), 1054-8.
16. Kaneko, T.; Schmitz, H.; Essery, J. M.; Rose, W.; Howell, H. G.; O'Herron, F. A.; Nachfolger, S.; Huftalen, J.; Bradner, W. T.; Partyka, R. A.; Doyle, T. W.; Davies, J.; Cundliffe, E., *J Med Chem* **1982,** 25, (5), 579-89.
17. Corbett, T. H.; Griswold, D. P., Jr.; Roberts, B. J.; Peckham, J. C.; Schabel, F. M., Jr., *Cancer* **1977,** 40, (5 Suppl), 2660-80.
18. Kimura, M.; Tokai, T.; Takahashi-Ando, N.; Ohsato, S.; Fujimura, M., *Biosci Biotechnol Biochem* **2007,** 71, (9), 2105-23.
19. Hohn, T. M.; Vanmiddlesworth, F., *Arch Biochem Biophys* **1986,** 251, (2), 756-61.
20. Hohn, T. M.; Desjardins, A. E.; McCormick, S. P., *Mol Gen Genet* **1995,** 248, (1), 95-102.
21. McCormick, S. P.; Alexander, N. J.; Proctor, R. H., *Can J Microbiol* **2006,** 52, (7), 636-42.
22. Kimura, M.; Shingu, Y.; Yoneyama, K.; Yamaguchi, I., *Biosci Biotechnol Biochem* **1998,** 62, (5), 1033-6.
23. Alexander, N. J.; Hohn, T. M.; McCormick, S. P., *Appl Environ Microbiol* **1998,** 64, (1), 221-5.
24. Kimura, M.; Kaneko, I.; Komiyama, M.; Takatsuki, A.; Koshino, H.; Yoneyama, K.; Yamaguchi, I., *J Biol Chem* **1998,** 273, (3), 1654-61.
25. Lee, T.; Han, Y. K.; Kim, K. H.; Yun, S. H.; Lee, Y. W., *Appl Environ Microbiol* **2002,** 68, (5), 2148-54.
26. Dennis, D.; Robertson, D.; Curtis, L.; Black, J., *Toxicol Ind Health* **2009**.
27. Pestka, J. J.; Yike, I.; Dearborn, D. G.; Ward, M. D.; Harkema, J. R., *Toxicol Sci* **2008,** 104, (1), 4-26.
28. Jarvis, B. B., *Phytochemistry* **2003,** 64, (1), 53-60.

29. Trapp, S. C.; Hohn, T. M.; McCormick, S.; Jarvis, B. B., *Mol Gen Genet* **1998,** 257, (4), 421-32.
30. Amagata, T.; Rath, C.; Rigot, J. F.; Tarlov, N.; Tenney, K.; Valeriote, F. A.; Crews, P., *J Med Chem* **2003,** 46, (20), 4342-50.
31. Hohn, T. M.; Krishna, R.; Proctor, R. H., *Fungal Genet Biol* **1999,** 26, (3), 224-35.
32. Caffrey, P., *Chembiochem* **2003,** 4, (7), 654-7.
33. Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R., *Biochemistry* **2003,** 42, (1), 72-9.
34. Reeves, C. D.; Hu, Z.; Reid, R.; Kealey, J. T., *Appl Environ Microbiol* **2008,** 74, (16), 5121-9.
35. Cryle, M. J.; Schlichting, I., *Proc Natl Acad Sci U S A* **2008,** 105, (41), 15696-701.
36. Garvey, G. S.; McCormick, S. P.; Alexander, N. J.; Rayment, I., *Protein Sci* **2009,** 18, (4), 747-61.
37. Garvey, G. S.; McCormick, S. P.; Rayment, I., *J Biol Chem* **2008,** 283, (3), 1660-9.
38. Alexander, N. J.; McCormick, S. P.; Hohn, T. M., *Yeast* **2002,** 19, (16), 1425-30.
39. Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W., *Proc Natl Acad Sci U S A* **2001,** 98, (24), 13543-8.
40. Christianson, D. W., *Chem Rev* **2006,** 106, (8), 3412-42.
41. Vedula, L. S.; Jiang, J.; Zakharian, T.; Cane, D. E.; Christianson, D. W., *Arch Biochem Biophys* **2008,** 469, (2), 184-94.
42. Sambrook, J.; Russel, D. W., *Molecular Cloning - A Laboratory Manual*. 3rd ed.; Cold Spring Harbor Laboratory Press: New York, 2001.
43. Greenhagen, B. T.; Griggs, P.; Takahashi, S.; Ralston, L.; Chappell, J., *Arch Biochem Biophys* **2003,** 409, (2), 385-94.

Notes:

Yousong Ding and David H. Sherman designed the experiments and Yousong Ding performed the experiments.

Dr. Pamela Schultz, Dr. Zachary Beck, and Dr. Clem Fortman initiated the project by preparing the first genomic DNA library, and identifying and sequencing fosmids N1 and C10.

Chapter 5

Summary and Future: Natural Product Biosynthesis and Drug Development

Natural products contribute to the majority of currently used drugs in the market. However, new and drug-resistant diseases have been reported more frequently in recent decades. Along with the advances in identification and characterization of novel targets, numerous secondary metabolites have been isolated, purified, and characterized from different environmental niches. For example, 961 new natural products were isolated from marine organisms in 2007, an increase of 24% from 2006 [1]. These new natural products not only attracted more efforts to understand their production in Nature, but also provided a tremendous source of novel chemical architectures for the development of new analogs with the hope to identify useful drugs.

Undoubtedly, chemical synthesis is the primary method for generating the most drugs used to treat various diseases. However, the required asymmetric synthesis steps remain challenging for many complex natural product drug candidates or analogs. Moreover, chiral centers in drugs critically contribute to their bioactivities and specificities. For example, over 54% of current drugs contain at least one stereogenic centre, and most must be developed as single stereoisomers [2]. As a result, the chemical route leads to high cost and lengthy separation process, and also produces undesired side products. On the other hand, the unbeatable selectivity of biocatalysts makes chemoenzymatic approaches to be energy efficient and environment friendly. The development of biocatalysts has been further boosted by the exponential growth in publicly available genome sequences and the recent advances in metagenomics, high throughput screening technology, and directed enzyme evolution [3-5]. As one important and integral part of new drug development, studies of natural product biosynthesis

have characterized a considerable number of excellent biocatalysts.

In my studies, several enzymes from natural product biosynthetic pathways have been identified and characterized. These studies not only deeply probed into the natural product assembly, but also provided a basis for generating both natural and unnatural analogs. For example, the regio- and stereo-selective cryptophycin P450 epoxidase catalyzes a β-epoxidation reaction in the maturation of bioactive cryptophycins in cyanobacterium *Nostoc* sp. More importantly, this enzyme was utilized in a chemoenzymatic approach to overcome the most challenging step in the natural product chemical synthetic route. Similar to the chemical syntheses of many other macrolactones and macrolactams, inefficient and unspecific macrocyclization challenged organic chemists to properly produce cryptophycins with traditional methods. The cryptophycin TE domain exhibited excellent catalytic ability to overcome this issue in cryptophycin production. Moreover, the substrate tolerance of this enzyme offers more opportunities to generate novel analogs for clinical evaluation. Besides these biocatalysts from bacterial strains, fungi also are rich sources for pharmaceutically valuable natural products and practical biocatalysts. In numerous fungal alkaloid biosyntheses and chemical syntheses, deoxybrevianamide E and its analogs serve as the key precursors. NotD identified in my studies is a putative biocatalyst to specifically and efficiently produce this compound for pharmaceutical production [6].

To be considered as practical industrial biocatalysts, enzymes need to contain a number of features such as high catalytic turnover (>500 min^{-1}), strict selectivity for the particular transformation, and good process stability under the non-native conditions [7]. In most cases, the enzymes must be optimized through direct evolution to improve their thermostability, pH tolerance, and catalytic capability [5]. For many enzymes that I studied, further optimization may be necessary before their practical application in large scale. For example, the inconsistent performance of CrpE may be derived from its instability. Thus, one important future direction in cryptophycin studies will be to improve CrpE stability and catalytic efficiency. Based on my studies, the MBP tag seemed to be required to maintain the solubility of enzyme in the reaction. Although the tag itself

causes no harm to enzyme activity, the codon optimized *CrpE* gene may increase enzyme overexpression level in *E. coli* strain, which may be inversely correlated to proper folding of the nascent polypeptide in heterologous host. For cryptophycin TE, avoiding of hydrolytic activity may be critical for its application in generating the large quantities of anticancer agent analogs. NotD showed very strict substrate selectivity to brevianamide F. To make it to be more applicable, its crystal structure will be important for understanding the structural determinants contributing to enzyme specificity. These determinants will be manipulated in the next step to expand the NotD substrate library to generate more reversibly prenylated cyclo-dipeptide analogs for drug development.

Besides identification of biocatalysts from natural product biosynthetic pathways, another valuable contribution is to produce natural products and their analogs in heterologous hosts. Complex secondary metabolites are produced by their producers mainly to compete with other habitants and/or to sense surroundings in natural environments. This can be a potential limitation in obtaining large quantity of natural products as leads in drug discovery and development. Moreover, more than 99 % of microorganisms in the environment fail to grow in the laboratory while the potential to find pharmaceutically important natural products from this unexplored source seems limitless. Thus, it is critical in natural product drug development to introduce natural product gene clusters into more technically- and industrially-amenable microorganisms such as *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas putida*, *Saccharomyces cerevisiae*, or *Streptomyces coelicolor* [8]. Furthermore, heterologous product of secondary metabolites offers possibility to manipulate their gene clusters to produce novel analogs or fine tune their production yields. The cryptophycin gene cluster represents an idea model for future hetereologous production. The codon bias between the cyanobacterium *Nostoc* sp. and a foreign strain such as *E. coli* will be considered. Moreover, optimizing the promoter for use in the *E. coli* system will also be important to improve product yield. Other strategies to facilitate cryptophycin production in *E. coli* may be to amplify rate-controlling reactions and to remove unnecessary genes and/or regulators. Heterologous production of fungal secondary metabolites may be more challenging, but yeast strains have been industrially applied to produce many pharmaceuticals and their precursors

such as taxol precursor and nicotianamine [9]. The identification of the notoamide gene cluster sheds lights on how unique prenylated indole alkaloids with diversified bioactivities are produced. Moreover, it will be one of the future directions to continue this study by heterologous production of these fungal alkaloids in yeast strain.

Recent advances in system biology and synthetic biology make it possible to perform metabolic engineering at the whole cell level, thus enabling more optimal design of a microorganism for the efficient production of drugs and drug precursors. Along with the developments in other fields, natural product studies will play increasingly important roles in new drug discovery and will be an integrated component in the battle against old and new human diseases.

Reference

1. Blunt, J. W.; Copp, B. R.; Hu, W. P.; Munro, M. H.; Northcote, P. T.; Prinsep, M. R., *Nat Prod Rep* **2009,** 26, (2), 170-244.
2. Carey, J. S.; Laffan, D.; Thomson, C.; Williams, M. T., *Org Biomol Chem* **2006,** 4, (12), 2337-47.
3. Patel, R. N., *Curr Opin Drug Discov Devel* **2006,** 9, (6), 741-64.
4. Pollard, D. J.; Woodley, J. M., *Trends Biotechnol* **2007,** 25, (2), 66-73.
5. Turner, N. J., *Nat Chem Biol* **2009,** 5, (8), 567-73.
6. Williams, R. M.; Cox, R. J., *Acc Chem Res* **2003,** 36, (2), 127-39.
7. Fox, R. J.; Clay, M. D., *Trends Biotechnol* **2009,** 27, (3), 137-40.
8. Zhang, H.; Wang, Y.; Pfeifer, B. A., *Mol Pharm* **2008,** 5, (2), 212-25.
9. Lee, S. Y.; Kim, H. U.; Park, J. H.; Park, J. M.; Kim, T. Y., *Drug Discov Today* **2009,** 14, (1-2), 78-88.