Structural and Biochemical Characterization of the Early Steps in Fungal Indole Alkaloid Biosynthesis

Qingyun Dan

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

Doctoral Committee:

Professor Janet L. Smith, Chair
Emeritus Professor David P. Ballou
Professor David H. Sherman
Professor John J.G. Tesmer
Associate Professor Zhaohui Xu

Dedication

To my parents, my grandpa Yanzhong Yu,

and my beloved wife Qiong.

Acknowledgements

During my Ph.D. study, a lot of people have helped, encouraged and supported me. Here I would like to take this opportunity to express my most sincere gratitude to them. First and most importantly, I would like to thank my mentor, Professor Janet L. Smith, for her endless support and guidance. Her enthusiasm for science has greatly inspired me and her selfless dedication strongly firmed my determination to pursue a scientific career. Janet is a great scientist and mentor to me, and I feel delighted and grateful to have had the chance to work with and learn from her.

I would also like to thank professor David H. Sherman, for allowing me to collaborate with his laboratory on this highly cooperative, exciting and challenging project, and providing lots of useful suggestions. Four collaborators in the Smith group and the Sherman group, Dr. Shengying Li, Dr. Hong T. Tran, Dr. Sean A. Newmister and Amy E. Fraley have been very helpful to me. Shengying started the fungal indole alkaloid project. When I started my Ph.D. study, I was inspired by his enthusiasm for the project and finally decided to join him. After Shengying left for China to start his own laboratory, Hong and Sean have been collaborating with me closely during the last three years, and nothing could have been achieved without them. Amy is a very talented graduate student in our lab. She has made a significant breakthrough on the project and helped a lot on thinking towards the big picture of the pathways that we study.

Moreover, many previous and current members in the Smith lab have helped me a lot, and I am grateful for all their help. Dr. Steffen M. Bernard and Dr. Dheeraj Khare helped me to settle in when I first came to Ann Arbor. Dr. David L. Akey, Dr. Donald D. Raymond, Dr. Jonathan R. Whicher and Dr. Jennifer G. McCarthy taught me a lot on X-ray crystallography and numerous other scientific matters. I would also like to thank Meredith A. Skiba, whom I think is the most indispensable current member in our lab, for making sure that everything is in the right order.

I have met many old and new friends in Ann Arbor, and their friendship is one of the most valuable treasure that I will cherish. Hongyang Li, Bing Yang, Jiyuan Yang, Yanxiao Zhang, Shuai Niu, Zhiyuan Yao, Long Chen, Andrew P. Sikkema, Eric Tse, and many others that I do not have the space to list here, I would like to thank them all for making Ann Arbor my second hometown.

Finally, I would like to thank my family, especially my parents, my grandparents and my wife for their support regardless of return. From the first day beginning, my parents have always given me the uttermost freedom to chase whatever I would like to do, and have supported and trusted me all the time. My grandpa, Yanzhong Yu, was a very talented student in school. However, he could not continue study due to political reasons and poverty back then. He brought me up, introduced science to me with interesting stories of famous scientists, and taught me that dedication and hardworking are necessities for success. I could not have been who I am without his tutoring and care.

I first met my wife, Qiong, when I was in Peking University, and was immediately attracted by her vivacity and optimism. We started to date eight years ago, and have

been together ever since. Despite all the ups and downs, our love only got strengthened and we got married last summer. We have already spent one thirds of our life together, and the percentage is only going to be higher and higher. From the bottom of my heart, I would like to thank you for your endless love and support, and for helping take care of our two lovely cats, Ciel and Crystal.

Table of Contents

List of Figures

List of Tables

Abstract

Prenylated indole alkaloids are a class of natural products with great structural diversity and pharmaceutical potential. These alkaloids, which are isolated from various fungi, mostly *Aspergillus* and *Penicillium*, are often produced by homologous dimodular non-ribosomal peptide synthetase (NRPS) pathways that combine two amino acids, typically tryptophan, proline, histidine or phenylalanine, to form the alkaloid skeleton. A unique bicyclo[2.2.2]diazaoctane group is a distinctive feature of the NRPS pathway of malbrancheamide, a calmodulin inhibitor produced by *Malbranchea aurantiaca*. The bicyclo[2.2.2]diazaoctane is proposed to form via an intramolecular Diels-Alder reaction, but the protein that ensures stereospecificity of the reaction is unknown. This thesis describes research focused on the structural and biochemical characterization of the early steps in the malbrancheamide biosynthetic pathway, which precedes the proposed Diels-Alder reaction. In collaborative studies, I solved the first crystal structure of a fungal NRPS terminal reductase domain, PhqB R in the homologous paraherquamide pathway, which indicates that it functions as a 2-electron or 4-electron reductase. I also solved a 1.6 Å crystal structure of MalC, a candidate for re-oxidation of a potential 4-electron reduction product. However, the MalC structure strongly indicates that it cannot catalyze a redox reaction and its function remains to be characterized. Furthermore, MalB and MalE, two

prenyltransferases in the pathway, were characterized in detail. In summary, the dissertation research provides the first structural and biochemical insights into the early steps of malbrancheamide biosynthesis and will guide protein engineering and chemoenzymatic synthesis of related compounds in the future.

Chapter I

Introduction

1.1 Natural Products

Natural products are secondary metabolites produced by microbes and plants [1-4]. With complex and diversified structures, they are not essential for survival of the producing organisms in absence of environmental stress. However, many natural products have potent biological activities, which are often utilized by the producing organisms to their advantage, for example by suppressing the growth of pathogenic and rival organisms [2]. Many natural products and their derivatives have been identified and further developed as antibiotics, anti-fungal agents, immuno-suppressants and anti-cancer drugs [3-5].

There is a long history of using plants and herbs that produce natural products as medicine, toxins, etc. [6] *Artemisia annua*, an annual ephemeral plant, was used in traditional Chinese medicine for treatments of fever and malaria, a fatal mosquito-borne infectious disease, over 1500 years ago [7] (Fig. 1.1.A). In Europe willow bark extract was the major medication for pain relief and fever treatment for a long period [8] (Fig. 1.1.B). However, it was unknown how and why these substances were effective, due to lack of experimental methods for small molecule isolation and identification. We now know that salicin, a natural product produced by willow bark,

can be enzymatically hydrolyzed to salicylic alcohol, which is further oxidized to salicylic acid (aspirin), the active ingredient [8]. As for *Artemisia annua*, the compound that kills the malaria protozoa is artemisinin (qinghaosu) [7]. With development of modern science, we now understand that not only plants, but also bacteria and fungi are abundant sources of natural products with great pharmaceutical potential. For example, penicillins, which were initially discovered and isolated from *Penicillium notatum* by Fleming, Florey and other researchers [9-10], are a group of widely used natural product drugs with potent anti-bacterial activities.

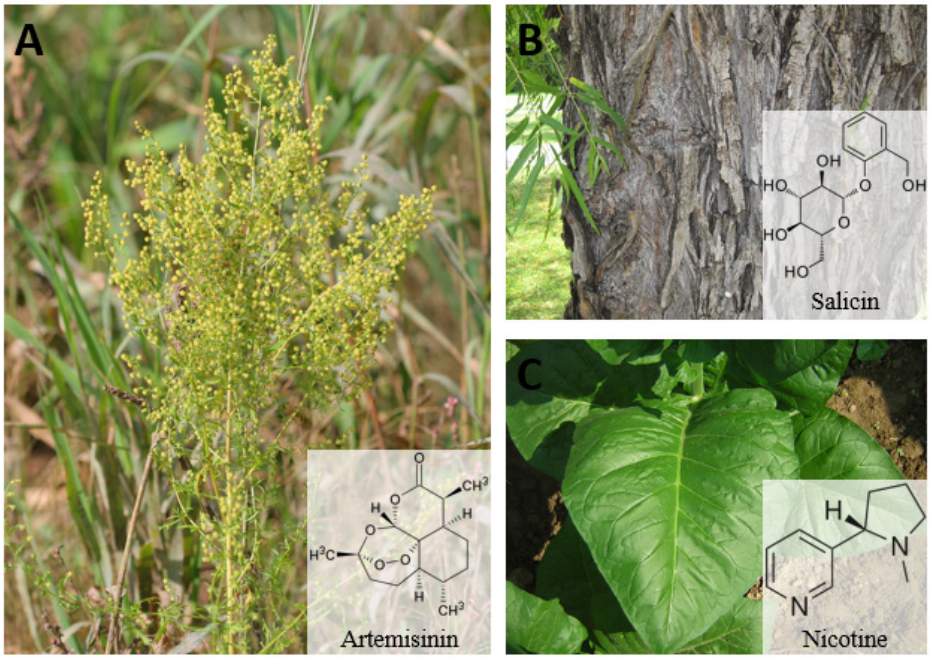


Figure 1.2. Examples of plant natural products and their producing organisms. A. *Artemisia annua*, an annual ephemeral plant for malaria treatment, and its biologically active compound artemisinin. B. Willow bark produces anti-inflammatory salicin. C. Structure of the nicotine alkaloid and its natural source tobacco leaves.

Studies of natural products are crucial for many reasons. First, natural products are now the major source for developing new drugs. About 50% of drugs approved by the

U.S. Food and Drug Administration (FDA) in the most recent 25 years are either natural products or their derivatives/mimics [11] (Fig. 1.2). Studies of new natural products will greatly benefit the discovery and development of novel drugs. Many natural products have also led to the discovery and characterization of important regulatory systems throughout life, for example, the mechanistic target of rapamycin (mTOR) [12-14]. Many others can function as biolabels and allow scientists to target macromolecules and pathways of interest.

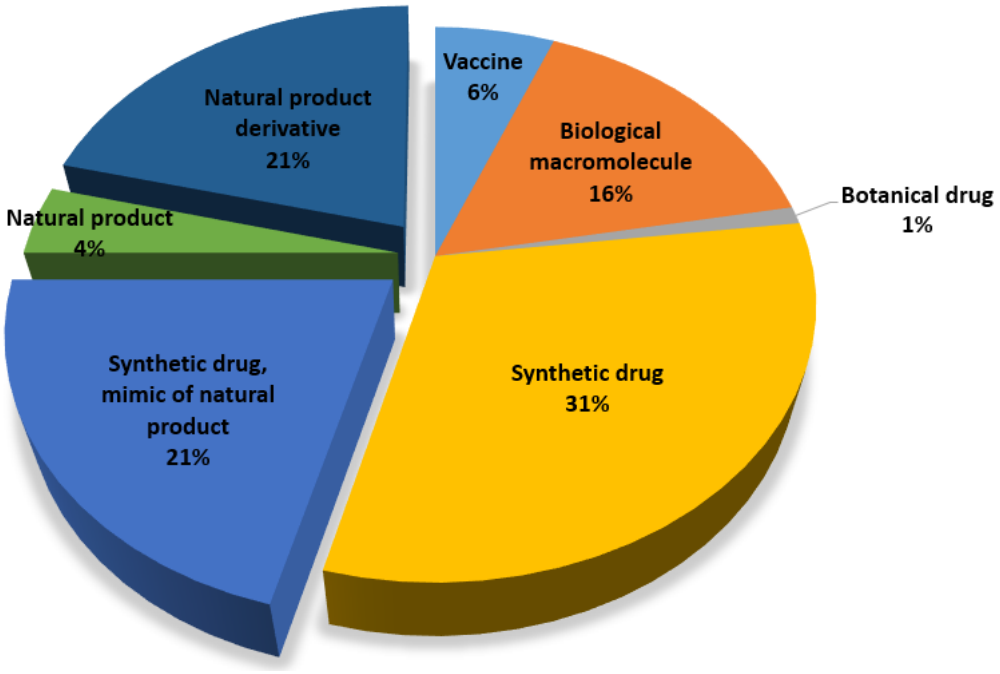


Figure 1.2. Source of FDA-approved drugs in 1981 – 2014 [11]. Natural products, mimics and derivatives contribute to 46% of the total, being the major source.

Direct isolation or extraction from producing organisms was the major method to obtain natural product. However, this can be difficult and inefficient, with complicated processes and extremely low yields. For example, bryostatin 1, a biologically potent marine macrolide, can be extracted from marine-derived *Bugula*

neritina in only about 1.3×10^{-6} % yield [15-16]. Furthermore, many bacteria and fungi that produce the target compounds cannot be cultured on a large scale under laboratory conditions. For plants with a long life span, for example, *Taxus brevifolia* (the pacific yew), the producer of the anti-cancer paclitaxel, a shortage of the natural resource is also a major limitation for large-scale production. In contrast, for many years, chemical synthesis (total synthesis and semi synthesis) has been applied with great success to the production of desired compounds at a lower cost than extraction from the natural source. Penicillin V, for example, can be chemically synthesized from penicilloic acids in five steps [17]. In fact, natural products have been a driving force for the development of modern organic chemistry and synthetic chemistry by providing intriguing chemistry, challenging targets, as well as novel synthetic schemes. One example is the developing field of biomimetic synthesis, which targets synthesis of natural products through biosynthetically related intermediate structures [18].

Despite rapid development of synthetic chemistry, there are still limitations for generating desired natural products by chemical synthesis. To date many compounds cannot be chemically synthesized in a productive way. Synthesis of others requires multiple steps and/or extreme conditions for catalysis. For example, total synthesis of bryostatins takes over 40 steps [19]. Furthermore, a single modification to the target molecule may disable the original synthetic scheme, in which case an entirely different scheme must be investigated. Also many natural products have stereo centers and various tailoring groups, which are often essential for biological activity.

Structural complication generally increases the difficulty of developing novel natural product total synthetic schemes. Studies of natural product biosynthetic pathways provide an alternative approach: chemoenzymatic synthesis.

1.2 Natural Product Biosynthetic Pathways

All natural products are produced by enzymes of secondary metabolic pathways in the producing organisms. In microbes the biosynthetic enzymes are often encoded in gene clusters, which can be identified by whole genome sequencing and genomic data mining. In addition, many enzymes involved in natural product biosynthesis catalyze unique chemistry with uncommon mechanisms. Identification of genes and, biochemical and structural characterization of the encoded enzymes can greatly expand the limits of enzymology, as well as guide protein engineering and chemoenzymatic synthesis of novel compounds.

Many interesting secondary metabolic pathways are modular, and can be further classified as type I polyketide synthase (PKS) pathways and non-ribosomal peptide synthetase (NRPS) pathways [20-21]. Type I PKS pathways are homologous with the metazoan type I fatty acid synthase (FAS) pathway, an essential primary metabolite pathway for *de novo* synthesis of fatty acids. PKS pathways are composed of sequential modules, each containing a set of elongation and modification domains. The substrate or intermediate is attached to an acyl carrier protein (ACP) domain via a 4′-phosphopantetheine linker (Ppant arm) [22-24]. Each module accepts the intermediate from the ACP of its upstream module, elongates the chain by two carbon

atoms, catalyzes certain modification reactions, and passes the product to the ACP of the downstream module. When the small molecule cargo reaches the last (terminal) module, the thioester bond between the intermediate and the Ppant arm is cleaved and the cargo is released. In some cases, the offloaded product goes through another series of modifications catalyzed by other enzymes in the pathway to generate the final natural product.

Non-ribosomal peptide synthetase (NRPS) pathways are enzyme complexes utilizing amino acids to generate dipeptides or polypeptides. Although evolutionarily not related to PKS pathways, NRPSs also use peptidyl carrier protein (PCP, also named as a thiolation or T domain), a similar carrier system to load and transport cargo [25-26]. The substrate (amino acid) is loaded onto the starting module, passed to the downstream modules through cycles of elongation and modification dependent on the domain composition of each module, and offloaded in the terminal module to generate the product.

These assembly-line like modular systems have drawn great interest from researchers since they were identified, due to the fact that each domain catalyzes only one reaction in the biosynthetic scheme, with the reaction order in the assembly-line generally following the domain order. This provides a hope for building an efficient chemoenzymatic synthesis platform, where synthesis of a target molecule can be achieved by simple genetic manipulation to swap, delete, insert and/or modify target domains. Numerous studies have been done to understand the reaction mechanisms, substrate specificities and domain-domain interactions of these modular biosynthetic

pathways [22-26]. In favorable cases, researchers are now capable of engineering a natural pathway to produce desired novel compounds. However, to date the efficiency is still far below an industrially applicable level, with very low yields (μg - mg level) and low percent conversion of starting materials (< 1%) [27]. Identifying and studying new modular secondary metabolite pathways, may help to add promising novel "enzymatic toolboxes" to our current synthetic schemes, as well as improve our understanding of modular assembly-line systems.

1.3 Prenylated fungal indole alkaloids

Prenylated indole alkaloids are a class of natural products isolated from different fungi, mostly *Aspergillus* and *Penicillium* [28]. They are often produced by homologous dimodular NRPS pathways, utilizing proline, histidine, phenylalanine or tryptophan as substrates [29]. Prenylated indole alkaloids are unique for several reasons. First, many of them have complex ring systems and stereocenters. Stephacidin B, for example, contains two bicyclo[2.2.2]diazaoctane rings, a nitrone, a hydroxyindole, and nine stereocenters, with fifteen rings in total [30] (Fig. 1.3). The unique bicyclo[2.2.2]diazaoctane ring system, a structural feature common to this type of prenylated indole alkaloid, was not reported previously.

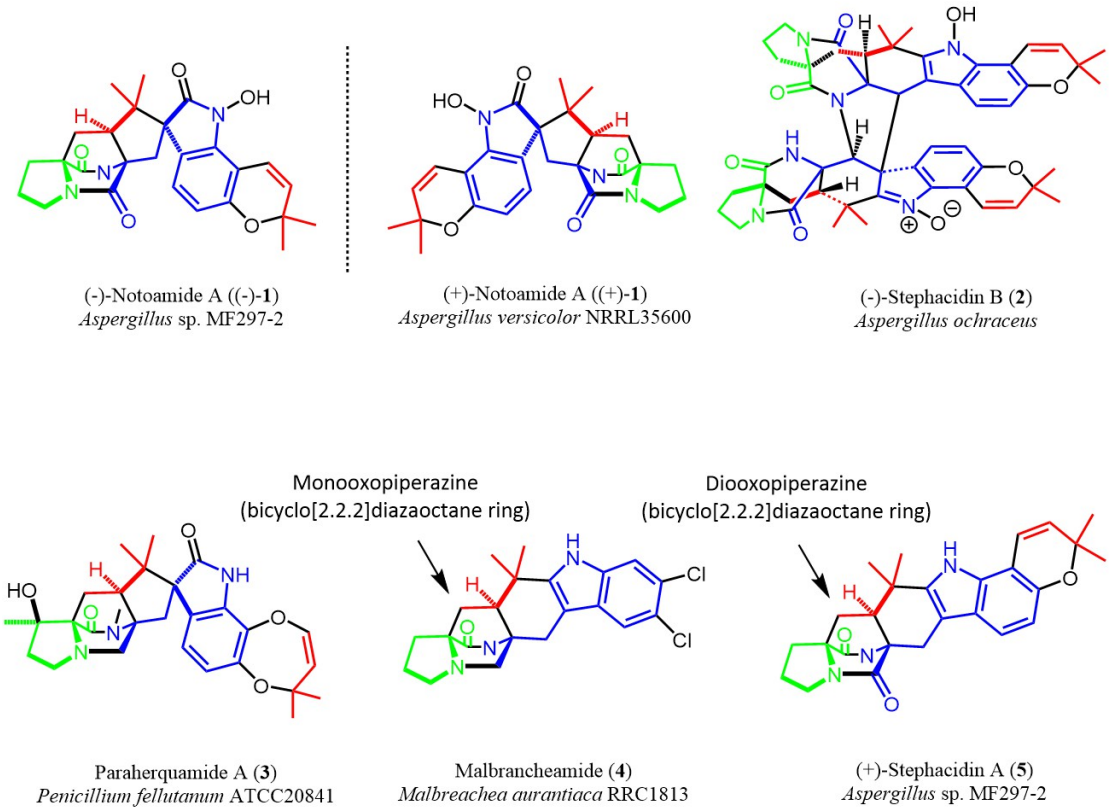


Figure 1.3. Examples of prenylated fungal indole alkaloids and their producing organisms. The bicyclo[2.2.2]diazaoctane ring is a common feature, and total enantiomers ((-)-**1**, (+)-**1**) have been discovered in different producing organisms. The natural products are colored according to the starting materials for biosynthesis (blue L-Trp, green L-Pro, red isoprene).

Prenylation of aromatics/indoles is another key feature of this group of secondary metabolites. Aromatic prenyltransferases exist widely in different species of bacteria and fungi, catalyzing transfer of isoprenyl moieties from donors such as dimethylallyl diphosphate (DMAPP) and geranyl diphosphate (GPP) to specific aromatic acceptors, tryptophan for example [31]. Prenylation is a primary source of the structural complexity of prenylated indole alkaloids [32]. For fungal indole alkaloids, DMAPP is usually the prenyl donor, and the added prenyl group is often incorporated into specific rings in later steps.

Most previously identified aromatic prenyltransferases have broad substrate selectivity. For instance, CdpNPT from *Aspergillus fumigatus* can catalyze reverse prenylation of (*S*)-benzodiazepinedione, (*R*)-benzodiazepinedione, several cyclic peptides, and tryptophan [33]. This substrate promiscuity can be explained by the nature of the substrate binding pockets [34-36], which include a largely hydrophobic area accommodating the substrate primarily via nonspecific hydrophobic interactions. However, several fungal indole alkaloid pathways seem to be exceptions. Two fungal aromatic prenyltransferases, NotC and NotF, have been characterized from the (-)-notoamide pathway in *Aspergillus* sp. MF297-2, and show narrow substrate selectivity and prenylation site specificity [37]. Enzymatic activity tests of NotF with different indole compounds show that its natural substrate (brevianamide F) is the only compound that triggers catalysis. Whether the strict substrate selectivity is characteristic of these fungal pathways requires further investigation.

In recent decades, more prenylated fungal indole alkaloids have been isolated and identified, for example, the brevianamides, aspergamides, sclerotamides, macfortines, stephacidins, notoamides, paraherquamides and malbrancheamides [38]. They share similar core structural features, while having different tailoring groups and drastically different biological activities, ranging from anti-cancer cytotoxicity to calmodulin inhibition. Studies of this group of compounds and their biosynthetic pathways may reveal biochemical and structural basis for their peculiarity, providing innovative ideas and thoughts for synthesis of similar complicated ring systems.

1.4 Notoamide, paraherquamide and malbrancheamide pathways

Notoamides are a group of prenylated indole alkaloids isolated from the marine-derived fungus *Aspergillus* sp. MF297-2. Notoamide A-C are anticancer agents, showing cytotoxicity to HeLa and L1210 cell lines, with moderate IC_{50} values (22 - 52 μg/mL) [39]. Very interestingly, enantiomers of notoamides have been found in different *Aspergillus* strains [40] (Fig. 1.3). In *Aspergillus* sp. MF297-2, (-)-notoamide A is produced as the final pathway product. However, in *Aspergillus versicolor*, only (+)-notoamide A has been isolated. One or several enantiodivergent steps are proposed to be present in the biosynthetic pathways to produce either set of the enantiomers.

Paraherquamides are isolated from *Penicillium paraherquei* in 1981 [41]. Paraherquamide has been shown to possess anthelmintic activity, with 91% efficacy against *Strongyloides stercoralis*, a common gastrointestinal nematode of dogs [42]. In calves the activity is more robust, with >95% efficacy against 8 different species of nematodes: *Haemonchus placei, Ostertagia ostertagi, Trichostrongylus axei, Trichostrongylus colubriformis, Cooperia oncophora, Nematodirus helvetianus, Oesophagostomum radiatum*, and *Dictyocaulus viviparous* [43].

Malbrancheamide is a novel phytotoxic alkaloid isolated from marine-derived *Malbranchea aurantiaca* in 2006 [44]. It was shown to inhibit radical growth of *Amaranthus hypochondriacus* using a Petri dish bioassay. Further kinetic studies with the calmodulin (CaM)-sensitive phosphodiesterase PDE-1 showed that

malbrancheamide is a moderately strong competitive inhibitor of calmodulin (K_i = 47.4±5.63 μM).

Through whole genome sequencing, genome mining and deep gene annotation, David Sherman (University of Michigan) identified the gene clusters for (-)-notoamide A in *Aspergillus* sp. MF297-2, (+)-notoamide A in *Aspergillus versicolor* NRRL35600, paraherquamide A in *Penicillium fellutanum* ATCC20841 and malbrancheamide in *Malbranchea aurantiaca* RRC1813 [40] (Fig. 1.4).

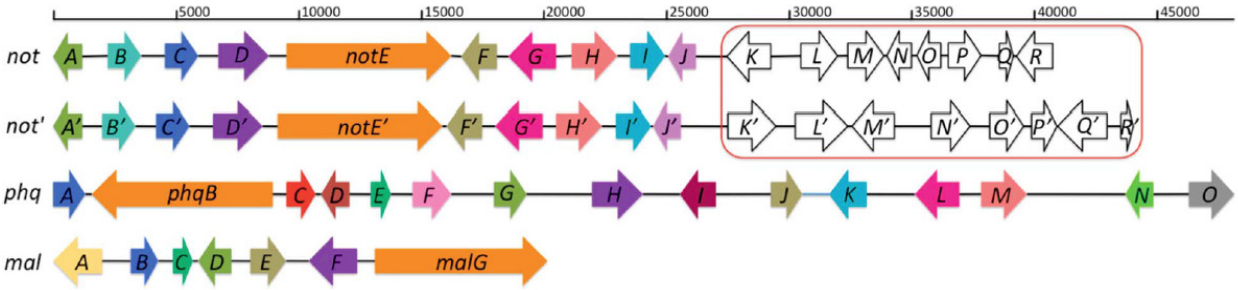


Figure 1.4. Gene cluster scheme for (-)/(+)-notoamide (not/not'), paraherquamide (phq) and malbrancheamide (mal) pathways [40]. Comparative analysis suggests a common mechanism for producing precursors. Colors indicate common predicted functions for the encoded enzymes in these pathways.

Comparative analysis indicates that these pathways are homologous, based on high sequence similarities and close relations of core genes in the pathways. Successful total synthesis and biomimetic synthesis for several prenylated indole alkaloids, pioneered by Robert Williams (Colorado State University), for example stephacidins A and B, notoamide B-D [45-49], allowed isotope-labeled feeding studies that significantly improved the understanding of these pathways by identifying common precursors and upstream-downstream relationships of labeled molecules [50-52].

Combined with BLAST-based function prediction [68] for each gene product, a scheme was proposed to explain how the pathways synthesize their final products [40].

For the (-)-notoamide pathway, the early steps have been established by the Sherman group [37, 50, 53] (Fig. 1.5). The first step is an NRPS-catalyzed reaction in which the NotE NRPS dimodule takes L-Pro and L-Trp as substrates and produces brenvianamide F, a common precursor of notoamides and stephacidins. Brenvianamide F is then prenylated by NotF, a prenyltransferase, to generate deoxybrevianamide E, which can be further oxidized to 6-OH-deoxybrevianamide E. NotC, the other prenyltransferase in the pathway, prenylates 6-OH-deoxybrevianamide E to form notoamide S.

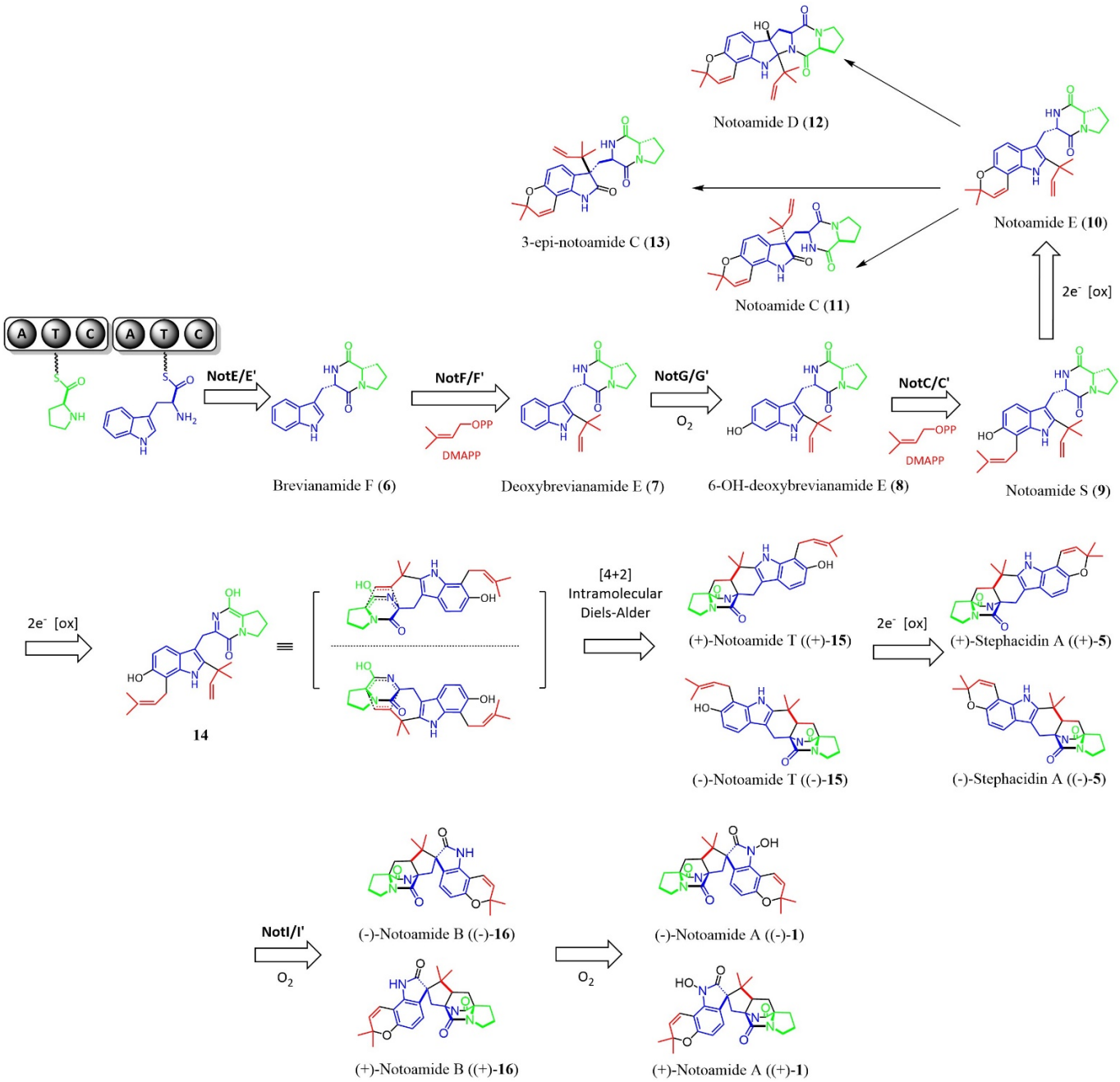


Figure 1.5. Proposed scheme for (-)/(+)-notoamide biosynthesis [40].

From notoamide S, the pathway branches. NotD, an FAD-dependent oxidase, catalyzes the pyran ring closure on notoamide S to generate notoamide E. Notoamide T, a dioxopiperazine, is also generated from notoamide S, and the bicyclo[2.2.2]diazaotane ring of notoamide T is proposed to arise from a putative intramolecular Diels-Alder cycloaddition reaction (IMDA). D,L-[^{13}C]$_2$- notoamide T

was synthesized chemically and fed to both *Aspergillus* sp. MF297-2 and *Aspergillus versicolor*, the strains that produce the notoamide enantiomers [54]. In *Aspergillus versicolor*, ^{13}C labeled (+)-stephacidin A, a non-endogenous natural metabolite, was isolated, as well as the endogenous natural metabolite (+)-notoamide B. This suggests that notoamide T is the precursor for stephacidin A, and (-)-stephacidin A is converted in *A. versicolor* to (+)-notoamide B by a stereo-selective enzyme, probably via an oxidative pinacol ring rearrangement. This enzyme was later revealed to be a FAD-dependent monooxygenase NotI' (personal communication from Hong Tran in the Sherman group). In *Aspergillus* sp. MF297-2, both (+), (-)-$[^{13}C]_2$-stephacidin A and (-), (+)-$[^{13}C]_2$-notoamide B were detected, indicating the same reaction scheme, but no evident stereoselectivity was observed. It is still unknown how many step(s) in both (+)- and (-)-notoamide pathways provide strict stereocontrol, and which enzyme is the (first) determining factor.

Compared to the (+)-/(-)-notoamide pathways, the paraherquamide and malbrancheamide pathways have one major difference: instead of forming dioxopiperazines, all identified products are monooxopiperazines (Fig. 1.3). The reductive state difference is proposed to arise from different domain compositions of the starting NRPS [40]. In the notoamide pathway, the terminal module of the NRPS (NotE) is a condensation domain, while in the other two pathways, a terminal reductase domain is proposed to reduce the dipeptide substrate to an aldehyde or alcohol.

Another important difference among these homologous pathways is that, compared to the notoamide and paraherquamide pathways, which both produce more than eight different products, the malbrancheamide pathway is much simpler. Only four natural products, premalbrancheamide, malbrancheamide B, isomalbrancheamide B and malbrancheamide, were isolated from *Malbranchea aurantiaca*, and they differ only in the number of halides within the structures. The gene cluster encodes seven proteins in total. Based on knowledge of the homologous notoamide and paraherquamide pathways, a 4-step reaction scheme (Fig. 1.6) is proposed [40]: MalG, an NRPS, utilizes L-Pro and L-Trp to produce L-Pro-L-Trp aldehyde. A prenyltransfer reaction then leads to a prenylated dipeptidyl aldehyde, which would go through a hypothetical Diels-Alder reaction to form premalbrancheamide, the monooxopiperazine with a bicyclo[2.2.2]diazaoctane ring. Premalbrancheamide and its analogs have been isolated from extracts of several different fungi, suggesting that its formation may be a common mechanism for biosynthesis of many fungal indole alkaloids. The final step is halogenation to install two chlorine atoms on the indole, producing malbrancheamide as the final product of the pathway.

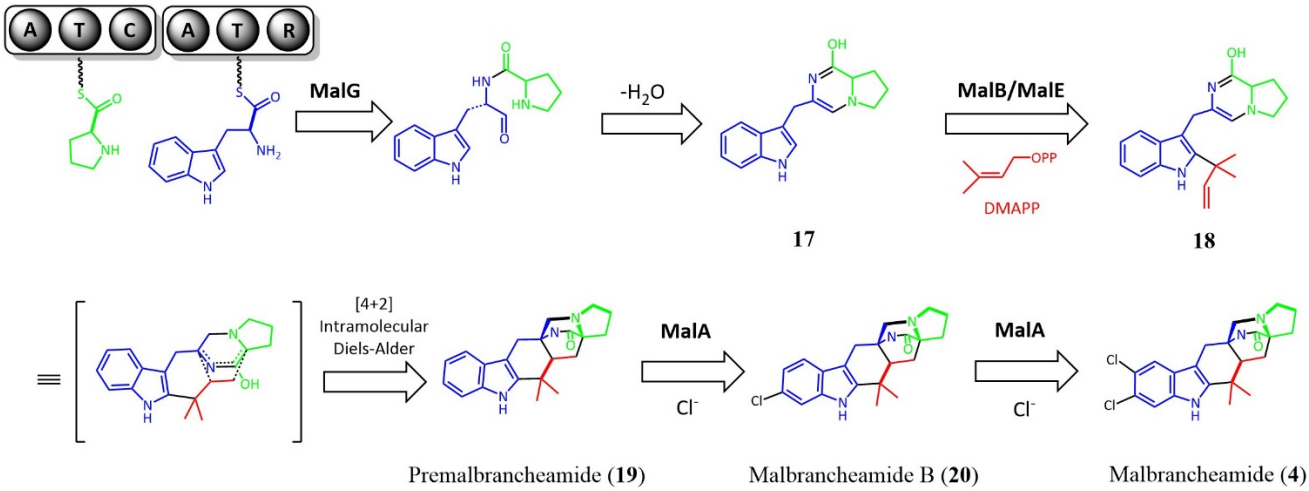


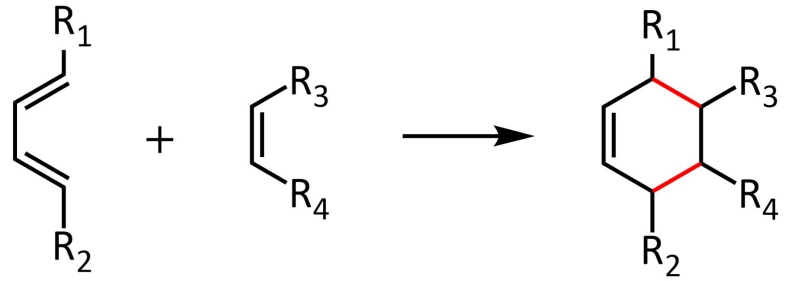
Figure 1.6. Proposed scheme for malbrancheamide biosynthesis [40].

To date, the biggest question for fungal indole alkaloid biosynthesis has not been fully addressed, which is how the unique bicyclo[2.2.2]diazaoctane ring is synthesized enzymatically. The diazaoctane ring was proposed to arise from a Diels-Alder reaction, yet no direct evidence has been reported. As the shortest pathway in the family, the malbrancheamide pathway provides a great platform to answer this question.

1.5 Intramolecular [4+2] Diels-Alder Reactions

The Diels-Alder reaction (D-A) is a [4+2] cycloaddition reaction between a conjugated diene and dienophile that forms up to four stereocenters in one step (Fig. 1.7). Depending on whether the diene and the dienophile come from the same molecule/substrate, Diels-Alder reactions are classified as intermolecular or intramolecular. As a textbook reaction in organic chemistry and synthetic chemistry, the Diels-Alder reaction is one of the most important ring forming reactions, with

broad application in total synthesis of numerous compounds [55-57]. When addressing complicated synthetic puzzles, most of which are delivered by natural products, Diels-Alder reactions sometimes offer solutions and shortcuts that no other reaction scheme can provide. One major reason is that Diels-Alder reactions are energetically favorable, and many bimolecular Diels-Alder reactions proceed at detectable rates without a catalyst, providing high efficiency.



Diene Dienophile

Figure 1.7. Diels-Alder reaction. The newly formed bonds are colored in red.

In nature, different groups of Diels-Alder product-like structures have been isolated and identified, for example, the bicyclo[2.2.2]diazaoctane ring. Whether a "Diels-Alderase", an enzyme catalyzing Diels-Alder reactions, exists has been debated over decades [58-61]. The chemistry of the reaction is relatively simple and well understood. Frontier molecular orbital theory [55] predicts that improvement of the reaction rate can be achieved simply by narrowing the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), suggesting the possibility that such enzymes may exist. More interestingly, without a catalyst, a Diels-Alder reaction can occur in two ways to generate a mixture

of enantiomers. However, all proposed natural Diels-Alder products possess only one configuration, indicating involvement of at least one enzyme to provide stereo control.

Before any Diels-Alderase was identified in nature, the Baker group computationally designed a group of unnatural bimolecular Diels-Alderases with high stereospecificity and substrate selectivity [62]. Using the ROSETTA computational design methodology [69], they successfully converted diisopropylfluorophosphatase and ketosteroid isomerase scaffolds into Diels-Alder catalysts, with fewer than fifteen amino acid substitutions in each case. Further modification by site-directed mutagenesis gave rise to a Diels-Alder catalyst (DA_20_10) possessing both strict stereoselectivity and substrate specificity as predicted.

Known enzymes that solely catalyze Diels-Alder reactions are very rare, although catalytic antibodies with Diels-Alderase activity have been generated [137]. Fungal macrophomate synthase (MPS), a Mg^{2+}-dependent enzyme that synthesizes benzoate from 2-pyrone and oxalacetate in a five-step reaction, was proposed to catalyze an inverse-electron demand Diels-Alder reaction in the second step [58] (Fig. 1.8). The crystal structure of MPS in complex with pyruvate was reported by the Tanaka group in 2003 [59]. However, MPS was subsequently shown to be a Michael-aldolase but not a Diels-Alderase [60]. Using mixed quantum and molecular mechanics (QM/MM) with Monte Carlo simulations and free-energy perturbation (FEP) calculations, the transition state (TS) energy of the reaction was calculated, either through a concerted Diels-Alder reaction model or a stepwise Michael-aldol reaction model. The

computational calculation showed that the TS energy for the concerted model was at least 20 kcal/mol higher than that of the Michael-aldol model, indicating that the reaction is not catalyzed by a concerted mechanism, which is a defining characteristic of Diels-Alder reaction.

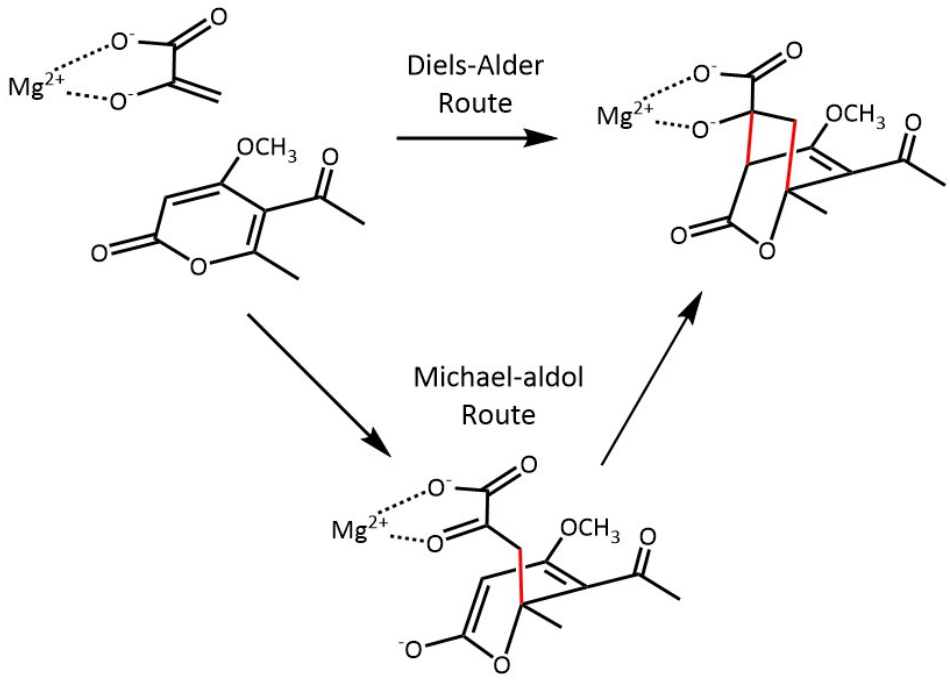


Figure 1.8. Proposed reaction pathway for the second step of fungal macrophomate synthase catalysis: Diels-Alder route and Michael-aldol route [60]. Newly formed bonds are colored in red.

Despite the many ongoing debates, the first confirmed intramolecular Diels-Alder enzyme, SpnF, was found in the spinosyn A biosynthetic pathway [63]. It catalyzes an intramolecular [4+2] cycloaddition reaction from an alkenyl to a dienyl group, increasing the reaction rate by 500-fold. A high-resolution crystal structure of SpnF was reported in 2015 [64] (Fig. 1.9.A). Bearing an *S*-adenosylmethionine (SAM)-dependent methyltransferase (MT) fold, the structure resembles Class I SAM-

dependent MTs. It is unknown whether SAM has a primarily structural role, although no single point mutation was identified that totally abolished cyclization activity. No catalytic residue has been identified, suggesting that the primary role of SpnF may be as an entropy trap, holding the substrate in an optimal conformation so that the 1,3-diene and the alkene group are in proximity and positioned for the [4+2] cycloaddition reaction.

In 2015, another group of homologous Diels-Alder enzymes was reported to catalyze stereoselective IMDA reactions to generate spirotetronate natural products [65]. The crystal structure of one member, PyrI4, in complex with its product, was reported in 2016 [66] (Fig. 1.9.B). With no sequence identity to SpnF, PyrI4 is a homodimer, with each chain forming an antiparallel β barrel, and is structurally unrelated to SpnF. The product sits at the bottom of the β barrel pocket, and an initially disordered N-terminal region packs into a highly bent α helix (α0), closing the active site pocket. α0 is essential for enzymatic activity, as deletion led to the loss of Diels-Alder activity.

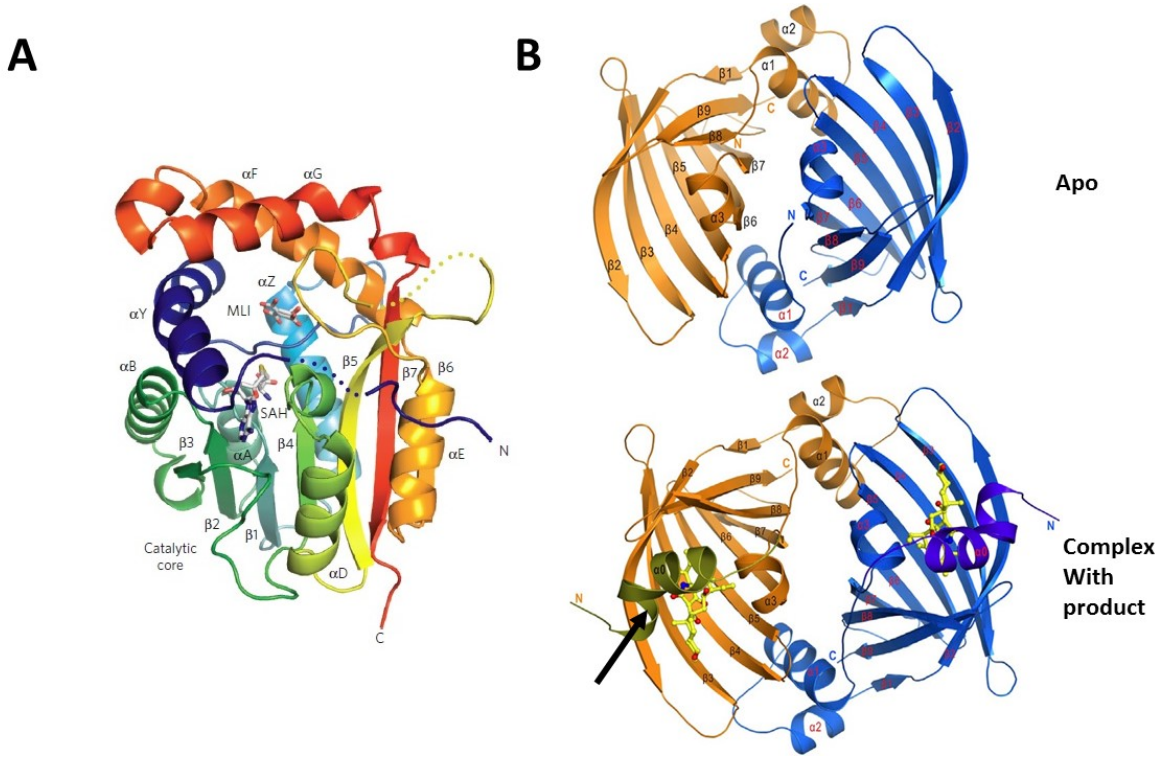


Figure 1.9. Structures of two Diels-Alder enzymes: SpnF (A) and PyrI4 (B) (figure from references [64] and [66]). These two enzymes are structurally unrelated. No catalytic residues were identified, suggesting an entropy trapping mechanism.

In summary, all evidence suggest that Diels-Alder enzymes do exist in nature, utilizing an entropy trapping mechanism to achieve catalysis of a concerted [4+2] cycloaddition reaction [63-67]. High-resolution crystal structures have also provided insights into how nature evolved the enzymes, which intriguingly agrees well with the ROSETTA *de novo* design methodology [69]. It appears that in both cases, random scaffolds suitable for accommodating the substrates are selected and evolved towards lowering the transition state energy. Further research may provide additional details of this unique type of "catalytic" reaction, as well as lead to further optimization of the current computational protein design scheme. No enzyme in the malbrancheamide pathway shows similarity to either group of identified Diels-Alder enzymes. The malbrancheamide pathway provides a perfect object for novel Diels-Alder enzyme

investigation, and is the focus of my dissertation research.

Chapter II

MalG: A dimodular NRPS with a terminal reductase domain

2.1 Introduction

Non-ribosomal peptide synthetases (NRPSs) are multi-domain enzymes that use amino acids to produce various peptidyl secondary metabolites such as cyclosporin and tyrocidine [4, 5]. Although NRPSs form the same kind of bonds (peptide bonds) as ribosomes, their substrate recognition and loading, intermediate transfer and product release processes are largely different [70-72]. Because they can use modified or unnatural amino acids [73], NRPSs generate peptidyl natural products of great structural diversity that ribosomes can never achieve. Moreover, NRPSs contain different sets of domains that can further modify peptides by methylation, acylation, ring cyclization, etc. NRPS and hybrid polyketide synthase (PKS) – NRPS pathways exist in various types of bacteria and fungi, many NRPS products have potent biological activities, and some have been widely used as antibiotics, anti-fungal agents, immuno-suppressants and anti-cancer drugs [74-78] (Fig. 2.1). Well-known examples include aminoadipoyl-cysteinyl-valine (ACV) synthetase, which produces the ACV tripeptide, a precursor of penicillins and cephalosporins [74], and cyclosporin [75] from the fungus *Tolypocladium inflatum*, which produces the immuno-suppressant drug cyclosporine.

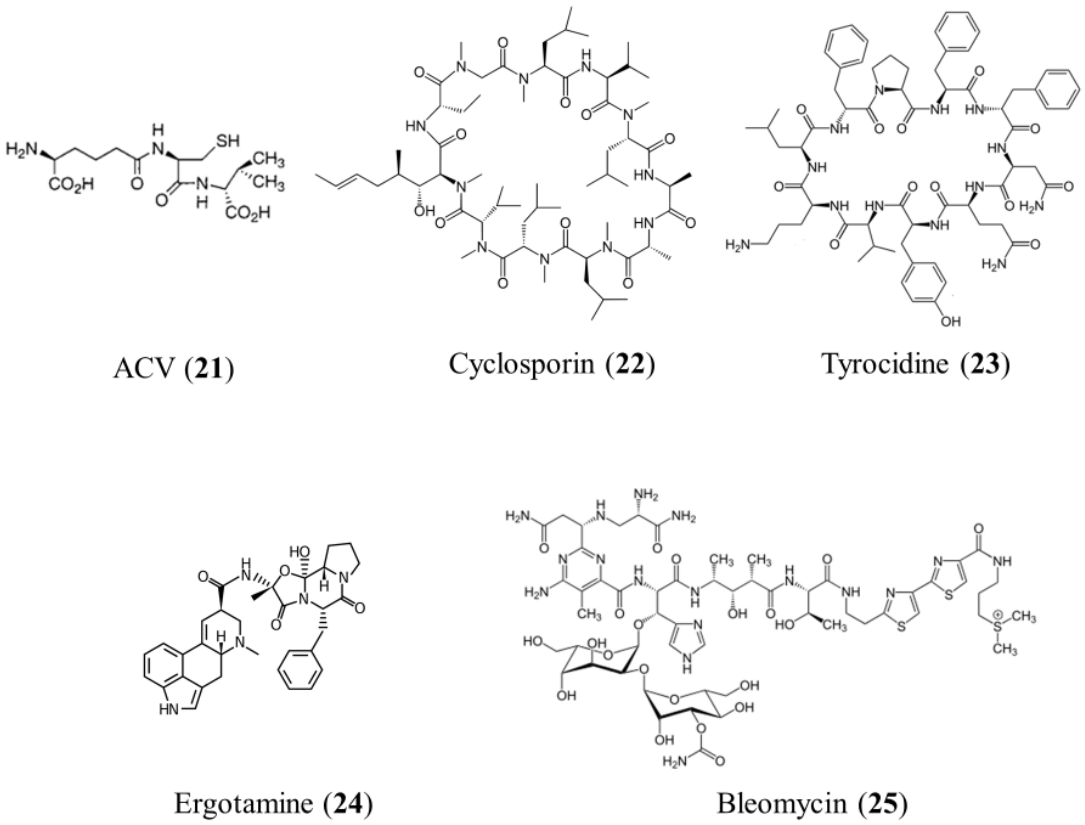


Figure 2.1. Examples of NRPS natural products. ACV is an intermediate of penicillin biosynthesis; cyclosporin is a widely-used immuno-suppressant; tyrocidine was the first commercially available antibiotic; ergotamine was an anti-migraine drug; bleomycin is an anti-cancer drug.

Working as an assembly line, NRPSs are often modular, with a minimal module composed of an adenylation (A) domain, a thiolation (T or PCP) domain and a condensation (C) domain [79-83] (Fig. 2.2). The A domain selectively recognizes and binds its amino acid substrate based on specificity-conferring residues in the substrate binding pocket. The signature sequence motif in the binding pocket region is indicative of substrate preference and has been widely used for predicting substrates of NRPSs of unknown function, as well as guiding site-directed mutagenesis to alter substrate specificity of the A domain [80, 81]. After substrate binding, the A domain

activates the amino acid via adenylation from ATP. The aminoacyl-AMP is then transferred to a T domain through a covalent thioester bond to 4'-phosphopantetheine (Ppant arm). The T domain performs a carrier function in NRPS, covalently tethering the intermediate and carrying it to catalytic domains for elongation or modification. This mechanism facilitates the efficiency of the multi-enzyme assembly line, as covalent linkage between the intermediate and the T domain increases the local substrate concentration during reaction. In an "A-T-C" module, the C domain accepts an amino acylthioester intermediate from the preceding T domain and another amino acid building block from the following T domain, and forms a peptide bond [82]. The intermediate is transferred to the downstream module after the upstream module finishes its catalytic cycle, providing strict accuracy.

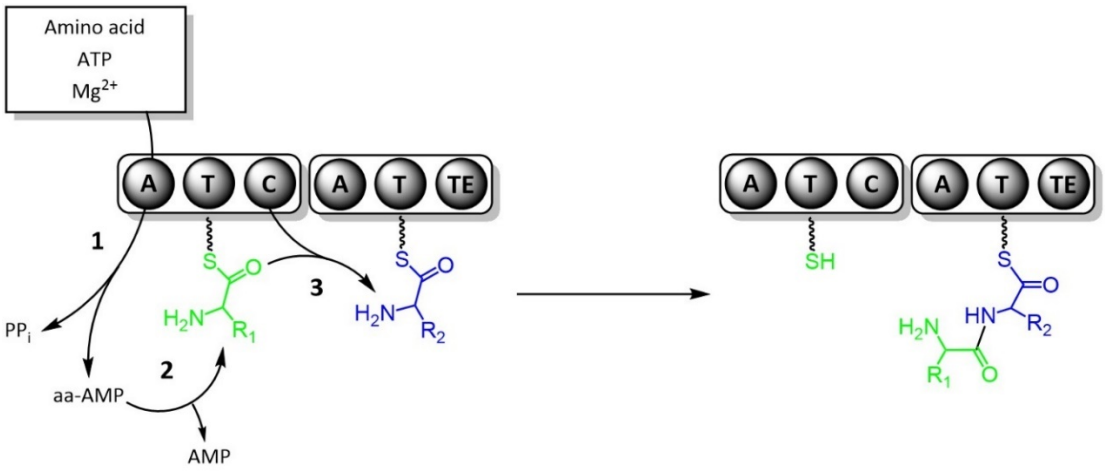


Figure 2.2. Reaction scheme of a dimodular NRPS. The first module has a canonical "A-T-C" domain composition. The A domain activates its amino acid substrate via adenylation, and loads it onto the T domain. The T domain transfers the intermediate to the C domain, where a peptide bond is formed and the dipeptidyl intermediate is transferred to the next module.

In the final synthetic step, the terminal domain for product release can vary, but is most commonly a thioesterase (TE) domain (Fig. 2.3) [84]. TEs are derived from serine hydrolases, where a catalytic Ser takes the mature peptide from its cognate T domain, forming an acyl-O-Ser-TE intermediate. The TE domain then catalyzes hydrolysis to release the product with a terminal carboxyl group. In some pathways, intramolecular cyclization can form a cyclic lactam or lactone. The difference between hydrolysis and cyclization can be explained simply by different solvation status of the active site during catalysis [85]. If no solvent molecule is present, then intramolecular nucleophilic attack can produce a cyclized product. In fungi, the terminal TE domain is often substituted with a specialized C domain. The C domain catalyzes a direct nucleophilic attack of an intermolecular or intramolecular amino group on the thioester bond and forms a peptide bond [86]. The mature peptide is released and forms a linear (intermolecular attack) or cyclized (intramolecular attack) product.

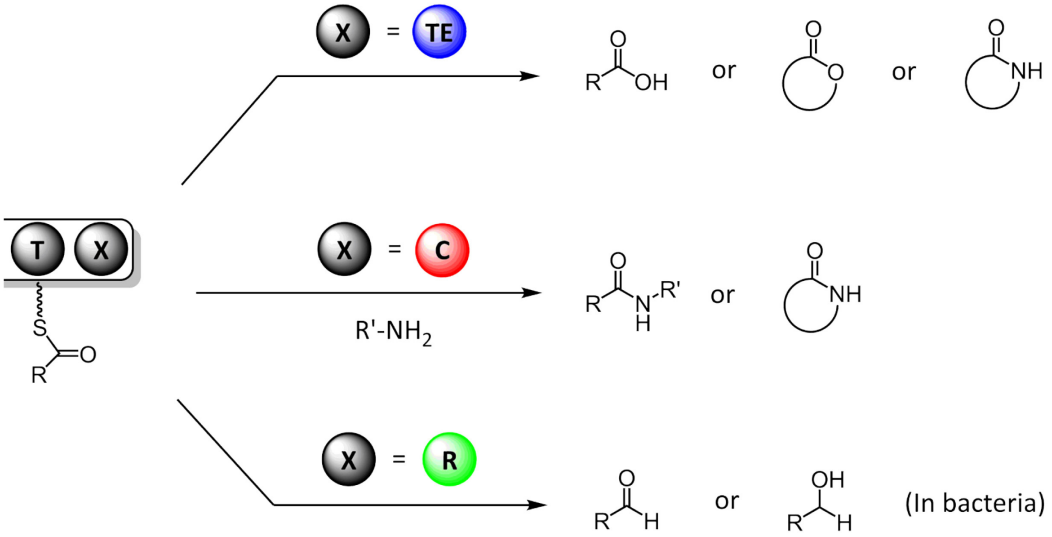


Figure 2.3. Examples of NRPS termination domains. TE, C or R domains can be found as terminal domains of NRPS, generating different products.

In some NRPS pathways, a terminal reductase (R) domain catalyzes a reductive release reaction. In bacteria, three types of reactions have been identified from different NRPS pathways with R domain termination: 2-electron reduction, 4-electron reduction and Dieckmann condensation [87-88] (Fig. 2.4). 2-electron reduction R domains produce products with a terminal aldehyde, while the 4-electron ones further reduce the aldehyde to an alcohol. For each step, the reaction is catalyzed by a canonical tyrosine-dependent short-chain dehydrogenase (SDR), with a "Ser/Thr-Tyr-Lys" catalytic triad. The Tyr is the proton donor, and an NAD(P)H cofactor transfers hydride. The Lys forms a hydrogen bond to the ribose hydroxyl of the cofactor, and lowers the pK_a of the catalytic Tyr. The Ser/Thr residue hydrogen bonds to a substrate hydroxyl or carbonyl group, orienting and stabilizing it to facilitate catalysis (Fig. 2.5.C). Some R domains produce tetramic acids or derivatives via Dieckmann condensation. In these cases, no reduction is carried on the substrates and no cofactor is needed for catalysis.

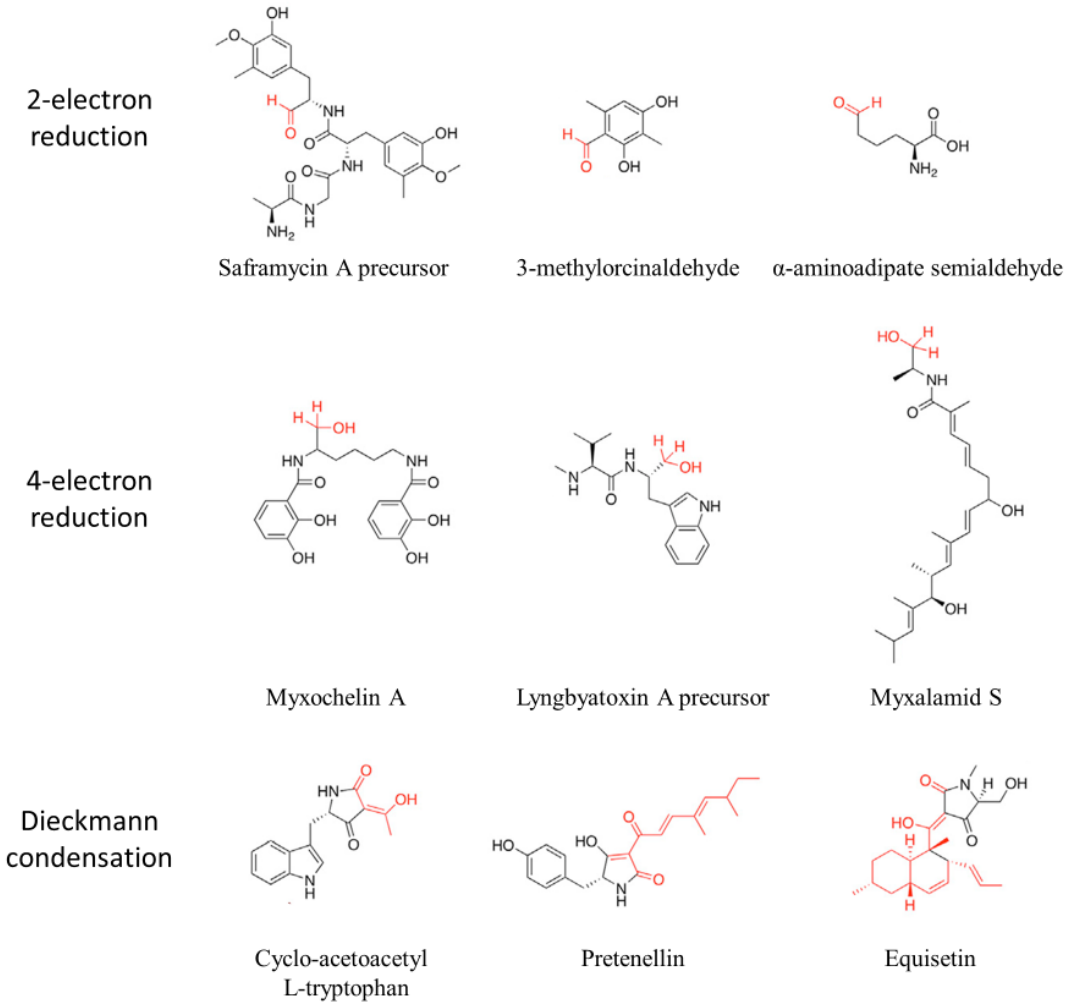


Figure 2.4. Products of terminal reductase domains [87]. The group added by the terminal R domain is colored in red.

Three crystal structures of bacterial NRPS terminal reductase domains have been reported, AusA R from *Staphylococcus aureus*, R_{NRP} from *Mycobacterium tuberculosis* and MxaA R from *Stigmatella aurantiaca* Sga15 [87-89]. AusA, a dimodular NRPS in the aureusimine biosynthetic pathway [89], has an "A-T-C-A-T-R" architecture and consumes L-Val and L-Tyr as substrates. The terminal AusA R domain catalyzes a 2-electron reduction reaction, producing a dipeptidyl aldehyde,

which is further cyclized to a dihydropyrazinone (Fig. 2.5.A). R_{NRP}, the terminal domain from a *Mycobacterium tuberculosis* NRPS, catalyzes the release of a lipopeptide intermediate through a 4-electron reduction, producing an alcohol [88]. The reaction proceeds in two steps, the first generating an aldehyde intermediate, and the second reducing the aldehyde to an alcohol. The second step was reported to be 15 times faster than the first reaction, yet no structural basis for facilitating the second half reaction was revealed, and no sequence motif was identified to distinguish 2-electron from 4-electron reductases. It was proposed that the 4-electron mechanism is a simple repeat of 2-electron reduction, based on the fact that the AusA and R_{NRP} active sites are quite similar, and share a common set of catalytic residues, which are derived from their SDR ancestors. MxaA is a NRPS from *Stigmatella aurantiaca* Sga15 and its terminal R domain reduces its acylthioester substrate to myxalamid S, a primary alcohol, via 4-electron reduction mechanism (Fig. 2.4). Compared with SDRs, bacterial R domains have two major structural differences: a unique N-terminal helix-turn-helix in the N-terminal nucleotide binding subdomain, and another helix-turn-helix near the C-terminus, which is proposed to interact with the T domain (Fig. 2.5.B).

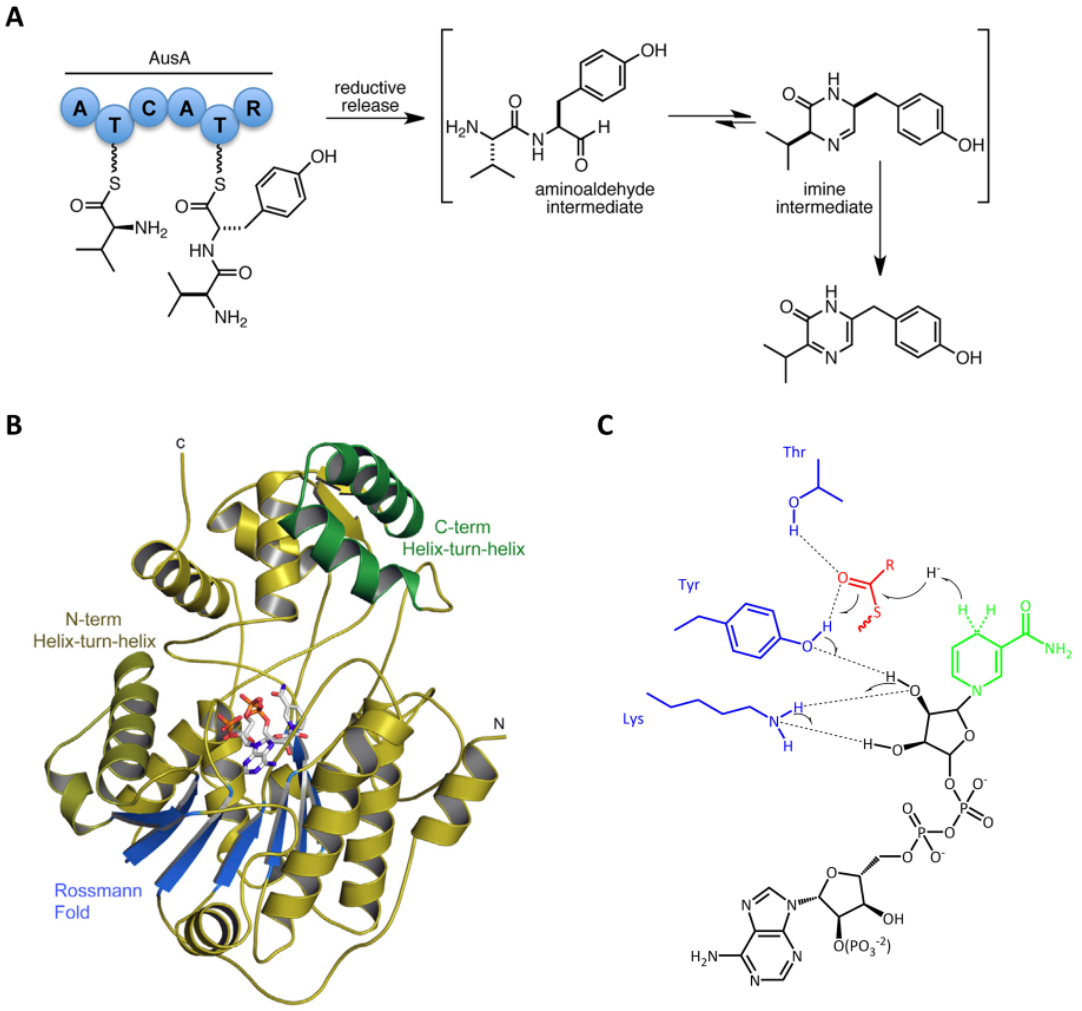


Figure 2.5. Terminal reductase domains of bacterial NRPS [87-89]. A. The initiation step of aureusimine biosynthesis (figure from reference [89]). B. Crystal structure of a bacterial NRPS (MxaA R from *Stigmatella aurantiaca* Sga15, PDB ID: 4DQV, figure from reference [87]). The central β sheet of the Rossmann fold is colored in blue, and the C-term helix-turn-helix is colored in green. The cofactor, NADPH, is shown as stick. C. Proposed reaction mechanism of bacterial NRPS R. The nicotinamide ring (green) of the NADPH cofactor transfers a hydride to the substrate (red), the catalytic Tyr (blue) donates a proton, and the double bond is reduced.

Fungal indole alkaloids are a large group of NRPS-produced products [90]. Most have been identified in ascomycetes, *Aspergillus* and *Penicillium* for example. In fungal indole alkaloid biosynthesis, most NRPSs are dimodular, producing dipeptidyl natural compounds. Four such pathways have been studied in collaboration with the Sherman

group: (-)/(+)-notoamide pathways, malbrancheamide pathway and paraherquamide pathway [40]. In the notoamide pathways, the NRPS (NotE/NotE') domain composition is "A-T-C-A-T-C". The first module accepts L-Pro and the second recognizes L-Trp, producing brevianamide F as the NRPS product [37] (Fig. 2.6.A). In the homologous malbrancheamide pathway, however, the proposed domain composition of MalG is "A-T-C-A-T-R" based on sequence analysis, suggesting a terminal reductive hydrolysis mechanism. Based on knowledge of bacterial NRPS R domains, the product was proposed to be a dipeptidyl aldehyde by the Sherman group [40] (Fig. 2.6.B). However, the sequence identity between this predicted fungal R domain and known bacterial R domains is quite low (15 - 22 %), and no structure of a fungal NRPS terminal reductase has been reported. Biochemical and structural investigation of these terminal enzymes will help to elucidate the products and the reaction mechanisms, and comparison with bacterial homologs may provide interesting discoveries and thoughts for future research.

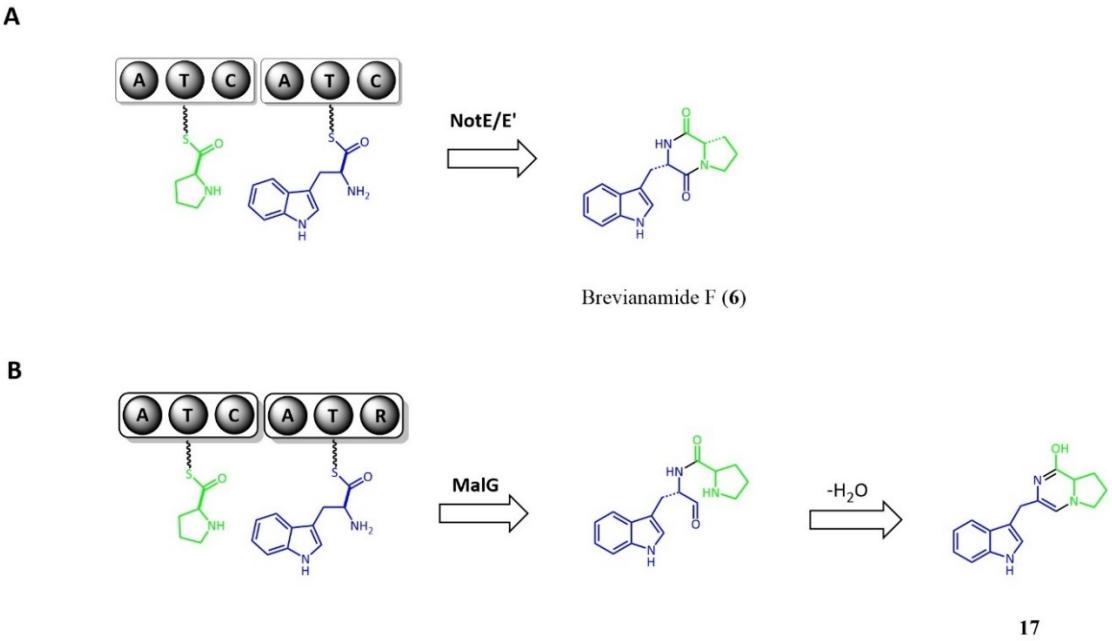


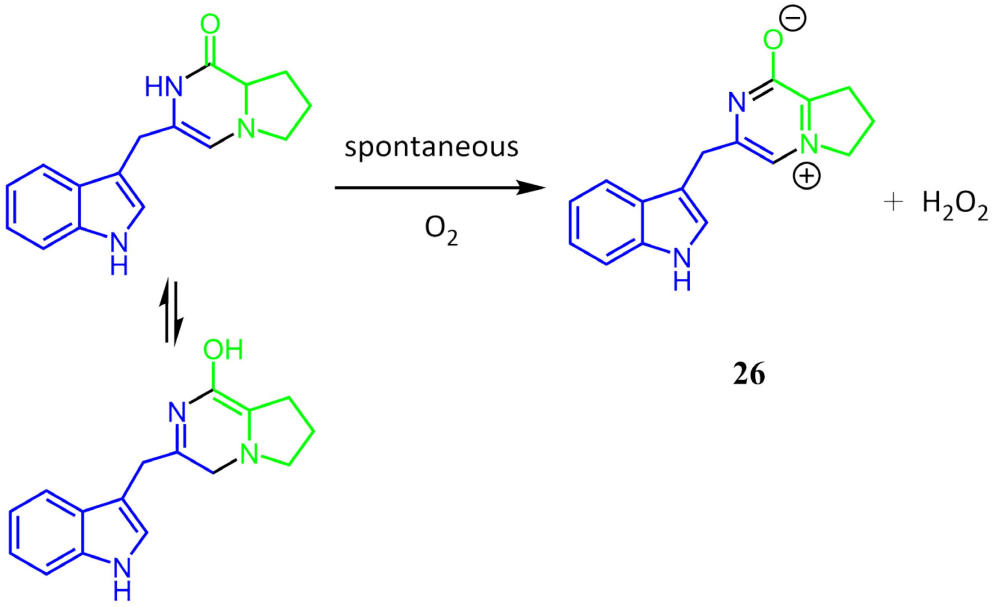
Figure 2.6. The initiation step of notoamide biosynthesis (A) and the proposed initiation step of malbrancheamide biosynthesis (B). The (-) notoamide pathway NRPS (NotE) terminates with a C domain, producing brevianamide F, while the malbrancheamide pathway NRPS (MalG) R domain is proposed to catalyze a reductive release reaction.

2.2 Results

2.2.1 Proposed MalG product showed instability *in vitro*.

To test the hypothesis that malbrancheamide NRPS (MalG) R catalyzes a reductive release reaction, L-Pro-L-Trp dipeptidyl aldehyde (**17**), the proposed malbrancheamide NRPS (MalG) product, was chemically synthesized by the Williams group at Colorado State University. Hong Tran in the Sherman group tested whether the synthetic substrate could be consumed by MalE, which was proposed to prenylate the NRPS product. However, the majority of **17** was spontaneously and irreversibly converted to **26** without any enzyme (Fig. 2.7). The structure of **26** was determined by LC/MS and NMR analysis, leading to the conclusion that **17** was unstable *in vitro* and rapidly oxidized to **26**, as the reaction was much slower when

carried out in an anaerobic reaction chamber. However, **26** is unlikely to be a pathway intermediate in the malbrancheamide pathway, because the aromatic ring is highly stable and cannot easily form a diene group, which is proposed to be a key factor for synthesizing premalbrancheamide, the common pathway intermediate (Fig. 1.6).



17 **26**

Figure 2.7. Proposed reaction scheme for spontaneous conversion of **17** to **26**. The structure of **26** was validated by LC/MS and ^{1}H NMR by Hong Tran.

The fact that **17** was highly unstable *in vitro* suggests that it might not be the MalG product. To test this possibility, I sought to produce a MalG R domain to perform an *in vitro* assay. I analyzed the MalG sequence to identify domain boundaries for the T and R domains, and developed expression and purification protocols for both domains and the T-R didomain. Hong Tran incubated the MalG R with several dipeptidyl analogs. However, none of tested compounds was consumed directly by the MalG R

domain and no product was detected. Instead, the amino groups of the analogs were highly reactive and spontaneously cyclized to produce cyclo-dipeptides.

2.2.2 Diffusive loading of substrates onto the MalG T domain

In vivo, the MalG R substrate is carried on the MalG T domain with a Ppant arm, which may contribute to the observed inefficiency of MalG R recognizing and acting on unlinked dipeptidyl analogs. Thus, it seems necessary to acquire a MalG T-loaded substrate to identify the terminal R domain product. For that purpose, a group of dipeptidyl thiophenol analogs were synthesized by the Williams group (Fig. 2.S1). Thiophenol is an excellent leaving group and thiophenol compounds can spontaneously acylate protein thiols, including the Ppant thiol. In these analogs, the N atom of the prolyl group was replaced by a C or O atom to prevent spontaneous cyclization, which may compete with substrate loading.

The MalG T domain was used for substrate loading. MalG T was co-expressed with sfp (a nonspecific 4'-phosphopantetheinyl transferase) [91] to transfer the Ppant arm onto MalG T *in vivo*. Intact protein mass spectrometry was used to identify the efficiency of Ppant transfer. Almost all (> 99%) MalG T was loaded with Ppant (Fig. 2.8).

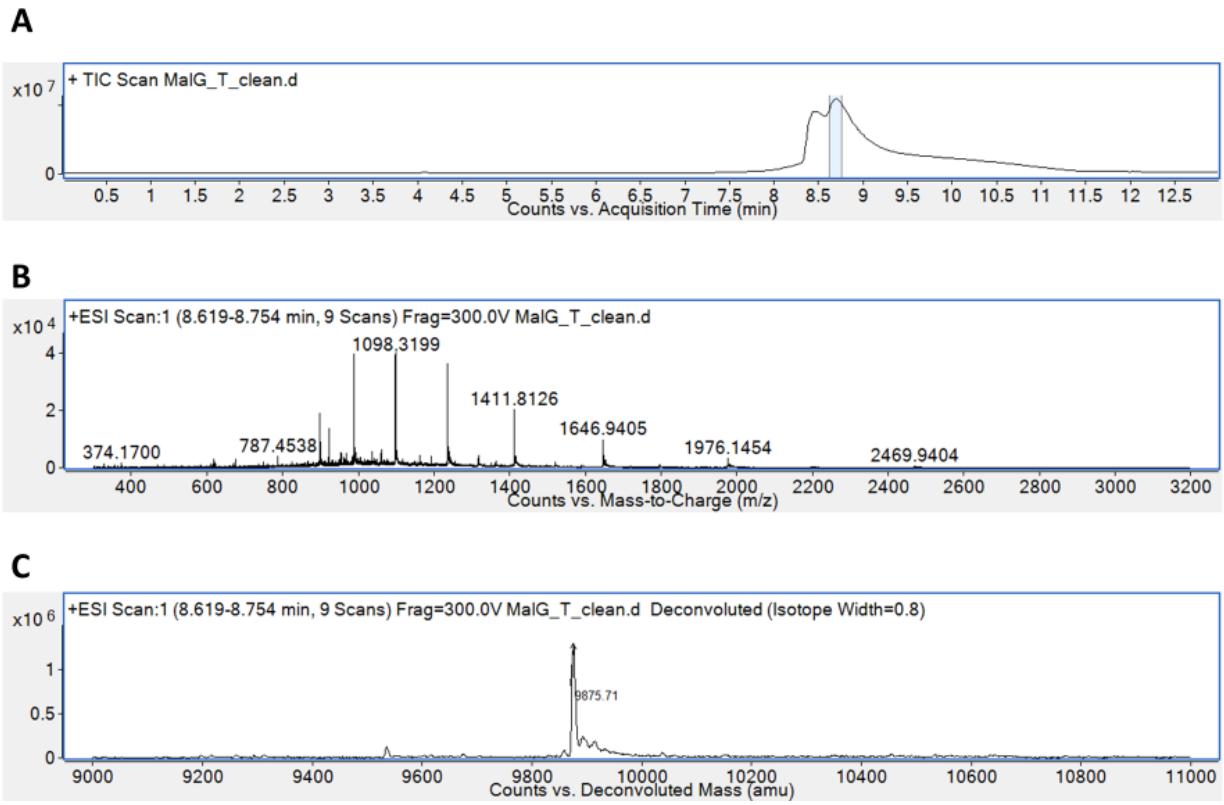


Figure 2.8. Test of the Ppant loading status of MalG T. A. HPLC elution profile of MalG T. B. Mass spectrum of the HPLC profile peak. C. Deconvoluted mass of MalG T (observed mass: 9875.71 Da, calculated mass of holo MalG T: 9875.8 Da).

To load the substrate onto MalG T, holo MalG T (MalG T-Ppant) was incubated with the thiophenol analog under different pH (8.1, 9.5 and 10.0) and buffer (HEPPS, CHES and borate) conditions. LC/MS was performed to check the substrate loading efficiency (Fig. 2.9). The majority of MalG T (~50%) remained unloaded with substrate (observed mass: 9875.71 Da; calculated mass: 9875.8 Da), while ~10% was loaded with one substrate molecule (observed mass: 10158.57 Da; calculated mass: 10159.8 Da). However, ~10% contained two copies of substrates (observed mass: 10444.86 Da; calculated mass: 10443.8 Da), and ~20% had three substrates attached to the T domain (observed mass: 10729.43 Da; calculated mass: 10727.8 Da). The additional substrate loading probably occurred on one or both of the two Cys residues

of MalG T, as Cys is nucleophilic at high pH. Loaded species were given to MalG R, but no offloaded product was detected by LC/MS. In summary, the efficiency of substrate diffusive loading onto the MalG T domain was very low, and off-target loading was observed to a certain extent, which may have hampered the efficiency of substrate recognition and catalysis of the terminal R domain. To fix that problem, dipeptidyl-CoA substrates are now being synthesized chemically, and sfp-catalyzed enzymatic loading will be tested soon.

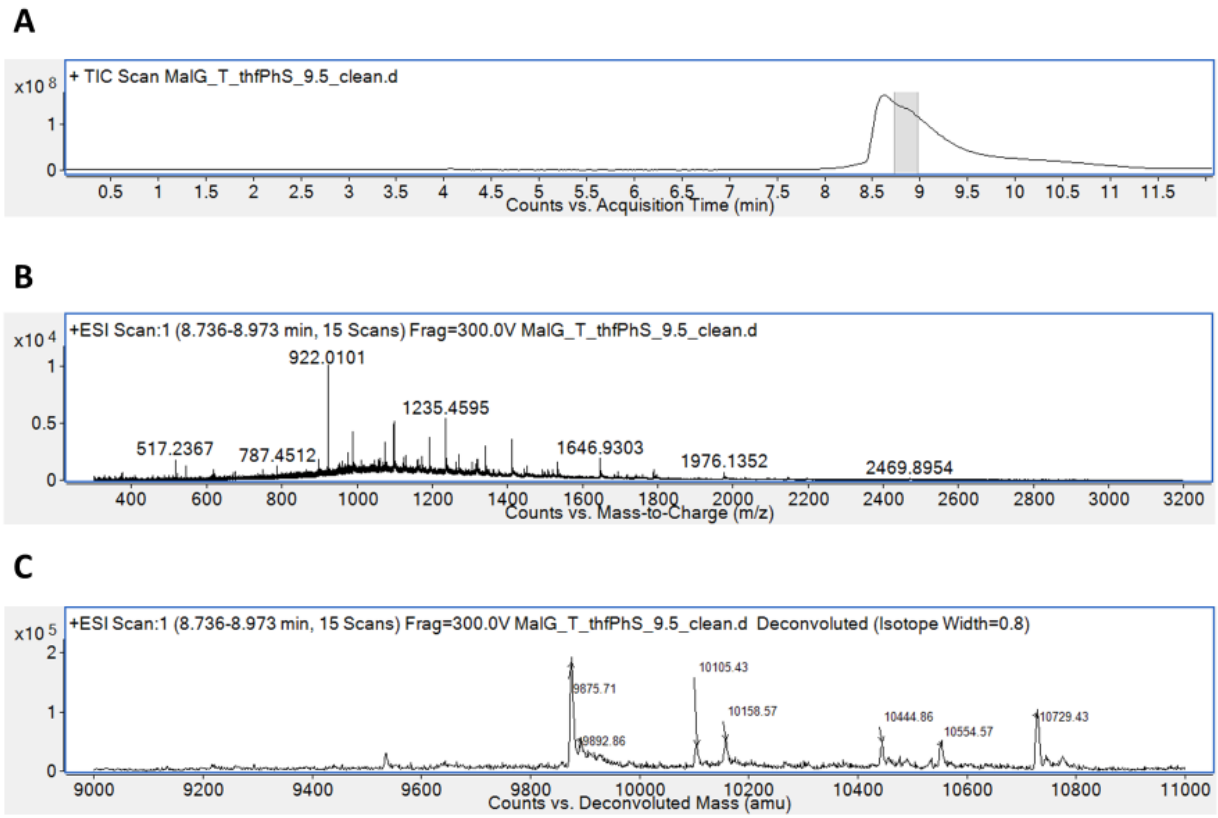


Figure 2.9. Test of thiophenol substrate loading of MalG T. A. HPLC elution profile of substrate loaded MalG T. B. Mass spectrum of the peak of the HPLC profile. C. Deconvoluted mass of substrate loaded MalG T (observed mass: 9875.71 Da, 10105.43 Da, 10158.57 Da, 10444.86 Da, 10554.57 Da, 10729.43 Da).

2.2.3 Structural analysis of fungal NRPS terminal reductase domain

To help understand the function of the fungal NRPS terminal reductase domain, structural studies were performed on MalG R. However, the MalG R domain was recalcitrant to crystallization, leading to investigation of a homologous paraherquamide NRPS (PhqB) reductase domain. As an NRPS in the paraherquamide pathway, PhqB is also proposed to terminate with a reductase domain. PhqB R shares 37% sequence identity with MalG R, and they catalyze similar reactions, based on the similarity of both intermediates and products of the two pathways (Fig. 2.10).

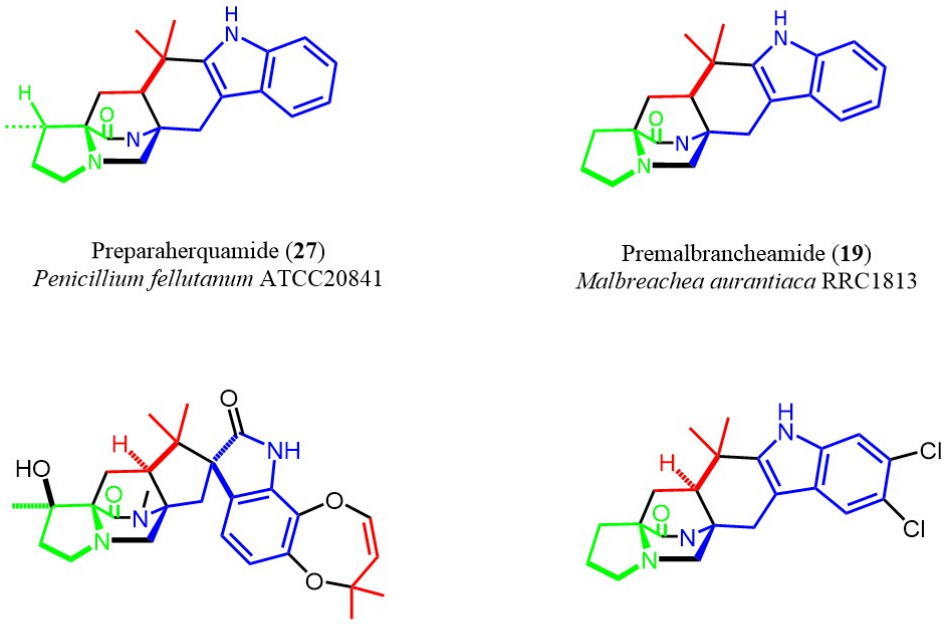


Figure 2.10. Structures of paraherquamides and malbrancheamides. The common pathway intermediates, preparaherquamide and premalbrancheamide, are very similar and only differ by a methyl group, suggesting a similar pathway scheme.

PhqB R crystals were obtained, which diffracted to 2.6 Å to the best. Published bacterial R domain structures have low sequence identity to PhqB R (22% at most), so the initial plan to solve the structure was to produce selenomethionyl (SeMet) protein for phase determination by anomalous diffraction. However, the yield of soluble SeMet R domain was vanishingly small. A series of heavy atom soaks was carried out and data were collected from several crystals soaked in KI; no usable data were obtained from crystals soaked in other compounds. The overall poor quality of crystals greatly complicated the heavy atom studies. Nevertheless, five iodine sites were identified that appeared to be consistent in data from a few crystals, but data from the derivatized crystals was poorly isomorphous with the native data. Additionally, phase improvement from non-crystallographic symmetry (NCS) averaging was not available due to the single copy of PhqB R in the asymmetric unit, further complicating the structure determination. Attempts to solve the structure by molecular replacement (MR) with bacterial R domain structures finally succeeded using MR-ROSETTA [97], which rebuilds the initial MR probe structures with ROSETTA, and further rebuilds the resulting models with PHENIX.autobuild. After hundreds of cycles of rebuilding, more than 100 residues that were not in the initial MR models could be modeled in the density map. After several cycles of further refinement, a 2.65 Å crystal structure of paraherquamide NRPS (PhqB) reductase domain was obtained.

The PhqB R is a tetramer in solution and in crystals, which differs from the MalG R (dimer) and the characterized bacterial NRPS R domains, which are often monomers

or dimers. Although tetrameric NRPS R domains are rare, their SDR ancestors are commonly tetramers. The PhqB R tetramer is located at a position of D2 point symmetry in crystals of space group *I*222, as each subunit interacts with all three other subunits, forming three different binding interfaces (Fig. 2.11). PDBePISA [92] was used to calculate the interface areas. The sizes of interfaces 1, 2 and 3 were determined to be 918 $Å^2$, 545 $Å^2$ and 433 $Å^2$, covering 5.8 %, 3.5 % and 2.8 % of the total subunit surface area, respectively. The overall percentage of covered surface area is 12.1%, which is within the range of subunit interfaces of oligomeric proteins (>10%). The interfaces are stabilized by both hydrophobic and hydrophilic interactions, and the calculated ΔG of tetrameric packing is -9.3 kcal/mol. The tetrameric oligomer state may be inconsistent with the general observation that NRPSs function as monomers. However, the N-termini of the subunits are at the exterior of the tetramer and well separated from one another in an arrangement that would allow flexible tethering of a "monomeric" NRPS module.

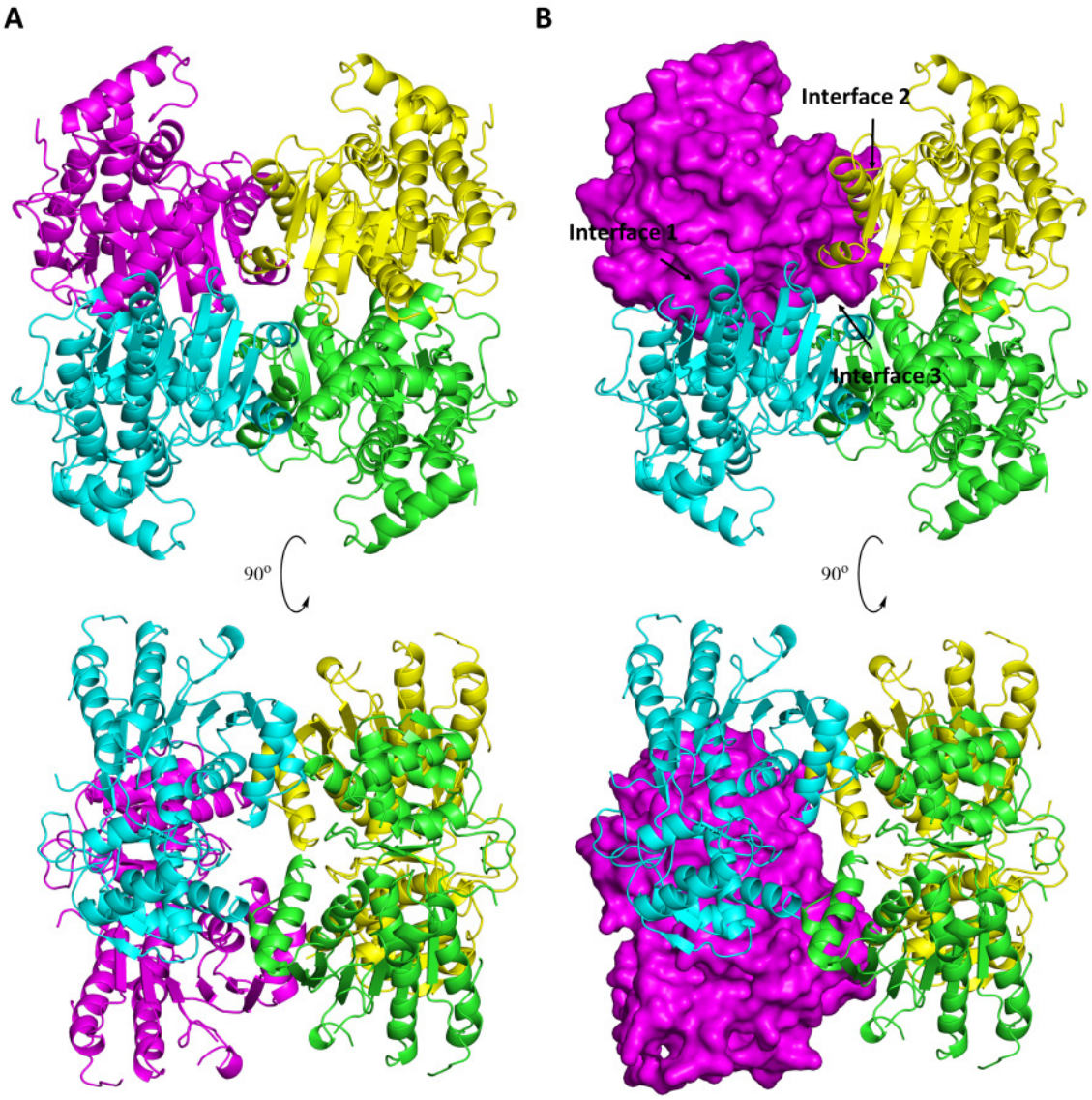


Figure 2.11. Structure of the PhqB R tetramer. Each subunit contacts three other subunits, forming three distinct interfaces (interface 1, 2 and 3).

Despite very low sequence identity (15 - 20%), the PhqB R structure resembles its bacterial homologs, with an N-terminal nucleotide binding subdomain and a C-terminal substrate binding subdomain (Fig. 2.12). The N-terminal nucleotide binding subdomain has a typical Rossmann fold, with a parallel β sheet (β1, β2, β3, β4, β5, β6 and β10) flanked by six α helices (α1, α2, α3, α4, α5 and α7). No cofactor is bound in the crystal structure, which is presumed to be NADPH. This hypothesis is further

supported by the observation of two conserved Arg residues (Arg65 and Arg75) in the structure, which may be responsible for selectively coordinating the extra phosphate group of NADPH over NADH (Fig. 2.13). The C-terminal subdomain, which is unique to NRPS terminal reductases, is composed of five α helices (α6, α8, α9, α10 and α11) and covers the active site. The proposed active site contains three conserved residues (Ser177 after β5, Tyr213 and Lys217 on α5), suggestive of a canonical "Ser-Tyr-Lys" catalytic triad and a similar reaction mechanism to bacterial NRPS terminal reductase domains and other Tyr-dependent SDRs (Fig. 2.13).

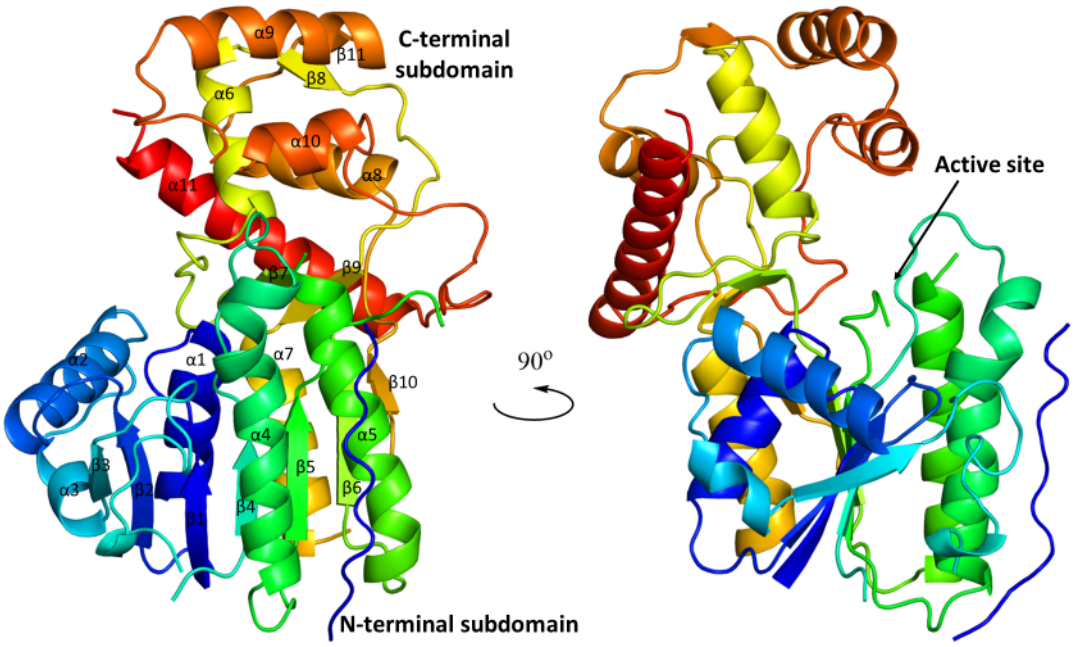


Figure 2.12. Structure of the PhqB R subunit in a rainbow color scheme (N-terminus in blue, C-terminus in red).

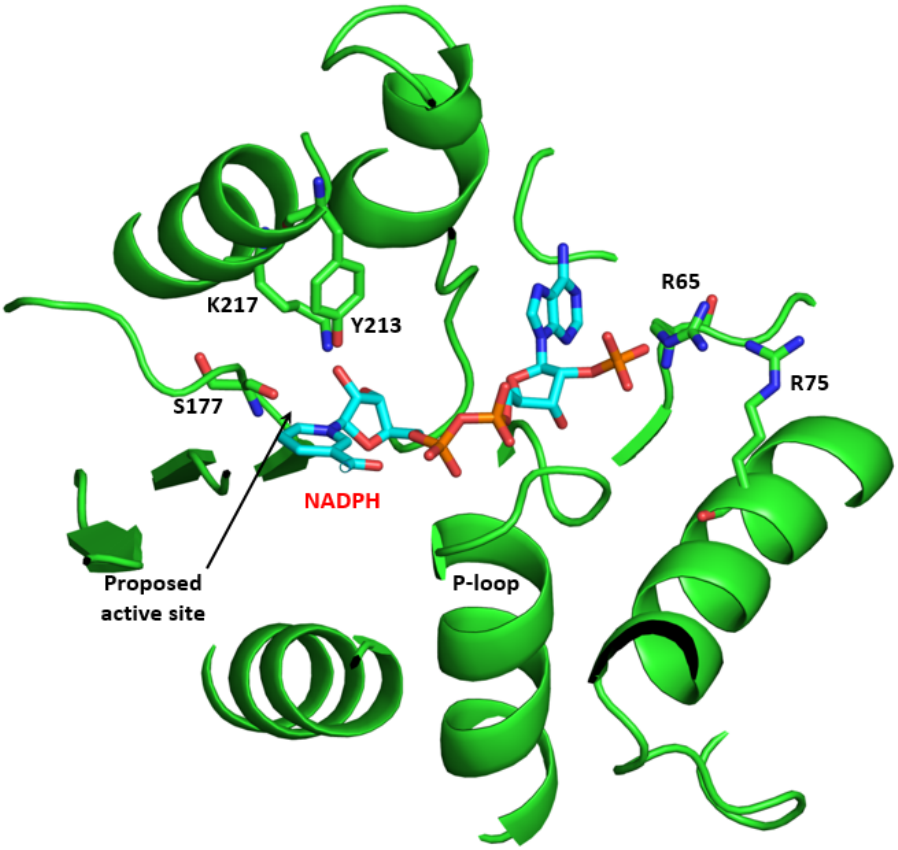


Figure 2.13. Proposed nucleotide binding site and active site in the PhqB R structure. NADPH was modeled in the structure using AutoDock [136]. Two Arg residues (R65 and R75) are proposed to coordinate the NADPH 2'-phosphate. Three conserved residues (S117, Y213 and K217) were found in the proposed active site.

A search for structural homologs using DALI [93] revealed the three bacterial NRPS terminal reductase domains as closest structural homologs of PhqB R, with Z scores of 31.2 (MxaA R, PDB ID: 4DQV), 26.5 (R_{NRP}, PDB ID: 4U7W) and 22.8 (AusA R, PDB ID: 4F6C) [87-89]. Superposition of these structures shows high levels of structural conservation, with several minor differences. Relative to the other structures, the PhqB R C-terminal α helix (α11) is tilted towards the core, and the loop preceding α11 is significantly shorter (Fig. 2.14). In bacterial NRPS R domains, the loop is extended and an extra α helix is present. This short loop feature is

characteristic of the PhqB R tetramer, as the loop would clash with neighboring subunit (magenta and cyan subunits in Fig. 2.11) if structured as in the bacterial homologs. Another major difference is the helix-turn-helix (α9 and α10) region of the C-terminal domain, which is predicted to bind the T domain. The structural difference may suggest different fungal and bacterial T-R interaction modes, yet more evidence is needed to conclude this. As to the catalytic residues, the bacterial NRPS R domains use "Thr-Tyr-Lys" catalytic triads to catalyze a 2-electron reduction reaction in one step (AusA R), or a 4-electron reduction reaction in two steps (MxaA R and R_{NRP}). In the PhqB R structure, both the Tyr and Lys are conserved, while the Thr was replaced by a Ser, forming a canonical "Ser-Tyr-Lys" catalytic triad. The loop between the proposed catalytic Ser and Tyr is disordered in the PhqB R structure, suggesting the possibility of a conformational change upon cofactor or substrate binding. Taking all evidence into consideration, the fungal NRPS R domains in the malbrancheamide and homologous paraherquamide pathways produce either dipeptidyl aldehydes or dipeptidyl alcohols. However, no significant structural difference was identified to distinguish 2-electron reduction from 4-electron reduction, and the PhqB R structure offers no hint on the exact fungal NRPS reaction or products. Enzymatic characterization of different substrates is needed to determine the final products.

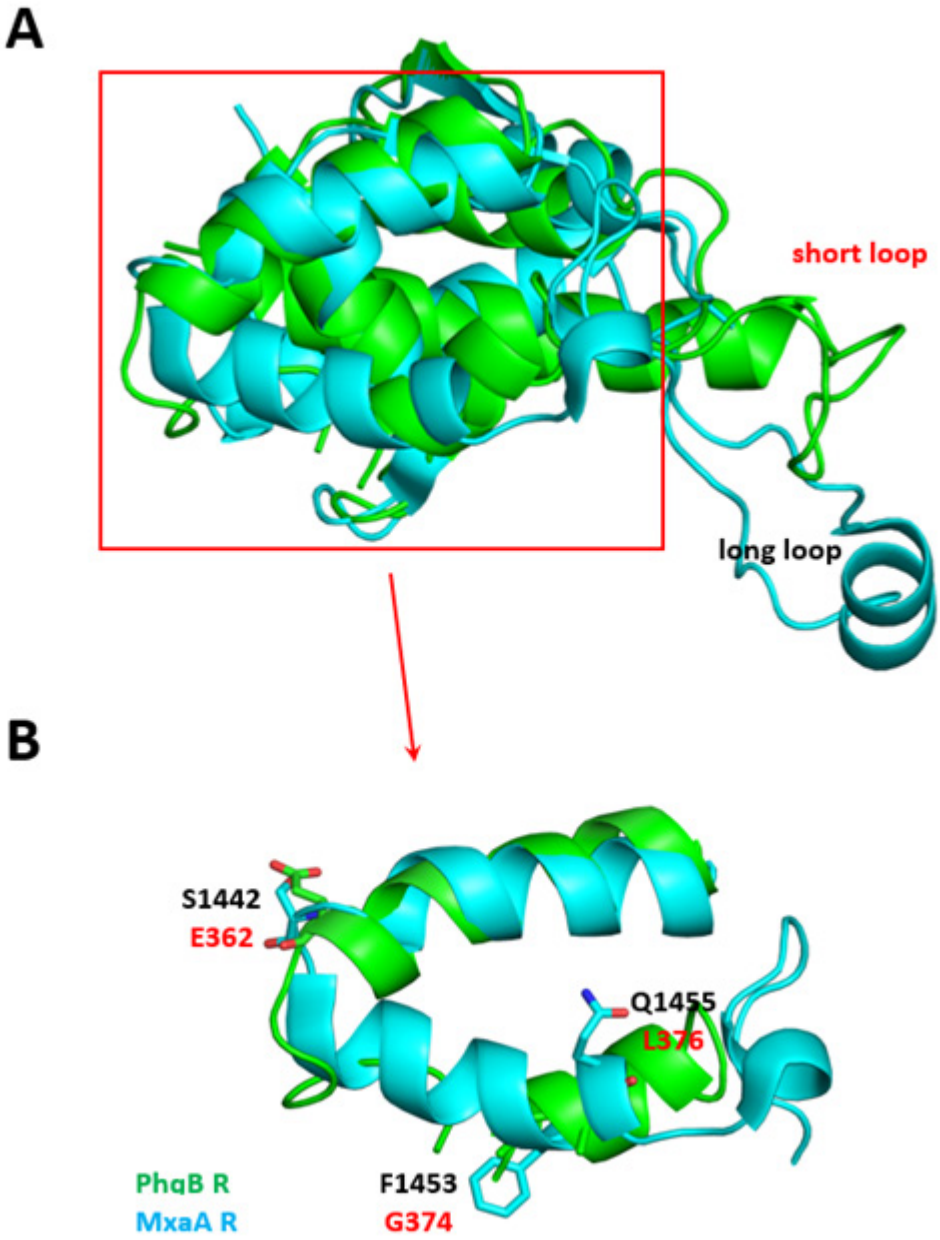


Figure 2.14. Superposition of PhqB R (green) and MxaA R (cyan) C-terminus. A. The loop that precedes the last α helix is much shorter in PhqB R (labeled in red) than in MxaA R (black). B. Close view of C-terminal helix-turn-helix. In MxaA R, three residues (S1442, F1453 and Q1455) were proposed to directly contact the MxaA T domain. None of these residues is conserved in PhqB R.

2.3 Discussion

The structure of PhqB R resembles bacterial NRPS reductase domains, and suggests a canonical "Ser-Tyr-Lys" catalytic triad and NADPH-dependent catalytic mechanism. Based on that, the NRPS terminal reductase domain in the malbrancheamide pathway should catalyze either a two-electron or a four-electron reduction reaction, producing a dipeptidyl aldehyde or an alcohol. Sequence alignments and protein BLAST were performed to predict functions of proteins encoded by the malbrancheamide gene cluster by Shengying Li in the Sherman group [40] (Table 2.1).

Table 2.1. Predicted functions of proteins in malbrancheamide biosynthesis

Mal Proteins (AA)	**Proposed Function**
MalA (667)	FAD-dependent halogenase
MalB (369)	Indole prenyltransferase
MalC (264)	Short-chain dehydrogenase
MalD (336)	Negative transcription regulator
MalE (438)	Indole prenyltransferase
MalF (590)	FAD-dependent oxidoreductase
MalG (2345, A-T-C-A-T-R)	Dimodular NRPS

To produce premalbrancheamide, the proposed Diels-Alder product, an NRPS (MalG), a prenyltransferase (MalB or MalE) and maybe a short-chain dehydrogenase (MalC) are needed to function in defined order, as shown in Figure 2.15. To identify the reaction sequence of malbrancheamide biosynthesis and determine the NRPS

product, MalB, MalC and MalE were expressed and purified to homogeneity, activity tests of which will be discussed in later chapters. The aldehyde is more likely to be the NRPS product based on the fact that its tautomeric form **18** provides the diene needed for the proposed Diels-Alder reaction (Fig 2.15). No further reductive step is needed, and MalC was shown to be unable to oxidize dipeptidyl alcohol, the 4-electron product, back to dipeptidyl aldehyde (Fig. 4.3). However, the aldehyde product was shown to be unstable *in vitro* (Fig. 2.7), and it is unknown how the fungi might overcome the problem of an unstable intermediate.

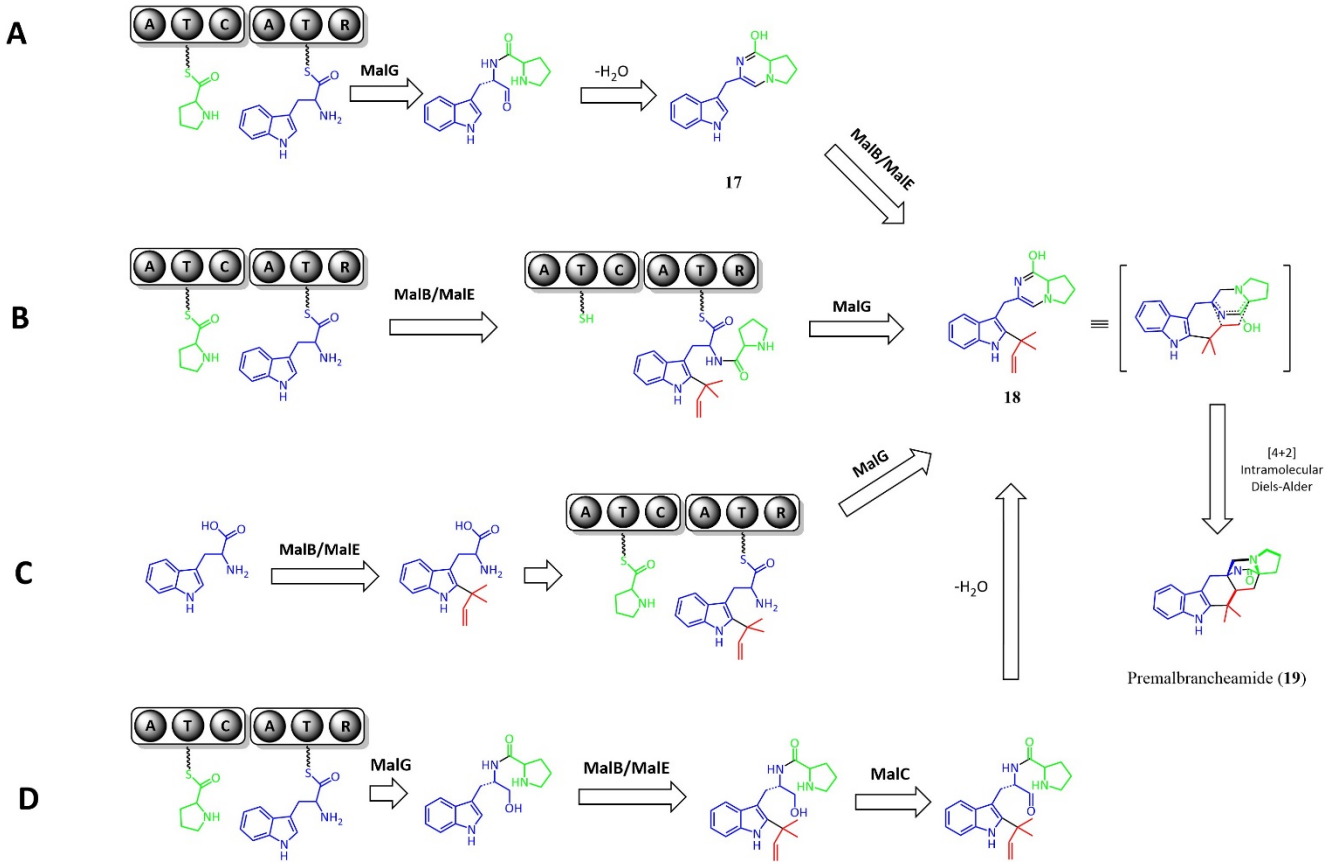


Figure 2.15. Possible reaction sequences of the early steps in malbrancheamide biosynthesis. MalG, MalB, MalE and perhaps MalC may function in the sequences illustrated to produce premalbrancheamide.

Another interesting feature of the NRPS R is the T-R didomain interaction. Structural studies of bacterial NRPS R domains suggest that the T domain binds the C-terminal helix-turn-helix region of R. This hypothesis is reasonable considering the position of the R domain N-terminus, which connects to the T domain by a linker of fewer than 10 amino acids. The C-terminal helix-turn-helix is near the N-terminal residue, and such an interaction pattern was supported by docking studies of MxaA T-R domain interactions [87], however, no crystal structure of the T-R didomain has been reported. The predicted MxaA T and R residues on the proposed binding interface are not conserved in MalG R or PhqB R, and structure of the C-terminal helix-turn-helix also differs (Fig. 2.14). Thus, it is possible that the terminal R domain interacts with T in a different way in the fungal and bacterial systems. Once efficient substrate loading onto the T domain can be achieved, the T-R didomain interaction can be tested by comparing substrate or cofactor consumption rates with wild-type and mutagenized R domains.

2.4 Materials and Methods

2.4.1 Materials and strains

NADH and NADPH were purchased commercially. *E. coli* XL1-Blue was used for vector storage, *E. coli* pRare, Bap1 and pGro7 were used for production of recombinant protein. The thiophenol compounds (Fig. 2.S1) were synthesized and provided by the Williams group.

2.4.2 Cloning of *malG T*, *R*, *T-R* and *phqB R*

The cDNA library of *Malbranchea aurantiaca* was generated by Shengying Li and Hong Tran. The gDNA of *Penicillium fellutanum* ATCC20841 was extracted by Sean Newmister. For cloning of *malG R*, PCR was used to amplify the cDNA template, followed by a ligation independent cloning (LIC) procedure [94, 95] to insert the genes into the pMCSG7 vector. For *malG T*, *malG T-R* and *phqB R*, the pMCSG9 vector was used to insert the gene. PCR primers are listed in Table 2.2. The plasmids (pMCSG9-MalG T, pMCSG7-MalG R, pMCSG9-MalG TR and pMCSG9-PhqB R) were then transformed into *E. coli* XL1-Blue cells for plasmid storage and harvest.

Table 2.2. Primers used for cloning

Gene	Oligo Sequence (5’ -> 3’)
malG T (Forward)	TACTTCCAATCCAATGCCACACTTCAACCTCACGAAAGCAC
malG T (Reverse)	TTATCCACTTCCAATGCTAAACCCCTTCAATGAGCCTGG
malG R (Forward)	TACTTCCAATCCAATGCCATGTCTGATGATCCGCTTCTGTC
malG R (Reverse)	TTATCCACTTCCAATGCTATCACAGGACGCGTCTAAAAATACG
malG T-R (Forward)	TACTTCCAATCCAATGCCACACTTCAACCTCACGAAAGCAC
malG T-R (Reverse)	TTATCCACTTCCAATGCTATCACAGGACGCGTCTAAAAATACG
phqB R (Forward)	TACTTCCAATCCAATGCCTGGTGGGAGAGGGTGCAA
phqB R (Reverse)	TTATCCACTTCCAATGCTATTAAGAGTTGATAAGACCATTCCC

2.4.3 Overexpression and purification

For expression of *malG R*, *E. coli* pRare cells were transformed with pMCSG7-MalG R and grown in Terrific Broth medium (30 µg/ml ampicillin and 100 µg/ml spectinomycin added) at 37 °C to a O.D. 600 = ~1.0. The culture was then transferred to 20 °C over 1 hour, and induced with 0.4 mM IPTG. After 18-20 hours of incubation (20°C, 225 rpm shake), cells were harvested by centrifugation and the cell pellet was stored at -20 °C. *MalG T* and *T-R* were overexpressed in *E. coli* Bap1 cells with the same protocol as *malG R*. *PhqB R* was overexpressed in *E. coli* pGro7 cells. The culture was induced with 0.4 mM IPTG and 1 mg/mL L-arabinose.

For purification, the cell pellet was resuspended in lysis buffer (10% v/v glycerol, 500 mM NaCl, 20 mM Tris buffer pH 7.9, 20 mM imidazole, 5 mg lysozyme, 2 mg

DNAse and 1 mM $MgCl_2$), and vortexed to mix for 30 min. Sonication and high speed centrifugation (16000 rpm, 30 min) were applied to obtain the lysate soluble fraction. The soluble fraction was filtered and loaded on a GE Ni-NTA HisTrap column (Ni-NTA buffer: 10% glycerol, 500 mM NaCl, 20 mM imidazole pH 7.9, 20 mM Tris pH 7.9; flow rate: 3 mL/min), and was eluted with a 20 – 600 mM imidazole gradient (Ni-NTA buffer; flow rate: 3 mL/min; gradient time: 12 min). Fractions containing the R or T-R domain were pooled and incubated with TEV protease in a 1:50 w/w ratio at 20 °C for 2 hours to remove the N-terminal His-tag or His-MBP-tag. The tag-free protein was dialyzed overnight at 4 °C into 10% glycerol, 2 mM DTT, 500 mM NaCl, 20 mM Tris pH 7.9, and eluted from a Ni-NTA HisTrap column (Ni-NTA buffer; flow rate: 3 mL/min). Further homogeneity was achieved by size-exclusion chromatography equilibrated with 10% v/v glycerol, 300 mM NaCl, 20 mM Tris pH 7.9 (GE HiLoad 16/60 Superdex 200 prep grade column; flow rate: 0.5 mL/min). SDS-PAGE was used to assess protein homogeneity, confirming >95% purity. Yields of purified proteins were 50 mg MalG R domain per L of *E. coli* culture, 40 mg/L MalG T-R, 5 mg/L MalG T, and 20 mg/L PhqB R.

2.4.4 Crystallization and Structural Determination

For crystallization, wild-type PhqB R protein was mixed with precipitant solution (10 % PEG 8000, 200 mM $MgCl_2$, 100 mM Tris pH 7.0) in a 1:1 v/v ratio. Crystals grew at 4 °C within 24-48 hours, and were harvested into precipitant solution with 25 % glycerol for cryoprotection, and flash cooled in liquid nitrogen. Diffraction data

were collected at the Advanced Photon Source (GM/CA beamline 23-ID-D) at an X-ray wavelength of 1.033 Å (360° of data, 100 K, 0.2° image width). Crystals grew reproducibly, but had generally poor diffraction quality, most with d_{min} poorer than 4 Å. The data described in Table 2.3 are the best obtained from more than 400 crystals screened. Data were processed with XDS [96]. The crystal structure was solved by MR-ROSETTA [97] using the structure of MxaA R as an initial probe. Model building was carried out with Coot [98]. Refinement was carried out with PHENIX refine [99].

2.4.5 Diffusive loading of substrates onto MalG T domain

MalG T (50 μM) was added to 10% v/v glycerol, 50 mM NaCl, 50 mM HEPES pH 8 or CHES pH 9.5 or borate pH 10. The loading reaction was initiated by addition of 1 mM thiophenol substrate. The reaction mix was incubated at 20°C for 12 hours, and quenched with 50% v/v methanol. Centrifugation was used to remove denatured protein, which was resuspended in $CHCl_3$ and analyzed by LC/MS. (Phenomenex Kinetix reverse-phase C18 column (40 mm × 2.1 mm, 2.6 μm), Buffer A: 0.2% v/v formic acid in water, Buffer B: 0.2% v/v formic acid in acetonitrile. HPLC protocol: 5% Buffer A for 4 min, 20-100% Buffer B gradient for 4 min, 100% Buffer B for 2 min. Flow rate: 0.5 mL/min.)

Author Contributions

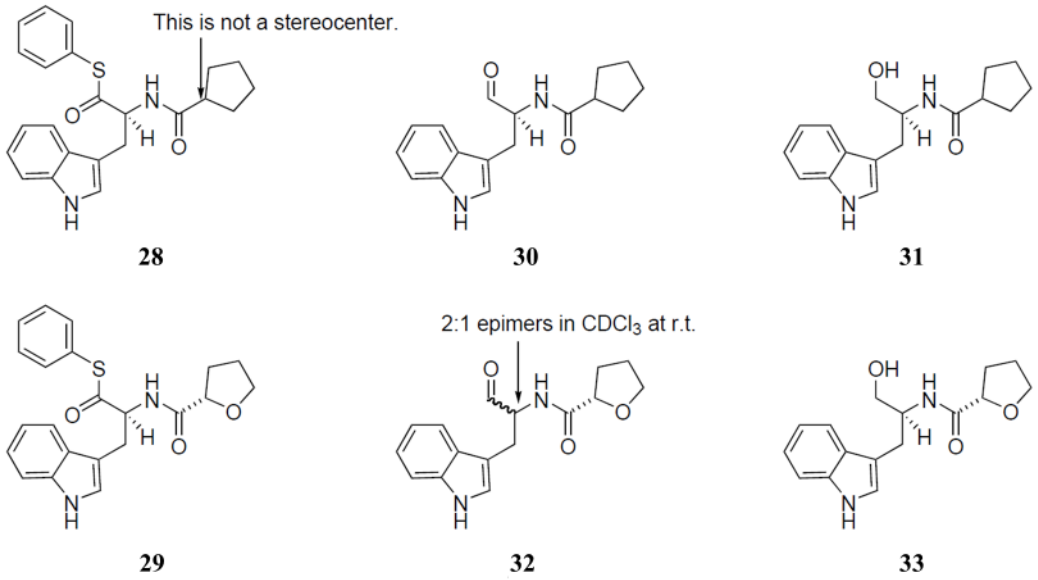
Qingyun Dan, Sean Newmister, Hong Tran, Robert Williams, David Sherman and Janet Smith contributed to the experimental design. Shengying Li and Hong Tran generated the cDNA library of *Malbranchea aurantiaca* RRC1813. Sean Newmister generated the gDNA of *Penicillium fellutanum* ATCC20841. Le Zhao synthesized thiophenol compounds. Qingyun Dan cloned and expressed *malG T-R*, *malG R*, *malG T* and *phqB R*, purified all proteins, and carried out all crystallographic and analysis steps. Hong Tran performed the MalG R enzymatic assay, Sean Newmister and Qingyun Dan performed the substrate loading assay. Qingyun Dan, Sean Newmister, David Sherman and Janet Smith evaluated the data.

Appendix

Table 2.3. Data collection and refinement statistics.

	PhqB R
Space group	*I* 2 2 2
Unit cell parameters (Å, °)	80.84 90.10 125.49 90 90 90
Wavelength (Å)	1.033
Resolution (Å)	43.43 – 2.65 (2.74 – 2.65)
Completeness	1.00 (1.00)
Reflections	179882 (18233)
Unique reflections	13669 (1340)
Multiplicity	13.2 (13.6)
Mean I/σ	19.74 (1.03)
R_{meas}	0.074 (2.88)
$CC_{1/2}$	1 (0.66)
CC*	1 (0.89)
Reflections (working set)	13640
Reflections (test set)	685
R_{work}	0.247
R_{free}	0.270
No. of chains per AU	1
No. of cofactors	0
No. of non-hydrogen atoms	2467
Avg B-value (Å^2)	131.4
Ramachandran plot: favored/allowed/outliers (%)	95.4/3.8/0.8
RMSD bonds (Å)	0.016
RMSD angles (°)	1.70

Figure 2.S1. Structures of thiophenol substrate analogs (**28**, **29**) and product standards (**30** - **33**) synthesized by the Williams group.



Chapter III

MalB and MalE: Redundant prenyltransferases in malbrancheamide biosynthesis

3.1 Introduction

Fungal aromatic prenyltransferases are important enzymes in various kinds of fungi, generating a variety of prenylated secondary metabolites, such as indole alkaloids and polyketides. Prenylation has been shown to contribute greatly to structural diversity of natural products, many of which are potential candidates for anti-cancer, anti-fungal and anthelmintic drug design [101, 102]. Often prenylated compounds show distinct biological activities compared with their non-prenylated counterparts, and the prenyl group is essential for maintaining bioactivity [103-105]. For example, resveratrol suppresses the growth of *Staphylococcus aureaus* only when it is geranylated at the C4 position [103]. Since it can be difficult to synthesize these natural products chemically due to slow reaction rates and lack of prenylation site control, protein engineering and chemoenzymatic synthesis are attractive routes to produce these compounds for research purposes and drug design.

Aromatic prenyltransferases can be divided into two subgroups based on which carbon atom of the prenyl group is involved in bond formation during catalysis: regular prenyltransferases form a C1-aromatic carbon bond, while reverse

prenyltransferases catalyze formation of a C3-aromatic carbon bond (Fig. 3.1). Previously whole genome sequencing, gene annotation, and enzymatic characterization led to identification of dozens of aromatic prenyltransferases in different fungal genera, including *Aspergillus* and *Penicillium*. Often the prenyl acceptor is the indole ring of L-tryptophan, and the primary prenyl donor is dimethylallyl pyrophosphate (DMAPP) [103].

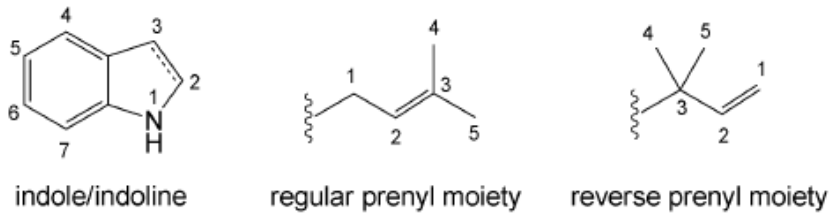


Figure 3.1. Structures of the prenyl acceptor indole, and the appended prenyl groups, showing the difference between "regular" and "reverse" prenylation products (figure from Hong Tran). Dimethylallyl pyrophosphate (DMAPP) is the prenyl donor.

One major feature of this type of modification is that the indole can be enzymatically prenylated at any available position (Fig. 3.2). For example, within *Aspergillus fumigatus*, FgaPT2 catalyzes C4 prenylation of L-tryptophan, while FtmPT1 prenylates brevianamide F at the C2 position [34, 107]. NotF in *Aspergillus* sp. MF297-2, shares 30% sequence identity with FtmPT1 and acts on the same substrate at the same carbon atom [13]. However, FtmPT1 forms a regular carbon-carbon bond, while NotF catalyzes a reverse prenyltransfer reaction only.

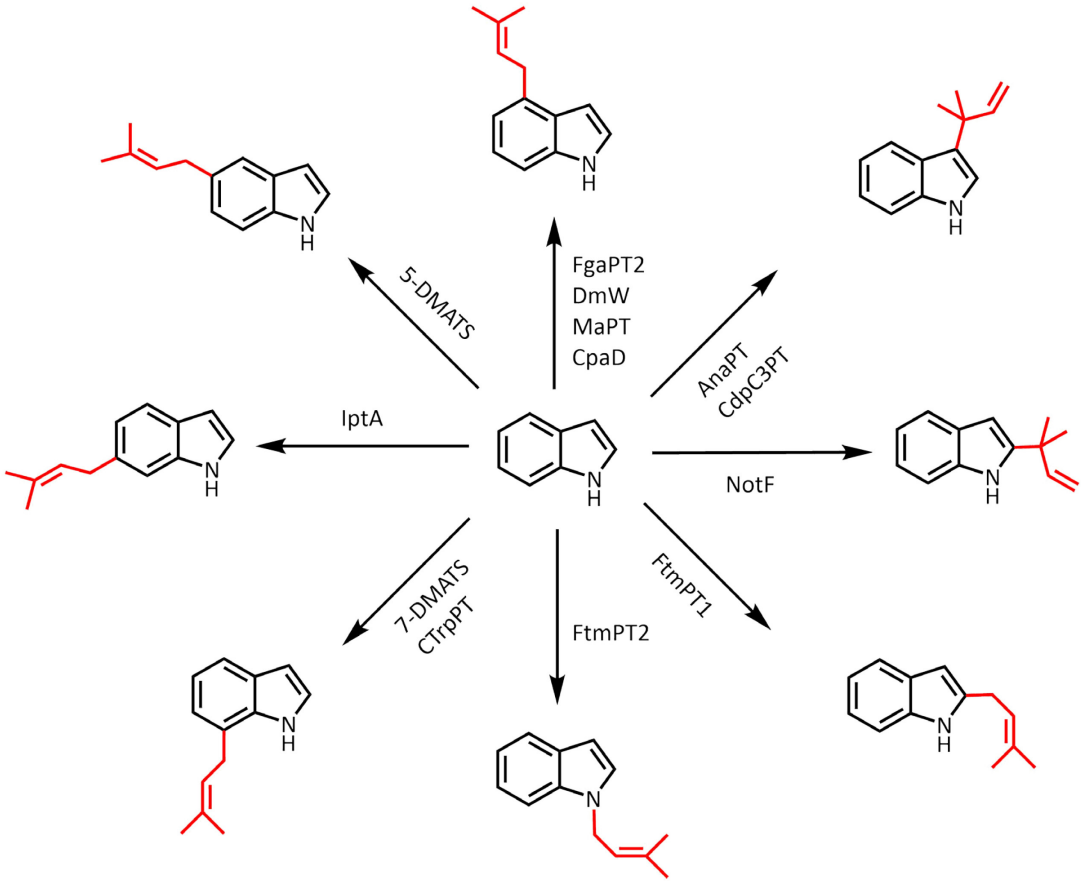


Figure 3.2. Examples of indole prenyltransferases with different prenylation sites. The added prenyl group is colored in red.

No sequence motif has been identified to distinguish any subtype of the reaction, and the structural basis for prenylation site control remains unclear. Also, despite broad substrate promiscuity, there are no known cases of fungal aromatic prenyltransferases that can catalyze both regular and reverse prenyltransfer reactions. To reveal the catalytic mechanism, as well as the structural difference between regular and reverse prenyltransferases, biochemical, structural and kinetic investigations have been performed on many characterized enzymes. For example, Fig. 3.3 shows FgaPT2, which catalyzes transfer of the prenyl moiety from DMAPP to the C4 position of L-tryptophan to produce dimethylallyl tryptophan (DMAT) in the initial step of ergot

alkaloid biosynthesis in *Aspergillus fumigatus*.

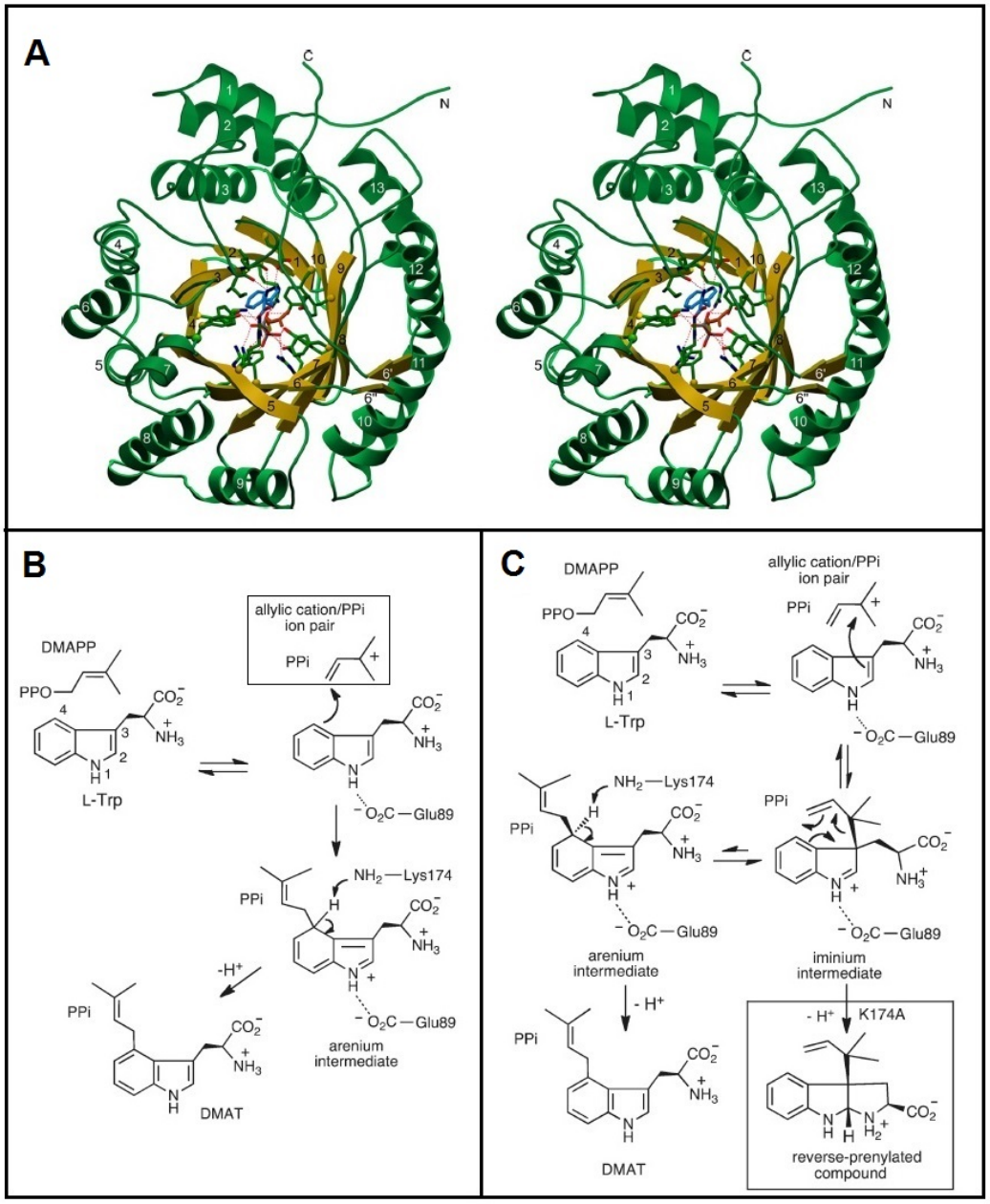


Figure 3.3. Structure and reaction mechanism of a typical regular prenyltransferase, FgaPT2. FgaPT2 catalyzes transfer of the prenyl moiety from DMAPP to the C4 position of L-tryptophan, producing dimethylallyl tryptophan (DMAT) as the initial step of ergot alkaloid biosynthesis in *Aspergillus fumigatus* [34, 109, 110]. A. 3D Stereo view of an FgaPT2 subunit (figure from reference [34]). B. Proposed electrophilic alkylation mechanism for FgaPT2-catalyzed reaction, including formation of a carbocation intermediate and an arenium ion intermediate (figure from reference [109]). C. Cope mechanism proposed for the FgaPT2-catalyzed reaction (figure from reference [110]). The K174A mutant produced DMAT and a reverse-prenylated product, suggesting a Cope rearrangement during the reaction.

For regular prenyltransferases, crystal structures of FgaPT2 and FtmPT1 from *Aspergillus fumigatus* have revealed a common architecture (Fig. 3.3A), represented by a novel β barrel topology (PT fold) [34, 36, 107, 108]. The inner barrel is formed by 10 anti-parallel β strands, and the outer barrel is filled with α helices, connected by αββα repeats [34]. For FgaPT2, two catalytic residues E89 and K174 were identified, as well as four tyrosine residues that are key for accommodating DMAPP. Multiple sequence alignments showed that most of these residues are conserved throughout this family of enzymes, and site-directed mutagenesis experiments have confirmed their necessity for maintaining enzymatic function.

As to catalytic mechanism, thorough biochemical and structural characterization suggests a common dissociative (S_N1) electrophilic alkylation mechanism (Fig. 3.3B) [34, 106]. According to this proposal, the first step of the prenyl transfer reaction is pyrophosphate dissociation from DMAPP, resulting in formation of a carbocation/pyrophosphate ion pair [109]. The carbocation intermediate then attacks the indole ring and forms an arenium intermediate. After deprotonation, the final product is generated and released from the enzyme. The observation of a positional isotope exchange when using 1-[^{18}O]-DMAPP in the reaction is consistent with the existence of an allylic cation intermediate, and kinetic isotope effect (KIE) measurements also agree with the proposed mechanism. Further studies on the FgaPT2 K174A mutant completed the proposed mechanism, indicating the presence of a Cope rearrangement [110] (Fig. 3.3C). In summary, although some details are still unclear, prior evidence indicates that regular prenyltransferases catalyze the

reaction via an electrophilic alkylation mechanism.

For reverse prenyltransferase, the crystal structure of CdpNPT from *Aspergillus fumigatus* shows a similar β barrel topology [108], and includes the same set of conserved residues as regular prenyltransferases, mutations of which greatly decrease the enzymatic activity. All evidence suggests a similar electrophilic alkylation mechanism, however, the existence of a carbocation intermediate has yet to be tested. It is hypothesized that the prenylation site control is at least partially provided by the active site arrangement, placing the target carbon atom closer to the prenyl donor.

In collaboration with the Sherman group, studies have been done on prenyltransferases in homologous pathways in the biosynthesis of four fungal indole alkaloids: (-)/(+)-notoamide, paraherquamide and malbrancheamide [37, 40]. Two prenyltransferases, NotC and NotF, in the (-)-notoamide pathway, have been characterized [37]. NotF reverse prenylates brevianamide F at the C2 position, while NotC catalyzes a regular C7 prenyltransfer reaction on 6-OH-deoxybrevianamide E (Fig. 1.5).

The malbrancheamide biosynthetic gene cluster encodes two putative prenyltransferases, MalB and MalE, of which MalE was shown to catalyze reverse prenyltransfer reaction of a dipeptidyl aldehyde analog (**34**) to produce the C2 prenylated product **35** (Fig. 3.S1). Based on sequence comparison to characterized prenyltransferases, MalB contains a full set of catalytic residues for a classic fungal aromatic prenyltransfer reaction, suggesting that it may be a functional prenyltransferase as well. However, only one prenylation step is required to

synthesize malbrancheamide, the final pathway product. It is unknown whether the redundancy is due to gene duplication or the fact that either MalB or MalE has an unidentified distinct function in the pathway. It is also unclear at which step of the biosynthetic pathway the prenyltransfer reaction occurs. At least three possibilities exist (Fig. 3.4) and can be tested with proposed substrates or analogs (dipeptidyl aldehyde, L-Trp or T-domain linked L-Trp or dipeptide) in hand. Characterization and comparative analysis of MalB and MalE have helped to solve the puzzle.

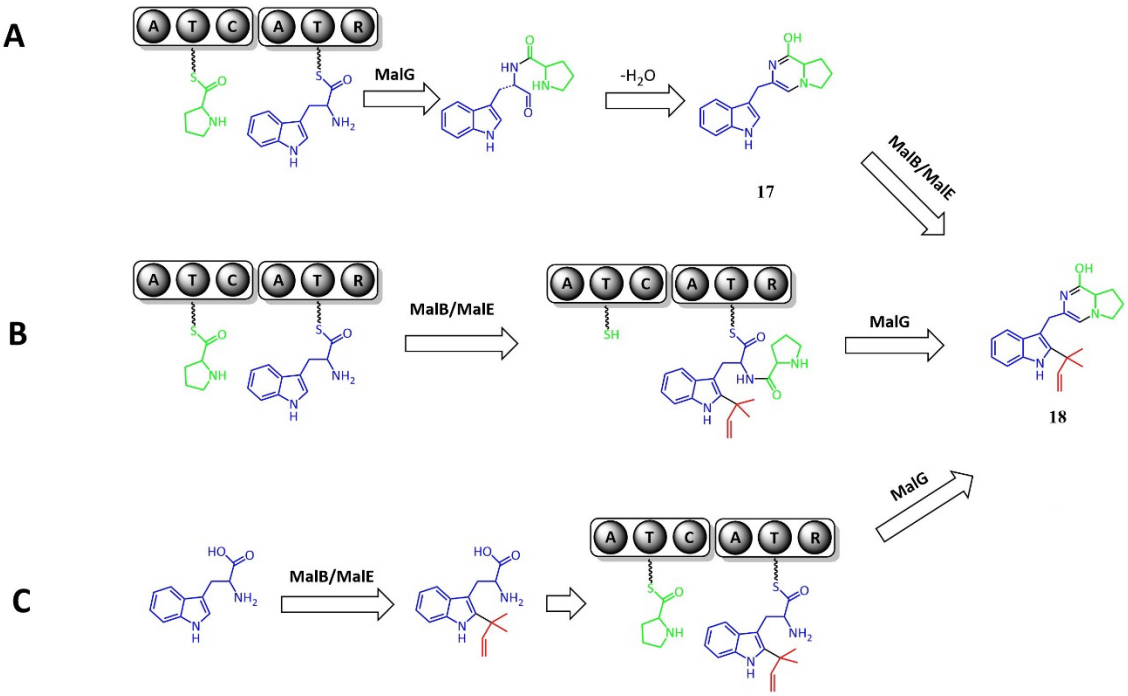


Figure 3.4. Possible schemes of prenyltransfer reactions catalyzed by MalB or MalE in malbrancheamide biosynthesis. A. The NRPS (MalG)-catalyzed reaction precedes the prenyltransfer reaction. B. The prenyltransfer reactions takes place when the substrate is loaded on the T domain of MalG. Prenylation could occur before or after the condensation (C) domain forms the Pro-Trp dipeptide. C. MalB or MalE prenylates L-Trp before the NRPS-catalyzed reaction.

3.2 Results

3.2.1 Biochemical characterization of MalB as an indole prenyltransferase

The MalB coding sequence was poorly expressed in *E. coli*, despite trials of various culture conditions, fusion partners and purification protocols. Heterologous expression in a baculovirus expression system finally yielded soluble MalB, which was purified to homogeneity. To test the enzymatic activity of MalB, a pathway intermediate (MalG product) candidate, L-Pro-L-Trp dipeptidyl alcohol (**36**), was used as substrate. MalB was shown to convert the dipeptidyl alcohol to a prenylated product, confirming its function as a prenyltransferase.

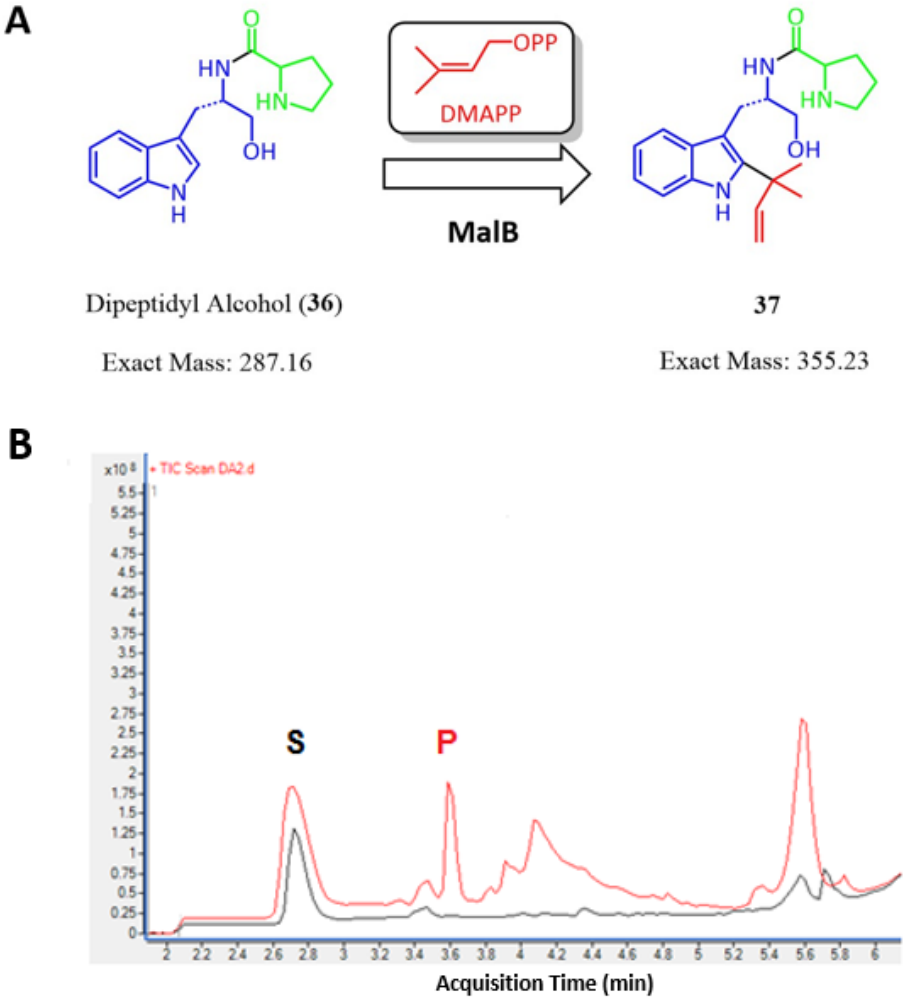


Figure 3.5. Enzymatic assay of MalB. A. Proposed reaction scheme of MalB converting **36** to **37**. B. HPLC elution profile of the reaction (black: no enzyme control; red: reaction with MalB. S stands for substrate, and P is the product peak.). Detected mass spectra of the substrate and the product are shown in Fig. 3.S3.

3.2.2 Substrate determination of MalB and MalE

A hypothesis that the prenyltransfer reaction occurs on L-Trp (Fig. 3.4.C) before it is consumed by MalG, the non-ribosomal peptide synthetase (NRPS), was tested. Neither MalB nor MalE prenylated L-Trp, indicating that the prenyltransfer reaction is not upstream of L-Trp loading onto the NRPS T domain (Fig. 3.S5).

The NRPS product in the pathway was proposed to be either L-Pro-L-Trp dipeptidyl aldehyde or dipeptidyl alcohol (**36**). MalE has been shown to reverse prenylate the C2

position of a dipeptidyl aldehyde analog **34** (Fig. 3.S1). Here, the dipeptidyl alcohol (**36**) was also tested, and MalE was shown to prenylate **36**, generating a prenylated product (Fig. 3.6). Since MalE is capable of prenylating both the aldehyde analog and the alcohol, the result provides no hint on the exact malbrancheamide NRPS product.

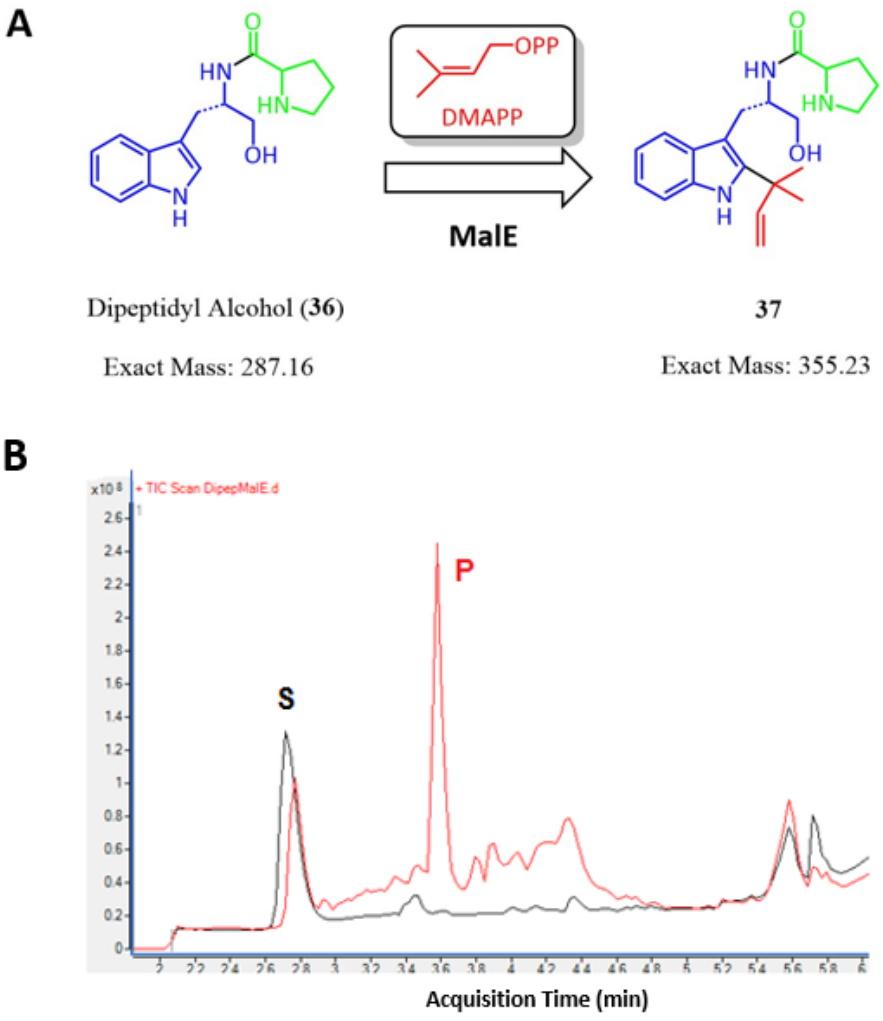


Figure 3.6. Prenyltransfer reaction of dipeptidyl alcohol (**36**) catalyzed by MalE. A. The proposed reaction scheme; B. HPLC elution profile of the reaction (black: no enzyme control; red: reaction with enzyme. S stands for substrate, and P is the product peak.). Detected mass spectra of the substrate and the product are shown in Fig. 3.S4.

3.2.3 Comparative analysis of MalB and MalE

It was shown by Hong Tran that MalE catalyzed the C2 reverse prenyltransfer reaction on either the dipeptidyl aldehyde analog (**34**) or brevianamide F, the natural substrate of NotF. Incubation of **34** with MalE and DMAPP generated a product with a mass consistent with **35** (Fig. 3.S1). The structure was confirmed by COSY NMR analysis, demonstrating that MalE reverse prenylates L-Pro-L-Trp dipeptidyl analogs at the C2 position.

To investigate whether MalB and MalE catalyze the same reaction, I set up separate reactions for both enzymes. The products were co-injected onto LC/MS. The result clearly showed that the MalB and MalE prenylated products co-eluted, confirming that MalB catalyzes the same C2 reverse prenyltransfer reaction as MalE (Fig. 3.7).

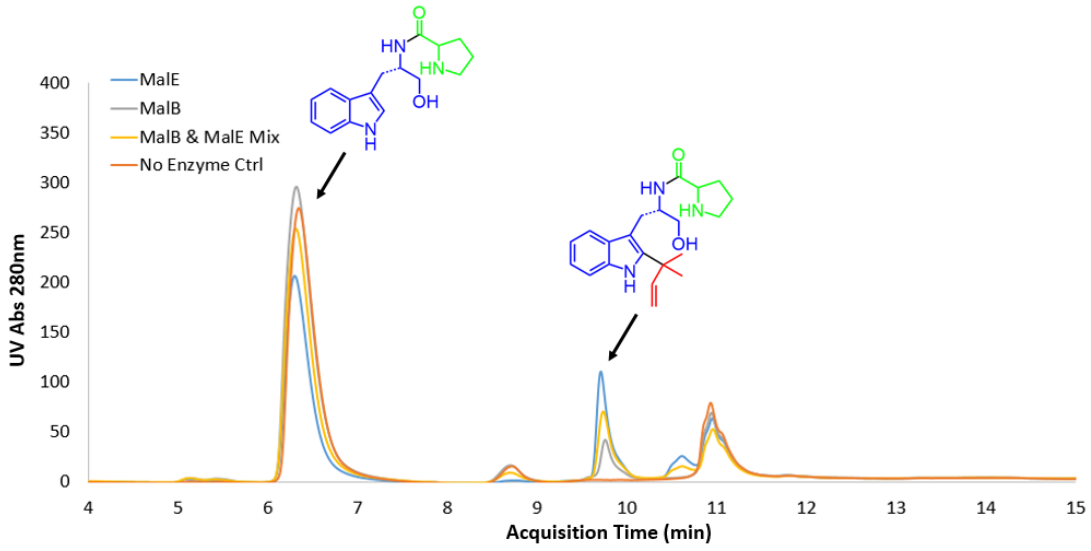


Figure 3.7. HPLC trace of MalB and MalE reactions. The dipeptidyl alcohol (**36**) was used as a substrate, and the MalB and MalE products co-eluted.

3.2.4 Genetic disruption in the homologous paraherquamide pathway

In the paraherquamide pathway (Fig. 1.4), three putative prenyltransferases PhqA, PhqI and PhqJ are encoded by the gene cluster, even though paraherquamide A, the final pathway product, has only two prenyl groups. *In vivo* genetic disruption of *phqI* and *phqJ* were performed by the Sherman group to address this question (personal communication from Fengan Yu). The *phqJ* deletion mutant profile led to an accumulation of preparaherquamide, a pathway intermediate, suggesting that PhqJ is involved in the biosynthetic pathway, and its catalyzed reaction is downstream of preparaherquamide production.

PqhI on the other hand, is essential for the early steps of the pathway: Mutation of the coding sequence led to detection of no pathway intermediate or product. Previously PhqI was proposed to be the MalB homolog in the paraherquamide pathway based on Softberry gene annotation [136], due to the prediction that neither *malB* nor *phqI* contained an intron. However later sequencing results for *phqI* clearly showed the existence of a short (60-bp) intron, which is commonly found in genes that encode fungal indole prenyltransferases. This left *malB* as the only intron-free prenyltransferase gene among all identified fungal indole alkaloid pathways. It remains unclear whether this feature of *malB* is suggestive of an independent evolutionary origin or functional uniqueness. MalB has the highest sequence identity (34%) to PhqA in the paraherquamide pathway. Genetic disruption of *phqA* will be tested, which may provide insight into the function of the extra prenyltransferase in the malbrancheamide and paraherquamide pathways.

3.3 Discussion

Previously it was unknown which step occurs first in the malbrancheamide pathway: the NRPS-catalyzed dipeptide formation, or the prenyltransfer reaction. Studies on the (-)-notoamide pathway showed that both prenylation steps took place after formation of brevianamide F, the NRPS product. NotF, the first prenyltransferase in the pathway, possesses strict substrate selectivity and acts only on brevianamide F, indicating a reaction scheme with an NRPS reaction followed by a prenyltransfer reaction.

However, two major differences exist between the (-)-notoamide and the malbrancheamide pathways. First, only one prenyltransfer reaction is proposed to occur in malbrancheamide biosynthesis, compared with two confirmed prenylation steps in the (-)-notoamide pathway. Secondly, the domain composition of notoamide NRPS (NotE) and malbrancheamide NRPS (MalG) differs. The terminal domain of NotE is a condensation domain, producing brevianamide F as the final product; while MalG ends with a reductase domain, which catalyzes reduction and hydrolysis. The NRPS product of the malbrancheamide pathway is either a dipeptidyl aldehyde or dipeptidyl alcohol, with different structural features from brevianamide F.

Thus, enzymatic tests of MalB/MalE with different putative substrates (L-Trp, **34** and **36**) are necessary to identify the correct reaction sequence of the malbrancheamide pathway. Our data shows that both MalB and MalE accept dipeptidyl analogs but not L-Trp, suggesting that the prenyltransfer reaction does not precede L-Trp loading onto

the NRPS T domain. It is unknown whether prenylation can occur when the dipeptide is attached to the NRPS, or the malbrancheamide pathway adopts the same scheme as the (-)-notoamide pathway.

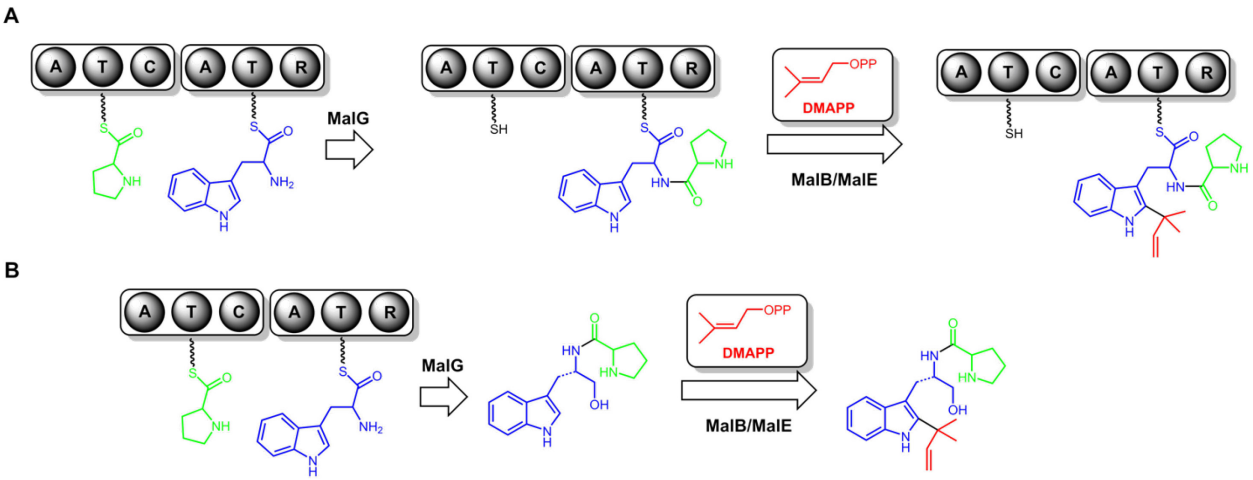


Figure 3.8. Two proposed mechanisms for the early steps of malbrancheamide biosynthesis. A. The prenyltransfer reaction occurs when the dipeptidyl intermediate is attached to the NRPS; B. The prenyltransfer reaction occurs after the NRPS-catalyzed reaction.

Another question we attempted to address is the identity of the NRPS product of malbrancheamide pathway. Experimental data shows that both L-Pro-L-Trp alcohol (**36**) and the dipeptidyl aldehyde analog (**34**) can be consumed by MalE (Fig. 3.6 and 3.S1), providing no hint on the natural product. Direct characterization of MalG and its terminal reductase domain is required to identify the structure of the NRPS product.

Another surprising discovery is that MalB and MalE catalyze the same reaction, which is unprecedented in any related fungal indole alkaloid pathway. The two enzymes share 25% sequence identity, suggesting they may result from gene

duplication, yet whether one or both enzymes has an unidentified function beyond prenyltransfer requires further investigation. Interestingly a similar puzzle exists in the homologous paraherquamide pathway, where two prenyl groups are built into the final pathway product and three putative prenyltransferases, PhqA, PqhI and PhqJ, were annotated within its gene cluster. Genetic deletion of *phqJ* indicates its function being downstream of preparaherquamide formation. However, enzymatic assays of PhqJ *in vitro* showed no activity on the putative substrate preparaherquamide (personal communication from Sean Newmister). PhqI was shown to reverse prenylate the dipeptidyl alcohol at the C2 position, and its deletion led to production of no pathway product. Taking all available data into consideration, PhqA and PhqI are proposed to be MalB and MalE homologs in the paraherquamide pathway, and genetic disruption of *phqA* may provide important insights into elucidating MalB and MalE functions in the future.

3.4 Materials and Methods

3.4.1 Materials and strains

The dipeptide aldehyde analog (**34**) was synthesized by the Williams group at Colorado State University. L-Pro-L-Trp dipeptidyl alcohol (**36**) was synthesized by Sean Newmister in the Sherman group. L-tryptophan and DMAPP were purchased commercially. *E. coli* XL1-Blue and DH10 Bac cells were used for plasmid storage, *E. coli* BL21(DE3) pRare and Insect High-five cells (BTI-TN-5B1-4) were used for protein overexpression.

3.4.2 Cloning of *malB* and *malE*

The cDNA library of *Malbranchea aurantiaca* RRC1813 was generated by Shengying Li and Hong Tran in the Sherman group. For cloning of *malE*, PCR was used to amplify the cDNA template, followed by a ligation independent cloning (LIC) procedure [94, 95] to insert the gene into the pMCSG7 vector. PCR primers are listed in Table 3.1. The plasmid (pMCSG7-MalE) was then transformed into *E. coli* XL1-Blue cells for plasmid storage and harvest.

W. Clay Brown in the Smith group sub-cloned *malB* into a baculovirus expression vector that encoded an N-terminal His tag and the fusion partner maltose binding protein (MBP). LIC was carried out to insert *malB* into the vector. The plasmid was then transformed to DH10 Bac cells (Invitrogen). Competent cells (20 μL) were incubated on ice in the presence of the DNA for 30-60 minutes. SOC media (80 μL) was added to each tube and incubated at 37 °C for 3 hours. Each sample was plated

on a Q-tray well containing kanamycin, tetracycline, gentamycin, IPTG and Bluo-Gal. The tray was incubated for 48 hours at 37 °C. Two white colonies from each well were picked and patch streaked onto an indicator plate, which was incubated overnight at 37 °C. Colonies that remained white were considered positive. Patch colonies were used to inoculate overnight cultures for bacmid preps. Cultures were grown in 5 mL of LB with kanamycin in a 24-well block shaking at 300 RPM at 37 °C. Cells were pelleted and then subjected to alkaline lysis. Samples were spun in a microfuge for 10 minutes at top speed (13000 rpm). An 800 µL aliquot was pipeted off and added to 800 µL of isopropanol in a fresh tube. Samples were mixed by inversion and then spun for 30 minutes at top speed in a microfuge. The supernatant was decanted and the tubes were air-dried. The pellets were resuspended in 40 µL of sterile water. A 7 µL aliquot of each bacmid DNA was mixed with 150 µL of media and 10 µL of transfection reagent in a well of a sterile 24-well block. These were allowed to stand for 30 minutes at room temperature in a biosafety cabinet. High-five cells at 2 x 10^6 cells/ml (850 µL) were added to each well. The block was covered with a sterile breather film and incubated at 27 °C with shaking at 120 RPM. After 4 hours, an additional 3.5 mL of media with 10% FBS was added, the block was resealed and incubated at 27 °C with shaking at 300 RPM for 6-7 days. The block was spun at 1000 x g at 4 °C for 10 minutes. In a biosafety cabinet, the media was drawn off and placed in a cryovial and stored at 4 °C until used for infection.

Table 3.1. Primers used for cloing of *malB* and *malE*

Gene	Oligo Sequence (5' -> 3')
malB (Forward)	TACTTCCAATCCAATGCCATGCCTTCACAAAGCCCATATCAT
malB (Reverse)	TTATCCACTTCCAATGCTACTAGTAAGCTGACAAGTTGGTTCG
malE (Forward)	TACTTCCAATCCAATGCCATGACAGCAGGTCCGATGG
malE (Reverse)	TTATCCACTTCCAATGCTATCAAGCACCATCTCCTTGACC

3.4.3 Overexpression and purification

malE was overexpressed in *E. coli* BL21(DE3) pRare. The cells were transformed with pMCSG7-MalE and grown in Terrific Broth medium (30 µg/ml ampicillin and 100 µg/ml spectinomycin added) to a O.D. 600 = ~1.0 at 37 °C. The culture was then transferred to 20 °C over 1 hour, and induced with 0.4mM IPTG. After 18 - 20 hours of incubation (20°C, 225 rpm shake), cells were harvested by centrifugation and the cell pellet was stored at -20°C. For purification, the cell pellet was resuspended with lysis buffer (10% v/v glycerol, 500 mM NaCl, 20 mM Tris buffer pH 7.9, 20 mM imidazole, 5 mg lysozyme, 2 mg DNAse and 1 mM $MgCl_2$), and vortexed to mix for 30 min. Sonication and high speed centrifugation (16000 rpm, 30 min) were applied to obtain the lysate soluble fraction. The soluble fraction was filtered and loaded on a GE Ni-NTA HisTrap column (Ni-NTA buffer: 10% glycerol, 500 mM NaCl, 20 mM imidazole pH 7.9, 20 mM Tris pH 7.9; flow rate: 3 mL/min), and was eluted with a 20 – 600 mM imidazole gradient in Ni-NTA buffer (flow rate: 3 mL/min; gradient time:

12 min). Fractions containing MalE were pooled and incubated with TEV protease in a 1:50 w/w ratio at 20 °C for 2 hours to remove the N-terminal His-tag. The tag-free protein was dialyzed overnight at 4 °C into 10% glycerol, 2 mM DTT, 500 mM NaCl, 20 mM Tris pH 7.9, and collected as the flow-through from a Ni-NTA HisTrap column (Ni-NTA buffer; flow rate: 3 mL/min). Further homogeneity was achieved by size-exclusion chromatography (GE HiLoad 16/60 Superdex 200 prep grade column equilibrated with 10% v/v glycerol, 500 mM NaCl, 20 mM Tris pH 7.9; flow rate: 0.5 mL/min). SDS-PAGE was used to assess protein homogeneity, confirming >95% purity. Yield of purified protein was 50 mg MalE per L of culture.

malB was overexpressed in High-five cells. 1L volumes of High-five cells were seeded at 2 x 10^6 cells/ml in Insect X-press media (Lonza) in 2.8 L Fernbach flasks. These were infected at an MOI of 2. The flasks were incubated at 20 °C with shaking at 140 rpm for 72 hours. The cells are harvested by centrifugation at 1000 g and 4 °C for 40 minutes in 1 L bottles. The media was decanted and the pellets were removed from the bottles with a plastic spatula and placed into Zip-loc freezer bags. The pellets were stored at -80 °C until purification. For purification, pellets were resuspended in 100 mL of lysis buffer (to 1 ml PBS add 250 µL of insect Pop-culture (Novagen), $MgCl_2$ to 2 mM final, 10 µL benzonase (Novagen) and Roche EDTA free protease inhibitor tablet) and incubated at 25 °C with shaking for 30 min for lysis. The sample was centrifuged at 13000 rpm for 10 min to collect the soluble fraction. The soluble fraction was filtered and loaded on a GE Ni-NTA HisTrap column (Ni-NTA buffer: 10% glycerol, 500 mM NaCl, 20 mM imidazole pH 7.9, 20 mM Tris pH

7.9; flow rate: 3 mL/min), and was eluted with a 20 – 600 mM imidazole gradient in Ni-NTA buffer (flow rate: 3 mL/min; gradient time: 12 min). Fractions containing MalB were pooled and incubated with TEV protease in a 1:50 w/w ratio at 20 °C for 2 hours to remove the N-terminal His-MBP-tag. The tag-free protein was dialyzed overnight at 4 °C into 10% glycerol, 2 mM DTT, 500 mM NaCl, 20 mM Tris pH 7.9, and collected as the flow-through from a Ni-NTA HisTrap column (Ni-NTA buffer; flow rate: 3 mL/min). Further homogeneity was achieved by size-exclusion chromatography (GE HiLoad 16/60 Superdex 200 prep grade column equilibrated with 10% v/v glycerol, 50 mM NaCl, 20 mM Tris pH 7.9; flow rate: 0.5 mL/min). SDS-PAGE was used to assess protein homogeneity, confirming >95% purity. Yield of purified protein was 5 mg MalB per L of culture.

3.4.4 Characterization of enzymatic activities

Enzyme (50 μM) and substrate (500 μM) were added to 10% v/v glycerol, 50 mM NaCl, 50 mM HEPES pH 7.5, 1 mM DTT, 1 mM EDTA. The reaction was initiated by addition of 250 μM DMAPP. The reaction mix was incubated at 25°C for 6 hours, and quenched with 50% v/v methanol. Centrifugation was used to remove denatured protein, and the soluble fraction was collected and analyzed by LC/MS (Phenomenex Kinetix reverse-phase C18 column (40 mm × 2.1 mm, 2.6 μm), Buffer A: 0.2% v/v formic acid in water, Buffer B: 0.2% v/v formic acid in acetonitrile. HPLC protocol: 5% Buffer A for 2 min, 5-100% Buffer B gradient for 4 min, 100% Buffer B for 2 min. Flow rate: 0.5 mL/min.).

3.4.5 Protein mass spectrometry

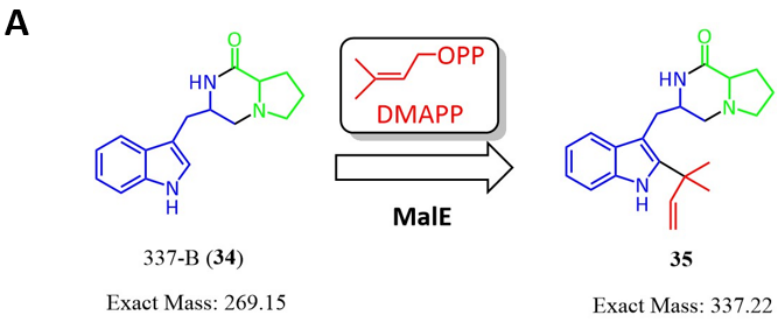
Protein (5 μM) was added to buffer (10% v/v glycerol, 20 mM Tris pH 7.9). LC/MS was used to identify the mass of the protein sample (Aeris widepore C4 column (3.6 μm, 50 × 2.10 mm), Buffer A: 0.2% v/v formic acid in water, Buffer B: 0.2% v/v formic acid in acetonitrile. HPLC protocol: 5% Buffer A for 2 min, 5-100% Buffer B gradient for 4 min, 100% Buffer B for 2 min. Flow rate: 0.5 mL/min.).

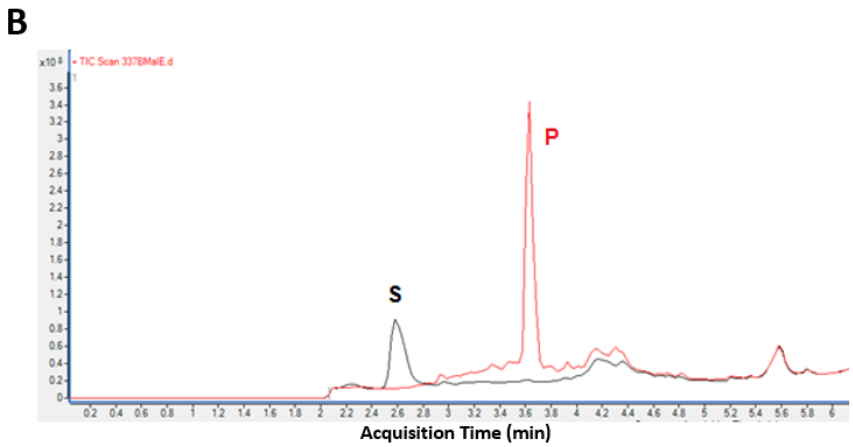
Author Contributions

Qingyun Dan, Sean Newmister, Hong Tran, Robert Williams, David Sherman and Janet Smith contributed to the experimental design. Shengying Li and Hong Tran generated the cDNA library. Qingyun Dan cloned and expressed *malE*. Hong Tran performed protein purification, enzymatic assays and mass spectrometry. W. Clay Brown cloned and expressed *malB*, and Qingyun Dan performed protein purification, enzymatic assays and protein mass spectrometry. Sean Newmister synthesized dipeptidyl alcohol (**36**), and James Sunderhaus synthesized the dipeptidyl aldehyde analog (**34**). Fengan Yu performed genetic disruption of the paraherquamide pathway. Qingyun Dan, Sean Newmister, David Sherman and Janet Smith evaluated the data.

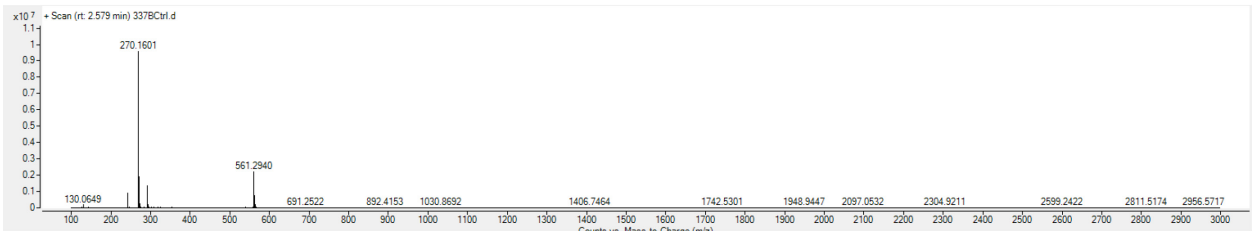
Appendix

Figure 3.S1. Enzymatic assay test of MalE. A. The reaction scheme of MalE, catalyzing **34** to **35**. The HPLC elution profile (B), detected substrate mass (C, observed mass: 270.16 Da, calculated mass: 270.15 Da) and product mass (D, observed mass: 338.23 Da, calculated mass: 338.22 Da) are presented as below. (Data and figure from Hong Tran in the Sherman group)





C



D

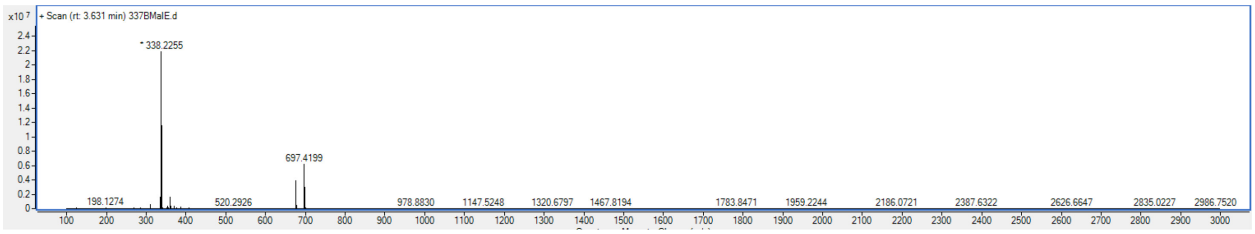

x10 7
+ Scan (rt: 3.631 min) 337BMalE.d
338.2255
697.4199
198.1274
520.2926
978.8830
1147.5248
1320.6797
1467.8194
1783.8471
1959.2244
2186.0721
2387.6322
2626.6647
2835.0227
2986.7520

Figure 3.S2. Deconvoluted mass spectrum of MalB. Soluble MalB can only be obtained when expressed in insect cells. To investigate whether post-translational modification was present, intact protein mass spectrometry was applied. The experimental mass of the MalB subunit was determined to be 41839.8 Da, 342.8 Da higher than the calculated mass. The mass difference does not match with any common single type of post-translational modification, suggesting a possibility of multiple post-translational modifications on the protein.

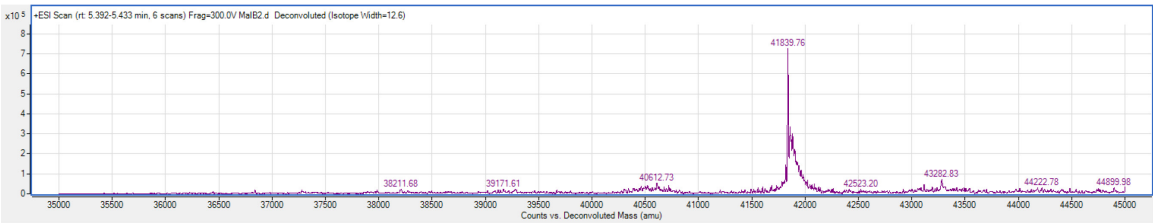
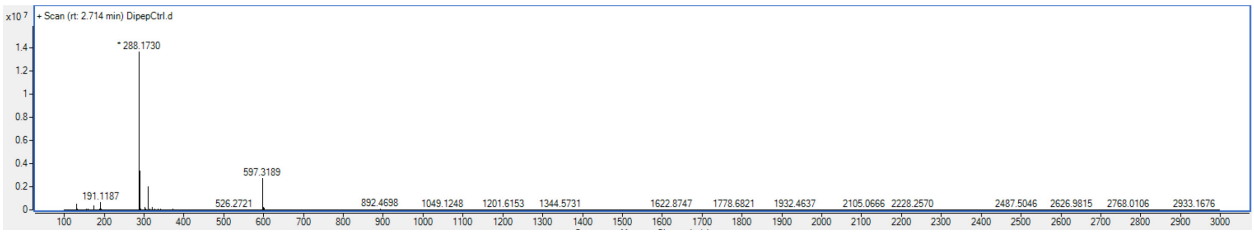


Figure 3.S3. Detected mass spectrum of the MalB reaction, converting **36** to **37**. The reaction scheme and HPLC profile of the reaction are shown in Fig. 3.5. A. Detected substrate mass: 288.17 Da (calculated mass: 288.16 Da). B. Detected product mass: 356.24 (calculated mass: 356.23 Da).

A



B

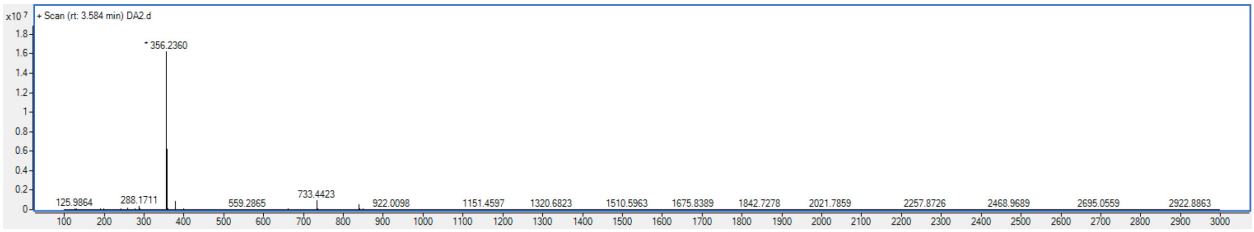
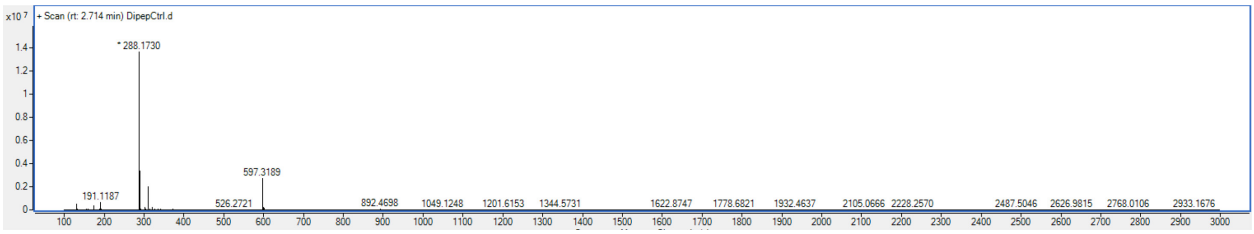


Figure 3.S4. Detected mass spectrum of the MalE reaction, converting **36** to **37**. The reaction scheme and HPLC profile of the reaction are shown in Fig. 3.6. A. Detected substrate mass: 288.17 Da (calculated mass: 288.16 Da). B. Detected product mass: 356.24 (calculated mass: 356.23 Da).

A



B

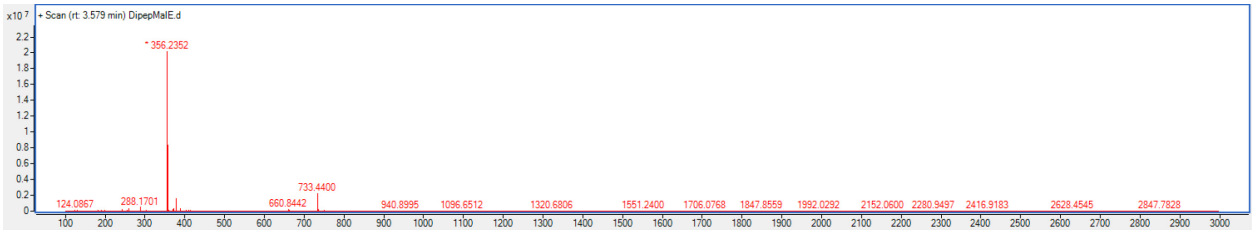
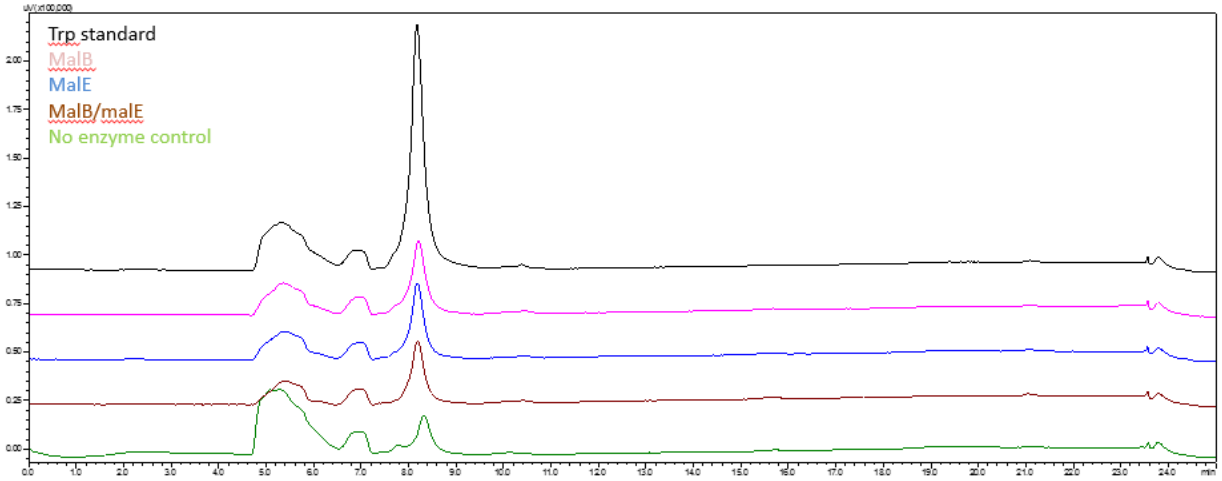
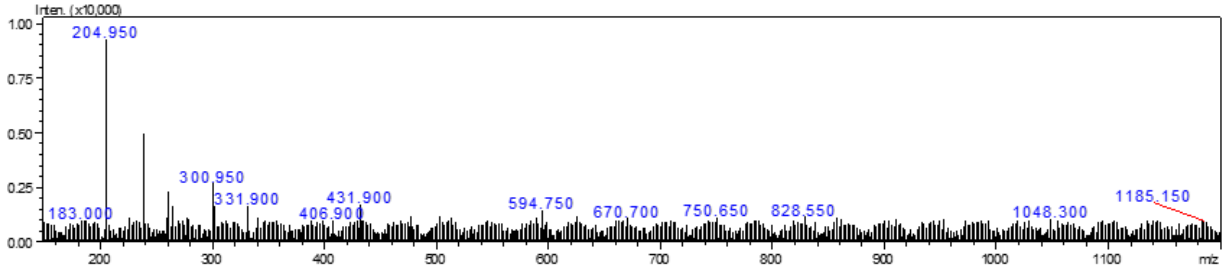


Figure 3.S5. Enzymatic activity test of MalB/MalE with L-Trp. A. HPLC elution profile of the reaction. B. Detected substrate mass (observed mass: 204.95 Da, calculated mass: 205.09 Da). No prenylated product was detected (calculated mass: 273.15 Da).

A



B



Chapter IV

MalC: A protein of unknown function in the malbrancheamide pathway

4.1 Introduction

4.1.1 Short-chain dehydrogenases/reductases

Short-chain dehydrogenases/reductases (SDRs) are an enzyme family that exists in almost all forms of life, catalyzing oxidative-reductive reactions to form various metabolites, hormones, etc. [111, 112]. Most SDRs are classified as oxidoreductases, with a few exceptions of lyses and isomerases [112]. So far over 3000 genes have been annotated as SDRs in databases, and in the human genome, more than 60 genes have been identified or proposed to be SDRs [113, 114]. The number goes up to ~150 in certain plant species, implying their abundance and importance. In this class of enzymes, many have been reported to act on aliphatic alcohols, steroids, prostaglandins and xenobiotics [112], suggesting great potential in chemoenzymatic synthesis of related compounds. For sustainable chemoenzymatic synthesis, SDRs are also useful for cofactor regeneration, which can be costly if purchased commercially [116]. In fact, SDRs are now widely used for synthesis of alcohol products in industry [115]. SDRs are also interesting pharmacological drug targets based on their physiological functions. Some SDRs, such as hydroxysteroid dehydrogenases, affect the cellular level of their hormone products and their malfunction may lead to

a series of metabolic syndromes [117]. Developing drugs that target related SDRs are of great interest. Other SDRs in microbes have been shown to participate in nutrient synthesis and growth regulation, studies of which may inspire the design of novel antibiotics [118 – 120].

Short-chain dehydrogenases generally have low sequence identity (15-30%) to other family members, yet all of them share a common structure, featured by a nucleotide binding subdomain typical of the Rossmann fold, a parallel β sheet flanked by α helices, and a C-terminal substrate binding subdomain/site [111-113] (Fig. 4.1). The substrate binding site is often covered by a flexible lid, which becomes ordered upon cofactor binding. Based on sequence lengths and cofactor preferences, SDRs can be further divided into two subgroups: classical SDRs and extended SDRs. Classical SDRs usually contain ~250 amino acids, while extended SDRs are often ~100 amino acids longer. For SDRs, NADH or NADPH is the most common cofactor [121].

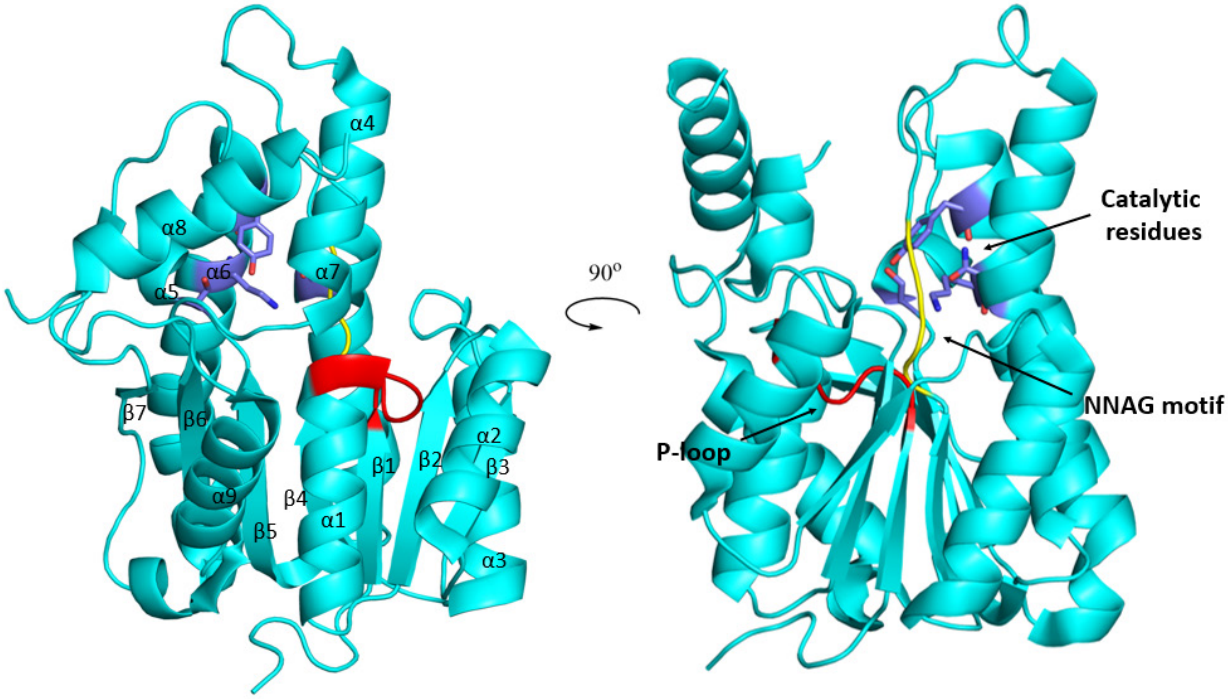


Figure 4.1. Structure of a classic SDR (RasADH from *Ralstonia* sp. DSM 6428, PDB ID: 4BMS) [126-128]. Conserved motifs are colored differently (TGX_3GXG motif or the P-loop in red, NNAG motif in yellow), and active site residues are colored in purple.

Sequence analysis, structural investigation and site-directed mutagenesis have identified conserved sequence motifs and residues that are characteristic of this family of enzymes [122-125] (Table 4.1). An N-terminal TGX_3GXG motif forms part of the nucleotide binding domain (the P-loop), and is responsible for coordinating the phosphate group of the cofactor [113]. A conserved NNAG motif follows β4, the central β strand of the β sheet and stabilizes the β sheet by a hydrogen bonding network [122]. Three conserved active site residues have been identified, forming a "Ser-Tyr-Lys" catalytic triad [124, 125] (Fig. 2.5.C). The Lys forms a hydrogen bond with a ribose hydroxyl group of the cofactor and lowers the pK_a of the catalytic Tyr [124]. The Tyr residue is nearly invariant throughout the entire family. During the

reaction, it serves as a proton donor, and the C4 atom of nicotinamide transfers a hydride to reduce the double bond of substrates. The reaction can be catalyzed in either direction. The Ser is hydrogen bonded to the hydroxyl or carbonyl group of the substrate, orienting and stabilizing it to facilitate catalysis. Often an Asn residue near the active site is also conserved, which occupies a water molecule, which further forms a hydrogen bond with the Lys. Thus, the catalytic residues are called "Asn-Ser-Tyr-Lys" catalytic tetrads. In the C-terminal substrate binding region of SDRs, the sequence identity drastically decreases as each SDR evolved to bind a particular substrate. Most SDRs have strict substrate selectivity and stereo specificity.

Table 4.1. Sequence motifs and catalytic residues in SDRs

Motif	Position	Function
TGX$_3$GXG	12-19	P-loop, cofactor phosphate binding
NNAG	86-89	Stabilization of central β-sheet
N-S-Y-K	111, 138 ,151, 155	Active site residues

4.1.2 SDR in the malbrancheamide pathway

In the malbrancheamide biosynthetic pathway, MalC is the only gene product annotated as an SDR (Table 2.1) [40]. The pathway requires only one reductive step, which is expected to be delivered by the terminal reductase domain of MalG, an NRPS, to form premalbrancheamide, the proposed Diels-Alder product. However, the proposed reductive NRPS product, L-Pro-L-Trp dipeptidyl aldehyde, was shown to be unstable *in vitro* (Fig. 2.7). Based on this instability, Sean Newmister proposed that

the NRPS reductase domain catalyzes a 4-electron reduction reaction to form a stable L-Pro-L-Trp dipeptidyl alcohol, which can be prenylated and subsequently oxidized back to an aldehyde at later steps (Fig. 4.2). MalC was proposed to catalyze the oxidation reaction, producing a dipeptidyl aldehyde or premalbrancheamide. In collaboration with the Sherman group, we have shown that the dipeptidyl alcohol can be prenylated by MalB or MalE, the prenyltransferases in the malbrancheamide pathway (Fig. 3.5 and 3.6). Biochemical and structural characterization of MalC will be a test of the proposal, as well as the possible function of MalC in the pathway.

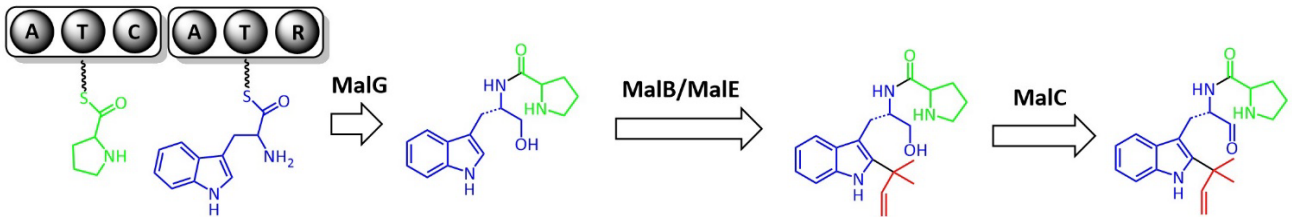


Figure 4.2. Proposed function of MalC in the early steps of malbrancheamide biosynthesis. Prenylated L-Pro-L-Trp dipeptidyl alcohol is the proposed native substrate for the MalC-catalyzed reaction.

4.2 Results

4.2.1 Test of MalC activity as short-chain dehydrogenase

To test the hypothesis that MalC functions as short-chain dehydrogenase and oxidizes a prenylated dipeptide alcohol to produce an aldehyde, L-Pro-L-Trp dipeptidyl alcohol was incubated with a MalB, MalE and MalC enzyme mixture. Using dimethylallyl pyrophosphate (DMAPP) as the prenyl donor, MalB and MalE prenylated the dipeptidyl alcohol; however, no prenylated dipeptidyl aldehyde was

detected, in the presence of MalC and NAD as cofactor (Fig. 4.3). Trials of NADP as a cofactor candidate and buffer screens (HEPES 7.5, HEPES 8 and Tris 9) yielded the same result. The result indicates that MalC does not accept either the dipeptidyl alcohol or its prenylated product as a substrate, and MalC shows no activity of short-chain dehydrogenase with the tested substrates. Its potential function or substrate in the malbrancheamide pathway requires further investigation.

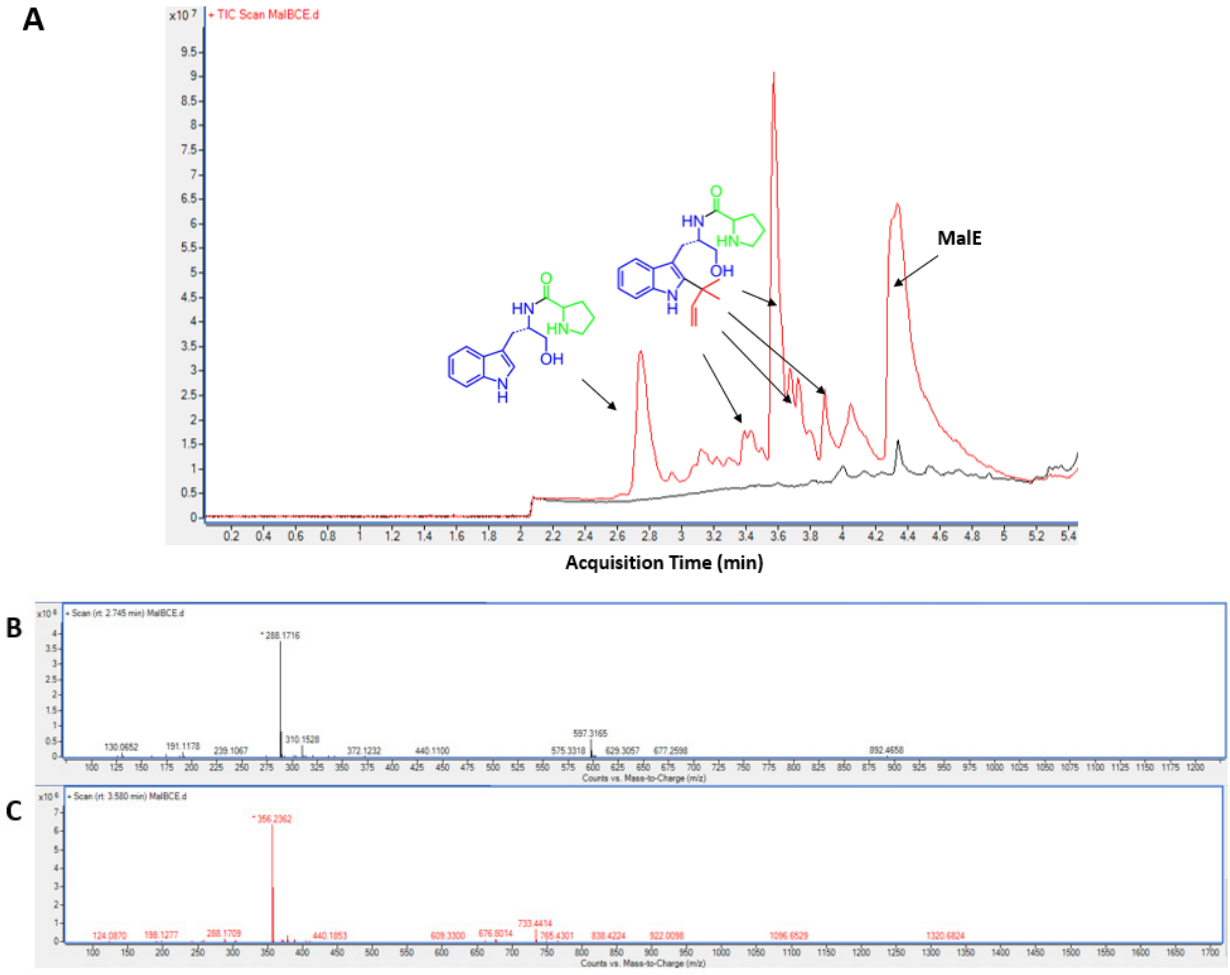


Figure 4.3. Test of potential oxidative activity of MalC. A. HPLC elution profile of the reaction of L-Pro-L-Trp dipeptidyl alcohol with a MalB, MalE and MalC enzyme mix (black: buffer control; red: reaction with enzyme mix). B. Mass spectrum of the substrate (observed mass: 288.17 Da; calculated mass: 288.16 Da). C. Mass spectrum

of detected product (observed mass: 356.24 Da; calculated mass of prenylated dipeptidyl alcohol: 356.23 Da). No masses were detected for the prenylated dipeptidyl aldehyde (336.20 Da) or premalbrancheamide (335.19 Da), the proposed Diels-Alder product.

4.2.2 MalC structure disproves its function as a short-chain dehydrogenase

To help understand the function of MalC, a 1.6 Å crystal structure was solved (Table 4.3). MalC is a tetramer in solution, as shown by size-exclusion chromatography and multi-angle light scattering (Fig. 4.S2). Each subunit interacts with other two subunits, forming two distinct binding interfaces (Fig. 4.4). The tetrameric interfaces are stabilized primarily via hydrophobic interactions, and tetrameric association is commonly observed in other SDR structures.

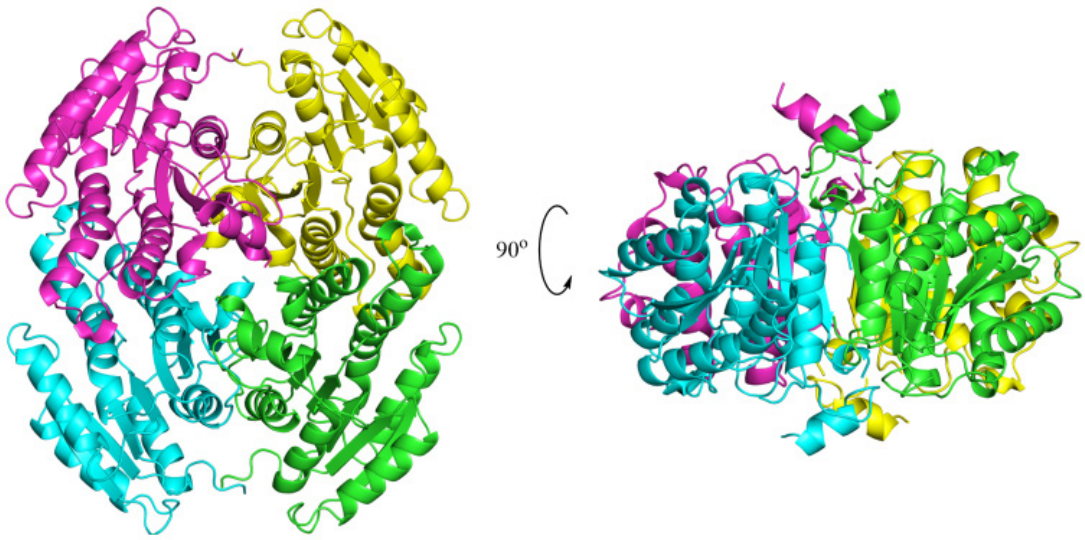


Figure 4.4. Structure of the MalC tetramer. Each subunit (colored differently) interacts with two other subunits in a dimer of dimers organization with D2 molecular symmetry.

The structure has a classical SDR fold with a Rossmann-fold nucleotide-binding subdomain, and a C-terminal substrate binding region (Fig. 4.5). The proposed substrate binding site is covered by a lid region, which is composed of α8 and an ~30 amino acid loop preceding α8. 10 residues (206 - 215) within the loop had no electron density in the MalC map, indicating flexibility in the region. Weak electron density was also present for residues 106 - 110 in the loop region following β4, the central β strand of the nucleotide binding subdomain. The loop region connects β4 to α4 (part of the active site pocket) and separates the substrate binding site from the cofactor binding site. Its flexibility may be required for possible structural rearrangement upon cofactor or substrate binding.

A search for structural homologs using DALI [93] revealed a group of bacterial short-chain dehydrogenases, including RasADH [126-128]. With a “Asn-Ser-Tyr-Lys” catalytic tetrad, RasADH uses NADP(H) as cofactor and catalyzes oxidation of bulky-bulky or small-bulky secondary alcohols to aldehydes. Superposition of the MalC and RasADH structures shows almost identical folds (RMSD = 1.2 Å), with 25% sequence identity (Fig. 4.5).

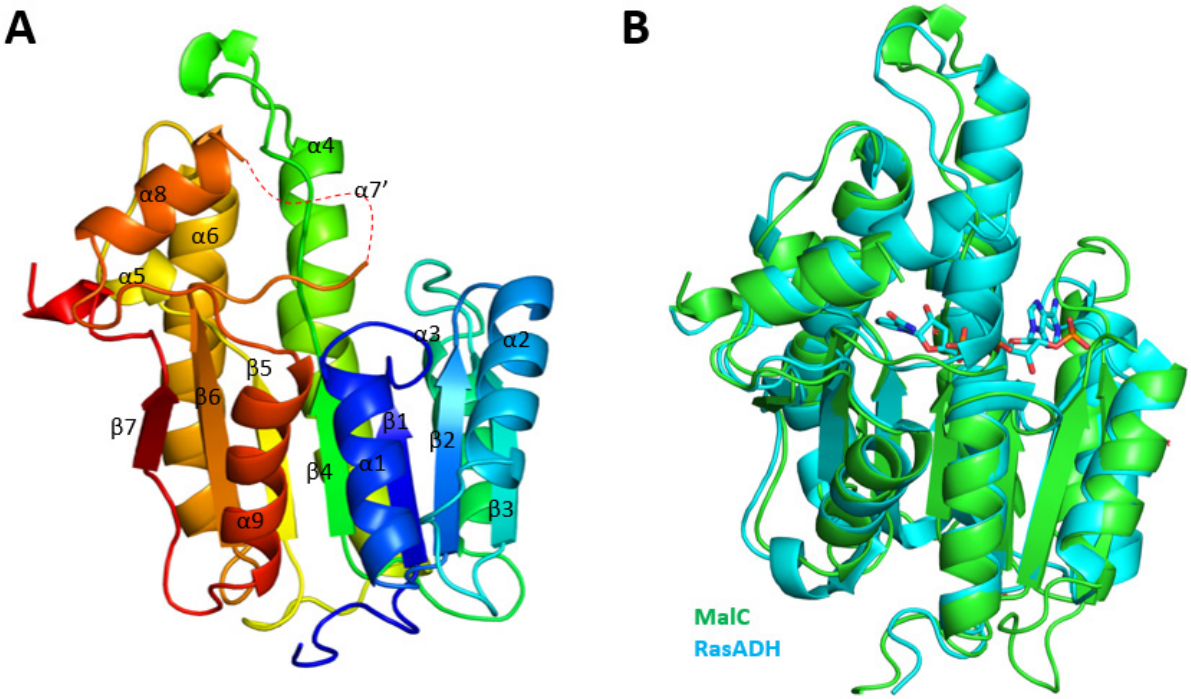


Figure 4.5. Structure of the MalC subunit. A. Rainbow color scheme presentation: N-terminus in blue, C-terminus in red. 10 residues (206 - 215) with no electron density are indicated as a dashed line. B. Superposition of MalC (green) and RasADH structure (cyan, PDB ID: 4BMS, RMSD = 1.2 Å). Structural topology of MalC and RasADH is almost identical. NADPH is shown in sticks.

Although structurally similar to RasADH, MalC lacks the "Asn-Ser-Tyr-Lys" catalytic tetrad (Fig. 4.6). No other amino acids surrounding the putative active site are candidates as a proton donor in substitution for the missing Tyr. This finding strongly discourages the proposal of MalC as a short-chain dehydrogenase/reductase.

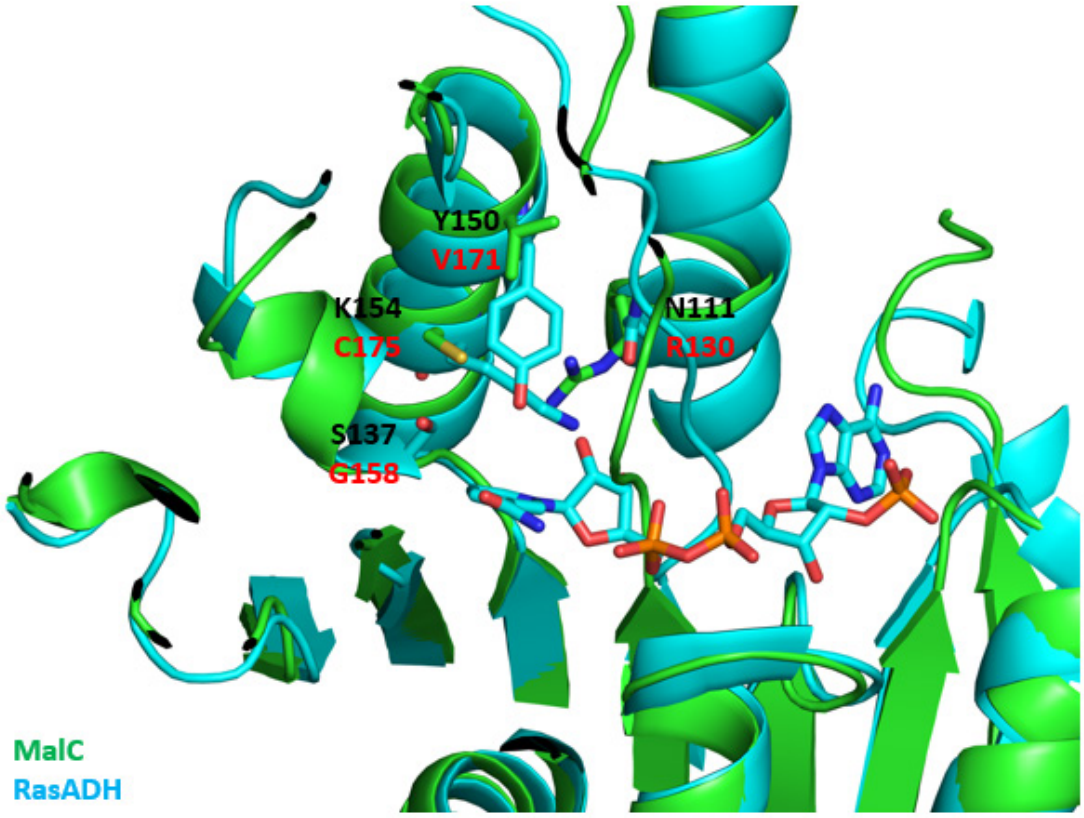


Figure 4.6. Superposition of MalC proposed active site (green) and RasADH active site (cyan). RasADH contains the classic Asn-Ser-Tyr-Lys catalytic tetrad (black), while in MalC these residues are Arg-Gly-Val-Cys (red).

4.2.3 MalC does not bind to NAD(P)H at µM concentration range.

Soaking experiments of MalC crystals with NAD(P) yielded no new density in the proposed nucleotide binding site or elsewhere. To investigate whether MalC binds to NADH or NADPH, fluorescence polarization (FP) was applied (Fig. 4.7). The result clearly shows that MalC does not bind to the reduced form of NADH or NADPH at µM concentration, which is the typical K_d range for SDRs (~1 - 50 µM) [121]. Taking all evidence into consideration, it is conclusive that MalC does not function as a short-chain dehydrogenase.

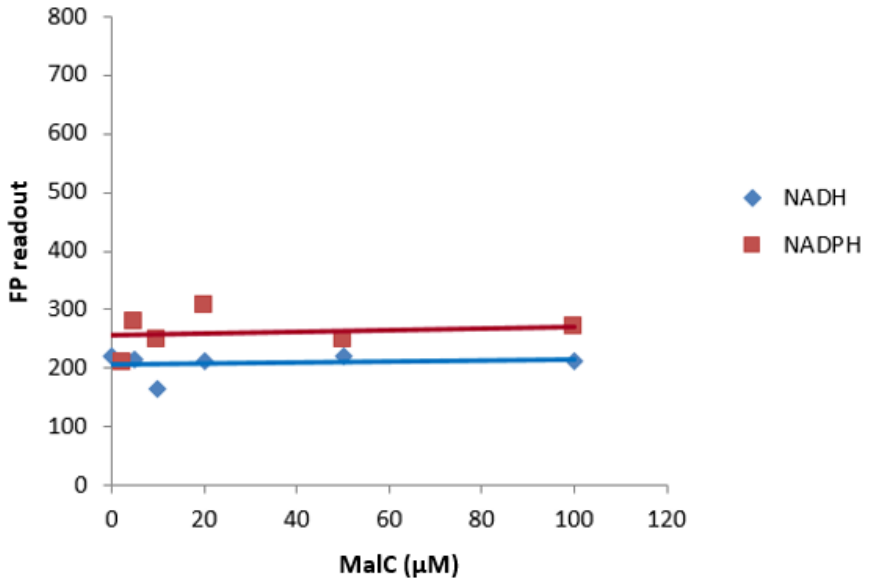


Figure 4.7. Binding assays of MalC with NADH/NADPH measured by fluorescence polarization. For each measurement, concentration of the cofactor was 20 μM, and the concentration of MalC varied from 0 to 100 μM (NADH binding in blue, NADPH binding in brown).

4.2.4 Multiple sequence alignment of MalC suggests a biological function.

The MalC structure clearly indicates that MalC is not a short-chain dehydrogenase. Moreover, MalC does not bind to NADH or NADPH at a μM concentration range. These data led to the hypothesis that MalC does not function as an oxidoreductase in the pathway, and that the *malC* gene was inherited from ancestors and has not yet been eliminated from the fungal genome through evolution. To address that hypothesis, a multiple sequence alignment of MalC was carried out (Fig. 4.S1). The result revealed that MalC has homologs in fungal indole alkaloid biosynthesis that share high sequence identity (40 – 50% pairwise identity), suggestive of a common role in the producing organisms. Sequence conservation was mapped to the MalC structure using ConSurf (Fig. 4.8) [129]. The proposed cofactor binding region is

highly conserved and contains an invariant "TGX$_3$GXG" motif (P-loop), suggesting the capability of binding a nucleotide-containing molecule. The proposed substrate binding site is less conserved. The "Asn-Ser-Tyr-Lys" catalytic tetrad is missing in all MalC homologs, suggesting the possibility that the protein is redesigned to perform a function other than oxidation or reduction.

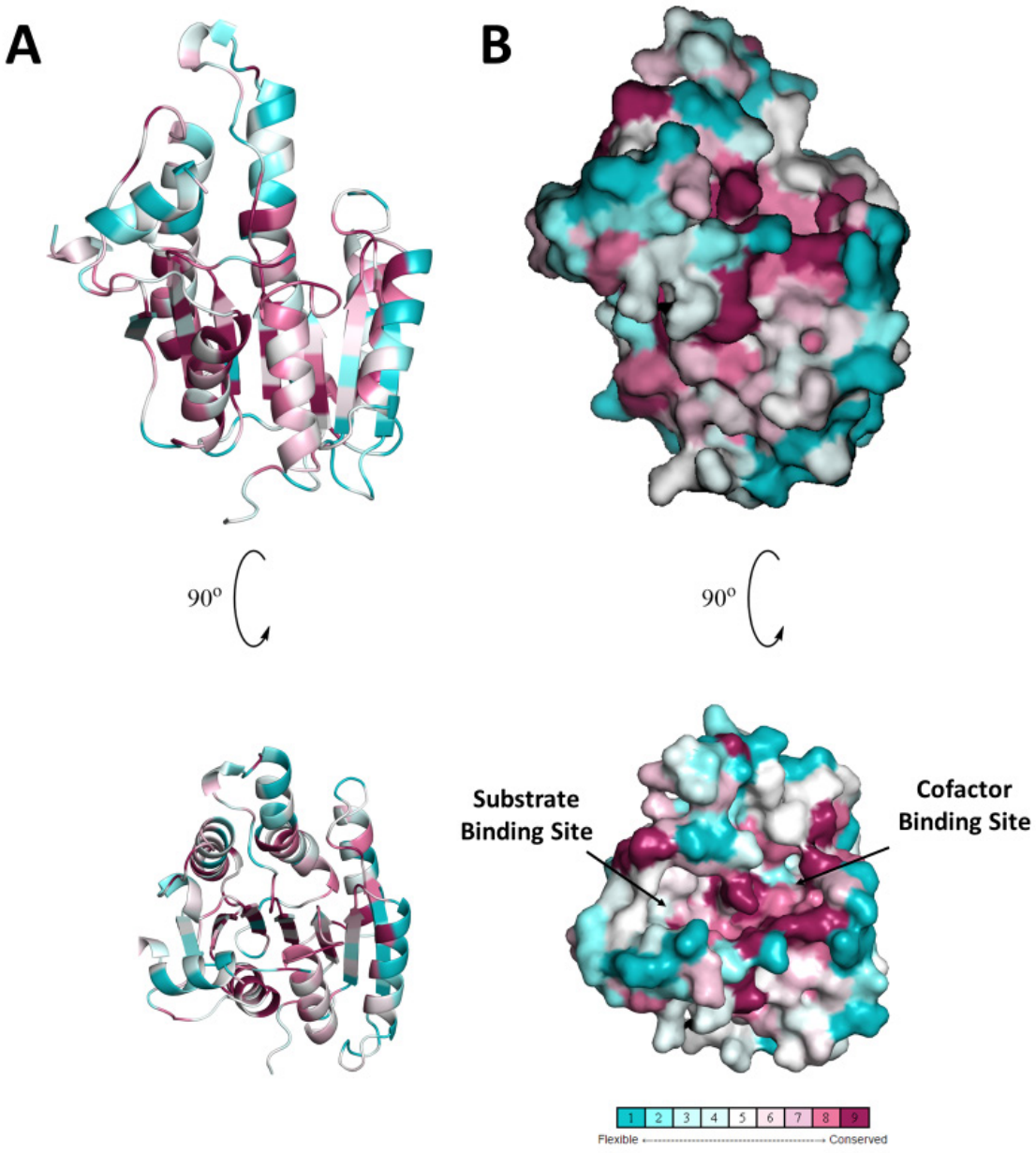


Figure 4.8. MalC sequence conservation mapped onto the structure. Most of the conservation falls in the nucleotide binding subdomain, and the binding cleft.

4.3 Discussion

Biochemical and structural data clearly indicate that MalC is not a short-chain dehydrogenase. Multiple sequence alignment shows conservation of most amino acids, at the positions where catalytic residues are located in SDR enzymes, suggestive of a conserved role for MalC and its homologs in the producing fungi. It is possible that MalC functions in the malbrancheamide pathway, maybe catalyzing or facilitating reactions other than SDR reactions. Previous studies showed that the L-Pro-L-Trp dipeptidyl aldehyde, the proposed pathway intermediate, was highly unstable and spontaneously oxidized *in vitro* (Fig. 2.7). Perhaps a protein binds to the unstable dipeptidyl aldehyde and catalyzes or facilitates the proposed Diels-Alder reaction to produce premalbrancheamide. MalC may serve as such function. However, soaking of MalC crystals with premalbrancheamide, revealed no new density at the proposed active site pocket.

An interesting feature of MalC and its homologs is that they co-exist with NRPS modules having a terminal reductase domain. In the malbrancheamide, paraherquamide and citrinalin biosynthetic pathways, where the terminal domain of an NRPS is proposed to be a reductase, a homolog of MalC is also encoded in the gene cluster [40]. No such SDR-annotated gene product is present in other pathways, (-)/(+)-notoamide pathways for example, that contain an NRPS with a terminal condensation domain. This suggests that MalC may function in cooperation with the NRPS terminal reductase domain. Previous studies have revealed groups of proteins

that bind and facilitate the NPRS-catalyzed reactions. A well-studied example is the MbtH-like protein, which serves as an integral component of bacteria NRPS [130]. MbtH co-purifies with its cognate NRPS *in vitro*, binds to the adenylation domain via π-stacking interactions and facilitates amino acid activation. MalC may perform a similar function, although there is no evidence that MalC binds to MalG, the NRPS in the pathway, and MalC is not related to any type of known NRPS binding motifs/proteins.

Another possibility is that MalC functions as a regulatory protein, not an enzyme in the pathway. A well-known example is NmrA, a negative transcription regulator [131, 132]. NmrA evolved from an SDR ancestor, yet the catalytic Tyr was missing, and the lid of substrate binding pocket was redesigned to bind the GATA-type transcription factor AreA, which regulates the expression of target genes. NmrA strongly prefers binding to the oxidized form of cofactors, NAD^+ or $NADP^+$, whereas the reduced form of the cofactors bind with a K_d in the mM range (6.0 mM in average) [133]. This piece of evidence is consistent with the observation that MalC does not (tightly) bind to NADH or NADPH. However, superposition of MalC and NmrA structures shows distinct differences in the transcription factor binding region, and MalC does not contain α11, which directly forms part of the NmrA-AreA binding interface (Fig. 4.9). MalC also does not contain the same set of NmrA residues responsible for binding the AreA zinc finger, suggesting a different role. Genetic disruption of *malC* is now under way in collaboration with the Sherman group to further investigate its potential function in the malbrancheamide pathway.

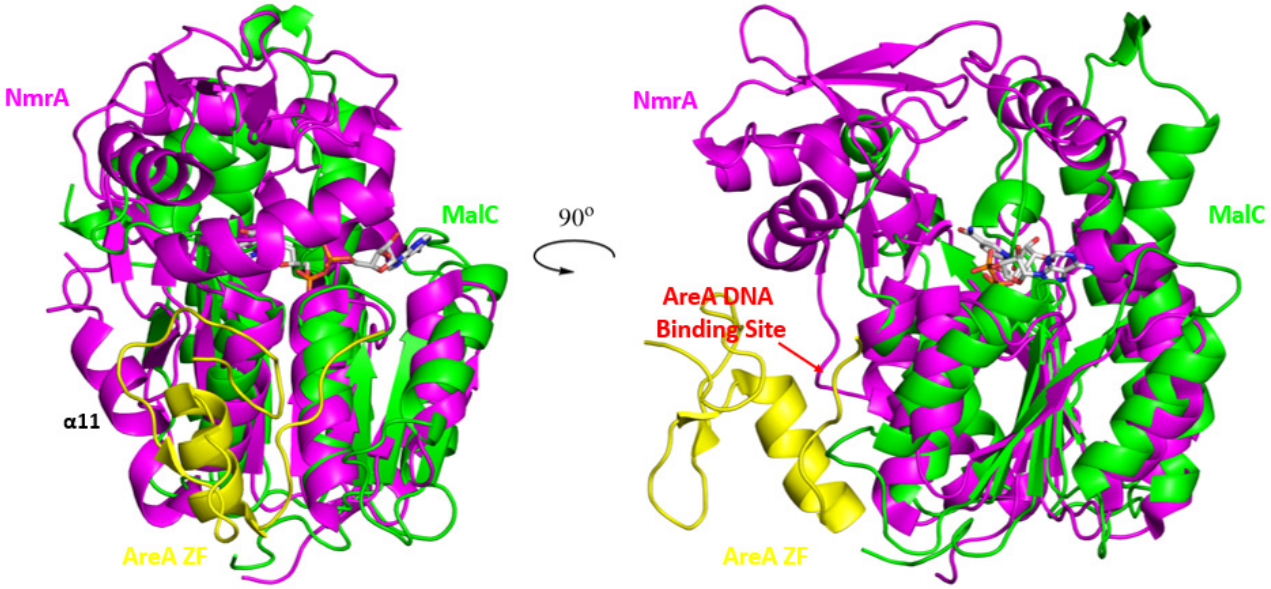


Figure 4.9. Superposition of MalC (green) and NmrA-AreA zinc finger (ZF) complex (NmrA magenta, AreA ZF yellow, PDB ID: 2VUT, RMSD= 3.8 Å). NmrA binds to GATA-type transcription factor AreA ZF, and negatively regulates its gene target expression by competing with AreA-DNA binding. α1, α6 and α11 of NmrA interact with AreA ZF, occupying its DNA binding site.

4.4 Materials and Methods

4.4.1 Materials and strains

NADH and NADPH were purchased commercially. *E. coli* XL1-Blue was used for plasmid storage, E. *coli* BL21(DE3) was used for production of recombinant protein.

4.4.2 Cloning of *malC*

The cDNA library of *Malbranchea aurantiaca* was generated by Shengying Li and Hong Tran. For cloning of *malC*, PCR was used to amplify the cDNA template, followed by a ligation independent cloning (LIC) procedure [94, 95] to insert the gene into the pMCSG7 vector. PCR primers are listed in Table 4.2. The plasmid (pMCSG7-MalC) was then transformed into *E. coli* XL1-Blue cells for storage and harvest.

Table 4.2. Primers used for cloning of *malC*

Gene	Oligo Sequence (5' -> 3')
malC (Forward)	TACTTCCAATCCAATGCCATGGCACCTACCAGGAGATC
malC (Reverse)	TTATCCACTTCCAATGCTATCAGCGCAAAAGCATCCCC

4.4.3 Overexpression and purification

For expression of wild-type *malC*, *E. coli* BL21(DE3) cells were transformed with pMCSG7-MalC and grown in Terrific Broth medium (30 μg/ml ampicillin added) at 37 °C to a O.D. 600 = ~1.0. The culture was then transferred to 20 °C over 1 hour,

and induced with 0.4 mM IPTG. After 18 - 20 hours of incubation (20°C, 225 rpm shake), cells were harvested by centrifugation and the cell pellet was stored at -20 °C. For production of selenomethionyl (SeMet) MalC, *E. coli* BL21(DE3) cells were transformed with pMCSG7-MalC and grown in M9 minimal medium (30 µg/ml ampicillin added) to a O.D. 600 = ~1.0 at 37 °C. SeMet was added to a final concentration of 50 mg/L, and the culture was then transferred to 20 °C over 1 hour, and induced with 0.4 mM IPTG. After 18 - 20 hours, the culture was harvested by centrifugation and the cell pellet was stored at -20 °C.

For purification, the cell pellet was resuspended in lysis buffer (10% v/v glycerol, 500 mM NaCl, 20 mM Tris buffer pH 7.9, 20 mM imidazole, 5 mg lysozyme, 2 mg DNAse and 1 mM $MgCl_2$), and vortexed to mix for 30 min. Sonication and high speed centrifugation (16000 rpm, 30 min) were applied to obtain the lysate soluble fraction. The soluble fraction was filtered and loaded on a GE Ni-NTA HisTrap column (Ni-NTA buffer: 10% glycerol, 500 mM NaCl, 20 mM imidazole pH 7.9, 20 mM Tris pH 7.9; flow rate: 3 mL/min), and was eluted with a 20 – 600 mM imidazole gradient (Ni-NTA buffer; flow rate: 3 mL/min; gradient time: 12 min). Fractions containing wild-type or SeMet MalC were pooled and incubated with TEV protease in a 1:50 w/w ratio at 20 °C for 2 hours to remove the N-terminal His-tag. The tag-free protein was dialyzed overnight at 4 °C into 10% glycerol, 2 mM DTT, 500 mM NaCl, 20 mM Tris pH 7.9, and the flow-through collected from a Ni-NTA HisTrap column (Ni-NTA buffer; flow rate: 3 mL/min). Further homogeneity was achieved by size-exclusion chromatography equilibrated with 10% v/v glycerol, 300 mM NaCl, 20 mM

Tris pH 7.9 (GE HiLoad 16/60 Superdex 200 prep grade column; flow rate: 0.5 mL/min). SDS-PAGE was used to assess protein homogeneity, confirming >95% purity. Yields of purified proteins were 100 mg wild-type MalC per L of *E. coli* culture and 20 mg/L SeMet MalC.

4.4.4 Crystallization and Structural Determination

For crystallization, wild-type MalC or SeMet MalC was mixed with precipitant solution (32% PEG 2K MME, 0.1 M sodium acetate, 0.1 M MES pH 6.5) in a 1:1 v/v ratio. Crystals grew at 20 °C within 24-48 hours, and were harvested without additional cryoprotection and flash cooled in liquid nitrogen. Data from crystals of wild-type MalC were collected at the Advanced Photon Source (GM/CA beamline 23-ID-D) at an X-ray wavelength of 1.033 Å with a Pilatus3 6M detector (360° of data, 100 K, 0.2° image width). Data from crystals of SeMet MalC were collected at an X-ray wavelength of 0.979 Å. Data were processed with XDS [96], and the SeMet MalC crystal structure was solved by single-wavelength anomalous diffraction (SAD) phasing with AutoSol [135]. Model building was carried out with Coot [98], and refinement was carried out with PHENIX [99]. Statistics for the crystallographic data and refinement are shown in Table 4.3.

4.4.5 Test for enzymatic activity

MalC (50 μM), MalE (50 μM), MalB (20 μM), DMAPP (250 μM) and NAD or NADP (1 mM) were added to 10% v/v glycerol, 50 mM NaCl, 50 mM HEPES pH

7.5. The reaction was initiated by addition of 500 μM substrate. The reaction mix was incubated at 25°C for 6 hours, and quenched with 50% v/v methanol. Denatured protein was removed by centrifugation and the soluble fraction was analyzed by LC/MS (Phenomenex Kinetix reverse-phase C18 column (40 mm × 2.1 mm, 2.6 μm), Buffer A: 0.2% v/v formic acid in water, Buffer B: 0.2% v/v formic acid in acetonitrile. HPLC protocol: 5% Buffer A for 2 min, 20-100% Buffer B gradient for 4 min, 100% Buffer B for 2 min. Flow rate: 0.5 mL/min.)

Author Contributions

Qingyun Dan, Sean Newmister, David Sherman and Janet Smith contributed to the experimental design. Shengying Li and Hong Tran generated the cDNA library. Qingyun Dan cloned and expressed MalC, performed protein purification and carried out all crystallographic and analysis steps. Qingyun Dan, Sean Newmister, David Sherman and Janet Smith evaluated the data.

Appendix

Table 4.3. Data collection and refinement statistics.

	MalC wild-type
Space group	P 4_2
Unit cell parameters (Å, °)	79.4, 79.4, 133.6, 90, 90, 90
Wavelength (Å)	1.033
Resolution (Å)	26.46 - 1.59 (1.63 - 1.59)
Completeness (%)	99 (90.3)
Reflections	665318 (103821)
Unique reflections	116831 (18041)
Multiplicity	5.7 (5.8)
Mean I/σ	13.8 (1.1)
R_{meas}	0.077 (0.78)
$CC_{1/2}$	1 (0.78)
CC*	1 (0.89)
Reflections (working set)	107788 (7657)
Reflections (test set)	1884 (126)
R_{work}	0.18 (0.23)
R_{free}	0.21 (0.25)
No. of chains per AU	4
No. of cofactors	0
No. of atoms protein/waters	7812/679
Avg B-value ($Å^2$)	32.6
Ramachandran plot: favored/allowed/outliers (%)	97.7/1.8/0.4
RMSD bonds (Å)	0.008
RMSD angles (°)	1.063

Figure 4.S1. Multiple sequence alignment of MalC and its homologs. None of the SDR catalytic tetrad (black box) is present. The "TGX$_3$GXG" motif (the first red box) in conserved, while the "NNAG" motif (the second red box) after β4 is missing in MalC homologs.

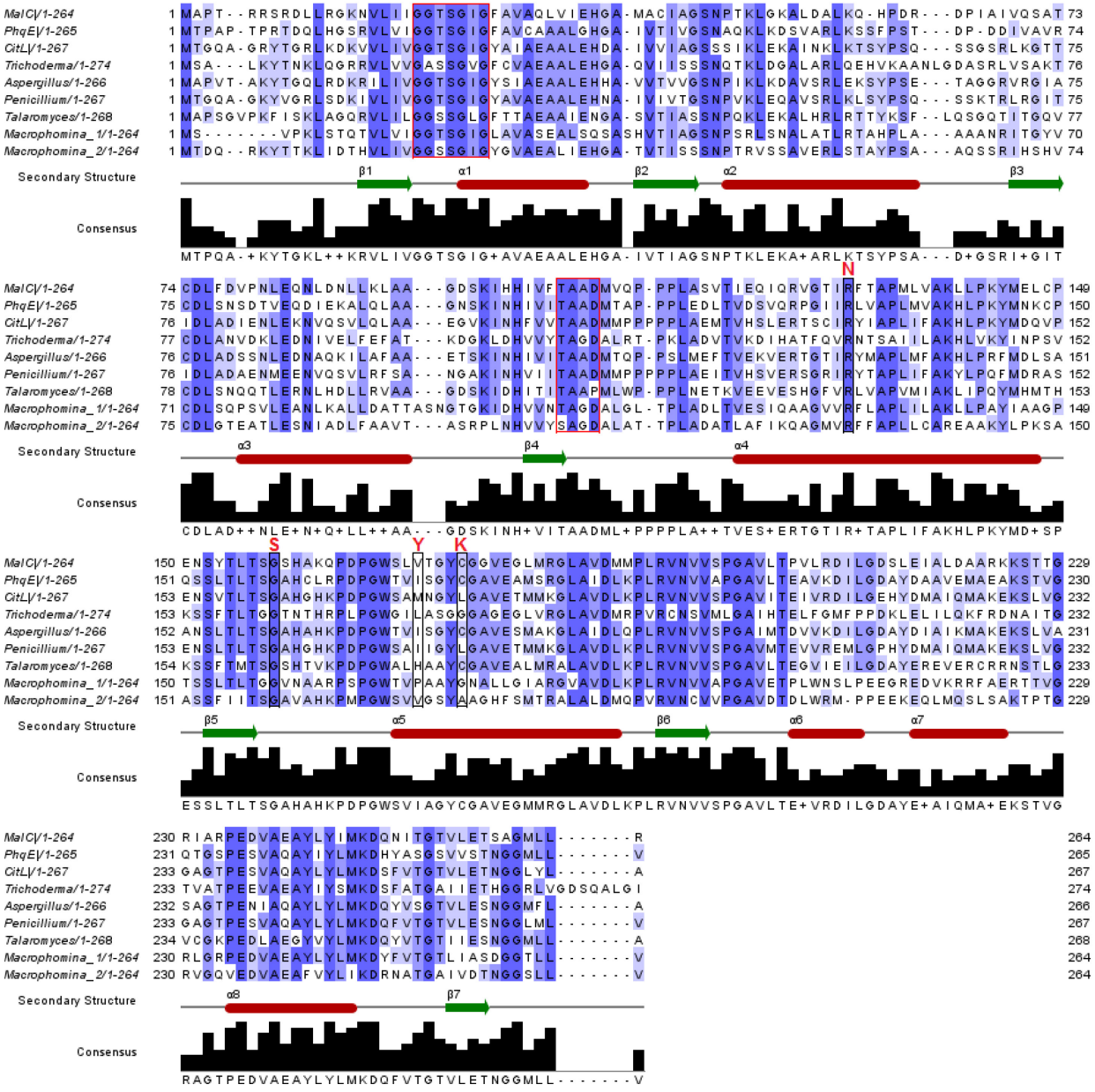
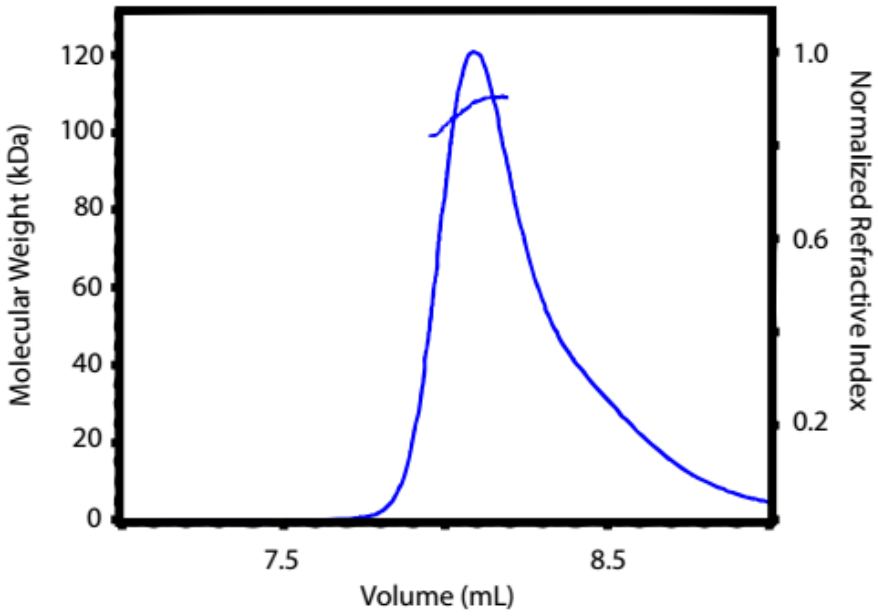


Figure 4.S2. Multi-angle light scattering analysis of purified MalC. Three statistical calculation methods were applied to measure the molar mass moments (M_w: weight-average molecular weight, 105.8 ± 1.7 Da. M_n: number-average molecular weight, 105.9 ± 1.7 Da. M_z: size-average molecular weight, 106.0 ± 6.1 Da.). The calculated mass of MalC subunit is 28.3 Da, and the result indicates that MalC is a tetramer in solution.



Chapter V

Future Directions

5.1 MalG: The NRPS

The structure of the terminal reductase domain of PhqB, a MalG homolog in the paraherquamide pathway (Figure 1.12), resembles structures of bacterial NRPS terminal reductase domains, strongly indicating a similar reductive hydrolysis mechanism of termination. What is still unknown is whether MalG catalyzes a 2-electron reduction to produce a peptidyl aldehyde, or a 4-electron reduction forming a peptidyl alcohol. Identifying the NRPS product will be a major target for future studies, particularly since it is the first step in the pathway. The peptidyl aldehyde was shown to be unstable *in vitro* (Fig. 2.7), suggesting the possibility of a more stable alcohol NRPS product. Consistent with this, the dipeptidyl alcohol has been shown to be stable *in vitro*, and able to be efficiently prenylated at the C-2 position by MalE or MalB, the prenyltransferases in the pathway (Fig. 3.5 and 3.6). The only missing piece in this scheme is a dehydrogenase to oxidize the prenylated alcohol to an aldehyde, which is proposed to tautomerize and produce a Diels-Alder product [40]. Nevertheless, a detailed enzymatic characterization of the NRPS is needed. Despite considerable effort, soluble full-length MalG was not obtained by heterologous expression in either *E. coli* or insect cells. A good alternative approach is to seek a

homolog. Currently three gene clusters (paraherquamide, citrinalin and chrysogenamide) homologous to the malbrancheamide gene cluster have been sequenced by the Sherman group, and cDNA libraries are being generated. With cDNAs in hand, production of recombinant MalG homologs can be tested, and successful enzymatic characterization with known amino acid substrates can help to reveal the actual NRPS product in these homologous fungal indole alkaloid pathways.

5.2 Crystal engineering of PhqB R

PhqB R crystals are very difficult to reproduce, with less than 1% of crystals having high-quality diffraction (better than 3.5 Å). A 2.65 Å crystal structure was solved, and three loops are missing in the electron density map, including the loop connecting the catalytic Ser and Tyr. Better diffracting crystals that can be reproduced more robustly are essential to future structural investigations. Crystal engineering has been used widely to improve reproducibility and diffraction quality of target crystals. With the crystal structure of PhqB R in hand, I will examine the crystal lattice contacts to identify any contacts that contribute to poor crystal packing, and use mutagenesis to create better contacts. Crystals that grow more reproducibly and have better diffraction quality are essential to obtaining cofactor complex.

Also, no structure of an NRPS terminal T-R didomain has been reported and it is unclear how reductase domains interact with T domains. A helix-turn-helix region in the C-terminal domain was proposed as a T domain binding site [87]. In the PhqB R crystal structure, this region is largely solvent exposed, and there appears to be

enough space to accommodate a T domain. The purified PhqB T domain will be used in co-crystallization experiments. A co-crystal structure of PhqB T-R would significantly improve our understanding of NRPS T-R interactions and functions.

5.3 MalB and MalE: Prenyltransferases

It is intriguing that two prenyltransferases are encoded within the malbrancheamide gene cluster, yet only one prenylation step is required for the pathway. The same situation exists in the homologous paraherquamide pathway, where the gene cluster encodes three hypothetical prenyltransferases, while only two prenylations are proposed to occur. It is possible that the existence of an extra prenyltransferase gene is due to gene duplication, which the fungal genomes have not yet eliminated through evolution. It is also possible that either MalE or MalB is multi-functional and catalyzes other reactions in addition to prenylation. Insights should come from genetic disruption experiments, which are under way in collaboration with the Sherman group.

Prenyltransferases are also very good candidates for protein engineering, with broad substrate specificity and strict prenylation site control. The structural basis for substrate promiscuity is the greasy pocket, which accommodates its substrate mainly via hydrophobic interactions. Structure-guided site-directed mutagenesis within the pocket can alter the substrate specificity of prenyltransferases [35]. In addition to the two prenyltrasferases characterized in my thesis research (MalB and MalE, Fig. 3.5 and 3.6), three other homologous fungal indole prenyltransferases in three different

pathways have exhibited tight substrate specificity (personal communication from Sean Newmister). It is unknown whether these fungal indole alkaloid prenyltransferases have unique pockets that bind substrates through specific interactions in addition to hydrophobic contacts. Comparative structural analysis and site-directed mutagenesis could provide useful information on the structural basis for the strict substrate specificity, and guild protein engineering and chemoenzymatic synthesis of related novel compounds.

5.4 Function of MalC

The high sequence conservation of MalC and its homologs (40 – 50% pairwise identity) suggests a common role in the producing organisms. However, the crystal structure of MalC clearly indicates that it is not a short-chain dehydrogenase. The active site does not resemble any known type of catalytic chamber, as none of the catalytic amino acids in the SDR family is present in MalC. Thus it is unclear whether MalC functions as an enzyme in the malbrancheamide pathway. Genetic disruption of MalC or its homologs may provide useful information. Genome mining of MalC will also help to identify its potential partners. With the limited information available, MalC homologs co-locate with an NRPS having a terminal reductase domain, suggesting the possibility that MalC may function together with the NRPS. However, MalC has no sequence or structural similarity to known NRPS binding peptides/proteins. Once enzymatic characterization of any NRPS homologous to

MalG has been achieved, MalC or its holomogs can be tested in enzymatic assays to see if their presence affects the function of the cognate NRPSs.

References

1. Foye, W. O. (2008). *Foye's principles of medicinal chemistry*. T. L. Lemke, & D. A. Williams (Eds.). Lippincott Williams & Wilkins.
2. Dixon, R. A. (2001). Natural products and plant disease resistance. *Nature*, *411*(6839), 843-847.
3. Newman, D. J., Cragg, G. M., & Snader, K. M. (2000). The influence of natural products upon drug discovery. *Natural product reports*, *17*(3), 215-234.
4. Walsh, C.T. (2003). *Antibiotics: Actions, Origins, Resistance. ASM Press, Washington*.
5. Stiller, C. R., Dupre, J., Gent, M., Jenner, M. R., Keown, P. A., Laupacis, A., ... & Wolfe, B. M. (1984). Effects of cyclosporine immunosuppression in insulin-dependent diabetes mellitus of recent onset. *Science*, *223*(4643), 1362-1367.
6. Butler, M. S. (2004). The role of natural product chemistry in drug discovery. *Journal of natural products*, *67*(12), 2141-2153.
7. Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nature medicine*, *17*(10), 1217-1220.
8. Vane, J. R., & Botting, R. M. (2003). The mechanism of action of aspirin. *Thrombosis research*, *110*(5), 255-258.
9. Alder, A. L. (1970). *The History of Penicillin Production.* American Institute of Chemical Engineers: New York.
10. Lax, E. (2004). *The mold in Dr. Florey's coat: the story of the penicillin miracle*. Macmillan.
11. Newman, D. J., & Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *Journal of natural products*, *79*(3), 629-661.
12. Brown, E. J., Albers, M. W., Shin, T. B., Ichikawa, K., Keith, C. T., Lane, W. S., & Schreiber, S. L. (1994). A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature*, *369*(6483), 756-758.
13. Sabatini, D. M., Erdjument-Bromage, H., Lui, M., Tempst, P., & Snyder, S. H. (1994). RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell*, *78*(1), 35-43.

14. Sabers, C. J., Martin, M. M., Brunn, G. J., Williams, J. M., Dumont, F. J., Wiederrecht, G., & Abraham, R. T. (1995). Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells. *Journal of Biological Chemistry*, *270*(2), 815-822.

15. Pettit, G. R., Herald, C. L., Doubek, D. L., Herald, D. L., Arnold, E., & Clardy, J. (1982). Isolation and structure of bryostatin 1. *Journal of the American Chemical Society*, *104*(24), 6846-6848.

16. Schaufelberger, D. E., Koleck, M. P., Beutler, J. A., Vatakis, A. M., Alvarado, A. B., Andrews, P., ... & Lebherz, W. B. (1991). The large-scale isolation of bryostatin 1 from Bugula neritina following current good manufacturing practices. *Journal of natural products*, *54*(5), 1265-1270.

17. Sheehan, J. C., & Henery-Logan, K. R. (1959). The total synthesis of penicillin V. *Journal of the American Chemical Society*, *81*(12), 3089-3094.

18. Nicolaou, K. C., & Chen, J. S. (2010). *Classics in total synthesis III*. Wiley-VCH.

19. Evans, D. A., Carter, P. H., Carreira, E. M., Charette, A. B., Prunet, J. A., & Lautens, M. (1999). Total synthesis of bryostatin 2. *Journal of the American Chemical Society*, *121*(33), 7540-7552.

20. Weissman, K. J., & Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nature reviews microbiology*, *3*(12), 925-936.

21. Marahiel, M. A., Stachelhaus, T., & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical reviews*, *97*(7), 2651-2674.

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

23. Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., ... & Skiniotis, G. (2014). Structure of a modular polyketide synthase. *Nature*, *510*(7506), 512.

24. Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., ... & Skiniotis, G. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*, *510*(7506), 560-564.

25. Strieker, M., Tanović, A., & Marahiel, M. A. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Current opinion in structural biology*, *20*(2), 234-240.

26. Gulick, A. M. (2016). Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Current Opinion in Chemical Biology*, *35*, 89-96.

27. Menzella, H. G., Carney, J. R., & Santi, D. V. (2007). Rational design and assembly of synthetic trimodular polyketide synthases. *Chemistry & biology*, *14*(2), 143-151.

28. Li, S. M. (2010). Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. *Natural product reports*, *27*(1), 57-78.

29. Williams, R. M., Stocking, E. M., & Sanz-Cervera, J. F. (2000). Biosynthesis of prenylated alkaloids derived from tryptophan. In *Biosynthesis* (pp. 97-173). Springer Berlin Heidelberg.

30. Qian-Cutrone, J., Huang, S., Shu, Y. Z., Vyas, D., Fairchild, C., Menendez, A., ... & Gao, Q. (2002). Stephacidin A and B: two structurally novel, selective inhibitors of the testosterone-dependent prostate LNCaP cells. *Journal of the American Chemical Society*, *124*(49), 14556-14557.

31. Heide, L. (2009). Prenyl transfer to aromatic substrates: genetics and enzymology. *Current opinion in chemical biology*, *13*(2), 171-179.

32. Steffan, N., Grundmann, A., Yin, W. B., Kremer, A., & Li, S. M. (2009). Indole prenyltransferases from fungi: a new enzyme group with high potential for the production of prenylated indole derivatives. *Current medicinal chemistry*, *16*(2), 218-231.

33. Yin, W. B., Ruan, H. L., Westrich, L., Grundmann, A., & Li, S. M. (2007). CdpNPT, an N-Prenyltransferase from Aspergillus fumigatus: Overproduction, Purification and Biochemical Characterisation. *Chembiochem*, *8*(10), 1154-1161.

34. Metzger, U., Schall, C., Zocher, G., Unsöld, I., Stec, E., Li, S. M., ... & Stehle, T. (2009). The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proceedings of the National Academy of Sciences*, *106*(34), 14309-14314.

35. Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, X., Li, S. M., & Stehle, T. (2010). Structure− function analysis of an enzymatic prenyl transfer reaction identifies a reaction chamber with modifiable specificity. *Journal of the American Chemical Society*, *132*(50), 17849-17858.

36. Schuller, J. M., Zocher, G., Liebhold, M., Xie, X., Stahl, M., Li, S. M., & Stehle, T. (2012). Structure and catalytic mechanism of a cyclic dipeptide prenyltransferase with broad substrate promiscuity. *Journal of molecular biology*, *422*(1), 87-99.

37. Ding, Y., Wet, J. R. D., Cavalcoli, J., Li, S., Greshock, T. J., Miller, K. A., ... & Williams, R. M. (2010). Genome-based characterization of two prenylation steps in the assembly of the stephacidin and notoamide anticancer agents in a marine-derived Aspergillus sp. *Journal of the American Chemical Society*, *132*(36), 12733-12740.

38. Finefield, J. M., Frisvad, J. C., Sherman, D. H., & Williams, R. M. (2012). Fungal origins of the bicyclo [2.2. 2] diazaoctane ring system of prenylated indole alkaloids. *Journal of natural products*, *75*(4), 812-833.
39. Kato, H., Yoshida, T., Tokue, T., Nojiri, Y., Hirota, H., Ohta, T., ... & Tsukamoto, S. (2007). Notoamides A–D: Prenylated Indole Alkaloids Isolated from a Marine-Derived Fungus, Aspergillus sp. *Angewandte Chemie*, *119*(13), 2304-2306.
40. Li, S., Srinivasan, K., Tran, H., Yu, F., Finefield, J. M., Sunderhaus, J. D., ... & Sherman, D. H. (2012). Comparative analysis of the biosynthetic systems for fungal bicyclo [2.2. 2] diazaoctane indole alkaloids: The (+)/(−)-notoamide, paraherquamide and malbrancheamide pathways. *MedChemComm*, *3*(8), 987-996.
41. Yamazaki, M., Okuyama, E., Kobayashi, M., & Inoue, H. (1981). The structure of paraherquamide, a toxic metabolite from penicilliumparaherquei. *Tetrahedron Letters*, *22*(2), 135-136.
42. Shoop, W. L., Eary, C. H., Michael, B. F., Haines, H. W., & Seward, R. L. (1991). Anthelmintic activity of paraherquamide in dogs. *Veterinary parasitology*, *40*(3-4), 339-341.
43. Shoop, W. L., Michael, B. F., Haines, H. W., & Eary, C. H. (1992). Anthelmintic activity of paraherquamide in calves. *Veterinary parasitology*, *43*(3), 259-263.
44. Martínez-Luis, S., Rodríguez, R., Acevedo, L., González, M. C., Lira-Rocha, A., & Mata, R. (2006). Malbrancheamide, a new calmodulin inhibitor from the fungus Malbranchea aurantiaca. *Tetrahedron*, *62*(8), 1817-1822.
45. Herzon, S. B., & Myers, A. G. (2005). Enantioselective synthesis of stephacidin B. *Journal of the American Chemical Society*, *127*(15), 5342-5344.
46. Baran, P. S., Guerrero, C. A., Ambhaikar, N. B., & Hafensteiner, B. D. (2005). Short, enantioselective total synthesis of stephacidin A. *Angewandte Chemie International Edition*, *44*(4), 606-609.
47. Greshock, T. J., Grubbs, A. W., Tsukamoto, S., & Williams, R. M. (2007). A concise, biomimetic total synthesis of stephacidin A and notoamide B. *Angewandte Chemie International Edition*, *46*(13), 2262-2265.
48. Greshock, T. J., Grubbs, A. W., Jiao, P., Wicklow, D. T., Gloer, J. B., & Williams, R. M. (2008). Isolation, Structure Elucidation, and Biomimetic Total Synthesis of Versicolamide B, and the Isolation of Antipodal (−)-Stephacidin A and (+)-Notoamide B from Aspergillus versicolor NRRL 35600. *Angewandte Chemie International Edition*, *47*(19), 3573-3577.
49. Grubbs, A. W., Artman, G. D., Tsukamoto, S., & Williams, R. M. (2007). A concise total synthesis of the notoamides C and D. *Angewandte Chemie*, *119*(13), 2307-2311.
50. Finefield, J. M., Kato, H., Greshock, T. J., Sherman, D. H., Tsukamoto, S., & Williams, R. M. (2011). Biosynthetic studies of the notoamides: isotopic

synthesis of stephacidin A and incorporation into notoamide B and sclerotiamide. *Organic letters*, *13*(15), 3802-3805.
51. Stocking, E. M., Sanz-Cervera, J. F., & Williams, R. M. (2001). Studies on the biosynthesis of paraherquamide: synthesis and incorporation of a hexacyclic indole derivative as an advanced metabolite. *Angewandte Chemie International Edition*, *40*(7), 1296-1298.
52. Ding, Y., Greshock, T. J., Miller, K. A., Sherman, D. H., & Williams, R. M. (2008). Premalbrancheamide: synthesis, isotopic labeling, biosynthetic incorporation, and detection in cultures of Malbranchea aurantiaca. *Organic letters*, *10*(21), 4863-4866.
53. Li, S., Finefield, J. M., Sunderhaus, J. D., McAfoos, T. J., Williams, R. M., & Sherman, D. H. (2011). Biochemical characterization of NotB as an FAD-dependent oxidase in the biosynthesis of notoamide indole alkaloids. *Journal of the American Chemical Society*, *134*(2), 788-791.
54. Sunderhaus, J. D., McAfoos, T. J., Finefield, J. M., Kato, H., Li, S., Tsukamoto, S., ... & Williams, R. M. (2012). Synthesis and bioconversions of notoamide T: a biosynthetic precursor to stephacidin A and notoamide B. *Organic letters*, *15*(1), 22-25.
55. Evans, D. A., & Johnson, J. S. (1999). Diels-Alder Reactions. *Comprehensive asymmetric catalysis*, *3*, 1177.
56. Ahrendt, K. A., Borths, C. J., & MacMillan, D. W. (2000). New strategies for organic catalysis: the first highly enantioselective organocatalytic Diels-Alder reaction. *Journal of the American Chemical Society*, *122*(17), 4243-4244.
57. Nicolaou, K. C., Snyder, S. A., Montagnon, T., & Vassilikogiannakis, G. (2002). The Diels–Alder reaction in total synthesis. *Angewandte Chemie International Edition*, *41*(10), 1668-1698.
58. Watanabe, K., Oikawa, H., Yagi, K., Ohashi, S., Mie, T., Ichihara, A., & Hornma, M. (2000). Macrophomate synthase: characterization, sequence, and expression in Escherichia coli of the novel enzyme catalyzing unusual multistep transformation of 2-pyrones to benzoates. *Journal of biochemistry*, *127*(3), 467-473.
59. Ose, T., Watanabe, K., Mie, T., Honma, M., Watanabe, H., Yao, M., ... & Tanaka, I. (2003). Insight into a natural Diels–Alder reaction from the structure of macrophomate synthase. *Nature*, *422*(6928), 185-189.
60. Guimarães, C. R. W., Udier-Blagovic, M., & Jorgensen, W. L. (2005). Macrophomate synthase: QM/MM simulations address the Diels-Alder versus Michael-Aldol reaction mechanism. *Journal of the American Chemical Society*, *127*(10), 3577-3588.
61. Serafimov, J. M., Gillingham, D., Kuster, S., & Hilvert, D. (2008). The putative Diels− Alderase macrophomate synthase is an efficient aldolase. *Journal of the American Chemical Society*, *130*(25), 7798-7799.
62. Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., Clair, J. L. S., ... & Houk, K. N. (2010). Computational design of an enzyme catalyst

for a stereoselective bimolecular Diels-Alder reaction. *Science*, *329*(5989), 309-313.
63. Kim, H. J., Ruszczycky, M. W., Choi, S. H., Liu, Y. N., & Liu, H. W. (2011). Enzyme-catalysed [4+ 2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature*, *473*(7345), 109-112.
64. Fage, C. D., Isiorho, E. A., Liu, Y., Wagner, D. T., Liu, H. W., & Keatinge-Clay, A. T. (2015). The structure of SpnF, a standalone enzyme that catalyzes [4+ 2] cycloaddition. *Nature chemical biology*, *11*(4), 256-258.
65. Hashimoto, T., Hashimoto, J., Teruya, K., Hirano, T., Shin-ya, K., Ikeda, H., ... & Kuzuyama, T. (2015). Biosynthesis of versipelostatin: identification of an enzyme-catalyzed [4+ 2]-cycloaddition required for macrocyclization of spirotetronate-containing polyketides. *Journal of the American Chemical Society*, *137*(2), 572-575.
66. Zheng, Q., Guo, Y., Yang, L., Zhao, Z., Wu, Z., Zhang, H., ... & Jiang, H. (2016). Enzyme-Dependent [4+ 2] Cycloaddition Depends on Lid-like Interaction of the N-Terminal Sequence with the Catalytic Core in PyrI4. *Cell chemical biology*, *23*(3), 352-360.
67. Hashimoto, T., & Kuzuyama, T. (2016). Mechanistic insights into Diels-Alder reactions in natural product biosynthesis. *Current Opinion in Chemical Biology*, *35*, 117-123.
68. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.
69. Rohl, C. A., Strauss, C. E., Misura, K. M., & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods in enzymology*, *383*, 66-93.
70. Schwarzer, D., Finking, R., & Marahiel, M. A. (2003). Nonribosomal peptides: from genes to products. *Natural product reports*, *20*(3), 275-287.
71. Tanovic, A., Samel, S. A., Essen, L. O., & Marahiel, M. A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science*, *321*(5889), 659-663.
72. Marahiel, M. A., & Essen, L. O. (2009). Nonribosomal peptide synthetases: mechanistic and structural aspects of essential domains. *Methods in enzymology*, *458*, 337-351.
73. Konz, D., & Marahiel, M. A. (1999). How do peptide synthetases generate structural diversity? *Chemistry & biology*, *6*(2), R39-R48.
74. Banko, G., Demain, A. L., & Wolfe, S. (1987). δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine synthetase (ACV synthetase): a multifunctional enzyme with broad substrate specificity for the synthesis of penicillin and cephalosporin precursors. *Journal of the American Chemical Society*, *109*(9), 2858-2860.
75. Borel, J. F. (1986). Editorial: Ciclosporin and its future. In *Ciclosporin* (pp. 9-18). Karger Publishers.

76. Hotchkiss, R. D. (1944). Gramicidin, tyrocidine, and tyrothricin. *Adv. Enzymol. Relat. Areas Mol. Biol*, *4*, 153-199.
77. Graham, J. R., & Wolff, H. G. (1938). Mechanism of migraine headache and action of ergotamine tartrate. *Archives of Neurology & Psychiatry*, *39*(4), 737-763.
78. Blum, R. H., Carter, S. K., & Agre, K. (1973). A clinical review of bleomycin—a new antineoplastic agent. *Cancer*, *31*(4), 903-914.
79. Mootz, H. D., Schwarzer, D., & Marahiel, M. A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem*, *3*(6), 490-504.
80. Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & biology*, *6*(8), 493-505.
81. Challis, G. L., Ravel, J., & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & biology*, *7*(3), 211-224.
82. Keating, T. A., Marshall, C. G., Walsh, C. T., & Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural & Molecular Biology*, *9*(7), 522-526.
83. Marahiel, M. A., & Essen, L. O. (2009). Nonribosomal peptide synthetases: mechanistic and structural aspects of essential domains. *Methods in enzymology*, *458*, 337-351.
84. Walsh, C. T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*, *303*(5665), 1805-1810.
85. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A., & Walsh, C. T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature*, *407*(6801), 215-218.
86. Du, L., & Lou, L. (2010). PKS and NRPS release mechanisms. *Natural product reports*, *27*(2), 255-278.
87. Barajas, J. F., Phelan, R. M., Schaub, A. J., Kliewer, J. T., Kelly, P. J., Jackson, D. R., ... & Tsai, S. C. (2015). Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. *Chemistry & biology*, *22*(8), 1018-1029.
88. Chhabra, A., Haque, A. S., Pal, R. K., Goyal, A., Rai, R., Joshi, S., ... & Gokhale, R. S. (2012). Nonprocessive [2+ 2] e-off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences*, *109*(15), 5681-5686.
89. Wilson, D. J., Shi, C., Teitelbaum, A. M., Gulick, A. M., & Aldrich, C. C. (2013). Characterization of AusA: a dimodular nonribosomal peptide synthetase responsible for the production of aureusimine pyrazinones. *Biochemistry*, *52*(5), 926-937.

90. Xu, W., Gavia, D. J., & Tang, Y. (2014). Biosynthesis of fungal indole alkaloids. *Natural product reports*, *31*(10), 1474-1487.
91. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., & Walsh, C. T. (1998). Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*, *37*(6), 1585-1595.
92. Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, *372*(3), 774-797.
93. Holm, L., & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends in biochemical sciences*, *20*(11), 478-480.
94. Aslanidis, C., and de Jong, P. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research* 18, 6069-6074.
95. Donnelly, M., Zhou, M., Millard, C., Clancy, S., Stols, L., Eschenfeldt, W., Collart, F., and Joachimiak, A. (2006). An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein expression and purification* 47, 446-454.
96. Kabsch, W. (2010). Xds. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 125-132.
97. DiMaio, F., Echols, N., Headd, J. J., Terwilliger, T. C., Adams, P. D., & Baker, D. (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nature methods*, *10*(11), 1102-1104.
98. Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2126-2132.
99. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., ... & McCoy, A. J. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213-221.
100. Keating, T. A., Ehmann, D. E., Kohli, R. M., Marshall, C. G., Trauger, J. W., & Walsh, C. T. (2001). Chain Termination Steps in Nonribosomal Peptide Synthetase Assembly Lines: Directed Acyl-S-Enzyme Breakdown in Antibiotic and Siderophore Biosynthesis. ChemBioChem, 2(2), 99-107.
101. Gallagher, R., and Latch, G. (1977). Production of the Tremorgenic Mycotoxins Verruculogen and Fumitremorgin B by Penicillium piscarium Westling. *Applied and environmental microbiology 33*, 730-731.
102. López-Gresa, M., González, M., Ciavatta, L., Ayala, I., Moya, P., and Primo, J. (2006). Insecticidal activity of Paraherquamides, including paraherquamide H and paraherquamide I, two new alkaloids isolated from Penicillium cluniae. *Journal of agricultural and food chemistry 54*, 2921-2925.
103. Telloa, M., Kuzuyamab, T., Heidec, L., Noela, J. P. and Richarda, S. B. (2008) The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cellular and Molecular Life Sciences 65*, 1459–1463.

104. Usui, T., Kondoh, M., Cui, C., Mayumi, T., and Osada, H. (1998). Tryprostatin A, a specific and novel inhibitor of microtubule assembly. *The Biochemical journal 333 (Pt 3)*, 543-548.
105. Botta, B., Vitali, A., Menendez, P., Misiti, D., Delle, M.G. (2005). Prenylated flavonoids: pharmacology and biotechnology. *Current Medicinal Chemistry 12*, 717-739.
106. Kuzuyama, T., Noel, J.P. and Richard, S.B. (2005). Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature 435*, 983–987.
107. Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, L., Li, S.M., and Stehle, T. (2010). Structure-Function Analysis of an Enzymatic Prenyl Transfer Reaction Identifies a Reaction Chamber with Modifiable Specificity. *J Am Chem Soc 132*, 17849-17858.
108. Metzger, U., Keller, S., Stevenson, C., Heide, L., and Lawson, D. (2010). Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. *Journal of molecular biology 404*, 611-626.
109. Luk, L., and Tanner, M. (2009). Mechanism of dimethylallyltryptophan synthase: evidence for a dimethylallyl cation intermediate in an aromatic prenyltransferase reaction. *J Am Chem Soc 131*, 13932-13933.
110. Luk, L., Qian, Q., and Tanner, M. (2011). A cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase? *J Am Chem Soc 133*, 12342-12345.
111. Joernvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., & Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, *34*(18), 6003-6013.
112. Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., ... & Jörnvall, H. (2003). Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chemico-biological interactions*, *143*, 247-253.
113. Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., ... & Oppermann, U. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry*, *277*(28), 25677-25684.
114. Kallberg, Y., Oppermann, U., Jörnvall, H., & Persson, B. (2002). Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes. *Protein Science*, *11*(3), 636-641.
115. Hummel, W. (1997). New alcohol dehydrogenases for the synthesis of chiral compounds. In *New Enzymes for Organic Synthesis* (pp. 145-184). Springer Berlin Heidelberg.
116. Liu, W., & Wang, P. (2007). Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnology advances*, *25*(4), 369-384.

117. Nobel, S., Abrahmsen, L., & Oppermann, U. (2001). Metabolic conversion as a pre-receptor control mechanism for lipophilic hormones. *European Journal of Biochemistry*, *268*(15), 4113-4125.
118. Gourley, D. G., Schüttelkopf, A. W., Leonard, G. A., Luba, J., Hardy, L. W., Beverley, S. M., & Hunter, W. N. (2001). Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites. *Nature Structural & Molecular Biology*, *8*(6), 521-525.
119. Thompson, J. E., Fahnestock, S., Farrall, L., Liao, D. I., Valent, B., & Jordan, D. B. (2000). The Second Naphthol Reductase of Fungal Melanin Biosynthesis inMagnaporthe grisea TETRAHYDROXYNAPHTHALENE REDUCTASE. *Journal of Biological Chemistry*, *275*(45), 34867-34872.
120. Thompson, J. E., Basarab, G. S., Andersson, A., Lindqvist, Y., & Jordan, D. B. (1997). Trihydroxynaphthalene reductase from Magnaporthe grisea: realization of an active center inhibitor and elucidation of the kinetic mechanism. *Biochemistry*, *36*(7), 1852-1860.
121. Kallberg, Y., Oppermann, U., Jörnvall, H., & Persson, B. (2002). Short-chain dehydrogenases/reductases (SDRs). *European Journal of Biochemistry*, *269*(18), 4409-4417.
122. Oppermann, U. C., Filling, C., Berndt, K. D., Persson, B., Benach, J., Ladenstein, R., & Jörnvall, H. (1997). Active site directed mutagenesis of 3β/17β-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. *Biochemistry*, *36*(1), 34-40.
123. Filling, C., Nordling, E., Benach, J., Berndt, K. D., Ladenstein, R., Jörnvall, H., & Oppermann, U. (2001). Structural role of conserved Asn179 in the short-chain dehydrogenase/reductase scaffold. *Biochemical and biophysical research communications*, *289*(3), 712-717.
124. Ghosh, D., & Vihko, P. (2001). Molecular mechanisms of estrogen recognition and 17-keto reduction by human 17β-hydroxysteroid dehydrogenase 1. *Chemico-biological interactions*, *130*, 637-650.
125. Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S., & Duax, W. L. (2001). Porcine carbonyl reductase structural basis for a functional monomer in short chain dehydrogenases/reductases. *Journal of Biological Chemistry*, *276*(21), 18457-18463.
126. Lavandera, I., Kern, A., Ferreira-Silva, B., Glieder, A., de Wildeman, S., & Kroutil, W. (2008). Stereoselective bioreduction of bulky-bulky ketones by a novel ADH from Ralstonia sp. *The Journal of organic chemistry*, *73*(15), 6003-6005.
127. Man, H., Kędziora, K., Kulig, J., Frank, A., Lavandera, I., Gotor-Fernández, V., ... & Grogan, G. (2014). Structures of Alcohol Dehydrogenases from Ralstonia and Sphingobium spp. Reveal the Molecular Basis for Their Recognition of 'Bulky–Bulky'Ketones. *Topics in Catalysis*, *57*(5), 356-365.
128. Lerchner, A., Jarasch, A., Meining, W., Schiefner, A., & Skerra, A. (2013). Crystallographic analysis and structure-guided engineering of NADPH-

dependent Ralstonia sp. Alcohol dehydrogenase toward NADH cosubstrate specificity. *Biotechnology and bioengineering*, *110*(11), 2803-2814.
129. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, gkq399.
130. Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., & Thomas, M. G. (2010). MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*, *49*(41), 8815-8817.
131. Stammers, D. K., Ren, J., Leslie, K., Nichols, C. E., Lamb, H. K., Cocklin, S., ... & Hawkins, A. R. (2001). The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. *The EMBO journal*, *20*(23), 6619-6626.
132. Kotaka, M., Johnson, C., Lamb, H. K., Hawkins, A. R., Ren, J., & Stammers, D. K. (2008). Structural analysis of the recognition of the negative regulator NmrA and DNA by the zinc finger from the GATA-type transcription factor AreA. *Journal of molecular biology*, *381*(2), 373-382.
133. Lamb, H. K., Leslie, K., Dodds, A. L., Nutley, M., Cooper, A., Johnson, C., ... & Hawkins, A. R. (2003). The negative transcriptional regulator NmrA discriminates between oxidized and reduced dinucleotides. *Journal of Biological Chemistry*, 278(34), 32107-32114.
134. Muro-Pastor, M., Gonzalez, R., Strauss, J., Narendja, F., & Scazzocchio, C. (1999). The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter. *The EMBO Journal*, 18(6), 1584-1597.
135. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., ... & Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D: Biological Crystallography*, *65*(6), 582-601.
136. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455-461.
137. Hilvert, D., Hill, K. W., Nared, K. D., & Auditor, M. T. M. (1989). Antibody catalysis of the Diels-Alder reaction. *Journal of the American Chemical Society*, 111(26), 9261-9262.