CHARACTERIZATION OF BIOSYNTHETIC ENZYMES OF THIOMARINOL AND OXYVINYLGLYCINES

Zachary D. Dunn

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill
in partial fulfilment of the requirements for the degree of Doctor of Philosophy
in the Department of Chemistry.

Chapel Hill
2018

Approved by:

Bo Li

Matthew R. Redinbo

Eric M. Brustad

David S. Lawrence

Albert A. Bowers

ABSTRACT

Zachary D. Dunn: Characterization of Biosynthetic Enzymes of Thiomarinol and Oxyvinylglycines
(Under the direction of Bo Li)

Natural products produced by bacteria have been an excellent source of bioactive small molecules. The diverse structures and bioactivates of natural products make them useful compounds in multiple fields, ranging from biochemistry, to medicine, to agriculture. Two interesting classes of enzymes that produce a diverse set of natural products are the assembly line like proteins nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). The thiomarinol pathway contains both of these classes of enzymes, and I characterized the enzymes from this pathway responsible for linking two separate NRPS and PKS antibiotics into the hybrid antibiotic thiomarinol. The biosynthetic scheme utilized by the thiomarinol pathway represents a novel strategy to make NRPS/PKS hybrids in nature. I also developed a new method for studying NRPS systems using the molecular probe cysteamine to capture and detect intermediates produced during *in vitro* reactions. I then used this method to elucidate the biosynthesis of the NRPS pathway that produces the oxyvinylglycine L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB). Since cysteamine captures intermediates that are linked through a thioester bond utilized by both NRPS and PKS enzymes, future studies will be able to use the cysteamine *in vitro* reaction to characterize both classes of enzymes. Lastly, I identified two new oxyvinylglycines produced by the *gvg* gene cluster, which was already known to produce 4-formylaminooxyvinylglycine

(FVG). Through metabolomic and biochemical studies, I provided new insights into the biosyntheses of these three oxyvinylglycine metabolites and provided a background for future studies to characterize the activity of new enzymatic reactions from the *gvg* pathway. Together, this work characterizes and provides insights into the activities of ten enzymes throughout three separate pathways, which will aid in the characterization of homologous pathways found throughout nature. A better understanding of these natural product pathways and enzymes will afford the ability to modulate the pathways to produce non-natural products with altered bioactivities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2,4-DA	(2*E*,4*E*)-2,4-decadienoic acid.
2,4-DDA	(2*E*,4*E*)-2,4-dodecadienoic acid
α-KG	α-ketoglutarate
A	Adenylation
ACN	Acetonitrile
ACP	Acyl carrier protein
Ala	Alanine
AMB	L-2-amino-4-methoxy-*trans*-3-butenoic acid
AMP	Adenosine monophosphate
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
AOVG	Aminooxyvinylglycine
Arg	Arginine
AT	Acyltransferase
ATP	Adenosine triphosphate
AVG	Aminoethoxyvinylglycine
BLAST	Basic Local Alignment Search Tool
BME	β-mercaptoethanol
Boc	tert-butyloxycarbonyl
C	Condensation
CDD	Conserved Domain Database
CO	Carbon monoxide
CoA	Coenzyme A

DCM	Dichloromethane
DIPEA	*N*,*N*-diisopropylethylamine
DTP	Dithiolopyrrolone
DTT	Dithiothreitol
EA	Ethyl acetate
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
FAD	Flavin adenine dinucleotide
Fmoc-Cl	9-Fluorenylmethoxycarbonyl chloride
FP-LC	Fast protein liquid chromatography
FVG	4-Formylaminooxyvinylglycine
Glu	Glutamate
GOVG	Guanidinooxyvinylglycine
GTP	Guanosine triphosphate
gvg	Growth arrest factor vinylglycine
HATU	Hexafluorophosphate azabenzotriazole tetramethyl uranium
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
Hser	Homoserine
IAA	Iodoacetamide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KS	Keto-synthease
LC	Liquid chromatography

LC-HRMS	Liquid chromatography high resolution mass spectrometry
MOPS	3-(N-morpholino)propanesulfonic acid
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass spectrometry
MS^2	Tandem mass spectrometry
MT	Methyltransferase
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance spectroscopy
NRPS	Nonribosomal peptide synthetases
ORF	Open reading frame
O-SH	*O*-succinyl-L-homoserine
PAA	Pseudomonic acid A
PAC	Pseudomonic acid C
PCR	Polymerase chain reaction
PKS	Polyketide synthases
PLP	Pyridoxal 5'-phophate
PMS	*Pseudomonas* minimal salts media
Ppant	Phosphopantetheine
Q-TOF	Quadrupole time of flight
SAM	*S*-adenosyl-L-methionine
SNAC	*N*-acetyl cysteamine
T	Thiolation

TCEP	tris(2-carboxyethyl)phosphine
TE	Thioesterase
TFA	Trifluoroacetic acid
TGH	tansglutaminase-like domain
THF	Tetrahydrofuran
UV	Ultraviolet
WT	Wildtype

CHAPTER 1: CHARACTERIZING SECONDARY METABOLITE PATHWAYS LEADS TO THE IDENTIFICATION OF NOVEL NATURAL PRODUCTS AND ENZYMATIC REACTIONS

Enzyme Introduction

Within living cells, thousands of chemical reactions are occurring simultaneously, and enzymes catalyze virtually all of these reactions. Thus, understanding the role and function of enzymes is paramount to understanding biological systems and processes. The first studies that attempted to characterize enzymes began over 100 years ago, when it was discovered that biological catalysis can occur outside of cells.[1] Since the first identification of these biological catalysts, countless studies have been performed to determine how they function and what roles they play in biochemical processes. From these studies many fundamental enzyme features have been discovered, such as the existence of cofactors and inhibitors, allosteric regulation, and the basics of enzyme kinetics.[2] One of the driving motivations behind these studies has been the desire to harvest the ability of enzymes to enhance reaction rates under relatively mild or physiologic conditions.

Compared to traditional organic catalysts, enzymes have many advantages. First, enzymes perform regiospecific and stereospecific reactions due to the chiral nature of their amino acid building blocks. Second, enzymes perform chemistry in relatively mild aqueous conditions compared to organic chemical syntheses, eliminating the need for organic solvents or the energy needed for heating or cooling of solutions. Third, enzymes perform chemistry with reactive intermediates in the presence of other functional groups.[3] Together these traits provide enzymes with the potential to be efficient and green alternatives to organic chemical

syntheses. In fact, many enzymes have already been used as biocatalysts in the synthesis of pharmaceuticals.[4–6] For example, a monooxygenase from *Rhodococcus* catalyzes the chiral hydroxylation of an indole precursor of Crixivan an HIV protease inhibitor. The monooxygenase installs two chiral hydroxyls that are required for activity in the final compound.[4,5] Also, enzymes are used to make modifications to bioactive molecules to improve pharmacokinetic profiles. For example, Ribavirin is an antiviral used in the treatment of hepatitis C, and the conversion of ribavirin into an improved prodrug is done by selectively modifying one of three hydroxyls of the parent compound. The prodrug has been produced and isolated in 82% yield on the kilogram scale using *Candidia antarctica* lipase B, eliminating the need for costly and time consuming selective protection and deprotections steps.[6] While some classes of enzymes are characterized and can be used in these syntheses, there are still many classes that are uncharacterized and have the potential to expand the biocatalytic toolbox. One source of uncharacterized enzymes are microbial pathways that produce molecules that are not necessary for reproduction but provide the organism a selective advantage; known as secondary metabolites.

Enzymes involved in microbe secondary metabolism have been a focus of many studies in recent years. The metabolites from these pathways represent a wealth of bioactive compounds that contain functional groups that are rarely found in primary metabolism, including diazo groups (azaserine), enediynes (calicheamicin), and thiosulfinates (leinamycin) among others (Figure 1.1). While these bioactive metabolites were first identified decades ago, as far back as 1954 in the case of azaserine,[7] the biosynthesis of many secondary metabolites have only recently been elucidated. The expansion of genome sequencing in the 1990's caused a paradigm shift in the field and allowed for new

bioinformatic techniques to analyze the genes responsible for secondary metabolite production. Prior to this period, it was challenging for researchers to identify a single gene of interest and purify the corresponding protein for *in vitro* studies. Now, with the availability of many microbial genomic sequences, computational programs have been developed that are able to connect isolated compounds to putative biosynthetic gene clusters, and vice versa.[8] In a few short years, the ability of investigators to characterize enzymes involved in microbial secondary metabolism took a giant step forward.

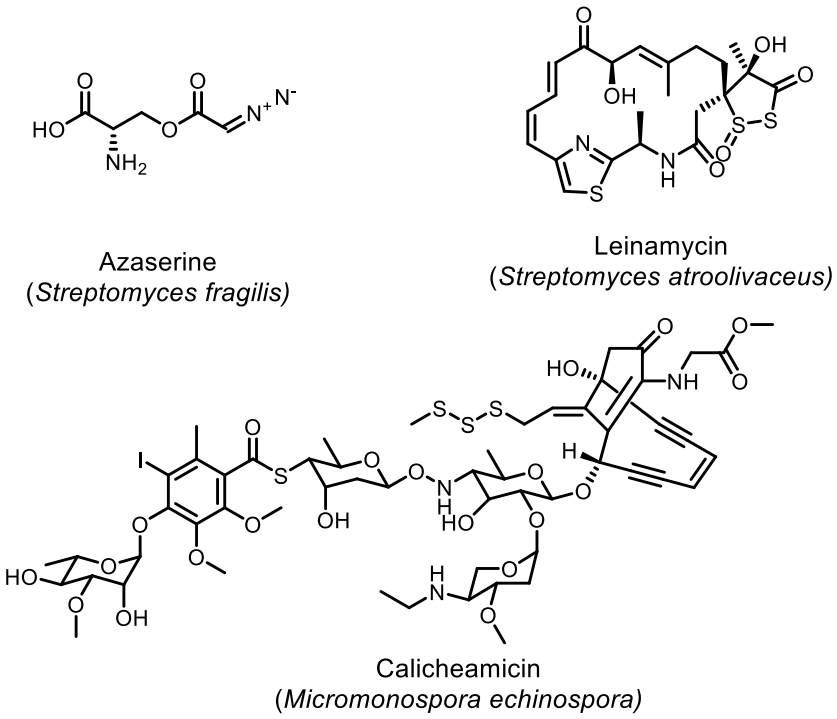


Figure 1.1 – Structures of microbial secondary metabolites that have interesting functional groups, with the producing strain for each listed.[7,9,10]

Genome guided strategies for characterizing secondary metabolite enzymes

In bacteria, the genes responsible for a single metabolic pathway often cluster together. This was originally observed in primary metabolism in the 1970's,[11] and it holds true for secondary metabolites. Therefore, researchers can search the genome of any

sequenced microbial strain to identify possible secondary metabolite gene clusters in order to discover the full potential of secondary metabolites each strain can produce. It is estimated that many actinobacteria can produce 20 to 30 different secondary metabolites, approximately 10 times the number identified in culture extracts.[12] Under the current paradigm, the general workflow to identify new secondary metabolites and characterize the enzymes in each pathway is shown in Figure 1.2.

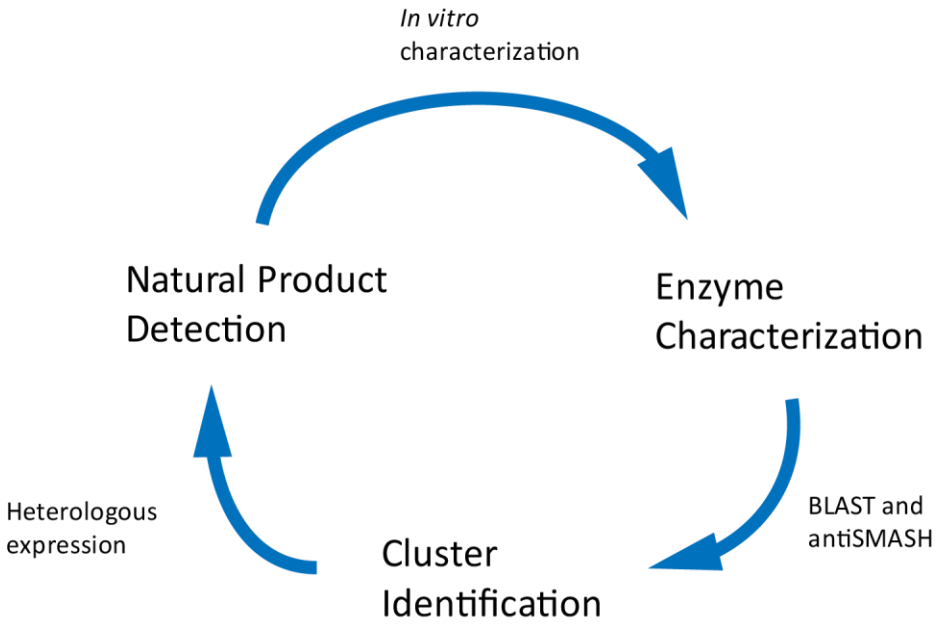


Figure 1.2- The workflow of characterizing enzymes of secondary metabolite natural product pathways.

The first step involves the identification of potentially new biosynthetic gene clusters by leveraging multiple computational software programs to parse through the extensive amounts of genomic data available. The two most commonly used computational search tools are Basic Local Alignment Search Tool (BLAST) and antibiotics & Secondary Metabolite Analysis Shell (antiSMASH).[13,14] BLAST uses a gene or protein sequence of interest, also known as a "biosynthetic hook", as a query to search any of the genomes available through NCBI. Combined with the collective information of the Conserved Domain

Database (CDD), putative enzymatic functions can be assigned to genes solely from the genetic or protein sequence.[15,16] As a complementary program, antiSMASH is able to identify all gene clusters within a given genome that are predicted to produce a secondary metabolite and compare them against other characterized clusters.[14] Using BLAST and antiSMASH together, researchers can predict a reasonable pathway for a given gene cluster if the product is known, without any additional experimental data. For some well-studied classes of secondary metabolites even the structure of the product can be predicted with some accuracy, however there are still limitations to these predictions.

The most successful predicted functions are from nonribosomal peptide synthetase (NRPS) and polyketide synthase (KS) containing gene clusters. Both of these classes of proteins are large modular assembly-line-like enzymes. Substrates are covalently tethered to the protein and undergo reactions by modifying domains and separate enzymes, called tailoring enzymes.[17,18] For PKS and NRPS pathways, even the substrates of the enzymes can be accurately predicted, which is a task that is difficult for other classes of enzymes. In nature, PKS enzymes use either malonyl-CoA or methylmalonyl-CoA as substrates which can be predicted by sequence similarities of certain domains.[17] NRPS enzymes are more difficult to predict since the 20 proteinogenic amino acids as well as hundreds of non-proteinogenic amino acids can be used as substrates. Although, the high structural similarity between NRPS domains allowed Stachelhaus et al. to identify 10 key residues that determine substrate specificity, known as the Stachelhaus code.[19] The Stachelhaus code is now incorporated into multiple free online programs that will predict NRPS amino acids substrates from genomic data, such as antiSMASH, NaPDoS, and NRPSPredictor.[14,20,21] This work was advanced further in 2016, when researchers were able to successfully connect

previously isolated PKS and NRPS products to uncharacterized gene clusters using only bioinformatics.[8] While bioinformatic analysis of NRPS and PKS pathways can accurately predict product scaffolds, many other classes of enzymes still need more experimental characterization to reach this level of knowledge. However, this body of work demonstrates how the characterization of a single enzyme can lead to a broader understanding of an entire class of enzymes with the aid of computational tools.

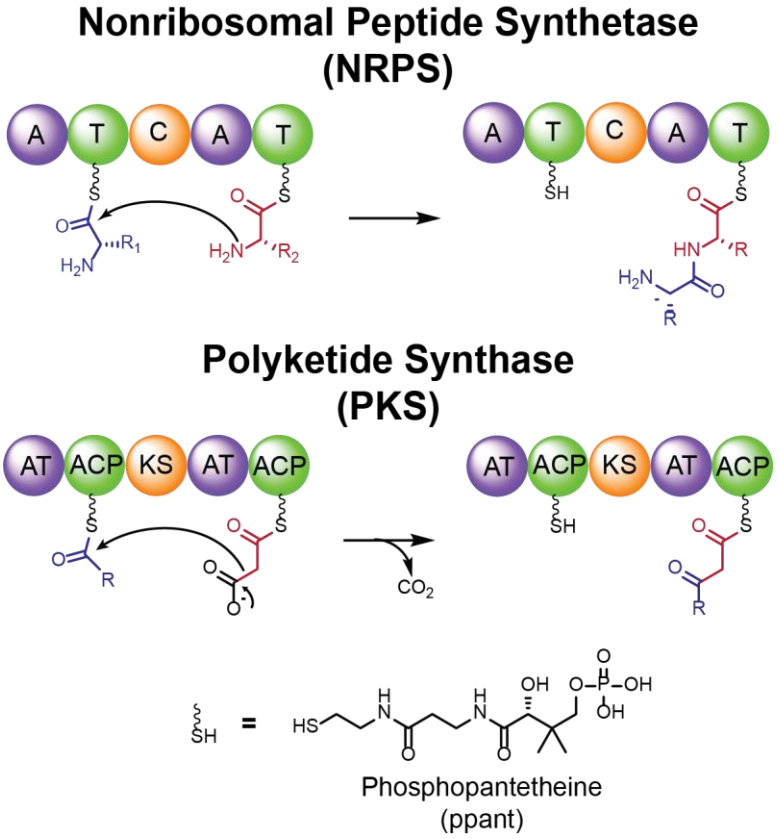


Figure 1.3 – Example domains and reactions of nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). The substrates are covalently linked to the proteins through the phosphopantetheine (Ppant) arm. The domains of the NRPS pathway are: adenylation (A), thiolation (T), and condensation (C). The domains of the KPS pathway are: acyltransferase (AT), acyl carrier protein (ACP), and keto-synthase (KS).

Bioinformatic analysis for most enzymes in secondary metabolism is still limited. While software such as BLAST is able to predict the enzyme class and cofactors, the native substrates and chemical transformations are usually difficult to predict. For example, in the mannopeptimycin pathway the *mppP* gene was predicted to encode a PLP (pyridoxal 5'-phosphate)-dependent aminotransferase.[22] However upon *in vitro* analysis of the enzyme, MppP was actually shown to act as an ornithine hydroxylase, which was a novel function for a PLP-dependent enzyme.[23,24] The characterization of MppP was then used to identify a homologue that performs similar chemistry within a different cluster.[25] The example of MppP highlights that each experimental characterization of an enzyme with unknown function helps aid the characterization of additional enzymes and the global understanding of enzymatic pathways. However, to accurately predict substrates and products similar to the NRPS and PKS systems using bioinformatics, many more enzymes of each class need to be characterized. Especially in secondary metabolism, many predicted proteins do not have close homologues that have been characterized, which can lead to incorrect functional assignments. Currently, the limitation of bioinformatically predicted functions from genomic data necessitate *in vitro* analyses to elucidate most natural product biosynthetic pathways.

Discovering the products of secondary metabolite gene clusters

The second step in the workflow for characterizing the enzymes of secondary metabolite natural product pathways is identifying the natural product of the proposed gene cluster being studied (Figure 1.2). To understand the functions of a class of enzymes in secondary metabolism one must first identify and characterize their native functions. Although, this is complicated by the fact that enzymes from the same class can perform a variety of chemical reactions using the same cofactors. Thus, without any information about

the native pathway, discovering the function of enzymes can be a difficult task. Therefore, as shown in figure 1.2, the most common starting point in characterizing the function of an enzyme from a gene cluster is the isolation and characterization of the final product of the pathway, which can be achieved using a variety of molecular biology and analytical chemistry techniques.

The isolation of microbial natural products is made difficult by the fact that the majority of these natural product pathways are not expressed under typical laboratory conditions. While some researchers have found that modifying culture conditions[26–28] or including multiple strains in a single culture can change the expression pattern of bacteria,[29,30] most investigators rely on molecular biology to identify natural products from target gene clusters. The two most common techniques are heterologous expression and gene knockouts. Each of these techniques has its advantages and disadvantages, and when used together can elucidate the final products of pathways as well as important intermediates.

Most native producers of natural products are difficult to work with in the laboratory. They either do not grow under laboratory settings, have low levels of expression of the desired pathway, or are not genetically tractable. Therefore, heterologous expression is used to overcome these obstacles by expressing the target cluster in an organism with desired traits. The most common strains used for heterologous expression are *Escherichia coli* for bacterial gene clusters and *Saccharomyces cerevisiae* for fungal gene clusters, since these are both well studied and compatible with most genetic manipulation techniques.[31] These strains can be transformed using plasmids or cosmids that contain the entire biosynthetic gene cluster. Additionally, during the cloning process the gene cluster can be put under the control of strong or inducible promoters in order to overexpress the enzymes of the pathway, leading

to greater production of the final product. The DNA assembler method is able to introduce promoters before each gene of a cluster in a single step to ensure equal overexpression of each gene.[32,33] Once the heterologous host is transformed with the desired gene cluster, culture extracts can be analyzed by NMR or LC-MS to detect new products and intermediates. For most pathways, heterologous expression leads to greater production of natural products than in the native organism, although some natural products are still not produced at all. These pathways may require substrates or cofactors not normally produced in the heterologous host, which will prevent the biosynthesis of the final product.[34] Thus the choice of heterologous host is crucial, and the decision is dependent on the enzymes of the pathway being expressed. Although heterologous hosts are useful for characterizing gene clusters, in certain circumstances it can be more beneficial to study pathways using knockouts in the native host.

When the native producer is genetically tractable, gene deletion and inactivation are powerful techniques to connect a gene cluster to its respective natural product or vice versa. In both cases, the function of the encoded protein is disrupted to prevent the production of the natural product. One common way to inactivate the genes responsible for the synthesis of a target natural product is transposon mutagenesis. Transposon mutagenesis randomly disrupts genes in the native producer thereby creating a library of mutants, each with a different gene inactivated.[35] The bottleneck in transposon mutagenesis is identifying the mutants that have the target pathway inactivated, which is often identified by phenotypic screens of thousands of mutants.[35,36] After identifying the mutants that no longer produce the target compound, sequencing of the transposon insertion also identifies the sequence of the gene that has been disrupted, thus elucidating a gene necessary for its biosynthesis.[35]

Alternatively, genetic knockouts can be a more efficient way of identifying the product of a target gene cluster. The first step of this two-step knock-out process utilizes homologous recombination; by creating plasmid vectors that can insert themselves into targeted locations in the bacterial genome by incorporating homologous regions of the targeted genomic sequence that have been amplified by PCR.[37,38] Then a second round of homologous recombination removes the plasmid along with a portion of the gene cluster leaving either a selective marker for screening or no insert, to reduce polarization effects on the surrounding genes.[37] By knocking out targeted genes in a cluster and then performing comparative metabolomics between the wild type and knock out strains, potential natural products can be identified.[39] Also, when genes responsible for late stage modifications are knocked-out, intermediates of the pathway can accumulate for easier detection. Together these two techniques, transposon mutagenesis and genetic knockout, can be extremely useful molecular biology techniques to connect gene clusters and their natural products, an important step in identifying the specific proteins worth further characterization.

***In vitro* characterization of biosynthetic enzymes**

Once a gene cluster is shown to produce a natural product, the next step is the characterization of the enzymes in the biosynthetic pathway (Figure 1.2). While many techniques have been used to characterize enzymes, the most definitive experiments are *in vitro* assays performed using purified proteins. By reproducing enzyme activity outside of cells, the presence and concentration of each individual component can be systematically controlled, and the necessary cofactors and substrates can be identified. Characterizing multiple enzymes of a natural product pathway in this way reveals the biosynthetic logic of converting simple building blocks into complex scaffolds. However, even when the purified

enzymes are active *in vitro*, the detection and identification of resultant products can still prove difficult.

Since the first observance of fermentation in the absence of cells in 1894,[1] the use of *in vitro* assays to characterize enzyme function has become increasingly common. Most initial methods for detecting enzyme activity were absorbance or radioactivity based. For example, as early as 1946, phosphatase activity has been measured using p-nitrophenyl phosphate, a non-native, colorless substrate that generates a yellow color after phosphate cleavage that can be readily detected using UV-Vis absorbance.[40] While the use of non-native substrates has been useful in characterizing some enzymes, many enzymes are not capable of reacting with non-native substrates. As an alternative, many *in vitro* assays have been developed that measure byproduct formation in order to detect activity. For example, peroxidase enzymes create reactive oxygen species from hydrogen peroxide, which in turn are able to react with exogenously added reagents sensitive to oxidation. These added reagents, such as 3-methyl-2-benzothiazolinone hydrazine and 3-(dimethylamino)benzoic acid, produce colored products when oxidized that can be used to follow the progress of the enzyme reaction using UV absorbance.[41] Although these byproduct detection strategies allow for fast analysis and activity detection of enzyme activity, the use of non-natural substrates and the indirect measurement of enzyme activity are significant limitations of these approaches. In recent years, advancements in liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance spectroscopy (NMR) methods have allowed for direct detection of the products of enzymatic reactions.

Detecting enzymatic products directly eliminates the need for additional non-native reagents that may interfere with enzyme function during *in vitro* reactions. Entire pathways

have been reconstituted in one pot assays *in vitro*, where reaction intermediates as well as the final products can be detected by LC-MS.[42–45] Additionally, by sequentially removing individual cofactors and enzymes from the assay, researchers can determine the necessary components of each chemical transformation. This approach allows for the elucidation of the roles of each enzyme in a pathway from one set of assay conditions. Even in cases where detection of products is difficult, the addition of derivatizing agents can improve both product/byproduct separation by LC, and detection by MS. Especially for small polar compounds, such as amino acids, derivatization allows for better detection of low abundance products.[46,47] For example, O-phthalaldehyde has been used to derivatize primary amines to improve separation on reverse phase columns and give compounds UV absorbance.[46] Also, derivatizing agents are typically added after the enzymatic assays are quenched, so that they do not interfere with enzyme activity. Additionally, chemical probes can be used to release covalently bound intermediates from enzymes to allow the characterization of individual steps and reaction intermediates of a single enzyme. Most commonly, this strategy has been utilized for NPRS and PKS systems, as the substrates of these pathways are covalently bound to the enzymes and undergo multiple sequential reactions before the release of the final product. For example, the nucleophile cysteamine mimics the natural phosphopantetheine arm of PKS and NRPS systems (Figure 1.4) and was used as a chemical probe to release intermediates that were covalently linked to the PKS CalE8 of the calicheamicin pathway. CalE8 was heterologously expressed in *E. coli* and purified with intermediates covalently linked, which were cleaved by cysteamine for analysis by LC-MS.[48] Essentially chemical reactions that occurred *in vivo* were able to be identified using a chemical probe *in vitro.* In addition to being useful in characterizing enzyme function, chemical probes and substrate

analogues have also been useful tools in understanding the chemical mechanism of enzyme reactions.

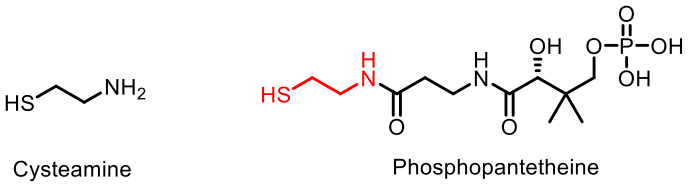


Figure 1.4 – Comparison of the chemical probe cysteamine and the posttranslational modification phosphopantetheine.

Elucidating Enzyme Mechanisms using Chemical Probes

In order to fully characterize an enzyme, it is also necessary to determine the chemical mechanism of the reaction. While the products and byproducts of enzyme reactions can be observed by LC-MS, often there are multiple reasonable chemical mechanisms by which those products can be produced. Therefore, further studies are necessary to determine which mechanism is actually correct. Elucidating an enzyme mechanism allows researchers to understand the possibilities and limitations of an enzyme as a biocatalyst for organic schemes. One of the oldest and most common techniques to probe the mechanism is to use isotope-labeled substrates and cofactors to track the movement of atoms through the reactions.

As early as 1947 isotope-labeled substrates were used by Doudoroff *et al.* to observed the exchange between radioactive ^{32}P enriched inorganic phosphate and the unlabeled phosphate of glucose-1-phosphate by sucrose phosphorylase.[49] By observing the rapid exchange of the radioactive ^{32}P into glucose-1-phosphate, it was determined that there must be a glucosyl-enzyme intermediate. While many of the first isotope labeling studies used radioactive isotopes, more recent studies use stable isotopes and detect exchange of atoms

using MS. For example, the source of the oxygen for the hydroxylation of arginine performed by MppP was determined by separately testing activity in the presence of either $^{18}O_2$ or $H_2{}^{18}O$. Only in the presence of $H_2{}^{18}O$ was ^{18}O incorporation into the product ever observed, meaning the hydroxyl in the product was from water, and not an activated oxygen species as previously hypothesized.[23] A similar method can also be used to determine regioselectivity of modifications, and allows for the detection of mechanistic steps that occur but are not present in the final product. One of the first examples used ATP labeled with ^{18}O at the phosphate linking oxygens to detect where cleavage of ATP occurs in the glutamine synthetase mechanism. By using differentially labeled isotope substrates, the reaction was shown to proceed through the addition of inorganic phosphate onto glutamate, since cleavage of ATP occurred between the beta and gamma phosphates before ATP regeneration.[50] Similarly, the oxidative decarboxylation activity of the nonheme oxidase UndA was revealed using differentially labeled lauric acid substrates. Substrates enriched with either deuterium or ^{13}C were used to definitively determine which atoms were lost through catalysis to understand the complex mechanism involving radical chemistry.[51] In addition to isotope enriched substrates, non-natural substrates can be used to further dissect the chemical transformations that occur.

By using non-native substrates that have altered functional groups, the course of an enzymatic reaction can be slightly changed to give more information about the chemical mechanism. Even in complex radical *S*-adenosyl-L-methionine (SAM)-dependent mechanisms non-native substrates have been used. In the biosynthesis of molybdopterin, MoaA is a radical SAM enzyme that uses GTP as a substrate. By using both chlorinated and deoxy analogues of GTP, products with different ring formations were observed due to the

trapping of early and late stage radical intermediates. The insight gained from the two non-native products allowed researchers to propose a mechanism for the native function of MoaA.[52] Using chemical probes and unnatural substrates to understand radical enzyme mechanisms truly shows the utility of these types of experiments. With clever experimental design, and access to analogues of substrates and intermediates, researchers can now understand enzyme functions down to the chemical mechanism, which unlocks new opportunities for using enzymes in future synthetic applications.

Summary and Perspectives

Enzymes from secondary metabolite pathways perform some of the most diverse and interesting chemistry found in nature and many have the potential to be useful in establishing efficient and green alternatives to standard organic chemical syntheses. The diversity of products synthesized by these enzymatic pathways exemplifies the wide range of chemical functional groups that are synthesized in secondary metabolism (Figure 1.1). Importantly, from a research perspective, the biosyntheses of several of these functional groups are still unknown because many of the enzymatic pathways that produce them are still uncharacterized. Even with the increase of genomic guided studies it is difficult to accurately predict activities for multiple enzyme classes. The exhaustive work done characterizing NRPS and PKS systems has led to the development of several bioinformatic prediction models, but even the these systems are limited to the canonical NRPS and PKS pathways.[8] Thus, the characterization of additional enzymes in PKS and NRPS pathways, as well as other enzymatic classes, will be needed to expand the scope of information used to support homology searches to lead to better predictive functions for future enzyme characterization studies. Also, characterizing additional classes of enzymes will expand the biocatalytic

toolbox available for semi-synthetic routes. Being able to employ well-characterized enzymes to perform difficult chemistry with almost absolute stereochemistry represents a green alternative to typical organic synthesis reactions. The work described herein leverages the iterative cycle shown in Figure 1.2 to advance the field of characterizing secondary metabolite pathways in multiple ways. First, the characterization of an enzyme pair that uses a novel strategy to produce an NRPS/PKS hybrid antibiotic (Chapter 2). Second, the use of new methods to study non-canonical NRPS pathways (Chapter 3). And finally, identification of multiple products from a single gene cluster and insights into the biosynthesis of these metabolites (Chapter 4).

CHAPTER 2: CHARACTERIZATION OF THE LATE STAGE MODIFICATION ENZYMES IN THE BIOSYNTHESIS OF THE HYBRID ANTIBIOTIC THIOMARINOL

Introduction

In recent years, the increasing prevalence of antibiotic resistant infections has underscored the critical need for the development of new antibiotics.[53,54] While significant efforts have been put into searching for novel molecular scaffolds that are able to inhibit unexplored cellular targets, the generation of hybrid antibiotics represents an alternative approach to this growing problem. Hybrid antibiotics combine the antibiotic warheads from two different molecules into one compound via covalent linkages. These hybrid antibiotics tend to have greater bioactivity compared to each individual component, and although combination drug therapies are frequently employed in cancer treatment, such combination treatments are rare in antimicrobial therapies. It is postulated that hybrid antibiotics would reduce the likelihood for resistance to develop as well as expand the spectrum of susceptible bacteria over that of either individual component.[55,56] However, there are potential challenges to generating these molecules synthetically. For example, it is possible that the covalent linker used to combine the two antibiotics will hinder the ability of one or both of the warheads to bind to their targets. Additionally, if the targets of each warhead are not located in the same cellular compartment, the effective concentration of each warhead would be reduced.[57,58]

While these are significant issues for synthetically generated hybrid antibiotics, naturally occurring molecules have been refined by nature to minimize these issues. One

example is thiomarinol, which is produced by *Pseudoalteromonas* sp. SANK73390. Thiomarinol is a combination of marinolic acid, which is structurally similar to pseudomonic acid A (PAA), the active molecule of the FDA-approved drug mupirocin, and holothin, which belongs to the dithiolopyrrolone (DTP) class of antibiotics.[59] PAA is known to inhibit bacterial isoleucine-tRNA synthetase and thereby prevent protein synthesis,[60] but the mechanism of action of holothin is not as well characterized. Holomycin, the acetylated form of holothin, is thought to inhibit bacterial transcription,[61,62] and recently the method of action was found to be through the disruption of metal homeostasis, and inhibition of a subset of metalloenzymes.[63] Interestingly, the isoleucine-tRNA synthetase target of PAA contains zinc in the active site of one of its domains.[60] Even though individually these antibiotics affect different pathways, the combined hybrid is more effective against a variety of bacterial strains, including methicillin-resistant *Staphylococcus aureus* (MRSA).[59]

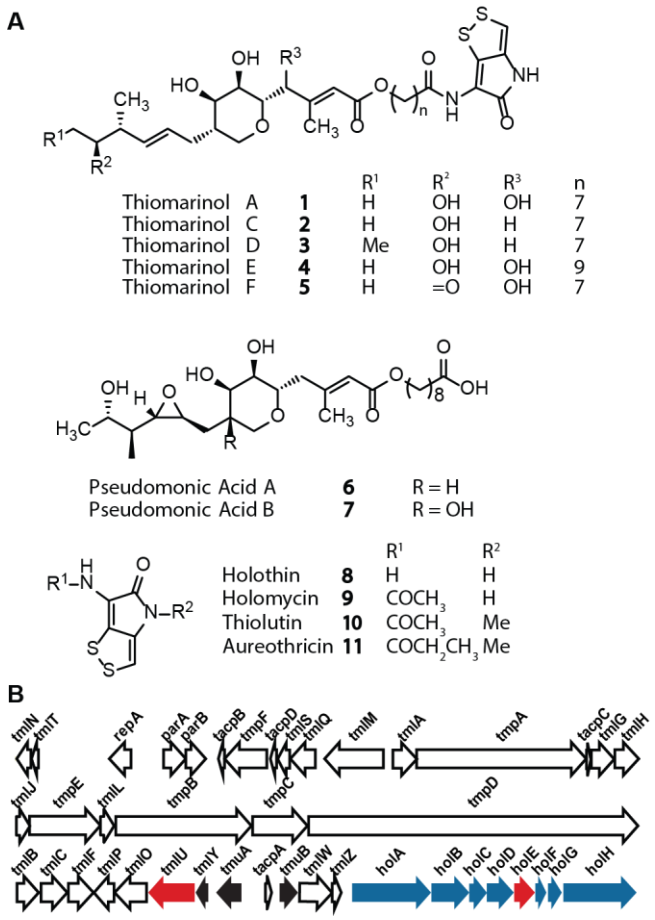


Figure 2.1 – Structure of thiomarinol analogues and components and the thiomarinol gene cluster. A) Structures of thiomarinols, pseudomonic acids, and dithiolopyrrolones. B) Gene cluster for thiomarinol. Open arrows indicate open reading frames (ORFs) with homology to the mupirocin pathway; blue ORFs are homologous to DTP biosynthetic genes; black ORFs are unique to the thiomarinol pathway; the red ORFs, TmlU and HolE, are the targets of this study and have counterparts in the mupirocin and holomycin pathways, respectively.

The gene cluster responsible for thiomarinol production from *Pseudoalteromonas* sp. SANK73390 was identified via whole genome sequencing of the producing bacterium. The cluster is harbored by a 97 kb cosmid that contains polyketide synthase (PKS) genes responsible for the production of marinolic acid, and nonribosomal peptide synthetase (NRPS) genes responsible for the production of holothin. Previous genetic knockout studies

have identified the enzyme TmlU (WP_013933291.1) as a necessary component in linking the two antibiotic warheads of thiomarinol. When the gene *tmlU* is deleted, thiomarinol is no longer produced, and instead both marinolic acid and holothin are produced.[64] Previous studies have also identified TmlU as a putative amide-ligase due to its homology to NovL (20.7%), CouL (21.1%), and SimL (18.5%), amide-ligases found in gene clusters responsible for the biosynthesis of aminocoumarin-containing antibiotics.[65–67] Each of these characterized amide ligases activate a carboxylic acid by creating an acyl-AMP intermediate before an amine attacks the activated ester to create an amide bond, however, this mechanism has not been proven for TmlU.

Instead, it is possible that TmlU work through a different mechanism. The thiomarinol cluster also contains HolE (WP_013933302.1), which shares homology with HlmA of the holomycin pathway from *Streptomyces clavuligerus.*[68,69] In holomycin synthesis, HlmA acetylates the amine of holothin to create holomycin using acetyl-CoA as a substrate.[68] The gene encoding HolE is found near the NRPS and tailoring genes responsible for holothin production, but was not predicted to have a specific role in holothin or thiomarinol biosynthesis; and instead it was thought to be responsible for only the background acylation of holothin. Here we considered the possibility that HolE may work with TmlU in linking together the two antibiotic warheads encoded by this cluster, which would be a novel strategy to produce NPRS/PKS hybrids in nature.

Materials and Methods

TmlU Expression and Purification

N-His TmlU was expressed in BAP1 *E. coli* cells in LB Miller media containing 50 µg/mL kanamycin sulfate. All chemicals were purchased from Fisher Scientific unless

otherwise noted. The cultures were grown at 37 °C and when the OD_{600} reached 0.6 they were cooled on ice, induced with 0.5 mM IPTG and grown for 18–20 hours at 16 °C. Unless otherwise stated, all purification steps were carried out at 4 °C or on ice. Cells were pelleted by centrifugation at 3000 $x\ g$ for 30 minutes. Pellets were resuspended in Buffer A (50 mM MOPS, 500 mM NaCl, 20 mM imidazole, 10 mM $MgCl_2$, 1 mM tris(2-carboxyethyl) phosphine (TCEP), 10% glycerol, pH 7.5) that contained complete EDTA-free protease inhibitor cocktail (Roche). The resuspension was lysed using a Fisher Scientific Sonic dismembrator model 500 (1.5 minute cycles of 30% amplitude with pulse on 0.5 seconds and pulse off 1.5 seconds) and centrifuged for 30 minutes at 34,500 $x\ g$ at 4 °C. The supernatant was filtered using a 0.45 µm filter, and then applied to a GE HisTrap HP Ni column which was equilibrated in Buffer A. His-tagged TmlU was eluted using a linear gradient of Buffer B (50 mM MOPS, 500 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$, 1 mM TCEP, 10% glycerol, pH 7.5). Pooled fractions were concentrated with a Millipore centrifugal filter with a 10 kDa molecular weight cut off at 3500 $x\ g$ to a final volume of 2.5 mL. The sample was desalted with a PD-10 column pre-equilibrated in Buffer C (20 mM potassium phosphate, 400 mM NaCl, 5% glycerol, pH 7.0). The flow-through was then applied to GE HiLoad 16/600 Superdex 200 pg column that was equilibrated with Buffer C. The pooled fractions were concentrated using a Millipore centrifugal filter at 3500 $x\ g$ to a concentration of 2–5 mg/mL.

HolE Synthesis and Subcloning

The complete HolE gene was synthesized by Life technologies in a pMA-T plasmid. DH5α chemically competent cells were transformed with the constructed plasmid and grown overnight on LB ampicillin plates (100 µg/mL) at 37 °C. A single colony was picked and

grown overnight in a culture volume of 5 mL LB with 100 µg/mL ampicillin. Plasmid was isolated from the culture using the Zyppy Plasmid Miniprep Kit. Isolated pMA-T-HolE and pET28a were digested with NdeI and XhoI at 37 °C for 3 hours then separated on an agarose gel. The resulting bands of digested HolE and pET28a were excised and extracted using the Zymoclean DNA Recovery Kit. The purified HolE and pET28a digested fragments were combined in a 1:3 vector:insert molar ratio. This mixture was ligated overnight using T4 DNA Ligase at 4 °C. The ligation product was cleaned and purified by butanol precipitation and resuspended in 10 µL of distilled water, which was used to transform Rosetta chemically competent cells. The pET28a-HolE plasmid was isolated from the transformed Rosetta cells as above and confirmed by analytical digestion by NdeI and XhoI.

HolE Expression and Purification

HolE expression was carried out similarly to TmlU except that protein expression was induced with 0.1 mM IPTG. The same purification procedure was followed using modified Buffer A (20 mM potassium phosphate, 500 mM NaCl, 50 mM imidazole, 10 mM $MgCl_2$, 10% glycerol, pH 7.5) and modified Buffer B (20 mM potassium phosphate, 500 mM NaCl, 500 mM imidazole, 10 mM $MgCl_2$, 10% glycerol, pH 7.5).

In vitro Reconstitution of TmlU and HolE Activities

Each reaction was initiated by addition of TmlU and/or HolE to final concentrations of 1–5 µM in 20 mM potassium phosphate pH 7.5. The remainder of the assay components were 2 mM $MgCl_2$, 1 mM ATP, 1 µM CoA, 0.1–1 mM of acyl carboxylic acid substrate, and 0.1–1 mM of amine substrate. Enzymatic reactions were incubated for 5–20 minutes at room temperature and quenched by adding an equal volume of acetonitrile, inverting and incubating at −20 °C for at least 20 minutes. Samples were spun at 20800 $x\ g$ for 10 minutes

to remove protein precipitant. The supernatants were diluted 2-fold and analyzed using Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS. Samples were separated using a Kinetex 2.6 µm C18 150 mm column. Solvent A consisted of 0.1% formic acid in water, and solvent B consisted of 0.1% formic acid in acetonitrile. Positive ion mode ESI mass spectrometry was carried out using the following parameters: gas temperature 300 °C, drying gas 10 L/min, nebulizer 45 lb/in^2, fragmentor 175 V, skimmer 65 V.

TmlU Kinetics

Solutions were prepared from the same stock solutions as *in vitro* reconstitution described above. Each reaction was initiated by addition of TmlU to a final concentration of 5 nM in 20 mM potassium phosphate at pH 7.5. The rest of the assay components were 10 µM HolE, 2 mM $MgCl_2$, 1 mM ATP, 1 µM CoA, 2–100 µM of PAC or PAA or 0.1–5 mM octanoic acid, and 500 µM 3-aminocoumarin. Aliquots of the reaction were added to an equal volume of acetonitrile every three minutes for PAC and PAA, and every 5 minutes for octanoic acid. The samples were incubated at room temperature for 20 minutes and spun at 20800 $x\ g$ for 10 minutes to remove protein precipitants. The samples were then analyzed undiluted by LC-MS as described above.

HolE Acyl-CoA Substrate Scope

Each reaction was initiated by addition of HolE to a final concentration of 5 µM in 20 mM potassium phosphate at pH 7.5. The rest of the assay components were 2 mM $MgCl_2$, 200 µM Acyl-CoA substrate, and 100 µM holothin. Samples were diluted 2-fold and analyzed as described above. These enzymatic reactions were incubated for 10 minutes at room temperature and quenched, centrifuged, and analyzed as described above.

Results

TmlU acts as an CoA-ligase and HolE acts an amide-ligase to produce thiomarinol

The N-terminal His-tagged form of TmlU was overexpressed and purified to $\geq$ 95% purity. When pseudomonic acid C (PAC) and holothin were incubated with TmlU and ATP, no product was formed. Further examination of the thiomarinol and mupirocin biosynthetic pathways showed that TmlU shared 14% homology with MupU, which suggested that TmlU may act as a CoA-ligase to make an activated intermediate that another gene in the cluster would use as a substrate. The putative amide ligase HolE was tested with TmlU, because it is also found in the thiomarinol gene cluster, and its role in the biosynthesis of thiomarinol had yet to be identified. When both TmlU and HolE were incubated with PAC, holothin, ATP, and CoA, the product PAC-holothin was detected using LC-MS (Figure 2.2A). The product had identical retention time and mass to the synthetic standard (Figure 2.2B), and removal of ATP, CoA, or either enzyme from the assay abolished PAC-holothin production. To confirm that TmlU and HolE were following the proposed mechanism, the PAC-AMP and PAC-CoA intermediates were identified using LC-MS (Figures 2.2B and D). The PAC-CoA complex would often be found either bound to either Al^{3+}, or Fe^{3+} in acidic conditions, or Na^{+} and K^{+} in basic conditions. However, under neutral conditions the $[M + H]^{+}$ species was observed (Figure 2.2B).

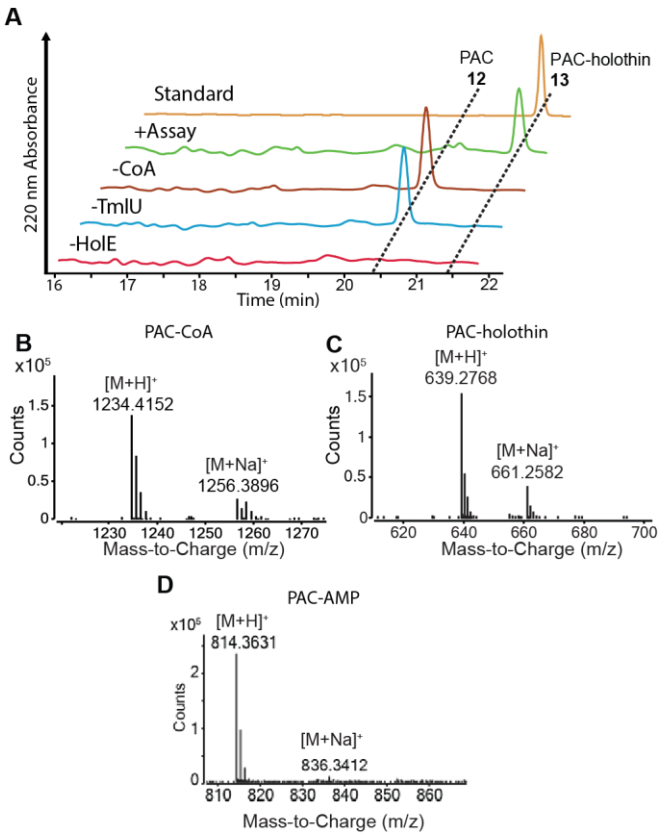


Figure 2.2 – Enzymatic production of PAC-holothin, a thiomarinol analogue, in vitro by TmlU and HolE. A) *In vitro* reconstitution of TmlU and HolE activity compared to a PAC-holothin synthetic standard and controls. B) Mass spectrum of PAC-CoA generated by TmlU (calculated [M+H]+, 1234.4155). C) Mass spectrum of PAC-holothin generated enzymatically by TmlU and HolE (calculated [M+H]+, 639.2768). D) Mass spectrum of PAC-AMP generated by TmlU (calculated [M+H] 814.3634)

TmlU has strict substrate specificity, while HolE is a promiscuous amide-ligase

After confirming the mechanisms of TmlU and HolE, we tested the substrate specificity of each enzyme. TmlU was found to be fairly selective and had only minor activity with substrates other than PAC and PAA (Figure 2.3A). However, HolE showed a greater promiscuity and demonstrated high levels of activity using a variety of acyl-CoA substrates, and moderate activity using amines that contained an adjacent ketone (Figure

2.3B and C). After determining the substrate specificity of each enzyme, the kinetics of each were probed. However, when HolE was incubated with high concentrations of holothin, HolE showed reduced activity, preventing determination of K_m and k_{cat} values. Kinetic studies of TmlU to produce the acyl-CoA product were probed by coupling this activity with an excess of HolE and 3-aminocouamrin in order to form the acyl-aminocoumarin product. Testing the kinetics of PAC found K_m and k_{cat} to be 6 ± 1 μM 3.2 ± 0.1 s^{-1} respectively (Figure 2.4A). Experiments using PAA as the substrate yielded similar values, K_m of 5.2 ± 0.5 μM and k_{cat} of 3.0 ± 0.1 s^{-1} (Figure 2.4B), but the affinity and efficiency of TmlU to activate octanoic acid was much lower (Figure 2.4C).

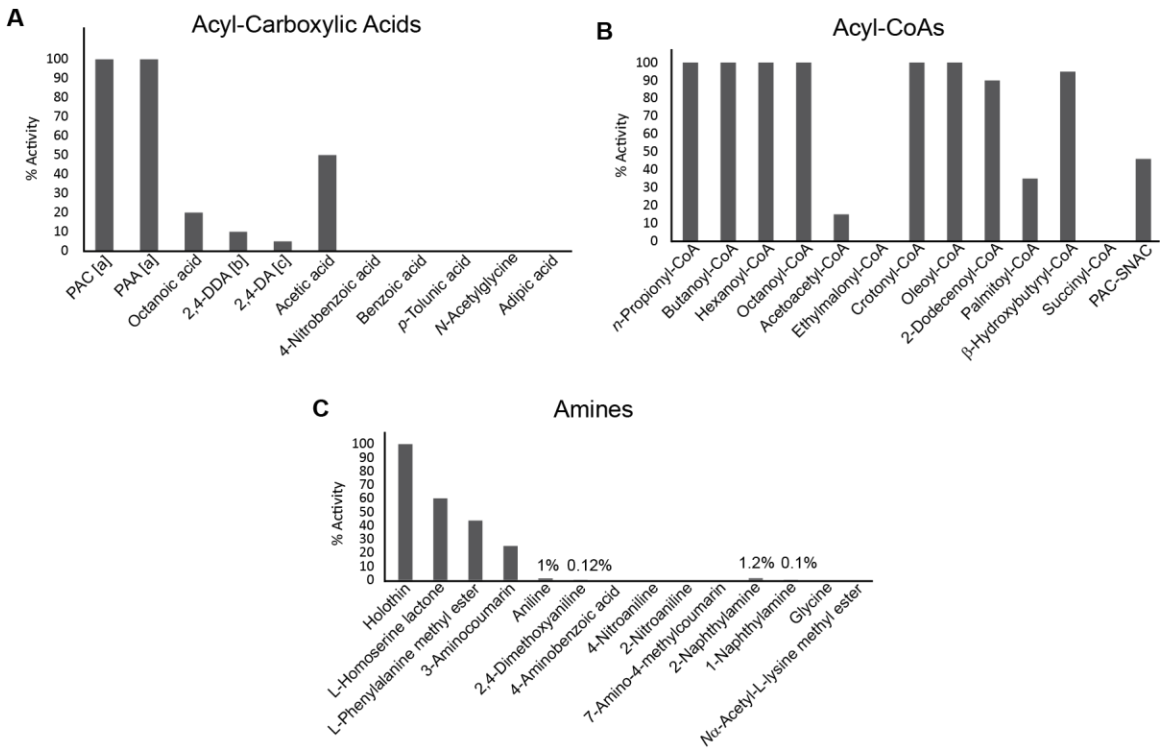


Figure 2.3 – Substrate promiscuity of TmlU and HolE. A) Incorporation of carboxylic acids into thiomarinol by 5 µM TmlU in the presence of saturating concentrations of HolE and holothin. Activity was measured by the formation of the acyl-holothin products at 360nm and normalized to activity with PAC as substrate. [a] 1 µM TmlU was used in this assay. [b] 2,4-DDA: (2*E*,4*E*)-2,4-dodecadienoic acid. [c] 2,4-DA: (2*E*,4*E*)-2,4-decadienoic acid. B) Activity of HolE with acyl-CoA derivatives as substrates to generate acyl-holothin products. PAC-SNAC - N-acetyl cysteamine pseudomonic acid C. C) Activity of HolE with different amines as substrates. PAC-CoA was generated by TmlU. Activity was measured by integration of ion intensities in mass spectra.

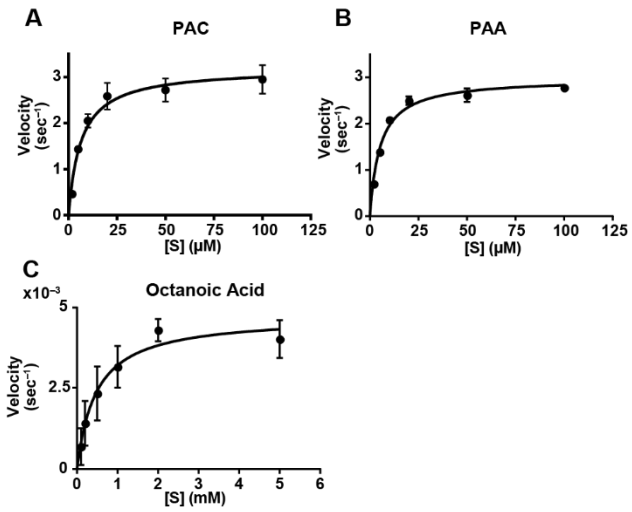


Figure 2.4 – Kinetic measurement of TmlU with different substrates. A) PAC, K_m=6±1 μM and k_{cat}=3.2±0.1 s^{-1}. B) PAA K_m=5.2±0.5 μM and k_{cat}=3.0±0.1 s^{-1}. C) Octanoic acid, K_m=0.5±0.1 mM and k_{cat}=(5.0±0.3) x 10^{-3} s^{-1}.

HolE is unable to produce a PAC-aminopenicillin hybrid

6-aminopenicillinic acid belongs to the beta-lactam family of antibiotics that inhibit bacterial cell wall biosynthesis, and we proposed it would be a substrate for HolE due to the similarities between 6-aminopenicillinic acid and holothin. However, when we included 6-aminopenicillinc acid in assays with HolE, we detected the production of the hydrolyzed product but not the PAC-6-aminopenicllin hybrid (Figure 2.5). Thus, the beta-lactam family of antibiotics appears to be incompatible with the HolE and TmlU assay to make new PAC containing hybrid antibiotics.

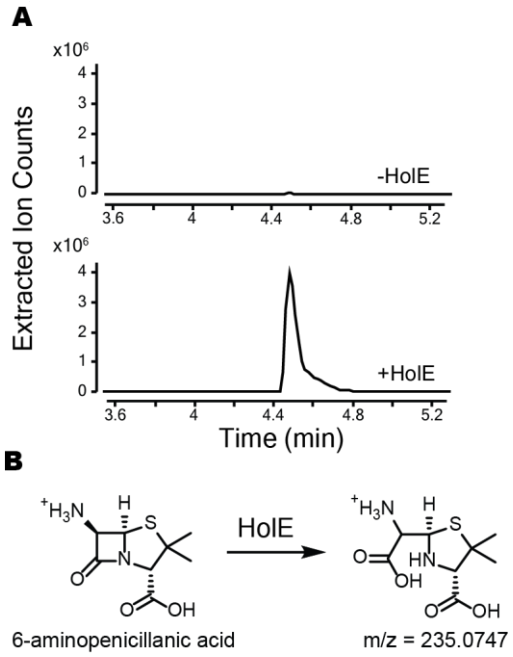


Figure 2.5 – HolE degrades 6-aminopenicillanic acid and is unable to form PAC analogue. A) Extracted ion chromatograms of the hydrolyzed 6-aminopenicillanic acid product, [M+H]+ 235.0747. B) Scheme of cleavage of 6-aminopenicillanic acid by HolE.

Discussion

Using *in vitro* reactions, we tested the hypothesis that TmlU acted as an amide ligase, and was the sole enzyme needed to link monic acid and holothin to create thiomarinol. In all *in vitro* reactions pseudomonic acid C (PAC) was used as a substrate instead of the native substrate marinolic acid, since PAC is much easier to synthesize and it only differs from the native substrate by one carbon in the acyl acid chain. When TmlU was incubated with holothin, PAC, and ATP, no PAC-holothin was detected, suggesting that TmlU is not an amide ligase and may instead have a different function. The presence of the putative amide ligase HolE in the thiomarinol cluster suggested that both HolE and TmlU may work together to produce thiomarinol. We hypothesized that TmlU may be an acyl-CoA ligase due to its homology to MupU, and that HolE would use an activated PAC-CoA intermediate for the

production of PAC-holothin (Figure 2.6). Incubation of HolE and TmlU with PAC, holothin, ATP, and CoA produced PAC-holothin. The identity of PAC-holothin was confirmed by comparing the elution time and m/z to a synthetic standard (Figure 2.2A and D). Removing HolE from the *in vitro* assay abolished PAC-holothin formation, but PAC was still consumed by TmlU, and the PAC-CoA intermediate was identified and confirmed by MS^2 (Figures 2.2B). Removal of CoA from the assay also resulted in the loss of PAC-holothin formation, and instead there was a slight accumulation of the PAC-AMP intermediate (Figure 2.2D). When either TmlU or ATP were not present there was no consumption of PAC; all of which supports the mechanism that TmlU acts as an acyl-CoA ligase to produce a PAC-CoA that HolE then uses as a substrate to create PAC-holothin (Figure 2.6).

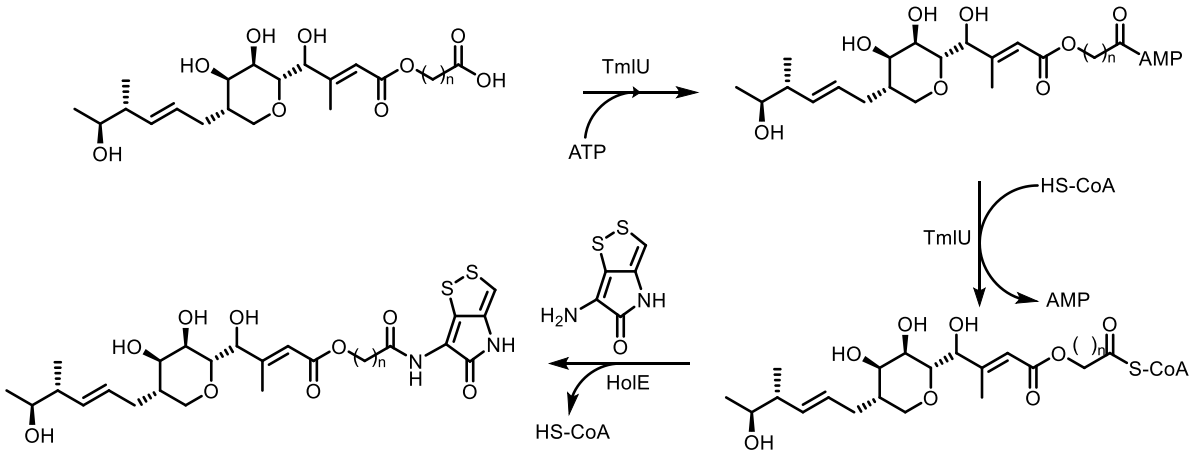


Figure 2.6 – Scheme of TmlU and HolE production of thiomarinol.

After the functions of both enzymes were determined, we investigated the kinetics of TmlU. The formation of acyl-CoA products by TmlU were measured in the presence of excess ATP and CoA. When PAC was used as substrate, TmlU had a K_m of 6 ± 1 µM and a k_{cat} of 3.2 ± 0.1 s^{-1} (Figure 2.4A). PAA had very similar values for K_m and k_{cat}, (Figure 2.4B) suggesting that TmlU is not responsible for the selectivity of *P. spp.* SANK73390 to produce thiomarinol without the epoxide, even though the corresponding pseudomonic acid is the

preferred product of the mupirocin pathway. Instead, we hypothesized that the thiomarinol pathway is missing the tailoring enzymes responsible for epoxide formation in PAA. We also tested the ability of TmlU to activate octanoic acid, to test the importance of the monic acid portion of PAC on TmlU activity. The catalytic efficiency (k_{cat}/K_m) of octanoic acid was 50,000-fold less than PAC or PAA, indicating that monic acid is necessary for greater TmlU activity (Figure 2.4C). To test if monic acid is able to activate TmlU by inducing a conformational change or if monic acid is important for substrate recognition, we measured the formation of octanoyl-CoA in the presence and absence of methyl monic ester. No difference in activity was observed, which indicates that interaction between monic acid and TmlU are involved in substrate recognition and not TmlU activation. We were unable to perform similar kinetic analysis of HolE because the native substrate holothin displayed inhibitory activity against HolE at concentrations above 20 μM.

The substrate specificity of each enzyme was investigated to test the potential ability of the TmlU/HolE system to produce new hybrid antibiotic molecules. In general, TmlU was found to be more selective than HolE. TmlU was only able to activate fatty acid chains and did not tolerate the presence of additional functional groups (Figure 2.3A). On the other hand, HolE was able to utilize many different acyl-CoA substrates, most with over 80% conversion, and only two of the acyl-CoA's tested had zero activity (Figure 2.3B). Additionally, HolE had significantly lower activity when incubated with the truncated N-acetyl cysteamine (SNAC) analogue of PAC-CoA that is missing the adenosine moiety, which suggests that the CoA portion is important for substrate recognition by HolE. To investigate amine specificity of HolE, a variety of scaffolds were tested with HolE, and the four compounds with greatest activity contained a cyclic structure with a carbonyl adjacent to

a primary amine (Figure 2.3C). The importance of the ring for substrate recognition was highlighted by the conversion of phenylalanine methyl ester being significantly better than that of glycine. A variety of aniline derivatives had slight activity, but all were far less than ketone containing substrates and would require modifications of HolE to make substantial amounts of product. Due to the inherent promiscuity of HolE, the enzyme pair of TmlU and HolE could be a useful system to create multiple non-natural PAC and PAA hybrid molecules, potentially with new bioactivities. Unfortunately, our attempt to use 6-aminopenicillanic acid as a substrate for HolE did not lead to hybrid formation, and instead 6-aminopenicillanic was hydrolyzed by HolE

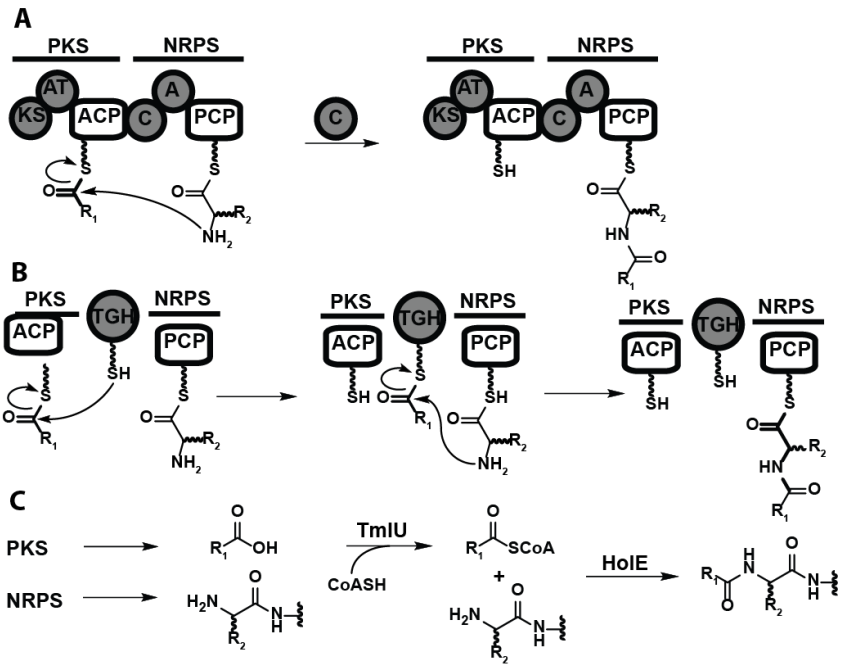


Figure 2.7 – Mechanism of thiomarinol production by TmlU and HolE compared to assembly-line tethered mechanisms to create PKS/NRPS hybrids. A) The condensation domain of a NRPS directly adds a PKS product to a growing peptide chain. B) A stand-alone transglutaminase-like domain (TGH) catalyzes the transfer between tethered PKS and NRPS products in the biosynthesis of andrimid.[70] C) Tailoring enzymes TmlU and HolE create a NRPS/PKS hybrid from released products of discrete PKS and NRPS pathways. KS: ketosynthase, AT: acyltransferase, ACP: acyl carrier protein, C: condensation domain, A: adenylation domain, PCP: peptidyl carrier protein.

Conclusion and Future Directions

The thiomarinol gene cluster is an interesting evolutionary product that appears to have combined two separate antibiotic pathways to produce a single hybrid antibiotic. The roles TmlU and HolE have been repurposed from the biosynthesis of either their respective PKS or NRPS pathways to instead catalyze the linkage of two bioactive molecules to produce thiomarinol, which has increased bioactivity and reduced antibiotic resistance. TmlU

and HolE utilize a novel strategy to produce NRPS/PKS hybrids, which has the advantage of using substrates that are free in solution instead of bound to an NRPS or PKS scaffold (Figure 2.7). Since both enzymes act on substrates free in solution, we were able to rapidly test various substrates for each enzyme *in vitro*. HolE was found to be a promiscuous amide ligase that was able to accept a variety of acyl-CoA and amine substrates. In this study, we were able to use the TmlU/HolE system to link two non-native substrates and create completely non-natural products. However, our attempt to create a PAC-penicillin hybrid was unsuccessful since HolE degraded the beta-lactam core of 6-aminopenicillanic acid. While beta-lactams were not compatible with HolE, it is possible that other amine-containing antibiotics could be used to make novel PAC containing hybrid antibiotics with improved bioactivity. Currently, the stringent specificity of TmlU limits the variety of acids that currently can be utilized to make hybrid molecules, the intrinsic promiscuity of HolE for various substrates is a valuable trait for future directed evolution and modification in order to incorporate a wider variety of amines into hybrid antibiotics.

CHAPTER 3: CHARACTERIZATION OF NONRIBOSOMAL PEPTIDE SYTHETASES BIOSYNTHETIC PATHWAYS USING CYSTEAMINE AS A MOLECULAR PROBE

Introduction

Two of the most well studied classes of secondary metabolite enzymes are nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). Both classes are large modular assembly-line-like enzymes that use a 4'-phosphopantetheine (Ppant) arm to covalently tether a growing natural product scaffold to the protein. In canonical NRPS pathways, each module will add one amino acid to the growing scaffold. The smallest complete module of an NRPS contains 3 domains: an adenylation (A) domain, a thiolation (T) domain, and a condensation (C) domain. The A domain activates the amino acid that is loaded onto the Ppant arm of the T domain, where the growing peptide chain can be elongated by the C domain and/or modified by tailoring enzymes in the pathway.[18] Tailoring enzymes are separate proteins from the NRPS that modify the natural product scaffold being constructed. The roles of tailoring enzymes range from cyclopropanations to hydroxylations to oxidation and reductions.[71–73] However, elucidating the function of tailoring enzymes is challenging since the native substrates for these enzymes are covalently linked to large NRPS enzymes. Additionally, previous studies have shown that many tailoring enzymes will not modify intermediates free in solution, and instead require the substrate to be linked to the native NRPS domain.[74] The difficulty of linking intermediates to the T domains of NRPSs has prevented the characterization of many classes of tailoring enzymes, some of which are predicted to perform novel chemistry never before seen in nature. While methods have been

developed to study tailoring enzymes, these techniques are typically cumbersome and require the chemical synthesis of putative intermediates.

One technique involves loading synthetic intermediates onto truncates of the native NRPS enzyme that consist of only a single T domain. This approach has two advantages. First, by loading the synthetic intermediate onto the T domain, the substrate more closely resembles the native substrate of the tailoring enzyme. Second, because the T domains are significantly smaller than intact NRPS enzymes, and the intact T domains, with covalently linked substrates and products, can be analyzed by LC-MS. Cryle *et al.* have successfully linked intermediates of the glycopeptide antibiotics onto relevant T domains; however, they only applied this approach to the study of one class of antibiotics and primarily focused on the modifications catalyzed by p450 enzymes.[74–76] The loading of synthetic intermediates onto T domains is accomplished by leveraging the substrate promiscuity of the 4'-phosphopantetheine transferase Sfp (Figure 3.1A), which will recognize a wide range of synthetic intermediates covalently linked to coenzyme A (CoA) as substrates (Figure 3.1 B). However, the direct synthetic coupling of the synthetic intermediates to CoA has two disadvantages: it requires the use of the expensive reagent CoA, and the coupling reaction can be hindered by the presence of functional groups on the target intermediate leading to low yields of the intermediate-linked CoA-analogue.[74,77]

One way to avoid the direct CoA-coupling is to synthesize a pantetheine analogue linked to the desired intermediate, and then use the CoA-forming enzymes CoaA, CoaD, and CoaE to produce the corresponding intermediate-linked CoA analogues. It has previously been shown that the corresponding CoA-analogues can be produced and loaded onto a T domain in a one-pot reaction (Figure 3.1 C).[78] In addition, this chemo-enzymatic approach

has also been used to load more stable amide-containing analogues, which are less susceptible to hydrolysis than the natural thioester-containing Ppant substrates.

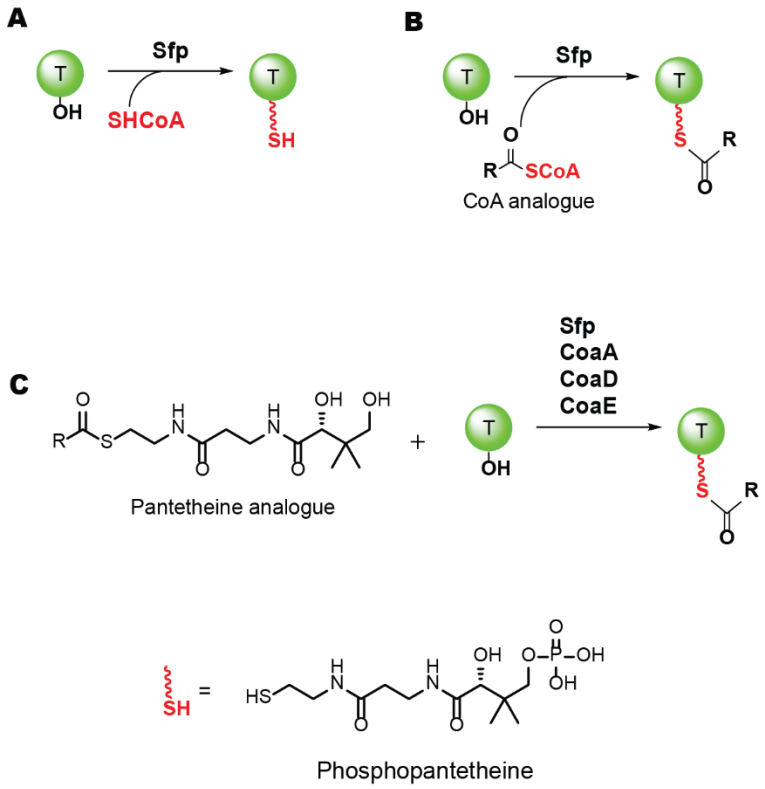


Figure 3.1 – Schemes to load intermediates onto T domains of NRPSs using the promiscuous 4'-phosphopantetheine transferase Sfp. A) The native activity of Sfp using coenzyme A (CoA) as a substrate B) The loading of a CoA-linked synthetic intermediate onto a T domain using Sfp. C) The loading of a pantetheine-linked synthetic intermediate in the presence of enzymes CoaA, CoaD, CoaE, and Sfp. R represents the structure of the intermediate being loaded.

The latter chemoenzymatic approach has been used to label T domains with fluorescent and chemical tags,[78,79] but to our knowledge it has never been used to load a biosynthetic intermediate onto a T domain for enzyme characterization. The modular nature of the synthesis of pantetheine analogues along with the promiscuity of the CoaA, CoaD,

CoaE, and Sfp enzymes mean that this approach is amenable to loading intermediates from various enzymatic pathways onto their native T domains. However, a significant barrier to using this method for enzyme characterization, especially for large and highly modified NRPS products, is the need to chemically synthesize each individual intermediate before subsequent enzymatic assays can be performed; and if the predicted substrate is incorrect, a new synthetic scheme must be devised and carried out before the next enzymatic assay can be performed. To avoid the need to chemically synthesize each intermediate, chemical probes can be used to cleave biosynthetic intermediates from NRPS enzymes.

While many chemical probes have been used to cleave native intermediates from NRPS enzymes, the effectiveness of each chemical probe varies among metabolic pathways. The development of a chemical probe method that is effective across most NRPS pathways would be a significant step forward for the field. Cysteamine is a chemical probe that was shown to successfully cleave intermediates from a PKS system for identification by LC-MS.[48] Cysteamine acts as a nucleophile that is able to cleave the thioester bond that links the natural product being synthesized to the PKS enzyme. Since the structure of cysteamine mimics the natural Ppant arm of both PKS and NRPS enzyme systems, it was hypothesized that cysteamine may be able to access the substrate binding site, and release intermediates from both PKS and NRPS. However, to date, the only example in the literature of cysteamine being used as a chemical probe to release intermediates from either PKS or NRPS enzymes studies the PKS CalE8.

The PKS CalE8 from the calicheamicin pathway was heterologously expressed in *E. coli* and purified with covalently linked intermediates that were cleaved by cysteamine for analysis by LC-MS.[48] This study provided a useful starting point for developing a method to

cleave intermediates using cysteamine during *in vitro* assays, which would expand the possible utility of cysteamine as a chemical probe. Being able to use cysteamine to cleave and identify intermediates produced from *in vitro* enzymatic assays would provide researchers a system for characterizing enzyme pathways and would establish cysteamine as a significant improvement versus previous cleavage methods. By controlling which enzymes and cofactors are included in the *in vitro* assays, the activity of both tailoring enzymes and the NRPS enzyme can be elucidated.

Two previously uncharacterized NRPS pathways were studied in this work to compare the effectiveness of different methods to characterize tailoring enzymes and non-canonical NRPS domains. One NRPS pathway that is predicted to have interesting tailoring enzymes is encoded by the *hlm* gene cluster from *Streptomyces clavuligerus*.[68] The *hlm* gene cluster has been shown to be necessary for the production of holomycin, which belongs to the dithiolopyrrolone (DTP) family of antibiotics. Members of the DTP family of antibiotics have been isolated from both soil and marine bacteria, and DTPs are effective against both Gram-positive and Gram-negative bacteria, which makes it a promising starting point for modification to create a clinically relevant analogue.[80] In studies of the *hlm* cluster, it has been shown that a cysteine dipeptide is loaded onto the T domain of the NRPS HlmE, but the chemical reactions necessary to form the aromatic ene-dithiol-disulfide core are unknown (Figure 3.2).[80] It is predicted that two putative flavoprotein tailoring enzymes, HlmB and HlmD, perform the oxidations necessary to produce holomycin, but the activity of these enzymes have not been characterized *in vitro*. In the proposed mechanism of holomycin biosynthesis, both enzymes are predicted to catalyze the formation of thioaldehydes, which would be a novel chemical transformation for enzymes.[68] The *hlm* gene cluster is just one

example of an NRPS pathway that is predicted to perform interesting chemistry, but the exact enzymatic reactions are still unknown.

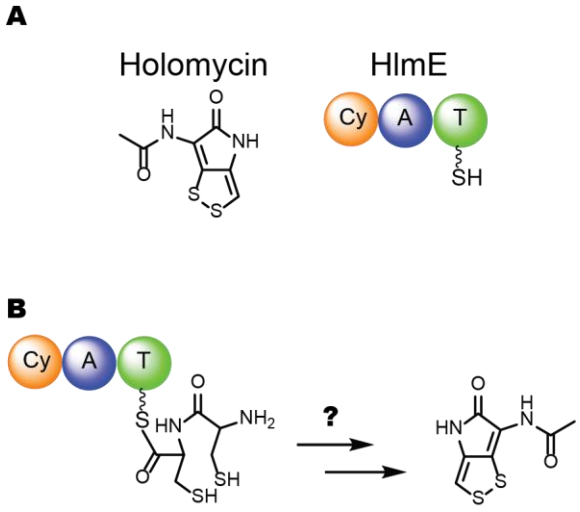


Figure 3.2 – Introduction to the biosynthesis of holomycin. A) The structure of holomycin and the domains of the HlmE enzyme. B) Enzymatic modifications convert the cysteine dipeptide loaded onto HlmE into holomycin, but the specific biosynthetic steps are still uncharacterized.

In addition to tailoring enzymes, the chemical diversity of NRPS products is expanded by noncanonical NRPS enzymes that do not follow the standard A-T-C domain structure described previously. One such example is encoded by the *amb* gene cluster from *Pseudomonas aeruginosa* PAO1, which produces the oxyvinylglycine L-2-amino-4-methoxy-*trans*-3-butenoic acid (AMB) (Figure 3.3). The oxyvinylglycines are a family of nonproteinogenic amino acids with various substitutions at the side chain ether (Figure 3.3). Oxyvinylglycines irreversibly inhibit PLP-dependent enzymes by covalently binding to the PLP cofactor.[81] The most well studied member of the family is aminoethoxyvinylglycine (AVG), which is used commercially as the active ingredient in ReTain® to halt the ripening of fruits by inhibiting ethylene biosynthesis in plants.[81] AMB inhibits the growth of multiple

strains of bacteria, including the human pathogen *Staphylococcus aureas.*[72] Although the bioactivities of oxyvinylglycines are potent and varied, few studies have examined the biosynthesis of these compounds. In the AMB biosynthetic pathway, the NRPS AmbE contains multiple domains that diverge from canonical NRPS logic (Figure 3.3). First, the second module of AmbE lacks an A domain to load an amino acid onto the second T domain (T2), Second, the C* domain has homology to C domains that are thought to catalyze modifications of the scaffold in addition to chain elongation.[72] Lastly, the N-terminal Q domain has no homology to any characterized NPRS domains and has an unknown role in the AMB pathway. The unusual domain architecture of AmbE makes it a challenging enzyme to study, but the elucidation of its biosynthetic steps will further the understanding of noncanonical NRPS pathways.

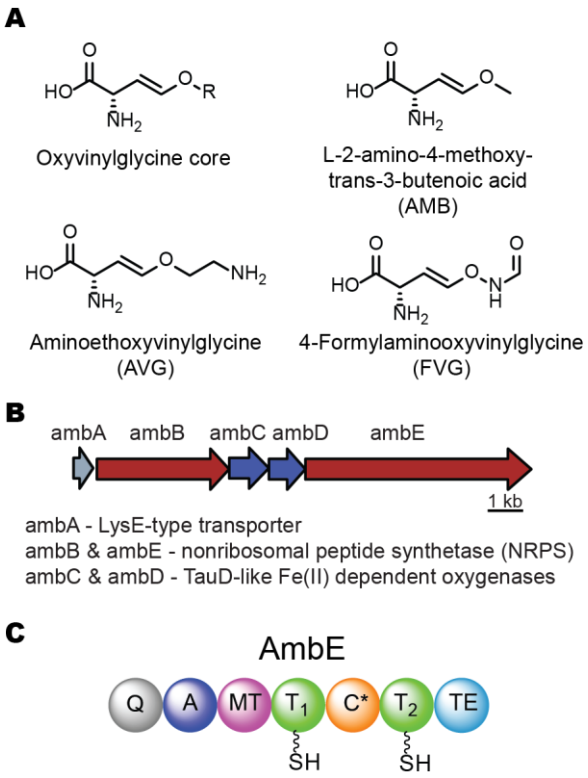


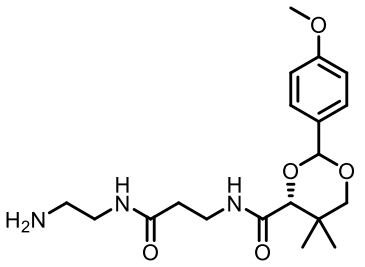
Figure 3.3 – Oxyvinylglycine core and amb gene cluster. A) The structures of L-2-amino-4-methoxy-trans-3-butenoic acid (AMB) and three other oxyvinylglycines found in nature. B) The *amb* biosynthetic gene cluster that produces AMB. C) The domains of the noncanonical NRPS AmbE: adenylation (A), thiolation (T), condensation (C*), thioesterase (TE), O-methyltransferase (MT), and a domain of unknown function (Q).

In a previous study the activity of the AMB pathway was reconstituted *in vitro*,[72] but many questions still remain. Murcia *et al.* used a mutant AmbE with an inactivated thioesterase (TE) domain, which is normally responsible for the release of products from NRPSs. When the AmbE TE mutant, AmbB, AmbC, and AmbD were incubated with substrates and cofactors, the assay accumulated an Ala-AMB-Ala tripeptide covalently linked to T2 of AmbE.[72] However, many questions about the pathway still remained, such as the roles of the tailoring enzymes AmbC and AmbD, and the noncanonical domains of

AmbE. Elucidating the full biosynthetic pathway of AMB will give a better understanding of how oxyvinylglycines are made by NRPS enzymes and provide additional insights into the mechanisms of non-canonical NRPS enzymes.

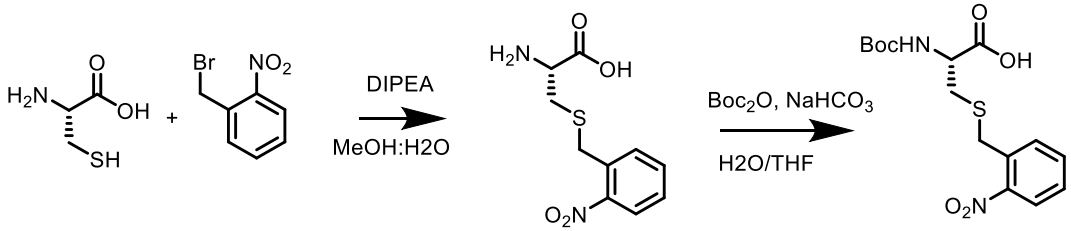
Materials and Methods

*Synthesis of protected amino pantetheine intermediate (**1**)*



3a

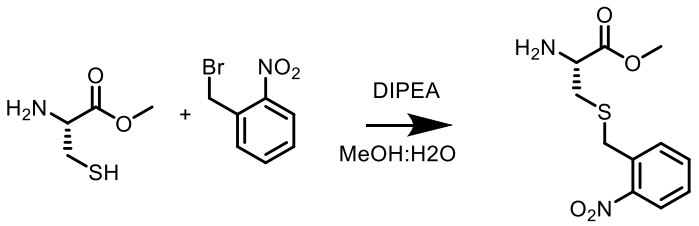
(4*R*)-*N*-(3-((2-aminoethyl)amino)-3-oxopropyl)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamide. 3a was synthesized following the procedure of Beld et al.[79]



1a

***N*-(*tert*-butoxycarbonyl)-*S*-(2-nitrobenzyl)-*L*-cysteine**. *S*-(2-nitrobenzyl)-*L*-cysteine was prepared according to previous literature (ref). *L*-cysteine (0.625 g, 5.15 mmol) was dissolved in H_2O (10 mL). *N*,*N*-diisopropylethylamine (DIPEA, 0.831 mL) was added to the solution dropwise on ice. To this reaction, a solution of 2-nitrobenzyl bromide (1.014 g, 4.7 mmol) in methanol (10 mL) was added dropwise on ice. The reaction was allowed to stir for 30 min before it was allowed to warm to room temperature. After 1 hour at room temperature, the precipitate was filtered, washed with ethyl acetate (2 x 10 mL), and

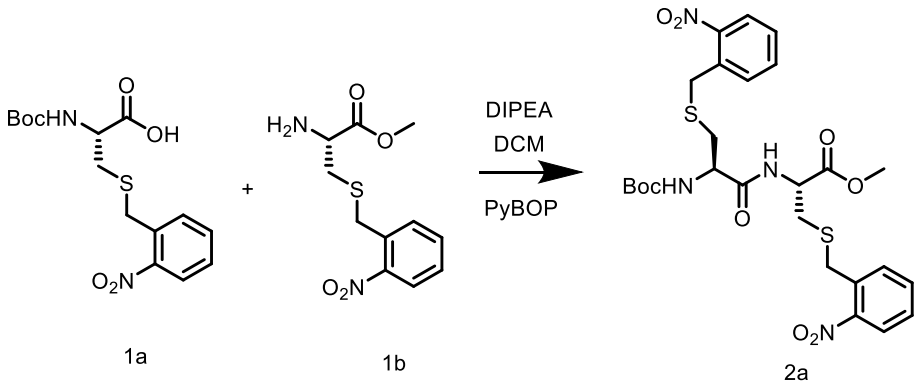
collected. The collected precipitate was then dried under reduced pressure to give a pale-yellow powder (0.773 g). This crude product was added to 5 mL of H_2O, and sodium bicarbonate (0.512 g, 6.1 mmol) was added to the solution. Boc anhydride (Boc_2O, 0.557 mL, 2.4 mmol) dissolved in THF (2 mL) was added dropwise on ice. After 30 min the reaction was allowed to warm to room temperature and stirred overnight. The reaction was then dried under reduced pressure and resuspended in ethyl acetate. The pH was adjusted to ~3.0, and then the solution was washed with brine. The organic layer was dried with sodium sulfate and then dried under reduced pressure to give a yellow oil (0.77 g, 46.0% yield over 2 steps). 1H NMR (400 MHz, Chloroform-d) δ 9.74 (s, 1H), 8.00 (d, J = 8.1 Hz, 1H), 7.57 (td, J = 7.6, 1.4 Hz, 1H), 7.51 – 7.40 (m, 2H), 5.37 (d, J = 7.3 Hz, 1H), 4.59 – 4.53 (m, 1H), 4.20 – 4.06 (m, 3H), 3.00 (dd, J = 14.0, 4.9 Hz, 1H), 2.90 (dd, J = 14.0, 5.6 Hz, 1H), 2.07 (s, 1H), 1.48 (s, 0H), 1.47 (s, 9H), 1.32 – 1.23 (m, 1H).



1b

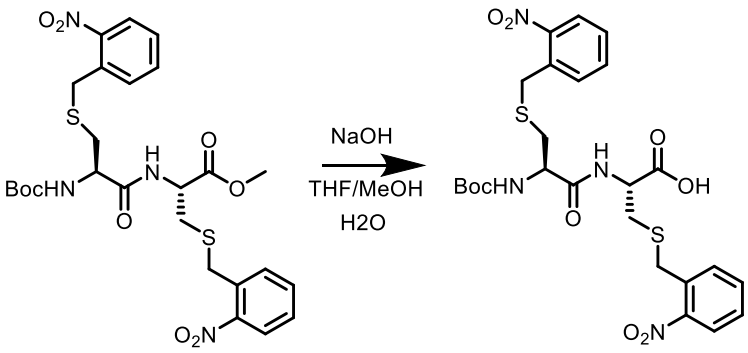
Methyl *S*-(2-nitrobenzyl)-*L*-cysteinate. *L*-cysteine hydrochloride (0.884 g, 5.15 mmol) was dissolved in H_2O (10 mL). DIPEA (1.662 mL) was added to the solution dropwise on ice. To this reaction, a solution of 2-nitrobenzyl bromide (1.014 g, 4.7 mmol) in methanol (10 mL) was added dropwise on ice. The reaction was allowed to stir for 1 h before it was allowed to warm to room temperature. The reaction stirred for 30 min at room temperature before the product was extracted with ethyl acetate (3 x 20 mL). The organic layer was dried over sodium sulfate then dried under reduced pressure to give a yellow oil (1.635 g). 1H NMR

(400 MHz, Chloroform-d) δ 7.92 (dt, J = 8.1, 1.7 Hz, 1H), 7.59 – 7.46 (m, 2H), 7.38 (ddd, J = 8.6, 6.8, 2.1 Hz, 1H), 4.11 (s, 2H), 4.17 – 3.99 (m, 2H), 3.82 (dd, J = 7.2, 4.7 Hz, 1H), 3.68 (d, J = 3.0 Hz, 3H), 2.92 (dd, J = 13.8, 4.8 Hz, 1H), 2.85 – 2.74 (m, 1H), 1.20 (td, J = 7.1, 6.4, 3.3 Hz, 1H).



1a 1b 2a

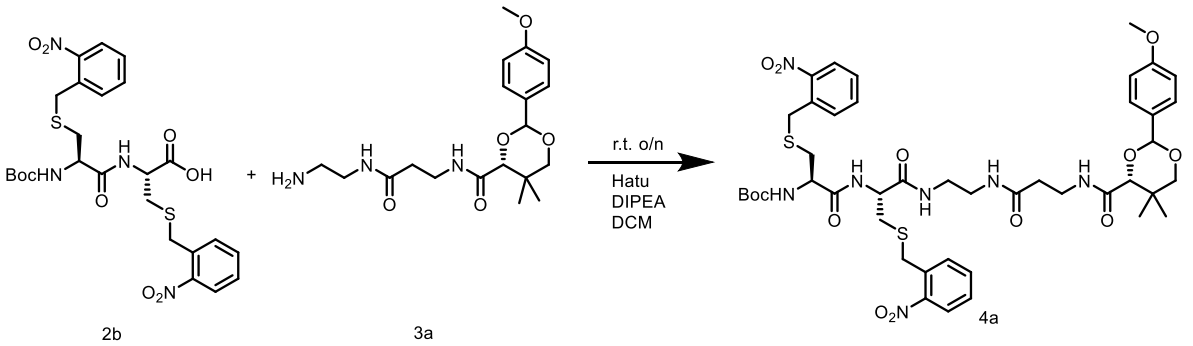
Methyl *N*-(*N*-(tert-butoxycarbonyl)-*S*-(2-nitrobenzyl)-*L*-cysteinyl)-*S*-(2-nitrobenzyl)-*L*-cysteinate. Compound 1a (0.536 g, 1.5 mmol) was dissolved in DCM (11.2 mL). The solution was placed on ice, and DIPEA (0.4 mL, 2.25 mmol) and PyBOP (0.859 g, 1.65 mmol) were added. The reaction was allowed to stir for 2 min before a solution of compound 1b (0.446 g, 1.65 mmol) dissolved in DCM (4.8 mL) was added dropwise. After 30 min the reaction was warmed to room temperature and allowed to stir for 1.5 h longer. The reaction was dried under reduced pressure and then purified by normal phase flash chromatography using Hexane/EA solvent system to afford compound 2a (0.546 g, 59.8% yield) as a yellow oil. 1H NMR (400 MHz, Chloroform-d) δ 8.06 – 7.97 (m, 2H), 7.63 – 7.53 (m, 3H), 7.51 – 7.40 (m, 3H), 7.16 (d, J = 7.9 Hz, 1H), 5.41 (s, 1H), 4.77 (dt, J = 7.7, 5.3 Hz, 1H), 4.34 (s, 1H), 4.22 (d, J = 13.6 Hz, 1H), 4.19 – 4.01 (m, 3H), 3.77 (d, J = 5.6 Hz, 3H), 3.04 – 2.86 (m, 3H), 2.78 (dd, J = 14.2, 6.7 Hz, 1H).



2a 2b

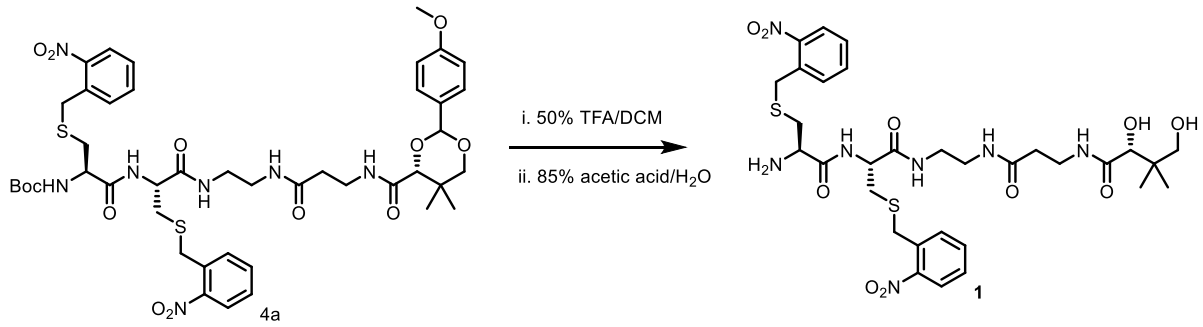
***N*-(*N*-(tert-butoxycarbonyl)-*S*-(2-nitrobenzyl)-*L*-cysteinyl)-*S*-(2-nitrobenzyl)-*L*-cysteine.** Compound 2a (500 mg, 0.821 mmol) was dissolved in THF (7.4 mL) and methanol (6.57 mL). The reaction was cooled on ice, and then 0.4 M NaOH in H_2O (6.43 mL) was added dropwise. The reaction was taken off ice and allowed to stir for 1 hour. The reaction was then acidified using HCl to pH 3.0, and product was extracted using EA (3 x 10 mL). The organic layer was washed with H2O (15 mL) and then brine (15 mL). The organic layer was then dried over sodium sulfate and dried under reduced pressure to afford compound 2b (0.438 g, 90% yield) as a yellow solid.

1H NMR (400 MHz, Chloroform-d) δ 8.04 – 7.96 (m, 3H), 7.62 – 7.53 (m, 5H), 7.50 – 7.38 (m, 6H), 5.47 (s, 1H), 4.82 – 4.74 (m, 1H), 4.47 (s, 1H), 4.22 – 4.04 (m, 6H), 3.06 (dd, J = 14.0, 5.0 Hz, 1H), 2.96 (d, J = 11.4 Hz, 1H), 2.84 (s, 4H), 2.07 (s, 1H), 1.46 (s, 9H), 1.41 (s, 1H), 1.28 (t, J = 7.2 Hz, 1H), 1.27 (s, 1H).



2b 3a 4a

Tert-butyl ((11*R*,14*R*)-1-((4*R*)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxan-4-yl)-11-(((2-nitrobenzyl)thio)methyl)-17-(2-nitrophenyl)-1,5,10,13-tetraoxo-16-thia-2,6,9,12-tetraazaheptadecan-14-yl)carbamate. Compound 2b (154 mg, 0.26 mmol) was dissolved in DCM (8 mL). On ice, HATU (108.6 mg, 0.29 mmol) and DIPEA (68 µL, 0.39 mmol) were added to the reaction. The reaction was allowed to stir for 10 min before the dropwise addition of compound 3a (108.6 mg, 0.29 mmol). After 15 min the reaction was warmed to room temperature and stirred overnight. The reaction was dried under reduced pressure and purified by prep HPLC using a Phenomenex Luna C18 column using H2O with 0.1% TFA as solvent A and ACN with 0.1% TFA as solvent B to afford compound 4a (167.4mg 67.3% yield). 1H NMR (400 MHz, Chloroform-d) δ 8.02 (td, J = 8.2, 7.8, 4.4 Hz, 2H), 7.91 – 7.82 (m, 1H), 7.71 (s, 0H), 7.65 – 7.40 (m, 8H), 7.23 (s, 1H), 7.19 – 7.06 (m, 1H), 7.06 – 7.00 (m, 1H), 6.92 (dd, J = 8.7, 1.9 Hz, 2H), 6.68 (s, 1H), 5.48 (s, 1H), 4.54 – 4.42 (m, 1H), 4.25 – 3.96 (m, 8H), 3.92 (s, 1H), 3.82 (s, 2H), 3.79 – 3.61 (m, 2H), 3.59 – 3.47 (m, 2H), 3.38 (s, 4H), 3.22 (pd, J = 7.7, 4.7 Hz, 4H), 2.87 (dddd, J = 29.5, 22.8, 13.8, 6.8 Hz, 2H), 2.47 (dt, J = 12.7, 5.9 Hz, 2H), 2.06 (s, 3H), 1.55 – 1.37 (m, 20H), 1.28 (t, J = 7.2 Hz, 4H), 1.15 – 0.99 (m, 6H), 0.94 (s, 1H).



(*R*)-*N*-((4*R*,7*R*)-4-amino-7-(((2-nitrobenzyl)thio)methyl)-1-(2-nitrophenyl)-5,8,13-trioxo-2-thia-6,9,12-triazapentadecan-15-yl)-2,4-dihydroxy-3,3-dimethylbutanamide. On ice, compound 4a (48 mg, 0.050 mmol) was dissolved in 4 mL of 1:1 DCM:TFA and 300 µL of water. After 1 h the reaction was dried under reduced pressure, and the resulting oil was dissolved in 85% acetic acid:water and stirred overnight. The reaction was dried under reduced pressure and purified by prep HPLC using a Phenomenex Luna C18 column using H2O with 0.1% TFA as solvent A and ACN with 0.1% TFA as solvent B to afford compound **1** (1.8 mg, 4.9% yield). Deprotection was confirmed by LC-HRMS.

Photodeprotection of cysteine intermediate

A solution of 250 µM compound **1**, 1 mM Ascorbic Acid, 2.5 mM semicarbazide in potassium phosphate buffer pH 6.0 in borosilicate glass was illuminated by UV A. The extent of deprotection was measured by LC/MS, and >90% deprotection was observed in 3 hours.

In vitro HlmE T domain loading of CoA analogue

The photodeprotected product in the photodeprotection solution was adjusted to pH 7.5 using 0.1 M NaOH. To 400 µL of photodeprotection solution 9 mM ATP, 10 mM $MgCl_2$, 0.5 µM CoAA, 0.7 µM CoAD, and 1.5 µM CoaE were added to a final volume of 500 µL. The assay was then incubated at 37 ºC overnight to form the CoA intermediate. To the complete CoA-analogue-forming assay, 100 µM HlmE T domain and 1 µM of Sfp were

added and allowed to mix at 37 C for 1 h. LC-MS was used to determine the extent of substrate loading onto the T domain.

HlmB and HlmD in vitro activity assays

To 50 µL of T domain loaded substrate, 1 mM of Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and 1 mM of ascorbic acid were added to the reaction and allowed to mix for 10 min at r.t. Then 20 µM of HlmB or HlmD or both were added and mixed for 1 h at r.t. Finally, 12.5 mM of iodoacetamide was added and allowed to mix at r.t. for 10 min before T domain analysis by LC/MS. Additional cofactors such as NADH, NAD+ were added at 1 mM concentration with TCEP when included.

Protein Expression and Purification

For protein purification, *E. coli* BAP-1 were transformed with pLIC-His-ambB. *E. coli* BL-21(DE3) were transformed with pLIC-His-ambC and pLIC-His-ambE individually. *E. coli* BL21-CodonPlus(DE3) RIPL cells were transformed with pLIC-His-ambD. All transformations were performed by electroporation. *E. coli* were routinely cultured at 37 ºC.

AmbB was expressed in 1L cultures of LB Miller media containing 100 µg/mL ampicillin. The cultures were grown at 37 °C with shaking to an OD_{600} of 0.6, which were then induced by the addition of 0.5 mM IPTG, and grown for another 18 hours at 16 °C. The cells were then pelleted by centrifugation at 3000 x g for 30 min. The cells were resuspended for lysis in Buffer A (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol (BME), and 30 mM imidazole) and sonicated using a Fisher Scientific Sonic dismembrator model 500 (1.5 min total on time of cycles of 30% amplitude with 0.5 sec pulse on, and 1.5 sec pulse off). The lysed cells were centrifuged at 34,500 x g for 30 min at 4 °C. The supernatant was filtered through a 0.45 µm filter before purification by FP-LC. The filtered

supernatant was purified by Ni affinity chromatography using a GE HisTrap HP Ni column. Buffer A was used as wash buffer, and Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM BME, and 300 mM imidazole) was used as elution buffer. Fractions containing AmbB were combined and concentrated to a final volume of 3.5 mL using a Millipore centrifugal filter with a 30 kDa molecular weight cut off. The combined fractions were then applied to a GE HiLoad 16/600 Superdex 200 pg column that had been equilibrated with Buffer C (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM BME, and 10% glycerol). Fractions containing pure AmbB were pooled and again concentrated using a Millipore centrifugal filter with a 30 kDa molecular weight cut off to a final concentration of 50–200 µM. AmbB was then flash frozen in liquid nitrogen and stored at −80 °C until needed.

AmbE, AmbE S1958A (TE mutant), AmbE S1819A (T2 mutant), and AmbE T domains were expressed and purified following the method above. AmbC and AmbD were expressed and purified by Ni affinity chromatography as described above, and then AmbC and AmbD were desalted using a GE PD-10 column equilibrated in Buffer C. AmbC and AmbD were then concentrated using a Millipore centrifugal filter with a 10 kDa molecular weight cut off, and frozen and stored similar to AmbB. AmbE T domains, AmbES1958A (TE mutant), and AmbE S1819A (T2 mutant) plasmids were expressed in BL21.

Reconstitution of AMB enzymes in vitro

Enzyme activity was reconstituted *in vitro* by incubating 1.25 µM AmbB and 2.5 µM AmbE in 25 mM potassium phosphate buffer pH 7.5 with 0.5 µM Sfp, 1 mM CoASH, and 4 mM $MgCl_2$ at 28 °C for one hour. To this solution 1 mM alanine, 1 mM glutamic acid, 1 mM α-KG, 1 mM sodium ascorbate, 320 µM *S*-adenosylmethionine, 2 mM ATP, and 200 µM ammonium iron(II) sulfate were added. The reaction was initiated by the addition of 7.2 µM

AmbD and 13.7 µM AmbC, which brought the final volume to 100 µL. The assay was incubated at 28 °C for 2 hours. To quench the reaction, 100 µL of acetonitrile was added and the assay was cooled to −20 °C for 20 min. The assays were then centrifuged at 14,000 x g for 5 min, and then the supernatant was taken for analysis by LC/MS.

AmbE T domain 1 glutamate-loading assays and AmbC and AmbD activity

To load the phosphopantetheine (Ppant) arm onto T domains in *trans*, 2 µM Sfp was added to a solution containing 25 mM potassium phosphate buffer pH 7.5, 2 mM $MgCl_2$, 1 mM CoA, and 150 µM AmbE T domain 1 (T1). The assay was shaken at 28 °C for 1 hour. The greatest amount of amino acid loading was observed when 10 µL of Ppant loading assay was added to 25 mM potassium phosphate buffer pH 7.5, 2 mM ATP, 1 mM amino acid, and 2 µM AmbE, or AmbE S1958A to a final volume of 50 µL. The assay was incubated at 28 °C for 2 hours. AmbE was used to load glutamate onto T1.

To test tailoring enzyme activity, 50 µM AmbC and AmbD were separately incubated with 1 mM ammonium iron(II) sulfate and 1 mM ascorbic acid in 25 mM potassium phosphate buffer pH 7.5 for 10 min prior to activity assays. For each assay, 5 µL of iron incubated AmbC or AmbD and 1 mM α-KG were added to 50 µL of T1 loading assay described previously. The assay was allowed to incubate at 28 °C for 2 hours. Samples were spun down at 20,000 x g and the supernatant was taken for analysis by LC/MS.

Chemical capture of NRPS-bound intermediates using cysteamine

The one-pot in vitro AMB reconstitution reaction was conducted using either AmbES1819A or AmbE-S1958A in place of wild-type AmbE to accumulate biosynthetic intermediates bound to AmbE. A sample of freshly prepared 50 mM cysteamine hydrochloride was added to the one-pot assay as the final component immediately after

adding all other substrates and cofactors. Individual enzymes (AmbB, AmbC, or AmbD) or substrates (Ala or SAM) were excluded from the reaction mixture in separate reactions. Isotopically labeled glutamic acids were used in place of L-glutamic acid for corresponding reactions. The modified assays containing cysteamine were incubated at room temperature for 2 hours and quenched and spun down as described previously. Supernatants were Fmoc derivatized by adding 25 µL of samples to 90 µL of ACN and 50 µL of 200 mM sodium borate buffer pH 9.0. To this solution, 90 µL of 18 mM Fmoc-Cl dissolved in acetonitrile was added and samples were incubated at room temperature for 10 min before analysis by LC-MS.

Analysis by LC/MS

Supernatants from small molecule *in vitro* samples were analyzed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 300 °C, fragmentor 175 V, and skimmer 65 V. Underivatized samples were separated using a 150 mm Phenomenex Kinetex C18 column with 2.6 µm particle size and 100 Å pore size. Solvent A consisted of 0.1% formic acid in water (Fisher), and solvent B consisted of 0.1% formic acid in acetonitrile (Fisher). Fmoc derivatized samples were separated using a 50 mm Phenomenex Gemini C18 column with 5 µm particle size and 110 Å pore size. Solvents A and B were same as above.

Supernatants from T domain studies were analyzed using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS/MS. Data were collected using positive ion mode ESI mass spectrometry with the following parameters: gas temperature 350 °C, fragmentor 250 V, skimmer 65 V. Masses chosen for MS^2 were selected automatically based on abundance

and charge state over 3, and masses were fragmented with a collision energy of 35 V. The protein samples were separated using a 150 mm Restek Viva C4 column with 5 µm particle size. Solvents used for separation are the same as above. The MS data were used for protein deconvolution and whole protein analysis. The MS^2 data were used for Ppant ejection analysis.[82]

Results

*HlmE T domain was loaded with the synthetic compound **1***

Compound 1 was synthesized and photodeprotected using UV light in the presence of ascorbic acid and semicarbazide and then converted to the respective CoA analogue using CoaA, CoaD, CoaE and Sfp, (Figure 3.4). However, the photodeprotection required acidic conditions, but the enzymes of the CoA forming assay required pH 7.5. Thus, to complete the entire deprotection and loading of substrate onto the HlmE T domain 1, we first photodeprotected at pH 6.0, then without purification, the pH was adjusted to 7.5 for the CoA forming and T domain loading assays.

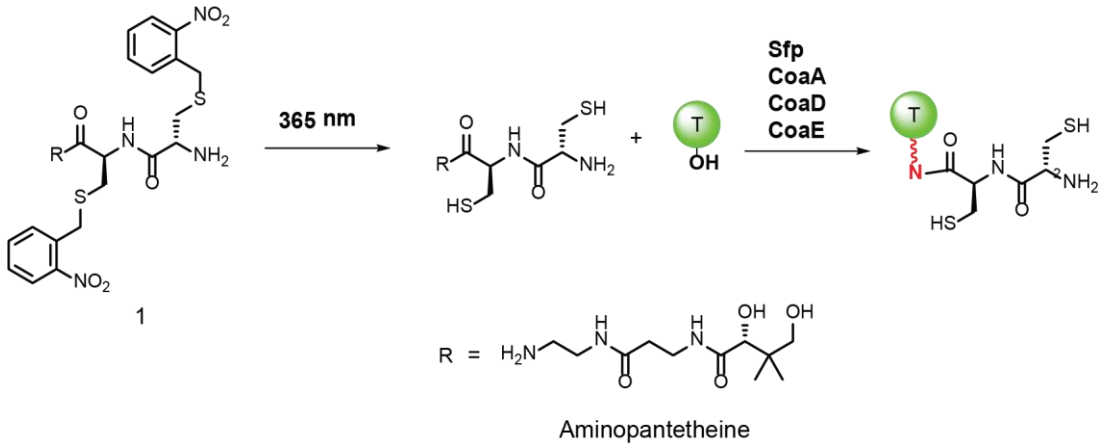


Figure 3.4 – Scheme for in vitro loading of HlmE T domain with synthetic cysteine dipeptide intermediate. Compound **1** was photodeprotected and then loaded onto HlmE T domain.

HlmB and HlmD were not active using T domain loaded substrate

Figure 3.5 shows that compound **1** was successfully loaded onto HlmE T domain after photodeprotection, but neither HlmB nor HlmD showed any activity with this substrate. A variety of cofactors known to be associated with flavoproteins were added to the assays, such as NADH, NAD^+, and ascorbic acid, but no modification of substrate was detected by LC/MS. Additionally, to determine whether the lack of detection was due to insufficient changes in mass, iodoacetamide (IAA) was used to derivatize free thiols in order to increase the mass shift that would occur if the thiols were modified. Again, no difference was observed in whole protein MS or MS^2 Ppant ejection between the samples with or without tailoring enzymes (Figure 3.5). Attempts to use cysteamine cleavage to identify intermediates were also unsuccessful, and activated cysteine was never observed by LC-MS. It is likely that the thiols of cysteine prevent the expected dimerization of cysteamine and complicated the analysis by LC-MS.

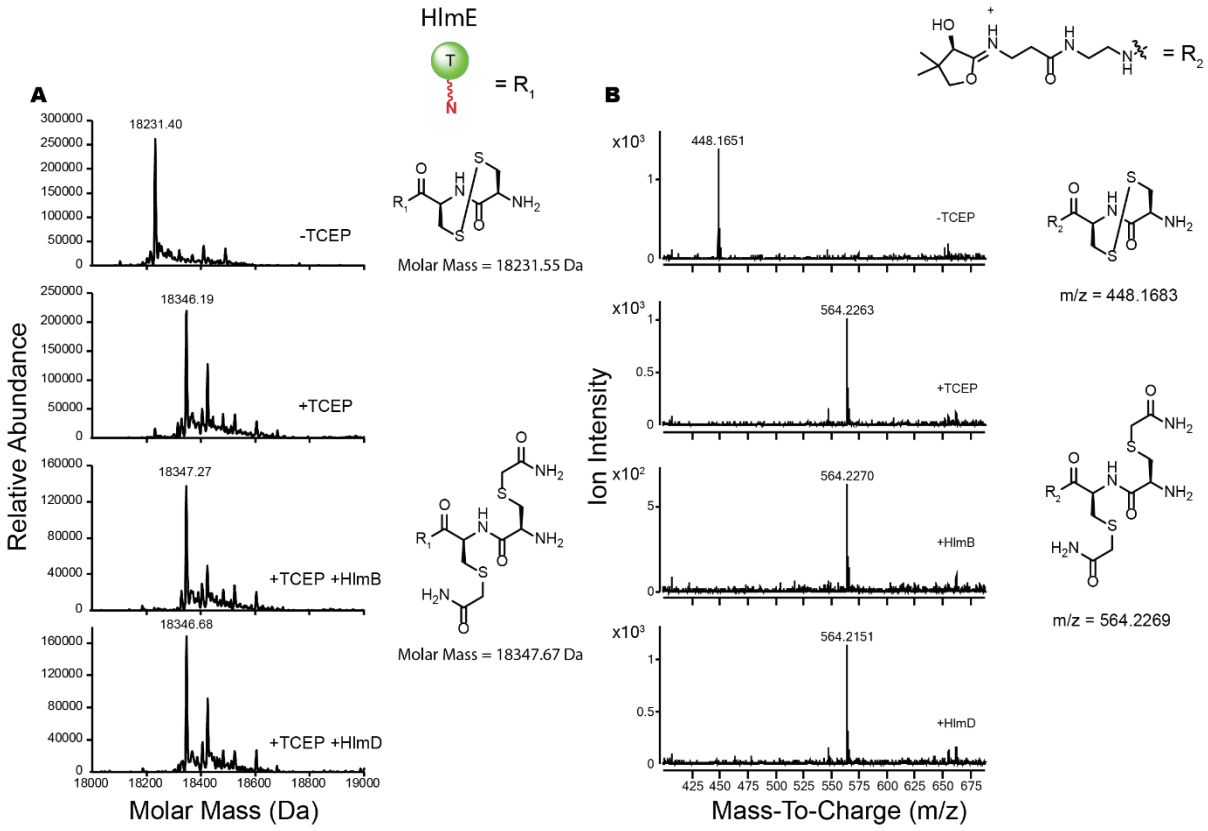


Figure 3.5 – HlmB and HlmD do not show activity when incubated with the synthetic cysteine dipeptide substrate linked to the T domain of HlmE. TCEP reduced disulfide bond, and iodoacetamide was used to derivatize free thiols for greater mass shifts. A) Analysis of *in vitro* reactions by whole protein mass spec. B) Analysis of *in vitro* reactions by MS^2 Ppant ejection.

Reconstitution of AMB biosynthetic enzymes produced Ala-AMB in vitro

We cloned and purified all four biosynthetic enzymes from the *amb* gene cluster and reconstituted activity in a one-pot assay by incubating these enzymes with their predicted substrates and cofactors in a single reaction. Surprisingly, LC-MS analysis of the assay detected a product with a mass that corresponds to an ala-AMB dipeptide (**2**) rather than an Ala-AMB-Ala tripeptide observed previously (Figure 3.6).[72] Ala-AMB was not produced when enzymes were denatured before the assay, or when the cofactor α-ketoglutarate (α-KG)

was absent from the reaction. To confirm its structure, the Ala-AMB product was Fmoc derivatized and then purified for analysis by NMR.

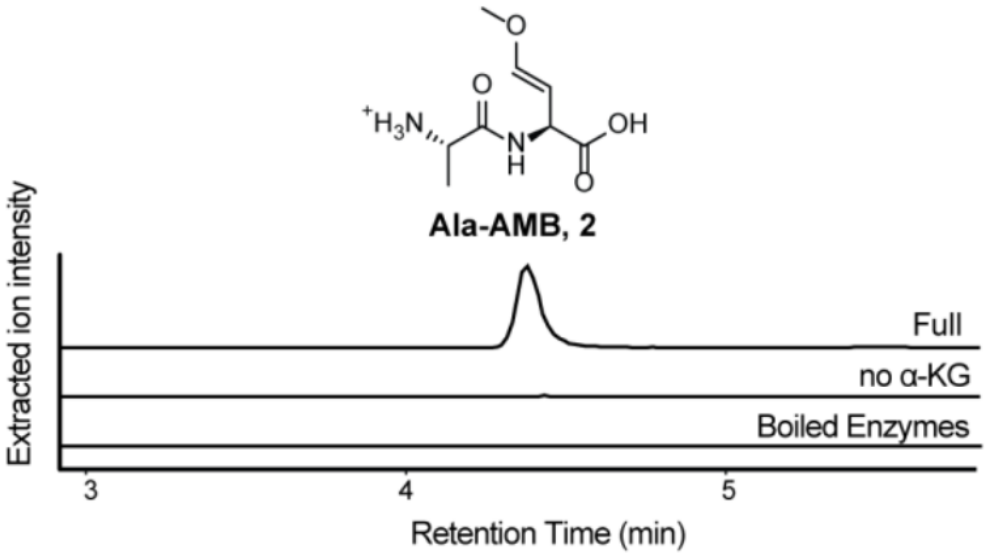


Figure 3.6 – Ala-AMB was the product of in vitro reactions using the four biosynthetic Amb enzymes. Extracted ion chromatograms from the one-pot assay and negative controls. α-ketoglutarate (α-KG) is a necessary cofactor of AmbC and AmbD.

Truncated AmbE T1 domain assays elucidate AmbC activity

Next, we characterized the chemical reactions that convert glutamate to AMB. In order to elucidate the modifications that occur while glutamate is tethered to AmbE, we loaded glutamate onto the first T domain of AmbE (T1) to act as a substrate for the tailoring enzymes. We performed glutamate loading onto the T1 by using Sfp to add the Ppant arm to T1, and then full length AmbE to activate glutamate and load it onto the Ppant arm of T1 (Figure 3.7A). Incubating T1 with AmbC led to a 16 Da increase in mass, suggesting AmbC hydroxylates glutamate and that AmbC is the first enzyme to modify glutamate in the pathway. However, we did not observe AmbD activity on glutamate or the hydroxylated glutamate substrate bound to T1 during sequential assays (Figure 3.7B). Next, we used the

tandem MS technique known as Ppant ejection,[82] to confirm that the hydroxylation had occurred on glutamate, and not another residue of T1. In addition, we used differentially deuterated glutamates as substrates to demonstrate that AmbC selectively hydroxylated the glutamate at C_3 (Figure 3.7C).

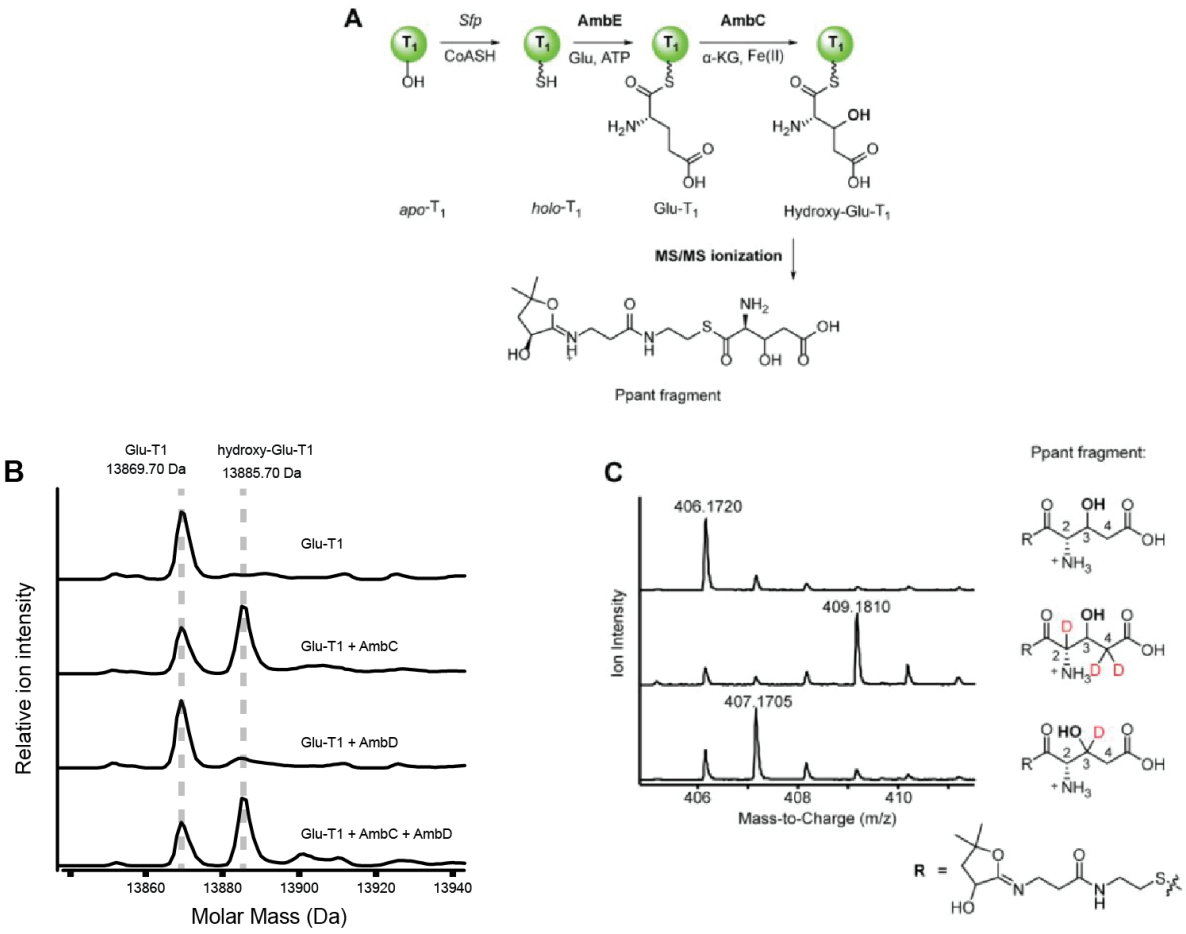


Figure 3.7 – AmbC hydroxylated glutamate loaded on AmbE T1, but AmbD showed no activity. A) The scheme of loading glutamate onto AmbE T1 for whole protein analysis by LC-MS, and the structure of Ppant fragment detected in MS^2. B) Deconvoluted protein mass spectra of glutamate modification on AmbE T1. AmbC hydroxylates glutamate, but AmbD did not show any activity. C) Mass spectra of Ppant ejection assays using deuterated glutamate substrates to determine AmbC regioselectivity. Substrates were: L-Glu (top), 2,4,4-D_3-L-Glu (middle), and 3,3-D_2-L-Glu (bottom).

Cysteamine cleavage identifies multiple intermediates of AMB biosynthesis

Next, we used cysteamine as a chemical probe to identify additional intermediates of the pathway. Cysteamine assays were performed by adding cysteamine to the one-pot assays

that included the TE mutant of AmbE. Since the TE domain cleaves products and intermediates from AmbE, we used the TE mutant to buildup intermediates on AmbE that could then only be cleaved by cysteamine. Additionally, the assays were Fmoc derivatized before analysis by LC-MS to increase the retention time of intermediates on reverse-phase columns and to achieve better signal intensity by MS (Figure 3.8A). Figure 3.8B shows 5 different intermediates identified by cysteamine capture across 5 assay conditions, a significant improvement over the T domain approach, which only proceeded through the first hydroxylation of AmbC. We were able to detect the final product, Ala-AMB, in the full assay, which confirms that the enzymes are still active in the presence of cysteamine. When AmbC was excluded, unmodified glutamate (**4**) was detected on the assembly line, suggesting that AmbC is responsible for the first modification, which confirmed the T domain result. Exclusion of AmbD led to the detection of monohydroxy-glutamate (**5**), suggesting that AmbD catalyzes the second modification after AmbC hydroxylation. Omission of *S*-adenosyl methionine (SAM), which is required for the methyltransferase domain, led to detection of dihydroxy-glutamate (**6**), suggesting that *O*-methylation occurs after both hydroxylations. Lastly, omission of AmbB resulted in accumulation of a dihydroxylated and methylated glutamate (**7**), but not AMB, suggesting that condensation with alanine occurs after the hydroxylations and O-methylation, but prior to the dehydration and decarboxylation necessary to form AMB.

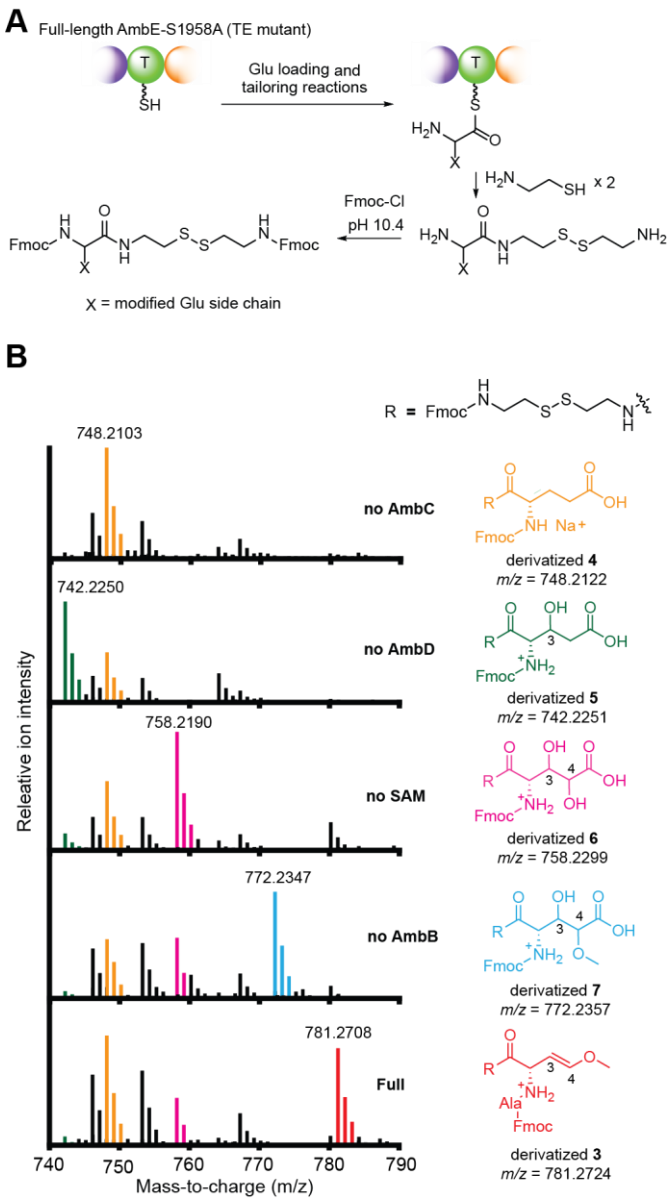


Figure 3.8 – Capture and detection of AMB biosynthetic intermediates using cysteamine cleavage, which revealed the order of biosynthetic steps. A) Scheme of cysteamine cleavage of covalently linked intermediates in NRPS pathways, followed by Fmoc derivatization. B) Mass spectra captured intermediates during reconstitution of AMB biosynthesis using the TE mutant AmbE S1958A. While the proton adduct of derivatized **4** is detected, it has been excluded for clarity.

To determine the regioselectivity of each modification, we used 2,4,4-D_3-glutamate as a substrate for the cysteamine cleavage assay (Figure 3.9). We observed that AmbC hydroxylation did not remove a deuterium, which confirmed the T domain assay result (Figure 3.7) demonstrating that AmbC hydroxylates glutamate at the C_3 position. AmbD did remove a deuterium, and it is assumed from the final structure of AMB that this must have occurred on C_4 of hydroxy-glutamate. Interestingly, one of the deuteriums of **7** was lost during the conversion to ala-AMB, and initially it was unclear whether this was the C_2 or the remaining C_4 deuterium (Figure 3.9).

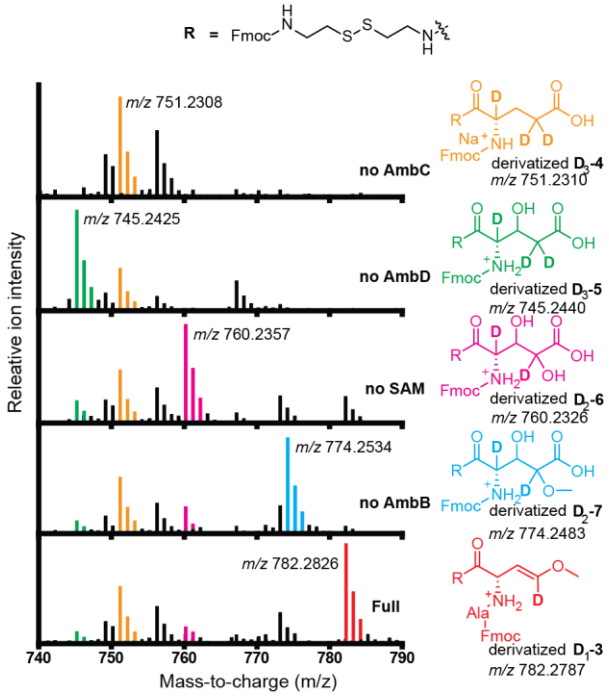


Figure 3.9 – The cysteamine cleavage assay was repeated using 2,4,4-D3-L-Glu as a substrate to confirm regioselectivity of modifications. Assay conditions and analysis were the same as Figure 3.8.

We were able to determine which deuterium remained by feeding differentially deuterated glutamates as substrates for the one-pot reaction. When 3,3-D_2-glutamate, 4,4-D_2-glutamate and 2,4,4-D_3-glutamate were used as substrates, the ala-AMB product contained

only one deuterium (Figure 3.10). Since one of the deuteriums of 4,4-D2-glutamate remained in the substrate, we can infer that the deuterium remaining from the 2,4,4-D3-glutamate substrate was at the C_4 position, which meant that during the conversion of compound **7** to **3**, the C_2 deuterium was lost. We hypothesized that this was due to a 2,3-dehydro intermediate resulting from the dehydration of the hydroxyl at C3.

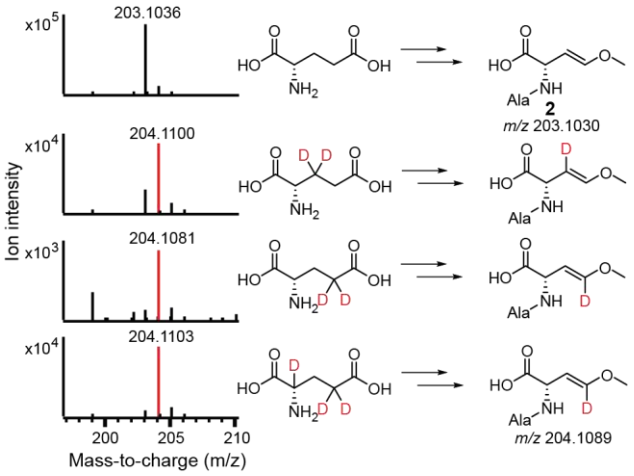


Figure 3.10 – AMB products of differentially deuterated glutamate substrates show loss of C_2 deuterium.

After elucidating multiple steps of the biosynthesis of AMB, we sought to identify where on AmbE these modifications were taking place. For this experiment we used the mutant AmbE S1819A, which lacks the serine necessary for Ppant modification at T2, leaving T1 as the only domain where the substrate can be tethered (Figure 3.11A). The full one-pot cysteamine cleavage assays using AmbE S1819A did not produce ala-AMB, and instead a small amount of alanyl-3-hydroxy-4-methoxy-Glu (**8**) was produced. This result suggests that both hydroxylations, the methylation, and the alanine condensation all occurred on T1, but the remaining dehydration and decarboxylation necessary to form AMB required T2.

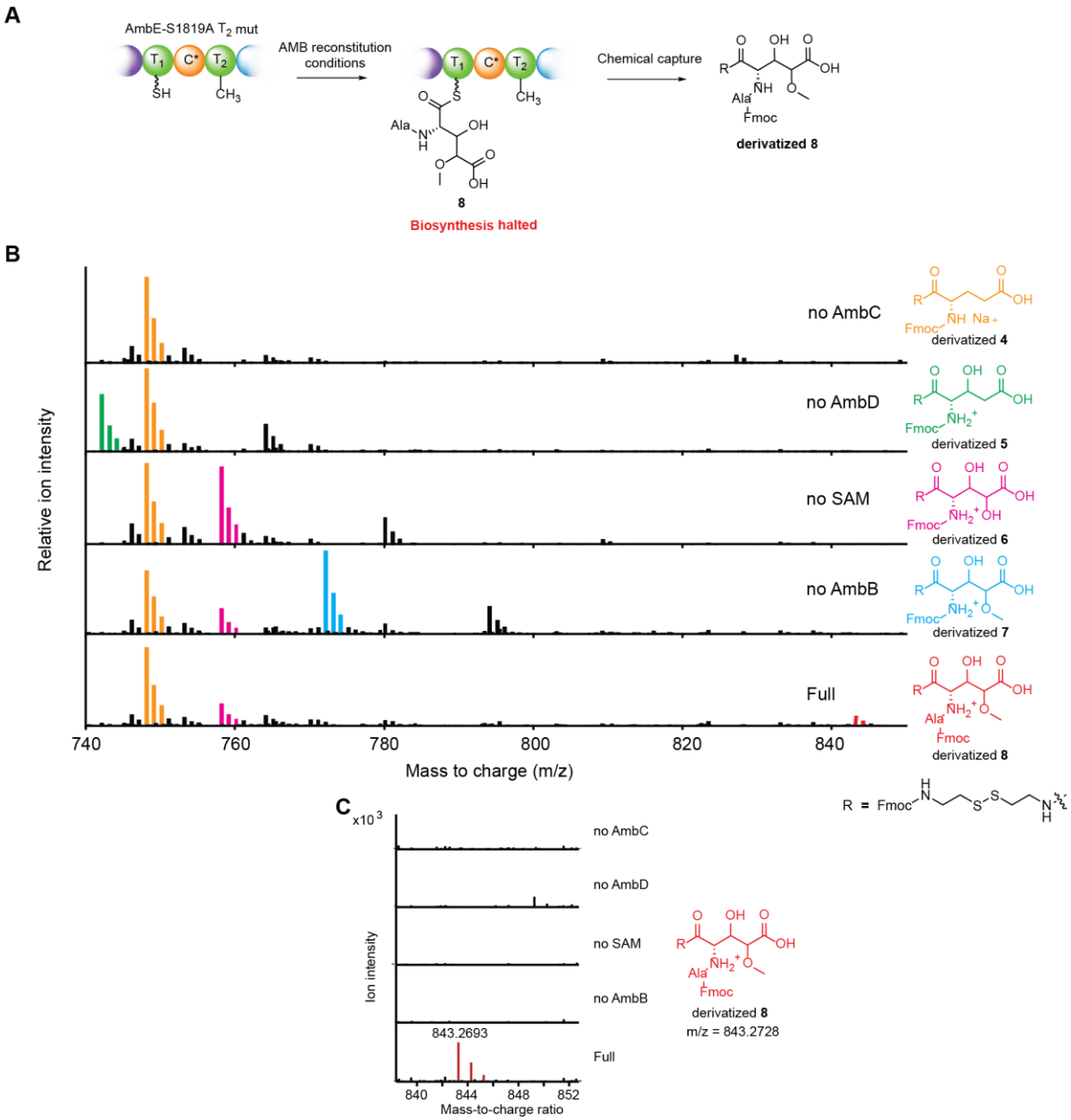


Figure 3.11 – The cysteamine cleavage assays were repeated using AmbE S1819A (T2 mutant) in place of AmbE S1958A (TE mutant). A new intermediate was detected, derivatized **8**, which was only present in the full assays. A) Scheme showing halting of biosynthesis due to mutation in T2 domain. B) Mass spectra of captured intermediates using AmbE S1819A in cysteamine cleavage assays. C) Expansion of derivatized **8**.

Discussion

The fact that intermediates of NRPS pathways are covalently bound to assembly-line-like enzymes has made elucidating the roles of tailoring enzymes and noncanonical NRPS domains inherently difficult. Here three different methods of studying NRPS pathways were tested and each had varying degrees of success. To study holomycin biosynthesis, we utilized a chemoenzymatic approach that has previously been used to load fluorescent tags and markers to T domains.[78,79] Using this chemoenzymatic approach, a cysteine dipeptide intermediate was successfully loaded onto the HlmE T domain (Figure 3.4), but neither HlmB nor HlmD showed any activity with this substrate *in vitro* (Figure 3.5). This result highlighted one of the key drawbacks of this chemoenzymatic approach: the intensive effort necessary to test a single substrate compared to typical *in vitro* reaction assays. Each substrate must be individually synthesized, making it difficult to rapidly test many possible substrates. Also, without a positive control, it is unclear which part of the assay was responsible for the lack of activity. It may be possible that the enzymes were inactive, necessary cofactors were missing, or the substrate was incorrect. Additionally, the native substrate of tailoring enzymes are linked to the full NRPS, not just the T domain, and it is has previously been shown that protein-protein interactions with additional domains of an NRPS are necessary for some tailoring enzymes.[74,75] While HlmB and HlmD are interesting targets to study since they may catalyze novel sulfur chemistry in biological systems, the chemoenzymatic approach was not suitable for identifying possible substrates without further information about the pathway. One way to ensure that the substrates of tailoring enzymes are correct would be to use the NRPS from the pathway to load an amino acid onto the T domain. Unfortunately, it has been shown that the thiols of cysteine will cleave the thioester bond of the Ppant arm releasing the cysteine dipeptide, making the holomycin pathway

difficult to study using this technique as well.[68] Additionally, attempts to use cysteamine to cleave intermediates were also unsuccessful (data not shown). It is possible that the thiols of cysteine may form mixed disulfides with cysteamine forming multiple different compounds for any intermediate cleaved from the NPRS, which would complicate the analysis by LC-MS. However, we were able to demonstrate that the AMB pathway is amenable to both of these methods.

The activity of AmbC was identified using truncated AmbE T1 loaded with glutamate as a substrate. To our knowledge, this was the first example of using the full length NRPS to load the native substrate onto a truncated T domain to be used as a substrate for tailoring enzymes (Figure 3.7). While this method was used to successfully identify the activity and regioselectivity of AmbC, no activity was observed for AmbD. From additional cysteamine cleavage assays, we have shown that the substrate of AmbD was in fact hydroxylated glutamate, which indicates that other interactions between AmbD and AmbE might be necessary for hydroxylation by AmbD. Similar to the chemoenzymatic approach, a significant limitation of using these truncated T domain studies is that the T domain alone is not the native substrate of the tailoring enzymes, and it appears as though the entire NRPS may often be necessary. One method that uses complete NPRS enzymes and does not suffer from this problem is the *in vitro* cysteamine cleavage assay.

The only example of cysteamine being used as a thioester probe in the literature is with a PKS from the calicheamicin pathway.[48] Belecki and Townsend used cysteamine to cleave intermediates that covalently bound to the PKS CalE8 during purification, which lead to extremely low concentrations of intermediates. We improved this technique by including cysteamine during AMB production in *in vitro* assays; thus, every time cysteamine cleaved

off an intermediate, the pathway continued to activate and modify additional substrates resulting in greater concentrations of cleaved intermediates. Additionally, the free amine present in cysteamine after cleavage and dimerization allowed for Fmoc derivatization for better sensitivity regardless of the intermediate being cleaved (Figure 3.8A). Using cysteamine as a chemical probe, multiple intermediates were cleaved from AmbE and detected by LC-MS. By altering the enzymes, cofactors, and substrates present in the assays, various intermediates were detected, and each chemical modification could be assigned to a specific enzyme (Figure 3.8B). AmbC and AmbD hydroxylated glutamate at carbons C_3 and C_4, respectively. The methyltransferase domain then methylated the hydroxyl on C_4, followed by the condensation of alanine onto the modified glutamate (**7**) by the AmbB (Figure 3.12). Additionally, using the AmbE T2 mutant in the cysteamine cleavage assays demonstrated that all of these reactions occurred on the T1 domain of AmbE (Figure 3.11). The regioselectivity of AmbC and AmbD were also determined by using the 2,4,4-D_3-glutamate substrates in the cysteamine cleavage assay (Figure 3.9). The only reaction that remains uncharacterized is the dehydration and decarboxylation of compound **8** that ultimately produces the final ala-AMB product (Figure 3.12). Phylogenetic analysis of the AmbE C* domain identifies homology to other C domains that are predicted to catalyze dehydrations of serine and threonine to produce dehydroalanine or dehydrobutyrine.[83,84] Thus it is possible that the C* domain may catalyze the dehydration of C3 to make the 2,3-dehydro intermediate **9** before decarboxylation. However, additional activity studies are necessary to characterize the final two transformations to form AMB.

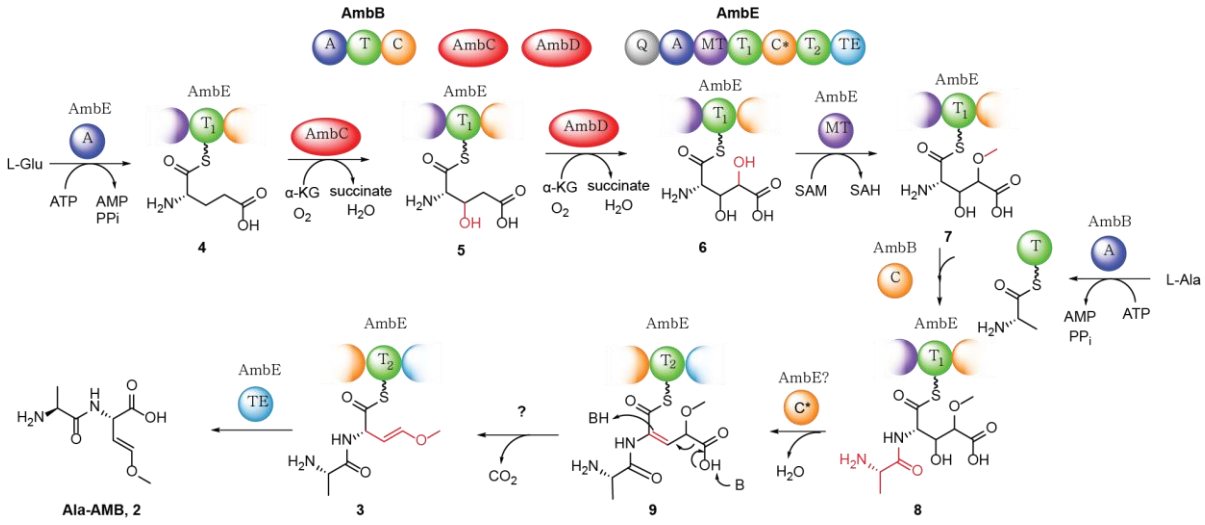


Figure 3.12 – Proposed enzymatic biosynthesis of Ala-AMB

Conclusion and Future Directions

Comparing the three methods described in this study, cysteamine cleavage was shown to be the most effective at characterizing an NRPS biosynthetic pathway, including the identification of intermediates and their step-by-step modifications. Using cysteamine cleavage *in vitro* followed by Fmoc derivatization led to increased intermediate concentrations; thereby allowing the detection of multiple intermediates of the AMB pathway and the determination of the order and timing of the biosynthetic steps necessary to produce AMB. Using a combination of deuterium labeled substrates, we discovered multiple unexpected chemical transformations, including the hydroxylation of C_3 by AmbC before subsequent dehydration to the 2,3-dehydro intermediate **9**. While loading glutamate onto AmbE T1 was a useful technique for determining the activity of AmbC, the fact that AmbD was not able to hydroxylate its substrate again showed the limitation of this method; in order to study tailoring enzymes, the full NRPS is often necessary for activity, not just the T domain. While cysteamine cleavage has only been used to study the AMB and calicheamicin pathways thus far, it has the potential to serve as a useful chemical probe in both NRPS and

PKS pathways. In pathways where tailoring enzymes have previously been shown to be active, cysteamine cleavage could be used to detect intermediates and definitively characterize the reactions catalyzed by these tailoring enzymes. However, one disadvantage of cysteamine cleavage assays is the fact that free thiols from substrates or cofactors may inhibit the expected dimerization of cysteamine and thus complicate the LC-MS analysis, potentially limiting the types of pathways that cysteamine may be used to study. Overall, cysteamine cleavage *in vitro* has been shown to be a promising technique to study NRPS and PKS systems and for characterizing tailoring enzymes with novel activities, which in turn can be used as biocatalysts in synthetic schemes or be used to identify additional enzymes that perform similar chemistry in other natural product pathways.

CHAPTER 4: METABOLOMICS AND BIOCHEMICAL STUDIES OF 4-FORMYLAMINOOXYVINYLGLYCINE BIOSYNTHESIS

Introduction

Pseudomonas fluorescens WH6 is a bacterial strain that was identified as a promising biocontrol agent to prevent the growth of grassy weeds. Grassy weeds are a problem for both recreational and professional lawns, especially weed strains that show resistance to chemical herbicides such as annual bluegrass weeds.[85] Additional methods to control grassy weed growth are needed to combat the growing prevalence of resistance. Through phenotypic screens the culture extracts of WH6 were found to inhibit the germination of bluegrass weeds.[85] The active metabolite was isolated and characterized as 4-formylaminooxyvinylglycine (FVG), a member of the oxyvinylglycine class of secondary metabolites.[36,86,87] The gene cluster responsible for the biosynthesis of FVG in WH6 was identified using transposon mutagenesis.[88] The cluster was named the growth arrest factor vinylglycine (*gvg*) cluster, and consists of 6 putative enzyme encoding genes, 2 small open reading frames (ORFs), 3 putative LysE type transporters, and one transcription regulator (Figure 4.1). Interestingly, the cluster encodes a protein GvgC, which belongs to a class of proteins with an unknown function. While the *C*-terminus of GvgC shares 20% homology with heme-oxygenases, the larger *N*-terminus portion of the protein does not have homology to any characterized proteins. Also, the putative amidino transferase from the cluster, *gvgD*, is not necessary for FVG biosynthesis, and its purpose in the cluster is unknown.[89] *GvgD*

may instead be involved in the biosynthesis of a different metabolite also produced by the *gvg* cluster.

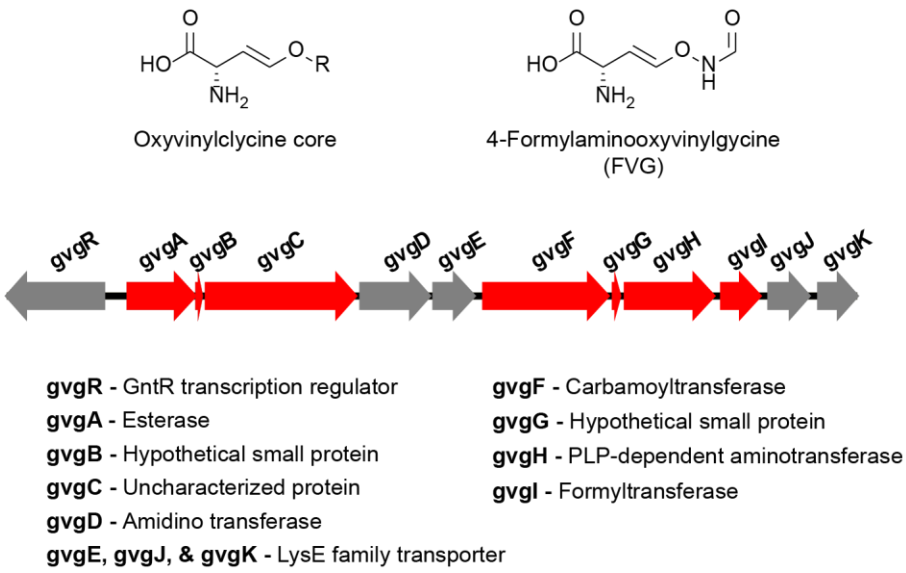


Figure 4.1- *Gvg* gene cluster and structure of FVG produced by Pseudomonas fluorescens WH6. Genes colored red are necessary for FVG production, and the putative function of the protein encode by each gene is listed. The core of all oxyvinylglycines is shown in comparison to FVG.

In addition to FVG, two other metabolites that contain the aminooxyvinylglycine core have been isolated from *Streptomyces*. Aminooxyvinylglycine (AOVG) and guanidinooxyvinylglycine (GOVG) were also identified for their antifungal activity, but no information about their biosynthesis is known.[90] Interestingly, AOVG and GOVG are structurally similar to canaline and canavanine from plant metabolism, which are used for nitrogen storage.[91,92] AOVG and GOVG differ from canaline and canavanine by only a beta desaturation, suggesting that they may act as antimetabolites to modulate plant metabolism (Figure 4.2). Additionally, the biosynthesis of most oxyvinylglycines are not well

understood, and those that have been elucidated proceed through different chemical transformations.[81,93,94] Feeding studies using stable isotope labeled amino acids in growth cultures that produce oxyvinylglycines have suggested homoserine is an important precursor for the aminoethoxyvinylglycine and rhizobitoxine oxyvinylglycine pathways, which may serve as a starting point for studying FVG.[93,95] In primary metabolism, homoserine is an intermediate in the aspartate-threonine pathway. In bacteria, such as *Pseudomonads*, the conversion of aspartic acid to threonine requires 5 enzymatic steps and contains branching points to other pathways at the intermediates homoserine and aspartate semialdehyde. It is possible that a metabolite in the aspartate-threonine pathway is a direct precursor for FVG biosynthesis. In addition to the oxyvinylglycine core, the presence of an N-O bond in FVG can give some insight into the biosynthesis.

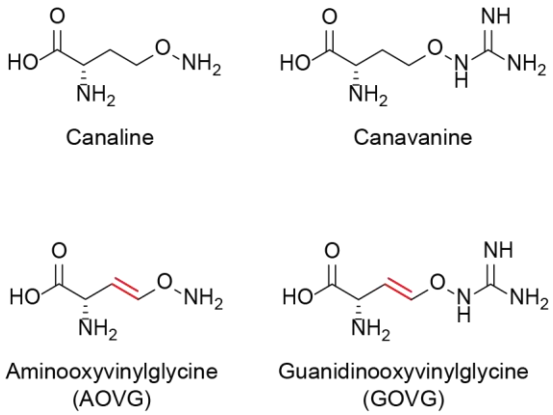


Figure 4.2 – Comparing the structures of canaline and canavanine, two plant metabolites, to AOVG and GOVG. AOVG and GOVG were previously isolated from a *Streptomyces* strain for their antifungal activity. The only difference between the plant metabolites and the oxyvinylglycines is the beta desaturation, which is highlighted in red.

The antibiotic D-cycloserine was isolated from *Streptomyces garyphalus*, and inhibits the D-ala-D-ala ligase necessary for bacterial cell wall biosynthesis.[96,97] While D-cycloserine

is not an oxyvinylglycine, the D-cycloserine cluster has many similarities to the *gvg* cluster and generates an N-O bond. In D-cycloserine biosynthesis, a heme-dependent enzyme (DcsA) hydroxylates the guanidine group of arginine, and then an arginase (DcsB) cleaves the modified guanidine group to produce hydroxyurea. The released hydroxyurea then acts as a substrate for a pyridoxal-5'-phosphate (PLP)-dependent enzyme (DcsD) that converts acetyl-L-serine to O-ureido-L-serine by first eliminating the acetyl group of acetyl-L-serine (Figure 4.3).[98,99] Similarly, the *gvg* cluster has multiple genes that encode proteins with analogous putative functions; GvgC is predicted to have a heme binding domain, GvgH is predicted to be a PLP-dependent enzyme, and GvgA is predicted to be an esterase, which have similar mechanisms to arginases. Due to these similarities, we hypothesized that GvgC may form an N-O bond, which is released from a larger metabolite by the esterase GvgA. Then, GvgH could utilize a modified homoserine substrate and the product of GvgA to produce AOVG. While some of the biosynthesis of FVG remains unknown, the extraction and *in vitro* data presented here provides insights into the biosynthesis of FVG by *P. fluorescens* WH6.

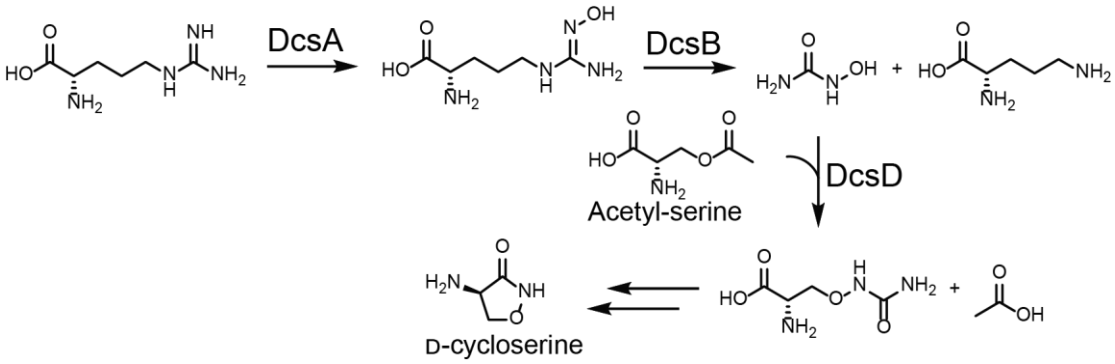


Figure 4.3 – Scheme of D-cycloserine biosynthesis.[98,99] DcsA is a heme dependent enzyme, DcsB is an arginase, and DcsD is a PLP-dependent enzyme.

Materials and Methods

Culture and extraction conditions of P. fluorescens WH6

Wildtype (WT) and mutant strains of WH6 were grown under the same conditions. Bacterial seed cultures of WH6 were started from a single colony and grown in LB medium supplemented with 100 µg/mL ampicillin. 50 µL of seed culture was used to inoculate 50 mL of modified *Pseudomonas* minimal salts media (PMS) as described previously.[85] PMS media consists of 0.2 g potassium chloride and 1.0 g ammonium phosphate monobasic in 50 mM sodium phosphate at pH 7.5 in 1 L. The media was autoclaved before the addition of 20 mL of sterile 2% magnesium chloride and 20 mL of sterile 10% glucose. After 24 hours of growth, supplements were added to the cultures as needed. The final concentrations of supplements were: 5 mM homoserine, 5 mM arginine, 2.5 mM folic acid, 5 mM aspartic acid. After 72 hours of growth, cells were pelleted at 4000 x g and cell-free supernatants were used for extractions. The cell-free supernatants were dried under vacuum and resulting dry solids were extracted with 90% ethanol in water as described previously.[86] Extractions were then dried under vacuum and resuspended in water at 1/20th the volume of the original culture.

Isotope feeding of WH6 cultures

Isotope feeding experiments were performed as described above, except that the desired amino acids were added to the media before cultures were inoculated with WH6. The isotope labeled amino acids were used in place of their corresponding non-labeled amino acids as needed. Also, all cultures were supplemented with either 5 mM homoserine or aspartic acid in addition to 5 mM arginine. To determine the percent incorporation of isotopes, the relative abundances of the expected mass shifts were compared to the parent

peak. We then subtracted the natural isotope distribution of the parent peak to find a corrected percent incorporation. Since the aspartic acid trials had multiple isotope incorporations, we subtracted the natural distribution of GOVG with fewer isotopes incorporated when calculating the percentage of GOVG with more isotopes incorporated.

Derivatization of samples for LC-MS analysis

Extraction samples were derivatized by 9-Fluorenylmethoxycarbonyl chloride (Fmoc-Cl) or pentafluorophenylhydrazine before analysis by LC-MS. Underivatized samples were analyzed by LC-MS without dilution. For Fmoc derivatization, 25 µL of extraction samples was added to 50 µL of 200 mM sodium borate buffer pH 9.0. To this solution, 40 µL of 7.5 mM Fmoc-Cl dissolved in acetonitrile was added and samples were incubated at room temperature for 10 min. Undiluted samples were then analyzed by positive mode LC-ESI-MS. For pentafluorophenylhydrazine derivatization, 2.5 µL of 50 mM Tris pH 7.5 and 0.5 µL of pentafluorophenylhydrazine were added to 22 µL of extraction samples. Samples were then incubated at 42 ºC for 1.5 hours. After derivatization samples were diluted 1:2 before analysis by negative mode LC-ESI-MS.

Protein Expression and Purification

GvgA, gvgC, gvgF, and *gvgH* individually cloned into the pLIC-His vector using the ligation independent method described previously.[100] *Escherichia coli* BL21 (DE3) cells were transformed with pLIC-His containing a single gene of the *gvg* cluster. Proteins were expressed and purified as described in Chapter 2. The lysis/wash buffer used for Gvg protein Ni columns contained 50 mM tris pH 7.5, 100 mM NaCl, 10% glycerol, and 15 mM imidazole. The elution buffer contained 50 mM tris pH 7.5, 100 mM NaCl, 10% glycerol, and 500 mM imidazole. The storage buffer, which was also used for size exclusion

chromatography, contained 50 mM tris pH 7.5, 150 mM NaCl, and 10% glycerol. GvgC, GvgF, and GvgG were purified by both Ni-affinity and size exclusion chromatography. GvgA was purified only by Ni-affinity chromatography, and buffer exchanged into storage buffer using a PD-10 desalting column.

GvgC CO binding assays

Purified GvgC was diluted to a final concentration of 10 µM in 200 mM potassium phosphate buffer at pH 8.0. An absorbance scan from 380–500 nm was recorded. A sample of 50 µL of 0.4 M sodium dithionite was added to the sample. Another absorbance scan was recorded. The sample of GvgC was then placed in a chamber containing carbon monoxide for 30 min. Then the final absorbance scan was recorded.

GvgC NADH consumption assays

The GvgC activity shown in Figure 4.8C was measured by first preparing a solution of 50 mM tris at pH 7.5 containing100 µM FAD, 1 mM NADH, 1 mM NADPH, and 1 mM NH_4Cl. Addition of 10 µM GvgC initiated the reaction and brought the final volume to 200 µL. The absorbance from 300-400 nm was recorded every 2 minutes using a Tecan Infinite M1000 Pro. Controls were performed lacking FAD and GvgC. Additional controls shown in Figure 4.8D were measured as described above, but these samples had either NADH or NADPH, or neither cofactor.

GvgF carbamoyl transferase assays

Each reaction contains 5 mM $MgCl_2$, 5 mM ATP, 2 mM DTT, 2.5 mM carbamoyl phosphate, and 1 mM of an amino acid substrate. The reaction was initiated by addition of GvgF to a final concentration of 10 µM in 50 mM Tris at pH 7.5. The 20 proteinogenic amino acids as well as ornithine were tested as substrates. The samples were incubated for 1

hour at 28 ºC and quenched by adding an equal volume of acetonitrile and incubating at −20 °C for 20 minutes. Samples were spun at 20800 x g for 10 minutes to remove protein precipitant. The supernatant was derivatized using Fmoc-Cl by adding 25 µL of reaction mixture to 140 µL of 53 mM borate buffer at pH 9.0 in 5:9 water:acetonitrile, and then adding 25 µL of 18 mM Fmoc-Cl dissolved in acetonitrile. The samples were incubated at room temperature for 10 min and then analyzed by LC-MS. The same assays were also performed using GvgC and GvgA in place of GvgF. These samples were quenched and derivatized and analyzed using the same procedures.

GvgH in vitro activity assays

Each reaction was initiated by the addition of GvgH to a final concentration of 10 µM. The other components of the reaction were 50 mM tris at pH 7.5, 1 mM *O*-succinylhomoserine (*O*-SH), and 1 mM α-ketoglutarate. The reactions were incubated at 28 ºC for up to 2 hours before either Fmoc or pentafluorophenylhydrazine derivatization. Fmoc derivatization was performed as described in GvgF quenching. Pentafluorophenylhydrazine derivatization involves adding 2 µL of 100 mM pentafluorophenylhydrazine to 50 µL of unquenched sample and incubating the mixture at 42 ºC for 1 hour. Pentafluorophenylhydrazine derivatized samples were analyzed by negative mode LC-ESI-MS.

GvgH activity was also tested using WH6 WT extracts as substrate. Supernatants of WT cultures were extracted as described above. 25 µL of the extracts were buffered with a final concentration of 50 mM tris pH 7.5 before the addition of 10 µM GvgH to a final volume of 50 µL. The extracts were then quenched and derivatized by Fmoc-Cl, as described in the GvgF assays, before analysis by LC-MS.

Results

P. fluorescens WH6 cultures produce AOVG and GOVG in addition to FVG

We confirmed that cultures of *P. fluorescens* WH6 grown in modified *Pseudomonas* minimal salts media (PMS) produce FVG. After cultures were extracted as previously described,[87] they were derivatized using Fmoc-Cl to achieve better separation and signal for analysis by LC-MS. In addition to FVG, we detected aminooxyvinylglycine (AOVG) and guanidinooxyvinylglycine (GOVG) by LC-MS (Figure 4.4), two new products associated with the *gvg* cluster. We also found that supplementing the PMS media with homoserine, arginine, and folic acid modulated the production of all three oxyvinylglycines. The addition of homoserine and arginine increased the overall production of all three metabolites, while folic acid increased the relative amount of FVG with respect to GOVG (Figure 4.4).

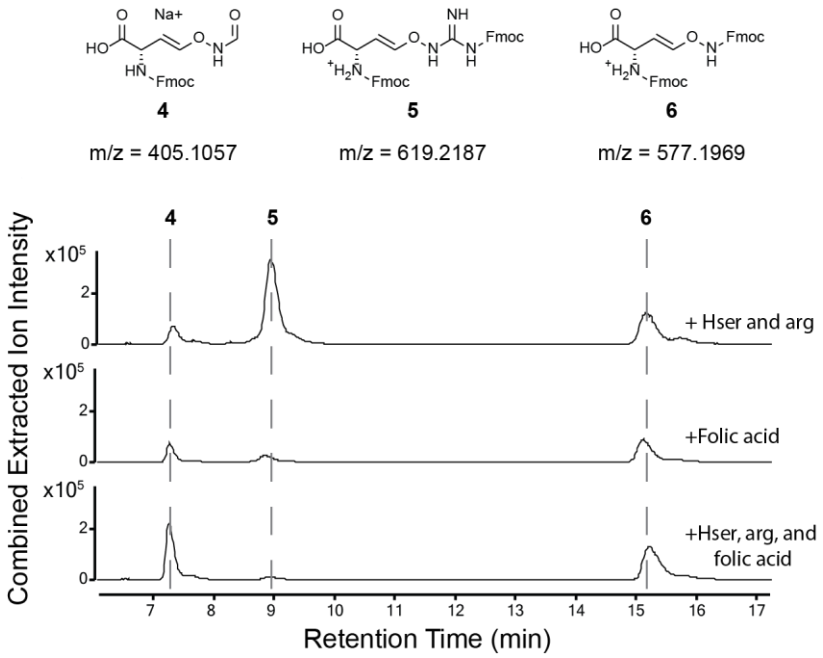


Figure 4.4 – *P. fluorescens* WH6 produced AOVG and GOVG in addition to FVG. Additional supplements added to the cultures changed the relative amounts of each oxyvinylglycine produced. Homoserine (Hser) and arginine (arg) were added at 5 mM when included, and folic acid was added at 2.5 mM when included. Samples were Fmoc derivatized before analysis. The traces shown are combined extracted ion chromatograms of all three metabolites.

To determine which enzymes are necessary for FVG biosynthesis, and identify possible intermediates in the pathway, we performed extractions on mutant strains of WH6 provided by our collaborators.[89] Each mutant contains a deletion of a single gene of the gvg cluster and was growth for metabolite extractions. We confirmed that *gvgA*, *gvgC*, *gvgF*, *gvgH*, and *gvgI* were necessary to produce FVG (Figure 4.5). While most mutant strains did not produce any oxyvinylglycines, the Δ*gvgI* mutant produced AOVG and GOVG (Figure 4.5), suggesting that GvgI catalyzes a late stage formylation of AOVG to FVG. We did not

detect any new intermediates in the other mutant strains by LC/MS in underivatized, Fmoc derivatized, or pentafluorophenylhydrazine derivatized extracts.

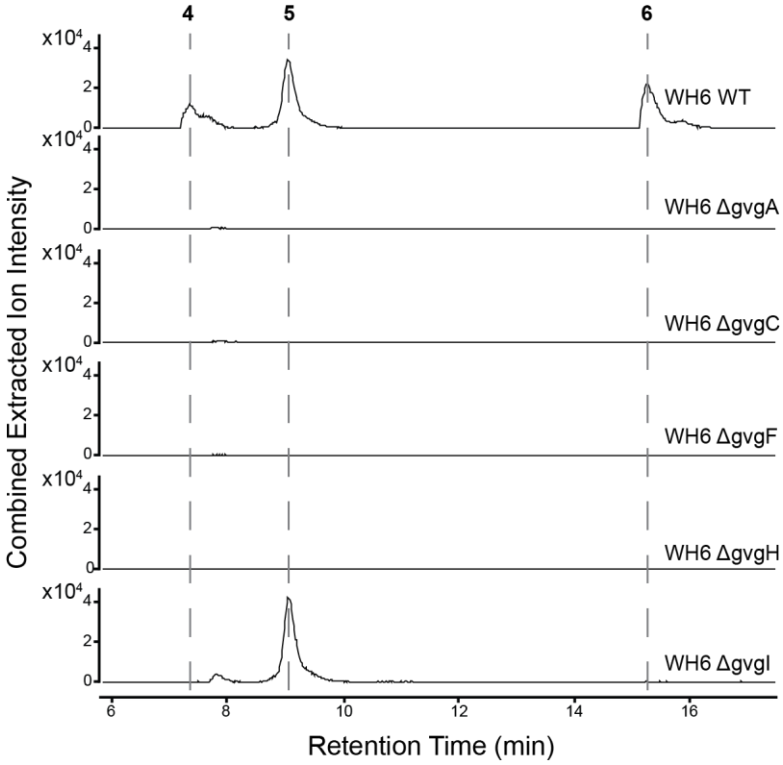


Figure 4.5- The *P. fluorescens* WH6 WT strain produces AOVG and GOVG and is the only strain to produce FVG. The ΔgvgI strain produces GOVG, and a very small amount of AOVG. None of the other knockouts produced any oxyvinylglycines. Each culture was grown with 5 mM homoserine and 5 mM arginine supplements. The extracts were derivatized by Fmoc-Cl before LC/MS analysis.

Homoserine and aspartic acid are related to the precursor of FVG

We then performed isotope feeding studies to uncover the precursors of the oxyvinylglycine core of FVG. When ^{15}N homoserine was added to the cultures, ^{15}N was incorporated into 40% of AOVG based on analysis by MS. Similarly, when arginine with both amidino nitrogens labeled with ^{15}N was supplemented in cultures, a single ^{15}N was incorporated into 34% of AOVG (Figure 4.6, Table 4.1). The marginal incorporation of both

homoserine and arginine when such an excess of each was supplemented to cultures suggests that homoserine and arginine are related biosynthetically to the precursors, but not the native substrates of the pathway. Additionally, we performed MS/MS on the isotope enriched AOVG products, which revealed the location of ^{15}N incorporation was dependent on which isotope labeled substrate was added (Figure 4.6). The ^{15}N of homoserine was incorporated at the backbone nitrogen of AOVG and the ^{15}N of arginine was incorporated at the side chain nitrogen.

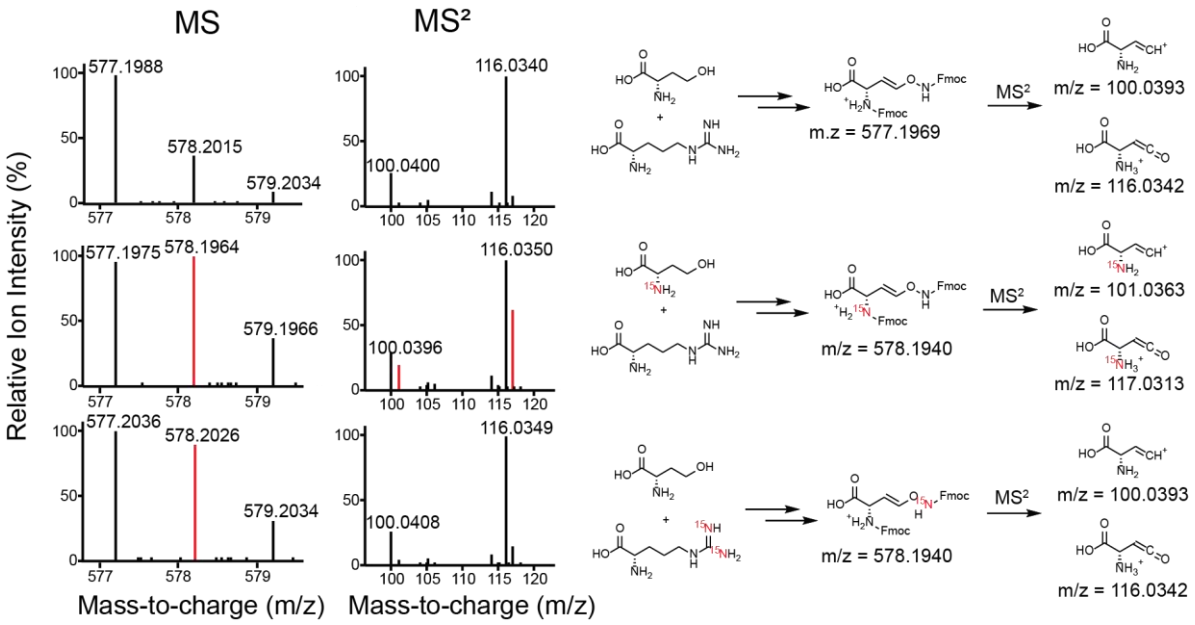


Figure 4.6 – Feeding of *P. fluorescens* WH6 cultures with isotope labeled homoserine and arginine shows that backbone of FVG metabolites is from homoserine related metabolites, and the side chain nitrogen is from arginine guanidine related metabolites. The MS traces show mass shifts of AOVG from extractions using isotope enriched homoserine or arginine as supplements. The MS^2 traces are tandem data of labeled dehydro-canaline, which show the location of isotope incorporation. The data is summarized by the schemes on the right.

Labeled amino acid	Unlabeled	^{15}N incorporation
None	99%	1%
Homoserine	60%	40%
Arginine	66%	34%

Table 4.1 – Isotope incorporation from homoserine and arginine into AOVG. Percentages are the amount of AOVG that is labeled by each isotope. The 1% labeling without the addition of labeled amino acids shows the uncertainty of measurements. The values listed are from a single trial.

Isotope labeled aspartic acid was also used to determine the origin of the carbon backbone of FVG. We used aspartic acid in place of homoserine because ^{13}C labeled homoserine is not commercially available. With aspartic acid being supplemented instead of homoserine, the production of GOVG was the greatest among the three metabolites, therefore, this product was used for analysis of percentage incorporation. Aspartic acid with the backbone ^{15}N was incorporated at a similar percentage to ^{15}N homoserine, suggesting that it was also labeling the alpha nitrogen of GOVG, although MS2 experiments were not performed to confirm the location. When aspartic acid with all carbons labeled with ^{13}C and the nitrogen labeled with ^{15}N was fed into cultures of WH6, we observed a complex isotope distribution (Figure 4.7), suggesting that the ^{15}N and carbon backbone are not always incorporated together. For GOVG, 24% had just the ^{15}N incorporated, 7% had only the ^{13}C backbone incorporated, and 17% had both the ^{15}N and ^{13}C isotopes incorporated (Table 4.2). Since the carbon backbone of aspartic acid is never partially incorporated but can be separately incorporated compared to the labeled nitrogen, aspartic acid is likely metabolized before entering FVG biosynthesis. Thus, the precursor for GOVG is predicted to be a metabolite related to aspartic acid.

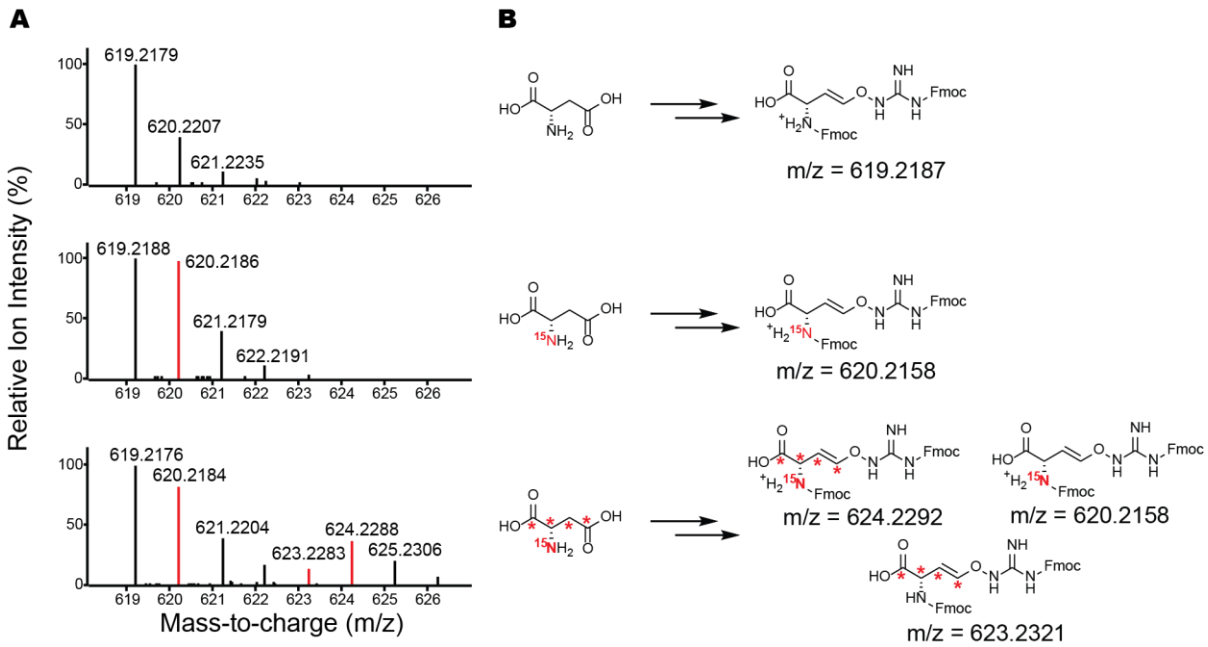


Figure 4.7 - Feeding of *P. fluorescens* WH6 cultures with ^{13}C and ^{15}N isotope labeled aspartic acid shows that the oxyvinylglycine backbone comes from aspartic acid related metabolites. With aspartic acid feeding, GOVG had the highest production of the oxyvinylglycines, and was used for analysis. Each culture contained 5 mM aspartic acid and 5 mM arginine and samples were Fmoc derivatized before analysis by LC-MS. ^{13}C location are marked by *.

Label of aspartic acid	Unlabeled	^{15}N	^{13}C	^{15}N and ^{13}C
None	98%	2%	0%	0%
15N labeled	62%	37%	1%	0%
13C and 15N labeled	52%	24%	7%	17%

Table 4.2 – Isotope incorporation from aspartic acid into GOVG. Percentages are shown for GOVG that is labeled by each isotope. The 2% labeling without the addition of labeled amino acids shows the uncertainty of measurements. The values listed are from a single trial.

In vitro analysis of Gvg enzymes

We expressed the Gvg biosynthetic enzymes in *E. coli* and purified them to reconstitute their activities *in vitro*. Purified GvgC in solution was red with a UV peak at 420 nm, and analysis by LC-MS identified that GvgC copurified with heme (Figure 4.8A and B). A portion of GvgC has homology to the heme oxygenase class of enzymes, which catalyze the breakdown of heme into biliverdin. However, GvgC contains an intact heme B, suggesting that it performs different chemistry than heme oxygenases in which the heme is degraded. We then tested whether GvgC could bind to carbon monoxide similar to p450 heme proteins. Here, we observed an absorbance shift in the presence of the reducing agent dithionite, suggesting that the iron center was reduced from Fe(III) to Fe(II). However, we did not detect an absorbance shift after the addition of carbon monoxide, suggesting that GvgC has different reactivity than p450 enzymes and globin domains, and may not bind oxygen (Figure 4.8B). Next, we tested the activity of GvgC in the presence of the cofactors NADH, NADPH, and FAD. In this assay, both NADH and NADPH exhibited maximum absorbance at 372 nm, and their oxidized forms did not. GvgC catalyzed the oxidation of NADH or NADPH in the presence of a catalytic amount of FAD (Figure 4.8C). Since the assays in Figure 4.8C contained both NADH and NADPH, to determine which cofactor was being oxidized the assays were repeated with a single cofactor in each sample, which showed that GvgC selectively oxidizes NADH in the presence of FAD (Figure 4.8D). However, when possible precursors for FVG biosynthesis were included, such as amino acids and nitrogen sources, we did not detect any oxidative modifications to the possible substrates. Together, this data shows that GvgC is active *in vitro*, and the heme center is capable of oxidative chemistry, but we have not identified the physiological substrate of the enzyme.

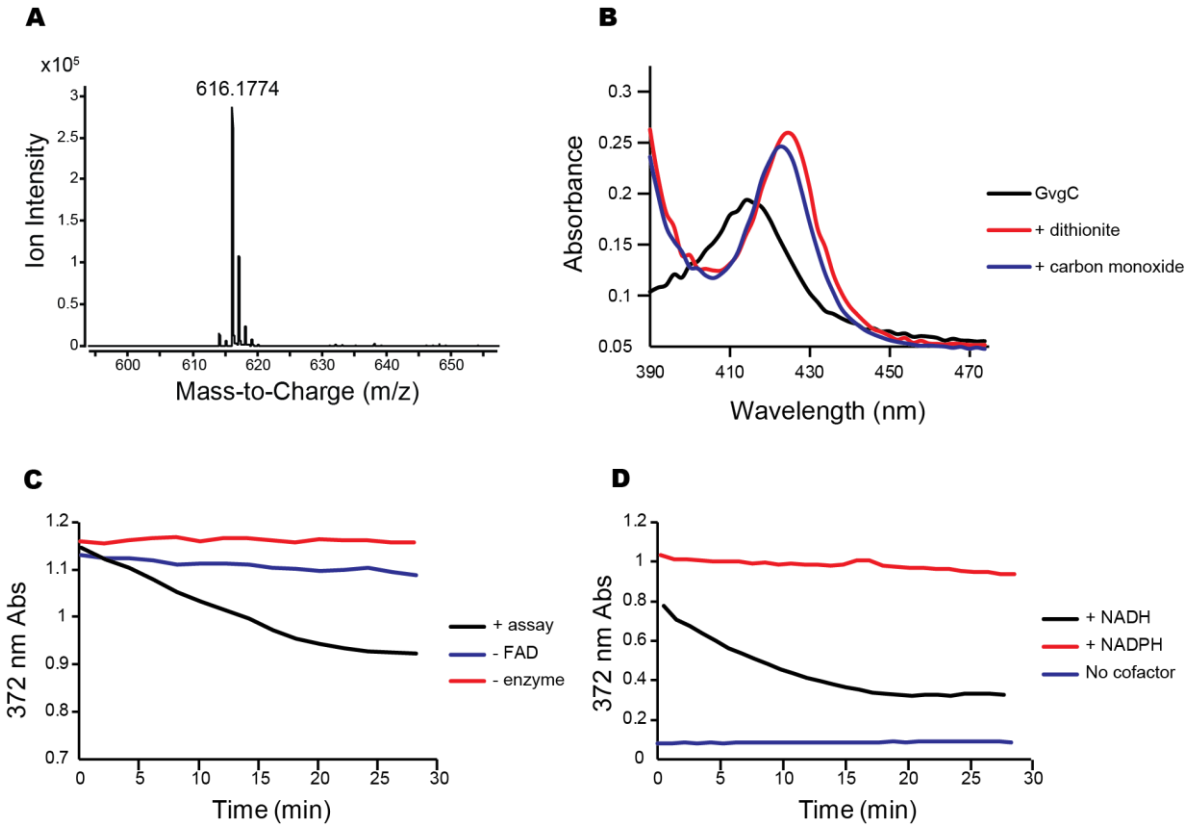


Figure 4.8 – GvgC does not bind carbon monoxide, but consumes NADH in the presence of FAD. A) Mass spectrum of heme b that copurified with GvgC. Calculated m/z = 616.1676 B) The iron center of GvgC is reduced by dithionite as seen by a change in absorbance. However, the reduced GvgC does not bind carbon monoxide since there is no absorbance shift when carbon monoxide is introduced. C) NADH and NADPH have absorbance at 372 nm. Loss of absorbance in the "+ assay" is due to oxidation of cofactors by GvgC. Both FAD and GvgC are necessary for oxidation to occur. D) The loss of absorbance is only in samples containing NADH. There is no change in samples with NADPH or samples with neither cofactor.

GvgH is a homolog of PLP-dependent aminotransferases. Purified GvgH was yellow in solution and displayed a UV absorbance spectrum consistent with PLP. When GvgH was incubated individually with each of the 20 canonical amino acids or homoserine, no amino

transferase activity was observed. However, when *O*-succinyl-L-homoserine (*O*-SH) was incubated with GvgH, consumption of *O*-SH was observed, but no new products were observed after Fmoc derivatization (Figure 4.9A). Since GvgH is predicted to be an aminotransferase, the product is expected to be an alpha-keto acid lacking the primary nitrogen necessary for Fmoc derivatization. Instead, we use the derivatizing agent pentafluorophenylhydrazine, which has been shown to derivatize alpha-keto acids.[47] Using the new derivatization method, we detected the cleavage product α-ketobutyric acid (Figure 4.9A). While the activity is clearly due to GvgH, the rate of the reaction was very slow, taking over two hours to reach completion, suggesting that this may not be the native substrate of the enzyme. Instead, other naturally occurring homoserine analogues such as *O*-acetyl-homoserine or aspartate semialdehyde may be the native substrate. Additionally, we found that GvgH could consume AOVG as a substrate from WH6 extracts (Figure 4.9B), however, due to the low concentration of AOVG in these extracts, we could not detect a new product in the extract mixture. GvgH is likely responsible for the production of AOVG in *P. fluorescens* WH6, and the substrate for this reaction is chemically similar to *O*-SH, but the origin of the N-O bond in AOVG is still unclear.

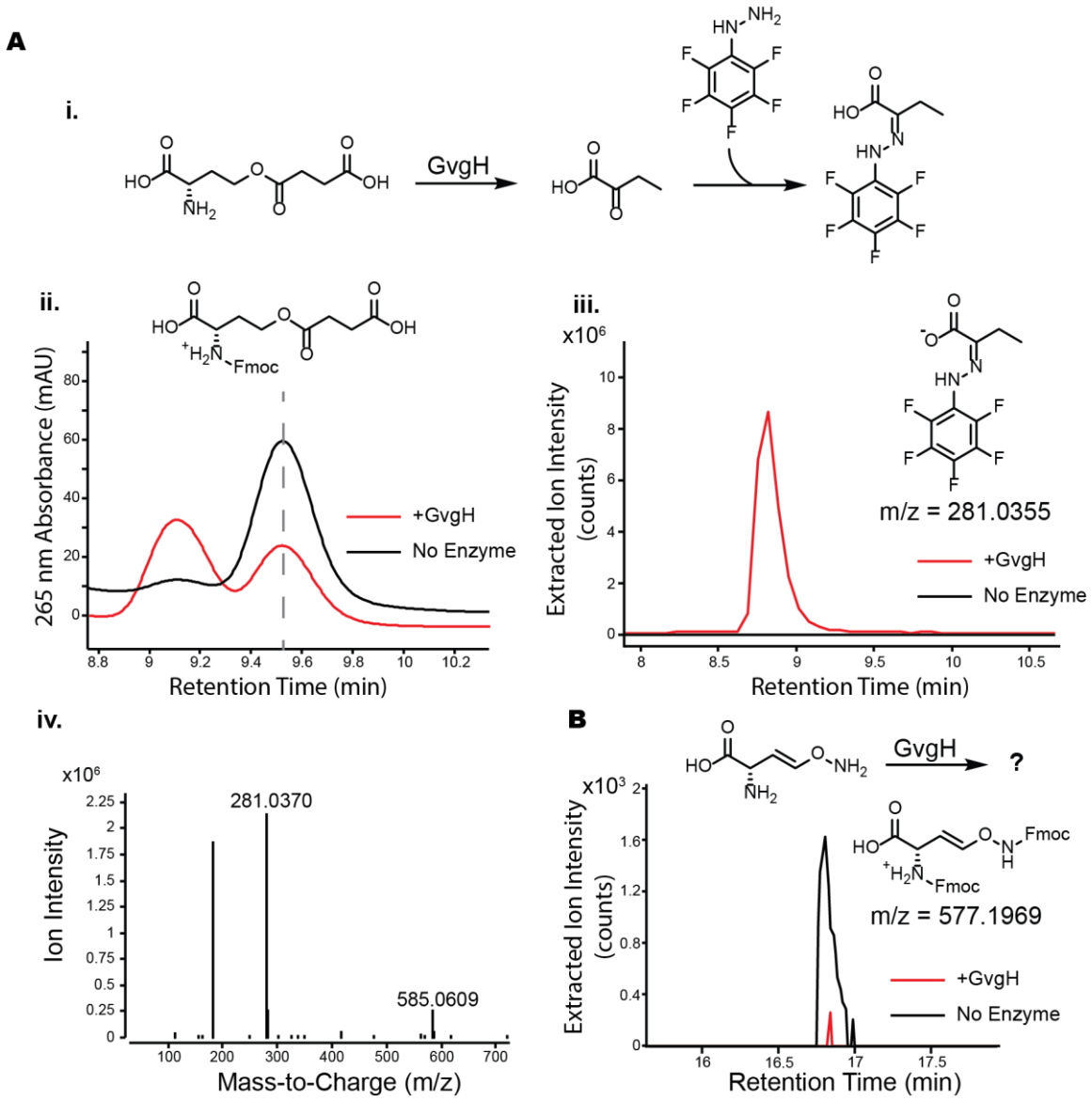


Figure 4.9 – Measuring GvgH activity using *O*-succinyl homoserine and culture extracts as substrates. In all traces, red traces are assays containing GvgH and black traces are controls without enzyme. A) i) Scheme of *in vitro* GvgH reaction and derivatization. ii) Consumption of O-succinyl homoserine by GvgH shown by absorbance after derivatization by Fmoc-Cl. iii) Conversion of O-succinyl homoserine to the cleaved alpha-keto product derivatized by pentafluorophenylhydrazine, show by EIC. iv) Mass spectrum of GvgH reaction. The labeled peaks correspond to [M-H]- = 281.0355 and [2M-2H+Na]- = 585.0602. B) Consumption of

AOVG from *P. fluorescens* WH6 extracts in the presence of GvgH. Extracts were derivatized by Fmoc-Cl for analysis.

We next tested the carbamoyl transferase activity of GvgF *in vitro* to identify a possible substrate for the *gvg* cluster and a substrate for GvgC or GvgH. Using each of the 20 proteinogenic amino acids and ornithine as substrates, we observed that GvgF was selective for ornithine as a substrate. In the assay, GvgF catalyzed the transfer of the carbamoyl group of carbamoyl-phosphate to ornithine to form citrulline (Figure 4.10B). Even the amino acid lysine, which has a side chain with one more carbon, showed no activity. However, GvgA and GvgC each produced significantly more citrulline than GvgF when incubated with ornithine, carbamoyl phosphate, and magnesium chloride (Figure 4.10C). The fact that all three enzymes catalyze the production of citrulline suggests that the observed activity is not the native function of these enzymes, and instead weak binding of ornithine and cosubstrates to the enzymes may be enough to catalyze the formation of citrulline. Although, the fact that all three enzymes interact with ornithine suggests the native substrates of these enzymes may be chemically similar to ornithine. Therefore, ornithine is likely not the native substrate of GvgF, and a metabolite with similar functional groups could be one of the precursors for FVG.

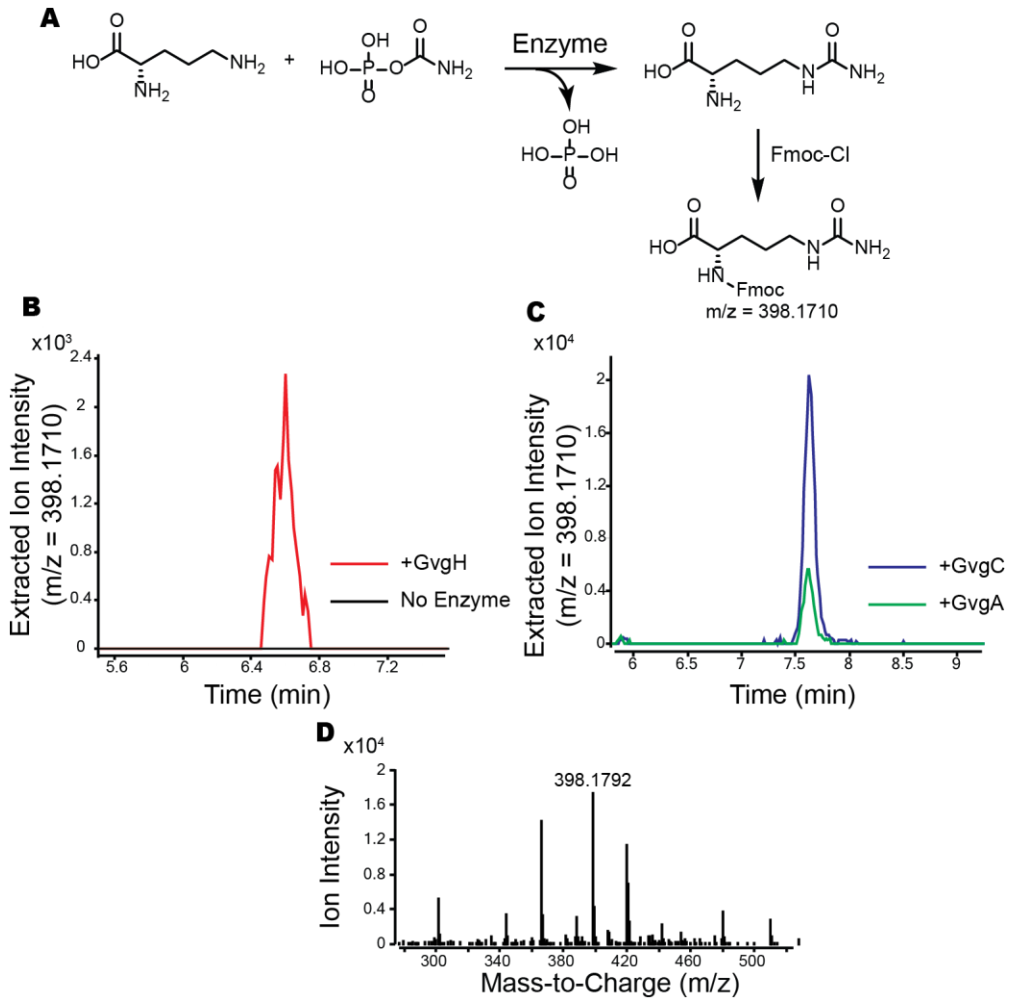


Figure 4.10 – GvgF catalyzes the formation of citrulline from ornithine over 1 hour, but the rate is slow. A) Scheme of *in vitro* citrulline formation and derivatization. B) Assays containing GvgF (red) produce citrulline, while assays lacking GvgF (black) do not. C) Assays containing GvgA (green) and GvgC (blue) produce more citrulline than GvgF (red) even though GvgA and GvgC are not predicted to have carbamoyl transferase activity. D) Mass spectrum of citrulline formation by GvgC, calculated [M+H] = 387.1710

Discussion

The *gvg* gene cluster has been shown to be responsible for FVG production, but the biosynthesis of FVG had been unknown.[89] By performing metabolomic analyses of *P.*

fluorescens WH6 WT and mutant strains we have revealed new insights into the biosynthesis of FVG. We identified two new products of the *gvg* gene cluster, GOVG and AOVG, and the structures of these metabolites in comparison to FVG suggest that AOVG is an intermediate in the production of FVG and GOVG (Figure 4.4). In each case, the conversion of AOVG into either product would require the addition of a single functional group, a formyl group for FVG or amidino group for GOVG. These additions are possible due to the *gvg* cluster containing both a putative formyl transferase and amidino transferase, GvgI and GvgD, respectively (Figure 4.1). The ability of the Δ*gvgI* mutant to produce a greater amount of GOVG further supports the hypothesis that AOVG is a branching point in biosynthesis of FVG and GOVG. Also, the cofactor for formyl transferases is 10-formyltetrahydrofolate, which is derived from folic acid. Therefore, when folic acid was added to WH6 cultures, the production of FVG increased and the production of GOVG decreased, indicating that there are two pathways competing for the same precursor, AOVG. While it is rare, this is not the first example in the literature of a single gene cluster producing multiple metabolites.[101,102]

While the metabolomics analysis supports the conclusion that the final biosynthetic step in the production of FVG involves formylation of AOVG, the biosynthesis of the oxyvinylglycine core is still unknown. Previous studies have implicated homoserine as a precursor for the oxyvinylglycine AVG,[95] however the results of the isotope enriched amino acid feeding studies of the *gvg* were inconclusive. When ^{15}N enriched homoserine was added to cultures of WH6, the backbone nitrogen of AOVG was only 40% labeled with ^{15}N, indicating that homoserine may only be metabolically close to the precursor of FVG (Table 4.1). It is possible that homoserine was being metabolized, and then the labeled nitrogen was incorporated after separating from the rest of the homoserine carbon backbone. This

possibility could not be directly tested due to the fact that ^{13}C labeled homoserine is not commercially available, and instead ^{13}C labeled aspartic acid was used as a substitute.

Aspartic acid is metabolically related to homoserine, and within cells the two metabolites can be interconverted enzymatically in two steps. Importantly, the carbon backbone of aspartic acid is conserved during the biosynthesis of homoserine. Additionally, ^{15}N labeled aspartic acid was incorporated into the oxyvinylglycine core of GVG at a similar rate to ^{15}N homoserine. When ^{13}C and ^{15}N labeled aspartic acid was supplemented into cultures of WH6, the labeled carbons were always incorporated together, but could be separately incorporated from the ^{15}N (Figure 4.7, Table 4.2), supporting the hypothesis that transformation of aspartic acid needs to occur before modification by the enzymes from the FVG pathway. Since both homoserine and aspartic acid appear to be metabolically related to the precursor of FVG, and the entire carbon backbone of aspartic acid is incorporated into the product, the precursor of FVG is expected to be in the aspartate-threonine pathway. To elucidate the biosynthesis and precursor of the AOVG core a series of *in vitro* studies were performed using purified enzymes from the *gvg* cluster.

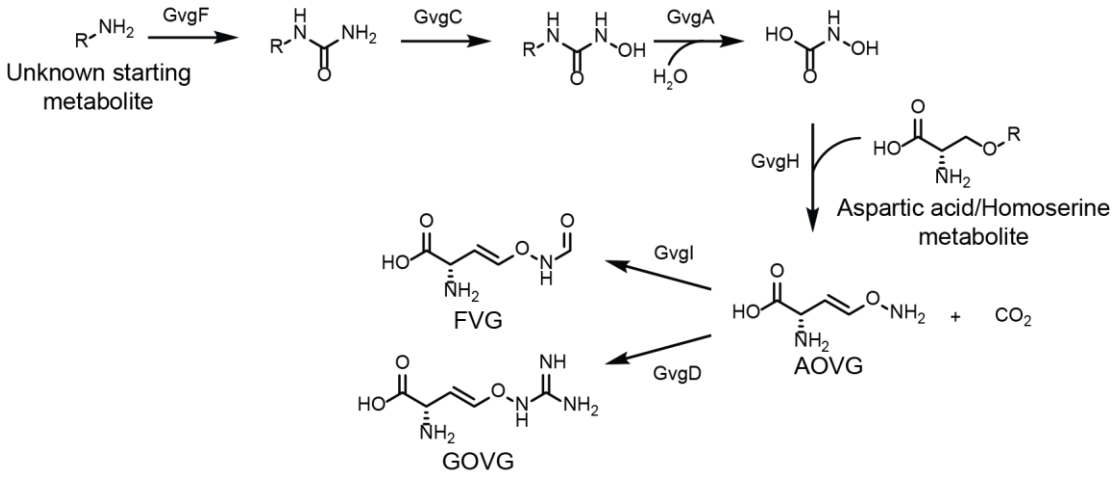


Figure 4.11 – Proposed biosynthesis of FVG and GOVG from AOVG by enzymes from the *gvg* cluster.

The Δ*gvgA*, Δ*gvgC*, Δ*gvgF*, and Δ*gvgH* mutants did not produce any AOVG, GOVG, or FVG, suggesting that all four encoded enzymes are necessary to produce the oxyvinylglycine core. Of these enzymes, GvgC copurified with heme and was found to catalyze the oxidation of NADH in the presence of FAD (Figure 4.8). Since heme enzymes are capable of forming heteroatom-heteroatoms bonds, including N-O and N-N bonds,[99,103] we propose that GvgC performs redox chemistry in the formation of FVG. However, no modification to substrates or formation of heteroatom bonds was observed when GvgC was incubated with possible amino acid and nitrogen containing substrates *in vitro*. GvgF is putatively assigned as a carbamoyl transferase and we demonstrated that this enzyme exhibited some activity to convert ornithine to citrulline, but the rate of this reaction was very slow. Both GvgC and GvgA also performed this reaction *in vitro* to a greater extent than GvgF, suggesting that ornithine is not the native substrate of GvgF (Figure 4.10). GvgF may instead utilize a substrate that is chemically similar to ornithine in FVG biosynthesis, but additional amino acids and primary amine containing substrates tested with GvgF did not show any activity. For GvgH, we confirmed that it is a PLP-dependent enzyme and showed that it catalyzed two reactions *in vitro.* First, GvgH uses AOVG from extracts of WH6 as a substrate, but no product was detected due to the low starting concentration of AOVG in the complex mixture (Figure 4.9B). PLP-dependent enzymes are reversible catalysts, and this *in vitro* activity suggests the native product of GvgH is AOVG. Second, GvgH catalyzed the conversion of *O*-SH to α-ketobutyric acid, but this reaction was not a productive step in the biosynthesis of FVG (Figure 4.9A), thus *O*-SH may not be the native substrate of GvgH. It is possible that GvgH did not show productive catalysis *in vitro* because a second substrate is required to react with the elimination product of *O*-SH that is still bound to GvgH. We

hypothesize the unknown second substrate is the source of the N-O bond found in AOVG. However, the addition of common metabolites that contain N-O bonds, such as hydroxyurea, nitrate, and nitrite had no effect on the assay, and hydroxylamine actually inhibited the activity. Because homologs of the *gvgA, gvgC,* and *gvgF* genes often cluster in pathways other than the *gvg* cluster, these genes may encode enzymes that synthesize a different and rare N-O bond containing compound. These enzymes likely function to produce a metabolite for uses beyond FVG biosynthesis (Figure 4.11). Additional *in vitro* experiments will be necessary to characterize the native functions of enzymes encoded by the *gvg* cluster and determine the structures of intermediates of the pathway.

Conclusion and Future Directions

Analysis of extractions of *P. fluorescens* WH6 revealed that in addition to FVG, the *gvg* cluster produces the oxyvinylglycines AOVG and GOVG, and that AOVG is likely a common intermediate in the biosynthesis of FVG and GOVG. However, elucidating the biosynthesis of AOVG has been more challenging. Isotope labeling data suggest the precursor of AOVG is part of the aspartate-threonine pathway, but *in vitro* assays were unable to confirm this hypothesis. *In vitro* assays of GvgC were able to identify the cofactors of an uncharacterized class of heme-dependent enzymes. Also, GvgH may catalyze the formation of the oxyvinylglycine core, but additional studies are needed. The data shown here provide a foundation to continue studying FVG biosynthesis, and identify the precursors of the *gvg* pathway, as well as the source of the N-O bond. Once the native functions of these enzymes are elucidated, additional clusters that produce uncharacterized oxyvinylglycines or N-O bond containing compounds can be identified through bioinformatic searches. Additionally, introducing non-native substrates *in vitro* would allow for the biosynthesis of

unnatural oxyvinylglycines with different bioactivates. Also, GvgC represents an uncharacterized class of heme-dependent enzymes, and understanding the chemistry it performs would allow the class to be utilized for novel biocatalytic reactions in synthetic schemes. Overall, there is still much to be discovered from studying the *gvg* gene cluster that would have impacts across chemistry, biochemistry, and enzymology.

REFERENCES

(1) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.

(2) Bugg, T. D. H. The Development of Mechanistic Enzymology in the 20th Century. *Nat. Prod. Rep.* **2001**, *18* (5), 465–493.

(3) Patel, R. N. Synthesis of Chiral Pharmaceutical Intermediates by Biocatalysis. *Coord. Chem. Rev.* **2008**, No. 252, 659–701.

(4) Zhang, N.; Stewart, B. G.; Moore, J. C.; Greasham, R. L.; Robinson, D. K.; Buckland, B. C.; Lee, C. Directed Evolution of Toluene Dioxygenase from Pseudomonas Putida for Improved Selectivity toward Cis-Indandiol during Indene Bioconversion. *Metab. Eng.* **2000**, *2* (4), 339–348.

(5) Chartrain, M.; Jackey, B.; Taylor, C.; Sandford, V.; Gbewonyo, K.; Lister, L.; Dimichele, L.; Hirsch, C.; Heimbuch, B.; Maxwell, C.; et al. Bioconversion of Indene to Cis (1S , 2R) Indandiol and Tram (1R , 2R) Indandiol by Rhodococcus Species. - *journal Ferment. Bioeng.* **1998**, *86* (6), 550–558.

(6) Tamarez, M.; Morgan, B.; Wong, G. S. K.; Tong, W.; Bennett, F.; Lovey, R.; McCormick, J. L.; Zaks, A. Pilot-Scale Lipase-Catalyzed Regioselective Acylation of Ribavirin in Anhydrous Media in the Synthesis of a Novel Prodrug Intermediate. *Org. Process Res. Dev.* **2003**, *7* (6), 951–953.

(7) Fusari, S. A.; Frohardt, R. P.; Ryder, A.; Haskell, T. H.; Johannessen, D. W.; Elder, C. C.; Bartz, Q. R. Azaserine, a New Tumor-Inhibitory Substance. Isolation and Characterization [1]. *J. Am. Chem. Soc.* **1954**, *76* (11), 2878–2881.

(8) Dejong, C. A.; Chen, G. M.; Li, H.; Johnston, C. W.; Edwards, M. R.; Rees, P. N.; Skinnider, M. A.; Webster, A. L. H.; Magarvey, N. A. Polyketide and Nonribosomal Peptide Retro-Biosynthesis and Global Gene Cluster Matching. *Nat. Chem. Biol.* **2016**, *12* (12), 1007–1014.

(9) Hara, M.; Asano, K.; Kawamoto, I.; Takiguchi, T.; Katsumata, S.; Takahashi, K.; Nakano, H. Leinamycin, a New Antitumor Antibiotic from Streptomyces: Producing Organism, Fermentation and Isolation. *J. Antibiot. (Tokyo).* **1989**, *42* (12), 1768–1774.

(10) Maiese, W. M.; Lechevalier, M. P.; Lechevalier, H. A.; Korshalla, J.; Kuck, N.; Fantini, A.; Wildey, M. J.; Thomas, J.; Greenstein, M. Calicheamicins, a Novel Family of Antitumor Antibiotics: Taxonomy, Fermentation, and Biological Properties. *J. Antibiot. (Tokyo).* **1989**, *42* (4), 558–563.

(11) Chakrabarty, A. M. PLASMIDS IN PSEUDOMONASl. *Annu. Rev. Genet.* **1976**, 7–30.

(12) Walsh, C. T.; Fischbach, M. A. Natural Products Version 2_0 Connecting Genes to

Molecules - Journal of the American Chemical Society (ACS Publications). *J. Am. Chem. Soc.* **2010**, 2469–2493.

(13) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402.

(14) Medema, M. H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 2), 339–346.

(15) Marchler-Bauer, A.; Lu, S.; Anderson, J. B.; Chitsaz, F.; Derbyshire, M. K.; DeWeese-Scott, C.; Fong, J. H.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; et al. CDD: A Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 1), 225–229.

(16) Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C. J.; Lu, S.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; et al. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45* (D1), D200–D203.

(17) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chemie - Int. Ed.* **2009**, *48* (26), 4688–4716.

(18) Strieker, M.; Tanović, A.; Marahiel, M. A. Nonribosomal Peptide Synthetases: Structures and Dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 234–240.

(19) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chem. Biol.* **1999**, *6* (8), 493–505.

(20) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. Specificity Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Transductive Support Vector Machines (TSVMs). *Nucleic Acids Res.* **2005**, *33* (18), 5799–5808.

(21) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2 - A Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 2), 362–367.

(22) Magarvey, N. A.; Haltli, B.; He, M.; Greenstein, M.; Hucul, J. A. Biosynthetic Pathway for Mannopeptimycins, Lipoglycopeptide Antibiotics Active against Drug-Resistant Gram-Positive Pathogens. *Antimicrob. Agents Chemother.* **2006**, *50* (6), 2167–2177.

(23) Han, L.; Vuksanovic, N.; Oehm, S. A.; Fenske, T. G.; Schwabacher, A. W.; Silvaggi,

N. R. *Streptomyces Wadayamensis* MppP Is a PLP-Dependent Oxidase, Not an Oxygenase. *Biochemistry* **2018**, acs.biochem.8b00130.

(24) Han, L.; Schwabacher, A. W.; Moran, G. R.; Silvaggi, N. R. Streptomyces Wadayamensis MppP Is a Pyridoxal 5'-Phosphate-Dependent l -Arginine α-Deaminase, γ-Hydroxylase in the Enduracididine Biosynthetic Pathway. *Biochemistry* **2015**, *54* (47), 7029–7040.

(25) Hedges, J. B.; Kuatsjah, E.; Du, Y.-L.; Eltis, L. D.; Ryan, K. S. Snapshots of the Catalytic Cycle of an O_2 , Pyridoxal Phosphate-Dependent Hydroxylase. *ACS Chem. Biol.* **2018**, acschembio.8b00039.

(26) Kawai, K.; Wang, G.; Okamoto, S.; Ochi, K. The Rare Earth, Scandium, Causes Antibiotic Overproduction in Streptomyces Spp. *FEMS Microbiol. Lett.* **2007**, *274* (2), 311–315.

(27) Moore, J. M.; Bradshaw, E.; Seipke, R. F.; Hutchings, M. I.; McArthur, M. *Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in Streptomyces Bacteria*, 1st ed.; Elsevier Inc., 2012; Vol. 517.

(28) Tanaka, Y.; Hosaka, T.; Ochi, K. Rare Earth Elements Activate the Secondary Metabolite-Biosynthetic Gene Clusters in Streptomyces Coelicolor A3(2). *J. Antibiot. (Tokyo).* **2010**, *63* (8), 477–481.

(29) Rigali, S.; Titgemeyer, F.; Barends, S.; Mulder, S.; Thomae, A. W.; Hopwood, D. A.; van Wezel, G. P. Feast or Famine: The Global Regulator DasR Links Nutrient Stress to Antibiotic Production by Streptomyces. *EMBO Rep.* **2008**, *9* (7), 670–675.

(30) Milshteyn, A.; Schneider, J. S.; Brady, S. F. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chem. Biol.* **2014**, *21* (9), 1211–1223.

(31) Wade, M.; Li, Y.-C.; M. Wahl, G. NIH Public Access. *Nat. Rev. Cancer* **2013**, *13* (2), 83–96.

(32) Shao, Z.; Luo, Y.; Zhao, H. Rapid Characterization and Engineering of Natural Product Biosynthetic Pathways via DNA Assembler. **2015**, *25* (8), 713–724.

(33) Shao, Z.; Zhao, H.; Zhao, H. DNA Assembler, an in Vivo Genetic Method for Rapid Construction of Biochemical Pathways. *Nucleic Acids Res.* **2009**, *37* (2), 1–10.

(34) Galm, U.; Shen, B. Expression of Biosynthetic Gene Clusters in Heterologous Hosts for Natural Product Production and Combinatorial Biosynthesis. *Expert Opin. Drug Discov.* **2006**, *1* (5), 409–437.

(35) Way, J. C.; Davis, M. A.; Morisato, D.; Roberts, D. E.; Kleckner, N. New Tn10 Derivatives for Transposon Mutagenesis and for Construction of LacZ Operon Fusions by Transposition. *Gene* **1984**, *32* (3), 369–379.

(36) Armstrong, D.; Azevedo, M.; Mills, D.; Bailey, B.; Russell, B.; Groenig, A.; Halgren, A.; Banowetz, G.; McPhail, K. Germination-Arrest Factor (GAF): 3. Determination That the Herbicidal Activity of GAF Is Associated with a Ninhydrin-Reactive Compound and Counteracted by Selected Amino Acids. *Biol. Control* **2009**, *51* (1), 181–190.

(37) Datta, S.; Costantino, N.; Court, D. L. A Set of Recombineering Plasmids for Gram-Negative Bacteria. *Gene* **2006**, *379* (1–2), 109–115.

(38) Sawitzke, J. A.; Thomason, L. C.; Bubunenko, M.; Li, X.; Costantino, N.; Court, D. L. *Recombineering: Using Drug Cassettes to Knock out Genes in Vivo*, 1st ed.; Elsevier Inc., 2013; Vol. 533.

(39) Covington, B. C.; McLean, J. A.; Bachmann, B. O. Comparative Mass Spectrometry-Based Metabolomics Strategies for the Investigation of Microbial Secondary Metabolites. *Nat. Prod. Rep.* **2017**, *34* (1), 6–24.

(40) Bessey, O.; Lowry, O.; Brock, M. A Method for the Rapid Determination of Alkaline Phosphatase with Five Cubic Millimeters of Serum. **1946**, No. 5, 321–329.

(41) Ngo, T. T.; Lenhof, H. M. A Sensitive and Versatile Chromogenic and Peroxidase-Coupled Assay for Peroxidase Reactions. **1980**, *397*, 389–397.

(42) Sun, Y.; Hahn, F.; Demydchuk, Y.; Chettle, J.; Tosin, M.; Osada, H.; Leadlay, P. F. In Vitro Reconstruction of Tetronate RK-682 Biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (2), 99–101.

(43) Sattely, E. S.; Fischbach, M. A.; Walsh, C. T. Total Biosynthesis: In Vitro Reconstitution of Polyketide and Nonribosomal Peptide Pathways. *Nat. Prod. Rep.* **2008**, *25* (4), 757.

(44) Roberts, A.; Ryan, K.; Moore, B.; Gulder, T. Total (Bio) Synthesis: Strategies of Nature and of Chemists. *Topp Curr Chem* **2010**, *297* (August), 149–203.

(45) Pahari, P.; Kharel, M. K.; Shepherd, M. D.; Van Lanen, S. G.; Rohr, J. Enzymatic Total Synthesis of Defucogilvocarcin M and Its Implications for Gilvocarcin Biosynthesis. *Angew. Chemie - Int. Ed.* **2012**, *51* (5), 1216–1220.

(46) Ju, J.; Ozanick, S. G.; Shen, B.; Thomas, M. G. Conversion of (2S)-Arginine to (2S,3R)-Capreomycidine by VioC and VioD from the Viomycin Biosynthetic Pathway of Streptomyces Sp. Strain ATCC11861. *ChemBioChem* **2004**, *5* (9), 1281–1285.

(47) Song, H.; Hu, W.; Naowarojna, N.; Her, A. S.; Wang, S.; Desai, R.; Qin, L.; Chen, X.; Liu, P. Mechanistic Studies of a Novel C-S Lyase in Ergothioneine Biosynthesis: The Involvement of a Sulfenic Acid Intermediate. *Sci. Rep.* **2015**, *5* (July), 1–13.

(48) Belecki, K.; Townsend, C. A. Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Enediyne

Polyketide Synthase CalE8. *J. Am. Chem. Soc.* **2013**, *135* (38), 14339–14348.

(49) Hamilton, G.; Westheimer, F. H. On the Mechanism of the Enzymatic Decarboxylation of Acetoacetate. *J. Am. Chem. Soc.* **1959**, *81*, 6332–6333.

(50) Midelfort, C. F.; Rose, I. A. A Stereochemical Method for Detection of ATP Terminal Phosphate Transfer in Enzymatic Reactions. Glutamine Synthetase. *J. Biol. Chem.* **1976**, *251* (19), 5881–5887.

(51) Rui, Z.; Li, X.; Zhu, X.; Liu, J.; Domigan, B.; Barr, I.; Cate, J. H. D.; Zhang, W. Microbial Biosynthesis of Medium-Chain 1-Alkenes by a Nonheme Iron Oxidase. *Proc. Natl. Acad. Sci.* **2014**, *111* (51), 18237–18242.

(52) Mehta, A. P.; Abdelwahed, S. H.; Xu, H.; Begley, T. P. Molybdopterin Biosynthesis: Trapping of Intermediates for the MoaA-Catalyzed Reaction Using 2′-DeoxyGTP and 2′-ChloroGTP as Substrate Analogues. *J. Am. Chem. Soc.* **2014**, *136* (30), 10609–10614.

(53) Ventola, C. L. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *P T A peer-reviewed J. Formul. Manag.* **2015**, *40* (4), 277–283.

(54) Tornimbene, B.; Eremin, S.; Escher, M.; Griskeviciene, J.; Manglani, S.; Pessoa-Silva, C. L. *WHO Global Antimicrobial Resistance Surveillance System Early Implementation 2016-17*; 2018.

(55) Chu, C. K.; Bardos, T. J. Synthesis and Inhibition Analysis of 2(4)-Imino-4(2)-Amino-2,4-Dideoxyriboflavin, a Dual Antagonist of Riboflavin and Folinic Acid. *J. Med. Chem.* **1977**, *20* (2), 312–314.

(56) Karoli, T.; Mamidyala, S. K.; Zuegg, J.; Fry, S. R.; Tee, E. H. L.; Bradford, T. A.; Madala, P. K.; Huang, J. X.; Ramu, S.; Butler, M. S.; et al. Structure Aided Design of Chimeric Antibiotics. *Bioorganic Med. Chem. Lett.* **2012**, *22* (7), 2428–2433.

(57) Pokrovskaya, V.; Baasov, T. Dual-Acting Hybrid Antibiotics: A Promising Strategy to Combat Bacterial Resistance. *Expert Opin. Drug Discov.* **2010**, *5* (9), 883–902.

(58) Shapiro, S. Speculative Strategies for New Antibacterials: All Roads Should Not Lead to Rome. *J. Antibiot. (Tokyo).* **2013**, *66* (7), 371–386.

(59) Shiozawa, H.; Kagasaki, T.; Kinoshita, T.; Haruyama, H.; Domon, H.; Utsui, Y.; Kodama, K.; Takahashi, S. Thiomarinol, a New Hybrid Antimicrobial Antibiotic Produced by a Marine Bacterium. Fermentation, Isolation, Structure, and Antimicrobial Activity. *J. Antibiot. (Tokyo).* **1993**, *46* (12), 1834–1842.

(60) Hughes, J.; Mellows, G. Inhibition of Isoleucyl-Transfer Ribonucleic Acid Synthetase in Escherichia Coli by Pseudomonic Acid. *Biochem. J.* **1978**, *176* (1), 305–318.

(61) Li, B.; Wever, W. J.; Walsh, C. T.; Bowers, A. A. Dithiolopyrrolones: Biosynthesis,

Synthesis, and Activity of a Unique Class of Disulfide-Containing Antibiotics. *Nat. Prod. Rep.* **2014**, *31* (7), 905–923.

(62) Oliva, B.; O'Neill, A.; Wilson, J. M.; O'Hanlon, P. J.; Chopra, I. Antimicrobial Properties and Mode of Action of the Pyrrothine Holomycin Antimicrobial Properties and Mode of Action of the Pyrrothine Holomycin. *Antimicrob. Agents Chemother.* **2001**, *45* (2), 532–539.

(63) Chan, A. N.; Shiver, A. L.; Wever, W. J.; Razvi, S. Z. A.; Traxler, M. F.; Li, B. Role for Dithiolopyrrolones in Disrupting Bacterial Metal Homeostasis. *Proc. Natl. Acad. Sci.* **2017**, *114* (10), 2717–2722.

(64) Murphy, A. C.; Fukuda, D.; Song, Z.; Hothersall, J.; Cox, R. J.; Willis, C. L.; Thomas, C. M.; Simpson, T. J. Engineered Thiomarinol Antibiotics Active against MRSA Are Generated by Mutagenesis and Mutasynthesis of Pseudoalteromonas SANK73390. *Angew. Chemie - Int. Ed.* **2011**, *50* (14), 3271–3274.

(65) Freel Meyers, C. L.; Oberthür, M.; Heide, L.; Kahne, D.; Walsh, C. T. Assembly of Dimeric Variants of Coumermycins by Tandem Action of the Four Biosynthetic Enzymes CouL, CouM, CouP, and NovN. *Biochemistry* **2004**, *43* (47), 15022–15036.

(66) Schmutz, E.; Steffensky, M.; Schmidt, J.; Porzel, A.; Li, S. M.; Heide, L. An Unusual Amide Synthetase (CouL) from the Coumermycin A1 Biosynthetic Gene Cluster from Streptomyces Rishiriensis DSM 40489. *Eur. J. Biochem.* **2003**, *270* (22), 4413–4419.

(67) Steffensky, M.; Li, S. M.; Heide, L. Cloning, Overexpression, and Purification of Novobiocic Acid Synthetase from Streptomyces Spheroides NCIMB 11891. *J. Biol. Chem.* **2000**, *275* (28), 21754–21760.

(68) Li, B.; Walsh, C. T. Identification of the Gene Cluster for the Dithiolopyrrolone Antibiotic Holomycin in Streptomyces Clavuligerus. *Proc. Natl. Acad. Sci.* **2010**, *107* (46), 19731–19735.

(69) Murphy, A. C.; Gao, S.-S.; Han, L.-C.; Carobene, S.; Fukuda, D.; Song, Z.; Hothersall, J.; Cox, R. J.; Crosby, J.; Crump, M. P.; et al. Biosynthesis of Thiomarinol A and Related Metabolites of Pseudoalteromonas Sp. SANK 73390. *Chem. Sci.* **2014**, *5* (1), 397–402.

(70) Fortin, P. D.; Walsh, C. T.; Magarvey, N. A. A Transglutaminase Homologue as a Condensation Catalyst in Antibiotic Assembly Lines. *Nature* **2007**, *448* (7155), 824–827.

(71) Mihara, K.; Tanabe, T.; Yamakawa, Y.; Funahashi, T.; Nakao, H.; Narimatsu, S.; Yamamoto, S. Identification and Transcriptional Organization of a Gene Cluster Involved in Biosynthesis and Transport of Acinetobactin, a Siderophore Produced by Acinetobacter Baumannii ATCC 19606T. *Microbiology* **2004**, *150* (8), 2587–2597.

(72) Murcia, N. R.; Lee, X.; Waridel, P.; Maspoli, A.; Imker, H. J.; Chai, T.; Walsh, C. T.;

Reimmann, C. The Pseudomonas Aeruginosa Antimetabolite L -2-Amino-4-Methoxy-Trans-3-Butenoic Acid (AMB) Is Made from Glutamate and Two Alanine Residues via a Thiotemplate-Linked Tripeptide Precursor. *Front. Microbiol.* **2015**, *6* (MAR), 1–13.

(73) Gao, S. S.; Hothersall, J.; Wu, J.; Murphy, A. C.; Song, Z.; Stephens, E. R.; Thomas, C. M.; Crump, M. P.; Cox, R. J.; Simpson, T. J.; et al. Biosynthesis of Mupirocin by Pseudomonas Fluorescens NCIMB 10586 Involves Parallel Pathways. *J. Am. Chem. Soc.* **2014**, *136* (14), 5501–5507.

(74) Brieke, C.; Peschke, M.; Haslinger, K.; Cryle, M. J. Sequential in Vitro Cyclization by Cytochrome P450 Enzymes of Glycopeptide Antibiotic Precursors Bearing the X-Domain from Nonribosomal Peptide Biosynthesis. *Angew. Chemie - Int. Ed.* **2015**, *54* (52), 15715–15719.

(75) Haslinger, K.; Peschke, M.; Brieke, C.; Maximowitsch, E.; Cryle, M. J. X-Domain of Peptide Synthetases Recruits Oxygenases Crucial for Glycopeptide Biosynthesis. *Nature* **2015**, *521* (7550), 105–109.

(76) Kokona, B.; Winesett, E. S.; Nikolai Von Krusenstiern, A.; Cryle, M. J.; Fairman, R.; Charkoudian, L. K. Probing the Selectivity of β-Hydroxylation Reactions in Non-Ribosomal Peptide Synthesis Using Analytical Ultracentrifugation. *Anal. Biochem.* **2016**, *495*, 42–51.

(77) Brieke, C.; Kratzig, V.; Haslinger, K.; Winkler, A.; Cryle, M. J. Rapid Access to Glycopeptide Antibiotic Precursor Peptides Coupled with Cytochrome P450-Mediated Catalysis: Towards a Biomimetic Synthesis of Glycopeptide Antibiotics. *Org. Biomol. Chem.* **2015**, *13* (7), 2012–2021.

(78) Agarwal, V.; Diethelm, S.; Ray, L.; Garg, N.; Awakawa, T.; Dorrestein, P. C.; Moore, B. S. Chemoenzymatic Synthesis of Acyl Coenzyme A Substrates Enables in Situ Labeling of Small Molecules and Proteins. *Org. Lett.* **2015**, *17* (18), 4452–4455.

(79) Beld, J.; Cang, H.; Burkart, M. D. Visualizing the Chain-Flipping Mechanism in Fatty-Acid Biosynthesis. *Angew. Chemie - Int. Ed.* **2014**, *53* (52), 14456–14461.

(80) Qin, Z.; Huang, S.; Yu, Y.; Deng, H. Dithiolopyrrolone Natural Products: Isolation, Synthesis and Biosynthesis. *Mar. Drugs* **2013**, *11* (10), 3970–3997.

(81) Berkowitz, D. B.; Charette, B. D.; Karukurichi, K. R.; McFadden, J. M. α-Vinylic Amino Acids: Occurrence, Asymmetric Synthesis, and Biochemical Mechanisms. *Tetrahedron Asymmetry* **2006**, *17* (6), 869–882.

(82) Meluzzi, D.; Zheng, W. H.; Hensler, M.; Nizet, V.; Dorrestein, P. C. Top-down Mass Spectrometry on Low-Resolution Instruments: Characterization of Phosphopantetheinylated Carrier Domains in Polyketide and Non-Ribosomal Biosynthetic Pathways. *Bioorganic Med. Chem. Lett.* **2008**, *18* (10), 3107–3111.

(83) Moffitt, M. C.; Neilan, B. A. Characterization of the Nodularin Synthetase Gene Cluster and Proposed Theory of the Evolution of Cyanobacterial Hepatotoxins. **2004**, *70* (11), 6353–6362.

(84) Tillett, D.; Dittmann, E.; Erhard, M.; Do, H. Von; Neilan, B. A. Structural Organization of Microcystin Biosynthesis in Microcystis Aeruginosa PCC7806 : An Integrated Peptide ^ Polyketide Synthetase System. 753–764.

(85) Banowetz, G. M.; Azevedo, M. D.; Armstrong, D. J.; Halgren, A. B.; Mills, D. I. Germination-Arrest Factor (GAF): Biological Properties of a Novel, Naturally-Occurring Herbicide Produced by Selected Isolates of Rhizosphere Bacteria. *Biol. Control* **2008**, *46* (3), 380–390.

(86) Banowetz, G. M.; Azevedo, M. D.; Armstrong, D. J.; Mills, D. I. Germination Arrest Factor (GAF): Part 2. Physical and Chemical Properties of a Novel, Naturally Occurring Herbicide Produced by Pseudomonas Fluorescens Strain WH6. *Biol. Control* **2009**, *50* (2), 103–110.

(87) McPhail, K. L.; Armstrong, D. J.; Azevedo, M. D.; Banowetz, G. M.; Mills, D. I. 4-Formylaminooxyvinylglycine, an Herbicidal Germination-Arrest Factor from Pseudomonas Rhizosphere Bacteria. *J. Nat. Prod.* **2010**, *73* (11), 1853–1857.

(88) Okrent, R. A.; Halgren, A. B.; Azevedo, M. D.; Chang, J. H.; Chang, J. H.; Mills, D. I.; Maselko, M.; Armstrong, D. J.; Banowetz, G. M.; Trippe, K. M. Negative Regulation of Germination-Arrest Factor Production in Pseudomonas Fluorescens WH6 by a Putative Extracytoplasmic Function Sigma Factor. *Microbiol. (United Kingdom)* **2014**, *160* (2014), 2432–2442.

(89) Okrent, R. A.; Trippe, K. M.; Maselko, M.; Manning, V. Functional Analysis of a Biosynthetic Cluster Essential for Production of 4-Formylaminooxyvinylglycine, a Germination-Arrest Factor from Pseudomonas Fluorescens WH6. *Microbiol. (United Kingdom)* **2017**, *163* (2), 207–217.

(90) Haruhisa, H.; Azusa, K.; Seigo, N.; Takeshi, A.; Yoshitami, O.; Masayoshi, G. ANTIBIOTICS WAP-5044C AND WAP-5044A, DERIVATIVE FROM WAP-5044 C TREATED WITH ALGINASE, THEIR PRODUCTION AND USE. 05-105657, 1993.

(91) Rosenthal, G. A. Investigations of Canavanine Biochemistry in the Jack Bean Plant, Canavalia Ensiformis (L.) DC: II. Canavanine Biosynthesis in the Developing Plant. *Plant Physiol.* **1972**, *50* (3), 328–331.

(92) Leguminosae, D. C.; Rosenthal, G. A. L-Canavanine Metabolism. *Plant Physiol.* **1982**, *1* (69), 1066–1069.

(93) Yasuta, T.; Okazaki, S.; Mitsui, H.; Yuhashi, K. I.; Ezura, H.; Minamisawa, K. DNA Sequence and Mutational Analysis of Rhizobitoxine Biosynthesis Genes in *Bradyrhizobium Elkanii*. *Appl. Environ. Microbiol.* **2001**, *67* (3–12), 4999–5009.

(94) Patteson, A. J. B.; Dunn, Z. D.; Li, B. In Vitro Biosyntehsis of the Nonproteinogenic Amino Acid Methoxyvinylglycine. *Angew. Chemie - Int. Ed.* **2018**.

(95) Fernández, M.; Cuadrado, Y.; Aparicio, J. F.; Martín, J. F. Role of Homoserine and Threonine Pathway Intermediates as Precursors for the Biosynthesis of Aminoethoxyvinylglycine in Streptomyces Sp. NRRL 5331. *Microbiology* **2004**, *150* (5), 1467–1474.

(96) Lowther, J.; Yard, B. A.; Johnson, K. A.; Carter, L. G.; Bhat, V. T.; Raman, M. C. C.; Clarke, D. J.; Ramakers, B.; McMahon, S. A.; Naismith, J. H.; et al. Inhibition of the PLP-Dependent Enzyme Serine Palmitoyltransferase by Cycloserine: Evidence for a Novel Decarboxylative Mechanism of Inactivation. *Mol. Biosyst.* **2010**, *6* (9), 1682–1693.

(97) Svensson, M. L.; Valeria, K.; Gatenbeck, S. Hydroxyurea, a Natural Metabolite and an Intermediate in d-Cycloserine Biosynthesis in Streptomyces Garyphalus. *Arch. Microbiol.* **1981**, *129* (3), 210–212.

(98) Kumagai, T.; Koyama, Y.; Oda, K.; Noda, M.; Matoba, Y.; Sugiyama, M. Molecular Cloning and Heterologous Expression of a Biosynthetic Gene Cluster for the Antitubercular Agent D-Cycloserine Produced by Streptomyces Lavendulae. *Antimicrob. Agents Chemother.* **2010**, *54* (3), 1132–1139.

(99) Kumagai, T.; Takagi, K.; Koyama, Y.; Matoba, Y.; Oda, K.; Noda, M.; Sugiyama, M. Heme Protein and Hydroxyarginase Necessary for Biosynthesis of D-Cycloserine. *Antimicrob. Agents Chemother.* **2012**, *56* (7), 3682–3689.

(100) Stols, L.; Gu, M.; Dieckman, L.; Raffen, R.; Collart, F. R.; Donnelly, M. I. A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site. *Protein Expr. Purif.* **2002**, *25* (1), 8–15.

(101) Delaney, S. M.; Mavrodi, D. V.; Bonsall, R. F.; Thomashow, L. S. PhzO, a Gene for Biosynthesis of 2-Hydroxylated Phenazine Compounds in Pseudomonas Aureofaciens 30-84. *J. Bacteriol.* **2001**, *183* (1), 318–327.

(102) Maddula, V. S. R. K.; Pierson, E. A.; Pierson, L. S. Altering the Ratio of Phenazines in Pseudomonas Chlororaphis (Aureofaciens) Strain 30-84: Effects on Biofilm Formation and Pathogen Inhibition. *J. Bacteriol.* **2008**, *190* (8), 2759–2766.

(103) Du, Y. L.; He, H. Y.; Higgins, M. A.; Ryan, K. S. A Heme-Dependent Enzyme Forms the Nitrogen-Nitrogen Bond in Piperazate. *Nat. Chem. Biol.* **2017**, *13* (8), 836–838.