ABSTRACT

Title of Document: DIHYDROPYRROLE FORMATION DURING SIBIROMYCIN BIOSYNTHESIS

Shalini Saha, Doctor of Philosophy
2015

Directed By: Professor Steven E. Rokita, Department of Chemistry and Biochemistry

A description of pyrrolo[1,4]benzodiazepine (PBD) biosynthesis in actinomycetes is a prerequisite for engineering production of analogs with enhanced antitumor activity. Several proteins expected to synthesize the PBD's dihydropyrrole moiety were heterologously expressed, purified and assayed for activity. UV-visible spectroscopy revealed that predicted dioxygenases SibV and homolog Orf12 associated with PBDs sibiromycin and anthramycin, respectively, catalyze the regiospecific 2,3-extradiol dioxygenation of L-3,4-dihydroxyphenylalanine (L-DOPA) to form L-2,3-*seco*dopa ($\lambda_{max}$ = 368 nm). $^{1}$H NMR spectroscopy indicated that L-2,3-*seco*dopa then spontaneously cyclizes into the α-keto acid tautomer of 4-(2-oxo-3-butenoic-acid)-4,5-dehydro-L-proline **1.1** ($\lambda_{max}$ = 414 nm). Thus, the dioxygenases establish the scaffold of the dihydropyrrole moiety. Both the quaternary structure and product formed by dioxygenases are conserved in

dihydropyrrole biosynthesis within both PBD and non-PBD pathways. Stability studies suggest that **1.1** is relatively labile and is likely consumed rapidly by subsequent biosynthetic steps.

Hydrolysis and methylation steps were proposed to modify the dihydropyrrole scaffold in PBDs. The predicted proteins SibS and homolog TomK associated with sibiromycin and PBD tomaymycin biosynthesis, respectively, were assayed for hydrolysis activity. The predicted protein SibZ associated with sibiromycin biosynthesis was assayed for methyltransferase activity. The proposed SibZ substrate 4-vinyl-4,5-dehydro-L-proline **1.2** was synthesized. For these three proteins, no catalytic activity was observed with their proposed substrates or substrate precursors under a range of conditions. However, SibS binds **1.1** ($K_D$ = 64 ± 2 μM) suggesting it participates in dihydropyrrole biosynthesis. HPLC-MS indicated that SibS catalyzes the depurination of *S*-adenosylmethionine although it is unlikely this reaction is involved in sibiromycin biosynthesis. These findings suggest that the pathway requires revision.

The adenylation and thiolation didomain of the predicted non-ribosomal peptide synthetase SibD associated with sibiromycin biosynthesis was expressed and purified. To test if SibD incorporates the dihydropyrrole moiety into sibiromycin, its proposed substrate 4-propenyl-4,5-dehydro-L-proline **1.5** was synthesized. A radioactivity exchange assay and peptide analysis by MS revealed that SibD does not adenylate and thiolate **1.5** or its precursors L-DOPA, **1.1** or **1.2**. However, L-threonine and the metabolic precursor L-tyrosine are substrates for these reactions. SibB promotes adenylation catalyzed by SibD and represents one of two proteins, distinct from MbtH-like proteins, capable of promoting adenylation.

# DIHYDROPYRROLE FORMATION DURING SIBIROMYCIN BIOSYNTHESIS

by

Shalini Saha

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:

Professor Steven. E. Rokita, Chair
Professor Dorothy Beckett
Professor Douglas A. Julin
Professor Jason D. Kahn
Professor Daniel Stein

# Dedication

To my father, mother and sister, I am grateful for your love and support.

# Acknowledgements

This project would not have not have been possible without assistance from many people. I am grateful for the guidance my advisor, Dr. Steve Rokita, provided me throughout my graduate career. He was the first to teach me about NRPSs in his class although at the time I did not realize how crucial they would be in my research. He gave me an opportunity explore the biosynthesis of natural products. My research was filled with challenges but his enthusiasm and feedback always made me eager to figure out the answers. He went above and beyond reaching out to other researchers that could help me and meticulously reviewing the many drafts of my papers, reports and presentations. For that and so much more, I thank him.

I thank Dr. Barbara Gerratana for introducing me to pyrrolo[1,4]benzodiazepines, a set of compounds I found fascinating to study. She taught me enzymology and laid a lot of the foundation for this project. I appreciate her continued advice on the project throughout the years. I also thank Drs. Lee Chuenchor, Isaac Yonemoto and especially Katie Connor for teaching me fundamental lab techniques. You were great role models.

I thank my committee Professors Beckett, Julin, Kahn and Stein for letting me discuss my research with them for several hours. They asked thoughtful questions and gave me helpful edits on this dissertation. I thank Professors DeShong, Sintim and Davis for teaching me physical organic chemistry and synthetic methodology in their classes.

I thank Dr. Chris Whitman, Dr. Craig Townsend and their research groups for providing me plasmids, compounds and help discussions. I thank Dr. Gary Posner for helping me develop the synthetic methodologies used in this work. I thank Drs. Yui-Fai Lam and Gene Mazzola for their expertise in NMR. I thank Dr. Phil Mortimer for his help

with the mass spectrometry work. I thank the rest of the support staff at both UMD and JHU.

Thank you to the past and present Rokita lab members for welcoming me into the group. Although we did not share projects, I always felt part of the group. It was a pleasure both learning from you and teaching you. I thank Blessing Deeyaa for inviting me to NOBCChE meetings, reminding me there is life outside of the lab. I thank Mark Hutchinson helping me with synthesis and taking on so many lab responsibilities. I thank Patt Ingavat for answering my constant questions about crystallography and feeding me dessert on the long days. I thank Abhishek Phatarphekar for teaching me so much about fly genetics and always being available to discuss anything. I thank Petrina Boucher for grabbing coffee with me while we navigated through the final years. I thank Chris Su and Zuodong Sun for being so enthusiastic about their research making me enthusiastic about mine.

I thank my friends and family for providing all the emotional support I needed to get through this. I thank my parents and my sister helping me throughout the years and feeding me along the way. And Jesse, I cannot thank you enough.

# Table of Contents

# List of Tables

## Chapter 1



## Chapter 2

None

## Chapter 3

None

## Chapter 4

None

## Chapter 5

None

## Chapter 6

None

# List of Figures

## Chapter 1



## Chapter 2



## Chapter 3



## Chapter 4

Chapter 5



Chapter 6

None

# List of Schemes

## Chapter 1



## Chapter 2



## Chapter 3

Chapter 4



Chapter 5



Chapter 6

# Abbreviations

aa – amino acid

ACP – acyl carrier protein

ADC – analog-to-digital converter

antiSMASH – Antibiotics & Secondary Metabolite Analysis Shell

ATP – adenosine triphosphate

BCA – bicinchoninic acid

BLAST – basic local alignment search tool

BOC – *tert*-butyloxycarbonyl

$BOC_2O$ – BOC anhydride

bp – base pair

bs – broad singlet

CBD – chitin binding domain

COSY – $^1H$-$^1H$ homonuclear correlation spectroscopy

CV – column volume

DMF – dimethylforamide

DMPU – *N*,*N*'-dimethylpropylene urea

DMSO – dimethyl sulfoxide

DMSO-$d_6$ – deuterated dimethyl sulfoxide

DOPA – 3,4-dihydroxyphenylalanine

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

$ESI^+$ – electrospray ionization under positive-ion mode

F – fluorescence emission intensity

$F_b$ – fluorescence emission intensity with saturating level of ligand

FAB – fast atom bombardment

FAD – flavin adenine dinucleotide (oxidized quinone form)

$FADH_2$ – FAD (reduced hydroquinone form)

$F_0$ – fluorescence emission intensity without ligand

F420-$Glu_N$ – 7,8-didemethyl-8-hydroxy-5-deazariboflavin-5'phosphate conjugated to glutamyl residues

$GI_{50}$ – concentration resulting in 50% cell growth corrected for cell count at time zero

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

$His_6$-SUMO – hexahistidine-small ubiquitin-like modifier

HMBC – heteronuclear multiple bond correlation

HP – high performance

HPLC – high performance liquid chromatography

HR – high resolution

HSQC – heteronuclear single quantum correlation

$IC_{50}$ – concentration that causes 50% growth inhibition

IPRP – ion-pair reverse phase

IPTG – isopropyl β-D-1-thiogalactopyranoside

$K_{av}$ – partition coefficient

kDa – kiloDalton

$K_D$ – dissociation constant

LB – lysogeny broth (Miller)

LC – liquid chromatography

LiHMDS – lithium bis(trimethylsilyl)amide

MLP – MbtH-like protein

$M_R$ – log of the molecular weight

MUSCLE – multiple sequence comparison by log-expectation

MWCO – molecular weight cutoff

NaPPi – tetrasodium pyrophosphate

NCBI – National Center for Biotechnology Information

NCI – National Cancer Institute

NMP – *N*-methyl-2-pyrrolidone

NMWL – nominal molecular weight limit

NOESY – nuclear overhauser effect spectroscopy

NRPS – non-ribosomal peptide synthetases

Ni-NTA – nickel-nitrolotriacetic acid

OD – optical density

PBD – pyrrolo[1,4]benzodiazepine

PCR – polymerase chain reaction

$Pd(dba)_2$ – palladium(0)bis(dibenzylideneacetone)

PDA – photodiode array

Pfam – protein family

PhNH(OTf) – *N*-phenyl-trifluoromethanesulfonamine

$PhN(OTf)_2$ – *N*-phenyl-bis(trifluoromethanesulfonimide)

PPi – inorganic pyrophosphate

PYBOP – benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate

$R_f$ – retention factor

RP – reverse phase

RT – room temperature

SAH - *S*-adenosyl homocysteine

SAM – *S*-adenosylmethionine

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

SNAC – *N*-acetylcysteamine

TCEP – tris(2-carboxyethyl)phosphine

TFA – trifluoroacetic acid

TGI – total growth inhibition

THF – tetrahydrofuran

TLC – thin layer chromatography

Tof – time of flight

$t_r$ – retention time

UMA – Udwary-Merski algorithm

UPLC – ultra performance liquid chromatography

$V_c$ – column volume

$V_e$ – elution volume

$V_0$ – void volume

v/v – volume by volume

w/v – weight by volume

w/w – weight by weight

$\Delta T_M$ – change in denaturation midpoint melting temperature

$\lambda_{max}$ – wavelength of maximum absorbance

$\lambda_{em}$ – emission wavelength

$\lambda_{ex}$ – excitation wavelength

$\bar{Y}$ – fractional saturation

$[L]_T$ – total ligand concentration

$[P]_T$ – total protein concentration

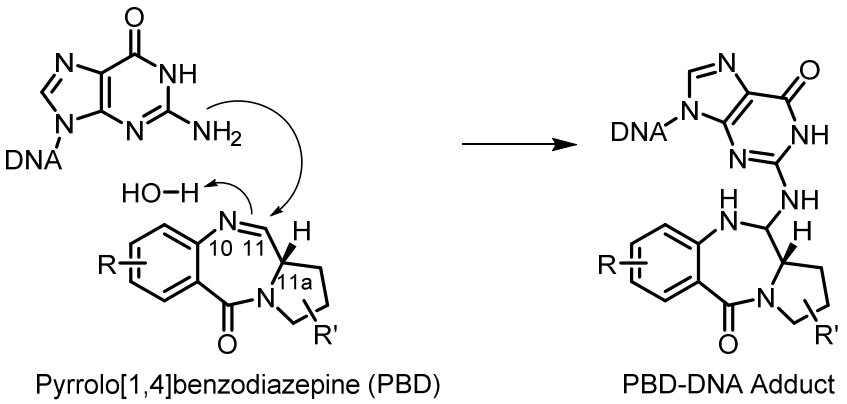
# Chapter 1: Introduction

## 1.1 Pyrrolo[1,4]benzodiazepines are of pharmacological interest

Natural products and their derivatives account for almost half of the small molecule anticancer drugs approved between 1940 and 2012,[1] motivating continued research in the field. Pyrrolo[1,4]benzodiazepines (PBDs) represents a class of interest as many of its members display antitumor activity. PBDs contain a 1,4-benzodiazepine ring also present in other classes of natural products that were developed into over 50 clinically approved psychoactive drugs such as diazepam (Valium®, Hoffman-La Roche, New Jersey). The presence of this pharmacophore makes PBDs a promising class of pharmaceuticals.[2]

PBDs are tricyclic secondary metabolites, the majority of which have emerged as anticancer drug candidates by virtue of their ability to alkylate double stranded DNA.[2-4] PBDs contain an anthranilate and dihydropyrrole or pyrrolidine moiety fused to each side of a diazepine ring. The diazepine ring bears an electrophilic imine carbon C11, the site of alkylation.[5-8] Alkylation occurs primarily at guanosine (G) (**Scheme 1-1**) in a sequence-specific manner.[9] 5'-Pyrimidine-G-pyrimidine-3' is the kinetic target and 5'-purine-G-purine-3' is the thermodynamic target.[10] The resulting DNA-PBD adduct inhibits DNA transcription and replication. Their efficacy is due in part to their ability to bind DNA in a sequence selective manner within the minor grove[11,12] and the resistance of their subsequent PBD-DNA adducts to proofreading machinery that repairs DNA.[13]

Actinomycetal genera *Micrococci* and *Streptomyces* produce a variety of PBDs **(Scheme 1-2)** with a range of antitumor or other bioactivity. For instance, anthramycin showed activity across 60 cell lines of an NCI screen (average concentration that inhibits

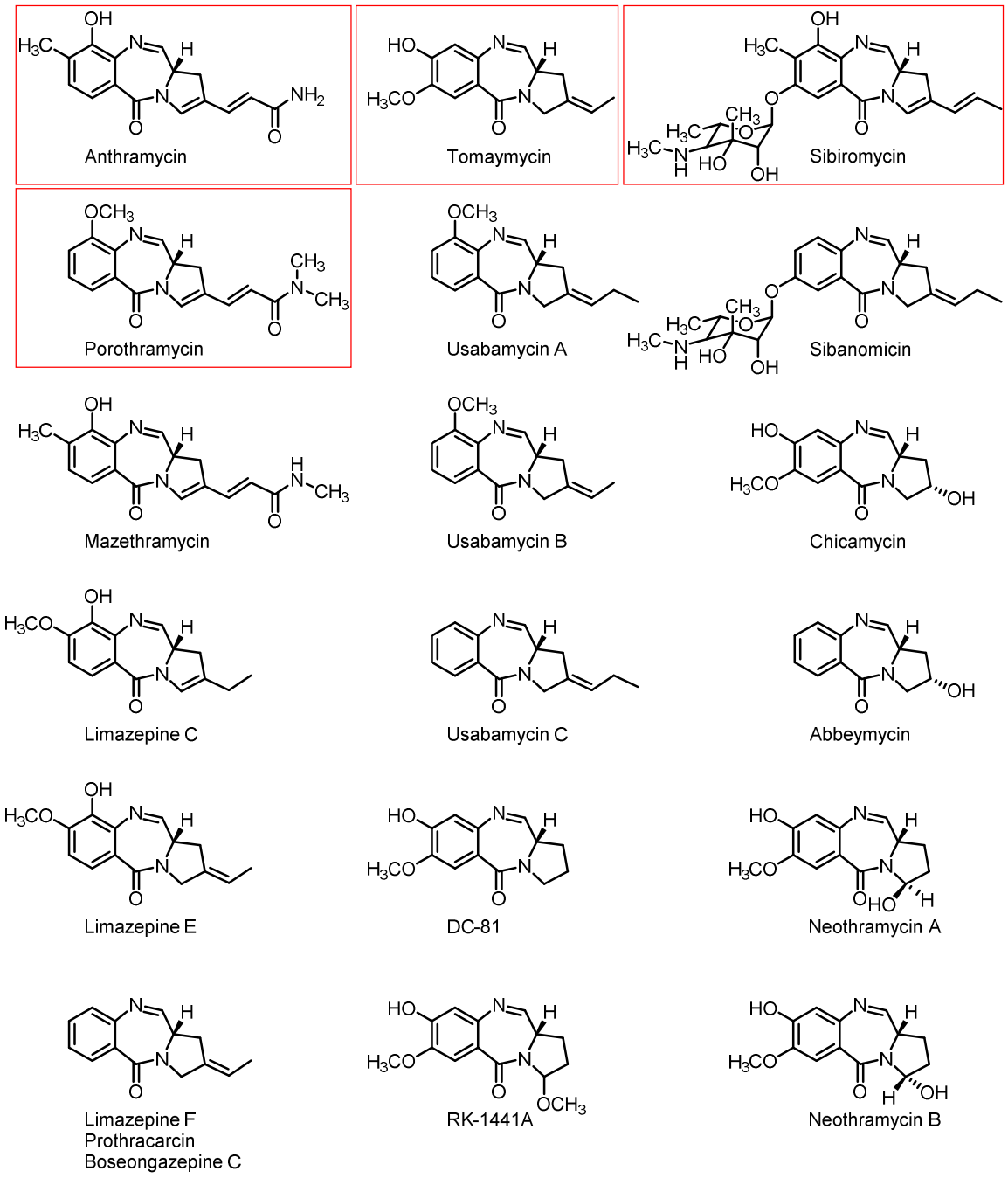



**Scheme 1-1.** Antitumor activity of PBDs results from the formation of PBD-DNA adducts via the proposed mechanism. R and R' represent a range of substituents.

50% cell growth ($GI_{50}$) = 0.029 μM and average concentration that inhibits 100% cell growth (TGI) = 0.61 μM).[14] In contrast, neothramycin specifically targets bladder cancer cells when used as a topical treatment eradicating tumors in 36% of patients and reducing tumor size by greater than 50% in 55% of patients in one trial.[15]

**1.2 Sibiromycin is the most potent naturally occurring PBD**

Sibiromycin displays the highest affinity for DNA and greatest cytotoxicity of all PBDs due to its unique presence of an appended amino sugar.[16] Thermal denaturation assays of calf thymus DNA-PBD complexes compared with DNA alone show sibiromycin is most effective at increasing the denaturation midpoint melting temperature of DNA ($\Delta T_M$) (**Table 1-1**).[11,17] The $\Delta T_M$ of a DNA-PBD complex compared with DNA is an indicator of DNA affinity of PBDs. Binding affinity was shown to correlate with the PBD's ability to inhibit transcription *in vitro*.[12] Sibiromycin is most cytotoxic to cells associated with leukemia and plasmacytoma (**Table 1-2**) and moderately cytotoxic to ovarian cancer cells compared with other PBDs.[17] Molecular dynamics and docking studies indicate that its sugar protrudes from the minor groove and likely blocks transcription factors from binding to their targets, thereby inhibiting transcription.[16] However, the pharmacological




**Scheme 1-2.** Naturally occurring PBDs with antibiotic and/or antitumor activity. Biosynthetic gene clusters have been identified for the boxed PBDs.

utility of this natural product is limited by its cardiotoxicity that has been traced to the hydroxyl group at C9 of the anthranilate moiety.[18,19]

**Table 1-1.** Stabilization of calf-thymus DNA by PBDs as assessed by the change in denaturation midpoint melting temperature ($\Delta T_M$) of DNA-PBD complex compared with DNA alone.[17]

| | $\Delta T_M$ (°C)[a] | | |
|---|---|---|---|
| **PBD** | **0 h** | **4 h** | **18 h** |
| Sibiromycin | 15.7 | 15.9 | 16.3 |
| Anthramycin | 9.4 | 11.2 | 13.0 |
| Tomaymycin | 1.0 | 2.4 | 2.6 |

[a] DNA-PBD at a 5:1 molar ratio was incubated at various times at 37 °C and pH 7.0.

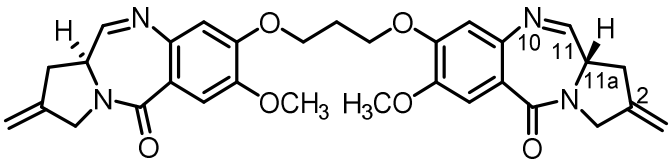
**Table 1-2.** Cytotoxicity of PBDs against various cancer cells as assessed by the half maximal inhibitory concentration ($IC_{50}$).[17]

| | $IC_{50}$ | | |
|---|---|---|---|
| **PBD** | **L1210 cells (leukemia)** | **ADJ/PC6 cells (plasmacytoma)** | **CH1 cells (ovarian)** |
| Sibiromycin | 2.9 nM | 0.017 nM | 40 nM |
| Anthramycin | 22 nM | 2.8 nM | 320 nM |
| Tomaymycin | 3.7 nM | 1.8 nM | 0.13 nM |

### 1.3 Efforts to develop PBD analogs into chemotherapeutic drugs

To circumvent the harmful side effect of sibiromycin, monomeric and dimeric PBD analogs were chemically synthesized.[20,21] SG2000 (formerly SJG-136) (**Figure 1-1**), the most potent synthetic PBD dimer is currently in Phase II trials for treating leukemia (AstraZeneca (Spirogen), London, UK). However, development of other synthetic analogs has been impeded by a number of issues. Installing the *C2-endo-exo* unsaturation is laborious and lengthy, particularly because a ring closure step that furnishes the diazepine ring through reductive cyclization removes the unsaturation in the dihydropyrrole. The unsaturation is necessary to flatten the ring allowing it to fit into the minor groove.[17] A fully unsaturated and saturated hydropyrrole ring or bulky substituents on the hydropyrrole ring inhibit PBD analogs from binding DNA.[22] Synthetic methodologies are also not easily compatible with modifying the substituents on the anthranilate ring without hindering the diazepine ring formation and perturbing two of its crucial structural attributes, the *S*-stereochemistry at C11a and the N10-C11 imine bond. Replacing the anthranilate ring with

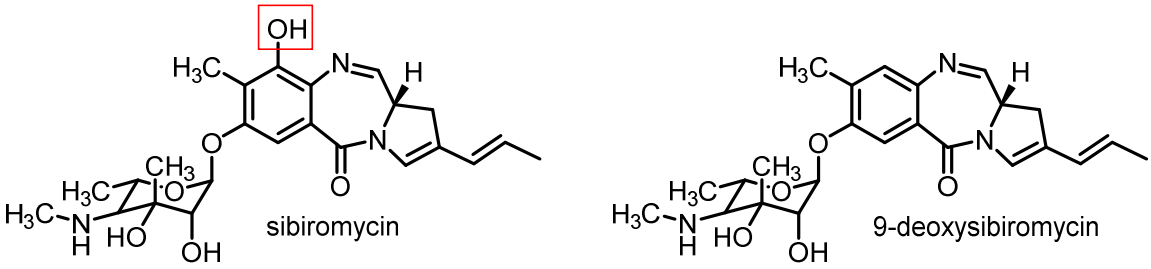
other heterocycles such as pyridine, pyrazine, pyrimidine or pyrazole decreases binding affinity and potency of the resulting analog[17] prompting interest in pursuing other avenues of PBD analog production.




SG2000

**Figure 1-1.** SG2000, a synthetic PBD dimer.

Biosynthesis offers a complementary and appealing strategy to generate new PBD candidates. For example, delineating the biosynthesis of the anthranilate moiety in PBDs[23,24] allowed for its reprogramming to create an analog of sibiromycin lacking its C9 hydroxyl group.[25] The resulting 9-deoxysibiromcyin **(Scheme 1-3)** minimizes cardiotoxicity as desired[25] and confirms the potential for creating a range of new PBDs that suppress unwanted side effects.




**Scheme 1-3.** Natural product sibiromycin and its bioengineered analog 9-deoxysibiromycin. The hydroxyl group inducing the cardiotoxic property of sibiromycin is indicated with a red box.

Creating a dihydropyrrole moiety within a PBD to abolish the abnormal DNA bending caused by its alkylation is an attractive goal since the resulting adduct has the potential to evade DNA repair enzymes that detect alteration in DNA curvature.[13] Some of the distortion caused by PBDs is already ameliorated by their right-handed twist that

complements right-handed duplex DNA.[7,8,26-29] Structure perturbation of the conjugate still persists and includes bending of the helix around the newly formed covalent bond. The degree of DNA bending induced by the adduct (5.0-8.9° with anthramycin and 8.2-14.5° with tomaymycin) inversely correlates with the degree of PBD twist (35.4° in anthramycin and 9.1° in tomaymycin).[26,27] Before natural processes can be harnessed to develop the new dihydropyrroles within PBDs to escape DNA repair, the biosynthetic pathway of these groups must be identified.

### 1.4 Current understanding of dihydropyrrole moiety biosynthesis

Previous feeding studies with radiolabeled amino acids first suggested that the dihydropyrrole moiety originates from L-tyrosine,[30] but few biochemical details of the intervening transformations are available. A comparison of the gene clusters responsible for producing the PBDs anthramycin,[23] sibiromycin **(Figure 1-2)**,[31] tomaymycin[32] and porothramycin[33] allowed a tentative assignment of gene function and revealed a common strategy for generating these PBDs. The biosynthetic gene clusters have not been identified for the remaining PBDs or a pair of pentacyclic PBD derivatives called circumdatins (**Scheme 1-4**). The biosynthetic gene cluster is available for tilivalline (**Scheme 1-4**), a toxic PBD derivative implicated in causing hemorrhagic colitis upon its overproduction by *Klebsiella oxytoca*, a member of the human intestinal microbiota.[34] However, no genes associated with pyrrolidine moiety biosynthesis are present in the gene cluster suggesting this moiety does not form from L-tyrosine. More information on the moiety was gleaned instead from biosynthetic gene clusters of non-PBD natural products lincomycin A[35,36] and hormaomycin[37] as both contain a pyrrolidine ring that also originates from L-tyrosine and supports the tentative gene assignments.

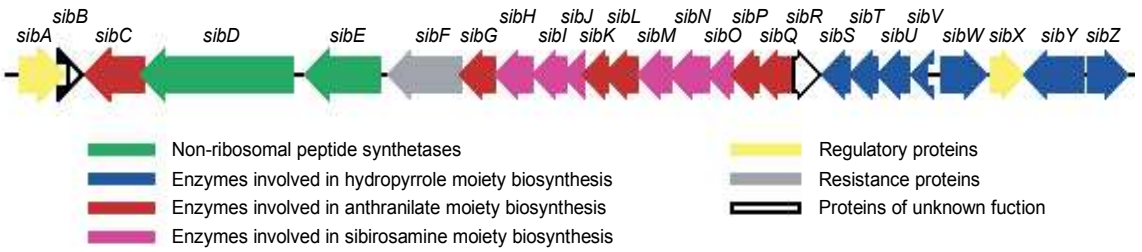
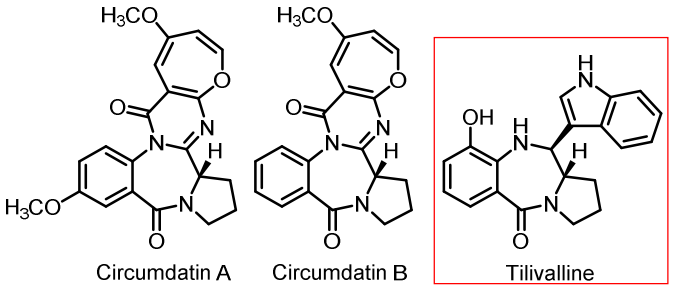



**Figure 1-2.** Genetic organization of the sibiromycin biosynthetic gene cluster in *Streptosporangium sibiricum*. Functional assignment of its gene products was based on the comparative analysis of the PBDs, lincomycin A and hormaomycin biosynthetic gene clusters. This figure was reproduced.[31]

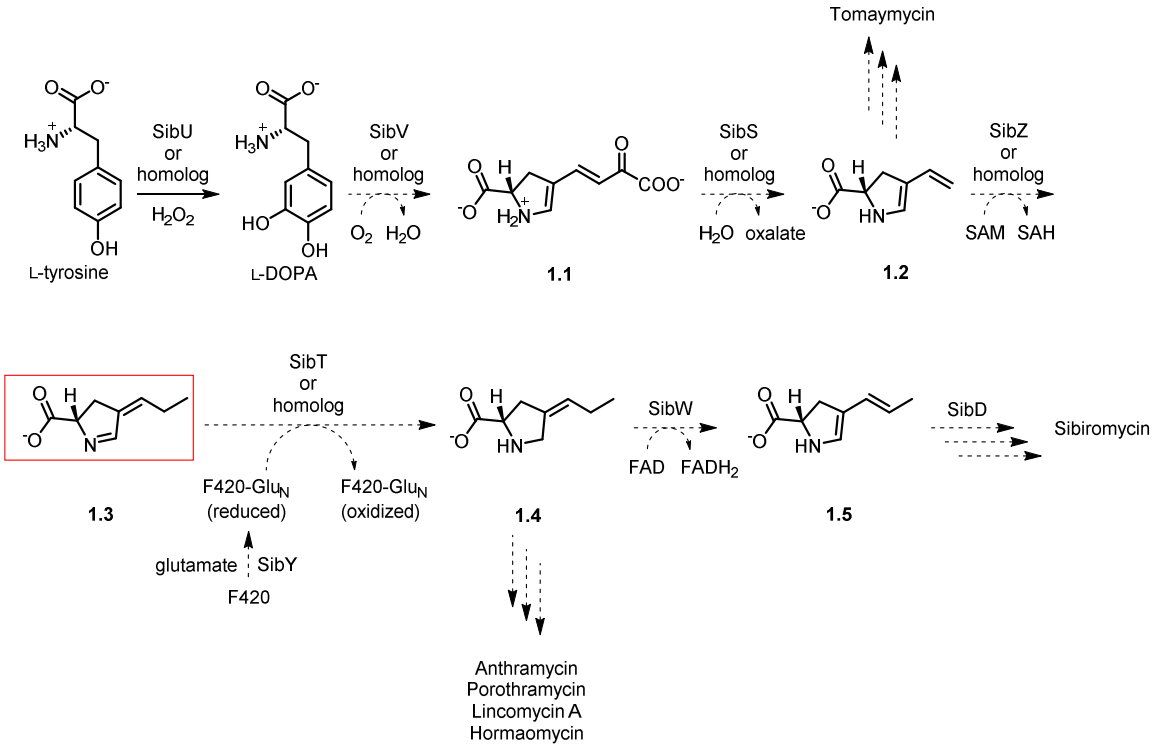



**Scheme 1-4.** Naturally occurring PBDs with no antitumor activity reported. Tilivalline is an enterotoxic PBD produced by a human gut microbe while no known bioactivity has been observed for the pair of circumdatins. The boxed PBD has an identified biosynthetic gene cluster.

*In vitro* activity assays confirmed that this pathway begins with an enzyme-catalyzed *ortho*-hydroxylation of L-tyrosine to form L-3,4-dihydroxyphenylalanine (L-DOPA).[38,39] L-[$^{14}$C]DOPA was partially incorporated into the PBD during feeding studies providing further evidence that L-DOPA is a pathway intermediate.[40] The subsequent steps have been tentatively proposed[31] (**Scheme 1-5**) based on feeding studies (**Scheme 1-6**), the comparison of available gene clusters (**Figure 1-2**) and functional assignment of gene products based on the identification of consensus sequence signatures using BLAST[41] (**Table 1-3**). These steps are expected to be common during the production of sibiromycin, anthramycin, porothramycin, tomaymycin, lincomycin A and hormaomycin as part of their gene clusters codes for similar a set of enzymes. The amino acid sequences of SibV and its

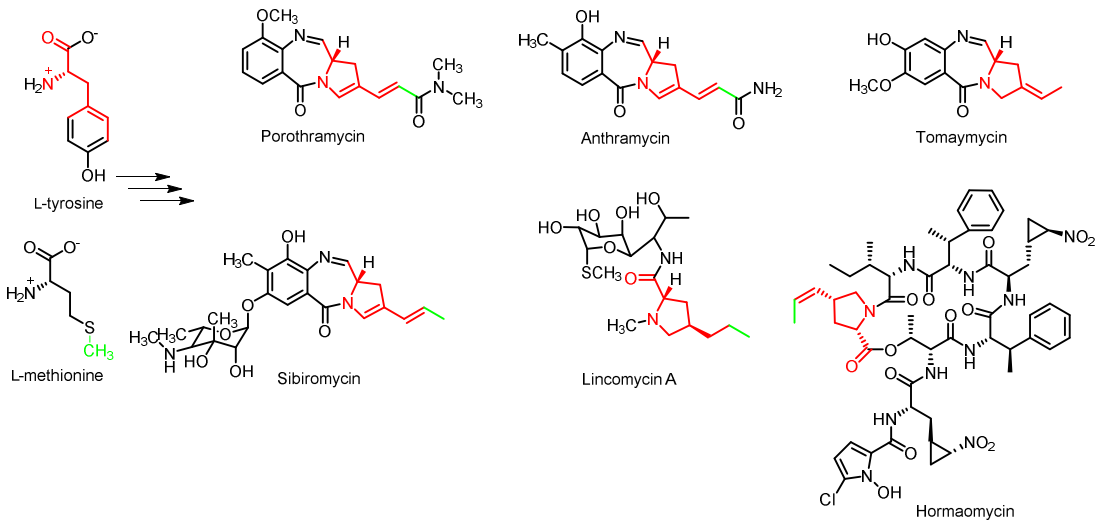
homologs are similar to dioxygenases suggesting they catalyze L-DOPA dioxygenation to form the dihydropyrrole scaffold **1.1**. The radiolabeling experiments indicated that two of the aromatic carbons in both L-[$^{14}$C]- tyrosine and L-[$^{14}$C]-DOPA did not get incorporated into the final PBD.[40] To account for this observation, a step is necessary to shorten the four-carbon side chain of the heterocycle in **1.1**. SibS and homologs are proposed to catalyze the hydrolysis of **1.1** to a vinyl dihydropyrrole **1.2**. SibZ and its homologs contain SAM-binding motifs and belong to a methyltransferase superfamily suggesting that they catalyze a methylation of **1.2** to form the 1-pyrroline **1.3**. The terminal carbon in the three-carbon side chain observed in the dihydropyrrole and pyrrolidine moieties derive from L-methionine (**Scheme 1-6**) presumably via the methyl donor *S*-adenosylmethionine (SAM). This supports the site of methylation. The SibZ homolog in tomaymycin biosynthesis is absent. This is consistent with the presence of a two-carbon side chain in its pyrrolidine moiety. SibT and its homologs belong to a flavin-utilizing superfamily, members of which resemble oxidoreductases. SibT is proposed to reduce the 1-pyrroline ring in **1.3** into a pyrrolidine derivative **1.4** using an F420-based cofactor. F420, structurally similar to flavin cofactors is a two-electron carrier that transfers hydride in a redox reaction. F420 is modified with a poly-glutamate tail (F420-$Glu_N$). SibY and its homologs belong to a γ-glutamyltranspeptidase superfamily suggesting they transfer an unknown number of glutamyl residues to F420. The proposed substrate **1.3** in the reduction reaction accumulated in a mutant strain of the lincomycin A producer *Streptomyces lincolnensis*, incapable of producing the F420 cofactor.[42] This supports the order of the hydrolysis and methylation steps as preceding the reduction of the ring. The reduction step represents a branching point in the pathways after which predicted enzymes unique to each natural

product impart structural differences in the moiety to produce distinct natural products. The resulting dihydropyrrole or pyrrolidine moiety is eventually incorporated into the natural product most likely by a non-ribosomal peptide synthetase (NRPS) or an NRPS-like protein present in the organism. For example, the putative NRPS SibD is proposed to incorporate the dihydropyrrole moiety into sibiromycin. This proposed pathway reconciles the putative roles of each enzyme with a chemically efficient biosynthetic route.




**Scheme 1-5.** Proposed biosynthetic pathway of the dihydropyrrole in sibiromycin. The boxed biosynthetic intermediate accumulated in a PBD producing strain (*S. lincolnensis*) that was mutated to incapacitate synthesis of the F420 cofactor.

Confirming the biosynthetic steps of the PBD dihydropyrrole moiety is a prerequisite to the development of engineered PBDs in the future. Of primary interest are the initial steps of the pathway as this facilitates engineering pathways in multiple PBD producers. As sibiromycin is the most potent naturally occurring PBD and was successfully engineered to 9-deoxysibiromycin, its biosynthesis was studied for the ultimate goal of




**Scheme 1-6.** The L-tyrosine and L-methionine precursors ultimately form the dihydropyrrole and pyrrolidine moiety depicted in red and green, respectively, in PBDs sibiromycin, tomaymycin, anthramycin and porothramycin and natural products lincomycin A and hormaomycin.

further engineering it in the future. This entailed the heterologous expression, purification and characterization of SibV, expected to form the dihydropyrrole scaffold, SibS and SibZ, expected to modify the dihydropyrrole exocyclic carbon chain and SibD, expected to incorporate the dihydropyrrole moiety into sibiromycin.

**Table 1-3.** Deduced function of proteins encoded by dihydropyrrole and pyrrolidine biosynthetic gene cluster. The sibiromycin dihydropyrrole biosynthetic gene cluster (GenBank accession number FJ768674) were compared with homologous ones involved in tomaymycin (GenBank accession number FJ768957), anthramycin (GenBank accession number EU195114), porothramycin (GenBank accession number HQ872605), lincomycin A (GenBank accession number EU124663) and hormaomycin (GenBank accession number HQ542230) biosynthesis. The highlighted enzymes have been reported to be active *in vitro*.

| Sibiromycin Biosynthetic Protein | Function | Homologs[a] | Identity/Similarity (%)[b] |
|---|---|---|---|
| SibD | Putative non-ribosomal peptide synthetase | TomB | 43/55 |
| | | Orf22 | 43/55 |
| | | Por21 | 43/56 |
| | | LmbC[c] | 32/42 |
| | | LmbN[d] | 34/48 |
| | | HrmP | 37/50 |
| SibU | Tyrosine hydroxylase | TomI | 48/60 |
| | | Orf13 | 44/55 |
| | | Por14 | 38/50 |
| | | LmbB2 | 44/56 |
| | | HrmE | 45/57 |
| SibV | Putative dioxygenase | TomH | 61/70 |
| | | Orf12 | 53/68 |
| | | Por13 | 47/64 |
| | | LmbB1 | 57/65 |
| | | HrmF | 52/63 |
| SibZ | Putative methyltransferase | Orf5 | 62/75 |
| | | Por10 | 61/73 |
| | | LmbW | 62/71 |
| | | HrmC | 51/68 |
| SibS | Putative hydrolase | TomK | 50/60 |
| | | Orf15 | 32/42 |
| | | Por16 | 31/43 |
| | | LmbX | 45/56 |
| SibT | Putative reductase | TomJ | 57/69 |
| | | Orf14 | 58/71 |
| | | Por15 | 56/69 |
| | | LmbY | 56/68 |
| | | HrmD | 45/57 |
| SibY | Putative γ-glutamyltransferase | TomL | 62/73 |
| | | Orf6 | 58/70 |
| | | Por11 | 57/69 |
| | | LmbA | 58/69 |
| | | HrmG | 56/68 |

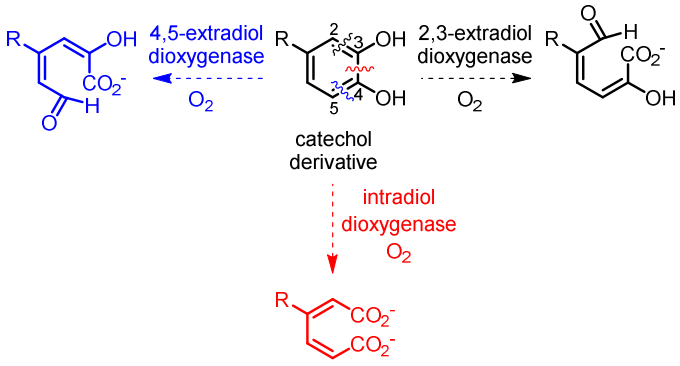
[a] Protein names of sibiromycin start with Sib, tomaymycin start with Tom, anthramycin start with Orf, porothramycin start with Por, lincomycin A start with Lmb and hormaomycin start with Hrm.
[b] Percent identity and similarity were determined using BLAST.
[c] Only the adenylation domain sequence in SibD was aligned with the standalone LmbC adenylase.
[d] Only the thiolation domain sequence in SibD was aligned with the thiolation domain sequence in LmbN.

# Chapter 2: Dihydropyrrole Scaffold Formation by SibV

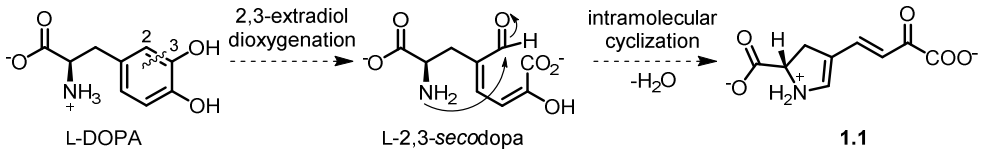
## 2.1 Introduction

The PBD dihydropyrrole moiety derives from L-tyrosine according to the incorporation pattern of radiolabeled amino acids fed to PBD producers.[30] Only the initial step of this pathway, an *ortho*-hydroxylase-catalyzed transformation of L-tyrosine to L-DOPA, has been verified experimentally *in vitro*.[38] A biosynthetic route that transforms L-DOPA to the dihydropyrrole ring was devised[23] based on chemistry commonly used by bacteria to degrade aromatic compounds.[43] Specifically, these catabolic pathways involve a cleavage of catechols similar to L-DOPA through a dioxygenation reaction. This step incorporates dioxygen into the aromatic ring to cleave a carbon-carbon bond either between the two hydroxyls (intradiol dioxygenation) to form a dicarboxylic acid or adjacent to one of the hydroxyls (extradiol dioxygenation) to form a semialdehyde (**Scheme 2-1**).[44] In the case of L-DOPA, extradiol dioxygenation was proposed to form the semialdehyde L-2,3-*seco*dopa that could then cyclize into the dihydropyrrole **1.1** (**Scheme 2-2**).[23] The action of dioxygenation is reasonable as the sibiromycin gene cluster, representative of the other PBD gene clusters, contain two genes that code for putative dioxygenases.[31]

The *sibV* gene from the sibiromycin gene cluster codes for SibV, a protein proposed to catalyze dioxygenation of L-DOPA.[31] SibV shares higher sequence homology with enzymes predicted to be dioxygenases (protein hits with 38-66% sequence identity and expect values from $10^{-26}$ to $10^{-65}$) and lower sequence homology with enzymes predicted to be a lactoylglutathione lyase (protein hits with 27-35% sequence identity and expect values from $10^{-4}$ to $10^{-7}$) according to BLAST analysis. The predicted amino acid sequence




**Scheme 2-1.** PBD biosynthesis is expected to involve oxidative cleavage of a catechol, a reaction observed during microbial degradation of aromatic compounds.[23]

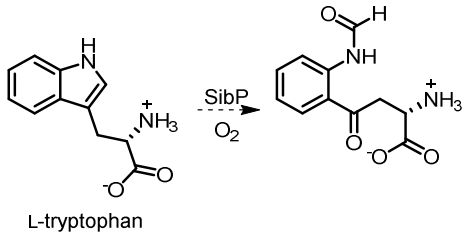



**Scheme 2-2.** The dihydropyrrole **1.1** may form via dioxygenation of L-DOPA into L-2,3-*seco*dopa followed by intramolecular cyclization to a common biosynthetic intermediate of PBD.[2]

of SibV contains part of the catechol 2,3-extradiol dioxygenase consensus sequence signature ($Hx_7FYx_2DPxGx_3E$).[45] The histidine and phenylalanine are both replaced with valine. Homologs of SibV are present in the anthramycin,[23] porothramycin,[33] tomaymycin,[32] lincomycin A[46-48] and hormaomycin[37] biosyntheses. All of their pathways are proposed to involve the equivalent dioxygenation step as the second step in the biosynthesis of their respective dihydropyrrole and pyrrolidine moieties. *In vitro* assays showed that LmbB1[46-48] from lincomycin A biosynthesis and HrmF[37] from hormaomycin biosynthesis catalyze a reaction with L-DOPA as the substrate to form a yellow product making SibV the likely dioxygenase of interest.

The sibiromycin gene cluster codes for another protein (SibP) with high sequence similarity to putative dioxygenases.[31] However, its amino acid sequence does not contain

any portion of the catechol 2,3-extradiol dioxygenase consensus signature sequence[45] suggesting its substrate is not a catechol like L-DOPA. SibP belongs to the tryptophan 2,3-dioxygenase superfamily according to BLAST analysis (domain hits with expect values from $10^{-46}$ to $10^{-87}$). Members of this superfamily use tryptophan or indolamine as substrates. Thus, SibP is proposed to cleave L-tryptophan, to initiate the transformation of the metabolic precursor to the anthranilate moiety (**Scheme 2-3**). The anthranilate moiety is common to PBDs and absent in lincomycin A and hormaomycin. This is in agreement with predicted SibP homologs present only in anthramycin, porothramycin and tilivalline biosynthetic pathways and absent in tomaymycin, lincomycin A or hormaomycin biosynthetic pathways. The anthranilate moiety in tomaymycin is derived from the shikimate pathway instead of L-tryptophan[32] explaining the absence of a SibP homolog in this particular PBD biosynthetic pathway. Taken together, the phylogenetic comparison shows that SibP is unlikely to catalyze the dioxygenation of L-DOPA, confirming that SibV is the most likely protein to catalyze the reaction.

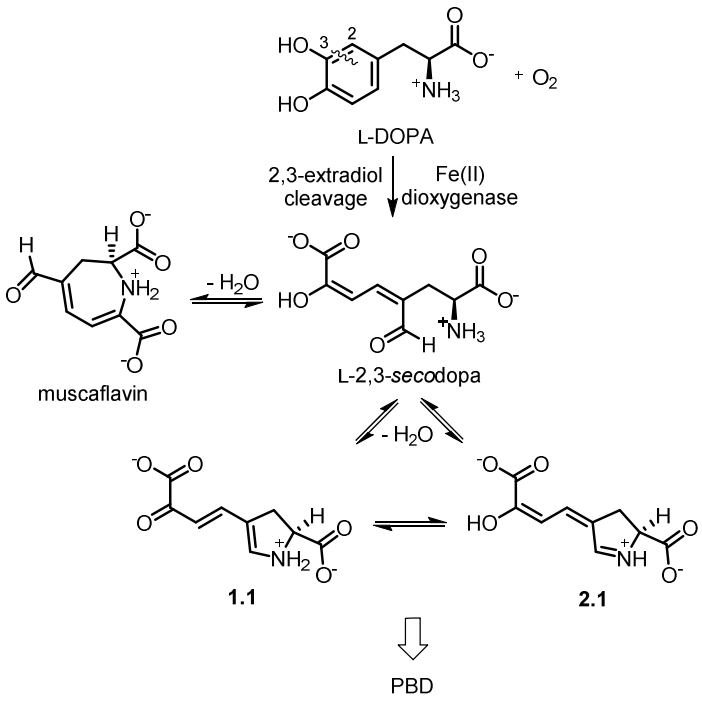


L-tryptophan

**Scheme 2-3.** SibP and its homologs are proposed to catalyze dioxygenation of L-tryptophan to initiate anthranilate moiety biosynthesis of PBDs sibiromycin, anthramycin, porothramycin and tilivalline.[2]

SibV and its homologs are expected to transform L-DOPA into L-2,3-*seco*dopa that has the potential to cyclize into the five-membered ring found in the PBD dihydropyrrole moiety (**Scheme 2-4**). However, multiple sites of cleavage and alternative cyclizations observed with other 2,3-extradiol dioxygenases could yield a plethora of possible products.

To confirm the dioxygenation step, SibV was produced and incubated with L-DOPA. The product(s) formed were identified and characterized. The SibV homolog in anthramycin (Orf12) was also similarly tested to determine the generality of this step along the path to the dihydropyrrole moiety.




**Scheme 2-4.** Two tautomers, **1.1** and **2.1** and a side product muscaflavin may form via L-2,3-*seco*dopa that is generated from oxidative cleavage of L-DOPA.

## 2.2 Results and Discussion

**Production of dioxygenases.** The putative dioxygenases associated with formation of the dihydropyrrole moiety in *Streptomyces refuineus's* anthramycin (Orf12) and *Streptosporangium sibiricum's* sibiromycin (SibV) were selected as representative of the general type of transformation. Orf12 and SibV were alternatively fused with an *N*-terminal $His_6$ and $His_6$-SUMO tag, respectively. Each was then expressed in BL21 (DE3) *Escherichia coli* and purified via Ni-NTA chromatography. The tag of SibV was removed by a SUMO-specific protease (Ulp1)[49] leaving behind a single non-native *N*-terminal

serine. The dioxygenases were purified to near homogeneity by size exclusion chromatography and reconstituted with $Fe^{2+}$ to yield light blue *holo*-dioxygenases. The final purification step provided *ca.* 150 mg $His_6$-Orf12 (Orf12) and 10 mg SibV per liter of growth media (**Appendix A-1**). Since $His_6$ tags sometimes interfere with protein oligomerization,[50,51] the oligomeric state of SibV, separated from its $His_6$-SUMO tag during purification, was examined by gel filtration. SibV was confirmed to form its expected dimer in solution (observed molecular mass of 34.6 kDa, theoretical monomeric mass of 17.1 kDa predicted by ExPASy ProtParam) (**Appendix A-2**). Gel filtration was used to determine that LmbB1 is also a dimer in solution[47] indicating that the oligomerization state of dioxygenases is conserved among the PBD and lincomycin A biosynthetic pathways.

**UV-visible spectroscopic characterization of the transient and final products formed by dioxygenase treatment of L-DOPA.** Transformation of L-DOPA by Orf12 generated a transient compound ($A_{max}$ at 378 nm) that subsequently diminished concurrent with formation of a yellow compound ($A_{max}$ at 414 nm) (**Figure 2-1**). SibV transformed L-DOPA to the same transient and final products as evidenced by equivalent changes in UV-visible absorbance (**Appendix A-3**). Thus, these two dioxygenases appear to promote identical reactions. The same $A_{max}$ at 414 nm was observed previously after turnover of dioxygenases LmbB1 and HrmF in lincomycin A[48] and hormaomycin[37] indicating that this transformation is not limited to PBD biosynthesis but general to pyrrolidine biosynthesis of many natural products.

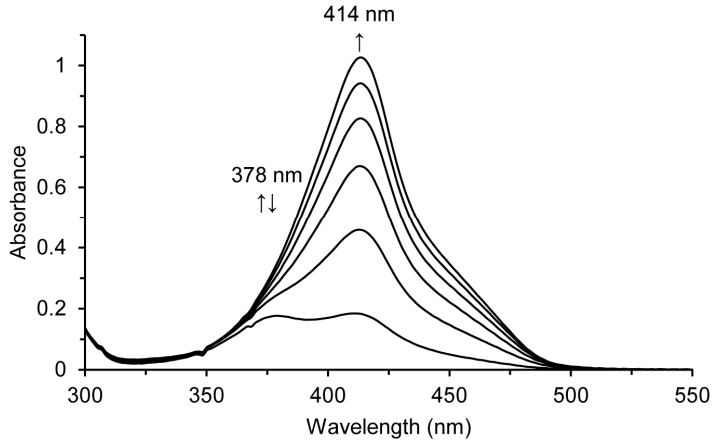



**Figure 2-1.** UV-visible spectroscopy of transient ($A_{max}$ = 378 nm) and yellow ($A_{max}$ = 414 nm) species during dioxygenation of L-DOPA catalyzed by Orf12. The reaction between L-DOPA (1.0 mM) and Orf12 (4.9 μM) was performed in sodium phosphate (250 mM) at pH 8.0. Scans were recorded every 20 sec for a total of 120 sec.

**$^1$H NMR spectroscopic analysis of the product formed by dioxygenation of L-DOPA catalyzed by Orf12.** The yellow compound generated in the dioxygenase-catalyzed transformation of L-DOPA was isolated by extraction and characterized by $^1$H NMR spectroscopy. This required a concentrated sample but oxygen-dependent inactivation of the enzyme made it necessary to add Orf12 in multiple aliquots to compensate for its loss of activity. The enzyme was subsequently removed by precipitating it with $CHCl_3$. The remaining reaction mixture in the aqueous phase was spiked with 10% $D_2O$ and then analyzed by NMR spectroscopy. No signals associated with L-DOPA were observed indicating its complete consumption (**Figure 2-2**). Assignment of the $^1$H signals was based on a model compound **1.3**, previously proposed as a downstream intermediate in the biosynthesis of the dihydropyrrole moiety (**Scheme 1-5**).[42] The chemical shifts, integration values and coupling constants were all consistent with 4-(2-oxo-3-butenoic-acid)-4,5-dehydro-L-proline, **1.1**. The signals located upfield of the HDO resonance were diagnostic of protons attached to $sp^3$ hybridized carbons and were assigned to H5α (3.11

ppm), H5β (2.72 ppm) and H4 (4.59 ppm) (**Figure 2-2**). These three signals are each a doublet of doublets arising from an AMX spin system that forms from coupling between the diastereotopic methylene protons H5α and H5β ($^2J_{HH}$ = 16 Hz) and are further split by H4 at the adjacent chiral center. Vicinal coupling constants between H5α and H4 ($^3J_{HH}$ = 12 Hz) and between H5β and H4 ($^3J_{HH}$ = 6.2 Hz) satisfied the Karplus relationship[52] that correlates vicinal coupling constants to dihedral angles (predicted to be 2.9° and 117.2° by Chem3D Pro, respectively). The resonance furthest downfield (7.59 ppm) was a singlet and consistent with proton H3 located on the unsaturated carbon in the dihydropyrrole. The remaining pair of doublets H1 (5.52 ppm) and H2 (7.58 ppm) were assigned to vicinal protons. The large HDO signal distorted the baseline beneath the H1 and H4 resonances and suppressed their integration from the expected value of one. $^1$H-$^1$H COSY analysis confirmed the expected connectivities between the protons with the exception of H4 (**Appendix A-4**). Its resonance is located near that of the solvent and was not observed due to solvent suppression. No aldehydic protons (9-10 ppm) were observed that would have indicated the formation of alternative products as discussed below (**Scheme 2-4**).

The multiplicity of signals for the pair of vicinal protons H1 and H2 were used to distinguish between the tautomers **1.1** and **2.1**. These protons produced a large coupling constant ($^3J_{HH}$ = 16 Hz) that is consistent with vicinal coupling constants of *trans*-olefinic protons ($^3J_{HH\ expected}$ = 12-18 Hz) such as those in the α-keto acid tautomer **1.1** but not in the α-enol acid tautomer **2.1.** This alternative would have exhibited a lower vicinal coupling constant ($^3J_{HH\ expected}$ < 11 Hz).[53] A pair of weaker signals (7.12 and 5.80 ppm) exhibits a coupling constant (8.5 Hz) consistent with the equivalent protons on **2.1**. Integration of these signals indicates the presence of approximately 5% **2.1**. The extinction

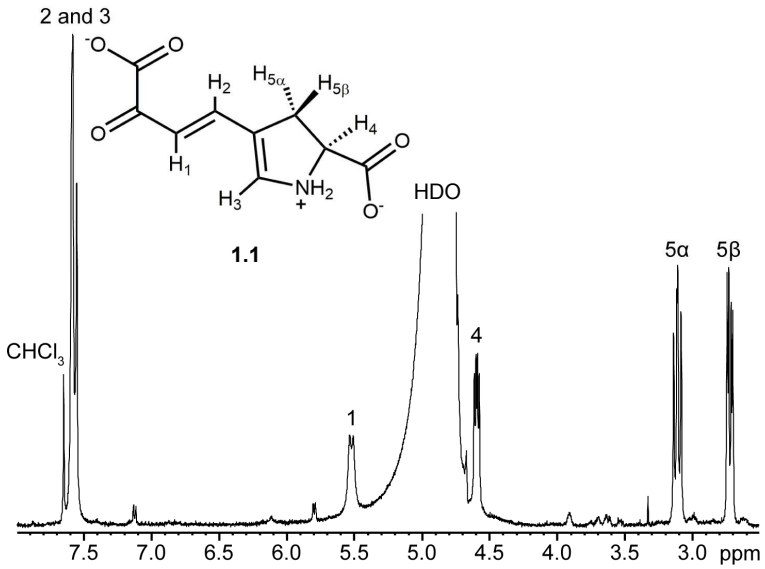


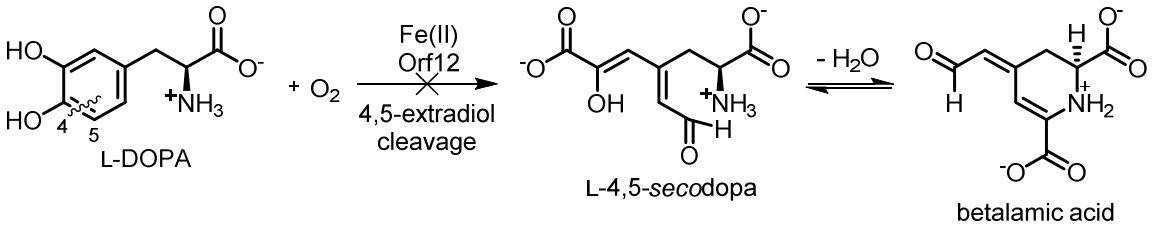

**Figure 2-2.** Annotated $^{1}$H NMR (500 MHz, 10% $D_2O$, $H_2O$ suppression with presaturation) spectrum of **1.1** formed by dioxygenation of L-DOPA catalyzed by Orf12. The solvent signal was truncated for clarity.

coefficient determined for **1.1** at pH 8.0 ($\varepsilon_{414\ nm}$ = 50 ± 3 $mM^{-1}cm^{-1}$) closely matches those reported for the LmB1 product ($\varepsilon_{413\ nm}$ = 48 ± 2 $mM^{-1}cm^{-1}$ and $\varepsilon_{414\ nm}$ = 45 ± 2 $mM^{-1}cm^{-1}$)[47,54] and provides confirmation that Orf12 generated the same product as LmbB1. However, LmbB1 was reported earlier to form **2.1**.[48] This was based on a $^{1}$H NMR spectrum of an equal mixture of product and starting material for which the coupling constant (8 Hz) of a substrate signal (6.94 ppm) was erroneously assigned to the product and used incorrectly to rule out the formation of **1.1**.[48] A $^{1}$H NMR spectrum of the starting material confirms the reassignment (**Appendix A-5**). Conditions used to isolate the enzyme product did not appear to alter the cyclized structure. NMR analysis of the crude product confirmed the predominance of **1.1** rather than **2.1** despite the greater conjugation of **2.1**. Differentiating between these tautomers is important to distinguish between potential substrates of the enzymes downstream in the biosynthetic pathways, particularly those of TomN, a tautomerase involved in tomaymycin biosynthesis.[55] We speculate that **1.1** is

similarly generated during lincomycin A and hormaomycin biosynthesis as well. Equivalent $ESI^+$-MS data were recorded for the product formed by both HrmF[37] and Orf12.

Previously, a 2,3-extradiol dioxygenase involved in betalain biosynthesis in the fungus *Amanita muscaria* was reported to form the seven-membered ring, muscaflavin after formation of L-2,3-*seco*dopa (**Scheme 2-4**).[56] This raises the possibility of a pathway that is non-productive for PBD formation may act competitively during dihydropyrrole biosynthesis. However, only **1.1** was detected from cyclization of L-2,3-*seco*dopa. A characteristic aldehydic signal expected in the $^1$H NMR spectrum of muscaflavin was not observed. The $^1$H NMR signals for the product assigned as **1.1** did not agree with the literature values reported for muscaflavin.[56] This indicates that the intramolecular cyclization of L-2,3-*seco*dopa to muscaflavin observed in betalain biosynthesis[56] does not occur in PBD dihydropyrrole biosynthesis. It is unclear if Orf12 prevents this process, which had been described as spontaneous.[56]

The *A. muscaria* 2,3-extradiol dioxygenase is known to cleave at the alternative C4-C5 bond of L-DOPA to form betalamic acid via L-4,5-*seco*dopa (**Scheme 2-5**).[56] This suggests that the dioxygenases involved in PBD dihydropyrrole biosynthesis might have the potential to cleave the aromatic ring of L-DOPA at multiple sites. However, no products of Orf12 resulting from 4,5-cleavage of L-DOPA were observed as indicated by the absence of $^1$H NMR signals for the aldehydic protons of L-4,5-*seco*dopa or betalamic acid. This indicates that Orf12 acts regiospecifically to cleave only the expected C2-C3 bond of L-DOPA. Alternative cleavage to form L-4,5-*seco*dopa and betalamic acid would have prevented formation of the five-membered dihydropyrrole.

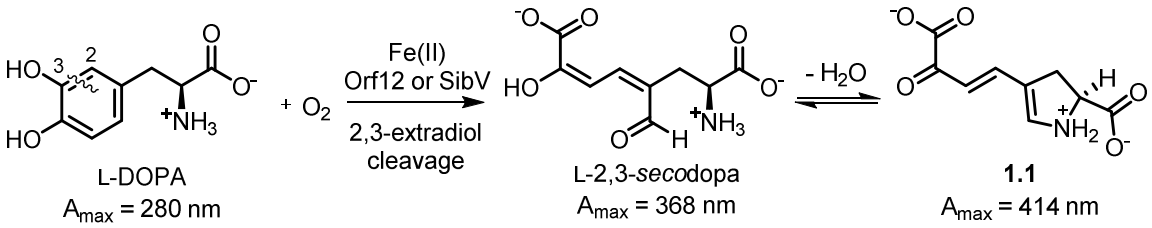



**Scheme 2-5.** PBD biosynthesis does not involve a competitive oxidation of L-DOPA to generate a non-productive derivative, betalamic acid.

**Is L-2,3-*seco*dopa a biosynthetic intermediate of PBD biosynthesis?** The transient product observed by UV-visible spectroscopy during dioxygenation is consistent with L-2,3-*seco*dopa since its absorbance maximum was similar to that exhibited by 2,3-*seco*caffeic acid ($A_{max}$ at 384 nm).[48] The extra conjugation of this analog explains the red-shift (16 nm) of its absorbance maximum relative to that of L-2,3-*seco*dopa. UV-visible spectroscopy revealed that the equilibrium between **1.1** and the colorless compound L-2,3-*seco*dopa ($A_{max}$ at 368 nm) can be shifted by increasingly acidic conditions (pH 3-5) (**Appendix A-6**). The absorbance maximum of L-2,3-*seco*dopa was slightly lower than that implied during its formation by dioxygenation ($A_{max}$ at 378 nm) of L-DOPA due to the additive effect of the absorbance from **1.1**. The transient compound was verified to be L-2,3-*seco*dopa ($\varepsilon_{368\ nm}$ = 27 ± 2 $mM^{-1}cm^{-1}$ at pH 4.0) by its ability to reconvert to **1.1** upon exposure to base (**Appendix A-6**) and by the presence of a downfield resonance (9.08 ppm) in its $^1H$ NMR spectrum (**Appendix A-7**) that corresponded to its aldehydic proton. The assignment was further confirmed by the coupling constant ($^3J_{HH}$ = 7.4 Hz) between vicinal protons H1 (6.01 ppm) and H2 (7.11 ppm) that is diagnostic of an anti-Karplus type arrangement of protons ($^3J_{HH\ expected}$ = <11 Hz) found in L-2,3-*seco*dopa and missing in **1.1**. Partial degradation of L-2,3-*seco*dopa occurred within the time (1 h) required for NMR characterization, as indicated by a decreasing signal at its maximal absorbance of 368 nm.

The degradation products likely contributed the small signals in the $^1$H NMR spectrum that remain unannotated.

The formation of L-2,3-*seco*dopa catalyzed independently by Orf12 and SibV was consistent with the products generated by turnover of other 2,3-extradiol dioxygenases that cleave catechols to form acyclic 2,3-*seco*catechols.[57] Further reaction of L-2,3-*seco*dopa begins with condensation of the nucleophilic amine with the proximal aldehyde and subsequently undergoes dehydration to form the dihydropyrrole **1.1** (**Scheme 2-6**). This cyclization occurs spontaneously since its rate was shown by others to be independent of enzyme concentration in the LmbB1-dependent production of L-2,3-*seco*dopa.[54] Whether downstream biosynthetic enzyme(s) prefer the acyclic compound L-2,3-*seco*dopa or the cyclic compound **1.1** as substrate is not yet known.




**Scheme 2-6.** Orf12 and SibV catalyzed the dioxygenation of L-DOPA to form L-2,3-*seco*dopa, a transient species observed only within the first minute of dioxygenation at pH >8 before it cyclized spontaneously to **1.1**.

The lifetime of L-2,3-*seco*dopa at pH 8 (250 mM sodium phosphate) is unusually short for further processing by subsequent biosynthetic enzymes. Exposure to a basic environment slows the cyclization of L-2,3-*seco*dopa[54] suggesting that dioxygenation reactions performed at pH > 8 may shift the equilibrium from **1.1** to L-2,3-*seco*dopa. To assess the effect of increasing basicity on the equilibrium between L-2,3-*seco*dopa and **1.1**, the cleavage of L-DOPA by either SibV or Orf12 was performed over a range of pH conditions (**Figure 2-1 and Appendix A-3 and A-8**). Not surprisingly, L-2,3-*seco*dopa

formed too transiently to be observed under neutral (pH 7) or slightly acidic conditions (pH 6). In contrast, L-2,3-*seco*dopa persisted longer when it formed at pH 9 ($A_{max}$ at 378 nm disappeared within 80 sec) rather than at pH 8 ($A_{max}$ at 378 nm disappeared within 40 sec). However, the transiency of L-2,3-*seco*dopa even at pH 9 suggests that **1.1** is the only candidate that would persist long enough for further processing during assembly of the PBD dihydropyrrole moiety.

**Lifetime of 1.1.** The lability and non-enzymatic degradation of **1.1** was apparent from the decrease of its absorbance at 414 nm over hours (**Appendix A-9**). At pH 8.0, its half-life was approximately 36 h at 25 °C and 25 h at 37 °C. A $^1$H NMR spectrum of the final degradation mixture showed a new downfield singlet (8.91 ppm) corresponding to an aldehydic proton but there were also at least another 16 new signals that could not be definitely assigned to any mixture of aldehyde containing isomers of **1.1** (muscaflavin, L-4,5-*seco*dopa or betalamic acid). The instability of **1.1** may also arise from its tendency to polymerize via nucleophilic addition at the α,β-unsaturated γ-keto acid. The observed lability of **1.1** suggests that the efficiency of the next enzyme in the biosynthetic pathway that consumes it is an important property to consider when manipulating this pathway in the future for engineering new PBD derivatives.

**2.3 Summary**

The dioxygenases Orf12 and SibV associated with the biosynthesis of PBDs were shown to catalyze the cleavage of the aromatic ring in L-DOPA to form L-2,3-*seco*dopa. This compound in turn spontaneously cyclized to **1.1** verified by $^1$H NMR spectroscopy. Together, these processes are crucial for generating the scaffold of the dihydropyrrole moiety common to all PBDs. The oligomeric state and enzymatic product generated by the

dioxygenases Orf12, SibV, LmbB1 and HrmF are conserved in the formation of PBDs, lincomycin A and hormaomycin. These observations suggest a likely commonality in all dihydropyrrole and pyrrolidine biosynthetic pathways originating from L-tyrosine.

**2.4 Experimental Procedures**

**Materials.** The pET28b plasmid and electrocompetent BL21(DE3) strain were obtained from Novagen (Darmstadt, Germany). The pSMT plasmid and Ulp1 expression vector were generously donated by Dr. C. Lima at the Sloan Kettering Institute (New York, NY).[49] Electrocompetent strain of GeneHogs *E. coli* was obtained from Invitrogen (Carlsbad, CA). Ni-NTA agarose was obtained from Qiagen (Valencia, CA). Sephacryl S-200 HR and a gel filtration calibration kit containing low molecular weight standards were obtained from GE Healthcare (Piscataway, NJ). An Econo-Pac 10DG column prepacked with Bio-Gel P-6DG gel was obtained from Bio-Rad (Hercules, CA). The Coomassie Brilliant Blue R-250 protein stain powder was obtained from Bio-Rad (Hercules, CA). L-DOPA was obtained from Sigma (St. Louis, MO). $H_2O$ was purified to a resistivity of 18.2 Ω-cm. All other reagents were purchased at the highest grade commercially available and used without further purification.

**General methods.** Protein purification was performed using an ÄKTA Prime Liquid Chromatography from GE Healthcare. Protein samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) made with 12% resolving and 4% stacking acrylamide layers. Protein samples were visualized after SDS-PAGE by staining the gel with Coomassie Brilliant Blue R-250. Protein concentrations were calculated using extinction coefficients (ε) at 280 nm predicted by ExPASy ProtParam.[58] All UV-visible absorption spectra were obtained using a Varian

monochromator UV-visible Cary 100 Spectrophotometer (Walnut Creek, CA). $^{1}$H NMR experiments were performed on a Bruker 500 MHz with $H_2O$ suppression using a presaturation pulse sequence and COSY experiments were performed on a Bruker 400 MHz with $H_2O$ suppression.[59] Chemical shifts ($\delta$) and coupling constants ($J$) are reported in parts per million (ppm) and Hertz (Hz), respectively. Mass spectra were obtained in the $ESI^+$ mode using a Waters Acquity UPLC-Xevo-G2-Q-Tof-MS (Milford, MA) equipped with a 2.1 mm x 50 mm BEH-C18 column (1.7 μM and 300 Å pore size) using a 0-80% aqueous acetonitrile gradient over 12 min (0.3 mL/min). Quaternary structure was characterized with a gel filtration column (Superdex 200 HR 10/3) generously provided by Dr. Z. Kelman formerly at University of Maryland Biotechnology Institute (UMBI) (Rockville, MD). Dihedral angles were predicted by Chem3D Pro 13.0.2 from PerkinElmer (Waltham, MA). All assays were performed at 25 °C unless stated otherwise. All plasmid constructs were confirmed by sequencing (UMBI).

**Expression of *apo*-dioxygenases.** BL21(DE3) starter cultures, transformed with either pSMT3/*sibV* or pET28b/*orf12*,[60] were diluted 200-fold in Miller LB broth supplemented with kanamycin (50 μg/mL) and grown at 37 °C with agitation until the $OD_{600\ nm}$ reached 0.9. The cell cultures were then induced with D-lactose (0.2%) at 30 °C with agitation (12 h), harvested by centrifugation at 5,000 rpm at 4 °C (15 min) and flash frozen in $N_2$ (l) for storage at -80 °C.

**Purification of *apo*-dioxygenases.** All purification steps were performed at 4 °C. Cells alternatively expressing $His_6$-Orf12 or $His_6$-SUMO-SibV were resuspended in wash buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 1 mM fresh DTT and 20 mM imidazole). Protease inhibitors benzamidine (1 mM) and phenylmethylsulfonyl fluoride (1

mM) were added and then the cells were lysed by four passages through a French Press at 1,000 psi. Cell debris was removed by centrifugation at 15,000 rpm (30 min) and the supernatant was loaded onto a Ni-NTA agarose (11 mL) column pre-equilibrated with wash buffer. The resin was washed with wash buffer and the dioxygenase was eluted with elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 1 mM fresh DTT and 250 mM imidazole) at 1 mL/min. The SibV fusion underwent additional steps to remove the SUMO tag via proteolysis by adding Ulp1 (1:100 w/w) to the eluate for 12 h. The digestion was dialyzed (2 x 1 L) (12-14 kDa MWCO) against wash buffer, reapplied to the Ni-NTA agarose column pre-equilibrated with wash buffer and eluted with wash buffer. $His_6$-Orf12 and SibV were then separately dialyzed (2 x 1 L) against gel filtration buffer (50 mM sodium phosphate pH 8.0, 10% glycerol and 1 mM fresh DTT), concentrated in a pressurized stirred cell (10,000 NMWL) (EMD Millipore) under $N_2$ (g) (40 bars) to a volume of 600 μL and applied onto a Sephacryl S-200 HR column (254 mL) pre-equilibrated with gel filtration buffer. The dioxygenases were then eluted with gel filtration buffer at 0.4 mL/min. The resulting *apo*-dioxygenases were ≥ 95% pure as estimated by ImageQuant analysis after their separation by denaturing gel electrophoresis and staining with Coomassie Brilliant Blue. Enzyme solutions in gel filtration buffer were flash frozen in droplets with $N_2$ (l) and stored at -80 °C until needed.

**Reconstitution of *apo*-dioxygenases with $Fe^{2+}$.** Each *apo*-dioxygenases (3.2 mM) was reconstituted in gel filtration buffer with $FeSO_4 \cdot 7H_2O$ (2.5 mM) in the presence of DTT (15.6 mM) and L-ascorbic acid (sodium salt, 15.6 mM) for 25 min at 4 °C producing red solutions. The reconstituting agents were removed via buffer exchange into desalting buffer (100 mM sodium phosphate pH 8.0) with a P-6DG desalting column (10 mL) pre-

equilibrated with desalting buffer. The resulting light blue *holo*-dioxygenases solutions in desalting buffer were flash frozen in droplets with $N_2$ (l) and stored at -80 °C. The *holo*-dioxygenases used for assays were subjected to a maximum of one freeze and thaw cycle.

**Determination of the oligomeric state of SibV.** Ribonuclease A (13.7 kDa), ovalbumin (44.0 kDa), conalbumin (75.0 kDa) and aldolase (158.0 kDa) were used as molecular weight standards and blue dextran (2,000.0 kDa) was used to determine the void volume ($V_o$). The molecular weight standards and *holo*-SibV (20 μg) were individually dissolved in size exclusion buffer (50 mM sodium phosphate pH 8.0, 150 mM NaCl, and 1 mM fresh DTT) and loaded onto the Superdex-200 resin column that had been pre-equilibrated with size exclusion buffer. Each sample was eluted with size exclusion buffer at 0.4 mL/min at 4 °C. The partition coefficient ($K_{av}$) was calculated (**Equation 2-1**) using elution volume ($V_e$), $V_o$, and packed resin volume ($V_c$). A calibration curve was generated by plotting $K_{av}$ against the log of the molecular weight ($M_R$).

**Equation 2-1.** $$K_{av} = (V_e - V_o) / (V_c - V_o)$$

**Generation and purification of the product 1.1 formed by Orf12.** Freshly thawed *holo*-Orf12 (0.4 mM) was added in aliquots (44) every 30 sec to a solution of L-DOPA (17.7 mM) in sodium phosphate (30 mM, pH 8.0). The enzyme was precipitated by washing the reaction mixture with chloroform (3 x 3 mL). Residual chloroform was removed from the aqueous layer under reduced pressure for 20 min to yield a yellow solution. Compound **1.1**: $^1$H NMR (10% $D_2O$, 500 MHz) $\delta$ 7.59 ppm (s, 1H), 7.58 ppm (d, $J$ = 16 Hz, 1H), 5.52 ppm (d, $J$ = 16 Hz, 0.5H), 4.59 ppm (dd, $J$ = 6.2, 12 Hz, 0.8H), 3.11 ppm (dd, $J$ = 12, 16 Hz, 1H), 2.72 ppm (dd, $J$ = 6.2, 16 Hz, 1H). ESI$^+$-MS: m/z 212.06 $(M)^+$, 166.05, 148.04.

**Acid-catalyzed hydrolysis of 1.1 to form L-2,3-*seco*dopa.** A solution of **1.1** (13 mM) in sodium phosphate (20 mM, pH 8.0) was spiked with deuterium chloride (DCl) (1.6% v/v) to form L-2,3-*seco*dopa. $^1$H NMR (10% $D_2O$, 400 MHz) $\delta$ 9.08 ppm (s, 1H), 7.11 ppm (d, $J$ = 7.4 Hz, 1H), 6.01 ppm (d, $J$ = 7.4 Hz, 1H), 3.44 ppm (m, 2H), 2.34 ppm (d, $J$ = 15 Hz, 1H).

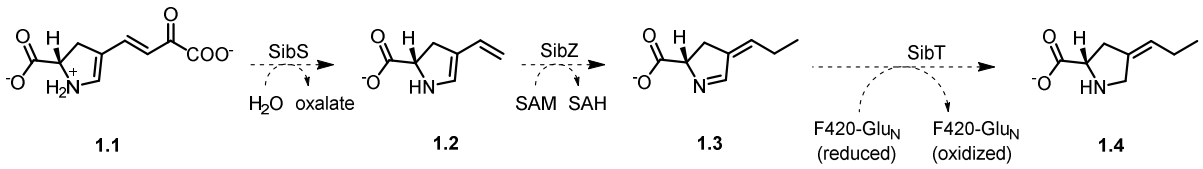
**Determination of extinction coefficients for 1.1 and L-2,3-*seco*dopa.** A fixed concentration of L-DOPA (7.6 mM) determined spectrophotometrically ($\varepsilon_{280\,nm}$ = 2.53 $mM^{-1}cm^{-1}$)[61] was transformed into **1.1** by batchwise addition of *holo*-Orf12 (0.2 mM) described above. $^1$H NMR analysis confirmed full conversion to **1.1** based on the lack of signals associated with L-DOPA. The absorbance of **1.1** at 414 nm was measured at pH 8.0 and Beer's law was used to calculate the $\varepsilon_{414\,nm}$ of **1.1**. A fixed concentration of **1.1** (15 μM) determined spectrophotometrically was then transformed into L-2,3-*seco*dopa with acid (a 1000-fold scaled down version of the procedure described above). Disappearance of an absorbance maximum at 414 nm and appearance of an absorbance maximum at 368 nm confirmed full conversion to L-2,3-*seco*dopa. The $\varepsilon_{368\,nm}$ of L-2,3-*seco*dopa was back-calculated at pH 4.0 using the $\varepsilon_{414\,nm}$ of **1.1** and the absorbance values at those respective wavelengths for both compounds (15 μM). These studies were done in triplicate and the error represented one standard deviation of uncertainty.

# Chapter 3: Probing the Roles of SibS and SibZ

## 3.1 Introduction

Upon formation of the dihydropyrrole **1.1**, a hydrolysis and a methylation step were proposed to modify its exocyclic carbon chain as the next steps in the biosynthetic pathway **(Scheme 3-1)**.[23] The hydrolysis step that removes two-carbon units from **1.1** was proposed to rationalize why PBD producers incorporated only seven out of the nine radiolabeled carbons from either [$^{14}$C]-L-tyrosine or [$^{14}$C]-L-DOPA into the PBD dihydropyrrole moiety **(Scheme 1-6)**.[30] The methylation step was used to explain why the third terminal carbon in the exocyclic chain in the final moiety originates from the methyl carbon in L-methionine **(Scheme 1-6)**.[30] L-methionine is a precursor of the biological methylating agent SAM. These two modifications could form **1.3**, a biosynthetic intermediate generated by a strain of the lincomycin A producer mutated to incapacitate synthesis of the F420 cofactor and prevent formation of **1.4**.[31,32,42] The F420 cofactor is a deazariboflavin derivative involved in redox reactions. The mutant study suggests that the hydrolysis and methylation steps precede the F420-dependent reduction step. Subsequent steps were proposed to impart the structural diversity observed among the dihydropyrrole and pyrrolidine moieties unique to each producer.[2,33]

The hydrolysis and reduction reactions expected to be common to the biosynthesis of the dihydropyrrole and pyrrolidine moiety are logically catalyzed by enzymes conserved throughout the pathways. The sibiromycin gene cluster contains two genes (*sibT* and *sibY*) that have orthologs in PBD, lincomycin A and hormaomycin gene clusters.[2] Both of these genes are expected to code for proteins involved in reducing **1.3** to **1.4** based on BLAST




**Scheme 3-1.** Biosynthetic steps proposed to modify the exocyclic carbon chain in the dihydropyrrole **1.1** to form sibiromycin, representative of transformations predicted throughout the syntheses of other PBDs as well as lincomycin A and hormaomycin.

analysis.[31] SibT, encoded by the *sibT* gene, contains a domain in its predicted amino acid sequence characteristic of F420-dependent oxidoreductases (expect value of $10^{-33}$).[31] SibY, encoded by the *sibY* gene, belongs to the γ-glutamyltranspeptide superfamily according to BLAST analysis (domain hits with expect values from $10^{-39}$ to $10^{-144}$).[31] It is proposed that SibY conjugates the requisite number of glutamyl residues onto the F420 cofactor to bind SibT allowing the cofactor to provide the hydride necessary for reducing **1.3**.[31]

Surprisingly no other gene is conserved throughout the PBD, lincomycin A and hormaomycin biosynthetic pathways, making it difficult to assign a gene product that catalyzes the hydrolysis of **1.1** to **1.2**. The *sibS* gene in the sibiromycin gene cluster is conserved throughout the other gene clusters except in hormaomycin's. Assuming another gene codes for the hydrolase used in hormaomycin biosynthesis, *sibS* was tentatively proposed to code for the hydrolase (SibS).[31] SibS shares sequence homology mostly with uncharacterized proteins, low (21%) sequence homology with the isomerase PhzF[62,63] and no sequence homology to known lyases according to BLAST analysis. However, the biosynthesis of the dihydropyrrole moiety appears to require SibS. A knockout of the *sibS* gene[64] prevented the production of sibiromycin in the resulting mutant *S. sibiricum* strain. The analogous knockout of the *sibS* ortholog *lmbX* resulted in a mutant *S. lincolnensis* strain incapable of producing lincomycin A without the addition of exogenous 4-

propylproline, a precursor of its pyrrolidine moiety.[65] The gene knockout and rescue studies indicate that SibS is essential in the biosynthesis of sibiromycin's dihydropyrrole moiety. The means by which the two terminal carbons from **1.1** are removed during hormaomycin biosynthesis is unclear, since no *sibS* ortholog could be located in its corresponding gene cluster or elsewhere on the genome using Southern hybridization and degenerate PCR.[37]

The methylation reaction is expected to be common throughout the pathways except in tomaymycin that lacks one carbon common to the other PBDs, lincomycin A and hormaomycin. The *sibZ* gene from the sibiromycin gene cluster is proposed to code for the methyltransferase (SibZ) based on the presence of a SAM binding motif ($AGCGCGQx_{17}DNx_{25}ADAx_{17}I$) in the predicted amino acid sequence according to BLAST analysis. This assignment is further supported by the conservation of SibZ homologs throughout the pathways except tomamycin's.[31-33]

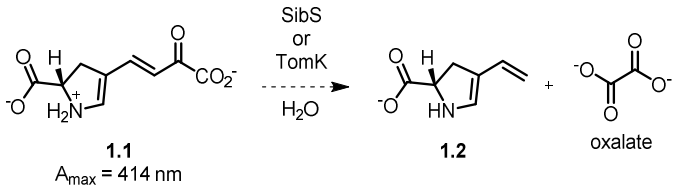
SibS and its homologs are tentatively proposed to hydrolyze **1.1** to **1.2** while SibZ and its homologs are expected to methylate **1.2** to **1.3** to continue the biosynthesis of the dihydropyrrole moiety (**Scheme 2-4**). To confirm the functional assignment of these proteins, SibS and SibZ were produced and treated with their expected substrates. As these transformations are expected to be common to other dihydropyrrole moiety biosyntheses, the SibS homolog in tomaymycin (TomK) was also produced and similarly tested.

### 3.2 Results and Discussion

**Production of SibS, TomK and SibZ.** SibZ expression and purification followed the cloning of the *sibZ* gene into the BamHI and HindIII sites of the pSMT3 vector. The equivalent vectors containing the *sibS*[66] and *tomK*[67] genes were already available. Each

was expressed in *E. coli* as $His_6$-SUMO fusion proteins and purified via Ni-NTA and size exclusion chromatography in a manner similar to the purification of SibV (Chapter 2). The final purification step provided *ca.* 11 mg SibS, 3 mg TomK and 4 mg SibZ per liter of growth media (**Appendix B-1**).

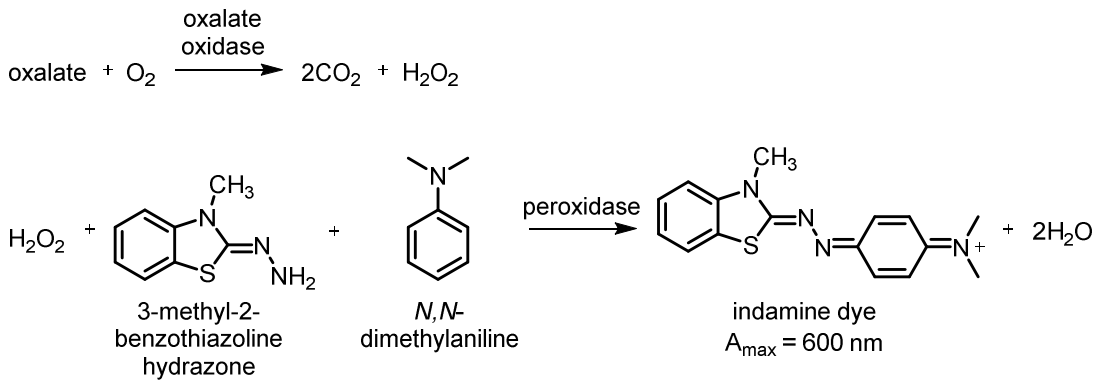
**SibS and TomK treatment with 1.1.** SibS was incubated with **1.1** and assayed for the hydrolysis product **1.2 (Scheme 3-2)**. UV-visible absorbance spectra of the **1.1** (20 μM) with and without SibS (1 μM) were monitored hourly up to 24 h. Formation of the catalytic product **1.2** was expected to be accompanied by an increase in absorbance at a wavelength lower than that of **1.1** (414 nm) since the product is less conjugated than the substrate. However, no new signal was observed suggesting hydrolysis to remove the α-keto acid from **1.1** did not occur. The absorbance maximum at 414 nm diminished over time for both the reaction and the negative control consistent with non-enzymatic decomposition of the substrate observed in Chapter 2.




**Scheme 3-2.** SibS and TomK were expected to catalyze the hydrolysis of **1.1** into **1.2** and oxalate as part of sibiromycin and tomaymycin biosynthesis, respectively.[31, 32]

Since hydrolysis was expected to form oxalate as the second product, a coupled enzyme assay[68] that detects oxalate was used as well to monitor the reaction **(Scheme 3-3)**. The assay relies on oxidizing oxalate to form hydrogen peroxide that is, in turn, converted to indamine. Indamine is a purple dye that can be detected spectrophotometrically ($A_{max}$ = 600 nm). An aliquot of the reaction mixture containing **1.1** (1 mM) and SibS (1 μM) was subjected to the coupling enzyme assay and compared with

an equivalent reaction mixture without SibS. The absorbance intensities at 600 nm were the same for both reactions suggesting the expected hydrolysis did not occur and produce oxalate. However, a positive control indicates that oxalate could only be detected in equivalent reactions spiked with at least 20 μM oxalate. Substrate turnover of at least 2% was necessary to obtain a detectable signal.




**Scheme 3-3.** Coupled enzyme assay to indirectly detect formation of oxalate.

Reaction mixtures with and without SibS were also subjected to separation by HPLC but their chromatograms appeared to be identical again suggesting no hydrolysis occurred **(Figure 3-1)**. To verify that no unexpected proteolysis occurred during the purification of SibS to prevent activity, the mass of purified SibS was verified. SibS had a parent ion experimental mass of 30790.95 Da in agreement with its theoretical mass of 30790.89 Da within a mass deviation of 1.9 ppm. Replacing SibS with homolog TomK did not change the results indicating the inactivity observed under the assay conditions is not unique to sibiromycin biosynthesis.

Hydrolysis of **1.1** was expected to involve an initial tautomerization, a step observed during the hydrolysis of another α-keto acid substrate between a carbon-carbon bond by another hydrolase (BphD)[69] (**Scheme 3-4**). The assays described above (UV-visible spectroscopy, coupled enzyme assay and HPLC analysis) are not sufficient to detect

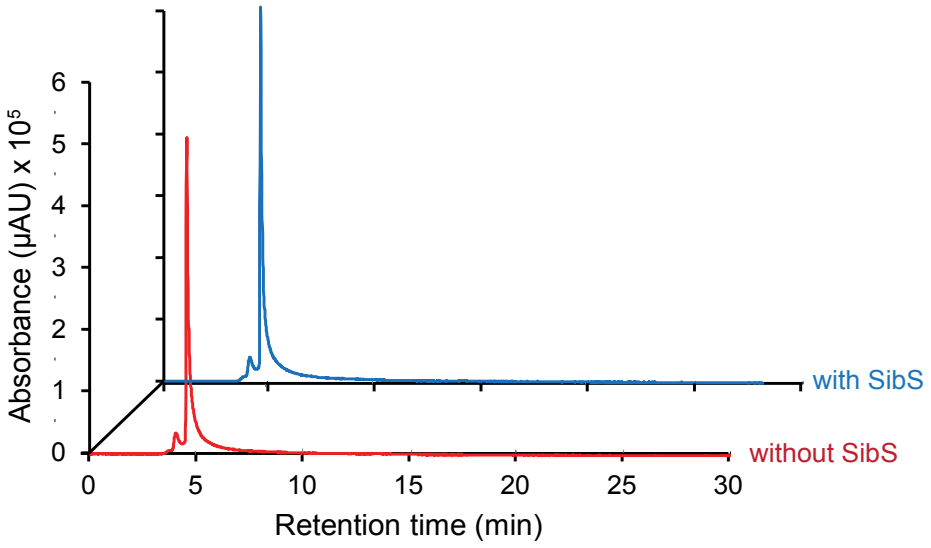
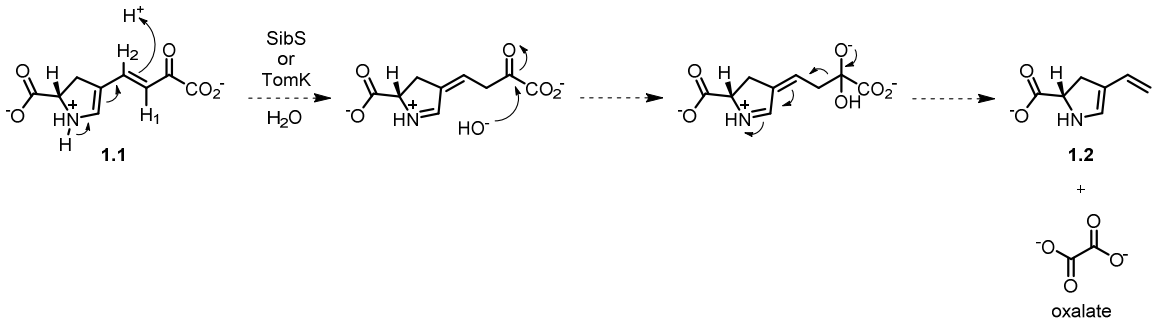



**Figure 3-1.** HPLC chromatogram (blue) monitoring SibS incubated with of **1.1** ($t_r$ = 4.9 min) expected to form **1.2** and oxalate upon hydrolysis. The reaction mixture was quenched at 4 h and subjected to HPLC separation. The additional chromatogram (red) depicts an equivalent reaction mixture lacking SibS as a negative control.

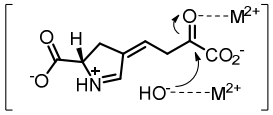
tautomerization of **1.1**. Thus a reaction mixture of SibS with **1.1** was additionally monitored by $^1$H NMR spectroscopy to detect the disappearance of vicinal coupling between H1 and H2 characteristic of the tautomer **1.1** as described in Chapter 2 and for the appearance of a methylene signal H1 upon tautomerization to the imine. However, all changes in proton signals observed over 24 h were also present in the spectrum of the equivalent reaction without SibS. These changes corresponded to the non-enzymatic degradation of **1.1** addressed in Chapter 2. SibS did not appear to catalyze tautomerization or hydrolysis of **1.1** under the assay conditions. Replacing SibS with TomK produced identical results.

The addition of a metal cofactor may be indispensable for catalytic activity as observed with other hydrolases.[70] The divalent metal cation may act as a Lewis acid to coordinate the carbonyl oxygen and reduce the pKa of $H_2O$ to generate the hydroxide nucleophile. This could promote the nucleophilic attack on the carbonyl carbon forming a




**Scheme 3-4.** A possible mechanism of SibS or TomK catalyzed hydrolysis of **1.1** initiates with a tautomerization step, in analogy to the mechanism of another hydrolase (BphD) that hydrolyzes a carbon-carbon bond of another α-keto acid substrate.[69]

tetrahedral intermediate before breaking down into **1.2** and oxalate (**Scheme 3-5**). To test the requirement of divalent metal cation, $Mg^{2+}$, $Zn^{2+}$ and $Mn^{2+}$ capable of promoting activity of other hydrolases[70] were considered. The reaction mixtures containing **1.1** (1 mM) with and without SibS (1 μM) were incubated with a divalent metal cation (1 mM). However, neither oxalate not **1.2** was detected using the oxalate and HPLC assay described above.



**Scheme 3-5.** Potential role of a divalent metal cation to facilitate hydrolysis of **1.1**.

**Binding of SibS with 1.1.** The inability of SibS to process **1.1** suggests that the putative substrate and role of SibS was incorrect. Since an enzyme-catalyzed reaction with its substrate entails binding the substrate prior to its turnover, it was instructive to determine if SibS had any affinity for **1.1**. Low affinity (> 1 mM) would further support a revision of SibS's role and substrate in the biosynthesis of sibiromycin. A decrease in the intrinsic fluorescence of SibS was observed during titration of **1.1** and consistent with **1.1** binding SibS to quench the fluorescence signal. The decrease in fluorescence intensity was used to fit for the dissociation constant ($K_D$ = 64 ± 2 μM) assuming a simple binding process.[71]

The ability of SibS to bind **1.1** suggests either SibS catalyzes a reaction with **1.1** but requires an unidentified cofactor or protein partner or **1.1** is structurally similar to the actual substrate. Either scenario still assumes that SibS participates in dihydropyrrole moiety biosynthesis.

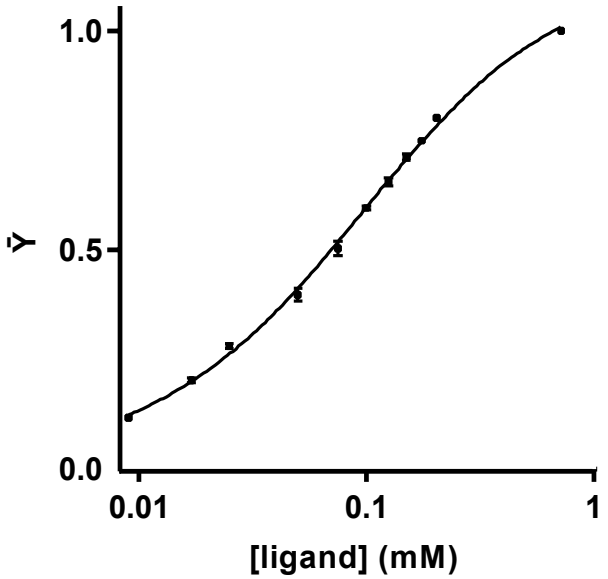



**Figure 3-2.** Intrinsic fluorescence quenching of SibS (3 μM) by **1.1**. Fluorescence emission intensity ($\lambda_{em}$ of 340 nm, $\lambda_{ex}$ of 290 nm) was measured during titration in 20 mM sodium phosphate pH 8.0 (25 °C). The fractional saturation ($\bar{Y}$) was calculated and the data was fit to an analysis described in the Experimental Procedures.

**Crystal structure of SibS with 1.1.** SibS belongs to the DAP epimerase superfamily according BLAST analysis. The substrate and function of most members of this superfamily have not been identified with a few exceptions known to catalyze prototropic shifts.[72] Among these, PhzF[62,63] is structurally related` to SibS[73] according to the superposition of their respective crystal structures (RMSD of 5.2 Å) (**Figure 3-3**). The active site for SibS was tentatively assigned at the interface of two subdomains from each monomer based on its location in the PhzF dimer. The oligomeric state of SibS was examined by gel filtration and appears to form a dimer in solution (observed molecular mass of 50.1 kDa, theoretical monomeric mass of 30.7 kDa predicted by ExPASy

ProtParam). Despite the structural similarity of SibS and PhzF, their sequence identity is low (21%) suggesting functional divergence between these enzymes. A crystal structure of SibS in complex with **1.1** could provide more information on the tentative active site. Specifically, this includes identifying residues that coordinate **1.1** to check if they are conserved in other members of the superfamily for binding and catalysis.

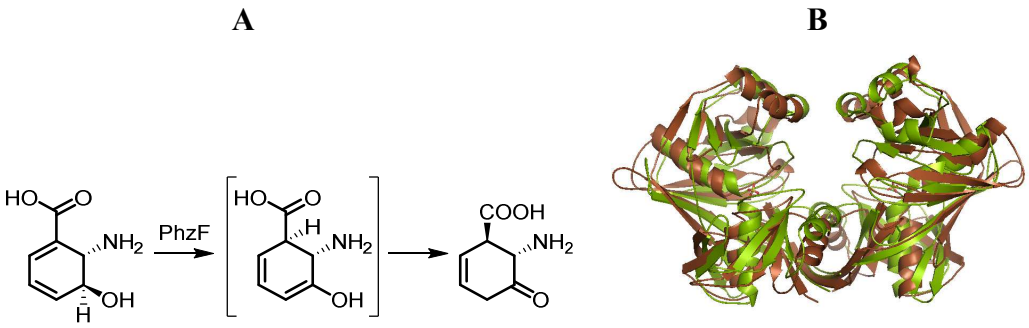



**Figure 3-3. A.** DAP-epimerase superfamily member PhzF catalyzes a pericyclic prototrophic shift followed by enol-keto tautomerization. **B.** Superimposed crystal structures of two SibS monomers, depicted in brown, obtained at 2.2 Å resolution and PhzF homodimer, depicted in green, with an RMSD of 5.2 Å based on an overlap of 462 $C^{\alpha}$ atoms in 598 total amino acids.

SibS crystals were transferred by Dr. Chuenchor into a soaking solution (1 mM DTT, 75 mM HEPES pH 7.5, 1.4 mM lithium sulfate and 15% glycerol) containing the ligand **1.1** (4.7 mM) for three hours. X-ray diffraction data was collected and processed by Dr. Chuenchor. Ligand density was observed at the proposed active site although it could not be refined to **1.1** (**Figure 3-4**) suggesting the ligand is dynamic. Instead, water molecules were built into the density of the ligand in order to identify the residues in proximity of the ligand.

The ligand coordinates to a catalytic residue conserved in DapF[74] and substrate binding residues conserved in PhzF[62,63] representing a new active site architecture in the DAP epimerase superfamily. Specifically, the ligand coordinates to C213 conserved in and used by DapF as a general acid. The ligand also coordinates to a pair of residues, H76 and

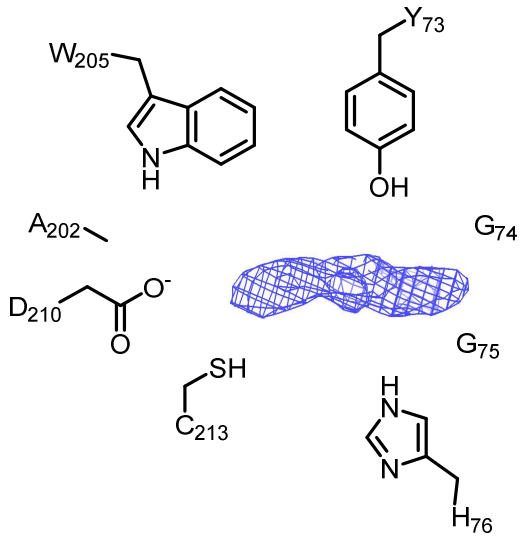



**Figure 3-4.** Schematic view of the SibS-**1.1** complex.

D210 that is conserved in PhzF. They are substrate binding residues, the latter of which is also essential for activating the substrate for PhzF prior to turnover. These residues are conserved in the predicted amino acid sequences of SibS homologs involved in the production of other dihydropyrrole and pyrrolidine moieties (**Figure 3-5**). They are also conserved in the amino acid sequence of a eukaryotic member of the superfamily (Aes1) although its substrate and biochemical function have yet to be identified.[75] Additional aromatic residues (Y73 and W205), a pair of glycines (G74 and G75) and A202 in SibS coordinate the ligand as well. These coordinating residues are conserved or mutated conservatively in the homologs except for G75. This glycine is replaced with a threonine in the homolog associated with anthramycin and porothramycin biosyntheses. The functional consequence of this new type of active site is unknown.

**Synthesis of SibZ's putative substrate 1.2.** SibS treatment of **1.1** did not form **1.2**, the proposed substrate for SibZ. If SibS requires an unidentified cofactor or protein partner to catalyze hydrolysis, the remaining part of the proposed pathway could still be correct. In order to test if SibZ catalyzes the SAM-dependent methylation of **1.2**, the proposed

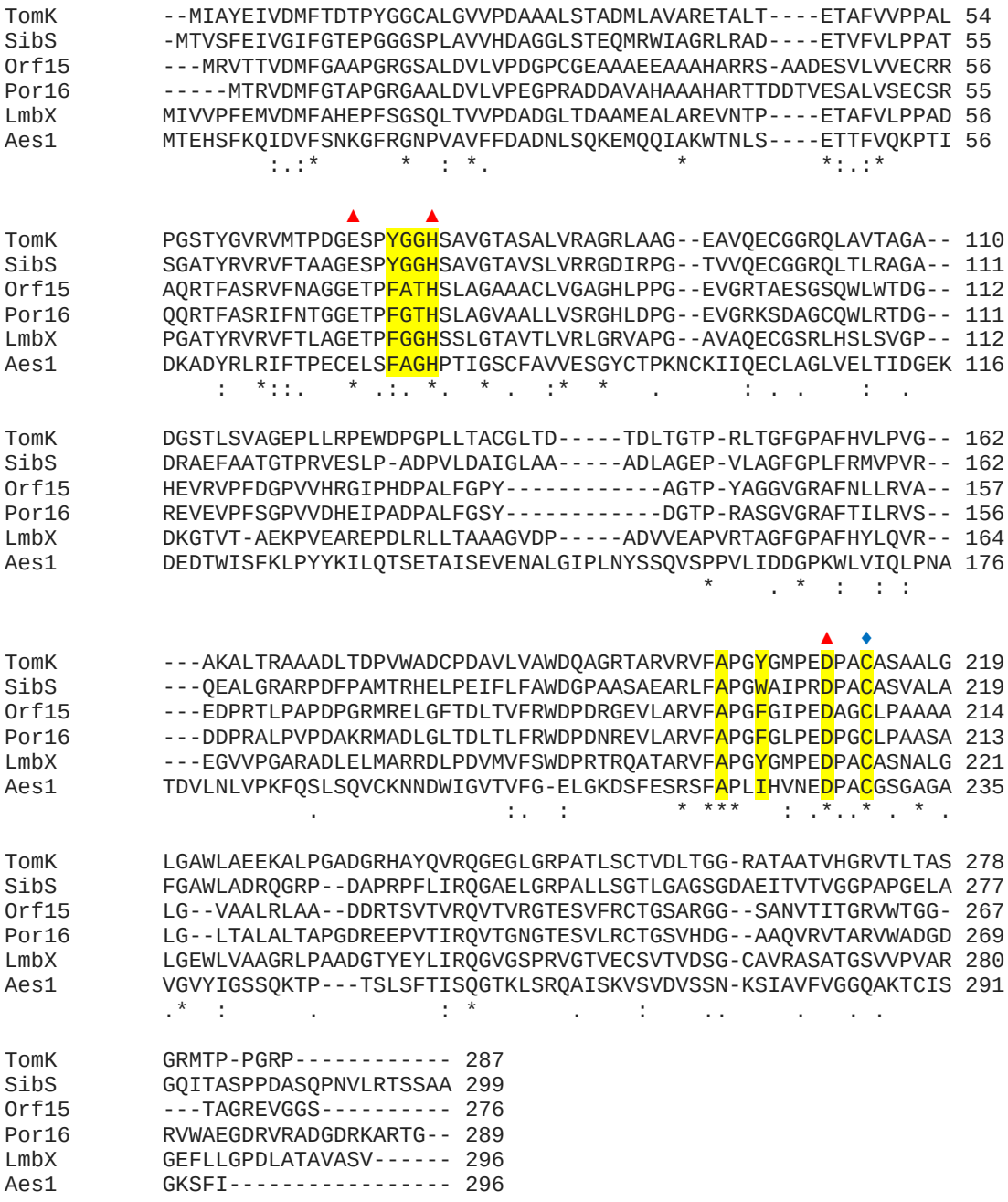
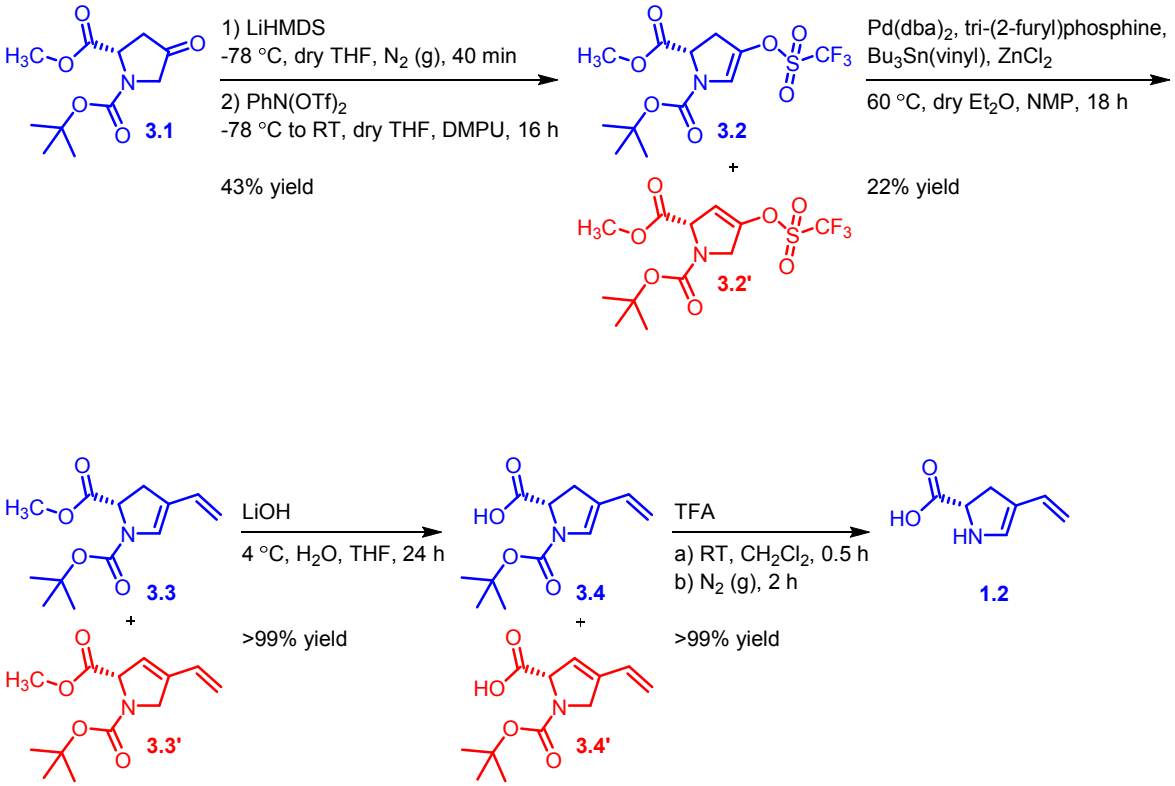



**Figure 3-5.** Sequence alignment of SibS with related proteins. The proposed binding and catalytic residues in SibS and its homologs based on the crystal structure of SibS in complex with **1.1** are highlighted. These residues include ones that are conserved in DAP epimerase superfamily members PhzF and DapF involved in binding (denoted with ▲) and catalyzing (denoted with ♦) their substrates, respectively. The alignment includes Aes1 (*Schizosaccharomyces pombe*) a eukaryotic member of the same superfamily with similar active site residues.

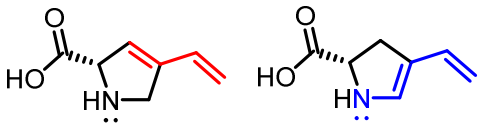
substrate was synthesized (**Scheme 3-6**). Briefly, enolization and triflation of **3.1** formed **3.2**. The enol triflate **3.2** was cross-coupled with tributytin-vinyl to install a vinyl group generating **3.3**. **3.3** was demethylated and deprotected to form the desired compound **1.2**. The product was assumed to be the desired enantiomer by literature precedent for related reactions.[76,77]




**Scheme 3-6.** Scheme for synthesizing **1.2**.

A pair of enol triflates formed during triflation as the starting material **3.1** possessed two sets of enolizable protons. This was consistent with observation of a duplicate set of signals by NMR spectroscopy (**Appendix B-2**). A $^{1}$H NMR spectrum of the starting material **3.1** coincidentally contained two sets of signals but these coalesced at a higher temperature (40 °C) suggesting the convergence of *cis* and *trans* amides into one rotamer. The isomers persisted through Stille cross-coupling (**Appendix B-4**) and

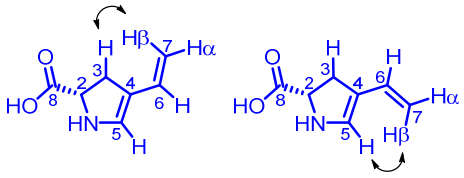
demethylation (**Appendix B-6**). In contrast, the last step produced only the desired isomer (**Appendix B-8**). The extended conjugated π-π system in **1.2** likely provides the driving force that fortuitously converts the undesired isomer to **1.2** instead of **1.2'** (**Scheme 3-7**). The removal of the electron withdrawing carbonyl group after boc deprotection allows the lone pair of electrons on the nitrogen to extend the conjugation formed between the vinyl and ring double bonds, an option only available in the desired isomer **1.2**.




**1.2'** **1.2**

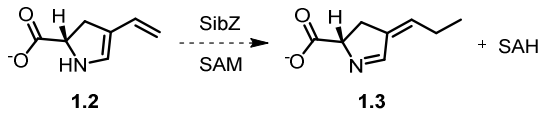
**Scheme 3-7.** Comparison of the conjugated π-π system in **1.2'** and **1.2** illustrated in red and blue, respectively.

The structures drawn throughout the synthetic scheme are consistent with NMR spectra ($^{1}H$, $^{13}C$, COSY, HSQC, HMBC) and MS (**Appendix B-3, B-5, B-7, B-9**) with the following exceptions. The amide carbonyl carbon signal (by $^{13}C$ NMR) and the C2-C4 correlation (by HMBC) were not observed in any of the compounds, a common issue with quaternary carbons. Two cross-peaks (H3-H7 and H5-H7) observed by NOESY (**Appendix B-8**) indicates that the product **1.2** was a mixture of geometric isomers (**Scheme 3-8**). This facile interconversion results from rotation around the single bond in the 1,3-butadiene group within **1.2**. For comparison, 1,3-butadiene has a rotational barrier of 5 kcal/mol at room temperature.[78]




**Scheme 3-8.** Two cross-peaks observed by NOESY consistent with the presence of geometrical isomers of **1.2**.

**SibZ treatment with 1.2.** SibZ (1 μM) was treated with **1.2** (86 μM) and SAM (166 μM) and assayed for methylation to **1.3 (Scheme 3-9)**. An HPLC based activity assay was used to monitor the formation of SAH, one of the expected products of the methylation of **1.2** where SAM acts as a methyl donor (**Scheme 3-9**). The reaction mixture was separated by HPLC for analysis at various times up to 24 h. The HPLC chromatograms each produced a signal ($t_r$ = 13 min) corresponding to SAH **(Figure 3-5)**. However, the signal integrated to a value equivalent to that of the corresponding signal produced by an equivalent reaction mixture without SibZ. This suggests that SibZ does not catalyze the methylation of **1.2** at rates above the control. A reaction mixture was spiked with SAH (86 μM) at the beginning of the reaction to indicate that it did not decompose over 24 h and could be detected via HPLC analysis.




**Scheme 3-9.** SibZ was expected to catalyze the methylation of **1.2** using SAM into **1.3** and SAH as part of sibiromycin biosynthesis.[31,32]

The addition of a metal cofactor may be required for catalytic activity. Many SAM-dependent methyltransferases use $Mg^{2+}$ as a cofactor to enhance methylation activity including one enzyme that methylates an α-keto acid substrate.[79] The reaction mixtures for SibZ were repeated in the presence of $Mg^{2+}$ (1 mM). However, HPLC analysis of the reaction mixtures again indicated that methylation did not occur at rates greater than the controls.

**SibZ activity in the absence and presence of SibS.** SibZ was unable to process **1.2** under the assay conditions suggesting that one of the precursors of **1.2** may instead be

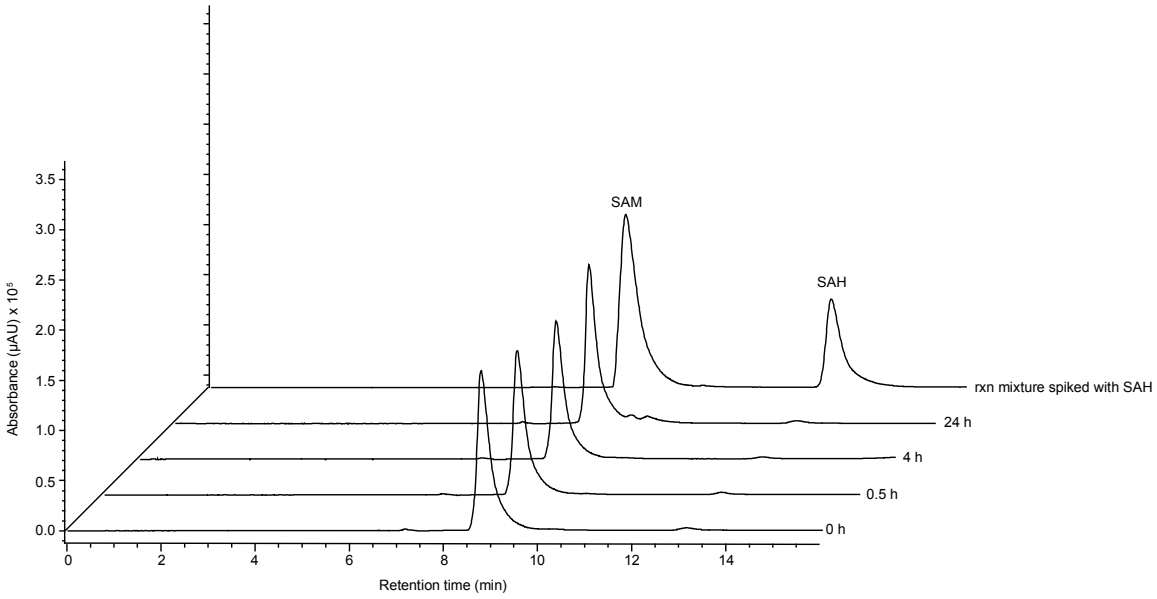



**Figure 3-6.** HPLC chromatogram monitoring SibZ-incubated with **1.2** and the methyl donor SAM ($t_r$ = 8.7 min) expected to form **1.3** and SAH ($t_r$ = 13.1 min) upon methylation. The reaction mixture was quenched at various times and subjected to HPLC separation. Chromatograms detected at 254 nm are shown. The back chromatogram depicts an equivalent reaction mixture spiked with SAH and incubated for 24 h as a positive control.

the substrate for SibZ. The four precursors of **1.2** (L-tyrosine, L-DOPA, L-2,3-*seco*dopa, **1.1**) were tested as substrates. L-2,3-*seco*dopa was generated *in situ* via SibV-catalyzed dioxygenation of L-DOPA. HPLC separation of each reaction mixture indicates a small signal for the product SAH. However, an equivalent signal was detected during the HPLC separation of an equivalent reaction mixture without SibZ. This indicates that the putative substrates failed to undergo methylation at rates greater than the controls.

The inactivity of both SibS and SibZ suggested each protein may require the other protein to form a functionally competent complex. Addition of SibS to the reaction mixture was analyzed by HPLC but did not appear to increase the SAH signal when compared with the equivalent reaction lacking SibZ.

**Unexpected ability of SibS to depurinate SAM.** The aforementioned studies that involved both SibS and SAM in the reaction mixture did produce a new compound detected by HPLC. Mixtures containing only SibS and SAM in buffer were monitored over time by HPLC and showed a time-dependent increase of a new product ($t_r$ = 4.3 min). The new product was collected and observed by $ESI^+$-MS to produce an ion with m/z of 136.0607. This suggests the formation of an $[M+H]^+$ ion of adenine with a theoretical mass of 136.0618 within a mass deviation of 8 ppm. Adenine likely results as a consequence of SibS-catalyzed deglycosylation of SAM. The $ESI^+$-MS of SAM itself also produces the same ion but this is accompanied by formation of other ions $[methylthioadenosine+H]^+$and $[adenosine-H_2O+H]^+$ all of which result via fragmentation. SibS activity is an example of depurinase activity observed almost exclusively in DNA repair with nucleotide substrates. No SAM-dependent enzymes have been reported to promote the same reaction. There is no obvious explanation of how depurination is a productive step in PBD biosynthesis suggesting it is a non-native function. This promiscuous activity may stem from the shared active site features of SibS with superfamily members such as Aes1. Depurination is a reasonable biochemical function of Aes1, a protein known to enhance antisense-RNA mediated gene silencing.[75]

**Attempt to express and isolate SibT.** SibS and SibZ either individually or together were unable to process a number of dihydropyrrole biosynthetic intermediates. If a multienzyme complex is necessary for SibS and SibZ to catalyze their respective reactions, SibT and SibY may be involved. These two proteins are the only predicted dihydropyrrole biosynthetic proteins conserved throughout the relevant natural products making them

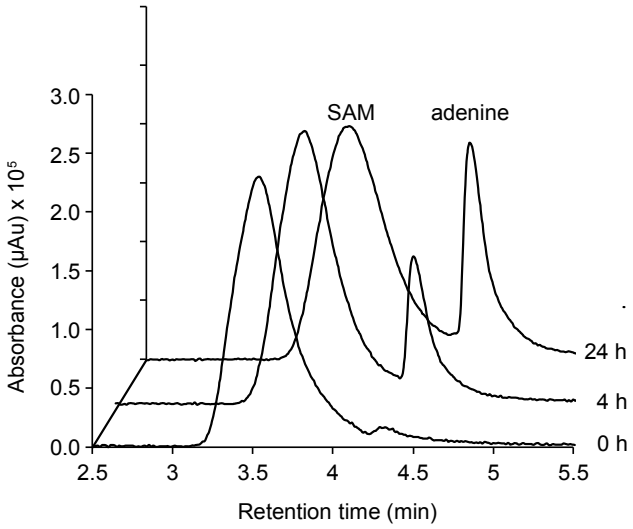



**Figure 3-7.** HPLC chromatogram monitoring SibS-catalyzed depurination of SAM ($t_r$ = 3.6 min) to form adenine ($t_r$ = 4.3 min). Chromatograms detected at 254 nm are shown.

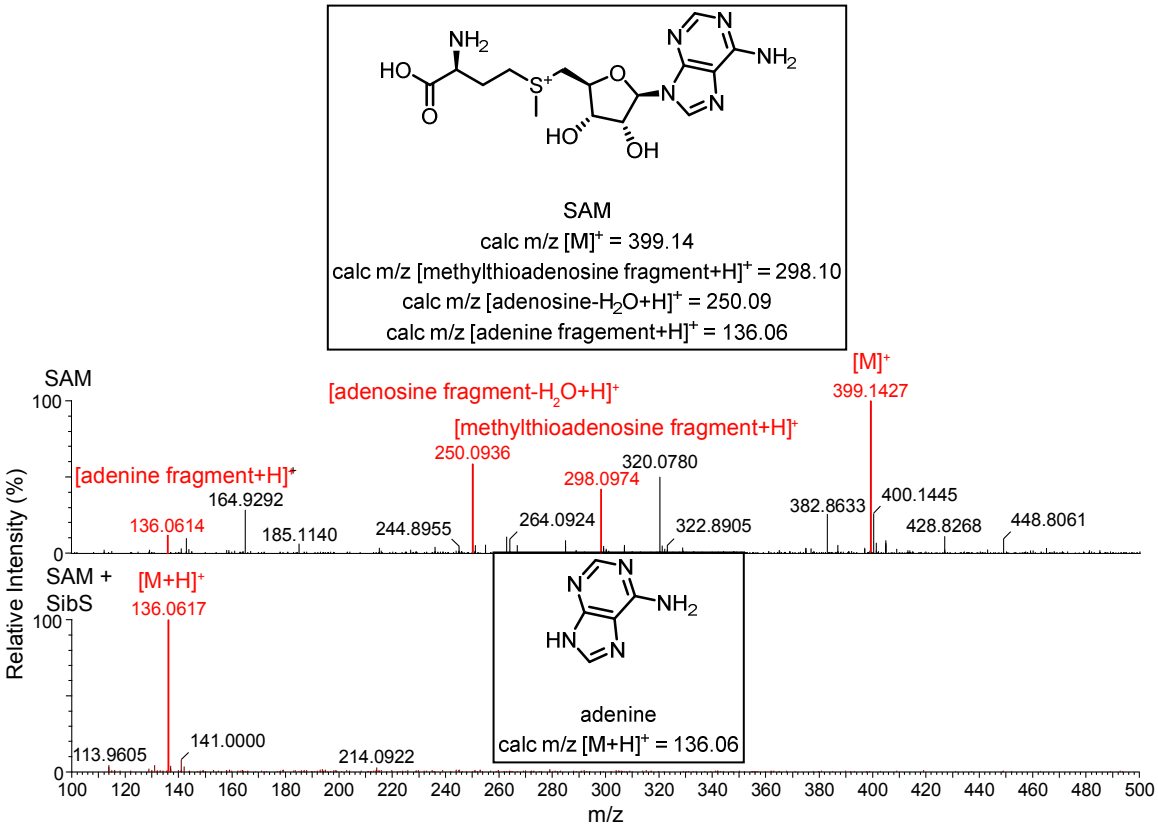



**Figure 3-8.** ESI$^+$-MS of the products formed after incubation of SAM reacted with SibS (bottom) compared with SAM in the reaction buffer (top).

available for forming a protein complex. SibT is expected to reduce **1.3** using a F420-$Glu_0$ cofactor derivative generated by SibY. To test if the absence of SibT impairs SibS and SibZ activity, the gene coding for SibT was cloned into a pSMT3 vector. BL21(DE3) *E. coli*, transformed with the vector, expressed SibT but in an insoluble form. The absence of its native cofactor may cause protein insolubility. Future attempts should include coexpression of SibT and SibY to see if this resolves the solubility issue.

**3.3 Summary**

SibS, TomK and SibZ were produced as soluble proteins. However, SibS and TomK incubated with **1.1** did not generate the expected hydrolyzed products **1.2** and oxalate in the absence or presence of a divalent metal ion. Hydrolysis of **1.1** was expected to proceed via tautomerization of **1.1** but neither SibS nor TomK promote tautomerization of **1.1**. SibS has affinity for **1.1** suggesting that either this ligand structurally resembles the substrate or is the substrate but requires an unidentified cofactor or protein for turnover. Surprisingly SibS did catalyze the depurination of SAM, likely a non-native reaction. The set of residues that coordinate **1.1** in the SibS crystal structure differs from the active site residues in functionally characterized DAP epimerase superfamily members suggesting SibS belongs to a new class of proteins within this superfamily.

SibZ was produced as a soluble protein and incubated with the chemically synthesized **1.2**. However, SibZ did not catalyze the methylation of **1.2** or its precursors L-tyrosine, L-DOPA, L-2,3-*seco*dopa and **1.1**. Addition of $Mg^{2+}$ or SibS did not activate SibZ-catalyzed methylation. To assess if SibS and SibZ form a functional complex with SibT, SibT was produced but it was insoluble.

### 3.4 Experimental Procedures

**Materials.** Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). *PfuTurbo*® DNA polymerase and its buffer were obtained from Stratagene (La Jolla, CA). Molecular biology grade DMSO was obtained from Sigma (St. Louis, MO). The QIAquick gel extraction kit was obtained from Qiagen (Valencia, CA). The GeneJet plasmid miniprep kit, dNTPs, T4 DNA ligase and its buffer were obtained from Fermentas (Glen Burnie, MD). SAM, restriction enzymes and their buffers were obtained from New England Biolabs (Ipswich, MA). The electrocompetent Rosetta2 strain was obtained from Novagen (Darmstadt, Germany). Oxalic acid dihydrate, barley seedling oxalate oxidase (0.71 U/mg), 3-methyl-2-benzothiazoline hydrazine hydrochloride monohydrate, *N,N*-dimethylaniline (99%), horseradish peroxidase P-8375 (250-330 U/mg), ethylenediamine tetraacetic acid tetrasodium salt dehydrate and sodium citrate hydrate were all obtained from Sigma. All other relevant materials are described in Chapter 2.

**General methods.** PCR was performed using an Eppendorf Mastercycler gradient (Hauppauge, NY). All plasmid constructs were confirmed by sequencing (UMBI or Genewiz). *E. coli* cells were lysed using an EmulsiFlex C3 homogenizer from Avestin (Ottawa, Canada). All other relevant general protocols are described in Chapter 2. Detection of oxalate using a coupled enzyme assay was performed according to the manufacturer's (Sigma) instructions. HPLC analysis employed an Agilent 1000 series pump coupled to an 1100 wavelength detector and Phenomex 250 x 4.6 mm Jupiter C18 analytical column with 5 μM particle size and 300 Å pore size (Torrance, CA). Fluorescence binding assays were conducted using a Hitachi F-4500 FL spectrophotometer generously provided by Dr. L. Isaacs at University of Maryland at College Park (UMCP)

(College Park, MD). The emission mode was selected in the FL solutions software and each scan was set to 2,400 nm/min with a 5 nm emission and excitation slot, 0.5 sec response time and 2 sec spin. THF and $Et_2O$ were dried over molecular sieves.

**Cloning of the *sibZ* and *sibT* genes.** The genes encoding SibZ and SibT were amplified from the pSuperSib2[31] cosmid using alternatively a forward primer of 5'-CCTTCTTGGATCCCGGGAATGGCAGGAC-3' and a reverse primer of 5'-CCTTCTTAAGCTTTCAATGGCGCTGGGTG-3' or a forward primer of 5'-AATTAATGGATCCCGGCACGGAGTCCTGATC-3' and a reverse primer of 5'-AATTATTAAGCTTTCATCCATCCCCCTGTAGGTG-3'. The PCR products were gel purified and digested with BamHI-HF and HindIII restriction enzymes. Each digested product was ligated with the linearized pSMT3 plasmid to generate pSMT3/*sibZ* or pSMT3/*sibT* alternatively.

**Expression of proteins.** Expression was carried out under similar conditions described in Chapter 2 except for the following changes. Starter cultures consisted of either BL21(DE3) transformed with pSMT3/*sibS*,[66] pSMT3/*tomK*[67] or pSMT3/*sibZ* or Rosetta2 transformed with pSMT3/*sibT*. An additional antibiotic, chloramphenicol (34 μg/mL), was added to Rosetta2 (DE3) cultures. The cell culture transformed with pSMT3/*sibS* was induced with D-lactose (0.2%) at 22 °C with agitation (12 h) while the cell culture transformed with pSMT3/*sibT* was induced with IPTG (50 μM) at 16 °C with agitation (15 h). The cell culture transformed with pSMT3/*sibZ* was induced under identical conditions as the one transformed with pSMT3/*sibV* (Chapter 2).

**Purification of proteins.** Purification of SibS, TomK and SibZ were carried out under similar conditions used for the purification of SibV described in Chapter 2 except

for the following changes. All purification steps were performed at 4 °C except for the chromatographic steps which were performed at room temperature while the buffers were kept on ice. TCEP (1 mM) was added to all the purification buffers instead of DTT. 10% glycerol was also included and the protease inhibitor benzamidine was excluded from all the purification buffers. Cells were lysed by four passages through an EmulsiFlex homogenizer at 15,000 psi. Proteins were concentrated in an Amicon ultra centrifugal filter (10,000 NMWL) (EMD Millipore) at 5,000 rpm. The resulting hydrolases were ≥ 95% pure as estimated by visual inspection after their separation by denaturing gel electrophoresis and staining with Coomassie Brilliant Blue. Enzyme solutions in gel filtration buffer were flash frozen in droplets with $N_2$ (l) and stored at -80 °C until needed.

**Solubility assessment of $His_6$-SUMO-SibT.** Cells expressing $His_6$-SUMO-SibT were lysed under identical conditions described above within the protein purification protocol. Protein solubility was analyzed after separation of the cell free extract by denaturing gel electrophoresis and staining with Coomassie Brilliant Blue.

**Determination of the oligomeric state of SibS.** Size exclusion chromatography to measure the oligomeric state of SibS in solution was carried out under identical conditions described in Chapter 2.

**Mass spectrometry of SibS.** Purified SibS was dialyzed (2 x 1 L) (12-14 kDa MWCO) against $H_2O$ for 4 h. The resulting desalted SibS was injected into a 100 mm x 2.1 mm Poroshell C3 column (14%-51% $CH_3CN$ gradient in 0.1% aqueous formic acid over 15 min, 150 µL/min). Intact protein mass measurement was obtained in a full scan MS mode using an LTP Orbitrap XL mass spectrometer with 60,000 resolution and *m/z* 400 in profile mode.

**HPLC assay to detect hydrolase activity.** SibS or TomK (1 μM) at 25 °C was incubated with **1.1** (1.0 mM) in sodium phosphate (25 mM pH 8.0) along with TCEP (1.0 mM). Either $MgCl_2$, $ZnCl_2$ or $MnCl_2$ (1.0 mM) was included in reaction mixtures testing for metal-dependence. Aliquots of the reaction mixture were quenched by addition of an equivalent volume of the HPLC elution buffer (aqueous 0.1% (v/v) formic acid pH 2.5) at various times (0, 2, 4, 6, 8, 10 and 24 h). Each sample was analyzed by HPLC as described in the general methods (30 min method, 1 mL/min). A negative control reaction was conducted excluding the protein.

**$^1$H NMR assay to detect SibS and TomK activity.** SibS and TomK (10 μM) were alternatively incubated at 25 °C with **1.1** (13 mM) in sodium phosphate (20 mM pH 8.0). The reaction mixtures were spiked with 10% $D_2O$ and monitored by $^1$H NMR spectroscopy for as long as 24 h. Negative control reactions were conducted by excluding either protein or substrate.

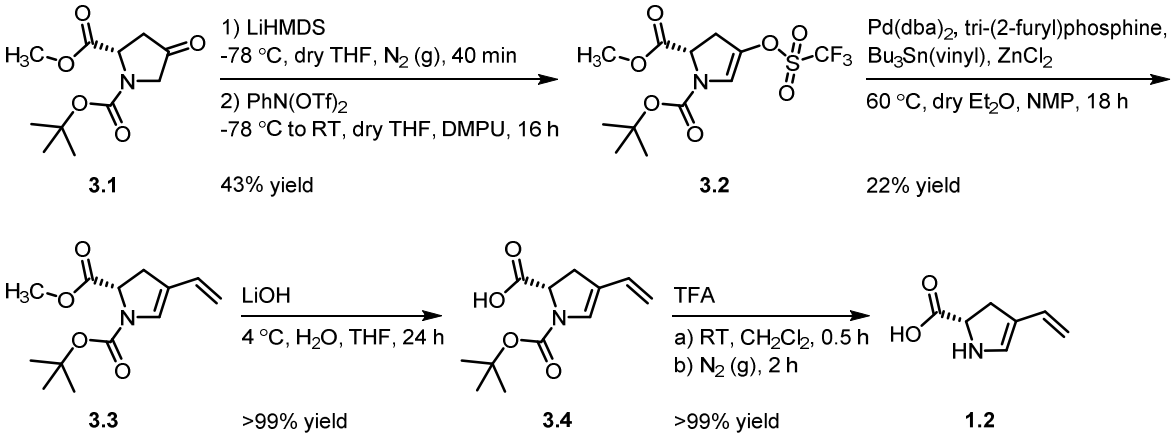
**Affinity of 1.1 for SibS.** SibS (3 μM) was titrated with **1.1** over a range of concentrations (9-717 μM) in sodium phosphate (20 mM pH 8.0) at 25 °C. The fluorescence emission intensity ($\lambda_{em}$ of 340 nm and $\lambda_{ex}$ of 290 nm) after each addition of ligand was recorded after the system reached equilibrium. An equivalent reaction where water was titrated as the ligand instead of **1.1** was used to account for changes in the fluorescence signals caused by dilution. An equation **(Equation 3-1)** that relates fluorescence emission intensity (F) to the endstate fluorescence emission intensities without ligand ($F_0$) and with saturating level of ligand ($F_b$), total ligand concentration ($[L]_T$), total protein concentration ($[P]_T$) and the equilibrium dissociation constant ($K_D$)

describes the simple binding process and was used to fit for $K_D$.[71] $\bar{Y}$, calculated using (**Equation 3-2**), was plotted against $[L]_T$ on a logarithmic scale (**Figure 3-2**).

**Equation 3-1.** $$F = F_o - (F_o - F_b)\frac{([P]_T+[L]_T+K_D)-\sqrt{([P]_T+[L]_T+K_D)^2-(4[P]_T[L]_T)}}{2[P]_T}$$

**Equation 3-2.** $$\bar{Y} = \frac{F_0-F}{F_o-F_b}$$

**Synthesis of 4-vinyl-4,5-dehydro-L-proline 1.2.**




***N*-Boc-4-triflate-4,5-dehydro-L-proline methyl ester 3.2 and *N*-Boc-4-triflate-3,4-dehydro-L-proline methyl ester 3.2' (Scheme 3-6).** Synthesis of the enol triflate was based on protocols reported previously.[77,80] A cold (-78 °C) solution of *N*-Boc-4-oxo-L-proline methyl ester **3.1** (204 mg, 0.84 mmol) dissolved in THF (5 mL) was added to a cold (-78 °C) solution of anhydrous LiHMDS (1 M in THF, 1.26 mL, 1.26 mmol) diluted in THF (5 mL). The reaction mixture was purged with $N_2$ (g) and stirred for 40 min at -78 °C. A cold (-78 °C) solution of $PhN(OTf)_2$ (370 mg, 0.84 mmol) dissolved in DMPU (302 µL, 2.5 mmol) and THF (5 mL) was injected into the reaction mixture and was stirred for 16 h during which the reaction temperature gradually increased from -78 °C to room temperature. The solvent in the reaction mixture was evaporated under reduced pressure and the remaining residue was dissolved in EtOAc (10 mL). The organic solution was washed with saturated $NaHCO_3$ (10 mL), $H_2O$ (10 mL) and brine (10 mL). The aqueous

fractions were combined and back extracted with EtOAc (5 mL). The organic fractions were combined, dried under $MgSO_4$ and evaporated under reduced pressure to yield a yellow residue. The residue was purified by silica gel column chromatography using 7.5% EtOAC and 7.5% $CH_2Cl_2$ in hexanes. The fractions containing desired material were evaporated under reduced pressure to yield a pale yellow oil (135 mg, 43% yield). $R_f$ 0.3 (7.5% EtOAC and 7.5% $CH_2Cl_2$ in hexanes). $^1H$ NMR (400 MHz, $CDCl_3$) δ 5.75 and 5.71 (s(2), 1H), 5.08 and 5.01 (t(2), 1H), 4.47-4.22 (m, *J*=14, 6.3, 2H), 3.80 and 3.78 (s(2), 3H), 1.49 and 1.44 (s(2), 9H). $^{13}C$ NMR ($CDCl_3$) δ 169.6 and 169.3, 146.3 and 145.8, 111.3 and 111.1, 81.5 and 81.4, 63.7 and 63.1, 52.8 and 52.6, 50.3 and 50.0, 28.3 and 28.2. HRMS (FAB): *m/z* 376.0676 $(M+H)^+$. Calcd for $C_{12}H_{16}F_3NO_7S$ $(M+H)^+$: 376.0672.

***N*-Boc-4-vinyl-4,5-dehydro-L-proline methyl ester 3.3 and *N*-Boc-4-vinyl-3,4-dehydro-L-proline methyl ester 3.-3' (Scheme 3-6).** Synthesis of the vinyl compound, specifically the coupling conditions used, was based on a protocol reported previously.[76] $Pd(dba)_2$ (4 mg, 7 μmol), tri-(2-furyl)phosphine (3.4 mg, 15 μmol), $ZnCl_2$ (34 mg, 0.25 mmol) and $Et_2O$ (245 μL) was added to a solution of *N*-Boc-4-triflate-dehydro-L-proline methyl esters **3.2** and **3.2'** (49 mg, 0.13 mmol) dissolved in anhydrous NMP (6 mL). The reaction mixture was purged with $N_2$ (g) and stirred for 10 min at room temperature. $Bu_3Sn$(vinyl) (40 μL, 0.14 mmol) was injected into the reaction mixture and stirred for 18 h at 60 °C under reflux. The reaction mixture was diluted in EtOAc (45 mL) and washed with $H_2O$ (2 x 45 mL) and brine (60 mL). The aqueous fractions were combined and back extracted with EtOAc (4 x 5 mL). The organic fractions were combined, dried under $MgSO_4$ and evaporated under reduced pressure to yield a yellow oil. The oil was dissolved in $CH_3CN$ (20 mL) and washed with petroleum ether 40-60 (2 x 20 mL). The petroleum

ether was back extracted with $CH_3CN$ (3 mL). The $CH_3CN$ fractions were combined and evaporated under reduced pressured to a dark yellow oil. The oil was purified by silica gel column chromatography using 7.5% EtOAC and 7.5% $CH_2Cl_2$ in hexanes. The solvent in the desired material was evaporated under reduced pressure to yield a clear oil (20 mg, 22% yield). $R_f$ 0.2 (7.5% EtOAC and 7.5% $CH_2Cl_2$ in hexanes). $^1H$ NMR (400 MHz, $CDCl_3$) δ 6.50 and 6.38 (m, 1H), 5.68 and 5.63 (s(2), 1H), 5.32-4.98 (m, 2H), 5.09 and 5.01 (s(2), 1H), 4.45-4.24 (m, 2H), 3.74 (s(2), 3H), 1.51 and 1.44 (s(2), 9H). $^{13}C$ NMR ($CDCl_3$) δ 171.0, 140.9, 130.0 and 129.8, 122.1 and 121.8, 118.2 and 117.7, 80.3, 66.8 and 66.4, 52.3 and 52.2, 52.2 and 52.0, 28.4 and 28.3. HRMS (FAB): *m/z* 254.1389 $(M+H)^+$. Calcd for $C_{13}H_{20}NO_4$ $(M+H)^+$: 254.1387.

***N*-Boc-4-vinyl-4,5-dehydro-L-proline 3.4 and *N*-Boc-4-vinyl-3,4-dehydro-L-proline 3.4' (Scheme 3-6).** *N*-Boc-4-vinyl-L-proline methyl esters **3.3** and **3.3'** (14 mg, 0.05 mmol) were dissolved in THF (4.8 mL) and treated with LiOH (1 M in $H_2O$, 1.6 mL). The reaction mixture was stirred for 24 h at room temperature. THF was evaporated under reduced pressure and the resulting white solid and aqueous solution was acidified with glacial acetic acid (58 μL, 1.00 mmol). The mixture above was diluted with $CH_3CN$ (0.1 mL) and subjected to solid-phase extraction using a Sep-Pak that was preconditioned by washing with acetonitrile (9 mL) and $H_2O$ (9 mL) successively. Bound material was eluted with 5% $CH_3CN$ in $H_2O$ (6 mL) and 75% $CH_3CN$ in $H_2O$ (18 mL). The desired material was lyophilized to yield a clear oil (13 mg, 99% yield). $^1H$ NMR (400 MHz, MeOD) δ 6.55-6.44 (m, 1H), 5.83-5.75 (m, 1H), 5.23-5.08 (m, 2H), 4.86 and 4.83 (s(2), 1H), 4.44-4.19 (m, 2H), 1.49 and 1.46 (s(2), 9H). $^{13}C$ NMR (MeOD) δ 177.3 and 177.2, 139.4, 131.9, 127.1 and 127.0, 116.6, 81.1, 71.4, 53.0, 28.9 and 28.7. HRMS ($ESI^+$): *m/z* 262.1055

$(M+Na)^+$. Calcd for $C_{12}H_{18}NO_4$ $(M+Na)^+$: 262.1050. $\lambda_{max}$ = 243 nm (diode array detector, 64% $CH_3CN$).

**4-Vinyl-4,5-dehydro-L-proline 1.2.** *N*-Boc-4-vinyl-dehydro-L-prolines **3.4** and **3.4'** (13 mg, 0.05 mmol) were dissolved in $CH_2Cl_2$ (0.1 mL) and treated with neat TFA (0.1 mL). The reaction mixture was stirred for 30 min in darkness at room temperature before it was placed under $N_2$ (g) for 2 h to yield a yellow oil. The oil was dissolved in MeOH (4 mL) and washed with hexanes (3 x 2 mL). MeOH was removed by evaporation under reduced pressure to yield a yellow oil (7 mg, 99% yield). $^1$H NMR (400 MHz, MeOD) δ 6.59 (dd, *J*=18, 11, 1H), 5.97 (s, 1H), 5.39 (d, *J*=11, 1H), 5.31 (d, *J*=18, 1H), 5.12 (s, 1H), 4.29 (dd, *J*=15, 2H). $^{13}$C NMR (MeOD) δ 139.3, 130.1, 123.1, 120.3, 68.4, 52.1. HRMS ($ESI^+$): *m/z* 140.0711 $(M+H)^+$. Calcd for $C_7H_{10}NO_2$ $(M+H)^+$: 140.0706. $\lambda_{max}$ = 243 nm (diode array detector, 8% MeOH).

**SAM-dependent methyltransferase activity.** SibZ (1 μM) was incubated at 25 °C with the putative substrate (86 μM) and the cosubstrate SAM (166 μM) in sodium phosphate (20 mM pH 8.0) along with TCEP (1.0 mM). $MgCl_2$ (1.0 mM) was included when assaying for metal-dependence. An aliquot of the reaction mixture was quenched by addition of TFA (10% v/v) at various times (0, 0.5, 4 and 24 h). 12 μL of the quenched solution was mixed with an equivalent volume of the HPLC elution buffer (9% aqueous MeOH and 0.01% (v/v) TFA pH 2.2) and analyzed by HPLC as described in the general methods (30 min method, 0.5 mL/min, 254 nm detection). Positive control reactions were conducted by including SAH (86 μM) while negative control reactions were conducted by excluding SibZ.

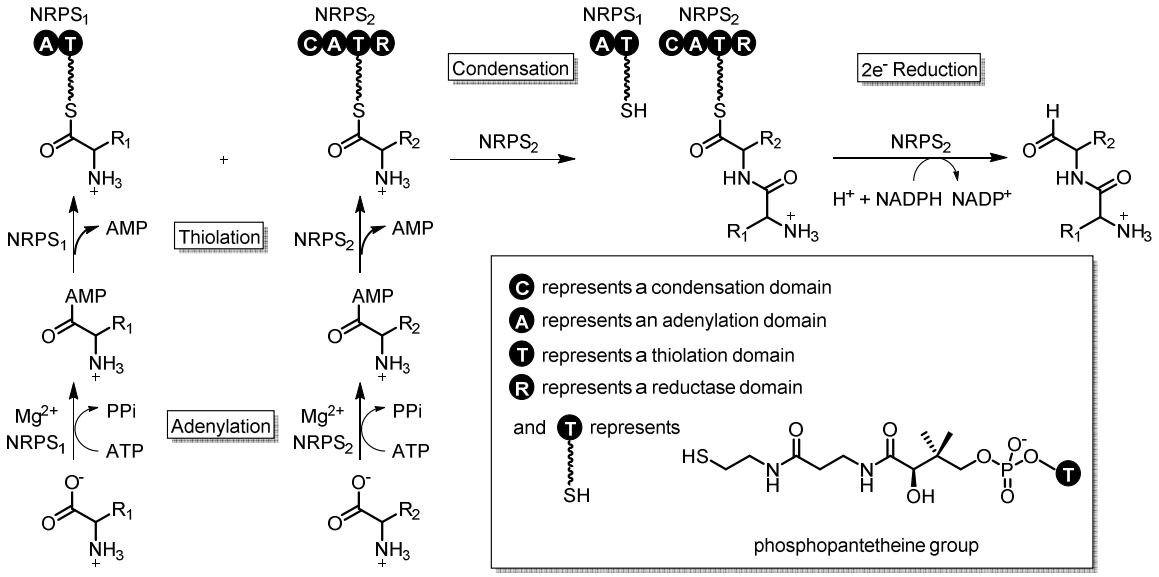
**HPLC assay to detect depurinase activity of SibS.** SibS (2.9 μM) was incubated at 25 °C with SAM (1.0 mM) in sodium phosphate (25 mM pH 8.0) along with TCEP (1.0 mM). An aliquot of the reaction mixture was quenched by addition of an equivalent volume of the HPLC elution buffer (aqueous 0.1% (v/v) formic acid pH 2.5) at various times (0, 2, 4 and 24 h). Each sample was analyzed by HPLC as described in the general methods (30 min method, 0.5 mL/min, 254 nm detection). A negative control reaction was conducted by replacing SibS with denatured SibS (heated to 98 °C for 2 min).

# Chapter 4: *In Vitro* Reconstitution of SibD's Adenylation and Thiolation Activities

## 4.1 Introduction

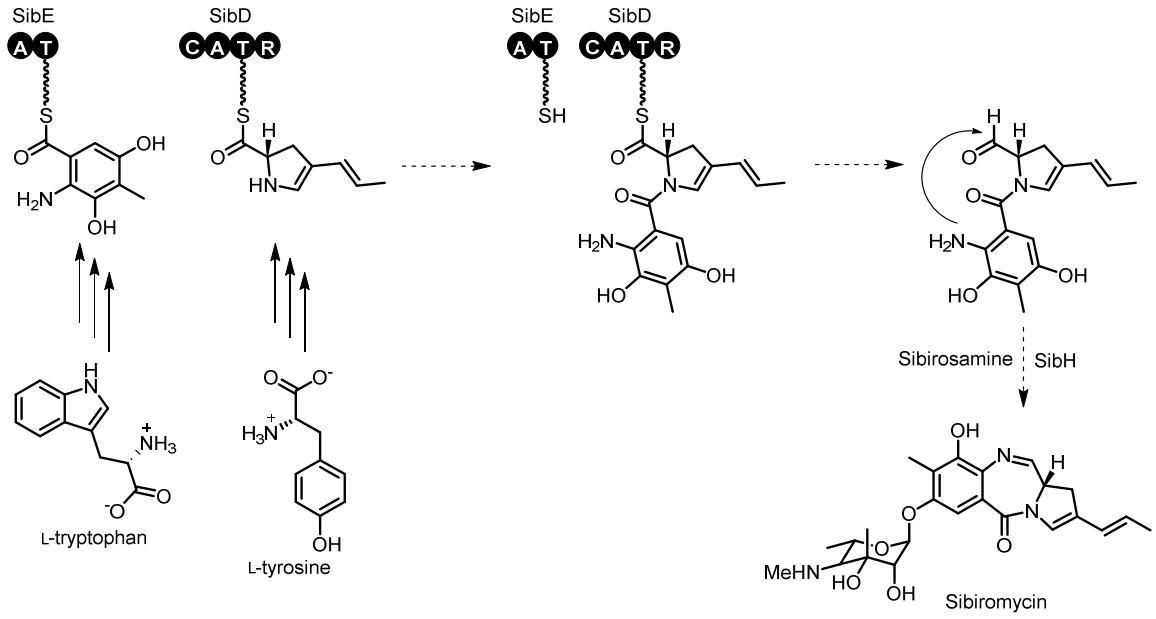
Understanding PBD biosynthesis involves studying not only the formation of the dihydropyrrole moiety but also its incorporation into the natural product. The PBD gene cluster contains a pair of biosynthetic genes predicted[31] to code for non-ribosomal peptide synthetases (NRPSs), a class of enzymes known to assemble moieties into natural products. NRPSs are large enzymes equipped with an assembly line of domains that link amino acid or carboxylic acid substrates into complex natural products. Before discussing specifically how the predicted NRPSs can assemble a PBD, it is necessary to understand the typical NRPS enzymatic steps (**Scheme 4-1**). It involves an assembly process that initiates with an $Mg^{2+}$-dependent adenylation domain selecting and activating an acid substrate using ATP. Adenylation forms an acyl adenylate, the substrate for a thiolation domain. The thiolation domain forms a thioester bond between its post-translationally loaded phosphopantetheine arm on a conserved serine and the acyl adenylate. A condensation domain forms a peptide bond between two thioester compounds each conjugated to a thiolation domain. The resulting peptide remains covalently tethered onto a single thiolation domain. The adenylation, thiolation and condensation steps repeat to grow the peptide chain until either a thioesterase or reductase domain terminates the assembly process by cleaving the peptide from the thiolation domain.

Sibiromycin is proposed to assemble in accordance with the NRPS paradigm (**Scheme 4-2**) using NRPSs SibE and SibD.[31] SibE contains two domains, an adenylation




**Scheme 4-1.** NRPS-mediated reactions with two amino acid substrates via their activation, thiolation and condensation to yield a dipeptide that is then reductively cleaved from the NRPS.

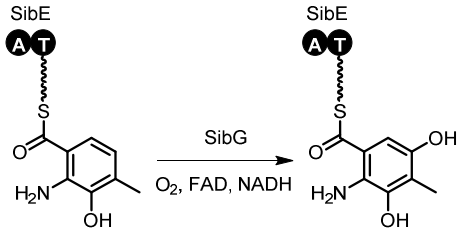
and thiolation domain. *In vitro* adenylation and thiolation assays showed that SibE adenylates an anthranilate compound derived from L-tryptophan and attaches the activated anthranilate onto its thiolation domain in preparation for its assembly into sibiromycin.[24] SibD and its homologs associated with related natural products each contain a condensation, adenylation, thiolation and reductase domain based on the presence of their respective signature motifs.[81] The following steps were proposed[31] taking into account the available domains and the structure of sibiromycin. SibD adenylates the dihydropyrrole moiety derived from L-tyrosine and the resulting activated dihydropyrrole forms a thioester linkage with the thiolation domain. The condensation domain uncouples the anthranilate compound from SibE and couples it to the dihydropyrrole moiety via a peptide bond. The reductase domain releases the dipeptide from SibD via a two-electron reduction using NADPH consistent with a Rossman fold present in the reductase, a signature nucleotide binding motif. An amino group on the anthranilate ring is positioned to attack the aldehyde




**Scheme 4-2.** SibE and SibD are proposed to activate, thiolate and condense the anthranilate and dihydropyrrole moieties derived from L-tryptophan and L-tyrosine, respectively. SibD likely reductively releases the dipeptide that undergoes a ring closure. Putative glycosyltransferase SibH catalyzes the transfer of the sibirosamine sugar onto the tricyclic aglycone to yield sibiromycin.

in the dihydropyrrole ring to form the diazepine ring. Putative glycosyltransferase SibH adds a sibirosamine sugar onto the tricyclic sibiromycin aglycone furnishing the PBD. The last two steps were proposed to occur post NRPS assembly of the dipeptide although the order and timing of these steps are speculative.

The typical NRPS steps are sometimes interrupted by additional modifications to the compound while it is loaded on the thiolation domain. For instance, SibG catalyzes the FAD/NADH-dependent hydroxylation of an anthranilate moiety precursor, 3-hydroxy-4-methylanthranilic acid, only once it is covalently attached to SibE (3-hydroxy-4-methylanthranilic acyl-*S*-SibE) (**Scheme 4-3**).[24,82] In analogy, SibD may activate and thiolate the final dihydropyrrole moiety or one of its precursors. The latter scenario would warrant a revision to the proposed dihydropyrrole biosynthetic pathway **(Scheme 1.6)** as adenylation and thiolation would have to precede steps modifying the dihydropyrrole.




**Scheme 4-3.** SibG-catalyzed hydroxylation of 3-hydroxy-4-methylanthranilic acyl-*S*-SibE.

Deletion of *orf22* and *npsB*, the *sibD* orthologs in the anthramycin and tilivalline producers, respectively, led to mutant strains that did not produce the expected PBD,[23,34] suggesting in analogy that SibD is also essential in sibiromycin biosynthesis. We chose to monitor SibD's adenylation and thiolation reactions *in vitro* to gain insight into the timing of these NRPS reactions relative to the dihydropyrrole modification steps. Answering this question involved production of an adenylation and thiolation didomain from SibD and testing each domain for activity using a panel of dihydropyrrole pathway intermediates as substrates.

## 4.2 Results and Discussion

**Production of the adenylation and thiolation didomain in SibD.** Amplification of the *sibD* gene (4512 bp) from cosmid DNA produced only truncated PCR products even under a condition that amplified a similarly sized (3252 bp) NRPS-encoded gene.[83] Although a synthetic gene could have been used in place of the PCR product, efforts to pursue heterologous expression of SibD were abandoned as solubility issues were anticipated due to the large size (1503 amino acids) of the protein. Expression of specific domains within the NRPS is a common strategy to overcome the lack of solubility of intact NRPSs. This requires identification of domain boundaries.

Many programs predict domain boundaries but require structural information of

related proteins.[84] No structures for multidomain intact NRPSs similar to SibD appear in the PDB[85] limiting the choice of methods to *ab initio* methods. Among these, the Udwary-Merski algorithm[86] (UMA) was a promising program as it extracts information from homologous sequences, available for SibD (TomB, Orf22, Por21, NpsB). Although initially developed to predict domain boundaries of polyketide synthases, another class of multidomain enzymes that assemble natural products, UMA has successfully predicted domain boundaries in NRPSs.[87] UMA defines boundaries by analyzing homologous sequences, primary sequence neighboring hydrophobicity and secondary structure of multidomain proteins. Low UMA scores for certain amino acids (504E, 1024R, 1123R, 1182A) in SibD predict they belong to a linker region (**Figure 4-1**). The predicted domains encompassed by the linker regions were verified by the presence of signature motifs of each domain.[81] This analysis indicated that 1182A does not belong to a linker region as it is between the R2 and R3 motifs in the reductase domain.

PCR primers were designed to anneal the nucleotides that code for the linker regions in SibD prior to the adenylation domain and past the thiolation domain. A truncated PCR product formed instead of the desired gene fragment coding for the adenylation and thiolation domains in SibD (*sibD*$_{AT}$). Two sets of PCR primers were used instead to amplify the gene fragment as two overlapping pieces and were cloned sequentially into the pSMT3 vector. This allowed for the production of SibD$_{AT}$ fused to an *N*-terminal His$_6$-SUMO tag.

The post-translation phosphopantetheinylation of the thiolation domain is necessary to render SibD$_{AT}$ fully functional. The sibiromycin gene cluster lacks a gene that codes for an enzyme that installs the phosphopantetheine arm on the thiolation domain. An *E. coli* strain modified to coexpress a promiscuous phosphopantetheinyl transferase (Sfp),

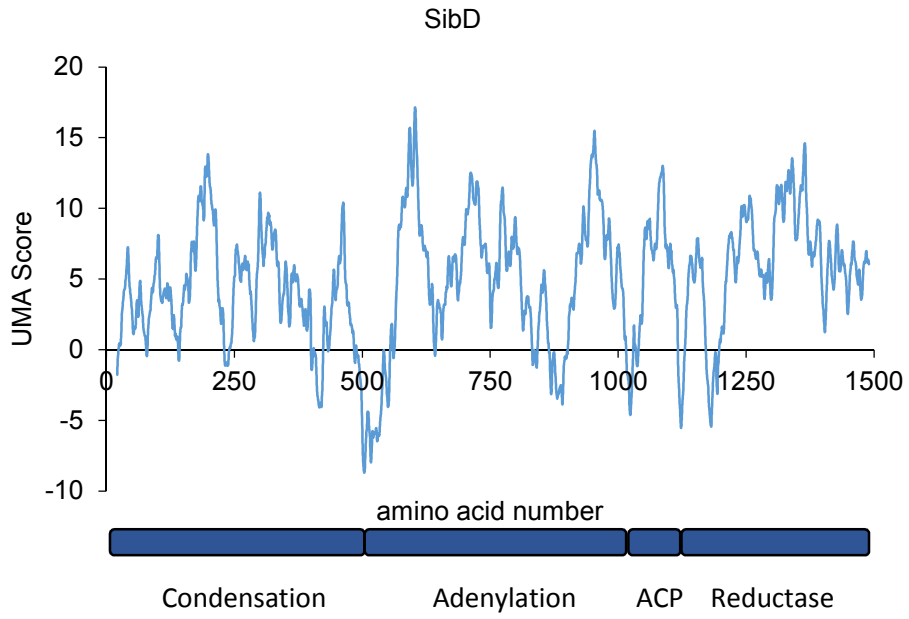


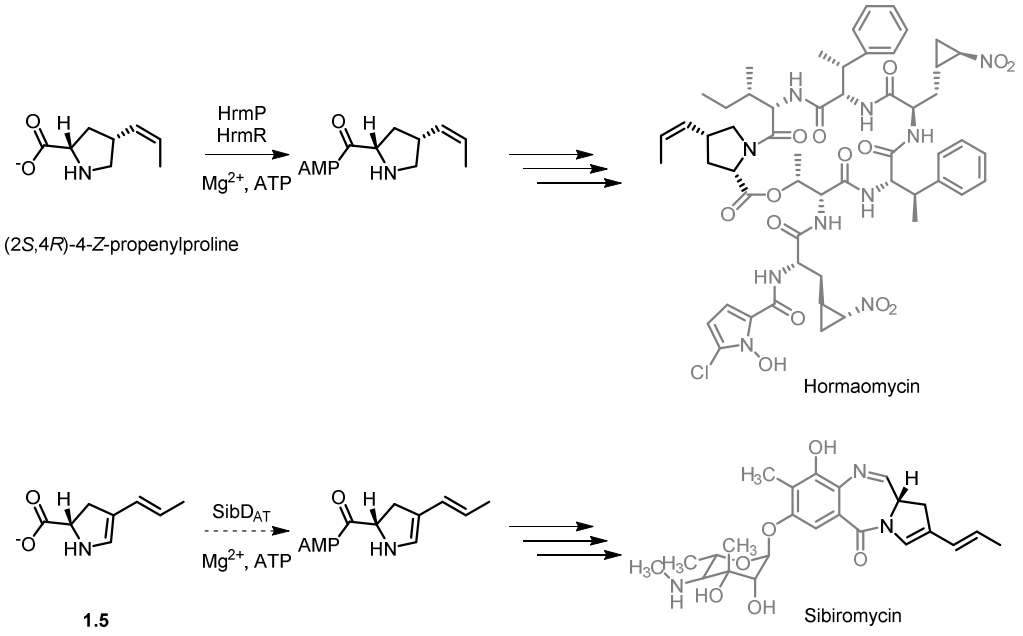

**Figure 4-1**. UMA prediction graph for SibD domain boundaries illustrated at low UMA scores. The thiolation domain is commonly referred to as the acyl carrier protein (ACP).

BAP-1,[88] was selected to express $SibD_{AT}$ and load the phosphopantetheine arm on the thiolation domain. Upon expression, the post-translational modification was assumed to have occurred at the only conserved serine (S1067 in SibD) in the thiolation domains of SibD and related NRPSs. $SibD_{AT}$ was purified via Ni-NTA affinity chromatography, subjected to Ulp1-catalyzed removal of the SUMO fusion and further purified via size exclusion chromatography. Purification yielded *ca*. 8 mg protein per liter of growth media (**Appendix C-1**). $SibD_{AT}$ was denatured, digested with trypsin and subjected to tandem MS analysis. Identification of the predicted peptides verified the presence of the desired post-translational modification at S1067 (**Appendix C-2**). The peptide of interest possessing the phoshopantetheine arm had a parent ion experimental mass of 3477.7141 Da in agreement with its theoretical mass of 3477.6296 Da within a mass deviation of 24 ppm. Identification of this peptide was further confirmed by detection of several of its daughter ions generated by fragmentation at amide C-N bonds extending from the amino

terminus (b ions) or carboxyl terminus (y ions) of the peptide (**Appendix C-3 and C-4**). For example the y23 ion containing the phosphopantetheinylated serine (y23*) had an experimental mass of 2584.1799 Da in agreement with its theoretical mass of 2584.1050 Da within a mass deviation of 29 ppm. The parent ion corresponding to the *apo* version of the peptide lacking the phosphopantetheine arm was not observed. These results indicate the successful expression and purification of the phosphopantetheinylated $SibD_{AT}$ didomain.

**Adenylation by $SibD_{AT}$.** With *holo*-$SibD_{AT}$ in hand, the next step was to determine the viability and substrates of the adenylation and thiolation domains. Since the final dihydropyrrole compound (2*S*,4*R*)-4-(*Z*)-propenylproline is incorporated into the structurally related natural product hormaomycin by action of its appropriate adenylase in the NRPS HrmP,[89] the final dihydropyrrole compound in sibiromycin, L-4-propenyl-4,5-dehydroproline **1.5,** is proposed to be the substrate for $SibD_{AT}$ (**Scheme 4-4**). This compound was synthesized under identical conditions that generated L-4-vinyl-4,5-dehydroproline **1.2** (Chapter 3) except propenyltributyltin was used as the coupling reagent. Although this substitution from its vinyl counterpart was tolerated by Stille coupling, the yield dropped to 11%. It is unknown if less starting material reacted, more side products formed or more product decomposed although the same number of compounds were present in the vinylation reaction as judged by UV irradiation and ninhydrin treatment of thin layer chromatography (TLC) plates spotted with the crude reaction.

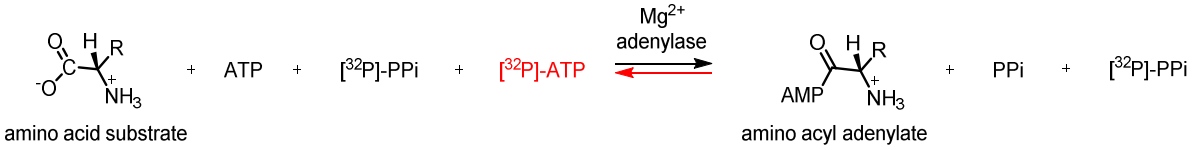
A radioactivity based activity assay routinely used to detect the reversible exchange of [$^{32}$P]PPi with ATP by capturing [$^{32}$P]ATP onto charcoal was used to indirectly monitor




**Scheme 4-4.** Adenylation reactions known or proposed to initiate incorporation of the pyrrolidine or dihydropyrrole moiety in hormaomycin and sibiromycin, respectively.

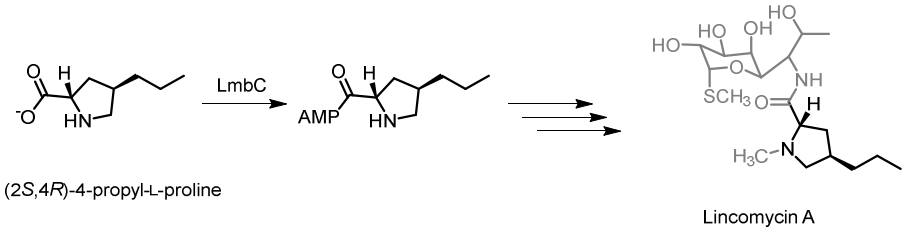
formation of the amino acyl-adenylate (**Scheme 4-5**). A reaction mixture consisting of $SibD_{AT}$ (0.25 μM), putative substrate L-4-propenyl-4,5-dehydroproline **1.5** (0.8 mM), cosubstrate ATP (3.7 mM) and $Mg^{2+}$ (7.5 mM) in the presence of PPi (1.0 mM) and [$^{32}$P]PPi (1 μCi) did not generate [$^{32}$P]ATP above background counts (**Figure 4-2**). The conditions were similar to those used previously[24] to detect adenylation by SibE suggesting either the substrate was incorrect or $SibD_{AT}$ was inactive.

One of the precursors of L-4-propenyl-4,5-dehydroproline **1.5** may instead be the adenylase substrate. Lincomycin A established a precedent for a precursor ((2*S*,4*R*)-4-propyl-L-proline) to be incorporated into the natural product[90] prior to an *N*-methylation step that furnishes its final pyrrolidine moiety (**Scheme 4-6**).[91] Among the proposed sibiromycin dihydropyrrole intermediates in the pathway, the commercially available ones (L-tyrosine and L-DOPA), enzymatically accessible one (L-4-(2-oxobutenoate)-4,5-dehydroproline **1.1**) and synthetically accessible one (L-4-vinyl-4,5-dehydroproline **1.2**)




**Scheme 4-5.** Radioactive adenylation assay that monitors amino acyl-adenylate formation indirectly though the reversible exchange of [$^{32}$P]PPi with ATP followed by the capture of the resulting [$^{32}$P]ATP onto charcoal.

were tested as substrates. Only L-tyrosine was adenylated by $SibD_{AT}$ generating [$^{32}$P]ATP above background counts (**Figure 4-2A**). Often an adenylase can activate some common amino acids in addition to its native substrate. For example the SibD homolog HrmP activates L-threonine at 10% of the efficiency of its native substrate.[90] L-threonine, an amino acid with no obvious place in the pathway for sibiromycin was tested as a substrate. Using L-threonine, equivalent counts of [$^{32}$P]ATP formed within error (680 ± 253 cpm) compared with L-tyrosine (589 ± 224 cpm). This suggests that L-tyrosine may also be a non-native substrate. For comparison, much higher counts (30962 ± 9064 cpm) are observed when monitoring the adenylation of L-serine, the native substrate for another NRPS NocB(A4T4) and its partner protein NocI (**Figure 4-2B**).[99]




**Scheme 4-6.** Adenylation of a pyrrolidine precursor before assembly into lincomycin A.

Two intermediates expected to form prior and subsequent to the SibT-catalyzed reduction remain untested as adenylation substrates (**Scheme 4-7**). The former, **1.3**, consists of a 1-pyrroline ring while the latter consists of a proline ring **1.4**. The closest commercially available analogs of **1.4** (L-proline, L-*trans*-4-methylproline, L-*cis*-4-

hydroxyproline and L-3,4-dehydroproline) (**Scheme 4-8**) were also tested but none generated [$^{32}$P]ATP above background in this partial reaction. This is surprising given that

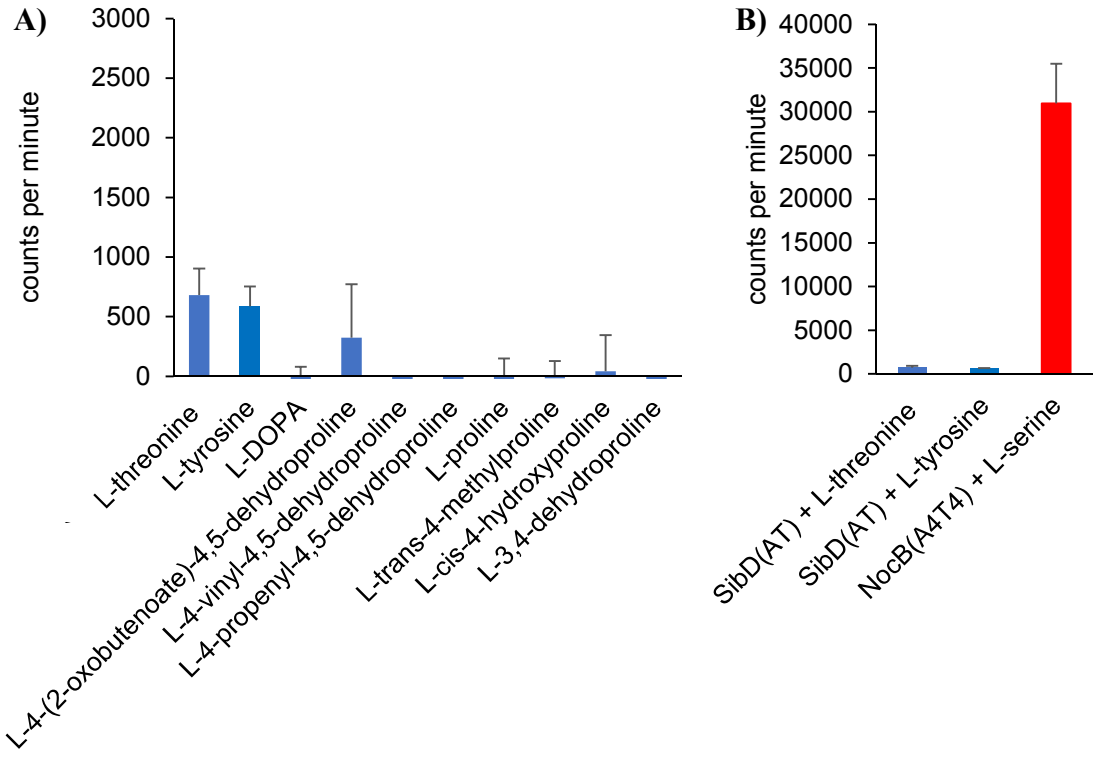
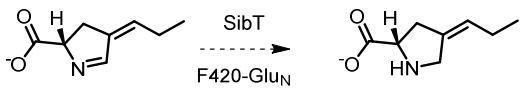


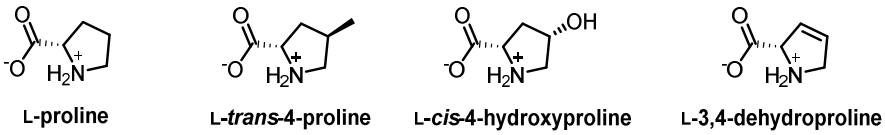

**Figure 4-2.** Probing for substrate activation by the adenylation domain in SibD$_{AT}$ via monitoring [$^{32}$P]ATP formation as a consequence of ATP/[$^{32}$P]PPi exchange during adenylation. Each measurement represents the average of three independent trials. The error bars represent one standard deviation of uncertainty. **A)** Adenylation by SibD$_{AT}$ is shown. **B)** The adenylation of L-serine by another NRPS NocB(A4T4) and its partner protein NocI is shown in red for comparison.




**Scheme 4-7.** The substrate and product in the proposed SibT-catalyzed reduction step are possible substrates for the adenylation domain in SibD$_{AT}$.

adenylases often activate substrate analogs, albeit less efficiently. For example, HrmP activates L-*trans*-4-methylproline in addition to its native substrate.[92] The lack of activity

might be explained by a highly specific adenylation domain but this is contradicted by the ability of $SibD_{AT}$ to activate at least two common amino acids as substrates.




**Scheme 4-8.** Available analogs for the proposed SibT product **1.4** that were tested as substrates for the adenylation domain in $SibD_{AT}$.

The adenylation domain in $SibD_{AT}$ is active but the absence of the condensation ($SibD_C$) and reductase ($SibD_R$) domains resulting in missing interdomain interactions may attenuate activity with the native substrate below levels of detection. To confirm that SibD's other domains are not necessary to detect adenylation of the native substrate, new plasmid constructs were cloned for expression and purification of the two missing domains as separate polypeptides. The two gene fragments *sibD*$_C$ and *sibD*$_R$ coding for $SibD_C$ and $SibD_R$, respectively, were predicted with aid of UMA (**Figure 4-1**). The gene fragment *sibD*$_C$ was synthesized (GenScript) as PCR amplification failed. Both gene fragments were cloned into a single pACYCDuet-1 plasmid utilizing both multiple cloning sites. The vectors pACYCDuet-1/*sibD*$_C$/*sibD*$_R$ and pSMT3/*sibD*$_{AT}$ containing compatible replicons and independent antibiotic selections were sequentially transformed and co-expressed in BAP-1 *E. coli* cells. The expressed proteins were simultaneously purified via Ni-NTA affinity chromatography, Ulp1-catalyzed proteolysis and size exclusion chromatography and yielded *ca*. 2 mg pure proteins per liter of growth media. An aliquot containing the purified proteins was denatured, digested with trypsin and subjected to tandem MS analysis. Detection of multiple peptides belonging to $SibD_C$ and $SibD_{AT}$ confirmed their successful co-purification. No peptides belonging to $SibD_R$ were detected. Unlike $SibD_C$

and $SibD_{AT}$, $SibD_R$ was expressed without the *N*-terminal $His_6$-SUMO-tag. The purification conditions appear to disrupt the interdomain interactions necessary for $SibD_R$ to co-purify with $SibD_C$ and $SibD_{AT}$. Instead the *sibD*$_R$ gene fragment was subcloned for separate expression and purification as an *N*-terminal $His_6$-SUMO fusion protein.

Adenylation by $SibD_{AT}$ with or without purified $SibD_C$ and $SibD_R$ added *in vitro* was tested against all the previously tested substrate candidates. The presence of molar equivalents of $SibD_C$ and/or $SibD_R$ to $SibD_{AT}$ during adenylation of all substrates except L-tyrosine did not generate [$^{32}$P]ATP counts above background counts. L-tyrosine adenylation by $SibD_{AT}$ with a molar equivalent of $SibD_C$ and/or $SibD_R$ generated equivalent [$^{32}$P]ATP counts within error compared to $SibD_{AT}$ alone (**Figure 4-3**). This suggests that the addition of the condensation and reductase domains are no effective for optimal adenylation activity although the native confirmation of SibD may require production of it as a single polypeptide.

**Thiolation by $SibD_{AT}$.** The adenylation reaction is only the first step toward acylating SibD's substrate. In the second step, the amino acyl adenylate is expected to form a thioester bond with the phosphopantetheine arm on the thiolation domain. The radioactivity assay monitors only the first of two partial reactions that constitute acylation. Adenylation alone may not reflect substrate selectivity of the acylation reaction. A mixture consisting of the reaction components necessary to form the L-tyrosinyl-adenylate was assayed to determine if it was subsequently covalently attached to $SibD_{AT}$ to form L-tyrosinyl-*S*-$SibD_{AT}$.

$SibD_{AT}$ from the adenylation reaction mixture was denatured and digested with trypsin before being subjected to tandem MS analysis. Both types of phosphopantetheine

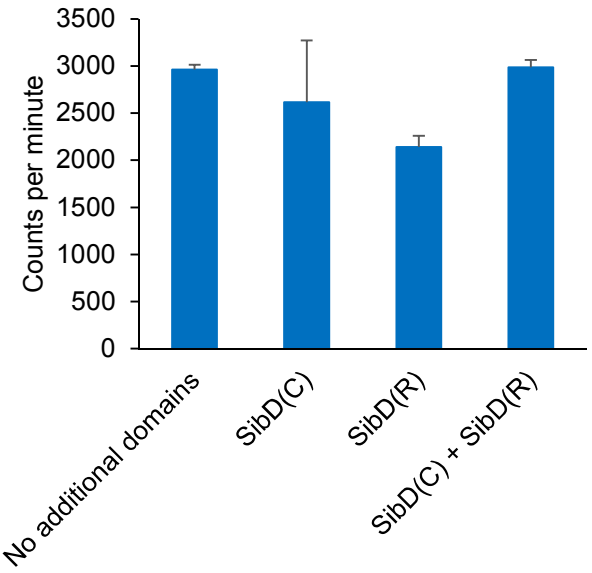
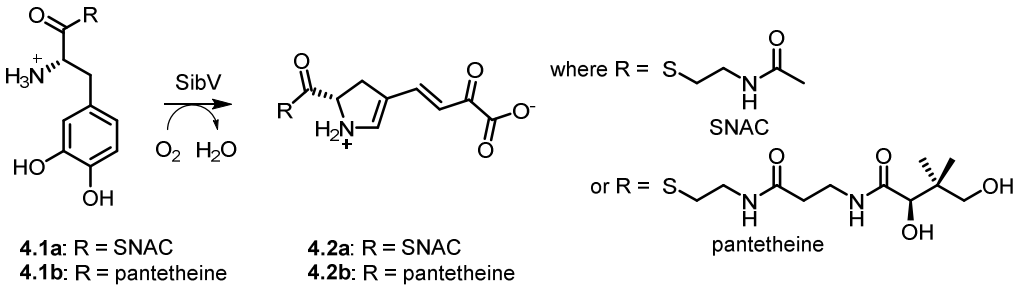



**Figure 4-3.** Effect of including the $SibD_C$ and/or $SibD_R$ domain during L-tyrosine adenylation by $SibD_{AT}$ assessed by the ATP/[$^{32}$P]PPi exchange assay. Each measurement represents the average of three independent trials. The error bars represent one standard deviation of uncertainty.

bearing peptides either conjugated to (**Appendix C-5 and C-6A**) or free of (**Appendix C-5 and C-7A**) L-tyrosine were detected. The peptide of interest bearing the phosphopantetheine arm conjugated to L-tyrosine was detected with an experimental mass of 3639.7837 Da in agreement with its theoretical mass of 3639.8298 Da within a mass deviation of 13 ppm. Several daughter b and y ions that differentiate between the acylated and unacylated peptides further supported their identities. For example the detection of acylated fragment peptide ions (y17*, y23* and y24*) with a mass deviation of 29, 18 and 15 ppm, respectively, indicated the formation of the acylated peptide (**Appendix C-6B**) while the detection of an unacylated fragment peptide ion (y28*) with a mass deviation of 28 ppm indicated also the presence of the unacylated peptide (**Appendix C-7B**). The presence of both sets of peptides indicate a mixture of L-tyrosinyl-*S*-$SibD_{AT}$ and unacylated $SibD_{AT}$ species were formed during adenylation and thiolation. The identical experiment in the absence of ATP generated unacylated peptides, serving as the negative control.

Replacing L-tyrosine with L-threonine, L-4-vinyl-4,5-dehydroproline **1.2** or L-4-propenyl-4,5-dehydroproline **1.5** as the substrate only generated unacylated peptides (**Appendix C-8, C-9 and C-10)**. We conclude that $SibD_{AT}$ is capable of catalyzing both adenylation and thiolation of L-tyrosine.

**Dioxygenation of substrate analogs mimicking L-DOPA tethered onto SibD.** If the native substrate for SibD is L-tyrosine, it follows that formation and modification of the dihydropyrrole moiety occurs after L-tyrosine is tethered to the thiolation domain. If this is the case, SibU should preferentially catalyze the hydroxylation of L-tyrosinyl-*S*-SibD over L-tyrosine. Similarly, SibV should preferentially catalyze the dioxygenation of L-DOPA-*S*-SibD over L-DOPA. Heterologous expression of SibU produced only insoluble protein even when expressed as an $His_6$-SUMO fusion protein[93] preventing study of the former reaction so the latter reaction was investigated instead. L-DOPA-*N*-acetylcysteamine (L-DOPA-SNAC) **4.1a** and L-DOPA-pantetheine **4.1b** were synthesized to serve as an analogs for L-DOPA-*S*-SibD and be tested as substrates for SibV **(Scheme 4-8)**. The SNAC moiety resembles the extremity of the phosphopantetheine arm. Replacing the acylated NRPS with small molecule thioester analogs is a commonly used strategy when the acylated NRPS is unavailable.[94,95]




**Scheme 4-9.** Enzymatic dioxygenation of L-DOPA and its SNAC and pantetheine derivatives.

A published protocol[96] was used to generate L-DOPA-SNAC **4.1a** and also adapted to generate L-DOPA-pantetheine **4.1b**. In both syntheses, L-DOPA was protected with an *N*-Boc group **4.3** (**Appendix C-11**) before converting it into a thioester **4.4** using PyBOP and benzenethiol. These steps proceeded under $N_2$ (g) to prevent polymerization of L-DOPA. This modification was not reported previously and appeared to increase the yield over the two steps from the reported[96] 6% to 63%. The desired thiophenolate ester **4.4** was confirmed by UPLC-ESI$^+$-MS. An experimental parent ion mass corresponded to the deprotected compound **4.5** (*m/z* 290.0850 [M-Boc+H]$^+$) expected under the acidic conditions used in the UPLC (**Appendix C-12**) and in agreement with the theoretical mass (m/z 290.0845 [M-Boc+H]$^+$) within a mass deviation of 1.7 ppm. The mass of the most intense signal corresponds to a fragment ion (experimental m/z = 152.0722) that forms upon dehydration and decarbonylation of the L-DOPA fragment ion (theoretical m/z = 152.0706) reported previously.[97] The $^1$H and $^{13}$C NMR signals for *N*-Boc-L-DOPA-thiophenol **4.4** (**Appendix C-13**) agreed with the reported[96] literature values. Transthioesterification with the SNAC thiol proceeded after the Boc deprotection step. Similarly, transthioesterification with the pantetheine dimethyl ketal thiol proceeded prior to the Boc deprotection step to prevent premature acetal deprotection. TFA removed the acid-labile protecting groups to form L-DOPA-SNAC **4.1a** and L-DOPA-pantetheine **4.1b**. The products were partially purified by RP-HPLC as they co-eluted with their respective deprotected thiol starting material (SNAC or pantetheine). Using ammonium formate/formic acid instead of the reported ammonium acetate/acetic acid as an HPLC buffer removed an interfering signal during NMR analysis. Two sets of $^1$H NMR signals were observed in the spectrum of each partially purified product confirming a mixture of

the desired product **4.1a** or **4.1b** and the thiol (**Appendix C-14 and C-15**) at a 4:1 ratio. Assignment of the $^{1}H$ signals was based on a model compound, L-tyrosine-SNAC as they were not reported for **4.1a**.[94] Product identities were further confirmed by UPLC-ESI$^{+}$-MS that generated experimental parent ion masses (*m/z* 299.1068 and 458.1949 $[M+H]^{+}$) in agreement with the theoretical masses (m/z 299.1060 and 458.1955 $[M+H]^{+}$) within a mass deviation of 2.7 and 1.3 ppm corresponding to L-DOPA-SNAC **4.1a** and L-DOPA-pantetheine **4.1b**, respectively (**Appendix C-16 and C-17**).

L-DOPA-SNAC **4.1a** and L-DOPA-pantetheine **4.1b** were tested as alternative substrates of SibV to probe if dioxygenation of L-DOPA-*S*-SibD occurs. L-DOPA-SNAC **4.1a** treated with SibV generated an experimental mass ion (*m/z* = 312.9641 $[M^{+}]$) in agreement with the theoretical mass (*m/z* = 313.0853 $[M^{+}]$) of **1.1**-SNAC **4.2a** expected to form upon dioxygenation and ring closure **(Figure 4-4)**. This mass signal was not observed if SibV or the substrate **4.1a** were left out of the reaction mixtures, serving as negative controls. However L-DOPA dioxygenation required a lower amount of SibV (0.1 mol %) to detect an observable product **1.1** signal in the HPLC spectrum whereas L-DOPA-SNAC **4.1a** required more SibV (27 mol %) to detect an observable product **4.2a** signal. SibV treatment of L-DOPA-pantetheine **4.1b** generated mass ions corresponding to unidentifiable compounds. Neither the expected dioxygenation product **1.1**-pantetheine **4.2b** nor its fragment generated via the equivalent fragmentation pattern (**Appendix C-17**) observed in **4.1b** were detected. SibV appears to favor catalysis of free L-DOPA over its conjugated thioester counterparts. This suggests SibV favors dioxygenation of free L-DOPA over L-DOPA-*S*-SibD although the latter may still prove to be the biosynthetically productive step.

One explanation for the absence of activity upon SibS or SibZ treatment of L-4-(2-oxobutenoate)-4,5-dehydroproline **1.1** (Chapter 3) is that these reactions only proceed if

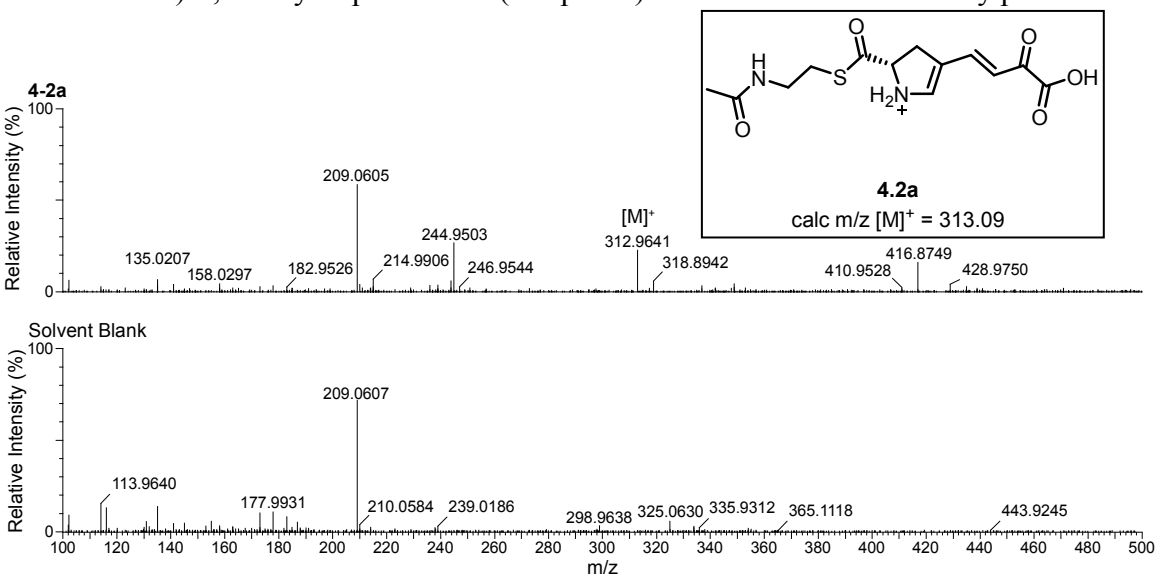



**Figure 4-4.** ESI$^+$-MS of solvent with (top) or without (bottom) **1.1**-SNAC **4.2a**. Both y-axes are the same scale with respect to the total ion count. The parent ion $[M]^+$ is detected.

their substrate is covalently attached to SibD. SibS and SibZ were incubated separately or as a putative protein complex with the **1.1**-SNAC **4.2a** generated *in situ* by SibV-catalyzed L-DOPA-SNAC **4.1a** dioxygenation. These reactions were performed in the absence and presence of $Mg^{2+}$ to see if it was necessary. The enzymatic mixtures were fractionated via ion-pair RP-HPLC (IPRP-HPLC). IPRP-HPLC prevented the substrate and potential products from co-eluting at the void volume with buffering reagents and metals unlike alternative used of C4, C8, C14-amide or C18 RP-HPLC. In each case, the HPLC spectrum signals appeared identical to the negative control in the absence of SibS, SibZ or both enzymes. No new signal corresponding to product formation was detected. It is important to note that SNAC compounds often fail to replace their NRPS counterparts as substrates[96] so such studies cannot conclusively rule out the possibility of SibS and SibZ catalyzing reactions with SibD conjugated compounds.

## 4.3 Summary

$SibD_{AT}$, the truncated version of SibD containing the adenylation and thiolation didomain was produced as a soluble and active protein. The thiolation domain contained the post-translationally installed phosphopantetheine group on its serine residue necessary for thiolation activity. $SibD_{AT}$ was expected to adenylate the product **1.5** of dihydropyrrole moiety biosynthesis. $SibD_{AT}$ adenylated L-tyrosine instead, the metabolic precursor in the pathway. The added presence of SibD's condensation and reductase domains did not affect adenylation activity as evident by a radioactive adenylation assay. Following adenylation, L-tyrosine was covalently conjugated to the thiolation domain as a thioester on the phosphopantetheine arm.

L-tyrosine may be a non-native substrate. If L-tyrosine is SibD's native substrate, L-DOPA-*S*-SibD should be an intermediate of the biosynthetic pathway. However, SibV prefers to catalyze the dioxygenation of L-DOPA over two L-DOPA-*S*-SibD analogs (L-DOPA-SNAC **4.1a** and L-DOPA-pantetheine **4.1b**). $SibD_{AT}$ also adenylates L-threonine, suggesting adenylation of certain common amino acids represents ancestral traits of SibD and not native activity. Two putative biosynthetic intermediates **1.3** and **1.4** are not currently available and may instead be the native substrate although surprisingly $SibD_{AT}$ is incapable of adenylating structurally similar compounds.

## 4.4 Experimental Procedures

**Materials.** The pACYCDuet-1 plasmid was obtained from Novagen (Darmstadt, Germany). Phusion high-fidelity DNA polymerase and its HF buffer were obtained from Finnzyme (Fair Lawn, NJ). T4 DNA ligase and its buffer were obtained from Promega (Madison, WI). The BAP-1[88] strain of *E. coli* was generously donated by Dr. C. Khosla at

Stanford University (Stanford, CA). Sequencing grade trypsin was obtained from Promega (Madison, WI) and 10 μL reverse-phase ZipTip$_{\mu\text{-C18}}$ was obtained from EMD Millipore (Darmstadt, Germany). [$^{32}$P]-Tetrasodium pyrophosphate was obtained from PerkinElmer (Waltham, MA). PyBOP was obtained from EMD Millipore. Pantetheine dimethyl ketal[98] and the expression vectors for NocB(A4T4) and NocI[99] were generously donated by Dr. C. Townsend at Johns Hopkins University (Baltimore, MD). All other relevant materials are described in previous chapters.

**General methods.** Codon optimization and DNA synthesis were performed by GenScript (Piscataway, NJ). Domain boundaries in SibD were predicted using the Udwary-Merski (UMA) algorithm.[86] NRPS protein domains were analyzed by SDS-PAGE made with 14% resolving and 4% stacking acrylamide layers. Peptide mass measurements were obtained in the ESI$^+$ mode using the Waters Acquity UPLC-Xevo-G2-aTof-MS (Milford, MA) equipped with a 2.5 mm x 150 mm BEH-C4 column (1.7 μM and 300 Å pore size) using a 5-80% aqueous acetonitrile gradient in 0.1% formic acid over 60 min (0.3 mL/min). Sequences for peptides were identified using BioPharmaLynx 1.4.3 and with a mass tolerance of 0.25 Da, MS and MS$^E$ mass match tolerance of 30.0 ppm and trypsin digest with 0 missed cleavages. Scintillation counting was performed with a PerkinElmer TriCarb 2910 TR liquid scintillation counter. NMR experiments were performed on a Bruker 300 MHz for the L-DOPA thioesters. L-DOPA thioesters concentrations were estimated using the $\varepsilon_{280\ nm}$ of L-DOPA (Chapter 2). HPLC analysis employed a Jasco PU-980 intelligent HPLC pump equipped with a HG-903-31 solvent mixing module, MD-1510 diode array (PDA) detection and Phenomex 250 x 4.6 mM Jupiter C18 analytical column with 5 μM particle size and 300 Å pore size (Torrance, CA). IPRP-HPLC was performed with

triethylamine (15 mM) and acetic acid (21 mM) (pH 5.0) as the aqueous mixture. Separation consisted of isocratic elution over 15 min, then a 5-95% methanol gradient over 10 min, followed by a 95% methanol isocratic step over 10 min. All other relevant general protocols are described in previous chapters.

**Cloning of the *sibD$_{AT}$* gene fragment.** The gene fragment encoding SibD$_{AT}$ was cloned sequentially in two pieces using an internal restriction site. The *sibD$_{AT(1700-3366)}$* gene fragment corresponding to nucleotides 1700-3366 in the *sibD* gene was amplified from the pSuperSib1[31] cosmid using a forward primer of 5'-CCTTCTTGGATCCCCGGCGCGGCAGTC-3' and a reverse primer of 5'-CCTTCTTAAGCTTTCACGTGTCGGCCCACAGG-3'. The PCR product was gel purified and digested with BamHI-HF and HindIII restriction enzymes. The resulting fragment was ligated with the linearized pSMT3 plasmid to generate pSMT3/*sibD$_{AT(1700-3366)}$*. The *sibD$_{A(1465-2100)}$* gene fragment harboring the 5' adenylase-encoding nucleotides 1465-2100 in the *sibD* gene was amplified from the pSuperSib1 cosmid using a forward primer of 5'-CCTTCTTGGATCCCAGCGGGTGCACGACC-3' and a reverse primer of 5'-CCTTCTTGGTCTCGCCGAGTCCG-3'. The PCR product was gel purified and digested with BamHI-HF and BsaI restriction enzymes. The fragment was then ligated with the linearized pSMT3/*sibD$_{AT(1700-3366)}$* vector to generate pSMT3/*sibD$_{AT}$*.

**Cloning of the *sibD$_C$* and *sibD$_R$* gene fragments.** The gene fragment encoding SibD$_C$ was synthesized with a 3'-opal stop codon and the remaining codons were optimized for expression in *E. coli* (**Appendix C-18**). This was cloned into the pUC57-Kan plasmid flanked by 5'-BamHI and 3'-HindIII restriction sites (pUC57/*sibD$_C$*) (GenScript). The gene fragment encoding SibD$_R$ was amplified from the pSuperSib1[31] cosmid using a

forward primer of 5'-AATTAATGGATCCATGCGTCTGCTGGACGGGTTG-3' and a reverse primer of 5'-AATTAATAAGCTTTCATGGCGCCGAGGTAGG-3' and the PCR product was gel purified. The corresponding gene fragments were digested with BamHI-HF and HindIII restriction enzymes and ligated with the linearized pSMT3 plasmid to generate pSMT3/*sibD*$_C$ and pSMT3/*sibD*$_R$ alternatively.

**Subcloning of the *sibD*$_C$ and *sibD*$_R$ gene fragments into a co-expression vector.** The *sibD*$_C$ gene fragment from the pUC57/*sibD*$_C$ vector was subcloned into multiple cloning site #1 in the pACYCDuet-1 plasmid using the 5'-BamHI and 3'-HindIII restriction sites to generate pACYCDuet-1/*sibD*$_{C(off\text{-}frame)}$. The codons in *sibD*$_C$ were off-frame from its upstream start codon, an unavoidable consequence of using the restriction sites available for subcloning. The *sibD*$_R$ gene fragment was amplified from the pSMT3/*sibD*$_R$ vector using a forward primer of 5'-AATTAATCATATGCGTCTGCTGGACGGGTTG-3' and a reverse primer of 5'-AATTAATCAATTGTCATGGCGCCGAGGTAGG-3'. The PCR product was gel purified and digested with NdeI and MfeI restriction enzymes. It was ligated with the linearized pACYCDuet-1/*sibD*$_{C(off\text{-}frame)}$ vector into multiple cloning site #2 to generate pACYCDuet-1/*sibD*$_{C(off\text{-}frame)}$/*sibD*$_R$. To correct the reading frame of the *sibD*$_C$ codons, both the pACYCDuet-1/*sibD*$_{C(off\text{-}frame)}$/*sibD*$_R$ and pSMT3/*sibD*$_C$ vectors were digested with NcoI, taking advantage of an internal restriction site within the *sibD*$_C$ gene fragment and then religated to generate pACYCDuet-1/*sibD*$_C$/*sibD*$_R$.

**Expression of proteins.** Expression was carried out under similar conditions described in Chapter 2 except for the following changes. Starter cultures consisted of either BAP-1 transformed with pSMT3/*sibD*$_{AT}$ or sequentially transformed with pSMT3/*sibD*$_{AT}$

and pACYCDuet-1/*sibD*$_C$/*sibD*$_R$ or Rosetta2(DE3) transformed with pSMT3/*sibD*$_R$. An additional antibiotic, chloramphenicol (17 μg/mL) was included in addition to kanamycin (25 μg/mL) in the BAP-1 cell culture containing two plasmids and the Rosetta2(DE3) cell culture. All cell cultures were ultimately induced at 16 °C with IPTG (20 μM). Expression of Noc(A4T4) and NocI is reported.[99]

**Purification of proteins.** Purification was carried out under similar conditions for purification of SibV described in Chapter 2 except for the following changes. All purification steps were performed at 4 °C except for the chromatographic steps which were performed at room temperature while the buffers were kept on ice. Cells were lysed by four passages through an EmulsiFlex homogenizer at 15,000 psi. Proteins were concentrated in an Amicon ultra centrifugal filter (10,000 NMWL) (EMD Millipore) at 5,000 rpm. All proteins were ≥ 95% pure as estimated by visual inspection after their separation by denaturing gel electrophoresis and staining with Coomassie Brilliant Blue. Purification of Noc(A4T4) and NocI is reported.[99]

**Synthesis of L-4-propenyl-4,5-dehydroproline 1.5.** Synthesis was carried out under identical conditions to those described in Chapter 3 except $Bu_3Sn$(vinyl) was replaced with $Bu_3Sn$(propenyl) (43 μL, 0.14 mmol).

***N*-Boc-4-propenyl-4,5-dehydro-L-proline methyl ester 4.6 and *N*-Boc-4-propenyl-3,4-dehydro-L-proline methyl ester 4.6'.** $^1$H NMR (400 MHz, $CDCl_3$) δ 6.12 and 5.91 (m, 1H), 5.68 and 5.67 (s(2), 1H), 5.61-5.41 (m, 1H), 5.04 and 4.95 (s(2), 1H), 4.55-4.16 (m, 2H), 3.72 (s(2), 3H), 1.79 (s(2), 3H), 1.48 and 1.42 (s(2), 9H). $^{13}$C NMR ($CDCl_3$) δ 171.3, 153.4, 140.6 and 139.8, 130.4 and 129.7, 122.6 and 121.8, 119.0 and

118.7, 80.2, 66.7 and 66.0, 55.0 and 54.6, 52.5 and 52.2, 28.3. HRMS ($ESI^+$): *m/z* 290.1364 $(M+Na)^+$. Calcd for $C_{14}H_{21}NO_4$ $(M+Na)^+$: 290.1363.

***N*-Boc-4-propenyl-4,5-dehydro-L-proline 4.7 and *N*-Boc-4-propenyl-3,4-dehydro-L-proline 4.7'.** $^1H$ NMR (400 MHz, MeOD) δ 6.26-5.91 (m, 1H), 5.81-5.55 (m, 2H), 4.50-4.18 (m, 2H), 1.84 and 1.82 (s(2), 3H), 1.50 and 1.45 (s(2), 9H). $^{13}C$ NMR (MeOD) δ 174.7 and 174.3, 141.0 and 140.0, 131.0 and 130.2, 125.9 and 124.1, 121.0, 81.6, 68.5 and 67.9, 55.8 and 53.3, 28.8 and 28.6, 18.5 and 15.1. HRMS (FAB): *m/z* 276.1213 $(M+Na)^+$. Calcd for $C_{13}H_{19}NO_4$ $(M+Na)^+$: 276.1206.

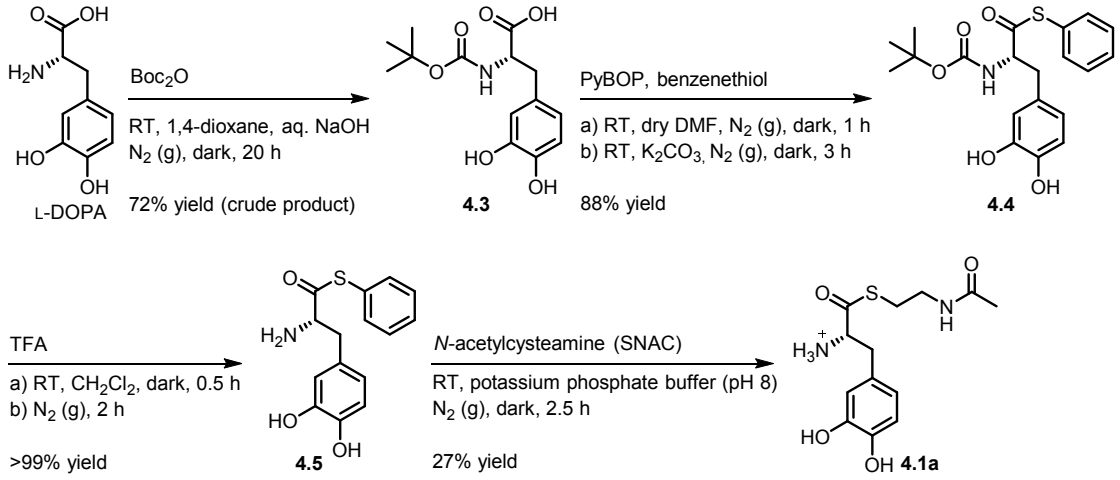
**4-Propenyl-4,5-dehydro-L-proline 1.5.** $^1H$ NMR (400 MHz, MeOD) δ 6.33-5.99 (dd, *J*=16, 12, 1H), 5.93-5.75 (m, 2H), 5.10 (s, 1H), 4.43-4.16 (m, 2H), 1.85 (s, 3H). $^{13}C$ NMR (MeOD) δ 138.0, 132.7, 124.4, 121.8, 119.4, 67.4, 54.2, 15.4. HRMS ($ESI^+$): *m/z* 154.0862 $(M+H)^+$. Calcd for $C_8H_{11}NO_2$ $(M+H)^+$: 154.0863.

**Adenylation activity.** The adenylation assay was based on a radioactive assay reported previously.[24] The $SibD_{AT}$ (0.25 μM) was incubated at 37 °C for 50 min with the putative amino acid substrate (0.8 mM) and cosubstrate ATP (3.7 mM) in HEPES (35 mM pH 7.5) along with TCEP (1.0 mM), $MgCl_2$ (7.5 mM), glycerol (7.5% v/v), NaPPi (1.0 mM) and [$^{32}P$]-PPi (1 μCi). $SibD_C$ (0.25 μM) and $SibD_R$ (0.25 μM) was included in reaction mixtures testing for condensation and reductase domain effects. The 100 μL reaction was quenched with 400 μL 500 mM $HClO_4$, chased with 400 μL 100 mM NaPPi and added to 200 μL 4% (w/v) activated charcoal. The charcoal was pelleted by centrifugation (14,000 rpm, 5 min) and washed 3x with 1 mL $H_2O$. The charcoal was resuspended in 500 μL $H_2O$ and transferred into a scintillation vial containing 5 mL scintillation fluid. Radioactivity was measured using a scintillation counter. Negative

control reactions were conducted by either excluding $SibD_{AT}$ or the putative amino acid substrate.

**Thiolation activity.** Pure $SibD_{AT}$ or an adenylation reaction mixture with $SibD_{AT}$ was denatured and then digested with trypsin according to the manufacturer's (Promega) instructions at a final protease to protein ratio of 1:100 (w/w). The reaction, denaturation and digestion steps were performed at pH 7.0 to prevent base-catalyzed hydrolysis of the thioester group from the phosphopantetheinylated peptide of interest. The tryptically digested peptides were desalted using a ZipTip and separated by the UPLC-MS for analysis as described in the general methods.

**Synthesis of L-DOPA-SNAC 4.1a.**



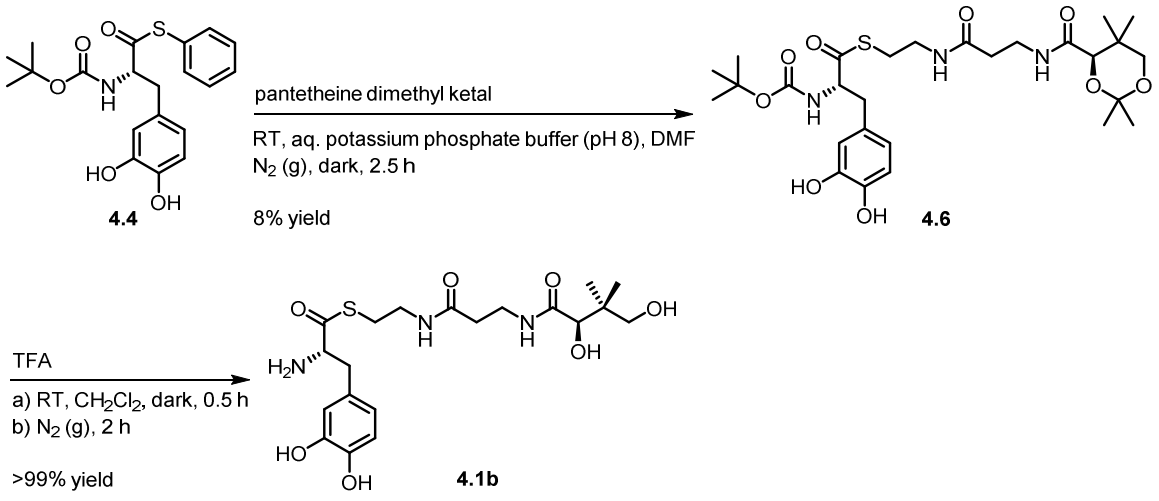

***N*-Boc-L-DOPA 4.3.** The synthesis was based on a reported protocol.[96] L-DOPA (250 mg, 1.25 mmol) and $Boc_2O$ (278 mg, 1.26 mmol) were dissolved in 1,4-dioxane (2.50 mL), $H_2O$ (1.25 mL) and NaOH (1.25 mL, 1 M). The reaction mixture was purged with $N_2$ (g) and stirred for 20 h in darkness at room temperature. The reaction mixture was acidified to pH 3 with $H_2SO_4$ (0.06 mL, 18 M) and extracted with EtOAc (3 x 3 mL). The organic fractions were combined, washed with $H_2O$ (3 x 3 mL) and brine (3 x 3 mL), dried over

$Na_2SO_4$ and evaporated under reduced pressure to yield a yellow oil (270 mg, 72% yield). $^1$H NMR (300 MHz, DMSO-$d_6$) δ 8.72 (bs, 1H), 6.93 (d, *J*=8.3, 1H), 6.65-6.56 (m, *J*=4.1, 2H), 6.47 (*J*=7.9, 1H), 4.09-3.85 (m, 1H), 2.95-2.45 (m, 2H), 1.33 (s, 9H). $^{13}$C NMR (DMSO-$d_6$) δ 173.8, 155.5, 144.9, 143.8, 128.7, 119.8, 116.5, 115.3, 78.1, 55.5, 36.0, 28.2. The crude product was used directly without purification in the next synthetic step.

***N*-Boc-L-DOPA-thiophenol 4.4.** Crude Boc-L-DOPA **4.3** prepared above (250 mg, 1.25 mmol) was dissolved in anhydrous DMF (1.0 mL) and combined with PyBOP (674 mg, 1.25 mmol) in anhydrous DMF (0.7 mL). This mixture was stirred for 10 min in darkness at room temperature under $N_2$ (g). Benzenethiol (152 µL, 1.25 mmol) was injected into the reaction mixture and stirred for 1 h. $K_2CO_3$ (86 mg, 0.63 mmol, 0.5 eq.) was added and the reaction mixture was stirred for an additional 3 h. The reaction mixture was combined with 10 mM potassium phosphate pH 6.0 (2.3 mL) and extracted with EtOAc (3 x 3 mL). The organic fractions were combined and evaporated under reduced pressure to yield a yellow oil. The oil was purified by silica gel radial chromatography (chromatotron) using a gradient of 30%-50% EtOAc in hexanes. The fractions containing desired material were evaporated under reduced pressure to yield a yellow oil (325 mg, 88% yield). $R_f$ 0.4 (40% EtOAc in hexanes). $^1$H NMR (300 MHz, $CDCl_3$) δ 7.46-7.29 (bs, 5H), 6.79 (d, *J*=8.1, 1H), 6.71 (d, *J*=1.5, 1H), 6.56 (d, *J*=7.5, 1H), 5.25 (d, *J*=8.7, 1H), 4.65 (m, 1H), 3.10-2.81 (m, 2H), 1.45 (s, 9H) are consistent with published values.[96] $^{13}$C NMR ($CDCl_3$) δ 200.0, 155.4, 144.2, 143.3, 134.6, 129.4, 129.1, 127.5, 127.1, 121.3, 116.2, 115.3, 80.8, 60.6, 37.6, 28.2 are consistent with published values.[96] HRMS (ESI$^+$): *m/z* 290.0850 (M-Boc+H)$^+$. Calcd for $C_{15}H_{15}NO_3S$ (M-Boc+H)$^+$: 290.0845.

**L-DOPA-SNAC 4.1a.** Boc-L-DOPA-thiophenol **4.4** (4.8 mg, 0.010 mmol) was dissolved in $CH_2Cl_2$ (0.1 mL) and treated with neat TFA (0.1 mL). The reaction mixture was stirred for 30 min in darkness at room temperature before it was placed under $N_2$ (g) for 2 h. *N*-Acetylcysteamine (SNAC) (5.5 µL, 0.050 mmol) was dissolved in 10 mM potassium phosphate pH 8.0 (0.2 mL), added to the reaction mixture and stirred for 2.5 h in darkness at room temperature under $N_2$ (g). The reaction was mixed with 5% $CH_3CN$ in 10 mM aqueous ammonium formate (1.0 mL) and fractionated by analytical C18 RP-HPLC as described in the general methods (5%-95% $CH_3CN$ gradient in 10 mM ammonium formate over 30 min, 1 mL/min). L-DOPA-SNAC **4.1a** was collected ($t_r$ = 8.6 min) and lyophilized to yield a white solid. $^1$H NMR (400 MHz, $D_2O$) δ 6.88 (d, *J*=8.1, 1H), 6.82 (d, *J*=2.3, 1H), 6.72 (dd, *J*=8.1, 2.0, 1H), 4.42 (t, *J*=7.9(2), 1H), 3.37 (m, 2H), 3.21 (m, 2H), 3.08 (m, 2H), 1.95 (s, 3H). HRMS (ESI$^+$): *m/z* 299.1068 $(M)^+$. Calcd for $C_{13}H_{19}N_2O_4S$ $(M)^+$: 299.1060. $\lambda_{max}$ = 239 and 283 nm (diode array detector, 18% $CH_3CN$ in 0.1% formic acid).

**Synthesis of L-DOPA-pantetheine 4.1b.**



**L-DOPA-pantetheine 4.1b.** Boc-L-DOPA-thiophenol **4.4** (4.8 mg, 0.01 mmol) was dissolved in DMF (50 µL) and mixed with 20 mM potassium phosphate pH 8.0 (100 µL). The solution was treated with pantetheine dimethyl ketal (3.2 mg, 0.010 mmol) also dissolved in DMF (50 µL) and mixed with 20 mM potassium phosphate pH 8.0 (100 µL). The reaction mixture was stirred for 2.5 h in darkness at room temperature under $N_2$ (g) and then treated with neat TFA (0.1 mL) before stirring for another 2 h in darkness at room temperature under $N_2$ (g). The reaction was mixed with 5% $CH_3CN$ in 10 mM aqueous ammonium formate pH 4.5 (1.0 mL) and fractionated by analytical C18 RP-HPLC (5%-95% $CH_3CN$ gradient in 10 mM ammonium formate over 30 min, 1 mL/min). L-DOPA-pantetheine **4.1b** was collected ($t_r$ = 9.0 min) and lyophilized to yield a white solid. HRMS ($ESI^+$): *m/z* 458.1949 $(M+H)^+$. Calcd for $C_{20}H_{31}N_3O_7S$ $(M+H)^+$: 458.1955. $\lambda_{max}$ = 243 and 283 nm (diode array detector, 21% $CH_3CN$ in 0.1% formic acid).

# Chapter 5: SibB Promotion of SibD's Adenylation Activity

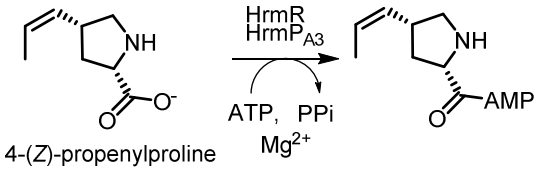
## 5.1 Introduction

The success of PBD engineering in the future relies not only on deciphering the dihydropyrrole biosynthetic steps but also achieving industrial-scale (multi-gram per liter) production of the desired PBD.[100] Improving the efficiency of biosynthetic enzymes is one strategy used to increase metabolite yields in engineered biosynthetic pathways.[101-104] This also appears to be used by nature as illustrated by a growing number of MbtH-like proteins (MLPs), a class of protein that promote adenylation reactions catalyzed by NRPSs.[105-112] Specifically MLPs, designated as members of the MLP superfamily, form complexes with adenylation domains to increase amino acyl adenylate production. The mechanism for this is unknown although mutant proteins implicate involvement of a hydrophobic interface that forms once the MLP and adenylase associate.[83,113-116] A search for an MLP in sibiromycin biosynthesis was undertaken by the prospect of increasing PBD yields in engineered pathways.

None of the genes within the sibiromycin biosynthetic cluster had previously been predicted[31] to code for proteins containing a Pfam domain (PF03621), a domain characteristic of MLP superfamily members. This domain contains three conserved tryptophans that are present within two universally conserved sequences (S$\phi$WP and P*X*GW where $\phi$ represents an aliphatic amino acid such as I, L or V) and one moderately (75%) conserved sequence (WTD*X*RP).[113] A recent BLAST search confirms that sibiromycin's 26 biosynthetic proteins share no sequence homology with predicted or known MLPs (evaluating hits with expect values < 0.1). However, such a gene could reside

elsewhere on the genome, a phenomenon observed for MLP encoding genes involved in the biosynthesis of other natural products.[99,117,118] However, locating an MLP in *S. sibiricum* would be difficult as its genome has not been sequenced. Although degenerate primers can amplify a target gene from genomic DNA via PCR, this technique is not amenable for MLP-encoding genes due to their low sequence homology. In some cases, MLPs from one biosynthetic pathway promotes an adenylation domain found in a biosynthetic pathway from another bacteria.[119-123] Expanding the search for candidate MLPs in other bacteria that produce related natural products pointed to three possible candidates.

Hormaomycin biosynthesis uses an MLP (HrmR) to activate the activity of all the adenylases embedded within its two NRPSs.[90] The adenylation domain in the third module of one of its NRPSs ($HrmP_{A3}$) co-expresses with HrmR and adenylates 4-(*Z*)-propenylproline (**Scheme 5-1**), a pyrrolidine moiety synthesized similarly to the dihydropyrrole moiety in sibiromycin.[90] In addition, HrmR appears to be a promiscuous MLP, activating seven adenylases with seven structurally distinct substrates. These two features make HrmR an attractive candidate for promoting SibD's adenylation activity.




**Scheme 5-1.** The adenylation domain in the NRPS involved in hormaomycin's pyrrolidine biosynthesis ($HrmP_{A3}$) catalyzes the adenylation of 4-(*Z*)-propenylproline only in the presence of HrmR, an MLP.

Mining the genomes of other PBD producers presents an opportunity to find additional candidates. However, none of the sequenced genes from the remaining related natural products (tomaymycin, anthramycin, porothramycin, tilivalline or lincomycin A)

are thought to code for MLPs and like sibiromycin, the genomes of their producers have not been sequenced. Instead, the Antibiotics & Secondary Metabolite Analysis Shell 3.0 (antiSMASH) program,[124] designed to find related biosynthetic gene clusters, was used to locate a new one in *Nocardia brasiliensis* (**Figure 5-1**), a fully sequenced human pathogen. Two MLP-encoding genes (NCBI accession numbers WP_014988415 and WP_011207082) were found outside that gene cluster and added to the list of candidates.

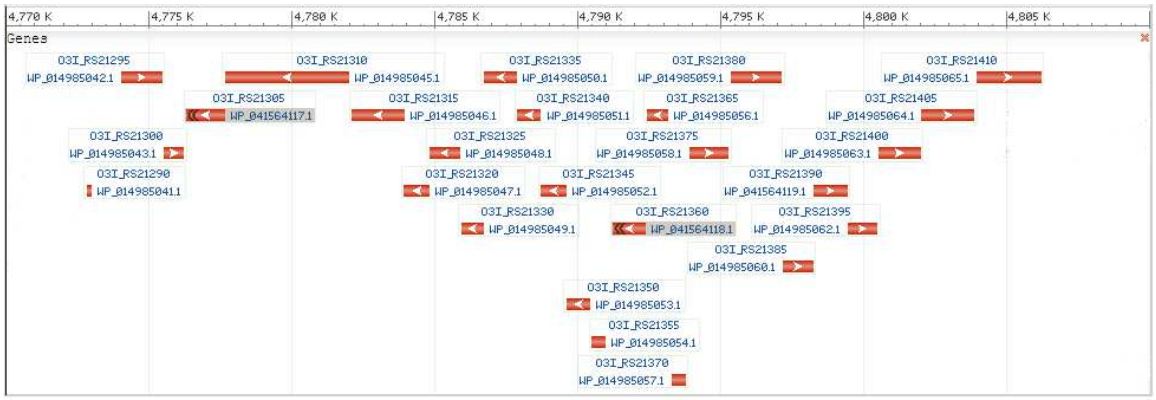



**Figure 5-1.** Newly identified PBD biosynthetic gene cluster in *N. brasiliensis* ATCC 700358 (GenBank accession number NC_018681).

One last protein was added to the list upon performing a rudimentary analysis of non-MLP proteins associated with sibiromycin biosynthesis. In rubradirin biosynthesis, a 180 amino acid domain devoid of consensus sequence signatures of MLPs still activates its downstream adenylation domain.[125] Neither the $SibD_C$ nor $SibD_R$ domains cause an observable effect on $SibD_{AT}$'s adenylation activity (Chapter 4) so different proteins within the sibiromycin biosynthetic pathway with no known function (SibB and SibR) were considered instead. SibB was a promising candidate as its encoding-gene (*sibB*) is expected to code for a small (93 aa) protein similarly sized to MLPs (~ 70 aa). Thus SibB was included in the list of candidates. According to BLAST analysis, SibB shares no sequence homology with MLP superfamily members. Although SibB has three tryptophans, none of

them are found in the same conserved sequences observed in MLPs. The SibB amino acid sequence shares low sequence similarity (<10%) with MLPs HrmR, WP_014988415 and WP_011207082.

The search for possible promoters of adenylation by SibD led to four possible candidates (HrmR, WP_014988415, WP_011207082 and SibB). Each protein was produced and then assayed to see if any could promote adenylation of L-tyrosine, a model substrate of $SibD_{AT}$. In addition, these proteins were assayed to determine if any could activate the adenylation of one of the other dihydropyrrole pathway intermediates, previously (Chapter 4) not adenylated by $SibD_{AT}$.

**5.2 Results and Discussion**

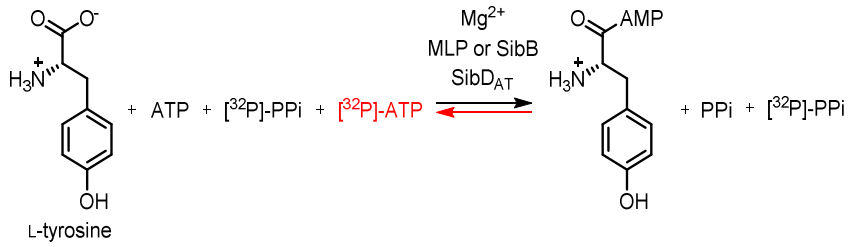
**Production of MLPs.** MLP coding genes that had been synthesized and optimized for expression in *E. coli* (GenScript) were subcloned into the pTYB12 vector. Each MLP fused with an *N*-terminal chitin binding domain-intein tag (CBD-intein) was expressed in BL21(DE3) *E. coli*. They were purified via chitin binding affinity chromatography. The tag was removed from the protein by DTT-induced CBD cleavage that resulted in a conformation change triggering intein cleavage. The final purification step provided *ca.* 8, 4, and 21 mg of HrmR, WP_014988415 and WP_011207082, respectively, per liter of growth media **(Appendix D-1 and D-2)**. Use of this purification tag was favored over the $His_6$-SUMO tag as the latter tag would pass its serine residue onto the *N*-terminus of the MLP. Any non-native residue has the potential to disrupt protein-protein interactions between each candidate MLP and $SibD_{AT}$, and thus non-native residues were avoided.

**Production of SibB.** The CBD-intein tagged SibB was expressed both in BL21(DE3) and Rosetta2 *E. coli*. The Rosetta2 strain is a derivative of the BL21(DE3) *E.*

*coli* strain that supplies additional tRNA codons rarely used by *E. coli* but present in the SibB amino acid sequence. Both sets of the fusion protein bound the chitin resin but in both cases, DTT treatment did not elute SibB from the resin based on SDS-PAGE analysis of the elution fraction. The likely explanation is that the protein misfolds during expression, consistent with a failed attempt to produce soluble native SibB without purification tags. Instead, soluble SibB was expressed as a fusion with an *N*-terminal $His_6$-SUMO tag in Rosetta2 *E. coli* and purified via Ni-NTA affinity chromatography using imidazole (250 mM) as the eluant. The tag was removed by Ulp1. SibB was separated from the free $His_6$-SUMO tag via a second round of Ni-NTA affinity chromatography. The final purification step provided *ca.* 0.5 mg of protein per liter of growth media **(Appendix D-3)**. SibB precipitated at concentrations higher than 0.1 mg/mL. This instability precluded use of size exclusion chromatography to obtain homogenous protein with respect to its oligomeric states as this step requires the application of a low volume of highly concentrated protein.

**SibB promotes $SibD_{AT}$ adenylation activity.** The effects of each MLP and SibB on L-tyrosine adenylation by $SibD_{AT}$ were tested using the ATP-[$^{32}$P]PPi exchange assay **(Scheme 5-2 and Figure 5-2)**. Addition of four molar equivalents of each of the three candidate MLPs to one molar equivalent of $SibD_{AT}$ during L-tyrosine adenylation did not increase the radiolabel incorporated into ATP compared with equivalent reactions lacking the MLP. All three MLPs consequently did not promote adenylation by $SibD_{AT}$. Addition of four molar equivalents of SibB in the equivalent adenylation reaction increased [$^{32}$P]-ATP yield nearly 6-fold (**Figure 5-2**). Analogous results were obtained when the substrate L-tyrosine was replaced with L-threonine, a non-native substrate of $SibD_{AT}$ (**Figure 5-3**).

This provided the first *in vitro* evidence that SibB promoted the adenylation reaction of $SibD_{AT}$.




**Scheme 5-2.** Radioactive L-tyrosine adenylation assay that monitors amino acyl-adenylate formation indirectly though the reversible exchange of [$^{32}$P]PPi with ATP followed by the capture of the resulting [$^{32}$P]ATP onto charcoal.

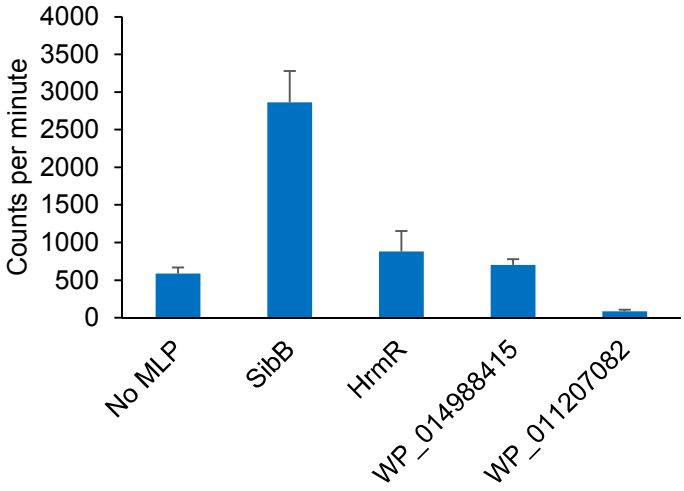



**Figure 5-2.** Effect of putative adenylase protein promoters on L-tyrosine adenylation by $SibD_{AT}$ assessed by the ATP/[$^{32}$P]PPi exchange assay. Each measurement represents the average of three independent trials. The error bars represent one standard deviation of uncertainty.

All the available sibiromycin dihydropyrrole pathway compounds and analogs were retested as substrates for adenylation by $SibD_{AT}$ in the presence of each MLP and SibB. With the exception of L-tyrosine and L-threonine, the remaining compounds treated with $SibD_{AT}$ did not increase the radiolabel incorporated into ATP compared with equivalent reactions lacking the MLP or SibB **(Figure 5-3)**. L-tyrosine and L-threonine remained the only substrates for $SibD_{AT}$. Among the candidate proteins, only SibB

promoted $SibD_{AT}$ adenylation of L-threonine and L-tyrosine, evident by an increase of radiolabel incorporation into ATP compared with equivalent reactions lacking SibB.

**Detection of a $His_6$-SUMO-SibB and $SibD_{AT}$ complex.** Decreasing the molar equivalents of SibB used from four to one in the L-tyrosine adenylation reaction described above resulted in a 54% drop in the level of radiolabel incorporation into ATP. A comparable result was observed with an MLP (NocI) that promotes multiple NRPS-catalyzed adenylation reactions during the assembly of the natural product nocardicin A.[99] NocI, representative of MLPs from other biosynthetic pathways, forms a protein complex with the NRPS,[99] suggesting SibB probably forms a protein complex with $SibD_{AT}$ as well. If SibB relies on strong protein-protein interactions with SibD for proper function, $SibD_{AT}$ might be expected to co-purify with SibB upon their co-expression. To test this theory, $SibD_{AT}$, as a fusion with an *N*-terminal $His_6$-SUMO tag, was co-expressed with native SibB in *E. coli*. The expressed proteins were purified via Ni-NTA chromatography, incubated with Ulp1 to remove the $His_6$-SUMO tag and purified further via size exclusion chromatography. However, the eluate produced by both chromatographic steps consisted of $His_6$-SUMO-$SibD_{AT}$ but not SibB, as determined by gel electrophoresis. An aliquot of the purified protein sample was denatured, digested with trypsin and subjected to UPLC-MS analysis. Only peptides derived from $SibD_{AT}$ were detected. It is likely that *E. coli* expressed an improperly folded SibB that is incapable of binding $SibD_{AT}$. The propensity for native SibB to misfold is supported by our earlier finding that SibB precipitates after removal of the $His_6$-SUMO tag. Improper folding of SibB also offers an explanation as to why cleavage of the intein tag did not occur in CBD-intein tagged SibB, a reaction sensitive

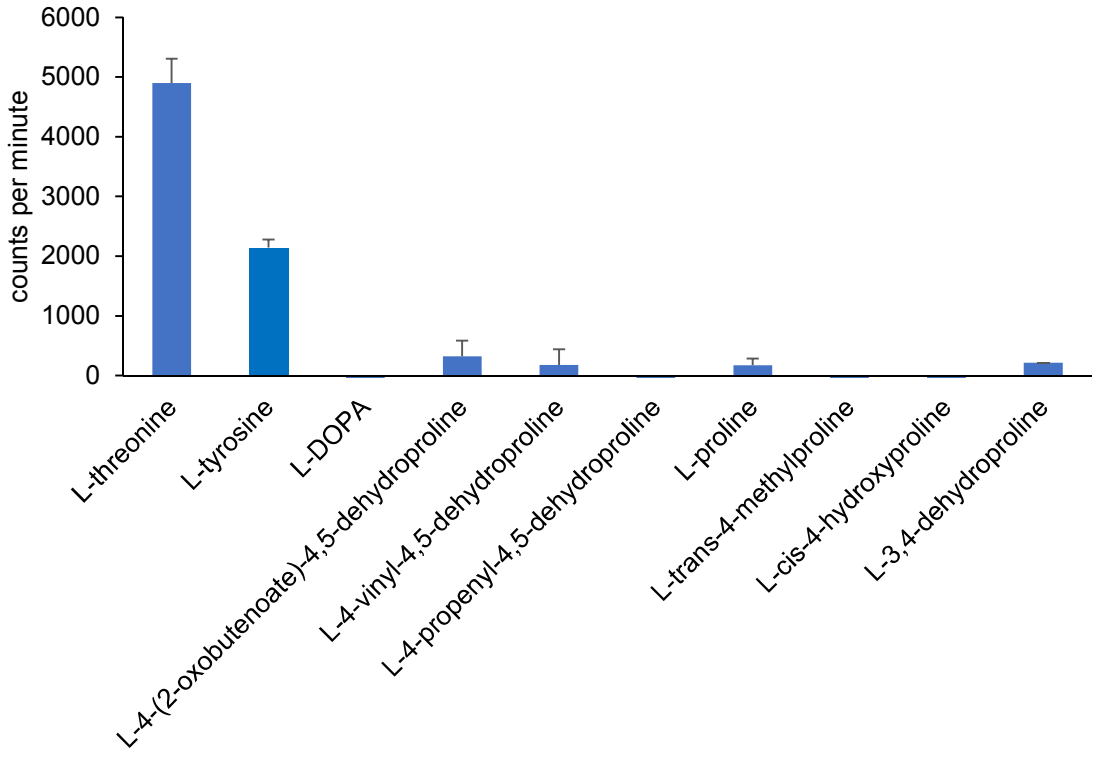



**Figure 5-3.** Probing for putative substrate activation by the adenylation domain in $SibD_{AT}$ in the presence of SibB determined by the ATP/[$^{32}$P]PPi exchange assay. Each measurement represents the average of three independent trials. The error bars represent one standard deviation of uncertainty.

to the quaternary structure of the protein. It is also possible that the co-purification conditions inhibited the protein-protein interactions necessary to form the protein complex.

Since we were unable to detect the protein complex *in vivo*, an *in vitro* approach was used instead. A binding assay was devised to see if the stable $His_6$-SUMO-SibB could act as a 'bait' protein and bind $SibD_{AT}$, the anticipated 'prey' protein. The ratio of $His_6$-SUMO-SibB and $SibD_{AT}$ in the protein complex was investigated more rigorously testing various molar equivalents of $SibD_{AT}$ **(Figure 5-4)**. At least up to four molar equivalents $SibD_{AT}$ in the presence of one molar equivalent of $His_6$-SUMO-SibB bound the nickel chelate resin. The two proteins co-eluted in cleavage buffer consistent with protein-protein interactions holding the complex together. A negative control shows that $SibD_{AT}$ does not

bind the nickel chelate resin in the absence of $His_6$-SUMO-SibB. Although the results suggest that SibD and SibB form a protein complex at a 4:1 ratio, this ratio remains speculative without knowledge of the oligomeric states of $SibD_{AT}$ and SibB. During its purification, $SibD_{AT}$ eluted from the size exclusion chromatography column as multiple peaks suggesting $SibD_{AT}$ exists in multiple oligomeric states. It is unclear if this heterogeneity persists with native SibD or if all the oligomeric states are functionally active.

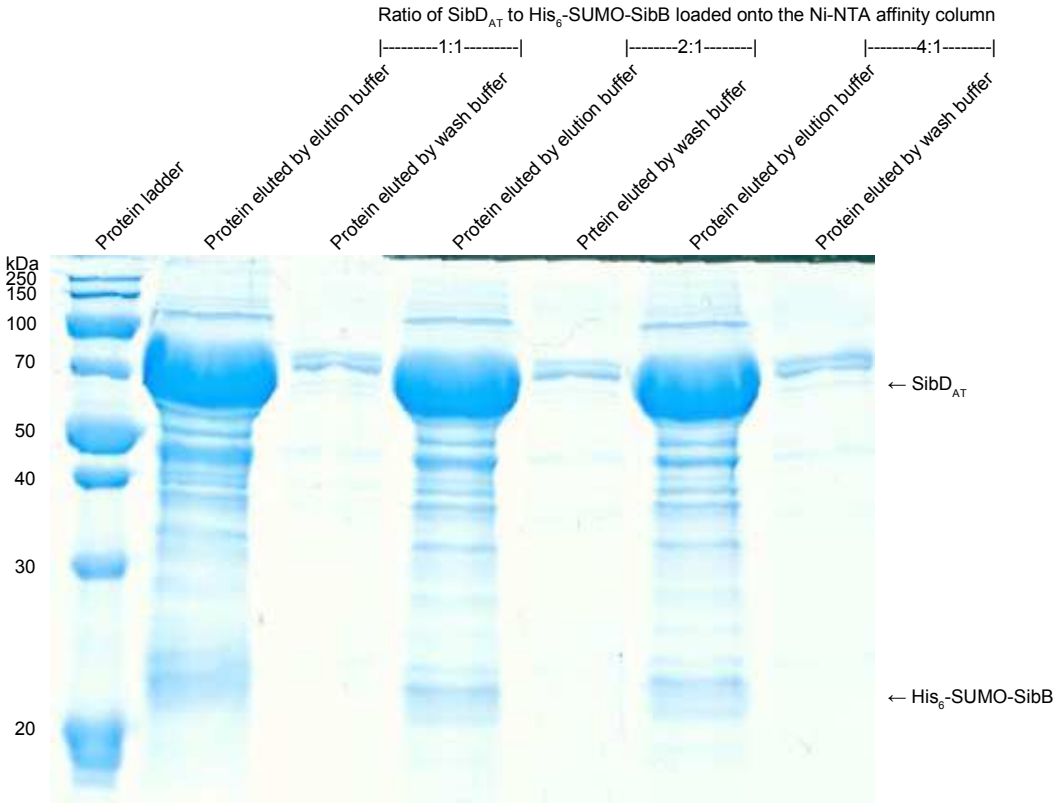



**Figure 5-4.** SDS-PAGE gel image of fractions from the *in vitro* binding assay to detect binding between $SibD_{AT}$ and $His_6$-SUMO-SibB. SDS-PAGE gel of the eluate consisting of the bait protein $His_6$-SUMO-SibB and the prey protein $SibD_{AT}$ loaded after application onto a Ni-NTA agarose column and elution with sodium phosphate buffer containing imidazole (250 mM). Mixtures of $SibD_{AT}$ to $His_6$-SUMO-SibB at a range of molar ratios were incubated together at 4 °C for 12 min prior to column loading.

**Prediction of residues expected to constitute the SibB-SibD interface.** The functional similarity between SibB and MLPs raises a possibility that they share a similar

mode of action. Crystal structures and sequence alignments of MLP-dependent adenylases from NRPSs indicate a conserved alanine or proline interacts with a hydrophobic cleft on the MLP composed of a conserved tryptophan dyad.[113,114] In contrast, the MLP-independent adenylases carry other amino acids at this site.[116] The adenylation domains in HrmP and LmbC, related to the adenylation domain in SibD, conform to this trend. HrmP contains an MLP-dependent adenylase[90] in its third module with an alanine (A2977) expected to interact with the MLP while LmbC is an MLP-independent standalone adenylase[89,126,127] with a threonine (T376) at the equivalent position (**Figure 5-5**). A proline occupies the same position in SibD (P887), initially suggesting it was an MLP-dependent protein. SibD deviates from the trend as it invokes the action of SibB, a non-MLP, to promote the adenylation reaction. P887 in SibD is conserved in other PBD synthesizing homologs TomB (P1012), Orf22 (P930), Por21 (P837) as well as the homolog (NCBI accession number WP_014985045) in the putative PBD producer *N. brasiliensis* (P815). By extrapolation, their adenylation reactions are expected to be promoted by similar proteins. In contrast, the SibD homolog NpsB predicted to assemble the PBD tilivalline likely catalyzes its adenylation reaction without using another protein as a serine (S844) is present instead of the proline. The three tryptophans present in SibB could form the hydrophobic interface and interact with the P887 in SibD although mutagenesis experiments are necessary to validate the requirement and role of all the aforementioned residues.

The *in vitro* studies show that SibB promotes the adenylation of L-tyrosine by $SibD_{AT}$. Blast analysis of SibB reveals that homologs are coded by genes in the pathogen *Mycobacterium abscessus* (NCBI accession number CRG59121) and antitumor antibiotic

```
SibD            QRPV-PFGVRGELYVGGPVVGRGYLHLPDQTRERFRPDPYAGRPGGRMYRTGDL  904
TomB            RELL-PPGVTGELYIGGPYLSPGYRGRLEETARRFLPDPYGG-DGEVIYRTGDL 1028
Orf22           DEPA-PFGVEGELCVAGPVVSPGYIGLPEKTRQAMVPDPFV--PGQLMYRTGDV  945
Por21           AQRA-PFGVDGELCVAGPVVSPGYIGLPEKTSRAMVPDPFV--PGQLMYRTGDI  852
WP_014985045    AQRV-PFGVEGELCVAGPVISPGYLGLPEKTREAMVPDPFV--AGQLMYRTGDV  830
NpsB            LQLV-PAGVAGELLIVSDFLSPGYHNKSEESARRFISAEGFATQSCHAYRTGDL  861
LmbC            GRPVDAVGSVGELVVAGPTVAAGYWGVEGHGAERFRTGETC--PGGRAYATGDQ  391
HrmPA3          LQAV-PPGVPGELCIAGPQLARGYLGRPGTTAERFVADPFGP-SGARMYRTGDR 2993
                     . *  *** : .  :. **          : .       .   * ***
```

**Figure 5-5**. Sequence alignment of SibD with biosynthetically relevant homologs. Any proline or alanine at the highlighted position is predicted to interact with an MLP.[116] The NCBI accession number of SibD homolog suspected of producing the PBD in *N. brasiliensis* is written as this protein remains unnamed. The last residue shown is numbered on the right.

producer *Streptomyces globisporus* (NCBI accession number WP_030583800) but none appear in other PBD biosynthetic pathways. However, functionally similar proteins may exist as the anthramycin gene cluster contains genes predicted to code for small proteins (89 aa length for Orf11 and 58 aa length for Orf18) and the putative PBD gene cluster in *N. brasiliensis* contains a gene predicted to code for a small protein (62 aa length for WP_014985041), all of unknown function.

### 5.3 Summary

The *sibB* gene, present within the sibiromycin biosynthetic gene cluster, codes for a small protein. Its role was not known until now. SibB promotes the adenylation activity of $SibD_{AT}$. $His_6$-SUMO-SibB co-elutes with $SibD_{AT}$ suggesting SibD recruits SibB *in vivo* to form a functionally active protein complex. SibB's promotion of adenylation activity is likely enhanced in the presence of SibD's unknown native substrate. Although SibB behaves similarly to MLPs, differences in their primary sequences expands the list of adenylase promoters. MLPs known (HrmR) or expected (WP_014988415 and WP_011207082) to promote adenylation reaction during assembly of related natural products did not replace SibB functionally.

## 5.4 Experimental Procedures

**Materials.** The pTYB12 plasmid and chitin resin were obtained from New England Biolabs (Ipswich, MA). The Pierce BCA protein assay kit was obtained from Thermo Scientific (Grand Island, NY). A HiTrap chelating HP column was obtained from GE Healthcare (Piscataway, NJ). The Coomassie Brilliant Blue G-250 protein stain powder was obtained from Bio-Rad (Hercules, CA). All other relevant materials are described in previous chapters.

**General methods.** A 15% resolving acrylamide layer was used to separate small (<15 kDa) proteins by SDS-PAGE for analysis of expression and purification. Protein samples from *in vitro* protein-protein binding assays were visualized after SDS-PAGE by staining the gel with colloidal-based Coomassie dye using a protocol[128] in which ethanol replaced methanol in the solvent. Sequence alignments were generated using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) program.[129,130] All other relevant general protocols are described in previous chapters.

**Cloning of the MLP-coding and *sibB* genes.** The nucleotide sequence used to express HmrR previously[90] was synthesized (GenScript) and cloned into the pUC57-Kan plasmid flanked by 5'-NcoI and 3'-HindIII restriction sites (GenScript). Non-native 5'-CG bases were included to keep the codons in-frame with the upstream start codon. The *hrmR* gene from the pUC57/*hrmR* vector was subcloned into the first multiple cloning site in the pACYCDuet-1 plasmid (Novagen) using the 5'-NcoI and 3'-HindIII restriction sites to generate pACYCDuet-1/*hrmR*. The *hrmR* gene was amplified from the pACYCDuet-1/*hrmR* vector using a forward primer of 5'-AAAAAAGAATGCTCGAGATCGAGAAAAGAATGGA-3' and a reverse primer of

5’-AAAAAACTCGAGTCATGCCGCGCTCCCTTC-3’. The PCR product was gel purified. The genes encoding WP_014988415 and WP_011207082 were each synthesized (GenScript) with codons optimized for expression in *E. coli* (**Appendix D-4**) and cloned into the pUC57-Kan plasmid flanked by 5’-BsmI and 3’-XhoI restriction sites (GenScript). For both genes, a non-native 5’-T base was included to keep the codons in-frame with the upstream start codon of its destination vector (GenScript). The gene encoding SibB was amplified from the pSuperSib1[31] cosmid using a forward primer of 5’-AATTAATGAATGCTACCCGGACCGCCTCGC-3’ and a reverse primer of 5’-AATTAATCTCGAGCTACAGCCACCCGGGC-3’ and the PCR product was gel purified. All four genes were digested with BsmI and XhoI restriction enzymes and each ligated with the linearized pTYB12 plasmid to generate pTYB12/*hrmR*, pTYB12/*wp_014988415*, pTYB12/*wp_011207082* and pTYB12/*sibB*. The genes each contained chitin binding domain encoding gene fragment within an intein encoding gene fragment on the 5’ side of the gene.

**Subcloning of the *sibB* gene.** The gene encoding SibB was amplified from the pSuperSib1[31] cosmid using a forward primer of 5’-AATTAAAGGATCCACCCGGACCGCCTCGC-3’ and a reverse primer of 5’-AATTAAAAAGCTTTCACAGCCACCCGGGCGC-3’. The PCR product was gel purified and digested with BamHI-HF and HindIII restriction enzymes. It was ligated with the linearized pSMT3 plasmid to generate pSMT3/*sibB*.

**Subcloning of the *sibB* gene into a co-expression vector.** The gene encoding SibB was amplified using 5’-BsmI-*sibB*-XhoI-3’ as a template that was previously amplified, purified and digested as described above. A forward primer of 5’-

AATTAAACCATGGGCACCCGGACCGCCTCGC-3' and a reverse primer of 5'-AATTAAAAAGCTTTCACAGCCACCCGGGCGC-3' were used in the PCR reaction. The resulting PCR product was gel purified, digested with NcoI and HindIII restriction enzymes and ligated with the linearized pACYCDuet-1 plasmid into the first multiple cloning site to generate pACYCDuet-1/*sibB*.

**Expression of proteins.** Expression was carried out under similar conditions described in Chapter 2 except for the following changes. *E. coli* cultures transformed with the pTYB12-based plasmids were supplemented with ampicillin (100 μg/mL) instead of kanamycin (50 μg/mL). The pTYB12/*sibB* and pSMT3/*sibB* were each transformed into Rosetta2(DE3) and their cultures were supplemented with ampicillin (50 μg/mL) and kanamycin (25 μg/mL), respectively. An additional antibiotic, chloramphenicol (17 μg/mL) was included in the Rosetta2(DE3) cell cultures. Cell cultures transformed with pTYB12-based plasmids were induced with IPTG (500 μM) while the cell cultures transformed with pSMT3/*sibB* were induced with IPTG (50 μM). Inductions were performed at 16 °C and the induction time was shortened to 4 h for WP_011207082 compared with 12 h for the remaining proteins. Co-expression of $His_6$-SUMO-$SibD_{AT}$ with SibB was carried out under identical co-expression conditions described in Chapter 4.

**Purification of proteins.** All purification steps were performed at 4 °C. Cells alternatively expressing CBD-intein-HrmR, CBD-intein-WP_014988415, CBD-intein-WP_011207082 or CBD-intein-SibB were resuspended in resuspension buffer (20 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA and 1 mM fresh TCEP). The cells were lysed by four passages through an EmulsiFlex homogenizer at 15,000 psi. Cell debris was removed

by centrifugation at 40,000 x g (60 min) and the supernatant was loaded onto a chitin resin (10 mL) column pre-equilibrated with wash buffer (50 mM Tris pH 8.0, 50 mM NaCl, 10% glycerol and 1 mM fresh TCEP). The resin was washed with wash buffer, resuspended in cleavage buffer (100 mM DTT in wash buffer) and incubated at 4 °C for three days to remove the intein tag by proteolysis. The protein was eluted with cleavage buffer via gravity. High molecular weight impurities were removed using an Amicon ultra centrifugal filter at 5,000 rpm with a membrane (30,000 NMWL). The protein in the filtrate was concentrated in an Amicon ultra centrifugal filter at 5,000 rpm with a membrane (3,500 NMWL). Protein solutions in gel filtration buffer were flash frozen in droplets with $N_2$ (l) and stored at -80 °C until needed. The attempted co-purification and purification of $His_6$-SUMO fusion proteins were carried out under similar conditions for purification of $SibD_{AT}$ described in Chapter 4 except for the following changes. SibB was susceptible to aggregation after its $His_6$-SUMO tag was removed so SibB was concentrated in an Amicon ultra centrifugal filter with a membrane (3,500 NMWL) only to < 0.2 mg/mL because SibB precipitated at higher concentrations. All isolated proteins were ≥ 95% pure as estimated by visual inspection after their separation by denaturing gel electrophoresis and staining with Coomassie Brilliant Blue.

**Adenylation activity.** The adenylation assay was monitored under identical conditions described in Chapter 4. An MLP or SibB (1 μM) was included in the reaction mixture.

**Protein-protein binding assay.** $His_6$-SUMO-SibB and $SibD_{AT}$ were quantified using the BCA protein assay kit. A mixture of both proteins at the indicated ratios in wash buffer (50 mM sodium phosphate pH 8.0, 175 mM NaCl, 10% glycerol and 1 mM fresh

TCEP) was loaded onto a Hi-Trap chelating HP (1 mL) column charged with $Ni^{2+}$ and pre-equilibrated with wash buffer. The resin was washed with wash buffer (5 CV) before the protein complex was eluted with elution buffer (1.5 CV) (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol and 1 mM fresh TCEP). The eluate from the wash (3.0 mL) and elution (1.5 mL) steps were concentrated to 150 µL in an Amicon ultra centrifugal filter (3,000 NMWL) at 13,000 rpm. The protein in each concentrated sample was separated by SDS-PAGE and visualized by staining the gel with a colloidal-based Coomassie stain.

# Chapter 6: Conclusions

The ultimate goal of studying PBD dihydropyrrole biosynthesis is to engineer new drugs capable of alkylating DNA and evading DNA error-correcting proteins. This work focused on the production and study of biosynthetic enzymes that are thought to be common in the formation, modification and assembly of the dihydropyrrole moiety within PBDs.

Unwanted byproducts and competing pathways are always issues of concern in engineered pathways.[100] For example, extradiol dioxygenases such as L-DOPA dioxygenases have the potential to cleave their catechols at multiple sites to form multiple acyclic compounds. The acyclic compounds themselves can then cyclize into multiple compounds. However, only a single regiochemistry is promoted by SibV and its homolog Orf12 to establish the dihydropyrrole scaffold **1.1** from L-DOPA. The lability of **1.1** ($t_{1/2}$ of 36 h at 25 °C) indicates that the next step in the engineered pathway must be efficient to avoid decomposition of this key pathway intermediate.

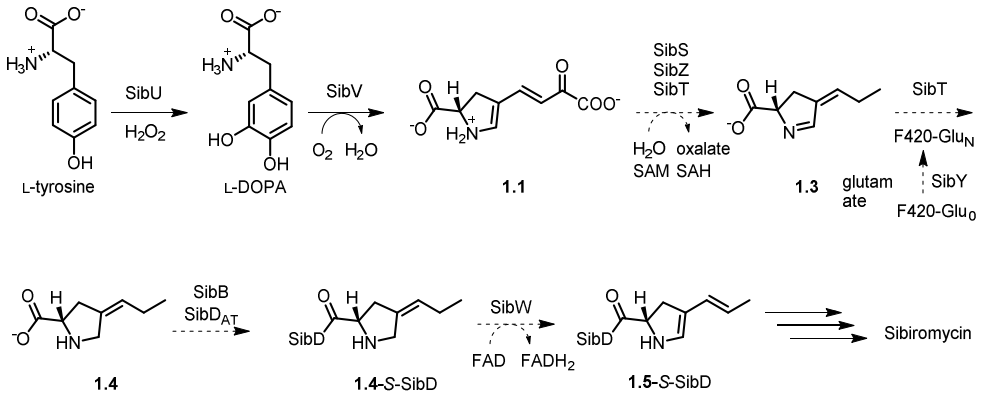
Maximizing product yields is crucial for engineered pathways to become the source of pharmaceuticals. The ability for MLPs to promote adenylation activity of NRPSs prompted a search for a functionally similar protein involved in sibiromycin biosynthesis. SibD catalyzed adenylation is promoted by SibB, a previously uncharacterized protein. Engineered pathways, particularly those based on PBDs that lack the SibB homolog may benefit by incorporating *sibB* into the gene cluster. This also encourages the search for functionally similar proteins beyond MLPs in other natural product biosynthetic pathways.

Before engineering efforts may begin, the function and/or substrates of many biosynthetic enzymes still need to be identified. SibS's affinity for **1.1** ($K_D$ = 64 ± 2 μM) supports its involvement in dihydropyrrole biosynthesis. However, the absence of observable hydrolase activity with SibS and methyltransferase activity with SibZ indicate that modifications to the dihydropyrrole scaffold are more complex than initially proposed. The ability for SibD to adenylate and thiolate L-tyrosine but not the final dihydropyrrole moiety **1.5** also indicates that the proposed pathway needs revision. The following hypotheses aim to provide alternative steps that are consistent with prior knowledge of the pathways and findings in this work.

**Hydrolysis and methylation of 1.1 are catalyzed by a protein complex.** One possible revision of the pathway involves formation of a protein complex including SibT, conserved throughout the biosynthetic pathways, to catalyze the hydrolysis and methylation of **1.1**. The use of a complex, although unprecedented for these types of reactions, could account for the absence of activity observed for reaction mixtures with SibS and/or SibZ. Upon formation of **1.1**, a complex of SibS, SibZ and SibT are proposed to hydrolyze and methylate **1.1** to **1.3** (**Scheme 6-1**). SibS and SibZ are still expected to participate in these reactions based on SibS's affinity for **1.1** and the SAM-binding motif in SibZ's amino acid sequence. In the presence of the requisite F420-GluN cofactor, SibT is still proposed to reduce **1.3** to **1.4**.[31] These steps are consistent with the accumulation of **1.3** in the absence of the F420-GluN cofactor in a mutant strain of the lincomycin A producer.[42]

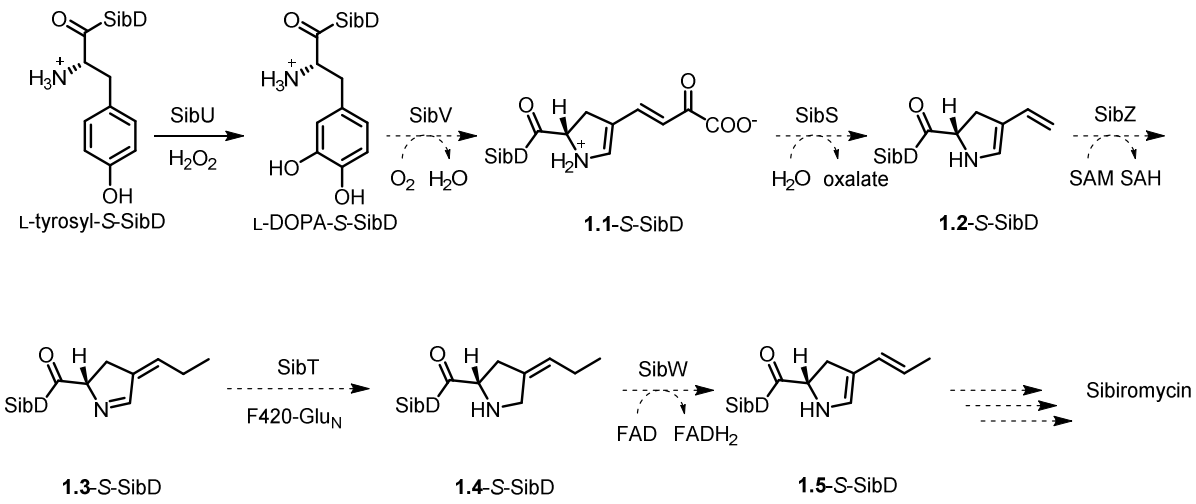
SibD is then proposed to adenylate and acylate **1.4**. Enzymes unique to each dihydropyrrole and pyrrolidine pathway are proposed to catalyze reactions with the

acylated **1.4** (**1.4**-*S*-SibD). For example, SibW in sibiromycin is proposed to catalyze an oxidation to form **1.5**-*S*-SibD. This accounts for the inability for SibD to adenylate or thiolate **1.5**. Once formation of the dihydropyrrole moiety is complete, the condensation and reductase domains of SibD are proposed to incorporate it into the natural product. Testing this hypothesis minimally requires the production of SibT, SibY and SibW and the synthesis of **1.4**.



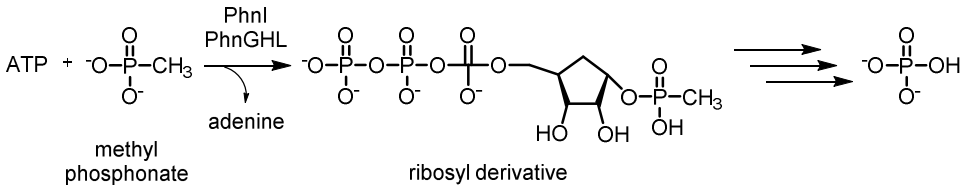

**Scheme 6-1.** Dihydropyrrole biosynthetic pathway in sibiromycin proposed based on the findings in this work.

**All biosynthetic steps occur with compounds loaded on the NRPS.** An alternative revision of the pathway involves SibD-catalyzed adenylation and thiolation of L-tyrosine observed *in vitro*. All the dihydropyrrole biosynthetic enzymes are then proposed to catalyze reactions with substrates conjugated to the ACP (**Scheme 6-2**). Thus, SibU, SibV, SibS, SibZ, SibT and SibW are proposed to catalyze reactions with biosynthetic intermediates that are loaded onto SibD. If this is correct, this would be the first reported example of a hydrolase and *C*-methyltransferase capable of processing substrates that are loaded on an ACP. SibS and SibZ were unable to process **1.1**-SNAC although these thioester substrate analogs often fail to replicate activity observed with their native NRPS counterparts.[96]

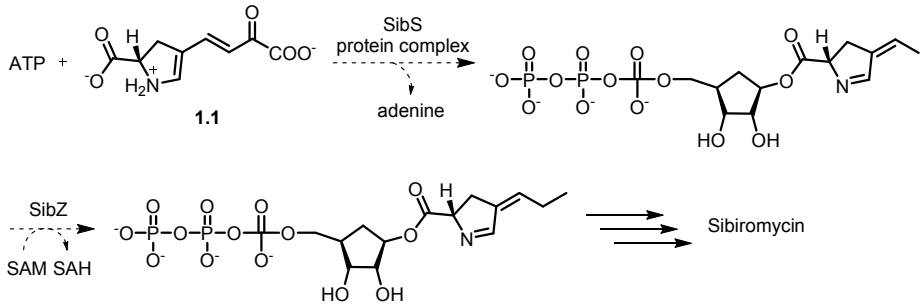



**Scheme 6-2.** An alternative biosynthetic pathway for dihydropyrrole moiety in sibiromycin.

**An alternative pathway that involves depurination and use of a protein complex.** Both proposals discussed above dismiss SibS's ability to depurinate SAM as an unproductive reaction. Although the SibS-catalyzed depurination of SAM does not have an obvious place in the biosynthesis of PBDs, it resembles a step during the biosynthesis of phosphate in *E. coli* derived from methylphosphonate.[131] Specifically, PhnI catalyzes the depurination of ATP to form a ribosyl derivative that is then ligated to methylphosphonate by action of an enzyme complex, PhnGHL (**Scheme 6-3**). By analogy, some of dihydropyrrole biosynthetic intermediates may exist as ribose intermediates as well (**Scheme 6-3**). This could involve the displacement of adenine with **1.1** on a ribose derivative. To verify this theory, SAM or another donor such as ATP will need to provide the ribose acceptor in the presence of a protein complex that transfers **1.1** onto the ribose derivative. Candidates for the protein complex include SibS, SibZ and SibT common in the majority of dihydropyrrole biosynthetic pathways.




**Scheme 6-3.** The formation of phosphate proceeds via the PhnI-catalyzed depurination of ATP to form a ribosyl group. The PhnGHL enzyme complex catalyzes the transfer of phosphonate onto the ribosyl acceptor.




**Scheme 6-4.** An alternative dihydropyrrole biosynthetic pathway in sibiromycin that accounts for the SibS-catalyzed depurinase activity in analogy to PhnI-catalyzed depurinase activity. To extend the analogy, this pathway needs an enzyme complex to catalyze the transfer of **1.1** onto the ribosyl acceptor where SibZ can methylate the vinyl group and subsequent enzymes can further tailor the dihydropyrrole ring.

**The search for genes involved in PBD biosynthesis outside of the gene cluster.**

The search for MLPs inadvertently led to the discovery of a new PBD gene cluster in *N. brasiliensis*. *N. brasiliensis* not only has the genetic potential to express a PBD, it has a fully sequenced genome unlike the other PBD producers. This facilitates the search for biosynthetically relevant genes missing within the PBD gene cluster. For example, the assignment of SibS as a hydrolase remains speculative, especially given its absence in hormaomycin biosynthesis, so other hydrolases predicted to be coded for by genes outside the gene cluster could be assayed for activity. In addition, the gene that codes for the phosphopantetheinyl transferase is missing in all PBD gene clusters but present in *N. brasiliensis* outside its PBD gene cluster (NCBI accession number WP_041564023). This gene is expected to code for the protein that post-translationally modifies the thiolation

domains of NRPSs responsible for assembling the PBDs. Testing of such genes will be crucial to engineer new PBDs in the future.

# Appendices

## Appendix A. Supporting Information for Chapter 2

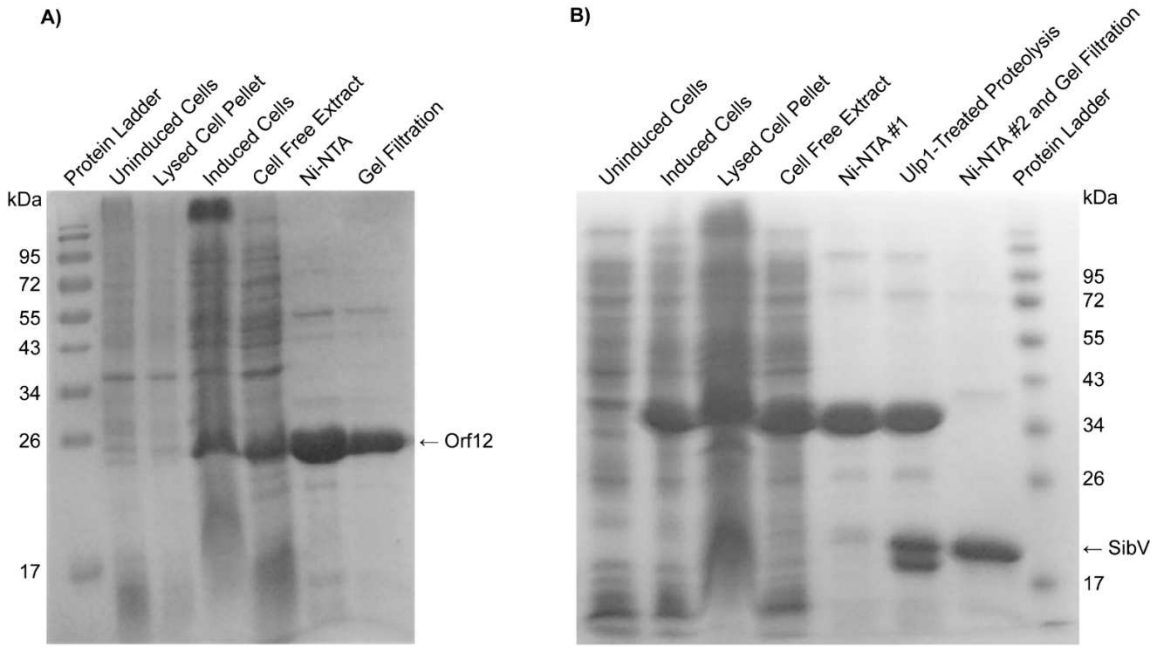



**Figure A-1.** SDS-PAGE gel images of PBD dioxygenases **A)** Orf12 and **B)** SibV at various expression and purification stages.

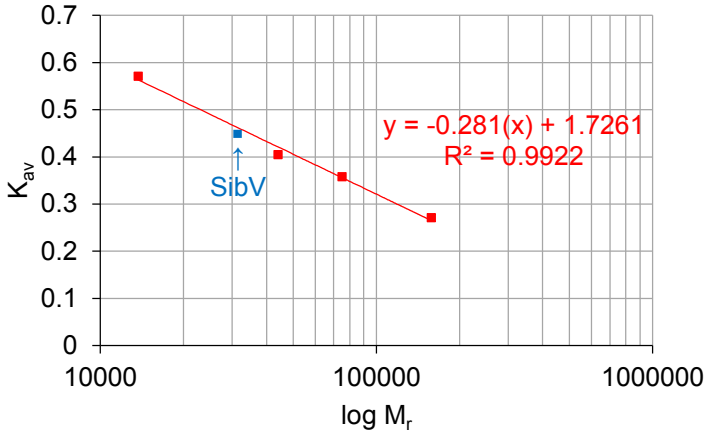



**Figure A-2.** The oligomeric state of SibV was determined by size exclusion chromatography. The experimentally determined partition coefficient $K_{av}$ of SibV, illustrated in blue, was superimposed onto the calibration curve, illustrated in red. Molecular weight standards ribonuclease A (13.7 kDa), ovalbumin (44.0 kDa), conalbumin (75.0 kDa) and aldolase (158.0 kDa) were used to establish the calibration curve.

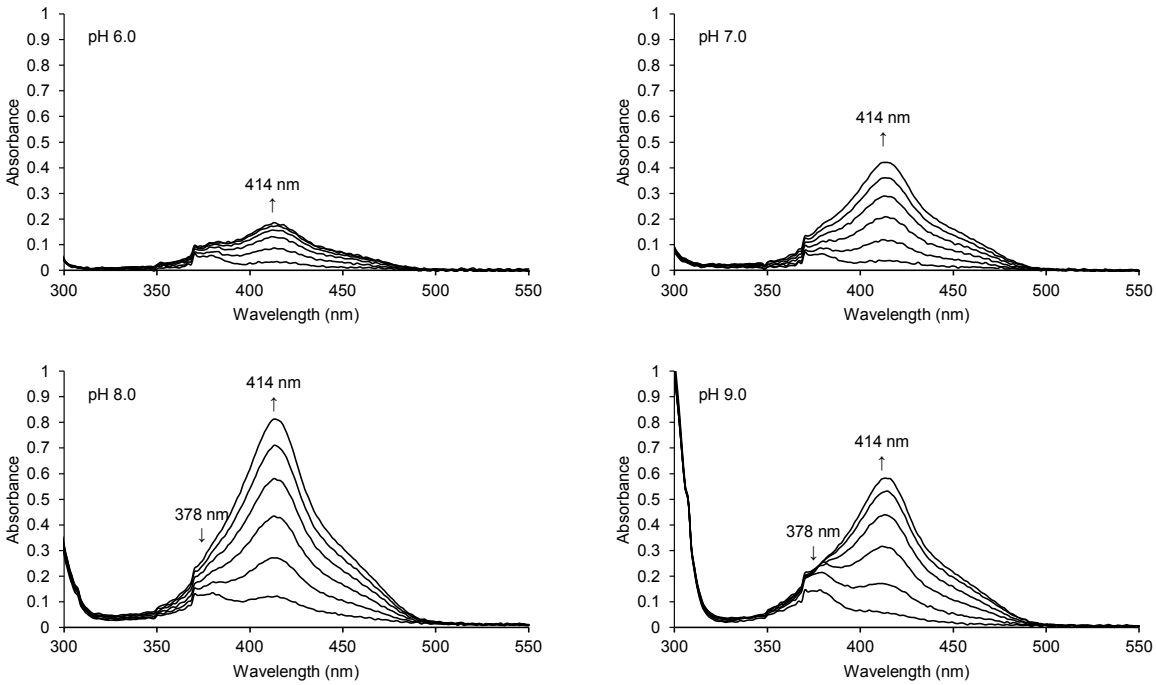



**Figure A-3.** UV-visible spectra during dioxygenation of L-DOPA catalyzed by SibV at the indicated pHs. Reaction between L-DOPA (1.0 mM) and SibV (4.9 μM) was performed in sodium phosphate (250 mM) at pH 6.0, 7.0 and 8.0 or glycine-NaOH (250 mM) at pH 9.0. Scans were recorded every 20 sec for a total of 120 sec. The $A_{max}$ at 414 nm increased and, when detected, the $A_{max}$ at 378 nm increased and subsequently decreased over time.

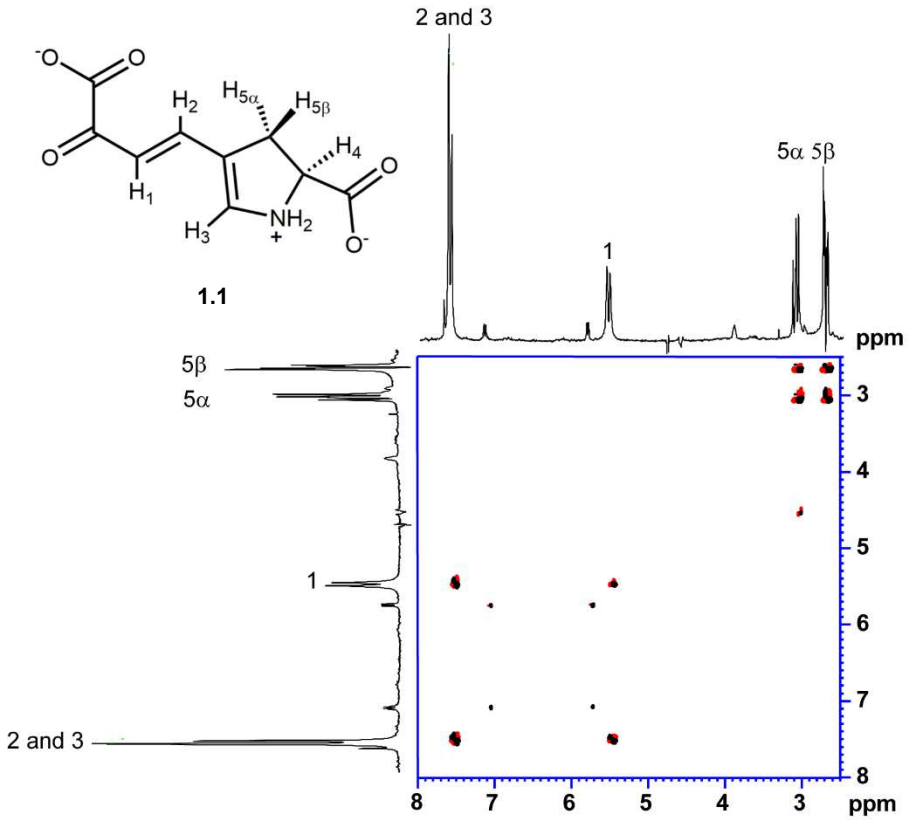



**Figure A-4**. Annotated COSY (400 MHz, 10% $D_2O$, $H_2O$ suppression) spectrum of **1.1** formed by dioxygenation of L-DOPA catalyzed by Orf12.

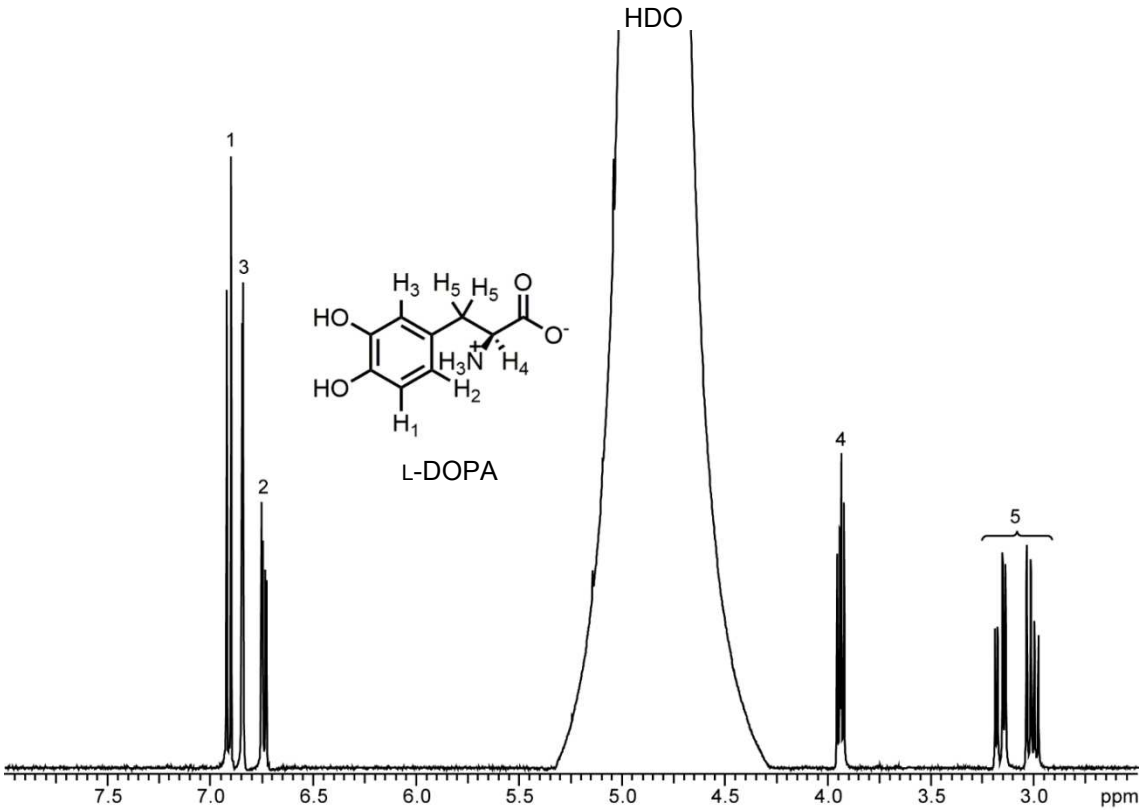



**Figure A-5.** Annotated $^{1}$H NMR (400 MHz, 10% $D_2O$, $H_2O$ suppression with presaturation) spectrum of L-DOPA in sodium phosphate (250 mM) at pH 8.0.

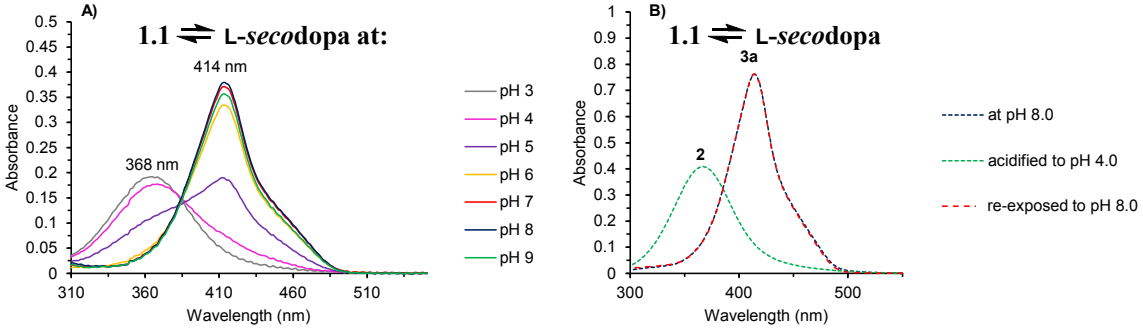



**Figure A-6.** UV-visible spectra monitoring interconversion of **1.1** and L-*seco*dopa, the reversible hydrolysis product formed under acidic conditions. **A)** Absorbance spectra were recorded when no more changes were observed after **1.1** (7.5 µM) was incubated at indicated pHs. Each absorbance spectrum represents an average of three independent determinations. **B) 1.1** (15.0 µM) in sodium phosphate (20 mM) at pH 8.0 was acidified to pH 4.0 and re-exposed to pH 8.0 to confirm the reversibility of the hydrolysis reaction. After each treatment, a UV-visible spectrum was recorded after no more changes were observed.

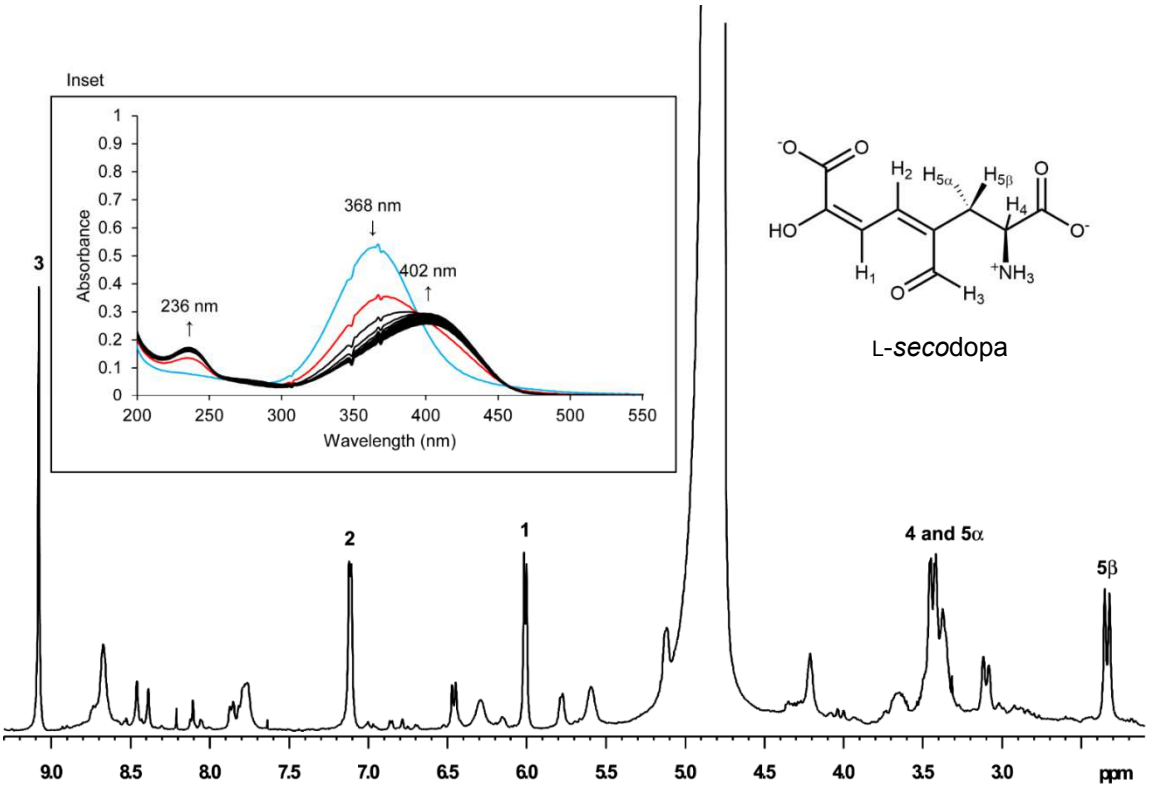



**Figure A-7.** Annotated $^{1}$H NMR spectrum of L-*seco*dopa formed by acid-catalyzed hydrolysis of **1.1** (400 MHz, 10% $D_2O$, $H_2O$ suppression with presaturation). The inset shows a UV-visible spectra of L-*seco*dopa and its decomposition at 1 h intervals including spectra, indicated in blue and red, that show the state of L-*seco*dopa during the first and last NMR scan, respectively.

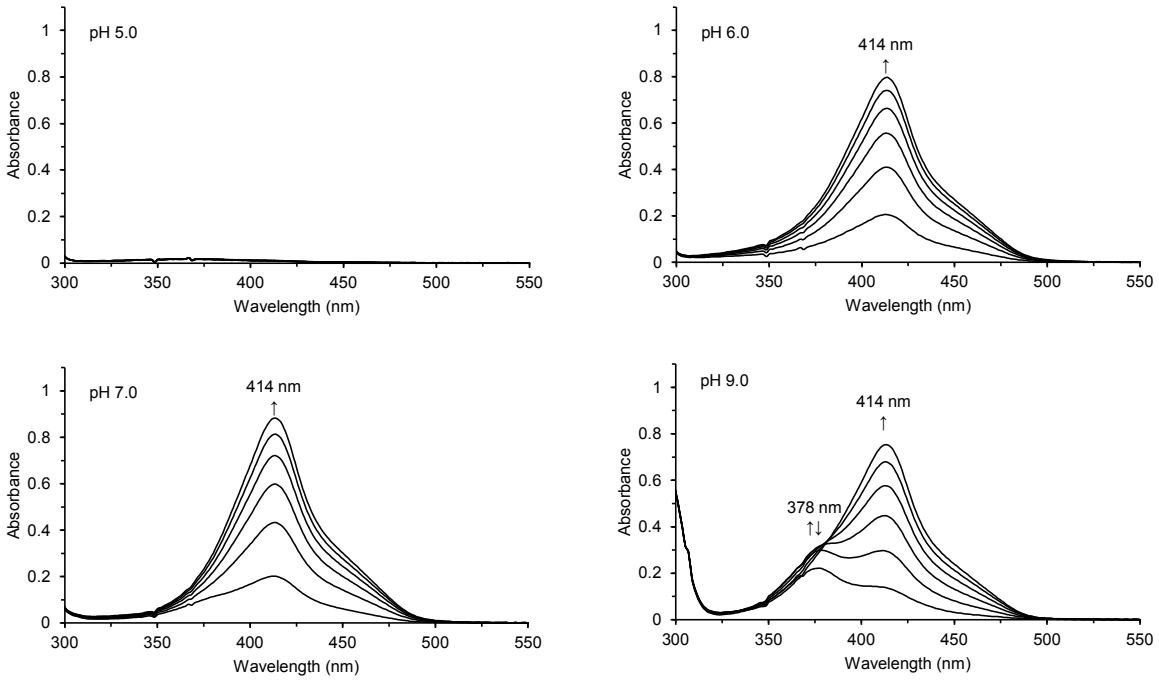



**Figure A-8.** UV-visible spectra during dioxygenation of L-DOPA catalyzed by Orf12 under the indicated pH conditions. Reaction between L-DOPA (1.0 mM) and Orf12 (4.9 μM) was performed in sodium phosphate-citrate (250 mM) at pH 5.0, sodium phosphate (250 mM) at pH 6.0 and 7.0 or glycine-NaOH (250 mM) at pH 9.0. Scans were recorded every 20 sec for a total of 120 sec. The $A_{max}$ at 414 nm increased and, when detected, the $A_{max}$ at 378 nm increased and subsequently decreased over time.

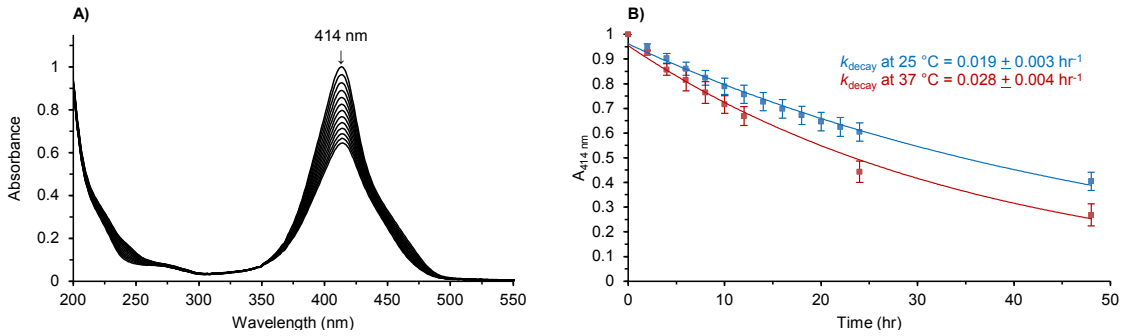



**Figure A-9.** Lifetime of **1.1** determined by visible spectroscopy. **A)** UV-visible spectra of **1.1** (20 μM) measured at 2 h intervals for 24 hours at 25 °C in sodium phosphate (20 mM) at pH 8.0. Each scan represents the average of three independent trials. **B)** $A_{414\ nm}$ vs. time at 25 and 37 °C as measured from the spectra in (A). Each data point represents an average absorbance value at 414 nm. The error bars represents one standard deviation of uncertainty. The solid line represents the best fit of the data to a first order decay.

## Appendix B. Supporting Information for Chapter 3

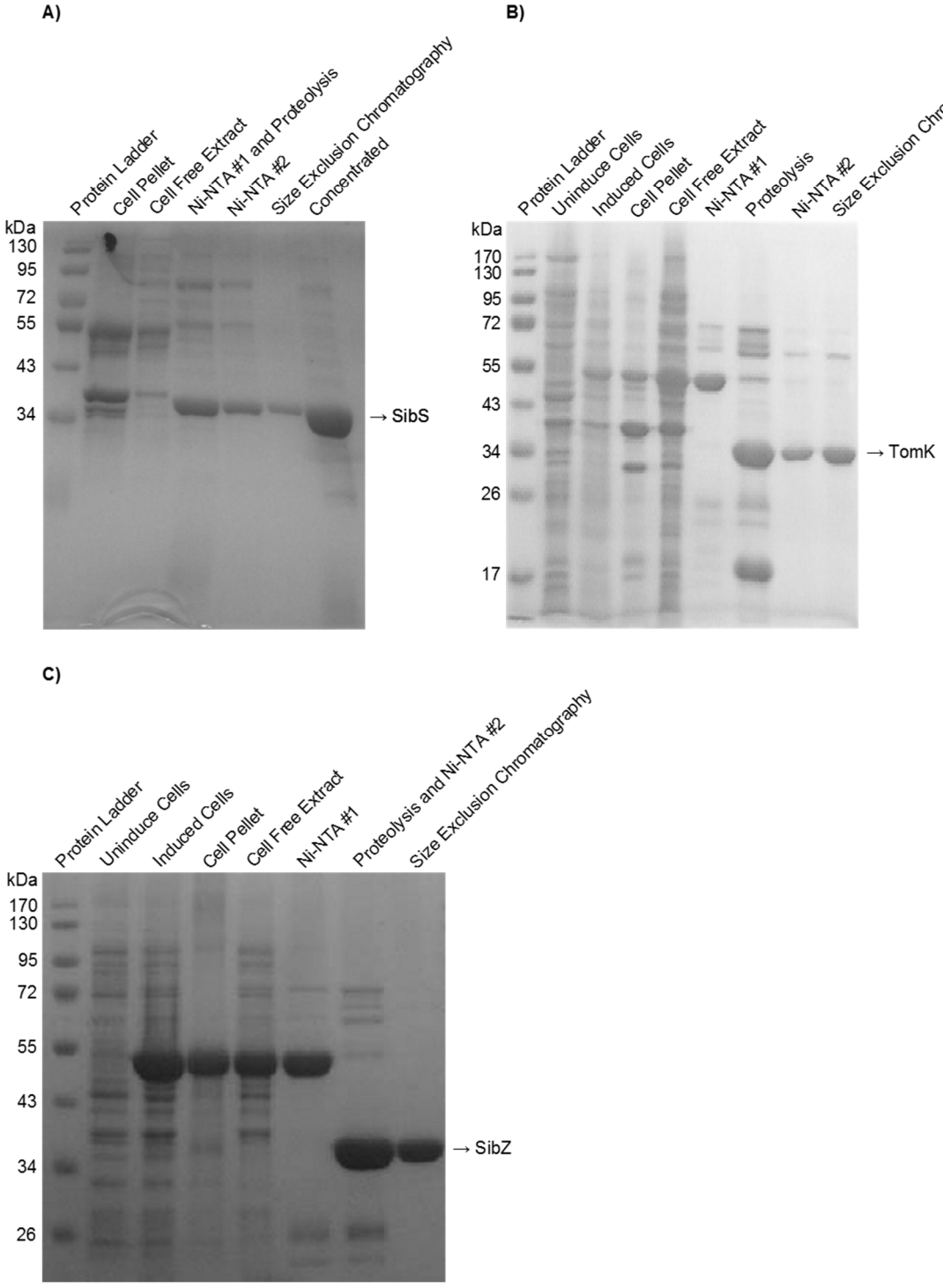
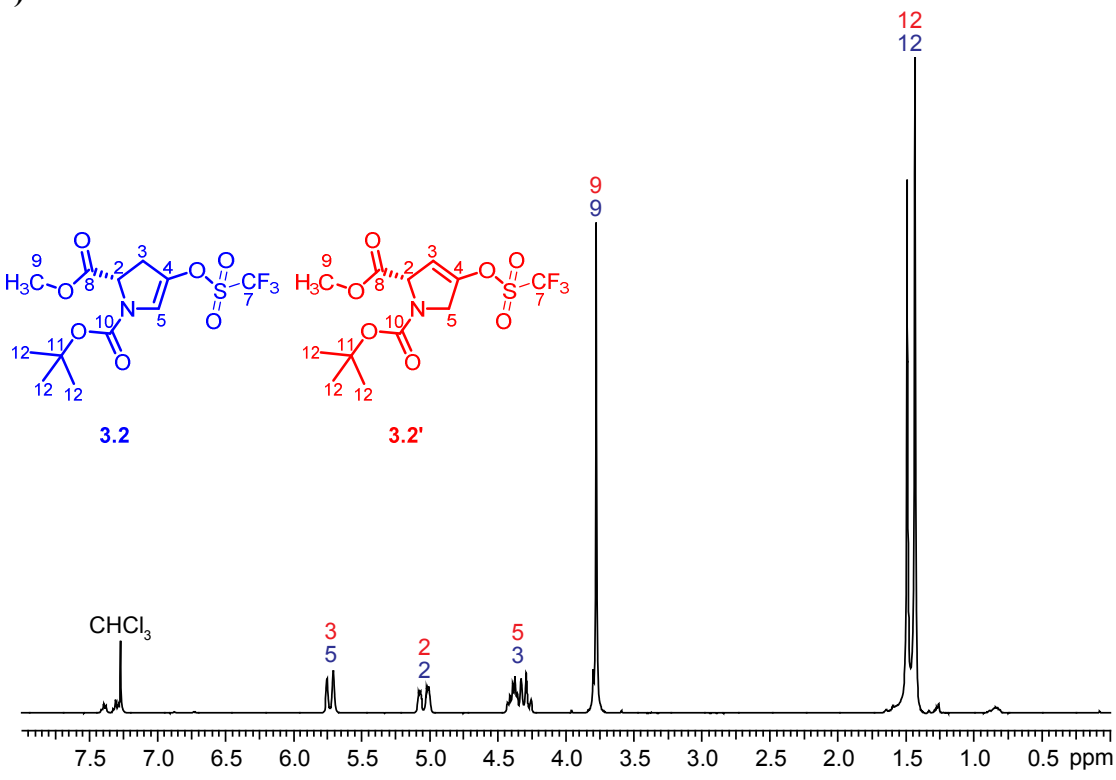
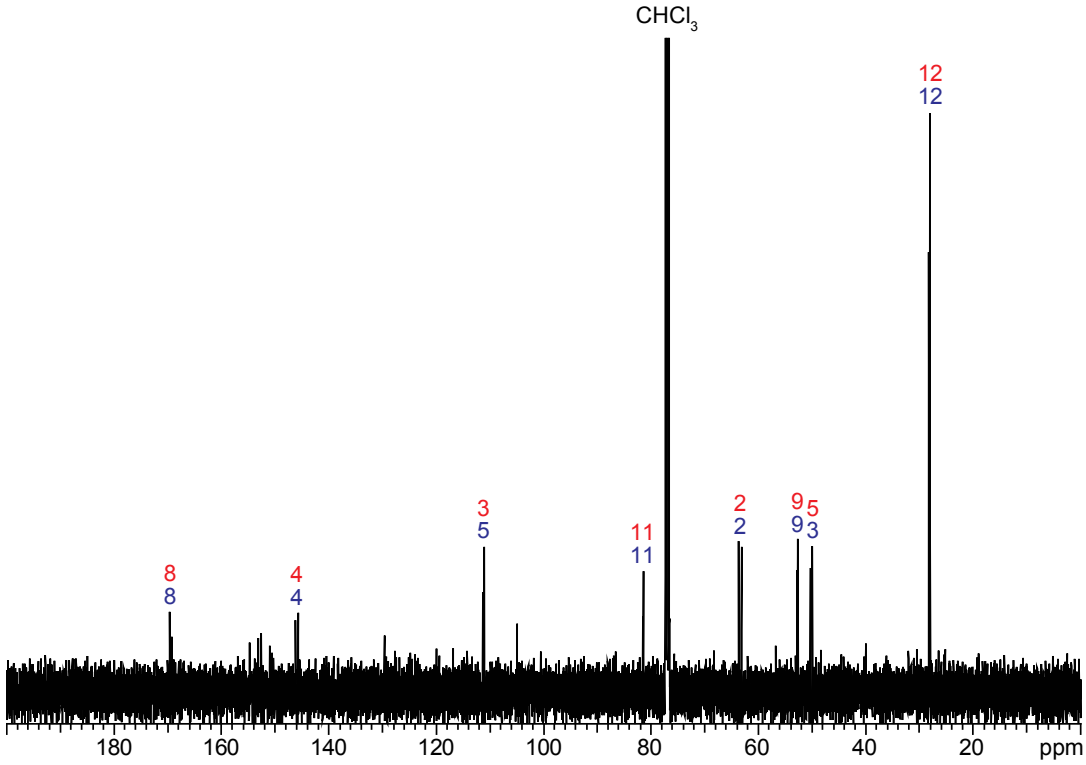



**Figure B-1.** SDS-PAGE gel images of **A)** SibS **B)** TomK and **C)** SibZ at various expression and purification stages.

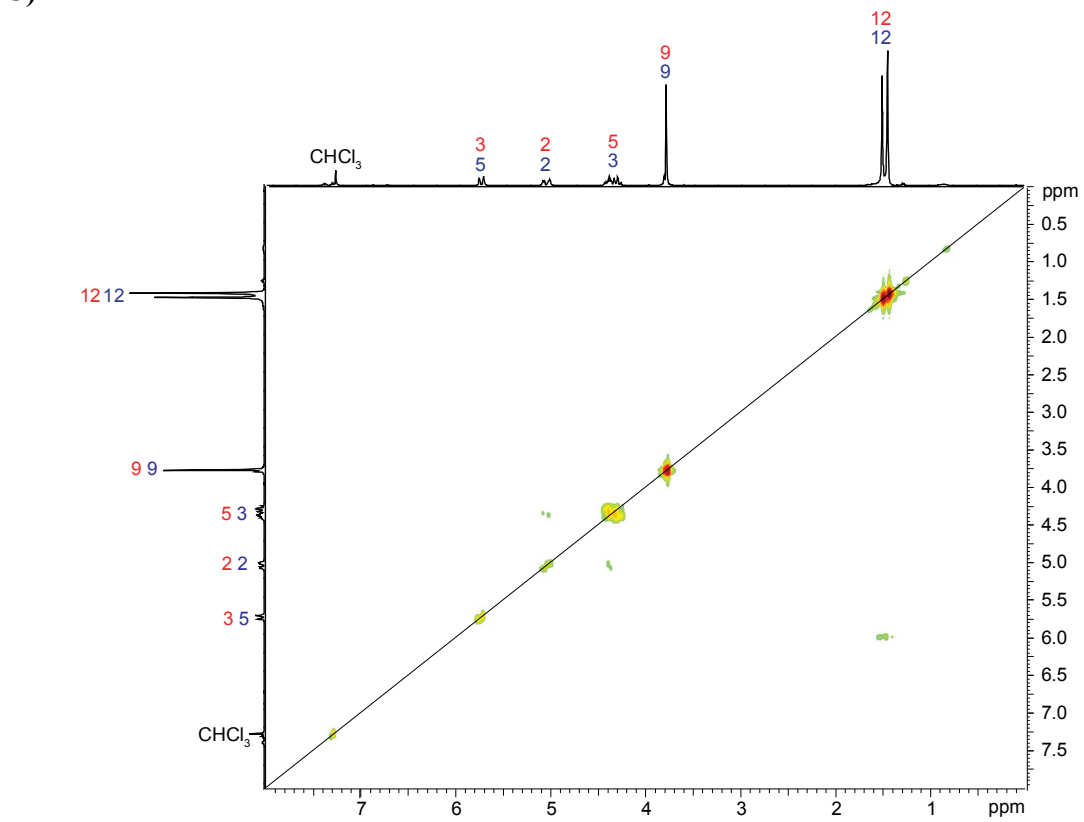
**A)**


12
12
9
9
3.2
3.2'
CHCl3
3
5
2
2
5
3
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

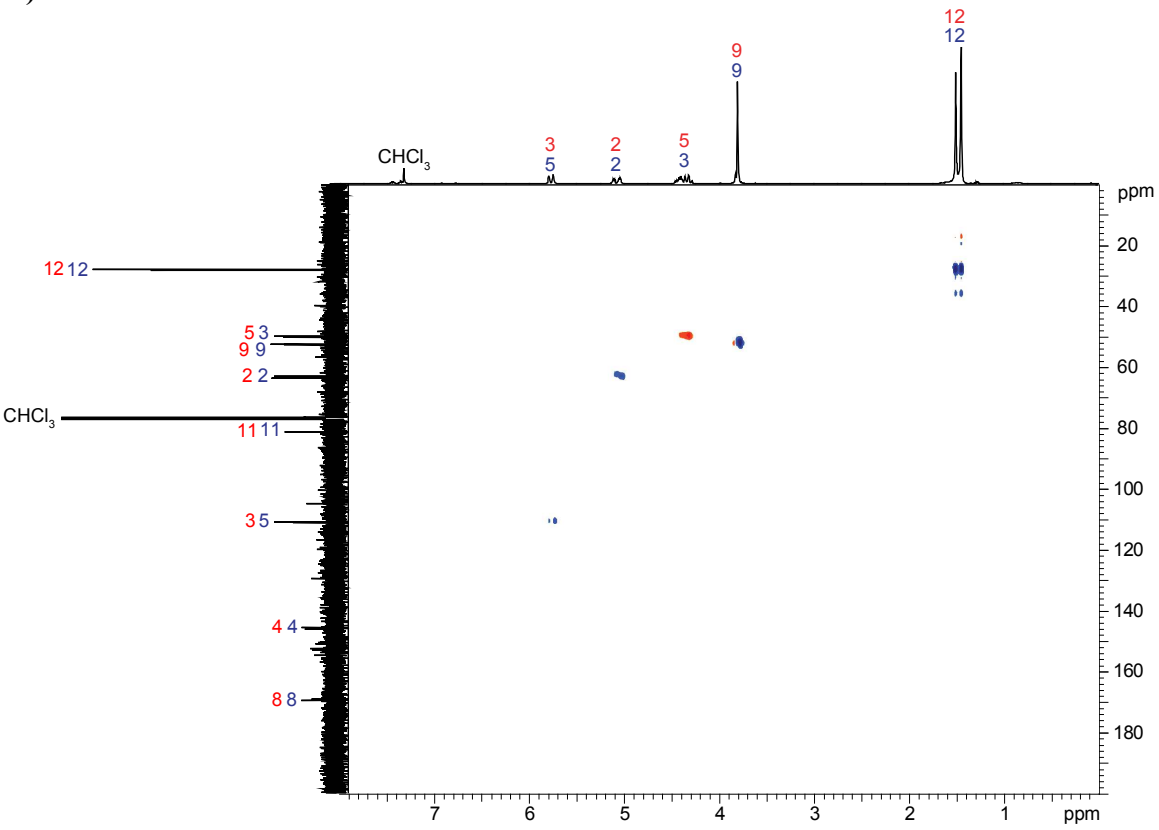
**B)**


CHCl3
12
12
3
5
11
11
2
2
9
9
5
3
8
8
4
4
180 160 140 120 100 80 60 40 20 ppm

**C)**


12
12
9
9
CHCl3
3
5
2
2
5
3
ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
12 12
9 9
5 3
2 2
3 5
CHCl3
7
6
5
4
3
2
1
ppm

**D)**


12
12
9
9
CHCl3
3
5
2
2
5
3
ppm
20
40
60
80
100
120
140
160
180
12 12
5 3
9 9
2 2
CHCl3
11 11
3 5
4 4
8 8
7
6
5
4
3
2
1
ppm

**E)**

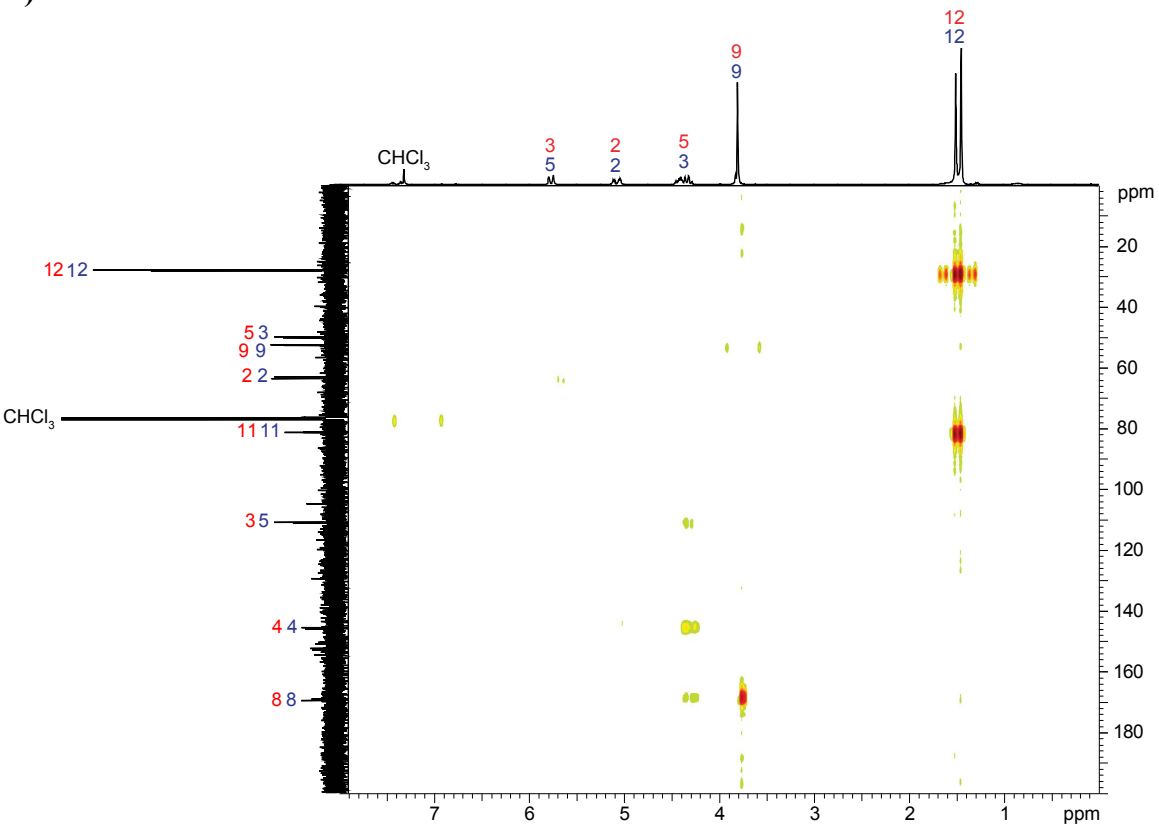



**Figure B-2.** Annotated **A)** $^{1}$H NMR spectrum, **B)** $^{13}$C NMR spectrum, **C)** COSY, **D)** HSQC and **E)** HMBC of *N*-Boc-4-triflate-4,5-dehydro-L-proline methyl ester **3.2** and *N*-Boc-4-triflate-3,4-dehydro-L-proline methyl ester **3.2'** in $CDCl_3$.

**Table B-1.** $^{13}$C and $^{1}$H NMR of *N*-Boc-4-triflate-4,5-dehydro-L-proline methyl ester **3.2** and *N*-Boc-4-triflate-3,4-dehydro-L-proline methyl ester **3.2'**.

| Position | | $^{13}$C shift (ppm) | $^{1}$H shift (ppm) | HMBC Connectivity |
|---|---|---|---|---|
| **3.2** | **3.2'** | | | |
| 2 | 2 | 63.7, 63.1 | 5.08, 5.01 (t(2), 1H) | 3, (4, 4), 5 |
| 3 | 5 | 50.3, 50.0 | 4.47-4.22 (m, *J*=14, 6.3, 2H) | 3, (4, 4), 5, (8, 8) |
| 4 | 4 | 146.3, 145.8 | | (2, 2), 3, 5 |
| 5 | 3 | 111.3, 111.1 | 5.75, 5.71 (s(2), 1H) | (2, 2), 3, 5 |
| 7 | 7 | | | |
| 8 | 8 | 169.6, 169.3 | | (3), (9, 9) |
| 9 | 9 | 52.8, 52.6 | 3.80, 3.78 (s(2), 3H) | (8, 8) |
| 10 | 10 | | | |
| 11 | 11 | 81.5, 81.4 | | (12, 12) |
| 12 | 12 | 28.3, 28.2 | 1.49, 1.44 (s(2), 9H) | (11, 11), (12, 12) |

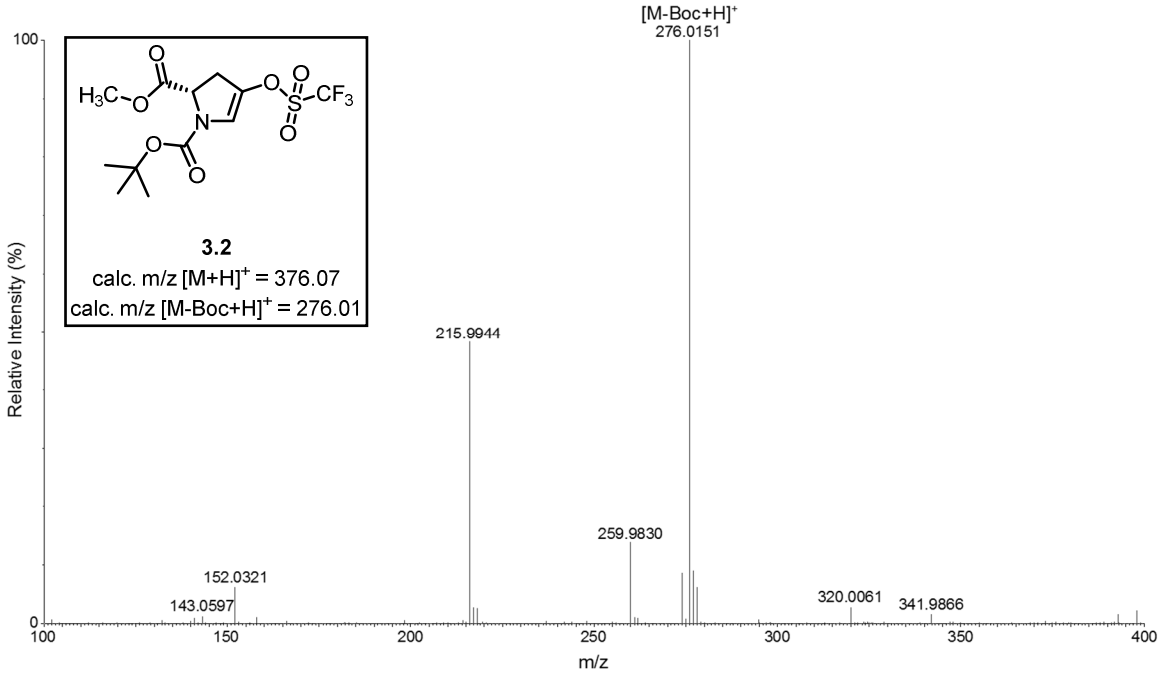
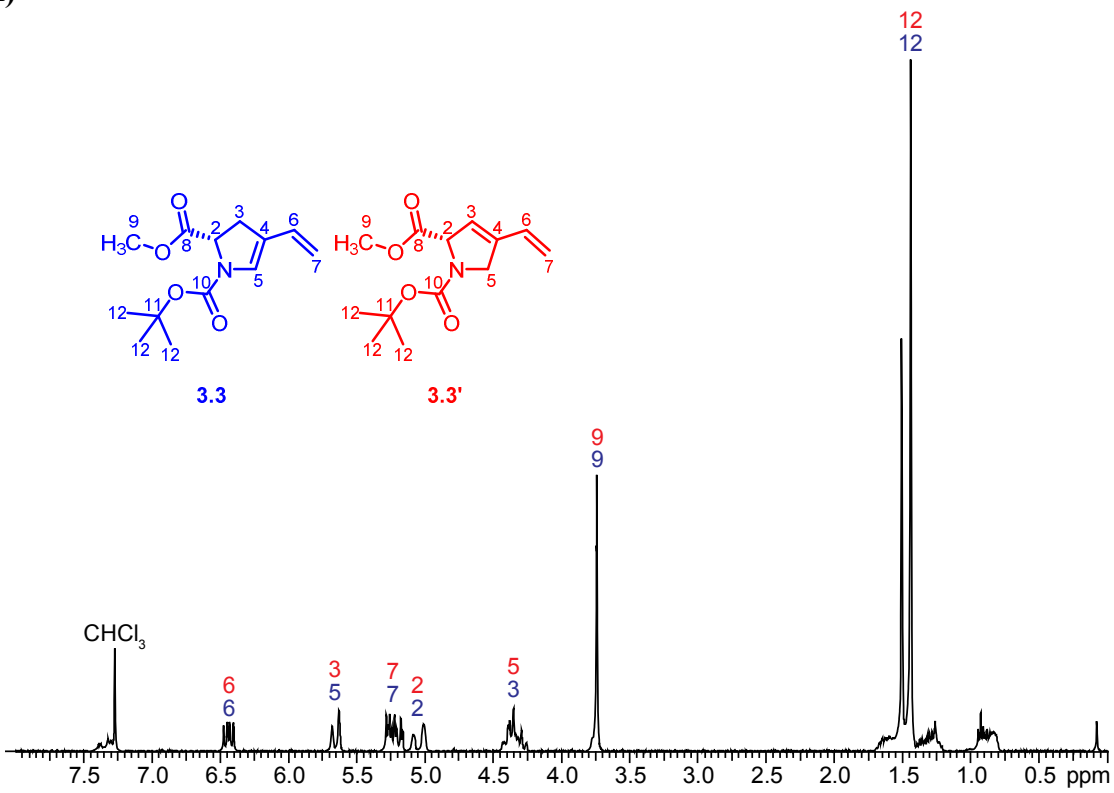
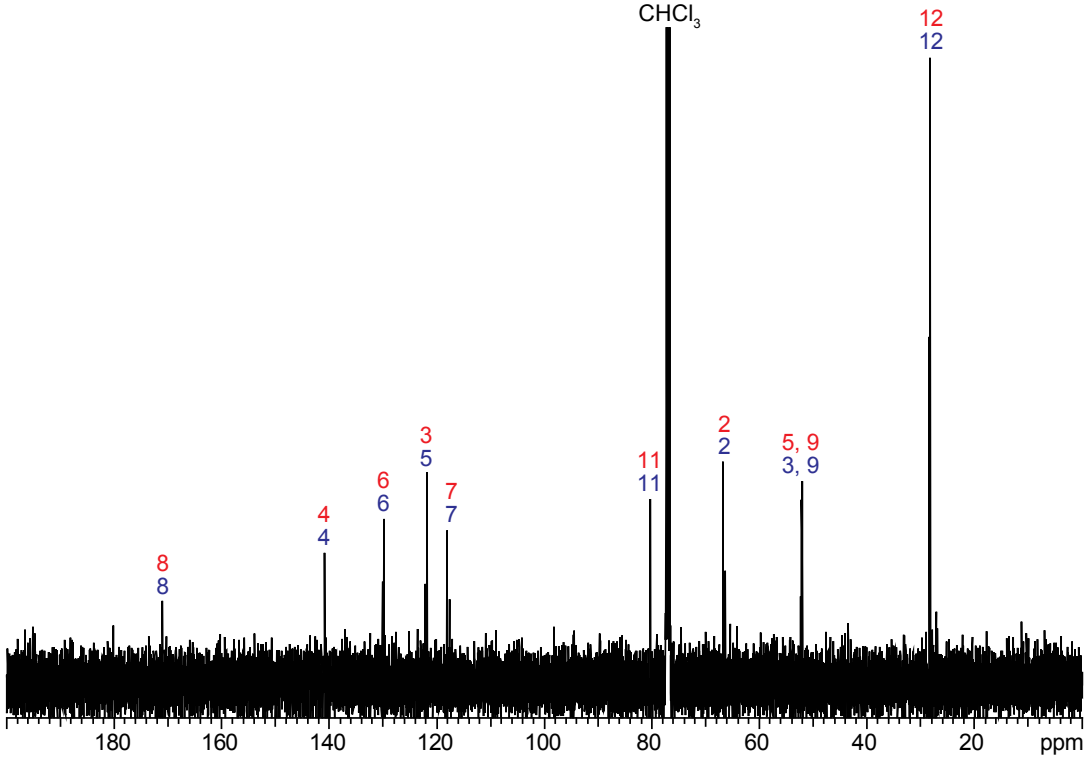



**Figure B-3.** ESI$^+$-MS of *N*-Boc-4-triflate-4,5-dehydro-L-proline methyl ester **3.2**. The fragment ion $[M\text{-}Boc+H]^+$ is detected.

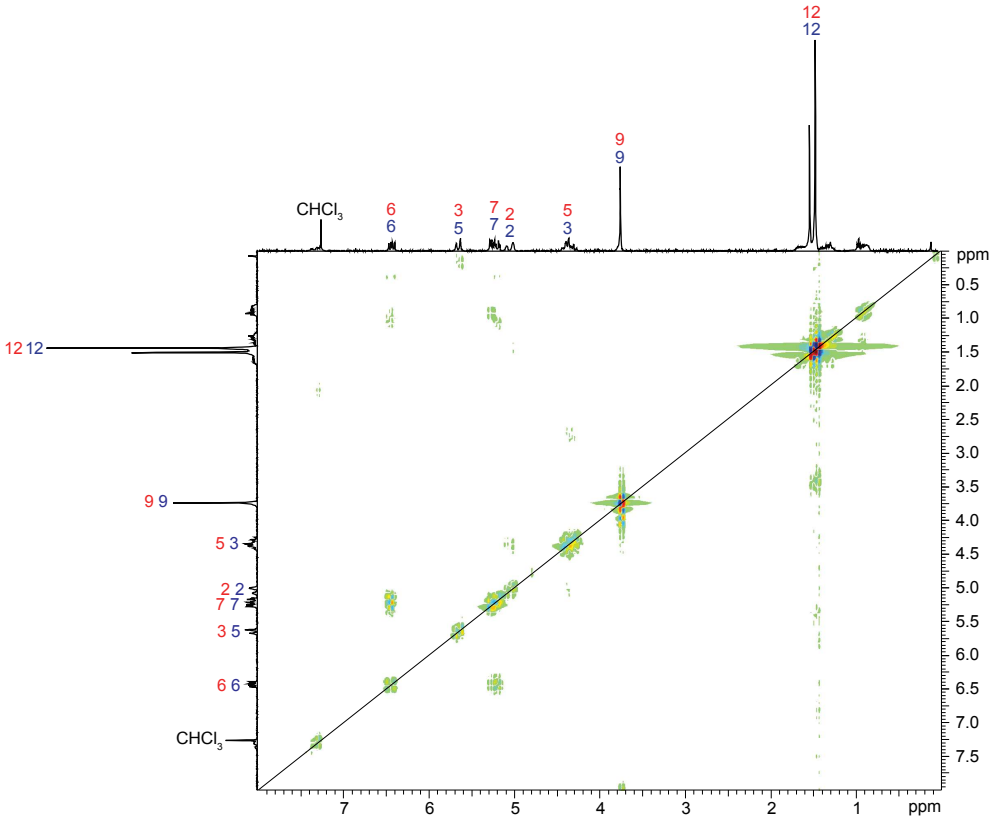
**A)**


3.3
3.3'
CHCl3
12
12
9
9
6
6
3
5
7
7
2
2
5
3
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

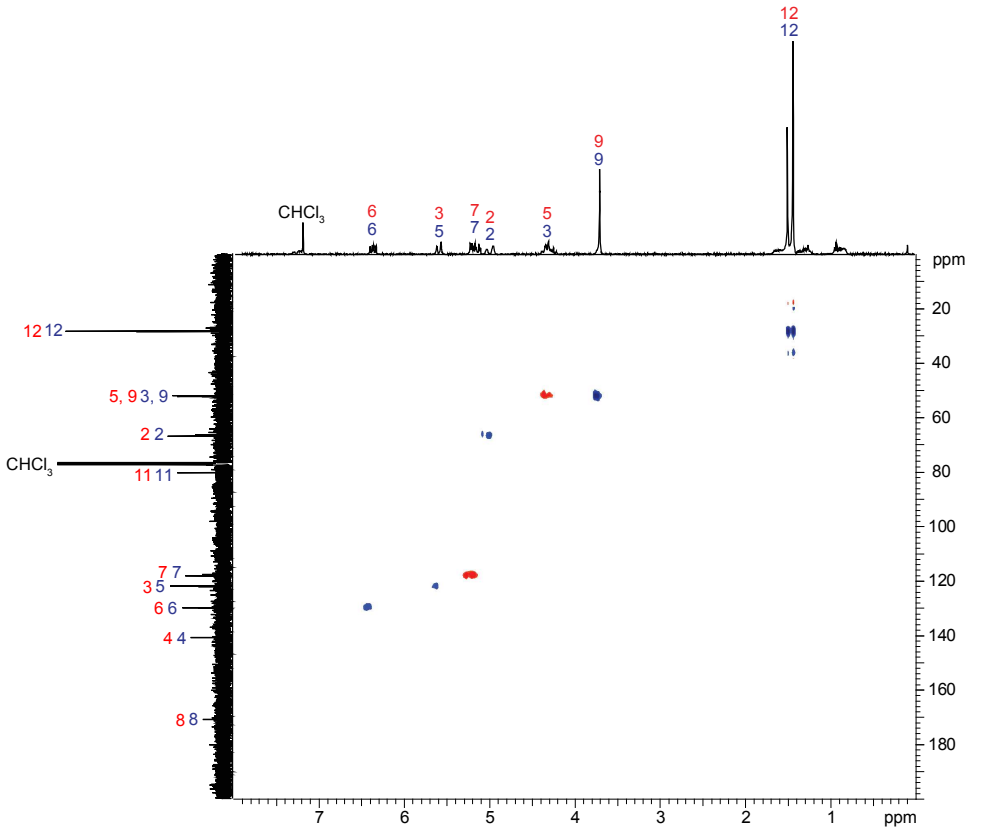
**B)**


CHCl3
12
12
2
2
5, 9
3, 9
11
11
3
5
6
6
7
7
4
4
8
8
180 160 140 120 100 80 60 40 20 ppm

**C)**


12
12
9
9
CHCl3
6
6
3
5
7
7
2
2
5
3
ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
12 12
9 9
5 3
2 2
7 7
3 5
6 6
CHCl3
7
6
5
4
3
2
1
ppm

**D)**


12
12
9
9
CHCl3
6
6
3
5
7
7
2
2
5
3
ppm
20
40
60
80
100
120
140
160
180
12 12
5, 9 3, 9
2 2
CHCl3
11 11
7 7
3 5
6 6
4 4
8 8
7
6
5
4
3
2
1
ppm

**E)**

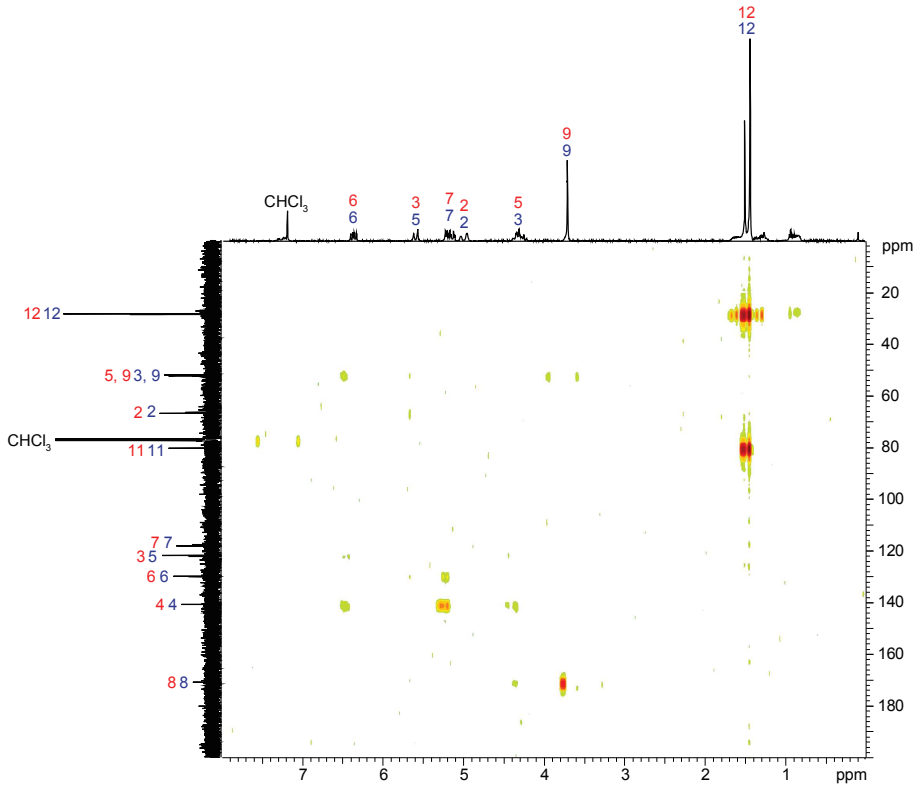



**Figure B-4.** Annotated **A)** $^{1}$H NMR spectrum, **B)** $^{13}$C NMR spectrum, **C)** COSY, **D)** HSQC and **E)** HMBC of *N*-Boc-4-vinyl-4,5-dehydro-L-proline methyl ester **3.3** and *N*-Boc-4-vinyl-3,4-dehydro-L-proline methyl ester **3.3'** in $CDCl_3$.

**Table B-2.** $^{13}$C and $^{1}$H NMR of *N*-Boc-4-vinyl-4,5-dehydro-L-proline methyl ester **3.3** and *N*-Boc-4-vinyl-3,4-dehydro-L-proline methyl ester **3.3'**.

| Position | | $^{13}$C shift (ppm) | $^{1}$H shift (ppm) | HMBC Connectivity |
|---|---|---|---|---|
| **3.3** | **3.3'** | | | |
| 2 | 2 | 66.8, 66.4 | 5.09, 5.01 (s(2), 1H) | 3, 5 |
| 3 | 5 | 52.2, 52.0 | 4.45-4.24 (m, 2H) | (2, 2), 3, (4, 4), 5, (6, 6), (8, 8) |
| 4 | 4 | 140.9 | | 3, 5, (6, 6), (7, 7) |
| 5 | 3 | 122.1, 121.8 | 5.68, 5.63 (s(2), 1H) | (2, 2), 3, 5, (6, 6) |
| 6 | 6 | 130.0, 129.8 | 6.50-6.38 (m, 1H) | (3, 3) , (4, 4), (5, 5), (7, 7) |
| 7 | 7 | 118.2, 117.7 | 5.32-4.98 (m, 2H) | (4, 4), (6, 6) |
| 8 | 8 | 171.0 | | (3, 3), (9, 9) |
| 9 | 9 | 52.3, 52.2 | 3.74 (s, 3H) | (8, 8) |
| 10 | 10 | | | |
| 11 | 11 | 80.3 | | (12, 12) |
| 12 | 12 | 28.4, 28.3 | 1.51, 1.44 (s(2), 9H) | (11, 11), (12, 12) |

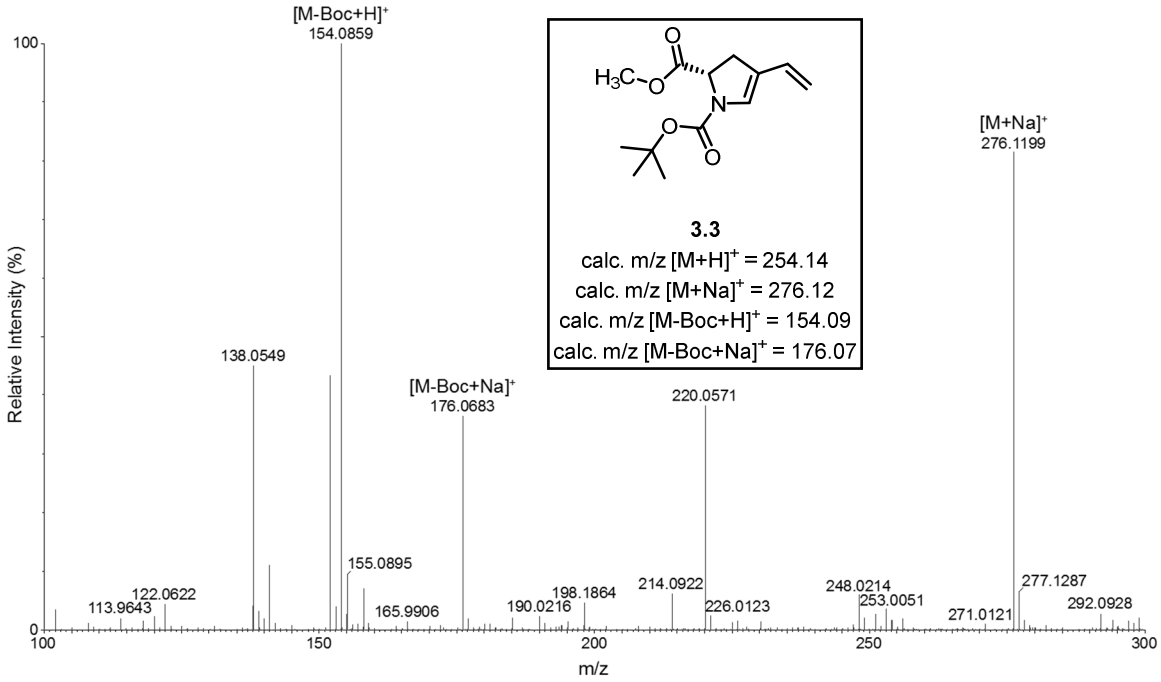
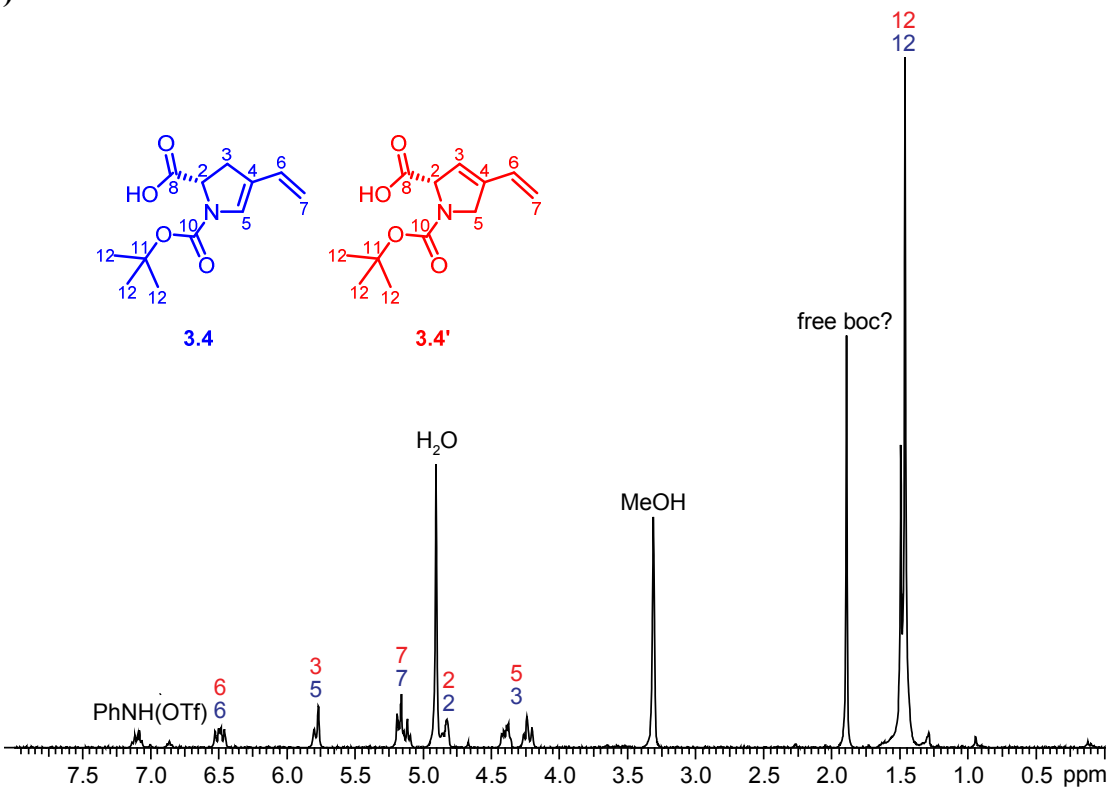
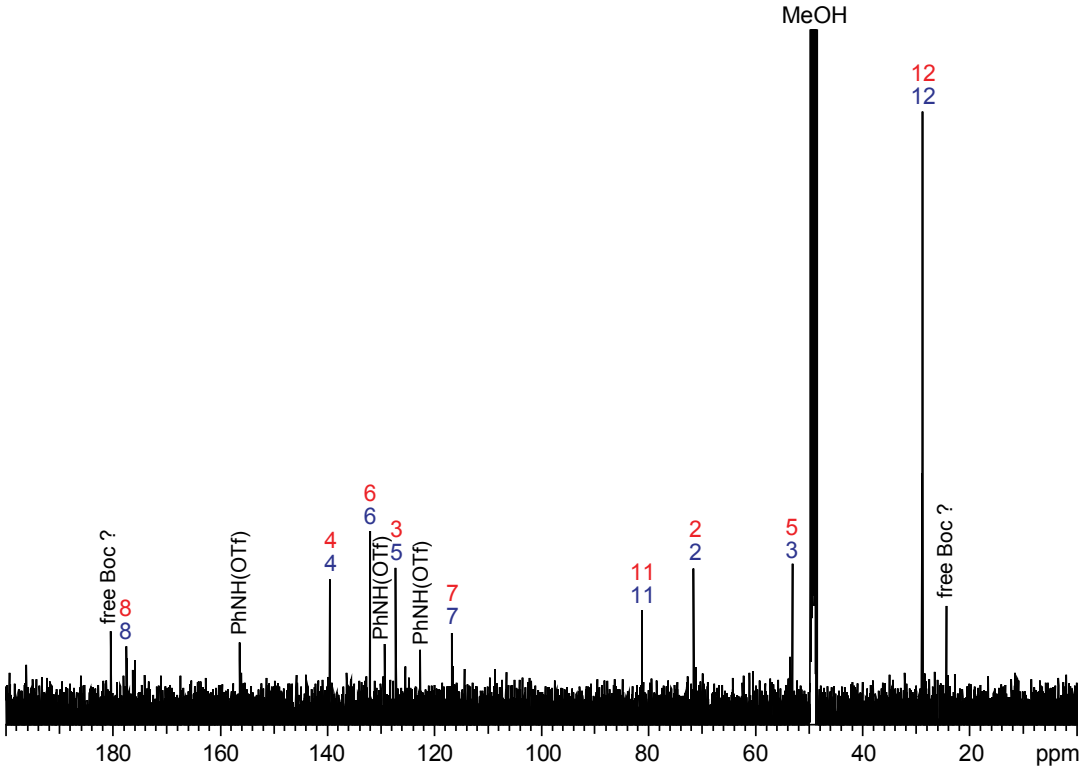



**Figure B-5.** ESI$^+$-MS of *N*-Boc-4-vinyl-4,5-dehydro-L-proline methyl ester **3.3**. The adduct $[M+Na]^+$ and fragment ions $[M\text{-}Boc+Na]^+$and $[M\text{-}Boc+H]^+$ are detected.

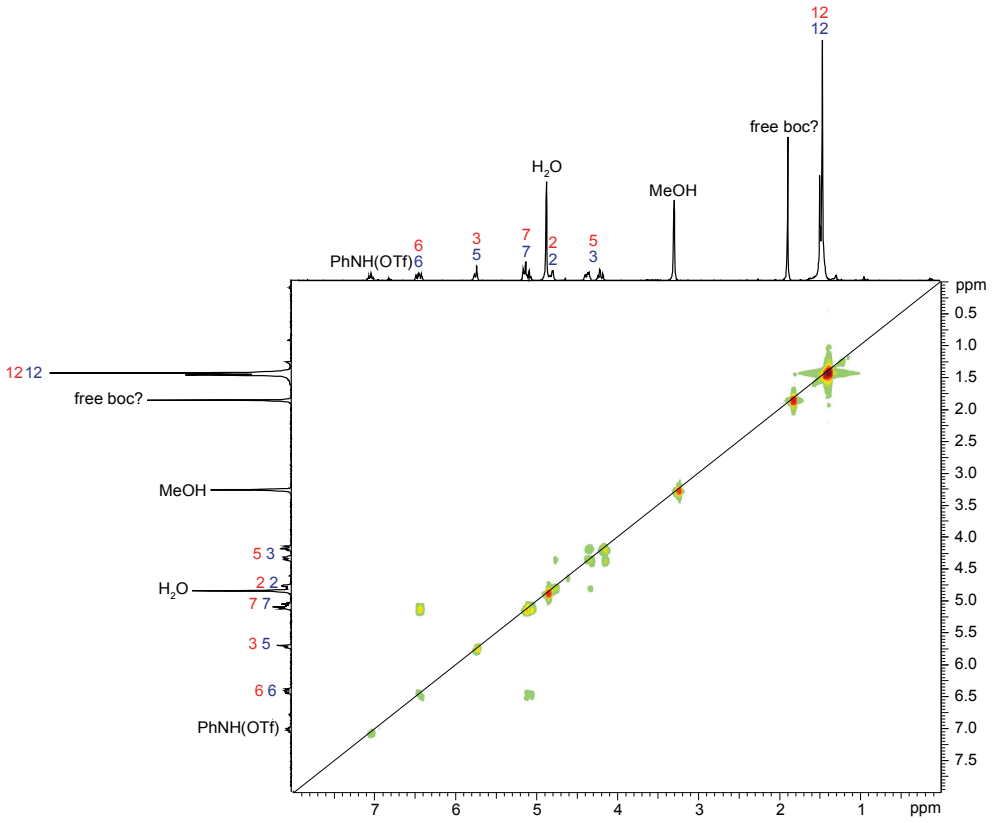
**A)**


3.4
3.4'
free boc?
H2O
MeOH
PhNH(OTf)
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

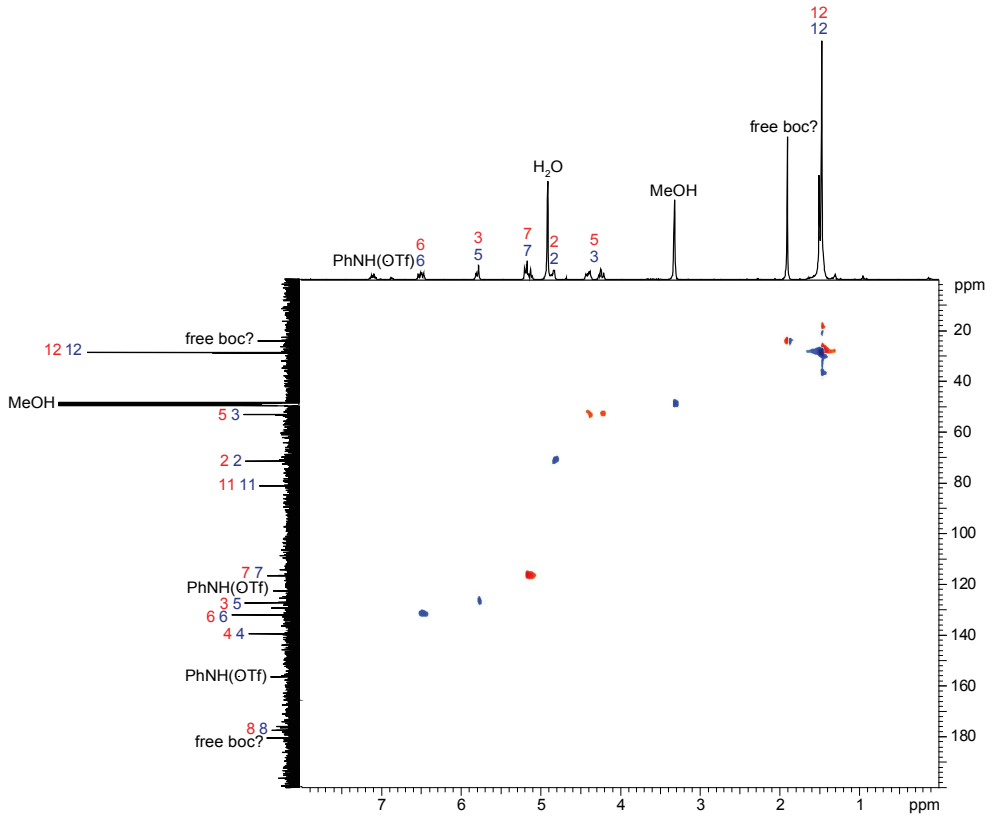
**B)**


MeOH
free Boc ?
PhNH(OTf)
PhNH(OTf)
PhNH(OTf)
free Boc ?
180 160 140 120 100 80 60 40 20 ppm

**C)**


12 12
free boc?
H2O
MeOH
PhNH(OTf) 6 6
3 5
7 7
2 2
5 3
ppm
12 12
free boc?
MeOH
5 3
H2O
2 2
7 7
3 5
6 6
PhNH(OTf)
7 6 5 4 3 2 1 ppm

**D)**


12 12
free boc?
H2O
MeOH
PhNH(OTf) 6 6
3 5
7 7
2 2
5 3
ppm
free boc?
12 12
MeOH
5 3
2 2
11 11
7 7
PhNH(OTf)
3 5
6 6
4 4
PhNH(OTf)
8 8
free boc?
7 6 5 4 3 2 1 ppm

**E)**

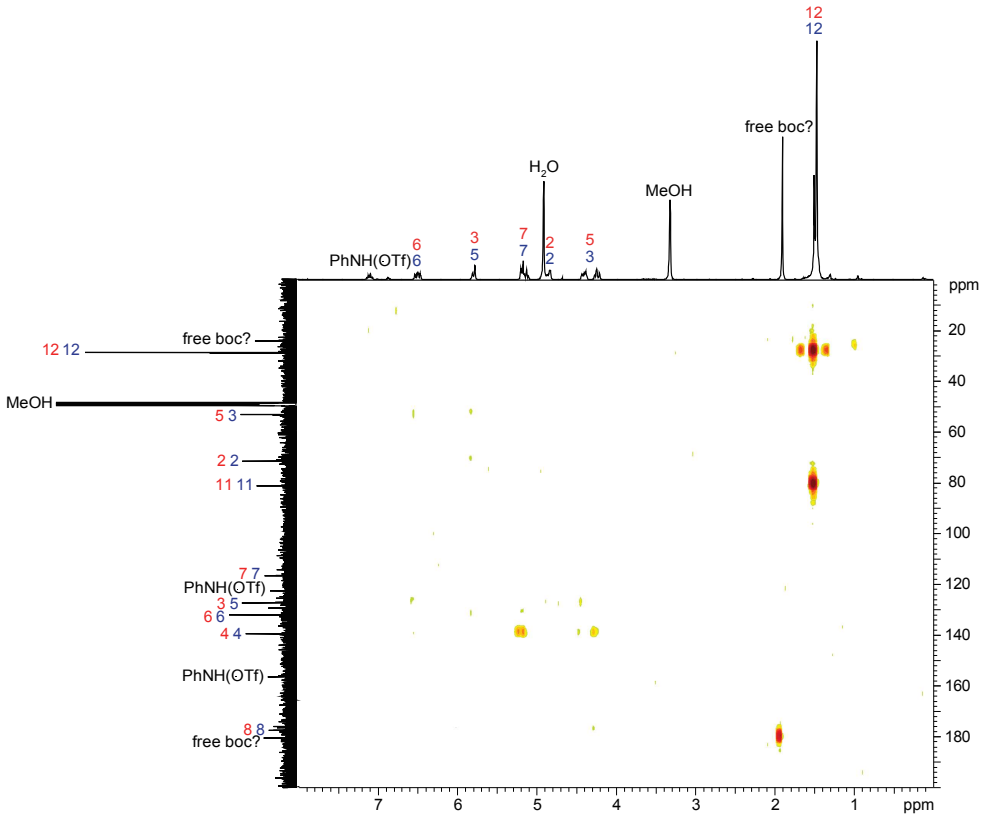



**Figure B-6.** Annotated **A)** $^{1}$H NMR spectrum, **B)** $^{13}$C NMR spectrum, **C)** COSY, **D)** HSQC and **E)** HMBC of *N*-Boc-4-vinyl-4,5-dehydro-L-proline **3.4** and *N*-Boc-4-vinyl-3,4-dehydro-L-proline **3.4'** in MeOD.

**Table B-3.** $^{13}$C and $^{1}$H NMR of *N*-Boc-4-vinyl-4,5-dehydro-L-proline **3.4** and *N*-Boc-4-vinyl-3,4-dehydro-L-proline **3.4'** (MeOD).

| Position | | $^{13}$C shift (ppm) | $^{1}$H shift (ppm) | HMBC Connectivity |
|---|---|---|---|---|
| **3.4** | **3.4'** | | | |
| 2 | 2 | 71.4 | 4.86, 4.83 (s(2), 1H) | 3, 5 |
| 3 | 5 | 53.0 | 4.44-4.19 (m, 2H) | 3, (4, 4), 5, (6, 6), (8, 8) |
| 4 | 4 | 139.4 | | 3, 5, (6, 6), (7, 7) |
| 5 | 3 | 127.1, 127.0 | 5.83-5.75 (m, 1H) | (2, 2), 3, 5, (6, 6) |
| 6 | 6 | 131.9 | 6.55-6.44 (m, 1H) | (3, 3) , (4, 4), (5, 5), (7, 7) |
| 7 | 7 | 116.6 | 5.23-5.08 (m, 2H) | (4, 4), (6, 6) |
| 8 | 8 | 177.3, 177.2 | | 3 |
| 10 | 10 | | | |
| 11 | 11 | 81.1 | | (12, 12) |
| 12 | 12 | 28.9, 28.7 | 1.49, 1.46 (s(2), 9H) | (11, 11), (12, 12) |

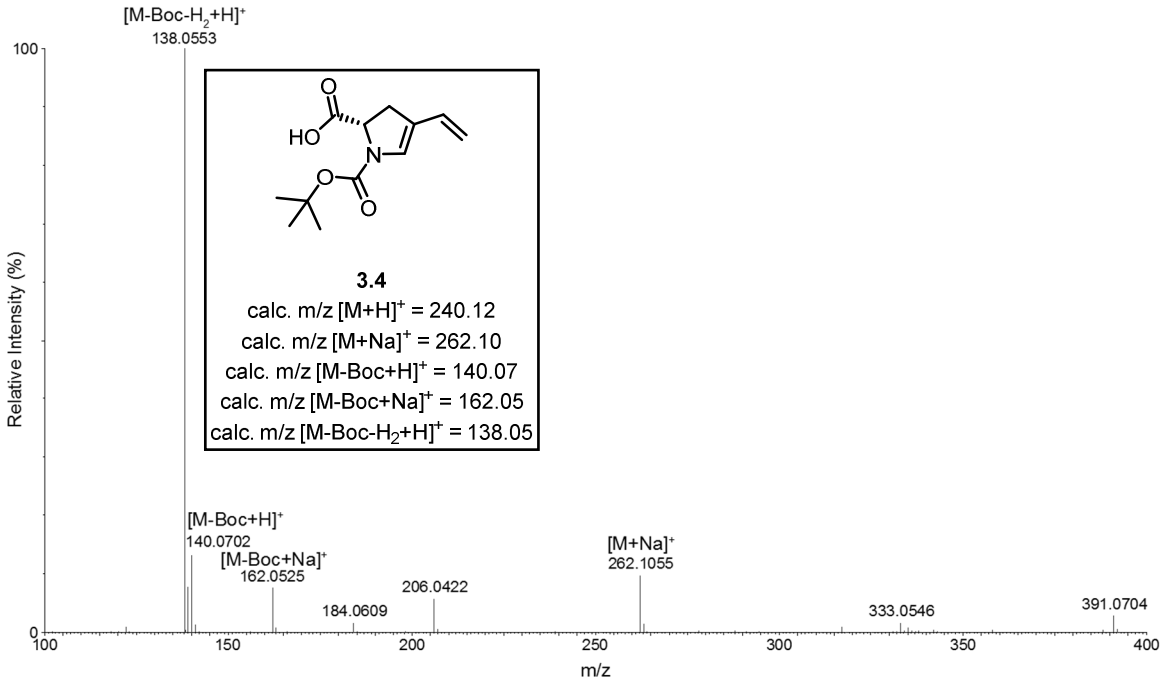
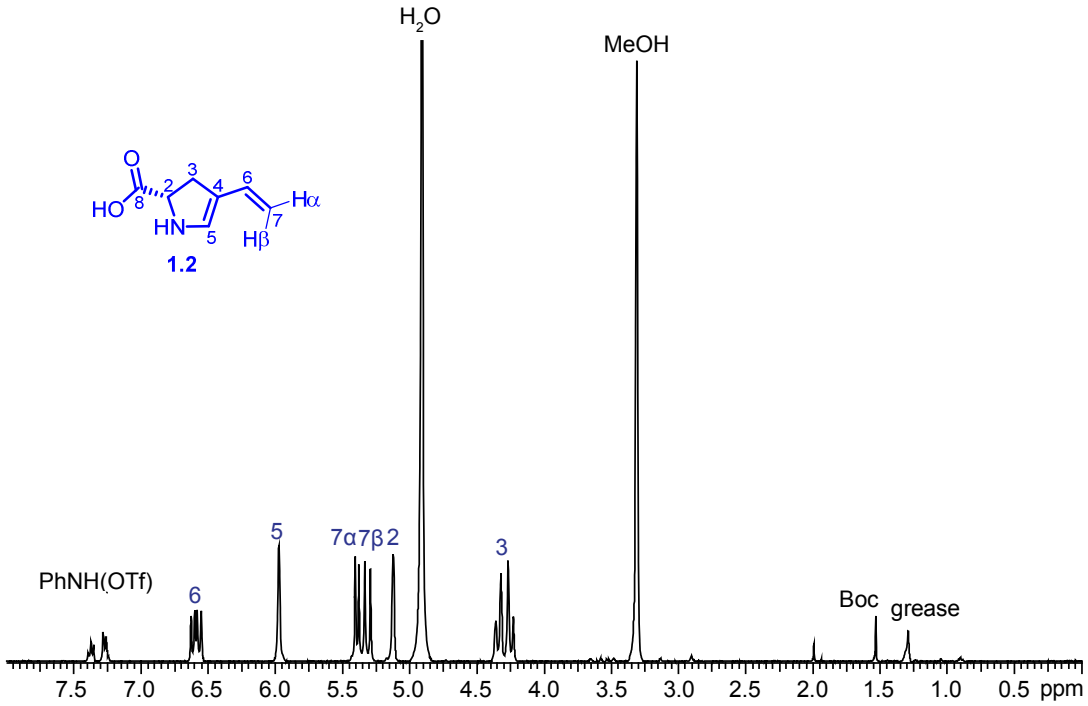
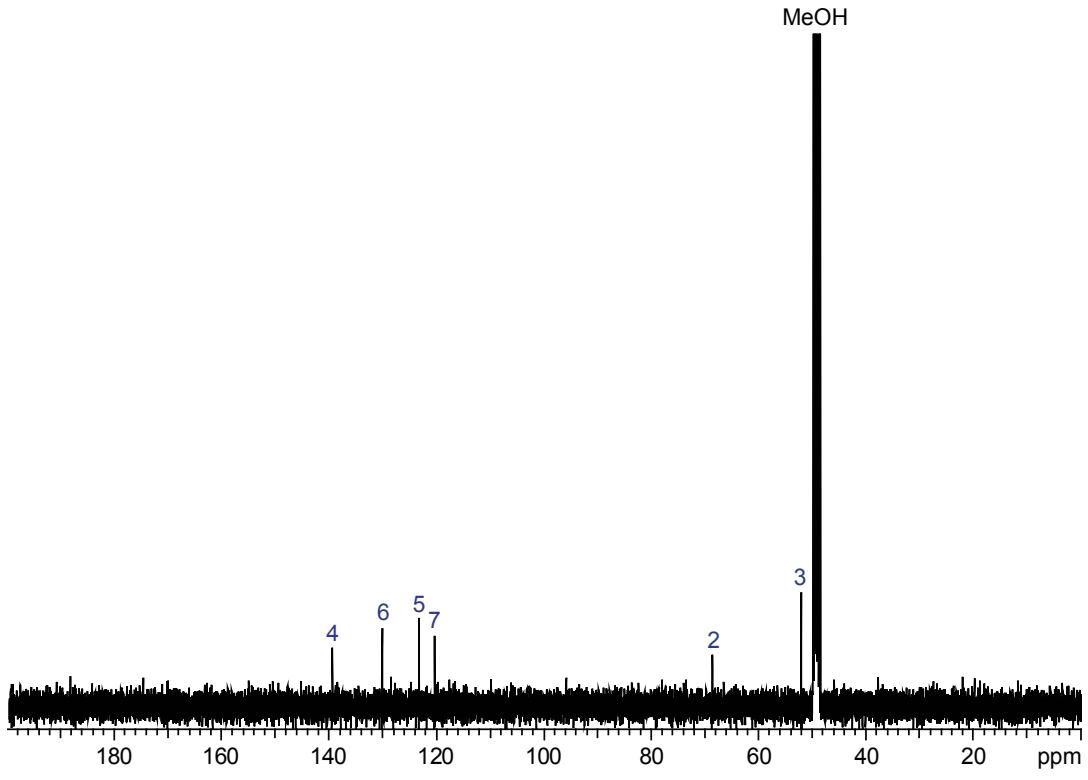



**Figure B-7.** ESI$^+$-MS of *N*-Boc-4-vinyl-4,5-dehydro-L-proline **3.4**. The adduct $[M+Na]^+$ and fragment ions $[M\text{-}Boc+Na]^+$, $[M\text{-}Boc+H]^+$ and $[M\text{-}Boc\text{-}H_2+H]^+$ are detected.

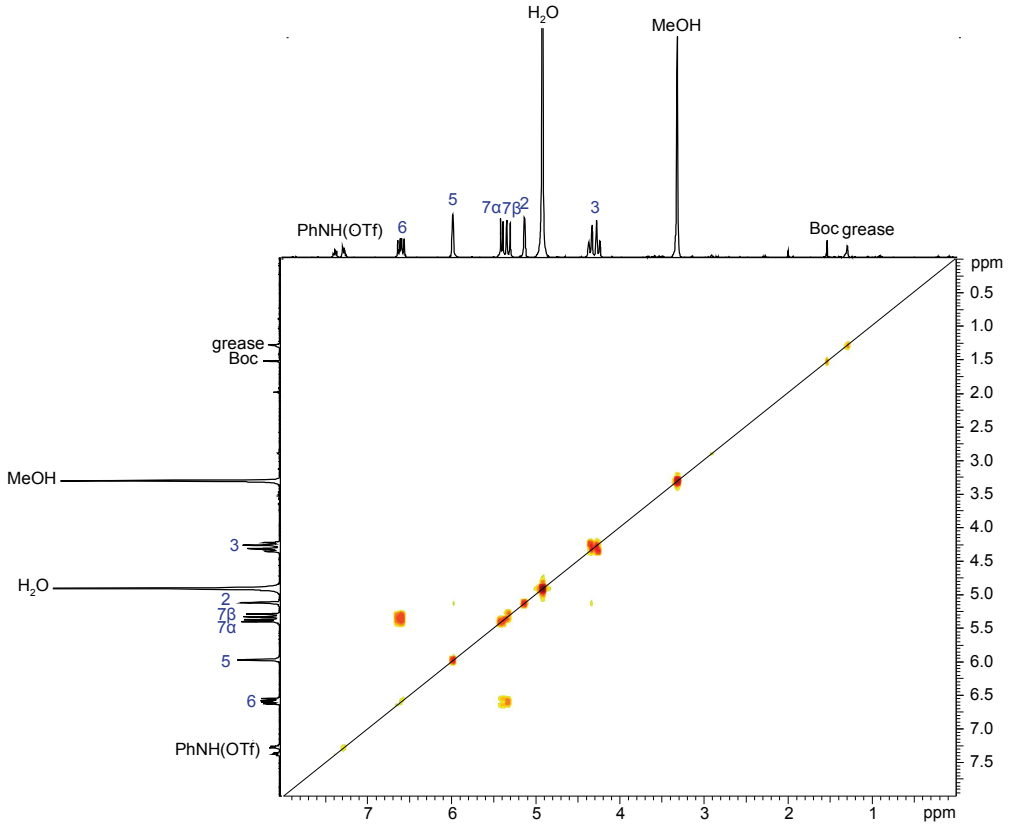
**A)**


H2O
MeOH
O
HO
8
2
3
4
6
Hα
HN
5
7
Hβ
1.2
5
7α7β2
3
PhNH(OTf)
6
Boc
grease
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

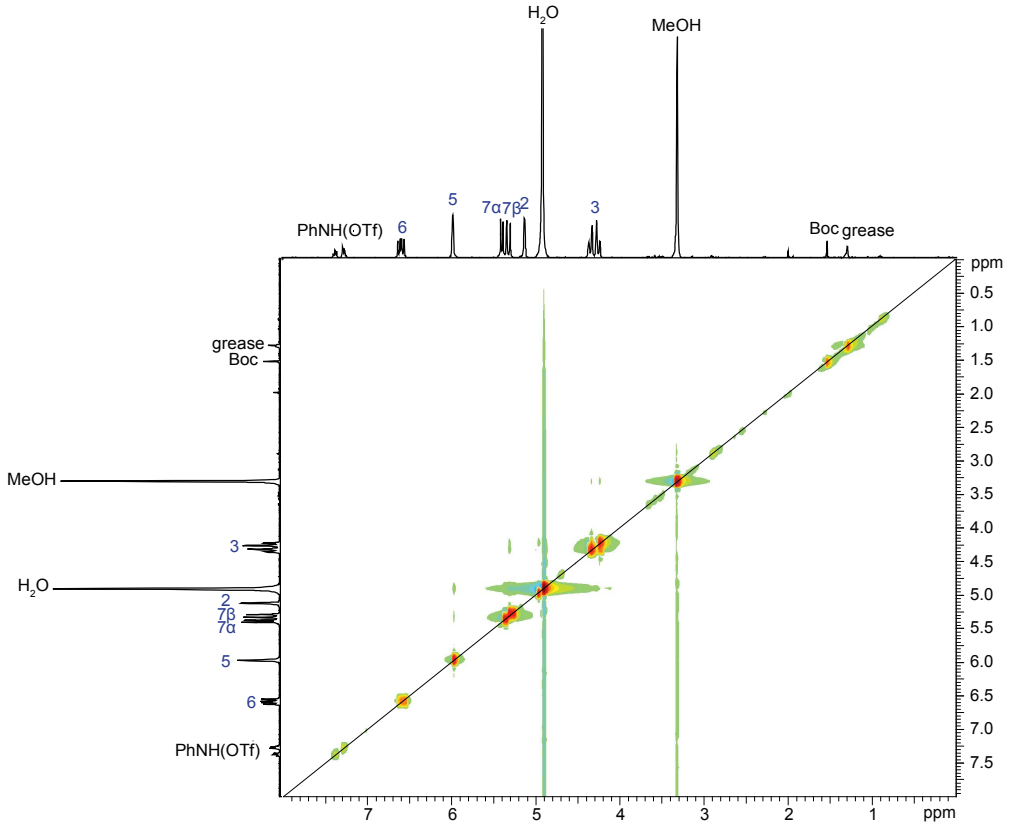
**B)**


MeOH
3
4
6
5
7
2
180 160 140 120 100 80 60 40 20 ppm

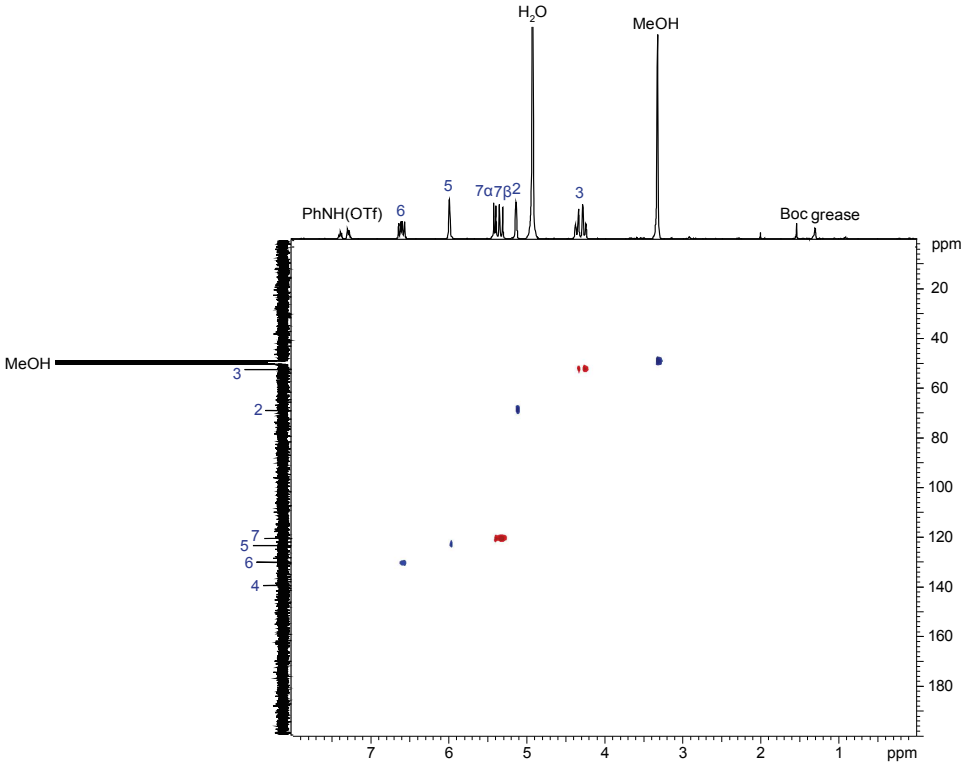
**C)**


H2O
MeOH
PhNH(OTf)
6
5
7α7β
2
3
Boc grease
grease
Boc
MeOH
3
H2O
2
7β
7α
5
6
PhNH(OTf)
ppm

**D)**


H2O
MeOH
PhNH(OTf)
6
5
7α7β
2
3
Boc grease
grease
Boc
MeOH
3
H2O
2
7β
7α
5
6
PhNH(OTf)
ppm

**E)**




**F)**

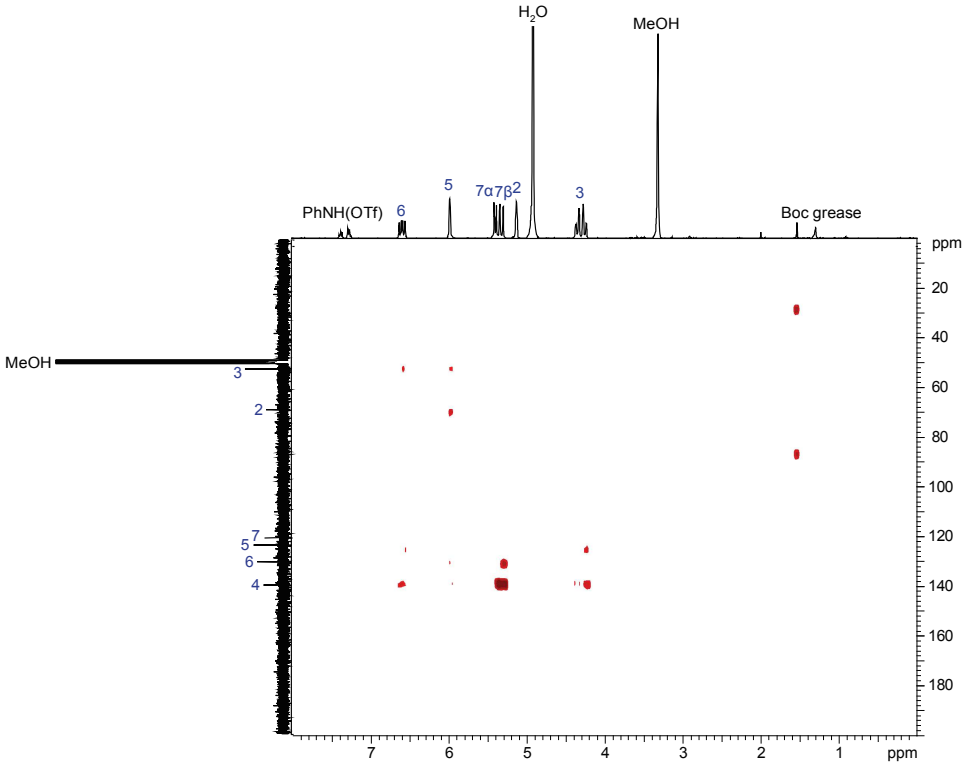



**Figure B-8.** Annotated **A)** $^{1}$H NMR spectrum, **B)** $^{13}$C NMR spectrum, **C)** COSY, **D)** NOESY, **E)** HSQC and **F)** HMBC of 4-vinyl-4,5-dehydro-L-proline **1.2** in MeOD.

**Table B-4.** $^{13}C$ and $^{1}H$ NMR of 4-vinyl-4,5-dehydro-L-proline **1.2**.

| Position | $^{13}C$ shift (ppm) | $^{1}H$ shift (ppm) | HMBC Connectivity |
|---|---|---|---|
| 2 | 68.4 | 5.12 (s, 1H) | 5 |
| 3 | 52.1 | 4.29 (dd, *J*=15, 2H) | 4, 5, 6 |
| 4 | 139.3 | | 3, 5, 6, 7 |
| 5 | 123.1 | 5.97 (s, 1H) | 2, 3, 4, 6 |
| 6 | 130.1 | 6.59 (dd, *J*=18, 11, 1H) | 3, 4, 5, 7 |
| 7α | 120.3 | 5.39 (d, *J*=11, 1H) | 4, 6 |
| 7β | 120.3 | 5.31 (d, *J*=18, 1H) | 4, 6 |
| 8 | | | |

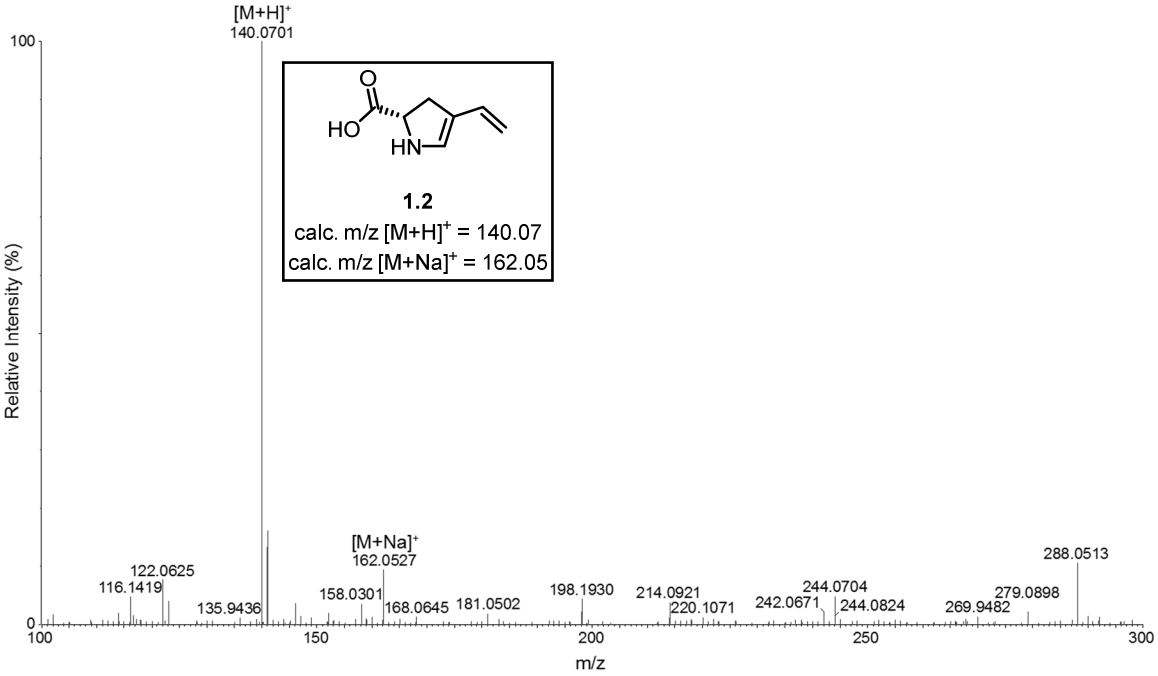



**Figure B-9.** $ESI^{+}$-MS of 4-vinyl-4,5-dehydro-L-proline **1.2**. The adduct $[M+Na]^{+}$ and parent ion $[M+H]^{+}$ are detected.

## Appendix C. Supporting Information for Chapter 4

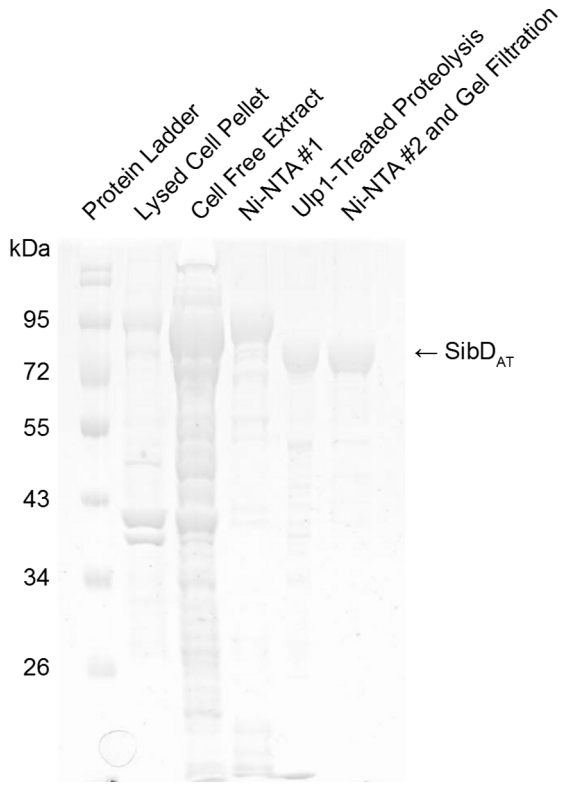



**Figure C-1.** SDS-PAGE gel image of NRPS didomain SibD$_{AT}$ at various purification stages.

SQRVHDLPWIPEDELRELRAAWSGPVDERPARQSLPQMFEARVREHPDRTAVRSSAGELTYRQLNSAANRL
ARALRRHGARRGTRVALRLERGPDAVVGLLAVLKSGASFVPLDPAYPPQRLALMLRDANPEVLLVHGGSPV
EVPAGTTVVDFGSVELPADDTDPAGQRAADDEMYVVYTSGSTGRPKGVIINDITITNLVHRQDDLSGLGET
ARTLQYMSISFDVSFMEIFCTLCAGGTVVVPDEQMRTDLRRLAGHLREQRVNRIFLPYVALQELATVLTRE
DIHLPDLAEVYTTGEALVVTAQIRDMFRRCTPAVLINAYGPSEAHLVSARRLPGDPSSWPERPAIGEVAGN
VRAYVLDQQQRPVPFGVRGELYVGGPVVGRGYLHLPDQTRERFRPDPYAGRPGGRMYRTGDLVMLTAPAGL
VHLGRADEQIKIRGYRVEPGEVEGALNDLPGITASAVVAAELEAGGRDLVAFVCTGSTLIDPRQVRDRLRA
SLPAYMVPSRIVTLDRLPVAPSGKTDRAALAALASRSPDRPADVSDDRPLGETEQRVAALWAGLLPGGGIG
PDTDFYSAGGHSLLAVRLRQAVEDEFGMELPLSALLATPTVAGMAARVDAVRAGHDEVDGPDLWADT
*

**Figure C-2.** Protein coverage (84.1%) map for SibD$_{AT}$ digested with trypsin. The observed digested peptides are highlight in yellow while the phosphopantetheinylated peptide is highlighted in black. S$^{*}$ denotes the phosphopantetheinylated serine.

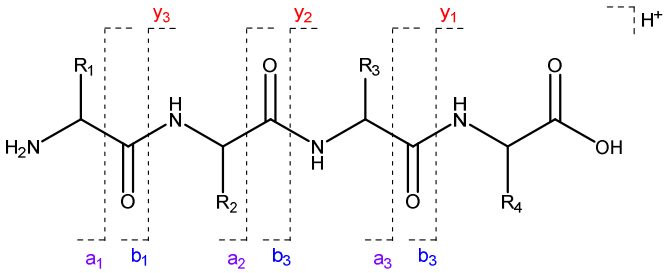



**Figure C-3.** Commonly observed peptide backbone cleavage sites in tandem MS experiments and notation of the peptide fragment ions that form as a result.

**A)**

| Phosphopantetheinylated peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHSLLAVR* |
|---|---|
| retention time | 32 min |
| intensity: | 5433 counts |
| assigned intensity: | 34.1 % |
| m/z: | 1160.2446 |
| charge state: | 3 |
| theoretical mass: | 3477.6296 Da |
| experimental mass: | 3477.7102 Da |
| mass error: | 23.2 ppm |
| analog-to-digital converter (ADC) response: | 180305 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b2 | VA | 171.1133 | 171.1151 | -0.0018 | 135 |
| b3 | VAA | 242.1505 | 242.1498 | 0.0007 | 177 |
| b4 | VAAL | 355.2345 | 355.2339 | 0.0006 | 174 |
| b5 | VAALW | 541.3138 | 541.3072 | 0.0067 | 111 |
| b8 | VAALWAGL | 782.4565 | 782.4548 | 0.0017 | 585 |
| b9 | VAALWAGLL | 895.5405 | 895.5353 | 0.0052 | 962 |
| b13 | VAALWAGLL | 1163.6577 | 1163.6411 | 0.0166 | 120 |
| y23* | PGGGIGPDTDFYSAGGHS*LLAVR | 2584.1050 | 2584.1799 | -0.0750 | 1794 |
| y32* | VAALWAGLLPGGGIGPDTDFYS*AGGHSLAVR (Parent ion) | 3478.6375 | 3478.7012 | -0.0637 | 1714 |

**Figure C-4.** Mass spectral analysis of phosphopantetheinylated peptide. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

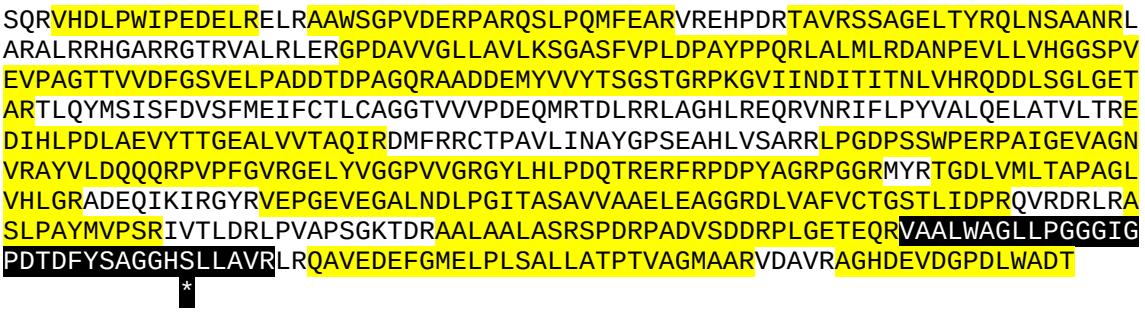



**Figure C-5.** Protein coverage (72.1%) map for $SibD_{AT}$ digested with trypsin. The observed digested peptides are highlight in yellow while the phosphopantetheinylated peptide free of and conjugated to tyrosine is highlighted in black. $S^*$ denotes the phosphopantetheinylated serine free of or conjugated to tyrosine.

**A)**

| Peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHSLLAVR<br>* |
|---|---|
| retention time | 30 min |
| intensity: | 26045 counts |
| assigned intensity: | 59.3 % |
| m/z: | 910.9539 |
| charge state: | 4 |
| theoretical mass: | 3639.8298 Da |
| experimental mass: | 3639.7837 Da |
| mass error: | -12.7 ppm |
| analog-to-digital converter (ADC) response: | 1591799 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b3 | VAA | 242.1505 | 242.1507 | -0.0002 | 190.0 |
| b4 | VAAL | 355.2345 | 355.2368 | -0.0023 | 120.0 |
| b5 | VAALW | 541.3138 | 541.3128 | 0.0010 | 190.0 |
| b6 | VAALWA | 612.3510 | 612.3490 | 0.0020 | 168.0 |
| b7 | VAALWAG | 669.3724 | 669.3807 | -0.0083 | 117.0 |
| b8 | VAALWAGL | 782.4565 | 782.4554 | 0.0011 | 527.0 |
| b9 | VAALWAGLL | 895.5405 | 895.5379 | 0.0026 | 892.0 |
| b11 | VAALWAGLLPG | 109.6147 | 1049.5889 | 0.0259 | 107.0 |
| b13 | VAALWAGLLPGGG | 1163.6577 | 1163.6541 | 0.0037 | 139.0 |
| b14 | VAALWAGLLPGGGI | 1276.7417 | 1276.7415 | 0.0002 | 113.0 |
| y17* | PDTDFYSAGGHS*LLAVR | 2308.0823 | 2308.0149 | 0.0674 | 1336.0 |
| y23* | PGGGIGPDTDFYSAGGHS*LLAVR | 2746.3049 | 2746.2559 | 0.0491 | 3955.0 |
| y24* | LPGGGIGPDTDFYSAGGHS*LLAVR | 2859.3889 | 2859.3464 | 0.0425 | 687.0 |
| y32* | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR (Parent ion) | 3640.8376 | 3640.7769 | 0.0608 | 5740.0 |

**Figure C-6.** Mass spectral analysis of phosphopantetheinylated peptide conjugated to tyrosine. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

**A)**

| | |
|---|---|
| Peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR |
| retention time | 32 min |
| intensity: | 60894 counts |
| assigned intensity: | 40.3 % |
| m/z: | 1160.2460 |
| charge state: | 3 |
| theoretical mass: | 3477.6296 Da |
| experimental mass: | 3477.7141 Da |
| mass error: | 24.3 ppm |
| analog-to-digital converter (ADC) response: | 3666720 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b2 | VA | 171.1133 | 171.1137 | -0.0004 | 194 |
| b3 | VAA | 242.1505 | 242.1505 | -0.0000 | 394 |
| b4 | VAAL | 355.2345 | 355.2348 | -0.0003 | 225 |
| b5 | VAALW | 541.3138 | 541.3117 | 0.0021 | 306 |
| b6 | VAALWA | 612.3510 | 612.3511 | -0.0001 | 469 |
| b7 | VAALWAG | 669.3724 | 669.3695 | 0.0029 | 408 |
| b8 | VAALWAGL | 782.4565 | 782.4546 | 0.0019 | 1243 |
| b9 | VAALWAGLL | 895.5405 | 895.5372 | 0.0034 | 2659 |
| b11 | VAALWAGLLPG | 1049.6147 | 1049.5836 | 0.0311 | 355 |
| b13 | VAALWAGLLPGGG | 1163.6577 | 1163.6504 | 0.0073 | 438 |
| b14 | VAALWAGLLPGGGI | 1276.7417 | 1276.7308 | 0.0109 | 525 |
| b15 | VAALWAGLLPGGGIG | 1333.7632 | 1276.7308 | 0.0016 | 236 |
| y1 | R | 175.1195 | 175.1198 | -0.0003 | 156 |
| y3 | AVR | 345.2250 | 345.2259 | 0.0009 | 154 |
| y4 | LAVR | 458.3091 | 458.3069 | 0.0022 | 133 |
| y5 | LLAVR | 571.3931 | 571.3926 | 0.0005 | 139 |
| y28* | WAGLLPGGGIGPDTDFYSAGGHS*LLAVR | 3124.4109 | 3124.4993 | -0.0884 | 151 |
| a4 | VAAL | 327.2396 | 327.2376 | 0.0020 | 137 |
| a5 | VAALW | 513.3189 | 513.3156 | 0.0033 | 112 |
| a8 | VAALWAGL | 754.4615 | 754.4587 | 0.0029 | 118 |
| a9 | VAALWAGLL | 867.5456 | 867.5405 | 0.0051 | 120 |
| y32* | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR (Parent ion) | 3478.6375 | 3478.7224 | -0.0850 | 15288 |

**Figure C-7.** Mass spectral analysis of phosphopantetheinylated peptide not conjugated to tyrosine. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

**A)**

| | |
|---|---|
| Peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHSLLAVR<br>* |
| retention time | 32 min |
| intensity: | 291178 counts |
| assigned intensity: | 40.7 % |
| m/z: | 1160.2478 |
| charge state: | 3 |
| theoretical mass: | 3477.6296 Da |
| experimental mass: | 3477.7195 Da |
| mass error: | 24.3 ppm |
| analog-to-digital converter (ADC) response: | 1613350 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b2 | VA | 171.1133 | 171.1127 | 0.0006 | 530 |
| b3 | VAA | 242.1505 | 242.1499 | 0.0005 | 1322 |
| b4 | VAAL | 355.2345 | 355.2334 | 0.0011 | 909 |
| b5 | VAALW | 541.3138 | 541.3110 | 0.0029 | 1638 |
| b6 | VAALWA | 612.3510 | 612.3513 | -0.0004 | 2150 |
| b7 | VAALWAG | 669.3724 | 669.3721 | 0.0003 | 1221 |
| b8 | VAALWAGL | 782.4565 | 782.4539 | 0.0026 | 5089 |
| b9 | VAALWAGLL | 895.5405 | 895.5386 | 0.0020 | 11397 |
| b11 | VAALWAGLLPG | 1049.6147 | 1049.5830 | 0.0315 | 1603 |
| b12 | VAALWAGLLPGG | 1106.6362 | 1106.6400 | -0.0040 | 128 |
| b13 | VAALWAGLLPGGG | 1163.6577 | 1163.6540 | 0.0035 | 1769 |
| b14 | VAALWAGLLPGGGI | 1276.7417 | 1276.7408 | 0.0009 | 2141 |
| b15 | VAALWAGLLPGGGIG | 1333.7632 | 1333.7632 | 0.0000 | 1215 |
| y1 | R | 175.1195 | 175.11989 | 0.0006 | 593 |
| y2 | AV | 274.1879 | 274.1863 | 0.0016 | 268 |
| y3 | AVR | 345.2250 | 345.2241 | 0.0009 | 693 |
| y4 | LAVR | 458.3091 | 458.3079 | 0.0012 | 310 |
| y5 | LLAVR | 571.3931 | 571.3933 | -0.0002 | 529 |
| y26* | GLLPGGGIGPDTDFYSAGGHS*LLAVR | 2867.2944 | 2867.3801 | -0.0857 | 4215 |
| y27* | AGLLPGGGIGPDTDFYSAGGHS*LLAVR | 2938.3315 | 2938.4187 | -0.0872 | 3107 |
| y28* | WAGLLPGGGIGPDTDFYSAGGHS*LLAVR | 3124.4109 | 3124.5012 | -0.0903 | 1777 |
| a4 | VAAL | 327.2396 | 327.2396 | -0.0003 | 517 |
| a5 | VAALW | 513.3189 | 513.3167 | 0.0022 | 403 |
| a6 | VAALWA | 584.3560 | 584.3550 | 0.0010 | 466 |
| a7 | VAALWAG | 641.3375 | 641.3783 | -0.0008 | 262 |
| a8 | VAALWAGL | 754.4615 | 754.4586 | 0.0029 | 694 |
| a9 | VAALWAGLL | 867.5456 | 867.5458 | -0.0002 | 722 |
| y32* | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR (Parent ion) | 3478.6375 | 3478.7271 | -0.0896 | 73771 |

**Figure C-8.** Mass spectral analysis of phosphopantetheinylated peptide not conjugated to threonine. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

**A)**

| | |
|---|---|
| Peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHSLLAVR<br>* |
| retention time | 32 min |
| intensity: | 121946 counts |
| assigned intensity: | 38.4 % |
| m/z: | 1160.2476 |
| charge state: | 3 |
| theoretical mass: | 3477.6296 Da |
| experimental mass: | 3477.7188 Da |
| mass error: | 25.6 ppm |
| analog-to-digital converter (ADC) response: | 6771313 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b2 | VA | 171.1133 | 171.1139 | -0.0006 | 200 |
| b3 | VAA | 242.1505 | 242.1494 | 0.0011 | 523 |
| b4 | VAAL | 355.2345 | 355.2339 | 0.0006 | 323 |
| b5 | VAALW | 541.3138 | 541.3110 | 0.0029 | 664 |
| b6 | VAALWA | 612.3510 | 612.3487 | 0.0023 | 871 |
| b7 | VAALWAG | 669.3724 | 669.3707 | 0.0016 | 577 |
| b8 | VAALWAGL | 782.4565 | 782.4544 | 0.0021 | 2269 |
| b9 | VAALWAGLL | 895.5405 | 895.5391 | 0.0014 | 4707 |
| b13 | VAALWAGLLPGGG | 1163.6577 | 1163.655 | 0.0023 | 737 |
| b14 | VAALWAGLLPGGGI | 1276.7417 | 1276.729 | 0.0125 | 690 |
| b15 | VAALWAGLLPGGGIG | 1333.7632 | 1333.762 | 0.0009 | 113 |
| y1 | R | 175.1195 | 175.1192 | 0.0003 | 286 |
| y2 | AV | 274.1879 | 274.1887 | -0.0008 | 104 |
| y3 | AVR | 345.2250 | 345.2245 | 0.0005 | 243 |
| y4 | LAVR | 458.3091 | 458.3086 | 0.0005 | 159 |
| y5 | LLAVR | 571.3931 | 571.3932 | -0.0001 | 182 |
| y22* | LPGGGIGPDTDFYSAGGHS*LLA | 2487.0520 | 2487.124 | -0.0720 | 385 |
| y26* | WAGLLPGGGIGPDTDFYSAGGHS*LLA | 2867.2944 | 2867.373 | -0.0789 | 1618 |
| a4 | VAAL | 327.2396 | 327.2394 | 0.0002 | 248 |
| a5 | VAALW | 513.3189 | 513.3183 | 0.0006 | 185 |
| a6 | VAALWA | 584.3560 | 584.3541 | 0.0019 | 140 |
| a7 | VAALWAG | 641.3375 | 641.3749 | 0.0026 | 120 |
| a8 | VAALWAGL | 754.4615 | 754.4566 | 0.0049 | 182 |
| a9 | VAALWAGLL | 867.5456 | 867.5392 | 0.0064 | 223 |
| y32* | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR (Parent ion) | 3478.6375 | 3478.7261 | -0.0866 | 30888 |

**Figure C-9.** Mass spectral analysis of phosphopantetheinylated peptide not conjugated to **1.2**. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

**A)**

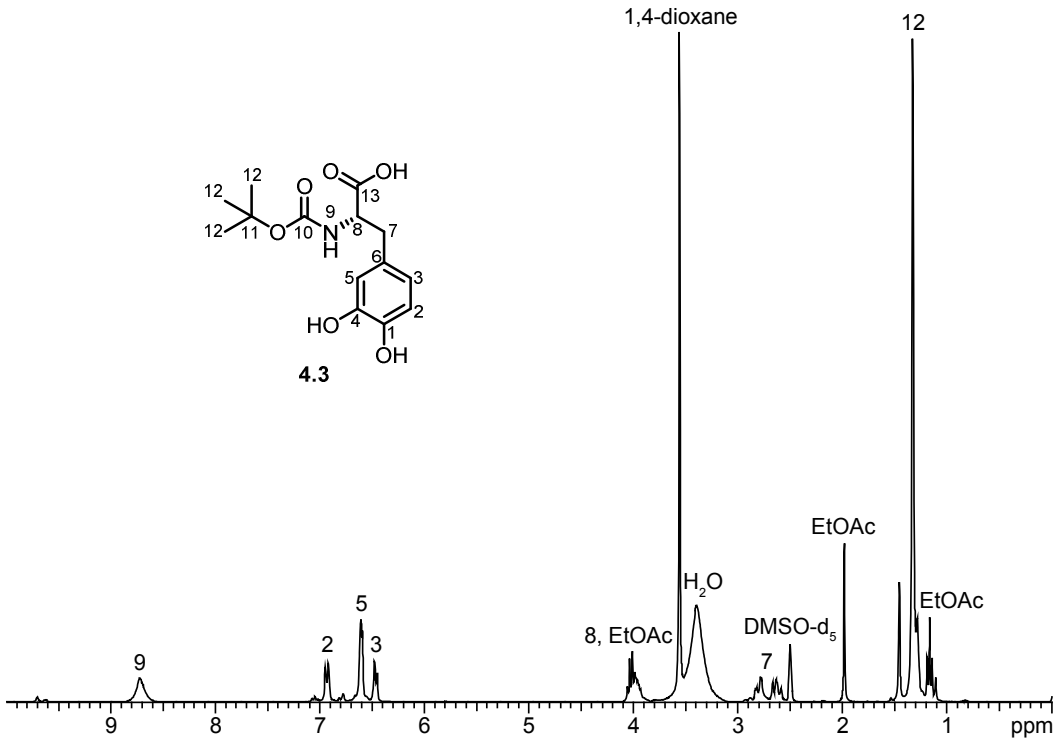
| | |
|---|---|
| Peptide of interest: | VAALWAGLLPGGGIGPDTDFYSAGGHSLLAVR<br>* |
| retention time | 32 min |
| intensity: | 163214 counts |
| assigned intensity: | 39.4 % |
| m/z: | 1160.2471 |
| charge state: | 3 |
| theoretical mass: | 3477.6296 Da |
| experimental mass: | 3477.7173 Da |
| mass error: | 25.2 ppm |
| analog-to-digital converter (ADC) response: | 9084524 |

**B)**

| Fragment detected | Fragment sequence | Theoretical mass (Da) | Experimental mass (Da) | Mass error (Da) | Intensity (counts) |
|---|---|---|---|---|---|
| b2 | VA | 171.1133 | 171.1122 | 0.0011 | 276 |
| b3 | VAA | 242.1505 | 242.1507 | -0.0002 | 702 |
| b4 | VAAL | 355.2345 | 355.2344 | 0.0001 | 451 |
| b5 | VAALW | 541.3138 | 541.3106 | 0.0032 | 920 |
| b6 | VAALWA | 612.3510 | 612.3497 | 0.0013 | 1092 |
| b7 | VAALWAG | 669.3724 | 669.3719 | 0.0005 | 696 |
| b8 | VAALWAGL | 782.4565 | 782.4551 | 0.0014 | 2929 |
| b9 | VAALWAGLL | 895.5405 | 895.5376 | 0.0029 | 6347 |
| b11 | VAALWAGLLPG | 1049.6147 | 1049.588 | 0.0265 | 890 |
| b13 | VAALWAGLLPGGG | 1163.6577 | 1163.658 | -0.0007 | 960 |
| b14 | VAALWAGLLPGGGI | 1276.7417 | 1276.738 | 0.0042 | 1087 |
| b15 | VAALWAGLLPGGGIG | 1333.7632 | 1333.765 | -0.0018 | 590 |
| y1 | R | 175.1195 | 175.1188 | 0.0007 | 343 |
| y2 | AV | 274.1879 | 274.1886 | -0.0007 | 141 |
| y3 | AVR | 345.2250 | 345.2240 | 0.0010 | 320 |
| y4 | LAVR | 458.3091 | 458.3046 | 0.0045 | 195 |
| y5 | LLAVR | 571.3931 | 571.3920 | 0.0011 | 236 |
| y26* | WAGLLPGGGIGPDTDFYSAGGHS*LLA | 2867.2944 | 2867.379 | -0.0845 | 2443 |
| y27* | LWAGLLPGGGIGPDTDFYSAGGHS*LLA | 2938.3315 | 2938.418 | -0.0864 | 1792 |
| a4 | VAAL | 327.2396 | 327.2379 | 0.0017 | 288 |
| a5 | VAALW | 513.3189 | 513.3146 | 0.0043 | 228 |
| a6 | VAALWA | 584.3560 | 584.3513 | 0.0047 | 201 |
| a7 | VAALWAG | 641.3375 | 641.3756 | 0.0019 | 155 |
| a8 | VAALWAGL | 754.4615 | 754.4585 | 0.0031 | 463 |
| a9 | VAALWAGLL | 867.5456 | 867.5478 | -0.0022 | 272 |
| y32* | VAALWAGLLPGGGIGPDTDFYSAGGHS*LLAVR (Parent ion) | 3478.6375 | 3478.7250 | -0.0874 | 42775 |

**Figure C-10.** Mass spectral analysis of phosphopantetheinylated peptide not conjugated to **1.5**. **A)** Mass spectral details on the parent peptide ion. **B)** Mass spectral details on the peptide fragment ions detected.

**A)**




**B)**

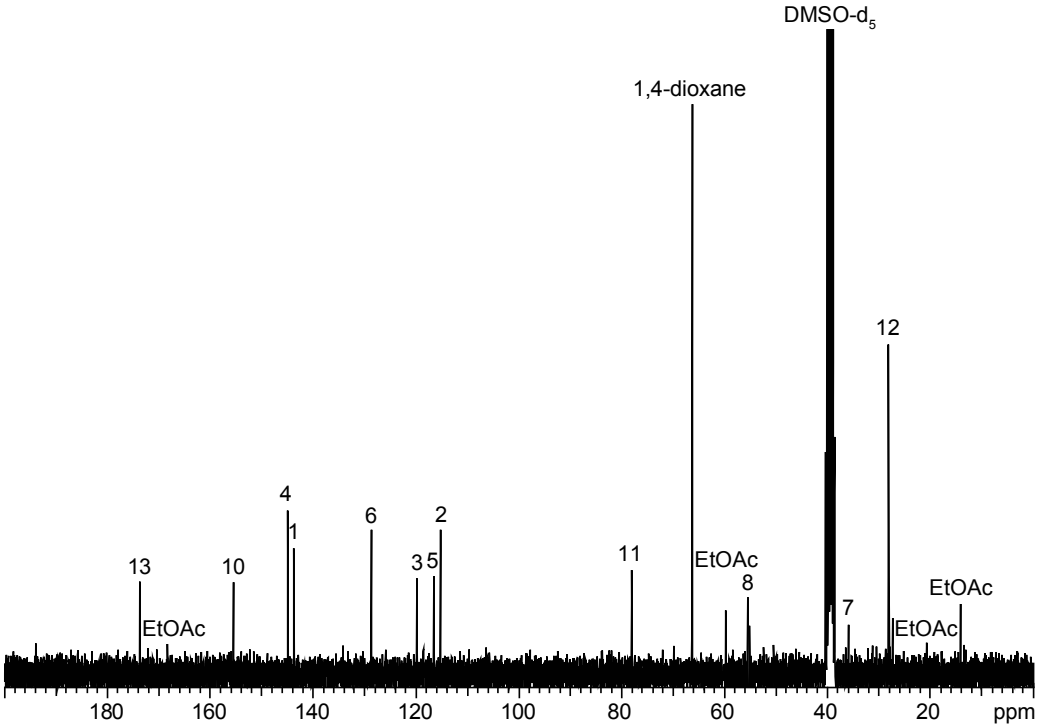



**Figure C-11.** Annotated **A)** $^{1}$H and **B)** $^{13}$C NMR spectra of *N*-Boc-L-DOPA **4.3** in DMSO-$d_6$.

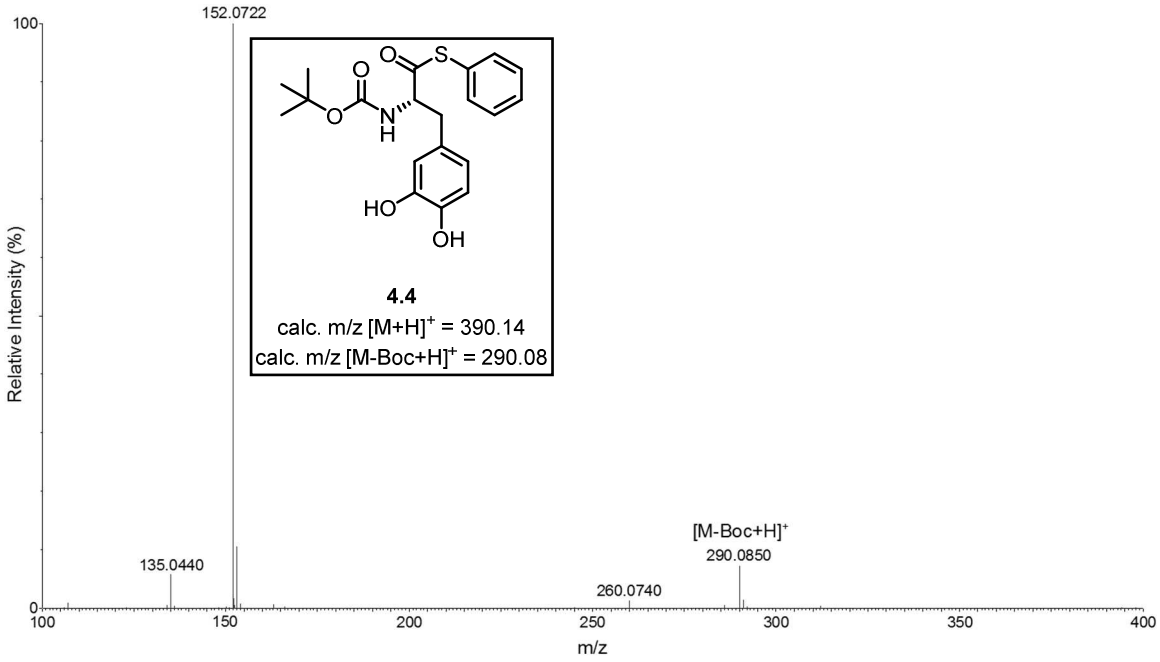



**Figure C-12.** ESI$^+$-MS of *N*-Boc-L-DOPA-thiophenol **4.4**.

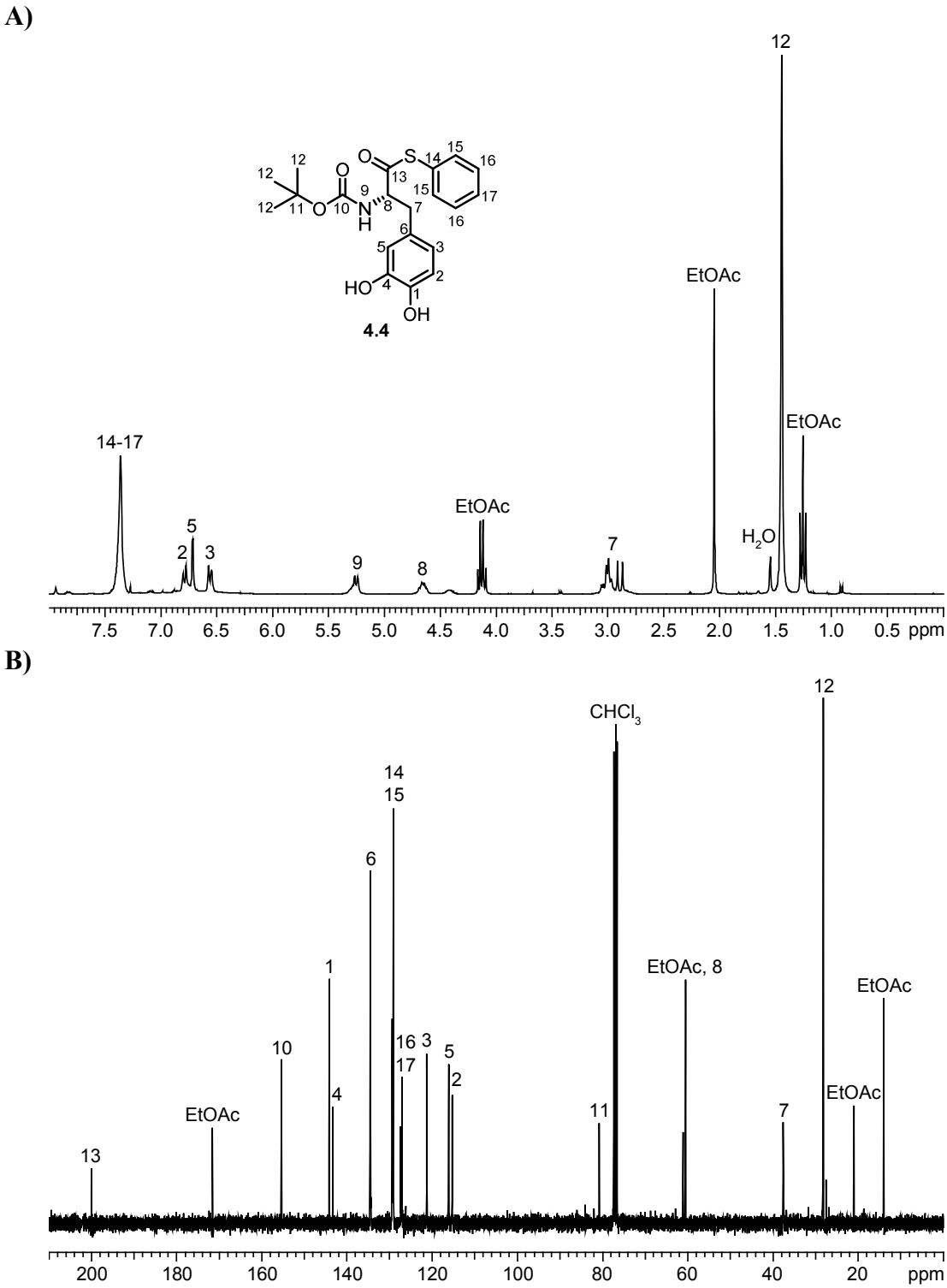



**Figure C-13.** Annotated **A)** $^{1}$H and **B)** $^{13}$C NMR spectra of *N*-Boc-L-DOPA-thiophenol **4.4** in $CDCl_3$.

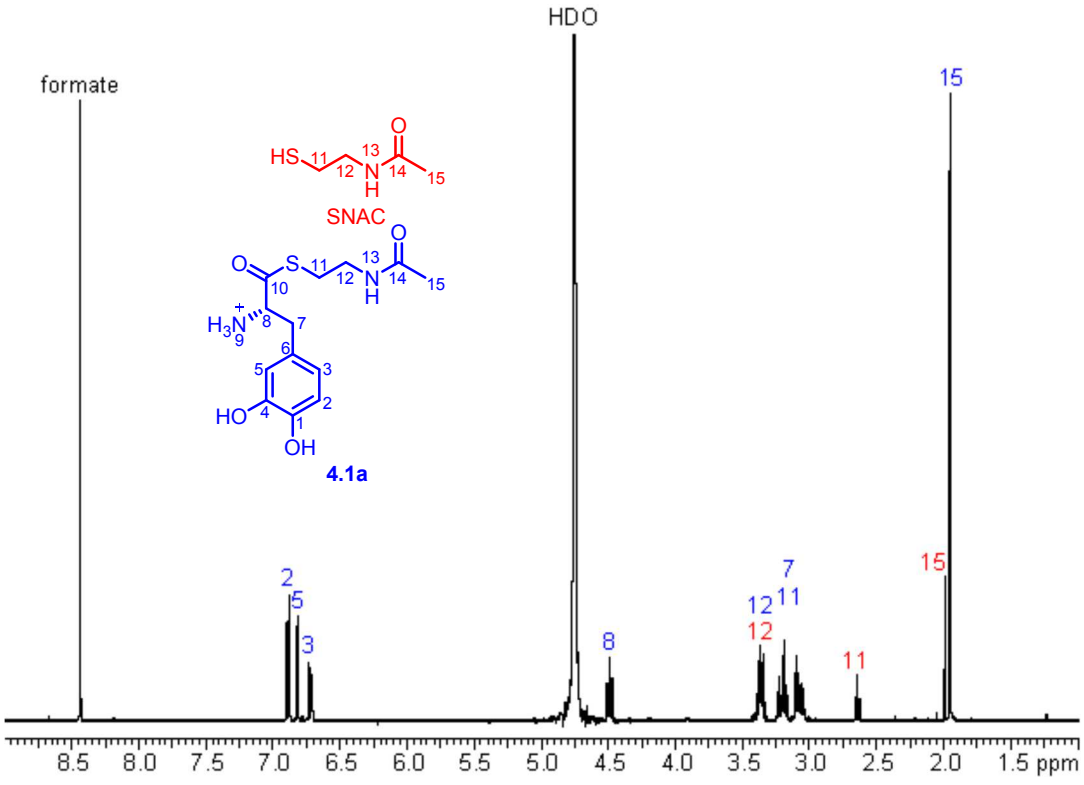



**Figure C-14.** Annotated $^{1}$H NMR spectrum of L-DOPA-SNAC **4.1a** and SNAC in $D_2O$.

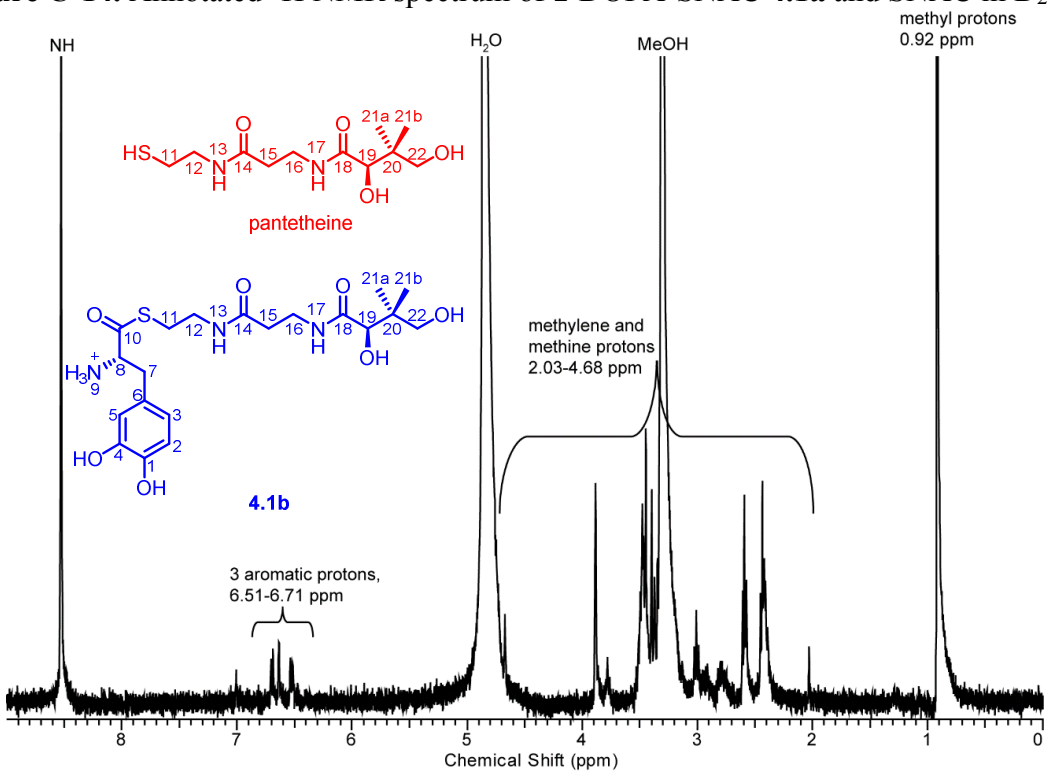



**Figure C-15.** Partially annotated $^{1}$H NMR spectrum of L-DOPA-pantetheine **4.1b** and pantetheine in MeOD.

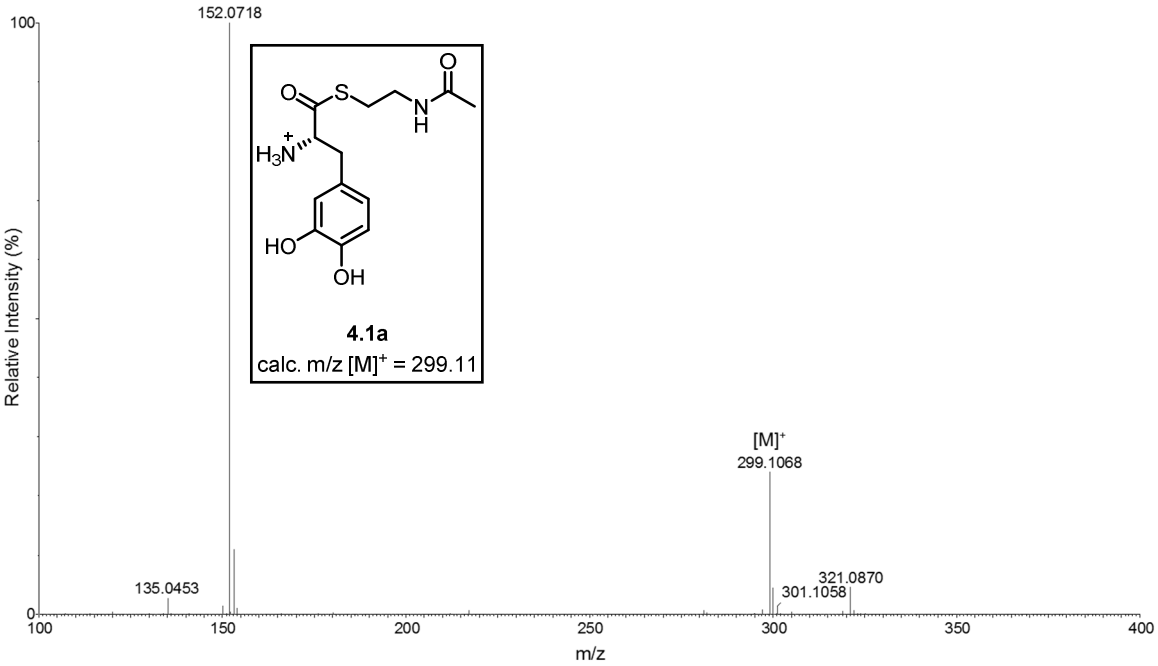



**Figure C-16.** ESI$^+$-MS of L-DOPA-SNAC **4.1a**. The parent ion $[M]^+$ is detected.

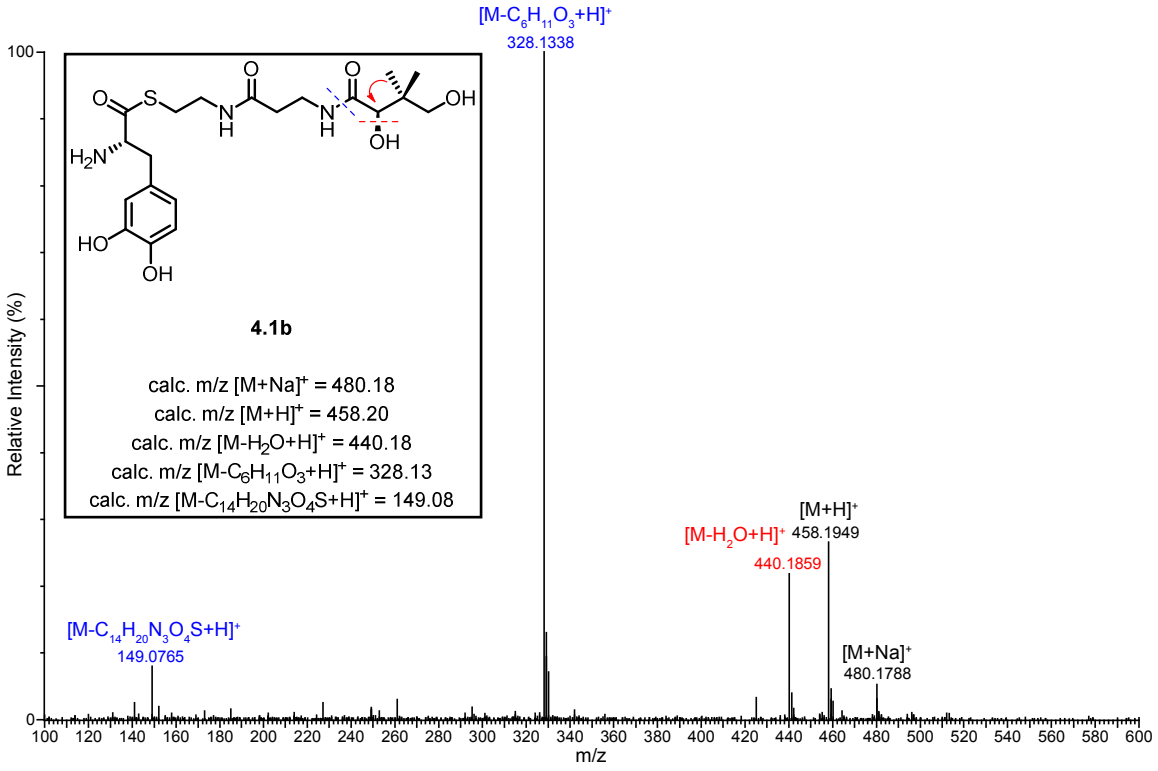



**Figure C-17.** ESI$^+$-MS of L-DOPA-pantetheine **4.1b**. The parent ion $[M+H]^+$, adduct $[M+Na]^+$, and fragment ions $[M\text{-}H_2O+H]^+$ , $M\text{-}C_6H_{11}O_3+H]^+$, and $[M\text{-}C_{14}H_{20}N_3O_2S+H]^+$ are detected. Formation of the $[M\text{-}H_2O+H]^+$ fragment ion involves a methyl shift indicated by an arrow. Formation of the other two fragment ions involves cleavage at sites indicated by dashed lines.

AGCGATGTCCTGCTGACCGTCGTGCGTGACCGTTGGCGTACCTCTTCTGGCCGCCCGGAAATTCTGGCTGC
TGAACTGGCCGATATTCTGGGTCGTGCGGTTGATCCGGAAGACATTGCCGCGGCAGGCAGTCCGGAATCCC
TGGTGGCGCGTCTGACCACCGGTTCAGTGCCGGCGGTTGGTACCCCGGGTACCGAACTGCGCAGCGCACCG
GCATGGCCGGCAACCCCGGGTCAGACGGGTATTTGGTATCTGAGTGAATTTGGCGAAGCTCGTACCGCGTA
TAATTCCCCGATCCTGCTGCGTGTGCCGGGTGCCCTGGACCCGGAAGCCCTGCGCGCTGCGCTGGAACATG
TGGTTCGTCGCCACGAAAGTCTGCGTACCACGTTCGAAATGCGCGATGGCACCCTGACGCAACTGGTTGCC
GCAGAACCGCGTTTTGGTTTTCGTACCGCTGCCGTGGGCTCCGCGGGTGAAGCGGGCCGTCTGGCGGCAGA
ACTGGCTGCCGAACATCTGGATCTGACCGATGGTCCGCTGCTGCGTGCCGTCTGTGTGGATGTTGGCGGTG
ACAATACGATTCTGCTGGTCAGCGTGCATCACGCAGTCTTTGATGGCTTCAGCTGGTCTGTGCTGCTGTCT
GAATGGCTGACCGCCTATCGTCGCGCGGCAGCTGGTGAACGTCTGCCGCATGATGCACCGCCGGTCCAGTT
TCGTCAAGCTGTTGCCGCACTGCGTCCGGCTGGTGAAGCCGGCCTGGCCTACTGGACCGGTCATCTGGCGG
GCGTTCCGGTCCTGGATCTGCCGCTGGACCACCCGGCGGGTCCGCGTGGCCGCGTTGCAGCTGCGAGCGTC
CATACCCGTCTGACCGCCTCTGAAGTCGTGCGTCTGCGTCGCCAGGCAGCATCTCATGGTGCAACCCCGGC
TGTTGCCCTGCTGGCTGCGTATGCACTGCTGATGCATCGCCACACCGGTCAGCCGGATCTGACCGTTGGTC
TGCCGGTGTCACTGCGTGATCGCTCGGCCGACCAACATGTGATCGGTCACCTGGTTAACACCGTTGTCCTG
CGTCATCGCGTTACGCCGGGTAGCACCGGCACGGATCTGCTGACCGCAACCCGCGATGAAGTGCGTGCAGC
ACTGGCTCACAAAAATGTTCCGTTTGAACAGATTGTCGAACGTCTGGCACCGGATCGTGCTGATGGTCGTC
CGCCGCTGTTCCAAACCATGGTGACGATCATGCCGGCGGAACGTCGCGATCTGCGCCATCTGGGTCTGGGT
GCAGATGCATGGCAGCACCTGGGTAGTACCCCGAAGTTTGATCTGGCCCTGGTGGTTGAAGAAGCTGCGGA
ACATCTGGGTCTGACCTTCGAATACGATCCGGCAGTGCTGGGCCACGACACGGTTACCCGTATGGCAGCCC
GTTTTCGCACCGTCCTGGCAGCCCTGATGGATGACCCGGATTGA

**Figure C-18.** Nucleotide sequence used to express the condensation domain (SibD$_C$) in SibD. This *sibD$_C$* synthetic gene fragment was optimized (GenScript) for expression in *E. coli*.

## Appendix D. Supporting Information for Chapter 5

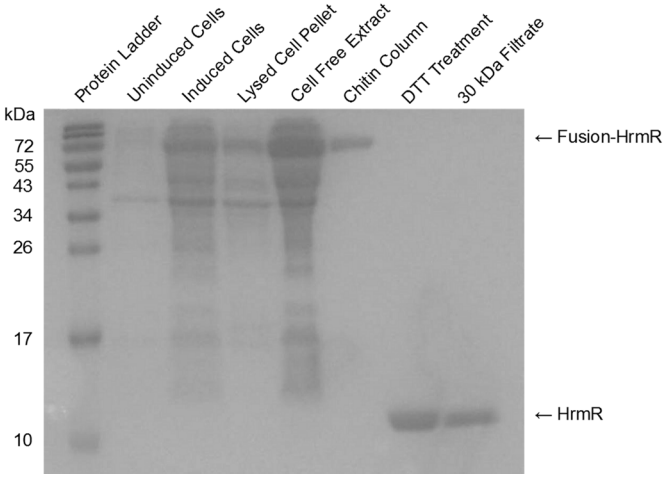
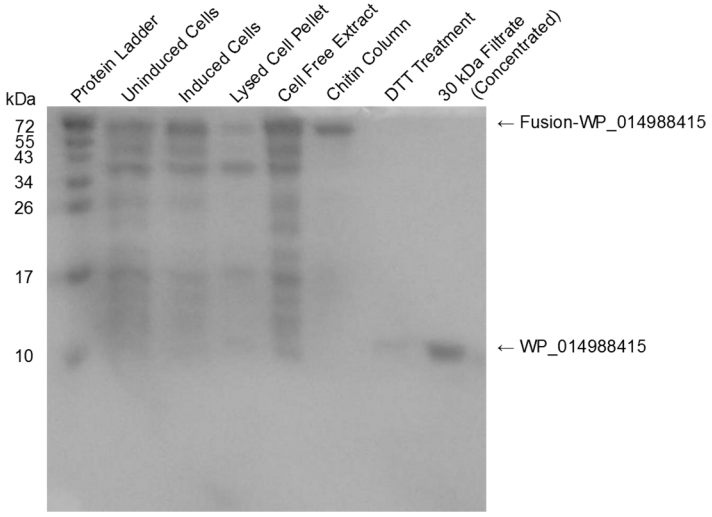



**Figure D-1.** SDS-PAGE gel image of MLP HrmR at various expression and purification stages.

**A)**




**B)**

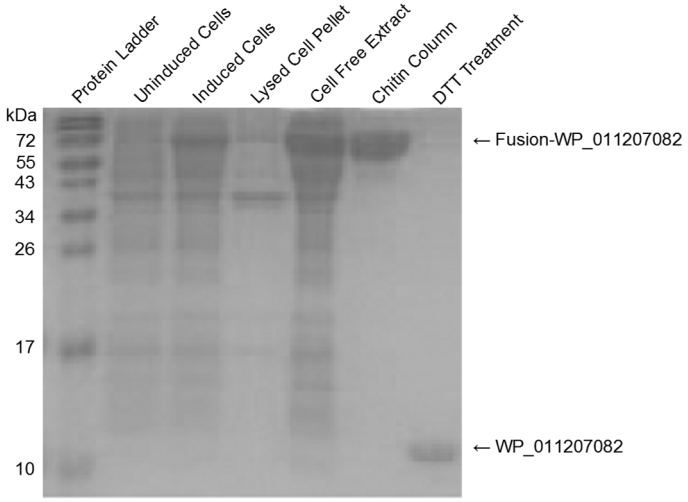



**Figure D-2.** SDS-PAGE gel images of MLPs **A)** WP_014988415 and **B)** WP_011207082 at various expression and purification stages.

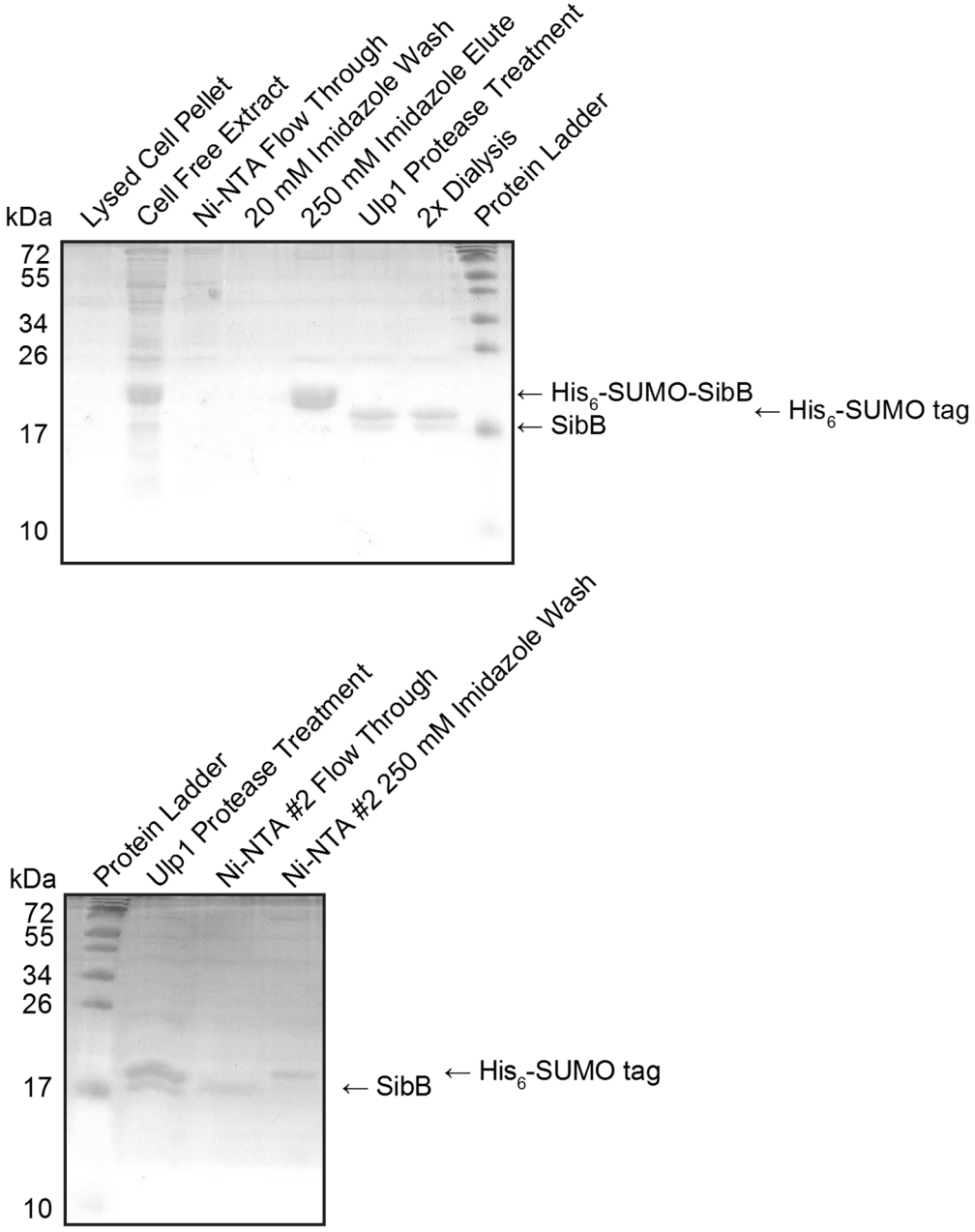



**Figure D-3.** SDS-PAGE gel image of SibB at various expression and purification stages.

**A)**

```
TAATAACCCGTTTGACAATGACGACGCTCAGTTCTATGTGCTGATTAATGCCGATGGTGAACACTCGCTGT
GGCCGATTTTTGCCGCAGTGCCGGGCGGTTGGACCATTGCTTATGGCGCGGCCGCACGTAAAGCGTGCCTG
CATTACGTTGAAACCCACTGGACGGATATGCGCCCGAAAAGTCTGATCGAAGCAGGTAGCAACGAAACGTC
TTAA
```

**B)**

```
TAAACTGTCCACGAACCCGTTTGATGACGAAGATGGTCGCTTCTTTGTGCTGGTGAATGACGAAGAACAAC
ACTCGCTGTGGCCGGCATTTGCAGAAGTTCCGGCTGGCTGGCGTGTGGTTTTCGGTGAAGATAGTCGCGCG
GCCTGCGTCGAATATGTGGAGAAAAACTGGACCGATATGCGTCCGAAAAGCCTGCGCGACGCAATGGCAGC
TGATGACGCAGCACGTCAGGGTGCCCAATCTTAA
```

**Figure D-4.** Nucleotide sequences of the genes coding for MLPs **A)** WP_014988415 and **B)** WP_011207082. These synthetic genes were optimized (GenScript) for expression in *E. coli*.

# References


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

2. Gerratana, B. (2012) Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines, *Med. Res. Rev. 32*, 254-293.

3. Hurley, L. H., and Thurston, D. E. (1984) Pyrrolo(1,4)benzodiazepine antitumor antibiotics: chemistry, interaction with DNA, and biological implications, *Pharm. Res. 1*, 52-59.

4. Thurston, D. E., and Bose, D. S. (1994) Synthesis of DNA-interactive pyrrolo[2,1-c][1,4]benzodiazepines, *Chem. Rev. 94*, 433-465.

5. Graves, D. E., Pattaroni, C., Krishnan, B. S., Ostrander, J. M., Hurley, L. H., and Krugh, T. R. (1984) The reaction of anthramycin with DNA. Proton and carbon nuclear magnetic resonance studies on the structure of the anthramycin-DNA adduct, *J. Biol. Chem. 259*, 8202-8209.

6. Krugh, T. R., Graves, D. E., and Stone, M. P. (1989) Two-dimensional NMR studies on the anthramycin-d(ATGCAT)$_2$ adduct, *Biochemistry 28*, 9988-9994.

7. Boyd, F. L., Stewart, D., Remers, W. A., Barkley, M. D., and Hurley, L. H. (1990) Characterization of a unique tomaymycin-d(CICGAATTCICG)$_2$ adduct containing two drug molecules per duplex by NMR, fluorescence, and molecular modeling studies, *Biochemistry 29*, 2387-2403.

8. Kopka, M. L., Goodsell, D. S., Baikalov, I., Grzeskowiak, K., Cascio, D., and Dickerson, R. E. (1994) Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C-C-A-A-C-G-T-T-G-G and a molecular explanation of specificity, *Biochemistry 33*, 13593-13610.

9. Barkley, M. D., Cheatham, S., Thurston, D. E., and Hurley, L. H. (1986) Pyrrolo[1,4]benzodiazepine antitumor antibiotics: evidence for two forms of tomaymycin bound to DNA, *Biochemistry 25*, 3021-3031.

10. Rahman, K. M., Vassoler, H., James, C. H., and Thurston, D. E. (2010) DNA sequence preference and adduct orientation of pyrrolo[2,1-c][1,4]benzodiazepine antitumor agents, *ACS Med. Chem. Lett. 1*, 427-432.

11. Puvvada, M. S., Hartley, J. A., Jenkins, T. C., and Thurston, D. E. (1993) A quantitative assay to measure the relative DNA-binding affinity of pyrrolo[2,1-c][1,4]benzodiazepine (PBD) antitumour antibiotics based on the inhibition of restriction

endonuclease BamHI, *Nucleic Acids Res. 21*, 3671-3675.

12. Puvvada, M. S., Forrow, S. A., Hartley, J. A., Stephenson, P., Gibson, I., Jenkins, T. C., and Thurston, D. E. (1997) Inhibition of bacteriophage T7 RNA polymerase *in vitro* transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines, *Biochemistry 36*, 2478-2484.

13. Petrusek, R. L., Uhlenhopp, E. L., Duteau, N., and Hurley, L. H. (1982) Reaction of anthramycin with DNA - biological consequences of DNA damage in normal and xeroderma pigmentosum cell lines, *J. Biol. Chem. 257*, 6207-6216.

14. Chen, Z., Gregson, S. J., Howard, P. W., and Thurston, D. E. (2004) A novel approach to the synthesis of cytotoxic C2–C3 unsaturated pyrrolo[2,1-c][1,4]benzodiazepines (PBDs) with conjugated acrylyl C2-substituents, *Bioorg. Med. Chem. Lett. 14*, 1547-1549.

15. Tsugaya, M., Washida, H., Hirao, N., Hachisuka, Y., Sakagami, H., and Iwase, Y. (1986) The treatment of bladder cancer by neothramycin, *Hinyokika Kiyo 32*, 1443-1448.

16. Jackson, P. J., James, C. H., Jenkins, T. C., Rahman, K. M., and Thurston, D. E. (2014) Computational studies support the role of the C7-sibirosamine sugar of the pyrrolobenzodiazepine (PBD) sibiromycin in transcription factor inhibition, *ACS Chem. Biol. 9*, 2432-2440.

17. Thurston, D. E., Bose, D. S., Howard, P. W., Jenkins, T. C., Leoni, A., Baraldi, P. G., Guiotto, A., Cacciari, B., Kelland, L. R., Foloppe, M.-P., and Rault, S. (1999) Effect of A-ring modifications on the DNA-binding behavior and cytotoxicity of pyrrolo[2,1-c][1,4]benzodiazepines, *J. Med. Chem. 42*, 1951-1964.

18. Cargill, C., Bachmann, E., and Zbinden, G. (1974) Effects of daunomycin and anthramycin on electrocardiogram and mitochondrial metabolism of the rat heart, *J. Natl. Cancer Inst. 53*, 481-486.

19. Lubawy, W. C., Dallam, R. A., and Hurley, L. H. (1980) Protection against anthramycin-induced toxicity in mice by coenzyme Q10, *J. Natl. Cancer Inst. 64*, 105-109.

20. Kumar, R., and Lown, J. W. (2003) Recent developments in novel pyrrolo[2,1-c][1,4]benzodiazepine conjugates: synthesis and biological evaluation, *Mini-Rev. Med. Chem. 3*, 323-339.

21. Kamal, A., Rao, M. V., Laxman, N., Ramesh, G., and Reddy, G. S. (2002) Recent developments in the design, synthesis and structure-activity relationship studies of pyrrolo[2,1-c][1,4]benzodiazepines as DNA-interactive antitumour antibiotics, *Curr. Med. Chem.: Anti-Cancer Agents 2*, 215-254.

22. Alley, M. C., Hollingshead, M. G., Pacula-Cox, C. M., Waud, W. R., Hartley, J. A., Howard, P. W., Gregson, S. J., Thurston, D. E., and Sausville, E. A. (2004) SJG-136 (NSC 694501), a novel rationally designed DNA minor groove interstrand cross-linking agent with potent and broad spectrum antitumor activity: part 2: efficacy evaluations, *Cancer Res. 64*, 6700-6706.

23. Hu, Y., Phelan, V., Ntai, I., Farnet, C. M., Zazopoulos, E., and Bachmann, B. O. (2007) Benzodiazepine biosynthesis in *Streptomyces refuineus*, *Chem. Biol. 14*, 691-701.

24. Giessen, T. W., Kraas, F. I., and Marahiel, M. A. (2011) A four-enzyme pathway for 3,5-dihydroxy-4-methylanthranilic acid formation and incorporation into the antitumor antibiotic sibiromycin, *Biochemistry 50*, 5680-5692.

25. Yonemoto, I. T., Li, W., Khullar, A., Reixach, N., and Gerratana, B. (2012) Mutasynthesis of a potent anticancer sibiromycin analogue, *ACS Chem. Biol. 7*, 973-977.

26. Mostad, A., Romming, C., and Storm, B. (1978) Structure of the DNA complexing agent anthramycin, *Acta Chem. Scand. 32*, 639-645.

27. Arora, S. K. (1981) Structure of tomaymycin, a DNA binding antitumor antibiotic, *J. Antibiot. 34*, 462-464.

28. Hurley, L. H., Reck, T., Thurston, D. E., Langley, D. R., Holden, K. G., Hertzberg, R. P., Hoover, J. R. E., Gallagher, G., Faucette, L. F., Mong, S. M., and Johnson, R. K. (1988) Pyrrolo[1,4]benzodiazepine antitumor antibiotics: relationship of DNA alkylation and sequence specificity to the biological activity of natural and synthetic compounds, *Chem. Res. Toxicol. 1*, 258-268.

29. Vargiu, A. V., Ruggerone, P., Magistrato, A., and Carloni, P. (2006) Anthramycin-DNA binding explored by molecular simulations, *J. Phys. Chem. B 110*, 24687-24695.

30. Hurley, L. H. (1980) Elucidation and formulation of novel biosynthetic pathways leading to the pyrrolo[1,4]benzodiazepine antibiotics anthramycin, tomaymycin, and sibiromycin, *Acc. Chem. Res. 13*, 263-269.

31. Li, W., Khullar, A., Chou, S., Sacramo, A., and Gerratana, B. (2009) Biosynthesis of sibiromycin, a potent antitumor antibiotic, *Appl. Environ. Microbiol. 75*, 2869-2878.

32. Li, W., Chou, S., Khullar, A., and Gerratana, B. (2009) Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an SJG-136 monomeric analog, *Appl. Environ. Microbiol. 75*, 2958-2963.

33. Najmanova, L., Ulanova, D., Jelinkova, M., Kamenik, Z., Kettnerova, E., Koberska, M., Gazak, R., Radojevic, B., and Janata, J. (2014) Sequence analysis of porothramycin biosynthetic gene cluster, *Folia Microbiol. (Prague, Czech Repub.) 59*, 543-552.

34. Schneditz, G., Rentner, J., Roier, S., Pletz, J., Herzog, K. A. T., Bücker, R., Troeger, H., Schild, S., Weber, H., Breinbauer, R., Gorkiewicz, G., Högenauer, C., and Zechner, E. L. (2014) Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis, *Proc. Natl. Acad. Sci. U.S.A. 111*, 13181-13186.

35. Peschke, U., Schmidt, H., Zhang, H. Z., and Piepersberg, W. (1995) Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11, *Mol. Microbiol. 16*, 1137-1156.

36. Koberska, M., Kopecky, J., Olsovska, J., Jelinkova, M., Ulanova, D., Man, P., Flieger, M., and Janata, J. (2008) Sequence analysis and heterologous expression of the lincomycin biosynthetic cluster of the type strain *Streptomyces lincolnensis* ATCC 25466, *Folia Microbiol. (Prague, Czech Repub.) 53*, 395-401.

37. Hofer, I., Crusemann, M., Radzom, M., Geers, B., Flachshaar, D., Cai, X., Zeeck, A., and Piel, J. (2011) Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite, *Chem. Biol. 18*, 381-391.

38. Connor, K. L., Colabroy, K. L., and Gerratana, B. (2011) A heme peroxidase with a functional role as an L-tyrosine hydroxylase in the biosynthesis of anthramycin, *Biochemistry 50*, 8926-8936.

39. Novotna, J., Olsovska, J., Novak, P., Mojzes, P., Chaloupkova, R., Kamenik, Z., Spizek, J., Kutejova, E., Mareckova, M., Tichy, P., Damborsky, J., and Janata, J. (2013) Lincomycin biosynthesis involves a tyrosine hydroxylating heme protein of an unusual enzyme family, *PLoS One 8*, 79974.

40. Hurley, L. H., Zmijewski, M., and Chang, C.-J. (1975) Biosynthesis of anthramycin. Determination of the labeling pattern by the use of radioactive and stable isotope techniques, *J. Am. Chem. Soc. 97*, 4372-4378.

41. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389-3402.

42. Kuo, M., Yurek, D., Coats, J., Chung, S., and Li, G. (1992) Isolation and identification of 3-propylidene-Δ-1-pyrroline-5-carboxylic acid, a biosynthetic precursor of lincomycin, *J. Antibiot. 45*, 1773-1777.

43. Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds — from one strategy to four, *Nat. Rev. Microbiol. 9*, 803-816.

44. Krastanov, A., Alexieva, Z., and Yemendzhiev, H. (2013) Microbial degradation of phenol and phenolic derivatives, *Eng. Life Sci. 13*, 76-87.

45. Harayama, S., and Rekik, M. (1989) Bacterial aromatic ring-cleavage enzymes are

classified into two different gene families, *J. Biol. Chem. 264*, 15328-15333.

46. Neusser, D., Schmidt, H., Spizek, J., Novotna, J., Peschke, U., Kaschabeck, S., Tichy, P., and Piepersberg, W. (1998) The genes *lmbB1* and *lmbB2* of *Streptomyces lincolnensis* encode enzymes involved in the conversion of L-tyrosine to propylproline during the biosynthesis of the antibiotic lincomycin A, *Arch. Microbiol. 169*, 322-332.

47. Novotna, J., Honzatko, A., Bednar, P., Kopecky, J., Janata, J., and Spizek, J. (2004) L-3,4-dihydroxyphenyl alanine-extradiol cleavage is followed by intramolecular cyclization in lincomycin biosynthesis, *Eur. J. Biochem. 271*, 3678-3683.

48. Colabroy, K. L., Hackett, W. T., Markham, A. J., Rosenberg, J., Cohen, D. E., and Jacobson, A. (2008) Biochemical characterization of L-DOPA 2,3-dioxygenase, a single-domain type I extradiol dioxygenase from lincomycin biosynthesis, *Arch. Biochem. Biophys. 479*, 131-138.

49. Mossessova, E., and Lima, C. D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast, *Mol. Cell 5*, 865-876.

50. Amor-Mahjoub, M., Suppini, J.-P., Gomez-Vrielyunck, N., and Ladjimi, M. (2006) The effect of the hexahistidine-tag in the oligomerization of HSC70 constructs, *J. Chromatogr. B 844*, 328-334.

51. Esbelin, J., Jouanneau, Y., Armengaud, J., and Duport, C. (2008) *Apo*Fnr binds as a monomer to promoters regulating the expression of enterotoxin genes of *Bacillus cereus*, *J. Bacteriol. 190*, 4242-4251.

52. Karplus, M. (1959) Contact electron-spin coupling of nuclear magnetic moments, *J. Chem. Phys. 30*, 11-15.

53. Bystrov, V. F. (1972) Spin-spin coupling between geminal and vicinal protons, *Russ. Chem. Rev. 41*, 281-304.

54. Colabroy, K. L., Smith, I. R., Vlahos, A. H. S., Markham, A. J., and Jakubik, M. E. (2014) Defining a kinetic mechanism for L-DOPA 2,3 dioxygenase, a single-domain type I extradiol dioxygenase from *Streptomyces lincolnensis*, *Biochim. Biophys. Acta 1844*, 607-614.

55. Burks, E. A., Yan, W., Johnson, W. H., Jr., Li, W., Schroeder, G. K., Min, C., Gerratana, B., Zhang, Y., and Whitman, C. P. (2011) Kinetic, crystallographic, and mechanistic characterization of TomN: elucidation of a function for a 4-oxalocrotonate tautomerase homologue in the tomaymycin biosynthetic pathway, *Biochemistry 50*, 7600-7611.

56. Terradas, F., and Wyler, H. (1991) 2,3- and 4,5-*seco*DOPA, the biosynthetic

intermediates generated from L-DOPA by an enzyme system extracted from the fly agaric, *Amanita muscaria* L., and their spontaneous conversion to muscaflavin and betalamic Acid, respectively, and betalains, *Helv. Chim. Acta 74*, 124-140.

57. Bugg, T. D. H., and Ramaswamy, S. (2008) Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations, *Curr. Opin. Chem. Biol. 12*, 134-140.

58. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A., (Eds.) (2005) *Protein identification and analysis tools on the ExPASy server*, John M. Walker ed., The Proteomics Protocols Handbook, Humana Press: Totowa, NJ.

59. Hwang, T. L., and Shaka, A. J. (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients, *J. Magn. Reson. Ser. A 112*, 275-279.

60. Saha, S., Li, W., Gerratana, B., and Rokita, S. E. (2015) Identification of the dioxygenase-generated intermediate formed during biosynthesis of the dihydropyrrole moiety common to anthramycin and sibiromycin, *Bioorg. Med. Chem. 23*, 449-454.

61. O'Neil, M. N., Smith, A., Heckelman, P. E., and Budavari, S., (Eds.) (2001) *The Merck index: an encyclopedia of chemicals, drugs, and biologicals*, 13th ed., Merck & Co.: Whitehouse Station, NJ.

62. Blankenfeldt, W., Kuzin, A. P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L. S., and Mavrodi, D. V. (2004) Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16431-16436.

63. Parsons, J. F., Song, F., Parsons, L., Calabrese, K., Eisenstein, E., and Ladner, J. E. (2004) Structure and function of the phenazine biosynthesis protein PhzF from *Pseudomonas fluorescens* 2-79, *Biochemistry 43*, 12427-12435.

64. Yonemoto, I. T., and Gerratana, B. Unpublished work.

65. Ulanova, D., Novotná, J., Smutná, Y., Kameník, Z., Gažák, R., Šulc, M., Sedmera, P., Kadlčík, S., Plháčková, K., and Janata, J. (2010) Mutasynthesis of lincomycin derivatives with activity against drug-resistant *Staphylococci*, *Antimicrob. Agents Chemother. 54*, 927-930.

66. Gerratana, B. Unpublished work.

67. Min, C., and Gerratana, B. Unpublished work.

68. Laker, M. F., Hoffman, A. F., and Meeuse, J. D. (1980) Spectrophotometric

determination of urinary oxalate with oxalate prepared from moss., *Clin. Chem. 26*, 827-830.

69. Li, J.-J., and Bugg, T. D. H. (2007) Investigation of a general base mechanism for ester hydrolysis in C-C hydrolase enzymes of the α/β-hydrolase superfamily: a novel mechanism for the serine catalytic triad, *Org. Biomol. Chem. 5*, 507-513.

70. Pircher, H., Straganz, G. D., Ehehalt, D., Morrow, G., Tanguay, R. M., and Jansen-Dürr, P. (2011) Identification of human fumarylacetoacetate hydrolase domain-containing protein 1 (FAHD1) as a novel mitochondrial acylpyruvase, *J. Biol. Chem 286*, 36500-36508.

71. Beckett, D. (2011) Measurement and analysis of equilibrium binding titrations: a beginner's guide, *Methods Enzymol. 488*, 1-16.

72. Herde, P., and Blankenfeldt, W. (2006) The purification, crystallization and preliminary strutural characterization of human MAWDBP, a member of the phenazine biosynthesis-like protein family, *Acta. Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. F62*, 546-549.

73. Chuenchor, W., Doukov, T. I., and Gerratana, B. Unpublished work.

74. Usha, V., Dover, L. G., Roper, D. L., and Besra, G. S. (2008) Characterization of *Mycobacterium tuberculosis* diaminopimelic acid epimerase: paired cysteine residues are crucial for racemization, *FEMS Microbiol. Lett. 280*, 57-63.

75. Raponi, M., and Arndt, G. M. (2002) Dominant genetic screen for cofactors that enhance antisense RNA-mediated gene silencing in fission yeast, *Nucleic Acids Res. 30*, 2546-2554.

76. Baldwin, J. E., Bamford, S. J., Fryer, A. M., Rudolph, M. P., and Wood, E. M. (1997) Towards a versatile synthesis of kainoids I: introduction of the C-3 and C-4 substituents *Tetrahedron 53*, 5233-5254.

77. O'Connell, C. E., Rowell, C. A., Ackermann, K., Garcia, A. M., Lewic, M. D., and Kowalczyk, J. J. (2000) Synthesis and evaluation of some hydroxyproline-derived peptidomimetics as isoprenyltransferase inhibitors., *Chem. Pharm. Bull. 48*, 740-742.

78. Murcko, M. A., Castejon, H., and Wiberg, K. B. (1996) Carbon−carbon rotational barriers in butane, 1-butene, and 1,3-butadiene, *J. Phys. Chem. 100*, 16162-16168.

79. Huang, Y.-T., Lyu, S.-Y., Chuang, P.-H., Hsu, N.-S., Li, Y.-S., Chan, H.-C., Huang, C.-J., Liu, Y.-C., Wu, C.-J., Yang, W.-B., and Li, T.-L. (2009) *In vitro* characterization of enzymes involved in the synthesis of nonproteinogenic residue (2S,3S)-β-methylphenylalanine in glycopeptide antibiotic mannopeptimycin, *ChemBioChem 10*, 2480-2487.

80. Baldwin, J. E., Bamford, S. J., Fryer, A. M., and Wood, M. E. (1995) A versatile approach to acromelic acid analogues, *Tetrahedron Lett. 36*, 4869-4872.

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

82. Phelan, V. V., Du, Y., McLean, J. A., and Bachmann, B. O. (2009) Adenylation enzyme characterization using γ-$^{18}O_4$-ATP pyrophosphate exchange, *Chem. Biol. 16*, 473-478.

83. Imker, H. J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C. T. (2010) *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF, *Chem. Biol. 17*, 1077-1083.

84. Kong, L., and Ranganathan, S. (2004) Delineation of modular proteins: domain boudnary prediction from sequence information, *Brief. Bioinform 5*, 179-192.

85. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank., *Nucleic Acids Res. 28*, 235-242.

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

87. Xu, Y., Orozco, R., Kithsiri Wijeratne, E. M., Espinosa-Artiles, P., Leslie Gunatilaka, A. A., Patricia Stock, S., and Molnár, I. (2009) Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of *Beauveria bassiana*, *Fungal Genet. Biol. 46*, 353-364.

88. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D., and E. Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*, *Science 291*, 1790-1792.

89. Kadlcík, S., Kucera, T., Chalupská, D., Gazák, R., Koberská, M., Ulanová, D., Kopecký, J., Kutejová, E., Najmanová, L., and Janata, J. (2013) Adaptation of an L-proline adenylation domain to use 4-propyl- L-proline in the evolution of lincosamide biosynthesis, *PLoS One 8*.

90. Crüsemann, M., Kohlhaas, C., and Piel, J. (2013) Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains, *Chemical Sci. 4*, 1041-1045.

91. Najmanová, L., Kutejová, E., Kadlec, J., Polan, M., Olšovská, J., Benada, O., Novotná, J., Kameník, Z., Halada, P., Bauer, J., and Janata, J. (2013) Characterization of *N*-demethyllincosamide methyltransferases LmbJ and CcbJ, *ChemBioChem 14*, 2259-

2262.

92. Cai, X., Teta, R., Kohlhaas, C., Crüsemann, M., Ueoka, R., Mangoni, A., Freeman, Michael F., and Piel, J. (2013) Manipulation of regulatory genes reveals complexity and fidelity in hormaomycin biosynthesis, *Chem. Biol. 20*, 839-846.

93. Connor, K. L., and Gerratana, B. Unpublished work.

94. Ehmann, D. E., Trauger, J. W., Stachelhaus, T., and Walsh, C. T. (2000) Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases, *Chem. Biol. 7*, 765-772.

95. Lautru, S., and Challis, G. L. (2004) Substrate recognition by nonribosomal peptide synthetase multi-enzymes, *Microbiol. 150*, 1629-1636.

96. Nelson, J. T., Lee, J., Sims, J. W., and Schmidt, E. W. (2007) Characterization of SafC, a catechol 4-*O*-methyltransferase involved in saframycin biosynthesis, *Appl. Environ. Microbiol. 73*, 3575-3580.

97. Varga, M., and Varga, E. (2014) Development and validation of an LC-MS/MS method for the analysis of L-DOPA in oat, *Acta Biol. Szegad 58*, 133-137.

98. Gaudelli, N. M., and Townsend, C. A. (2013) Stereocontrolled syntheses of peptide thioesters containing modified seryl residues as probes of antibiotic biosynthesis, *J. Org. Chem. 78*, 6412-6426.

99. Davidsen, J. M., Bartley, D. M., and Townsend, C. A. (2013) Non-ribosomal oropeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI, *J. Am. Chem. Soc. 135*, 1749-1759.

100. Pickens, L. B., Tang, Y., and Chooi, Y.-H. (2011) Metabolic engineering for the production of natural products, *Annu. Rev. Chem. Biomol. Eng. 2*, 211-236.

101. Leonard, E., and Koffas, M. A. G. (2007) Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*, *Appl. Environ. Microbiol. 73*, 7246-7251.

102. Chang, M. C. Y., Eachus, R. A., Trieu, W., Ro, D.-K., and Keasling, J. D. (2007) Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s, *Nat. Chem. Biol. 3*, 274-277.

103. Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W.-H., Mo, J. D., Tidor, B., Stephanopoulos, G., and Prather, K. L. J. (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control, *Proc. Natl. Acad. Sci. U.S.A. 107*, 13654-13659.

104. Wang, C.-w., Oh, M.-K., and Liao, J. C. (2000) Directed evolution of metabolically engineered *Escherichia coli* for carotenoid production, *Biotechnol. Prog. 16*, 922-926.

105. McMahon, M. D., Rush, J. S., and Thomas, M. G. (2013) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*, *J. Bacteriol. 194*, 2809-2818.

106. Zhang, C., Kong, L., Liu, Q., Lei, X., Zhu, T., Yin, J., Lin, B., Deng, Z., and You, D. (2013) *In vitro* characterization of echinomycin biosynthesis: formation and hydroxylation of L-tryptophanyl-*S*-enzyme and oxidation of (2S,3S) β-hydroxytryptophan, *PLoS One 8*, e56772.

107. Bosello, M., Zeyadi, M., Kraas, F. I., Linne, U., Xie, X., and Marahiel, M. A. (2013) Structural characterization of the heterobactin siderophores from *Rhodococcus erythropolis* PR4 and elucidation of their biosynthetic machinery, *J. Nat. Prod. 76*, 2282-2290.

108. Hiratsuka, T., Koketsu, K., Minami, A., Kaneko, S., Yamazaki, C., Watanabe, K., Oguri, H., and Oikawa, H. (2013) Core assembly mechanism of quinocarcin/SF-1739: bimodular complex nonribosomal peptide synthetases for sequential Mannich-type reactions, *Chem. Biol. 20*, 1523-1535.

109. Al-Mestarihi, A. H., Villamizar, G., Fernández, J., Zolova, O. E., Lombó, F., and Garneau-Tsodikova, S. (2014) Adenylation and *S*-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis, *J. Am. Chem. Soc. 136*, 17350-17354.

110. Zolova, O., E., and Garneau-Tsodikova, S. (2014) KtzJ-dependent serine activation and *O*-methylation by KtzH for kutznerides biosynthesis, *J. Antibiot. 67*, 59-64.

111. Baltz, R. H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery, *J. Ind. Microbiol. Biotechnol. 38*, 1747-1760.

112. Baltz, R. H. (2014) MbtH homology codes to identify gifted microbes for genome mining, *J. Ind. Microbiol. Biotechnol. 41*, 357-369.

113. Drake, E. J., Cao, J., Qu, J., Shah, M. B., Straubinger, R. M., and Gulick, A. M. (2007) The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*, *J. Biol. Chem. 282*, 20425-20434.

114. Buchko, G. W., Kim, C.-Y., Terwilliger, T. C., and Myler, P. J. (2010) Solution structure of Rv2377c-founding member of the MbtH-like protein family, *Tuberculosis 90*, 245-251.

115. Zolova, O. E., and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK, *Med. Chem. Commun. 3*, 950-955.

116. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *J. Biol. Chem. 288*, 1991-2003.

117. Lautru, S., Oves-Costales, D., Pernodet, J.-L., and Challis, G. L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145, *Microbiol. 153*, 1405-1412.

118. Al-Mestarihi, A. H., Garzan, A., Kim, J. M., and Garneau-Tsodikova, S. (2015) Enzymatic evidence for a revised congocidine biosynthetic pathway, *ChemBioChem 16*, 1307-1313.

119. Wolpert, M., Gust, B., Kammerer, B., and Heide, L. (2007) Effects of deletions of *mbtH*-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*, *Microbiol. 153*, 1413-1423.

120. Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases, *Biochemistry 49*, 8815-8817.

121. Zhang, W., Heemstra, J. R., Walsh, C. T., and Imker, H. J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins, *Biochemistry 49*, 9946-9947.

122. Boll, B., Taubitz, T., and Heide, L. (2011) Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis, *J. Biol. Chem. 286*, 36281-36290.

123. Tatham, E., sundaram Chavadi, S., Mohandas, P., Edupuganti, U., Angala, S., Chatterjee, D., and Quadri, L. E. (2012) Production of mycobacterial cell wall glycopeptidolipids requires a member of the MbtH-like protein family, *BMC Microbiol. 12*, 118.

124. Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters, *Nucleic Acids Res. 43*, W237-W243.

125. Boll, B., and Heide, L. (2013) A domain of RubC1 of rubradirin biosynthesis can functionally replace MbtH-like proteins in tyrosine adenylation, *ChemBioChem 14*, 43-44.

126. Janata, J., Kadlcik, S., Koberska, M., Ulanova, D., Kamenik, Z., Novak, P., Kopecky, J., Novotna, J., Radojevic, B., Plhackova, K., Gazak, R., and Najmanova, L. (2015) Lincosamide synthetase - a unique condensation system combining elements of nonribosomal peptide synthetase and mycothiol metabolism, *PLoS One 10*, e0118850.

127. Zhao, Q., Wang, M., Xu, D., Zhang, Q., and Liu, W. (2015) Metabolic coupling of two small-molecule thiols programs the biosynthesis of lincomycin A, *Nature 518*, 115-119.

128. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G. M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P. G. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis, *Electrophoresis 25*, 1327-1333.

129. Edgar, R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinformatics 5*, 113.

130. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res. 32*, 1792-1797.

131. Kamat, S. S., Williams, H. J., and Raushel, F. M. (2011) Intermediates in the transformation of phosphonates to phosphate by bacteria, *Nature 480*, 570-573.