UNIVERSITY OF CALIFORNIA, SAN DIEGO

Biosynthesis and Engineering of Cyclomarin and Cyclomarazine: Prenylated, Non-Ribosomal Cyclic Peptides of Marine Actinobacterial Origin

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

in

Oceanography

by

Andrew William Schultz

Committee in charge:

Professor Bradley Moore, Chair
Professor Eric Allen
Professor Pieter Dorrestein
Professor William Fenical
Professor William Gerwick

2010

The Dissertation of Andrew William Schultz is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2010

DEDICATION

To my wife Elizabeth and our son Orion

and

To my parents Dale and Mary

Thank you for your never ending love and support

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenylation domain
ACP	Acyl carrier protein
ADH	2-amino-3,5-dimethylhex-4-enoic acid
AT	Acyltransferase domain
B:	Basic residue
BLAST	Basic local alignment search tool
C	Condensation domain
CHA	L-3-cyclohex-2′-enylalanine
CoA	Coenzyme A
Cy	Cyclase domain
cym	Cyclomarin biosynthetic gene cluster
CYP	Cytochrome P450
Da	Dalton
DKP	Diketopiperazine
DMAPP	Dimethylallyl diphosphate
E. coli	*Escherichia coli*
EtBr	Ethidium bromide
EtOAc	Ethyl acetate
ER	Enoyl reductase domain
FMN	Flavin mononucleotide
HMBC	Heteronuclear multiple bond correlation
HPLC	High pressure liquid chromatography
HSQC	Heteronuclear multiple quantum correlation

kb	Kilobase
KS	Ketosynthase domain
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometery
MeCN	Acetonitrile
MeOH	Methanol
MS	Mass spectrometery
MT	Methyltransferase domain
NAD^+	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
NPAA	Non-proteinogenic amino acid
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
ORF	Open reading frame
Ox	Oxidation domain
P450	Cytochrome P450
PCR	Polymerase chain reaction
PPant	Phosphopantetheinyl
PK	Polyketide
PKS	Polyketide synthase
PTase	Prenyltransferase
RBS	Ribosome binding site
Red	Reductive offloading domain
SAM	*S*-Adenosyl methionine

T	Thiolation domain (peptidyl carrier protein)
TE	Thioesterase domain

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to sincerely thank everyone, mentioned here or not, who has developed my interest in science, been there to support me, or has directly mentored me throughout the process of acquiring my doctorate. I would like to thank and acknowledge several people in particular:

I would like to thank Professor Bradley Moore for his dedication to my development as a scientist and an individual. I thank you for your patience and your enthusiasm, your willingness to talk at any time, and your flexibility while maintaining a dedication to excellence. You were there when I needed guidance, but gave me the freedom to manage my project, which has been instrumental in my development as an independent researcher. I thank you for giving me the opportunity to join your lab and for all your mentoring both in science and life.

I would also like to thank the rest of my dissertation committee, Professors Eric Allen, Pieter Dorrestein, William Fenical, and William Gerwick for the guidance and assistance you have provided. Special thanks go to Professors Fenical and Jensen along with Dong-Chan Oh and Ratnakar Asolkar for discovering cyclomarin and cyclomarazine, for providing the genome sequence along with cultures of *Salinispora arenicola* CNS-205, and for advice on fermenting and isolating analogs of these compounds. Additional thanks goes to Professor Dorrestein and Wei-Ting Liu for your sincere interest in my work and for our collaboration on sequencing cyclomarin analogs. I would like to thank Professor Joseph Noel, Stephanie Richard, and Monica Tello of the Salk Institute

for training and advice in protein expression and purification. I would also like to thank our collaborators Professor Phil Baran, Chad Lewis, Mike Luzung, Professor Steven Gould, John Carney, and Thomas Williamson.

I also would like to thank everyone I have worked with in the Moore lab over the last 5 years. Every single one of you has been instrumental in my development as a scientist and as a person. I will always have fond memories of my time working with you. Everyone has been willing to help with a technique or give some advice at a moment's notice, which has been and still is greatly appreciated. I will miss the coffee breaks and happy hours! A special thanks goes to Diana Plutchak for all your help, I hope you enjoyed learning form me as much as I enjoyed mentoring you.

Last but not least, I would like to thank Elizabeth for her support throughout this process. You have been there for me through the ups and downs. This challenge has made our relationship stronger instead of weaker, and without your moral support to keep me going, I don't know what I would do. This process has really been a team effort! I love you!

Chapter 2, in full, is a reprint of the material as it appears in Biosynthesis and Structures of Cyclomarins and Cyclomarazines, Prenylated Cyclic Peptides of Marine Actinobacterial Origin (2008). Schultz, Andrew W.; Oh, Dong-Chan; Carney, John R.; Williamson, R. Thomas; Udwary, Daniel W.; Jensen, Paul R.; Gould, Steven J.; Fenical, William; Moore, Bradley S., Journal of the American Chemical Society, 130, 4507–4516. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Functional Characterization of the Cyclomarin/Cyclomarazine Prenyltransferase CymD Directs the Biosynthesis of Unnatural Cyclic Peptides (2010). Schultz, Andrew W.; Lewis, Chad A.; Luzung, Michael R.; Baran, Phil S.; Moore, Bradley S., Journal of Natural Products, 73, 373–377. The dissertation author was the primary investigator and author of this paper.

VITA

2002 Bachelor of Science, Purdue University, West Lafayette, IN

2002-2005 Advanced Product Specialist, Abbott, Abbot Park, IL

2005-2010 Research Assistant, University of California, San Diego, CA

2010 Doctor of Philosophy, University of California, San Diego, CA

PUBLICATIONS

A. W. Schultz, D.-C. Oh, J. R. Carney, R. T. Williamson, D. W. Udwary, P. R. Jensen, S. J. Gould, W. Fenical, and B. S. Moore. “Biosynthesis and Structures of Cyclomarins and Cyclomarazines, Prenylated Cyclic Peptides of Marine Actinobacterial Origin.” J. Am. Chem. Soc., 131, 4507-4516 (2008)

W.-T. Liu, J. Ng, D. Meluzzi, N. Bandeira, M. Gutierrez, T. L. Simmons, A. W. Schultz, R. Linington, B. S. Moore, W. Gerwick, D. Pevzner, and P. C. Dorrestein, “The Structural Characterization of Cyclic Non-Ribosomal Peptides by Tandem Mass Spectrometry.” Anal. Chem., 81, 4200-4209 (2009)

A. W. Schultz, C. A. Lewis, M. R. Luzung, P. S. Baran and B. S. Moore, "Functional Characterization of the Cyclomarin/Cyclomarazine Prenyltransferase CymD Directs the Biosynthesis of Unnatural Cyclic Peptides."J. Nat. Prod., 73, 373-377 (2010)

A. A. Roberts, A. W. Schultz, R. D. Kersten, P. C. Dorrestein, and B. S. Moore, "Iron Acquisition in the Obligate Marine Actinomycete Genus Salinispora." in preparation (2010)

FIELD OF STUDY

Major Field: Biosynthesis of Marine Derived Natural Products

ABSTRACT OF THE DISSERTATION

Biosynthesis and Engineering of Cyclomarin and Cyclomarazine: Prenylated, Non-Ribosomal Cyclic Peptides of Marine Actinobacterial Origin

by

Andrew William Schultz

Doctor of Philosophy in Oceanography

University of California, San Diego, 2010

Professor Bradley Moore, Chair

Cyclomarins are potent anti-inflammatory cyclic peptides originally encountered in the estuarine streptomycete, strain CNB-982, and subsequently isolated from the marine obligate actinobacterium *Salinispora arenicola* CNS-205 along with the truncated antibacterial diketopiperazine cyclomarazines. The cyclomarins are composed of seven amino acid residues, two of which are of unique biosynthetic interest, 2-amino-3,5-dimethylhex-4-enoic acid (ADH), which has never been reported as a constituent of any peptide, and *N*-(1,1-dimethyl-1-allyl)- tryptophan, which is a rare example of the prenylation of a bacterial non-ribosomal peptide. The *N*-prenylated tryptophan residue, along with δ-

hydroxyleucine, is present in both the cyclomarins and cyclomarazines, suggesting a common biosynthetic origin.

Because of the unique structural features of ADH and *N*-(1,1-dimethyl-1-allyl)- tryptophan, along with the extensive tailoring of the remaining residues of cyclomarin, we initiated the interrogation of the biosynthetic machinery behind cyclomarin and cyclomarazine synthesis. We were able to determine the basic steps of cyclomarin and cyclomarazine production, along with insights into how the two unique amino acids are synthesized. Furthermore, we were able to identify the oxidative enzymes involved in tailoring both peptides. Analysis of the genome sequence of *S. arenicola* CNS-205 revealed the *cym* biosynthetic gene cluster, which is dominated by *cymA*, encoding for a seven module non-ribosomal peptide synthetase. Downstream of *cymA* are the genes for *N*-(1,1-dimethyl-1-allyl)- tryptophan synthesis via the prenyltransferase CymD, and ADH via *cymE-H*. CymEF is an aldolase/dehydrogenase pair utilized for the key C-C bond forming step of ADH synthesis, along with a methyltransferase CymG, and a dehydrogenase CymH. Upstream of *cymA* is the Fe^{2+}/α-ketoglutarate dependent leucine δ-oxygenase CymW, and three cytochrome P450s oxidases, CymO, CymS, and CymV. Bioinformatic analysis coupled with in vivo mutagenesis confirmed the role of Cym W along with CymS as a *N*-(1,1-dimethyl-1-allyl)- tryptophan β-hydroxylase, and CymV as a *N*-(1,1-dimethyl-2,3-epoxypropyl)-tryptophan epoxidase, and suggested the function of CymO as a phenylalanine β-hydroxylase . In addition to establishing the function of the

aforementioned enzymes, we were able to generate the novel *N*-(1-methyl)-tryptophan and *N*-(1-propargyl)-tryptophan analogs of both sized peptides in the *cymD* deficient mutant, confirming the preference for *N*-alkylated tryptophan residues and the amenability of cyclomarin and cyclomarazine to derivatization via mutasynthesis.

Chapter 1: Introduction

The Chemical Diversity of Non-ribosomal Peptides

1.1: Introduction to non-ribosomal peptides

Non-ribosomal peptides (NRPs) have saved innumerable lives. Penicillin has been used for nearly three quarters of a century to treat bacterial infections. Nearly half of all antibiotics prescribed worldwide are β-lactams,[1] a structural class of non-ribosomal peptides which includes the penicillins (Figure 1.1) and cephalosporins. The non-ribosomal glycopeptide vancomycin (Figure 1.2) has been utilized to treat methicillin resistant *Staphylococcus aureus* (MRSA) and other serious infections[2] as an "antibiotic of last resort." One of the most recently approved antibiotics is daptomycin (Figure 1.3), a non-ribosomal lipopeptide, which represents the first new structural class in antibiotics in the last thirty years.[3] Life saving NRPs are not exclusively antibiotics; for instance, the immunosuppressant cyclosporine (Figure 1.1) has given many people a new lease on life by preventing the rejection of donor organs and tissues.[4]

Humans have found a use for NRPs in medicine. Yet where do these natural products come from, and what are their roles in nature? Although derived from amino acids, NRPs are not synthesized by the ribosome, but are instead synthesized via modular megasynthetases. They fall into a class of compounds deemed "secondary metabolites," originally believed to have non-essential, or secondary, roles in the metabolism of their host producers.[5,6] This is in contrast to primary metabolites, which serve a common, well conserved role in the majority of organisms, supporting the most basic of life's functions such as cellular respiration and energy storage, cellular structure, and genetic information storage and transfer. The roles of secondary metabolites in nature are much less

well understood, but play a role in chemical defense against predators[7] or competitors[8] or intra- or interspecies communication.[9] The current view of secondary metabolites is that they are not “secondary”, but are integral to the survival and/or reproduction of an organism. Without these metabolites to facilitate communication or prevent predation, the organism (or species) would not be able survive in their environment, and thus are essential for their survival. These roles can be quite complex, one example of which is the oxylipins produced by the eggs of several brown algal species.[10] These compounds act as a chemoattractant for spermatozoids, facilitating sexual reproduction. Degradants of these compounds act as a feeding deterrent, making the sessile brown algae unpalatable or toxic to predators, such sea urchins and amphipods.[11] This dual functionality exemplifies how a compound can have a signaling role in its host, while having a strong bioactivity in an animal, which is an important parallel to medicinal use of secondary metabolites by humans.

The marine environment provides numerous examples of NRPs (Figure 1.4, Table 1.1), such as the actinobacterial compounds cyclomarin,[12] cyclomarazine,[13] and salinosporamide.[14] Cyanobacteria are known as prolific producers of NRPs, such as lyngbyatoxin,[15] nostocyclopeptide,[16] nostopeptolides,[17] and microcystin.[18] Marine invertebrates are also represented by the sponge-derived jaspamide,[19] the sea hare compound dolastatin[20] and the tunicate-derived ET-743,[21] although these compounds are most likely provided via symbiosis or dietary consumption of bacterial producers of these

compounds.[22,23] ET-743[21,24] has been approved as an anti-cancer agent, while salinosporamide[25] and dolastatin[24] are in clinical trials as such (Table 1.1).

Although several marine NRPs are known and a few are in development as drugs,[24] the most famous and medicinally relevant examples come from the terrestrial environment. In addition to penicillin, cephalosporin, vancomycin, and daptomycin, the NRPs chloramphenicol,[26] teichoplanin,[27] and novobiocin[28] have been or are in use as antibiotics. Other indications are also provided by terrestrial NRPs, such as the approved anti-cancer agents bleomycin[24] while epothilone B[29] and its analog ixabepilone[30] are in clinical trials. The immunosuppressant activities of cyclosporine[27] are utilized in anti-rejection therapies in transplant recipients and nikkomycin[31] has entered clinical trials as an anti-fungal.

Although many NRPs are famous for curing diseases, some are medically relevant as potent toxins or virulence factors, or act as surfactants. The microcystins are powerful hepatotoxins,[18] produced by various cyanobacteria. Lyngbyatoxin,[15] a protein kinase C activating compound, is the causative agent of swimmer's itch. Pyochelin,[32] yersiniabactin[33] vibriobactin[34] are iron chelating siderophores that have been implicated as virulence factors[35] in their host organisms, leading way to opportunistic infections, the plague, and dysentery. Syringomycin[36] and surfactin[37] both have detergent properties, and syringomycin has been noted as a virulence factor in *Pseudomonas syringae* infections in stone fruit crop plants.[36]

One of the impediments to the development of natural products as drugs is the difficulty in obtaining sufficient quantities for development and clinical use. This is even more problematic for marine compounds, especially those sourced from marine invertebrates.[24,38] In addition to difficulties sourcing some of these compounds, there is a constant need to find analogs to thwart antibiotic resistance mechanisms[39-41] or the need to generate derivatives to make the compound a better drug by optimizing solubility or bioavailablity.[42] One of the most promising solutions to these issues is to gain an understanding how these compounds are produced at the genetic and biochemical levels, and then utilize this knowledge to manipulate the biosynthesis of the compound of interest in the host organism, or to engineer another organism to produce it.

As shown in Figures 1.1-1.5 and Table 1, NRPs are very diverse in their structures and their bioactivities. How does nature provide so much chemical diversity?[43] What is the mechanism by which is NRPs are biosynthesized? How are the numerous chlorinations, oxidations, prenylations, glycosylations, and other modifications installed on NRPs? How are some of the unusual amino acids, unavailable to ribosomal peptide assembly, produced and incorporated into NRPs? How has the knowledge of these processes been utilized to manipulate the structure of NRPs? The current state of knowledge in NRP biosynthesis as it pertains to these questions will be explored in this chapter. The remainder of this dissertation will focus on efforts toward understanding and manipulating the mechanisms by which cyclomarin and cyclomarazine are produced by the marine actinobacterium *Salinispora arenicola* CNS-205, not in

order to develop these compounds as drugs, but to gain a greater understanding on how these compounds are produced. This knowledge can be utilized by other researchers to discover and manipulate the biosynthesis of other NRPs in an effort to provide new or better drugs or to understand the roles of these compounds in the environment.

Table 1.1: Non-ribosomal peptides, their sources, bioactivities, and clinical use

compound	source	bioactivity	clinical use?	ref
teichoplanin	actinobacteria	anti-bacterial	yes, EU only	[27]
novobiocin	actinobacteria	anti-bacterial	yes[a]	[27]
chloramphenicol	actinobacteria	anti-bacterial	yes	[26,40]
vancomycin	actinobacteria	anti-bacterial	yes	[27]
bleomycin	actinobacteria	anti-cancer	yes	[27]
daptomycin	actinobacteria	anti-bacterial	yes	[27]
nikkomycin	actinobacteria	anti-fungal, insecticide	phase I[b]	[44]
salinosporamide	actinobacteria	anti-cancer	phase II	[14]
cyclomarin	actinobacteria	anti-inflammatory	no	[12]
cyclomarazine	actinobacteria	anti-bacterial	no	[13]
chlorobiocin	actinobacteria	anti-bacterial	no	[28]
kutzneride	actinobacteria	anti-fungal	no	[45]
balhimycin	actinobacteria	anti-bacterial	no	[46]
chloroeremomycin	actinobacteria	anti-bacterial	no	[47]
A54145	actinobacteria	anti-bacterial	no	[48]
CDA	actinobacteria	anti-bacterial	no	[49]
surfactin	bacilli	surfactant	no	[37]
microcystin	cyanobacteria	hepatotoxin	no	[18]
lyngbyatoxin	cyanobacteria	dermal irritant	no	[15]
nostocyclopeptide	cyanobacteria	cytotoxic	no	[16]
nostopeptolide	cyanobacteria	not reported	no	[17]
cyclosporine	fungi	immunosuppressant	yes	[27]
penicillin	fungi	anti-bacterial	yes	[27]
safracin	gammaproteobacteria	anti-bacterial, anti-cancer	no	[50]
syringomycin	gammaproteobacteria	plant pathogenicity	no	[51]
pyochelin	gammaproteobacteria	siderophore	no	[32]
yersiniabactin	gammaproteobacteria	siderophore	no	[33]
vibriobactin	gammaproteobacteria	siderophore	no	[34]
epothilone	myxobacteria	anti-cancer	phase III	[52]
chondramide	myxobacteria	anti-cancer	no	[53]
dolastatin	sea hare	anti-cancer	phase II[b],	[20]
jaspamide	sponge	anti-cancer	no	[19]
ET-743 (Yondelis)	tunicate	anti-cancer	yes	[24]

[a]Removed from market and/or very restricted use
[b]Clinical trials discontinued for various reasons, but analogs are continuing

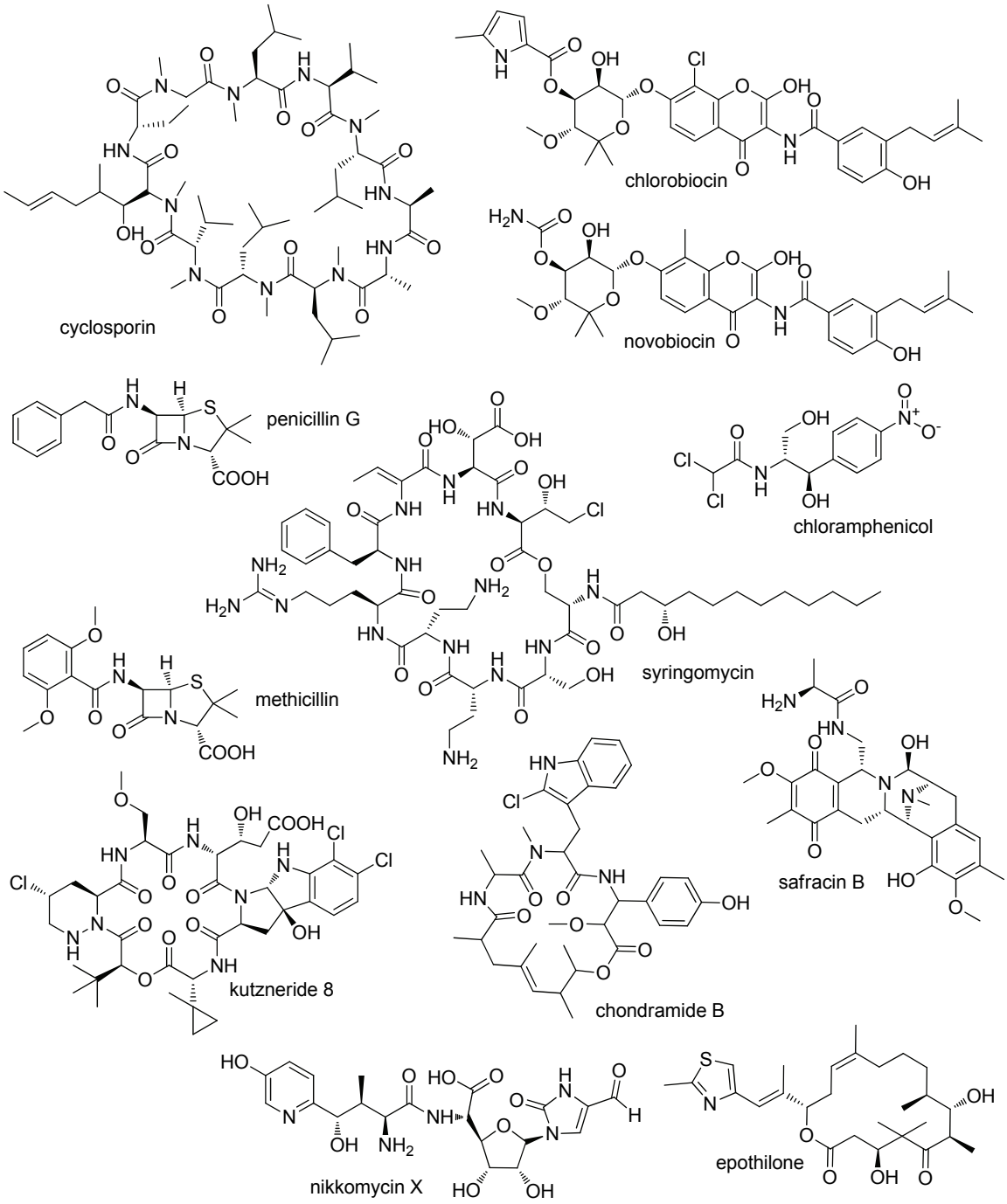


Figure 1.1: Terrestrial non-ribosomal peptides. The stereochemistry of chondramide B has not been reported.

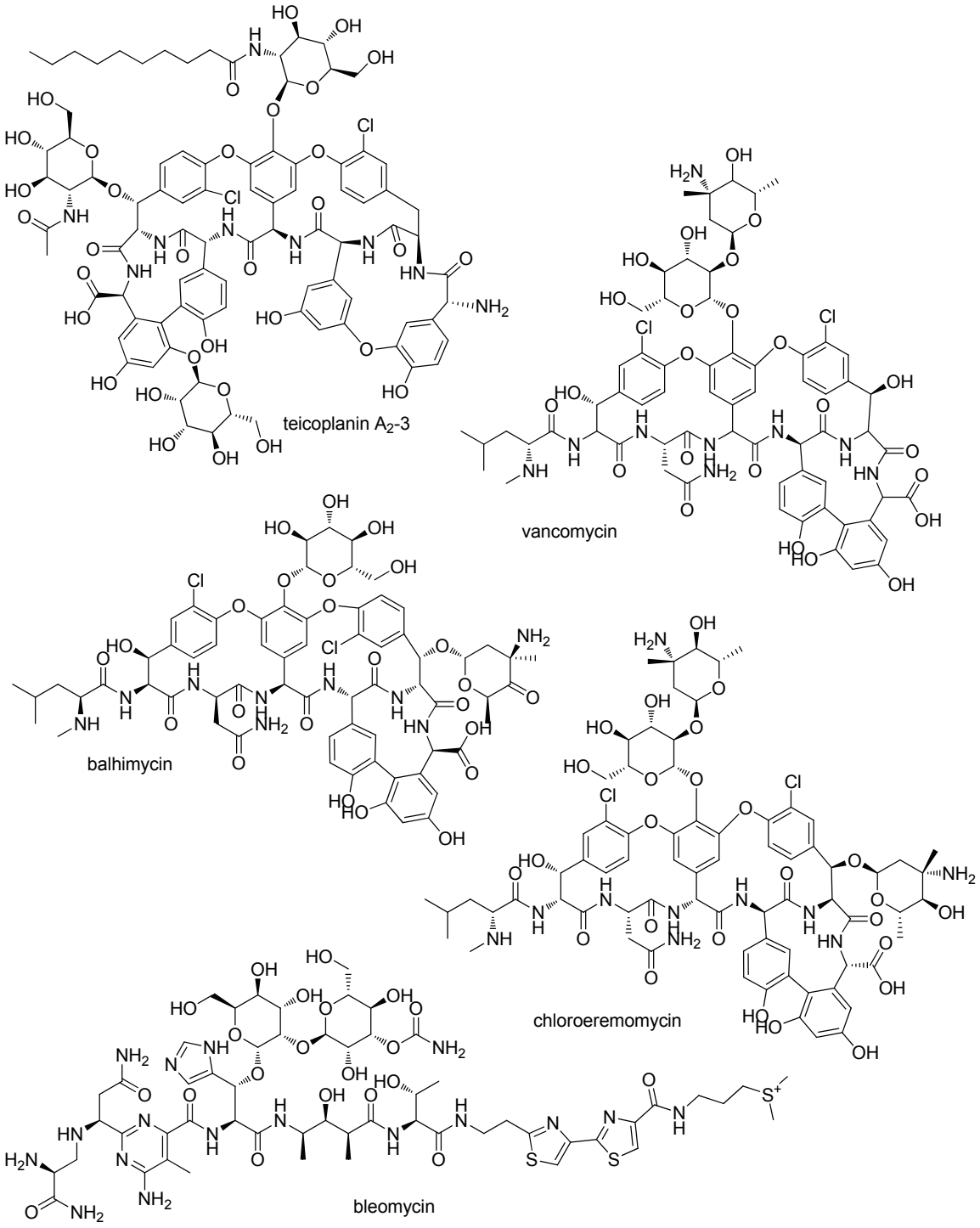


Figure 1.2: Glycopeptide antibiotics

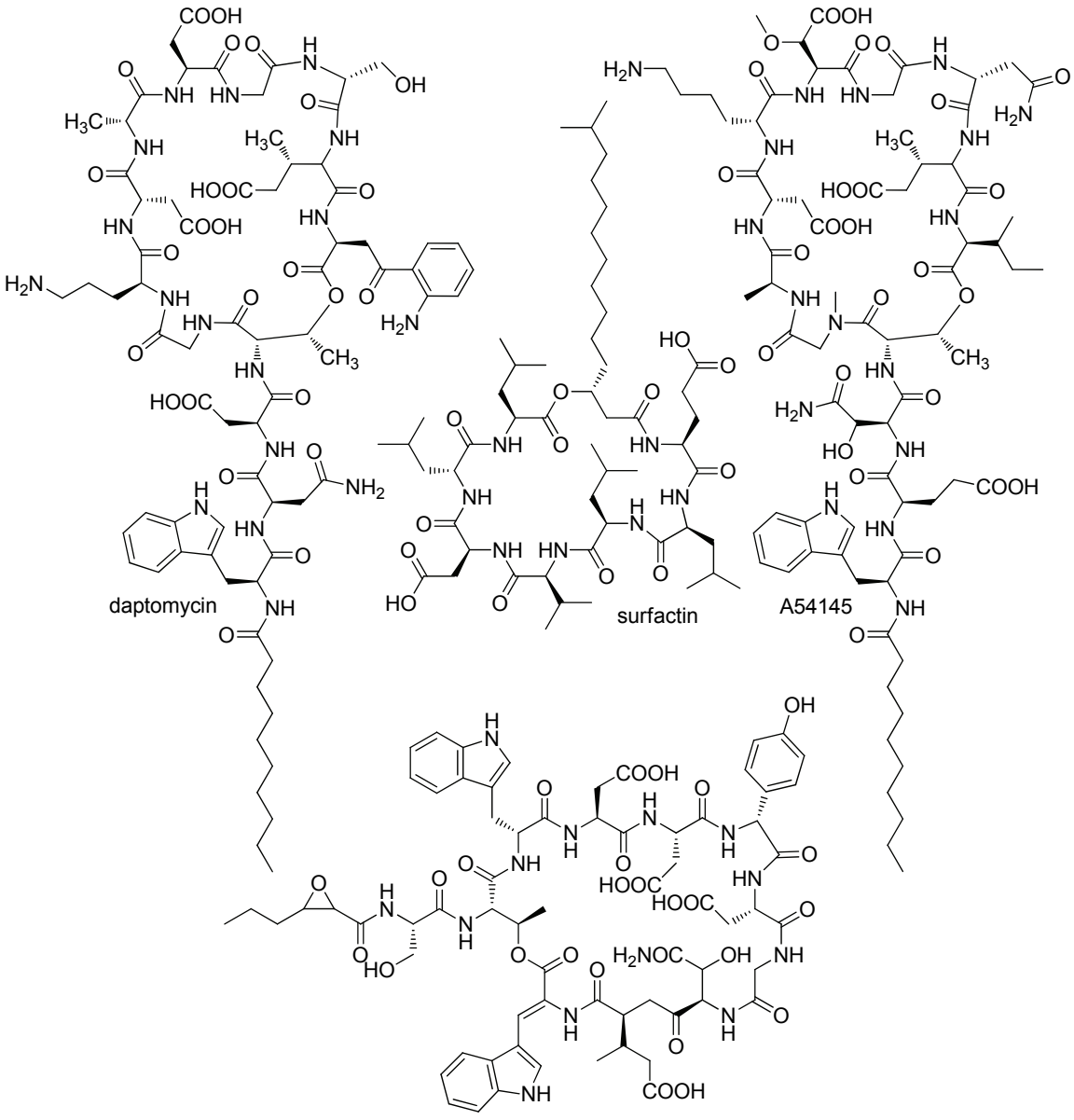


Figure 1.3: Lipopeptide antibiotics

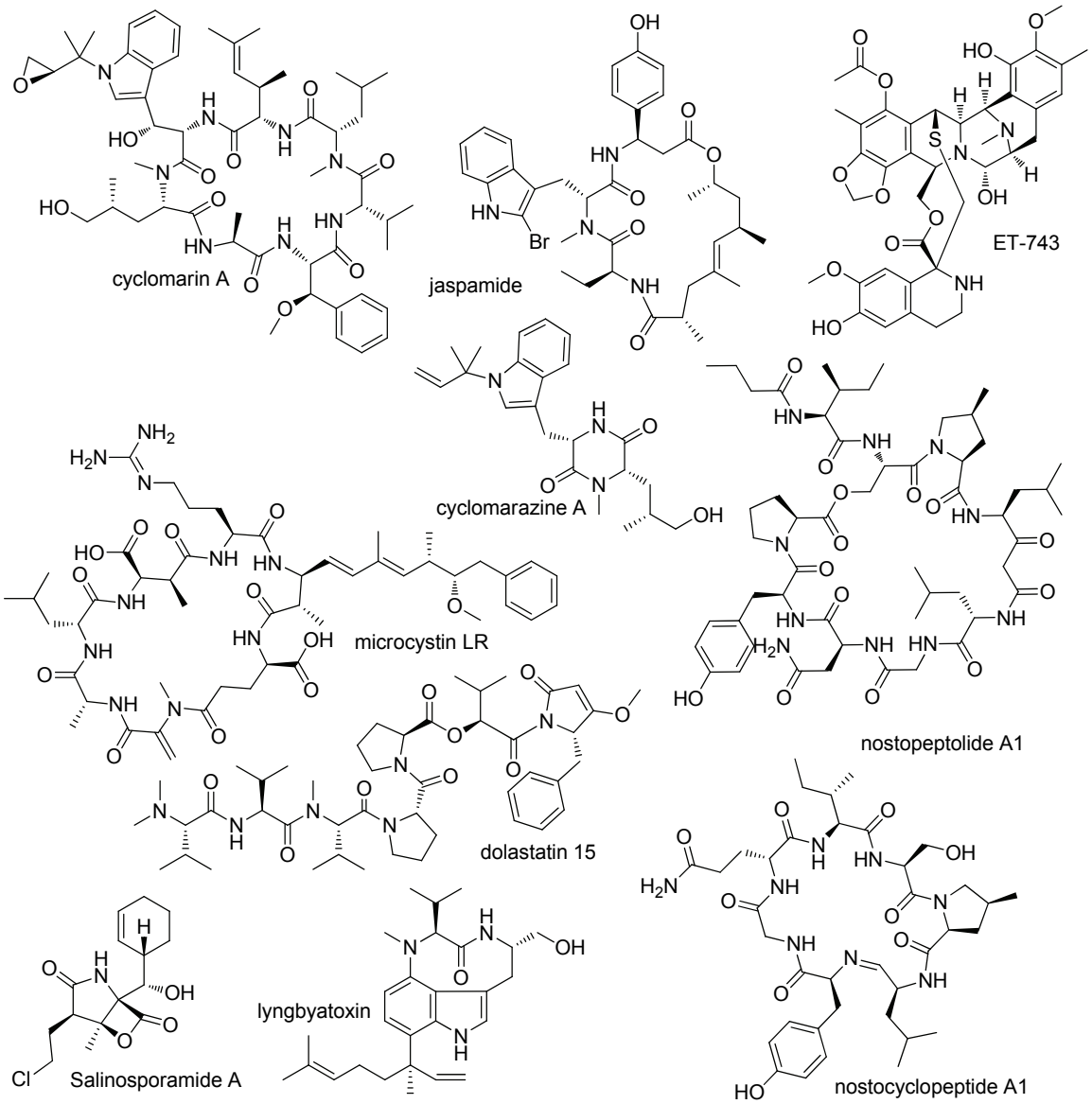


Figure 1.4: Marine non-ribosomal peptides

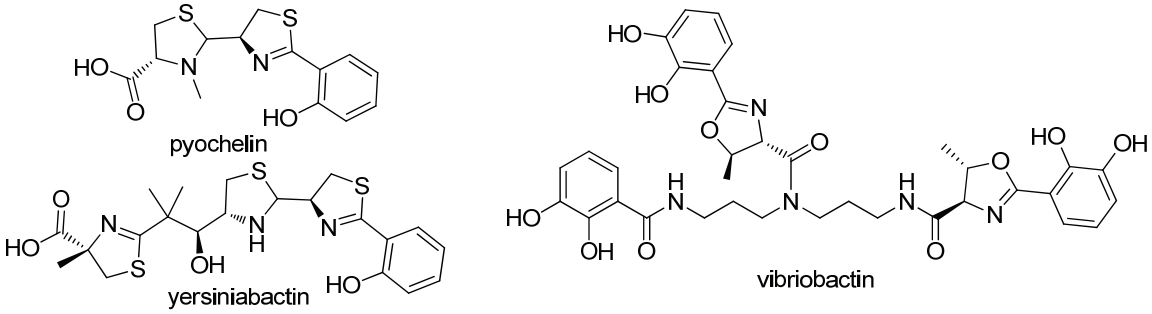


Figure 1.5: Non-ribosomal siderophores

1.2: NRPs are assembled via non-ribosomal peptide synthetases

As their moniker suggests, non-ribosomal peptides (NRPs) are assembled utilizing a mechanism that does not rely on the mRNA, tRNA, or the ribosome for the selection of amino acids and synthesis of the peptide bonds. Instead, amino acids are selected for incorporation into NRPs and peptide bonds are formed utilizing a large, modular enzyme complex, the non-ribosomal peptide synthetase (NRPS). These megasynthetases are ribosomally synthesized, and genes coding for them are often amongst the largest open reading frames (ORF) in the host organism.[5,54,55]

Figure 1.6 depicts non-ribosomal peptide synthesis of cyclo-L-Trp-L-*N*-methylleucine via a hypothetical NRPS. Each module is responsible for selecting and incorporating a single amino acid into the growing NRP, and furthermore, each module comprises several domains that are responsible for different functions in the process of selecting, incorporating, tailoring, and cleaving the NRP from the NRPS. The adenylation domain (A, Figure 1.6 panel A, highlighted in red) selects for and forms the adenylate of a specific amino acid, activating it for Claisen condensation with the phosphopantetheinyl (PPant) arm

of the thiolation domain (T, also referred to as a peptidyl carrier protein). The PPant group is attached to the NRPS via a phosphopantetheinyl transferase utilizing coenzyme A (CoA) as the donor for this prosthetic group.[39] The PPant arm provides flexibility, allowing the growing peptide to interact with other domains. The condensation domain (C, Figure 1.6 panel B) then catalyzes the condensation of the first amino acid with the second, cleaving the thioester bond between the first amino acid (Trp) and the T domain on module 1. Modifications to NRPs can occur while they are tethered to the T domain, and domains that can catalyze these modifications are often observed in NRPSs. In this example (Fig 1.6, panel C), a methyltransferase domain (MT) catalyzes the *S*-adenosyl methionine (SAM) dependent *N*-methylation of the leucine residue to provide the methylated product (Fig 1.6, panel D). To release the peptide product, a thioesterase domain (TE) domain catalyzes cyclization and thioester bond cleavage, providing the product peptide (cyclo-L-Trp-L-*N*-methylleucine) with the NRPS ready for another round of peptide synthesis.

So why would nature rely on such a metabolically expensive process to produce NRPs when the cell already has an efficient process to assemble peptides via the ribosome? The NRPS provides an assembly mechanism that allows for greater diversity in the amino acids that can be incorporated into the peptide and a greater ability to interact with other enzymes that may modify the growing peptide. Tailoring domains, such as MT domains, can modify the structure of amino acids after incorporation, while other tailoring enzymes can function in trans to modify tethered NRPs or act post-NPRS assembly. Individual

TE domains form a specific product, but as an enzyme class, are capable of forming both linear and cyclic peptides where the cyclization can form between any free amide or hydroxyl and the C-terminus or the NRP. Adenylation domains are not restricted to the 20 proteinogenic amino acids and can be specific for any amino acid that the organism has the machinery to produce.

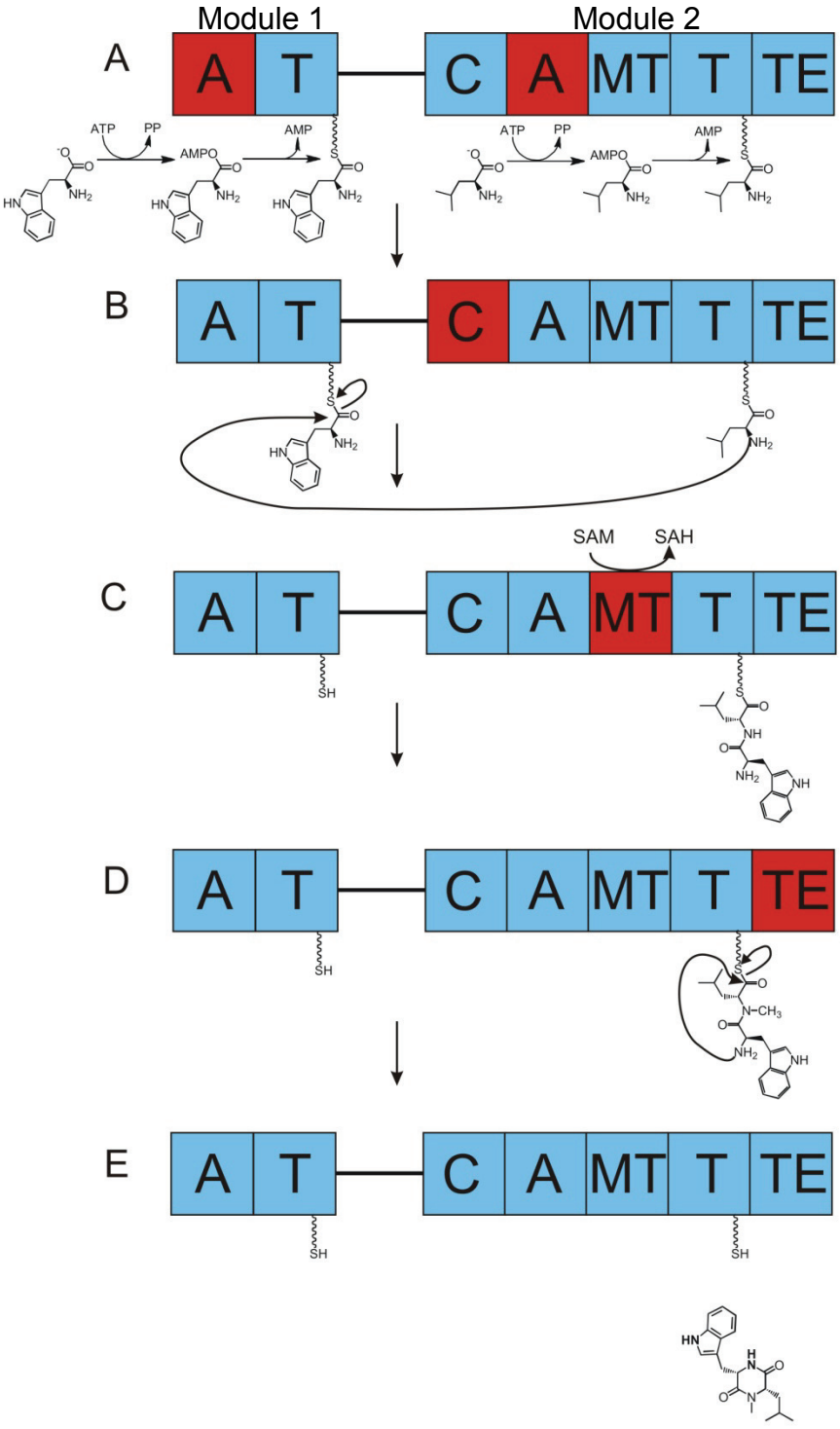


Figure 1.6: Assembly mechanism of non-ribosomal peptides

1.3: Tailoring

One way by which nature is able to increase the chemical diversity of NRPs is to perform tailoring reactions to the amino acid residues of the peptide.[56,57] These reactions can occur to the amino acids before being incorporated into the growing peptide, while tethered to the NRPS, or post-assembly. Some common modifications include epimerization, oxidation, halogenation, methylation, prenylation, and glycosylation, and cyclization of serine, threonine, or cysteine to form oxazoline or thiazoline heterocycles.

1.3.1: Epimerization

Proteinogenic amino acids can be incorporated into an NRP in their native L-configuration or can be converted to the corresponding D-configuration utilizing an epimerase or a racemase. This conversion can be catalyzed by an epimerase domain contained within the NRPS, such as in daptomycin[58] and A54145[48] biosynthesis or via a free enzyme, as in the case of cyclosporin in which D-alanine is created from L-alanine via an alanine racemase,[59,60] before being incorporated into the growing NRP.

1.3.2: Oxidation

Oxidation is a reoccurring theme in NRP biosynthesis, the most common of which is the installation of a hydroxyl group at the beta position of the amino acid residue. By 2001 at least 10 out of the 20 proteinogenic amino acids have been observed as a β-OH derivative in a natural product, two more of which are naturally β-hydroxy, serine and threonine, and glycine lacks a β-carbon to be oxidized.[61] One prominent example is in the biosynthesis of the glycopeptide

antibiotic balhimycin (Figure 1.2), where a stand-alone NRPS, the AT didomain BpsD, accepts tyrosine, which is in turn β-oxidized by the cytochrome P450 (CYP) OxyD, followed by release by the thioesterase Bhp, yielding β-hydroxytyrosine. This oxidized tyrosine is then selected for by the appropriate A domain of the balhimycin NRPS for incorporation into the growing peptide.[62] Tyrosine is also modified in a similar fashion by the CYP NovI after forming the tyrosyl-*S*-NovH aminoacyl enzyme, before being incorporated into novobiocin.[63] Histidine is converted to β-hydroxyhistidine[64] via the CYP NikP2 as an aminoacyl NRPS intermediate (Figure 1.7). The freed, oxidized amino acid is further modified to form the imidazolone base of nikkomycin X. CYP catalyzed β-oxidization of a tethered intermediate has also been implicated in the biosynthesis of salinosporamide (Figure 1.16 panel C).[65,66]

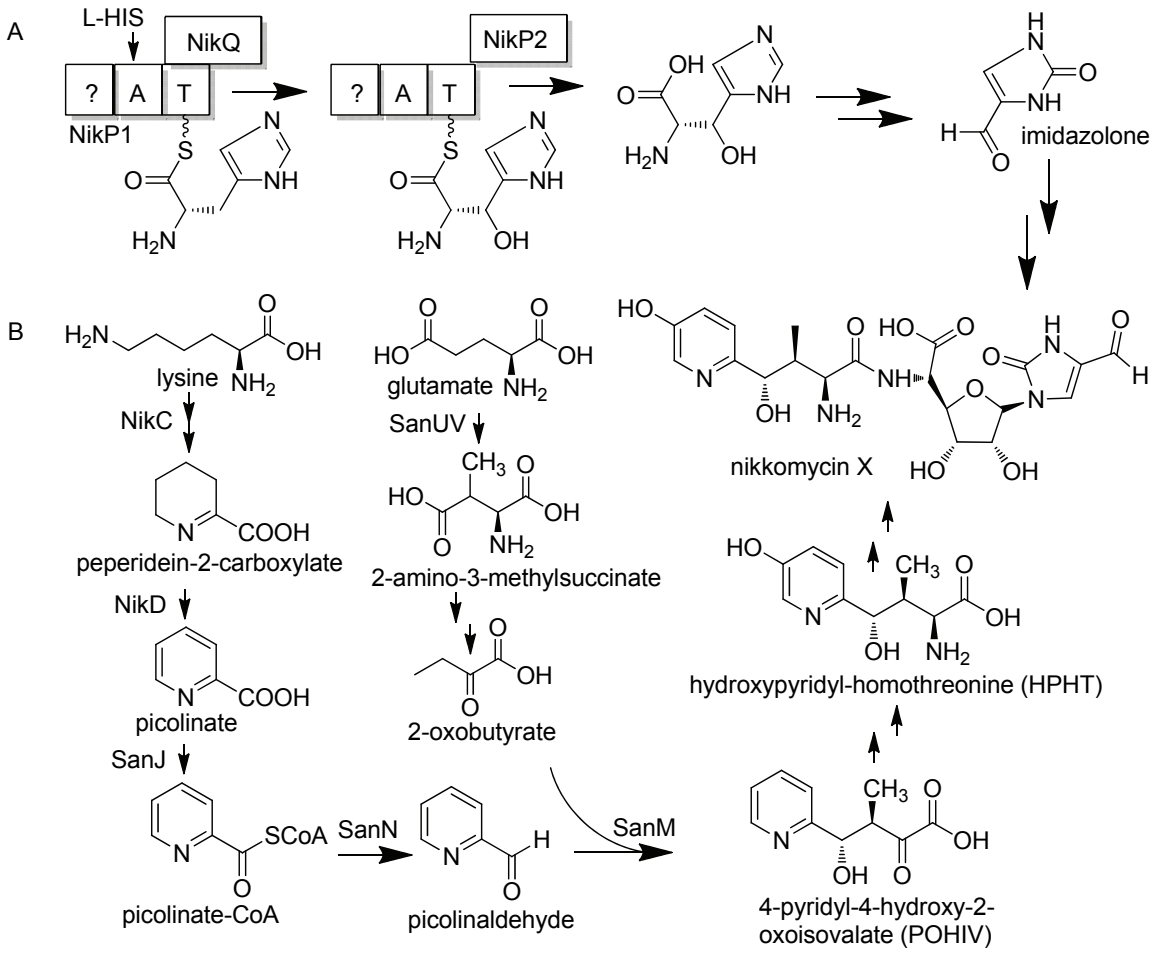


Figure 1.7: Abbreviated pathway to nikkomycin X. Panel A depicts β-oxidation of histidine while tethered to the NRPS NikP1 via the cytochrome P450 NikP2. The pathway to the NPAA HPHT utilizes an aldolase/dehydrogenase pair SanNM followed by transamination and oxidation to form HPHT.

β-oxidation can also be catalyzed by by Fe^{2+}/α-ketoglutarate-dependent oxygenase, one case of which is observed in CDA biosynthesis. Inactivation of the Fe^{2+}/α-ketoglutarate-dependent oxygenase SCO3236 in *S. colecolor* yielded the novel CDA analogs lacking a β-OH functionality on asparagine, establishing its function as a asparagine oxidase.[67] Additionally, the Fe^{2+}/α-ketoglutarate-dependent oxygenases KtzO and KtzP have been implicated in creating the

threo and *erythro* diastereomers of *β*-OH-glutamate residue of the cyclic depsipeptides kutznerides.[68,69]

Recently, a novel monoxygenase, CmlA[70] has been discovered from the pathway to chloramphenicol[26] that installs the β-hydroxy group on the NRPS tethered NPAA L-*p*-aminophenylalanine. This enzyme utilizes an oxo-bridged dinuclear iron cluster to provide L-*p*-amino-β-hydroxyphenylalanine utilizing O_2 as the oxygen source. BLAST analysis of bacterial genomes revealed at least 40 genes with significant identity to CmlA,[26,70] four of which are located in gene clusters that encode known β-hydroxylated NRPs.

Oxidation is not exclusive to the β position or to tethered peptidyl intermediates. The structurally related compounds safracin and saframycin both contain a 3-hydroxy-5-methyl-*O*-methyltyrosine residue. In vivo gene inactivation of Sac D in the safracin B producer *Pseudomonas fluorescens* suggested its role in the oxidation of this tailored tyrosine precursor (Figure 1.8).[50] The SacD homolog from the *Streptomyces lavendulae* pathway to saframycin[71] was also characterized, confirming the role of this enzyme.

Another important role of oxidation in tailoring NRPs is coupling. LtxB, a CYP, (Figure 1.9) facilitates the formation of the nitrogen-carbon bond between the *N*-methyl valine residue and the indole during the biosynthesis of lyngbyatoxin.[72,73] Oxidative coupling is key to forming the rigid peptide aglycone of the glycopeptide antibiotics.[47,57]

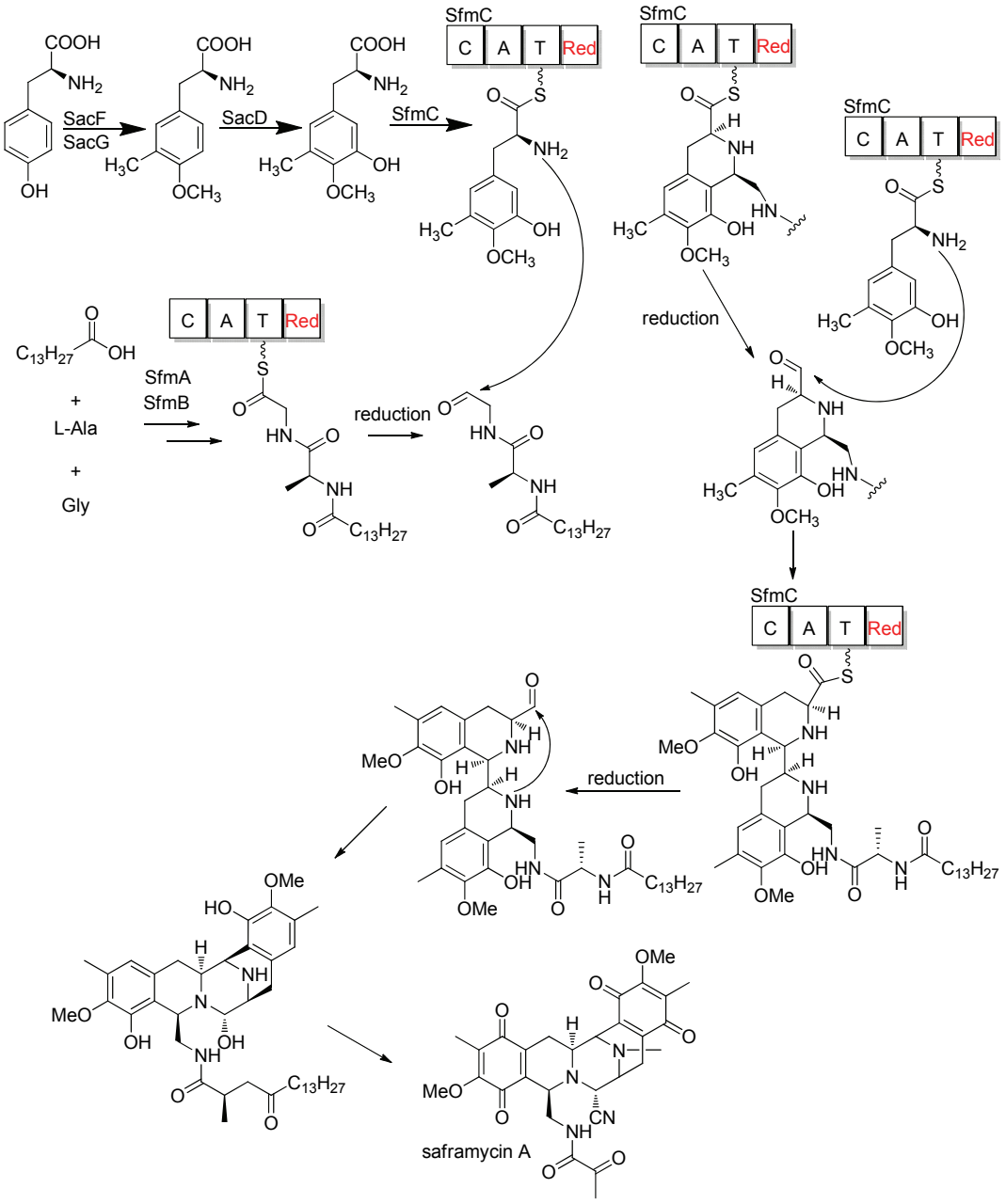


Figure 1.8: Pathway to saframycin A requires two tyrosine residues that have been *O*- and *C*-methylated and oxidized, condensed with alanine and glycine via the NRPS SacA, SacB and SacC. Numerous reductive offloading concerted with biological Pictet-Spengler reactions facilitate the formation of the highly cyclized saframycin A product. Both saframycin and safacin are produced by a similar pathway and the safacin enzymes are depicted for the production of the highly oxidized and methylated tyrosine residue, while the saframycin enzymes are depicted for the remainder of the pathway.

1.3.3: Halogenation

Halogenation of secondary metabolites is catalyzed by one of four classes of enzymes, the $FADH_2$-dependent, non-heme Fe^{2+}/α-ketoglutarate dependent, SAM-dependent, or the haloperoxidases.[74] $FADH_2$-dependent halogenases have been implicated in the chlorination of a tryptophan residue at the C2 position in the NRPS-PKS hybrids chondramide B and D,[53] the 6,7-dichloro-tryptophan residue of kutzneride,[75] and on the β-hydroxytyrosine residue of the glycopeptide antibiotic balhimycin,[62,76] either while tethered to the NRPS, or after peptide release. The non-heme Fe^{2+}/α-ketoglutarate dependent chlorinase installs a chlorine atom on the γ methyl group of the L-threonine residue of syringomycin.[51] Of the final two types of halogenases, only the SAM dependent class has been shown to modify amino acids or non-ribosomal peptide products. 5'-FDAS catalyzes the S_N2 type displacement of methionine, providing the 5'-fluoro-5'-deoxyadenosine precursor to 4-fluorothreonine,[77] although it is uncertain if this product is a component of any NRP produced by the host of 5'-FDAS, *Streptomyces cattleya*. Also of note is SalL,[78] the enzyme responsible for producing the 5'-chloro-5'-deoxyadenosine precursor to the chloroethylmalonyl-CoA PKS extender utilized in the biosynthesis of the PKS-NRPS hybrid salinisporamide A.

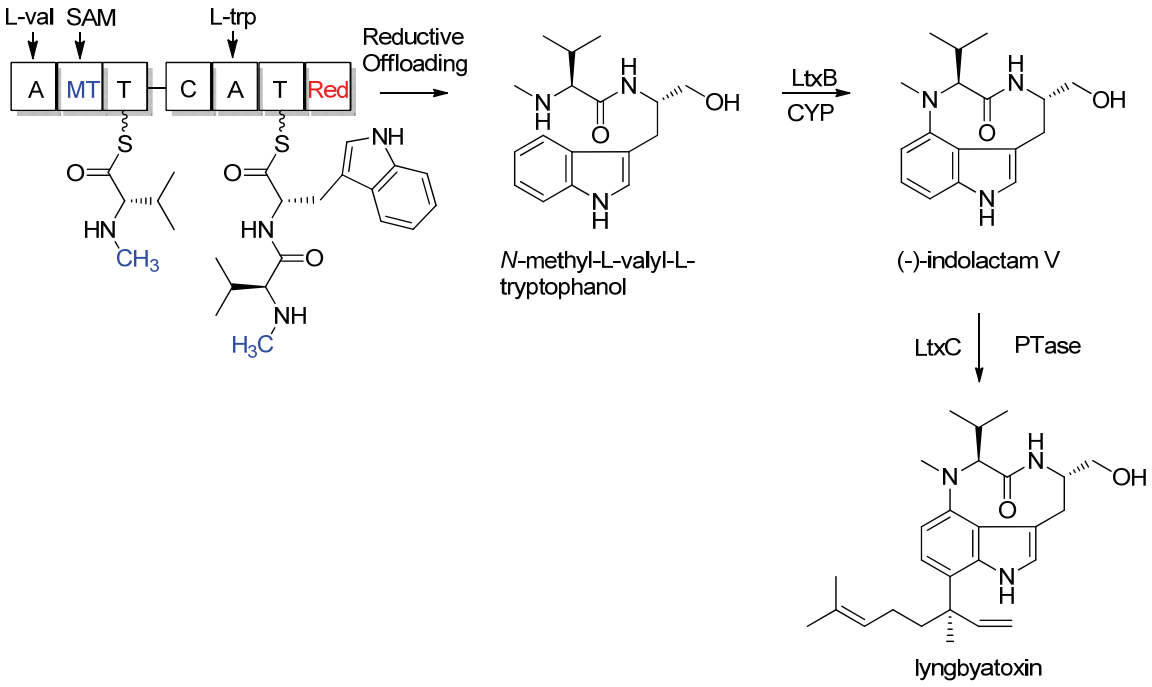


Figure 1.9: Abbreviated pathway to lyngbyatoxin

1.3.4: Methylation

Numerous examples of tailoring via methylation occur in the assembly of NRPs, the majority of which are *N*-methylations, but there are examples at *C* and *O* positions.[57] Generally, these methylations are catalyzed by SAM dependent methyltransferases, which can be found as free enzymes that methylate intermediates tethered to the NRPS, but are most often seen as cis acting modules embedded in the megasynthase[57]. One example of an embedded *N*-methyltransferase is seen in lyngbyatoxin biosynthesis (Figure 1.9), where a SAM dependent methyltransferase domain methylates the free amine of the NRPS tethered valinyl thioester.[72] The *N*-methylation of the amide nitrogen is a common theme, although it can vary from no occurrences as in the structure of daptomycin[58] to seven out of eleven residues being *N*-methylated as seen in

cyclosporine.[59] An additional example of *N*-methylation of a peptide amide, this time by a free enzyme on a post-NRPS product, is SacI (Figure 1.8) in the biosynthesis of safracin[50] and the structurally related saframycin. *C*-methyltransferase and *O*-methyltransferase activity is suggested for SacF and SacG, respectively, in the biosynthesis of safracin. SafC, involved with the biosynthesis of the same 3-hydroxy-5-methyl-*O*-methyltyrosine residue as SacF and SacG in saframycin biosynthesis was characterized biochemically, confirming its role as a *O*-methyltransferase.[79] Surprisingly, SafC and SacG do not share significant identity,[71] providing an interesting example of how two pathways towards two closely related secondary metabolites can utilize different enzymes to form an identical residue.

1.3.5: Prenylation

Another tailoring reaction seen in NRPs is prenylation, albeit rarely. Prenyltransferases catalyze an aromatic substitution reaction between an aromatic acceptor and a prenyl donor. Prenylation of an NRP (Figure 1.9) is exemplified in the biosynthesis of lyngbyatoxin. LtxA, the NRPS, catalyzes the formation of *N*-methyl-L-valyl-L-tryptophanol via condensation of Trp and Leu, *N*-methylation of the valine residue, followed by reductive offloading.[72,80] The CYP LtxB[73] catalyzes the cyclization to provide indolactam V, most likely via an epoxide intermediate.[72] Prenylation is then catalyzed via LtxC, utilizing indolactam V as the aromatic acceptor and geranyl pyrophosphate as the prenyl donor.[72]

1.3.6: Glycosolation

Tailoring by glycosylation also occurs in the biosynthesis of non-ribosomal peptides, the most prominent of which is the glycopeptides. Vancomycin, teicoplanin, and chloroeremomycin all contain two or more glycosylations that are required for their antibiotic activities against Gram-postitive pathogens.[81] Several of these glycosyltransferases have been cloned from strains of *Amycolatopsis orientalis* that produce vancomycin or chloroeremomycin and were shown to glycosylate the vancomycin or teicoplanin aglycone or the glucosylated vancomycin or teicoplanin aglycone, depending on their native substrate specificity.[81] It was also shown that novel hybrids of teicoplanin/vancomycin glycopeptides[81] or a glucosyl A47934 derivative[82] could be produced by mixing and matching these glycosyltransferases in vitro or in vivo.

1.3.7: Cyclization of serine, threonine, or cysteine to form oxazoline or thiazoline

Heterocycle formation is prominent in the formation of the siderophores pyochelin,[83,84] yersiniabactin[33,85,86] and vibriobactin,[87] and can be observed in other classes of compounds, such as epothilone[52,88] and bleomycin.[89-91] These heterocycles are derived from serine, threonine, or cysteine in all of these cases, where NRPSs are utilized to create the heterocycle and form all or part of the final product.[92] Cysteine cyclization to form the thiazoline ring of epothilone is catalyzed by the cyclase domain (Cy) of EpoB, which condenses the acetyl

group tethered to EpoA with cysteine to yield the acetylcystenyl intermediate similar to any other NRPS condensation domain.[88] The Cy domain then catalyzes cyclization and dehydration to from a thiazolinyl intermediate. Further tailoring is carried out via the flavin mononucleotide dependent oxidation domain (Ox), which converts the thiazoline ring to a thiazole.[89] Conversely, reduction could occur to provide the thiazolidine ring as observed in yersiniabactin. A Cy domain is present in the vibriobactin[87] NRPS biosynthesis, and is responsible for forming the oxazoline rings from threonine.

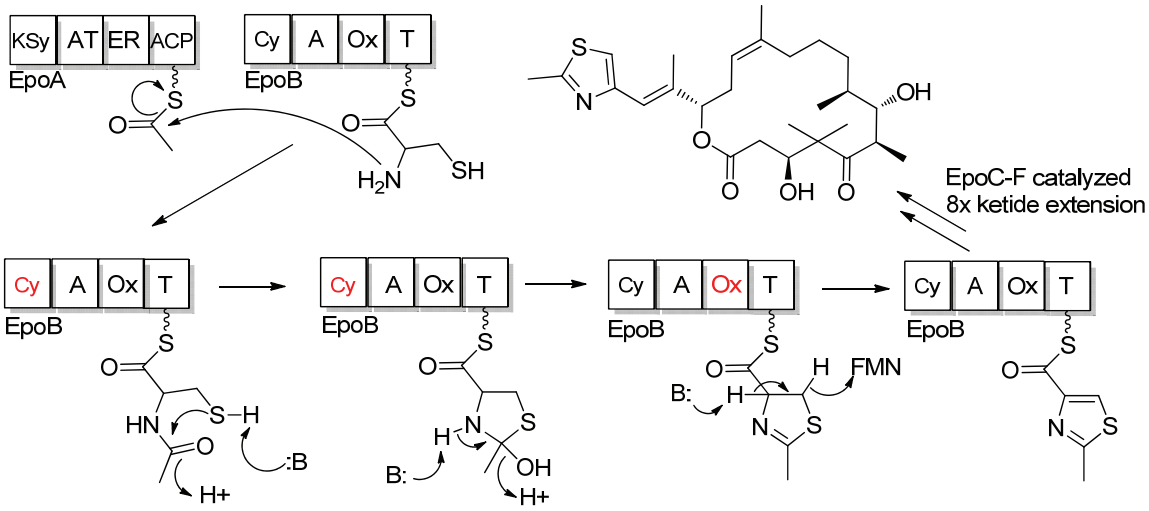


Figure 1.10: Summarized scheme towards the biosynthesis of epothilone, highlighting the formation of the thiazole ring

1.3.8: Reductive offloading

The majority of NRPs are released from the synthetase via the action of a thioesterase as a linear or cyclic product. Some biosynthetic pathways do not contain a terminal thioesterase, and in its place is a domain with significant structural similarity to short chain dehydrogenases/reductases.[80] The majority of the examples of these reductive offloading domains (Red) come from

cyanobacterial pathways, such as those to lyngbyatoxin (Figure 1.9)[72,80] and nostocyclopetide,[93] although the pseudomonad NRPS utilized for safracin production and the actinobacterial saframycin NRPS (Figure 1.8) contains a reductive domain.[50,94]

Both two and four electron reductions can be catalyzed by Red domains to provide a linear peptidyl alcohol or a reactive aldehyde that can undergo intermolecular condensation. During lyngbyatoxin biosynthesis, a four electron reduction occurs, reducing the carboxy terminus of *N*-methyl-L-valyl-L-tryptophan to the corresponding alcohol.[80] A two electron Red catalyzed reduction to the aldehyde followed by nucleophilic attach by the *N*-terminal amine to form the cyclic imine has been implicated in the biosynthesis of the nostocyclopeptides.[93] It has been proposed that the SafC NRPS is utilized iteratively to add both modified tyrosine residues. The Saf C red domain catalyzes a two-electron reduction, releasing peptidic safracin precursors as aldehydes that participate in Pictet-Spengler-like reactions to form the heterocyclic core of safracin B.[50]

1.4: Biosynthetic pathways to non-proteinogenic amino acids

A second way in which NRPS structures can be diversified by nature is by incorporation of non-proteinogenic amino acids (NPAAs), which are specialized amino acids that go beyond the standard 20 amino acids utilized in both ribosomal and NRP synthesis. Often, the genes required for the biosynthesis of NPAAs are clustered with the NRPS. Adenylation domains (A-domains), although specific for one amino acid or closely related amino acids, are not

restricted to the standard 20 amino acids, providing a mechanism by which NPAAs are incorporated.

Pathways to non-proteinogenic amino acids are diverse. Some of these pathways utilize construction mechanisms similar to those for proteinogenic amino acids, occasionally borrowing enzymes utilized by the host organism to provide the standard 20 amino acids. On the other hand, some NPAAs are produced by pathways quite divergent from those utilized to produce the proteinogenic amino acids.

Classifying the amino acids incorporated by NRPS as tailored proteinogenic amino acids or as NPAAs can be tricky, especially when a proteinogenic amino acid is extensively tailored by multiple enzymes. One example of this is the pathway to 4-methylproline (Figure 1.11), a constituent of the nostopeptolide[17] and the nostocyclopeptides.[16] The pathway to nostopeptolide[95] and nostocyclopeptides[93] contains two enzymes, NosE and NosF, which have been characterized biochemically to confirm their role in the formation of 4-methylproline.[96] Similarly, the peptidic precursor of penicillin contains the non-proteinogenic L-α-aminoadipic acid residue that is produced from lysine and involves a piperidine intermediate (Figure 1.17).[5]

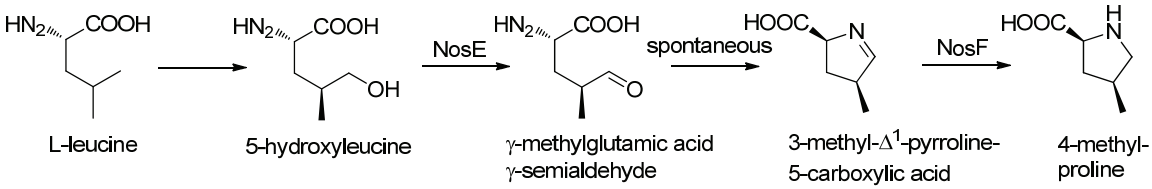


Figure 1.11: Pathway to 4-methylproline. The enzyme required to form 5-hydroxyleucine has not been characterized.

1.4.1: Direct intermediates of primary metabolic pathways

One common source of non-proteinogeinc amino acids is the direct intermediates of primary metabolic pathways, such as kynurinine and ornithine, both of which are residues in the structure of daptomycin. Kynurinine is an intermediate of tryptophan degradation, and *Streptomyces roseosporus*, the producer of daptomycin, carries two copies of the tryptophan 2,3-dioxygenase (TDO), which is the first, committed step of tryptophan degradation.[58] One copy of TDO is clustered with the pathway to daptomycin, while the other copy resides elsewhere in the genome, and surprisingly is only 29% identical. Additionally, ornithine, a NPAA intermediate of arginine biosynthesis is a residue of daptomycin.

1.4.2: Transamination of α-keto acids

One example of the production of NPAAs comes from the non-ribosomal lipopeptide CDA in *Streptomyces coelicolor*. By inspection of the structure of CDA, one would expect 3-methyl-glutamic acid to come from the direct methylation of glutamic acid or an NRPS tethered intermediate of CDA production. Surprisingly, the methyltransferase GlmT from the CDA pathway

acts on α-ketoglutarate (Figure 1.12) to form 3-methyl-2-oxoglutarate, which in turn is transaminated via IlvE, the *S. coelicolor* transaminase borrowed from the biosynthesis of branched chain proteinogenic amino acids, to yield 3-methyl-glutamic acid.[97] The adenylation domain responsible for the incorporation of 3-methyl-glutamic acid does show some substrate promiscuity, allowing for the incorporation of methyl and desmethyl glutamic acid, yielding the corresponding CDA analogs.[98]

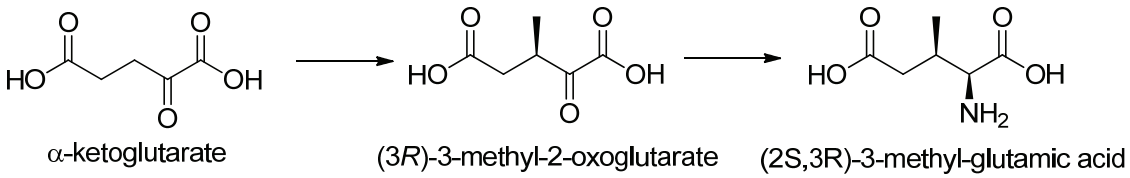


Figure 1.12: Pathway to 3-methyl-glutamic acid

1.4.3: Polyketide synthase derived

PKS derived amino acids are also incorporated into NRPs, one of the most prominent examples of which is the 4-(2-butenyl)-4,*N*-dimethyl-L-threonine component of cyclosporine. Although the exact details of the pathway have not been worked out,[5,60] experimental evidence suggests that this amino acid is derived from acetyl-CoA and malonyl-CoA, including a SAM-derived methylation and an oxidation to incorporate the alpha-keto group (Figure 1.13). After release from the PKS as a CoA thioester, it is transaminated to the amino acid at the alpha position before or after being converted to the free carboxylic acid.

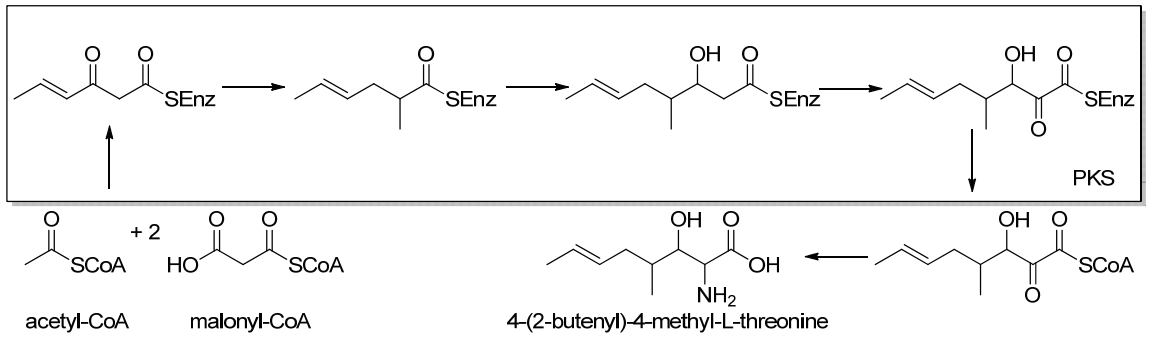


Figure 1.13: Suggested pathway to the 4-(butenyl)-4 -methyl-L-threonine component of cyclosporine. Boxed reactions are catalyzed by the PKS or by associated enzymes reacting in trans on PKS tethered intermediates.

Another example of a PKS derived NPAA is the 3,5-dihydroxyphenylglycine residue of vancomycin, which is formed by the condensation of 4 malonyl-CoA units via a type III PKS,[99-101] followed by oxidation[102] and transamination to the final product.

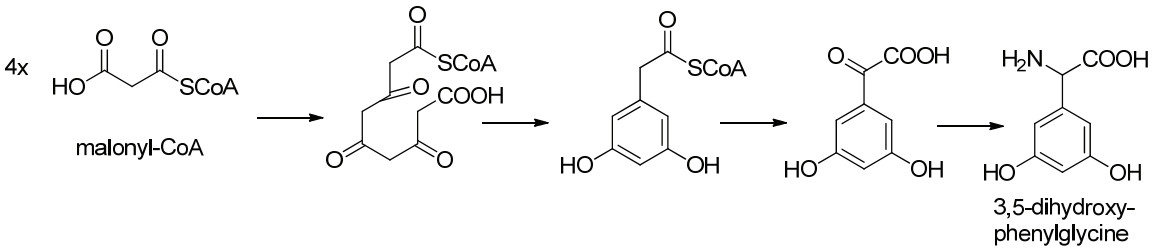


Figure 1.14: Pathway to 3,5-dihydroxyphenylglycine in vancomycin biosynthesis

1.4.5: Shikimate pathway

Often intermediates of aromatic amino acid biosynthesis, which are derived from the shikimate pathway, are utilized as the building blocks of cyclic or aromatic NPAAs. This is evident in the biosynthesis of 4-hydroxyphenylglycine (Figure 1.15), an NPAA component found in both vancomycin[102] and CDA,[98] which is created from prephenic acid via 4-hydroxyphenylpyruvate. There are obvious similarities in how 4-hydroxyphenylglycine and 3,5-dihydroxy-

phenylglycine are created, both requiring oxidation[102] to create an alpha-keto group which can be transaminated to give the amino acid, although the two pathways start from different precursors.

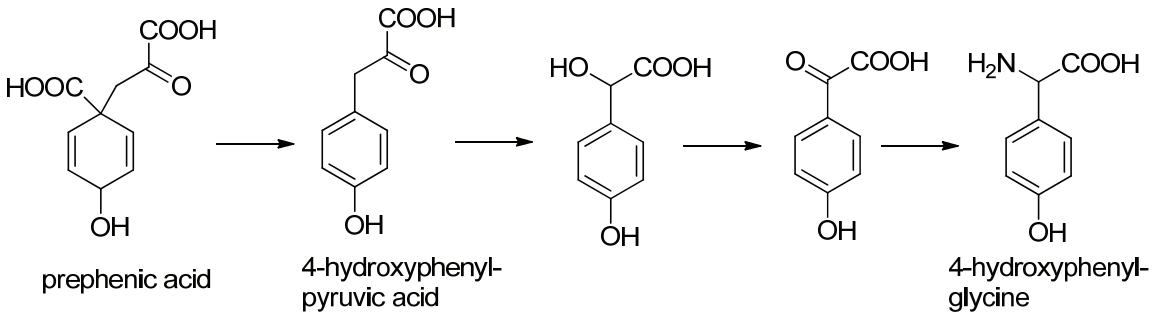


Figure 1.15 Pathway to 4-hydroxyphenylglycine

Another prominent example of a shikimate derived non-proteinogenic amino acid is L-3-cyclohex-2′-enylalanine (CHA), a component of the PK-NRP hybrid proteasome inhibitor salinosporamide.[65,103,104] Stable isotope incorporation studies coupled with bioinformatic predictions and in vitro enzymology support a pathway by which prephenic acid is decarboxylated to give the endocyclic diene dihydro-4-hydroxyphenylpyruvate, which is non-enzymatically isomerized to form the exocyclic diene, followed by reduction and transamination to yield 3-cyclohex-2-enyl-L-alanine (Figure 1.16).

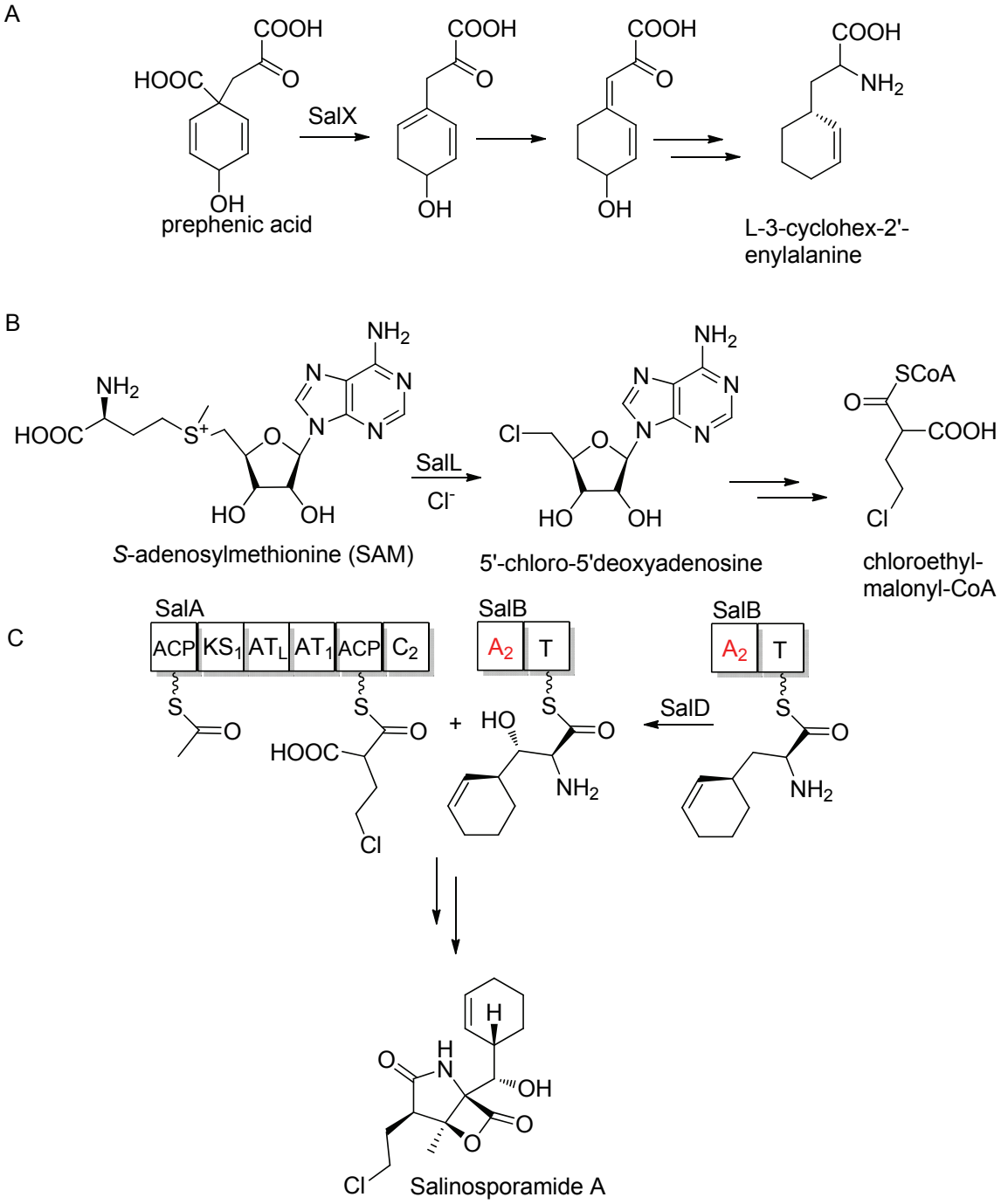


Figure 1.16 Pathway to salinosporamide A, highlighting the biosynthesis of the NPAA precursor cyclohexenylalanine and the unusual PK extender chloroethylmalonyl-CoA.

1.4.6: Aldolase/dehydrogenase heterodimers

NPAA hydroxypyridyl-homothreonine (HPHT), a component of nikkomycin X, is constructed from picolinate-CoA and 2-oxobutyrate through the action of SanN and SanM.[44,105] These two enzymes comprise an aldolase/dehydrogenase heterodimer that catalyze an aldol condensation yielding 4-pyridyl-4-hydroxy-2-oxoisovalate (Figure 1.7). Further oxidation and transamination provides HPHT. Various bacteria, including certain pseudomonads[106] and strains of *E. coli*,[107] use a aldolase/dehydrogenase heterodimer pair to degrade phenolic compounds, while *Burkholderia cepacia* J2315 utilizes a similar enzymatic pair to degrade tryptophan. Thus, the pathway to HPHT provides an interesting example of where nature has evolved a degradative pathway to synthesize NPAAs.

1.5: Interfacing non-ribosomal peptide synthetases with polyketide synthases adds further chemical diversity

The final way to be discussed in which nature increases the structural diversity of NRPs is to interface with polyketide synthases (PKS) to produce hybrid NRP/PK molecules.[108,109] Two of the numerous examples of this are salinosporamide and epothilone. During the biosynthesis of salinosporamide A (Figure 1.16), acetate is extended via the activity of the PKS SalA with chloroethylmalonyl-CoA to form the PK portion of this molecule.[65,103,110] Sal B accepts L-3-cyclohex-2′-enylalanine,[66] which is β-oxidized via SalD and then condensed with the PKS portion of this molecule. Further tailoring steps establish the highly functionalized γ-lactam-β-lactone core.

Epothilone (Figure 1.10) represents an extreme example in which an NRPS is utilized to provide a methylthiazolylcarboxy starter unit to a PKS system. Cysteine and acetate (via malonyl-CoA) are condensed by the PKS EpoA and the NRPS EpoB. The thiazole functionality of the methylthiazolylcarboxy starter is provided by cyclization and oxidation via EpoB, the eight rounds of polyketide extension via EpoC-F followed by additional tailoring steps provides epothilone.[88]

1.6: Engineering of nonribosomal peptides to increase nature's peptide diversity

Although nature has developed numerous mechanisms to provide structural diversity in NRPs via tailoring reactions, use of non-proteinogenic amino acids, and by interfacing with other compound classes, humanity has developed clever means to further diversify NRPs. The need to engineer structural diversity beyond what is provided by nature has been driven by society's pharmaceutical needs. Indeed, nature may provide a compound that is highly bioactive in a petri dish or test tube, but certain features of a compound may make it a poor drug in vivo. Solubility, bioavailability, stability, and metabolism[42,111] are all features of a compound that may need to be altered to transform a bioactive compound into a useful drug. A common approach to diversify compounds is to produce analogs via total synthesis or semi-synthesis. This can be extremely difficult and time consuming when it comes to NRPs because of their size and the high level of functionalization. One solution is to manipulate the biosynthesis of the NRP to provide the desired analogs. Precursor directed biosynthesis, mutasynthesis, and combinatorial biosynthesis

are all biosynthetic techniques that have been utilized to manipulate the structures of NRPs.[42,112] In addition to increasing the “drugability” of a compound, these approaches can teach us about how these compounds are made. This knowledge can then be utilized to manipulate the biosynthesis of other compounds or to direct further rounds of structural optimization.

1.6.1: Precursor directed biosynthesis and mutasynthesis

Precursor directed biosynthesis is performed by supplementing the culture media of a wild-type organism with a substrate that is similar to a natural precursor of a compound you wish to engineer. Depending on the specificities of the enzymatic machinery, this substrate may be incorporated into the structure of the final product, generating a novel analog.[42] In addition to generating novel analogs, this approach can be used to probe the mechanism by which the compound is produce by providing predicted precursors that have been isotopically labeled with stable isotopes. Isolation of the compound followed by spectrometric analysis to observe the pattern of incorporation can provide significant clues to the biosynthetic process involved in making the wild type compound.[111]

Precursor directed biosynthesis has been utilized to diversify the structure of penicillin (Figure 1.17). The first penicillin utilized clinically was penicillin G, which is produced by the fungus *Penicillium chrysogenum* when it is fermented on corn-steep liquor, a substrate high in phenylethylamine, which is converted into phenylacetic acid by *P. chrysogenum*. When the fermentation broth is supplemented with phenoxyacetic acid, penicillin V is produced, which is more

stable via oral administration, in contrast to penicillin G, which must be administered via injection or with an antacid because it degrades in the acidic environment of the stomach.[5]

The inherent disadvantage of precursor directed biosynthesis is the production of the wild type compound and the novel analog as a mixture. An additional limitation of this approach is that the unnatural substrate must compete for incorporation into the final product. One way to circumvent these disadvantages is to generate a mutant unable to produce the precursor of the wild-type compound, and then supplement this mutant with the novel precursor, making a novel compound via a mutasynthetic approach.[42,111]

There are numerous examples in which mutasynthesis has been utilized to diversify the structures of NRPs and NRP/PKS hybrid molecules. One such example includes diversification of pyochelin, which was accomplished by providing the appropriate aza or aryl acid analogs to a culture of *Pseudomonas aeruginosa* deficient in the biosynthesis of the salicylic acid starter. Analysis of the fermentation broth from the supplemented mutant revealed the production of new pyochelin analogs 5-fluoropyochelin, 6-azapyochelin, and 4-methylpyochelin. All three of these compounds retained their ability to chelate iron, the latter of which was more active than the natural pyochelin.[83]

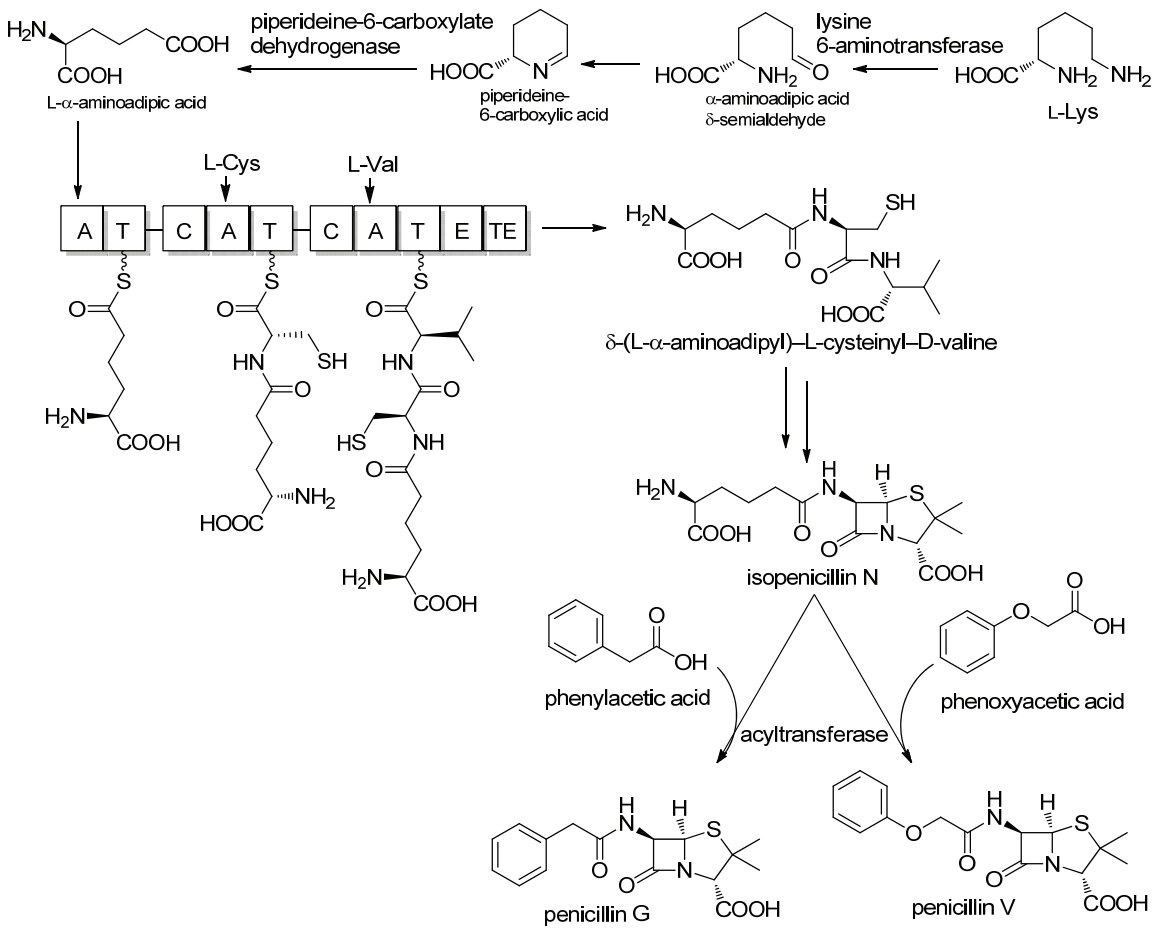


Figure 1.17: Summarized pathway to penicillin highlighting the pathway to the NPAA L-α-aminoadipic acid and the precursor directed biosynthesis of penicillin G and V.

Nikkomycin has also been diversified using a similar approach in both the nucleoside moiety and the NPAA.[111] To diversify the NPAA, a *Streptomyces tendae* Tü901 mutant deficient in lysine-2-aminotransferase (NikC, Figure1.7) was supplemented with benzoic acid, which was accepted by the rest of the pathway to provide 2-amino-4-hydroxy-3-methyl-4-(4'-hydroxyphenyl)butanoic acid, replacing HPHT in the structure of nikkomycin X.[113]

Clorobiocin has also been diversified via a mutasynthetic approach (Figure 1.18) in which the gene encoding the prenyltransferase *cloQ* was

inactivated[114] in *Streptomyces roseochromogenes*, abolishing production of known clorobiocin analogs. A series of 4-hydroxyphenylpyruvate analogs were provided to this mutant, which in turn produced novel clorobiocin analogs with substitutions on ring A where a prenyl group would have been installed in the wild type organism.[115,116]

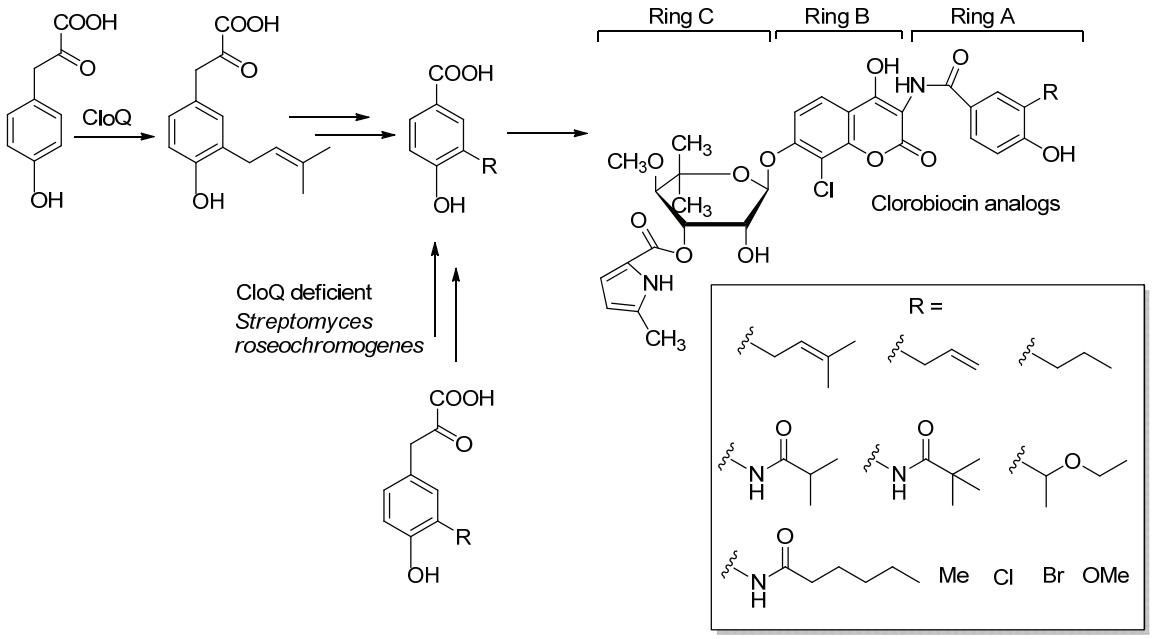


Figure 1.18: Mutasynthesis of clorobiocin

Inactivating 4-hydroxymandelate synthase in the pathway to 4-hydroxyphenylglycine in *Streptomycetes coelicolor* eliminated the production of CDA. Providing intermediates downstream of the 4-hydroxymandelate synthase, such as 4-hydroxymandelate, 4-hydroxyphenylglyoxylate, or 4-hydroxyphenylglycine restored CDA production. Supplementing this mutant with 4-fluorophenylglycine, 4-fluorophenylpyruvate, or 4-fluoromandelic acid provided a novel CDA analog in which 4-hydroxyphenylglycine was replaced with 4-fluorophenylglycine.[49]

A series of analogs of the glycopeptide antibiotic balhimycin were created when the pathways to 3,5-dihydroxyphenylglycine and β-hydroxytyrosine where inactivated in the host producer *Amycolatopsis balhimycina*. When the 3,5-dihydroxyphenylglycine mutant was provided 3,5-dimethoxyphenylgycine, 3-methoxy,5-hydroxyphenylglycine, 3-methoxyphenylglycine, or 3-hydroxyphenylglycine, the corresponding substitutions were observed in the structure of balhimycin. In a similar fashion, 3-fluoro-β-hydroxytyrosine, 3,5-difluoro-β-hydroxytyrosine, and 2-fluoro-β-hydroxytyrosine were accepted, creating fluorinated balhimycin analogs.[76]

The structure of salinosporamide has also been successfully engineered utilizing mutasynthesis[66,117]. Providing a series of cyclic or linear alkyl or alkenyl amino acids to the *S. tropica salX* mutant, deficient in the production of L-3-cyclohex-2′-enylalanine, provided a series of novel salinosporamide analogs. Interestingly, the cyclopentenyl substituted derivative salinosporamide X7 was found to be nearly as potent as the natural product salinosporamide A in the proteasome inhibition assay, and even more potent that salinosporamide A against the human colon cancer cell line HCT-116.

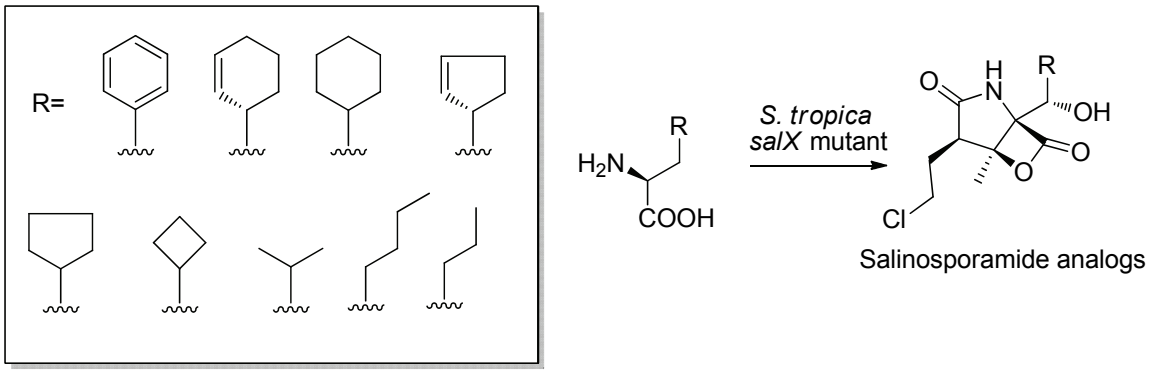


Figure 1.19: Mutasynthesis of salinosporamide

1.6.2: Combinatorial biosynthesis

One of the most ambitious and exhaustive uses of combinatorial biosynthesis applied to NRPs was utilized to diversify the structure of daptomycin (A2197C) and the related lipopeptide A54145 in order to develop a drug effective against community-acquired pneumonia.[118,119] Daptomycin is effective in vitro against the causative agent of community-acquired pneumonia (CAP), S*treptococcus pneumoniae*, but failed to meet noninferiority criteria in clinical trials. This in vivo failure is believed to be caused by sequesterization of daptomycin in lung lipids. The goal of these studies was to create a daptomycin analog that was less susceptible to sequesterization while maintaining a level of efficacy against *S. pneumonia* in vivo.

A domain swapping approach was utilized to create a series of hybrid NRPSs containing domains from the A2197C NRPS DptABCD and the A54145 NRPS LptABCD in either the host producer of A2197C,[118] *Streptomyces roseosporus,* or A54145,[119] *Streptomyces fradiae,* depending on which NRPS was utilized as the host for the swapped domains. In the example depicted in (Figure 1.20), modules 2 and 3 of LptA were swapped with modules 2 and 3 of DptA to create LptA'BCD and then expressed in *S. fradiae*.[119] Fermentation provided the novel compound CB-182,561. At least 24 different lipopeptides were produced by swapping different modules between the two systems,

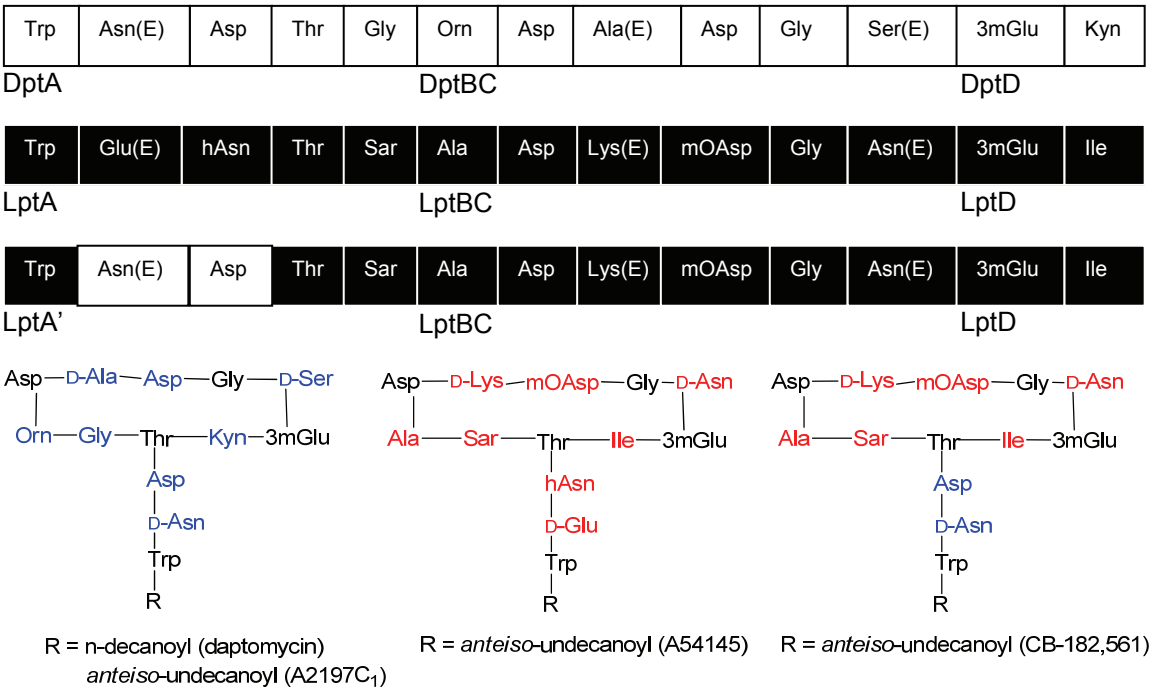


Figure 1.20: Modular architecture of the daptomycin (A2197C) NRPS DptABCD and the A54145 NRPS LptABCD, and the hybrid NRPS LptA'BCD created by swapping out modules 2 and 3 of LptA and replacing them with modules 2 and 3 of DptA. Modules depicted as boxes, amino acid specificities denoted within. Modules containing epimerase domains denoted via an E in brackets. NPAA ammino acid abbreviations used: L-ornithine (Orn), L-*threo*-3-methyl-glutamic acid (3mGlu), and L-kynurenine (Kyn), methoxy-Asp (moAsp), hydroxyasparagine (hAsp), sarcosine (Sar). The structure of the natural product of DptABCD, A2197C and of LptABCD, A54145, are shown below the domain diagram. The product of the combitorial NRPS LptA'BCD, CB-182,561, is also shown. Amino acids in black denote invariable residues between the two natural products, blue represent amino acids native to the A2197C series, and red represents amino acids native to the A54145 series.

expressing them in *S. fradiae* or *S. roseosporus,* and isolating the resultant lipopeptides from the fermentation broth. CB-182,561 was found to be 32 times more effective than daptomycin against *S. pneumonia* in vitro in the presence of bovine pulmonary surfactant, and was found to be efficacious without lethal toxicities in a mouse model of CAP in which daptomycin was ineffective.

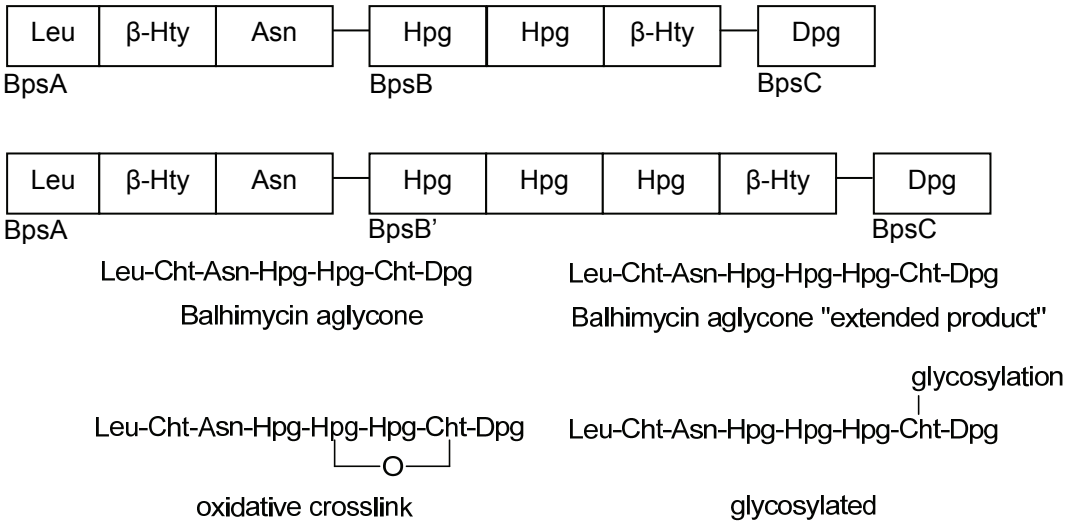


Figure 1.21: Modular architecture of the balhimycin NRPS BpsABC and the modified balhimycin NRPS BpsAB'C. Modules depicted as boxes, amino acid specificities denoted within. NPAA abbreviations used: β-chlorohydroxytyrosine (Cht), 4-hydroxyphenylglycine (Hpg), 3,5-dihydroxyphenylglycine (Dpg). The structure of the natural product of BpsABC, the balhimycin aglycone is depicted below the module diagrams. The engineered product of BpsAB'C, the balhimycin aglycone "extended product" and the detected tailored products are also depicted.

The structure of the balhimycin aglycone[120] has been modified to incorporate an additional 4-hydroxyphenylglycine residue into the structure of this NRP (Figure 1.21). Inserting an additional module in between the other two modules specific for this NPAA provided an aglycone containing three 4-hydroxyphenylglycine residues instead of the typical two. Although most of the oxidative tailoring and gylcosylation that one would have expected to occur to the balhimycin aglycone did not occur, some peptides were observed where the β-chlorohydroxytyrosine residue was glycosylated or the oxidative linkage was established between the side chains of β-chlorohydroxytyrosine and 4-hydroxyphenylglycine.

Creation of a novel surfactin analog[121] was realized by deleting the second domain from the surfactin NRPS in *Bacillus subtilis* ATCC 21332 (Figure 1.22). Deletion of this domain, followed by LC-MS analysis of the resultant mutant revealed the production of Δ2-surfactin, an analog of surfactin lacking the Leu residue. It was reported that Δ2-surfactin was still hemolytic, but not as potent as the natural product surfactin.

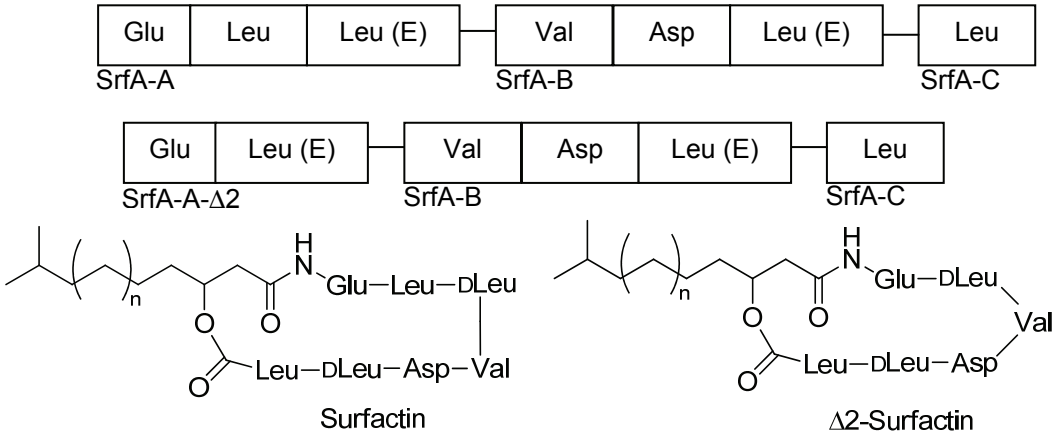


Figure 1.22: Modular architecture of the surfactin NRPS SrfA-ABC and the modified surfactin NRPS SrfA-A-Δ2BC. Modules depicted as boxes, amino acid specificities denoted within. Modules containing epimerase domains denoted via an E in brackets. The structure of the natural product of SrfA-ABC, surfactin and the engineered product of SrfA-A-Δ2, SrfA-B, and SrfA-C, Δ2-surfactin are also denoted.

1.7: Conclusion

The NRPs are a class of secondary metabolites that are very structurally diverse. Because of the non-ribosomal assembly process, they may contain both proteinogenic and non-proteinogenic amino acids, the latter of which are provided by unique biosynthetic pathways. Furthermore, extensive tailoring before, during, or after peptide assembly provides even greater structural diversity. This diversity is paramount to their roles in nature and their use by man

to treat and cure disease. In the drive to find new or better drugs, man has developed techniques to further diversify these compounds by manipulating their biosynthesis.

Although numerous advances have been made in understanding how non-ribosomal peptides are assembled, there is still a lot to learn. Genome projects have provided a wealth of information about the prevalence of NRP biosynthetic pathways in nature, how NRPs are assembled and how NPAAs are biosynthesized from primary metabolites.[54] A lot of work still needs to be done to characterize these mechanisms beyond in silico predictions, and to harness this knowledge to further understand the roles of NRPs in nature and to manipulate these biosynthetic mechanisms to provide new medicines.

The remainder of this dissertation addresses understanding the biosynthesis of cyclomarin and cyclomarazine, the pathway for which was found via sequencing of the genome of the marine actinobacterium *Salinispora arenicola* CNS-205. The focus is to understand how the non-proteinogenic and highly functionalized amino acid components of these related peptides are assembled and how this pathway can be manipulated to create analogs of these NRPs. Chapter 2 highlights the discovery and bioinformatic analysis of the *cym* biosynthetic pathway, providing a plausible biosynthetic route to cyclomarin and cyclomarazine and insights into the prenylation and oxidative tailoring of these peptides. Chapter 3 further explores the *N*-prenylation of the tryptophan residue contained in both sized peptides, and their amenability to manipulation via mutasynthesis. Chapter 4 documents efforts to understand and manipulate the

biosynthesis of 2-amino-3,5-dimethyl-4-hexenoic acid, the non-proteinogenic amino acid residue of cyclomarin.

1.8: References

(1) Elander, R. Industrial production of ß-lactam antibiotics. *Appl. Microbiol. Biotechnol.* **2003**, *61*, 385-392.

(2) Mulligan, M.; Murray-Leisure, K.; Ribner, B.; Standiford, H.; John, J.; Korvick, J.; Kauffman, C.; Yu, V. Methicillin-resistant *Staphylococcus aureus*: a consensus review of the microbiology, pathogenesis, and epidemiology with implications for prevention and management. *Am. J. Med.* **1993**, *94*, 313-328.

(3) Raja, A.; Labonte, J.; Lebbos, J.; Kirkpatrick, P. Daptomycin. *Nat. Rev. Drug Discov.* **2003**, *2*, 943-944.

(4) Stähelin, H. The history of cyclosporin A (Sandimmune®) revisited: another point of view. *Cell. Mol. Life Sci.* **1996**, *52*, 5-13.

(5) Dewick, P. *Medicinal natural products: a biosynthetic approach*; John Wiley & Sons Inc, 2009.

(6) Amsler, C. D.; Fairhead, V. A.; Callow, J. A. In *Adv. Bot. Res.*; Academic Press: **2005**; Vol. Volume 43, p 1-91.

(7) Pawlik, J. Marine invertebrate chemical defenses. *Chem. Rev.* **1993**, *93*, 1911-1922.

(8) Teasdale, M.; Liu, J.; Wallace, J.; Akhlaghi, F.; Rowley, D. Secondary metabolites produced by the marine bacterium *Halobacillus salinus* that inhibit quorum sensing-controlled phenotypes in gram-negative bacteria. *Appl. Environ. Microbiol.* **2009**, *75*, 567.

(9) Williams, P.; Winzer, K.; Chan, W.; Cámara, M. Look who's talking: communication and quorum sensing in the bacterial world. *Philos. Trans. R. Soc. London, Ser. B.* **2007**, *362*, 1119.

(10) Hay, M. E. Marine chemical ecology: chemical signals and cues structure marine populations, communities, and ecosystems. *Annu. Rev. Mar. Sci.* **2009**, *1*, 193-212.

(11) Pohnert, G.; Boland, W. The oxylipin chemistry of attraction and defense in brown algae and diatoms. *Nat. Prod. Rep.* **2002**, *19*, 108-122.

(12) Renner, M.; Shen, Y.-C.; Cheng, X.-C.; Jensen, R.; Frankmoelle, W.; Kauffman, A.; Fenical, W.; Lobkovsky, E.; Clardy, J. Cyclomarins A-C, new antiinflammatory cyclic peptides produced by a marine bacterium (*Streptomyces* sp.). *J. Am. Chem. Soc.* **1999**, *121*, 11273-11276.

(13) Schultz, A. W.; Oh, D.-C.; Carney, J. R.; Williamson, R. T.; Udwary, D. W.; Jensen, P. R.; Gould, S. J.; Fenical, W.; Moore, B. S. Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **2008**, *130*, 4507-4516.

(14) Feling, R.; Buchanan, G.; Mincer, T.; Kauffman, C.; Jensen, P.; Fenical, W. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed.* **2003**, *42*, 355-357.

(15) Cardellina, J.; Marner, F.; Moore, R. Seaweed dermatitis: structure of lyngbyatoxin A. *Science.* **1979**, *204*, 193.

(16) Golakoti, T.; Yoshida, W.; Chaganty, S.; Moore, R. Isolation and structure determination of nostocyclopeptides A1 and A2 from the terrestrial cyanobacterium *Nostoc* sp. ATCC53789. *J. Nat. Prod.* **2001**, *64*, 54-59.

(17) Golakoti, T.; Yoshida, W.; Chaganty, S.; Moore, R. Isolation and structures of nostopeptolides A1, A2 and A3 from the cyanobacterium *Nostoc* sp. GSV224. *Tetrahedron.* **2000**, *56*, 9093-9102.

(18) Tillett, D.; Dittmann, E.; Erhard, M.; Von Döhren, H.; Börner, T.; Neilan, B. A. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem. Biol.* **2000**, *7*, 753-764.

(19) Zabriskie, T.; Klocke, J.; Ireland, C.; Marcus, A.; Molinski, T.; Faulkner, D.; Xu, C.; Clardy, J. Jaspamide, a modified peptide from a *Jaspis* sponge, with insecticidal and antifungal activity. *J. Am. Chem. Soc.* **1986**, *108*, 3123-3124.

(20) Pettit, G.; Kamano, Y.; Herald, C.; Tuinman, A.; Boettner, F.; Kizu, H.; Schmidt, J.; Baczynskyj, L.; Tomer, K.; Bontems, J. The isolation and structure of a remarkable marine animal antineoplastic constituent: dolastatin 10. *J. Am. Chem. Soc.* **1987**, *109*, 6883-6885.

(21) Cuevas, C.; Pérez, M.; Martín, M.; Chicharro, J.; Fernández-Rivas, C.; Flores, M.; Francesch, A.; Gallego, P.; Zarzuelo, M.; De La Calle, F. Synthesis of ecteinascidin ET-743 and phthalascidin Pt-650 from cyanosafracin B. *Org. Lett.* **2000**, *2*, 2545-2548.

(22) Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **2004**, *21*, 519-538.

(23) Harrigan, G.; Goetz, G. Symbiotic and dietary marine microalgae as a source of bioactive molecules–experience from natural products research. *J. Appl. Phycol.* **2002**, *14*, 103-108.

(24) Molinski, T. F.; Dalisay, D. S.; Lievens, S. L.; Saludes, J. P. Drug development from marine natural products. *Nat. Rev. Drug Discov.* **2009**, *8*, 69-85.

(25) Chauhan, D.; Hideshima, T.; Anderson, K. C. A novel proteasome inhibitor NPI-0052 as an anticancer therapy. *Br. J. Cancer.* **2006**, *95*, 961-965.

(26) He, J.; Magarvey, N.; Piraee, M.; Vining, L. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiology.* **2001**, *147*, 2817.

(27) Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; Mcmahon, M. D.; Thomas, M. G. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **2008**, *5*, 191-211.

(28) Maxwell, A. The interaction between coumarin drugs and DNA gyrase. *Mol. Microbiol.* **1993**, *9*, 681-686.

(29) Fogh, S.; Machtay, M.; Werner-Wasik, M.; Curran Jr, W. J.; Bonanni, R.; Axelrod, R.; Andrews, D.; Dicker, A. P. Phase I trial using Patupilone (epothilone B) and concurrent

radiotherapy for central nervous system malignancies. *Int. J. Radiat. Oncol. Biol. Phys.* **2009**, *77*, 1009-1016.

(30) Huang, H.; Menefee, M.; Edgerly, M.; Zhuang, S.; Kotz, H.; Poruchynsky, M.; Huff, L.; Bates, S.; Fojo, T. A phase II clinical trial of ixabepilone (Ixempra; BMS-247550; NSC 710428), an epothilone B analog, in patients with metastatic renal cell carcinoma. *Clin. Cancer Res.*, *16*, 1634.

(31) Clinicaltrials.gov **2010**; Vol. 2010.

(32) Cox, C.; Rinehart, K.; Moore, M.; Cook, J. Pyochelin: novel structure of an iron-chelating growth promoter for *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4256-4260.

(33) Gehring, A.; Demoll, E.; Fetherston, J.; Mori, I.; Mayhew, G.; Blattner, F.; Walsh, C.; Perry, R. Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **1998**, *5*, 573-586.

(34) Wyckoff, E.; Stoebner, J.; Reed, K.; Payne, S. Cloning of a *Vibrio cholerae* vibriobactin gene cluster: identification of genes required for early steps in siderophore biosynthesis. *J. Bacteriol.* **1997**, *179*, 7055-7062.

(35) Miethke, M.; Marahiel, M. A. Siderophore-based iron acquisition and pathogen control. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413-451.

(36) Hutchison, M.; Gross, D. Lipopeptide phytotoxins produced by *Pseudomonas syringae* pv. syringae: comparison of the biosurfactant and ion channel-forming activities of syringopeptin and syringomycin. *Mol. Plant-Microbe Interact.* **1997**, *10*, 347-354.

(37) Cosmina, P.; Rodriguez, F.; De Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; Van Sinderen, D. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **1993**, *8*, 821-831.

(38) Blunt, J. W.; Copp, B. R.; Munro, M. H. G.; Northcote, P. T.; Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **2010**, 165-237.

(39) Peláez, F. The historical delivery of antibiotics from microbial natural products--Can history repeat? *Biochem. Pharmacol.* **2006**, *71*, 981-990.

(40) Falagas, M. E.; Grammatikos, A. P.; Michalopoulos, A. Potential of old-generation antibiotics to address current need for new antibiotics. *Expert. Rev. Anti. Infect. Ther.* **2008**, *6*, 593-600.

(41) Donadio, S.; Maffioli, S.; Monciardini, P.; Sosio, M.; Jabes, D. Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J. Antibiot.* **2010**, *63*, 423-430.

(42) Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat. Prod. Rep.* **2008**, *25*, 25-34.

(43) Konz, D.; Marahiel, M. How do peptide synthetases generate structural diversity? *Chem. Biol.* **1999**, *6*, R39-R48.

(44) Bormann, C.; Mohrle, V.; Bruntner, C. Cloning and heterologous expression of the entire set of structural genes for nikkomycin synthesis from *Streptomyces tendae* Tü901 in *Streptomyces lividans*. *J. Bacteriol.* **1996**, *178*, 1216-1218.

(45) Broberg, A.; Menkis, A.; Vasiliauskas, R. Kutznerides, Depsipeptides from the actinomycete *Kutzneria* sp. 744 inhabiting mycorrhizal roots of *Picea abies* seedlings. *J. Nat. Prod.* **2006**, *69*, 97-102.

(46) Pelzer, S.; Sussmuth, R.; Heckmann, D.; Recktenwald, J.; Huber, P.; Jung, G.; Wohlleben, W. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob. Agents Chemother.* **1999**, *43*, 1565.

(47) Van Wageningen, A.; Kirkpatrick, P.; Williams, D.; Harris, B.; Kershaw, J.; Lennard, N.; Jones, M.; Jones, S.; Solenberg, P. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* **1998**, *5*, 155-162.

(48) Miao, V.; Brost, R.; Chapple, J.; She, K.; Gal, M.-F.; Baltz, R. The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 129-140.

(49) Hojati, Z.; Milne, C.; Harvey, B.; Gordon, L.; Borg, M.; Flett, F.; Wilkinson, B.; Sidebottom, P.; Rudd, B.; Hayes, M. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **2002**, *9*, 1175-1187.

(50) Velasco, A.; Acebo, P.; Gomez, A.; Schleissner, C.; Rodríguez, P.; Aparicio, T.; Conde, S.; Muñoz, R.; De La Calle, F.; Garcia, J. L.; Sánchez-Puelles, J. M. Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. *Mol. Microbiol.* **2005**, *56*, 144-154.

(51) Vaillancourt, F.; Yin, J.; Walsh, C. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α–ketoglutarate-and O_2-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 10111.

(52) Molnar, I.; Schupp, T.; Ono, M.; Zirkle, R.; Milnamow, M.; Nowak-Thompson, B.; Engel, N.; Toupet, C.; Stratmann, A.; Cyr, D. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **2000**, *7*, 97-109.

(53) Rachid, S.; Krug, D.; Kunze, B.; Kochems, I.; Scharfe, M.; Zabriskie, T. M.; Blöcker, H.; Müller, R. Molecular and biochemical studies of chondramide formation highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem. Biol.* **2006**, *13*, 667-681.

(54) Walsh, C. T.; Fischbach, M. A. Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc.* **2010**, *132*, 2469-2493.

(55) Marahiel, M.; Essen, L. Nonribosomal peptide synthetases: mechanistic and structural aspects of essential domains. *Methods Enzymol.* **2009**, *458*, 337-351.

(56) Samel, S.; Marahiel, M.; Essen, L. How to tailor non-ribosomal peptide products—new clues about the structures and mechanisms of modifying enzymes. *Mol. BioSyst.* **2008**, *4*, 387-393.

(57) Walsh, C.; Chen, H.; Keating, T.; Hubbard, B.; Losey, H.; Luo, L.; Marshall, C.; Miller, D.; Patel, H. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **2001**, *5*, 525-534.

(58) Miao, V.; Coeffet-Legal, M.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M. Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology.* **2005**, *151*, 1507.

(59) Hoppert, M.; Gentzsch, C.; Schörgendorfer, K. Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in *Tolypocladium inflatum*. *Arch. Microbiol.* **2001**, *176*, 285-293.

(60) Offenzeller, M.; Santer, G.; Totschnig, K.; Su, Z.; Moser, H.; Traber, R.; Schneider-Scherzer, E. Biosynthesis of the unusual amino acid (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine of cyclosporin A: enzymatic analysis of the reaction sequence including identification of the methylation precursor in a polyketide pathway *Biochemistry.* **1996**, *35*, 8401-8412.

(61) Chen, H.; Thomas, M.; O'connor, S.; Hubbard, B.; Burkart, M.; Walsh, C. Aminoacyl-S-enzyme intermediates in β-hydroxylations and α,β-desaturations of amino acids in peptide antibiotics. *Biochemistry.* **2001**, *40*, 11651-11659.

(62) Puk, O.; Bischoff, D.; Kittel, C.; Pelzer, S.; Weist, S.; Stegmann, E.; Sussmuth, R.; Wohlleben, W. Biosynthesis of chloro-β-hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J. Bacteriol.* **2004**, *186*, 6093.

(63) Chen, H.; Walsh, C. Coumarin formation in novobiocin biosynthesis:β-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem. Biol.* **2001**, *8*, 301-312.

(64) Chen, H.; Hubbard, B. K.; O'connor, S. E.; Walsh, C. T. Formation of β-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **2002**, *9*, 103-112.

(65) Eustáquio, A.; Mcglinchey, R.; Liu, Y.; Hazzard, C.; Beer, L.; Florova, G.; Alhamadsheh, M.; Lechner, A.; Kale, A.; Kobayashi, Y. Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12295.

(66) Mcglinchey, R. P.; Nett, M.; Eustáquio, A. S.; Asolkar, R. N.; Fenical, W.; Moore, B. S. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **2008**, *130*, 7822-7823.

(67) Neary, J.; Powell, A.; Gordon, L.; Milne, C.; Flett, F.; Wilkinson, B.; Smith, C.; Micklefield, J. An asparagine oxygenase (AsnO) and a 3-hydroxyasparaginyl phosphotransferase (HasP) are involved in the biosynthesis of calcium-dependent lipopeptide antibiotics. *Microbiology.* **2007**, *153*, 768.

(68) Fujimori, D.; Hrvatin, S.; Neumann, C.; Strieker, M.; Marahiel, M.; Walsh, C. Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16498.

(69) Strieker, M.; Nolan, E. M.; Walsh, C. T.; Marahiel, M. A. Stereospecific synthesis of threo- and erythro-β-hydroxyglutamic acid during kutzneride biosynthesis. *J. Am. Chem. Soc.* **2009**, *131*, 13523-13530.

(70) Makris, T.; Chakrabarti, M.; Münck, E.; Lipscomb, J. A family of diiron monooxygenases catalyzing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15391.

(71) Fu, C.; Tang, M.; Peng, C.; Li, L.; He, Y.; Liu, W.; Tang, G. Biosynthesis of 3-hydroxy-5-methyl-o-methyltyrosine in the saframycin/safracin biosynthetic pathway. *J. Microbiol. Biotechnol.* **2009**, *19*, 439.

(72) Edwards, D. J.; Gerwick, W. H. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J. Am. Chem. Soc.* **2004**, *126*, 11432-11433.

(73) Huynh, M. U.; Elston, M. C.; Hernandez, N. M.; Ball, D. B.; Kajiyama, S.-I.; Irie, K.; Gerwick, W. H.; Edwards, D. J. Enzymatic production of (-)-indolactam V by LtxB, a cytochrome P450 monooxygenase. *J. Nat. Prod.* **2009**, *73*, 71-74.

(74) Wagner, C.; El Omari, M.; Ko Nig, G. Biohalogenation: nature's way to synthesize halogenated metabolites. *J. Nat. Prod.* **2009**, *72*, 540-553.

(75) Heemstra, J. R.; Walsh, C. T. Tandem action of the O2- and FADH2-dependent halogenases KtzQ and KtzR produce 6,7-dichlorotryptophan for kutzneride assembly. *J. Am. Chem. Soc.* **2008**, *130*, 14024-14025.

(76) Bister, B.; Bischoff, D.; Nicholson, G. J.; Stockert, S.; Wink, J.; Brunati, C.; Donadio, S.; Pelzer, S.; Wohlleben, W.; Süssmuth, R. D. Bromobalhimycin and chlorobromobalhimycins—illuminating the potential of halogenases in glycopeptide antibiotic biosyntheses. *ChemBioChem.* **2003**, *4*, 658-662.

(77) Dong, C.; Huang, F.; Deng, H.; Schaffrath, C.; Spencer, J.; O'hagan, D.; Naismith, J. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature.* **2004**, *427*, 561-565.

(78) Eustáquio, A.; Pojer, F.; Noel, J.; Moore, B. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2007**, *4*, 69-74.

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

(80) Read, J.; Walsh, C. The lyngbyatoxin biosynthetic assembly line: chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. *J. Am. Chem. Soc.* **2007**, *129*, 15762-15763.

(81) Losey, H.; Peczuh, M.; Chen, Z.; Eggert, U.; Dong, S.; Pelczer, I.; Kahne, D.; Walsh, C. Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. *Biochemistry.* **2001**, *40*, 4745-4755.

(82) Solenberg, P.; Matsushima, P.; Stack, D.; Wilkie, S.; Thompson, R.; Baltz, R. Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*. *Chem. Biol.* **1997**, *4*, 195-202.

(83) Ankenbauer, R.; Staley, A.; Rinehart, K.; Cox, C. Mutasynthesis of siderophore analogues by *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 1878-1882.

(84) Serino, L.; Reimmann, C.; Visca, P.; Beyeler, M.; Chiesa, V.; Haas, D. Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron- regulated pchDCBA operon in *Pseudomonas aeruginosa*. *J. Bacteriol.* **1997**, *179*, 248-257.

(85) Miller, D.; Walsh, C. Yersiniabactin synthetase: probing the recognition of carrier protein domains by the catalytic heterocyclization domains, Cy1 and Cy2, in the chain-initiating HMWP2 subunit. *Biochemistry.* **2001**, *40*, 5313-5321.

(86) Gehring, A.; Mori, I.; Perry, R.; Walsh, C. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry.* **1998**, *37*, 11637-11650.

(87) Marshall, C.; Burkart, M.; Keating, T.; Walsh, C. Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate. *Biochemistry.* **2001**, *40*, 10655-10663.

(88) Chen, H.; O'connor, S.; Cane, D.; Walsh, C. Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem. Biol.* **2001**, *8*, 899-912.

(89) Schneider, T. L.; Shen, B.; Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry.* **2003**, *42*, 9722-9730.

(90) Shen, B.; Du, L.; Sanchez, C.; Chen, M.; Edwards, D. Bleomycin biosynthesis in *Streptomyces verticillus* ATCC15003: A model of hybrid peptide and polyketide biosynthesis. *Bioorg. Chem.* **1999**, *27*, 155-171.

(91) Shen, B.; Du, L.; Sanchez, C.; Edwards, D. J.; Chen, M.; Murrell, J. M. The biosynthetic gene cluster for the anticancer drug bleomycin from *Streptomyces verticillus* ATCC15003 as a model for hybrid peptide–polyketide natural product biosynthesis. *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 378-385.

(92) Roy, R.; Gehring, A.; Milne, J.; Belshaw, P.; Walsh, C. Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Nat. Prod. Rep.* **1999**, *16*, 249-263.

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

(94) Koketsu, K.; Watanabe, K.; Suda, H.; Oguri, H.; Oikawa, H. Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nature chemical biology.* **2010**.

(95) Hoffmann, D.; Hevel, J.; Moore, R.; Moore, B. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. *Gene.* **2003**, *311*, 171-180.

(96) Luesch, H.; Hoffmann, D.; Hevel, J.; Becker, J.; Golakoti, T.; Moore, R. Biosynthesis of 4-methylproline in cyanobacteria: cloning of *nosE* and *nosF* genes and biochemical characterization of the encoded dehydrogenase and reductase activities. *J. Org. chem.* **2003**, *68*, 83-91.

(97) Mahlert, C.; Kopp, F.; Thirlway, J.; Micklefield, J.; Marahiel, M. A. Stereospecific enzymatic transformation of α-ketoglutarate to (2S,3R)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J. Am. Chem. Soc.* **2007**, *129*, 12011-12018.

(98) Kempter, C.; Kaiser, D.; Haag, S.; Nicholson, G.; Gnau, V.; Walk, T.; Gierling, K.; Decker, H.; Zähner, H.; Jung, G.; Metzger, J. CDA: Calcium-dependent peptide antibiotics from *Streptomyces coelicolor* A3(2) containing unusual residues. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 498-501.

(99) Chen, H.; Tseng, C. C.; Hubbard, B. K.; Walsh, C. T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14901-14906.

(100) Li, T.; Choroba, O.; Hong, H.; Williams, D.; Spencer, J. Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (S)-3, 5-dihydroxyphenylglycine. *Chem. Commun.* **2001**, *2001*, 2156-2157.

(101) Pfeifer, V.; Nicholson, G. J.; Ries, J.; Recktenwald, J.; Schefer, A. B.; Shawky, R. M.; Schroder, J.; Wohlleben, W.; Pelzer, S. A polyketide synthase in glycopeptide biosynthesis. *J. Biol. Chem.* **2001**, *276*, 38370-38377.

(102) Tseng, C. C.; Vaillancourt, F. H.; Bruner, S. D.; Walsh, C. T. DpgC Is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic pathway. *Chem. Biol.* **2004**, *11*, 1195-1203.

(103) Beer, L. L.; Moore, B. S. Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica*. *Org. Lett.* **2007**, *9*, 845-848.

(104) Mahlstedt, S.; Fielding, E. N.; Moore, B. S.; Walsh, C. T. Prephenate decarboxylases: a new prephenate-utilizing enzyme family that performs nonaromatizing decarboxylation en route to diverse secondary metabolites. *Biochemistry.* **2010**, *49*, 9021-9023.

(105) Ling, H.; Wang, G.; Tian, Y.; Liu, G.; Tan, H. SanM catalyzes the formation of 4-pyridyl-2-oxo-4-hydroxyisovalerate in nikkomycin biosynthesis by interacting with SanN. *Biochem. Biophys. Res. Commun.* **2007**, *361*, 196-201.

(106) Arai, H.; Yamamoto, T.; Ohishi, T.; Shimizu, T.; Nakata, T.; Kudo, T. Genetic organization and characteristics of the 3-(3-hydroxyphenyl) propionic acid degradation pathway of *Comamonas testosteroni* TA441. *Microbiology.* **1999**, *145*, 2813.

(107) Ferrandez, A.; Garcia, J.; Diaz, E. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of *Escherichia coli* K-12. *J. Bacteriol.* **1997**, *179*, 2573-2581.

(108) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468-3496.

(109) Du, L.; Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 215.

(110) Liu, Y.; Hazzard, C.; Eustaquio, A.; Reynolds, K.; Moore, B. Biosynthesis of salinosporamides from α,β-unsaturated fatty acids: implications for extending polyketide synthase diversity. *J. Am. Chem. Soc.* **2009**, *131*, 10376-10377.

(111) Weist, S.; Süssmuth, R. Mutational biosynthesis—a tool for the generation of structural diversity in the biosynthesis of antibiotics. *Appl. Microbiol. Biotechnol.* **2005**, *68*, 141-150.

(112) Wilkinson, B.; Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* **2007**, *3*, 379-386.

(113) Bormann, C.; Kálmánczhelyi, A.; Süßmuth, R.; Jung, G. Production of nikkomycins Bx and Bz by mutasynthesis with genetically engineered *Streptomyces tendae* TÜ901. *J. Antibiot.* **1999**, *52*, 102-108.

(114) Pojer, F.; Wemakor, E.; Kammerer, B.; Chen, H.; Walsh, C.; Li, S.; Heide, L. CloQ, a prenyltransferase involved in clorobiocin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2316.

(115) Galm, U.; Heller, S.; Shapiro, S.; Page, M.; Li, S.; Heide, L. Antimicrobial and DNA gyrase-inhibitory activities of novel clorobiocin derivatives produced by mutasynthesis. *Antimicrob. Agents Chemother.* **2004**, *48*, 1307.

(116) Galm, U.; Dessoy, M.; Schmidt, J.; Wessjohann, L.; Heide, L. In vitro and in vivo production of new aminocoumarins by a combined biochemical, genetic, and synthetic approach. *Chem. Biol.* **2004**, *11*, 173-183.

(117) Nett, M.; Gulder, T. A. M.; Kale, A. J.; Hughes, C. C.; Moore, B. S. Function-oriented biosynthesis of β-lactone proteasome inhibitors in *Salinispora tropica*. *J. Med. Chem.* **2009**, *52*, 6163-6167.

(118) Nguyen, K.; Ritz, D.; Gu, J.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 17462.

(119) Nguyen, K.; He, X.; Alexander, D.; Li, C.; Gu, J.; Mascio, C.; Van Praagh, A.; Mortin, L.; Chu, M.; Silverman, J. Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties. *Antimicrob. Agents Chemother.* **2010**, *54*, 1404.

(120) Butz, D.; Schmiederer, T.; Hadatsch, B.; Wohlleben, W.; Weber, T.; Süssmuth, R. D. Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem.* **2008**, *9*, 1195-1200.

(121) Mootz, H.; Kessler, N.; Linne, U.; Eppelmann, K.; Schwarzer, D.; Marahiel, M. Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **2002**, *124*, 10980-10981.

Chapter 2:

Biosynthesis of Cyclomarins and Cyclomarazines, Prenylated Cyclic Peptides of Marine Actinobacterial Origin

2.1: Abstract

The heptapeptide cyclomarin and the dipeptide cyclomarazine are structurally related cyclic peptides containing modified amino acid residues, including derivatives of *N*-(1,1-dimethylallyl)-tryptophan and δ-hydroxyleucine, which are common in the di- and heptapeptide series. Stable isotope incorporation studies in *Streptomyces* sp. CNB-982, which was first reported to produce the cyclomarin anti-inflammatory agents, illuminated the biosynthetic building blocks associated with the major metabolite cyclomarin A, signifying that this marine microbial peptide is nonribosomally derived. DNA sequence analysis of the 5.8 Mb *S. arenicola* circular genome and PCR-targeted gene inactivation experiments identified the 47 kb cyclomarin/cyclomarazine biosynthetic gene cluster (*cym*) harboring 23 open reading frames. The cym locus is dominated by the 23 358 bp *cymA*, which encodes a 7-module nonribosomal peptide synthetase (NRPS) responsible for assembly of the full-length cyclomarin heptapeptides as well as the truncated cyclomarazine dipeptides. The unprecedented biosynthetic feature of the megasynthetase CymA to synthesize differently sized peptides in vivo may be triggered by the level of β oxidation of the priming tryptophan residue, which is oxidized in the cyclomarin series and unoxidized in the cyclomarazines. Biosynthesis of the *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-hydroxytryptophan residue of cyclomarin A was further illuminated through gene inactivation experiments, which suggest that the tryptophan residue is reverse prenylated by CymD prior to release of the cyclic peptide from the CymA megasynthetase, whereas the cytochrome P450 CymV

installs the epoxide group on the isoprene of cyclomarin C post-NRPS assembly. Additional gene inactivation experiments confirmed the role of CymS in tryptophan β oxidation and CymW in leucine δ oxidation. Last, the novel amino acid residue 2-amino-3,5-dimethylhex-4-enoic acid in the cyclomarin series was shown by bioinformatics and stable isotope experiments to derive from a new pathway involving condensation of isobutyraldehyde and pyruvate followed by *S*-adenosylmethionine methylation. Assembly of this unsaturated, branched amino acid is unexpectedly related to the degradation of 3-(3-hydroxyphenyl)propionic acid.

2.2: Introduction

2.2.1: Cyclomarin consists of several modified and non-proteinogenic amino acids of significant biosynthetic interest

Cyclomarins A-C (**1**-**3**) are potent anti-inflammatory marine natural products first described from an estuarine streptomycete, strain CNB-982 (Figure 1.1).[1] Chemotyping analysis of *Salinispora arenicola* CNS-205, an actinobacterial isolate from Palau,[2] revealed the production of potent secondary metabolites belonging to the rifamycin and staurosporin structural families in addition to the known anti-inflammatory cyclic peptides cyclomarin A (**1**) and C (**3**).[3] The novel analog, cyclomarin D (**4**, Figure 1), was discovered by Dong-Chan Oh[4] in the fermentation broth of this marine isolate, further extending the structural diversity of this peptide natural product. Slight variations in oxidation and methylation differentiate these 21-membered cyclic heptapeptides, which are comprised of a number of highly tailored proteinogenic and nonproteinogenic

amino acid residues, including *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-hydroxytryptophan, *N*-methyl-δ-hydroxyleucine, β-methoxyphenylalanine, *N*-methylleucine, and 2-amino-3,5-dimethylhex-4-enoic acid (ADH) in addition to unmodified alanine and valine moieties. The syntheses of these unusual residues and the total synthesis of **3** have been described.[5] In addition to the cyclomarins, the related diketopiperazines (DKPs) cyclomarazine A (**5**) and B (**6**) were isolated from *S. arenicola* CNS-205[4] by Dong-Chan Oh, which contain *N*-(1,1-dimethyl-1-allyl)-tryptophan and *N*-methyl-δ-hydroxyleucine residues. Because of the similarity between the first two residues of cyclomarin and cyclomarazine, it is likely that they share a common biosynthetic origin.

There is an interesting structural similarity between cyclomarin and rufomycin[6] (Figure 2.2) class of compounds, which includes the ilamycins isolated from *Streptomyces islandicus*[7,8] and K95-5076 isolated from *Actinoplanes* sp. K95-5076.[9] Both sets of cyclic hepapeptides contain the *N*-(1,1-dimethyl-1-allyl)-tryptophan residue, which like cyclomarin is twice oxidized in some analogs, once at the β-position and once across the prenyl olefin. δ-oxidation of leucine is also observed in both compound classes, forming a hydroxyl functionality in **1**-**3**,**4** and a aldehyde function in ilamycin, and is likely to facilitate the formation of the lactam ring of K95-5076. Along with the extensive presence of aliphatic proteinogenic amino acids, both contain a non-proteinogenic hexenoic amino acid containing an unsaturation between the γ- and δ-positions of the side chain, in a position *N* terminal to the Trp derived residue. ADH is a novel structural unit of **1**-**3** and lacks a clear biosynthetic

origin. Although the structure of this eight-carbon, nonproteinogenic amino acid residue is characteristic of proteinogenic branched chain amino acids, it has not been encountered in other natural products. Cyclomarin's δ-hydroxyleucine and reverse *N*-prenylated β-hydroxytryptophan units are rare peptidic building blocks with limited distribution beyond the rufomycin compounds. δ-Hydroxyleucine was previously described in depsipeptides from *Paecilomyces lilacinus*[10] and *Biploris zeicola*[11] fungal metabolites where preliminary enzymological studies suggest Fe(II)/α-ketoglutarate-dependent hydroxylases and related enzymes catalyze hydroxylation at aliphatic carbons of amino acid residues.[12] The prenylation of indoles is commonplace in fungal metabolites such as the ergot alkaloids ergotamine and fumigaclavine, the diketopiperazines fumitremorgin B and brevianamide A and the indole diterpenes paxilline and lolitrem B,[13] the *N*-prenylation of the indole nitrogen of tryptophan is a rare biosynthetic feature in fungi and bacteria. One of the few examples of *N*-prenylated tryptophan derivatives, in addition to the cyclomarins and rufomycins, is flustramine A from the marine bryozoan *Flustra foliacea*.[14] Several enzymes responsible for the transfer of an isoprene to an indole ring have been characterized from bacteria and fungi, including LtxC from the cyanobacterium *Lyngbya majuscula*, which catalyzes the transfer of geranyl pyrophosphate to indolactam-V during the biosynthesis of lyngbyatoxin.[15] Bioinformatic analysis of the genome sequence of *Aspergillus fumigatus* revealed several prenyltransferases (PTases),[16] four of which have been characterized as of 2007. Three of these enzymes, FtmPT1,[17] FgaPT2,[18] and 7-DMATS,[19] prenylate the indole ring at C2, C4, and C7,

respectively, while the other example, catalyzes a “forward” prenyl transfer to the indole nitrogen.[20]

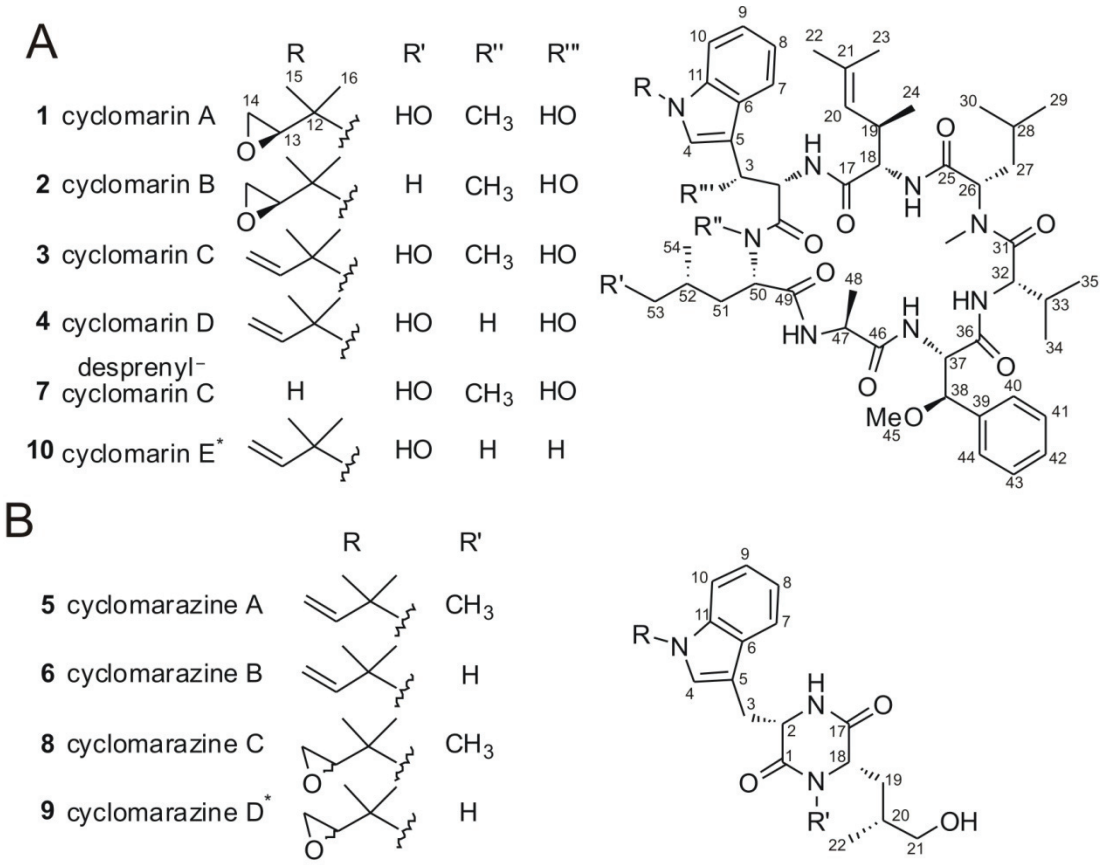


Figure 2.1: Structures of cyclomarin (1-4, 7, 10) and cyclomarazine (5, 6, 8, 9) cyclic peptides from *Salinispora arenicola* CNS-205 and *Streptomyces* sp. CNB-982. *Structures inferred by experimental evidence, but have not been fully characterized.

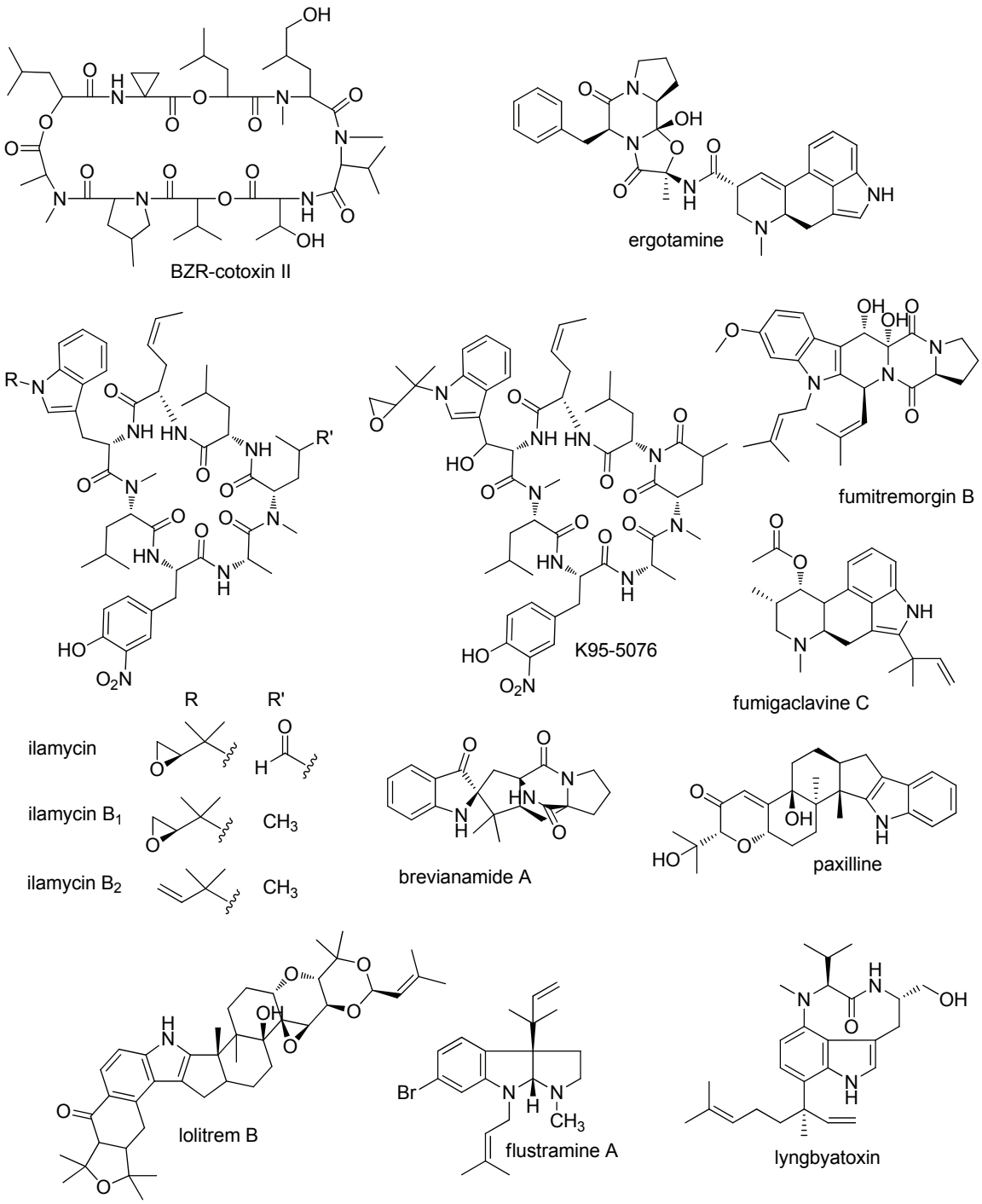


Figure 2.2: Compounds sharing structural features with cyclomarin and cyclomarazine. The stereochemistry of BZR-cotoxin II was not reported.

2.2.2: Specific Aims

On the basis of these unusual structural features, we set out to explore the biosynthesis of the cyclomarin peptides in *S.arenicola* CNS-205 in collaboration with Dong-Chan Oh, Daniel Udwary, William Fenical and Paul Jensen at Scripps Institution of Oceanography with a focus of understanding the origins of the unique structural features of these cyclic heptapeptides. In addition, we collaborated with John Carney, Thomas Williamson, and Steven Gould, formally of Oregon State University, who probed the biosynthesis of cyclomarin in *Streptomyces* sp. CNB-982 with stable isotopes. Through these collaborative studies, it was revealed that cyclomarin A, C and D along with the truncated diketopiperazine (DKP) cyclomarazines A-B, are produced by the marine actinomycete, *Salinispora arenicola* CNS-205,[21,22] and appear to share a common biogenesis. Sequence analysis of the finished genome of *S. arenicola* CNS-205 (GenBank accession number NC_009953) provided the cyclomarin/cyclomarazine nonribosomal peptide synthetase (NRPS) biosynthetic gene cluster, which furthermore permitted us to explore the biosynthesis of these cyclic peptides at the biochemical and genetic levels. Re-analysis of the data provided from stable isotope feeding studies in *Streptomyces* sp. CNB-982, guided by bioinformatic analysis of the cyclomarin/cyclomarazine (*cym*) gene cluster, provided insights into the biosynthesis of both sized peptides. These insights directed gene inactivation studies along with new stable isotope incorporation studies in *S. arenicola*, confirming several bioinformatic predictions

and solidified the role of the *cym* cluster in the biosynthesis of both sized peptides. Over the course of these studies, several new cyclomarin analogs were generated in mutants deficient in specific *cym* genes involved in prenylation and oxidative tailoring, and two previously unreported cyclomarazine analogs were observed. Here we report a comprehensive investigation of cyclomarin and cyclomarazine biosynthesis utilizing numerous techniques from the chemistry/biology interface.

2.3: Results

2.2.1: Molecular basis for the biosynthesis of the cyclomarins

Structural inspection of the cyclomarins and cyclomarazines suggested that these cyclic peptides are biosynthesized nonribosomally[23] since they harbor modified amino acid residues (*N*-methylation, *N*-prenylation, and side chain oxidation) as well as the novel nonproteinogenic amino acid residue 2-amino-3,5-dimethyl-4-hexenoic acid (ADH). Feeding experiments with [U-^{13}C]glucose, [methyl-^{13}C]methionine, and [3'-^{13}C]tryptophan in *Streptomyces* sp. CNB-982, performed by J. Carney, R. T. Williamson and S. Gould and reinterpreted here, established the biosynthetic origins of the precursor building blocks in **1**, thereby confirming that all of the modified residues except for ADH are derived from standard amino acids (Figure 2.3, Table 2.2). Four methyl carbons in **1**, namely, the two *N*-methyl amides, the *O*-methyl group at C38, and the C24 methyl of the ADH residue, are derived from methionine. Taken together, the biosynthesis of **1** is expected to involve a seven-module NRPS that assembles largely modified amino acid residues involving four methyltransferases (MTs), four oxygenases, a

prenyltransferase (PTase), and a dedicated pathway to ADH. Several proposals for the ADH pathway based on the biosynthesis of branched chain amino acids were put forth by Williamson[24] and reproduced as Figure 2.4. The initial interpretation of the [U-^{13}C]glucose and [methyl-^{13}C]methionine feeding studies, in which the INADEQUATE ^{13}C-^{13}C correlation between C18 and C19 was dismissed as an interunit coupling artifact, supported a pathway that incorporates leucine, acetate, and methionine. Therefore, a pathway similar to that for isoleucine biosynthesis is expected to provide the ADH residue of **1**.

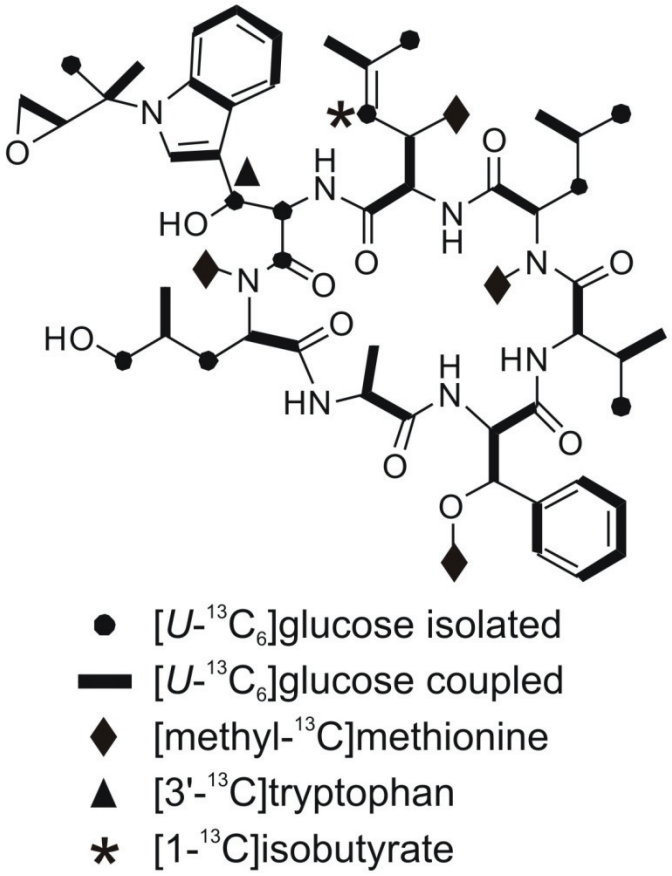


Figure 2.3: Biosynthetic origin of the carbons in cyclomarin A (1). ^{13}C-Labeling patterns from the incorporation of [U-$^{13}C_6$]glucose (bold lines and dots), [methyl-^{13}C]methionine (diamonds), [3'-13C]tryptophan (triangle), and [1-^{13}C]-isobutyrate (asterisk) are depicted. Intact incorporation of ^{13}C-^{13}C units was determined by INADEQUATE NMR.

Table 2.1: NMR Data for [U-^{13}C]glucose-derived cyclomarin A (1)

residue	C no.	^{13}C	$^{1}J_{cc}$ (Hz)	INADEQUATE
N-(1,1-dimethyl-2,3-epoxypropyl)-β-OH-Trp	1	171.0		
	2	52.9		
	3	68.6		
	4	123.4	71	C5
	5	111.9	71	C4
	6	127.0	56	C11
	7	119.2	a	
	8	119.8	*a*	
	9	122.1	*a*	
	10	113.7	*a*	
	11	136.1	56	C6
	12	58.1	38	C16, C15[b]
	13	57.8	29	C14
	14	45.4	29	C13
	15	23.1	38	C12[b]
	16	24.5	38	C12
ADH	17	172.5	54	C18
	18	58.1	55	C17, C19[c]
	19	35.5	36	C18[d]
	20	124.8		
	21	134.4	42	C23
	22	25.7		
	23	18.9	42	C21
	24	18.5		
N-Me-Leu	25	168.4	52	C26
	26	58.6	52	C25
	27	38.9		
	28	25.0	35	C29
	29	24.4	35	C28
	30	23.5		
	26-*N*-Me	29.6		
Val	31	170.6	55	C32
	32	55.2	55	C31
	33	30.8	35	C35
	34	19.3		
	35	20.0	35	C33
β-OMe-Phe	36	169.6	50	C37
	37	55.9	50, 40	C36, C38
	38	80.0	40	C37
	39	135.1	57	C44
	40	127.8	57	
	41	128.3	*a*	
	42	128.7	*a*	
	43	128.3	*a*	
	44	127.8	57	
	45	57.8	*a*	
Ala	46	171.6	55	C47
	47	50.6	55, 35	C46, C48
	48	20.8	35	C47
N-Me-δ-OH-Leu	49	168.8	52	C50
	50	59.3	52	C49
	51	33.1		
	52	33.2	35	C54
	53	66.5		
	54	17.6	35	C52
	N-8-Me	29.2		

[a] Low-intensity coupled signals were observed, but ^{1}J was not determined. [b] Minor coupling species indicative of the MEP pathway. [c] Single enrichment also evident. [d] Two coupling species identified: doublet ($J_{C17-C18}$) and doublet of doublets ($J_{C17-C18}/J_{C18-C19}$).

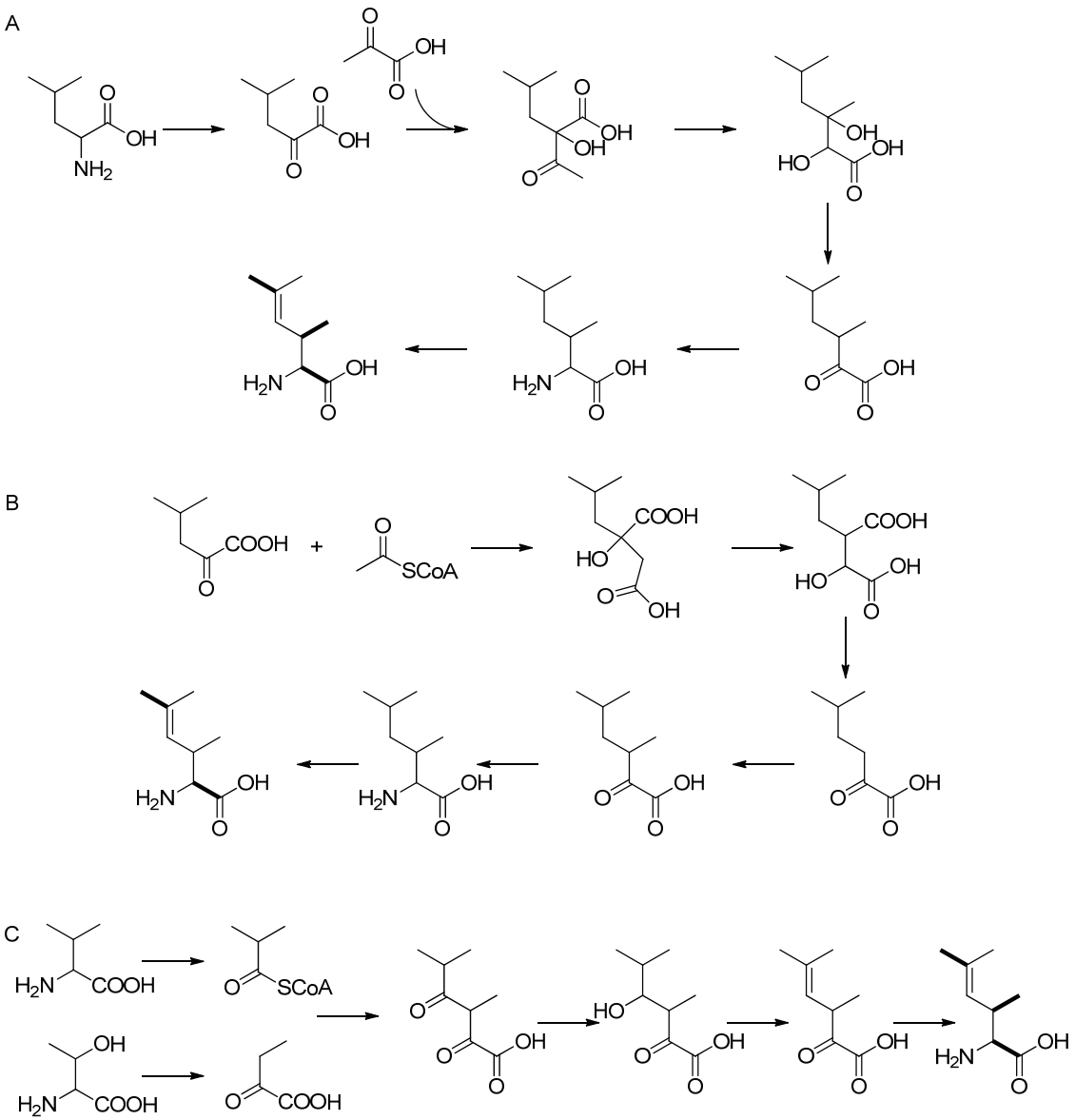


Figure 2.4: Potential pathways to ADH based on branched chain amino acid biosynthesis. Two pathways are depicted here that are analogous to isoleucine (A), and leucine (B) biosynthesis. The third pathway(C) depicts a pathway utilizing valine intermediates to arrive at ADH. Intact two carbon units from glucose are depicted with bold lines. Figure adapted from Williamson[24].

We recently completed the genome sequence analysis of the cyclomarin/cyclomarazine-producing bacterium *S. arenicola* CNS-205, which revealed 10 NRPS-related biosynthetic gene clusters in its 5,786,361 bp

genome.[25] Further inspection identified a 47,477 bp biosynthetic gene cluster (*cym*) harboring 23 open reading frames (ORFs) whose deduced functions are consistent with cyclomarin biosynthesis. The *cym* locus is dominated by the largest ORF in the *S. arenicola* genome, the 23,358 bp cymA, which codes for a heptamodular NRPS (Figure 2.5). Its domain architecture is consistent with the cyclomarin heptapeptides in which modules-2 and -6 additionally harbor MT domains, suggesting that the cyclic peptide framework is wholly assembled by this single megasynthetase. Bioinformatic analysis of the seven adenylation (A) domains in CymA strongly correlated module-4 to the activation of L-phenylalanine (Table 2.2) while the specificities of the remaining A domains were less clear.[26,27] Taken together, these data suggested the order of incorporation of the amino acid residues in cyclomarin biosynthesis starting with the tryptophan derivative and ending with ADH to yield a linear heptapeptide intermediate bound to the thiolation (T) domain in module-7 (Figure 2.5). Subsequent release and macrocyclization is then putatively achieved by the C-terminal thioesterase (TE). The cyclomarazine DKPs are consistent with this biosynthetic model in which the module-2-bound diketide is cleaved from CymA to yield the diketopiperazine. This premature cleavage reaction may be facilitated by the type II TE[28] CymQ.

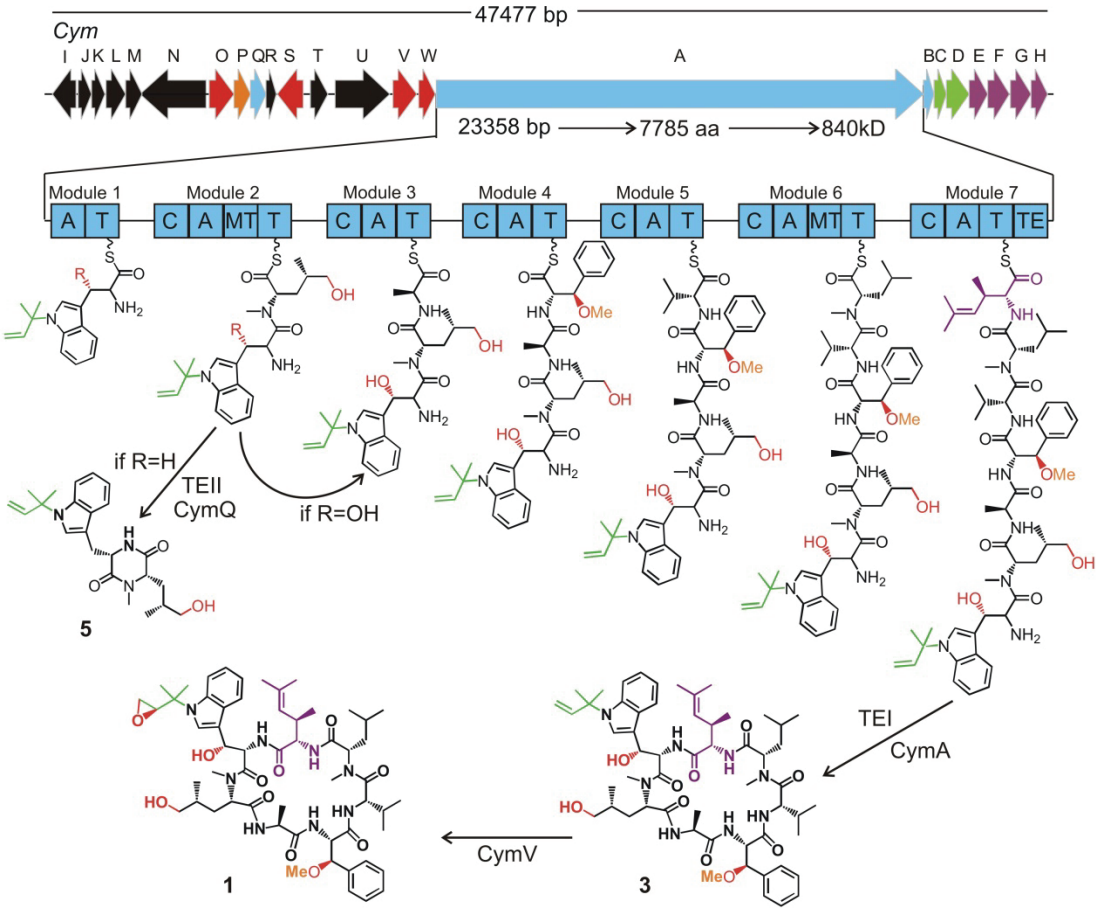


Figure 2.5: Biosynthetic gene cluster organization of cym and proposed biosynthesis of cyclomarin A (1) and cyclomarazine A (5). Each arrow represents the direction of transcription of an ORF and is color coded to signify enzyme function which is further reflected chemically. NRPS-related genes are colored blue with enzymatic domain abbreviations as follows: A, adenylation; T, thiolation (peptidyl carrier protein); C, condensation; MT, methyltransferase; and TE, thioesterase. Oxidative genes are colored red, *O*-methyltransferases in orange, prenylation related genes in green, ADH biosynthetic genes in purple, and regulatory/transport/other genes in black.

The cym cluster further harbors four oxygenases (*cymO*, *cymS*, *cymV*, and *cymW*), a putative PTase (*cymD*), an *O*-MT (*cymP*), a four-gene operon associated with ADH biosynthesis (*cymE-H*), four genes putatively involved in regulation and resistance, and two others in phosphoenolpyruvate metabolic flux (Table 2.3). Pseudogenes containing fragments of integrases and transposases flank the *cym* locus, suggesting that this cluster may have been acquired

horizontally. This observation may explain the co-occurrence of this pathway in *S*. sp. CNB-982.

Table 2.2: Cyclomarin synthetase (CymA) adenylation domain specificity

A-domain	sequence	closest match	% identity	predicted specificity
1	DVWTHGGVAK	DAWTYGGVIK	70	Bmt
2	DALHSGAVAK	DALVMGAVMK	70	Phe
3	DVFFVGLVAK	DVFSVGLVMK	80	D-lyserg
4	DAWTVAAVCK	DAWTVAAVCK	100	Phe
5	DALFVGLVAK	DALFFGLVDK	80	Leu
6	DALLIGAVVK	DAALIGAVFK	80	Val or Leu
7	DAMVLGVVAK	DAQDLGVVDK	70	Gln

Bmt = (4R)-(4E)-2-butenyl-4-methyl-L-threonine
D-lyserg = D-lysergic acid

2.2.2: 2-Amino-3,5-dimethyl-4-hexenoic acid (ADH) biosynthesis

Six of seven amino acid constituents in **1** are clearly recognizable as derived from common protein amino acids modified by methylation, prenylation, and/or hydroxylation. The remaining residue ADH has no proteinogenic counterpart. A number of scenarios involving analogous reactions in branched chain amino acid biosynthesis were envisaged and first probed in *S*. sp. CNB-982 with stable isotopes. The labeling pattern resulting from [U-^{13}C]glucose and [methyl-^{13}C]methionine assimilation (Figure 2.2) initially suggested the involvement of leucine, acetate, and methionine building blocks (Figure 2.4 panel B). The C24 methyl group is clearly derived from methionine, whereas the remaining ADH carbons originate from glucose fragments. The origin of the C17-C19 three-carbon unit in **1** initially led us astray due to two discernible ^{13}C-labeling patterns involving intact and fragmented incorporations of glucose units in which the ^{13}C NMR signal of C19 was singly enriched as well as coupled to

C18, which in turn couples with C17. At the time, we incorrectly assumed that the coupling between C18 and C19 was an artifact arising from interunit coupling. Hence, our working hypothesis to ADH biosynthesis involved a pathway by analogy to leucine biosynthesis (Figure 4.2 panel B) in which α-ketoisocaproic acid could condense with acetyl-CoA followed by isomerization, oxidative decarboxylation, β-methylation, transamination, and desaturation.

Inspection of the *S. arenicola* cym cluster revealed the four gene operon *cymE-H* downstream of the cymA NRPS, which by the process of elimination was predicted to be involved in the biosynthesis of ADH. Indeed, the presence of a SAM dependent methyltrasferase in this operon corresponded well with the [methyl-^{13}C]methionine labeling data. Unfortunately, the three remaining genes have little in common with leucine biosynthesis, nor would be sufficient to support a pathway similar to those depicted in Figure 4.2, panel A or C.

When the proposed functions of CymE-H were taken individually, along with the original interpretation of the [U-^{13}C]glucose, it was difficult to propose a reasonable pathway to ADH. Reanalysis of CymE-H as an operon, not as individual genes, along with reanalysis of the stable isotope incorporation data suggested an entirely different pathway to ADH involving isobutyrate, pyruvate, and methionine precursors (Figure 2.4a). Three of the four genes were homologous to those involved in the degradation of 3-(3-hydroxyphenyl)propionic acid (3-HPP) in *Comamonas testosteroni* TA441.[29]

Table 2.3: Deduced functions of the open reading frames in the *cym* gene cluster

protein	amino acid	accession	protein family (pfam)	proposed function	Mutant chemotype
CymI	386	SARE4547	phosphoenolpyruvate synthase	PEP synthesis from pyruvate	
CymJ	214	SARE4548	no conserved domain	unknown	
CymK	200	SARE4549	fructose-2,6-bisphosphatase	glycolysis regulation	
CymL	314	SARE4550	ABC-type multidrug transporter	cyclomarin transporter	
CymM	264	SARE4551	ABC-type multidrug transporter	cyclomarin transporter	
CymN	1043	SARE4552	DnrI, transcriptional activator	regulation	
CymO	404	SARE4553	cytochrome P450	phenylalanine β-hydroxylation	
CymP	263	SARE4554	*O*-methyltransferase	*O*-methylation of β-hydroxyphenylalanine	
CymQ	242	SARE4555	thioesterase	thioesterase[1]	wild type
CymR	163	SARE4556	no conserved domain	unknown	
CymS	420	SARE4557	cytochrome P450	tryptophan β-hydroxylation[1]	loss of **1**,**3** gain of **10**
CymT	274	SARE4558	transposase	transposase	
CymU	877	SARE4559	LuxR	regulation	
CymV	395	SARE4560	cytochrome P450	epoxidation of prenylated tryptophan[1]	loss of **1** increase of **3**
CymW	264	SARE4561	dioxygenase	leucine δ-hydroxylation[1]	loss of **1**,**3**,**4**,**5**,**8** gain of **2**
CymA	7785	SARE4562	NRPS	cyclomarin synthetase	
CymB	71	SARE4563	MbtH-like protein	NRPS associated	
CymC	196	SARE4564	IPP isomerase type I	isopentyenyl diphosphate isomerase	
CymD	373	SARE4565	no conserved domain[2]	*N*-prenylation of tryptophan residue[1]	loss of **1**,**3**,**4**,**5**,**8** gain of **7**
CymE	295	SARE4566	acetaldehyde dehydrogenase	ADH[3] biosynthesis	
CymF	348	SARE4567	4-hydroxy-2-ketovalerate aldolase	ADH biosynthesis	
CymG	327	SARE4568	*O*-methyltransferase	ADH biosynthesis	
CymH	264	SARE4569	2-keto-4-pentenoate hydratase	ADH biosynthesis	

[1]Function interrogated in this work via PCR directed gene inactivation and chemotype analysis
[2]No entry reported in pfam database for soluble aromatic prenytransferases
[3]2-amino-3,5-dimethyl-4-hexenoic acid

In the catabolism of 3-HPP, the degradant 2-hydroxypenta-2,4-dienoic acid is hydrated by MhpD to 4-hydroxy-2-oxopentanoic acid to facilitate the reverse aldol reaction to pyruvate and acetaldehyde (Figure 2.6b). Acetaldehyde is then subsequently converted to acetyl-CoA by the NAD+-dependent dehydrogenase MhpF. In an analogous manner yet operating in the reverse biosynthetic direction, valine-derived isobutyraldehyde via isobutyryl-CoA is homologated with pyruvate to give 4-hydroxy-5-methyl-2-oxohexanoic acid via CymE and CymF, respectively (Figure 2.6a). Reinterpretation of the [U-^{13}C]glucose INADEQUATE data as a legitimate coupling between C17-C18 and C18-C19 supported pyruvate as a precursor. CymE and CymF share high sequence identity with MhpF (36%; BAA82883) and MhpE (41%; BAA82884), respectively. Dehydration by CymH, which is 35% identical to MhpD (BAA82880), putatively yields 2-hydroxy-5-methylhexa-2,4,-dienoic acid. Methylation by the SAM-dependent MT CymG followed by transamination utilizing a branched-chain amino acid transaminase such as SARE1130 would provide ADH. These last two enzymatic reactions are homologous to those in the biosynthesis of the (2S,3R)-3-methylglutamate residue of CDA from *Streptomyces coelicolor* A3(2),[30] although CymG shares no significant structural identity with the CDA MT (NP_627429). Support for this revised biosynthetic pathway to ADH was achieved in *S. arenicola* CNS-205 with the administration of sodium [1-^{13}C]isobutyrate, which specifically enriched C20 of **1** at 29%.

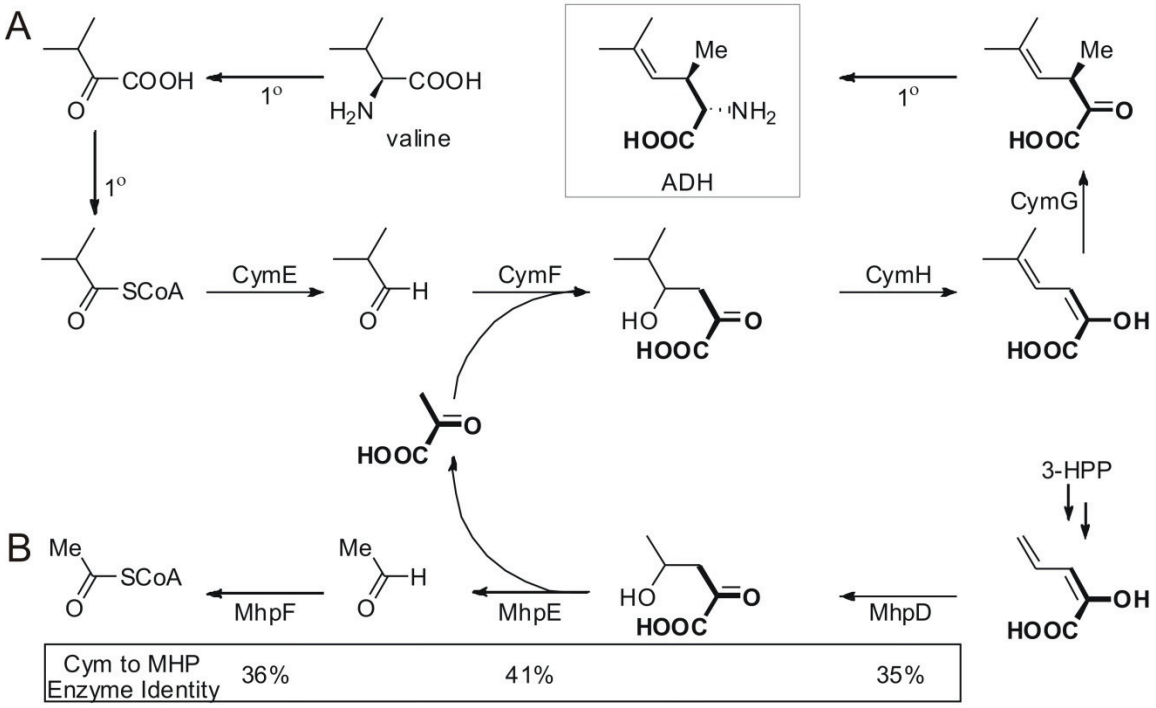


Figure 2.6: Proposed pathway for the biosynthesis of (A) 2-amino-3,5-dimethyl-4-hexenoic acid (ADH) from valine and pyruvate (bold) in *S. arenicola* CNS-205 and its relatedness to the catabolism of (B) 3-(3-hydroxyphenyl)propionic acid (3-HPP) in *C. testosteroni* TA441 (Mhp pathway). Boxed percentages denote % identity of Cym to corresponding enzymes in the Mhp pathway. 1° denotes primary enzymatic reactions in *S. arenicola* encoded by genes outside the *cym* cluster.

2.2.3: Reverse *N*-prenylation of the tryptophan residue

The [U-^{13}C]glucose feeding experiment revealed that the isopentenyl pyrophosphate (IPP)-derived unit attached to the indole nitrogen atom originates in *S*. sp. CNB-982 predominantly from the mevalonate pathway with minor incorporation from the methylerythritol (MEP) pathway (Figure 2.3). While the presence of both isoprenoid pathways has been demonstrated in other streptomycetes,[31] genomic analysis of *S. arenicola* exclusively revealed the complete MEP pathway. Although the *S. arenicola* MEP pathway genes are dispersed in the genome, the cym cluster harbors two terpenoid-related genes,

namely, the type I IPP isomerase *cymC* and the PTase *cymD* (Figure 2.5, Table 2.3). CymD shares 23% identity to the dimethylallyltyrosine synthase SirD (AAS92554) involved in the biosynthesis of sirodesmin PL from the plant pathogenic fungus *Leptosphaeria maculans*[32] and 24% identity to the PTase LtxC (AAT12285), which is involved in the *C*-geranylation of a tryptophan residue during lyngbyatoxin biosynthesis in the cyanobacterium *Lyngbya majuscula*.[15]

To characterize the function of CymD, we inactivated the encoding gene by PCR-targeted gene replacement utilizing a protocol previously adapted in *Salinispora tropica*.[33] Replacement of *cymD* with an apramycin resistance/oriT cassette in the pCC1Fos-based clone BPPW9569 harboring a 38 kb fragment of the cym cluster gave the *S. arenicola cymD*$^-$ knockout mutant. HPLC-MS analysis of the mutant verified elimination of cyclomarin and cyclomarazine production (Figure 2.7). We anticipated the formation of new desprenyl cyclomarin and cyclomarazine analogs in comparable quantities to the wildtype compounds assuming that these molecules would be the substrates for CymD. Although we did isolate and partially characterize desprenylcyclomarin C (**7**) by HR-FTMS (m/z) 981.5408 $[M + Na]^+$, 981.5420 calcd), it was produced in the mutant at approximately 100-fold less than the corresponding **1** in the wild-type strain, thereby impeding its complete chemical characterization via NMR. Further analysis by Wei-Ting Liu, Julio Ng, et al.[34] utilizing MS based cyclic peptide sequencing confirmed the structure of **7** as the desprenyl analog.

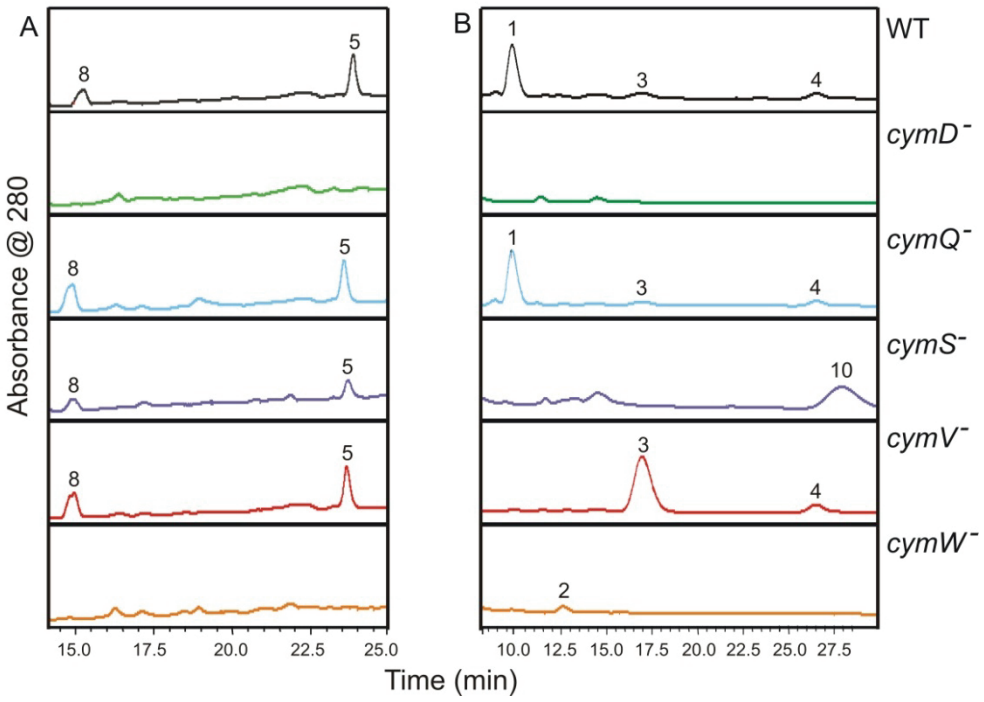


Figure 2.7: Figure 5. LC-MS analysis of the organic extracts of *S. arenicola* CNS-205 (WT) and the *S. arenicola* CNS-205 mutants *cymD*⁻, *cymQ*⁻, *cymS*⁻, *cymV*⁻, and *cymW*⁻. Chromatograms recorded at 210 nm. Panel A, dipeptide analysis, panel B, heptapeptide analysis. See Materials and Methods section for details. Note the complete loss of peaks corresponding to **1, 3**-**5, 8** in the *cymD* replacement mutant. Inactivation of *cymQ* had no significant effect on *cym* related metabolites. Inactivation of *cymS* or *cymV* had no effect on the production of **5** or **8**, while inactivation of *cymW* abolished cyclomarazine (**5**, **8**) production. The loss of production of **1** was observed in *cymS*⁻, or *cymV*⁻, or *cymW*⁻ and a 10-fold concomitant increase in production of **3** in the *cymV* mutant. Also of note is the appearance of a novel metabolite predicted to be cyclomarin E (**10**) in the *cymS*⁻ mutant and the appearance of cyclomarin B (**2**) in the *cymW*⁻ mutant, which has not been observed in *S. arenicola* CNS-205 WT.

2.2.4: Oxidative tailoring

In agreement with the chemical structure of cyclomarin A (**1**), the *cym* gene cluster harbors four oxygenase-encoding genes. BLAST analysis of the Fe(II)/α-ketoglutarate-dependent dioxygenase CymW and combined BLAST and phylogenetic analysis (Appendix 2.7.1) of the three cytochrome P450s CymO, CymS, and CymV, provided insight into their putative biosynthetic functions in amino acid modification (Table 2.3). While there are no ferrodoxins associated with the cym locus to accompany the three P450s, there are eight ferrodoxin encoding genes distributed throughout the *S. arenicola* genome. CymO belongs to a family of NRPS-associated amino acid β-hydroxylases that oxidize the β-position of PCP-bound amino acid residues.[35] Cyclomarin A harbors two such residues, β-methoxyphenylalanine and β-hydroxytryptophan. Immediately downstream of *cymO* is the putative MT *cymP*, which strongly suggests that their protein products are jointly involved in the β-oxidation/methylation of CymA-(T_4)-bound phenylalanine prior to the condensation with the CymA-module-3 tripeptide precursor (Figure 2.5). The coding specificity of the module-4 A domain is strongly consistent with an unmodified phenylalanine substrate (Table 2.2),[26,27] further supporting the hypothesis that these amino acid modifications transpire while tethered to CymA.

Bioinformatic analysis of CymV suggested that this cytochrome P450 is involved in olefin epoxidation. CymV shows sequence similarity to SalD (42% identity; ABG02269) and LtxB (31% identity; AAT12284), which are putative epoxidases involved in the biosynthesis of salinomycin[36] and lyngbyatoxin,[15]

respectively. To probe the function of CymV as the oxygenase responsible for the epoxidation of the isoprene olefin of **3** to **1**, we similarly inactivated the encoding gene cymV by PCR-based mutagenesis. HPLC-MS analysis of the resulting knockout mutant extract showed the loss of **1** with concomitant formation of **3**, which was produced at trace levels in the wild type (Figure 2.7). The structure of the mutant-derived **3** was verified by NMR comparison with published spectra.[1] Since **3** is produced in the mutant at levels comparable to the production of **1** in the wildtype strain, the CymV-catalyzed installation of the epoxide group probably occurs post-NRPS assembly. The wild-type and mutant strains both synthesize the olefinic **4**-**6** as anticipated.

The remaining two oxygenases CymS and CymW are homologous to cytochrome P450s and α-ketoglutarate-dependent iron dioxygenases, respectively, which hydroxylate aliphatic residues. CymS is homologous to CYP107Z4 (51% identity, AAT45275) from *Streptomyces lydicus*, which converts avermectin to 4'-oxo-avermectin at an aliphatic position.[37] CymW has 41% identity to a characterized L-proline 4-hydroxylase[38] (accession BAA20094) from a member of the *Dactylosporangium* genus. PCR targeted gene replacement of *cymS* and *cymW* confirmed their roles in β-hydroxylation of the tryptophan and stereospecific δ-hydroxylation of leucine, respectively. Inactivation of *cymS* abolished production of **1, 3-4** while the DKP series was unaffected. Metabolic profiling (Figure 2.7) revealed the presence of a new compound, cyclomarin E (**10**), with two pseudomolecular ions, $[M+Na]^+$ = 1033 and $[M+H]^+$ = 1011 and a retention time longer than that for **4**, which is consistent with an analog of **4**

lacking an oxidation at the β position of tryptophan. The ion $[M+H-H_2O]^+$ was not observed for **10**, also supporting the lack of the hydroxyl group on Trp since all known cyclomarins produce this characteristic ionization due to elimination of water from this position.[34] A new peak with a mass and retention appropriate for **2** was observed in the cymW$^-$ mutant, a compound not previously observed in *S. arenicola* CNS-205 but previously described in *Streptomyces* sp. CNB-982.[1] This mutant did not produce **1, 3-4,** nor **5-6,** all of which contain δ-hydroxyleucine.

Comparative metabolite profiling of *S. arenicola* CNS-205 wild type and *cym* mutants revealed the presence of two additional cyclomarazine analogs, cyclomarazine C (**8**) and D (**9**), containing the epoxide functionality characteristic of **1**. These compounds are not present in the *cymD*$^-$ or *cymW*$^-$ mutants, but surprisingly were still produced in the epoxidase mutant *cymV*$^-$. Both **8** (ESI-MS (m/z) 436.2 $[M + Na]^+$, 436.2 calcd) and **9** (ESI-MS (m/z) 422.2 $[M + Na]^+$, 422.2 calcd) eluted as two peaks with equivalent peak areas and MS spectra, indicative of a diastereomeric pair of compounds. Repeated rounds of preparative chromatography enriched the purity of the diastereomers of **8**, providing **8a** and **8b**. Full characterization of **8a** via interpretation of ^{1}H, HMBC and HSQC NMR data (Table 2.4) confirmed the presence of the epoxide functionality between C13 and C14. The ^{1}H chemical shift values for the 1,1-dimethyl-2,3-epoxypropyl functionality of **8a** [(δ_H, assignment), (3.25,13), (2.86,14), (1.46,15), (1.61,16)] compared favorably with those reported[1] for **1** [(δ_H, assignment), (3.22,13), (2.90,14), (1.57,15), (1.66,16)]. Similarly, the ^{13}C values for **8a** [(δ_C, assignment),

(57.6,12), (58.8,13), (44.7,14), (22.8,15), (23.8,16)] were similar to those for **1** [(δ_C, assignment), (58.1,12), (57.7,13), (45.4,14), (23.1,15), (24.4,16)] Comparison of **8a** with **8b** via comparison of ^{1}H-NMR data confirmed subtle differences in the chemical shifts for H4, H7, H8, and H13-16, validating their diastereomeric relationship with an inversion of stereochemistry at C13. The mass spectrum and reduced chromatographic retention time of **9** is consistent with an analog of **8** lacking the *N*-methylation on the δ-hydroxyleucine residue, although **9** was produced in quantities that impeded full structural analysis. The presence of the epoxides **8-9** in the *cymV*$^-$ deficient mutant suggests that olefin oxidation in the dipeptide series is a non-enzymatic process, or is a side reaction of an unrelated oxidase. This assumption is supported by the presence of both diastereomers of **8-9** in equivalent quantities.

2.4: Discussion

In this work we employed a multidisciplinary approach involving natural product discovery, stable isotope incorporation studies, genomics, and in vivo mutagenesis to unravel the biosynthesis of the related cyclomarazine dipeptides and the cyclomarin heptapeptides from two unrelated marine actinomycetes. These complementary experiments revealed an intriguing biosynthetic mechanism for nonribosomal peptides involving new pathways to nonproteinogenic amino acid residues as well as the in vivo formation of different size cyclic peptide products from the same NRPS assembly line.

Table 2.4 NMR Spectroscopic Data for Cyclomarazine C (8) in DMSO-d_6

	cyclomarazine C (**8a**)			cyclomarazine C (**8b**)
position	δ_C[a]	δ_H[b], (*J* in Hz)	HMBC	δ_H[b,c], (*J* in Hz)
1	165.6			
2	56.0	4.14, d (3.4)	1, 3, 5, 17	
NH-2		8.24, d (2.9)		
3a	30.1	3.15, dd (14.4, 5.4)	1, 2, 4, 5, 6	
3b	30.1	3.05, dd (14.4, 4.2)	1, 2, 4, 5, 6	
4	125.21	7.18, s	2, 3, 5, 6, 7, 10, 11, 12	7.19, s
5	107.5			
6	129.7			
7	119.1	7.54, d (7.8)	5, 6, 9, 10, 11	7.54, d (7.3)
8	118.6	7,00, t (7.4)	5, 6, 9, 10, 11	
9	120.8	7.07, t (7.6)	6, 7, 10, 11	
10	112.8	7.70, d (8.4)	6, 8	7.71, d (8.5)
11	135.0			
12	57.6			
13	58.8	3.25, m	12, 14, 15, 16	3.27, m
14	44.7	2.86, s	12	2.89, s
15	22.8	1.46, s	12, 16	1.45, s
16	23.8	1.61, s	12, 15	1.62, s
17	166.8			
18	58.8	3.62, dd (9.0, 3.0)	1, 17, 19, 20	
NMe-18	31.9	2.73, s	1, 18	
19a	36.0	0.71, m	17, 18, 20, 21, 22	
19b	36.0	0.55, ddd (13.7,10.4, 5.9)	17, 18, 20, 21, 22	
20	32.4	1.39, m		
21a	66.0	2.83, m	19, 20, 22	
21b	66.0	2.75, m	19, 20, 22	
OH-21		4.29, m	20, 21	
22	16.2	0.61, d (6.6)	19, 20, 21	

[a]Assignment by HMQC and HMBC correlations. [b]600 MHz.
[c]Only shifts that differentiate the two stereoisomers listed here

Analysis of the structural features of the cyclomarin and cyclomarazine analogs along with analysis of stable isotope incorporation patterns into **1** from *S*.sp. CNB-982 were the starting point for developing biosynthetic proposals that

were further evaluated genetically upon completion of the genome sequence of *S. arenicola* CNS-205. Bioinformatic and in vivo mutagenesis experiments correlated the cym locus to cyclomarin and cyclomarazine biosynthesis, thereby providing the molecular basis for the assembly of their amino acid components and their cyclic structures.

The molecular architecture of the NRPS megasynthetase CymA perfectly correlates with the synthesis of the cyclomarin heptapeptides starting with the tryptophan residue and ending with ADH. The majority of the cyclomarin building blocks are modified proteinogenic amino acids that have undergone methylation, oxidation, and prenylation reactions catalyzed by tailoring enzymes encoded by genes flanking the massive 23 kb *cymA* gene. Tryptophan, which is the priming amino acid residue, undergoes a series of modification events involving reverse *N*-prenylation by CymD, β-hydroxylation by CymS, and epoxidation by CymV to yield *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-hydroxytryptophan as in **1**. The timing of these biosynthetic events relative to peptide synthesis was probed in vivo, suggesting that prenylation and β-hydroxylation occur either on the free amino acid or while it is covalently tethered via a thioester linkage to CymA, whereas epoxidation occurs post NRPS assembly. Precedence for the prenylation reaction, however, suggested that the reaction would rather occur post NRPS assembly based on the biosynthesis of other prenylated NRPS products such as terrequinone A[39] and lyngbyatoxin.[15] Gene inactivation of the PTase *cymD* eliminated the production of the natural cyclomarins and cyclomarazines in *S. arenicola*, thereby establishing the biosynthetic linkage of these differently sized

peptides. Furthermore, production of desprenylcyclomarin C (**7**) at a substantially lower concentration in the mutant relative to the corresponding cyclomarin and cyclomarazines in the wild-type organism suggested that prenylation is an integral biosynthetic reaction and that it must occur prior to the stage of the diketide intermediate common to the cyclomarazines and cyclomarins. Chapter 3 reports further vivo and in vitro characterization of CymD, which confirms that prenylation occurs on free Trp before incorporation by CymA. In contrast to the low production of **7** in the PTase mutant, inactivation of the cytochrome P450 epoxidase *cymV* resulted in a mutant that efficiently shifted cyclomarin production from the epoxide **1** to the olefin **3**, which indicates that the liberated **3** is the substrate of CymV.

Intriguingly, the co-occurrence of the two cyclic peptide series derived from the same NRPS system is to the best of our knowledge unprecedented in nature. The key structural difference between the common amino acid residues in these peptides is the oxidation state of the β-position of the tryptophan moiety, which is oxidized in the heptapeptides and unoxidized in the diketide DKPs. DKPs have been reported to derive from NRPSs as well as NRPS-independent systems.[40] However, the majority of those deriving from NRPSs harbor a proline residue at the second amino acid position and are proposed to form spontaneously resulting from conformational constraints induced by the proline residue.[41,42] Few examples exist contrary to the necessity of proline for DKP formation by NRPSs, and in those cases,[43,44] the associated NRPS is programmed to produce a diketide. Whether formation of the cyclomarazine

DKPs is facilitated by the type II TE CymQ serving an editing function to hydrolyze the incompletely processed dipeptide or its cleavage is nonenzymatic due to ineffective further processing by module 3 of the NRPS CymA is not clear (Figure 2.5). Since inactivation of CymQ had little effect on the production ratio of the di- and heptapeptides (Figure 2.7), it is unlikely that it is responsible for catalyzing DKP formation dependent on the oxidation state of Trp. In addition, inactivation of the tryptophan β-hydroxylase did not increase cyclomarazine production and did provide a putative cyclomarin lacking this oxidation. These results support, but do not confirm, that partitioning to the diketide is nonenzymatic. Surprisingly, a β-oxidized cyclomarazine has not been observed in any wild type or mutant extract.

A further unusual feature identified in this study involves the unexpected formation of the nonproteinogenic amino acid residue ADH. While our initial isotope labeling experiments led us astray in the formulation of a biosynthetic model, bioinformatic analysis of the *cym* locus suggested an alternative pathway related to the degradation of 3-HPP, yet operating in the reverse direction (Figure 2.6). The acquisition and evolution of the ADH biosynthetic pathway represents an intriguing example of pathway creation, which is being further explored biochemically, the results of which are reported in chapter 4.

2.5: Experimental Section

2.5.1: General experimental procedures

Low-resolution LC/MS data were acquired using a Hewlett-Packard series 1100 LC/MS system utilizing electrospray ionization in positive mode, with a

linear gradient of 10–90% aqueous MeCN at a flow rate of 0.7 mL/min over 24 min on a reversed-phase C_{18} column (Phenomenex Luna, 4.6 mm x 100 mm, 5 µm) for general extract screening. High-resolution mass spectra were collected by HR-ESI-TOFMS at the Scripps Research Institute, La Jolla, CA and ESI-HR-FTMS at the Chemistry & Biochemistry Mass Spectrometry Facility, University of California San Diego. NMR spectral data was obtained on a Bruker 600 MHz spectrometer equipped with a 1.7 mm cryoprobe.

2.5.2: Bacterial strains, plasmids, culture conditions, and DNA manipulations

The marine actinomycete strain CNS-205 was isolated from a sediment sample collected at a depth of 20 m off Palau in 2004 by the Fenical and Jensen laboratories at Scripps Institution of Oceanography, UCSD. The strain was identified as *S. arenicola* by the Fenical and Jensen laboratories based on 16S rDNA analysis (100% identity; NC_009953) and DNA–DNA hybridization studies with the *S. arenicola* type strain CNH-643.[22] *Streptomyces* sp. CNB-982 was previously described.[1] Both bacteria were grown in A1 media (10 g starch, 4 g yeast extract, 2 g peptone, and 10 mL of 1 M Tris (pH 8.0) per liter seawater) at 28 °C and 200 rpm unless otherwise noted. *Escherichia coli* DH5α (Invitrogen) was used for cloning experiments as described,[45] and *E. coli* S17-1[46] and ET12567/pUZ8002[47] were used for conjugation experiments. Previously adapted for use in *Salinispora*,[33] REDIRECT® (Plant Bioscience Limited, Norwich, UK) technology was utilized for PCR targeting.[47] The λ-Red function was provided by pKD20 (bla).[46] Fosmid BPPW9569 containing a fragment of the *cym* gene cluster

from *cymS* and a transposase directly downstream of *cymH* in the pCC1FOS vector (Epicentre) was obtained from the Joint Genome Institute, Walnut Creek, CA and utilized for both *cymD* and *cymV* deletions. Apramycin (100 µg/mL for *S. arenicola*; 50 µg/mL for *E. coli*), chloramphenicol (2.5 µg/mL for *S. arenicola*; 12.5-25.0 µg/mL for *E. coli*), carbenicillin (100 µg/mL), and nalidixic acid (100 µg/mL) were used for selection of recombinants. DNA purification and manipulation was performed according to standard procedures.[45,48]

2.5.3: Cultivation of *S arenicola* and isolation of 1, 8, and 9

S. arenicola CNS-205 was cultured at 27 °C with shaking at 250 rpm in 10 2.8-L Fernbach flasks containing 1 L of the medium A1+BFe (10 g starch, 4 g yeast extract, 2 g peptone, 40 mg $Fe_2(SO_4)_3$•$4H_2O$, 100 mg KBr, and 1 L seawater). After 10 days organic chemical constituents from 10 L culture were collected by solid phase extraction using Amberlite XAD-7 resin. The crude extract was prepared by washing the resin with acetone and evaporating the solvent in vacuo. The pH of the remaining aqueous residue was adjusted to pH ~12 with NaOH and partitioned with EtOAc. The EtOAc layer was dried under vacuum to yield ~1 g of crude material, which was fractionated utilizing Sephadex LH-20 resin (GE Healthcare, 35 g resin in a 30 mm wide column) with MeOH as the solvent. Fractions were screened by LC-MS, and those containing cyclomarin or cyclomarazines were pooled. The cyclomarin pool was injected on a Phenomenex Synergi 10µm 250 x 21.2 mm column with a gradient of 48-90% MeCN in H_2O over 30 min at a flow rate of 13 mL/min. Cyclomarin A was eluted at 20.6 min. Cyclomarin A was further purified utilizing a Phenomenex Luna

C_{18}(2) semi-preparative column C_{18} 10 mm x 250 mm column at 2.5 mL/min) using a isocratic solvent of 57.5% MeCN in H_2O. 3.2 mg of **1** eluted at 28 min from WT *S. arenicola* CNS-205 per liter of culture.

To purify cyclomarazine C and D, the cyclomarazine fraction from above was injected on a Phenomenex Synergi 10µm 250 x 21.2 mm column with a gradient of 20-40% MeCN in H_2O over 30 min at a flow rate of 13 mL/min. Cyclomarazine C and D coeluted at 22.5 min. The fraction containing **8** and **9** were injected on a Phenomenex Onyx C_{18} 10 x 100 mm column and separated with 15% MeCN in H_2O at a flow rate of 3.5 mL/min. Compounds **8** and **9** both eluted as a pair of peaks that could not be resolved to baseline (Appendix figure 2.S2), suggesting that they are diasterioisomers, with **9a** and **9b** eluting at 9.5 and 10.5 min, and **8a** and **8b** eluting at 37.5 min and 39.7 min, respectively. Repetitive rounds of chromatography were utilized to separate the stereoisomers of **8a** and **8b**, yielding 2.1 mg of **8a**, and 0.6 mg of **8b.** Note that these compounds were of equal peak area in the first injection and the yield of **8b** was low because of experimental error.

Cyclomarazine C (8a). NMR data see Table 2.1. MS (ESI^+) 436.2 *m/z* $[M+Na]^+$ (calcd for $C_{21}H_{25}N_3O_4Na$ 436.2).

Cyclomarazine C (8b). NMR data see Table 2.1. MS (ESI^+) 436.2 *m/z* $[M+Na]^+$ (calcd for $C_{21}H_{25}N_3O_4Na$ 436.2).

Cyclomarazine D (9a&b). MS (ESI^+) 422.2 *m/z* $[M+Na]^+$ (calcd for $C_{20}H_{23}N_3O_3Na$ 422.2).

2.5.4: Labeling experiments

Initial feeding experiments were carried out in cultures of *Streptomyces* sp. CNB-982 and isolation of the resulting labeled cyclomarin A was carried out by J. Carney, R. T. Williamson and S. Gould.[4] Production cultures in A1 media were incubated in Fernbach flasks for 6-7 d at 230 rpm and 23 ºC. A mixture of 1 g of D-[U-$^{13}C_6$]glucose and 1 g D-glucose was pulse-fed at 48, 62, 73, 88, and 96 h to a 1 L production culture. Pulse-feeding [methyl-^{13}C]methionine (52.5 mg) at 48, 60, 72, 85, and 94 h and DL-[3'-^{13}C]tryptophan (90.1 mg) at 40, 52, 64, 76, and 88 h to two separate 1L production cultures was also carried out. An additional feeding experiment was carried out with cultures of *S. arenicola* CNS-205. Production cultures consisting of A1+BFe media (1 L) were incubated in Fernbach flasks for 10 d at 230 rpm and 28 °C. Sodium [1-^{13}C]isobutyrate[49] (25 mg) was fed at 60 h.

2.5.5: DNA sequence analysis of the cyclomarin gene cluster (*cym*)

The putative cyclomarin NRPS *cymA* was identified from a draft genome assembly of *S. arenicola* CNS-205 (GenBank accession CP00850) provided by the Joint Genome Institute (JGI) through the Community Sequencing Program. NRPS domain structure for *cymA* was determined as described,[50] and adenylation domain specificity was predicted using NRPSPredictor.[51] Sequence analysis and putative functions for enzymes encoded by genes surrounding *cymA* were predicted using the suite of tools available at NBCI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). Open reading frames were further analyzed by FramePlot.[52]

2.5.6: Conjugation protocol for *S. arenicola*

S. arenicola CNS-205 was cultivated for 5 days in 50 mL A1 in a 250 mL flask containing a stainless steel spring. Mycelium was harvested from 25 mL and resuspended in 2 mL A1. 0.5 mL of *E. coli* S17-1 suspension[48] was mixed with 0.5 mL of the *S. arenicola* suspension and then plated on A1 agar plates. After 1 d incubation at 34 °C, the plates were overlaid with 2.5 mg each apramycin and nalidixic acid. Exconjugates were visible 10 days after plate inoculation.

2.5.7: Inactivation of *cymD, cymQ, cymS, cymV* and *cym W*

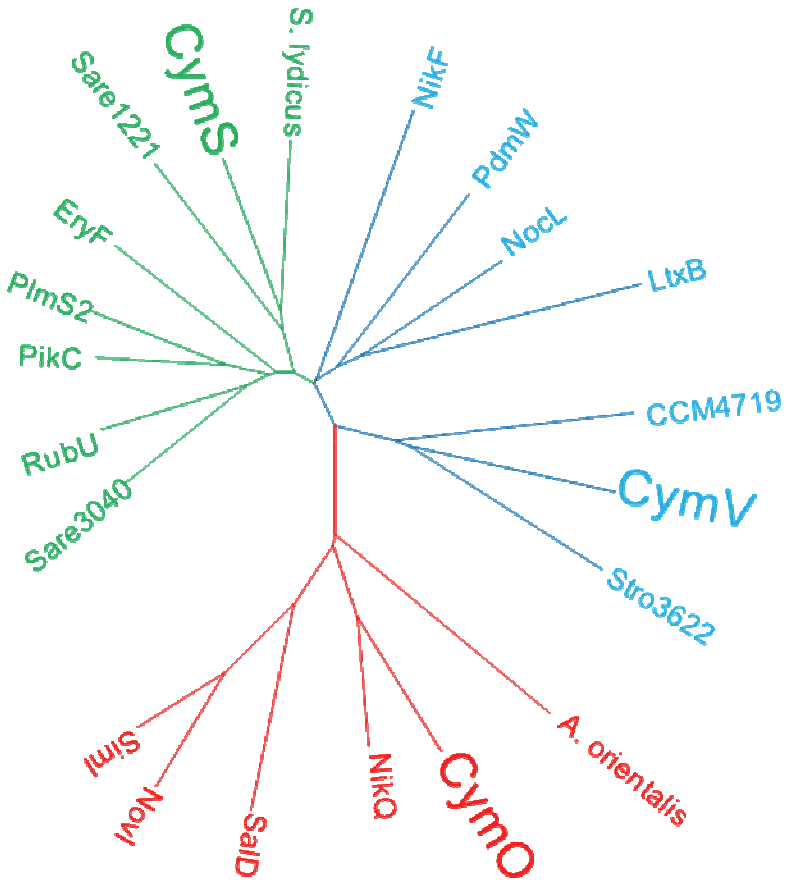
cymD was inactivated utilizing the modified PCR targeting system previously described for *S. tropica*.[33] In fosmid BPPW9569, *cymD* was replaced by the apramycin resistance (*acc(3)IV*) cassette from pIJ773 utilizing λ-Red-mediated recombination. The *cymD* replacement cassette was generated by PCR using primers cymDkoF1 and cymDkoR1 (See appendix table 2.S1). Lowercase letters represent 39 nt homologuous extensions immediately upstream and downstream of *cymD*, respectively, including the putative start and stop codons. This cassette was introduced into *E. coli* BW25113/pKD20[46] containing the chloramphenicol-resistant fosmid BPPW9569. Gene replacement was verified using PCR primers cymDckF1 and cymDckR1. The mutated fosmid was introduced into *S. arenicola* CNS-205 by conjugation from *E. coli* S17-1 and confirmed by PCR analysis. Identical methodology was utilized to inactivate *cymQ, cymS, cymV* and *cym W* utilizing primers listed in appendix table 2.S1.

2.5.8: Analysis of the *S. arenicola cym* mutants

One hundred µL of 4 d cultures of the *S. arenicola* WT or mutants were used to inoculate duplicate 50 mL cultures in A1+BFe containing 100 µg/mL apramycin (muntants only) in 250 mL Erlenmeyer flasks containing stainless steel springs. On day 10, the cultures were extracted with EtOAc, dried in vacuo and combined with ~40 mg Celite. The sample and Celite were loaded on a 1g C_{18} solid phase extraction column (Thermo) and eluted with a step gradient of 20, 40, 60, 80, and 100% MeCN in H_2O (6 mL steps). Eighty % MeOH fractions were dried and reconstituted in 700 µL MeOH and analyzed for cyclomarazine analogs with a gradient of 20-40% MeCN over 30 min. One Hundred % MeOH fractions were dried and reconstituted in 100 µL and analyzed for cyclomarin analogs, utilizing isocratic 57.5% MeCN in H_2O. For both analyses, a reversed-phase C_{18} column (Phenomenex Luna, 4.6 mm x 100 mm, 5 µm) at 0.7 ml/min was used. Larger scale fermentations of the *cymV*$^-$ (1 L) and *cymD*$^-$ (10 L) mutants were cultured, extracted, and the products isolated as described above. The *S. arenicola cymV* knockout mutant yielded 2.1 mg of **3**, whereas the *cymD*$^-$ mutant gave 0.5 mg of a crude fraction containing **7** (HR-FTMS $[M+Na]^+$ *m/z* 981.5408, calcd 981.5420).

2.6: Appendix

2.6.1: Phylogenetic analysis of the cytochrome P450s CymO, CymS, CymV to predict their role in cyclomarin and cyclomarazine biosynthesis



Appendix figure 2.S1: Neighbor joining phylogenetic tree of CymO, CymS, and CymV

Deduced protein sequences for putative cytochrome P450s (CYP) from the cyclomarin (*cym*) cluster from *Salinaspora arenicola* CNS-205 were blasted against the NCBI non-redundant (nr) database using the blastp algorithm.[53] BLAST hits with E values less than 10^{-50} in which the function of the sequence has been characterized or can be deduced were utilized. A multiple sequence

alignment of the CYP protein sequences was carried out using an implementation of clustalW algorithms in Molecular Evolutionary Genetics Analysis (MEGA) version 3.1[54]. The Gonnet protein weight matrix was utilized with GOP at 15 and GEP at 0.3 for both pairwise and multiple alignment.[55] After the first round of alignment, overhanging N and C termini were deleted, and GOP and GEP was set at 10 and 0.2 for the final round of alignment.[55,56] Alignment data from the CYP sequences was also fed into the neighbor-joining (NJ) algorithim (figure 3) including a 1000 repetition bootstrap analysis (figure 4). Default parameters included complete deletion of gaps with the Poisson correction and the assumption of uniform rates of mutation. The MEGA tree visualizer, part of MEGA 3.1, was utilized for displaying and editing trees.

Figure 2.S1 depicts the unrooted NJ phylogenetic tree derived from this analys of the CYP protein sequence data. The red clade contains members involved in β oxidation of peptidyl carrier protein tethered amino acids. NikQ[57] converts histidine to β-hydroxy-histidine. SimI[58], NovI[59], and the CYP from *A. orientalis* (13) catalyze the conversion of tyrosine to β-hydroxy-tyrosine and SalD[60] catalyzes a similar reaction on CHA to β-OH-CHA. Because all of the P450s in the red clade are β-hydroxylating enzymes, we can deduce that CymO is a β-hydroxylating P450, likely to oxidize the phenylalanine or the *N*-(1,1-dimethylallyl)-tryptophan residue of cyclomarin.

CymV is a member of two clades labeled blue in Figure 2.S1, the members of which install epoxide or hydroxyl groups on aromatic carbons. It is possible that this enzyme is forming the epoxide moiety on the isoprene unit of

cyclomarin, since Stro3622 from *S. tropica*[50] is a component of an enediyne biosynthetic cluster. In the same clade is CCM4719, which may provide the epoxide needed to form the macolide structure of salinomycin[36]. LtxB is theorized to form an epoxidated cryptic intermediate in the biosynthesis of lyngbyatoxin[15] NocL[61] and NikF[62] are involved in installing hydroxyl groups in aromatic systems.

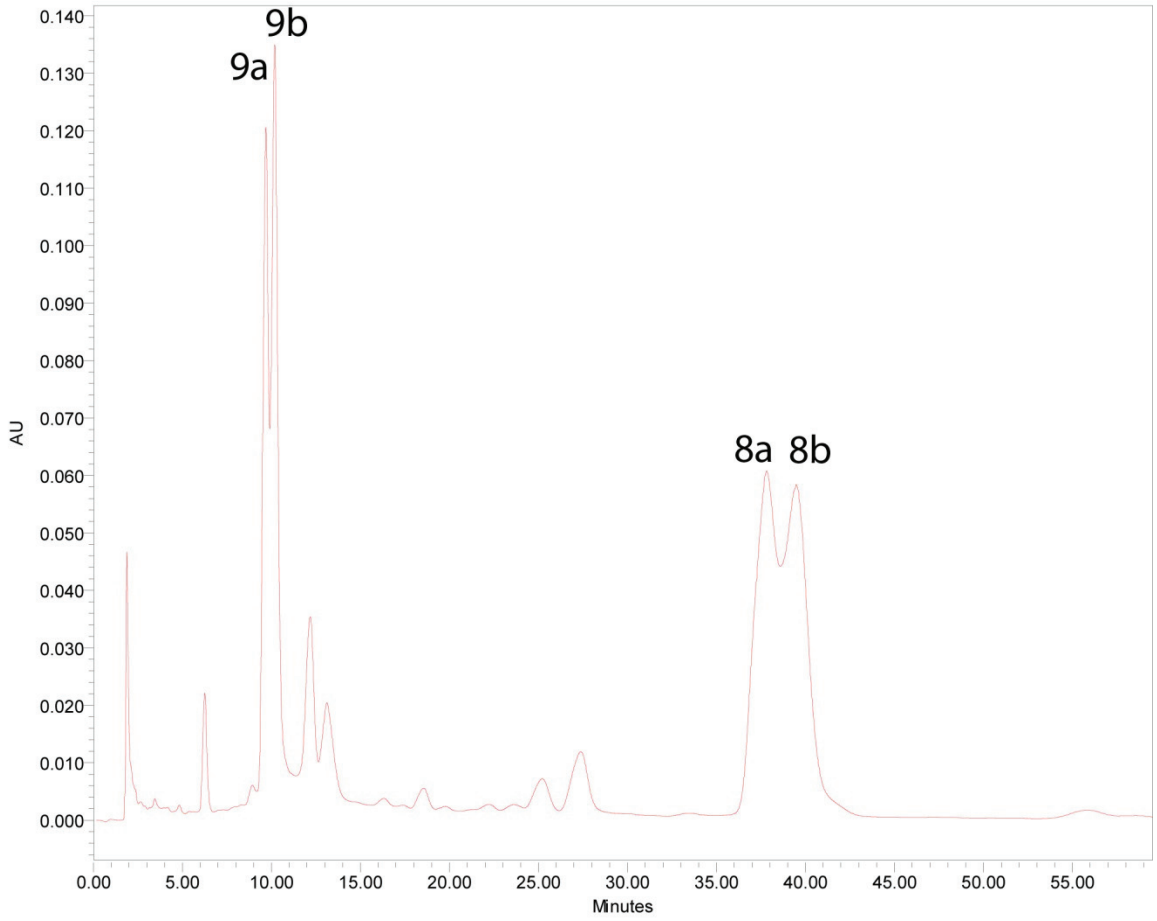
CymS may be involved in hydroxylating the isoleucine residue in cyclomarin A, or the β postion of the prenylated tryptophan residue. Most of the members of the green clade oxidize an aliphatic carbon. The *S. lydicus* derived P450 is known to convert avermectin to oxo-avermectin.[37] Sare1221 is involved in rifamycin production.[63] EryF hydroxylates an aliphatic position of the erythromycin aglycone,[64] while PikC also catalyzes the hydroxylation on the alkane side chain of pikromycin.[65] Phoslactomycin B is hydroxylated on an alkane ring via PlmS2.[66] RubU[67] and Sare3040[25] are also expected to oxidize at aliphatic positions.

Appendix table 2.S1: Primers for PCR directed gene inactivation

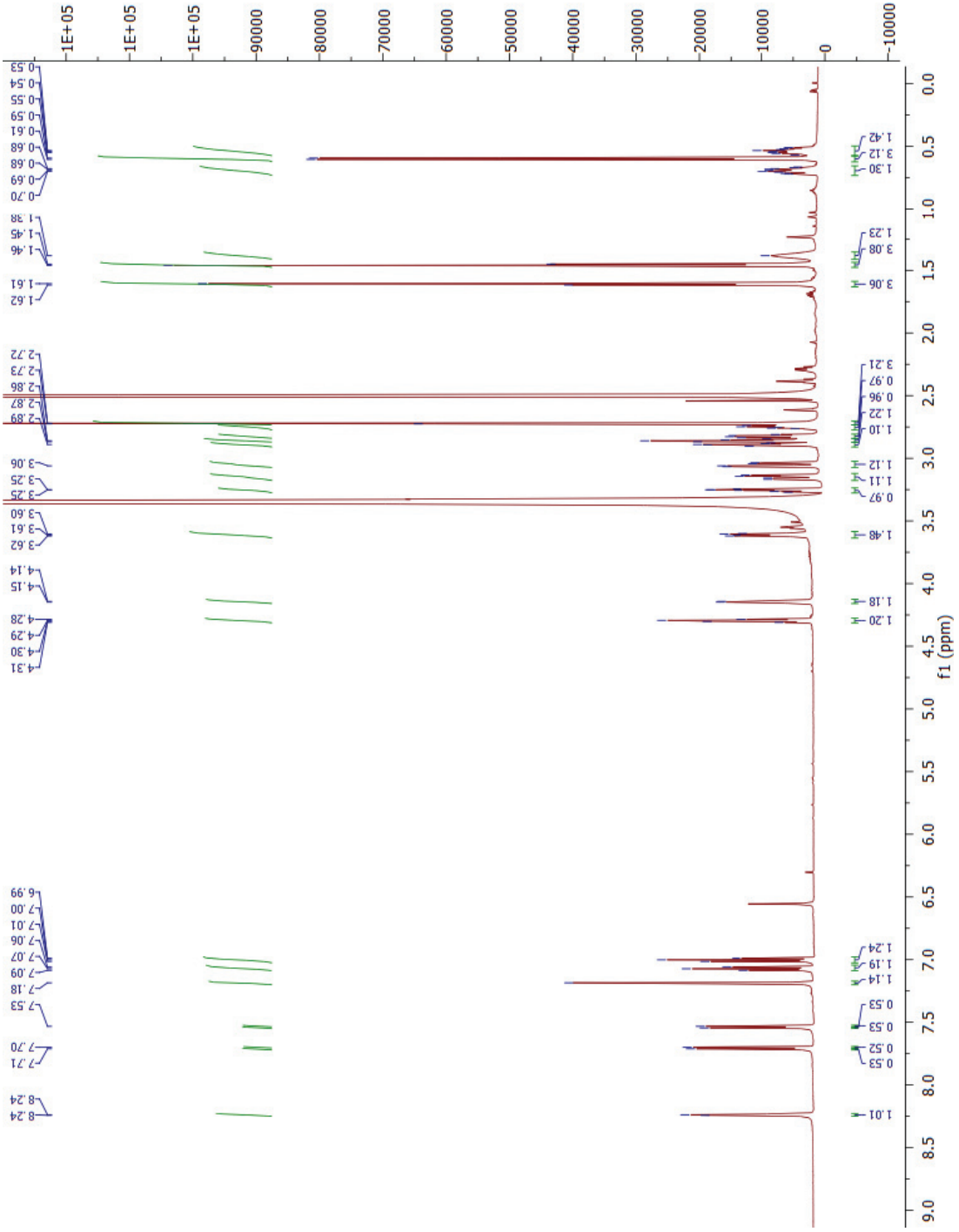
Replacement cassette primers[a,b]	
cymDKOF1	cgcgcctcgacgtggattctgacaaggagaaccgggttgATTCCGGGGATCCGTCGACC
cymDKOR1	gacgtcgtgcatctccggtgacaggtagaccttcatccgTGTAGGCTGGAGCTGCTTC
cymQKOF1	ctcgggggcagcccggcggcccgcacactggtctgcgtgATTCCGGGGATCCGTCGACC
cymQKOR1	gtccgcagtggtccgggcatagtcggcgaccaactcccgTGTAGGCTGGAGCTGCTTC
cymSKOF1	atatgatcggacagcgaatagaaattcgaggacgccatgATTCCGGGGATCCGTCGACC
cymSKOR1	gagcttgcgcatgccggatactgcgtgccattgcagctgTGTAGGCTGGAGCTGCTTC
cymVKOF1	cttccgacgacacgacgacggggcggcgtgagcggtatgATTCCGGGGATCCGTCGACC
cymVKOR1	gatcgcgtgacgtgcggcgaccggcgaggtgcccgatcaTGTAGGCTGGAGCTGCTTC
cymWKOF1	gaacagctcgcgagctaccacaacgacgggtacgtgctgATTCCGGGGATCCGTCGACC
cymWKOR1	gagcggggaggtgtcgcggccggccaggtgctccgcgcgTGTAGGCTGGAGCTGCTTC

Verification primers[a]	
cymDckF1	TGGTTCCCGGCCGTGATCGG
cymDckR1	CGGCGAACAGGACGCACGTC
cymQckF1	TCGCCGAGGTGCACGACGAA
cymQckR1	GCGAGCGCTGACCGTCGT
cymSckF1	ATGGGTGCCGGCAGCGTCG
cymSckR1	GGAGGAACCCGGCGAGGAC
cymVckF1	AGCGTGAGATCGGCCTAGGG
cymVckR1	TTGTGGTAGCTCGCGAGCTG
cymWckF1	CGGTCCGCGTTCCCTTCG
cymWckR1	CGACTCGTACAACCGGACGA

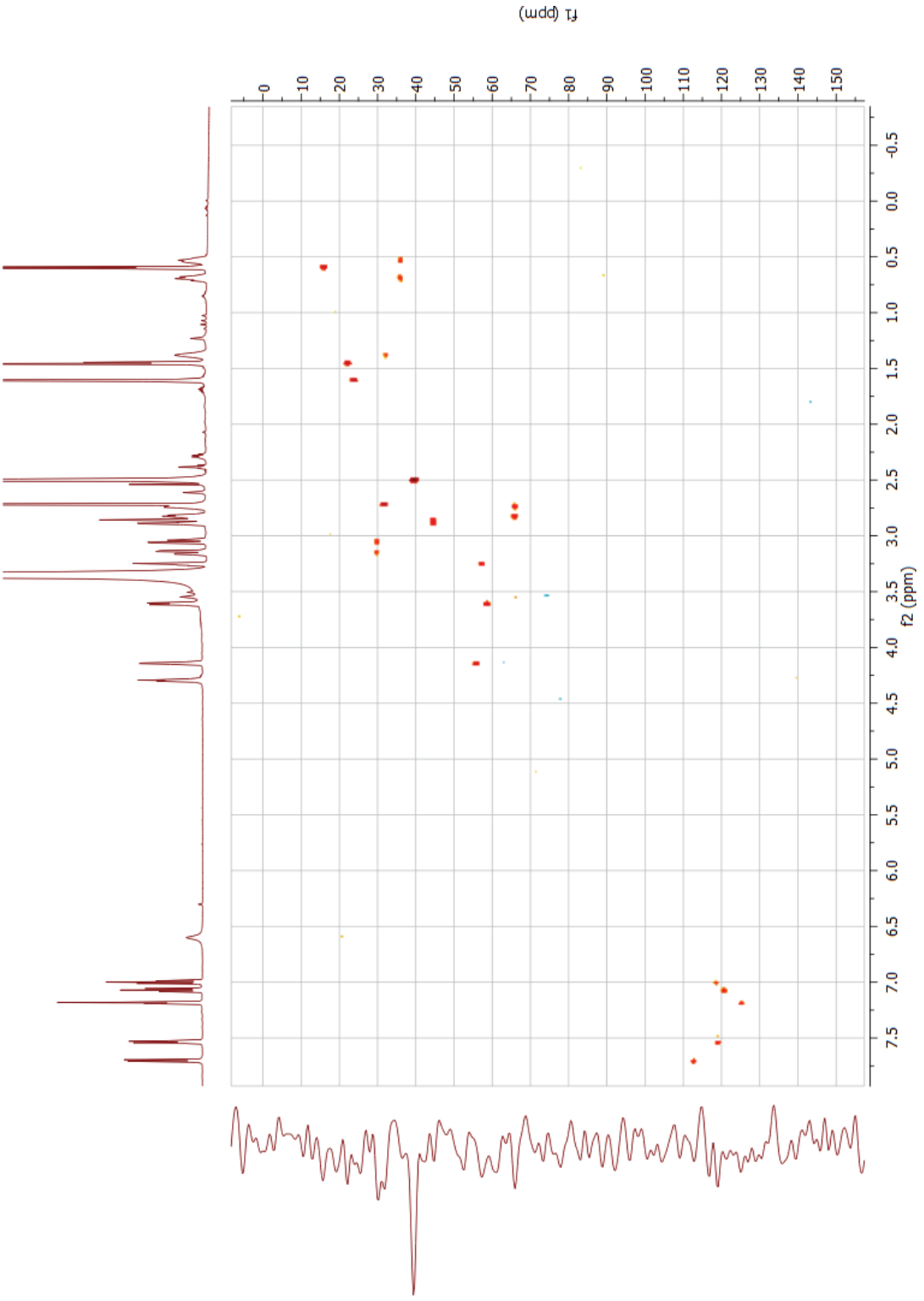
[a]All primers denoted 5' to 3'. [b]Capital letters denote sequence homologous to apramycin resistance cassette, lower case letters homologous to gene being replaced.



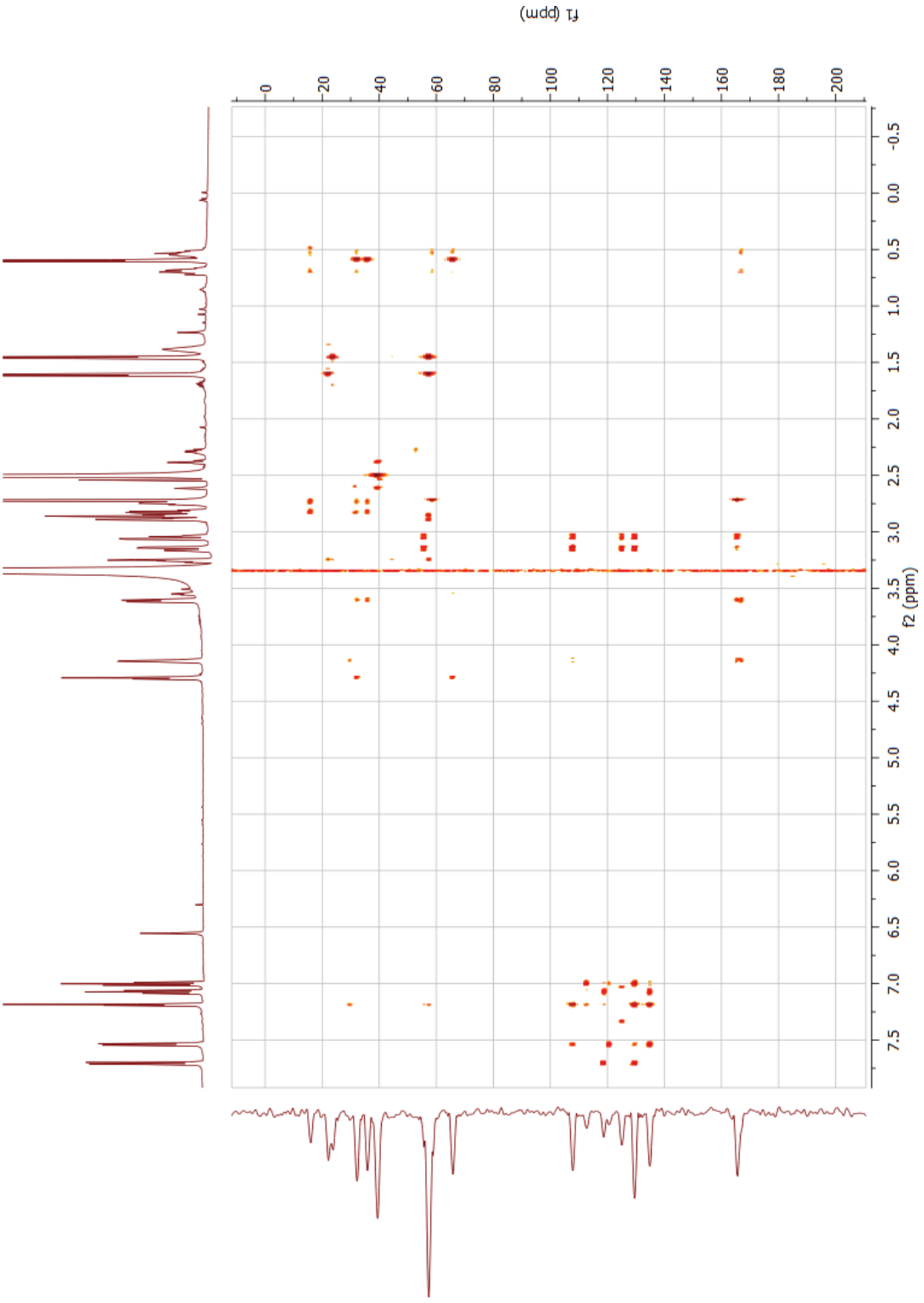
Appendix figure 2.S2: Preparative HPLC chromatogram of cyclomarazine C (8a,8b) and D (9a,9b)



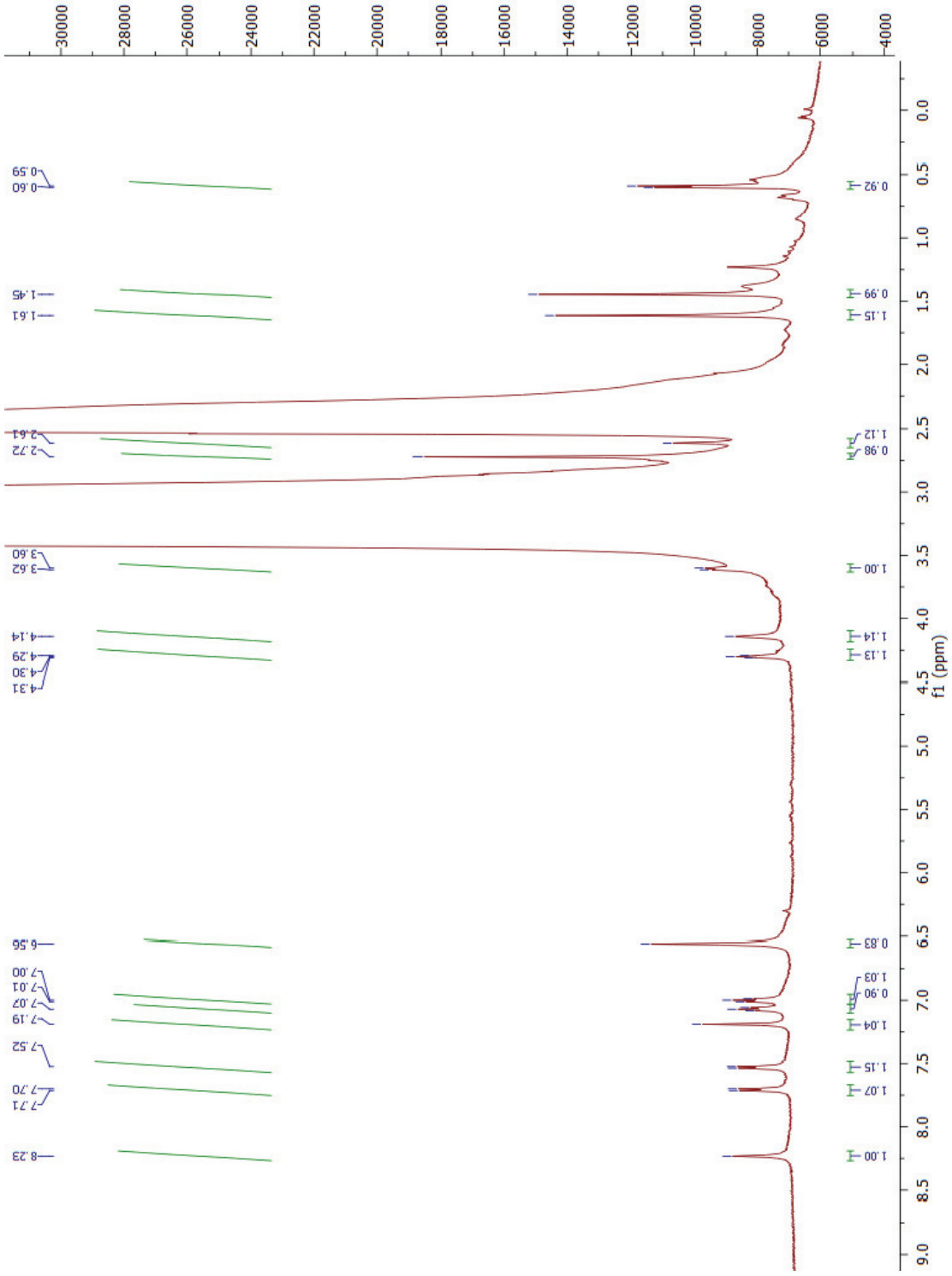
Appendix figure 2.S3: ^{1}H spectrum of cyclomarazine C (8a) in DMSO-d_6



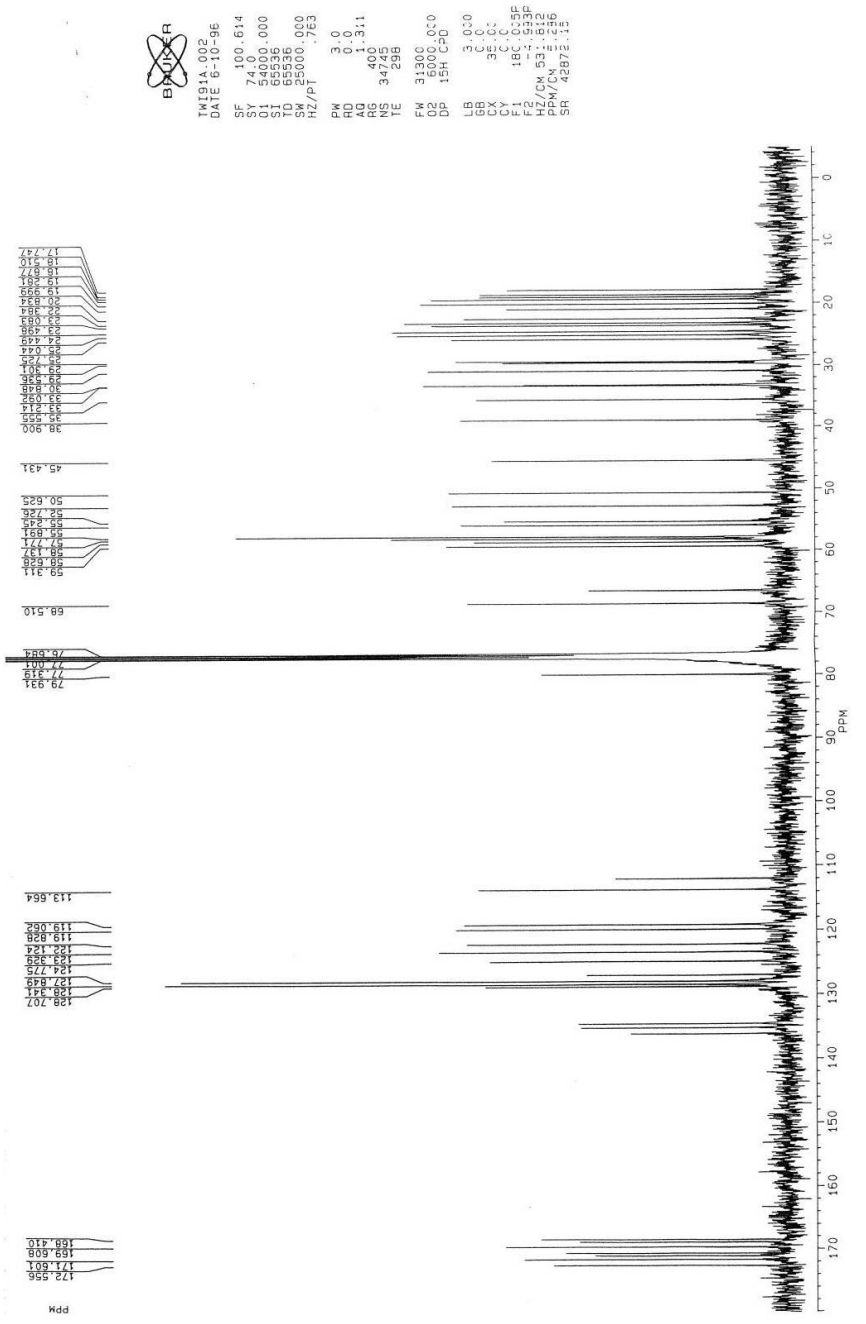
Appendix figure 2.S4: HSQC spectrum of cyclomarazine C (8a) in DMSO-d_6



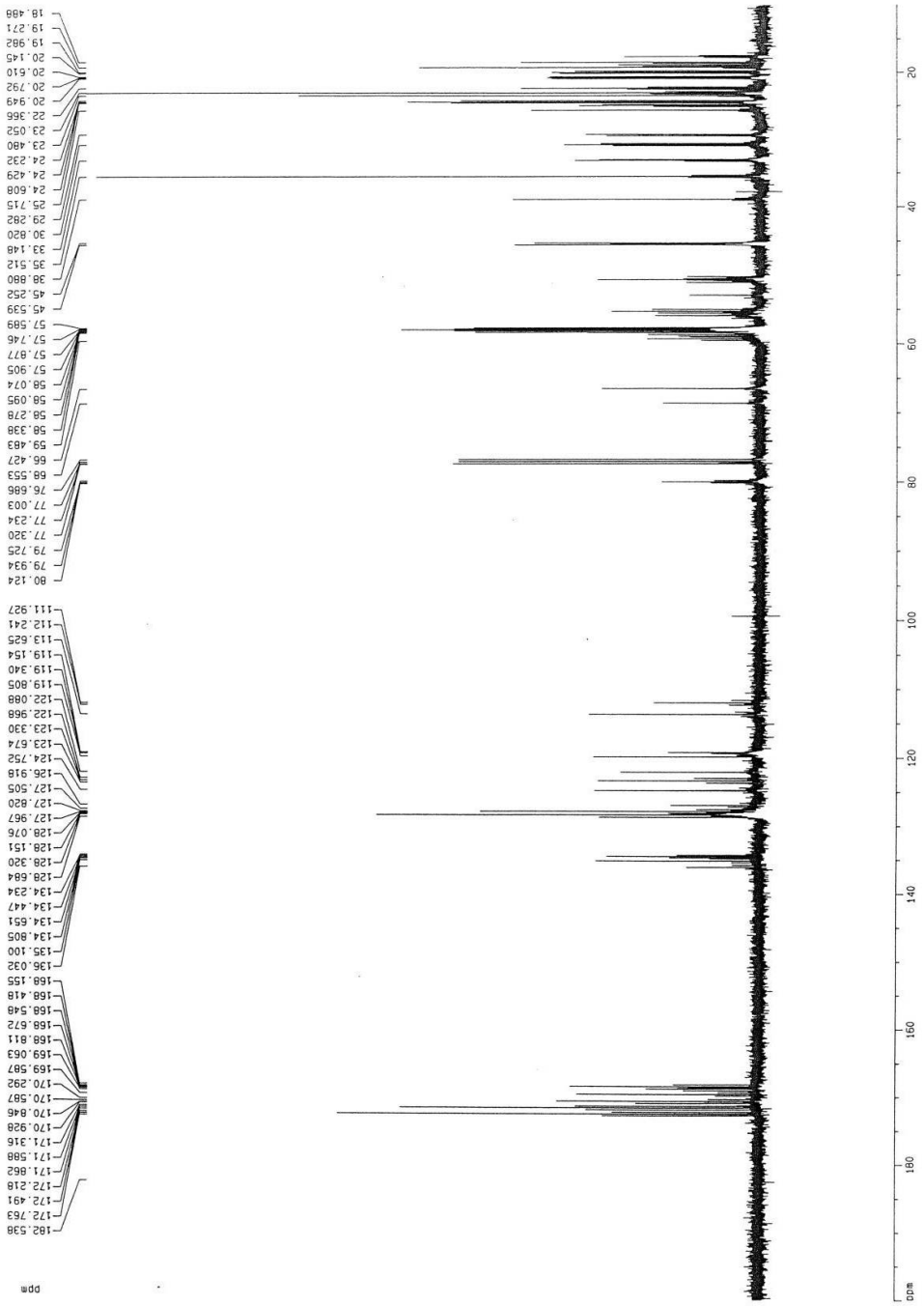
Appendix figure 2.S5: HMBC spectrum of cyclomarazine C (8a) in DMSO-d_6



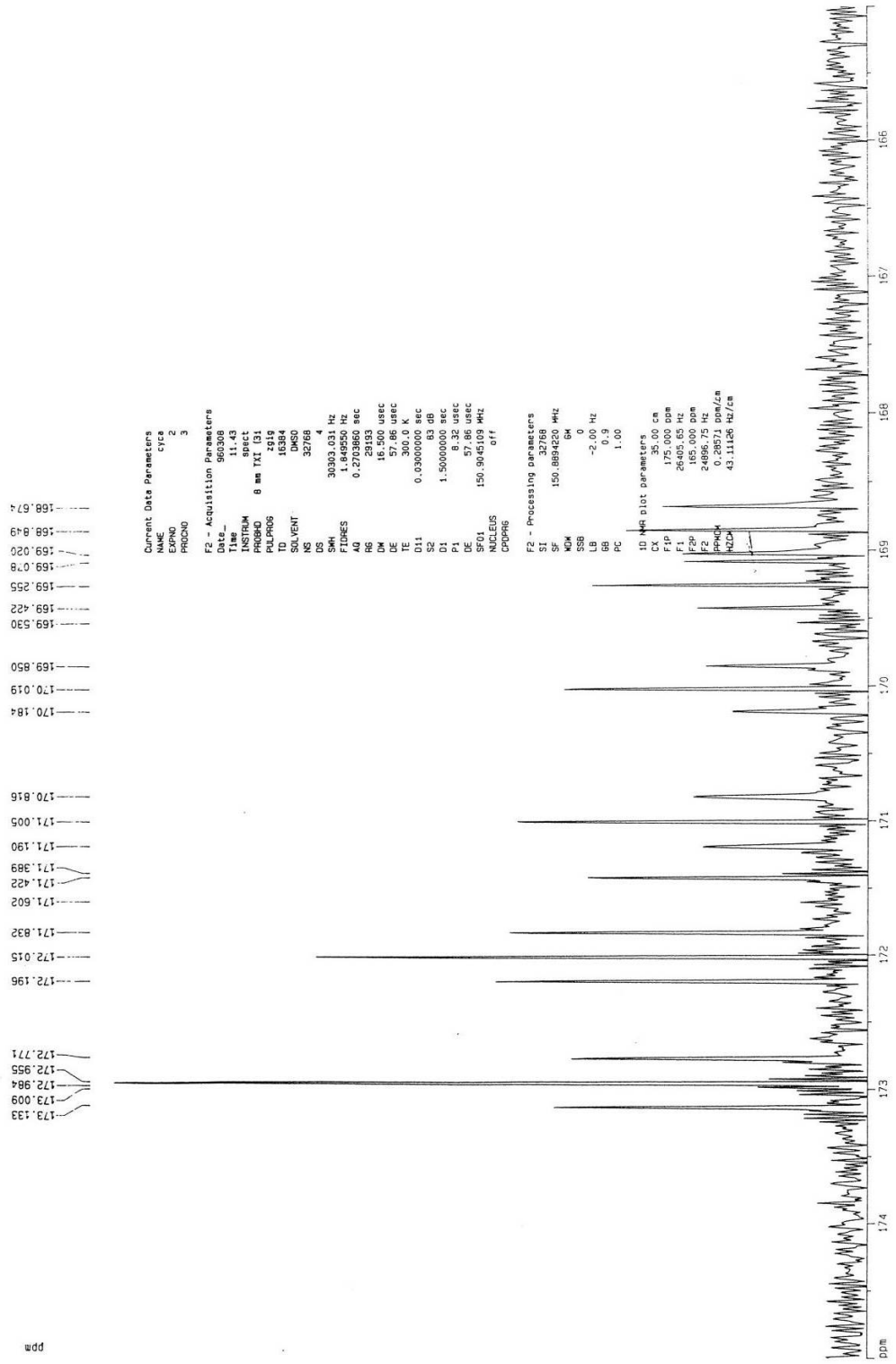
Appendix figure 2.S6: ^{1}H spectrum of cyclomarazine C (8b) in DMSO-d_6



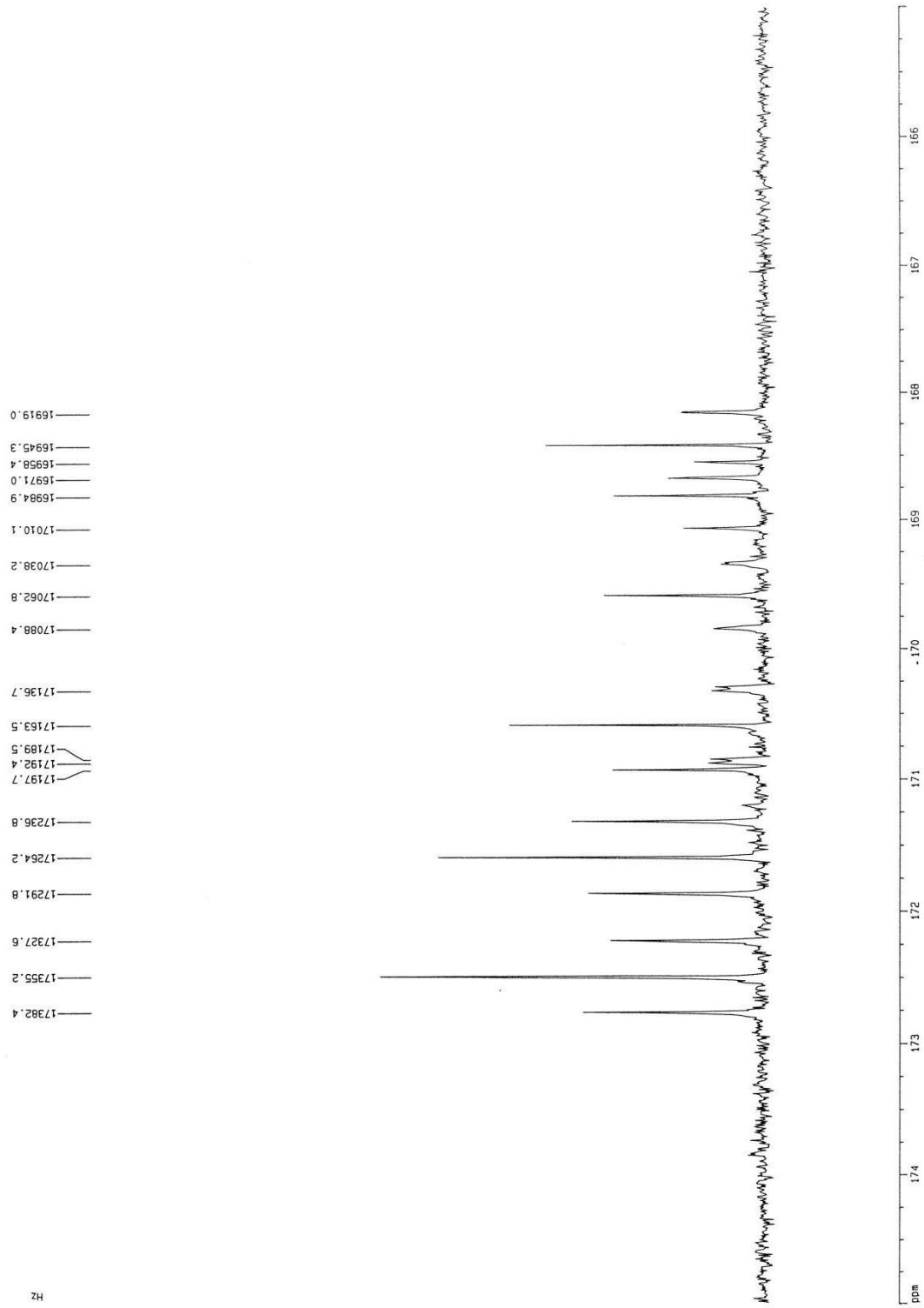
Appendix figure 2.S7: ^{13}C spectrum of natural abundance cyclomarin A (1) in $CDCl_3$



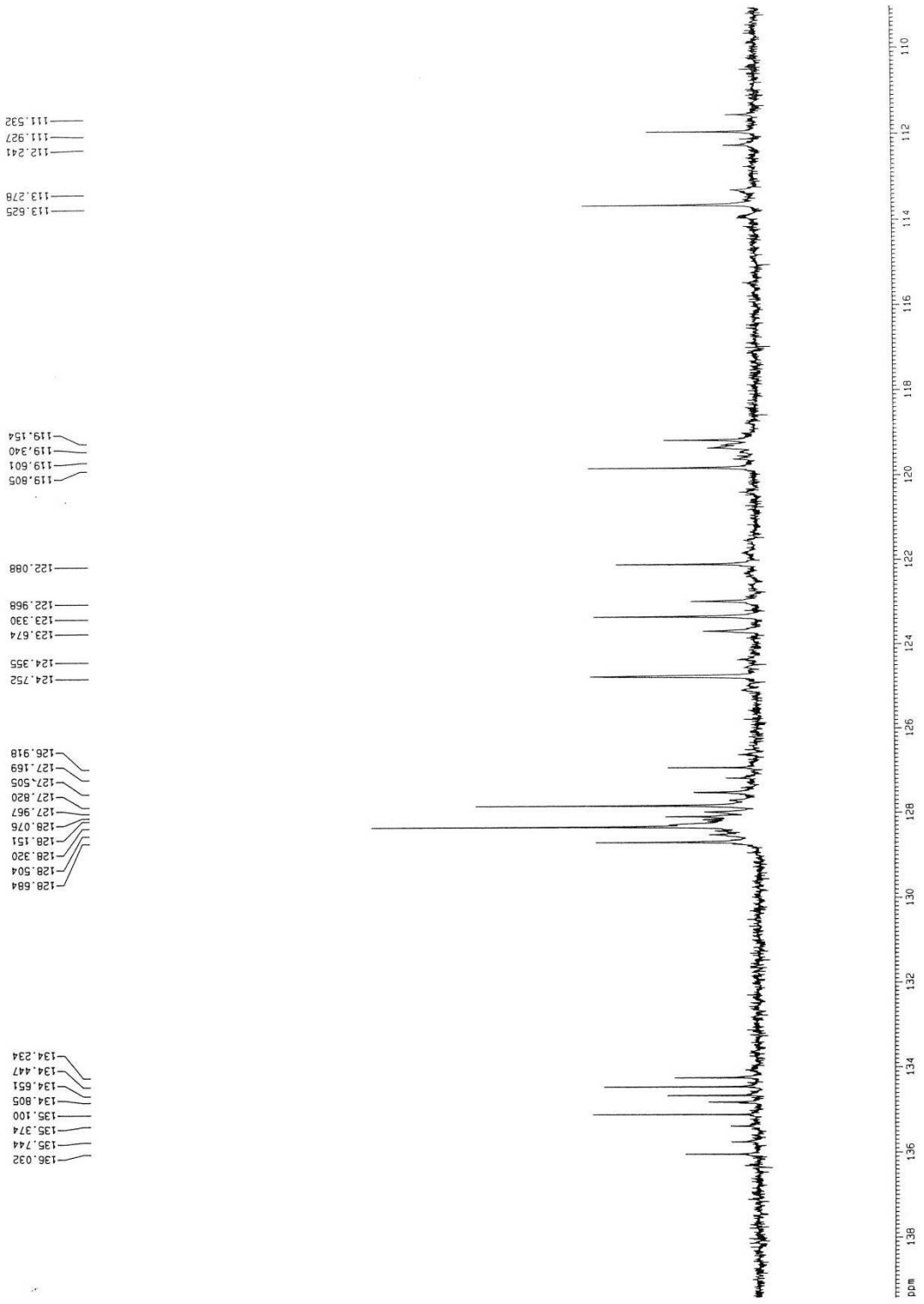
Appendix Figure 2.S8: ^{13}C spectrum of cyclomarin A from [U-$^{13}C_6$] glucose feeding (1a) in $CDCl_3$



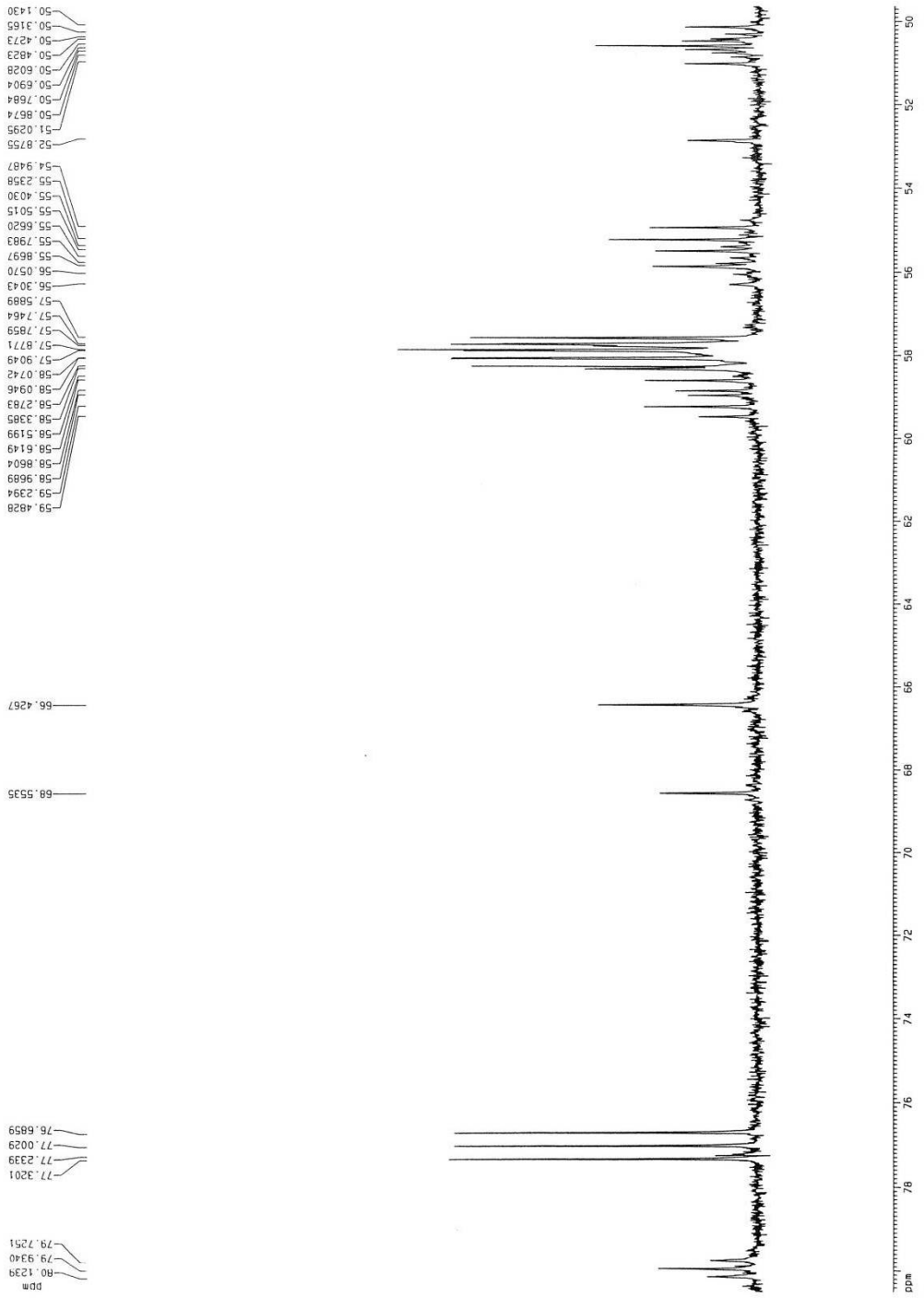
Appendix Figure 2.S9: ^{13}C spectrum (150 MHz) of 1a in $CDCl_3$, 165-175 ppm



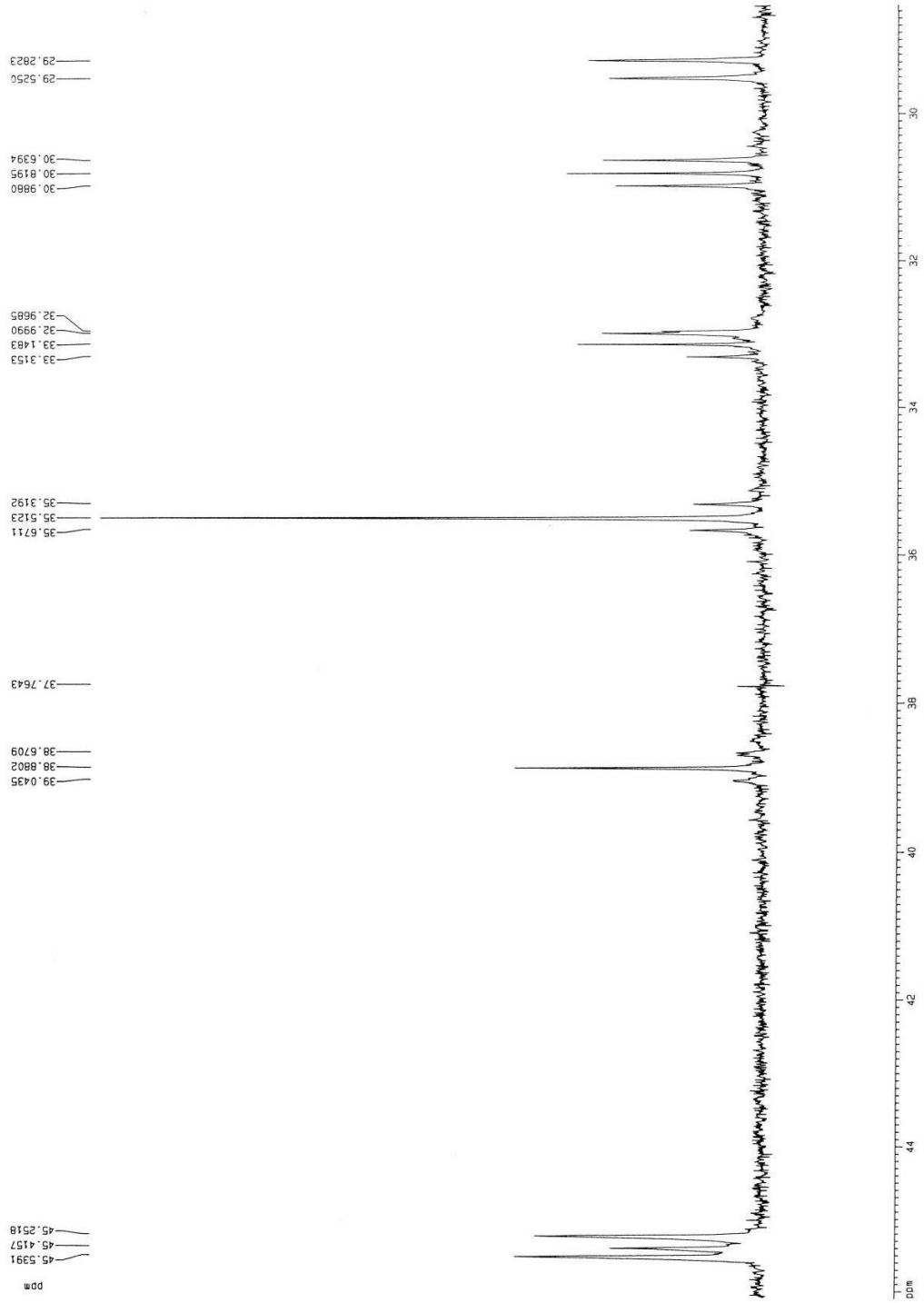
Appendix Figure 2.S10: ^{13}C spectrum of 1a in $CDCl_3$, 165-175 ppm



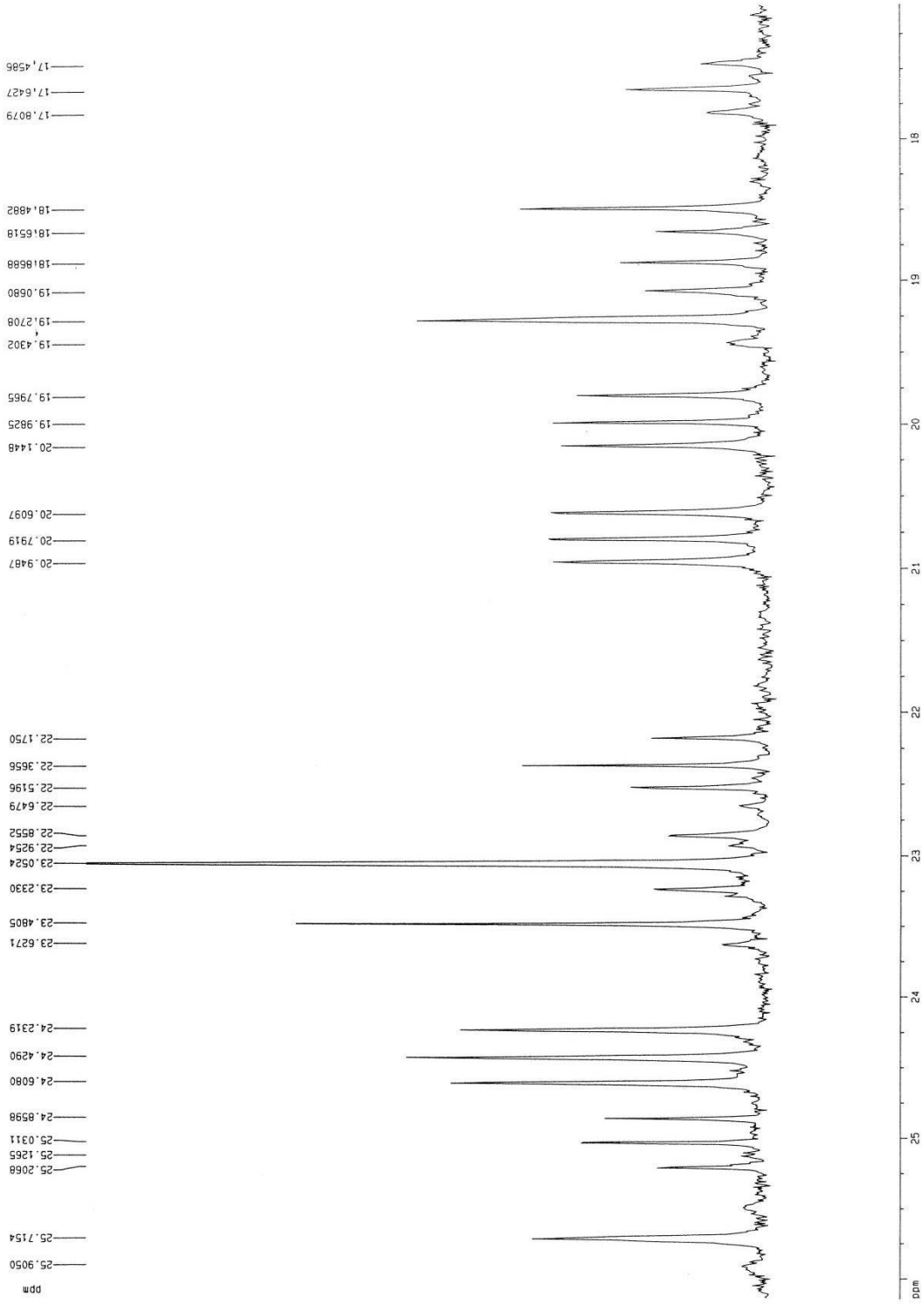
Appendix Figure 2.S11: ^{13}C spectrum of 1a in $CDCl_3$, 110-138 ppm



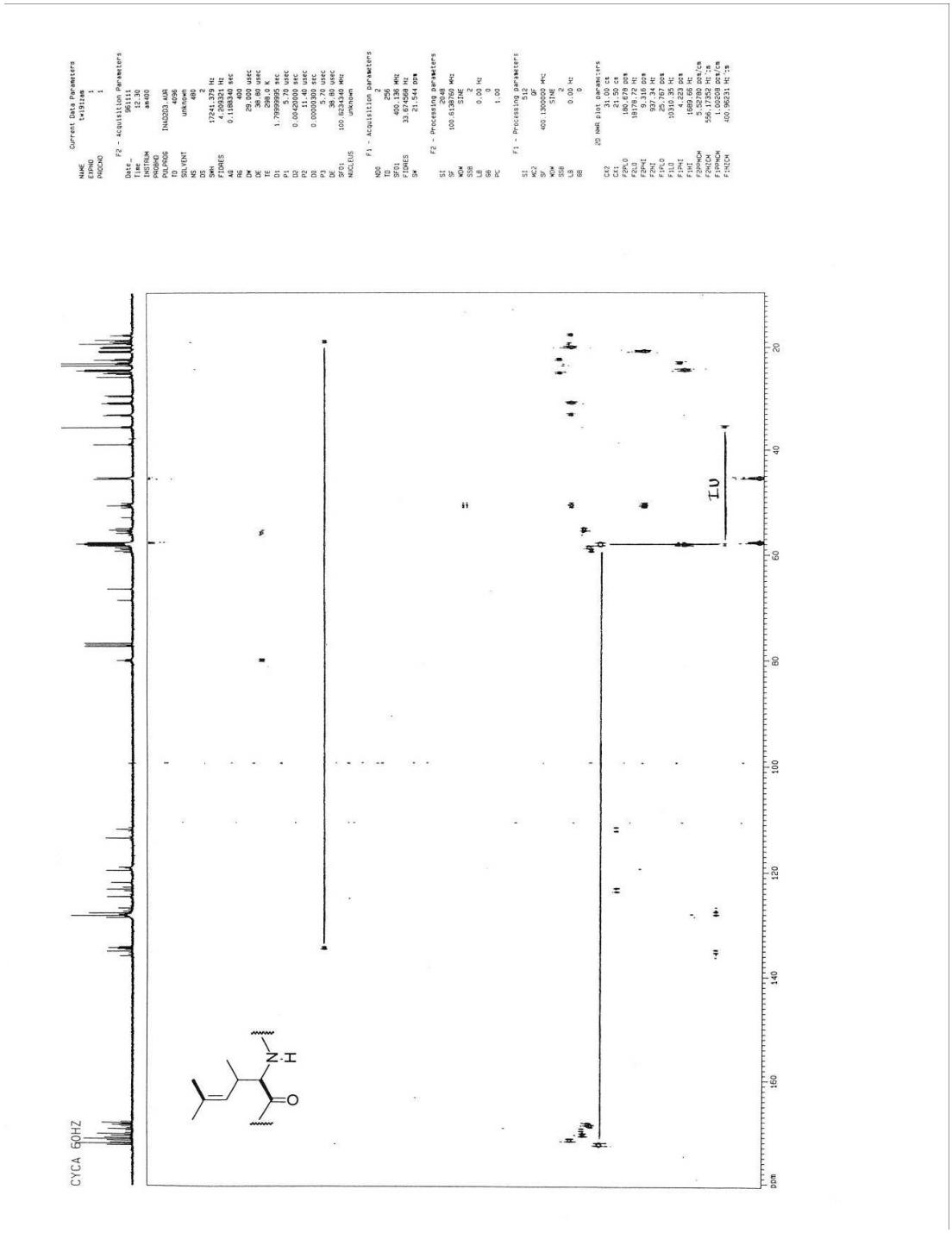
Appendix Figure 2.S12: ^{13}C spectrum of 1a in $CDCl_3$, 50-80 ppm



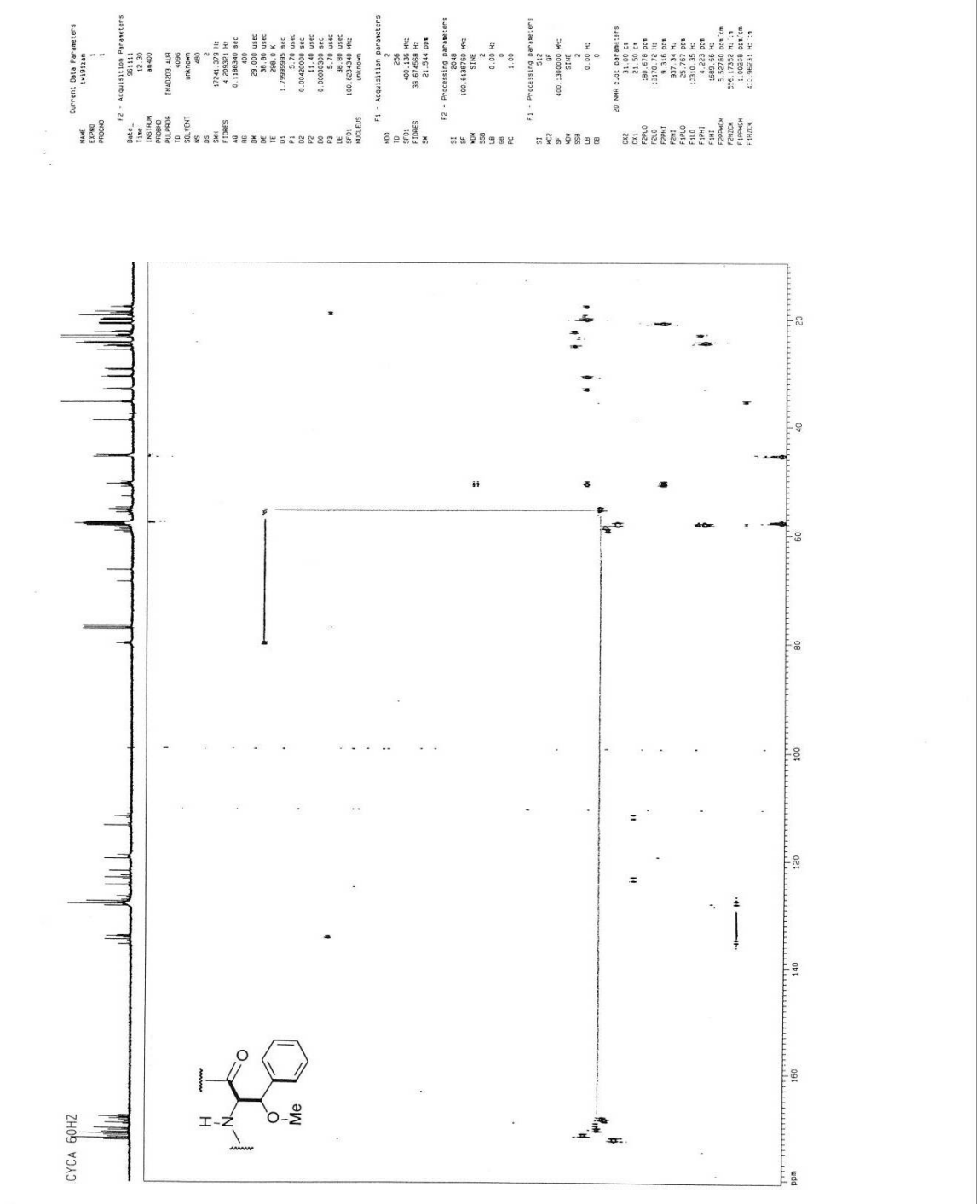
Appendix Figure 2.S13: ^{13}C spectrum of 1a in $CDCl_3$, 29-45 ppm



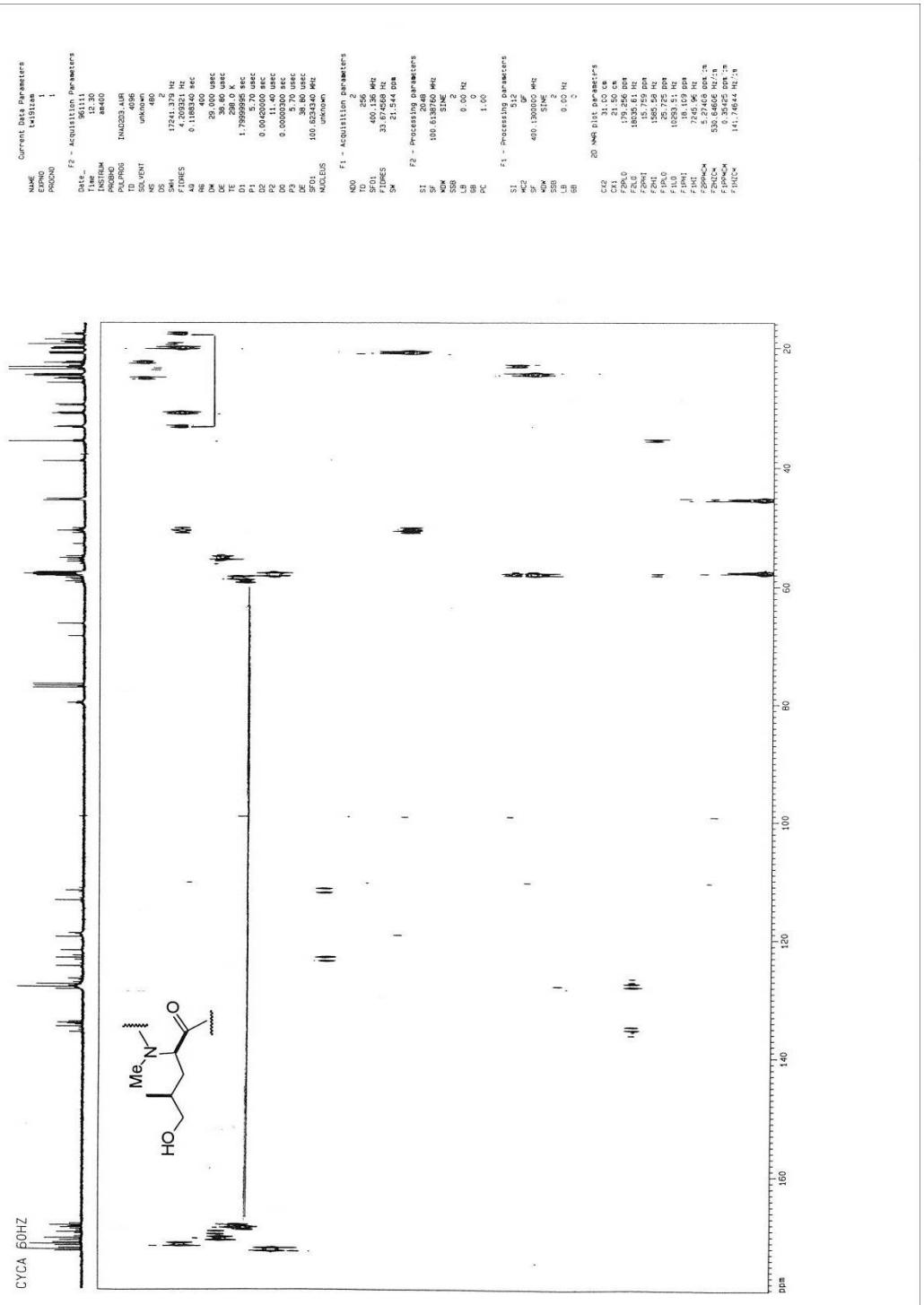
Appendix Figure 2.S14: ^{13}C spectrum of 1a in $CDCl_3$, 17-26 ppm



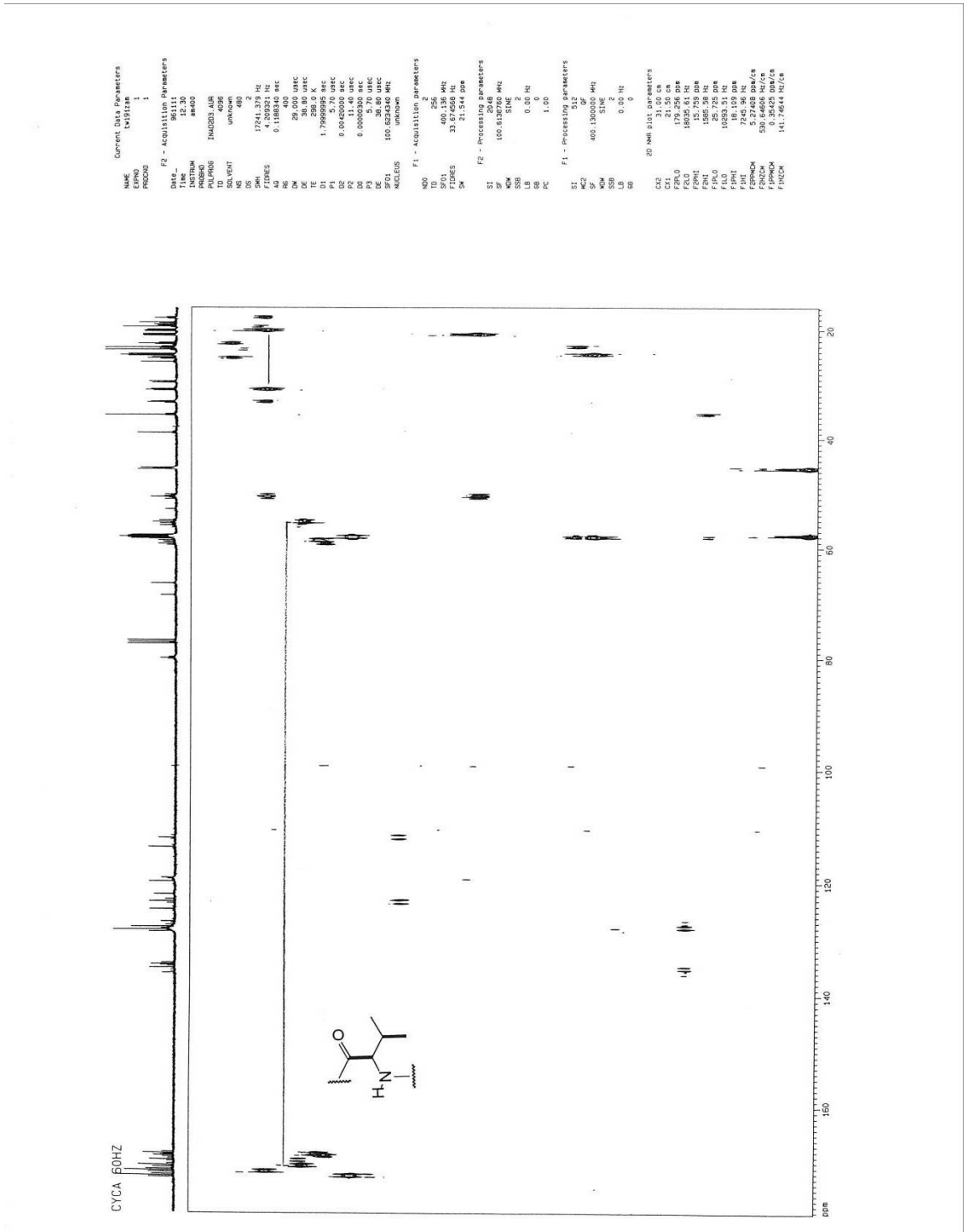
Appendix figure 2.S15: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing ADH residue couplings



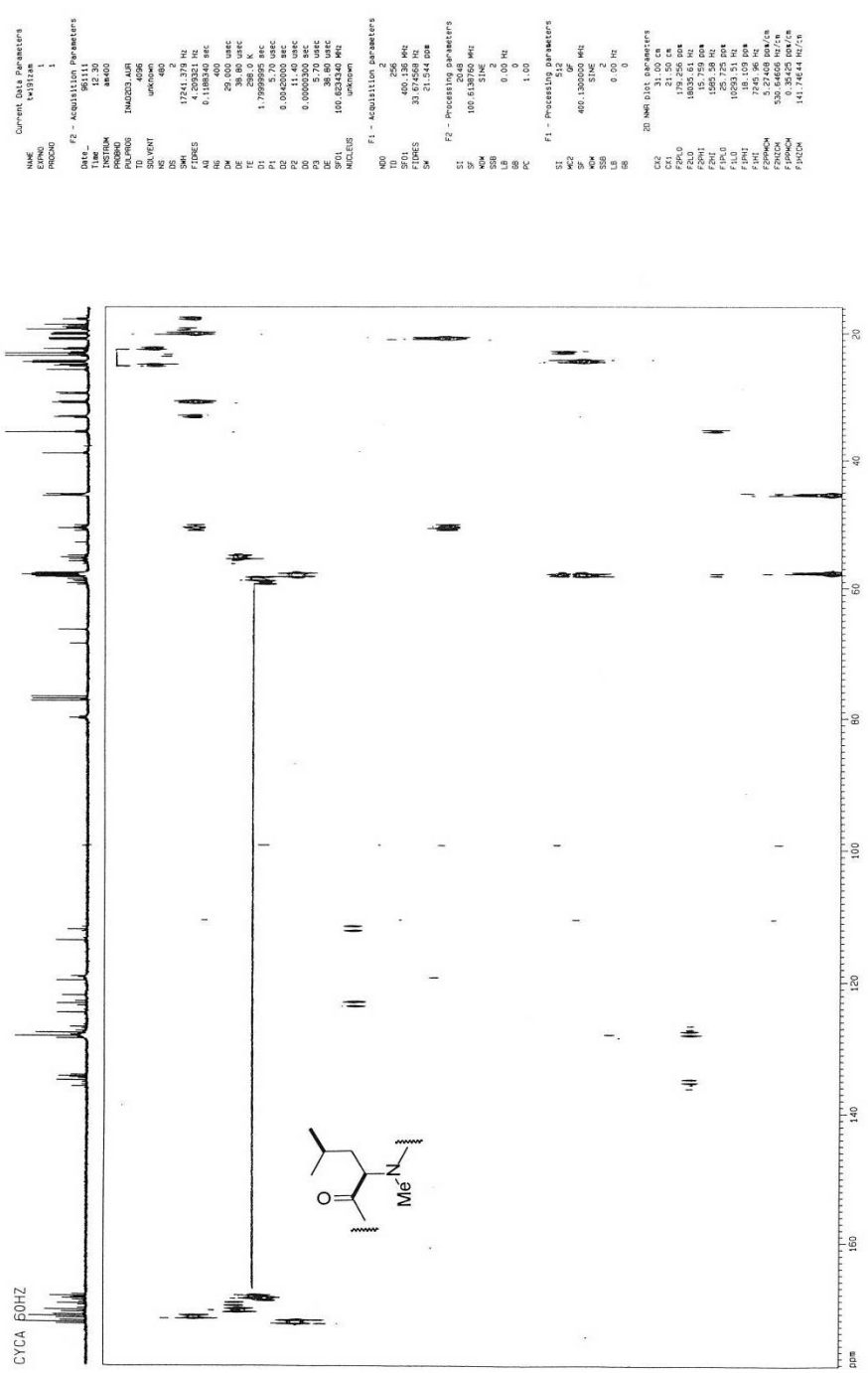
Appendix figure 2.S16: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing β-OMe-Phe residue couplings



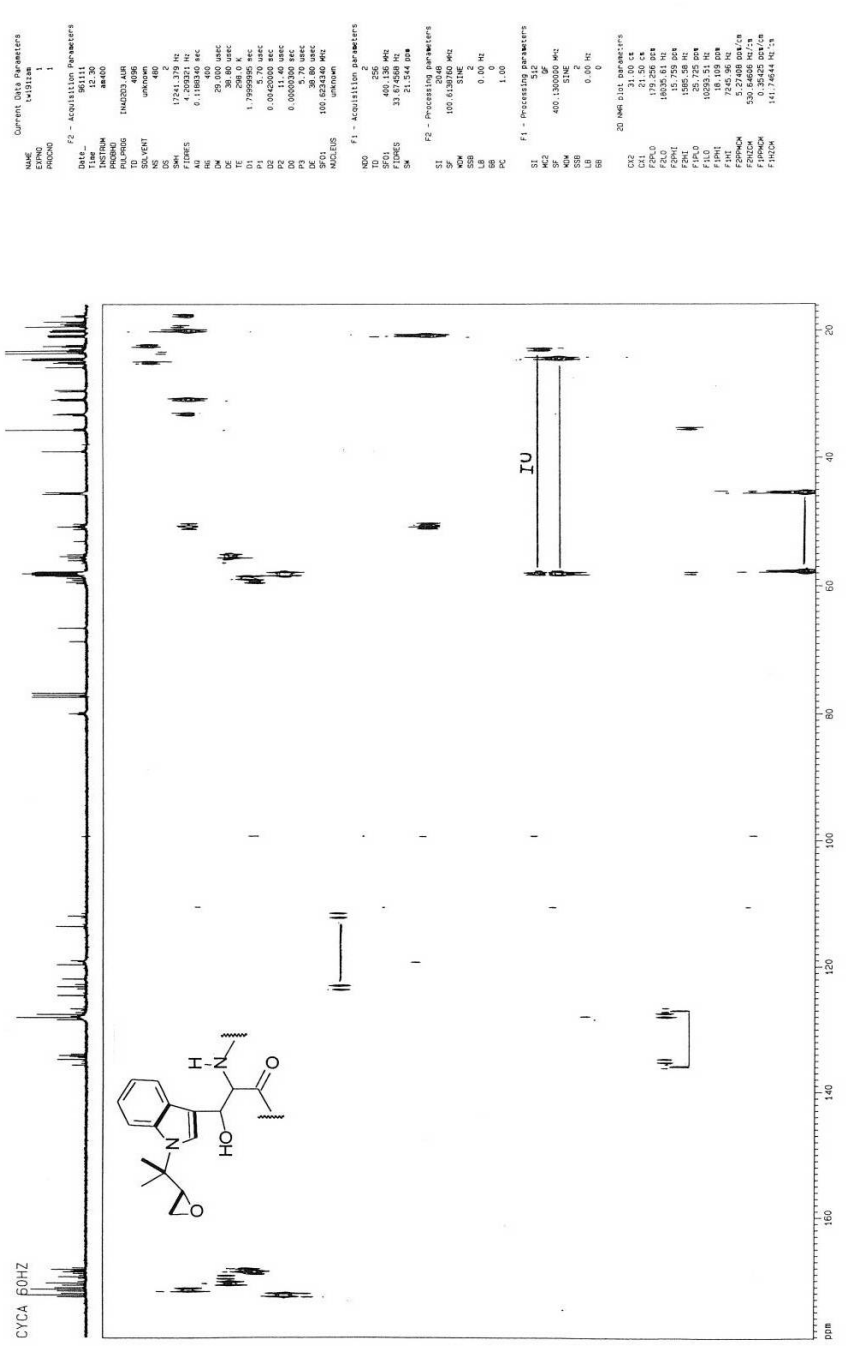
Appendix figure 2.S17: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing *N*-Me-δ-OH-Leu residue couplings



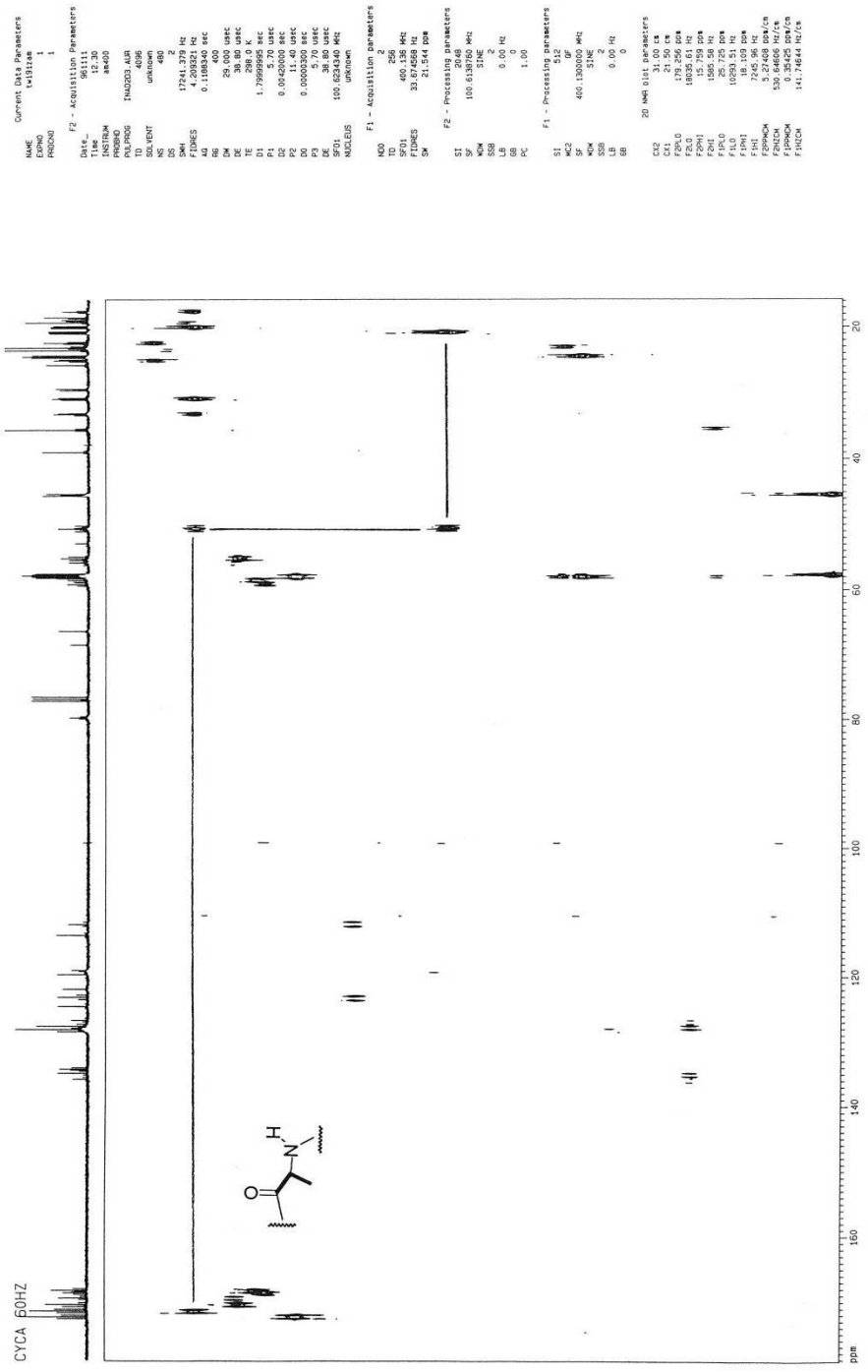
Appendix figure 2.S18: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing Val residue couplings



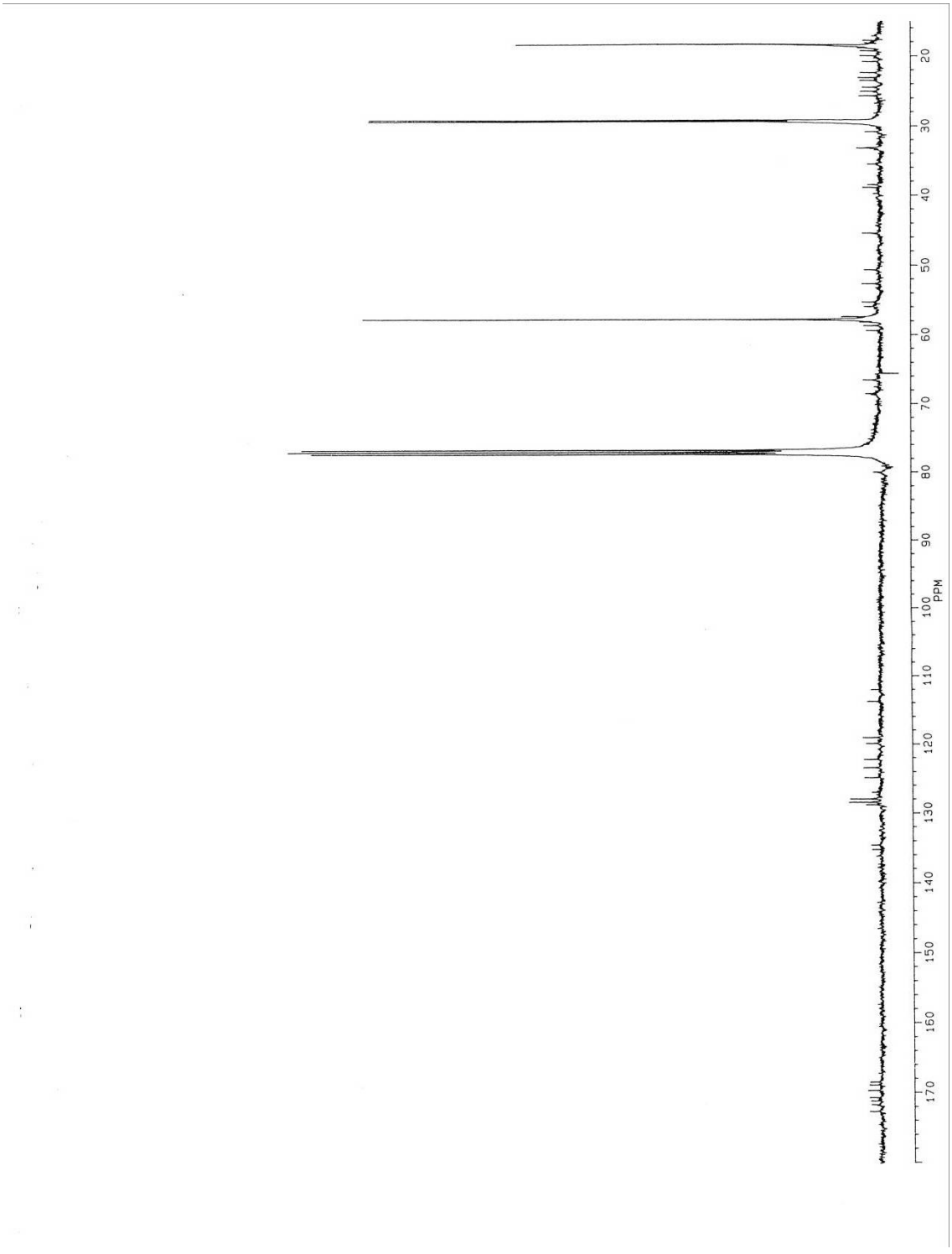
Appendix figure 2.S19: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing *N*-Me-Leu residue couplings



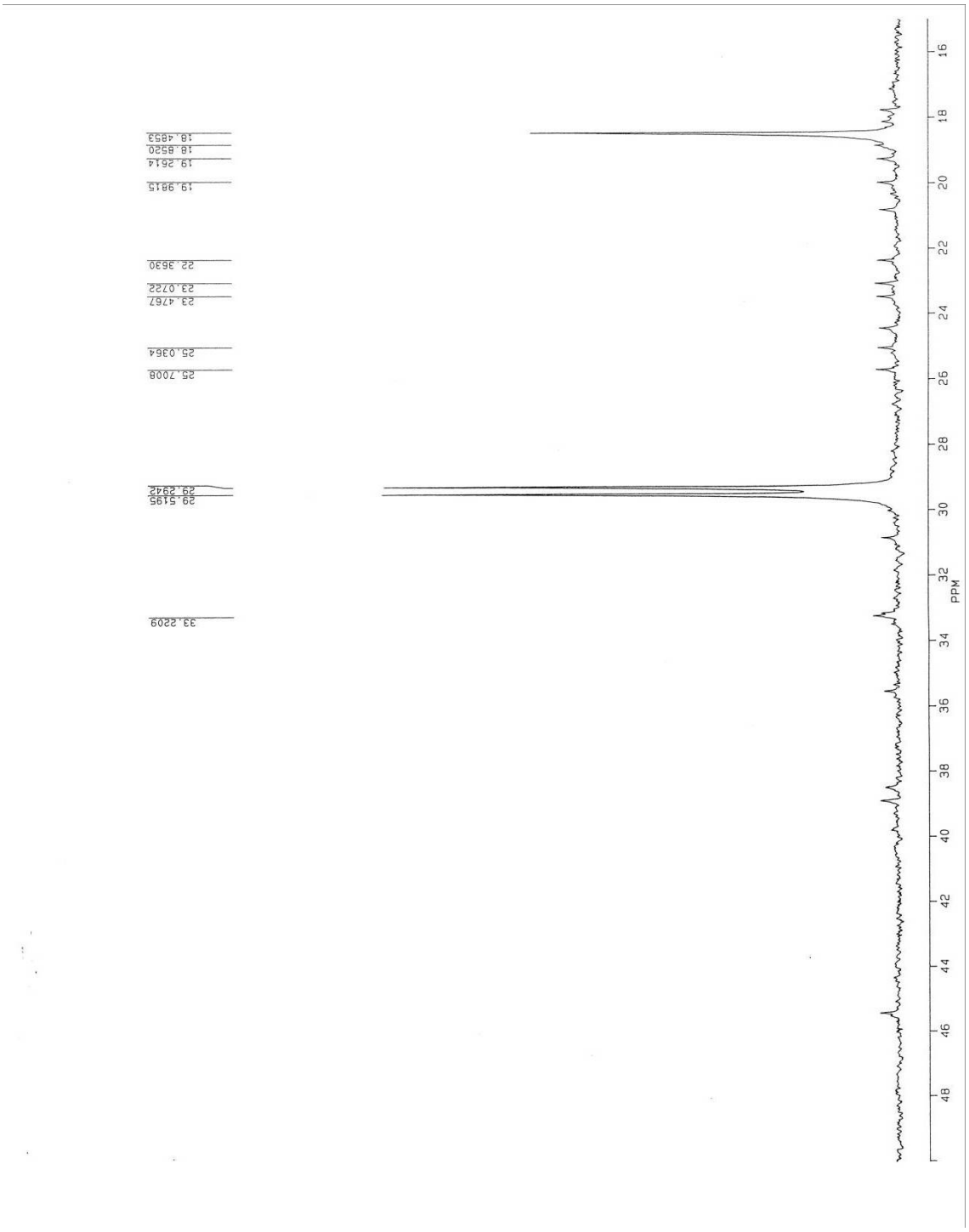
Appendix figure 2.S20: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-OH-Trp residue couplings



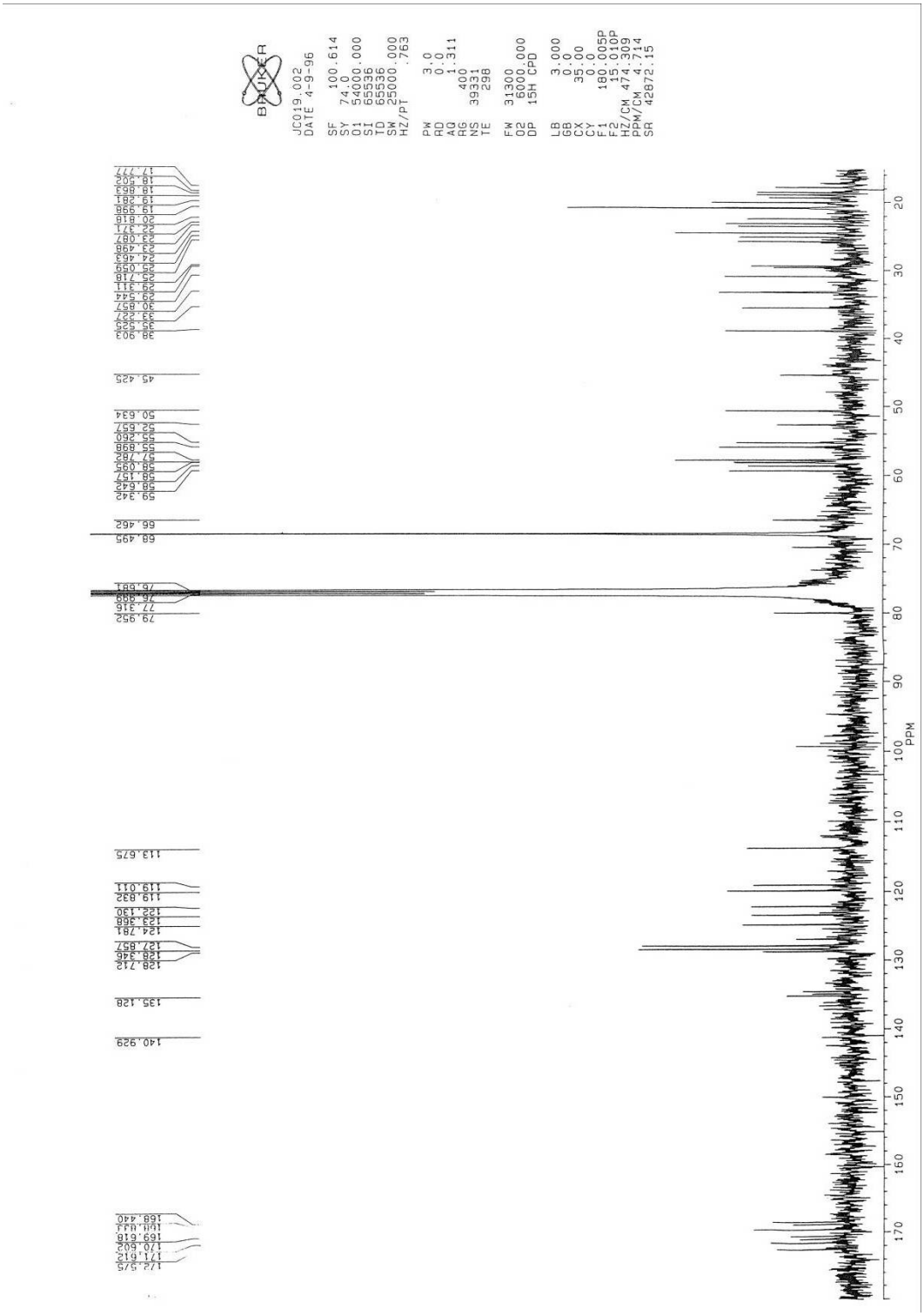
Appendix figure 2.S21: 2D INADEQUATE NMR spectrum of 1a in $CDCl_3$ showing Ala residue couplings



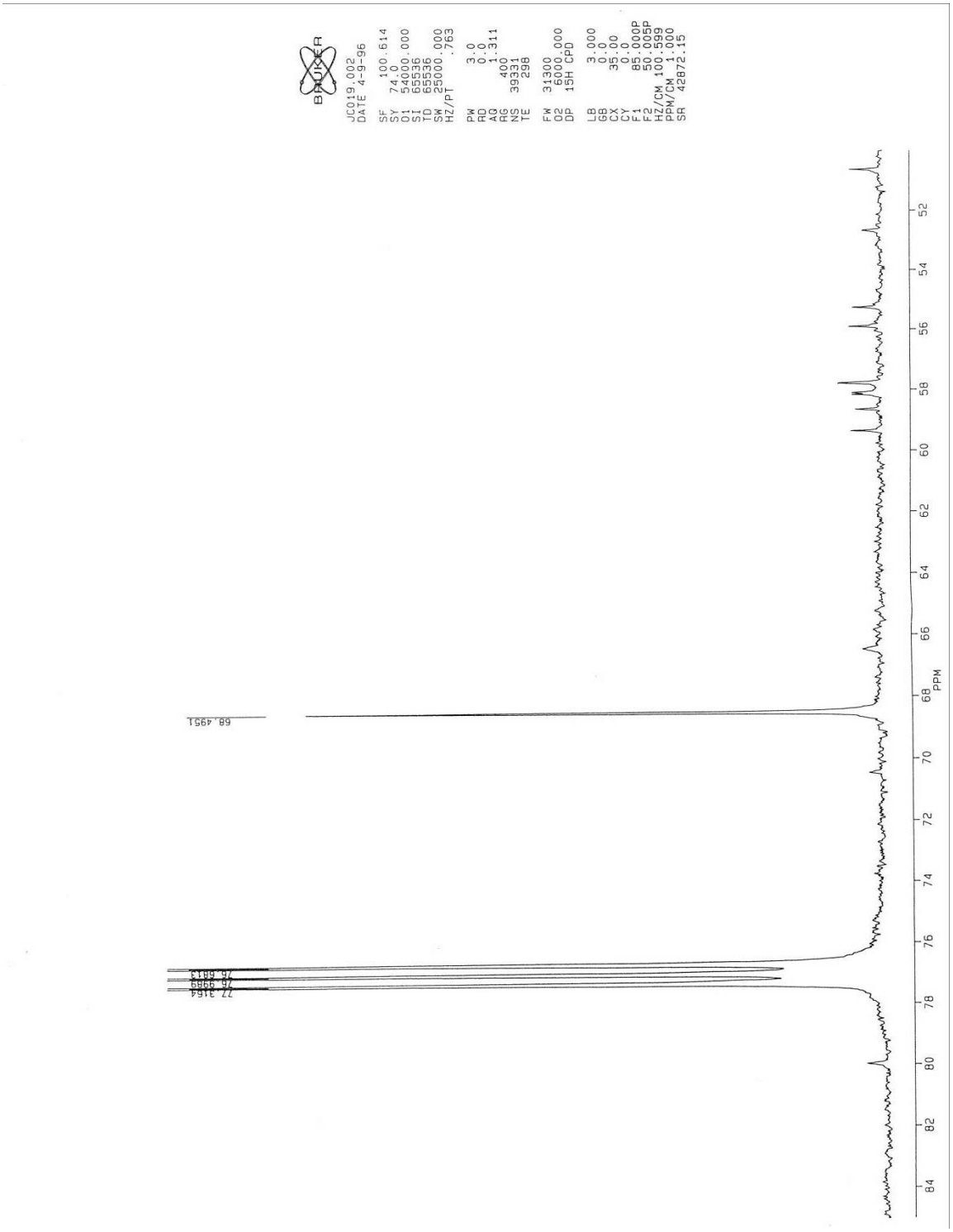
Appendix figure 2.S22: ^{13}C NMR spectrum of cyclomarin A from [methyl-^{13}C]methionine feeding (1b) in $CDCl_3$



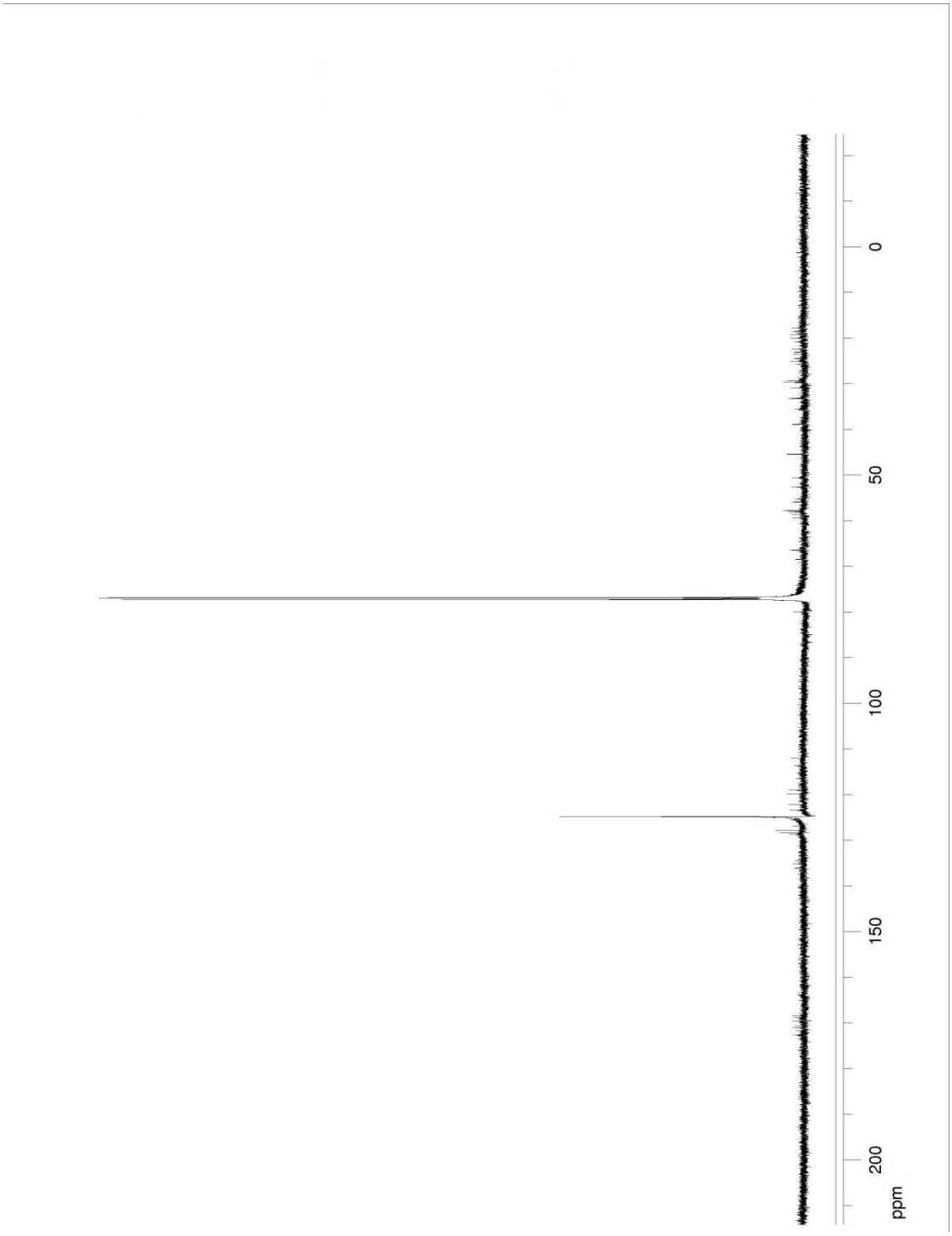
Appendix figure 2.S23: ^{13}C NMR spectrum of 1b in $CDCl_3$, expansion



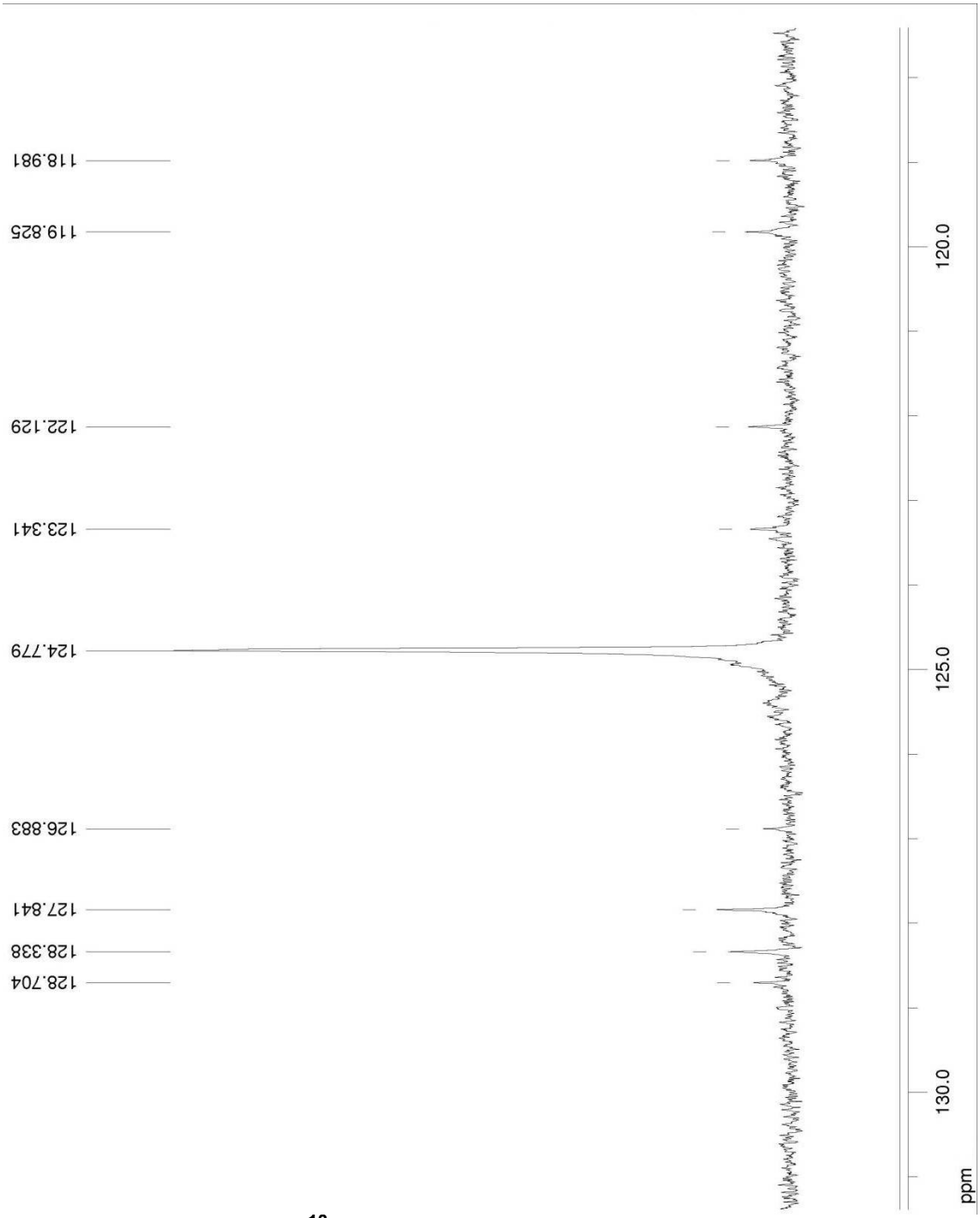
Appendix figure 2.S24: ^{13}C NMR spectrum of cyclomarin A from [3-^{13}C]tryptophan feeding (1c) in $CDCl_3$



Appendix figure 2.S25: ^{13}C NMR spectrum of 1c in $CDCl_3$, expansion



Appendix figure 2.S26: ^{13}C NMR spectrum of cyclomarin A from sodium [1-^{13}C]isobutyrate feeding (1d) in $CDCl_3$



Appendix figure 2.S27: ^{13}C NMR spectrum of 1d in $CDCl_3$, expansion

2.7: References

(1) Renner, M.; Shen, Y.-C.; Cheng, X.-C.; Jensen, P.; Frankmoelle, W.; Kauffman, C.; Fenical, W.; Lobkovsky, E.; Clardy, J. Cyclomarins A-C, new antiinflammatory cyclic peptides produced by a marine bacterium (*Streptomyces* sp.). *J. Am. Chem. Soc.* **1999**, *121*, 11273-11276.

(2) Gontang, E.; Fenical, W.; Jensen, P. Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. *Appl. Environ. Microbiol.* **2007**, *73*, 3272-3282.

(3) Fenical, W.; Jensen, P. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* **2006**, *2*, 666-673.

(4) Schultz, A.; Oh, D.-C.; Carney, J.; Williamson, R.; Udwary, D.; Jensen, P.; Gould, S.; Fenical, W.; Moore, B. Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **2008**, *130*, 4507-4516.

(5) Hamada, Y.; Shioiri, T. Recent progress of the synthetic studies of biologically active marine cyclic peptides and depsipeptides. *Chem. Rev.* **2005**, *105*, 4441-4482.

(6) Kulanthaivel, P.; Vasudevan, V.; WO Patent WO/2000/078,797: **2000**.

(7) Iitaka, Y.; Nakamura, H.; Takada, K.; Takita, T. An X-ray study of ilamycin B1, a cyclic heptapeptide antibiotic. *Acta. Crystallogr. B Struct. Crystallogr. Cryst. Chem.* **1974**, *30*, 2817-2825.

(8) Takita, T.; Ohi, K.; Okami, Y.; Maeda, K.; Umezawa, H. New antibiotics, ilamycins. *J. Antibio. Ser. A.* **1962**, *15*, 46-48.

(9) Takahashi, Y.; Omura, S. Isolation of new actinomycete strains for the screening of new bioactive compounds. *J. Gen. Appl. Microbiol.* **2003**, *49*, 141-154.

(10) Isogai, A.; Nakayama, J.; Takayama, S.; Kusai, A.; Suzuki, A. Structural elucidation of minor components of peptidyl antibiotic P168s (leucinostatins) by tandem mass spectrometry. *Biosci., Biotechnol., Biochem.* **1992**, *56*, 1079.

(11) Ueda, K.; Xiao, J.-Z.; Doke, N.; Nakatsuka, S. Structure of BZR-cotoxin II produced by *Bipolaris zeicola* race 3, the cause of leaf spot disease in corn. *Tetrahedron Lett.* **1992**, *33*, 5377-5380.

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

(13) Hoffmeister, D.; Keller, N. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat. Prod. Rep.* **2007**, *24*, 393-416.

(14) Carle, J.; Christophersen, C. Marine alkaloids. 2. Bromo alkaloids from a marine bryozoan *Flustra foliacea*. Isolation and structure elucidation. *J. Org. Chem.* **1980**, *45*, 1586-1589.

(15) Edwards, D.; Gerwick, W. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J. Am. Chem. Soc.* **2004**, *126*, 11432-11433.

(16) Nierman, W. C.; Pain, A.; Anderson, M. J.; Wortman, J. R.; Kim, H. S.; Arroyo, J.; Berriman, M.; Abe, K.; Archer, D. B.; Bermejo, C.; Bennett, J.; Bowyer, P.; Chen, D.; Collins, M.; Coulsen, R.; Davies, R.; Dyer, P. S.; Farman, M.; Fedorova, N.; Fedorova, N.; Feldblyum, T. V.; Fischer, R.; Fosker, N.; Fraser, A.; Garcia, J. L.; Garcia, M. J.; Goble, A.; Goldman, G. H.; Gomi, K.; Griffith-Jones, S.; Gwilliam, R.; Haas, B.; Haas, H.; Harris, D.; Horiuchi, H.; Huang, J.; Humphray, S.; Jimenez, J.; Keller, N.; Khouri, H.; Kitamoto, K.; Kobayashi, T.; Konzack, S.; Kulkarni, R.; Kumagai, T.; Lafton, A.; Latge, J.-P.; Li, W.; Lord, A.; Lu, C.; Majoros, W. H.; May, G. S.; Miller, B. L.; Mohamoud, Y.; Molina, M.; Monod, M.; Mouyna, I.; Mulligan, S.; Murphy, L.; O'neil, S.; Paulsen, I.; Penalva, M. A.; Pertea, M.; Price, C.; Pritchard, B. L.; Quail, M. A.; Rabbinowitsch, E.; Rawlins, N.; Rajandream, M.-A.; Reichard, U.; Renauld, H.; Robson, G. D.; De Cordoba, S. R.; Rodriguez-Pena, J. M.; Ronning, C. M.; Rutter, S.; Salzberg, S. L.; Sanchez, M.; Sanchez-Ferrero, J. C.; Saunders, D.; Seeger, K.; Squares, R.; Squares, S.; Takeuchi, M.; Tekaia, F.; Turner, G.; De Aldana, C. R. V.; Weidman, J.; White, O.; Woodward, J.; Yu, J.-H.; Fraser, C.; Galagan, J. E.; Asai, K.; Machida, M.; Hall, N.; Barrell, B.; Denning, D. W. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature.* **2005**, *438*, 1151-1156.

(17) Grundmann, A.; Li, S.-M. Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology.* **2005**, *151*, 2199-2207.

(18) Unsold, I. A.; Li, S.-M. Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. *Microbiology.* **2005**, *151*, 1499-1505.

(19) Kremer, A.; Westrich, L.; Li, S.-M. A 7-dimethylallyltryptophan synthase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterization. *Microbiology.* **2007**, *153*, 3409-3416.

(20) Yin, W.; Ruan , H.; Westrich, L.; Grundmann, A.; Li, S.-M. CdpNPT, an *N*-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation. *ChemBioChem.* **2007**, *8*, 1154-1161.

(21) Jensen, P. R.; Williams, P. G.; Oh, D.-C.; Zeigler, L.; Fenical, W. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* **2007**, *73*, 1146-1152.

(22) Maldonado, L. A.; Fenical, W.; Jensen, P. R.; Kauffman, C. A.; Mincer, T. J.; Ward, A. C.; Bull, A. T.; Goodfellow, M. *Salinispora arenicola* gen. nov., sp. nov. and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family *Micromonosporaceae*. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 1759-1766.

(23) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468-3496.

(24) Williamson, R. T. Development and application of NMR spectroscopy to marine natural products structure and biosynthesis. Ph.D. Dissertation, Oregon State University, 2001.

(25) Penn, K.; Jenkins, C.; Nett, M.; Udwary, D. W.; Gontang, E. A.; Mcglinchey, R. P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E. E.; Moore, B. S.; Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine actinobacteria. *ISME J.* **2009**, *3*, 1193-1203.

(26) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **2000**, *7*, 211-224.

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

(28) Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14083-14088.

(29) Arai, H.; Yamamoto, T.; Ohishi, T.; Shimizu, T.; Nakata, T.; Kudo, T. Genetic organization and characteristics of the 3-(3-hydroxyphenyl)propionic acid degradation pathway of *Comamonas testosteroni* TA441. *Microbiology.* **1999**, *145*, 2813-2820.

(30) Mahlert, C.; Kopp, F.; Thirlway, J.; Micklefield, J.; Marahiel, M. A. Stereospecific enzymatic transformation of α-ketoglutarate to (2S,3R)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J. Am. Chem. Soc.* **2007**, *129*, 12011-12018.

(31) Kuzuyama, T.; Seto, H. Diversity of the biosynthesis of the isoprene units. *Nat. Prod. Rep.* **2003**, *20*, 171-183.

(32) Gardiner, D. M.; Cozijnsen, A. J.; Wilson, L. M.; Pedras, M. S. C.; Howlett, B. J. The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol. Microbiol.* **2004**, *53*, 1307-1318.

(33) Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2008**, *4*, 69-74.

(34) Liu, W.-T.; Ng, J.; Meluzzi, D.; Bandeira, N.; Gutierrez, M.; Simmons, T. L.; Schultz, A. W.; Linington, R. G.; Moore, B. S.; Gerwick, W. H.; Pevzner, P. A.; Dorrestein, P. C. Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Anal. Chem.* **2009**, *81*, 4200-4209.

(35) Chen, H.; Thomas, M. G.; O'connor, S. E.; Hubbard, B. K.; Burkart, M. D.; Walsh, C. T. Aminoacyl-S-enzyme intermediates in β-hydroxylations and α,β-desaturations of amino acids in peptide antibiotics. *Biochemistry.* **2001**, *40*, 11651-11659.

(36) Izumikawa, M.; Murata, M.; Tachibana, K.; Ebizuka, Y.; Fujii, I. Cloning of modular type I polyketide synthase genes from salinomycin producing strain of *Streptomyces albus*. *Biorg. Med. Chem.* **2003**, *11*, 3401-3405.

(37) Jungmann, V.; Molnar, I.; Hammer, P. E.; Hill, D. S.; Zirkle, R.; Buckel, T. G.; Buckel, D.; Ligon, J. M.; Pachlatko, J. P. Biocatalytic conversion of avermectin to 4"-oxo-avermectin: characterization of biocatalytically active bacterial strains and of cytochrome P450 monooxygenase enzymes and their genes. *Appl. Environ. Microbiol.* **2005**, *71*, 6968-6976.

(38) Shibasaki, T.; Mori, H.; Chiba, S.; Ozaki, A. Microbial proline 4-hydroxylase screening and gene cloning. *Appl. Environ. Microbiol.* **1999**, *65*, 4028-4031.

(39) Balibar, C. J.; Howard-Jones, A. R.; Walsh, C. T. Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation. *Nat. Chem. Biol.* **2007**, *3*, 584-592.

(40) Lautru, S.; Gondry, M.; Genet, R.; Pernodet, J.-L. The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* **2002**, *9*, 1355-1364.

(41) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. Peptide bond formation in nonribosomal peptide biosynthesis. *J. Biol. Chem.* **1998**, *273*, 22773-22781.

(42) Stachelhaus, T.; Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry.* **2000**, *39*, 5775-5787.

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

(44) Healy, F. G.; Wach, M.; Krasnoff, S. B.; Gibson, D. M.; Loria, R. The txtAB genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol.* **2000**, *38*, 794-804.

(45) Sambrook, J.; Russell, D. *Molecular cloning: a laboratory manual*; Cold Spring Harbor Laboratory Press, 2001.

(46) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640-6645.

(47) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1541-1546.

(48) Kieser, T.; Bibb, M.; Buttner, M.; Chater, K.; Hopwood, D. *Practical streptomyces genetics*; The John Innes Foundation Norwich, 2000.

(49) Moore, B. S.; Seng, D. Biosynthesis of the bicyclic depsipeptide salinamide A in *Streptomyces* sp. CNB-091: Origin of the carbons. *Tetrahedron Lett.* **1998**, *39*, 3915-3918.

(50) Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10376-10381.

(51) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **2005**, *33*, 5799-5808.

(52) Ishikawa, J.; Hotta, K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett.* **1999**, *174*, 251-253.

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

(54) Kumar, S.; Tamura, K.; Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **2004**, *5*, 150.

(55) Hall, B. *Phylogenetic trees made easy. A how-to manual for molecular biologists.* ; 2nd ed.; Sinauer Associates: Sunderland, Mass, 2004.

(56) Baxevanis, A.; Ouellette, B. *Bioinformatics: a practical guide to the analysis of genes and proteins*; 3rd ed.; John Wiley and sons: Hoboken, N.J., 2001.

(57) Chen, H.; Hubbard, B.; O'Connor, S.; Walsh, C. Formation of β-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* **2002**, *9*, 103-112.

(58) Galm, U.; Schimana, J.; Fiedler, H.; Schmidt, J.; Li, S.; Heide, L. Cloning and analysis of the simocyclinone biosynthetic gene cluster of *Streptomyces antibioticus* Tü 6040. *Arch. Microbiol.* **2002**, *178*, 102-114.

(59) Steffensky, M.; Muhlenweg, A.; Wang, Z.; Li, S.; Heide, L. Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. *Antimicrob. Agents Chemother.* **2000**, *44*, 1214.

(60) Mcglinchey, R. P.; Nett, M.; Eustáquio, A. S.; Asolkar, R. N.; Fenical, W.; Moore, B. S. Engineered biosynthesis of antiprotealide and other unnatural salinosporamide proteasome inhibitors. *J. Am. Chem. Soc.* **2008**, *130*, 7822-7823.

(61) Kelly, W.; Townsend, C. Mutational analysis of *nocK* and *nocL* in the nocardicin A producer *Nocardia uniformis*. *J. Bacteriol.* **2005**, *187*, 739.

(62) Lauer, B.; Russwurm, R.; Schwarz, W.; Kalmanczhelyi, A.; Bruntner, C.; Rosemeier, A.; Bormann, C. Molecular characterization of co-transcribed genes from *Streptomyces tendae* Tü901 involved in the biosynthesis of the peptidyl moiety and assembly of the peptidyl nucleoside antibiotic nikkomycin. *Mol. Gen. Genet.* **2001**, *264*, 662-673.

(63) Wilson, M.; Gulder, T.; Mahmud, T.; Moore, B. Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the *sare1259*-encoded cytochrome P450. *J. Am. Chem. Soc.* **2010**, *132*, 12757-12765.

(64) Andersen, J. F.; Hutchinson, C. R. Characterization of *Saccharopolyspora erythraea* cytochrome P-450 genes and enzymes, including 6-deoxyerythronolide B hydroxylase. *J. Bacteriol.* **1992**, *174*, 725-735.

(65) Sherman, D.; Li, S.; Yermalitskaya, L.; Kim, Y.; Smith, J.; Waterman, M.; Podust, L. The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J. Biol. Chem.* **2006**, *281*, 26289.

(66) Ghatge, M.; Reynolds, K. The plmS2-encoded cytochrome P450 monooxygenase mediates hydroxylation of phoslactomycin B in *Streptomyces* sp. strain HK803. *J. Bacteriol.* **2005**, *187*, 7970.

(67) Li, A.; Piel, J. A gene cluster from a marine *Streptomyces* encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A. *Chem. Biol.* **2002**, *9*, 1017-1026.

2.8: Acknowledgements

We would like to thank Dr. A. S. Eustáquio (UCSD) for valuable discussions regarding PCR-targeted mutagenesis and Dr. J. A. Kalaitzis (UCSD) for assistance with NMR experiments.

Chapter 2, in full, is a reprint of the material as it appears in Biosynthesis and Structures of Cyclomarins and Cyclomarazines, Prenylated Cyclic Peptides of Marine Actinobacterial Origin (2008). Schultz, Andrew W.; Oh, Dong-Chan; Carney, John R.; Williamson, R. Thomas; Udwary, Daniel W.; Jensen, Paul R.; Gould, Steven J.; Fenical, William; Moore, Bradley S., Journal of the American Chemical Society, 130, 4507–4516. The dissertation author was the primary investigator and author of this paper.

Chapter 3:

Functional Characterization of the Cyclomarin/Cyclomarazine Prenyltransferase CymD Directs the Biosynthesis of Unnatural Cyclic Peptides

3.1: Abstract

In vitro and in vivo characterization of the cyclomarin/cyclomarazine prenyltransferase CymD revealed its ability to prenylate tryptophan prior to incorporation into both cyclic peptides by the non-ribosomal peptide synthetase CymA. This knowledge was utilized to bioengineer novel derivatives of these marine bacterial natural products by providing synthetic *N*-alkyl tryptophans to a prenyltransferase deficient mutant of *Salinispora arenicola* CNS-205.

3.2: Introduction

The anti-inflammatory cyclomarin A[1] (Figure 3.1, **8**) and antibacterial cyclomarazine A[2] (**5**) are cyclic hepta- and dipeptides produced by the marine actinobacterium *Salinispora arenicola* CNS-205.[3] Common to both peptides is a *N*-(1,1-dimethyl-1-allyl)-tryptophan (**2**) residue, which is further oxidized in **8** and cyclomarin C (**9**). While prenylated indoles, such as **2**, are common structural features in fungal and plant natural products,[4-8] these tryptophan-derived moieties are rare in bacteria. Indeed, there are numerous examples of prenylated indoles of fungal origin, such as fumigaclavine C, fumitremorgin B, and terrequinone A isolated from various *Aspergillus* species[4] (Figure 3.2). Through sequencing efforts,[9] many of the prenyltransferases involved in the biosynthesis of these compounds and other fungal indole-prenyl hybrids have been identified, and their functions characterized.[4-8,10-13] In both fumitremorgin B and terrequinone A (Figure 3.2)[4], prenylation occurs after the assembly of the amino acid derived core. Prenylation post NRPS assembly is also observed in the biosynthesis of lyngbyatoxin A (Figure 1.9),[14] a rare example of a prenyl-indole

hybrid of bacterial origin. The majority of PTase reactions follow this trend, although it has been suggested that prenylation of aromatic amino acid residues precedes the assembly of sirodesmin PL[8] as well as an unknown *Aspergillus fumigatus* peptide containing a C-7 prenylated tryptophan residue provided by the PTase 7-DMATS (Figure 3.3).[7]

We recently characterized the 47,477-bp cyclomarin/cyclomarazine gene cluster *cym*, which is dominated by the 23,358-bp *cymA* gene encoding a heptamodular nonribosomal peptide synthetase (NRPS) responsible for the dual assembly of **5** and **8**.[2] The discovery of the *cym* locus led to the in vivo characterization of the prenyltransferase (PTase) CymD, which in turn suggested its role in prenylating a common biosynthetic precursor to both peptides prior to assembly[2] in contrast to the majority of peptidyl PTase reactions. This implies that CymD prenylates a NRPS tethered intermediate, or is a *N*-(1,1-dimethyl-1-allyl)-tryptophan synthase and this prenylated amino acid is selected for and incorporated by module 1 of CymA. To the best of our knowledge, this would be the only known example of a **2** synthase, although it has been shown that FtmPT1 and CdpNPT can convert tryptophan to **2** (Figure 3.4) in vitro at a diminished catalytic efficiency in comparison to their preferred DKP substrates (Figure 3.2 and 3.4).[5] In this study, we established the biological function of CymD as **2** synthase, which provided a general strategy to readily generate unnatural *N*-alkylated tryptophan analogs of **5** and **8** by a unified mutasynthetic approach.

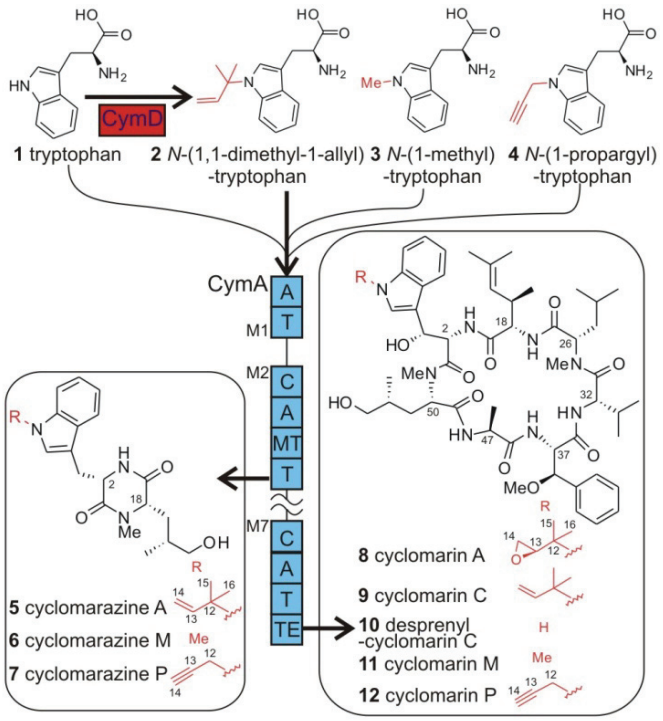


Figure 3.1: Structures and biosynthesis of natural and unnatural cyclomarin and cyclomarazine analogs. Supplementing cultures of the *S. arenicola cymD*$^{-}$ mutant with tryptophan analogs yields known and novel cyclomarins and cyclomarazines, thereby establishing **2** as the in vivo substrate for the heptamodular CymA NRPS. Thick arrows denote the pathway operative in wild type *S. arenicola* CNS-205, while thin lines represent pathways accessible via the *cymD*-deficient mutant. Abbreviations: A, adenylation domain; C, condensation domain; M1–M7, modules 1–7; MT, methyltransferase; T, thiolation domain; and TE, thioesterase.

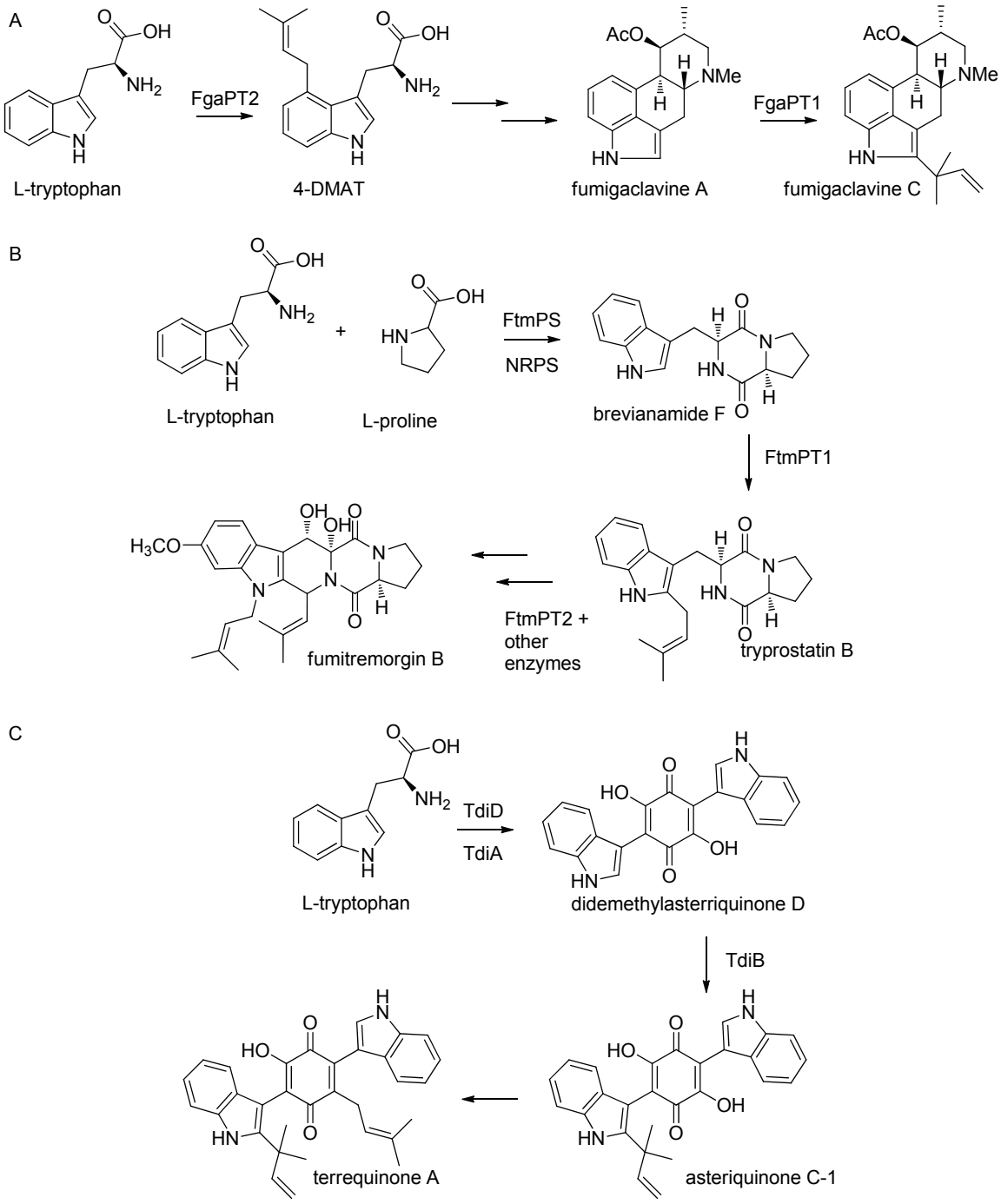


Figure 3.2: Pathways to fungal prenyl-indole hybrids highlighting the prenyltransferases integral to their biosynthesis

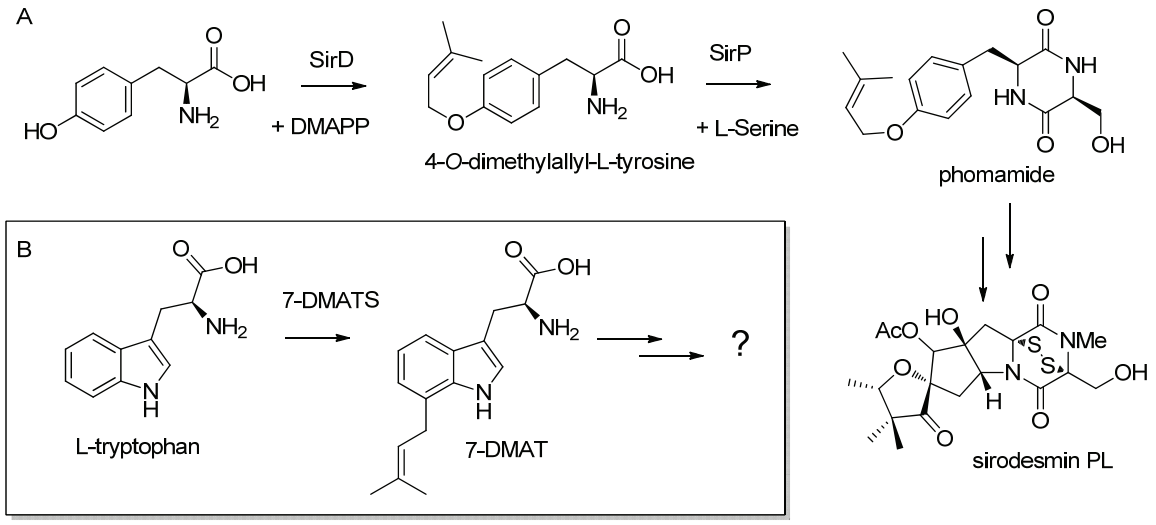


Figure 3.3: Two fungal pathways in which prenylation is suggested to occur prior to peptide assembly

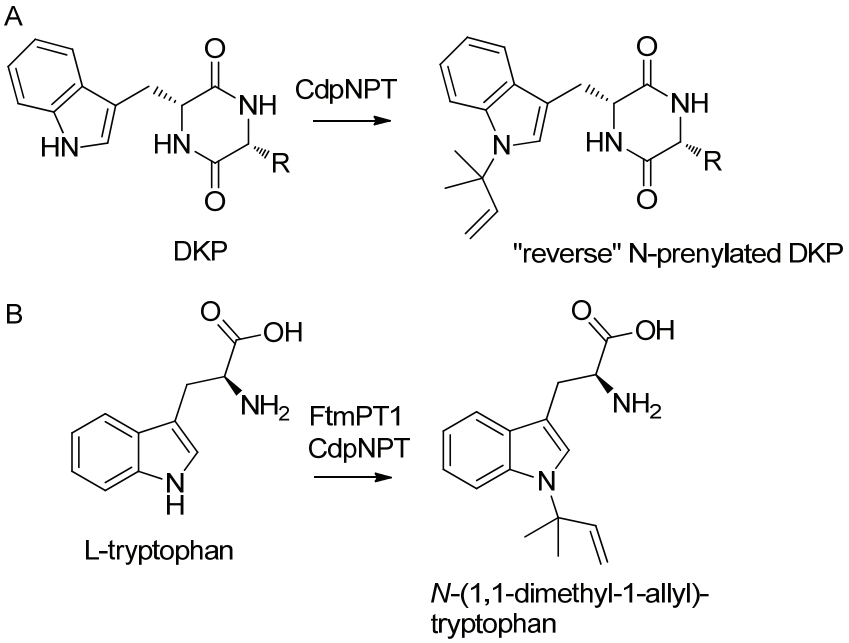


Figure 3.4: CdpNPT and FtmPT1 fugal PTases and their promiscuity towards tryptophan derived indoles. (A) The preferred substrate for CdpNPT is DKPs and FtmPT1 is the DKP brevianamide F (Figure 3.2), (B) It has been shown[5] that both PTases are promiscuous and will convert tryptophan to *N*-(1,1-dimethyl-1-allyl)-tryptophan, albeit at significantly reduced catalytic efficiency.

3.3: Results

The targeted disruption of the PTase gene *cymD* in *S. arenicola* CNS-205 previously led to a mutant deficient in the known cyclomarin and cyclomarazine chemistry and provided a novel analogue, desprenylcyclomarin C (**10**).[2,15] The >100-fold lower production of **10** in the *cymD* knockout mutant not only implied that prenylation occurs pre-NRPS assembly, but also suggested that tryptophan is a poor substrate for CymA.[2] We thus set out to explore the in vitro function of CymD, which was prepared as a 42 kD octahistidyl-tagged recombinant protein in *Escherichia coli*. The affinity-purified protein (see Appendix Figure 3.S16) was incubated with L-tryptophan (**1**) and dimethylallyl pyrophosphate (DMAPP) to yield **2** (Figure 3.5), which was identical chromatographically and spectroscopically to synthetic **2**. Reactions in which DMAPP was replaced by isoprenyl, geranyl, or farnesyl pyrophosphate yielded no detectable products, suggesting a high level of specificity for the isoprene donor. Moreover, addition of Ca^{2+} or Mg^{2+} to the reaction mixture did not significantly alter yield, nor did addition of the cation-chelating agent ethylenediaminetetraacetic acid, suggesting that CymD functions cation independently. Alignment of CymD with the fungal indole PTase FgaPT2 (Figure 3.6) revealed the presence of two conserved lysine residues at positions 146 and 207 consistent with cation-independent fungal PTases[13] that belong to an emerging group of divergent prenylating enzymes from bacteria and fungi that share a α/β barrel fold.[16,17]

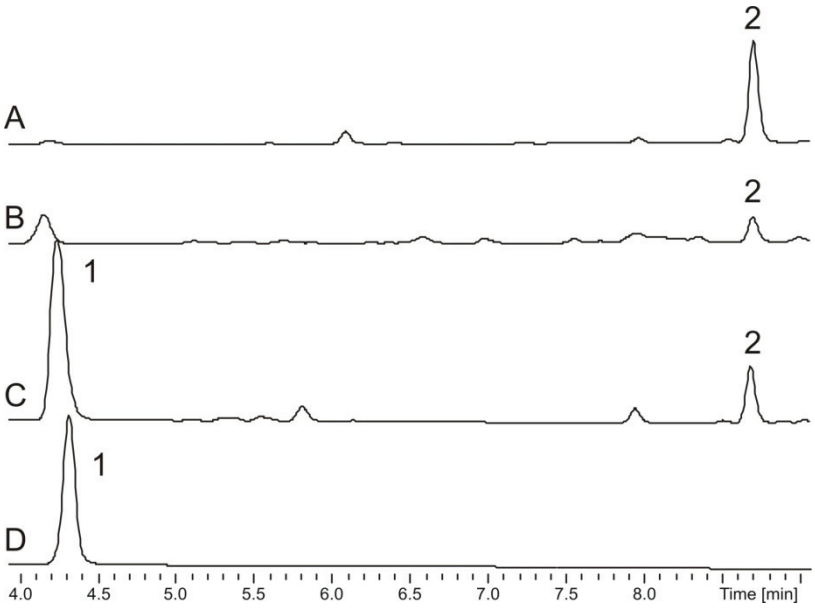


Figure 3.5: LC-DAD analysis of *N*-(1,1-dimethyl-1-allyl)-tryptophan (2) at 210 nm. Trace A, synthetic **2**; Trace B, natural **2** in the crude organic extract of *S. arenicola* CNS-205; Trace C, enzymatic **2** from the CymD-catalyzed reaction of DMAPP and tryptophan (**1**); and Trace D, the CymD boiled control of the enzymatic reaction in C.

Analysis of the *S. arenicola* CNS-205 fermentation broth revealed that **2** is indeed a natural product produced at 2 mg/L (Figure 3.5) and that its assembly ceases in the *cymD*$^{-}$ mutant. The structure of natural **2** was identical in all regards to enzymatic and synthetic **2** (Appendix Figure 3.S3). These data suggested that **2** is preformed and selected by the initiating NRPS module M1 on the heptamodular CymA synthetase to give a common biosynthetic intermediate that ultimately gives rise to natural cyclomarin and cyclomarazine cyclic peptides (Figure 3.1). We verified this conclusion by the chemical complementation of the *cymD*$^{-}$ mutant with synthetic **2** that restored the in vivo production of **5** and **8** to wild type levels (Appendix Figures 3.S1 and 3.S2), establishing **2** as the endogenous substrate of the CymA loading didomain M1.

		Motif I			Motif II	
		*			*	
FgaPT2	186	L**K**T**Y**IY**P**AL**K**AVV**TG**K	201	247	SC**DL**TS**P**AKS**R**I**K**I**Y**L	262
CdpNPT	218	A**K**E**Y**FF**P**GI**K**CAA**TG**Q	223	274	CC**DL**VD**P**AHT**R**F**K**V**Y**I	289
FtmPT2	191	A**K**A**Y**FF**P**GM**K**SLA**TG**L	206	254	GV**DL**CT**P**ERS**R**L**K**F**Y**V	269
CymD	145	F**K**AWFYLNVTGPDGAF	160	195	SL**DL**SDDPAA**R**V**K**V**Y**F	210
LtxC	187	M**K**T**Y**LNLNIAGFESGP	202	238	AL**DL**AHSDHP**R**L**K**I**Y**L	253
NphB	149	AELFARYGLDK-----	164	185	EAESVLALVRELGLHV	200
CloQ	148	L**K**DFLALGLAH-----	163	187	ARIHALSGSTPPAAHV	202

Figure 3.6: Portion of an alignment of CymD with the active site of FgaPT2 and other prenyltransferases. Fungal indole prenyltransferases (PTases) from *Aspergillus fumigatus* FgaPT2 (accession no. AAX08549), CdpNPT (ABR14712), and FtmPT2 (AFUA_8G00250) were aligned with bacterial indole PTases CymD from *Salinispora arenicola* CNS-205 (SARE_4565) and LtxC (*Lyngbya majuscula,* AAT12285) and bacterial aromatic PTases NphB (*Streptomyces* sp. CL190, AB187169) and CloQ (*Streptomyces roseochromogenes* AF329398). Li and coworkers[13] have identified the conserved residues in bold, two of which are lysines (*) responsible for the non-metal dependent activity of fungal indole PTases.

Since prenylation in the cyclomarin/cyclomarazine series occurs on free tryptophan rather than on a peptide precursor as in precedent examples,[4-6,14] it provided us an opportunity to employ a mutasynthesis approach[18] to explore whether *N*-alkyl tryptophan analogues could be simultaneously assimilated into both natural product classes to give rise to novel analogues. Commercially available *N*-(1-methyl)-tryptophan (**3**) was first evaluated by administering it to a culture of the *cymD*$^{-}$ mutant. LC-MS analysis of the resulting organic extract revealed anticipated products that were subsequently isolated and fully characterized by NMR and high-resolution MS to give the novel *N*-methyl indoles

Table 3.1: NMR Spectroscopic Data for Cyclomarins M (11) and P (12) in $CDC1_3$

	cyclomarin M (**11**)			cyclomarin P (**12**)		
position[a]	δ_C[b]	δ_H[c], (*J* in Hz)	HMBC	δ_C[b]	δ_H,[c] (*J* in Hz)	HMBC
1	170.8			170.5		
2	52.9	4.57, t (3.9)	1, 3, 5, 17	53.3	4.58, t (4.1)	1, 3, 5, 17
NH-2		6.70, d (2.9)	3, 17		6.75, d (1.9)	2, 3, 17
3	67.9	5.29, d (3.9)	1, 4, 5, 6	68.3	5.28, d (4.4)	1, 4, 5, 6
OH-3		4.13, s	2, 3		4.04[d], m	2, 3, 5
4	126.9	7.08[d], m	3, 5, 6, 7, 10, 11	125.4	7.21, s	2, 3, 5, 10, 11, 12
5	112.3			113.5		
6	125.3			127.9		
7	118.8	7.50, d (8.0)	5, 6, 8, 9, 10, 11	119.4	7.58, d (8.0)	4, 5, 8, 9, 10, 11
8	119.7	7.08[d], m	3, 5, 6, 7, 10, 11	120.4	7.11[d], m	4, 9, 10, 11
9	122.5	7.22, m	8, 10, 11	122.9	7.24, m	6, 7, 11
10	109.9	7.28, d (8.3)	6, 8	110.0	7.36, d (8.3)	4, 8, 9
11	137.0			136.2		
12	32.8	3.72, s	4, 11	35.8	4.79, d (2.4)	4, 11, 13, 14
13				77.1		
14				74.2	2.4, t (2.5)	12
17	172.4			172.6		
18	58.1	4.04, m	17, 19, 20, 24, 25	58.2	4.05[d], m	17, 19, 20, 24, 25
NH-18		8.00, d (9.4)	17, 18, 25		8.00, d (9.4)	17, 18, 25
19	35.5	1.60, s	18, 20, 21, 24	35.6	1.61, m	18, 20, 21, 23
20	124.7	4.71[d], m	22, 23, 24	124.8	4.72, d (9.9)	19, 22, 23
21	134.6			134.6		
22	25.7	1.23, s	18, 20, 21, 24	25.8	1.24, s	18, 20, 21, 24
23	19.5	1.70, s	18, 20, 21, 22, 24	18.9	1.72, s	18, 20, 21, 22, 24
24	18.6	0.62, d (6.5)	18, 19, 20	18.5	0.62, d (6.5)	18, 19, 20
25	168.5			168.5		
26	58.6	4.82, m	NMe-26	58.6	4.85, m	25, 27, 28, 31, NMe-26
NMe-26	29.5	2.82, s	26, 31	29.5	2.82, s	26, 31
27a	38.8	2.28, ddd (13.3, 10.9, 4.4)	25, 26, 28	38.9	2.25, m	25, 26, 28, 29
27b	38.8	1.02, m	25, 26, 28	38.9	1.03, m	25, 26. 28, 29, 30
28	25.2	1.43, m		25.0	1.45, m	27, 30
29	23.3	0.83, d (6.6)	27, 28	22.8	0.85, d (6.6)	27, 28
30	22.4	0.85, d (6.6)	27, 28	23.4	0.88, d (6.6)	27, 28, 29
31	170.4			171.4		
32	55.3	4.34, t (8.8)	31, 33, 34	55.3	4.35, t (8.7)	33, 34, 35, 36
NH-32		7.93, d (8.0)	32, 33, 36		7.93, d (8.0)	32, 33, 36
33	30.7	2.21, m	32	30.8	2.21, m	32, 34, 35
34	19.3	1.04, d (6.7)	32, 33, 35	19.2	1.04, d (6.7)	32, 33, 35
35	19.9	0.92, d (6.6)	32, 33, 34	20.5	0.92, d (6.7)	32, 33, 34
36	169.6			169.6		
37	55.9	4.88, m	36, 38, 39, 46	55.9	4.88, t (4.9)	36, 38, 39, 46
NH-37		7.14, d (4.7)	36, 37, 46		7.13[d], m	36, 37, 46
38	79.8	5.04, d (5.3)	37, 39, 45, 40−44	79.9	5.05, d (5.3)	36, 37, 39, 40−44, 45
39	135.1			135.2		38
40−44	127−128	7.23−7.25, m	38, 43	127−128	7.23−7.25, m	38, 39
45	57.8	3.34, s	38	57.8	3.34, s	38
46	171.7			171.8		
47	50.7	4.86, m	46, 48	50.6	4.86, m	46, 48
NH-47		8.19, d (10.3)	47, 49		8.17, d (10.2)	47, 49
48	20.8	1.29, d (7.3)	46, 47	20.8	1.28, d (7.3)	46, 47
49	169.7			169.1		
50	59.3	4.72[d], m	51, 52	59.3	4.68, dd (10.5, 3.0)	1, 49, 52, NMe-8
NMe-50	29.4	2.66, s	1, 50	29.3	2.64, s	1,50
51a	32.9	2.24, m	49, 52, 53, 54	32.8	2.16, m	49, 52, 53, 54
51b	32.9	0.58, ddd (13.9, 5.7, 3.2)	49, 52, 53, 54	32.8	0.41, m	49, 52, 53, 54
52	33.3	1.32, m	51, 54	33.2	1.26, m	50, 51, 53, 54
53a	66.1	3.23, m		66.3	3.12, m	51, 54
53b	66.1	3.16, m		66.3	3.18, m	51, 54
54	17.9	0.74, d (6.8)	51, 52, 53	17.7	0.65, d	52, 53

[a] Numbering based on ref 2. [b] Assignment by HMQC and HMBC correlations. [c] 600 MHz. [d] Overlapping signals.

cyclomarin M (Figure 3.1, **11**) and cyclomarazine M (**6**). NMR analyses (Tables 3.1 and 3.2) verified the loss of the respective prenyl groups of **8** and **5** while

maintaining the remainder of the peptidic structures. HMBC correlations and chemical shift analyses further established the *N*-methyl substitution (C-12) on the tryptophan indole ring.

Table 3.2: NMR Spectroscopic Data for Cyclomarazines M (6) and P (7) in DMSO-*d*$_6$

	cyclomarazine M (**6**)			cyclomarazine P (**7**)		
position[a]	δ_C[b]	δ_H[c], (*J* in Hz)	HMBC	δ_C[b]	δ_H[c] (*J* in Hz)	HMBC
1	166.2			165.5		
2	55.7	4.10, dd (7.6, 4.3)	1, 3, 5, 17	55.6	4.12, dd (7.7, 4.3)	3, 5
NH-2		8.19, d (2.4)	1, 2, 18		8.20, d (2.5)	
3a	29.7	3.19, dd (14.4, 4.9)	1, 2, 4, 5, 6	29.7	3.18, dd (14.4, 4.9)	1, 2, 4, 5, 6
3b	29.7	3.00, dd (14.4, 4.5)	1, 2, 4, 5, 6	29.7	3.02, dd (14.4, 4.4)	1, 2, 4, 5, 6
4	128.7	7.02, s	3, 5, 6, 7, 11, 12	127.2	7.11, s	2, 3, 5, 6, 11, 12
5	108.4			109.3		
6	128.6			128.4		
7	118.9	7.50, d (7.9)	5, 6, 8, 11	119.1	7.53, d (7.9)	5, 6, 8, 9, 11
8	118.5	6.98, dd (7.5, 7.5)	4, 10	118.9	7.02, dd (7.3, 7.3)	6, 7, 10
9	121.0	7.11, dd (7.6, 7.6)	6, 8, 10, 11	121.3	7.14, dd (7.6, 7.6)	7, 11
10	109.4	7.34, d (8.2)	6, 8, 9	109.7	7.44, d (8.2)	8, 6, 9, 11
11	136.9			135.5		
12	32.1	3.72, s	4, 11	34.6	5.01, m	4, 11, 13, 14
13				79.1		
14				75.4	3.39, t (2.4)	12, 13
17	167.5			166.7		
18	58.6	3.57, dd (9.0, 3.2)	1, 17, 19, NMe-18	58.6	3.57, dd (9.0, 3.1)	1, 17, 19, 20
NMe-18	31.7	2.66, s	1, 18	31.6	2.68, s	1, 17, 18
19a	35.7	0.52, ddd (13.1, 9.5, 3.3)	17, 18, 20, 21, 22	35.0	0.55[d], m	17, 18, 20, 21, 22
19b	35.7	0.38, ddd (13.7, 9.1, 4.2)	17, 18, 20, 21, 22	35.0	0.41, ddd (13.7, 9.0, 4.3)	17, 18, 20, 21, 22
20	32.0	1.32, m	19, 21, 22	32.2	1.34, m	
21a	66.0	2.83, dd (10.3, 5.3)	19, 20, 22	65.9	2.81, m	19, 20
21b	66.0	2.72, dd (10.1, 7.0)	19, 20, 22	65.9	2.74, m	19, 20
OH-21		4.34, m			4.28, m	
22	15.9	0.57, d (6.6)	19, 20	15.9	0.58[d], d (6.6)	17, 19, 20, 21

[a] Numbering based on ref 2. [b] Assignment by HSQC and HMBC. [c] 600 MHz. [d] Overlapping signals.

We next explored incorporating an *N*-propargyl group to these cyclic peptides in order to prepare analogs appropriate for biological evaluation via semi-synthesis utilizing Click chemistry.[19] Upon the addition of **4** to the *S. arenicola cymD* knockout mutant, new *cym* analogs cyclomarin P (**12**) and cyclomarazine P (**7**) carrying the anticipated propargyl side chains (C-12 through C-14) were similarly produced by fermentation. The yields of unnatural **11** and **12** were comparable to that of natural **8** in *S. arenicola* CNS-205 at 1–3 mg/L with a similar trend observed in the cyclomarazine series. These observations coupled with the inability of tryptophan to restore wild type biosynthetic levels[2] suggest that the native CymA loading adenylation (A) domain in module 1 (M1) can

accommodate varied *N*-1 substituted tryptophan substrates for assimilation into the *cym* hepta- and dipeptides.

3.4: Discussion

Herein we report the characterization of CymD as a bacterial *N*-(1,1-dimethyl-1-allyl)-tryptophan synthase, which alkylates tryptophan with DMAPP as the prenyl donor in a cation-independent manner prior to NRPS assembly. Although several biochemical studies have been performed on bacterial[5] and fungal indole PTases,[4-8] few studies have probed the in vivo timing of the prenyltransfer reaction. In vitro kinetic analysis suggests tryptophan and tyrosine are prenylated by 7-DMATS[7] and SirD,[8] respectively, prior to incorporation into non-ribosomal peptides, although the final product of the 7-DMATS containing pathway is unknown and in vivo radioisotope incorporation studies contradict the in vitro kinetic analysis of SirD.[20] Numerous studies have revealed the promiscuity of bacterial[16] and fungal[4] prenyltransferases towards related aromatic substrates, making it difficult to solely extend in vitro observations to in vivo biological relevance. *N*-(1,1-Dimethyl-1-allyl)-tryptophan synthase activity was recently reported for the fungal PTase CdpNPT, which functions to reverse prenylate various tryptophan-containing DKPs at the indole nitrogen yet exhibits a high level of in vitro promiscuity toward indoles such as tryptophan.[5] In contrast to CdpNPT's preference to preformed peptides, CymD prenylates free tryptophan, providing both in vitro and in vivo evidence that prenylation occurs prior to NRPS-mediated incorporation into the cyclomarin and cyclomarazine peptides. The preference of CymA module 1 for **2** over tryptophan provided us a

unique opportunity to prepare novel *N*-alkylated cyclomarin and cyclomarazine analogs, which in turn supported the pre-NRPS timing of CymD catalyzed prenyl transfer. This work demonstrates the amenability of cyclomarin/cyclomarazine to diversification via mutasynthesis, providing novel analogues of these di- and heptapeptides and furthering our general understanding of NPRS specificity and amenability to engineering.

3.5: Experimental section

3.5.1: General experimental procedures

Chemicals were acquired from Sigma-Aldrich or Fisher Scientific unless noted otherwise. NMR spectroscopic data were obtained on a Bruker 600 MHz spectrometer equipped with a 1.7 mm cryoprobe. LC analysis was performed on a Agilent 1200 series LC or a HP series 1100 LC-MS system utilizing electrospray ionization in positive mode with a linear gradient of 10-90% acetonitrile at 0.7 mL/min over 24 min utilizing a Luna 4.6 x 100 mm, 5 µm C_{18}(2) column (Phenomenex). High-resolution mass spectra were collected by ESI-HR-FTMS at the Chemistry & Biochemistry Mass Spectrometry Facility, University of California San Diego.

3.5.2: Bacterial strains and culture conditions

Wild type *S. arenicola* CNS-205 and the prenyltransferase deficient mutant *S. arenicola cymD*$^-$ were grown in A1+BFe media (10 g starch, 4 g yeast extract, 2 g peptone, 40 mg of $Fe_2(SO_4)_3$•$4H_2O$, 100 mg KBr, 1 L seawater) at 28 °C and 200 rpm.[2] Production cultures were grown for 10 days in 2.8 L Fernbach flasks containing 1 L of media unless stated otherwise. *Escherichia coli* DH5α

was used for cloning and expression experiments as described.[21] DNA purification and manipulation was performed according to standard procedures.[21,22]

3.5.3: CymD purification

The *cymD* gene (GenBank protein accession SARE_4565, genome ascension CP00850) was amplified from genomic *S. arenicola* CNS-205 DNA via PCR utilizing the forward primer cymDF1 5'-CGTGGTTCgagctcTTGACCGAGGAGTTGACGACGGTCCG (SacI site underlined) and the reverse primer cymDR1 5'-GCTCGAATTCaagcttTCATTCGGTTCTCCCTCTCG (HindIII site underlined). Ligation into pHIS8[23] yielded plasmid pHIS8-*cymD*, which was transformed into *E. coli* BL21(DE3) (Invitrogen) for expression. Transformants were grown in ZYP-5052 autoinduction media[24] containing 50 µg/mL kanamycin at 16 °C for two days. Cells were harvested and lysed by sonication in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 2 mM β-mercaptoethanol) with the addition of 0.5 g/L lysozyme. The lysate was cleared by centrifugation at 20,000 rpm for 50 min, and the resulting supernatant was loaded onto a polypropylene column containing 6 mL Ni-NTA agarose (Qiagen). The column was washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 2 mM β-mercaptoethanol), and eluted in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 10% glycerol, 2 mM β-mercaptoethanol). The resulting eluate was placed in a 10 kD cut off dialysis tube and dialyzed overnight against 2 L of phosphate buffer (50 mM sodium

phosphate pH 8.0, 300 mM NaCl, 2 mM DTT). Octahistidyl-tagged CymD was further purified by FPLC (GE Healthcare) utilizing a Superdex 200 26/60 size exclusion column with phosphate buffer as the mobile phase, and fractions containing CymD were pooled and concentrated to a final concentration of 2.34 mg/mL using a Centriprep Ultracel YM-10 (10 kD cutoff, Millipore).

3.5.4: CymD prenyltransferase assay

His_8-tagged CymD (1 μg, 0.2 μM) was incubated with 1 mM L-tryptophan and 1 mM dimethylallyl pyrophosphate (DMAPP) in 100 μL 50 mM imidazole buffer, pH 6.8, and incubated for 2 h at 37 °C. The reaction was quenched with 10 μL 1.5 M TCA, and centrifuged for 10 min at 13,000 rpm. The supernate was then analyzed by LC or LC-MS. To assay for isoprene pyrophosphate specificity, DMAPP was replaced with 1 mM isopentenyl pyrophosphate, geranyl pyrophosphate, or farnesyl pyrophosphate. Metal dependence was interrogated by including 10 mM Ca^{2+}, 10 mM Mg^{2+}, or no metal added with or without 50 mM EDTA. To characterize the product of of CymD-catalyzed tryptophan conversion, 3.25 mmol L-tryptophan, 3.25 mmol DMAPP, and 380 μg enzyme were incubated in 3.25 ml 50 mM imidazole buffer, pH 6.8, overnight at 37 °C. The reaction was dried in vacuo, redissolved in a minimal volume of MeOH, and was fractionated by preparative HPLC utilizing an Onyx C_{18} 100 x 10 mm column (Phenomenex) in isocratic 40% MeOH at 5 mL/min, with *N*-(1,1-dimethyl-1-allyl)-tryptophan eluting at 9.5 min, yielding 2.3 mg product after removing the mobile phase in vacuo.

3.5.5: Purification of *N*-(1,1-dimethyl-1-allyl)-tryptophan from *S. arenicola* CNS-205

A 4 L production culture of *S. arenicola* CNS-205 was incubated with XAD-7HP resin for 2 h. The resin and cells were collected by filtration and extracted in acetone, which in turn was evaporated in vacuo. The pH of the remaining aqueous residue was adjusted to ~14 with 10 N NaOH and washed with EtOAc. The aqueous phase was then adjusted to ~pH 2 with concentrated HCl and extracted with EtOAc. The EtOAc fraction was dried over $MgSO_4$ and evaporated in vacuo. The resulting residue was fractionated using YMC ODS-A resin (60 mm x 100 mm) with a step gradient of aqueous 20% MeOH followed by 50% MeOH. Fractions containing *N*-(1,1-dimethyl-1-allyl)-tryptophan (**2**) were concentrated in vacuo, and the remaining aqueous residue was acidified with glacial acetic acid to pH ~4. Compound **2** was absorbed onto an equilibrated 1 g SCX cation-exchange SPE cartridge (Phenomenex) from the acidified fractions, which in turn was eluted with 3 M NH_4OH. The eluate was desalted using a 1 g C_{18} SPE cartridge (Thermo), yielding 2.3 mg **2** HRMS (ESI^+) *m/z* 273.1599 [$M+H^+$] (calcd for $C_{16}H_{21}N_2O_2$, 273.1597).

3.5.6: Chemical complementation of the *cymD*- mutant and the production of novel cyclomarin and cyclomarazine analogues via mutasynthesis

Duplicate 50 mL cultures of *S. arenicola* CNS-205 and the *cymD*- mutant strain were grown in 250 mL Erlenmeyer flasks containing stainless steel springs. To duplicate mutant cultures, 4 mg tryptophan, *N*-(1,1-dimethyl-1-allyl)-tryptophan (**2**), *N*-(1-methyl)-tryptophan (**3**), or *N*-(1-propargyl)-tryptophan (**4**) were added as 1 mg aliquots daily from day 6 through and including day 9. On the following day 10, the cultures were extracted with EtOAc, and the crude residues analyzed by analytical LC-MS as described above. For the production of analogs, 1 L cultures of the *cymD*- mutant were administered compounds **3** (1 g) or **4** (50 mg) dissolved in H_2O in four equal daily aliquots starting on the 6^{th} day of growth. The cultures were extracted by EtOAc partitioning to yield approximately 130 mg crude organic extracts, which were fractionated utilizing Sephadex LH-20 resin (GE Healthcare, 35 g resin in a 30 mm wide column) with MeOH as the mobile phase. Fractions containing the target compound, as judged on the basis of ESI-MS were pooled, dried in vacuo, and subjected to purification via preparative HPLC utilizing a Luna C_{18}(2) 250 x 10 mm column (Phenomenex) under isocratic conditions with a flow rate of 2.5 mL/min and monitored at 210 nm. Methyl indole derivatives cyclomarazine **6** (0.4 mg) eluted at 33 min with 21% MeCN and cyclomarin **11** (2.6 mg) eluted at 36 min with 52% MeCN. Propargyl derivatives cyclomarazine **7** (1.5 mg) eluted at 33 min with 24% MeCN, while cyclomarin **12** (1.0 mg) eluted at 29 min with 56% MeCN.

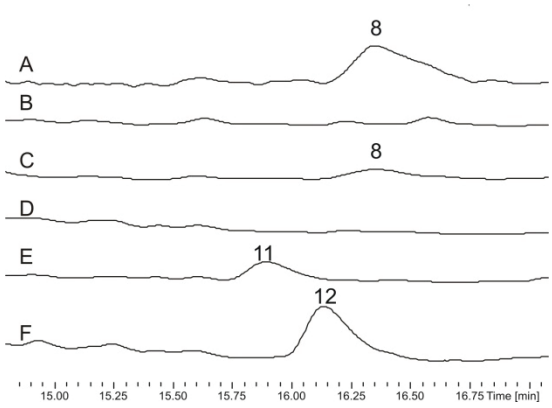
Cyclomarazine M (6). NMR data, see Table 3.2; HRMS (ESI^+) *m/z* 366.1790 $[M+Na]^+$ (calcd for $C_{19}H_{25}N_3O_3Na$, 366.1788).

Cyclomarazine P (7). NMR data see Table 3.2. HRMS (ESI^+) 390.1790 *m/z* $[M+Na]^+$ (calcd for $C_{21}H_{25}N_3O_3Na$ 390.1788).

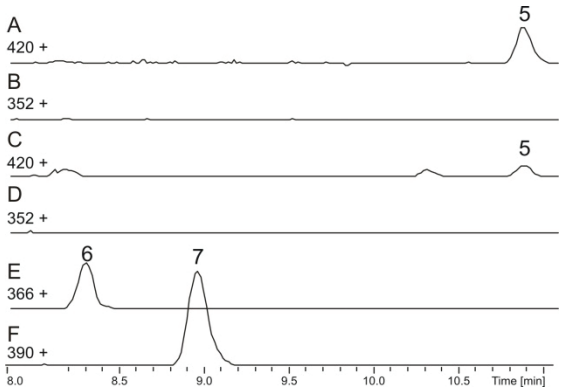
Cyclomarin M (11). NMR data see Table 3.1. HRMS (ESI^+) 995.5587*m/z* $[M+Na]^+$ (calcd for $C_{52}H_{76}N_8O_{10}Na$ 995.5577).

Cyclomarin P (12). NMR data see Table 3.1. HRMS (ESI^+) 1019.5582*m/z* $[M+Na]^+$ (calcd for $C_{54}H_{76}N_8O_{10}Na$ 1019.5577).

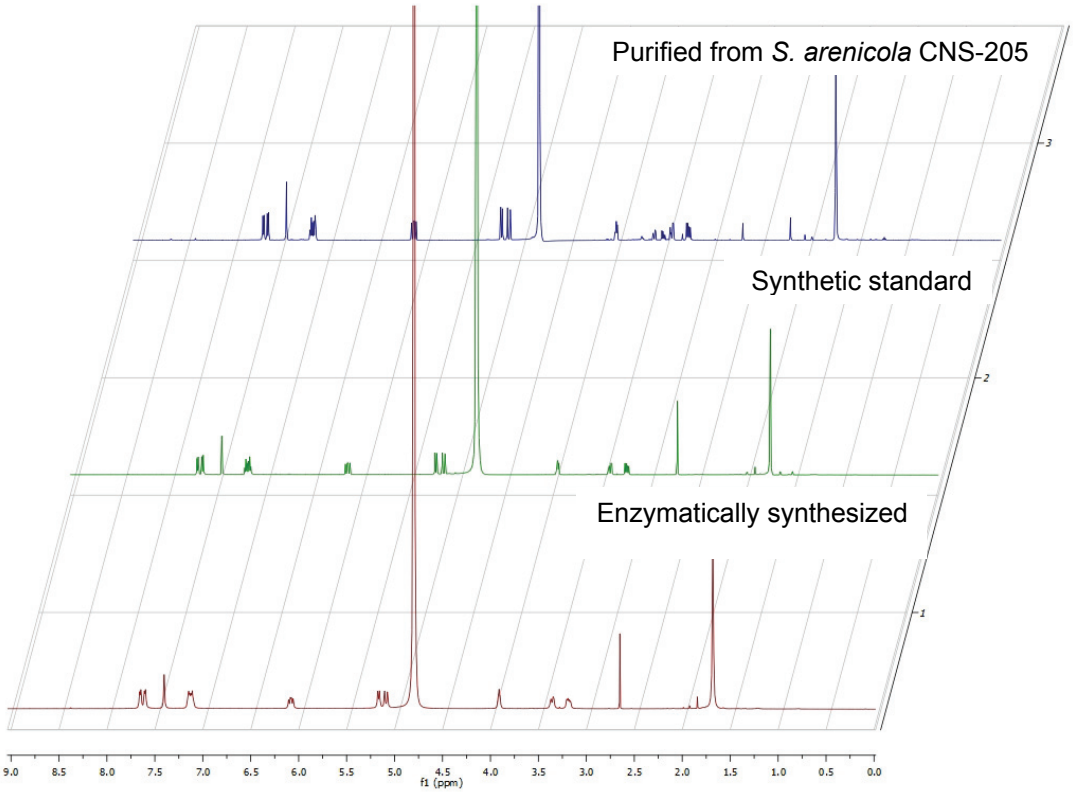
3.6: Appendix

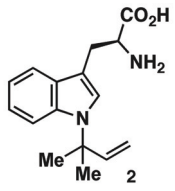


Appendix Figure 3.S1: Analysis of relative yield of cyclomarin analogs in *Salinispora arenicola* CNS-205 wild type and *cymD*- mutant. Trace A: Wild type *S. arenicola* CNS-205. Trace B-F, *S. arenicola cymD*$^-$ mutant. Traces C-F represent cultures supplemented with 80 mg/L tryptophan analogs. Trace C: *N*-(1,1-dimethyl-1-allyl)-tryptophan, D: tryptophan, E: *N*-(1-methyl)-tryptophan, F: *N*-(1-propargyl)-tryptophan. Chromatograms recorded at 210 nm. Note similar production levels of cyclomarin M (**11**, trace E) and cyclomarin P (**12**, trace F) in comparison to cyclomarin A (**8**) in trace A and C. Desprenylcyclomarin C (**10**, RT=15.2 min) is produced at levels below the detection limit of this analysis in the *cymD*$^-$ mutant (trace B), and did not increase to a detectable level with the addition of tryptophan (trace D).

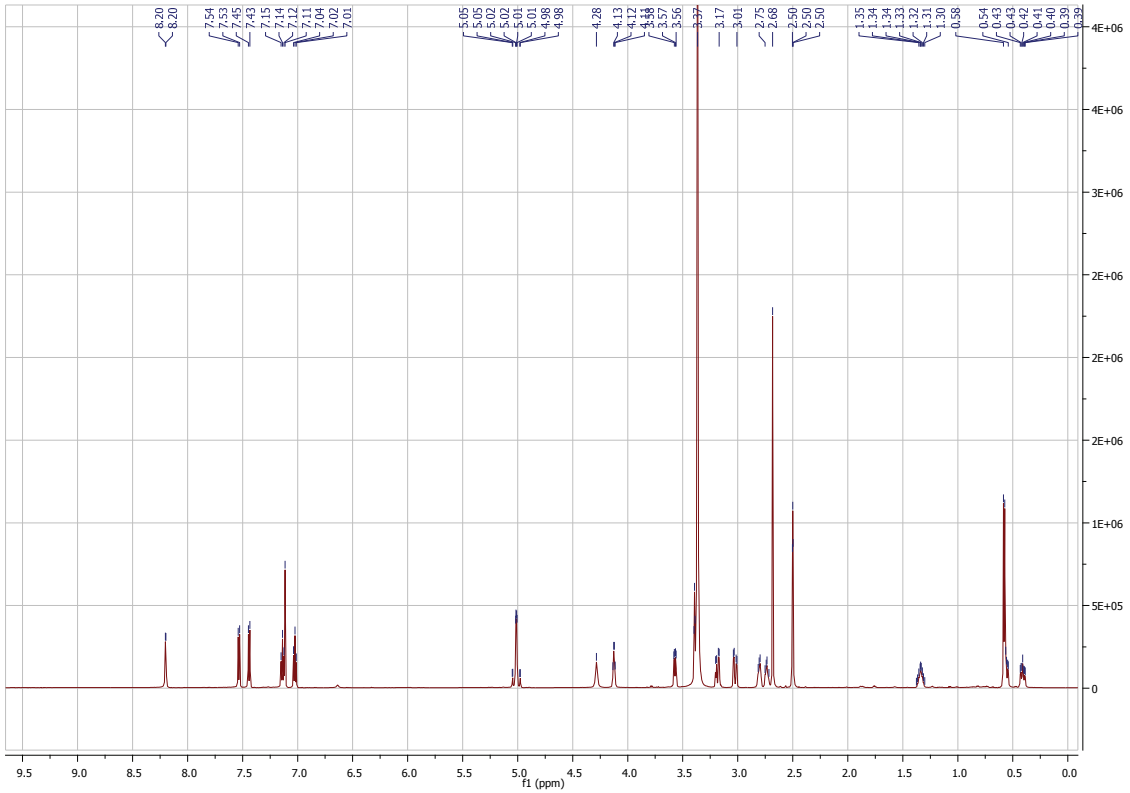


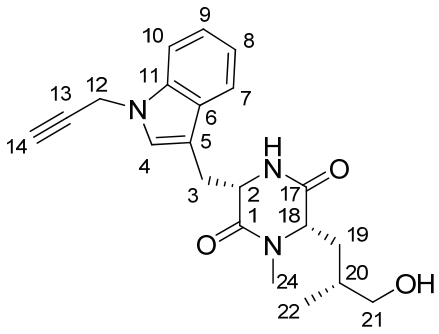
Appendix Figure 3.S2: Analysis of production levels of cyclomarazine analogs in *Salinispora arenicola* CNS-205 wild type and *cymD-* mutant. Trace A: Wild type *S. arenicola* CNS-205. Trace B-F, *S. arenicola cymD*$^{-}$ mutant. Traces C-F represent cultures supplemented with 80 mg/L tryptophan analogs. Trace C: *N*-(1,1-dimethyl-1-allyl)-tryptophan, D: tryptophan, E: *N*-(1-methyl)-tryptophan, F: *N*-(1-propargyl)-tryptophan. Chromatograms represent the ESI-MS extracted ion for the $[M+Na]^+$ species as labeled. Note similar production levels of cyclomarazine M (**6,** trace E) and cyclomarazine P (**7**, trace F) in comparison to cyclomarazine A (**5**) in trace A and C. Production of desprenylcyclomarazine (expected $[M+Na]^+$ = 352) in the *cymD*$^{-}$ mutant has never been observed utilizing our assay conditions.



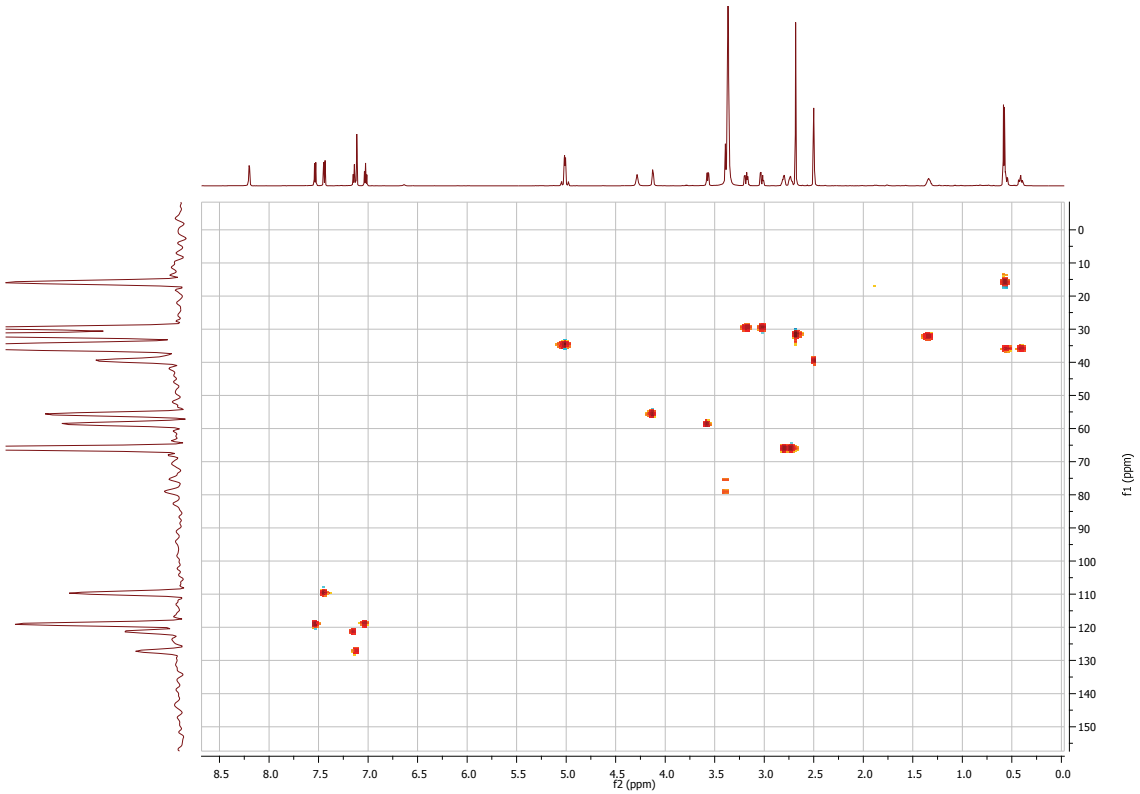


Appendix Figure 3.S3: ^{1}H-NMR comparison of *N*-(1,1-dimethyl-1-allyl-tryptophan (2) in DMSO-d_6

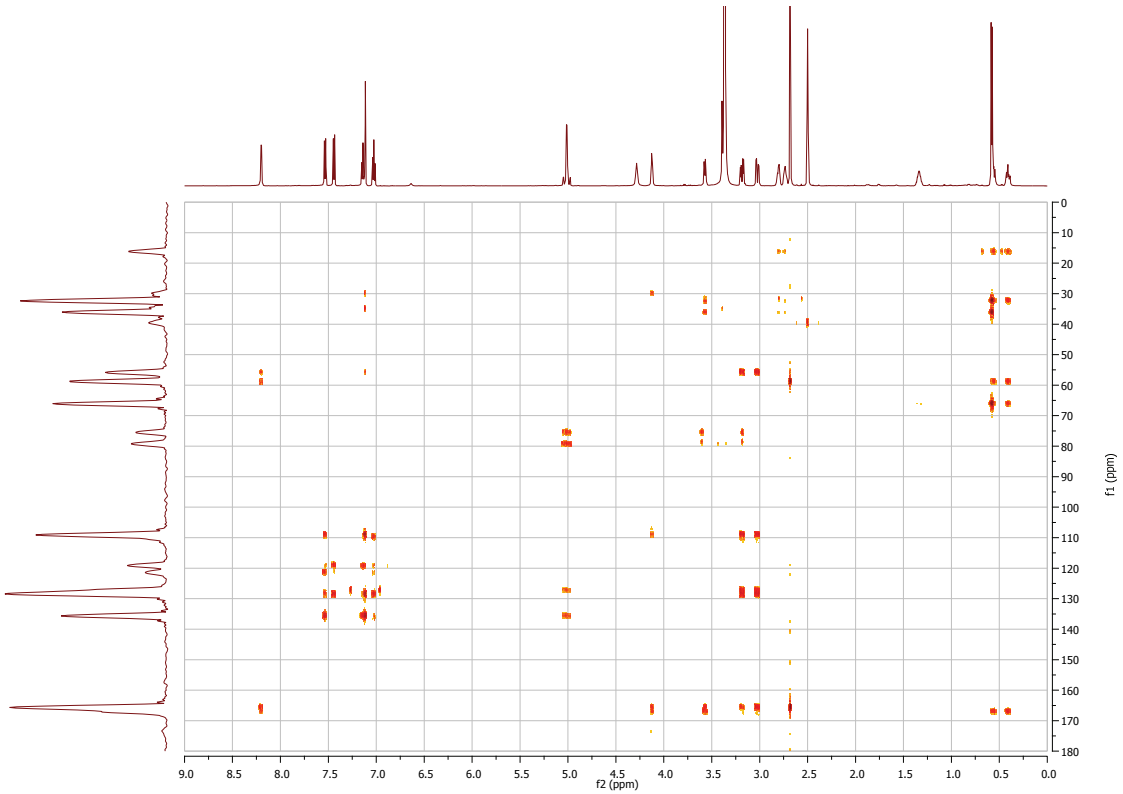




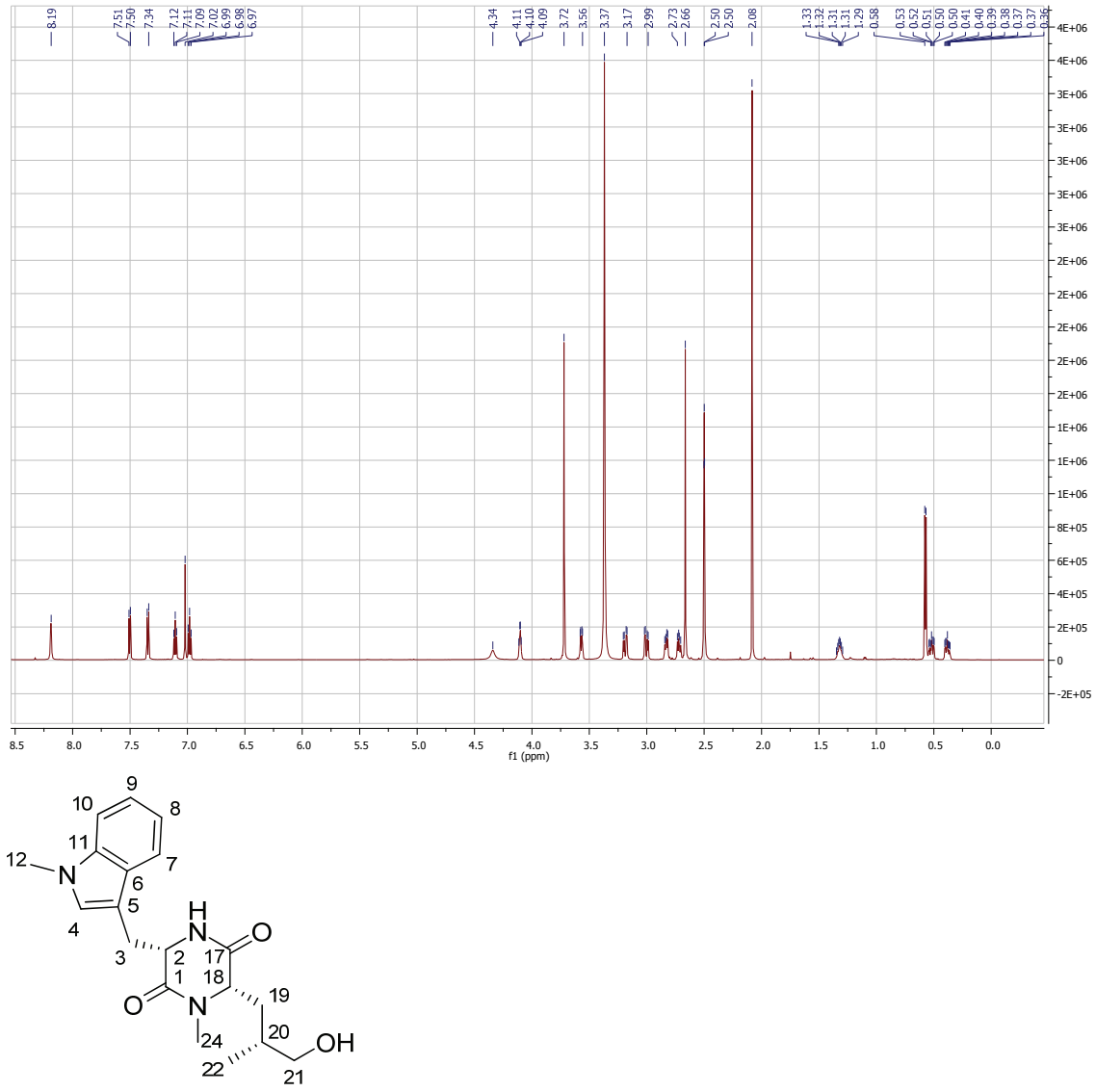
Appendix Figure 3.S4: ^{1}H-NMR of cyclomarazine P (7) in DMSO-d_6



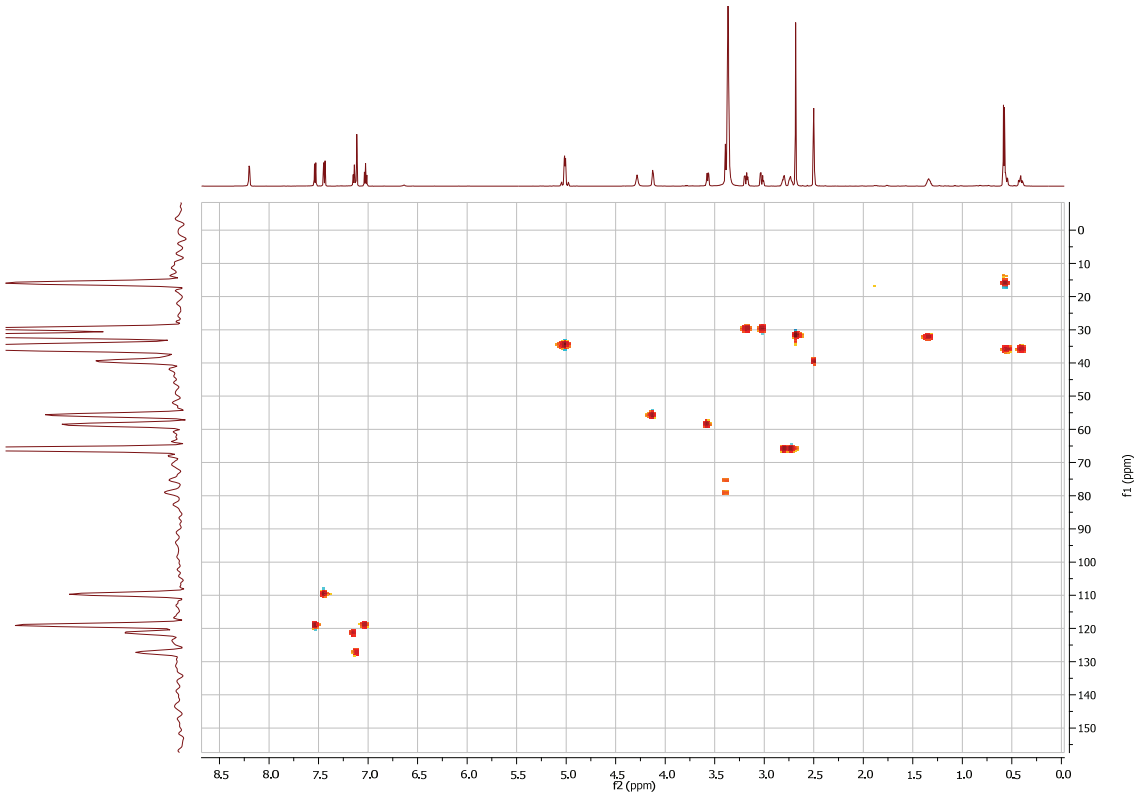
Appendix Figure 3.S5: HSQC NMR of cyclomarazine P (7) in DMSO-d_6



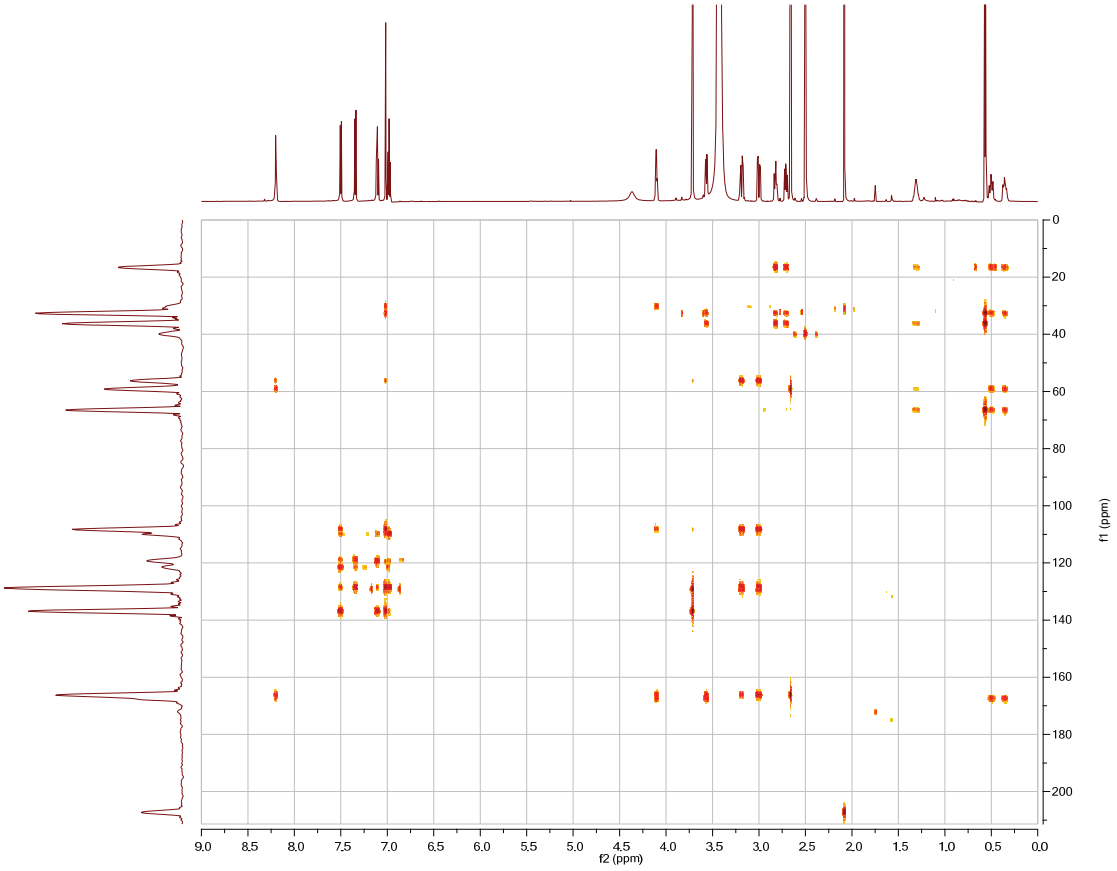
Appendix Figure 3.S6: HMBC NMR of cyclomarazine P (7) in DMSO-d_6



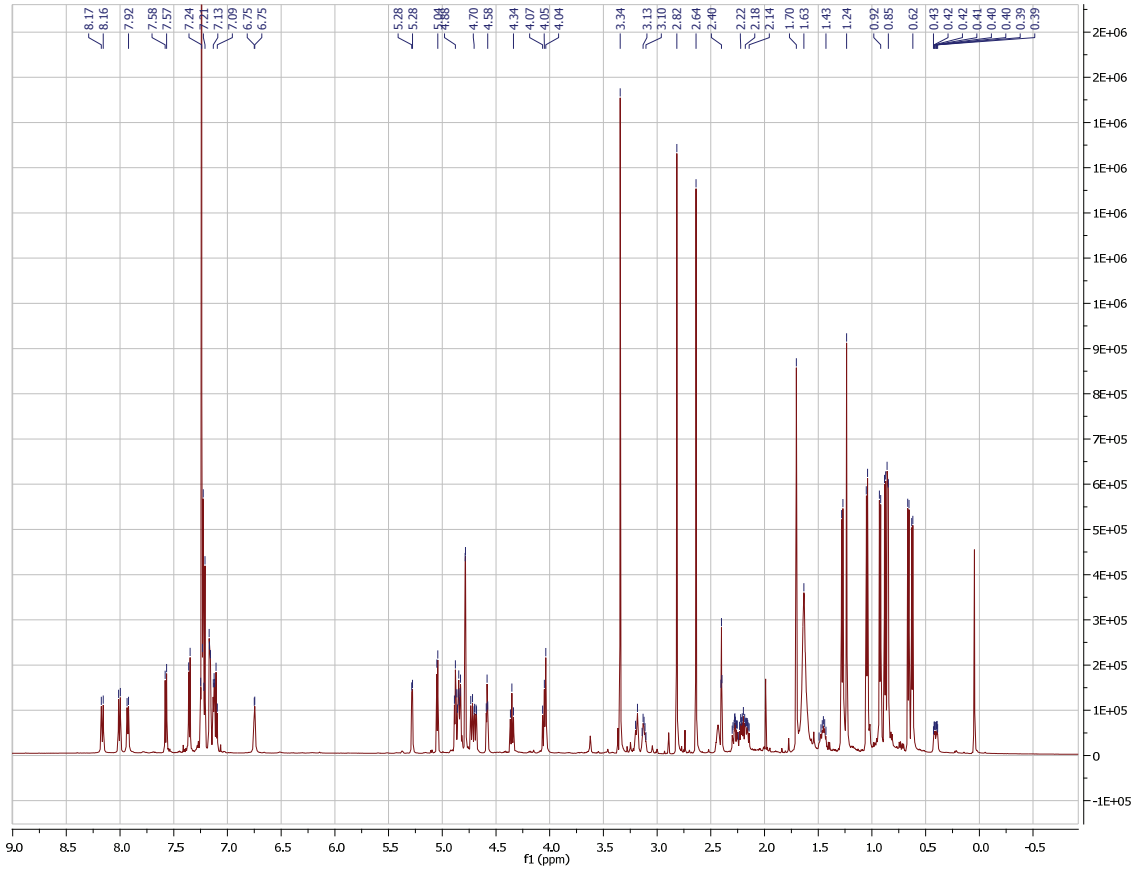
Appendix Figure 3.S7: ^{1}H-NMR of Cyclomarazine M (6) in DMSO-d_6

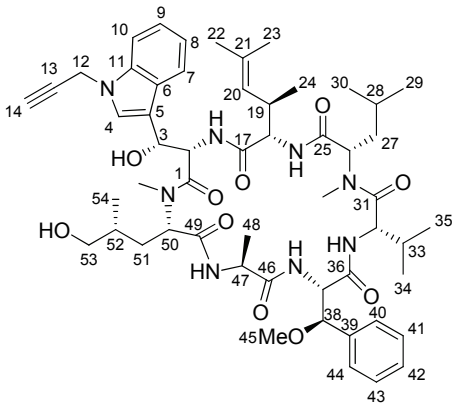


Appendix Figure 3.S8: HSQC NMR of cyclomarazine M (6) in DMSO-d_6

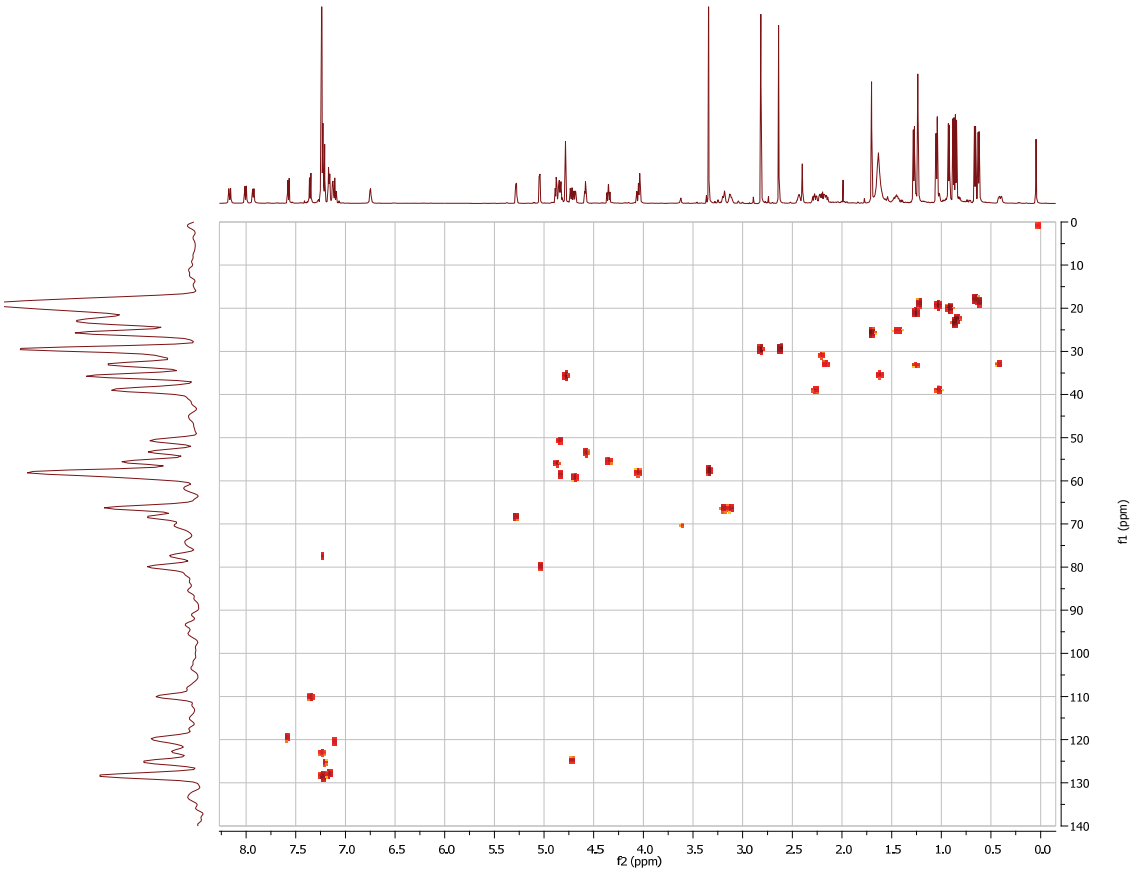


Appendix Figure 3.S9: HMBC NMR of cyclomarazine M (6) in DMSO-d_6

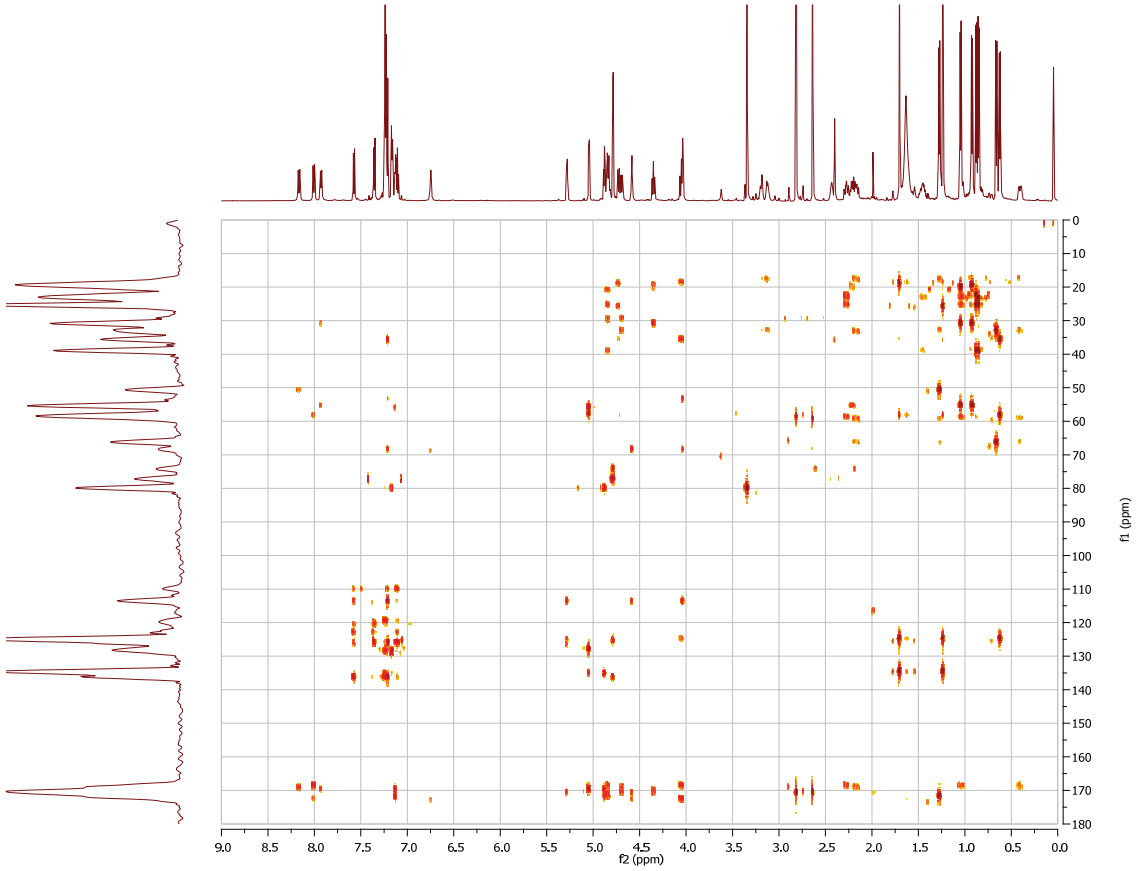




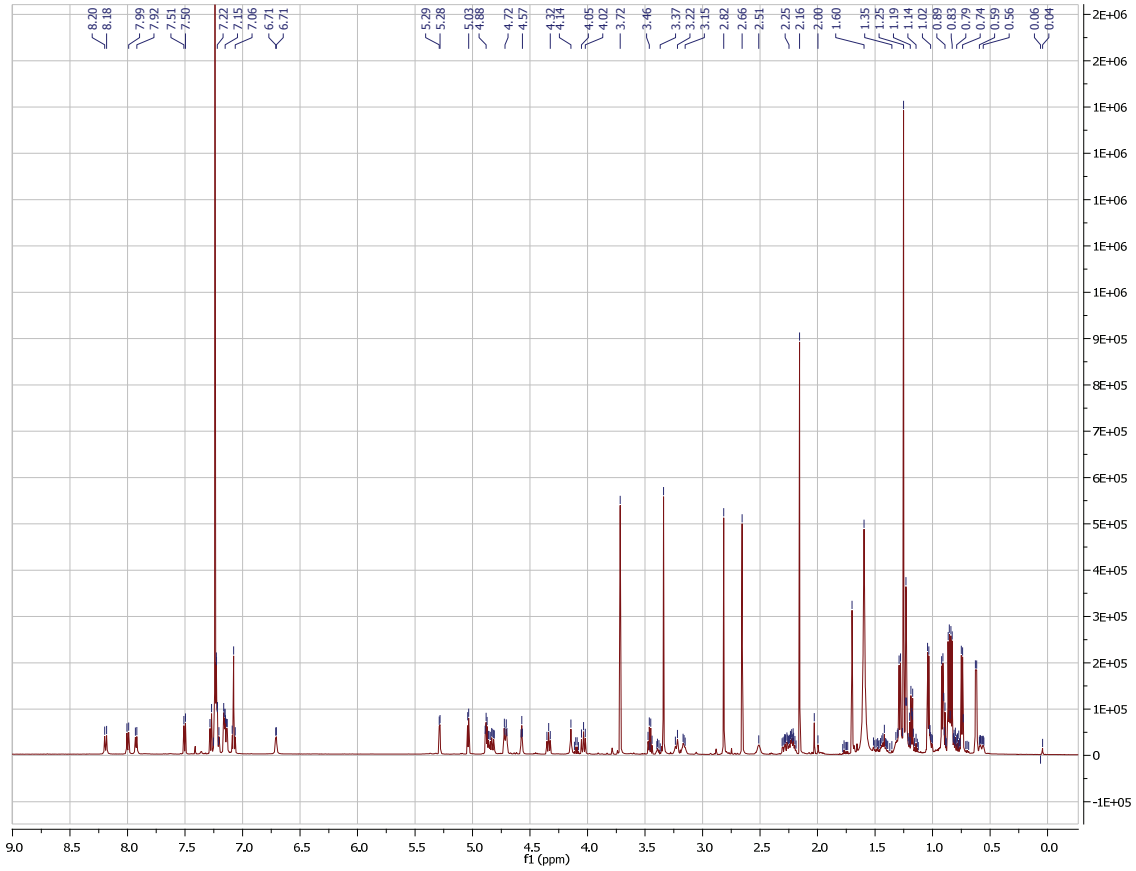
Appendix Figure 3.S10: ^{1}H-NMR of cyclomarin P (12) in $CDCl_3$

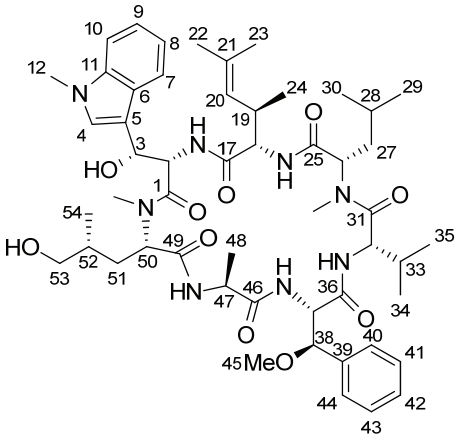


Appendix Figure 3.S11: HSQC NMR of cyclomarin P (12) in $CDCl_3$

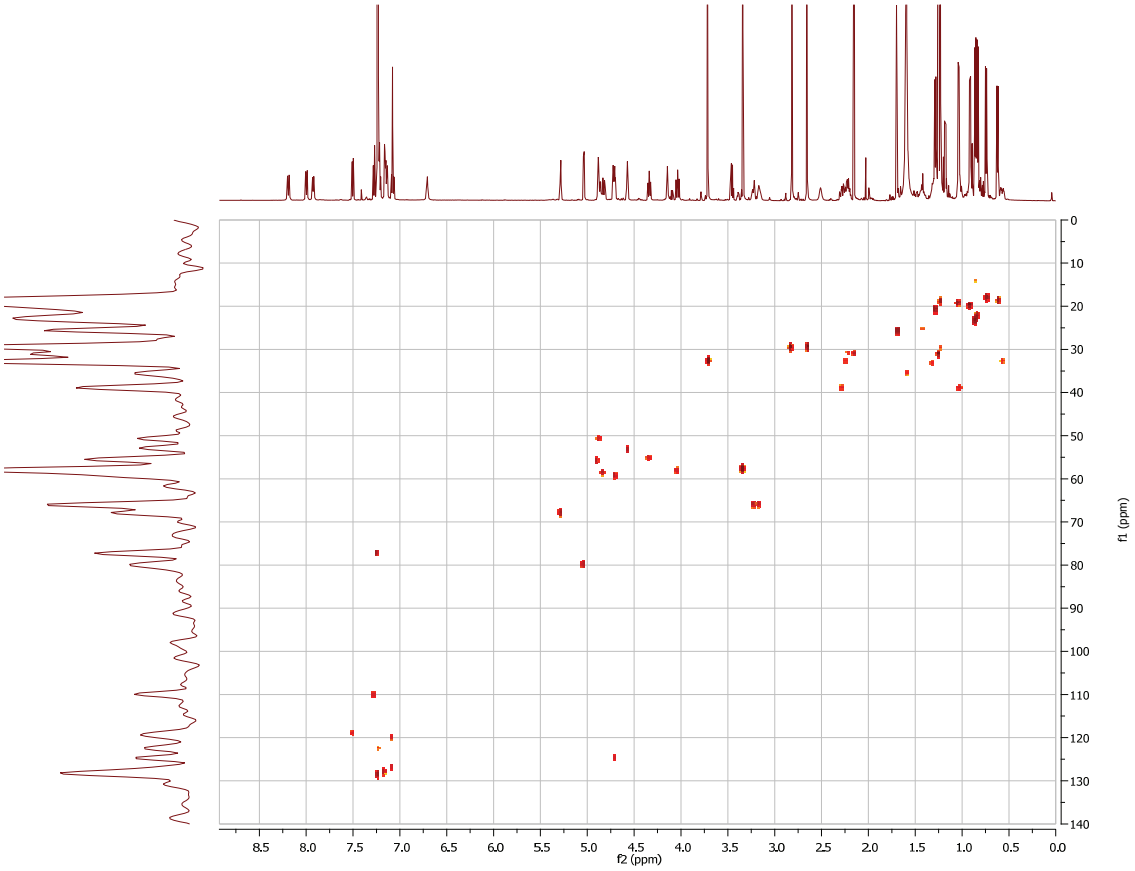


Appendix Figure 3.S12: HMBC NMR of cyclomarin P (12) in $CDCl_3$

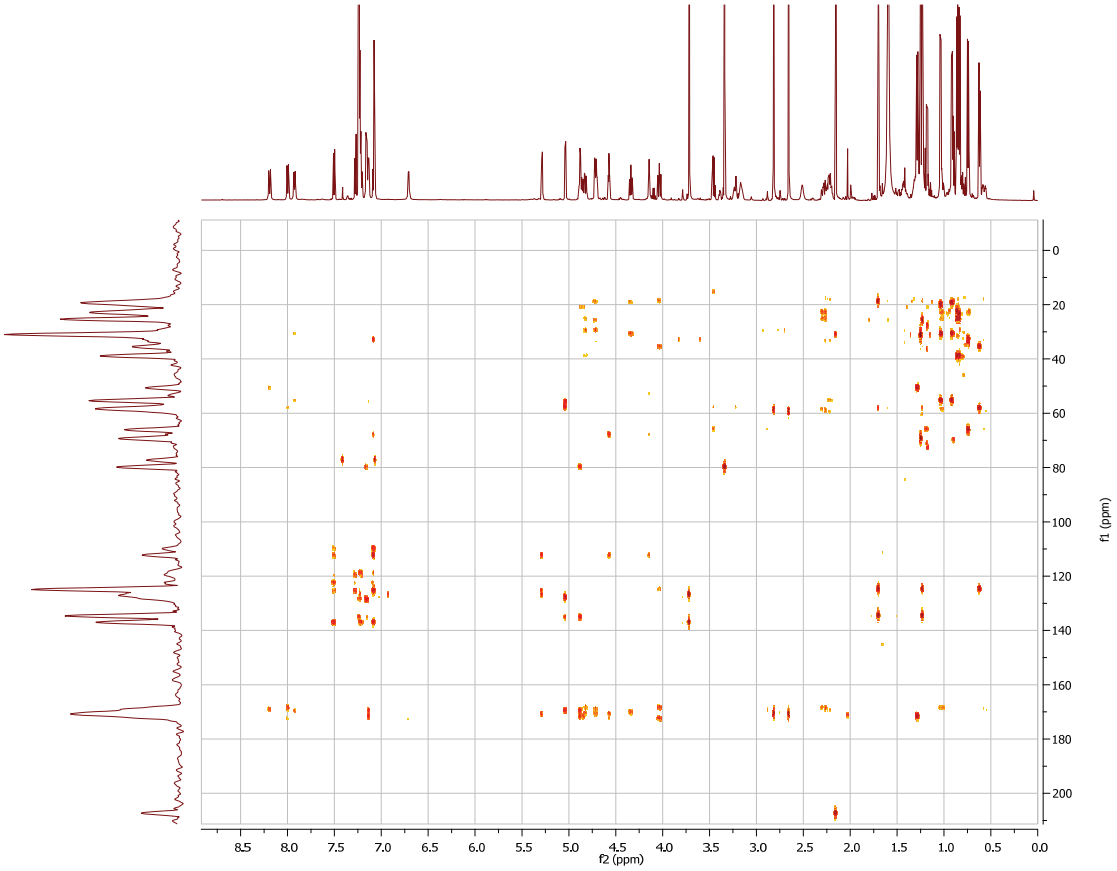




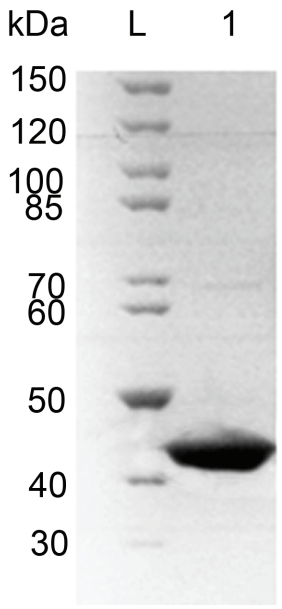
Appendix Figure 3.S13: ^{1}H-NMR of cyclomarin M (11) in $CDCl_3$



Appendix Figure 3.S14: HSQC NMR of cyclomarin M (11) in $CDCl_3$



Appendix Figure 3.S15: HMBC NMR of cyclomarin M (11) in $CDCl_3$



Appendix Figure 3.S16: Ni^{2+} affinity purified CymD. Image depicts the SDS-PAGE gel analysis of Ni^{2+} affinity purified CymD provided by *E. coli* BL21(DE3) harboring the plasmid pHIS8*cymD.* L: Fisher EZ-Run Rec Protein Ladder 1: Purified CymD. Gel is 4-20% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.7: References

(1) Renner, M.; Shen, Y.-C.; Cheng, X.-C.; Jensen, P.; Frankmoelle, W.; Kauffman, C.; Fenical, W.; Lobkovsky, E.; Clardy, J. Cyclomarins A-C, new antiinflammatory cyclic peptides produced by a marine bacterium (*Streptomyces* sp.). *J. Am. Chem. Soc.* **1999**, *121*, 11273-11276.

(2) Schultz, A.; Oh, D.-C.; Carney, J.; Williamson, R.; Udwary, D.; Jensen, P.; Gould, S.; Fenical, W.; Moore, B. Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **2008**, *130*, 4507-4516.

(3) Penn, K.; Jenkins, C.; Nett, M.; Udwary, D. W.; Gontang, E. A.; Mcglinchey, R. P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E. E.; Moore, B. S.; Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine actinobacteria. *ISME J.* **2009**, *3*, 1193-1203.

(4) Steffan, N.; Grundmann, A.; Yin, W.; Kremer, A.; Li, S. Indole prenyltransferases from fungi: a new enzyme group with high potential for the production of prenylated indole derivatives. *Curr. Med. Chem.* **2009**, *16*, 218-231.

(5) Zou, H.; Zheng, X.; Li, S.-M. Substrate promiscuity of the cyclic dipeptide prenyltransferases from *Aspergillus fumigatus*. *J. Nat. Prod.* **2008**, *72*, 44-52.

(6) Ruan, H. L.; Yin, W. B.; Wu, J. Z.; Li, S. M. Reinvestigation of a cyclic dipeptide *N*-prenyltransferase reveals rearrangement of N-1 prenylated indole derivatives. *ChemBioChem.* **2008**, *9*, 1044-1047.

(7) Kremer, A.; Westrich, L.; Li, S.-M. A 7-dimethylallyltryptophan synthase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterization. *Microbiology.* **2007**, *153*, 3409-3416.

(8) Kremer, A.; Li, S.-M. A tyrosine *O*-prenyltransferase catalyses the first pathway-specific step in the biosynthesis of sirodesmin PL. *Microbiology.* **2010**, *156*, 278-286.

(9) Nierman, W. C.; Pain, A.; Anderson, M. J.; Wortman, J. R.; Kim, H. S.; Arroyo, J.; Berriman, M.; Abe, K.; Archer, D. B.; Bermejo, C.; Bennett, J.; Bowyer, P.; Chen, D.; Collins, M.; Coulsen, R.; Davies, R.; Dyer, P. S.; Farman, M.; Fedorova, N.; Fedorova, N.; Feldblyum, T. V.; Fischer, R.; Fosker, N.; Fraser, A.; Garcia, J. L.; Garcia, M. J.; Goble, A.; Goldman, G. H.; Gomi, K.; Griffith-Jones, S.; Gwilliam, R.; Haas, B.; Haas, H.; Harris, D.; Horiuchi, H.; Huang, J.; Humphray, S.; Jimenez, J.; Keller, N.; Khouri, H.; Kitamoto, K.; Kobayashi, T.; Konzack, S.; Kulkarni, R.; Kumagai, T.; Lafton, A.; Latge, J.-P.; Li, W.; Lord, A.; Lu, C.; Majoros, W. H.; May, G. S.; Miller, B. L.; Mohamoud, Y.; Molina, M.; Monod, M.; Mouyna, I.; Mulligan, S.; Murphy, L.; O'neil, S.; Paulsen, I.; Penalva, M. A.; Pertea, M.; Price, C.; Pritchard, B. L.; Quail, M. A.; Rabbinowitsch, E.; Rawlins, N.; Rajandream, M.-A.; Reichard, U.; Renauld, H.; Robson, G. D.; De Cordoba, S. R.; Rodriguez-Pena, J. M.; Ronning, C. M.; Rutter, S.; Salzberg, S. L.; Sanchez, M.; Sanchez-Ferrero, J. C.; Saunders, D.; Seeger, K.; Squares, R.; Squares, S.; Takeuchi, M.; Tekaia, F.; Turner, G.; De Aldana, C. R. V.; Weidman, J.; White, O.; Woodward, J.; Yu, J.-H.; Fraser, C.; Galagan, J. E.; Asai, K.; Machida, M.; Hall, N.; Barrell, B.; Denning, D. W. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature.* **2005**, *438*, 1151-1156.

(10) Grundmann, A.; Li, S.-M. Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology.* **2005**, *151*, 2199-2207.

(11) Unsold, I. A.; Li, S.-M. Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. *Microbiology.* **2005**, *151*, 1499-1505.

(12) Yin, W.; Ruan , H.; Westrich, L.; Grundmann, A.; Li, S.-M. CdpNPT, an *N*-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation. *ChemBioChem.* **2007**, *8*, 1154-1161.

(13) Stec, E.; Steffan, N.; Kremer, A.; Zou, H.; Zheng, X.; Li, S. M. Two lysine residues are responsible for the enzymatic activities of indole prenyltransferases from fungi. *ChemBioChem.* **2008**, *9*, 2055-2058.

(14) Edwards, D.; Gerwick, W. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J. Am. Chem. Soc.* **2004**, *126*, 11432-11433.

(15) Liu, W.-T.; Ng, J.; Meluzzi, D.; Bandeira, N.; Gutierrez, M.; Simmons, T. L.; Schultz, A. W.; Linington, R. G.; Moore, B. S.; Gerwick, W. H.; Pevzner, P. A.; Dorrestein, P. C. Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Anal. Chem.* **2009**, *81*, 4200-4209.

(16) Kuzuyama, T.; Noel, J. P.; Richard, S. B. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature.* **2005**, *435*, 983-987.

(17) Metzger, U.; Schall, C.; Zocher, G.; Unsöld, I.; Stec, E.; Li, S.-M.; Heide, L.; Stehle, T. The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14309-14314.

(18) Kirschning, A.; Taft, F.; Knobloch, T. Total synthesis approaches to natural product derivatives based on the combination of chemical synthesis and metabolic engineering. *Org. Biomol. Chem.* **2007**, *5*, 3245-3259.

(19) Kolb, H. C.; Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today.* **2003**, *8*, 1128-1137.

(20) Ferezou, J.-P.; Quesneau-Thierry, A.; Servy, C.; Zissmann, E.; Barbier, M. Sirodesmin PL biosynthesis in *Phoma lingam* tode. *J. Chem. Soc., Perkin Trans. 1.* **1980**, 1739-1746.

(21) Sambrook, J.; Russell, D. *Molecular cloning: a laboratory manual*; Cold Spring Harbor Laboratory Press, 2001.

(22) Kieser, T.; Bibb, M.; Buttner, M.; Chater, K.; Hopwood, D. *Practical streptomyces genetics*; The John Innes Foundation Norwich, 2000.

(23) Jez, J. M.; Ferrer, J.-L.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry.* **2000**, *39*, 890-902.

(24) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **2005**, *41*, 207-234.

3.8: Acknowledgment

The detection of **2** in *S. arenicola* was originally observed by our UCSD colleagues R. Asolkar and W. Fenical (personal communication). We thank S. Richard and J. Noel for assistance with purification of CymD. **2** and **4** synthesized and provided by C. Lewis, M. Luzung, and P. Baran.

Chapter 3, in full, is a reprint of the material as it appears in Functional Characterization of the Cyclomarin/Cyclomarazine Prenyltransferase CymD Directs the Biosynthesis of Unnatural Cyclic Peptides (2010). Schultz, Andrew W.; Lewis, Chad A.; Luzung, Michael R.; Baran, Phil S.; Moore, Bradley S., Journal of Natural Products, 73, 373–377. The dissertation author was the primary investigator and author of this paper.

Chapter 4:

Efforts towards Functional Characterization and Biosynthetic Manipulation of the 2-Amino-3,5-dimethyl-4-hexenoic Acid Residue of Cyclomarin

4.1: Abstract

In vivo characterization of *cymF*, *cymG*, and *cymH* via PCR directed gene inactivation confirmed their role in the production of 2-amino-3,5-dimethyl-4-hexenoic acid and provided a novel cyclomarin in which this amino acid residue was replaced by phenylalanine. Extensive attempts to heterologously overexpress CymEF failed to provide soluble protein, impeding in vitro characterization of this pathway, although soluble CymG, CymH and Sare1130 were obtained. An attempt to reconstruct the ADH pathway in *Streptomyces coelicolor* is also described within.

4.2: Introduction

As mentioned in chapter 2, all of the known cyclomarins (Figure 4.1, **1**-**4**) contain the non-proteinogenic amino acid residue 2-amino-3,5-dimethyl-4-hexenoic acid (ADH). This aliphatic amino acid is similar to the branched chain amino acids leucine, valine and isoleucine, although ADH contains two more carbon atoms than the largest example of this class. Another unique feature of ADH is the unsaturation between the γ and δ carbons of the side chain, which is not observed in any aliphatic proteinogenic amino acid. Although ADH is unique to the cyclomarin natural products, a related amino acid, 2-amino-4-hexenoic acid, is a residue of the rufomycin[1] compounds K95-5076[2] and ilamycin[3] and is highlighted in Figure 4.2. Previous bioinformatic analysis coupled with stable isotope precursor incorporation studies (Chapter 2) confirmed a biosynthetic

origin that has more in common with the catabolism of 3-(3-hydroxyphenyl) propanoic acid[4] than the anabolism of branched chain amino acids.

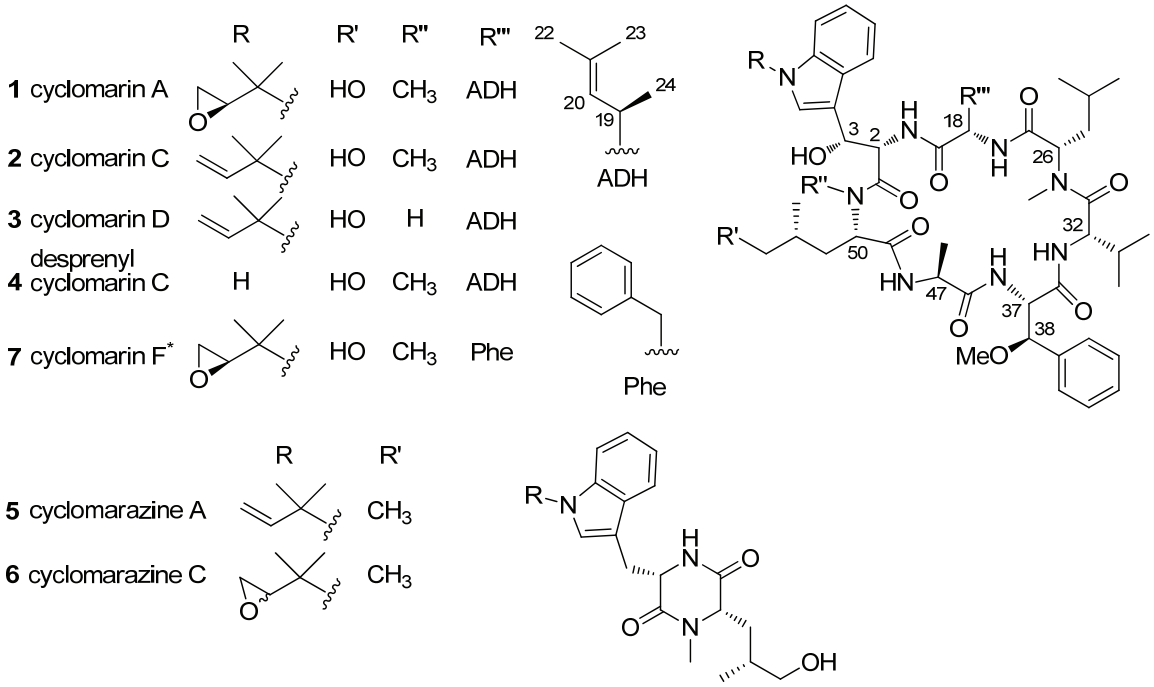


Figure 4.1: Structures of cyclomarin (1-4, 7) and cyclomarazine (5 and 6) cyclic peptides from *Salinispora arenicola* CNS-205 *Structures inferred by experimental evidence, but have not been fully characterized.

The putative pathway for the biosynthesis of ADH (Figure 4.3) begins with the conversion of the valine catabolite isobutyryl-CoA to isobutyraldehyde via CymE (Sare4566), which is channeled to its heterodimeric partner CymF (Sare4567) that catalyzes an aldol condensation with pyruvate to yield 4-hydroxy-5-methyl-2-oxo-hexanoic acid. Dehydration via CymH (Sare4569) yields 5-methyl-2-hydroxy-2,4-dihexenoic acid, which is methylated by the SAM-dependent enzyme CymG (Sare4568) to provide 3,5-dimethyl-2-oxo-4-hexenoic acid. Transamination via the branched chain amino acid aminotransferase Sare1130 would provide ADH. Precedence for “borrowing” the branched chain

amino acid aminotransferase from primary metabolism is observed in the biosynthesis of many nonproteinogenic amino acids, including the 3-methylglutamate residue[5] of calcium dependent antibiotic in *Streptomyces coelicolor* A3(2).

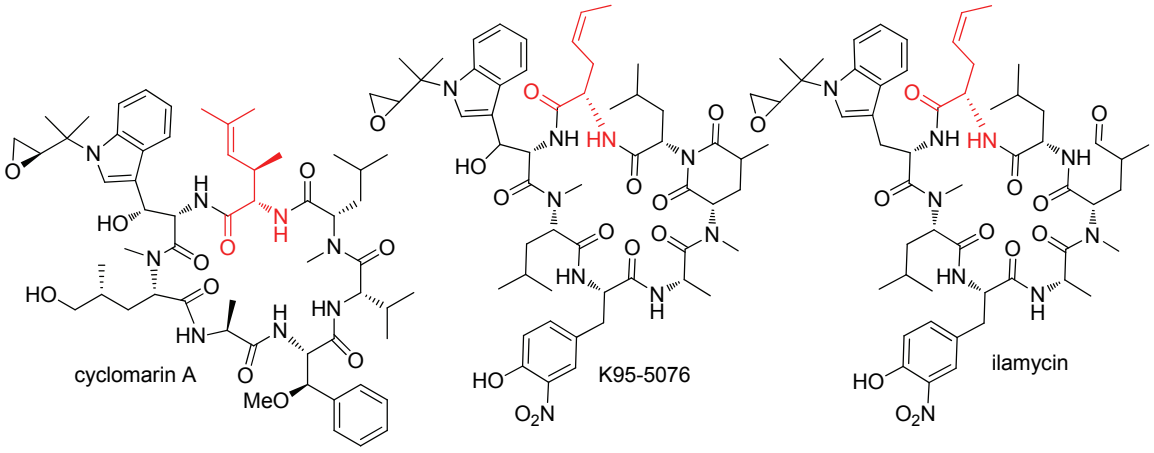


Figure 4.2: Structures of cyclomarin, K95-5076, and ilamycin, highlighting the non-proteinogenic branched chain amino acids contained within. The amino acid residues, highlighted in red, all contain a desaturation between the γ and δ carbons of the side chain, and are N-terminal to the *N*-(1,1-dimethyl-2,3-epoxypropyl)-tryptophan residue common to all three cyclic heptapeptides.

Bioinformatic analysis coupled with stable isotope precursor incorporation studies detailed in Chapter 2 provided support for this biosynthetic pathway to ADH. CymE, CymF, and CymH are significantly similar to MhpF, MhpE, and MhpD which are utilized by *Comamonas testosteroni* TA441 to degrade 3-hydroxyphenylpropionic acid via hydration and reverse aldol cleavage (Figure 4.3 A and B). The high level of similarity between the three enzymes and their proximity in their respective genomes implied similar functionality between the two pathways. Providing [U-^{13}C]glucose and [methyl-^{13}C]methionine to the cyclomarin producer *Streptomycetes*. sp. CNB-982 followed by NMR analysis of

the corresponding ^{13}C labeled cyclomarin A revealed an isotope incorporation pattern consistent with pyruvate, valine-derived isobutyryl-CoA, and SAM-dependent methylation. Supplementing cultures of *S. arenicola* CNS-205 with sodium [1-^{13}C]isobutyrate, likely converted to isobutyl-CoA via a native CoA-ligase, provided cyclomarin A in which the γ carbon (C20) of the ADH residue was solely and specifically enriched, further supporting this hypothetical pathway.

Hydroxypyridyl-homothreonine (HPHT), a component of nikkomycin X (Figure 1.7), is the only other non-proteinogenic amino acid known to be produced via an aldolase/dehydrogenase pair. In the biosynthesis of this amino acid, the dehydrogenase SanN converts picolinate-CoA to picolinaldehyde, which is condensed with 2-oxobutyrate via the SanM aldolase, providing the α-ketoacid 4-pyridyl-4-hydroxy-2-oxoisovalate (POHIV)[6]. Yeast two-hybrid assays and immunoprecipitation confirmed that SanMN forms a heterodimer. The SanMN complex was shown to condense picolinaldehyde and 2-oxobutyrate to form POHIV in vitro, and SanN showed dehydrogenase activity towards the structural analog of picolinate-CoA, benzoate-CoA, to provide benzaldehyde. Further oxidation and transamination yielded HPHT. Blast analysis of CymE to SanN and CymF to SanM reveal 40% and 50% identity, respectively, suggesting these enzyme pairs are functionally related.

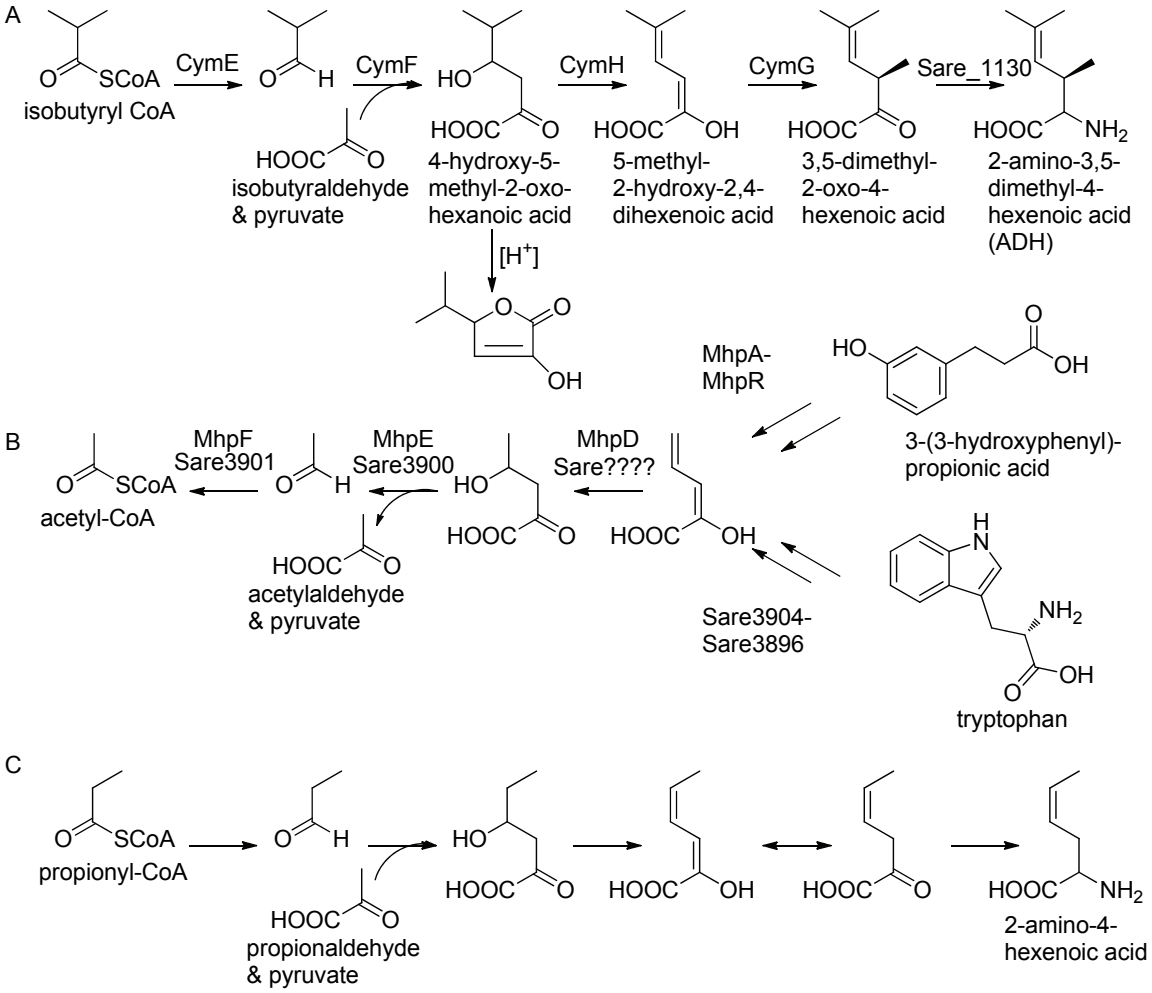


Figure 4.3: Aldolase/dehydrogenase pathways related to ADH biosynthesis. (A): Isobutyryl-CoA from valine metabolism is converted to isobutyraldehyde via CymE, which in turn is channeled to CymF, which catalyzes the aldol condensation between isobutyraldehyde and pyruvate to yield 4-hydroxy-5-methyl-2-oxo-hexanoic acid. The dehydratase CymH catalyzes the formation of the diene 5-methyl-2-hydroxy-2,4-dihexenoic acid, which in turn is methylated by the SAM dependent methyltransferase CymG. Transamination by the branched chain amino acid aminotransferase Sare1130 provides the amino acid, ADH. Acid catalyzed lactonization could be utilized to convert 4-hydroxy-5-methyl-2-oxo-hexanoic acid to the corresponding lactone, facilitating analysis. (B) depicts the 3-hydroxyphenylpropionic acid degradation pathway from *Comamonas testosteroni* TA441and the tryptophan degradation pathway present in *S.arenicola* CNS-205 which relies on an aldolase/dehydrogenase pair with similar function to CymEF. (C) depicts the proposed pathway to the 2-amino-4-hexenoic acid, a constituent of K95-5076 and the ilamycins.

4.2.1: Hydratases, aldolases, and dehydrogenases and their role in aromatic degradation

Several studies have been conducted on MhpD-F and their homologs, which operate downstream of the *Meta* fission product (MFP) hydrolases[7] in aromatic degradation pathways. MhpD from *Escherichia coli* W3110/pTB7 was characterized as a Mn^{2+}-dependent 2-hydroxypentadienoic acid hydratase[8] to provide 4-hydroxy-2-keto-pentanoate. A crystal structure of HpcG,[9] which catalyzes the hydration of 2-ketohepta-4-ene-1,7-dioate to 4-hydroxy-2-ketoheptane-1,7-dioate, supported the role of Mn^{2+} in catalysis.

Functional characterization of MhpE confirmed its role as a 4-hydroxy-2-keto-pentanoic acid aldolase.[10] Expression of MhpE and MhpF in *E. coli* is coupled and MhpF is required for proper folding of MhpE[11]. The related 4-hydroxy-2-ketovalerate aldolase DmpG and acylating acetaldehyde dehydrogenase DmpF have been shown to form a heterodimer when co-crystallized, and a channel appropriate for the transfer of acetaldehyde between the active sites of DmpG and DmpF is present[12], preventing the exposure of the cell to this reactive intermediate. BphI is also related to DmpG and MhpE and was shown via in vitro enzymology to channel acetylaldehyde to its dehydrogenase partner BphJ[13]. All three are type II, Mn^{2+}-dependent aldolases.[10,11,13]

The MhpF related dehydrogenase BphJ[14] and DmpF[12] utilize the coenzyme nicotinamide adenine dinucleotide (NAD^+) as an oxidizing agent in the conversion of acetyladehyde to acetyl-CoA. BphIJ was shown to accept longer

chain aldehydes, providing the corresponding acyl-CoA, albeit with a higher K_m than that observed when 4-hydroxy-2-oxopentanoate was provided BphIJ. This study confirmed the role of BphIJ as an acylating aldehyde dehydrogenase and supporting a substrate channeling mechanism[14].

The recently discovered pathway for tryptophan catabolism characterized in *Burkholderia cepacia* J2315[15] utilizes a 2-ketopentenoate hydratase, 2-keto-4-hydroxypentanoate aldolase, and an acetaldehyde dehydrogenase to provide pyruvate and acetate in an analogous manner to the Mhp pathway. Interestingly, *Salinispora arenicola* and *Salinispora tropica,* along with some streptomycetes, contain the genes that constitute this pathway, as highlighted in Figure 4.3B. The presence of this route for Trp degradation in actinobacteria may suggest its role as the progenitor to the pathways for ADH and HPHT production.

There is significant interest in utilizing aldolases in asymmetric synthesis[16-18] due to their ability to form carbon-carbon bonds in a stereospecific manner without the production of side products. In a study comparing BphI to HpaI[19], another type II aldolase that converts 4-hydroxy-2-oxo-1,7-heptanedioate to pyruvate and succinic semialdehyde, the earlier was shown to catalyze aldol condensations with a higher level of stereospecificity, while the latter has broader substrate specificity and a higher turnover.[13] Aldolases are generally specific for their donor compound, such as pyruvate, but have relaxed specificity in the chain length acceptor molecule, as long as they contain the necessary aldehyde functionality[14,17].

Because of the similarity between the ADH and 2-amino-4-hexenoic acid residues in the structurally related heptapeptides cyclomarin and rufomycin, one could also speculate on the involvement of a similar pathway towards this non-proteinogenic amino acid (Figure 4.3c). Replacing isobutyryl-CoA with propionyl-CoA in a pathway lacking a methyltransferase could provide 2-amino-4-hexenoic acid. Indeed, BphI has been shown to catalyze the formation of propionyl-CoA from propionaldehyde[14], suggesting that longer chain aldehydes could be accommodated by this class of enzymes.

4.2.2: Specific aims

On basis of the similarity between the ADH and HPHT anabolic pathways and the numerous aromatic degradation pathways, we set forth to characterize the biosynthesis of this unique, unsaturated, branched chain amino acid component of cyclomarin. Building upon the bioinformatic analysis and precursor incorporation studies reported in chapter 2, we set forth to further characterize the substrate specificity of this pathway with a goal to harness it to generate unnatural amino acids and unnatural cyclomarin analogs. Here we report a comprehensive investigation of the ADH pathway and its role in cyclomarin biosynthesis involving in vivo functional analysis and heterologous protein expression.

4.3: Results

4.3.1: Inactivation of ADH biosynthetic genes and partial characterization of the novel cyclomarin F.

To characterize the in vivo function of CymF, CymG, and CymH, we inactivated each encoding gene by PCR-targeted gene replacement[20] as individual *S. arenicola* CNS-205 clones. Replacement of *cymF* with an apramycin resistance/oriT cassette in the pCC1Fos-based clone BPPW978 harboring a fragment of the *cym* cluster containing *cymF-cymH* gave the *S. arenicola cymF*$^-$ knockout mutant. A similar procedure provided the *S. arenicola cymG*$^-$ *and cymH*$^-$ mutants. HPLC-MS analysis of the mutants verified elimination of production of the known cyclomarins A, C, and D (**1**-**3**) while maintaining cyclomarazine A and C (**5**,**6**) production (Figure 4.4), supporting their putative role in ADH biosynthesis. Interestingly, we saw the production of a novel peak in all three ADH deficient mutants with a ESI-MS ionization pattern of $[M + Na]^+$ 1073 and $[M\text{-}H_2O + H]^+$ of 1033, which is diagnostic for cyclomarin compounds containing the β-hydroxylated *N*-prenylated tryptophan residue[21]. This chemical species, deemed cyclomarin F (Figure 4.1, **7**), is 8 Da heavier than **1** that corresponds well with the substitution of ADH with phenylalanine. Replacement of ADH with a proteinogenic amino acid might be expected, since inactivation of the prenyltransferase CymD provided desprenylcyclomarin C (**4**) where Trp was incorporated when the preferred substrate, *N*-(1,1-dimethyl-1-allyl)-tryptophan, was no longer produced (Chapters 2 and 3).

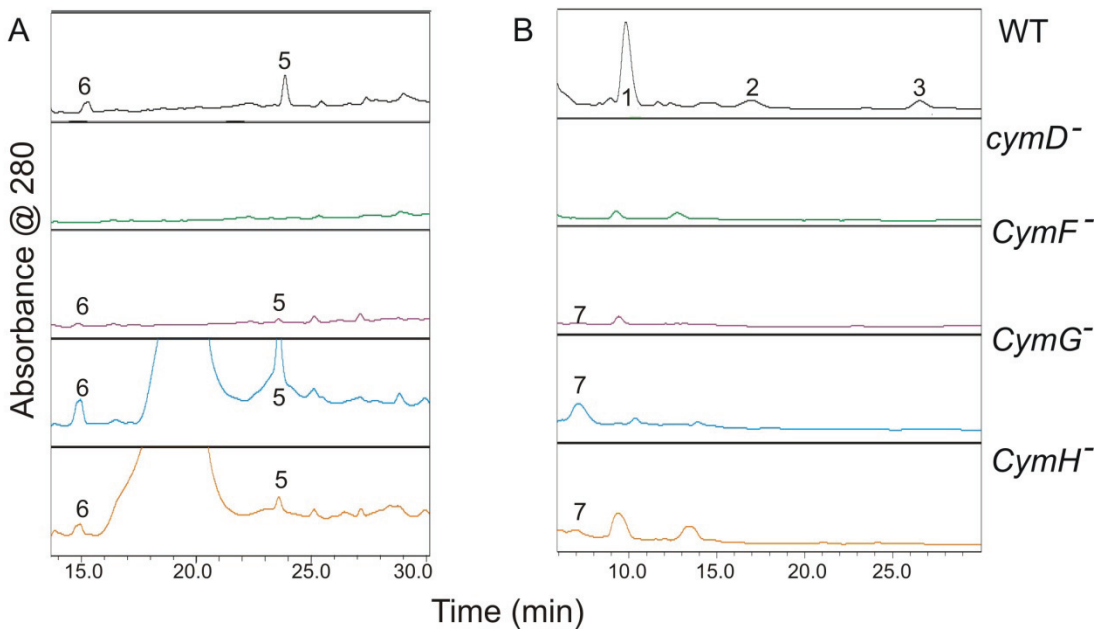


Figure 4.4: LC-MS analysis of the organic extracts of *S. arenicola* CNS-205 (WT) and the *S. arenicola* CNS-205 mutants *cymD*$^-$, *cymF*$^-$, *cymG*$^-$, *cymH*$^-$. Chromatograms recorded at 210 nm. (A) cyclomarazine analysis, (B) cyclomarin analysis. See materials and methods section for details. Note the complete loss of peaks corresponding to **1**-**3, 5** and **6** in the *cymD* replacement mutant. Inactivation of *cymF, cymG* or *cymH* had no effect on the production the cyclomarazines **5** or **6**, while abolishing the production of known cyclomarins **1**-**3**. Note the appearance of the novel metabolite predicted to be cyclomarin F (**7**) in all three mutants of the ADH biosynthetic pathway with a mass of 1073 for the $[M+Na]^+$ ion.

Although we did isolate and partially characterize **7** by ESI-MS (m/z) 1073.5 $[M + Na]^+$, 1073.6 calcd), and NMR, complete structural characterization could not be completed due to insufficient material of adequate purity which degraded upon standing in $CDCl_3$. **7** was produced in the *cymG*$^-$ mutant at approximately 10-fold lower concentration than that of the corresponding **1** in the wild-type strain. ^{1}H-NMR (Appendix figure 4.S1) confirmed **7**'s identity as a cyclomarin analog containing resonances appropriate for amide, indole, and aromatic protons in the 8.5-6.5 ppm region which compared favorably to the

reported structure of **1**.[22] Interference due to the chloroform solvent peak and impurities made precise integration of these resonances impossible, which would have confirmed the presence of Phe in addition to the *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-hydroxytryptophan and β-methoxyphenylalanine residues. Analysis of the 5.5-4.5 ppm region contained two resonances appropriate for protons connected to oxidized β-carbons and seven resonances for α-protons, two sets of which were overlapping (Table 4.1). HSQC analysis was utilized to determine the chemical shifts of the carbons attached to these protons, the values of which compare favorably to values determined for **1**[22] (See Table 4.1). By elimination, the amino acid replacing ADH has a δ_H of 4.48 and a δ_C of 54.8 for C18, which is significantly different than the values reported for the ADH residue (C18, δ_H = 3.22 and δ_C = 57.7) of **1**. Also of note is the lack of a resonance corresponding to the proton attached to γ position of ADH (C20, expected δ_H = 4.77 and δ_C = 124.7), in this region confirms the loss of this residue in **7**. Signals appropriate for H19 (δ_H = 1.66) H22 (δ_H = 1.26) and H23 (δ_H = 1.73) of **1** were not observed in the 2.0-0.0 ppm region of the 1D ^{1}H and HSQC spectra. Surprisingly, a signal appropriate for position 24 was observed (δ_H = 0.63 and δ_C = 17.7 observed, δ_H = 0.64 and δ_C = 18.5 reported for **1**), although the quantitation is below what would be expected for a methyl group and is most likely an impurity peak, since numerous superfluous signals appear in this region. An HMBC data set was collected (data not shown), but it was of insufficient quality to confirm the identity of any of the amino acid residues of cyclomarin.

Table 4.1: Partial NMR data for cyclomarin F in $CDCl_3$, characterization of α- and select β- resonances

	cyclomarin F (**7**)		cyclomarin A (**1**)[d]	
position	δ_C[a]	δ_H[b], (*J* in Hz)	δ_C[a]	δ_H[b], (*J* in Hz)
2	53.4	4.62 (m)	53.3	4.58 (t, 3)
3	68.7	5.19 (d, 4.8)	68.7	5.31 (d, 2.5)
18	54.8	4.48 (m)[c]	58.1	4.08 (t, 10)
26	58.8	4.71 (m)[c]	58.7	4.83 (t, 3.5)
32	54.8	4.48 (m)[c]	55.3	4.36 (t, 8.5)
37	56.0	4.93 (t, 4.8)	55.9	4.90 (t, 5)
38	79.9	5.09 (d, 5.3)	80.1	5.08 (d, 5.5)
47	50.6	4.87 (ddd,14.3, 10.0, 7.1)	50.6	4.88 (m)
50	58.8	4.71 (m)[c]	59.2	4.81 (t, 8)

[a]Assignment by HMQC correlations. [b]600 MHz. [c]overlapping resonances
[d]as reported in Renner et al.[22]

4.3.2: Progress towards the heterologous overexpression and in vitro reconstitution of the ADH pathway

In order to further characterize CymE-H and Sare1130 in vitro, we set forth to heterologously overexpress each of these enzymes as histidyl-tagged recombinant proteins in *E. coli*. Each of the five corresponding genes were individually cloned into the pHIS8 vector[23]. Expression of pHIS8*cymG* and pHIS8*sare1130* in *E.coli* followed by Ni^{2+}-affinity purification successfully provided soluble octahistydyl-tagged Sare1130 (40 kDa) and CymG (35 kDa). Optimal expression was achieved in the *E. coli* strain BL21(DE3) utilizing ZYP-5052 autoinduction media[24] at 37 °C and 30 °C respectively. Coomassie stained SDS-PAGE gel analysis of these purified proteins is shown in Figure 4.5. Sare1130 was produce at 8 mg/L, while CymG was produced in a concentration of 6 mg/L.

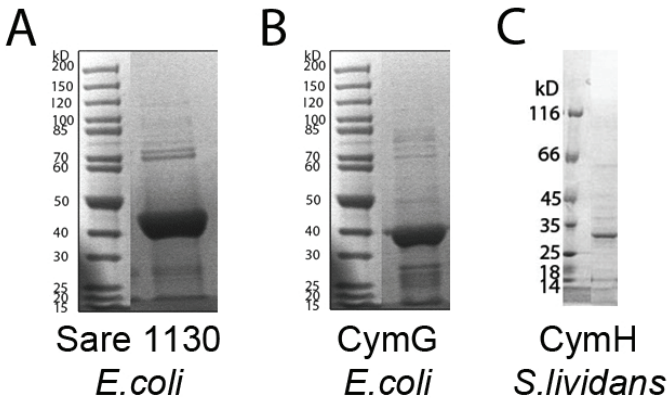


Figure 4.5: Ni^{2+} affinity purified recombinant proteins. (A) depicts the SDS-PAGE gel analysis of Ni^{2+} affinity purified Sare 1130 provided by *E. coli* BL21(DE3) harboring the plasmid pHIS8*sare1130*. (B) CymG provided by pHIS8*cymG* in *E. coli* BL21(DE3). (C) CymH from heterologous expression of the *E. coli- Streptomyces* shuttle vector pXY200oriT*cymH* in *Streptomyces lividans*TK24.

Unfortunately, overexpression of pHIS8*cymE*, pHIS8*cymF*, and pHIS8*cymH* in BL21(DE3) did not provide soluble protein (data not shown). In the case of CymE and CymF, this was not surprising, since it has been shown that production of MhpF and MhpE is transcriptionally coupled and that this heterodimeric pair must be produced together to provide soluble protein.[11] To attempt co-expression of *cymE* and *cymF*, two plasmids were generated, pHIS8*cymEF* and pETDuet*cymEF*, in which *cymE* and *cymF* is produced under the control of the same and individual *lac*T7 promoters, respectively.[25] Expression of pHIS8*cymEF* and pETDuet*cymEF* in BL21(DE3) cells in utilizing autoinduction at various temperatures unfortunately did not provide soluble protein (data not shown).

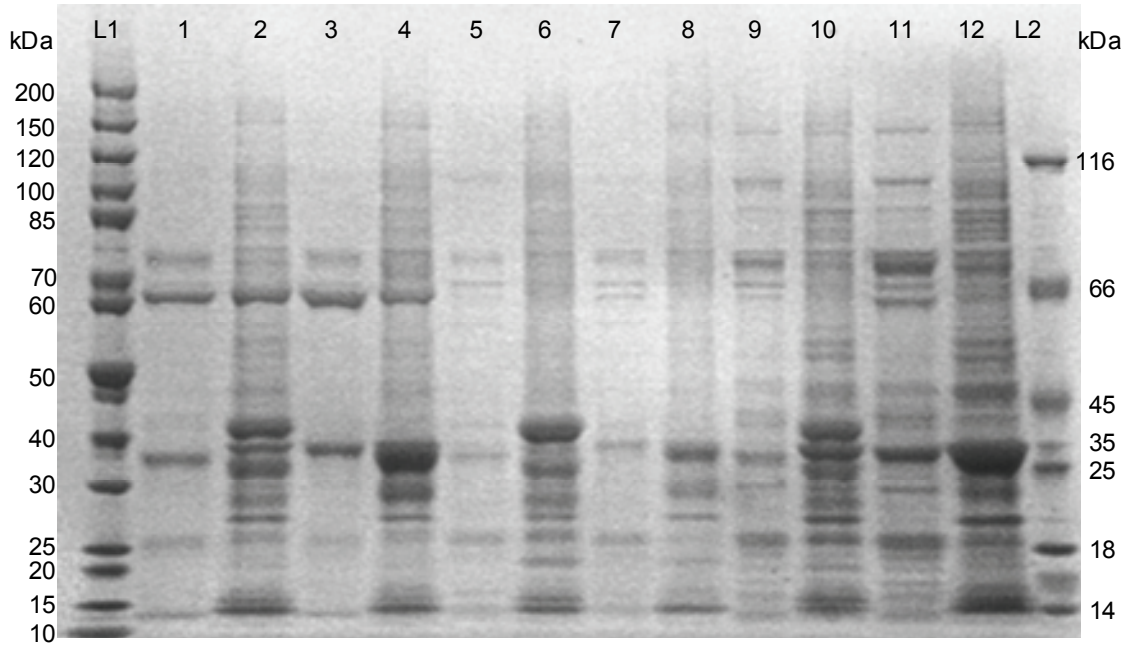


Figure 4.6: SDS-PAGE gel of pilot expression experiments to provide soluble CymE and CymF recombinant proteins. See Table 4.2 for sample identities and expected protein sizes. Note the production of both CymE and CymF in the insoluble fractions in the majority of the samples, but the lack of soluble CymF in the Ni^{2+} affinity chromatography. It is expected that CymE and CymF form a heterodimer that would co-purify if they were soluble and properly folded.

A large array of different *E. coli* expression strains, optimized for low temperature production (Arctic Express), reducing cytoplasmic conditions (SHuffle T7), or optimized codon usage and repressed basal production (Rosetta 2 pLysS) were explored in attempts to provide soluble CymEF. Figure 4.6 depicts a Coomassie stained SDS-PAGE gel utilized in analyzing some of these experiments, the samples for which are described in Table 4.2. Note in lane 1 and 3, which contain the Ni^{2+} purified soluble fraction obtained via 10 °C expression in Arctic Express cells, contain bands appropriate for soluble CymE. If CymF was soluble and properly folded, we would expect it to co-purify with CymE in nickel affinity chromatography,[11] which is not observed in this gel. No

products were observed when these fractions were assayed for aldolase activity (data not shown), further supporting that soluble, functional CymF is not being produced in *E. coli*.

Table 4.2: Combinations of *E. coli* strains and *cymE* and *cymF* containing vectors utilized in heterologous overexpression experiments

lane	*E. coli* strain	vector	sample	CymE[a] (kD)	Cym F[a] (kD)	temp °C	IPTG (mM)
1	Arctic Express	pETDuetcymEF-R	Ni^{2+} eludate	32.8	39.8	10	1.0
2	Arctic Express	pETDuetcymEF-R	insoluble	32.8	39.8	10	1.0
3	Arctic Express	pHIS8cymEF	Ni^{2+} eludate	33.8	36.6	10	1.0
4	Arctic Express	pHIS8cymEF	insoluble	33.8	36.6	10	1.0
5	SHuffle T7	pETDuetcymEF-R	Ni^{2+} eludate	32.8	39.8	16	0.4
6	SHuffle T7	pETDuetcymEF-R	insoluble	32.8	39.8	16	0.4
7	SHuffle T7	pHIS8cymEF	Ni^{2+} eludate	33.8	36.6	16	0.4
8	SHuffle T7	pHIS8cymEF	insoluble	33.8	36.6	16	0.4
9	Rosetta 2 pLysS	pETDuetcymEF-R	Ni^{2+} eludate	32.8	39.8	16	0.4
10	Rosetta 2 pLysS	pETDuetcymEF-R	insoluble	32.8	39.8	16	0.4
11	Rosetta 2 pLysS	pHIS8cymEF	Ni^{2+} eluate	33.8	36.6	16	0.4
12	Rosetta 2 pLysS	pHIS8cymEF	insoluble	33.8	36.6	16	0.4

[a]CymE contains a N-terminal His_8 tag from the pHIS8 construct, and a His_6 tag from the pETduet construct, while CymF contains a C-terminal S-tag from the pETduet construct. This accounts for the expected differences in peptide size.

Attention shifted towards utilizing *Streptomyces lividans* TK24 to produce soluble CymE, CymF, and CymH. Co-expression of SanNM in *S. lividans* TK24 was successfully utilized by Lee et al.,[11] and we expected a similar approach would provide CymEF. The pXY200[26] derivative pXY200oriT was provided by Tobias Guilder (Moore Laboratory, Scripps Institution of Oceanography), which allows for conjugation from *E. coli* into streptomycetes and provides a His_6 tag and thiostrepton inducible overexpression. Cloning of *cym* genes into pXY200oriT provided pXY200oriT*cymE*, pXY200oriT*cymF,* pXY200oriT*cymEF*, and pXY200oriT*cymH*. Expression of pXY200oriT*cymH* provided a soluble

product of the expected size (29 kDa, Figure 4.5 panel C, yield not determined), while expression of *cymE* and *cymF* together and individually in *S. lividans* failed to provide soluble recombinant enzyme.

4.3.3: Progress towards in vivo heterologous production of ADH

Since attempts to produce recombinant CymEF via heterologous overexpression in *E. coli* and *S. lividans* for in vitro analysis was unfruitful, other approaches towards characterization of the ADH pathway were explored. We decided to pursue heterologous in vivo pathway expression in *Streptomyces coelicolor* YU105[27] utilizing pRM5[28] as a host vector. This vector has been utilized successfully to produce actinorhodin precursors,[28] enterocin,[29] and the rifamycin[30] precursor 3-amino-5-hydroxybenzoic acid. We created two derivatives, pRM5*cymEF* and pRM5*cymEFGH* (Figure 4.7) in which the *act* genes from *act*KS/At through *act*IV, contained within PacI and EcoRI sites, were replaced with *cymEF* or *cymEFGH* containing the native *cymE* ribosome binding site , and transformed these vectors into protoplasts of *Streptomyces coelicolor* YU105.[27] We assumed that the native branched chain amino acid aminotransferase could convert 3,5-dimethyl-2-oxo-4-hexenoic acid to ADH, and

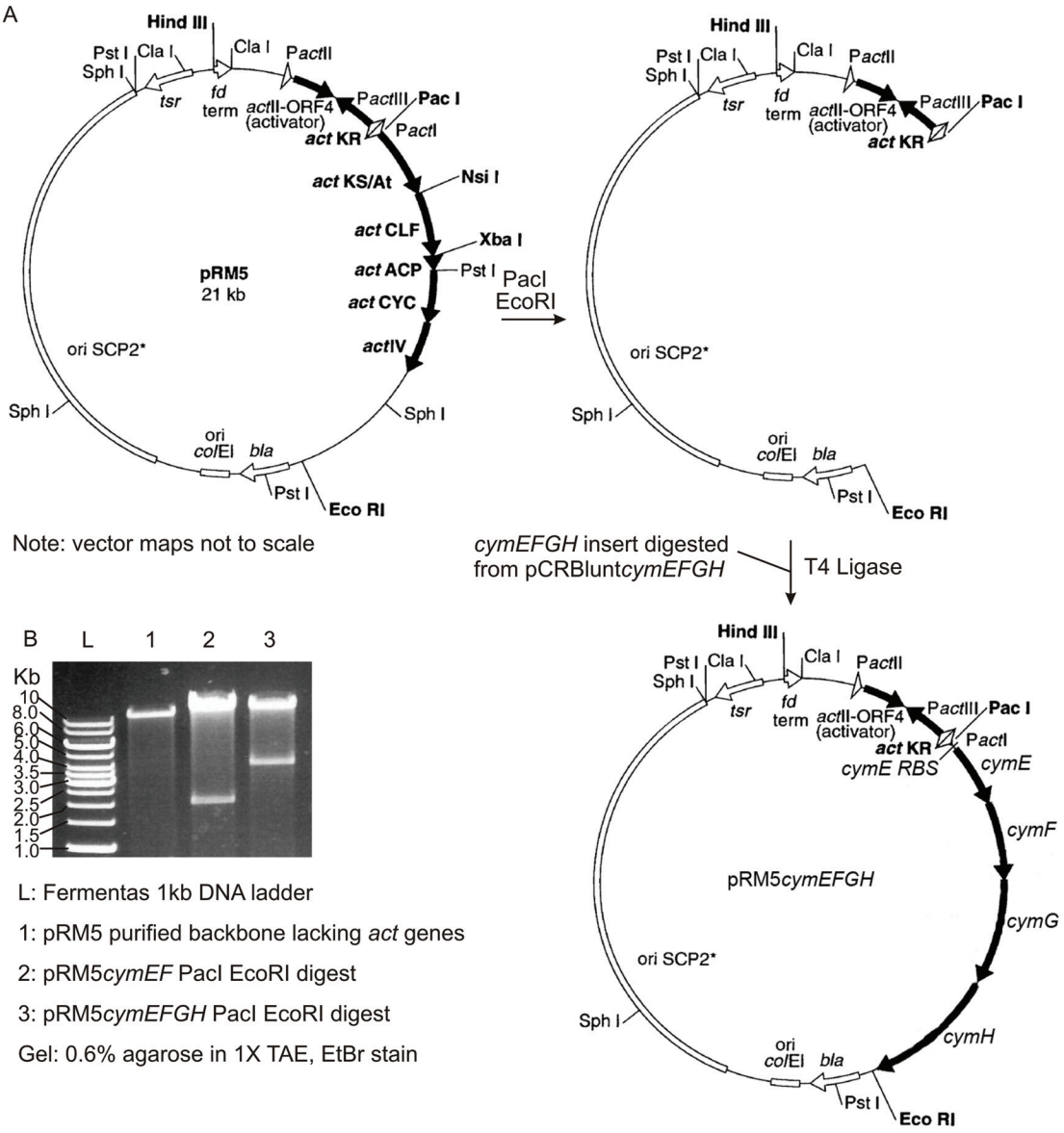


Figure 4.7: Cloning scheme for generation of pRM5*cymEFGH* from pRM5 for in vivo heterologous production of ADH in *S. coelicolor* YU105. (A) The plasmid pRM5 was digested with PacI and EcoRI to remove the *act* genes from *act*KS/At through *act*IV. *cymEFGH* was amplified from genomic DNA, subcloned into pCR®-Blunt II-TOPO®, and digested with PacI and EcoRI to liberate the fragment *cymEFGH.* *cymEFGH* was then ligated into the pRM5 backbone to yield pRM5*cymEFGH*. A similar strategy was utilized to create pRM5*cymEF*. See materials and methods for details. pRM5 vector map reproduced and adapted from McDaniel et al.[28] (B) Agarose gel of digested pRM5*cymEFGH* and pRM5*cymEF* confirms the correct insert size of 2 kb for *cymEF* and 4kb for *cymEFGH*.

therefore we did not clone *sare1130* into the pRM5 based constructs. Recovered transformants were assayed for the production of ADH via aqueous normal phase LC-MS and for 3,5-dimethyl-2-oxo-4-hexenoic acid by acid-catalyzed lactonization followed by extraction by reversed phase LC-MS. The results of these studies were inconclusive (data not shown), and require further analysis that is ongoing.

4.4: Discussion

Herein we report the partial characterization of *cymEFGH* as a four gene operon required for the production of 2-amino-3,5-dimethyl-4-hexenoic acid, a non-proteinogenic amino acid component of cyclomarin. Inactivation of *cymF*, *cymG*, or *cymH* abolished production of the known cyclomarins and provided a novel heptapeptide, cyclomarin F, which lacks the ADH residue, confirming the necessity of each of these genes in the production of ADH. The production of cyclomarin F, in which ADH was replaced by another endogenous amino acid (most likely Phe), suggests that a mutasynthetic[31] approach could be utilized to further diversify cyclomarin at this position. The aldolase/dehydrogenase pair CymEF likely converts isobutyryl-CoA to isobutyraldehyde, which is in turn condensed with pyruvate, forming a new C–C bond to provide 4-hydroxy-5-methyl-2-oxo-hexanoic acid. The similarity to other aldolase/dehydrogenase pairs such as SanNM[6] and MhpEF[4,11] along with previous precursor incorporation studies support this function for CymEF. Unfortunately, attempts to heterologously overexpress CymEF to provide enzyme for in vitro conformation of their function failed in both *E. coli* and *S. lividans*. The remaining components

of this pathway, CymG, CymH, and Sare1130 were easily provided as soluble, polyhistydyl-tagged proteins. Attempts to provide 4-hydroxy-5-methyl-2-oxo-hexanoic acid via heterologous pathway reconstruction in *S. coelicolor* is ongoing, and could provide confirmation of the function of CymEF. This product could then be incubated with recombinant CymG, CymH and Sare1130 to further characterize this pathway. Additionally, co-expression of *cymEFGH* with the aid of the endogenous branched chain amino acid aminotransferse should provide ADH.

4.5: Experimental section

4.5.1: General experimental procedures

Low-resolution LC-MS data was acquired using a Hewlett-Packard series 1100 LC/MS system utilizing electrospray ionization in positive mode, with a linear gradient of 10–90% aqueous MeCN at a flow rate of 0.7 mL/min over 24 min on a reversed phase C_{18} column (Phenomenex Luna, 4.6 mm x 100 mm, 5 µm) for general extract screening. High-resolution mass spectra were collected by ESI-HR-FTMS at the Chemistry & Biochemistry Mass Spectrometry Facility, University of California San Diego. NMR spectral data was obtained on a Bruker 600 MHz spectrometer equipped with a 1.7 mm cryoprobe.

4.5.2: Bacterial strains, plasmids, culture conditions, and DNA manipulations

Salinispora arenicola CNS-205 was grown in A1 media (10 g starch, 4 g yeast extract, and 2 g peptone) at 28 °C and 200 rpm unless otherwise noted. *Escherichia coli* DH5α (Invitrogen) or *E. coli* Top10 (Invitrogen) was used for cloning experiments as described,[32] and *E. coli* S17-1[33] and ET12567/pUZ8002[34] were used for conjugation experiments. Previously adapted for use in *Salinispora*,[20] REDIRECT® (Plant Bioscience Limited, Norwich, UK) technology was utilized for PCR targeting.[35] The λ-Red function was provided by pKD20 (bla).[36] Fosmid BPPW978 containing a fragment of the *cym* gene cluster from *cymA* to several kb beyond *cymH* in the pCC1FOS vector (Epicentre) was obtained from the Joint Genome Institute, Walnut Creek, CA and utilized for both *cymF, cymG* and *cymH* deletions. Apramycin (100 µg/ml for *S. arenicola*; 50 µg/ml for *E. coli*), chloramphenicol (2.5 µg/ml for *S. arenicola*; 12.5-25.0 µg/ml for *E. coli*), carbenicillin (100 µg/ml), and nalidixic acid (100 µg/ml) were used for selection of recombinants. DNA purification and manipulation was performed according to standard procedures.[32,37] Identities of constructed plasmids were verified by restriction digest and sequencing.

4.5.3: Conjugation protocol for *S. arenicola*

S. arenicola CNS-205 was cultivated for 5 days in 50 mL A1 in a 250 mL flask containing a stainless steel spring. Mycelium was harvested from 25 mL and resuspended in 2 mL A1. 0.5 mL of *E. coli* S17-1 suspension[37] was mixed

with 0.5 mL of the *S. arenicola* suspension and then plated on A1 agar plates. After 1 d incubation at 34 °C, the plates were overlaid with 2.5 mg each apramycin and nalidixic acid. Exconjugates were visible 10 days after plate inoculation.

4.5.4: Inactivation of *cymF*, *cymG*, and *cymH*

cymF was inactivated utilizing the modified PCR targeting system previously described for *S. tropica*.[20] In fosmid BPPW978, *cymF* was replaced by the apramycin resistance (*acc(3)IV*) cassette from pIJ773 utilizing λ-Red-mediated recombination. The *cymF* replacement cassette was generated by PCR using primers cymFkoF1 and cymFkoR1 (See appendix Table 4.S1). Lowercase letters represent 39 nt homologuous extensions immediately upstream and downstream of *cymF*, respectively, including the putative start and stop codons. This cassette was introduced into *E. coli* BW25113/pKD20[36] containing the chloramphenicol-resistant fosmid BPPW978. Gene replacement was verified using PCR primers cymFckF1 and cymFckR1. The mutated fosmid was introduced into *S. arenicola* CNS-205 by conjugation from *E. coli* S17-1 and confirmed by PCR analysis. Identical methodology was utilized to inactivate *cymG, cymH,* utilizing primers listed in appendix Table 4.S1.

4.5.5: Analysis of the *S. arenicola cym* mutants

100 µL of 4 d cultures of the *S. arenicola* WT or mutants was used to inoculate duplicate 50 mL cultures in A1+BFe containing 100 µg/ml apramycin (muntants only) in 250 mL Erlenmeyer flasks containing stainless steel springs. On day 10, the cultures were extracted with EtOAc, dried in vacuo and combined

with ~40 mg celite. The sample and celite was loaded on a 1g C_{18} solid phase extraction column (Thermo) and eluted with a step gradient of 20, 40, 60, 80, and 100% MeCN in H_2O (6 mL steps). 80% MeOH fractions were dried and reconstituted in 700 µL MeOH and analyzed for cyclomarazine analogs with a gradient of 20-40% MeCN over 30 min. 100% MeOH fractions were dried and reconstituted in 100 µL and analyzed for cyclomarin analogs, utilizing isocratic 57.5% MeCN in H_2O. For both analyses, a reversed phase C_{18} column (Phenomenex Luna, 4.6 mm x 100 mm, 5 µm) at 0.7 ml/min was used.

4.5.6: Purification of cyclomarin F from *S. arenicola* CNS-205 *cymF*⁻ mutant

S. arenicola CNS-205 was cultured at 27 °C with shaking at 250 rpm in 10 2.8-L Fernbach flasks containing 1 L of the medium A1+BFe. After 10 days organic chemical constituents from 10 L culture were collected by solid phase extraction using Amberlite XAD-7HP resin. The crude extract was prepared by washing the resin with acetone and evaporating the solvent in vacuo. The pH of the remaining aqueous residue was adjusted to pH ~12 with NaOH and partitioned with EtOAc. The EtOAc layer was dried under vacuum to yield ~1 g of crude material, which was fractionated utilizing Sephadex LH-20 resin (GE Healthcare, 35 g resin in a 30 mm wide column) with MeOH as the solvent. Fractions were screened by LC-MS, and those containing cyclomarin F were pooled. The cyclomarin pool was injected on a Phenomenex Synergi 10µm 250 x 21.2 mm column with a gradient of 48-90% MeCN in H_2O over 30 min at a flow rate of 13 mL/min. Cyclomarin F was eluted at 19.5 min. Cyclomarin F was further purified utilizing a Phenomenex Luna C_{18}(2) semi-preparative column (10

mm x 250 mm column at 2.5 mL/min) using a isocratic solvent of 60% MeCN in H_2O. 0.8 mg of **7** eluted at 23.2 min.

4.5.7: Cloning, heterologous expression and purification of ADH related enzymes in *E.coli*

The *cymE* gene (GenBank protein accession SARE4566, genome ascension CP00850) was amplified from genomic *S. arenicola* CNS-205 DNA via PCR utilizing the forward primer cymEF1 and reverse primer cymER1 (Appendix Table 4.S2). Ligation into pHIS8[15] yielded plasmid pHIS8-cymE, which was transformed into *E. coli* BL21(DE3) (Invitrogen) for expression. Transformants were grown in ZYP-5052 autoinduction media[16] containing 50 µg/mL kanamycin at 16 °C for two days. Cells were harvested and lysed by sonication in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol) with the addition of 0.5 g/L lysozyme. The lysate was cleared by centrifugation at 20,000 rpm for 50 min, and the resulting supernatant was loaded onto a polypropylene column containing 0.5 mL Ni-NTA agarose (Qiagen). The column was washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol), and eluted in elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol). A similar cloning and expression strategy was utilized to express *cymF*, *cymG*, *cymH*, and *sare1130* individually, the primers for which are denoted in Appendix Table 4.S2). To express *cymEF* together, primers cymEF1 and cymFR1 were utilized and cloned into pHIS8 under control of the singular T7*lac* promoter of

pHIS8. SDS-PAGE analysis followed by Coomassie staining was utilized to assay for expression and solubility of the recombinant enzymes.

The *E. coli* expression vector pETDuet-1 (Invitrogen) was utilized for co-expression of *cymE* and *cymF* under the control of individual promoters, and was selected for using 50 µg/ml ampicillin. The primers cymEduetF1 and cymEduetR1 were utilized to clone *cymE* into the first multiple cloning site, cymFduetF1 and cymFduetR1 was utilized to clone *cymF* into the second site without a stop codon to include the C-terminal s-tag to create pETDuet*cymEF* (Appendix Table 4.S2).

Additional *E. coli* expression strains, chaperone plasmids, and culturing conditions for heterologous expression were screened. Both pHIS8*cymEF* and pETDuet*cymEF* were transformed into Arctic Express (DE3) (Agilent, 100 µg/ml gentamicin), Rosetta 2 (DE3) pLysS (Merck, 68 µg/ml chloramphenicol), and SHuffle T7 (NEB, no selection) utilizing the manufacturer's protocols. For IPTG induction experiments, 50 mL culture with appropriate antibiotics in a 250 mL Erlenmeyer flask was inoculated with 100 µL overnight culture and grown at 37° C for 3-4 hrs to an OD of ~0.8. When at the appropriate OD, the culture temperature was adjusted to 10°C or 16°C and induced with IPTG at concentrations form 0.1 µM to 1.0 µM and grown overnight. Cells were harvested by centrifugation and screened for soluble protein expression as above.

4.5.8: Cloning, heterologous expression and purification of ADH related enzymes in *Streptomycetes lividans* TK24

The *cymE* gene was amplified from genomic *S. arenicola* CNS-205 DNA via PCR utilizing the forward primer cymEactF1 and reverse primer cymEactR1 (Appendix Table 4.S2). Ligation into pXY200oriT[15] yielded plasmid pXY200oriT*cymE*, which was transformed into *S. lividans* TK24 spores via conjugation as described elsewhere,[37] and was plated on MS agar and incubated at 30°C. Ex-conjugants were selected by flooding with apramycin to a concentration of 50 µg/ml 18 hrs after conjugation. Individual colonies were inoculated into 50 mL TSB, and grown for 3-4 days. A 1 ml aliquot was utilized to inoculate 50 mL of YEME containing 50 ug/mL apramycin and 25 ug/mL nalidixic acid in 250 mL Erlenmeyer flasks containing stainless steel springs. After 24 hrs, expression was induced with 5 µg/ml thiostrepton. Cells where harvested 1-2 d later, and recombinant proteins where isolated via Ni^{2+} affinity chromatography and SDS-PAGE as described for *E. coli*. A similar strategy was utilized to clone and express *cymF* and *cymH.* As in pHIS8 expression, *cymEF* was also cloned together under control of the same P_{tipA} promoter to yield pXY200oriTcymEF.

4.5.9: Generation of plasmids for In vivo heterologous production of ADH in *S. coelicolor* YU105

To construct *E. coli-Streptomyces* shuttle plasmids appropriate for production of ADH, *cymEFGH* was amplified from genomic *S. arenicola* CNS-205 DNA via PCR utilizing the reverse primer cymHRM5R3 (Appendix Table 4.S2) and forward primer cymEFRBS3, including the ribosome binding site directly upstream of *cymE*. The *cymEFGH* amplicon was subcloned into pCR®-Blunt II-TOPO® (Invitrogen), digested with PacI and EcoRI and then ligated into pRM5[28], which was previously digested with PacI and EcoRI to remove the *act* genes from *act*KS/At through *act*IV, to form pRM5*cymEFGH*. A similar approach was utilized to produce an expression plasmid, pRM5*cymEF* with primers cymEFRBS3 and cymFRM5R2 (Appendix Table 4.S2) for production of 4-hydroxy-5-methyl-2-oxo-hexanoic acid. The resultant plasmids were utilized to transform protoplasts of *Streptomycetes coelicolor* YU105[27], which were plated on R5YE. After ~18 h, transformants were selected for by flooding with 12.5 µg/ml thiostrepton.

4.5.10: Assay for In vivo heterologous production of ADH in *S. coelicolor* YU105

1 mL of a 4 d starter was used to inoculate single 50 mL cultures in modified R5YE lacking sucrose to prevent catabolite repression[30]. Utilizing a variation of the pure MeOH extraction method,[38] cells were harvested by centrifugation after 10 d of growth. 25 mL MeOH was added to the cell pellet, vortexed, frozen in N_2, thawed at room temperature, and centrifuged at 1000 *g*

for 20 min. This was repeated once. The MeOH fractions were pooled, and two 4 mL aliquots were concentrated to dryness in vacuo, and the remaining MeOH extract retained for further analysis. Two one dried aliquot 3 ml dd H_2O was added with 0.5 mL concentrated HCl and heated at 100 °C for 10 min[8]. When cool, the solution was extracted with 2 x 3 mL EtOAc and the organic phase dried in vacuo. This was reconstituted in MeOH, filtered, and analyzed by reverse phase LC-MS. To the second aliquot of the sample, 200 µL ddH_2O + 800 µL MeCN +0.1% formic acid was added, vortexed and centrifuged at 13,000 kpm for 10 min, and filtered. The samples were then analyzed by aqueous normal phase LC-MS utilizing a gradient from 5% to 50% A in B, where A is 0.1% formic acid in ddH_2O, B is 0.1% formic acid in MeCN on a Cogent diamond hydride 4.6 x 100 mm, 4 µm column at 1.0 mL/min and monitored by ESI-MS.

4.5.11: Assay for aldolase activity

Aldolase activity was assayed using procedures similar to those by Pollard et al.[10] Protein fractions potentially containing soluble, active CymF were concentrated with Nanosep 10 kD spin columns (Pall) and reconstituted in a minimal amount of 50 mM phosphate buffer. 50 µL concentrated enzyme was incubated at 37° C for 3-4 hrs with 2 mM isobutyraldehyde, 4 mM pyruvate, 1 mM $MnCl_2$, 1 mM DTT, 50 mM sodium phosphate buffer, pH 8.0 in a total volume of 1 mL. For the control, the buffered enzyme was boiled for 10 min before the addition of substrates. 100 µL of this reaction was mixed with 400 µL MeCN + 0.1% formic acid, vortexed, and centrifuged at 13,000 rpm for 10 min. The supernate was filtered and analyzed by aqueous normal phase LC-MS as

described above. Additionally, a 100 µL sample was mixed with 10 µL concentrated HCl, boiled for 10 min[8], and centrifuged. The supernate was analyzed by reversed phase LC-MS.

4.6: Appendix

Appendix table 4.S1: Primers for PCR directed gene inactivation

Replacement cassette primers[a,b]	
cymFKOF1	tgtgactcgacgttgcgcgacggcaaccacgccgtcgctATTCCGGGGATCCGTCGACC
cymFKOR1	cgtcaggtccgcggacgccgcctcggcggccagggccatTGTAGGCTGGAGCTGCTTC
cymGKOF1	tttgttgaccacatctgacggttgagggtgagaatcatgATTCCGGGGATCCGTCGACC
cymGKOR1	gccgggcaggccgatcggaacgaccttggtttgcgagaaTGTAGGCTGGAGCTGCTTC
cymHKOF1	gtggagctctgggaggccgcccggacccgtcggccggtcATTCCGGGGATCCGTCGACC
cymHKOR1	aacctgttcccccggcaccagatccaccgcgccggcccaTGTAGGCTGGAGCTGCTTC

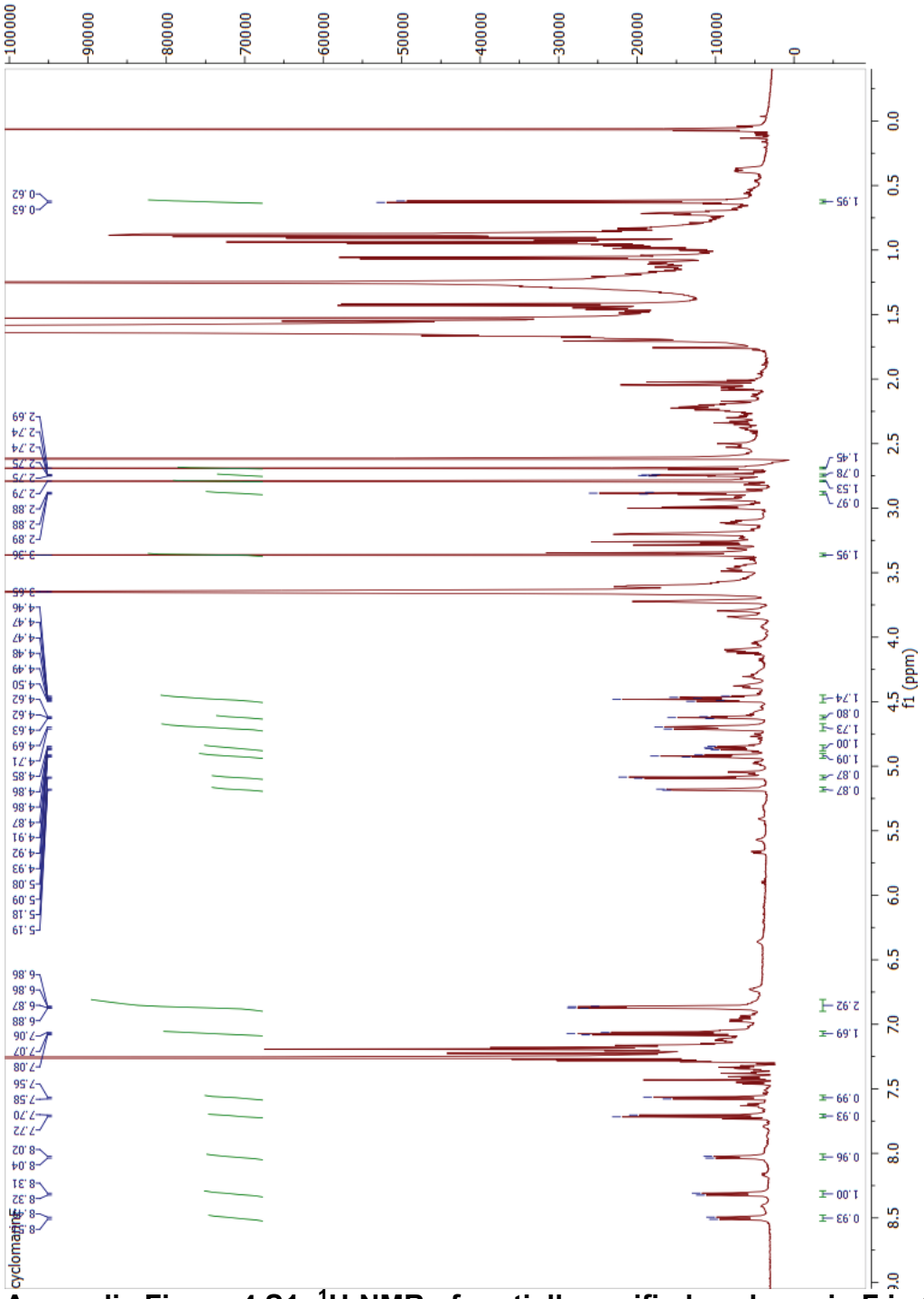
Verification primers[a]	
cymFckF1	GACTGCTACCTGCTCACCGT
cymFckR1	GGCTCGGCCGCGTCCAG
cymGckF1	TGGCCGGCCAGGAGGACAC
cymGckR1	CGTCCGCCAGGTCGAGTCG
cymHckF1	CTTCATGAGCGTCACCCTCC
cymHckR1	CACCCACATCGCACCCACG

[a]All primers denoted 5' to 3'. [b]Capital letters denote sequence homologous to apramycin resistance cassette, lower case letters homologous to gene being replaced.

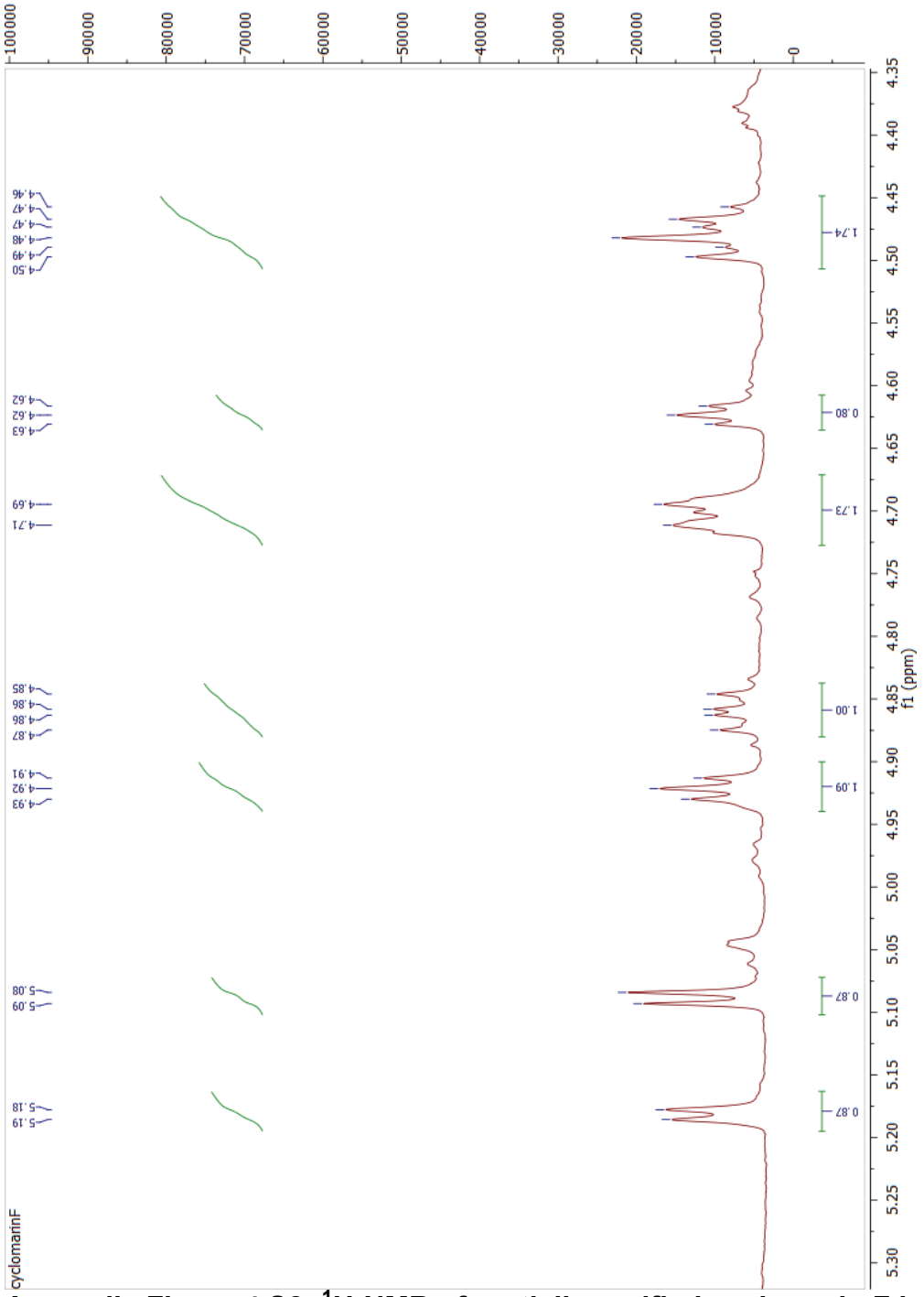
Appendix table 4.S2: Primers for PCR gene amplification utilized in heterologous expression experiments

name	RE site	sequence[a]
pHIS8 expression primers		
cymEF1	**BamHI**	*CGTGGTTC*ggatccATGAGTACGCAGACGCATCCAGCTCC
cymER1	**HindIII**	*GCTCGAATTC*aagcttTCAACGCCTCAGCCCGCTGCG
cymFF1	**BamHI**	*CGTGGTTC*ggatccATGAGCGTGAAACTGCAGATCTGTGACTCG
cymFR1	**HindIII**	*GCTCGAATTC*aagcttTCAGATGTGGTCAACAAAGGCGTGCGTC
cymGF1	**NcoI**	*CGTGGTTC*cccatggATGGAACGGAAAGACCCGGAAGTCG
cymGR1	**HindIII**	*GCTCGAATTC*aagcttTCATCTGGTCGCCACGAAGAGCCG
cymHF1	**NcoI**	*CGTGGTTC*cccatggATGACGGCCGTCGCGCGGACG
cymHR1	**HindIII**	*GCTCGAATTC*aagcttCTAACCTTCTGATCCGTAGGTCATGGAG
sare1130F1	**EcoRI**	*CGTGGTTC*gaattcATGAGCGGTGGTGACAAGCTCGAC
sare1130R1	**HindIII**	*GCTCGAATTC*aagcttTCAGAGCACGTGGTGGGCCC
pETDuet-1 expression primers		
cymEduetF1	**BamHI**	*CGTGGTTC*ggatccAATGAGTACGCAGACGCATCCAGCTCC
cymEduetR1	**HindIII**	*GCTCGAATTC*aagcttTCAACGCCTCAGCCCGCTGCG
cymFduetF1	**NdeI**	*CGTGGTTC*catatgATGAGCGTGAAACTGCAGATCTGTGACTCG
cymFduetR1	**EcoRV**	*GCTCGAATTC*gatatcCCGATGTGGTCAACAAAGGCGTGCGTCAG
pXY200oriT expression primers		
cymEactF1	**NdeI**	*CGTGGTTC*catatgATGAGTACGCAGACGCATCCAGCTCC
cymEactR1	**EcoRI**	*GCTCGAATTC*gaattcTCAACGCCTCAGCCCGCTGCG
cymFactF1	**NdeI**	*CGTGGTTC*catatgATGAGCGTGAAACTGCAGATCTGTGACTCG
cymFactR1	**EcoRI**	*GCTCGAATTC*gaattcTCAGATGTGGTCAACAAAGGCGTGCGTC
cymHactF1	**NdeI**	*CGTGGTTC*catatgATGACGGCCGTCGCGCGGACG
cymHactR1	**EcoRI**	*GCTCGAATTC*gaattcCTAACCTTCTGATCCGTAGGTCATGGAG
pRM5 expression primers		
cymEFRBS3	**PacI**	ttaattaaCAGCATCACCTCAGTGGCCAGACC
cymFRM5R2	**EcoRI**	gaattcCTAACCTTCTGATCCGTAGGTCATGGAG
cymHRM5R3	**none**	CAGACCCTTTATGATCTTGCGCGCCCG

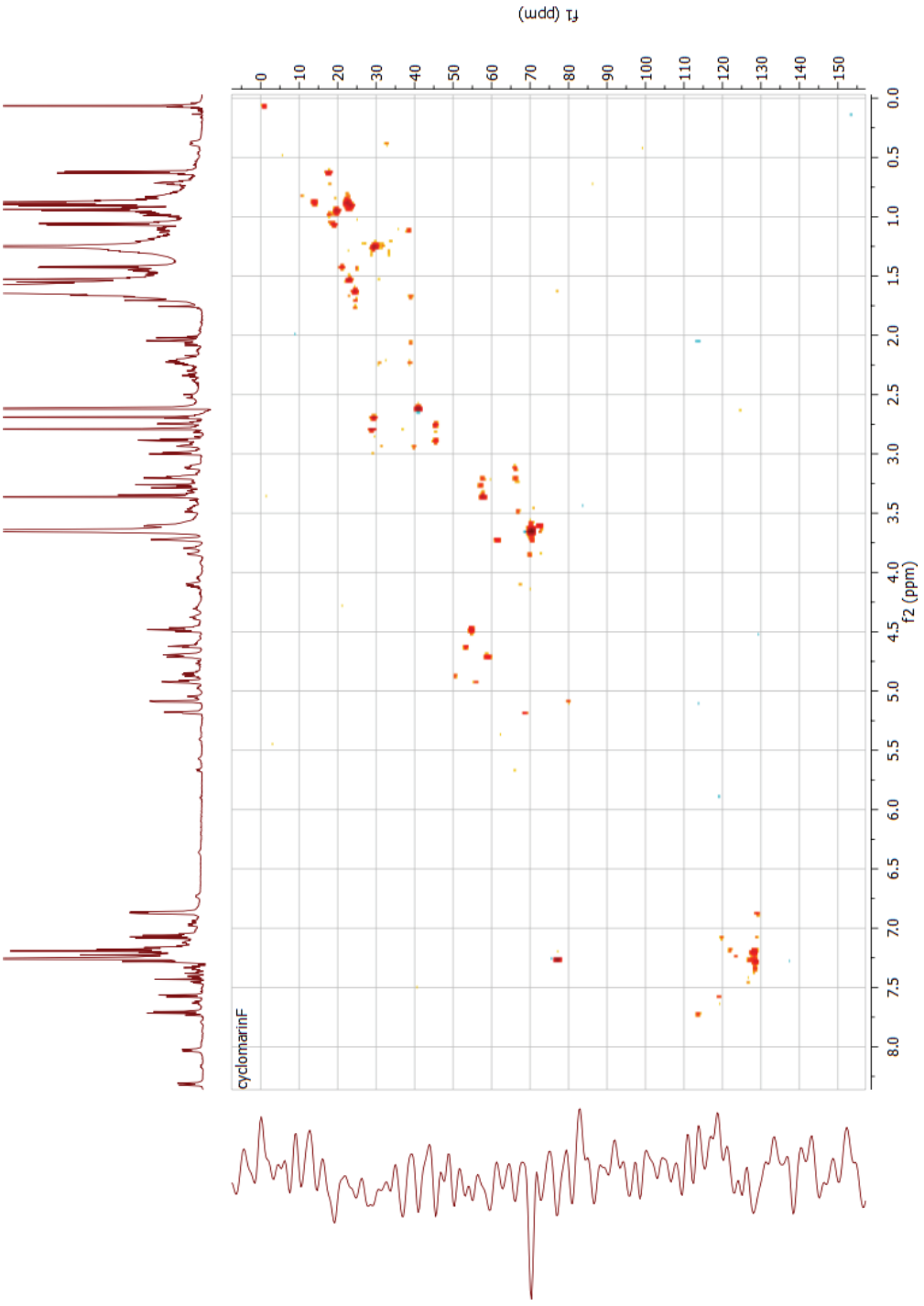
[a]Italics: random sequence, underline: restriction site, standard uppercase: sequence homologous to gene of interest



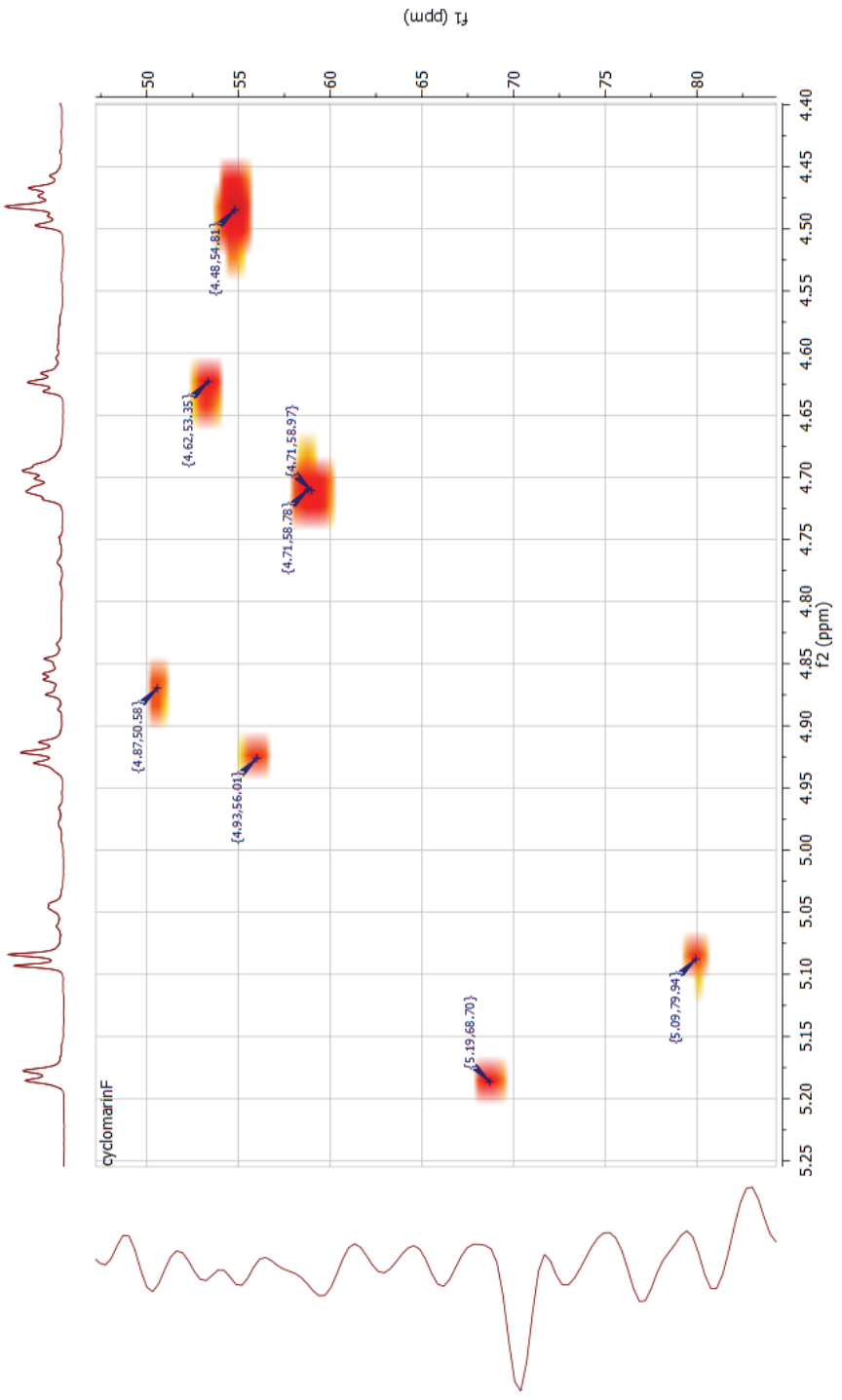
Appendix Figure 4.S1: ^{1}H-NMR of partially purified cyclomarin F in $CDCl_3$



Appendix Figure 4.S2: ^{1}H-NMR of partially purified cyclomarin F in $CDCl_3$, expansion from 5.3 ppm to 4.35 ppm



Appendix Figure 4.S3: HSQC of partially purified cyclomarin F in $CDCl_3$



Appendix Figure 4.S4: HSQC of partially purified cyclomarin F in $CDCl_3$, expansion from 5.25 to 4.40 ppm and 52.5 to 90.0 ppm

4.7: References

(1) Kulanthaivel, P.; Vasudevan, V.; WO Patent WO/2000/078,797: **2000**.

(2) Takahashi, Y.; Omura, S. Isolation of new actinomycete strains for the screening of new bioactive compounds. *J. Gen. Appl. Microbiol.* **2003**, *49*, 141-154.

(3) Takita, T.; Ohi, K.; Okami, Y.; Maeda, K.; Umezawa, H. New antibiotics, ilamycins. *J. Antibiot. Ser. A.* **1962**, *15*, 46-48.

(4) Arai, H.; Yamamoto, T.; Ohishi, T.; Shimizu, T.; Nakata, T.; Kudo, T. Genetic organization and characteristics of the 3-(3-hydroxyphenyl)propionic acid degradation pathway of *Comamonas testosteroni* TA441. *Microbiology.* **1999**, *145*, 2813-2820.

(5) Mahlert, C.; Kopp, F.; Thirlway, J.; Micklefield, J.; Marahiel, M. Stereospecific enzymatic transformation of α-ketoglutarate to (2*S*, 3*R*)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J. Am. Chem. Soc.* **2007**, *129*, 12011-12018.

(6) Ling, H.; Wang, G.; Tian, Y.; Liu, G.; Tan, H. SanM catalyzes the formation of 4-pyridyl-2-oxo-4-hydroxyisovalerate in nikkomycin biosynthesis by interacting with SanN. *Biochem. Biophys. Res. Commun.* **2007**, *361*, 196-201.

(7) Khajamohiddin, S.; Repalle, E. R.; Pinjari, A. B.; Merrick, M.; Siddavattam, D. Biodegradation of aromatic compounds: an overview of *meta*-fission product hydrolases. *Crit. Rev. Microbiol.* **2008**, *34*, 13-31.

(8) Pollard, J.; Bugg, T. Purification, characterisation and reaction mechanism of monofunctional 2-hydroxypentadienoic acid hydratase from *Escherichia coli*. *Eur. J. Biochem.* **1998**, *251*, 98-106.

(9) Izumi, A.; Rea, D.; Adachi, T.; Unzai, S.; Park, S.-Y.; Roper, D. I.; Tame, J. R. H. Structure and mechanism of HpcG, a hydratase in the homoprotocatechuate degradation pathway of *Escherichia coli*. *J. Mol. Biol.* **2007**, *370*, 899-911.

(10) Pollard, J.; Rialland, D.; Bugg, T. Substrate selectivity and biochemical properties of 4-hydroxy-2-keto-pentanoic acid aldolase from *Escherichia coli*. *Appl. Environ. Microbiol.* **1998**, *64*, 4093.

(11) Lee, S.-J.; Ko, J.-H.; Kang, H.-Y.; Lee, Y. Coupled expression of MhpE aldolase and MhpF dehydrogenase in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **2006**, *346*, 1009-1015.

(12) Manjasetty, B. A.; Powlowski, J.; Vrielink, A. Crystal structure of a bifunctional aldolase–dehydrogenase: sequestering a reactive and volatile intermediate. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6992-6997.

(13) Wang, W.; Baker, P.; Seah, S. Y. K. Comparison of two metal-dependent pyruvate aldolases related by convergent evolution: substrate specificity, kinetic mechanism, and substrate channeling. *Biochemistry.* **2010**, *49*, 3774-3782.

(14) Baker, P.; Pan, D.; Carere, J.; Rossi, A.; Wang, W.; Seah, S. Characterization of an aldolase-dehydrogenase complex that exhibits substrate channeling in the polychlorinated biphenyls degradation pathway. *Biochemistry.* **2009**, *48*, 6551-6558.

(15) Colabroy, K.; Begley, T. Tryptophan catabolism: identification and characterization of a new degradative pathway. *J. Bacteriol.* **2005**, *187*, 7866.

(16) Lopez-Sant N, J.; Alvaro, G.; Clapes, P. Use of aldolases for asymmetric synthesis. *Enzyme biocatalysis: principles and applications.* **2008**, 333.

(17) Samland, A.; Sprenger, G. Microbial aldolases as C–C bonding enzymes—unknown treasures and new developments. *Appl. Microbiol. Biotechnol.* **2006**, *71*, 253-264.

(18) Takayama, S.; Mcgarvey, G. J.; Wong; Chi-Huey Microbial aldolases and transketolases:new biocatalytic approaches to simple and complex sugars. *Annu. Rev. Microbiol.* **1997**, *51*, 285-310.

(19) Wang, W.; Seah, S. Purification and biochemical characterization of a pyruvate-specific class II aldolase, HpaI. *Biochemistry.* **2005**, *44*, 9447-9455.

(20) Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2008**, *4*, 69-74.

(21) Liu, W.-T.; Ng, J.; Meluzzi, D.; Bandeira, N.; Gutierrez, M.; Simmons, T. L.; Schultz, A. W.; Linington, R. G.; Moore, B. S.; Gerwick, W. H.; Pevzner, P. A.; Dorrestein, P. C. Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Anal. Chem.* **2009**, *81*, 4200-4209.

(22) Renner, M.; Shen, Y.; Cheng, X.; Jensen, P.; Frankmoelle, W.; Kauffman, C.; Fenical, W.; Lobkovsky, E.; Clardy, J. Cyclomarins A- C, new antiinflammatory cyclic peptides produced by a marine bacterium (*Streptomyces* sp.). *J. Am. Chem. Soc.* **1999**, *121*, 11273-11276.

(23) Jez, J. M.; Ferrer, J.-L.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase *Biochemistry.* **2000**, *39*, 890-902.

(24) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **2005**, *41*, 207-234.

(25) Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 871-876.

(26) Izumikawa, M.; Cheng, Q.; Moore, B. S. Priming type II polyketide synthases via a type II nonribosomal peptide synthetase Mechanism. *J. Am. Chem. Soc.* **2006**, *128*, 1428-1429.

(27) Yu, T.; Hopwood, D. Ectopic expression of the *Streptomyces coelicolor* whiE genes for polyketide spore pigment synthesis and their interaction with the act genes for actinorhodin biosynthesis. *Microbiology.* **1995**, *141*, 2779.

(28) Mcdaniel, R.; Ebert-Khosla, S.; Hopwood, D.; Khosla, C. Engineered biosynthesis of novel polyketides. *Science.* **1993**, *262*, 1546.

(29) Hertweck, C.; Xiang, L.; Kalaitzis, J. A.; Cheng, Q.; Palzer, M.; Moore, B. S. Context-dependent behavior of the enterocin iterative polyketide synthase: a new model for ketoreduction. *Chem. Biol.* **2004**, *11*, 461-468.

(30) Yu, T.-W.; Müller, R.; Müller, M.; Zhang, X.; Draeger, G.; Kim, C.-G.; Leistner, E.; Floss, H. G. Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. *J. Biol. Chem.* **2001**, *276*, 12546-12555.

(31) Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat. Prod. Rep.* **2008**, *25*, 25-34.

(32) Sambrook, J.; Russell, D. W. *Molecular Cloning, a Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2001.

(33) Simon, R.; Priefer, U.; Pühler, A. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram negative bacteria. *Bio/Technology.* **1983**, *1*, 784-791.

(34) Paget, M. S. B.; Chamberlin, L.; Atrih, A.; Foster, S. J.; Buttner, M. J. Evidence that the extracytoplasmic function sigma factor E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **1999**, *181*, 204-211.

(35) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1541-1546.

(36) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6640-6645.

(37) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; The John Innes Foundation: Norwich, 2000.

(38) Villas Bôas, S.; Højer Pedersen, J.; Åkesson, M.; Smedsgaard, J.; Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast.* **2005**, *22*, 1155-1169.

4.8: Acknowledgment

We would like to thank Diana Plutchak for her assistance with chemical extraction of the ADH deficient mutants and heterologous expression in *E. coli.,* and Tobias Guilder for the generous gift of pXY200oriT. We would also like to thank Alex Roberts for her advice on cloning experiments and heterologous expression.

Chapter 5:

Conclusion and Future Directions

Countless lives have been saved because of the ability of penicillin and vancomycin to cure bacterial infections, and the ability of cyclosporine to suppress immune functions allowing for organ transplantation. These drugs are just three examples of the numerous clinically relevant non-ribosomal peptides produced by certain bacteria and fungi. Indeed, microbial fermentation is utilized, and will continue to be utilized for the foreseeable future, to provide these useful compounds due to their structural complexity and useful biological activity. Because these compounds are difficult to make or derivatize by chemical synthesis, there is significant interest in creating analogs of these compounds by manipulating their biosyntheses[1]. There are numerous reasons to develop analogs of these clinical agents; one of the most significant is to combat the growing threat posed by antibiotic resistant pathogens[2]. Knowledge gained by understanding the biosynthesis of non-ribosomal peptides, even those not used clinically, could be applied to engineer producers of the compound of interest to provide new analogs. This knowledge can also be applied to de novo drug discovery efforts by screening genomes for the genetic potential to produce novel compounds[3].

Developing drugs useful for treating disease is the most compelling reason to study how these peptides are biosynthesized, but this is not the only reason. Beyond the pure intellectual curiosity about how these compounds are made, there is interest in understanding the enzymology behind assembly and tailoring of the amino acid components, along with the details of how the peptides are assembled. This knowledge in turn can be utilized to develop enzymes for

industrial catalysis[4] and chemoenzymatic synthesis[5]. Understanding the biosynthesis of non-ribosomal peptides could provide insight into the roles of these compounds in their host producers.

This dissertation has chronicled efforts to understand and manipulate the biosynthesis of the non-ribosomal peptides cyclomarin and cyclomarazine. This work was not focused on developing these compounds as drugs, but instead in understanding how they are biosynthesized so that this knowledge can in turn be applied to engineer other non-ribosomal peptides or to develop tools for chemoenzymatic synthesis. Indeed, several features of these cyclic, non-ribosomal peptides are of significant biosynthetic interest. The *N*-(1,1-dimethyl-1-allyl)-tryptophan residue shared between cyclomarin and cyclomarazine is a rare example of prenylation of non-ribosomal peptide in a bacterial system. In addition, cyclomarin contains the non-proteinogenic amino acid 2-amino-3,5-dimethyl-4-hexenoic acid (ADH), which has not been observed in any other secondary metabolite. Beyond these key structural features, four oxidations and three methylations are present on the fully fictionalized cyclomarin A, some of the genes for which have been characterized during this study.

Cyclomarin and cyclomarazine biosynthesis begins with the tryptophan derived residue *N*-(1,1-dimethyl-1-allyl)-tryptophan, and the mechanism by which tryptophan is tailored to provide this residue was determined. We found that tryptophan is prenylated by the prenyltransferase CymD prior to being incorporated by the cyclomarin synthetase CymA. Purified, recombinant CymD was shown to convert tryptophan and DMAPP to *N*-(1,1-dimethyl-1-allyl)-

tryptophan in a cation-independent manner. The preference for *N*-alkylated tryptophan over endogenous tryptophan was established by providing *N*-(1,1-dimethyl-1-allyl)-tryptophan, *N*-(1-methyl)-tryptophan, and *N*-(1-propargyl)-tryptophan to a *cymD* deficient mutant of *S. arenicola* CNS-205, which efficiently restored the production of the known cyclomarin A and cyclomarazine A and provided the corresponding *N*-methyl or *N*-propargyl mutasynthetic analogs at wild type levels. Further oxidative tailoring of this amino acid is provided by two cytochrome P450s, CymV and CymS, the first of which installs the epoxide, and the second of which installs the β-hydroxyl group, forming the fully functionalized *N*-(1,1-dimethyl-2,3-epoxypropyl)-β-hydroxytryptophan residue of cyclomarin A. Bioinformatic analysis coupled with in vivo mutagenesis established the functions for these two CYPs.

The second residue contained in both cyclomarin and cyclomarazine is δ-hydroxyl-*N*-methyl-leucine. Inactivation of the Fe(II)/α-ketoglutarate-dependent hydroxylase CymW supported its role in providing the δ-hydroxyl functionality by eliminating the production of all known cyclomarins and cyclomarazines containing the δ-hydroxylation on leucine. Module 2 of CymA contains a methyltransferase domain that likely catalyzes the methyl transfer to this residue from the methyl donor SAM.

Modules 3 through 6 are likely to incorporate the proteinogenic amino acids alanine, phenylalanine, valine, and leucine into the structure of cyclomarin. Tailoring of phenylalanine is most likely provided by the cytochrome P450 CymO while tethered to the NRPS followed by SAM dependent methylation via CymP to

provide the β-methoxyphenylalanine residue. Although the functions of these enzymes were not probed in this work, bioinformatic analysis along with in vivo characterization of the remaining oxidases and methyltransfereases in the pathway supports the functions of CymO and CymP. To provide the final *N*-methyl group of cyclomarin, the methyltransferase domain in module 6 would convert the leucine residue to *N*-methylleucine.

The final amino acid incorporated is ADH, a non-proteinogeic amino acid provided by the four gene operon *cymE-H*, three genes of which share significant identity with enzymes from *meta* cleavage pathways utilized to degrade phenolic compounds in some bacteria. Stable isotope incorporation studies coupled with bioinformatic analysis establish the origin of ADH from isobutyryl-CoA, pyruvate and SAM in a pathway operating in an analogous, yet reverse manner to the degradative pathways for phenolic compounds. Inactivation of *cymF*, *cymG*, or *cymH*, in *S. arenicola* abolished the production of the known cyclomarins and provided a new analog in which ADH was replaced by an endogenous substrate, most likely phenylalanine. These inactivation studies confirmed the role of *cymF-H* in the production of ADH and suggest that this residue of cyclomarin may be amendable to modification via mutasynthesis. Key to forming the carbon framework of this amino acid is the aldolase/dehydratase pair CymEF, which convert isobutyryl-CoA to isobutyraldehyde, which is in turn condensed with pyruvate to give 4-hydroxy-5-methyl-2-oxohexanoic acid. Dehydration via CymH provides a substrate appropriate for SAM dependent methylation by CymG followed by transamination via Sare1130 to yield ADH. Numerous attempts were

made to provide soluble CymEF via heterologous overexpression in *E. coli* and *Streptomyces lividans* for in vitro characterization and to probe their usefulness in chemoenzymatic synthesis. Unfortunately, the appropriate conditions to provide active, soluble CymEF were not discovered. Current studies to produce ADH in *Streptomyces coelicolor* via in vivo heterologous pathway reconstitution are underway.

This work has highlighted the power of an integrated, interdisciplinary approach on the chemistry/biology interface to elucidate the details of cyclomarin and cyclomarazine biosynthesis. Stable isotope incorporation studies coupled with bioinformatic analysis established the origins of the residues of these cyclic peptides and provided a putative pathway to ADH. In vivo and in vitro characterization of several Cym enzymes established their role in tailoring of the proteinogenic amino acid residues and solidified the pathway to the non-proteinogenic amino acid. In addition, we were able to engineer both the cyclomarin and cyclomarazine series to provide analogs in which the prenyl group on tryptophan was missing or replaced by a methyl or propargyl group, or the ADH residue was replaced by another endogenous substrate. Metabolic profiling of the various mutants highlighted the presence of a previously uncharacterized cyclomarazine analog containing the epoxide functionality on the tryptophan derived residue.

5.1: References

(1) Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat. Prod. Rep.* **2008**, *25*, 25-34.

(2) Donadio, S.; Maffioli, S.; Monciardini, P.; Sosio, M.; Jabes, D. Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J. Antibiot.* **2010**, *63*, 423-430.

(3) Wilkinson, B.; Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* **2007**, *3*, 379-386.

(4) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial biocatalysis today and tomorrow. *Nature.* **2001**, *409*, 258-268.

(5) Kopp, F.; Marahiel, M. A. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Curr. Opin. Biotechnol.* **2007**, *18*, 513-520.