UNIVERSITY OF CALIFORNIA SAN DIEGO

Biosynthetic strategies for chemodiversity generation in marine filamentous cyanobacteria

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

Doctor of Philosophy

in

Marine Biology

by

Nathan Aaron Moss

Committee in charge:

Professor William H. Gerwick, Chair
Professor Eric Allen
Professor Lena Gerwick
Professor James W. Golden
Professor Bradley S. Moore

2019

The dissertation of Nathan Aaron Moss is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2019

DEDICATION

This dissertation is dedicated to my unstoppable parents, David and Jane Moss, my peerless sister Anica and intrepid brother Daniel, and to my wife Sophia, who is wonderful beyond words.

And to our grandparents.

EPIGRAPH

There are the rushing waves
mountains of molecules
each stupidly minding its own business
trillions apart
yet forming white surf in unison.

—Richard Feynman

There is no excellent beauty that hath not some strangeness in the proportion.

—Francis Bacon

Nothing in the world is worth having or worth doing unless it means effort, pain, difficulty… I have never in my life envied a human being who led an easy life. I have envied a great many people who led difficult lives and led them well.

—Theodore Roosevelt

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A – Adenylation domain
AA – Amino acid
ACP – Acyl carrier protein
AcT – Acetyltransferase
Ahp – 3-amino-6-hydroxy-2-piperidone
AmT – Amidotransferase
ASI – *Moorea producens* ASI16Jul14-2
AT – Acyltransferase
BGC – Biosynthetic gene cluster
C – Condensation domain
CH_2Cl_2 – Dichloromethane/methylene chloride
CH_3OH – Methanol
$CHCl_3$ – Chloroform
CoA – Coenzyme A
COM – Communication domain
COSY – Correlation Spectroscopy
DCM – Dichloromethane/methylene chloride
DH – Dehydratase domain
DMAPP – Dimethylallyl pyrophosphate
DMSO – Dimethyl sulfoxide
E – Epimerase domain
ECH – Enoyl CoA hydratase domain
ER – Enoyl reductase domain
ESI – Electrospray ionization
EtOH – Ethanol
FACL – Fatty acid-CoA ligase
FAS – Fatty acid synthase
D-FDAA – 1-fluoro-2-4-dinitrophenyl-5-D-alanine amide
L-FDAA – 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide
gDNA – Genomic DNA
GNAT – GCN5-acetyltransferase
GNMT – Glycine-*N*-methyltransferase
Hal – Halogenase
HCS – Hydroxymethylglutaryl-CoA synthase
HMBC – Heteronuclear multiple-bond correlation spectroscopy
HPLC – High-performance liquid chromatography
HR-ESI-MS – High resolution electrospray ionization mass spectrometry

HSQC – Heteronuclear single-quantum correlation spectroscopy
IPP – Isopentenyl pyrophosphate
KR – Ketoreductase domain
KR^0 – Inactive ketoreductase domain
KS – Ketosynthase domain
LC – Liquid chromatography
LD – Loading domain or lipoyl domain
Lip – Lipocalin
MeCN – Acetonitrile
MS – Mass spectrometry
MT – Methyltransferase
MTT – MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NMR – Nuclear magnetic resonance
NOESY – Nuclear Overhauser effect spectroscopy
NRPS – Non-ribosomal peptide synthetase
ORF – Open reading frame
OT – Octanoyltransferase
PAB – *Okeania hirsuta* PAB10Feb10-1
PAP – *Okeania hirsuta* PAP21Jun06-1
PBDE – Polybrominated diphenyl ether
PCP – Peptidyl carrier protein
PKS – Polyketide synthase
R – Thioester reductase
RiPPs – Ribosomally synthesized and post-translationally modified peptides
ROESY – Rotating-frame nuclear Overhauser effect correlation spectroscopy
SAM – S-adenosyl methionine
SDR – Short chain dehydrogenase/reductase
SW BG-11 – Seawater BG-11 media
TE – Thioesterase
Tle – *tert*-leucine

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation would not be possible without the selfless and patient guidance of several postdoctoral researchers and graduate students in the Gerwick Lab and its affiliates, to whom I owe a tremendous debt of gratitude: Cameron Coates, Emily Mevers, Arnaud Taton, Karin Kleigrewe, Matt Bertin, and Jehad Almaliti. I would like to thank all current Gerwick Lab members and undergraduate volunteers past and present for their input, advice, contributions and scientific discussion, in particular Tiago Leão and Evgenia Glukhov.

I would like to thank Professors Bill and Lena Gerwick for providing opportunities and support to make my research possible, including the Training Grant in Marine Biotechnology and the Microbial Sciences Initiative Graduate Fellowship, as well as offering countless pieces of advice and guidance for several projects over my graduate research career. I would like to thank my committee members as well, Professors Brad Moore, Eric Allen, and James Golden for their time and guidance towards meaningful projects and outcomes.

I thank many members of CMBB, past and present, for their advice and counsel on many scientific topics: Marco Allemann, Shaun McKinnie, Vinayak Agarwal, Tristan de Rond, Peter Jordan, Krystle Chavarria, Nastassia Patin, Grant Seiler, Gabriel Castro-Falcòn, Trevor Purdy, and Professors Chambers Hughes and Paul Jensen.

I also am thankful for the invaluable contribution of collaborators at other institutions: Meredith Skiba, Andy Sikkema, and Janet Smith at the University of Michigan, Pingping Qu and Feixue Fu at the University of Southern California, and Anton Korobeynikov at the University of St. Petersburg.

I am grateful to the research-support staff in and around SIO and UCSD: dive safety officers Rich Walsh and Christian McDonald, without whom much of my research would not be possible, as well as Thomas Arrhenius, whose technical wizardry facilitated countless experiments, and Brendan Duggan, Anthony Mrse, and Yongxuan Su for their analytical expertise.

I thank all of my friends for the many visits, companionship, humor, and support, and who despite my disappearance to San Diego and our diaspora across the country, remain hilarious, excellent, steadfast, and true.

I would like to thank my incredible parents for selflessly giving me an irreplaceable foundation of education, support, and love, and teaching the value of hard work and dedication. To my brother Daniel and sister Anica, thank you for the visits, the perpetual outdoorsmanship and the respite it affords.

I finally thank my wife, Sophia, for being an endless wellspring of inspiration, positivity, support, and love. This would not have happened without you.

Chapter 2, in full, is a reprint, with permission, of the material as it appears in ACS Chemical Biology, 2018, Moss, Nathan A.; Leão, Tiago F.; Rankin, Michael R.; McCullough, Tyler M.; Qu, Pingping; Korobeynikov, Anton; Smith, Janet L.; Gerwick, Lena; and Gerwick, William H. 2018, *13*, pages 3385-3395. The dissertation author is the primary investigator and author of this material.

Chapter 3, in full, has been submitted for publication of the material. Moss, Nathan A.; Seiler, Grant; Leão, Tiago F.; Castro-Falcòn, Gabriel; Gerwick, Lena; Hughes, Chambers; Gerwick, William H. "Nature's combinatorial biosynthesis produces vatiamides A-F." 2019. The dissertation author is the primary investigator and author of this material.

Chapter 4, in part, is reproduced, with permission, from ACS Chemical Biology; Biosynthesis of *t*-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module. Skiba, Meredith A.; Sikkema, Andrew P.; Moss, Nathan A.; Lowell, Andrew N.; Su, Min; Sturgis, Rebecca M.; Gerwick, Lena; Gerwick, William H.; Sherman, David H.; Smith, Janet L. 2018, *13*, 1640-1650.

VITA

Academic

2019	Ph.D, Marine Biology Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA
2013 – 2019	Graduate student, Marine Biology Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA
2007	B.S., Chemical Biology Department of Chemistry, University of California, Berkeley, Berkeley, CA

Professional

2007 – 2013	Associate Scientist I-III Amyris Inc, Emeryville, CA

Selected Publications

Moss, N.A., Leão, T., Rankin, M.R., McCullough, T.M., Qu, P., Korobeynikov, A., Smith, J.L., Gerwick, L. and Gerwick, W.H., 2018. Ketoreductase domain dysfunction expands chemodiversity: malyngamide biosynthesis in the cyanobacterium Okeania hirsuta. ACS chemical biology, *13*(12), 3385-3395.

Moss, N. A., Leao, T., Glukhov, E., Gerwick, L., and Gerwick, W. H., 2018. "Collection, Culturing, and Genome Analyses of Tropical Marine Filamentous Benthic Cyanobacteria." Methods in enzymology, *604*, 3-43.

Skiba, M. A., Sikkema, A. P., **Moss, N. A**., Lowell, A. N., Su, M., Sturgis, R. M., Gerwick, L., Gerwick, W.H., Sherman, D. H. and Smith, J. L., 2018. "Biosynthesis of t-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module." ACS Chemical Biology, *13*(6), 1640-1650.

Skiba, M. A., Sikkema, A. P., **Moss, N. A**., Tran, C. L., Sturgis, R. M., Gerwick, L., Gerwick, W.H., Sherman, D. H. and Smith, J. L., 2017. "A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit." ACS Chemical Biology, *12*(12), 3039-3048.

Kinnel, R. B., Esquenazi, E., Leao, T., **Moss, N. A.**, Mevers, E., Pereira, A. R., Monroe, E. A., Korobeynikov, A., Murray, T. F., Sherman, D. H., Gerwick, L., Dorrestein, P. C., and Gerwick, W. H., 2017. A Maldiisotopic approach to discover natural products: Cryptomaldamide, a hybrid tripeptide from the marine cyanobacterium Moorea producens. Journal of Natural Products, *80*(5), 1514-1521.

Moss, N.A., Bertin, M.J., Kleigrewe, K., Leão, T.F., Gerwick, L. and Gerwick, W.H., 2016. "Integrating mass spectrometry and genomics for cyanobacterial metabolite discovery." Journal of industrial microbiology & biotechnology, *43*(2-3), 313-324.

Bertin, M.J., Demirkiran, O., Navarro, G., **Moss, N.A.**, Lee, J., Goldgof, G.M., Vigil, E., Winzeler, E.A., Valeriote, F.A. and Gerwick, W.H., 2016. Kalkipyrone B, a marine cyanobacterial γ-pyrone possessing cytotoxic and anti-fungal activities. *Phytochemistry*, *122*, 113-118.

Kirby, J., Dietzel, K.L., Wichmann, G., Chan, R., Antipov, E., **Moss, N.A.**, Baidoo, E.E., Jackson, P., Gaucher, S.P., Gottlieb, S. and LaBarge, J., 2016. "Engineering a functional 1-deoxy-D-xylulose 5-phosphate (DXP) pathway in Saccharomyces cerevisiae." Metabolic Engineering, *38*, 494-503.

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., **Moss, N. A.**, Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature, *496*, 528–532.

Fellowships and Grants

2017-2018 – Microbial Sciences Initiative Graduate Fellowship
2015-2016 – Trainee, Training Grant in Marine Biotechnology
2013-2014 – Regents Fellowship (UCSD/SIO), Nora and Alan Jaffe Fellowship, Robert L. Cody Memorial Fellowship, Henry L. and Grace Doherty Fellowship

Awards

2017 – Claude E. Zobell Fellowship, Marine Biology Division, SIO Department
2017 – Lynn Brady Student Travel Award, American Society of Pharmacognosy
2016 – Graduate Student Excellence Travel and Research Award, SIO Department

ABSTRACT OF THE DISSERATION

Biosynthetic strategies for chemodiversity generation in marine filamentous cyanobacteria

by

Nathan Aaron Moss

Doctor of Philosophy in Marine Biology

University of California San Diego, 2019

Professor William H. Gerwick, Chair

Organisms throughout the natural world have evolved different methods for biosynthesis of natural products, utilized in their native environs for defense, structure and signaling. One such method is the employment of modular polyketide synthase and non-ribosomal peptide synthetase enzymes to produce molecules from acetate and amino acid building blocks. The combinatorial capacity of these modules, combined with associated tailoring enzymes and unique protein-protein interaction strategies, has enabled a wide diversity of chemical structures and scaffolds. Many secondary metabolites or semi-synthetic derivatives have strong affinity for enzymatic targets and are used therapeutically to treat certain diseases. Photosynthetic marine filamentous

cyanobacteria are prominent sources of bioactive natural products and present unique enzymatic methods for generating new chemical structures. This thesis describes the use of chemical biology techniques, including biochemistry, genomics, stable-isotope labeled feeding studies, NMR and mass spectrometry-based isolation and structure elucidation to explore and understand these novel methods for generating chemical structure diversity in cyanobacteria. Following an introduction and background in chapter one, chapter two describes new strategies for biosynthesis in the production of type A malyngamides in the species *Okeania hirsuta* sp. PAB10Feb10-1 and sp. PAP21Jun06-1, including neofunctionalization of a lipoic acid synthesis enzyme and ketoreductase domain inactivation. Chapter three of this thesis describes an unprecedented combinatorial approach to natural products biosynthesis in *Moorea producens* sp. ASI16Jul14-2: identical protein-protein interaction motifs guide a polyketide synthase-generated fatty acid tail to disparate downstream NRPS modules in the same biosynthetic gene cluster, representing a new mechanism for generating combinatorial chemical diversity in the productions of vatiamides A-F. Chapter four of this thesis highlights investigations toward understanding the biosynthesis of the *t*-butyl group in apratoxin A, a cyclic lipopeptide derived from cyanobacterium *Moorea bouillonii* PNG5-198. Lastly, chapter five presents a comprehensive overview of the findings described here, and discusses future avenues for research in microbial biosynthesis and natural products drug discovery as a whole.

Chapter 1: Introduction

1.1 Natural product therapeutic use

Secondary metabolites have been defined as molecules which are not critical for immediate survival of an organism under laboratory conditions, but likely confer selective fitness to the organism in its natural environment.[1] Microbes, plants, and fungi frequently produce specialized secondary metabolites evolved for their defense and communication properties. In the pre-synthetic chemistry era, nearly all medicinal treatments were based on applications of rarified extracts of natural products or ingestion of a producing organism, frequently plant-derived.[2] Early 20th century developments in microbiology and an expanded chemistry research apparatus enabled the discovery of disease therapeutics based on single active naturally occurring molecules, including penicillin and streptomycin.[3] These initial developments blossomed into the discovery and characterization of many additional natural product therapies or derivative therapies of diverse molecular structure and function.[4] The terpene artemisinin, for example, is used to treat malaria, while the sponge-derived polyether macrolide halichondrin is forms the pharmacological basis of drug Eribulin, active against breast cancer.[5–7] Aminoglycoside neomycin, originally isolated from actinomycetes, is a potent antibiotic, while sponge-derived nucleoside cytarabine is used to treat cancer.[8–10] Many lipopeptide structures have been isolated with useful medicinal properties, such as the antibiotic daptomycin, while lipids like docosahexaenoic acid are used to treat inflammation.[11,12]

However, in the 1990s, modern pharmaceutical research largely departed from natural product-based discovery in favor of derivatives of pharmacophores developed via synthetic chemical libraries, an approach which lent itself to scalability and high-throughput medicinal chemistry workflows.[1,13] Many pharmaceutical companies reduced natural product screening capabilities and focus. Despite this, as of 2014 75% of anticancer drugs developed from 1940-2014 are based on a natural product pharmacophore or some iteration of a biological derivative.[14] Nevertheless, the combination of novel extraction and

screening techniques, needs for new targets, and modern health challenges such as multidrug-resistant bacteria are renewing interest in research into natural product-derived therapies.[15–18]

While historical natural products discovery has focused largely on easily culturable microbial organisms from soils and other terrestrial habitats, as well as plants and their extracts, the marine realm has also proved to be an excellent source of natural products.[19,20] Several compounds originating from the marine realm are currently approved for sale or in clinical trials (Figure 1.1).[21] Plinabulin, based on

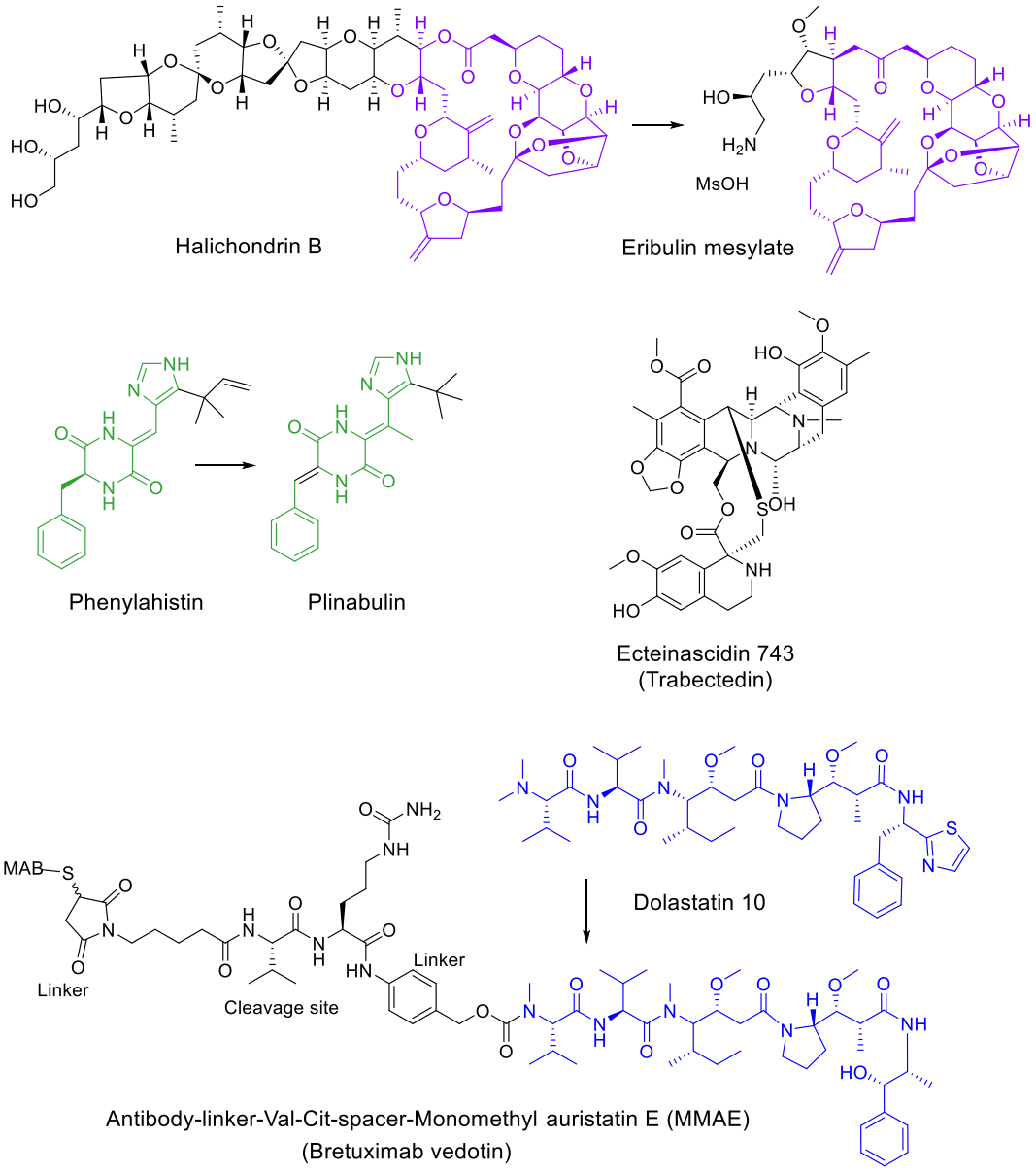


Figure 1.1. Drugs or drug candidates from marine sources. Colored molecular elements indicate natural product origin.

marine fungal-derived (-)-phenylahistin, is currently in phase III clinical trials.[22] Ecteinascidin 743, from a bacterial tunicate symbiont, is used to treat sarcomas in several different cancers.[23] Monomethyl auristatin E (MMAE), an antibody-drug conjugate based on the structure of dolastatin 10 (a mollusk derived peptide believed to originate from its cyanobacterial diet), was approved for clinical treatment of lymphoma and other cancers in 2017.[24,25]

As the pharmacopeia of useful natural products expands, both marine and otherwise, it is important to expand knowledge of their structures, origins, and applications. One important aspect governing the production of natural products is the mechanisms and enzymology by which the native producing organism biosynthesizes these unique molecules. A vast array of strategies is employed across the kingdom of life in the native production of natural products. The biosynthesis of natural products from a particular phylum, the marine cyanobacteria, is explored in greater depth in the ensuing thesis.

1.2 Biosynthesis of natural products

The diversity of natural product molecular structure is due to disparate strategies in biosynthesis, which is defined as the enzymatic transformations imparted by proteins to form a molecule within an organism. Significant evolutionary pressure has favored certain categories of genes which result in the ability to combine molecular building blocks and then differentiate them via tailoring and modification.[26] The long term evolution of microbes and plants combined with differences in structural morphology, local chemical environment, redox capacity, and pH of the producing system has resulted in a vast diversification in chemical species based on the framework of fundamental molecular building blocks.

This concept is manifested in several categories of natural product biosynthesis. Isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) via the mevalonate (in plants and eukaryotes) and non-mevalonate pathway (bacteria) can be combined by terpene synthases to form terpene isoprenoids. Different numbers of IPP and DMAPP may be combined, rearranged, oxidized, and reduced to form novel molecular structures. Isoprene is comprised of a single unit originating from DMAPP, while hundreds of combined IPP/DMAPP units form natural rubber. Tens of thousands of isoprenoid compounds, including steroids, hormones, and other volatile compounds have been

characterized from across the fungal, bacterial, and plant kingdom.[27–30] Pharmaceutically, this class is exemplified by taxol (Paclitaxel; Abraxane), an anticancer agent isolated from the Pacific Yew tree.[31]

Proteinogenic amino acids may also serve as molecular building blocks. In ribosomally synthesized and post-translationally modified peptides (RiPPS), a leader peptide signals for expression of short segments of DNA, which are then translated into peptide small molecules, which can be modified thereafter, such as nosiheptide from *Streptomyces*.[32,33] Related, non-ribosomal peptide synthetases (NRPS) biosynthetic pathways take an alternate approach, whereby single proteinogenic or non-standard amino acids are first activated with ATP by an adenylation (A) enzyme domain, and then combined with subsequent NRPS-activated amino acids downstream to form a polypeptide.[34] These amino acid A domains are contained within standalone enzyme modules which must also contain a condensation (C) domain and peptidyl-carrier protein (PCP). Combinations of these modules produce a variety of molecular scaffolds and may be modified by in-module epimerase (E) and reduction domains, or standalone modification enzymes such as acyltransferases and methyltransferases. Didemnin, an NRPS-encoded molecule originating from a marine tunicate *Tridemnum solidum*, was the first marine-derived drug to be tested in human cancer clinical trials.[35–37]

Another key biosynthetic scheme involves the use of polyketide synthases (PKS), of which there are historically three different major classifications: type I, with individual ketosynthase (KS) domains contained within separate interacting modules, type II, which employs iterative use of standalone heterodimeric enzymes, and type III, which uses an iterative KS homodimer to extend starter units with successive rounds of acetate or propionate extension.[36,38–40] However, additional, unclassified structural varieties have suggested that it may be more effective to avoid specific classification of the vast evolutionary diversity of PKS biosynthetic types.[41] Type I KS modules must contain a KS, acyltransferase (AT) and acyl carrier protein (ACP) domain, but frequently also feature NADPH-dependent KR and enoyl reductase (ER) domains, as well as redox-neutral dehydratases (DH) and MTs. Type I KS modules can often be organized into large biosynthetic gene clusters (BGCs), with several KS modules strung together in series to create a lipid chain possessing varying level of saturation and oxidation.[42] An

additional classification of PKS biosynthetic schemes include *trans*-AT pathways, whereby AT domains are not encoded within the KS modules, but rather, on a separate protein and act in-*trans* to sequences of KS-containing modules.[43] Used as an antibiotic, erythromycin from *Saccharopolyspora erythraea*, is a classic and well-studied example of a type I PKS-derived therapeutic molecule.[44]

In addition to biosynthesis schemes that contain only either NRPS or PKS modules, numerous pathways also contain both types of modules, referred to as hybrid PKS/NRPS pathways. In such a case the resulting molecule is frequently a lipopeptide, containing reduced units derived from acetate/propionate and amino acids. Often such molecules are cyclized to a macrocyclic lactone and, may contain genes encoding for 5-6 membered heterocycle formation. Combined with unique loading modules and tailoring domains, the combination thereof has led to a highly diverse set of molecular scaffolds. The anticancer compound bleomycin, isolated from *Streptomyces*, is an example of a hybrid PKS-NRPS pathway.[45,46]

1.3 Cyanobacterial biosynthesis

Due to selective pressures over time, some classes of organisms have evolved significant capacity for natural products biosynthesis. This manifests itself in the diversity of biosynthetic classes, number of different metabolites produced, and overall biosynthetic potential; that is, the number of complete pathways identified that do not have a known, isolated natural product but which may be activated under certain environmental conditions. Diversity in natural products is frequently furcated by molecule class – plants and fungi, for example, have evolved tremendous capacity for the biosynthesis of different volatile terpene molecules, while marine sponges and their symbionts produce a wide variety of indole alkaloids.[27,47,48] Actinobacteria produce a wide variety of PKS, NRPS, PKS/NRPS, and modified peptide natural products of clinical relevance, particularly antibiotics.[49,50] Possibly owing to their ancient origins and diverse habitat preference, as well as their capacity for symbiosis and endosymbiosis, the phylum cyanobacteria have evolved a broad capacity across several structural classes to encode for the production of bioactive natural products (Figure 1.2).[51,52]

Cyanobacteria produce many modified tetracyclic terpene compounds known as hopanoids, with

a similar structure as eukaryotic steroids, with broad function in contributing to membrane fluidity, permeability, and modulating external stresses. Several species also produce sesquiterpene geosmin and related analogs, as well as diterpenes noscomin and tolypodiol, though typically cyanobacterial terpenes have not been shown to be bioactive.[53,54] Alternatively, terpene building blocks are frequently incorporated into other pathway schemes to produce bioactive molecules: biosynthesis of the ambiguines, a unique indole alkaloid family, fuses terpene building blocks with tryptophan, forming a conserved core that is the basis for over a dozen analogs (Figure 1.2A).[55] In similar fashion, lyngbyatoxin biosynthesis features a fusion of geranyl-PP with a modified tryptophan-alanine dipeptide.[56–58] Saxitoxin, a modified alkaloid, is a notable and frequently encountered cyanobacterial and dinoflagellate neurotoxin that causes paralytic shellfish poisoning (PSP).[59] Saxitoxin biosynthesis involves the condensation of arginine with propionate, followed by methylation, amidinotransfer, oxidation, and carbamoylation in a highly unusual series of transformations initiated by a PKS loading monomodule (Figure 1.2B).[60] The widespread cyanobacterial alkaloid scytonemin, shown to be a strong UV protectant, also features an interesting biosynthetic origin. L-tryptophan is first converted to 3-indole pyruvic acid, followed by TPP-dependent coupling to *p*-hydroxyphenylpyruvic acid, cyclization to form a cyclopentanone ring and lastly a series of oxidations and dimerization to form the final scytonemin product.[61]

Cyanobacteria also feature RiPPs-based biosynthesis, which enables a combinatorial-style ability to generate cyclic peptide diversity. The family of patellamides, RiPPs-derived octopeptides, were found to be biosynthesized by *Prochloron didemnii*, an as-yet unculturable cyanobacterial endosymbiont of a marine ascidian.[62] These are part of a broader class of cyclic peptides, the cyanobactins, which have been isolated from many cyanobacterial species and are produced in part by a unique set of proteases which cleave the precursor peptide and subsequently foster cyclization.[32,63–65]

Another highly unique cyanobacterial biosynthetic pathway has been characterized in the production of polybrominated diphenyl ethers (PBDEs), which mimic the structures of synthetic flame retardant molecules and are found in great quantity distributed throughout marine environment.[66] Metagenomics and reconstitution of pathway enzymes illuminated the *bmp* pathway in *Hormoscilla*

spongeliae, a cyanobacterial obligate endosymbiont of *Dysideidae* sponges.[67] In this pathway, chorismic acid is converted to hydroxybenzoic acid, followed by bromination by a unique flavin-dependent enzyme Bmp5, then oxidatively coupled by Bmp7 to form an array of bicyclic polybrominated products and their respective isomers (Figure 1.2C).[68]

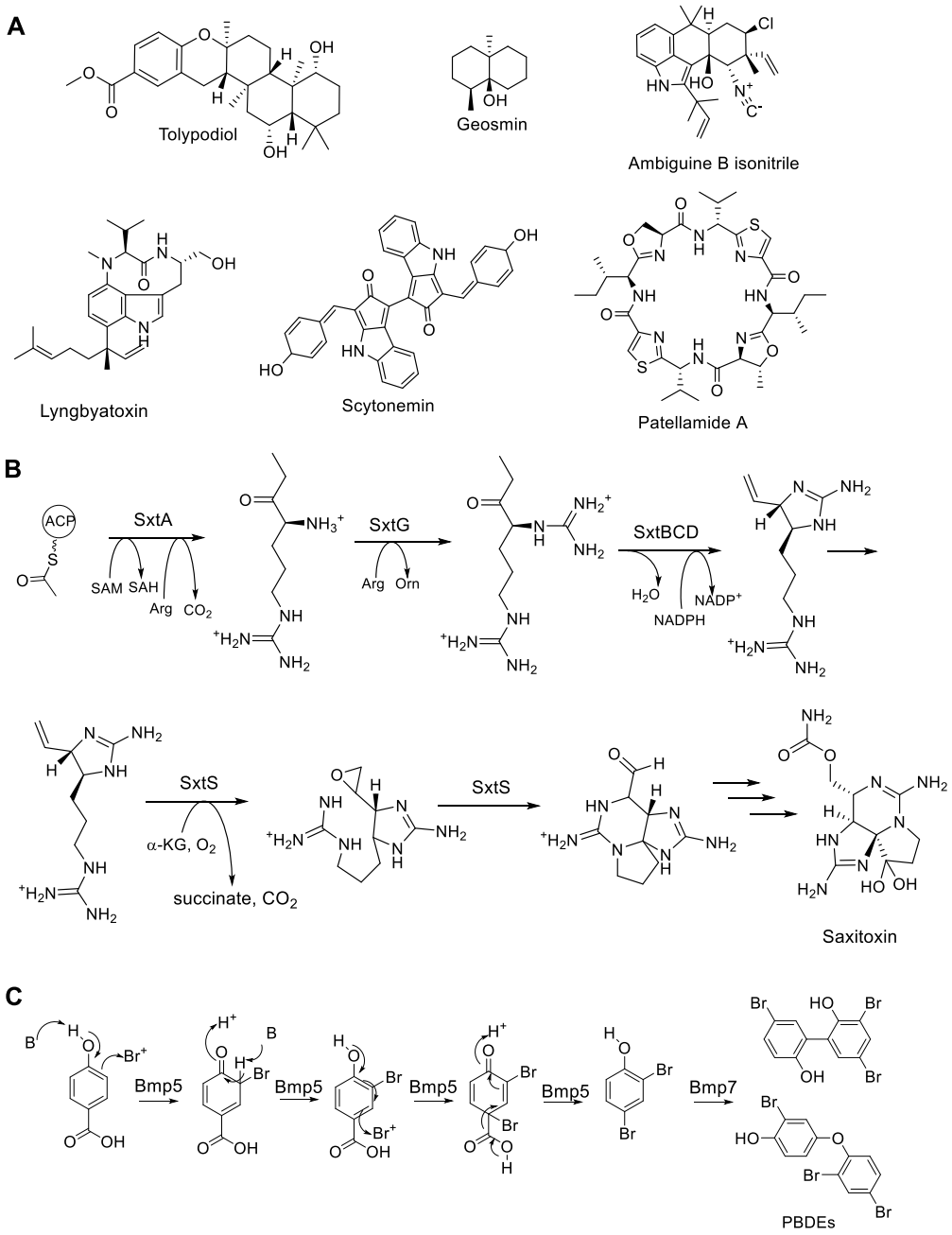


Figure 1.2. Cyanobacterial natural products and biosynthesis examples. A) Terpenoids, fused terpenes, alkaloids, RiPPs B) Saxitoxin biosynthesis. C) PBDE biosynthesis.

1.4 PKS and NRPS biosynthesis in cyanobacteria

A key area in which cyanobacterial biosynthetic diversity is displayed is in type I *trans*-AT PKS and NRPS pathways. *Trans*-AT PKS biosynthesis differs from *cis*-AT PKS biosynthesis largely by the exclusion of AT domains from individual KS modules.[69] Rather, the AT reaction is carried out by a unique standalone AT-domain. In such pathways, numerous KS modules that contain KR, DH, ER or MT domains may be encoded and acted upon sequentially by the catalytically active AT domain in *trans*. Additional interesting biosynthetic differences between *trans* and *cis* AT pathways include standalone reductive domains which are not embedded in a module, split modules, frequent introduction of heterocycles, and even "non-elongating" KS modules which contain KS modules but impart only reductive or methyltransfer functions.[43] Indeed, *trans*-AT pathways are found throughout the microbial kingdom, and there are several examples of this biosynthetic pathway in cyanobacteria (Figure 1.3). These include two close analogs of pederin: the first, cusperin, is from a non-symbiotic marine *Nostoc*, and a second, nosperin, is from a terrestrial cyanobacteria closely associated with lichen *Peltigera membranacea.*[70,71] Cyanobacterial isolate borophycin, a highly interesting boron-binding molecule, shares structural homology with boron-binding tartrolons likely produced by a trans-AT pathway in *Teredinibacter turnerae* T7901, a marine shipworm bacterial symbiont. However, biosynthesis of borophycins in the cyanobacteria was demonstrated through feeding studies only in the pre-genomics era, and the pathway has not been characterized.[72–74] Stable isotope feeding studies and genome analysis of *Leptolyngbya sp.* ISB3Nov94-8, illuminated the phormidolide pathway. This *trans*-AT PKS pathway results in the highly modified, cyclized, and polyhydroxylated 44-carbon polyketide known as phormidolide, a compound noted for its toxicity to brine shrimp.[75] Another *trans*-AT pathway originates from a terrestrial cyanobacterial source of the swinholides, a large family of macrocyclic analogs long isolated from environmental samples of *Theonella* sponges and diverse cyanobacteria, for example *Geitlerinema* and *Scytonema pseudohofmanni*.[76–79]

Cyanobacteria also feature NRPS pathways. Some examples of purely NRPS-derived modules include the cyanopeptolin family, "lariat"-style cyclic peptides which contain 3-amino-6-hydroxy-2-

piperidone (Ahp) as a unifying feature.[80] This pathway is loaded via a FkbH domain which activates glyceric acid for procession by subsequent NRPS modules. An additional example includes the antifungal hassallidins, glycosylated lariat-style peptides with modified fatty acid additions.[81] The latter group of molecules is widely distributed among heterocyst-forming cyanobacteria.[81] Similarly structured pathways

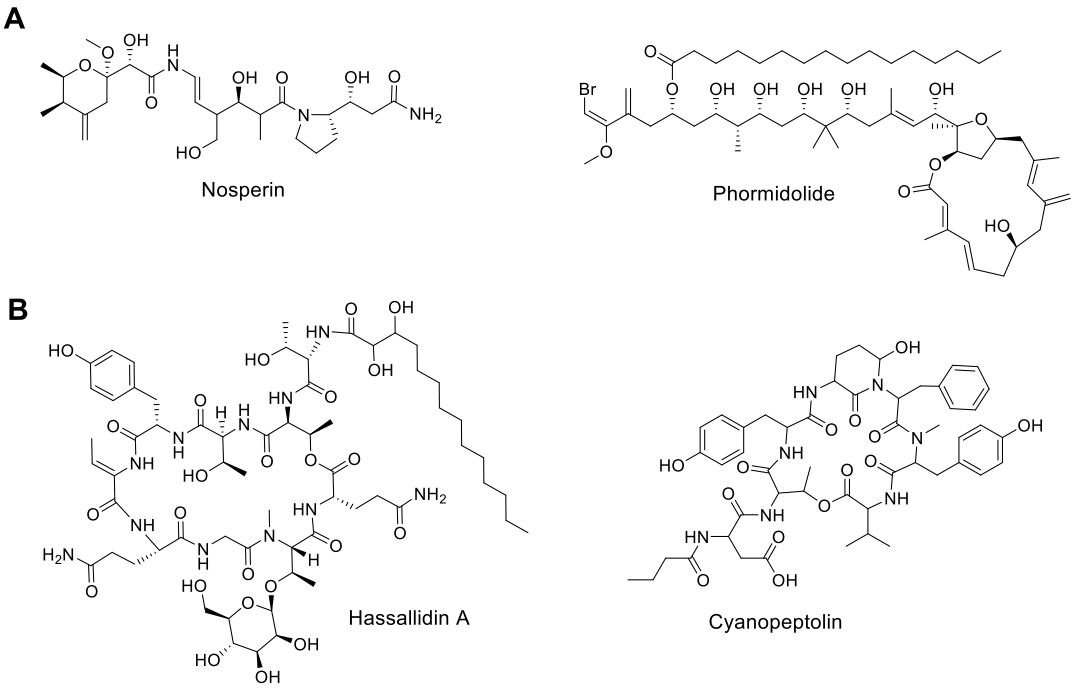


Figure 1.3. Cyanobacterial molecular diversity. A) *Trans*-AT and B) NRPS pathways from cyanobacteria.

include those for the nodularins and balticidins.[82,83] However, a survey of cyanobacterial chemodiversity reveals that cyanobacterial molecules rarely derive solely from canonical NRPS modules, for example, in tyrocidine biosynthesis in the gram-positive firmicute *Bacillus brevis*.[84] Frequently at least one KS module or non-amino-acid activating starter unit is involved in the pathway in some fashion. One might speculate that this is because canonical peptides may be susceptible to endogenous or exogenous protease degradation, thus reducing the evolutionary selection of exclusively multimodular NRPSs. It is more energetically efficient in terms of biosynthesis to rely on RiPPs-derived peptide molecule formation from canonical amino acids, followed by methylation, fatty-ester addition, or other manipulations to prevent degradation. By comparison, energetically and genomically expensive NRPS modules may be reserved

for pathways which more heavily modify and their amino acid cores, thus potency may be selected for with lower risk of defensive adaptation.

1.5 Hybrid PKS/NRPS biosynthesis in cyanobacteria

Dozens of bioactive molecules isolated from cyanobacteria contain, or are predicted to contain, both NRPS and PKS modules in their biosynthetic pathways. The multitude of these pathways and structures demonstrate that the inclusion of acetate subunits and their reduced derivatives along with a diverse array of amino acids enables the production of a wide array of structures capable of imparting biological activity. Indeed, given the relative dearth of exclusively type I *cis*-AT PKS-only and relatively sparse examples of NRPS-only derived molecules in cyanobacteria, one may speculate that significant evolutionary advantage is generated by the combination of PKS and NRPS modules in a single biosynthesis pathway.[85] Fundamental principles in pharmacology suggest that bioactive molecules must bridge a fatty-acid based cellular membrane and then subsequently bind to, react with, or block a polar reactive or structural target. As such, bioactive molecules are well served by possessing both nonpolar and polar moieties. This concept is captured in part by the calculation of a molecule's partition coefficient, logP, which is a measure of the concentration of a molecule in a mixture of immiscible aqueous and nonpolar solvents. An initial range of cLogP from -0.4 to 5.6 or <5.0 has been suggested as optimal for discovery of drug candidates.[86,87] Thus, one reason this class of molecule is discovered with such regularity in bioactivity screens is that they combine aliphatic moieties generated by PKS modules and fatty-acid activating starter units with the polar character imparted by the inclusion of amino acid residues.

As many cyanobacterial natural products were isolated prior to the genomic era, no biosynthetic pathway has yet been concretely assigned to every molecule, although the combination of acetate and amino-acid derived moieties infers a PKS/NRPS origin. These molecular families include analogs of tumonoic acid, aplysiatoxin, majusculamide, antillatoxin, microcolin, dolastatin, ulongamide and lyngbyastatins, among many others.[88]

One of the most biochemically well studied cyanobacterial PKS/NRPS pathways is that which encodes for curacin A, an anti-tubulin agent isolated from *Moorea producens* 3L in 1994 and whose BGC was characterized in 2004 (Figure 1.4).[89] Structurally, there are several unique features in curacin A, including a terminal alkene and a cyclopropyl ring from which one carbon is derived from C2 of acetate, unusual for a PKS/NRPS pathway. Several individual protein domains have been expressed and explored biochemically from this pathway. These including characterization of a cyclopropyl-forming multifunctional chlorination and β-branching gene cassette found in numerous natural product pathways,[90] a sulfotransferase-thioesterase domain pair which generates the terminal alkene,[91] and a dual function GCN5-acetyltransferase (GNAT) domain which carries out both acetyl transfer and decarboxylation, forming the basis for pathway loading in a several microbial pathways.[92] Additionally, a PKS module C-MT was purified and used to understand C-MT active site structure and stereocontrol.[93]

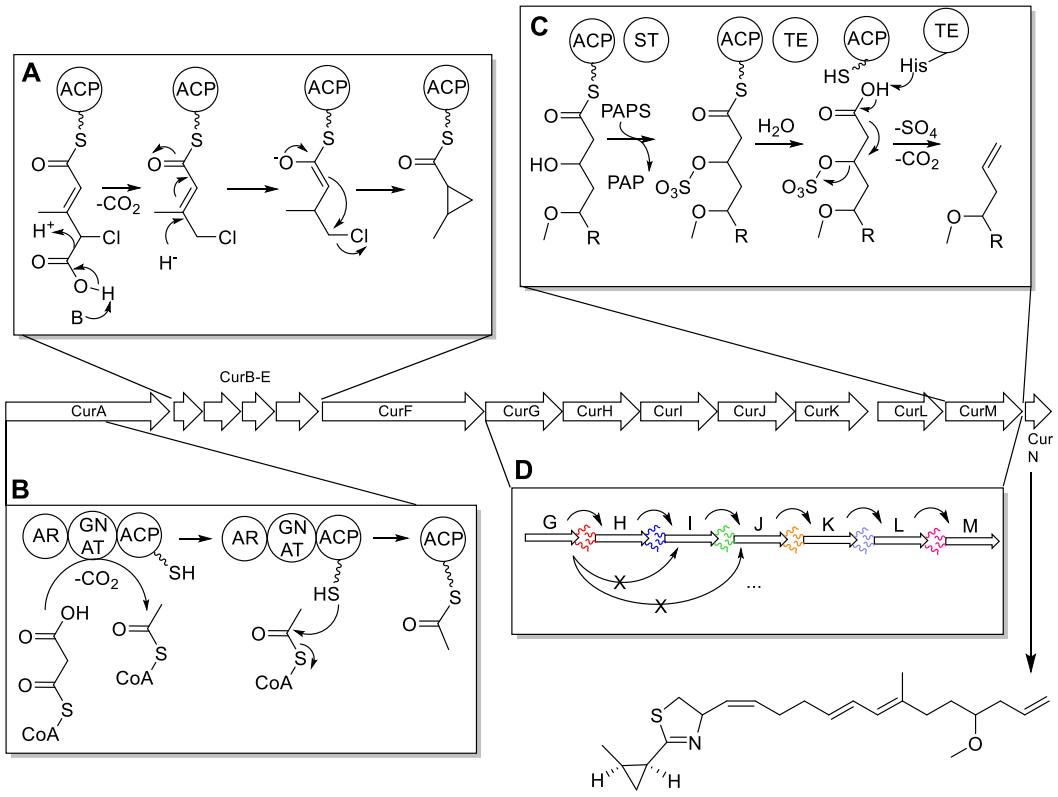


Figure 1.4. Curacin A biosynthesis. A) β-branching to create cyclopropyl ring by CurB-E (Gu *et al.*, *Nature*, 2009). B) Decarboxlaytion and acetyl transfer to initiate CurA ACP (Gu *et al.*, *Science*, 2007). C) Sulfotransfer and thioesterase activity to create terminal alkene via CurM (Gehret *et al.*, *J. Biol. Chem.*, 2011, Gu *et al.*, *JACS*, 2009). D) Type II docking domains enable selectivity between CurG-CurM (Whicher *et al.*, *Chem and Biol.*, 2013)

Additional PKS/NRPS pathways reveal uniquely functionalized domains utilized by cyanobacteria for diversification of their natural products. For example, in cylindrospermopsin biosynthesis, amidinotransfer onto free glycine is followed by several cyclization and extension transformations. CyrG and H then mediate a unique reaction to form a uracil ring by introduction of a guanidine moiety to the carbonyl-carbons of a bound thioester and β-ketone.[94] Amidinotransfer is mediated by CyrA, the first example of an arginine-glycine amidinotransferase in a prokaryote (Figure 1.5A).[95] Another example is provided by BarB1 and BarB2, chlorination enzymes involved in the barbamide pathway from *Moorea producens* 3L. These were shown to trichlorinate isoleucine in stepwise fashion for incorporation into barbamide.[96–98] Anatoxin, a potent neurotoxin isolated from several freshwater cyanobacterial strains, also bears unique features in its biosynthesis. After activation of proline, it is reduced to form a positively charged nitrogen, and then extended by acetate, followed by a TE-mediated intramolecular cyclization to form an unusual 8-C ring structure. Orf1 of the anatoxin pathway, suspected to also contribute to cyclization, shares homology with fatty-acid binding lipocalin-like proteins from eukaryotes, as well as enzymes suspected in heterocyclization in stigmatellin and alnumycin, among others.[99–101] Discussed later, a similar protein is suspected to participate in malyngamide head group cyclization, a focus of Chapter two in this thesis.[102]

Another well-studied loading scheme is encountered in the terminal alkyne and alkynylbromide forming loading enzymes of the jamaicamide pathway, a molecule originally isolated from *Moorea producens* JHB (Figure 1.5B).[103] Terminal alkynes and their brominated analogs are widespread in cyanobacterial natural products, also being encountered in the veraguamides,[104] carmabins,[105] dragomabins,[106] and antanapeptins.[107] The first three proteins, JamABC, of the jamaicamide pathway were expressed and experimentally determined that JamA activates a hexenoic and transfers it to JamC ACP, while a membrane-bound desaturase, JamB, converts the initial alkene to an alkyne.[108] The JamABC cassette provides further evidence of the flexibility and novelty of cyanobacterial biosynthetic initiation schemes.

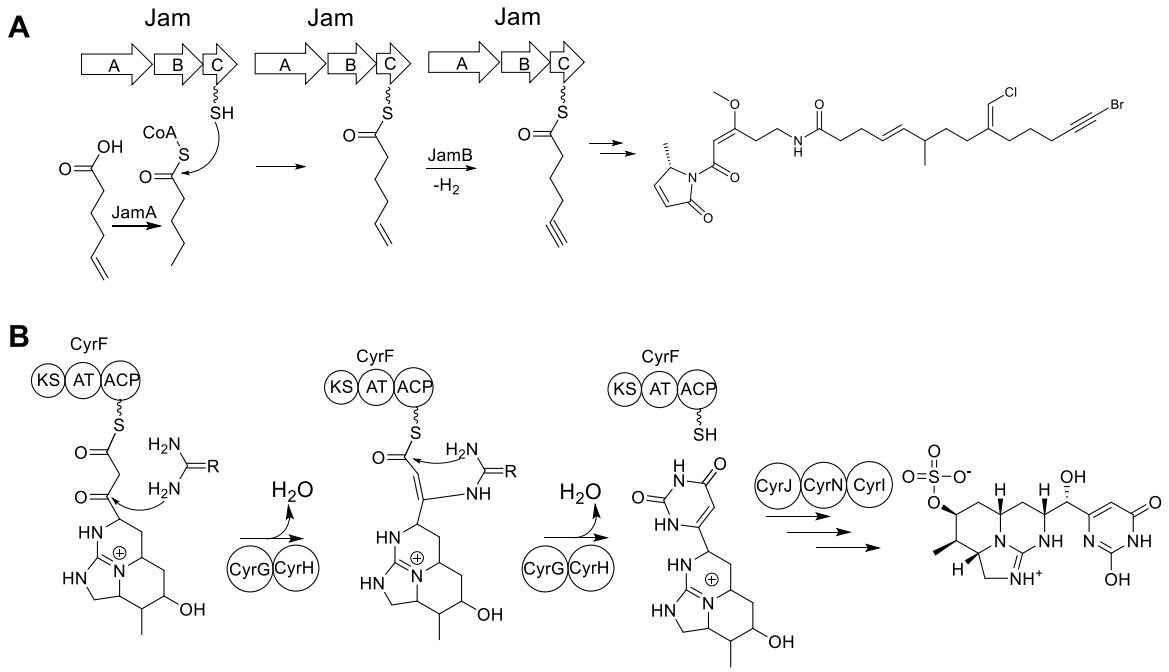


Figure 1.5. JamABC and cylindrospermopsin biosynthesis. A) JamABC cassette activates hexenoic acid to hexenoyl-CoA, followed by transfer to JamC ACP, then JamB enables desaturation. B) CyrGH form uracil de novo on CyrF-bound intermediate.

1.6 Docking and collinearity in *cis* assembly line biosynthesis

A general term for PKS, NRPS, and PKS/NRPS gene clusters that add molecular building blocks in a defined order to form a final molecule is termed "assembly line" biosynthesis. A key principle in PKS/NRPS assembly line biosynthesis is the coordination of individual protein modules with one another in a concerted order that enables the assembly of building blocks in a defined and exact sequence. Each enzyme and domain performs a specific biosynthetic step in the process with each enzymatic reaction cycle, a concept also known as collinearity.[109] This specific intermodular interaction is facilitated by noncovalent interactions between a subset of amino acid residues located on docking domains: small, nonenzymatic motifs that flank the N and C terminus of KS and NRPS modules in *cis* assembly line biosynthesis, discussed below in greater detail.[110]

1.6.1 Docking domains in *cis*-AT PKS/NRPS biosynthesis

Type I docking domains were first characterized in the 6-deoxyerythronolide B synthase (DEBS) pathway by X-ray crystallography and biochemical complementation of chimeric modules.[111] The

concept of the selectivity of docking was expanded in examining the curacin A pathway, where seven consecutive KS modules, CurG though CurM, interact in exactly the same sequence to produce the methylated and methoxylated fatty acid portion of curacin A; this section also bears a terminal alkene. Remarkable specificity for each docking domain's cognate downstream partner was demonstrated via chimeric recombination of Cur docking domains with pikromycin KS modules. These docking domains were shown to bear a distinct structure from type I docking domains by crystallography, and were thus categorized as type II docking domains.[112] A third type of docking domain, called "COM" or communication domains, enable the junction between NRPS modules, exemplified in the tyrocidine, gramicidin S, and surfactin biosynthetic pathways.[113] A set of β-sheets form a hand-like structure on the downstream N-terminal docking domain, while a single helix on the upstream interacting thiolase noncovalently interacts with it to foster interdomain selectivity.[113] An additional type of docking domain, distinct from type I, II, and COM domains has been observed both between hybrid pathway PKS and NPRS module junctions as well as NRPS-NRPS junctions. While the C-terminal thiolase docking domain in this interaction has been structurally characterized only in the rhapdopeptide pathway, the downstream N-terminus has been explored more thoroughly in the biosynthesis of tubulysin, epothilone, and rhabdopeptide/xenortide-like peptides (RXPs).[114–116] By solution NMR and crystallography, it was found to displays an $\alpha_1\beta_1\beta_2\alpha_2\alpha_3$ tertiary structure 3-5 residues on the β2-sheet in this latter type of docking domain interacts noncovalently with the same number of residues on a single upstream helix.[114,116,117]The upstream C-terminal thiolase docking domain was theorized to be largely "natively unfolded" and thus unable to undergo structural analysis.[114] Nevertheless, a recent study was able to structurally characterize the sequence of the five terminal residues of the C-terminal docking domain as contributing to a "specificity code" by interaction with residues 24-30 on the β_2 sheet of the N-terminal docking domain by constructing a Cdd-$(Gly\text{-}Ser)_n$-ddN protein and analyzed via solution NMR.[116]

Lastly, another type of docking domain which has been observed for upstream NRPS module interaction with downstream KS modules between EpoB and EpoC in the epothilone pathway.[118] This interaction is characterized by a series of positively charged residues at the C terminus of the upstream

docking domain and negatively charged residues on the downstream module, with electrostatic interactions thought to modulate communication. This type of domain is also encountered in the bleomycin and yersiniabactin pathways.[119,120] Other motifs for structural interaction between individual *trans*-AT pathway domains and modules exist, but they are not discussed here.

1.6.2 Collinearity in natural products biosynthesis

It is clear that numerous strategies have been developed in natural products biosynthesis pathways to ensure selectivity of assembly line module interaction. These include both the use of structurally different docking domains in individual pathways, such as in the epothilone pathway, as well as developing an amino acid docking "code" to ensure cognate pairs of the same structural class match in a pathway with numerous modules, as in curacin A biosynthesis (Figure 1.4D). As a result, natural product molecules are nearly always produced identically with regards to the incorporation of backbone KS and NRPS-derived moieties. While minor structural variations occur frequently, such as differing amino acid incorporation owing to adenylation domain promiscuity, methylation, acetylation, and halogenation, rarely is structural variation observed as a result of the fundamental abandoning of the collinearity in assembly line biosynthesis.

However, a few notable exceptions to collinearity have been studied. For example, in pikromycin biosynthesis, occasional skipping of the ketide extension occurring in the final module, pikAIV, leads to thioester cleavage and macrocyclization of a truncated molecule, 10-deoxymethynolide, as well as a pikAIV-generated cyclic product, narbonolide.[121] Both analogs are then further processed and glycosylated to their final products. An unusual bi-product non-collinear pathway is also observed in the biosynthesis of surugamides in *Streptomyces* sp. JAMM992. In a linear pathway containing four tightly spaced NRPS modules, *surA*-*surD*, it was found that *surA* and *surD* are responsible for biosynthesis of surugamides A-E, while *surB* and *surC*, despite being flanked by *surA* and *surD*, independently produce surugamide F (Figure 1.6A).[122] In the biosynthesis of myxobacterial stigmatellin, "stuttering" of modules, whereby a molecule chain is re-transferred to a prior module, in this case the KS of StiH or StiJ, is observed (Figure 1.6B).[99] Lastly, in a striking and unique example of non-collinearity in PKS/NRPS

biosynthesis, the many analogs generated by the thalassospiramide pathway arise as a process of multiple iterations of stuttering, skipping, and assembly line re-processing to form a diverse family of compounds (Figure 1.6C).[123] However, these examples remain rare compared to the hundreds of described

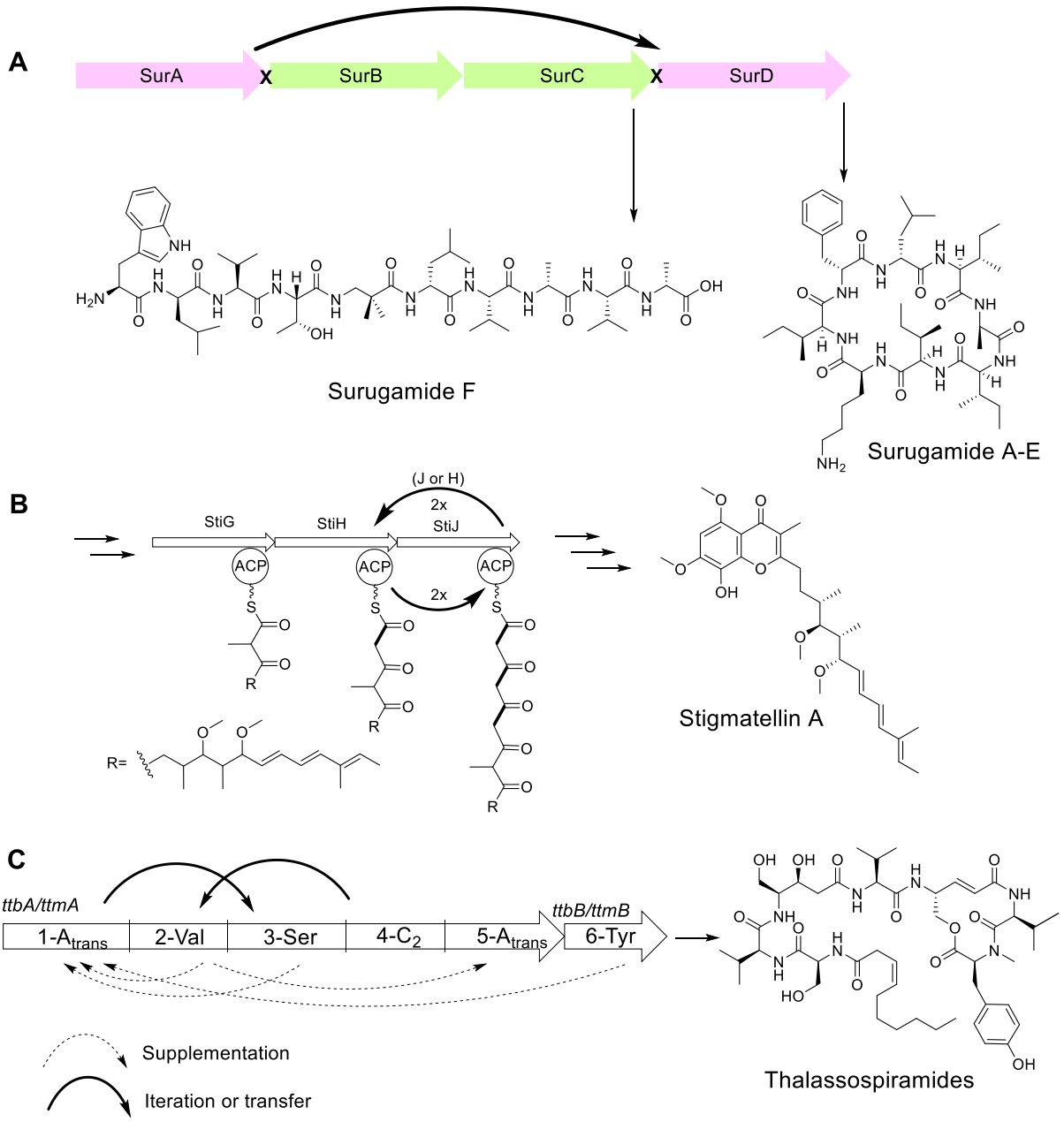


Figure 1.6. Non-collinear biosynthetic pathways. A) Surugamide biosynthesis in *Streptomyces* sp. JAMM992. Bold arrow indicates surugamide A-E production via module skipping from SurA to SurD. SurBC, sandwiched between SurA and SurD, produces surugamide F. (Ninomiya *et al.*, *ChemBioChem*, 2016). B) Stuttering in stigmatellin A biosynthesis in *Stigmatella aurantiaca*. Stuttering at StiH or StiJ creates extra acetate extensions. (Gaitatzis *et al.*, *J. Biol. Chem.*, 2002). C) Biosynthetic "multitasking" in thalassospiramide biosynthesis in *Tistrella* and *Thalassospira*. Solid arrow indicates iteration, dashed arrow indicates amino acid supplementation to modules missing A domains. (Ross *et al.*, *JACS*, 2013).

biosynthetic pathways which do follow the pattern of collinearity.[124] Thus, discoveries of non-collinear methods for natural products biosynthesis represent an important expansion of our knowledge of biosynthetic strategies.

1.7 Analyzing biosynthesis: new technologies and strategies

Vastly increased amounts of genome and pathway information is available as more organisms and environmental samples are sequenced at a greater rate due to lower sequencing costs.[125] As such, computational tools are being deployed with increasing frequency and fidelity to predict the molecular products of biosynthesis pathways.[126] The prediction of pathways by software tools provides a top-down approach to create molecular libraries for screening in health-related assays and complements traditional methods for isolation of natural products isolated from environmental collections. Understanding new methods for biosynthesis of natural products, assembly-line pathway or otherwise, adds to the body of knowledge of natural products and their evolutionary origins. These biosynthetic strategies may subsequently be understood in a genomic context, whereby the use of computational algorithms such as DELTA-BLAST or pHMMs can be employed to predict gene function based on sequence comparison of biosynthetic domains and modules to previously characterized domains and modules and the proteins they produce.[127,128] Homology-based search programs that predict molecular substructures from genomic data, such as NAPDOS, SANDPUMA, and NRPSPredictor, have been developed with the advent of increasing computational power and data collection in the last decade.[129–131] These tools have been employed in bioinformatic BGC analysis platforms, exemplified by AntiSMASH, which enables facile analysis of genomic samples for BGCs by employing and synchronizing an array of predictive tools encompassing several different classes of biosynthesis.[132] Other tools, such as GNPS, attempt to use analytical data to inform molecular structure prediction by visualization of spectral data.[133] More comprehensive genomes-to-natural products efforts such as GNP attempt to offer a seamless path from genomics to natural products prediction by connecting LC/MS/MS data to genomic data.[134] However, these approaches are limited by the need for meticulously well-curated data and are limited to easily interpreted BGCs with well-known modules and domains, as well as certain classes of easily ionizable molecules.

Another mechanism for the discovery of novel biosynthetic schemes is the use of heterologous expression of BGCs. In this method, biosynthetic genes of unknown function may be transferred from their original nucleic acid context, whether synthesized based on sequencing data, from an environmental metagenome library, or native organism, into an easily manipulated heterologous host.[135–139] This transfer invites manipulation of the gene cluster to produce a heretofore uncharacterized product, as well as knockout analyses to determine gene function. This approach offers a systematic approach to identifying unknown gene and proteins, a gap which cannot be adequately fulfilled by computational predictive tools which are based specifically on known gene function.

In summation, expanding our knowledge of biosynthetic strategies improves computational predictive algorithms and informs heterologous expression strategies. Ongoing technological advancements such as these ultimately will enable utilization of sequenced BGCs for the generation of new natural products, whether in a semi-synthetic manner from purified recombinant enzymes, environmental or culture isolation, or heterologous expression. These natural products may be screened in high-throughput fashion to search for assay hits, leading to more rapid identification of hits to prioritize in the drug development pipelines, for utilization in improving human health.

1.8 Dissertation contents

Owing to the long history of isolation of unique PKS/NRPS-derived metabolites from cyanobacteria, I undertook research to further understand the sources and mechanisms of generating molecular structure diversity in their natural products. More specifically, this scholarship concerns the biosynthesis of malyngamide C acetate and malyngamide I, and the discovery of the vatiamides and elucidation of their biosynthetic pathway, and finally a study of the biosynthesis of the initial steps of the apratoxin A pathway. As benthic filamentous cyanobacteria have largely been resistant to genetic manipulations,[140] strategies which did not employ genetic manipulation of the producing organisms were deployed to study the interesting natural products emanating from these organisms.

Chapter 2 describes an in-depth examination of evolutionary radiation in the malyngamide biosynthetic pathway. Over two dozen type A malyngamides have been isolated since their discovery in

1978, without firm assignment of the producing organism. Mass spectrometry-based observations of characteristic masses of malyngamide C and malyngamide I, combined with genome sequencing of two laboratory cultures of *Okeania hirsuta*, prompted a full analytical characterization of the produced molecules from these two strains. This ultimately enabled assignment of the *mgc* and *mgi* pathways in *Okeania hirsuta* PAB10Feb10-1 and PAP21Jun06-1, respectively. Phylogenetic comparison, a pharmacological knockout, and expression and biochemical examination of the penultimate KR domain in each pathway, implicated a mutated, inactive KR (KR^0) as the source of the cyclohexanone ring characteristic to all type A malyngamides. The strong evolutionary selection which generates this cyclohexanone ring, which is common to all 28 type A malyngamide analogs, appears to result from the inactivation of a single KR domain. This understanding enables theorization of a comprehensive model for the biosynthesis of all type A malyngamides, and is an excellent example of the unique ways by which cyanobacteria have generated significant chemical diversity.[102]

The discovery of six lipopeptide analogs, vatiamides A-F, along with their biosynthetic pathway, is presented in Chapter 3. Vatiamides A-D were initially isolated from a cultured of *Moorea producens* ASI16Jul14-2 collected from Vatia Bay, American Samoa. All analogs share a PKS-derived alkynyl or bromoalkynyl vinyl chloride fatty acid tail common to the jamaicamides.[103] A bioassay-guided approach initially led to the nonbrominated vatiamide A/C analogs, and mass spectrometry uncovered their brominated pairs: vatiamide B/D. However the head group of vatiamides A/B featured a trimethylated tyrosine, while the head group of vatiamides C/D featured a bicyclic ketide-extended glycine derivation with methyl and methoxy modifications. Genome assembly of combined short and long-read sequencing uncovered a 90-kb hybrid PKS/NRPS pathway. This pathway features proteins predicted to encode for the biosynthesis of vatiamides A/B via a series of PKS modules and a tyrosine-incorporating adenylation module. Separately, the pathway also appears to use a module-skipping event to generate vatiamides C/D via a glycine-incorporating A domain and four subsequent KS modules. However, two unspoken-for NRPS modules in the middle of the pathway were predicted to incorporate alanine and valine, but initially appeared not to be functionally expressed. An analysis of the N-terminal docking domains of the A-

domains incorporating tyrosine, glycine, and alanine revealed 100% identity in the first 45 amino acids, implying a loss of assembly line selection and further implicating an Ala-Val containing product with the alkynyl jamaicamide tail. It was hypothesized that a fifth and sixth analog would be generated via these modules, but the exact structure was unknown. Therefore, a search was made for this theoretical dipeptide using a fluorescent coumarin-core probe with azide functionality designed to undergo a click reaction with any alkyne-bearing molecules. This was applied to crude ASI16ul14-2 as a screening tool, and HPLC-UV-ESI-MS detection at 350 nm revealed unbound probe, free alkynyl fatty acid tail, vatiamides A and C, and an additional uncharacterized peak with a $[M+H]^+$ of 466. This latter compound ionized relatively poorly and was in lower abundance than vatiamides A/B and C/D. Subsequent purification of this compound and MS of a brominated analog revealed vatiamides E and F, containing a head group of D-alanine and L-*N*-methyl-valinamide. Thus, all three identical docking domains functionally interact with the upstream Jam-tail encoding sequence, and result in the production of a total of six vatiamides produced from a single pathway. The interaction between an upstream C-terminal ACP docking domain with three identical downstream docking domains on three distinct NRPS modules represents a new non-collinear "multichannel" or "multiplexing" combinatorial biosynthetic scheme.

Chapter 4 features a discussion of *t*-butyl biosynthesis in apratoxin A. Stable isotope labeled feeding to the producing organism, *Moorea bouillonii* PNG 5-198, provided the groundwork for later analysis of the recombinant AprA protein, which serves as the loading module for the multimodular apratoxin PKS/NRPS pathway. ^{13}C-NMR analysis of cultures fed with labeled methionine, propionate, or acetate compared to an unlabeled control indicated significant enrichment of all of the predicted SAM-derived methyl groups in apratoxin A, including the three methyl groups of the pivalyl moiety generated by AprA. Additionally, the FabD fatty acid synthase (FAS) acyltransferase was found to initiate malonate transfer to the AprA ACP domain. Ultimately, this research contributed to uncovering an iron-dependent methyltransferase which both methylates and decarboxylates dimethylmalonyl-ACP to the pivalyl-ACP moiety processed by the subsequent PKS genes.

Finally in chapter 5, a synopsis of the research will be presented along with an outlook for future endeavors for the projects described here. A summation of this research and its impact on our understanding of natural product biosynthesis in cyanobacteria will be discussed, as well as justification for continuing scholarship in the field of the biosynthesis of natural products as well as their applications in human medicine and biotechnology.

Chapter 2: Ketoreductase domain dysfunction expands chemodiversity: malyngamide biosynthesis in the cyanobacterium *Okeania hirsuta*

2.1 Abstract

Dozens of type A malyngamides, principally identified by a decorated six-membered cyclohexanone head group and methoxylated lyngbic acid tail, have been isolated over several decades. Their environmental sources include macro- and microbiotic organisms, including sea hares, red alga, and cyanobacterial assemblages but their true producing organism has remained enigmatic. Many type A analogs display potent bioactivity in human-health related assays, spurring an interest in this molecular class and its biosynthetic pathway. Here we present the discovery of the type A malyngamide biosynthetic pathway in the first sequenced genome of the cyanobacterial genus *Okeania*. Bioinformatic analysis of two cultured *Okeania* genome assemblies identified 62 and 68 kb polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) pathways with unusual loading and termination genes. NMR data of malyngamide C acetate derived from ^{13}C-substrate-fed cultures provided evidence that an intact octanoate moiety is transferred to the first KS module via a LipM homolog originally associated with lipoic acid metabolism and implicated an inactive ketoreductase (KR^0) as critical for six-membered ring formation, a hallmark of the malyngamide family. Phylogenetic analysis and homology modeling of the penultimate KR^0 domain inferred structural cofactor-binding and active site alterations as contributory to domain dysfunction, which was confirmed by recombinant protein expression and NADPH binding assay. The carbonyl retained from this KR^0 ultimately enables an intramolecular Knoevenagel condensation to form the characteristic cyclohexanone ring. Understanding this critical step allows assignment of a biosynthetic model for all type A malyngamides, whereby well-characterized tailoring modifications explain the surprising proliferation and diversity of analogs.

2.2 Introduction

Marine microorganisms have been extraordinarily rich sources of bioactive natural products, and several have been approved as drug treatments.[20] A modified synthetic derivative of dolastatin 10, originally isolated from the sea hare *Dolabella auricularia* but whose true source is likely cyanobacteria,

has been approved to treat lymphoma.[24] A family of frequently isolated marine bioactive molecules, the type A malyngamides, present an array of bioactive properties including anticancer and anti-inflammatory activity.[141,142] The first type A malyngamides were isolated and elucidated in 1978, and in the intervening decades, 28 analogs have been characterized.[143] Type A and B malyngamides differ in the head group structure; type A mostly features a modified six-membered ring with a ketone functionality, while type B analogs contain a pyrrolidone head group (Note A1).[144] Related, hermitamides and serinol-derived analogs contain a lyngbic acid tail with a modified amino acid head group.[145,146] Type A malyngamides differ from one another principally in the tailoring of the head group, with methylation and oxidations found disparately at nearly all carbon positions about the cyclohexanone ring, and include additional features often deriving from epoxidation or acetylation. Predicted amino acid linkage (Gly, Ser, Thr, or β-Ala), and lyngbic acid chain length (12–16 carbon) can vary as well.[143,147–152] However, the majority of the malyngamides possess a 14-carbon lyngbic acid tail and glycine linkage as those in this study. Thus, the core structure of these malyngamide analogs is remarkably conserved across biogeographically diverse locales and sources.

The attributed environmental sources of the malyngamides include red algae,[153] invertebrates,[150] and most commonly, assemblages of filamentous cyanobacteria of the genus *Moorea* (previously *Lyngbya*), isolated in pan-tropical distribution.[154,155] While a range of biological activities have been reported for the 28 analogs, no single target has been identified, though the epoxyketone and enone moieties present in most analogs are a well-studied pharmacophore for proteasome inhibition.[156]

Previous research suggested that non-axenic cultures of *Okeania hirsuta* PAB10Feb10-1 ("PAB") and PAP21Jun06-1 ("PAP") were producers of this natural product class, however, a detailed analytical characterization and genetic context for their production was not included in those studies (Note A2).[157–159] Here we establish PAP as the producer of malyngamide I, and PAB as the producer of malyngamide C and C acetate, and identify and characterize their complete polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) biosynthetic pathways from their genomic sequences (Figure 2.1).

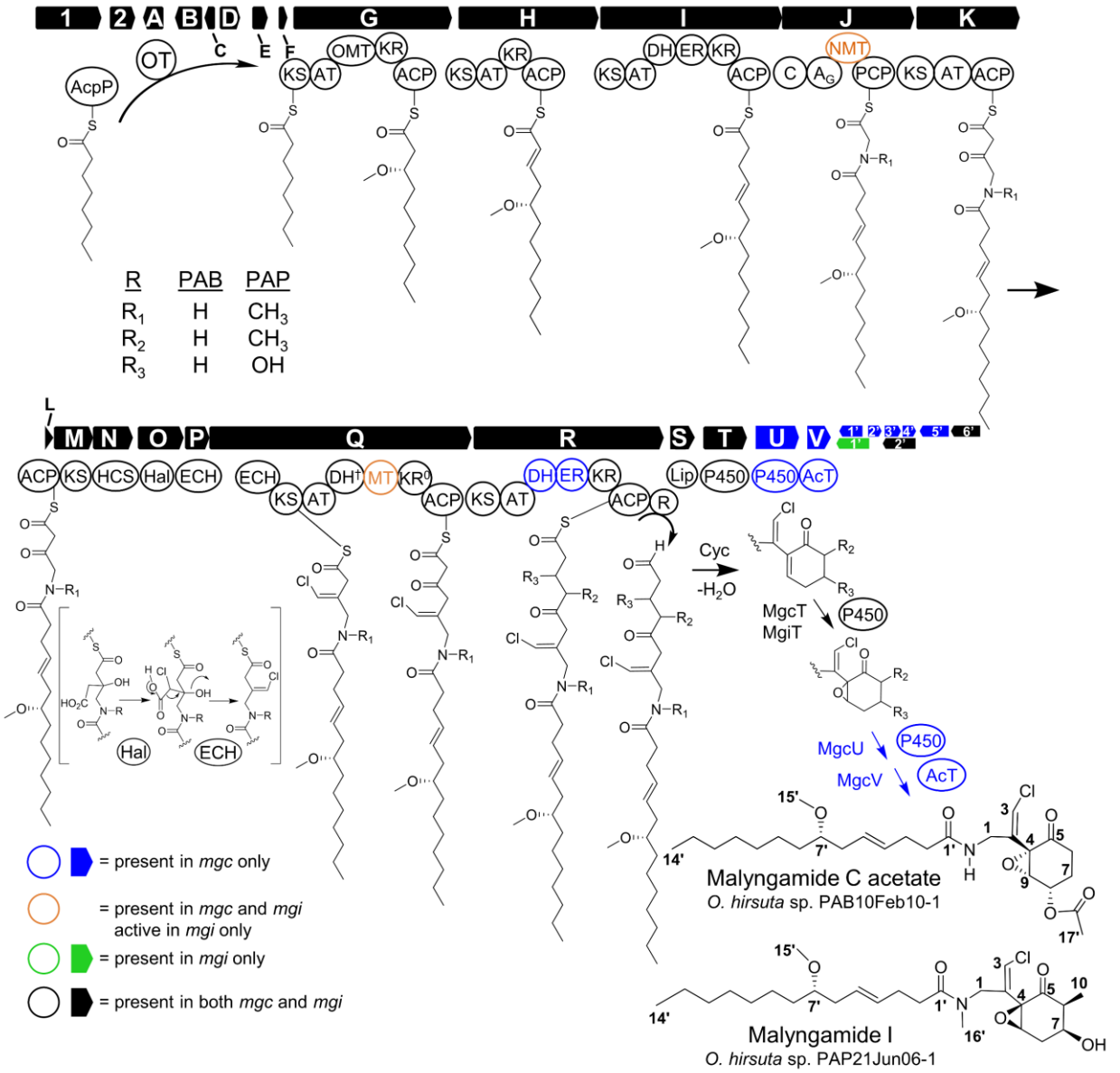


Figure 2.1. Combined malyngamide C acetate (mgc) and malyngamide I (mgi) biosynthesis pathways from Okeania hirsuta PAB and PAP. Abbreviations: acyl carrier protein (ACP), AcpP (fatty acid synthase acyl carrier protein), acetyltransferase (AcT), adenylation (A), acyltransferase (AT), condensation (C), C-methyltransferase (MT), dehydratase (DH), enoyl-CoA hydratase (ECH), enoyl reductase (ER), halogenase (Hal), hydroxymethylglutaryl-CoA synthase (HCS), ketosynthase (KS), ketoreductase (KR), lipocalin-like (Lip), N-methyltransferase (NMT), O-methyltransferase (OMT), octanoyltransferase (OT), cytochrome P450 (P450), peptidyl carrier protein (PCP), thioester reductase (R). Blue color indicates genes or domains found only in PAB. Orange indicates domains that are structurally present in both PAB and PAP, but active in PAP only. Green genes are present in PAP only. Black colored genes are present in both pathways. ^{0} – inactive domain. † – structurally present in PAP, catalytic residue absent in PAB, no activity predicted due to KR^{0}.

Typically, cyclic cyanobacterial lipopeptides are loaded by either fatty acid CoA-ligases,[108] amino acid or acyl-AMP synthetases,[160] GCN5-acetyltransferase (GNAT) domains,[89] or pseudo-GNAT

domains.[161] However, both malyngamide pathways are predicted to initiate biosynthesis using MgcA/MgiA, a LipM-like octanoyltransferase (Note A3).[162] As filamentous marine cyanobacteria have not yet proven to be genetically tractable, we sought to examine malyngamide biosynthesis via several alternative methods, including genomic analysis of the genus *Okeania*, stable-isotope labeled feeding of malyngamide subunits, pharmacological knockouts, bioinformatics and phylogeny of the MgcA octanoyltransferase, and homology modeling and biochemical analysis of the MgcQ inactive ketoreductase (KR^0) domain.

Ultimately, J_{CC} calculated between ^{13}C-enriched cyclohexanone carbons clarified the biosynthesis of the head group in both pathways, which is not eminently predicted by the domain order within the PKS/NRPS modules. We deduced that a thioester reductase (R) domain in the final PKS module generates a terminal aldehyde, and its electrophilicity allows a Knoevenagel condensation by nucleophilic attack from the alpha carbon of the MgcK-installed acetate unit. Subsequent P450 activity and acetyltransfer is proposed to generate the final characteristic head groups in each molecule. We present a compelling model for the biosynthesis of type A malyngamides wherein the biosynthetic keystone is the KR^0-enabled Knoevenagel condensation. With this new insight, well understood KS, NRPS, and tailoring enzymatic reactions easily explain all of the type A malyngamide analogs. Implications for PKS/NRPS biosynthesis strategies are discussed, as this pathway demonstrates a unique example of module dysfunction leading to a striking radiation of chemical diversity.

2.3 Results

2.3.1 Structure confirmation.

Chemical extracts of *Okeania hirsuta* strains PAB and PAP were analyzed by HPLC-ESI-MS/MS. Peaks with $[M+H]^+$ 456.2 and 498.2 were identified in PAB, and one major peak with $[M+H]^+$ 484.2 was found in PAP. MS/MS fragmentation patterns of $[M+H]^+$ 456.2 and of $[M+H]^+$ 484.2 matched the MS/MS fragmentation patterns of pure compound standards of malyngamide C and malyngamide I (Figure A2.1A, A2.1B), respectively, while the MS/MS pattern of $[M+H]^+$ 498.2 suggested an acetylated analog of malyngamide C (Figure A2.1C, Note A4-A5). Putative malyngamide C acetate HR-ESI-MS

was $[M+H]^+$ = 498.2630, corresponding to molecular formula of $C_{26}H_{41}ClNO_6$, and optical rotation (OR) $[\alpha]_D^{25}$ –34.6 (*c* 0.1, CH_3OH). Putative malyngamide I HR-ESI-MS was $[M+H]^+$ = 484.2840, corresponding to formula of $C_{26}H_{43}ClNO_5$, and OR $[\alpha]_D^{25}$ + 40.0 (*c* 0.1, CH_3OH). HR-ESI-MS values and OR for each accord with previously obtained values.[154,155] Malyngamide C acetate structure was confirmed by 1D ^1H NMR and ^{13}C NMR (Table A2.1, Figure A2.9-A2.10), and malyngamide I structure was confirmed by ^1H NMR and ^{13}C NMR, COSY, HSQC, and HMBC (Table A2.2, Figures A2.14-2.17).

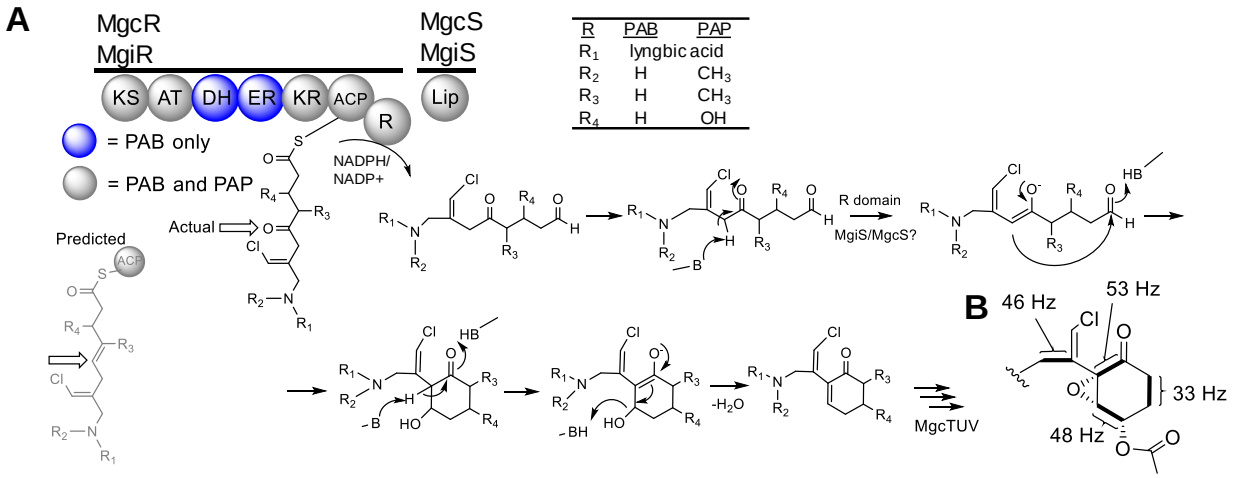
2.3.2 Genomic information and pathway analysis.

PAB and PAP genome assembly information is summarized in Table A2.3. A phylogenomic tree generated from conserved genes in diverse cyanobacteria indicates *Okeania* is closely related to *Trichodesmium* and has considerable evolutionary distance from planktonic *Synechococcus* and *Prochlorococcus* (Figure A2.2). Unlike *Moorea*, the *Okeania* genome contains the nitrogen-fixing *nif* genes, and both *Okeania* were experimentally determined to fix nitrogen under N-starvation using the acetylene reduction assay (Figure A2.3). Two closely related PKS/NRPS pathways were detected in PAB and PAP via antiSMASH[132] and DELTA-BLAST,[128] and were predicted to encode for malyngamide C acetate ("*mgc*" cluster, 68 kb in length) and malyngamide I ("*mgi*" cluster, 62 kb in length), respectively. The predicted open reading frames and proteins for the biosynthesis of these pathways are discussed below (Figure 2.1, Table A2.4A-A2.4B). PCR was used to confirm correct pathway assembly and close short gaps (Table A2.5).

A LipM-like octanoyltransferase, MgcA, is predicted to initiate each pathway by transferring an octanoyl bound to fatty acid synthase (FAS) acyl carrier protein (AcpP, generally ACP) to the MgcG KS active site cysteine. MgcA is slightly promiscuous toward decanoyl-ACP, as trace amounts of compounds with +28 AMU, $[M+H]^+$ 526 in PAB and $[M+H]^+$ 512 in PAP, respectively, were observed at a ratio of approximately 20:1 via MS (Figure A2.4). This substrate flexibility is not entirely surprising, as malyngamides D and E feature homologs of lyngbic acid that are extended by two saturated carbon atoms.[143] Downstream of mgcA, the genes mgcB and mgcC appear to encode for transposases, possibly

reflective of the combinatorial diversity of PKS/NRPS pathway loading genes. MgcD is an *sfp*-type phosphopantetheinylase that activates ACPs involved in PKS pathways with phosphopantetheine.[163] MgcE is another transposase, and MgcF is a small membrane-bound protein with homology to a short noncatalytic portion of JamB in the jamaicamide pathway. MgcG is the first PKS module, whose KS active site is initially octanoyl-bound by action of MgcA. This octanoyl moiety undergoes acetate extension by the MgcG AT (acyltransferase)-catalyzed condensation with malonyl-CoA. The resulting beta-carbonyl functionality is first reduced to a hydroxyl group by the MgcG KR, and then methylated by an *O*-methyltransferase (*O*-MT). MgcH adds an acetate unit and reduces the prior carbonyl to a *trans* double bond via a KR followed by dehydration; the latter transformation is likely performed by the downstream MgcI dehydratase (DH) domain, as observed in other PKS/NRPS pathways.[89,164,165] MgcI lengthens the bound molecule by an acetate unit and fully reduces the beta-carbonyl by activity of a KR, DH, and enoylreductase (ER), thus completing the synthesis of ACP-bound lyngbic acid.

MgcJ is an NRPS module with an adenylation (A) domain that activates and then installs a glycine residue, and in the PAP pathway also contains an *N*-methyltransferase (*N*-MT) domain that methylates the newly introduced amino acid using S-adenosyl methionine (SAM). In the PAB pathway, the *N*-MT is structurally present but inactive. Subsequently, MgcK installs another intact acetate unit with no reduction of the preceding beta-carbonyl. MgcL through MgcP install a chlorinated *exo*-cyclic methylene group via an HMG-CoA synthase-like mechanism. [90,166] Another malonyl-CoA is condensed to an acetate by MgcQ, but unexpectedly, the MgcK-installed acetate carbonyl at C-5 is not reduced by the MgcQ KR. Therefore, the KR is an inactive KR^0, discussed subsequently. The *C*-methyltransferase (*C*-MT) within MgcQ/MgiQ is active in the *mgi* pathway (in PAP), but not in the *mgc* pathway, (in PAB) which is missing a required active site His (Figure A2.5A). Lastly, the PAB MgcQ DH lacks a catalytic site His and thus is not predicted to be active; while the same domain in PAP does contain the active site His, this is a null point due to the pathway's inactive KR^0 (Figure 2.1, Figure A2.5B).



R	PAB	PAP
R_1	lyngbic acid	
R_2	H	CH_3
R_3	H	CH_3
R_4	H	OH

Figure 2.2. Cyclization scheme. a) Illustration of predicted cyclization mechanism of malyngamide C acetate, catalyzed by MgcR and MgcS. Blue colored domains are found only in PAB and not PAP. Active site acid-base chemistry deprotonates the carbon adjacent to the C-5 carbonyl, forming an oxyanion enoyl intermediate, which then collapses to nucleophilically attack the C-9 aldehyde, forming a transient alcohol at C-9. A second deprotonation at C-4 forms a second enolate intermediate, which collapses to lose hydroxide and forms the cyclohexenone core structure of malyngamide K. P450 and acetylation tailoring follows to form the final head group depicted in b) which indicates J_{CC} values derived from malyngamide C acetate extracted and purified from [1,2-$^{13}C_2$]acetate-fed and [1,2-$^{13}C_2$]glycine-fed PAB cultures.

The final KS module, MgcR and MgiR, are structurally divergent in the two organisms; in the PAB pathway, the module reduces the MgcQ-installed carbonyl with a full reductive cassette (KR, ER, DH) whereas the PAP pathway contains solely a KR, resulting in a hydroxyl group at C-7. Crucially, in both strains, MgcR terminates with an R domain and is predicted to generate a terminal aldehyde via a $2e^-$ transfer, thereby generating a suitable electrophile for cyclization, though there is some ambiguity as to which gene is specifically responsible for the mechanism, as discussed below in feeding studies and J_{CC} analysis section (Figure 2.1, Figure 2.2A). MgcT, a P450 with significant homology between pathways, is then predicted to catalyze epoxidation of the newly generated double bond. Here, biosynthesis halts in PAP, producing malyngamide I. In PAB, MgcU, an additional P450 is predicted to oxidize C-8 to an alcohol, followed by acetylation with O-acetyltransferase MgcV, thereby producing malyngamide C acetate. Orf 1 and 2 are enzymes of unknown function found just upstream of MgcA, and are homologous to those in the *mgi* pathway. Several ORFs of unknown function are found downstream of each pathway, flanked downstream by a kinase regulator common to both pathways.

2.3.3 ^{13}C-octanoate feeding and MgcA phylogeny.

LipM-like octanoyltransferase functionality has not been observed before in the context of initiating a PKS or hybrid biosynthetic pathway. MgcA was examined by feeding of [1-^{13}C]octanoate to *O. hirusta* PAB over a growth period of two weeks, followed by purification and ^{13}C-NMR comparison of resonance integrals to unlabeled malyngamide C acetate. A 3x enrichment of the ω-8 (C-7') resonance was observed compared to unlabeled malyngamide C acetate, indicating direct utilization of the intact

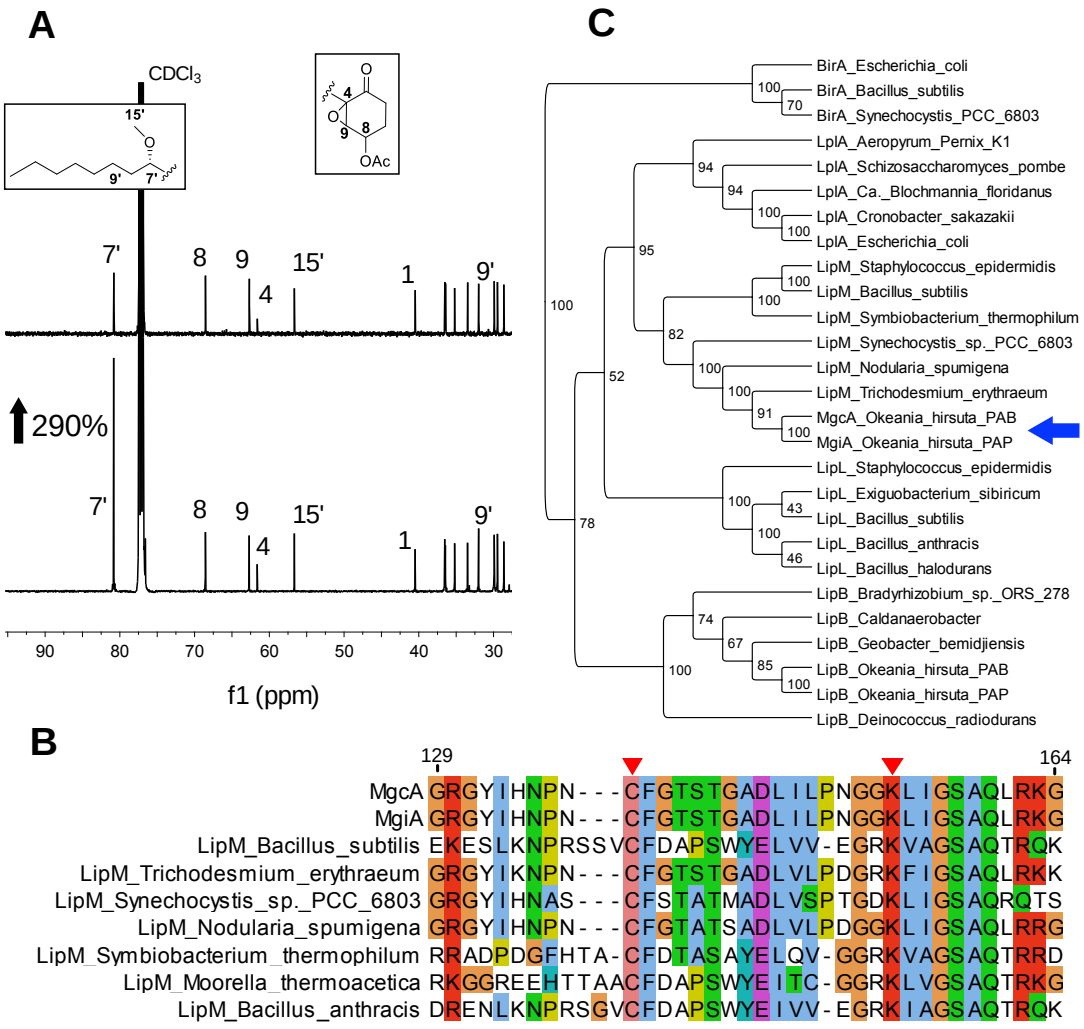


Figure 2.3. [1-^{13}C]octanoate fed batch enrichment and LipM phylogeny. a) Comparison of ^{13}C-NMR C-7', C-8, C-9, C-4, and C-15' resonances from malyngamide C acetate isolated from native abundance sample (top) and sample fed with [1-13C]octanoate (bottom). Percentage increase in intensity of C-7' signal depicted next to C-7' signal. b) alignment of LipM octanoyltransferases with MgcA/MgiA. Active site cysteine and lysine are indicated with red triangles. c) Phylogenetic tree of select PFAM03099 protein sequences. Topology threshold percentages depicted at node junctions.

octanoate moiety into the pathway (Figure 2.3A, A2.11, Table A2.6). The modest enrichment of C-1' (~50%) [1-^{13}C]octanoate supplementation may results from beta-oxidation of C-1 – C-2 of the labeled octanoate into [1-^{13}C]acetate building blocks.

MgcA and MgiA were analyzed by phylogenetic comparison within PFAM03099, the biotin/lipoate A/B protein ligase family. Key active site cysteine and lysine residues are conserved between MgcA, MgiA, *B. subtilis* LipM octanoyltransferase, and other putative cyanobacterial LipM analogs (Figure 2.3B). A phylogenetic tree was created based on the core structure of PFAM03099, and MgcA and MgiA clade with other putative LipM enzymes with a consensus support value of 95% vs. LipL, the nearest PFAM03099 clade (Figure 2.3C). LipM clearly falls outside the alternate family clades of BirA, LipB, Lpl, and LplA. In *B. subtilis*, LipM acts as an intermediate, transferring octanoate between octanoyl-FAS ACP and a lysine active site on the E2 lipoyl domain (LD) of 2-oxoacid dehydrogenase complexes.[162] However, no such LD is found between *mgc* pathway Orf1 and MgcG. Therefore, it is unclear specifically how MgcA transfers the C-8 moiety from octanoyl-ACP to the MgcG KS active site. However, based on results of the ^{13}C-labeled acetate and octanoate feeding experiments, phylogenetic analysis, and close proximity to MgcG PKS machinery, we assign the activity of MgcA and MgiA as responsible for initiating the malyngamide pathways. Interestingly, both *O. hirsuta* strains also contain LipB; while this is a functionally homologous lipoic acid synthesis enzyme, it differs in structure and sequence homology to LipM.[167] Indeed, many cyanobacteria contain both LipB and LipM homologs, including closely related *Trichodesmium erythraeum*, where a homologous LipM is not found adjacent to any biosynthesis pathway. In *Okeania*, we speculate that LipB satisfies the essential requirement for octanoyl transfer in lipoic acid synthesis, thus allowing the reprogramming of LipM to initiate a hybrid PKS/NRPS pathway; to our knowledge, this neofunctionalization of LipM is without precedent and represents a unique involvement of a member of this enzyme family to contribute to secondary metabolism.

2.3.4 Ancymidol pharmacological knockout.

A P450 is theorized to epoxidize C-4 – C-9 in malyngamide I and malyngamide C acetate, and an additional P450 in PAB is predicted to hydroxylate C-8. To evaluate this possibility, we provided cultures of *Okeania hirsuta* PAB with ancymidol, a P450 inhibitor, for a period of two weeks.[168] LC/MS results indicated a relative decrease in malyngamide C acetate production and a concomitant relative increase in the production of shunt product $[M+Na]^+$ peaks (Figure 2.4). To confirm that shunt products were indeed malyngamides, MS^2 data was collected for each shunt product and compared to fragments derived from known masses malyngamide C and malyngamide C acetate. Differences in product ion masses by the same molecular weight change as the parent ion, an MS^1 elevated [M+2] peak corresponding to chlorination, as well as MS^2-based networking confirmed that the shunt products were malyngamides (Figure A2.6). A species corresponding to the mass of didehydro-deoxy-malyngamide C, *m/z* 474.2,

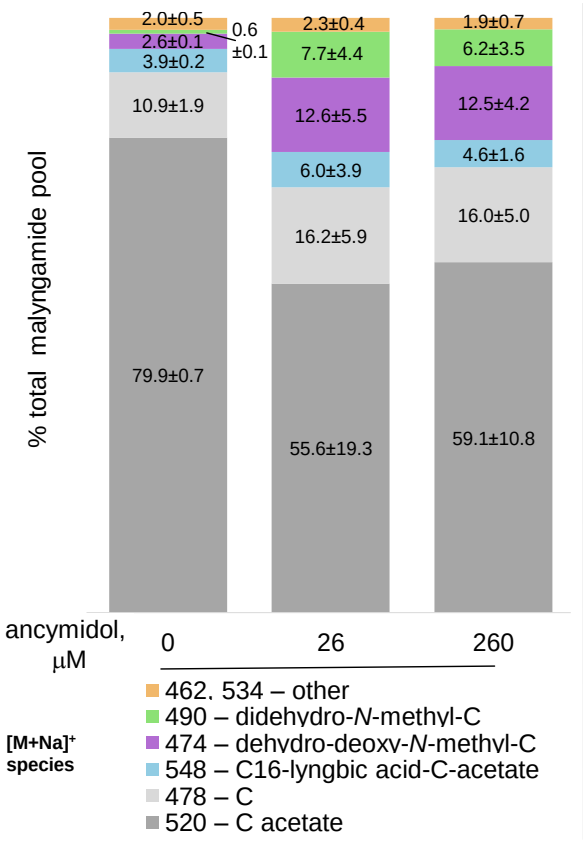


Figure 2.4. Ancymidol pharmacological knockout. Bar segments are labeled with the respective percentages of the total pool of each chemical species with standard deviation.

showed a relative increase from near nonexistent to approximately 10% of total malyngamide pool, while didehydro-*N*-methyl-malyngamide C, *m/z* 490.2, increased from near baseline to approximately 5% of total malygamide pool. The relative increase of these species when treated with a P450 inhibitor is intriguing; we speculate that this may occur as a result of kinetic retardation of pathway throughput in which intermediate species are held longer in each KS/NRPS turnstile.[169] This would enable domains with a structurally inefficient *N*-MT (the *C*-MT is catalytically inert) to have greater access to the MgcJ-bound intermediate, generating more of the *N*-methylated species.

2.3.5 ^{13}C-labeled acetate, glycine feeding studies.

To explore KS loading, biosynthesis of the head group, and confirm construction of the skeletal core as predicted by informatics, [1,2-$^{13}C_2$]acetate and [1,2-$^{13}C_2$]glycine were provided to live cultures of *O. hirsuta* PAB. Incorporations of labeled acetate and glycine units were determined by the appearance of flanking shifts in the ^{13}C NMR, indicating *J*-coupling between intact pairs of ^{13}C-enriched atoms (Figure 2.2B, A2.12-A2.13, Table A2.7). [1,2-$^{13}C_2$]glycine supplementation resulted in the emergence of flanking peaks only between C-1 and C-2, as expected. By contrast, [1,2-$^{13}C_2$]acetate supplementation showed flanking peaks for all intact acetate-derived carbon atoms, with the exception of C-3 which is derived from C-2 of acetate via the MgcN HCS-mediated reaction. As expected, the appearance in malyngamide C acetate of flanking ^{13}C NMR shifts at positions C-14' – C-7' from [1,2-$^{13}C_2$]acetate provision indicates that acetyl-CoA subunits are extended via FAS machinery until octanoate is preferentially transferred by MgcA to the first KS module of the pathway. Differing levels of [1,2-$^{13}C_2$]acetate incorporation support the conclusion that different biosynthetic processes are responsible for C-7' to C-14' versus the other acetate-deriving carbon atoms in the molecule. The coupled signals for C-4 – C-9, C-1', C-3' – C-5', and C-16' – C-17' averaged 15.1% ± 3.0% of the unenriched signal intensity whereas those for C-7' – C-14' averaged 20.8% ± 2.2 % of the unenriched signal (P-value = 0.001 by one-way ANOVA) (Figure A2.7). The intensity of some flanking signals could not be accurately measured due to overlap with neighboring peaks and these were not included in the analysis. The results suggest higher enzyme processivity and turnover associated with the biosynthesis of the ω-1 to ω-8 section of the lyngbic acid tail versus the

remainder of the malyngamide chain. This is possibly due to different processivity and turnover rates in standalone type II FAS systems versus type I PKS megasynthase modules, consistent with the hypothesis that MgcA directs a FAS-derived octanoyl-ACP to MgcG.

An analysis of J_{CC} coupling values between adjacent carbons was pivotal to understanding the biosynthesis of the cyclohexanone head group. Intact acetate incorporations were identified for C-4 – C-5, C-6 – C-7, and C-8 – C-9, installed by MgcK, MgcQ, and MgcR by collinearity (Table A2.7). These results were not in accordance with initial predictions for formation of the cyclohexanone ring. Bioinformatic analysis of the gene cluster indicated that C-5 – C-6 should be reduced and converted to an alkene by a theoretically functional KR/DH domain in MgiQ (PAP strain), whereas C-5 should be reduced to a hydroxy group by a theoretically functional KR/DH^0 (PAB strain). However, these predicted products did not provide for a reasonable mechanism for carbocyclization that involves the alpha carbon (C-4) of the MgcK-derived acetate unit. One possibility considered was that perhaps the MgcQ KR was a vestigial, inactive KR^0, in which case the C-5 carbonyl would remain intact. In fact, this appears to be the case, as is detailed in the subsequent section (Figure 2.2).

The coupling of C-8 – C-9 infers that an intact acetate unit is loaded by MgcR, followed by chain release due to the R domain subsequent to the MgcR ACP. Some R domains such as CpaS are redox incompetent, and solely catalyze Dieckmann cyclization, a non-redox intramolecular Claisen condensation.[170] In this latter case, these R domains substitute an L for Y in the YXXXK short-chain dehydrogenase (SDR) catalytic motif, and possess a conserved Asp residue which is required for cyclization. However, the MgcR/MgiR R domains contain the catalytic residues required for SDR reduction and NADPH binding (Figure A2.5C), as in LtxA and Lys2.[56,171] Therefore, it is expected that an aldehyde is generated at C-9.[56,171,172] As a result of the carbonyl derived from MgcK and MgcQ KR^0, the pK_a of the C-4 proton is significantly reduced, making it an attractive reactive carbon for cyclization with the C-9 aldehyde electrophile generated by MgcR R. Cyclization is proposed as follows: deprotonation of C-4 leads to a stabilized enolate anion species, which collapses back to reform the C-5 ketone such that the enolate double bond electrons attack the terminal C-9 aldehyde, forming a C–C bond between C-4 and

C-9 and a C-9 hydroxyl, which is then eliminated, representing a complete intramolecular Knoevenagel condensation (Figure 2.2).

The exact responsible proteins and timing of cyclization may be due to a combination of MgcR R domain and MgcS, a small lipocalin-like protein that has homologs implicated in PKS/NRPS cyclization in several microbial natural products, including the cyanobacterial metabolite anatoxin A.[99,173,174] As lipocalins have been identified as mostly non-enzymatic fatty acid-binding and transport proteins of eukaryotes, we speculate their role in malyngamide biosynthesis is that of a stabilizing entity, positioning or conforming the molecular chain such that C-4 is proximal to C-9 so as to promote carbon-carbon bond formation. The intermediate product resulting from cyclization is malyngamide K, an analog independently isolated from environmental sources.[152,159] Subsequent tailoring produces malyngamide C acetate and I, respectively, from the *mgc* and *mgi* pathways.

2.3.6 MgcQ KR expression, phylogeny, and modeling.

The hallmark C-5 ketone which is present in all type A malyngamides is installed by MgcK; however, as discussed, an initial bioinformatic analysis of MgcQ/MgiQ inferred C-4 – C-5 reduction to an alkene in PAB and generation of a C-5 hydroxyl group in PAP. A series of active site and tertiary structure alterations in the MgcQ/MgiQ KR^0 appears to have rendered this domain nonfunctional. We explored the viability of the *mgc* and *mgi* KR catalytic domain (KRc) via phylogeny and homology modeling. Key KRc regions and active site residues have been studied in detail previously and indicate regions important for activity and stereochemistry: the NADPH binding site, stereocontrol by the presence of either the "LDD" motif or the "W" motif, catalytic site, and "lid" helix region which encloses the active site.[175–177] Alignment of MgcQ domains from PAB and PAP to known active microbial KR domains as well as MgcG, H, and I show an unusual E variation in the first glycine of the conserved GGXGXXG diphosphate-binding "P-loop" in the NADPH binding site (Figure 2.5), in addition to other, unshared variations in residues 7–13. The inactive domain with the most similar NADPH binding site variant that has been characterized is EpoE,[178] which contains a G7D mutation at the same location, though EpoE also contains an Arg at the position of the Tyr proton donor. An analysis of the PAB MgcQ

KR catalytic site compared to active and inactive KR domains reveals some heterogeneity in sequence: the PAB MgcQ KRc contains a catalytic Tyr155, while the PAP MgcQ KRc contains a His at the same position. However, active KRc domains from CurJ, JamJ, AprI also contain the His alteration, implying that the His may also serve as a proton donor in KRc reduction (Figure 2.5). KRc domains from AmphI, NysI, and PimS2 are inactive examples as well; they contain a catalytic site Glu among other noncanonical residue changes. StiH contains an active site His, yet is inactive, possibly due to a Glu substitution at the Lys118 position (MgcQ KRc numbering) required for catalysis.[99,179] Overall, catalytic activity is known to be dependent on several steric and electrostatic interactions between enzyme and cofactor reaction participants. As such, alignment of the active site lid sequence in MgcQ indicates a deletion of several residues relative to active KRc domains, a modification that is also observed in the inactive KRc Raps2-6 involved in the biosynthesis of rapamycin.[180] However, the MgiQ KR in PAP shows no such truncation. To summarize, while MgcQ and MgiQ contain catalytic site Tyr155/His155 and Lys118, they lack either the conserved LDD or W stereocontrol motifs, a bulky Glu7 in the NADPH binding P-loop, and MgcQ contains a truncated lid.

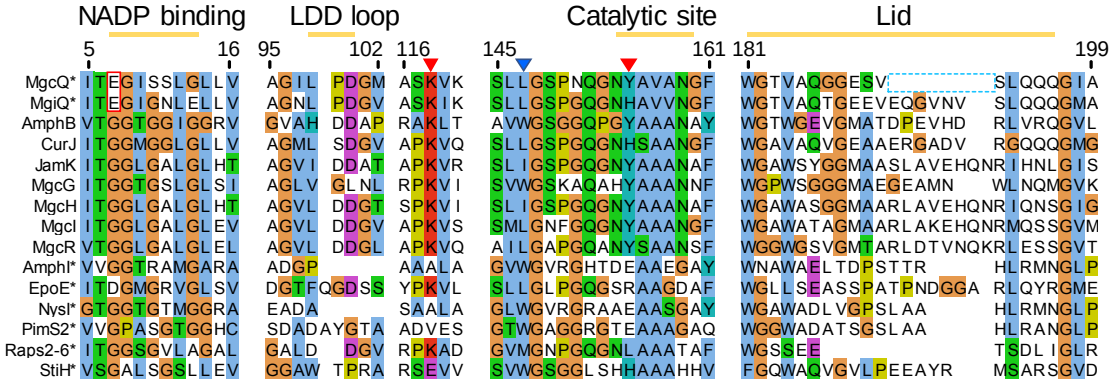


Figure 2.5. Sequence alignment of key regions of the KRc domain. Numbering accords with the sequence of the KRc^0 of MgcQ. G7E highlighted in red box. Catalytic site residues Lys118 and Tyr155 indicated with red triangles. Stereocontrol Trp site indicated with blue triangle. * indicates inactive domains. MgcQ lid residue alignment gap in light blue dashed box.

We initially used homology modeling to examine the NADPH binding site and other tertiary structure elements as potential sources of MgcQ KRc inactivity. A homology model of the MgcQ and

MgiQ KRc were generated based on the X-ray crystal structure of the active fungal amphotericin AmphB KRc, with bound NADPH (Figure 2.6). In both MgcQ and MgiQ, models revealed that the G7E

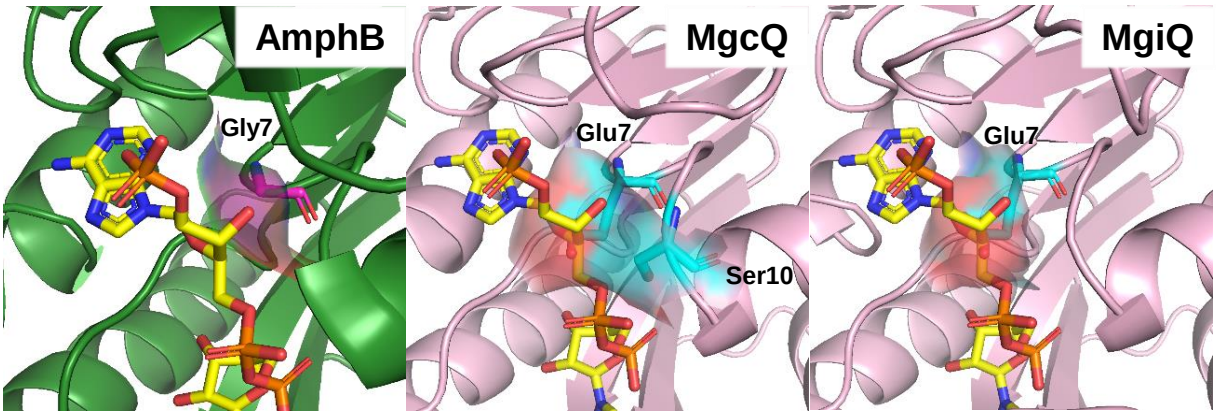


Figure 2.6. Homology modeling of MgcQ and MgiQ KRc. From left to right: AmphB KRc crystal structure (green cartoon, Gly7 backbone and transparent surface in magenta), MgcQ homology model (pink cartoon, Glu7/Ser10 rendered as sticks and transparent surface in cyan/red), and MgiQ homology model (pink cartoon, Glu7 rendered as sticks and transparent surface in cyan/red). NADPH rendered in stick form.

substitution of a polar glutamate residue within the NADPH binding site introduces considerable steric and electrostatic bulk, potentially hindering binding; Ser10 in MgcQ also appears to contribute minor steric bulk. Modeling also suggested tertiary structure changes to the positioning of the active site lid; however, the amino acid heterogeneity in lid sequence does not allow for reliable prediction of specific residues or motifs responsible for inactivity in this region.

To explore the hypothesized inactivity of MgcQ KR^0, we expressed and purified residues 1751-2679 of native MgcQ as a His-tagged MT-KR^0 didomain and assessed NADPH binding with fluorescence polarization.[181] Residues encompassing the catalytic KR domain (KRc) are numbered 1–245 in this study, corresponding to residues 2391–2635 of MgcQ and 662–906 of the recombinant MT-KR^0 didomain. To understand the role of the NADPH binding site in KR^0 inactivity, we additionally prepared two didomain constructs with mutations in the KR^0 NADPH binding sites: E7G, imparting a reversion to a canonical binding site in MgiQ and MgcQ KRc, and double mutant E7G S10G, with the Ser10 found only in MgcQ as comparison. Based on these unusual modifications to the P-loop, we tested our hypothesis that cofactor

binding to MgcQ KR is impaired by directly measuring NADPH binding to intact didomain via fluorescence polarization. Results showed that compared to the CurJ MT-KR control didomain, NADPH showed no significant increase in polarization in the presence of increasing concentration of the native or altered PAB MgcQ MT-KR^0 didomains, indicating total lack of measurable NADPH binding to the KRc (Figure 2.7). Thus, we theorize that Glu7 and Ser10 are likely not the only modifications to MgcQ and

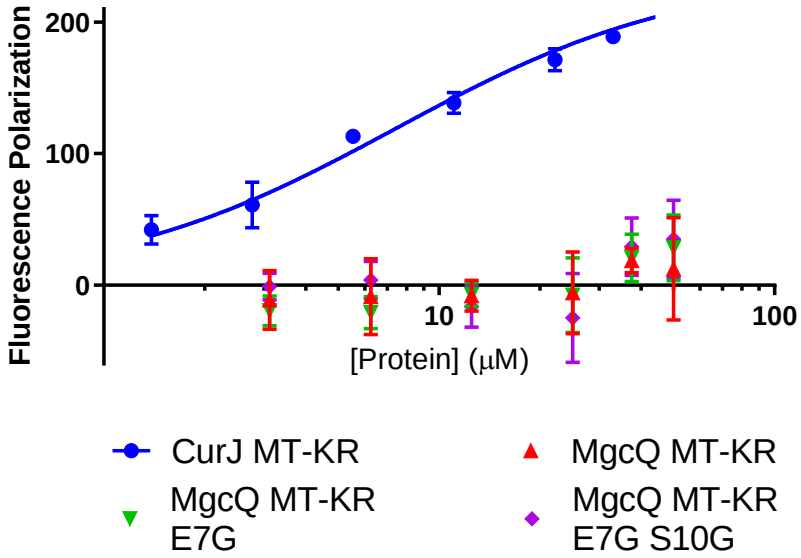


Figure 2.7. Fluorescence polarization when NADPH solution incubated with increasing concentrations of native MgcQ MT-KR^0, single mutant McgQ MT-KR^0 E7G, and double mutant MgcQ MT-KR^0 E7G S10G compared to CurJ MT-KR positive control.

MgiQ which confer inactivity: this likely results from uncharacterized changes in tertiary domain structure as well. Overall, evidence strongly suggests NADPH cannot bind to MgcQ, rendering the KR^0 redox incompetent.

2.4 Discussion

From two strains of the tropical marine cyanobacterium *Okeania hirsuta* (PAB and PAP), a highly similar pair of hybrid PKS/NRPS biosynthetic pathways were discovered that are predicted to encode for malyngamide C acetate and malyngamide I. Biosynthesis is initiated through a LipM octanoyltransferase from lipoic acid synthesis, then extended with PKS/NRPS modules, including a chlorinated β-branching methylation. In the penultimate PKS module, an unexpected KR^0 generates a

carbonyl which later contributes to carbocyclization via attack on a terminal aldehyde generated by an R-domain in the final PKS module. P450 and acetyltransfer reactions complete the biosynthesis.

Examining the 28 type A malyngamides characterized to date, we have assigned each molecule to one of four pathway groups, A1–A4 (Figure 2.8), with numbering of the 6-membered ring using equivalent positions to figure 1. Groupings are assigned from the predicted methylation, reductive domain activity, and chain release mechanism of the predicted final two modules, which define the PKS-derived architecture of the head group. To summarize, group A1 molecules have no methylations on the head group and a saturated C-7 position. Group A2 molecules are methylated at C-6 and feature either a hydroxyl, acetyl, or unsaturated C-7 position. Group A3 malyngamides are methylated at C-9, which offers the possibility of an intriguing potential alternate cyclization strategy (Figure A2.8), while group A4 molecules are methylated at C-8 only, and not at C-9, implying retention of Knoevenagel cyclization, but with C-MT activity on the head group. Each pathway group produces a base malyngamide species: A1; K, A2; L, A3; G, A4; E, upon which head group tailoring proceeds. Some analogs, such as malyngamide C and malyngamide N, may in fact be shunt products from malyngamide C acetate or malyngamide I, or whose presence is induced by environmental or stability factors, or a combination of both.

Most of the components of malyngamide biosynthesis, namely short chain fatty acid and amino acid activation and incorporation, polyketide assembly, chlorination, P450 activity, β-branching, *O*-methylation and acetylation, and glycosylation, are well-understood tailoring reactions in PKS and NRPS pathways. Thus, it is surprising that the "keystone" of type A malyngamide biosynthesis and radiative chemogenomics in *Okeania* is not a biochemical transformation at all, but rather the lack thereof. The altered MgcQ KR^0 and R domains are required to enable the critical Knoevenagel cyclization. In this regard, the MgcQ module and domains appear to be in a state of catalytic disrepair, yet with retention of genetic information and structural integrity. Though MgcQ by sequence contains KS-AT-DH-MT-KR-ACP, the DH, MT, and KR are catalytically inactive in PAB, while only the KR is catalytically inactive

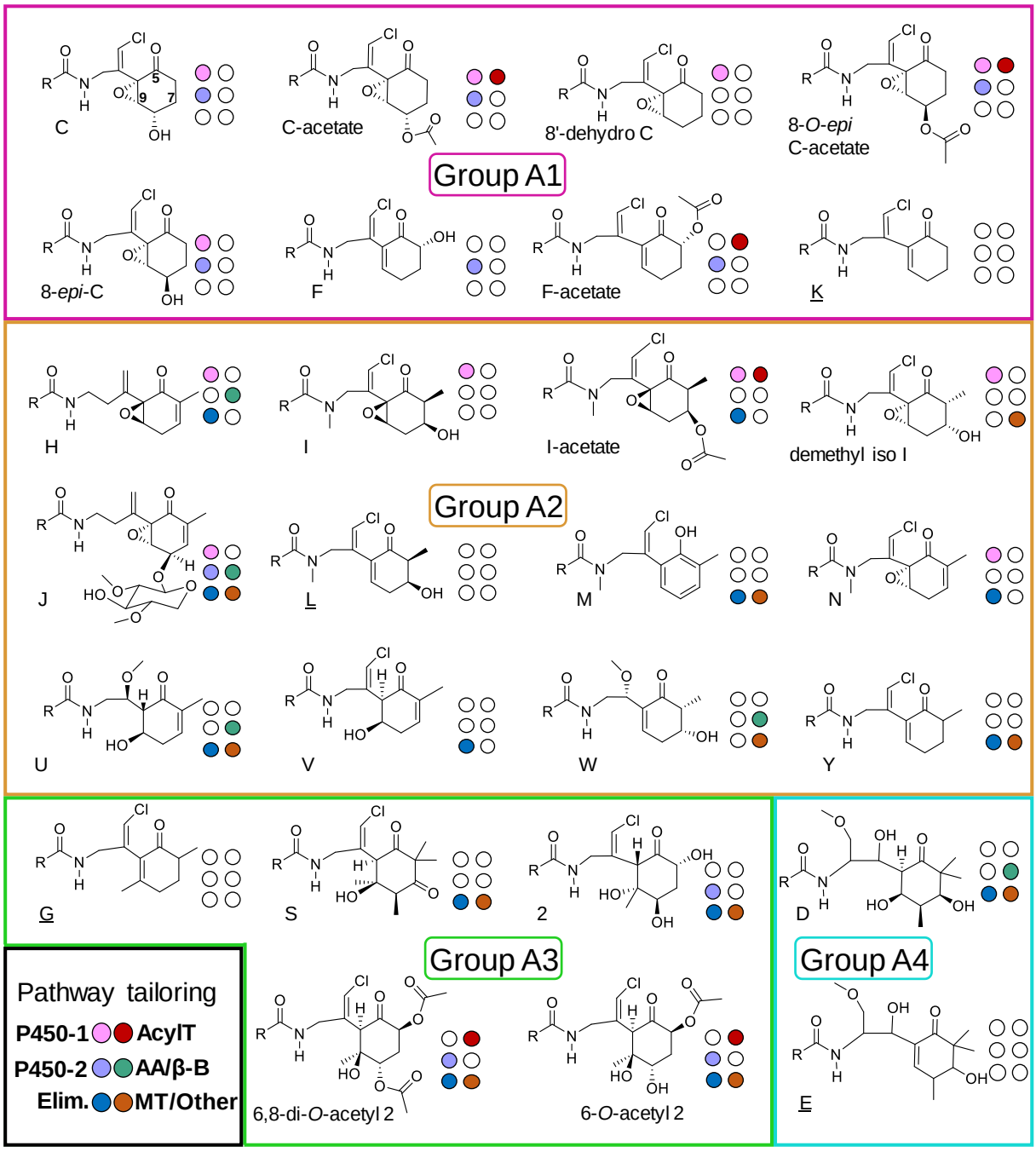


Figure 2.8. Biosynthetic model for type A malyngamides: Group A1 in pink, Group A2 in orange, Group A3 in light green, and group A4 in aqua. Untailored "base" chemical species are noted in each category by six uncolored circles and underline. Predicted tailoring differences from base species due to pathway diversity are indicated by coloration of the six circles next to each head group, with key to tailoring predictions in lower left. Differences to "base" group molecule are organized into six main categories: P450-1; P450 equivalent to MgcT (epoxidation), P450-2; P450 equivalent to MgcU (hydroxylation), Elim; pattern of post-cyclization hydroxyl elimination, AcylT; acetyltransferase activity, AA/β-B; amino acid incorporation or β-branching/halogenation cassette, and MT/Other; methyltransferase, cyclization, or other transformations.

in the PAP strain. Because inactive domains are normally excised over evolutionary time as they do not contribution to fitness, it is curious that *Okeania* has retained these partially non-functional MgcQ and MgiQ modules. As all type A malyngamides possess cyclohexyl ring structures, there appears to be considerable selective pressure to maintain this moiety, as its presence requires both a partially non-functional MgcQ KR^0 and a functional MgcR R. Thus, we speculate that excision of individual domains from MgcQ may render the module structurally and functionally defunct and unable to interact with the downstream MgcR. Further, since biosynthesis groups A1-A4 varyingly contain such diversity in tailored motifs, it would appear that the KR^0 has persisted through evolutionary time. Indeed, the global distribution of *Okeania*, widespread isolation of malyngamides, and shared bioactivity among numerous of analogs speaks to the KR^0 representing an ancient and important dysfunctionality that gave rise to this intriguing family of molecules.

2.5 Experimental methods

2.5.1 Collection, culturing and gDNA extraction

Okeania hirsuta PAB10Feb10-1 was collected by hand in 1–2 m depth subtidal mangrove benthic habitat in Bocas del Toro National Park,[157] Panama, coordinates 9°17'24.7"N 82°11'23.2"W. *O. hirsuta* PAP21Jun06-1 was collected by hand at depth of 1–2 m on mixed substrate in Portobelo National Park, Panama, coordinates 9°32'27"N, 79°40'33"W. Both under the auspices of the Panama International Cooperative Biodiversity Groups program. Cultures were transported to laboratory and propagated at 27-28°C in a 16/8 light/dark cycle under 5.5-10 μmol photons $m_{-2}S_{-1}$ in seawater BG-11 (SW BG-11) containing a final concentration of 1.5 g L^{-1} NO_3, with the exception of N_2 fixation experiments (see below section).[182,183] An approximate volume of 0.5 mL of packed wet biomass for each organism was harvested for gDNA extraction and washed with DI water over a paper filter and Büchner funnel to remove salt. Biomass was immediately frozen in liquid nitrogen and rapidly pulverized to a fine powder in a small mortar and pestle and transferred with a cold plastic scoop to a 2.0 mL Eppendorf tube. Frozen biomass powder was immediately extracted using the bacterial gDNA isolation protocol (Qiagen) and cleaned with G20 genomic tip (Qiagen) purification per manufacturer's instructions.

2.5.2 Assembly, bioinformatics, phylogeny, and modeling

PAB gDNA was submitted for MiSeq 300 bp paired-end sequencing (Illumina), and a second batch to Pacific Biosciences long-read sequencing. Reads were assembled using SPAdes 3.6 hybrid assembly.[184] PAP gDNA was submitted for HiSeq (Illumina), and assembled using SPAdes 3.6 assembly. DarkHorse (v 1.0) was used to remove contigs containing no cyanobacterial genes.[185] The PAB assembly contained a smaller number of contigs of generally larger size than the binned PAP draft, thus PAP contigs were scaffolded using PAB as a reference via Medusa scaffolder 1.0 (http://combo.dbe.unifi.it/medusa). Both draft genomes harbored putative malyngamides pathways and all housekeeping genes previously reported from complete cyanobacterial genomes.[186,187] Contigs were submitted to AntiSMASH 4.0 and NCBI DELTA-BLAST and analyzed for domain composition.[128,132] MgcA and MgcS protein sequences were submitted to Phyre2 protein structure prediction server.[188]

Sequences were aligned with MUSCLE algorithm using default settings in Geneious (v. 8.1.9, Biomatters) and Jalview (v. 2.10.5).[189,190] The LipM phylogenetic tree was calculated using Geneious Tree Builder by Unweighted Pair Group Method with Arithmetic Mean, using the Jukes-Cantor genetic distance model, re-sampled with 2,000 bootstrap iterations, and a topology threshold of 70%. KR domain homology models were generated using SWISS-MODEL using PDB entry 3MJS monomer as a model for PAB MgcQ KR, with all entries trimmed to the KRc portion of the MgcQ KR domain.[191] Modeling images were made using Pymol (v. 2.2.0, Schrodinger).[192]

2.5.3 Phylogenomics

Phylogenomic analysis was performed using 29 different conserved cyanobacterial genes as assessed previously,[186] including genes from 110 complete cyanobacterial genomes currently in the Joint Genomes Institute (JGI IMG) database,[187] in addition to the two draft genomes for *Okeania hirsuta* reported in this study. The genes were concatenated, then aligned via MUSCLE 3.8.31 and trimmed via TrimAl 1.2 (standard settings). The tree was built by the tool Geneious Tree Builder, Geneious® 8.1.2, (BioMatters) using the Jukes-Cantor genetic building model, and the Neighbor-Joining method, with 1,000 bootstrap repetitions and one outgroup genome Melainabacteria *Vampirovibrio chlorellavorus* SM1D11, the furthest cyanobacterial organism from PAP/PAB.

2.5.4 Biomass chemistry extraction and malyngamide isolation

Samples for chemical extraction were harvested over a Büchner funnel with paper filter and rinsed with DI water to remove salt. Biomass was subjected to 3–4 iterations of 2:1 DCM:MeOH sequential extraction and dried. Crude extract was reconstituted in acetonitrile and filtered through a C18 cartridge (Agilent), eluted with ten column volumes of 100% acetonitrile or methanol. Samples were then applied to semipreparative HPLC at 3 mL min^{-1} through a 150 mm X 10 mm, 5 μm Kinetex C18 column on a 65% to 99% (v/v) acetonitrile/water gradient over 12 minutes. Pure compounds were collected by monitoring UV peaks at 209, 228 and 234 nm. 0.8 mg of unlabeled malyngamide C acetate from PAB, and 1.9 mg of unlabeled malyngamide I from PAP were isolated for structural characterization.

2.5.5 Stable isotope feeding studies

PAB was chosen for feeding experiments due to its slightly faster growth rate (laboratory observations). PAB was scaled up for 8-12 weeks from a finger-sized clump of biomass in 30 mL of media, to two 2.8 L Fernbach flasks with 1 L of media each. Samples were concentrated to 0.75 L of media in a single 2.8 L Fernbach flask for feeding experiments. A total of 47.0 mg [1,2-$^{13}C_2$]acetate plus 46.0 mg of unlabeled acetate were dissolved in 5 mL sterile water and added in three equimolar batches over 10 days to a final concentration of 0.75 mM of labeled substrate, followed by harvest after 15 total days of growth with labeled substrate. 28 mg of [1,2-$^{13}C_2$]glycine, for a final concentration of 0.5 mM, was added to a separate culture and harvested in similar fashion. 62.0 mg of [1-^{13}C]octanoate, to a final concentration of 0.5 mM, was added to a separate culture and harvested in similar fashion. All labeled substrates were purchased from Cambridge Isotopes. Pure labeled malyngamide C acetate was recovered as follows: 0.2 mg from the [1-^{13}C]octanoate-fed batch, 0.1 mg from the [1,2-$^{13}C_2$]acetate-fed batch, and 0.2 mg from the [1,2-$^{13}C_2$]glycine-fed batch.

2.5.6 Ancymidol feeding

Nine replicates of PAB were inoculated from 10-20 0.5 cm filaments into 30 mL of SW BG-11 with final concentration of 1.5 gL^{-1} NO_3. Three replicates were grown in standard media, three were dosed ancymidol to a final concentration of 26 μM, and three were dosed ancymidol to a final concentration of 260 μM. Flasks were incubated for 14 days, then harvested, extracted, and submitted to LC/MS/MS as described above. MS spectra were converted to .mzxml using MSConvert (ProteoWizard), and submitted to Global Natural Product Social network (GNPS, https://gnps.ucsd.edu/) using default parameters.[133,193] Results were analyzed in Cytoscape (v. 3.6.1).[194] Mass chromatogram signals for each malyngamide shunt product identified were averaged across biological triplicates and normalized to the total pool of malyngamides and malyngamide shunt products identified in the extract.

2.5.7 Mass spectrometry, polarimetry, NMR and enrichment analysis

HPLC-MS/MS of culture extracts and isolation/purification steps was performed on an LCQ Advantage Max ion trap MS with Finnigan Surveyor HPLC system (Thermo Fisher Scientific). Ancymidol-fed culture extracts were analyzed on an Agilent 1200-series HPLC/6530 qToF (Agilent). Both systems employed data dependent acquisition in positive mode, using a 4.6 mm x 100 mm Kinetex 5 μm C18 column for separations on a gradient of 50:50 water:acetonitrile 100% acetonitrile using a 3:1 split flow rate of 0.6 mL min^{-1} (0.15 mL min^{-1} to the MS) on the LCQ system and 0.7 mL min^{-1} on the Agilent system. All solvents contained 0.1% (v/v) formic acid. High resolution MS data was acquired by direct infusion on an LTQ Orbitrap XL MS with Orbitrap Fourier Transform mass analyzer (Thermo Fisher Scientific).

Optical rotation was obtained on a Jasco P-2000 polarimeter, using a 1-dm chamber. Compounds were dissolved in methanol at 1 mg/mL and 160 μL were loaded into the chamber. An average of ten readings was taken for each compound and a solvent blank was subtracted from each.

NMR analysis of malyngamide C acetate for both structural confirmation and labeling experiments was carried out on a 500 MHz magnet VX system (Varian) in $CDCl_3$ at 125 Hz for CNMR and 499.825 MHz for HNMR; shifts were referenced to the center $CHCl_3$ peak at 77.16 ppm for CNMR and 7.24 ppm for HNMR. 2D Structure confirmation for malyngamide I was performed on a 500 MHz JEOL ECZ500 system (JEOL) in $CDCl_3$ and DMSO-d_6, at room temperature. NMR of all enriched species were performed in Shigemi tubes (Shigemi). Enrichment analysis of [1-^{13}C]octanoate, as well as J_{CC}-values, were calculated as previously described.[89,161] Flanking acetate enrichment percentage was derived by calculating the ratio of the sum of the intensity of coupled ^{13}C acetate peaks to the intensity of the central ^{13}C peak for each resonance. Overlapping signals were not analyzed. JEOL data was processed using Delta software (JEOL), and Varian data was processed in Mestrenova (Mestrelab).

2.5.8 Recombinant protein expression

Generally, KR domains adjacent to C-MT domains require expression of both the KR, C-MT, and a short β-strand prior to the C-MT in order to be stable.[93] An amplicon including the PAB MgcQ β-ribbon, MT, KRc, and a portion of the module ACP were amplified from genomic DNA by PCR using Phusion polymerase (New England Biolabs), and cloned into a pET-28b vector using PCR-overlap extension Gibson assembly to make vector B050V. The region of *mgcQ* encoding the MT-KR didomain was amplified from the B050V template using the KOD HotStart Polymerase (Sigma-Aldrich) and inserted into pMCSG7 by Ligation Independent Cloning (LIC). Mutants were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent). The sequence coding for the CurJ MT-KR didomain was cloned into pMCSG7 from the pLM9 cosmid.[89] Plasmids were transformed into *E. coli* XL1-Blue cells for plasmid extraction while selecting with 100 μg mL^{-1} ampicillin.

The *E. coli* pGro7 chaperone-producing strain (Takara) was used to produce MgcQ MT-KR, while mutants and the CurJ MT-KR were generated in *E. coli* pRare2-CDF cells.[112] 500 mL TB expression cultures were grown at 37 °C to an OD_{600} of 1.0. Cultures of *E. coli* pGro7 cells were supplemented with 100 μg mL^{-1} ampicillin and 35 μg mL^{-1} chloramphenicol, while cultures of *E. coli* pRare cells contained 100 μg mL^{-1} ampicillin and 50 μg mL^{-1} spectinomycin. After cooling to 20 °C, the cultures were induced with 50 μM IPTG and 1 g of L-arabinose. Cultures were incubated at 20 °C for 18–20 hours with shaking at 225 rpm, then harvested.

Cells were resuspended in lysis buffer (for MgcQ MT-KR and its mutants): 50 mM HEPES pH 7.5, 300 mM NaCl. 10% (v/v) glycerol/water, 20 mM imidazole; for CurJ MT-KR: 50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol/water, 20 mM imidazole) with 5 mg DNase, 2 mg lysozyme, and 1 mM $MgCl_2$. Complete lysis was achieved with sonication.

The soluble fraction was collected by centrifugation and filtered before loading onto a 5 mL HisTrap HP (GE Healthcare) column with Ni-NTA buffer (for MgcQ MT-KR and its mutants: 50 mM HEPES pH 7.5, 300 mM NaCl. 10% (v/v) glycerol/water, 20 mM imidazole; for CurJ MT-KR: 50 mM

Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol/water, 20 mM imidazole) at a flow rate of 2.5 mL min^{-1}, followed 5 column volumes of Ni-NTA buffer. Proteins were eluted with a 20 mM to 655 mM imidazole gradient over 12 min.

The eluate was assessed by SDS-PAGE and fractions containing the protein of interest were concentrated to 5 mL. MgcQ MT-KR produced in *E. coli* pGro7 cells was incubated on ice with 5 mM ATP for 30 min to stimulate chaperone dissociation. The protein solutions were then loaded onto a Hiload 16/60 Superdex 200 prep grade gel-filtration column (GE Healthcare) pre-equilibrated with size-exclusion buffer (for MgcQ MT-KR and its mutants: 50 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol/water; for CurJ MT-KR: 50 mM Tris pH 7.4, 150 mM NaCl, 10% (v/v) glycerol/water). MgcQ MT-KR and CurJ MT-KR didomains eluted with apparent molecular weights of 98 kDa and 94 kDa, respectively, compared to their true molecular weights of 105 kDa and 102 kDa. This suggests that both are monomeric in solution. Fractions containing the MT-KR didomain were pooled, flash frozen in liquid nitrogen, and stored at –80 °C.

2.5.9 Fluorescence polarization

Protein samples were diluted to 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.175 μM, and 0 μM in a 50 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol/water buffer with 25 μM NADPH. Fluorescence polarization (FP) data were recorded at 340/450 nm on a Spectramax M5 spectrophotometer (Molecular Devices) and analyzed with software from the vendor. K_D values were calculated using Prism (GraphPad).

2.5.10 Cyanobacterial culturing for nitrogen fixation

Six biological replicates of PAB and PAP each were scaled up from mechanically separated single filaments to approximately 0.5 g of wet packed biomass in 300 mL of media in a sterile tissue culture flask with vented caps over a period of 10 weeks. Three replicates of each species were grown in saltwater BG-11 media containing 1.5 g L^{-1} NO_3, and three were grown in saltwater BG-11 media containing no added exogenous nitrate. For initial inoculation, single filaments were washed with nitrate-free media twice before adding to media wells.

2.5.11 Nitrogen fixation measurements

N_2-fixation rate was determined using the Acetylene Reduction Assay by gas chromatography with a Shimadzu gas chromatograph GC-8a (Shimadzu Scientific Instruments).[195] 20 mL of culture containing the cyanobacterial biomass from each replicate bottle was added to an airtight 40 mL serum vial. For each vial, 2 mL of air was extracted with a syringe and 2 mL of acetylene (C_2H_2) was injected to the headspace. Samples were incubated in the dark at 25°C, for 14 hours. Afterward, ethylene production was measured by injecting 200 μL of culture headspace to the GC device. 200 μl of 100 ppm ethylene (base code: GMT10325TC, Matheson Gas Products) was used as a retention time standard. Relative N_2 fixation was measured by calculating the ppm of ethylene generated in 200 μL headspace injection divided by dry biomass.

2.6 Appendix

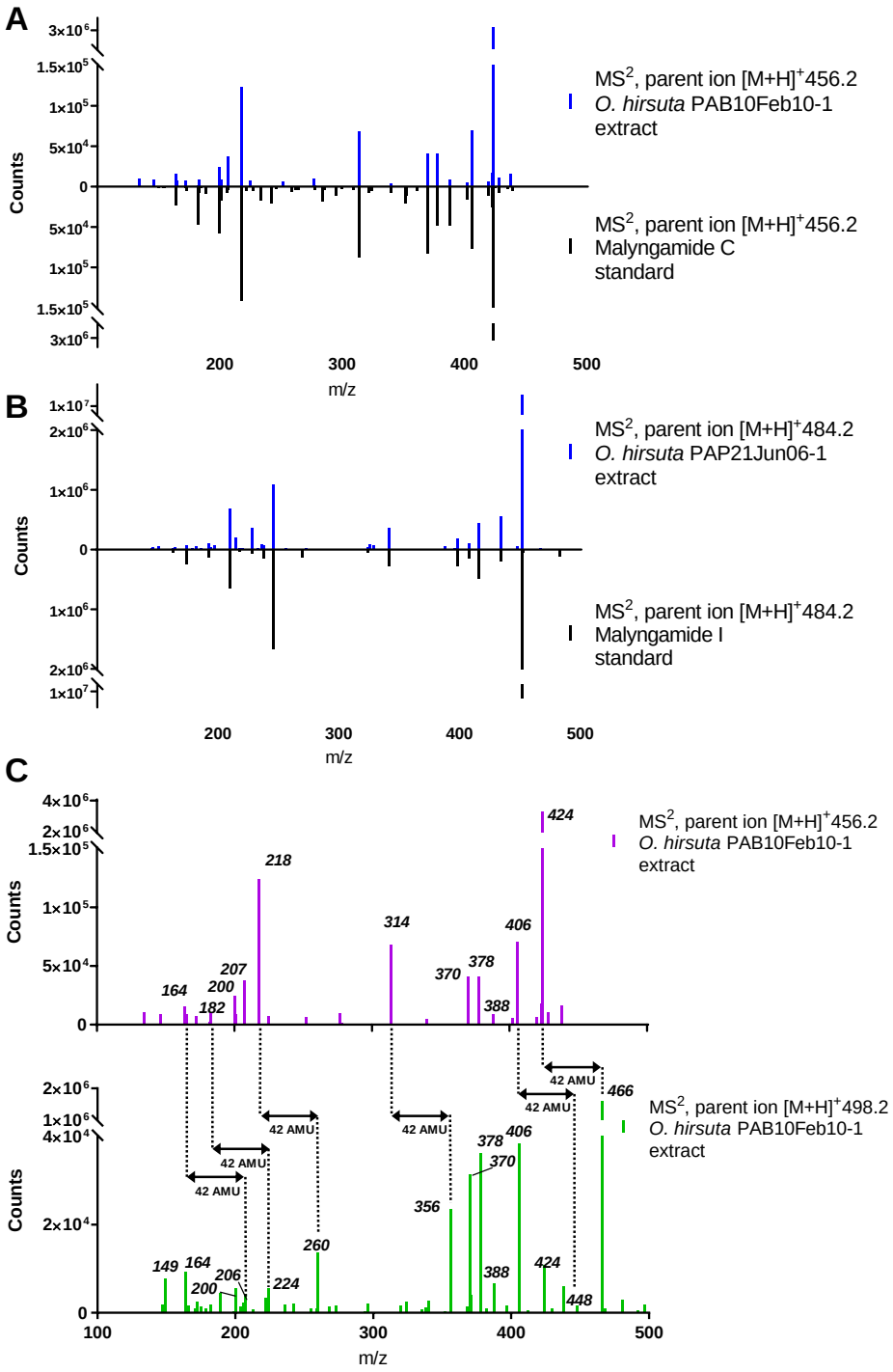


Figure A2.1. Mass spectrometry analysis of malyngamides. a) Mass spectrum comparison of MS^2 of a) pure compound standard of malyngamide C (black) compared with $[M+H]^+$ 456.2 in extract from *Okeania hirsuta* PAB10Feb10-1 (blue) and b) pure compound standard of malyngamide I (black) compared with $[M+H]^+$ 484.2 in extract from *Okeania hirsuta* PAB21Jun06-1 (blue). c) Mass spectrum comparison of MS2 of $[M+H]^+$ 456.2 (violet) and $[M+H]^+$ 498.2 (green) from extract of *Okeania hirsuta* PAB10Feb10-1.

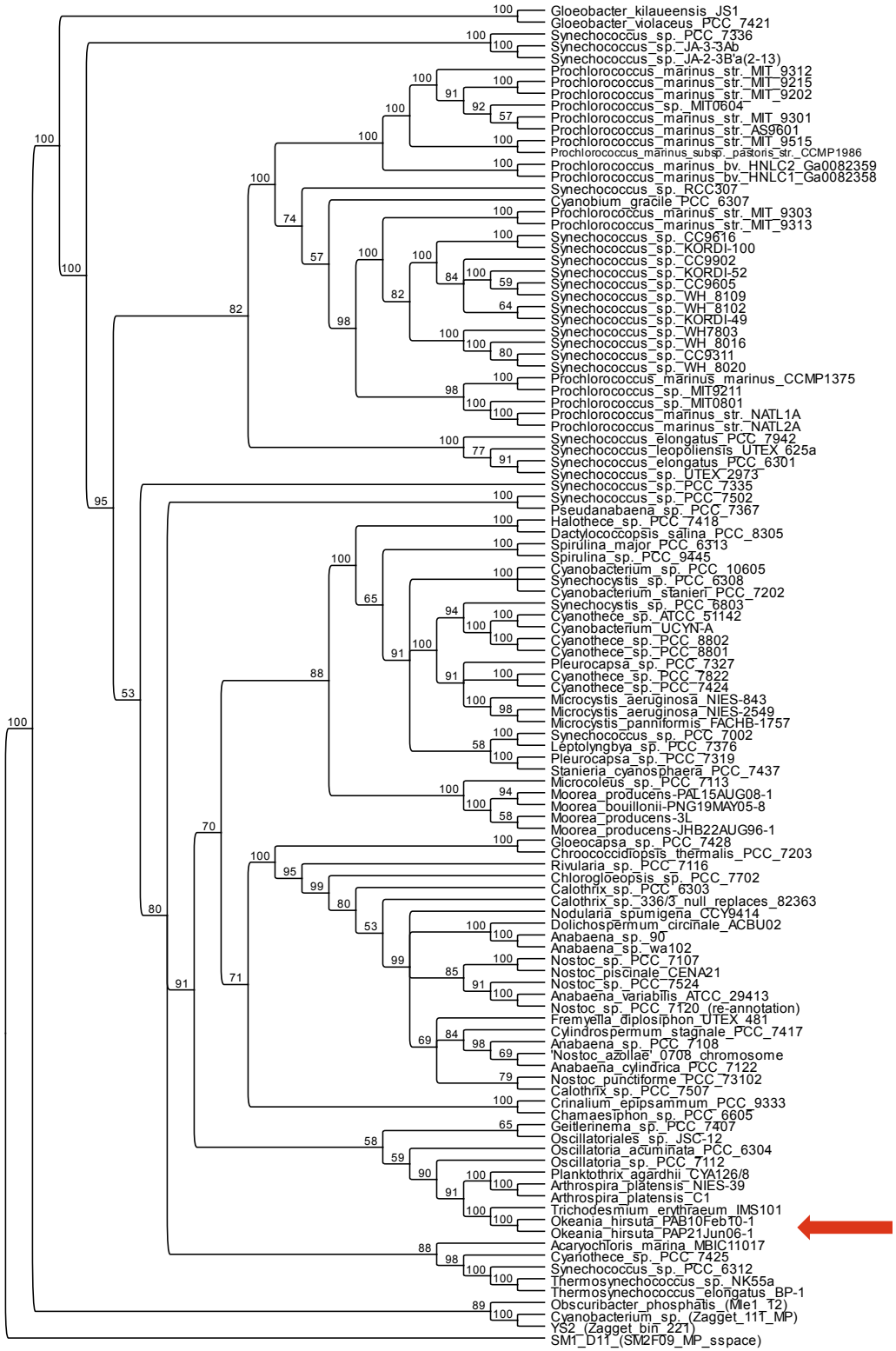


Figure A2.2. Phylogenomic tree of *Okeania hirsuta* PAB10Feb10-1 and *Okeania hirsuta* PAP21Jun06-1. Consensus support is displayed at node junctions. Tree generation detailed in experimental methods.

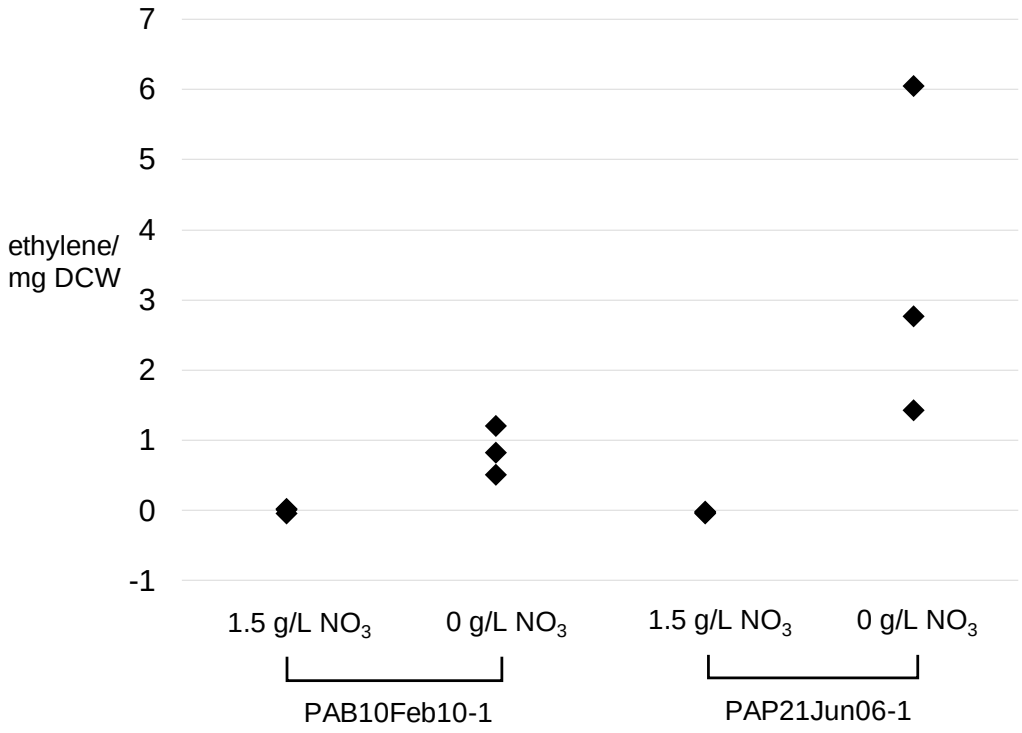


Figure A2.3. Acetylene reduction assay results. Each point represents a biological replicate. Ethylene signal is normalized to mg dried biomass.

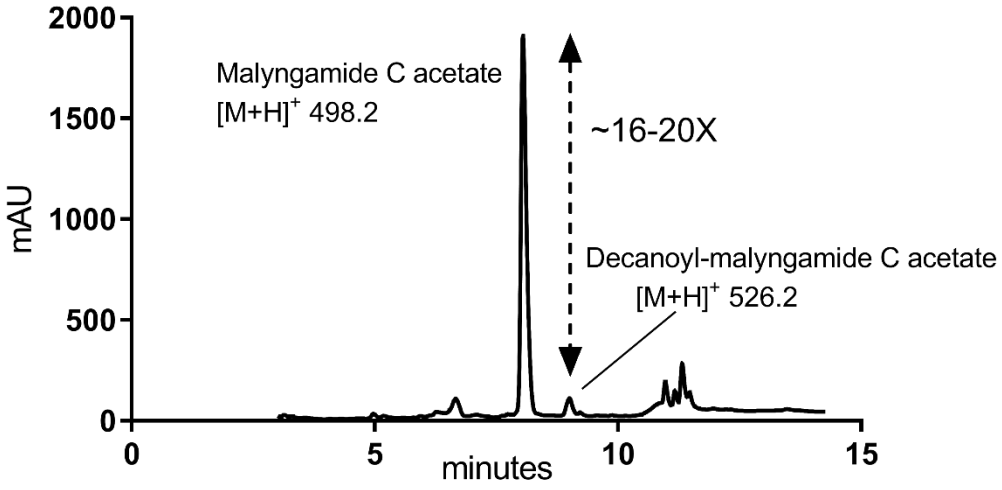


Figure A2.4. MS chromatogram comparison of malyngamide C acetate signal with $[M+H]^+$ 526, corresponding to AMU +28, or C16-lyngbic acid-malyngamide C acetate. Exact ratio varied with extract and culture conditions in the range of 16–20:1, *m/z* 498.2:*m/z* 526.2.

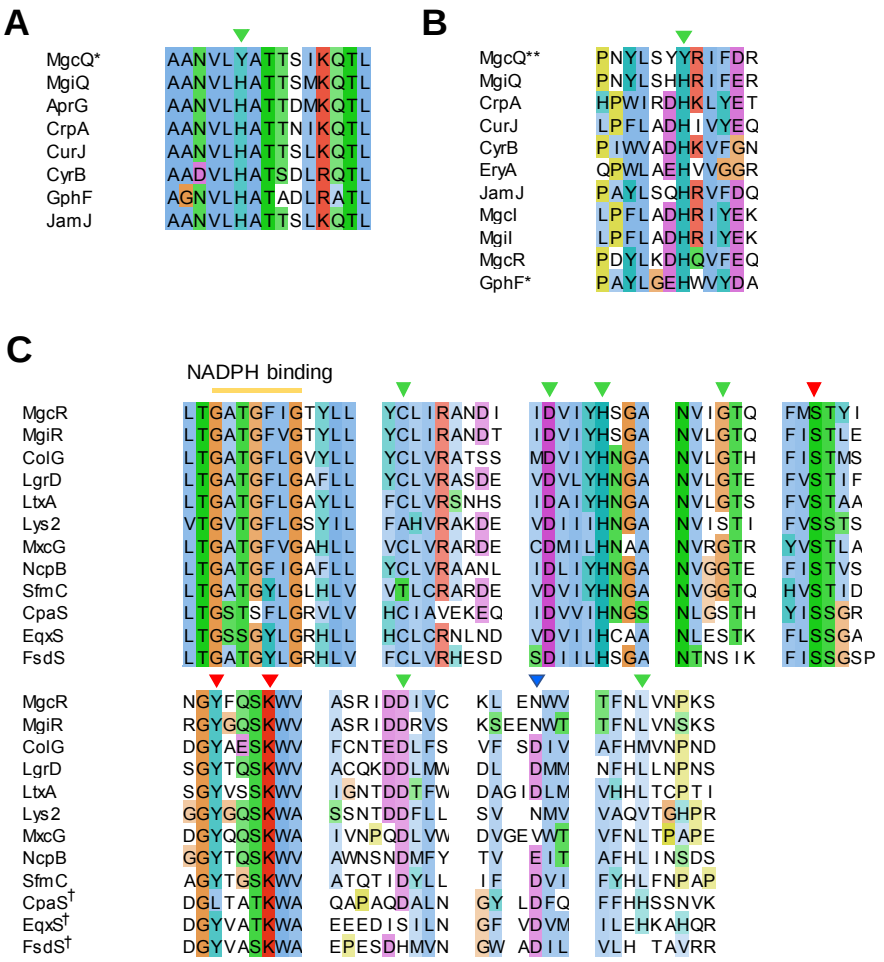


Figure A2.5. Alignment of key residues in MgcQ/MgiQ catalytic domains with comparable cyanobacterial and microbial domains. Double asterisk indicates predicted inactive made null by MgcQ KR^0. Single asterisk indicates inactive domain. a) C-MT - green triangles indicate active site residues b) DH - green triangle indicates active site residue c) R domains. Crosses indicate Dieckmann cyclization-only domains, red triangles indicate SDR Ser-Tyr-Lys catalytic triad, blue triangle indicates $R^{\dagger}$-critical Asp, green triangles indicate polar residues important in $R^{\dagger}$ domain activity.[170]

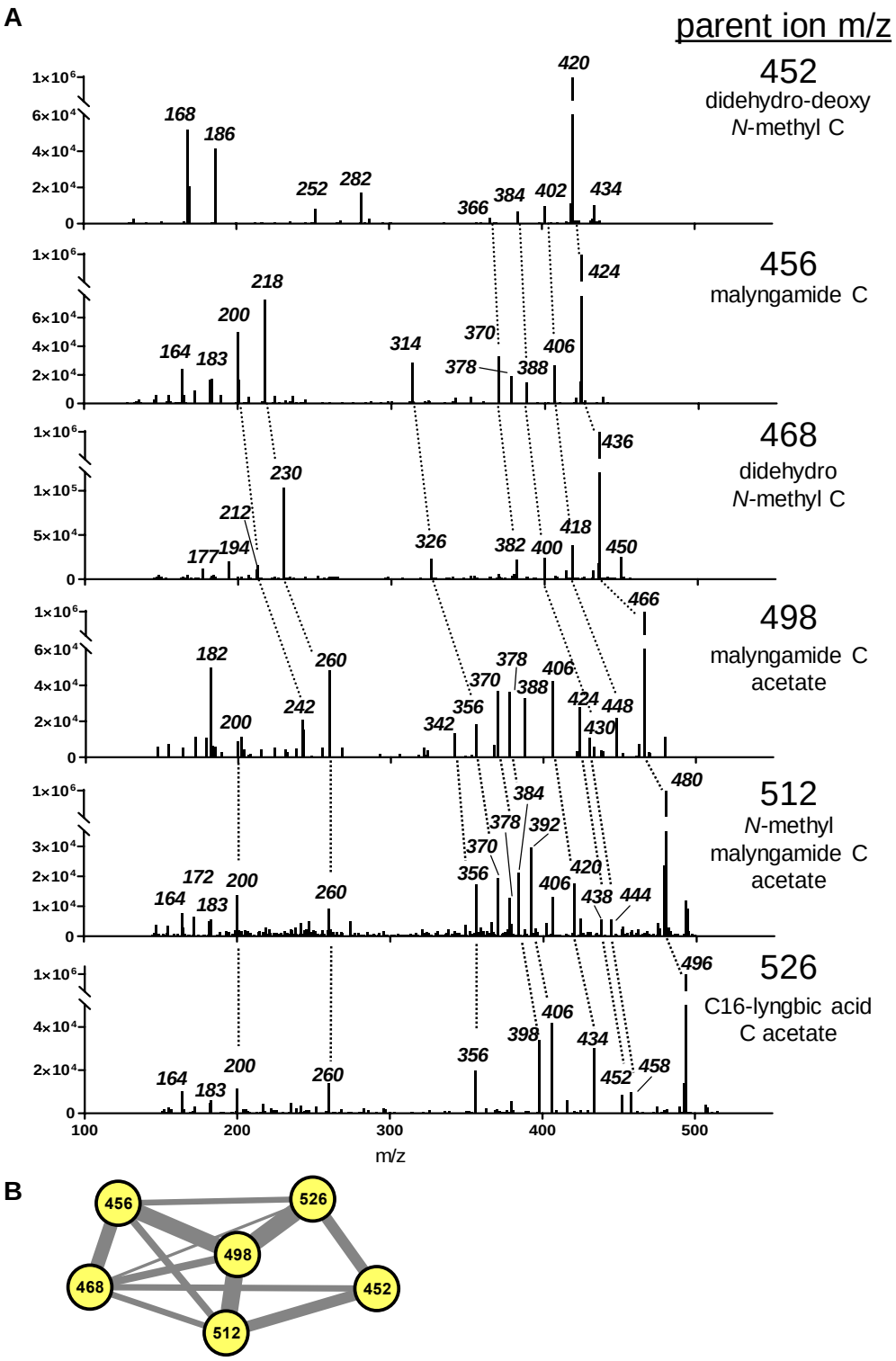


Figure A2.6. MS^2 fragmentation data of ancymidol derived shunt products. a) MS^2 spectra of parent ion $[M+H]^+$at right of each spectrum. Dashed lines connect fragment ions which differ by same mass difference between parent ion molecular weight of next-highest analog. b) Global Natural Product Spectral networking (GNPS)-derived network of MS^2 fragmentation patterns. Line thickness represents cosine score calculated between fragment ions, indicating similarity of fragment patterns.[196]

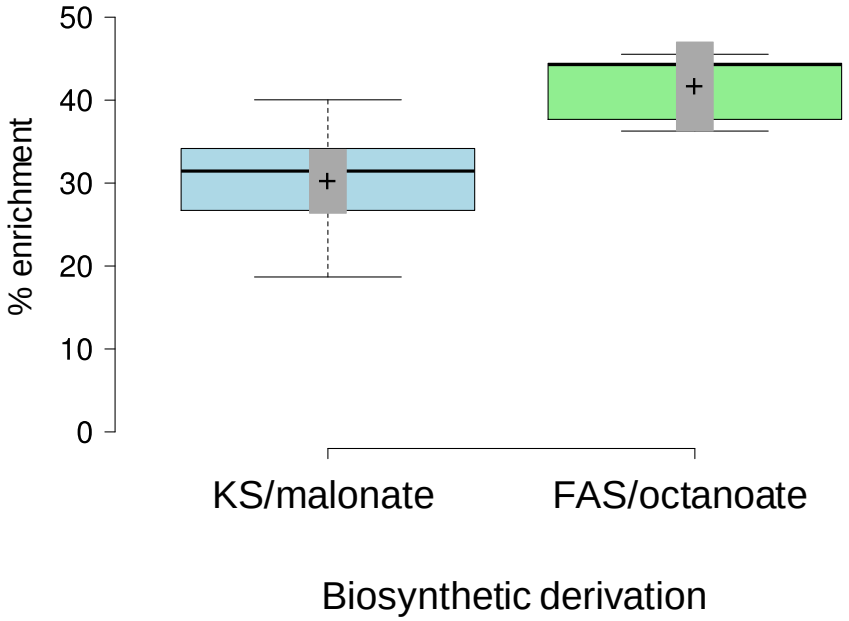


Figure A2.7. Variance of % enrichment of acetate-derived carbon resonances as indicated by summed coupled signal intensity divided by uncoupled signal intensity. KS/malonate values are derived from C-4–C-9, C-8', C-10', C-13' – C-14', and C-16' – C-17' (n = 12) flanking coupled acetate carbon resonances. FAS/octanoate values are derived from C-1' and C-3' – C-6', and C7' (n = 5). Relative signal heights were measured for both the coupled resonances and the uncoupled resonance for each acetate-derived carbon in order to evaluate levels of ^{13}C enrichment. One-way ANOVA indicates P-value <.001 between the two categories. Grey bars indicate 95% confidence intervals.

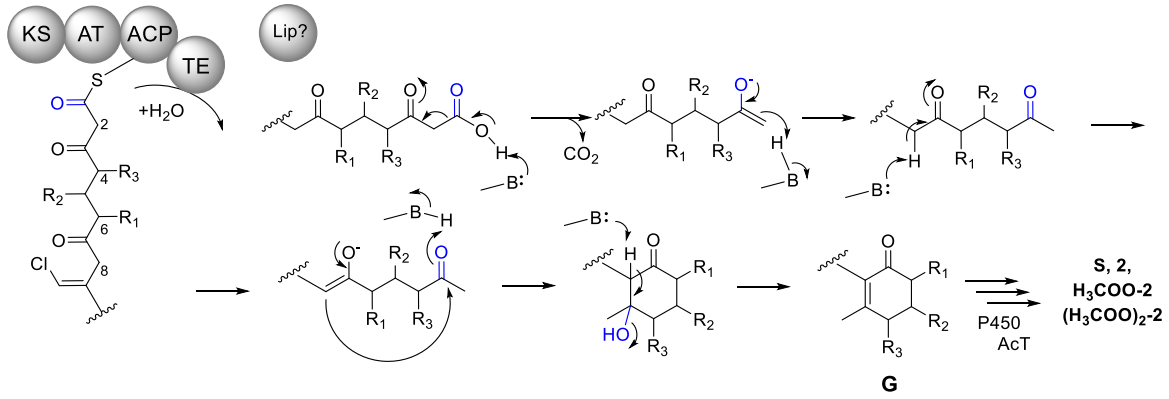


Figure A2.8. Theoretical cyclization mechanism for generation of group A3 malyngamides, displaying only final KS module. Hypothetical thioesterase (TE) domain is proposed to generate terminal carboxylic acid. Decarboxylation reaction would drive electrons to form enolate anion, which then abstracts a proton to form a ketone at C-3. At this point, the C-8 proton is abstracted, forming another enolate anion, which then pushes back to attack the newly generated C-3 ketone, forming a C-C bond and hydroxyl group at C-3. Elimination of water creates the A4 base molecule malyngamide G, which may then be processed by P450 tailoring enzymes and acetyltransferases to form other A3 malyngamides.

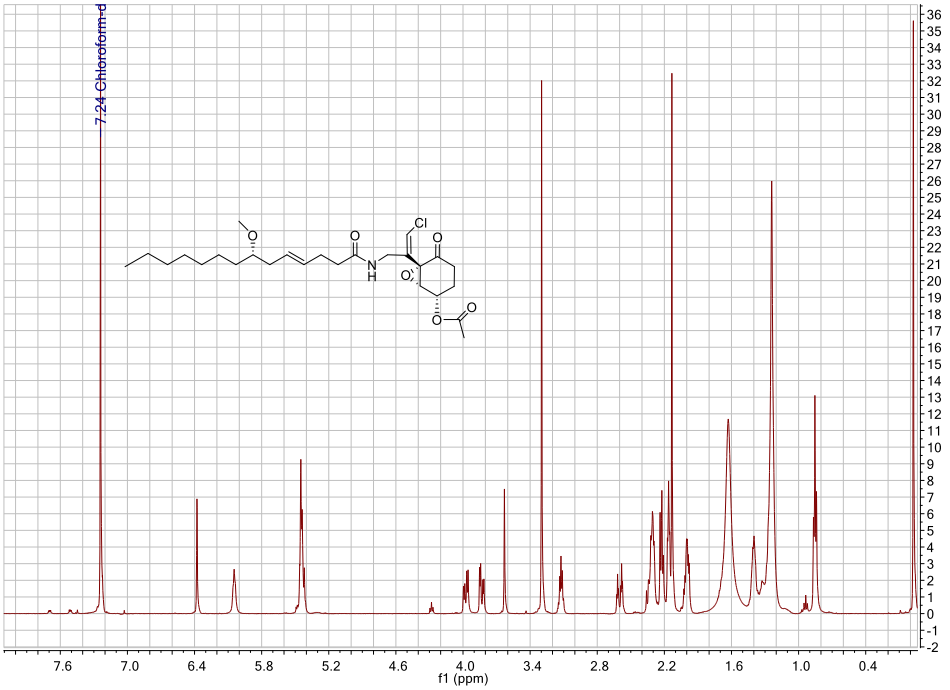


Figure A2.9. ^{1}H-NMR of malyngamide C acetate. (500 MHz, $CDCl_3$)

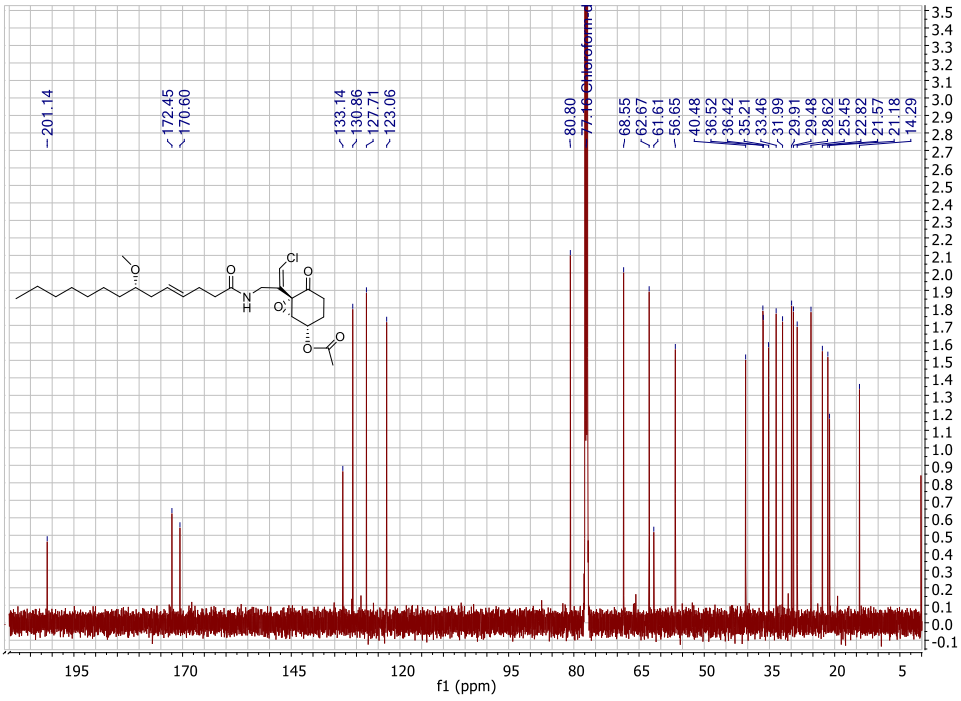


Figure A2.10. ^{13}C-NMR of malyngamide C acetate, unlabeled. (125 MHz, $CDCl_3$)

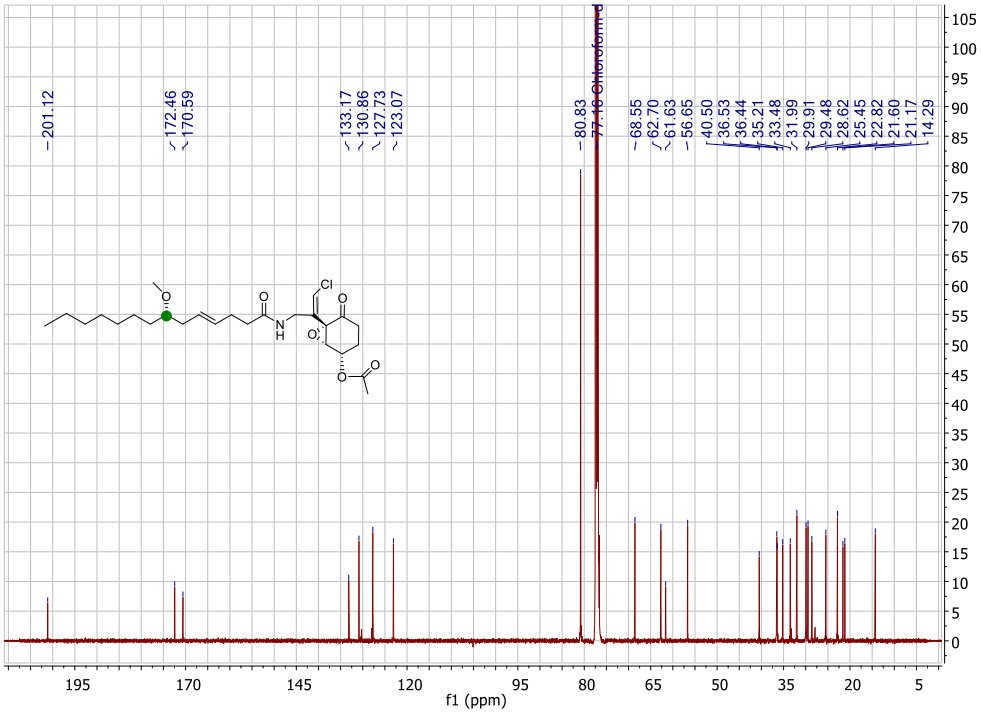


Figure A2.11. ^{13}C-NMR of malyngamide C acetate, [1-^{13}C]octanoate-fed sample. (125 MHz, $CDCl_3$). Green circle indicates ^{13}C label incorporation carbon.

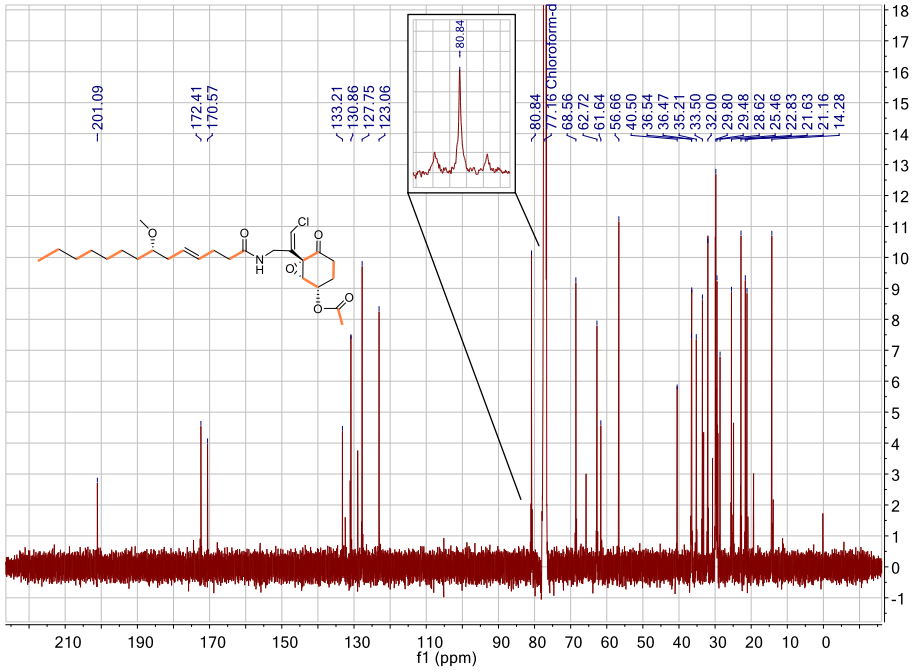


Figure A2.12. ^{13}C-NMR of malyngamide C acetate, [1,2-$^{13}C_2$]acetate-fed sample. (125 MHz, $CDCl_3$). Orange bold bonds indicate intact labeled acetate incorporation carbons. Inset spectrum indicates example of flanking ^{13}C-labeled resonances.

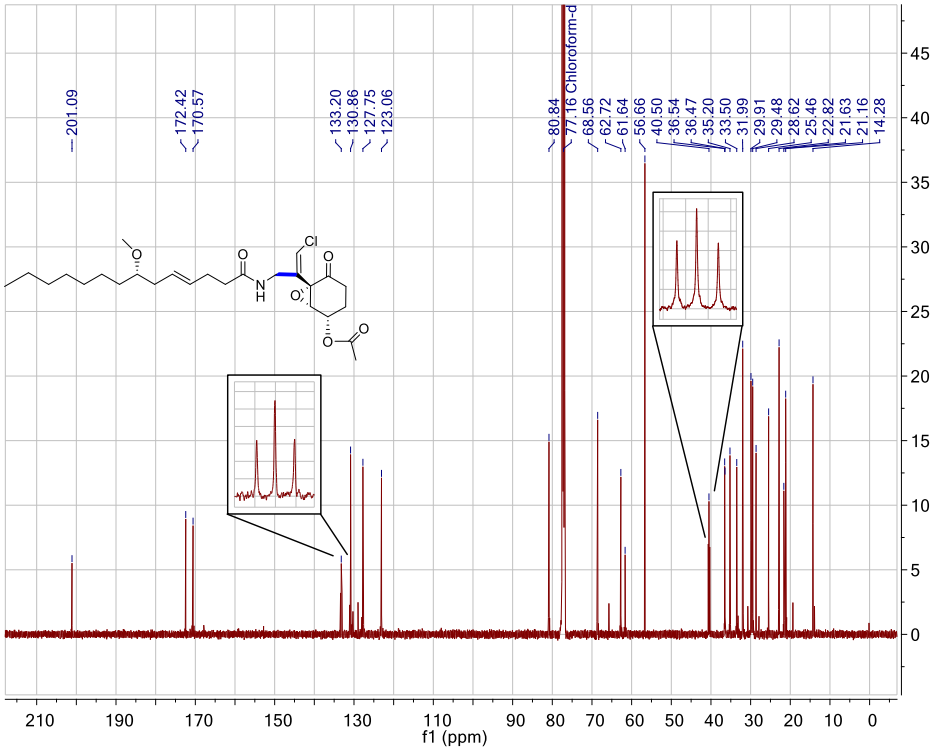


Figure A2.13. ^{13}C-NMR of malyngamide C acetate, [1,2-$^{13}C_2$]glycine-fed sample. (125 MHz, $CDCl_3$). Blue bold bond indicates intact labeled glycine incorporation carbons. Inset spectra indicate flanking ^{13}C-labeled resonances from glycine-derived carbons.

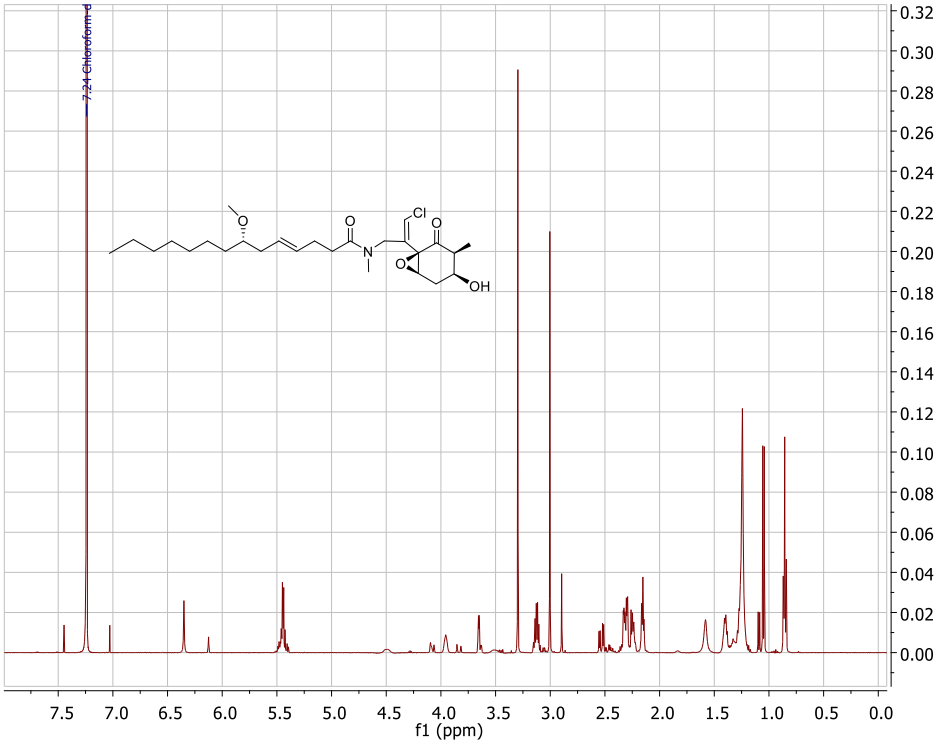


Figure A2.14. ^{1}H-NMR of malyngamide I. (500 MHz, $CDCl_3$).

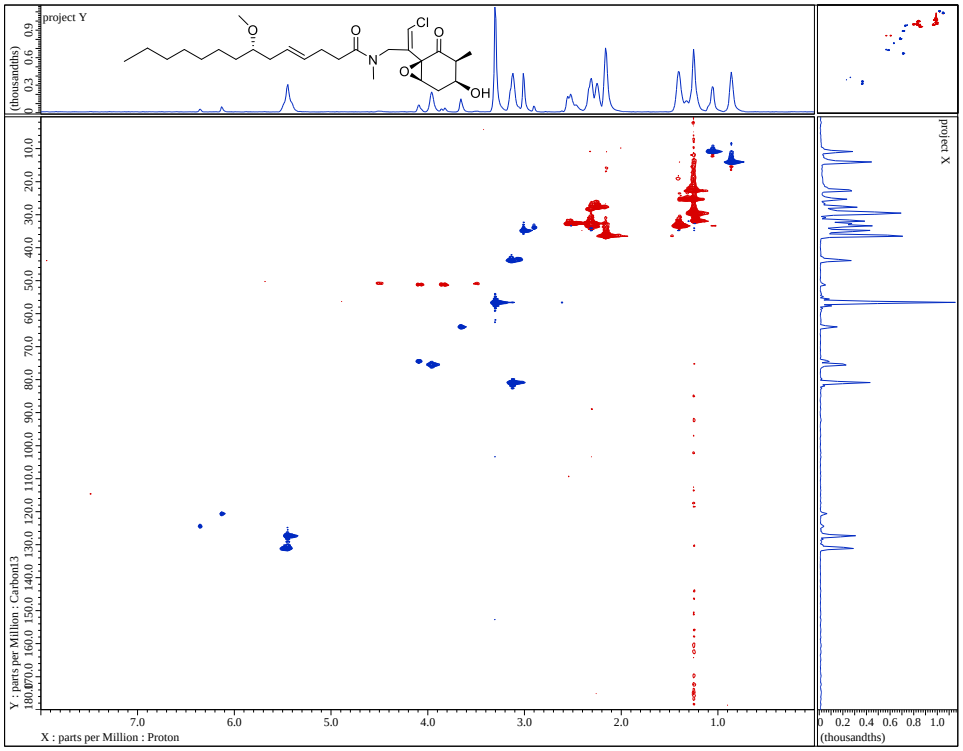


Figure A2.15. HSQC of malyngamide I. (500 MHz, $CDCl_3$).

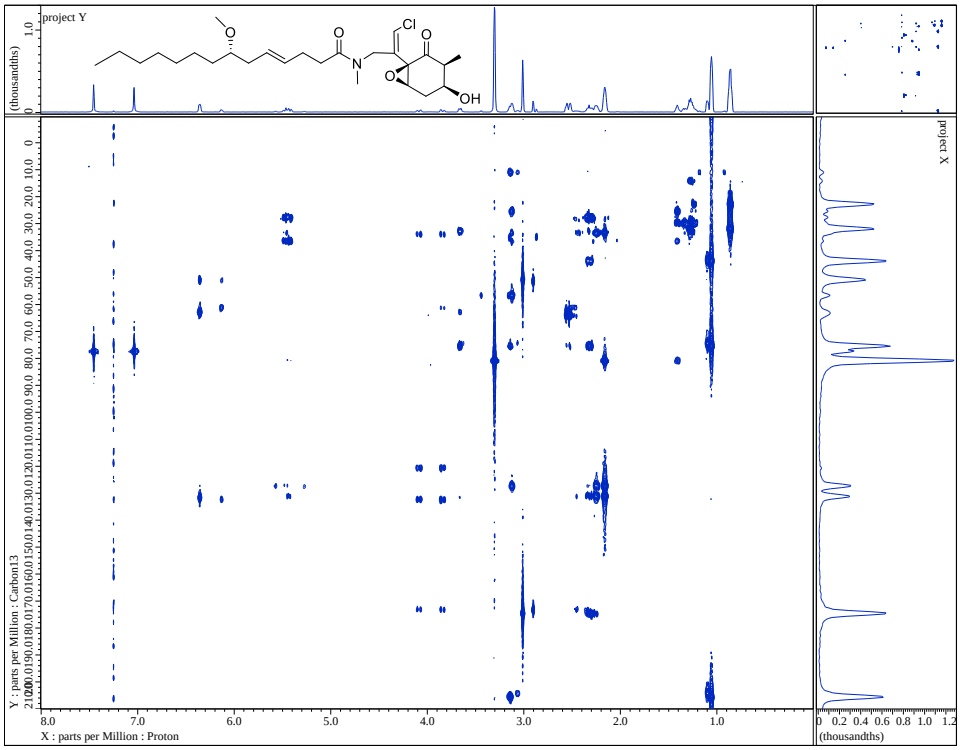


Figure A2.16. HMBC of malyngamide I. (500 MHz, $CDCl_3$).

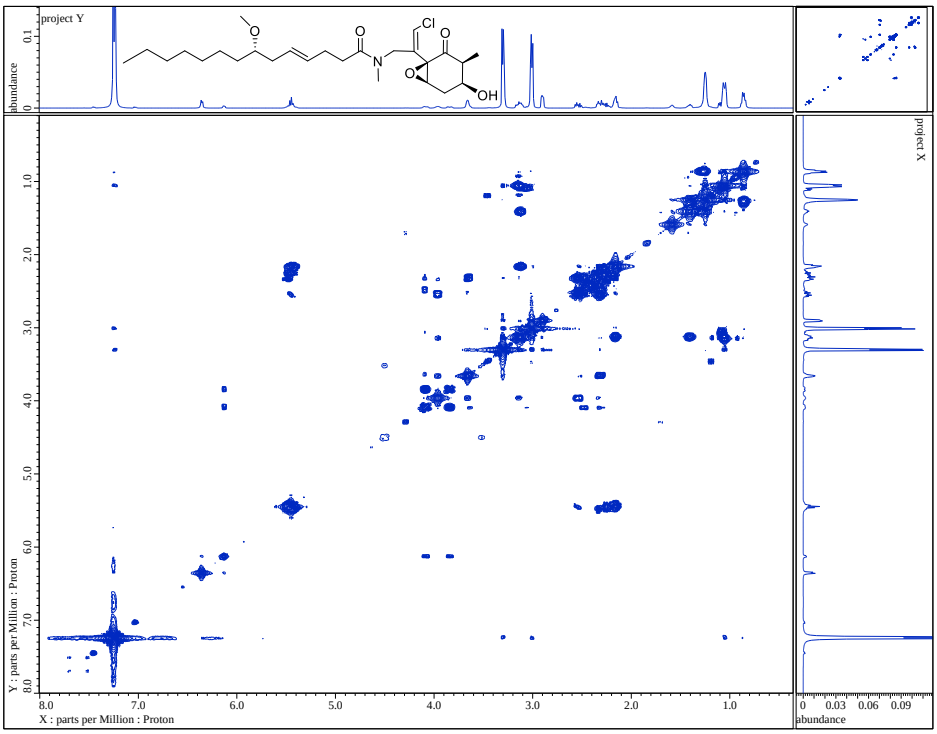


Figure A2.17. COSY of malyngamide I. (500 MHz, $CDCl_3$).

Table A2.1. ^{13}C-NMR and ^{1}H-NMR comparison of malyngamide C acetate isolated in previous study (left) vs. malyngamide C acetate isolated from *O. hirsuta* PAB10Feb10-1 (right).

Malyngamide C acetate, previous isolation[197]				Malyngamide C acetate, this study	
#	δ_H, type, *J* (Hz) $CDCl_3$	δ_C, $CDCl_3$		δ_H, type, *J* (Hz) $CDCl_3$, 25°C	δ_C $CDCl_3$, 25°C
14'	0.86, m	14.1		0.87, t, (7.2)	14.2
13'	1.25, m	22.6		1.26, m	22.7
12'	1.25, m	29.7		1.26, m	29.8
11'	1.25, m	29.3		1.26, m	29.4
10'	1.25, m	25.3		1.26, m	25.3
9'	1.25, m	31.8		1.26, m	31.9
8'	1.42, m	33.3		1.42, m	33.3
7'	3.12, m	80.7		3.14, m	80.7
6'	2.25, m	36.3		2.24, m	36.3
5'	5.46, m	127.6		5.48, m	127.6
4'	5.46, m	130.7		5.48, m	130.7
3'	2.29, m	28.4		2.32, m	28.5
2'	2.17, m	36.3		2.18, br t, (4.5)	36.4
1'	-	172.4		-	172.3
1	3.85, dd, (4.8, 14.7) 3.97, dd, (4.8, 14.7)	40.3		3.85, dd, (5.8, 15) 4, dd, (5.8,15)	40.4
2	-	133.1		-	133
3	6.39, s	122.8		6.4, s	122.9
4	-	61.5		-	61.5
5	-	200.9		-	201
6	2.35, m 2.6, dt, (4.0, 17.8)	35		2.35, m 2.62, dt, (4.0, 17.9)	35.1
7	2.01, m	21.5		2.03, m	21.4
8	5.44, m	68.4		5.48, m	68.4
9	3.64, s	62.6		3.65, s	62.6
15' (*O*-Me)	3.3, s	56.4		3.32, s	56.5
16'	-	170.4		-	170.5
17'	2.14	20.9		2.15, s	21.1
NH	6.1 (4.8)	-		6.07, br t, (4.8)	-

Table A2.2. ^{13}C-NMR and ^{1}H-NMR comparison of malyngamide I isolated in previous study (left) vs. malyngamide I isolated from *O. hirsuta* PAP21Jun06-1 (right), including HMBC and COSY correlations. Double peaks due to presence of minor *N*-amide structural conformer. *N.D.*= peak not detected due to overlapping signals in 2D plots.

Malyngamide I, original isolation[155]					Malyngamide I, this study			
#	δ_H, type, *J* (Hz) DMSO-d_6 97°C	δ_C DMSO-d_6 25°C	δ_C $CDCl_3$ 25°C		δ_H, type, *J* (Hz) $CDCl_3$ 25°C	δ_C $CDCl_3$ 25°C	HMBC $CDCl_3$ 25°C	COSY $CDCl_3$ 25°C
14'	0.875, bt, (6.6)	13.9	13.9		0.84, bt, (7.0)	14.0	13', 15"	13', 15"
13'	1.3, m	22.0	22.5		1.26, m	22.3	14', 15"	14', 15"
12'	1.3, m	29.1	29.6		1.22, m	31.6	13', 14', 16'	13'
11'	1.3, m	27.8	28.2		1.26, m	29.8	*N.D.*	*N.D.*
10'	1.3, m	28.6	29.1		1.26, m	29.5	*N.D.*	*N.D.*
9'	1.3, m	24.6	25.1		1.26, m	25.3	8', 10', 16'	8'
8'	1.4, m	32.6, 32.4	33.2, 33.1		1.41, m	33.3	6', 7', 9', 10'	7', 9'
7'	3.14, dddd (5.9, 5.8, 5.7, 5.5)	79.8	80.6		3.12, m	80.8	5', 6', 8', 9', 15'	6', 8'
6'	2.12, m	35.8	36.2		2.15, m	36.1	4', 5', 7', 8'	5', 7'
5'	5.46, m	126.55, 126.48	127.1, 123.7		5.45, m	130.7	2', 3', 4', 6', 7'	4', 6'
4'	5.46, m	131.4, 131.3	131.0, 130.8		5.45, m	127.3	2', 3', 5', 6', 7'	2', 3', 5'
3'	2.21, m	27.5	27.5		2.23, m	27.4	1', 2', 4', 5'	2', 4'
2'	2.35, m	31.6, 31.2	31.6		2.31, m	33.1	1', 3', 4', 5'	3', 4'
1'	-	171.8, 171.4	174.0, 173.0		-	173.3, 174.9	-	-
1	3.96, m	50.5, 47.9	51.0, 50.2		3.8, dd, (17.8, 2.0) 4.1, dd (17.8, 2.0)	50.6	1', 2, 3, 4, 16'	16', 1
2	-	132.8, 132.3	132.2, 131.5		-	132.4, 131.9	-	-
3	6.37, 6.13, bt	120.7, 124.7	120.2		6.12, 6.38, bt	120.8, 124.7	1, 2, 9, 15'	1
4	-	61.2, 60.9	62.3, 60.9		-	62.8, 61.1	-	-
5	-	204.6, 204.5	205.2, 204.3		-	204.5, 205.6	-	-
6	2.92, qd, (6.8, 2.5)	42.8	43.6, 43.4		3.13, m	43.6	5, 7, 10	7, 10
7	3.96, m	72.7	74.8, 73.5		3.95, 4.09, m	75.4, 74.5	6, 9	6, 8, 9
8	2.35, 2.21, m	31.6, 31.2	32.5, 32.3		2.31,2.52, m	32.6	4, 6, 7, 9	7, 9
9	3.63, dd, (3.6, 1)	64, 63.5	63.7		3.65, m	64.5	2, 4, 7, 8	7, 8
10	0.978, d, (6.7)	11.0	10.7		1.05, 1.09, d, (7.0)	10.7, 10.9	5, 6, 7	6
15' (*O*-Me)	3.24, s	55.7	56.3		3.3, s	56.4	7'	-
16' (*N*-Me)	2.86, bs	34.5, 33.1	34.7, 33.8		3.0, 2.9, s	34	1, 1'	-
OH	2.12, m	-	-		3.5,4.5, bs	-	-	-

Table A2.3. *Okeania* genome assembly statistics.

	O. hirsuta PAB10Feb10-1	*O. hirsuta* PAP21Jun06-1
Genbank Accession #	RCBY00000000	RCBZ00000000
Pathway Accession #	MK142792	MK142793
# contigs	716	392
Total length	8.8 Mb	8.1 Mb
Largest contig	114658 bp	152661 bp
N50	27938	37228
GC%	36.6	36.4
Coverage	37.97	26.64
# N's/100 kbp	0	439
Sequencing method	Illumina MiSeq®, Pacific Biosciences	Illumina MiSeq®
Assembly method	SPAdes 3.6 hybrid assembly, Medusa 1.0	SPAdes 3.6 assembly

Table A2.4. Proposed functions of ORFs in A) *mgc* gene cluster, PAB and B) *mgi* gene cluster, PAP

A)

Protein	Length	Proposed Function	Similarity	Identity	Similarity	Accession
Orf1	779		Heterocyst differentiation protein, *Planktothrix serta* PCC 8927	46%	60%	CUR22335.1
Orf2	313		Heterocyst differentiation protein, *Trichodesmium erythraeum*	76%	85%	WP_011613549.1
MgcA	249	Octanoyltransferase	Octanoyltransferase, *Bacillus subtilis*	36%	49%	KJJ40644.1
MgcB	297	Transposase	IS4 family transposase, *Blastocatellia bacterium* AA13	38%	55%	PYP85004.1
MgcC	184	Transposase	Transposase, *Paenibacillus alvei*	42%	60%	WP_051135135.1
MgcD	245	sfp-type phosphopantetheinylase	4'-phosphopantetheinyl transferase, *Hydrocoleum* sp. CS-953	80%	87%	WP_094673082.1
MgcE	87	Transposase	Transposase, *Trichodesmium erythraeum*	84%	90%	WP_047155615.1
MgcF	102	Unknown	JamB, *Lyngbya majuscula (Moorea producens)*	54%	69%	AAS98775.1
MgcG	1903	KS AT KR O-MT ACP	CurL, *Lyngbya majuscula (Moorea producens)*	48%	66%	AAT70107.1
MgcH	1681	KS AT KR (DH) ACP	CurM, *Lyngbya majuscula (Moorea producens)*	64%	77%	AAT70108.1
MgcI	2184	KS AT DH ER KR ACP	JamL, *Lyngbya majuscula (Moorea producens)*	76%	86%	AAS98783.1
MgcJ	1587	C A(Gly) N-MT ACP	PuwA, *Cylindrospermum alatosporum* CCALA 988	46%	63%	AIW82277.1
MgcK	1241	KS AT ACP ACP	AprB, *Moorea bouillonii*	62%	75%	WP_075900458.1
MgcL	82	ACP	AprC, *Moorea bouillonii*	89%	96%	WP_075900456.1
MgcM	407	KS	AprD, *Moorea bouillonii*	91%	94%	WP_075900454.1
MgcN	419	HMG-CoA Synthase	AprE, *Moorea bouillonii*	87%	92%	WP_075900452.1
MgcO	480	Halogenase	JamE, *Lyngbya majuscula (Moorea producens)*	61%	74%	AAS98777.1
MgcP	255	ECH	Enoyl-CoA hydratase/isomerase, *Moorea producens*	77%	86%	WP_071106602.1
MgcQ	2790	ECH KS AT DH^0 MT^0 KR^0 ACP	AprG, Moorea bouillonii	71%	83%	WP_081431224.1
MgcR	2577	KS AT DH ER KR ACP R	JamL, *Lyngbya majuscula (Moorea producens)*	56%	73%	AAS98783.1
MgcS	169	Cyclase	Cyclase, *Kamptonema*	29%	47%	WP_007358359.1
MgcT	451	P450	Cytochrome P450, *Calothrix* sp. HK-06	59%	76%	WP_073621730.1
MgcU	466	P450	Cytochrome P450, *Hassallia byssoidea* VB512170	58%	73%	KIF36330.1
MgcV	443	Acetyltransferase	JamQ, *Lyngbya majuscula (Moorea producens)*	75%	85%	AAS98788.1
Orf1'	344		Phospholipase/lecithinase/hemolysin, *Rivularia* sp. PCC 7116	67%	80%	WP_015122243.1
Orf2'	154	Transposase	Transposase, *Moorea producens* PAL	69%	78%	OLT65278.1
Orf3'	124	Transposase	Transposase, *Moorea producens* JHB	74%	87%	AOY84486.1
Orf4'	96	Transposase	Transposase, *Moorea producens* PAL	60%	72%	OLT65278.1
Orf5'	434		Linear amide C-N hydrolase, *Moorea producens*	54%	66%	WP_071106594.1
Orf6'	595		AarF/ABC1/UbiB, *Trichodesmium erythraeum*	63%	81%	WP_047156081.1

B)

Protein	Length	Proposed Function	Similar Sequence	Identity	Similarity	Accession
Orf1	779		Heterocyst differentiation protein, *Planktothrix serta* PCC 8927	46%	60%	CUR22335.1
Orf2	305		Heterocyst differentiation protein, *Trichodesmium erythraeum*	76%	85%	WP_011613549.1
MgiA	249	Octanoyltransferase	Octanoyltransferase, *Bacillus subtilis*	36%	49%	KJJ40644.1
MgiB	325	Transposase	Transposase, *Victivallales bacterium* CCUG 44730	34%	53%	WP_106052395.1
MgiC	124	Transposase	Transposase, *Paenibacillus alvei*	35%	57%	WP_051135135.1
MgiD	245	sfp-type phosphopantetheinylase	4'-phosphopantetheinyl transferase, *Hydrocoleum* sp. CS-953	80%	87%	WP_094673082.1
MgiE	178	Transposase	Transposase, *Trichodesmium erythraeum*	83%	90%	WP_047155615.1
MgiF	101	Unknown	JamB, *Lyngbya majuscula (Moorea producens)*	54%	69%	AAS98775.1
MgiG	1902	KS AT KR O-MT ACP	CurL, *Lyngbya majuscula (Moorea producens)*	48%	66%	AAT70107.1
MgiH	1681	KS AT KR (DH) ACP	CurM, *Lyngbya majuscula (Moorea producens)*	63%	77%	AAT70108.1
MgiI	2184	KS AT DH ER KR ACP	JamL, *Lyngbya majuscula (Moorea producens)*	77%	86%	AAS98783.1
MgiJ	1586	C A(Gly) N-MT ACP	PuwA, *Cylindrospermum alatosporum* CCALA 988	47%	64%	AIW82277.1
MgiK	1238	KS AT ACP ACP	AprB, *Moorea bouillonii*	63%	76%	WP_075900458.1
MgiL	79	ACP	AprC, *Moorea bouillonii*	91%	96%	WP_075900456.1
MgiM	406	KS	AprD, *Moorea bouillonii*	90%	94%	WP_075900454.1
MgiN	419	HMG-CoA Synthase	AprE, *Moorea bouillonii*	91%	95%	WP_075900452.1
MgiO	480	Halogenase	JamE, *Lyngbya majuscula (Moorea producens)*	62%	76%	AAS98777.1
MgiP	254	ECH	Enoyl-CoA hydratase/isomerase, *Moorea producens*	80%	87%	WP_071106602.1
MgiQ	2772	ECH KS AT DH MT KR^0 ACP	AprG, *Moorea bouillonii*	73%	84%	WP_081431224.1
MgiR	2008	KS AT KR ACP R	CurM, *Lyngbya majuscula (Moorea producens)*	58%	75%	AAT70108.1
MgiS	166	Cyclase	Cyclase, *Kamptonema*	27%	48%	WP_007358359.1
MgiT	451	P450	Cytochrome P450, *Calothrix* sp. HK-06	56%	77%	WP_073621730.1
Orf1'	436		Hypothetical protein, *Trichodesmium erythraeum*	29%	51%	WP_047156304.1
Orf2'	593		AarF/ABC1/UbiB kinase, *Trichodesmium erythraeum*	66%	82%	WP_047156081.1

Table A2.5. Primers used in this study.

Primers used in PAB and PAP sequencing and assembly confirmation			
Name	Direction	Sequence	Description
P093F3	Forward	CTATTATCGTGGCATCCG	mgcF to mgcG AT domain, PAB
P093R2	Reverse	CTCGCTGAAATGGATAAGTT	mgcF to mgcG AT domain, PAB
P093F4	Forward	CTACTGAGAACTTTGTACTGTCCAG	Inter mgcC-mgcD region to mgcG AT domain, PAB
P093R4	Reverse	TACTTGAGATTCTGATGCTAAC	Inter mgcC-mgcD region to mgcG AT domain, PAB
P093F5	Forward	CCCAATGAATGATGCTGAAG	mgcD to mgcF, PAB
P093R5	Reverse	GATTTATCTGATTGTTTCACTG	mgcD to mgcF, PAB
P093R6	Reverse	ATCGTATCTTGGTGTCCTATG	Inter mgcC-mgcD region or mgcD to intra mgcE-mgcF region, PAB
P093R7	Reverse	CCATTAGAGATCGGTAAAGC	Inter mgcC-mgcD region or mgcD to intra mgcE-mgcF region, PAB
P094F	Forward	AAGTAGAGTATTAAGCAGGC	Inter Orf1-Orf2 region to inter mgcC-mgcD region, PAB
P094R	Reverse	CTGGACAGTACAAAGTTCTCAGTAG	Inter Orf1-Orf2 region to inter mgcC-mgcD region, PAB
P108F	Forward	CCGAACAGAAACCCGTTAAAC	mgcT to mgcV, PAB
P108R	Reverse	GACGCTCTTGAATAATATCCAGC	mgcT to mgcV, PAB
P114F	Forward	GTCCTCTACAAACTTTTGACATTTC	Orf1' to Orf5', PAB
P114R	Reverse	AACTCCTGCTTGATAAGTCC	Orf1' to Orf5', PAB
P115F	Forward	GAATTGGATGGGTGAAAACTTACG	mgiE to mgiG, PAP
P115R	Reverse	CTGGGCATCAAACTGGTCTATAC	mgiE to mgiG, PAP
P115F2	Forward	GCATTGATTCGTACTCTTG	mgiB to inter mgiC-mgiD, PAP
P115R2	Reverse	AACTGGACAGTACAAAGTTCTC	mgiB to inter mgiC-mgiD, PAP
Primers used in MgcQ recombinant protein expression			
P111F	Forward	CAGCAGCGGCCTGGTGCCGCGCGGCAGCCATTTGTATGAAGTAGAATGGAGATCC	amplify PAB β-ribbon-MT-KR, overlap with pET28B, cut with NdeI and xhoI
P111R5	Reverse	GGATCTCAGTGGTGGTGGTGGTGGTGTCATTCTTCCAAGTAATCAGTTAGTGACTG	amplify PAB β-ribbon-MT-KR, stop codon overlap with pET28B, cut with NdeI and xhoI
P1	Forward	TTGTATGAAGTAGAATGGAGA	Amplification of mgcQ MT-KR from B050V
P2	Reverse	ATGAAATAGATTCAGGATTTTT	Amplification of mgcQ MT-KR from B050V
P3	Forward	GATGCTACCTACTTAATTACAGGAGGTATTAGCAGTTTAGGGTTA	MgcQ MT-KR E7G Mutagenesis
P4	Reverse	TAACCCTAAACTGCTAATACCTCCTGTAATTAAGTAGGTAGCATC	MgcQ MT-KR E7G Mutagenesis
P5	Forward	ATGCTACCTACTTAATTACAGGAGGTATTGGCAGTTTAGGGTTATTGGTGG	MgcQ MT-KR E7G/ S10G Mutagenesis
P6	Reverse	CCACCAATAACCCTAAACTGCCAATACCTCCTGTAATTAAGTAGGTAGCAT	MgcQ MT-KR E7G/ S10G Mutagenesis
P7	Forward	TACTTCCAATCCAATGCCATGACTCTGGGTTCAGATTTG	Amplification of curJ MT-KR
P8	Reverse	TTATCCACTTCCAATGCTAAGAAGTAGTTTGGATTGTCTCC	Amplification of curJ MT-KR

Table A2.6. Enrichment calculations for [1-^{13}C]octanoate labeled malyngamide C acetate.[89,161] C-7' octanoate enriched carbon depicted in bold underline.

		[1-^{13}C]octanoate labeled malyngamide C acetate						Native abundance malyngamide C acetate											
			Integral normalized to:						Normalization factor:					% enrichment normalized to:					Mean % enrichment
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
C#	δ_C	Integral	C15'	C1	C3	C2'	C12'	Integral	C15'	C1	C3	C2'	C12'	C15'	C1	C3	C2'	C12'	
14'	14.1	0.98	0.95	1.00	0.92	1.01	0.78	0.77	0.92	0.70	0.69	0.74	0.71	3.80	42.86	33.22	36.46	9.96	25.26
13'	22.7	1.06	1.03	1.08	0.99	1.09	0.85	0.89	1.06	0.81	0.79	0.86	0.82	-2.87	33.68	24.67	27.70	2.90	17.22
12'	29.8	1.25	1.21	1.28	1.17	1.29	1.00	1.08	1.29	0.98	0.96	1.04	1.00	-5.61	29.91	21.15	24.09	0.00	13.91
11'	29.3	1.23	1.19	1.26	1.15	1.27	0.98	0.97	1.15	0.88	0.87	0.93	0.90	3.41	42.33	32.73	35.95	9.56	24.80
10'	25.3	1.05	1.02	1.07	0.98	1.08	0.84	1.08	1.29	0.98	0.96	1.04	1.00	-20.71	9.13	1.77	4.24	-16.00	-4.32
9'	31.9	1.02	0.99	1.04	0.95	1.05	0.82	0.95	1.13	0.86	0.85	0.91	0.88	-12.44	20.52	12.39	15.12	-7.23	5.67
8'	33.3	1.1	1.07	1.12	1.03	1.13	0.88	1.09	1.30	0.99	0.97	1.05	1.01	-17.70	13.27	5.63	8.20	-12.81	-0.68
7'	80.7	4.6	4.47	4.69	4.30	4.74	3.68	1.16	1.38	1.05	1.04	1.12	1.07	223.40	345.11	315.08	325.17	242.62	**290.28**
6'	36.4	1.15	1.12	1.17	1.07	1.19	0.92	0.97	1.15	0.88	0.87	0.93	0.90	-3.31	33.07	24.10	27.11	2.43	16.68
5'	127.6	1	0.97	1.02	0.93	1.03	0.80	1	1.19	0.91	0.89	0.96	0.93	-18.45	12.24	4.67	7.22	-13.60	-1.58
4'	130.7	0.92	0.89	0.94	0.86	0.95	0.74	0.97	1.15	0.88	0.87	0.93	0.90	-22.65	6.46	-0.72	1.69	-18.05	-6.66
3'	28.5	0.98	0.95	1.00	0.92	1.01	0.78	0.97	1.15	0.88	0.87	0.93	0.90	-17.61	13.40	5.75	8.32	-12.71	-0.57
2'	36.3	0.97	0.94	0.99	0.91	1.00	0.78	1.04	1.24	0.95	0.93	1.00	0.96	-23.94	4.69	-2.37	0.00	-19.42	-8.21
1'	172.3	0.45	0.44	0.46	0.42	0.46	0.36	0.3	0.36	0.27	0.27	0.29	0.28	22.33	68.37	57.01	60.82	29.60	47.63
1	40.3	0.98	0.95	1.00	0.92	1.01	0.78	1.1	1.31	1.00	0.98	1.06	1.02	-27.34	0.00	-6.75	-4.48	-23.03	-12.32
2	122.9	1.07	1.04	1.09	1.00	1.10	0.86	1.12	1.33	1.02	1.00	1.08	1.04	-22.09	7.23	0.00	2.43	-17.46	-5.98
3	133.0	0.45	0.44	0.46	0.42	0.46	0.36	0.47	0.56	0.43	0.42	0.45	0.44	-21.92	7.47	0.22	2.65	-17.28	-5.77
4	61.5	0.33	0.32	0.34	0.31	0.34	0.26	0.24	0.29	0.22	0.21	0.23	0.22	12.14	54.34	43.93	47.42	18.80	35.32
5	201.0	0.31	0.30	0.32	0.29	0.32	0.25	0.36	0.43	0.33	0.32	0.35	0.33	-29.77	-3.34	-9.87	-7.67	-25.60	-15.25
6	35.0	0.96	0.93	0.98	0.90	0.99	0.77	0.95	1.13	0.86	0.85	0.91	0.88	-17.59	13.43	5.77	8.35	-12.69	-0.55
7	21.4	0.99	0.96	1.01	0.93	1.02	0.79	0.9	1.07	0.82	0.80	0.87	0.83	-10.29	23.47	15.14	17.94	-4.96	8.26
8	68.4	1.08	1.05	1.10	1.01	1.11	0.86	1.17	1.39	1.06	1.04	1.13	1.08	-24.72	3.61	-3.38	-1.03	-20.25	-9.15
9	62.5	1.1	1.07	1.12	1.03	1.13	0.88	1.18	1.40	1.07	1.05	1.13	1.09	-23.98	4.64	-2.42	-0.05	-19.46	-8.25
15'	56.5	1.03	1.00	1.05	0.96	1.06	0.82	0.84	1.00	0.76	0.75	0.81	0.78	0.00	37.63	28.35	31.47	5.94	20.68
16'	170.5	0.33	0.32	0.34	0.31	0.34	0.26	0.34	0.40	0.31	0.30	0.33	0.31	-20.85	8.94	1.59	4.06	-16.14	-4.48
17'	21.0	0.8	0.78	0.82	0.75	0.82	0.64	0.69	0.82	0.63	0.62	0.66	0.64	-5.45	30.14	21.36	24.31	0.17	14.11

Table A2.7. NMR enrichment values for δ_C shifts in malyngamide C acetate from label-fed cultures. Column 3: J_{CC} values calculated between flanking carbon resonances of malyngamide C acetate for acetate-derived carbon from [1,2-$^{13}C_2$]acetate-fed PAB. Hexanone carbons indicated in bold/italic. Column 4: J_{CC} values between flanking carbon resonances of malyngamide C acetate observed for glycine-derived carbon from [1,2-$^{13}C_2$]glycine-fed PAB.

Labeled substrate feeding to PAB			
		[1,2-$^{13}C_2$]acetate	[1,2-$^{13}C_2$]glycine
1	2	3	4
Carbon #	δ_C	J_{CC} (Hz)	J_{CC} (Hz)
14'	14.2	35	-
13'	22.7	35	-
12'	29.8	35	-
11'	29.4	35	-
10'	25.3	37	-
9'	31.9	35	-
8'	33.3	39	-
7'	80.7	39	-
6'	36.3	43	-
5'	127.6	44	-
4'	130.7	45	-
3'	28.5	43	-
2'	36.4	50	-
1'	172.3	49	-
1	40.4	-	46
2	133	-	46
3	122.9	-	-
4	61.5	***52***	-
5	201.0	***53***	-
6	35.1	***31***	-
7	21.4	***33***	-
8	68.4	***48***	-
9	62.6	***48***	-
15'	56.5	-	-
16'	170.5	60	-
17'	21.1	59	-

Table A2.8. Accession numbers and databases used in alignments and PFAM03099 phylogeny

Protein	Alignment, figure	Source	Accession #
AmphB	KR, 5	Genbank	AAK73513.1
AmphI	KR, 5	Genbank	AAK73501.1
CurJ	KR/DH/MT, 5/S5a/S5b	Genbank	AEE88280.1
EpoE	KR, 5	Genbank	AAF62884.1
JamL	KR, 5	Genbank	AAS98783.1
NysI	KR, 5	Genbank	AVX51098.1
PimS2	KR, 5	Genbank	CAC20921.1
Raps2-6	KR, 5	Genbank	CAA60459.1
StiH	KR, 5	Genbank	CAD19092.1
EryA	DH, S5a	Genbank	AAA26493.2
CrpA	DH/MT, S5a/S5b	Genbank	EF159954.1
GphF	DH/MT, S5a/S5b	Genbank	AHA38199.1
JamJ	DH/MT, S5a/S5b	Genbank	AAS98781.1
CyrB	MT, S5b	Genbank	AHN91582.1
AprG	MT, S5b	Genbank	WP_081431224.1
EqxS	R, S5c	Genbank	S4W172.1
Lys2	R, S5c	Genbank	AJQ13116.1
SfmC	R, S5c	Genbank	ABI22133.1
MxcG	R, S5c	Genbank	AAG31130.1
LtxA	R, S5c	Genbank	AAT12283.1
LgrD	R, S5c	Genbank	WP_081437244.1
ColG	R, S5c	Genbank	WP_081431359.1
NcpB	R, S5c	Genbank	WP_114082684.1
CpaS	R, S5c	Genbank	RAQ57019.1
Fsds	R, S5c	Genbank	Q5SBL6.2
LipM, *B. subtilis*	PFAM03099, 3	Uniprot	P54511
LipM, *S. thermophilum*	PFAM03099, 3	Uniprot	Q67N40
LipM, *M. thermoacetica*	PFAM03099, 3	Uniprot	Q2RH52
LipM, *B. anthracis*	PFAM03099, 3	Uniprot	Q81M24
LipM, *T. erythraeum*	PFAM03099, 3	Genbank	WP_011613204.1
LipM, *S.* sp. PCC6803	PFAM03099, 3	Genbank	WP_010872753.1
LipM, *N. spumigena*	PFAM03099, 3	Genbank	WP_063871123.1
LipM, *S. epidermidis*	PFAM03099, 3	Uniprot	Q5HP16
BirA, *E. coli*	PFAM03099, 3	Uniprot	P06709
BirA, *S.* sp. PCC6803	PFAM03099, 3	Uniprot	P72816
BirA, *B. subtilis*	PFAM03099, 3	Uniprot	P0CI75
LipL, *S. epidermidis*	PFAM03099, 3	Uniprot	Q5HRF6
LipL, *E. sibiricum*	PFAM03099, 3	Uniprot	B1YHU0
LipL, *B. subtilis*	PFAM03099, 3	Uniprot	P39648
LipL, *B. haloduruans*	PFAM03099, 3	Uniprot	A0A0F7HN74
LipL, *B. anthracis*	PFAM03099, 3	Uniprot	Q81JR5
LplA, *A. pernix*	PFAM03099, 3	Uniprot	Q9Y9E6
LplA, *S. pombe*	PFAM03099, 3	Uniprot	O13629
LplA, *Candidatus B. floridanus*	PFAM03099, 3	Uniprot	Q7VR65
LplA, *C. sakazakii*	PFAM03099, 3	Uniprot	A7MIG6
LplA, *E. coli*	PFAM03099, 3	Uniprot	P32099
LipB, *B.* sp. ORS 278	PFAM03099, 3	Uniprot	A4YT30
LipB, *G. bemidjiensis*	PFAM03099, 3	Uniprot	B5EDZ5
LipB, *D. radiodurans*	PFAM03099, 3	Uniprot	Q9RWA5
LipB, *C. subterraneus*	PFAM03099, 3	Uniprot	Q8R9E0

2.7 Notes

Note A1: malygnamide T contains an unsaturated pyran head group and O, P, and 3 are uncyclized, thus they do not fall into either category. M, characterized with a phenolic head group, likely undergoes keto-enol tautomerization and is therefore classified as type A.

Note A2:an environmental collection of *O. hirsuta* PAB was originally attributed as the source of the macrolides bastimolides A and B, however laboratory cultures do not produce these compounds, and genome sequencing suggests that they lack the PKS pathway predicted to encode for bastimolide biosynthesis. An environmental collection of *O. hirsuta* PAP was attributed as the producer of tetrapeptides venturamide A and B, however laboratory cultures do not produce these compounds, and genome sequencing offers no obvious pathway.[198,199]

Note A3: with the exception of mgcR/mgiR, all KS and NRPS module organization is identical between PAB/mgc and PAP/mgi pathways – references to mgc genes and proteins are equivalent to mgi genes and proteins unless otherwise explicitly noted.

Note A4: standards from in-house pure compound library derived from earlier isolations from environmental sources.

Note A5: repeated LC/MS injections of crude PAB extract over time showed an increase in $[M+H]^+$ 456.2 and decrease in $[M+H]^+$ 498.2, suggesting that the presence of malyngamide C in this organism may be a degradation product of malyngamide C acetate.

2.8 Acknowledgments

We would like to acknowledge the help of M. A. Skiba and Q. Dan for construct design and recombinant protein biochemistry, F. Fu for advisement on the acetylene reduction assay, C. Larsen for genomic DNA extraction, Y. Su and L. Gross for high resolution mass spectrometry, A. Mrse for NMR consultation and instruction, and M. Kissick, A.-M. Hoskins, P. Kanjanakantorn, and B. Ni for cyanobacterial culturing and extraction.

This chapter, in full, is a reprint, with permission, of the material as it appears in ACS Chemical Biology, 2018, Moss, Nathan A.; Leão, Tiago F.; Rankin, Michael R.; McCullough, Tyler M.; Qu, Pingping; Korobeynikov, Anton; Smith, Janet L.; Gerwick, Lena; and Gerwick, William H. 2018, *13*, pages 3385-3395. The dissertation author is the primary investigator and author of this material.

Chapter 3: Combinatorial assembly-line biosynthesis produces vatiamides A-F

3.1 Abstract

Hybrid type I PKS/NRPS pathway biosynthesis typically proceeds in a collinear manner: one molecular building block is enzymatically attached to its predecessor in an ordered sequence coincident with gene arrangement. Here, genome mining combined with use of a synthetic coumarin-based click probe led to the discovery and characterization of vatiamides A-F: three alkynylated lipopeptides and their brominated analogs from the cyanobacterium *Moorea producens* ASI16Jul14-2. Vatiamides A-F are predicted to derive from a unique multichannel non-collinear 90 kb biosynthesis *vat* pathway assembled from long and short-read genome sequencing. Within this pathway, an upstream PKS cassette VatA-VatM interacts with three separate downstream cognate NRPS partners, VatN, VatQ, and VatS to form brominated and non-brominated versions of a molecule comprised of a common fatty-acid tail and three disparate head groups. An uncharacterized aminotransferase is proposed to generate the terminal amide in vatiamide E/F biosynthesis. The non-collinearity in this system is facilitated by a promiscuous intermodule PKS-NRPS docking scheme; VatM and VatP C-terminal ACP-bound docking domains are predicted to interact with VatN, VatR, and VatS N-terminal docking domains of identical sequence and without any selectivity. This mode enables a combinatorial biosynthetic capacity unprecedented in prokaryotic systems and represents an entirely novel scheme for diversification of natural product biosynthesis in a single organism.

3.2 Introduction

Numerous microbial species contain hybrid polyketide synthase/non ribosomal peptide synthetase (PKS/NRPS) biosynthetic gene clusters which produce diverse modified lipopeptide natural products (NP).[200] As discussed in chapter one, the study of saxitoxin, cylindrospermopsin, curacin A, and apratoxin A pathways, among others, has provided much knowledge of protein-protein interaction, novel gene function, and loading module diversity in NP biosynthesis.[60,89,94,95,112,161,201] In typical PKS/NRPS systems, activated acyl-CoAs and amino acids are sequentially coupled by individual PKS or NRPS modules. This

occurs collinearly such that the order in which their genes are present and then transcribed is coincident with the functioning of proteins that selectively interact to elongate NPs with high fidelity. Initially, bioassay-guided fractionation of two modestly active fractions prompted characterization of vatiamides A-D (**1-4**) (Figure 3.1). Each analog bore identical fatty acid tails (e.g. 7/**8**) which structurally match that

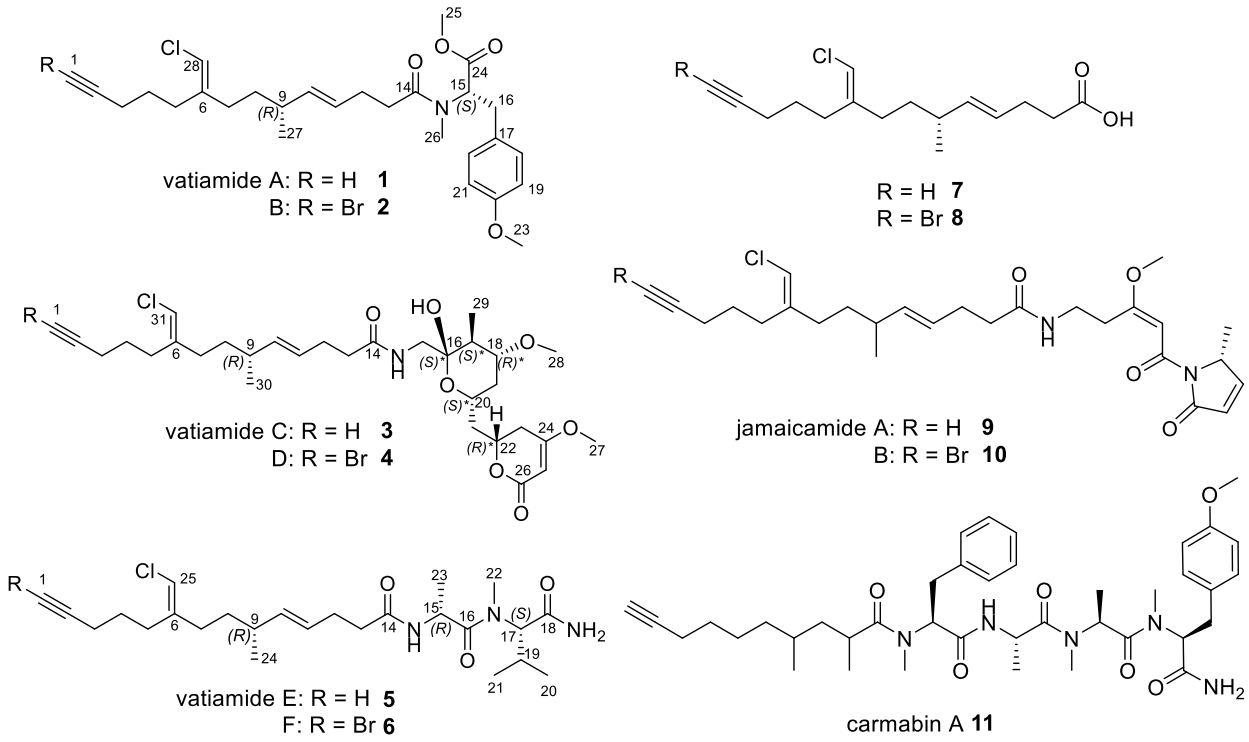


Figure 3.1. Vatiamides **1-6** from *Moorea producens* ASI16Jul14-2 "ASI" isolated in this study, jamaicamides **9/10** from *Moorea producens* JHB, and carmabin (**11)** from *Moorea producens* 3L.

portion of jamaicamides A-B from *Moorea producens* JHB (**9**/**10**). This discovery prompted a second-generation whole genome sequencing of *Moorea producens* ASI16Jul14-2. Hybrid assembly of short and long reads led to the discovery of a 90 kb hybrid PKS/NRPS pathway, which contains several unique features. First, a JamABC analog cassette, VatABC, is followed by modules that exactly match the modular order of the jamaicamide pathway in VatD-M. The biosynthesis of **1** and **2** is predicted to involve the use NRPS VatN and *O*-MT VatO, respectively. The biosynthesis of **3** and **4** occurs by way of module skipping, whereby **7** and **8** are abstracted from VatM (or standalone homolog, VatP) to VatS, an NRPS, and then processed by VatT-VatX, breaking the standard model of collinearity.

3.2.1 β−αββαα-type docking domain background

As discussed in chapter 1, collinearity is an important aspect of type I PKS/NRPS biosynthetic pathways in cyanobacteria. Indeed, all characterized PKS/NRPS pathways in cyanobacteria appear to follow this procession of modules: one gene, one protein, one enzymatic reaction, in the same order as the genes are arranged and transcribed to form a molecule with an identical molecular backbone with each enzymatic cycle. A high degree of selectivity has been exemplified in the study of the curacin A pathway, wherein type II docking domains equip CurG – CurM, seven consecutive PKS modules, with the ability to only interact with their cognate partners in the production of curacin A.[112] In other pathways, the interactions between modules frequently occur by contacts of COM domains between NRPS-NRPS junctions and type I docking domains between KS domains.[111,113]

A fourth type of docking domain, found between the C-terminal thiolase and N-terminus of some PKS modules, NRPS modules, and tailoring domains has been characterized in the epothilone, tubulysin, and most thoroughly in a rhapdopeptide/xenorpeptide-like (RXP) pathway.[114–116] Here we refer to this docking domain as a β−αββαα-type docking domain, denoting its two main structural features: the upstream C-terminal docking domain (Cdd) beta sheet "β" and its cognate downstream partner "αββαα," the N-terminal docking domain (Ndd) comprised of an alpha helix, two consecutive beta sheets, and two more alpha helices, forming a hairpin in an αββαα arrangement.[114–118] This type of docking domain was biochemically characterized as a "recognition sequence" in the interaction between EpoA and EpoB in epothilone biosynthesis, via complementation studies.[115] Subsequently, the N-terminal docking motif in the aforementioned docking interaction was characterized by solution NMR in the study of the TubB-TubC interaction system in the biosynthesis of tubulysin; this research led to the discovery of the $\alpha_1\beta_1\beta_2\alpha_2\alpha_3$ fold structure of the docking interaction, as well as the identification of a putative docking "key" encoded by polar residues of the β_2 sheet of the Ndd.[114] A later study analyzed the tri-module NRPS RXPs pathway in *Xenorhabdus stockiae* KJ12.1. The authors were able to generate an NMR-solution based structure of the final residues of the C-terminal dd interaction with the N-terminal dd by

constructing a flexible $(Gly\text{-}Ser)_n$ linker between the two interacting domains. The final five residues of the C-terminal docking domain were found to form a β-sheet, dubbed β_3 in the β–αββαα docking domain type, and interact with a set of 4-5 residues of the β_2 sheet of the N-terminal docking domain at amino acid positions 24-29, and depicted by cartoon in Figure 3.2. These "recognition rules" were observed in

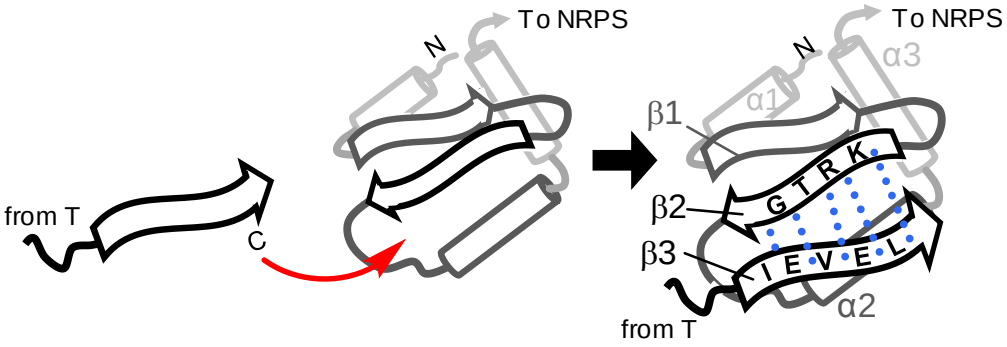


Figure 3.2. Depiction of β-αββαα docking domain interaction. C: C-terminus of upstream module, N: N-terminus of downstream module. Blue dots represent electrostatic and dipole interaction between β-sheets, conferring specificity.

similar docking domains across bacterial taxa. The RXPs pathway highlights a method by which structural diversity is generated by an unusual concentration-dependent module processing order, leading to repeated stuttering and intermodule transfer. This enables biosynthesis of several consecutive residues of Val or *N*-Me-Val, terminating in a phenylethylamine substituent. However, it is the docking interaction specifically which has been leveraged by ASI to generate chemodiversity in a new and unprecedented fashion. In the *vat* pathway, a sequence analysis of three NRPS modules, VatN, VatQ, and VatS, revealed three identical 45-amino acid αββαα-type docking domains at the N-terminus of each NRPS module. As VatA-P and VatS-VatX have predicted enzymatic roles that lead to the production of **1**-**4**, we hypothesized that VatQ and VatR also result in a lipopeptide product by transfer from VatM or VatP to VatQ-R, but was produced in lower abundance or ionized poorly, evading obvious detection. Thus, we decided to use a reactivity-guided approach to probe for the product of VatQ-VatR.

3.2.2 Reactivity guided isolation illuminates VatQ-VatR

There are many strategies for guiding the isolation of single bioactive natural products from crude fractionated extracts of cultured or environmental biomass. One of the most common is bioactivity-guided fractionation; the application of increasingly rarified extracts to cell lines or proteins with an analytical readout to indicate an effect on the targeted biological entity of interest. In this way, extracts can be refined to a single compound which is responsible for the activity of the extract.[202–204] Another strategy has been termed reactivity guided isolation. Broadly, this refers to the application of a reactive chemical substance to a crude extract to bind, activate, or in some way modify molecules containing a biologically or pharmacologically relevant functional group or moiety to make it more easily detectable or isolable (Figure 3.3).[205–207]

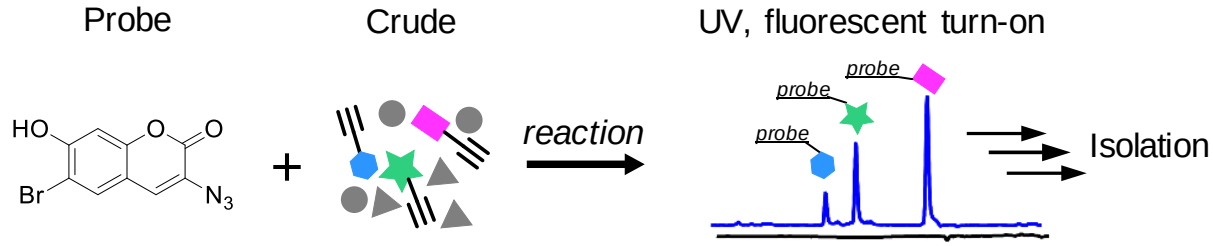


Figure 3.3. Depiction of reactivity-guided isolation scheme used to identify **5,6** from ASI.

A reactivity-guided approach was employed to search for the predicted product of *vatQ-vatR*. A coumarin-core probe containing an azide substituent to enable click-chemistry binding with free alkynes was synthesized in support of this goal. This approach allowed us to clearly identify an additional analog, vatiamide E, and its brominated analog vatiamide F, which enabled reconciliation of all pathway genes to their produced natural products. A closer look at the C-terminus of VatR revealed an NADPH-binding domain and homology to some amine-functionalizing enzymes. This domain is predicted to form the terminal amide on vatiamide E/F, although the mechanism for this transformation is not known.

3.2.3 Synopsis

We used a bioactivity and reactivity-guided approach to characterize six new lipopeptides, vatiamides A-F, and their biosynthetic pathway from ASI. Analysis of the pathway and its products

revealed a combinatorial strategy that employs a promiscuous β–αββαα docking domain interaction to produce a lipopeptide tail that is then connected to three distinct peptidic head groups. This is in marked contrast to the tightly selective nature of the curacin pathway, and all other known cyanobacterial PKS-NRPS pathways. Compared to previously characterized uses of non-collinear biosynthesis such as in stigmatellin, surugamide, and thalassospiramide production, the vatiamides pathway strategy is without precedent. The abstraction of a multimodular PKS product to three separate NRPS modules that then separately process a thioester-bound molecule until termination is unique in the NP biosynthetic literature to date. As such it represents a significant new mechanism by which prokaryotes may generate NP diversity.

3.3 Results

An environmentally-derived cyanobacterium with mat-forming thick reddish-brown filaments was hand-collected by snorkel in Vatia Bay, American Samoa, and designated with the collection code ASI16Jul14-2, referred to in this chapter as "ASI." After transport back to the laboratory, successive propagation and isolation of single filaments in seawater BG-11 media yielded a non-axenic single filamentous cyanobacterium by microscopic examination. Biomass was repeatedly propagated and scaled up over several months, and then harvested in successive batches. Harvests were pooled, extracted with 2:1 dichloromethane:methanol, dried, and then fractionated over silica gel, eluting with mixtures of solvents to make fractions A (nonpolar) to I (polar), as detailed in the experimental methods. Fractions were dried and submitted for bioassay on *Artemia salina* brine shrimp, as well as the H-460 NCI human lung cancer cell line. Significant mortality was found on brine shrimp in fractions C – E, and modest mortality rates were observed in fractions F – G; low toxicity against H-460 cells was observed in the D, E, H, and I fractions. (Figure 3.4).

3.3.1 Structure characterization of vatiamides A-D

Bioactivity from the *A. salina* and H-460 assays prompted a closer analysis of the major fraction peaks in active fractions D – H. Sub-fractionation and purification of the major compounds via semi-preparative C18-HPLC resulted in the isolation of four pure major compounds. 1D and 2D NMR, as well

as high-resolution mass spectrometry analysis, were employed to establish the planar structures (**1-4)** (Table A3.2-A3.8, Figures A3.4-3.20). As discussed previously, each molecule features an identical fatty-acid tail that possesses a terminal alkyne or bromoalkyne, a polyketide chain bearing a secondary methyl group and a vinyl chloride moiety (**7**/**8**), identical to that found in the jamaicamides (**9**/**10).**[103] This tail

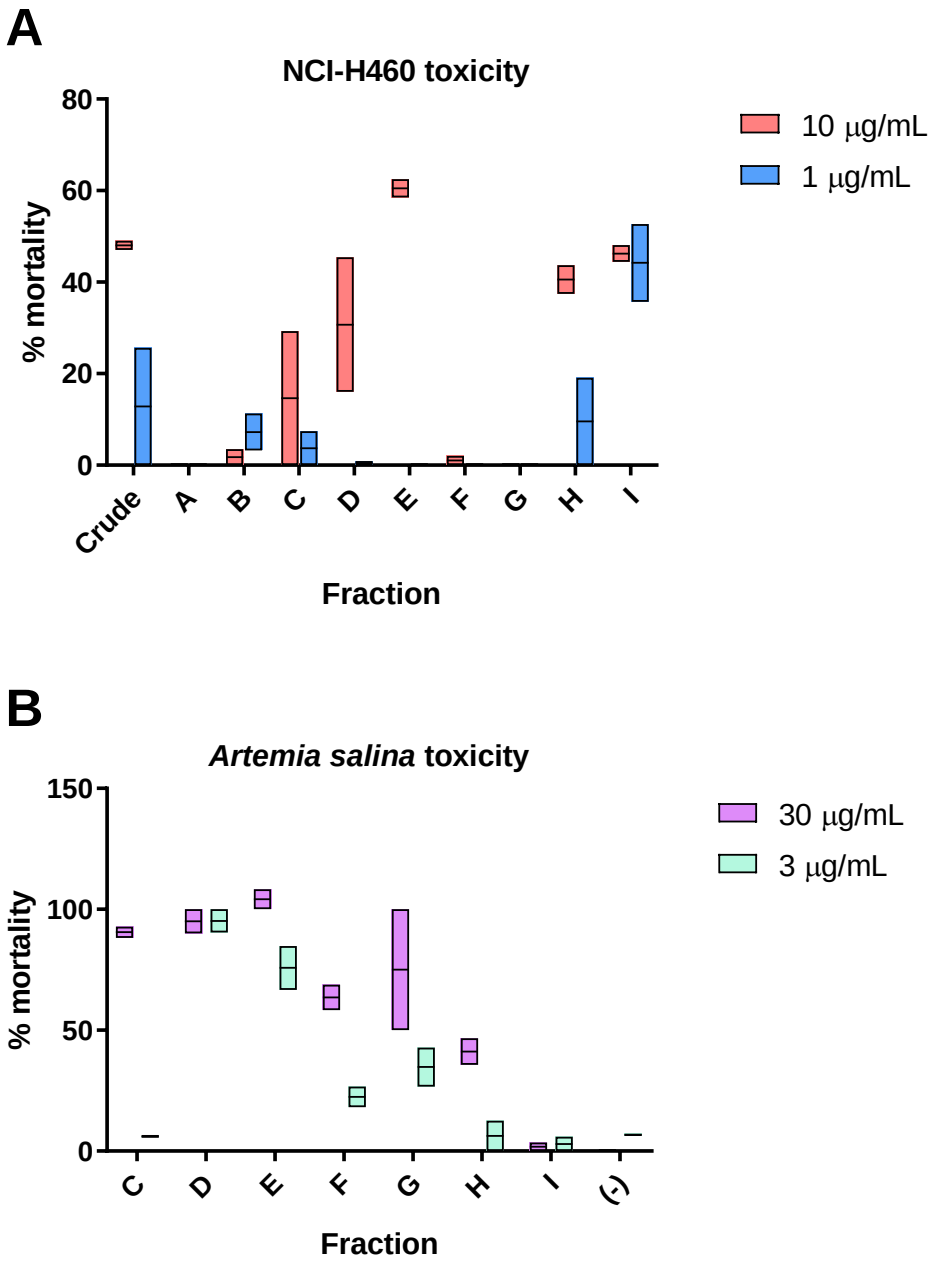


Figure 3.4. Bioassay of fractionated extracts. A) NCI-H-460 MTT fluorometric assay results for fractions A-I and crude. B) *A. salina* brine shrimp toxicity for fractions C – I. DMSO used as negative control.

was found to be extended with an *N,O,O*-trimethyl-L-tyrosine in **1/2.** However in **3/4,** it was appended by a glycine-derived amide and is then extended with a bicyclic C10 polyketide-derived structure. This latter portion of the molecule begins with a hemiketal at C16, with a hydroxy originally deriving from the glycine carbonyl, and an ether bond to C20, forming a pyran ring that bears a secondary methyl group at C17 and methoxy group at C18. C19 – C20 are formed from another acetate incorporation, with C20 carbonyl reduced to an alcohol by KR activity, which attacks C16 to form the pyran. In addition to the ether, C20 forms a branch via an aliphatic one-carbon bridge (C21) to C22, which is reduced by a KR to an alcohol, and extended via another acetate unit, C23 – C24. After further extension by one more acetate group, C25 – C26, the carbonyl at C24 is enolized, and by action of an *O*-MT forms another *O*-methyl group. The C26 carbonyl carbon is then nucleophilically attacked by the hydroxy of C22 to form the terminal unsaturated δ-lactone. Use of DMSO-d_6 was key in establishing the presence of the hemiketal and pyran ether, as in this solvent long range ^{1}H-^{13}C coupling of the OH proton on the C16 hemiketal was observed to the quaternary carbon C16, C15, and C17 by HMBC (Table A3.6). The chirality of the amino acid component of **1/2** was determined to be L (*S*) by Marfey's analysis compared to an FDAA-derived standard of *N,O,O*-Me-L-tyrosine (Figure A3.30). Stereocenters C16, 17, 18, 20, and 22 were assigned in **3/4** as S*,S*,R*,S*,R* by NOESY, molecular modeling, and *J*-based coupling analysis (Figures A3.15, A3.20-3.21, A3.32-3.33). Again, NMR analysis in DMSO-d_6 played a key role as it helped to establish relative 3-D structure by enabling the C16 hemiketal proton to interact via nOe with the C20 methine proton, indicating that the two protons were likely on the same face of the pyran ring.

3.3.2 Sequence comparison of VatK and C9 stereochemistry

The C9 stereochemistry of the fatty acid tail, common to all vatiamides, was assigned via sequence analysis of the VatK KR. The stereochemistry of α-methyl groups installed by SAM-utilizing C-MTs contained within PKS modules may be predicted by sequence analysis of the active KR.[175,208,209] Based on several key amino acid motifs within the KR domain, the KR may be assigned as either the A1, A2, B1, or B2 type, with each type conferring either *R* or *S* configuration to the α-carbon or the

β-hydroxy group after the PKS has installed the acetate unit and is still bound to the ACP module via a thioester. In particular, the presence of an LDD loop motif located in close physical proximity to the catalytic active site denotes A vs. B-type KR activity, while the absence or presence of a proline which modifies the activity of the active site tyrosine denotes 1 vs. 2-type KR activity.[175] In the VatK KR, the LDD loop is present, but modified to LSD, while the active site lacks a proline (Figure 3.5A). A similar LSD motif is found in several other active KR domains in PKS modules, with that from CrpA in cryptophycin biosynthesis resulting in a B1-type stereochemistry and *R* assignment of the methyl-

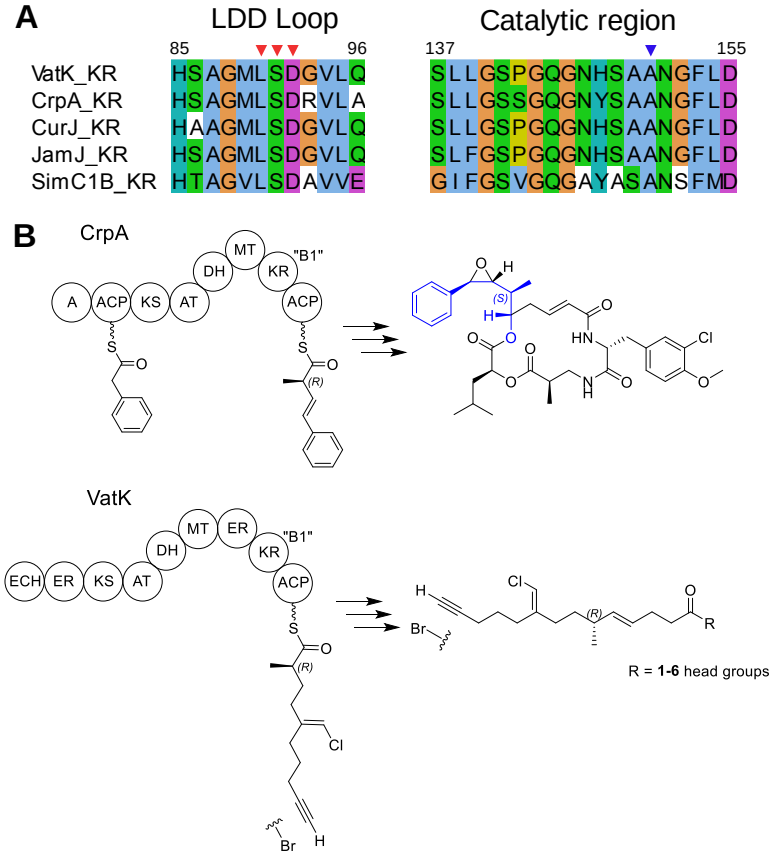


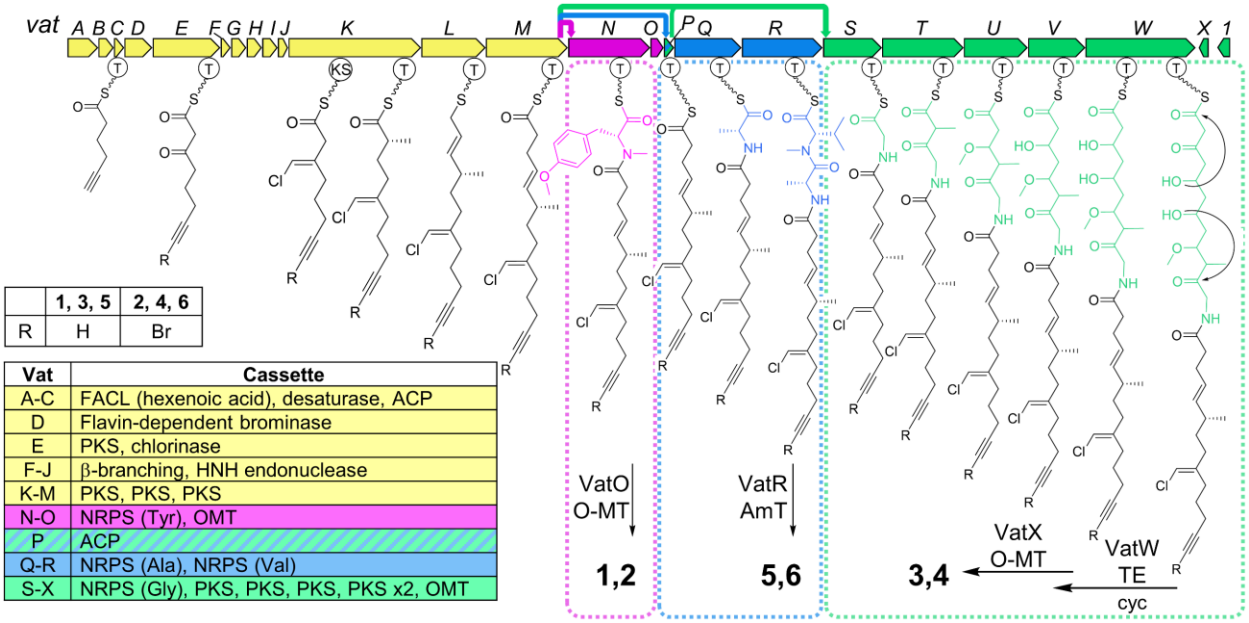
Figure 3.5. VatK KR analysis assigns C9 stereochemisty. Red triangles denote LDD loop, blue triangle indicates "P" location within active site A) Alignment of LDD loop and catalytic region of VatK with close homologs with D to S substitution. Red triangles indicate LDD motif, dictating A vs. B-type KR, and blue triangle indicates "P" position, dictating B1 (No P) vs. B2 (+P) type KR confirmation. Pathway and MiBIG codes as follows: CrpA: cryptophycin, BGC0000975.1, *Nostoc* sp. ATCC 53789; CurJ: curacin, BGC0000976.1, *Moorea producens* 3L; JamJ: jamaicamide, BGC0001001.1, *Moorea producens* JHB; SimC1B: simocyclinone, BGC0001072.1, *Streptomyces antibioticus*. B) CrpA activity and cryptophycin (top), VatK activity and vatiamide fatty acid tail **7**/**8** (bottom).[210]

branching carbon on the elongating molecular chain, and *S* stereochemistry in the final molecule owing to the installation of an epoxide proximal to the branching methyl. Thus, owing to the similar polarity and bulk of LSD vs. LDD as a tripeptide and validated by the presence of a B1-type KR in CrpA, we predict that VatK contains a B1-type KR, conferring *R* stereochemistry to C9 of the final biosynthesized molecules **1-6** (Figure 3.5B).

3.3.3 Genome sequencing and pathway analysis

The unusual nature of the bifurcated head group structure of **1/2** vs. **3/4** prompted a bioinformatic analysis of the ASI genome. Genomic DNA was extracted from cultures and submitted to next-generation sequencing using paired-end 250 bp reads via the Illumina HiSeq ® platform. By 16S rRNA sequence analysis and morphology characteristics, the organism was identified as *Moorea producens* (Figure A3.1, Table A3.10).[182,211] An assembly derived from short-read sequence data enabled assignment of the biosynthesis of **1/2**, but not **3/4** due to the fragmented nature of the assembly, as high coverage at the contig termini prevented assembly due to similar DNA sequences of the pathway adenylation domains. However, the **1/2** biosynthetic pathway was evident in a 50 kb PKS/NRPS BGC *vatA-M* which contained the biosynthetic machinery homologous to that of the chloro-fatty-acid moiety of **9**/**10** biosynthesis, but diverging at VatN, an NRPS module with an adenylation (A) domain predicted to activate a tyrosine residue. This is followed by an in-module *N*-methyltransferase (MT) and *O*-MT, peptidyl carrier protein (PCP), thioesterase (TE), and downstream standalone *O*-MT enzyme, VatO, in accordance with predicted biosynthesis of **1/2**. No biosynthetic genes in an orientation or domain structure matching that of **3/4** were found, although the assembly contained several BGC fragments and contigs containing partial PKS and NRPS modules. In order to investigate this further, additional high-quality gDNA was extracted, and submitted for sequencing using Pacific Biosciences (PacBio) technology, which generates longer DNA molecule reads of 5 kb – 20 kb. These longer reads increase the ability of assembler programs to correctly piece together stretches of DNA which contain high homology, such as adenylation domains and repeat regions. This facilitated a hybrid assembly with the previously obtained HiSeq-derived reads. Submission of the resulting assembly to AntiSMASH illuminated the complete 90 kb *vat* pathway, with the original

50 kb pathway extended by 40 kb of NRPS and PKS modules immediately downstream of VatO (Figure 3.6, Table A3.9).[132] Continuing from VatO, VatP is a standalone ACP with high homology to the ACP of VatM, followed by VatQ, an NRPS module predicted to encode for a condensation domain (C), A(Ala), a PCP and an epimerase domain (E). Module VatR is an NRPS domain with C, A(Val), an *N*-MT, PCP, and terminal domain bearing homology to an NADPH-binding motif and putative aminotransferase of



	1, 3, 5	2, 4, 6
R	H	Br

Vat	Cassette
A-C	FACL (hexenoic acid), desaturase, ACP
D	Flavin-dependent brominase
E	PKS, chlorinase
F-J	β-branching, HNH endonuclease
K-M	PKS, PKS, PKS
N-O	NRPS (Tyr), OMT
P	ACP
Q-R	NRPS (Ala), NRPS (Val)
S-X	NRPS (Gly), PKS, PKS, PKS, PKS x2, OMT

Figure 3.6. Vatiamide biosynthetic pathway *vat* in *Moorea producens* ASI16Jul14-2, "ASI". Biosynthetic genes for **7/8** highlighted in yellow. Three VatM/VatP cognate NRPS module partners and their subsequent PKS or NRPS cassettes highlighted as follows: pink **1/2**, green **3/4**, and blue **5/6**, with colored arrows at top indicating transmittal of **7**/**8** to VatN, VatP, and VatS, respectively. ACP; acyl carrier protein, AmT; amidotransferase, FACL; fatty-acid CoA ligase, OMT; O-methyltransferase, T; thiolase, TE; thioesterase (Table A3.9).

uncharacterized function. COM-type dds mediate interaction of VatQ/VatR. Downstream of VatR, the VatS NRPS module contains a C, A(Gly), PCP, and is then followed by five PKS modules, VatT-W. In addition to the required KS, AT, and ACP in type I PKS modules, VatT contains a DH, *C*-MT, and KR, though the KR is predicted to be an inactive KR^0. This KR^0 plays a pivotal role in generating the pyran ring and hemiketal functionality: as the KR^0 in VatT would leave a carbonyl intact from the VatS-incorporated glycine, it provides a suitable electrophile for attack from the subsequently derived C20 hydroxy group. VatU contains an active KR, in addition to an *O*-MT, which together are responsible for

the creation of the C18 methoxy group. VatV contains only a KR, which produces the C20 hydroxy which later attacks the remnant C16 to form the pyran/hemiketal, while VatW contains two PKS modules: one contains a KR used to form the terminal δ-lactone, and the second module has no oxidoreductive modifiers, resulting in a carbonyl which undergoes attack from C20. A thioesterase (TE) embedded in VatW is believed to participate in one or two intramolecular cyclizations to create the bicyclic ring structure observed in **3/4** by facilitating esterase activity on the terminal thioester C26 as well as the VatS glycine-derived carbonyl C16, left unreduced due to the VatT KR^0. However, an additional uncharacterized gene, Orf1, possesses homology by BLAST to various hydrolases including cholylglycine hydrolase, penicillin V acylase, and C-N amide hydrolase, all of which catalyze hydrolytic attack of a carbonyl; thus, Orf1 may be responsible for one or more of the heterocyclization events. As C26 and C24 each bear carbonyls, this significantly lowers the pKa of the C25 protons. This primes the ability of VatX, an *O*-MT, to methylate the C24 carbonyl; deprotonation of C25 forms an enol anion which attacks the methyl group of SAM. Owing to the exactly synchronous domain-to-structure accord between predicted and observed functionalities, we propose a non-collinear transfer of **7/8** from VatM to VatS, where it is then processed by VatT-VatW to form compounds **3/4**.

3.3.4 *Vat* docking domain promiscuity

An alignment of the intermodule docking motifs revealed that the first 45 amino acids of the N-terminal NRPS docking motif (Ndd) on VatN, VatQ, and VatS were 100% identical, and that the final six residues of the C-terminal dd (Cdd) on VatM and VatP were identical as well, closely matching previously characterized β–αββαα type dds (Figure 3.7).[114,116] Generally, as residues 24-29 on the β_2

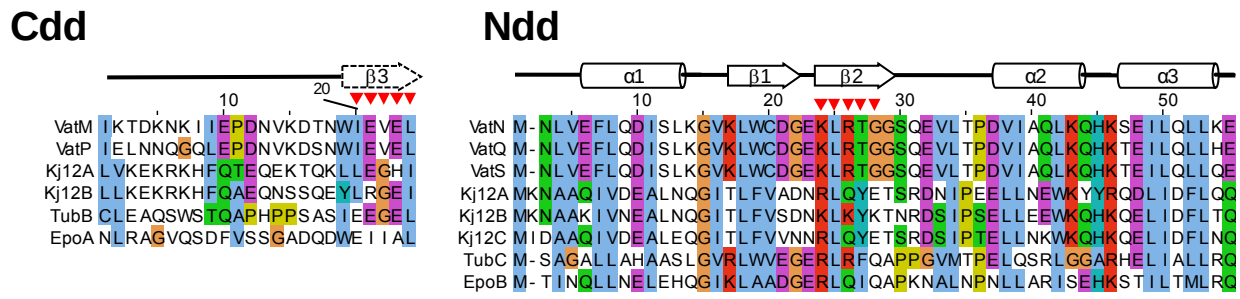


Figure 3.7. β–αββαα docking domain alignment from *vat* pathway. C-terminal docking domains VatM and VatP (left) and N-terminal αββαα docking domains (right) aligned with previously established docking domains from KJ12 RXPs, tubulysin, and epothilone pathways. Depiction of tertiary structure above each alignment. Red triangles indicate key specificity code interaction residues.

sheet of the N-terminal NRPS dd confer interaction with the terminal 5-6 residues of the "β_3" sheet of the upstream thiolase C-terminal dd, we theorized that there is no selective preference by the upstream C-terminal VatM/P dds to downstream N-terminal VatQ/N/S dds. Thus, VatM or VatP thioester-bound **7/8** can interact with the VatN, VatQ, or VatS Ndd stochastically. This was validated by the observed product composition wherein **7**/**8** appear in **1-4** but the 'head' group of **1/2** (VatM – VatN transfer) differs from that of **3/4** (VatM – VatS transfer). However, the biosynthetic assignment of all genes except for *vatQ* and *vatR*, for which there was no identified product, resulting in a conundrum.

3.3.5 Reactivity guided detection and isolation

To sensitively probe for these predicted products from *vatQ* and *vatR*, we employed a reactivity-guided approach in order to more easily discern any additional alkynes that may be transferred to the products of VatQ-VatR.[212,213] To facilitate this approach, we utilized a fluorogenic "click"-based probe **12,** which undergoes a copper-catalyzed azide-alkyne cycloaddition (CuAAC) with terminal alkynes (Figure 3.8A).[214–216] Incorporation of a bromine atom on the coumarin ring endows the probe with a characteristic isotopic pattern, aiding in identification of tagged compounds during LC/MS analysis of probed extracts.[217] In addition, the probe exhibits "turn-on" fluorescence at 490 nm upon formation of a triazole adduct, enabling benchtop screening for terminal alkyne-bearing NPs in crude extracts. Lastly, the conjugated coumarin core of the probe significantly increases absorption at the UV range of 350 – 400

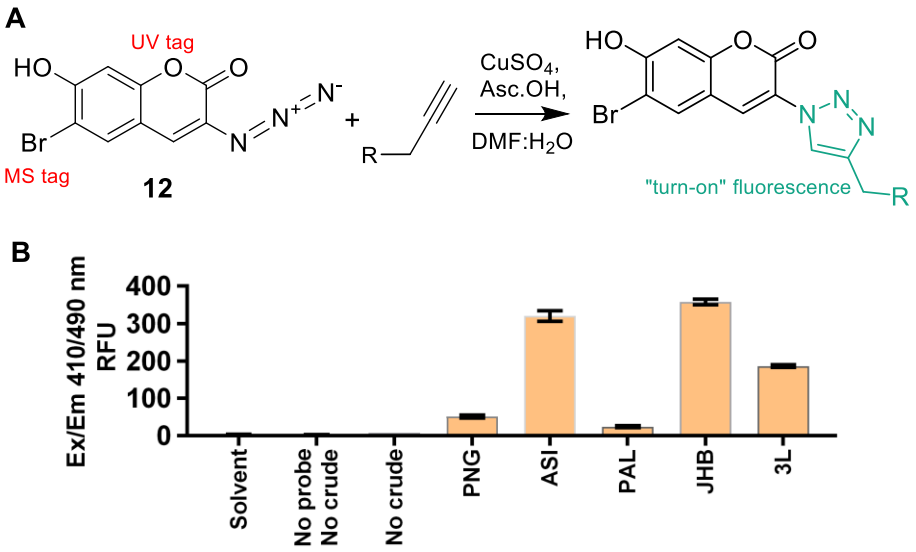


Figure 3.8. Probe mechanism and *Moorea* screening. A) Coumarin-based click-reactive probe **12** undergoes a copper-catalyzed "click" reaction with alkynes to form a fluorescent triazole adduct product. B) Application of **12** to crude extracts of *Moorea* indicate fluorescence increase upon binding alkynes in ASI (vatiamides **1,3,5**), JHB (jamaicamide **9**), and 3L (carmabin **11)**. PNG and PAL show minimal binding.

nm, making detection during separation much easier, as absorptions are generated in a frequently sparse UV range. Proof of concept for this aim was demonstrated by reaction of **12** with purified **1** and **3,** followed by HR-ESI-MS (Figure A3.34-3.35). Probing the extracts of the *Moorea producens* strains ASI, JHB (**9-10** producer),[103] 3L (**11** producer),[105,218] PAL15Aug08-1 (no known alkynes), and *Moorea bouillonii* PNG5-198 (no known alkynes) clearly indicated the presence of alkynes in ASI, JHB, and 3L but not PNG5-198 or PAL15Aug08-1, which displayed baseline fluorescence (Figure 3.8B).

This experiment was followed by application of **12** to crude extracts of ASI in order to determine if an additional alkynylated product was being biosynthesized by utilizing VatQ-VatR. Analysis of crude ASI-**12** reaction mixture by HPLC-UV-HRESI-MS/MS showed five major peaks by UV at 350 nm with significantly increased signal intensity when compared to un-probed crude ASI (Figure 3.9). These peaks, by *m/z* values, were found to be **12**, **1+12**, **3+12**, **7+12**, and lastly an as-yet unidentified sixth peak (**5+12**) with an $[M+H]^+$ calculated to be 466 when probe mass of 281 amu is subtracted from the HR-ESI-$[M+H]^+$ *m/z* 747 (Figure A3.36). Discovery of this additional peak prompted a re-examination of

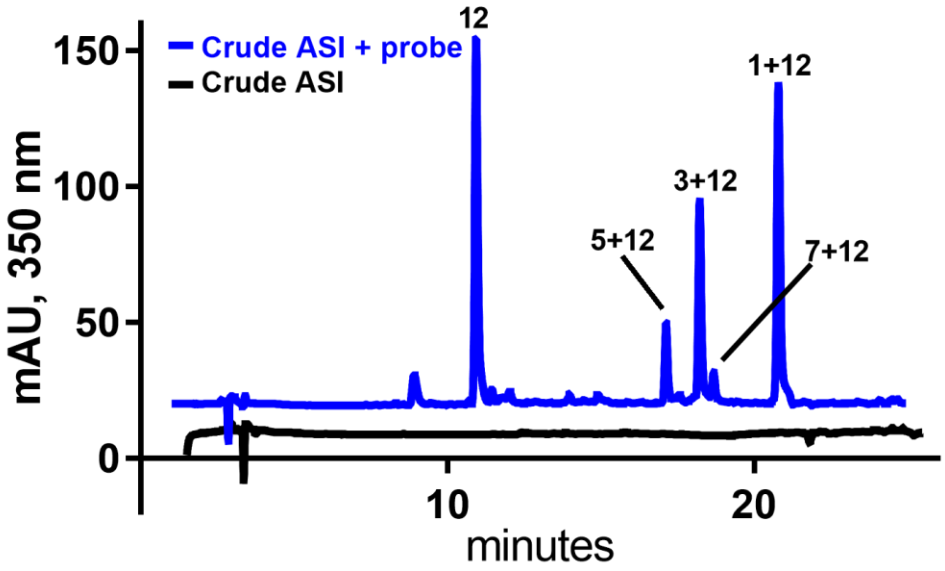


Figure 3. 9. HPLC-UV of ASI crude extract with (blue) and without addition of probe **12** (black). Probe plus alkyne adduct peaks indicated above each peak.

the un-probed H and I fractions from the ASI extract. This survey revealed a minor, near-baseline peak with a principal ion $[M+Na]^+$ *m/z* 488, equating to the $[M+H]^+$ *m/z* 466 indicated by probe **12** binding in the ASI crude. Additionally, a peak with $[M+H]^+$ *m/z* 544, $[M+Na]^+$ *m/z* 566 by LC/MS was found, suggesting a potentially brominated analog of $[M+H]^+$ *m/z* 466. This brominated analog was not evident in initial crude MS/MS runs due to its low abundance and poor ionization compared to **1-4**.

3.3.6 Isolation and structural analysis of vatiamides E and F

Extraction, fractionation, and separation of additional biomass yielded sufficient material to characterize $[M+H]^+$ *m/z* 544 (**6**), but only trace quantities of $[M+H]^+$ *m/z* 466 (**5**). 1D and 2D NMR revealed a partial structure **8** with an amide linkage to an alanyl residue followed by an *N*-methyl-valinamide residue, defining compound **6**. By HR-ESI-MS, the non-brominated analog **5** was also observed. Elucidation of the structure of **6** proved non-trivial, as it is poorly soluble in both methanol and chloroform. It is soluble in DMSO, however, but analysis of **6** in DMSO-d_6 rendered two conformers, present in a 2:1 ratio, despite the appearance of a single peak in the HPLC-UV purification procedure. The conformer is likely due to hindered rotation about the C22 *N*-Me amide, locking the molecule into two structural states and thus altering the chemical shifts of individual atoms as observed by ^{1}H-NMR and

^{13}C-NMR. Compounding this, the terminal amide shows split resonances for each proton, likely due to H-bonding with either the terminal or alanyl carbonyl.

To indicate that compound **6** was indeed a pure single entity, two strategies were employed: first, NMR analysis in d_3-acetonitrile was used to alter the conformer population about the amide, and second, elevated temperature NMR was employed to increase the rate of interconversion between the two conformers such that a single time-averaged molecule could be observed. The conformer ratio was altered from 2:1 to 8:1 by ^{1}H-NMR analysis in d_3-acetonitrile, which increased confidence in the presence of a single entity and enabled confirmation of resonances and coupling in DMSO-d_6 (Figure A3.28). Elevated temperature ^{1}H-NMR in DMSO-d_6 resolved the split terminal NH_2 resonance, as rapid interconversion between the conformers coalesced their two resonances into a single peak, (Figure A3.29), again suggesting that this preparation of compound **6** was a single pure compound. Marfey's analysis of **6** indicated D-alanine and L-*N*-Me-valinamide, thus defining stereocenters 15*R* and 17*S* (Figure A3.31). The presence of the epimerase in VatQ is therefore validated by the D-alanyl moiety in **6.** Thus, the presence of **5** and **6** reconciles the presence and function of VatQ-VatR, lending evidence to the notion that **7/8** are stochastically dispensed from VatM and VatP to VatN, VatR, and VatS, in sharp contrast to collinear PKS-NRPS biosynthesis.

3.3.7 Terminal amide incorporation to vatiamides E-F

We theorize that the terminal amide functionality in compounds **5** and **6** is incorporated by the undescribed domain embedded in VatR, downstream of the PCP. A homologous domain is also found at the C-terminus of the carmabin A (**11)** biosynthetic pathway in *Moorea producens* 3L; carmabin A aslo contains a terminal amide functionality.[187,218] A terminal amide is also observed in dragomabin, as well as some analogs of dolastatin; however, the biosynthetic pathway for these later molecules has not yet been described. We suspect it likely that the terminal amide in these metabolites is installed by a similar domain. This domain contains an NADPH binding pocket, and bears homology to other amine-functionalizing enzymes such as PigE, an aminotransferase in *Serratia* bacteria, and ornithine cyclodeaminases, as well as the 2,3-diaminopropionate biosynthesis protein SbnB in bacteria. This latter

protein is frequently paired with or fused to SbnA, a PLP-binding aminotransferase domain with a well-characterized mechanism that facilitates amino transfer. A BLAST search of SbnA versus the hybrid ASI assembly in search of a biosynthetic pair produced five hits with significant homology to the query; however, none of them were in close genetic proximity to the *vat* pathway, or even on the same contig. Therefore, we speculate that the amino transfer activity to the terminal carbonyl of **6** may be mediated by a mechanism that pairs the VatR amino functionalization SbnB homolog with one of the non-pathway SbnA homologs in the ASI genome.

3.4 Discussion

The biosynthesis of **1-6** by abstraction of a PKS-derived modified fatty acid starter unit onto separate NRPS cognate partners to form three different head groups is a unique and heretofore uncharacterized method of generating chemodiversity with type I PKS/NRPS systems, depicted in Figure 3.10. In synthetic and semi-synthetic chemistry, researchers aim to take a common pharmacophore and modify it with different functional groups in order to expand or modulate its biological activity. Here, we reveal

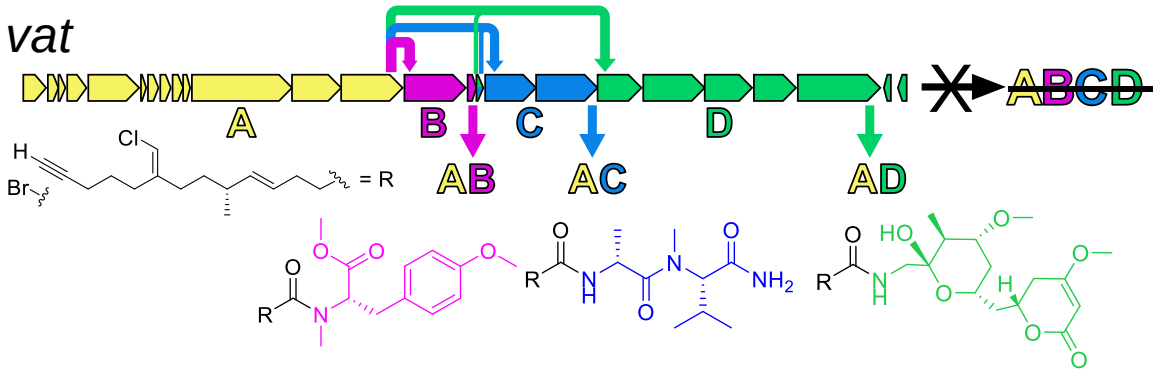


Figure 3.10. Depiction of the molecular assembly of the vatiamide pathway. PKS/NRPS modules, in atypical fashion, shunt a modified fatty acid from PKS module A onto cassettes B, C, and D separately to form molecules depicted as AB, AC, and AD, as opposed to the expected representative molecule from a collinear ABCD biosynthetic assembly line.

that nature has already selected for a version of native, *in vivo* combinatorial biosynthesis. (Figure 3.11). We postulate that this interaction is mediated by *N*-terminal NRPS docking domains with identical sequence and structure, enabling non-selective interaction with VatM and VatP ACP Cdd.

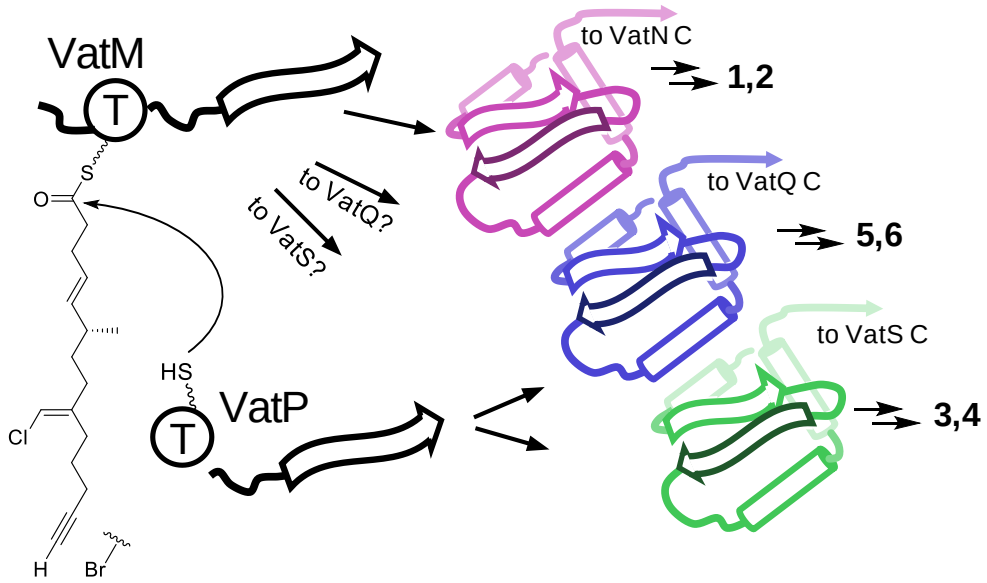


Figure 3.11. Depiction of stochastic interaction between upstream VatM and VatP Cdd-bearing modules and downstream VatN, VatQ, and VatS NRPS modules. T; thiolase, C; condensation domain.

The plethora of collinear pathways that produce a single product is evidence that significant evolutionary pressure has been placed on eliminating promiscuity; that is, ensuring that biosynthetic modules interact with their cognate dd partner through unique dd sequences and thus produce a particular product with high fidelity.[112,200] One way to test this concept is to broadly search for pathways which also contain two or more identical docking domains; if a single molecule is produced from such a pathway, this would present a challenge to our hypothesis that identical dds foster combinatorial interaction. In order to survey known PKS/NRPS pathways for potential avenues to dd promiscuity, a BLAST search of the MIBiG pathway database was undertaken.[124] Using the VatM/P Cdd and VatN/R/S Ndd to query the MIBiG database revealed seven additional distinct pathways with at least two β–αββαα dds between separate modules; apratoxin A was added via NCBI-derived sequence data (Figure 3.12). However, none of the pathways contain greater than one set of dds that are 100% homologous. One of the interesting findings is that the dd is not limited to just PKS or NRPS modules: in aeruginosin biosynthesis, for example, a standalone halogenase AerJ in aeruginosin biosynthesis has both a β–sheet Cdd interacting with subsequent protein AerB, and a β–αββαα–Ndd interacting with AerA selectively.[124] This observation further bolsters the case for the uniqueness of the *vat* pathway in ASI.

However, biosynthesis using PKS and NRPS modules are not explicitly bound by collinearity rules, as discussed in the introduction chapter of this dissertation. Alternate strategies for biosynthesis in a PKS/NRPS context exist, including stuttering, halting, module skipping and bidirectionality.[103,122,123,219,220] However, we report here the first characterized example to date of a reversal of selectivity in a PKS/NRPS context based on identical docking domains. The promiscuity enabled by these domains infuses a new element into the combinatorial possibilities of both heterologous and native pathways; there could be numerous combinations of genes given the diversity of horizontal and vertical branching within the microbial kingdom.

One may speculate that the origin of this combinatorial pathway in ASI is synergistic defensive cytotoxicity imparted by the combination of more than one of the three head groups, or separate environmental targets for each. To test for this, we attempted to observe synergistic cytotoxicity upon mixing together various of the vatiamides in the H-460 lung cancer assay. However, to date we have only observed modest H-460 human lung cancer cytotoxicity with both pure compounds and combinations of compounds, with the brominated analogs **2** and **4** being slightly more cytotoxic than the non-brominated versions **1** and **3** (Figure A3.2-A3.3). Further bioassay should undoubtedly be undertaken to further understand the environmental and evolutionary roles played by these compounds in providing a selective advantage for ASI.

Perhaps one of the more interesting questions surrounding this system arises in understanding how such a protein structure is assembled over evolutionary time: either the docking motif was replicated and inserted in front of three pre-existing NRPS modules, or a single NRPS module with this dd was duplicated twice and each evolved different amino acid specificities followed by further tailoring divergence. There are numerous transposases located upstream and downstream of the main pathway genes, and there is an endonuclease of unknown function within the β-branching cassette, which may play a role in pathway recombination to assemble this arrangement of genes. However, VatM ACP and VatP homology suggests that gene duplication was responsible for VatP.

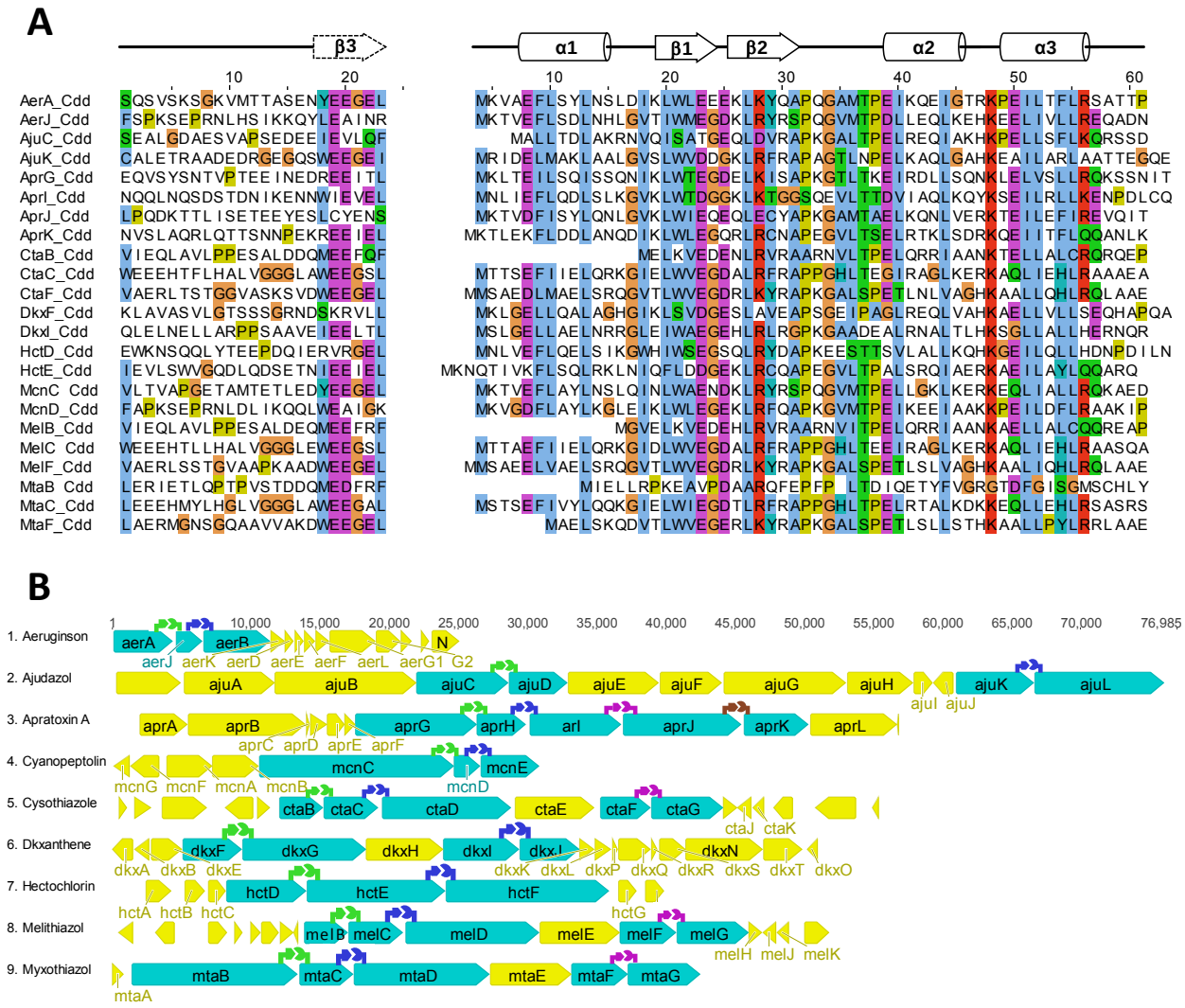


Figure 3.12. Docking domain alignments and selective PKS/NRPS pathways from MIBiG and NCBI. Pathways contain >1 Cdd-Ndd interactions of the β–αββαα type within a single biosynthetic gene cluster. A) Alignment of C-terminal docking domains on left and N-terminal docking domains on right, with tertiary protein structure graphics above each alignment. B) Graphic of each pathway PKS/NRPS core, with docking domain cognate pairs depicted by colored green (first), blue (second), purple (third), or brown (fourth) cartoon above interacting genes. Separate interactions are indicated by different colors. Pathway and organism, and MIBiG/NCBI identifier is as follows: aeruginson – *Microcystis aeruginosa* NIES-98, BGC0000298; ajudazol – *Chondromyces crocatus*, BGC0000954; apratoxin A – *Moorea bouillonii* PNG5-198, NCBI: txid568701; cyanopeptolin – *Microcystis* sp. NIVA-CYA 172/5, BGC0000332; cystothiazole – *Cystobacter fuscus*, BGC0000982; dkxanthene – *Stigmatella aurantiaca* DW4/3-1, BGC0000986; hectochlorin – *Moorea producens* JHB, BGC000100; melithiazol – *Melittangium lichenicola*, BGC0001010; myxothiazol – *Stigmatella aurantiaca* DW4/3-1, BGC0001024.

Lastly, discovery of this full pathway only occurred after employing long-read and short-read sequencing with hybrid assembly. Long-reads alone would likely have had too many errors to correctly

analyze these natural product pathways, while short reads alone originally produced a stunted pathway that left several modules as orphans. This is because assembler graphs may not be able to reconcile three nearly identical 150-bp stretches in a microbial genome without breaking continuity and generating "orphan" clusters. As many microbial sequencing efforts have been performed with short read sequencing only, it is possible that there are many more orphan gene clusters which were not correctly assembled into full pathways, combinatorial or otherwise. Therefore, we predict multichannel combinatorial biosynthesis may be more widespread than currently recognized and may be present in numerous different cyanobacterial and bacterial clades.

3.5 Experimental methods

3.5.1 Organism collection and culturing

A thumb-sized agglomeration of mat-forming cyanobacterial filaments was collected by hand on shallow rock substrate environs at a depth of 1-2 m in Vatia Bay, American Samoa, coordinates 14°14'45.8"S 170°40'24.3"W and given the environmental collection code ASI16Jul14-2. The filaments were transferred to breathable 30 mL culture flasks with local seawater until transport to laboratory. Repeated culturing of this environmental collection of >1 cyanobacterial species saw emergence of a single large diameter (60-100 μM) filamentous cyanobacterium with thick rust-colored sheath and tightly stacked cells, matching the morphology of the genus *Moorea*.[211] Single filaments of this strain were separated, drawn through agar to isolate the unicyanobacterial filament, and propagated in modified seawater BG-11 media containing a final concentration of 0.5 g/L $NaNO_3$.[182] The organism was grown at 27-28°C in 16h/8h light/dark cycle under 5.5-10 μmol photons/m^2S and split for subculturing every 3-5 weeks. Repeated culturing and iterative scaleup of this species rendered enough biomass for chemical and genomic DNA extraction. Microscopy was performed on an Evos XL light microscope (Life Technologies) with 10X, 20X, and 40X objectives.

3.5.2 DNA extraction and sequencing

A loose clump of *Moorea* sp. ASI16Jul14-2 filaments approximately 2 cm in diameter was transferred to a Büchner funnel with filter paper and washed with DI water, then immediately transferred to mortar and pestle and pulverized under liquid nitrogen to a fine frozen powder. This powder was immediately transferred to a cold 2.0 mL Eppendorf tube on ice and immediately processed using the Qiagen bacterial gDNA isolation protocol (Qiagen), followed by cleanup using a G20 tip (Qiagen). Eluent was precipitated with isopropanol and then washed with EtOH, followed by drying in a laminar flow hood and reconstitution in 200 μL of DNAse free water.

For Illumina HiSeq sequencing, resulting gDNA was processed using a miniaturized version of the Kapa HyperPlus Illumina-compatible library prep kit (Kapa Biosystems), used for library generation.

DNA extracts were normalized to 5 ng total input per sample in an Echo 550 acoustic liquid handling robot (Labcyte Inc). A Mosquito HTS liquid-handling robot (TTP Labtech Inc was used for 1/10 scale enzymatic fragmentation, end-repair, and adapter-ligation reactions carried out using). Sequencing adapters were based on the iTru protocol,[221] in which short universal adapter stubs are ligated first and then sample-specific barcoded sequences added in a subsequent PCR step. Amplified and barcoded libraries were then quantified by the PicoGreen assay and pooled in approximately equimolar ratios before being sequenced on an Illumina HiSeq 4000 instrument to >30X coverage.

For Pacific Biosciences long-read sequencing, prior to submission, gDNA was assessed for quality and quantity using an Agilent 4200 Tapestation with Genomic Tape and a Qubit (Thermo Fischer Scientific). Sequencing libraries were generated using SMRTbell Template Preparation Reagent Kits (Pacific Biosciences) following the 20 Kb library protocol. Libraries were size selected to >6 kb using a PippinHT (Sage Sciences). Libraries were sequenced on a PacBio RS II sequencer (UCSD IGM Genomics Center, La Jolla, CA) via 4-hour movies using the DNA/Polymerase Binding Kit Version P6 V2 with C4 sequencing chemistry.

3.5.3 Assembly and pathway analysis

Initial Illumina HiSeq short-read output data was assembled via SPAdes 3.6.[184] Darkhorse was used to bin out contigs containing non-cyanobacterial genes,[185] followed by dissolution of contigs and reassembly of those containing only cyanobacterial genes. Submission to AntiSMASH revealed biosynthetic gene clusters and domain analysis, while NCBI DELTA PBLAST was used to further analyze individual modules for their domain composition, including the initial truncated **1/2** PKS/NRPS gene cluster.[128] Hybrid assembly of the short and long-read data was performed by SPAdes 3.6 hybrid assembler tool and analyzed similarly via bioinformatics tools, yielding longer and more complete contigs. The putative VatR C-terminal aminotransferase domain as well as Orf1 were analyzed using Phyre2 protein prediction server.[222]

3.5.4 Chemical extraction, fractionation, sub-fractionation and HPLC purification

Approximately 80 mL of packed wet biomass was collected over several growth harvests, rinsed with DI water over a Büchner funnel with glass fiber filter, and stored at -20°C until extraction. Biomass was extracted using successive rounds of 2:1 methylene chloride:methanol which were combined and dried by rotary evaporator. Crude extract was separated by vacuum liquid chromatography (VLC) consisting of hexane-packed TLC-grade (H) silica (Sigma-Aldrich) over a ceramic filter funnel. Fractions were eluted using gentle vacuum, from nonpolar to polar, in nine fractions ("A-I") as follows: 100% hexane, 9:1 hexane:EtOAc, 8:2 hexane:EtOAc, 6:4 hexane:EtOAc, 4:6 hexane:EtOAc, 2:8 hexane:EtOAc, 100% EtOAc, 7:3 EtOac:MeOH, 100% MeOH, which were dried via rotovap.

Prior to individual compound purification, fractions were subfractionated in 10% stepwise MeCN/H_2O reverse phrase gradients using 500 mg or 1 g Hypersep SPE cartridges (Agilent). Gradients and eluent % for each molecule pair are as follows: **1/2** 70 to 100% acetonitrile in H_2O, elution at 90-100% acetonitrile, **3/4** 50 to 100% acetonitrile in H_2O, elution at 70-90% acetonitrile, **5/6** 50 to 100% acetonitrile in H_2O, elution at 50-70% acetonitrile. Subfractions were analyzed for **1-6** content by LC/MS as detailed in subsequent section. Subfractions were dried via rotovap or N_2, lyophilization, reconstituted in MeCN/H_2O mixtures and submitted for semipreparative HPLC.

Purification of individual compounds from fractions was performed by semipreparative HPLC using a C18 Kinetex 5 μm 10x150 mm column using a reverse phase gradient of acetonitrile in H_2O, all solvents +0.1% (v/v) formic acid. HPLC gradient conditions are as below.

Table A3.1**:** Gradient conditions for semipreparative HPLC purification of **1-6**

Compounds	**1, 2**		**3, 4**		**5, 6**	
Fraction	D, E		G, H		H, I	
Flow rate (mL/min)	4.0		4.0		3.0	
UV monitor (nm)	209, 236, 276, 380		209, 228, 236, 276		209, 228, 254, 276	
	Time (min)	% MeCN	Time (min)	% MeCN	Time (min)	% MeCN
Gradient	0	80	0	65	0	80
	1	80	1	65	1	80
	5	99	16	99	22	99
	12	99	20.8	99	22.5	99
	12.5	80	21	99	29.4	80
	16.5	80	25	65	29.5	80

3.5.5 NMR, mass spectrometry, and polarimetry

^{13}CNMR and ^{1}HNMR was carried out on a Varian VX500 (500 MHz - ^{1}H, 125 MHz - ^{13}C). 2D experiments including HSQC, HMBC, COSY, NOESY, ROESY, and additional ^{1}HNMR experiments were carried out on a JEOL ECZ500 500 MHz system (500 MHz - ^{1}H, 125 MHz - ^{13}C). NOESY for **3** was also collected on a Varian NPA600 600 MHz system with 1.7 mm cryoprobe using 600 MHz for ^{1}H nuclei. All NMR measurements were taken at room temperature except for the analysis of **6**, which included a ^{1}HNMR comparison in DMSO-d_6 at 20°C, 40°C, 60°C,and 80°C. References were made in respective experiments to solvent signals $CHCl_3$ (δ_C = 77.16, δ_H = 7.24), DMSO (δ_C = 39.52, δ_H = 2.50), and acetonitrile (δ_C = 118.26, δ_H = 1.94). HSQC, NOESY, and ROESY data from the JEOL system was processed using Delta software (JEOL), while HMBC, COSY, and ^{13}CNMR and ^{1}HNMR data from both the NPA600, Varian VX500 and JEOL ECZ500 was processed in Mestrenova (Mestrelab).

Polarimetry values were obtained on a Jasco P-2000 polarimeter in a 1 dm chamber using a sodium lamp at 589 nm, at 25°C. Compounds were dissolved at 0.5 mg/mL **(2)**, 1.0 mg/mL **(1,4)** or 2.0 mg/mL **(3,6)** in methanol and 130 μL was used for each measurement. Ten readings were recorded and averaged for each compound, and subtracted from the mean of ten readings of a solvent blank for each sample.

Fractions, subfractions, and semipreparative HPLC flow-through fractions were analyzed via HPLC-ESI-MS/MS on a Finnigan Surveyor HPLC coupled to an LCQ Advantage Max ion trap mass spectrometer (Thermo Fisher Scientific) using data-dependent acquisition. High-resolution mass spectrometry data of **1-6** and **12+5** adduct was obtained on a 1200-series HPLC (Agilent) coupled to a 6530 qToF mass spectrometer (Agilent). Each system employed data dependent acquisition in positive ESI mode on a 4.6 mm x 100 mm Kinetex 5 μm C18 column with a gradient of 50% acetonitrile to 99% acetonitrile in H_2O using a 3:1 split flow rate of 0.6 mL/min (0.15 mL/min to the MS source) on the LCQ system and 0.7 mL/min (unsplit) on the Agilent system. All solvents contained 0.1% (v/v) formic acid.

High-resolution data of **12+1** and **12+3** adducts were obtained via direct infusion into a LTQ Orbitrap XL mass spectrometer in negative ESI mode with Orbitrap Fourier Transform mass analyzer (Thermo Fisher Scientific).

3.5.6 Computational chemical modeling

The preferred conformation of C22 of **3** was analyzed by energy minimization using MM2 and MMFF94 force field algorithms within Chem3D (PerkinElmer).[223,224]

3.5.7 NCI H-460 MTT-stain bioassay and brine shrimp assay

Cells from human lung cancer cell line NCI-H-460 were added at 3.33 x 10^4 cells/mL to a 96-well plate in Roswell Park Memorial Institute (RPMI) 1640 media, containing fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cells were incubated for recovery O/N at 37°C in 5% CO_2 in 180 μL per well. Two replicates of fractions C-I were tested at 1 μg/mL and 10 μg/mL. Pure compounds were initially dissolved in DMSO at 1 mg/mL, and diluted to initial working concentrations as follows: 41 μM **(1)** – 2 dilution series of 3 replicates; 17.8 μM **(2)** – 3 dilution series of 2 replicates; 35 μM **(3)** – 2 dilution series of 3 replicates; 31 μM **(4)** – 2 dilution series of 3 replicates; 29 μM **(6)** – 2 dilution series of 3 replicates, 1 dilution series of two replicates. Synergistic initial working concentrations of combined pure compounds were as follows: 40 μM **(1+2) –** 1 dilution series of 3 replicates; 90 μM **(1+2+6) –** 2 dilution series of 3 replicates; 72 μM **(2+4+6) –** 2 dilution series of 3 replicates; 60 μM **(4+6) –** 1 dilution series of 3 replicates; 60 μM **(1+3) –** 1 dilution series of 3 replicates. Working solutions were made through serial dilution eight times by a factor of 0.3164 in RPMI 1640 media without FBS, with 20 μL added to each well. An equal volume of RPMI 1640 media without FBS was added to wells designated as negative controls for each plate. Plates were incubated for approximately 48 h before MTT staining. Doxorubicin was used as a positive control, while an equivalent volume of DMSO was used as negative control. Plates were read at 570 and 630 nm using a SpectraMax M2 microplate reader (Molecular Devices) to determine cell viability.[225]

Toxicity towards brine shrimp (*Artemia salina*) was assayed as per previous research.[204,226,227] Brine shrimp eggs were incubated to hatching in a dark chamber for 48 h, after which 20-40 live brine shrimp in ~0.25-0.5 mL artificial seawater were transferred to 4.5 mL of artificial seawater plus 3 or 30 μg/mL chemical extract fraction, initially dissolved in DMSO. Conditions were run in duplicate. After 24 h at 27-28 °C, the number of dead brine shrimp was counted. Acetone was used to kill the remaining brine shrimp, and % mortality was generated by dividing initial dead by final dead quantity of brine shrimp.

3.5.8 Marfey's analysis of vatiamide A and F

A standard of *N,O,O*-L-tyrosine was synthesized as previously described.[228,229] 0.5 mg of **1** was dissolved in 1 mL 6N HCl in a glass pear flask, sealed with parafilm, and incubated at 110°C O/N for 16 h. The flask was cooled and dried via rotovap. The contents were resuspended in 2 mL of 1M $NaHCO_3$, and 1 mL each was transferred to two separate 2.0 mL glass vials. Marfey's reagent L-FDAA (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) and D-FDAA were separately dissolved in acetone and added in 4:1 molar excess to the two separate vials, as well as to two separate vials of *N,O,O*-L-tyrosine dissolved in 1 mL 1M $NaHCO_3$. The vials were incubated at 40-50°C for one hour. 1N HCl was added dropwise until the solution was neutral by pH paper. The solution was diluted 10-fold in acetonitrile, and dispensed to LC/MS vial with syringe and needle through a nylon filter. A similar procedure was repeated for 0.5 mg of **6**, using only D-FDAA. D-FDAA was also reacted with standards of L-alanine and D-alanine, as well as standards of *N*-Me-L-valine and *N*-Me-DL-valine for retention time comparison. HPLC/MS/MS conditions for Marfey's analysis employed a 4.6 mm x 100 mm Kinetex 5 μm C18 column using a 3:1 split flow rate of 0.6 mL/min (0.15 mL/min to the MS source) attached to a Thermo LCQ Advantage Max ion trap mass spectrometer (Thermo Fisher Scientific), using a reverse phase gradient of acetonitrile in H_2O, each solvent with 0.1% formic acid (v/v). The gradient for analysis of **1** (% acetonitrile in H_2O): 0 min – 20%, 1 min – 20%, 46 min – 99%, 52 min – 99%, 52.5 min – 20%, 60 min – 20%. The gradient for

analysis of **6** (% acetonitrile in H_2O): 0 min – 20%, 1 min – 20%, 40 min – 50%, 45 min – 50%, 46 min – 99%, 50 min – 99%, 51 min – 20%, 60 min – 20%.

3.5.9 *Moorea* 16S rRNA comparison

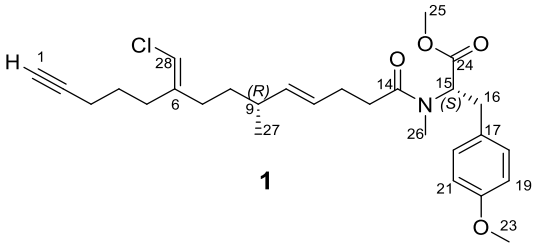
All alignments of docking domains, 16S rRNA sequences, and the VatK KR sequence comparison were performed using MUSCLE algorithm in Geneious (Biomatters).[190] Alignment images were prepared using Jalview.[189] Identity matrix reflects percent homology of nucleotide sequences spanning entire 16S rRNA sequence of selected *Moorea producens*.

3.5.10 Fluorescence measurements and imaging

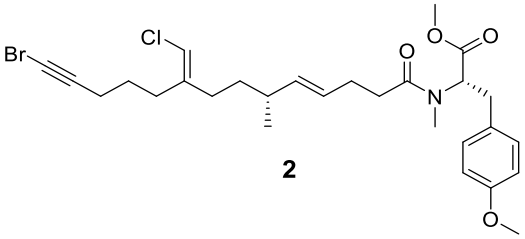
Small-scale *Moorea* crude extractions, without fractionation, were performed for fluorometric probe analysis. Solvent (2 x 10 mL of 2:1 DCM:MeOH) was sequentially applied to a thumb-sized clump of harvested and rinsed biomass of *Moorea producens* ASI16Jul14-2 (ASI), *Moorea producens* 3L (3L), *Moorea producens* PAL15Aug08-1 (PAL), *Moorea producens* JHB (JHB), and *Moorea bouillonii* PNG5-198 (PNG). Extracts were dried and resuspended to 10 mg/mL in DMF. To carry out the click reaction, the following solutions were prepared: 1.0 mg/mL **12** in DMF, 2.8 mg/mL $CuSO_4$ in H_2O, 6.2 mg/mL ascorbic acid in H_2O. In this order, 10 μL of each of the above solutions was added to an LC vial insert: $CuSO_4$, then asc.OH, then crude, then probe with mixing by pipette. A negative control containing no crude or azide was created, whereby 10 μL of DMF blank was used in place of each of the omitted reagents. An additional negative control containing no crude was prepared, whereby 10 μL of DMF blank was added in place of a crude. After 1 hour, these solutions were transferred to wells in a 96-well plate, and fluorescence was measured on a SPECTRAmax M2 (Molecular devices). Excitation was at 410 nm, emission was measured at 490 nm.

3.6 Appendix

Table A3.2. Vatiamide A (**1**) NMR spectral data (500 MHz, $CDCl_3$)

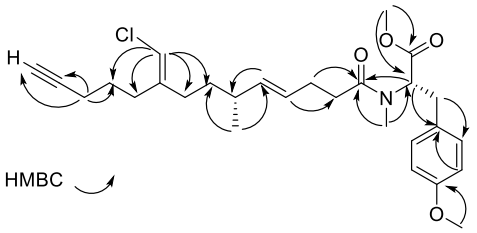


1



2

C#	δ_C	δ_H, *J* in Hz	type	HMBC	COSY	key NOE
1	68.4	1.95	s		3	
2	84.2					
3	18.1	2.18	m	1,2,4,5	1,4,5	
4	25.9	1.62	m	2,3,5,6	2,3,5	
5	29.2	2.26	m	3,4,6,7,28	3,4,28	
6	141.9					
7	32.5	1.98	m	6,8,9,28	8,9,28	28
8	34.7	1.28	m	6,7,9,10,27	7,9,27	28
9	36.3	1.98	m	7,10,11,27	8,10,11,27	
10	136.3	5.21	m	9,11,12	9,11,12,27	
11	127.7	5.28	m	10,12,13	9,10,12	
12	27.6	2.18	m	10,11,13,14	10,11,13	
13	33.4	2.28	m	10,11,12,14	11,12	
14	173					
15	58.2	5.25	m	14,16,17,26	16,25	
16	33.8	3.27 (5.5,14.8), 2.93 (11.3,14.5)	dd, dd	15,17,18/21,24	15,17, 18/21	
17	128.9					
18/22	129.7	7.05 (7.9)	d	16,17,19/21,20	16,19/22	
19/21	113.9	6.78 (8.0)	d	17,18/22,20	18/22,23	
20	158.4					
23	55	3.75	s	20		
24	171.7					
25	52.7	3.73	s	15,24		
26	32.9	2.80	s	14,15		
27	20.7	0.91 (6.0)	d	8,9,10	9	
28	112.9	5.76	s	4,5,6,7	5,7	7,8



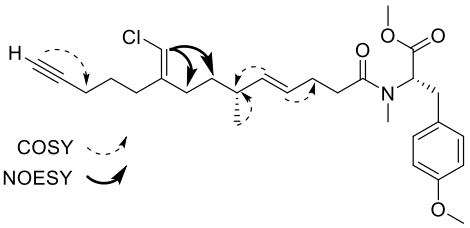
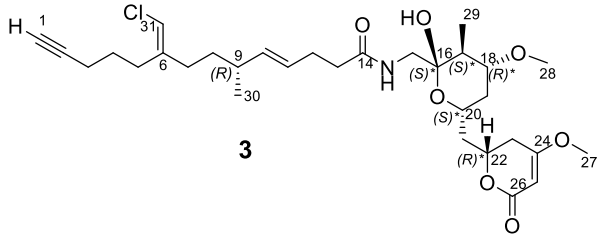


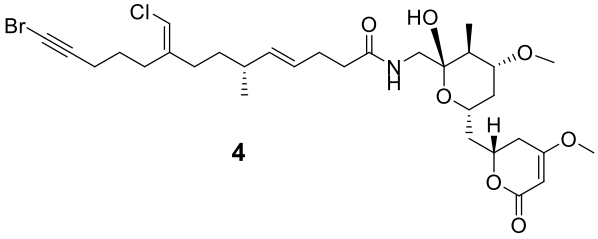
Table A3.3. Vatiamide A/B (**1,2**) mass spectrometry and optical rotation

Molecule, charge	**1,** $[M+H]^+$	**2,** $[M+H]^+$
Formula	$C_{28}H_{39}ClNO_4$	$C_{28}H_{38}BrClNO_4$
Calculated	488.2568	566.1673
Result	488.2557	566.1666
Δ ppm	-2.25	-1.24
$[\alpha]_D^{25}$ (c 0.1, CH_3OH)	7.0	-54.0

Table A3.4. Vatiamide C (**3**) NMR spectral data (^{1}H, ^{13}C, HSQC, HMBC, COSY: 500 MHz, $CDCl_3$; NOESY: 600 MHz, $CDCl_3$)

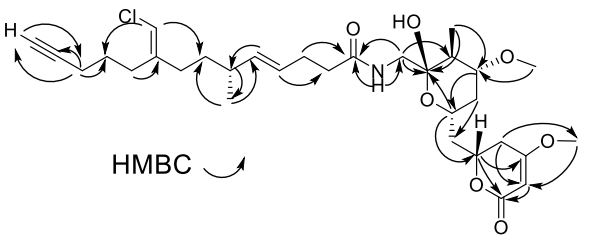


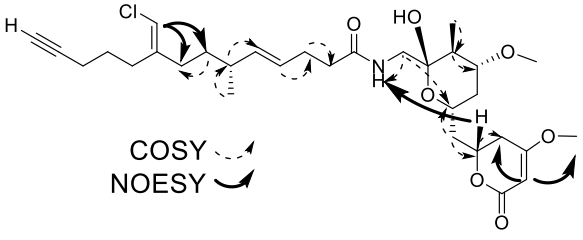
3



4

C#	δ_C	δ_H, *J* in Hz	type	HMBC	COSY	key NOE
1	68.8	1.96	t	3	3	
2	84.28					
3	18.5	2.2 (2.6, 7.2, 7.2)	dt	1, 2, 4, 5	1, 4	
4	26.2	1.63	m	2, 3, 5, 6	3, 5	
5	29.45	2.30	m	4, 6, 31	4	
6	142.02					
7	32.74	2.0	m	5, 6, 8, 9, 31	8	31
8	34.85	1.33	m	6, 7, 9, 10, 30	7, 9	31
9	36.45	2.0	m	7, 8, 10, 11, 30	7, 8, 10	
10	137.16	5.29 (7.5, 15.3)	dd	8, 9, 11, 12, 30	9, 11, 12	
11	127.45	5.34	m	9, 10, 12	10, 12	
12	28.74	2.30	m	10, 11, 13	10, 11, 13	
13	36.21	2.31	m	11, 12, 14	11, 12, 13	
14	175.26					
15	48	3.88 (7.4, 13.8), 2.91 (4.8, 13.8)	dd, dd	14, 16, 17, 18 (3.9)	20, NH	29 (3.88), 19 (2.91), NH
16	98.56					
17	43.68	1.38	m	15, 16, 18, 19, 29	18, 29	19 (1.1), 15 (2.9)
18	78.09	3.25 (4.5, 10.7, 10.7)	dt	16, 17, 20, 28, 29	19, 29	17, 19, (2.05)
19	36.98	1.12 (12,12), 2.05	dd, m	17, 18, 20, 21 (1.1)	18, 20	17 (1.1), 18, 20, 28 (2.05)
20	65.35	4.07 (2.5, 9.3, 2.5, 9.3)	tt	16, 18, 21	19, 21	19, 21, 22, 23
21	40.14	2.08, 1.66 (3, 5.4)	m, dd	19, 20, 22, 23	20, 22	20 (1.66), 22
22	74.3	4.54	m	20, 21, 23, 24, 26	21, 23	20, 21, 23
23	32.63	2.48 (4.9, 17), 2.41 (9.3, 17.2)	dd, dd	21, 22, 24, 25, 26	22, 25	20, 21 (1.66), 27
24	172.89					
25	90.46	5.12	s	22, 23, 24, 26	23, 27	23, 27
26	167.45					
27	56.21	3.72	s	23, 24, 25		
28	56.82	3.34	s	18		
29	11.57	1.06 (6.6)	d	16, 17, 18	17	
30	20.9	0.93 (6.7)	d	8, 9, 10, 11	9	
31	112.87	5.78	s	4, 5, 6, 7, 8	5, 7	7, 8
NH		6.4 (5.6, 5.6)	t	14, 15	15	14, 15
OH		4.25	bs			





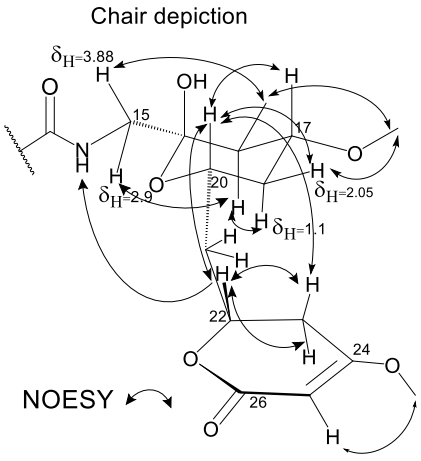


Table A3.5. Vatiamide C (**3**) NMR spectroscopy with key correlations (500 MHz, DMSO-d_6)

C#	δ_C	δ_H, *J* in Hz	type	HMBC	key NOE
1	71.57	2.79 (3.2, 3.2)	t	2	
2	84.07				
3	17.74	2.16	d	1, 2, 4, 5	
4	25.65	1.54 (7.5, 7.5, 7.5)	dq	2, 3, 5, 6	
5	28.78	2.23	m	3, 4, 6, 7, 31	
6	142.11				
7	31.86	2.01	m	5, 6, 8, 31	
8	34.32	1.33	m	6, 7, 9, 10, 30	
9	35.68	2	m	6, 7, 8, 10, 11, 30, 31	
10	135.79	5.26 (7.3, 15.4)	dd	9, 11, 12, 30	
11	127.69	5.34	m	12, 19	
12	28.31	2.13	m	10, 11, 13, 14	
13	35.26	2.15	m	10, 11, 12, 14	
14	173.52				
15	45.47	2.89 (4.2,13.5), 3.45 (7.8, 13.5)	dd, dd	14, 16, 17 (2.89)	
16	98.45				
17	40.44	1.33	m	15, 16, 18, 20, 21, 29	
18	77.86	3.15 (4.5, 10.7, 10.8)	td	17, 27, 29	
19	36.1	0.95, 2.1 (2.8, 5.6, 15)	m, ddd	18, 20, 21	
20	63.12	3.89	m		OH
21	40.3	1.74 (4.3, 7.7, 14.0), 1.91 (5.6, 8.7, 13.9)	ddd, ddd	19 (1.74), 20, 22, 23	
22	72.67	4.55	m		
23	31.61	2.46 (4.0, 17.2), 2.56 (1.3, 11.7, 17.1)	dd, ddd	22, 24, 25, 26 (2.56), 21 (2.56)	
24	171.96				
25	89.94	5.16 (1.3)	d	23, 24, 26	
26	166.44				
27	55.68	3.23	s	18	
28	56.24	3.72	s	24	
29	11.66	0.91 (6.6)	d	16, 18, 21	
30	20.55	0.92 (6.7)	d	8, 10	
31	112.36	6.01	s	5, 6, 7	
NH		7.52 (4.3, 7.6)	dd	14	
OH		5.73 (1.3)	d	15, 16, 21	20

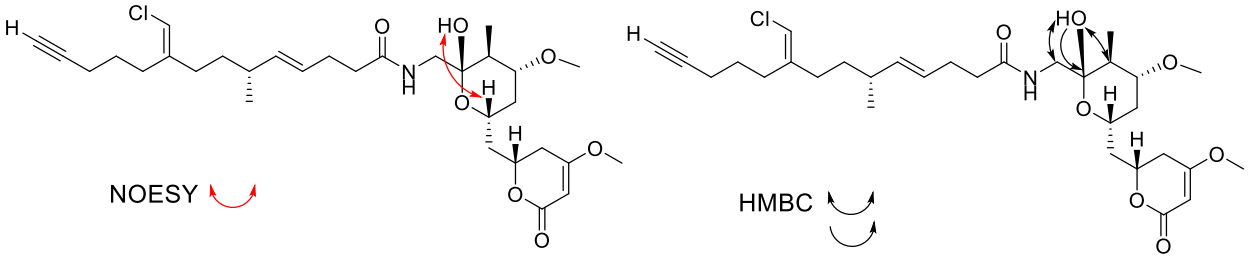
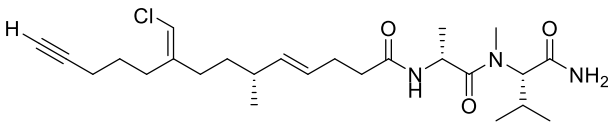


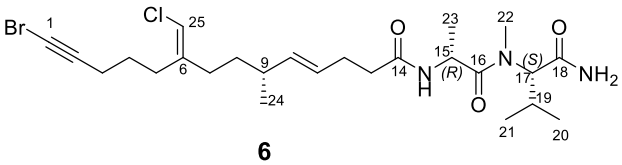
Table A3.6. Vatiamide C, D (**3, 4**) mass spectrometry and optical rotation

	3, $[M+Na]^+$	**4**, $[M+Na]^+$
Formula	$C_{31}H_{46}ClNO_7Na$	$C_{31}H_{45}BrClNO_7Na$
Calculated	602.2855	680.1960
Result	602.2834	680.1953
Δ ppm	3.49	1.03
$[\alpha]_D^{25}$ (c 0.1, CH_3OH)	-0.3	33.3

Table A3.7.Vatiamide F (**6**) NMR spectroscopy with key correlations (500 MHz, DMSO-d_6)



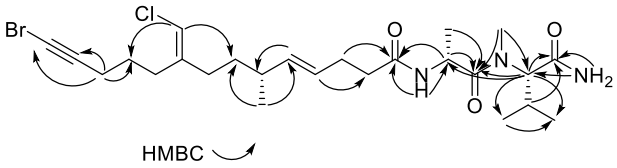
5



6

C#	δ_C	δ_H, *J* in Hz	type	HMBC	COSY	key NOE
1	40.46					
2	80.05					
3	18.83	2.24	m	1, 2, 4, 5, 6	4, 5	
4	25.43	1.54 (6.6, 6.6, 6.6, 6.6)	p	3, 5	3, 5	
5	28.66	2.20	m	3, 4, 6, 7, 25	3, 4	
6	142					
7	31.82	2.03	m	6, 8, 25	8, 25	25
8	34.3	1.33	m	7, 9, 10, 11, 24	7, 9	25
9	35.73	1.98	m	6, 7, 8, 10, 11, 24, 25	8, 10, 24	
10	135.81	5.24 (7.3, 15.3)	dd	9, 11, 12, 24	9, 11	
11	127.63	5.33	m	9, 10, 12	10, 12	
12	28.16	2.16	m	10, 11, 13, 14	11, 13	
13	35	2.12	m	10, 11, 12, 14	12	NH
14	171.13					
15	44.81	4.70 (6.9, 6.9, 6.9, 6.9)	p	14, 16, 23	NH, 23	22, 23
16	172.93					
17	60.95	4.53 (10.6)	d	15, 16, 18, 19, 20, 21, 22	19, 22	NH_2(b), 20, 21, 22
18	171.52					
19	26.27	2.08	m	17, 18, 20, 21	17, 20, 21	22
20	19.82	0.91 (6.3)	d	17, 19, 21	17, 19	17, 19, 22
21	18.65	0.73 (6.8)	d	17, 19, 20	17, 19, 20	17, 19, 22
22	30.38	2.98	s	16, 17	17	15, 19, 20, 21, 23
23	17.47	1.16 (7.4)	d	15, 16	15	NH, 15, 20
24	20.58	0.92 (6.3)	d	8, 9, 10	8, 9	
25	112.45	6.03	s	4, 5, 6, 7, 8, 25	5, 7	7, 8
NH		8.09 (54)	d	14, 15, 23	15, 23	13, 23
NH_2		7.05(a),7.16(b)[a]	s,s	17(b), 18(a, b)	17(b)	17(b), 20(a, b)

[a] Note: Presence of additional "shadow" peaks due to conformer from rotation around *N*-methyl amide bond. At 25°C, peaks appear at 2:1 ratio in *DMSO-d_6* and 8:1 ratio in acetonitrile-d_3; single conformer resolved at 80°C in DMSO-d_6. Major conformer peak at 25°C reported above.



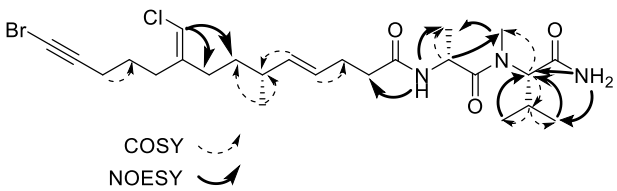


Table A3.8. Vatiamide E, F (**5, 6**) mass spectrometry and optical rotation

Molecule, charge	**5**, $[M+Na]^+$	**6**, $[M+Na]^+$
Formula	$C_{25}H_{40}ClN_3O_3Na$	$C_{25}H_{39}BrClN_3O_3Na$
Calculated	488.2650	566.1756
Result	488.2641	566.1742
Δ ppm	1.84	2.47
$[\alpha]_D^{25}$ (c 0.1, CH_3OH)	N/A	-25.0

Table A3.9. Proposed functions of open reading frames in *vat* pathway

Protein	Length	Proposed Function	Similarity	Identity	Similarity	Accession
VatA	598	Fatty-acyl AMP ligase	JamA, *Lyngbya majuscula* (*Moorea producens* JHB)	92%	95%	AAS98774.1
VatB	321	Desaturase	JamB, *Lyngbya majuscula* (*Moorea producens* JHB)	97%	98%	AAS98775.1
VatC	100	ACP	JamC, *Lyngbya majuscula* (*Moorea producens* JHB)	93%	95%	AAS98798.1
VatD	683	Flavin-dependent brominase	JamD, *Lyngbya majuscula* (*Moorea producens* JHB)	94%	97%	AAS98776.1
VatE	1737	KS AT halogenase ACP ACP ACP	JamE, *Lyngbya majuscula* (*Moorea producens* JHB)	92%	94%	AAS98777.1
VatF	79	ACP	JamF, *Lyngbya majuscula* (*Moorea producens* JHB)	94%	94%	AAS98799.1
VatG	409	KS	JamG, *Lyngbya majuscula* (*Moorea producens* JHB)	91%	94%	AAS98778.1
VatH	419	HCS	JamH, *Lyngbya majuscula (Moorea producens)*	98%	99%	AAS98779.1
VatI	434	Unknown	HNH endonuclease, *Moorea producens* PAL-8-15-08-1	96%	97%	WP_070395762.1
VatJ	254	ECH	JamI, *Lyngbya majuscula* (*Moorea producens* JHB)	96%	99%	AAS98780.1
VatK	3277	ECH ER KS AT DH CMT ER KR ACP	JamJ, *Lyngbya majuscula* (*Moorea producens* JHB)	94%	96%	AAS98781.1
VatL	1659	KS AT KR ACP	JamK, *Lyngbya majuscula* (*Moorea producens* JHB)	92%	94%	AAS98782.1
VatM	2187	KS AT DH ER KR ACP	JamL, *Lyngbya majuscula* (*Moorea producens* JHB)	79%	86%	AAS98783.1
VatN	2153	C A(Tyr) N-MT O-MT PCP TE	Nonribosomal peptide synthetase, *Moorea producens* 3L	72%	81%	AEF01451.1
VatO	271	O-MT	O-methyltransferase protein, *Planktothrix serta* PCC 8927	65%	77%	CUR19792.1
VatP	79	ACP	ColF, *Moorea bouillonii* PNG5-198	94%	96%	AKQ09583.1
VatQ	1582	C A(Ala) PCP E	PuwG, *Cylindrospermum alatosporum* CCALA 988	52%	68%	AIW82284.1
VatR	1895	C A(Val) N-MT PCP AmT	BarG, *Lyngbya majuscula* (*Moorea producens*)	50%	64%	AEE88297.1
VatS	1585	C A(Gly) N-MT PCP	MgcJ, *Okeania hirsuta* PAB10Feb10-1	84%	92%	AZH23792.1
VatT	2260	KS AT DH C-MT KR ACP	MgcQ, *Okeania hirsuta* PAB10Feb10-1	67%	80%	AZH23817.1
VatU	1950	KS AT O-MT KR ACP	CurL, *Moorea producens* 3L	67%	81%	AEE88278.1
VatV	1610	KS AT KR ACP	CurG *Moorea producens* 3L	62%	75%	AEE88283.1
VatW	2888	KS AT KR ACP KS AT ACP TE	ColF, *Moorea bouillonii* PNG5-198	64%	77%	AKQ09583.1
VatX	219	O-MT	StfMI, *Streptomyces steffisburgensis*	42%	60%	CAJ42328.1
Orf1	378	Putative hydrolase	Linear amide C-N hydrolase, *Moorea producens*	86%	93%	WP_071106594.1
Orf2	354	unknown	tRNA 2-selenouridine synthase MnmH *Moorea producens*	96%	97%	WP_070395342.1
Orf3	85	transposase	Transposase, *Leptolyngbya* sp. PCC 7376	59%	78%	AFY39744.1
Orf4	86	transposase	ISAs1 family transposase *Moorea producens*	88%	91%	WP_083373584.1
Orf5	185	transposase	DDE transposase family protein *Moorea producens*	85%	91%	WP_070395360.1
Orf6	1155	Regulatory	Response regulator *Moorea producens*	98%	98%	WP_071106591.1

Table A3.10. Organisms and accession numbers for 16S sequence comparison (Figure A3.1).

Organism	Accession #
Moorea producens 3L	NR_116539.1
Moorea producens JHB	CP017708.1
Moorea producens PAL15Aug08-1	CP017599.1

A

	ASI16Jul14-2	3L	JHB	PAL15Aug08-1
ASI16Jul14-2	-	97.7	98.6	98.9
3L	97.7	-	98.6	98.4
JHB	98.6	98.6	-	98.7
PAL15Aug08-1	98.9	98.4	98.7	-

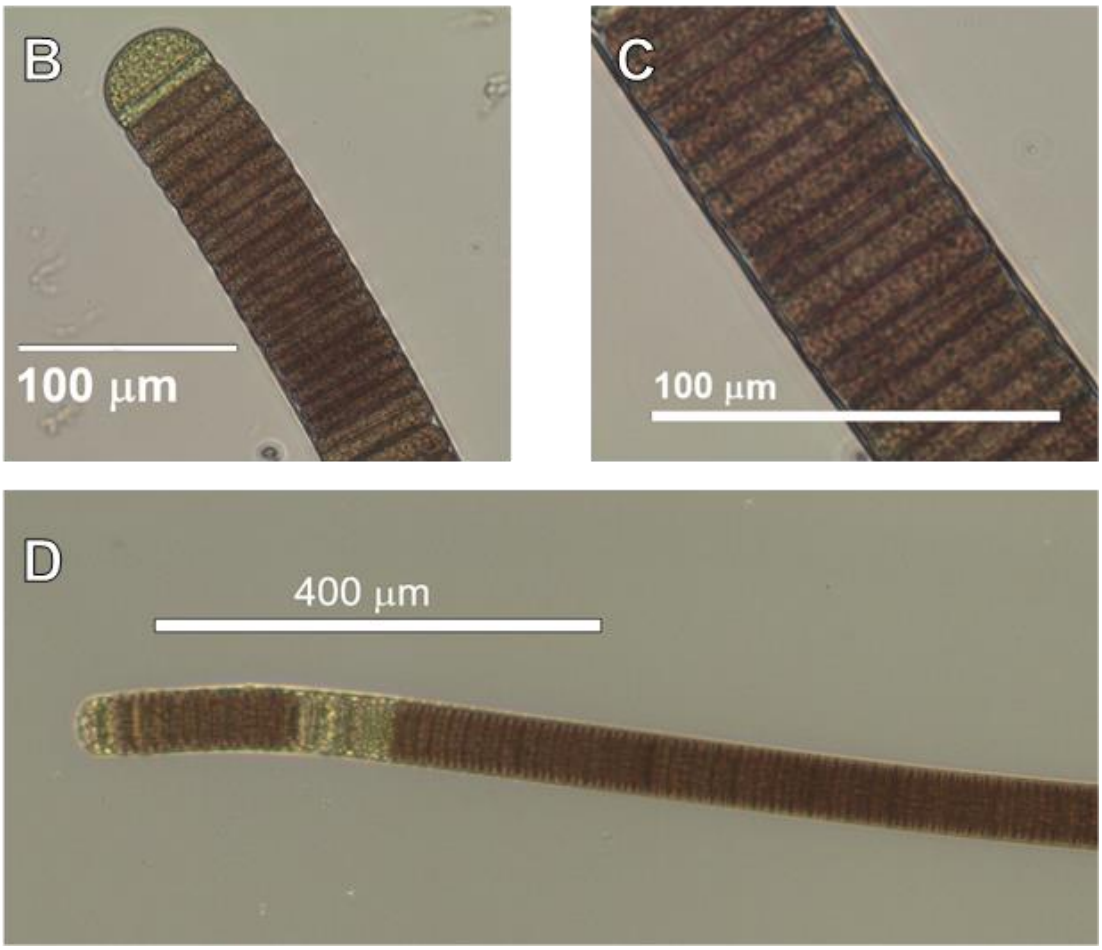


Figure A3.1. Analysis of 16S rRNA sequence and morphology of *Moorea producens* ASI16Jul14-2. A) Identity matrix of 16S rRNA sequences of *Moorea producens*, aligned by MUSCLE, reflecting % nucleotide identity. B) Microscopy image of ASI at 40X objective C) Microscopy image of ASI at 20X objective D) Microscopy image of ASI at 10X objective

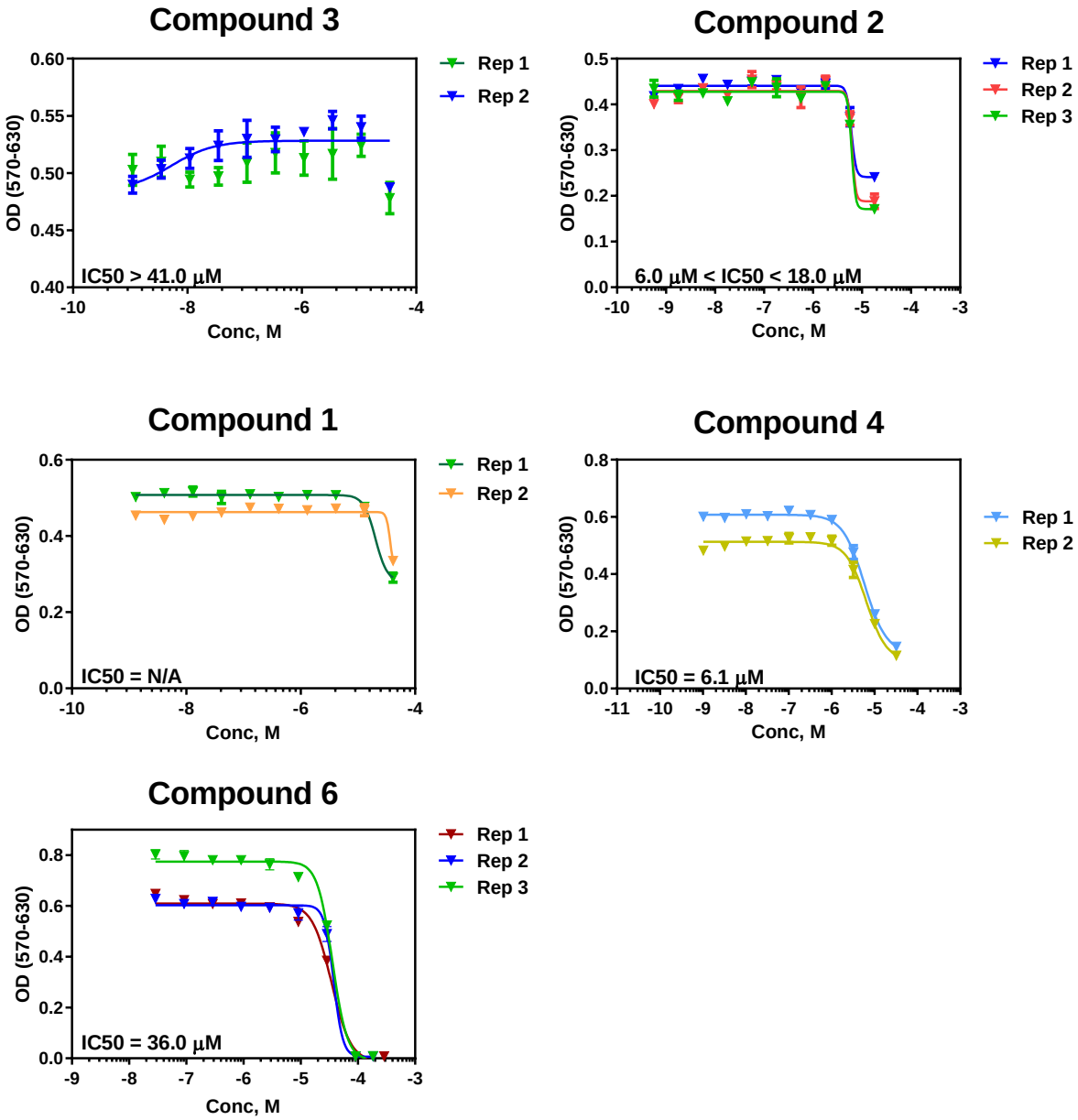


Figure A3.2. NCI-H460 MTT fluorometric assay results for pure compounds **1-4, 6**. Each replicate represents a dilution curve generated from three biological replicates. Compound **5** was not recovered in quantity sufficient for bioassay testing.

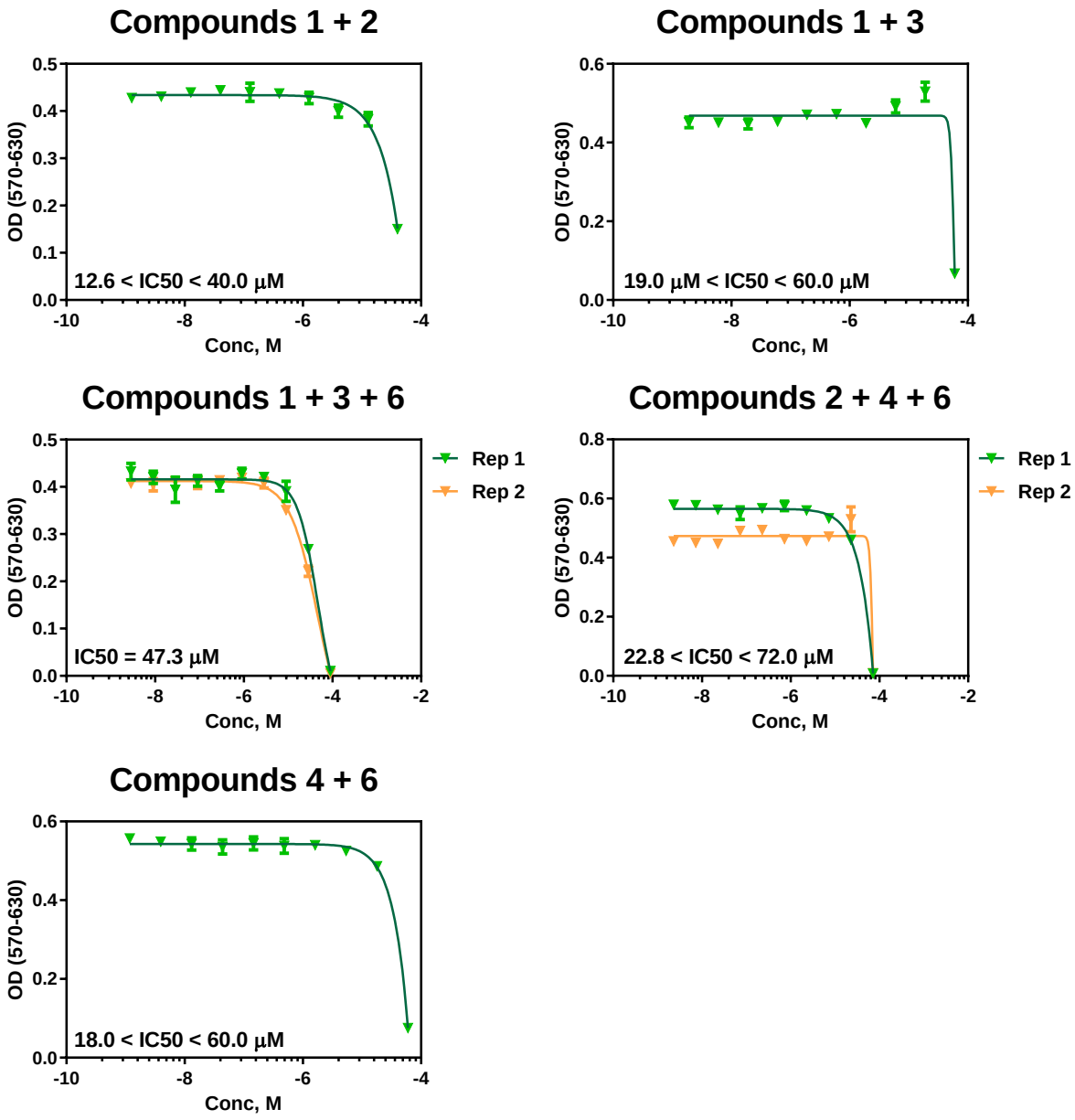


Figure A3.3. NCI-H460 MTT fluorometric assay results for mixtures of compounds **1-4, 6** to assess synergy. Each replicate represents a dilution curve generated from three biological replicates. Compound **5** was not recovered in quantity sufficient for bioassay testing.

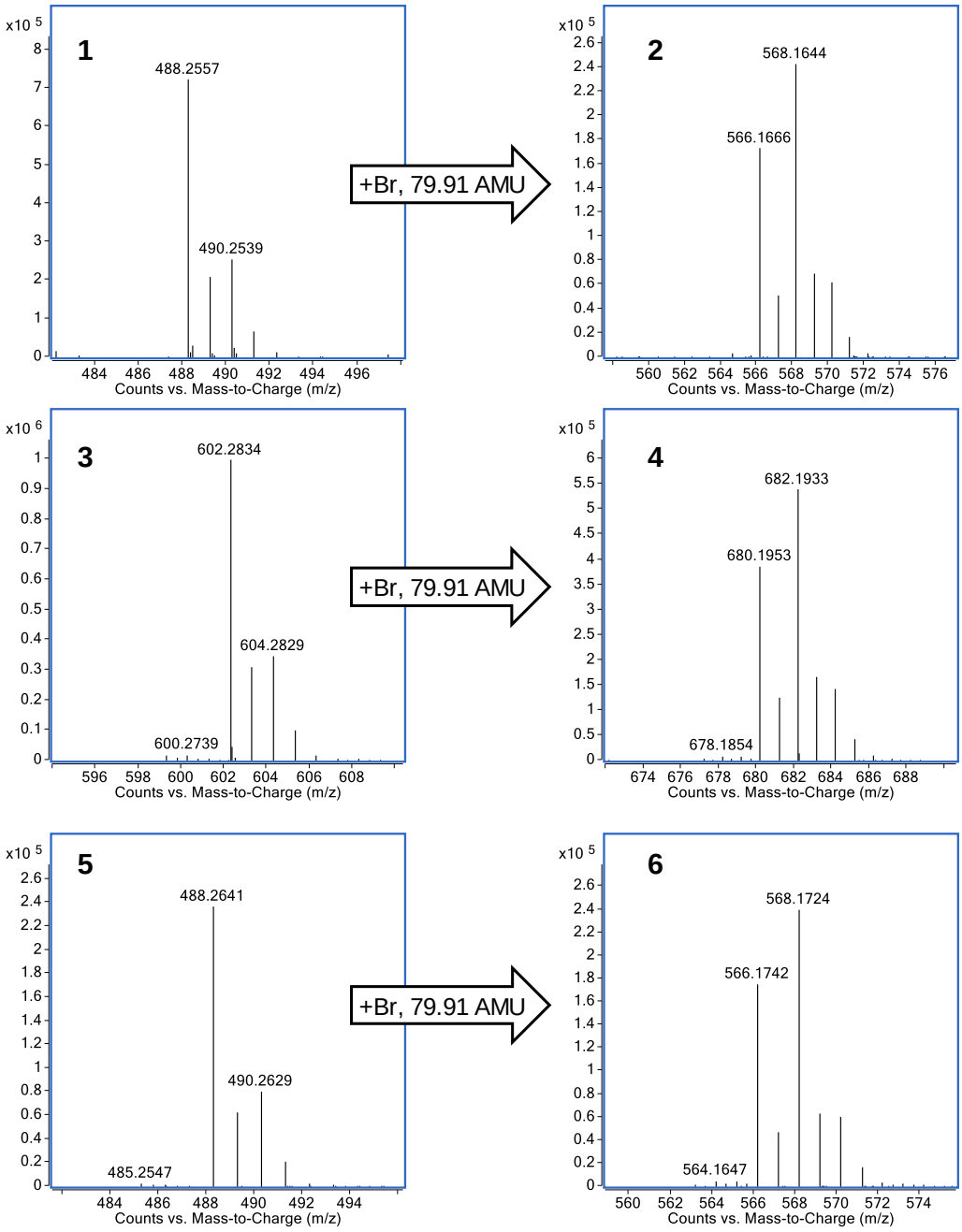


Figure A3.4. High-resolution mass spectrometry of **1**-**6**.

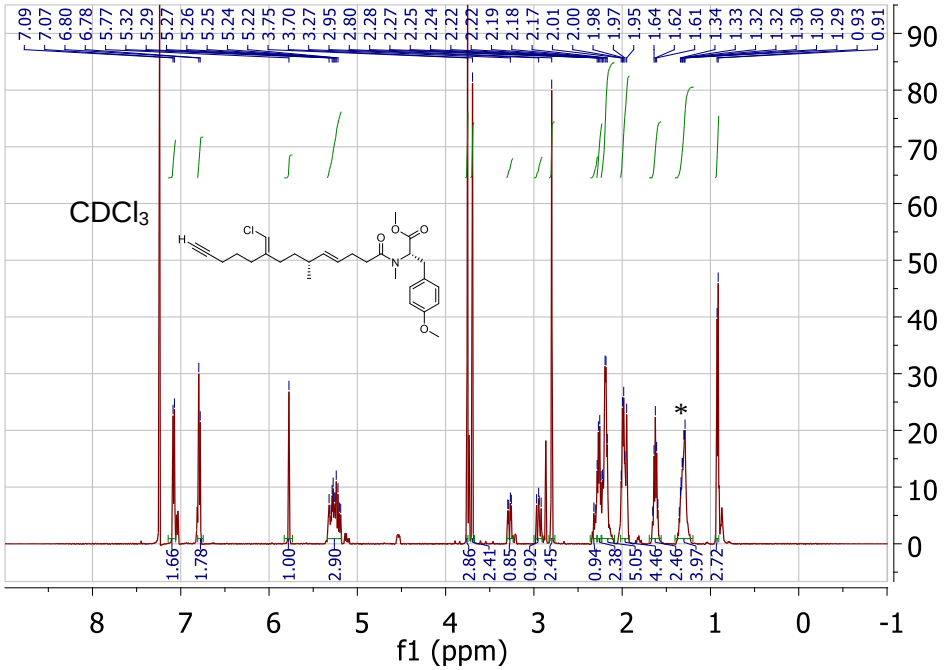


Figure A3.5. ^{1}H-NMR spectrum of **1** (500 MHz, $CDCl_3$). * minor grease contaminant.

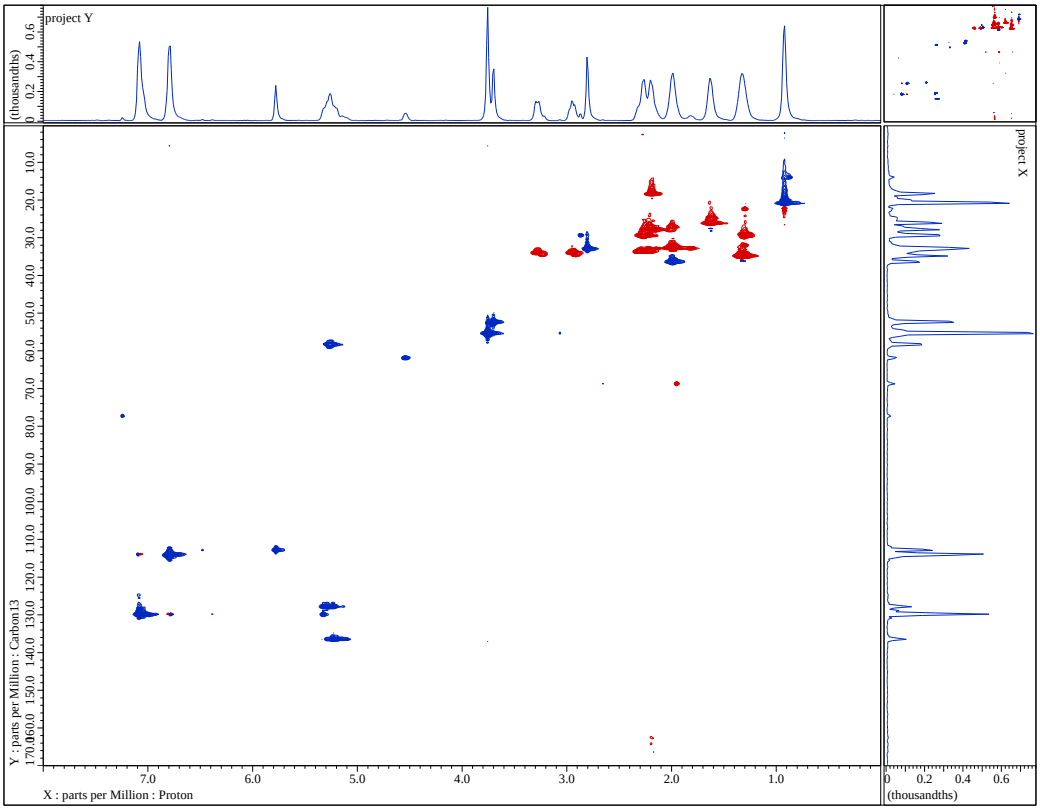


Figure A3.6. HSQC spectrum of **1** (500 MHz, $CDCl_3$).

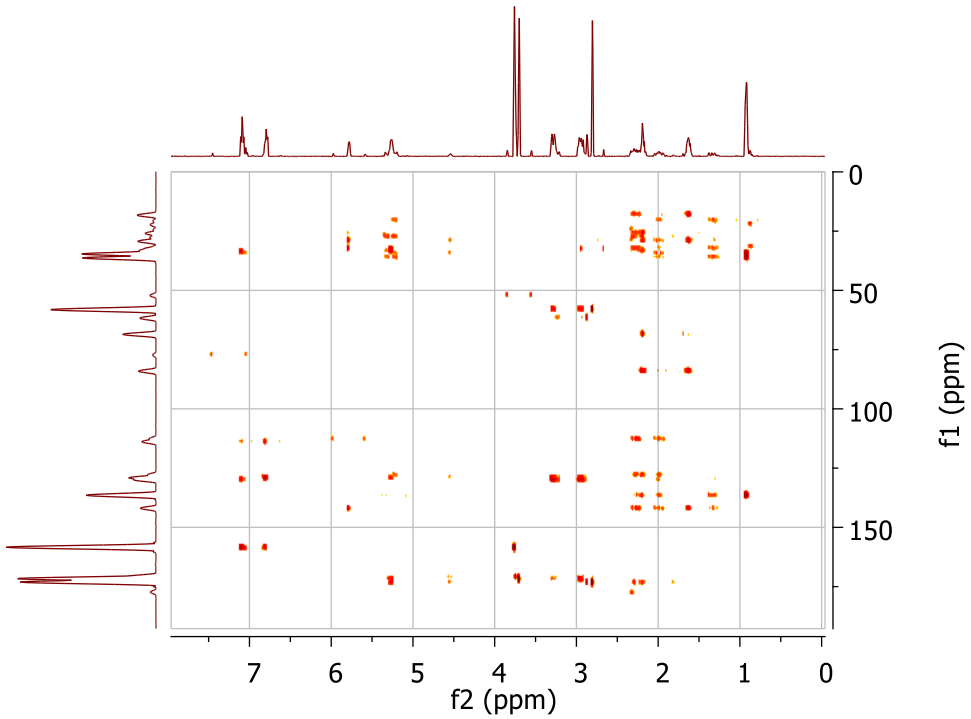


Figure A3.7. HMBC spectrum of **1** (500 MHz, $CDCl_3$).

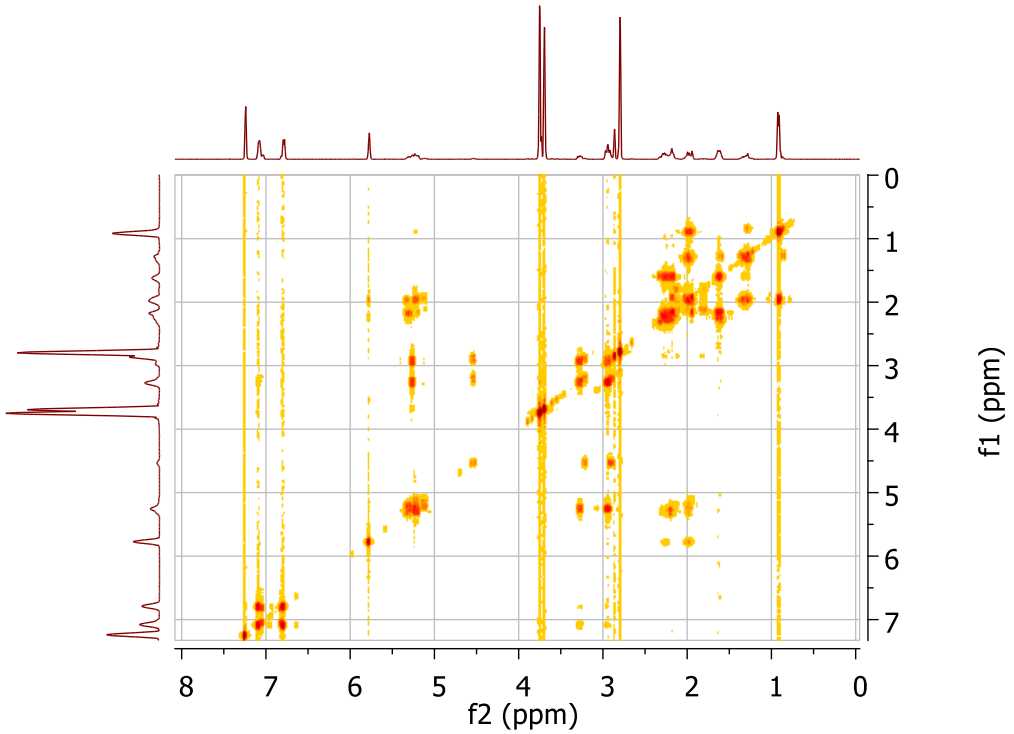


Figure A3.8. COSY spectrum of **1** (500 MHz, $CDCl_3$).

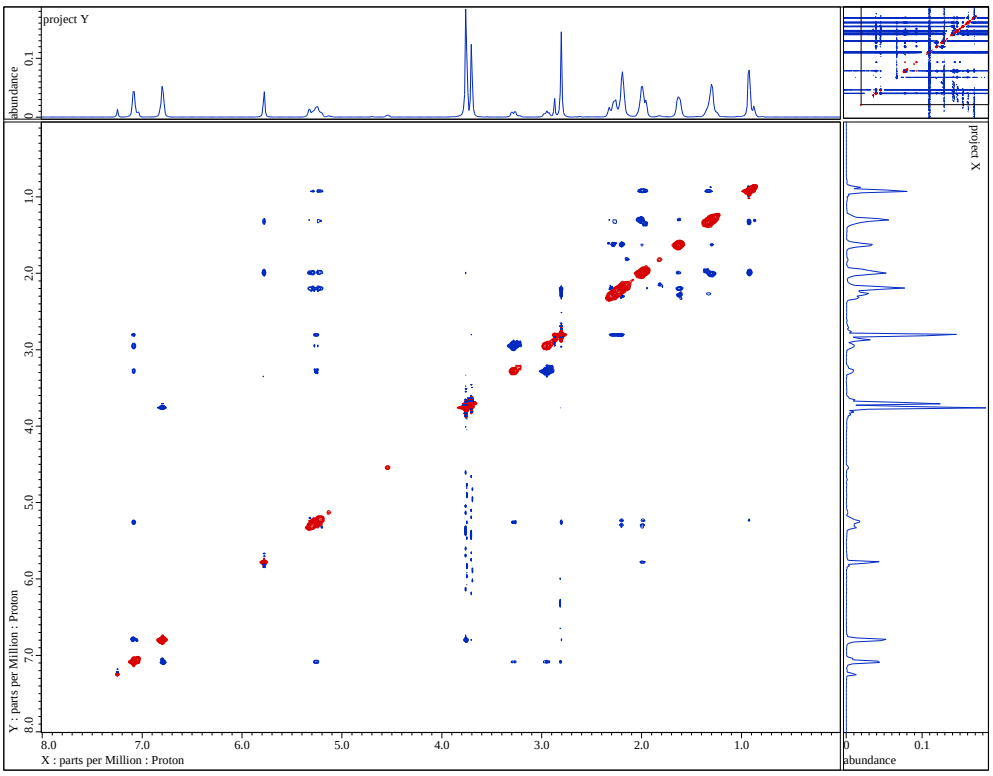


Figure A3.9. NOESY spectrum of **1** (500 MHz, $CDCl_3$).

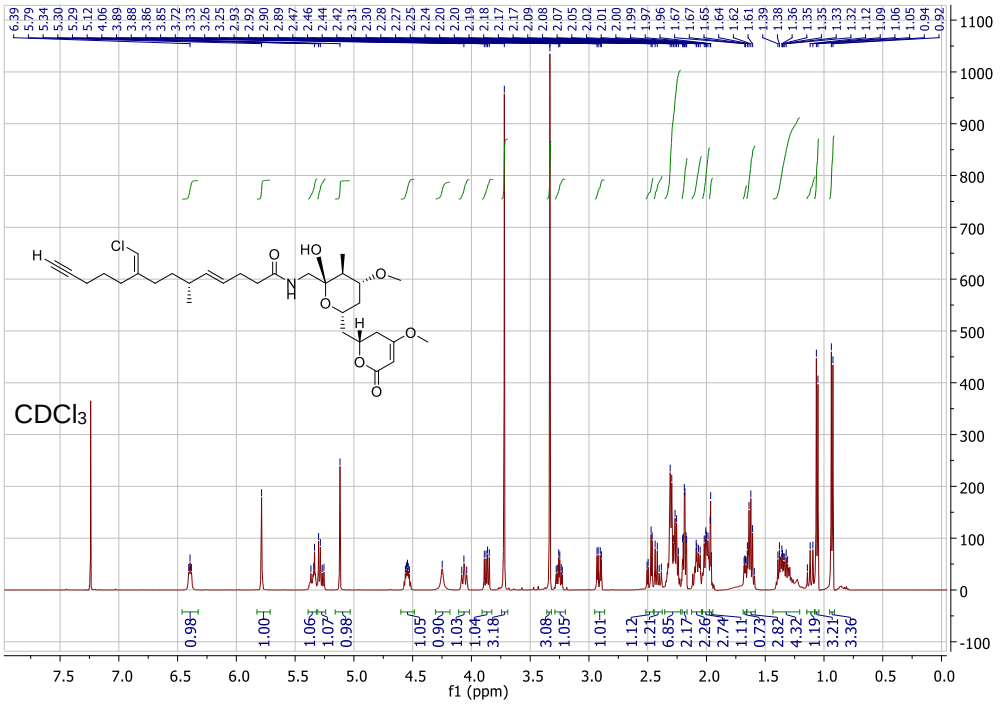


Figure A3.10. ^{1}H-NMR spectrum of **3** (500 MHz, $CDCl_3$).

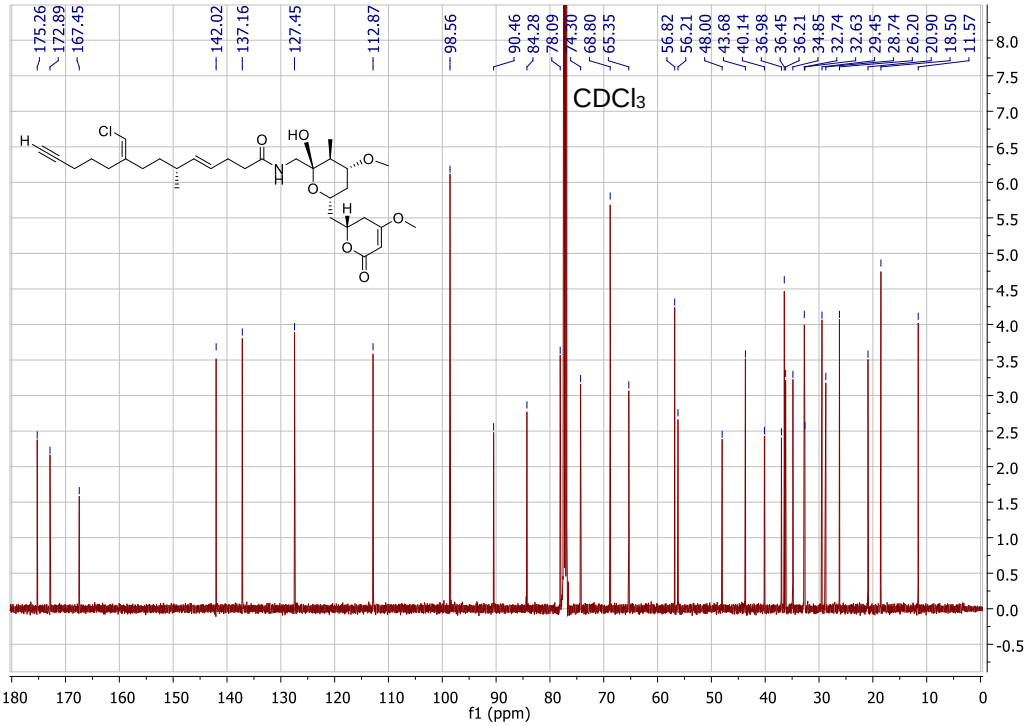


Figure A3.11. ^{13}C-NMR spectrum of **3** (500 MHz, $CDCl_3$).

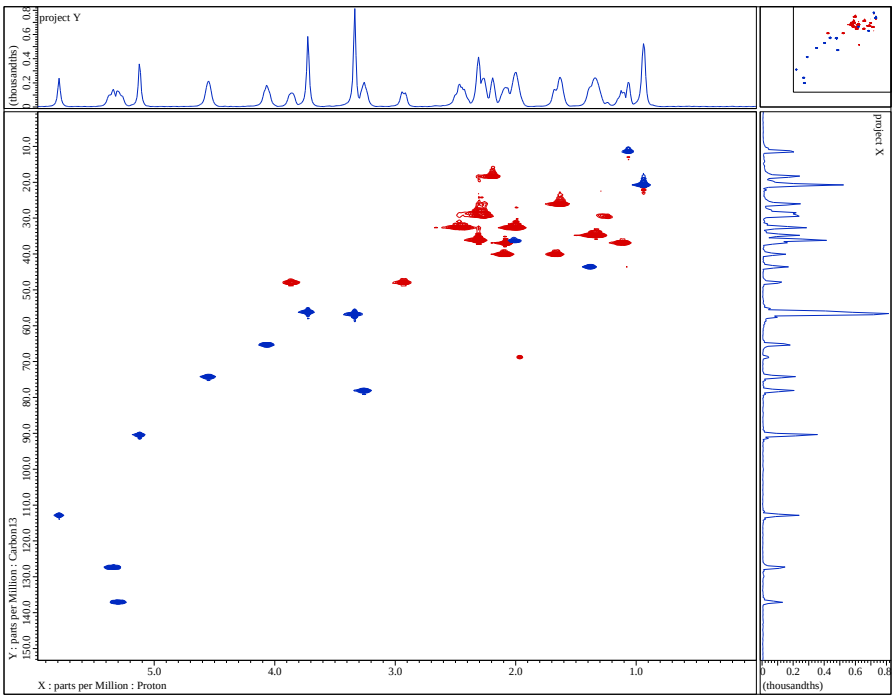


Figure A3.12. HSQC spectrum of **3** (500 MHz, $CDCl_3$).

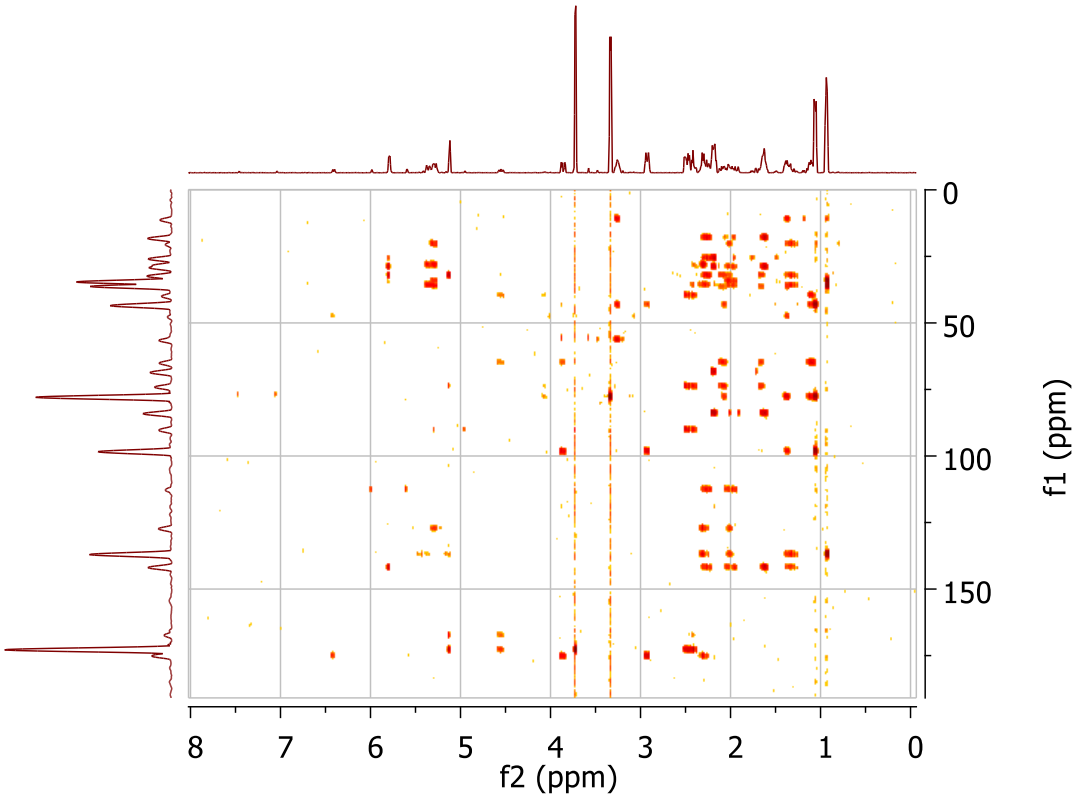


Figure A3.13. HMBC spectrum of **3** (500 MHz, $CDCl_3$).

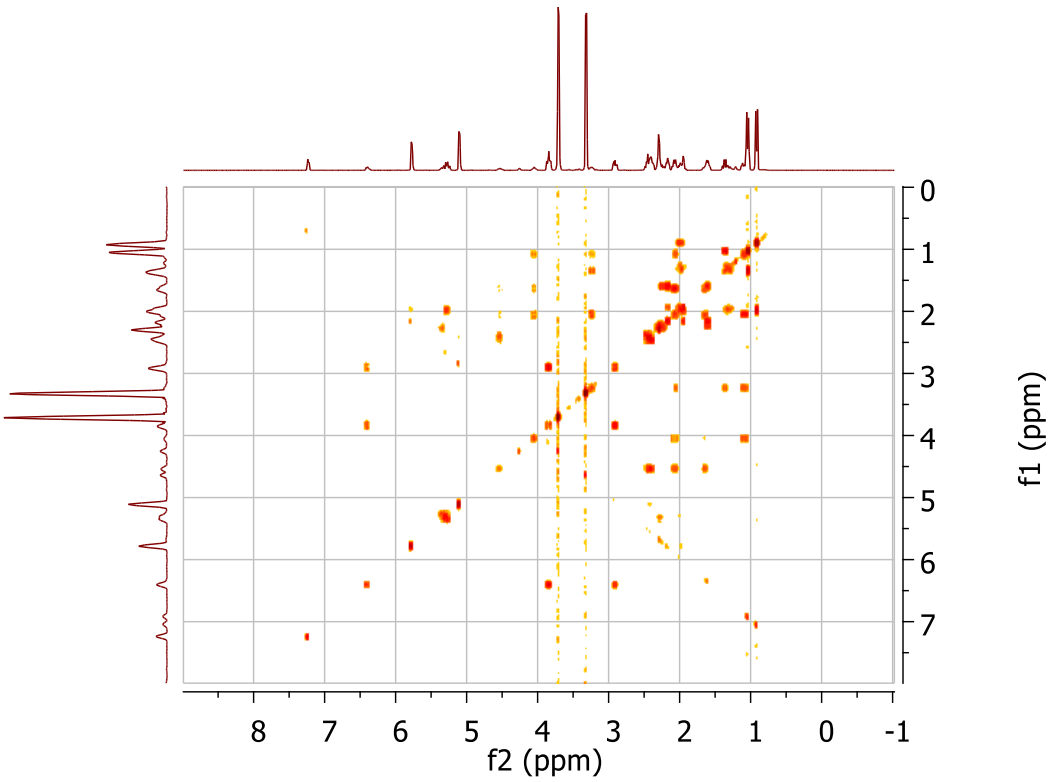


Figure A3.14. COSY spectrum of **3** (500 MHz, $CDCl_3$).

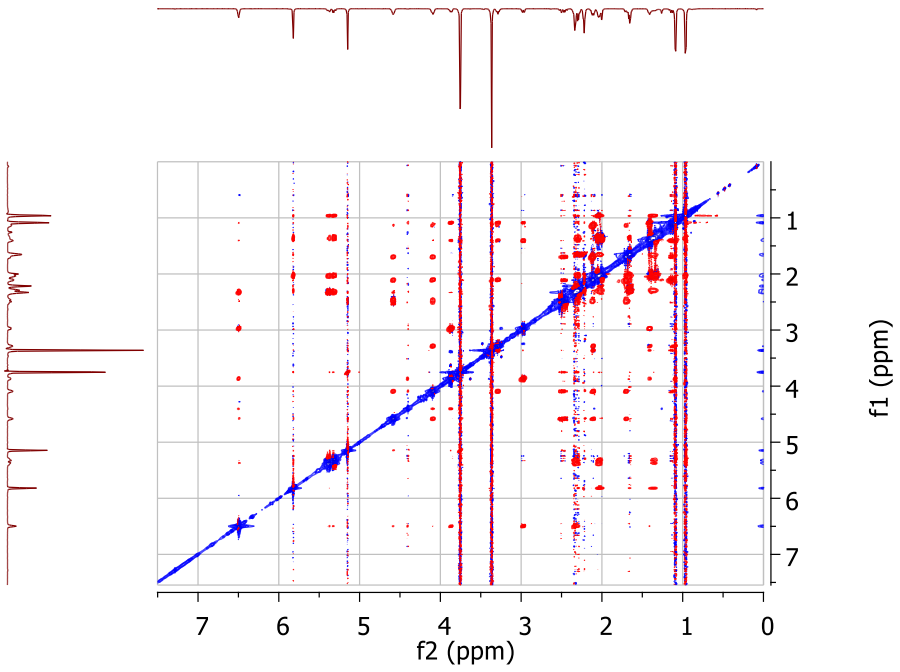


Figure A3.15. NOESY spectrum of **3** (500 MHz, $CDCl_3$).

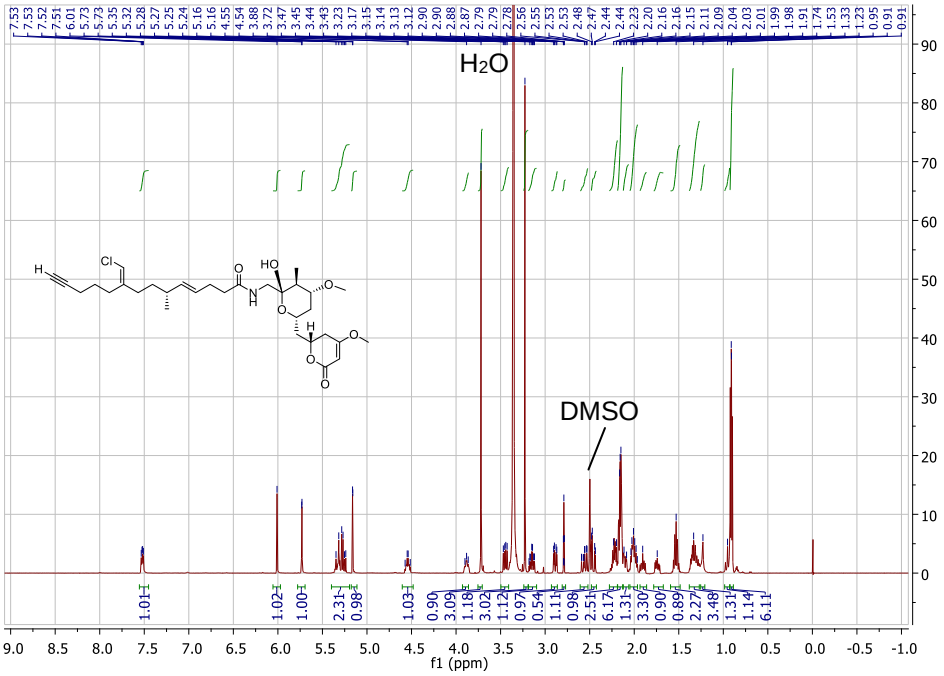


Figure A3.16. ^{1}H-NMR spectrum of **3** (500 MHz, DMSO-d_6.)

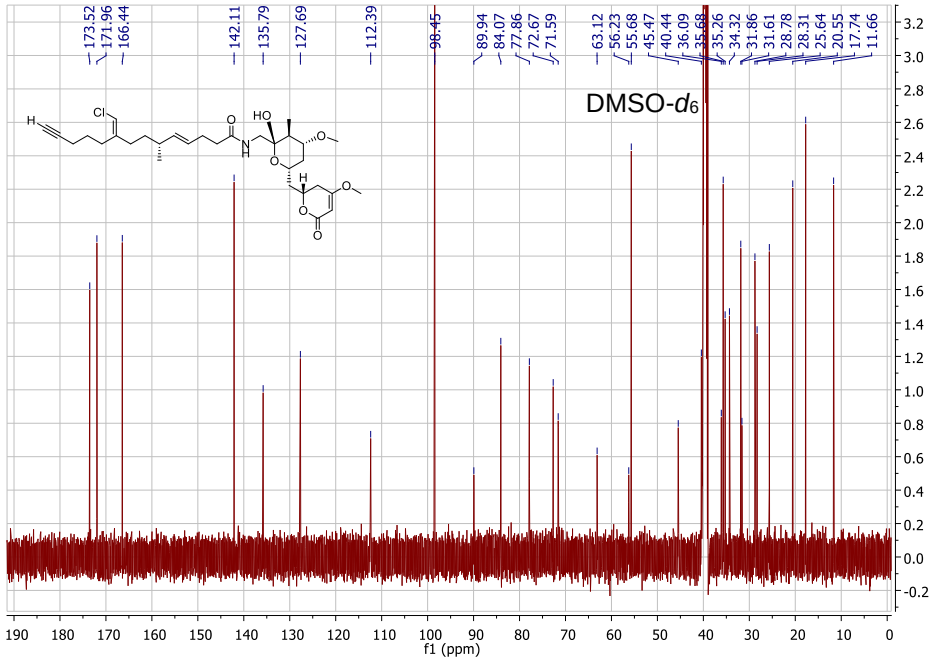


Figure A3.17. ^{13}C-NMR spectrum of **3** (500 MHz, DMSO-d_6.)

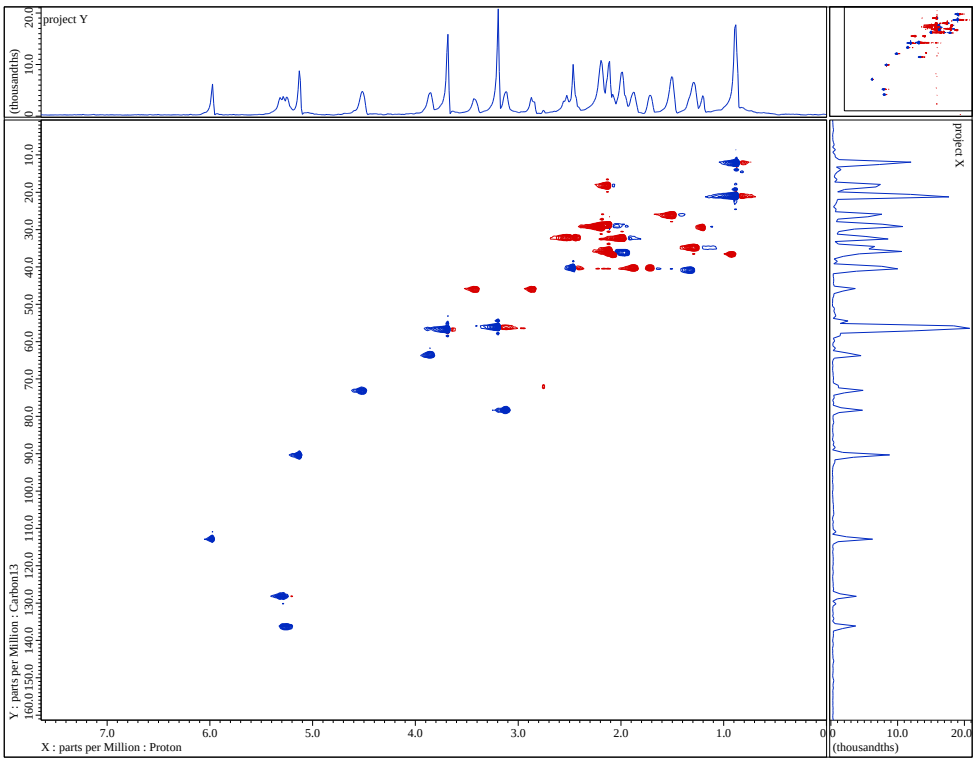


Figure A3.18. HSQC spectrum of **3** (500 MHz, DMSO-d_6.)

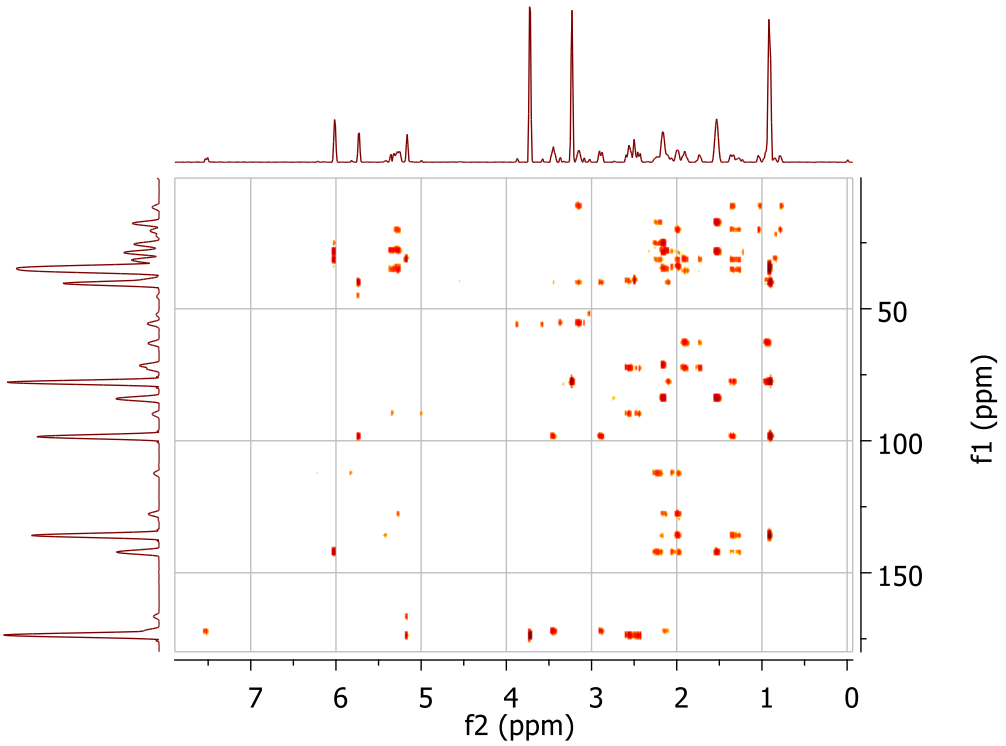


Figure A3.19. HMBC spectrum of **3** (500 MHz, DMSO-d_6.)

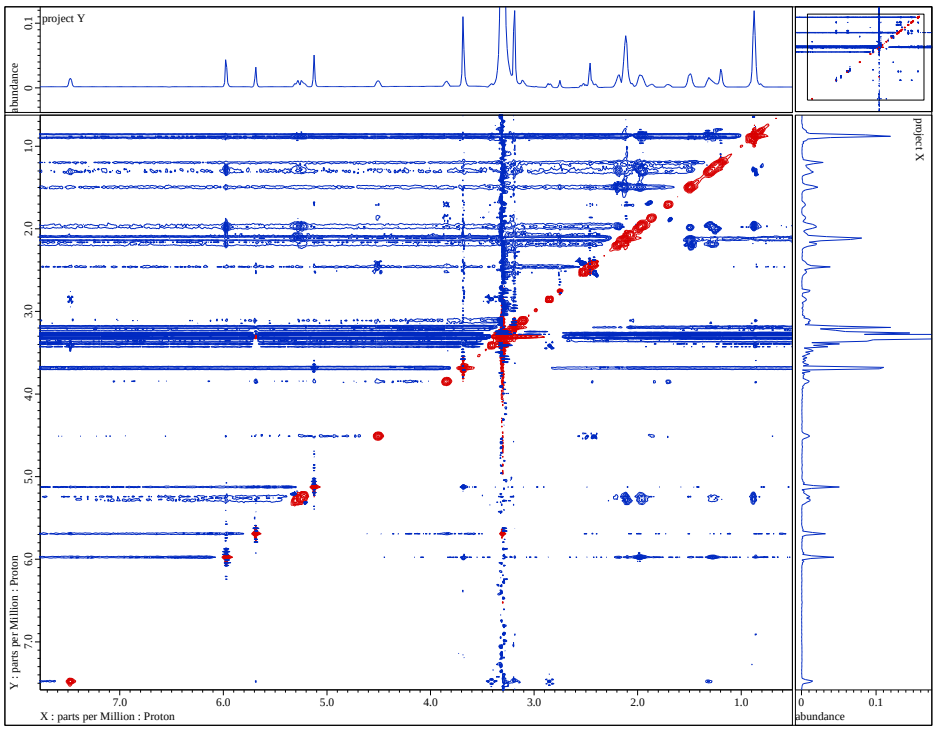


Figure A3.20. NOESY spectrum of **3** (500 MHz, DMSO-d_6.)

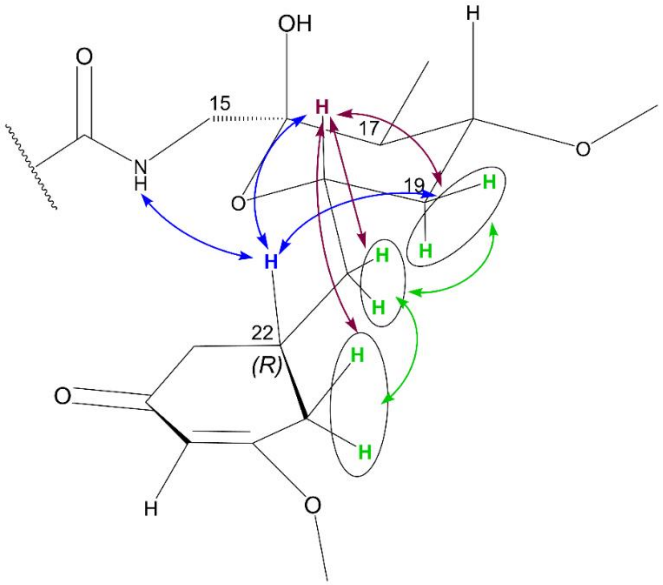


Figure A3.21. NOESY correlation superimposed on three dimensional structure of compound **3** in the chair conformation (500 MHz, $CDCl_3$).

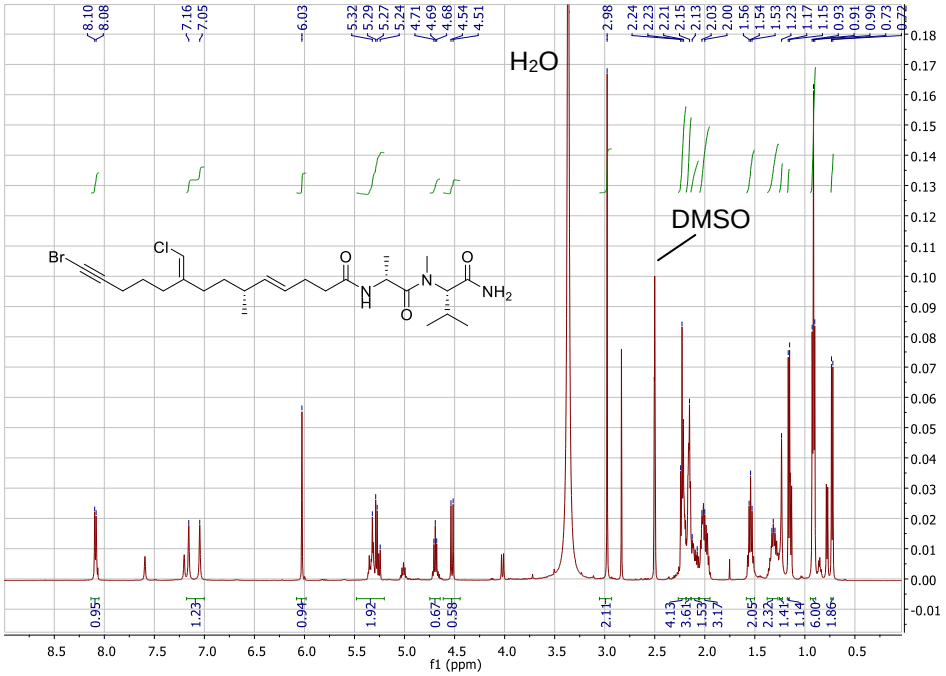


Figure A3.22. ^{1}H-NMR spectrum of **6** (500 MHz, DMSO-d_6).

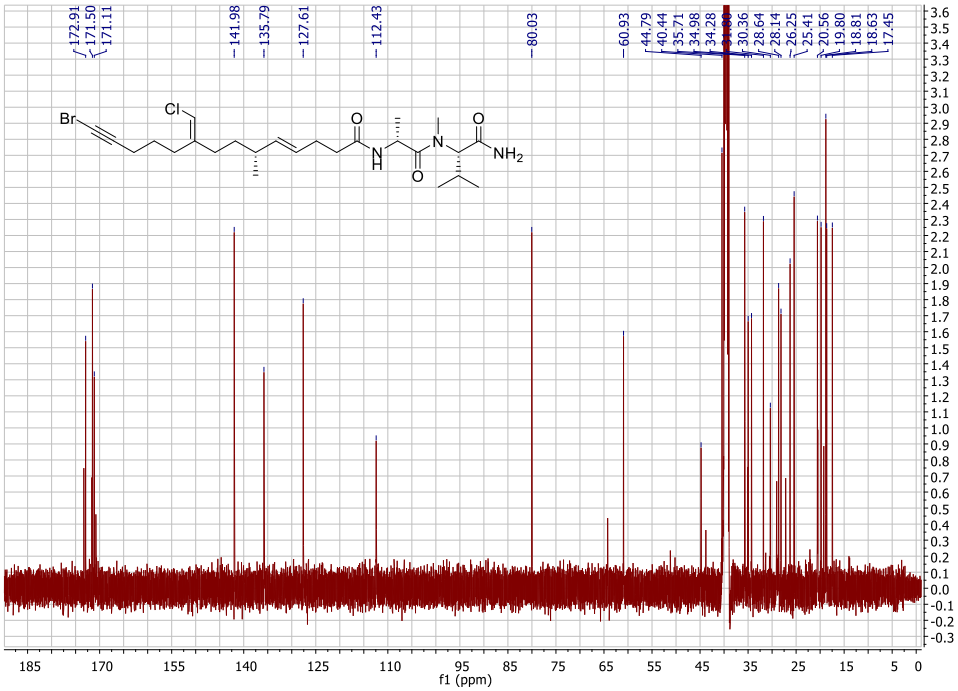


Figure A3.23. ^{13}C-NMR spectrum of **6** (500 MHz, DMSO-d_6).

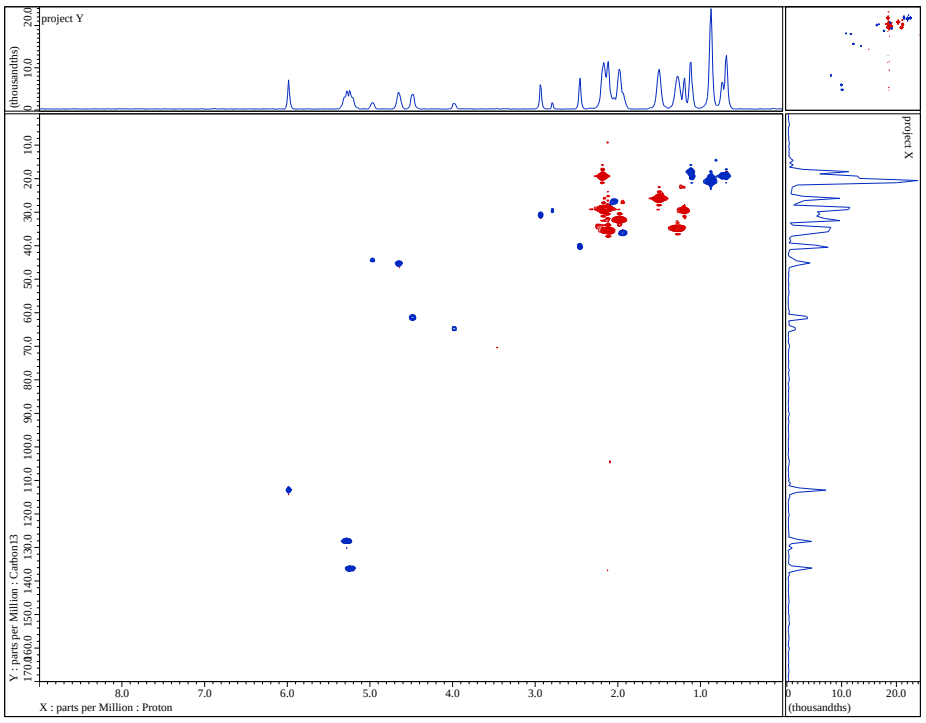


Figure A3.24. HSQC spectrum of **6** (500 MHz, DMSO-d_6).

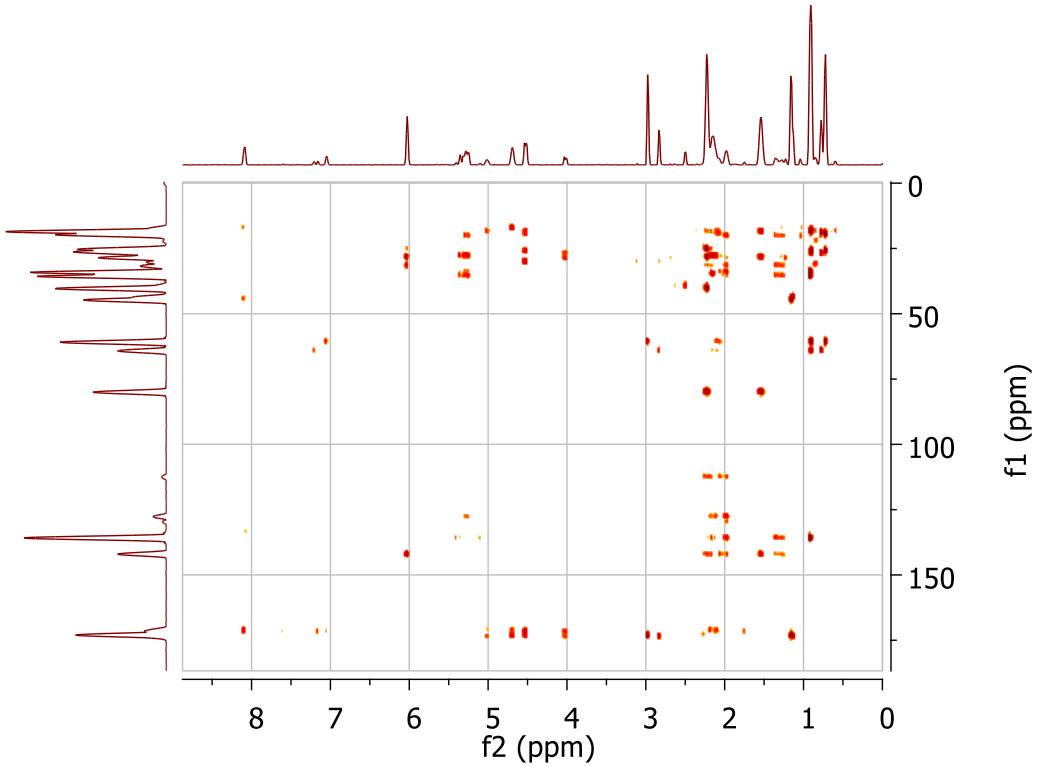


Figure A3.25. HMBC spectrum of **6** (500 MHz, DMSO-d_6).

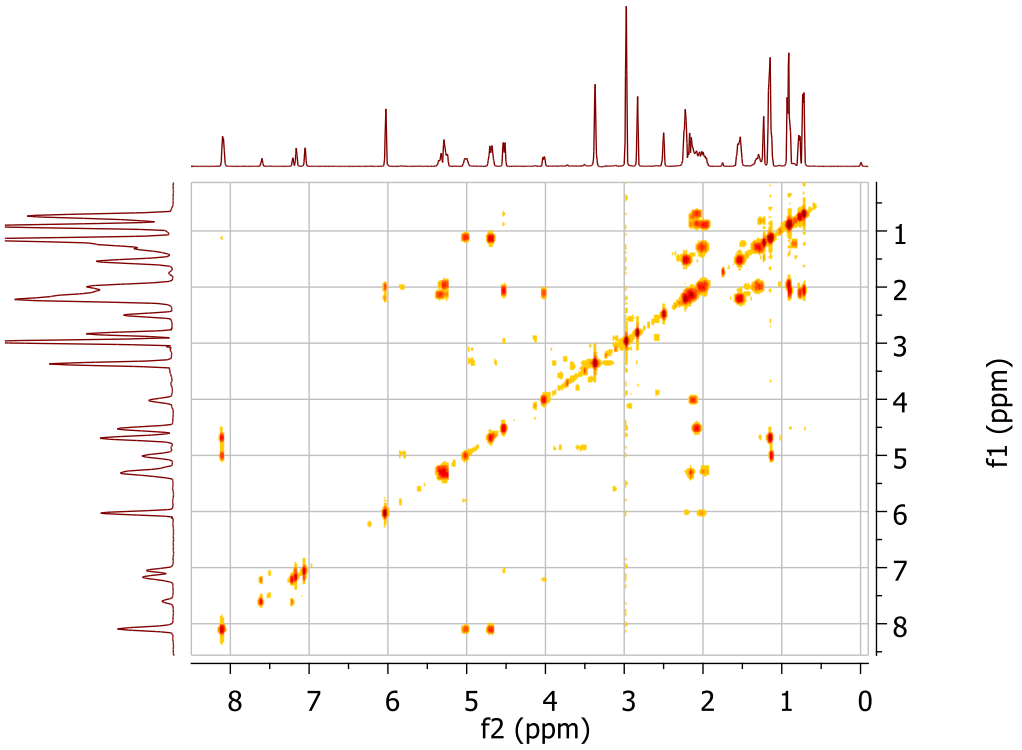


Figure A3.26. COSY spectrum of **6** (500 MHz, DMSO-d_6).

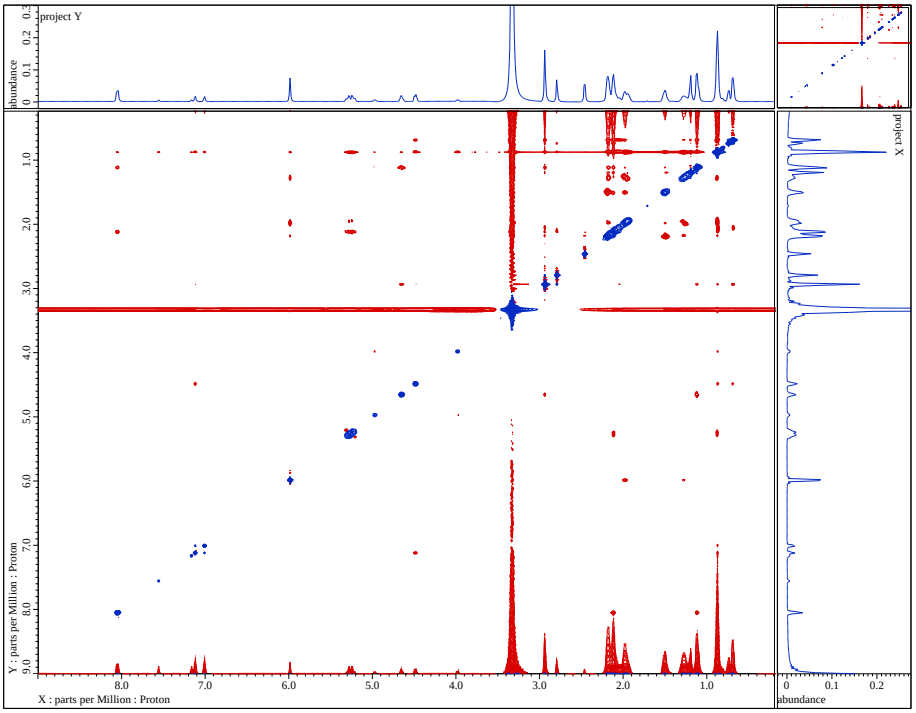


Figure A3.27. NOESY spectrum of **6** (500 MHz, DMSO-d_6).

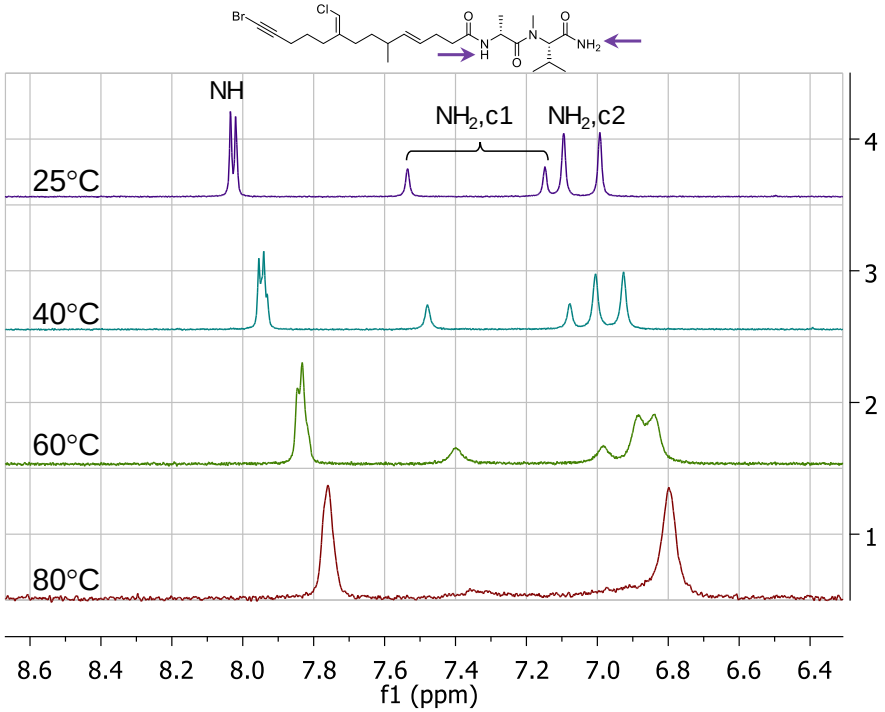


Figure A3.28. ^{1}H-NMR spectrum of **6** in DMSO-d_6 under increasing temperature. Terminal amide peaks labeled; c1 indicates conformer 1 and c2 indicates conformer 2

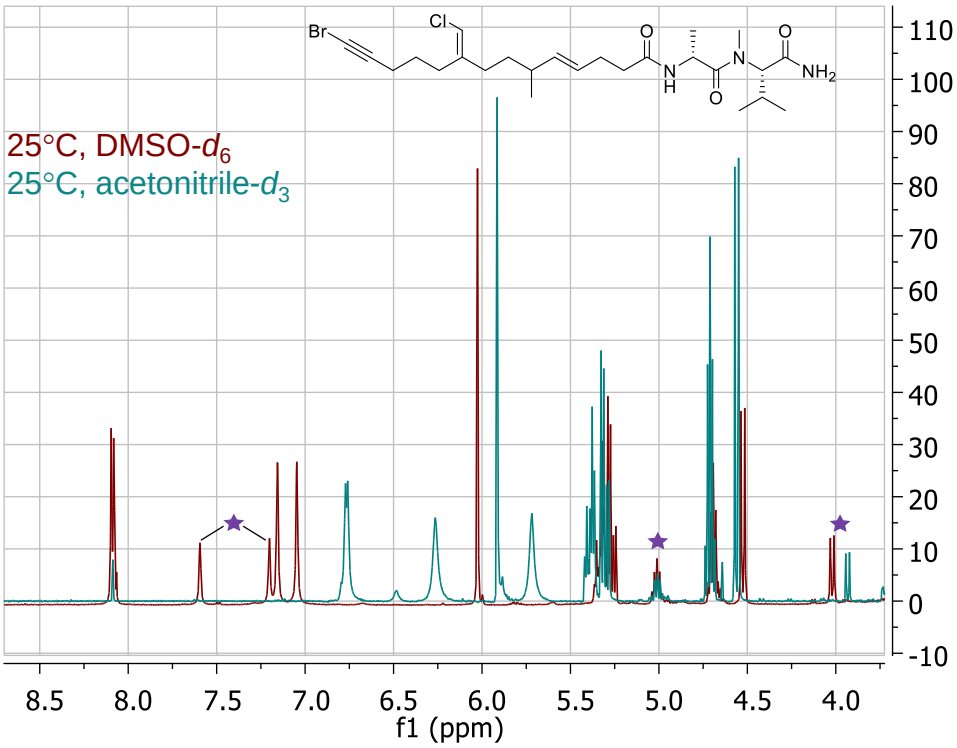


Figure A3.29. ^{1}H-NMR spectral comparison of **6** in DMSO-d_6 and acetonitrile-d_3. Minor conformer peaks indicated with purple stars.

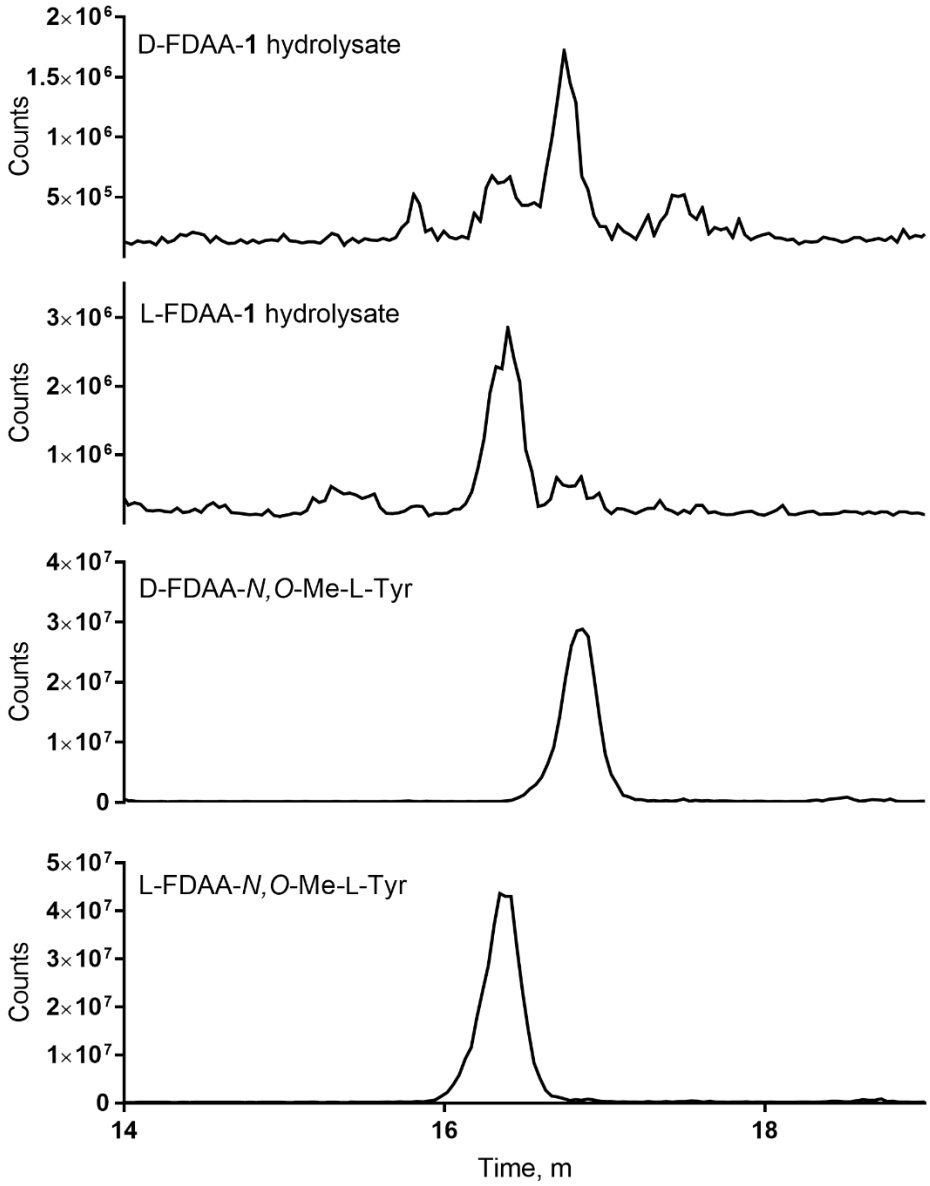


Figure A3.30. LC/MS chromatogram depicting Marfey's analysis of **1**. $[M+H]^+$ = 462 for each plot corresponding to L- and D-FDAA reaction with hydrolysis products of *N*,*O*,*O*-Me-L-Tyr, or **1**.

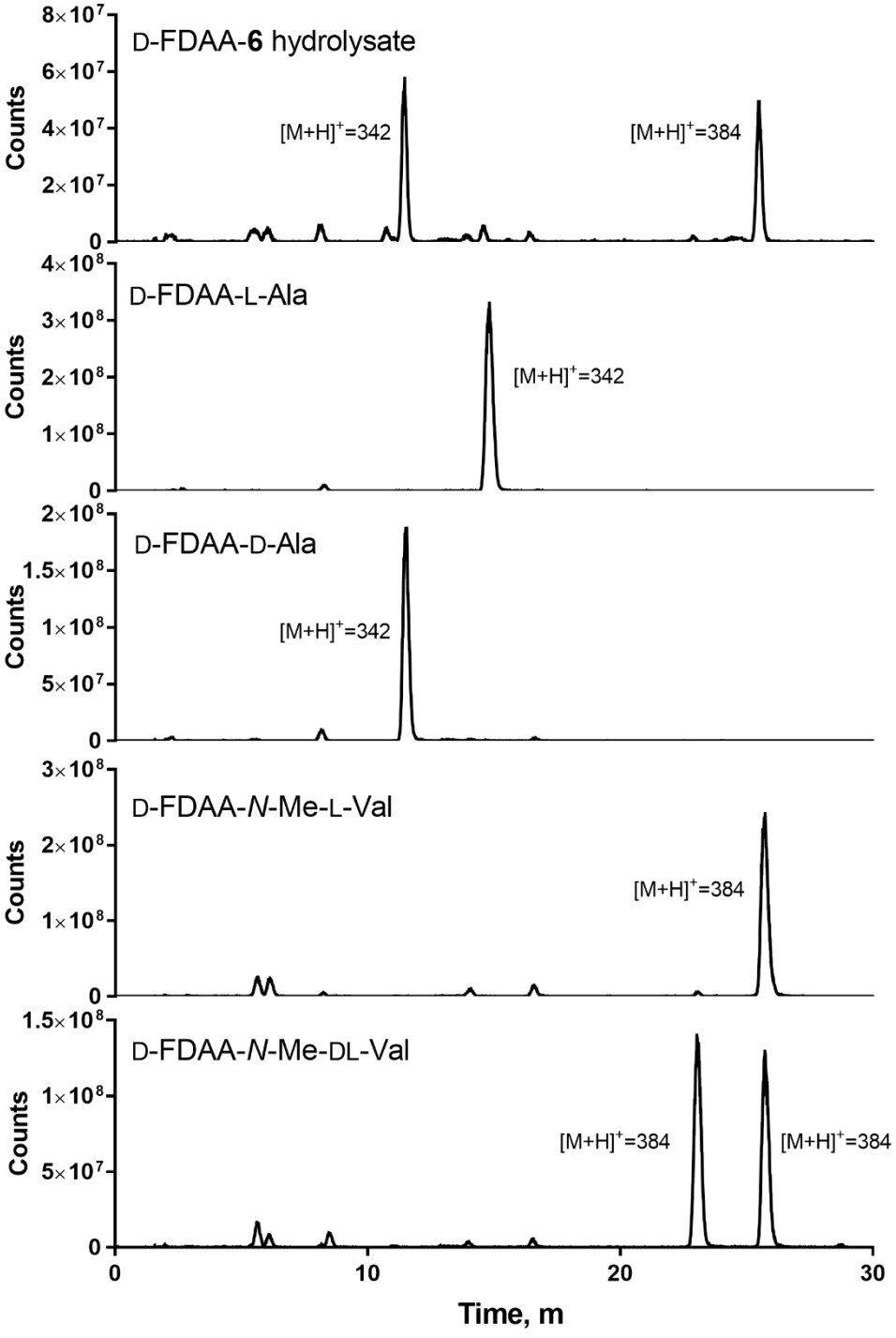


Figure A3.31. LC/MS chromatogram depicting Marfey's analysis of **6**. $[M+H]^+$ values indicated next to each peak, corresponding to D-FDAA reaction with Ala (342) or *N*-Me Val (384). Hydrolysis of **6** reacted with D-FDAA shown on top plot.

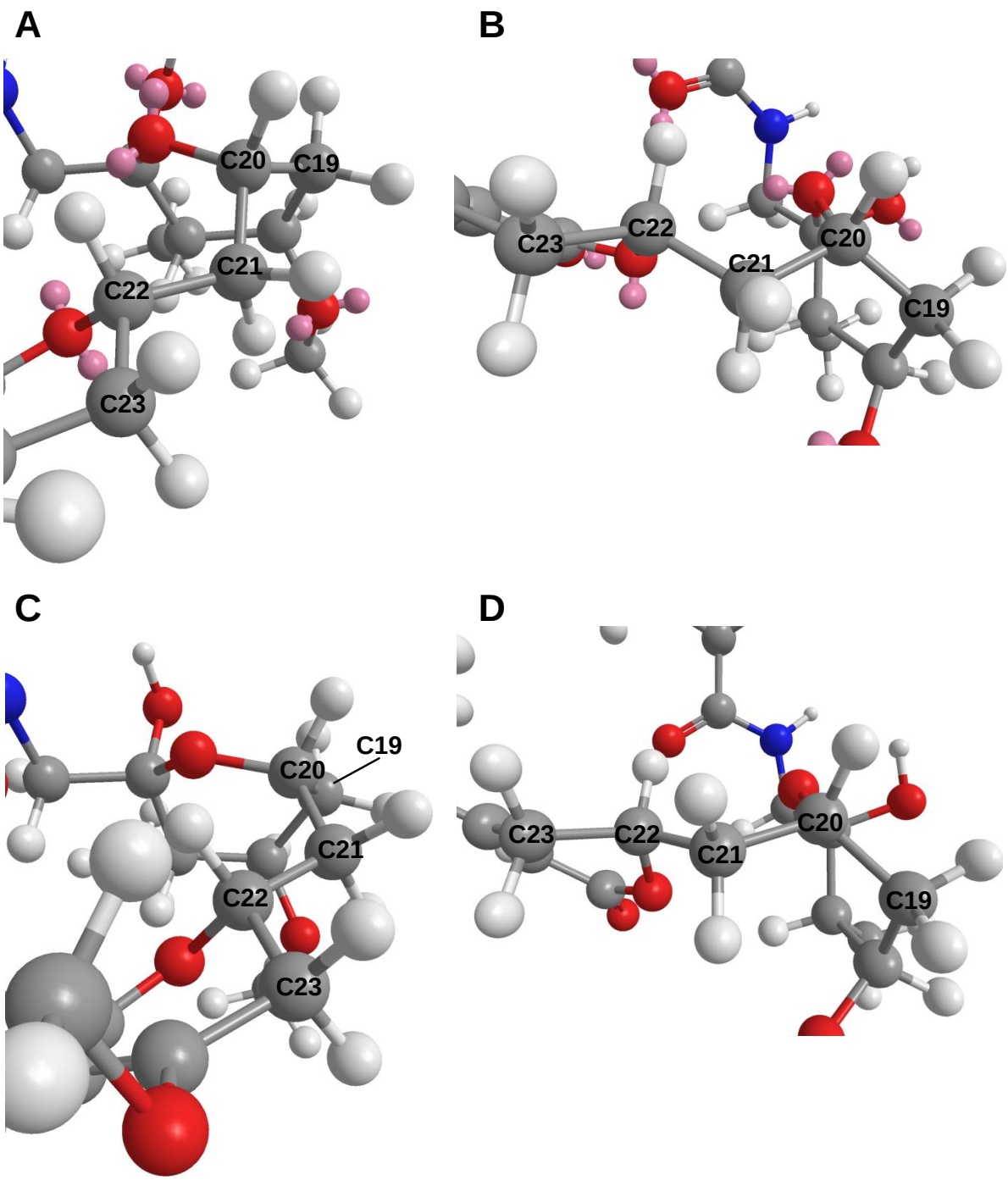


Figure A3.32. Molecular modeling analysis of C22 stereocenter of **3** in *R* position. A) MM2 force field, face view. B) MM2 force field, profile. C) MMFF94 force field, face view. D) MMFF94 force field, profile. Axial orientation and R configuration of C22 is supported by NOESY in contrast to S configuration. (Figure A3.15, A3.20-21, A3.33).

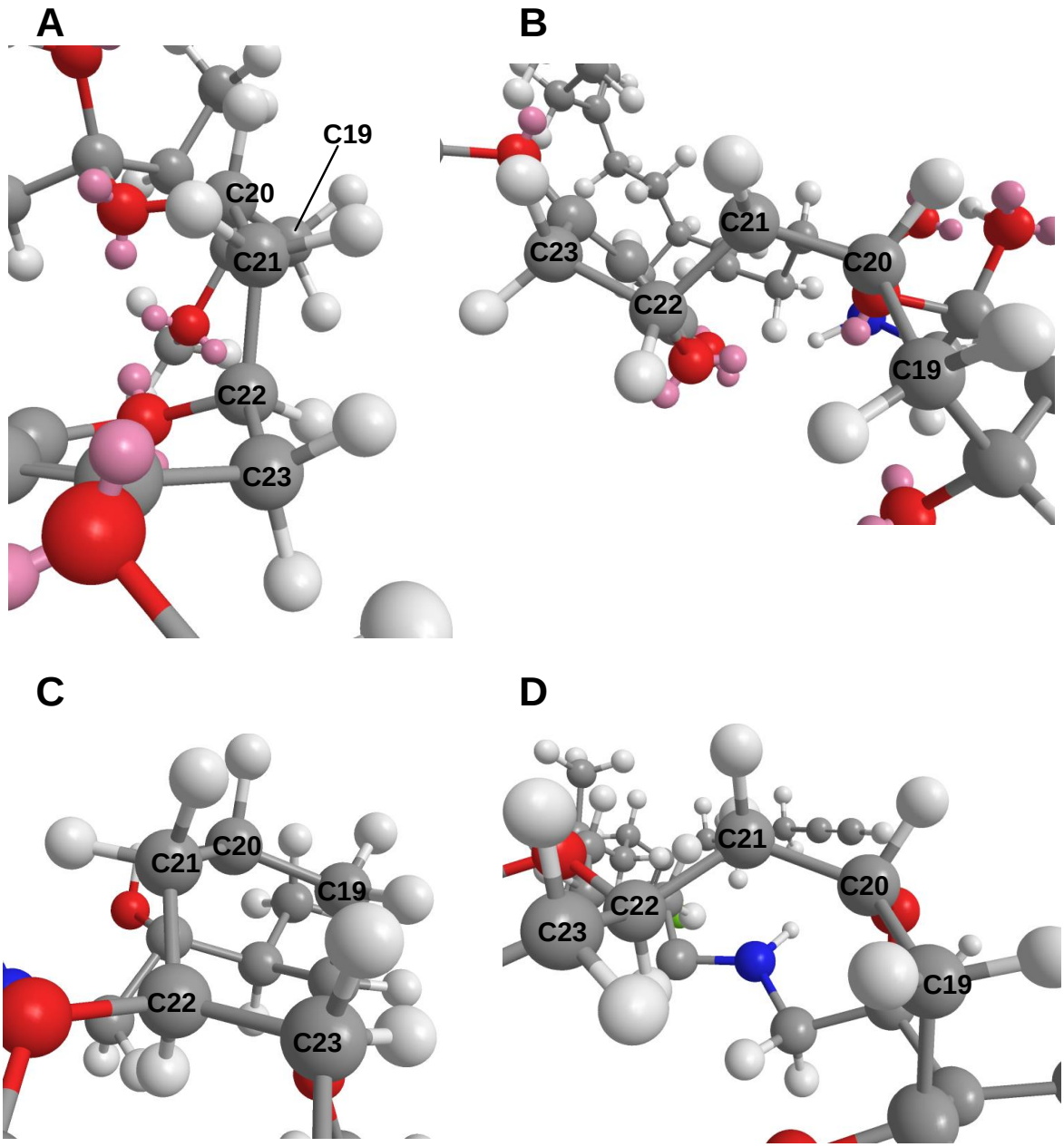


Figure A3.33. Molecular modeling analysis of C22 stereocenter of **3** in *S* position. A) MM2 force field, face view. B) MM2 force field, profile. C) MMFF94 force field, face view. D) MMFF94 force field, profile. Axial orientation and R configuration of C22 is supported by NOESY in contrast to S configuration; particularly in trans positioning of C20-C22. (Figure A3.15, A3.20-21, A3.32).

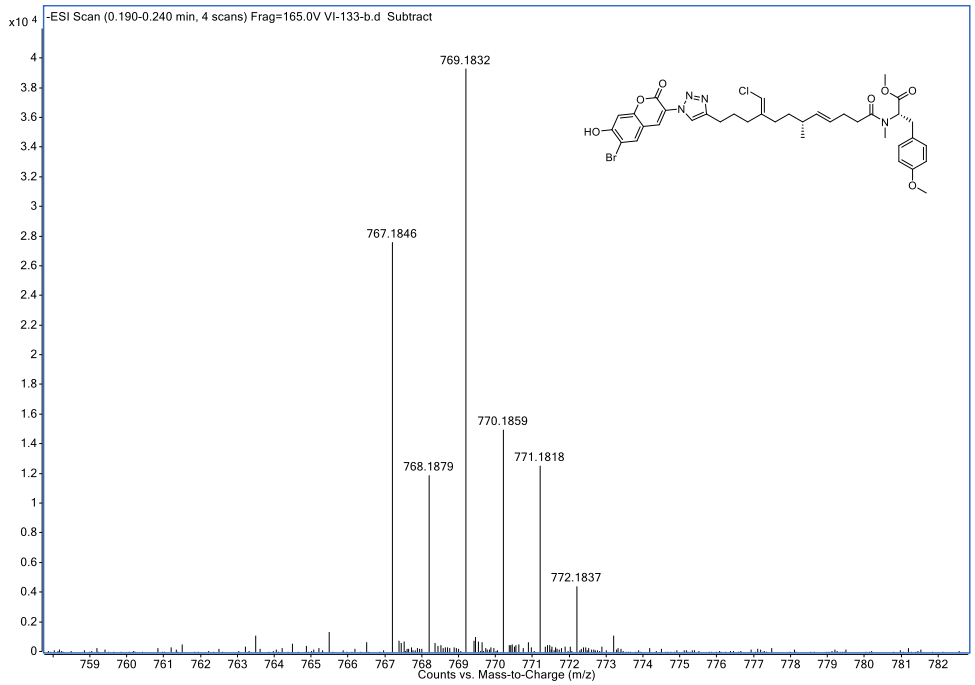


Figure A3.34. HRMS spectrum of **12**+**1**.

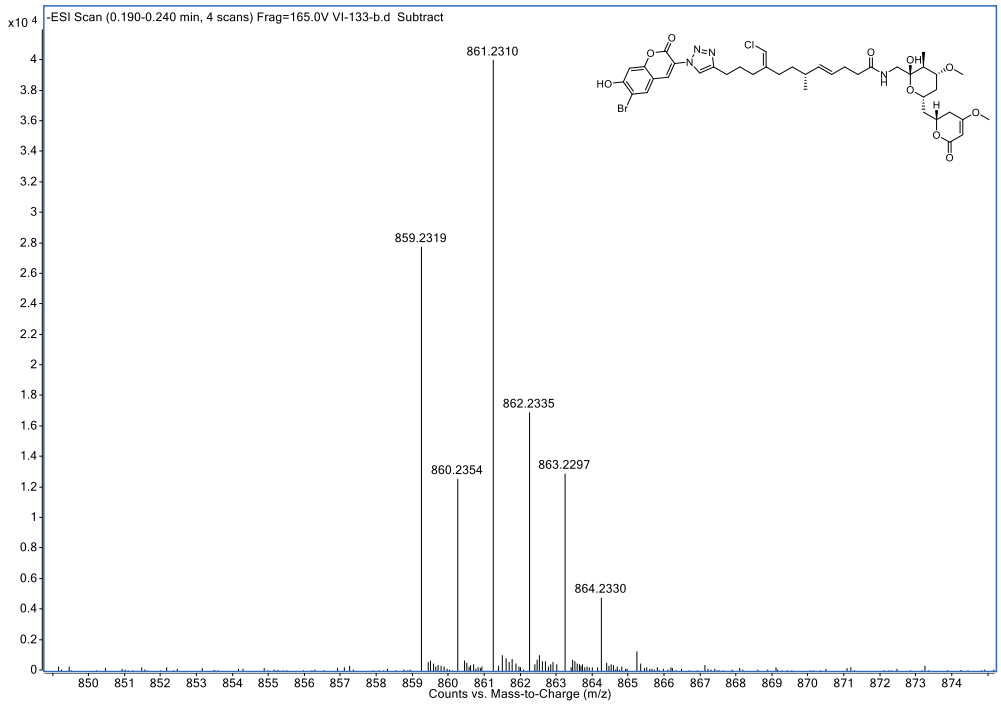


Figure A3.35. HRMS spectrum of **12**+**3**.

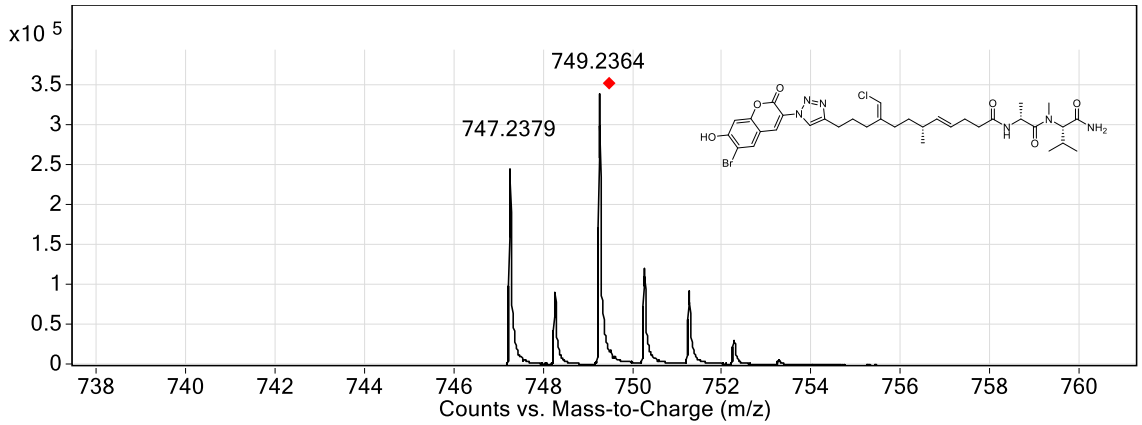


Figure A3.36. HRMS spectrum of **12+5.**

3.7 Acknowledgments

I would like to acknowledge J. G. Sanders, G. Humphrey, J. Gaffney, R. A. Salido Benitez, J. J. Minich, C. Brennan, K. Sanders, and R. Knight for their expertise in library preparation and short-read genome sequencing. I would like to acknowledge E. Ricciardelli, B. A. Henson, and K. Jepsen for their expertise in library preparation for long-read genome sequencing.

We thank E. Glukhov, P. Kanjanakantorn, M. Kissick, B. Ni, and A. M. Hoskins for their contributions to cyanobacterial culturing, bioassay and extraction, and B. Miller for harvesting and collection. We thank C. Leber for his assistance in running Marfey's analysis. We would like to acknowledge Y. Su, L. Gross, A. Mrse and B. Duggan for analytical expertise. We thank M. Pierce and T. Murray for bioassay testing. This work was supported by NIH GM107550, NIH GM118815 to LG and WHG. NAM was supported by NIH training grant in Marine bio-technology and Microbiome and Microbial Sciences Initiative Graduate Student Fellowship.

This chapter, in full, has been submitted for publication of the material. Moss, Nathan A.; Seiler, Grant; Leão, Tiago F.; Castro-Falcòn, Gabriel; Gerwick, Lena; Hughes, Chambers; Gerwick, William H. "Nature's combinatorial biosynthesis produces vatiamides A-F." 2019. The dissertation author is the primary investigator and author of this material.

Chapter 4: *tert*-butyl biosynthesis in the apratoxin A pathway from *Moorea bouillonii* PNG 5-198

4.1 Abstract

The *t*-butyl group is a four-carbon trimethyl alkyl substituent found in numerous natural products of disparate origin and biosynthesis. One such *t*-butyl group is found in apratoxin A, a potent anticancer compound isolated from the marine cyanobacterium *Moorea bouillonii*. This moiety is hypothesized to derive from loading module AprA in the biosynthetic pathway of apratoxin A. Sequence alignment and comparison to related loading modules inferred AprA activity to catalyze three sequential methylations of a thioester-bound intermediate using S-adenosyl methionine (SAM); the only such module found in a cultured organism thus far in a type I PKS/NRPS context. The biosynthesis of the *t*-butyl group in cultured *Moorea bouillonii* PNG5-198 (PNG) was analyzed in separate experiments via biochemistry of domain constructs from AprA. In order to validate such findings in an *in vivo* context, *t*-butyl biosynthesis in PNG was assessed by stable isotope labeled feeding studies. [1-^{13}C]propionate and [methyl-^{13}C]methionine were fed to cultures of PNG, and apratoxin A purified from each was compared to apratoxin A from unlabeled cultures via ^{13}C-NMR. A 99% increase in signal was observed for the *t*-butyl moiety of apratoxin A in the sample fed [methyl-^{13}C]methionine, while no increase in signal was observed in the labeled propionate-fed sample. These findings added an *in vivo* complementation to in-depth biochemistry experiments that clarified the full biosynthetic mechanism of AprA, expanding our knowledge of cyanobacterial loading schemes.

4.2 Introduction

4.2.1 *tert*-butyl group in natural products and pharmacology

Natural selection often imbues organisms with proteins that incorporate certain functional groups into molecules they produce for its reactivity or structural properties. One such intriguing functional group found in nature is the *t*-butyl group. Owing to its steric bulk and nonreactivity, the *t*-butyl group is frequently used in synthetic chemistry as a protecting element in the form of the *tert*-butyloxycarbonyl (BOC) group.[230] This functionality has the effect of preventing unwanted reactions on

centers susceptible to reactivity, such as amines, as well as fostering directed molecular conformations.[231] Additionally, *t*-butylated molecules and the non-standard amino acid *tert*-leucine (Tle) have also been incorporated into chiral catalysts, enabling asymmetric oxidation and cross-coupling reactions.[232] Tle has been incorporated into synthetic bioactive molecules such as VX-950, a hepatitis C protease inhibitor, and several hydroxamic acids bearing a Tle substituent were assessed to target MMP-3, -6 and -9 matrix metalloproteinases, a cancer target.[233–235]

In the natural world, the *t*-butyl group is found in several natural products and has been isolated in molecules from a variety of different phyla. Frequently it is encountered as Tle and incorporated into peptide chains. The hemiasterlins, halicylindramide B, milnamide A, the criamides, as well as massive polytheonamide have all been isolated from marine sponges and feature one or more Tle residues.[236–240] The *t*-butyl group is also found as a modification to terpene-derived steroid molecules from sponges and their endosymbionts, including epipolasterol, four topsentinol analogs, and several of the erylosides.[241–243] Additional terpene-derived *t*-butyl containing molecules are found terrestrially, such as in the ginkgolides, bilobalide, and butyrolactol A.[244–246] Coumarin-derived swietenone and a similar analog of pranferin from plant roots are another interesting set of molecules bearing *t*-butyl moieties, which contain a ketone and derive from a terpene-modified phenylpropanoid core.[247–249] The PKS-derived bryostatins, protein kinase C inhibitors and anticancer agents, feature a pivalic acid ester in several analogs (Figure 4.1).[250,251]

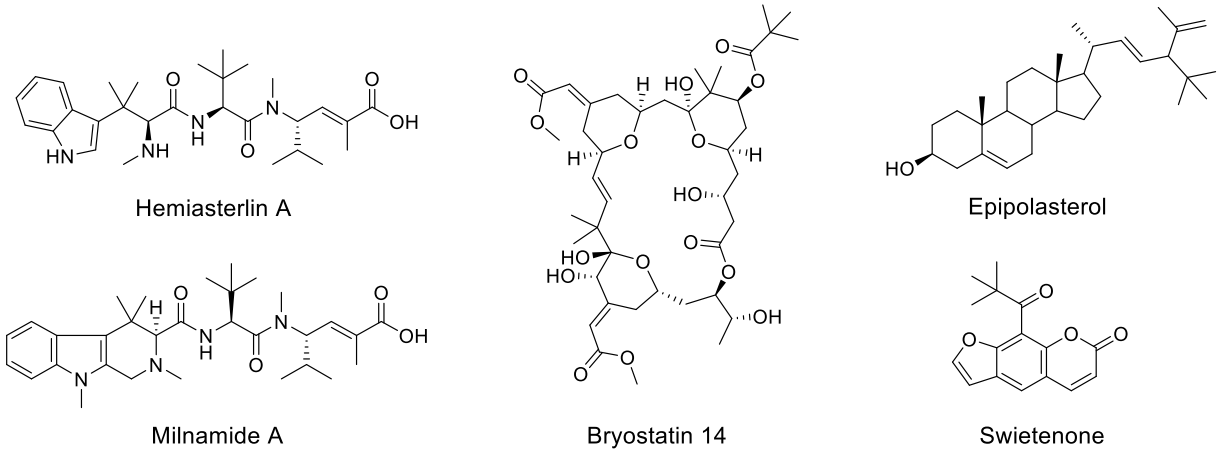


Figure 4.1. *t*-butyl containing molecules and their sources. Hemiasterlin A; sponge *Hemiasterella minor*, milnamide A; sponge *Auletta* sp., bryostatin 14; bacterium *Candidatus Endobugula sertula,* epipolasterol; sponge *Epipolasis* sp., swietenone; plant *Chloroxvlon swietenia* DC.

The presence of distinctly bulky Tle, alkanyl, and pivaloyl *t*-butyl moieties infers a steric, rather than reactivity-based contribution to the pharmacology of the molecules in which they appear. Studies performed on the ginkgolides indicated that the *t*-butyl moiety may exclusively be responsible for wedging the molecule into position to bind to its membrane target, the platelet-activating factor receptor.[252,253] Structural analysis of small molecules containing Tle found that the *t*-butyl moiety imbued a preferred conformation due to increased steric bulk and structural rigidity, while studies of peptide substitutions of small molecule hepatitis C protease inhibitors, as well as cyanobacterially derived palmyrolide indicated that Tle contributed to a protection of adjacent reactivity-prone carbonyls from protease activity.[231,234,254,255] Generally, the distribution and pharmacology of the molecules which contain *t*-butyl groups makes them interesting topics for further study.

4.2.2 *t*-Butyl groups in cyanobacteria

An outsized contribution to the pharmacopeia of *t*-butyl containing natural products in the marine realm comes from filamentous cyanobacteria of the genus *Moorea* (formerly *Lyngbya)* and *Okeania* (Figure 4.2). Several molecules bearing the pivaloyl group have been described, and include several analogs of closely related structures. Two neurotoxic antillatoxin analogs, cyclic lipopeptides featuring a terminal *t*-butyl on a branched dienyl hydrocarbon chain, were isolated from a collection of *Moorea producens* from Curaçao.[256,257] Laingolide and madangolide, two small, lightly modified cyclic hydrophobic amides, were isolated from a *"Lyngbya"* species in Papua New Guinea and contain

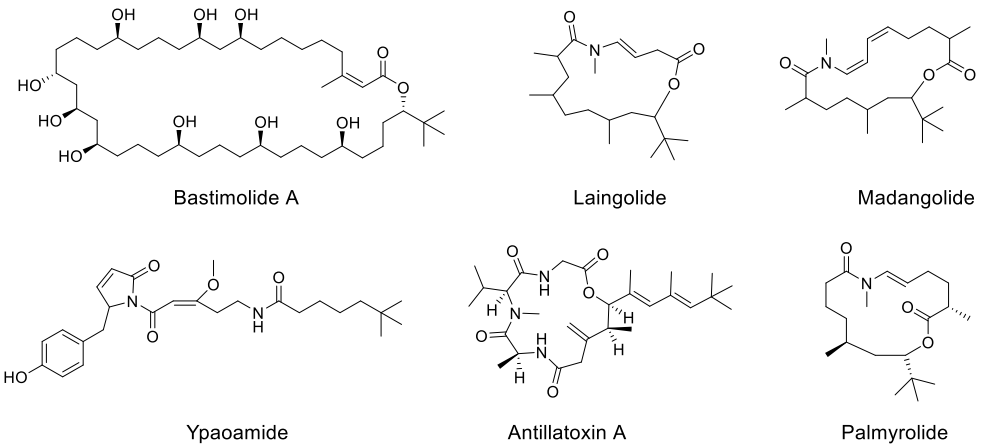


Figure 4.2. *Tert*-butyl containing compounds isolated from filamentous cyanobacteria.

a *t*-butyl adjacent to a cyclic ester bond.[258] Antimalarial compounds bastimolides A and B feature an initial *t*-butyl in a similar motif, attributed to *Okeania* but likely truly sourced from a *Leptolyngbya*.[198,259] Additional compounds include palmyrolide, a sodium channel blocker, the anti-Chagas Disease janadolides from *Okeania*, and ypaoamide, from a Guamanian "*Lyngbya"* collection.[255,260,261] To date, the most studied *t*-butyl containing cyanobacterial compound is the potent anticancer compound apratoxin A, originally isolated from a Guamanian environmental collection of *Lyngbya (Moorea)*, and later found to be produced by a culture of *Moorea bouillonii* PNG5-198 from Papua New Guinea, along with apratoxin B.[201,262] Subsequently, eight additional analogs, B-H, and A-sulfoxide, have been isolated since the original molecule was described, with all but apratoxin C containing a terminal *t*-butyl or pivaloyl group (Figure 4.3).[263–267] Apratoxin A was found to be a potent cytotoxin against several human cancer cell lines

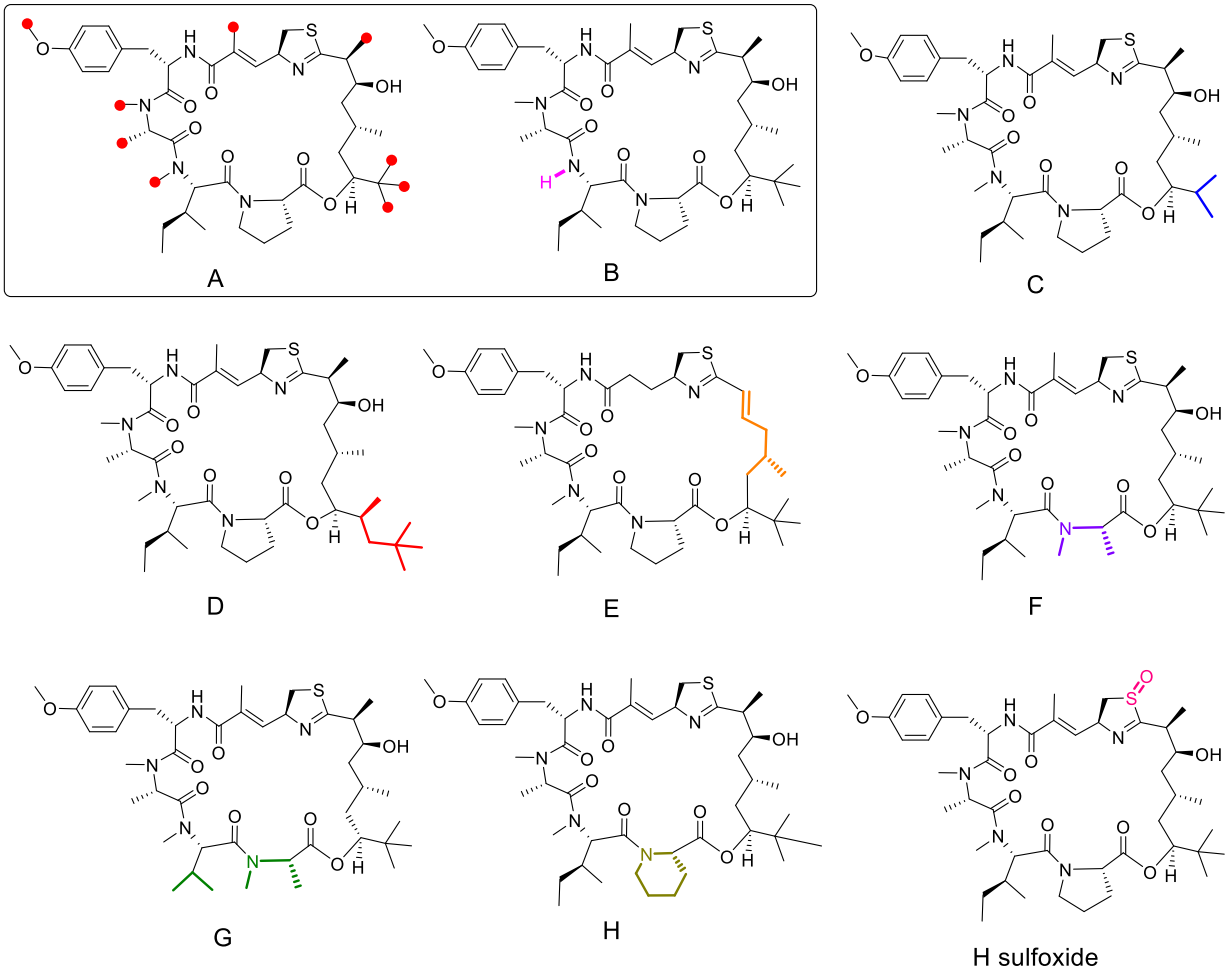


Figure 4.3. Analogs of apratoxin. Predicted SAM-derived methyl carbons highlighted with red circles. Analogs A and B, isolated from *Moorea bouillonii* PNG5-198 surrounded by box.

including H-460 and HCT-116. The mechanism of action was determined to be facilitated by luminal inhibition of the Sec61 protein, which is a complex that facilitates translocation of proteins into and out of the endoplasmic reticulum (ER).[268,269]

4.2.3 Biosynthesis of *t*-butyl moiety in Tle and terpenoids

Though a moderate number of molecules have been isolated and contain the *t*-butyl moiety, the number of biosynthetic pathways that have been determined for these molecules is far sparser, due to the only recent availability of inexpensive genome sequencing technology. One such molecule is polytheonamide, isolated from sponge *Theonella swinhoei*, the biosynthetic pathway of which was determined by metagenomic sequencing of a microbial consortia from the organism.[270] Polytheonamide is a RiPPs-derived methylated polypeptide that was found to create membrane pores as its mode of toxicity. The

study of PoyC protein, in the biosynthesis of Tle groups within the molecule, uncovered the use of a cobalamin, or B12-dependent reaction, to generate Tle, thus paving the way for understanding the biosynthesis of this *t*-butyl substituent (Figure 4.4A).[271,272]

Terpenoids are another molecular type to feature *t*-butyl groups and for which there is a suggested biosynthetic route. In terpene and sterol biosynthesis, the generation of a carbocation followed by methyl and hydride shifts is quite common, and is a primary driver of terpene structural diversity. The low sequence homology and steric differences of the active site of terpene synthases may initiate unique and unexpected mechanisms. It has been suggested that in the biosynthesis of t-butylated sterols, a SAM methylation followed by deprotonation and methyl shift results in *t*-butylation (Figure 4.4B).[273,274] A separate route involving a cyclopropyl ring intermediate has been proposed for the biosynthesis of swietenone (Figure 4.4C).[247]

In cyanobacteria, all of the structures characterized to date likely originate from type I PKS/NRPS pathways, with the exception of bastimolide, thought to originate from a *trans*-AT PKS pathway. While it is possible that some *t*-butyl containing molecules originally assigned to sponges such as hemiasterlin or epipolasterol may eventually be known to come from cyanobacterial endosymbionts, to date no such formal assignment has been made. Thus far, neither Tle-containing peptides, *t*-butyl

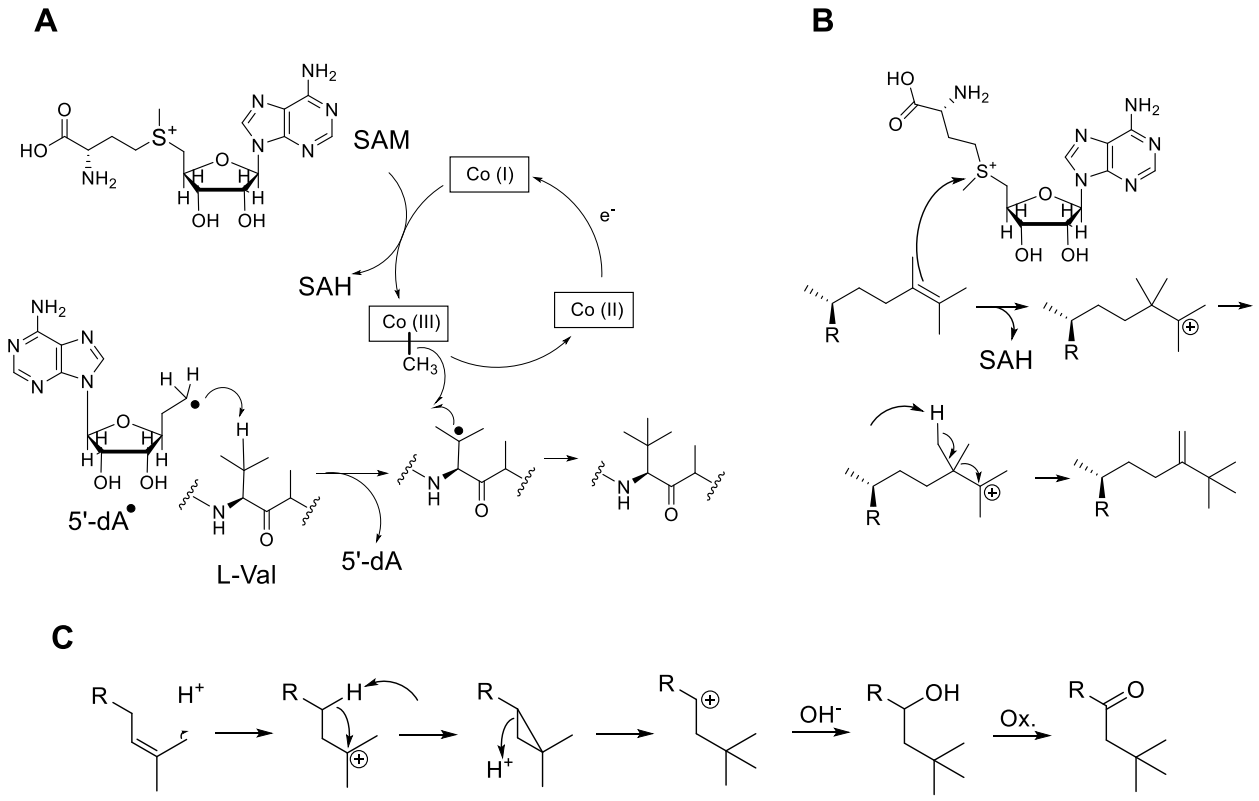


Figure 4.4. Biosynthesis of t-butyl groups. A) PoyC catalyzes cobalamin-dependent radical SAM mechanism (scheme adapted from Benjdia et al., 2017). B) Methyl shift in t-butylated sterols (scheme adapted from Giner and Djerassi, 1991). C) Cyclopropyl ring intermediate proposed in swietenone biosynthesis (scheme adapted from Mujumdar et al., 1975)

modified sterols, or any terpene-derived molecules containing *t*-butyl groups have been isolated from cyanobacteria. Indeed, to date the apratoxin A pathway remains the only published type I PKS/NRPS route to *t*-butyl biosynthesis from cyanobacteria.

4.2.4 Stable isotope labeled feeding experiments to probe *t*-butyl biosynthesis in apratoxin A

AprA is the initial module found upstream of a series of PKS and NRPS modules that comprise the apratoxin A pathway. Initially, limited biochemical or chemical-biology experiments were able to be performed on the entire module owing to protein solubility and purification issues. Therefore, we decided to approach the issue using stable isotope labeled feeding experiments with live cultures of *Moorea bouillonii* PNG5-198 in order to understand possible biosynthetic routes using AprA. Generally, in this method ^{12}C atoms within apratoxin A are replaced by stable and NMR-active ^{13}C isotopes by way of

incorporation of individual molecular building blocks. Possible biosynthetic units may thus be investigated systematically by observation of enrichment compared to molecules without any enrichment, the site specificity of which is conveniently determined be ^{13}C NMR analysis.

4.3 Results

4.3.1 Phylogeny and domain organization of AprA

The apratoxin A pathway was initially investigated and characterized using multiple displacement amplification (MDA) of an isolated single cell of *Moorea bouillonii* PNG5-198. PCR to bridge contigs containing PKS and NRPS contigs ultimately revealed the apratoxin A pathway, which was found to consist of three PKS modules, a β-branching methylation cassette, and four NRPS modules incorporating cysteine, tyrosine, isoleucine, and finally proline, before a condensation domain enables cyclization.[201] The pathway is initiated by the AprA protein, an ~1100 amino acid module that bears little homology to known initiation modules and PKS modules. However, bioinformatics tools allowed for prediction of a conserved adapter region of unclear function (AR), glycine-*N*-methyltransferase-(GNMT)-like domain ("MT1"), followed by a domain with some similarity to the trans-acylating/decarboxylating GNAT domain in curacin A biosynthetic initiation, as well as a CMT ("MT2") and terminal ACP. Similar domain arrangements in other known pathways produce starter molecules with zero (CurA, curacin A), one (SxtA, saxitoxin), or two (GphF, gephyronic acid) methyl groups bonded to C2 of acetate, as well as the predicted *t*-butyl-producing BryX, although the bryostatin producer is not yet cultivable (Figure 4.5).[161]

Here, we evaluated several possibilities for the generation of the *t*-butyl group in apratoxin A, with the GNMT and CMT domains implying SAM as the source of methyl groups to thioester-bound malonate. Another potential route for one of the methyl groups could involve decarboxylation of ACP-bound methylmalonyl CoA; however propionyl-CoA carboxylase, the enzyme responsible for biosynthesis of methylmalonyl CoA, was not found in the PNG genome. In addition, no cyanobacteria is known to biosynthesize methylmalonyl CoA. Therefore we deemed that it far less likely but nonetheless

possible by some unknown mechanism within AprA that C3 of propionate may represent one of the three methyl groups. The use of other starter units such as derivatives of valine were considered possibilities to

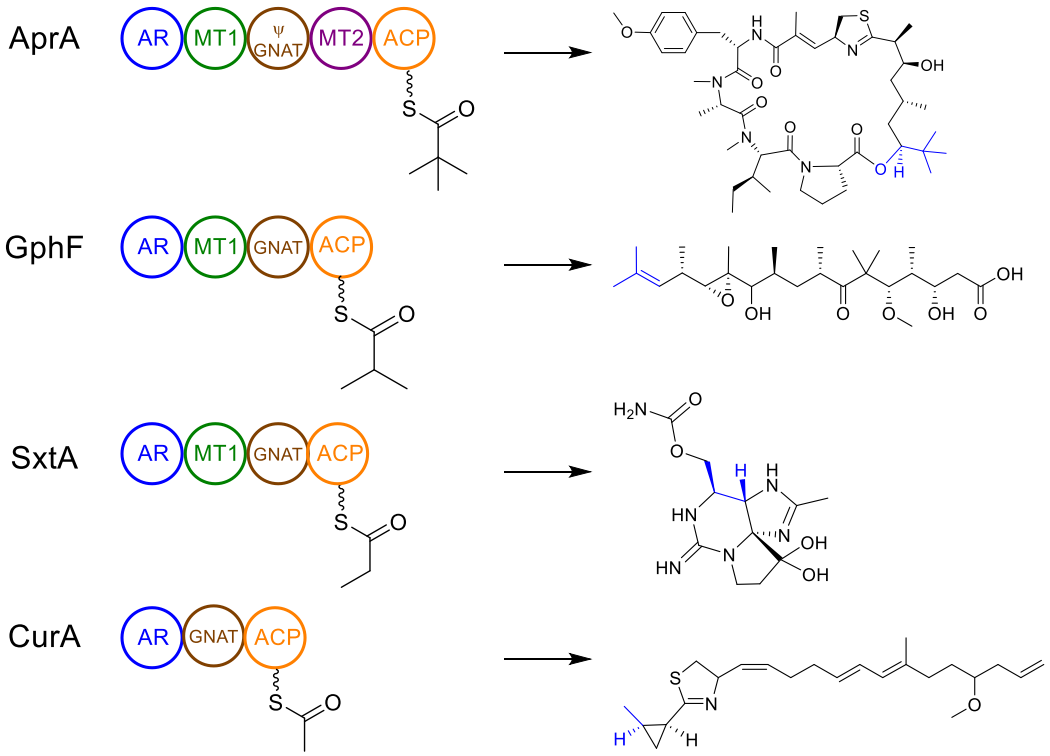


Figure 4.5. AprA and related proteins.

load AprA; however, stable isotope labeled feeding studies of the biosynthesis of gephyronic acid, which features a similar loading module minus the MT2, found that the methyl donors to the starter unit were SAM, as opposed to valine.[275] Therefore, we fed labeled [methyl-^{13}C]methionine and [1-^{13}C]propionate to separate live cultures of *Moorea bouillonii* PNG5-198 in order to probe the methyl origins of *t*-butyl in apratoxin A. (Figure 4.6).

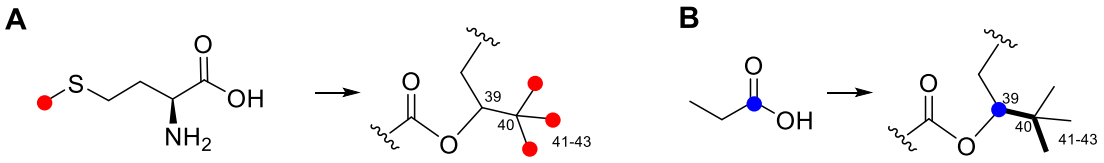


Figure 4.6. Labeling scheme of *t*-butyl group of apratoxin A. A) feeding of [^{13}C-methyl]methionine. Label indicated in red circle. Predicted incorporation in blue circle. B) feeding of labeled [1-^{13}C]propionate, labeled indicated in blue circle. Predicted carbon atoms from propionate depicted in bold bond.

4.3.2 Feeding, extraction, and purification.

Moorea bouillonii PNG5-198 was scaled up from finger-sized clumps of cyanobacterial filaments in SW BG-11 media over a period of 10-12 weeks. Cyanobacterial filaments reached optimal growth rate in 250 mL culture flasks and were subsequently transferred to 1 L of media in two 2.8 L glass Fernbach flasks for each feeding condition as well as the unlabeled control. Cultures were concentrated to 0.75 L in one 2.8 L Fernbach flask and subsequently fed with labeled substrate over 15 days. Biomass was harvested and extracted, then apratoxin A was purified as detailed in Experimental Methods (4.5).

4.3.3 ^{13}C-NMR of [1-^{13}C]propionate – fed batch

In the case of [1-^{13}C]propionate fed batches, methanol-d_4 was used as the NMR solvent, as the C39 peak of apratoxin A is found at δ_C=77.1 and thus would be buried under $CDCl_3$ solvent peak in the ^{13}C-NMR spectrum. As such, peak identity in methanol-d_4 was confirmed by HSQC and HMBC experiments (Table A4.1, Figure A4.9-A4.10). All peaks were identified and corresponded to similar

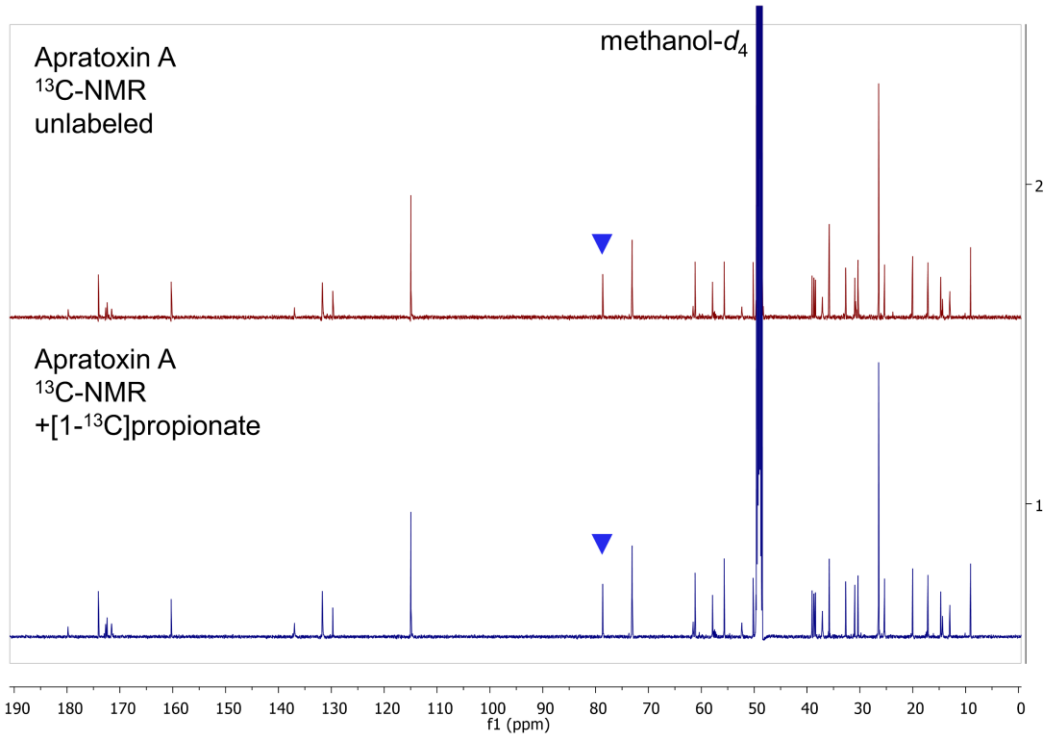


Figure 4.7. Comparison of ^{13}C-NMR spectra of apratoxin A fed with labeled propionate (bottom) vs unlabeled (top). (125 MHz, methanol-d_4). C39 marked with blue triangle.

Table 4.1. Enrichment of apratoxin A in cultures of *M. bouillonii* PNG 5-198 supplemented with [1-^{13}C]propionate. Final enrichment values are derived by normalization of enriched ^{13}C-NMR resonances to five different carbon atoms. Calculation steps are detailed in the Experimental Methods section 4.5.3. The enrichment value of C39 is bolded/italicized/red in column 21 and 22.

Column # →																						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Biosynthetic		Integral of unlabeled	Normalization factor using:					Integral of enriched	Normalized integral of enriched sample using:					Enrichment compared to:					Average %	Fold
C#	source	ppm	resonance	C7	C13	C20	C30	C37	resonance	C7	C13	C20	C30	C37	C7	C13	C20	C30	C37	enrichment	Change
1	Pro	174.1	45.5	0.5	0.9	0.9	0.2	0.3	56.0	44.5	63.0	50.2	50.6	52.1	25.8	-11.2	11.6	10.7	7.4	8.8	1.1
2	Pro	61.2	91.4	0.9	1.8	1.7	0.4	0.5	97.9	89.6	126.8	100.9	101.8	104.9	9.3	-22.8	-3.0	-3.8	-6.6	-5.4	0.9
3	Pro	29.8	2.2	0.0	0.0	0.0	0.0	0.0	2.9	2.1	3.0	2.4	2.4	2.5	37.2	-3.1	21.8	20.8	17.2	18.8	1.2
4	Pro	N.P.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Pro	N.P.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	Ile	172.7	32.0	0.3	0.6	0.6	0.2	0.2	38.7	31.4	44.4	35.3	35.6	36.7	23.5	-12.8	9.6	8.7	5.5	6.9	1.1
7	Ile	57.9	97.8	1.0	1.9	1.8	0.5	0.6	95.8	95.8	135.6	107.9	108.8	112.1	0.0	-29.4	-11.3	-12.0	-14.6	-13.4	0.9
8	Ile	31.0	65.8	0.7	1.3	1.2	0.3	0.4	79.9	64.4	91.2	72.6	73.2	75.4	24.0	-12.4	10.1	9.2	5.9	7.4	1.1
9	Ile	25.3*	173.2	1.8	3.4	3.2	0.8	1.0	198.6	169.6	240.2	191.1	192.7	198.6	17.1	-17.3	3.9	3.0	0.0	1.3	1.0
10	Ile	9.1	79.2	0.8	1.6	1.5	0.4	0.5	91.9	77.6	109.9	87.4	88.2	90.8	18.5	-16.3	5.1	4.3	1.2	2.5	1.0
11	Ile	14.7	73.3	0.7	1.4	1.4	0.4	0.4	85.5	71.8	101.6	80.9	81.6	84.0	19.1	-15.9	5.7	4.8	1.7	3.1	1.0
12	SAM	30.4	93.6	1.0	1.8	1.8	0.5	0.5	91.5	91.7	129.8	103.3	104.2	107.4	-0.2	-29.6	-11.5	-12.2	-14.8	-13.7	0.9
13	Ala	171.6	51.0	0.5	1.0	1.0	0.2	0.3	70.7	50.0	70.7	56.3	56.8	58.5	41.6	0.0	25.7	24.6	20.9	22.6	1.2
14	Ala	61.6	58.7	0.6	1.2	1.1	0.3	0.3	71.3	57.5	81.4	64.8	65.3	67.3	24.0	-12.4	10.1	9.1	5.9	7.4	1.1
15	Ala	14.4	64.4	0.7	1.3	1.2	0.3	0.4	70.6	63.1	89.3	71.1	71.7	73.9	11.8	-21.0	-0.7	-1.6	-4.5	-3.2	1.0
16	SAM	35.8	82.8	0.8	1.6	1.6	0.4	0.5	86.3	81.1	114.8	91.4	92.2	95.0	6.5	-24.8	-5.5	-6.3	-9.1	-7.9	0.9
17	Tyr	172.4	43.3	0.4	0.8	0.8	0.2	0.3	51.2	42.4	60.1	47.8	48.2	49.7	20.6	-14.8	7.0	6.2	3.0	4.4	1.0
18	Tyr	52.3	61.2	0.6	1.2	1.1	0.3	0.4	84.0	60.0	84.9	67.6	68.1	70.2	40.1	-1.0	24.3	23.3	19.7	21.3	1.2
19	Tyr	37.1	130.5	1.3	2.6	2.4	0.6	0.8	155.9	127.8	181.0	144.0	145.2	149.6	22.0	-13.9	8.2	7.3	4.2	5.6	1.1
20	Tyr	129.7	53.3	0.5	1.0	1.0	0.3	0.3	58.9	52.2	74.0	58.9	59.4	61.2	12.7	-20.4	0.0	-0.8	-3.8	-2.5	1.0
21/25	Tyr	131.7	214.8	2.2	4.2	4.0	1.0	1.2	243.0	210.4	297.9	237.1	239.1	246.3	15.5	-18.4	2.5	1.6	-1.4	0.0	1.0
22/24	Tyr	115.0	206.4	2.1	4.0	3.9	1.0	1.2	215.3	202.1	286.2	227.8	229.7	236.7	6.5	-24.8	-5.5	-6.3	-9.0	-7.8	0.9
23	Tyr	160.3	50.0	0.5	1.0	0.9	0.2	0.3	48.6	48.9	69.3	55.2	55.6	57.3	-0.7	-29.9	-11.9	-12.7	-15.2	-14.1	0.9
26	SAM	55.7	100.0	1.0	2.0	1.9	0.5	0.6	100.0	97.9	138.7	110.4	111.3	114.7	2.1	-27.9	-9.4	-10.2	-12.8	-11.6	0.9
27	Acetate	171.3	4.8	0.0	0.1	0.1	0.0	0.0	4.6	4.7	6.6	5.3	5.3	5.5	-1.5	-30.5	-12.6	-13.4	-15.9	-14.8	0.9
28	Acetate	N.P.	8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	Cys	137.0	58.0	0.6	1.1	1.1	0.3	0.3	80.2	56.8	80.5	64.0	64.6	66.5	41.2	-0.3	25.3	24.2	20.6	22.2	1.2
30	Cys	73.1*	207.4	2.1	4.1	3.9	1.0	1.2	230.8	203.1	287.6	228.9	230.8	237.8	13.6	-19.7	0.8	0.0	-2.9	-1.6	1.0
31	Cys	38.4	76.2	0.8	1.5	1.4	0.4	0.4	94.2	74.7	105.7	84.1	84.9	87.4	26.2	-10.9	12.0	11.0	7.8	9.2	1.1
32	SAM	13.0	72.5	0.7	1.4	1.4	0.3	0.4	87.3	71.0	100.5	80.0	80.7	83.1	23.0	-13.1	9.2	8.3	5.1	6.5	1.1
33	Acetate	179.8	43.7	0.4	0.9	0.8	0.2	0.3	43.7	42.8	60.6	48.2	48.6	50.1	2.1	-27.9	-9.4	-10.2	-12.8	-11.6	0.9
34	Acetate	50.2	76.2	0.8	1.5	1.4	0.4	0.4	92.0	74.6	105.6	84.1	84.8	87.4	23.3	-12.9	9.4	8.5	5.3	6.7	1.1
35	Acetate	73.1*	207.4	2.1	4.1	3.9	1.0	1.2	230.8	203.1	287.6	228.9	230.8	237.8	13.6	-19.7	0.8	0.0	-2.9	-1.6	1.0
36	Acetate	39.1	79.8	0.8	1.6	1.5	0.4	0.5	94.7	78.2	110.7	88.1	88.8	91.5	21.1	-14.5	7.5	6.6	3.4	4.8	1.0
37	Acetate	25.3*	173.2	1.8	3.4	3.2	0.8	1.0	198.6	169.6	240.2	191.1	192.7	198.6	17.1	-17.3	3.9	3.0	0.0	1.3	1.0
38	Acetate	38.7	77.6	0.8	1.5	1.5	0.4	0.4	82.8	76.0	107.6	85.6	86.4	89.0	9.0	-23.0	-3.3	-4.1	-6.9	-5.7	0.9
39	**Acetate**	**78.7**	**87.4**	**0.9**	**1.7**	**1.6**	**0.4**	**0.5**	**102.0**	**85.6**	**121.2**	**96.5**	**97.3**	**100.2**	**19.2**	**-15.9**	**5.7**	**4.9**	**1.8**	***3.1***	***1.0***
40	Acetate	32.7	93.5	1.0	1.8	1.8	0.5	0.5	93.8	91.6	129.7	103.2	104.1	107.3	2.3	-27.7	-9.2	-9.9	-12.6	-11.4	0.9
41-43	SAM	26.5	404.2	4.1	7.9	7.6	1.9	2.3	487.6	395.8	560.5	446.1	449.8	463.5	23.2	-13.0	9.3	8.4	5.2	6.6	1.1
44	SAM	17.1	88.5	0.9	1.7	1.7	0.4	0.5	95.7	86.7	122.7	97.7	98.5	101.5	10.4	-22.0	-2.0	-2.8	-5.7	-4.4	1.0
45	Acetate	20.0	89.7	0.9	1.8	1.7	0.4	0.5	103.2	87.8	124.4	99.0	99.8	102.8	17.5	-17.0	4.2	3.4	0.3	1.7	1.0

Note: N.P. = Obscured. C28 – weak signal - observed in natural abundance, but not enriched sample. C5 obscured by methanol-d_4 solvent peak. Asterisks indicate overlapping ^{13}C-NMR shifts – the enrichment values cannot be distinguished, so both are reported.

shifts in $CDCl_3$. In ^{13}C-NMR samples, peaks were analyzed by integration, as opposed to intensity, as this measure more effectively captures small shifts in signal due to magnetic field perturbations over the course of the run. In comparison of unlabeled apratoxin A and apratoxin A harvested from PNG fed with [1-^{13}C]propionate, no enrichment of the C39 peak was observed; (Figure 4.7, Table 4.1, Figures A4.1-A4.4). Median absolute change in integrated signal was calculated to be 6.6%, while standard deviation of all changes in all signals was calculated to be 6.1%; this indicates that the signal change observed from

this propionate feeding experiment is not distinguishable from noise. The observed change in signal for C39 was 3.1%, within the margin of error, and again, was not indicative of enrichment.

4.3.4 ^{13}C-NMR of [Methyl-^{13}C]methionine – fed batch

^{13}C-NMR data was collected in $CDCl_3$ for apratoxin A fed with [methyl-^{13}C]methionine. (Figure 4.8, Table 4.2, Figures A4.5-A4.8). As methionine is the precursor to SAM, and SAM is the predicted methyl donor in nearly all PKS/NRPS modular methyltransferase domains, it is expected that cultures fed [methyl-^{13}C]methionine should show substantial increases in ^{13}C-NMR signal integral for carbon positions which have incorporated the donor ^{13}C methyl group. In the case of [methyl-^{13}C]methionine-fed

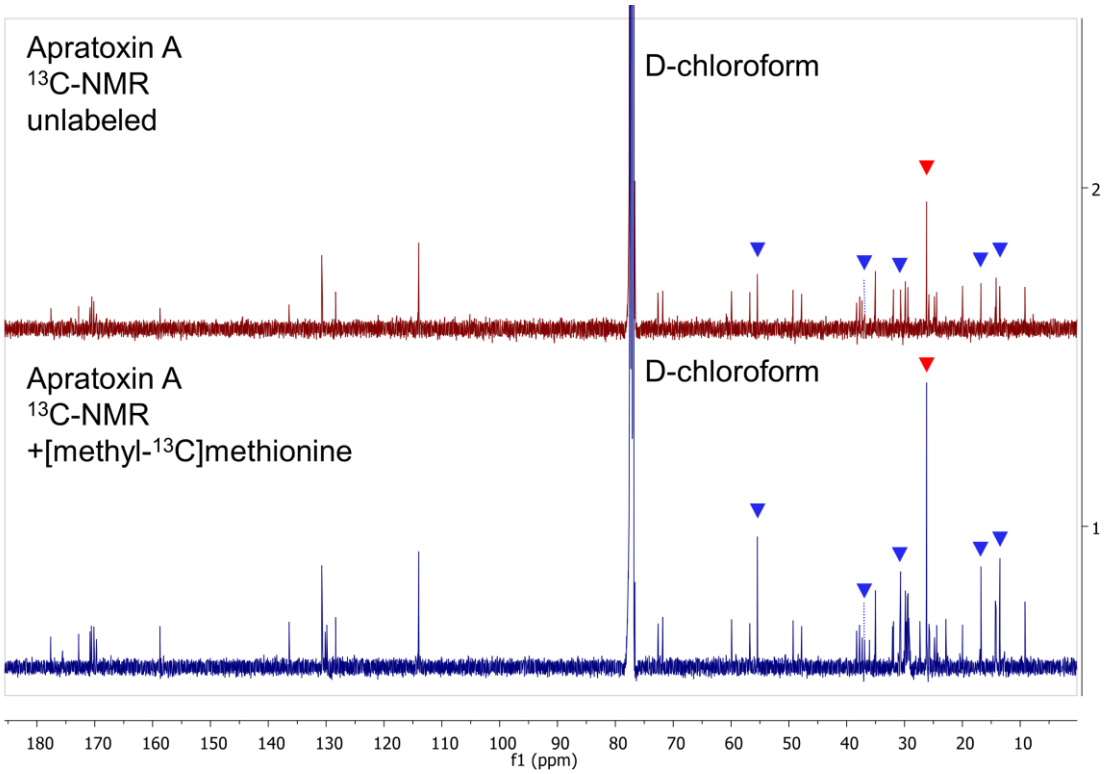


Figure 4.8. Comparison of ^{13}C-NMR spectra of apratoxin A fed with labeled methionine (bottom) vs unlabeled (top). (125 MHz, $CDCl_3$). SAM-enriched carbons marked with blue triangles. C41-43 marked with red triangle.

PNG, significant enrichment of several peaks was observed in the apratoxin A ^{13}C-NMR spectrum. All of the peaks predicted to derive from SAM showed a significant increase in their ^{13}C-NMR signals when compared to the unlabeled version (Figure 4.9). Enrichment percentages for non-*t*-butyl SAM-derived

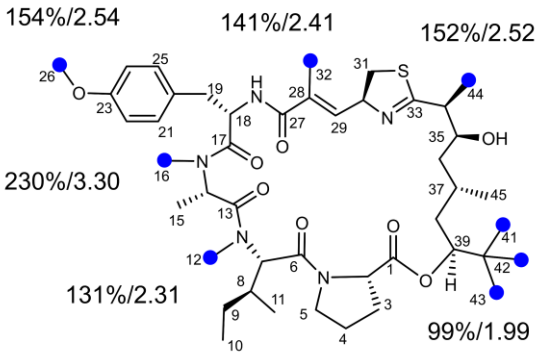


Figure 4.9. Apratoxin A molecule with SAM derived methyl groups indicated in blue circles. Percentage enrichment is displayed next to each SAM-derived carbon, along with fold increase.

Table 4.2. Enrichment of apratoxin A by culture of *M. bouillonii* PNG 5-198 supplemented with [methyl-^{13}C]methionine. Final enrichment values are derived by normalization of enriched ^{13}C-NMR resonances to five different carbon atoms. Calculation steps are detailed in Experimental Methods 4.5.3. SAM-derived enrichment percentages are bolded/italicized/red in column 21 and 22.

Column # →																					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
			Integral of						Integral of		Normalized integral of										
	Biosynthetic		unlabeled	Normalization factor using:					enriched	enriched sample using:					Enrichment compared to:					Average %	Fold
C#	source	ppm	resonance	C7	C13	C20	C30	C37	resonance	C7	C13	C20	C30	C37	C7	C13	C20	C30	C37	enrichment	change
1	Pro	172.8	42.7	0.4	0.7	0.4	0.4	0.6	45.9	37.2	40.2	30.6	36.2	40.1	23.3	14.1	50.0	26.7	14.5	25.7	1.26
2	Pro	59.9	112.6	1.1	1.8	1.1	1.0	1.6	79.2	98.1	106.1	80.7	95.5	105.7	-19.3	-25.3	-1.9	-17.1	-25.1	-17.7	0.82
3	Pro	29.5	119.5	1.1	2.0	1.2	1.1	1.6	126.5	104.1	112.6	85.6	101.3	112.2	21.5	12.4	47.7	24.8	12.8	23.8	1.24
4	Pro	25.8	91.9	0.9	1.5	0.9	0.8	1.3	81.7	80.1	86.6	65.9	78.0	86.3	1.9	-5.7	24.0	4.8	-5.4	3.9	1.04
5	Pro	47.8	133.9	1.3	2.2	1.3	1.2	1.8	88.1	116.7	126.2	96.0	113.6	125.7	-24.5	-30.2	-8.2	-22.5	-30.0	-23.1	0.77
6	Ile	170.8	37.4	0.4	0.6	0.4	0.3	0.5	51.4	32.6	35.2	26.8	31.7	35.1	57.9	46.1	92.1	62.3	46.6	61.0	1.61
7	Ile	56.8	104.0	1.0	1.7	1.0	0.9	1.4	90.7	90.7	98.0	74.6	88.2	97.7	0.0	-7.5	21.6	2.8	-7.2	1.9	1.02
8	Ile	31.9	82.9	0.8	1.4	0.8	0.7	1.1	68.2	72.3	78.1	59.4	70.3	77.9	-5.7	-12.7	14.7	-3.1	-12.4	-3.8	0.96
9	Ile	24.8	93.7	0.9	1.5	0.9	0.8	1.3	66.0	81.7	88.3	67.2	79.5	88.0	-19.2	-25.3	-1.7	-17.0	-25.0	-17.6	0.82
10	Ile	9.2	124.2	1.2	2.0	1.2	1.1	1.7	83.4	108.3	117.0	89.0	105.4	116.6	-23.0	-28.8	-6.4	-20.9	-28.5	-21.5	0.79
11	Ile	14.2	145.1	1.4	2.4	1.4	1.3	2.0	86.1	126.5	136.7	104.0	123.1	136.3	-31.9	-37.0	-17.2	-30.1	-36.8	-30.6	0.69
12	SAM	30.7	89.0	0.9	1.5	0.9	0.8	1.2	175.7	77.6	83.8	63.8	75.5	83.6	126.5	109.6	175.5	132.8	110.3	***130.9***	***2.31***
13	Ala	170.2	61.1	0.6	1.0	0.6	0.5	0.8	57.6	53.3	57.6	43.8	51.8	57.4	8.1	0.0	31.5	11.1	0.3	10.2	1.10
14	Ala	60.9	76.9	0.7	1.3	0.8	0.7	1.1	90.6	67.0	72.5	55.1	65.2	72.2	35.1	25.0	64.3	38.9	25.5	37.8	1.38
15	Ala	14.1	62.4	0.6	1.0	0.6	0.6	0.9	70.6	54.4	58.8	44.7	52.9	58.6	29.9	20.1	57.9	33.5	20.6	32.4	1.32
16	SAM	36.9	38.8	0.4	0.6	0.4	0.3	0.5	109.7	33.8	36.6	27.8	32.9	36.4	224.3	200.0	294.4	233.2	201.0	***230.6***	***3.31***
17	Tyr	170.6	58.8	0.6	1.0	0.6	0.5	0.8	58.5	51.3	55.4	42.2	49.9	55.2	14.1	5.6	38.8	17.3	5.9	16.3	1.16
18	Tyr	50.6	97.8	0.9	1.6	1.0	0.9	1.3	13.8	85.2	92.1	70.1	82.9	91.8	-83.8	-85.0	-80.3	-83.4	-85.0	-83.5	0.17
19	Tyr	37.3	116.5	1.1	1.9	1.2	1.0	1.6	95.5	101.6	109.8	83.5	98.8	109.4	-6.0	-13.0	14.4	-3.4	-12.7	-4.1	0.96
20	Tyr	128.4	100.6	1.0	1.6	1.0	0.9	1.4	72.1	87.7	94.8	72.1	85.3	94.5	-17.8	-23.9	0.0	-15.5	-23.7	-16.2	0.84
21/25	Tyr	130.8	293.1	2.8	4.8	2.9	2.6	4.0	237.2	255.4	276.1	210.0	248.6	275.2	-7.2	-14.1	12.9	-4.6	-13.8	-5.4	0.95
22/24	Tyr	114.0	254.6	2.4	4.2	2.5	2.3	3.5	225.7	221.9	239.9	182.5	215.9	239.0	1.7	-5.9	23.7	4.5	-5.6	3.7	1.04
23	Tyr	158.8	43.0	0.4	0.7	0.4	0.4	0.6	46.0	37.5	40.5	30.8	36.5	40.4	22.6	13.5	49.1	26.0	13.8	25.0	1.25
26	SAM	55.4	124.2	1.2	2.0	1.2	1.1	1.7	270.1	108.2	117.0	89.0	105.3	116.6	149.6	130.9	203.5	156.5	131.7	***154.5***	***2.55***
27	Acetate	169.7	44.5	0.4	0.7	0.4	0.4	0.6	50.1	38.8	41.9	31.9	37.7	41.7	29.2	19.6	57.2	32.8	20.0	31.8	1.32
28	Acetate	130.6	42.1	0.4	0.7	0.4	0.4	0.6	56.6	36.7	39.6	30.1	35.7	39.5	54.4	42.8	87.7	58.7	43.3	57.4	1.57
29	Cys	136.4	90.2	0.9	1.5	0.9	0.8	1.2	98.7	78.6	85.0	64.6	76.5	84.7	25.6	16.2	52.7	29.1	16.6	28.0	1.28
30	Cys	72.6	111.7	1.1	1.8	1.1	1.0	1.5	94.8	97.4	105.3	80.1	94.8	104.9	-2.7	-10.0	18.3	0.0	-9.7	-0.8	0.99
31	Cys	37.7	71.5	0.7	1.2	0.7	0.6	1.0	37.6	62.3	67.4	51.3	60.7	67.2	-39.7	-44.2	-26.7	-38.0	-44.0	-38.5	0.62
32	SAM	13.5	81.2	0.8	1.3	0.8	0.7	1.1	167.6	70.8	76.5	58.2	68.9	76.2	136.8	119.1	188.0	143.4	119.8	***141.4***	***2.41***
33	Acetate	177.6	57.3	0.6	0.9	0.6	0.5	0.8	65.3	50.0	54.0	41.1	48.6	53.8	30.8	21.0	59.0	34.4	21.4	33.3	1.33
34	Acetate	49.3	76.3	0.7	1.2	0.8	0.7	1.1	84.9	66.5	71.9	54.7	64.7	71.7	27.6	18.0	55.2	31.1	18.5	30.1	1.30
35	Acetate	71.8	116.2	1.1	1.9	1.2	1.0	1.6	122.4	101.3	109.5	83.3	98.6	109.1	20.8	11.8	46.9	24.1	12.1	23.2	1.23
36	Acetate	38.3	94.7	0.9	1.6	0.9	0.8	1.3	104.0	82.6	89.3	67.9	80.3	89.0	25.9	16.5	53.1	29.4	16.9	28.4	1.28
37	Acetate	24.4	72.6	0.7	1.2	0.7	0.6	1.0	68.4	63.3	68.4	52.0	61.6	68.2	8.0	-0.1	31.4	11.0	0.3	10.1	1.10
38	Acetate	37.8	111.1	1.1	1.8	1.1	1.0	1.5	129.8	96.8	104.7	79.6	94.2	104.3	34.0	24.0	63.0	37.7	24.4	36.6	1.37
39	Acetate	77.4	N.P.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	Acetate	35.0	93.4	0.9	1.5	0.9	0.8	1.3	94.2	81.4	88.0	66.9	79.2	87.7	15.8	7.1	40.8	19.0	7.5	18.0	1.18
41-43	SAM	26.2	387.6	3.7	6.3	3.9	3.5	5.3	659.6	337.8	365.2	277.8	328.7	363.9	95.2	80.6	137.4	100.6	81.2	***99.0***	***1.99***
44	SAM	16.8	94.1	0.9	1.5	0.9	0.8	1.3	203.0	82.0	88.6	67.4	79.8	88.3	147.5	129.0	201.0	154.4	129.8	***152.3***	***2.52***
45	Acetate	20.0	109.2	1.0	1.8	1.1	1.0	1.5	70.7	95.2	102.9	78.3	92.6	102.5	-25.7	-31.2	-9.6	-23.6	-31.0	-24.2	0.76

Note: N.P. = No Peak - peak not observed due to $CDCl_3$ peak overlap

methyl groups were found to be as follows: C12 – 131%, 16 – 230%, 26 – 154%, 32 – 141%, 44 – 152%.

C41-43, the three peaks representing the *t*-butyl group, increased by 99%, or a nearly twofold increase

over the unlabeled sample of apratoxin A. The median absolute change for all remaining non-SAM derived peaks was 23.5%. No systematic bias was observed in enrichment based on biosynthetic origin aside from the SAM-derived peaks, indicating no major scrambling of labeled substrate.

4.4 Discussion

The results observed provided an *in vivo* validation of our expected findings. Hypothesizing only two reasonable routes to methylation of the malonate starter unit bound to AprA ACP, we sought to examine the possibilities for biosynthesis via stable isotope labeled enrichment. We expected that no incorporation of propionate would be observed, as cyanobacteria lack the enzymatic machinery to biosynthesize methylmalonyl CoA. We also hypothesized that substantial enrichment of the ^{13}C signal would take place on predicted SAM-derived methyl groups when cultures of *Moorea bouillonii* PNG 5-198 were fed [methyl-^{13}C]methionine. This was observed as a twofold (99%) increase in signal was observed for carbons 41-43. No significant enrichment of C39, the predicted point for incorporation of an enriched ^{13}C signal from [1-^{13}C]propionate was observed. These findings provide a backdrop and support to the significant biochemistry and structural biology investigations using purified recombinant AprA protein domains (Figure 4.10). With stability and solubility challenges reconciled experimentally, AprA

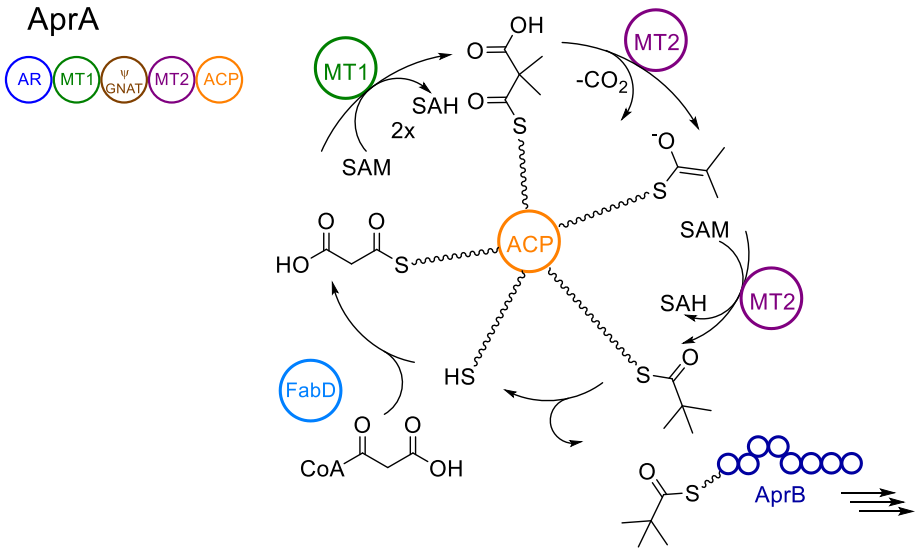


Figure 4.10. AprA from *Moorea bouillonii* PNG5-198 activity characterized by biochemistry experiments and structural biology, adapted from Skiba et al., 2018. SAM = S-adenosyl methionine. SAH = S-adenosyl homocysteine. FabD = fatty acid synthase acyltransferase from *Moorea boullonii* PNG.

domain constructs were produced and allowed exploration of the relationship of AprA MT1, the ψ-GNAT, and MT2 to the AprA ACP.[161] Ultimately, a suite of experiments showed that MT1 is a unique iron-dependent methyltransferase, which requires binding of Fe^{3+} to provide an electron acceptor for enolized ACP-bound malonate, which then reacts with SAM to methylate the C2 position of malonate, followed by a second α-deprotonation and attack of another SAM to form ACP-bound dimethylmalonyl-ACP.[276] The pseudo-GNAT, unlike active CurA GNAT, was not found to have any enzymatic role. The MT2 was found, remarkably, to first catalyze decarboxylation of the dimethylmalonyl-ACP, generating an enolic oxyanion, which then shunts electrons back to the α-carbon to attack a final SAM donor to generate ACP-bound pivalic acid, or the *t*-butyl group. This is then processed by AprB PKS and the remainder of the apratoxin A pathway. This system is an excellent example of the unique and highly evolved ways in which cyanobacteria have diversified their secondary metabolism, particularly with initiation modules in PKS/NRPS pathways.

Based on these findings, computational algorithms may be utilized to identify *t*-butyl containing natural products based on gene sequence alone, which aids in structure prediction and target identification. It is highly likely that the remainder of the type I PKS/NRPS derived cyanobacterial natural products utilize an AprA homolog in their biosynthetic pathways. One question that remains is how widespread the module may be; while it has been found in cyanobacteria, it was also found through sequencing the metagenome of *Candidatus Endobugula sertula* that is responsible for biosynthesizing the protein kinase C inhibitor bryostatin. Indeed, there may be many more biologically active natural products bearing a *t*-butyl group yet to discover.

4.5 Experimental methods

4.5.1 Culturing and feeding experiments of *Moorea bouillonii* PNG5-198

Similar culture conditions were utilized for both cultures of *Moorea bouillonii* PNG5-198 fed with [methyl-^{13}C]methionine or labeled [1-^{13}C]propionate. The apratoxin A producer was cultured in SW-BG11 media between 27-28°C, under light of 5.4-10.8 μmol photons/m^2S. Two boluses of 0.5 g of wet

filaments approximately 1 cm in length were inoculated into 250 mL Erlenmeyer flasks containing 75 mL of media and incubated for 14 days, followed by sub-culturing of each biomass flask in two 2.8 L Fernbach flasks containing 1 L of media. Growth was continued for ten days, and then the biomass from each flask was combined into one 2.8 L Fernbach flask per feeding condition containing 750 mL of media, and the culture equilibrated for an additional 7 days to ensure growth stability. [Methyl-^{13}C]methionine (Cambridge Isotopes) was added to the culture in three batches of 20 mg each over 13 days, to a final concentration of 0.53 mM, and harvested after an additional 6 days of growth. [1-^{13}C]Propionate (Cambridge Isotopes) was added to a separate culture flask with a similar quantity of biomass in three batches of 28.8 mg over period of 15 days, to a final concentration of 1.2 mM, and harvested after an additional 15 days of growth. Cultures were harvested and washed with DI water over a Büchner funnel with filter paper and stored at -20°C until extraction.

4.5.2 Extraction and purification of apratoxin A

Extraction proceeded by addition of 100 mL 2:1 CH_2Cl_2:MeOH in four sequential batches, and harvesting the extract over a Büchner funnel from each batch. Crude extract was dried via rotary evaporation. Extracts were fractionated by C18 Hypersep SPE (Agilent) and eluted with 40%, 80%, and 100% MeCN, then 100% CH_2Cl_2. The 80% and 100% MeCN fractions were found to contain apratoxin A by LC/MS. Fractions were combined and subsequently purified via semi-preparative HPLC using a gradient of 65% to 99% MeCN in H_2O + 0.1% formic acid at 3 mL/min with a 150 x 10 mm, 5 μM Kinetex C18 semi-preparative column. Subsequent semi-pure apratoxin A plus demethylated analogs were re-submitted for HPLC using a gradient of 70% to 81% MeCN in H_2O + 0.1% formic acid at 4 mL/min, rendering 1.2 mg of apratoxin A from the [1-^{13}C]propionate experiment and 1.8 mg of apratoxin A from the [methyl-^{13}C]methionine experiment. Approximately 2 mg of apratoxin A harvested from previous non-labeled batches was purified identically and used for NMR enrichment comparisons.

4.5.3 ^{13}C-NMR, ^{1}H-NMR and incorporation calculations

All ^{13}C NMR data was collected on a Varian VX500 spectrometer at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR. A JEOL ECZ500 spectrometer was used to obtain HSQC and HMBC spectra in

order to confirm ^{13}C and ^{1}H NMR shifts of apratoxin A in methanol-d_4 solvent which was used in the analysis of the [1-^{13}C]propionate feeding experiment (Table A4.1, Figures A4.9-A4.10). Peaks were referenced to residual solvent signals in each sample: $CDCl_3$ δ_C=77.16, δ_H=7.24; methanol-d_4 δ_C=49.00, δ_H=3.31. Sample processing was performed using MestReNova for Varian VX500 data and Delta for JEOL data. Table 4.1 indicates ^{13}C-enrichment by carbon number in the [methyl-^{13}C]methionine feeding experiment, using a normalization method adapted from previous studies.[89] The calculation method for deriving percentage enrichment is as follows for table 4.1 and 4.2: integrated values of natural abundance apratoxin A ^{13}C NMR signals are indicated in column 4. First, five carbon atoms with disparate biosynthetic origins and a range of chemical shifts were used for normalization comparison of unenriched and ^{13}C-labeled spectra: C7, C13, C20, C30, and C37. Columns 5-9 indicate normalization factors, which were obtained by dividing the integrals of the ^{13}C NMR signals in column 4 by those of C7, C13, C20, C30, and C37 (underlined in column 4). Following this, the integral for the individual enriched signals of the five carbon atoms used in the normalization procedure (underlined in column 10) were multiplied by their normalization factors (columns 5-9) to obtain an expected signal integral for all carbon signals if no enrichment had occurred (columns 11-15). Finally, percent enrichment was calculated by subtracting columns 11-15 from column 10, then dividing the result by the column 11-15 values *((enriched – expected unenriched)/expected unenriched)*, then multiplying by 100 to yield the values in columns 16-20. The results in columns 16-20 were averaged to yield the values in column 21. Column 22 is another expression of enrichment, namely the fold change of the enriched vs. unenriched sample. Median and standard deviation of change in enrichment was calculated by taking the median of the absolute value of all enrichment values, while standard deviation was calculated by standard means.

4.6 Appendix.

Table A4.1. ^{13}C-NMR shifts of native abundance apratoxin A in $CDCl_3$ and methanol-d_4, [methyl-^{13}C]methionine labeled apratoxin A in $CDCl_3$, and [1-^{13}C]propionate-labeled apratoxin A in methanol-d_4.

	[methyl-^{13}C]methionine labeled $CDCl_3$	Native abundance $CDCl_3$	[1-^{13}C]propionate labeled methanol-d_4	Native abundance methanol-d_4
C#	ppm	ppm	ppm	ppm
1	172.8	172.8	174.1	174.1
2	59.9	59.9	61.2	61.2
3	29.5	29.4	29.8	29.8
4	25.8	25.8	26.5	26.5
5	47.8	47.8	47.5	47.5
6	170.8	170.8	172.7	172.7
7	56.8	56.7	57.9	57.9
8	31.9	31.9	31.0	31.0
9	24.8	24.8	25.3	25.3
10	9.2	9.2	9.1	9.1
11	14.2	14.2	14.7	14.7
12	30.7	30.7	30.4	30.4
13	170.2	170.2	171.6	171.6
14	60.9	60.8	61.6	61.6
15	14.1	14.1	14.4	14.4
16	36.9	36.9	35.8	35.8
17	170.6	170.5	172.4	172.4
18	50.6	50.6	52.3	52.3
19	37.3	37.4	37.1	37.1
20	128.4	128.4	129.7	129.7
21/25	130.8	130.8	131.7	131.7
22/24	114.0	114.0	115.0	115.0
23	158.8	158.8	160.3	160.3
26	55.4	55.4	55.7	55.7
27	169.7	169.7	171.3	171.3
28	130.6	130.6	N.P.	130.8
29	136.4	136.4	137.0	137.0
30	72.6	72.6	73.1	73.1
31	37.7	37.7	38.4	38.4
32	13.5	13.5	13.0	13.0
33	177.6	177.6	179.8	179.8
34	49.3	49.3	50.2	50.2
35	71.8	71.8	73.1	73.1
36	38.3	38.3	39.1	39.1
37	24.4	24.4	25.3	25.3
38	37.8	37.8	38.7	38.7
39	N.P.	N.P.	78.7	78.7
40	35.0	35.0	32.7	32.7
41/42/43	26.2	26.2	26.5	26.5
44	16.8	16.8	17.1	17.1
45	20.0	20.0	20.0	20.0

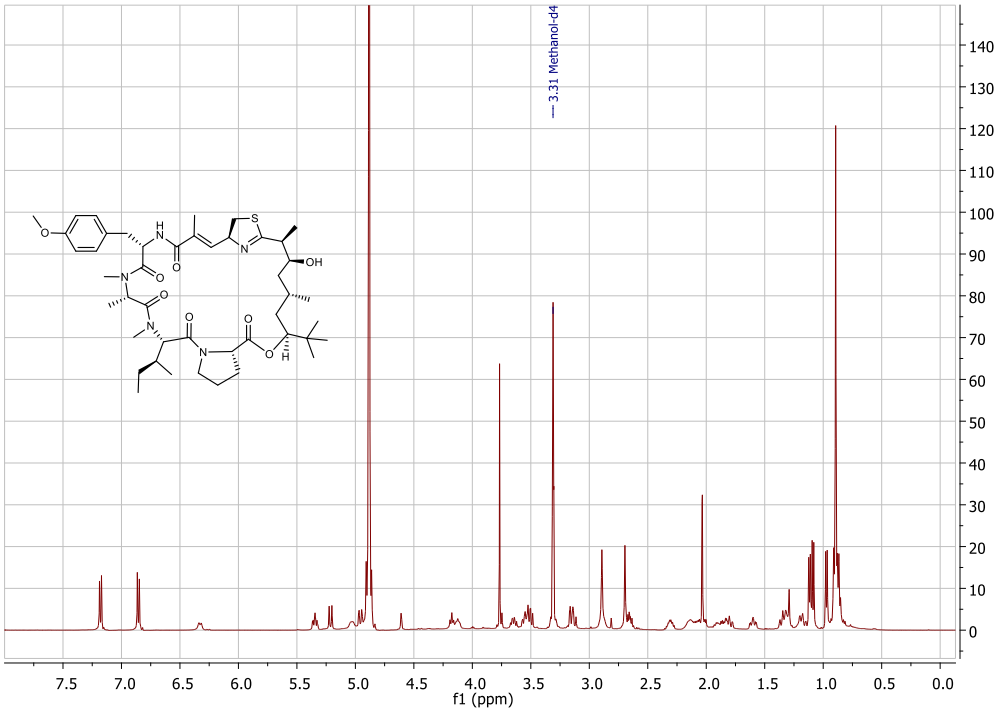


Figure A4.1. ^{1}H-NMR of apratoxin A, unlabeled (500 MHz, methanol-d_4)

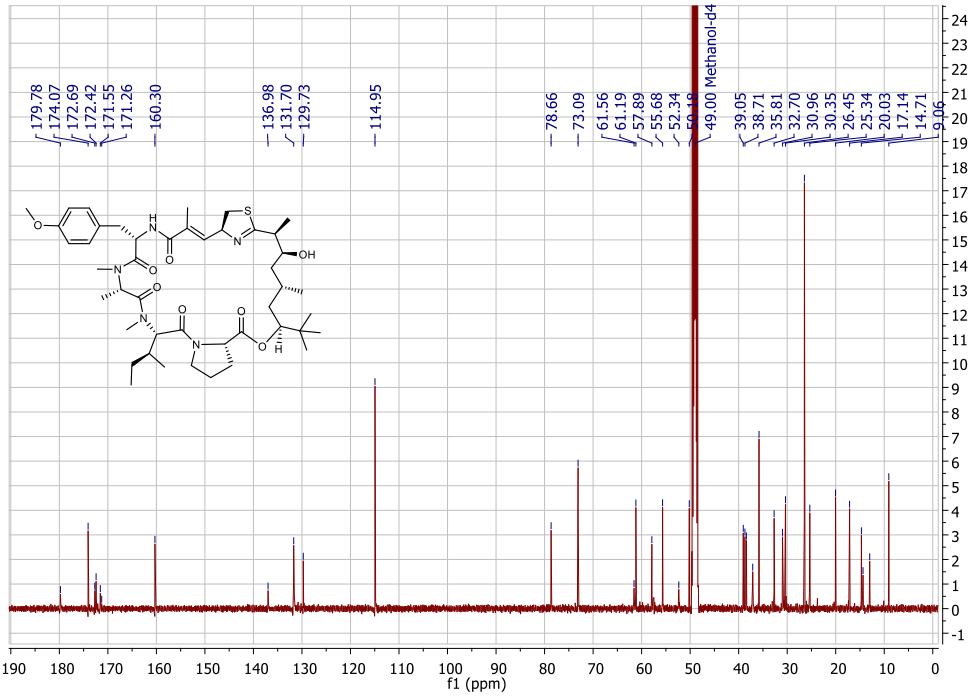


Figure A4.2. ^{13}C-NMR of apratoxin A, unlabeled (125 MHz, methanol-d_4)

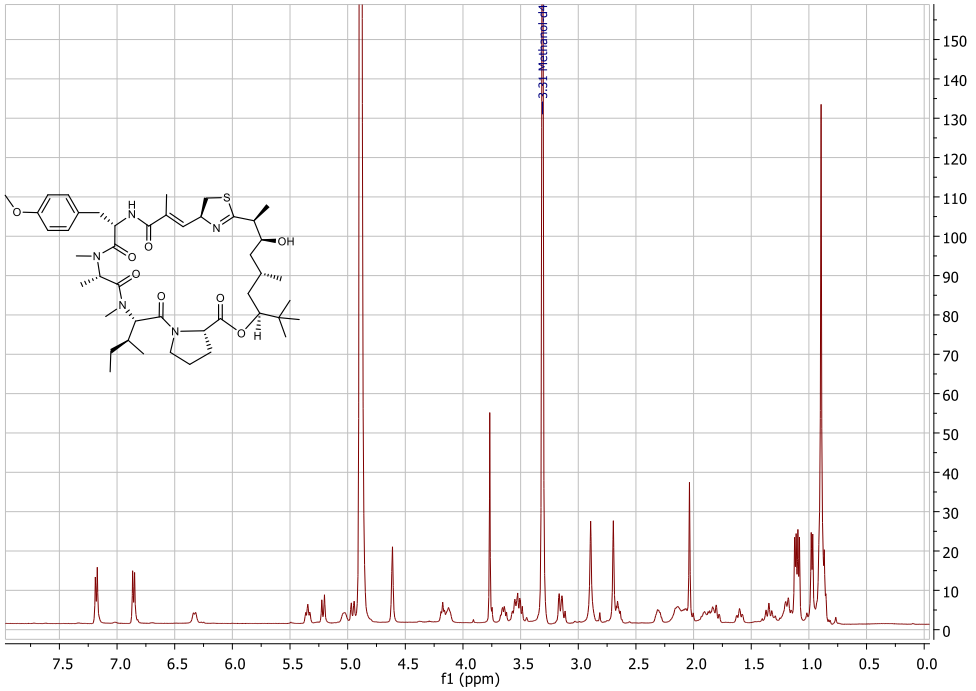


Figure A4.3. ^{1}H-NMR of apratoxin A, fed [1-^{13}C]propionate (500 MHz, methanol-d_4)

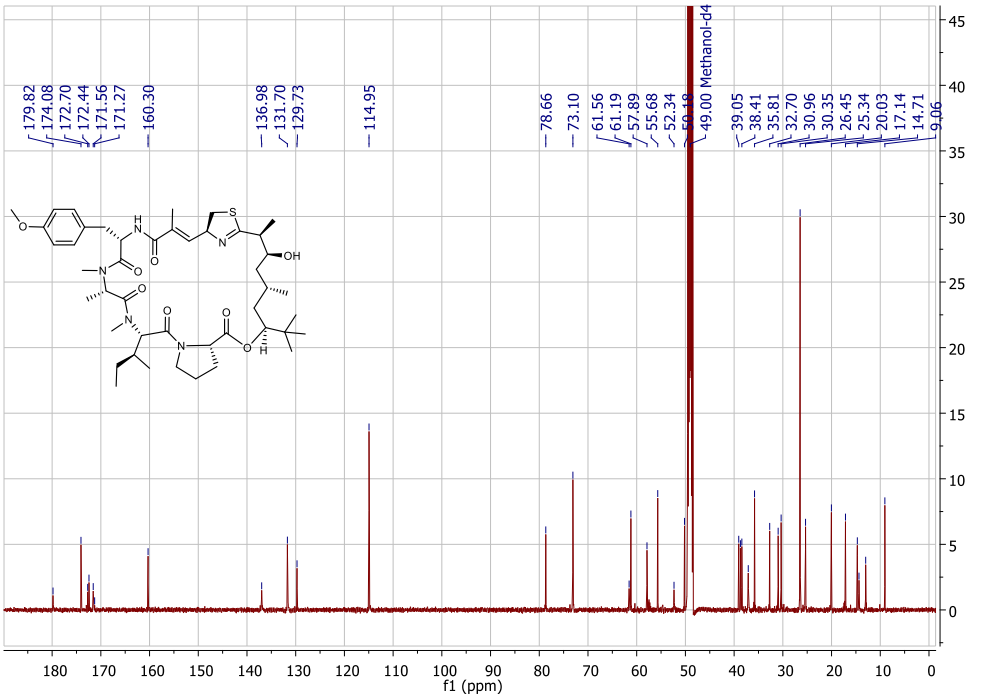


Figure A4.4. ^{13}C-NMR of apratoxin A, fed [1-^{13}C]propionate (125 MHz, methanol-d_4)

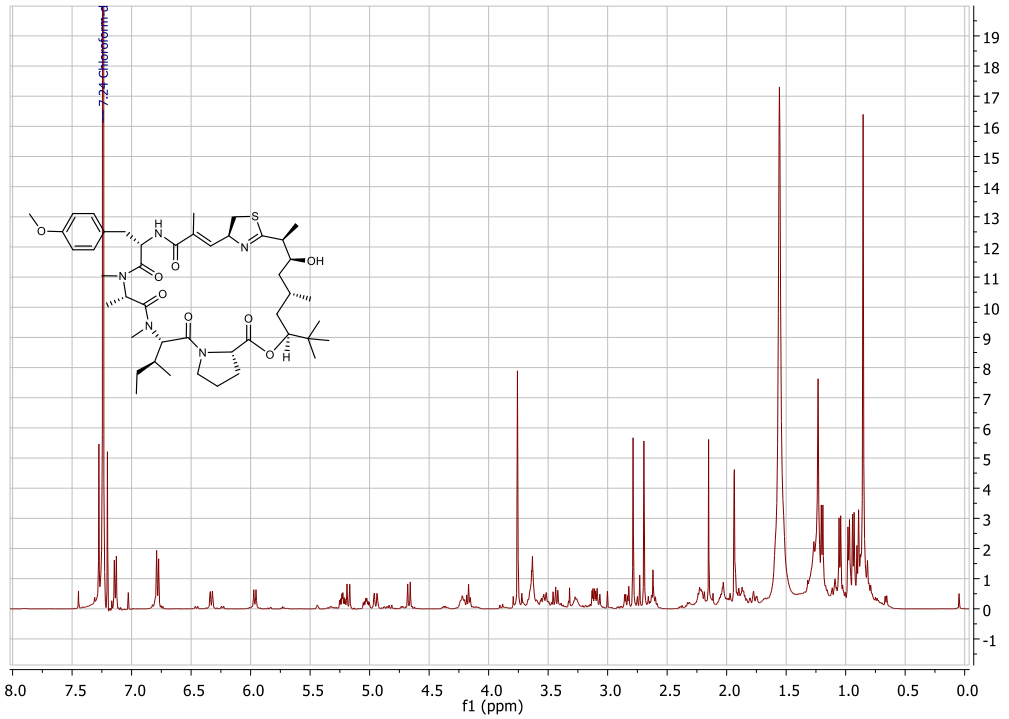


Figure A4.5. ^{1}H-NMR of apratoxin A, unlabeled (500 MHz, $CDCl_3$).

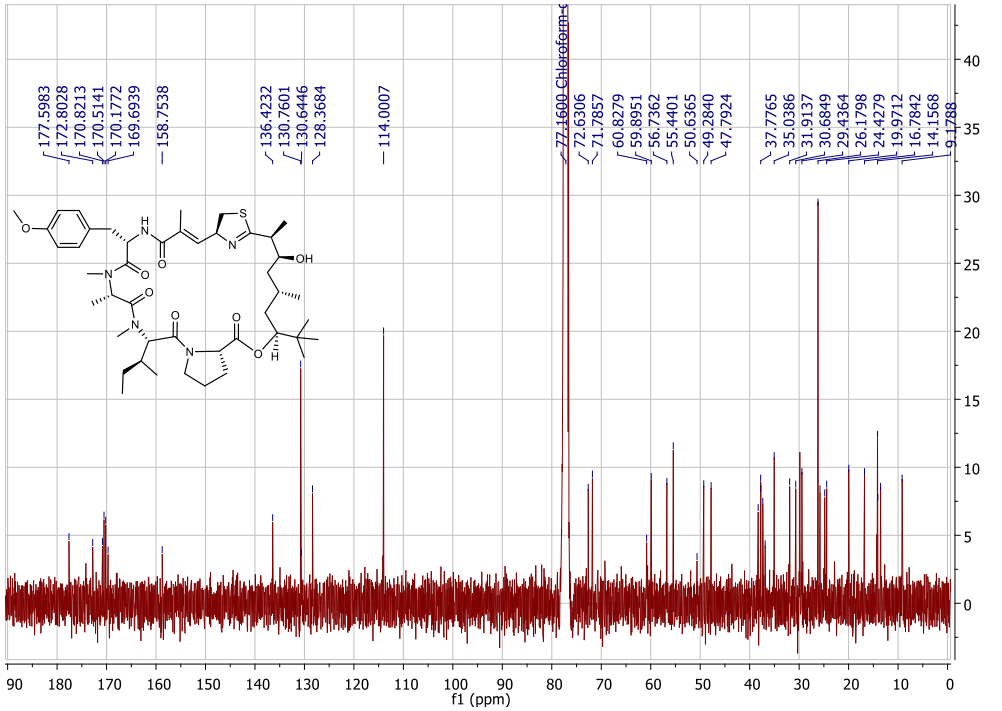


Figure A4.6. ^{13}C-NMR of apratoxin A, unlabeled (125 MHz, $CDCl_3$).

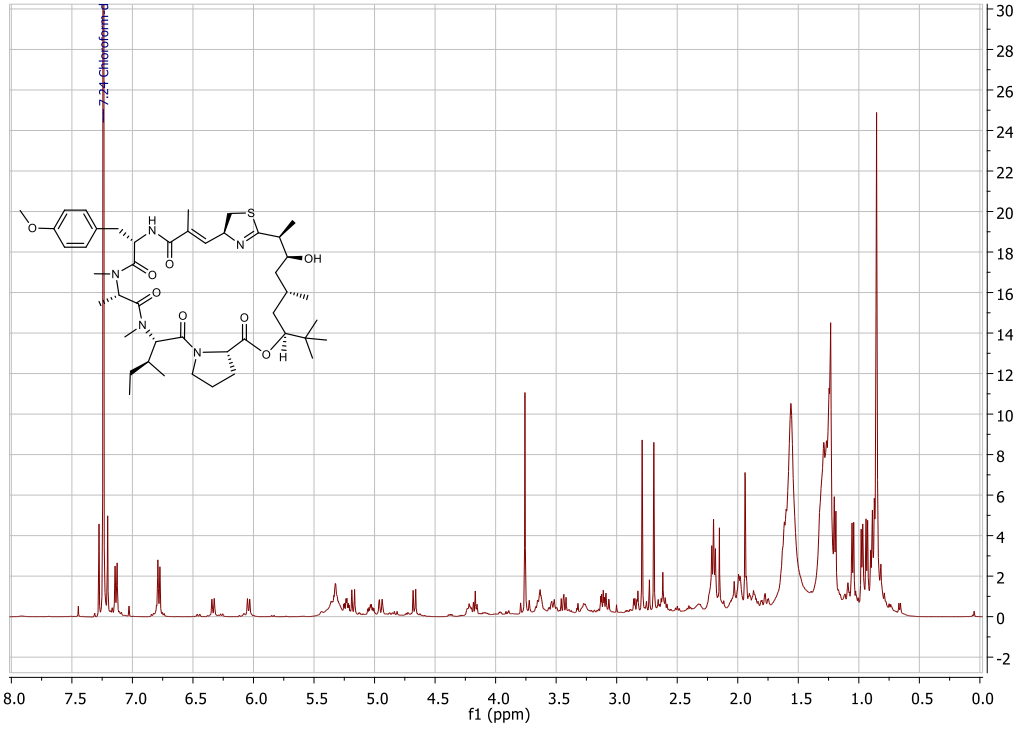


Figure A4.7. ^{1}H-NMR of apratoxin A, fed [methyl-^{13}C]methionine (500 MHz, $CDCl_3$).

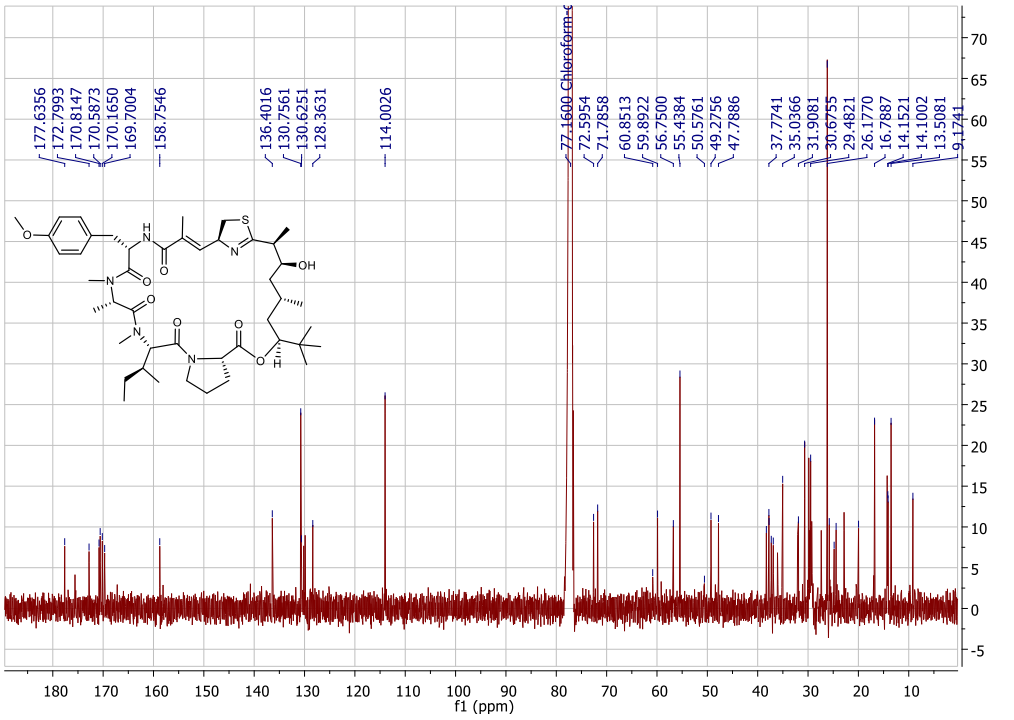


Figure A4.8. ^{13}C-NMR of apratoxin A, fed [methyl-^{13}C]methionine (125 MHz, $CDCl_3$).

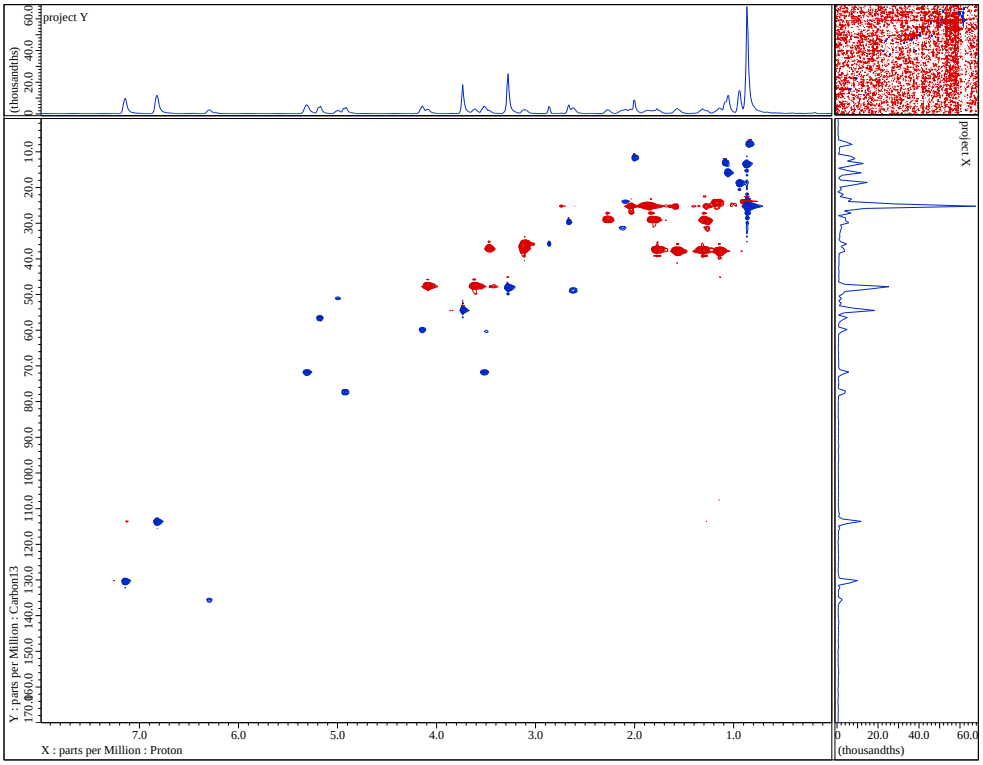


Figure A4.9. HSQC of unlabeled apratoxin A, (500 MHz, methanol-d_4)

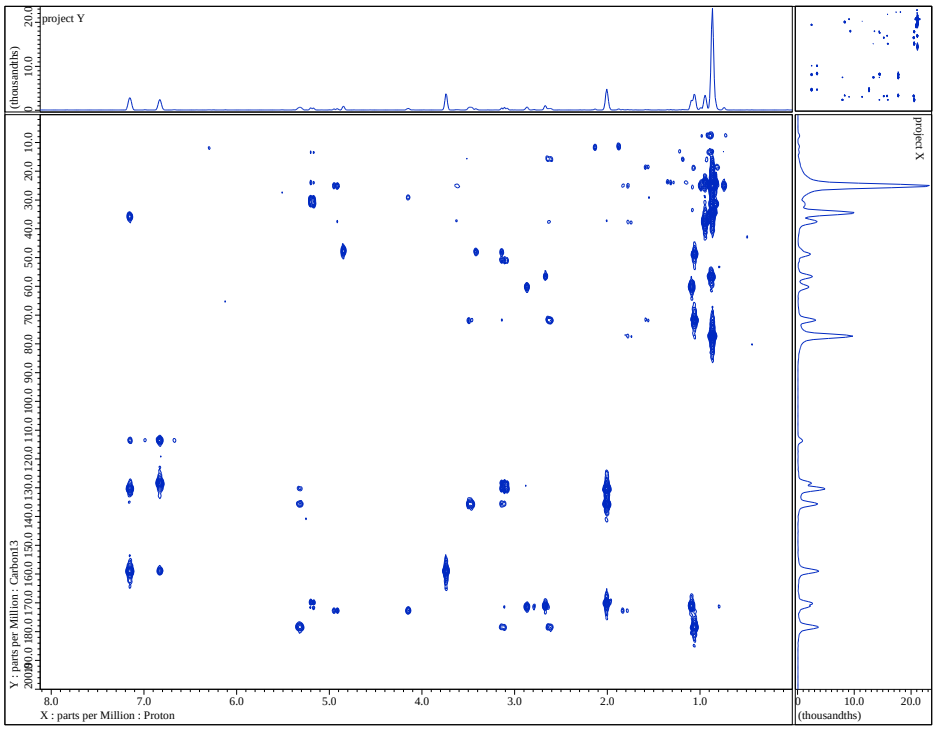


Figure A4.10. HMBC of unlabeled apratoxin A, (500 MHz, methanol-d_4)

4.7 Acknowledgments.

I would like to acknowledge E. Mevers, K. Kleigrewe, M. Bertin, E. Glukhov and C. Coates for their contribution to this project with regard to cyanobacterial culturing, extraction and NMR instruction. I acknowledge M. Skiba, A. Sikkema, and J. Smith, for their tireless work and dedication to high standards in determining the biochemistry of AprA. I acknowledge W. H. Gerwick for his input to the ideation of this project. We thank the following funding sources for their support of this project: NIH; Training Grant in Marine Biotechnology, UCSD/SIO; Regents Fellowship, Nora and Alan Jaffe Fellowship, Robert L. Cody Memorial Fellowship, and the Henry L. and Grace Doherty Fellowship.

This chapter, in part, is reproduced, with permission, from ACS Chemical Biology; Biosynthesis of t-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module. Skiba, Meredith A.; Sikkema, Andrew P.; Moss, Nathan A.; Lowell, Andrew N.; Su, Min; Sturgis, Rebecca M.; Gerwick, Lena; Gerwick, William H.; Sherman, David H.; Smith, Janet L. 2018, *13*, 1640-1650.

Chapter 5: Continuing research in natural products biosynthesis

5.1 Research synopsis

This dissertation encapsulated several lines of research into understanding how tropical marine filamentous cyanobacteria generate chemical diversity through novel biosynthetic processes. Several chemical biology techniques were employed to study the production of bioactive molecules from PKS/NRPS pathways discovered within filamentous cyanobacteria, primarily centered upon mass spectrometry and NMR, and were complemented by genomics. Owing to the currently intractable genetic nature of marine filamentous cyanobacteria, manipulation of the producing organisms as well as transformation of intact pathways into heterologous hosts was not accomplished. However, tremendous advances in genomic sequencing and its analysis tools, as well as computational modeling of enzyme domains, enabled a discovery-based characterization of two large BGCs, the malyngamides from *Okeania hirsuta* PAB10Feb10-1 and PAP21Jun06-1, and the vatiamides *Moorea producens* ASI16Jul14-2. Additionally, NMR techniques provided the groundwork for a comprehensive biochemical analysis of the loading module of the apratoxin A domain. In the analysis of the vatiamides and malyngamides in particular, molecular characterization by NMR was accompanied by genome sequencing to answer questions about unknown characteristics in the biosynthesis of these two classes of molecules. In each project, findings illuminated unique enzymes and protein interaction mechanisms by which cyanobacteria impart structural diversity to molecular scaffolds. These mechanisms, specifically neofunctionalization of lipoic acid metabolic enzymes and conserved mutation of a KR domain in the malyngamides, and the interaction of nonselective docking domains in vatiamides, present unorthodox alternative strategies to previously held paradigms for PKS/NRPS biosynthesis.

5.2 Project summaries and future directions

5.2.1 Malyngamide biosynthesis

A longstanding mystery in cyanobacterial natural products concerned the source of the type A malyngamides. Isolated for decades from filamentous marine cyanobacteria as well as from their

predators and their environmentally close consortia, the true source and the biosynthetic logic behind their occurrence remained a mystery. Chapter two described genome sequencing of two separate cultures of *Okeania hirusta*, "PAB" and "PAP" followed by domain and module predictive analysis to discover an approximately 60 kb biosynthetic gene cluster in each organism. This gene cluster was analyzed and found to contain the modular machinery necessary to produce the backbone core of malyngamide C acetate in PAB and malyngamide I in PAP. However, key intermediate steps, specifically loading and cyclization, were not apparent by genomic sequence alone and required further tools to deduce. Specifically, a phylogenetic analysis of the genes upstream of the first ketosynthase module in each pathway identified a gene predicted to encode for a LipM-like enzyme, MgcA/MgiA. This enzyme is predicted to transfer a FAS ACP-bound octanoate moiety to the first KS of malyngamides, though its specific activity is not known. The octanoyltransfer activity of MgcA was suggested by observance of an increase in the specific NMR signal at C-7' after extraction of *Okeania* biomass fed with [1-^{13}C]octanoate. Similarly, analysis of malyngamide C acetate isolated from batches of *Okeania* PAB fed with [1,2-$^{13}C_2$]acetate offered structural information as well. J_{CC} coupling constants between C-4 – C-9 of the six-membered cyclohexanone head group indicated that the characteristic C-5 ketone came from MgcK module. Persistence of the ketone was theorized to occur due to a mutated or otherwise inactivated KR^0 domain in MgcQ. This was supported by fluorescence polarization analysis *in vitro* of recombinant MgcQ MT-KR^0 didomain, as well as homology modeling. This ketone, combined with a terminal aldehyde generated by a thioester reductase, provides a logical path to cyclohexanone formation by a Knoevenagel condensation. The understanding of these two key transformations, loading and cyclization, in the production of malyngamide C acetate, inferred a biosynthetic model for all 28 type A malyngamides predicated on the inactivity of the MgcQ KR^0. Thus, genomics and NMR and MS-based analysis of this pathway illuminated that cyanobacteria generate a radiation of molecular diversity following evolutionary selection for this inactive domain.[102]

One invitation for subsequent research into this pathway is the activity of the LipM enzyme. Though it is suggested to initiate octanoyltransfer to the first pathway KS by analysis via NMR, a

thorough *in vitro* study of its biochemical activity is warranted, especially given that MgcA/MgiA uniquely appear to activate a PKS module, which is not similar to the native substrate of LipM.[162,167] Lipoic acid biosynthesis and scavenging has been well-characterized in bacteria, and a number of the pathway members, including LipB, are found in *Okeania* as well as other cyanobacteria. Future work may investigate the substrate specificity of this enzyme, as well as its transfer to the first KS module, and whether any cofactors or additional protein activity is required. This may be accomplished by an experiment involving chemical crosslinking probes, or radiolabeling of expressed proteins, followed by analysis via electrophoresis.

An additional point of study could involve the small lipocalin-like protein that follows MgcR/MgiR. This protein is generally reported in higher organisms to be responsible for binding and transport of fatty acids. After MgcR/MgiR generate a terminal aldehyde, cyclization is predicted to occur. It is unclear which domain, the reductase or lipocalin, is immediately responsible. As homologs to this domain also appear in a number of cyclization mechanisms throughout the microbial kingdom, including in the potently neurotoxic anatoxin, further characterization of this domain is warranted as it may illuminate a widespread biosynthetic strategy. Lastly, a phylogenomic analysis of additional malyngamide producers is warranted. Study of the tailoring and loading enzymes of various as-yet undiscovered malyngamide pathways would validate the model presented here, and possibly reveal additional new methods of biosynthetic diversification.

5.2.2 Vatiamide biosynthesis

In chapter 3, bioassay guided fractionation led to the isolation of the vatiamides A – D, bearing an alkynyl or bromoalkynyl jamaicamide-like fatty acid tail linked via amide to a trimethylated tyrosine head group (A/B) or a glycine-bicyclic ketide head group (C/D). Hybrid sequencing and genome analysis of producing organism *Moorea producens* ASI showed a 90 kb PKS/NRPS pathway containing the core molecular structure of vatiamides A – D. This pathway was shown to contain separate cassettes encoding for the jamaicamide-like tail, and separately, the A/B and C/D head groups, which each flanked a central cassette that did not appear to produce a product. A click-chemistry based coumarin-like probe was used

to identify and pull from crude extract an additional pair of vatiamide analogs bearing a modified alanyl-valinamide dipeptide head group not immediately evident in LC/MS: vatiamides E/F. Analysis of the entire pathway indicated an unprecedented promiscuous utilization of docking domains, short structural proteins whose amino acid sequence fosters selective interaction between PKS and/or NRPS modules. This promiscuity results in *in vivo* combinatorial biosynthesis, wherein the fatty acid tail is abstracted from its producing VatA – VatM cassette onto three separate downstream NRPS modules to produce the suite of six compounds. To date this is the only example of combinatorial biosynthesis enabled by docking domain promiscuity in the microbial kingdom. Additionally, a cryptic domain at the terminus of module VatR appears to endow VatR with a unique aminotransferase activity previously observed only in the biosynthesis of carmabin, and is not yet biochemically characterized.[218]

Significant future scholarship should be undertaken on this project owing to the uniqueness of the pathway and its potential applications to biotechnology. First, the basis for lack of selectivity between VatM/P and VatN/Q/S should be probed and capitalized upon with regard to NRPS module diversity. The ability to connect an upstream series of KS modules to several different downstream NRPS modules could be highly valuable for the ability to test the pharmacology of different molecular head groups. If this concept could be harnessed, it could be employed in a high-throughput synthetic biology setting, for example, by attaching this cognate pair of docking domains to many different upstream PKS modules and many different downstream NRPS modules.

From an evolutionary standpoint, the unusual mechanism by which three distinct NRPS modules acquired the exact same docking domain is also worth study. Which evolved first, three separate NRPS modules which then acquired three identical docking domains by recombination? Or was a single NRPS domain duplicated, and then duplicated again, followed by slow mutation of the adenylation domain Stachelhaus residues which confer amino acid activation? Or a third, less likely but not impossible concept: that the docking domains of the three NRPS modules started out as heterogeneous sequences, but were slowly selected to be identical so that the organism could produce three different molecular classes? How did this duplication event occur, and how did all three NRPS modules then acquire distinct NRPS or

PKS modules to follow, creating distinct “cassettes”? What native transformation and recombination enzymes, such as transposases, mediated these duplications or exchanges?

Additionally, questions remain regarding the evolutionarily selective pressures which resulted in the production of six different peptidic head groups with a common fatty acid tail. The energy demand associated with activation of amino acids, loading and desaturation of a hexenoyl-CoA starter unit, and reduction of acetate units is significant. Therefore, the benefit to the organism from the production of these molecules is likely considerable in its native environs. But further, how and why were the separate NRPS cassettes selected for? Is the target of the six molecules the same, and each molecule binds or interacts with a separate chemical entity on the target? Or are there different targets within the same organism or class of organisms (for example, fish), such that each head group class binds to a different target, thus increasing the toxicity of the system? Or, are the targets of each molecule more disparate in the phylogenetic kingdom; for example, vatiamides A/B may target fish, while vatiamides C/D are anti-fungal, while vatiamides E/F are molluscicidal to prevent three avenues of predation? Moreover, why was this particular PKS cassette selected for in (including the jamaicamides) eight different natural products? Surely the modified aliphatic chain plays a crucial role in the activity of these molecules, be it by intercalating membranes or positioning the head group into an appropriate enzyme pocket. Considering that other bioactive cyanobacterial molecules, including the type A and type B malyngamides, columbamides, and antillatoxin, among many others, feature modified and substituted fatty acid tail sections, it is likely that this structural feature plays an important role in the pharmacology of the natural product. Considerable bioassay testing on a range of targets would be required to assess all of these options and understand the production of these molecules in an ecological context. The results may indicate routes to clinical relevance if a particular target is found to be relevant to human health.

Lastly, from a phylogenomic standpoint, it seems unlikely that this type of pathway is truly unique in the kingdom of life, as the evolution of bacteria and cyanobacteria has many vertical and horizontal branch points. However, it is quite possible that poor sequencing coverage and quality has prevented the accurate assembly of these types of pathways, owing to the high levels of conservation

between modules and docking domains. Therefore, a broad long-read sequencing project of filamentous cyanobacteria or other species suspected of producing such combinatorial pathways should be undertaken to understand the phylogenetic diversity of the pathway. Until then, this pathway remains the sole example of native combinatorial biosynthesis in a PKS/NRPS context.

5.2.3 *t*-Butyl biosynthesis in apratoxin A

In chapter four, stable isotope labeled feeding studies followed by analysis by NMR lent clarity to the understanding of apratoxin A biosynthesis. [^{13}C-Methyl]methionine fed in batches to *Moorea bouillonii* PNG5-198 indicated specific uptake of the labeled methyl group into specific positions of apratoxin A as predicted by biosynthetic tools. The increase in signal of 99% in C-41 – C-43 showed that SAM resulted in the methylation of the *t*-butyl group in cyanobacteria. Complementarily, the feeding of [1-^{13}C]propionate, theorized as a potential alternative supplier of one of the methyl groups in the *t*-butyl group of apratoxin A, did not lead to a significant enrichment of C-39, the predicted incorporation point for [1-^{13}C]propionate. Biochemical experiments performed by collaborators on recombinant domains from AprA fully explored the mechanism of AprA. These experiments clarified the participation of an iron-dependent methyltransferase acting twice on the C-2 of ACP-bound malonate, followed by decarboxylation and finally methyl transfer by a type I in-module C-MT. This resulted in the pivaloyl-ACP that is passed on to AprB and the remainder of apratoxin biosynthesis.[161,276]

While we believe that there is not a great deal of future work to be done on *t*-butyl biosynthesis leading to the apratoxins, there are further steps that may be undertaken to understand the system in its entirety. Compared to other similar modules, such as in CurA, the GNAT in AprA is found to be an inactive "pseudo-GNAT" that cannot catalyze acyltransferase activity, instead relying on FabD acyltransferase to load the AprA ACP. The evolutionary selection towards the inactive pseudo-GNAT is highly interesting, as it appears to foster a structure/activity tradeoff between the ability of the GNAT to load the modular ACP versus the structural inclusion of the second C-MT and therefore creation of the pivaloyl group. Additionally, the distribution of the *t*-butyl group in PKS/NRPS derived molecules from

both *Moorea* and species such as *Okeania* and *Bryozoa* symbionts invites a biogeographical analysis of its distribution and phylogenetic occurrence.

5.3 Future perspectives

The study of natural products has evolved considerably over several decades, adopting and absorbing new technologies into workflows that enable discovery of new molecules, protein function, and characterization of targets.[1,135,277] Philosophically, in terms of academic research, there has never been a better time to be a natural products researcher. The confluence of funding for automation, genetic manipulation, and genomics, offer advances in the short to medium term for the discovery of new screening possibilities, gene and protein function, and novel BGCs, respectively. In addition, the cost of DNA synthesis continues to fall, and technologies that enable assembly of nucleic acid molecules for introduction into a model organism are flourishing.[278,279] Genome assembly is likely just a few years away from cloud computing interfaces that enable single-button *de novo* assembly of new microbial organisms, followed by analysis for biosynthetic content. In terms of molecule characterization, computational analysis has made MS and NMR-based structural characterization available to non-specialists, making structure solutions easier than ever to discern. If its promise is realized, cryo-EM offers revolutionary capabilities in solving molecular structures.[280] Combining these new techniques with robust legacy separation methods allows molecules to be found and solved with an ease like never before. Put simply, the technological black box has largely been removed from three major aspects of natural products research: assembly of genomes, creation of and assembly of DNA parts, and the solution of molecular structures. With enough effort, it is likely that any organism may reveal its genetic treasures, any gene can be inserted into a subset of model organisms, and any molecule can be solved in a reasonable research timeline.

However, there remain significant uncertainties that must also be addressed. Namely, we are still woefully deficient in understanding most protein function, structural biology, and the regulatory elements that govern transcription of RNA and expression and folding of proteins, particularly concerning the interaction of PKS/NRPS megasynthase complexes with tailoring enzymes. One can spend an enormous

effort attempting to purify a protein to study its function, but if it is too large, insoluble, and cannot be expressed properly, or requires expression with additional complex enzymes in order to be active in a stable form, acquiring a deep understanding of its role becomes improbable. Similarly, massive effort may be undertaken to introduce a piece of DNA carrying a BGC into a host organism, but if the host organism lacks the promoters, transcription factors, precursors and transporters to produce the molecule in reasonable titer, production may not occur for entirely unknown reasons. While entire research facilities and grants are funded with the aim of utilizing advanced predictive algorithms and AI/machine learning to predict molecular structure from genes, there still remains tremendous uncertainty in the functions of the genes themselves. Thus, molecular structure prediction from genes is still unreliable at best, unless the BGC being analyzed is comprised of a small subset of very well-understood domains. Therefore, I believe that understanding protein function and regulation in microbial secondary metabolism must be prioritized; without accurate, "clean," and well-understood inputs, opening the genomic treasure chest promised by machine learning and AI will only reveal ore.

The second, longstanding barrier to a fruitful natural products research ecosystem is translation from benchtop to bedside. Conferences in this field frequently feature a global array of dozens of posters advertising an interesting (or not so interesting) molecule isolated from an organism. Many of these molecules were tested in one or two broad bioassays such as plate-based antibiotic assays, cancer cells or brine shrimp; they may be tested in a protein-based assay or a more involved mammalian cell-based assay if the lab has advanced culturing capabilities or a relationship with a well-funded medical research institution. But frequently this marks the end of the development of a natural product: a published scientific paper. In the absence of a steady source of compound, there may not be enough for a broad panel of bioassay testing. Therefore, synthesis should be emphasized, if not explicitly incentivized via commercial sources, as any compound is with certainty impossible to develop as a drug without a reliable source. To this end, the field should focus its attention on reducing "paper molecules" and establishing a systematic methodology of channeling newly discovered natural products towards appropriate and applicable assays, in order to fully assess the potential of molecules. Although isolated molecules may

display varying levels of toxicity against targets, the gap between the discovery stage and the clinic, or even the discovery stage and more advanced screening of the molecule and semisynthetic analogs, is huge. Investing the time and energy to complete the life cycle of a natural product, from discovery to clinic, will ultimately pay dividends via funding and commercial interest in the field.

This is not just a problem limited to traditional isolation and structural characterization schemes: the products of the most complex and intriguing cloned pathways must also be extensively assayed and screened if they are to be effectively channeled through the drug development pipeline. Ultimately, the entire natural products community must reconcile the tremendous gap between academic ideation and actual application of clinical treatments derived therefrom. The former Chief Medical Information Officer of Merck, Dr. Sachin Jain, put it succinctly:

> There is a disconnect between the conferences I attend, the journals and blogs I read, and the reality of medical practice on the frontlines of healthcare delivery. There is a 'change layer' – the cloud in which visionary ideas about transforming health care resides. But there is also a 'reality layer' – the place where most care is delivered. Both are necessary, but there is little mixing between them. So while there is a booming innovation industry – a new startup being created every day, a new app being launched every minute – the actual experience of delivering or receiving care is changing scarcely.[281]

As with healthcare, the natural products world must address the gap between the "change layer" and the "reality layer" of the discovery and development of new natural products as therapies. One way this is possible is through new combinations of high-content screening platforms and computational analysis techniques, such as FUSION, "Function-First" lead generation, and image-based screening using disease models.[16,17,282,283] These efforts offer a much more thorough interrogation of the potential of molecular hits and offer a staircase between discovery and the clinic, albeit a very preliminary one.

Yet, this might strike some in the natural products academic field as an undue junket into an area of the field that belongs outside the university setting and squarely on the shoulders of for-profit pharmaceutical companies. Many have different, contrasting visions of the purposes of academic institutions with regard to drug discovery. Some see laboratory training of future scholars as a driving component of research. Others seek only to understand the most fundamental, basic scientific properties of natural products and the systems that generate them in order to inform and encourage future

applications of the studied system. Some leaders believe immediate application and translation of findings to societal benefit is critical to the field, especially in maintaining consistent funding sources. Frequently, these contrasting views seem to be distilled into two main camps: is the purpose of natural products academia to discover new drugs? Or alternatively, is the purpose to discover new *methods* for discovery of new drugs? It is the opinion of this author that there is a place and need for both as long as there is an earnest avoidance of derivative modes of operation and experimental thought, and a serious dedication by the individual researcher to seek new, novel, interesting, and meaningful methods, techniques, and exploration of the world of natural products.

References

(1) Katz, L., and Baltz, R. H. (2016) Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol. 43*, 155–176.

(2) Farnsworth, N. R., Akerele, O., Bingel, A. S., Soejarto, D. D., and Guo, Z. (1985) Medicinal plants in therapy. *Bull. World Health Organ. 63*, 965–981.

(3) Demain, A. L. (1999) Pharmaceutically active secondary metabolites of microorganisms. *Appl. Microbiol. Biotechnol. 52*, 455–63.

(4) Demain, A. L. (2006) From natural products discovery to commercialization: A success story. *J. Ind. Microbiol. Biotechnol. 33*, 486–495.

(5) Miller, L. H., and Su, X. (2011) Artemisinin: Discovery from the Chinese herbal garden. *Cell 146*, 855–858.

(6) Vahdat, L. T., Pruitt, B., Fabian, C. J., Rivera, R. R., Smith, D. A., Tan-Chiu, E., Wright, J., Tan, A. R., DaCosta, N. A., Chuang, E., Smith, J., O'Shaughnessy, J., Shuster, D. E., Meneses, N. L., Chandrawansa, K., Fang, F., Cole, P. E., Ashworth, S., and Blum, J. L. (2009) Phase II Study of Eribulin Mesylate, a Halichondrin B Analog, in Patients With Metastatic Breast Cancer Previously Treated With an Anthracycline and a Taxane. *J. Clin. Oncol. 27*, 2954–2961.

(7) Uemura, D., Takahashi, K., Yamamoto, T., Katayama, C., Tanaka, J., Okumura, Y., and Hirata, Y. (1985) Norhalichondrin A: an antitumor polyether macrolide from a marine sponge. *J. Am. Chem. Soc. 107*, 4796–4798.

(8) Bergmann, W., and Feeney, R. J. (1950) The isolation of a new thymine pentoside from sponges. *J. Am. Chem. Soc. 72*, 2809–2810.

(9) Löwenberg, B., Pabst, T., Vellenga, E., van Putten, W., Schouten, H. C., Graux, C., Ferrant, A., Sonneveld, P., Biemond, B. J., Gratwohl, A., de Greef, G. E., Verdonck, L. F., Schaafsma, M. R., Gregor, M., Theobald, M., Schanz, U., Maertens, J., Ossenkoppele, G. J., and Dutch-Belgian Cooperative Trial Group for Hemato-Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK) Collaborative Group. (2011) Cytarabine dose for acute myeloid leukemia. *N. Engl. J. Med. 364*, 1027–36.

(10) Waksman, S. A., and Lechevalier, H. A. (1949) Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, including Tuberculosis Organisms. *Science (80-.). 109*, 305–307.

(11) Calder, P. C. (2010) Omega-3 fatty acids and inflammatory processes. *Nutrients 2*, 355–374.

(12) Debono, M., Abbott, B. J., Molloy, R. M., Fukuda, D. S., Hunt, A. H., Daupert, V. M., Counter, F. T., Ott, J. L., Carrell, C. B., and Howard, L. C. (1988) Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J. Antibiot. (Tokyo). 41*, 1093–105.

(13) Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A., and Gordon, E. M. (1994) Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *J. Med. Chem. 37*, 1233–1251.

(14) Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod. 79*, 629–661.

(15) Demain, A. L. (2014) Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol. 41*, 185–201.

(16) Schulze, C. J., Bray, W. M., Woerhmann, M. H., Stuart, J., Lokey, R. S., and Linington, R. G. (2013) "Function-first" lead discovery: Mode of action profiling of natural product libraries using image-based screening. *Chem. Biol. 20*, 285–295.

(17) Potts, M. B., Kim, H. S., Fisher, K. W., Hu, Y., Carrasco, Y. P., Bulut, G. B., Ou, Y. H., Herrera-Herrera, M. L., Cubillos, F., Mendiratta, S., Xiao, G., Hofree, M., Ideker, T., Xie, Y., Huang, L. J. S., Lewis, R. E., MacMillan, J. B., and White, M. A. (2013) Using functional signature ontology (FUSION) to identify mechanisms of action for Natural products. *Sci. Signal. 6*, 1–14.

(18) Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov. 14*, 111–129.

(19) Gulder, T. A. M., and Moore, B. S. (2009) Chasing the treasures of the sea - bacterial marine natural products. *Curr. Opin. Microbiol. 12*, 252–60.

(20) Gerwick, W. H., and Moore, B. S. (2012) Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol. 19*, 85–98.

(21) Rangel, M. (2015) An overview of the marine natural products in clinical trials and on the market. *J. Coast. Life Med. 3*, 421–428.

(22) Kanoh, K., Kohno, S., Asari, T., Harada, T., Katada, J., Muramatsu, M., Kawashima, H., Sekiya, H., and Uno, I. (1997) (-)-phenylahistin: A new mammalian cell cycle inhibitor produced by Aspergillus ustus. *Bioorganic Med. Chem. Lett. 7*, 2847–2852.

(23) Rinehart, K. L., Holt, T. G., Fregeau, N. L., Stroh, J. G., Keifer, P. A., Sun, F., Li, L. H., and Martin, D. G. (1990) Ecteinascidins 729, 743, 745, 759A, 759B, and 770: Potent Antitumor Agents from the Caribbean Tunicate Ecteinascidia turbinata. *J. Org. Chem. 55*, 4512–4515.

(24) Francisco, J. A., Cerveny, C. G., Meyer, D. L., Mixan, B. J., Klussman, K., Chace, D. F., Rejniak, S. X., Gordon, K. A., DeBlanc, R., Toki, B. E., Law, C. L., Doronina, S. O., Siegall, C. B., Senter, P. D., and Wahl, A. F. (2003) cAC10-vcMMAE, an anti-CD30-monomethyl auristatin E conjugate with potent and selective antitumor activity. *Blood 102*, 1458–1465.

(25) Pettit, G. R., Kamano, Y., Herald, C. L., Tuinman, A. A., Boettner, F. E., Kizu, H., Schmidt, J. M., Baczynskyi, L., Tomer, K. B., and Bontems, R. J. (1987) The Isolation and Structure of a Remarkable Marine Animal Antineoplastic Constituent: Dolastatin 10. *J. Am. Chem. Soc. 109*, 6883–6885.

(26) Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science (80-.). 282*, 63–68.

(27) Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., and Stuppner, H. (2015) Discovery

and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol. Adv. 33*, 1582–1614.

(28) Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., and Ikeda, H. (2015) Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. 112*, 857–862.

(29) Bugni, T. S., and Ireland, C. M. (2004) Marine-derived fungi: A chemically and biologically diverse group of microorganisms. *Nat. Prod. Rep. 21*, 143–163.

(30) Ebel, R. (2010) Terpenes from Marine-Derived Fungi. *Mar. Drugs 8*, 2340–2368.

(31) Keane, K. M., Bell, P. G., Lodge, J. K., Constantinou, C. L., Jenkinson, S. E., Bass, R., and Howatson, G. (2016) Phytochemical uptake following human consumption of Montmorency tart cherry (L. Prunus cerasus) and influence of phenolic acids on vascular smooth muscle cells in vitro. *Eur. J. Nutr. 55*, 1695–705.

(32) Sardar, D., and Schmidt, E. W. (2016) Combinatorial biosynthesis of RiPPs: docking with marine life. *Curr. Opin. Chem. Biol. 31*, 15–21.

(33) Mocek, U., Knaggs, A. R., Tsuchiya, R., Nguyen, T., Beale, J. M., and Floss, H. G. (1993) Biosynthesis of the modified peptide antibiotic nosiheptide in Streptomyces actuosus. *J. Am. Chem. Soc. 115*, 7557–7568.

(34) Finking, R., and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides1. *Annu. Rev. Microbiol. 58*, 453–88.

(35) Rinehart, K., Gloer, J., Hughes, R., Renis, H., McGovren, J., Swynenberg, E., Stringfellow, D., Kuentzel, S., and Li, L. (1981) Didemnins: antiviral and antitumor depsipeptides from a caribbean tunicate. *Science (80-.). 212*, 933–935.

(36) Jiang, T. L., Liu, R. H., and Salmon, S. E. (1983) Antitumor activity of didemnin B in the human tumor stem cell assay. *Cancer Chemother. Pharmacol. 11*, 1–4.

(37) Stewart, J. A., Low, J. B., Roberts, J. D., and Blow, A. (1991) A phase i clinical trial of didemnin B. *Cancer 68*, 2550–2554.

(38) Shimizu, Y., Ogata, H., and Goto, S. (2017) Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem 18*, 50–65.

(39) Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol. 7*, 285–295.

(40) Moore, B. S., and Hopke, J. N. (2001) Discovery of a New Bacterial Polyketide Biosynthetic Pathway. *ChemBioChem 2*, 35–38.

(41) Muller, R. (2004) Don't Classify Polyketide Synthases. *Chem. Biol. 11*, 4–6.

(42) Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases. *Nat. Prod. Rep. 29*, 1050–1073.

(43) Piel, J. (2010) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep. 27*, 996–1047.

(44) Pieper, R., Kao, C., Khosla, C., Luo, G., and Cane, D. E. (1996) Specificity and versatility in erythromycin biosynthesis. *Chem. Soc. Rev. 25*, 297–302.

(45) Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., and Murrell, J. M. (2001) 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. 27*, 378–385.

(46) Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., and Murrell, J. M. (2002) Cloning and Characterization of the Bleomycin Biosynthetic Gene Cluster from Streptomyces v erticillus ATCC15003. *J. Nat. Prod. 65*, 422–431.

(47) Schmidt-Dannert, C. (2014) Biosynthesis of Terpenoid Natural Products in Fungi, in *Educational Technology and Society*, pp 19–61.

(48) Gul, W., and Hamann, M. T. (2005) Indole alkaloid marine natural products: An established source of cancer drug leads with considerable promise for the control of parasitic, neurological and other diseases. *Life Sci. 78*, 442–453.

(49) Manivasagan, P., Venkatesan, J., Sivakumar, K., and Kim, S.-K. (2014) Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiol. Res. 169*, 262–278.

(50) Baltz, R. H. (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol. 8*, 557–63.

(51) Jones, A. C., Gu, L., Sorrels, C. M., Sherman, D. H., and Gerwick, W. H. (2009) New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Curr. Opin. Chem. Biol. 13*, 216–223.

(52) Cameron Coates, R., Trentacoste, E., and Gerwick, W. H. (2013) Bioactive and Novel Chemicals from Microalgae, in *Handbook of Microalgal Culture*, pp 504–531. John Wiley & Sons, Ltd, Oxford, UK.

(53) Agger, S. A., Lopez-Gallego, F., Hoye, T. R., and Schmidt-Dannert, C. (2008) Identification of sesquiterpene synthases from Nostoc punctiforme PCC 73102 and Nostoc sp. strain PCC 7120. *J. Bacteriol. 190*, 6084–6096.

(54) Pattanaik, B., and Lindberg, P. (2015) Terpenoids and Their Biosynthesis in Cyanobacteria. *Life 5*, 269–293.

(55) Hillwig, M. L., Zhu, Q., and Liu, X. (2014) Biosynthesis of ambiguine indole alkaloids in cyanobacterium Fischerella ambigua. *ACS Chem. Biol. 9*, 372–377.

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

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

(58) Ongley, S. E., Bian, X., Zhang, Y., Chau, R., Gerwick, W. H., Müller, R., and Neilan, B. A. (2013) High-Titer Heterologous Production in E. coli of Lyngbyatoxin, a Protein Kinase C Activator from an Uncultured Marine Cyanobacterium. *ACS Chem. Biol. 8*, 1888–1893.

(59) Pearson, L., Mihali, T., Moffitt, M., Kellmann, R., and Neilan, B. (2010) On the Chemistry, Toxicology and Genetics of the Cyanobacterial Toxins, Microcystin, Nodularin, Saxitoxin and Cylindrospermopsin. *Mar. Drugs 8*, 1650–1680.

(60) Kellmann, R., Mihali, T. K., Young, J. J., Pickford, R., Pomati, F., and Neilan, B. A. (2008) Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Appl. Environ. Microbiol. 74*, 4044–4053.

(61) Balskus, E. P., Case, R. J., and Walsh, C. T. (2011) The biosynthesis of cyanobacterial sunscreen scytonemin in intertidal microbial mat communities. *FEMS Microbiol. Ecol. 77*, 322–332.

(62) Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc. Natl. Acad. Sci. 102*, 7315–7320.

(63) Sivonen, K., Leikoski, N., Fewer, D. P., and Jokela, J. (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl. Microbiol. Biotechnol. 86*, 1213–1225.

(64) Lee, J., Mcintosh, J., Hathaway, B. J., and Schmidt, E. W. (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J. Am. Chem. Soc. 131*, 2122–2124.

(65) Gu, W., Dong, S. H., Sarkar, S., Nair, S. K., and Schmidt, E. W. (2018) The Biochemistry and Structural Biology of Cyanobactin Pathways: Enabling Combinatorial Biosynthesis, in *Methods in Enzymology* 1st ed., pp 113–163. Elsevier Inc.

(66) Wiseman, S. B., Wan, Y., Chang, H., Zhang, X., Hecker, M., Jones, P. D., and Giesy, J. P. (2011) Polybrominated diphenyl ethers and their hydroxylated/methoxylated analogs: Environmental sources, metabolic relationships, and relative toxicities. *Mar. Pollut. Bull. 63*, 179–188.

(67) Agarwal, V., El Gamal, A. A., Yamanaka, K., Poth, D., Kersten, R. D., Schorn, M., Allen, E. E., and Moore, B. S. (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol. 10*, 640–647.

(68) Agarwal, V., Blanton, J. M., Podell, S., Taton, A., Schorn, M. A., Busch, J., Lin, Z., Schmidt, E. W., Jensen, P. R., Paul, V. J., Biggs, J. S., Golden, J. W., Allen, E. E., and Moore, B. S. (2017) Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol. 13*, 537–543.

(69) Helfrich, E. J. N., and Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep. 33*, 231–316.

(70) Kust, A., Mareš, J., Jokela, J., Urajová, P., Hájek, J., Saurav, K., Voráčová, K., Fewer, D. P., Haapaniemi, E., Permi, P., Řeháková, K., Sivonen, K., and Hrouzek, P. (2018) Discovery of a Pederin Family Compound in a Nonsymbiotic Bloom-Forming Cyanobacterium. *ACS Chem. Biol. 13*, 1123–1129.

(71) Kampa, A., Gagunashvili, A. N., Gulder, T. A. M., Morinaka, B. I., Daolio, C., Godejohann, M., Miao, V. P. W., Piel, J., and Andresson, O. S. (2013) Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci. 110*, E3129–E3137.

(72) Elshahawi, S. I., Trindade-Silva, A. E., Hanora, A., Han, A. W., Flores, M. S., Vizzoni, V., Schrago, C. G., Soares, C. A., Concepcion, G. P., Distel, D. L., Schmidt, E. W., and Haygood, M. G. (2013) Boronated tartrolon antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm gills. *Proc. Natl. Acad. Sci. 110*, E295–E304.

(73) Banker, R., and Carmeli, S. (1998) Tenuecyclamides A-D, cyclic hexapeptides from the cyanobacterium Nostoc spongiaeforme var. tenue. *J. Nat. Prod. 61*, 1248–1251.

(74) Hemscheidt, T., Puglisi, M. P., Larsen, L. K., Patterson, G. M. L., Moore, R. E., Rios, J. L., and Clardy, J. (1994) Structure and biosynthesis of borophycin, a new boeseken complex of boric acid from a marine strain of the blue-green alga Nostoc linckia. *J. Org. Chem. 59*, 3467–3471.

(75) Bertin, M. J., Vulpanovici, A., Monroe, E. A., Korobeynikov, A., Sherman, D. H., Gerwick, L., and Gerwick, W. H. (2016) The Phormidolide Biosynthetic Gene Cluster: A trans-AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide. *ChemBioChem 17*, 164–173.

(76) Humisto, A., Jokela, J., Liu, L., Wahlsten, M., Wang, H., Permi, P., Machado, J. P., Antunes, A., Fewer, D. P., and Sivonen, K. (2017) The Swinholide Biosynthesis Gene Cluster from a Terrestrial Cyanobacterium, Nostoc sp. Strain UHCC 0450. *Appl. Environ. Microbiol.* (Drake, H. L., Ed.) *84*, e02321-17.

(77) Carmely, S., and Kashman, Y. (1985) Structure of swinholide-a, a new macrolide from the marine sponge. *Tetrahedron Lett. 26*, 511–514.

(78) Andrianasolo, E. H., Gross, H., Goeger, D., Musafija-Girt, M., McPhail, K., Leal, R. M., Mooberry, S. L., and Gerwick, W. H. (2005) Isolation of swinholide A and related glycosylated derivatives from two field collections of marine cyanobacteria. *Org. Lett. 7*, 1375–8.

(79) Ishibashi, M., Moore, R. E., Patterson, G. M. L., Xu, C., and Clardy, J. (1986) Scytophycins, cytotoxic and antimycotic agents from the cyanophyte Scytonema pseudohofmanni. *J. Org. Chem. 51*, 5300–5306.

(80) Rounge, T. B., Rohrlack, T., Tooming-Klunderud, A., Kristensen, T., and Jakobsen, K. S. (2007) Comparison of cyanopeptolin genes in Planktothrix, Microcystis, and Anabaena strains: Evidence for independent evolution within each genus. *Appl. Environ. Microbiol. 73*, 7322–7330.

(81) Vestola, J., Shishido, T. K., Jokela, J., Fewer, D. P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L., and Sivonen, K. (2014) Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. *Proc. Natl. Acad. Sci. 111*, E1909–E1917.

(82) Moffitt, M. C., and Neilan, B. A. (2004) Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl. Environ. Microbiol. 70*, 6353–6362.

(83) Bui, T. H., Wray, V., Nimtz, M., Fossen, T., Preisitsch, M., Schröder, G., Wende, K.,

Heiden, S. E., and Mundt, S. (2014) Balticidins A-D, antifungal hassallidin-like lipopeptides from the Baltic Sea cyanobacterium Anabaena cylindrica Bio33. *J. Nat. Prod. 77*, 1287–1296.

(84) Mootz, H. D., and Marahiel, M. A. (1997) The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol. 179*, 6843–50.

(85) Pancrace, C., Gugger, M., and Calteau, A. (2017) Genomics of NRPS/PKS Biosynthetic Gene Clusters in Cyanobacteria, in *Cyanobacteria: Omics and Manipulation* (Los, D., Ed.), pp 55–74.

(86) Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem. 1*, 55–68.

(87) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev. 64*, 4–17.

(88) Tan, L. T. (2010) Filamentous tropical marine cyanobacteria: a rich source of natural products for anticancer drug discovery. *J. Appl. Phycol. 22*, 659–676.

(89) Chang, Z., Sitachitta, N., Rossi, J. V, Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 67*, 1356–67.

(90) Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., Hkansson, K., Wipf, P., Smith, J. L., Gerwick, W. H., and Sherman, D. H. (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature 459*, 731–735.

(91) Gehret, J. J., Gu, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. (2011) Terminal alkene formation by the thioesterase of curacin A biosynthesis: Structure of a decarboxylating thioesterase. *J. Biol. Chem. 286*, 14445–14454.

(92) Gu, L., Geders, T. W., Wang, B., Gerwick, W. H., Håkansson, K., Smith, J. L., and Sherman, D. H. (2007) GNAT-like strategy for polyketide chain initiation. *Science (80-.). 318*, 970–974.

(93) Skiba, M. A., Sikkema, A. P., Fiers, W. D., Gerwick, W. H., Sherman, D. H., Aldrich, C. C., and Smith, J. L. (2016) Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol. 11*, 3319–3327.

(94) Mihali, T. K., Kellmann, R., Muenchhoff, J., Barrow, K. D., and Neilan, B. A. (2008) Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. *Appl. Environ. Microbiol. 74*, 716–722.

(95) Muenchhoff, J., Siddiqui, K. S., Poljak, A., Raftery, M. J., Barrow, K. D., and Neilan, B. A. (2010) A novel prokaryotic l-arginine: Glycine amidinotransferase is involved in cylindrospermopsin biosynthesis. *FEBS J. 277*, 3844–3860.

(96) Galonić, D. P., Vaillancourt, F. H., and Walsh, C. T. (2006) Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *J. Am. Chem. Soc. 128*, 3900–3901.

(97) Flatt, P. M., O'Connell, S. J., McPhail, K. L., Zeller, G., Willis, C. L., Sherman, D. H., and Gerwick, W. H. (2006) Characterization of the initial enzymatic steps of barbamide biosynthesis. *J. Nat. Prod. 69*, 938–944.

(98) Orjala, J., and Gerwick, W. H. (1996) Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 59*, 427–430.

(99) Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem. 277*, 13082–13090.

(100) Méjean, A., Mann, S., Maldiney, T., Vassiliadis, G., Lequin, O., and Ploux, O. (2009) Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-a in the cyanobacterium Oscillatoria PCC 6506 occurs on a modular polyketide synthase initiated by L-proline. *J. Am. Chem. Soc. 131*, 7512–7513.

(101) Oja, T., Palmu, K., Lehmussola, H., Leppäranta, O., Hännikäinen, K., Niemi, J., Mäntsälä, P., and Metsä-Ketelä, M. (2008) Characterization of the Alnumycin Gene Cluster Reveals Unusual Gene Products for Pyran Ring Formation and Dioxan Biosynthesis. *Chem. Biol. 15*, 1046–1057.

(102) Moss, N. A., Leão, T., Rankin, M. R., McCullough, T. M., Qu, P., Korobeynikov, A., Smith, J. L., Gerwick, L., and Gerwick, W. H. (2018) Ketoreductase Domain Dysfunction Expands Chemodiversity: Malyngamide Biosynthesis in the Cyanobacterium Okeania hirsuta. *ACS Chem. Biol. 13*, 3385–3395.

(103) Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. (2004) Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium Lyngbya majuscula. *Chem. Biol. 11*, 817–833.

(104) Mevers, E., Liu, W. T., Engene, N., Mohimani, H., Byrum, T., Pevzner, P. A., Dorrestein, P. C., Spadafora, C., and Gerwick, W. H. (2011) Cytotoxic veraguamides, alkynyl bromide-containing cyclic depsipeptides from the marine cyanobacterium cf. Oscillatoria margaritifera. *J. Nat. Prod. 74*, 928–936.

(105) Hooper, G. J., Orjala, J., Schatzman, R. C., and Gerwick, W. H. (1998) Carmabins A and B, new lipopeptides from the Caribbean cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 61*, 529–533.

(106) McPhail, K. L., Correa, J., Linington, R. G., González, J., Ortega-Barría, E., Capson, T. L., and Gerwick, W. H. (2007) Antimalarial linear lipopeptides from a panamanian strain of the marine cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 70*, 984–988.

(107) Nogle, L. M., and Gerwick, W. H. (2002) Isolation of four new cyclic depsipeptides, antanapeptins A-D, and dolastatin 16 from a Madagascan collection of Lyngbya majuscula. *J. Nat. Prod. 65*, 21–24.

(108) Zhu, X., Liu, J., and Zhang, W. (2015) De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol. 11*, 115–120.

(109) Robbins, T., Liu, Y. C., Cane, D. E., and Khosla, C. (2016) Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol. 41*, 10–18.

(110) Weissman, K. J. (2015) The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol. 11*, 660–670.

(111) Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. (2003) The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol. 10*, 723–731.

(112) Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013) Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chem. Biol. 20*, 1340–1351.

(113) Dehling, E., Volkmann, G., Matern, J. C. J., Dörner, W., Alfermann, J., Diecker, J., and Mootz, H. D. (2016) Mapping of the Communication-Mediating Interface in Nonribosomal Peptide Synthetases Using a Genetically Encoded Photocrosslinker Supports an Upside-Down Helix-Hand Motif. *J. Mol. Biol. 428*, 4345–4360.

(114) Richter, C. D., Nietlispach, D., Broadhurst, R. W., and Weissman, K. J. (2008) Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol. 4*, 75–81.

(115) O'Connor, S. E., Walsh, C. T., and Liu, F. (2003) Biosynthesis of Epothilone Intermediates with Alternate Starter Units: Engineering Polyketide–Nonribosomal Interfaces. *Angew. Chemie Int. Ed. 42*, 3917–3921.

(116) Hacker, C., Cai, X., Kegler, C., Zhao, L., Weickhmann, A. K., Wurm, J. P., Bode, H. B., and Wöhnert, J. (2018) Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nat. Commun. 9*, 4366.

(117) Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T., and Drennan, C. L. (2016) Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. 113*, 12432–12437.

(118) Liu, F., Garneau, S., and Walsh, C. T. (2004) Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chem. Biol. 11*, 1533–42.

(119) Miller, D. A., Luo, L., Hillson, N., Keating, T. A., and Walsh, C. T. (2002) Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem. Biol. 9*, 333–344.

(120) Du, L., Sánchez, C., Chen, M., Edwards, D. J., and Shen, B. (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol. 7*, 623–642.

(121) Beck, B. J., Yoon, Y. J., Reynolds, K. A., and Sherman, D. H. (2002) The hidden steps of domain skipping: Macrolactone ring size determination in the pikromycin modular polyketide synthase. *Chem. Biol. 9*, 575–583.

(122) Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., Wakimoto, T., Ohnishi, Y., Matsunaga, S., and Takada, K. (2016) Biosynthetic Gene Cluster for Surugamide A Encompasses an Unrelated Decapeptide, Surugamide F. *ChemBioChem 17*, 1709–1712.

(123) Ross, A. C., Xu, Y., Lu, L., Kersten, R. D., Shao, Z., Al-Suwailem, A. M., Dorrestein, P. C., Qian, P., and Moore, B. S. (2013) Biosynthetic Multitasking Facilitates Thalassospiramide Structural Diversity in Marine Bacteria. *J. Am. Chem. Soc. 135*, 1155–1162.

(124) Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J. N., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kötter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gómez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J. M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Müller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B., Payne, S. M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Süssmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J. M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., and Glöckner, F. O. (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol. 11*, 625–631.

(125) Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet. 17*, 333–51.

(126) Medema, M. H., and Fischbach, M. A. (2015) Computational approaches to natural product discovery. *Nat. Chem. Biol. 11*, 639–648.

(127) Minowa, Y., Araki, M., and Kanehisa, M. (2007) Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol. 368*, 1500–1517.

(128) Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., and Madden, T. L. (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct 7*, 1–14.

(129) Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012) The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One* (de Crécy-Lagard, V., Ed.) *7*, e34064.

(130) Chevrette, M. G., Aicheler, F., Kohlbacher, O., Currie, C. R., and Medema, M. H. (2017) SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* (Birol, I., Ed.) *33*, 3202–3210.

(131) Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res. 39*, W362–W367.

(132) Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., De Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., and Medema, M. H. (2017) AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res. 45*, W36–W41.

(133) Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., Boya, C. A. P., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B., Pogliano, K., Linington, R. G., Gutiérrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., and Bandeira, N. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol. 34*, 828–837.

(134) Johnston, C. W., Skinnider, M. A., Wyatt, M. A., Li, X., Ranieri, M. R. M., Yang, L., Zechel, D. L., Ma, B., and Magarvey, N. A. (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun. 6*, 8421.

(135) Kim, E., Moore, B. S., and Yoon, Y. J. (2015) Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol. 11*, 649–59.

(136) Gomez-Escribano, J. P., and Bibb, M. J. (2014) Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol. 41*, 425–31.

(137) Ongley, S. E., Bian, X., Neilan, B. A., and Müller, R. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep. 30*, 1121–1138.

(138) Brady, S. F., Chao, C. J., Handelsman, J., and Clardy, J. (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett. 3*, 1981–4.

(139) Nah, H.-J., Pyeon, H.-R., Kang, S.-H., Choi, S.-S., and Kim, E.-S. (2017) Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Front. Microbiol. 8*, 394.

(140) Videau, P., Wells, K. N., Singh, A. J., Gerwick, W. H., and Philmus, B. (2016) Assessment of Anabaena sp. Strain PCC 7120 as a Heterologous Expression Host for Cyanobacterial Natural

Products: Production of Lyngbyatoxin A. *ACS Synth. Biol. 5*, 978–988.

(141) Gross, H., McPhail, K. L., Goeger, D. E., Valeriote, F. A., and Gerwick, W. H. (2010) Two cytotoxic stereoisomers of malyngamide C, 8-epi-malyngamide C and 8-O-acetyl-8-epi-malyngamide C, from the marine cyanobacterium Lyngbya majuscula. *Phytochemistry 71*, 1729–1735.

(142) Villa, F. A., Lieske, K., and Gerwick, L. (2010) Selective MyD88-dependent pathway inhibition by the cyanobacterial natural product malyngamide F acetate. *Eur. J. Pharmacol. 629*, 140–146.

(143) Mynderse, J. S., and Moore, R. E. (1978) Malyngamides D and E, Two trans-7-Methoxy-9-methylhexadec-4-enamides from a Deep Water Variety of the Marine Cyanophyte Lyngbya majuscula. *J. Org. Chem. 43*, 4359–4363.

(144) Tidgewell, K., Clark, B. R., and Gerwick, W. H. (2010) The Natural Products Chemistry of Cyanobacteria, in *Comprehensive Natural Products II* (Lewis, M., Liu, H.-W., Townsend, C., and Ebizuka, Y., Eds.), pp 141–188. Elsevier.

(145) Tan, L. T., Okino, T., and Gerwick, W. H. (2000) Hermitamides A and B, Toxic Malyngamide-Type Natural Products from the Marine Cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 63*, 952–955.

(146) Wan, F., and Erickson, K. L. (1999) Serinol-Derived Malyngamides from an Australian Cyanobacterium. *J. Nat. Prod. 62*, 1696–1699.

(147) Praud, A., Valls, R., Piovetti, L., and Banaigs, B. (1993) Malyngamide G : Proposition de structure pour un nouvel amide chloré d'une algue bleu-verte epiphyte de Cystoseira crinita. *Tetrahedron Lett. 34*, 5437–5440.

(148) McPhail, K. L., and Gerwick, W. H. (2003) Three new malyngamides from a Papua New Guinea collection of the marine cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 66*, 132–135.

(149) Malloy, K. L., Villa, F. A., Engene, N., Matainaho, T., Gerwick, L., and Gerwick, W. H. (2011) Malyngamide 2, an oxidized lipopeptide with nitric oxide inhibiting activity from a Papua New Guinea marine cyanobacterium. *J. Nat. Prod. 74*, 95–98.

(150) Appleton, D. R., Sewell, M. A., Berridge, M. V., and Copp, B. R. (2002) A new biologically active malyngamide from a New Zealand collection of the sea hare Bursatella leachii. *J. Nat. Prod. 65*, 630–631.

(151) Orjala, J., Nagle, D., and Gerwick, W. H. (1995) Malyngamide H, an ichthyotoxic amide possessing a new carbon skeleton from the caribbean cyanobacterium lyngbya majuscula. *J. Nat. Prod. 58*, 764–768.

(152) Wu, M., Milligan, K. E., and Gerwick, W. H. (1997) Three new malyngamides from the marine cyanobacterium Lyngbya majuscula. *Tetrahedron 53*, 15983–15990.

(153) Kan, Y., Fujita, T., Nagai, H., Sakamoto, B., and Hokama, Y. (1998) Malyngamides M and N from the Hawaiian red alga Gracilaria coronopifolia. *J. Nat. Prod. 61*, 152–155.

(154) Ainslie, R. D., Barchi, J. J., Kuniyoshi, M., Moore, R. E., and Mynderse, J. S. (1985)

Structure of Malyngamide C. *J. Org. Chem. 50*, 2859–2862.

(155) Todd, J. S., and Gerwick, W. H. (1995) Malyngamide I from the tropical marine cyanobacterium Lyngbya majuscula and the probable structure revision of stylocheilamide. *Tetrahedron Lett. 36*, 7837–7840.

(156) Meng, L., Mohan, R., Kwok, B. H. B., Elofsson, M., Sin, N., and Crews, C. M. (1999) Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proc. Natl. Acad. Sci. 96*, 10403–10408.

(157) Engene, N., Paul, V. J., Byrum, T., Gerwick, W. H., Thor, A., and Ellisman, M. H. (2013) Five chemically rich species of tropical marine cyanobacteria of the genus Okeania gen. nov. (Oscillatoriales, Cyanoprokaryota). *J. Phycol. 49*, 1095–1106.

(158) Petitbois, J. G., Casalme, L. O., Lopez, J. A. V., Alarif, W. M., Abdel-Lateff, A., Al-Lihaibi, S. S., Yoshimura, E., Nogata, Y., Umezawa, T., Matsuda, F., and Okino, T. (2017) Serinolamides and Lyngbyabellins from an Okeania sp. Cyanobacterium Collected from the Red Sea. *J. Nat. Prod. 80*, 2708–2715.

(159) Winnikoff, J. R., Glukhov, E., Watrous, J., Dorrestein, P. C., and Gerwick, W. H. (2014) Quantitative molecular networking to profile marine cyanobacterial metabolomes. *J. Antibiot. (Tokyo). 67*, 105–112.

(160) Kleigrewe, K., Almaliti, J., Tian, I. Y., Kinnel, R. B., Korobeynikov, A., Monroe, E. A., Duggan, B. M., Di Marzo, V., Sherman, D. H., Dorrestein, P. C., Gerwick, L., and Gerwick, W. H. (2015) Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod. 78*, 1671–1682.

(161) Skiba, M. A., Sikkema, A. P., Moss, N. A., Lowell, A. N., Su, M., Sturgis, R. M., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2018) Biosynthesis of t-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module. *ACS Chem. Biol. 13*, 1640–1650.

(162) Christensen, Q. H., and Cronan, J. E. (2010) Lipoic acid synthesis: A new family of octanoyltransferases generally annotated as lipoate protein ligases. *Biochemistry 49*, 10024–10036.

(163) Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., and Burkart, M. D. (2014) The phosphopantetheinyl transferases: Catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep. 31*, 61–108.

(164) Edwards, D. J., Marquez, B. L., Nogle, L. M., Mcphail, K., Goeger, D. E., Roberts, M. A., Gerwick, W. H., Table, S. H., Spectral, C. N. M. R., and Jamaicamide, B. (2004) Structure and Biosynthesis of the Jamaicamides , New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium Lyngbya majuscula. *Chem. Biol. 11*, 817–813.

(165) Fiers, W. D., Dodge, G. J., Sherman, D. H., Smith, J. L., and Aldrich, C. C. (2016) Vinylogous Dehydration by a Polyketide Dehydratase Domain in Curacin Biosynthesis. *J. Am. Chem. Soc. 138*, 16024–16036.

(166) Geders, T. W., Gu, L., Mowers, J. C., Liu, H., Gerwick, W. H., Håkansson, K., Sherman, D. H., and Smith, J. L. (2007) Crystal structure of the ECH2 catalytic domain of CurF from Lyngbya majuscula: Insights into a decarboxylase involved in polyketide chain β-branching. *J. Biol. Chem. 282*,

35954–35963.

(167) Cronan, J. E. (2016) Assembly of Lipoic Acid on Its Cognate Enzymes: an Extraordinary and Essential Biosynthetic Pathway. *Microbiol. Mol. Biol. Rev. 80*, 429–50.

(168) Park, N. S., Myeong, J. S., Park, H. J., Han, K., Kim, S. N., and Kim, E. S. (2005) Characterization and culture optimization of regiospecific cyclosporin hydroxylation in rare actinomycetes species. *J. Microbiol. Biotechnol. 15*, 188–191.

(169) Lowry, B., Li, X., Robbins, T., Cane, D. E., and Khosla, C. (2016) A turnstile mechanism for the controlled growth of biosynthetic intermediates on assembly line polyketide synthases. *ACS Cent. Sci. 2*, 14–20.

(170) Liu, X., and Walsh, C. T. (2009) Cyclopiazonic acid biosynthesis in Aspergillus sp.: Characterization of a reductase-like R* domain in cyclopiazonate synthetase that forms and releases cyclo-acetoacetyl-L-tryptophan. *Biochemistry 48*, 8746–8757.

(171) Ehmann, D. E., Gehring, A. M., and Walsh, C. T. (1999) Lysine biosynthesis in Saccharomyces cerevisiae: Mechanism of alpha- aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry 38*, 6171–6177.

(172) Li, L., Deng, W., Song, J., Ding, W., Zhao, Q. F., Peng, C., Song, W. W., Tang, G. L., and Liu, W. (2008) Characterization of the saframycin a gene cluster from Streptomyces lavendulae NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. *J. Bacteriol. 190*, 251–263.

(173) Liu, X., Biswas, S., Berg, M. G., Antapli, C. M., Xie, F., Wang, Q., Tang, M. C., Tang, G. L., Zhang, L., Dreyfuss, G., and Cheng, Y. Q. (2013) Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from Burkholderia thailandensis MSMB43. *J. Nat. Prod. 76*, 685–693.

(174) Jiang, Y., Song, G., Pan, Q., Yang, Y., and Li, R. (2015) Identification of genes for anatoxin-a biosynthesis in Cuspidothrix issatschenkoi. *Harmful Algae 46*, 43–48.

(175) Keatinge-Clay, A. T. (2007) A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol. 14*, 898–908.

(176) Keatinge-Clay, A. T. (2016) Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep. 33*, 141–149.

(177) Caffrey, P. (2003) Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *Chembiochem 4*, 654–7.

(178) Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. *Gene 249*, 153–160.

(179) Aparicio, J. F., Caffrey, P., Gil, J. A., and Zotchev, S. B. (2003) Polyene antibiotic biosynthesis gene clusters. *Appl. Microbiol. Biotechnol. 61*, 179–188.

(180) Aparicio, J. F., Molnár, I., Schwecke, T., König, A., Haydock, S. F., Khaw, L. E., Staunton,

J., and Leadlay, P. F. (1996) Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: Analysis of the enzymatic domains in the modular polyketide synthase. *Gene 169*, 9–16.

(181) Jarabak, J., and Sack, G. H. (1969) A Soluble 17β-Hydroxysteroid Dehydrogenase From Human Placenta. The Binding of Pyridine Nucleotides and Steroids. *Biochemistry 8*, 2203–2212.

(182) Castenholz, R. W. (1988) Culturing Methods for Cyanobacteria. *Methods Enzymol. 167*, 68–93.

(183) Moss, N. A., Leao, T., Glukhov, E., Gerwick, L., and Gerwick, W. H. (2018) Collection, Culturing, and Genome Analyses of Tropical Marine Filamentous Benthic Cyanobacteria. *Methods Enzymol.* 1st ed. Elsevier Inc.

(184) Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol. 19*, 455–477.

(185) Podell, S., and Gaasterland, T. (2007) DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol. 8*, R16.

(186) Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C. A., Sivonen, K., Piel, J., and Gugger, M. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics 15*, 1–14.

(187) Leao, T., Castelão, G., Korobeynikov, A., Monroe, E. A., Podell, S., Glukhov, E., Allen, E. E., Gerwick, W. H., and Gerwick, L. (2017) Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus *Moorea*. *Proc. Natl. Acad. Sci. 114*, 3198–3203.

(188) Kelly, L. A., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc. 10*, 845–858.

(189) Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191.

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

(191) Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res. 46*, W296–W303.

(192) Schrodinger, L. (2010) The PyMOL Molecular Graphics System, version 1.4. New York City, NY.

(193) Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M. Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S. L., Nuwaysir, L. M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E. W., Moritz, R. L.,

Katz, J. E., Agus, D. B., MacCoss, M., Tabb, D. L., and Mallick, P. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol. 30*, 918–920.

(194) Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics 27*, 431–432.

(195) Capone, D. G. (1993) Determination of nitrogenase activity in aquatic samples using the acetylene reduction procedure, in *Handbook of methods in aquatic microbial ecology* (Kemp, P. F., Sherr, B. F., Sherr, E. B., and Cole, J. J., Eds.), pp 621–631. Lewis, Boca Raton, FL.

(196) Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., and Dorrestein, P. C. (2012) Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. 109*, E1743–E1752.

(197) Wright, A. D., Coll, J. C., and Price, I. R. (1990) Tropical Marine Algae, VII. The Chemical Composition of Marine Algae from North Queensland Waters. *J. Nat. Prod. 53*, 845–861.

(198) Shao, C. L., Linington, R. G., Balunas, M. J., Centeno, A., Boudreau, P., Zhang, C., Engene, N., Spadafora, C., Mutka, T. S., Kyle, D. E., Gerwick, L., Wang, C. Y., and Gerwick, W. H. (2015) Bastimolide A, a Potent Antimalarial Polyhydroxy Macrolide from the Marine Cyanobacterium Okeania hirsuta. *J. Org. Chem. 80*, 7849–7855.

(199) Linington, R. G., González, J., Ureña, L. D., Romero, L. I., Ortega-Barría, E., and Gerwick, W. H. (2007) Venturamides A and B: Antimalarial constituents of the Panamanian marine cyanobacterium Oscillatoria sp. *J. Nat. Prod. 70*, 397–401.

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

(201) Grindberg, R. V., Ishoey, T., Brinza, D., Esquenazi, E., Coates, R. C., Liu, W., Gerwick, L., Dorrestein, P. C., Pevzner, P., Lasken, R., and Gerwick, W. H. (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One 6*, e18565.

(202) Kinghorn, A., Su, B.-, Lee, D., Gu, J.-, and Pezzuto, J. (2003) Cancer Chemopreventive Agents Discovered by Activity-Guided Fractionation: An Update. *Curr. Org. Chem. 7*, 213–226.

(203) Yang, X., Summerhurst, D. K., Koval, S. F., Ficker, C., Smith, M. L., and Bernards, M. A. (2001) Isolation of an antimicrobial compound fromImpatiens balsamina L. using bioassay-guided fractionation. *Phyther. Res. 15*, 676–680.

(204) Meyer, B., Ferrigni, N., Putnam, J., Jacobsen, L., Nichols, D., and McLaughlin, J. (1982) Brine Shrimp: A Convenient General Bioassay for Active Plant Constituents. *Planta Med. 45*, 31–34.

(205) Jeon, H., Lim, C., Lee, J. M., and Kim, S. (2015) Chemical assay-guided natural product isolation via solid-supported chemodosimetric fluorescent probe. *Chem. Sci. 6*, 2806–2811.

(206) Odendaal, A. Y., Trader, D. J., and Carlson, E. E. (2011) Chemoselective enrichment for natural products discovery. *Chem. Sci. 2*, 760.

(207) Carlson, E. E., and Cravatt, B. F. (2007) Chemoselective probes for metabolite enrichment and profiling. *Nat. Methods 4*, 429–35.

(208) Keatinge-Clay, A. T., and Stroud, R. M. (2006) The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure 14*, 737–48.

(209) Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J., and Leadlay, P. F. (2005) Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. *Chem. Biol. 12*, 1145–1153.

(210) Magarvey, N. A., Beck, Z. Q., Golakoti, T., Ding, Y., Huber, U., Hemscheidt, T. K., Abelson, D., Moore, R. E., and Sherman, D. H. (2006) Biosynthetic Characterization and Chemoenzymatic Assembly of the Cryptophycins. Potent Anticancer Agents from Nostoc Cyanobionts. *ACS Chem. Biol. 1*, 766–779.

(211) Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., and Gerwick, W. H. (2012) Moorea producens gen. nov., sp. nov. and Moorea bouillonii comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int. J. Syst. Evol. Microbiol. 62*, 1171–1178.

(212) Castro-Falcón, G., Millán-Aguiñaga, N., Roullier, C., Jensen, P. R., and Hughes, C. C. (2018) Nitrosopyridine Probe To Detect Polyketide Natural Products with Conjugated Alkenes: Discovery of Novodaryamide and Nocarditriene. *ACS Chem. Biol. 13*, 3097–3106.

(213) Castro-Falcón, G., Hahn, D., Reimer, D., and Hughes, C. C. (2016) Thiol Probes To Detect Electrophilic Natural Products Based on Their Mechanism of Action. *ACS Chem. Biol. 11*, 2328–36.

(214) Sivakumar, K., Xie, F., Cash, B. M., Long, S., Barnhill, H. N., and Wang, Q. (2004) A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Org. Lett. 6*, 4603–4606.

(215) Ross, C., Scherlach, K., Kloss, F., and Hertweck, C. (2014) The molecular basis of conjugated polyyne biosynthesis in phytopathogenic bacteria. *Angew. Chemie - Int. Ed. 53*, 7794–7798.

(216) Zhu, X., Shieh, P., Su, M., Bertozzi, C. R., and Zhang, W. (2016) A fluorogenic screening platform enables directed evolution of an alkyne biosynthetic tool. *Chem. Commun. 52*, 11239–11242.

(217) Yang, L., Chumsae, C., Kaplan, J. B., Moulton, K. R., Wang, D., Lee, D. H., and Zhou, Z. S. (2017) Detection of Alkynes via Click Chemistry with a Brominated Coumarin Azide by Simultaneous Fluorescence and Isotopic Signatures in Mass Spectrometry. *Bioconjug. Chem. 28*, 2302–2309.

(218) Jones, A. C., Monroe, E. A., Podell, S., Hess, W. R., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R. S., Yates, J. R., Reinhardt, R., Kube, M., Burkart, M. D., Allen, E. E., Dorrestein, P. C., Gerwick, W. H., and Gerwick, L. (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium Lyngbya majuscula. *Proc. Natl. Acad. Sci. 108*, 8815–8820.

(219) Wenzel, S. C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blöcker, H., and Müller, R. (2005) Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: Evidence for an

iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *ChemBioChem 6*, 375–385.

(220) Li, S., Wu, X., Zhang, L., Shen, Y., and Du, L. (2017) Activation of a Cryptic Gene Cluster in Lysobacter enzymogenes Reveals a Module/Domain Portable Mechanism of Nonribosomal Peptide Synthetases in the Biosynthesis of Pyrrolopyrazines. *Org. Lett. 19*, 5010–5013.

(221) Glenn, T. C., Nilsen, R., Kieran, T. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S., Louha, S., Garcia-De-Leon, F. J., Angel del Rio Portilla, M., Reed, K., Anderson, J. L., Meece, J. K., Aggery, S., Rekaya, R., Alabady, M., Belanger, M., Winker, K., and Faircloth, B. C. (2016) Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv (Cold Spring Harb. Labs Journals, 2016)*.

(222) Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc. 10*, 845–858.

(223) Allinger, N. L. (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc. 99*, 8127–8134.

(224) Halgren, T. A. (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem. 17*, 490–519.

(225) Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods 65*, 55–63.

(226) Austin, B., and Allen, D. A. (1982) Microbiology of laboratory-hatched brine shrimp (Artemia). *Aquaculture 26*, 369–383.

(227) Gerwick, W. H., Proteau, P. J., Nagle, D. G., Hamel, E., Blokhin, A., and Slate, D. L. (1994) Structure of Curacin A, a Novel Antimitotic, Antiproliferative and Brine Shrimp Toxic Natural Product from the Marine Cyanobacterium Lyngbya majuscula. *J. Org. Chem. 59*, 1243–1245.

(228) Zhao, J.-C., Yu, S.-M., Liu, Y., and Yao, Z.-J. (2013) Biomimetic Synthesis of ent -(−)-Azonazine and Stereochemical Reassignment of Natural Product. *Org. Lett. 15*, 4300–4303.

(229) Zhao, J., Yu, S., Qiu, H., and Yao, Z. (2014) Total synthesis of ent-(−)-azonazine using a biomimetic direct oxidative cyclization and structural reassignment of natural product. *Tetrahedron 70*, 3197–3210.

(230) Sureshbabu, V. V., and Narendra, N. (2011) Amino Acids, Peptides and Proteins in Organic Chemistry. *Amin. Acids, Pept. Proteins Org. Chem.* (Hughes, A. B., Ed.). Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

(231) Formaggio, F., Baldini, C., Moretto, V., Crisma, M., Kaptein, B., Broxterman, Q. B., and Toniolo, C. (2005) Preferred conformations of peptides containing tert-leucine, a sterically demanding, lipophilic alpha-amino acid with a quaternary side-chain Cbeta atom. *Chemistry 11*, 2395–404.

(232) Bisel, P., Al-Momani, L., and Müller, M. (2008) The tert-butyl group in chemistry and biology. *Org. Biomol. Chem. 6*, 2655.

(233) Jacobson, I. C., Reddy, P. G., Wasserman, Z. R., Hardman, K. D., Covington, M. B., Arner,

E. C., Copeland, R. A., Decicco, C. P., and Magolda, R. L. (1998) Structure-based design and synthesis of a series of hydroxamic acids with a quaternary-hydroxy group in P1 as inhibitors of matrix metalloproteinases. *Bioorg. Med. Chem. Lett. 8*, 837–842.

(234) Örtqvist, P., Peterson, S. D., Åkerblom, E., Gossas, T., Sabnis, Y. A., Fransson, R., Lindeberg, G., Helena Danielson, U., Karlén, A., and Sandström, A. (2007) Phenylglycine as a novel P2 scaffold in hepatitis C virus NS3 protease inhibitors. *Bioorganic Med. Chem. 15*, 1448–1474.

(235) Johansson, P., Bäck, M., Kvarnström, I., Jansson, K., Vrang, L., Hamelink, E., Hallberg, A., Rosenquist, Å., and Samuelsson, B. (2006) Potent inhibitors of the hepatitis C virus NS3 protease: Use of a novel P2 cyclopentane-derived template. *Bioorg. Med. Chem. 14*, 5136–5151.

(236) Talpir, R., Benayahu, Y., Kashman, Y., Pannell, L., and Schleyer, M. (1994) Hemiasterlin and geodiamolide TA; two new cytotoxic peptides from the marine sponge Hemiasterella minor (Kirkpatrick). *Tetrahedron Lett. 35*, 4453–4456.

(237) Li, H. yu, Matsunaga, S., and Fusetani, N. (1995) Halicylindramides A-C, antifungal and cytotoxic depsipeptides from the marine sponge Halichondria cylindrata. *J. Med. Chem. 38*, 338–43.

(238) Crews, P., Farias, J. J., Emrich, R., and Keifer, P. A. (1994) Milnamide A, an unusual cytotoxic tripeptide from the marine sponge Auletta cf. constricta. *J. Org. Chem. 59*, 2932–2934.

(239) Coleman, J. E., Dilip de Silva, E., Kong, F., Andersen, R. J., and Allen, T. M. (1995) Cytotoxic peptides from the marine sponge Cymbastela sp. *Tetrahedron 51*, 10653–10662.

(240) Hamada, T., Matsunaga, S., Yano, G., and Fusetani, N. (2005) Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, Theonella swinhoei. *J. Am. Chem. Soc. 127*, 110–8.

(241) Gulavita, N. K., Wright, A. E., Kelly-Borges, M., Longley, R. E., Yarwood, D., and Sills, M. A. (1994) Eryloside E from an atlantic sponge Erylus goffrilleri. *Tetrahedron Lett. 35*, 4299–4302.

(242) Kerr, R. G., Foss, C., Matsunaga, S., and Fusetani, N. (1997) Isolation and Structure Elucidation of Epipolasterol and 22,23-Dihydroepipolasterol from the Marine Sponge Epipolasis sp. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 117*, 561–563.

(243) Dai, J., Sorribas, A., Yoshida, W. Y., Kelly, M., and Williams, P. G. (2010) Topsentinols, 24-isopropyl steroids from the marine sponge Topsentia sp. *J. Nat. Prod. 73*, 1597–600.

(244) Sakabe, N., Takada, S., and Okabe, K. (1967) The structure of ginkgolide A, a novel diterpenoid trilactone. *Chem. Commun.* 259–261.

(245) Nakanishi, K., Habaguchi, K., Nakadaira, Y., Woods, M. C., Maruyama, M., Major, R. T., Alauddin, M., Patel, A. R., Weinges, K., and Baehr, W. (1971) Structure of bilobalide, a rare tert-butyl containing sesquiterpenoid related to the C20-ginkgolides. *J. Am. Chem. Soc. 93*, 3544–3546.

(246) Kotake, C., Yamasaki, T., Moriyama, T., Shinoda, M., Komiyama, N., Furumai, T., Konishi, M., and Oki, T. (1992) Butyrolactols A and B, new antifungal antibiotics. Taxonomy, isolation, physico-chemical properties, structure and biological activity. *J. Antibiot. (Tokyo). 45*, 1442–50.

(247) Mujumdar, R. B., Rao, A. V. R., Rathi, S. S., and Venkataraman, K. (1975) Swietenone, the

first natural t-butyl ketone, from Chloroxylon swietenia. *Tetrahedron Lett. 11*, 867–868.

(248) Harborne, J. B. (1999) Classes and functions of secondary products from plants, in *Chemicals from Plants*, pp 1–25. World Scientific/Imperial College Press.

(249) Wu, T. S. (1987) Alkaloids and coumarins of Skimmia reevesiana. *Phytochemistry 26*, 873–875.

(250) Sudek, S., Lopanik, N. B., Waggoner, L. E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D. H., and Haygood, M. G. (2007) Identification of the putative bryostatin polyketide synthase gene cluster from "Candidatus *Endobugula sertula*", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod. 70*, 67–74.

(251) Kamano, Y., Zhang, H., Morita, H., Itokawa, H., Shirota, O., Pettit, G. R., Herald, D. L., and Herald, C. L. (1996) Conformational analysis of a marine antineoplastic macrolide, bryostatin 10. *Tetrahedron 52*, 2369–2376.

(252) Vogensen, S. B., Strømgaard, K., Shindou, H., Jaracz, S., Suehiro, M., Ishii, S., Shimizu, T., and Nakanishi, K. (2003) Preparation of 7-substituted ginkgolide derivatives: potent platelet activating factor (PAF) receptor antagonists. *J. Med. Chem. 46*, 601–8.

(253) Braquet, P., Esanu, A., Buisine, E., Hosford, D., Broquet, C., and Koltai, M. (1991) Recent progress in ginkgolide research. *Med. Res. Rev. 11*, 295–355.

(254) Orvieto, F., Koch, U., Matassa, V. G., and Muraglia, E. (2003) Novel, potent phenethylamide inhibitors of the hepatitis C virus (HCV) NS3 protease: Probing the role of P2 aryloxyprolines with hybrid structures. *Bioorganic Med. Chem. Lett. 13*, 2745–2748.

(255) Pereira, A. R., Cao, Z., Engene, N., Soria-Mercado, I. E., Murray, T. F., and Gerwick, W. H. (2010) Palmyrolide A, an unusually stabilized neuroactive macrolide from palmyra atoll cyanobacteria. *Org. Lett. 12*, 4490–4493.

(256) Orjala, J., Nagle, D. G., Hsu, V. L., and Gerwick, W. H. (1995) Antillatoxin: An Exceptionally Ichthyotoxic Cyclic Lipopeptide from the Tropical Cyanobacterium Lyngbya majuscula. *J. Am. Chem. Soc. 117*, 8281–8282.

(257) Nogle, L. M., Okino, T., and Gerwick, W. H. (2001) Antillatoxin B, a neurotoxic lipopeptide from the marine cyanobacterium Lyngbya majuscula. *J. Nat. Prod. 64*, 983–985.

(258) Klein, D., Braekman, J. C., Daloze, D., Hoffmann, L., Castillo, G., and Demoulin, V. (1999) Madangolide and laingolide A, two novel macrolides from Lyngbya bouillonii (Cyanobacteria). *J. Nat. Prod. 62*, 934–936.

(259) Shao, C.-L., Mou, X.-F., Cao, F., Spadafora, C., Glukhov, E., Gerwick, L., Wang, C.-Y., and Gerwick, W. H. (2018) Bastimolide B, an Antimalarial 24-Membered Marine Macrolide Possessing a tert -Butyl Group. *J. Nat. Prod. 81*, 211–215.

(260) Ogawa, H., Iwasaki, A., Sumimoto, S., Kanamori, Y., Ohno, O., Iwatsuki, M., Ishiyama, A., Hokari, R., Otoguro, K., Omura, S., and Suenaga, K. (2016) Janadolide, a Cyclic Polyketide-Peptide Hybrid Possessing a tert-Butyl Group from an Okeania sp. Marine Cyanobacterium. *J. Nat. Prod. 79*, 1862–1866.

(261) Nagle, D. G., Paul, V. J., and Roberts, M. A. (1996) Ypaoamide, a new broadly acting feeding deterrent from the marine cyanobacterium Lyngbya majuscula. *Tetrahedron Lett. 37*, 6263–6266.

(262) Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J., and Corbett, T. H. (2001) Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium Lyngbya majuscula. *J. Am. Chem. Soc. 123*, 5418–5423.

(263) Thornburg, C. C., Cowley, E. S., Sikorska, J., Shaala, L. A., Ishmael, J. E., Youssef, D. T. A., and McPhail, K. L. (2013) Apratoxin H and Apratoxin A Sulfoxide from the Red Sea Cyanobacterium Moorea producens. *J. Nat. Prod. 76*, 1781–1788.

(264) Luesch, H. (2002) New apratoxins of marine cyanobacterial origin from guam and palau. *Bioorg. Med. Chem. 10*, 1973–1978.

(265) Wu, P., Cai, W., Chen, Q.-Y., Xu, S., Yin, R., Li, Y., Zhang, W., and Luesch, H. (2016) Total Synthesis and Biological Evaluation of Apratoxin E and Its C30 Epimer: Configurational Reassignment of the Natural Product. *Org. Lett. 18*, 5400–5403.

(266) Tidgewell, K., Engene, N., Byrum, T., Media, J., Doi, T., Valeriote, F. A., and Gerwick, W. H. (2010) Evolved Diversification of a Modular Natural Product Pathway: Apratoxins F and G, Two Cytotoxic Cyclic Depsipeptides from a Palmyra Collection of Lyngbya bouillonii. *ChemBioChem 11*, 1458–1466.

(267) Gutiérrez, M., Suyama, T. L., Engene, N., Wingerd, J. S., Matainaho, T., and Gerwick, W. H. (2008) Apratoxin D, a potent cytotoxic cyclodepsipeptide from papua new Guinea collections of the marine cyanobacteria Lyngbya majuscula and Lyngbya sordida. *J. Nat. Prod. 71*, 1099–1103.

(268) Paatero, A. O., Kellosalo, J., Dunyak, B. M., Almaliti, J., Gestwicki, J. E., Gerwick, W. H., Taunton, J., and Paavilainen, V. O. (2016) Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. *Cell Chem. Biol. 23*, 561–566.

(269) Huang, K.-C., Chen, Z., Jiang, Y., Akare, S., Kolber-Simonds, D., Condon, K., Agoulnik, S., Tendyke, K., Shen, Y., Wu, K.-M., Mathieu, S., Choi, H., Zhu, X., Shimizu, H., Kotake, Y., Gerwick, W. H., Uenaka, T., Woodall-Jappe, M., and Nomoto, K. (2016) Apratoxin A Shows Novel Pancreas-Targeting Activity through the Binding of Sec 61. *Mol. Cancer Ther. 15*, 1208–16.

(270) Freeman, M. F., Gurgui, C., Helf, M. J., Morinaka, B. I., Uria, A. R., Oldham, N. J., Sahl, H.-G., Matsunaga, S., and Piel, J. (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science (80-.). 338*, 387–390.

(271) Parent, A., Guillot, A., Benjdia, A., Chartier, G., Leprince, J., and Berteau, O. (2016) The B 12 -Radical SAM Enzyme PoyC Catalyzes Valine C β -Methylation during Polytheonamide Biosynthesis. *J. Am. Chem. Soc. 138*, 15515–15518.

(272) Benjdia, A., Balty, C., and Berteau, O. (2017) Radical SAM Enzymes in the Biosynthesis of Ribosomally Synthesized and Post-translationally Modified Peptides (RiPPs). *Front. Chem. 5*, 87.

(273) Li, X., and Djerassi, C. (1983) Minor and trace sterols in marine invertebrates 40. Structure and synthesis of axinyssasterol, 25-methylfucosterol and 24-ethyl-24-methylcholesterol -- novel sponge sterols with highly branched side chains. *Tetrahedron Lett. 24*, 665–668.

(274) Giner, J. L., and Djerassi, C. (1991) Biosynthesis of 24-methylene-25-methylcholesterol in Phaseolus vulgaris. *Phytochemistry 30*, 811–814.

(275) Young, J., Stevens, D. C., Carmichael, R., Tan, J., Rachid, S., Boddy, C. N., Müller, R., and Taylor, R. E. (2013) Elucidation of gephyronic acid biosynthetic pathway revealed unexpected sam-dependent methylations. *J. Nat. Prod. 76*, 2269–2276.

(276) Skiba, M. A., Sikkema, A. P., Moss, N. A., Tran, C. L., Sturgis, R. M., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2017) A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit. *ACS Chem. Biol. 12*, 3039–3048.

(277) Bachmann, B. O., Van Lanen, S. G., and Baltz, R. H. (2014) Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J. Ind. Microbiol. Biotechnol. 41*, 175–184.

(278) Lino, C. A., Harper, J. C., Carney, J. P., and Timlin, J. A. (2018) Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv. 25*, 1234–1257.

(279) Carlson, R. (2018) Competition and the Future of Reading and Writing DNA, in *Synthetic Biology: Parts, Devices and Applications* (Smolke, C., Ed.) Volume 8., pp 3–11. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

(280) Gruene, T., Wennmacher, J. T. C., Zaubitzer, C., Holstein, J. J., Heidler, J., Fecteau-Lefebvre, A., De Carlo, S., Müller, E., Goldie, K. N., Regeni, I., Li, T., Santiso-Quinones, G., Steinfeld, G., Handschin, S., van Genderen, E., van Bokhoven, J. A., Clever, G. H., and Pantelic, R. (2018) Rapid Structure Determination of Microcrystalline Molecular Compounds Using Electron Diffraction. *Angew. Chemie Int. Ed. 57*, 16313–16317.

(281) Jain, S. H. (2017) The health care innovation bubble. *Healthcare*.

(282) Kurita, K. L., Glassey, E., and Linington, R. G. (2015) Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc. Natl. Acad. Sci. U. S. A. 112*, 11999–2004.

(283) Siqueira-Neto, J. L., Moon, S., Jang, J., Yang, G., Lee, C., Moon, H. K., Chatelain, E., Genovesio, A., Cechetto, J., and Freitas-Junior, L. H. (2012) An Image-Based High-Content Screening Assay for Compounds Targeting Intracellular Leishmania donovani Amastigotes in Human Macrophages. *PLoS Negl. Trop. Dis. 6*, e1671.