Phenotypic Screening and Genome Mining for the Discovery of Natural Products from

the Genus *Nostoc*

By:

Daniel Steven May

B.S., Alma College 2013

DISSERTATION

Submitted as partial fulfillment of the

requirements for the degree of Doctor of

Philosophy in Pharmacognosy in the Graduate

College of the University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Dr. Jimmy Orjala, Chair and Advisor

Dr. Joanna Burdette

Dr. Brian Murphy

Dr. Alessandra S. Eustaquio

Dr. Leslie Aldrich, Department of Chemistry

ACKNOWLEDGEMENTS

I first want to thank my PhD advisor Dr. Jimmy Orjala for his patience, guidance, and support during my time at UIC. I am incredibly grateful for all the opportunities that he has provided me from professional connections to the freedom to pursue a variety of different directions in my research. I greatly appreciate his patience and discussions on my writing and regular discussions on my future as a scientist.

I would also like to thank Dr. Brian Doyle, my research advisor at Alma College, who first introduced me to the exciting field of natural products and allowed me to begin my research as a natural product chemist in his lab.

I would like to thank the members of my preliminary committee, Dr. Shura Mankin and Dr. Jeremy Johnson for their guidance and support early in my graduate career. I would also like to thank the members of my dissertation committee. Dr. Joanna Burdette for her mentorship and help in my development as a researcher. Dr. Brian Murphy for his good-natured humor and helpful insight into many of my projects. Dr. Alessandra Eustaquio for her support in building my enthusiasm for biosynthesis and the opportunities to present my work that she has provided. Dr. Leslie Aldrich for insightful conversations about my future as an academic professional.

I would like to thank the members of the Orjala Lab for their friendship and support over the last five years. Dr. Shangwen Luo helped me get acclimated to the lab and proceeded to teach me everything she knew about cyanobacteria and structure elucidation before finishing her own dissertation. I appreciate the time that she took to

answer my questions and her patience with me as a young graduate student. Dr. George Chlipala was a wealth of knowledge on nearly every topic that came up in the lab and I am thankful for his experience and insightful discussions we had. The help provided by Dr. Aleksej Krunic in acquiring and teaching me about NMR experiments greatly aided in my dissertation work. Camila Crnkovic has been a great friend and lab bench-mate as well as my favorite mass spectrometry collaborator. Conversations with her about my projects and posters have always provided improvements and creative insight. Peter Sullivan and I have had countless wonderful conversations on cyanobacteria and the world at large which have provided a nice respite from days of grueling data analysis. I am grateful to the other members of the lab, both past and present for their friendship and help over the last five years, Rojin Ahadi, Mark Sadek, Kyle Mathes, and Sean Romanowski.

I would like to thank the many other friends that I have made at UIC: Brian Guo, Maryam Elfeki, Skylar Carlson, Michael Mullowney, Rhea Bovee, Vanessa Nepomuceno, Sezen Meydan, James Marks, Sofia Costa, Chase Clark Sylvia Kunakom, Wei-Lun Chen, Tom Speltz, Daniel Nosal, and many others for making it easy to switch from a tiny town in Michigan to a huge city.

I would like to thank my family for their support and understanding when I decided to move to Chicago to pursue a PhD in a field that none of them had ever heard about. The conversations and visits with them have always helped to keep me grounded and their support means so much to me. I especially want to thank Sarah Proudfoot whose emotional support and friendship carried me through numerous challenging times during my graduate career.

ACKNOWLEDGEMENTS (continued)

I would like to thank the faculty and staff in the department of medicinal chemistry and pharmacognosy for their help in teaching me and supporting me through my graduate career. The lectures, access to equipment, administrative help, and support from the department was integral to my success as a graduate student.

Thank you.

DSM

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF ABBREVIATIONS

BLAST   Basic Local Alignment Search Tool

AntiSMASH   Antibiotic and Secondary Metabolite Analysis Shell

NRP   Nonribosomal Peptide

NRPS   Nonribosomal Peptide Synthetase

PK   Polyketide

PKS   Polyketide Synthase

RiPPs   Ribosomally Synthesized & Post-Translationally Modified Peptide

A domain   Adenylation Domain

PCP domain   Peptide Carrier Protein Domain

C domain   Condensation Domain

TE domain   Thioesterase Domain

R domain   Reductase Domain

E domain   Epimerase Domain

Ox domain   Oxidase Domain

Cl domain   Halogenase Domain

Fm domain   Formylation Domain

Cyc domain   Cyclase Domain

MT domain   Methyl Transferase Domain

AT domain   Acyl Transferase Domain

ACP domain   Acyl Carrier Protein Domain

KS domain   Keto Synthase Domain

KR domain   Keto Reductase Domain

DH domain   Dehydratase Domain

ER domain   Enoyl Reductase Domain

CoA   Coenzyme A

rRNA   Ribosomal Ribonucleic Acid

RuBisCO   Ribulose-1,5-bisphosphate carboxylase/oxygenase

MMAE   Mono-Methyl Auristatin E

LIST OF ABBREVIATIONS (continued)

MAB	Monoclonal Antibody
NCBI	National Center for Biotechnology Information
NIES	National Institute for Environmental Studies
PCC	Pasteur Culture Collection of Cyanobacteria
CENA	Centro de Energia Nuclear na Agricultura
ATCC	American Type Culture Collection
UTEX	University of Texas Culture Collection
Choi	2-carboxy-6-hydroxyoctahydroindole
NF-KB	Nuclear Factor Kappa-light-chain-enhancer of Activated B Cells
Ahp	3-amino-6-hydroxy-2-piperidone
PP	Protein Phosphatase
Hmp	3-hydroxy-4-methylproline
OATP	Organic-Anionic-Transporting-Polypeptide
Ahoa	3-amino-2,5-dihydroxy-8-phenyloctanoic acid
Hph	Homophenylalanine
Hty	Homotyrosine
PNV	5-Phenylnorvaline
PNL	6-Phenylnorleucine
PCR	Polymerase Chain Reaction
ANOVA	Analysis of Variants
DCM	Dichloromethane
MeOH	Methanol
IPA	Isopropyl Alcohol
DMSO	Dimethyl Sulfoxide
EtoAc	Ethyl Acetate
ACN	Acetonitrile
HPLC	High Performance Liquid Chromatography
UPLC	Ultra High Performance Liquid Chromatography
IC_{50}	Half Maximal Inhibitory Concentration
MIC	Minimal Inhibitory Concentration

LIST OF ABBREVIATIONS (continued)

COSY	Correlation Spectroscopy
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
MHBC	Heteronuclear Multiple Bond Correlation Spectroscopy
ROESY	Rotating-frame Overhauser Effect Spectroscopy
DEPTQ	Distorsionless Enhancement by Polarization Transfer Quaternary
ECD	Electronic Circular Dichroism
BGC	Biosynthetic Gene Cluster
IT-ToF	Ion Trap Time-of-Flight
Q-ToF	Quadrupole Time-of-Flight
MALDI-ToF	Matrix Assisted Laser Desorption Ionization Time-of-Flight
HRESIMS	High Resolution Electrospray Ionization Mass Spectrometry
EDTA	Ethylenediaminetetraacetic Acid
SDS	Sodium dodecyl sulfate
OD_{600}	Optical Density at 600 nm
SMURF	Secondary Metabolite Unique Regions Finder
NaPDoS	Natural Product Domain Seeker
FDLA	1-fluoro-2-4-dinitrophenyl-5-L-leucine amide
ORF	Open Reading Frame

SUMMARY

Cyanobacteria are a good source of natural products and the genus *Nostoc* has been shown to be an especially prolific producer of biologically active metabolites. The UIC cyanobacteria culture collection contains over one thousand cyanobacterial strains that have mostly been isolated from visible macroscopic colonies. A cyanobacterial specific primer set was developed to assess the taxonomic diversity of cyanobacteria in these macroscopic collections and environmental samples with no visible cyanobacterial growth. Amplicon sequencing with the cyanobacterial specific primer set allowed identification of low abundance cyanobacteria in soil samples. Incubation of soil samples in cyanobacterial specific media with no nitrate led to the growth of nitrogen fixing cyanobacteria, including *Nostoc* spp., in sufficient amounts to isolate strains for the UIC cyanobacteria culture collection. Extracts of Nostoc spp. from the culture collection were assessed for cytotoxicity against a panel of cancer cell lines. This phenotypic screening approach led to the identification of a series of cytotoxic [7,7] paracyclophanes from six *Nostoc* strains. Two new merocyclophanes and five ribocyclophanes, a new class of [7,7] paracyclophanes, were among the [7,7] paracyclophanes identified. The structures were solved by HRESIMS, 1D/2D NMR, X-ray crystallography and ECD spectroscopy. A putative biosynthetic gene cluster (BGC) was identified for the merocyclophanes and a biosynthetic mechanism was proposed that explains the differences in the structures of merocyclophanes and other [7,7] paracyclophanes. To perform a more comprehensive survey of *Nostoc* natural products, a genome mining approach was used to identify BGCs from a cultured Nostoc sp. from the UIC cyanobacteria culture collection. Genome mining and bioinformatic

SUMMARY (continued)

predictions combined with isotopic labeling and comparative metabolomics allowed

identification of several compounds produced by a single *Nostoc* sp. isolated from a soil

collection. The isotopic labeling aided in identification of the compounds in an extract

as well as the structure elucidation. Two of the identified compounds were new and the

structures were elucidated by HRESIMS, 1D/2D NMR, MS/MS, Marfey's analysis, and

bioinformatic predictions. Utilizing both phenotypic screening and genome mining

approaches allowed a more comprehensive study of the natural products produced by

cultured *Nostoc* spp. from the UIC cyanobacteria culture collection.

1 INTRODUCTION

Natural Products

Natural Products, or secondary metabolites, are compounds produced by an organism to provide the organism with a competitive advantage. These metabolites can be metal chelators, quorum sensing molecules, or antibiotics; all of which provide the producing organism with an edge over other organisms in their environment. These compounds are often structurally complex, with multiple chiral centers, and have been optimized by natural selection over millions of years to interact with specific molecular targets.[1] For these reasons natural products have been used as medicines since antiquity. Written records from ancient cultures around the world describe using natural products for various ailments. For example, Opium from the opium poppy and willow tree bark are both natural products that have been used by humans as analgesics.[2]

The discovery of penicillin, and its use during the second world war to combat infection, revealed the immense potential of natural products as commercial pharmaceuticals.[3,4] The use of natural products quickly expanded from antibiotics to anticancer drugs, antifungal drugs, immunosuppressants, and more.[5–7] Today, natural products are well represented among FDA approved drugs. From 1981 to 2014, 33% of all small molecule approved drugs were natural products or natural product derivatives. If synthetic drugs with a pharmacophore or mechanism of action inspired by natural products are included, the percentage increases to 68% of all FDA approved small molecule drugs from 1981 to 2014 (Figure 1).[8] Natural products are chemically diverse and will continue to be used for drug discovery either as inspiration for synthetic compounds or as pharmaceuticals themselves.[9–11]

Small Molecule Approved Drugs from 1981-2014
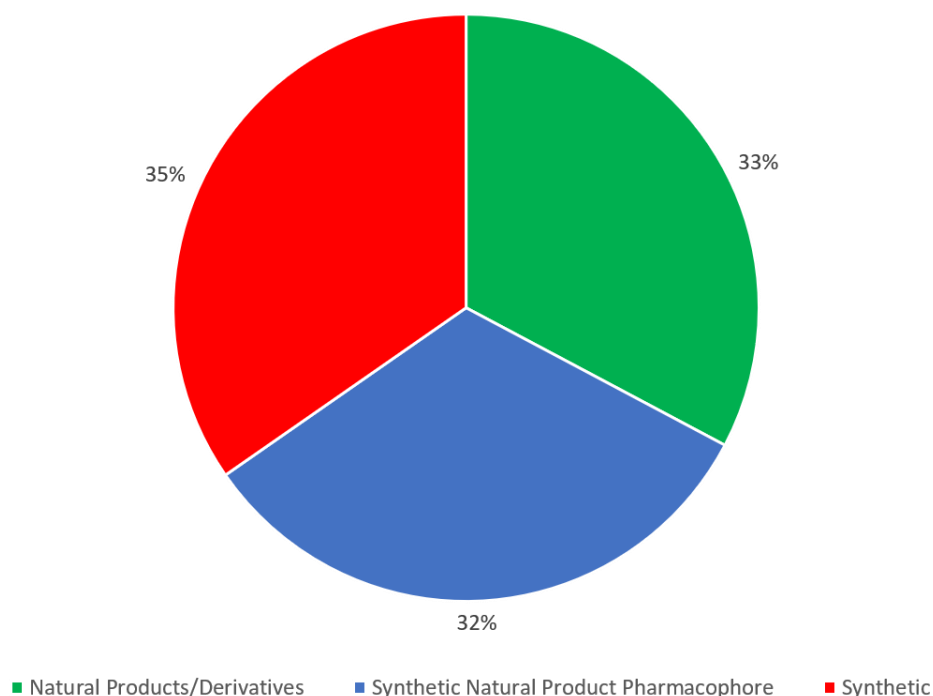
Figure 1: Small molecule approved drugs from 1981-2014, adapted from Newman and Cragg J Nat Prod 2016

Natural Product Discovery Platforms

Two widely used natural product discovery platforms are phenotypic screening and genome mining.[12] Phenotypic screening approaches have been the traditional route for natural product discovery. This approach takes advantage of the biological and chemical diversity of organisms to identify compounds that elicit a desired biological response. This often involves the creation of large libraries of organisms, extracts, and pure compounds that can be screened in a high-throughput manner against assays for any chosen disease state. This paradigm has proved fruitful; identifying many of the

natural products currently used as pharmaceuticals, and was the basis of the Waksman platform of antibiotic discovery.[13,14] The longstanding use of this paradigm however has led to a high rate of rediscovery of known compounds and has necessitated the development of many dereplication strategies. Dereplication aims to quickly identify whether a compound of interest is known or is new, in order to focus materials and efforts on new compounds. Most dereplication protocols involve a sensitive tandem analytical technique such as liquid chromatography-mass spectrometry or liquid chromatography-UV spectroscopy. These types of protocols utilize a chromatographic step to separate individual compounds by their polarity and then characterize them by some intrinsic property such as the molecular mass or chromophore, which can then be searched in databases. Newer techniques include mass spectrometry based molecular networking. This method uses similarities in molecular fragmentation patterns to group similar compounds based on an intrinsically populated database or a crowd-sourced type database like GNPS.[15–19]

Genome mining approaches have becoming increasingly popular with the surge of publicly available genomic data.[10] These approaches use bioinformatic techniques to identify the genes that encode biosynthetic enzymes responsible for producing secondary metabolites. The predicted secondary metabolites can then be identified through a variety of analytical techniques. In bacteria and fungi, genome mining approaches are aided by the existence of biosynthetic gene clusters (BGCs). In these organisms, the genes responsible for producing a secondary metabolite tend to be located together in a cluster in one locus within the genome. Using bioinformatic techniques such as BLAST and libraries of conserved biosynthetic enzyme motifs,

BGCs can be identified and even annotated.[20] One of the more popular bioinformatic suites used for identification of BGCs is AntiSMASH which uses BLAST and hidden markov models to screen genomic data for matches or partial matches against a library of described BGCs.[20,21] Genome mining approaches have the benefits of identifying novel compounds as well as matching compounds to their BGCs. However, it provides little information on biological effect of the identified compound. As full genome sequencing becomes more accessible and affordable, genome mining approaches will likely become more widely used, and should lead to the discovery of more new natural products.[12,22,23]

Biosynthesis of Natural Products

Natural products are produced by a diverse set of biosynthetic pathways which explains the observed chemical diversity. Natural products can be organized by their biosynthetic origin as described by the Genomic Standards Consortium. These include Non-Ribosomal Peptides (NRP), Polyketides (PK), NRP-PK hybrids, Ribosomally synthesized and Post-translationally modified Peptides (RiPPs), terpenes, shikimate derived compounds, alkaloids, and others that do not fit neatly into any of the above categories.[24]

Nonribosomal peptide synthetases (NRPSs)
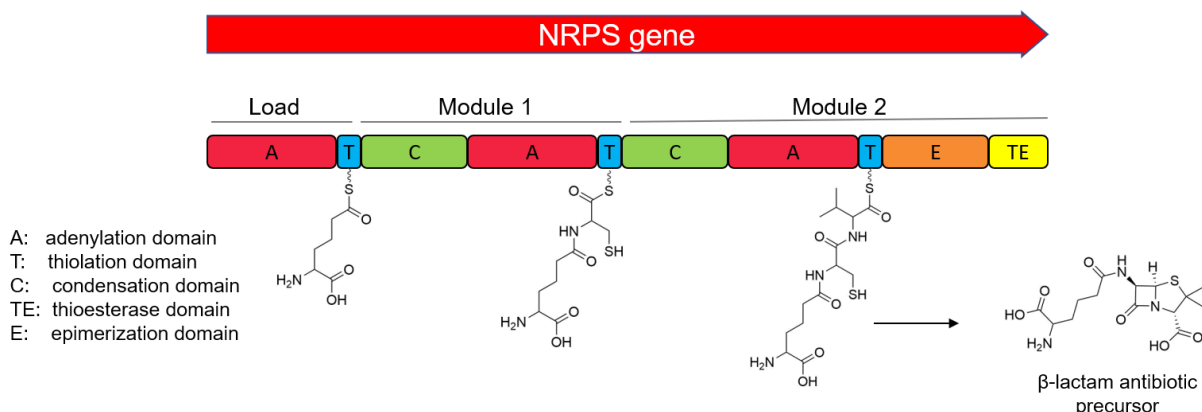
Figure 2: Representation of Nonribosomal Peptide Synthetase biosynthetic gene cluster and biosynthetic mechanism

Non-Ribosomal Peptides represent a group of structurally diverse peptides natural products. These peptides are not produced by the ribosome; rather, they are produced by multi-domain, modular enzyme complexes called Non-Ribosomal Peptide Synthetases (NRPS). The domains of NRPS are grouped into modules. Each module is responsible for the incorporation of one amino acid and contains individual domains that are responsible for the enzymatic steps in the incorporation and processing of the amino acids. Each module is made up of at least three domains: an adenylation domain, a peptide carrier protein domain, and a condensation domain (Figure 2).[25,26,27] Adenylation domains (A) are responsible for selecting and activating amino acids, requiring each adenylation domain to have a selective substrate specificity for one amino acid. Adenylation domains are not limited to the 20 standard ribosomal amino acids and can select for non-standard amino acids, allowing for much greater chemical

diversity in NRPs as opposed to ribosomally produced peptides.[28] Adenylation domains

activate amino acids with ATP, forming an aminoacyl adenylate molecule which is then

loaded onto the sulfide of the phospho-pantetheine unit of the peptide carrier protein.

Peptide carrier proteins (PCP), also known as thiolation domains (T), are covalently

primed with a 4'-phospho-pantetheine unit derived from Acyl-CoA. These PCP domains

are then used to tether the growing compound to the enzyme during each elongation

step. Elongation steps are performed by the Condensation domain (C) which forms the

amide bond between the amine of the downstream phospho-pantetheine tethered

amino acid and the upstream thioester of the growing peptide.[25,26,27] These elongation

steps occur until a Thio-esterase domain (TE) is reached. This domain hydrolyzes the

growing peptide from the predecessor PCP domain and can cyclize the peptide chain to

form cyclic peptides. NRPSs can also be terminated through the use of a reduction

domain (R) which can reduce the thioester to the corresponding aldehyde.[26] NRPSs

can contain a variety of other tailoring domains such as epimerization domains (E),

methyl transferase domains (MT), reduction domains (R), cyclization domains (Cyc),

formylation domains (Fm), halogenation domains (Cl), and oxidation domains (Ox).[29]

These tailoring domains and the incorporation of non-standard amino acids allow NRPS

to produce structurally diverse natural products.

Polyketide Synthase (PKS)

AT: acetyltransferase domain
ACP: acyl carrier protein domain
KS: ketosynthase domain
TE: thioesterase domain
KR: ketoreductase domain
DH: dehydratase domain
ER: enoylreductase domain
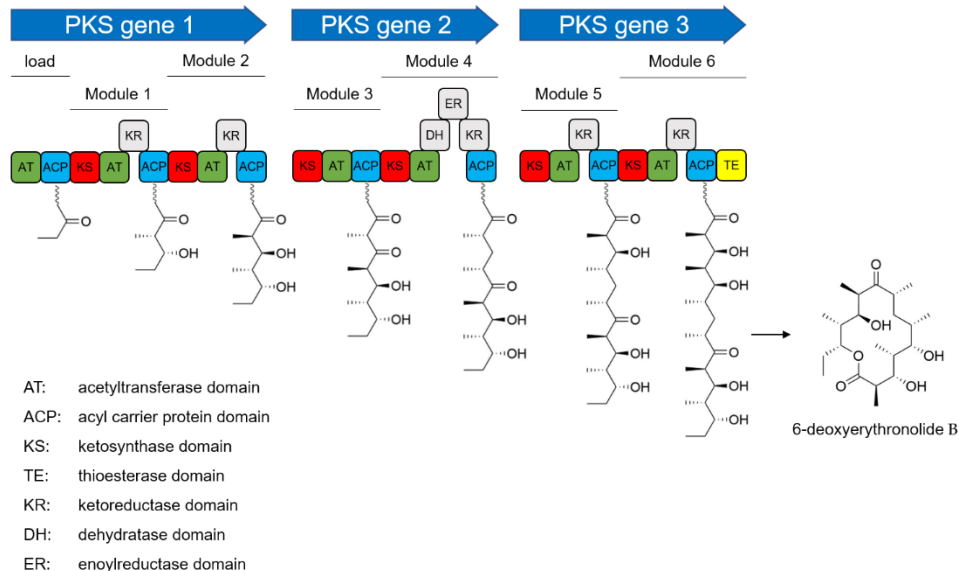
Figure 3: Representation of the three Polyketide Synthase biosynthetic gene cluster types and biosynthetic mechanism of Type I PKS

The production of Polyketides is catalyzed by Polyketide Synthases (PKS). PKS are more diverse than NRPS and can be categorized broadly into three groups: Type I PKS, Type II PKS, and Type III PKS (Figure 3).[31] Type I PKS are the most similar to NRPS with genes that encode large multi-domain, modular enzyme complexes.[30] The modular logic of PKS is similar to NRPS. Each module in a Type I PKS contains at least three domains: An acyl transferase domain, an acyl carrier protein domain, and a keto synthase domain.[30,32] Acyl transferase domains (AT) are responsible for selecting and loading each ketide unit. Similar to the A domains of NRPS, AT domains have a specificity for particular ketide units such as acetate, malonate, or methylmalonate.[33] The ketide units are covalently attached to the sulfide of a 4'-phospho-pantetheine

cofactor on an acyl carrier protein (ACP) domain by the AT domain. The ACP domains

tether the growing compound to the enzyme for each elongation step. The elongation

step is completed by a keto synthase (KS) domain which catalyzes a Claisen

condensation reaction to form a carbon-carbon bond with the loss of carbon dioxide.[32]

The modular elongation steps continue until the compound reaches a Thio-esterase

domain (TE). Similar to NRPS, the TE domain hydrolyzes the compound from the

enzyme, often cyclizing the compound in the process.[30] PKS can also contain tailoring

domains such as ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and

methyl transferase domains (MT) which can provide varying degrees of oxidation or

reduction, resulting in extensive structural diversity.[30]

In general, Type II PKS synthesize polyaromatic compounds, such as

tetracycline, and have a more streamlined gene cluster, containing two genes encoding

two KS proteins, KS_α and KS_β, and an ACP domain. Malonyl-CoA tethered to an ACP

is iteratively acted upon by the KS_α and KS_β heterodimer, with the KS_α performing

decarboxylative condensation and KS_β thought to be involved in controlling the length of

the polyketide chain. [31,34] Although spontaneous aldol reactions can occur to form the

polycyclic structures, cyclase enzymes are often encoded within the gene cluster and

act as chaperones, guiding the formation of the polycyclic rings.[34] Many tailoring

enzymes have also been found to be associated with Type II PKS genes including

glycosyl transferases, prenyl transferases, and various oxidation and reduction

enzymes.[34,35]

Multifunctional Type III PKS also produce aromatic polyketide compounds. Type

III PKS can use acyl CoA substrates directly by utilizing an iteratively acting

homodimeric keto synthase enzyme.[36] The same decarboxylative condensation

reactions lead to chain elongation, and termination typically occurs through either

facilitated or spontaneous aldol reactions, leading to aromatic ring formation.[37]
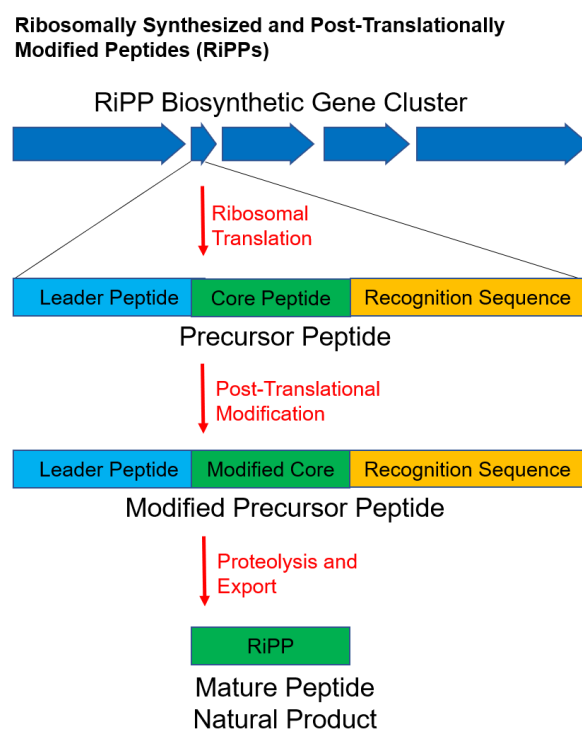
Figure 4: Representation of RiPP biosynthetic mechanism

RiPPs or Ribosomally synthesized and Post-translationally modified Peptides

are, as the name suggests, peptides that are produced by the ribosome, thus their

sequences are directly encoded in the DNA. RiPPs are a diverse group of compounds

that have historically been categorized into many different subcategories. Recently, an attempt was made by van der Donk et al. (2013) to standardize the terminology used to describe and classify RiPPs.[38] Regardless of subclass or producing organism, each RiPP is encoded in a larger peptide called a precursor peptide. The precursor peptide consists of the core peptide, leader peptide, and sometimes a recognition sequence peptide. The core peptide is the amino acid sequence that will be post-translationally modified and become the natural product (Figure 4).[38,39] The leader peptide is important for recognition of post-translational modifying enzymes as well as export from the cell.[40] Post-translational modification of the core peptide can occur both pre-excision from the precursor peptide and post-excision. A core peptide that has been modified, but not yet excised from the precursor peptide, is termed the modified core peptide. Once the modified core peptide has been excised, it is termed the mature peptide. [38,39] Due to the leader peptide acting as the recognition sequence for post-translational modification enzymes, the core peptide can be naturally diverse, leading to a wide diversity of structures.[40] Modifying enzymes can include enzymes for heterocyclization of serine, threonine, and cysteine to form oxazolines and thiazolines, prenylation enzymes, glycosylation enzymes, and enzymes responsible for intramolecular disulfide bond formation.[41]

Sources of Natural Products

Many sources of natural products have been and are still being studied. Organisms such as higher plants and fungi have been well studied due to their ease of access, while other sources of secondary metabolites have only recently become available.[2,42] Marine organisms, such as deep sea sponges, tunicates, and marine

bacteria became important sources of secondary metabolites with the advent of scuba gear and the submersible research vessel.[43,44] Niche environments such as symbiotic bacteria or cave dwelling bacteria are increasingly being studied as sources of unique secondary metabolites that evolved under heavy evolutionary pressure.[45–51] The recent discovery of teixobactin also shows the potential for discovering new secondary metabolites from "uncultured" bacteria.[52] Over the last seventy years, bacteria have been a primary source of natural products.[53] The prokaryotes are present in higher abundances on the planet than any other group of organisms and represent an immense amount of biological diversity. It is estimated that only a small fraction of bacteria can be cultivated in laboratory settings, leaving much more work to be done in discovering new bacterial secondary metabolites.[54,55] However, new compounds are still being found at an impressive rate from bacteria that can be readily cultured.[56]

Cyanobacteria: A Diverse and Ancient Phylum

Three of the most widely studied bacteria for secondary metabolite production are Actinobacteria, Myxobacteria, and Cyanobacteria. All three groups have large genomes with the genetic capacity to produce secondary metabolites.[57–61] Cyanobacteria represents an attractive phylum of bacteria for natural product discovery. They are some of the oldest bacteria on the planet, with fossil evidence, in the form of stromatolites and oncolites, of filamentous cyanobacteria being present 3.5 billion years ago.[62] The great oxygenation event that changed the atmosphere from a reducing atmosphere to an oxidizing atmosphere and killed much of the anaerobic life on earth is thought to be caused primarily by ancient cyanobacteria approximately 2.6 billion years

ago.[62] Over time cyanobacteria have evolved into a vast array of different taxonomic

groups and occupy almost every type of environment on the planet.

Cyanobacterial taxonomy is complex. As one of the oldest phyla of bacteria on

the planet, they are incredibly diverse and have a long evolutionary history which

complicates taxonomic efforts. Like many bacteria, they were originally classified using

microscopic techniques.[63] This led to the creation of many polyphyletic taxonomic

assignments that are still being parsed today. Current taxonomic descriptions are

based on a polyphasic approach that utilizes 16S rRNA analysis, morphological

analysis (including natural variants), ecological niche and distribution, and ultrastructural

studies.[64] Currently, cyanobacteria are split into four taxonomic subclasses: the

Gloeobacteriophycidae, Synechococcophycidae, Oscillatoriophycidae, and

Nostochophycidae.[64] Gloeobacteriophycidae contains a single genus of cyanobacteria,

Gloeobacter, which is thought to be the most primitive cyanobacteria since it lacks

thylakoids, the internal membranes used for photosynthesis.[65] Oscillatoriophycidae and

Synechococcophycidae are both polyphyletic subclasses and contain many genera with

similar morphological structures and 16S rRNA sequences. Both subclasses are

currently undergoing revision.[64] Nostochophycidae contains the orders

Chroococcidiopsidales and Nostocales, the latter is one of the few monophyletic orders

supported by 16S rRNA analysis, and contains the subject genus of this dissertation,

Nostoc.[64,66]

Cyanobacterial Metabolism

A unique metabolism has allowed cyanobacteria to become a widespread and

successful group of bacteria. The ability to perform oxygenic photosynthesis provides

cyanobacteria with the distinct advantage of fixing atmospheric carbon dioxide to create energy rich carbohydrates. The abundance of carbon dioxide, water, and sunlight have allowed cyanobacteria to colonize much of the planet and allow them to thrive wherever these resources are present.[67,68] Cyanobacteria use internal membranes, separate from the cell membrane, called thylakoids to perform photosynthesis.[62] Embedded within the thylakoids are the photosystems II and I as well as phycobilisomes which act as light harvesting complexes. The phycobilisomes are made of phycobiloproteins which contain the pigments that provide cyanobacteria with their blue-green color, phycocyanin and chlorophyll.[69,70] Carbon fixation presents inherent problems, namely keeping carbon dioxide from diffusing out of the cell and concentrating the carbon dioxide near photosynthetic enzymes to make photosynthesis more efficient. The first problem is overcome by utilizing a bicarbonate buffer for carbon dioxide that is transported into the thylakoid membranes. The cyanobacteria sequester the membrane impermeable bicarbonate ion within the thylakoid membrane, concentrating the inorganic carbon near the photosynthetic enzymes.[70] However, bicarbonate is unable to be used directly by the carbon fixation enzymes and is required to be changed back into carbon dioxide to be incorporated into useable sugars. For this, cyanobacteria utilize protein complexes called carboxysomes. These protein complexes house groupings of carbonic anhydrase and RuBisCO, which are responsible for conversion of bicarbonate to carbon dioxide and fixation of carbon dioxide to 3-phosphoglycerate, respectively.[71] By organizing enzymes responsible for carbon fixation and the photosystems together within the cell, and by concentrating inorganic carbon within the

thylakoids with the use of a bicarbonate buffer, cyanobacteria are able to enhance the efficiency of their photosynthesis.

Another significant metabolic process utilized by cyanobacteria is the ability to fix atmospheric nitrogen by utilizing nitrogenase enzymes. Although nitrogen fixation is not unique to cyanobacteria, they are the only known organisms to perform both photosynthesis and nitrogen fixation, making them vital parts of both the carbon and nitrogen cycles.[62] This co-existence of nitrogenase enzymes and photosynthetic enzymes is perplexing as oxygen, the product of photosystem II, inhibits the function of nitrogenases.[72] Cyanobacteria have accordingly evolved methods to utilize both sets of enzymes in one organism.[73–75] One of the more common methods for separating the processes of photosynthesis and nitrogen fixation is to perform them in accordance with a circadian rhythm. Many cyanobacteria perform photosynthesis during the day, when the necessary sunlight is plentiful, and perform nitrogen fixation during the night when the photosynthetic enzymes would be useless due to the lack of sunlight.[75] Other cyanobacteria, specifically members of the Nostocales, utilize cell differentiation to separate the mutually exclusive processes. These differentiated cells, called heterocysts, lack the photosynthetic enzymes of photosystem II and contain specific heterocyst glycolipids within their cell walls to limit the diffusion of oxygen into these specialized cells. Heterocysts then distribute fixed nitrogen to the vegetative cells in the filament and the vegetative cells provide organic carbon sources to the heterocysts.[76,74] The nitrogenase enzymes exist in an enzyme complex that consists of two parts: the dinitrogenase reductase, a smaller homodimer of Fe proteins, and the dinitrogenase, a larger $\alpha_2\beta_2$ tetramer that contains two FeMo-cofactors. The complex utilizes 16

molecules of ATP to catalyze the reaction of one molecule of atmospheric dinitrogen into two molecules of ammonia and one molecule of dihydrogen.[77] Through this process cyanobacteria are thought to perform most of the nitrogen fixation on the planet. Due to the high energy cost of this process, nitrogen fixation only occurs in nitrate limiting conditions, and cyanobacteria will preferentially take up nitrate from their surroundings when it is available.[74,76]

Cyanobacteria as a Source of Natural Products

Cyanobacteria are known to produce a diverse number of natural products.[59,78,79] Freshwater cyanobacteria are best known as producers of cyanotoxins, compounds, that when present in drinking water, cause toxicity to both humans and livestock.[80] Cyanotoxins were first linked to freshwater cyanobacteria in a publication in *Nature* in 1879, which reported the death of livestock that drank from water during a cyanobacterial bloom.[81] Cyanotoxins are a group of compounds with various structures and physiological effects and are produced by a wide range of the cyanobacterial phylum.[82] Anatoxin-a (**1**), saxitoxin (**2**), cylindrospermopsin (**3**), microcystin LR (**4**), and nodularin R (**5**) are the most widely identified cyanotoxins (Figure 5).
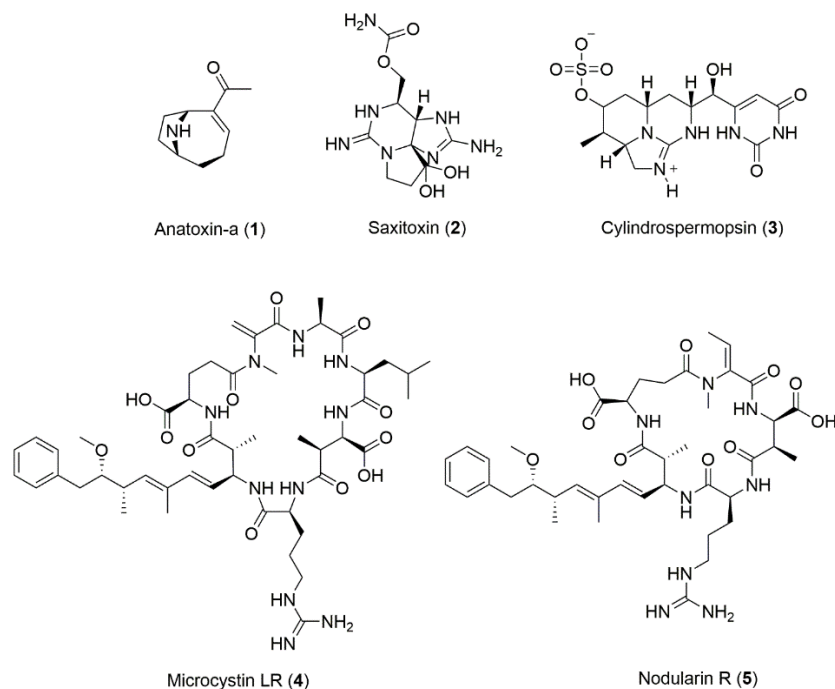
Figure 5: Representative group of cyanotoxins produced by cyanobacteria

Anatoxin-a represents one of the best known neurotoxic cyanotoxins. Anatoxin-a is a

pyrrolidine alkaloid associated with planktonic *Anabaena* species, such as the frequent

harmful algal bloom forming *Anabaena flos-aquae.* The neurotoxin acts on post-

synaptic nicotinic acetylcholine receptors in the central nervous system which leads to a

loss in neuromuscular transmission.[80,82] Microcystins (**4**) and nodularins (**5**) are

NRPS/PKS produced peptides containing non-proteinogenic amino acids and act as

potent liver toxins. Microcystins are most often associated with the cyanobacterium

Microcystis aeruginosa and are often found contaminating drinking water, such as the

regular cyanobacterial blooms that occur in Lake Erie near Toledo, Ohio.[83] Both

microcystins and nodularins are known to inhibit protein phosphatase 1 and 2A which

leads to increased phosphorylation of proteins within liver cells and eventually leads to apoptosis and hepatotoxicity. They are also classified as tumor promoters due to the inhibition of protein phosphatase 1 and 2A which help regulate cell growth and differentiation.[80,82] The wide variety of structures and activities of cyanotoxins provides evidence that cyanobacteria are highly capable producers of biologically active secondary metabolites.

In addition to the cyanotoxins, many diverse secondary metabolites have been reported from cyanobacteria.   A recent publication by Shih et al. demonstrated the extensive genetic capacity of cyanobacteria to produce secondary metabolites by identifying characteristic biosynthetic genes for each of the previously described biosynthetic pathways (Figure 6).[84] The identification of NRPS, PKS, and genes for a variety of RiPPs across the full phylum of cyanobacteria, reveals the capacity of these bacteria to produce a large number of secondary metabolites. This has also been shown by a variety of research groups through the isolation of an increasingly diverse number of structures from cyanobacteria.[56,78,84]
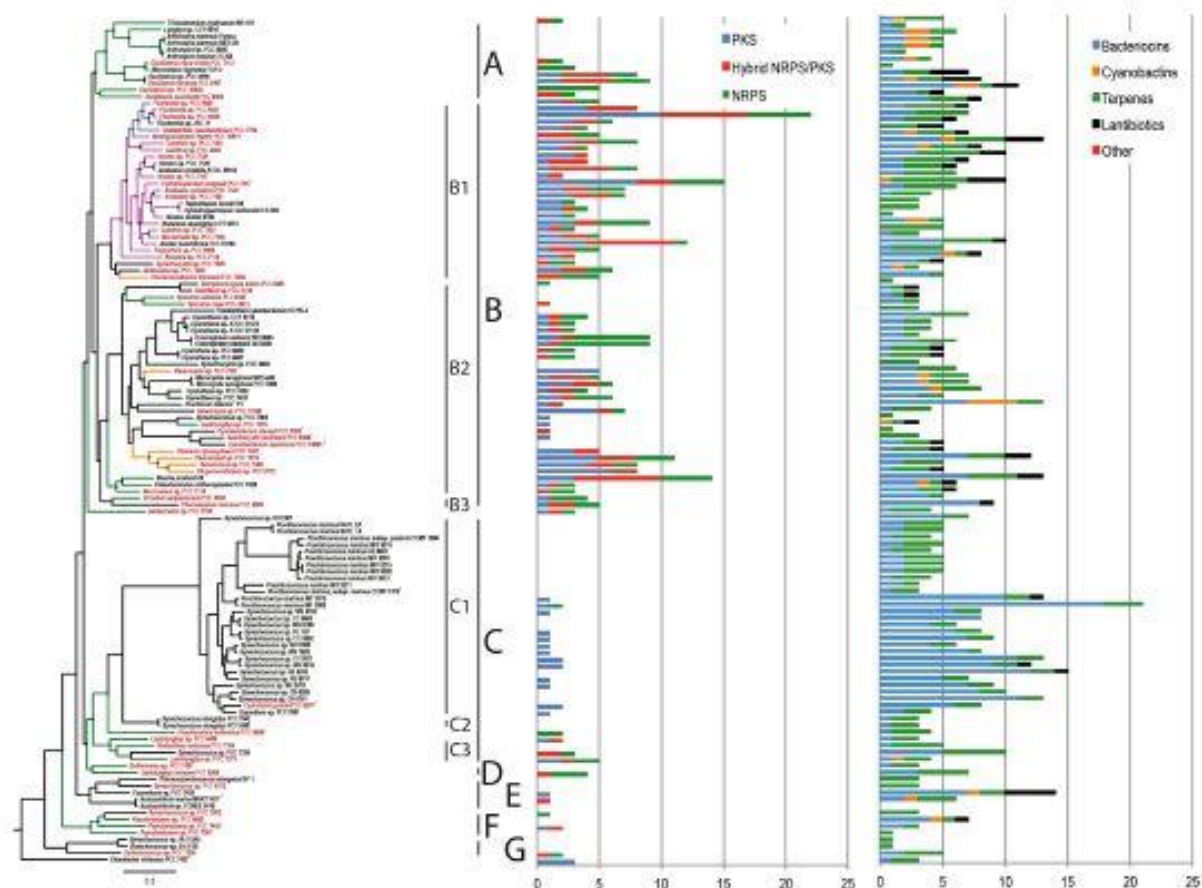
Figure 6: Portrayal of the biosynthetic capacity of cyanobacteria across the full phylum using the Castenholz taxonomic system described in Bergey's Manual of Systematic Bacteriology, adapted from Shih et al. [83] with permission from PNAS.

One compound based on a cyanobacterial natural product has been approved by the FDA as an anticancer drug. A synthetic analog of the cyanobacterial compound dolastatin 10 (**6**), is currently on the market for the treatment of relapsed Hodgkins lymphoma and anaplastic large cell lymphoma. This represents the culmination of decades of work to identify the bioactive compound, optimize the structure of the synthetic compound, and tether it to a monoclonal antibody to produce an effective drug.[85] Dolastatin 10 was originally isolated from a sea hare, *Dolabella auricularia*, and at the time was one of the most potent anticancer natural products identified, with an IC_{50} concentration in the picomolar range against cultured cancer cells.[86] Later it was found that the producing organism was a cyanobacteria, *Symploca*, which was a part of the sea hare's diet.[87] Due to the difficulty of isolating sufficient amounts of the compound from either the sea hare or the cyanobacterium, many attempts were made to synthesize the compound and improve on the structure. Clinical trials were performed for the treatment of a variety of cancers, but Dolastatin 10 and its synthetic analogs proved too toxic to act as a single administered drug.[88,89] However, an analog, monomethyl auristatin E (**7**) found success as an antibody drug conjugate (Figure 7). This formulation was submitted to clinical trials as brentuximab vedotin by Seattle Genetics and was approved by the FDA in 2012 for the treatment of relapsed Hodgkins Lymphoma and anaplastic large cell lymphoma.[85,90] This example demonstrates the potential of cyanobacterial metabolites as leads for effective drugs.

Dolastatin 10 (**6**)

Monomethyl Auristatin E (**7**)

Brentuximab Vedotin
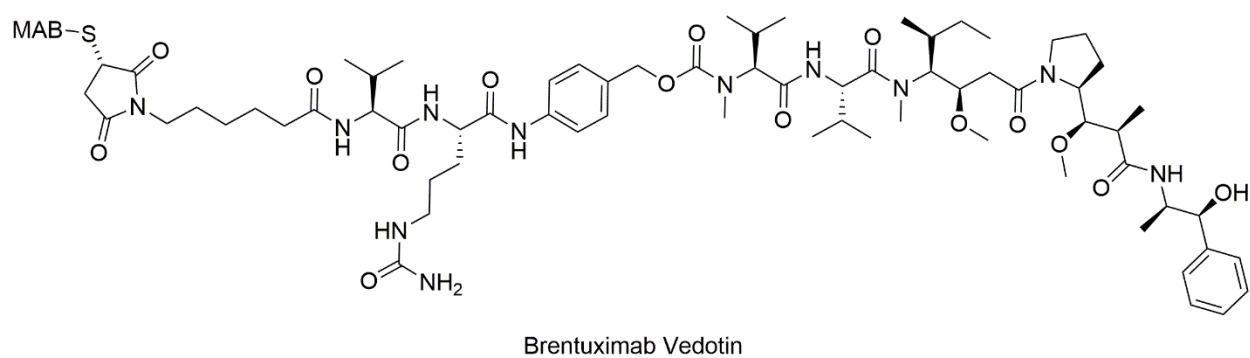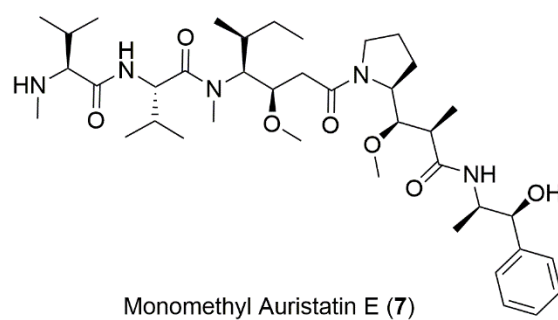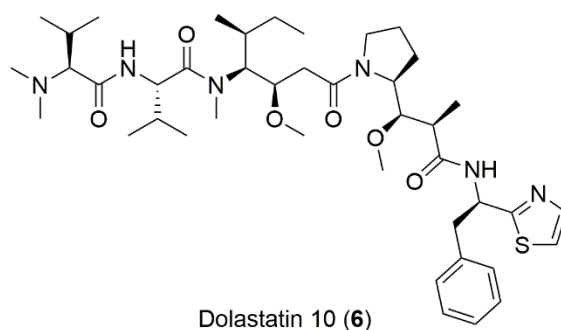
Figure 7: Structures of the natural product dolastatin 10, the synthetic analog MMAE, and the FDA approved drug Brentuximab vedotin

Nostoc: A Widespread and Well-Adapted Cyanobacterium

Nostoc is a genus within the Nostochophycidae, that has both unique biology and prolific secondary metabolite producing capabilities. *Nostoc* has been well known since the early days of taxonomy and has undergone numerous taxonomic revisions. *Nostoc* was initially positioned in the tree of life as a primitive relative of plants, and although cyanobacteria partook in the endosymbiosis that led to the development of chloroplasts, this placement was incorrect.[91] This placement among algae is one reason cyanobacteria have persisted under the moniker of blue-green algae.[92] As microscopy and molecular techniques became more prevalent, *Nostoc* was placed correctly within the prokaryota and more specifically as a member of the phylum Cyanobacteria.[91] Within Cyanobacteria, *Nostoc* occupies multiple lineages in the Nostocaceae family. As with many cyanobacterial genera, *Nostoc* is polyphyletic, meaning the genus consists of multiple non-clading groups of organisms which should be separated into their own genera. Bergey's Manual of Systematic Bacteriology, Second Edition, recognizes five *Nostoc* clusters.[93] One of these clusters, *Nostoc* Cluster II was renamed as the *genus novum Desmonostoc* in 2013.[94] The other clusters will likely need to be further studied and designated as their own separate genera, but that is beyond the scope of this dissertation. This dissertation will focus on the two large taxonomic clusters, *Nostoc sensu stricto*, or *Nostoc* Group I, and *Nostoc* Group III and will colloquially refer to them together as *Nostoc*.
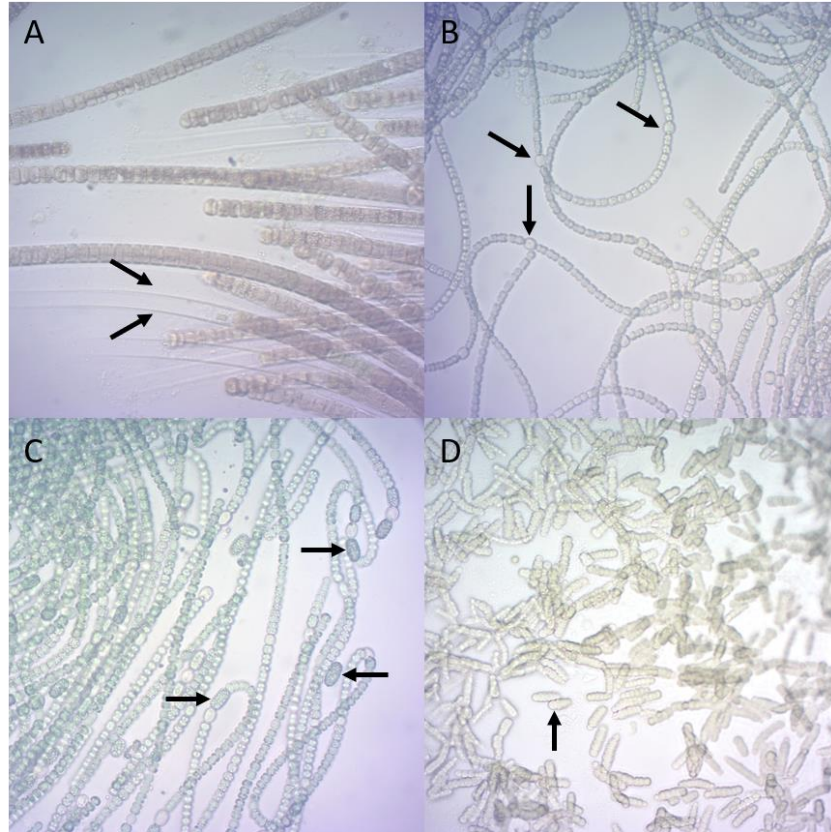
Figure 8: Cell and filament types of cyanobacteria: A) Sheath B) Heterocysts C) Akinetes D) Hormogonia. All images taken at 40X.

The genus *Nostoc* can be found in soil, rocks, freshwater, and brackish water.[91] It has a global distribution, thriving in locations from the tropics to the Arctic and Antarctic.[91,95,96] The cells of *Nostoc* are round or barrel-shaped and exist in isopolar trichomes. The cells divide in a single plane and are sometimes enveloped by a mucilaginous sheath (Figure 8A).[93] *Nostoc* has a complex life cycle, not commonly seen in other cyanobacteria, which involves cellular differentiation as well as specialized filament formation. *Nostoc* has three distinct cell types that are present at different points in the life cycle: vegetative cells, heterocysts, and akinetes.[97,98] Vegetative cells

make up most cells present in a filament and are responsible for performing

photosynthesis. These cells are constitutively present during cell replication and can

break from the filament to form new filaments.[98] Heterocysts are specialized,

differentiated cells that perform nitrogen fixation (Figure 8B). Heterocysts, in *Nostoc*,

are formed at regular junctions within the filament based on a diffused nitrogen gradient

and do not actively replicate.[74,76,99] The heterocysts possess a thickened cell wall

composed of a specialized type of glycolipid, known as heterocyst glycolipids, which

makes them appear larger than the vegetative cells.[100] Akinetes are a torpid spore-like

cell that are produced during nutrient limiting conditions, stationary phase growth, or

consistent low light levels (Figure 8C).[99,101] This cell type allows the bacterium to

remain dormant until conditions are more favorable for growth. During akinete

formation, granules of cyanophycin and glycogen begin forming as storage molecules

for fixed nitrogen and carbon respectively, the cell wall thickens due to the production of

heterocyst glycolipids, and metabolism slows drastically.[99,101] Akinetes can withstand

desiccation and low temperatures, but not temperatures higher than a vegetative cell,

differentiating them from other types of bacterial spores.[101] *Nostoc* also form

hormogonia, motile filaments that allow for the dispersal of the colony and are integral in

the formation of symbioses (Figure 8D).[76,98] The symbiosis of *Nostoc* with lower plants

and one genus of angiosperm has been well studied.[76] In the case of members of the

angiosperm genus *Gunnera*, *Nostoc* has been shown to enter the stem and form an

intracellular symbiosis with the plant.[102] Recently, it was shown that the hormogonia

formation and cellular invasion of *Gunnera* species is mediated by the cyanobacterial

natural product nostopeptolide A1.[103] It is thought this symbiosis has developed due to

the cyanobacteria's ability to fix nitrogen for the plant, while the plant provides safety and fixed carbon to the cyanobacteria.[76] *Nostoc* is also known to form symbioses with lichen forming fungi to form several types of lichens.[104,105] The ability to differentiate cells to simultaneously perform photosynthesis and nitrogen fixation within the same filament, enter symbiosis with a variety of other organisms, and utilize cell differentiation to survive low nutrient conditions has made *Nostoc* a widespread and well-adapted organism.

    Nostoc has been found to have a comparably large genome for a bacterium, approximately 8 Mb, and a larger genome than many cyanobacteria. This is likely due to a variety of factors including: the complex life cycle and cellular differentiation observed in members of *Nostoc*, genes for photosynthesis and nitrogen fixation, as well as genes required for secondary metabolite production.[106–108] As of this dissertation, there are currently fifteen sequenced *Nostoc* genomes on NCBI, at varying levels of completeness. The seven complete genomes range in size from 6.3 Mb to 9.1 Mb with an average genome size of 7.5 Mb and an average GC content of 41.25% (Table I). Only a few sequenced strains have been analyzed for their ability to produce natural products, with strains from PCC and NIES being the most well studied. These studies have found that *Nostoc* has a large potential to produce natural products, with each genome containing multiple diverse types of BGCs. As full genome sequencing becomes more accessible and less expensive, the metabolic potential of *Nostoc* will become clearer. Described and published BGCs will be detailed further, along with their respective products in the next section.

TABLE I: SEQUENCED GENOMES OF NOSTOC SPP. AVAILABLE FROM GENBANK INCLUDING SIZE AND COMPLETENESS.

Assembly Number	Strain ID	Size (Mb)	Completeness Level
GCA_000020025	PCC 73102	9.1	Complete
GCA_001904715	FACHB-389	9.0	Complete
GCA_001298445	CENA21	7.1	Complete
GCA_000009705	PCC 7120	7.2	Complete
GCA_000316625	PCC 7107	6.3	Complete
GCA_000316645	PCC 7524	6.7	Complete
GCA_001548375	NIES-3756	7.0	Complete
GCA_001712795	KVJ20	9.2	Scaffold
GCA_002154725	106C	8.7	Contigs
GCA_002154695	T09	8.2	Scaffold
GCA_002155185	RF31Y	9.2	Scaffold
GCA_002245975	213	8.3	Scaffold
GCA_002245985	232	9.2	Scaffold
GCA_002246015	210A	8.3	Scaffold
GCA_001698435	MBR210	6.9	Scaffold

Nostoc: A Source of Natural Products

The secondary metabolites produced by *Nostoc* are also complex. *Nostoc* produces many NRPs, PKs, RiPPs, terpenes, and shikimate derived compounds that

have a wide range of biological activities.[78,79] The order Nostocales appears to be a

particularly prolific source of secondary metabolite gene clusters in the analysis

performed by Shih et al.[84] Several structurally diverse natural products have been

identified from *Nostoc* strains.[78] BGCs have only been identified for some of the

compounds produced by *Nostoc*, and there is still much work to be done to match

known compounds to their BGCs. Likewise, with the growing use of genome mining in

natural products research, many orphan BGCs have been identified in *Nostoc* genomes

that do not match any known compounds produced by *Nostoc*, emphasizing the amount

of chemical diversity still to be discovered from this genus. The following section

describes compounds identified from *Nostoc* and represents many, but not all the

compounds that have been identified to date. They are classified by biosynthetic origin

or likely biosynthetic origin, if no BGC has been linked to the production of the

compound.

Compounds produced by NRPS
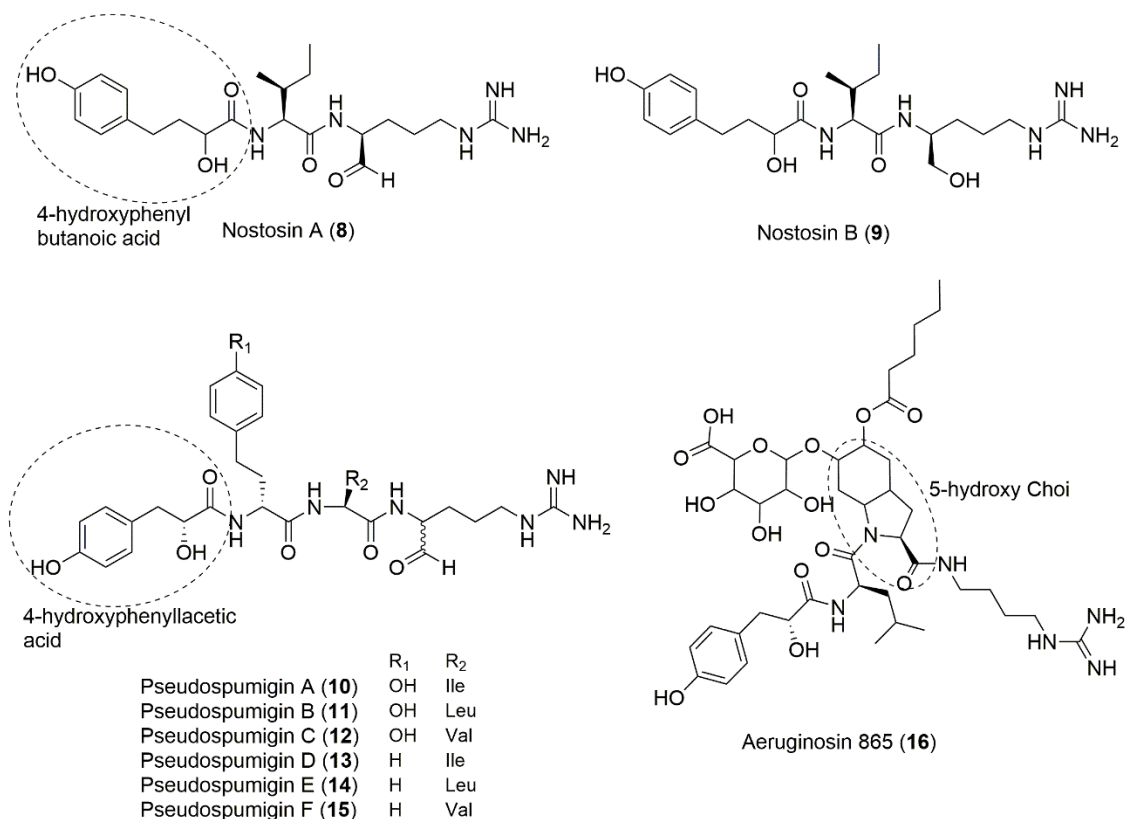
Figure 9: Aeruginosins and analogs produced by Nostoc spp.

Aeruginosins and their structural analogs are diverse NRPs with protease inhibitory activities. They are typically associated with bloom forming cyanobacteria such as *Microcystis* and *Nodularia*.[109] However, the genus *Nostoc* has also been identified as a source of structurally unique and highly modified aeruginosin analogs (Figure 9).[110] The smallest of these are the nostosins. Nostosins A (**8**) and B (**9**) lack the characteristic 2-carboxy-6-hydroxyoctahydroindole (Choi) or proline/methylproline amino acids that are commonly found in other aeruginosin analogs, although they

conserve the 4-hydroxyphenyl acid, a hydrophobic amino acid, and a modified arginine.[111] Nostosin A was found to inhibit trypsin (IC$_{50}$ of 0.35 µM) while **9** had a two orders of magnitude higher IC$_{50}$ value, suggesting the modified arginine with an aldehyde is important for the activity. Indeed, it has been shown that a similar compound, leupeptin, which also has an arginal, forms a covalent bond with a serine in the active site of trypsin.[111,112] Molecular docking experiments provided evidence that **8** also covalently binds to the same serine in the trypsin active site.[111] Pseudospumigins A-F (**10-15**) are another example of aeruginosin analogs found in *Nostoc* spp. They also lack the characteristic Choi or proline/methylproline amino acids. Pseudospumigins, however, do maintain the characteristic size of the aeruginosins with a 4-hydroxyphenyllacetic acid and three amino acids, one of which is a modified arginine.[112] The pseudospumigins showed lower inhibition of serine proteases than other arginal containing aeruginosin analogs. It is possible that the lack of a cyclized amino acid, like Choi or proline, and/or the increased molecular size hinder binding to the active site; however there is currently no clear explanation to the wide range of activities found in the aeruginosin analogs.[112] Finally, aeruginosin 865 (**16**) is a NRP with a more standard aeruginosin scaffold, containing 4-hydroxyphenyllacetic acid, a hydrophobic amino acid, a Choi, and a modified arginine. Aeruginosin 865 contains a 5-hydroxy-Choi amino acid that is decorated with both a fatty acid unit and a uronic acid carbohydrate.[113] Like most aeruginosin analogs, **16** is active against serine proteases, but studies have shown that it also inhibits translocation of NF-κB into the nucleus through an undescribed mechanism.[110,113] Aeruginosin 865 is the first aeruginosin with this activity, making it of interest as a potential immunomodulatory drug, due to the

central role NF-κB plays in regulation of the immune response. Currently, the only BGC published of an aeruginosin analog from a *Nostoc* is that of the pseudospumigins (MF668123.2). The BGCs for other aeruginosin analogs have been identified in other non-*Nostoc* cyanobacterial genomes (AM071396.2, FJ609416.1).[109,112] The genes *spuA* and *spuB* from the spumigin BGC (CP007203.2) identified in *Nodularia spumigena* share high similarity to the homologous pseudospumigin genes, with only one adenylation domain in *spuB* selecting for a different amino acid.[112,114]

	R_1	R_2	R_3	R_4	R_5
Anabaenopeptin SA1 (**17**)	Arg	Ile	PNV	Asn	Phe
Anabaenopeptin SA2 (**18**)	Arg	Val	Hty	Ser	Phe
Anabaenopeptin SA3 (**19**)	Lys	Ile	Hty	Ala	Phe
Anabaenopeptin SA4 (**20**)	Lys	Ile	PNV	Asn	Phe
Anabaenopeptin SA5 (**21**)	Ile	Val	PNV	Asn	Phe
Anabaenopeptin SA6 (**22**)	Ile	Ile	Hph	Asn	Phe
Anabaenopeptin SA7 (**23**)	Ile	Ile	PNV	Asn	Phe
Anabaenopeptin SA8 (**24**)	Ile	Ile	PNL	Asn	Phe
Anabaenopeptin SA9 (**25**)	Phe	Ile	HtyCl	Gly	Hph
Anabaenopeptin SA10 (**26**)	Phe	Val	Hty	Gly	HtyCl
Anabaenopeptin SA11 (**27**)	Phe	Ile	Hty	Gly	HtyCl
Anabaenopeptin SA12 (**28**)	Phe	Val	Hty	Gly	Hty
Anabaenopeptin SA13 (**29**)	Tyr	Val	Hty	Ser	Phe
Anabaenopeptin A (**30**)	Tyr	Val	Hty	Ala	Phe
Anabaenopeptin D (**31**)	Phe	Val	Hty	Ala	Phe
Nostamide A (**32**)	Phe	Ile	Hph	Gly	Hty
Anabaenopeptin 896 (**33**)	Phe	Ile	PNL	Asn	Phe
Anabaenopeptin 882 (**34**)	Phe	Ile	PNV	Asn	Phe
Anabaenopeptin 871 (**35**)	Phe	Ile	Hty	Hty	Hty
Anabaenopeptin 857 (**36**)	Phe	Val	Hty	Ala	Hty
Anabaenopeptin 855 (**37**)	Phe	Ile	Hph	Ala	Hty
Anabaenopeptin 841 (**38**)	Phe	Val	Php	Ala	Hty
Anabaenopeptin KVJ827 (**39**)	Tyr	Val	Hph	Gly	Hph
Anabaenopeptin KVJ841 (**40**)	Tyr	Ile	Hph	Gly	Hph
Anabaenopeptin KVJ811 (**41**)	Phe	Val	Hph	Gly	Hph
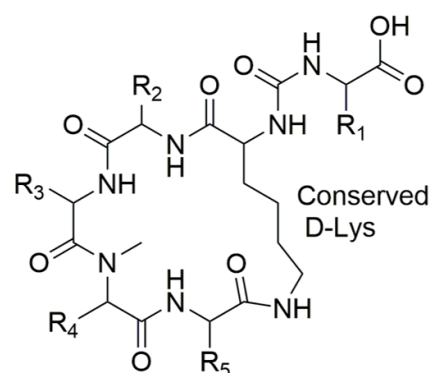
Figure 10: Anabaenopeptins produced by Nostoc spp.

Anabaenopeptins are cyclic hexapeptides with serine protease and phosphatase inhibition activities. The structures of all known anabaenopeptins contain a ureido bond between a conserved D-Lysine and a variable amino acid outside the cyclic peptide (Figure 10).[115,116] The D-Lysine is the only conserved amino acid in the structure of anabaenopeptins; however, there appear to be broad preferences for particular types of amino acids at each of the other positions. Shishido et al. studied the biosynthetic gene clusters of anabaenopeptin producing strains, and the identity of the amino acids incorporated in each position of 115 described anabaenopeptins from cyanobacteria. They found that the identity of the amino acid outside the ring was correlated with the taxonomy of the producing strain, suggesting a long evolutionary history for this BGC.[117]

	R₁	R₂	R₃
Namalide B (**42**)	Ile	Ile	OH
Namalide C (**43**)	Ile	Val	OH
Namalide D (**44**)	Ile	Ile	H
Namalide E (**45**)	Ile	Val	H
Namalide F (**46**)	Val	Ile	H
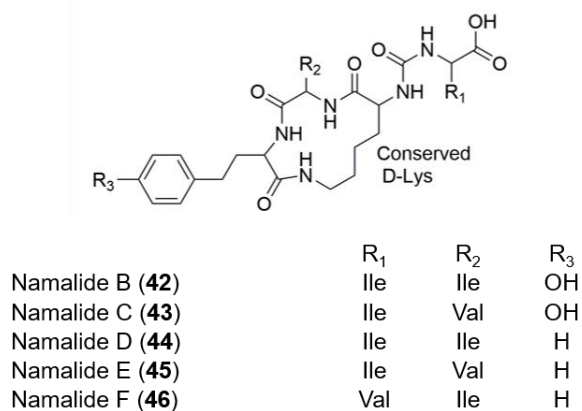
Figure 11: Namalides produced by Nostoc spp.

In the same publication, namalides B-F (**42-46**), cyclic tetrapeptides that appear to be structurally similar to anabaenopeptins were also reported from a *Nostoc* that produces anabaenopeptins (Figure 11). The anabaenopeptin BGC was identified in the producing organism, but no separate gene cluster was identified for the namalides, suggesting that they are produced from the same BGC by skipping two modules (MF741679.1, MF741700.1).[117] This indicates that the chemical diversity of the anabaenopeptins could likely be due to the genetic flexibility of the BGCs.[118] Over twenty anabaenopeptins (**17-41**) have been isolated from *Nostoc*, and due to the high capacity for chemical diversity produced by these flexible gene clusters, it is likely more will be obtained in the future.[115,116,119]
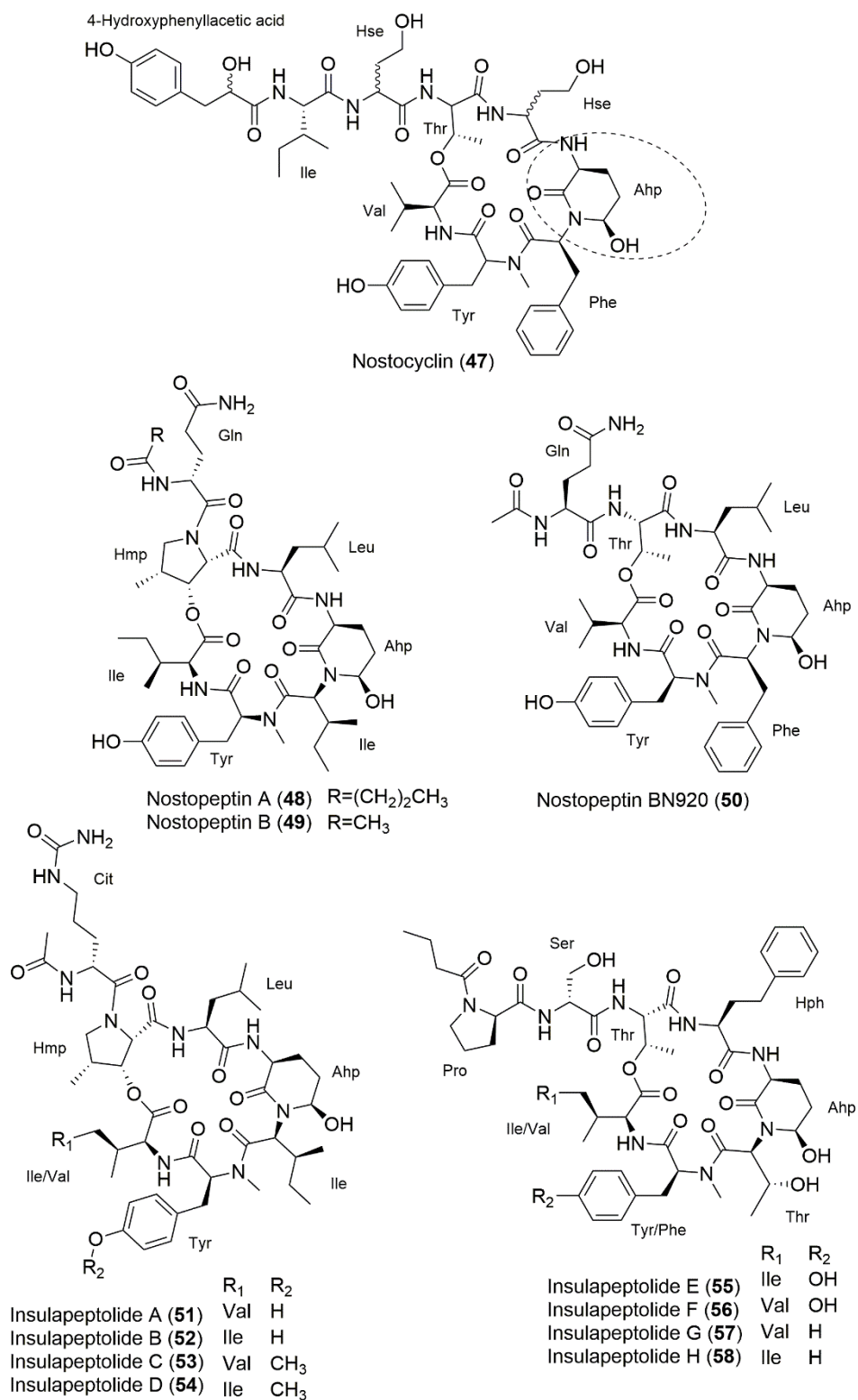
Figure 12: Cyanopeptolin analogs produced by Nostoc spp. with hemiaminal forming Ahp amino acid highlighted on nostocyclin (**47**)**.**

Another group of protease inhibiting NRPs produced by *Nostoc* are analogs of cyanopeptolin (Figure 12). Each of these compounds is a cyclic hexa-depsipeptide with a variable length tail and a conserved 3-amino-6-hydroxy-2-piperidone (Ahp) amino acid. The Ahp forms a hemiaminal with the adjacent amino acid, typically isoleucine or phenylalanine. It is believed that this conserved Ahp is important for the protease inhibition.[120] The first described cyanopeptolin analog from *Nostoc* was nostocyclin (**47**) which contains two homoserines as well as a unique 4-hydroxyphenyllacetic acid tail in place of the more typical acetate or butyrate moieties.[121] In a separate report, **47** was shown to inhibit the same protein phosphatases as the hepatotoxic microcystins, Protein Phosphatase 1 (PP1). Nostocyclin was found to inhibit the activity of the catalytic subunits PP2A and PPP4 of PP1 *in vitro* with IC$_{50}$s of 35 nm and 50 nm respectively, approximately 600 times less potent than microcystin LR.[122] Nostopeptin A (**48**) and B (**49**) are similar in structure to **47** and other cyanopeptolin analogs, however, they have been shown to inhibit both chymotrypsin and elastase with modest activity of about 1 µM each. The Nostopeptins all have glutamine as the tail amino acid, outside the cyclic peptide. Nostopeptin A and B both have a 3-hydroxy-4-methylproline (Hmp) amino acid which closes the cyclic peptide, while Nostopeptin BN920 (**50**) has a threonine, similar to **47**.[120,123] More recently the insulapeptolides A-H (**51**-**58**) were identified from *Nostoc insulare*. These compounds were identified as human leukocyte elastase inhibitors, but also had varying inhibition of other serine proteases. Insulapeptolides A-D contain an Hmp amino acid to form the macrocyclic ring, while insulapeptolides E-H contain a threonine and have longer amino acid tails. Interestingly, the differences in structure appear to relate to the inhibition of human

leukocyte elastase, with **51**-**54** being approximately 10 times more potent.[124] Currently

no BGCs have been published for the cyanopeptolin analogs produced by *Nostoc.*

	R$_1$	R$_2$	R$_3$	R$_4$	R$_5$	R$_6$
Nostoweipeptin W1 (**62**)	CH$_3$CO	CH$_3$	CH$_3$	OH	OH	CH$_3$
Nostoweipeptin W2 (**63**)	CH$_3$CO	CH$_3$	CH$_3$	OH	H	CH$_3$
Nostoweipeptin W3 (**64**)	CH$_3$CO	CH$_3$	CH$_3$	H	OH	CH$_3$
Nostoweipeptin W4 (**65**)	CH$_3$CO	H	CH$_3$	OH	OH	CH$_3$
Nostoweipeptin W5 (**66**)	CH$_3$CO	CH$_3$	CH$_3$	OH	OH	H
Nostoweipeptin W6 (**67**)	CH$_3$CO	CH$_3$	H	OH	OH	CH$_3$
Nostoweipeptin W7 (**68**)	CH$_3$CO	CH$_3$	H	OH	OH	CH$_3$
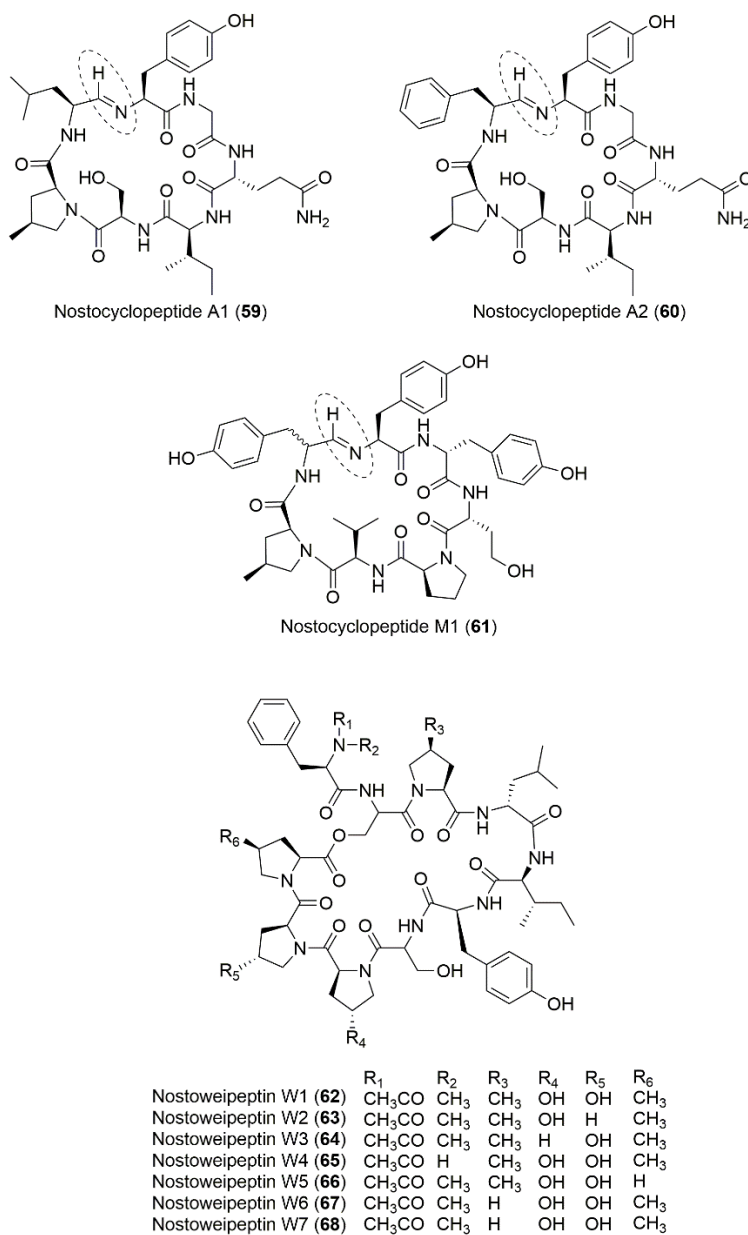
Figure 13: Other NRPS produced compounds isolated from Nostoc spp. with imino bond in
nostocyclopeptides highlighted.

While the genus *Nostoc* produce a diverse group of protease and phosphatase inhibitors, it has also been found to produce many other NRPs with wide ranging bioactivities. Nostocyclopeptides A1 (**59**), A2 (**60**), and M1 (**61**) and the nostoweipeptins W1-W7 (**62-68**) are produced by NRPSs and although they have different structures, they have similar activities (Figure 13).[125,126] Both sets of compounds are 4-methylproline containing cyclic peptides, however, nostocyclopeptides are smaller and have an imino bond while the larger nostoweipeptins contain multiple 4-methylprolines as well as 4-hydroxyproline.[127] [1-^{13}C, ^{18}O$_2$] acetate feeding experiments by Golakoti et al. provided evidence that the imino bond of the nostocyclopeptides is derived from the carbonyl of the amino acid directly following the 4-methylproline.[128] The BGC for the nostocyclopeptides has been identified (AY167420.1) and consists of two large NRPS genes (composed of three and four elongation modules respectively), an ABC transporter gene, and a beta-lactamase type gene. Genes encoding an alcohol dehydrogenase and a carboxylate reductase were also present and were shown to be responsible for 4-methylproline biosynthesis.[126,129] A series of studies has shown that both nostocyclopeptides and nostoweipeptins inhibit the hepatocyte drug transporters OATP1B1 and OATP1B3. These drug transporters are found in liver tissue where they allow uptake of xentobiotics, including the hepatotoxic cyanobacterial compounds **4** and **5**. Both nostocyclopeptides and nostoweipeptins are nontoxic, making them potential antitoxins for highly hepatotoxic compounds.[125–127]

Compounds Produced by PKS
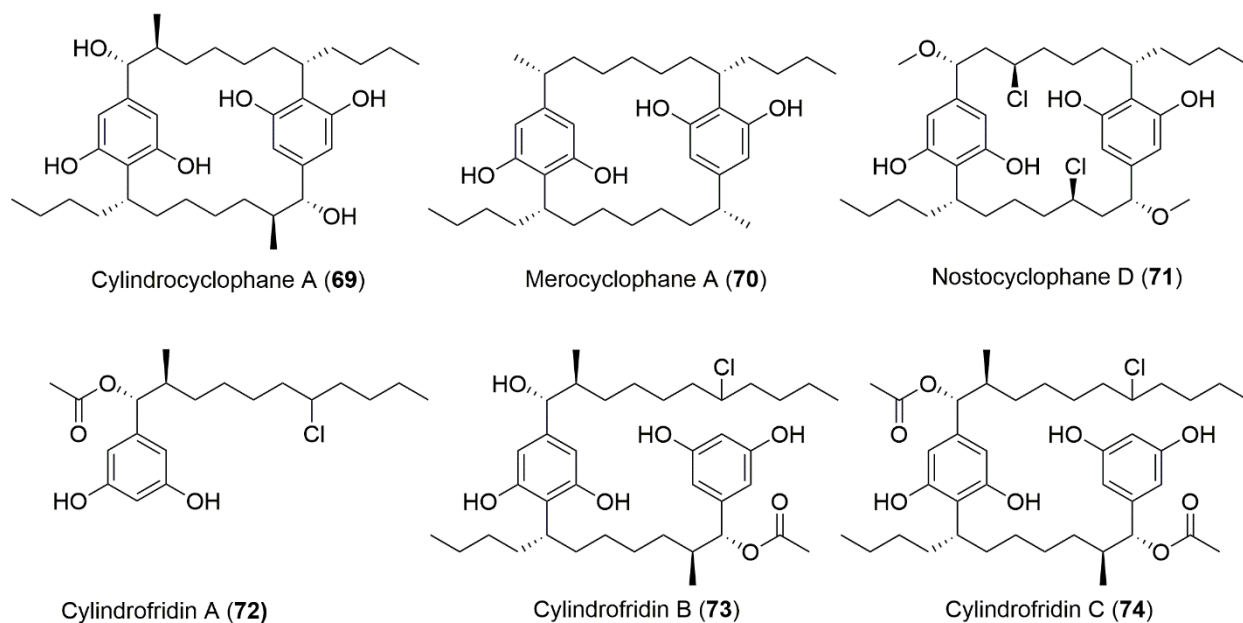
Figure 14: Representatives of the three [7,7] paracyclophane core structures and the biosynthetic intermediate cylindrofridins

[7,7] paracyclophanes are the most common PKS products found in *Nostoc*. They are a group of structurally diverse compounds that can be classified into three categories based on their core structures: the cylindrocyclophanes (represented by cylindrocyclophane A (**69**)) which contain a branched *β*-methyl group, the merocyclophanes (represented by merocyclophane A (**70**)) which contain a branched *α*-methyl group, or the nostocyclophanes (represented by nostocyclophane D (**71**)) which do not contain either a *β*-methyl or an *α*-methyl group in the core structure (Figure

14).[130–132] These compounds are produced by a type I/type III PKS hybrid BGC that

has been identified and described for two compounds with the cylindrocyclophane core,

the cylindrocyclophanes (JX477167.1), and the carbamidocyclophanes

(KT826756.1).[133,134]
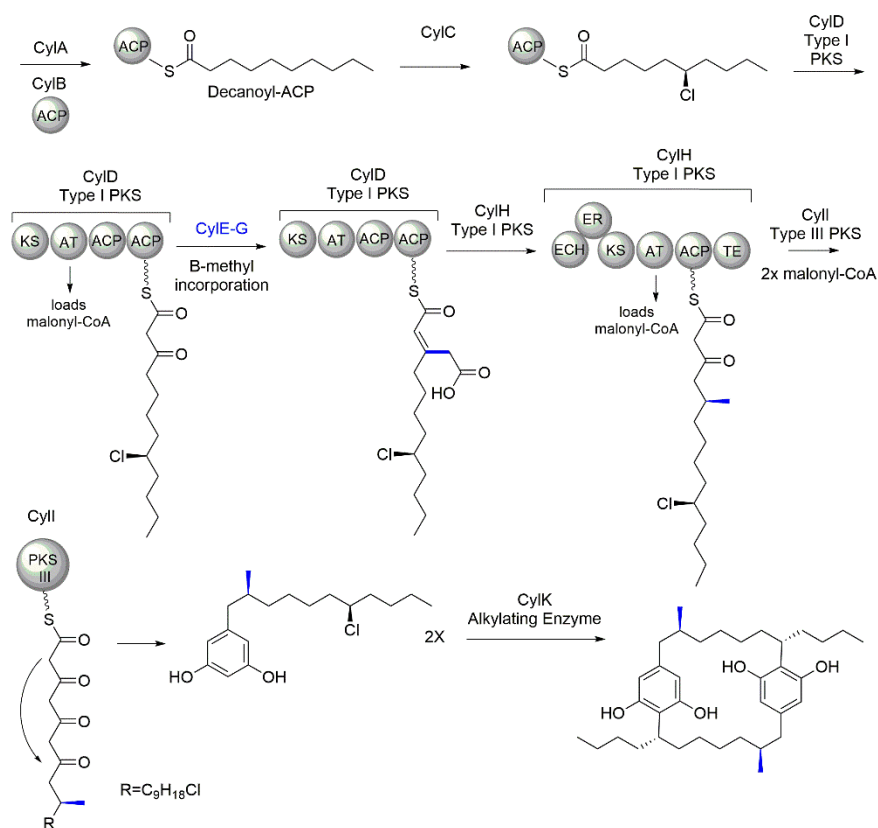
Figure 15: Biosynthesis of cylindrocyclophanes as described by Nakamura et al. 2012 and Nakamura et al 2017. Enzymes involved in beta-methyl installation and beta-methyl groups are highlighted in blue.[133,135]

The biosynthesis involves production of two units of a chlorinated alkyl resorcinol structure by the type I/type III PKS, β-methyl group installation through an HMG-CoA Synthase type mechanism, followed by a subsequent dimerization event to form the macrocycle. This dimerization event involves carbon-carbon bond formation between the resorcinol ring on one monomer and the chlorinated carbon on the aliphatic carbon chain of the other monomer (Figure 15).[133,135,136] Modifying enzymes have also been identified such as carbamido-transferases and halogenases that further diversify the structures produced by these BGCs.[134,137] Cylindrocyclophanes are the most well studied class of cyclophanes with thirty structures identified to date. This class has been identified in both *Nostoc* and Cylindrospermum.[130,137–140] Merocyclophanes and nostocyclophanes have each only been reported once, both from *Nostoc* strains. The strain that produces merocyclophanes is currently available in the UIC culture collection, while the strain that produces nostocyclophanes, UTEXB1932, is no longer available at the UTEX culture collection due to contamination.[131,132] [7,7] paracyclophanes have been reported to have a variety of biological activities including anti-fungal, anti-tuberculosis, anti-bacterial, protease inhibition, and cytotoxicity.[125-127,132-135] It has been shown that cyclophane biosynthetic intermediates, such as the linear cylindrofridin B (**73**) and C (**74**), retain no biological activity while the monomer, cylindrofridin A (**72**), is mildly anti-bacterial.[141]
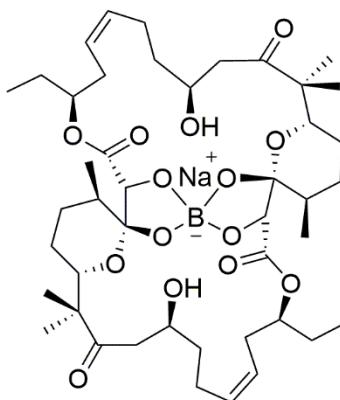
Borophycin (**75**)

Figure 16: Structure of boron salt containing borophycin

Borophycin (**75**) is another PKS produced compound identified from *Nostoc.* Borophycin represents a rare group of natural products that contain a boron and sodium salt, other examples being boromycin and aplasmomycin, both produced by *Streptomyces* spp (Figure 16).[142,143] Borophycin, like the [7,7] paracyclophanes, appears to be biosynthesized by the dimerization of two PKS derived units. [13]C acetate feeding experiments suggested the likely PKS origin of the compound, although the methyl groups found along the macrocycle did not show any ^{13}C enrichment. The source of the methyl groups was identified in a separate feeding experiment with methyl-[13]C L-methionine, in which ^{13}C enrichment was observed in each of the methyl groups surrounding the macrocycle. This suggests a type I methyl transferase, using S-Adenosyl methionine, installs the methyl groups around the core borophycin structure. Borophycin, like other boron-sodium salt containing natural products, was found to be cytotoxic with an IC$_{50}$ of 77 nM to LoVo human colon cancer cells.[144]

Compounds Produced by NRPS/PKS Hybrids
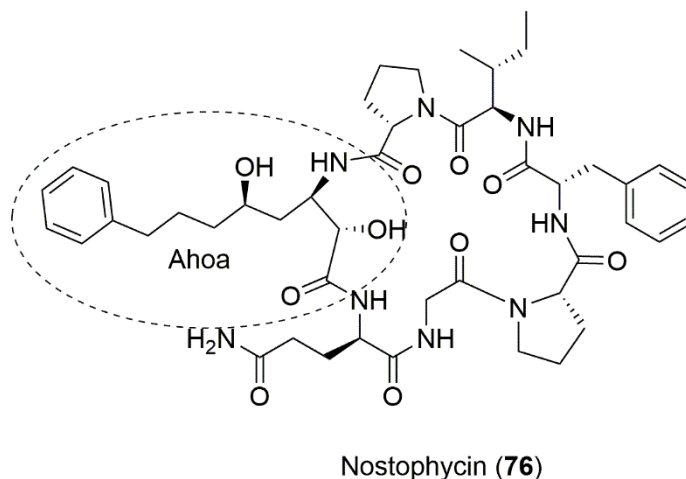
Nostophycin (**76**)

Figure 17: Structure of nostophycin with the beta amino acid Ahoa highlighted

Nostophycin (**76**) is a cyclic NRPS/PKS hybrid with a PKS tail in the form of the β amino acid 3-amino-2,5-dihydroxy-8-phenyloctanoic acid (Ahoa) (Figure 17). Similar to microcystins, that are produced by the same strain, *Nostoc* sp. 152, nostophycin (**76**) contains an elongated β amino acid tail as well as two D-amino acids (Figure 5).[145] Fewer et al. identified the nostophycin BGC (JF430079.1) and noted the similarities of the adenylation loading domains between nostophycin, microcyctin, nodularin, and cryptophycin BGCs. A maximum likelihood tree formed from the sequence of those adenylation domains, however, showed that they do not form a monophyletic clade, suggesting the BGCs are not immediately related and may have evolved separately.[146] The nostophycin BGC consists of three genes: a type I PKS gene, a hybrid type I PKS/NRPS gene with a monooxygenase domain, and an NRPS gene. The first two genes are responsible for production of the Ahoa, while the NRPS modules are

responsible for production of the cyclic peptide.[146] Unlike the microcystins, nostophycin

has only been shown to be weakly cytotoxic. Both nostophycin and the microcystins,

however, have been shown to be produced in higher abundance when the strain is

grown under low light or low phosphate conditions.[147]

	R$_1$	R$_2$	R$_3$
Nostopeptolide A1 (**77**)	Leu	H	Ile
Nostopeptolide A2 (**78**)	Leu	H	Val
Nostopeptolide A3 (**79**)	H	Leu	Ile

	R$_1$	R$_2$
Nostopeptolide L1 (**80**)	CH$_3$	OH
Nostopeptolide L2 (**81**)	CH$_3$	H
Nostopeptolide L3 (**82**)	H	OH
Nostopeptolide L4 (**83**)	H	H
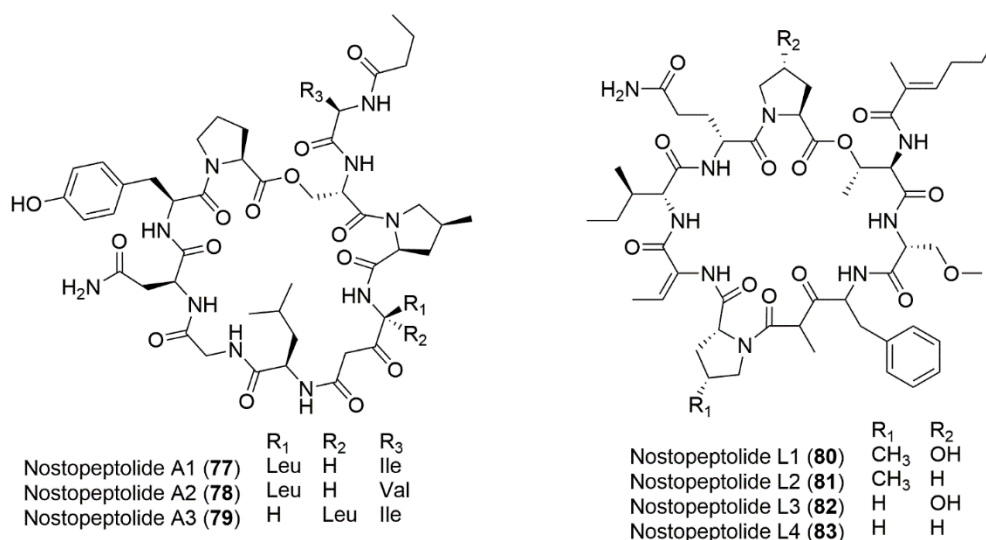
Figure 18: Nostopeptolides produced by Nostoc spp.

Nostopeptolides are a series of NRPS/PKS hybrid cyclic depsipeptides found in

low abundances in terrestrial *Nostoc* spp., especially those often found in symbiosis

with other organisms. Each analog contains nine amino acids, one polyketide unit, and

an acyl tail (Figure 18). Like many other depsipeptides produced by *Nostoc*,

nostopeptolides A1-A3 (**77-79**) and L1-L4 (**80-83**) contain either 4-methylproline or 4-

hydroxyproline.[126,148] The BGC for nostopeptolide A1 has been identified (AF204805.2)

and is composed of three NRPS genes, one PKS gene, and genes that are responsible

for production of 4-methylproline.[149,126] The 4-methylproline moiety appears to be

important in the blocking of OATP1B1 and OATP1B3 hepatocyte drug transporters, similar to the nostocyclopeptides and nostoweipeptins.[126] Another recent study showed that nostopeptolides are likely involved in cellular differentiation and plant symbiosis. Liaimer et al. were able to show that under standard growing conditions nostopeptolide A1 (**77**) is constitutively produced and located in the sheath of the *Nostoc* strain ATCC 29133. When production of nostopeptolide A1 (**77**) was downregulated due to the knock-out of an undescribed and unrelated PKS gene, filaments differentiated into motile and invasive hormogonia. They also showed that in the wild-type strain this motile phenotype could be recapitulated by incubating the cells with the exudate of plants with which *Nostoc* forms symbiosis.[103] This suggests that chemical cross-talk between the species likely leads to a change in the secondary metabolite profile of the strain, and ultimately a change in biological life cycle.
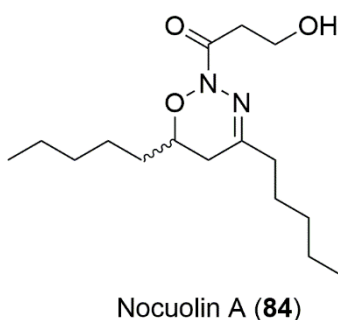
Nocuolin A (**84**)

Figure 19: Structure of Nocuolin A

Due to the incorporation of diverse units by NRPS/PKS hybrid gene clusters, structurally unique natural products such as nocuolin A (**84**) can be produced. Nocuolin A was identified in three different cyanobacterial strains, *Nostoc*, *Nodularia*, and *Anabaena*, as an inducer of apoptosis in HeLa cells. The compound is the first known naturally occurring 1,2,3 oxadiazine, containing a direct O-N-N bond within a heterocycle (Figure 19). A conserved BGC was reported in two of the three strains that produced the compound (KP143720.1, KY594676.1).[150] The putative BGC is an NRPS/PKS hybrid with one PKS gene and one NRPS gene. The BGC also contains a variety of enzymes responsible for fatty acid processing and incorporation, columbamide-like halogenation, and enzymes for benzene oxidative degradation. Curiously, the adenylation domain of the NRPS was predicted to incorporate an aromatic amino acid, which is not present in the final compound. Stable isotope labeling experiments with ^{15}N labeled tryptophan suggested that tryptophan may be the source of one or both Nitrogens in the molecule.[150]
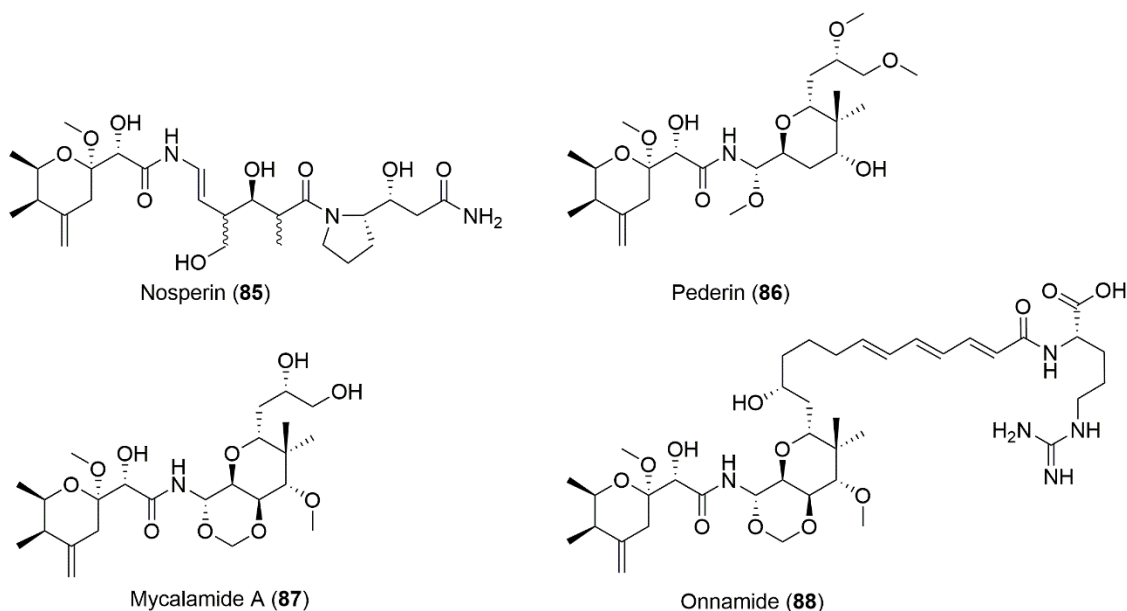
Figure 20: Nosperin and similar compounds produced by other symbiotic bacteria

Nosperin (**85**) represents a rare form of NRPS/PKS hybrid produced compound

in cyanobacteria, a composite Type I-*trans* AT PKS/Type I-*cis* AT PKS/NRPS hybrid.

Nosperin was isolated from the photobiont of a lichen, a *Nostoc*, and the BGC

(GQ979609.2) was used to prioritize the study of the compound due to its unusual

biosynthetic architecture.[151] The BGC contains one *trans* AT type I PKS gene, meaning

the AT domains are not included in each module and rather act iteratively from a single

AT gene outside the type I PKS gene, one *trans* AT type I PKS/NRPS hybrid gene, and

one *cis* AT type I PKS gene. The genes *nspA* and first half of *nspC* had high homology

to the biosynthetic genes of pederin (**86**), an insect associated bacterial natural product,

and onnamide (**88**), a sponge associated bacterial natural product. The second half of

nspC and *nspD* had high homology to cyanobacterial genes. This data led Kampa et al.

to hypothesize that this composite BGC may be the result of horizontal gene transfer,

leading to a part other-bacterial, part cyanobacterial BGC.[151] The genes immediately

upstream and downstream from the nosperin BGC contain transposases as well as

other plasmid associated genes, suggesting that the unique biosynthetic architecture is

likely the result of horizontal gene transfer. The compound was identified by culturing

the *Nostoc* photobiont separately from the lichen and using ^{13}C stable isotope labeling

to identify the compound in the cell extract. While **85** has not been evaluated for

bioactivity, it contains the same core N-acyl linked tetrahydropyran structure as other

cytotoxic compounds such as **88** and mycalamide A (**87**), a sponge associated bacterial

natural product (Figure 20). This portion of the structure was found to be critical to the

activity associated with **87**, suggesting that **85** could have similar bioactivity.[110,151] This

is especially interesting as *Nostoc* is the only source of these types of compounds that

can be readily cultured, making it a potential source of these unique cytotoxic

compounds for future studies.

Cryptophycin A (**89**)
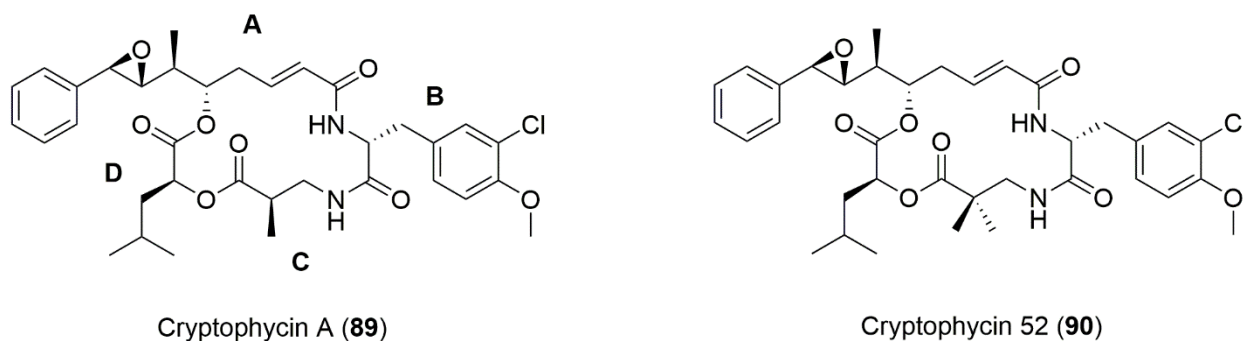
Cryptophycin 52 (**90**)

Figure 21: Structures of two cryptophycins produced by Nostoc spp. with the four units, A, B, C, and D labeled around the structure.

Cryptophycins are NRPS/PKS produced depsipeptides that contain a 7,8-epoxy-5-hydroxy-6-methyl-8-phenyl-2-octenoic acid (phenyloctenoic acid) (A), a modified tyrosine (B), methyl β-alanine (C), and modified leucine (D) (Figure 21).[152,153] The observed cytotoxicity and the structure containing a macrocycle, epoxide, and highly modified amino acids made cryptophycin A (**89**), and analogs, the subject of many studies.[154] A BGC has been described for the cryptophycins (EF159954.1) and consists of two PKS genes and two NRPS genes as well as genes encoding a cytochrome P450, and enzymes for deamination, decarboxylation, and halogenation. The two PKS genes are responsible for incorporation of the phenyloctenoic acid (A), while the NRPS genes incorporate a methoxy-tyrosine (B), a methyl β-alanine (C), and leucinic acid (D).[155] Studies found that cryptophycins interact with the vinca alkaloid domain of tubulin, inhibiting depolymerization of microtubules.[156,157] Many analogs were produced by the cyanobacteria or were synthesized and were used in structure activity relationship studies. These studies found that the cyclic lactone, methyl group on the β-alanine, and 3-chloro-methoxy-tyrosine were all necessary for activity. The epoxide on the phenyloctenoic acid, however, could be opened to form the chlorohydrin or bromohydrin, but if it was reduced to the diol or double bond activity was lost.[153] The promising bioactivity and wealth of knowledge about the structure led to a synthetic analog, cryptophycin 52 (**90**), which entered clinical trials against non-small cell lung cancer (Figure 21). Toxicity and limited efficacy at safe dosing levels caused **90** to fail a phase II trial, however it was suggested that other dose schedules may improve the prospects of future studies.[158]

Ribosomally Synthesize and Post-translationally modified Peptides

	R₁	R₂
Tenuecyclamide A (**91**)	Ala	Ala
Tenuecyclamide B (**92**)	Ala	Ala
Tenuecyclamide C (**93**)	Gly	Met
Tenuecyclamide D (**94**)	Gly	MetSO
Nostocyclamide A (**95**)	Gly	Val
Nostocyclamide M (**96**)	Gly	Met

Banyascyclamide A (**97**)

Banyascyclamide B (**98**) R= Leu
Banyascyclamide C (**99**) R= Phe
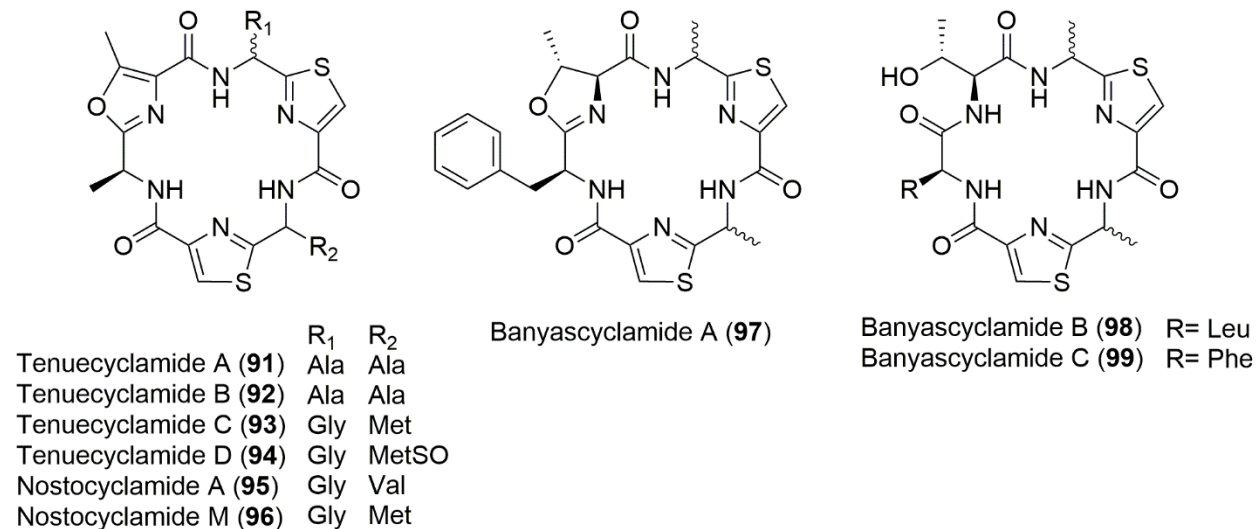
Figure 22: Cyanobactins produced by Nostoc spp.

Most RiPPs that have been identified from *Nostoc* are cyanobactins, small cyclic peptides that typically contain heterocycles in the form of thiazoles, oxazoles, and methyloxazoles.[159] The cyanobactins found in *Nostoc* are tenuecyclamides A-D (**91-94**), nostocyclamides A (**95**) and M (**96**), and banyascyclamides A-C (**97-99**) (Figure 22). Each compound is a cyclic hexapeptide with either two or three heterocycles that are either thiazoles or methyloxazoles.[123,160–162] Interestingly, **93**, **94**, and **96** contain methionine or a methionine sulfoxide, which are not often found in cyanobactins.[160,161] Only the tenuecyclamide BGC has been identified to date (EU290741.1). It contains two alkaline serine proteases, a heterocyclase, and a precursor peptide. The core peptides for both **91-92** and **93-94** are encoded in the precursor peptide.[163] The nostocyclamides have been reported to inhibit the growth of other cyanobactiera, while

the tenuecyclamides and banyascyclamides were reported to have mild antibiotic

activity.[160–162]

Compounds Produced by the Terpene Pathway

Geosmin (**100**)

Germacradienol (**101**)

Comnostin A (**102**) R= CH$_2$OH
Comnostin B (**103**) R= CHO
Comnostin C (**104**) R= COOH

Comnostin D (**105**)

Comnostin E (**106**)

107

Noscomin A (**108**)

Noscomin B (**109**)
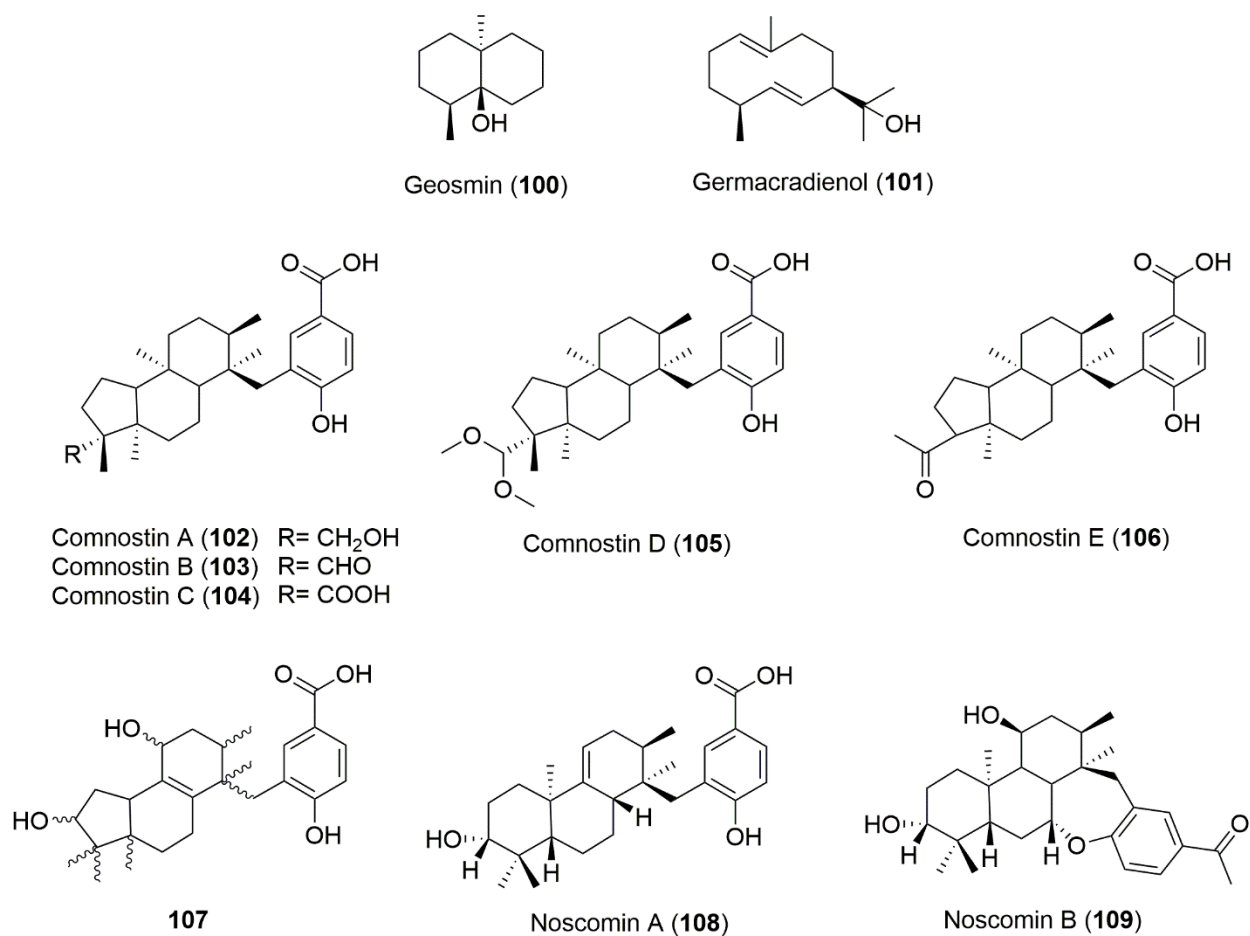
Figure 23: Terpene compounds produced by Nostoc spp.

Several terpenes have been reported from *Nostoc* including sesquiterpenes and diterpenes. *Nostoc* and other cyanobacteria are known to produce the "earthy smelling" sesquiterpene compounds geosmin (**100**) and germacradienol (**101**) (Figure 23). Two publications simultaneously reported the presence of the gene encoding the terpene cyclase enzyme responsible for production of **100**, in *Nostoc* (FJ010203.1).[164,165] Like all terpenes, **100** is formed from five carbon isoprene units that are linked together to form isoprene diphosphate chains.[166] The single terpene cyclase performs both the cyclization of the sesquiterpene chain, as well as the enzymatic transformation of **101** to **100** and acetate.[164,165] Diterpenes have also been isolated from *Nostoc* strains. The diterpenes noscomin A (**108**) and B (**109**) and the comnostins A-E (**102**-**106**) along with an unnamed diterpene (**107**) from the Antarctic strain *Nostoc* sp. CCC 537, are all tricyclic terpenes with a 4-hydroxy benzoic acid moiety bound to the C ring (Figure 23).[167–169] The compounds differ in level and placement of unsaturated bonds within the rings as well as hydroxylation pattern. The comnostins contain a five-membered ring as the A ring while noscomins have six-membered rings as the A ring.[167–169] Noscomin B (**109**) contains an ether bond between the B ring and the hydroxy group of the 4-hydroxy benzoic acid, forming a fourth ring in the diterpene core.[170] No BGC has been identified for these diterpenes although it is likely similar to other diterpene cyclases, containing one cyclase enzyme and tailoring enzymes. All diterpenes isolated from *Nostoc* have been shown to have broad spectrum antibiotic activity.[167–170]

Compounds Produced by the Shikimate Pathway
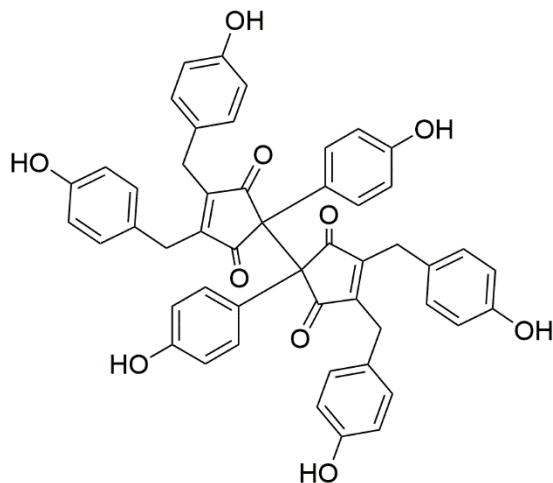
Nostotrebin 6 (**110**)

Figure 24: Structure of nostotrebin 6

Shikimate derived compounds are produced through the same biosynthetic pathway that creates aromatic amino acids, and as such, the compounds produced are highly aromatic. Nostotrebin 6 (**110**) is a unique 2,2-bis-cyclopentenedione core with two 4-hydroxybenzyl groups on each cyclopentene ring (Figure 24).[171] Due to the structural similarities to the amino acid tyrosine it is likely that **110** is produced through the shikimate pathway, although no BGC has been identified. Nostotrebin 6 was initially identified as a cholinesterase inhibitor with modest, low micromolar activity *in vitro* against both acetylcholinesterase and butylcholinesterase.[171] Other studies showed that **110** had antibacterial activity against Gram-positive and antibiotic resistant Gram-positive *Staphylococcus* and *Enterococcus* strains.[172] The unique poly-phenol structure of **110** also has implications for the development of permselective layers applied to

electrodes. Hrbac et al. were able to use electropolymerization to form uniform polymer layers of **110** on both carbon and gold-plated electrode probes. Electrodes are used in the detection of neurotransmitters and the permeability of the nostotrebin 6 layer was found to be selective for dopamine. The nostotrebin 6 layered probe was used to identify cationic dopamine in the presence of interfering neutral and anionic compounds, by selectively allowing only the cationic compounds pass through to the probe.[173] This demonstrates a unique use of a biologically active natural product in electrochemical sensing.
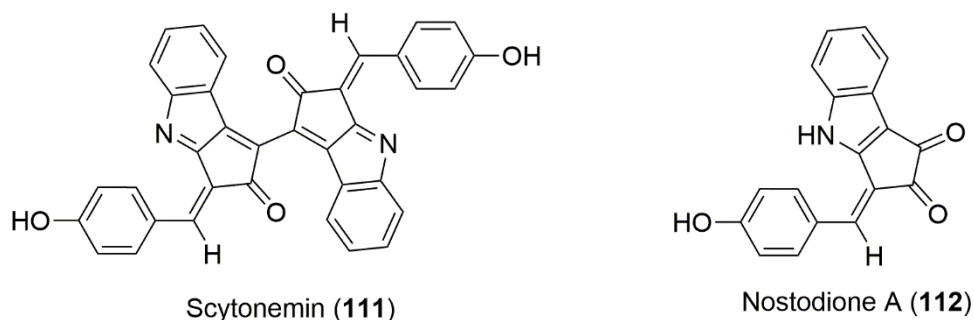
Figure 25: Structures of scytonemin and nostodione A

Scytonemin (**111**) is a shikimate pathway derived compound that acts as a cyanobacterial sunscreen (Figure 25). The compound is produced in response to UV-A radiation exposure and absorbs UV light in the range of 360 – 385nm. By concentrating **111** within the sheath of filamentous bacteria, the cyanobacterium can protect itself from DNA damaging UV light in harsh environments such as the open ocean, soil crusts, and deserts.[174,175] Scytonemin had been known as a cyanobacterial pigment for over one hundred years, however the structure was only solved about twenty-five years ago.[176] Approximately ten years ago a BGC was identified (CP001037.1), which initiated a series of studies into the biosynthesis of this UV absorbing sunscreen compound.[177] Scytonemin was confirmed to be produced by joining of tryptophan and 4-hydroxyphenylpyruvic acid, both products of the shikimate pathway.[178] The gene cluster consists of six genes, *scyA*-*scyF*, that are thought to encode enzymes responsible for biosynthesis of scytonemin, along with several enzymes responsible for aromatic amino acid biosynthesis.[179] ScyB is responsible for converting tryptophan to 3-indole pyruvic acid, which is then coupled to 4-hydroxyphenylpyruvic acid by ScyA.[178,180] Cyclization of the resulting β-ketoacid occurs through ScyC which creates the tricyclic indole that appears to be one half of the scytonemin structure.[180] It is still unclear which enzymes are responsible for the subsequent oxidation and dimerization of the tricyclic monomer. Ferreira et al. showed that *scyD* is not essential for scytonemin production, and is likely a pseudogene, while *scyE* and *scyF* gene knock-outs lead to no production and decreased production of **111**, respectively.[181] However, ScyE does not appear to produce an oxidized intermediate such as nostodione A (**112**) which led the group to suggest it is not immediately responsible for the oxidation or

dimerization.[181,182] Studies are still being carried out to identify the remaining steps in the biosynthesis of this unique compound.

Prospects for *Nostoc* Natural Products

The cryptophycins have proven to be the most promising pharmaceutical compounds produced by *Nostoc* sp. to date. Initially identified by the Merck Company in 1990, the compounds were highly cytotoxic and subsequent studies identified a diverse set of chemical analogs.[153,183,184] The cryptophycins cause cytotoxicity through binding to tubulin and inhibiting the depolymerization of microtubules.[156,185] Interestingly, the cryptophycins were found to be less susceptible to p-glycoprotein efflux pumps, a common drug resistance mechanism seen in cancers, than other anti-microtubule treatments such as vincristine and vinblastine. This made **89** a promising candidate as a new anticancer drug for increasingly resistant cancers.[154] Naturally occurring analogs and synthesis of the compound led to extensive structure activity relationship studies and a greater understanding of the precise mechanism of action.[153,184,186] Cryptophycin 52, a synthetic analog, was developed by Eli Lily and Company and moved forward into clinical trials for the treatment of non-small cell lung cancer. Unfortunately **90** showed peripheral neurotoxicity during the course of the trial and the trial was stopped.[158] The cryptophycins are still being pursued as anticancer drugs in a few studies using antibody drug conjugates to attempt to limit the damaging side effects.[184] Antibody drug conjugates were effective in obtaining approval for brentuximab vedotin after the cyanobacterial metabolites failed phase II clinical trials, and it is likely that a similar approach could be used with the cryptophycins.

Rationalization of the Study

The long evolutionary history of cyanobacteria has allowed them to become taxonomically diverse and adapt to a wide range of habitats.[62] While adapting to many of these habitats, the development of secondary metabolites to provide a competitive edge over competitors would seem likely. Indeed, cyanobacteria have been found to be a reliable source of new secondary metabolites with a wide range of biological activities.[59,78,84] Many of these secondary metabolites have activities that could make them useful as pharmaceuticals or molecular probes. Marine cyanobacteria have been studied extensively; however, freshwater and terrestrial cyanobacteria have been gaining more attention and seem to be just as prolific producers of natural products as their marine counterparts.[78,187,188] *Nostoc* is a widespread, freshwater and terrestrial cyanobacterial genus that has a unique biological life cycle. It also has the ability to enter symbiosis with many organisms in different taxonomic kingdoms.[76,91,98] Chemically diverse secondary metabolites have been identified from *Nostoc*, including the cryptophycins and scytonemin which show promise as an anticancer drug and a UV absorbing pigment, respectively.[115,154,176] Genomic studies and a steady reporting of new compounds produced by *Nostoc*, suggest that there are more secondary metabolites yet to be discovered from this prolific genus.[78,84]

We hypothesize that new secondary metabolites can be discovered from the genus *Nostoc* by utilizing both traditional phenotypic screening and modern genome mining approaches.

The UIC culture collection contains over 1000 mostly freshwater and terrestrial cyanobacterial strains, 200 of which are preliminarily labeled as *Nostoc*.[189] The

taxonomic identification of cyanobacteria from field collections using 16S amplicon sequencing, along with the use of nitrogen free media, will allow selection of nitrogen fixing cyanobacteria, such as *Nostoc*. *Nostoc* strains can be evaluated for their secondary metabolite producing capabilities using both traditional phenotypic screening approaches, as well as genome mining approaches. Utilizing both approaches to natural product discovery will allow for a more comprehensive survey of the secondary metabolites produced by *Nostoc*. *Nostoc* strains UIC 10110, UIC 10250, UIC 10366, UIC 10448, and UIC 10534 were all chosen for further investigation based on their activity in an antiproliferative screen. Strain UIC 10630 was chosen for investigation by genome mining, based on metabolomic and genomic prioritization. The following chapters detail the identification and addition of new *Nostoc* strains to the UIC library, the isolation and structure elucidation of new bioactive secondary metabolites from cultured *Nostoc* spp., and a method to link biosynthetic gene clusters to their produced compounds in cyanobacteria.

2 EXPLORING CYANOBACTERIAL DIVERSITY IN ENVIRONMENTAL COLLECTIONS

Cyanobacterial specific primers, database, and bioinformatic pipeline were developed

by George Chlipala, sequencing was performed by DNAS at UIC.

Identification of Environmental Cyanobacteria for Inclusion in a Culture Collection

The UIC cyanobacteria culture collection contains over one thousand cyanobacterial strains and represents a considerable resource for drug discovery. These strains were added to the culture collection by identifying and collecting macroscopic colonies in the environment and isolating individual strains from these collections.[190,191] This practice can be inherently biased, as only strains that can grow to sufficient densities in the environment will be identified and selected for inclusion in the culture collection. Macroscopic colonies can also exist as cyanobacterial assemblages and may contain multiple different taxa, some of which may not be isolated and added to the collection. Furthermore, isolating cyanobacteria solely from macroscopic colonies neglects the wealth of cyanobacteria isolated from other environments such as soil crusts, symbionts, and other underexplored environments.[95,98,151,192,193] A 16S rRNA amplicon sequencing method was developed to explore the diversity of cyanobacteria in our macroscopic collections and to identify other sources of cyanobacteria for collection and isolation.

16S rRNA amplicon sequencing has become a standard method for describing the taxonomic distribution and diversity of bacteria and archaea in the environment.[194] Polymerase chain reaction (PCR) is used to amplify a portion of the 16S rRNA gene from environmental genomic DNA, and the resulting amplicons are sequenced by next-generation sequencing. The sequenced reads are clustered based on sequence similarity and annotated by comparison to a 16S rRNA reference database.[194,195] This method allows for taxonomic identification of cyanobacterial sequences from

macroscopic collections as well as other environments. However, using standard bacterial PCR primers for amplicon sequencing for these purposes would present several challenges. A universal 16S rRNA primer set is typically used for amplification, to access the greatest diversity of bacteria present in a sample.[196] However, cyanobacteria grow in assemblages with many epiphytic bacteria, which would also be amplified by a universal 16S rRNA primer set. These non-cyanobacterial amplicons would compete with the cyanobacterial amplicons for sequencing and decrease the effective sequencing coverage for the cyanobacterial amplicons. This challenge could in theory be overcome by achieving higher coverage of each amplicon through deep sequencing techniques. However, in samples with low abundance cyanobacteria it is impossible to initially know how deep to sequence to obtain optimum coverage.

Another challenge is the annotation of cyanobacterial sequences. Cyanobacteria are a diverse group of bacteria with a complicated taxonomy due to a long evolutionary history. New taxa are regularly described, and reference strains are often re-assigned to new taxa.[63,64] An accurate and modern reference database is required for correct annotation of cyanobacterial sequences. To meet both these challenges, a cyanobacterial specific primer set and a curated cyanobacterial 16S rRNA reference database were developed (Figure 26).
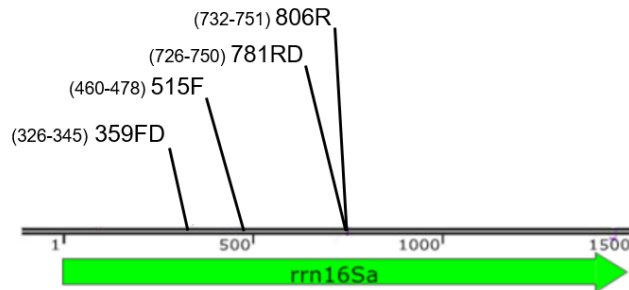
Figure 26: Annotated map of rRNA enconding region with annealing sites of universal primers 515F and 806R and cyanobacterial specific primers 359FD and 781RD

The cyanobacterial specific primer set would selectively amplify cyanobacterial sequences, resulting in an increased relative abundance of cyanobacterial amplicons in a sample. This increases the effective sequencing capacity of cyanobacterial amplicons and provide more accurate relative abundance values for cyanobacterial taxa. Using these primers to analyze our collections would provide a better understanding of the taxonomic distribution within macroscopic cyanobacterial collections and allow identification of low abundance cyanobacteria in other environments.

Primer Verification with Macroscopic Collections

Figure 27: Example of a macroscopic collection.

Macroscopic collections were analyzed to confirm the specificity of the cyanobacterial specific primer set for cyanobacterial sequences and to identify any cyanobacterial taxonomic biases caused by this primer set. Each macroscopic collection had visible cyanobacterial growth that would be easily detectable by either universal or specific primer sets (Figure 27). This allowed confirmation that the specific primer set enriched cyanobacterial amplicons without significantly skewing the detected relative abundances of the different cyanobacterial taxa present. Genomic DNA from each of the collections was amplified with the universal primer set 515F-806R and the cyanobacterial specific primer set 359FD-781RD. The amplicons were sequenced using an Illumina MiSeq and annotated with an inhouse curated reference database. To control for the differences in size between the two sets of amplicons, the

cyanobacterial specific amplicon sequences were trimmed to the same length as the universal primer amplicon sequences and annotated again. The cyanobacterial specific primer set was able to increase the raw cyanobacterial sequence counts and increase the relative abundance of cyanobacterial sequences in each of the macroscopic collections (Figure 28). This demonstrated the specific primer set was selective for cyanobacterial template DNA over other bacterial sequences. However, since the cyanobacterial specific primer set was designed based on known 16S rRNA sequences it is possible that this primer set could skew the taxonomic diversity by more readily amplifying some taxa over others.
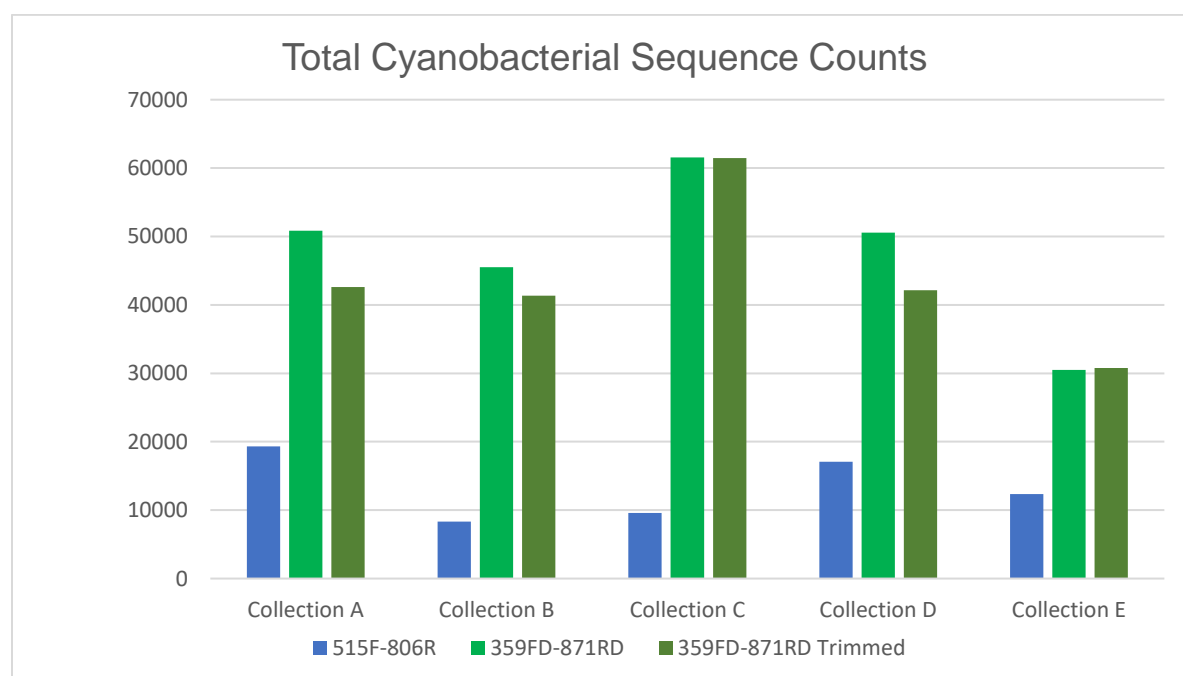
Figure 28: Comparison of cyanobacterial sequence selectivity in the universal primer set 515F-806R and the cyanobacterial specific primer set 359FD-781RD for collections with macroscopic cyanobacterial colonies

To determine whether the specific primer set was skewing the observed cyanobacterial taxa, Bray-Curtiss dissimilarity indices were calculated to compare the overall cyanobacterial community detected using the universal primer set to the specific primer set at both the trimmed and untrimmed lengths. Neither the change to the cyanobacterial specific primer nor the extended length of the cyanobacterial specific amplicons significantly affected the observed cyanobacterial community in the macroscopic collections (Figure 29 Collections A-D). This suggests that the cyanobacterial specific primer set can selectively amplify cyanobacterial template DNA from environmental samples without altering the cyanobacterial community as compared to a universal primer set.
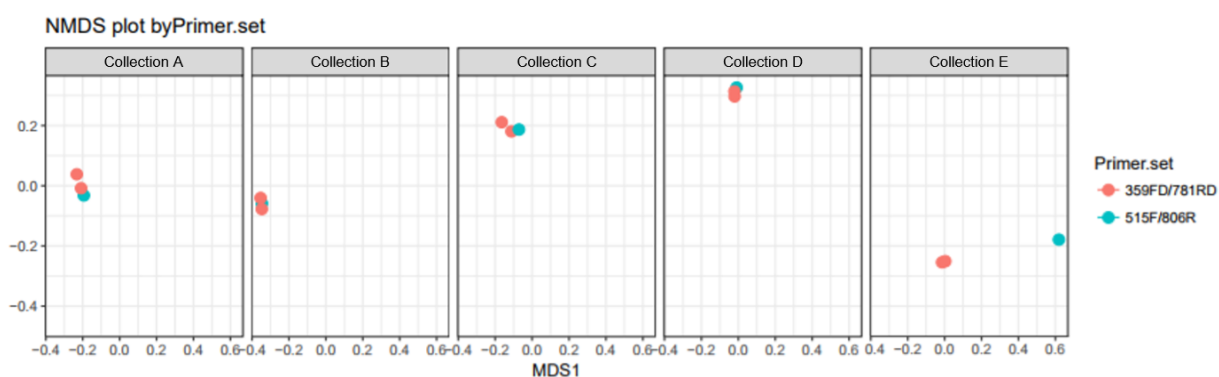
Figure 29: Bray-Curtiss plots displaying the dissimilarity in the beta diversity in each macroscopic collection amplified with the universal primer set 515F-806R and cyanobacterial specific primer set 359FD-781RD

Another common challenge to analyzing environmental samples for cyanobacterial 16S sequences is the abundance of chloroplastic DNA in many collections. Cyanobacteria are evolutionarily related to chloroplasts present in plants and algal cells, and the 16S sequences of chloroplastic DNA form a closely related clade to cyanobacterial 16S sequences.[197] To assess the specificity of the primers for cyanobacterial 16S sequences rather than chloroplastic sequences, a green algae collection (Figure 28 and 29 Collection E) was analyzed using both the universal primer set and the cyanobacterial specific primer set. The universal primer set amplified almost entirely chloroplastic 16S sequences with 98.9% of the sequences being annotated as chloroplasts. In contrast, the cyanobacterial specific primer set was able to selectively amplify the low abundance cyanobacterial amplicons that were overwhelmed by the chloroplastic amplicons produced by the universal primer set. The amplicons produced by the specific primers were annotated as being almost 80% cyanobacteria as opposed to the 1.1% identified with the universal primer set (Figure 30). This demonstrates that the cyanobacterial specific primer set is both selective for cyanobacterial sequences and sensitive enough to identify low abundance cyanobacteria in a sample dominated by closely related sequences. This result prompted us to analyze other environmental samples with low abundance cyanobacteria.
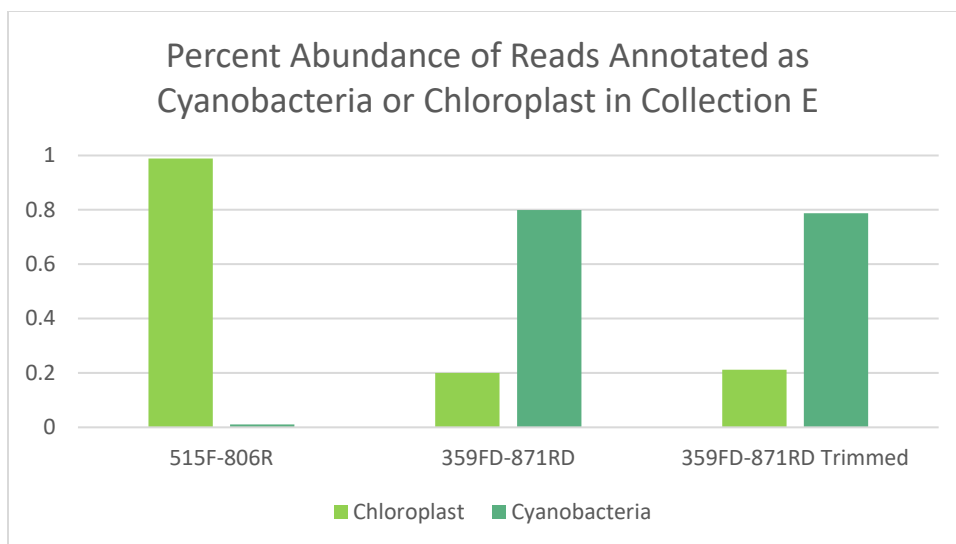
Figure 30: The cyanobacterial specific primer set selectively amplifes cyanobacterial 16S rRNA sequences over chloroplastic 16S rRNA sequences

Analysis of Low-Abundance Cyanobacteria in Environmental Samples

Figure 31: Example of soil collection with no macroscopic cyanobacterial growth

The UIC cyanobacteria culture collection mainly contains strains that have been isolated from macroscopic visible collections. Cyanobacteria are also known to be present in soil, therefore the cyanobacterial specific primer set was used to identify cyanobacteria in collections with no visible macroscopic colonies (soil collections) (Figure 31). Each soil sample was analyzed with both the universal primer set and the cyanobacterial specific primer set. Both primer sets were able to identify cyanobacterial sequences in each of the soil collections. The specific primer set, however, was able to selectively amplify the low abundance cyanobacterial sequences leading to both higher total sequence counts and higher relative abundance values (Figure 32).
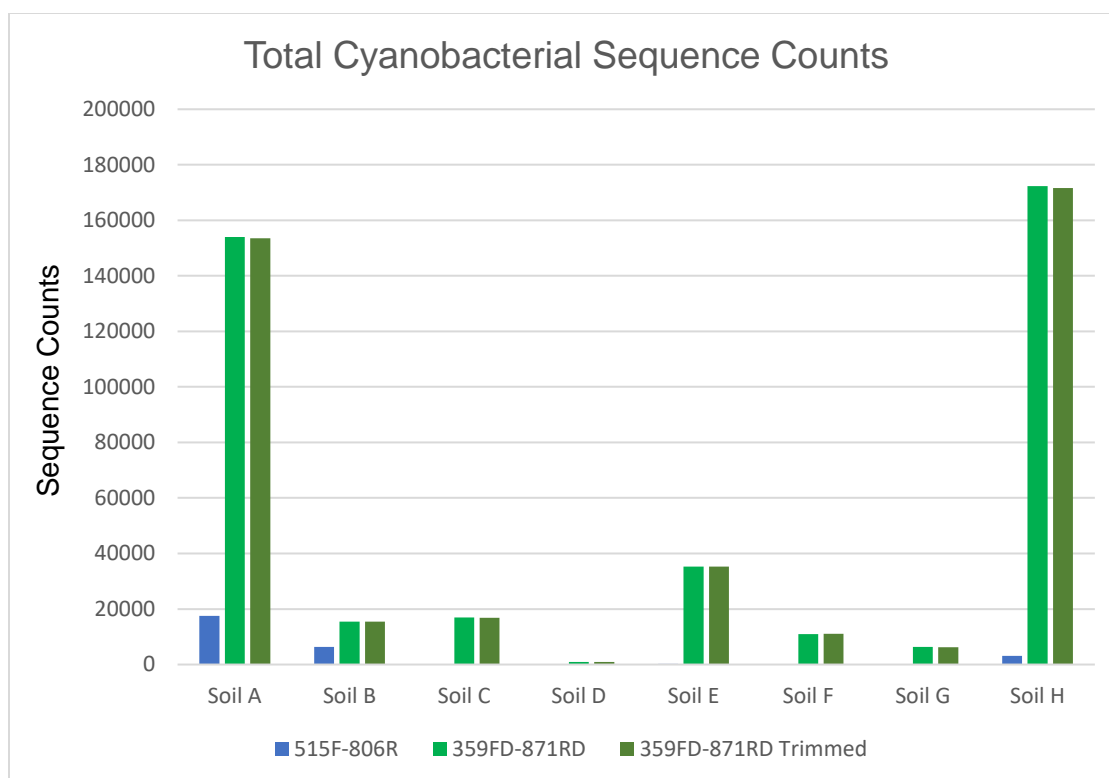
Figure 32: Comparison of total sequence counts amplified by both the universal primer set 515F-806R and the cyanobacterial specific primer set 359FD-871RD for soil collections with no macroscopic cyanobacterial colonies

Bray-Curtiss dissimilarity indices were calculated to assess the consistency of the observed cyanobacterial community between the two primer sets. Initial analysis of the Bray-Curtiss dissimilarity indices revealed slight changes in the observed cyanobacterial community present in three of the soil collections (Figure 33). The difference observed in Soil G and Soil D can be explained by the very low sequence counts for cyanobacteria when using the universal primer set. The universal primer amplicons had 55 reads and 32 reads annotated as cyanobacteria for Soil G and Soil D respectively, while the cyanobacterial specific amplicons had 7,839 reads and 2,541

reads annotated as cyanobacteria, respectively (Figure 32). The increase in the

number of reads likely resulted in more accurate relative abundance values and allowed

detection of lower abundance amplicons which led to a change in the diversity indices.

The differences observed in Soil B appear to be caused by an increase in the

abundance of the genus *Hapalosiphon* by the specific primer set. However, it is unlikely

that the increase occurred due to a preferential amplification of this genus by the

cyanobacterial specific primer set, as other soil collections containing the *Hapalosiphon*

genus did not see corresponding increases. Any primer set will have inherent biases,

but most of the soil collections analyzed in this study have been shown to have

consistent observed cyanobacterial communities regardless of the primer used.[198] This

demonstrates the selectivity and sensitivity of the cyanobacterial specific primer set

when studying cyanobacteria in low abundance environments.
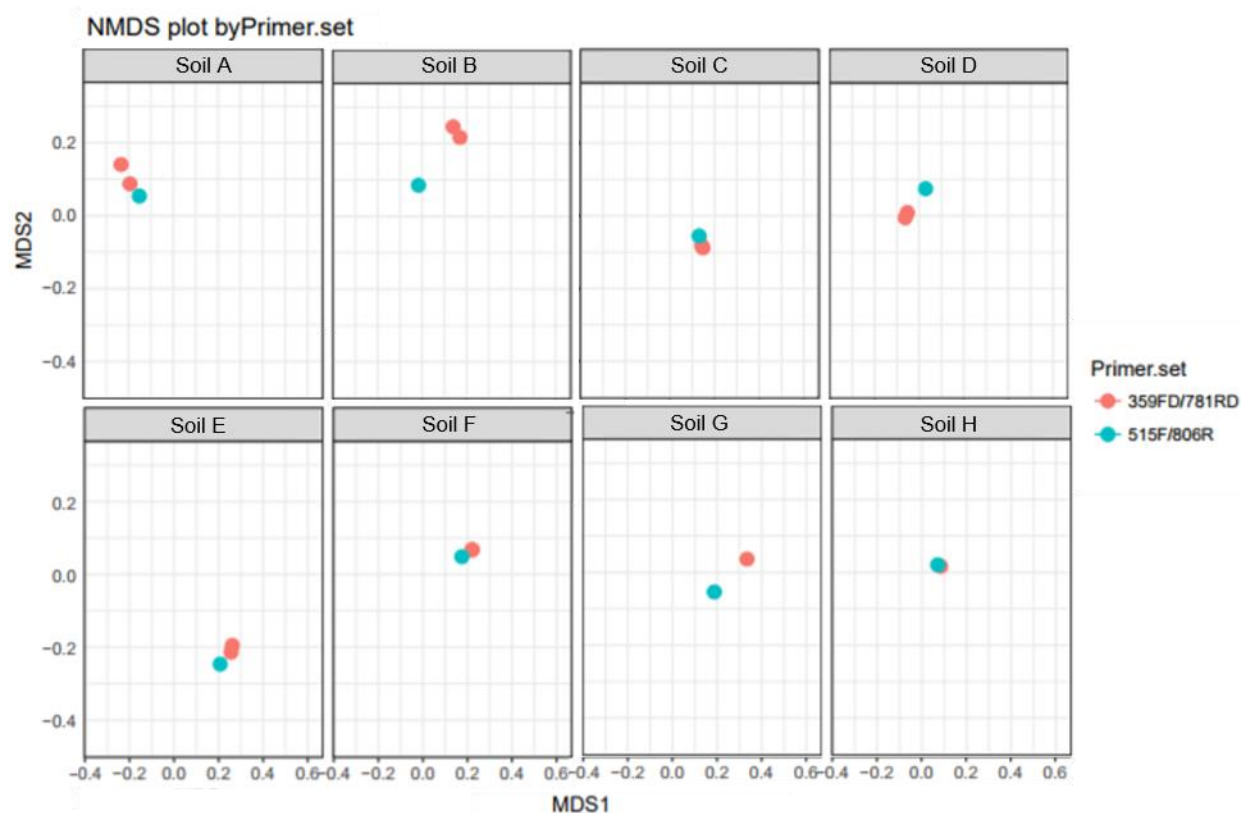
Figure 33:Bray-Curtiss plots displaying the dissimilarity in the beta diversity in each soil collection amplified with the universal primer set 515F-806R and cyanobacterial specific primer set 359FD-781RD

Analysis of the Bray-Curtiss dissimilarity indices and amplicon sequencing data demonstrated the wealth of cyanobacterial diversity present in the soil collections that are different from the taxa typically encountered in macroscopic collections (Figure 34 and Figure 35). Some of these taxa that are currently missing from the UIC cyanobacteria culture collection have been reported to produce natural products and should be included in the collection to maximize the taxonomic and chemical diversity. Additionally, most of the soil collections analyzed contained *Nostoc* spp., known

producers of biologically active natural products. To access these strains, methods to

obtain cyanobacteria from soil collections were developed.
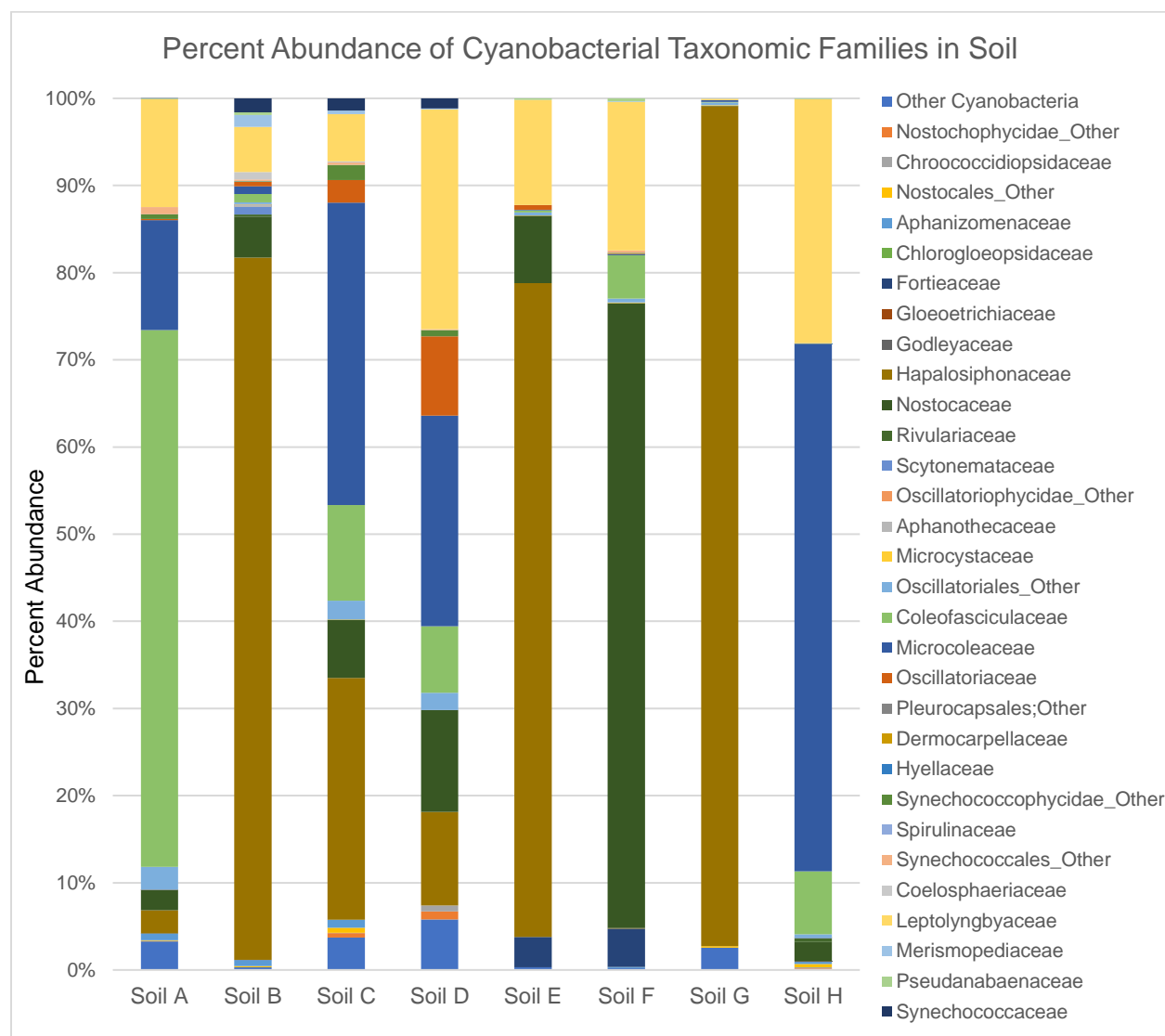
Figure 34: Percent Abundnaces of soil cyanobacterial families demonstrating number of families present in each soil collection. Nostocaceae in dark green, Hapalosiphonaceae in brown, Leptolyngbyaceae in yellow, and Microcoleaceae in blue.
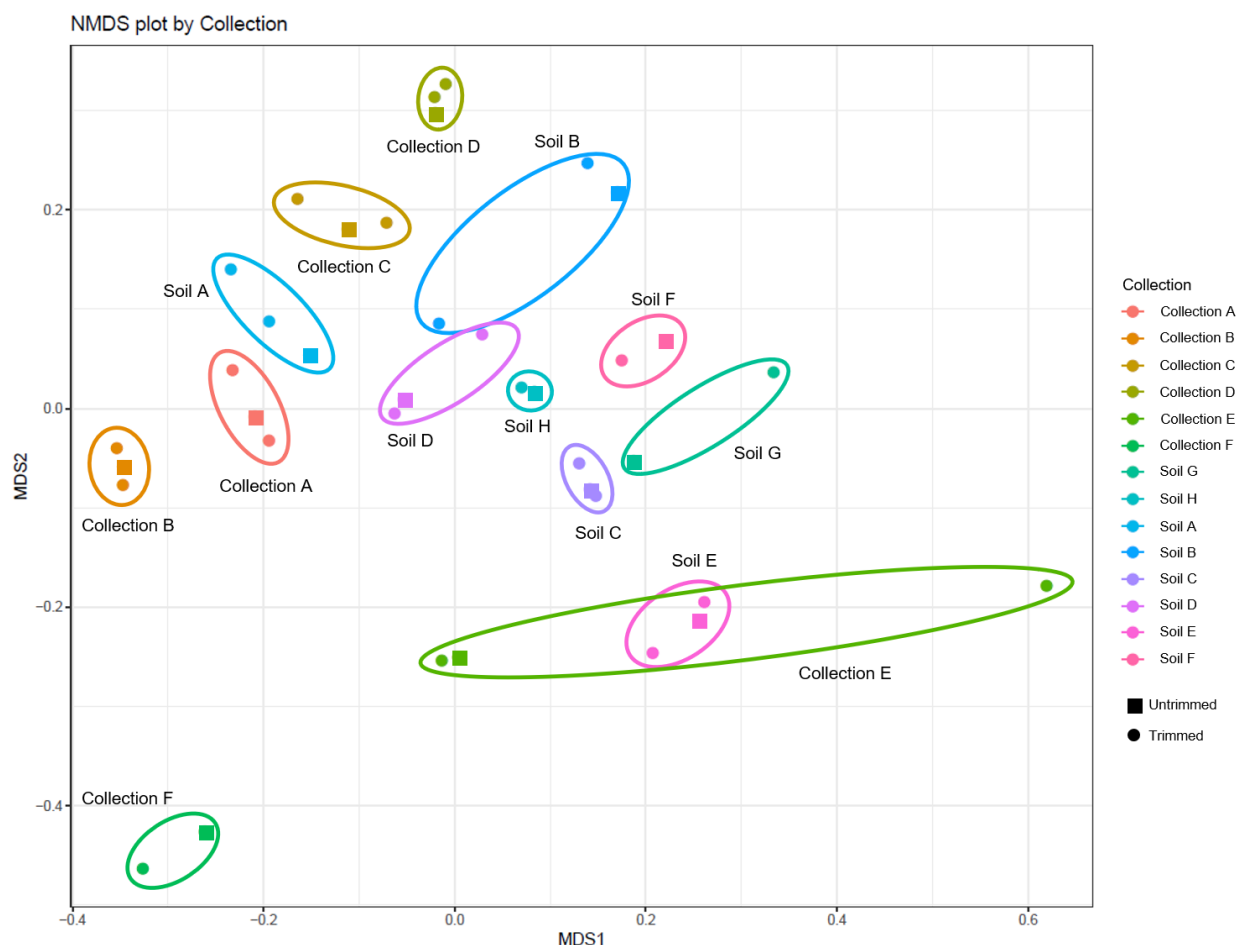
Figure 35: Multi-Dimensional Scaling plots displaying the Bray-Curtiss dissimilarity in the beta diversity in all collections amplified with the universal primer set 515F-806R and cyanobacterial specific primer set 359FD-781RD

Obtaining Cyanobacteria from Soil Collections

The abundance of cyanobacterial taxa in soil samples, identified using the specific primer set, provides an abundant source of cyanobacteria for addition to the UIC cyanobacteria culture collection. Soil samples are easy to collect and store and appear to contain multiple cyanobacterial taxa. Many of the taxa observed in the soil collections are either taxa not typically identified in macroscopic collections, such as

Microcoleus and *Tychonema*, or taxa that are well known to produce natural products, such as *Hapalosiphon* and *Nostoc*.[78,115] Regardless of the interest in these strains, the challenge remained obtaining the low abundant cyanobacteria from the soil collections.

Incubating the soil collections in cyanobacteria specific media under light, produced cyanobacterial growth sufficient to isolate individual strains using micro-isolation methods.[189,199,200] These cyanobacterial growths contained multiple taxa in assemblages; as such, they were termed artificial blooms due to their resemblance to the naturally occurring cyanobacterial blooms (Figure 36). Multiple cyanobacterial strains could be isolated from each artificial bloom; however, it was noticed that certain cyanobacteria, especially members of the Leptolyngbyaceae, grew quickly and could overwhelm the artificial bloom and make isolation of other strains difficult. Changing the media used to generate the blooms to $BG12_o$ media, a growth media without a source of nitrogen, decreased the growth of the Leptolyngbyaceae members and allowed other cyanobacterial genera to grow. Understanding the effect different media have on the growth of certain taxa from soil collections will allow better access to the full diversity of cyanobacteria present in soil.
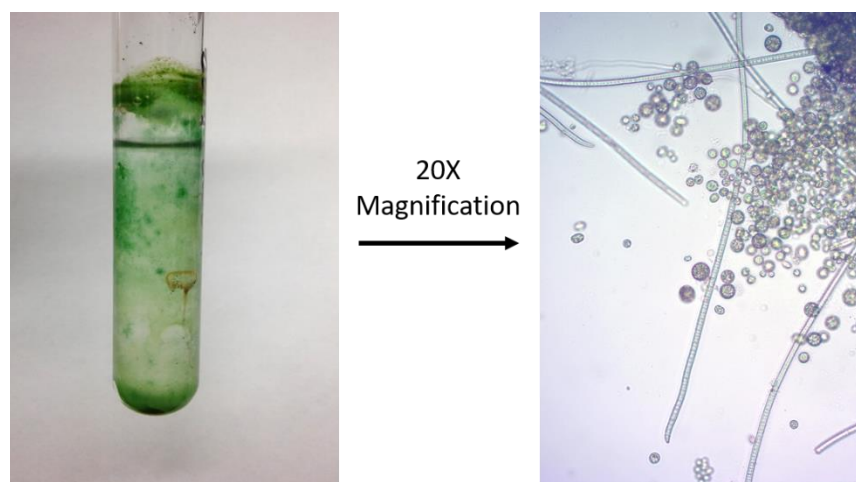
Figure 36: Macroscopic and Microscopic images of artificial bloom Soil I.

Artificial Blooms were created from four soil collections using three different

media, Z, BG12, and BG12$_o$, and were grown, processed, and sequenced in triplicate.

The three media were chosen due to their predominant use for culturing strains in the

UIC cyanobacteria culture collection. The artificial bloom replicates and the original soil

collections were all analyzed by the amplicon sequencing method using the

cyanobacterial specific primer set. The resulting relative abundance values for each soil

sample and each of the corresponding artificial blooms were analyzed in a two-way

ANOVA test to identify statistically significant changes in taxonomic relative abundance

due to media choice. Three of the four soil collections showed statistically significant

changes in the relative abundances of the taxa after five weeks of growth: Soil I, Soil K,

and Soil L. *Leptolyngbya* and *Nodosilinea*, members of the family Leptolyngbyaceae,

were consistently and significantly increased in Z and BG12 media. Leptolyngbyaceae

members are known to grow quickly compared to other cyanobacteria, therefore it is

unsurprising that they are overrepresented in each of our artificial blooms.[64] However, the family is not well known for producing natural products, and the presence of these small filaments complicates isolation of pure strains of more desirable cyanobacteria.[78] *Nostoc*, was consistently and significantly increased in $BG12_o$ media. *Nostoc* is a member of the Nostocaceae family and these well-known natural product producers are often grown in $BG12_o$ in the UIC cyanobacteria culture collection. Nostocaceae members all produce heterocysts, which allow the strains to fix atmospheric nitrogen and grow in the absence of nitrate, thus their enrichment in $BG12_o$ media. Microcoleaceae, a family containing taxa not well represented in the UIC cyanobacteria culture collection, had consistently and significantly decreased relative abundance in each of the three media compared to the original soil collection (Figure 37, Appendix C). Members of the Microcoleaceae are known to be integral in soil crust communities, and it is possible that they require a solid substrate on which to grow, rather than liquid media.[201,202] Based on this small sample size, the artificial blooms appear to enrich only certain taxa that were initially abundant in the soil sample. Many of the taxa not commonly isolated by traditional micro-isolation methods are not enriched in the artificial blooms, likely due to the growth conditions that were selected. Repeating these experiments with solid agar media and other media types might allow for growth of these strains. *Nostoc* strains, which are known for producing natural products, appear to be easily enriched by generating artificial blooms in $BG12_o$ media, providing a reliable source of new *Nostoc* strains for the culture collection and ultimately natural product discovery.

Soil K Artificial Bloom Analysis

Taxon	Z Media	BG12 Media	BG12$_o$ Media
Hapalosiphonaceae	-0.00014	9.4E-05	-1.1E-05
Gloeoetrichiaceae	-4.7E-05	-4.7E-05	-4.7E-05
Nostocaceae	-0.021	-0.01139	0.1436
Rivulariaceae	0.04467	0.01314	0.02653
Scytonemataceae	-0.00024	-9.8E-05	8.04E-05
Tolypothrichaceae	-0.00014	-0.00014	0.004767
Fortieaceae	-4.7E-05	8.45E-06	0.03468
Aphanizomenaceae	-0.00014	-5.4E-05	0.000298
Nostocales	-9.4E-05	2.24E-06	0.00053
Chroococcidiopsidaceae	-0.00019	-9.2E-05	0.004894
Oscillatoriales	-4.7E-05	-4.7E-05	-4.7E-05
Coleofasciculaceae	-0.00108	-0.00104	0.006236
Gomontiellaceae	-4.7E-05	-4.7E-05	-4.7E-05
Microcoleaceae	-0.2969	-0.2962	-0.2885
Dermocarpellaceae	4.62E-06	-5.9E-06	3.87E-05
Aphanothecaceae	-9.9E-05	-0.00014	0.000344
Spirulinaceae	-0.00038	-0.00017	-1.7E-05
Synechococcophycidae	-9.4E-05	-9.4E-05	-9.4E-05
Synechococcales	-0.00105	-0.00095	-0.00032
Chamaesiphonaceae	-4.7E-05	-4.7E-05	-4.7E-05
Coelosphaeriaceae	-4.7E-05	-4.7E-05	-4.7E-05
Leptolyngbyaceae	0.3255	0.3489	0.000117
Merismopediaceae	-0.00024	-0.00027	0.000225
Pseudanabaenaceae	-0.04856	-0.05171	0.06642
Synechococcaceae	-0.00013	0.000131	-0.00015

Figure 37: Heat map displaying the percent change in the relative abundances of cyanobacterial taxa due to artificial bloom growth in differing media types for soil collection K. The more red a color the greater the percent decrease in relative abundance, the more green a color the greater the percent increase in relative abundance. Bold outlines represent statistically significant changes p < 0.05.

Conclusions

A cyanobacterial specific primer set for next-generation sequencing was developed along with a curated cyanobacterial 16S sequence reference database to explore the diversity of cyanobacterial taxa available in the environment. Analysis of macroscopic cyanobacterial collections allowed a comparison of the specific primer set to a universal primer set. The specific primer set was shown to be selective for

cyanobacterial sequences, and importantly did not excessively skew the observed cyanobacterial taxonomic diversity when compared to the universal primer set. Analysis of a green algae collection further demonstrated the specificity of the specific primer set by selectively amplifying low abundance cyanobacterial 16S sequences in the presence of abundant chloroplastic 16S sequences. The sensitivity of the specific primer set was further explored by analysis of soil collections with no macroscopic cyanobacterial colonies. The specific primer set was able to enrich the low abundance cyanobacterial sequences and increase relative abundance of the cyanobacterial sequences compared to the universal primer set. Interestingly, analysis of the soil collections revealed the presence of both taxa that are not represented in the UIC cyanobacteria culture collection and taxa that are well known to produce natural products.

Artificial blooms were created by incubating the soil collections in cyanobacterial specific media under light to access the cyanobacteria detected in soil collections. Three cyanobacterial specific media were tested for their ability to enrich for certain taxa in the artificial blooms. Z media and BG12 media both enriched for members of the Leptolyngbyaceae, a group of cyanobacteria that are well represented in the culture collection, but not reported to produce many natural products. Incubation of soil collections with each of the three media led to a decrease in the percent abundance of members of the Microcoleaceae. This family is not well represented in the UIC cyanobacteria culture collection but has been reported to produce natural products. To access these cyanobacteria, the experiment should be repeated with solid agar media to attempt to enrich for less frequently isolated cyanobacteria. Incubation of soil with nitrate free $BG12_o$ media led to enrichment of *Nostoc* strains. This finding was

unsurprising due to the ability of *Nostoc* strains to produce heterocysts, which facilitate efficient nitrogen fixation in the absence of available nitrate. Incubation of soil collections with BG12$_o$ media appears to be an ideal way to isolate *Nostoc* strains for inclusion in the cyanobacterial culture collection.

Methods

Environmental Collections

Collection locations and descriptions of macroscopic growth field collections and field collections with no macroscopic growth (soil collections) are detailed in appendix C. DNA extraction of macroscopic growth collections occurred within 24 hours of collection or the collection was placed into RNA later and stored at 10°C until DNA extraction could be performed. DNA extraction of soil collections was performed within 24 hours of collection or stored at 10°C until DNA could be extracted. Each soil collection consisted of topsoil collected from an area of approximately 5 cm^2.

Artificial Bloom Generation

Three mg of each soil collection were incubated in cyanobacterial specific media (Z, BG12, and BG12$_o$) under light for 5 weeks. Formulation of each media can be found in appendix D. After 5 weeks, blooms were transferred to a sterile tube and centrifuged to pellet the total cell mass. Metagenomic DNA was then extracted from the pellet.

DNA Extraction and PCR

Metagenomic DNA was extracted using a Macherey-Nagel NucleoSpin Soil Kit following the manufacturer's protocol. PCR mixtures contained 36.5 µL nuclease free water, 10 µL 5X Phusion HF buffer, 1 µL dNTP mix (10 µM), 0.5 µL of each primer (5 µM), 1 µL metagenomic DNA (100 ng), and 0.5 µL Phusion high-fidelity DNA

polymerase for a final volume of 50 µL. The reactions were performed on a Bio-Rad

C1000 Thermal Cycler with a reaction program of: denaturation for 60 s at 95 °C; 5

cycles of amplification with 10 s at 95 °C, 20 s at 55 °C, and 15 s at 72 °C; then 20

cycles of amplification with 10 s at 95 °C, 20 s at 65 °C, and 15 s at 72 °C; and a final

elongation of 5 min at 72 °C. PCR products were purified using a ThermoFisher

Scientific GeneJet PCR Purification Kit.

Cyanobacterial Specific 16S rRNA Primers for Next-Generation Amplicon Sequencing

PCR of the 16S rRNA gene from metagenomic DNA was performed using the

cyanobacterial specific primer set based on the 341F and 781R primers described by

Nubel et al.[203] A multiple sequence alignment of 137 cyanobacterial 16S rRNA

sequences obtained from Genbank was used to add degeneracy in the 341F and 781R

primers. The list of accession numbers used, and the resulting multiple sequence

alignment used can be found in appendix C. The new cyanobacterial specific primers

named 359FD and 781RD were modified to add the fluidigm CS1 and CS2 adapters,

respectively, to yield the finished primers: 359FD-CS1 (5'-ACA CTG ACG ACA TGG

TTC TAC A-GG GRA WTY TTY CGC AAT GGG-3') and 781RD-CS2 (5'- TAC GGT

AGC AGA GAC TTG GTC T-GA CTA CWG GGG TAY CTA ATY CCW TT-3'). The

primer and fluidigm adapters are separated in the previous sequences by a hyphen.

Next-Generation Sequencing

DNA Services facility at the University of Illinois at Chicago performed library

preparation and sequencing services. The purified PCR products were amplified using

the 359FD-CS1 and 781RD-CS2 primers and sequenced using an Illumina MiSeq with

2 X 350 chemistry.

<u>Sequence Processing</u>

Forward and reverse reads were merged using the software package PEAR.[204] Any ambiguous nucleotides occurring at the ends of reads were trimmed and reads with internal ambiguous nucleotides were discarded. Smith-Watermann alignment was used to identify and trim the primer sequences from the reads. Reads lacking either primer sequence were discarded. Reads were trimmed based on quality score using a modified Mott algorithm with PHRED quality threshold of $p = 0.01$. Resulting sequences that were shorter than 225 bases were discarded. Chimeric sequences were identified using USEARCH with the GreenGenes 13_8 reference sequences.[205,206]

A standard QIIME pipeline was modified to create taxonomic summaries using sub-OUT resolution of the sequence dataset.[207,208] The sequence files were merged with sample information and all sequences were de-replicated to produce a list of unique sequences. All sequences that had an abundance of at least 10 counts were designated seed sequences. USEARCH was used to find the nearest seed sequence for any non-seed sequence with a minimum identity threshold of 98%.[205] Non-seed sequences that matched seed-sequences had their counts merged. If non-seed sequences did not match a seed sequence, they were retained as independent sequences.

Taxonomic annotations for seed sequences and unmatched non-seed sequences were assigned using USEARCH and the modified GreenGenes database described in May DS, Chlipala GE et al., with a minmum similarity threshold of 90%.[205] To improve the depth of annotation, the standard QIIME assignment algorithm was modified to only consider hits at each taxonomic level that had an assigned name. If a

sequence matched a reference to the family taxonomic level, but not to genus and

species, the taxonomic assignment would stop at the family level. A minimum

sequence identity of 97% or 99% was required for a reference database match to be

considered for genus or species level assignment, respectively. All taxonomic

annotations and sequence abundances were merged into a single table.

3 PHENOTYPIC SCREENING LEADS TO DISCOVERY OF NEW [7,7] PARACYCLOPHANES FROM UIC 10110

Antiproliferation assay performed by Wei-Lun Chen, hollow fiber tumor assay performed by Dan Lantvit, DEPTQ experiment acquired by Aleksej Krunic, sequencing was performed by DNAS at UIC

Phenotypic Screening of Strains in the UIC Culture Collection

Cancer is the second leading cause of death in the United States.[209] Phenotypic screening methods have long been used as reliable approaches for identifying anticancer natural products, such as paclitaxel and daunorubicin.[6,8,210,211] The Orjala lab uses a phenotypic screening approach to discover new anticancer natural products from the UIC cyanobacteria culture collection, which contains over 1000 cultured cyanobacterial strains.[189] Each strain is grown for six to eight weeks under light and filtered aeration. The cell mass is harvested by centrifugation and lyophilized. The dried cell mass is extracted via maceration with a 1:1 mix of dichloromethane (DCM) and MeOH. The resulting extracts are fractionated by vacuum liquid chromatography with a step-wise gradient of isopropyl alcohol (IPA) in water over a Diaion HP20 stationary phase to yield six fractions. Fractions are dried, and one mg of each fraction is stored in DMSO at 10 mg/ml as part of a fraction library, which is regularly screened in antiproliferation assays (Figure 38). Phenotypic screening provides high throughput screening capabilities and identifies compounds with immediate biological relevance, but often results in the re-identification of known compounds. Dereplication is an efficient process to identify whether an active compound in a complex mixture, such as an extract or crude fraction, is known or new. Compounds that can not be matched to known compounds are prioritized for isolation and structure elucidation.
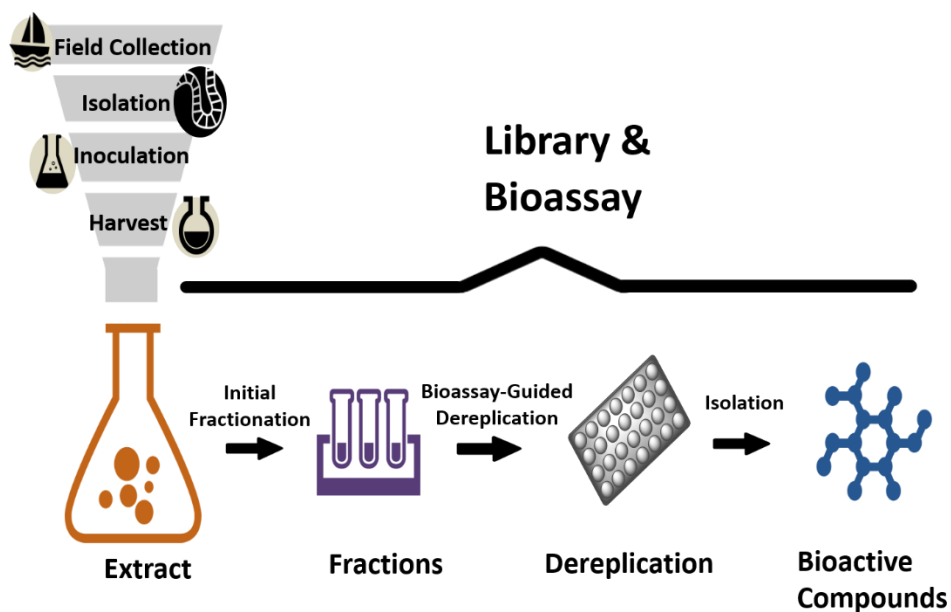
Figure 38: Orjala Lab activity guided natural product discovery workflow

Dereplication of UIC 10110 *Nostoc* sp.

The re-discovery obstacle demonstrates the necessity for a robust dereplication protocol to be combined with phenotypic screening approaches for drug discovery. Many dereplication protocols have been published using UV databases or mass spectrometry techniques that rely on the strength of well-characterized spectroscopic databases.[18] The Orjala lab dereplication protocol, however, uses intrinsic characteristics that can be searched in any publicly available chemical database.[16] The dereplication protocol uses a chromatographic HPLC separation step with a monolithic column and an automatic fraction collector, which together provide a high level of

reproducibility. Monolithic columns allow for large injection loads with minimal back pressure due to the greater connectivity of the pores in the stationary phase. The large injection volumes allow sufficient amounts of an active fraction to be separated to enable biological evaluation as well as ^{1}H NMR spectroscopy and mass spectrometry of the active subfractions. Mass spectrometry is used to obtain accurate mass and molecular formula which can be readily searched in publicly available databases. ^{1}H NMR spectra provide characteristic structural moieties that can be used to manually eliminate potential matches from a molecular formula search (Figure 39). Molecular formulae and characteristic structural moieties searches can be performed in tandem with the dereplication tool available in the CYANOS data management system, which searches a database of published cyanobacterial compounds.[212]
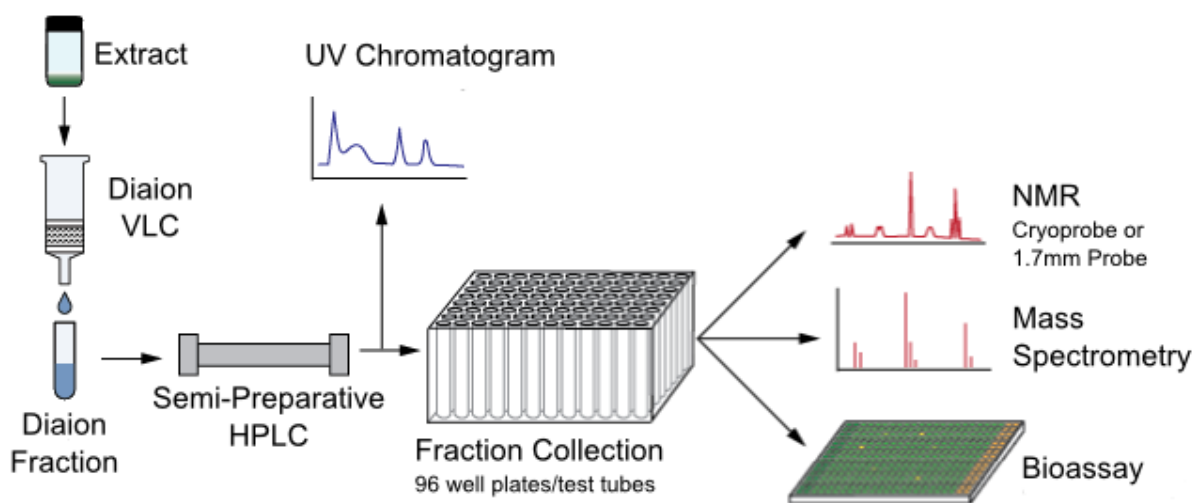
Figure 39: Depiction of the Orjala Lab dereplication protocol

The cell extract of UIC 10110, *Nostoc* sp. was found to be active in an antiproliferation assay against MDA-MB-435, a melanoma cell line. After an initial fractionation step, fraction four eluting with 70% IPA, was found to be the most active in the antiproliferation assay (Figure 40). Fraction four was then assessed using our dereplication method. Five mg of the active fraction was separated by semi-preparative HPLC with a monolithic C18 column, using a gradient of 55% MeOH in water to 100% MeOH in 13 minutes at 4 mL/min. Subfractions were collected at 30 second intervals and dried. Half of each resulting subfraction was submitted for bioassay to determine which peaks in the resulting chromatogram corresponded to the observed activity. Subfractions 13, 14, and 15, and subfraction 20 were the most active in the antiproliferation assay and corresponded with peaks at seven and ten minutes, respectively. These subfractions were analyzed using ^{1}H NMR spectroscopy and mass spectrometry. ^{1}H NMR spectra of the peaks showed subfractions 13, 14, and 15 contained the same compound, while subfraction 20 contained a different but similar compound. The spectra for both compounds contained aromatic signals near δ_H 6.0, methyl signals around δ_H 1.0, and many signals in the region for aliphatic methylene protons. Mass spectrometry was performed using an IT-ToF mass spectrometer, which provided accurate mass that was used to determine the molecular formula. The active component in subfraction 13, 14, and 15 had a molecular weight of 568.416 Da and a molecular formula of $C_{36}H_{56}O_5$, while the compound in subfraction 20 had a molecular weight of 552.417 Da and a molecular formula of $C_{36}H_{56}O_4$. The molecular formula of the compound in subfraction 20, $C_{36}H_{56}O_4$, matched the known compound merocyclophane A, and the ^{1}H NMR spectrum of the subfraction was identical to the

published data.[131] The molecular formula of the compound in subfractions 13, 14, and

15 did not match any compounds in online databases or a literature search. We

therefore determined that the active components of UIC 10110 were merocyclophane A

and a new antiproliferative compound (Figure 40).
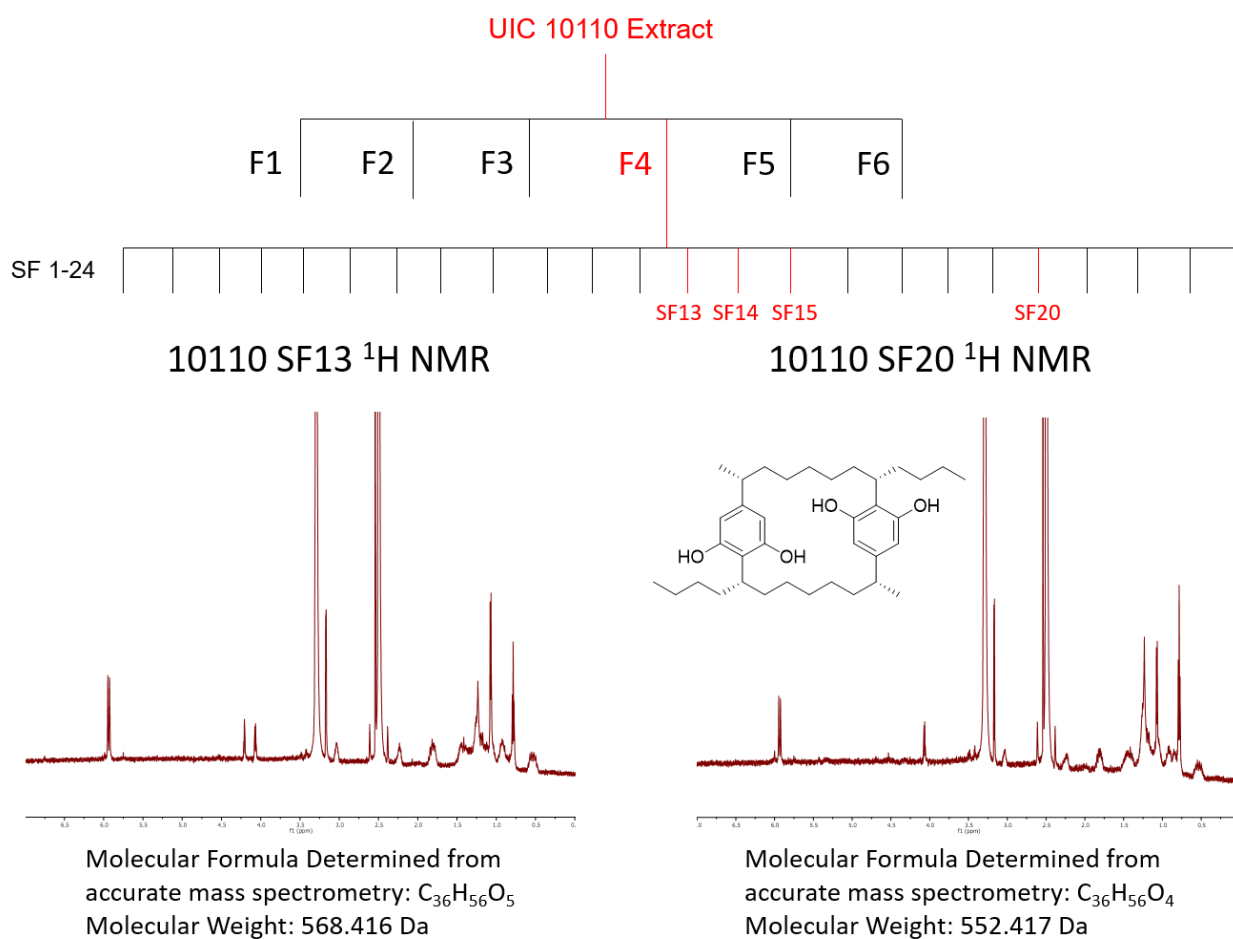
Figure 40: Depiction of the dereplication process for the extract of UIC 10110 and dereplication of merocyclophane A, red lines indicate bioactivity.

Structure Elucidation of New [7,7] paracyclophanes from UIC 10110 Nostoc sp.

The new compound identified by dereplication was isolated via semi-preparative reversed-phase HPLC. A gradient from 65% acetonitrile (ACN) in water to 100% ACN in 35 minutes at 4 mL/min was used to isolate the compounds. The compound eluted at 19 minutes and constituted a substantial portion of the parent fraction. The known merocyclophane A eluted at 27 minutes. A peak eluting at 6 minutes was also isolated and based on ^{1}H NMR and mass spectrometry it was determined to be a new compound with a similar structure to merocyclophane A and the new compound eluting at 19 minutes.

The structures of the two new compounds were solved by ^{1}H NMR and mass spectrometry. The molecular formula of the two new compounds were $C_{36}H_{56}O_5$ and $C_{36}H_{56}O_6$ respectively, while the molecular formula of the previously identified merocyclophane A was $C_{36}H_{56}O_4$. Both compounds had ^{1}H NMR spectra similar to that of merocyclophane A, suggesting the two compounds were analogs of merocyclophane A. The molecular formula of the two compounds suggested the presence of one and two additional hydroxy groups, respectively. The structure of merocyclophane C, eluting at 19 minutes, was determined by the presence of a triplet at δ_H 3.47 in the ^{1}H NMR spectrum that was not present in the ^{1}H NMR spectrum of merocyclophane A. The deshielded signal is indicative of protons adjacent to an oxygen, and the triplet coupling pattern required the oxygen be placed on the end of the aliphatic chain at C-30. The placement of this oxygen atom in merocyclophane C breaks the symmetry observed in merocyclophane A. This can be seen in the ^{1}H NMR spectrum of

merocyclophane C, by the separation of the doublet methyl signals at δ_H 1.14 and 1.15 (Figure 41).
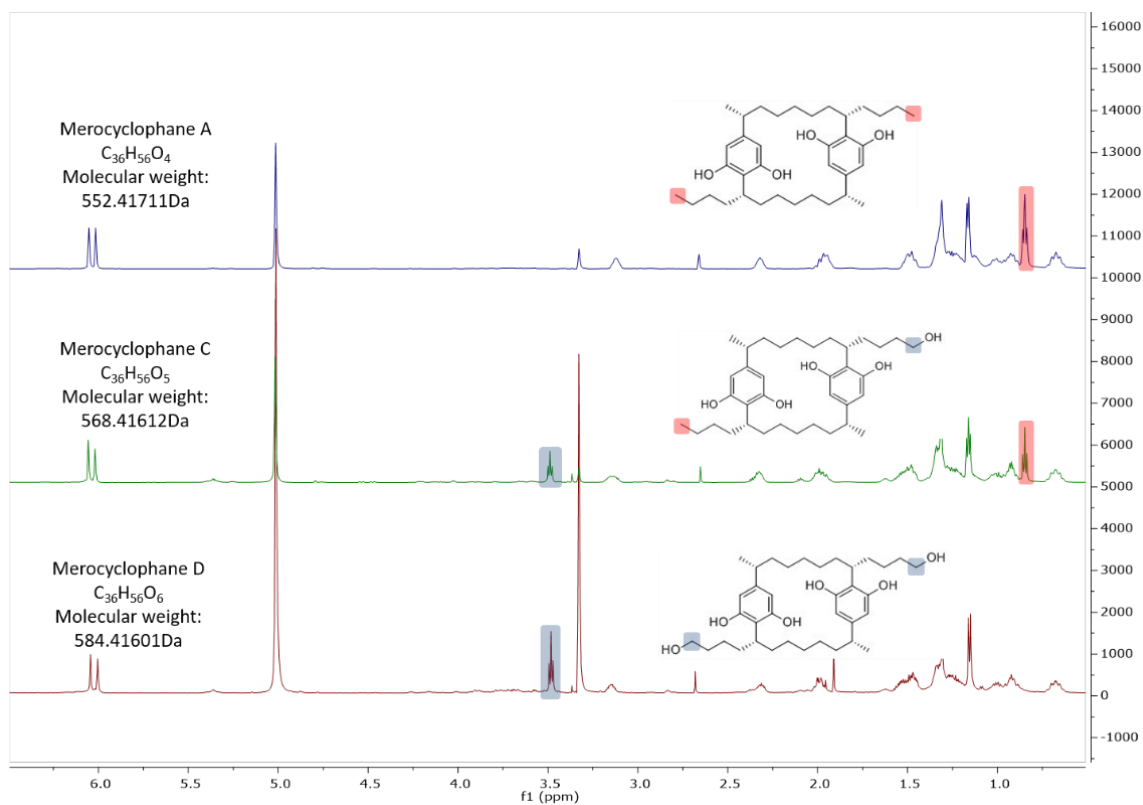
Figure 41: Structure elucidation of the new merocyclophanes C and D, colored boxes on structure correspond to colored boxes on ^1H NMR spectra

The structure of merocyclophane C was confirmed by COSY correlations, which allowed the identification of the aliphatic chain spin system and verified the location of the oxygen atom with correlations from H-29 to H-30 at δ_H 3.47. The compound was confirmed to contain α-methyl groups by COSY correlations from H-35 to H-1 and H-36 to H-14. HMBC correlations from H-27 and H-6 to C-8 and from H-7 to C-9 and C-13 connected one resorcinol to the aliphatic chain, while correlations from H-1 to C-23 and C-25 and from H-25 and H-23 to C-21 connected the other resorcinol to the aliphatic chain. A similar set of HMBC correlations was found on the other half of the molecule, which confirmed the two-dimensional structure (Figure 42 and Table II).

The structure of merocyclophane D was solved similarly. The symmetrical appearance of the ^{1}H NMR spectrum was restored suggesting that the compound had one hydroxy group on the end of each aliphatic chain. The H-30 triplet at δ_H 3.47 integrated to four protons instead of two, as seen in merocyclophane C, and the signals for the methyl groups at δ_H 0.83 were no longer present. The placement of these hydroxy groups was confirmed by COSY correlations from H-29 to H-30 and H-33 to H-34. COSY correlations from H-35 to H-1 and H-36 to H-14 confirmed the branched α-methyl groups and allowed assignment of the aliphatic chains. HMBC correlations similar to those observed in merocyclophane C were used to connect the resorcinol units to the aliphatic chains (Figure 42 and Table II).

Merocyclophane C (**113**)

Merocyclophane D (**114**)

Merocyclophane A (**70**)
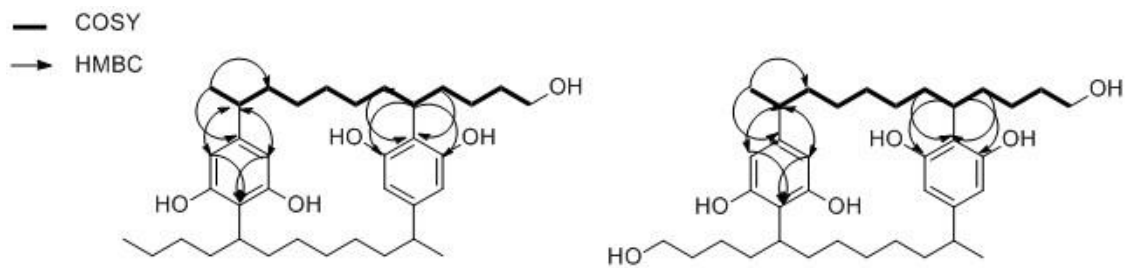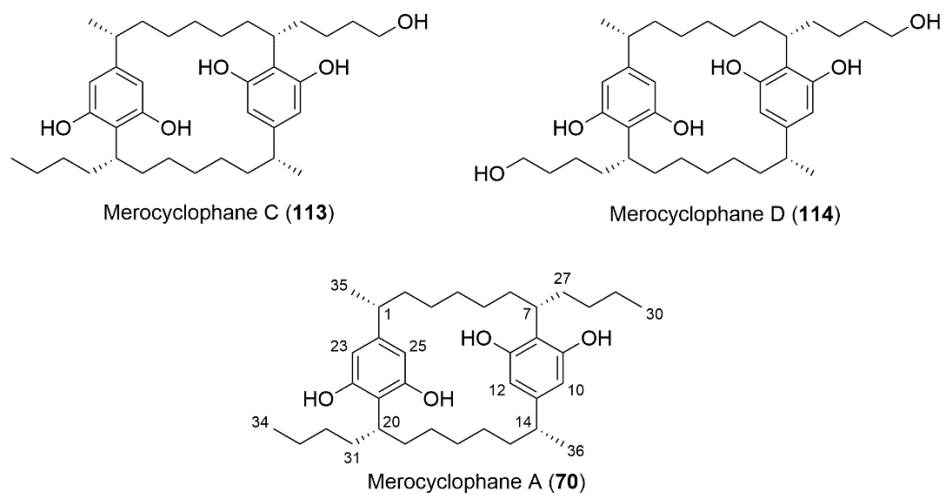
Figure 42: Key 2D NMR correlations of merocyclophane C and D

TABLE II: SPECTROSCOPIC DATA FOR MEROCYCLOPHANES C AND D (^{1}H 900 MHZ, ^{13}C 225 MHZ IN MEOH-D4).

	merocyclophane C (**113**)		Merocyclophane D (**114**)	
Position	δ_C, type	δ_H (*J* in Hz)	δ_C, type	δ_H (*J* in Hz)
1	42.1, CH	2.30, m	42.1, CH	2.3, m
2	40.9, CH$_2$	1.45/1.32, m	40.9, CH$_2$	1.32/1.45, m
3	31.1, CH$_2$	0.91/0.67, m	31.1, CH$_2$	0.93/0.65, m
4	32.8, CH$_2$	1.27/0.88, m	32.8, CH$_2$	1.27/0.93, m
5	30.9, CH$_2$	1.3/0.99, m	30.8, CH$_2$	0.67/0.98, m
6	35.52, CH$_2$	1.31, m	35.5, CH$_2$	1.32/1.98, m
		3.14, dddd (10.1,	37.0, CH	3.14, m
7	37.0, CH	12.1, 5.9, 4.1)		
8	116.4, C		116.1, C	
9	158.78, C		158.8, C	
10	104.7, CH	6.04, s	104.6, CH	6.03, s
11	147.0, C		147.0, C	
12	109.4, CH	6.0, s	109.4, CH	5.99, s
13	157.14, C		157.1, C	
14	42.1, CH	2.30, m	42.1, CH	2.3, m
15	40.9, CH$_2$	1.45/1.32, m	40.9, CH$_2$	1.32/1.45, m
16	31.1, CH$_2$	0.91/0.67, m	31.1, CH$_2$	0.93/0.65, m
17	32.8, CH$_2$	1.27/0.88, m	32.8, CH$_2$	1.27/0.93, m
18	30.8, CH$_2$	1.3, 0.99, m	30.8, CH$_2$	0.67/0.98, m
19	35.5, CH$_2$	1.31, m	35.5, CH$_2$	1.32/1.98, m
		3.14, ddd (10.1,	37.0, CH	3.14, m
20	37.0, CH	12.1, 5.9, 4.1)		
21	116.1, C		116.1, C	
22	158.8, C		158.8, C	
23	104.7, CH	6.04, s	104.6, CH	6.03, s
24	146.9, C		147.0, C	
25	109.4, CH	6.0, s	109.4, CH	5.99, s
26	157.08, C		157.1, C	
27	35.1, CH$_2$	1.96, m	35.1, CH$_2$	1.94, m
28	31.9, CH$_2$	1.19, 1.11 m	25.8, CH$_2$	1.22/1.13, m
29	34.1, CH$_2$	1.53/1.45, m	34.1, CH$_2$	1.52/1.31, m
30	63.5, CH$_2$	3.47, t (6.8)	63.5, CH$_2$	3.47, t (6.8)
31	35.0, CH$_2$	1.92, m	35.1, CH$_2$	1.94, m
32	25.8, CH$_2$	1.24/1.21, m	25.8, CH$_2$	1.22/1.13, m
33	24.2, CH$_2$	1.30/1.22, m	34.1, CH$_2$	1.52/1.31, m
34	14.7, CH$_3$	0.83, t (6.8)	63.5, CH$_2$	3.47, t (6.8)
35	23.8, CH$_3$	1.14, d (7.2)	23.8, CH$_3$	1.14, d (7.2)
36	23.8, CH$_3$	1.15, d (7.2)	23.8, CH$_3$	1.14, d (7.2)

The stereoconfiguration of merocyclophanes C and D was determined using electronic circular dichroism spectroscopy (ECD) and comparing the spectra to that of merocyclophane A. The structure of merocyclophane A was solved using single crystal x-ray crystallography and ECD, which allowed for the assignment of the absolute configuration of the four chiral centers.[131] The ECD spectra of merocyclophanes C and D matched that of merocyclophane A, with negative cotton effects observed at Δε 216 and Δε 278, suggesting that all three merocyclophanes had the same absolute configuration at each of the chiral centers (Appendix A). The stereoconfiguration of the four chiral centers is consistent with all other published [7,7] paracyclophanes.[130,137–140]

Biosynthesis of [7,7] Paracyclophanes

Merocyclophanes belong to a group of bioactive compounds produced by members of the Nostocaceae family, the [7,7] paracyclophanes. These macrocyclic polyketides were first identified by Moore et al. in 1991.[130,132] They identified the cylindrocyclophanes and nostocyclophanes and correctly predicted their biosynthetic origin as polyketides, using a ^{13}C labeled acetate feeding experiment.[213] Additional classes of [7,7] paracyclophanes that have been discovered include the carbamidocyclophanes and merocyclophanes, each with different moieties attached at the alpha position of the resorcinol core and the terminal carbon of the aliphatic chain.[131,138–140] The merocyclophanes differ from other [7,7] paracyclophanes by the presence of a branched α-methyl group, and while the biosynthesis of the branched β-methyl group [7,7] paracyclophanes has been well described, the biosynthetic origin of the branched α-methyl group is unknown.

Biosynthetic gene clusters of both cylindrocyclophane and carbamidocyclophane biosynthesis have been identified by Nakamura et al. and Presitch et al., respectively.[133,134] The complete biosynthesis of the cylindrocyclophane β-methyl containing structure was elucidated in a series of papers by Nakamura et al. (Figure 15).[133,135,136] They found that biosynthesis begins with recruitment of a decanoic acid by CylA and CylB, which is then chlorinated at the C-6 position by a unique halogenase, CylC, in a stereospecific reaction. This halogenated decanoic acid is then loaded onto the first of two type I PKS enzymes. The first type I PKS, CylD, adds one malonate unit, and the chain is then acted on by a group of three enzymes, CylE-CylG, in a mechanism similar to HMG-CoA synthase, to install the β-methyl group. The second type I PKS, CylH, adds a second malonate unit and transfers the compound to a type III PKS enzyme, CylI, which adds two more malonate units and cyclizes the chain. This is then cleaved from the type III PKS to form a halogenated alkyl-resorcinol. The macrocycle is formed by homodimerization of the halogenated alkyl-resorcinol by the alkylating enzyme CylK, using the chlorines as leaving groups (Figure 15).[133,135] Various tailoring enzymes add carbamate units, acetate units, or terminal alkyl halogenation to form the wide chemical diversity seen in the [7,7] paracyclophanes.[134,135,137]
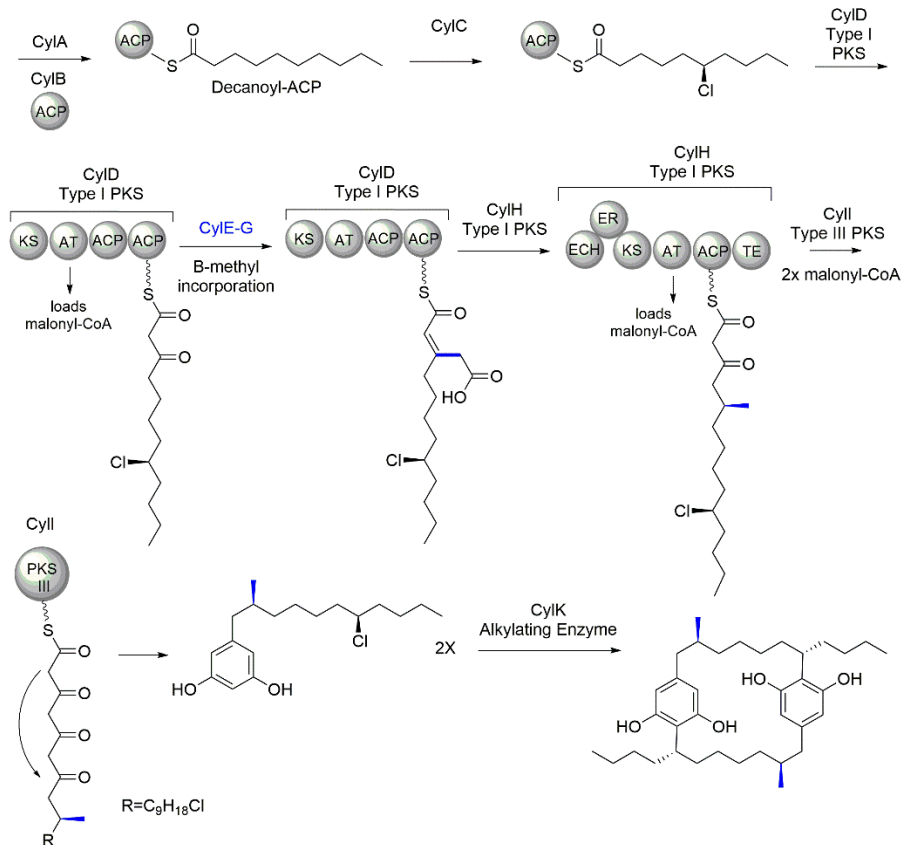
Figure 15: Biosynthesis of cylindrocyclophanes as described by Nakamura et al. 2012 and Nakamura et al. 2017. Enzymes responsible for beta-methyl incorporation are colored blue along with the beta-methyl group on the structure.[133,135]

Identification of the Merocyclophane Gene Cluster and a Proposed Biosynthesis

The Orjala lab is the only group to have reported the isolation of the merocyclophanes. This provided a unique opportunity to identify the biosynthetic gene cluster that produces the merocyclophanes, and specifically to find a biosynthetic source for the α-methyl group, the major structural difference seen in the merocyclophanes.[131] It has been shown in both cylindrocyclophane and carbamidocyclophane biosynthesis that the β-methyl group is installed by a trio of

enzymes, CylE-CylG/CabE-CabG, that act in an HMG-CoA synthase-type mechanism.[133,134] Seeing as the merocyclophanes have an α-methyl group in place of a β-methyl group, but still retain the alkyl-resorcinol macrocycle, it was hypothesized that the merocyclophane gene cluster could be identified by the type I and type III PKS genes, but would lack homologs to CylE-CylG.

Genomic DNA from UIC 10110 was isolated and prepared for sequencing on an Illumina Miseq, using a Nexterra DNA Library Preparation Kit. The resulting sequences were assembled using the SPAdes assembler algorithm and the draft genome was submitted to AntiSMASH.[21,214] The AntiSMASH results identified a BGC of approximately 23kb that was annotated as a type I/type III PKS hybrid. This gene cluster was further evaluated and genes similar to both type I PKS genes, and the type III PKS gene from the cylindrocyclophane BGC were identified. This putative merocyclophane gene cluster also contained genes similar to the decanoic acid recruitment and activation genes *cylA* and *cylB*, the halogenase *cylC*, and the aromatic ring alkylation gene *cylK* (Figure 43 and Table III). BLAST analysis was performed on the genes surrounding the putative gene cluster to define the likely limits of the BGC. Comparison of the individual domains of the first type I PKS in each BGC, *cylD* and *merC*, identified the biosynthetic source of the α-methyl group in the merocyclophanes. As predicted, the putative merocyclophane BGC does not contain homologs for β-methyl installation, *cylE-cylG*, instead it has an extended first type I PKS. The *merC* gene is more than twice as large the corresponding *cylD* and is composed of seven domains, rather than the four domains in *cylD*. The α-methyl group is likely installed by

these extra domains, which consist of a KR, DH, ER, and C-methyl transferase, and would completely reduce the decanoic acid ketone and attach the α-methyl group.

TABLE III: PROTEINS ENCODED IN THE MEROCYCLOPHANE BIOSYNTHETIC GENE CLUSTER INCLUDING LENGTH, BLAST PROPOSED FUNCTIONS, AND PERCENT IDENTITY TO THE CORRESPONDING CYL PROTEINS.

Protein	Length (aa)	Percent Identity to Cyl BGC Proteins	BLASTP Proposed Function Based on Identity Scores
MerA	104	56%	Acyl Carrier Protein
MerB	468	72%	Halogenase
MerC	256	71%	Type I PKS
MerD	1396	59%	Type I PKS
MerE	374	77%	Type III PKS
MerF	199	75%	Methyl Transferase
MerG	460	0%	Cytochrome P450
MerH	921	66%	Alkylating Dimerization Enzyme
MerI	612	61%	Acyl-CoA Synthase
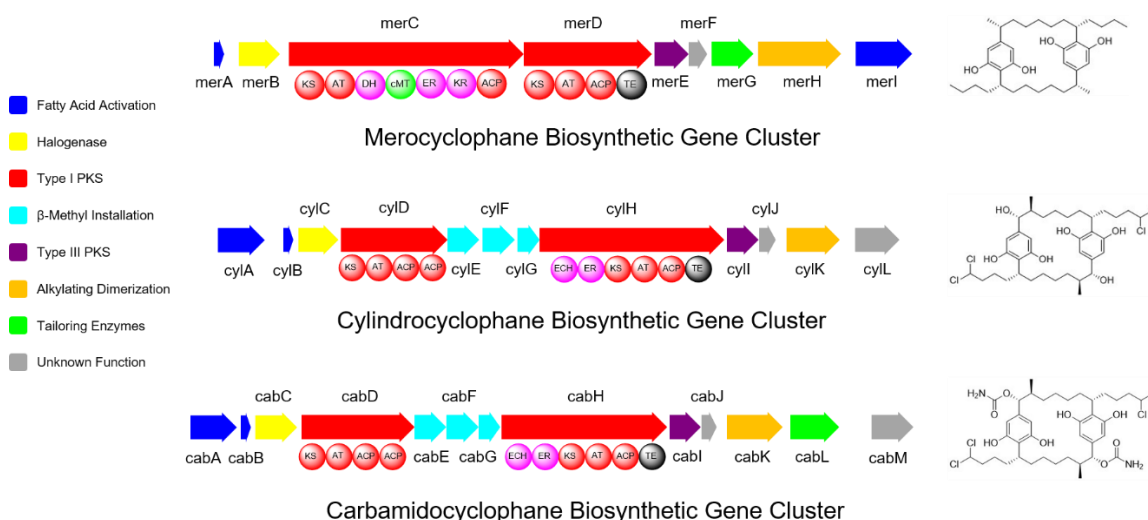
Figure 43: Comparison of the biosynthetic gene cluster organization of the three known [7,7] paracyclophane biosynthetic gene clusters. Similarly colored genes perform similar functions in the biosynthesis.

The putative BGC identified from the genomic DNA of UIC 10110 appeared to use a similar biosynthetic logic to that of cylindrocyclophane biosynthesis (Figure 43). We proposed a mechanism for the merocyclophane biosynthesis based on the well described biosynthesis of the cylindrocyclophanes by Nakamura et al.[133,135] Biosynthesis of the merocyclophanes likely begins with the recruitment and activation of decanoic acid by MerI and MerA. The activated decanoic acid is chlorinated at the C-6 position by the enantiospecific halogenase MerB. This halogenated decanoic acid is then transferred to the first type I PKS, MerC, which adds one malonate unit. The previous ketone is acted on by the KR, DH, ER, and C-methyl transferase domains to fully reduce the carbonyl and add the α-methyl group. The chain is moved to the second type I PKS, MerD, which adds a second malonate unit and transfers the growing

chain to the type III PKS. The type III PKS, MerE, iteratively adds two more malonate units and aids in cyclization of the resorcinol core as described by Posehn et al.[215] The individual halogenated monomers are then homodimerized by MerH to create the final macrocycle. The cytochrome P450, MerG, potentially leads to oxidation at the terminal carbon of the aliphatic chain, leading to merocyclophane C and D, but it is unknown which substrate MerG might act on (Figure 44).
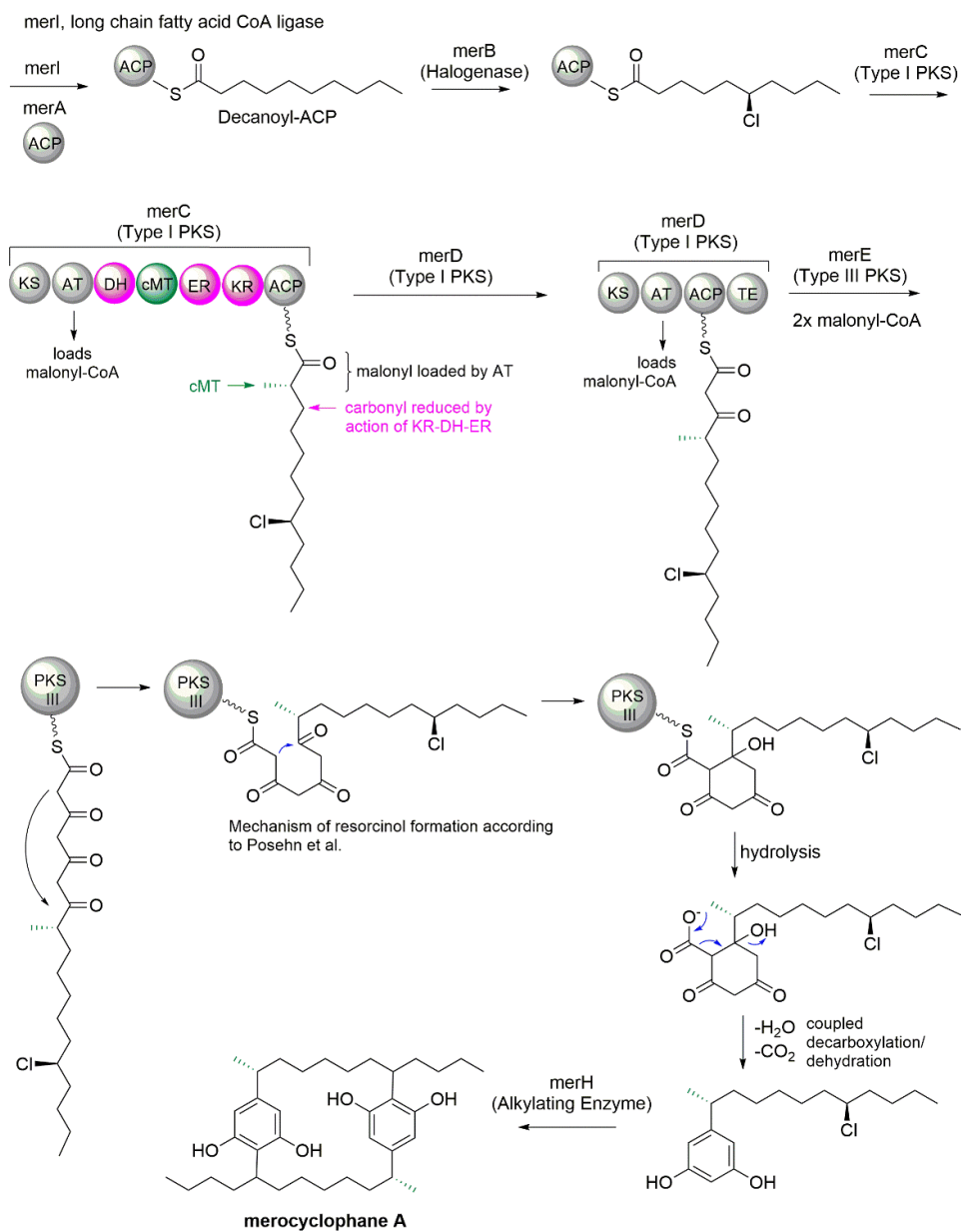
Figure 44: Proposed biosynthetic mechanism for merocyclophanes based on the biosynthetic mechanism of cylindrocyclophanes published by Nakamura et al.

Taxonomy of UIC 10110

UIC 10110 was given a preliminary taxonomic determination of *Nostoc* sp. based on microscopic analysis (Figure 105 Appendix B). The cells were round and spherical and showed differentiation. Intercalary and terminal heterocysts were observed when the strain was grown in nitrate-free media, but no akinetes were observed in later phases of growth; rather, the filaments fragmented, and vegetative cells were released to spread the colony. The cells divided along one plane within the filament and the filaments were flexuous, with no branching, and surrounded by a thin sheath.

The preliminary taxonomy was verified by 16S rRNA analysis. DNA was isolated from UIC 10110 and a portion of the 16S rRNA sequence was amplified via PCR, using the cyanobacterial specific primers 109F and 1509R. The resulting amplicon was purified and submitted for Sanger sequencing. The sequence was manually inspected and trimmed to a length of 1,228 bp. A neighbor-joining tree was produced from a multiple sequence alignment with other 16S rRNA sequences of Nostocaceae strains. UIC 10110 clades with other members of *Nostoc* group I, *Nostoc sensu stricto,* and interestingly clades with the other known producer of merocyclophane A, UIC 10062 (Figure 45).
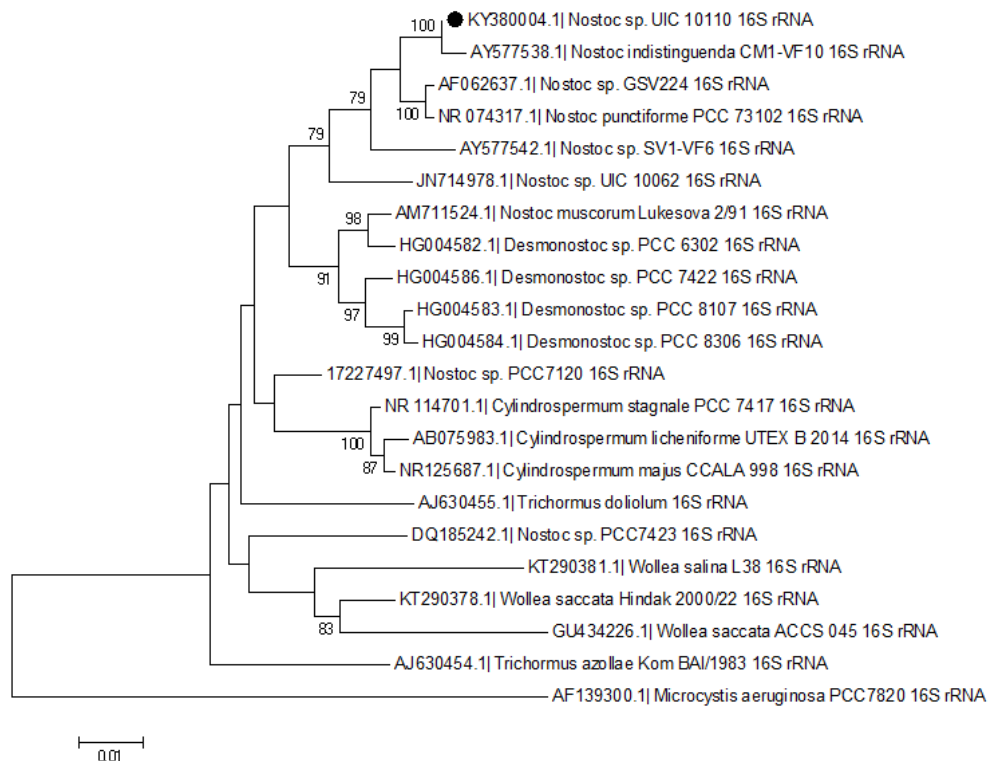
Figure 45: Neighbor-joining tree representing the phylogenetic distances of 16S rRNA sequences of UIC 10110 with other Nostocaceae strains, values at the branches were determined by 1000 bootstrap replicates

Biological Activity of Merocyclophanes

The cytotoxicity of the newly isolated merocyclophanes C and D was assessed using an antiproliferation assay, along with the previously reported merocyclophane A, carbamidocyclophanes A and F, and cylindrocyclophane D. The [7,7] paracyclophanes were screened against three cell lines: MDA-MB-435 a melanoma cell line, MDA-MB-231 a breast cancer cell line, and OVCAR3 an ovarian cancer cell line. Merocyclophane D was found to be the most active of the [7,7] paracyclophanes

screened with an IC_{50} of 0.9 μM against MDA-MB-435, 1.0 μM against MDA-MB-231,

and 2.0 μM against OVCAR3 (Table IV).

TABLE IV: IC_{50} VALUES OF [7,7] PARACYCLOPHANES AND PACLITAXEL AGAINST MDA-MB-435 (MELANOMA), MDA-MB-231 (BREAST CANCER), AND OVCAR3 (OVARIAN CANCER).

Compound	MDA-MB-435 IC_{50} (μM)	MDA-MB-231 IC_{50} (μM)	OVCAR3 IC_{50} (μM)
Merocyclophane A (**70**)	9.8	6.2	5.1
Merocyclophane C (**113**)	1.6	1.4	1.4
Merocyclophane D (**114**)	0.9	1.0	2.0
Carbamidocyclophane A	2.4	1.0	4.3
Carbamidocyclophane F	3.2	1.2	4.2
Cylindrocyclophane D	12	13	10
Paclitaxel	0.0031	0.0026	0.0075

Cyanobacterial [7,7] paracyclophanes have been described to have wide ranging

in vitro biological activities, including cytotoxic, antibacterial, antifungal, and proteasome

inhibition. A hollow fiber tumor assay was performed to evaluate the *in vivo* efficacy of

the [7,7] paracyclophanes as anticancer agents. Merocyclophane C was chosen for the

assay, rather than the more potent merocyclophane D, due to its larger abundance in

the cell extract and ease of isolation. The hollow fiber tumor assay was performed

according to protocols described in Ren et al.[216] Merocyclophane C was dissolved in

30% DMSO, 40% H_2O, and 30% Cremaphor EL as an excipient to aid in solubility.

Solubility was achieved through sonication and heating to 35 °C, and it remained

soluble even when cooled. Hollow fibers containing cultured OVCAR3 or MDA-MB-231

cancer cells were inserted interperitoneally in Ncr *nu*/*nu* mice. The mice were treated

with the merocyclophane C formulation for four days at 10 mg/kg or 15 mg/kg before being sacrificed. The hollow fibers were recovered, and the cancer cells were assessed for viability in an antiproliferation assay. Neither cell line showed statistically significant inhibited growth at either concentration; however, the MDA-MB-231 cancer cells showed inhibited growth with a p-value of 0.051 at 15 mg/kg when compared to the negative control. This prompted a repeat of the experiment against the MDA-MB-231 cells at two concentrations of 15 mg/kg and 17 mg/kg with a newly isolated batch of merocyclophane C. During this second experiment, however, the mice lost significant weight and the treatment proved too toxic, so the experiment was terminated.

Conclusions

Two new merocyclophane analogs were isolated from the freshwater cyanobacterium UIC 10110 *Nostoc* sp. Merocyclophane C and D are the first reported analogs of the merocyclophane core structure reported by Kang et al.[131] Both contained the characteristic α-methyl group on the [7,7] paracyclophane core. The structures were elucidated by ^{1}H NMR and mass spectrometry and confirmed by 2D NMR experiments. The stereoconfiguration of the four chiral centers was determined using ECD and comparing the spectra to that of merocyclophane A. A type I/type III PKS hybrid gene cluster was identified as a putative merocyclophane BGC in the sequenced genome of UIC 10110. The putative merocyclophane BGC has high homology to the described cylindrocyclophane and carbamidocyclophane BGCs. The biosynthesis of the merocyclophanes was proposed based on the architecture of the putative gene cluster and the biosynthesis of the cylindrocyclophanes. This proposed biosynthesis provided a biosynthetic source for the α-methyl group that differentiates the

merocyclophanes from the β-methyl group containing cylindrocyclophanes. The precise

functions of each of the enzymes encoded in the putative gene cluster will need to be

verified to confirm the identity of the gene cluster as well as the accuracy of the

proposed biosynthesis. UIC 10110 was determined to be a *Nostoc* sp. via

morphological analysis and partial 16S rRNA sequence analysis. It was found to clade

with members of *Nostoc* group I, *Nostoc sensu stricto,* which further demonstrates the

ability of this genus to produce biologically active natural products. The biological

activity of the newly isolated merocyclophanes was assessed in both *in vitro*

antiproliferation assays and an *in vivo* hollow fiber tumor assay. Merocyclophane D was

found to be the most potent of all the [7,7] paracyclophanes examined in the

antiproliferation assay, but due to its low abundance in the extract merocyclophane C

was chosen for the hollow fiber tumor assay instead. No statistically significant growth

inhibition of the cell lines was seen in the hollow fiber tumor assay at either

concentration of the merocyclophane C treatment, and repeated experiments with

higher concentrations led to mouse toxicity. The differences in toxicity between the two

experiments could be due to the use of two separately isolated batches of

merocyclophane C. The two batches could have had differences in purity leading to

imprecise concentrations. It is also likely that the low solubility of merocyclophane C

had adverse effects on the experimental outcome. It is possible that using more

hydrophilic [7,7] paracyclophanes could prove more successful in future hollow fiber

tumor assays.

Methods

Spectroscopic Instrumentation

UV spectra were recorded on Shimadzu UV spectrometer UV2301 on scanning mode from 190 to 360 nm. ECD spectra were recorded with a JASCO J-710 ECD spectrometer. IR spectra were recorded with a JASCO FTIR-410 Fourier transform infrared spectrometer. All NMR spectra were recorded on a Bruker Avance 900 MHz spectrometer. Chemical shifts in all NMR spectra were referenced to the residual solvent peak (MeOH-d_4 δ_H 3.31 and δ_C 49.15). HMBC spectra were acquired with average $^3J_{CH}$ of 8 Hz. HRESIMS was recorded on a Shimadzu LC-IT-ToF mass spectrometer.

Biological Material

An environmental sample containing UIC 10110 was collected outside Denver, Colorado, from the Colorado River in 2008 (40°17'32" N, 105°51'02" W). Microisolation techniques were used to isolate an individual filament of the cyanobacterium, which was cultured in Z medium, into the unialgal strain UIC 10110.[189] UIC 10110 was grown in two 10 L cultures of Z medium, with constant sterile aeration and a 18/6 hour light/dark cycle for 8 weeks. All fluorescent lights emitted 2.6 klx and the culture room was kept at 22°C. The cell mass was harvested by centrifugation after 8 weeks and lyophilized.

Microscopic Taxonomic Analysis

UIC 10110 was given a preliminary taxonomic identification of *Nostoc* sp. based on microscopic analysis of morphological indicators. The trichomes were isopolar, flexible, had no branching, and were coated in a thin sheath, with cells dividing in a single plane. The cells were spherical and had individual intercalary heterocysts, with

no heterocysts observed at the end of filaments. No akinetes were observed in the filament; rather, the filament appeared to reproduce through breakage and dispersal of the filaments. Modern taxonomic systems from Komarek et al. 2014, and Bergey's Manual of Systematic Bacteriology 2001 were followed in determining the preliminary taxonomy of UIC 10110.[64,93] Micrograph can be found in Appendix B.

<u>Genome Sequencing</u>

Genomic DNA was extracted from approximately 250 mg wet cell mass of UIC 10110. The cells were centrifuged, and the cell pellet was suspended in lysis buffer (2.5 mL, 10 mM Tris, 0.1 EDTA, 0.5% w/v SDS, 20 µg/mL pancreatic RNase pH 8.0) and chicken egg white lysozyme (1 mg/mL). The cell and lysozyme mixture was incubated at 37 °C for one hour in a water bath. Proteinase K, at a final concentration of 100 µg/mL, was added to the mixture and incubated at 50 °C for one hour in a water bath. The cells were pelleted via centrifugation and the supernatant was removed. After the pre-treatment, the Wizard Genomic DNA purification kit (Promega) was used to extract genomic DNA from the cells, and concentration and purity were assessed using a nanodrop UV spectrophotometer. The Nextera DNA Library Preparation Kit was used to create sequencing libraries and the sequencing was performed with an Illumina MiSeq. The resulting reads were trimmed using a quality score trimmer with a 0.01 limit and all reads below 50 base pairs were discarded. The assembly was performed with the Spades assembler through Illumina Basespace with all default settings selected.[214] The resulting draft assembly of approximately 4,000 contigs was submitted to AntiSMASH 3.0.[21] The putative merocyclophane gene cluster can be found on GenBank with the accession code: KY379971.

16S rRNA Amplification and Taxonomic Analysis

     PCR was used to amplify a portion of the 16S rRNA gene from the genomic DNA isolated from UIC 10110. The cyanobacterial specific 16S rRNA primers 109F and 1509R were used in the PCR reaction. The reaction contained 2 µL of both primers at 10 µM, 1 µL of dNTPs at 10 µM, 0.5 µL of Phusion high-fidelity polymerase, 2 µL of the genomic DNA at 25 ng/µL, 10 µL of the Phusion Buffer, and 32.5 µL of nuclease-free water. The reaction was performed on a Bio-Rad C1000 Thermocycler with a program of denaturation for 2 minutes at 95 °C, followed by 35 cycles of 95 °C for 30 seconds, 49 °C for 30 seconds, 72 °C for 2 minutes, and a final extension of 72 °C for 5 minutes. Agarose gel electrophoresis was used to confirm the amplification and a MinElute kit (Qiagen) was used to purify the product. Sanger sequencing was performed with the cyanobacterial specific primers 109F and 1509R, as well as an internal primer 359F. The resulting sequence was manually examined and trimmed to 1,228 bp. The partial 16S rRNA sequence used in the phylogenetic analysis can be found on GenBank with the accession code: KY380004. Cyanobacterial 16S sequences from the Nostocaceae family of at least 1Kb, were obtained from GenBank, including several Bergey's Manual reference strains.[93] The 16S rRNA sequence of *Microcystis aeruginosa* was used as an outgroup. These were aligned using ClustalW with default gap opening and extension penalties. The multiple sequence alignment was used in conjunction with a neighbor-joining tree, using 1000 bootstrap replicates, to infer the evolutionary distances between strains.

Extraction and Isolation of Merocyclophanes from UIC 10110

The dry cell mass of the two 10 L cultures was 9.22 g which was extracted via maceration three times with a mixture of 1:1 DCM and MeOH. The dried extract yield was 1.268 g and was fractionated by vacuum liquid chromatography on a Diaion HP20 column with a stepwise gradient of IPA in water to produce six fractions. The 70:30 IPA in water fraction, fraction four, was the most active against MDA-MB-435 cancer cell line. ^{1}H NMR and HRESIMS based dereplication of fraction four revealed the presence of the known merocyclophane A, as well as a potentially new [7,7] paracyclophane. The dereplication separation was performed on a semipreparative HPLC with a reversed-phased monolithic column (Phenomenex Onyx Monolithic C18 column 4.6 X 100 mm) with a gradient from 55% MeOH to 100% MeOH in 23 minutes at 4 mL/min.. A third [7,7] paracyclophane was identified during the isolation process. The three [7,7] paracyclophanes were isolated via semipreparative HPLC with a reversed-phase column (Varian Microsorb Dynamax C18 semipreparative column 250 mm X 10 mm) with a gradient from 65% to 100% ACN in 35 minutes at 4 mL/min. Merocyclophane D eluted at 6 minutes, merocyclophane C eluted at 19 minutes, and merocyclophane A eluted at 27 minutes. The percent yield of the three [7,7] paracyclophanes from the 1.268 g of extract were: merocyclophane A: 18.2 mg (1.4%), merocyclophane C: 21.7 mg (1.7%), and merocyclophane D: 2.1 mg (0.16%).

Spectroscopic Data

Merocyclophane C (**113**): Pale amorphous powder; [α]^{25}$_D$ -32 (*c*0.15, MeOH); UV (MeOH) λmax (log ε) 207 (4.55), 221 (4.02), 272 (3.24) nm; ECD (MeOH) λmax (Δε) 205.4 (-4.9), 215.8 (-6.0), 277.8 (0.02); IR (neat) vmax 3371 (br), 2972, 2931, 2865,

1621, cm^{-1}; ^{1}H and ^{13}C NMR spectroscopic data, Table II; HRESIMS m/z 567.4173 [M-H]$^-$ (calcd for C$_{36}$H$_{55}$O$_5$, 567.4169), spectra available in Appendix A.

Merocyclophane D (**114**): White amorphous powder; [α]$^{25}_D$ -26 (*c*0.23, MeOH); UV (MeOH) λmax (log ε) 208 (4.6), 221 (4.02), 272 (3.24) nm; ECD (MeOH) λmax (Δε) 205.8 (-3.6), 216.2 (-4.5), 279.6 (0.1); IR (neat) vmax 3378, 2970, 2932, 2859, 1656, cm^{-1}; ^{1}H and ^{13}C NMR spectroscopic data, Table II; HRESIMS m/z 583.4154 [M-H]$^-$ (calcd for C$_{36}$H$_{55}$O$_6$, 583.4159), spectra available in Appendix A.

Cytotoxicity Assay

MDA-MB-435, MDA-MB-231, and OVCAR3 cells were harvested by trypsinization during log phase growth. Cells were grown in 96-well plates at 37 °C in 5% CO$_2$. Cultures were incubated overnight and treated with samples, positive control (paclitaxel), or negative control (DMSO) for 72 hours. Cell viability was assessed by CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega). IC$_{50}$ values were calculated by comparison to the negative control.

Animals

All protocols for the Ncr *nu/nu* mice were followed from previously described protocols.[216] The protocols were approved by the University of Illinois at Chicago Animal Care and Use Committee (protocol number 13-057), and university guidelines for animal care were followed.

In Vivo Hollow Fiber Assay

Protocols were followed from the previously published Ren et al.[216] Merocyclophane C was dissolved 30% DMSO, 40% water, and 30% Cremaphor EL, as

an excipient. Ncr *nu*/*nu* mice were prepared with hollow fibers, inserted into the

peritoneum, containing either OVCAR2 or MDA-MB-231 cancer cell lines. The mice

were injected intraperitoneally with the merocyclophane C formulation at 10 mg/kg or 15

mg/kg, paclitaxel at 5 mg/kg, or vehicle for four days, after which the mice were

sacrificed, and the hollow fibers were recovered. The cancer cells were then analyzed

for growth inhibition using the antiproliferation assay described above.

4   PHENOTYPIC SCREENING OF *NOSTOC* SPP. LEADS TO DISCOVERY OF

CHEMICALLY DIVERSE [7,7] PARACYCLOPHANES

[7,7] paracyclophanes Identified in *Nostoc* spp.

Many chemically diverse [7,7] paracyclophanes have been found from a variety of *Nostoc* spp. The nostocyclophanes were the first [7,7] paracyclophanes identified from a *Nostoc* sp., UTEX B1932. These compounds are the only [7,7] paracyclophanes identified with chlorine within the macrocycle as well as a unique glycosylation patterns on the resorcinol hydroxy groups.[132] However, the strain that produces these unique compounds has since been lost and the compounds have not been reported again. UIC 10022a was the first *Nostoc* sp. strain to be reported to produce cylindrocyclophanes with terminal aliphatic chain chlorination. Chlipala et al. showed that chlorines on the terminal aliphatic chains could be replaced with bromine by growing the strain in media with bromine in place of chlorine.[137] The vietnamese *Nostoc* sp. strains CAVN2 and CAVN10 were found to produce carbamidocyclophanes as well as cylindrocyclophanes.[138,140] The carbamidocyclophanes have an uncommon carbamate group at the α position as well as terminal aliphatic chain chlorination. Preisitsch et al. were able to use the same approach as Chlipala et al. to add bromine to the terminal aliphatic chains to further increase the chemical diversity produced by CAVN2.[134] Further study of CAVN2 and CAVN10 led to a [7,7] paracyclophane enriching extraction method that led to the discovery of the cylindrofridins A-C, biosynthetic intermediates of the cylindrocyclophanes, from *Cylindrospermum stagnale* PCC 7417.[141] The genome of CAVN2 was also the source of the carbamidocyclophane gene cluster, the first [7,7] paracyclophane BGC identified in a *Nostoc* sp.[134] The merocyclophanes were found in strains UIC 10062 and UIC 10110, both *Nostoc* spp. Merocyclophanes contained an α-

methyl group which differentiated them from the previously identified nostocyclophanes and cylindrocyclophane core structures.[131,217] These structurally diverse [7,7] paracyclophanes were all identified through phenotypic screening approaches. Additional [7,7] paracyclophane producing strains including a group of [7,7] paracyclophanes containing a new moiety at the α position were identified through continued screening of *Nostoc* sp. cell extracts from the UIC cyanobacteria culture collection.

Dereplication of [7,7] paracyclophanes from *Nostoc* spp.

Using the previously described phenotypic screening workflow, extracts of *Nostoc* sp. strains were screened in antiproliferation assays to identify potential anticancer compounds. The extracts and fractions of five strains were active against MDA-MB-435 and MDA-MB-231 cell lines. Using the previously described dereplication protocol, all five strains were found to contain [7,7] paracyclophanes as the active components. Dereplication of the active component in UIC 10250 revealed the production of merocyclophane A. HRESIMS analysis of the active subfraction identified a molecular weight that matched the molecular weight of merocyclophane A, and the ^{1}H NMR spectrum of the subfraction showed signals similar to the published ^{1}H NMR spectrum of merocyclophane A (Figure 46).[131] Microscopic morphological analysis was performed, and the morphology was consistent with *Nostoc* sp. (Appendix B). However, the cultured strain was lost before DNA could be extracted and 16S rRNA analysis could be performed.
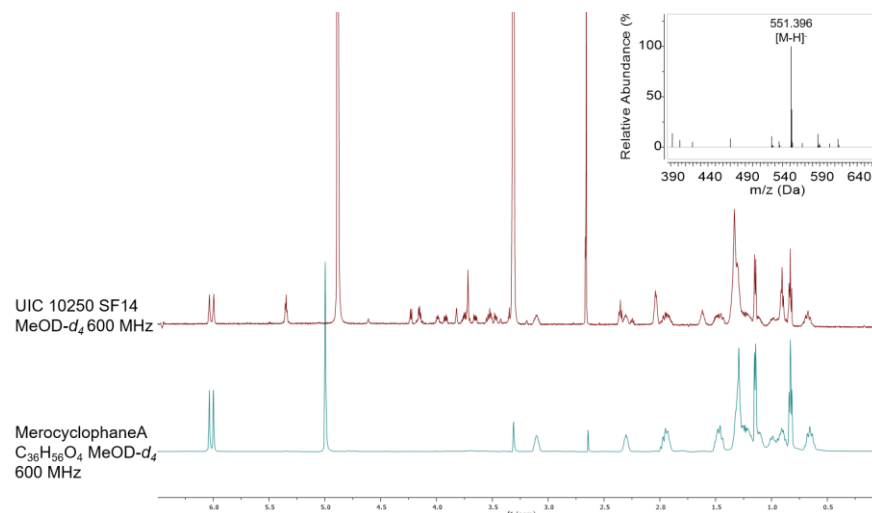
Figure 46: Dereplication of merocyclophane A from UIC 10250 Nostoc sp. showing a deprotonated ion [M-H]⁻ and ¹H NMR spectrum that match that of the published merocyclophane A.

Dereplication of the active components of UIC 10448 revealed the presence of cylindrocyclophanes A_1-A_4 and C_1-C_4. The chlorine isotope patterns in the mass spectra of the active subfractions were able to discern the presence of each of the chlorinated compounds and the ¹H NMR spectra were representative of the published spectra for cylindrocyclophanes A_1-A_4 and C_1-C_4 (Figure 47).[137] 16S rRNA analysis confirmed the microscopic morphological analysis and designated the strain as a *Nostoc* sp., clading within *Nostoc* group III (Figure 50 Appendix B).
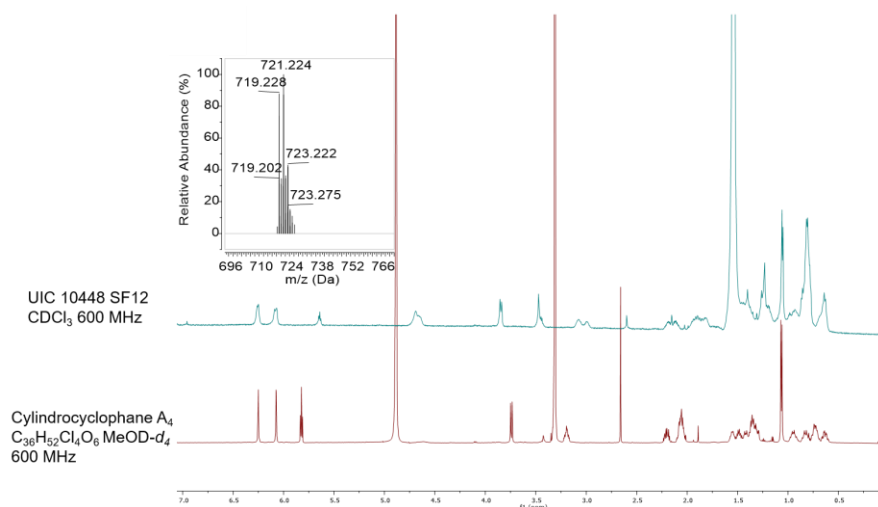
Figure 47: Dereplication of cylindrocyclophane A₄ from UIC 10448, including the isotope pattern in the HRESIMS consistent with a tetrachlorination pattern
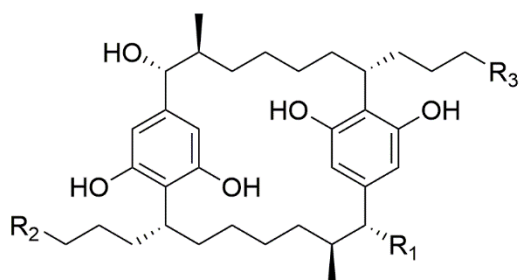
	R₁	R₂	R₃
Cylindrocyclophane A₄ (**115**)	OH	CHCl₂	CHCl₂
Cylindrocyclophane A₃ (**116**)	OH	CH₂Cl	CHCl₂
Cylindrocyclophane A₂ (**117**)	OH	CH₃	CHCl₂
Cylindrocyclophane A₁ (**118**)	OH	CH₃	CH₂Cl
Cylindrocyclophane C₄ (**119**)	H	CHCl₂	CHCl₂
Cylindrocyclophane C₃ (**120**)	H	CH₂Cl	CHCl₂
Cylindrocyclophane C₂ (**121**)	H	CH₃	CHCl₂
Cylindrocyclophane C₁ (**122**)	H	CH₃	CH₂Cl

UIC 10279, UIC 10366, and UIC 10534 were found to produce additional [7,7] paracyclophanes. Dereplication of the active fractions of UIC 10279 and UIC 10534 identified three new [7,7] paracyclophanes with molecular weights of 848.492 Da, 716.449 Da, and 700.445 Da. Dereplication of the active fractions of UIC 10366 identified new [7,7] paracyclophanes with molecular weights of 848.492 Da and 758.457 Da, as well as the known compound cylindrocyclophane D. All ^{1}H NMR spectra of the new [7,7] paracyclophanes contained aromatic signals near δ_H 6.0, methyl signals near δ_H 1.0, and carbohydrate-like signals between δ_H 3.0 and δ_H 4.0, suggesting the presence of a glycoside. Glycosylated [7,7] paracyclophanes were obtained from both UIC 10279 and UIC 10366. Dereplication of the active components in UIC 10534 suggested the presence of compounds identical to those found in UIC 10279. During isolation of the compounds from UIC 10366, another minor compound was obtained with a molecular weight of 794.437 Da.

Isolation and Structure Elucidation of Ribocyclophanes A-E

Ribocyclophanes A, B, and C were isolated from UIC 10279 and the structures were elucidated using single crystal x-ray crystallography, mass spectrometry, and NMR techniques. Ribocyclophanes A, D, E, and cylindrocyclophane D were isolated from UIC 10366 and the structures were elucidated using mass spectrometry and NMR techniques. All compounds were isolated via HPLC on a C8 column with a gradient of 60% MeOH in water to 80% MeOH in water in 40 minutes.

The molecular formula of ribocyclophane A was identified as $C_{46}H_{72}O_{14}$. The ^{1}H NMR spectrum showed a high degree of similarity to known cylindrocyclophane spectra but differed in the presence of signals between δ_H 3.0 and δ_H 4.0. These signals along

with a signal of an anomeric proton at δ_H 4.85 suggested the presence of a carbohydrate in ribocyclophane A. The compound was determined to be symmetrical based on analysis of the ^{1}H NMR and DEPTQ spectra, where half of the expected signals were observed. COSY correlations from the anomeric H-1'/6' to H-2'/7' and sequential correlations from H-2'/7' to H-3'/8' to H-4'/9' to the diastereotopic H-5'/10', suggested that the carbohydrate was a pentopyranose. HMBC correlations from H-5'/10' to the anomeric C-1'/6' confirmed the carbohydrate to be a pentopyranose and correlations from H-1'/6' to C-1/14 placed the carbohydrate at the α position. The cylindrocyclophane core was determined similarly to the previously discussed merocyclophanes: COSY correlations identified the aliphatic chain spin system and β-methyl group, and HMBC correlations connected the resorcinol groups to the aliphatic chains (Figure 48, Table V). The absolute configuration of the glycoside was determined by single crystal x-ray crystallography and optical rotation spectroscopy. X-ray crystallography determined the glycone to be a β-ribopyranose with the oxygen atoms at C-1, C-2, and C-4 in the axial position (Appendix A). Optical rotation spectroscopy of the acid hydrolyzed carbohydrate showed negative rotation, indicating the absolute configuration of the glycone to be β-D-ribopyranose with a $^1C_4$ configuration. The absolute configuration of the acid hydrolyzed aglycone [7,7] paracyclophane was found using ECD. Negative Cotton effects at 230 nm and 280 nm suggested the same stereoconfiguration of the six stereocenters as all other described [7,7] paracyclophanes (Appendix A).
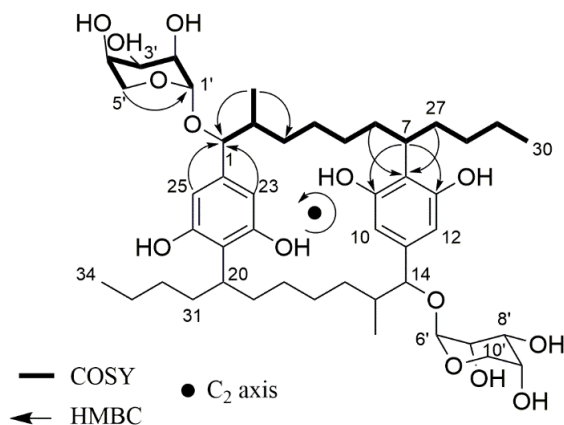
Figure 48: Key 2D NMR correlations of ribocyclophane A, with similar correlations used to solve B-D

TABLE V: SPECTROSCOPIC DATA FOR RIBOCYCLOPHANE A (^{1}H 900 MHZ, ^{13}C 225 MHZ IN MEOH-D_4).

No.	Ribocyclophane A (123)	
	δ_C, type	δ_H, mult. (*J* in Hz)
1/14	90.1, CH	3.65, d (12.2
2/15	41.8, CH	1.65, m
3/16	34.9, CH$_2$	0.66, m 0.75, m
4/17	30.0, CH$_2$	0.82, m 1.44, m
5/18	30.9, CH$_2$	0.71, m 0.96, m
6/19	35.5, CH$_2$	1.32, m 2.03, ddd (4.2)
7/20	37.0, CH	3.16, m
8/21	117.9, C	
9/22	158.9, C	
10/23	105.5, CH	6.19, s
11/24	142.1, C	
12/25	109.3, CH	6.09, s
13/26	157.1, C	
27/31	34.8, CH$_2$	1.47, m 1.95, m
28/32	31.9, CH$_2$	1.04, m 1.17, m
29/33	24.0, CH$_2$	1.18, m 1.28, m
30/34	14.8, CH$_3$	0.79, t (7.2)
35/36	16.7, CH$_3$	1.05, d (6.0)
1'/6'	104.6, CH	4.85, d (1.8)

2'/7'	72.8, CH	3.73, d (1.8, 3.0)
3'/8'	66.9, CH	3.70, t (3.0)
4'/9'	71.6, CH	3.59, ddd (2.4, 3.0)
5'/10'	65.6, CH_2	3.19, dd (12.6, 2.4) 3.30, dd (12.6)

Ribocyclophanes B and C were similarly solved by mass spectrometry and NMR techniques. The molecular formulae of ribocyclophanes B and C were found to be $C_{41}H_{64}O_{10}$ and $C_{41}H_{64}O_9$, respectively. The symmetrical appearance of the ^1H NMR spectrum was no longer apparent, as it had been in ribocyclophane A, suggesting that ribocyclophane B and C were asymmetrical. In the ^1H spectra of both compounds, the signals corresponding to the glycone integrated to one hydrogen each, suggesting the presence of only one ribopyranose. H-14, at δ_H 3.74, in the ^1H NMR spectrum of ribocyclophane B, was shifted slightly from the spectrum of ribocyclophane A, and showed no HMBC correlations to anything at the α position, suggesting the β-D-ribopyranose in ribocyclophane A was replaced with a hydroxy group. This was also supported by the change of 132 Da in the molecular weight. Ribocyclophane C was solved by the presence of a diastereotopic CH_2 group at C-14 with signals at δ_H 2.61 and δ_H 1.81 that both showed COSY correlations to H-15. This indicated the hydroxy group at C-14 in ribocyclophane B was replaced with a methylene group at C-14 in ribocyclophane C. The molecular formula and the change of 16 Da between the two compounds supported this structure (Table VI). Both compounds had similar ECD spectra to ribocyclophane A, with positive Cotton effects at approximately 270 nm and 220 nm, suggesting that they had the same absolute configuration as ribocyclophane A.

TABLE VI: SPECTROSCOPIC DATA FOR RIBOCYCLOPHANES B AND C (^{1}H 900 MHZ, ^{13}C 225 MHZ IN MEOH-D$_4$).

No.	Ribocyclophane B (**124**)		Ribocyclophane C (**125**)	
	δ_C, [a, b] type	δ_H[b], mult. (*J* in Hz)	δ_C, [a,b] type	δ_H[b], mult. (*J* in Hz)
1	88.5, CH	3.65, d (10.2)	88.4, CH	3.66, d (10.2)
2	40.1, CH	1.66, m	40.1, CH	1.66, m
3	33.4, CH$_2$	0.65, m 0.76, m	33.3, CH$_2$	0.66, m 0.79, m
4	28.6, CH$_2$	0.83, m 1.42, m	28.6, CH$_2$	0.81, m 1.43, m
5	29.0, CH$_2$	0.72, m 0.95, m	29.3, CH$_2$	0.74, m 0.97, m
6	33.9, CH$_2$	1.32, m 2.04, m	34.0, CH$_2$	1.32, m 2.00, m
7	35.4, CH	3.16, m	35.4, CH	3.11, m
8	116.1, C		114.7, C	
9	157.7, C		156.9, C	
10	103.5, CH	6.24, s	106.5, CH	6.02, s
11	142.2, C		139.4, C	
12	107.6, CH	6.07, s	108.5, CH	5.97, s
13	156.0, C		155.7, C	
14	80.2, CH	3.74, d (9.6)	44.3, CH$_2$	1.81, t (12.6) 2.61, dd 13.2, 3.6)
15	40.7, CH	1.56, m	35.4, CH	1.57, m
16	33.4, CH$_2$	0.65, m 0.76, m	35.3, CH$_2$	0.62, m 1.01, m
17	28.6, CH$_2$	0.83, m 1.42, m	28.8, CH$_2$	0.81, m 1.34, m
18	29.0, CH$_2$	0.72, m 0.95, m	29.3, CH$_2$	0.74, m 0.97, m
19	33.9, CH$_2$	1.32, m 2.04, m	34.0, CH$_2$	1.32, m 2.00, m
20	35.4, CH	3.16, m	35.4, CH	3.16, m
21	116.1, C		116.4, C	
22	157.7, C		157.2, C	
23	103.7, CH	6.20, s	103.9, CH	6.20, s
24	140.4, C		140.3, C	
25	107.8, CH	6.09, s	107.6, CH	6.09, s
26	155.0, C		155.7, C	
27	33.3, CH$_2$	1.47, m 1.95, m	33.3, CH$_2$	1.48, m 1.94, m
28	30.2, CH$_2$	1.04, m 1.17, m	30.3, CH$_2$	1.06, m 1.17, m
29	22.4, CH$_2$	1.19, m 1.29, m	22.3, CH$_2$	1.20, m 1.29, m
30	13.1, CH$_3$	0.79, t (7.2)	13.1, CH$_3$	0.82, t (7.2)
31	33.3, CH$_2$	1.47, m 1.95, m	33.3, CH$_2$	1.48, m 1.94, m
32	30.2, CH$_2$	1.04, m 1.17, m	30.3, CH$_2$	1.06, m 1.17, m
33	22.4, CH$_2$	1.19, m 1.29, m	22.3, CH$_2$	1.20, m 1.29, m
34	13.1, CH$_3$	0.79, t (7.2)	13.3, CH$_3$	0.79, t (7.2)
35	15.4, CH$_3$	1.05, d (7.2)	15.4, CH$_3$	1.05, d (6.6)
36	15.4, CH$_3$	1.06, d (6.6)	19.3, CH$_3$	0.95, d (6.6)

1'	103.0, CH	4.85, d (2.4)	103.0, CH	4.85, d (2.4)
2'	71.1, CH	3.73, brs	71.1, CH	3.73, brs
3'	65.2, CH	3.70, t (3.0)	65.3, CH	3.70, t (3.0)
4'	70.1, CH	3.60, brs	70.0, CH	3.59, brs
5'	63.9, CH_2	3.20, dd (12.6, 2.4) 3.31, overlapped	64.0, CH_2	3.20, dd (12.6, 2.4) 3.31, overlapped

Ribocyclophane D and E were isolated from the extract of UIC 10366 using semi-preparative C8 HPLC. Ribocyclophane D was found to have a molecular formula of $C_{43}H_{66}O_{11}$ and the ^1H NMR spectrum was similar to that of ribocyclophane B and ribocyclophane C. The signals corresponding to the glycone integrated to one hydrogen each, suggesting the presence of one carbohydrate and a different moiety at the other α position that would break the symmetry observed in ribocyclophane A. A sharp singlet at δ_H 2.0 and a change in mass of 42 Da compared to ribocyclophane B, suggested the presence of an acetate moiety at the α position. An HMBC correlation from H-14 to C-37 at δ_c 172.6 confirmed the presence and location of the acetate group. ^1H NMR signals for the carbohydrate matched those of ribocyclophanes A-C, suggesting the sugar to be a β-D-ribopyranose (Table VII). This was confirmed by 2D NMR experiments as well as optical rotation spectroscopy of the liberated glycone, after acid hydrolysis. The ECD spectrum of ribocyclophane D also matched that of the other ribocyclophanes, indicating the same absolute configuration at the six [7,7] paracyclophane core stereocenters.

HRESIMS analysis of ribocyclophane E showed an isotope pattern consistent with the presence of chlorine, with isotope peaks of the deprotonated molecule at 793.9098 Da and 795.9101 Da in a 3:1 ratio. Accordingly, it was found to have a

molecular formula of $C_{43}H_{67}ClO_{11}$. The ^{1}H spectrum shared some characteristics of the other ribocyclophanes. The spectrum maintained the same methyl group pattern, similar carbohydrate signals, and a strong singlet near δ_H 2.0, however other signals did not indicate a [7,7] paracyclophane structure. The ^{1}H spectrum also shared similarities with the cylindrofridins B and C, isolated by Preisitsch et al.[141] The highly overlapped CH_2 signals complicated structure elucidation of the aliphatic chains, but careful analysis of the COSY spectrum revealed the presence of two aliphatic chains. The placement of the chlorine atom was also determined by analysis of the COSY spectrum, with correlations from H-7, at δ_H 3.83 to H-6 and H-27. HMBC signals similar to those found in the other ribocyclophanes were identified to connect resorcinol moieties to the aliphatic chains, however, no signals were observed from H-7 to C-9 and C-13 or H-6 and H-27 to C-8 (Table VII). This indicated that the macrocycle was not closed and ribocyclophane E was a linear compound similar to cylindrofridin B and C. The linear structure was also confirmed by a calculation of the double bond equivalents from the molecular formula. The β-D-ribopyranose and acetate moiety were both connected by HMBC signals from H-1 to C-1' and H-14 to C-37, respectively. Positive Cotton effects around 220 nm and 270 nm suggested the same absolute configuration of the stereocenters in ribocyclophane E as the other isolated ribocyclophanes. The absolute configuration of C-7 could not be determined, but based on studies of the enantioselective halogenase by Nakamura et al. the likely absolute configuration of C-7 is R.[135] This could be confirmed by incubating ribocyclophane E with the alkylating enzyme CylK and assessing whether ribocyclophane D is formed. Nakamura et al. demonstrated that the alkylating enzyme CylK closes the macrocycle only when the

absolute configuration of C-7 is R.[135] The structure of ribocyclophane E suggested that

it was likely a biosynthetic intermediate of ribocyclophane D.

TABLE VII: SPECTROSCOPIC DATA FOR RIBOCYCLOPHANES D AND E (^1H 900 MHZ, ^{13}C 225 MHZ IN MEOH-D$_4$).

No.	Ribocyclophane D (**126**)		Ribocyclophane E (**128**)	
	δ_C,[a,] type	δ_H[c], mult. (J in Hz)	δ_C,[b] type	δ_H[c], mult. (J in Hz)
1	90.1, CH	3.65, d (10.0)	87.6, CH	4.0, m
2	41.7, CH	1.64, m	40.4, CH	1.77, m
3	34.9, CH$_2$	0.65, m 0.76, m	33.9, CH$_2$	0.97, m 1.26, m
4	30.0, CH$_2$	0.84, m 1.43, m	22.7, CH$_2$	1.32, m
5	30.8, CH$_2$	0.7, m 0.94, m	27.6, CH$_2$	1.33, m 1.44, m
6	35.5, CH$_2$	1.33, m 2.03, m	39.4, CH$_2$	1.54, m 1.62, m
7	37.0, CH	3.16, m	65.1, CH	3.83, m
8	118.7, C		102.7, CH	6.15, s
9	159.0, C		159.9, C	
10	109.6, CH	6.13, s	106.0, CH	6.17, s
11	139.3, C		144.1, C	
12	105.3, CH	6.18, s	106.0, CH	6.17, s
13	157.1, C		159.9, C	
14	83.6, CH	5.0, d (10.4)	80.8, CH	5.4, m
15	40.2, CH	1.77, m	39.9, CH	1.79, m
16	34.6, CH$_2$	0.71, m 0.78, m	30.6, CH$_2$	1.29, m
17	29.8, CH$_2$	0.84, m 1.43, m	29.8, CH$_2$	1.05, m 1.34, m
18	30.7, CH$_2$	0.7, m 0.94, m	29.6, CH$_2$	1.09, m
19	35.4, CH$_2$	1.33, m 2.03, m	34.5, CH$_2$	1.47, m 1.92, m
20	37.0, CH	3.16, m	36.4, CH	3.19, m
21	117.8, C		118.2, C	
22	159.0, C		158.3, C	
23	105.5, CH	6.2, s	107.3, CH	6.22, br
24	142.1, C		141.9, C	
25	109.2, CH	6.08, s	107.3, CH	6.22, br
26	157.1, C		158.3, C	
27	35.0, CH$_2$	1.49, m 1.91, m	39.4, CH$_2$	1.6, m 1.68, m
28	31.8, CH$_2$	1.05, m 1.17, m	27.6, CH$_2$	1.29, m 1.34, m
29	24.0, CH$_2$	1.19, m	23.4, CH$_2$	1.27, m

		1.29, m		1.33, m
30	14.8, CH_3	0.79, t (7.3)	14.3, CH_3	0.92, t (7.3)
31	34.8, CH_2	1.49, m 1.95, m	34.5, CH_2	1.5, m 1.94, m
32	31.9, CH_2	1.05, m 1.17, m	31.7, CH_2	1.07, m 1.17, m
33	24.1, CH_2	1.22, m 1.29, m	23.8, CH_2	1.22, m 1.30, m
34	14.7, CH_3	0.82, t (7.3)	15.1, CH_3	0.82, t (7.2)
35	17.0, CH_3	1.05, d (6.4)	16.5, CH_3	0.99, d (6.5)
36	16.7, CH_3	0.96, d (6.4)	14.8, CH_3	0.85, d (6.5)
37	172.6, C		173.1, C	
38	21.3, CH_3	2.0, s	21.2, CH_3	2.07, s
1'	104.5, CH	4.85, d (2.0)	103.8, CH	4.89, br
2'	72.8, CH	3.72, brs	72.6, CH	3.76, br
3'	67, CH	3.7, t (3.2)	66.9, CH	3.78, br
4'	71.6, CH	3.6, brs	71.3, CH	3.67, br
5'	65.7, CH_2	3.2, dd (12.3, 2.5) 3.32, overlapped	65.6, CH_2	3.32, overlapped 3.47, overlapped
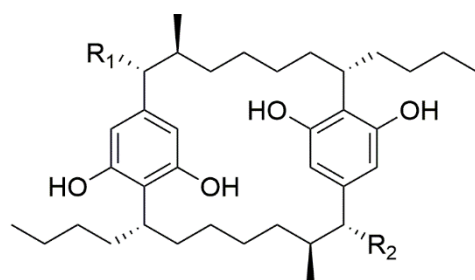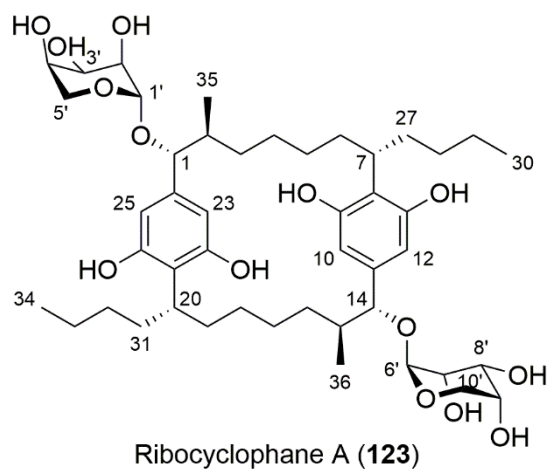
Ribocyclophane A (**123**)

	R_1	R_2
Ribocyclophane B (**124**)	Ribose	OH
Ribocyclophane C (**125**)	Ribose	H
Ribocyclophane D (**126**)	Ribose	OAc
Cylindrocyclophane D (**127**)	OAc	OAc

Ribocyclophane E (**128**)
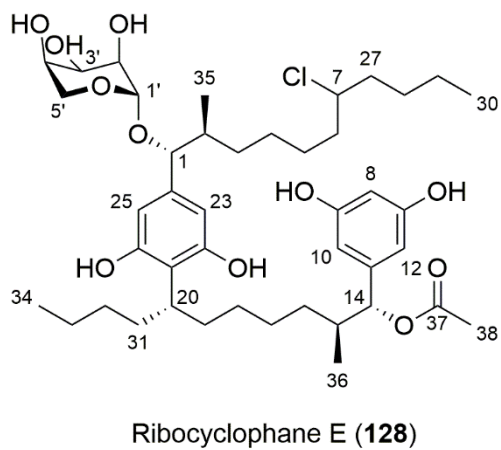
The most prominent feature of these new [7,7] paracyclophanes is the β-D-ribopyranose that adopts a $^1C_4$ conformation. This conformation is more thermodynamically unfavorable compared to the $^4C_1$ conformation, with a potential energy approximately four times higher.[218] The more unfavorable conformation is likely adopted due to steric interactions with the [7,7] paracyclophane structure. The $^4C_1$ conformation occupies a larger space compared to the $^1C_4$ conformation and would likely interact with the resorcinol or the β-methyl group (Figure 49). The $^1C_4$ conformation is also supported by the anomeric effect, which causes heteroatom substituents to adopt an axial conformation when next to heteroatoms in a heterocycle. These ribocyclophanes represent a rare case of a $^1C_4$ β-D-ribopyranose being reported in a natural product.

Favored

$^1C_4$ Conformation

Steric Effect
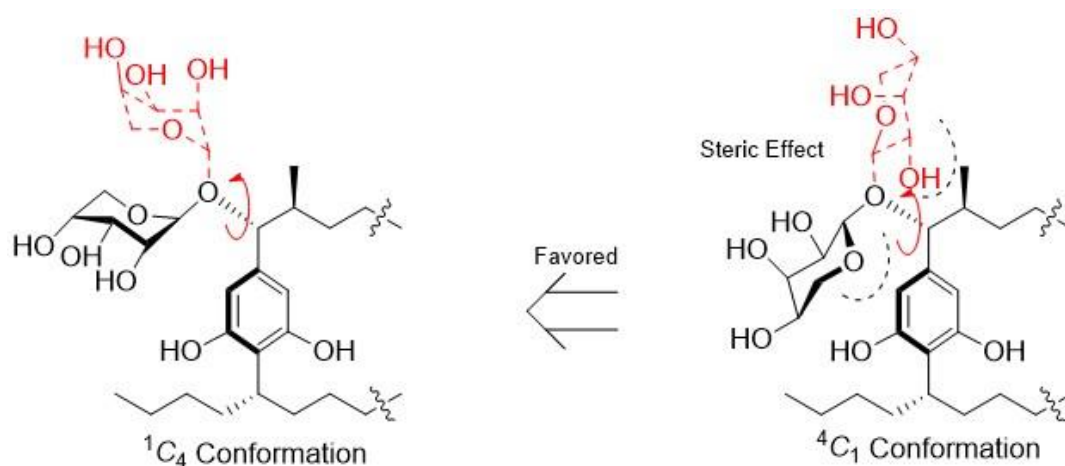
$^4C_1$ Conformation

Figure 49: Representation of the steric effects likely forcing the β-D-Ribopyranose into a thermodynamically unfavorable $^1C_4$ conformation

Biological Activity of Ribocyclophanes A-E

The bioactivity of the ribocyclophanes and cylindrocyclophane D was assessed by antiproliferation assays. The pure compounds were screened against MDA-MB-435, a melanoma cell line, and MDA-MB-231, a breast cancer cell line. Ribocyclophane A and cylindrocyclophane D were the least active of the [7,7] paracyclophanes screened in the assay with IC_{50} values of 6.4 µM and 5.2 µM against MDA-MB-435 and 3.0 µM and 2.3 µM against MDA-MB-231, respectively. The most potent of the [7,7] paracyclophanes was ribocyclophane D, with an IC_{50} of 0.6 µM against MDA-MB-435 and 0.8 µM against MDA-MB-231. Ribocyclophane E was found to be inactive against either cell line (Table VIII). This is consistent with the findings of Preisitsch et al., who found that cylindrofridin B and C were inactive against cancer cells.[141] The [7,7] paracyclophanes obtained from UIC 10366, ribocyclophanes A, D, E, and cylindrocyclophane D, were also evaluated in antibacterial assays. The four compounds were screened against *Enterococcus faecalis*, *Pseudomonas aeruginosa*, and a lptD knockout of *Escherichia coli* - which inhibits transport of lipopolysaccharides to the outer membrane of *E. coli*, making it easier for compounds to penetrate the Gram-negative cell wall. The same pattern of activity was observed in the antibacterial screen as the antiproliferation screen for both the Gram-positive *E. faecalis* and ΔlptD *E. coli*. Ribocyclophane A was the least active, while ribocyclophane D was the most active with an MIC of 0.3 µM against both *E. faecalis* and ΔlptD *E. coli*. Ribocyclophane E was inactive against all the bacteria. None of the compounds were active against the Gram-negative *Pseudomonas aeruginosa*; however, they were able to inhibit the growth of the ΔlptD *E. coli* (Table VIII). This indicates that the intact Gram-negative cell wall is likely the major barrier to

[7,7] paracyclophanes inhibiting the growth of Gram-negative bacterial cells. The

bioactivity of the ribocyclophanes is similar to those of other reported [7,7]

paracyclophanes despite the large glycoside moiety of the ribocyclophanes that would

potentially cause steric interactions with a biological target. It is likely that the greater

hydrophilicity and similar level of activity these compounds exhibit could be

advantageous of in future biological studies.

TABLE VIII: IC_{50} AND MIC VALUES FOR [7,7] PARACYCLOPHANES ISOLATED FROM UIC 10279 AND UIC 10366 AGAINST MDA-MB-435 (MELANOMA), MDA-MB-231 (BREAST CANCER), E. FACAELIS, P. AERUGINOSA, AND E. COLI.

Compound	MDA-MB-435 IC_{50} (µM)	MDA-MB-231 IC_{50} (µM)	*Enterococcus facaelis* MIC (µM)	*Pseudomonas aeruginosa* MIC (µM)	*Eschericia coli* ΔlptD MIC (µM)
Ribocyclophane A (**123**)	6.4	3.0	1.5	>50	0.7
Ribocyclophane B (**124**)	4.1	2.5	N/A	N/A	N/A
Ribocyclophane C (**125**)	1.2	1.3	N/A	N/A	N/A
Ribocyclophane D (**126**)	0.8	0.6	0.3	>50	0.3
Ribocyclophane E (**128**)	>30	>30	>50	>50	>50
Cylindrocyclophane D (**127**)	5.2	2.3	0.7	>50	0.3

Chemotaxonomic Analysis of [7,7] Paracyclophane Producing _Nostoc_ spp.

The dereplication and isolation of [7,7] paracyclophanes from _Nostoc_ spp. in the UIC

cyanobacteria culture collection and elsewhere has shown that production of [7,7]

paracyclophanes may be loosely correlated with taxonomy. _Cylindrospermum_ and

Nostoc are currently the only two genera that have been found to produce [7,7]

paracyclophanes. Within _Nostoc_, production of [7,7] paracyclophanes appears

relegated to two distinct clades, _Nostoc_ group I and _Nostoc_ group III (Figure 50). Partial

16S rRNA sequences are available for nearly all the strains known to produce [7,7]

paracyclophane, except for the Nostocyclophane producer, UTEX B 1932, and the

merocyclophane producer, UIC 10250. A neighbor-joining tree was created from a

multiple sequence alignment of the partial 16S rRNA sequences of [7,7]

paracyclophane producing strains, using 1000 bootstrap replicates. The strains that

produce [7,7] paracyclophanes with a branched β-methyl group, the

cylindrocyclophanes, carbamidocyclophanes, and ribocyclophanes, all clade within the

genus _Cylindrospermum_ or _Nostoc_ group III. The strains that produce [7,7]

paracyclophanes with a branched α-methyl group, the merocyclophanes, clade together

in _Nostoc_ group I. It was previously shown that the putative BGC for merocyclophanes

has distinctly different PKS gene architecture from the gene clusters that produce

cylindrocyclophanes.[217] Interestingly, the strains that harbor these different gene

clusters fall into different phylogenetic groups. The BGCs identified to produce

cylindrocyclophanes, from the genomes of _Cylindrospermum stagnale_ PCC 7417,

Cylindrospermum lichenforme UTEX 2014b, and _Nostoc_ sp. CAVN2, have high

homology, but are in different phylogenetic groups.[133,134] This could suggest the

possibility of a common cylindrocyclophane producing ancestor, or potentially,

horizontal gene transfer as the two genera are evolutionarily removed from one another.

As more strains and BGCs are identified that produce [7,7] paracyclophanes, these

patterns will become clearer and it may be possible to draw further conclusions as to

the evolution of the biosynthesis of these polyketide compounds. Current numbers are

insufficient to make formal predictions; however, these trends could be used to help

narrow the number of genera necessary to screen to identify new [7,7]

paracyclophanes.
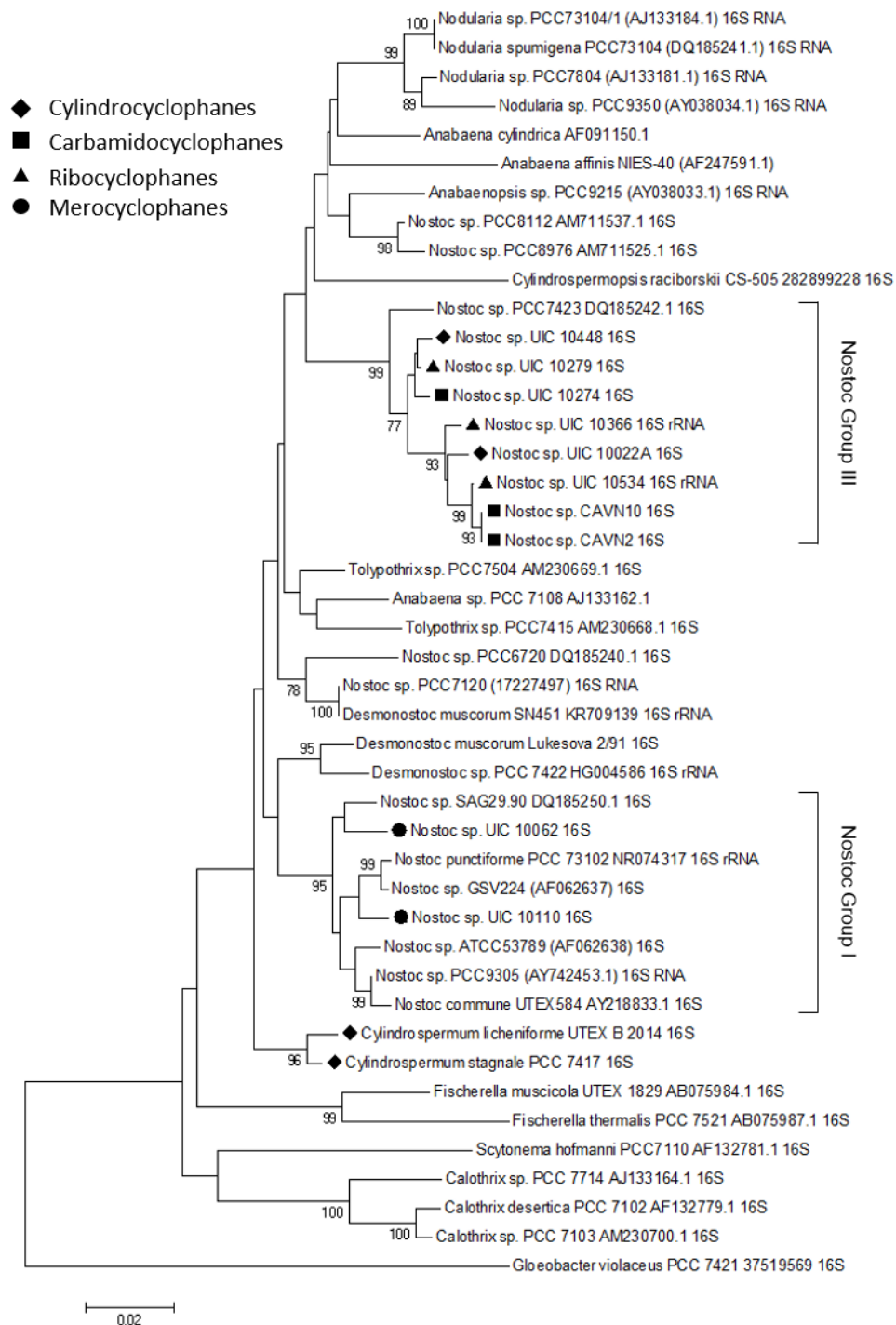
Figure 50:Neighbor-joining tree representing the phylogenetic distances of 16S rRNA sequences of all known [7,7] paracyclophane producers with 16S rRNA sequences, with other Nostocaceae strains. Values at the branches were determined by 1000 bootstrap replicates. Shapes represent the type of [7,7] paracyclophane produced by the strains, circles are merocyclophanes, diamonds are cylindrocyclophanes, squares are carbamidocyclophanes, and triangles are ribocyclophanes.

Conclusions

     Continued screening of the UIC cyanobacteria culture collection in antiproliferation assays resulted in the identification of additional strains that produce merocyclophanes and cylindrocyclophanes, as well as three strains that produce a new class of [7,7] paracyclophane named ribocyclophanes. Ribocyclophanes A-C were isolated from the cell extract of UIC 10279 and ribocyclophanes A, D, and E were isolated from the cell extract of UIC 10366. UIC 10534 was also found to produce ribocyclophanes A-C. Ribocyclophanes represent a new group of compounds with the cylindrocyclophane core structure. A β-D-ribopyranose adopting a thermodynamically unfavorable $^1C_4$ conformation was identified as the glycoside in all the ribocyclophanes. The unique $^1C_4$ conformation of the glycoside has rarely been seen in other natural products. The sugar was identified by 2D NMR, single crystal x-ray crystallography, and optical rotation spectroscopy. The new ribocyclophanes were evaluated in antiproliferation assays and were found to have similar activity levels as other [7,7] paracyclophanes, except for ribocyclophane E, which has an unclosed macrocycle and lacks activity. Ribocyclophanes A, D, E, and cylindrocyclophane D were also screened in an antibacterial assay and were found to be active against Gram-positive bacteria, but not Gram-negative bacteria, likely due to the structure of the Gram-negative cell wall. The increased hydrophilicity of the ribocyclophanes and the potent activity of ribocyclophane D would likely make them better candidates for screening in a hollow fiber tumor assay or use in mechanism of action studies. 16S rRNA sequences of the known [7,7] paracyclophane producers were aligned and used to create a phylogenetic tree. Currently, only three clades appear to produce [7,7] paracyclophanes:

Cylindrospermum, *Nostoc* group I, and *Nostoc* group III. *Cylindrospermum* and *Nostoc* group III appear to produce compounds with the cylindrocyclophane core structure, while *Nostoc* group I contains the two strains that produce the merocyclophanes. These studies provide more evidence that screening *Nostoc* strains in phenotypic assays is an effective way to identify chemically diverse analogs of bioactive natural products.

Methods

Spectroscopic Instrumentation

UV spectra were recorded on a Shimadzu UV spectrometer UV2301, scanning from 190 to 360 nm. IR spectra were recorded on a JASCO FTIR-410 Fourier transform infrared spectrometer. Optical rotations were recorded on a Perkin-Elmer 241 polarimeter. ECD spectra were recorded on a JASCO J-710 ECD spectrometer. NMR spectra were obtained on a Bruker Avance 600 MHz NMR spectrometer containing a 5 mm CPTXI Z-gradient probe. DEPTQ and NMR spectra for ribocyclophane D and E were obtained on a Bruker Avance II 900 MHz NMR spectrometer containing a 5 mm ATM CPTCI Z-gradient probe. The chemical shifts of all spectra were referenced to residual solvent signals for DMSO-d_6 (δ_H 2.49 and δ_C 39.7) or MeOH-d_4 (δ_H 3.31 and δ_C 49.15). The TOCSY experiment used a mixing time of 60 ms and the T-ROESY experiment used a mixing time of 200 ms. HMBC spectra were recorded with average $^3J_{CH}$ of 8 Hz. HRESIMS spectra were recorded on a Shimadzu LC-IT-ToF mass spectrometer.

<u>Biological Material</u>

UIC 10279 was isolated from an environmental sample collected in the

southwest suburbs of Chicago near Joliet, Illinois. UIC 10366 was isolated from an

environmental sample collected in the northwest suburbs of Chicago near Elgin, Illinois.

UIC 10534 was isolated from an environmental sample collected from the Mississippi

River in Savanna, Illinois. UIC 10250 was isolated from an environmental sample

collected from the Colorado River outside of Austin, Texas. UIC 10448 was isolated

from an environmental sample collected from Rockford, Illinois. All strains were isolated

using micropipette isolation techniques. The taxonomic identification of each strain was

performed in accordance to modern taxonomic schemes as stated by Komarek et al.

and partial 16S rRNA analysis, except UIC 10250 which was subsequently lost. Partial

16S rRNA sequences can be found on GenBank under the following accession

numbers: UIC 10279 JX962720, UIC 10366 MF622942, UIC 10534 MG874686, UIC

10448 MG874685. Strains were each cultured in four 3 L Fernbach flasks containing Z

media with a 18/6 hour light/dark cycle, and sterile aeration. The culture room was

maintained at 22 $^{\circ}$C and lights emitted 2.6 klx. Cultures were grown for 7 weeks before

being harvested and lyophilized.

<u>Microscopic Taxonomic Analysis</u>

UIC 10279, UIC 10366, UIC 10534, UIC 10250 and UIC 10448 were all identified

as *Nostoc* sp. based on microscopic analysis of morphological indicators. All strains

had isopolar, flexible trichomes with no branching. Cells divided in a single plane and

were spherical. Cellular differentiation was observed within the trichome. All strains

produced heterocysts when grown in media lacking nitrate and akinetes were not

observed in any of the strains. Sheaths were only observed in UIC 10250 (Appendix B).

Modern taxonomic systems from Komarek et al. and Bergey's Manual of Systematic

Bacteriology 2001 were followed in determining the preliminary taxonomy of the strains.

16S rRNA Amplification and Taxonomic Analysis

Genomic DNA was isolated from each of the strains as described in the previous

chapter. PCR was used to amplify a portion of the 16S rRNA gene, using the

cyanobacterial specific primers 109F and 1509R. The PCR reaction was performed on

a Bio-Rad C1000 Thermocycler and contained 2 µL of each primer at 10 µM, 1 µL of

dNTPs at 10 µM, 0.5 µL of Phusion high-fidelity polymerase, 2 µL of the genomic DNA

at 25 ng/µL, 10 µL of the Phusion Buffer, and 32.5 µL of nuclease-free water. The

thermocycler program was denaturation for 2 minutes at 95 °C, followed by 35 cycles of

95 °C for 30 seconds, 49 °C for 30 seconds, 72 °C for 2 minutes, and a final extension of

72 °C for 5 minutes. The amplification was confirmed by agarose gel electrophoresis

and the amplicons were purified using a MinElute kit (Qiagen). The products were

sequenced by Sanger sequencing with the cyanobacterial specific primers 109F and

1509R, and an internal primer 359F. The sequences were manually examined and

quality trimmed to at least 1Kb. Partial 16S rRNA sequences for each strain, except

UIC 10250, can be found on GenBank. Nostocaceae 16S sequences of at least 1 Kb

were obtained from GenBank and used to create a multiple sequence alignment, with

Gloeobacter violaceus PCC 7421 used as an outgroup. The multiple sequence

alignment was created using ClustalW with default gap opening and extension

penalties. A neighbor-joining tree was created using the multiple sequence alignment

with 1000 bootstrap replicates to infer evolutionary distances between the [7,7] paracyclophane producing strains.

Extraction and Isolation

Lyophilized cells were extracted with DCM and MeOH (1:1) three times and dried using a rotary evaporator. Extracts were fractionated over a Diaion HP-20 solide phase with a step-wise gradient of IPA in water. ^{1}H NMR and LC-MS based dereplication identified merocyclophane A in UIC 10250 and cylindrocyclophanes A_1-A_4 and C_1-C_4 in UIC 10448. The new compounds ribocyclophanes A-C were identified in UIC 10279 and UIC 10534 and ribocyclophane A, D, E, and the known cylindrocyclophane D in UIC 10366. The new compounds were isolated by reversed-phase HPLC on a Varian Microsorb Dynamax C8 column 250mm X 10 mm with a gradient from 60% MeOH in water to 80% MeOH in water over 40 minutes at 4 mL/min. The yields of the new compounds from the active fractions were ribocyclophane A: 15 mg (0.5%), ribocyclophane B: 3 mg (0.1%), and ribocyclophane C: 10 mg (0.3%) from UIC 10279 and ribocyclophane D: 15 mg (0.5%) ribocyclophane E: 1 mg (>0.1%) and cylindrocyclophane D: 6 mg (0.2%) from UIC 10366.

Spectroscopic Data

Ribocyclophane A (**123**): white amorphous powder; $[\alpha]^{25}_D$ - 6 (*c* 0.12, MeOH); UV (MeOH) λ_{max} (log ε) 208 (4.41), 223 (3.88), 276 (3.13) nm; ECD (MeOH 1.0 mM) λ_{max} (Δε) 208 (2.87), 224 (1.27), 282 (1.04) nm; IR (neat) v_{max} 3365, 2930, 2857, 1628, 1432, 1381, 1321 cm^{-1}; ^{1}H and ^{13}C NMR, COSY, and HMBC data, Table V; HRESIMS *m/z* 847.4894 [M-H]$^-$ (calcd for $C_{46}H_{71}O_{14}$, 847.4844), spectra available in Appendix A.

Ribocyclophane B (**124**): white amorphous powder; $[\alpha]^{25}_D$ - 6 (*c* 0.07, MeOH); UV (MeOH) λ_{max} (log ε) 207 (4.10), 221 (3.65), 275 (2.90) nm; ECD (MeOH 0.7 mM) λ_{max} (Δε) 208 (1.62), 222 (0.99), 284 (0.86) nm; IR (neat) v_{max} 3378, 2930, 2858, 1631, 1469, 1432, 1372 cm^{-1}; ^1H and ^{13}C NMR data, Table VI; HRESIMS *m/z* 715.4471 [M-H]$^-$ (calcd for $C_{41}H_{63}O_{10}$, 715.4421), spectra available in Appendix A.

Ribocyclophane C (**125**): white amorphous powder; $[\alpha]^{25}_D$ + 6 (*c* 0.16, MeOH); UV (MeOH) λ_{max} (log ε) 209 (4.54), 223 (4.03), 275 (3.23) nm; ECD (MeOH 0.7 mM) λ_{max} (Δε) 211 (2.59), 224 (1.44), 282 (2.00) nm; IR (neat) v_{max} 3382, 2926, 2856, 1691, 1620, 1592, 1431, 1369 cm^{-1}; ^1H and ^{13}C NMR data, Table VI; HRESIMS *m/z* 699.4533 [M-H]$^-$ (calcd for $C_{41}H_{63}O_9$, 699.4472), spectra available in Appendix A.

Ribocyclophane D (**126**): white amorphous powder; $[\alpha]^{25}_D$ - 3 (*c* 0.14, MeOH); UV (MeOH) λ_{max} (log ε) 208 (4.38), 220 (3.85), 280 (3.05) nm; ECD (MeOH 1.0 mM) λ_{max} (Δε) 211(2.82), 228 (1.21), 284 (0.9); IR (neat) v_{max} 3380, 2931, 2862, 1625, 1433, 1372, 1256 cm^{-1}; ^1H and ^{13}C NMR data, Table VII; HRESIMS *m/z* 757.4548 [M-H]$^-$ (calcd for $C_{43}H_{65}O_{11}$, 757.4526), spectra available in Appendix A.

Ribocyclophane E (**128**): white amorphous powder; $[\alpha]^{25}_D$ + 7 (*c* 0.14, MeOH); UV (MeOH) λ_{max} (log ε) 211 (3.82), 223 (3.61), 278 (2.98) nm; ECD (MeOH 0.7 mM) λ_{max} (Δε) 208 (2.12), 226 (1.12), 277 (-0.8); IR (neat) v_{max} 3405, 2930, 2858, 1567, 1426, 1387, 1257, 830 cm^{-1}; ^1H and ^{13}C NMR data, Table VII; HRESIMS *m/z* 793.4310 [M-H]$^-$ (calcd for $C_{43}H_{66}ClO_{11}$, 793.4293), spectra available in Appendix A.

Single Crystal X-Ray Crystallogrpahy

A single crystal, approximately 10 X 10 X 30 μM, was encased in paratone-N oil and cooled 10 100K in a nylon loop. Data was collected at the Advanced Photon

source, LC-CAT, sector 21, on a MAR 300 mm CCD detector. 240 images with width of 1° were collected at a crystal-to-detector distance of 91 mm and wavelength of 0.77 Å, to a resolution of 0.8 Å. Images were indexed, and intensities were integrated and scaled with XDS.[219] The unit cell is monoclinic, space group P2$_1$, with two molecules in the asymmetric unit, and eight disordered MeOH molecules. SHELXS was used to solve and refine the structure.[220] Crystal data: 2C$_{46}$H$_{72}$O$_{14}$ ·8C$_2$H$_6$O, monoclinic, space group *P2$_1$* (No. 4), *a* = 10.752 (2) Å, *b* = 34.601 (4) Å, *c* = 16.762 (2) Å, β = 90.480 (8)°, V = 6235.7 (16) Å^3, Z = 4, Dc = 1.10, μ = 0.081, F000 = 2256. Reflections collected/unique = 55588/23169 (Rint = 0.0514), final R1 = 0.0996, wR2 = 0.2478 for reflections with I>2σI; R1 = 0.1063, wR2 = 0.2552 for all reflections; goodness-of-fit = 1.019 for all unique data. Bijvoet analysis: Flack x = -.4(6), Parsons z = -.24(15), Probability P2(true) = 1.000. Crystallographic data (ID entry CCDC 898167) was deposited in the Cambridge Crystallographic Data Centre.

Acid Hydrolysis

Acid hydrolysis was performed on approximately 4 mg of ribocyclophane A, isolated from both UIC 10279 and UIC 10366. The compounds were hydrolyzed with 1N HCl (1 mL) at 100 °C for 1 hour. The hydrolysate was dried under filtered air and extracted with water and ethyl acetate (EtOAc) to yield the water fraction (1 mg) and the EtOAc fraction (3 mg). The negative specific rotation found in the water fraction indicated the glycone was D-ribopyranose: [α]^{25}D − 9 (*c* 0.13, H$_2$0). Isolation of the aglycone was performed by reverse-phased HPLC, as previously described. The pure aglycone had a deprotonated molecular weight of 583.4113 [M-H]⁻. The aglycone was

analyzed by ECD and displayed negative Cotton effects at 230 and 280, indicating the same absolute configuration reported for other [7,7] paracyclophanes.

Antiproliferation Assay

MDA-MB-435 and MDA-MB-231 cells were harvested by trypsinization during log phase growth. Cells were grown in 96-well plates at 37 $^{\circ}$C in 5% CO_2. Cultures were incubated overnight and treated with samples, positive control (paclitaxel) or negative control (DMSO) for 72 hours. Cell viability was assessed by CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega). IC_{50} values were calculated by comparison to the negative control.

Antibacterial Assay

Bacterial cultures were grown at 37 $^{\circ}$C until log phase growth, determined by OD_{600} measurements. Bacterial cells were diluted to 0.1 OD_{600} and seeded into 96 well plates. The bacteria were treated with a serial dilution of the [7,7] paracyclophane compounds or positive control (streptomycin or kanamycin) and incubated overnight. MIC values were determined by incubation of the wells with alamar blue for two hours and visual inspection of metabolism of the dye.

5 GENOME MINING UNLOCKS THE NATURAL PRODUCT PRODUCING CAPABILITIES OF UIC 10630 *NOSTOC* SP.

Genome Mining

      Phenotypic screening has long been used as an effective approach to identify natural products. The compounds isolated through phenotypic screening are biologically relevant to the activity being probed and many have become staples of modern medicine.[8,53,210] As the previous two chapters demonstrated, phenotypic screening readily identifies bioactive natural products, but often the compounds identified belong to a single chemical scaffold. While the bioassay identifies compounds immediately relevant to the activity being probed, it inherently limits the diversity of compounds that can be identified at one time. Different compounds can be identified from the same strain by changing the bioassay used to detect them, however a different bioassay simply changes the narrow window used to identify compounds. As the price of genome sequencing has fallen and bioinformatic tools for analyzing genomes have become more prevalent, genome mining has become a viable approach for assessing the full natural product producing potential of an organism.[221–223]

      Genome mining is a bottom-up natural product discovery platform that uses genes that encode for biosynthetic enzymes to drive the identification and isolation of new natural products. Vast amounts of publicly available genomic data and decreasing costs of genome sequencing have resulted in a shift toward this discovery paradigm.[224,225] Next-generation sequencing and real-time long read sequencing technologies have greatly improved the sequencing throughput and allowed more complete genomes to be obtained.[223,226] This genomic data can be "mined" using bioinformatic tools that use conserved motifs from described biosynthetic genes to identify potential biosynthetic genes as parts of BGCs. Many of these bioinformatic

tools are publicly available such as BAGEL, SMURF, and NaPDoS, which greatly

decreases the costs and time required to perform bioinformatic analyses on large

genomic datasets.[227] AntiSMASH is a comprehensive bioinformatic platform and

biosynthetic gene cluster database that has become a widely used tool in genome

mining approaches. AntiSMASH utilizes a well curated database of verified BGCs with

known products to identify putative BGCs, based on their similarity to well described

biosynthetic gene motifs.[20,24,228] BGCs in the NRPS, PKS, RiPP, terpene, saccharide,

and other classes are all identified by AntiSMASH. Putative BGCs can also be

compared to known BGCs using BLAST, providing a percent similarity score to

described gene clusters. NRPS and PKS BGCs can be analyzed further to predict the

corresponding structures based on a suite of algorithms that use the adenylation and

acyl transferase domain substrate specificities to predict the individual amino acids or

acyl groups incorporated into the compounds.[24,228]

Matching Biosynthetic Gene Clusters to Natural Products

Despite the improvements in sequencing technology and bioinformatic tools,

predicting the complete structures of compounds encoded in BGCs remains

problematic. Deviations to the rule of collinearity, such as domains being skipped or

iteratively used, complicate structure predictions. In addition, accurately predicting the

substrate preferences of adenylation or acyl transferase domains is not always possible.

Furthermore, many biosynthetic enzymes and tailoring enzymes remain poorly

described. This is especially true of orphan BGCs for which the respective natural

product is unknown. This can lead to incomplete or conservative structural predictions

which makes identifying the produced compounds by mass spectrometry nearly

impossible. Approaches to match BGCs to their respective compounds typically involve genetic manipulation of the BGC or the use of a heterologous host, both of which are not easily scalable and can be time consuming.[60] Several approaches to match cyanobacterial BGCs to their respective compounds in a cell extract have been described.[18,229–231]

Molecular networking has been used for identifying the products of BGCs that are homologous to known BGCs. Two BGCs with high homology to each other would potentially produce compounds with similar MS/MS fragmentation patterns. In a molecular network, the *m/z* of the known compound and the unknown compound would likely group together, facilitating identification of the new natural product.[231,232] Untargeted MS/MS and molecular networking can also be used to find the products of orphan BGCs; however, the reference database must be well populated and annotated. Nodes in a molecular network with few connections to other known compounds are prioritized and analyzed as potential matches for the BGC, but this does little more than narrow the number of *m/z* values to investigate.[231,232]

Other approaches have used isotopic labeling to identify the products of BGCs by NMR or comparative metabolomics approaches. The genomisotopic approach described by Gross et al. uses isotopically labeled precursors, predicted by A and AT domain analysis, in the media of the bacterium to label the product of a BGC. These labeled compounds can then be identified using NMR.[229] This approach, however is limited in several ways. The isotopically labeled precursor must be unique to the BGC being analyzed, stable in culture media over time, and non-toxic to the producing

organism. Confidence in the precursor prediction also must be high before purchasing expensive isotopically labeled precursors.

A method utilized by Kinnel et al. is practical only for cyanobacteria but offers more promise as an effective way to identify the products of many BGCs at one time.[230] The MALDIisotopic approach takes advantage of the unique culturing conditions of cyanobacteria to identify compounds that contain nitrogen. Cyanobacteria are grown in media that contain few components, consisting of water, numerous salts, and a few metals. Cyanobacteria are also known to incorporate nitrogen provided in the media into all primary and secondary metabolites.[233] When ^{15}N nitrate is used as the only source of nitrogen in the media, cyanobacteria will exclusively incorporate the heavier ^{15}N into the natural products they produce.[230] The MALDIisotopic approach uses a MALDI-ToF mass spectrometer to detect the change in mass and identify the compound as a nitrogen containing natural product. The MALDI-ToF allows analysis of a single filament or small portion of a culture, allowing *in situ* analysis of the compounds that are produced. This approach could be coupled with genome sequencing and bioinformatic approaches such as AntiSMASH to match compounds with their BGCs. However, the MALDIisotopic approach suffers from the effects of the matrix used to assist in ionization, as smaller molecular weight compounds or compounds produced at low abundances may not be identified due to the noise from the matrix. In order to overcome some of the challenges associated with each of these approaches a combination of the genomisotopic approach and MALDIisotopic approach was used to match BGCs identified from a terrestrial *Nostoc* sp. to their cognate natural products.

Genome Mining and Comparative Metabolomic Approach

A method to identify the products of BGCs identified by genome mining was developed based on the studies of Gross et al., Kinnel et al., and the natural product workflow developed in the Orjala Lab.[190,229,230] The method involves full genome sequencing of cyanobacteria and analysis by AntiSMASH to identify putative BGCs. Concurrently, the cyanobacterium is grown in two culture media, one containing ^{15}N nitrate as the exclusive source of nitrogen and one containing natural isotopic abundance nitrate. The cell masses of the two cultures are extracted and the labeled and unlabeled extracts are analyzed by comparative metabolomics with UPLC-HRESIMS. Differences in the molecular weight between compounds with the same retention time in the labeled and unlabeled extracts correspond to the number of nitrogens present in the compound. This number can then be compared to the number of nitrogens predicted by adenylation domain substrate predictions from AntiSMASH (Figure 51). This allows matching of *m/z* values in the cell extract of a cyanobacterium to the identified BGC, aiding in identification and isolation of new compounds. Several *Nostoc* spp. were sequenced by next-generation sequencing for identification of the natural product producing potential of the strains via genome mining. *Nostoc* sp., UIC 10630, isolated from the soil collection Soil F (Chapter 2), was investigated as a proof of concept.
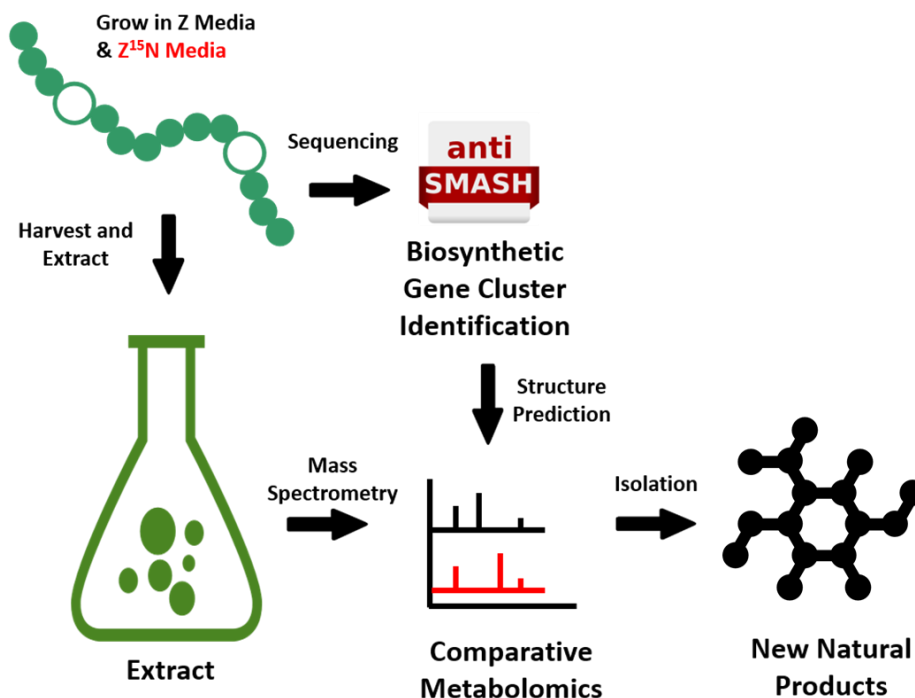
Figure 51: Diagram of the genome mining and comparative metabolomic approach used to identify the natural products encoded in biosynthetic gene clusters

Sequencing and Bioinformatic Identification of BGCs from UIC 10630

UIC 10630 *Nostoc* sp. was sequenced by both Illumina MiSeq and PacBio SMRT Cell sequencing to obtain more complete coverage of the genome. Epiphytic bacterial DNA contaminated the Illumina and PacBio sequencing data, causing the hybrid assembly to be highly fragmented. Filtering the sequence reads by coverage level allowed most of the cyanobacterial sequences to be separated from the contaminating epiphytic bacterial sequences. The filtered sequences from both sequencing technologies were assembled into a hybrid assembly. The draft genome of UIC 10630 was submitted to AntiSMASH 4.0 for identification of putative BGCs. The resulting AntiSMASH data was manually curated to identify BGCs that appeared to be complete.

Nine BGCs were identified as complete, with at least two others split between contigs.

Two of the BGCs, BGC 1 and 2, matched known BGCs for geosmin and heterocyst

glycolipids, both compounds that contain no nitrogen and are often associated with

Nostoc spp.[100,164,165] Another BGC, a type I PKS designated BGC 3, had low similarity

to other BGCs, but was predicted to not contain any nitrogens, rendering the isotopic

labeling approach impractical for identifying the natural product. The remaining six

BGCs were designated as BGC 4-9 and were prioritized for identification of the

respective natural products using the new method. Two of the BGCs had high similarity

to know BGCs, with BGC 4 having 100% similarity to the nostopeptolide BGC, and BGC

5 having 87% similarity to the anabaenopeptin BGC. Four of the BGCs had low

similarity or no similarity to other know BGCs, with BGC 6 having 41% similarity to the

aeruginoside BGC and the others having no similarity to known gene clusters. A list of

the type of each BGC and similar BGCs is shown in Table IX along with the number of

predicted nitrogens for each natural product.

TABLE IX: LIST OF COMPLETE BIOSYNTHETIC GENE CLUSTERS IDENTIFIED BY ANTISMASH IN UIC 10630. TYPE OF GENE CLUSTER, SIMILARITY TO ANY KNOWN BIOSYNTHETIC GENE CLUSTER AND NUMBER OF NITROGENS PREDICTED BY ADENYLATION DOMAINS ARE INCLUDED.

Cluster Number	BGC Type	Known BGCs (% Similarity)	Adenylation Domain Predicted # of Nitrogens
BGC 1	Type I PKS	Heterocyst Glycolipid (100%)	0
BGC 2	Terpene Synthase	Geosmin (100%)	0
BGC 3	Type I PKS	Heterocyst Glycolipid (57%)	0
BGC 4	NRPS-PKS Hybrid	Nostopeptolide (100%)	10
BGC 5	NRPS	Anabaenopeptin (87%)	7
BGC 6	NRPS	Aeruginoside (41%)	6
BGC 7	NRPS-PKS Hybrid	None	3
BGC 8	NRPS-PKS Hybrid	None	11
BGC 9	NRPS-PKS Hybrid	None	2

^{15}N Labeling and Comparative Metabolomics

Studies have shown that the metabolome of cyanobacteria can change depending on the growth conditions.[234] To access the full potential of the cyanobacterial metabolome, we used three different media, Z, BG12, and BG12$_o$. Z media is a lean media, containing fewer salts and minerals than the rich media BG12. BG12$_o$ is a derivative of BG12 that contains no nitrogen source. Z media contains two sources of nitrate, $NaNO_3$ and $Ca(NO_3)_2$. To create a ^{15}N derivative of Z media, both nitrate sources were exchanged in equal molar amounts for $Na^{15}NO_3$. The Calcium ion from $Ca(NO_3)_2$ was supplied through equal molar amounts of $CaCl_2$. BG12 media contains only one source of nitrate, $NaNO_3$, which was exchanged for $Na^{15}NO_3$ to

create a ^{15}N derivative of BG12. UIC 10630 was grown in five media, Z, Z^{15}N, BG12,

BG12^{15}N, and BG12$_o$, in biological duplicates. To avoid bias in the extraction, two

different cell mass extraction methods were employed: extraction with MeOH:CH$_2$Cl$_2$

(1:1) or H$_2$O:MeOH (1:1). To identify compounds that may have been exuded into the

media, a direct analysis of the lyophilized spent media via MALDI-ToF was also

performed (Figure 52). The two cell extracts were analyzed by HRESIMS using UPLC-

IT-ToF with polarity switching. Both extraction methods were able to detect the same

four compounds with incorporated ^{15}N label. The analysis of the spent media detected

only two of the ^{15}N labeled compounds previously identified in the cell extracts. *In situ*

analysis of the labeled and unlabeled filaments of UIC 10630 was performed via

MALDI-ToF to identify the number of compounds that could be detected by the

MALDIisotopic approach compared to our method. The MALDIisotopic approach

described by Kinnel et al. was only able to detect three of the previously identified

compounds with the ^{15}N label.[230]
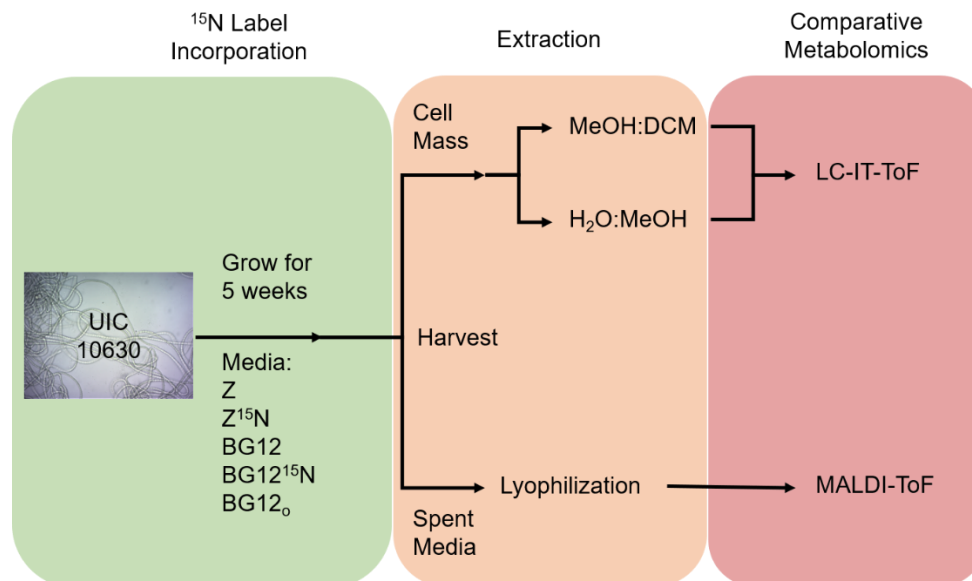
Figure 52: Flowchart demonstrating the experimental set-up and extraction methods used in identifying the natural products produced by UIC 10630 Nostoc sp

Products from four of the six remaining BGCs were identified by detecting shifts in the molecular weight between peaks with the same retention time in the labeled and unlabeled extracts. All four identified compounds were observed in each of the media types and in both cell mass extractions. Compounds containing, ten, seven, six, and three nitrogens were identified in the mass spectra with protonated ions of 1081.591 Da, 828.421 Da, 865.461 Da, and 604.380 Da respectively (Figure 53).

Figure 53: Mass spectra from the comparative metabolomics method showing the shift in mass due to incorporation of ^{15}N label into the compounds. Mass spectra from unlabeled compounds are shown in black and mass spectra from labeled compounds are shown in red. The corresponding BGC was determined based on AntiSMASH adenylation domain predictions.

Dereplication of the Known Compound Nostopeptolide A1

The compound containing ten nitrogens corresponded to BGC 4, the BGC that had 100% similarity to the BGC of nostopeptolide A1 (**77**). Nostopeptolide A1 is a known natural product isolated from *Nostoc* sp. GSV224, that is known to aid in cyanobacterial invasion of plant tissue.

Nostopeptolide A1 (**77**)

The protonated ion and molecular formula of $[M+H]^+ = 1081.593$ and $C_{53}H_{80}N_{10}O_{14}$

matched the protonated ion and formula reported for Nostopeptolide A1.[148,149] The

compound was further analyzed by MS/MS and showed a fragmentation pattern

identical to that of nostopeptolide A1, confirming the presence of **77** in the extract of UIC

10630. The observation of a molecular ion of $[M+H]^+ = 1091.556$ in the labeled extract

suggests that the ^{15}N label was fully incorporated into the natural product and

demonstrates that this method can be used to match BGCs to their respective nitrogen

containing compounds.

De-Orphaning Biosynthetic Gene Cluster of Aeruginosin 865

The compound containing 6 nitrogens corresponded to BGC 6, the BGC with low

similarity to the aeruginoside BGC. A protonated ion of $[M+H]^+ = 865.461$ was observed

and a molecular formula of $C_{41}H_{64}N_6O_{14}$ was determined by HRESIMS. A literature

search for aeruginosins and aeruginosides with a molecular formula of $C_{41}H_{64}N_6O_{14}$

yielded a match for aeruginosin 865 (**16**), an aeruginoside produced by various *Nostoc*

spp. with a unique uronic acid glycoside and fatty acid chain attached to the Choi amino acid.[113] One mg of the compound was isolated by semi-preparative HPLC on a C18 column, eluting at 15.5 minutes with a gradient of 10% ACN in water to 100% ACN for 30 minutes at 4 mL/min. The compound was fragmented by MS/MS on a Q-ToF mass spectrometer and the spectrum of the isolated compound matched that of the published aeruginosin 865 (Figure 54). This confirmed presence of **16** in the extract and suggested BGC6 to be the biosynthetic gene cluster for aeruginosin 865.
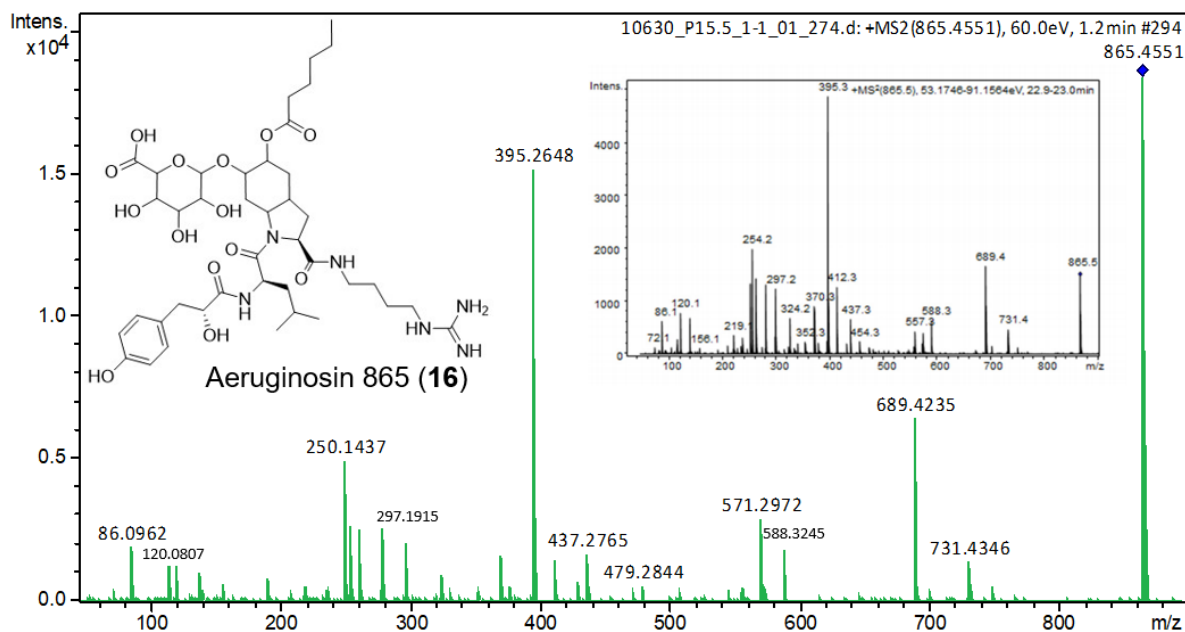
Figure 54: Dereplication of aeruginosin 865 based on HRESIMS and MS/MS spectra

Like other aeruginosins, **16** is a serine protease inhibitor, but it has also been shown to inhibit NF-κB translocation into the nuclease, a unique activity amongst this class of compounds.[113] The uronic acid carbohydrate and a fatty acid on a 4,5, hydroxy-Choi amino acid is also unique within the aeruginosin structural class. No BGC has ever been reported to be responsible for the production of **16**, so a preliminary analysis of the BGC identified by this experiment was performed (Figure 55 and Table X). The NRPS genes with sequence similarity to *aerA*, *aerB*, and *aerG* in aeruginosin and aeruginoside biosynthesis, were predicted to select for phenyl-acetate, leucine, and proline. The three genes homologous to aerD-aerF are likely responsible for choi biosynthesis from arogenate, as suggested by Ishida et al.[109,235] The third NRPS gene, *aerG*, likely selects for the completed choi amino acid and not proline as initially predicted. The other genes are not well described in aeruginosin or aeruginoside biosynthesis, however, the presence of ORF11 and ORF12, which are predicted to encode a mannose synthase and mannosyltransferase gene, suggest they could be responsible for addition of the uronic acid carbohydrate. Likewise, ORF14 is predicted to encode a MBOAT O-acyltransferase type protein which may be involved in addition of the choi-bound fatty acid chain. The similarity to the core biosynthetic genes of aeruginosin and aeruginoside BGCs and the presence of genes likely responsible for addition of the uronic acid carbohydrate and fatty acid, suggest that this BGC may be responsible for the production of aeruginosin 865.[109] The highly conserved protein sequences in the genome of the *Nostoc* photobiont of the lichen *Peltigera malacae*, suggests that the BGC for aeruginosin 865 may also be present in these symbiotic cyanobacteria (Table X).

Putative Aeruginosin 865 BGC
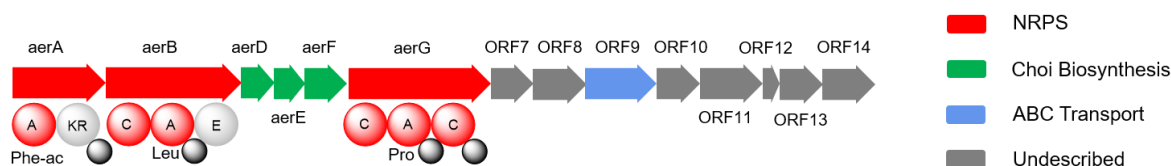
Figure 55: Putative biosynthetic gene cluster of aeruginosin 865 identified by AntiSMASH and the comparative metabolomics method

TABLE X: PROTEINS ENCODED IN THE PUTATIVE AERUGINOSIN 865 BIOSYNTHETIC GENE CLUSTER INCLUDING LENGTH, CLOSEST HOMOLOGOUS PROTEIN BASED ON BLASTP, AND HOMOLOGY TO CORRESPONDING AERUGINOSIDE BIOSYNTHETIC GENES.

Protein	Length (aa)	Closest BlastP Similar Protein [Organism] (% Identity)	% Identity to Aeruginoside Biosynthesis Proteins
AerA	1433	KR domain-containing protein [Nostoc sp. Peltigera malacea cyanobiont] (**93%**)	AerA 54%
AerB	1581	Non-Ribosomal Peptide Synthase [Nostoc sp. Peltigera malacea cyanobiont] (**93%**)	AerB 59%
AerD	198	Bacilysin biosynthesis protein bacA [Nostoc sp. Peltigera malacea cyanobiont] (**97%**)	AerD 59%
AerE	233	Cupin domain-containing protein [Nostoc sp. Peltigera malacea cyanobiont] (**95%**)	AerE 55%
AerF	264	KR domain-containing protein [Nostoc sp. Peltigera malacea cyanobiont] (**94%**)	AerF 73%
AerG	1638	Hypothetical protein [Nostoc sp. Peltigera malacea cyanobiont] (**94%**)	AerG 59%
ORF7	354	Hypothetical protein [Nostoc sp. Peltigera malacea cyanobiont] (**93%**)	None
ORF8	399	Aldo/Keto reductase [Nostoc sp. Peltigera malacea cyanobiont] (**93%**)	ORF8 74%
ORF9	619	Peptide ABC transport substrate binding protein [Nostoc sp. Peltigera malacea cyanobiont] (**89%**)	None

ORF10	311	Isopenicillin N synthase family oxygenase [Nostoc sp. Peltigera malaceae cyanobiont] (**97%**)	AerH 38%
ORF11	547	Dolichyl-phosphate-mannose-protein mannosyltransferase [Nostoc sp. Peltigera malacea cyanobiont] (**90%**)	None
ORF12	164	GtrA family protein [Nostoc sp. Peltigera malacea cyanobiont] (**93%**)	None
ORF13	369	Hypothetical protein [Nostoc sp. Peltigera malacea cyanobiont] (**90%**)	None
ORF14	506	MBOAT family protein [Nostoc sp. Peltigera malacea cyanobiont] (**85%**)	None

Identification and Structure Elucidation of a New Anabaenopeptin

The compound with seven nitrogens corresponded to BGC 5, which was similar to a published anabaenopeptin BGC (Table XI).[118] The protonated ion of $[M+H]^+$ = 828.421 is found in three anabaenopeptins: anabaenopeptin D, anabaenopeptin 827, and anabaenopeptin KVJ827 (Figure 56).[117,119,236] MS/MS on a Q-ToF mass spectrometer and analysis of the adenylation domain predictions was used to determine if the compound was known or new. The adenylation domain predictions of the BGC suggested the incorporation of an N-Me glycine at the R_5 position, eliminating the possibility of the compound being anabaenopeptin D or anabaenopeptin 827. The adenylation domain prediction at the R_1 position suggested incorporation of phenylalanine, eliminating the possibility of the compound being anabaenopeptin KVJ827.
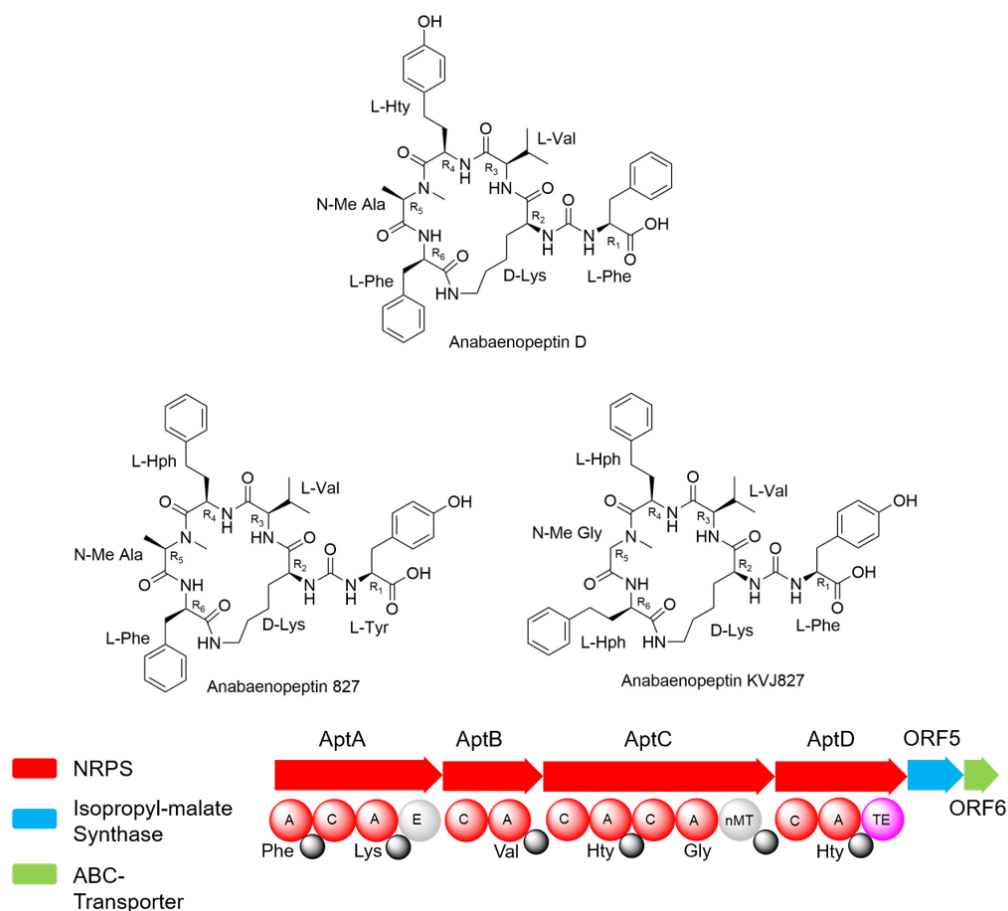
Figure 56: Dereplication of the new anabaenopeptin UIC827 by HRESIMS and adenylation domain prediction. BGC of new anabaenopeptin with adenylation domain predictions shown for dereplication purposes.

This suggested that BGC 5 produced a new anabaenopeptin analog. Analysis of the MS/MS spectrum confirmed the dereplication performed by adenylation domain predictions and demonstrated the compound was a new anabaenopeptin. Sequential losses of amino acids in the structure of the new anabaenopeptin allowed for a preliminary determination of the structure. Protonated ions at *m/z* 332.196 Da and 249.121 Da corresponded to the masses of Val-Hph-N-Me Gly and Hty-N-Me Gly,

respectively. A protonated ion of *m/z* 635.354 Da corresponds to the mass of the

compound with loss of the R_1 position of Phe and a carbonyl. Finally, a protonated ion

of *m/z* 84.081 Da, the immonium ion of Lys, confirmed the presence of the final amino

acid (Figure 57). The structure of the new compound is suggested to be a cyclic

peptide containing Lys-Val-Hph-N-Me Gly-Hty, with a urido bond between the Lys and

Phe located outside the ring. This structure is consistent with the well studied scaffold

of all other anabaenopeptin compounds, hexapeptides with a five amino acid ring and a

ureido bond to a sixth amino acid.
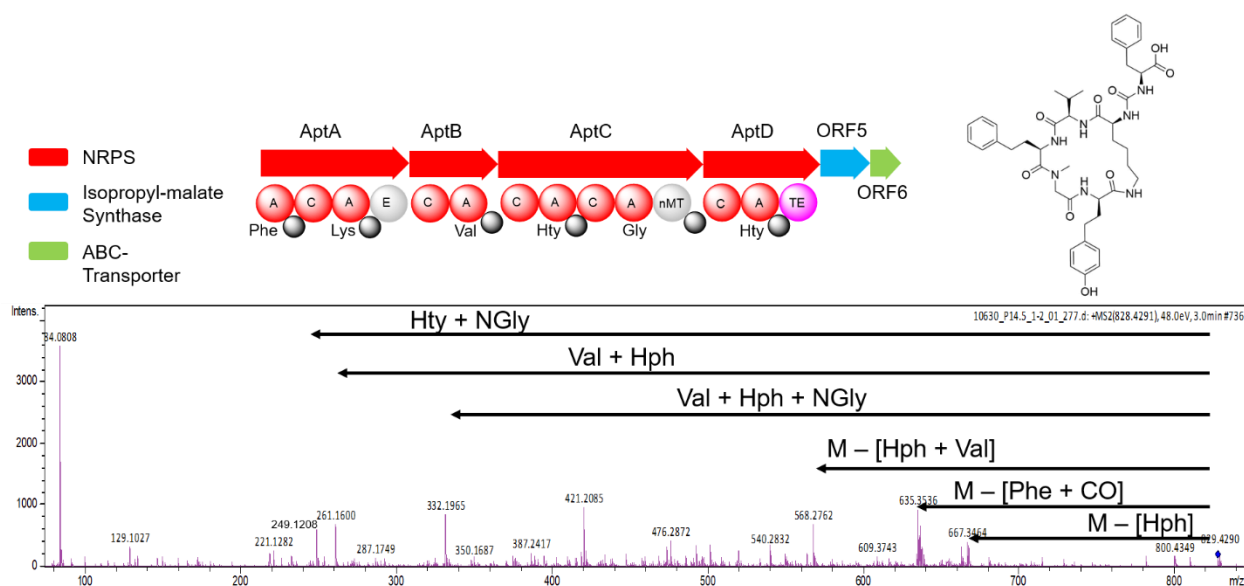
Figure 57: Structure elucidation of anabaenopeptin UIC827 based on adenylation domain substrate prediction and MS/MS fragmentation analysis

TABLE XI: PROTEINS ENCODED IN THE ANABAENOPEPTIN UIC827 BIOSYNTHETIC GENE CLUSTER INCLUDING LENGTH AND HOMOLOGY TO CORRESPONDING ANABAENOPEPTIN BIOSYNTHETIC GENES

Protein	Length	% Identity to Anabaenopeptin Biosynthesis Proteins
AptA	2196 aa	AptA 87%
AptB	1069 aa	AptB 88%
AptC	2578 aa	AptC 87%
AptD	1402 aa	AptD 88%
AptE	392 aa	AptE 84%
AptF	675 aa	AptF 84%

This new anabaenopeptin was present in low abundances in the extracts, and less than one mg was isolated from the cell extract of UIC 10630. Anabaenopeptins that contain an N-Me glycine at the R_5 position were reported by Grach-Pogrebinsky et al. to have two distinct conformational isomers due to the nearly equal energy states of the cis and trans isomer around the glycine residue.[237] The low abundance and the duplicated signals of the conformational isomers complicated structure confirmation of the new anabaenopeptin by NMR. Rather, Marfey's analysis was used to confirm the identity as well as the stereoconfiguration of each amino acid.[238] The compound was hydrolyzed by acid hydrolysis for 24 hours at 110 °C and subsequently reacted with L-FDLA to yield the liberated amino acids derivatized with the L-Marfey's reagent. The retention time of each of the derivatized amino acids was compared to L-amino acid standards reacted with both L-FDLA and D-FDLA to determine the stereoconfiguration of each of the sample amino acids.[238] All amino acids identified in the MS/MS analysis

were accounted for through Marfey's analysis, except glycine which was not examined

due to its lack of a stereocenter, and all except the lysine were in the L configuration

(Table XII). The stereoconfiguration is consistent with the structures of other

anabaenopeptins as well as the presence of an epimerization domain adjacent to the

Lys-selecting adenylation domain in the BGC. The MS/MS analysis and Marfey's

analysis confirmed the identity of each amino acid, while the adenylation domain

predictions and MS/MS analysis verified the amino acid sequence of the peptide and

confirmed the structure of the new anabaenopeptin UIC827 (**129**).
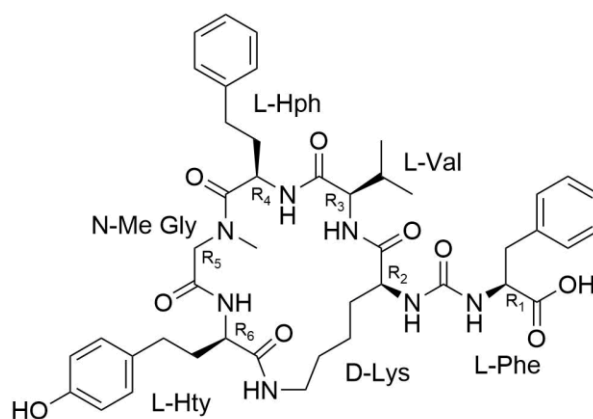
Anabaenopeptin UIC827 (**129**)

TABLE XII: MARFEY'S ANALYSIS OF HYDROLYZED AMINO ACIDS FROM ANABAENOPEPTIN UIC827. RETENTION TIMES FOR THE L AND D FDLA-L-AMINO ACID STANDARDS AND L-FDLA HYDROLYZED AMINO ACIDS ARE SHOWN WITH MATCHING RETENTION TIMES HIGHLIGHTED IN GREEN.

Amino Acid	Standard L Retention Time	Standard D Retention Time	Sample Retention Time
FDLA-Phe	5.14 min	6.27 min	5.16 min
FDLA-Lys	1.92 min	1.66 min	1.62 min
FDLA-Val	4.35 min	5.76 min	4.30 min
FDLA-Hph	5.59 min	6.80 min	5.61 min
FDLA-Hty	3.57 min	4.92 min	3.55 min

^{15}N Labeling Identifies and Aids in Structure Elucidation of Nostopyrrolidonamide

The compound containing three nitrogens corresponded to BGC7, a BGC with no known homologous gene clusters, making it especially interesting. The adenylation domain predictions suggested the compound was a small peptide and the molecular weight obtained by HRESIMS suggested a molecular formula of $C_{36}H_{49}N_3O_5$ with 14 double bond equivalents. Database and literature searches did not identify any matches for a compound with this molecular weight or formula, suggesting it was a new compound. The compound was isolated by reversed phase HPLC on a C18 column with a gradient of 85% ACN in water to 100% ACN for 15 minutes at 4 mL/min and eluted at 13.5 minutes. The prevalence of the compound in the extract and the ease of isolation allowed structure elucidation to be performed by NMR.

Adenylation and acyl transferase domain predictions suggested a fatty acid chain, an N-Me leucine, two phenylalanines, and a malonate unit to be present in the structure. A gene adjacent to the PKS gene encoded for an O-methyltransferase,

suggesting the presence of a methoxy group in the final compound (Figure 58 and

Table XIII).
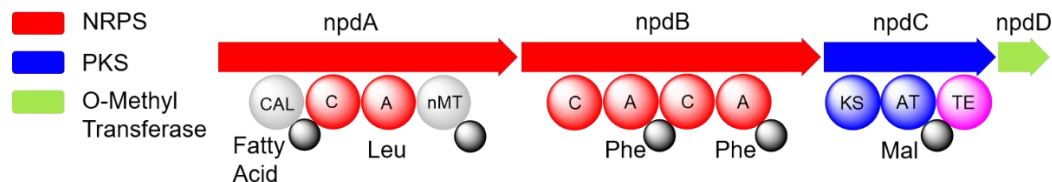
Figure 58: Putative biosynthetic gene cluster of nostopyrrolidonamide

TABLE XIII: *PROTEINS ENCODED IN THE NOSTOPYRROLIDONAMIDE BIOSYNTHETIC GENE CLUSTER INCLUDING LENGTH AND PROPOSED FUNCTION BASED ON BLASTP IDENTITY SCORE.*

Protein	Length	BLASTP Proposed Function Based on Identity Score
NpdA	2201 aa	NRPS
NpdB	2223 aa	NRPS
NpdC	1295 aa	Type I PKS
NpdD	345 aa	O-Methyl transferase

The length of the fatty acid chain was inferred to be an eight-carbon chain by

subtracting the mass of the N-Me leucine, two phenylalanines, malonate unit, and an

extra methyl group from the molecular weight. The eight-carbon chain was confirmed

by sequential COSY correlations from the diastereotopic H-7 to the terminal H-1 methyl

group. The leucine was likewise confirmed by sequential COSY correlations from the α-

proton H-9 to the two terminal methyl groups H-12 and H-13. HMBC correlations were used to confirm the N-Me leucine moiety and connect the fatty acid to the N-Me leucine with correlations from the methyl group H-35 to C-9 and C-8. A combination of COSY and HMBC correlations confirmed the presence of two phenylalanines, however the shifts for the α-protons, H-15 δ_H 5.86 and H-24 δ_H 4.9, were quite different from each other, suggesting unique chemical environments for each of them. HMBC correlations determined the methoxy group was placed on C-32 of the second phenylalanine with correlations from H-25 and H-36 to C-32, providing a potential explanation for the differences in the shifts of the two phenylalanine α-protons. HMBC correlations from H-24 to C-33 and from H-33 to C-34 connected the malonate unit to the second phenylalanine. The integration of H-33 and chemical shifts of H-33 and C-33 confirmed the presence of a double bond between C-32 and C-33 (Figure 59 and Table XIV). The linear structure had 13 double bond equivalents, one short of the required number for the molecular formula of $C_{36}H_{49}N_3O_5$ and suggested the presence of a ring.

Nostopyrrolidonamide (**130**)
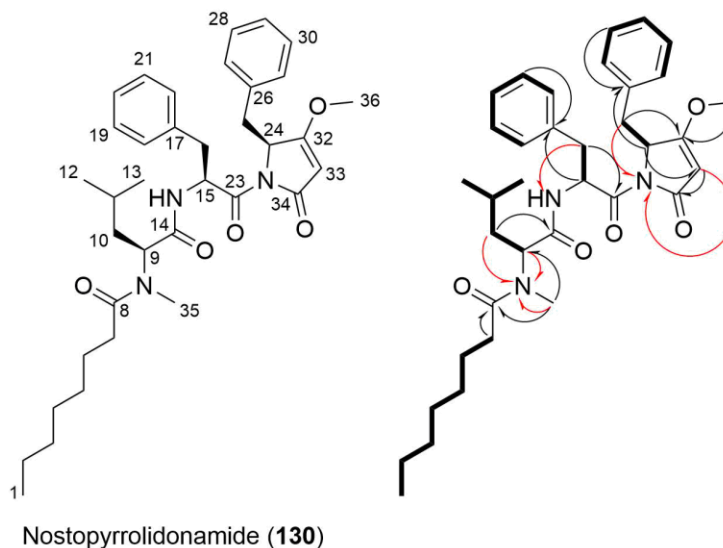
Figure 59: Key 2D NMR correlations of nostopyrrolidonamide. Black arrows indicate ^{13}C HMBC correlations, and red arrows indicate ^{15}N HMBC correlations.

A ring could reasonably be formed in one of two ways, both involving the formation of a fourth amide bond between the malonate carbonyl and one of the two phenylalanine nitrogens. An amide bond between C-34 and the nitrogen of the first phenylalanine would form an eight-membered ring, while an amide bond between C-34 and the nitrogen of the second phenylalanine would form a five-membered ring (Figure 60). None of the HMBC or ROESY correlations observed provided evidence for which amide bond was formed. Opportunely, the ^{15}N labeling used to identify the natural product in the cell extract also aided in solving the structure. The ^{15}N labeled compound was isolated from the labeled extract using the same method described above and was analyzed using ^{15}N HSQC and ^{15}N HMBC. The ^{15}N HSQC identified one secondary amine with a shift of δ_N 118.9. ^{15}N HMBC correlations from the beta methylene of first phenylalanine, H-16, to the secondary amine suggested that the ring could not be formed by an amide bond between C-34 and the first phenylalanine nitrogen, eliminating the possibility of an eight-membered ring. Furthermore, ^{15}N HMBC

correlations from H-25 and H-33 to a tertiary amine with the shift of δ_N 114.7 confirmed

the presence of the five-membered ring, completing the 2D structure of the new

compound nostopyrrolidonamide (**130**) (Figure 59, Table XIV).
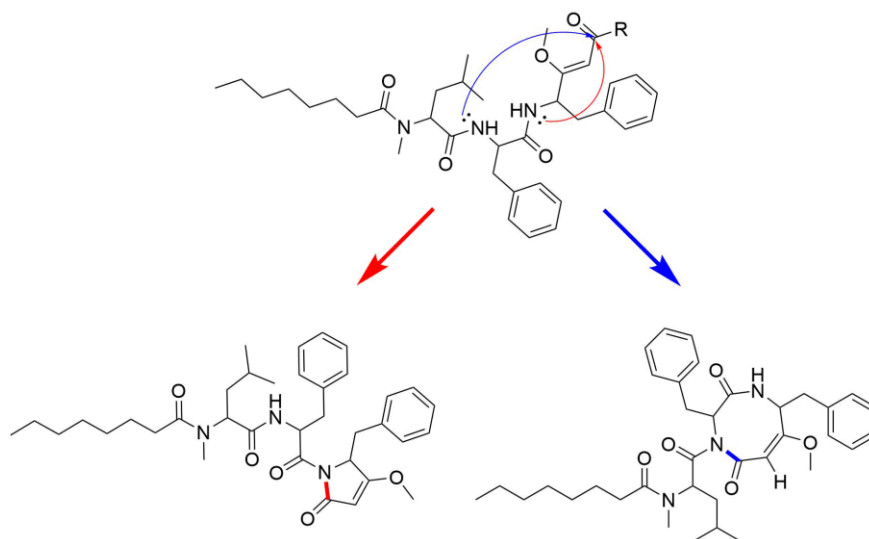
Figure 60: Diagram of two potential mechanisms for amide bond formation and ring closure.

The three stereocenters were determined using Marfey's method. Acid hydrolysis

followed by derivatization of the liberated amino acids with L-FDLA allowed a

comparison to standard amino acids derivatized with either L-FDLA or D-FDLA.[238] The

N-Me leucine and first phenylalanine were both found to be in the L configuration. Acid

hydrolysis was unable to open the pyrrolidinone ring and no molecular weight matching

a derivatized phenylalanine and malonate unit was observed. Ozonolysis followed by

oxidative work-up, acid hydrolysis, and derivatization with L-FDLA allowed identification

of only one stereoisomer for phenylalanine, suggesting both are L.

TABLE XIV: SPECTROSCOPIC DATA FOR NOSTOPYRROLIDONAMIDE (^{1}H 900 MHZ, ^{13}C 225 MHZ IN MEOH-D$_4$).

Position	δ_C, type	δ_H (*J* in Hz)
Nostopyrrolidonamide (**130**)
1	14.63, CH3	0.92, t (7.0)
2	23.9, CH2	1.35, m
3	33.1, CH2	1.33, m
4	30.6, CH2	1.28, m
5	30.5, CH2	1.37, m
6	26.31, CH2	1.59, m
7	34.6, CH2	2.33, 2.23, ddd (6.5, 8.6, 15.2)
8	176.5, C	
9	55.4, CH	5.22, dd (6.2, 9.2)
10	37.8, CH2	1.66, 1.51, m
11	26.1, CH	1.43, m
12	23.5, CH3	0.94, d (6.7)
13	22.4, CH	0.9, d (6.7)
14	172.6, C	
15	55.7, CH	5.86, dd (3.4, 10.8)
16	38.0, CH2	3.2, 2.67, dd (13.7, 3.5)
17	138.9, C	
18/22	130.52, CH	7.30, m
19/21	129.43, CH	7.25, m
20	127.86, CH	7.19, m
23	173.3, C	
24	61.3, CH	4.9, dd (3.1, 4.7)
25	35.1, CH2	3.57, 3.09, dd (14.0, 4.3)
26	135.1, C	
27/31	131.1, CH	7.0, m
28/30	129.47, CH	7.22, m
29	128.22, CH	7.19, m
32	180.7, C	
33	95.72, CH	4.98, s
34	172, C	
35	31, CH3	2.54, s
36	59.7, CH3	3.88, s

Nostopyrrolidonamide contains a pyrrolidinone ring, formed from a phenylalanine and a malonate unit. This moiety has been found in marine cyanobacterial natural products, appearing in the structures of dolastatin 15 (**131**) and belamide (**132**). A similar pyrrolidinone ring was also found in sintokamide A-E (**133**-**137**), sponge derived natural products that are thought to be produced by cyanobacteria (Figure 61). Both dolastatin 15 and belamide are cytotoxic compounds with microtubule depolymerization activity, although it is believed that the pyrrolidinone ring is not a critical part of the pharmacophore.[86,239] The sintokamides were reported to inhibit transactivation of the N-terminus of the androgen receptor in prostate cancer cells leading to inhibition of androgen receptor induced proliferation in androgen receptor sensitive cells.[240,241] Despite the pharmaceutically relevant activities of compounds with moieties similar to **130**, it was not found to be active in either an antiproliferative assay or an antibacterial assay. Future studies should be performed to determine the potential pharmaceutical relevance for this new compound.
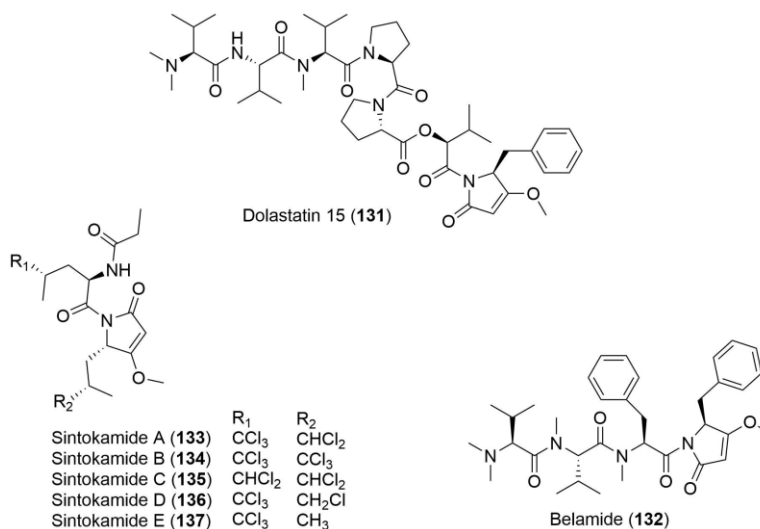
Figure 61: Bioactive natural products containing pyrrolidinone ring

Identification of Compounds with No Nitrogen and Orphan Natural Products

The products of the remaining BGCs could not be identified using the ^{15}N labeling and comparative metabolomics method described here. BGC1-3 were not predicted to produce compounds containing nitrogen, however, BGC1 and BGC2 have 100% homology to well-studied BGCs and the products may be identified by mass spectrometry. The presence of glycolipids, likely produced by BGC2, was confirmed during the comparative metabolomics study. An apparent increase in the production of a compound with a mass of 576.468 Da and a molecular formula of $C_{32}H_{64}O_8$ was observed in the cell extract of UIC 10630 when grown in $BG12_o$ media. This molecular formula is consistent with a heterocyst glycolipid known to be produced by *Nostoc* spp., 1-(O-hexose)-3,25-hexacosanediol.[100] The production of this heterocyst glycolipid from *Nostoc* spp. grown in media without a nitrogen source has been previously reported.[234]

The products of BGC 8 and 9 were not detected in either extraction method or the media when using the ^{15}N labeling and comparative metabolomic approach. It is possible that the compounds are not produced under the chosen growth conditions (Figure 62 and 63). For example, the predicted structure of BGC 8 contains many hydrophilic and charged amino acids, suggesting that it could be exuded into the media and used as a metal chelator, and could be induced by metal deprivation in the media (Figure 62). Future transcriptomic analysis should to be carried out to determine if the BGCs are being transcribed.
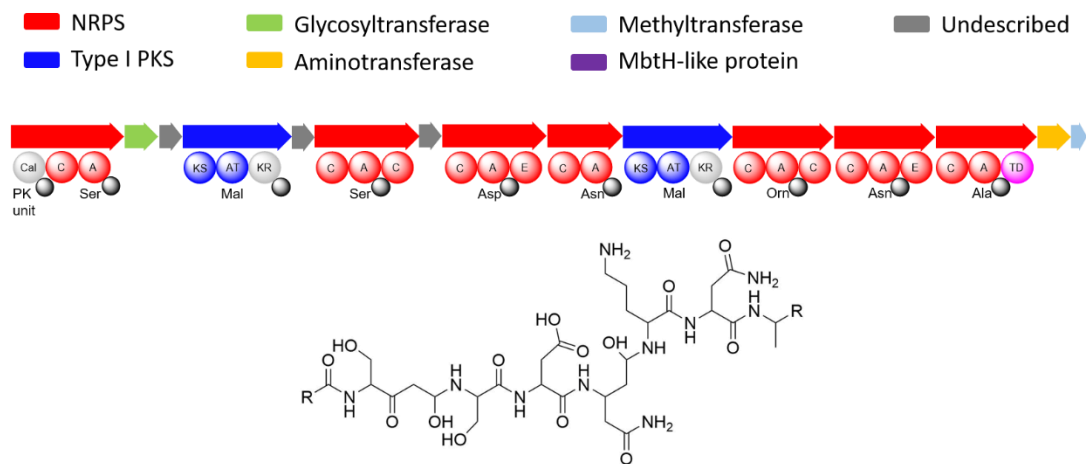
Figure 62: Orphan biosynthetic gene cluster 8 with predicted gene annotations and predicted structure of the produced compound.
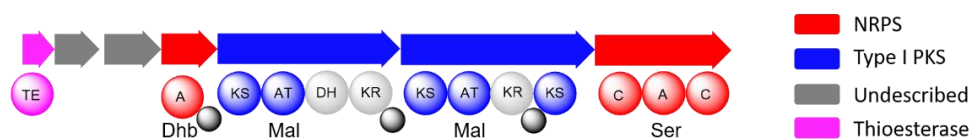
Figure 63: Orphan biosynthetic gene cluster 9 with predicted gene annotations

Conclusions

We developed an approach to match natural products to their cognate BGCs based on prior studies by Gross et al. and Kinnel et al.[229,230] Exchanging regular isotopic abundance nitrate with ^{15}N labeled nitrate in culture media, allowed full incorporation of ^{15}N into the produced natural products. Comparison of unlabeled cell extracts and ^{15}N labeled cell extracts allowed identification of compounds that contain nitrogen. In the case of UIC 10630, the number of nitrogens identified could be directly linked to the number of nitrogens predicted by the adenylation domains of BGCs. This approach was used to identify four compounds in the cell extract of UIC 10630 *Nostoc* sp., and each compound represents a useful aspect of this method. First, two of the identified compounds were known compounds, nostopeptolide A1 and aeruginosin 865. This method allowed rapid identification and dereplication of the known compounds within a cell extract. Second, the method was able to identify an orphan gene cluster, BGC 6, and match it with the production of a known natural product, aeruginosin 865. This demonstrated that this approach could be used for de-orphaning BGCs for which a known natural product exists. Third, the method was used to identify and isolate a new, low abundance analog of the anabaenopeptins, anabaenopeptin UIC827.

Anabaenopeptin UIC827 was not identified using the MALDIisotopic approach, likely due to the low abundance of the compound *in situ*. Finally, the ^{15}N labeling served a dual purpose in the identification and structure elucidation of the new compound nostopyrrolidonamide. This method was able to identify the product of a previously unknown BGC and aid in the structure elucidation of the isolated compound by enabling ^{15}N NMR experiments. Two compounds predicted to contain nitrogen could not be identified and further experiments should be performed to determine if they are transcribed or if they are produced under other unique culturing conditions. Similarly, compounds without nitrogen can not be detected using this approach. However, an approach similar to the one described by Gross et al., i.e. supplementation of culture media with predicted precursors, could help identify individual non-nitrogen containing compounds. In situations where a BGC does not follow the rule of collinearity or an adenylation domain prediction is ambiguous, a range of predicted nitrogens may need to be employed to link a BGC to a labeled compound.

Methods

Spectroscopic Instrumentation

UV spectra were recorded on Shimadzu UV spectrometer UV2301 on scanning mode from 190 to 360 nm. IR spectra were recorded with a JASCO FTIR-410 Fourier transform infrared spectrometer. COSY, HSQC, HMBC, and ROESY spectra were recorded on a Bruker Avance 800 MHz spectrometer, while ^{1}H, DEPTQ, ^{15}N HSQC, and ^{15}N HMBC were recorded on a Bruker Avance 900 MHz spectrometer. Chemical shifts in all NMR spectra were referenced to the residual solvent peak (MeOH-d_4 and MeOH-d_3 δ_H 3.31 and δ_C 49.15). HMBC spectra were acquired with average $^3J_{CH}$ of 8

Hz. ^{15}N HSQC spectrum was acquired with $^{1}J_{NH}$ of 90 Hz and ^{15}N HMBC spectrum was acquired with $^{3}J_{NH}$ of 6 Hz. Extract and pure compound HRESIMS spectra were recorded on a Shimadzu LC-IT-ToF mass spectrometer, in-situ and media MS spectra were recorded on a Bruker AutoFlex MALDI-ToF mass spectrometer, and MS/MS spectra were recorded on a Bruker Impact II LC-Q-ToF mass spectrometer. MS/MS spectra were obtained with a dissociation energy of 60 eV with stepping factors of 0.2, 0.8, and 1.

Biological Material

A soil collection was made at the University of Illinois at Chicago College of Pharmacy field station outside Downer's Grove, Illinois in 2015 (41° 48.8' N, 88° 2.3' W). The soil collection was incubated in BG12$_o$ media for four weeks before an individual filament was isolated via microisolation techniques and grown in Z media into the unialgal strain UIC 10630. UIC 10630 was grown in four 2 L cultures of Z medium, with constant sterile aeration and a 18/6 hour light/dark cycle for 8 weeks. The culture room was kept at 22 °C with all fluorescent lights emitting 2.6 klx. The cell mass was harvested by centrifugation and lyophilized.

Microscopic Taxonomic Analysis

UIC 10630 was preliminarily determined to be a *Nostoc* sp. based on microscopic analysis. The trichomes were unbranched, flexible, isopolar, and were coated in a thin sheath, with cells dividing in a single plane. The trichomes had intercalary heterocysts and barrel shaped vegetative cells. Hormogonia were observed and had terminal heterocysts. No akinetes were observed in the filament. Modern taxonomic systems from Komarek et al. 2014, and Bergey's Manual of Systematic

Bacteriology 2001 were followed in determining the preliminary taxonomy of UIC

10630.[64,93]

<u>DNA Extraction and Genome Sequencing</u>

  Genomic DNA was extracted from approximately 250 mg wet cell mass of UIC

10630. Cells were centrifuged, and the cell pellet was suspended in lysis buffer (2.5

mL, 10 mM Tris, 0.1 EDTA, 0.5% w/v SDS, 20 µg/mL pancreatic RNase pH 8.0) and

chicken egg white lysozyme (1 mg/mL). The mixture was incubated in a water bath at

37 $^{\circ}$C for one hour. Proteinase K, at a final concentration of 100 µg/mL, was added and

incubated in a water bath at 50 $^{\circ}$C for one hour. The pre-treated cells were pelleted via

centrifugation and the supernatant was removed. The pellet was resuspended in

nuclease free water and briefly vortexed to suspend epiphytic bacteria. The cells were

pelleted again, and the supernatant was removed. The Macherey-Nagel Nucleospin

Soil kit was used to extract genomic DNA from the pre-treated cells. Concentration and

purity were assessed using a nanodrop UV spectrophotometer. The genomic DNA was

sequenced by both Illumina and PacBio sequencing. A Nextera DNA Library

Preparation Kit was used to create sequencing libraries for the Illumina sequencing

which was performed with an Illumina MiSeq. PacBio sequencing was performed on

the original extracted genomic DNA. The resulting reads from both sequencing

technologies were trimmed using a quality score trimmer with a 0.01 limit and all reads

below 50 base pairs were discarded. A coverage filter was used to remove any

sequences with less than 50X coverage and a hybrid assembly was performed on the

remaining sequences. The resulting draft assembly was submitted to AntiSMASH 3.0.

<u>16S rRNA Amplification and Taxonomic Analysis</u>

A portion of the 16S rRNA gene was amplified by PCR from the genomic DNA isolated from UIC 10630. Cyanobacterial specific 16S rRNA primers 109F and 1509R were used in the PCR reaction. The reaction contained 2 µL of both primers at 10 µM, 1 µL of dNTPs at 10 µM, 0.5 µL of Phusion high-fidelity polymerase, 2 µL of the genomic DNA at 25 ng/µL, 10 µL of the Phusion Buffer, and 32.5 µL of nuclease-free water. The reaction was performed on an Applied Biosystems 2720 thermal cycler with a program of denaturation for 2 minutes at 95 ^{o}C, followed by 35 cycles of 95 ^{o}C for 30 seconds, 49 ^{o}C for 30 seconds, 72 ^{o}C for 2 minutes, and a final extension of 72 ^{o}C for 5 minutes. Agarose gel electrophoresis was used to confirm the amplification and a GeneJet PCR purification kit (Thermo) was used to purify the product. Sanger sequencing was performed with the cyanobacterial specific primers 109F and 1509R, as well as an internal primer 359F. The resulting sequence was manually examined and trimmed. The partial 16S rRNA sequence of UIC 10630 used in the phylogenetic analysis can be found on GenBank with the accession code: MG992485. Cyanobacterial 16S sequences from the Nostocaceae family of at least 1Kb, were obtained from GenBank, including several Bergey's Manual reference strains. The 16S rRNA sequence of *Gloeobacter violaceus* was used as an outgroup. These were aligned using ClustalW with default gap opening and extension penalties. The multiple sequence alignment was used in conjunction with a neighbor-joining tree, using 1000 bootstrap replicates, to infer the evolutionary distances between strains.

Labeling Methods and Extraction

Ingredient lists for the media used (Z, BG12, BG12$_o$, Z^{15}N, and BG12^{15}N can all be found in appendix D. UIC 10630 was grown in duplicate in each media for 5 weeks in 250 mL flasks in the culture room at 22 °C with all fluorescent lights emitting 2.6 klx. The cell masses were harvested by centrifugation and lyophilized. The dried cell masses of each duplicate were combined and extracted with a 1:1 mix of dichloromethane and methanol. The lipophilic extract was dried in vacuo. The spent cell mass was then dried and extracted again with a 1:1 mix of methanol and water to yield a hydrophilic extract that was also dried in vacuo. 150 mL of the spent media, post-harvest, was lyophilized to yield a media extract. Portions of the lipophilic extract and hydrophilic extract were filtered by Extract Clean C18 cartridges, prior to analysis by the LC-IT-ToF mass spectrometer.

Comparative Mass Spectrometry Parameters

Extracts for UIC 10630 in each of the five media were analyzed by HRESIMS on a Shimadzu LC-IT-ToF. A Phenomenex Kinetex C18 RPLC column (length 50 mm x 2.1 mm 1.7 μm particle size 100 Å pore size) was used for the analysis. The separation method used was a gradient from 5% ACN with 0.1% formic acid in water to 100% ACN with 0.1% formic acid for 7 minutes. The mass spectrometer parameters were event 1 (positive mode, scan range 250–1200 m/z, event time 127 ms); event 2 (positive mode, scan range 900–2000 m/z, event time 300 ms), event 3 (negative mode, scan range 250–1200 m/z, event time 127 ms), event 4 (negative mode, scan range 900–2000 m/z, event time 300 ms).

<u>Extraction and Isolation of Compounds from UIC 10630</u>

The four 2L flask culture of UIC 10630 was lyophilized and extracted by a 1:1

mixture of methanol and dichloromethane. The extract was separated by vacuum liquid

chromatography over a diaion HP20 column with a step-wise gradient of IPA in water.

The separation yielded six fractions in 100% water, 20% IPA, 40% IPA, 60% IPA, 70%

IPA, and 100% IPA. The masses identified by the comparative metabolomic approach

were used to identify and isolate the compounds from the fractions. The compounds

were isolated by reversed-phase HPLC on a Varian Microsorb Dynamax C18

semipreparative column 250 mm X 10 mm. The method used to isolate

nostopyrrolidonamide from the 70% IPA fraction was a gradient from 85% ACN to 100%

ACN for 15 minutes at 4 mL/min. Nostopyrrolidonamide eluted at 13.5 minutes and was

the major component of the fraction. Aeruginosin 865 and anabaenopeptin UIC827

were present in the 40% IPA fraction at low abundances and were isolated by a

gradient from 10% ACN to 100% ACN for 30 minutes at 4 mL/min. Anabaenopeptin

UIC827 eluted at 14.5 minutes and aeruginosin 865 eluted at 15.5 minutes. All peaks

were dried in vacuo.

<u>Spectroscopic Data</u>

Anabaenopeptin UIC827 white amorphous powder; $[\alpha]^{25}_D$ - 12 (*c* 0.09, MeOH); UV

(MeOH) λ_{max} (log ε) 280 (3.1) nm; ^{1}H see Appendix A; HRESIMS *m/z* 828.4213 [M+H] [+]

(calcd for $C_{44}H_{58}N_7O_9$, 828.4291).

Nostopyrrolidonamide amorphous solid; $[\alpha]^{25}_D$ + 8.3 (*c* 0.36, MeOH); UV (MeOH) λ_{max}

(log ε) 272 (2.4), 280 (2.2), 327 (1.8) nm; IR (neat) v_{max} 2954, 2930, 1727, 1680, 1627,

1378 1307, 996 cm^{-1}; ^{1}H and ^{13}C NMR data see Table XIV; HRESIMS *m/z* 604.3803 [M+H]$^+$ (calcd for $C_{36}H_{50}N_3O_5$, 604.3750), spectra available in Appendix A.

Cytotoxicity Assay

MDA-MB-435 and MDA-MB-231 cells were harvested by trypsinization during log phase growth. Cells were grown in 96-well plates at 37 °C in 5% CO_2. Cultures were incubated overnight and treated with samples, positive control (paclitaxel) or negative control (DMSO) for 72 hours. Cell viability was assessed by CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega). IC_{50} values were calculated by comparison to the negative control.

Antibacterial Assay

Bacterial cultures of *Pseudomonas aeruginosa*, *Enterococcus facaelis*, and *Escherichia coli* ΔlptD were grown at 37 °C until log phase growth, determined by OD_{600} measurements. Bacterial cells were diluted to 0.1 OD_{600} and seeded into 96 well plates. The bacteria were treated with a serial dilution of the compounds or positive control (streptomycin or kanamycin) and incubated overnight. MIC values were determined by incubation of the wells with alamar blue for two hours and visual inspection of metabolism of the dye.

6   CONCLUSIONS AND DISCUSSION

Conclusions

    Nostoc is a polyphyletic genus that encompasses a wide variety of natural product producing cyanobacteria. These *Nostoc* spp. can be found in freshwater, terrestrial, and symbiotic environments all around the planet. Their ability to perform photosynthesis, unique nitrogen fixation strategy, and genetic capacity to produce natural products have made them the focus of several biological studies. Despite the number of studies of *Nostoc* spp, they remain understudied as a source of natural products as compared to other bacteria such as *Streptomyces* or *Moorea*.

    Amplicon sequencing experiments with a cyanobacterial specific primer set demonstrated that cyanobacteria are abundant in soil samples (Figure 32 and Figure 34). Soil samples can be obtained at any time of the year and are easier to transport than macroscopic blooms of cyanobacteria. Incubation of soil samples in cyanobacterial specific media enabled growth of cyanobacteria from the soil. Different cyanobacterial specific media were capable of selecting for different genera. $BG12_o$ media consistently and with statistical significance increased the relative abundance of heterocystous genera of cyanobacteria, including *Nostoc*, a genus known for production of natural products (Figure 37 and Appendix C). One *Nostoc* strain, UIC 10630, isolated from a soil sample was studied for its genetic capacity to produce natural products and resulted in the isolation of two new compounds. Additional *Nostoc* strains obtained from soil collections during this project will continue to be screened in antiproliferative assays through the Orjala Lab workflow.

    Screening *Nostoc* spp. strains from the UIC cyanobacteria culture collection for antiproliferative activity in cell-based assays lead to the identification of a series of [7,7]

paracyclophanes. Known [7,7] paracyclophanes were identified from UIC 10448, UIC

10250, and UIC 10534, while new [7,7] paracyclophanes were identified from UIC

10110, UIC 10279, and UIC 10366. Ribocyclophanes A-E (**123**-**126**, **128**) were isolated

from the cell extract of UIC 10279 and UIC 10366 (Figure 48). The ribocyclophanes

represent a new group of [7,7] paracyclophanes in the cylindrocyclophane core

structural group, containing a β-D-ribose in the α position. The ribose adopts a unique

and thermodynamically unfavorable $^1C_4$ conformation that has not been reported in

other natural products. The unfavorable conformation is likely adopted due to the steric

effects of the branched β-methyl group and the resorcinol ring (Figure 49).

Ribocyclophane E (**128**) is likely a biosynthetic intermediate of ribocyclophane D (**126**),

as it has a similar structure to the cylindrofridins, and corroborates the reported

biosynthesis of the [7,7] paracyclophanes. New merocyclophanes were isolated from

UIC 10110 and represent the first reported analogs of the merocyclophane core

structural group (Figure 42). Merocyclophanes have only been reported by the Orjala

Lab, which provided a unique opportunity to identify the likely biosynthetic origin of the

merocyclophane α-methyl group. A putative merocyclophane BGC was identified in the

sequenced genome of UIC 10110, and a biosynthetic mechanism based on reports

from the Balskus lab was proposed (Figure 15 and 44).[133,135] Analysis of the type I PKS

genes identified a C-methyl transferase domain as the probable source of the α-methyl

group present in the merocyclophane core structure (Figure 43).

The [7,7] paracyclophanes isolated from the UIC culture collection have been

tested in antiproliferation and antibacterial assays. Most [7,7] paracyclophanes show

modest activity in these assays with low micromolar activity; however, some of the

compounds appear to be particularly active such as ribocyclophane D (**126**), merocyclophane C (**113**), and merocyclophane D (**114**). Merocyclophane C was further analyzed in a preliminary *in vivo* assay, the hollow fiber tumor assay. The initial assay did not show statistically significant antiproliferative activity with a p value of 0.051 for merocyclophane C and in a repeated experiment the compound proved too toxic to the mice to continue. Overall the diverse [7,7] paracyclophanes identified by a phenotypic screening approach have demonstrated the wealth of chemical diversity that can be produced by *Nostoc* spp. for a single biologically relevant chemical scaffold.

Reidentification of analogs of chemical scaffolds is one of the inherent downfalls of phenotypic screening natural product discovery platforms. Genome Mining represents a method to identify the full natural product producing potential of individual strains, however a common challenge of this method is the identification of the products of BGCs in a cell extract. An approach to identify the products of BGCs in a cyanobacterial cell extract was developed using ^{15}N labeling. [15]N labeling with labeled nitrate or ammonium has previously been used as a way to determine the nitrogen content of cyanobacterial natural products and more recently as a way to identify nitrogen containing compounds.[162,242,243] The full incorporation of the ^{15}N label into nitrogen containing compounds allowed matching of *m/z* values with their respective BGCs by comparing the predicted and observed number of nitrogens for each compound. The method was used to identify the products of four BGCs in the extract of UIC 10630 *Nostoc* sp. (Figure 53). Two of the compounds, nostopeptolide A1 (**77**) and aeruginosin 865 (**16**) were known compounds, however the BGC for aeruginosin 865 had never been described. A preliminary examination of the gene cluster that likely

encodes for aeruginosin 865 was performed and identified putative genes for incorporation of the glycoside and fatty acid (Figure 55). This demonstrated the dereplication and BGC de-orphaning abilities of this method.

The other two compounds identified by the method were new compounds. Anabaenopeptin UIC827 (**129**) is a new anabaenopeptin present in low abundance in the cell extract of UIC 10630. It could not be detected by the Maldiisotopic approach described by Kinnel et al. likely due to the low abundance *in situ*, however due to the enrichment in the extract method it was able to be detected by our method.[230] The new structure was determined by adenylation domain predictions, MS/MS and Marfey's method (Figure 57). The other new compound was more abundant in the extract and was solved using HRESIMS and NMR. The complete 2D structure of Nostopyrrolidonamide (**130**) could not be solved using a standard set of 2D NMR experiments due to the pyrrolidinone moiety. The ^{15}N label provided a dual purpose in both the identification of the compound in the extract and the structure elucidation in the form of ^{15}N NMR experiments. ^{15}N HSQC and ^{15}N HMBC using the labeled compound completed the structure of the pyrrolidinone ring and the 2D structure of the compound (Figure 59). The stereocenters were determined by Marfey's method. The genome mining approach revealed classes of compounds not previously identified in the Orjala lab: nostopeptolides, aeruginosins, and anabaenopeptins, as well as a new compound. Together, these chapters have shown that using both phenotypic screening and genome mining approaches for natural product discovery will allow a more comprehensive search for new natural products from freshwater cyanobacteria.

Discussion

New Sources of Cyanobacteria

Cyanobacteria are widespread and inhabit many different unique environments. Isolation of cyanobacteria solely from macroscopic environmental collections likely limits the taxonomic diversity of the UIC culture collection, which could in turn limit the chemical diversity isolated from the cultured strains. The cyanobacterial specific primer set used in the amplicon sequencing studies revealed the wealth of cyanobacteria present in environmental collections with no macroscopic colonies. Some of these cyanobacteria were able to be isolated from these soil collections and have since been added to the culture collection. However, it was shown that not all cyanobacteria present in the soil are easily grown by incubating the soil in media under light. The studies should be repeated using solid agar and other media types to see if changing these parameters could select for additional cyanobacteria. Other sources of cyanobacteria should also be sought out to add to the culture collection. Cyanobacteria, especially *Nostoc*, are known to be present in symbiosis with fungi, angiosperms, and bryophytes.[102,105,110,151] Many of these symbioses are facultative for the cyanobacterium, allowing it to be cultured without the symbiotic host.[110] The selectivity of the cyanobacterial specific primer set for cyanobacterial sequences over chloroplastic sequences should aid in the identification of cyanobacterial symbiotic partners. Symbiotic bacteria are also known have unique natural products not often found in non-symbiotic bacteria (see nosperin Figure 20).[151] Similar to the cyanobacteria present in the soil, some cyanobacteria are also known to grow as endoliths in harsh environments.[96,193] It is likely that these could also be detected using

the specific primer set and potentially grown from the rocks using cyanobacterial specific media. Isolating cyanobacteria from more varied environments should help increase the chemical diversity present in the cultured strains.

Cyanobacterial Taxonomy Revision

The polyphyletic nature of *Nostoc* is also common among other cyanobacterial genera. The long evolutionary history and a use of both botanical nomenclature and prokaryotic nomenclature led cyanobacterial taxonomy to be complicated.[64] Many genera that are morphologically similar clade separately when analyzed by 16S rRNA sequence. This can be seen clearly in the genus *Nostoc*. *Nostoc* strains are split into at least five recognized clades by the *Bergey's Manual of Systematic Bacteriology* including the two discussed in chapters 3-5.[93] *Nostoc* group I and *Nostoc* group III are two of the larger clades containing many strains available from strain libraries but are clearly two separate genera. 16S rRNA analysis provides evidence for the separation of these two clades as well as the chemotaxonomic analysis performed in chapter 4. *Nostoc* group II was recently designated as the *genus novum Desmonostoc* and it is likely that with future studies of these cyanobacteria, group I and group III will similarly be separated.[94] Work underway in the Orjala Lab to sequence a partial 16S rRNA sequence of each of the over one thousand strains in the UIC culture collection will provide an invaluable resource for the editing of the polyphyletic genera commonly encountered from our environmental collections.

Identification of Molecular Targets of Cyanobacterial Natural Products

The antiproliferative screening approach and genome mining approach utilized in this dissertation work are both unable to identify the molecular targets of the natural

products that they identify. The molecular target of the [7,7] paracyclophanes has proved especially elusive and is one of the last unknown aspects of these unique compounds. Attempts to select for spontaneous mutants of Δ*lptD E. coli* on [7,7] paracyclophane agar plates failed to produce any mutants that could be sequenced for identification of resistance conferring mutations in the genes of potential targets. Similarly, screening of several [7,7] paracyclophanes in the FuSiOn molecular profiling platform and a yeast knock-out library provided conflicting results, suggesting no clear consensus on a molecular target of these compounds.[244,245] It is possible that the [7,7] paracyclophanes act as detergents and disrupt the cell membranes of both bacteria and eukaryotic cells, but the rigid chemical scaffold and range of bioactivities strongly suggest a molecular target. Similarly, any compound identified by genome mining will have an unknown biological activity unless a close analog of the compound is known, as is the case with the protease inhibiting anabaenopeptins. Another potential way to identify targets of compounds discovered by genome mining is through the detection of duplicated genes near the BGC that could convey self-resistance, however this did not yield any results for [7,7] paracyclophanes or nostopyrrolidonamide (**130**). Other methods for discovering the molecular targets of these compounds should continue and will aid in determining how these natural products could be used as pharmaceuticals or as molecular probes.

Final Perspectives

Current and emerging technologies have been improving both the sequencing and bioinformatic technologies as well as analytical chemistry technologies in ways that will greatly increase the productivity of natural product discovery. Using both genome

mining approaches combined with traditional screening approaches and emerging analytical chemistry techniques, many new cyanobacterial natural products will be identified in the future. It is apparent that remaining challenges such as identifying the molecular targets of these natural products will need a renewed focus as many more compounds are discovered. The genus *Nostoc,* however, remains a prolific source of new natural products and continued investigation of this bacteria will likely produce more unique and biologically relevant natural products.

REFERENCES

(1)   Croteau, R.; Kutchan, T. M.; Lewis, N. G. *Biochem. Mol. Biol. Plants* **2000**, *7* (7), 1250–1318.

(2)   Balandrin, Manuel F, Kinghorn, A Douglas, Farnsworth, N. R. *J. Nat. Prod.* **1997**, *60* (1), 52–60.

(3)   Swann, J. P. *Br. Soc. Hist. Sci.* **1983**, *16* (2), 154–190.

(4)   Quinn, R. *Am. J. Public Health* **2013**, *103* (3), 426–434.

(5)   Koehn, F. E.; Carter, G. T. *Nat. Rev. Drug Discov.* **2005**, *4* (3), 206–220.

(6)   Newman, D. J.; Giddings, L. A. *Phytochem. Rev.* **2014**, *13* (1), 123–137.

(7)   Cragg, G. M.; Grothaus, P. G.; Newman, D. J. *Chem Rev* **2009**, *109* (7), 3012–3043.

(8)   Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79* (3), 629–661.

(9)   Shu, Y.-Z. *J. Nat. Prod.* **1998**, *61* (8), 1053–1071.

(10) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nat. Rev. Drug Discov.* **2015**, *14* (2), 111–129.

(11) Harvey, A. *Drug Discovery Today*. **2000**, 294–300.

(12) Luo, Y.; Cobb, R. E.; Zhao, H. *Curr. Opin. Biotechnol.* **2014**, 230–237.

(13) Shatz, A.; Bugle, E.; Waksman, S. A. *Clin. Orthop. Relat. Res.* **1944**, No. 437, 3.6.

(14) Lewis, K. *Nature* **2012**, *485* (7399), 439–440.

(15) Corley, D. G.; Durley, R. C. *J. Nat. Prod.* **1994**, *57* (11), 1484–1490.

(16) Cordell, G. A.; Shin, Y. G. *Pure Appl. Chem* **1999**, *71* (6), 1089–1094.

(17) Qiu, F.; Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S. N.; Pauli, G. F. *J. Nat. Prod.* **2012**, *75* (3), 432–443.

(18) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J. Nat. Prod.* **2013**, *76* (9), 1686–1699.

(19) Sica, V. P.; Raja, H. A.; El-elimat, T.; Kertesz, V.; Berkel, G. J. Van; Pearce, C. J.; Oberlies, N. H. *J. Nat. Prod.* **2015**.

(20) Medema, M. H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M. a.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* **2011**, *39* (SUPPL. 2), 339–346.

(21) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. a.; Muller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, 1–7.

(22) Brady, S. F.; Simmons, L.; Kim, J. H.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26* (11), 1488.

(23) Micallef, M. L.; D'Agostino, P. M.; Al-Sinawi, B.; Neilan, B. a.; Moffitt, M. C. *Mar. Genomics* **2014**, *21*, 1–12.

(24) 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.; Glöckner, F. O. *Nat. Chem. Biol.* **2015**, *11* (9), 625–631.

(25) Strieker, M.; Tanović, A.; Marahiel, M. A. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 234–240.

(26) Süssmuth, R. D.; Mainz, A. *Angew. Chemie - Int. Ed.* **2017**, *56* (14), 3770–3821.

(27) Sieber, S. A.; Marahiel, M. A. *Chem. Rev.* **2005**, *105* (2), 715–738.

(28) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6* (8), 493–505.

(29) Hur, G. H.; Vickery, C. R.; Burkart, M. D. **2012**, *829* (10), 1074–1098.

(30) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G.

R.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510* (7506), 512–517.

(31) Shen, B. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285–295.

(32) Khosla, C. *J. Org. Chem.* **2009**, *74* (17), 6416–6420.

(33) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. *Nat. Prod. Rep.* **2009**, *26* (1), 90–114.

(34) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. *Nat. Prod. Rep.* **2007**, *24* (1), 162–190.

(35) Olano, C.; Méndez, C.; Salas, J. A. *Nat. Prod. Rep.* **2010**, *27* (4), 571.

(36) Yu, D.; Xu, F.; Zeng, J.; Zhan, J. *IUBMB Life* **2012**, *64* (4), 285–295.

(37) Shimizu, Y.; Ogata, H.; Goto, S. *ChemBioChem* **2017**, *18* (1), 50–65.

(38) Van Der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30* (1), 108–160.

(39) Ortega, M. A.; Van Der Donk, W. A. *Cell Chem. Biol.* **2016**, *23* (1), 31–44.

(40) Sardar, D.; Pierce, E.; McIntosh, J. A.; Schmidt, E. W. *ACS Synth. Biol.* **2015**, *4* (2), 167–176.

(41) Sardar, D.; Lin, Z.; Schmidt, E. W. *Chem. Biol.* **2015**, *22* (7), 907–916.

(42) Verpoorte, R. *Drug Discovery Today.* **1998**, 232–238.

(43) Molinski, T. F.; Dalisay, D. S.; Lievens, S. L.; Saludes, J. P. *Nat. Rev. Drug Discov.* **2009**, *8* (1), 69–85.

(44) Gerwick, W. H.; Moore, B. S. *Chem. Biol.* **2012**, *19* (1), 85–98.

(45) Cardani, C.; Ghiringhelli, D.; Mondelli, R.; Quilioo, A. *Tetrahedron.* **1965**, 2537–2545.

(46) Piel, J. *Nat. Prod. Rep.* **2009**, *26* (3), 338–362.

(47) Oh, D.-C.; Poulsen, M.; Currie, C.; Clardy, J. *Nat. Chem. Biol.* **2009**, *5* (6), 391–393.

(48) Ochoa, J. L.; Sanchez, L. M.; Koo, B.-M.; Doherty, J. S.; Rajendram, M.; Huang, K. C.; Gross, C. A.; Linington, R. G. *ACS Infect. Dis.* **2017**, acsinfecdis.7b00105.

(49) Gabriel, C. R.; Northup, D. E.; Cheeptham, N.; Northup, D. E.; Dapkevicius, M. D. L. N. E. *Cave Microbiomes* **2013**, *1*, 85–108.

(50) Derewacz, D. K.; McNees, C. R.; Scalmani, G.; Covington, C. L.; Shanmugam, G.; Marnett, L. J.; Polavarapu, P. L.; Bachmann, B. O. *J. Nat. Prod.* **2014**, *77* (8), 1759–1763.

(51) Pawlowski, A. C.; Wang, W.; Koteva, K.; Barton, H. A.; McArthur, A. G.; Wright, G. D. *Nat. Commun.* **2016**, *7*, 1–10.

(52) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. *Nature* **2015**, *517* (7535), 455–459.

(53) Cragg, G. M.; Newman, D. J. *Biochim. Biophys. Acta - Gen. Subj.* **2013**, *1830* (6), 3670–3695.

(54) Lewis, K.; Epstein, S.; D'Onofrio, A.; Ling, L. L. *J. Antibiot. (Tokyo).* **2010**, *63* (8), 468–476.

(55) Stewart, E. J. *J. Bacteriol.* **2012**, *194* (16), 4151–4160.

(56) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. *Proc. Natl. Acad. Sci.* **2017**, *114* (22), 5601–5606.

(57) Fenical, W.; Jensen, P. R. *Nat. Chem. Biol.* **2006**, *2* (12), 666–673.

(58) Herrmann, J.; Fayad, A. A.; Müller, R. *Nat. Prod. Rep.* **2017**, *34* (2), 135–160.

(59) Welker, M.; Dittmann, E.; Von Döhren, H. *Methods Enzymol.* **2012**, *517*, 23–46.

(60) Winter, J. M.; Behnken, S.; Hertweck, C. *Curr. Opin. Chem. Biol.* **2011**, *15* (1), 22–31.

(61) Donadio, S.; Monciardini, P.; Sosio, M. *Nat. Prod. Rep.* **2007**, *24* (5), 1073.

(62) Whitton, B. A. *Ecology of Cyanobacteria II: Their Diverisity in Space and Time*, II.; Whitton, B. A., Ed.; Springer: New York, **2012**.

(63) Komarek, J. *Algae* **2006**, *21* (4), 349–375.

(64) Komarek, J.; Kastovsky, J.; Mares, J.; Johansen, J. R. *Preslia* **2014**, *86* (4), 295–335.

(65) Rippka, R.; Waterbury, J.; Cohen-Bazire, G. *Arch. Microbiol.* **1974**, *100* (1), 419–436.

(66) Korelusová, J. Phylogeny of heterocytous Cyanobacteria (Nostocales and Stigonematales), **2008**.

(67) Shestakov, S. V.; Karbysheva, E. A. *Biol. Bull. Rev.* **2017**, *7* (4), 259–272.

(68) Shih, P. M. *Curr. Biol.* **2015**, *25* (5), R192–R193.

(69) MacColl, R. *J. Struct. Biol.* **1998**, *124* (2–3), 311–334.

(70) Cohen, Y.; Gurevitz, M. *The Prokaryotes* **2006**, *4*, 1074–1098.

(71) Rae, B. D.; Long, B. M.; Whitehead, L. F.; Förster, B.; Badger, M. R.; Price, G. D. *J. Mol. Microbiol. Biotechnol.* **2013**, *23* (4–5), 300–307.

(72) Berman-Frank, I.; Lundgren, P.; Falkowski, P. *Res.Microbiol.* 2003, pp 157–164.

(73) Bryceson, I.; Fay, P. *Mar. Biol.* **1981**, *61* (2–3), 159–166.

(74) Zhang, C. C.; Laurent, S.; Sakr, S.; Peng, L.; Bédu, S. *Mol. Microbiol.* **2006**, *59* (2), 367–375.

(75) Mitsui, A.; Kumazawa, S.; Takahashi, A.; Ikemoto, H.; Cao, S.; Arai, T. *Nature* **1986**, *323* (6090), 720–722.

(76) Meeks, J. C.; Elhai, J.; Meeks, J. C.; Elhai, J. *Microbiol. Mol. Biol. Rev.* **2002**, *66* (1), 94–121.

(77) Esteves-Ferreira, A. A.; Cavalcanti, J. H. F.; Vaz, M. G. M. V.; Alvarenga, L. V.; Nunes-Nesi, A.; Araújo, W. L. *Genet. Mol. Biol.* **2017**, *40* (1), 261–275.

(78) Dittmann, E.; Gugger, M.; Sivonen, K.; Fewer, D. P. *Trends Microbiol.* **2015**, *23* (10), 642–652.

(79) Van Wagoner, R. M.; Drummond, A. K.; Wright, J. L. C. 2007; pp 89–217.

(80) Merel, S.; Walker, D.; Chicana, R.; Snyder, S.; Baurès, E.; Thomas, O. *Environment International.* **2013**, 303–327.

(81) FRANCIS, G. *Nature* **1878**, *18* (444), 11–12.

(82) Carmichael, W. W. *J. Appl. Bacteriol.* **1992**, *72* (6), 445–459.

(83) Steffen, M. M.; Davis, T. W.; McKay, R. M. L.; Bullerjahn, G. S.; Krausfeldt, L. E.; Stough, J. M. A.; Neitzey, M. L.; Gilbert, N. E.; Boyer, G. L.; Johengen, T. H.; Gossiaux, D. C.; Burtner, A. M.; Palladino, D.; Rowe, M. D.; Dick, G. J.; Meyer, K. A.; Levy, S.; Boone, B. E.; Stumpf, R. P.; Wynne, T. T.; Zimba, P. V.; Gutierrez, D.; Wilhelm, S. W. *Environ. Sci. Technol.* **2017**, *51* (12), 6745–6755.

(84) Shih, P. M.; Wu, D.; Latifi, A.; Axen, S. D.; Fewer, D. P.; Talla, E.; Calteau, A.; Cai, F.; Tandeau de Marsac, N.; Rippka, R.; Herdman, M.; Sivonen, K.; Coursin, T.; Laurent, T.; Goodwin, L.; Nolan, M.; Davenport, K. W.; Han, C. S.; Rubin, E. M.; Eisen, J. a; Woyke, T.; Gugger, M.; Kerfeld, C. a. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (3), 1053–1058.

(85) De Claro, R. A.; McGinn, K.; Kwitkowski, V.; Bullock, J.; Khandelwal, A.; Habtemariam, B.; Ouyang, Y.; Saber, H.; Lee, K.; Koti, K.; Rothmann, M.; Shapiro, M.; Borrego, F.; Clouse, K.; Chen, X. H.; Brown, J.; Akinsanya, L.; Kane, R.; Kaminskas, E.; Farrell, A.; Pazdur, R. *Clin. Cancer Res.* **2012**, *18* (21), 5845–5849.

(86) Pettit, G. R.; Kamano, Y.; Herald, C. L.; Fujii, Y.; Kizu, H.; Boyd, M. R.; Boettner, F. E.; Doubek, D. L.; Schmidt, J. M.; Chapuis, J. C.; Michel, C. *Tetrahedron* **1993**, *49* (41), 9151–9170.

(87) Luesch, H.; Moore, R. E.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H. *J. Nat. Prod.* **2001**, *64* (7), 907–910.

(88) Vaishampayan, U.; Glode, M.; Du, W.; Kraft, A.; Hudes, G.; Wright, J.; Hussain, M. *Clin. Can* **2000**, *6*, 4205–4208.

(89) Perez, E. a; Hillman, D. W.; Fishkin, P. a; Krook, J. E.; Tan, W. W.; Kuriakose, P.

a; Alberts, S. R.; Dakhil, S. R. *Invest. New Drugs* **2005**, *23* (3), 257–261.

(90) Senter, P. D.; Sievers, E. L. *Nat. Biotechnol.* **2012**, *30* (7), 631–637.

(91) Dodds, W. K.; Gudder, D. A.; Mollenhauer, D. *J. Phycol.* **1995**, *31* (1), 2–18.

(92) Palinska, K. A.; Surosz, W. *Hydrobiologia* **2014**, *740* (1), 1–11.

(93) Castenholz, R. Boone, D., Castenholz, R., Eds.; Springer: New York, **2001**; 473–599.

(94) Hrouzek, P.; Lukešová, A.; Mareš, J.; Ventura, S. *Fottea* **2013**, *13* (2), 201–213.

(95) Pandey, K. D.; Shukla, S. P.; Shukla, P. N.; Giri, D. D.; Singh, J. S.; Singh, P.; Kashyap, A. K. *Cell. Mol. Biol. (Noisy-le-grand).* **2004**, *50* (5), 575–584.

(96) Wynn-Williams, D. D.; Edwards, H. G. M.; Garcia-Pichel, F. *Eur. J. Phycol.* **1999**, *34* (4), 381–391.

(97) Mateo, P.; Perona, E.; Berrendero, E.; Leganés, F.; Martín, M.; Golubić, S. *FEMS Microbiol. Ecol.* **2011**, *76* (2), 185–198.

(98) Hrouzek, P.; Ventura, S.; Lukešová, A.; Mugnai, M. A.; Turicchia, S.; Komárek, J. *Arch. Hydrobiol. Suppl. Algol. Stud.* **2005**, *117* (1), 251–264.

(99) Meeks, J. C.; Campbell, E. L.; Summers, M. L.; Wong, F. C. *Arch. Microbiol.* **2002**, *178* (6), 395–403.

(100) Bauersachs, T.; Compaoré, J.; Hopmans, E. C.; Stal, L. J.; Schouten, S.; Sinninghe Damsté, J. S. *Phytochemistry* **2009**, *70* (17–18), 2034–2039.

(101) Sutherland, J. M.; Herdman, M.; Stewart, W. D. P. *J. Gen. Microbiol.* **1979**, *115*, 273–287.

(102) Rasmussen, U.; Bergman, B.; Johansson, C. *Molecular Plant-Microbe Interactions.* **1994**, 696–702.

(103) Liaimer, A.; Helfrich, E. J. N.; Hinrichs, K.; Guljamow, A.; Ishida, K.; Hertweck, C.; Dittmann, E. *Proc. Natl. Acad. Sci.* **2015**, *112* (6), 1862–1867.

(104) Bergman, B.; Hallbom, L. *Can. J. Bot.* **1982**, *60* (10), 2092–2098.

(105) Bergman, B.; Rai, A. *Physiol. Plant.* **1989**, *77* (2), 216–224.

(106) Herdman, M.; Janvier, M.; Rippka, R.; Stanier, R. Y. *J. Gen. Microbiol.* **1979**, *111* (1), 73–85.

(107) Videau, P.; Wells, K. N.; Singh, A. J.; Gerwick, W. H.; Philmus, B. *ACS Synth. Biol.* **2016**, *5* (9), 978–988.

(108) Beck, C.; Henningknoopbiologiehu-berlinde, H. K.; Axmann, I. M.; Steuer, R. **2012**, 1–17.

(109) Ishida, K.; Welker, M.; Christiansen, G.; Cadel-Six, S.; Bouchier, C.; Dittmann, E.; Hertweck, C.; De Marsac, N. T. *Appl. Environ. Microbiol.* **2009**, *75* (7), 2017–

2026.

(110) Liaimer, A.; Jensen, J. B.; Dittmann, E. *Front. Microbiol.* **2016**, *7* (NOV), 1–16.

(111) Liu, L.; Jokela, J.; Wahlsten, M.; Nowruzi, B.; Permi, P.; Zhang, Y. Z.; Xhaard, H.; Fewer, D. P.; Sivonen, K. *J. Nat. Prod.* **2014**, *77* (8), 1784–1790.

(112) Jokela, J.; Heinilä, L. M. P.; Shishido, T. K.; Wahlsten, M.; Fewer, D. P.; Fiore, M. F.; Wang, H.; Haapaniemi, E.; Permi, P.; Sivonen, K. *Front. Microbiol.* **2017**, *8* (OCT), 1–14.

(113) Kapuścik, A.; Hrouzek, P.; Kuzma, M.; Bártová, S.; Novák, P.; Jokela, J.; Pflüger, M.; Eger, A.; Hundsberger, H.; Kopecký, J. *ChemBioChem* **2013**, *14* (17), 2329–2337.

(114) Fewer, D. P.; Jokela, J.; Rouhiainen, L.; Wahlsten, M.; Koskenniemi, K.; Stal, L. J.; Sivonen, K. *Mol. Microbiol.* **2009**, *73* (5), 924–937.

(115) Drapalova, P.; Stys, D.; Lukesova, A.; Kopecky, J. *Arch. Hydrobiol. Suppl. Algol. Stud.* **2008**, *127* (August), 61–82.

(116) Sanz, M.; Andreote, A. P. D.; Fiore, M. F.; Dörr, F. A.; Pinto, E. *Mar. Drugs* **2015**, *13* (6), 3892–3919.

(117) Shishido, T. K.; Jokela, J.; Fewer, D. P.; Wahlsten, M.; Fiore, M. F.; Sivonen, K. *ACS Chem. Biol.* **2017**, *543*, acschembio.7b00570.

(118) Rouhiainen, L.; Jokela, J.; Fewer, D. P.; Urmann, M.; Sivonen, K. *Chem. Biol.* **2010**, *17* (3), 265–273.

(119) Guljamow, A.; Kreische, M.; Ishida, K.; Liaimer, A.; Altermark, B.; Bähr, L.; Hertweck, C.; Ehwald, R.; Dittmann, E. *Appl. Environ. Microbiol.* **2017**, *83* (23).

(120) Okino, T.; Qi, S.; Matsuda, H.; Murakami, M.; Yamaguchi, K. *J. Nat. Prod.* **1997**, *60* (2), 158–161.

(121) Kaya, K.; Sano, T.; Beattie, K. A.; Codd, G. A. *Tetrahedron Lett.* **1996**, *37* (37), 6725–6728.

(122) Hastie, C. J.; Borthwick, E. B.; Morrison, L. F.; Codd, G. A.; Cohen, P. T. W. *Biochim. Biophys. Acta - Gen. Subj.* **2005**, *1726* (2), 187–193.

(123) Pluotno, A.; Carmeli, S. *Tetrahedron* **2002**, *58*, 9949–9957.

(124) Mehner, C.; Müller, D.; Kehraus, S.; Hautmann, S.; Gütschow, M.; König, G. M. *Chembiochem* **2008**, *9* (16), 2692–2703.

(125) Herfindal, L.; Myhren, L.; Kleppe, R.; Krakstad, C.; Selheim, F.; Jokela, J.; Sivonen, K.; Døskeland, S. O. *Mol. Pharm.* **2011**, *8* (2), 360–367.

(126) Liu, L.; Jokela, J.; Herfindal, L.; Wahlsten, M.; Sinkkonen, J.; Permi, P.; Fewer, D. P.; Døskeland, S. O.; Sivonen, K. *ACS Chem. Biol.* **2014**, *9* (11), 2646–2655.

(127) Jokela, J.; Herfindal, L.; Wahlsten, M.; Permi, P.; Selheim, F.; Vasconçelos, V.;

Døskeland, S. O.; Sivonen, K. *ChemBioChem* **2010**, *11* (11), 1594–1599.

(128) Golakoti, T.; Yoshida, W. Y.; Chaganty, S.; Moore, R. E. *J. Nat. Prod.* **2001**, *64* (1), 54–59.

(129) Luesch, H.; Hoffmann, D.; Hevel, J. M.; Becker, J. E.; Golakoti, T.; Moore, R. E. *J. Org. Chem.* **2003**, *68* (1), 83–91.

(130) Moore, B. S.; Chen, J.; Patterson, G. M. L.; Moore, R. E. *Tetrahedron* **1992**, *48* (15), 3001–3006.

(131) Kang, H. S.; Santarsiero, B. D.; Kim, H.; Krunic, A.; Shen, Q.; Swanson, S. M.; Chai, H.; Kinghorn, a. D.; Orjala, J. *Phytochemistry* **2012**, *79*, 109–115.

(132) Chen, J. L.; Moore, R. E.; Patterson, G. M. L. *J. Org. Chem.* **1991**, *56* (10 mL), 4360–4364.

(133) Nakamura, H.; Hamer, H. a.; Sirasani, G.; Balskus, E. P. *J. Am. Chem. Soc.* **2012**, *134* (45), 18518–18521.

(134) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H. J.; Schneefeld, M.; Herrmann, J.; Kumpfmüller, J.; Thürmer, A.; Neidhardt, I.; Wiesner, C.; Daniel, R.; Müller, R.; Bange, F. C.; Schmieder, P.; Schweder, T.; Mundt, S. *Mar. Drugs* **2016**, *14* (1).

(135) Nakamura, H.; Schultz, E. E.; Balskus, E. P. *Nat. Chem. Biol.* **2017**, *13* (June).

(136) Nakamura, H.; Wang, J. X.; Balskus, E. P. *Chem. Sci.* **2015**, *6* (7), 3816–3822.

(137) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. *J. Nat. Prod.* **2010**, *73* (9), 1529–1537.

(138) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. *J. Nat. Prod.* **2006**, *69* (10), 1379–1383.

(139) Luo, S.; Kang, H.-S.; Krunic, A.; Chlipala, G. E.; Cai, G.; Chen, W.-L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. *Tetrahedron Lett.* **2014**, *55* (January), 686–689.

(140) Preisitsch, M.; Harmrolfs, K.; Pham, H. T.; Heiden, S. E.; Füssel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H.; Müller, R.; Mundt, S. *J. Antibiot. (Tokyo).* **2014**, No. May 2014, ASAP article, DOI: 10.1038/ja.2014.118.

(141) Preisitsch, M.; Niedermeyer, T. H. J.; Heiden, S. E.; Neidhardt, I.; Kumpfmüller, J.; Wurster, M.; Harmrolfs, K.; Wiesner, C.; Enke, H.; Müller, R.; Mundt, S. *J. Nat. Prod.* **2016**, *79* (1), 106–115.

(142) Hutter, R.; Keller-Schierlein, W.; Knusel, F.; Prelog, V.; Rodgers, G.; Suter, P.; Vogel, G.; Voser, W.; Zahner, H. *Helv. Chim. acta.* **1967**, *50* (6), 1533–1539.

(143) Okami, Y.; Okazaki, T.; Kithara, T.; Umezawa, H. *J. Antibiot. (Tokyo).* **1976**, *29* (10), 1019–1025.

(144) Hemscheidt, T.; Puglisi, M. P.; Larsen, L. K.; Patterson, G. M. L.; Moore, R. E.;

Rios, J. L.; Clardy, J. *J. Org. Chem.* **1994**, *59* (12), 3467–3471.

(145) Fujii, K.; Sivonen, K.; Kashiwagi, T.; Hirayama, K.; Harada, K. I. *J. Org. Chem.* **1999**, *64* (16), 5777–5782.

(146) Fewer, D. P.; Österholm, J.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Sivonen, K. *Appl. Environ. Microbiol.* **2011**, *77* (22), 8034–8040.

(147) Kurmayer, R. *J. Phycol.* **2011**, *47* (1), 200–207.

(148) Golakoti, T.; Yoshida, W. Y.; Chaganty, S.; Moore, R. E. *Tetrahedron.* **2000**, 9093–9102.

(149) Hoffmann, D.; Hevel, J. M.; Moore, R. E.; Moore, B. S. *Gene* **2003**, *311* (1–2), 171–180.

(150) Voráčová, K.; Hájek, J.; Mareš, J.; Urajová, P.; Kuzma, M.; Cheel, J.; Villunger, A.; Kapuscik, A.; Bally, M.; Novák, P.; Kabeláč, M.; Krumschnabel, G.; Lukeš, M.; Voloshko, L.; Kopecký, J.; Hrouzek, P. *PLoS One* **2017**, *12* (3), 1–20.

(151) Kampa, A.; Gagunashvili, A. N.; Gulder, T. A. M.; Morinaka, B. I.; Daolio, C.; Godejohann, M.; Miao, V. P. W.; Piel, J.; Andresson, O. S. *Proc. Natl. Acad. Sci.* **2013**, *110* (33), E3129–E3137.

(152) Trimurtulu, G.; Ohtani, I.; Patterson, G. M. L.; Moore, R. E.; Corbett, T. H.; Valeriote, F. A.; Demchik, L. *J. Am. Chem. Soc.* **1994**, *116* (11), 4729–4737.

(153) Golakoti, T.; Ogino, J.; Heltzel, C. E.; Husebo, T. Le; Jensen, C. M.; Larsen, L. K.; Patterson, G. M. L.; Moore, R. E.; Mooberry, S. L.; Corbett, T. H.; Valeriote, F. A. *J. Am. Chem. Soc.* **1995**, *117* (49), 12030–12049.

(154) Smith, C. D.; Zhang, X.; Mooberry, S. L.; Patterson, G. M. L.; Moore, R. E. *Cancer Res.* **1994**, *54* (14), 3779–3784.

(155) Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y. S.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. *1* (12).

(156) Zhang, X. *J. Biol. Chem.* **1996**, *271* (11), 6192–6198.

(157) Kerksiek, K.; Mejillano, M. R.; Schwartz, R. E.; Georg, G. I.; Himes, R. H. *FEBS Lett.* **1995**, *377* (1), 59–61.

(158) Edelman, M. *Lung Cancer* **2003**, *39* (2), 197–199.

(159) Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J. *Appl. Microbiol. Biotechnol.* **2010**, *86* (5), 1213–1225.

(160) Banker, R.; Carmeli, S. *J. Nat. Prod.* **1998**, *61* (10), 1248–1251.

(161) Jüttner, F.; Todorova, A. K.; Walch, N.; Von Philipsborn, W. *Phytochemistry* **2001**, *57* (4), 613–619.

(162) Todorova, A. K.; Jüttner, F.; Linden, A.; Plüiss, T.; von Philipsborn, W. *J. Org. Chem.* **1995**, *60* (24), 7891–7895.

(163) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4* (6), 341–343.

(164) Giglio, S.; Jiang, J.; Saint, C. P.; Cane, D. E.; Monis, P. T. *Environ. Sci. Technol.* **2008**, *42* (21), 8027–8032.

(165) Agger, S. A.; Lopez-Gallego, F.; Hoye, T. R.; Schmidt-Dannert, C. *J. Bacteriol.* **2008**, *190* (18), 6084–6096.

(166) Oldfield, E.; Lin, F. Y. *Angew. Chemie - Int. Ed.* **2012**, *51* (5), 1124–1137.

(167) Jaki, B.; Orjala, J.; Sticher, O. *J. Nat. Prod.* **1999**, *62* (3), 502–503.

(168) Jaki, B.; Orjala, J.; Heilmann, J.; Linden, A.; Vogler, B.; Sticher, O. *J. Nat. Prod.* **2000**, *63* (3), 339–343.

(169) Asthana, R. K.; Deepali; Tripathi, M. K.; Srivastava, A.; Singh, A. P.; Singh, S. P.; Nath, G.; Srivastava, R.; Srivastava, B. S. *J. Appl. Phycol.* **2009**, *21* (1), 81–88.

(170) Jaki, B.; Heilmann, J.; Sticher, O. *J. Nat. Prod.* **2000**, *63* (9), 1283–1285.

(171) Zelík, P.; Lukšeová, A.; Čejka, J.; Buděšínský, M.; Havlíček, V.; Čegan, A.; Kopecký, J. *J. Enzyme Inhib. Med. Chem.* **2010**, *25* (3), 414–420.

(172) Cheel, J.; Bogdanová, K.; Ignatova, S.; Garrard, I.; Hewitson, P.; Kolář, M.; Kopecký, J.; Hrouzek, P.; Vacek, J. *Algal Res.* **2016**, *18*, 244–249.

(173) Hrbac, J.; Jakubec, P.; Halouzka, V.; Matejka, P.; Pour, M.; Kopecky, J.; Vacek, J. *Electrochem. commun.* **2014**, *38*, 53–56.

(174) Ferroni, L.; Klisch, M.; Pancaldi, S.; Häder, D. P. *Mar. Drugs* **2010**, *8* (1), 106–121.

(175) Soule, T.; Garcia-Pichel, F.; Stout, V. *J. Bacteriol.* **2009**, *191* (14), 4639–4646.

(176) Proteau, P. J.; Gerwick, W. H.; Garcia-Pichel, F.; Castenholz, R. *Experientia* **1993**, *49* (9), 825–829.

(177) Sorrels, C. M.; Proteau, P. J.; Gerwick, W. H. *Appl. Environ. Microbiol.* **2009**, *75* (14), 4861–4869.

(178) Balskus, E. P.; Walsh, C. T. *J. Am. Chem. Soc.* **2008**, *130* (46), 15260–15261.

(179) Soule, T.; Stout, V.; Swingley, W. D.; Meeks, J. C.; Garcia-Pichel, F. *J. Bacteriol.* **2007**, *189* (12), 4465–4472.

(180) Balskus, E. P.; Walsh, C. T. *Biol. Chem.* **2009**, 14648–14649.

(181) Ferreira, D.; Garcia-Pichel, F. *Front. Microbiol.* **2016**, *7* (MAY), 1–10.

(182) Kobayashi, A.; Kaiiyama, S. ichiro; Inawaka, K.; Kanzaki, H.; Kawazu, K. *Zeitschrift fur Naturforsch. - Sect. C J. Biosci.* **1994**, *49* (7–8), 464–470.

(183) Sesin, D. F. United States patent 4,845,085, 1989.

(184) Weiss, C.; Figueras, E.; Borbely, A. N.; Sewald, N. *J. Pept. Sci.* **2017**, *23* (7–8),

514–531.

(185) Field, J. J.; Kanakkanthara, A.; Miller, J. H. *Bioorganic Med. Chem.* **2015**, *22* (18), 5050–5059.

(186) Liang, J.; Moher, E. D.; Moore, R. E.; Hoard, D. W. *J. Org. Chem.* **2000**, *65* (10), 3143–3147.

(187) Tan, L. T. *J. Appl. Phycol.* **2010**, *22* (5), 659–676.

(188) Tan, L. T. *Phytochemistry* **2007**, *68* (7), 954–979.

(189) Chlipala, G.; Mo, S.; Carcache de Blanco, E. J.; Ito, A.; Bazarek, S.; Orjala, J. *Pharm. Biol.* **2009**, *47* (1), 53–60.

(190) Chlipala, G.; Mo, S.; Carcache De Blanco, E. J.; Ito, A.; Bazarek, S.; Orjala, J. *Pharm. Biol.* **2009**, *47* (1), 53–60.

(191) Falch, BS; Konig, GM; Wright, AD; Sticher, O; Angerhofer, CK; Pezzuto, JM; Bachmann. *Planta Med.* **1995**, *61* (4), 321–328.

(192) De La Torre, J. R.; Goebel, B. M.; Friedmann, E. I.; Pace, N. R. *Appl Env. Microbiol* **2003**, *69* (7), 3858–3867.

(193) Dong, H.; Rech, J. A.; Jiang, H.; Sun, H.; Buck, B. J. *J. Geophys. Res. Biogeosciences* **2007**, *112* (2), 1–11.

(194) Fox, G.; Stackebrandt, E.; Hespell, R.; Gibson, J.; Maniloff, J.; Dyer, T.; Wolfe, R.; Balch, W.; Tanner, R.; Magrum, L.; Zablen, L.; Blakemore, R.; Gupta, R.; Bonen, L.; Lewis, B.; Stahl, D.; Luehrsen, K.; Chen, K.; Woese, C. *Science* **1980**, *209* (4455), 457–463.

(195) Paradis, E.; Claude, J.; Strimmer, K. *Bioinformatics* **2004**, *20* (2), 289–290.

(196) Yang, B.; Wang, Y.; Qian, P. Y. *BMC Bioinformatics* **2016**, *17* (1), 1–8.

(197) Turner, S.; Pryer, K. M.; Miao, V. P. W.; Palmer, J. D. *J. Eukaryot. Microbiol.* **1999**, *46* (4), 327–338.

(198) Wintzingerode, F.; Göbel, U. B.; Stackebrandt, E. *FEMS Microbiol. Rev.* **1997**, *21*, 213–229.

(199) Patterson, G. M. L.; Baldwin, C.; Bolis, C.; Caplan, F.; Karuso, H.; Larsen, L.; Levine, I.; Moore, R. E.; Nelson, C.; Tschappat, K.; Tuang, G. *J. Phycol.* **1991**, *27*, 530–536.

(200) Moore, R. E.; Patterson, G. M. L.; Carmichael, W. W. *Mem. Calif. Acad. Sci.* **1988**, *13*, 143–150.

(201) Lange, L.; Meyer, A.; Budel, B. **2018**, *8* (1), 52–57.

(202) Belnap, J.; Gardner, J. *West. North Am. Nat.* **1993**, *53* (1), 40–47.

(203) Nübel, U.; Muyzer, G.; Garcia-pichel, F.; Muyzer, G. *Microbiology* **1997**, *63* (8),

3327–3332.

(204) Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A. *Bioinformatics* **2014**, *30* (5), 614–620.

(205) Edgar, R. C. *Bioinformatics* **2010**, *26* (19), 2460–2461.

(206) McDonald, D.; Price, M. N.; Goodrich, J.; Nawrocki, E. P.; Desantis, T. Z.; Probst, A.; Andersen, G. L.; Knight, R.; Hugenholtz, P. *ISME J.* **2012**, *6* (3), 610–618.

(207) Bhute, S. S.; Pande, P.; Shetty, S. A.; Shelar, R.; Mane, S.; Kumbhare, S. V.; Gawali, A.; Makhani, H.; Navandar, M.; Dhotre, D.; Lubree, H.; Agarwal, D.; Patil, R.; Ozarkar, S.; Ghaskadbi, S. S.; Yajnik, C. S.; Juvekar, S.; Makharia, G. K.; Shouche, Y. S.; Thompson, J. D.; Higgins, D. G.; Gibson, T. J.; Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S.; Wattam, A. R.; Davis, J. J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E. M.; Disz, T.; Gabbard, J. L.; Gerdes, S.; Henry, C. S.; Kenyon, R. W.; Machi, D.; Mao, C.; Nordberg, E. K.; Olsen, G. J.; Murphy-Olson, D. E.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G. D.; Shukla, M.; Vonstein, V.; Warren, A.; Xia, F.; Yoo, H.; Stevens, R. L.; Langille, M. G. I.; Zaneveld, J.; Caporaso, J. G.; Mcdonald, D.; Knights, D.; Reyes, J. A.; Clemente, J. C.; Burkepile, D. E.; Vega Thurber, R. L.; Knight, R.; Beiko, R. G.; Huttenhower, C.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, K.; Gordon, J. I.; Huttley, G. a; Kelley, S. T.; Knights, D.; Jeremy, E.; Ley, R. E.; Lozupone, C. a; Mcdonald, D.; Muegge, B. D.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. a; Marathe, N.; Shetty, S. A.; Lanjekar, V. B.; Ranade, D. R.; Shouche, Y. S.; Suryavanshi, M. V.; Bhute, S. S.; Jadhav, S. D.; Bhatia, M. S.; Gune, R. P.; Shouche, Y. S.; Suryavanshi, M. V.; Joshi, S. M.; Yajnik, C. S.; Shouche, Y. S.; Ghaskadbi, S. S.; Turnbaugh, P. J.; Hamady, M.; Yatsunenko, T.; Cantarel, B. L.; Duncan, A.; Ley, R. E.; Sogin, M. L.; Jones, W. J.; Roe, B. A.; Affourtit, J. P.; Egholm, M.; Henrissat, B.; Heath, A. C.; Knight, R.; Gordon, J. I.; Steinhoff, U.; Sears, C. L.; Guarner, F.; Malagelada, J. R.; Björkstén, B.; Sepp, E.; Julge, K.; Voor, T.; Mikelsaar, M.; Cummings, J. H.; MacFarlane, G. T.; Flint, H. J.; Scott, K. P.; Louis, P.; Duncan, S. H.; Round, J. L.; Mazmanian, S. K.; Inman, M.; Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, S.; Manichanh, C.; Nielsen, T.; Pons, N.; Yamada, T.; Mende, D. R.; Li, J.; Xu, J.; Li, S. S.; Li, D.; Cao, J.; Wang, B.; Liang, H.; Zheng, H.; Xie, Y.; Tap, J.; Lepage, P.; Bertalan, M.; Batto, J.; Hansen, T.; Paslier, D. Le; Linneberg, A.; Nielsen, H. B.; Pelletier, E.; Renault, P.; Zhou, Y.; Li, Y.; Zhang, X.; Li, S. S.; Qin, N.; Yang, H.; Carabotti, M.; Scirocco, A.; Maselli, M. A.; Severi, C.; Ppatil, D.; Pdhotre, D.; Gchavan, S.; Sultan, A.; Jain, D. S.; Lanjekar, V. B.; Gangawani, J.; Sshah, P.; Stodkar, J.; Shah, S.; Ranade, D. R.; Patole, M. S.; Shouche, Y. S.; Furness, J. B.; Callaghan, B. P.; Rivera, L. R.; Cho, H.; Hooper, L. V; Littman, D. R.; Macpherson, A. J.; Program, M. P. *Nature* **2011**, *9* (5), 1–15.

(208) Tikhonov, M.; Leach, R. W.; Wingreen, N. S. *ISME J.* **2015**, *9* (1), 68–80.

(209) National Center for Health Statistics. *Cent. Dis. Control* **2017**, 314–317.

(210) Da Rocha, A. B.; Lopes, R. M.; Schwartsmann, G. *Curr. Opin. Pharmacol.* **2001**, *1* (4), 364–369.

(211) Altmann, K.-H.; Gertsch, J. *Nat. Prod. Rep.* **2007**, *24* (2), 327–357.

(212) Chlipala, G. E.; Krunic, A.; Mo, S.; Sturdy, M.; Orjala, J. *J. Chem. Inf. Model.* **2011**, *51* (1), 171–180.

(213) Bobzin, S.; Moore, R. *Tetrahedron* **1993**, *49* (35), 7615–7626.

(214) 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.; Pevzner, P. a. *J. Comput. Biol.* **2012**, *19* (5), 455–477.

(215) Posehn, S. E.; Kim, S. Y.; Wee, A. G. H.; Suh, D. Y. *ChemBioChem* **2012**, *13* (15), 2212–2217.

(216) Ren, Y.; Chen, W.-L.; Lantvit, D. D.; Sass, E. J.; Shriwas, P.; Ninh, T. N.; Chai, H.-B.; Zhang, X.; Soejarto, D. D.; Chen, X.; Lucas, D. M.; Swanson, S. M.; Burdette, J. E.; Kinghorn, A. D. *J. Nat. Prod.* **2016**, acs.jnatprod.6b00924.

(217) May, D. S.; Chen, W.; Lantvit, D. D.; Zhang, X.; Krunic, A.; Burdette, J. E.; Eustaquio, A.; Orjala, J. *J. Nat. Prod.* **2017**, *80* (4), 1073–1080.

(218) Vijayalakshmi, K. S.; Rao, V. S. R. *Carbohydr. Res.* **1972**, *22* (2), 413–424.

(219) Kabsch, W. *J. Appl. Crystallogr.* **1993**, *26* (6), 795–800.

(220) Sheldrick, G. M. *Acta Crystallogr.* **2008**, *64* (1), 112–122.

(221) Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 175–184.

(222) Vincent, A. T.; Derome, N.; Boyle, B.; Culley, A. I.; Charette, S. J. *J. Microbiol. Methods* **2017**, *138*, 60–71.

(223) Metzker, M. L. *Nat. Rev. Genet.* **2010**, *11* (1), 31–46.

(224) Calteau, A.; Fewer, D. P.; Latifi, A.; Coursin, T.; Laurent, T.; Jokela, J.; Kerfeld, C. A.; Sivonen, K.; Piel, J.; Gugger, M. *BMC Genomics* **2014**, *15* (1), 1–14.

(225) Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2017**, *44* (4–5), 573–588.

(226) Roberts, R. J.; Carneiro, M. O.; Schatz, M. C. *Genome Biol.* **2013**, *14* (6), 2–5.

(227) Ziemert, N.; Alanjary, M.; Weber, T. *Nat. Prod. Rep.* **2016**, *33* (8), 988–1005.

(228) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. a; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43* (W1), W237-43.

(229) Gross, H.; Stockwell, V. O.; Henkels, M. D.; Nowak-Thompson, B.; Loper, J. E.; Gerwick, W. H. *Chem. Biol.* **2007**, *14* (1), 53–63.

(230) Kinnel, R. B.; Esquenazi, E.; Leao, T.; Moss, N.; Mevers, E.; Pereira, A. R.; Monroe, E. A.; Korobeynikov, A.; Murray, T. F.; Sherman, D.; Gerwick, L.; Dorrestein, P. C.; Gerwick, W. H. *J. Nat. Prod.* **2017**, *80* (5), 1514–1521.

(231) 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.; Gerwick, W. H. *J. Nat. Prod.* **2015**, *78* (7), 1671–1682.

(232) Moss, N. A.; Bertin, M. J.; Kleigrewe, K.; Leão, T. F.; Gerwick, L.; Gerwick, W. H. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2–3), 313–324.

(233) Holl, C. M.; Montoya, J. P. *J. Phycol.* **2005**, *41* (6), 1178–1183.

(234) Crnkovic, C. M.; May, D. S.; Orjala, J. *J. Appl. Phycol.* **2017**, 1–10.

(235) Ishida, K.; Christiansen, G.; Yoshida, W. Y.; Kurmayer, R.; Valls, N.; Bonjoch, J.; Hertweck, C.; Börner, T. **2014**, *14* (5), 565–576.

(236) Häggqvist, K.; Toruńska-Sitarz, A.; Błaszczyk, A.; Mazur-Marzec, H.; Meriluoto, J. *Toxins (Basel).* **2016**, *8* (4), 1–17.

(237) Grach-Pogrebinsky, O.; Carmeli, S. *Tetrahedron* **2008**, *64* (44), 10233–10238.

(238) Fujii, K.; Harada, K. I. *Anal. Chem.* **1997**, *69* (24), 5146–5151.

(239) Simmons, T. L.; McPhail, K. L.; Ortega-Barría, E.; Mooberry, S. L.; Gerwick, W. H. *Tetrahedron Lett.* **2006**, *47* (20), 3387–3390.

(240) Banuelos, C. A.; Tavakoli, I.; Tien, A. H.; Caley, D. P.; Mawji, N. R.; Li, Z.; Wang, J.; Yang, Y. C.; Imamura, Y.; Yan, L.; Wen, J. G.; Andersen, R. J.; Sadar, M. D. *J. Biol. Chem.* **2016**, *291* (42), 22231–22243.

(241) Sadar, M. D.; Williams, D. E.; Mawji, N. R.; Patrick, B. O.; Wikanta, T.; Chasanah, E.; Irianto, H. E.; Soest, R. Van; Andersen, R. J. *Org. Lett.* **2008**, *10* (21), 4947–4950.

(242) Leikoski, N.; Liu, L.; Jokela, J.; Wahlsten, M.; Gugger, M.; Calteau, A.; Permi, P.; Kerfeld, C. a.; Sivonen, K.; Fewer, D. P. *Chem. Biol.* **2013**, *20* (8), 1033–1043.

(243) Esquenazi, E.; Jones, A. C.; Byrum, T.; Dorrestein, P. C.; Gerwick, W. H. *Proc. Natl. Acad. Sci.* **2011**, *108* (13), 5226–5231.

(244) 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.; White, M. A. *Sci. Signal.* **2013**, *6* (297), 1–14.

(245) Nitiss, J. L. *Cancer Chemother. Pharmacol.* **1994**, *34*, 6–13.

APPENDICES

APPENDIX A – Spectroscopic Data
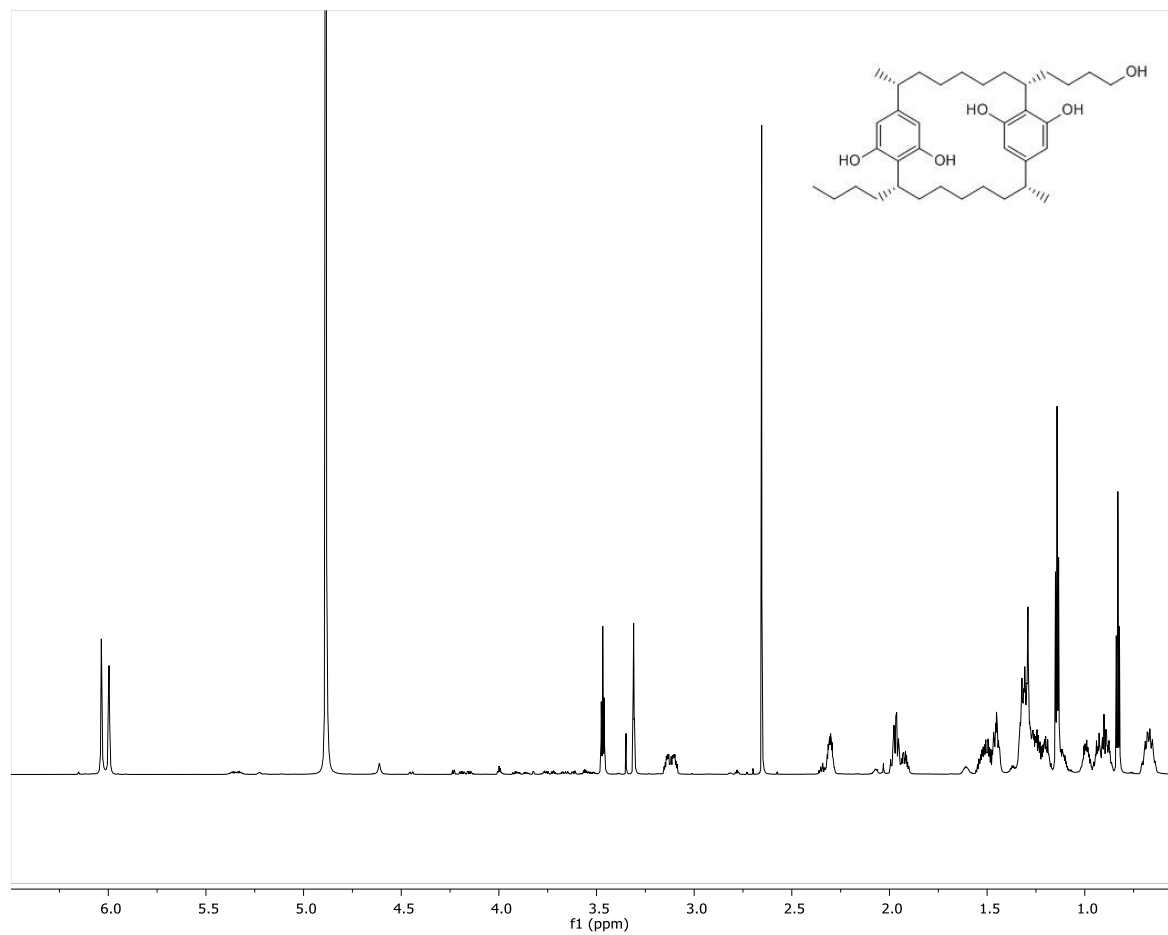
Figure 64: ^1H NMR spectrum (900 MHz MeOd-d$_4$) of merocyclophane C, see Table II
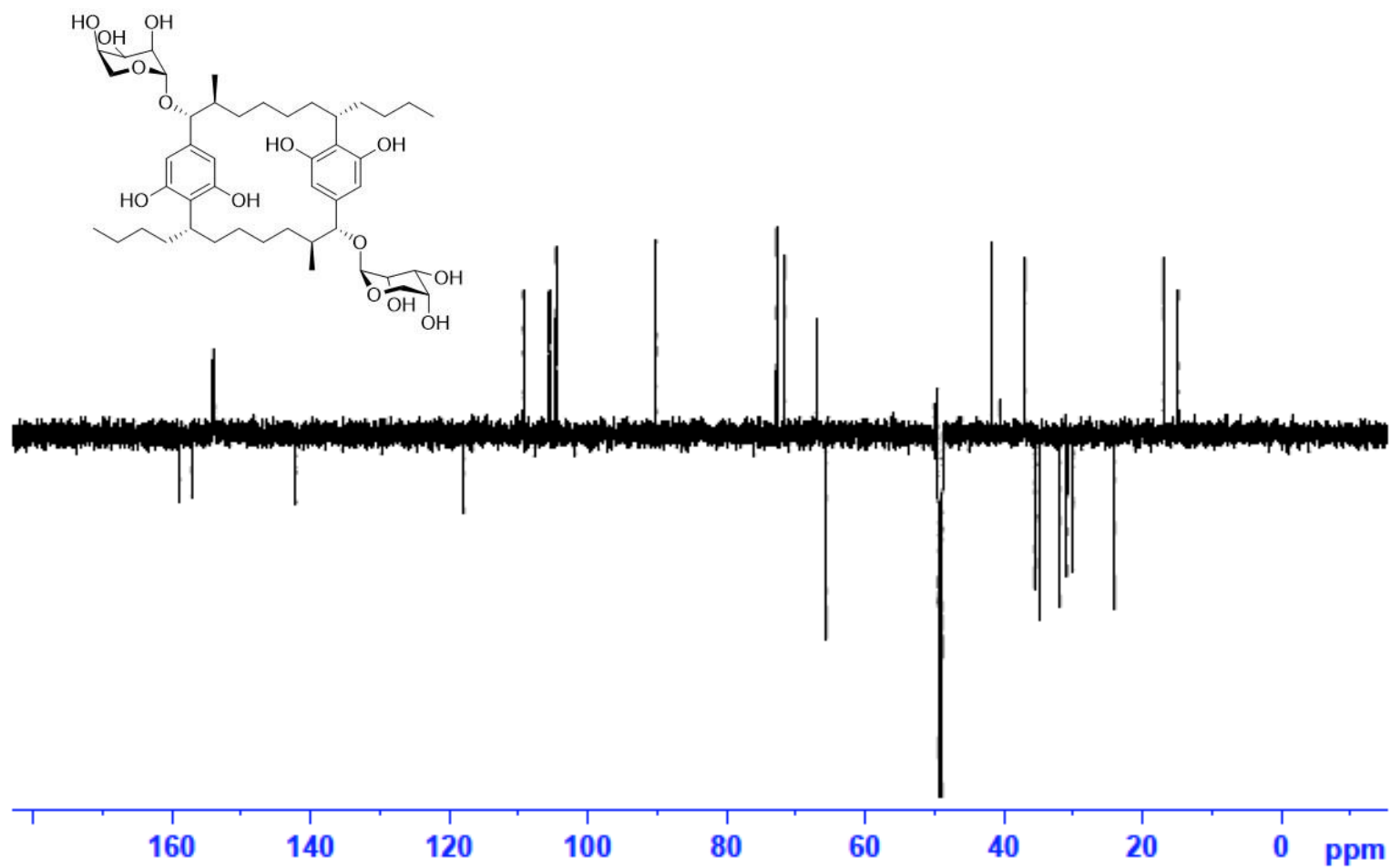
APPENDIX A (continued)
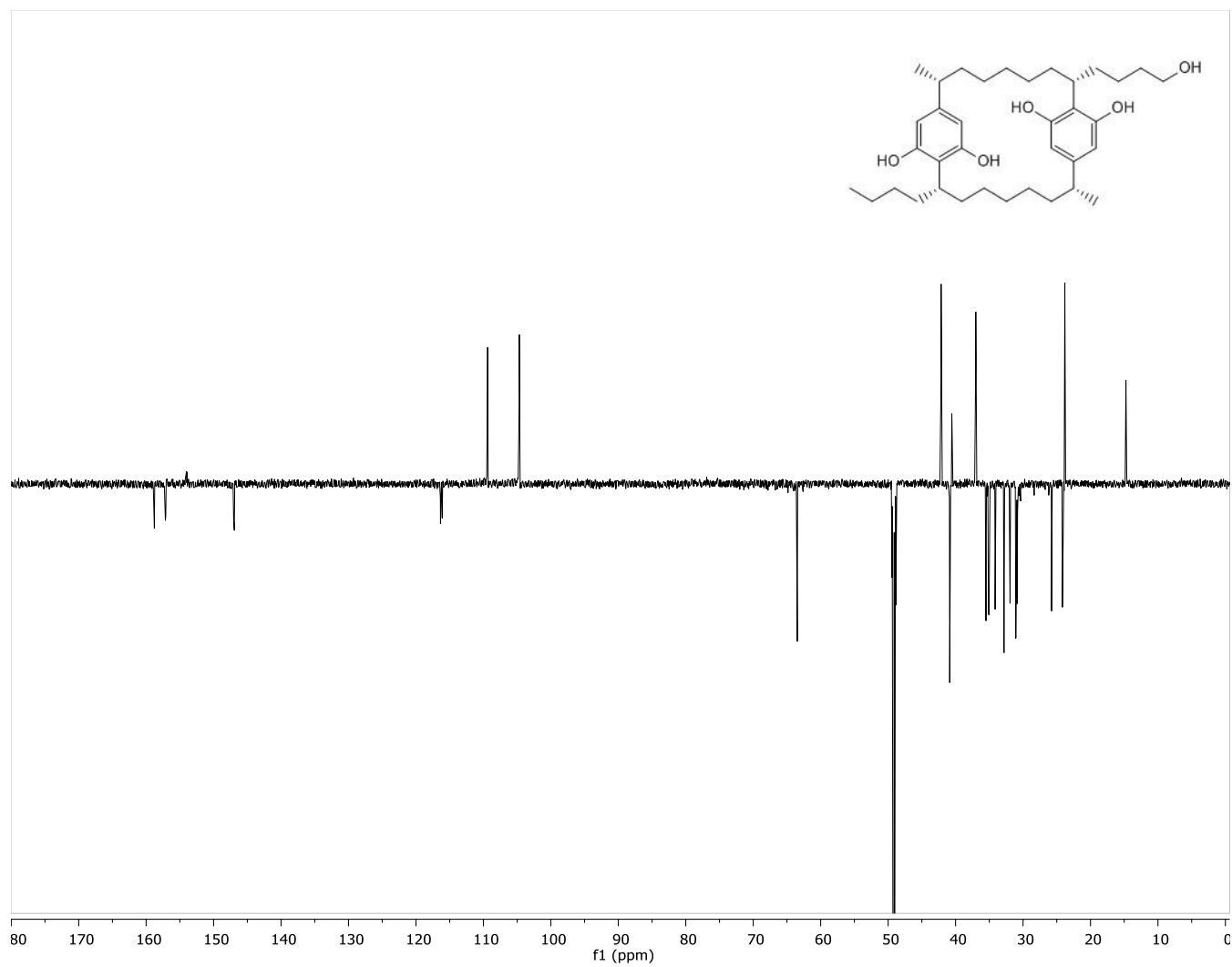
Figure 65: DEPTQ spectrum (225 MHz MeOd-d₄) of merocyclophane C, see Table II
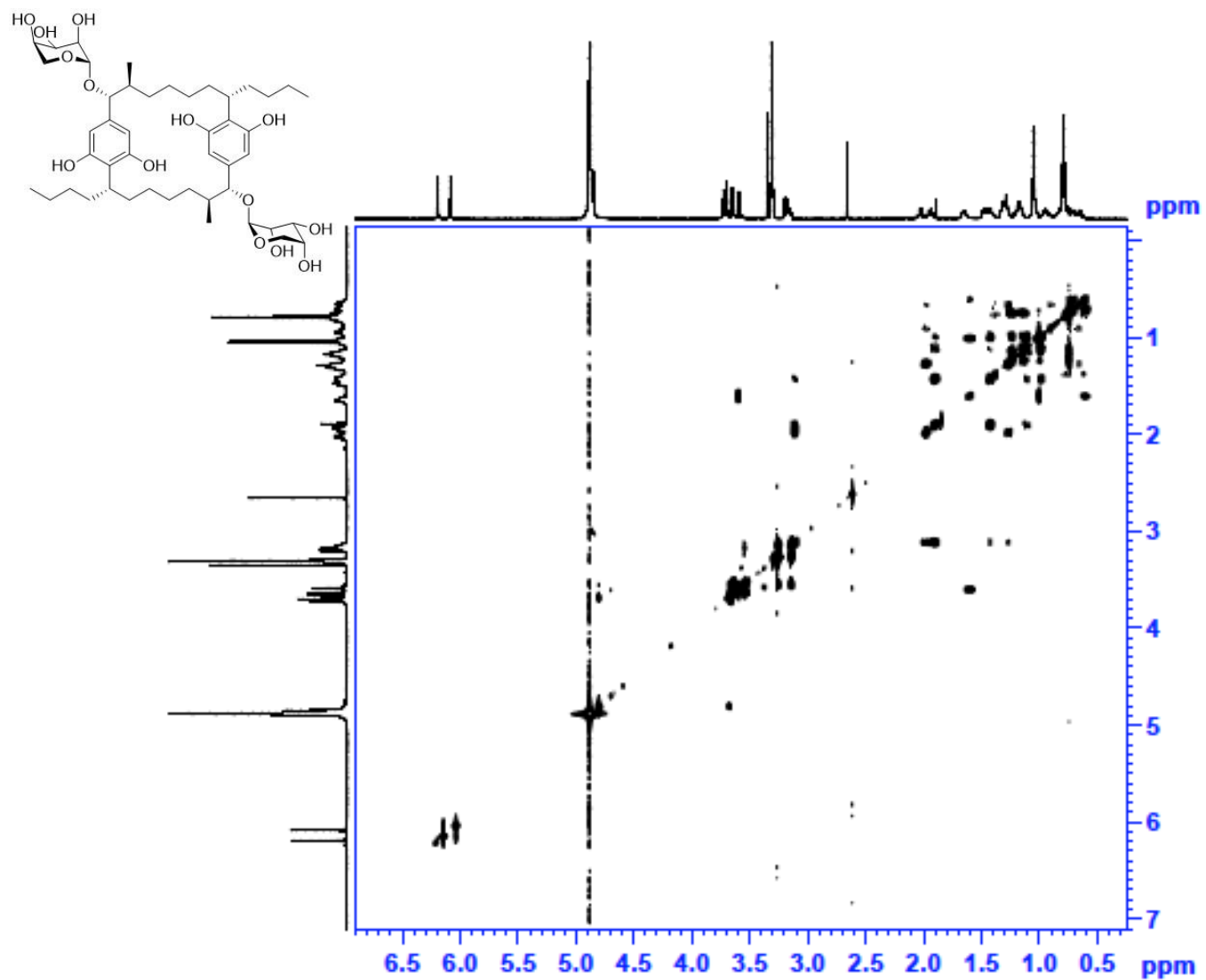
APPENDIX A (continued)
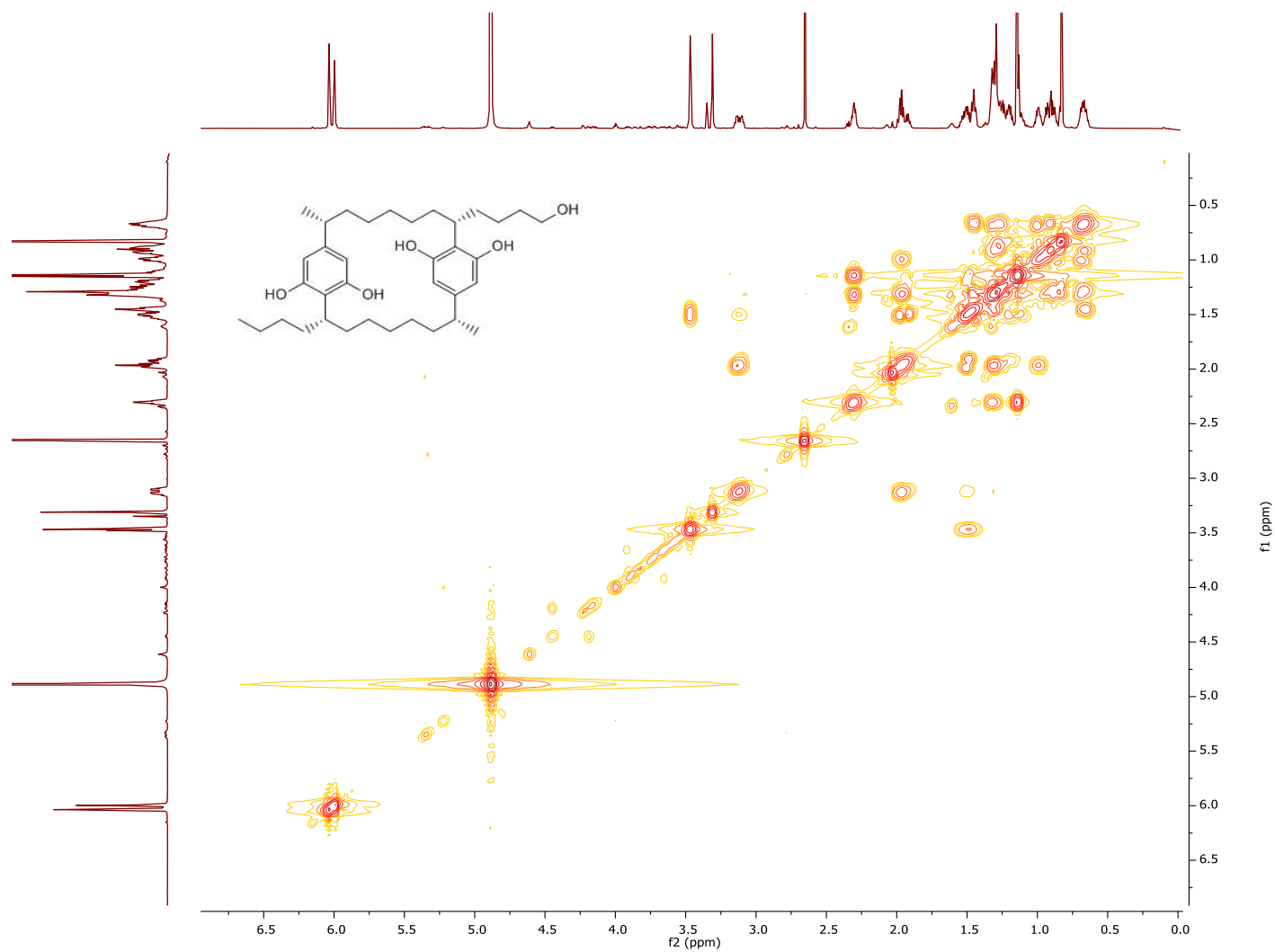
Figure 66: COSY spectrum (600 MHz MeOd-d$_4$) of merocyclophane C, see Table II
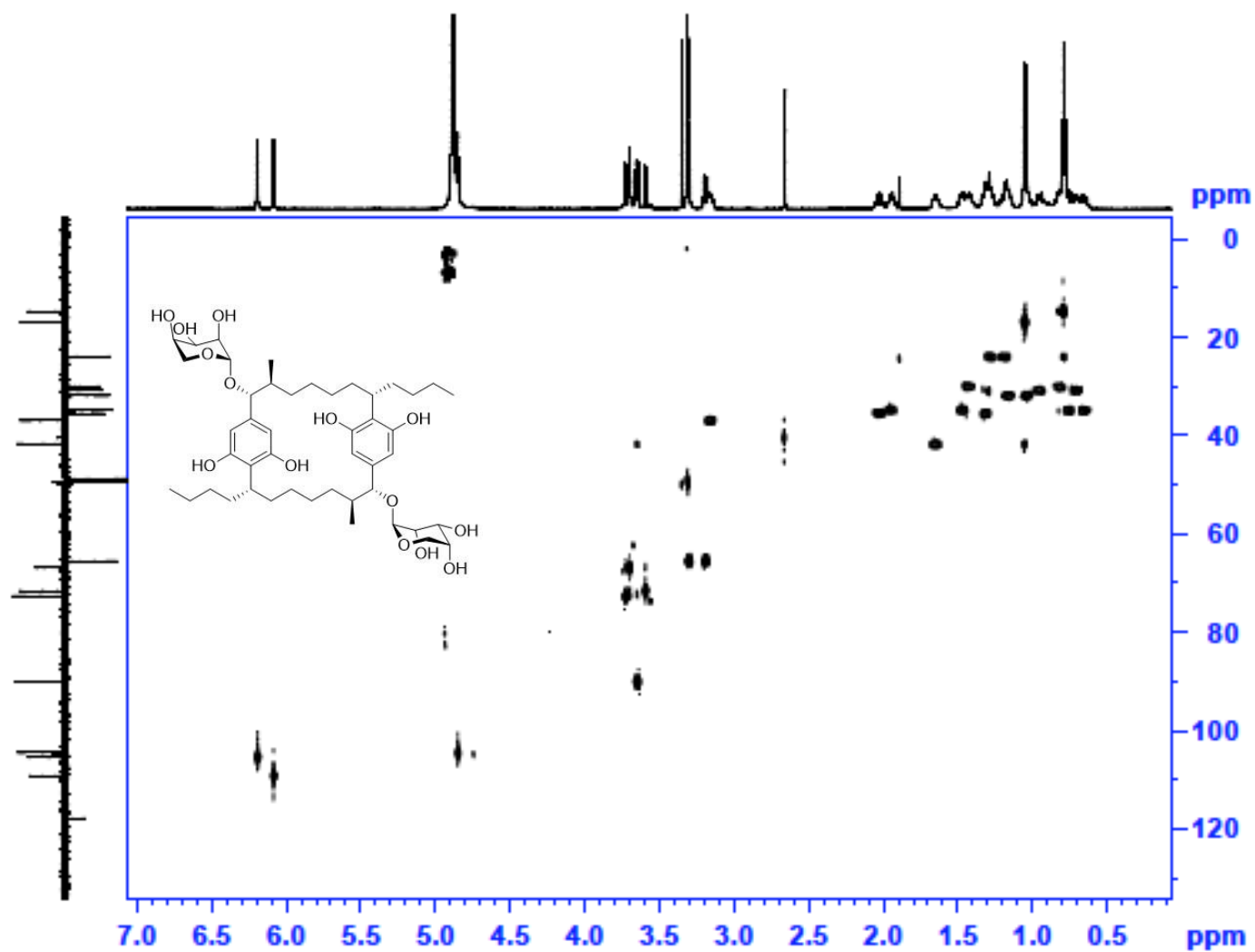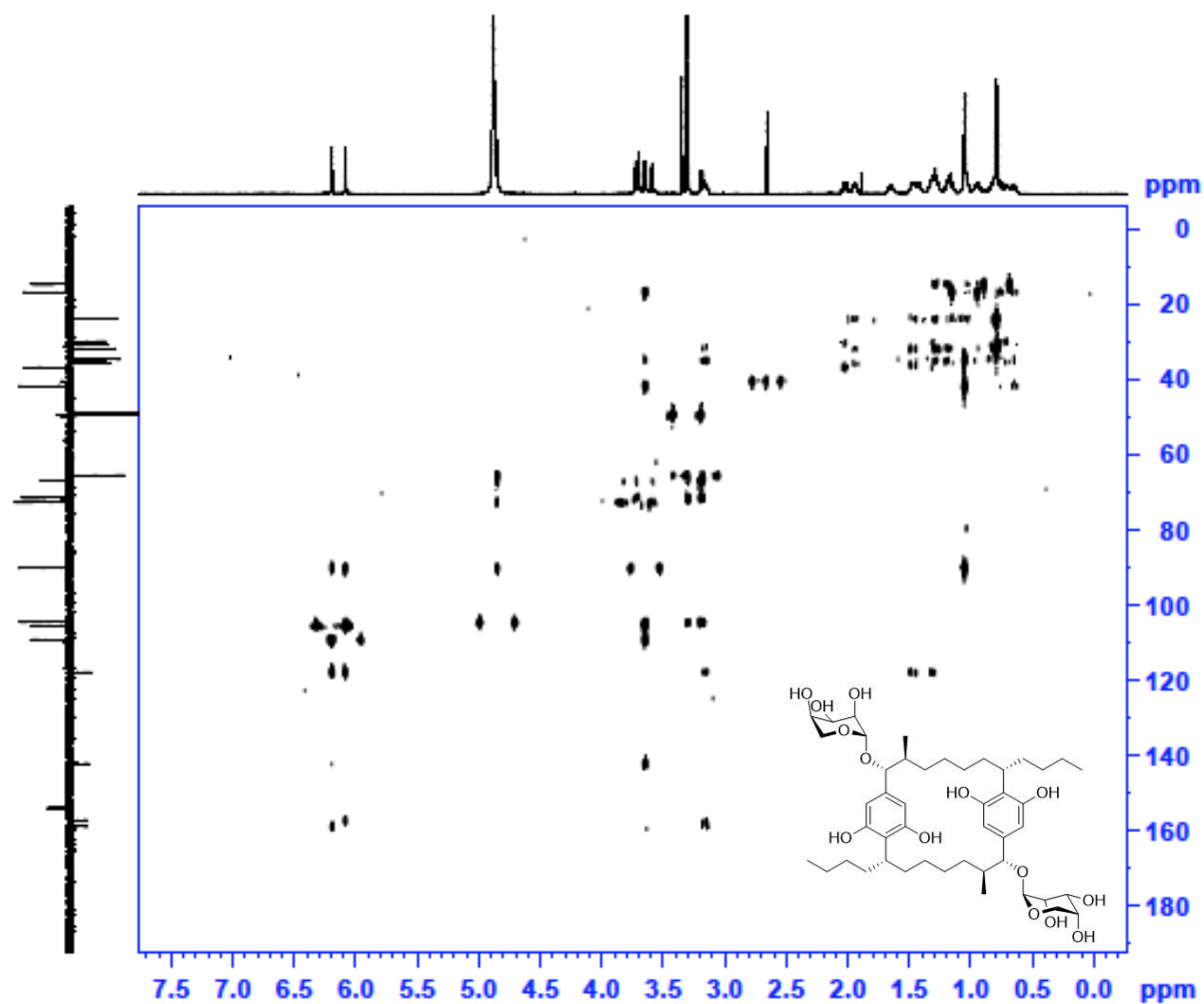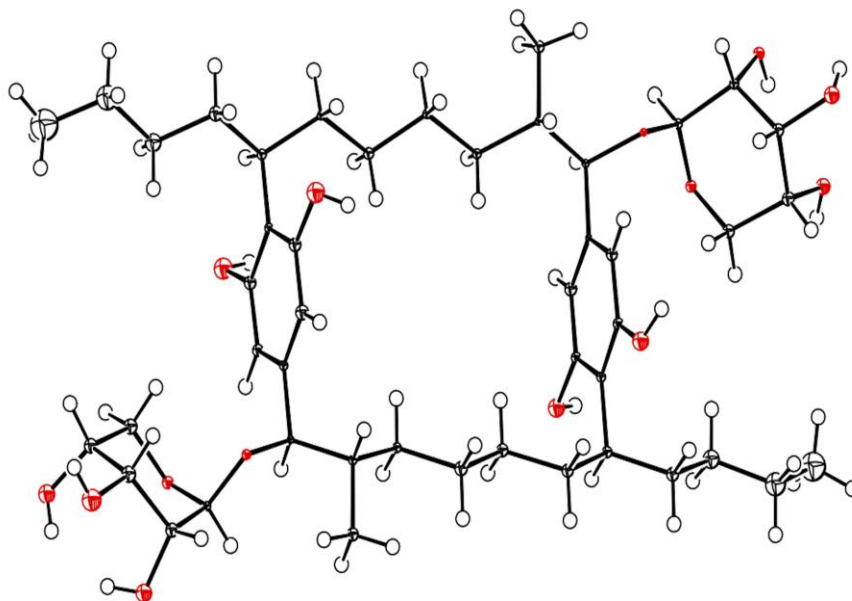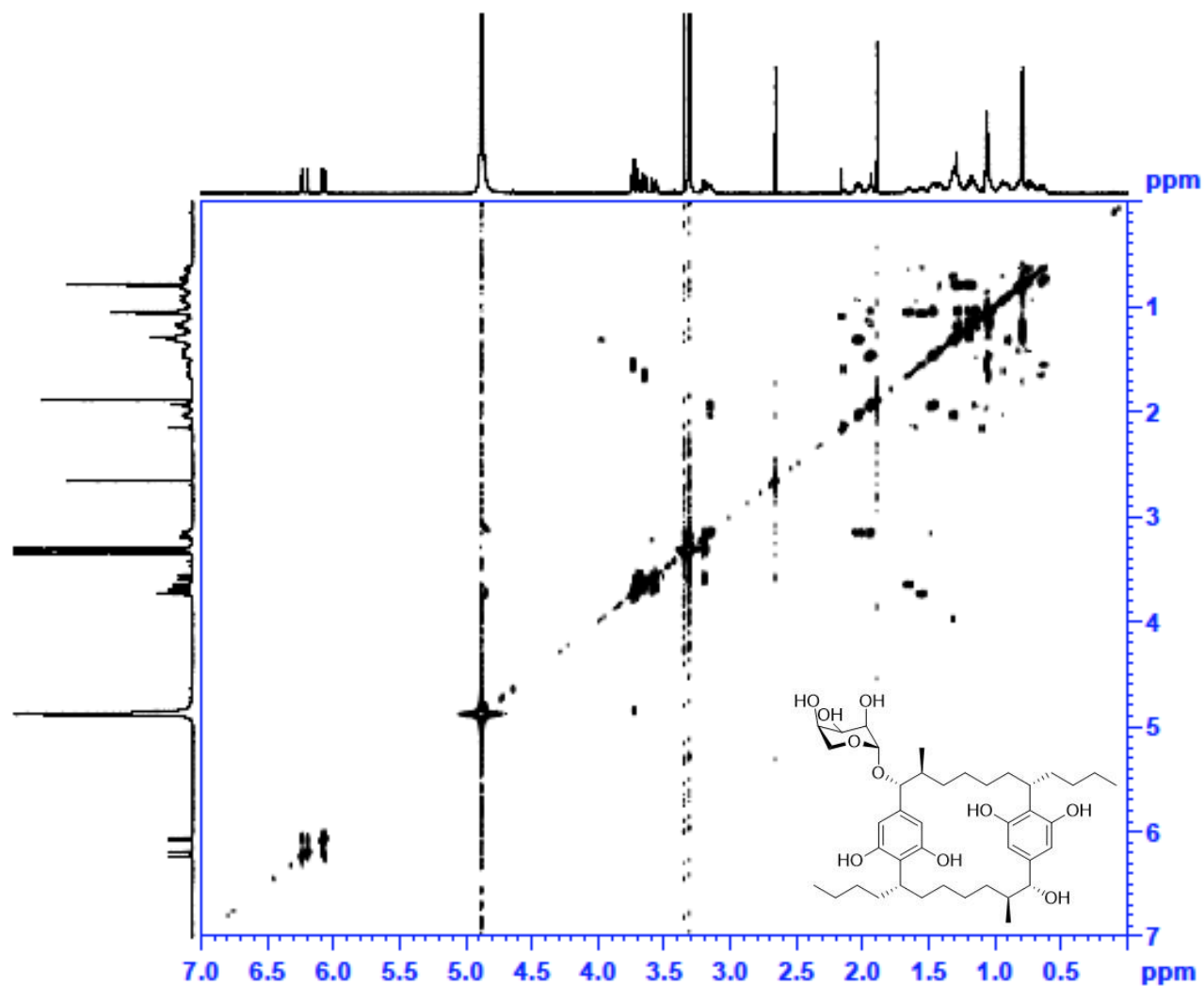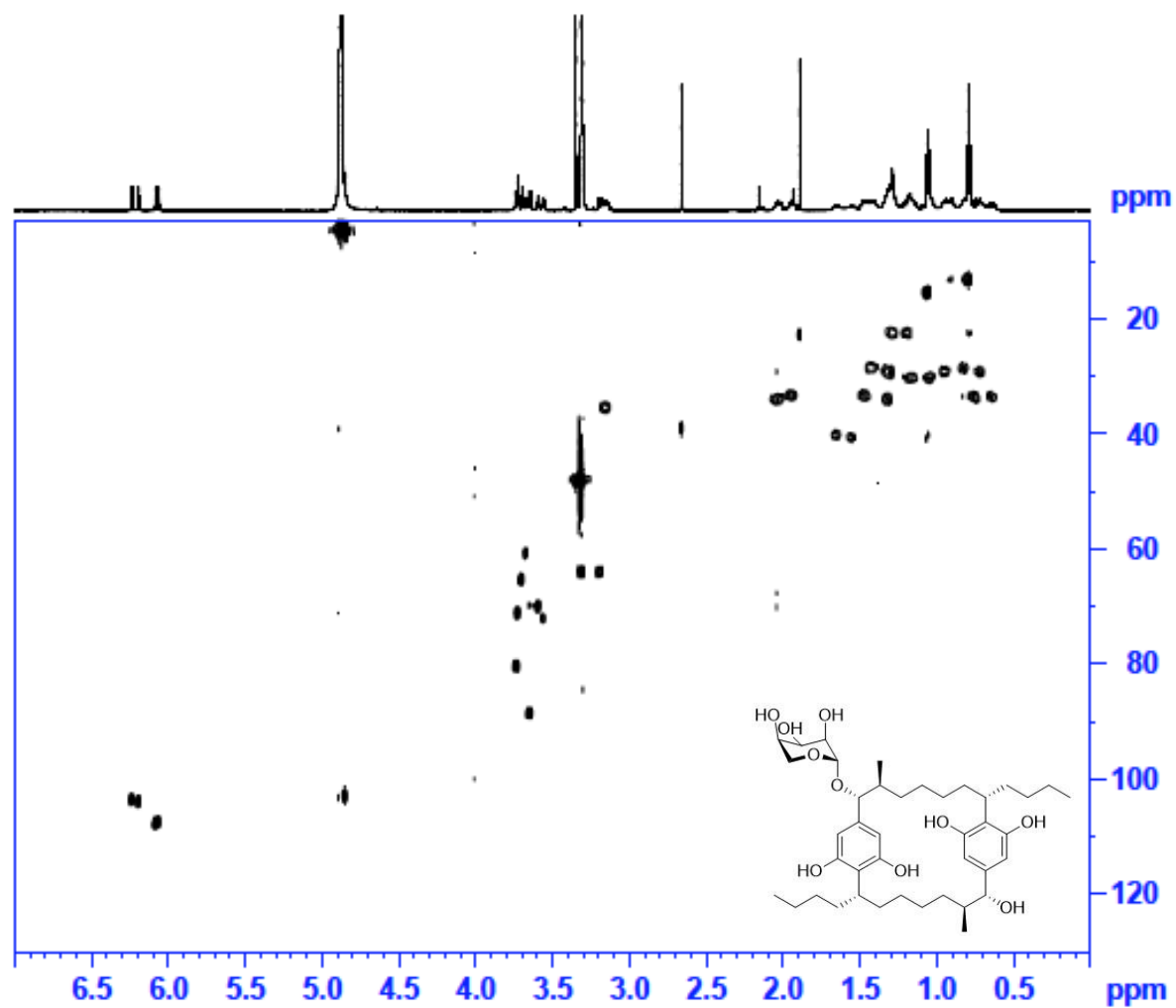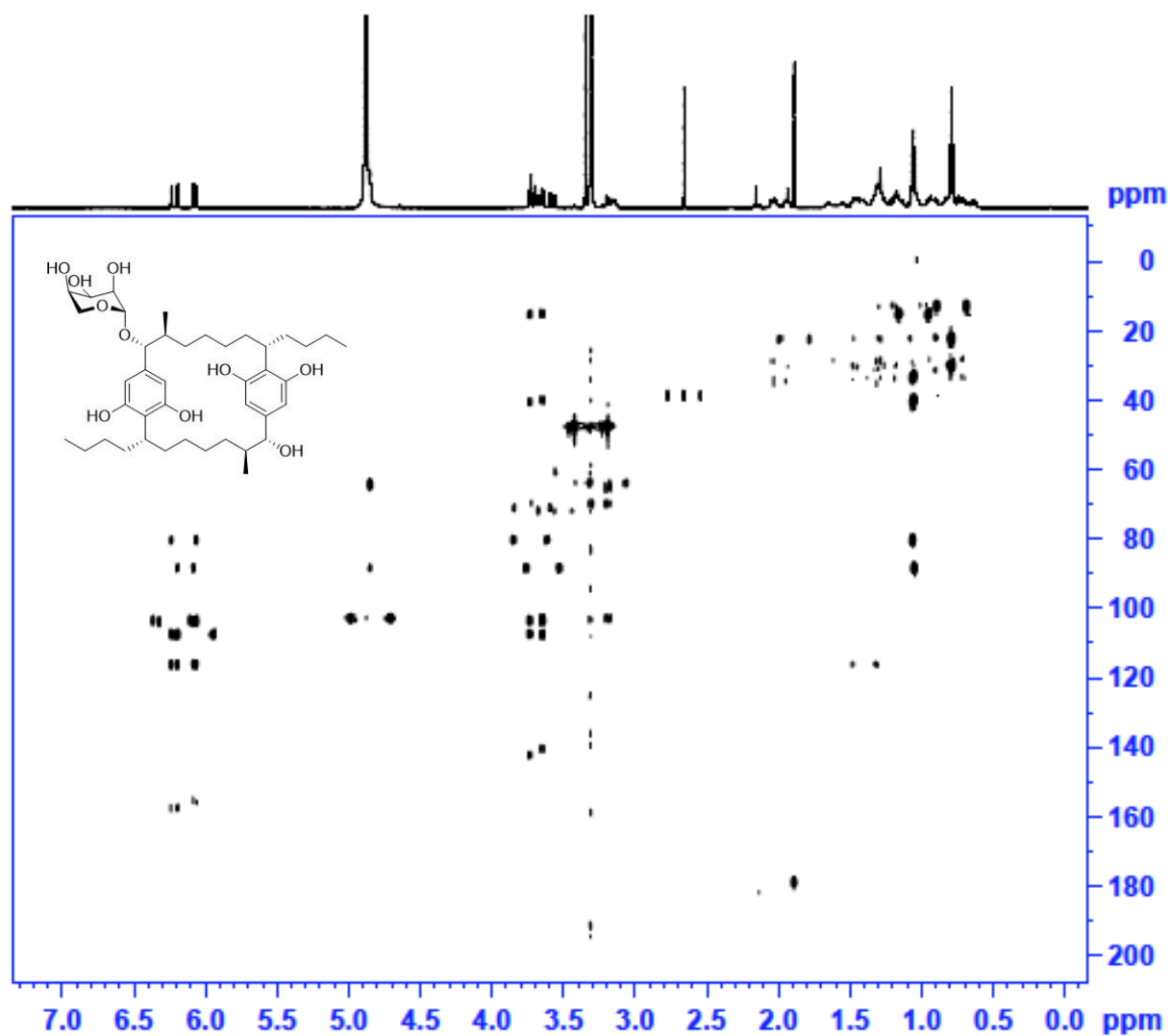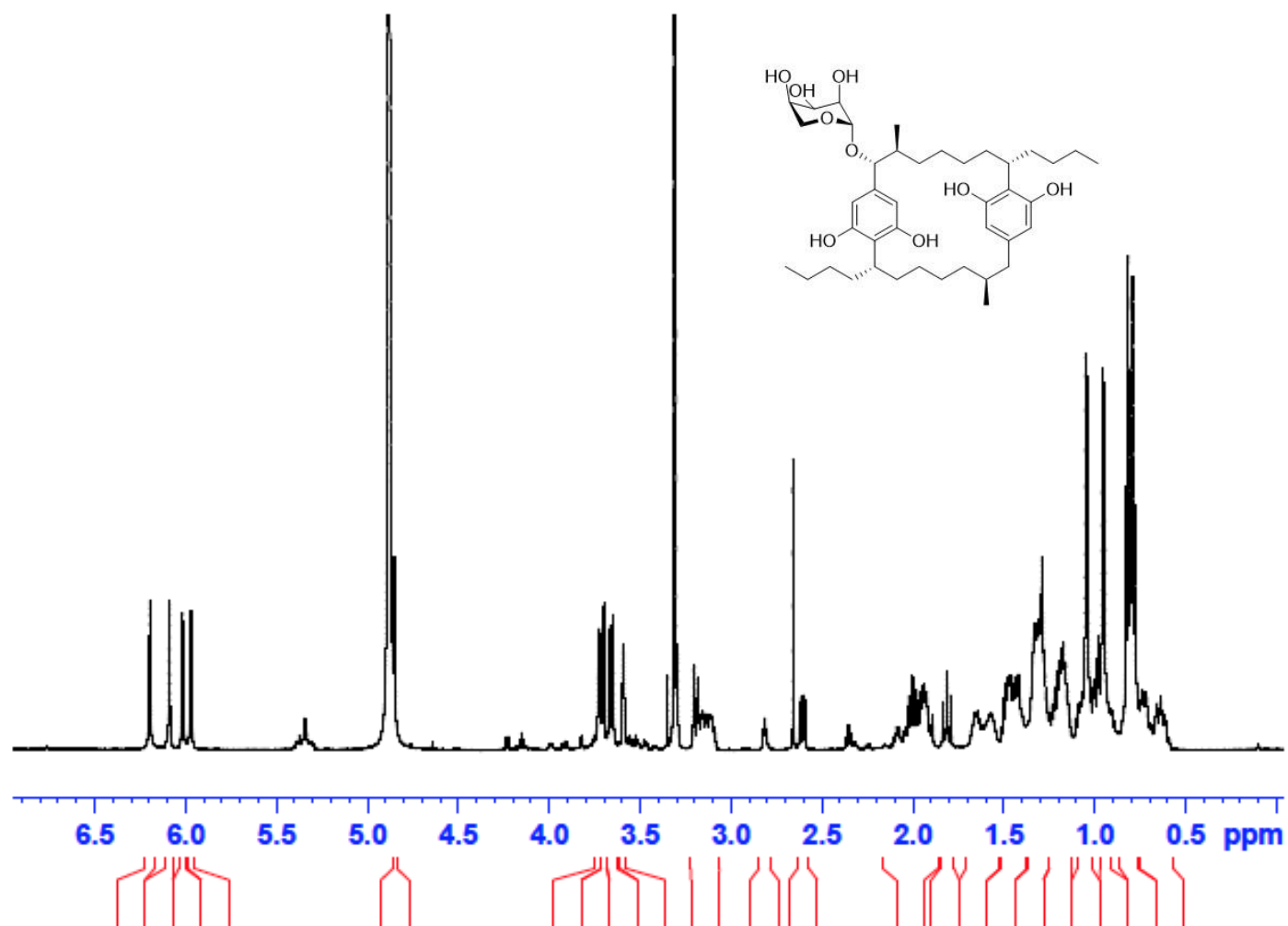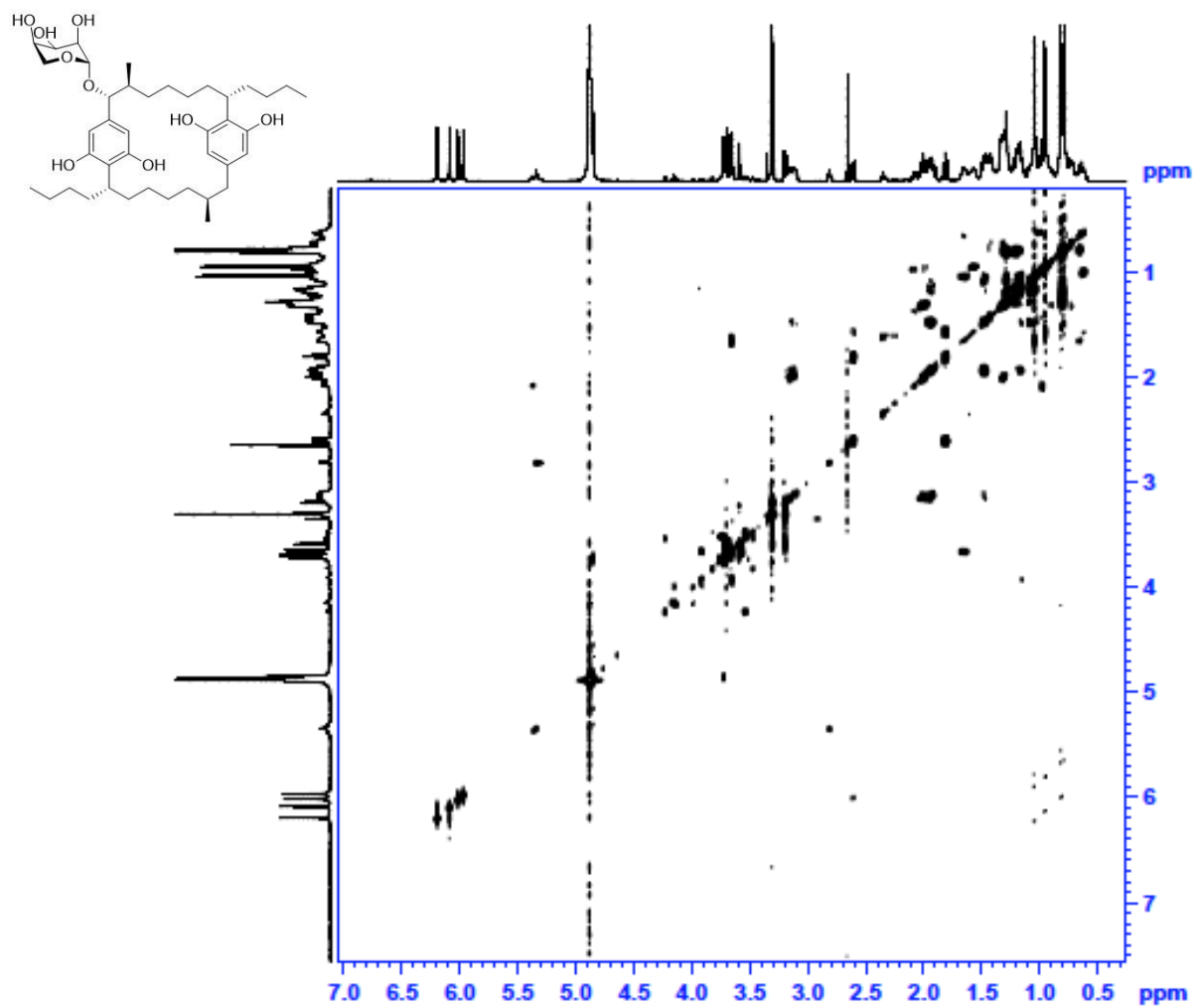
APPENDIX A (continued)
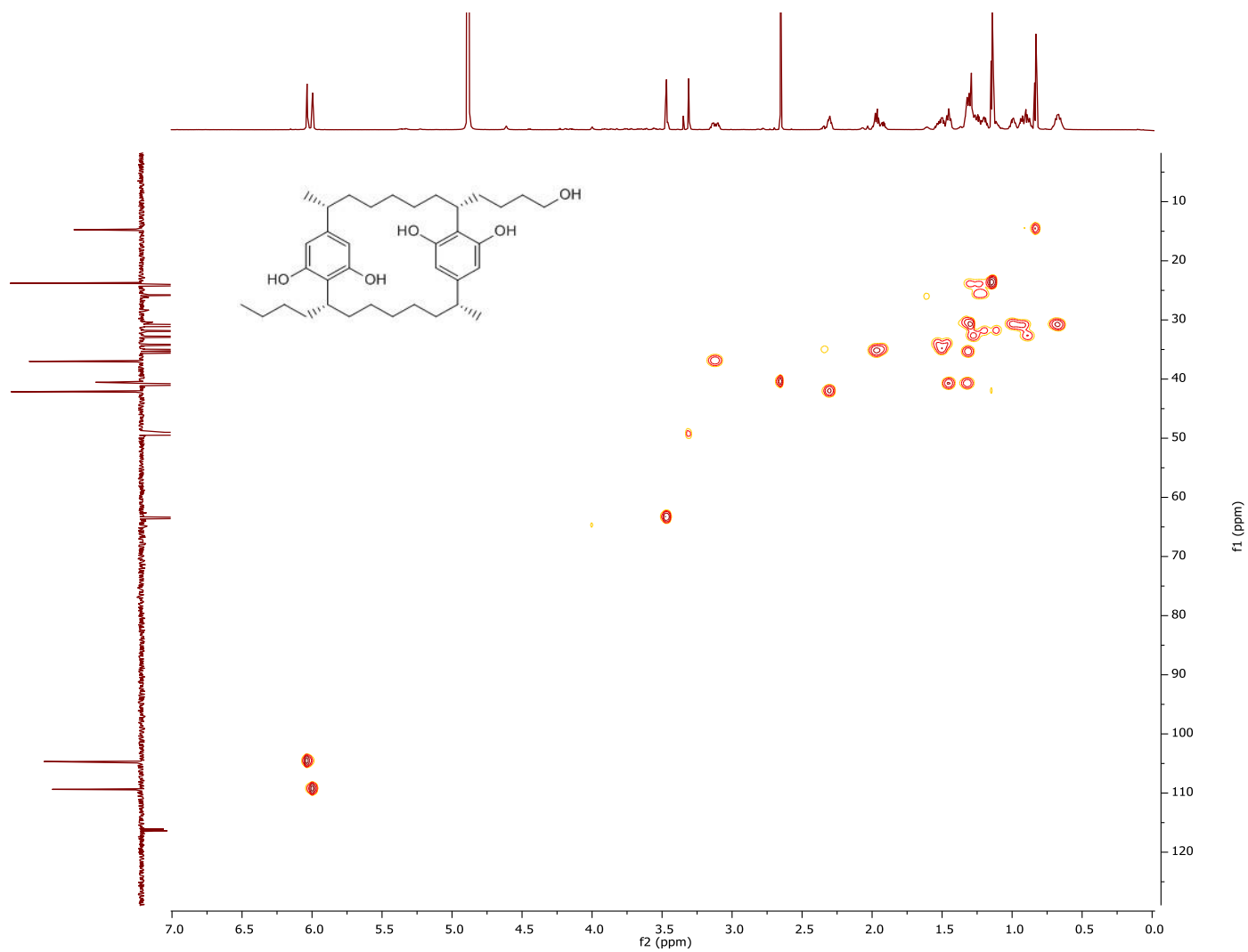
Figure 67: HSQC spectrum (600 MHz MeOd-d$_4$) of merocyclophane C, see Table II

APPENDIX A (continued)
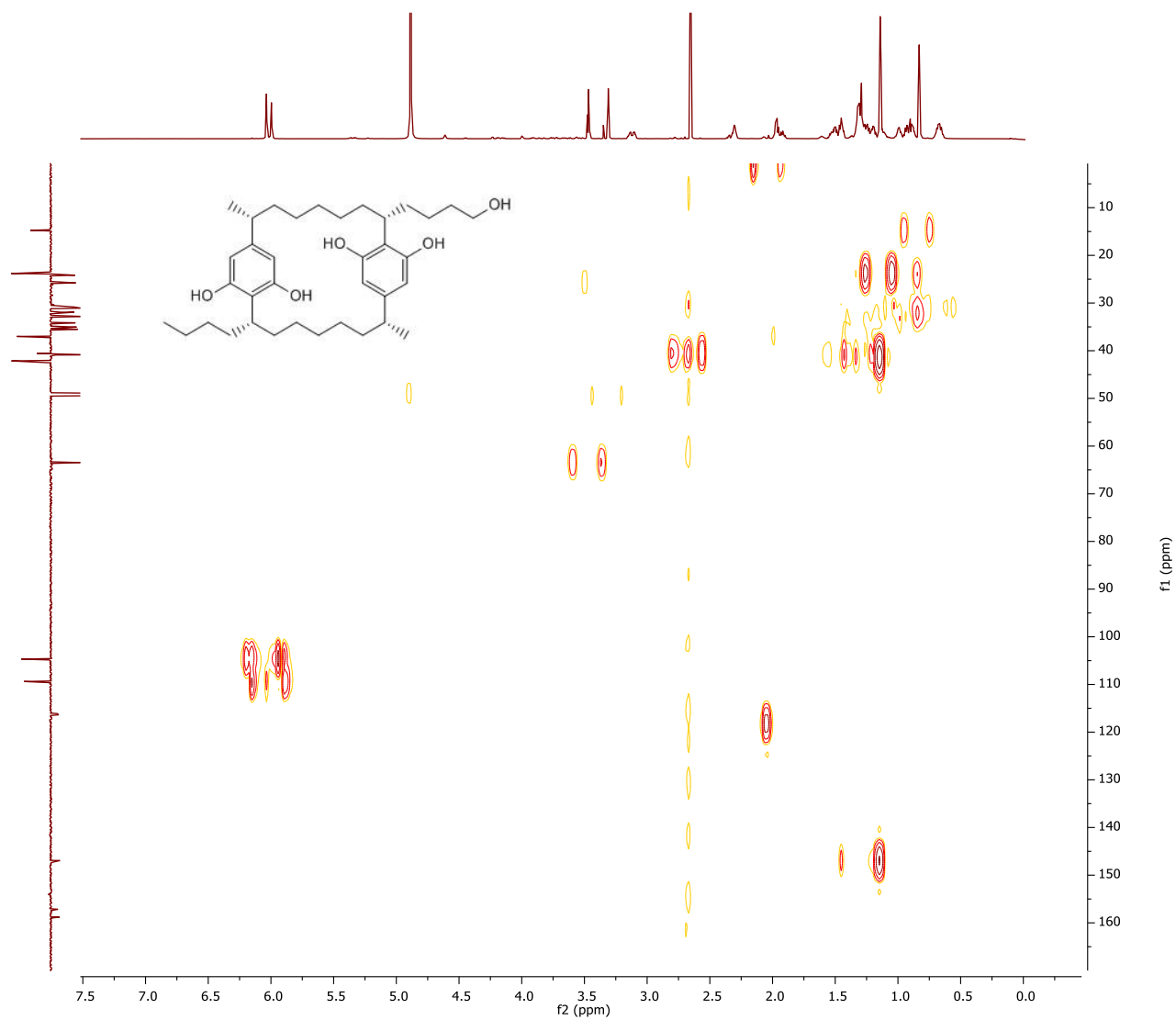
Figure 68: HMBC spectrum (600 MHz MeOd-d₄) of merocyclophane C, see Table II
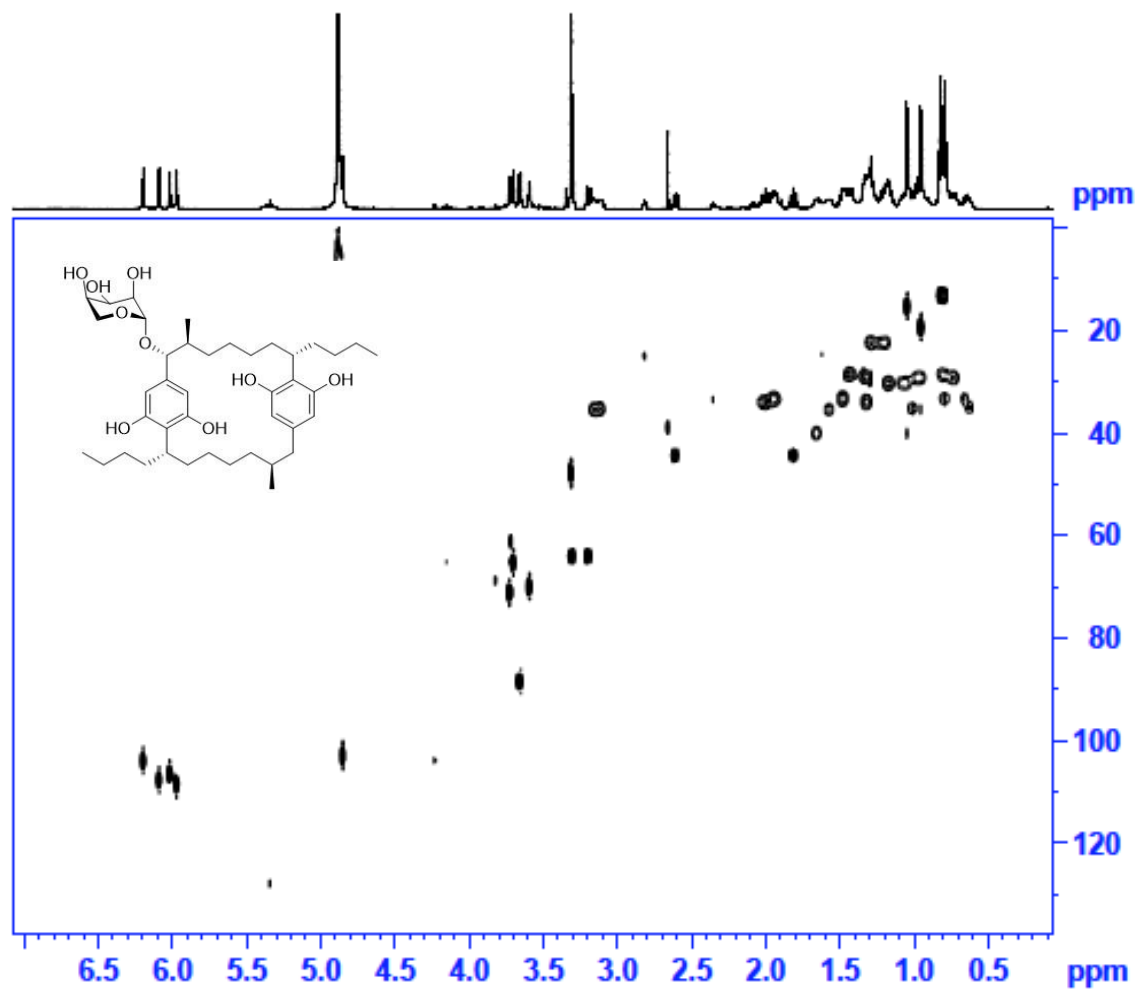
APPENDIX A (continued)
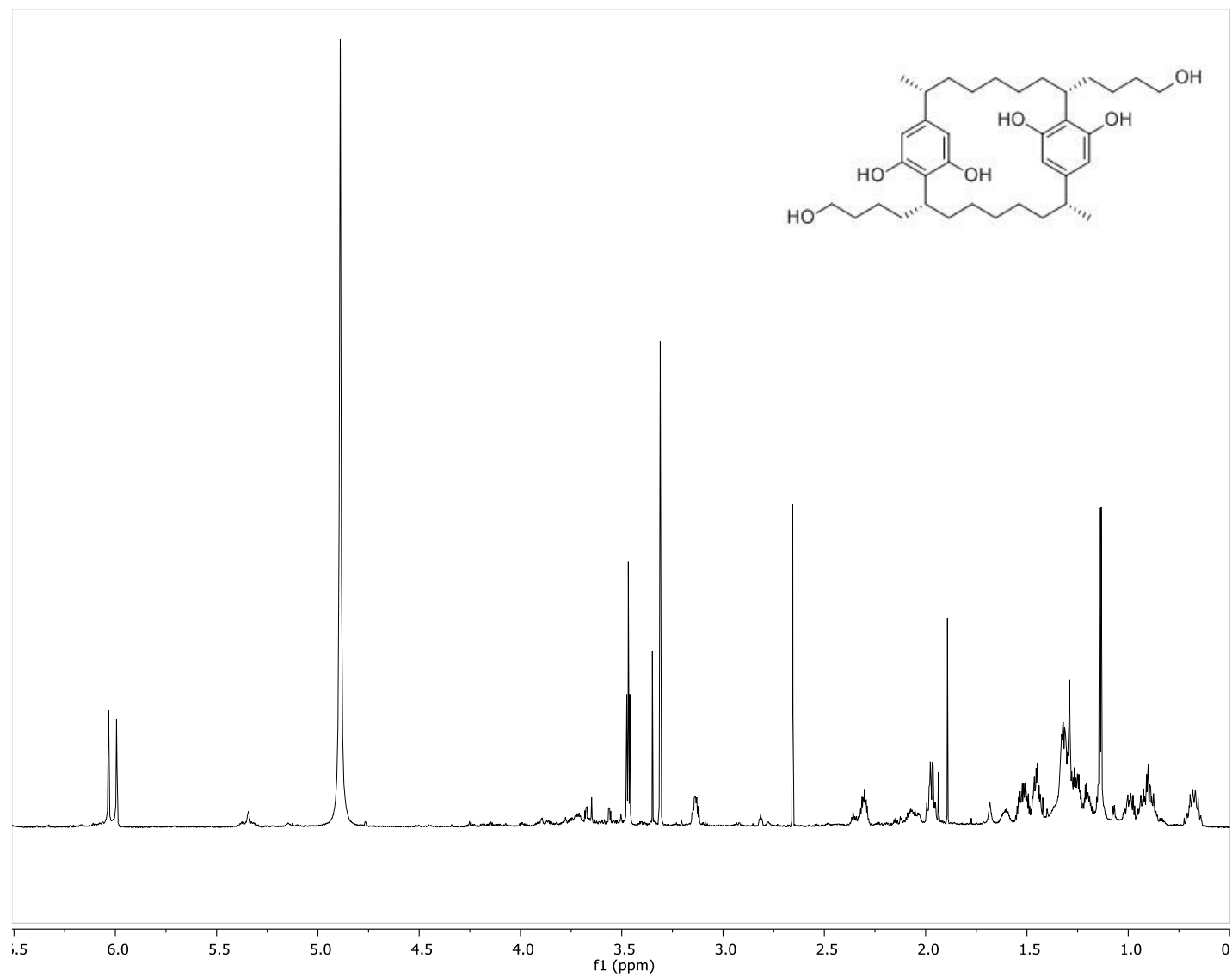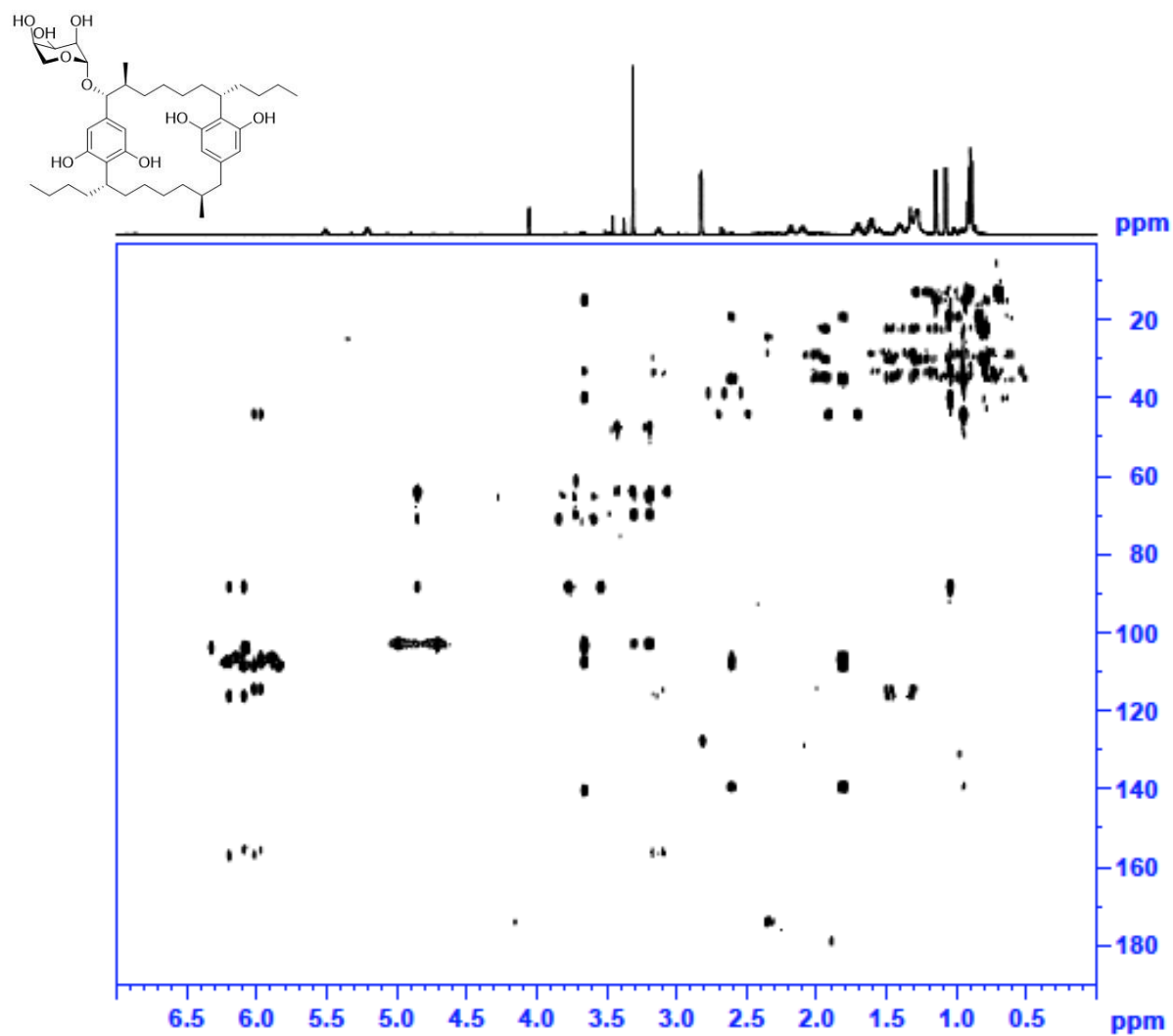

Figure 69: ^1H NMR spectrum (900 MHz MeOd-d$_4$) of merocyclophane D, see Table II

APPENDIX A (continued)
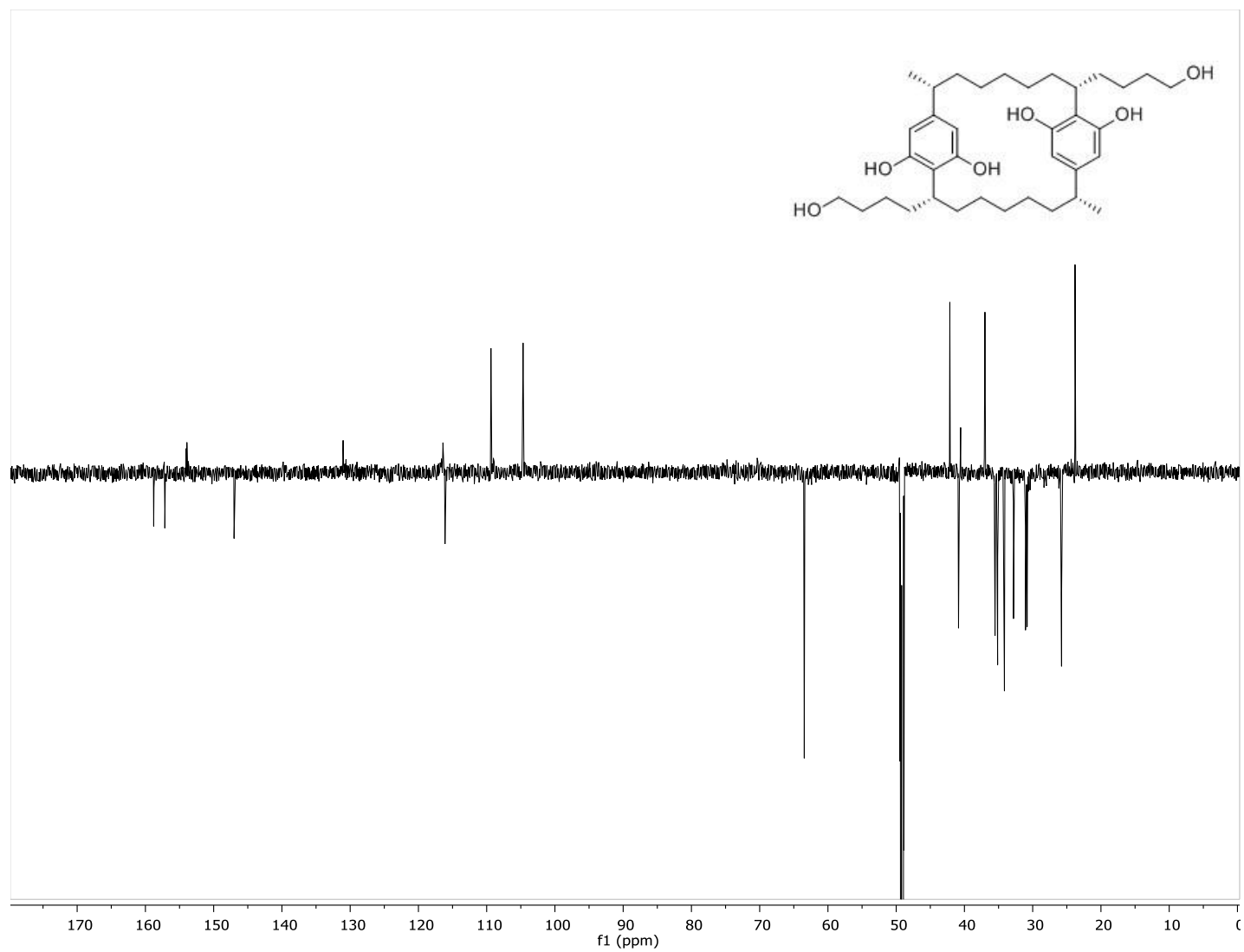
Figure 70: DEPTQ spectrum (225 MHz MeOd-d₄) of merocyclophane D, see Table II

APPENDIX A (continued)
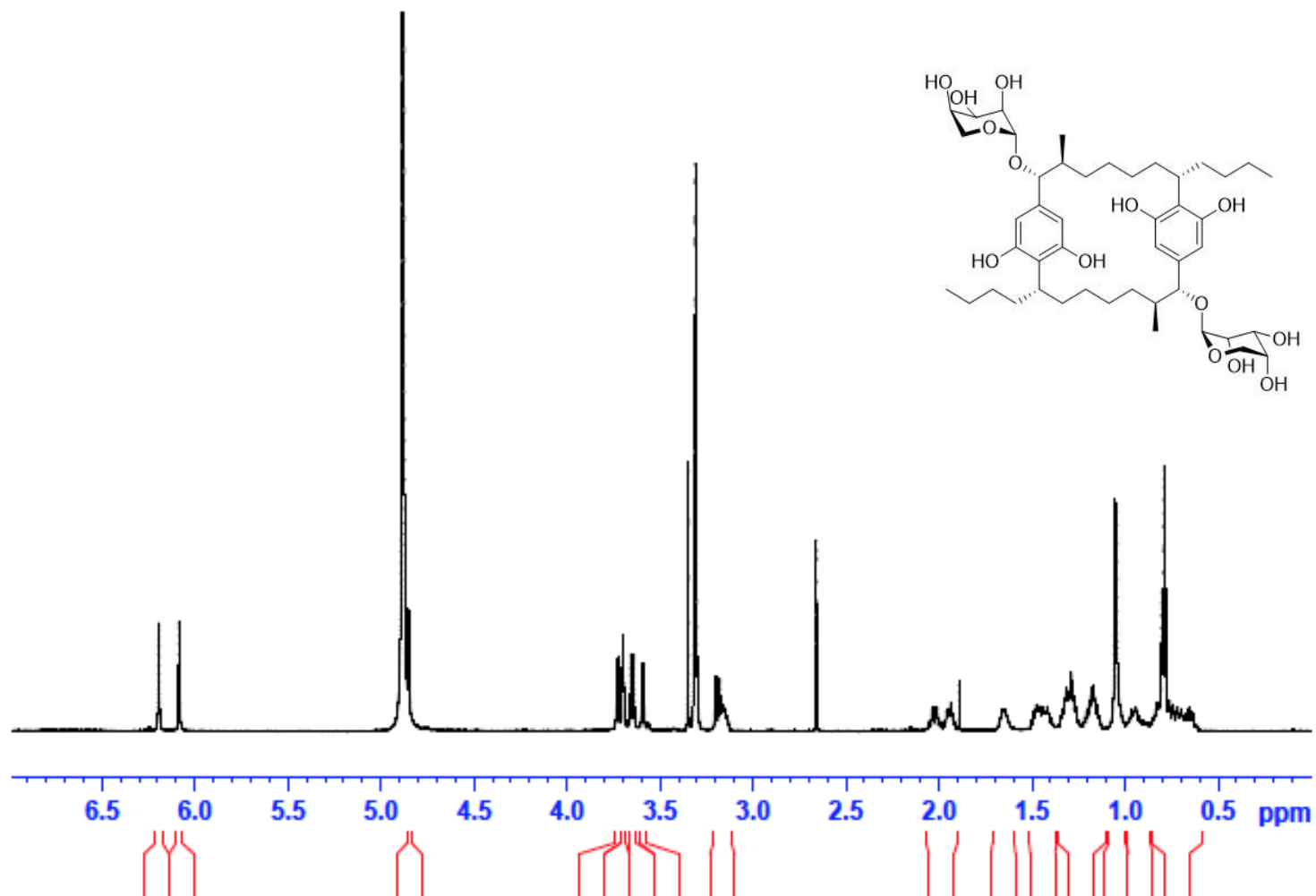
Figure 71: ¹H NMR spectrum (600 MHz MeOd-d₄) of ribocyclophane A, see Table V

APPENDIX A (continued)

Figure 72: DEPTQ spectrum (225 MHz MeOd-d₄) of ribocyclophane A, see Table V

APPENDIX A (continued)

Figure 73: COSY spectrum (600 MHz MeOd-d₄) of ribocyclophane A, see Table V
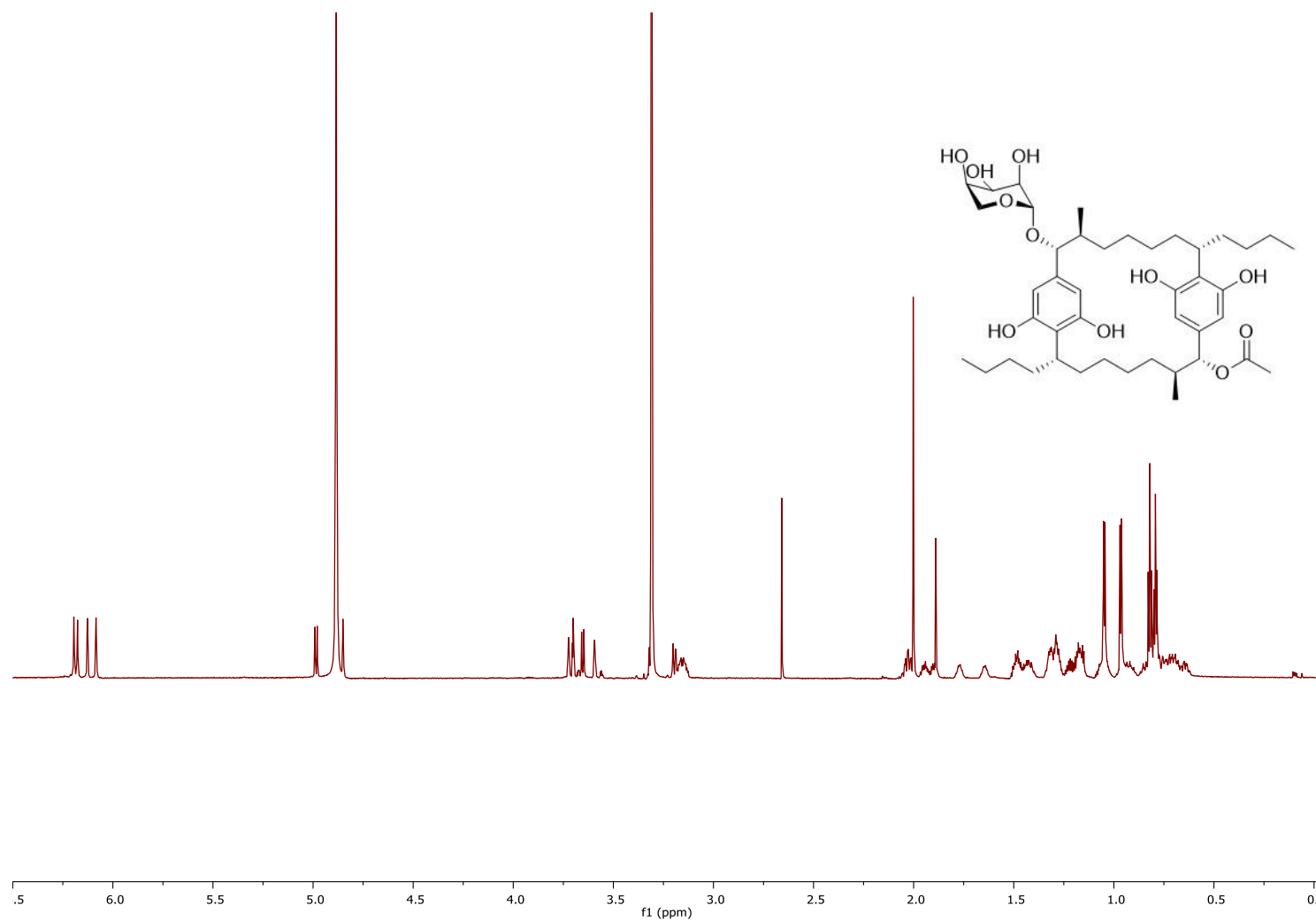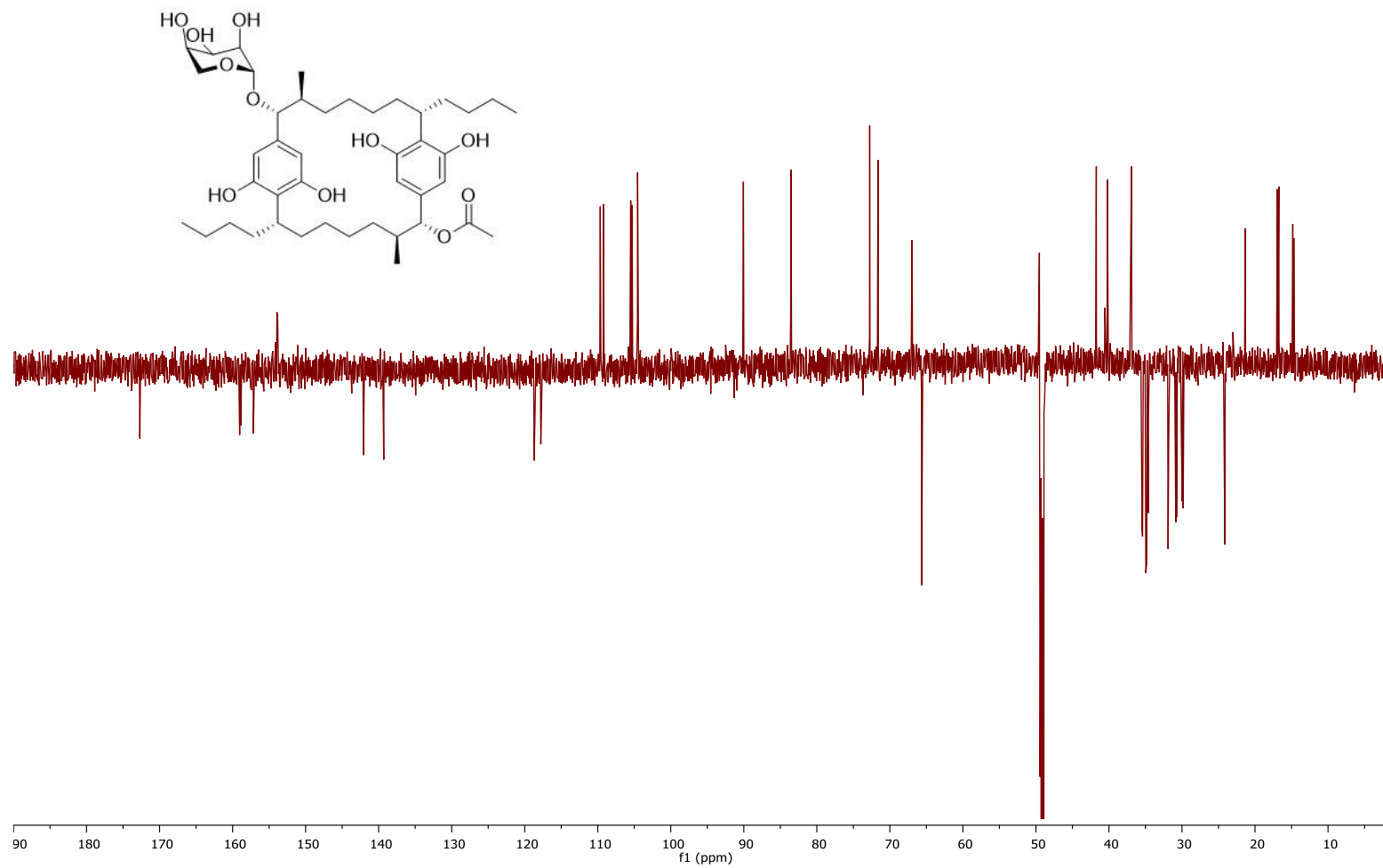
APPENDIX A (continued)

Figure 74: HSQC spectrum (600 MHz MeOd-d₄) of ribocyclophane A, see Table V

APPENDIX A (continued)

Figure 75: HMBC spectrum (600 MHz MeOd-d₄) of ribocyclophane A, see Table V

Figure 76: ORTEP drawing of ribocyclophane A

APPENDIX A (continued)
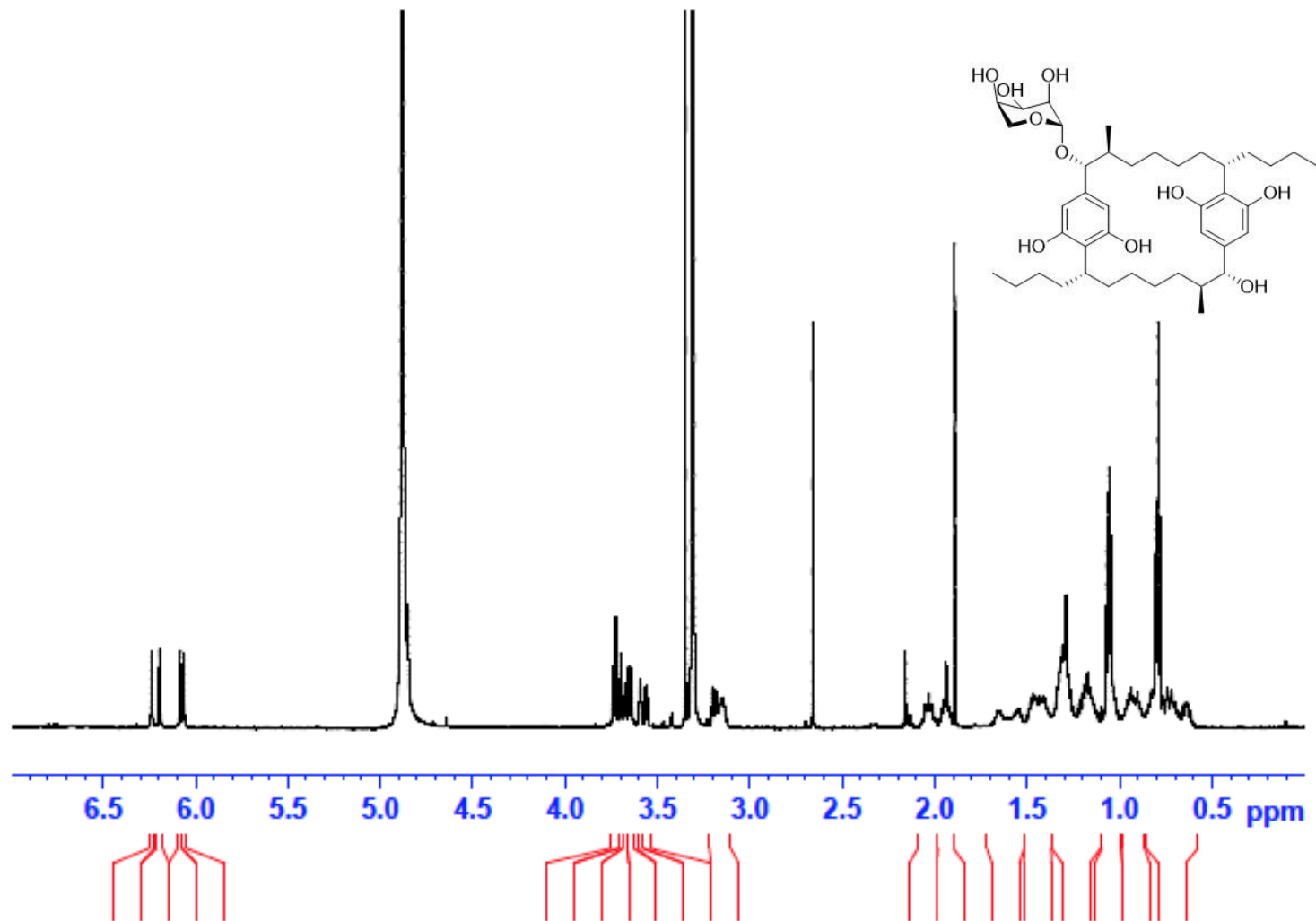
Figure 77: ¹H NMR spectrum (600 MHz MeOd-d₄) of ribocyclophane B, see Table VI

APPENDIX A (continued)

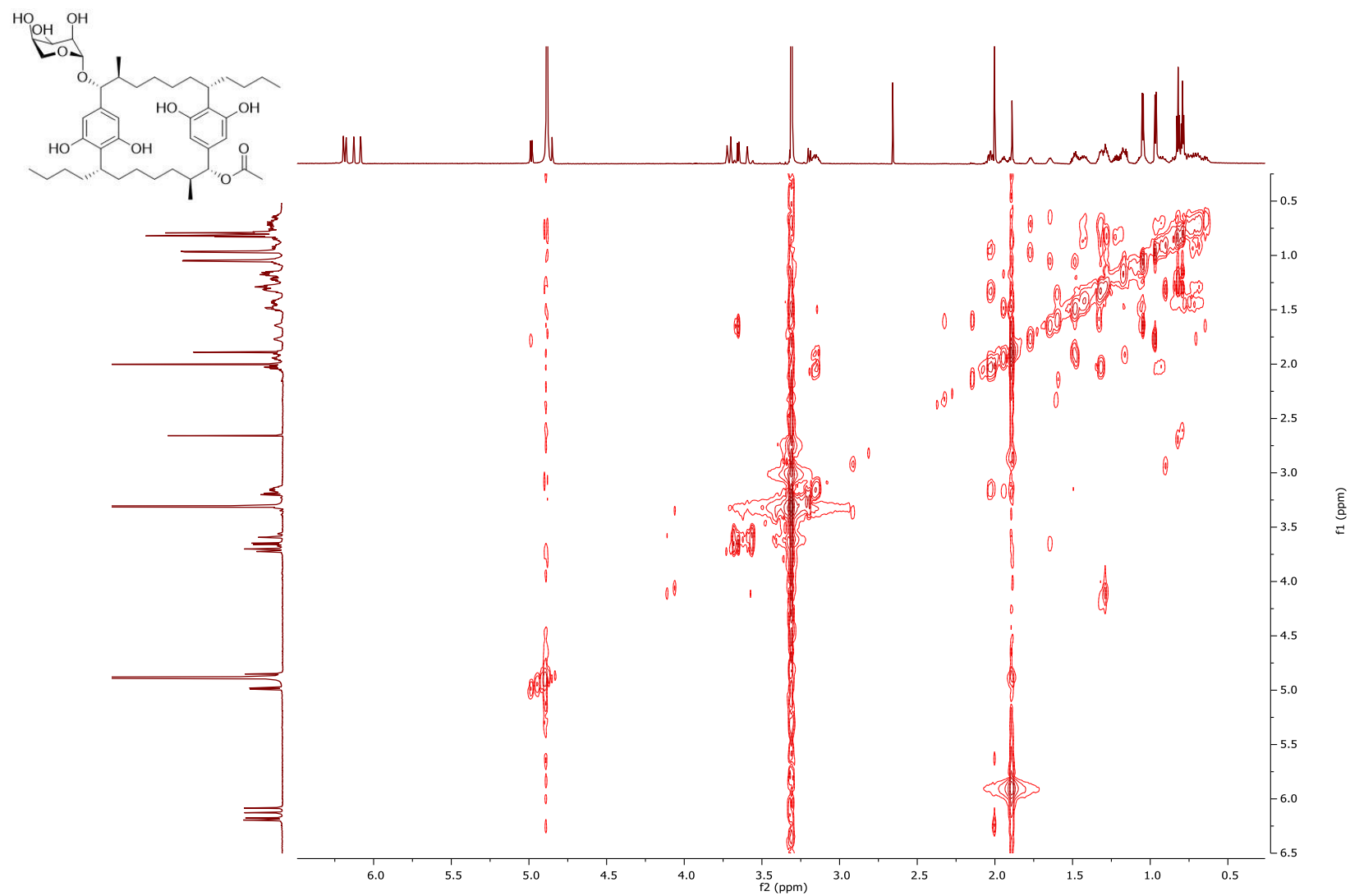*Figure 78: COSY spectrum (600 MHz MeOd-d₄) of ribocyclophane B, see Table VI*

APPENDIX A (continued)

Figure 79: HSQC spectrum (600 MHz MeOd-d₄) of ribocyclophane B, see Table VI

APPENDIX A (continued)

Figure 80: HMBC spectrum (600 MHz MeOd-d₄) of ribocyclophane B, see Table VI

APPENDIX A (continued)

Figure 81: ¹H NMR spectrum (600 MHz MeOd-d₄) of ribocyclophane C, see Table VI

APPENDIX A (continued)

Figure 82: COSY spectrum (600 MHz MeOd-d₄) of ribocyclophane C, see Table VI
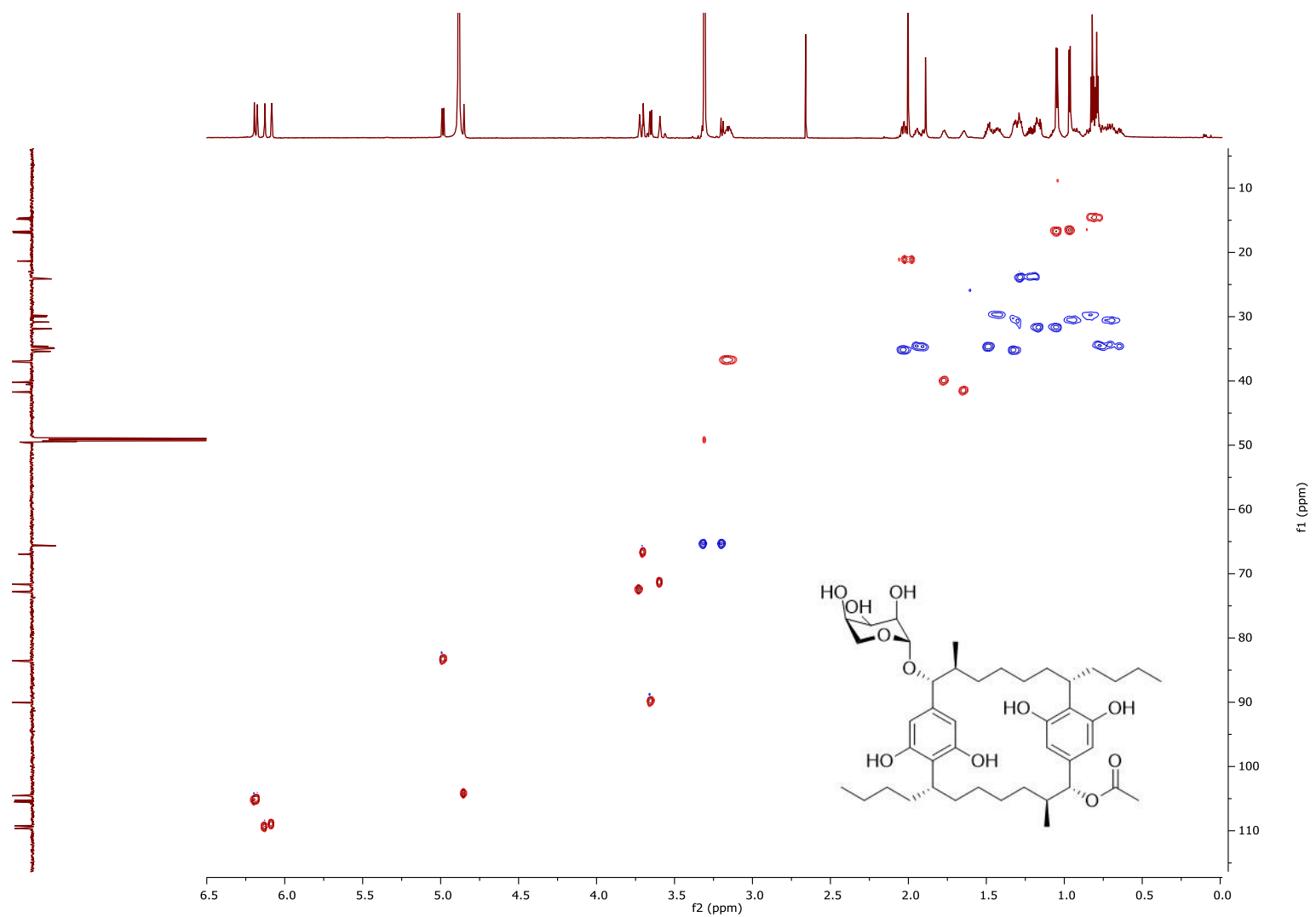
APPENDIX A (continued)

Figure 83: HSQC spectrum (600 MHz MeOd-d₄) of ribocyclophane C, see Table VI
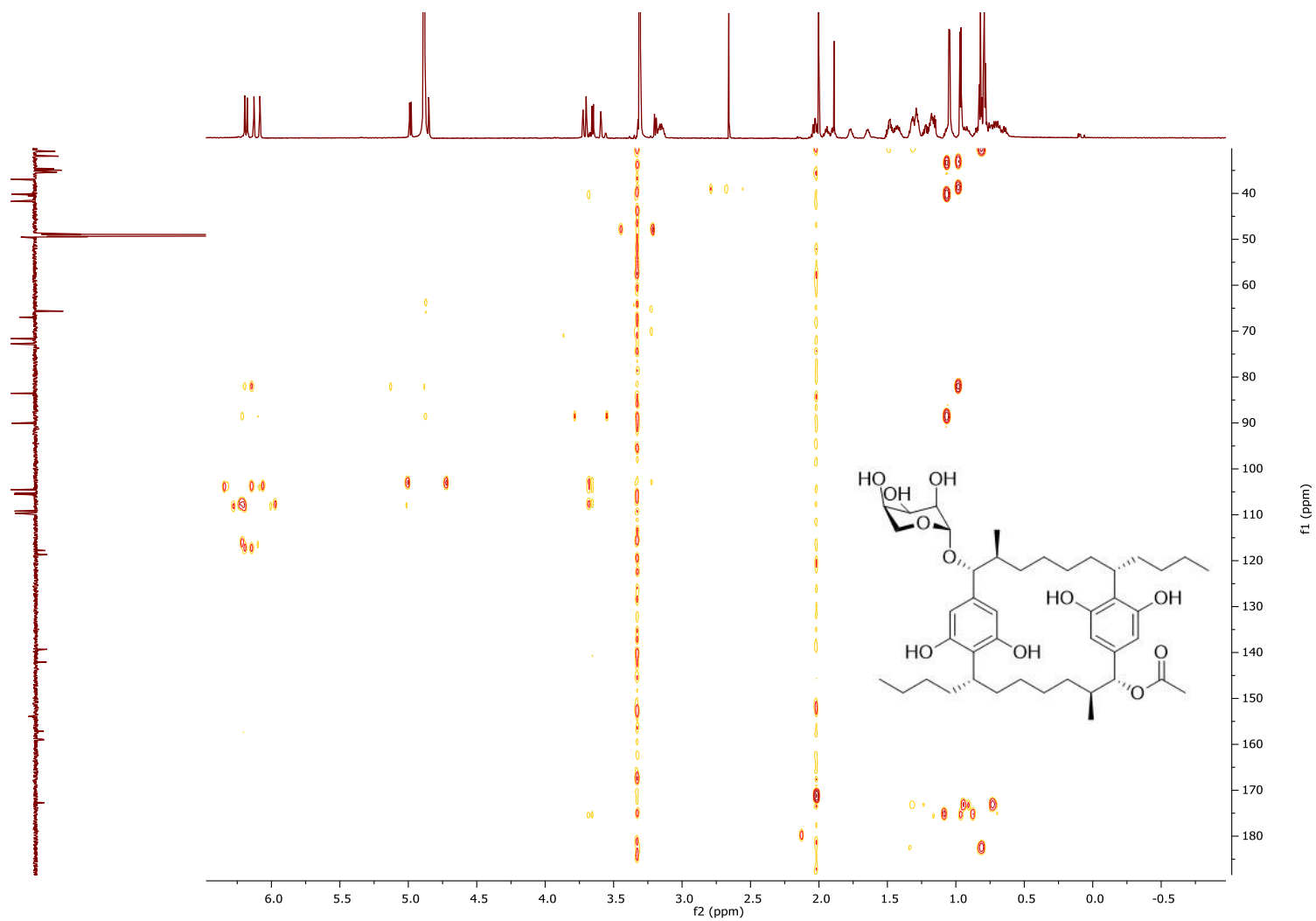
APPENDIX A (continued)
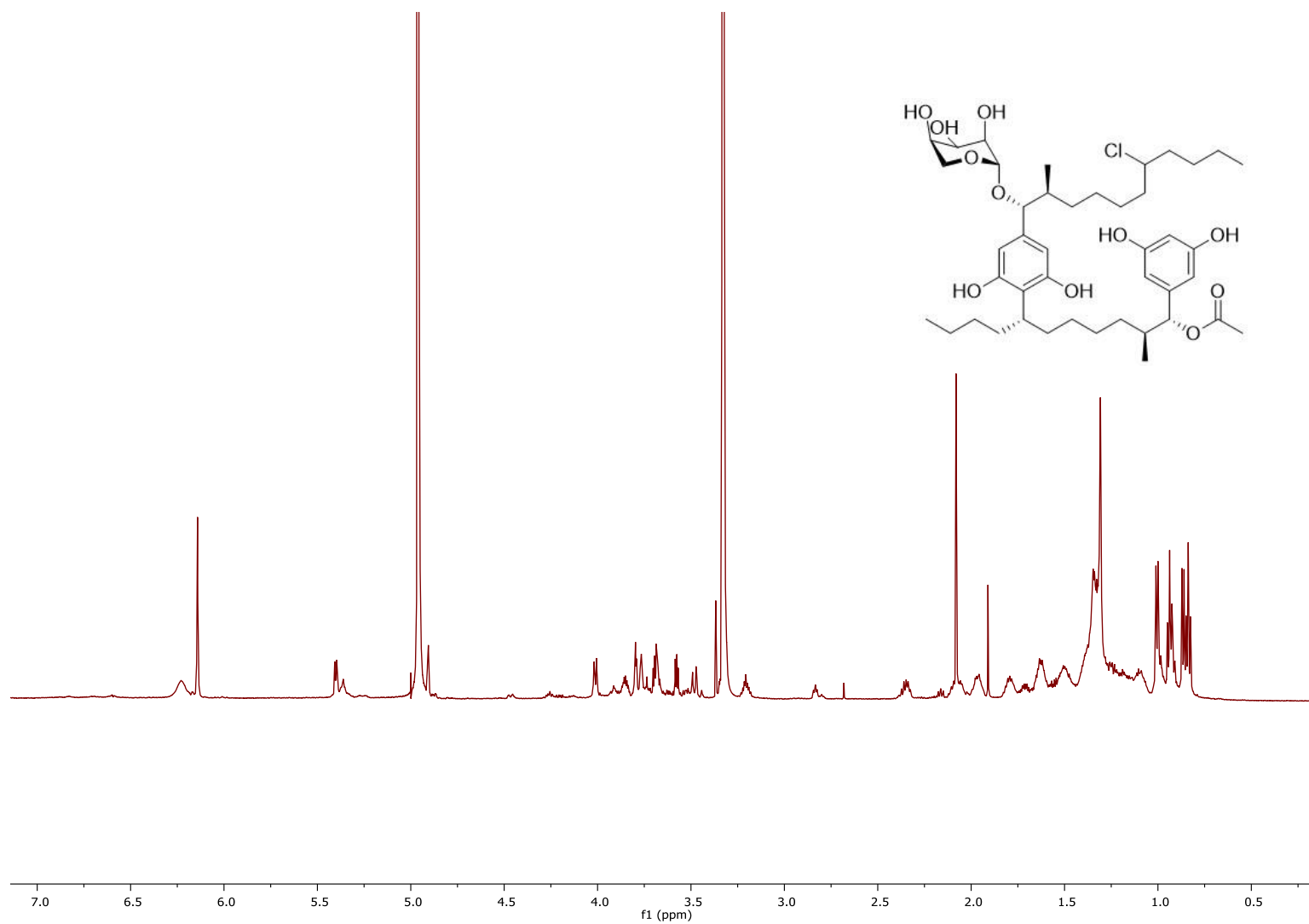

Figure 84: HMBC spectrum (600 MHz MeOd-d₄) of ribocyclophane C, see Table VI

APPENDIX A (continued)

Figure 85: ^1H NMR spectrum (900 MHz MeOd-d_4) of ribocyclophane D, see Table VII

APPENDIX A (continued)

Figure 86: DEPTQ spectrum (225 MHz MeOd-d₄) of ribocyclophane D, see Table VII
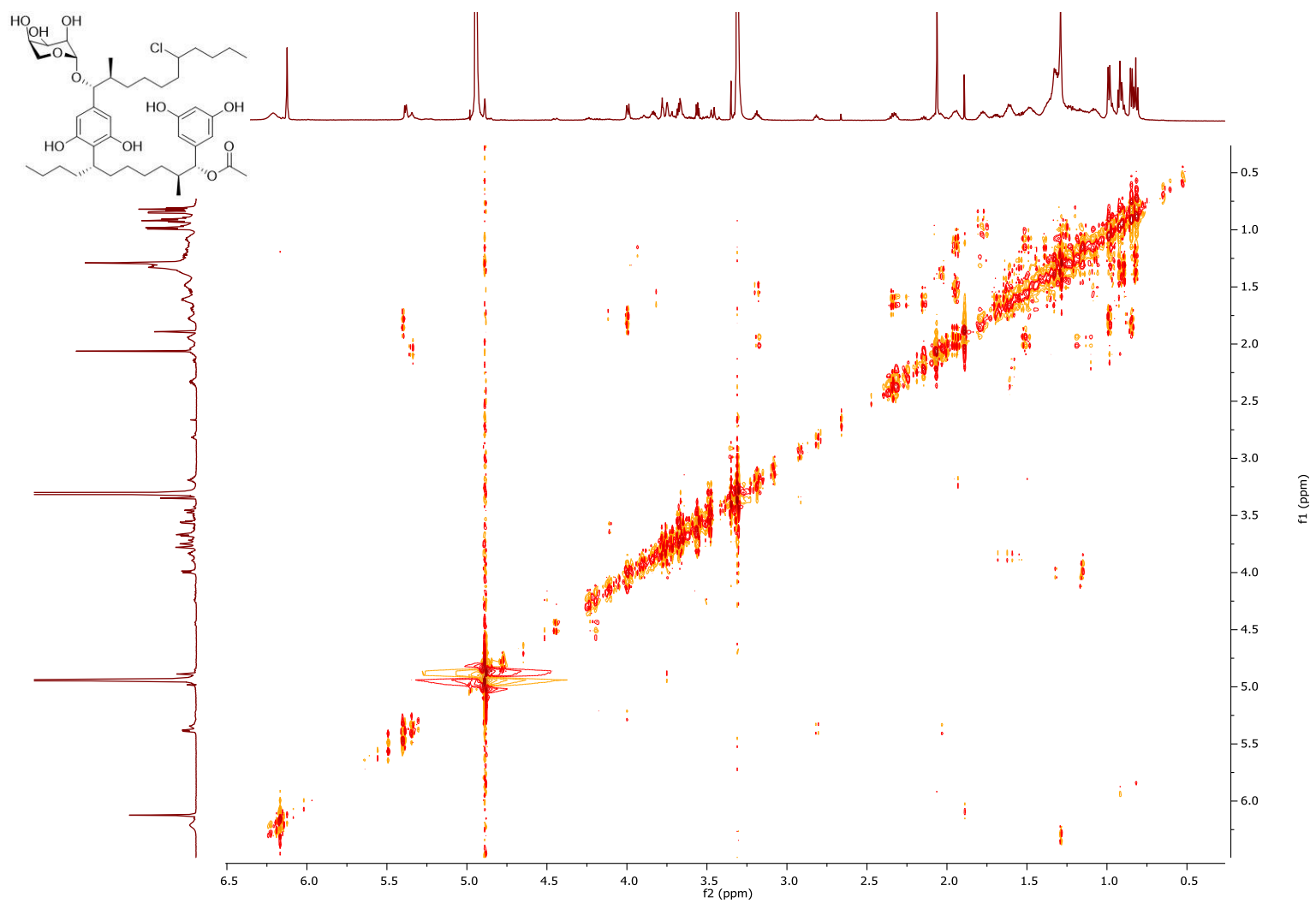
APPENDIX A (continued)

Figure 87: COSY spectrum (900 MHz MeOd-d₄) of ribocyclophane D, see Table VII
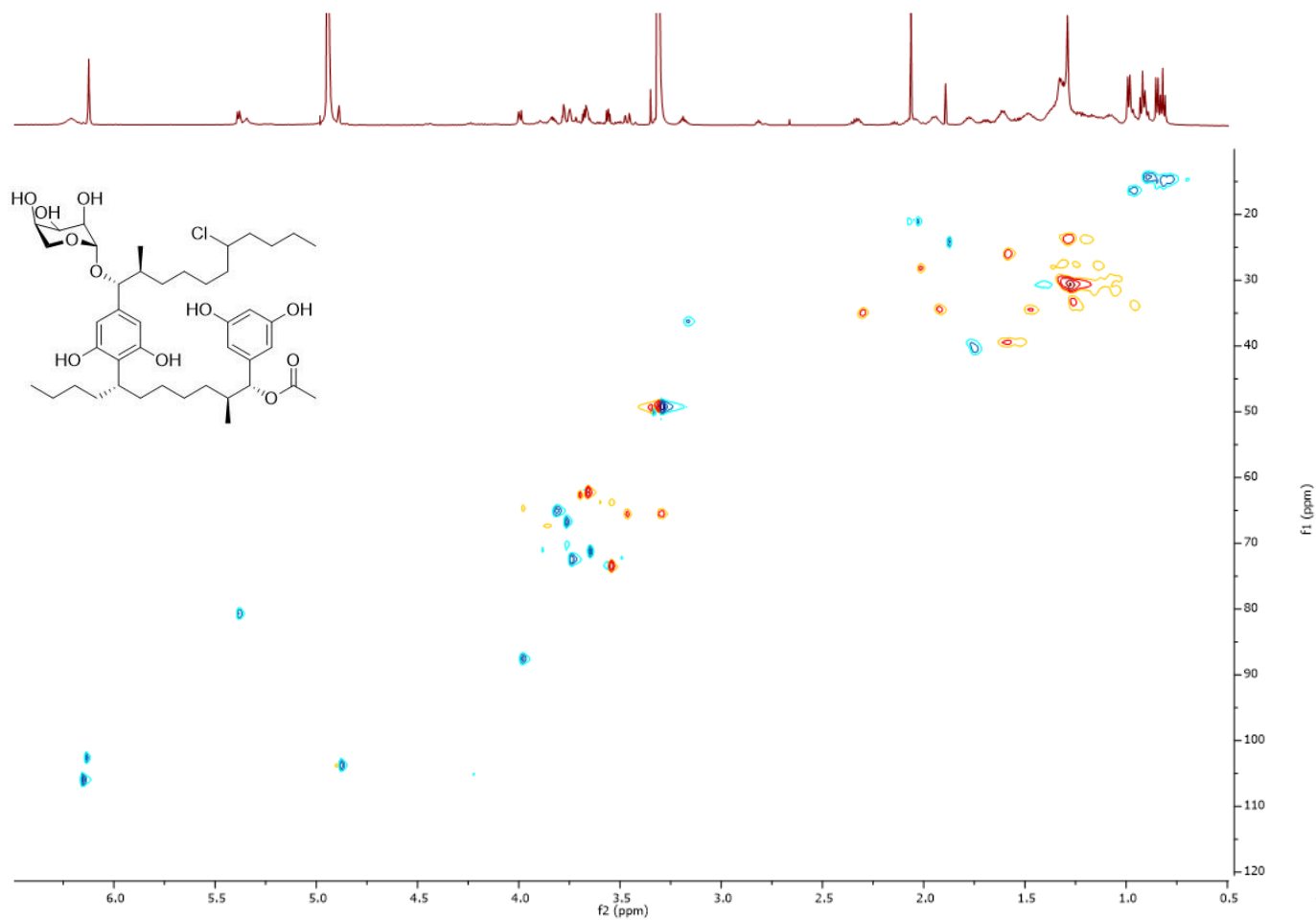
APPENDIX A (continued)

Figure 88:Edited HSQC spectrum (900 MHz MeOd-d$_4$) of ribocyclophane D, see Table VII

APPENDIX A (continued)

Figure 89: HMBC spectrum (900 MHz MeOd-d₄) of ribocyclophane D, see Table VII

APPENDIX A (continued)

Figure 90: ^1H NMR spectrum (900 MHz MeOd-d_4) of ribocyclophane E, see Table VII

APPENDIX A (continued)

Figure 91: COSY spectrum (900 MHz MeOd-d₄) of ribocyclophane E, see Table VII

APPENDIX A (continued)

Figure 92: Edited HSQC spectrum (900 MHz MeOd-d₄) of ribocyclophane E, see Table VII
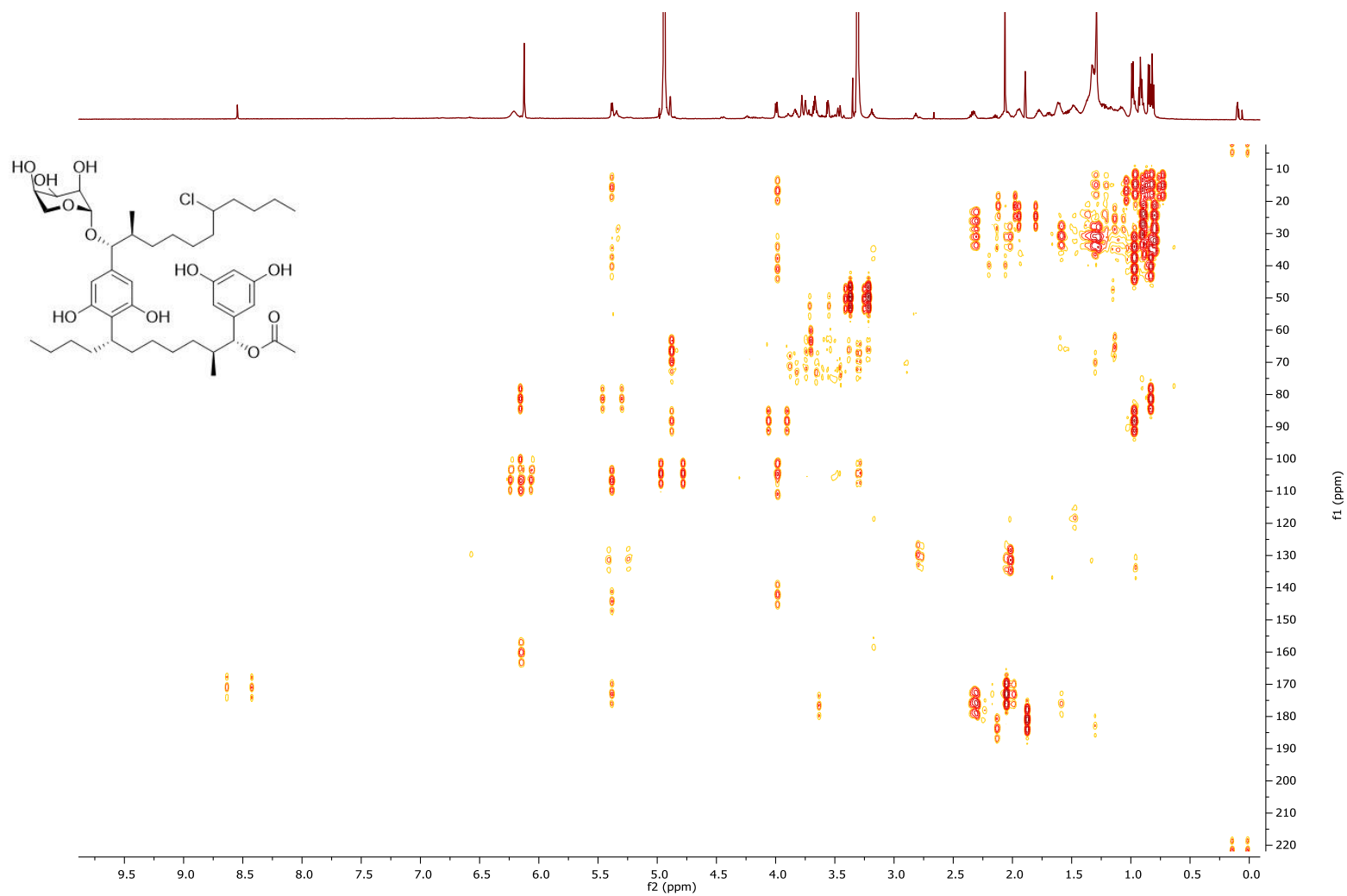
APPENDIX A (continued)

Figure 93: HMBC spectrum (900 MHz MeOd-d₄) of ribocyclophane E, see Table VII
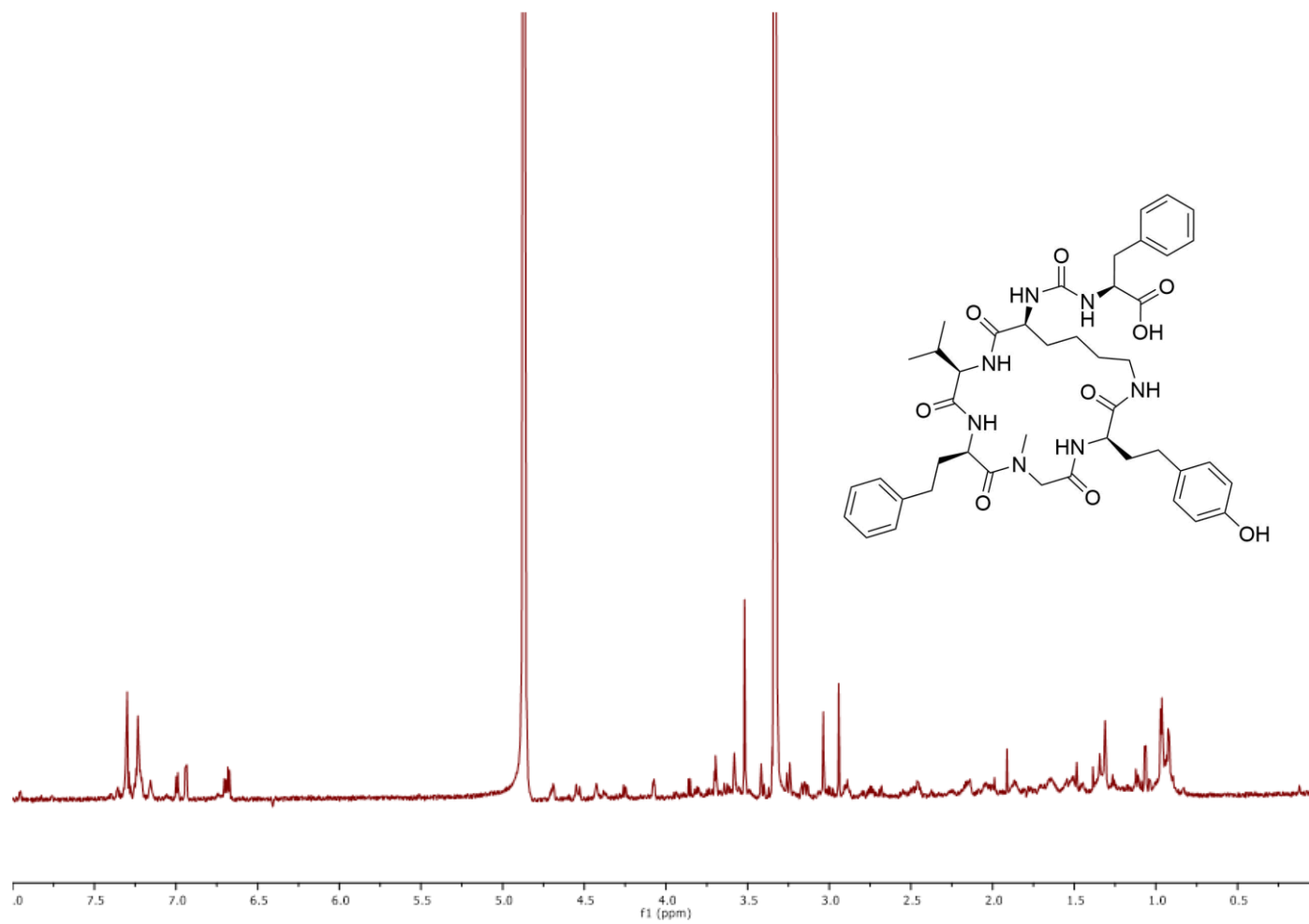
APPENDIX A (continued)

Figure 94: ^1H NMR spectrum (900 MHz MeOd-d$_4$) of anabaenopeptin UIC827
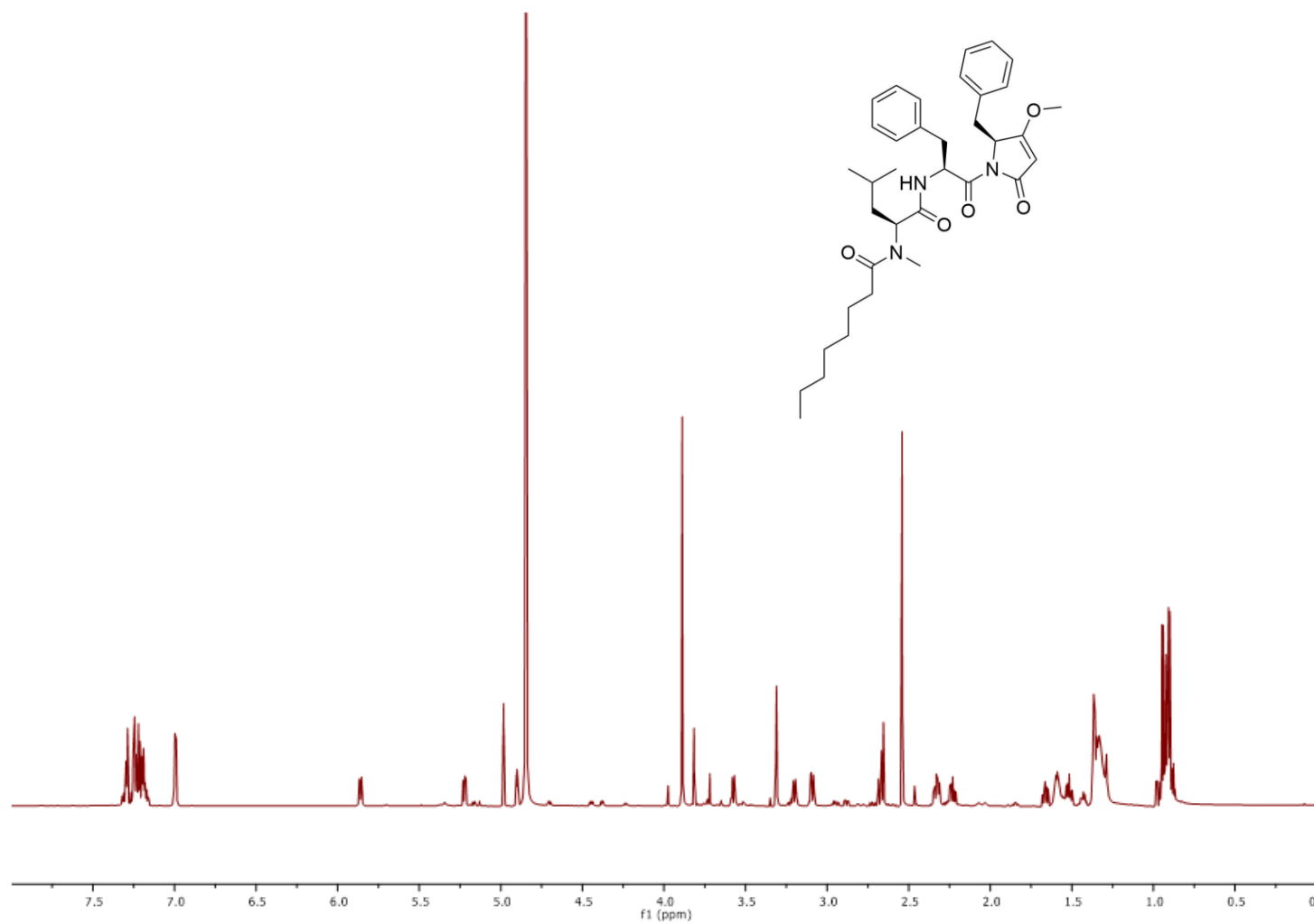
APPENDIX A (continued)

Figure 95:^{1}H NMR spectrum (900 MHz MeOd-d$_4$) of nostopyrrolidonamide, see Table XIV
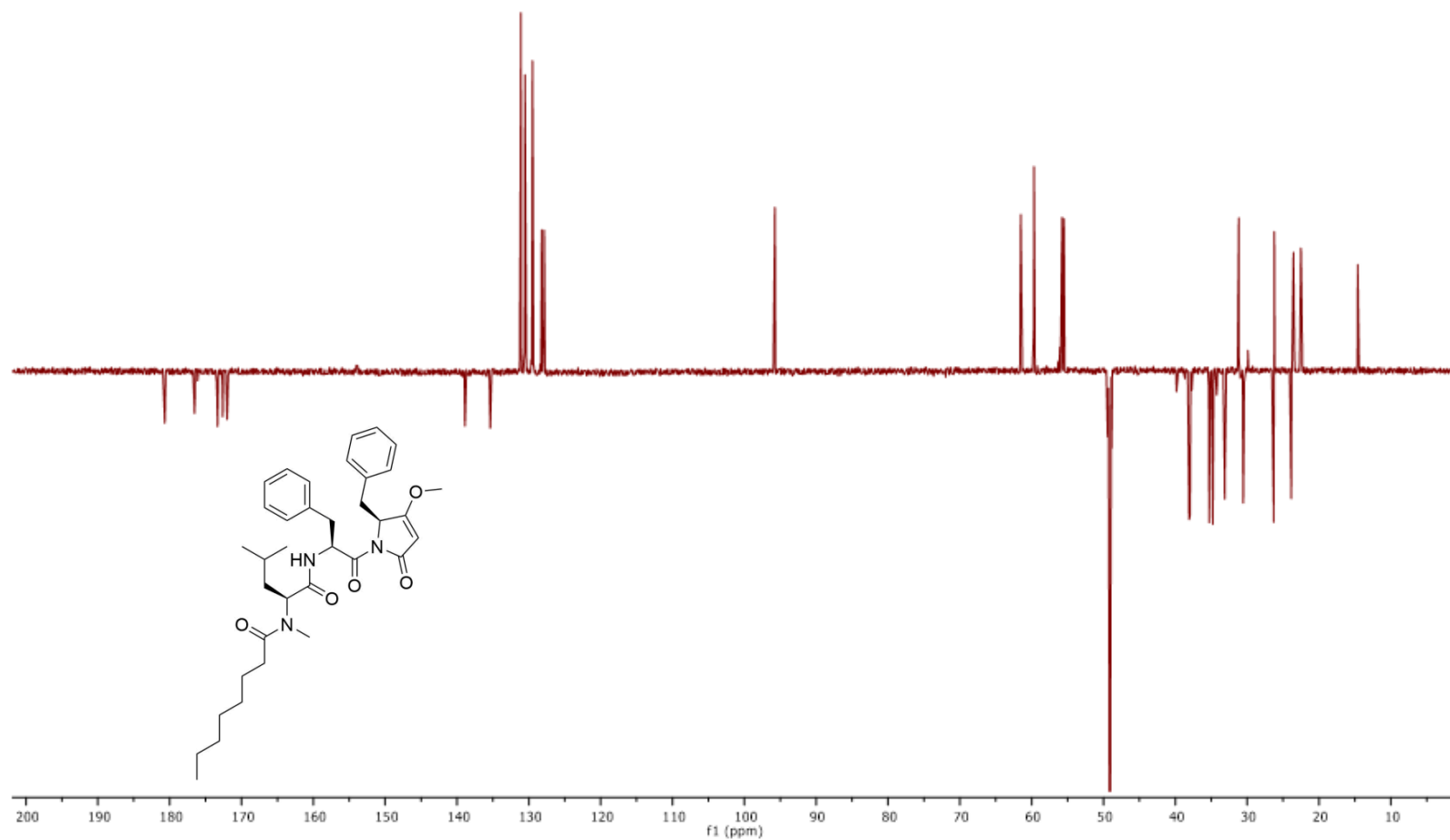
APPENDIX A (continued)

Figure 96: DEPTQ spectrum (225 MHz MeOd-d₄) of nostopyrrolidonamide, see Table XIV

APPENDIX A (continued)

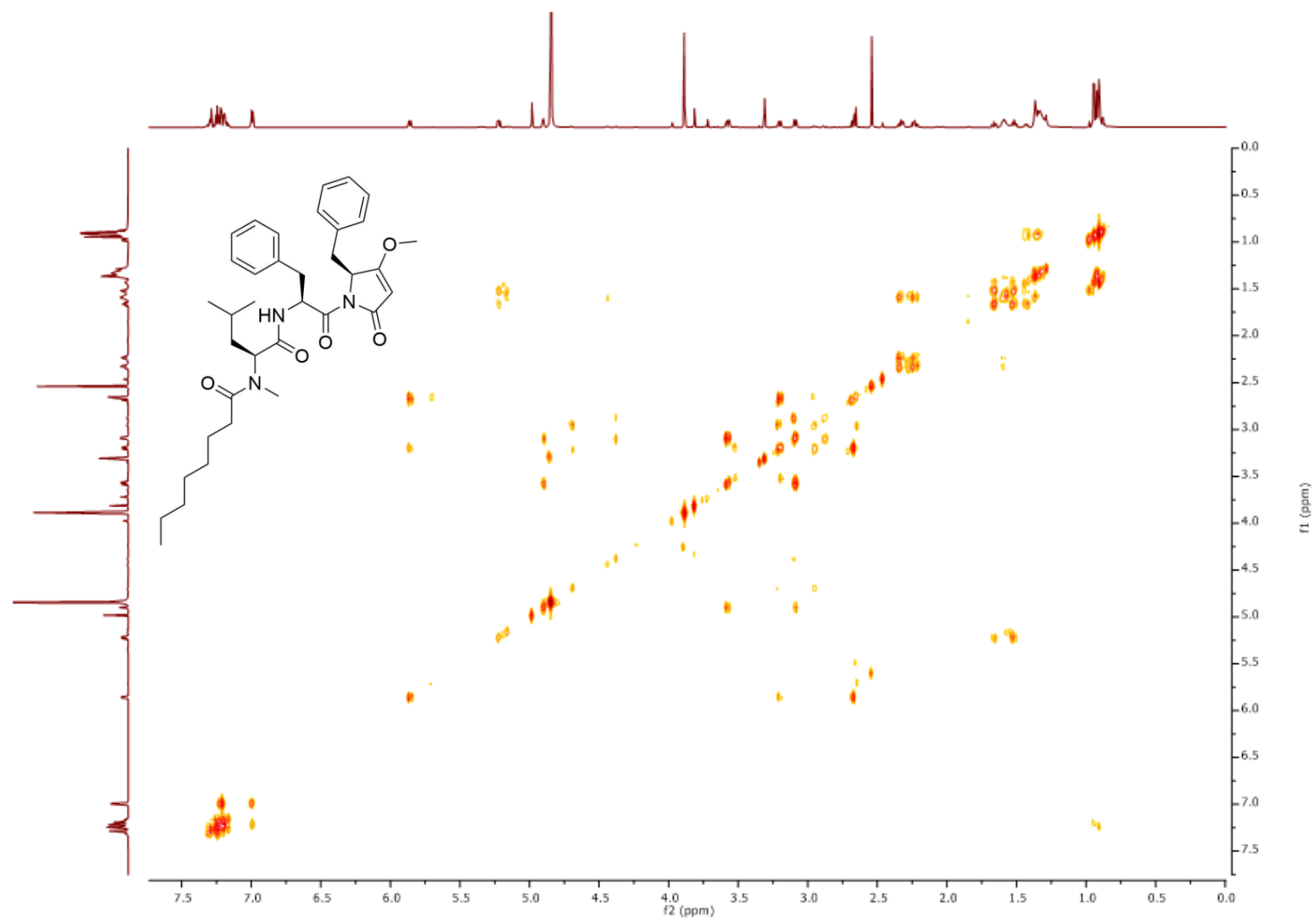*Figure 97: COSY spectrum (800 MHz MeOd-d₄) of nostopyrrolidonamide, see Table XIV*

APPENDIX A (continued)

Figure 98: HSQC spectrum (800 MHz MeOd-d₄) of nostopyrrolidonamide, see Table XIV

APPENDIX A (continued)

Figure 99: HMBC spectrum (800 MHz MeOd-d₄) of nostopyrrolidonamide, see Table XIV

APPENDIX A (continued)
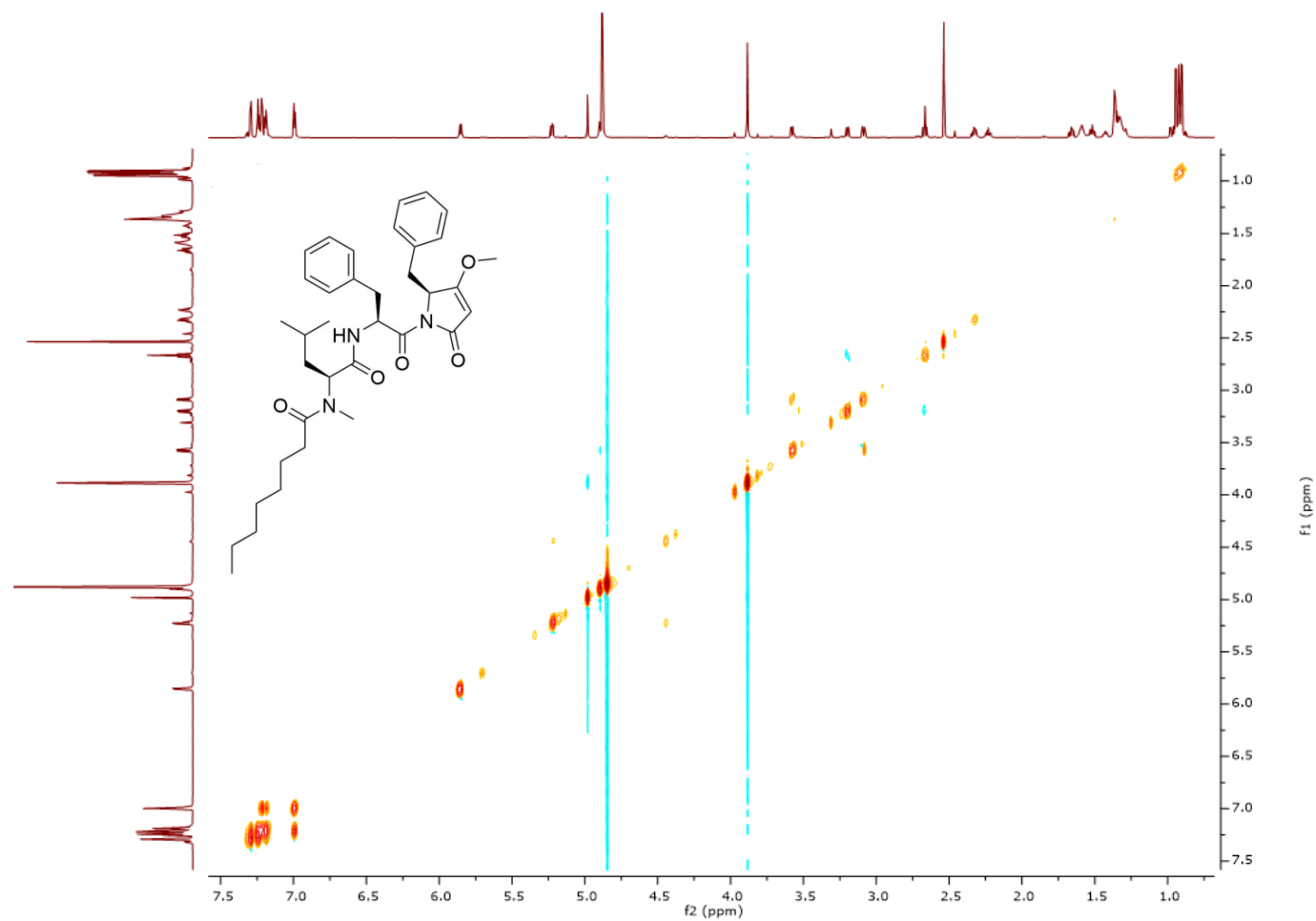
Figure 100: ROESY spectrum (800 MHz MeOd-d₄) of nostopyrrolidonamide, see Table XIV
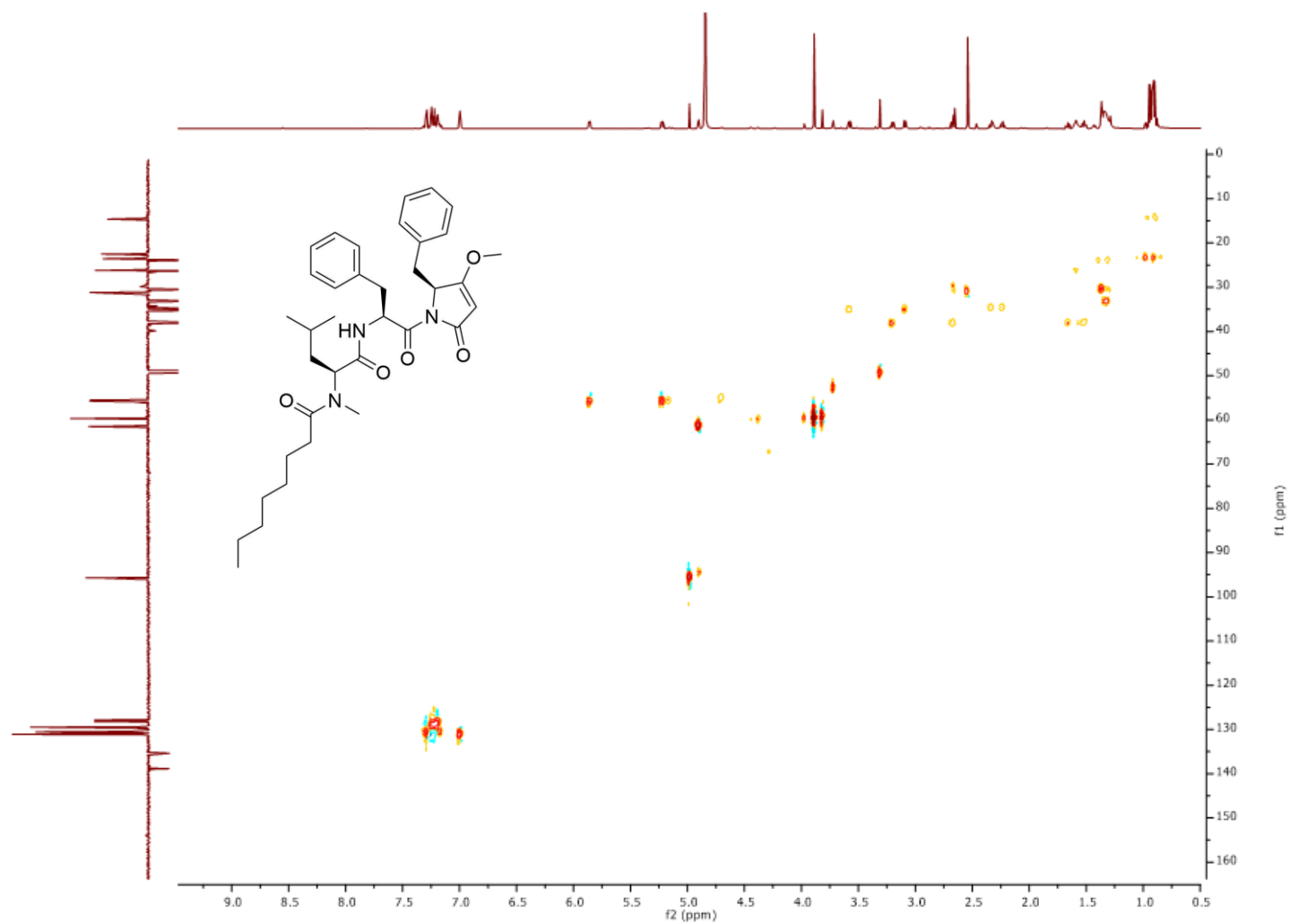
APPENDIX A (continued)
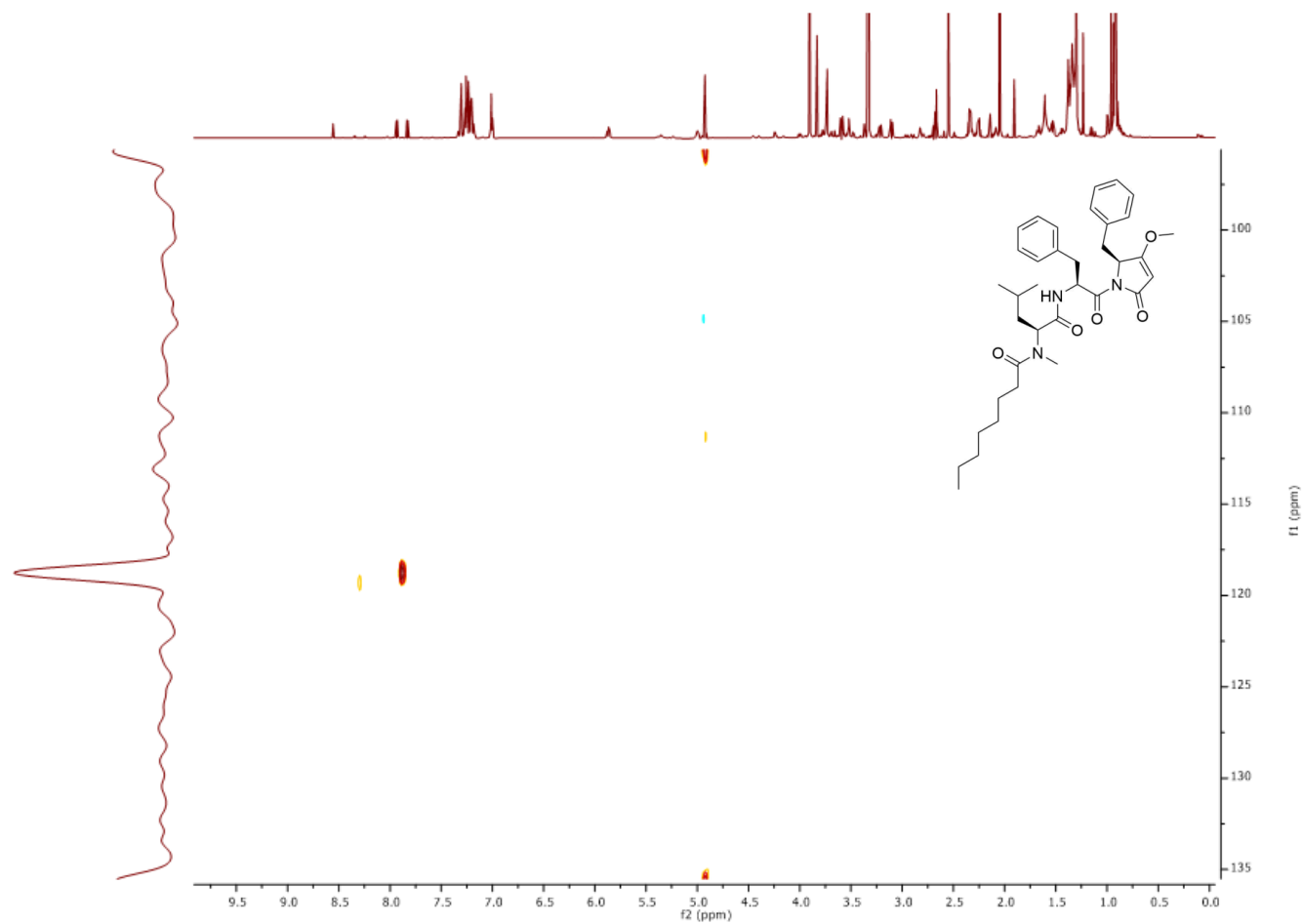
Figure 101: ^{15}NHSQC spectrum (900 MHz MeOH-d$_3$) of nostopyrrolidonamide, see Table XIV

APPENDIX A (continued)
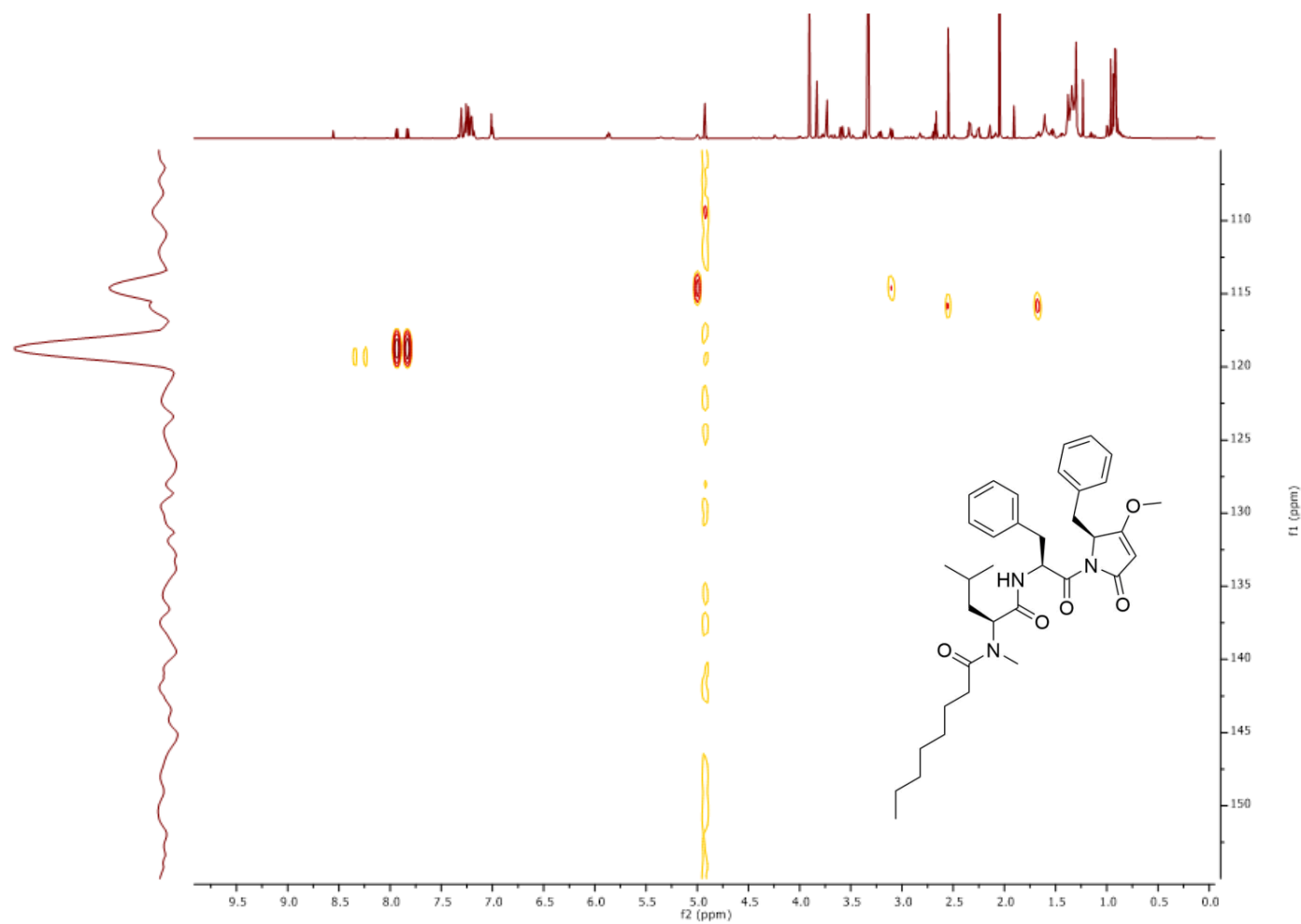
Figure 102: ^{15}NHMBC spectrum (900 MHz MeOH-d$_3$) of nostopyrrolidonamide, see Table XIV
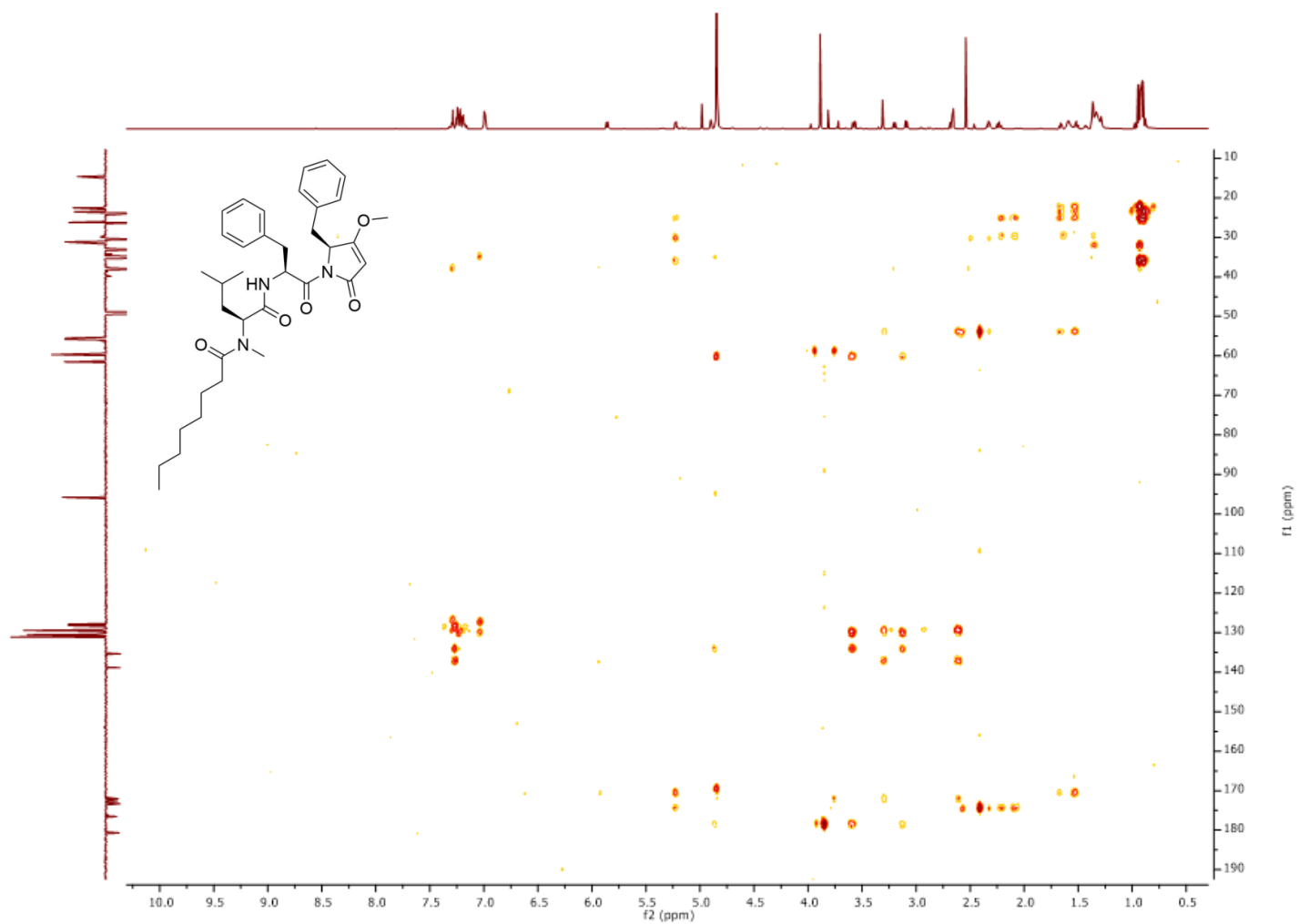
APPENDIX A (continued)
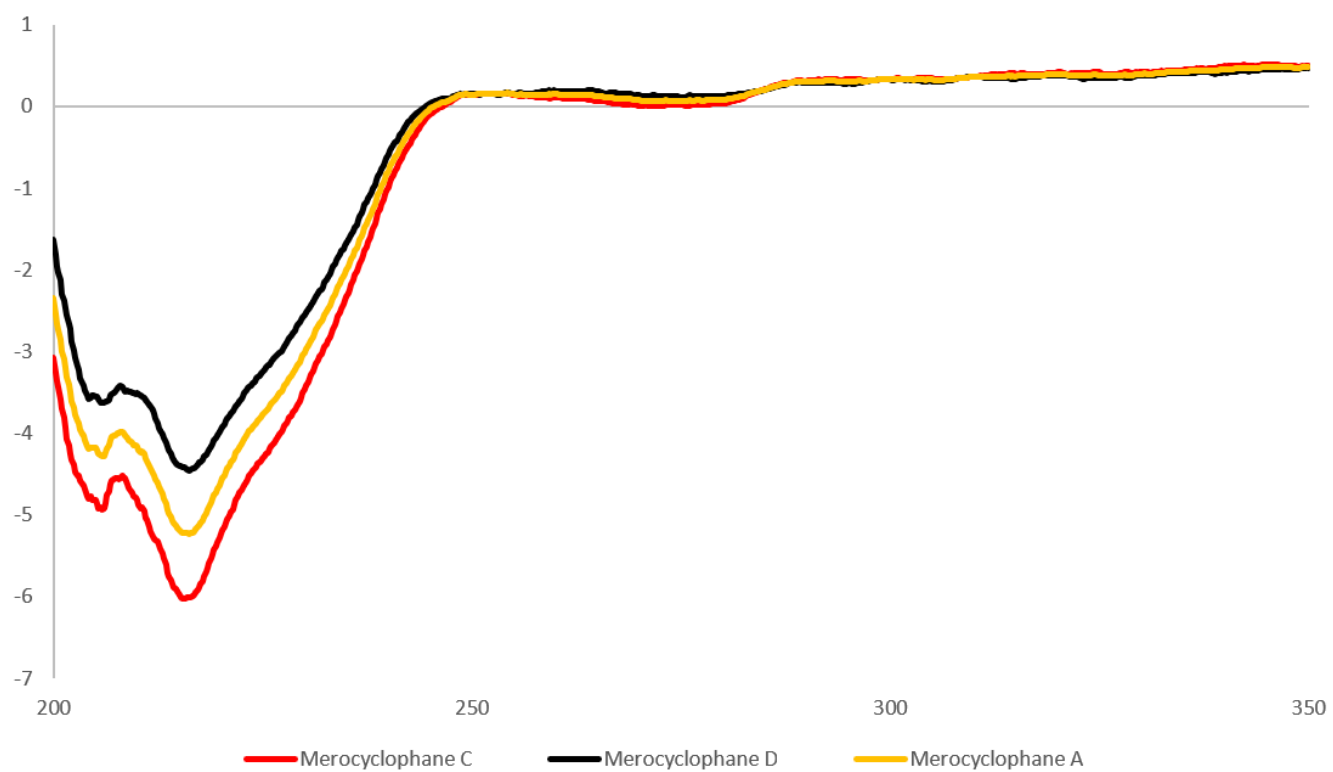
Figure 103: ECD spectra of merocyclophanes A, C, and D

APPENDIX A (continued)
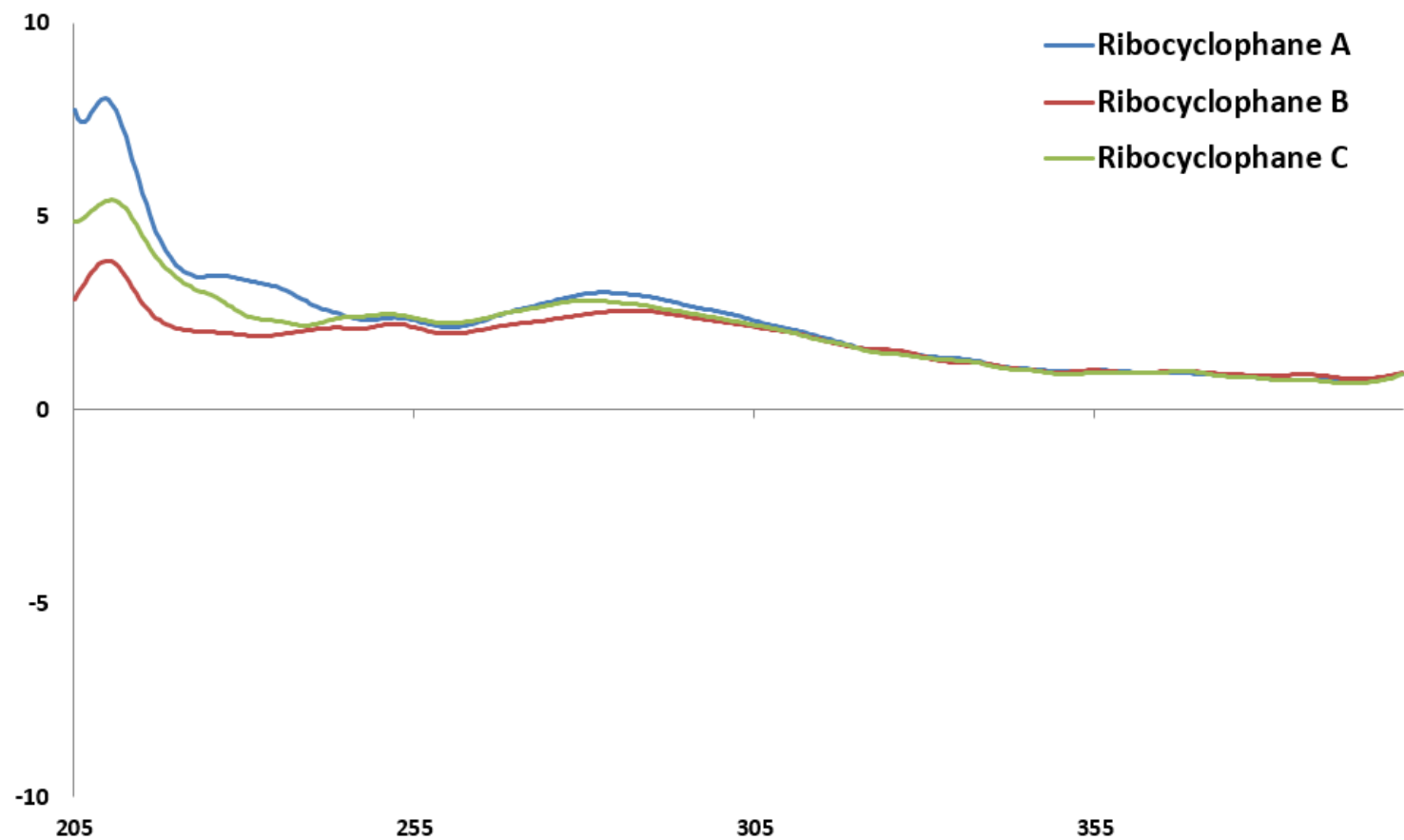
Figure 104: ECD spectra of ribocyclophanes A-C

APPENDIX A (continued)
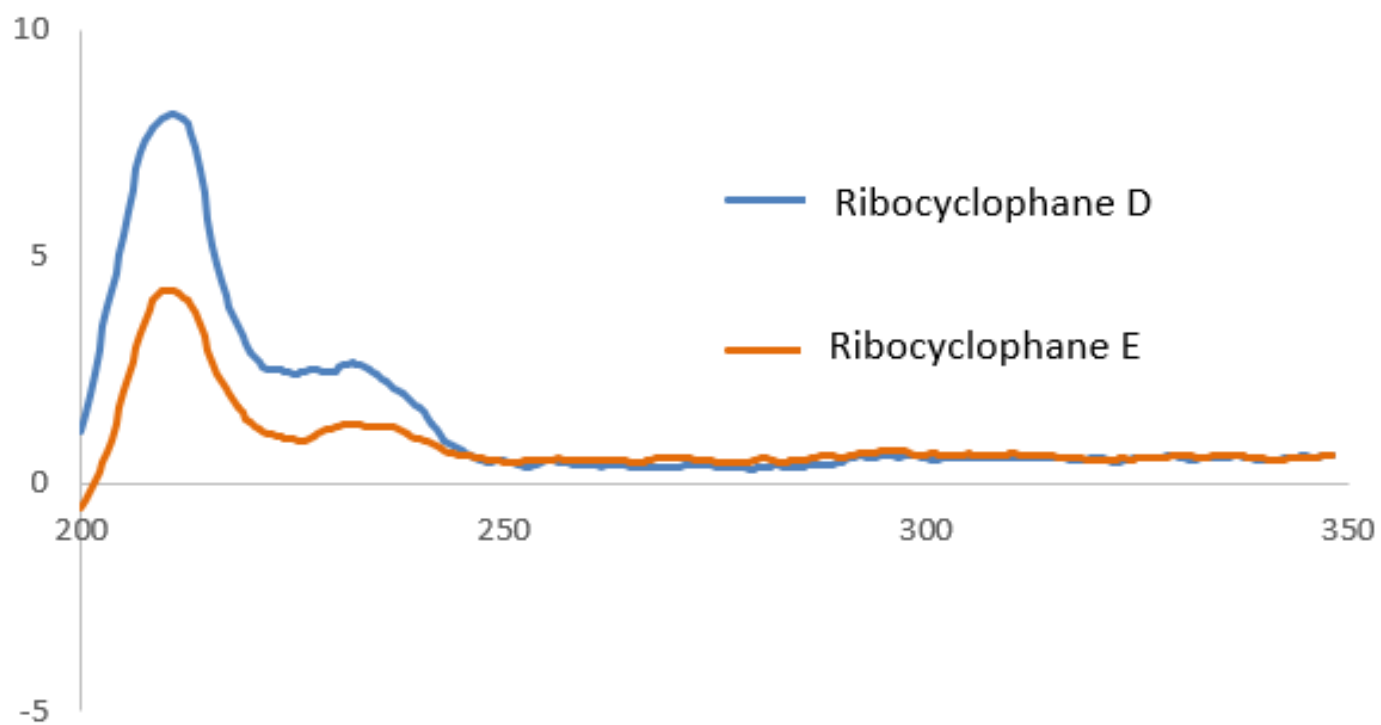
Figure 105: ECD spectra of ribocyclophanes D and E

APPENDIX B – Micrographs of strains studied in this dissertation
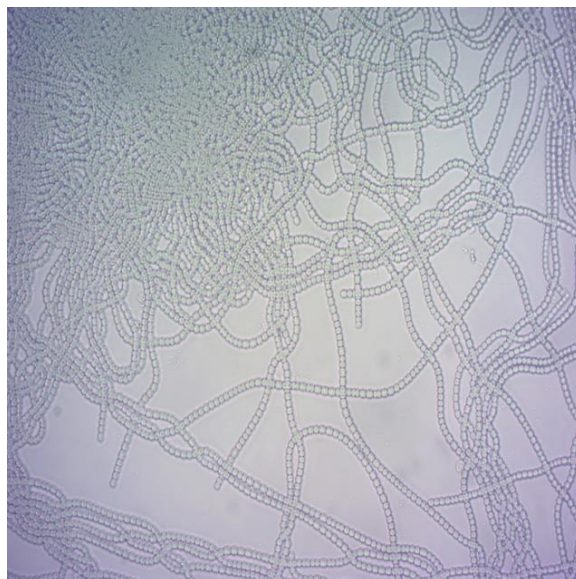
Figure 106: Micrograph of UIC 10110 Nostoc sp.
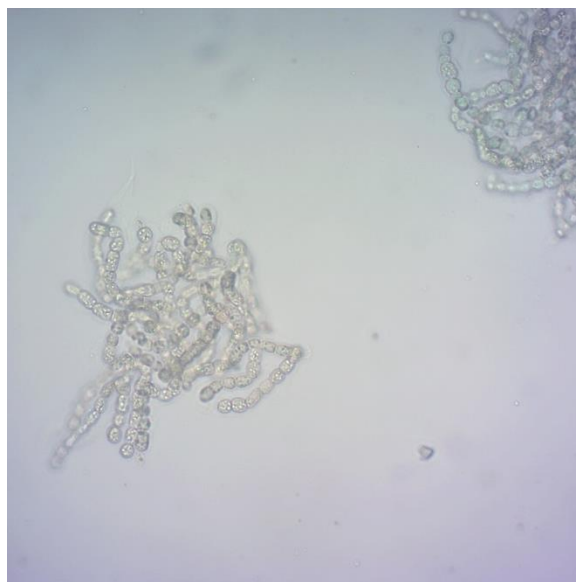
Figure 107: Micrograph of UIC 10250 Nostoc sp.

APPENDIX B (continued)

Figure 108:Micrograph of UIC 10448 Nostoc sp.

Figure 109: Micrograph of UIC 10366 Nostoc sp.

APPENDIX B (continued)
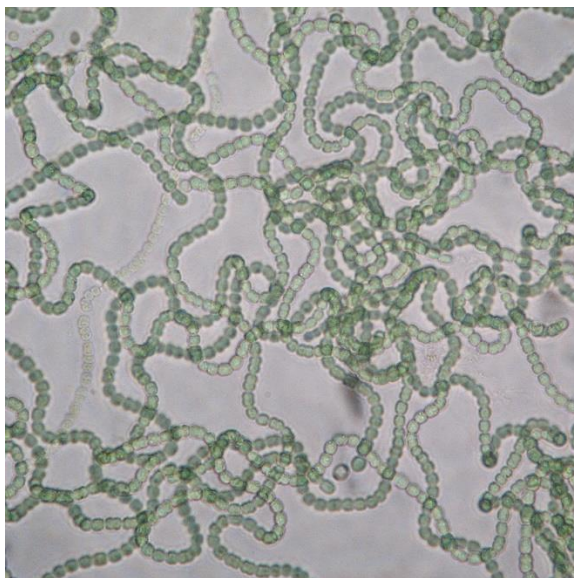
Figure 110: Micrograph of UIC 10279 Nostoc sp.
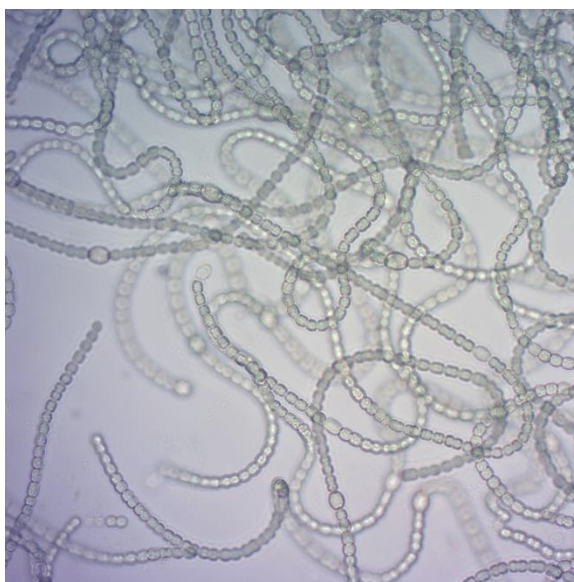
Figure 111: Micrograph of UIC 10534 Nostoc sp.

APPENDIX B (continued)

Figure 112: Micrograph of UIC 10630 Nostoc sp.
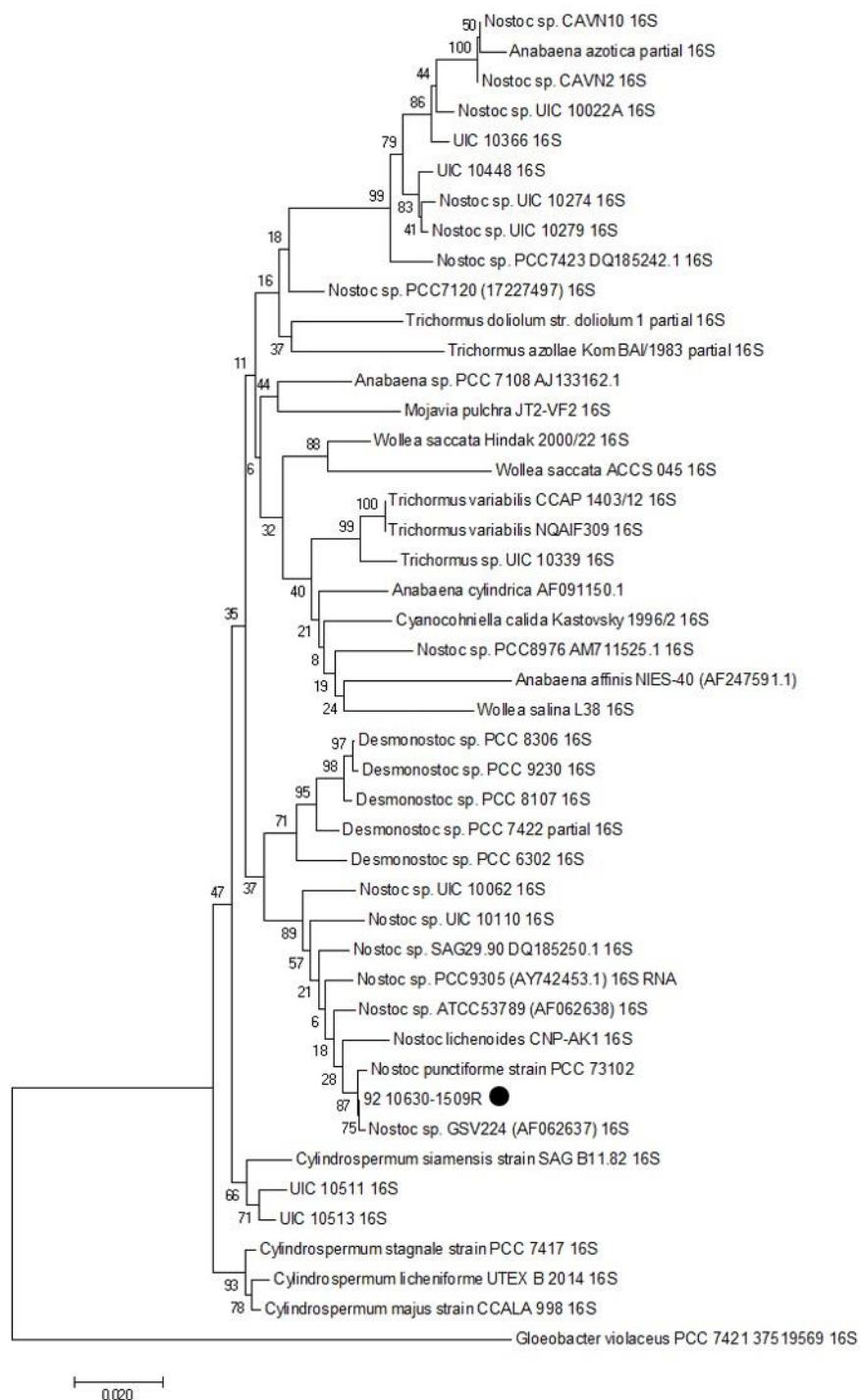
Figure 113: 16S rRNA neighbor-joining tree of Nostocaceae strains. Values at branches were determined by 1000 bootstrap replicates. UIC 10630 Nostoc sp. in Nostoc Group I highlighted with black circle.

APPENDIX C – Chapter 2 extra figures

Soil L Artificial Bloom Analysis

| Taxon	Z media	BG12 media	BG12$_o$ media
Tolypothrichaceae	-0.01812	-0.01816	-0.01655
Gloeoetrichiaceae	6.58E-05	0	0.000926
Fortieaceae	-0.06102	-0.06198	-0.03645
Nostocaceae	-0.2304	-0.2579	0.5595
Aphanizomenaceae	-0.01157	-0.01151	-0.01137
Rivulariaceae	0.2846	0.001934	-0.0036
Hapalosiphonaceae	-0.00232	-0.00232	-0.00232
Scytonemataceae	-0.00232	-0.00232	-0.00228
Chroococcidiopsidaceae	0	0	0.000227
Nostocales	-0.01153	-0.01154	-0.01151
Oscillatoriophycidae	-0.00232	-0.00232	-0.00232
Coleofasciculaceae	-0.0045	-0.00458	-0.00459
Microcoleaceae	-0.07455	-0.07859	-0.0772
Oscillatoriaceae	-0.00232	-0.00228	-0.00224
Dermocarpellaceae	0	0	6.63E-05
Spirulinaceae	-0.00926	-0.00926	-0.00926
Chamaesiphonaceae	0	0	6.06E-05
Leptolyngbyaceae	0.2303	0.546	-0.295
Merismopediaceae	-0.00232	-0.00232	-0.00226
Pseudanabaenaceae	-0.05552	-0.05613	-0.05663
Schizotrichaceae	0	0	0.000101
Synechococcaceae	-0.01389	-0.01378	-0.01389
Synechococcales	-0.00884	-0.00882	-0.00909
Gloeobacteraceae	-0.00463	-0.00463	-0.00463

Figure 114: Heat map displaying the percent change in the relative abundances of cyanobacterial taxa due to artificial bloom growth in differing media types for soil collection L. The more red a color the greater the percent decrease in relative abundance, the more green a color the greater the percent increase in relative abundance. Bold outlines represent statistically significant changes p < 0.05.

APPENDIX C (continued)

Soil I Artificial Bloom Analysis

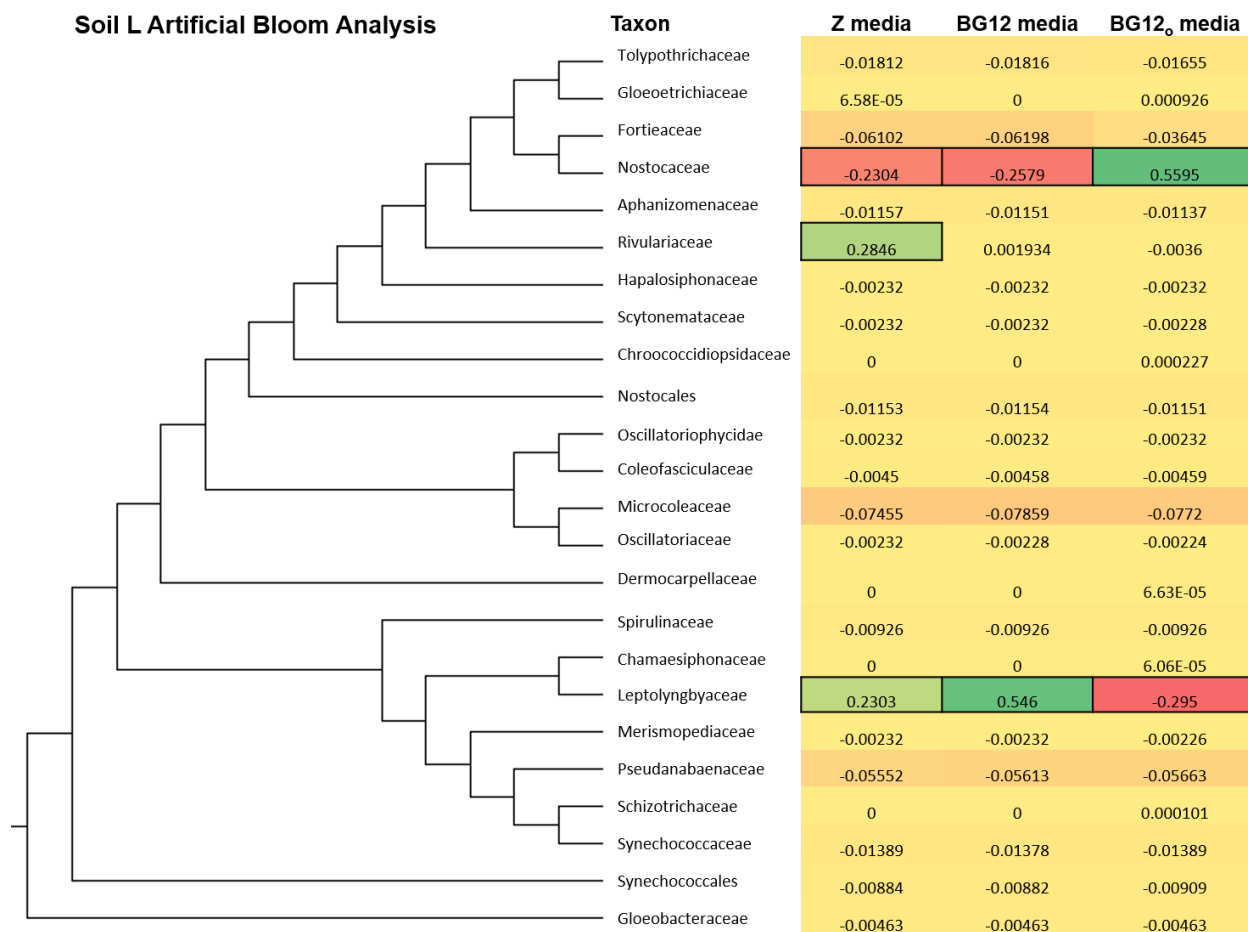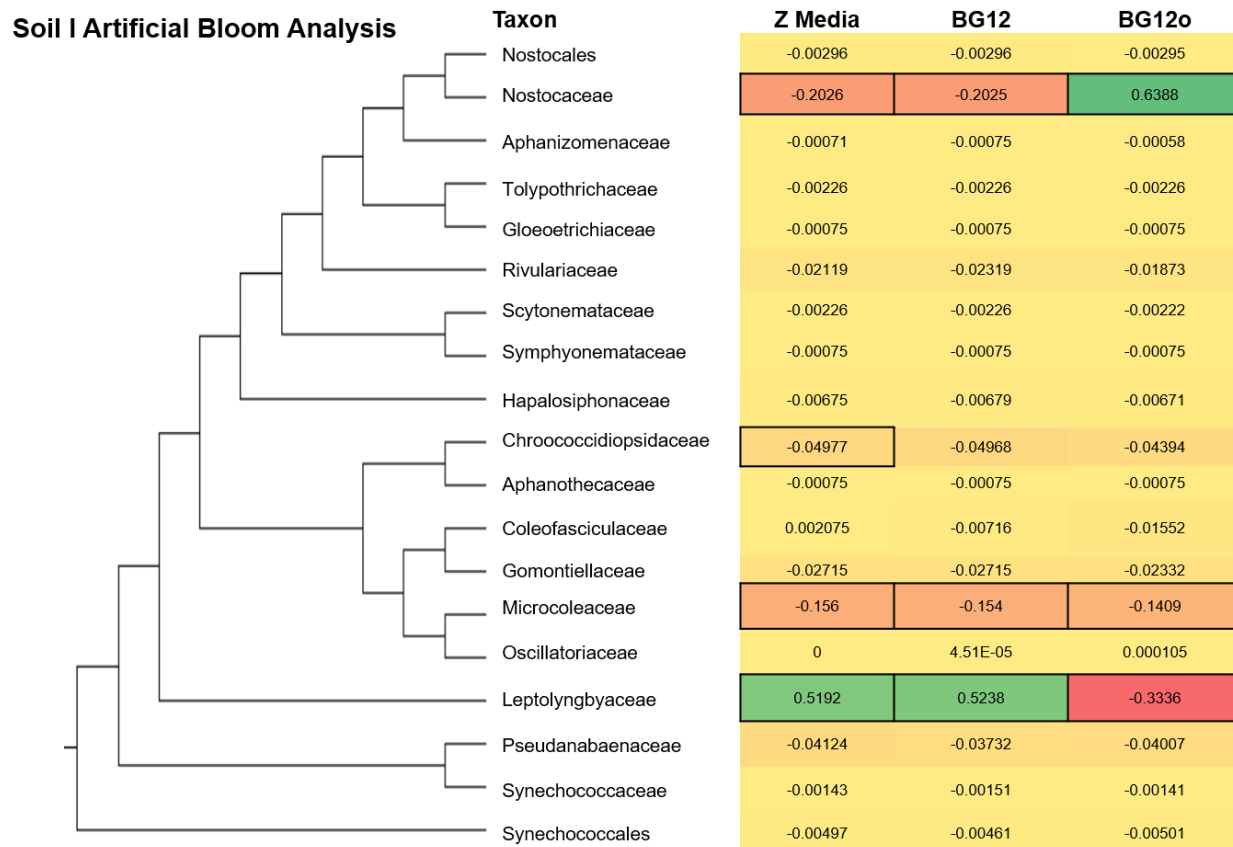| Taxon	Z Media	BG12	BG12o
Nostocales	-0.00296	-0.00296	-0.00295
Nostocaceae	-0.2026	-0.2025	0.6388
Aphanizomenaceae	-0.00071	-0.00075	-0.00058
Tolypothrichaceae	-0.00226	-0.00226	-0.00226
Gloeoetrichiaceae	-0.00075	-0.00075	-0.00075
Rivulariaceae	-0.02119	-0.02319	-0.01873
Scytonemataceae	-0.00226	-0.00226	-0.00222
Symphyonemataceae	-0.00075	-0.00075	-0.00075
Hapalosiphonaceae	-0.00675	-0.00679	-0.00671
Chroococcidiopsidaceae	-0.04977	-0.04968	-0.04394
Aphanothecaceae	-0.00075	-0.00075	-0.00075
Coleofasciculaceae	0.002075	-0.00716	-0.01552
Gomontiellaceae	-0.02715	-0.02715	-0.02332
Microcoleaceae	-0.156	-0.154	-0.1409
Oscillatoriaceae	0	4.51E-05	0.000105
Leptolyngbyaceae	0.5192	0.5238	-0.3336
Pseudanabaenaceae	-0.04124	-0.03732	-0.04007
Synechococcaceae	-0.00143	-0.00151	-0.00141
Synechococcales	-0.00497	-0.00461	-0.00501

Figure 115: Heat map displaying the percent change in the relative abundances of cyanobacterial taxa due to artificial bloom growth in differing media types for soil collection I. The more red a color the greater the percent decrease in relative abundance, the more green a color the greater the percent increase in relative abundance. Bold outlines represent statistically significant changes p < 0.05.

Table XV: List of macroscopic collections and soil collections and where they were collected from.

Collection Name	Collection Type	Collection Location
Collection A	Macroscopic	Cornell Ave Chicago, IL
Collection B	Macroscopic	Lincoln Park Chicago, IL
Collection C	Macroscopic	Sleepy Hollow State Park Lansing, MI
Collection D	Macroscopic	Parchment MI
Collection E	Macroscopic	White Lake, MI
Soil A	Soil	Museum of Science and Industry Chicago, IL
Soil B	Soil	Museum of Science and Industry Chicago, IL
Soil C	Soil	Museum of Science and Industry Chicago, IL
Soil D	Soil	Museum of Science and Industry Chicago, IL
Soil E	Soil	UIC Field Station Downers Grove, IL
Soil F	Soil	UIC Field Station Downers Grove, IL
Soil L	Soil	UIC Field Station Downers Grove, IL
Soil G	Soil	Devil's Lake State Park Baraboo, WI
Soil H	Soil	Devil's Lake State Park Baraboo, WI
Soil I	Soil	Alma College Alma, MI
Soil K	Soil	Arrigo Park Chicago, IL

List of strains and accession numbers used for cyanobacterial specific primer set development

Aerosakkonema_funiforme_Lao26_(AB686261)

Alkalinema_pantanalense_CENA528__(KF246494)

Anabaena__PCC7108_(AJ133162)

Anabaena_affinis_NIES-40_(AF247591)

Anabaena_cylindrica__(AF091150)

Anabaenopsis__PCC9215_(AY038033)

Annamia_toxica_HOs24_(HQ658457)

Aphanizomenon_flos-aquae__(AY038035)

Arthronema_africanum_SAG_1.89_(KM019913)

Arthronema_africanum_SAG_12.89_(KM019974)

Arthronema_gygaxiana_UTCC_393_(AF218370)

Arthrospira_platensis_PCC_7345_(JN831265)

Arthrospira_platensis_PCC_9108_(DQ393284)

Arthrospira_platensis_PCC_9223_(DQ393285)

Arthrospira_platensis_UTEX_2340_(DQ393280)

Blennothrix_ganeshii_Mex3_(KM276544)

Blennothrix_ganeshii_Mex8_(KM276546)

Calothrix__PCC7103_(AM230700)

Calothrix__PCC7714_(AJ133164)

Calothrix_desertica_PCC7102_(AF132779)

Chlorogloeopsis_fritschii__(AB093489)

Chroococcidiopsis___(AJ344557)

Coleofasciculus_chthonoplastes_SAG_2209_(NR_125521)

Crinalium_epipsammum_PCC_9333_(NR_102461)

Cronbergia_siamensis_SAG_11.82_(KM019950)

Cyanobium__PCC7001_(AB015058)

Cyanocohniella_calida_CCALA_1049_(KJ737427)

Cylindrospermopsis_raciborskii__(AF092504)

Cylindrospermum__PCC7417_(AJ133163)

Desertifilum_fontinale_KR2012/2_(KJ028038)

Desmonostoc_muscorum_NIVA-CYA_818_(AM711524)

Fischerella_muscicola__(AB075984)

Fischerella_thermalis__(AB075987)

Geitlerinema__PCC7105_(AF132780)

Geitlerinema_sp._PCC_7407_(NR_102448)

Gloeobacter_violaceus__(37519569)

Gloeocapsa__PCC73106_(AF132784)

Gloeothece_membranacea__(X78680)

Haloleptolyngbya_alcalis_KR2005/106_(JN712770)

Halomicronema_sp._TFEP1_(AF320093)

Jaaginema_homogeneum_PMC252.05_(GQ859646)

Johanseninema_constricta_K2_5P2_(KJ140089)

Johanseninema_constricta_K2_5P3_(KJ140087)

Johanseninema_constricta_K2_6P1_(KJ140088)

Kamptonema_animale_CCALA_139_(KP221931)

Kastovskya_adunca_ATA6-11-RM4_(NR_125700)

Leptolyngbya__PCC7104_(AB039012)

Leptolyngbya__PCC7375_(AF132786)

Leptolyngbya_boryana_PCC_6306_(KJ939014)

Limnoraphis_robusta_CCALA_966_(NR_118325)

Limnothrix__CENA109_(EF088335)

Limnothrix__CENA110_(EF088338)

Limnothrix_planktonica_CHAB759_(JQ004025)

Limnothrix_planktonica_CHAB763_(JQ004026)

Lyngbya_aestuarii_PCC7419_(AB075989)

Microcoleus__PC_7113_(NR_102467)

Microcoleus_vaginatus_PCC_9802_(AF284803)

Microcystis_aeruginosa__(AF139300)

Microcystis_aeruginosa__(U40340)

Moorea_bouillonii_PAL08-16_(GU111927)

Moorea_bouillonii_PNG5-198(R)_(FJ041298)

Moorea_bouillonii_PNG5-198(R)_(FJ041299)

Moorea_producens_3L(T)_(EU315909)

Moorea_producens_3L(T)_(FJ151527)

Moorea_producens_JHB_(FJ151521)

Moorea_producens_NAC8-48_(GU724200)

Moorea_producens_PAL-8-17-08-2_(GQ231522)

Moorea_producens_PNG6-221_(FJ356669)

Moorea_producens_PNG6-221_(FJ356670)

Myxosarcina__PCC7325_(AJ344562)

Neosynechococcus_sphagnicola_sy1_(KJ469129)

Nodosilinea_epilithica_Kovacik_1998/7_(HM918677)

Nodularia__PCC7804_(AJ133181)

Nodularia__PCC9350_(AY038034)

Nodularia_spumigena__(DQ185241)

Nostoc__PCC6720_(DQ185240)

Nostoc__PCC7120_(17227497)

Nostoc__PCC7423_(DQ185242)

Nostoc__PCC8112_(AM711537)

Nostoc__PCC8976_(AM711525)

Nostoc__PCC9305_(AY742453)

Oculatella_subterranea_SP301/Zammit_2007/2_(HQ917689)

Okeania_comitata_3L-OSC__(EU244875)

Okeania_hirsuta_PAB-10-Feb-10-1_(NR_118586)

Okeania_lorea_FK12-17_(NR_118545)

Okeania_plumata_FK12-27_(NR_118585)

Oscillatoria__PCC7112_(AB074509)

Oscillatoria_acuminata_PCC6304_(AB039014)

Oscillatoria_sancta_PCC7515_(AF132933)

Oxynema_thaianum_CCALA_960_(NR_125585)

Pantanalinema_rosaneae_CENA516__(KF246483)

Phormidesmis_priestleyi_ANT.L66.1_(AY493581)

Phormidesmis_priestleyi_CYN71_(JQ687335)

Phormidium_lucidum_CY-012_(KC217548)

Planktolyngbya_limnetica_PMC271.06_(GQ859645)

Planktolyngbya_limnetica_S14_(KF487299)

Planktothricoides_raciborskii_CHAB3331_(JF429938)

Planktothricoides_raciborskii_NIES-207_(NR_040858)

Planktothrix__PCC_7811_(GQ351564)

Plectolyngbya_hodgsonii_ANT.LPR2.2_(AY493583)

Plectonema__SAG_38.90_(KM019916)

Plectonema_calothricoides_SAG_1463-4_(KM019914)

Pleurocapsa__PCC7319_(AB039006)

Pleurocapsa__PCC7327_(AB039007)

Pleurocapsa__PCC7516_(X78681)

Prochlorothrix_hollandica_PCC_9006_(NR_126312)

Pseudanabaena__PCC6802_(AB039016)

Pseudanabaena__PCC7403_(AB075995)

Pseudanabaena__PCC7408_(AB039020)

Schizothrix_Schizothrix_sp._(EU445293)

Scytonema_hofmanni__(AF132781)

APPENDIX C (continued)

Spirulina_major_0BB36S18_(AJ639890)

Spirulina_major_PCC_6313_(AM709631)

Stanieria__PCC7301_(AB039009)

Stanieria_cyanosphaera__(AF132931)

Symploca_atlantica_PCC8002_(AB075997)

Synechococcus__PCC6312_(AF448081)

Synechococcus__PCC7003_(AB015059)

Synechococcus__WH5701_(87303064)

Tolypothrix__CCMP1185_(AB075998)

Tolypothrix__PCC7415_(AM230668)

Tolypothrix__PCC7504_(AM230669)

Trichocoleus__CNP1-B1-4_(AY239603)

Trichocoleus__CNP1-Z1-C2_(AY239602)

Trichocoleus_desertorum_ATA4-8-CV2_(NR_125697)

Trichocoleus_sociatus_SAG_26.92_(EF654080)

Trichodesmium__NIBB1067_(X70767)

Trichormus_azollae__(AJ630454)

Trichormus_variabilis__(AJ630457)

Tychonema_bornetii_NIVA-CYA_60_(LM651414)

Tychonema_bourrellyi_CCAP_1459/11B_(NR_112123)

Tychonema_bourrellyi_NIVA-CYA_33/1_(LM651410)

Tychonema_bourrellyi_NIVA-CYA_96/3_(LM651417)

Tychonema_tenue_SAG_4.82_(GQ324973)

Wilmottia_murrayi_KGI28_(HQ873481)

Xenococcus__PCC7305_(AF132783)

Methylobacterium goesingense strain iEII3 (NR_115219)

<div align="center"><u>APPENDIX D – Media Ingredients</u></div>

Z media ingredients

To approximately 950 mL distilled H_2O add:

Quantity	Compound	Stock Solutions	Final Concentration (in Media)
6.2 mL	$NaNO_3$	75 g/L H_2O	5.47×10^{-3} M
1 mL	$Ca(NO_3)_2 \cdot 4H_2O$	59 g/L H_2O	2.50×10^{-4} M
3.1 mL	K_2HPO_4	10 g/L H_2O	1.78×10^{-4} M
0.5 mL	$MgSO_4 \cdot 7H_2O$	50 g/L H_2O	1.01×10^{-4} M
1 mL*	Na_2CO_3	20 g/L H_2O	1.89×10^{-4} M
2 mL	Tris/HCl, pH 7.8	1.00 M	2.00×10^{-3} M
1 mL	FeEDTA solution	-	-
80 µL	Gaffron solution	-	-
0.5 mL	f/2 vitamin solution	-	-

Make final volume up to 1 L with distilled H_2O, adjust pH to 7.8, and autoclave.

* Should sterile filter in Na_2CO_3 after autoclaving the media.

APPENDIX D (continued)

BG12 media ingredients

To 950 mL distilled H_2add:

Quantity	Compound	Stock Solutions	Final Concentration
1.5 g	$NaNO_3$	-	1.76×10^{-2} M
3.75 mL	K_2HPO_4	10 g/L H_2O	2.15×10^{-4} M
1.5 mL	$MgSO_4 \cdot 7H_2O$	50 g/L H_2O	3.04×10^{-4} M
1.0 mL	Na_2CO_3	20 g/L H_2O	1.89×10^{-4} M
0.72 mL	$CaCl_2 \cdot 2H_2O$	50 g/L H_2O	2.45×10^{-4} M
117 µL	$Na_2EDTA \cdot 2H_2O$	8 g/L H_2O	2.79×10^{-6} M
10.0 mL	FeEDTA Solution	-	-
1.0 mL	A5+Co Metal Sol'n	-	-

Make final volume up to 1 L with distilled H_2O, adjust pH to 7.4, and autoclave.

BG12$_o$ media is made by omitting the $NaNO_3$ from the BG12 media.

Copyright Clearance Center

RightsLink®

Home | Account Info | Help

ACS Publications
Most Trusted. Most Cited. Most Read.

Title:	Merocyclophanes C and D from the Cultured Freshwater Cyanobacterium Nostoc sp. (UIC 10110)
Author:	Daniel S. May, Wei-Lun Chen, Daniel D. Lantvit, et al
Publication:	Journal of Natural Products
Publisher:	American Chemical Society
Date:	Apr 1, 2017

Copyright © 2017, American Chemical Society

Logged in as:
Daniel May
University of Illinois at Chicago

Account #:
3001260166

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK | CLOSE WINDOW

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Ribocyclophanes A–E, Glycosylated Cyclophanes with Antiproliferative Activity from Two Cultured Terrestrial Cyanobacteria

Author: Daniel S. May, Hahk-Soo Kang, Bernard D. Santarsiero, et al

Publication: Journal of Natural Products

Publisher: American Chemical Society

Date: Mar 1, 2018

Copyright © 2018, American Chemical Society

VITA

Daniel S May

Education:
2013 – Present: PhD Candidate University of Illinois at Chicago Expected Defense 2018
2009 – 2013:   BS Alma College   Major: Biology, Minor: Chemistry

Positions and Employment:
2013 – Present: Graduate Research Assistant, University of Illinois at Chicago,
                         Department of Medicinal Chemistry and Pharmacognosy,
                         Advisor: Dr. Jimmy Orjala
2013 – 2014:   Teaching Assistant
                         College of Pharmacy, University of Illinois at Chiago
2011 – 2013:   Research Assistant
                         Alma College, Department of Biology
                         Advisor: Dr. Brian Doyle
2010 – 2013:   Teaching Assistant
                         Alma College, Department of Biology
                         Advisor: Mike Bishop

Ongoing Research Support:
T32 Grant # - 1T32AT007533   Orjala (PI)   01/01/2014-12/31/2018

Honors and Awards:
2018   Al R. Langerman Memorial Scholarship Award
2016   W.E. van Doren Scholar Award
2015   MIKI Conference Poster Presentation Award
2014   T32 Training Grant Recipient
2013   Alma College Honors in Biology
2013   Alma College Honors Day Presenter

Professional Memberships:
2015 - Present   Member, American Society of Pharmacognosy
2015 - Present   Member, Rho Chi Society – International Pharmaceutical
                              Sciences Honorary
2012 – Present   Member, Beta Beta Beta – National Biology Honorary

Publications:

May, D. S.; Chen, W.; Lantvit, D. D.; Zhang, X.; Krunic, A.; Burdette, J. E.; Eustaquio, A.; Orjala, J. Merocyclophanes C and D from the Cultured Freshwater Cyanobacterium *Nostoc* sp. (UIC 10110), *J. Nat. Prod.* **2017**, *80* (4), 1073–1080.

May, D. S., Kang H.S., Santarsiero B.D., Krunic A., Shen Q., Burdette J.E., Swanson S.M., Orjala J. Ribocyclophanes A-E, Glycosylated Cyclophanes with Antiproliferative Activity from Two Cultured Terrestrial Cyanobacteria, *J. Nat. Prod.* **2018**, *81* (3), 572-578.

Crnkovic, C.M., **May, D.S.** & Orjala, J. The Impact of Culture Conditions on Growth and Metabolic Profiles of Freshwater Cyanobacteria, *J. Appl. Phycol.* **2018**, *30*, (1), 375-384.

May, D.S.; Chlipala, G.E.; Orjala, J. Cyanobacterial Specific Partial 16S Primers for the Identification of Cyanobacteria in Environmental Samples, In Preparation.

May, D.S.; Crnkovic, C.M.; Krunic, A.; Oberlies, N.; Orjala, J. Stable Isotope Labeling and Comparative Metabolomics Facilitates Genome Mining in Freshwater Cyanobacteria, In preparation

Oral Presentations:

Secondary Metabolites from Freshwater Cyanobacteria: Structure Elucidation, Biosynthesis, and Bioactivity of Merocyclophanes C and D, MIKI Medicinal Chemistry Conference, Minneapolis, Minnesota, April 7-9, 2017

A Bioinformatic and Metabolomic Strategy to Discover Natural Products from Cyanobacteria, Alma College Life Sciences Seminar Series, Alma, Michigan November 30, 2017

Poster Presentations:

May, Daniel; Crnkovic, Camila; Krunic, Aleksej; Orjala, Jimmy, A Bioinformatic and Metabolomic Strategy to Discover New Natural Products from Cultured Cyanobacteria, UIC Research Day 2018, Chicago, Illinois, February 9, 2018
        -Poster Award

May, Daniel; Chlipala, George; Crnkovic, Camila; Krunic, Aleksej; Orjala, Jimmy, Probing the Diversity of Environmental Cyanobacteria Using a New 16S rRNA Cyanobacterial Specific Primer Set, ASP Annual Meeting 2017, Portland, Oregon, July 29 – August 2, 2017

May, Daniel; Chlipala, George; Orjala, Jimmy, A new 16S rRNA Cyanobacterial Specific Primer Set to Probe the Diversity of Soil Cyanobacteria, Perlman Symposium 2017, Madison, Wisconsin, March 31, 2017

May, Daniel; Crnkovic, Camila; Sullivan, Peter; Krunic, Aleksej; Orjala, Jimmy, Isolation and Structure Elucidation of Ribocyclophanes D and G from the Cultured Freshwater Cyanobacterium UIC 10366 *Nostoc* sp., Slatkin Symposium 2016, Chicago, Illinois, November 18, 2016

May, Daniel; Crnkovich, Camila; Sullivan, Peter; Chen, Wei-Lun; Burdette, Joanna; Orjala, Jimmy, A Standardized Method for Dereplication of Bioactive Fractions of Freshwater Cyanobacteria, Annual ICSB Conference/Interim ASP 2016, Oxford, Mississippi, April 11-14, 2016

May, Daniel; Crnkovich, Camila; Sullivan, Peter; Chen, Wei-Lun; Burdette, Joanna; Orjala, Jimmy, A Standardized Method for Dereplication of Bioactive Fractions of Freshwater Cyanobacteria, MIKI Medicinal Chemistry Conference 2015, Iowa City, Iowa, April 8-10, 2016

May, Daniel; Luo, Shangwen; Krunic, Aleksej; Chlipala, George; Orjala, Jimmy, Isolation and Structure Elucidation of Merocyclophane C from the Cultured Cyanobacteria *Nostoc sp.* (UIC 10110) ASP Annual Meeting 2015, Copper Mountain, Colorado, July 24-29, 2015

May, Daniel; Luo, Shangwen; Krunic, Aleksej; Chlipala, George; Orjala, Jimmy, Isolation and Structure Elucidation of Merocyclophane C from the Cultured Cyanobacteria *Nostoc sp.* (UIC 10110) MIKI Medicinal Chemistry Conference 2015, Lawrence, Kansas, April 10-12, 2015

-Poster Award

May, Daniel; Luo, Shangwen; Krunic, Aleksej; Chlipala, George; Orjala, Jimmy, Isolation and Structure Elucidation of Merocyclophane C from the Cultured Cyanobacteria *Nostoc sp.* (UIC 10110) Slatkin Symposium 2014, Chicago, Illinois, November 14, 2014

Sullivan, Peter; **May, Daniel**; Crnkovic, Camila; Orjala, Jimmy, A Multifaceted Approach to Improve Upon the Freshwater Cyanobacteria Drug Lead Discovery Pipeline, Experimental Biology Conference 2017, Chicago, Illinois, April 22-26, 2017

Crnkovic Camila M, **May Daniel**, Orjala Jimmy. The Influence Of Phosphate And Nitrate On Growth And Chemical Diversity Produced By The Freshwater Cyanobacterium *Nostoc sp.* (UIC 10110), Annual ICSB/ Interim ASP 2016, Oxford, Mississippi. April 11th-14th, 2016

Chlipala, George; Naqib, Ankur; **May, Daniel**; Maienschein-Cline, Mark; Kanabar, Pinal; Lei, Zhengdeng; Hu, Vincent; Chuckman, Morris; Orjala, Jimmy; Green, Stefan; Neil, Bahroos, Improving Taxonomic Annotation of 16S Amplicon Sequencing Data from Environmental Samples, Great Lakes Bioinformatics/Canadian Computational Biology Conference 2016, Toronto, Canada, May 16-19, 2016

Sullivan, Peter**; May, Daniel**; Chen, Wei-Lun; Burdette, Joanna; Orjala, Jimmy, The Identification of Metabolites Active Against OVCAR3 Through Dereplication of Cyanobacteria UIC Strain 10480, College of Pharmacy Research Day 2016, Chicago, Illinois, February 26, 2016

University Service:

2015 – 2018	College of Pharmacy Graduate Student Committee member
2017 – 2018	Your Future in Science Seminar Organizing Committee member
2016 – 2017	Webster-Sibilsky Lecture Planning Committee member