UNIVERSITY OF CALIFORNIA, SAN DIEGO

Orthogonality in Natural Products Workflows

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

in

Marine Biology

by

Paul Davis Boudreau

Committee in charge:

Professor William H. Gerwick, Chair
Professor Lihini Aluwihare
Professor Pieter C. Dorrestein
Professor William Fenical
Professor Amro Hamdoun
Professor Bradley Moore

2015

The dissertation of Paul Davis Boudreau is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

No one finishes a Ph. D. alone, I am no exception to this rule. The long list of people to thank begins before I even started this endeavor, with my family who have always supported me, thanks Mom, Dad, and Eleanor. At MIT, I wouldn't have graduated without Zak Fallows and Jessica McKellar, my undergraduate lab partners; or First East, my residence hall on East Campus, who are all, simply put, awesome. Professor Rick Danheiser gave me my first opportunity to join a research project, which certainly led to me pursuing a Ph. D. I also thank the graduate students in the Danheiser lab, Xiao-Yin Mak is a tremendous scientist and incredible mentor who I learned a great deal from; Cindy Crosswhite, Shaun Fontaine, and Julia Robinson put up with me for so long, it's hard to believe their patience. Professor Bill Gerwick allowed me to earn my Ph. D. in his lab, and I must thank the other students who entered the lab with me, Sam Mascuch, Emily Mevers, and Emily Trentacoste. They are a great cohort to work with but also the best people I could have asked to slog through grad-school with. Bailey Miller is a great collaborator and comrade in arms. Other friends, mentors, classmates, and troublemakers to thank include Marcy Balunas, Matt Bertin, Kerri Dawn, Ed Esquenazi, Alan Foreman, Bill Jones, Cliff Kapono, Don Nguyen, Laura Sanchez, Mindi Summers, Kevin Tidgewell, and many others. I got where I am because of you all, your are my heroes.

EPIGRAPH

What do you mean by "Accept misfortune as the human condition"?

Misfortune comes from having a body.

Without a body, how could there be misfortune?

—Lao Tzu (Translation by Gia-fu Feng and Jane English)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in the Journal of Natural Products, Boudreau, Paul D.; Byrum, Tara; Liu, Wei-Ting; Dorrestein, Pieter C.; Gerwick, William H. **2012**, volume *75*, pages 1560-1570.

Chapter 3, in part, is a reprint of the material as it appears in the Journal of Natural Products, Yang, Jane Y.; Sanchez, Laura M.; Rath, Christopher M.; Liu, Xueting; Boudreau, Paul D.; Bruns, Nicole; Glukhov, Evgenia; Wodtke, Anne; de Felicio, Rafael; Fenner, Amanda; Wong, Weng Ruh; Linington, Roger G.; Zhang, Lixin; Debonsi, Hosana M.; Gerwick, William H.; Dorrestein, Pieter C. **2013**, volume *76*, pages 1686-1699.

Chapter 4, in full, is a reprint of the material as it appears in the Public Library of Science Online Edition, Boudreau, Paul D.; Monroe, Emily A.; Mehrotra, Suneet; Desfor, Shane; Korobeynikov, Anton; Sherman, David H.; Murray, Thomas F.; Gerwick, Lena; Dorrestein, Pieter C.; Gerwick, William H. **2015**, volume *10*, page e0133297.

Chapter 5, in part, is a reprint with permission from Phytochemistry Reviews Gerwick, L.; Boudreau, P.; Choi, H.; Mascuch, S.; Villa, F. A.; Balunas, M. J.; Malloy, K. and Teasdale, M. E.; Rowley, D. C.; Gerwick, W. H. **2013**, volume *12*, pages 459-465. It is aslo, in part, currently being prepared for submission for publication of the material. Boudreau, Paul D.; Miller, Bailey; Gerwick, William H. The dissertation author is the primary investigator and author of this material.

Chapter 6, in part, is currently being prepared for submission for publication of the material. Boudreau, Paul D.; Remmel, Ariana; Glukhov, Evgenia; Dorrestein, Pieter C.; Gerwick, William H. The dissertation author is the primary investigator and author of this material.

VITA

2009	B. S. in Chemistry, Minor in Biology, Massachusetts Institute of Technology, Cambridge, MA
2009-2015	Graduate Student in Marine Biology, Gerwick Lab, Scripps Institution of Oceanography, La Jolla, CA
2015	Ph. D. in Marine Biology, Scripps Institution of Oceanography, La Jolla, CA

PUBLICATIONS

Boudreau, P. D.; Byrum, T.; Liu, W. T.; Dorrestein, P. C.; Gerwick, W. H., "Viequeamide A, a Cytotoxic Memeber of the Kulolide Superfmaily of Cyclic Depsipeptides from a Marin Button Cyanobacterium" *J. Nat. Prod.*, **2012**, *75*, 1560-1570.

Gerwick, Lena; Boudreau, P.; Choi, Hyukjae; Mascuch, Samantha; Villa, Francisco A.; Balunas, Marcy J.; Malloy, Karla; Teasdale, Margaret E.; Rowley, David C.; Gerwick, William H. "Interkingdom signaling by structurally related cyanobacterial and algal secondary metabolites" *Phytochem. Rev.*, **2013**, *12*, 459-465.

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., "Molecular Networking as a Dereplication Strategy" *J. Nat. Prod.*, **2013**, *76*, 1686-1689.

Liu, W. T.; Lamsa, A.; Wong, W. R.; Boudreau, P. D.; Kersten, R.; Peng, Y.; Moree, W. J.; Duggan, B. M.; Moore, B. S.; Gerwick, W. H.; Linington, R. G. Pogliano, K; Dorrestein, P. C., "MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*" *J. Antibiot.*, **2014**, *67*, 99-104.

Mevers, E.; Haeckl, F. P. J.; Boudreau, P. D.; Byrum, T.; Dorrestein, P. C.; Valeriote, F. A.; Gerwick, W. H., "Lipopeptides from the Tropical Marine Cyanobacterium *Symploca* sp." *J. Nat. Prod.*, **2014**, *77*, 969-975.

Boudreau, P. D.; Monroe, E. A.; Mehrotra, S.; Desfor, S; Korobeynikov, A.; Sherman, D.H.; Murray, T. F.; Gerwick, L.; Dorrestein, P. C.; Gerwick, W. H., "Expanding the Described Metabolome of the Marine Cyanobacterium Moorea producens JHB through Orthogonal Natural Products Workflows" *PLoS ONE*, **2015**, *10*, e0133297.

Shao, C. L.; Linington, R. G.; Balunas, M. J.; Centeno, A.; Boudreau, P.; Zhang, C.; Engene, N.; Spadafora, C.; Mutka, T. S.; Kyle, D. E.; Gerwick, L.; Wang, C. Y.; Gerwick, W. H., "Bastimolide A, a Potent Antimalarial Polyhydroxy Macrolide from the Marine Cyanobacterium *Okeania hirsuta*" *Journal of Organic Chemistry*, **2015**, *80*, 7849-7855.

Willumstad, T. P.; Boudreau, P. D.; Danheiser, R. L., "Synthesis of Highly Substituted Quinolines via a Tandem Ynamide Benzannulation/ Iodocyclization Strategy" *Journal of Organic Chemistry*, **2015**, *ASAP*.

ABSTRACT OF THE DISSERTATION

Orthogonality in Natural Products Workflows

by

Paul Davis Boudreau

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2015

Professor William H. Gerwick, Chair

In 1969 Weinheimer and Spraggins reported the isolation of prostaglandins from the gorgonian coral *Plexaura homomalla* [Weinheimer and Spraggins, 1969]. This was a remarkable discovery for being one of the first reports of a marine natural product and for its significance to human health, where prostaglandins were being studied for their obstetric effects. It foreshadowed a new pipeline of drug discovery from the oceans; as well as highlighting many difficulties of drug discovery in general, and specific problems of working with marine samples. The isolation of a compound of interest from *P. homomalla* demanded that questions be asked about sustainability and ecology; were there other sources of the prostaglandins to be discovered, could this species be cultured or harvested sustainably?

Further work on *P. homomalla* revealed a minor analog, 5-*trans*-15(S)-PGA_2, demonstrating one of the most persistent problems in natural products chemistry, the presence of compounds that are difficult or impossible to detect at the low concentration which they are found in nature. Decades later, with thousands of marine natural products reported from sources across the tree of life, many of these problems still exist. New analytical tools have been developed, and old ones have become far more sensitive, but this has not erased these difficulties. In fact they have often simply shown that the unexpected diversity of biologically active compounds produced below the previous limit of detection. Every tool and technique developed, or improved, to survey the metabolome offers new data, but none provides a complete picture. To be successful a natural products workflow must incorporate the concept of 'orthogonality', application of multiple overlapping and complementing techniques to provide a more comprehensive survey of the target metabolome. The work in this dissertation exemplifies this concept, where the study of various cyanobacterial strains began with one technique, such as NMR structure guided isolation, which led to another, such as mass spectrometry fragmentation analysis, when new tools became necessary to describe an organism's metabolic capacity. This work has revealed novel compounds, shown unprecedented diversity in known families of compounds, and improved our understanding of the mechanism of action of an active compound.

Chapter 1

Marine Natural Products Chemistry, a Grand Balancing Act

1.1 Abstract

The first natural product chemists to investigate the marine world started their work at a time when the diversity of life in the oceans was poorly described, the biosynthetic pathways available to organisms were little understood or unknown, and what was then state of the art in instrumentation compares feebly with current equipment. Yet these early investigators had great success, which is not surprising knowing now that marine environments are prolific sources of a broad and diverse pharmacopeia of natural products that have equally diverse activities and potential applications. From those early efforts, often driven by feature-guided (bioassay-guided or structure-guided) isolation, have come thousands of compounds, and several clinically approved drugs. More researchers working in this field has increased its productivity but also raised the problem of compound rediscovery, leading many researchers to explore new environments, poorly studied taxa, or shift their focus to alternative workflows such as comparative metabolomics and genomics. In all cases, natural products discovery efforts have had to balance between the benefits of discovering

novel compounds, an improved understanding of chemical ecology, the cost of doing such work, and the challenges presented by isolation and structure elucidation; the grand balancing act of these factors defines the utility of natural products chemistry and outlines the directions it will take in the future.

1.2 Introduction

The traditional workflow of the Gerwick lab would be recognizable to the researchers who conducted the first isolations from the marine realm. As with the isolations of the prostaglandins from soft corals [Weinheimer and Spraggins, 1969], or the first sponge isolations of bromo-cyclohexenediones [Sharma and Burkholder, 1967b, Sharma and Burkholder, 1967a], our workflow begins with a lipophilic extraction, proceeds to targeted purification of pure compounds, and concludes with structure elucidation. Of course the methods involved in each step have improved or changed; dendrolasin (**1**), for example, was found following a scent based isolation targeting the characteristic smell of terpenes [Vanderah and Schmitz, 1975]. Presently we use generally more reliable (and safer!) methods for targeting compounds, but the overall workflow falls well within the same feature-guided isolation workflow.

Each step within the workflow, however, has been changed by the development of new techniques and workflows. One of the first isolations of a natural product from the oceans was 2,6-dibromo-4-acetamido-4-hydroxy-cyclohexadienone (**2**), a compound found to have broad spectrum antibiotic activity, from the sponge *Verongia cauliformis* [Sharma and Burkholder, 1967b, Sharma and Burkholder, 1967a]. A compound with a low hydrogen to carbon ratio, it would have been difficult to determine the structure of **2** with 2D-NMR spectroscopy, even if its isolation had not occurred almost a decade before the first 2D-NMR experiments were conducted [Aue et al., 1976]. Instead, a key component of the structure elucidation was chemical degradation and total synthesis. This workflow is a difficult and

unenviable one, as highlighted by the synthesis of bryostatin 1 (**3**), which was 30 steps in its longest linear sequence [Keck et al., 2011]; in general, it is not be feasible to attempt the synthesis of the most complex and exciting compounds isolated from nature as a structure elucidation tool.

Nor is the scale of previous isolations still necessary, collecting high quality NMR spectra for nitenin (**4**) and dihydronitenin (**5**) on a 100 MHz instrument was only possible because they were isolated on the mole-decamole scale [Fattorusso et al., 1971]; this can be juxtaposed with the isolation and structure elucidation of the sanguinamide A (**6**) and B (**7**) on the nanomole scale, which exemplifies the power of two-dimensional experiments, high-field magnets, and cryo-probes in NMR spectroscopy in structure elucidation [Dalisay et al., 2009, Molinski, 2010].

More than just improving on this workflow, marine natural products chemists have developed entirely novel workflows. When the genome of the bacterium *Haemophilus influenzae* Rd was sequenced in 1995 it was the first genome of a free-living organism [Fleischmann et al., 1995]; since then hundreds of bacterial genomes have been sequenced, the cost of sequencing has fallen, and the tools for genome annotation have improved [Medema et al., 2011, Blin et al., 2013, Gross et al., 2007, Ziemert et al., 2012]. The most striking result of this work is that the genomes of previously studied organism show a multitude of 'cryptic' or 'silent' biosynthetic pathways, for which a genetic sequence is available but not linked to a natural product [Challis, 2008]. It has also identified a new workflow for the discovery of natural products from bacteria. Instead of targeting a feature of a compound, such as an activity in a bioassay or the presence of a functional group as determined by NMR or MS, now biosynthetic pathways can be targeted that either might be expected to produce a compound with a desired bioactivity or structural feature by bioinformatics analysis.

The changes in the field of natural products isolation have allowed a sea change in

the broader natural products field. In years past the large diversity of chemistry produced by oceanic organisms was unknown, and as such the evolutionary reasons for the production of these compounds was unimagined. The ability to rapidly identify the taxonomic identity of organisms and elucidate the secondary metabolites produced by them has opened the field of marine chemical ecology, where researchers can begin to parse how chemical cues lead to higher level biological and ecological function [Hay, 2009].

1.3 Discussion

The often used canonical definition of a natural product, or secondary metabolite, is a compound produced apart from an organism's need for growth, energy metabolism, or reproduction. This definition explains the central component of natural products chemistry, that these compounds fulfill specialized needs such as communication and signaling, but misses the fact that in some cases biologically relevant products, or isolated compounds of interest for pharmacological properties, may in fact not be produced by the organism, they may be degradation products of the 'natural' species, or explicitly required by the organism to function and grow in its environment.

The didemnin story is interesting because it highlights some of these issues. Originally isolated from a Caribbean tunicate, the didemnins were shown to have activity in several biological assays [Rinehart et al., 1981b, Rinehart et al., 1981a], with didemnin B (**8**) being the first marine natural product to enter clinical trials [Lee et al., 2012]. Since its progression into clinical trials however, Xu and coworkers showed that didemnins are produced by *Tistrella* spp., a free living bacterial species [Xu et al., 2012]. In addition, using matrix assisted laser desorption ionization imaging mass spectrometry (MALDI-IMS) and other techniques, they were able to show that the chemical species originally produced by the bacteria are didemnin X (**9**) and Y (**10**) and that didemnin B is the product of an extracellular hydrolytic enzyme [Xu et al., 2012].

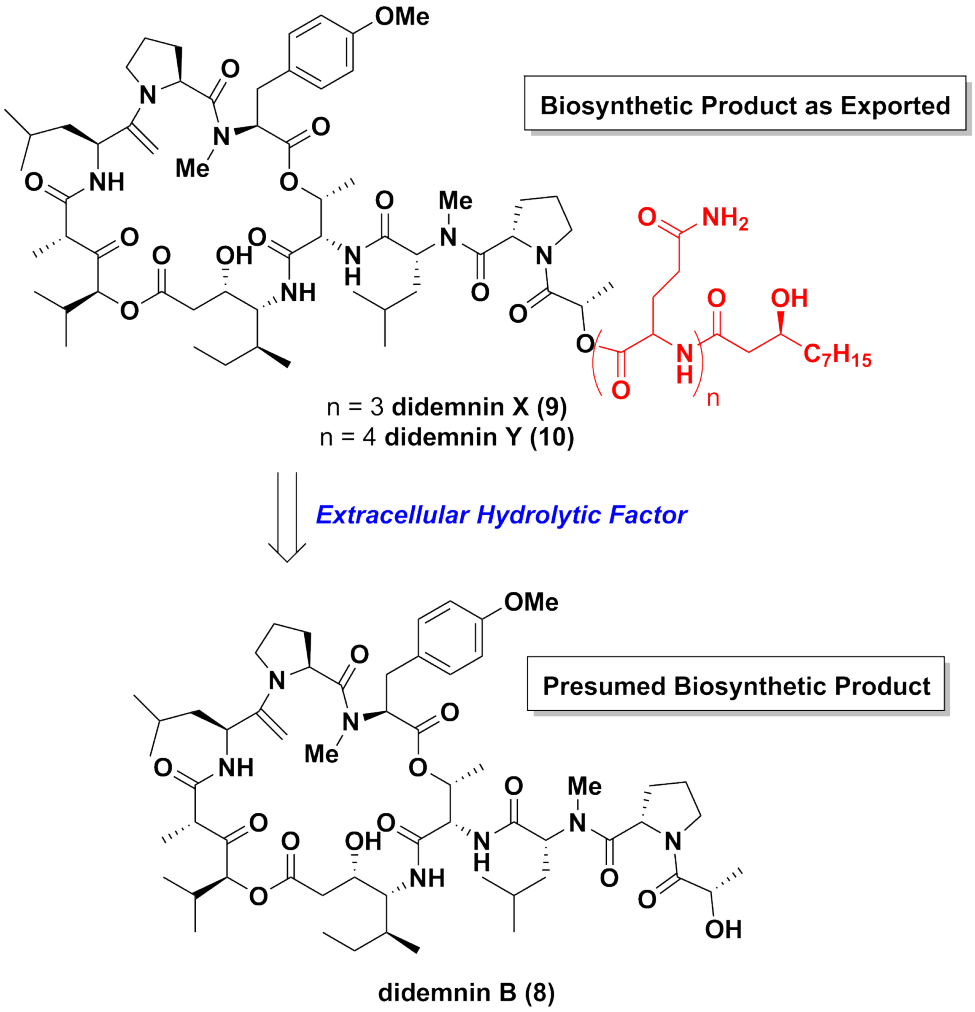


Figure 1.1: *Structures of the didemnins (**8-10**).* The first products discovered in this family were not the "natural" species, instead, they were derivatives formed after activation by an excreted hydrolytic factor.

The shift in secondary metabolite production regimes reported by Gallagher in the *Streptomyces* strain CNQ-525 under increasing oxygen stress was the first report of ecological plasticity for a natural product pathway [Gallagher, 2015]. Under normal growing conditions the biosynthetic pathway responsible for the production of napyradiomycin (**11**) functions normally, but under microaerophilic conditions the productions of **11** drops off and 8-amino-flavin (**12**) production increases. First isolated as a presumed biosynthetic intermediate of furaquinocin (**13**) [Isogai et al., 2012], also, as with **11**, a hybrid polyketide-isoprenoid natural product, this natural product is predicted to be an intermediate in the

synthesis of **11** as well. The loss of production of a secondary metabolite under stress to the organism, however, is not new or unprecedented. What is remarkable in this case is that it appears that **12**, which is redox active, is an endogenous electron shuttle used to reduce extracellular electron acceptors as molecular oxygen is diminished. Further work to validate the entire electron shuttle cascade is needed, but if firmly established, it would demonstrate the ability of an organism to regulate from a single biosynthetic pathway to different ecological roles [Gallagher, 2015]. **11** has classical secondary metabolite properties, mainly its potent antibacterial activity [Shiomi et al., 1986], which is markedly different from **12**, which is not bioactive and appears to be involved in energy metabolism, well within the bounds of primary not secondary metabolism.

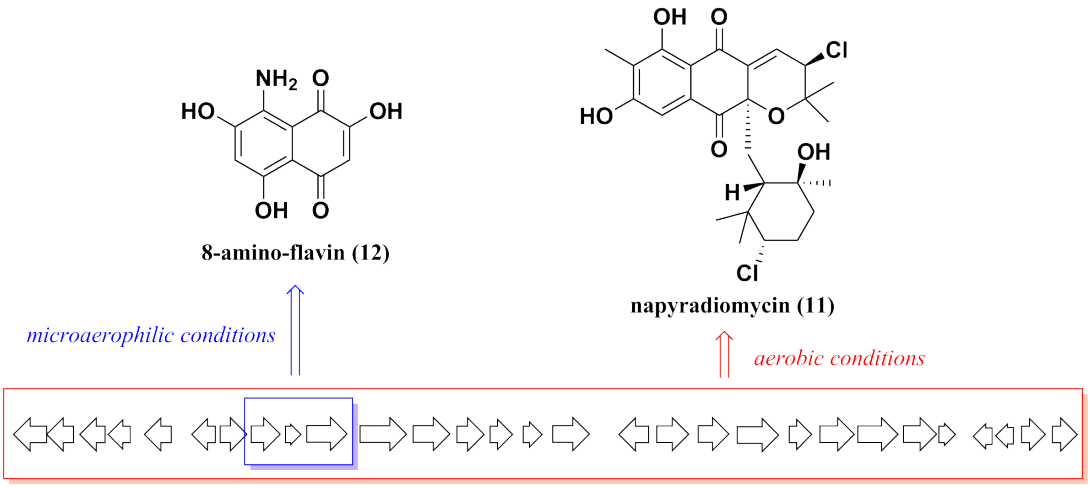


Figure 1.2: *Structure of 8-amino-flavin (**12**) and napyradiomycin (**11**).* Under microaerophilic condition the production the full biosynthetic product, napyradiomycin (**11**), is diminished and the genes *napB1*, *napB2*, and *napB3*, highlighted in blue, produce a significant increase in the intermediate compound 8-amino-flavin (**12**).

In the literature, reports of new natural products are often made without this ecological context. Presented as the products of an organism as found by some isolation workflow, or shown to be active in a biological assay (one usually largely unrelated to any environmental or ecological context) rather than provided with an explanation of their role for the organism in its environment. This is not an ideal situation. To work towards a productive

workflow that can both provide a pipeline for new natural products of pharmacologically novel structure *as well as* elucidating biological and ecological roles of these compounds, then the different isolation workflows must be compared to determine their most successful features and strategies that explicitly target a wider understanding of a compounds role in the environment must be developed.

1.3.1 Feature-Guided Isolation

Potency in biological assays is a feature that has been used to isolate active compounds from remarkably dilute sources. In the case of dolastatin 10 (**14**) in the original isolation from the sea hare *Dolabella auricularia* the natural product was present in concentrations of ca. 1 mg per 100 kg of organism [Pettit et al., 1989]. Simply because **14** is very potent, originally reported as single μg/kg activity against several nude mouse cell line xenographs and more recently in clinical drug trails [Pettit et al., 1989, Pitot et al., 1999, Perez et al., 2005], the compound could be traced through repeated solvent fractionations and chromatography steps to arrive at the active species.

With better performing purification systems, isolated compounds have been successfully purified in truly minute amounts; **6** and **7** were isolated from a single specimen of *H. sanguineus* in μg quantities [Dalisay et al., 2009]. The productivity of feature-guided isolation in the twenty years between these two studies has faced hurdles from diminishing interest from pharmaceutical companies to the increasing incidence of reisolation of known compounds from the drying up of the 'low hanging fruit' and a shifting interest to combinatorial chemistry [Koehn and Carter, 2005].

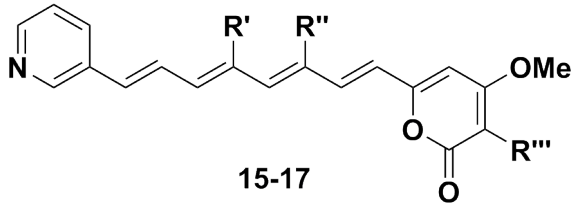
Feature-guided isolation, however, is still incredibly productive, year after year hundreds of isolations are reported in the literature from following a bioassay or structure guided isolation [Blunt et al., 2013, Blunt et al., 2014].

1.3.2 Comparative Metabolomics

To keep this pipeline productive new approaches are being developed that use alternative means to assess targets for isolation. Principle component analysis (PCA), and MALDI-IMS aim to provide insight into metabolites of interest not by targeting specific features, but instead by comparing metabolites against each other to determine their uniqueness.

Bungi and coworkers have used LCMS-based PCA to prioritize strains for further work drug discovery efforts [Hou et al., 2012]. The goal being that a comparison of PCA scores between strains could reveal strains with the least redundancy, and within those strains rapidly providing a list of targets for drug discovery that are distinct. Those distinct metabolites, responsible for making the metabolome different from closely related strains, can be expected to not be compounds common to all the strains or compounds known to be produced other the other strains.

For example, the profile of strain WMMB-272 was found to be different from other marine derived *Streptomyces* spp. because of the production of several polyenepyrones (**15-17**)containing a rare pyridine functional group. Showing that this workflow can provide access to distinct metabolites, even within a highly studied genus like *Streptomyces*.



15-17

Figure 1.3: *Structures of the polyenepyrones (**15-17**) targeted by principal component analysis.* For **15** R' = Me R'' = H R''' = Me, for **16** R' = H R'' = Me R''' = Me, and for **17** R' = H R'' = Me R''' = H.

Another novel approach to profiling metabolites is MALDI-MS imaging of organisms capable of growing on an agar surface, which is an exciting way to profile the interactions between organisms. By culturing on thin agar plates, the interactions between

multiple streaked organisms can be imaged in the MALDI spectrometer which not only shows the MS spectra of metabolites produced by an organism but the localization of these metabolites. Mainly, there are the metabolites excreted into the media, present in the culture itself, or localized to a specific zone of interaction. Meaning that when grown against a target strain, such as a pathogenic species, instead of profiling a fractionated extract for inhibitory activity, the metabolites present within a zone of inhibition can become targets for an isolation workflow [Yang et al., 2012].

1.3.3 Genomics

Since the development of next generation sequencers, the cost of gene sequencing has fallen dramatically, far faster than when compared to "Moore's Law", in computer science the rough doubling every year in the number of transistors per microchip and the halving of the cost per individual microchip [Hayden, 2014].

As with the proliferation of applications for computer power, gene sequencing has led to new techniques, new instruments, and new discoveries in genetics. Within the field of natural products one of the most important of these is 'genome mining' the ability to profile an entire organism's genome and by compare the biosynthetic pathways to known pathways allows the prediction of a natural product structure before any material has even been isolation. More importantly the widespread sequencing of productive bacteria has led to the discovery that productive strains contain the genetic potential to produce unobserved metabolites, these 'cryptic' or 'silent' pathways are a source of molecular diversity that was entirely unknown before the advent of genome sequencing [Challis, 2008].

To isolate the compounds putatively assumed to be produced by these pathways, many strategies have been developed, from predicting a property of the compound and analyzing the extract for those properties, gene modification to observe a gain or loss of production, or heterologous expression of the pathway in a different organism [Challis,

2008].

There are a great many examples of this work revealing new molecular diversity. Nygerones A (**18**) and B (**19**) were isolated from *Aspergillus niger* by epi-genetic tailoring using suberoylanilide hydroxamic acid (**20**) as a histone deacetylase inhibitor to modify the fungus' gene expression. Under these conditions the culture, when extracted, showed novel metabolites from a previously silent pathway [Henrikson et al., 2009].

Using ^{15}N-labeled leucine Gross and coworkers discovered the orfamides A-C (21-23) from an orphan pathway in *Pseudomonas fluorescens* Pf-5. The pathway responsible for the biosynthesis could be predicted to incorporate certain amino acids based on the NRPS genes involved. By feeding an amino acid with an ^{15}N-label it was possible to trace the resulting product through ^{15}N-NMR during the fractionation process [Gross et al., 2007].

To use predictions about a compound from a biosynthetic pathway can also take the form of predicting a halogenation pattern or other structural features to guide the isolation. By predicting that a product produced by *Streptomyces coelicolor* M145 would likely be a siderophore, Challis and coworkers could guide their isolation to the compound coelichelin (**24**) [Challis and Ravel, 2000, Lautru et al., 2005].

1.3.4 Chemical Ecology

To gain insight into marine realms, terrestrial analogies are often used [Pinet, 1992]. The same is true of the organization of trophic levels where terrestrial examples are used in comparing relationships between primary producers, herbivores, and predators in the oceans. Too often these descriptions are quite lacking, and fail to describe the complexities unique to the marine environment. The mid-ocean gyres are often compared to terrestrial deserts because the have low nutrient availability and low rates of primary productivity. A challenge to this comes from the observation that nutrient utilization rates are much higher in the gyre, where primary producers rapidly scavenge and consume available nutrients

[Garrison, 1999].

In exploring the connections between trophic levels in the marine realm, researchers have found surprising and unique relationships between primary producers and their consumers. The Hay lab described a system where the phytoplankton *Phaeocystis globosa* uses chemical cues from its predators to shift between morphologies to better protect itself from predation [Long et al., 2007]. *P. globsa* can grow as either solitary cells, which are preferred by ciliate predators but harder for copepods to consume, or a colonial form, that is preferred by copepods but disfavored by ciliates. A most exciting aspect of this research was that the switch between morphologies was controlled by chemical cues from the predators [Long et al., 2007]. Great insight was derived from understanding of the role of small molecules in signaling and communication in the sea, and this did not derive from insights gained from terrestrial ecosystems.

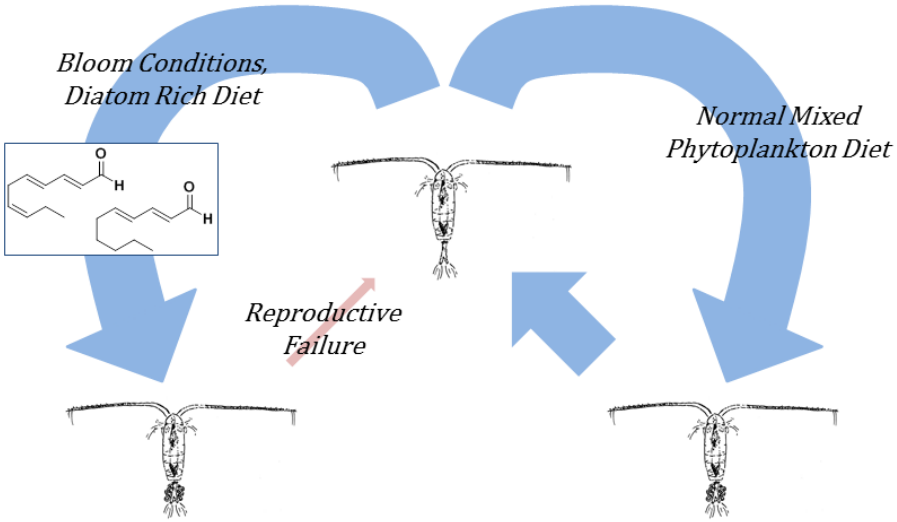


Figure 1.4: *Reporductive failure in the copepod life-cycle.* During bloom conditions a diet dominated by diatoms leads not to a toxicity effect (a loss of adult copepods which fed on the diatoms), but instead leads to a reduced viability of copepod eggs and an overall loss of fecundity.

Another study into the chemical ecology of phytoplankton and copepods found that the extracts of, or a diet including, certain diatom species acts to inhibit egg hatching in their copepod predators [Ianora et al., 1999, Miralto et al., 1999, Starr et al., 1999]. The diatom

species *Thalassiosira rotula*, *Skeletonema costatum*, and *Pseudo-nitzschia delicatissima* were shown to produce small unsaturated aldehydes which reduce egg viability in the copepods *Acartia clausi* and *Calanus helgolandicas* [Miralto et al., 1999]. This has been shown in other diatom-copepod systems, and it was found that the effect was seen to be both species dependent and proportional to consumption of the diatom prey [Starr et al., 1999]. The effect has also been shown to be robust in the laboratory and not related to anoxia or that the diatom-only diet lacked a required micro-nutrient, but rather, due to specific small molecule toxicity effects [Ianora et al., 1999].

1.4 Conclusions

The natural products field has changed dramatically as technological and methodological improvements have made it possible to elucidate the structure of targets that are more complex and/or available on a scale orders of magnitude less than in the first endeavors into the marine natural products field. This change has led to an explosion of thousands of new marine natural products, including several new drugs and drug candidates [Gerwick and Moore, 2012]. It has also caused a profound change in the study of both macro and microbial ecology; there it has be found that the small molecule chemical cues dominate the interactions of complex interspecies relationships.

Chapter 2

The Viequeamides, Cyclic Desipeptides from a Tropical Filamentous Cyanobacterial Strain

2.1 Abstract

The viequeamides, a family of 2,2-dimethyl-3-hydroxy-7-octynoic acid (Dhoya) containing cyclic depsipeptides, were isolated from a shallow subtidal collection of a 'button' cyanobacterium (Rivularia sp.) from near the island of Vieques, Puerto Rico. Planar structures of the two major compounds, viequeamide A (**25**) and viequeamide B (**26**), were elucidated by 2D-NMR spectroscopy and mass spectrometry, whereas absolute configurations were determined by traditional hydrolysis, derivative formation, and chromatography in comparison with standards. In addition, a series of related minor metabolites, viequeamide C-F (**27-30**), were characterized by high resolution mass spectroscopic (HRMS) fragmentation methods. **1** was found to be highly toxic to H460 human lung cancer cells (IC_{50} = 60 $\pm$ 10 nM), whereas the mixture of **26-30** was inactive. From a broader perspective,

the viequeamides help to define a "superfamily" of related cyanobacterial natural products, the first of which to be discovered was 'kulolide' (**31**). Within the kulolide superfamily, a wide variation in biological properties is observed, and the reported producing strains are also highly divergent, giving rise to several intriguing questions about structure-activity relationships and the evolutionary origins of this metabolite class.

2.2 Introduction

Marine cyanobacteria are increasingly being recognized as important sources of biologically active secondary metabolites [Tan, 2007, Tan, 2010, Gerwick and Moore, 2012]. While numerous natural products have been reported from marine invertebrates, subsequent work has shown that in many cases, the true sources of compounds of interest are bacteria [Piel, 2004, Piel, 2009]. Efforts by many research groups to directly focus on marine bacteria and cyanobacteria, have been rewarded by the discovery of an impressive number of highly cytotoxic compounds, such as curacin A (**32**) [Gerwick et al., 1994], largazole (**33**) [Taori et al., 2008], the carmaphycins A (**34**) and B (**35**) [Pereira et al., 2012], and salinosporamide A (**36**) [Feling et al., 2003]. More broadly, the diversity of marine natural products discovered so far, reveals an untapped wealth of structural diversity and activities well worth exploring.

Within marine natural products, depsipeptides produced by a hybrid pathway of nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), have been extensively reported. These biosynthetic pathways, as well as specific modifications, such as *N*-methylation, are characteristic of several groups of filamentous marine cyanobacteria [Tan, 2007, Tan, 2010]. For example since the isolation of **31** in 1993 by Scheuer and coworkers from the cephalaspidean mollusk *Philinopsis speciosa*, [Reese et al., 1996] a variety of structurally similar compounds have been isolated from diverse cyanobacteria. These related compounds contain either a Dhoya or 3-hydroxy-2-methyl-7-octynoic acid

(Hmoya) moiety within an overall cyclic depsipeptide framework, and suggests the existence of a 'kulolide superfamily' within the natural products of these filamentous prokaryotes [Reese et al., 1996, Nakao et al., 1998, Luesch et al., 2001a, Montaser et al., 2011, Medina, 2009, Malloy, 2011, Nogle and Gerwick, 2002, Bunyajetpong et al., 2006, Tripathi et al., 2009, Tripathi et al., 2010, Salvador et al., 2011, Mevers et al., 2011]. The striking structural similarities within this large metabolite class, as well as their diverse reported metabolic sources, suggests that the kulolide superfamily may have an ancient evolutionary origin within the cyanobacteria.

2.3 Results and Discussion

In the current study, an extraction of a marine cyanobacterium collected from the Puerto Rican island of Vieques, showed a ^{1}H NMR spectrum with peaks characteristic of a depsipeptide. Purification of this material afforded a major compound, **25**, that was found to be highly cytotoxic to the H460 cell line. The structure of this compound and a related minor metabolite were determined by 2D-NMR and mass spectroscopic (MS) methods, while the planar structures of a series of very minor metabolites were determined by MS fragmentation methods. Absolute configuration of the two more major compounds, **25** and B **26**, were determined by hydrolysis to fragments which were converted into various derivatives and analyzed by chromatography in comparison with standards.

A 'button' cyanobacterium was collected from the Puerto Rican island of Vieques, preserved in 1:1 isopropyl alcoholseawater, and later extracted for its lipid soluble metabolites. The extract was subjected to a preliminary Vacuum Liquid Chromatography (VLC) fractionation scheme. High Performance Liquid Chromatography (HPLC) purification of the relatively polar ethyl acetate eluted fraction afforded a single major compound, (**25**, 13.6 mg). Additionally, **26-28** and two unresolved compounds **29** and **30**, were obtained as inseparable mixtures, as described below.

HRMS analysis of **25** showed a compound with an $[M + Na]^+$ *m/z* of 826.4940, providing the molecular formula $C_{42}H_{69}N_5O_{10}$ (11 degrees of unsaturation). The ^{13}C NMR spectrum showed seven carbonyl peaks at δ 168 174, thus leaving unaccounted four additional double bonds or rings in the molecule (Table 1).

Table 2.1: *Summary of NMR data (in $CDCl_3$) for viequeamide A (**25**).* [a]The carbon resonances at 18.9, 19.1, 19.4, and 19.6 were very close together and their HSQC associations could not be confidently assigned; HMBC correlations failed to confidently resolve their identities as well. [b]These proton resonances overlapped. [c]Hmpa, 2-hydroxy-3-methyl-pentanoic acid. [d,e]These proton resonances overlapped. *(Continued on the next page.)*

Residue	Position	δ_C, type	δ_H (*J* in Hz)	HMBC	TOCSY	ROESY
Thr	1	169.2, C				
	2	58.1, CH	4.68, dd (10.3, 3.2)	1, 3, 5	Thr-NH, 3	
	3	68.2, CH	3.87, m		Thr-OH, 2, 4	
	4	19[a], CH_3	0.80, d (6.4)	2, 3		
	OH	-	4.48, d (12.4)	3	3, 4	Thr-NH
	NH	-	8.04, d (10.4)	5	2	Thr-OH, 6
N-Me-Val-1	5	168.2, C	-			
	6	67.4, CH	4.06, d (12.4)	5, 7, 10, 11	7, 8, 9	Thr-NH, 12
	7	26.0, CH	2.41, m	6	6, 8, 9	
	8	19[a], CH_3	0.84, d (6.7)	6, 7, 9	6, 7, 9	
	9	20.1, CH_3	0.93-1.00[b]			
	10	29.0, CH_3	2.77, s	6, 11		7
Pro	11	172.1, C	-			
	12	55.7, CH	5.01, dd (8.4, 3.1)		13-15	6
	13a	29.7, CH_2	2.09, m		12, 13b, 14, 15	
	13b	29.7, CH_2	2.03, m		12, 13a, 14, 15	
	14	25.4, CH_2	2.52, m		12, 13a/b, 15b,	
	15a	47.4, CH_2	3.95, m	13, 14	12, 13, 15b	15b, 17
	15b	47.4, CH_2	3.52, m	13, 14	12-15a,	15a, 17
Hmpa[c]	16	168.7, C	-			
	17	74.5, CH	4.82, d (2.0)	*18-21 weak*		15

Table 2.1 (Continued). *Summary of NMR data (in $CDCl_3$) for viequeamide A (**25**).* [a]The carbon resonances at 18.9, 19.1, 19.4, and 19.6 were very close together and their HSQC associations could not be confidently assigned; HMBC correlations failed to confidently resolve their identities as well. [b]These proton resonances overlapped. [c]Hmpa, 2-hydroxy-3-methyl-pentanoic acid. [d,e]These proton resonances overlapped. *(Continued from previous page.)*

Residue	Position	δ_C, type	δ_H (J in Hz)	HMBC	TOCSY	ROESY
Hmpa[c]	18	36.6, CH	1.72 m	19, 20, 21		
	19	13.8, CH_3	1.10, d (6.7)	18, 20, 21		
	20	27.3, CH_2	1.43-1.55[d]	18, 20, 21		
	21	12.0, CH_3	0.93-1.00[b]			
N-Me-Val-2	22	170.3, C	-			
	23	63.7, CH	3.91, d (10.7)	22, 24, 27, 28	24-26	29
	24	29.5, CH	2.28e		23, 25, 26	
	25	19[a], CH_3	0.93-1.00[b]			
	26	19[a], CH_3	0.93-1.00[b]			
	27	28.3, CH_3	2.98, s	23, 28		
Val	28	172.8, C	-			
	29	53.9, CH	4.89, dd (7.3, 2.3)	28, 30, 33	Val-NH	23
	30	31.7, CH	1.96, m		31, 32	
	31	20.7, CH_3	0.93-1.00[b]			
	32	16.1, CH_3	0.75, d (6.6)	29-31	30, 31	
	NH	-	6.83, d (7.4)	33	29	35
Dhoya	33	174.7, C	-			
	34	46.7, C				
	35	17.1, CH_3	1.36, s	33, 34, 36, 37		Val-NH
	36	25.6, CH_3	1.18, s	33-35, 37		
	37	77.4, CH	5.55, d (8.5)	1, 33, 34	38, 39, 40	
	38a	27.9, CH_2	1.76, m			
	38b	27.9, CH_2	1.43-1.55[d]			
	39	24.8, CH_2	1.43-1.55[d]	41		
	40	18.1, CH_2	2.19-2.32[e]	41, 42	37-39, 42	
	41	84.2, C	-			
	42	68.9, CH	1.93, t (2.6)		38, 39, 40	

Total Correlation Spectroscopy (TOCSY) experiments revealed eight proton spin systems. Four of these were for easily recognized amino acid residues, namely threonine, proline (accounting for one more degree of unsaturation), and two *N*-methyl valine residues.

Two of the remaining TOCSY-defined spin systems were connected by Correlation Spectroscopy (COSY), suggesting constrained rotation might be preventing strong TOCSY correlations within this valine residue. Furthermore, a combination of Heteronuclear Multiple Bond Correlation (HMBC) and Heteronuclear Single Quantum Correlation (HSQC) experiments confirmed that these two spin systems belonged to a valine residue wherein a very small ^{1}H-^{1}H coupling was present between the α and β-methine protons (J = 2.3). A third unassigned and very weak TOCSY spin system was better revealed by COSY correlations, and involved a doublet proton at δ_H 4.82 that showed a small ^{1}H-^{1}H coupling constant (J = 2.0) to an adjacent methine at δ_H 1.72 (m). The HSQC spectrum indicated that the methine proton at δ_H 4.82 was bonded to a carbon atom at δ_C 74.5, strongly suggesting that this carbon had an attached oxygen atom. From the COSY, HMBC, and HSQC spectra the δ_H 4.82 doublet was shown to be the α methine proton of an Hmpa residue, the hydroxy acid analog of isoleucine. The final TOCSY spin system was characteristic of a PKS derived residue, as it contained a chain of adjacent methylenes (δ_C 27.9, 24.8, 18.1) and an oxygen-bearing methine (δ_C 77.4). HMBC correlations from two singlet methyl groups (δ_H 1.18 and 1.36) to a quaternary carbon (δ_C 46.7), the final unaccounted for carbonyl carbon at δ_C 174.7, and the oxygen bearing carbon atom at δ_C 77.4, defined an α,α-dimethyl-β-hydroxy-carbonyl system. Additionally, the distal methylene group of the methylene chain (δ_H 2.19-2.32) displayed HMBC correlations to carbons at δ_C 84.2 and 68.9, revealing that this residue terminated with an alkyne group, and therefore accounted for the 9^{th} and 10^{th} degrees of unsaturation present in viequeamide A (**1**). With consideration of the molecular formula, this latter residue was defined as a derivative of Dhoya. The sequence of residues in **25** was elucidated by HMBC and Rotating-frame Overhauser Effect Spectroscopy (ROESY) experiments. Beginning with the Dhoya residue, both the valine α-methine proton and nitrogen bearing proton showed HMBC correlations to the Dhoya carbonyl carbon at δ_C 174.7, showing that an amide connected these two residues. In turn, the methyl group of *N*-methyl

valine-1 showed HMBC correlations to the valine carbonyl carbon at δ_C 172.8, showing an *N*-methyl amide group connected these two residues. The α-methine proton in Hmpa at δ_H 4.82 showed HMBC correlation to the carbonyl carbon of *N*-methyl valine-1 at δ_C 170.3, revealing that there was an ester linkage between these two residues. It was noted that the α-methine proton of Hmpa at δ_H 4.82 showed no HMBC correlations across its carbonyl carbon to the next residue in the sequence. By ROSEY, however, a correlation was observed between the α-methine proton of Hmpa and the δ-protons of the proline residue. We have previously observed in this class of compound that HMBC correlations to a proline residue are not easily observed [Mevers et al., 2011]; however, this ROESY correlation provided the amide connection of Hmpa to proline. To the other side of the proline, the *N*-methyl protons of *N*-methyl valine-2 showed strong HMBC correlation to the proline carbonyl carbon at δ_C 172.1, indicating the linkage of these two residues through an *N*-methyl amide. Both the amide nitrogen proton and the α-methine proton of threonine showed HMBC correlations to the carbonyl carbon of *N*-methyl valine-2 at δ_C 168.2, showing these residues were linked by an amide bond. Completing the residue sequence, and accounting for the final 11^{th} degree of unsaturation, the oxygen-bearing β-methine proton of the Dhoya residue showed HMBC correlation to the carbonyl of threonine at δ_C 169.2, indicating the linkage of these two residues through an ester, thus completing an overall macrocyclic structure for **25**. Summarizing, a combination of HMBC and ROESY correlation data allowed full definition of the residue sequence of **25** as cyclo-[DhoyaVal*N*-Me-Val-1HmpaPro*N*-Me-Val-2Thr] (Table 1, Figure 1).

The configurations of the amino acids in **25** were all determined to be L by hydrolysis and Marfey's analysis [Marfey, 1984, Bhushan and Brückner, 2004], whereas the configuration of the Hmpa residue was established to be 2*R*,3*S* by chiral gas-column mass spectrometry (GCMS) in comparison to synthetic standards [Mevers et al., 2011, Mamer, 2000]. The configuration of the Dhoya residue was determined by hydrogenation of **25** using

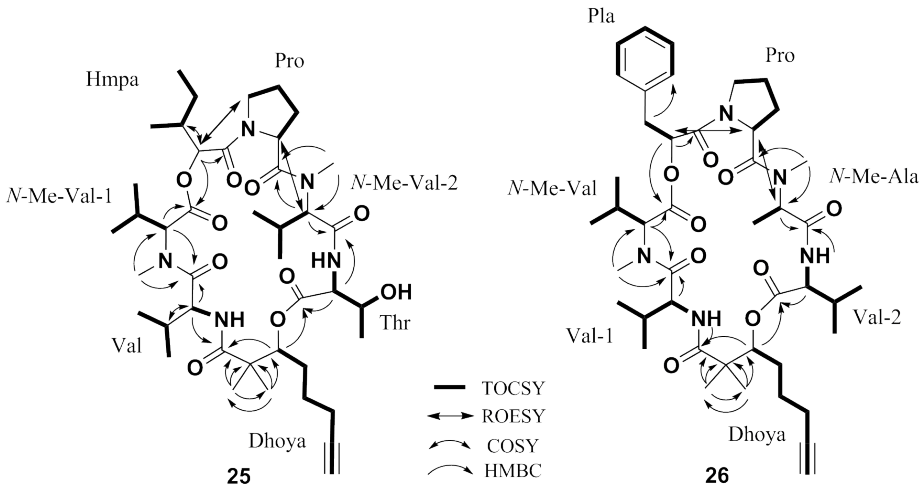


Figure 2.1: *Key 2D-NMR correlations of viequeamide A (**25**) and viequeamide B (**26**).*

hydrogen gas over palladium on carbon and then hydrolysis to release the 2,2-dimethyl-3-hydroxyoctanoic acid (Dhoaa) unit; this was compared by GCMS to standards previously prepared [Nunnery et al., 2011], thus establishing the *S* configuration and complete stereostructure of **25**.

Several analogs of **25**, namely **26-30**, were also isolated from this extract. Unfortunately, these latter derivatives were inseparable by HPLC possibly because of the presence of equilibrating rotomers of each resulting from multiple *N*-methyl amide functional groups as well as their very similar structures. While the planar structure of the major compound in the mixture, **26**, could be determined by 2D-NMR as described below, the minor constituents could only be characterized by the use of HRMS fragmentation experiments.

By inspection of the ^{1}H and ^{13}C NMR spectra for **26** in comparison with that of **25**, these metabolites were clearly very closely related with both possessing two *N*-methyl amide peaks, similar doublet methyl groups indicative of valine, leucine, or isoleucine residues, and the alkyl methylene and singlet methyl protons indicative of a Dhoya residue. The most striking difference between the two was the presence of aryl protons in **26** which were absent in **25**. By the interplay of TOCSY, COSY, HSQC, and HMBC experiments, the residues of **26** were elucidated as Dhoya, proline, valine, *N*-methyl valine, and two

residues not observed in viequeamide A, phenyl lactic acid (Pla) and *N*-methyl alanine. The sequence of residues of **26** was deduced from ROESY and HMBC correlations as cyclo-[Dhoya-Val-1-*N*-Me-Val-Pla-Pro-*N*-Me-Ala-Val-2] (Table 2, Figure 1).

Stereoanalysis of this **26** dominated mixture of compounds **26-30**, was undertaken with the assumption that the minor natural products **27-30** would contribute negligibly to the analysis (see analysis of molecular ions below for relative concentrations). Thus, the mixture was first hydrogenated to reduce the unstable acetylene functional group of the Dhoya residue and then subjected to microwave-assisted acid hydrolysis. A portion of the hydrolysate was converted with 1-fluoro-2,4-dinitrophenyl-5-L-valine amide (L-FDVA) to Marfey's-type derivatives and compared by LCMS with authentic amino acids similarly derivatized, revealing only L residues (L-*N*-Me-Ala, L-Val, L-*N*-Me-Val, and L-Pro). A second portion of the hydrolysate was converted to methyl esters with CH_2N_2 and then analyzed under two different GCMS run conditions to identify (*S*)-2,2-dimethyl-3-hydroxyoctanoic acid (*S*-Dhoaa, and by inference *S*-Dhoya) and *S*-phenyl lactic acid, thus completing the stereostructure of **26**.

Liquid Chromatography High Resolution Mass Spectrometry (LC-HRMS) of the mixture of compounds dominated by **26** gave the major molecular ions at *m/z* 830.4666, 832.4833, 834.4988, and 844.4833, present in the approximate ratio of 25:5:1:1, respectively. These corresponded to the $[M + Na]^+$ ions for compounds with the molecular formulas of $C_{44}H_{65}N_5O_9$, $C_{44}H_{67}N_5O_9$, and $C_{44}H_{69}N_5O_9$, for viequeamides B-D, and $C_{45}H_{67}N_5O_9$ for viequeamides E/F which are isobaric. The molecular formulas of compounds **3** and **4**, $C_{44}H_{67}N_5O_9$ and $C_{44}H_{69}N_5O_9$, suggested that they might be related to viequeamide B **26** by varying levels of saturation of the alkyne functional group in the Dhoya residue. These hypotheses were confirmed by MS^2 fragmentation experiments which showed that **27** had a fragmentation pattern consistent with the alkene analog 2,2-dimethyl-3-hydroxyoctenoic acid (Dhoea), in viequeamide B, whereas **28** agreed well with the corresponding alkane

Table 2.2: *Summary of NMR data (in $CDCl_3$) for viequeamide B (**26**).* [a]Tightly clustered carbon signals at 30.4, 30.3, 30.3 and 30.2 could not be resolved by HSQC or HMBC, although a DEPT-135 did reveal 30.4 to be the methylene carbon C-12 in proline. [b]Minor rotomer at 2.79, s. [c,d]Overlapping signals in the proton spectrum.

Residue	Position	δ_C, type	δ_H (J in Hz)	HMBC	TOCSY	ROESY
Val-1	1	169.8, C	-			
	2	57.1, CH	4.77, dd (9.5, 3.9)	1, 3-6	Val-1-NH, 3-5	
	3	29.6, CH	2.31, m		Val-1-NH, 2, 4, 5	
	4	20.4, CH_3	0.99, d (6.8)	2, 3, 5	Val-1-NH, 2, 3, 5	
	5	18.2, CH_3	0.73, d (6.9)	2, 3, 4	Val-1-NH, 2-4	
	NH	-	7.57, d (9.6)	2, 6	2-5	9
N-Me-Ala	6	170.4, C	-			
	7	55.9, CH	4.28, q (7.0)	6, 8, 9	8	9, 11
	8	15.1, CH_3	1.48, d (7.0)	7, 6	7	
	9	30[a], CH_3	2.80, s[b]	7, 10		Val-1-NH, 7, 8
Pro	10	171.0, C	-			
	11	57.4, CH	4.17, m	12, 13	12, 13, 14	12, 13
	12	30.4[a], CH_2	1.45-1.62, m[c]			
	13a	21.8, CH_2	1.71, m[d]			
	13b	21.8, CH_2	1.54, m[c]			
	14a	45.9, CH_2	3.61, m	12, 13	13, 14b	
	14b	45.9, CH_2	3.24, m	12, 13, 15	13, 14a	
Pla	15	167.0, C	-			
	16	70.9, CH	5.72, dd (11.3, 5.1)	15, 17, 18, 22	17a, 17b	11, 19
	17a	37.9, CH_2	3.73, dd (11.8, 5.1)	15, 16, 18, 19	16	19
	17b	37.9 CH_2	3.28, t (11.6)	15, 16, 18, 19	16	19
	18	135.7, C	-			
	19	129.9, CH	7.42, d (7.4)	17, 20, 21	20, 21	11, 16, 17
	20	128.3, CH	7.28, m	18, 19	19, 21	
	21	126.9, CH	7.20, m	18, 19	19, 20	
N-Me-Val	22	170.5, C	-			
	23	64.9, CH	4.48, d (9.0)	22, 24-26, 28	24-26	27, 29
	24	30[a], CH	2.36-2.27, m	23	23, 25, 26	27
	25	19.9, CH_3	1.18, d (6.7)	23, 24, 26	23, 24, 26	27
	26	21.3, CH_3	1.03, d (6.7)	23-25	23-25	27
	27	30[a], CH3	2.99, s	23, 28		23-26
Val-2	28	172.6, C	-			
	29	53.5, CH	4.71, t (8.7)	28, 30-33	Val-2-NH, 30-32	23
	30	32.6, CH	2.02, m	29	29, 31, 32	
	31	18.3, CH_3	0.93, d (6.7)	29, 30, 32	Val-2-NH, 29, 30, 32	
	32	19.8, CH_3	0.93, d (6.7)	29-31	Val-2-NH, 29-31	
	NH	-	6.16, d (8.8)	28, 30-33	29-32	35
Dhoya	33	174.7, C	-			
	34	46.2, C	-			
	35	17.0, CH_3	1.26, s	33, 34, 36, 37		Val-2-NH
	36	25.7, CH_3	1.15, s	33-35, 37		37
	37	77.9, CH	5.75, dd (11.3, 1.8)	1, 33-35, 39	38-40	
	38	27.5, CH_2	1.71, m[d]		37, 39, 40 ,42	
	39	24.6, CH_2	1.45-1.63, m[c]		37, 38, 40 ,42	
	40	17.9, CH_2	2.23, dt (6.6, 2.5)	38, 39, 41, 42	37-39, 42	
	41	83.7, C	-			
	42	69.1, CH	1.95, t (2.5)		38-40	

analog, Dhoaa.

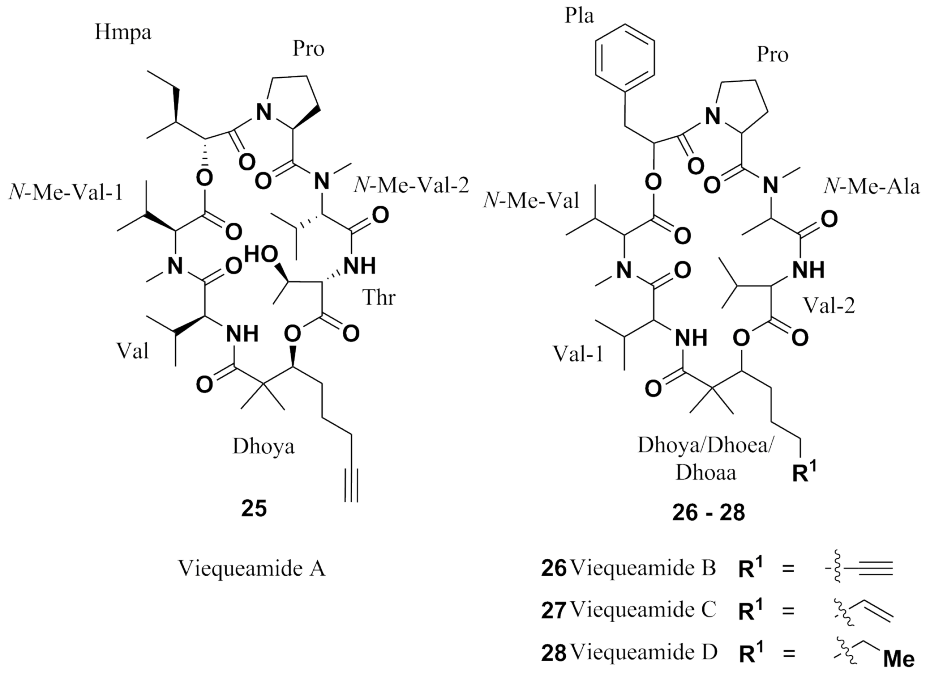


Figure 2.2: *Structures of the viequeamides (**25-28**).*

In the metabolites with heaviest *m/z* value, **29** and **30**, the additional 14 amu mass (relative to **26**) could be explained by elongation of the PKS chain by one carbon, methylation of an amide residue, or the exchange of a valine residue in **26** by either leucine or isoleucine. Surprisingly fragmentation of the parent ion of 822 *m/z* showed two overlapping fragmentation patterns, revealing that this peak was composed of a mixture of two isobaric compounds possessing the same molecular formula. Using the nonribosomal peptide sequencing tool previously developed by our laboratories [Nunnery et al., 2011], the structures of **29** and **30** were found to be analogous to **26**, with the addition of 14 mass units located alternately on Val-1 or Val-2 to either side of the Dhoya residue. In **29** Val-1 was modified whereas, in **30** Val-2 that was modified (see Figure 2). Unfortunately fragmentation experiments could not resolve the identity of the residues in these two compounds as either leucine, isoleucine, or *N*-methyl valine as all three have identical molecular formulas which would match the observed fragment masses. As such the structures of **29** and **30** are unresolved at

these residue positions.

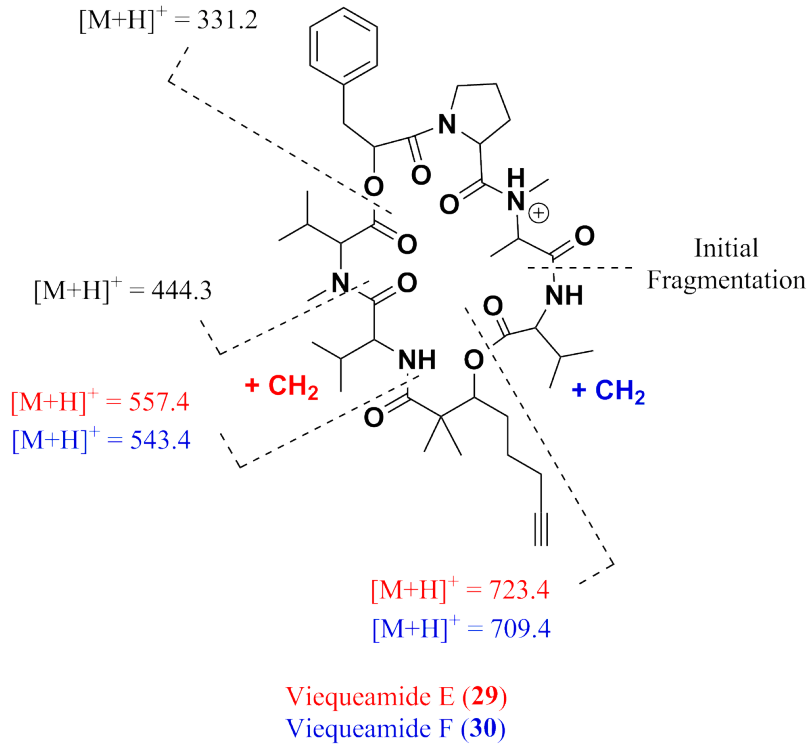


Figure 2.3: *MS^2 fragment ions for the isobaric compounds viequeamide E (**29**) and F (**30**).*

The viequeamides were evaluated for their cytotoxic properties using the H460 human lung cancer cell line. Surprisingly, while the viequeamides are structurally quite similar to one another, they showed dramatically different biological activities. **25** was observed to be potently cytotoxic with an $LD_{50} = 60 \pm 10$ nM, whereas a mixture of the minor compounds **26-30**; however, dominated by compound **26** as noted prior) showed no toxicity relative to control at the maximum dose evaluated (30 μg/mL).

The viequeamides are structurally related to **31**, a metabolite originally isolated by the Scheuer group from a predatory opistobranch mollusk [Reese et al., 1996]. Since that first report, there has been an increasing number of compounds reported of a related structural nature, we designate here these compounds as the "kulolide superfamily" which is recognizable by virtue of a β-hydroxy octanoic acid derivative along with a well defined sequence of amino acid or hydroxy acid residues in an overall cyclic arrangement. This

superfamily can be subdivided into two groups: one containing the Dhoya residue, or its more reduced equivalents, Dhoea, and Dhoaa, and the other possessing a Hmoya residue, or its brominated (Br-Hmoya) or reduced analogs, 3-hydroxy-2-methyl-7-octenoic acid (Hmoea), and 3-hydroxy-2-methyl-7-octanoic acid (Hmoaa). The structural trends between these subgroups are strongly conserved with the main exception being that the Hmoya subgroup always possesses a sequence of six residues, whereas the Dhoya subgroup usually contains one additional residue.

Within the kulolide superfamily the residue sequence can be summarized as: PKS residue (the presumed initiation of the biosynthetic pathway), valine, *N*-methylated non-polar amino acid, hydroxy acid analog of a non-polar amino acid, proline, non-polar amino acid which is often methylated at the amide nitrogen, and either another non-polar amino acid as in the case of the Dhoya subgroup or no further residues in the Hmoya subgroup. The most common non-polar amino acids observed are valine, isoleucine, and phenylalanine, and their modified *N*-methylated analogs, or the hydroxy acid analogs, 2-hydroxyisovaleric acid (Hiva), Hmpa, and phenyl lactic acid, respectively. A polar residue is only found in one example, the threonine of **25**, and thus may explain its exceptional cytotoxicity relative to other members of the class. The greatest variability in residues is for those furthest from the PKS-derived portion, which is intriguing as these are predicted to be the final residues of the mixed PKS-NRPS biosynthetic assembly process, and thus may represent the most recently introduced modules from an evolutionary perspective (e.g. a possible neofunctionalization of a duplicated NRPS module). For example, the third residue in the Dhoya subgroup is variable with either an *N*-methyl valine or an *N*-methyl phenylalanine residue whereas the final and seventh residue is alternately glycine, alanine, valine or threonine (Figure 3, Table 3).

There are many cyclic depsipeptides containing a Dhoya, Hmoya, or Dhoya/Hmoya analog residues that differ from the sequence of residues that defines the kulolide superfam-

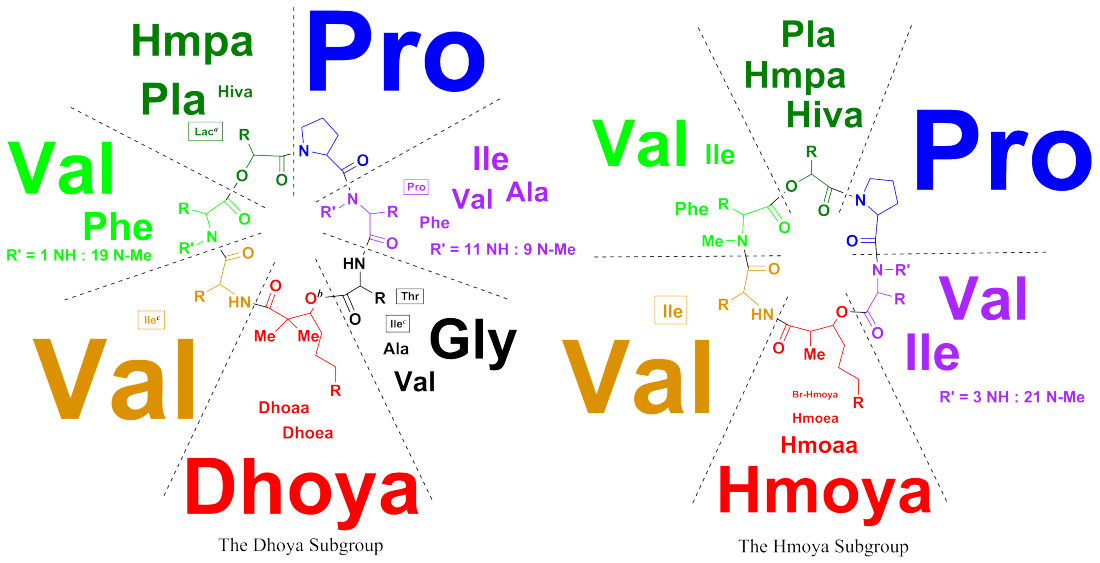


Figure 2.4: *MS*^{2} *The prevalence of residues found in the kulolide superfamily.* (Size of the font is proportional to the prevalence of those residues within the family. The residues in boxes are shown at twice their proportional size to enhance readability.) [a]Lac, lactic acid. [b]In kulokainalide-1 (**37**) there is a unique eighth residue, phenylalanine [Nakao et al., 1998]. [c]While shown as isoleucine, these residues were determined by fragmentation analysis and are unresolved between leucine, isoleucine, or *N*-methyl valine as noted in the text (see Figure 2, and S19).

ily. Examples include the Dhoya containing mantillamide (**38**) [Liu et al., 2009] and the wewakpeptins A-D (**39-42**) [Han et al., 2005], which are larger than the members of the Dhoya subgroup, and the yanucamide A (**43**) and B (**44**) [Sitachitta et al., 2000, Xu et al., 2003] and the cocosamide A (**45**) and B (**46**) [Gunasekera et al., 2011] which are shorter than the seven residue chain length common to the Dhoya subgroup. Georgamide (**47**) [Wan and Erickson, 2001] is an interesting case in that it would match well to the residue sequence trends observed in the Dhoya subgroup, but only if the sequence of residues was reversed in its order.

Ultimately, while trends can be deduced from the sequence and overall architecture of these metabolites, and we thus are able to infer associations between them, a compelling knowledge of their evolutionary relationships will require access to their biosynthetic genes and comprising nucleotide sequence information. For example, the dudawalamides are also a notable case because dudawalamide A (**48**) and B (**49**) do not fit into the kulolide

superfamily, while dudawalamide C-E (**50-52**) do [Malloy, 2011]. **48** has seven residues, as with most of the Dhoya Subgroup, but matches poorly to the trends in the residues observed for the kulolide superfamily. The sequence of the first three residues after the Dhoya residue is "lactic acid, alanine, *N*-methyl isoleucine". This does not match the trend of "amino acid (almost exclusively valine), *N*-methyl amino acid, hydroxy acid" which is observed in the rest of the Dhoya subgroup. In **49** the structure of cyclo-[Dhoya-Val-*N*-Me-Phe-Pro-*N*,*O*-Me-Tyr-Ala] is one residue shorter than is common in other compounds in the Dhoya subgroup. If a hydroxy acid were added before the proline residue, it would fit well within the kulolide superfamily. Likewise, if the lactic acid residue were moved to before the proline in **48**, then this compound would also fit well within the kulolide superfamily. Thus, while a possible conjecture is that **48** and **49** are members of the kulolide superfamily wherein the biosynthetic pathway has been modified by a rearrangement and a deletion, respectively, of the hybrid PKS-NRPS biosynthetic pathway, without sequence data for these biosynthetic genes, this conjecture is provocative, yet unsubstantiated.

The sources of metabolites in the kulolide superfamily are diverse, both in terms of geographical location and the nature of the organisms from which the compounds were first isolated. However, the fact that the kulolide superfamily bears the structural hallmarks of filamentous prokaryotic secondary metabolites as well as the actuality of isolation of this compound superfamily directly from benthic cyanobacteria, strongly suggests that the biosynthetic origin of all members of this superfamily derives from filamentous prokaryotes. The kulolides (**31**, **53**, **54**), kulomo'opunalides (**62**, **63**), and kulokainalide-1 (**37**) were all obtained from the predatory mollusk *Philinopsis speciosa* from Oahu; however, because of the distinctive structural features of these compounds, the Scheuer group drew the inference at that time that *P. speciosa* was likely accumulating these compounds through its diet of sea hares that in turn fed on cyanobacteria [Reese et al., 1996, Nakao et al., 1998]. All other members of the kulolide superfamily were isolated directly from filamentous cyanobacteria,

Table 2.3: *Metabolites of the kulolide superfamily and their residue sequences.* [a][Reese et al., 1996]. [b][Nakao et al., 1998]. [c]Kulokainalide-1 (**37**) is eight residues long, rather than the seven seen in most of the Dhoya Subgroup. [c][Luesch et al., 2001a]. [d][Montaser et al., 2011]. [e][Medina, 2009]. [f][Malloy, 2011]. [g]While shown as isoleucine, these residues were determined by fragmentation analysis and are unresolved between leucine, isoleucine, or *N*-methyl valine (see Figure 2, and S19). [h][Nogle and Gerwick, 2002]. [i][Bunyajetpong et al., 2006]. [j][Tripathi et al., 2009]. [k][Tripathi et al., 2010]. [l][Salvador et al., 2011]. [m][Mevers et al., 2011].

Compound Name	Residues 1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
Dhoya Subgroup							
Kulolide-1 (**31**)[a,b]	*S*-Dhoya	L-Val	D-*N*-Me-Val	*S*-Pla	L-Pro	L-Val	L-Ala
Kulolide-2 (**53**)[b]	*S*-Dhoea	L-Val	D-*N*-Me-Val	*S*-Pla	L-Pro	L-Val	L-Ala
Kulolide-3 (**54**)[b]	*S*-Dhoaa	L-Val	D-*N*-Me-Val	*S*-Pla	L-Pro	L-Val	L-Ala
Kulokainalide-1 (**37**)[b]	*S*-Dhoya	L-Val	D-*N*-Me-Val	*S*-Lac	L-Pro	L-Pro	L-Val-L-Phe[c]
Pitipeptolide A (**55**)[c]	*S*-Dhoya	L-Val	L-*N*-Me-Phe	2*S*,3*S*-Hmpa	L-Pro	L-Ile	Gly
Pitipeptolide B (**56**)[c]	*S*-Dhoea	L-Val	L-*N*-Me-Phe	2*S*,3*S*-Hmpa	L-Pro	L-Ile	Gly
Pitipeptolide C (**57**)[d]	*S*-Dhoaa	L-Val	L-*N*-Me-Phe	2*S*,3*S*-Hmpa	L-Pro	L-Ile	Gly
Pitipeptolide D (**58**)[d]	*S*-Dhoya	L-Val	L-Phe	2*S*,3*S*-Hmpa	L-Pro	L-Ile	Gly
Pitipeptolide E (**59**)[d]	*S*-Dhoya	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-Ile	Gly
Pitipeptolide F (**60**)[d]	*S*-Dhoya	L-Val	L-*N*-Me-Phe	2*S*,3*S*-Hmpa	L-Pro	L-Val	Gly
Radamamide A (61)[e]	*S*-Dhoya	L-Val	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Phe	Gly
Dudawalamide C (**50**)[f]	Dhoya	Val	*N*-Me-Val	Hiva	Pro	*N*-Me-Phe	Gly
Dudawalamide D (**51**)[f]	Dhoya	Val	*N*-Me-Val	Hmpa	Pro	*N*-Me-Phe	Gly
Dudawalamide E (**52**)[f]	*S*-Dhoya	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-Ile	Gly
Viequeamide A (**25**)	*S*-Dhoya	L-Val	L-*N*-Me-Val	2*R*,3*S*-Hmpa	L-Pro	L-*N*-Me-Val	L-Thr
Viequeamide B (**26**)	*S*-Dhoya	L-Val	L-*N*-Me Val	*S*-Pla	L-Pro	L-*N*-Me-Ala	L-Val
Viequeamide C (**27**)	Dhoea	Val	*N*-Me Val	Pla	Pro	*N*-Me Ala	Val
Viequeamide D (**28**)	Dhoaa	Val	*N*-Me Val	Pla	Pro	*N*-Me Ala	Val
Viequeamide E (**29**)	Dhoya	Ile[g]	*N*-Me Val	Pla	Pro	*N*-Me Ala	Val
Viequeamide F (**30**)	Dhoya	Val	*N*-Me Val	Pla	Pro	*N*-Me Ala	Ile[g]
Hmoya Subgroup							
Kulomo'opunalide-1 (**62**)[b]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Ile	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me Ile	–
Kulomo'opunalide-2 (**63**)[b]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Ile	*S*-Hiva	L-Pro	L-*N*-Me Ile	–
Antanapeptin A (**64**)[h]	Hmoya	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-*N*-Me-Ile	–
Antanapeptin B (**65**)[h]	Hmoea	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-*N*-Me-Ile	–
Antanapeptin C (**66**)[h]	Hmoaa	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-*N*-Me-Ile	–
Antanapeptin D (**67**)[h]	Hmoya	L-Val	L-*N*-Me-Phe	*S*-Hiva	L-Pro	L-*N*-Me-Val	–
Radamamide B (**68**)[e]	Hmoya	L-Ile	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Ile	–
Trungapeptin A (**69**)[i]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	*S*-Pla	L-Pro	L-*allo*-Ile	–
Trungapeptin B (**70**)[i]	2*S*,3*R*-Hmoea	L-Val	L-*N*-Me-Val	*S*-Pla	L-Pro	L-*allo*-Ile	–
Trungapeptin C (**71**)[i]	2*S*,3*R*-Hmoaa	L-Val	L-*N*-Me-Val	*S*-Pla	L-Pro	L-*allo*-Ile	–
Hantupeptin A (**72**)[j,k]	2*R*,3*S*-Hmoya	L-Val	L-*N*-Me-Ile	*S*-Pla	L-Pro	L-*N*-Me-Val	–
Hantupeptin B (**73**)[k]	2*R*,3*S*-Hmoea	L-Val	L-*N*-Me-Ile	*S*-Pla	L-Pro	L-*N*-Me-Val	–
Hantupeptin C (**74**)[k]	2*R*,3*S*-Hmoaa	L-Val	L-*N*-Me-Ile	*S*-Pla	L-Pro	L-*N*-Me-Val	–
Veraguamide A (**75**)[l,m]	8-Br-2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Val	–
Veraguamide B (**76**)[l,m]	8-Br-2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	*S*-Hiva	L-Pro	L-*N*-Me-Val	–
Veraguamide C (**77**)[l,m]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Val	–
Veraguamide D (**78**)[l]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Ile	–
Veraguamide E (**79**)[l]	2*S*,3*R*-Hmoya	L-Ile	L-*N*-Me-Ile	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Val	–
Veraguamide F (**80**)[l]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	*S*-Pla	L-Pro	L-*N*-Me-Val	–
Veraguamide G (**81**)[l]	2*S*,3*R*-Hmoya	L-Val	L-*N*-Me-Val	2*S*,3*S*-Hmpa	L-Pro	L-*N*-Me-Val	–
Veraguamide H (**82**)[m]	Hmoya	Val	*N*-Me Val	Hiva	Pro	*N*-Me Val	–
Veraguamide I(**83**)[m]	Hmoaa	Val	*N*-Me Val	Hmpa	Pro	*N*-Me Val	–
Veraguamide J (**84**)[m]	Hmoaa	Val	*N*-Me Val	Hiva	Pro	*N*-Me Val	–
Naopeptin (**85**)[f]	Hmoya	Val	*N*-Me-Val	Pla	Pro	*N*-Me-Ile	–

reported variably as *Lyngbya majuscula* (*Moorea producens*) [Luesch et al., 2001a, Montaser et al., 2011, Medina, 2009, Nogle and Gerwick, 2002, Bunyajetpong et al., 2006, Tripathi et al., 2009, Tripathi et al., 2010, Wan and Erickson, 2001], *Moorea sp.* [Malloy, 2011], *Symploca cf. hydnoides* [Salvador et al., 2011], and *Oscillatoria margaritifera* [Mevers et al., 2011]. However, these identifications, except for the identification of *Oscillatoria margaritifera* as the producer of **75-77** and **82-84**, were based on morphology and not 16S rRNA phylogenetic analysis. Considering the fact that the genera *Oscillatoria* and *Moorea* (previously *Lyngbya*) are morphologically very similar to one another, identifications based solely on this characteristic should be considered quite tenuous [Mevers et al., 2011, Engene et al., 2012, Castenholz et al., 2001].

The cyanobacterial strain producing the viequeamides was analyzed for its phylogenetic relationship to other cyanobacteria using the 16S rRNA gene sequence. This analysis revealed that it possessed a closer relationship to members of the family Nostocaceae than Oscillatoriaceae. Unfortunately, as with genera in the family Oscillatoriaceae, over-reliance on morphological characteristics has led some strains of *Calothrix*, *Rivularia*, *Tolypothrix*, and their closely related genera in the family Nostocaceae to be misidentified [Sihvonen et al., 2007, Domínguez-Escobar et al., 2011, Komárek and Anagnostidis, 1989]. In addition nutrient availability can produce morphological differences within a strain, further complicating identification of material collected from natural habitats when compared to culture collection strains grown under culture conditions [Berrendero et al., 2011].

The 16S rRNA phylogenetic analysis of this viequeamide-producing strain revealed that it aligned best with the genus *Rivularia*. This analysis further indicated that the two closely related genera *Calothrix* and *Rivularia* formed separate clades (see Figure A.33 and A.34 in Appendix). The *Calothrix* clade was well supported with the exception of three sequences reported as *Rivularia*, EU009147, EU009149, and EU009150 (which showed 100% co-identity of residues for overlapping portions of their sequences). However, a prior

report has suggested that the sequence EU009149 was incorrectly identified, and pertains to *Calothrix*, which is in good agreement with the analysis presented here [Domínguez-Escobar et al., 2011]. The *Calothrix* strains that fell outside of the *Calothrix* clade were most likely misidentified as *Calothrix*, probably as a result of a morphology based identification.

The *Rivularia* clade contained strains alternately reported as *Calothrix*, unclassified Nostocales strains, as well as the uncultured 'button' cyanobacterium previously reported as *Dichothrix sp.* from thrombolites in Highborne Cay, Bahamas [Berrendero et al., 2011].

The viequeamide producer clustered robustly to the *Dichothrix sp.* strains, but together these strains did not separate from the *Rivularia* clade. *Dichothrix* is a poorly represented genus in culture collections and among published 16S rRNA sequences, a confounding factor in this phylogenetic analysis of the viequeamides producer. However, with the viequeamides producing strain and the *Dichothrix* strains robustly falling within the *Rivularia* clade, this strongly suggests that all of these organisms should be assigned to the genus *Rivularia* and not to *Dichothrix*.

Most members of the kulolide superfamily were isolated from collections made in the Pacific Ocean: Oahu in Hawaii [Reese et al., 1996, Nakao et al., 1998], Thailand [Bunyajetpong et al., 2006], Papua New Guinea [Malloy, 2011], the Pacific coast of Panama [Mevers et al., 2011], Singapore [Tripathi et al., 2009, Tripathi et al., 2010], and Guam [Luesch et al., 2001a, Montaser et al., 2011, Salvador et al., 2011]. The veraguamides were isolated from both the Eastern and Western regions of the Pacific Ocean (Coiba, an island on the Pacific side of Panama and Guam) [Salvador et al., 2011, Mevers et al., 2011]. However, the antanapeptins and radamamides were isolated from Madagascar in the Indian Ocean [Medina, 2009, Nogle and Gerwick, 2002]. The isolation of the viequeamides from the Caribbean Sea (Atlantic Ocean) represents the first members of the kulolide superfamily to be isolated from this ocean. Clearly, the kulolide superfamily of metabolites is distributed pan-tropically and thus likely represents an ancient biosynthetic pathway present in many

different lineages of cyanobacteria.

A final point to note is the remarkable difference in biological properties within these newly isolated viequeamide natural products. Variation in biological properties has been previously noted in other members of the kulolide superfamily, such as the veraguamides [Salvador et al., 2011, Mevers et al., 2011]. However, a comprehensive understanding of the structural features dictating these varying biological properties is not possible now because different protocols and cell lines have been used in the various studies of the biological properties of the kulolide family of compounds, and the relevant biological target or targets have not yet been defined. This latter area seems an especially promising area for investigation and efforts are underway to discover the cytotoxicity-related biological target of **25**.

2.4 Experimental Section

2.4.1 General Experimental Procedures

Optical rotation was measured on a Jasco P-2000 Polarimeter. UV/visual-light spectra were measured on a Beckman Coulter DU 800 Spectrophotometer. IR spectra were measured on a Thermo Electron Corporation Nicolet IR 100 FT-IR. NMR spectra were collected on a Varian Unity 500 MHz (500 MHz and 125 MHz for the ^{1}H and ^{13}C nuclei respectively) using $CDCl_3$ from Cambridge Isotope Laboratories, Inc. 99.8% D containing 0.03% *v/v* trimethylsilane (δ_H 0.0 and δ_C 77.16 as internal standards using trimethylsilane and $CDCl_3$, respectively). Microwave heated reactions were run in a Biotage Initiator microwave synthesizer. LCMS data for stereochemical analysis of the hydrolysates of **25** and **26** were obtained with a Phenomenex Luna 5 μm C18(2) 100Å column (4.6 x 250 mm) with a Thermo Finnigan Surveyor Autosampler-Plus/LC-Pump-Plus/PDA-Plus system and a Thermo Finnigan LCQ Advantage Max mass spectrometer. The elution for this analysis

began with 10% CH_3CN/90% H_2O acidified with 0.1% *v/v* formic acid, and immediately ramped to 50% CH_3CN/50% H_2O acidified with 0.1% formic acid over 85 min at a flow rate of 0.4 mL/min (monitoring 200-600 nm and *m/z* 250-2000 in both positive and negative ion mode). LC-HRMS data for analysis of compounds **2-6** were obtained on an Agilent 6239 HR-ESI-TOFMS with a Phenomenex Luna 5 μm C18(2) 100Å column (4.6 x 250 mm) and a gradient starting at 60% CH_3CN/40% H_2O and immediately ramping to 100% CH_3CN over 20 min, then holding at 100% CH_3CN for 5 min. GCMS conducted with a Thermo Electron Corp. DSQ/TRACE-GC-Ultra GCMS system with a Cyclosil B column (Agilent Technologies J&W Scientific, 30 m x 0.25 mm). MS fragmentation experiments were run with a Biversa Nanomate (Advion Biosystems, Ithaca, NY) electrospray source for a Finnigan LTQ-FTICR-MS instrument (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0. HPLC purification was carried out with a Waters 515 HPLC Pump with a Waters 996 Photodiode Array Detector using Empower Pro software. All solvents were HPLC grade except for 99.8% acetone from Fisher which was distilled before use, and water which was purified by a Millipore Milli-Q system before use.

2.4.2 Collection and Identification of Cyanobacteria

Small 1-2 cm brown cyanobacterial puffballs were found growing in 0.5-1.5 m of water adhered to coral rubble, eelgrass, and dead gorgonian soft corals at Playa de la Chiva on Vieques Island in the Commonwealth of Puerto Rico, USA. A 1 L sample was preserved in 1:1 isopropanol/seawater solution, transported back to our laboratory in San Diego and stored at -20 ° C until it was extracted. A small sample of cyanobacterial biomass was also persevered in RNA stabilization reagent (RNAlater^{TM} from QIAGEN) for subsequent 16S rRNA sequencing and analysis. A voucher sample was stored in our laboratory (VQC-26/MAR/11-1).

2.4.3 Polymerase Chain Reaction (PCR) and Cloning

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Inc., Madison, WI, USA) following the manufacturer's specifications. DNA concentration and purity were measured on a DU® 800 spectrophotometer (Beckman Coulter). The 16S rRNA genes were PCR-amplified from isolated DNA using the cyanobacterial specific primers 27F 5'-AGAGTTTGATCCTGGCTCAG-3' and 809R 5'-GCTTCGGCACGGC TCGGGTCGATA-3'. The PCR reaction contained 1.0 L (ca. 100 ng) of DNA, 2.5 μL of 10 PfuUltra IV reaction buffer, 1.0 μL (10 mM) of dNTP mix, 1.0 μL of each primer (10 μM), 1.0 μL of PfuUltra IV fusion HS DNA polymerase and 17.5 μL of water for a total volume of 25 L. The PCR reactions were performed in an Eppendorf® Mastercycler® gradient as follows: initial denaturation for 4 min at 95 ° C, amplification by 30 cycles of 30 sec at 95 ° C, 30 sec at 50 ° C and 1 min at 72 ° C, and final elongation for 7 min at 72 ° C. PCR products were purified using a MinElute® PCR Purification Kit (Qiagen) before subcloning with the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen) following the manufacturer's specifications. Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen) and sequenced with M13 primers. The 16S rRNA gene sequence is available in the DDBJ/EMBL/GenBank databases under acc. No. JQ979179.

2.4.4 Phylogenetic Inferences.

All gene sequences were analyzed using Geneious Pro v.5.5.4 [Drummond et al., 2011]. The 16S rRNA gene sequences were aligned using the L-INS-I algorithm in MAFFT v6.814b [Katoh et al., 2002]. Best-fitting nucleotide substitution models optimized by maximum likelihood (ML) selected using corrected Akaike/Bayesian Information Criterion (AIC/BIC) in jModelTest v0.1.1 [Posada, 2008]. The evolutionary histories of the cyanobacterial genes were inferred using ML and Bayesian inference algorithms. The ML inference was performed using PhyML [Guindon and Gascuel, 2003] in Geneious

Pro v5.5.4. The analysis was run using the GTR+I+G model (selected by AIC and BIC criteria) assuming heterogeneous substitution rates and gamma substitution of variable sites (proportion of invariable sites (pINV) = 0.067, shape parameter (α) = 0.275, number of rate categories = 4). Bootstrap resampling was performed on 1,000 replicates. Bayesian analysis was conducted using MrBayes [Huelsenbeck and Ronquist, 2001] in Geneious Pro v5.5.4 with four Metropolis-coupled MCMC chains (one cold and three heated) ran for 3,000,000 generations. The first 10% were discarded as burn-in and data set was sampled with a frequency of every 200 generations. The molecular clock parameters were set for uniform branch lengths, tree height exponential = 1 and shape parameter exponential = 10.

2.4.5 Cytotoxicity Assay

H460 cells were added to 96-well plates at 3.33 x 104 cells/mL of Roswell Park Memorial Institute (RPMI) 1640 medium with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The cells, in a volume of 180 μL per well, were incubated overnight (37 $^{\circ}$ C, 5% CO_2) to allow recovery before treatment with the test compounds. Compounds were dissolved in dimethylsulfoxide to a stock concentration of 10 mg/mL. Working solutions of the compounds were made in RPMI 1640 medium without FBS, with a volume of 20 μL added to each well to give a final compound concentration of either 30 or 3 μg/mL. An equal volume of RPMI 1640 medium without FBS was added to wells designated as negative controls for each plate. Plates were incubated for approximately 48 h before staining with Thiazolyl Blue Tetrazolium Bromide (98% Sigma). Using a ThermoElectron Multiskan Ascent plate reader, plates were read at 570 and 630 nm. Concentration response graphs were generated using GraphPad Prism (GraphPad Software Inc., San Diego, CA).

2.4.6 Extraction and Isolation

The preserved biomass/alcohol solution was filtered through cheese cloth and the crude algal biomass extracted by soaking for 15 min in 600 mL of 2:1 CH_2Cl_2/MeOH with manual fragmentation of the cyanobacterial clumps. The organic layer was then filtered through cheese cloth and the algal biomass was repetitively re-extracted six times by submerging in 2:1 CH_2Cl_2/MeOH with mild warming ($\leq$ 30 ° C) for 30 min and then filtering through cheese cloth. The aqueous layer was concentrated to a reduced volume (ca. 500 mL) by rotory evaporation ($\leq$ 35 ° C) and then extracted with an equal volume of CH_2Cl_2 using 30 mL of brine to assist in phase separation. All of the organic layers were then combined and concentrated by rotory evaporation to afford 6.6 g of black-green solids. A small portion was preserved for future bioassays and the remainder was purified by VLC with 300 mL of silica gel (Type H, 10-40 μm, Sigma-Aldrich) in a 10 cm diameter by 9 cm glass vacuum armed frit; fractions of 300 mL volume using progressively more polar mixtures of hexanesEtOAcMeOH (nine fractions, 100% hexanes, 10% EtOAc/90% hexanes, 20% EtOAc/80% hexanes, then 20% increments to 100% EtOAc, 25% MeOH/75% EtOAc, and 100% MeOH).

The 100% EtOAc eluting fraction was concentrated by rotory evaporation to afford 0.245 g of a dark brown-orange oil. This material was purified by RP-SPE with a Strata 6 mL column with 1 g of C18-E (55 μm, 70 Å) packing material prepared with five column washes of CH_3CN followed by five column volumes of 50% CH_3CN/50% H_2O. The material from the EtOAc fraction was dissolved in 50% CH_3CN/50% H_2O, loaded onto the column, and six fractions obtained using 30 mL each of 50% CH3CN/50% H_2O, 60% CH_3CN/40% H_2O, 80% CH_3CN/20% H_2O, CH_3CN, MeOH, and CH_2Cl_2. The 50% CH_3CN/50% H_2O fraction was sequentially placed in 3 mL of EtOAc and 3 mL of 75% CH_3CN/25% H_2O, the insoluble material removed by filtration, and the filtrate purified by RP-HPLC using a Synergi Fusion RP-80 (250 x 10.00 mm, 4 μ, isocratic elution with 60% CH_3CN/40% H_2O at 3 mL/min

over 70 min). This afforded 14.8 mg of a white solid containing ca. 92% pure viequeamide A (**1**) (purity by ^{1}H NMR integration) which was used for NMR-based structure elucidation. Re-purification of this material by RP-HPLC using the same conditions gave 13.6 mg of a white solid which had no detectible impurities by NMR, but trace **2** when using selected ion chromatogram LCMS; this sample was used for bioassays and measurement of optical rotation.

The insoluble matter from the EtOAc and 75% CH_3CN/25% H_2O solubilization steps contained compounds similar to **25** by ^{1}H NMR and LCMS. These materials were combined, dissolved in pure CH_3CN, and purified by RP-HPLC using a Synergi Fusion RP-80 (250 x 10.00 mm, 4 μ column, isocratic elution of 2:1 CH_3CN/H_2O elution at 3 mL/min over 40 min). The HPLC fractions from this purification were still impure; however, LCMS and ^{1}H NMR analyses suggested the presence of a mixture of interconverting *N*-methyl amide rotomers that caused broad and overlapping elution peaks, thus yielding impure mixtures. Several fractions from this latter HPLC were combined to yield a simplified mixture with a single major compound ($\geq$ 95% compound **26**, by ^{1}H NMR), and this was analyzed by 2D NMR and mass spectrometry to provide the planar structure of **26**. The remaining uncombined fractions from the latter HPLC work (0.5 mg) contained a mixture of compounds **26-30** in which **27** and **28** predominated.

Viequeamide A (**25**): white amorphous solid; $[\alpha]23_D$ 32.6 (CH_2Cl_2); UV (MeOH) λ_{max} (log) 204 (4.30) nm; IR (neat) ν_{max} 3387, 2967, 1749, 1641, 1455, 1366, 1308, 1154, 1122, 733 cm^{-1}; see Table 1 for NMR data; HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 826.4940 (calculated for $C_{42}H_{69}N_5O_{10}Na$ 826.4937).

Viequeamide B (**26**): white amorphous solid; see Table 2 for NMR data; HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 830.4670 (calculated for $C_{44}H_{65}N_5NaO_9$ 830.4674).

2.4.7 LC-HRMS Fragmentation Analysis of Viequeamides B-F (26-30)

LCMS of a purified fraction containing viequeamides B-F (**26-30**) revealed four major masses at *m/z* 830.4666, *m/z* 832.4833, *m/z* 834.4988, and *m/z* 844.4833, corresponding to molecular formulas $C_{44}H_{65}N_5NaO_9$, $C_{44}H_{67}N_5NaO_9$, $C_{44}H_{69}N_5NaO_9$, and $C_{45}H_{67}N_5NaO_9$, respectively. To characterize the planar structures of 3-6, a separate experiment was conducted to fragment the major masses observed in the LC-HRMS analysis. The data were obtained on these four *m/z* values from two separate HPLC fractions at ca. 0.1 mM (CH_3CN), one of which was from HPLC fractions dominated by compound **26**, and the other from HPLC fractions dominated by **27** and **28**. A 50% MeOH/50% H_2O solution was acidified with 1% *v/v* formic acid (Acros Organics 98+ %) and used to dilute 10-fold the 0.1 mM samples (e.g. 1 μL of sample into 10 μL acidified MeOH/H_2O). Samples were subjected to electrospray ionization on the Nanomate nano-spray source (pressure: 0.3 psi, spray voltage: 1.4kV), then fragmented and analyzed on the LTQ-FT-MS. Generally, the instrument was first auto-tuned on the *m/z* value of the ion to be fragmented. Then, the $[M + H]^+$ ion of each compound was isolated in the linear ion trap and fragmented by collision induced dissociation. Sets of consecutive, high-resolution MS/MS scans were acquired in profile mode and averaged using QualBrowser software by Thermo (see Figure A.16-19 in Appendix).

2.4.8 Stereochemical Analysis of Viequeamide A (25) by Marfey's Analysis and Chiral GCMS

A 60 μL aliquot (600 μg) was transferred from a 10 mg/mL solution of **25** in CH_3CN to a conical microwave reaction tube with a stir bar. The solvent was removed under N_2 (*g*) flow and the resulting solid was dissolved in 600 μL of 6 M HCl (*aq*). This solution was

stirred and heated by microwave to 160 °C for 5 min, and the resulting material was divided into two equal portions and both were dried under N_2 (*g*) flow.

One portion was dissolved in 0.75 mL of a 1 M $NaHCO_3$ (*aq*) solution, and to this was added drop wise 71 μL of 1-fluoro-2,4-dinitrophenyl-D-alanine amide (D-FDAA) in acetone (1.0 mg/mL), incubated for 1 h at 35-45 °C, and then quenched with 370 μL of 2 M HCl (*aq*). The reaction mixture was transferred to fresh vial using CH_3CN/H_2O mixtures, reduced in volume under N_2 (g) flow, and diluted to ca. 1 mL with 50% CH_3CN/50% H_2O; 10 μL of this mixture was analyzed by HPLC-ESIMS comparison to authentic amino acid standards, all D-FDAA derivatized. Retention times for the authentic standards were as follows: D-Val (62.48 min), L-Val (71.58 min), D-Thr (41.30 min), D-*allo*-Thr (41.46 min), L-Thr (48.60 min), L-*allo*-Thr (44.74 min), D/L-*N*-Me-Val (69.92 and 74.73 min), L-*N*-Me-Val (74.70 min), D-Pro (51.55 min), and L-Pro (54.53 min). The hydrolysate peaks with the expected masses were found at 71.82, 48.79, 74.91, and 54.53 min which correspond to L-Val, L-Thr, L-*N*-Me-Val, and L-Pro.

The second portion of the acid hydrolysis product was dissolved in 0.6 mL of 1:1 Et_2O/MeOH, and CH_2N_2 in Et_2O was added by polished glass pipette until the solution was a persistent yellow color. The reaction mixture was allowed to stand for 0.5 h at room temperature and then dried under N_2 (*g*) flow. The resulting material was dissolved in ca. 12 μL of CH_2Cl_2 and 1 μL was injected into the GCMS system at 35 °C for 15 min followed by a ramp of 1.5 °C/min until 60 °C was reached at which time the GC was ramped at 50 °C/min until 170 °C where the temperature was held for 5 min. (For all GCMS runs, the ion source was held at 250 °C and the positive ion scan from 50 to 1000 *m/z* was collected from 7 min until the end of the run.) Retention times of previously prepared authentic standards,21 were measured for 2*S*,3*S*-Hmpa, 2*S*,3*R*-Hmpa, 2*R*,3*R*-Hmpa, and 2*R*,3*S*-Hmpa (40.4, 39.4, 40.0, and 39.8 min respectively) and compared with that of the natural product hydrolysate (39.8 min) showing the residue in viequeamide A to be 2*R*,3*S*-Hmpa. Coinjections of the

standards and the natural product hydrolysate confirmed this result.

To determine the configuration of the stereocenter in the Dhoya moiety, a stirred solution containing 400 μg of **1** in 400 μL of EtOH was treated with 1.25 mg of 10% palladium on carbon, and the atmosphere was replaced with H_2 (*g*) by balloon. The reaction mixture was stirred at room temperature for 3 h and filtered through glass wool into a microwave reaction tube with a stir bar, with ca. 2 mL CH_2Cl_2. The sample was then concentrated under N_2 (*g*) flow, redissolved in 800 μL of 2 M HCl (*aq*), and the resulting mixture stirred and heated by microwave at 160 °C for 5 min. The reaction mixture was partially dried under N_2 (*g*) flow and the remaining residue of water removed via azeotrope with 2 mL of benzene. The resulting solids were dissolved in 400 μL of 1:1 Et_2O/MeOH, and while stirring CH_2N_2 in Et_2O was added in excess. After 1.25 h significant color loss was observed, and thus another portion of CH_2N_2 in Et_2O was added and stirring was continued for 45 min. The reaction mixture was concentrated under N_2 (*g*) flow and the residue quickly dissolved in CH_2Cl_2, transferred to a GCMS vial, and re-concentrated under N_2 (*g*) flow to ca. 12 μL total volume.

Chiral GCMS analysis (Cyclosil B) of the above product and synthetic Dhoya standards based on a method recently described [Mamer, 2000]. Briefly, the initial temperature was 40 °C with an immediate ramp of 20 °C/min to 100 °C where upon the temperature was held for 4 min. The oven temperature was then again ramped at 2 °C/min until 130 °C was reached where upon the temperature was held for 5 additional min. The retention times of both the *S* and the *R* standards (24.3 and 24.5 min, respectively) were measured and compared with that of the natural product hydrolysate (24.3 min), thus revealing the presence of an *S*-Dhoya unit in viequeamide A. Coinjections of standards and the natural product hydrolysate confirmed this result.

2.4.9 Stereochemical Analysis of Viequeamide B (26) by Marfey's Analysis and Chiral GCMS

A 700 μg sample of **26** (from HPLC described above, minor analogs **27-30** present but mixture dominated by **26** by ^{1}H-NMR) was dissolved in 700 μL of ethanol, treated with 2.0 mg of 10% palladium on carbon, and the atmosphere replaced with H_2 (*g*) by balloon. The reaction mixture was stirred at room temperature for 1 h and filtered through glass wool with 3 x 0.5 mL of CH_2Cl_2. The filtrate was concentrated to dryness by rotory evaporation, and transferred to a microwave reaction tube with a stir bar, again through glass wool with 3 x 0.5 mL of CH_2Cl_2. The sample was concentrated under N_2 (*g*) flow, redissolved in 700 μL of 2 M HCl (*aq*), and the resulting mixture was stirred and heated by microwave at 160 C for 5 min. The reaction mixture was then concentrated to dryness under N_2 (*g*) flow, and the resulting solids were dissolved in 700 μL of CH_2Cl_2 and partitioned into a 400 μL and 300 μL sample.

The 300 μL sample was concentrated to dryness in a fresh vial and redissolved in 300 μL of 1 M $NaHCO_3$ (*aq*), and a stir bar and 1.0 mL of L-FDVA in acetone (at a concentration of 1.0 mg/mL) was added. The valine amide analog was selected because in testing the L-FDAA derivatized authentic *N*-Me-Ala standards, no separation was observed between diastereomers despite lengthening the elution gradient of the HPLC method. The reaction mixture was then stirred at 30 °C for 1 h, and then quenched with 150 μL of 2 M HCl (*aq*). The reaction mixture was then transferred to a fresh vial with 3 x 1.0 mL of CH_3CN through an Advantec syringe filter (HP020AN), concentrated to dryness under N_2 (*g*) flow, and transferred to an LCMS vial again through a syringe filter with 3 x 0.5 mL of CH_3CN; 10 μL of this mixture was analyzed by HPLC-ESIMS comparison to authentic amino acid standards, all L-FDVA derivatized. Retention times for the authentic standards were as follows: D/L-*N*-Me-Ala (76.85 and 77.91 min), L-*N*-Me-Ala (76.80), D-Val (101.73 min), D/L-Val (82.39 and 101.88 min), D/L-*N*-Me-Val (91.64 and 103.87 min), L-*N*-Me-Val

(91.69 min), D-Pro (79.10 min), and L-Pro (70.58 min). The hydrolysate peaks with the expected masses were found at 76.40, 82.24, 91.69, and 70.59 min which correspond to L-*N*-Me-Ala, L-Val, L-*N*-Me-Val, and L-Pro.

The 400 μL sample was concentrated to dryness redissolved in 400 μL of 1:1 Et_2O/MeOH and while stirring, CH_2N_2 in Et_2O was added in excess. After 1 h significant color loss was observed, and thus another portion of CH_2N_2 in Et_2O was added, and stirring continued for 1 h. The reaction was then concentrated to dryness under N_2 (*g*) flow and the residue quickly transferred to a GCMS vial with 5 x 50 μL of CH_2Cl_2.

Chiral GCMS analysis (Cyclosil B) of the above product and synthetic Dhoaa standards was performed with the identical method used for the hydrolysate of **25**. The retention times of the *S* and *R* standards (23.7 and 23.9 min, respectively) were measured and compared with that of the natural product hydrolysate (23.7 min), thus revealing the presence of an S-Dhoya unit in **26**. Coinjections of standards and the hydrogenated natural product hydrolysate confirmed this result.

A separate method was used to compare the hydrolysate to commercially available standards for Pla. Using the same column as above the initial temperature of 40 °C was held steady for 5 min, then ramped at 5 °C/min to 100 C, then immediately ramped to 130 °C at 1.5 °C/min. The retention times of the *S* and *R* standards (44.78 and 43.45 min, respectively) were measured and compared with that of the natural product hydrolysate (45.00 min), thus revealing the presence of an *S*-Pla unit in viequeamide B. Coinjections of standards and the hydrogenated natural product hydrolysate confirmed this result.

While compounds **27-30** were present in this mixture, because of their low concentration in the sample it is unlikely that products from the hydrolysis of **27-30** could be detected if they differed from the major peaks from **26**, and thus the stereoconfiguration of **27-30** must still be considered unresolved.

2.5 Acknowledgments

We thank H. Choi and E. Mevers for providing the Hpma and several Marfey's standards, J. Nunnery for providing synthetic Dhoya standards, A. Mrse and J. Rho for assistance in setting up NMR experiments, and E. Monroe for help with the phylogenetic analysis. We acknowledge E. Theodorakis and A. Saitman (UCSD Chemistry) for use of, and help with, the microwave reactor, and the UCSD Chemistry and Biochemistry Mass Spectrometry Facility for LC-HR-ESI-TOFMS experiments and analysis. Funding was provided by the E. W. Scripps Fellowship (PB), National Institutes of Health grant CA100851 (WHG), and R01 GM082683 (PCD).

This chapter, in full, is a reprint, with permission, of the material as it appears in the Journal of Natural Products, 2012, Boudreau, Paul D.; Byrum, Tara; Liu, Wei-Ting; Dorrestein, Pieter C.; Gerwick, William H. volume 75, pages 1560-1570.

Chapter 3

Difficulties in the Dereplication Workflow, the Need for Orthogonal Workflows

3.1 Abstract

Dereplication, the identification of known molecules during a drug discovery effort, is vital. Every step of the natural products drug discovery process, collection of samples, lipophilic extraction, purification, and structure elucidation bear a cost in terms of resources and work hours. If dereplication were only possible at the last stage by comparison of an elucidated structure to literature reports, then the cost of such an effort would be very high. Instead, an effort must be made to optimize each step along the workflow. For example, at the first step of a drug discovery effort, collection of samples can be targeted to new environments or species. Ideally dereplication can also be brought into each step of the workflow. When possible, our laboratory targets extractions to strains of interest as defined by 16S rRNA phylogeny. This protocol fails, however, when dealing with large numbers

of samples, where 16S isolation can be costly and expensive, and can miss new chemical diversity in cases where closely related organisms have differing secondary metabolomes, whether from horizontal gene transfer or high biosynthetic diversity within a closely related group. Phyloproteomics-based mass spectrometry screening is an attractive alternative, because it is relatively inexpensive, rapid, and easily scalable; however, it is also potentially susceptible to the same drawbacks as 16S phylogeny when dealing with closely related organisms with highly divergent metabolomes. Unfortunately, when this method was applied to filamentous marine cyanobacteria, the polysaccharide sheaths contaminated the protein extract and overwhelmed the necessary signals of the proteome. This failure forced us to consider other alternative workflows, and what data would be the most useful in a drug discovery effort. As our targets in this work are second metabolites, techniques that directly assess the metabolome are of great use. MS fragmentation based 'molecular networking' analysis, which compares compounds directly based on their MS^2 fragmentation patterns, is just such a method. Using an LCMS system also allows dereplication to be incorporated into earlier stages of the workflow. Crude extracts and preliminary fractions can easily be screened to create a molecular network where ions can be analyzed for similarity within and between extracts but also to libraries of other samples or pure compounds. Screening of a *Moorea bouillonii* strain from New Ireland, Papua New Guinea (the type strain PNG05-198^T [Engene et al., 2012]) revealed the known compounds lyngbabellin A and apratoxin A and B within the network, and demonstrated the ability 'molecular networking' to dereplicate known compounds within our laboratory's extracts.

3.2 Introduction

One of the first compounds ever isolated from a marine source was dendrolasin (**1**) [Vanderah and Schmitz, 1975]. Isolated from the sponge *Oligoceras hemorrhages*, it was a remarkable discovery and opened a new source for natural products; however, **1** was also a

known compound. Almost two decades earlier Quilico et al. isolated this compound from the ant *Lasius* (*Dendrolasius*) *fuliginosus* [Quilico et al., 1956, Quilico et al., 1957]. At the time there was no effective way to dereplicate compounds other than through a complete structure elucidation. In this case it revealed an interesting result, a common metabolite between an ant and a sponge, but it is not enviable in the amount of work required to determine the presence of this known compound.

Since then the tools involved in the dereplication workflow have improved. Commercialization of higher field magnets for NMR provides the benefit of improved sensitivity. Which in turn allows use of NMR as a dereplication step earlier in the workflow, when a compound might be present in an impure mixture, avoiding unnecessary purification steps. Similar improvements in MS analysis have come from both the development of softer ionization techniques and implementation of LCMS systems that use chromatography to separate compound. Which have likewise allowed implementation of MS earlier in the isolation workflow, and provided far more information than was available using only direct injection with electron impact ionization.

Regardless of these advances, few comprehensive approaches are available early in the workflow, affordable, available across laboratories, and easy to use. An ideal metabolite dereplication strategy would be broadly applicable to different sample types and laboratories, and usable with either unprocessed samples or at early steps in the workflow. There are alternatives to dereplicating the metabolites extracted from an organism, it is also possible to identify the producing organism and rely on previous characterization of its known metabolome in the literature.

The great difficulties in this process of dereplication are that producing organisms may be unknown or poorly studied in the literature, closely related organisms might produce very different compounds, or (as is very common in bacteria) it may be very difficult to rigorously characterize the species of interest. Constraining the discussion to filamentous

cyanobacteria these problems are very clear; it is well known that morphology is not tightly associated with phylogeny and closely related organisms can produce very different compounds [Engene et al., 2012]. To properly characterize the identity of a field collection of cyanobacteria requires a phylogenetic comparison of the 16S rRNA gene, and as a best practice, requires full genome sequencing.

However, the time, effort, and cost of sequencing a full genome from a field collection of cyanobacteria in most cases easily outstrips the comparative cost of doing an extraction, VLC purification and preliminary metabolome dereplication by MS or NMR. In this context, identifying a field collection by gene sequencing, while informative and necessary for publication of any analysis of the metabolome, is a poor candidate for a dereplication strategy. An idealized version of dereplication by organism identification should be fast, straightforward, and broadly applicable (as with morphological identification) but would also have a level of rigor in determining the relatedness of organisms (as with gene identification). There are different approaches possible, but phyloproteomics stands out as a candidate because it uses raw bacterial biomass with minimal processing, is widely available with commercially available platforms, and has sufficient accuracy as revealed by its clinical approval for the identification of pathogenic strains of bacteria [Lewtas and Young, 2015].

MALDI-based phyloproteomics is a well-established procedure for the identification of bacteria, and its used clinically to identify isolated pathogens. Filamentous marine cyanobacteria, however, are not important in a clinical setting, and have not been previously characterized in this manner. For our laboratory's drug discovery efforts, it would be very useful to have such a rapid and simple method for the identification of marine cyanobacteria. Relying on the close relationship of the proteome and the genome, phyloproteomics is essentially an inference that similar proteomes come from similar and related organisms. The sampling procedure for phyloproteomic analysis begins with direct protein extraction from an organism and then submission of that protein extract to MALDI-MS analysis. This

simple and efficient process makes it highly amenable to the workflow of dereplicating directly from a field collection or culture of cyanobacteria.

Unfortunately, efforts to profile the proteome of species within the genus *Moorea* (a good representative genus, and one of great interest to the Gerwick lab) were confounded by the polysaccharide sheath which surrounds the cyanobacterial filament. This material contaminated the protein extracts of *Moorea* and overwhelmed signals from the proteins when ionized on the MALDI-MS. As a result, one of the most interesting divisions of cyanobacteria, the sheathed filamentous strains, phyloproteomics could not be used as a dereplication strategy.

If the proteome could not readily be accessed to assign relatedness, and from there infer information about the secondary metabolome, then one of the first questions to be asked was: "Is there a way to directly assess and compare the secondary metabolomes of strains of interest?" MS fragmentation based 'molecular networking' analysis, which compares compounds directly based on their MS^2 fragmentation patterns, was explored as a method of profiling the metabolome of the Gerwick laboratory cyanobacterial collections. The initial experiments were quite successful. Not only were natural products known to be produced by a strain of *Moorea producens* observed in the network, but also observed were analogs of these compounds that the molecular networking algorithm sorted together into 'clusters'. This opened an avenue to both dereplicate the metabolomes of organisms as well as target novel analogs for a natural products discovery workflow.

3.3 Results and Discussion

3.3.1 MALDI-Phyloproteomics

MALDI-based phyloproteomics utilizes homology between the proteome to assign relatedness between strains. While the fact that closely related strains have highly similar

proteomes may be surprising at first, it is important to remember that the vast majority of the proteome is expressed constitutively, in the same "housekeeping" systems of primary metabolism that are used in genetics to canonically define phylogeny; specifically ribosomal proteins constitute a major source of the signals seen in the proteome extract [Murray, 2010, Sauer and Kliem, 2010]. Even in cases where strains produce different natural products, or have very well characterized mutations, those proteins tend to be a very small portion of the proteome. As such proteomic similarity can be used as an indicator of relatedness. Previous studies have shown that there are limits to this technique, for example, variation is seen based on culture conditions. While technical and biological reproducibility are high, allowing identification of species, intra-species strain resolution is often not feasible[Conway et al., 2001].

As a first effort to use this technique in our laboratory, literature procedures were employed in attempts to isolate the proteome of filamentous cyanobacterial strains from our culture collection. Unfortunately, the signal from the polysaccharide sheath surrounding the filaments of species in the genus *Moorea* overwhelmed signals from the proteome. As a control, there was little difficulty in collecting a proteome spectrum from *Escherichia coli*, a Gram-negative bacterium. This result was consistent with the explanation that the sheath and its high polysaccharide content were the problem. Optimization of the workflow was then carried out on a *Cyanobium sp.*, a single celled cyanobacterium, and led to a new method that isolated the proteome from this species. However, when applied to a *Moorea sp.*, the polysaccharide sheath still overwhelmed signals from the proteome.

In our workflow, bacteria are collected in the field, preserved in an alcohol/seawater mixture, treated with a lipophilic solvent extraction, and the metabolome is profiled by LCMS or NMR. Rarely the is 16S rRNA sequence used to identify a strain of interest before a determination has been made by another method that the strain is of interest for novel chemistry.

While using either phyloproteomics or 16S rRNA phylogenty to identify strains has the advantage of working directly with the cyanobacterial filament and takes place well before the majority of effort is put into the process of extracting and isolating compound. Phyloproteomics has the advantage of being faster and more easily scalable to dozens or hundreds of strains, thus potentially allowing an entire expedition's worth of extracts to be profiled before extraction.

There are many methodologies available in the literature to process samples for MALDI-phyloproteomics aw well as many software platforms available to process the resulting data. In this work the methods published by Freiwald and Sauer were used as a starting point to develop techniques for marine cyanobacteria [Freiwald and Sauer, 2009]. As with most work in the field of phyloproteomics, this method uses an acidic extraction of the lysed cell material into an organic solvent, and then this material is applied to a MALDI target plate and treated with MALDI matrix media.

To test this procedure a fresh sample of *Moorea producens* 3L was selected. An added difficulty of working with a Gram-positive bacterium, like *Moorea* is the presence of a robust cell wall which makes the cells harder to lyse, and in the case of *Moorea*, the additional polysaccharide sheath of the filaments. As such, the procedure had to be modified to incorporate either a manual pulverization of the filaments or a flash-freezing step where the filaments were frozen in N_2 (*l*) and then ground to a powder with a mortar and pestle to aid in breaking the cell wall and sheath apart. The extract was then acidified according to literature procedure with either trifluoroacetic acid (TFA) or formic acid (Figure 1).

The results of these "protein" extracts were disappointing, as they were dominated by a single major peak. Potential explanations for this included an instrument issue, a procedural error during the extraction, or an issue with the individual *Moorea producens* 3L culture. Experiments to tease apart which of these potential issues was responsible for the poor data quality began by making fresh technical replicates of extracts from *Moorea*

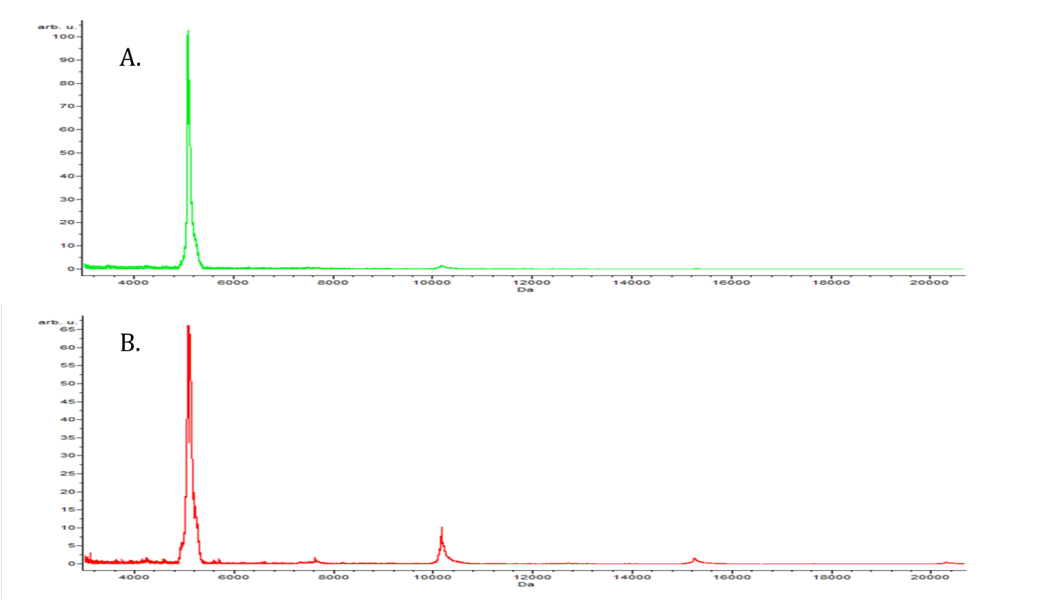


Figure 3.1: *MALDI-MS spectra of* Moorea producens *3L.* Prepared according to literature procedure and acidified with (A.) trifluoroacetic acid and (B.) formic acid.

producens 3L and *Moorea producens* JHB. These experiments showed the same result of poor protein coverage with the spectra dominated by a single major broad peak (Figure 2). This suggested that the root cause of this issue was not *Moorea producens* 3L specifically, or a single procedural problem occurring in a single extraction. A more broad instrument or procedural issue was ruled out by preparing an extract of *Escherichia coli* GC5 (without the additional N_2 (*l*) freezing-powderization step); this showed a good spectrum which hit to the library standard of *E. coli* in Biotyper.

A good result with a Gram-negative bacteria pointed towards an issue with the selected strains of the genus *Moorea*. If the modification of the literature procedure was not sufficient to lyse the cell wall of the bacterium, then this could explain the absence of strong protein signals in the extract. Using a free-living single celled cyanobacterium, *Cyanobium sp.*, as a test case, modifications of the procedure by Freiwald and Sauer[Freiwald and Sauer, 2009] were optimized to provide better coverage of the proteome from a Gram-positive bacterium. Eventually a modified version of the formic acid extract where both methanol and acetonitrile (rather that methanol alone) was found to provide a spectrum with multiple

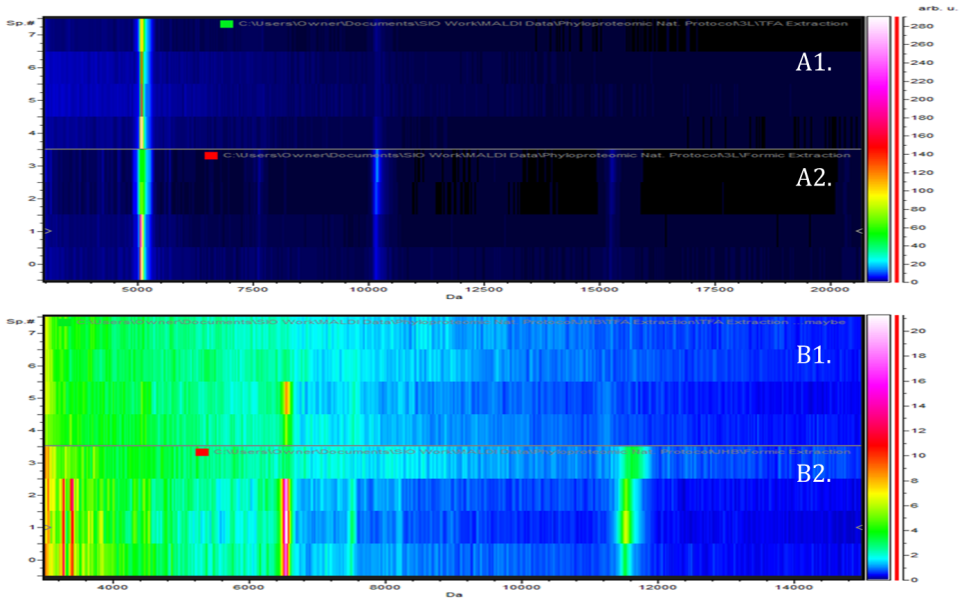


Figure 3.2: *MALDI-MS spectra of* Moorea producens *3L (A) and JHB (B).* Spectra for several technical replicates are presented in heat-map view. Samples were prepared with trifluoroacetic acid and formic acid (1 and 2, respectively).

protein signals in the desired mass window. An attempt to apply this new method to *Moorea*, however, resulted in little improvement to the spectrum, and this revealed that the issue was specific to the desired strains of filamentous cyanobacteria. The sheath polysaccharide, a major difference between *Cyanobium* and *Moorea*, is a likely explanation for why the spectra were unchanged. It appears that in both cases polysaccharide sheath material is abundant in the extract and preventing good ionization of the protein material.

Attempts to utilize direct plating of the bacteria and the MALDI plate instead of preparing a protein extract were also unsuccessful, suggesting that although MALDI-phyloproteomics is a useful tool to dereplicate the identity of strains, in the case of the sheathed filamentous cyanobacteria, it is not effective.

3.3.2 Molecular Networking

Given the unsuitability of 16S rRNA gene phylogeny or MALDI-phyloproteomics for profiling field collections of the filamentous marine cyanobacteria, it was clear that a

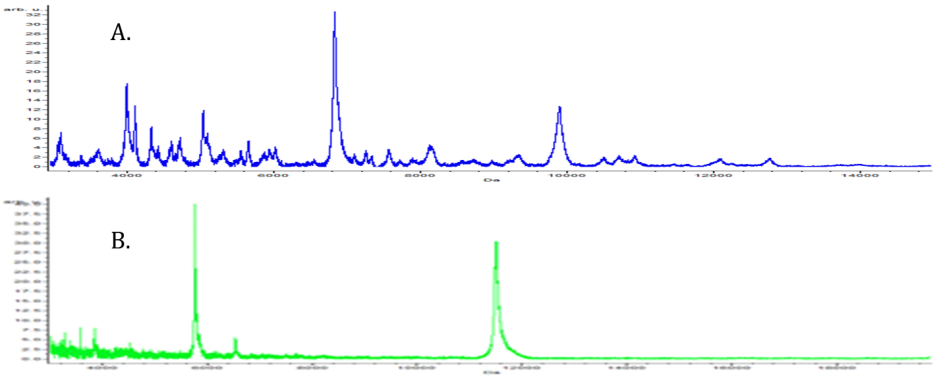


Figure 3.3: *MALDI-MS spectra of* Cyanobium *and* Moorea producens *JHB.* Extracts were prepared using the acetonitrile/methanol/formic acid protocol for both (A.) *Cyanobium* and (B.) *Moorea producens* 3L.

different approach was necessary. If no method was available to take a snapshot of the strains identity, then the necessary dereplication step would need to take place later in the workflow after the lipophilic extraction.

Ideally, this dereplication step should take place on the crude extract rather than requiring the additional time and resources that go into further purification steps. This of course is highly problematic because most dereplication techniques require more purified material, as in NMR analysis where clearly resolving the functional groups present in a sample is only possible if such features are detectable and discernible. In a crude extract, bioactive compounds are often present in minute amounts, while primary metabolites such as fats and chlorophyll derivatives are abundant. This results in signals from desirable compounds that are either overlapped with other metabolites or simply below the limit of detection.

As an alternative, LCMS analysis is a good choice for profiling crude extracts because the LC component allows for compound separation as part of the analysis and MS is suitable to the detection of major as well as trace constituents. There are many tools by which dereplicate compounds observed in a MS chromatogram, not the least of which is manual analysis of molecular formulas compared to the literature. This is a highly

effective tool when a high quality spectrum HRMS spectrum has provided a good prediction of the molecular formula, and other features, such as halogen isotope distribution can be informative in a similar manner. Difficulties arise when only LRMS data is available, resulting in molecular formula predictions that are less accurate. Even when using a HRMS system, there can be difficulties in profiling a crude extract because the effective dynamic range means that the most abundant compounds may saturate the detector, preventing the collection of HRMS data for both the most abundant as well as the most minor compounds.

Putting aside all considerations of data collection, it is of course also very labor intensive to dereplicate each signal manually. Molecular networking is a great alternative because instead of relying on MS^1 derived molecular formula or halogenation pattern, it relies on an automated MS^2 spectra comparison. Using this platform is advantageous because it is automated, is just as robust with LRMS as with HRMS data, and because the MS^2 fragmentation pattern represented in an MS^2 spectrum a proxy for structural relatedness. This is because bonds do not break randomly but rather, they fragment based on the dynamics of the parent molecule's internal bond strength and strain.

It is also possible to compare spectra of known compounds to crude extracts. This feature is key as a dereplication workflow because it allows ready comparison of known metabolites with otherwise unknown extracts, the ideal situation for a dereplication approach. To verify that molecular networking would be well suited to filamentous cyanobacterial strains, *Moorea bouillonii* PNG05-198 was selected as a candidate strain to profile by molecular networking because it is a known producer of lyngbyabellin A (**86**) and apratoxin A (**87**) and B (**88**) [Luesch et al., 2000, Luesch et al., 2001b, Luesch et al., 2002, Engene et al., 2012].

Using standard spectra for the known compounds **86** and **87**, a molecular network could be generated comparing the crude extract and these standards. This network showed in both cases a cluster of parent masses around the nodes coming from the $[M+H]^+$ of each

compound. These parent masses can come from source fragments of the major compound such as a loss of a water fragment, analogs structures such as **88**, or molecules which incorporate a heavy isotope such as with the heavy ^{37}Cl-bearing molecules of **86**.

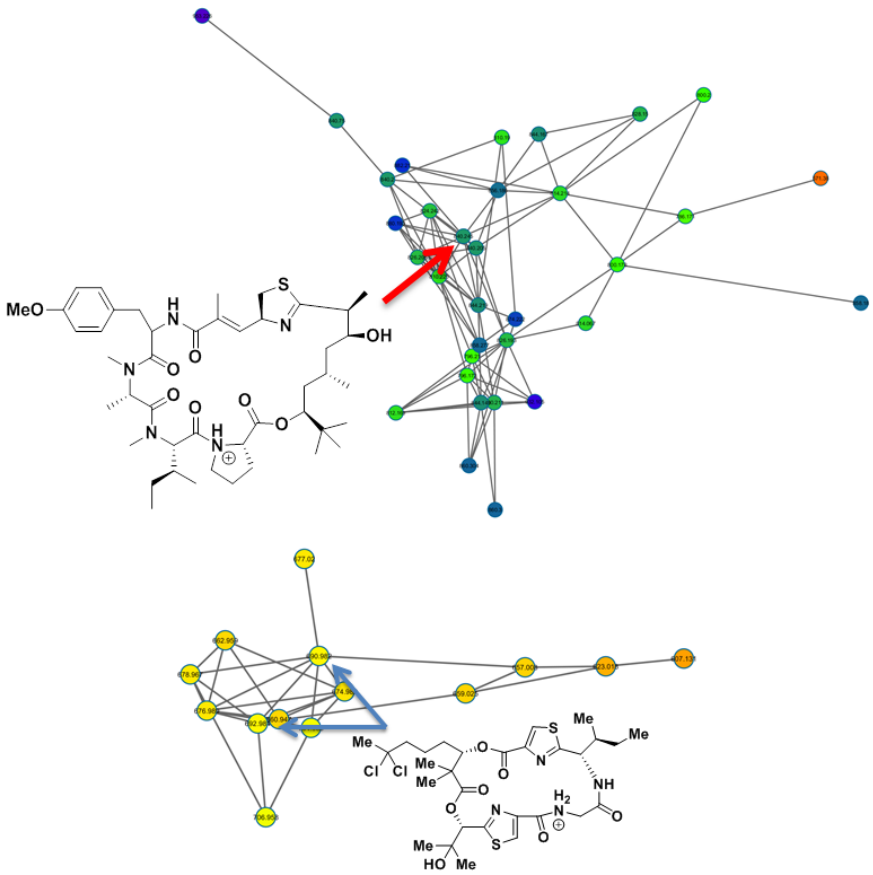


Figure 3.4: *Molecular networking clusters containing the major compounds lyngbyabellin A (**86**) and apratoxin A (**87**) from* Moorea bouillonii *PNG05-198.* Lyngbyabellin A (**86**) is present as a pair of nodes from the presence ions bearing the heavy ^{37}Cl atom.

The identification of these compounds from the crude extract using this workflow is very efficient and effective. It is worthwhile to point out the advantages of this strategy over alternatives. Unlike dereplication by NMR, no purification is necessary to arrive at a mixture sufficiently pure to deduce the presence of known compounds. The crude extract was neither fractionated by VLC or SPE. While it is likely that a HR-LCMS system could have provided

m/z values accurate enough to identify the molecular formulas of **86** and **87**, this data was collected on a LR-LCMS system. In molecular networking the MS^2 fragmentation provides an orthogonal analysis to the comparison of parent masses, no matter the accuracy with which those masses are collected, thus allowing effective comparisons and providing a highly effective dereplication platform.

3.4 Conclusions

The value of dereplication in a natural products workflow comes from circumventing steps in the discovery process which by necessity require effort in both researcher hours and resources involved in each step of purification. An idealized dereplication workflow would involve no cost of its own and would avoid all isolation/purification steps. This is not an impossible goal; for example, surveying of marine macroscopic algae armed only with a good reference book and knowledge of the literature, a talented researcher can identify a sample of red algae and predict the contents of its metabolome with high accuracy. Sadly, with bacteria, fungi, and poorly studied macroscopic organisms, neither the literature knowledge-base or the firm connection between morphology and taxonomy exists.

Accessing taxanomic identity is possible for bacteria, using 16S rRNA gene phylogeny or by MALDI-based phyloproteomics; however, this requires more cost and effort than a visual identification, but accesses taxanomic relatedness far before the extraction and isolation steps required for traditional methodologies of dereplication. Confounding factors when using filamentous cyanobacteria prevented the use of MALDI-based phyloproteomics, and 16S rRNA gene phylogeny is undesirable in terms of both cost and time investments. Molecular networking provides the ability to rapidly and reliably profile an unknonwn metabolome against standard libraries, and with minimal investment in an extraction and isolation workflow. This is an advantageous approach which other users have relied upon [Yang et al., 2013] and in a diverse set of cases, and with the recent creation of an online

platform for molecular networking [GNP, 2015] researchers worldwide will be able to access 'molecular networking' as a dereplication tool.

3.5 Materials and Methods

3.5.1 Protein Extraction and MALDI-Phyloproteomics

Initial attempts to profile the *Moorea* extract relied on the method of developed by Freiwald and Sauer [Freiwald and Sauer, 2009] directly using only the modification of manually pulverizing the filaments before extraction by glass pipette. A later attempt used an initial step that flash froze the filaments in N_2 (*l*) then pulverized them by mortar and pestle before proceeding with the extraction. Neither attempt saw good ionization for the extracted proteome.

To develop conditions using *Cyanobium* sp., samples were taken from the SW-BG11 media culture of *Cyanobium* sp. growing in the culture room in six 1200 μL aliquots. These samples were centrifuged in microcentrifuge tubes at 10,000 rpm for 5 min. The pellet weighed between 1 and 6 mg, this material was then extracted with either a formic acid or TFA protocol.

For the formic acid extractions the pellet was first washed with 1200 μL of 3:1 EtOH/H_2O by mixing with pipette then centrifuged twice at 10,000 rpm for 2 min and the supernatant was discarded. The washed pellet was treated with one of three extraction protocols: 1) 10 μL of 70% formic acid (*aq*) was added then mixed by pipette, then 10 μL of CH_3CN was added. 2) 10 μL of 70% formic acid (*aq*) was added then mixed by pipette, then 10 μL of CH_3CN adn 10 μL of MeOH were added. 3) 10 μL of 70% formic acid (*aq*) was added then mixed by pipette, then 10 μL of MeOH was added. Each of the extracts was then centrifuged at 10,000 rpm for 5 min and the protein extract supernatant was removed.

For the TFA extractions the pellet was treated with 80% TFA (*aq*) and mixed by

pipette then let stand for 30 min, then treated with one of three extraction protocols: 1) 150 μL of H_2O and 200 μL of CH_3CN were added. 2) 50 μL of MeOH and 50 μL of CH_3CN were added. 3) 50 μL of MeOH was added. Each of the extracts was then centrifuged at 10,000 rpm for 5 min and the protein extract supernatant was removed.

The exracts were then plated onto a steel MALDI target plate with 1 μL of sample and 1 μL of 35 mg/mL universal MALDI matric in 50% CH_3CN/2.5% TFA (*aq*).

The final extraction of *Moorea producens* JHB was made by taking a small sample of filaments flash freezing with N_2 (*l*) and pulverized with a mortar and pestle. This material was suspended in 1.0 mL of H_2O and a 100 μL portion was transfered to a microcentrifuge tube, centrifuged at 10,000 rpm for 3 min and the supernatant was removed to afford a pellet of 4.4 mg wet biomass. This was washed with 1200 μL of 3:1 H_2O/EtOH, vortexed, and centrifuged twice as before removing the supernatant each time. The washed pellet was then air dried for ca. 1 min and treated with 10 μL of 70% formic acid (*aq*), 10 μL of CH_3CN, and 10 μL of MeOH was added. The mixture was vortexed then centrifuged as before, the supernatant was collected and plated as before.

3.5.2 Lipophilic Extraction for Molecular Networking

A preserved collection of *Moorea bouillonii* growning in the culture room was extracted iteratively with 2:1 dichloromethane/methanol, the first extract with mild warming $\leq$35 °C. The ethanol-water mixture the material cyanobacteria had been stored in was concentrated to reduced volume by rotovap then extracted by phase separation iteratively into dichloromethane. All organic layers were combined and dried to afford the crude cyanobacterial extract as a dark-green solid.

3.5.3 MS Data Collection and Molecular Network Generation

The crude extract was prepared at a concentration of 1.0 mg/mL in 1:1 CH_3CN/H_2O and 1:1 MeOH/HO, and data was collected on a Thermo Finnigan LCQ Advantage Max mass spectrometer. Ran 10 μL of sample with a gradient starting with a 5 min hold at 1:1 CH_3CN/H_2O (water acidified at 0.1% with formic acid) to CH_3CN over 25 min with another 5 min hold, then reequilibrated by bringing elution back to 1:1 over 5 min with a final 5 min hold. The column used was a Phenomenex Kinetex C-18 100 Å 100 x 4.60 mm column with a 0.700 mL/min flow rate. Collected four scan events, the first a positive MS^1 scan over a 300-2000 *m/z* window the rest MS^2 scans of the first three most intense ions using dynamic exclusion. Resulting MS files were converted to .mzXML format and ran with the Spectral Networking algorithm [Watrous et al., 2012, GNP, 2015].

3.6 Acknowledgments

I thank E. Esquenazi, S. Mascuch, V. Phelan, C. Rath, and J. Watrous for their help in learning MALDI-MS techniques and analysis.

This chapter, in part, is a reprint, with permission, of the material as it appears in the Journal of Natural Products, Yang, Jane Y.; Sanchez, Laura M.; Rath, Christopher M.; Liu, Xueting; Boudreau, Paul D.; Bruns, Nicole; Glukhov, Evgenia; Wodtke, Anne; de Felicio, Rafael; Fenner, Amanda; Wong, Weng Ruh; Linington, Roger G.; Zhang, Lixin; Debonsi, Hosana M.; Gerwick, William H.; Dorrestein, Pieter C. volume 76, pages 1686-1699.

Chapter 4

Expanding the Described Metabolome of the Marine Cyanobacterium Moorea producens JHB through Orthogonal Natural Products Workflows

4.1 Abstract

Moorea producens JHB, a Jamaican strain of tropical filamentous marine cyanobacteria, has been extensively studied by traditional natural products techniques. These previous bioassay and structure guided isolations led to the discovery of two exciting classes of natural products, hectochlorin (**89**) [Marquez et al., 2002] and jamaicamides A (**90**) and B (**91**) [Edwards et al., 2004]. In the current study, mass spectrometry-based 'molecular networking' was used to visualize the metabolome of *Moorea producens* JHB, and both guided and enhanced the isolation workflow, revealing additional metabolites in these compound classes. Further, we developed additional insight into the metabolic capabilities

of this strain by genome sequencing analysis, which subsequently led to the isolation of a compound unrelated to the jamaicamide and hectochlorin families. Another approach involved stimulation of the biosynthesis of a minor jamaicamide metabolite by cultivation in modified media, and provided insights about the underlying biosynthetic machinery as well as preliminary structure-activity information within this structure class. This study demonstrated that these orthogonal approaches are complementary and enrich secondary metabolomic coverage even in an extensively studied bacterial strain.

4.2 Introduction

Moorea producens JHB is a strain of tropical filamentous marine cyanobacterium that has been in culture in our laboratory for nearly two decades since its collection from Hector's Bay, Jamaica in 1996. We have extensively studied its natural products by traditional isolation techniques [Marquez et al., 2002, Edwards et al., 2004]. An NMR-guided process led to the isolation and structure elucidation of **89**, a highly potent cytotoxin which enhances actin polymerization [Marquez et al., 2002]. In separate work, sodium channel blocking and fish toxicity assays guided the isolation and discovery of **90** and **91** [Edwards et al., 2004]. Since then, this strain has been extensively studied (under its name before reclassification *Lyngbya majuscula* JHB), with cultures from our laboratory being the subject of five more publications [Marquez et al., 2002, Edwards et al., 2004, Dorrestein et al., 2006, Ramaswamy et al., 2007, Gu et al., 2009, Esquenazi et al., 2011, Engene et al., 2012]. In this work we present the application of new methods for natural product discovery being applied to study the metabolome of *M. producens* JHB, from which several novel compounds, not observed in previous studies of this strain, were discovered (Figure 1).

Recent developments in the field of natural products have both improved the traditional activity/structure-based isolation workflows, and provided other orthogonal techniques to profile the metabolomes of organisms of interest. Traditional activity/structure-based

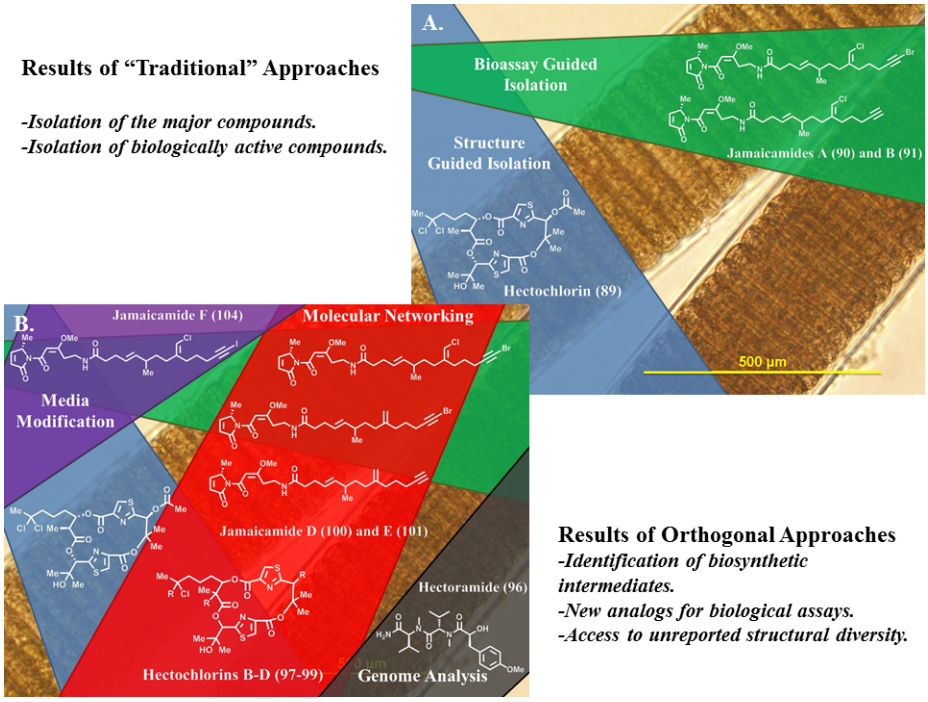


Figure 4.1: *The metabolome of* M. producens *JHB from prior studies (A.); and the expanded metabolome of* M. producens *JHB from approaches described in this study (B.).*

isolation procedures have been supplemented by the use of solvent extractors [Johnson et al., 2010], improved HPLC systems and columns to improve compound yields [Hayes et al., 2014], new high-throughput screening techniques [Hou et al., 2012]; or analytical techniques that make it possible to determine the structures of ever smaller quantities of compounds, such as more sensitive cryo-probe NMR systems and new NMR pulse sequences [Kupče, 2003, Molinski, 2010]. These improvements have enabled nanomole-scale structure elucidations such as were used to characterize **6** and **7** from the marine nudibranch *Hexabranchus sanguineus* [Dalisay et al., 2009]. Concurrently, the field of natural products is exploring novel methods to assess secondary metabolomes that are distinct from traditional isolation workflows, chief among them is genome mining [Challis, 2008]. As genome sequencing has become more rapid and less expensive, and the pipeline for assembling and annotating secondary metabolite pathways from gene sequence information, using programs such as antiSMASH and NaPDos, has become more efficient [Medema et al., 2011, Blin et al.,

2013, Gross et al., 2007, Ziemert et al., 2012, ant, 2015]. Consequently, applications of genome mining to the field of natural products have become increasingly diversified [Challis, 2008, Jensen et al., 2014, Udwary et al., 2007]. Genome mining has also shown that many strains possess far more biosynthetic gene clusters than previously expected, and indeed, these 'silent' pathways in some cases constitute the preponderance of a strain's biosynthetic capacity [Jensen et al., 2014, Udwary et al., 2007, Liu et al., 2013]. Work with the daptomycin (**92**) producing strain of *Streptomyces roseosporus* showed that it has the capacity to produce three other Non-Ribosomal Peptide Synthetase (NRPS) products, namely arylomycin (**93**), napsamycin (**94**), and stenothricin (**95**), all previously unreported from this organism [Liu et al., 2013].

Even with these impressive enhancements to our isolation methodologies, significant challenges remain to fully characterize the metabolic potential of an organism. Any single workflow has drawbacks; bioassay-guided isolation methods are inherently biased and overlook metabolites with differing activity. Structure-guided isolation schemes can miss compounds that are produced at low concentration and thus escape detection by normal methods, or if they have unremarkable spectroscopic properties. In addition, it is never known whether an organism is expressing all of its secondary metabolite biosynthetic pathways under a given set of environmental or culture conditions [Jones et al., 2011]. While a successful genome mining effort might fully characterize the biosynthetic pathways within an organism, connecting and coordinating this with an isolation workflow can be challenging [Terpe, 2006]. Without heterologous expression or knockout experiments, links between a biosynthetic pathway and a structure are often only tentative.

In our decision to reexamine the metabolome of *Moorea producens* JHB we were mindful of these challenges, but also of the many examples in the literature where novel techniques found previously unreported compounds from productive well studied strains. Since the first report from *M. producens* JHB, our laboratory has gained access to new

assays, such as modulation of cathepsin L activity [Miller et al., 2014], and new facilities on the UCSD campus for high-field NMR and High-Resolution Mass Spectrometry (HRMS) [Ska, 2015]. However, because **89** and **90** are the major secondary metabolites from this strain, we knew that any additional compounds would be minor metabolites, and hence, alternative techniques that could target these compounds were needed.

Mass spectrometry based-molecular networking is well suited to serve as an alternative metabolic profiling platform because it is highly sensitive, amenable to use with compound or culture libraries, and blind to traditional guideposts for isolation projects such as chromatographic retention time or halogen isotope patterns [Watrous et al., 2012, Yang et al., 2013]. Mass spectrometry is highly sensitive, and while no ionization technique is universal, electrospray ionization (ESI) effectively ionizes a wide range of structural classes providing good coverage of the metabolome [Fenn, 2003].

Molecular networking utilizes MS^2 data to sort parent ions based on their structural similarity. Secondary ion mass fragmentation data relate directly to molecular structure because chemical bonds break on the basis of bond strength, strain within a molecule, and ability of a fragment to stabilize charge [Watrous et al., 2012]. Fragmentation patterns are thus intimately related to molecular structure, but independent of other bases for assessing compound similarity, such as the parent ion mass, halogen isotope pattern, or LCMS retention time. The Spectral Networking algorithm normalizes the intensity of fragment ions, uses each as an independent axis to construct a multidimensional vector for each spectrum, and finally compares the similarity of these vectors using a cosine function [Watrous et al., 2012]. This cosine score is then used to plot the relationships between different parent ion masses with the open source software Cytoscape (Figure 2) [Cyt, 2015].

For *M. producens* JHB, a crude extract of the cyanobacterium was profiled by LCMS and data from these runs were constructed into a molecular network as described above. This network identified an array of previously unreported compounds that were related to

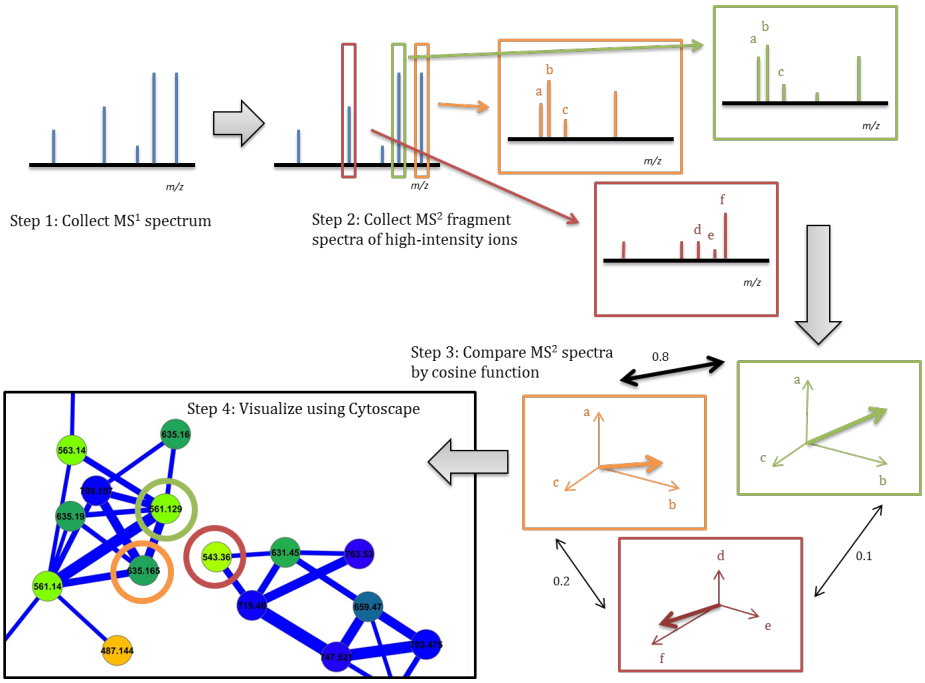


Figure 4.2: *Visualization of MS data with molecular networking.* The process of forming a molecular network begins with collection of MS^1 spectra (Step 1), from which parent ions are selected for MS^2 fragmentation (Step 2). After the data is acquired it is processed by the Spectral Networking algorithm which converts MS^2 data to vectors. The minimum number of peaks required to construct a vector is six; in this figure it is represented as only three so that the vector could be presented visually. It is important to note that when the multi-dimensional vectors are compared against all others by cosine function the vectors can have hundreds of dimensions (Step 3). This output is then visualized in Cytoscape with each node representing a MS^2 spectrum labeled and colored based on its parent mass (Step 4).

the major metabolites **89** and **90**, likely produced as intermediates along the biosynthetic pathway, or by promiscuity in the substrate binding sites of these modular-type biosynthetic enzymes. These analogs provide new insights into the biosynthetic machinery and a more thorough understanding of the true secondary metabolic capacity of *M. producens* JHB than accomplished in previous work.

Additionally, further insight into the biosynthesis of **90** was gained by growing *M. producens* JHB in seawater media supplemented with sodium iodide. Under these conditions a minor metabolite of the jamaicamide family was produced in large quantities, and allowed the isolation and characterization of the iodinated analog of **90**. This also provided insight

into the halogenation enzyme involved in this biosynthetic step as well as yielding sufficient compound for biological assays.

The two approaches described above revealed new compounds within the jamaicamide and hectochlorin families, products of the two biosynthetic gene clusters previously identified in this organism [Edwards et al., 2004, Ramaswamy et al., 2007, Gu et al., 2009]. To assess whether *M. producens* JHB possessed a broader biosynthetic capacity, we complemented the molecular networking approach with an in-depth bioinformatics interrogation of the sequenced genome. This led to the detection of numerous short NRPS pathways that were predicted by anti-SMASH to encode for small polar peptides, as well as one somewhat longer pathway for a predicted water soluble compound, the structure of which will be presented in a separate publication. An effort to isolate the products of these shorter NRPS pathways led to the isolation and characterization of a new compound, hectoramide (**96**), containing two N-methyl valine residues and a modified tyrosine residue. The structure of this new peptide was determined by 2D-NMR and HRMS analysis, and configuration of the residues by hydrolysis and chiral chromatographic analysis.

4.3 Results and Discussion

4.3.1 Molecular networking of *M. producens* JHB

The MS data from the crude extract LCMS chromatograms of *M. producens* JHB were analyzed using the Spectral Networking tool to create a molecular network and visualized in Cytoscape (Figure B.1 in Appendix) [Watrous et al., 2012, Cyt, 2015]. These data were also uploaded into the Global Natural Products Social Molecular Networking database where they are publicly available (Massive ID MSV000078990) [GNP, 2015].

The network was interrogated for the $[M+H]^+$ ions from the known major metabolites from this cyanobacterial strain, **89** and **90**. In both cases, this revealed a cluster of

parent masses from related analogs in the same family of compounds (Figures 3 and 4). One important caveat of the molecular networking process is that a single compound can give rise to multiple nodes. For example ions deriving from different adducts, from differing isotope composition, as well as source fragments, are all observed in the molecular network when present in sufficient intensity for the spectrometer to collect their MS^2 spectra. The hectochlorin cluster is clearly an illustration of a single molecule producing multiple nodes from different halogen isotopes. The presence of heavy chlorine atoms in some of the hectochlorin molecules results in three nodes with mass differences of two Daltons within the hectochlorin cluster (Figure 3).

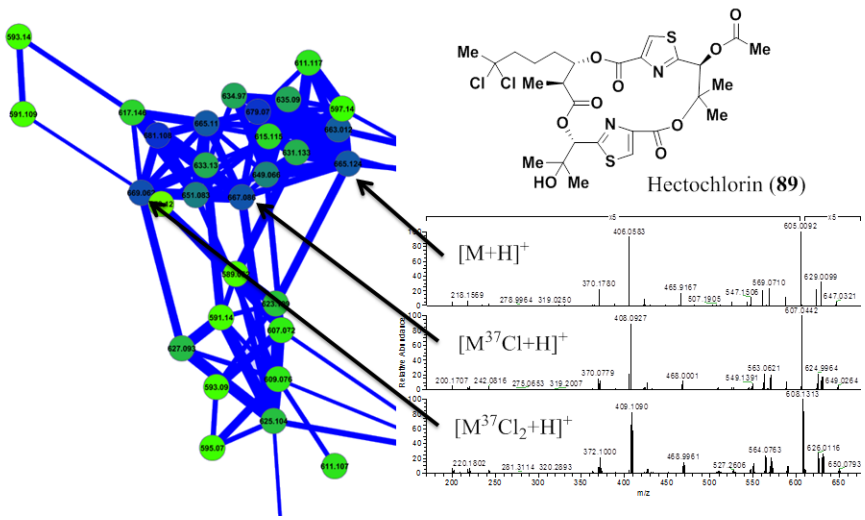


Figure 4.3: *Hectochlorin (**89**) isotope pattern within the* M. producens *JHB network.* The presence of ^{35}Cl or ^{37}Cl within specific hectochlorin molecules yielded different parent masses and fragment spectra. The species with both 35Cl atoms has an *m/z* of 665, the species with one ^{35}Cl and one ^{37}Cl atom has and *m/z* of 667, and the species with ^{37}Cl atoms has an *m/z* of 669. Because of this the fragment spectra share only masses from those fragments without chlorine atoms, and those fragments bearing the chlorine atoms show the same mass differences as their parent masses.

The hectochlorin cluster was also interrogated for new hectochlorin analogs; many of

the masses that clustered with the nodes from compound **89** were consistent with hypothetical analogs. The LCMS trace showed these species had distinct retention times from that of **89** (i.e. they were not source fragments), as well as isotope patterns that matched their variable levels of halogenation. Using the structure of **89** as a template, a series of hypothetical compounds could be proposed which were consistent with these masses (Table 1).

Table 4.1: *Analogs of hectochlorin (**89**) deduced from the molecular network.*

Compound Name and Number	Structural Features	Halogenation Pattern	Retention Time	Masses Observed in Network
Hectochlorin (**89**)	Original Compound	Dichloro	23.8 min	665/667/669
Hectochlorin B (**97**)	Deacyl-hectochlorin	Dichloro	20.3 min	623/625
Hectochlorin C (**98**)	Dechloro-hectochlorin	Monochloro	22.6 min	631/633
Hectochlorin D (**99**)	Methyl-hectochlorin	Dichloro	25.8 min	679/681

The network was also examined for new analogs within the jamaicamide cluster. Both **90** and **91** were present as a pair of $[M+H]^+$ ions separated by 2 Da, again due to incorporation of heavy halogen atoms (Cl, Br). Another pair of nodes, with parent masses also separated by 2 Da, clustered with high cosine scores to the known jamaicamides. The LCMS[1] chromatogram showed that this pair of nodes possessed a distinct retention time, and that it had an isotopic composition consistent with a singly brominated but non-chlorinated species. Thus, a dechloro-analog of **90**, jamaicamide D (**100**), was hypothesized to explain this pair of connected nodes and was supported by HRMS data [*m/z* 533.1983, retention time 29.9 min, $C_{27}H_{38}BrN_2O_4$ $[M+H]^+$ calculated 533.2009, -2.6 mamu].

The biosynthetic gene cluster and many of the individual biosynthetic steps for **90** are known, and previous work in our laboratory has suggested that **91** is the precursor for **90** with bromination occurring as the final step in the pathway [Edwards et al., 2004, Dorrestein et al., 2006, Gu et al., 2009, Esquenazi et al., 2011]. In this scenario, **100** would be formed by bromination of a non-chlorinated precursor (Figure B.27 in Appendix). Further inspection of the network revealed a mass consistent with such a non-halogenated species, termed jamaicamide E (**101**), also clustered in the jamaicamide family (Figure 4). The LCMS[1] trace again showed a distinctive retention time for **101**, and an HRMS formula supportive

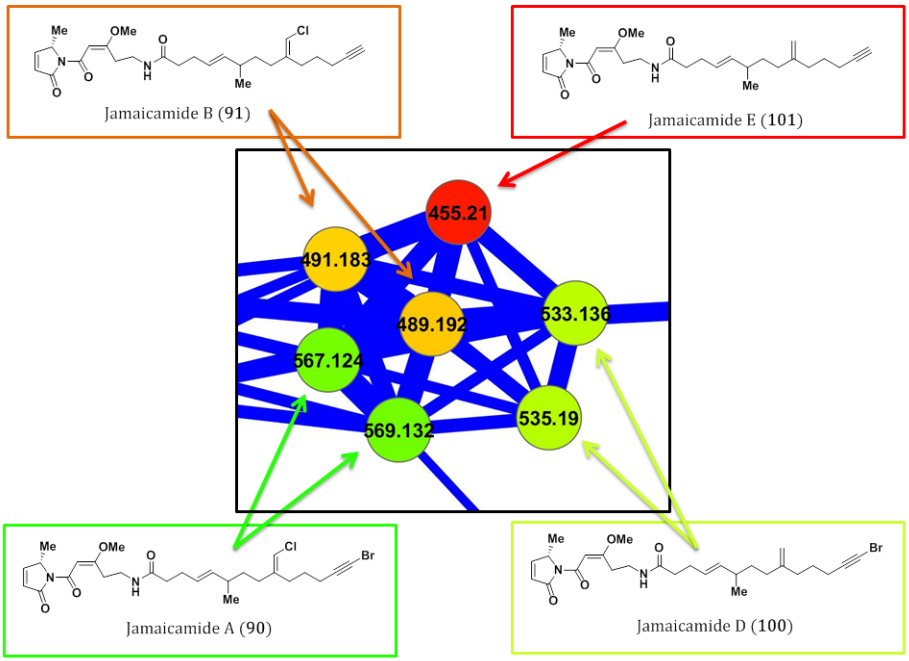


Figure 4.4: *The jamaicamide cluster of nodes within the molecular network of* Moorea producens *JHB.* As with the hectochlorins, shown in Figure 3, the jamaicamides are present in the network as multiple parent masses from the presence of heavy halogen atoms.

of this structure [*m/z* 455.2877, retention time 26.5 min, $C_{27}H_{39}N_2O_4$ $[M+H]^+$ calculated 455.2904, -2.7 mamu].

4.3.2 Purification of analogs from the Jamaicamide and Hectochlorin families

The LCMS analysis of this cyanobacterial extract used a serial set of five scans; one high-resolution, one low-resolution, and then three LTQ-MS^2 scans of the three most intense ions from the previous scan using dynamic exclusion. Low-resolution LTQ scans were used to generate the MS^2 data for molecular networking because of the higher sensitivity of this mode. This analysis generated MS^1 and MS^2 data consistent with the proposed structures for the jamaicamide and hectochlorin analogs. To rigorously characterize these new compounds, an extract that had been preparatively fractionated by Vacuum Liquid Chromatography

(VLC) was interrogated for analogs. Using the standard system for VLC in our laboratory uses fixed volumes of progressively more polar mixtures of hexanesEtOAcMeOH proportional to the size of the extraction. (nine fractions, 100% hexanes, 10% EtOAc/hexanes, 20% EtOAc/hexanes, then 20% increments to 100% EtOAc, 25% MeOH/EtOAc, and 100% MeOH). By LCMS analysis, fractions eluting with 80% EtOAc/hexanes and 100% EtOAc contained hectochlorin and the jamaicamide metabolites. These two VLC fractions were combined and further purified by Reverse-Phase Solid-Phase Extraction (RP-SPE) and RP-HPLC. The two VLC fractions eluting in 25% methanol to ethyl acetate, and in 100% methanol, were found to contain additional hectochlorin analog masses by LCMS, and were combined and purified in the same fashion as the less polar fractions.

These HPLC purifications afforded pure **89**, **90**, **91**, and several of the minor analogs proposed by molecular networking analysis of the crude extracts. The major metabolites **89** and **90**, along with the new derivatives **100** and **97**, were isolated in sufficient quantity to fully characterize by NMR spectroscopy. However, not all of the species identified from the network were present in isolable amounts, as in the case of **101**. To better characterize these other more minor analogs, they were subjected to $HRMS^2$ fragmentation analysis, and then assembled de novo using the previously reported cyclic peptide dereplication tool [Ng et al., 2009].

The observation that **97**, the deacetyl-analog, was the most abundant analog of compound **89**, is consistent with the conjecture that this is an intermediate in the biosynthesis of **89** [Ramaswamy et al., 2007]. The gene cluster encoding for hectochlorin biosynthesis shows no evidence of an acylation enzyme in the NRPS domain responsible for constructing the cyclic core of the molecule [Ramaswamy et al., 2007]. Indeed, in the original report on the biosynthetic gene cluster, it was noted that the *hct* gene cluster lacked an acylation enzyme; it was proposed that acylation occurs as a post-NRPS modification [Ramaswamy et al., 2007]. Moreover, **97** was previously reported along with **1** from the sea hare *Bursatella*

leachii. In that report the authors suggested that accumulation of **97** in *B. leachii* could be the result of metabolism of **89** by the sea hare [Suntornchashwej et al., 2005]. With direct isolation of **97** from the cyanobacterial source, however, it is clear that direct accumulation from its diet is also a possibility.

At the time of its isolation, **97** was found to be a more potent cytotoxin than **1** against the human carcinoma of the nasopharynx (KB) and human small cell lung cancer (NCI-H187) cell lines (LD_{50}'s of 0.31 μM and 0.32 μM for **97**, compared to 0.86 M and 1.20 μM for **89**) [Suntornchashwej et al., 2005]. Molecular networking groups compounds based on their similar structural frameworks irrespective of their abundance, polarity, or other characteristics such as halogenation pattern or biological activity. Therefore, metabolites with significantly different polarities within an extract can still be easily recognized as structural analogs by molecular networking. The majority of compound **97** was found in the last two and most polar VLC fractions, which are often complex mixtures with poorly soluble nuisance compounds such as salts and glycolipids, potentially masking compounds of interest. By MS-based profiling and then targeting the isolation process towards the isolation of a suite of structurally similar products, rather than a single compound, information about structure activity relationships can be developed, as well as insights gained into the biosynthetic process.

Besides **97**, the other analogs of **89** detected in this investigation are likely biosynthetic shunt products. They may be produced by errors in loading of the biosynthetic substrates or in the reactions the biosynthetic enzymes catalyze. For example, the monochlorinated analog of **89**, **98**, is likely produced by a process wherein the chlorination enzyme fails to catalyze the addition of a second chlorine atom to the substrate, and this modification is transparent to subsequent downstream biosynthetic steps. However, the mono-chloro group in **98** is curious because formation of the trichloromethyl group of barbamide (**102**) in a different strain of M. producens, which is catalyzed by a highly homologous enzyme

[Galonić et al., 2006, Chang et al., 2002, Flatt et al., 2006], occurs without going through a discrete monochlorinated intermediate. Rather, in the case of **102**, the biosynthesis proceeds through an initial dichlorination step catalyzed by the BarB2 halogenase protein, followed by a monochlorination step catalyzed by the BarB1 halogenase protein to yield the final trichlorinated product. The isolation of **98** reveals that the halogenation enzyme in hectochlorin biosynthesis is different from the barbamide halogenases, as neither barbamide enzyme was observed to catalyze the formation of a discrete monochlorinated species in a detectable yield, and as the barbamide proteins act on the less reactive terminal methyl species [Galonić et al., 2006, Chang et al., 2002, Flatt et al., 2006].

4.3.3 Genomic Insights into *M. producens* JHB Natural Products

Genome sequence analysis of *M. producens* JHB revealed the known biosynthetic pathways of hectochlorin and the jamaicamides [Edwards et al., 2004, Ramaswamy et al., 2007, Gu et al., 2009]. Other numerous short biosynthetic pathways were annotated by antiSMASH, and many of the predicted products of these pathways were dipeptides (Figure 5). Based on their predicted structures, we posited that if produced these compounds would be present in the more polar VLC fractions. A careful search of the metabolome contained in the polar fractions by LCMS analysis showed the dominant compound in the most polar eluting fractions was the hectochlorin analog **100**. However, during purification by RP-SPE followed by RP-HPLC, not only was **100** isolated, but also a novel compound, named here as **96**, which was thoroughly characterized by NMR, HRMS and chemical degradation to components that could be compared with standards.

Analysis of the TOtal Corrolation SpectroscopY (TOCSY) spectrum for compound **96** revealed four spin systems. Two of these were valine residues, seen by correlations of two doublet methyl peaks to their respective β-proton multiplets, and from there to their α-protons. Both of these residues were shown to be *N*-methyl valine residues by reciprocal

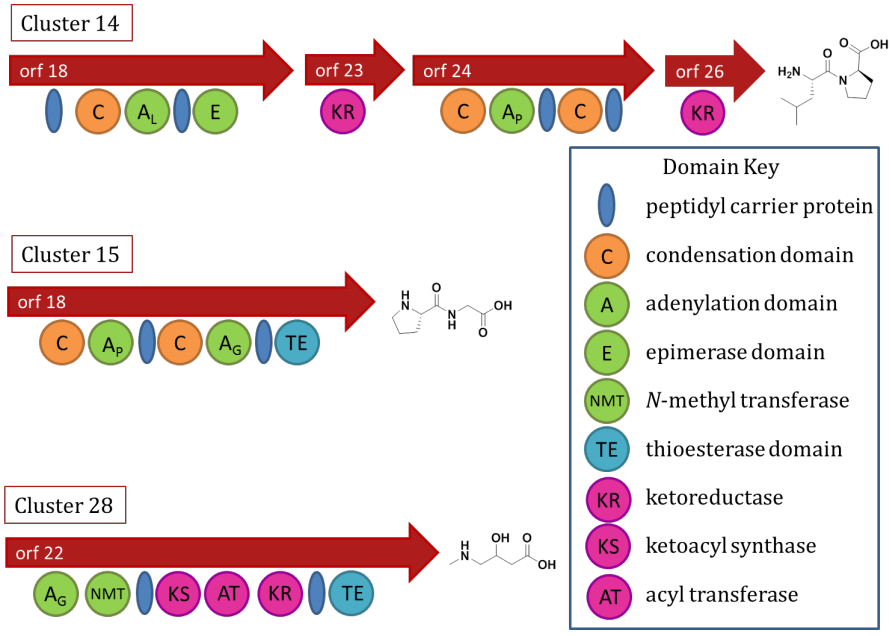


Figure 4.5: *Three NRPS or NRPS/PKS-type biosynthetic gene clusters from the* Moorea producens *JHB genome, annotated by antiSMASH for predicted products.*

Heteronuclear Multiple Bond Correlation (HMBC) correlations between the *N*-methyl singlets and the α-carbons. The presence of a para-substituted aromatic ring was revealed by a COSY correlation between two doublet 2H aromatic peaks; one of the substituents was revealed to be an *O*-methyl by HMBC signals between the singlet methyl protons at δ 3.79 and an aromatic ring carbon (C-19). The final COSY spin system connected an α-proton and a β-methylene group; the latter showed HMBC correlations to several of the aromatic ring carbons, and thus explained the other ring substituent. The downfield shift of the α-carbon (δ 69.9) and α-proton (δ 4.52) indicated this position was substituted with a hydroxy group rather than an amino group, revealing the final residue to be a 3-(4-methoxyphenyl)lactic acid (Mpla).

The linkage of these residues was determined by 2D-NMR and MS^2 fragmentation analysis. The *N*-methyl group of the first valine showed a correlation to the carbonyl of the second valine residue. In turn, this latter valine showed a correlation from its *N*-methyl to the carbonyl of the adjacent Mpla residue. As such, the initial structure proposed for

Table 4.2: *Data table of hectoramide (**96**). Spectra collected in $CDCl_3$ with 1.0%* v/v *TMS on a 600 MHz instrument and an inverse 500 MHz cryo-probe, for the carbon spectrum. [a]Represents cumulative data from two HMBC experiments with differing experimental parameters (Figures A.43 and A.44 in Appendix).*

Residue	Position	δ_C, type	δ_H (J in Hz)	HMBC[a]	TOCSY
N-Me-Val-1	1	171.3, C	–		
	2	62.2, CH	4.58 d (11.3)	1, 3, 5, 6	3-5
	3	25.4, CH	2.31 m	2, 5	2, 4,5
	4	18.2(8), CH_3	0.78 d (6.7)	2, 3, 5	2, 3, 5
	5	19.6, CH_3	1.00 d (6.4)	2-4	2-4
	6	30.8, CH_3	3.05 s	2, 7, self	
N-Me-Val-2	7	171.7, C	–		
	8	58.6, CH	5.25 d (10.9)	7, 9-13	9-11
	9	27.1, CH	2.40 m	8, 10, 11	8, 10, 11
	10	18.3(4), CH_3	0.88 d (6.9)	8, 9, 11	8, 9, 11
	11	19.4, CH_3	0.93 d (6.4)	8-10	8-10
	12	29.7, CH_3	2.99 s	8, 13, self	
Mpla	13	174.8, C	–		
	14	69.9, CH	4.52 bd (8.6)	OH, 15a,b	
	15a	40.5, CH_2	2.6 dd (8.8, 14.2)	13, 14, 17	14, 15b
	15b	40.5, CH_2	2.8 dd (3.3, 14.2)	17	14, 15a
	16	129.1, C	–		
	17	130.1, CH	7.16 d (8.6)	15, 19, self	18
	18	114.0, CH	6.84 d (8.5)	16, 19, self	17
	19	158.5, C	–		
	20	55.3, CH_3	3.79 s	19, self	
	OH	–	3.51 bs		14

this compound was (HOOC-*N*-Me-Val)-(*N*-Me-Val)-(Mpla-OH) (**103**); however, this was inconsistent with data from HRMS analysis. While the structure **103** would have a predicted molecular formula of $C_{22}H_{34}N_2O_6$, the MS^1 $[M+H]^+$ peak of this compound was at *m/z* 422.2648 Da, consistent with the molecular formula of $C_{22}H_{35}N_3O_5$. This was resolved by comparing predicted ^{13}C-NMR shifts with various candidate structures, and indicated that a terminal primary amide and hydroxy group were fully consistent with the C-1 and C-14 ^{13}C-NMR shifts (Figure B.41 in Appendix). In addition, MS^2 fragmentation showed

fragment ions diagnostic for this structure (Figure 6), explicitly (H_2NOC-*N*-Me-Val)-(*N*-Me-Val)-(Mpla-OH).

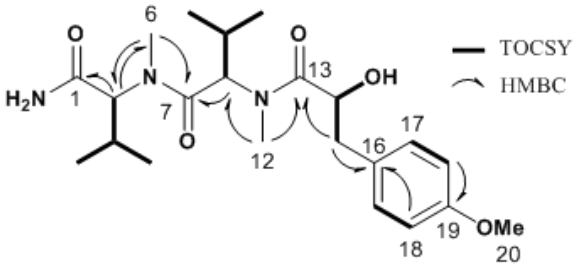


Figure 4.6: *TOCSY and key HMBC correlations in hectoramide (**96**).*

The absolute configuration of **96** was elucidated by traditional hydrolysis and derivatization methods. The two *N*-methyl-valine residues were characterized by Marfey's analysis [Marfey, 1984, Bhushan and Brückner, 2004], and the Mpla was derivatized with (2*S*)-2-octanol for chiral GCMS analysis [Nunnery et al., 2012]. These analyses established the configuration of both valine residues as L, and the Mpla moiety as *S*.

Based on the structure of **96**, its biosynthetic pathway would be expected to have three adenylation domains, coding for one tyrosine and two valine residues, and three methylation domains, two for amide *N*-methylations and one for the phenol *O*-methylation. In this regard, the planar structure of **96** is different from any of the predicted antiSMASH molecules for the pathways observed in the *M. producens* JHB genome (Figure 5). To refine the predicted structures of these pathways, the adenylation domains were submitted to NRPSpredictor2 [Rausch et al., 2005, Röttig et al., 2011, NRP, 2015]. This improved our confidence in the predictions from antiSMASH, but also strongly suggested that these pathways were not responsible for the production of **96**. The genome was further interrogated specifically for valine and tyrosine adenylation domains and methyltransferases. Searching the genome for these targets found several hits; however, each was inconsistent with the expected biosynthetic pathway of **96**. The bacteriocin pathways annotated by AntiSMASH were also inconsistent with the biosynthesis of **96** because the necessary sequence "YVV"

was not found in the leader peptide sequences, and because there were no methylation domains in these clusters.

If **96** was produced by a heterotrophic bacterial strain living in association with *M. producens* JHB, this might explain why the pathway was not found in the *M. producens* JHB genome. Because Moorea producens JHB is not axenic, the raw genome data is actually a rarefied metagenome of DNA from both the cyanobacteria and associated heterotrophic bacterial strains. After the DNA sequence data were assembled, it was binned into groups of high (generally heterotrophic) or low (generally cyanobacterial) G+C content. To continue the search for a **96** biosynthetic pathway, the high-G+C scaffolds were also evaluated with antiSMASH, and revealed three additional biosynthetic gene clusters. However, none of the latter pathways were NRPS pathways; and an additional bacteriocin pathway contained neither a leader sequence consistent with **96**, nor methylation domains. As none of the pathways found in the *M. producens* JHB metagenomic data could convincingly be attributed to **96** biosynthesis, this leaves the origin of this compound uncertain; possibilities include that the compound is produced by a fragmented pathway, a pathway that was not annotated by AntiSMASH, or by a heterotrophic bacterium in the culture but for which the genetic material was not captured by this genome sequencing work.

4.3.4 Jamaicamide Halogenation

Molecular networking analysis of the *M. producens* JHB extract revealed that the pathways for jamaicamide and hectochlorin had a greater capacity for analog production than was previously appreciated. To further probe this, specifically the process of halogenation in the jamaicamides, filaments of *M. producens* JHB were grown in SW-BG11 media that was supplemented with sodium bromide and sodium iodide in equimolar concentrations (10 mM each) above the natural abundance of bromide (0.84 mM) or iodide (0.2-0.5 μM, average of 0.4 μM) in seawater [Bruland, 1983]. Profiling of the crude extract of this small scale

experiment by LCMS with molecular networking analysis indicated an additional mass and fragmentation pattern consistent with an iodinated analog of **90**, named here as jamaicamide F (**104**) (Figures B.7 and B.8 in Appendix).

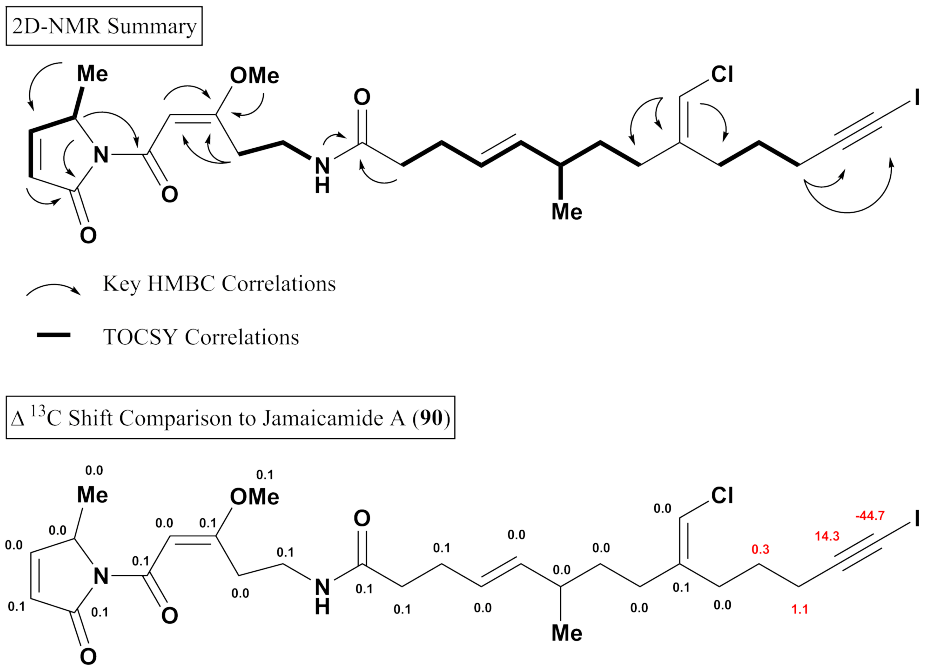


Figure 4.7: *NMR data for jamaicamide F (**104**).*

To confirm the structure of compound **104**, multiple larger scale cultures of *M. producens* JHB were grown in media supplemented solely with sodium iodide (10 mM). After growth for 8-9 weeks the cultures were extracted and the crude extract purified by sequential RP-SPE and RP-HPLC to afford **104**. Its structure was confirmed by NMR analysis, and generally showed a dataset very similar to that of 2. All carbon atoms in **104** were within 0.0-0.1 ppm of those in **90** except for those four closest to the terminal alkyne, suggestive of conservation of structure and configuration between these two analogs (Figure 7). The iodo-alkyne functionality possessed distinctive chemical shifts of -6.6 and 94.2 ppm for C-1 and C-2, respectively (Table B.5 and Figure B.32 in Appendix). The original crude extract chromatogram of *M. producens* JHB grown in normal SW-BG11 media was subsequently inspected for the presence of **104**. Indeed **104** could be observed in the MS^1

extracted ion chromatogram at *m/z* 615 with a distinct retention time (29.4 min), and is thus to our knowledge the first report of a naturally occurring iodo-alkyne. However, it was of too low intensity to be selected for fragmentation and observed by the MS^2-based molecular network. Despite the fact that iodide salts are not specifically included in SW-BG11 media, they are present as trace constituents in sodium chloride (bromide averaging 56 ppm and iodide averaging 0.24 ppm in Instant Ocean brand sea salt [Link, 2015]), thus explaining the source of iodine to produce compound **104** in cultures prepared in SW-BG11 media.

It was interesting that when *M. producens* JHB was grown in media with an equal molar abundance of bromide and iodide ions, the two compounds **90** and **104** were observed in the LCMS chromatogram with equal intensity. While ion intensity can be difficult to quantitate, in this case because the difference in structure is replacement of a bromine atom with an iodine atom, it is likely that compounds **90** and **104** ionize comparably. The production of **90** and **104** in equal amounts in this experiment suggests that the enzyme is not selective for bromine or iodine. Halogenation enzymes can derive their selectivity from differences of halide electronegativity, abundance in nature, ionic radius and reactivity [Butler and Walker, 1993, Fujimori and Walsh, 2007]. Fluoride's paramount electronegativity results in the fact that enzymes incorporating this halogen are both rare, and unusually for halogenases, non-oxidative [Fujimori and Walsh, 2007]. In seawater fluorine, bromine, and iodine are relatively rare with F/Cl/Br/I molar ratios of 2,000:1,000,000:2,000:1 [Bruland, 1983]. Thus, it appears that the enzyme responsible for bromination of jamaicamide excludes iodine on the basis of its relative scarcity in seawater, yielding **90** as the major product in nature.

Previous work on the jamaicamide biosynthetic pathway showed that hexanoic acid, 5-hexenoic acid, and 5-hexynoic acid but not 6-bromo-5-hexynoic acid were suitable substrates for JamA [Dorrestein et al., 2006]. By in vivo MALDI-TOF MS analysis following ^{15}N labelling, it was found that production of **91** is light dependent, whereas production of **90** continues in the dark, and that **90** can be produced without fresh production of **91**

[Esquenazi et al., 2011]. These data strongly support the proposal that the biosynthesis proceeds through a final bromination step on a pool of free **91** to afford **90**.

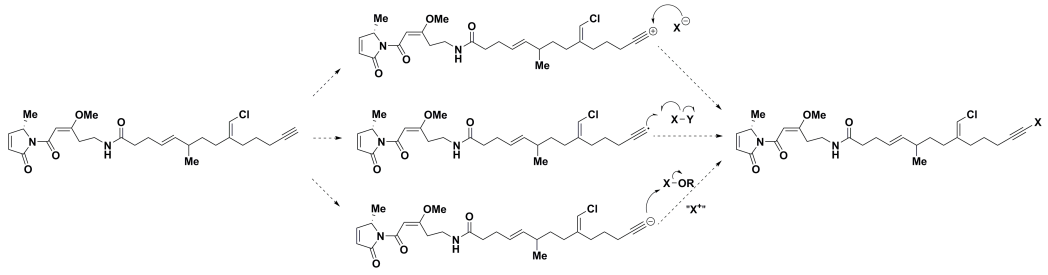


Figure 4.8: *The potential mechanisms of halogenation of the alkyne in the jamaicamides.*

Three mechanisms of bromination of **90** are conceivable using Br^{-} by a nucleophilic halogenase, $Br^{\cdot}$ by a non-heme iron O_2-dependent halogenase, or 'Br^{+}' by a haloperoxidase or an $FADH_2$-dependent halogenase (Figure 8) [Fujimori and Walsh, 2007, Blasiak and L., 2009]. Assuming bromination occurs on the terminal alkyne, as discussed above, then halogenation by a nucleophilic halogenase Br- is unlikely because this would putatively proceed through a high energy sp hybridized alkynyl carbocation species, which is highly disfavored over the resonance stabilized propargyl cation [Carey and Sundberg, 2007]. Bromination through the alkynyl radical is a conceivable mechanism, but the enzyme activity profile of reacting with both bromine and iodine and being unable to react with abundant chlorine is different from the reactivity profile of the known radical halogenases. Known non-heme iron O_2-dependent halogenase can catalyze both chlorination and bromination (with chlorination being highly favored), and no reported iodination activity [Blasiak and L., 2009]; this may be due to the steric constraint of the larger iodide or because of a difference in fundamental energetics, radical iodination with elemental iodine is an endothermic process [Carey and Sundberg, 2007]. Bromination through a hypohalite 'Br^{+}' species is compatible with the observation of an iodinated species and no chlorinated species because bromoperoxidases are capable of oxidizing the less electronegative iodine but not the more electronegative fluorine or chlorine [Butler and Walker, 1993]. Flavin-dependent

halogenases have not been observed to iodinate, hypothesized as a problem of sterics for the larger iodide or hydrolysis of the bound iodine intermediate, so they are a poorer candidate for producing the reactive "X^+" equivalent [Blasiak and L., 2009].

Walsh and coworkers suggested an alternative mechanism for the bromination of jamaicamide via radical bromination of a saturated precursor methyl terminus [Neumann et al., 2008]; however, this seems flawed for two possible reasons. First, as noted above, this terminal halogenation appears to occur with the alkyne rather than the alkane. Additionally, as stated above, with the discovery of **104**, the reactivity profile of acting on bromine and iodine but not chlorine does not match to any known non-heme iron and O_2-depenedent halogenase. Therefore, electrophilic addition by a haloperoxidase remains as the most likely mechanism for the bromination/iodination event in jamaicamide biosynthesis.

4.3.5 Ion Channel Pharmacology of the Jamaicamides

The ability of compounds **90**, **91**, and **104** to antagonize veratridine-stimulated Ca^{2+} influx was studied in neocortical neurons. All three jamaicamides tested (**90**, **91**, and **104**) produced concentration-dependent antagonism for the increase in neuronal $[Ca^{2+}]_i$ induced by veratridine (Figure B.37 in Appendix). The concentration-response curves were fit by a three-parameter logistic equation yielding IC_{50} values of 1.82 μM (95% CI = 0.9-3.3 μM), 6.88 μM (95% CI = 3.0-15.6 μM) and 4.3 μM (95% CI = 2.2-8.4 μM), respectively for **90**, **91**, and **104**. The results suggest that **90** is approximately 2-3 times more potent than the other two jamaicamides tested.

Given that **90** and **91** have previously been reported to have sodium channel blocking activity at a concentration of 5 μM in Neuro-2a mouse neuroblastoma cell lines [Edwards et al., 2004], we assessed their ability to block veratridine-stimulated sodium influx. We used primary cultures of mammalian neurons as a model system that is more relevant than transformed cell lines to mammalian neurotoxicology [Cao et al., 2011]. We determined the

ability of these compounds to antagonize veratridine-stimulated Na^+ influx in neocortical neurons. All three jamaicamides (**90**, **91**, and **104**) produced concentration-dependent antagonism of the increase in neuronal $[Na^+]_i$ induced by veratridine (Figure B.38 in Appendix). The concentration-response curves were best fit by a three-parameter logistic equation yielding IC_{50} values of 1.1 μM (95% CI = 0.5-2.5 μM), 3.6 μM (95% CI = 1.5-8.5 μM) and 2.3 μM (95% CI = 1.0-5.0 μM), respectively for **90**, **91**, and **104**. Again, our results suggest that **90** is approximately 2-3 times more potent than the other two jamaicamides as a sodium channel blocker in neocortical neurons. Collectively, these data indicate that the structural differences between **90**, **91**, and **104** have only minor roles in the interaction of these compounds with voltage-gated sodium channels.

4.4 Conclusions

The results of using molecular networking, genome analysis, and directed biosynthetic feeding studies with *M. producens* JHB illustrated that each approach has the potential to enhance the traditional structure or bioassay-guided natural products workflows. Just as importantly, it revealed that previous studies have overlooked metabolic diversity in the hectochlorin and the jamaicamide structural families. This is not because the previous studies were poorly designed or executed, but because these previous methods were inherently less comprehensive and thus do not fully explore the secondary metabolome of a given organism.

The current study demonstrated the power of molecular networking to guide the isolation of new analogs in desired natural product families; however, the greater lesson is that no single method is comprehensive, and that to truly appreciate the secondary metabolome of an organism, it is necessary to employ orthogonal methods. Such an approach reveals minor or overlooked metabolites, which may be intermediates or byproducts of the biosynthetic machinery responsible for the secondary metabolome, and is applicable to complex samples, such as field collections from the environment or, as in this case,

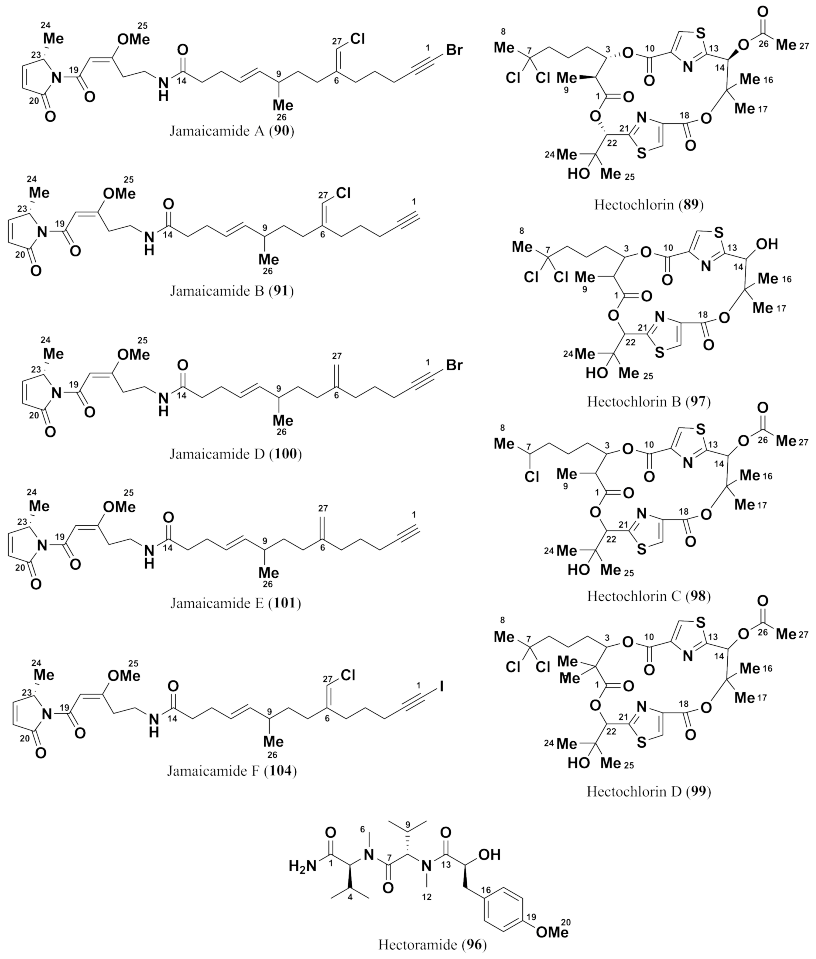


Figure 4.9: *The expanded metabolome of* M. producens *JHB described in this study.* The carbon atoms are numbered as they are in the text and supporting information.

non-axenic cultures. Compounds produced in only trace quantities, however, can still be extremely valuable for understanding the biosynthetic process, for providing analogs to gain an initial appreciation of structure-activity relationships, and to provide a holistic view of the secondary metabolome of an organism.

4.5 Materials and Methods

4.5.1 General Experimental Procedures

NMR spectra were collected on a Varian Unity 500 MHz (500 MHz and 125 MHz for the ^{1}H and ^{13}C nuclei respectively), a Varian VX 500 MHz with ^{13}C-optimized cryoprobe, and a Bruker 600 MHz (600 MHz and 150 MHz for the 1H and ^{13}C nuclei respectively) with 1.7 mm inverse cryo-probe. NMR experiments were conducted using CDCl3 from Cambridge Isotope Laboratories, Inc. 99.8% D, containing 0.03% or 1.0% *v/v* trimethylsilane (referencing δ_H 0.0 as the internal standard from trimethylsilane, and δ_C 0.0 or δ_C 77.16 as internal standards using trimethylsilane and $CDCl_3$, respectively). Microwave-heated reactions were run in a Biotage Initiator microwave synthesizer. LR-LCMS data were collected on a Thermo Finnigan Surveyor Autosampler/LC-Pump-Plus/PDA-Puls with a Thermo Finnigan Advantage Max mass spectrometer. HRMS data was collected on a Finnigan LTQ-FTICR-MS instrument (Thermo-Electron Corporation, San Jose, CA) fitted with either an Ion-Max ESI source for LCMS runs, or a Biversa Nanomate (Advion Biosystems, Ithaca, NY) electrospray source. HPLC purification was carried out with a Waters 515 HPLC Pump with a Waters 996 Photodiode Array Detector using Empower Pro software. All solvents were HPLC grade except for H_2O which was purified with a Millipore Milli-Q system before use, CH_3CN which was LCMS grade from J.T. Baker, and acetone which was distilled before use.

4.5.2 Crude Extraction and LCMS

M. producens JHB samples were prepared from material cultured in separate technical replicates in artificial seawater media using a 16-8 h standard light-dark cycle at 28 °C. The cultures were collected by rapid vacuum filtration from the media at room temperature, combined, and extracted iteratively with 2:1 dichloromethane/methanol to afford 0.81 g

crude extract. LC-HRMS data for the creation of the molecular network were collected using samples of the crude cyanobacterial extract prepared from stock solutions of the extract at 100 mg/mL in 1:1 EtOH/isooctane. Of the stock solution, 10 μL (1 mg crude extract) was diluted to 1 mg/mL in MeOH or CH_3CN, and insoluble solids were filtered with a Celltreat PTFE syringe filter before use. Samples and a blank of CH_3CN were run on a LC-LTQ-FTICR-MS (Liquid-column Chromatography Linear Triple Quadrupole Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with a Phenomenex Synergi 4 m Fusion-RP 80 Å 100 x 2.00 mm column with a Security Guard. The elution used a gradient that began with 70% H_2O (acidified 1% *v/v* with HCOOH)/CH_3CN for 5 min, then ramped to 1% H_2O/CH_3CN at 40 min, held there for 5 min, brought back 70% at 47 min, and re-equilibrated for 3 min. The flow was diverted to waste for the first 5.25 min. The MS detector used a serial set of five scans; one high-resolution FT-MS^1 scan was followed by one low-resolution LTQ-MS^1 scan, and then three LTQ-MS^2 scans of the three most intense ions from the LTQ-MS^1 scan.

4.5.3 HPLC Purification

The VLC fractions containing hectochlorin (80% EtOAc/hexanes and 100% EtOAc) were combined and fractionated with a Grace Pure C18-Max 1 g/6 mL reverse phase solid phase extraction column (RP-SPE). The column was prepared with three column volumes of CH_3CN and then run with a 30 mL elution each of CH_3CN, MeOH, and CH_2Cl_2. The latter VLC fractions which contained the hectochlorin analogs (25% MeOH/EtOAc and 100% MeOH) were combined and fractionated in the same fashion. From each the CH_3CN fraction was purified by HPLC, using a Synergi 4μ RP-80 250 x 10.00 mm column with an isocratic elution at 3.0 mL/min. The first fraction was eluted with 1.2:1.8 H_2O/CH_3CN for 70 min, the second with 1.25:1.75 H_2O/CH_3CN for 30 min. **89** was collected as 6.6 mg of a yellow-green oil, **90** as 8.9 mg of a light green oil, and **91** as 6.7 mg of a yellow-green

oil; each of these matched literature reports for the NMR and HRMS spectra of these compounds [Marquez et al., 2002, Edwards et al., 2004]. Also collected was **96** as 1.3 mg of a yellow oil, **98** as 0.2 mg of a yellow oil, **99** as 2.6 mg of a yellow-brown oil, and 8 as 0.5 mg of a light-green solids. Samples of **97** from the two HPLC purifications were combined and re-purified with a Jupiter 5μ C-18 300 Å 250 x 4.60 mm column eluted with 2:3 CH_3CN/H_2O at 1.5 mL/min and the purified **97** was collected as 1.1 mg of colorless oil.

The crude extract from the sodium iodide enrichment experiment was purified similarly with the same RP-SPE column but using an elution of 1:1 CH_3CN/H_2O, CH_3CN, MeOH, and CH_2Cl_2. The CH3CN eluting fraction was purified by HPLC using a Phenomenex Kinetex 5μ C-18 100 Å 100 x 4.60 mm column, eluted with an isocratic elution of 52% CH_3CN/H_2O at 1.0 mL/min. The **91** was collected as 1.2 mg of a yellow oil and **104** as 1.5 mg of a faint-yellow oil.

*Hectoramide (**96**)*: yellow oil; $[\alpha]25_D$ -11.1 (CH_2Cl_2, c = 3.5 mM); UV-Vis (MeOH) λ_{maxima} (log ε) 204 (2.(2)) nm, 262 (1.(7)) nm; IR (neat) ν_{max} 3396.4, 2961.8, 2929.7, 1681.1, 1629.4, 1466.9, 1249.5, 1073.9, 1036.7 cm^{-1}; see Table 2 for NMR data; HR-ESI-FT-MS $[M + H]^+$ *m/z* 422.2648 (calculated for $C_{22}H_{36}N_3O_5$ 422.2649, -0.1 mamu).

*Hectochlorin B (**97**)*: faint-yellow oil; $[\alpha]25_D$ -0.325 (MeOH, c = 47.2 mM); UV-Vis (MeOH) λ_{maxima} (log ε) 237 (4.(0)) nm; IR (neat) ν_{max} 3394.2, 2976.5, 2932.2, 1715.4, 1379.8, 1323.6, 1244.5, 1150.8, 1090.5, 741.1 cm^{-1}; see Table B.3 in Appendix for NMR data; HR-ESI-FT-MS $[M + H]^+$ *m/z* 623.1050 (calculated for $C_{25}H_{33}O_8Cl_2N_2S_2$ 623.1050, 0 mamu).

*Hectochlorin C (**98**)*: yellow oil; $[\alpha]25_D$ 4.35 (MeOH); UV-Vis (MeOH) λ_{maxima} (log) 232 (3.(9)) nm; IR (neat) νmax 2926.6, 2857.7, 1715.6, 1243.7, 1211.7, 1151.1, 1085.1, 1048.3 cm-1; HR-ESI-FT-MS $[M + H]^+$ *m/z* 631.1545 (calculated for $C_{27}H_{36}O_9ClN_2S_2$ 631.1545, 0 mamu).

*Hectochlorin D (**99**)*: yellow-brown oil; $[\alpha]25_D$ -1.53 (MeOH); UV-Vis (MeOH)

λ_{maxima} (log ε) 233 (3.(4)) nm; IR (neat) νmax 3302.8, 2928.1, 2861.1, 1712.2, 1449.0, 1242.5, 971.0 cm^{-1}; HR-ESI-FT-MS $[M + H]^+$ *m/z* 679.1306 (calculated for $C_{28}H_{37}O_9Cl_2N_2S_2$ 679.1312, -0.6 mamu).

*Jamaicamide D (**100**)*: faint-green, amorphous solids; see Table B.4 in Appendix for NMR data; HR-ESI-FT-MS $[M + H]^+$ *m/z* 533.2012 (calculated for $C_{27}H_{38}O_4BrN_2$ 533.2009, 0.3 mamu).

*Jamaicamide F (**104**)*: faint-yellow oil; $[\alpha]25_D$ 83.8 (CH_2Cl_2, c = 2.4 mM); UV-Vis (MeOH) λ_{maxima} (log ε) 266 (4.(1)) nm; IR (neat) ν_{max} 3318, 3116, 2933, 2863, 1716, 1656, 1599, 1545, 1439, 1396, 1335, 1296, 1204, 1171, 1137, 1082, 973, 823, and 754 cm^{-1}; see Table B.5 in Appendix for NMR data; HR-ESI-FT-MS $[M + H]^+$ *m/z* 615.1471 (calculated for $C_{27}H_{37}O_4ClIN_2$ 615.1481, -1.0 mamu).

4.5.4 Pure Compound HR MS/MS

Fractions from the HPLC purification were prepared for HRMS. Samples were diluted 1:10 with 50% H_2O/MeOH acidified 1% v/v with formic acid and subjected to electrospray ionization with a Nanomate nano-spray source (pressure: 0.3-0.4 psi, spray voltage: 1.3-1.4 kV), and then fragmented and analyzed on a LTQ-FT-MS. The instrument was first auto-tuned to a standard of 100 μM cytochrome C. The $[M + H]^+$ ion of each compound was isolated in the linear ion trap and fragmented by collision induced dissociation. Sets of consecutive, high-resolution MS/MS scans were acquired in profile mode and averaged using QualBrowser software by Thermo.

4.5.5 Molecular Network

MS spectra were converted to Mascot generic format (.mgf) files using MSCovert (from Proteowizard [52]) and then networked with the Spectral Networking algorithm using a minimum cosine of 0.7; the data were then viewed in Cytoscape [Cyt, 2015].

4.5.6 Preparation of Mpla Ester Standards

Samples of L and D *O*-methyl-tyrosine, 19.6 mg and 19.7 mg respectively, were charged into 8.0 mL vials and magnetic stir bars were added. These amino acids were dissolved in 1.00 mL of 0.9999(7) M $NaNO_2$ (*aq*) and cooled to 0 °C with stirring and 5.0 mL of 0.200(6) M perchloric acid (*aq*) was added. The reaction mixture was warmed to room temperature slowly over 30 min then sealed and heated to reflux by oil bath for 5 min. Afterwards, it was cooled to room temperature and diluted with 10 mL of 5% citric acid (*aq*) and extracted with 3 x 10 mL of CH_2Cl_2. The crude reaction mixture was concentrated to afford a colorless oil. This was placed in a fresh vial with a stir bar, then dissolved in 300 μL of (2*S*)-2-octanol while stirring under Ar. The acyl chloride (150 μL) was added dropwise, the vial sealed and heated to 115 °C in an oil bath. After 1.5 h the vials were cooled to room temperature and quenched with 1.0 mL of water and 2.0 mL of dichloromethane. The organic layer was passed over ca. 2 cm of $MgSO_4$ in a glass pipette, concentrated to afford 4.8 and 4.1 mg of the two esters as colorless oils for the L and D case, respectively.

4.5.7 Stereochemical Analysis of Hectoramide (96)

A 300. μL aliquot of a 1 mg/mL solution of **96** in CH_2Cl_2 was transferred to a 0.5-2 mL microwave reaction tube with a stir bar, dried under N_2 (*g*), and taken up in 600 μL of 6 M HCl (*aq*). This mixture was reacted in a microwave reactor at 110 °C for 5 min, then dried again under N_2 (*g*). The sample was redissolved in 600 μL of CH_2Cl_2, split into two equal portions, dried again under N_2 (*g*), charged with a magnetic stir bar, and put under an inert Ar atmosphere.

The first sample was dissolved in 150 μL of 1 M $NaHCO_3$ and 300.0 L of D-FDAA solution (1.00 mg/mL in acetone) was added dropwise with stirring. This reaction mixture was warmed to 45-50 °C for 1 h, then cooled to room temperature and neutralized with 75.0 μL of 2 M HCl (*aq*). This material was transferred to a fresh vial with 3 x 0.250 mL of

CH_3CN, concentrated to dryness under N_2 (g). The sample was transferred to an LCMS vial through a syringe filter with 3 x 0.5 mL of CH_3CN, re-concentrated under N_2 (g) and brought to a final volume of 150 μL CH_3CN. A 25.0 μL aliquot of the sample was injected on the LR-LCMS and run in comparison with previously prepared authentic standards [Mevers et al., 2011]. The samples were analyzed by LR-LCMS with a Phenomenex Kinetex 5 μ C18 100 Å 100 x 4.60 mm column, and a 75 min gradient elution beginning at 5% CH_3CN/95% H_2O acidified with 0.1% *v/v* with HCOOH (Acros) for 5 min, and then ramped up to 50% CH_3CN/50% H_2O acidified with 0.1% *v/v* with formic acid over 65 min, and then brought back to the starting condition over 1 min and re-equilibrated for 4 min.

The second portion of the hydrolysate was dissolved in 200 μL of (2*S*)-2-octanol and 100 μL of acyl chloride was added dropwise with stirring. The reaction vial was sealed and warmed to 110-120 °C and kept for 1 h, then cooled to room temperature and quenched with 0.50 mL of H_2O. The reaction was extracted with 2.0 mL of CH_2Cl_2 and the organic layer was separated and passed through $MgSO_4$. The organic layer was dried and reconstituted in a GCMS vial with 3 x 50 μL of CH_2Cl_2 and then analyzed by GCMS alone and in comparison with similarly derivatized D and L standards by co-injection.

4.5.8 Genome sequencing

Genomic DNA from cultured biomass of *M. producens* JHB was extracted using a standard phenol:$CHCl_3$:isoamyl alcohol extraction protocol. The genome was sequenced at the University of Michigan Microarray Core Facility (http://www.umich.edu/ caparray/), using Illumina HiSeq sequencing. Raw reads were corrected via BayesHammer [Nikolenko et al., 2013]. Assembly was performed using the SPAdes Genome Assembler version 3.1.1 [Nurk et al., 2013], followed by scaffolding with Opera and binning to obtain cyanobacterial-specific scaffolds [Gao et al., 2011], with an average G+C content of 44%, which is comparable to the related *Moorea producens* 3L genome G+C content of 44% [Jones et al.,

2011]. The sequences for the clusters detailed in Figure 5 were uploaded to GenBank with the accession numbers KP860346-48.

4.5.9 Media Experiments

Cultures of *Moorea producens* JHB were grown as previously described [Engene et al., 2012], in 2 L Erlyenmeyer flasks with 1 L of SWBG-11 media supplanted with equimolar amounts of NaI and NaBr at a concentration of 10 mM (1.5 g and 1.0 g into 1 L SWBG-11 media, respectively), with a control sample grown in unmodified SWBG-11 media. The inoculation was made with only a few short filaments, roughly 2-3 centimeters long. After 7 weeks of growth, the cultures were extracted and analyzed by LCMS as described above.

To provide compound **104** on a larger scale, *Moorea producens* JHB was grown in SWBG-11 media containing 10 mM NaI (15.0 g into 10 L SWBG-11). Multiple cultures were grown in shallow pans containing 5 L of media each and split into new pans with fresh media after 8 or 17 weeks of growth. A total of four pans were grown and the cyanobacteria were extracted and compounds purified as described above.

4.5.10 Neocortical Neuron Culture

Primary cultures of neocortical neurons were obtained from embryonic day 16 Swiss-Webster mice as described elsewhere [Cao et al., 2008, Jabba et al., 2010]. Briefly, pregnant mice were euthanized by CO_2 asphyxiation and embryos were removed under sterile conditions. Neocortices were collected, stripped of meninges, minced by trituration with a Pasteur pipette, and treated with trypsin for 25 min at 37 °C. The cells were then dissociated by two successive trituration and sedimentation steps in soybean trypsin inhibitor and DNase containing isolation buffer, centrifuged, and resuspended in Eagle's minimal essential medium with Earle's salt (MEM) and supplemented with 1 mM L-glutamine, 10%

fetal bovine serum, 10% horse serum, 100 IU/mL penicillin, and 0.10 mg/mL streptomycin (pH 7.4). Cells were plated onto poly-L-lysine-coated 96-well (9 mm), clear-bottomed, black-well culture plates (MidSci, St. Louis, USA) at a density of 1.5 105 cells/well. Cells were then incubated at 37 °C in a 5% CO_2 and 95% humidity atmosphere. The culture media was changed every other day, starting from day 5 in-vitro using a serum-free growth medium containing Neurobasal Medium supplemented with B-27, 100 IU/mL penicillin, 0.10 mg/mL streptomycin, and 0.2 mM L-glutamine. Neocortical cultures were used in experiments between 10-11 days in-vitro. All animal use protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Creighton University.

4.5.11 Intracellular Ca^{2+} Concentration Measurement

Cells grown in 96-well plates were used for determination of intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$). Briefly, the growth medium was removed and replaced with dye-loading medium (100 μL per well) containing 4 μM fluo-3 AM and 0.04% pluronic acid in Locke's buffer. After 1 h of incubation in dye-loading medium, the cells were washed four times in fresh Locke's buffer (8.6 mM HEPES, 5.6 mM KCl, 154 mM NaCl, 5.6 mM glucose, 1.0 mM $MgCl_2$, 2.3 mM $CaCl_2$, 0.1 mM glycine, pH 7.4) using an automated microplate washer (Bio-Tek Instruments Inc, VT, USA). Various concentrations of **90**, **91**, and **104** were then added to the cells at a rate of 26 μL/s, yielding a final volume of 200 μL/well. The cells were incubated in the presence or absence of jamaicamides for 5-7 minutes at 37 °C in a 5% CO_2 and 95% humidity atmosphere prior to transfer to a FLEX Station II (Molecular Devices) benchtop scanning fluorescence chamber. The fluorescence measurements were performed at 37 °C. The cells were excited at 488 nm and Ca^{2+}-bound fluo-3 emission was recorded at 538 nm at 2 s intervals. After recording baseline fluorescence for 60 s, either vehicle or veratridine was added to each well at a rate of 26 μL/s; the fluorescence was then monitored for an additional 200 s.

4.5.12 Intracellular Na^+ Concentration Measurement

$[Na^+]_i$ measurement and full in-situ calibration of sodium-binding benzofuran isophthalate (SBFI) fluorescence ratio were performed as described previously [Jabba et al., 2010]. Cells grown in 96-well plates were washed four times with Locke's buffer using an automated microplate washer (BioTek Instruments, Winooski, VT). After measuring the background fluorescence of each well, cells were incubated for 1 h at 37 °C with dye-loading buffer (100 L/well) containing 10 μM SBFI-AM (Invitrogen) and 0.02% Pluronic F-127 (Invitrogen). Cells were washed five times with Locke's buffer, leaving a final volume of 150 L in each well. Jamaicamide addition and preincubation was as described above for Ca^{2+} concentration determination. Plates were then placed in a FlexStation II (Molecular Devices, Sunnyvale, CA) chamber to detect Na^+-bound SBFI emission at 505 nm (excitation wavelength was 340 and 380 nm). Fluorescence readings were taken once every 5 s for 60 s to establish the baseline, then either vehicle (control) or veratridine was added and fluorescence was monitored for an additional 260 s. After correcting for background fluorescence, SBFI fluorescence ratios (340/380) and concentration-response graphs were generated.

4.5.13 Bioassay Data Analysis

The Fluo-3 and SBFI raw fluorescence emission data were exported to an Excel work sheet. The Fluo-3 fluorescence or SBFI fluorescence ratios (340/380) versus time were analyzed and concentration-response graphs generated using GraphPad Prism software (GraphPad Software Inc., San Diego, CA). The IC_{50} values of **90**, **91**, and **104** for antagonism of veratridine-stimulated calcium or sodium influx were determined by nonlinear regression analysis using a three-parameter logistic equation.

4.6 Acknowledgements

This work was supported by NIH grants NSO53398 (to W. H. G. and T. F. M.), CA100851 (to W.H.G.), CA108874 (to W. H. G., L. G., D. H. S.) and GM107550 (W. H. G., L. G., and P. C. D.), and the Russian Science Foundation grant 14-50-069 (A. K.). We thank the government of Jamaica for permission to make the original collection of *M. producens* JHB, T. Byrum for help in the culturing of the cyanobacterium and the photo used to create Figure 1, B. Duggan for help with the acquisition of the 600 MHz data sets, and J. Watrous, M. Meehan, C. Rath, and D. Nguyen for help with the LTQ-FTICR-MS and construction of the molecular networks, R. C. Coates for allowing us use of his Moorea producens JHB extracts and VLC fractions, and P. Fardella for help with secondary metabolite sequence analyses.

This chapter, in full, is a reprint of the material as it appears in the Public Library of Science Online Edition, Boudreau, Paul D.; Monroe, Emily A.; Mehrotra, Suneet; Desfor, Shane; Korobeynikov, Anton; Sherman, David H.; Murray, Thomas F.; Gerwick, Lena; Dorrestein, Pieter C.; Gerwick, William H. **2015**, volume *10*, e0133297.

Chapter 5

Synthesis of Marine Natural Products, Balancing Utility Against Availability and Structural Insight

5.1 Abstract

Total chemical synthesis is a tool used either late in natural products workflows, or not at all. The reasons for this are straight forward; a synthetic effort can only be directed after a structure is known, or at least mostly established (it can, for example, be used as a tool to resolve lingering questions about stereochemistry), so that a synthetic target can be identified. In addition, organic synthesis requires a different skill set and lab set up than most natural products work; meaning that research groups tend to focus on one or the other. All of this belies the great utility that total synthesis holds for natural products research, where synthesis can help confirm a structure, provide more of a rare or compound of interest with limited availability, and provide analogs that can help elucidate a mechanism of action. As with all techniques, total synthesis has costs and benefits that must be balanced against

the utility of having more material or analogs. Laurencione (**105**) and gallinamide A (**l06**) both provided synthetically tractable targets, but of differing degrees of difficulty; **105** is a one step synthesis from a commercially available starting material whereas **106** is many more. The reasons for embarking on these projects was weighed against the necessary effort to synthesize these compounds. While **105** is quite easy to produce, and having compound on hand was of interest for biological assay work; the questions surrounding **106** had to be deeper as the effort was considerably more. The questions concerning **106** centered on pharmacological issues could the selectivity between the different cathepsin protease targets be improved? What were the key features that determined potency in this system? The nature and impact of these questions justified the more significant synthetic effort required to produce the gallinamide analogs described herein.

5.2 Introduction

Laurencione (**105**) was first isolated from the marine red algae *Laurencia spectabilis* [Bernart et al., 1992], it has an unusual structure because it contains both a primary alcohol and a reactive α-diketone, which means that it exists as an equilibrium mixture of the open form (5-hydroxypentane-2,3-dione) and the cyclized ketal from (2-hydroxy-2-methyldihydrofuran-3(2H)-one). Lowery and coworkers subsequently showed that **105** is one of several small molecule analogs of (*S*)-4,5-dihydroxy-2,3-pentanedione (**107**) that can mimic the effects of **107** in signaling bacterial quorum sensing through the AI-2 system [Lowery et al., 2005]. Work by Lena Gerwick's group had suggested that there might be a connection between molecules involved in quorum sensing and those that can modulate the mammalian inflammatory system, however, there was no sample of **105** available to test this conjecture. A one step synthesis of **105** had been reported by Aelterman and coworkers [Aelterman et al., 1997], which is also how Lowery and coworkers provided **105** for their work, and offered the opportunity to synthesize more material to test in Lena Gerwick's

LPS-stimulated macrophage inflammation assay.

Gallinamide A (**106**) was isolated by Linington and coworkers from a Panamanian collection of cyanobacteria [Linington et al., 2009] as a compound active against the malaria parasite *Plasmodium falciparum*, and then independently isolated by Taori and coworkers from a Floridian cyanobacterial collection [Taori et al., 2009]. The stereochemistry was confirmed by a synthesis preformed by Conroy and coworkers [Conroy et al., 2010, Conroy et al., 2011], and it was also independently synthesized by Stolze and coworkers, who showed that it inhibited the *Plasmodium* falcipain cystine protease [Stolze et al., 2012]. Later work described the activity of natural **106** as a cathepsin inhibitor, and demonstrated that it was potent, irreversible, and selective for cathepsin L over cathepsin V [Miller et al., 2014].

Guided by previous syntheses of **106**, a retrosynthetic analysis fragmented the molecule into three components: an enamide core, the cyclized head group and the lipophilic tail. With the exception of L-leucic acid, all of the base starting materials were variously protected standard amino acids (Figure 1).

5.3 Results and Discussion

5.3.1 Synthesis of Laurencione

Following the synthesis of Aelterman [Aelterman et al., 1997], the **105** could be synthesized from the commercially available 5-hydroxypentan-2-one by seleneous acid oxidation. In the original isolation Bernart and coworkers derivitized the natural **105** with base and acetic anhydride to form laurencione diacetate (**108**) [Bernart et al., 1992], which could also be done with the synthetic product to afford this analog. Using a modification of the synthesis of **105** where the reaction was run in acetic acid, a new analog, the monoacetate form of laurencione (**109**), was also synthesized (Figure 2).

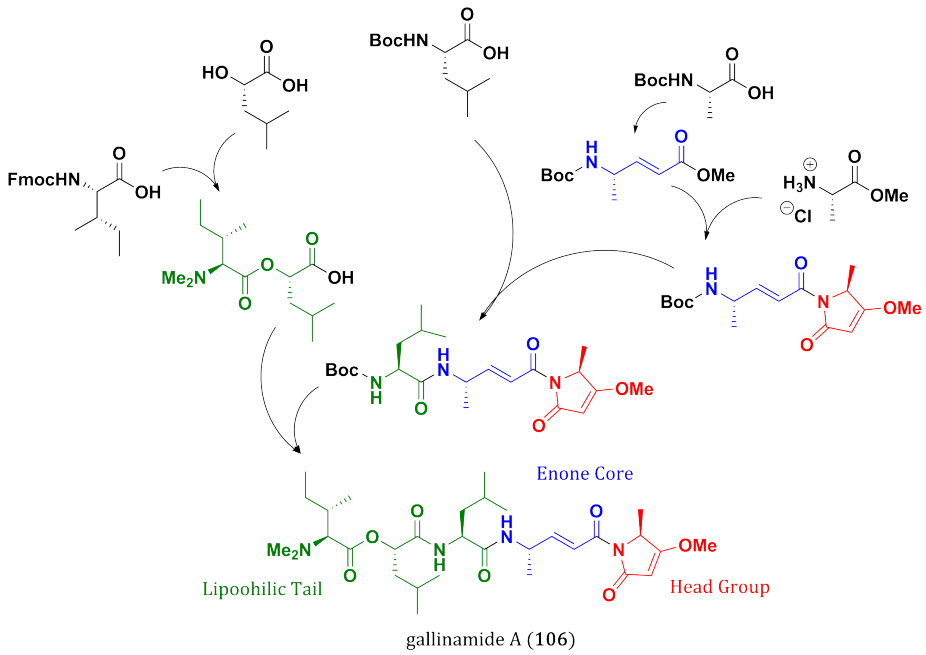


Figure 5.1: *Retrosynthetic analysis of gallinamide A (**106**) and the base components used in this synthesis.*

5.3.2 Synthesis of Gallinamide A

To accomplish the total synthesis of **106**, reactions were drawn from both of the published routes with additional modifications that were found to improve or facilitate the reactions, or necessary due to other constraints. An example of this can be seen in the first steps of the synthesis that involved forming the enamide core from Boc-L-Alanine-OH. The first reaction formed the Weinreb amide, which is then reduced to the aldehyde and reacted with an anylidene to form the conjugated system that will become the enamide. Emulating Conway's scheme was best for forming the Weinreb amide; however, in the next steps Stolze's route of reacting this with the ester anylidene without purification of the aldehyde proved more effective than purifying the aldehyde and then reacting it with free acid anylidene. It was also found that refluxing the Wittig reaction, rather than running it at room temperature, improved the yield and shortened the reaction time (Figure 3).

The synthetic route towards the headgroup was largely draw from Stolze's route,

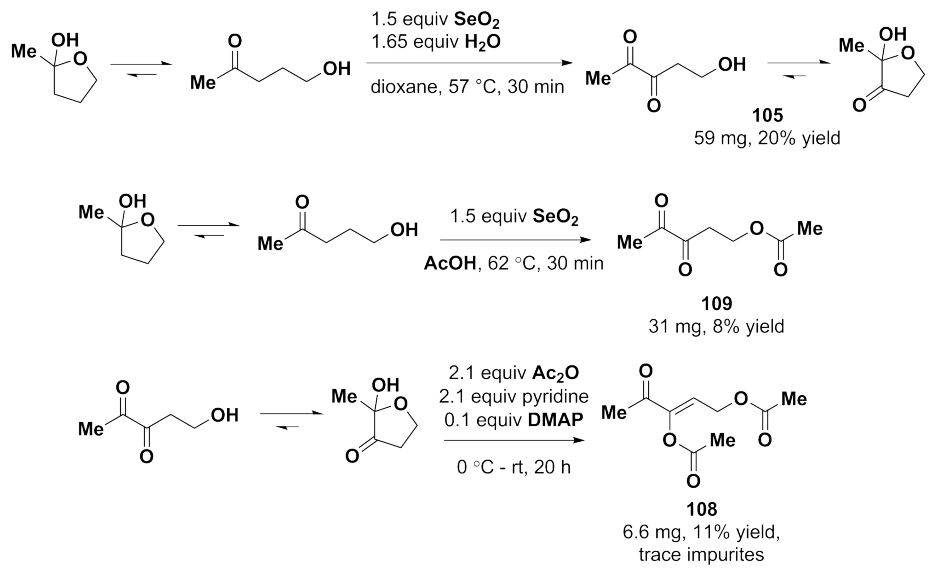


Figure 5.2: *Synthetic scheme for the laurenciones.*

using an 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) based coupling reaction to append an alanine onto the enamide core, which was subsequently deprotected, extended with Meldrum's acid, cyclized, and the enol was trapped via a Mitsunobu reaction (Figure 4).

In the final steps which formed the tail and then coupled it to the enamide core, the route largely followed Stolze's procedure, but again with modifications wherein the reactions conditions used in Conroy's and Stolze's schemes proved difficult to emulate or there was also an opportunity to improve the yield or ease of synthesis (Figure 5). Conroy, for example used piperidine to deprotect the Fmoc group, which is difficult to use because it is currently a controlled substance in California as is the thionyl chloride used in Stolze's route to the tail section.

In this work, to append the tail, Boc-L-Leu-OH was first attached to the free amine of the TFA deprotected amine in the enamide-containing residue via EDC coupling. This was then coupled to a dipeptide ester prepared similarly to Stolze's and Conroy's routes, but using different protecting groups. The L-leucic acid was protected with the benzyl ester rather than a *tert*-butyl ester or methyl ester because the synthesis of the *tert*-butyl

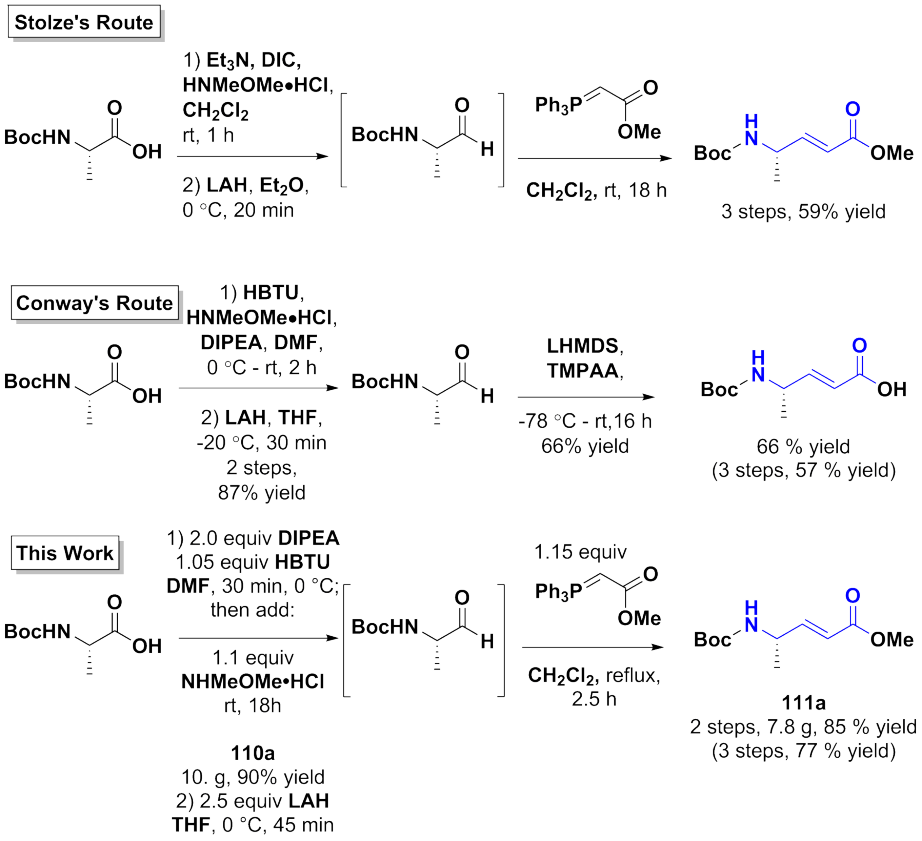


Figure 5.3: *Comparison of synthetic routes towards the gallinamide A (**106**) enamide core.*

ester was difficult to replicate and the cleavage of the methyl ester with lithium hydroxide risked cleaving the desired ester bond connecting the dimer. The benzyl ester of L-leucic acid was connected to the Fmoc-protected L-isoleucine via carbondiimide coupling using *N,N*-diisopropylcarbodiimide (DIC). The product was filtered to remove the DIC derived urea, but was not otherwise purified, as this was not shown to improve the yield before Fmoc deprotection. The Fmoc deprotection with diiospropyl amine gave a free amine that could be dimethylated with methyl iodide, and finally the desired dimer was produced by removing the benzyl group via hydrogenation. The free acid dimer was not found to be stable, so it was used immediately in an EDC coupling with the free amine of the leucine extended intermediate which was again produced by TFA deprotection to afford the complete molecule of **106**.

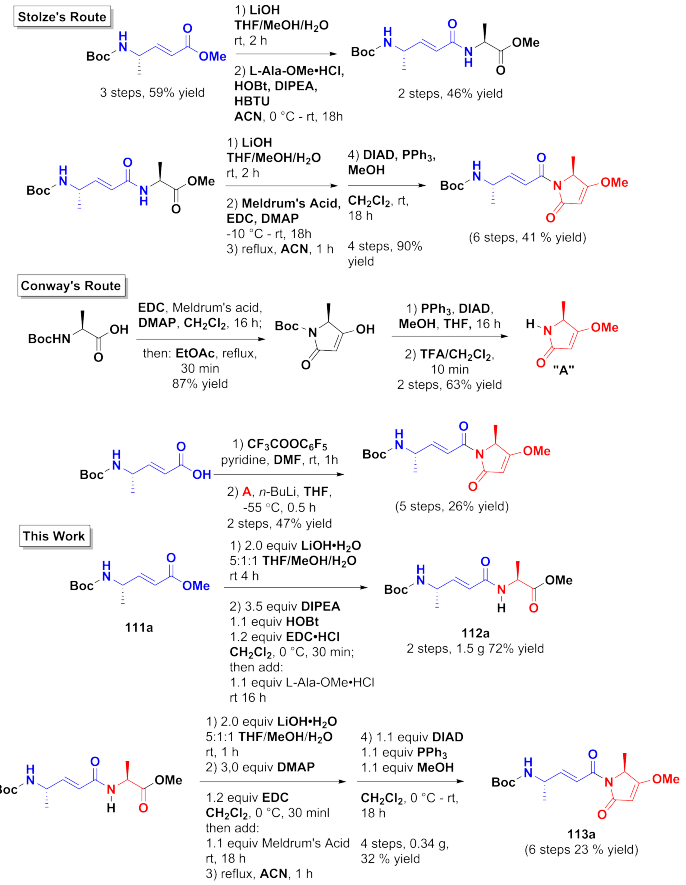


Figure 5.4: *Comparison of synthetic routes towards the gallinamide A (**106**) headgroup.*

5.3.3 Synthesis of the Gallinamide Analogs

By replacing the starting amino acids, a variety of analogs could be synthesized. The modifications focused on four components of (**106**); the final amino acid in the tail, the amino acid in the enamide core, the amino acid in the head group, and the alcohol used to trap the enol in the head group with the Mitsunobu reaction (Figure 6).

To modify the tail, attempts were made to use Fmoc-L-isoleucine, Fmoc-L-leucine, Fmoc-L-valine, and Fmoc-L-phenylalanine. Surprisingly repeated attempts to apply the reaction conditions from the isoleucine case to other amino acids resulted in a poor yield for the phenylalanine case and a low yield and poor purity for the leucine case at the methylation step. This is a remarkable difference in yield because of the nearly trivial difference between

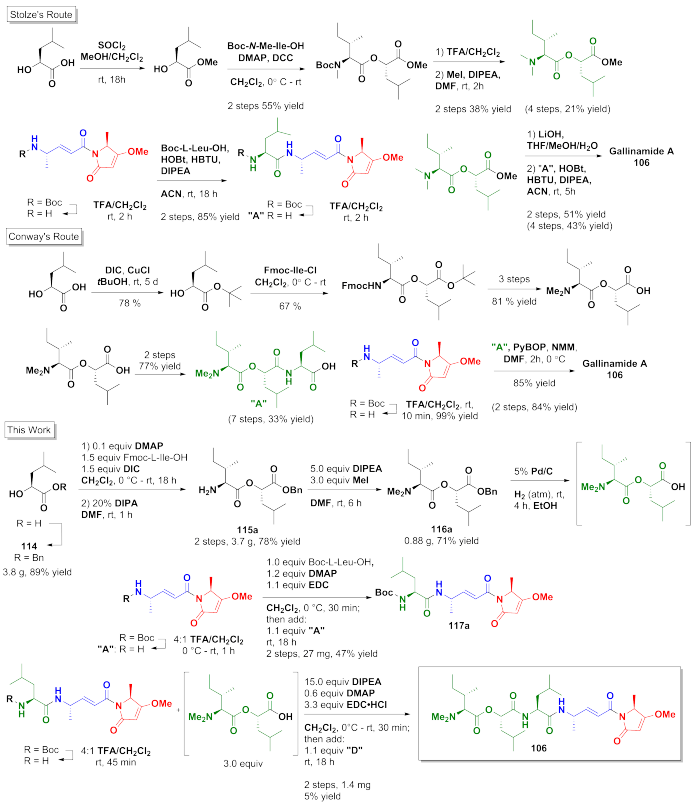


Figure 5.5: *Comparison of synthetic routes towards the gallinamide A (**106**) tail.*

isoleucine and leucine structures. One explanation for this might be a difference in volatility as the product at this step is fully protected and is thus more sensitive to being lost by evaporation during purification.

The enamide core could easily be modified by changing out the starting material Boc-L-alanine for the phenylalanine cases, however, the Weinreb amides derived from phenylalanine could not be crystallized out of *N,N*-dimethyl foramide (DMF), but had to be purified by flash chromatography. In the same fashion, the headgroup extension of the enamide core could be modified by using an amino acid methyl ester other than L-alanine methyl ester. In this work L-alanine, L-leucine, and D/L-phenyl alanine were used, which combined with different enamide starting materials led to nine different core-head group cases. The final modification was the use of an alcohol other than methanol to trap the enol in the Mitsunobu reaction. 5-Hexyn-1-ol was used to create the new analog at this position

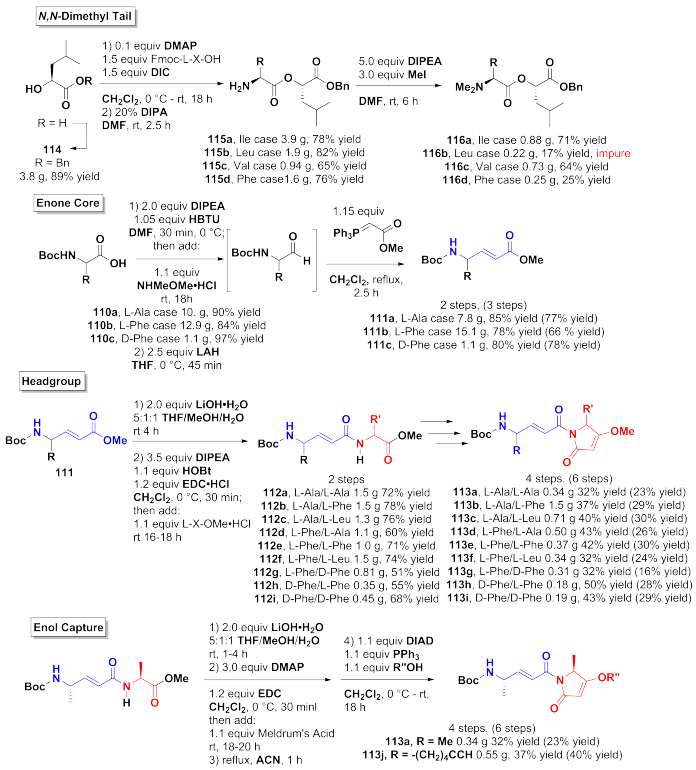


Figure 5.6: *Modifications to the synthesis to provide analog structures.*

as it provided the opportunity to attach a probe via click chemistry at a later step.

5.4 Conclusions and Future Work

The synthetic laurenciones have shown these species to be potent anti-inflammatory and quorum sensing activating activity. Roles which could prove to important in several disease models, by contributing this material to the Gerwick compound pure compound library, these compounds are also available to other researchers who may find related or entirely novel activities for them.

In the gallinamides project, the synthesis of these analogs sought to answer specific questions about the structure activity relationship between this and compound and its activity in the cathepsin inhibition assay. Perhaps the most pressing relates to whether our model of this compound acting as an irreversible covalently bound inhibitor through the attack of

the thiol in the active site to the Michael-accepting enamide is correct. This conjecture is compatible with the observation that **106** acts as a potent irreversible inhibitor of cathepsin L [Miller et al., 2014], but that observation is compatible with other potential mechanisms of action, including a similar thiol attack of the Michael-acceptor system in the trapped enolate within the headgroup (Figure 7). By creating the hydrogenated analog of **106** where the enamide system is removed this hypothesis can be tested, if our hypothesis is correct than removal of the Michael-acceptor would create a species that is no longer an irreversible inhibitor, though due to its high structural homology to the original species, it could be expected to still be a competitive reversible inhibitor.

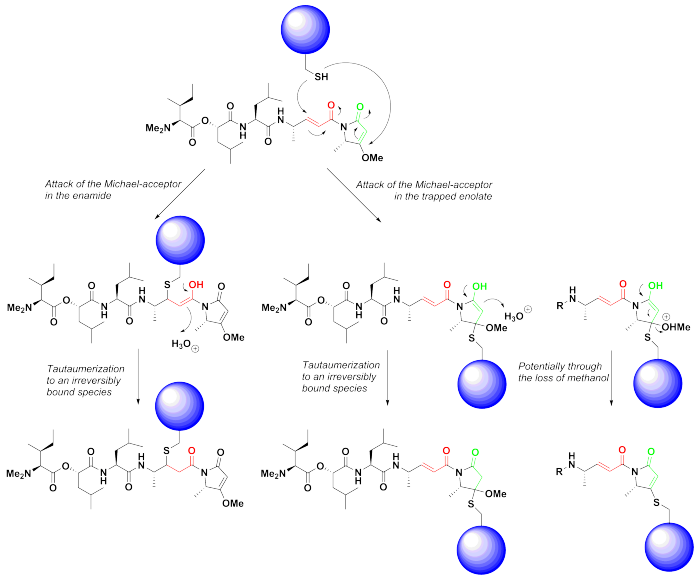


Figure 5.7: *Potential mechanisms of action to explain the irreversible binding of gallinamide A (**106**).*

A major theme of the modeling work carried out previously [Miller et al., 2014], was the suggestion that extension of the alanine groups into the hydrophobic pockets of the enzyme could improve binding. The analogs derived from phenyl alanine and leucine are intended to test this conjecture. The use of leucine and phenyl alanine containing groups will test both the potential of extended functional groups to extend into the enzyme pocket and whether the aromatic rings from phenyl alanine provide an additional benefit

through π-stacking. By using opposing stereochemistry, the modeled result, that *S* oriented stereocenters derived from L amino acids fit best in the enzyme, can be tested. A similar insight can be gained from the modified tail groups, going from the smallest valine derived case to the larger isoleucine and phenylalanine derived cases; where both size of the functional group and the potential of π-stacking to improve binding can be contrasted.

By testing these compounds against the other cathepsin enzymes, it will also be possible to see if any of the structural changes in the analogs improve the selectivity observed in **106** for cathepsin L over V [Miller et al., 2014]. Which is of great importance because, as discussed in that report, there are no good selective inhibitors of these enzymes which are highly homologous and, at present, have no tool available to study *in vivo* the activity of either enzyme in isolation.

Though not specifically designed for the purpose of anti-malarial activity, the original reports of potent activity against the malaria parasite *Plasmodium falciparum*, suggest that submitting the **106** analogs to the *Plasmodium* assay may be a fruitful endeavor. Though it is known that **106** acts against the parasite by inhibiting the falcipain cystine proteases, it remains to be seen how strong the correlation between cathepsin inhibition and falcipain inhibition is.

5.5 Material and Methods

5.5.1 General Procedures

Unless specifically noted otherwise reactions were carried out in flame or oven-dried glassware under Argon, and were stirred by magnetic stir bar, with the exception of reactions run in water as a solvent which were not dried.

5.5.2 Synthesis of the Laurenciones

Laurencione Synthesis. A 25 mL pear flask with stir bar under Ar was charged with 250 μL of 5-hydroxy-2-pentanone and dissolved in 2.20 mL of dioxane. To the reaction mixture, while stirring, was added 73 μL of water and 0.41 g of selenium (IV) oxide in one portion. The reaction mixture was warmed to 57 °C by oil bath, and stirred for 30 min. The reaction mixture was passed through a glass frit with multiple EtOAc washes. This mixture was then concentrated to a reduced volume and diluted into 15 mL of EtOAc. This mixture was washed with 15 mL 10% $NaHCO_3$ (*aq*), and then the aquesous later was extracted with four times equal volume EtOAc, and the combined organic layers were washed with 25 mL of brine and dried over $MgSO_4$. The filtrate was concentrated to an orange-red liquid and purified by flash chromatography with an EtOAc/hexanes elution to afford **105** as 59 mg of an amber oil, 20% yield. The material matched spectroscopically to the previously published reports [Aelterman et al., 1997, Bernart et al., 1992].

Laurencione Monoacetate Synthesis. Ran as with the synthesis of laurencione but dissolved in 2.5 mL of acetic acid without any water. Collected the crude as a red-brown oil and purified by flash chromatography with a EtOAc/hexanes elution to afford 41 mg of laurencione monoacetate (8% yield) with 31 mg of **109** (15% yield). IR (neat) ν_{max} 2970, 1739, 1718, 1421, 1366, 1242, 1090, 1042, 909, 818, 580 cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$) 4.37 (t, J = 5.0), 3.06 (t, J = 5.0), 2.37 (s), 2.03 (s); ^{13}C NMR (125 MHz, $CDCl_3$) 196.8, 196.1, 171.1, 59.1, 35.5, 23.5, 20.9; HR-ESI-TOF-MS $[M+MeOH+Na]^+$ *m/z* 213.0734 (calcd for $C_8H_{14}NaO_4$ 213.0733).

Laurencione Diacetate Synthesis. To a vial containing 35 mg of **105** with an oven-dried stir bar, septum under Ar was added pyridine and 4-dimethylamino-pyridine. This reaction mixture was cooled to 0 °C by ice bath and the acetic anhydride by syringe slowly. The reaction mixture was warmed to rt and stirred for 20 h. The reaction mixture was diluted with 5 mL of dichloromethane and washed with 5 mL sat. $NaHCO_3$ (*aq*). The aqueous layer

was then extracted four times with equal volume dichloromethane (DCM) and the combined organic layers were washed with 5 mL of 5% HCl (*aq*), this aqueous layer was extracted once as previously. The combined organic layers were then washed with 5 mL brine and dried over $MgSO_4$. The filtrate of this was concentrated to a brown oil by rotovap. This material was purified by RP-HPLC with a CH_3CN/H_2O elution. Collected **108** as 6.6 mg of yellow oil. The material matched spectroscopically to the previously published report [Bernart et al., 1992].

5.5.3 Synthesis of Gallinamide Analogs

Weinreb Amide Synthesis. As a general procedure, the Boc-protected amino acid was dissolved in DMF, the reaction mixture was cooled to 0 °C by ice bath and the *N,N*-diisopropyl ethyl amine (DIPEA) was added dropwise. The *N,N,N,N*-tetramethyl-O-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU) was then added in one portion and the reaction mixture was allowed to warm to rt overnight. The alanine product was collected as with the literature procedure by iterative crystallization from cold CH_3CN [Conroy et al., 2010], but the phenylalanine cases did not precipitate out of DMF. In the phenylalanine cases the reactions were worked up by diluting the reaction mixtures into diethyl ether. The dilute mixture was then washed twice with 5% citric acid (*aq*), then brine, and then finally dried over $MgSO_4$, and the filtrate was concentrated to dryness by rotovap. The crude residue was then taken up in hexanes and filtered through a diatomaceous earth plug, concentrated again by rotovap and purified by silica gel flash chromatography with an ethyl acetate/hexanes gradient elution.

tert-butyl (*S*)-(1-(methoxy(methyl)amino)-1-oxopropan-2-yl)carbamate (**110a**). As white crystals. ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.25 (bs, 1H), 4.68 (bs, 1H), 3.77 (s, 3H), 3.21 (s, 3H), 1.44 (s, 9H), 1.31 (d, J = 7.0 Hz, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.7, 155.3, 79.5, 61.7, 46.6, 32.2, 28.4, 18.7. IR (neat) ν_{max} 3295, 2976, 1705, 1660, 1545, 1451,

1363, 1298, 1183, 980, 585 cm^{-1}. λ_{max} (log ε) 205 (6.5) nm, maximum at the end of range. $[\alpha]_D^{24}$ +0.1 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 241.1154, -2.0 ppm (calculated for $C_9H_{18}N_2NaO_4^+$, 241.1159).

tert-butyl (*S*)-(1-(methoxy(methyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**110b**). As a viscous faint-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.28 (m, overlaps with chloroform signal, 2H) 7.22 (m, 1H), 7.17 (d, *J* = 7.4, 2H), 5.16 (bd *J* = 9.1, 1H), 4.95 (bd *J* = 8.21, 1H), 3.66 (s, 3H), 3.17 (s, 3H), 3.05 (dd, *J* = 13.8, 6.1, 1H), 2.87 (dd, *J* = 13.8, 7.2, 1H), 1.39 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.3, 155.1, 136.6, 129.4, 128.3, 126.7, 79.6 , 61.6, 51.5, 38.9 32.1 , 28.3. IR (neat) ν_{max} 3321, 2976, 2935, 1711, 1662, 1498, 1453, 1390, 1367, 1251, 1171 cm^{-1}. λ_{max} (log ε) 205 (6.8) nm, maximum at the end of range. $[\alpha]_D^{24}$ +21.8 (CH_2Cl_2).

tert-butyl (*R*)-(1-(methoxy(methyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**110c**). As a faint-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, overlaps with chloroform signal, 2H) 7.23 (m, 1H), 7.17 (m, 2H), 5.16 (d *J* = 8.6, 1H), 4.95 (q *J* = 7.08, 1H), 3.66 (s, 3H), 3.17 (s, 3H), 3.06 (dd, *J* = 13.8, 6.1, 1H), 2.88 (dd, *J* = 13.7, 7.1, 1H), 1.39 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.2, 155.1, 136.5, 129.4, 128.3, 126.7, 79.6 , 61.6, 51.5, 38.8 32.0 , 28.3. IR (neat) ν_{max} 3426.1, 3323.1, 2976.0, 2934.7, 1709.6, 1661.4, 1500.7, 1452.3, 1377.5, 1329.2, 1251.2, 1170.5, 1021.3, 988.6, 749.6, 700.6 cm^{-1}. λ_{max} (log ε) 205 (7.0) nm, maximum at the end of range. $[\alpha]_D^{25}$ -21.0 (CH_2Cl_2).

Enone Synthesis. As a general procedure, the Weinreb amide was dissolved in tetrahydrofuran (THF), cooled to 0 °C by ice bath, and the lithium aluminum hydride was added in five equal portions. After 45 min the reaction was quenched with minimal ethyl acetate and diluted into dichloromethane, this mixture was washed with 10% KH_2PO_4 (*aq*), using a small amount of brine to assist phase separation. The aqueous phase was back extracted iteratively with dichloromethane, and then the combined organic phases were washed with brine and dried over $MgSO_4$. The filtrate of this step was collected as a yellow

oil and dissolved in dichloromethane (DCM). While stirring the anylidene was added to the reaction mixture in one portion and the reaction was heated to reflux by oil bath for 2.5 h. The reaction mixture was then concentrated to dryness by rotovap and purified by silica gel flash chromatography with an ethyl acetate/hexanes gradient elution.

methyl (*S,E*)-4-((*tert*-butoxycarbonyl)amino)pent-2-enoate (**111a**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.89 (dd, *J* = 15.4, 4.7 Hz, 1H), 5.91 (dd, *J* = 15.7, 1.1 1H), 4.51 (bs, 1H), 4.40 (bs, 1H), 3.74 (s, 3H), 1.45 (s, 9H), 1.27 (d, *J* = 6.9 Hz, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.8, 154.9, 149.7, 119.7, 79.8, 51.6, 47.0, 28.4, 20.3. IR (neat) ν_{max} 3354, 2978, 1712, 1519, 1448, 1369, 1275, 1171, 1046 cm^{-1}. λ_{max} (log ε) 205 (6.4) nm, maximum at the end of range. $[\alpha]_D^{24}$ -18.5 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 252.1204, -0.8 ppm (calculated for $C_{11}H_{19}NNaO_4^+$, 252.1206).

methyl (*S,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoate (**111b**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (t, *J* = 7.4 Hz, 2H), 7.25 (m, overlaps with chloroform signal, 1H), 7.17 (d, *J* = 7.5, 2H), 6.92 (dd, *J* = 15.7, 5.1, 1H), 5.85 (d, *J* = 15.5, 1H), 4.62 (bs, 1H), 4.53 (bs, 1H), 3.72 (s, 3H), 2.89 (d, *J* = 6.7, 2H), 1.40 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 154.9, 147.9, 136.3, 129.4, 128.6, 126.9, 120.7, 79.9, 52.3, 51.7, 40.8, 28.3. IR (neat) ν_{max} 3352, 2978, 1712, 1516, 1439, 1367, 1282, 1168, 1023, 702 cm^{-1}. λ_{max} (log ε) 205 (7.7) nm, maximum at the end of range. $[\alpha]_D^{25}$ -1.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 328.1516, -0.9 ppm (calculated for $C_{17}H_{23}NNaO_4^+$, 328.1519).

methyl (*R,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoate (**111c**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (m, 2H), 7.24 (m, 1H), 7.17 (m, 2H), 6.91 (dd, *J* = 5.1, 15.7, 1H), 5.86 (dd, *J* = 1.8, 15.7, 1H), 4.62 (bs, 1H), 4.50 (bs, 1H), 3.73 (s, 3H), 2.89 (d, *J* = 6.8, 2H), 1.40 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 166.6, 154.9, 147.9, 136.3, 129.4, 128.6, 126.9, 120.7, 79.9, 52.3, 51.6, 40.8, 28.3. IR (neat) ν_{max} 3358.6, 2976.8, 2941.2, 1714.8, 1515.3, 1444.9, 1365.2, 1279.9, 1255.0, 1168.5, 1023.1 cm^{-1}. λ_{max}

(log ε) 205 (7.3) nm, maximum at the end of range. $[\alpha]_D^{25}$ -4.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 328.1517, -0.6 ppm (calculated for $C_{17}H_{23}NNaO_4^+$, 328.1519).

Headgroup Synthesis. As a general procedure, the starting ester was dissolved into the 5:1:1 THF/MeOH/H_2O solution and the lithium hydroxide was added in one portion, the reaction mixture was then vigorously stirred for 3.25-4 h during which the color of the solution changed from colorless to yellow-orange. The reaction mixture was then diluted with DCM and acidified by washing with 5% citric acid (*aq*); the aqueous layer was back extracted twice with DCM, and then the combined organic layers were washed with brine and then dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap and diluted into DCM. The DIPEA was then added dropwise, the hydroxybenzotriazole (HOBt) was added in one portion, and this solution was cooled to 0 °C by ice bath before the EDC was added in one portion. After 30 min at 0 °C the amino acid methyl ester, as a hydrochloride salt, was added in one portion. The reaction was then allowed to warm to rt overnight. The reaction mixture was then worked up by diluting into DCM and washed with with 5% citric acid (*aq*), back extracted three times with DCM, and then the organics were washed with brine and dried over $MgSO_4$. The filtrate of this step was then dried by rotovap and purified by iterative silica gel flash chromatography with a 100:1 DCM/methanol elution.

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)pent-2-enoyl)-*L*-alaninate (**112a**). As amorphous white solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.75 (dd, *J* = 5.7, 15.2 Hz, 1H), 6.14 (d, *J* = 6.6, 1H), 5.90 (d, *J* = 15.3, 1H), 4.67 (p, J = 7.2, 1H), 4.53 (m, 1H), 4.37 (m, 1H), 3.77 (s, 3H), 1.45 (m, 11H), 1.26 (d, *J* = 6.9, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.5, 164.8, 154.9, 145.7, 122.2, 79.7, 52.5, 48.1, 47.1, 28.4, 20.6, 18.5. IR (neat) ν_{max} 3347, 2978, 2937, 1722, 1628, 1516, 1454, 1330, 1292, 1247, 1175, 1051, 971 cm^{-1}. λ_{max} (log ε) 208 (6.9) nm. $[\alpha]_D^{25}$ -25.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 323.1575, -0.6 ppm (calculated for $C_{14}H_{24}N_2NaO_5^+$, 323.1577).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)pent-2-enoyl)-*L*-phenylalaninate (**112b**). As amorphous off-white solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.28 (m, overlaps with chloroform signal, 3H), 7.09 (d, *J* = 6.8, 2H), 6.74 (dd, *J* = 15.3, 5.4, 1H), 6.00 (d, *J* = 7.4, 1H), 5.86 (d, *J* = 15.3, 1H), 4.95 (dt, *J* = 7.6, 5.7, 1H), 4.52 (m, 1H), 4.38 (m, 1H), 3.73 (s, 3H), 3.16 (t, *J* = 6.2, 2H), 1.44 (s, 9H), 1.25 (d, *J* = 6.9, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.9, 164.9, 154.9, 146.1, 135.8, 129.3, 128.6, 127.2, 121.9, 79.7, 53.2, 52.4, 47.0, 37.9, 28.4, 20.6. IR (neat) ν_{max} 3345, 2977, 2937, 1722, 1630, 1629, 1515, 1452, 1329, 1247, 1170 cm^{-1}. λ_{max} (log ε) 208 (7.1) nm. $[\alpha]_D^{25}$ +30.8 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 399.1888, -0.5 ppm (calculated for $C_{20}H_{28}N_2NaO_5^+$, 399.1890).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)pent-2-enoyl)-*L*-leucinate (**112c**). As a light-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.75 (dd, *J* = 15.2, 5.7, 1H), 6.01 (d, *J* = 8.4, 1H), 5.92 (d, *J* = 15.2, 1H), 4.71 (td, *J* = 8.6, 4.4, 1H), 4.58 (m, 1H), 4.39 (m, 1H), 3.75 (s, 3H), 1.67 (m, 2H), 1.57 (m, 1H), 1.45 (s, 9H), 1.26 (d, *J* = 6.9, 3H), 0.95 (t, *J* = 6.5, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.6, 165.1, 154.9, 145.8, 122.1, 79.7, 52.9, 50.7, 47.0, 41.8, 28.4, 24.8, 22.8, 21.9, 20.5. IR (neat) ν_{max} 3308.0, 3058.2, 2965.8, 2877.6, 1741.1, 1682.6, 1641.1, 1528.9, 1449.8, 1368.2, 1249.6, 1169.0, 1055.3, 981.0 cm^{-1}. λ_{max} (log ε) 205 (6.8) nm, maximum at the end of range. $[\alpha]_D^{25}$ -25.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 365.2043, -1.1 ppm (calculated for $C_{17}H_{30}N_2NaO_5^+$, 365.2047).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*L*-alaninate (**112d**). As amorphous off-white solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (t, *J* = 7.4, 2H), 7.23 (t, *J* = 7.4, 1H), 7.18 (d, *J* = 7.5, 2H), 6.80 (m, 1H), 6.05 (d, *J* = 7.4, 1H), 5.85 (d, *J* = 15.4, 1H), 4.65 (p, *J* = 7.4, 1H), 4.59 (bs, 1H), 4.52 (bs, 1H), 3.76 (s, 3H), 2.90 (m, 2H), 1.42 (d, *J* = 7.2, 3H), 1.40 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.4, 164.5, 155.0, 144.0, 136.5, 129.4, 128.5, 126.8, 123.1, 79.8, 52.6, 52.3, 48.1, 41.0, 28.3, 18.6. IR (neat) ν_{max} 3334.4, 3023.7, 2983.5, 2936.2, 1738.7, 1676.5, 1634.7, 1528.4, 1448.5, 1370.8, 1318.3, 1267.3, 1214.4, 1166.7, 1049.8, 1024.0, 973.8, 699.4, 654.3 cm^{-1}. λ_{max} (log ε) 205 (6.6)

nm, maximum at the end of range. $[\alpha]_D^{25}$ -3.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 399.1889, -0.2 ppm (calculated for $C_{20}H_{28}N_2NaO_5^+$, 399.1890).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*L*-phenylalaninate (**112e**). As amorphous off-white solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.27 (m, 4H), 7.23 (m, 2H), 7.16 (m, 2H), 7.07 (m, 2H), 6.80 (dd, *J* = 15.2, 5.3, 1H), 5.97 (m, 1H), 5.79 (dd, *J* = 15.3, 1.6, 1H), 4.93 (dt, *J* = 7.7, 5.7, 1H), 4.59 (bs, 1H), (m, 1H), 3.72 (s, 3H), 3.14 (qd, J = 13.9, 5.7, 1H), 2.88 (m, 1H), 1.39 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.9, 164.6, 154.9, 144.2, 136.5, 135.8, 129.5, 129.3, 128.6, 128.5, 127.1, 126.8, 122.9, 79.8, 53.2, 52.4, 52.3, 41.0, 37.8, 28.3. IR (neat) ν_{max} 3338.5, 3028.4, 2976.6, 1741.4, 1680.9, 1638.1, 1528.6, 1445.9, 1365.5, 1325.6, 1256.8, 1212.1, 1170.4, 699.1 cm^{-1}. λ_{max} (log ε) 205 (7.1) nm, maximum at the end of range. $[\alpha]_D^{25}$ +48.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 475.2207, 0.8 ppm (calculated for $C_{26}H_{32}N_2NaO_5^+$, 475.2203).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*L*-leucinate (**112f**). As amorphous off-white solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 2H), 7.24 (m, 1H), 7.18 (m, 2H), 6.80 (m, 1H), 5.84 (m, 2H), 4.70 (td, *J* = 8.41, 4.6, 1H), 4.60 (m, 1H), 4,52 (m, 1H), 3.74 (s, 3H), 2.89 (m, 2H), 1.65 (m, 2H), 1.56 (m, 1H), 1.40 (s, 9H), 0.94 (t, *J* = 7.0, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.4, 164.7, 154.9, 144.1, 136.4, 129.5, 128.5, 126.8, 123.0, 79.8, 52.4, 52.3, 50.7, 41.8, 41.0, 28.3, 24.8, 22.8, 22.0. IR (neat) ν_{max} 3343.1, 3025.1, 2960.7, 2874.4, 1744.4, 1679.0, 1638.0, 1528.9, 1446.3, 1367.5, 1325.1, 1257.8, 1206.0, 1168.6, 1021.1, 979.5, 698.3, 651.1 cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at the end of range. $[\alpha]_D^{25}$ -6.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 441.2367, 1.6 ppm (calculated for $C_{23}H_{34}N_2NaO_5^+$, 441.2360).

methyl ((*S,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*D*-phenylalaninate (**112g**). As amorphous light-yellow solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, 4H), 7.25 (m, 2H), 7.17 (d, *J* = 7.2, 2H), 7.05 (m, 2H), 6.81 (dd, *J* = 15.2, 5.1, 1H), 5.92 (d, *J* = 7.7, 1H) 5.77 (dd, *J* = 15.3, 1.7, 1H), 4.95 (dt, *J* = 7.8, 5.5, 1H), 4.60 (m, 1H),

4.52 (m, 1H), 3.73 (s, 3H), 3.14 (dd, *J* = 5.7, 3.96, 2H), 2.88 (d, *J* = 6.7, 2H), 1.39, (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.8, 164.4, 155.0, 144.2, 136.4, 135.6, 129.5, 129.2, 128.6, 128.5, 127.1, 126.8, 122.8, 79.8, 53.1, 52.4, 52.3, 41.1, 37.8, 28.3. IR (neat) ν_{max} 3330.9, 3029.8, 2977.8, 1741.8, 1679.4, 1638.5, 1526.1, 1445.8, 1365.6, 1252.0, 1212.2, 1169.9, 1021.9, 742.2, 699.4 cm^{-1}. λ_{max} (log ε) 205 (7.2) nm, maximum at the end of range. $[\alpha]_D^{25}$ -54.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 475.2209, 1.3 ppm (calculated for $C_{26}H_{32}N_2NaO_5^+$, 475.2203).

methyl ((*R,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*L*-phenylalaninate (**112h**). As off-white amorphous solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, 4H), 7.24 (m, 2H), 7.17 (d, *J* = 7.3, 2H), 7.05 (d, *J* = 6.6, 2H), 6.81 (dd, *J* = 15.2, 5.2, 1H), 5.91 (d, *J* = 7.8, 1H), 5.77 (d, *J* = 15.4, 1H), 4.95 (dt, *J* = 7.8, 5.5, 1H), 4.60 (m, 1H), 4.52 (m, 1H), 3.73 (s, 3H), 3.15 (t, *J* = 4.8, 2H), 2.88 (d, *J* = 6.7, 2H), 1.39 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.8, 164.4, 155.0, 144.2, 136.4, 135.6, 129.5, 129.3, 128.6, 128.5, 127.1, 126.8, 122.8, 79.8, 53.1, 52.4, 52.3, 41.1, 37.8, 28.3. IR (neat) ν_{max} 3335.2, 3025.6, 2980.0, 2946.9, 1747.5, 1676.0, 1638.3, 1528.2, 1445.0, 1366.9, 1323.4, 1258.1, 1212.7, 1168.6, 698.9 cm^{-1}. λ_{max} (log ε) 205 (7.1) nm, maximum at the end of range. $[\alpha]_D^{25}$ +59.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 475.2207, 0.8 ppm (calculated for $C_{26}H_{32}N_2NaO_5^+$, 475.2203).

methyl ((*R,E*)-4-((*tert*-butoxycarbonyl)amino)-5-phenylpent-2-enoyl)-*D*-phenylalaninate (**112i**). As off-white amorphous solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 4H), 7.23 (m, 2H), 7.16 (d, *J* = 7.59, 2H), 7.06 (d, *J* = 7.51, 2H), 6.81 (dd, *J* = 15.1, 5.1, 1H), 5.90 (d, *J* = 7.2, 1H), 5.78 (d, *J* = 15.2, 1H), 4.93 (q, *J* = 5.9, 1H), 4.61 (m, 1H), 4.50 (m, 1H), 3.73 (s, 3H), 3.14 (m, 2H), 2.89 (m, 2H), 1.39 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 171.8, 164.5, 154.9, 144.2, 136.4, 135.7, 129.5, 129.2, 128.6, 128.5, 127.1, 126.8, 122.8, 79.8, 53.2, 52.4, 52.2, 41.0, 37.7, 28.3. IR (neat) ν_{max} 3336.1, 3029.0, 2976.2, 1741.0, 1679.8, 1637.8, 1528.0, 1445.6, 1365.5, 1324.0, 1256.0, 1211.6, 1170.1, 698.2 cm^{-1}. λ_{max}

(log ε) 206 (7.3) nm. $[\alpha]_D^{25}$ -55.5 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 475.2209, 1.3 ppm (calculated for $C_{26}H_{32}N_2NaO_5^+$, 475.2203).

Headgroup Cyclization and Mitsunobu. As a general procedure, the starting ester was dissolved into the 5:1:1 THF/MeOH/H_2O solution and the lithium hydroxide was added in one portion, the reaction mixture was then stirred vigorously for 1-2.5 h. The reaction mixture was then diluted into DCM and acidified by washing with 5% citric acid (*aq*); the aqueous layer was back extracted iteratively with DCM, and then the combined organic layers were washed with brine then dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap and diluted into DCM; the pyridine was then added in one portion and the reaction mixture was cooled to 0 °C by ice bath and the EDC was added in one portion. After at 0 °C 30 min the Meldrum's acid was added in one portion, and the reaction mixture was allowed to warm to rt overnight. The reaction mixture was then worked up by diluting into DCM, washing twice with 5% citric acid (*aq*), and once with brine before being dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap and dissolved in CH_3CN, this solution was brought to reflux by oil bath for 1 h. The reaction mixture was then concentrated to dryness by rotovap and dissolved in DCM, to this mixture was added PPh_3 in one portion and the alcohol dropwise. The reaction mixture was then cooled to 0 °C by ice bath and the diisopropyl azodicarboxylate (DIAD) was added dropwise, the reaction mixture was then allowed to warm to rt overnight. As a workup the reaction mixture was simply concentrated to dryness by rotovap and purified by silica gel flash chromatography with wa gradient elution of EtOAc/hexanes. The material collected from this column was further subjugated to reverse-phase solid phase extraction (RP-SPE) with a 10 g C-18 column using elutions of 20%/40%/60% CH_3CN/H_2O and CH_3CN. The fraction that eluted with 60% was collected and dried by rotovap to afford the product.

It was noted that after this reaction series the products showed an unusual 'twinning'

of some peaks on the ^{13}C-NMR spectrum, though initially expected to be from poor tuning of the proton channel leading to incomplete decoupling, this was not the case. As both quaternary carbons displayed pronounced splitting, and as extensive efforts to improve the tunning did not resolve the issue, it seemed reasonable to assume that something else was responsible for the *pseudo*-coupling of some of the carbon signals.

The *pseduo*-coupling may also be present in proton spectrum where under variable temperature the coupling constants of the signals change, at low temperature there are three distinct couplings to the signal at 7.1 ppm as a ddd, while at elevated temperature there are only two coupling as a dt; the doublet signal from the methyl enolate also converges with increasing temperature (Figure 8).

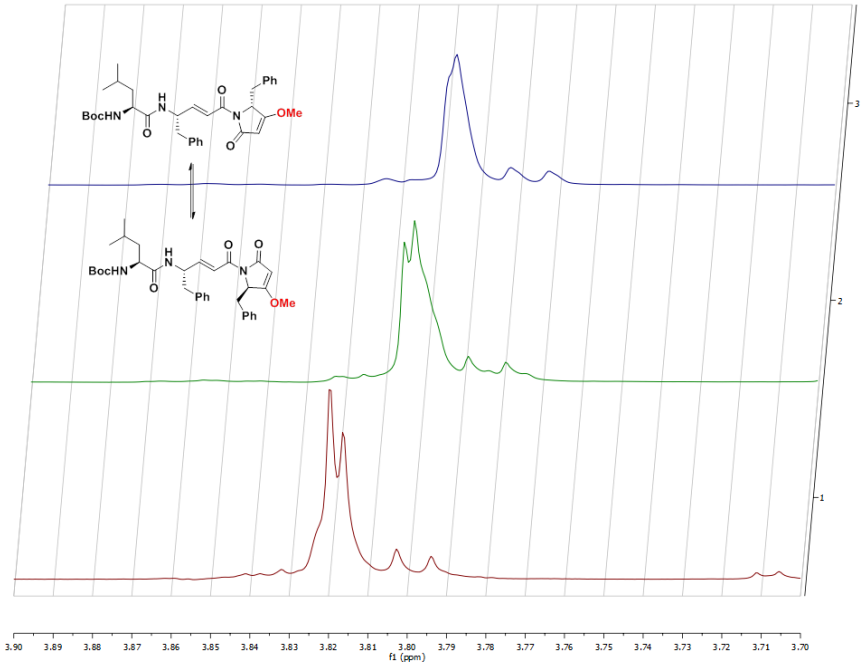


Figure 5.8: *Variable temperature* ^{1}*H-NMR of* ***117g***. *Top Spectrum* Sample at 55 °C in $CDCl_3$. *Middle Spectrum* Sample at 45 °C in $CDCl_3$. *Bottom Specturm* Sample at 22 °C in $CDCl_3$. The methyl signal doublet converges with increased temperature.

tert-butyl ((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)carbamate (**113a**). As a colorless foam. ^{1}H NMR (500 MHz, $CDCl_3$)

δ 7.41 (m, 1H), 7.02 (m, 1H), 5.62 (m, 1H), 4.48 (bs, 1H), 3.87 (s, 3H), 1.49 (dd, *J* = 6.5, 2.6, 3H), 1.46 (s, 9H), 1.30 (d, *J* = 6.9, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.7*, 198.8, 164.4*, 155.0*, 149.5*, 121.6*, 93.0, 79.6, 58.7, 55.7, 47.3*, 28.4, 20.4, 17.1*. *Showed twinning. ν_{max} 3345.3, 3107.4, 2978.3, 2936.7, 1718.7, 1627.8, 1516.8, 1453.9, 1332.0, 1290.1, 1246.5, 1174.9, 1053.5, 975.7, 811.6, 701.4 cm^{-1}. λ_{max} (log ε) 243 (6.8) nm. $[\alpha]_D^{25}$ -20.5 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 347.1575, -0.6 ppm (calculated for $C_{16}H_{24}N_2NaO_5^+$, 347.1577).

tert-butyl ((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)carbamate (**113b**). As faint-yellow semi-solids. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (m, 1H), 7.20 (m, 3H), 7.09 (m, 1H), 6.94 (m, 2H), 4.91 (m, 1H), 4.82 (s, 1H), 4.61 (s, 1H), 4.49 (s, 1H), 3.81 (s, 3H), 3.55 (m, 1H), 3.12 (dd, *J* = 14.0, 3.0, 1H), 1.45 (s, 9H), 1.32 (d, *J* = 6.9, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.8, 169.8, 164.7*, 155.0*, 149.8*, 134.3*, 129.6, 128.2, 127.0*, 121.5, 95.0*, 79.0, 59.7*, 58.4*, 47.3, 34.6*, 28.4*, 20.5*. *Showed twinning. ν_{max} 3343.1, 2976.8, 2935.2, 1718.8, 1629.4, 1512.2, 1452.1, 1337.0, 1246.5, 1171.0 cm^{-1} λ_{max} (log ε) 205 (6.9) nm, maximum at end of range. $[\alpha]_D^{25}$ -7.3 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 423.1895, 1.2 ppm (calculated for $C_{22}H_{28}N_2NaO_5^+$, 423.1890).

tert-butyl ((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)carbamate (**113c**). As a yellow foam. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 (ddd, *J* = 15.5, 10.7, 1.7, 1H), 7.00 (m, 1H), 5.05 (s, 1H), 4.69 (dt, *J* = 6.4, 3.0, 1H), 4.65 (m, 1H), 4.47 (s, 1H), 3.86 (s, 3H), 1.83 (m, 2H), 1.76 (m, 1H), 1.45 (d, *J* = 1.2, 9H), 1.30 (d, *J* = 6.9, 3H), 0.92 (d, *J* = 6.4, 3H), 0.88 (d, *J* = 6.4, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.5*, 170.1, 164.4*, 154.9*, 149.3*, 121.7*, 93.5*, 79.4*, 58.5*, 58.4, 47.2, 39.0, 28.3, 24.1*, 23.6*, 22.6*, 20.3*. *Showed twinning. ν_{max} 3349.5, 2968.7, 2876.5, 1718.8, 1626.5, 1515.4, 1455.6, 1330.9, 1245.9, 1173.5, 1053.8, 982.8 cm^{-1}. λ_{max} (log ε) 234 (6.6) nm. $[\alpha]_D^{27}$ -12.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 389.2046, -0.3 ppm

(calculated for $C_{19}H_{30}N_2NaO_5^+$, 389.2047).

tert-butyl ((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)carbamate (**113d**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 (ddd, *J* = 15.5, 9.1, 1.6, 1H), 7.29 (m, 2H), 7.20 (m, 3H), 7.06 (m, 1H), 5.03 (s, 1H), 4.68 (m, 1H), 4.61 (p, *J* = 6.3 1H), 4.54 (m, 1H), 3.86 (s, 3H), 2.92 (m, 2H), 1.48 (t, *J* = 6.4, 3H), 1.40 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.7, 169.7, 164.1*, 155.0, 147.8*, 136.7*, 129.4*, 128.5, 126.7, 122.5*, 93.0, 79.7, 58.7, 55.7, 52.6, 41.0, 28.3, 17.1*. *Showed twinning. ν_{max} 3345.6, 2977.7, 2935.3, 1717.1, 1627.2, 1509.5, 1452.6, 1333.7, 1290.1, 1247.9, 1172.5, 1097.9, 1039.9, 973.1, 701.1 cm^{-1}. λ_{max} (log ε) 210 (7.8) nm. $[\alpha]_D^{25}$ -2.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 423.1890, 0.0 ppm (calculated for $C_{22}H_{28}N_2NaO_5^+$, 423.1890).

tert-butyl ((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)carbamate (**113e**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (m, 3H), 7.19 (m, 5H), 6.93 (m, 2H), 4.90 (m, 1H), 4.80 (s, 1H), 4.61 (s, 1H), 3.82 (s, 3H), 3.52 (m, 1H), 3.12 (m, 1H), 2.90 (m, 2H), 1.40 (d, *J* = 2.1, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.7*, 169.6, 164.3, 154.8*, 148.0*, 136.6*, 134.1*, 129.5, 129.4, 128.4, 128.1, 126.9, 126.6*, 94.9*, 79.6*, 59.6(1), 59.5(9), 58.3, 53.1*, 52.3, 41.0, 40.8, 37.7*, 34.4, 28.2. ν_{max} 3343.1, 3027.9, 2976.7, 2936.3, 1716.8, 1629.9, 1505.5, 1449.3, 1358.1, 1248.4, 1169.6, 1020.8, 969.6, 737.7, 701.2 cm^{-1}. λ_{max} (log ε) 205 (7.4) nm, maximum at end of range. $[\alpha]_D^{25}$ 1.0 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 499.2205, 0.4 ppm (calculated for $C_{28}H_{32}N_2NaO_5^+$, 499.2203).

tert-butyl ((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)carbamate (**113f**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 (m, 1H), 7.29 (t, *J* = 7.4, 2H), 7.22 (m, 1H), 7.19 (m, 2H), 7.05 (m, 1H), 5.04 (s, 1H), 4.69 (dt, *J* = 6.9, 3.5, 1H), 4.62 (m, 1H), 4.57 (m, 1H), 3.85 (d, *J* = 1.2, 3H), 2.94 (m, 2H) 1.83 (m, 2H), 1.73 (m, 1H), 1.40 (s, 9H), 0.92 (dd, *J* = 6.5, 2.6, 3H), 0.87 (dd,

J = 6.5, 3.2, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.6*, 170.2*, 164.2, 155.0*, 147.7, 136.7*, 129.5*, 128.5*, 126.7, 122.5*, 93.6*, 79.7, 58.6*, 58.5(4), 58.5(0), 52.6*, 41.0*, 38.9*, 28.3, 24.2*, 23.7*, 22.7*. *Showed twinning. ν_{max} 3346.4, 2959.9, 2872.3, 1719.6, 1625.6, 1505.0, 1454.1, 1333.7, 1250.1, 1173.3, 980.9 cm^{-1}. λ_{max} (log ε) 210 (6.6) nm. $[\alpha]_D^{25}$ 6.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 465.2364, 0.9 ppm (calculated for $C_{25}H_{34}N_2NaO_5^+$, 465.2360).

tert-butyl ((*S,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)carbamate (**113g**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (m, 3H), 7.19 (m, 5H), 6.94 (m, 2H), 4.90 (td, *J* = 4.7, 3.1, 1H), 4.80 (s, 1H), 4.71 (s, 1H), 4.62 (s, 1H), 3.82 (s, 3H), 3.53 (m, 1H), 3.11 (m, 1H), 2.93 (m, 2H), 1.40 (d, *J* = 2.2, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.8*, 169.7, 164.4*, 155.0*, 148.1*, 136.7*, 134.2*, 129.6, 129.5, 128.5*, 128.2, 126.7*, 122.3*, 95.0*, 79.7, 59.7*, 58.4, 52.6*, 41.0*, 34.5, 28.3. *Showed twinning. ν_{max} 3338.0, 3026.0, 2976.9, 2934.4, 1718.5, 1628.8, 1502.2, 1451.0, 1346.7, 1248.4, 1170.6, 1120.5, 1020.7, 969.2, 737.9, 700.7 cm^{-1}. λ_{max} (log ε) 205 (7.2) nm, maximum at end of range. $[\alpha]_D^{25}$ 2.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 499.2205, 0.4 ppm (calculated for $C_{28}H_{32}N_2NaO_5^+$, 499.2203).

tert-butyl ((*R,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)carbamate (**113h**). As an amber oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 5H), 7.19 (m, 5H), 6.92 (m, 2H), 4.94 (m, 1H), 4.89 (m, 1H), 4.80 (s, 1H), 4.62 (m, 1H) 3.82 (s, 3H), 3.57 (m, 1H), 3.13 (m, 1H), 2.90 (m, 2H), 1.40 (d, *J* = 2.2, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.7, 169.6, 164.3, 154.8, 148.0*, 136.6*, 134.1*, 129.5*, 129.4, 128.4, 128.1, 126.9, 126.6*, 122.5*, 94.9*, 79.6, 59.6*, 58.3, 53.1*, 41.4*, 34.4, 28.2. *Showed twinning. ν_{max} 3336.1, 3028.5, 2977.4, 2936.2, 1718.2, 1630.1, 1505.5, 1450.2, 1356.0, 1248.0, 1170.5, 1117.1, 1021.2, 970.0, 737.2, 701.4 cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at end of range. $[\alpha]_D^{25}$ -6.5 (CH_2Cl_2).

tert-butyl ((*R,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-

oxo-1-phenylpent-3-en-2-yl)carbamate (**113i**). As an amber oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 5H), 7.19 (m, 5H), 6.92 (m, 2H), 4.94 (m, 1H), 4.89 (m, 1H), 4.80 (s, 1H), 4.62 (m, 1H), 3.82 (s, 3H), 3.57 (m, 1H), 3.13 (m, 1H), 2.90 (m, 2H), 1.40 (d, J = 2.2, 9H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.8, 169.7, 164.4, 155.0, 1481*, 136.5*, 134.2*, 129.6*, 129.5, 128.5*, 128.2, 127.0, 126.7*, 122.6*, 95.0*, 79.8*, 59.7*, 58.4, 53.2*, 41.4*, 34.5, 28.3. *Showed twinning. ν_{max} 3338.4, 3028.5, 2977.4, 2935.9, 1718.0, 1629.8, 1505.0, 1450.1, 1356.3, 1247.5, 1169.9, 1020.8, 969.5, 737.5, 700.8 cm^{-1}. λ_{max} (log ε) 205 (7.1) nm, maximum at end of range. $[\alpha]_D^{25}$ -7.8 (CH_2Cl_2).

tert-butyl ((*S,E*)-5-((*S*)-3-(hex-5-yn-1-yloxy)-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)carbamate (**113j**). As a light-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 (ddd, *J* = 15.5, 7.1, 1.7, 1H), 7.04 (m, 1H), 5.01 (s, 1H), 4.62 (m, 2H), 4.47 (bs, 1H), 4.06 (m, 1H), 4.01 (m, 1H), 2.28 (td, *J* = 6.9, 2.7, 2H), 1.99 (t, *J* = 2.7, 1H), 1.93 (m, 2H), 1.68 (m, 2H), 1.50 (dd, *J* = 6.6, 2.2, 3H), 1.46 (s, 9H), 1.30 (d, *J* = 6.9, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 179.7*, 169.9, 164.3*, 155.0*, 149.5*, 121.6*, 93.1, 83.4, 79.6, 71.4, 69.1, 55.8, 47.4*, 28.4, 27.4, 24.6, 20.4, 18.0, 17.2*. *Showed twinning. ν_{max} 3303.8, 3106.7, 2975.6, 2938.8, 1717.1, 1624.2, 1514.5, 1455.5, 1330.2, 1288.8, 1242.5, 1174.4, 1047.1, 976.3, 934.9, 811.4, 703.9, 638.3 cm^{-1}. λ_{max} (log ε) 244 (7.1) nm. $[\alpha]_D^{25}$ -19.8 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 413.2051, 1.0 ppm (calculated for $C_{21}H_{30}N_2NaO_5^+$, 413.2047).

Leucine Extension. As a general procedure, the carbamate starting material was dissolved in DCM and cooled to 0 °C by ice bath before the trifluoroacetic acid (TFA) was added dropwise to the solution. After 1 h the reaction mixture was concentrated to dryness by rotovap and then azeotroped three times with three drops of toluene by rotovap. In a separate flask, the Boc-L-Leu-OH was dissolved in DCM; to this mixture was added the DIPEA, dropwise, and the HOBt, in one portion. The leucine reaction mixture was cooled to 0 °C by ice bath and the EDC was added in one portion. After 30 min at 0 °C the deprotected

amine was added as a solution in DCM with iterative washes of DCM. The reaction mixture was then allowed to warm to rt overnight. The reaction mixture was worked up by dilution into DCM, washing twice with 5% citric acid (*aq*), and once with brine before drying over $MgSO_4$. The filtrate of this step was then purified by silica gel flash chromatography with an EtOAc/hexanes elution. The material collected from this column was further subjected to RP-SPE with a 10 g C-18 column using elutions of 25%/40%/75 or 80% CH_3CN/H_2O and CH_3CN. The fraction that eluted with 75 or 80% was collected and dried by rotovap.

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117a**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 (ddd, *J* = 15.6, 9.5, 1.7, 1H), 6.99 (m, 1H), 6.17 (m, 1H), 5.03 (s, 1H), 4.89 (m, 1H), 4.77 (m, 1H), 4.62 (qd, *J* = 6.6, 3.1, 1H), 4.08 (m, 1H), 3.87 (s, 3H), 1.69 (m, 2H), 1.49 (m, 4H), 1.44 (s, 9H), 1.32 (d, *J* = 7.0, 3H), 0.95 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.7, 171.8*, 169.7*, 164.2*, 155.8, 148.3*, 122.4*, 93.0, 80.0, 58.7, 55.7*, 53.0, 46.0*, 41.1*, 28.3, 24.7*, 23.0, 22.1*, 20.0*, 17.1*. *Showed twinning. ν_{max} 3313.0, 2968.7, 2876.5, 1719.8, 1666.9, 1629.1, 1529.4, 1455.8, 1358.7, 1330.6, 1288.5, 1245.5, 1173.4, 1039.2, 976.3 cm^{-1}. λ_{max} (log ε) 242 (7.0) nm. $[\alpha]_D^{25}$ -35.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 460.2425, 1.5 ppm (calculated for $C_{22}H_{35}N_3NaO_6^+$, 460.2418).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117b**). As a light amber-colored oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (t, J = 14.5, 1H), 7.19 (m, 3H), 7.07 (dd, J = 15.6, 4.9, 1H), 6.94 (m, 2H), 6.42 (m, 1H), 5.01 (m, 1H), 4.90 (m, 1H), 4.81 (m, 2H), 4.12 (m, 1H), 3.83 (s, 3H), 3.55 (m, 1H), 3.45 (m, 1H), 3.12 (m, 1H), 1.66 (m, 2H), 1.50 (m, 1H), 1.43 (s, 9H), 1.33 (d, J = 6.5, 1H), 0.92 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.8*, 172.0, 169.7, 164.6*, 155.9*, 148.9*, 134.2*, 129.5, 128.2*, 127.0*, 122.0*, 95.0*, 80.0, 59.7*, 58.4*, 53.0, 45.9, 41.2, 34.5*, 31.6, 28.3, 24.7, 22.9, 22.0, 19.9*. *Showed twinning.

ν_{max} 3308.2, 3062.9, 3032.2, 2968.2, 2875.7, 1722.1, 1668.2, 1631.9, 1530.2, 1452.8, 1357.5, 1246.7, 1171.6, 1118.9, 1025.6, 970.3, 806.7, 735.0, 702.7, 642.0 cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at end of range. $[\alpha]_D^{25}$ -20.0 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z* 536.2738, 1.3 ppm (calculated for $C_{28}H_{39}N_3NaO_6^+$, 536.2731).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117c**). As a light-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 (ddd, J = 15.6, 11.7, 1.8, 1H), 6.98 (ddd, J = 15.6, 7.0, 4.9, 1H), 6.2 (m, 1H), 5.05 (s, 1H), 4.92 (m, 1H), 4.78 (m, 1H), 4.69 (dt, J = 6.9, 3.5, 1H), 4.09 (m, 1H), 3.86 (s, 3H), 1.82 (m, 2H), 1.68 (m, 2H), 1.50 (m, 1H), 1.44 (s, 9H), 1.32 (d, J = 7.0, 3H), 1.27 (m, 1H), 0.95 (m, 6H), 0.92 (d, J = 6.4, 3H), 0.88 (dd, J = 6.5, 1.7, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.6*, 171.8*, 170.1*, 164.3*, 155.8, 148.4*, 122.3*, 93.6*, 80.0, 58.6, 58.4*, 52.9, 45.9*, 41.2*, 39.0*, 28.3, 24.7, 24.2*m 23.7*, 23.0, 22.7, 22.0, 20.0*. *Showed twinning. ν_{max} 3309.0, 3070.5, 2961.3, 2874.9, 1721.5, 1668.3, 1627.6, 1531.8, 1456.4, 1358.8, 1330.2, 1245.3, 1172.8, 1044.7, 982.8 cm^{-1}. λ_{max} (log ε) 233 (6.7) nm. $[\alpha]_D^{25}$ -23.3 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z* 502.2895, 1.4 ppm (calculated for $C_{25}H_{41}N_3NaO_6^+$, 502.2888).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117d**). As a light-yellow foam. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 (t, *J* = 16.3, 1H), 7.29 (m, 2H), 7.21 (m, 3H), 7.03 (m, 1H), 6.33 (m,1H), 5.03 (s, 1H), 4.98 (m, 1H), 4.85 (m, 1H), 4.60 (p, *J* = 6.5, 1H), 4.08 (m, 1H), 3.85 (s, 3H), 2.94 (m, 2H), 1.62 (m, 2H), 1.47 (m, 4H), 1.44 (s, 9H), 0.91 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 172.0*, 169.6*, 163.9*, 155.7, 146.6*, 136.5*, 129.4, 128.5, 126.8, 123.0*, 93.0*, 79.9, 58.7, 55.7*, 53.0, 51.1, 40.5, 28.3, 24.7, 22.9, 22.0, 17.1*. *Showed twinning. ν_{max} 3306.0, 2959.6, 2873.7, 1718.8, 1666.9, 1628.4, 1524.4, 1453.9, 1333.0, 1288.3, 1246.3, 1174.3, 700.9 cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at end of range. $[\alpha]_D^{25}$ -20.1 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z*

536.2735, 0.7 ppm (calculated for $C_{28}H_{39}N_3NaO_6^+$, 536.2735).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117e**). As a light-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.32 (m, 3H), 7.20 (m, 6H), 7.09 (m, 1H), 6.92 (m, 2H), 6.27 (m, 1H), 5.02 (m, 1H), 4.89 (m, 1H), 4.80 (m, 2H), 4.08 (m, 1H), 3.82 (d, *J* = 1.9, 3H), 3.54 (m, 1H), 3.11 (m, 1H), 2.96 (m, 2H), 1.62 (m, 2H), 1.43 (m, 10H), 0.90 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.7*, 171.9*, 169.5*, 164.2*, 155.6, 146.8*, 136.4*, 134.1*, 129.5*, 129.4*, 128.5*, 128.1*, 126.9*, 126.8*, 122.8*, 94.9*, 79.9, 59.5*, 58.3*, 53.0, 51.0*, 41.0*, 40.4*, 34.3*, 28.2, 24.6, 22.8, 22.0. *Showed twinning. ν_{max} 3307.6, 3062.1, 3030.2, 2957.2, 2871.1, 1718.6, 1666.6, 1629.8, 1523.2, 1450.7, 1354.0, 1247.3, 1170.4, 1117.9, 1024.7, 968.8, 736.0, 700.4 cm^{-1}. λ_{max} (log ε) 205 (7.2) nm, maximum at end of range. $[\alpha]_D^{25}$ -10.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 612.3047, 0.5 ppm (calculated for $C_{34}H_{43}N_3NaO_6^+$, 612.3044).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117f**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (m, 1H), 7.27 (m, 2H), 7.23 (m, 1H), 7.20 (m, 2H), 7.00 (ddd, *J* = 15.6, 12.1, 5.2, 1H), 6.20 (m, 1H), 5.03 (s, 1H), 4.98 (m, 1H), 4.79 (m, 1H), 4.67 (dt, *J* = 6.8, 3.5, 1H), 3.85 (s, 3H), 2.94 (m, 2H), 1.82 (m, 3H), 1.72 (m, 1H), 1.63 (m, 2H), 1.44 (d, *J* = 1.3, 9H), 0.91 (m, 9H), 0.87 (t, *J* = 6.8, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 180.5*, 171.9*, 170.0*, 164.0*, 155.6, 146.4*, 136.5*, 129.4*, 128.5*, 126.8, 123.2*, 98.6*, 80.0*, 58.6, 58.5*, 53.1, 51.1*, 40.6*, 39.0(0), 38.9(5), 28.3, 24.7*, 24.2*, 23.7(1), 23.6(7), 22.9, 22.7. *Showed twinning. ν_{max} 3305.4, 3063.9, 2958.6, 2873.5, 1718.7, 1668.4, 1627.1, 1525.6, 1453.6, 1332.8, 1247.3, 1172.6, 1042.6, 980.3, 811.5, 741.6, 699.2 cm^{-1}. λ_{max} (log ε) 205 (7.2) nm, maximum at end of range. $[\alpha]_D^{28}$ -19.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 578.3203, 0.3 ppm (calculated for $C_{31}H_{45}N_3NaO_6^+$, 578.3201).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117g**). As a light-brown foam. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 2H), 7.20 (m, 7H), 7.08 (ddd, *J* = 15.6, 7.8, 5.2, 1H), 6.91 (m, 2H), 5.02 (m, 1H), 4.89 (dt, *J* = 5.6, 3.0, 1H), 4.76 (m, 2H), 4.06 (m, 1H), 3.82 (d, *J* = 1.7, 3H), 3.54 (td, *J* = 13.8, 5.2, 1H), 3.10 (dt, *J* = 13.9, 3.5, 1H), 2.95 (m, 2H), 1.62 (m, 2H), 1.43 (d, *J* = 2.9, 9H), 1.40 (m, 1H), 0.90 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.7*, 171.8, 169.4, 164.2*, 155.6, 146.8*, 146.8*, 136.4*, 134.1*, 129.5*, 129.4*, 128.5*, 128.1*, 126.9*, 126.8*, 122.9*, 94.9*, 79.9, 59.6*, 58.3, 53.0, 51.1*, 41.0, 40.5*, 34.4*, 28.2, 24.6, 22.8, 21.9. *Showed twinning. ν_{max} 3307.5, 2955.0, 1719.9, 1667.4, 1629.5, 1522.8, 1451.1, 1343.2, 1246.9, 1170.6, 968.5, 700.6 cm^{-1}. λ_{max} (log ε) 205 (7.1) nm, maximum at end of range. $[\alpha]_D^{25}$ -9.6 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z* 612.3046, 0.3 ppm (calculated for $C_{34}H_{43}N_3NaO_6^+$, 612.3044).

tert-butyl ((*S*)-1-(((*R,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117h**). As a bright yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, 3H), 7.20 (m, 6H), 7.08 (m, 1H), 6.93 (m, 2H), 6.28 (m, 1H), 5.01 (m, 1H), 4.88 (m, 1H), 4.79 (s, 1H), 3.82 (s, 3H), 3.54 (m, 1H), 3.11 (m, 2H), 2.94 (m, 1H), 1.62 (m, 1H), 1.52 (m, 1H), 1.42 (m, 10H), 0.86 (t, J = 5.6, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.9*, 172.0, 169.6, 164.5*, 156.1, 147.2*, 136.5*, 134.2*, 129.6, 129.4*, 128.5, 128.2*, 127.0*, 126.9*, 122.8*, 94.9*, 80.2, 59.7*, 58.4, 53.0, 51.0*, 41.6, 40.4*, 34.6, 28.3, 24.7, 22.9, 21.8. *Showed twinning. ν_{max} 3305.3, 3062.5, 3031.1, 2957.7, 1718.7, 1666.4, 1632.6, 1529.5, 1449.8, 1359.6, 1246.7, 1170.1, 1118.7, 1026.4, 969.9, 737.6, 699.9 cm^{-1}. λ_{max} (log ε) 205 (7.2) nm, maximum at end of range. $[\alpha]_D^{25}$ -6.7 (CH_2Cl_2).

tert-butyl ((*S*)-1-(((*R,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117i**). As a bright yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.29 (m, 3H), 7.19 (m, 6H), 7.09 (m,

1H), 6.91 (m, 2H), 6.28 (m, 1H), 5.01 (m, 1H), 4.88 (dt, *J* = 5.5, 2.9, 1H), 4.79 (s, 1H), 3.82 (s, 3H), 3.54 (ddd, *J* = 15.7, 14.0, 5.3, 1H), 3.11 (m, 2H), 2.94 (m, 1H), 1.60 (m, 1H), 1.53 (m, 1H), 1.42 (m, 10H), 0.86 (t, *J* = 5.6, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 177.9*, 172.0, 169.6, 164.2, 155.9, 147.2*, 136.4*, 134.2*, 129.6, 129.4*, 128.6*, 128.2*, 127.0*, 126.9*, 122.8*, 94.9*, 80.1, 59.7*, 58.4, 53.0, 51.0*, 41.1, 40.3*, 34.6, 28.3*, 24.7, 22.9, 21.8. *Showed twinning. ν_{max} 3310.6, 3061.9, 3030.9, 2957.7, 2872.3, 1718.8, 1668.2, 1631.5, 1509.1, 1450.7, 1356.6, 1247.9, 1170.2, 1118.4, 1025.4, 969.8, 807.4, 736.2, 700.8 cm^{-1}. λ_{max} (log ε) 205 (7.0) nm, maximum at end of range. $[\alpha]_D^{25}$ -7.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 612.3045, 0.3 ppm (calculated for $C_{34}H_{43}N_3NaO_6$, 612.3044).

tert-butyl ((*S*)-1-(((*S,E*)-5-((*S*)-3-(hex-5-yn-1-yloxy)-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**117j**). As a colorless foam. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 (ddd, *J* = 15.6, 9.7, 1.9, 1H), 7.00 (m, 1H), 6.24 (bs, 1H), 5.01 (s, 1H), 4.92 (m, 1H), 4.77 (m, 1H), 4.62 (m, 1H), 4.11 (m, 1H), 4.03 (m, 2H), 2.28 (td, *J* = 6.9, 2.6, 2H), 1.99 (t, *J* = 2.5, 1H), 1.94 (m, 2H), 1.68 (m, 4H), 1.49 (m, 4H), 1.44 (s, 9H), 1.32 (d, *J* = 7.0, 3H), 0.95 (m, 6H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 179.7, 171.8*, 169.8*, 164.2*, 155.8, 148.4*, 122.3*, 93.1, 83.4, 80.0, 71.4, 69.1, 55.8*, 52.9, 46.0*, 41.2*, 28.3, 27.4, 24.7*, 24.6, 23.0, 22.0, 20.0*, 18.0 17.2*. *Showed twinning. ν_{max} 3303.1, 3095.8, 2961.3, 2877.4, 1720.4, 1668.0, 1625.2, 1531.5, 1457.4, 1356.3, 1329.3, 1287.5, 1241.3, 1174.3, 1037.9, 977.9, 940.4, 861.6, 810.0, 703.1, 637.1, 506.5 cm^{-1}. λ_{max} (log ε) 245 (7.0) nm. $[\alpha]_D^{25}$ -30.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 526.2894, 1.1 ppm (calculated for $C_{27}H_{41}N_3NaO_6^+$, 526.2888).

Benzyl Esterification The leucic acid was dissolved in the sulfoxide, and while stirring the carbonate was added in one portion, then the bromide was added dropwise. After stirring for 6 h at rt the reaction mixture was quenched with deionized water and extracted iteratively with diethyl ether. The combined organic phase were washed with brine and then dried over $MgSO_4$. The filtrate was then concentrated to an oil and purified by iterative flash

chromatography with gradient elution of EtOAc/hexanes to afford a colorless oil.

benzyl (*S*)-2-hydroxy-4-methylpentanoate (**114**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (m, 5H), 5.21 (d, *J* = 2.0, 2H), 4.25 (m, 1H), 2.62 (d, *J* = 6.1, 1H), 1.89 (m, 1H), 1.57 (m, 2H), 0.94 (t, *J* = 6.3, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 175.7, 135.2, 128.7, 128.5, 128.3, 69.1, 67.3, 43.4, 24.4, 23.2, 21.5. ν_{max} 3457.6, 2956.2, 2874.1, 1736.1, 1459.9, 1375.4, 1266.2, 1206.1, 1139.9, 1085.4, 1004.0, 744.4, 697.4, 590.7, 417.6 cm^{-1}. λ_{max} (log ε) 205 (6.8) nm, maximum at end of range. $[\alpha]_D^{25}$ -18.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 245.1149, 0.4 ppm (calculated for $C_{13}H_{18}NaO_3^+$, 245.1148).

Tail Dimer Synthesis. As a general procedure, the benzyl ester protected L-leucic acid was dissolved in DCM, then the pyridine and Fmoc-protected amino acid were added to the solution in one portion each. The reaction was then cooled to 0 °C by ice bath and the DIC was added dropwise. The reaction mixture was then allowed to warm to rt overnight, after which the reaction was worked up by diluting into DCM, washing with 10% $NaHCO_3$ (*aq*), 5% citric acid (*aq*), and brine before bring dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap and the DIC-derived urea salt was removed by filtration of this material through a glass frit with iterative EtOAc washes. The filtrate of this fractionation was then concentrated to dryness again by rotovap and taken up in DMF. To this solution was added the DIPA and the reaction was stirred for 2.5-2.75 h. Then the reaction mixture was diluted into diethyl ether, washed with 5% citric acid (*aq*), and brine (if by TLC the citric acid showed product it was back extracted with ether). Then the organic phase was dried over $MgSO_4$, and the filtrate of this step was then concentrated to dryness by rotovap. The crude residue was purified by silica gel flash chromatography with a gradient diethyl ether/hexanes elution.

benzyl (*S*)-2-((*L*-isoleucyl)oxy)-4-methylpentanoate (**115a**). As a faint-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (m, 5H), 5.17 (s, 2H), 5.11 (dd, *J* = 9.6, 4.0, 1H), 3.40 (d, *J* = 4.9, 1H), 1.81 (m, 2H), 1.74 (m, 1H), 1.66 (m, 1H), 1.51 (m, 3H), 1.21 (m, 1H),

0.92 (m, 12H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 175.2, 170.4, 135.3, 128.6, 128.4, 128.3, 71.2, 67.0, 59.1, 39.8, 39.0, 24.6, 24.4, 23.0, 21.6, 15.5, 11.6. ν_{max} 3383.7, 3035.2, 2962.1, 2876.9, 1744.5, 1459.6, 1380.2, 1273.4, 1171.2, 1073.9, 1011.8, 746.1, 698.9 cm^{-1}. λ_{max} (log ε) 205 (6.1) nm, maximum at end of range. $[\alpha]_D^{26}$ -11.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 358.1982, (-2.0) ppm (calculated for $C_{19}H_{29}NNaO_4^+$, 358.1989).

benzyl (*S*)-2-((*L*-leucyl)oxy)-4-methylpentanoate (**115b**). As a faint-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 (m, 3H*), 5.14 (m, 3H), 2.50 (dd, *J* = 9.0, 5.4, 1H), 1.81 (m, 2H), 1.74 (m, 1H), 1.68 (m, 1H), 1.60 (m, 1H), 1.41 (m, 3H), 0.94 (d, *J* = 6.4, 3H), 0.91 (m, 9H). *Value of integral was low, likely due to the acquisition time being too short to fully allow the benzyl group to relax. ν_{max} 3373.5, 3036.5, 2957.7, 2874.1, 1747.2, 1661.4, 1505.2, 1459.2, 1377.8, 1271.0, 1182.3, 1135.7, 1074.2, 1012.1, 745.7, 699.0 cm^{-1}. HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 358.1992, (0.8) ppm (calculated for $C_{19}H_{29}NNaO_4^+$, 358.1989).

benzyl (*S*)-2-((*L*-valyl)oxy)-4-methylpentanoate (**115c**). As a faint-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (m, 5H), 5.16 (s, 2H), 5.12 (dd, *J* = 9.7, 4.0, 1H), 3.35 (d, *J* = 4.7, 1H), 2.11 (m, 1H), 1.81 (m, 1H), 1.73 (m, 1H), 1.66 (m, 1H), 1.41 (dd, *J* = 6.7, 3.0, 1H), 0.99 (d, *J* = 6.9, 3H), 0.94 (d, *J* = 6.6, 3H), 0.91 (d, *J* = 6.7, 3H), 0.90 (d, *J* = 6.9, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 175.2, 170.4, 135.2, 128.6, 128.4, 128.3, 71.2, 67.0, 59.8, 39.7, 31.9, 24.6, 23.0, 21.5, 19.4, 16.7. ν_{max} 3387.8, 2961.5, 2877.6, 1744.6, 1461.7, 1381.4, 1275.4, 1168.5, 1073.3, 1009.9, 747.3, 698.7, cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at end of range. $[\alpha]_D^{26}$ -15.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 322.2014, 0.3 ppm (calculated for $C_{18}H_{28}NO_4$, 322.2013).

benzyl (*S*)-2-((*L*-phenylalanyl)oxy)-4-methylpentanoate (**115d**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 (m, 5H), 7.30 (m, 2H), 7.23 (m 3H), 5.16 (m, 3H), 3.79 (dd, *J* = 8.6, 4.7, 1H), 3.20 (dd, *J* = 13.8, 4.7, 1H), 2.76 (dd, *J* = 13.8, 8.6, 1H), 1.82 (m, 1H), 1.68 (m, 2H), 0.93 (d, *J* = 6.4, 3H), 0.91 (d, *J* = 6.3, 3H). ^{13}C NMR (500 MHz, $CDCl_3$)

δ 174.8, 170.5, 137.4, 135.3, 129.5, 128.8, 128.7, 128.6, 128.5, 126.9, 71.5, 67.3, 55.7, 40.9, 39.8, 24.7, 23.2, 21.7. ν_{max} 3031.8, 2957.7, 1745.5, 1457.1, 1379.9, 1273.5, 1177.9, 1074.7, 1009.4, 746.9, 699.5 cm^{-1}. λ_{max} (log ε) 205 (7.1) nm, maximum at end of range. $[\alpha]_D^{24}$ -22.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 370.2014, 0.3 ppm (calculated for $C_{22}H_{28}NO_4^+$, 370.2013).

Tail Dimer Methylation. As a general procedure, the free amine was dissolved in DMF and to this solution DIPEA and then MeI were added sequentially dropwise. After 6-6.25 h the reaction mixture was diluted into hexanes and washed with 5% citric acid (*aq*); the aqueous layer was back extracted three times with hexanes and the combined organics were washed with brine and dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap, with caution, as the product was observed to be mildly volatile. The crude residue was then purified by silica gel flash chromatography with a gradient diethyl ether/hexanes elution.

benzyl (*S*)-2-((dimethyl-*L*-isoleucyl)oxy)-4-methylpentanoate (**116a**). As a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 (m, 5H), 5.17 (s, 2H), 5.09 (dd, *J* = 10.3, 3.8, 1H), 2.93 (d, *J* = 10.5, 1H), 2.30 (s, 6H), 1.83 (m, 2H), 1.77 (m, 1H), 1.67 (m, 1H), 1.61 (m, 1H), 1.14 (m, 1H), 0.95 (d, *J* = 6.6, 3H), 0.91 (d, *J* = 6.5, 3H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.4, 170.8, 135.4, 128.7, 128.5(4), 128.5(1), 72.4, 70.8, 67.1, 41.5, 39.9, 33.4, 25.1, 24.7, 23.3, 21.3, 15.7, 10.5. ν_{max} 2959.1, 2874.8, 2788.3, 1739.4, 1460.6, 1374.1, 1271.9, 1150.4, 1070.6, 746.0, 698.9 cm^{-1}. λ_{max} (log ε) 205 (6.7) nm, maximum at end of range. $[\alpha]_D^{25}$ -44.1 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 350.2328, 0.6 ppm (calculated for $C_{20}H_{32}NO_4^+$, 350.2326).

benzyl (*S*)-2-((dimethyl-*L*-leucyl)oxy)-4-methylpentanoate (**116b**). As a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ Showed suspected product with unknown contaminant. HR-ESI-TOF-MS $[M + H]^+$ *m/z* 364.2485, 0.8 ppm (calculated for $C_{21}H_{34}NO_4$, 364.2482).

benzyl (*S*)-2-((dimethyl-*L*-valyl)oxy)-4-methylpentanoate (**116c**). As a colorless oil.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 (m, 5H), 5.17 (s, 2H), 5.09, (dd, *J* = 10.3, 3.8, 1H), 2.82 (d, *J* = 10.6, 1H), 2.31 (s, 6H), 2.01 (m, 1H), 1.83, (m, 1H), 176 (m, 1H), 1.61 (ddd, *J* = 13.2, 9.2, 3.8, 1H), 0.98 (d, *J* = 6.6, 3H), 0.95 (d, *J* = 6.7, 3H), 0.91 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.1, 170.6, 135.3, 128.6, 128.3(9), 128.3(6), 74.0, 70.6, 67.0, 41.2, 39.8, 27.6, 24.5, 23.1, 21.1, 19.5, 19.3. ν_{max} 2961.4, 2874.8, 2787.3, 1737.8, 1459.5, 1376.1, 1269.8, 1135.4, 1067.1, 745.2, 698.2 cm^{-1}. λ_{max} (log ε) 205 (6.9) nm, maximum at end of range. $[\alpha]_D^{25}$ -37.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 364.2485, 0.8 ppm (calculated for $C_{21}H_{34}NO_4^+$, 364.2482).

benzyl (*S*)-2-((dimethyl-*L*-phenylalanyl)oxy)-4-methylpentanoate (**116d**). As a light yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33 (m, 5H), 7.24 (m, 1H), 7.20 (m, 3H), 5.14 (m, 2H), 5.04 (dd, *J* = 9.9, 4.0, 1H), 3.58 (m, 1H), 3.05 (m, 1H), 2.96 (m, 1H), 2.40 (s, 6H), 1.70 (m, 1H), 1.61 (m, 2H), 0.92 (d, *J* = 6.5, 3H), 0.88 (d, *J* = 6.5, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 171.2, 170.5, 138.3, 135.5, 129.1, 128.7, 128.5(3), 128.4(7), 128.4, 126.5, 71.2, 68.9, 67.1, 41.7, 39.9, 35.9, 24.6, 23.2, 21.4. ν_{max} 3031.6, 2957.3, 2873.2, 2788.8, 1741.6, 1498.3, 1455.4, 1371.9, 1271.3, 1154.7, 1073.9, 1019.5, 744.5, 698.7 cm^{-1}. λ_{max} (log ε) 205 (7.5) nm, maximum at end of range. $[\alpha]_D^{25}$ -13.6 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 398.2329, 0.8 ppm (calculated for $C_{24}H_{32}NO_4^+$, 398.2326).

Final Coupling Reaction. As a general procedure, a vial was charged with the benzyl ester protected dimer then this material was dissolved in EtOH, the palladium (as a 1% loading on carbon) was added in one portion, and the atmosphere was replaced with hydrogen. The reaction was then stirred overnight and the reaction mixture was passed through a ca. 1 cm diatomaceous earth plug, prepared in a pipette with glass wool, into a 25 mL pear flask using three 1 mL DCM rinses. This material was then dried under N_2 (*g*) flow.

In a separate vial the carbamate of the leucine extended product was dissolved in DCM and cooled to 0 °C by ice bath and the TFA was added dropwise. After 0.5-1 h at 0 °C the reaction mixture was dried under N_2 (*g*) flow, then azeotroped three times, with three

drops of toluene.

The free acid dimer was dissolved DCM and to this solution was added the DIPEA, dropwise, and the HOBt, in one portion. The reaction mixture was cooled to 0 °C by ice bath and the EDC was added in one portion. After 30-45 min the deprotected lecuine extended product was taken up in DCM and added to the reaction mixture with two DCM rinses. The reaction mixture was then allowed to warm to rt overnight. To work up the reaction, the mixture was diluted into dichloromethane and washed with 5% citric acid (*aq*); the aqueous layer was then back extracted three times with hexanes and the combined organics were washed with brine and dried over $MgSO_4$. The filtrate of this step was then concentrated to dryness by rotovap and purified by a RP-SPE using elutions of 25%/60%/80% CH_3CN/H_2O and CH_3CN. The fractions containing product as detected by LCMS, usually the 60% and 80% fraction were combined and further purified by RP-HPLC using an CH_3CN/H_2O elution.

gallinamide A (**106**). Full 4 ItLLALAcM (Gallinamide A) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 (ddd, *J* = 15.4, 9.2, 1.8, 1H), 6.99 (ddd, *J* = 15.6, 7.9, 4.9, 1H), 6.53 (dd, *J* = 8.5, 5.3, 1H), 6.45 (dd, *J* = 14.8, 8.1, 1H), 5.17 (m, 1H), 5.05 (d, *J* = 3.1, 1H), 4.72 (m, 1H), 4.62 (m, 1H), 4.47 (tt, *J* = 8.7, 5.2 1H), 3.87 (s, 3H), 2.94 (d, *J* = 10.2, 1H), 2.31 (s, 6H), 1.85 (m, 1H), 1.75 (m, 2H), 1.64 (m, 4H), 1.49 (dd, *J* = 6.5, 2.5, 3H), 1.29 (dd, *J* = 6.9, 3H), 1.17 (m, 1H), 0.94 (m, 12H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.7*, 171.0*, 170.6*, 170.4*, 169.7*, 164.1*, 148.3*, 122.2*, 93.0*, 72.4(0), 72.3(8), 58.7*, 55.7*, 51.4*, 46.1*, 41.6, 41.1*, 40.8, 33.4, 25.0, 24.7, 24.4, 23.2, 23.0, 22.0*, 21.4, 19.9*, 17.1*, 15.8, 10.5. *Showed twinning. ν_{max} 3291.8, 3082.8, 2960.1, 2874.8, 2787.1, 1727.1, 1631.7, 1546.3, 1457.6, 1332.5, 1289.0, 1247.5, 1176.7, 1051.3, 977.8, 809.1, 700.9 cm^{-1}. $[\alpha]_D^{26}$ -64.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 593.3909, 0.0 ppm (calculated for $C_{31}H_{53}N_4O_7^+$, 593.3909).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-

5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118a**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.34 (dd, *J* = 15.5, 10.2, 1H), 7.20 (m, 3H), 7.06 (dd, *J* = 15.5, 4.9, 1H), 6.94 (m, 2H), 6.51 (m, 1H), 6.37 (m, 1H), 5.17 (m, 1H), 4.89 (m, 1H), 4.83 (d, *J* = 5.6, 1H), 4.77 (m, 1H), 4.56 (m, 1H), 3.82 (s, 3H), 3.54 (m, 1H), 3.12 (d, *J* = 13.8, 1H), 2.94 (d, *J* = 10.2, 1H), 2.32 (s, 6H), 1.84 (m, 1H), 1.78 (m, 3H), 1.65 (m, 3H), 1.56 (m, 1H), 1.31 (dd, *J* = 7.0, 3.2, 3H), 1.15 (m, 1H), 0.93 (m, 12H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 178.0*, 171.1*, 170.7*, 170.4*, 169.8*, 164.6*, 148.8*, 134.3*, 129.7, 128.3*, 127.1*, 122.3*, 95.0, 72.6, 72.5, 59.8*, 58.5, 51.6, 46.1*, 41.7, 41.2*, 41.0, 34.7*, 33.5, 25.2, 24.8, 24.5, 23.3, 23.1*, 22.2*, 21.5, 20.0*, 15.9, 10.6. *Showed twinning. ν_{max} 3293.5, 3075.9, 2959.4, 2873.8, 2787.4, 1727.6, 1634.6, 1545.4, 1455.5, 1347.1, 1247.7, 1166.4, 1052.2, 972.3, 731.9 cm^{-1}. $[\alpha]_D^{25}$ -59.0 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 669.4222, 0.1 ppm (calculated for $C_{37}H_{57}N_4O_7^+$, 669.4223).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-phenylalaninate (**118b**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 1H 7.35 (ddd, *J* = 15.5, 8.4, 1.7, 1H), 7.29 (m, 2H), 7.21 (m, 6H), 7.05 (dd, *J* = 15.5, 4.9, 2.5, 1H), 6.94 (m, 2H), 6.30 (dd, *J* = 20.1, 8.2, 1H), 6.21 (dd, *J* = 15.4, 8.1, 1H), 5.09 (dt, *J* = 8.3, 4.2, 1H), 4.89 (m, 1H), 4.81 (d, *J* = 3.1, 1H), 4.76 (m, 1H), 4.36 (m, 1H), 3.81 (s, 3H), 3.56 (m, 2H), 3.12 (m, 1H), 3.01 (m, 2H), 2.42 (d, *J* = 1.7, 6H), 1.67 (m, 4H), 1.57 (m, 1H), 1.46 (m, 1H), 1.33 (dd, *J* = 7.0, 3.1, 3H), 0.91 (m, 12H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.8, 170.9*, 170.5*, 170.4*, 169.7, 164.5*, 148.7*, 137.7*, 134.3*, 129.6, 128.9, 128.6, 128.2*, 127.0*, 126.8*, 122.2*, 94.9*, 72.9*, 69.2*, 59.7*, 58.4*, 51.4, 46.1*, 41.9, 40.8, 40.4*, 35.7, 34.6*, 24.8, 24.4, 23.1, 23.0*, 21.9*, 21.5, 20.0*. *Showed twinning. ν_{max} 3291.5, 3067.3, 2956.5, 2872.6, 2788.8, 1727.2, 1635.2, 1547.5, 1454.5, 1348.7, 1245.2, 1165.2, 1068.7, 1023.1, 972.1, 807.4, 735.3, 701.0 cm^{-1}. $[\alpha]_D^{25}$ -34.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z*

703.4063, -0.3 ppm (calculated for $C_{40}H_{55}N_4O_7^+$, 703.4065).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-L-isoleucinate (**118c**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 (ddd, *J* = 15.5, 9.6, 1.8, 1H), 6.98 (m, 1H), 6.53 (m, 1H), 6.36 (dd, *J* = 23.7, 8.0, 1H), 5.18 (m, 1H), 5.06 (d, *J* = 5.4, 1H), 4.73 (m, 1H), 4.68 (m, 1H), 4.46 (m, 1H), 3.86 (s, 3H), 2.94 (d, *J* = 10.1, 1H), 2.32 (s, 6H), 1.83 (m, 3H), 1.75 (m, 4H), 1.65 (m, 3H), 1.57 (m, 1H), 1.29 (d, *J* = 7.0, 3H), 1.18 (m, 1H), 0.93 (m, 18H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.6*, 171.0, 170.6*, 170.4*, 170.1*, 164.3*, 148.2*, 122.4*, 93.6, 72.5, 72.4, 58.6*, 51.5, 46.1*, 41.6, 41.1*, 40.9, 39.1*, 33.4, 25.1, 24.8, 24.4, 24.2, 23.7*, 23.2, 23.0, 22.7, 22.1*, 21.4, 19.9, 15.8, 10.5. *Showed twinning. ν_{max} 3292.0, 3086.1, 2959.5, 2874.3, 2788.1, 1727.5, 1630.3, 1546.5, 1456.9, 1336.3, 1223.6, 1172.7, 1051.9, 983.6, 809.3, 685.0 cm^{-1}. $[\alpha]_D^{25}$ -58.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 635.4376, -0.3 ppm (calculated for $C_{34}H_{59}N_4O_7^+$, 635.4378).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-valinate (**118d**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 (ddd, *J* = 15.6, 8.5, 1.7, 1H), 6.97 (ddd, *J* = 15.5, 7.4, 4.9, 1H), 6.39 (t, *J* = 7.3, 1H), 6.15 (m, 1H), 5.16 (dt, *J* = 9.4, 4.4, 1H), 5.03 (s, 1H), 4.72 (m, 1H), 4.61 (qd, *J* = 6.5, 2.7, 1H), 4.43 (tt, *J* = 8.9, 5.2, 1H), 3.86 (s, 3H), 2.83 (d, *J* = 10.4, 1H), 2.33 (s, 6H), 2.02 (m, 1H), 1.76 (m, 3H), 1.57 (m, 3H), 1.49 (dd, *J* = 6.6, 2.1, 1H), 1.30 (d, *J* = 6.2, 3H), 0.99 (d, *J* = 6.7, 3H), 0.95 (m, 12H), 0.90 (d, *J* = 6.5, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.7, 171.1*, 170.6*, 170.4*, 169.6, 164.2*, 148.1*, 122.3*, 93.0, 74.3, 72.5*, 58.7, 55.7*, 51.5, 46.1*, 41.5, 41.0, 40.9*, 27.7, 24.8, 24.5, 23.2, 23.0, 22.0*, 21.4, 19.9*, 19.7, 19.4, 17.1*. *Showed twinning. ν_{max} 3292.5, 3081.2, 2958.4, 2874.9, 2789.3, 1727.5, 1658.3, 1631.8, 1546.1, 1457.7, 1332.3, 1289.6, 1242.6, 1179.2, 1053.5, 977.7, 809.2, 700.7 cm^{-1}. $[\alpha]_D^{26}$ -33.5 (CH_2Cl_2). HR-ESI-TOF-MS

$[M + H]^+$ *m/z* 579.3751, -0.2 ppm (calculated for $C_{30}H_{51}N_4O_7^+$, 579.3752).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-3-(hex-5-yn-1-yloxy)-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118e**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 (ddd, *J* = 15.6, 8.9, 1.8, 1H), 6.99 (ddd, *J* = 15.6, 8.4, 5.0, 1H), 6.53 (m, 1H), 6.44 (dd, *J* = 14.3, 8.2, 1H), 5.17 (m, 1H), 5.02 (d, *J* = 3.8, 1H), 4.72 (m, 1H), 4.61 (m, 1H), 4.47 (tt, *J* = 8.7, 5.5, 1H), 4.06 (m, 1H), 4.01 (m, 1H), 2.94 (d, *J* = 10.2, 1H), 2.31 (s, 6H), 2.28 (td, *J* = 6.9, 2.6, 2H), 2.00 (t, *J* = 2.6, 1H), 1.93 (m, 2H), 1.85 (m, 1H), 1.76 (m, 3H), 1.66 (m, 5H), 1.59 (m, 1H), 1.49 (dd, *J* = 6.7, 2.8, 3H), 1.29 (d, *J* = 6.9, 3H), 1.17 (m, 1H), 0.94 (m, 12H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 179.7*, 171.0, 170.6*, 170.4*, 169.8*, 164.1*, 148.2*, 122.3*, 93.2, 83.4, 72.5, 72.3, 71.4, 69.1, 55.8*, 51.5, 46.1*, 41.6, 41.1*, 40.9, 33.4, 27.4, 25.1, 24.8, 24.6, 24.4, 23.2, 23.0, 22.1*, 21.5, 19.9*, 18.0, 17.2*, 15.8, 10.5. *Showed twinning. ν_{max} 3295.3, 3085.7, 2958.0, 2874.8, 2787.7, 1725.8, 1659.0, 1625.3, 1545.5, 1459.4, 1330.6, 1287.7, 1234.9, 1175.7, 1046.0, 977.7, 703.1, 636.9 cm^{-1}. $[\alpha]_D^{26}$ -50.7 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 659.4376, -0.3 ppm (calculated for $C_{36}H_{59}N_4O_7^+$, 659.4378).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118f**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 (ddd, *J* = 15.6, 12.2, 1.8, 1H), 7.27 (m, 2H), 7.21 (m, 1H), 7.18 (m, 2H), 7.02 (m, 1H), 6.47 (dd, *J* = 8.2, 3.7, 1H), 6.37 (dd, *J* = 20.7, 8.4, 1H), 5.16 (m, 1H), 5.02 (d, *J* = 4.0, 1H), 4.95 (m, 1H), 4.59 (p, *J* = 6.4, 1H), 4.40 (m, 1H), 3.85 (s, 3H), 2.92 (m, 3H), 2.30 (s, 6H), 1.84 (m, 1H), 1.74 (m, 3H), 1.64 (m, 3H), 1.53 (m, 1H), 1.47 (dd, *J* = 8.4, 6.5, 3H), 1.15 (m, 1H), 0.93 (m, 15H), 0.85 (d, *J* = 6.6, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.6*, 170.9*, 170.8, 170.3*, 169.6*, 163.9*, 146.5*, 136.5*, 129.3*, 128.6, 126.9, 123.0*, 93.0, 72.6, 72.3, 58.7*, 55.7*, 51.6*, 51.4*, 41.6, 41.1, 40.9*,

40.5*, 33.5, 252., 24.7*, 24.4, 23.2, 23.0, 22.1*, 21.5, 17.1*, 15.7, 10.5. *Showed twinning. $[\alpha]_D^{28}$ -35.2 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 669.4221, -0.3 ppm (calculated for $C_{37}H_{54}N_4O_7^+$, 669.4222).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-phenylalaninate (**118g**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 (ddd, *J* = 15.5, 10.4, 1.7, 1H), 7.24 (m, 6H) 7.17 (m, 4H), 7.02 (td, *J* = 15.7, 15.6, 5.2, 1H), 6.50 (m, 1H), 6.19 (dd, *J* = 17.6, 8.0, 1H), 5.10 (m, 1H), 5.02 (d, *J* = 2.9, 1H), 4.94 (m, 1H), 4.59 (m, 1H), 4.31 (m, 1H), 3.85 (s, 3H), 3.49 (m, 1H), 3.01 (m, 2H), 2.96 (m, 1H), 2.89 (m, 1H), 2.39 (s, 6H), 1.65(m, 4H), 1.53 (m, 1H), 1.47 (t, *J* = 6.6, 3H), 1.41 (m, 1H), 0.89 (m, 12H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.6*, 170.7(2)*, 170.6(8), 170.4*, 169.5*, 164.0*, 146.7*, 137.6*, 136.6*, 129.3*, 128.9*, 128.6*, 128.5, 126.9, 126.8, 123.0*, 93.0, 72.6*, 69.5*, 58.7*, 55.7, 515.*, 51.4, 41.9*, 40.7, 40.4*, 40.0*, 35.9*, 24.7*, 24.4, 23.1, 23.0, 21.9*, 21.6*, 17.1*. *Showed twinning. ν_{max} 3295.1, 3063.6, 3032.2, 2955.0, 2873.1, 2787.3, 1725.4, 1661.5, 1628.4, 1543.7, 1455.0, 1334.0, 1289.3, 1245.2, 1174.6, 1063.1, 971.4, 739.0, 699.8 cm^{-1}. $[\alpha]_D^{25}$ -44.3 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 725.3886, 0.1 ppm (calculated for $C_{40}H_{54}N_4NaO_7^+$, 725.3885).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118h**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 3H), 7.23 (m, 1H), 7.19 (m, 5H), 7.08 (ddd, *J* = 15.6, 7.8, 5.1, 1H), 6.90 (m, 2H), 6.46 (m, 1H), 6.27 (m, 1H), 5.16 (m, 1H), 4.98 (m, 1H), 4.88 (td, *J* = 5.0, 2.9, 1H), 4.80 (d, J = 2.8, 1H), 4.38 (tdd, *J* = 8.8, 5.7, 3.6, 1H), 3.81 (d, *J* = 1.2, 3H), 3.53 (ddd, *J* = 15.7, 14.0, 5.1, 1H), 3.10 (ddd, *J* = 13.9, 6.9, 2.9, 1H), 2.93 (m, 3H), 2.30 (d, *J* = 1.9, 6H), 1.84 (m, 1H), 1.76 (m, 2H), 1.65 (m, 4H), 1.51 (m, 1H), 1.15 (m, 1H), 0.95 (m, 6H), 0.90 (m, 9H), 0.86 (m, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.8, 170.9, 170.7, 170.2*, 169.5*, 164.2*, 146.8*,

136.4*, 134.2*, 129.6, 129.4*, 128.6, 128.2*, 127.0*, 126.9*, 122.9*, 94.9, 72.6, 72.3*, 59.7*, 58.4*, 51.6*, 51.3*, 41.6, 41.0*, 40.9, 40.6, 40.4, 34.5*, 33.5, 25.2, 24.7, 24.4, 23.2, 22.9*, 22.1*, 21.5*, 15.7, 10.5. *Showed twinning. $[\alpha]_D^{24}$ -11.9 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 745.4538, -0.3 ppm (calculated for $C_{43}H_{61}N_4O_7^+$, 745.4540).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-phenylalinate (**118i**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (m,2H), 7.20 (m, 12H), 7.09 (m, 1H), 6.91 (m, 2H), 6.46 (dd, *J* = 41.0, 8.4, 1H), 6.21 (dd, *J* = 11.3, 8.0, 1H), 5.09 (m, 1H), 4.96 (m, 1H), 4.87 (m, 1H), 4.79 (s, 1H), 4.30 (m, 1H), 3.80 (s, 3H), 3.50 (m, 2H), 3.10 (m, 1H), 3.00 (m, 3H), 2.93 (m, 1H), 2.39 (s, 6H), 1.65 (m, 4H), 1.52 (m, 1H), 1.40 (m, 1H), 0.88 (m, 12H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.7, 170.6(1)*, 170.5(7)*, 170.3*, 169.5*, 164.2*, 147.0*, 137.6*, 136.5*, 134.2*, 129.5*, 129.3*, 128.8, 128.5(1), 128.4(6), 128.1*, 126.9*, 129.8(1), 126.7(8)*, 122.9*, 94.9, 72.6*, 69.5*, 59.6*, 58.3*, 51.4*, 51.3*, 41.8, 40.7, 40.4*, 40.0*, 35.8*, 34.5*, 24.6*, 24.3, 23.1, 22.8*, 21.9*, 21.5*. *Showed twinning. ν_{max} 3297.8, 3030.8, 2954.4, 2872.2, 2789.2, 1726.1, 1662.3, 1629.9, 1542.7, 1453.8, 1342.8, 1246.9, 1163.9, 1068.3, 1023.2, 969.7, 737.5, 700.0 cm^{-1}. $[\alpha]_D^{27}$ -32.7 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 801.4196, -0.2 ppm (calculated for $C_{46}H_{58}N_4NaO_7^+$, 801.4198).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118j**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 2H), 7.23 (m, 1H), 7.19 (m, 6H), 7.09 (ddd, J = 15.5, 7.6, 5.1, 1H), 6.90 (m, 2H), 6.48 (dd, J = 8.2, 4.1, 1H), 6.35 (dd, J = 70.5, 8.5, 1H), 5.16 (m, 1H), 4.98 (m, 1H), 4.88 (m, 1H), 4.80 (d, J = 3.7, 1H), 4.40 (tt, J = 8.6, 5.6, 1H), 3.81 (d, J = 1.3, 1H), 3.53 (ddd, J = 17.2, 13.9, 5.1, 1H), 3.10 (ddd, J = 14.0, 7.4, 3.0, 1H), 2.93 (m, 3H), 2.30 (d, J = 2.05, 6H), 1.84 (m, 1H), 1.77 (m, 2H), 1.64 (m, 4H), 1.52 (m, 1H), 1.15 (m, 1H), 0.95 (m, 6H), 0.90 (m, 9H), 0,86 (dd,

J =6.7, 1.6, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.8*, 170.9*, 170.7*, 170.2*, 169.5*, 164.2*, 146.8*, 136.4*, 134.2*, 129.6, 129.4*, 128.6, 128.2*, 127.0*, 126.9*, 122.9*, 94.9, 72.6, 72.3*, 59.7*, 58.4*, 51.6*, 51.3*, 41.6, 41.0*, 40.9, 40.5*, 34.5*, 33.5, 25.2, 24.7, 24.4, 23.2, 22.9*, 22.1*, 21.5*, 15.8, 10.5. *Showed twinning. ν_{max} 3300.5, 3065.9, 3029.6, 2957.7, 2873.9, 2785.4, 1725.5, 1662.1, 1629.8, 1542.7, 1454.8, 1341.9, 1249.3, 1167.7, 1124.9, 1050.3, 970.5, 734.6, 700.0 cm^{-1}. $[\alpha]_D^{26}$ -48.2 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z* 767.4354, -0.2 ppm (calculated for $C_{43}H_{60}N_4NaO_7^+$, 767.4354).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-phenylalaninate (**118k**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.31 (m, 3H), 7.20 (m, 11H), 7.08 (ddd, *J* = 15.5, 8.2, 5.2, 1H), 6.91 (m, 2H), 6.41 (dd, J = 29.3, 8.3, 1H), 6.18 (t, *J* = 8.2, 1H), 5.09 (dd, *J* = 8.6, 3.9, 1H), 4.96 (m, 1H), 4.88 (m, 1H), 4.79 (s, 1H), 4.29 (q, *J* = 7.4, 1H), 3.81 (s, 3H), 3.52 (m, 2H), 3.10 (m, 1H), 3.00 (m, 3H), 2.93 (m, 1H), 2.39 (d, *J* = 2.1, 6H), 1.66 (m, 4H), 1.52 (m, 1H), 1.40 (m, 1H), 0.88 (m, 12H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.8, 170.7*, 170.6, 170.4*, 169.5*, 164.3*, 147.0*, 137.6*, 136.5*, 134.2*, 129.6*, 129.3*, 128.9, 128.6*, 128.2*, 127.0*, 126.9, 126.9(0)*, 122.8(8)*, 94.9, 72.6*, 69.5*, 59.7*, 58.3*, 51.5*, 51.4*, 41.9, 40.7, 40.5*, 40.0*, 35.9*, 34.5*, 24.7*, 24.4, 23.1, 22.9*, 22.0*, 21.6*. *Showed twinning. ν_{max} 3298.1, 3030.8, 2954.4, 2871.6, 2788.9, 1726.0, 1662.3, 1629.9, 1543.5, 1453.3, 1342.0, 1246.1, 1163.6, 1066.2, 1022.9, 969.3, 808.1, 737.1, 700.0 cm^{-1}. $[\alpha]_D^{26}$ -33.2 (CH_2Cl_2). HR-ESI-TOF-MS [M + Na]$^+$ *m/z* 801.4196, -0.2 ppm (calculated for $C_{46}H_{58}N_4NaO_7^+$, 801.4198).

(*S*)-1-(((*S*)-1-(((*R,E*)-5-((*S*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118l**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 3H), 7.24 (m, 1H), 7.20 (m, 5H), 7.08 (m, 1H), 6.92 (m, 2H), 6.45 (dd, *J* = 13.9, 8.1, 1H), 6.13 (dd, *J* = 25.5, 8.5, 1H), 5.16 (m, 1H), 4.98

(m, 1H), 4.87 (m, 1H), 4.79 (d, *J* = 5.1, 1H), 4.38 (m, 1H), 3.81 (d, *J* = 2.3, 3H), 3.52 (td, *J* = 13.5, 5.2, 1H), 3.11 (m, 1H), 3.03 (m, 1H), 2.92 (m, 2H), 2.29 (s, 6H), 1.82 (m, 1H), 1.73 (m, 2H), 1.61 (m, 3H), 1.42 (m, 2H), 1.13 (m, 1H), 0.87 (m, 18H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.9*, 171.1, 170.8*, 170.4*, 169.6*, 164.2*, 146.8*, 136.4*, 134.3*, 129.5, 129.4*, 128.6, 128.2, 127.0, 126.9*, 122.9*, 94.9*, 72.6*, 72.5*, 59.7*, 58.4, 51.5, 51.3*, 41.6*, 40.8*, 40.7*, 40.3*, 34.6*, 33.4*, 25.2*, 24.6*, 24.4*, 23.2*, 22.9*, 21.9*, 21.5*, 15.7*, 10.5*. *Showed twinning. ν_{max} 3291.7, 3066.9, 2957.5, 2872.1, 2787.1, 1727.4, 1632.9, 1542.0, 1454.0, 1345.1, 1250.2, 1165.1, 1050.0, 970.7, 806.0, 735.1, 699.8, 516.2 cm^{-1}. $[\alpha]_D^{26}$ +8.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 745.4535, -0.7 ppm (calculated for $C_{43}H_{61}N_4O_7^+$, 745.4540).

(*S*)-1-(((*S*)-1-(((*R,E*)-5-((*R*)-2-benzyl-3-methoxy-5-oxo-2,5-dihydro-1H-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118m**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.30 (m, 3H), 7.25 (m, 1H), 7.20 (m, 5H), 7.08 (ddd, *J* = 15.5, 14.0, 5.1, 1H), 6.91 (m, 2H), 6.49 (dd, *J* = 13.1, 8.2, 1H), 6.22 (dd, *J* = 28.5, 8.5, 1H), 5.16 (m, 1H), 4.98 (m, 1H), 4.86 (m, 1H), 4.79 (d, *J* = 6.2, 1H), 4.39 (m, 1H), 3.81 (d, *J* = 2.6, 3H), 3.53 (td, *J* = 13.7, 5.2, 1H), 3.11 (ddd, *J* = 13.9, 7.7, 3.0, 1H), 3.03 (ddd, *J* = 27.4, 13.9, 6.3, 1H), 2.91 (m, 2H), 2.28 (d, *J* = 2.2, 6H), 1.82 (m, 1H), 1.73 (m, 2H), 1.59 (m, 3H), 1.42 (m, 2H), 1.13 (m, 1H), 0.86 (m, 18H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 177.8*, 171.0, 170.7*, 170.3*, 169.5*, 164.1*, 146.8*, 136.4*, 134.2*, 129.4, 129.3*, 128.5, 128.1, 126.9, 126.8*, 122.8*, 94.8*, 72.5*, 72.4*, 59.6*, 58.3, 51.4, 51.2*, 41.5*, 40.8*, 40.6*, 40.2*, 34.5*, 33.3*, 25.1*, 24.5*, 24.3*, 23.1*, 22.8*, 21.9*, 21.4*, 15.6*, 10.4*. *Showed twinning. ν_{max} 3291.2, 3031.6, 2957.8, 2872.8, 2786.9, 1727.3, 1633.4, 1542.4, 1455.3, 1345.4, 1250.6, 1165.4, 1049.9, 971.0, 805.4, 735.2, 699.9 cm^{-1}. $[\alpha]_D^{26}$ -23.7 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 745.4538, -0.3 ppm (calculated for $C_{43}H_{61}N_4O_7^+$, 745.4540).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-

5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**118n**). ν_{max} 3299.0, 3068.7, 2958.0, 2872.5, 2786.3, 1724.5, 1660.4, 1626.7, 1543.4, 1457.7, 1335.0, 1249.6, 1173.4, 1051.1, 981.7, 740.2, 698.7 cm^{-1}. $[\alpha]_D^{27}$ -42.4 (CH_2Cl_2). HR-ESI-TOF-MS $[M + H]^+$ *m/z* 711.4691, 0.0 ppm (calculated for $C_{40}H_{63}N_4O_7^+$, 711.4691).

(*S*)-1-(((*S*)-1-(((*S,E*)-5-((*S*)-2-isobutyl-3-methoxy-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxo-1-phenylpent-3-en-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-valinate (**118o**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (ddd, *J* = 17.6, 15.5, 1.8, 1H), 7.27 (Overlaps with chloroform signal, m, 2H), 7.21 (m, 1H), 7.17 (m, 2H), 6.99 (ddd, *J* = 15.6, 9.1, 5.1, 1H), 6.42 (dd, *J* = 8.1, 5.8, 1H), 6.23 (dd, *J* = 30.8, 8.4, 1H), 5.16 (m, 1H), 5.03 (d, *J* = 2.8, 1H), 4.94 (m, 1H), 4.66 (m, 1H), 4.38 (m, 1H), 3.85 (s, 3H), 2.91 (m, 2H), 2.81 (dd, *J* = 10.2, 2.2, 1H), 2.31 (d, *J* = 1.9, 6H), 2.01 (m, 1H), 1.76 (m, 6H), 1.62 (m, 2H), 1.52 (m, 1H), 0.91 (m, 24H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.6, 171.0*, 170.8*, 170.2*, 170.0*, 164.0*, 146.4*, 136.4*, 129.3*, 128.6*, 126.9, 123.1*, 93.6, 74.3, 72.3*, 58.6*, 58.5*, 51.6*, 51.3*, 41.5, 41.0*, 40.9, 40.5*, 39.0*, 27.7, 24.7*, 24.4, 24.2*, 23.7*, 23.2, 22.9*, 22.7*, 22.1*, 21.5, 19.6*, 19.4. *Showed twinning. ν_{max} 3291.1, 3074.5, 2957.4, 2872.7, 2787.5, 1724.4, 1660.5, 1626.9, 1544.7, 1458.4, 1336.1, 1227.6, 1171.1, 1049.4, 981.8, 809.5, 739.6, 699.2 cm^{-1}. $[\alpha]_D^{26}$ -40.7 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 719.4353, -0.1 ppm (calculated for $C_{39}H_{60}N_4NaO_7^+$, 719.4354).

Hydrogenation of the Enamide. In a vial 15.2 mg of gallinamide A was dissolved in 1.00 mL ethanol and ca. 7.6 mg of palladium on carbon (1% loading) was added in one portion, and the atmosphere was replaced with hydrogen. After stirring overnight the reaction was quenched with 1.0 mL of dimethylsulfoxide (DMSO) and stirred for 6 h at which point it was passed through a ca. 1 cm diatomaceous earth plug, prepared in a pipette with glass wool, with 10 mL of dichloromethane and the filtrate was washed with 10 mL of brine. The aqueous layer was extracted three times with equal volume DCM and the

combined organics were dried over $MgSO_4$. This material was concentrated to dryness by rotovap and purified by RP-SPE using two elutions 10% CH_3CN/H_2O and CH_3CN. The CH_3CN was further purified by RP-HPLC using an CH_3CN/H_2O elution.

(*S*)-1-(((*S*)-1-(((*S*)-5-((*S*)-3-methoxy-2-methyl-5-oxo-2,5-dihydro-1*H*-pyrrol-1-yl)-5-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)amino)-4-methyl-1-oxopentan-2-yl dimethyl-*L*-isoleucinate (**119**). ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.49 (dd, *J* = 8.3, 4.3, 1H), 6.36 (dd, *J* = 11.8, 7.7, 1H), 5.20 (m, 1H), 5.04 (d, *J* = 2.6, 1H), 4.56 (q, *J* = 6.6, 1H), 4.40 (m, 1H), 3.92 (m, 1H), 3.87 (d, *J* = 1.4, 3H), 2.94 (d, *J* = 10.1, 1H), 2.88 (m, 2H), 2.32 (s, 6H), 1.81 (m, 5H), 1.65 (m, 4H), 1.53 (m, 1H), 1.47 (t, *J* = 6.2, 3H), 1.18 (m, 1H), 1.13 (d, *J* = 6.5, 3H), 0.94 (m, 12H), 0.88 (m, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ 180.7*, 172.3*, 170.9*, 170.8*, 170.2, 169.6, 92.8, 72.6, 72.5*, 58.7*, 55.6*, 51.6*, 45.5*, 41.6, 41.5*, 41.1*, 34.0*, 33.5*, 30.9*, 25.2, 24.8*, 24.5, 23.2, 23.0*, 22.1*, 21.5*, 20.8*, 17.1*, 15.8*, 10.5. *Showed twinning. ν_{max} 3291.6, 3083.2, 2959.5, 2874.3, 2787.9, 1728.4, 1642.1, 1548.4, 1457.1, 1377.0, 1329.0, 1294.4, 1247.5, 1155.9, 1052.2, 983.9, 810.4, 693.2 cm^{-1}. $[\alpha]_D^{26}$ -466 (CH_2Cl_2). HR-ESI-TOF-MS $[M + Na]^+$ *m/z* 617.3888, 0.5 ppm (calculated for $C_{31}H_{55}N_4NaO_7^+$, 617.3885).

5.6 Acknowledgments

This work would not have been possible without Bailey Miller who was involved in the planning of this synthesis project as well as conducting HPLC purifications of the full synthetic products.

This chapter, in part, is currently being prepared for submission for publication of the material. Boudreau, Paul D.; Miller, Bailey; Gerwick, William H. The dissertation author is the primary investigator and author of this material.

This chapter, in part, is reproduced, with permission, from Phytochemistry Reviews Gerwick, L.; Boudreau, P.; Choi, H.; Mascuch, S.; Villa, F. A.; Balunas, M. J.; Malloy, K.

and Teasdale, M. E.; Rowley, D. C.; Gerwick, W. H. **2013**, volume *12*, pages 459-465.

Chapter 6

Future Directions for Marine Natural Products Chemistry

6.1 Abstract

The multidimensional profiling of a single strain of cyanobacteria, *Moorea producens* JHB, with molecular networking provided a great deal of new insight into its metabolome. A core feature of MS-based molecular networking is that the data sets are eminently scalable, allowing the incorporation of data from multiple samplings, strains, time-points, or collections. Knowing both the insights provided by molecular networking to a single collection, and the ease in scaling this technique to multiple collections, an effort was made to profile not just one strain but dozens of strains by molecular networking. Very quickly a great deal of data was generated, and the first attempt to parse this data set was to use standards of known metabolites to identify strains producing a desired compound. This was effective but it also drew attention to several collections with high homology between their chromatograms. As an exemplary case, all of those strains producing palmyramide A were derived from field collections at the same site, Palmyra atoll. Probing further it appears that these collections may be composed, at least in part, of a closely related *Moorea producens*,

thus suggesting a reliable link between chemotype and biological identity. If validated, this information could provide an alternative to 16S rRNA gene sequence analysis to identify cyanobacterial collections.

6.2 Introduction

Chemotype, defined here as a working taxanomic definition based on the presence of one or more natural products, is commonly used in the Gerwick laboratory isolation workflow. When bioassay data is received from collaborators or in-house results, dereplication is usually necessary to ensure that any hits are not the result of previously known compounds. As an example, if a VLC fraction is shown to be cytotoxic in a cancer cell line assay but is later found by NMR or MS dereplication to possess dolastatin 12 (**120**), then that fraction is rejected from further study. **120** is a potent cytotoxin and likely responsible for the activity seen in the prior assay, and as such it is not worth the expense and effort to further purify it and verify **120** as the active species.

This chemotype concept can be expanded beyond the simple definition of 'a dolastatin 12 producer', by drawing on other sources of information such as a field identification or collection site. In this regard, the 'chemotype' concept can provide a narrower picture of an organism. For example, from several collections from Palmyra atoll that were field identified as *Lyngbya* (*Moorea*), the compounds malyngamide C (**121**) and palmyramide A (**122**) were reisolated. The original collection that led to the structure elucidation of **122** was a collection of *Lyngbya majuscula* (currently referred to as *Moorea producens* PAL) from Palmyra, the extract of which was noted to contain **122**, curacin D (**123**), and **121** [Taniguchi et al., 2010]. Since then cyanobacterial collections from Palmyra that were field identified as *Moorea* and found to contain a cytotoxic fractions are now expected to contain these compounds. To validate cytotoxic hits from these collections, often only the simplest of dereplication tools are used, a LR-LCMS run or a ^{1}H NMR of the crude extract or VLC

fraction. When compounds **121-123** are detected again, the producing organism is simply assumed to be the same *Moorea producens* strain as the original collection. However, due to expense and time considerations, these assumptions have not yet been verified by 16S rRNA gene comparisons.

When networked together the most striking result from analyzing the crude extract library was the way in which this chemotype appeared in the dataset. Several collections appeared to be related to this same chemotype, and from further inspection of their chromatograms, they showed incredibly high homology. This suggests a potential relatedness between these collections that with further effort may provide a new way to identify cyanobacterial strains.

6.3 Results and Discussion

6.3.1 Molecular Networking Profile of *Moorea producens* Strains

From the Gerwick lab extract library, samples of crude extracts were prepared by passing through a C-18 column with acetonitrile. These samples were compared by a standard LCMS method. Of the over dozen samples containing palmyramide A (**122**), manual verification (inspection of the chromatogram to check the retention time and abundance) showed that several of these extracts had highly homologous chromatograms that contained compounds **121-123**.

Included within the data set was the crude extract for the original collection (Extract 1859) that led to the isolate on palmyramide A. It was not surprising then, to observe in this chromatogram the presence of **121-123**. It was surprising, however, to observe several different extracts that had the nearly identical chromatograms, suggesting that these were collections of the same or a closely related cyanobacterium. Based on manual identification of these collections from MS-profiles, there appeared to be four different subtypes (Figure 1

and Table 1). Type-1 was essentially identical to the original collection, Type-2 was highly similar but less abundant in all peaks, Type 3 was the same but additional peaks unique to that chromatogram were present (mainly a 1197 $[M+H]^+$ at a retention time of 11 min), and Type-4 again with unique peaks, but different from those present in Type-3 (mainly a pair of peaks, *m/z* 360 and 326 at 18.6 and 18.8 min) (Figure 1).

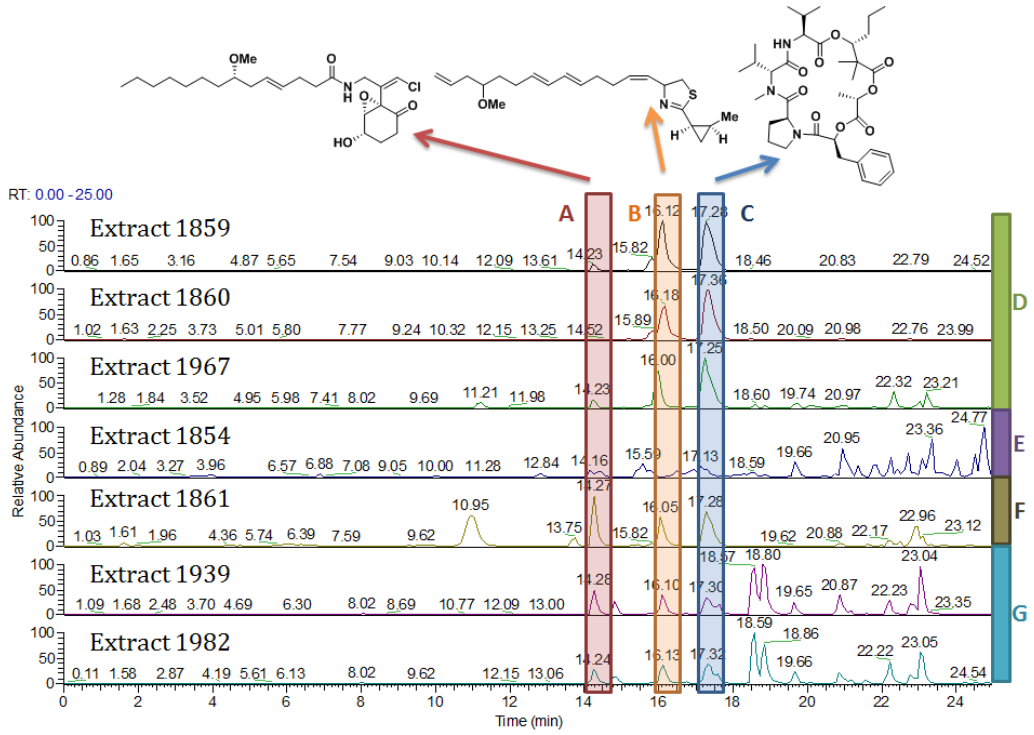


Figure 6.1: *Comparison of the chromatograms with high homology to the palmyramide A (**122**) producing strain.* (A-C) Location of the malyngamide C (**121**), palmyramide A (**122**), and curacin D (**123**) peaks within the chromatogram, respectively. The four chromatogram chemotypes observed: (D) Subtype-1 essentially identical to the original strain, (E) Subtype-2 highly similar but less abundant of all peaks, (F) Subtype-3 the same peaks are present but with additional peaks unique to the chromatogram, and (G) Subype-4 again with unique peaks, but those peaks are dissimilar to Subtype-3.

Knowing that many of the collections of filamentous cyanobacteria are in fact assemblages of multiple strains, the appearance of different chemotypes containing additional masses may be explained by the presence of other cyanobacteria in the collection. Subype-3 was field identified as *Hormothamnion enteromorphoides*, and it may be that in fact this is an assemblage of both that species and a strain related to *Moorea producens* PAL. Extract

1982 in Subtype-4 was identified as a mixture of a 'green algae' and *Lyngbya*, and the fact that both collections contain similar peaks may indicate the presence of another species with a strain related to *Moorea producens* PAL in both these collections, though we have too little information at present to determine if it is *Hormothamnion enteromorphoides*, a green algae, or something else.

Table 6.1: *Compared collection data of PAL-subtypes. [a]This extract led to the original isolation of palmyramide A (**122**).*

Extract Code	Collection Code	Collection Site	Field Identification	PAL-Subtype
1859[a]	PAL 15-AUG-08-1	Palmyra Atoll	Black *Lyngbya*	Subtype-1
1967	PAL 31-AUG-09-1	Palmyra Atoll	Cyanobacterial assemblage	Subtype-1
1860	PAL 15-AUG-08-11	Palmyra Atoll	*Oscillatoriales*	Subtype-1
1854	PAL 15-AUG-08-9	Palmyra Atoll	Red filamentous turf	Subtype-2
1861	PAL 17-AUG-08-1	Palmyra Atoll	*Hormothamnion enteromorphoides*	Subtype-3
1939	PAL 31-JUL-09-1	Palmyra Atoll	Brown *Lyngbya*	Subtype-4
1982	PAL 31-AUG-09-2	Palmyra Atoll	Green algae and *Lyngbya*	Subtype-4

Another explanation of these profiles is that the same organism was collected under conditions that led to differential expression of their metabolic pathways. To probe this further, another avenue of exploration was available. Several collections of field identified *Moorea producens* were collected alive from Palmyra Atoll and grown in the Gerwick lab culture collection, including the original collection that had led to the isolation of palmyramide A (**122**), *Moorea producens* PAL. An extract of each of these was also run by LCMS and these chromatograms were compared with one another as well as the field collected chemotypes. These appeared to be similar to the field collected chemotype-1; however, these cultures did not produce **121** in detectable amounts, and there was a new peak *m/z* 627 at a retention time of 2.5 min (Figure 2).

The loss of **121** from the extracts of the cultured cyanobacteria above could be explained by a change in gene expression triggered by growing under laboratory culture conditions. Alternatively, there could have been a loss of a cyanobacterium that was present in the field assemblage but not in the laboratory monocultures. Another possibility is that it might result from the loss of a heterotrophic bacterium that was growing in association with

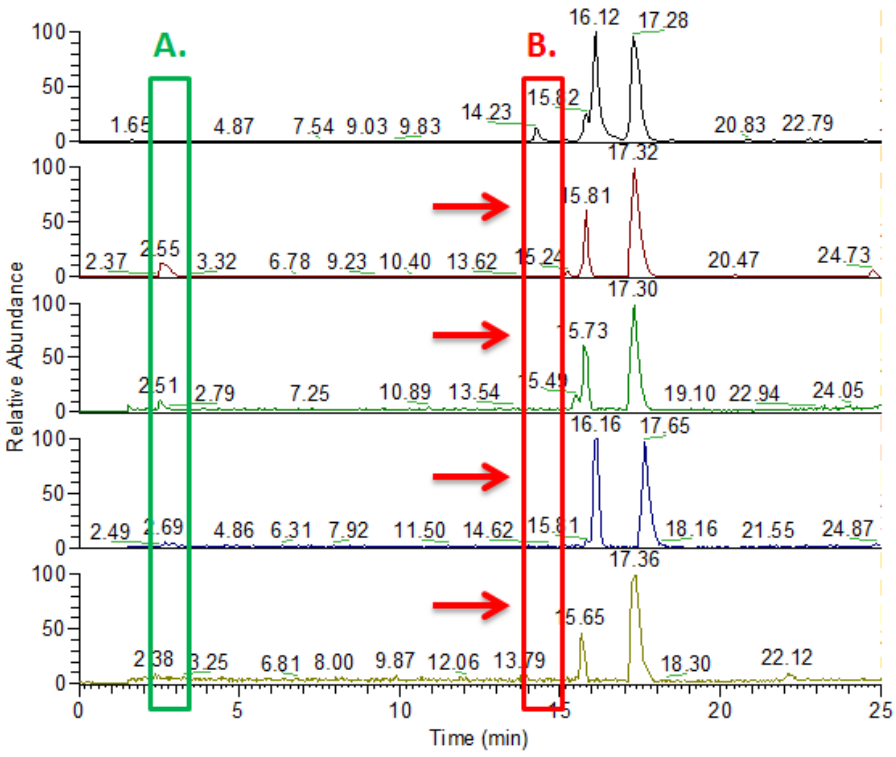


Figure 6.2: *Comparison of the chromatograms of field collected* Moorea producens *PAL and the cultured strains from Palmyra.* (A) The cultured strains have a much greater amount of an unknown mass at *m/z* 627, and (B) no longer produce malyngamide C (**121**) in detectable amounts.

the cyanobacterium in the wild.

The new peak at *m/z* 627 was also present in the original field isolations, however, in very small quantities just emerging from the baseline. This new shared metabolite had been observed in other work to be common to *Moorea producens* JHB [Kleigrewe et al., 2015], but its structure was unknown.

6.3.2 Structure Determination of the Shared Metabolite

Material from a fresh collection of *Moorea producens* JHB led to the isolation of the compound responsible for the *m/z* 627 peak. Using a combination of HRMS fragmentation and 2D-NMR analysis its structure could be determined (Table 2, Figure 3). By inspection of the proton spectrum, structural similarity to the previously described hectoramide (**96**)

was obvious (see Chapter 4) [Boudreau et al., 2015]. The presence of an additional TOCSY spin system in the aromatic region coming from two doublets and another *O*-methyl singlet, suggested another *O*-methyl tyrosine based residue. In both hectoramide and this compound, an Mpla residue was present. In the new compound the additional residue was clearly similar to Mpla, but HMBC correlations between the α position carbon of the new residue and an unusual species which was a singlet in the proton spectrum at δ_H 2.32 and had an HSQC correlation to δ_C 41.6 showed that it was in fact an *N,N*-dimethyl-*O*-methyl tyrosine (Dmty) residue. HMBC correlations between these residues and analysis of $HRMS^2$ fragmentation confirmed this structure to be (H2NOC-emphN-Me-Val)-(*N*-Me-Val)-(Mpla)-(Dmty), for which we give the trivial name 'taxamide' (**124**).

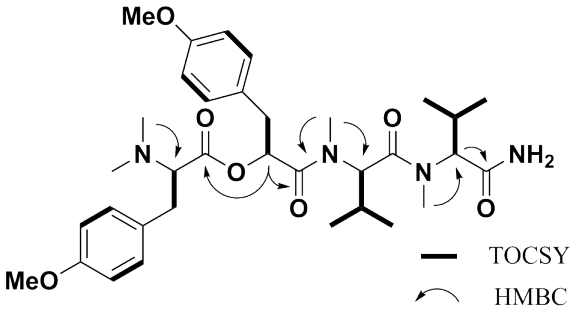


Figure 6.3: *Key NMR features of taxamide (**124**).*

It is possible that the **96** previously isolated is in fact a degradation product of **124**, produced by hydrolysis of the ester bond connecting the Mpla and Dmty residues. This is reasonable as ester hydrolysis is a common occurrence in natural products isolation schemes; this is especially likely given that the VLC fraction yielding **96** was several years old whereas the extract that provided **124** was fresh. If **124** is to be used as a taxonomic marker of the 'PAL-chemotype', then the extract must be created and analyzed before extensive decomposition can take place. This is especially the case because following the loss of the *N,N* dimethyl tail, the compound no longer easily ionizes by ESI MS (to ionize, **96** required a special buffer condition not commonly used in LCMS analysis). Thus, **96** cannot be used as a proxy for **124** in this chemotaxonomic analysis.

Table 6.2: *Data table of taxamide (**124**). [a]Carbon species did not resolve. [b]Minor rotomer at 2.98, s. [c]Minor rotomer at 2.81, s.*

Residue	Position	δ_C, type	δ_H (J in Hz)	HMBC	TOCSY
N-Me-Val-1	1	171.3, C	–		
	2	62.1, CH	4.53 d (11.2)	1, 3-6	3
	3	25.4, CH	2.22 m	2, 4, 5	2, 4, 5
	4	19.6[a], CH_3	0.97 d (6.4)	2, 3, 5	3
	5	18.1, CH_3	0.67 d (6.6)	2-4	3
	6	30.6, CH_3[b]	2.89 s	2, 7	
N-Me-Val-2	7	171.8, C	–		
	8	58.1, CH	5.14 d (10.8)	7, 9-13	9
	9	27.3, CH_2	2.28 m	8, 11	8, 10, 11
	10	19.6[a], CH_3	0.88 d (6.7)	8, 9, 11	9
	11	18.2, CH_3	0.80 d (6.7)	8-10	9
	12	29.8, CH_3[c]	2.88 s	8, 13	
Mpla	13	169.9, C	–		
	14	71.7, CH	5.33 m	13, 15, 16, 21	15
	15	36.4, CH_2	2.98 m	14, 16, 17	14
	16	128.0, C	–		
	17	130.4, CH	7.15 d (8.5)	self, 15, 19	18
	18	114.0, CH	6.83 d (8.5)	self, 16, 19	17
	19	158.8, C	–		
	20	55.3, CH_3	3.78 s	19	
Dmty	21	170.9, C	–		
	22	69.5, CH	3.41 m	21, 23, 24, 29	23
	23a	35.2, CH_2	2.91 m	22, 24	22
	23b	35.2, CH_2	2.82 m	22, 24	22
	24	129.7, C	–		
	25	129.9, CH	6.86 d (8.3)	self, 23, 27	26
	26	113.7, CH	6.71 d (8.2)	self, 24, 27	25
	27	158.1, C	–		
	28	55.1, CH_3	3.75 s	27	
	29	41.6, CH_3	2.32 s	self, 22	

Unfortunately, the observation that the taxamide-producing 'PAL-chemotype' that also contains **122**, and **123**, has not yet been backed up by the traditional method of 16S rRNA gene identification. Efforts are underway in the Gerwick lab to complete this

verification, comparing both the 16S gene sequence of the strains growing in culture as well as field collections of all four chemotypes. If verified by 16S rRNA gene sequence that the presence of these three compounds is rigorously linked with the strain identity of *Moorea producens* PAL, then this opens up the potential to identify field collections as *Moorea producens* PAL not by 16S gene sequencing alone but also by LCMS comparison against standards of the three compounds that define the chemotype. This technique holds the advantage of being much more rapid and easy, and also avoids the pitfall of defining the chemotype too broadly. There are many examples in the literature of different strains producing the same compound, but because **122**, **123**, and **124** are expected to come from three separate biosynthetic pathways, the likelihood of all three being heterologously transferred from another bacterium or all being maintained by a closely related strain after a phylogenetic divergence in their evolution is unlikely.

We did find that the closely related strains of *Moorea producens* JHB and *Moorea producens* PAL share the presence of taxamide, but they do not share any other major metabolites, specifically hectochlorin (**89**) and jamaicamide A (**90**) for *Moorea producens* JHB and **122** and **123** for *Moorea producens* PAL. Because the curacins and the jamaicamides are known to have somewhat related biosyntheses [Gu et al., 2009], it is possible that they result from a gene diversification or horizontal gene transfer event. This observation suggests that this chemotype may be well fitted to a subspecies definition within this genus. *Moorea producens* JHB is the only strain known to produce the jamaicamides or hectochlorin, as these compounds have not been observed in any of the PAL-chemotype strains or other *Moorea producens* strains; without other strains that compare more closely to JHB, the definition of production of **124**, **89**, and **90** as a rigorous definition for a 'JHB-chemotype' of *Moorea producens* will be tentative.

6.4 Conclusions

When conducting a field taxonomic identification of a collection of cyanobacteria, the best that can generally be anticipated is an indication of the genus. Morphology is a poor marker for species identity [Engene et al., 2012], but having few alternatives, it is relied upon. Knowing now the strain specific association of **124** to *Moorea* collections, it may be possible to replace morphology-based identifications with an LCMS-based chemotype identification. This might even be reducible to a TLC-based method that could be easily utilized in the field for these identifications. Knowing that the 'PAL-chemotype' and 'JHB-chemotype' appear to be specific to a subspecies definition, it may be possible to replace 16S rRNA gene identification with a chemotype identification when all of the known metabolites are detected by LCMS. The acceptance of this methodology by the broader natural products community will require more data that compares 16S rRNA gene data with LCMS metabolite profiles to validate these and other cases. However, if validated, then utilization of LCMS or TLC-based taxonomic identifications will be much more rapid than existing genetic methods and will be more usable in field situations.

6.5 Material and Methods

6.5.1 Molecular Network Generation

Samples of the crude extracts in the Gerwick lab fraction library are stored in 1:1 EtOH/iso-octane at either 10 mg/mL or 100 mg/mL concentration. Sufficient volume to provide 3 mg of total crude extract was loaded onto a 100 mg/1 mL C-18 RP-SPE column from Agilent, prepared with three column volumes of CH_3CN, then the crude extract was eluted off with 3 mL of CH_3CN to afford a solution of ca. 1 mg/mL concentration. A blank control was prepared using pure 1:1 EtOH/iso-octane, but otherwise prepared as with all the

cyanobacterial extracts. From all of the crude extract solutions, 10 μL of sample was profiled on a Thermo Finnigan LCQ Advantage Max mass spectrometer using a standard LCMS gradient. To reduce the risk of cross contamination a short method, with an accelerated gradient elution, was run with an injection of pure CH_3CN to reduce column bleed into the next injection. It should be noted that after the most concentrated samples column bleed in the next sample was noted including with **122**, but manually this was easily identified because in the first run **122** had a sharp peak with a retention time that matched other runs, but in the contaminating runs it was present at very low ion counts throughout the chromatogram.

6.5.2 NMR Acquisition

NMR spectra were collected on a Varian Unity 500 MHz (500 MHz and 125 MHz for the ^{1}H and ^{13}C nuclei respectively) and a Bruker 600 MHz (600 MHz and 150 MHz for the 1H and ^{13}C nuclei respectively) with 1.7 mm inverse cryo-probeusing using $CDCl_3$ from Cambridge Isotope Laboratories, Inc. 99.8% D containing 0.03% *v/v* trimethylsilane (δ_H 0.0 and δ_C 77.16 as internal standards using trimethylsilane and $CDCl_3$, respectively).

6.6 Acknowledgments

Ariana Remmel purified the taxamide (**124**) from *Moorea producens* JHB, Evgenia Glukhov helped to prepare and run samples of the crude extract library for networking.

This chapter, in part, is currently being prepared for submission for publication of the material. Boudreau, Paul D.; Remmel, Ariana; Glukhov, Evgenia; Dorrestein, Pieter C.; Gerwick, William H. The dissertation author is the primary investigator and author of this material.

Chapter 7

Conclusions and Final Notes

7.1 Introduction

Arguing that using orthogonality in natural products isolation workflows increases their productivity might strike some as a truism. If anyone who tries different techniques can reasonably expect to see different things, and naturally applying a new workflow can be expected to provide new insights. The greater insight is that all techniques and workflows are myopic, with their own specific limitations, that can often only be overcome by using complementing orthogonal approaches.

Rare in the natural products community are those who believe that research effort should not be spent on developing new techniques to isolate and identify compounds of interest. Strikingly common, however, are those who believe that the value of traditional techniques is bankrupt, even as every year researchers following traditional pathways continue to find new analogs and entirely novel structural scaffolds from nature. The future of natural products chemistry will not center on an abandonment of traditional isolation workflows, but rather, in developing new techniques to complement them, access new chemical space, and better understand the role of small molecule natural products in the environment and in pharmacology.

7.2 Discussion

Traditional extraction and structure or bioassay guided isolation is often pejoratively labeled ‘grind and find’. What this adage highlights, however, are not the worst but the best components of traditional natural products isolation work. It requires little training or experience to ‘grind’ up a sample and perform a simple lipophilic extract. It is also easy to ‘find’ compounds by screening fractions using ^{1}H NMR or bioassay guided fractionation (though more involved bioassays require more extensive facilities and experience, the brine shrimp or goldfish toxicity assays can be said to be available to almost any research group).

As discussed in Chapter 2, the isolation of the viequeamides followed a straight forward sequence of isolating from the VLC fraction those compounds responsible for the distinctive ^{1}H NMR signals that were characteristic of marine cyanobacterial metabolites. The best components of traditional natural products isolation are ‘grind and find’; the drawbacks of these approaches are the tendency to reisolate known compounds or to expend considerable effort to find that the active component of a mixture is present in such small titer that determining its structure is nearly impossible. In the viequeamides case, the presence of *N*-methyl amide rotomers confounded the purification of viequeamides B-F (**26-30**), making structure determination by NMR difficult. As such, an orthogonal technique, MS^{2}-based fragmentation pattern analysis, was used to elucidate the structures of these compounds.

The presentation of the viequeamides story in this way does not reveal the considerable additional effort spent prior to that work which led to the reisolation of several known compounds. Palmyramide A (**122**) [Taniguchi et al., 2010], octadec-5-yne-7Z,9Z,12Z-trienoic acid (**125**) [Paul and Fenical, 1980], dolastatin 12 (**120**) [Harrigan et al., 1998], and malyngamide C (**121**) [Ainslie et al., 1985], were found from several separate isolation and structure elucidation projects prior to working on the collection from Vieques island that led to the isolation of the viequeamides. As these other projects show, ‘grind and find’ leads to finding a lot of compounds; sadly, not all of which are new or noteworthy.

Taken as a whole, this experience shows both the most attractive and problematic features of the traditional natural products isolation: high productivity and high reisolation rates. Facing the prospect of having to repeat of this experience, an attempt was made to see if other techniques were available as dereplication tools. In Chapter 3 the use of molecular networking showed that this tool could effectively dereplicate known compounds in the extracts of marine filamentous cyanobacteria. The cluster of nodes present within the network of *Moorea bouillonii* successfully showed not only the known compounds lyngbyabellin A (**86**) [Luesch et al., 2000] and apratoxin A (**87**) [Luesch et al., 2001b], but also several nodes that suggested structural analogs of these compounds were present in the extract.

Following up on these results, as discussed in Chapter 4, a more rigorous and in depth attempt to profile the minor metabolites identified from the molecular network of *Moorea producens* JHB found several previously unreported analogs of the major compounds **89** and **90**. This profile of the strain's metabolome provided new insight, even though *Moorea producens* JHB has been extensively studied before. Which supports the broader conjecture that the application of novel techniques can provide valuable insight in what many might consider well-studied areas.

Research in the Gerwick lab on gallinamide A (**106**) faced an altogether different problem. Identified from a fraction screening effort that showed the parent fraction to be a potent inhibitor of the cysteine protease cathepsin L, the reisolated pure compound was found to be a potent, irreversible, and selective inhibitor of cathepsin L [Miller et al., 2014]. At the end of this effort, however, no material remained to continue on with *in vivo* or animal studies. In many cases, this is how isolation projects end, with little ability to continue a project or further collect material. In this case, total synthesis was deemed a favorable route to more material because it also offered the potential to provide synthetic analogs of **106** that might elucidate the structure-activity relationship of the functional groups within the

molecule that determine its striking activity in the cathepsin assay. The work covered in Chapter 6 sought to provide a suite of compounds to probe these relationships, and fill the gap left after the consumption of all the original material in the bioassay guided workflow.

Other tactics have been used when faced with challenges such as those seen in the viequeamides and gallinamides projects. In point of fact, approaches that seek to overcome the limitations of repeated reisolation of known metabolites or limited scale, are common. Many researchers are looking to new sources to reduce the chance of rediscovery by profiling the metabolomes of organisms from novel genera, or those from extreme or poorly explored environments, such as the work surrounding culturing and profiling organisms from the deep sea [Thronburg et al., 2010, Andrianasolo et al., 2011, Pettit, 2011]. Even the reisolation of a known metabolite, however, can be highly informative from an ecological perspective. A recent project in the Gerwick lab led to the reisolation of spongosine from a culture of *Vibrio harveyi* isolated from the sponge *Tectitethya crypta*, which was the original source for the isolations of the 'spongonucleosides'. Showing that at least some of the spongonucleosides are produced by bacterial sources suggests that perhaps there is also a bacterial source for the important pharmaceutical arabinonucleosides, and that ecologically the sponge may be in a symbiosis to accumulate these compounds from its bacterial occupants [Bertin et al., 2015].

Perhaps little has done more for the natural products community's understanding of the great diversity of small molecule metabolites than the explosion of research effort in gene sequencing and genome mining. Not only have entirely new workflows been developed that are divorced from tradition natural products isolation, but the sequencing of productive strains has revealed a hidden diversity of silent pathways that were wholly unknown before [Jensen et al., 2014, Udwary et al., 2007, Liu et al., 2013]. Productive as this work has been, it is important to keep the greater frame of natural products chemistry in mind. The ability to identify biosynthetic gene clusters within an organism's genome provides a means

to both answer difficult questions and overcome hurdles in isolation and production. The true producer of a compound in a close symbiosis between organisms can be elucidated from their respective genomes. Alternatively, through heterologous expression, a compound can be provided on large scale which either may be of limited availability, or is explicitly unavailable from an organism.

In the not too distant past the standard of taxonomic identification and classification to report an organism in the literature was a detailed morphological description. Now, the standard has shifted towards both the morphological description and a description of characteristic gene sequences, usually a portion of the ribosomal rRNA. It is not difficult to imagine, however, that there will come a time when, through the reduced cost of genome sequencing and processing, identifying an organism may require a whole genome sequence. As with the insights provided by a large scale effort to determine gene sequence similarity between organisms, the large scale effort to sequence genomes underway today will also likely provide new insights into evolutionary history and natural products biosynthesis.

The availability of these new and overlapping forms of data about an organism's identity offers new opportunities to researchers. In Chapter 6, preliminary efforts were described to link a metabolomic profile to a chemotype which could be rigorously linked to a taxonomic identity within the *Moorea* genus was detailed. In the future, other researchers in similar positions will be able to define similar chemotype-to-taxonomic links within or for other genera and other species. This work was only possible because of access to new tools, mainly molecular networking, available to the Gerwick lab as they were in development; illustrating both how quickly new approaches can be incorporated into the isolation workflow and the way in which new tools provide, and will continue to provide, previously unrealized insights into the broader field of marine natural products.

Appendix A

Supplementary Information for Chapter 2

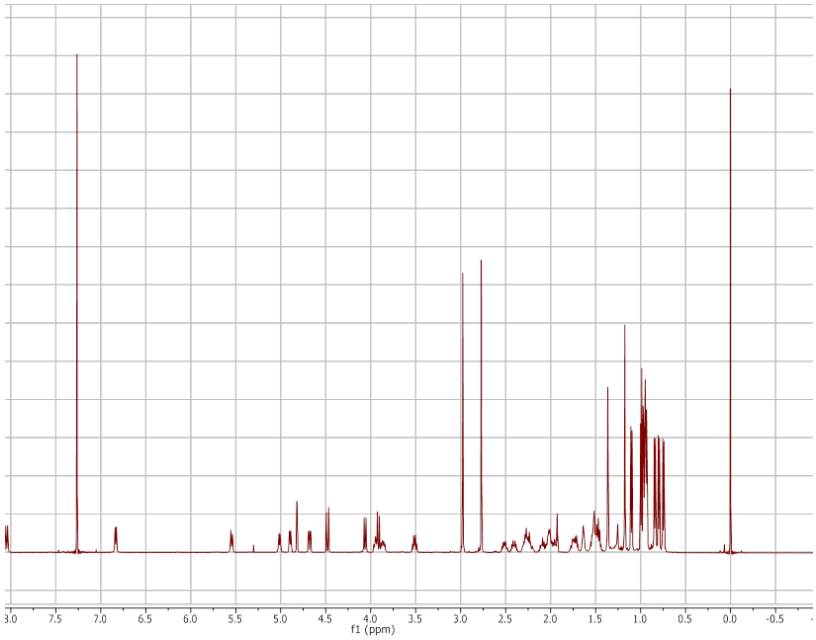


Figure A.1: ^{1}H *NMR (500 MHz,* $CDCl_3$*) spectrum of viequeamide A (**25**).*

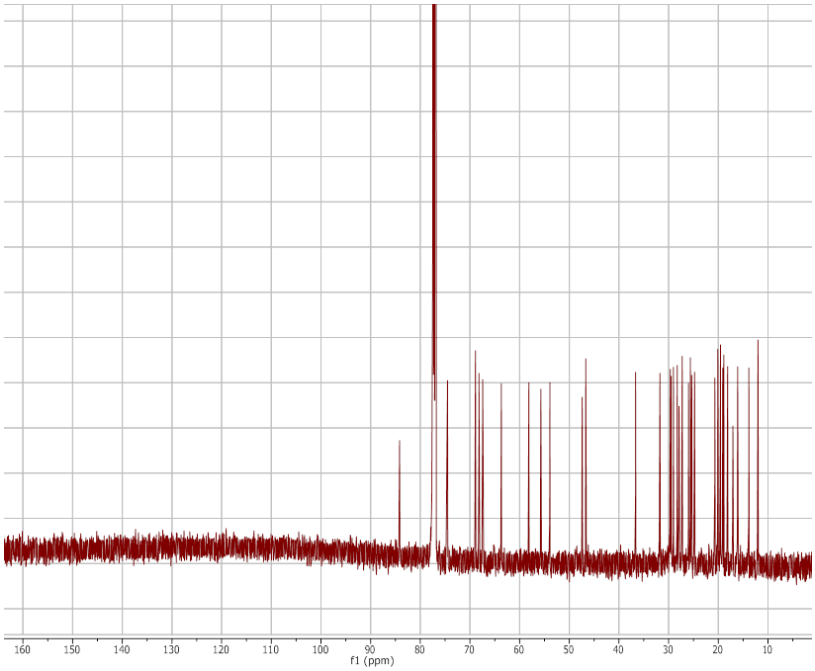


Figure A.2: *^{13}C NMR (125 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

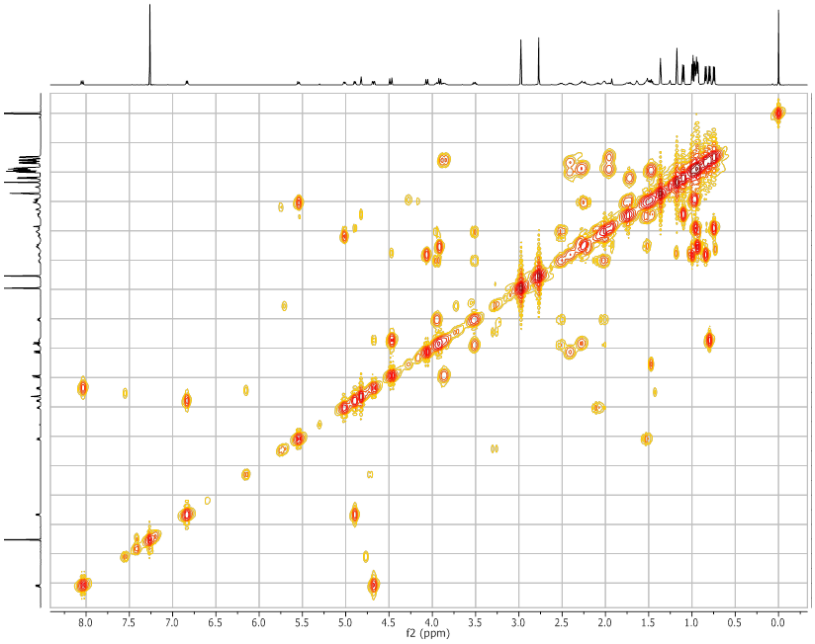


Figure A.3: *COSY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

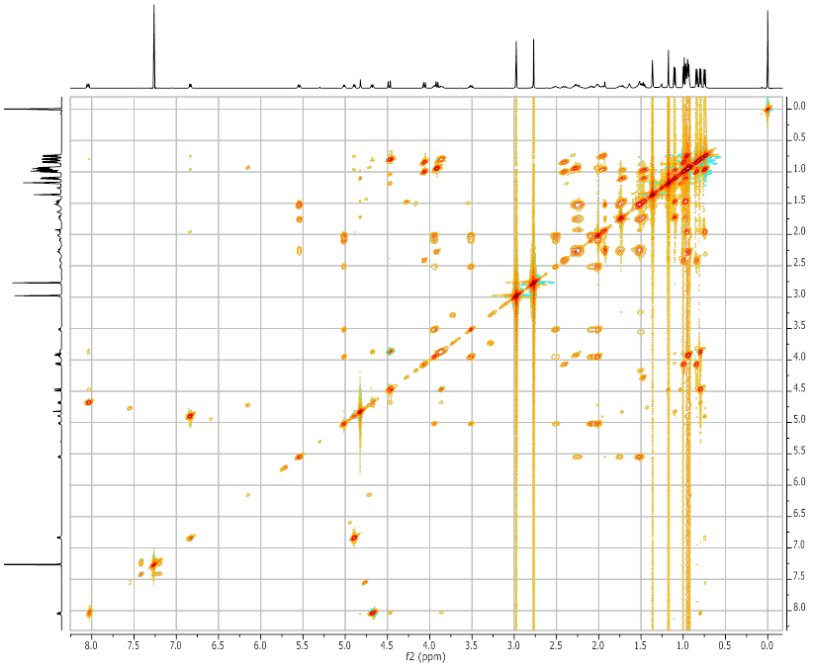


Figure A.4: *TOCSY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

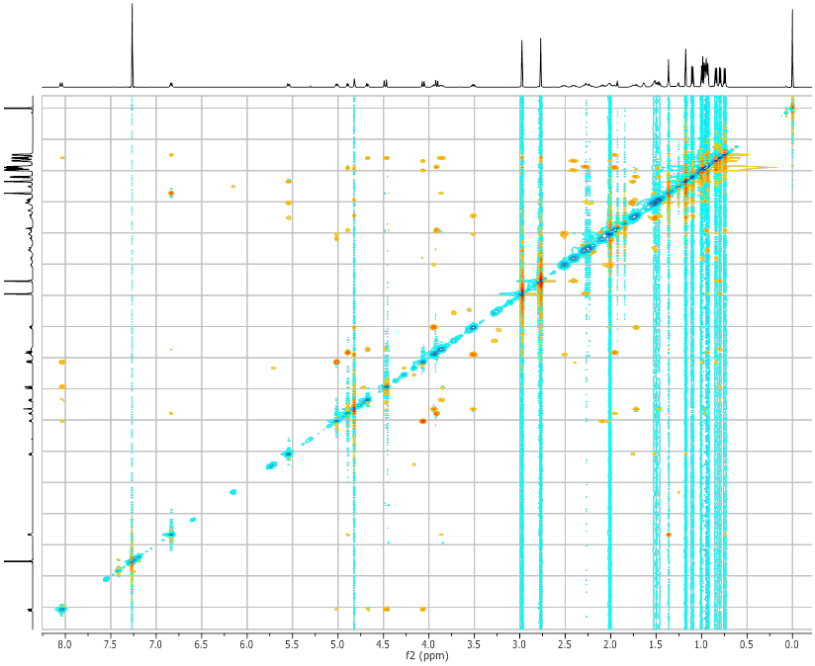


Figure A.5: *ROESY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

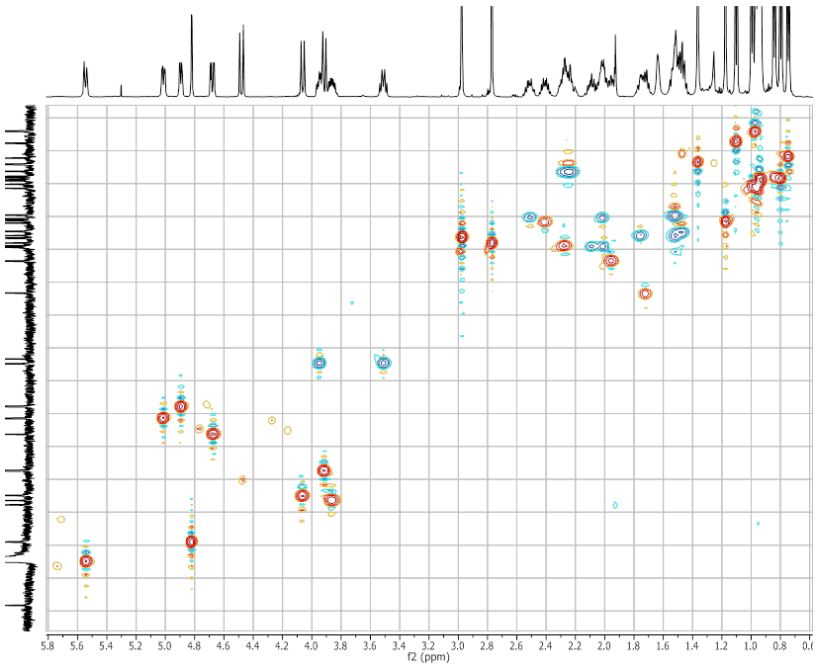


Figure A.6: *HSQY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

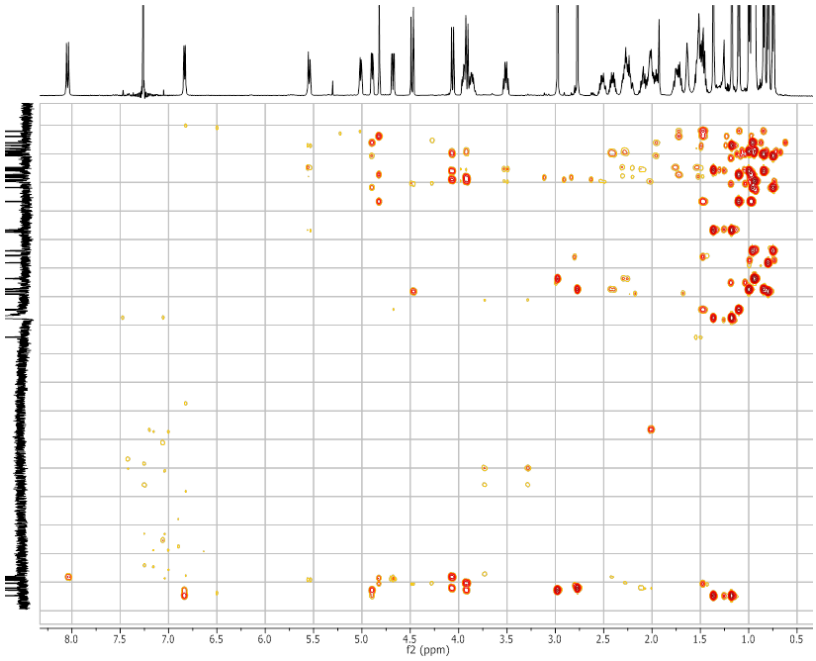


Figure A.7: *HMBC-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide A (**25**).*

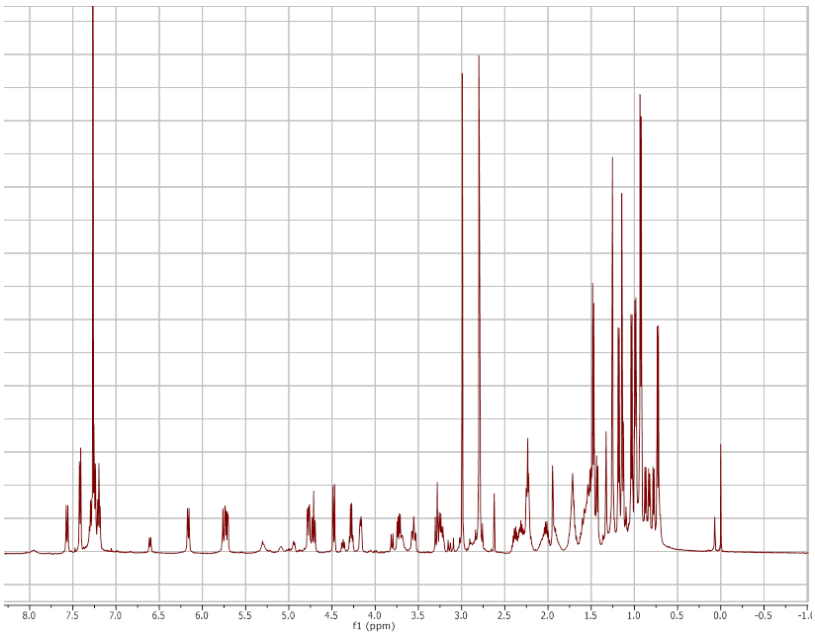


Figure A.8: ^{1}*H-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

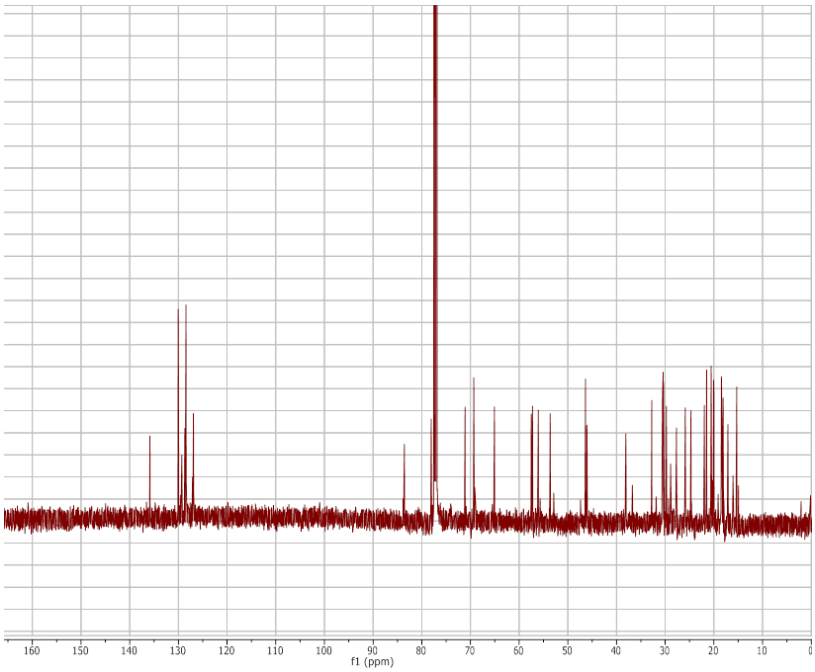


Figure A.9: ^{13}*C-NMR (125500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

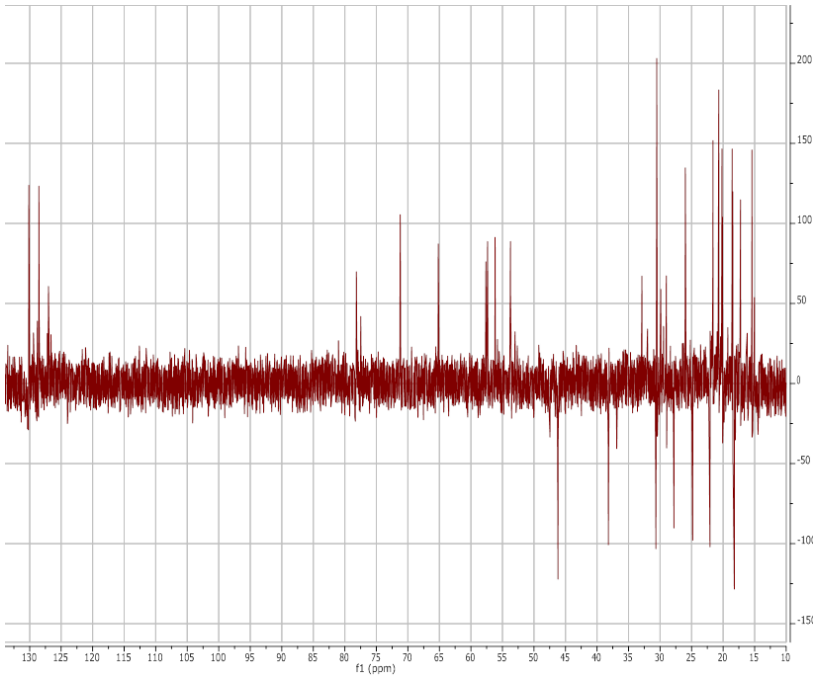


Figure A.10: *DEPT-135-NMR (125 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

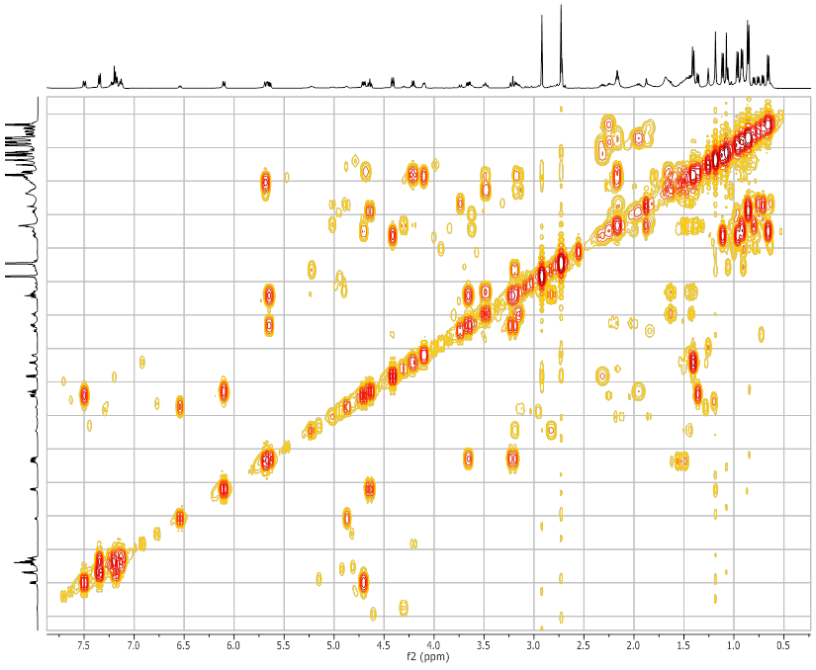


Figure A.11: *COSY-NMR (125 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

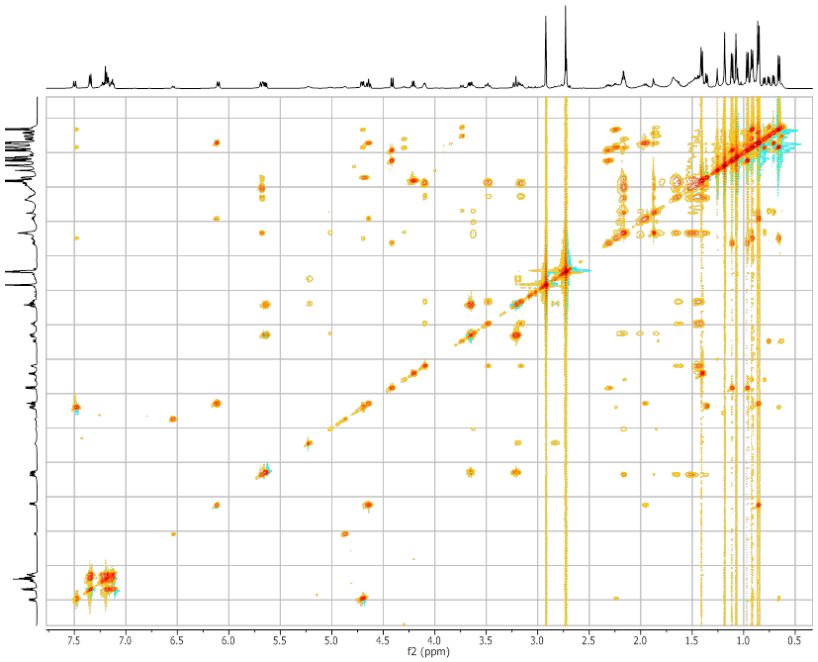


Figure A.12: *COSY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

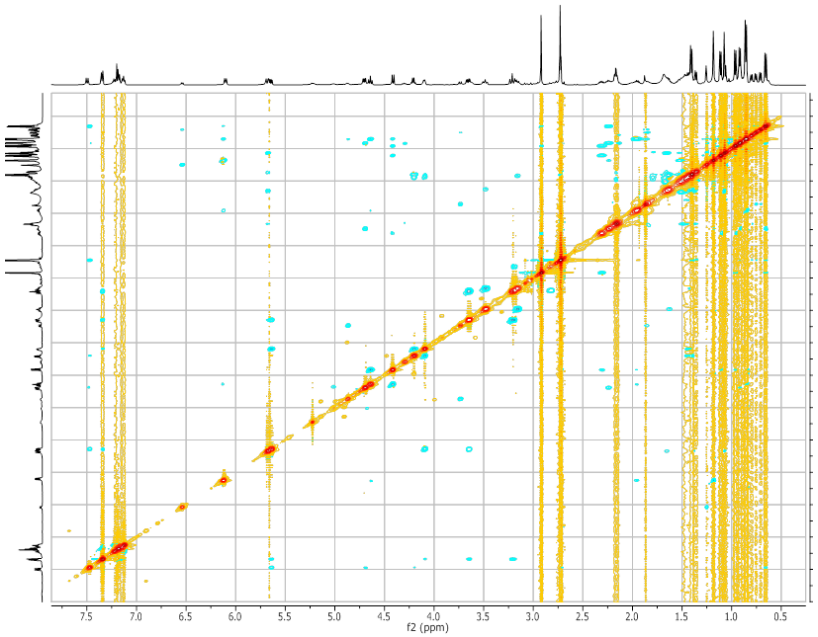


Figure A.13: *ROESY-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

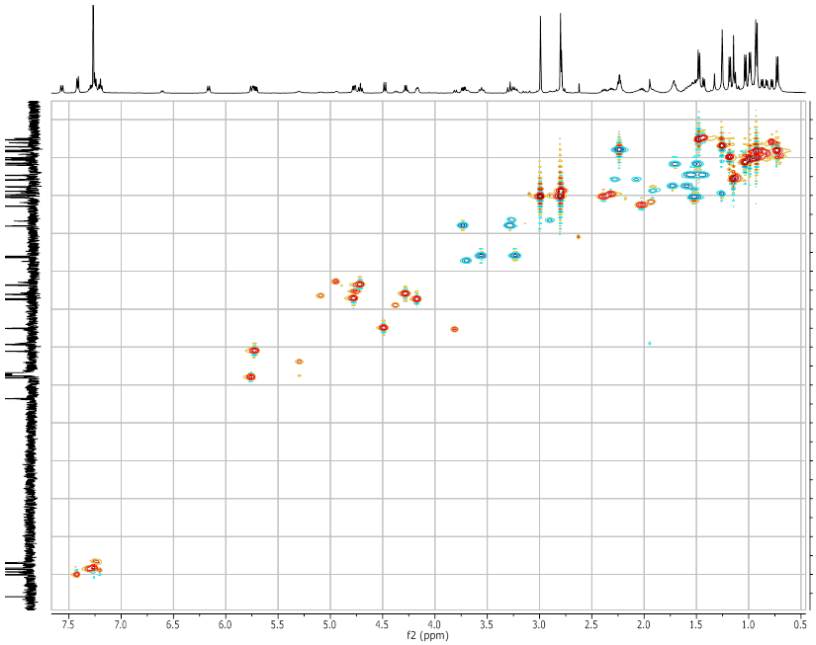


Figure A.14: *HSQC-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

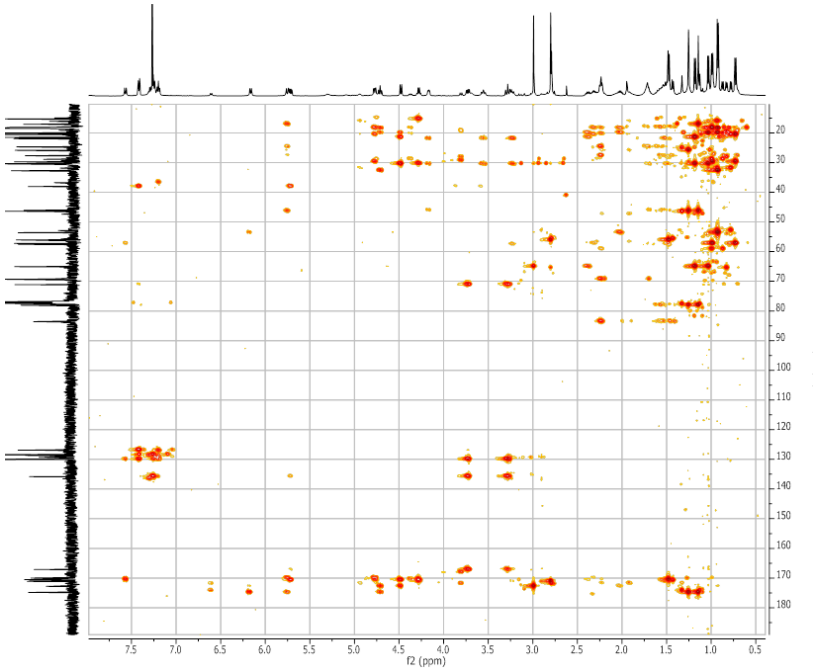


Figure A.15: *HMBC-NMR (500 MHz, $CDCl_3$) spectrum of viequeamide B (**26**).*

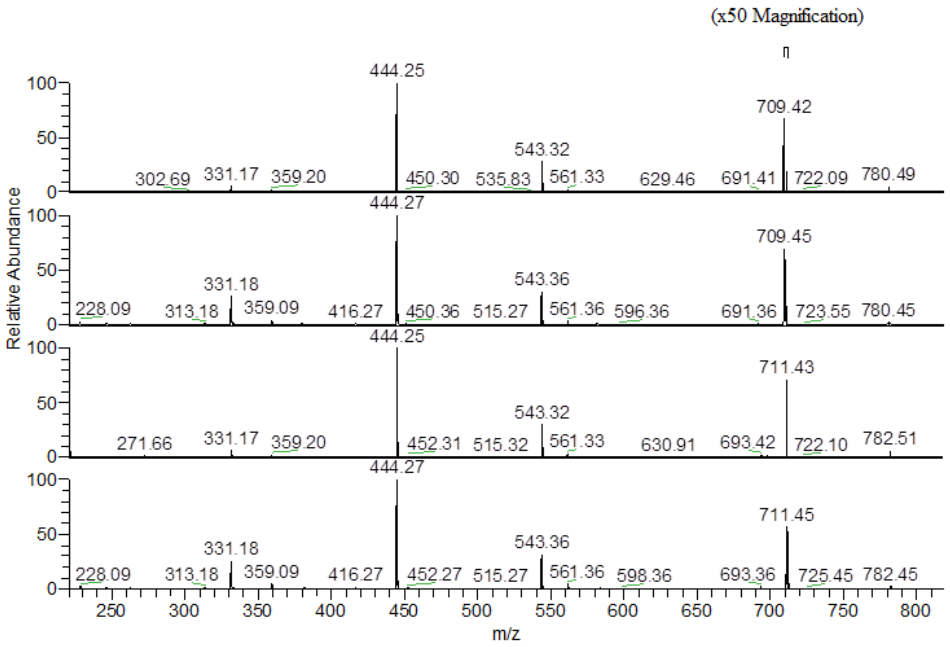


Figure A.16: *FT-MS/LTQ Ion-Trap MS*^{2} *spectra of viequeamides B (**26**) and C (**27**).* Spectra in descending order. *First Spectrum* Parent Mass 808, $[M+H]^+$ of **26**. *Second Spectrum* LTQ Ion-Trap MS Spectrum Parent Mass 808, $[M+H]^+$ of **26**. *Third Spectrum* Parent Mass 810, $[M+H]^+$ of **27**. *Second Spectrum* LTQ Ion-Trap MS Spectrum Parent Mass 810, $[M+H]^+$ of **27**.

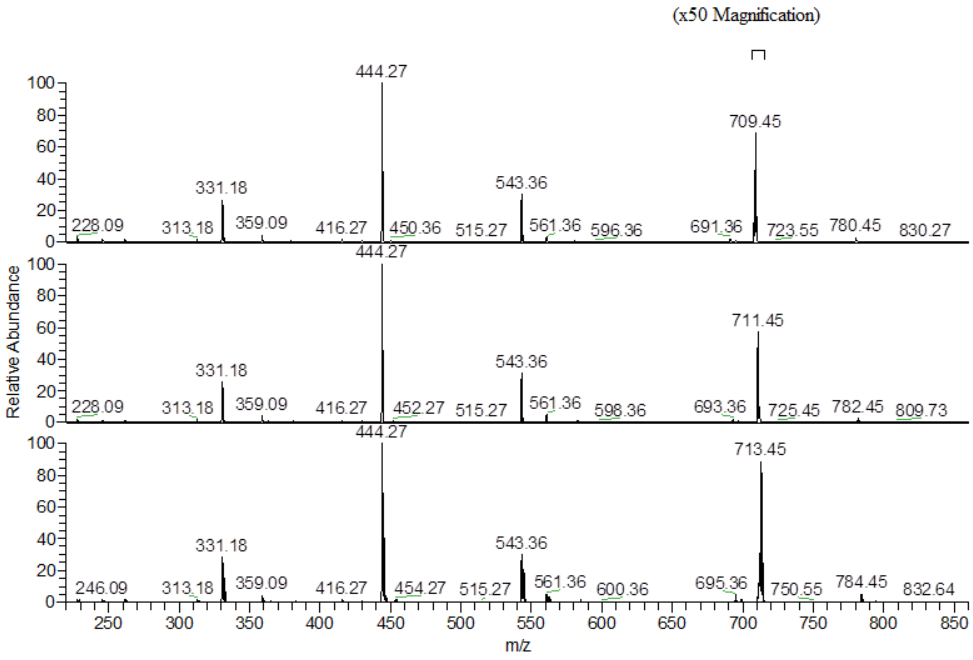


Figure A.17: *LTQ-MS Ion-Trap spectra of viequeamdies B-D ((**26-28**)).* Spectra in descending order. *First Spectrum* Parent Mass 808, $[M+H]^+$ of **26**. *Second Spectrum* Parent Mass 810, $[M+H]^+$ of **27**. *Third Spectrum* Parent Mass 812, $[M+H]^+$ of **28**.

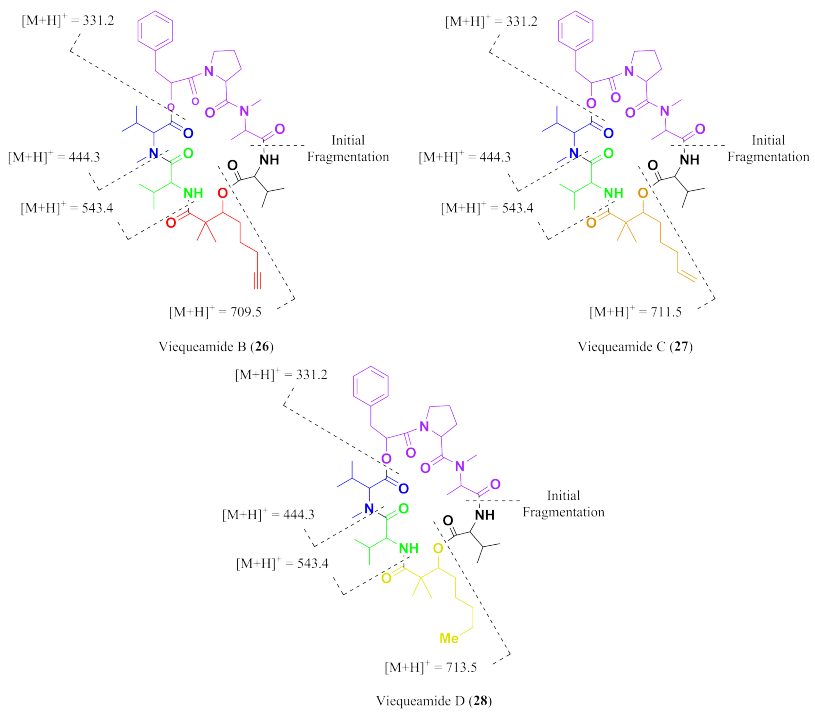


Figure A.18: *Comparison of LTQ Ion-Trap MS^2 fragments from viequeamides B-D (**26-28**).*

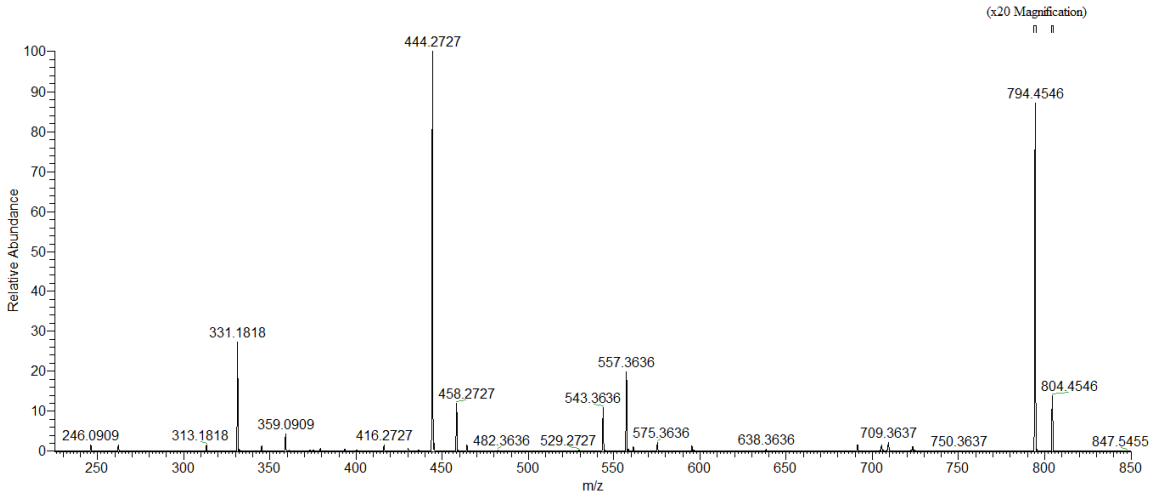


Figure A.19: *LTQ Ion-Trap MS^2 fragmentation spectrum of viequeamide E and F.* The parent mass of 822 was fragmented and showed two different series of fragment ions, for viequeamide E (**29**) the series was: *m/z* 331.18, 444.27, 557.36, and 723.45; and for viequeamide F (**30**) the series was: *m/z* 331.18, 444.27, 543.36, and 709.36. Indicating that this parent mass was in fact two isobaric compounds. Each series corresponds to the addition of CH_2 to the structure of Viequeamide B, one to the valine-1 residue ((**29**)), and the other to the valine-2 residue ((**30**)). Unfortunately the addition of CH_2 to valine cannot be distinguished by MS techniques between leucine, isoleucine, and *N*-methyl valine.

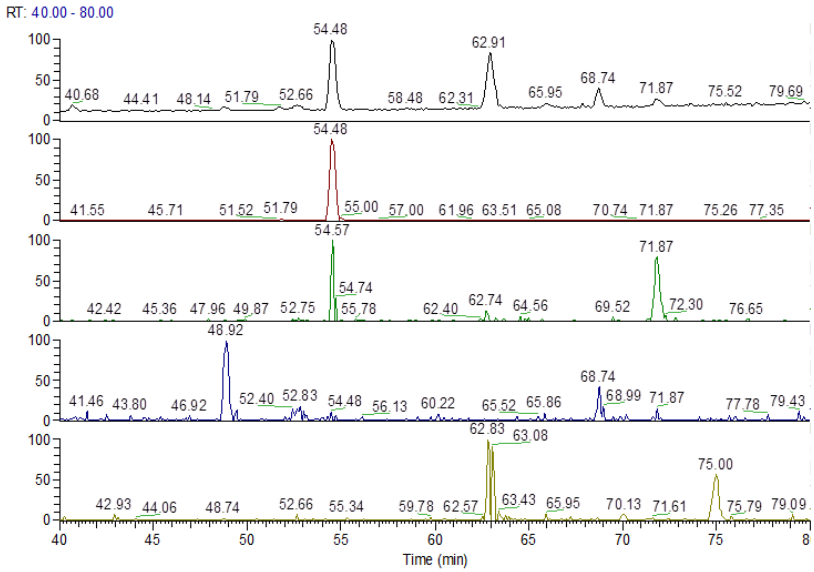


Figure A.20: *Marfey's L-FDAA derivatives of viequeamide A (**25**) hydrolysate via LCMS.*

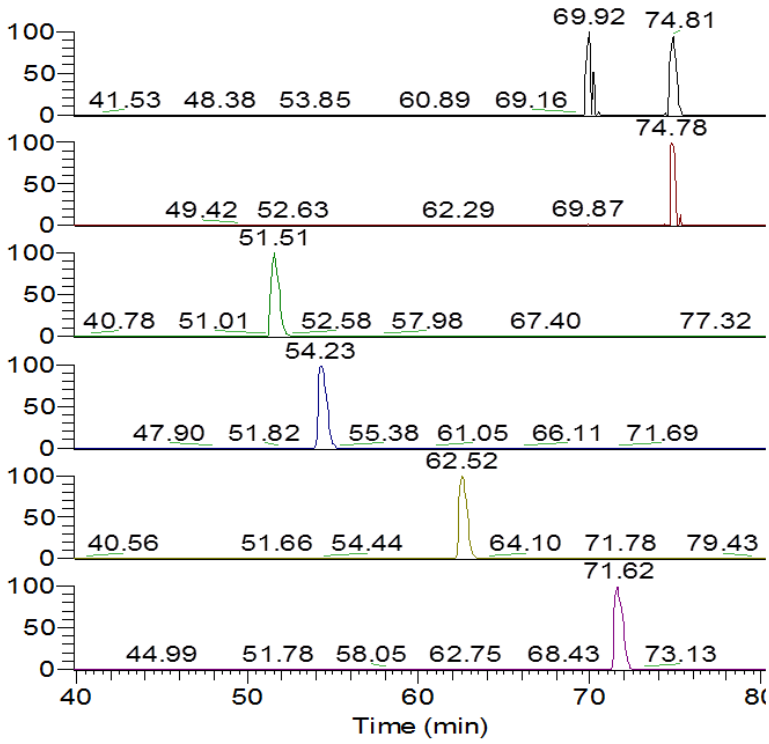


Figure A.21: *Marfey's L-FDAA derivatized amino acid standards (Val, Pro, and* N*-Me-Val).*

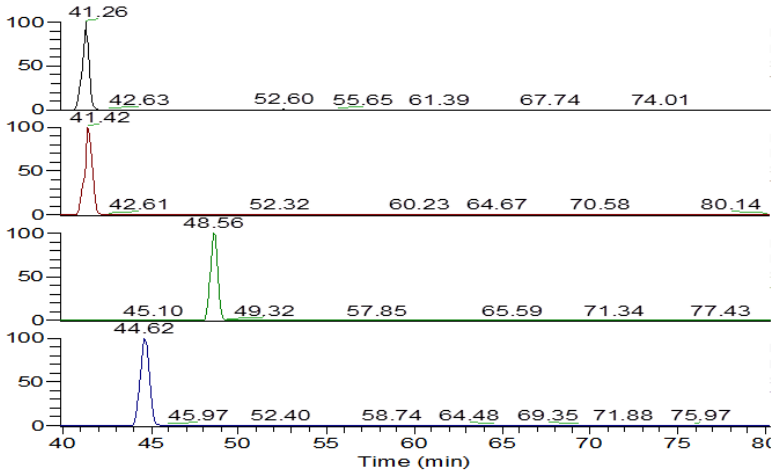


Figure A.22: *Marfey's L-FDAA derivatized amino acid standards (Thr).*

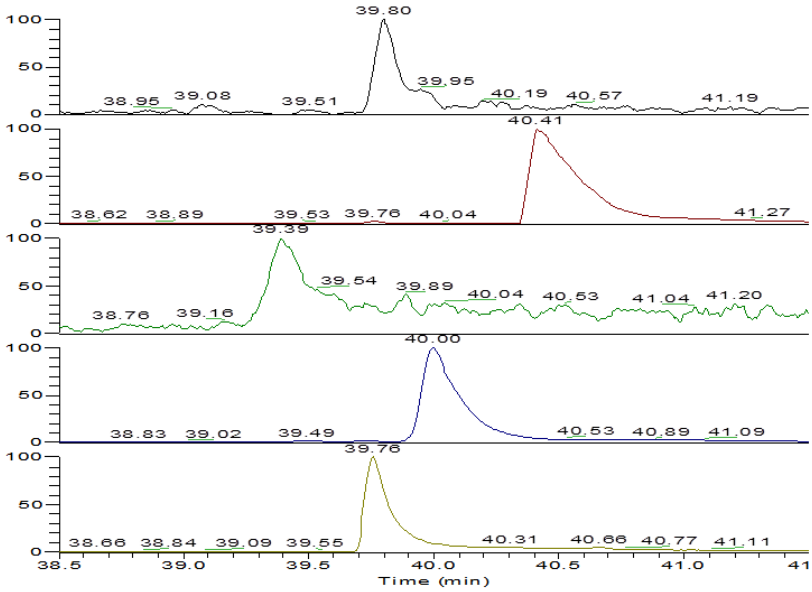


Figure A.23: *Stereoanalysis of viequeamide A (**25**) via GCMS (Hmpa residue).*

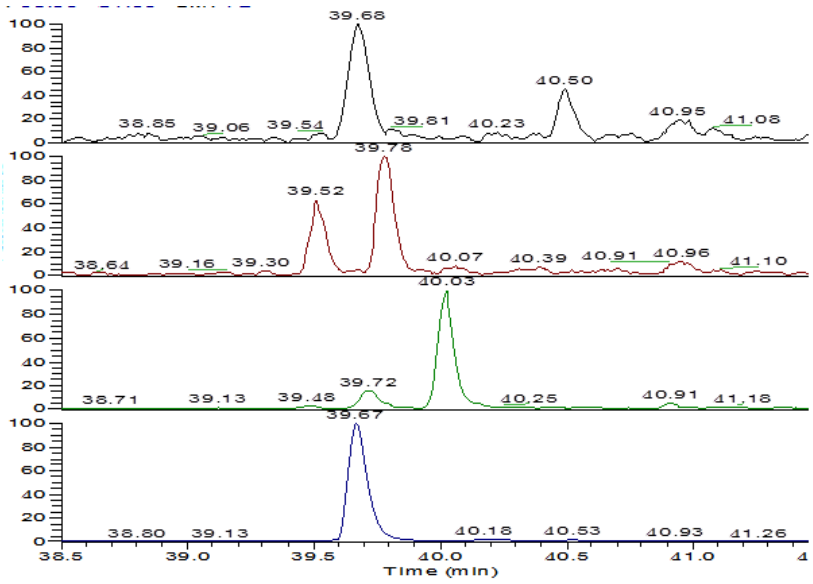


Figure A.24: *Stereoanalysis of viequeamide A (**25**) via GCMS (Hmpa standard coinjections).*

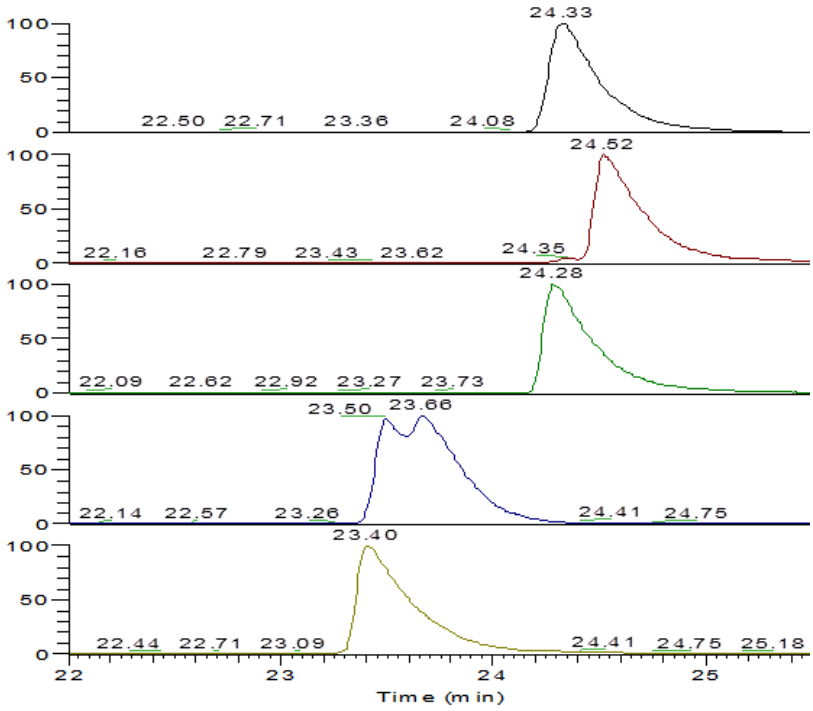


Figure A.25: *Stereoanalysis of viequeamide A (**25**) via GCMS (Dhoya residue).*

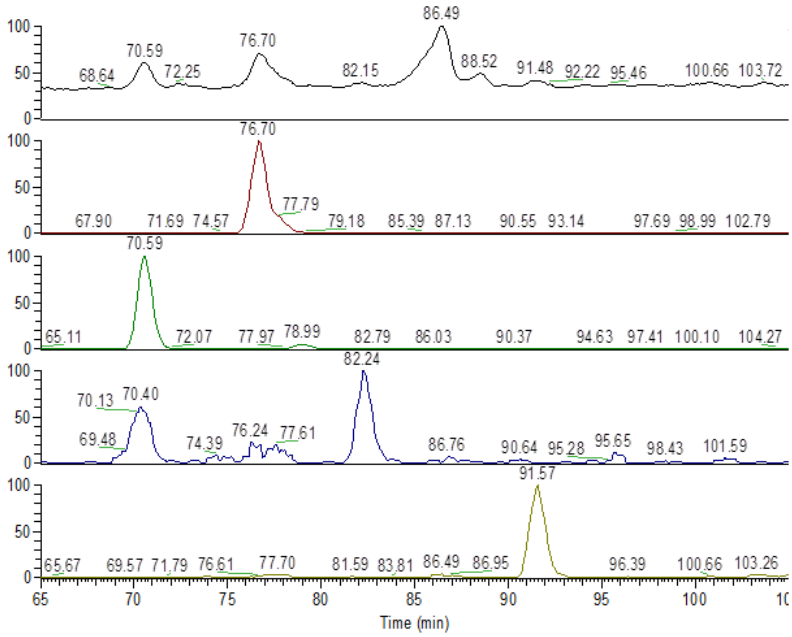


Figure A.26: *Marfey's L-FDVA derivatives of viequeamide B-F (**26-30**) hydrolysate via LCMS.*

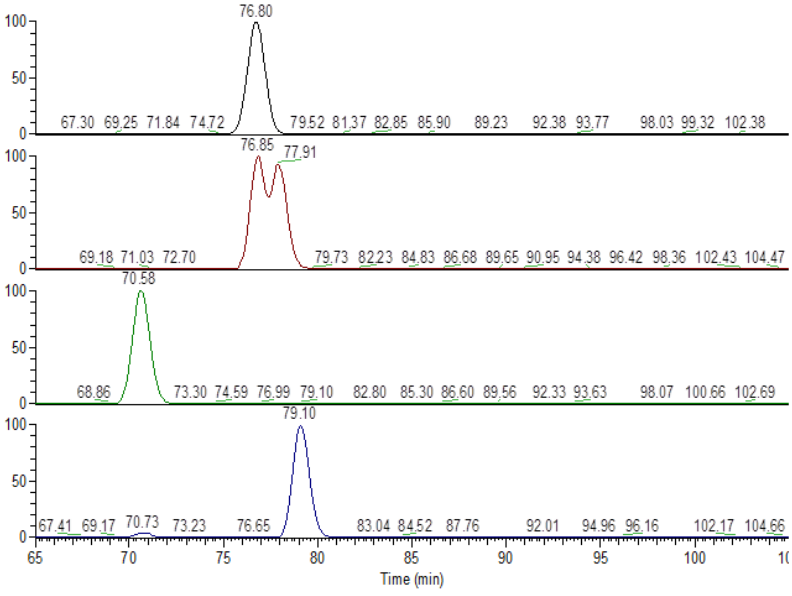


Figure A.27: *Marfey's L-FDVA derivatized amino acid standards (*N*-Me-Ala, Pro).*

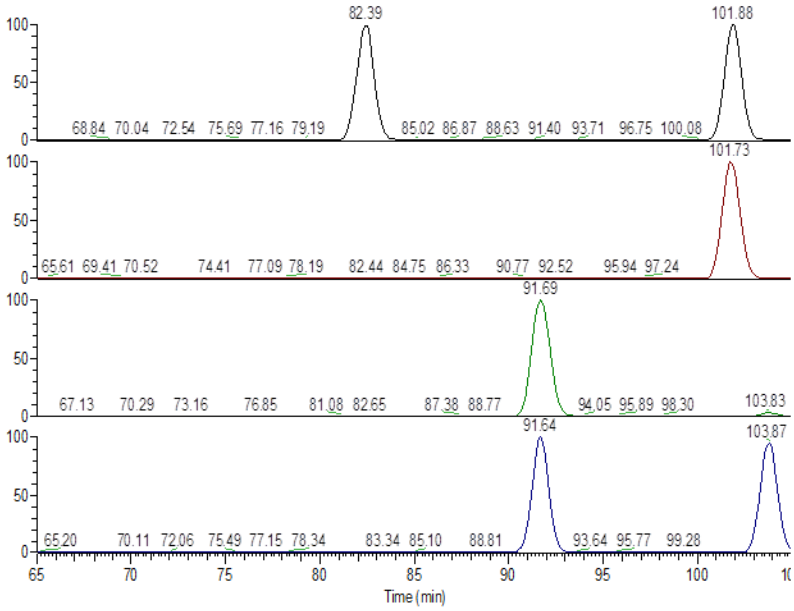


Figure A.28: *Marfey's L-FDVA derivatized amino acid standards (Val,* N-*Me-Val).*

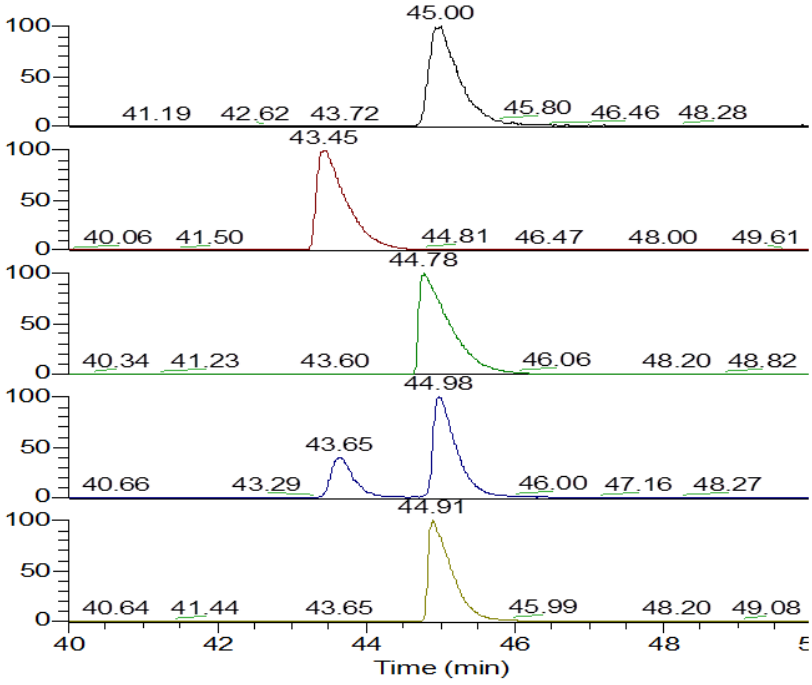


Figure A.29: *Stereoanalysis of viequeamide B (**26**) via GCMS (Pla residue).*

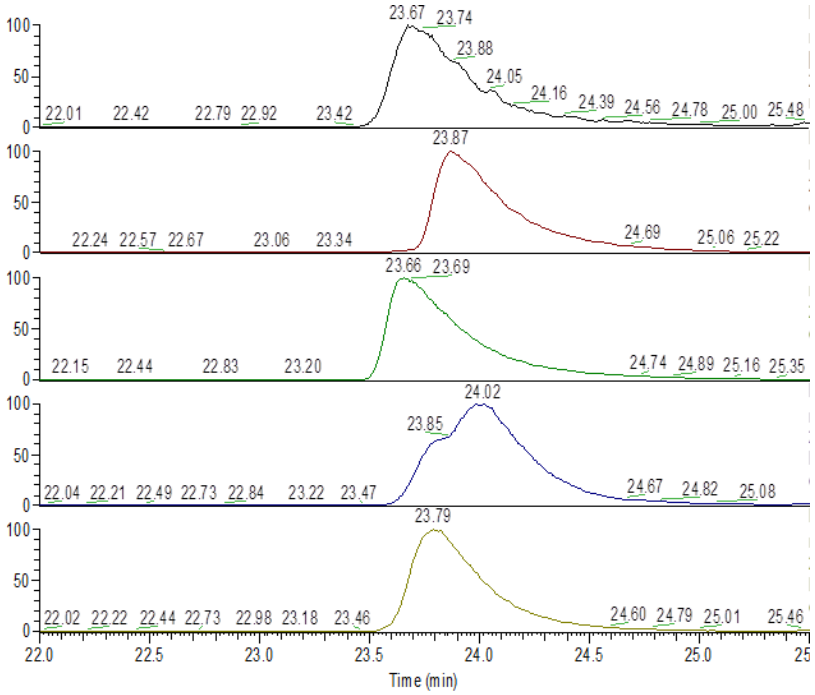
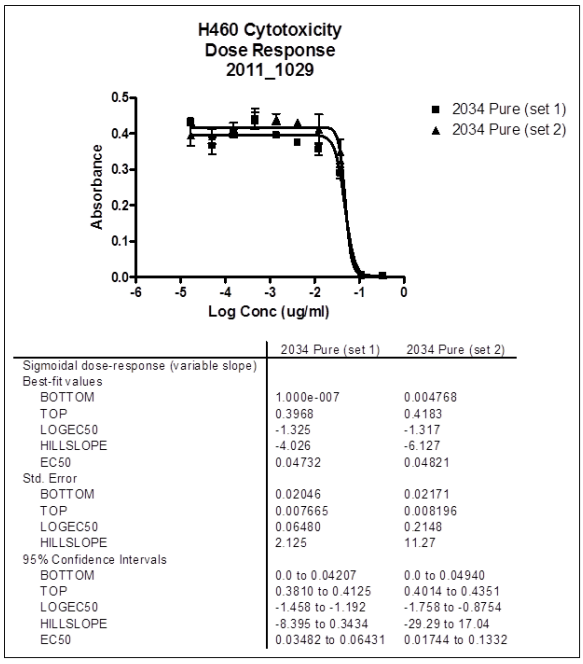


Figure A.30: *Stereoanalysis of viequeamide B (**26**) via GCMS (Dhoya residue).*



	2034 Pure (set 1)	2034 Pure (set 2)
Sigmoidal dose-response (variable slope)		
Best-fit values		
BOTTOM	1.000e-007	0.004768
TOP	0.3968	0.4183
LOGEC50	-1.325	-1.317
HILLSLOPE	-4.026	-6.127
EC50	0.04732	0.04821
Std. Error		
BOTTOM	0.02046	0.02171
TOP	0.007665	0.008196
LOGEC50	0.06480	0.2148
HILLSLOPE	2.125	11.27
95% Confidence Intervals		
BOTTOM	0.0 to 0.04207	0.0 to 0.04940
TOP	0.3810 to 0.4125	0.4014 to 0.4351
LOGEC50	-1.458 to -1.192	-1.758 to -0.8754
HILLSLOPE	-8.395 to 0.3434	-29.29 to 17.04
EC50	0.03482 to 0.06431	0.01744 to 0.1332

Figure A.31: *H460 bioassay dose response curve for viequeamide A (**25**).*

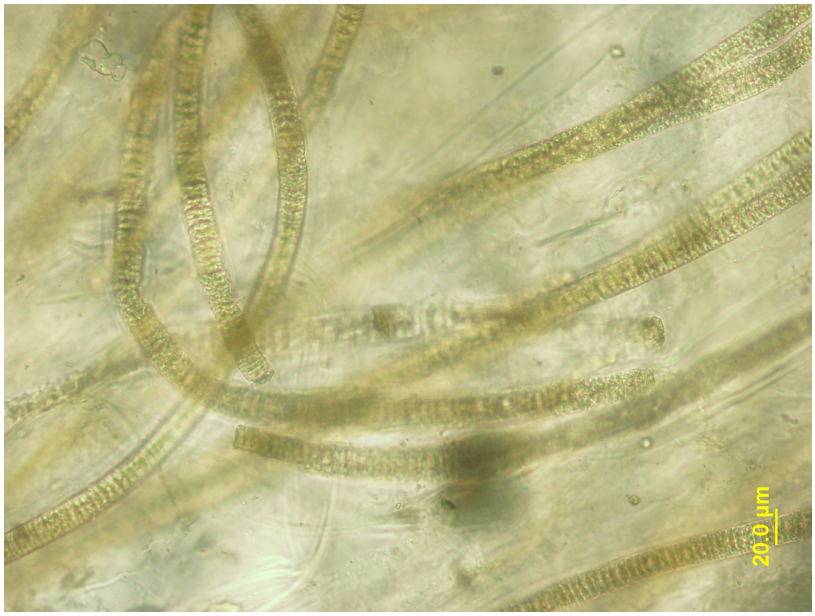


Figure A.32: *Morphological description of viequeamides producer (VQC-26/MAR/11-1).* Morphological characterization was performed using an epifluorescent microscope (1000) and descriptions were made in accordance with traditional phycological systems.[Komáek, 2005] Macroscopically, VQC-26/MAR/11-1 grew as a hemispherical colony on the coral rubble or other hard substrate with a crust-like thallus type. On a cellular level, VQC-26/MAR/11-1 had a distinct sheath with a single trichome per sheath. The cells were discoid in shape, 15 μm wide and 2.5 μm long. There were no constrictions at the crosswalls. The apical cells were discoid as well with no calyptras present. Filaments were not attenuated at the ends. Morphological comparison and putative taxonomic identification of the cyanobacterial specimen were performed in accordance with modern classification systems.

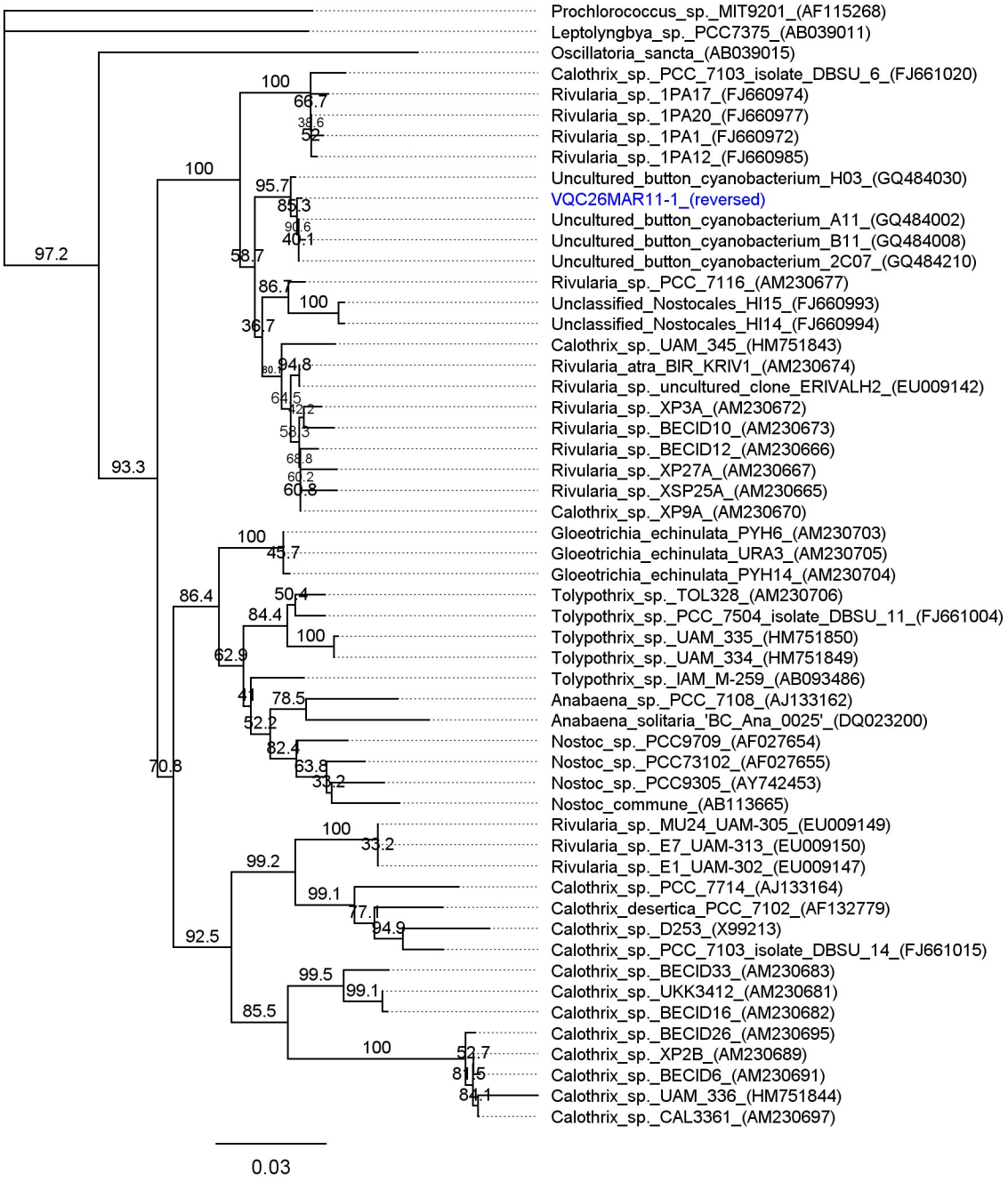


Figure A.33: *Phylogenetic tree (Maximum Likelihood) of the viequeamides producer (VQC-26/MAR/11-1).*

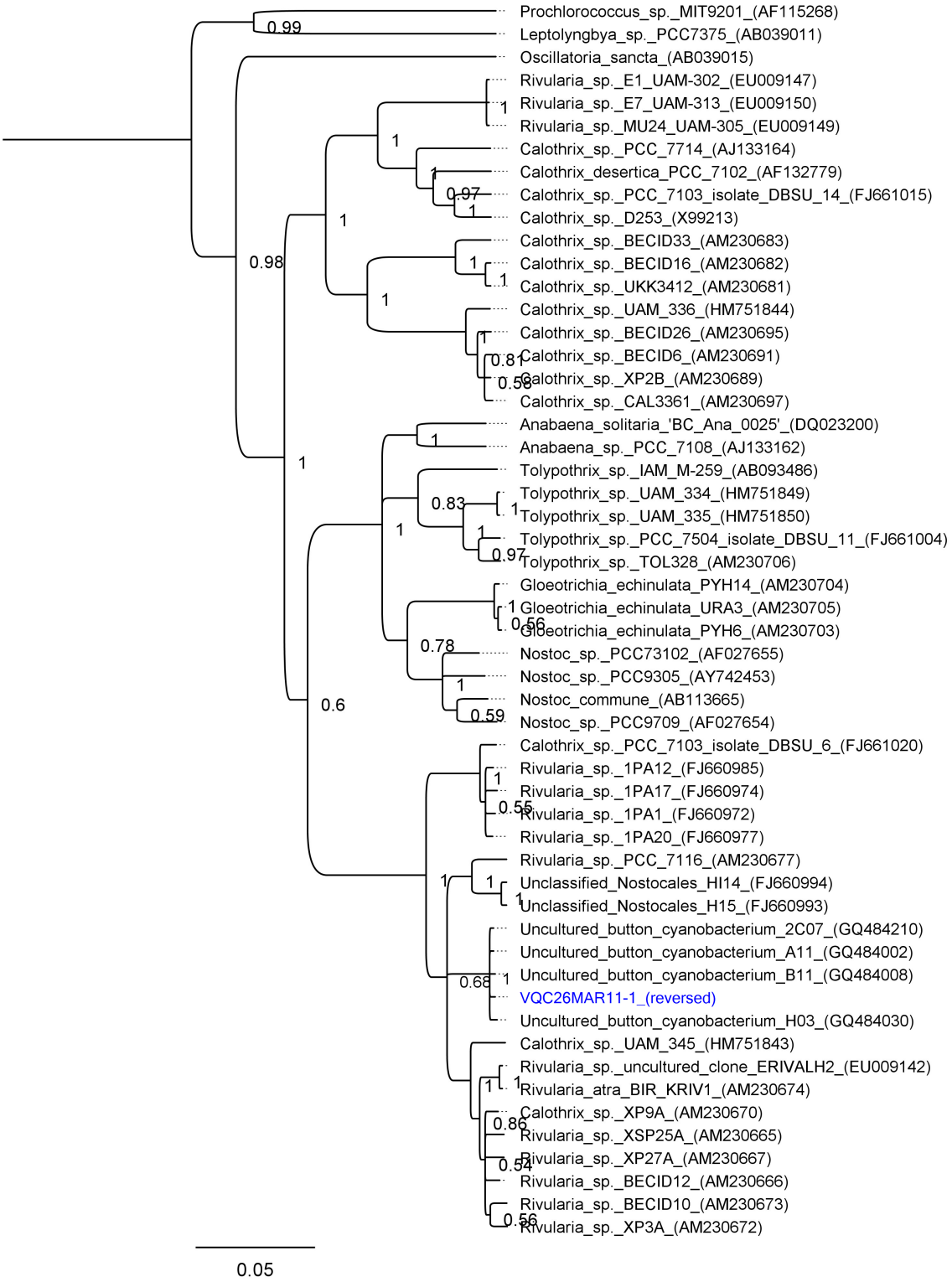


Figure A.34: *Phylogenetic tree (MrBayes) of the viequeamides producer (VQC-26/MAR/11-1).*

Appendix B

Supplementary Information for Chapter 4

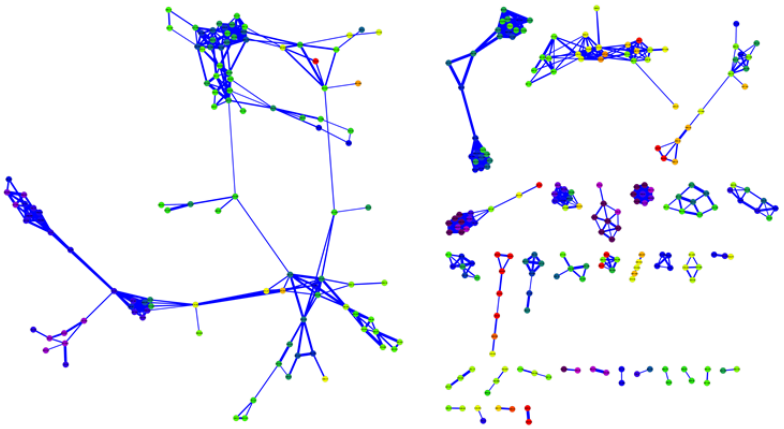


Figure B.1: *Full JHB molecular network from LTQ-FT data.*

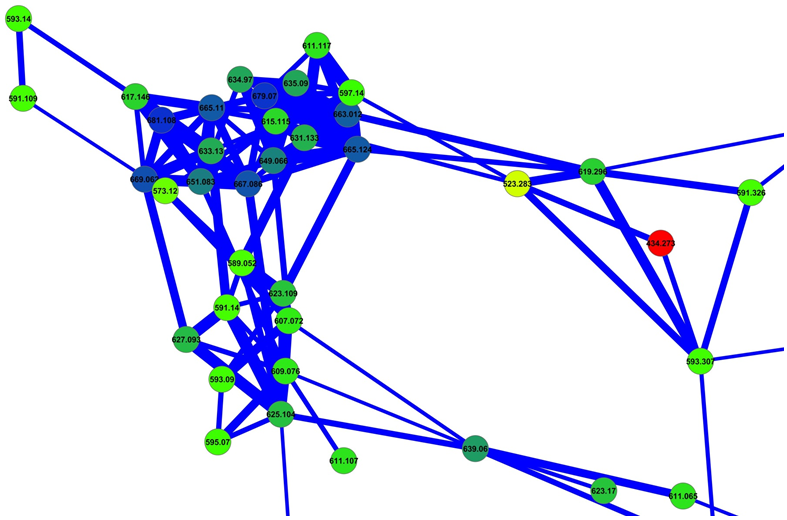


Figure B.2: *Hectochlorins cluster within the JHB molecular network from LTQ-FT data.*

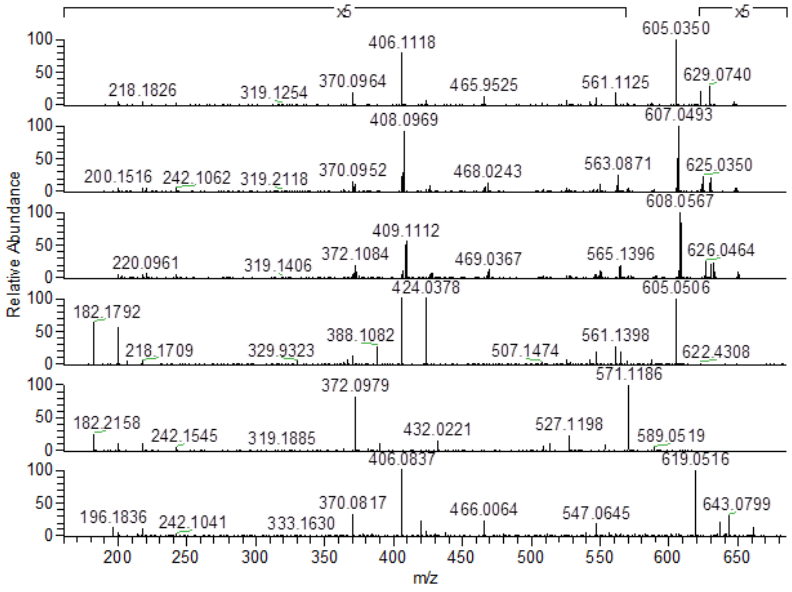


Figure B.3: *IT-MS*^{2} *fragment spectra for the hectochlorins from LTQ-FT data.* In descending order: *First Spectrum* Hectochlorin (**89**) $[M + H]^+$. *Second Spectrum* Hectochlorin (**89**) $[M^{37}Cl + H]^+$. Parent Mass 667 *m/z*. *Third Spectrum* Hectochlorin (**89**) $[M^{37}Cl_2 + H]^+$. Parent Mass 669 *m/z*. *Fourth Spectrum* Hectochlorin B (**97**) $[M + H]^+$, Parent Mass 623 m/z. *Fifth Spectrum* Hectochlorin C (**98**) $[M + H]^+$, Parent Mass 631 *m/z*. *Sixth Spectrum* Hectochlorin D (**99**) $[M + H]^+$, Parent Mass 679 *m/z*.

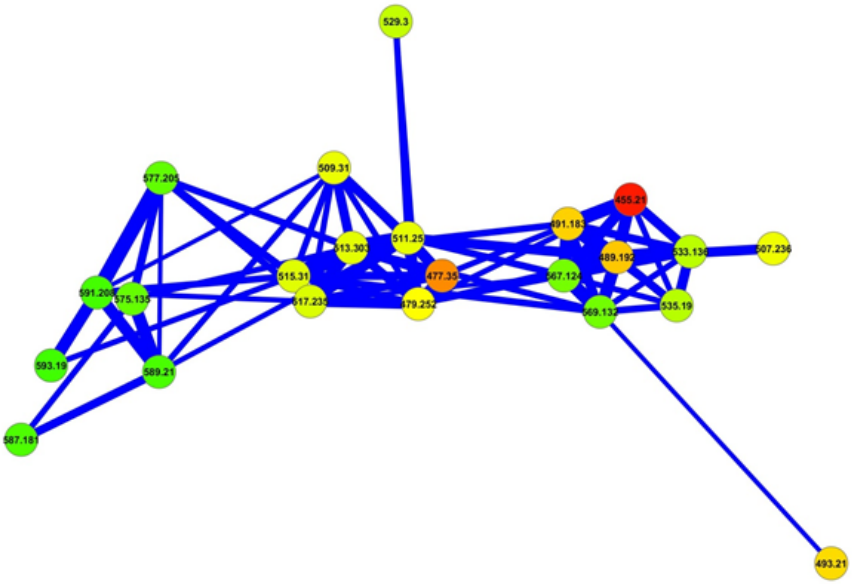


Figure B.4: *Jamaicamides cluster within the JHB molecular network from LTQ-FT data.*

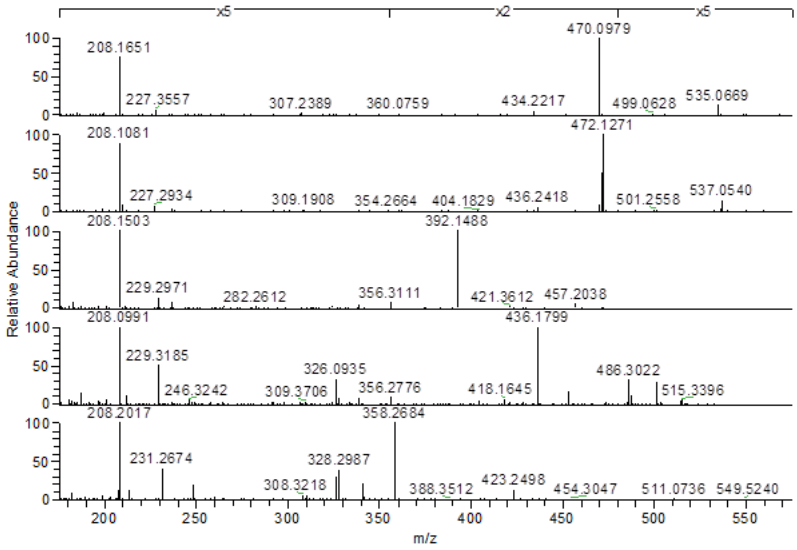


Figure B.5: *IT-MS*^{2} *fragment spectra for the jamaicamides from LTQ-FT data.* Fragment Spectra in descending order: *First* Jamaicamide A (**90**) $[M + H]^+$, Parent Mass 567 *m/z*. *Second* Jamaicamide A (**90**) $[M + H]^+$, Parent Mass 569 *m/z*. *Third* Jamaicamide B (**91**) $[M + H]^+$, Parent Mass 489 *m/z*. *Fourth* Jamaicamide D (**100**) $[M + H]^+$, Parent Mass 533 *m/z*. *Fifth* Jamaicamide E (**101**) $[M + H]^+$, Parent Mass 455 *m/z*

Table B.1: *Masses of observed for the hectochlorins and jamaicamides from LTQ-FT data.*

Compound	Calculated Formula	LCMS Observed HRMS[1] Mass *m/z*	Retention Time (min)	Isolated Compound HRMS[1] Mass *m/z*
Hectochlorin (**89**)	$C_{27}H_{35}Cl_2N_2O_9S_2$ [M + H]$^+$, 665.1156	665.1182, 3.9 ppm, 2.6 mamau	23.8	665.1154, -0.3 ppm, -0.2 mamu
	$C_{27}H_{35}Cl^{37}ClN_2O_9S_2$ [M + H]$^+$, 667.1126	667.1153, 4.0 ppm, 2.7 mamau		665.1126, 0 ppm, 0 mamu
	$C_{27}H_{34}Cl_2N_2NaO_9S_2$ [M + Na]$^+$, 687.0975	687.1070, 13.8 ppm, 9.5 mamau		687.0976, 0.1 ppm, 0.1 mamu
	$C_{27}H_{34}Cl^{37}ClN_2NaO_9S_2$ [M + Na]$^+$, 689.0945	689.1057, 16.2 ppm, 11.2 mamau		689.0948, 0.4 ppm, 0.3 mamu
Hectochlorin B (**97**)	$C_{25}H_{33}Cl_2N_2O_8S_2$ [M + H]$^+$, 623.1050	623.1041, -1.4 ppm, -0.9 mamau	20.3	623.1050, 0 ppm, 0 mamu
	$C_{25}H_{33}Cl^{37}ClN_2O_8S_2$ [M + H]$^+$, 625.1020	625.1019, -0.2 ppm, -0.1 mamau		625.1021, 0.2 ppm, 0.1 mamu
	$C_{25}H_{32}Cl_2N_2NaO_8S_2$ [M + Na]$^+$, 645.0869	645.0884, 2.3 ppm, 1.5 mamau		645.0870, 0.2 ppm, 0.1 mamu
	$C_{25}H_{32}Cl^{37}ClN_2NaO_8S_2$ [M + Na]$^+$, 647.0840	647.0840, 2.5 ppm, 1.6 mamau		647.0841, 0.2 ppm, 0.1 mamu
Hectochlorin C (**98**)	$C_{27}H_{36}ClN_2O_9S_2$ [M + H]$^+$, 631.1545	631.1516, -4.6 ppm, -2.9 mamau	22.6	631.1545, 0 ppm, 0 mamu
	$C_{27}H_{36}^{37}ClN_2O_9S_2$ [M + H]$^+$, 633.1516	633.1492, -3.8 ppm, -2.4 mamau		633.1519, 0.5 ppm, 0.3 mamu
	$C_{27}H_{35}ClN_2NaO_9S_2$ [M + Na]$^+$, 653.1356	653.1333, -4.9 ppm, -3.2 mamau		653.1360, -0.8 ppm, -0.5 mamu
	$C_{27}H_{35}^{37}ClN_2NaO_9S_2$ [M + Na]$^+$, 655.1335	655.1312, -3.5 ppm, -2.3 mamau		655.1337, 0.3 ppm, 0.2 mamu
Hectochlorin D (**99**)	$C_{28}H_{37}Cl_2N_2O_9S_2$ [M + H]$^+$, 679.1312	679.1273, -5.7 ppm, -3.9 mamau	25.8	679.1306, -0.9 ppm, -0.6 mamu
	$C_{28}H_{37}Cl^{37}ClN_2O_9S_2$ [M + H]$^+$, 681.1283	681.1245, -5.8 ppm, -3.8 mamau		681.1274, -1.3 ppm, -0.9 mamu
	$C_{28}H_{36}Cl_2N_2NaO_9S_2$ [M + Na]$^+$, 701.1131	701.1091, -5.7 ppm, -4.0 mamau		701.1129, -0.3 ppm, -0.2 mamu
	$C_{27}H_{36}Cl^{37}ClN_2NaO_9S_2$ [M + Na]$^+$, 703.1102	703.1065, -5.3 ppm, -3.7 mamau		703.1097, -0.7 ppm, -0.5 mamu
Jamaicamide A (**90**)	$C_{27}H_{37}BrClN_2O_4$ [M + H]$^+$, 567.1620	567.1631, 1.9 ppm, 1.1 mamau	30.5	567.1621, 0.2 ppm, 0.1 mamu
	$C_{27}H_{37}^{81}BrClN_2O_4$, 569.1599	569.1601, 0.3 ppm, 0.2 mamau		569.1597, -0.4 ppm, -0.2 mamu
	$C_{27}H_{36}BrClN_2NaO_4$ [M + Na]$^+$, 589.1439	589.1478, 6.6 ppm, 3.9 mamau		589.1442, 0.5 ppm, 0.3 mamu
	$C_{27}H_{36}^{81}BrClN_2NaO_4$ [M + Na]$^+$, 591.1419	591.1450, 5.2 ppm, 3.1 mamau		591.1418, -0.2 ppm, -0.1 mamu
Jamaicamide B (**91**)	$C_{27}H_{38}ClN_2O_4$ [M + H]$^+$, 489.2515	489.2515, 0 ppm, 0 mamau	27.2	489.2513, -0.4 ppm, -0.2 mamu
	$C_{27}H_{38}^{37}ClN_2O_4$, 491.2485	491.2510, 5.1 ppm, 2.5 mamau		491.2487, 0.4 ppm, 0.2 mamu
	$C_{27}H_{37}ClN_2NaO_4$ [M + Na]$^+$, 511.2334	511.2358, 4.7 ppm, 2.4 mamau		511.2333, -0.2 ppm, -0.1 mamu
	$C_{27}H_{37}^{37}ClN_2NaO_4$ [M + Na]$^+$, 513.2305	513.2336, 6.0 ppm, 3.1 mamau		513.2306, 0.2 ppm, 0.1 mamu
Jamaicamide D (**100**)	$C_{27}H_{38}BrN_2O_4$ [M + H]$^+$, 533.2009	533.1983, -4.9 ppm, -2.6 mamau	29.9	533.2012, 0.6 ppm, 0.3 mamu
	$C_{27}H_{38}^{81}BrN_2O_4$, 535.1989	535.1962, -5.0 ppm, -2.7 mamau		535.1991, 0.4 ppm, 0.2 mamu
	$C_{27}H_{37}BrN_2NaO_4$ [M + Na]$^+$, 555.1829	555.1799, -5.4 ppm, -3.0 mamau		555.1833, 0.7 ppm, 0.4 mamu
	$C_{27}H_{37}^{81}BrN_2NaO_4$ [M + Na]$^+$, 557.1808	557.1778, -5.4 ppm, 3.0 mamau		557.1812, 0.7 ppm, 0.4 mamu
Jamaicamide E (**101**)	$C_{27}H_{39}N_2O_4$ [M + H]$^+$, 455.2904	455.2877, -5.9 ppm, -2.7 mamau	26.5	Not isolated by HPLC
	$C_{27}H_{38}N_2NaO_4$ [M + Na]$^+$, 477.2724	477.2693, -6.5 ppm, -3.1 mamau		
Jamaicamide F (**104**)	$C_{27}H_{37}ClIN_2O_4$ [M + H]$^+$, 615.1481	615.1549, 11.1 ppm, 6.8 mamau	29.4	615.1471, -1.6 ppm, -1.0 mamu
	$C_{27}H_{37}^{37}ClIN_2O_4$, 617.1452	617.1516, 10.4 ppm, 6.4 mamau		617.1441, -1.8 ppm, -1.1 mamu
	$C_{27}H_{36}ClIN_2NaO_4$ [M + Na]$^+$, 637.1300	637.1363, 9.9 ppm, 6.3 mamau		637.1289, -1.7 ppm, -1.1 mamu
	$C_{27}H_{36}^{37}ClIN_2NaO_4$ [M + Na]$^+$, 639.1271	639.1333, 9.7 ppm, 6.2 mamau		639.1259, -1.9 ppm, -1.2 mamu

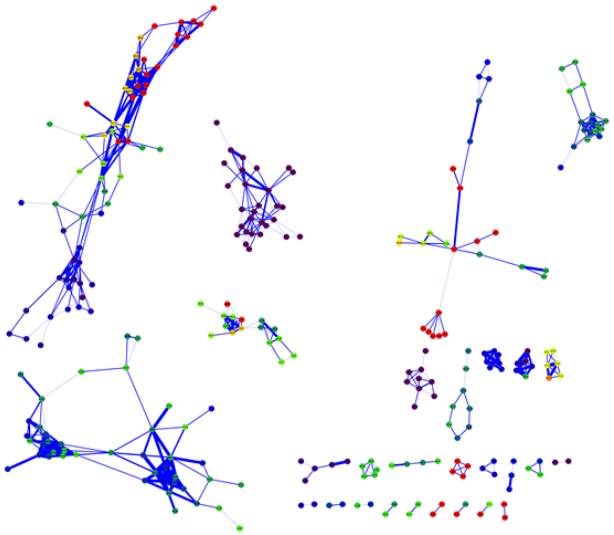


Figure B.6: *Full JHB molecular network, iodide rich media, from LCQ data.*

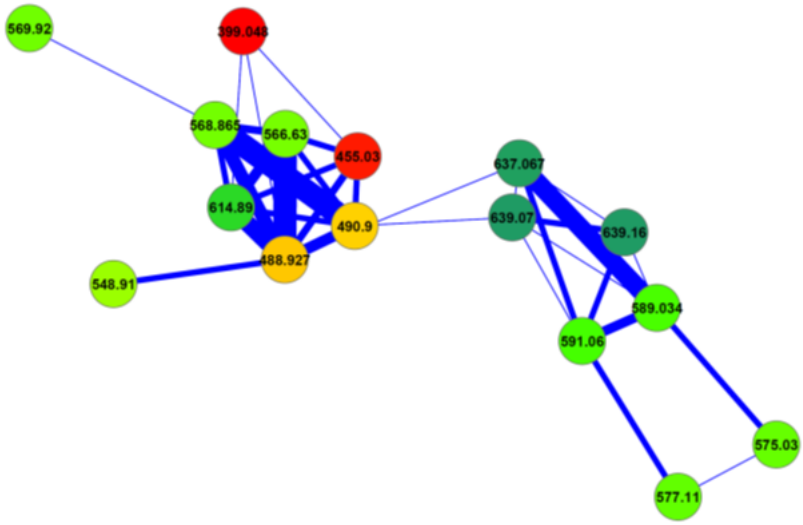


Figure B.7: *Jamaicamides cluster within the JHB molecular network, iodide rich media, from LCQ data.*

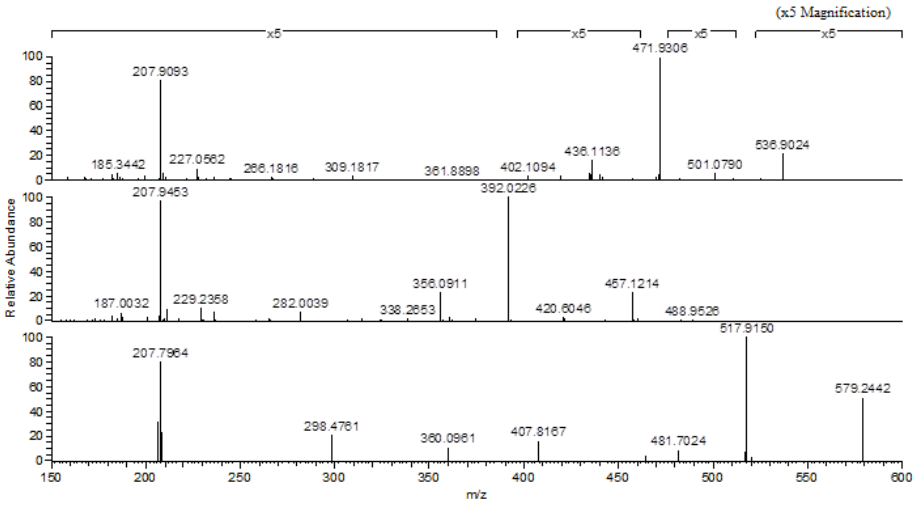


Figure B.8: *IT-MS*^{2} *fragment spectra for the jamaicamides, iodide rich media, from LCQ data.* Fragment Spectra in descending order: *Top Spectrum* Jamaicamide A (**90**) $[M^{81}Br + H]^+$, Parent Mass 569 *m/z*. *Middle Spectrum* Jamaicamide B (**91**), $[M + H]^+$, Parent Mass 489 *m/z*. *Bottom Spectrum* Jamaicamide F (**104**) $[M + H]^+$, Parent Mass 615 *m/z*.

Table B.2: *NMR data summary for hectochlorin (**89**).* All experiments in $CDCl_3$ with TMS standard, 1% *v/v* for ^{13}C experiment, and 0.03% *v/v* for the rest. [a]Specturm collected on a Varian VX 500 MHz with ^{13}C-optimized cryoprobe. [b]Spectra collected on a Bruker 600 MHz (600 MHz and 150 MHz for the ^{1}H and ^{13}C nuclei respectively) with 1.7 mm inverse cryo-probe. [c]Signal by projection from the HMBC experiment.

Position	δ_C, type[a]	δ_H (J in Hz)[b]	HMBC[b]	TOCSY[b]
1	172.9, C	–		
2	42.6, CH	3.16, quintet (7.4)	1, 3, 9	3-6, 9
3	75.2, CH	5.35, m		2, 4-6,9
4a	30.9, CH_2	1.71, m		2, 3, 4b, 5, 6, 9
4b	30.9, CH_2	1.86, m		2, 3, 4a, 5, 6, 9
5	20.9, CH_2	1.71, m		2, 4-6, 9
6a	49.3, CH_2	2.26, m		2-5, 6b, 9
6b	49.3, CH_2	2.15, m		2-5, 6a, 9
7	90.4,C	–		
8	37.2, CH_3	2.11, s	6, 7	
9	15.1, CH_3	1.29, d (7.4)	1-3, self	2-6
10	161.1, C	–		
11	147.0, C	–		
12	128.6, CH	8.17, s	11, 13	14
13	166.3, C[c]	–		
14	74.7, CH	6.84, s	13, 15-17, 26	12
15	82.0, C	–		
16	24.4, CH_3	1.84, s	14, 15, 17, self	
17	21.9, CH_3	1.61, s	14-16, self	
18	160.4, C	–		
19	147.5, C	–		
20	127.7, CH	7.94, s	19, 21	
21	165.1, C	–		
22	77.8, CH	5.67, s	1, 21, 23, 25	
23	71.6, C	–		
24	26.7, CH_3	1.31, s	22, 23, 25	
25	26.0, CH_3	1.37, s	22, 24	
26	168.7, C	–		
27	20.9, CH_3	2.19, s	26, self	

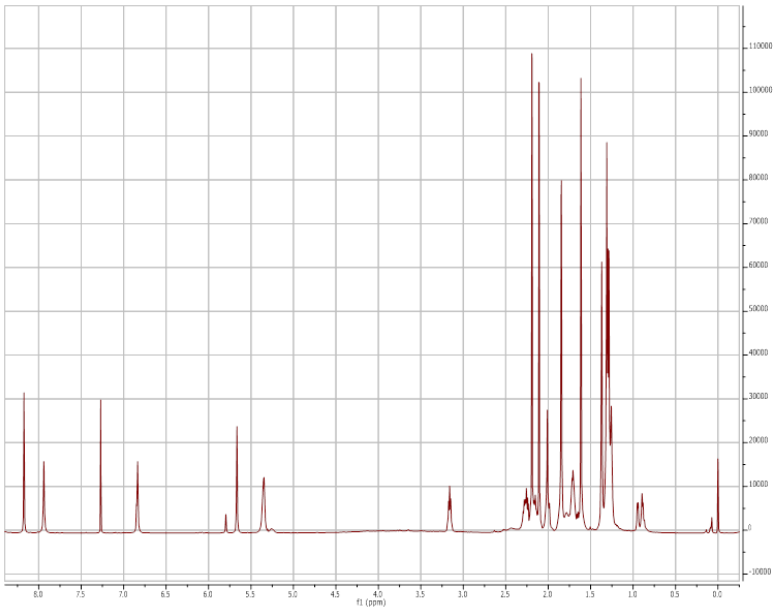


Figure B.9: *^{1}H-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin (**89**).*

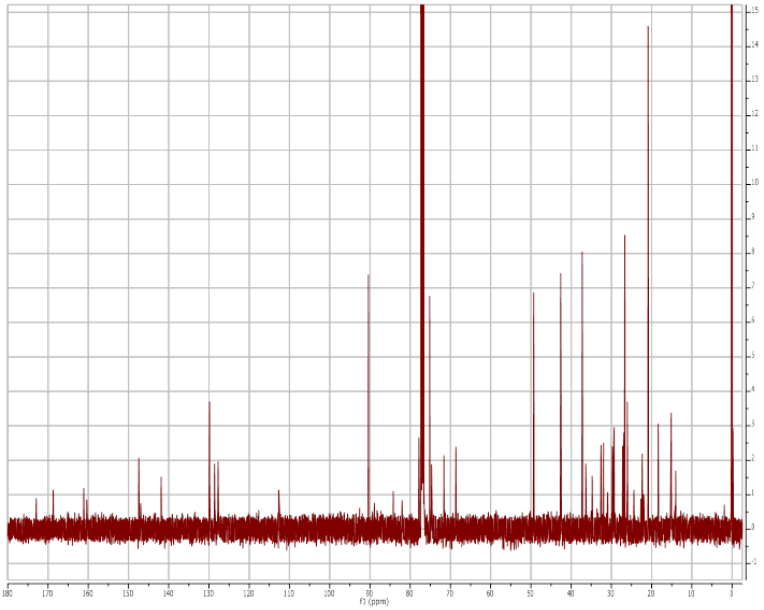


Figure B.10: *^{13}C-NMR (125 MHz, $CDCl_3$) spectrum of hectochlorin (**89**).* Benzene contamination at 129.8 ppm.

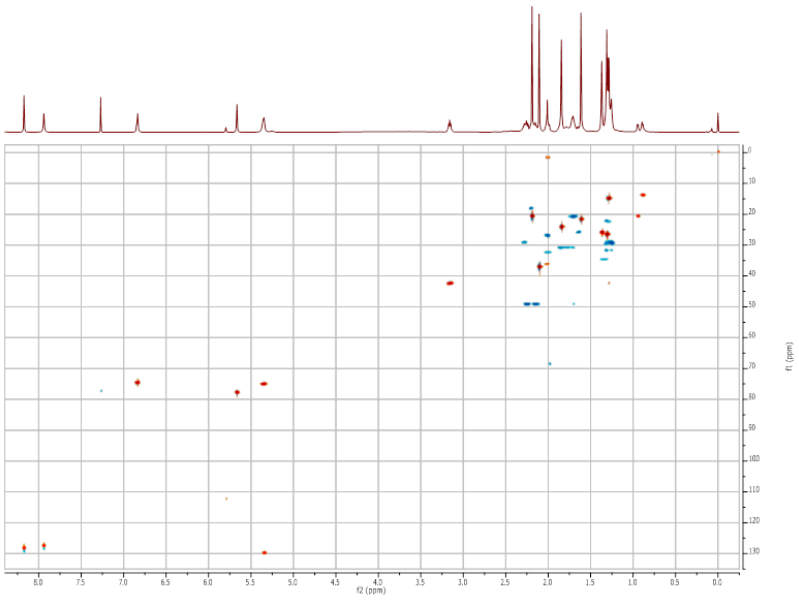


Figure B.11: *HSQC-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin (**89**).*

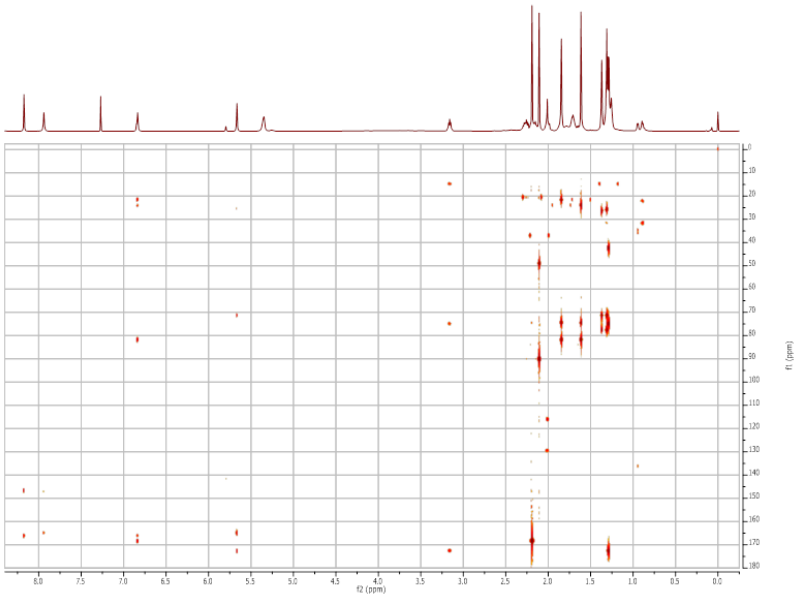


Figure B.12: *HMBC-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin (**89**).*

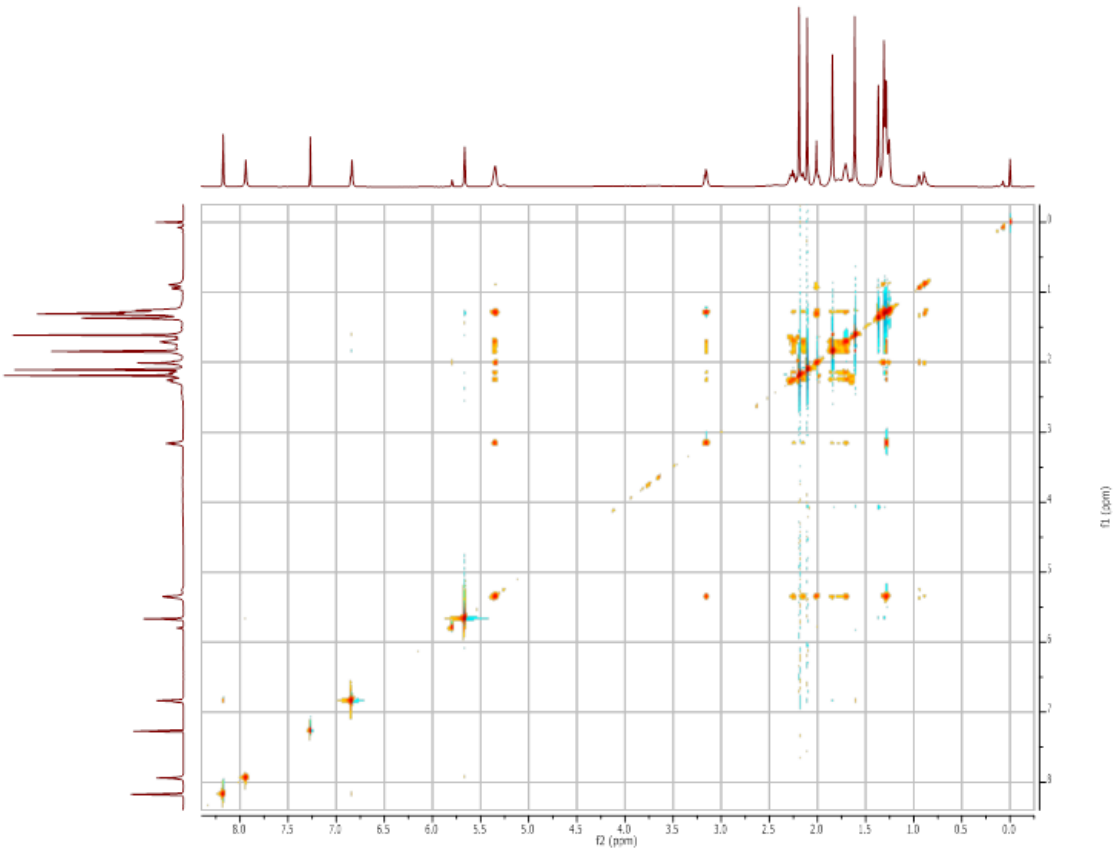


Figure B.13: *TOCSY-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin (**89**).*

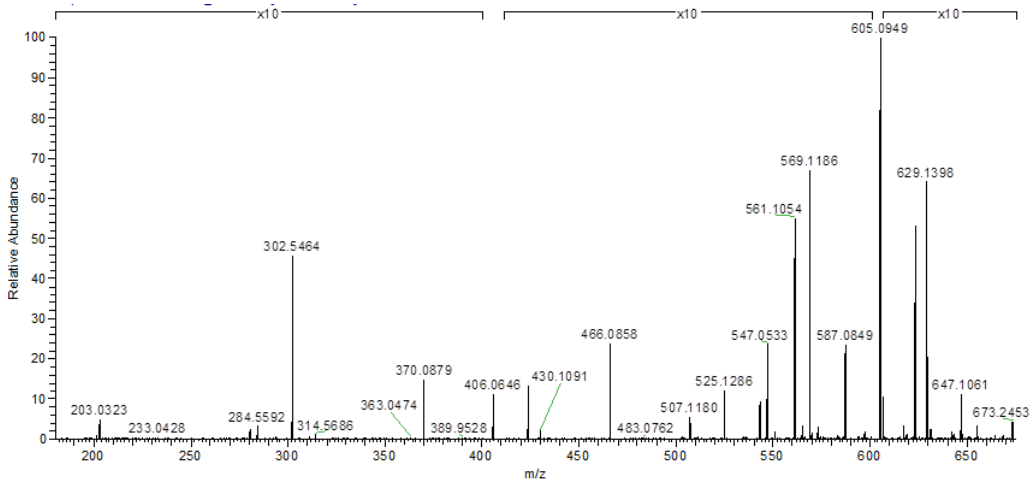


Figure B.14: *FT-MS*^{2} *fragment spectrum of hectochlorin (**89**).* Hectochlorin (**89**) $[M + H]^+$, Parent Mass 665 *m/z*.

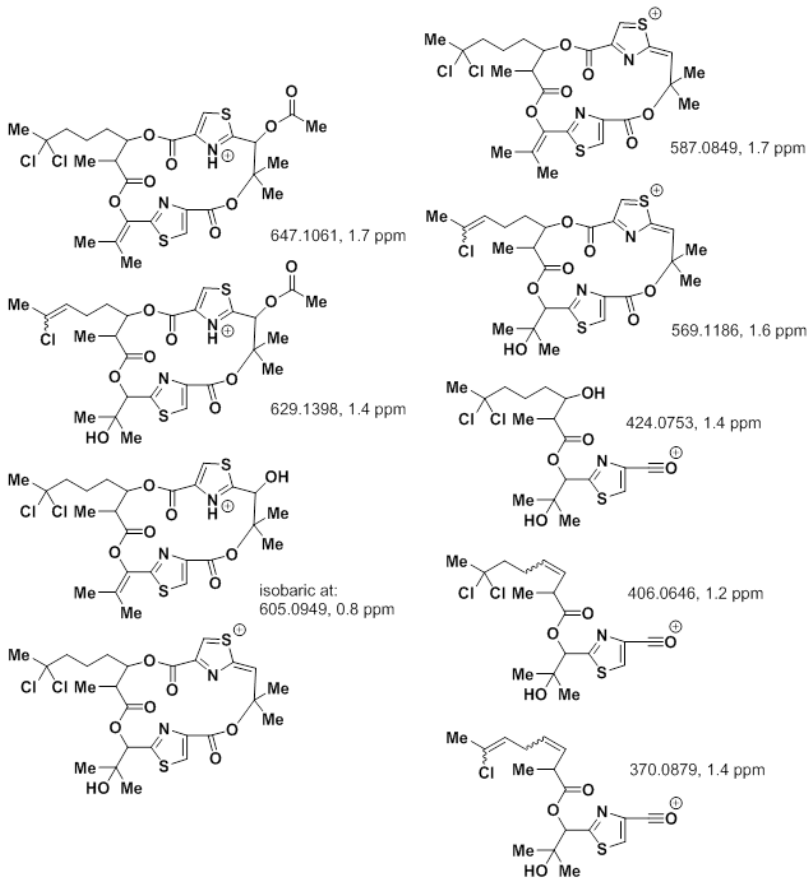


Figure B.15: *FT-MS*^{2} *fragment summary for hectochlorin (**89**).*

Table B.3: *NMR data summary for hectochlorin B (**97**).* All experiments in $CDCl_3$ with TMS standard, 1% *v/v* for ^{13}C experiment, and 0.03% *v/v* for the rest. [a]Carbon chemical shifts from 2D-NMR experiment extraction. [b]Spectra collected on a Bruker 600 MHz (600 MHz and 150 MHz for the ^{1}H and ^{13}C nuclei, respectively) with 1.7 mm inverse cryo-probe. [c]Proton species overlapped, shift and coupling extracted from the HSQC.

Position	δ_C, type[a]	δ_H (J in Hz)[b]	HMBC[b]	TOCSY[b]
1	173.2, C	–		
2	42.4, CH	3.14, quintet (7.6)	1, 3, *4 weak*, 9	3, 9
3	74.2, CH_2	5.32, m	*2,4 weak*	2, 4-6, 9
4a	30.8, CH_2	1.74, m		3
4b	30.8, CH_2	1.83, m		3
5	20.6, CH_2	1.73, m		3, 6
6a	49.2, CH_2	2.12, m	5, 7, *4, 8 weak*	3, 5, 6b
6b	49.2, CH_2	2.25, m	5, 7, *4, 8 weak*	3, 5, 6a
7	90.3, C	–		
8	37.2, CH_3	2.09, s	6, 7, *self*	
9	14.9, CH_3	1.28, d (7.2)c	1-3, *self*	2, 3
10	160.6, C	–		
11	146.3, C	–		
12	128.3, CH	8.11, s	10, 11, 13	
13	171.4, C	–		
14	76.3, CH	5.09, bs		
15	85.5, C	–		
16	26.7, CH_3	1.84, s	14, 15, 17	
17	22.2, CH_3	1.77, s	14-16	
18	162.3, C	–		
19	147.0, C	–		
20	128.5, CH	8.04, s	18, 19, 21	
21	164.9, C	–		
22	77.2, CH	5.58, s	1, 21, 23, 24	
23	71.6, C	–		
24	25.7, CH_3	1.35, s	22, 23, 25	
25	26.9, CH_3	1.28, sc	22-24	

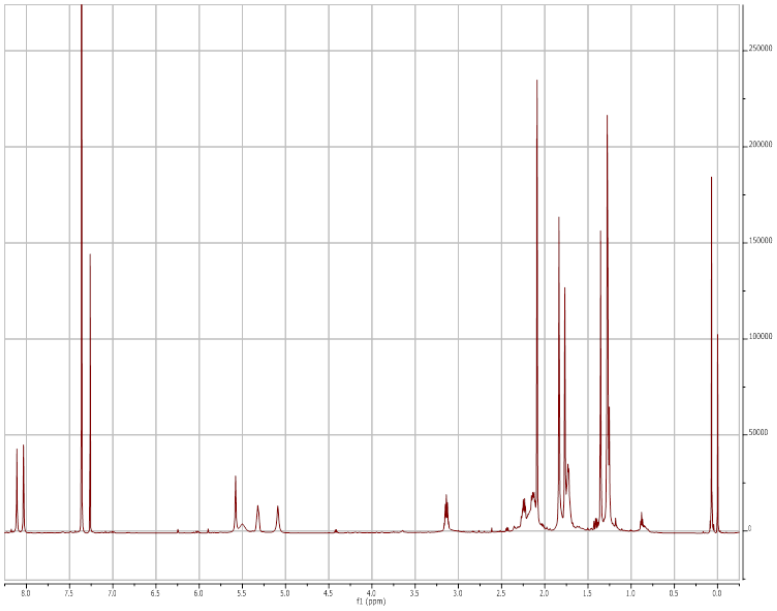


Figure B.16: *^{1}H-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin B (**97**).*

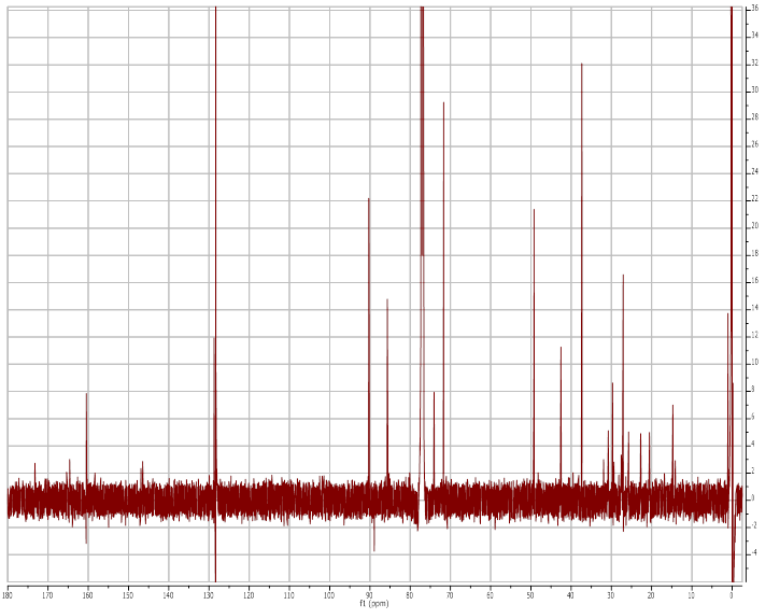


Figure B.17: *^{13}C-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin B (**97**).*

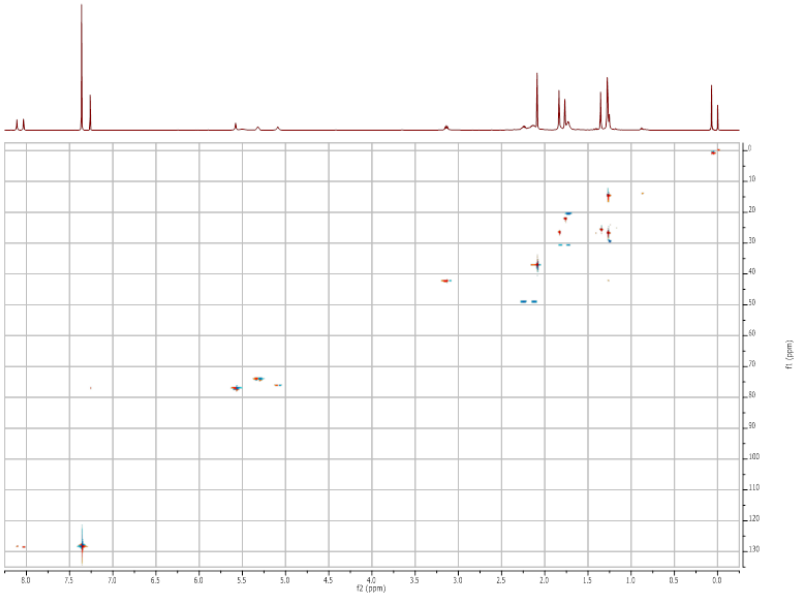


Figure B.18: *HSQC (600 MHz, $CDCl_3$) spectrum of hectochlorin B (**97**).*

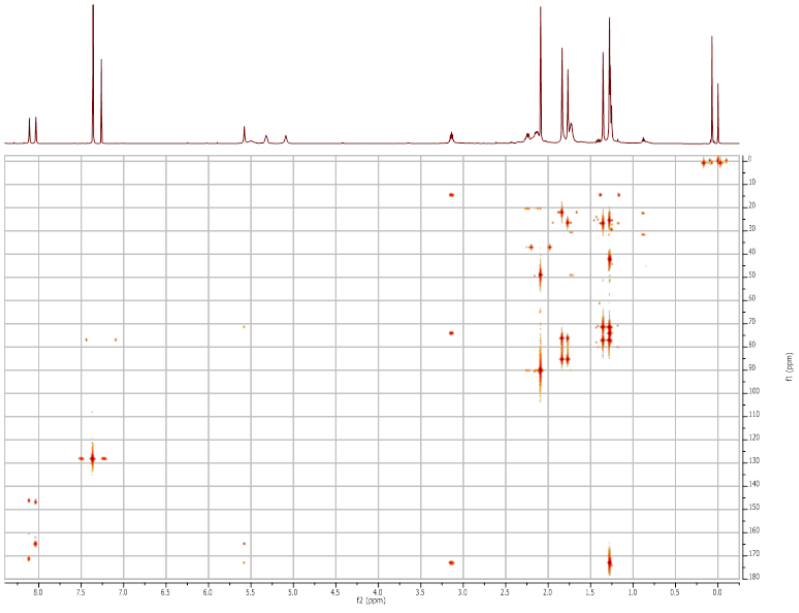


Figure B.19: *HMBC (600 MHz, $CDCl_3$) spectrum of hectochlorin B (**97**).*

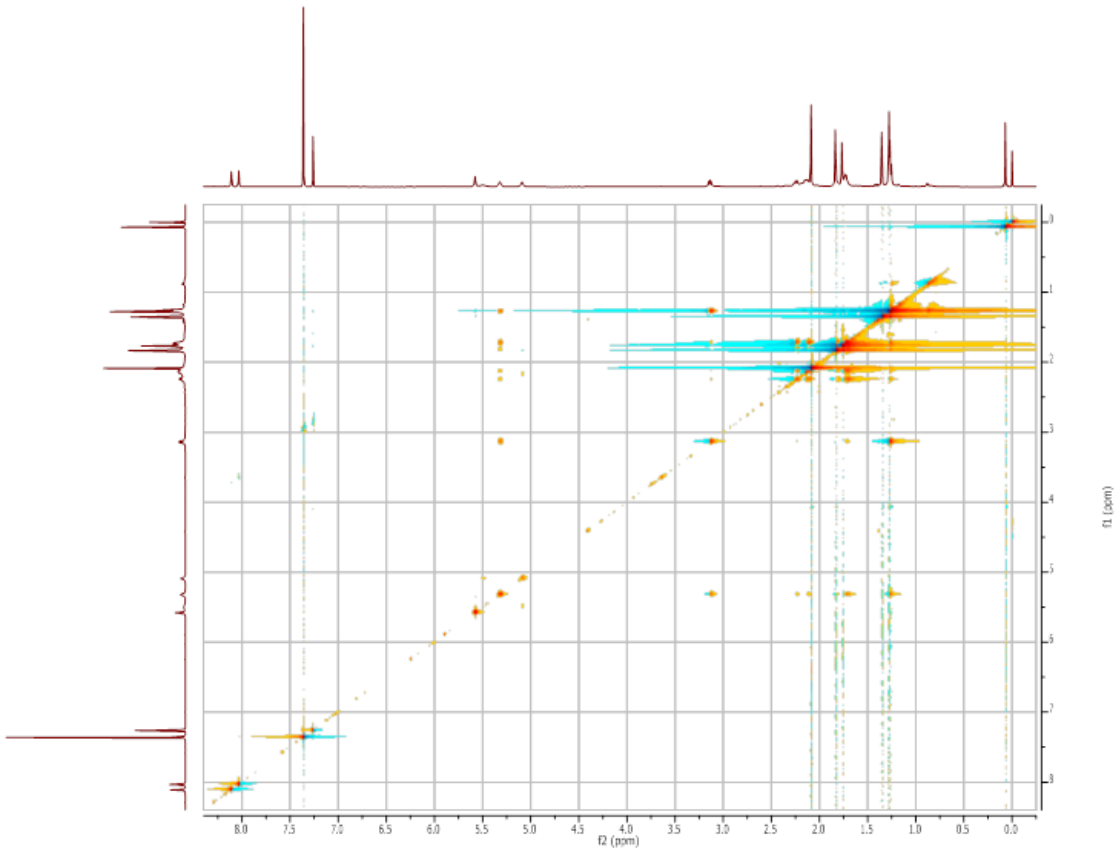


Figure B.20: *TOCSY-NMR (600 MHz, $CDCl_3$) spectrum of hectochlorin B (**97**).*

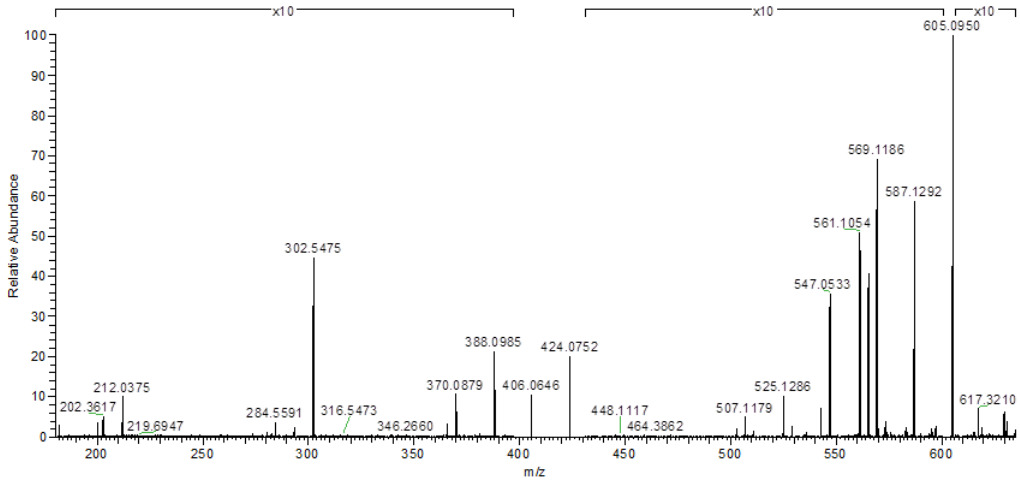


Figure B.21: *FT-MS*^{2} *fragment spectrum of hectochlorin B (**97**).* Hectochlorin B (**97**) $[M + H]^+$, Parent Mass 623 *m/z*.

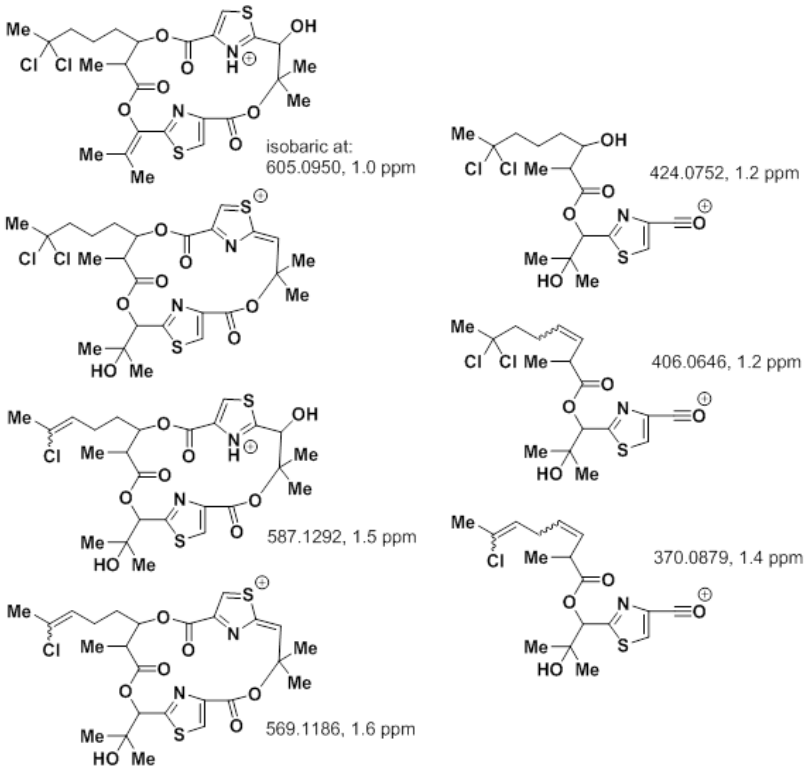


Figure B.22: *FT-MS*^{2} *fragment summary for hectochlorin B (**97**).*

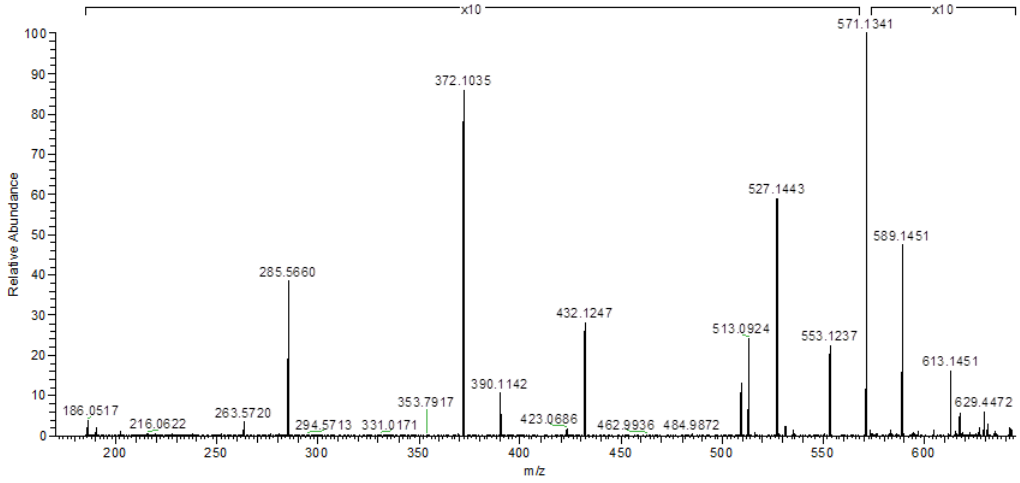


Figure B.23: *FT-MS*^{2} *fragment spectrum of hectochlorin C (**98**).* Hectochlorin C (**98**) $[M + H]^+$, Parent Mass 631 *m/z*.

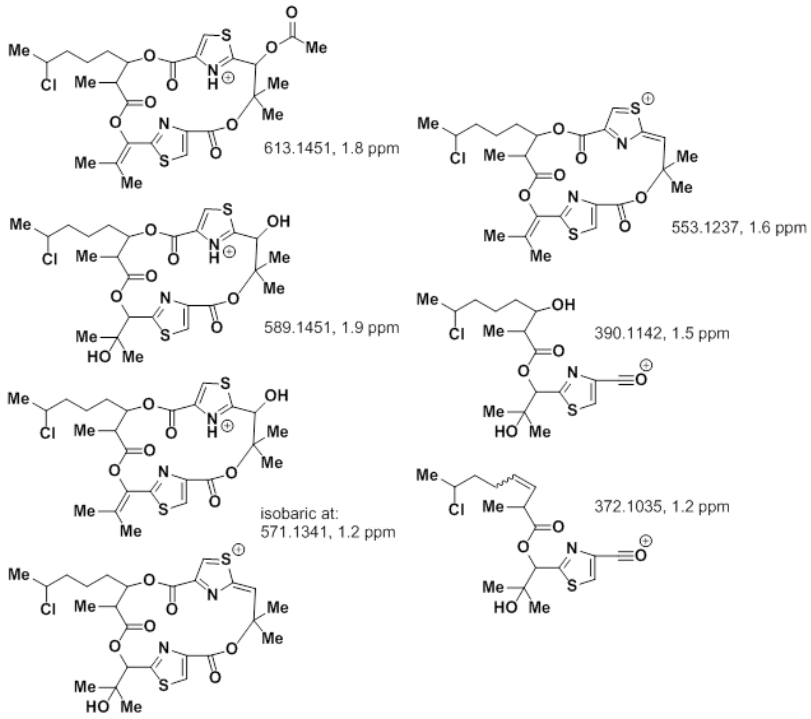


Figure B.24: *FT-MS*^{2} *fragment summary for hectochlorin C (**98**).*

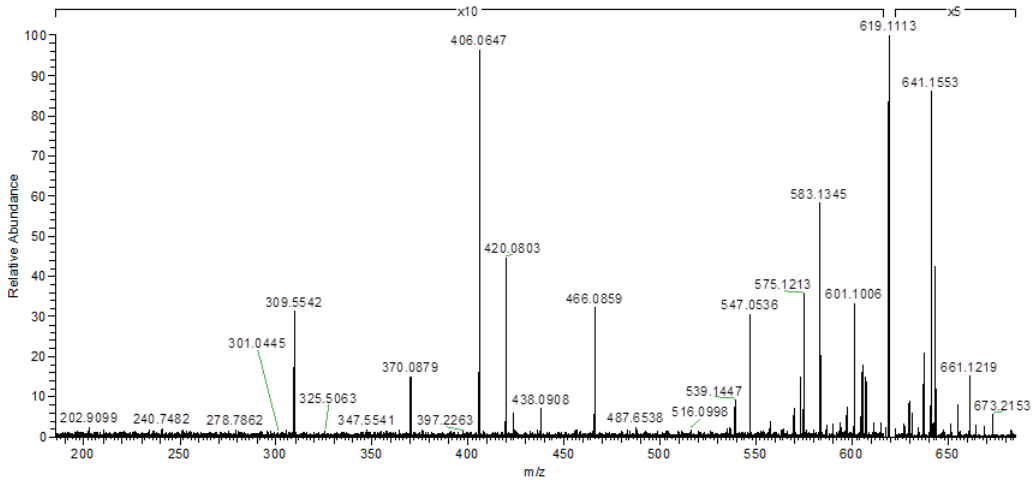


Figure B.25: *FT-MS*^{2} *fragment spectrum of hectochlorin D (**99**).* Hectochlorin D (**99**) $[M + H]^+$, Parent Mass 679 *m/z*.

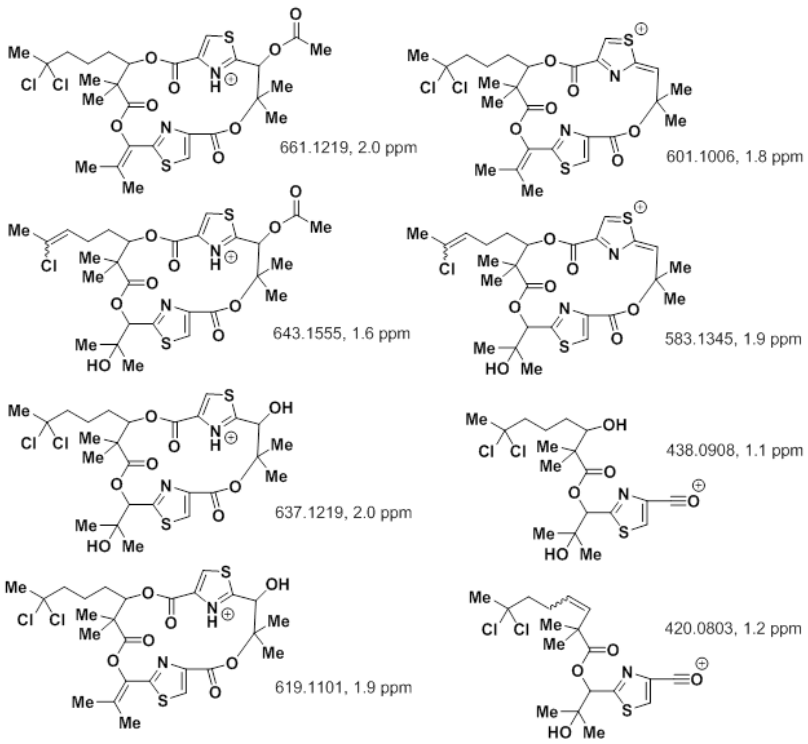


Figure B.26: *FT-MS*^{2} *fragment summary for hectochlorin D (**99**).*

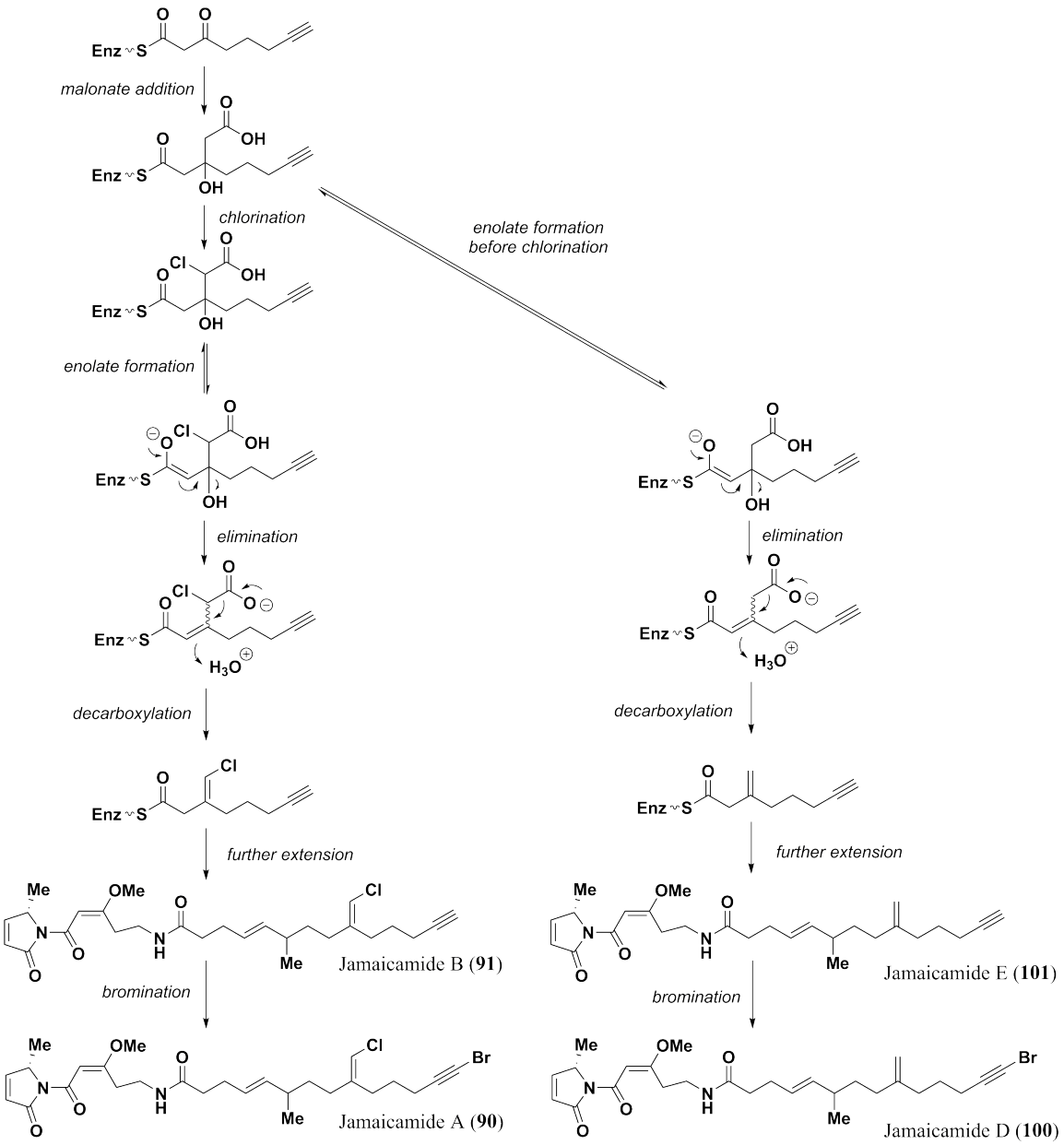


Figure B.27: *Proposed biosynthetic pathway divergence for jamaicamide A and D (**90** and **100**).*

Table B.4: *NMR data summary for jamaicamide D (**100**).* All experiments in $CDCl_3$ with 0.03% *v/v* TMS standard, on a Varian Unity 500 MHz (500 MHz and 125 MHz for the ^1H and ^{13}C nuclei respectively). [a]Carbon chemical shifts by projection from HMBC experiment. [b]Overlapping with water peak, shift determined by TOCY correlation. [c]Proton species overlapped.

Position	δ_C, type[a]	δ_H (J in Hz)	HMBC	TOCSY
1		1.50[b]		3
2		–		
3		2.20, t (7.1)		1, 4, 5
4		1.60-1.66, m		3, 5
5		2.06-2.10, m		3, 4
6		–		
7		2.01-2.05, m		8, 26
8	34.9	1.29-1.40, m		7, 9, 26
9	36.6	1.91-1.95, m		8, 26
10	137.1	5.26-5.39, m[c]		12, 13, 26
11		5.26-5.39, m[c]		12, 13, 26
12		2.26-2.30, m		10, 11, 13a
13a		2.17-2.20, m		10-12, 13b
13b		2.26-2.30, m		10, 11, 13a
14		–		
NH		6.69, bs		15
15		3.47-3.54, m		16, NH
16a		2.81-2.86, m		15, 16b
16b		2.98-3.51, m		15, 16a
17	175.6	–		
18		6.73, s		
19		–		
20		–		
21		6.08, dd (1.4, 6.0)		22
22	153.1	7.23, dd (2.0, 5.9)		21, 24
23	58.2	4.86, q (6.7)		24
24		1.46, d (6.7)	22, 23	22, 23
25		3.75, s	17	
26		0.95, d (6.7)	8-10	7-11
27		4.70, m		

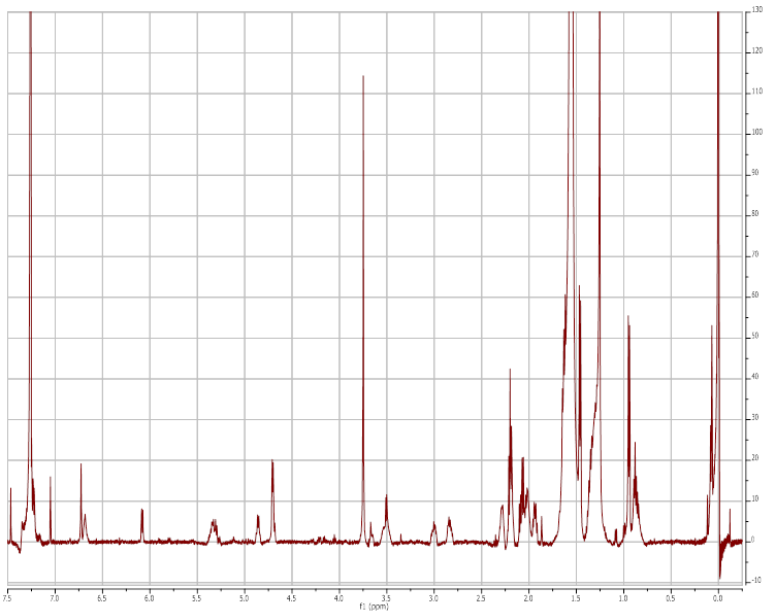


Figure B.28: *^{1}H-NMR (500 MHz, $CDCl_3$) spectrum of jamaicamide D (**100**).*

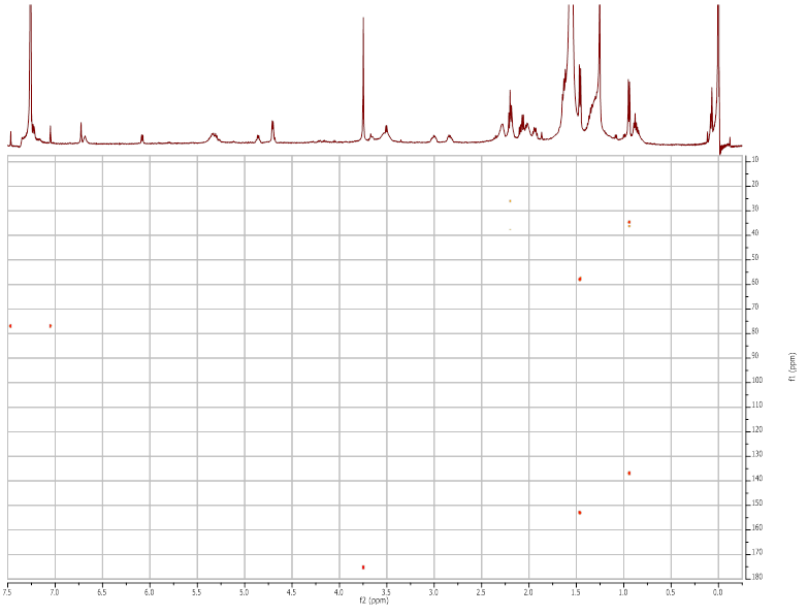


Figure B.29: *HMBC-NMR (500 MHz, $CDCl_3$) spectrum of jamaicamide D (**100**).*

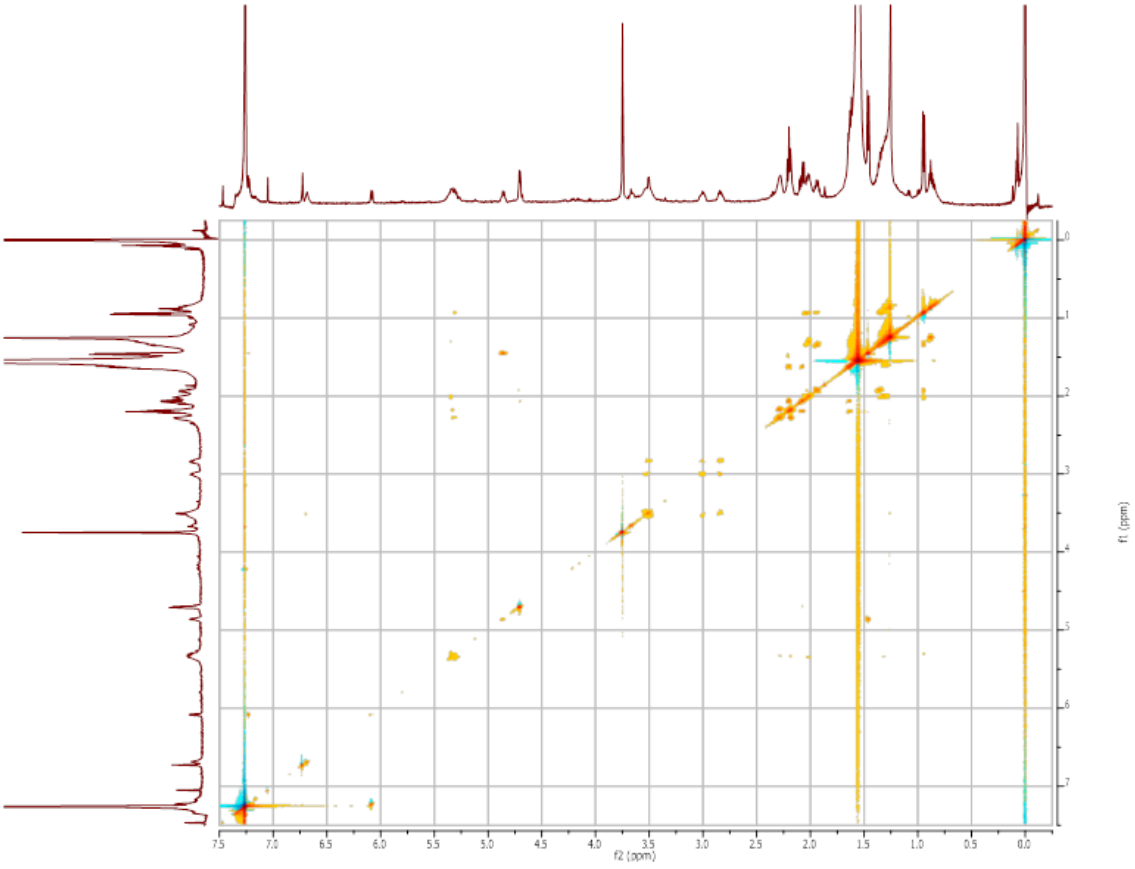


Figure B.30: *TOCSY-NMR (500 MHz, $CDCl_3$) spectrum of jamaicamide D (**100**).*

Table B.5: *NMR data summary for jamaicamide F (**104**).* All experiments in $CDCl_3$ with TMS standard, 1% *v/v* for ^{13}C experiment, and 0.03% *v/v* for the rest. [a]Specturm collected on a Varian VX 500 MHz with ^{13}C-optimized cryoprobe. [b]Spectra collected on a Bruker 600 MHz (600 MHz and 150 MHz for the ^{1}H and ^{13}C nuclei, respectively) with 1.7 mm inverse cryo-probe. [c,d]Proton species overlapped, shift assigned by HSQC. [e]Multiplet shift assigned by HSQC.

Position	δ_C, type[a]	δ_H (J in Hz)[b]	HMBC[b]	TOCSY[b]
1	-6.6, C	–		
2	94.2, C	–		
3	20.7, CH_2	2.36, t (7.2)	1, 2, 4, 5, self	4, 5
4	26.1, CH_2	1.64, quin (7.1)	2, 3, 5, 6, self	3, 5
5	29.2, CH_2	2.25,[c,e] m	3, 4, 6, 7, 27	3, 4
6	141.8, C	–		
7	32.6, CH_2	2.00,[d,e] m	5, 6, 8, 9, 27	8, 26
8	34.7, CH_2	1.34,[e] m	6, 10, 9, 7, 26	7, 9, 26
9	36.3, CH	2.02,[d,e] m	7, 10, 11, 26	8, 26
10	136.6, CH	5.27, dd (7.7, 15.4)	8, 9, 11, 12, 26, self	9, 11, 12, 13a, 26
11	127.5, CH	5.36, td (6.5, 15.2)	9, 10, 12	9, 10, 12, 13a, 26
12	28.6, CH_2	2.29,[c,e] m	10, 11, 13a	
13a	36.8, CH_2	2.18, t (7.6)	11, 12, 14	10, 11, 12, 13b
13b	36.8, CH_2	2.29,[c,e] m	10, 11, 14, self	13a
14	172.4, C	–		
NH	–	6.69, bs	14, 15 15, 16	
15a	38.3, CH_2	3.29,[e] m	17, self	NH, 15b, 16
15b	38.3, CH_2	3.54,[e] m	17, self	NH, 15a, 16
16a	32.3, CH_2	2.84,[e] m	15, 17, 18	NH, 15, 16b
16b	32.3, CH_2	3.00,[e] m	15, 17, 18	NH, 15, 16a
17	175.4, C	–		
18	95.0, CH	6.73, s	16, 17, self	
19	166.1, C	–		
20	170.1, C	–		
21	125.9, CH	6.08, dd (6.0, 1.6)	20, 22, 23, 24, self	22-24
22	153.2, CH	7.23, dd (2.0, 6.1)	20, 21, 23, 24, self	21, 23, 24
23	58.1, CH	4.86, tq (1.4, 7.0)	*19, 20 weak*, 21, 22, 24	21, 22, 24
24	17.9, CH_3	1.46, d (6.7)	22, 23, self	21-23
25	56.2, CH_3	3.75, s	17, self	
26	20.9, CH_3	0.95, d (6.7)	8-10, self	8-11
27	112.8, CH	5.79, s	5-7, self	

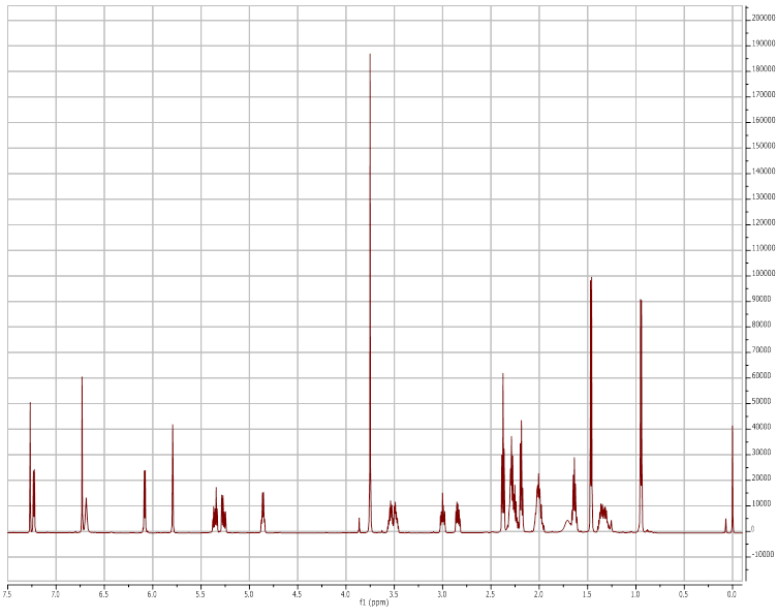


Figure B.31: ^{1}*H-NMR (600 MHz, $CDCl_3$) spectrum of jamaicamide F (**104**).*

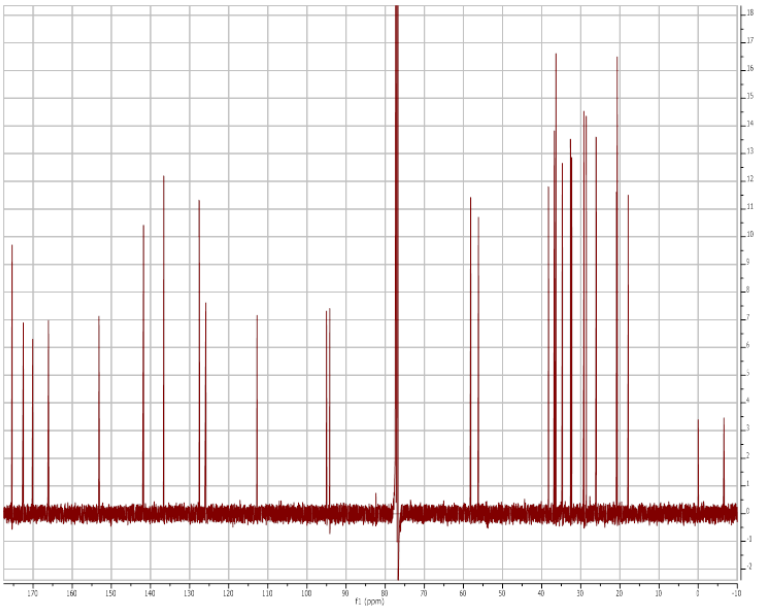


Figure B.32: ^{13}*C-NMR (125 MHz, $CDCl_3$) spectrum of jamaicamide F (**104**).*

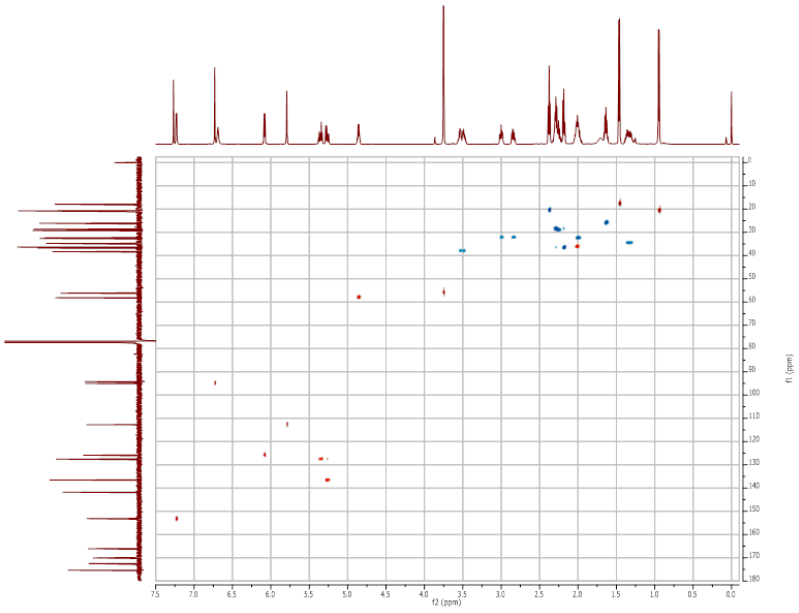


Figure B.33: *HSQC-NMR (600 MHz, $CDCl_3$) spectrum of jamaicamide F (**104**).*

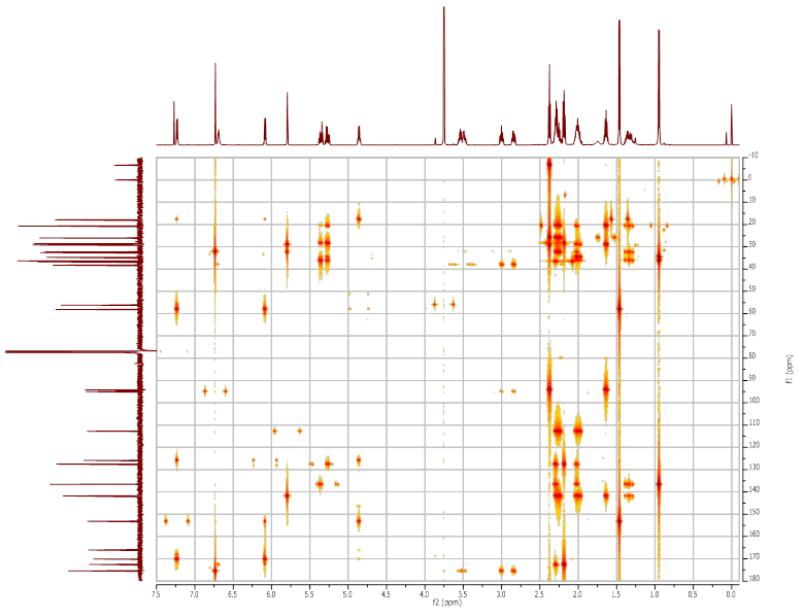


Figure B.34: *HMBC-NMR (600 MHz, $CDCl_3$) spectrum of jamaicamide F (**104**).*

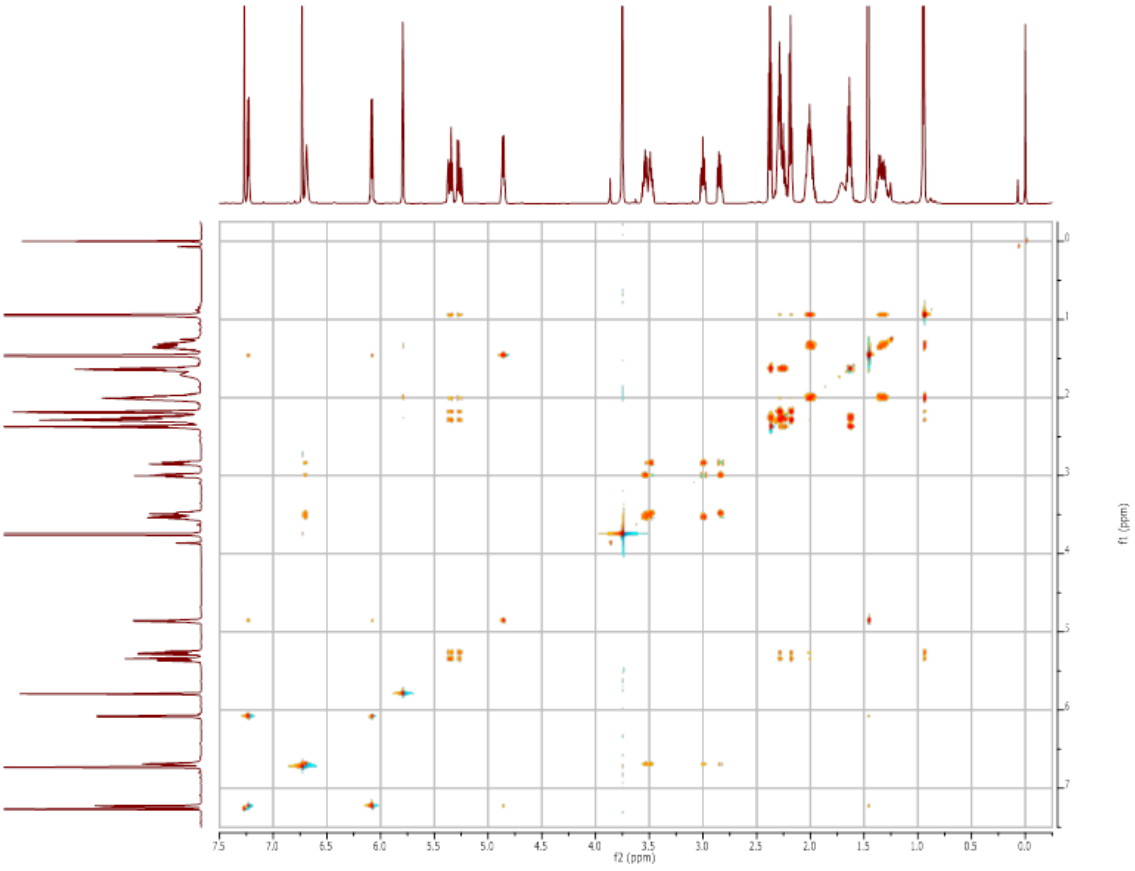


Figure B.35: *TOCSY-NMR (600 MHz, $CDCl_3$) spectrum of jamaicamide F (**104**).*

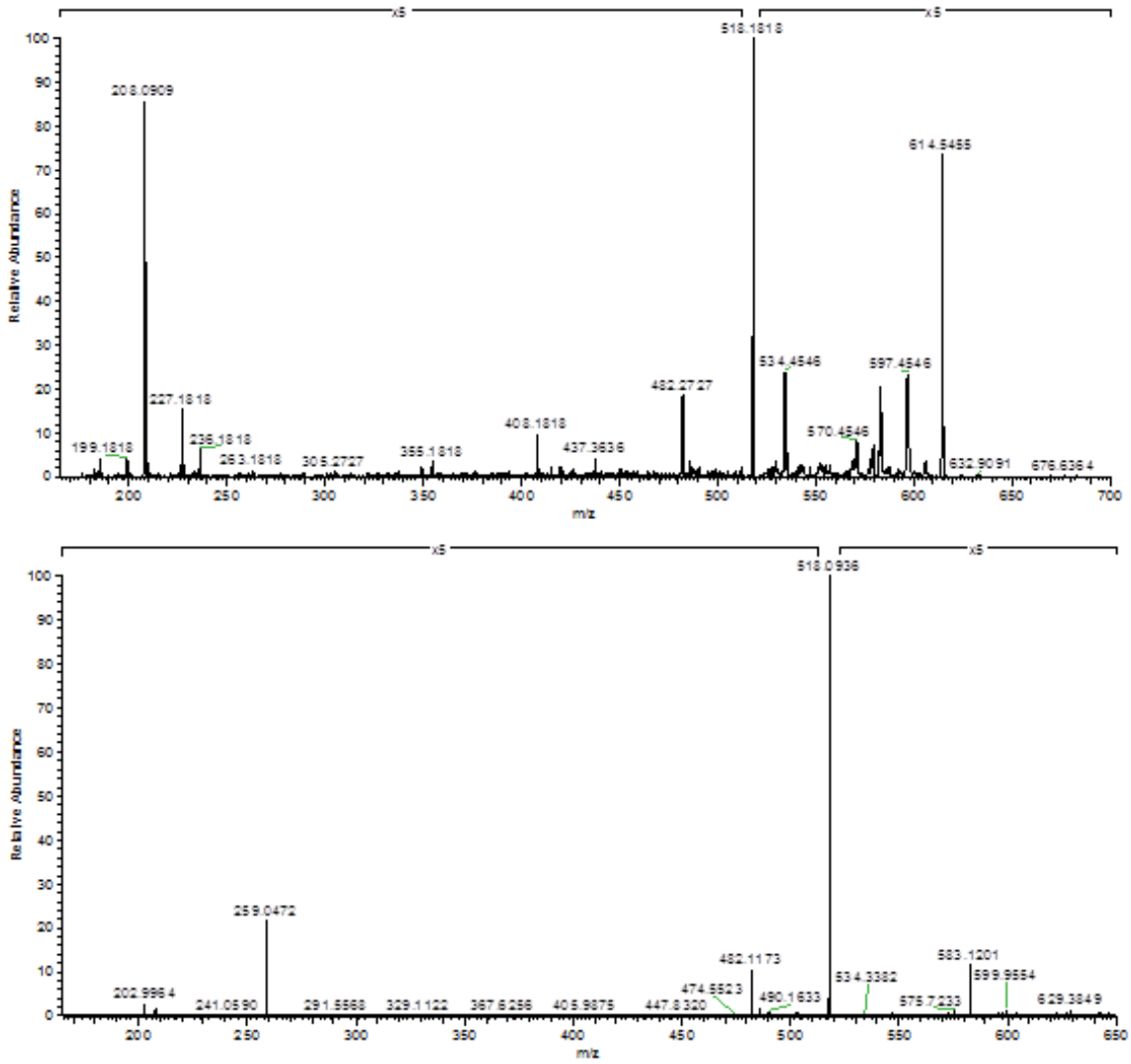


Figure B.36: *MS*^{2} *fragment spectra of jamaicamide F (**104**). Top Figure* ITMS^2 spectrum, parent mass of 615, $[M+H]^+$ of jamaicamide F (**104**). *Bottom Figure* FTMS^2 spectrum, parent mass of 615, $[M+H]^+$ of jamaicamide F (**104**).

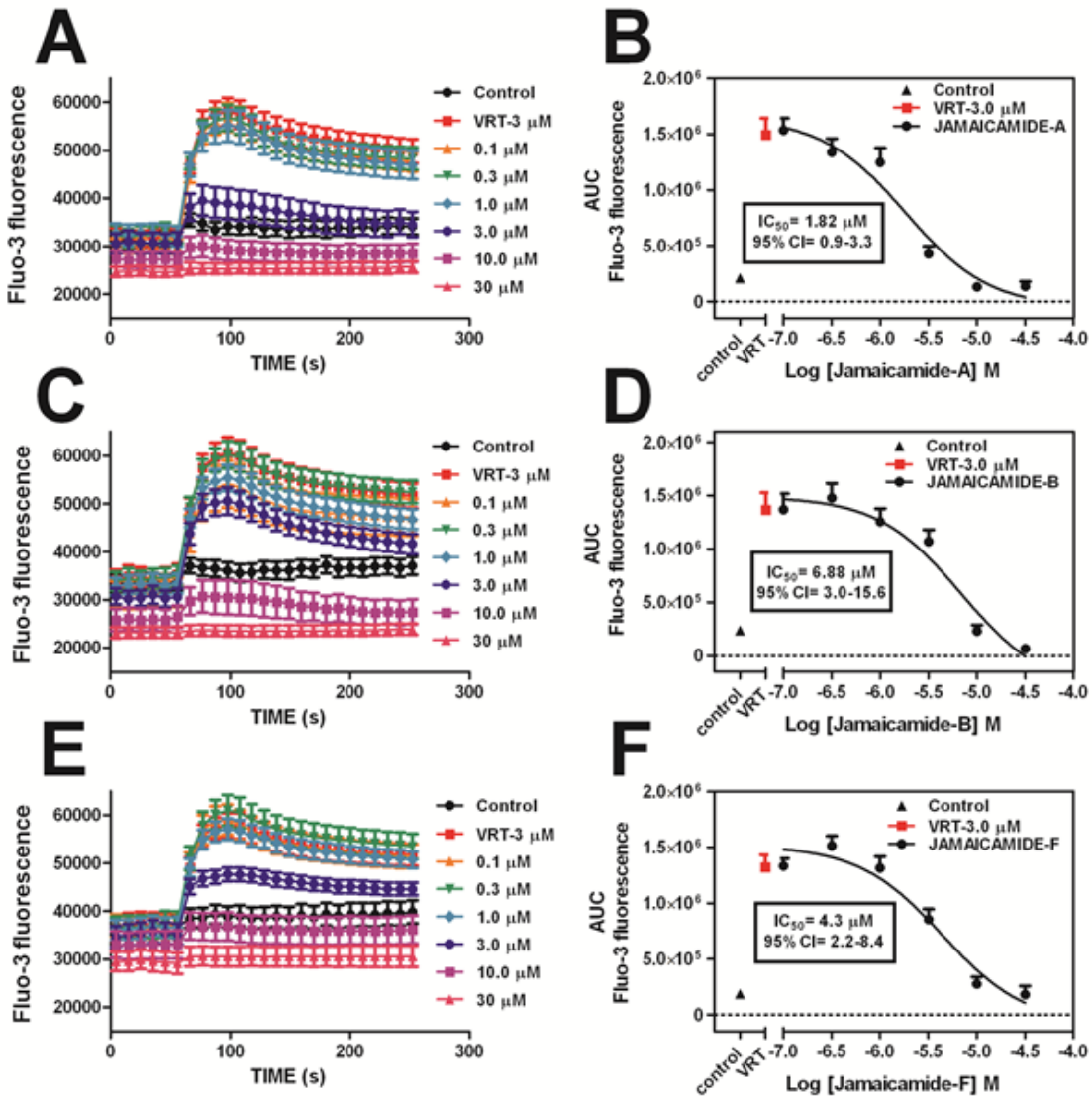


Figure B.37: *Effect of the jamaicamides on the veratridine-induced Ca^{2+} influx in murine neocortical neurons.*

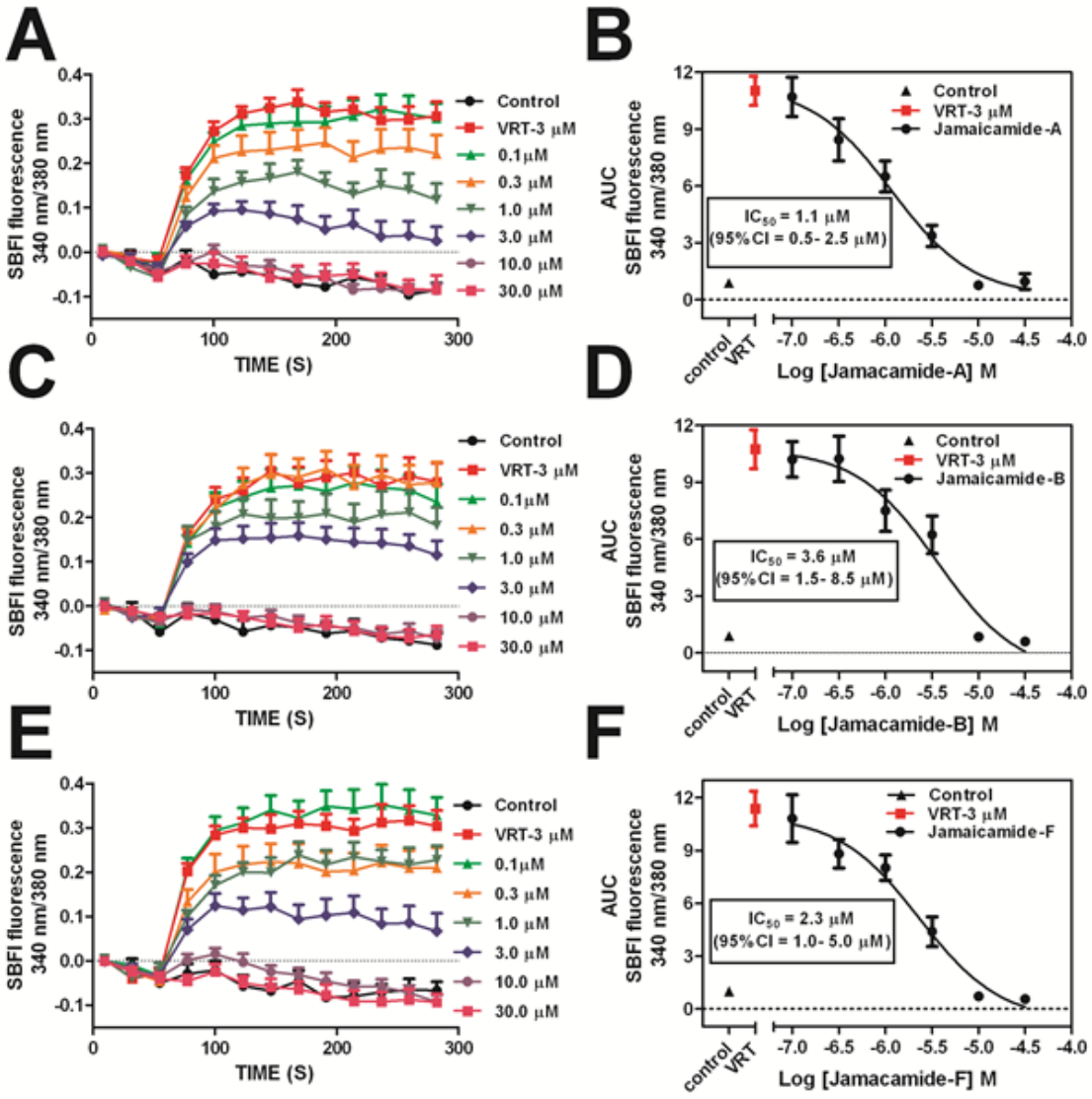


Figure B.38: *Effect of the jamaicamides on the veratridine-induced Na^+ influx in murine neocortical neurons.*

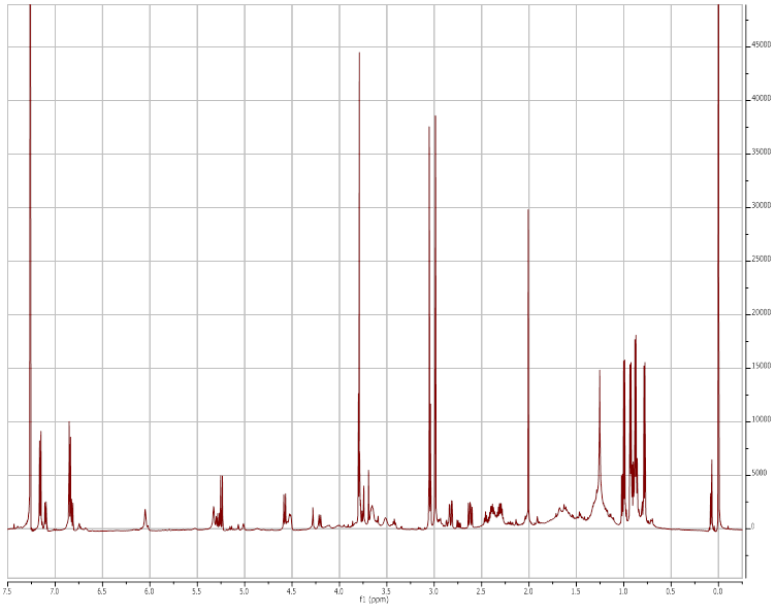


Figure B.39: ^{1}*H-NMR (600 MHz, $CDCl_3$) spectrum of hectoramide (**96**).*

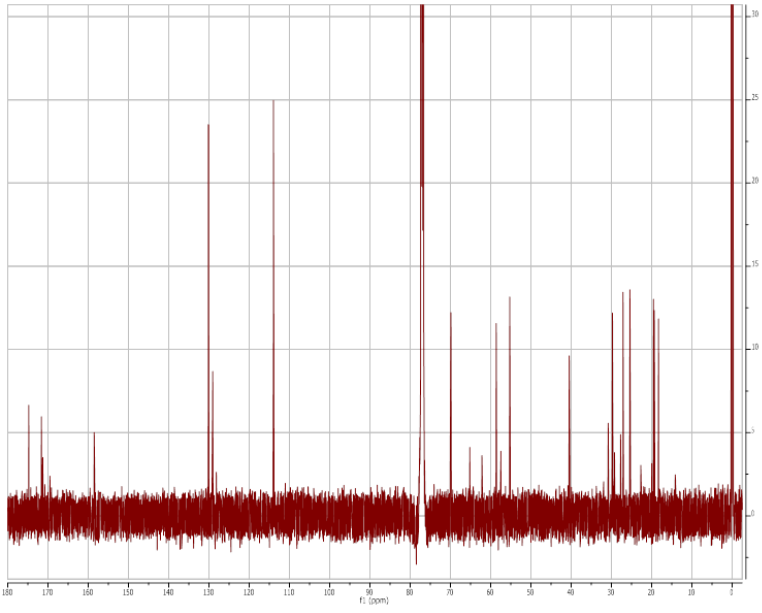
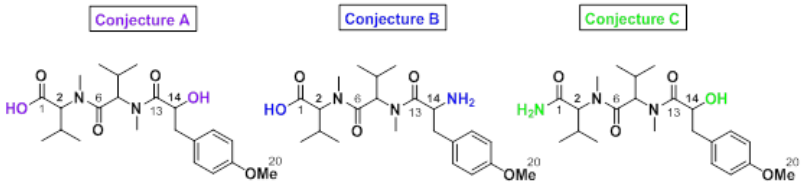


Figure B.40: ^{13}*C-NMR (125 MHz, $CDCl_3$) spectrum of hectoramide (**96**).*



Postion, δ_C	Conjecture A ChemDraw δ_C Prediction, Difference	Conjecture B ChemDraw δ_C Prediction, Difference	Conjecture C ChemDraw δ_C Prediction, Difference
1, 171.3	175.7, -4.4	175.7, -4.4	176.7, -5.4
2, 62.2	**80.4, -18.2**	**80.4, -18.2**	**72.5, -10.3**
3, 25.4	27.7, -2.3	27.7, -2.3	26.7, -1.3
4, 18.2(8)	19.2, -0.9	19.2, -0.9	18.8, -0.5
5, 19.6	19.2, 0.4	19.2, 0.4	18.8, 0.8
6, 30.8	32.5, -1.7	32.5, -1.7	32.1, -1.3
7, 171.7	172.7, -1.0	172.7, -1.0	172.7, -1.0
8, 58.6	68.1, -9.5	67.8, -9.2	67.7, -9.1
9, 27.1	27.3, -0.2	27.3, -0.2	27.3, -0.2
10, 18.3(4)	18.8, -0.5	18.8, -0.5	18.8, -0.5
11, 19.4	18.8, 0.6	18.8, 0.6	18.8, 0.6
12, 29.7	32.4, -2.7	32.1, -2.4	32.4, -2.7
13, 174.8	168.9, 5.9	170.0, 4.8	168.9, 5.9
14, 69.9	**70.4, -0.5**	**51.2, 18.7**	**70.4, -0.5**
15, 40.5	41.5, -1.0	39, 1.5	41.5, -1.0
16, 129.1	128.3, 0.8	128.9, 0.2	128.3, 0.8
17, 130.1	129.8, 0.3	129.8, 0.3	129.8, 0.3
18, 114.0	114.2, -0.2	114.2, -0.2	114.2, -0.2
19, 158.5	157.8, 0.7	157.8, 0.7	157.8, 0.7
20, 55.3	55.8, -0.5	55.8, -0.5	55.8, -0.5

Figure B.41: *Predicted* ^{13}*C-NMR shifts for potential structures of hectoramide (**96**).*

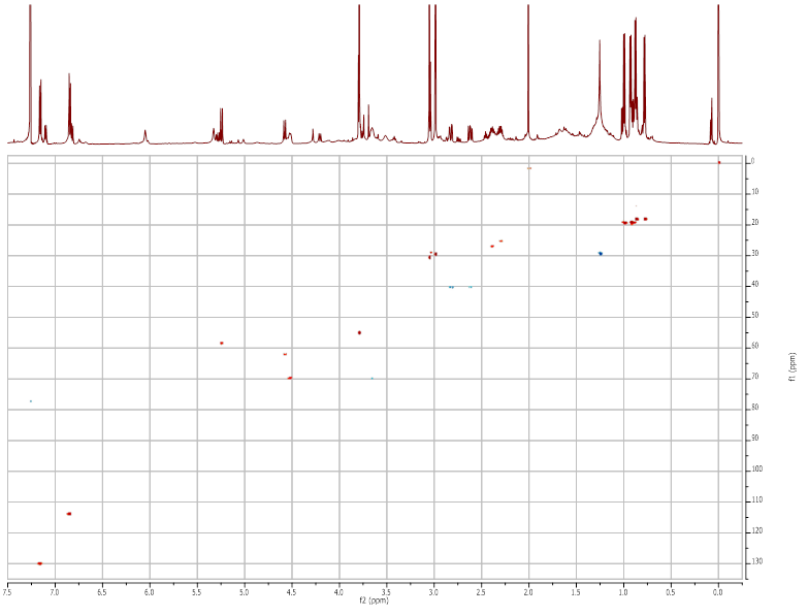


Figure B.42: *HSQC (600 MHz,* $CDCl_3$*) spectrum of hectoramide (**96**).*

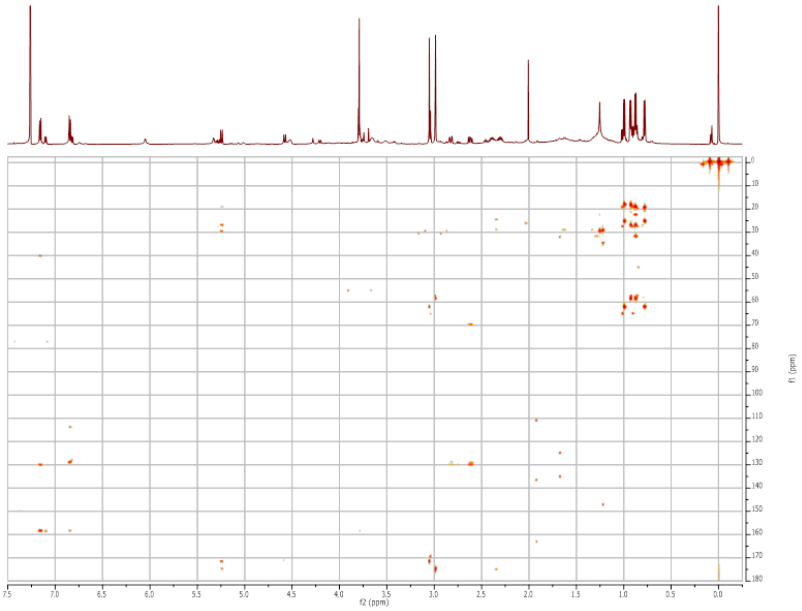


Figure B.43: *HSQC (600 MHz, $CDCl_3$) spectrum of hectoramide (**96**).*

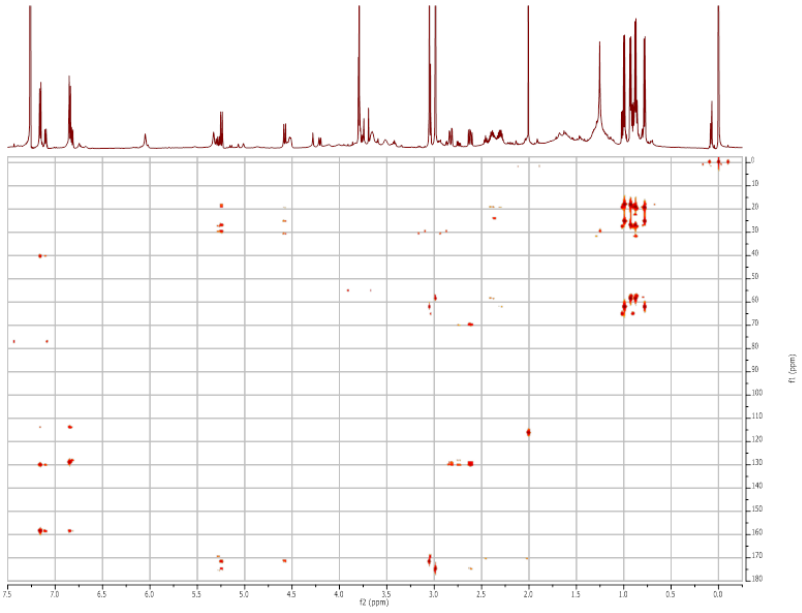


Figure B.44: *HMBC (600 MHz, $CDCl_3$) spectrum of hectoramide (**96**).*

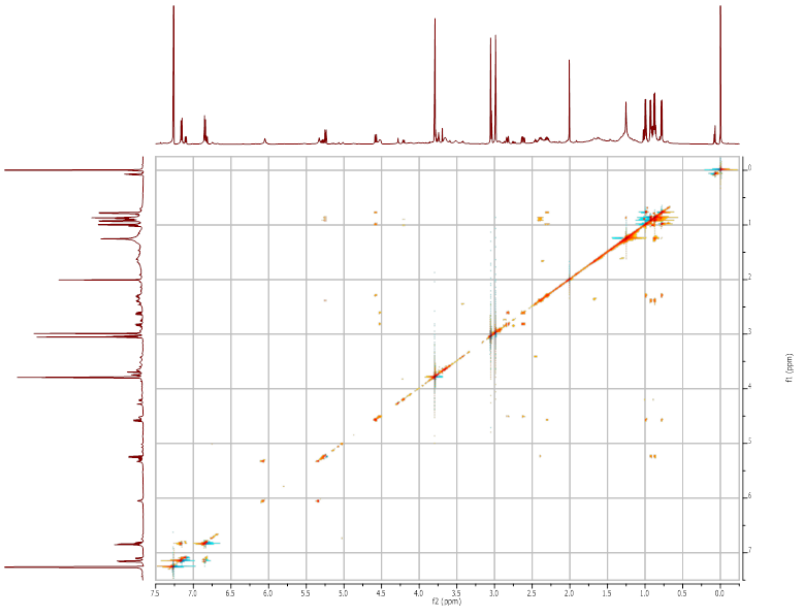


Figure B.45: *TOCSY (600 MHz, $CDCl_3$) spectrum of hectoramide (**96**).*

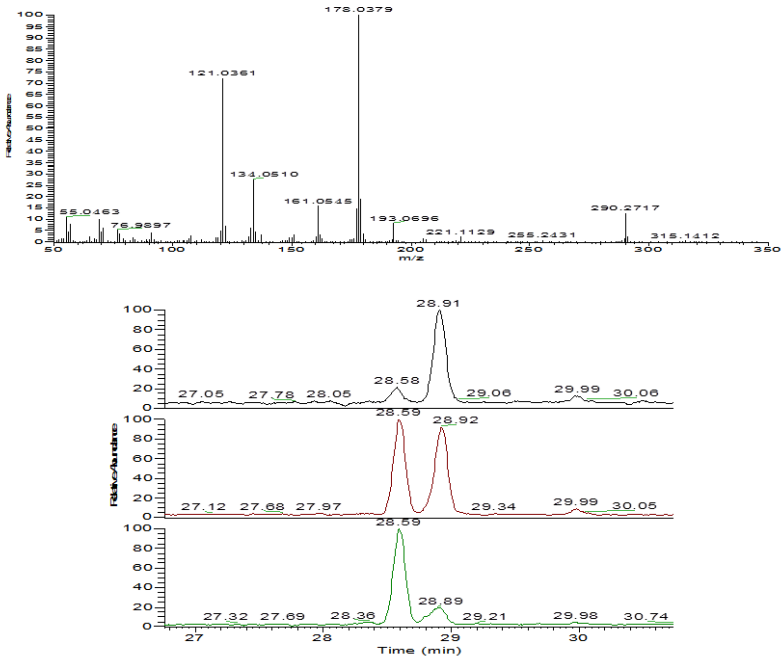


Figure B.46: *GC-MS analysis of 2-*S*-octonol ester standards. Top Figure* Spectrum of Derivatized *S* Standard. Extracted ion chromatograms at 177.75-178.25 *m/z*. *Top Chromatogram* Sample of *S* Standard. *Middle Chromatogram* Sample of 1:1 mix of *S/R* Standards. *Bottom Chromatogram* Sample of *R* Standard.

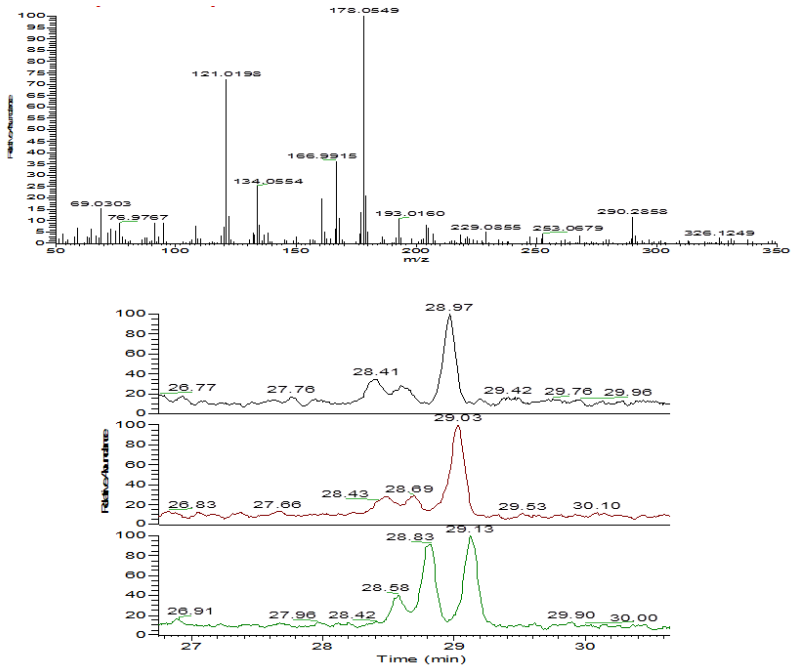


Figure B.47: *GC-MS analysis of* 2-S-*octonol ester derivatized hydrolysate of hectoramide (**96**). Top Figure* Spectrum of Derivatized Sample. Extracted ion chromatograms at 177.75-178.25 *m/z*. *Top Chromatogram* Sample alone. *Middle Chromatogram* Co-injection with *S* Standard. *Bottom Chromatogram* Co-injection with *R* Standard.

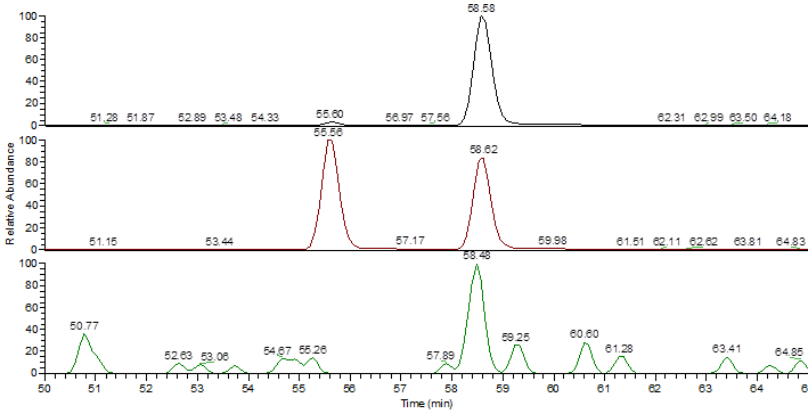


Figure B.48: *Marfey's analysis of D-FDAA derivatized hydrolysate of hectoramide (**96**).* Extracted ion chromatograms of 383.70-384.10 *m/z*. *Top Figure* D-FDAA derivatized L-*N*-Me-Val standard. *Middle Figure* D-FDAA derivatized D/L-*N*-Me-Val standard. *Bottom Figure* D-FDAA derivatized hydrolyaste.

Appendix C

Supplementary Information for Chapter 6

NMR Spectra of Gallinamide Synthetic Intermediates

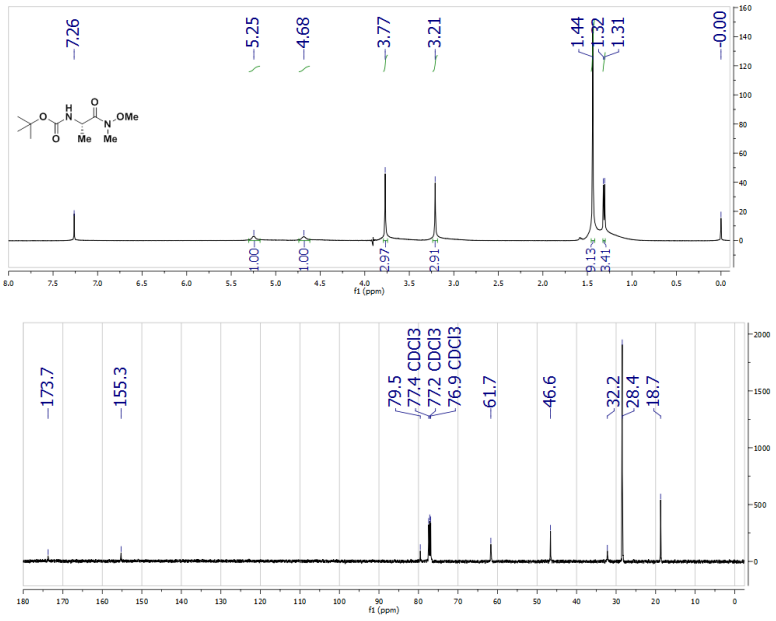


Figure C.1: *^{1}H and ^{13}C-NMR spectra of **110a**.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

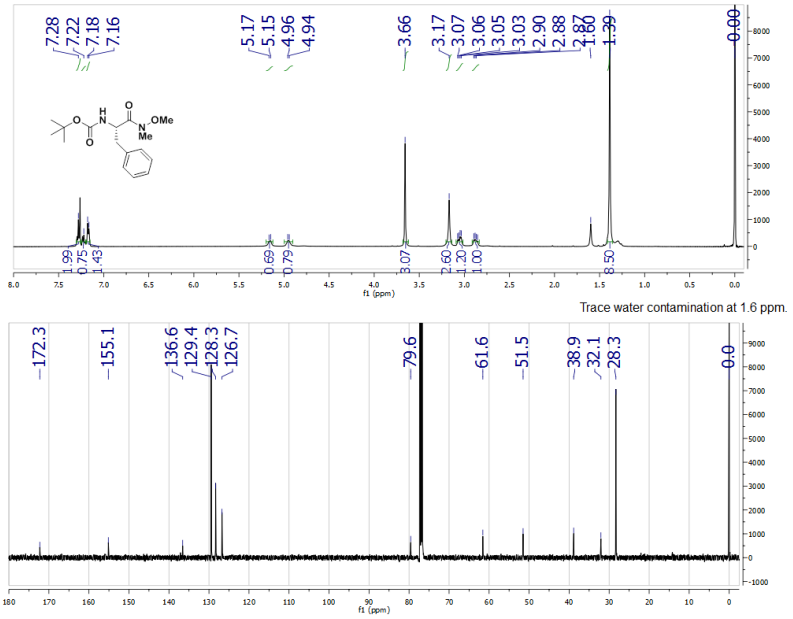


Figure C.2: ^{1}H *and* ^{13}C*-NMR spectra of* ***110b***. Spectra collected in $CDCl_3$ with 1.0% *v/v* TMS on an inverse 500 MHz cryo-probe.

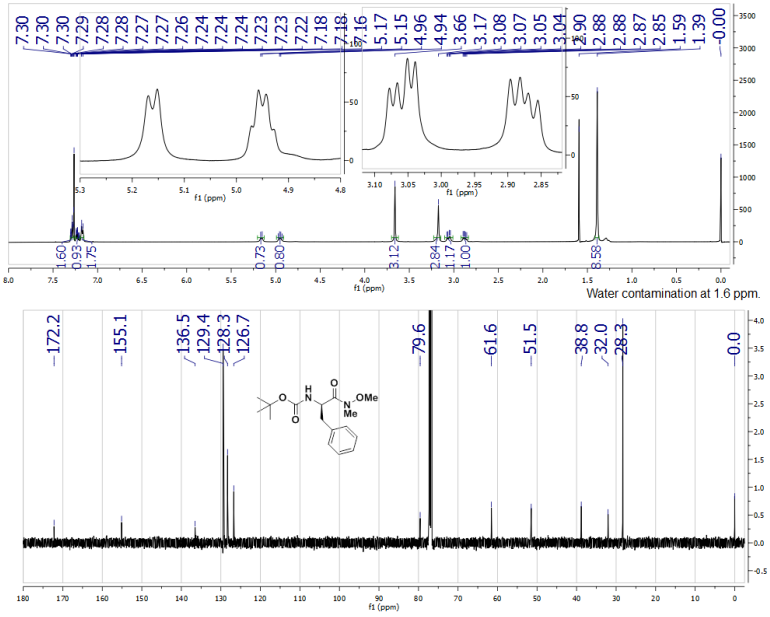


Figure C.3: ^{1}H *and* ^{13}C*-NMR spectra of* ***110c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

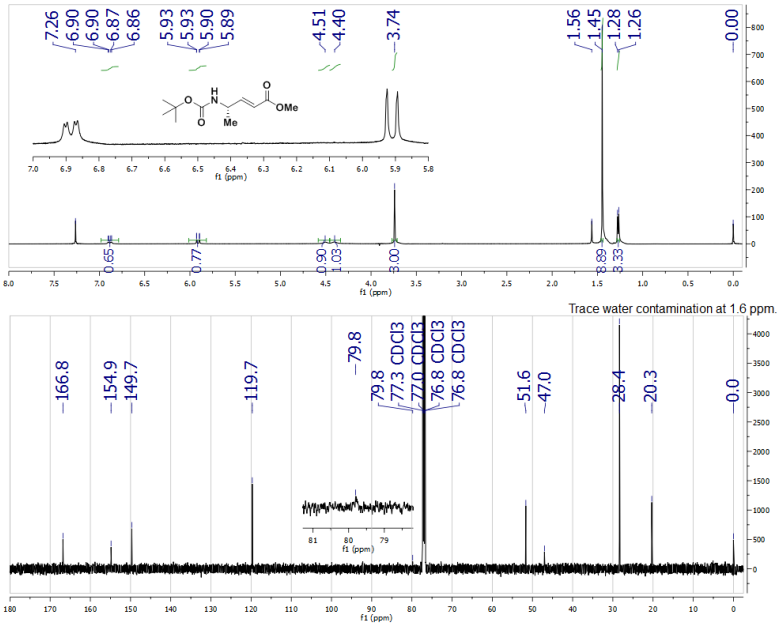


Figure C.4: ^{1}H *and* ^{13}C*-NMR spectra of* ***111a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

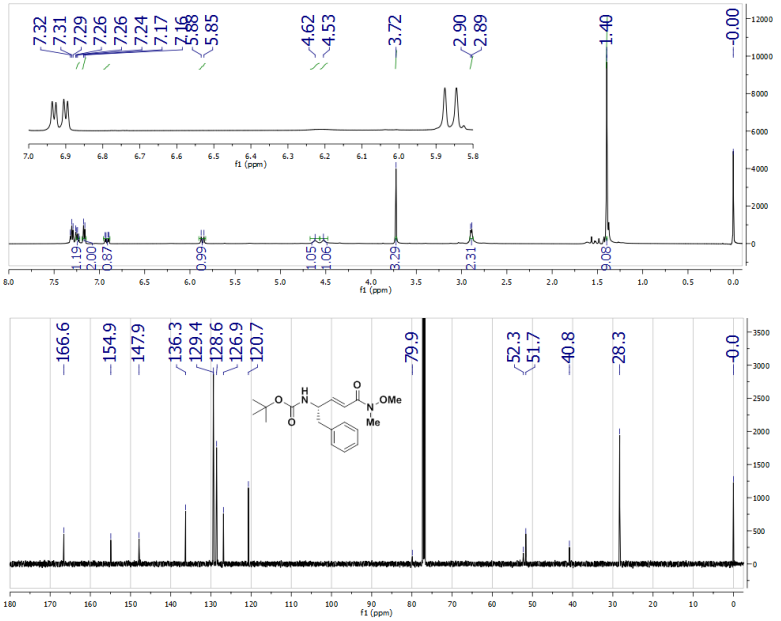


Figure C.5: ^{1}H *and* ^{13}C*-NMR spectra of* ***111b***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

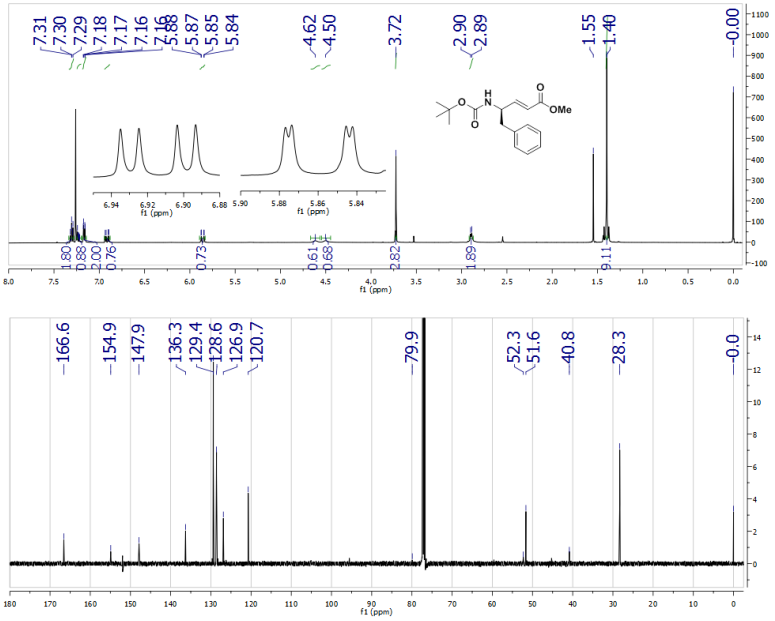


Figure C.6: ^{1}H *and* ^{13}C*-NMR spectra of* ***111c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

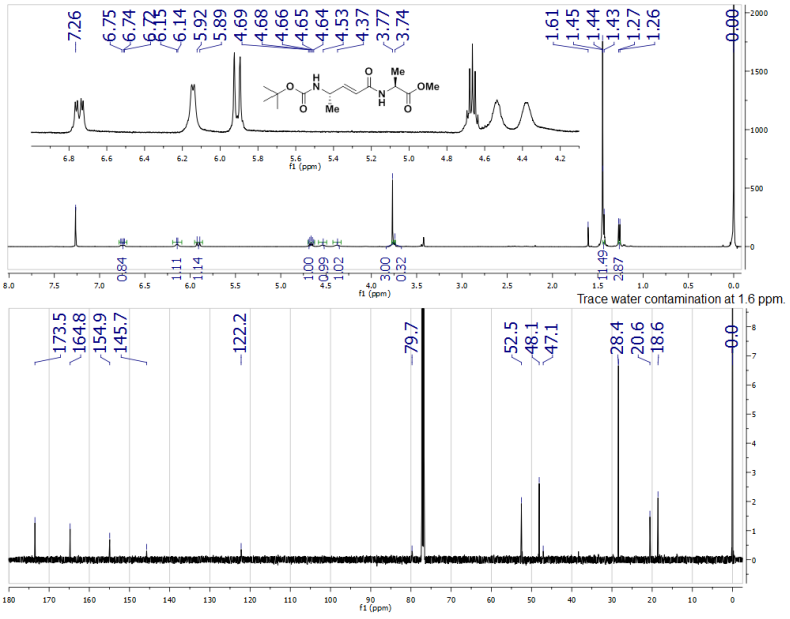


Figure C.7: ^{1}H *and* ^{13}C*-NMR spectra of* ***112a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

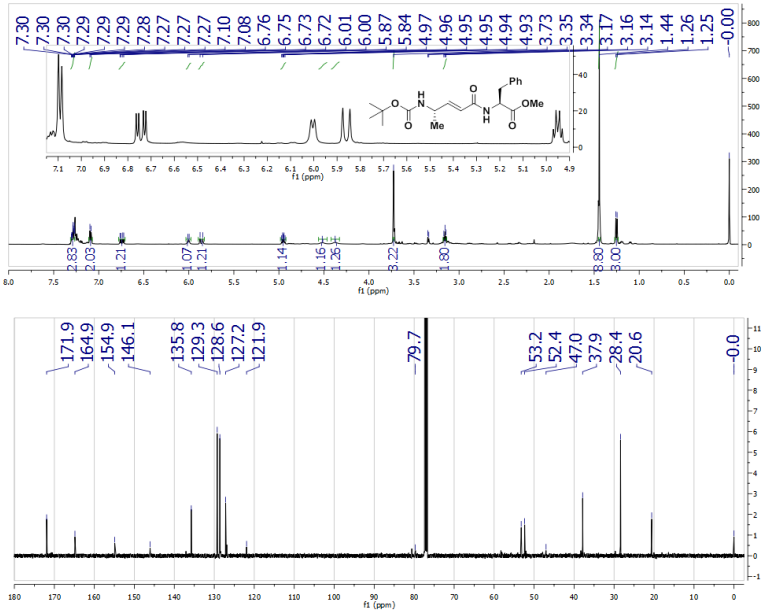


Figure C.8: ^{1}H *and* ^{13}C*-NMR spectra of* ***112b***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

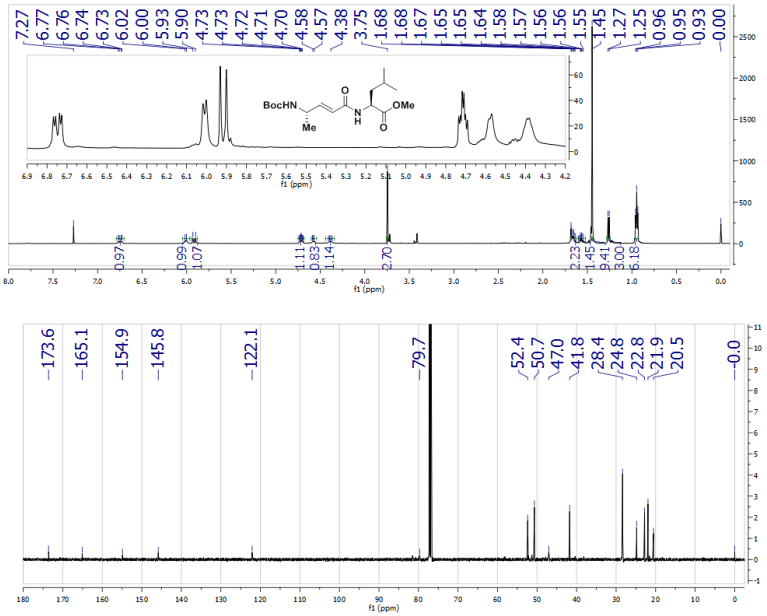


Figure C.9: ^{1}H *and* ^{13}C*-NMR spectra of* ***112c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

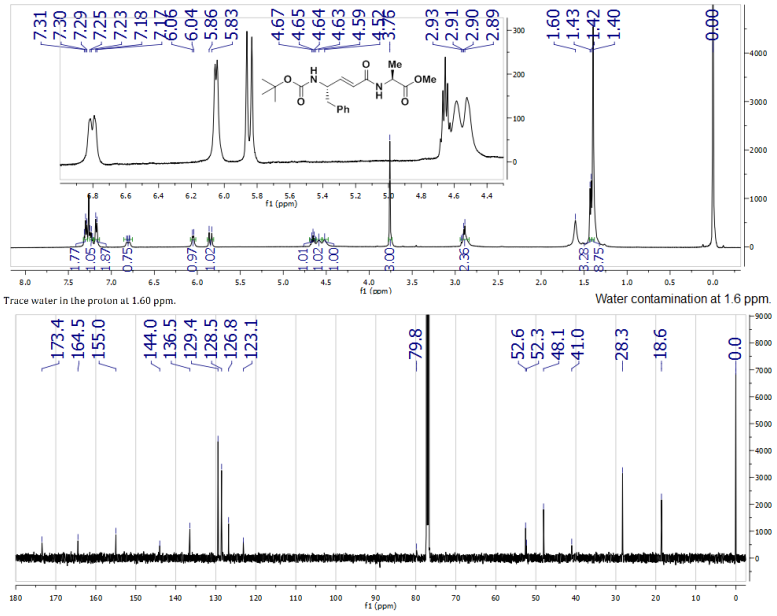


Figure C.10: ^{1}H *and* ^{13}C*-NMR spectra of* ***112d***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

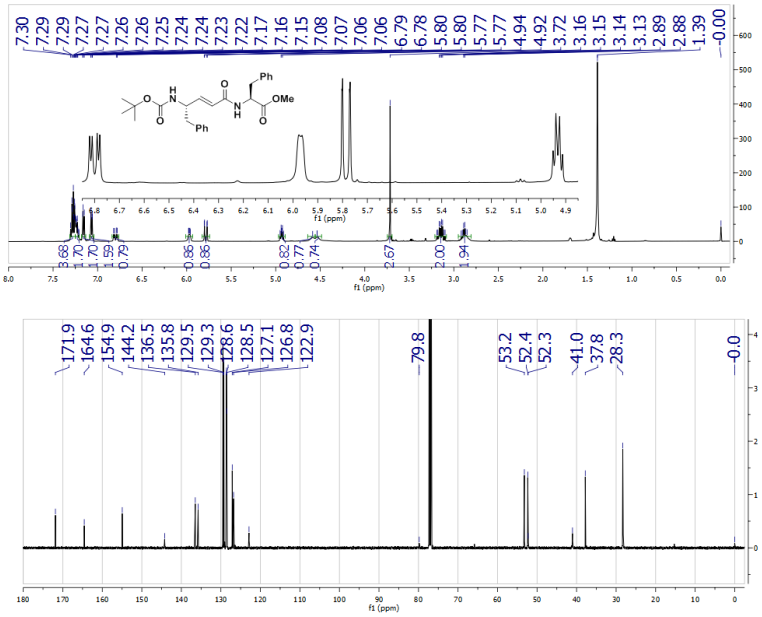


Figure C.11: ^{1}H *and* ^{13}C*-NMR spectra of* ***112e***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

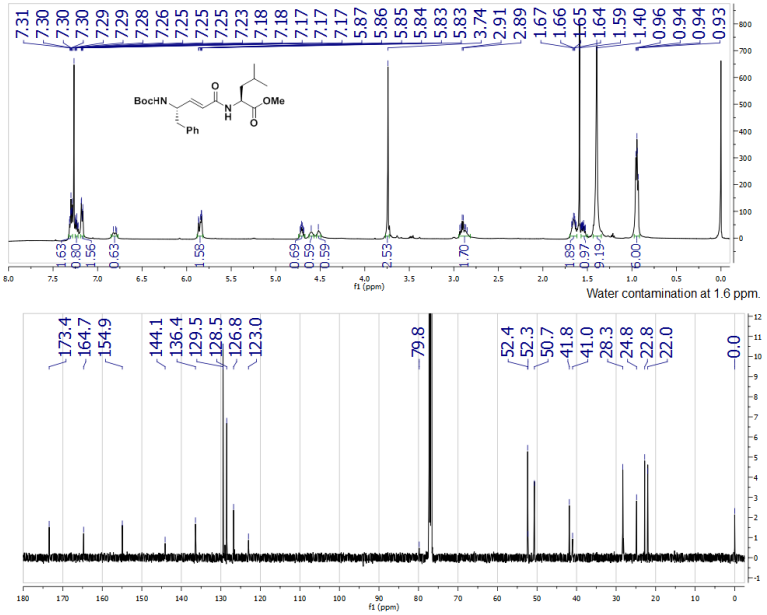


Figure C.12: ^{1}H *and* ^{13}C*-NMR spectra of* ***112f***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

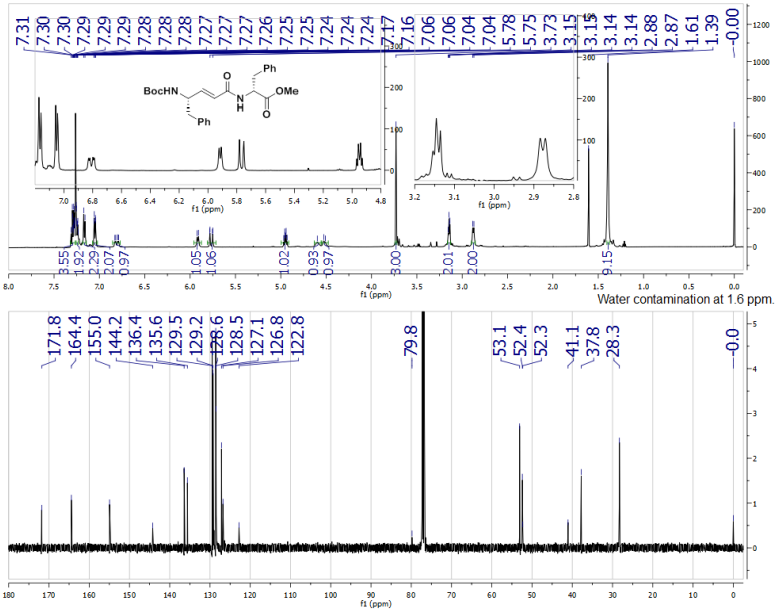


Figure C.13: ^{1}H *and* ^{13}C*-NMR spectra of* ***112g***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

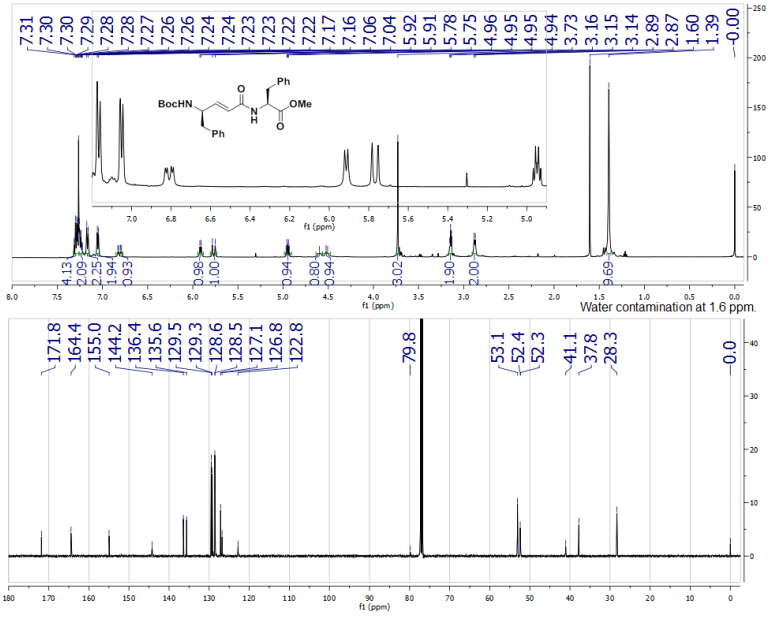


Figure C.14: ^{1}H *and* ^{13}C*-NMR spectra of* ***112h***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

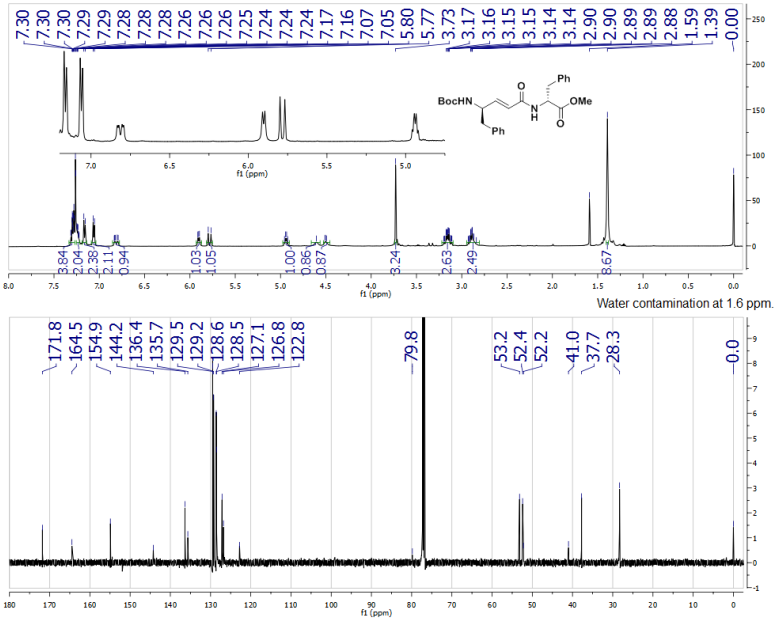


Figure C.15: ^{1}H *and* ^{13}C*-NMR spectra of* ***112i***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

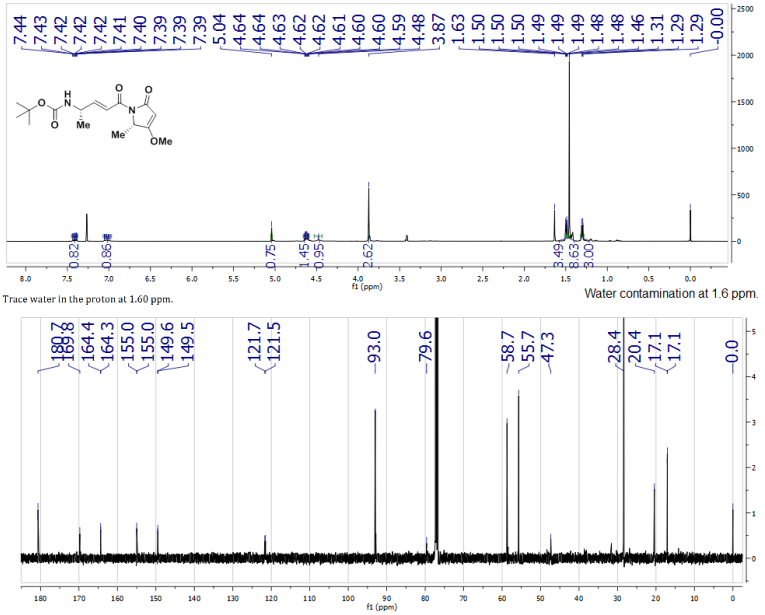


Figure C.16: ^{1}H *and* ^{13}C*-NMR spectra of* ***113a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

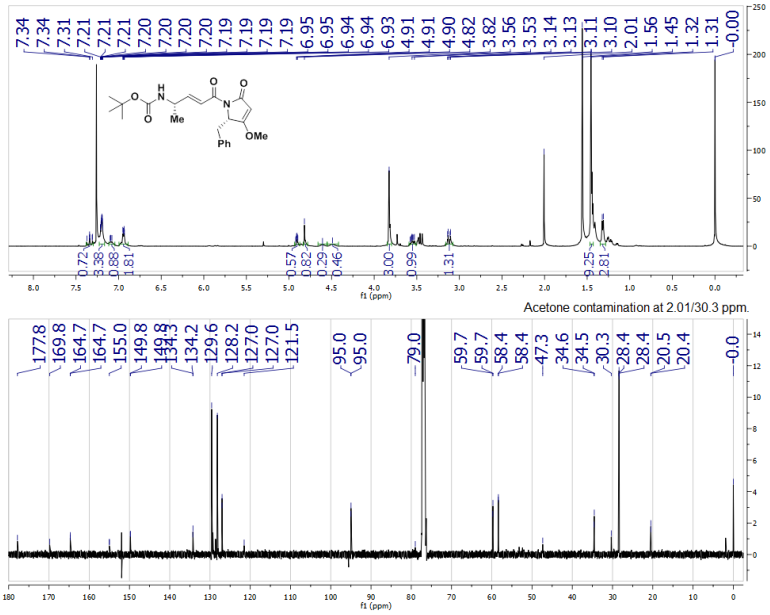


Figure C.17: ^{1}H *and* ^{13}C*-NMR spectra of* ***113b***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

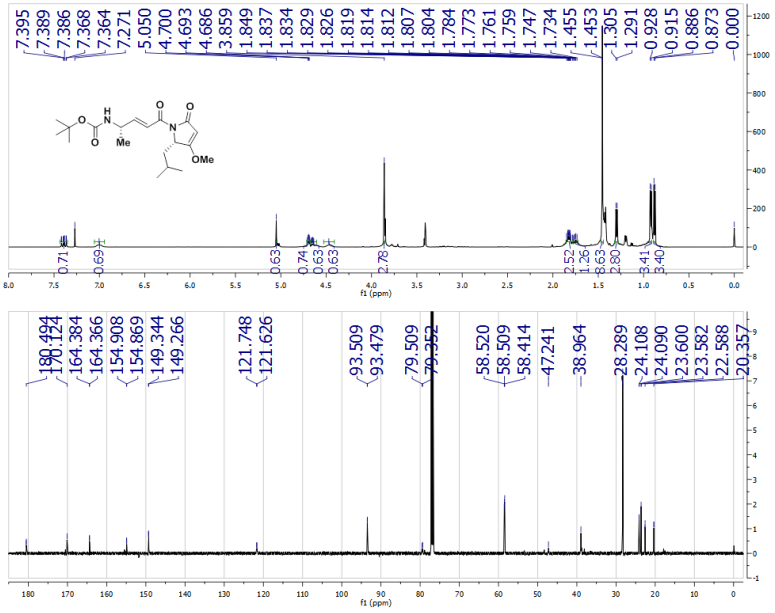


Figure C.18: ^{1}H *and* ^{13}C*-NMR spectra of* ***113c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

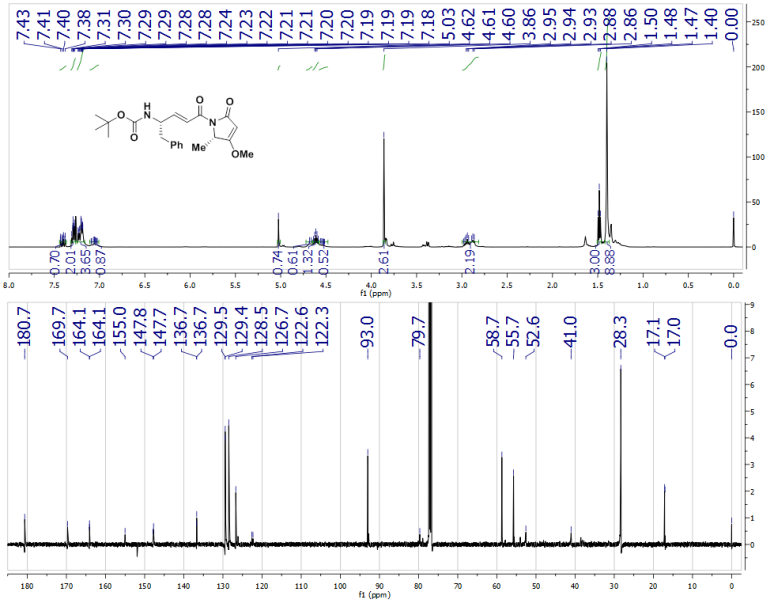


Figure C.19: ^{1}H *and* ^{13}C*-NMR spectra of* ***113d***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

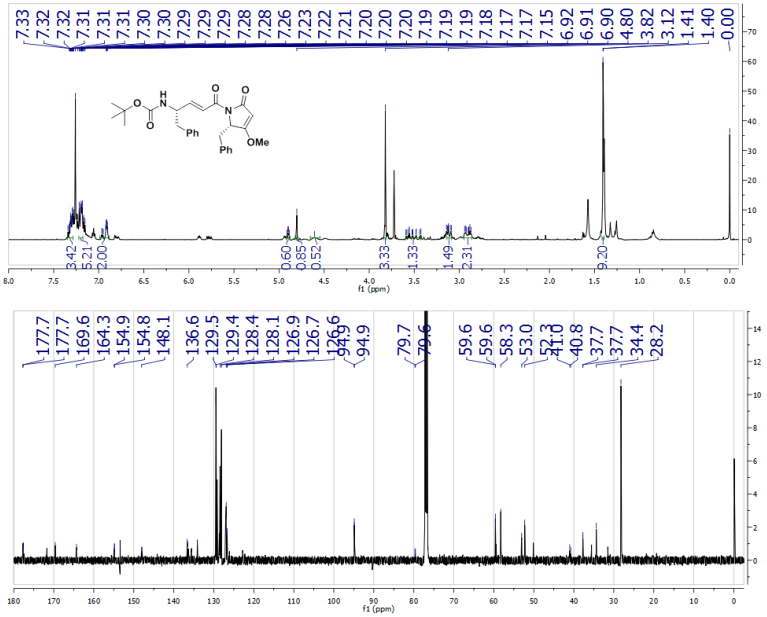


Figure C.20: ^{1}H *and* ^{13}C*-NMR spectra of* ***113e***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

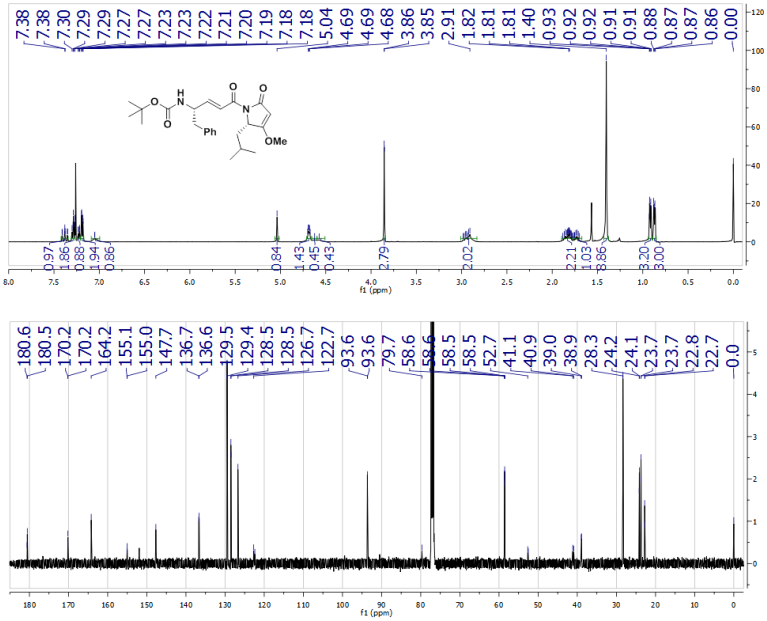


Figure C.21: ^{1}H *and* ^{13}C*-NMR spectra of* ***113f***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

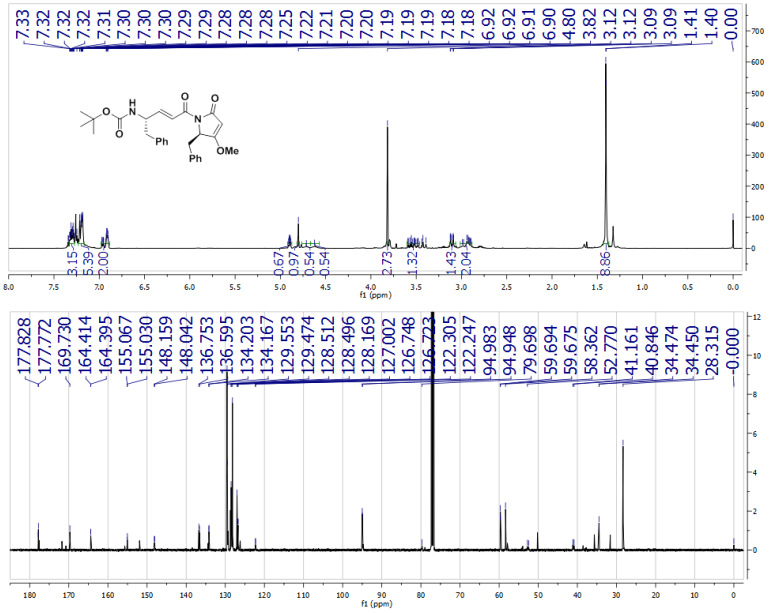


Figure C.22: ^{1}H *and* ^{13}C*-NMR spectra of* ***113g***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

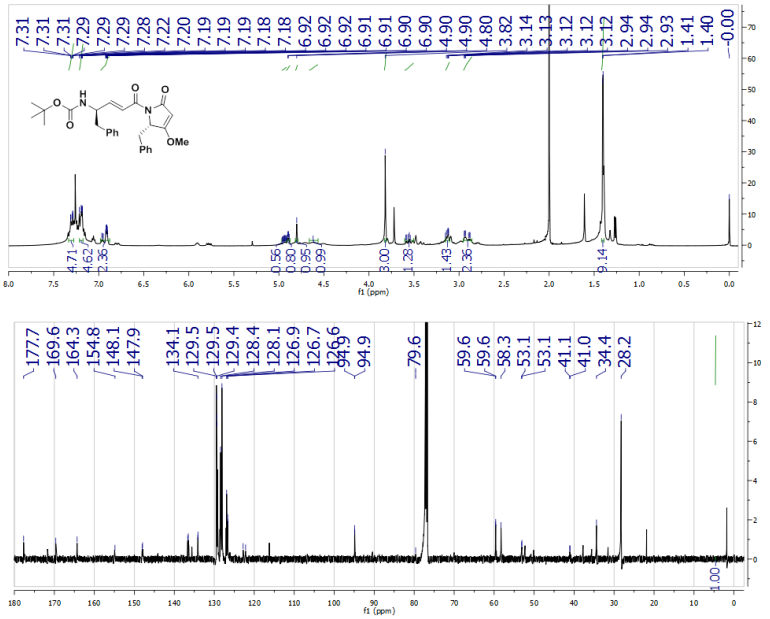


Figure C.23: ^{1}H *and* ^{13}C*-NMR spectra of* ***113h***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

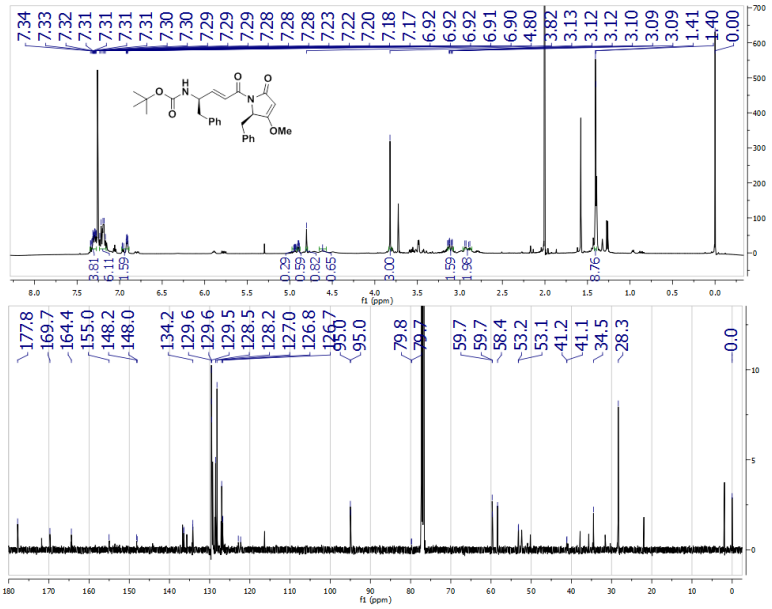


Figure C.24: ^{1}H *and* ^{13}C*-NMR spectra of* ***113i****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

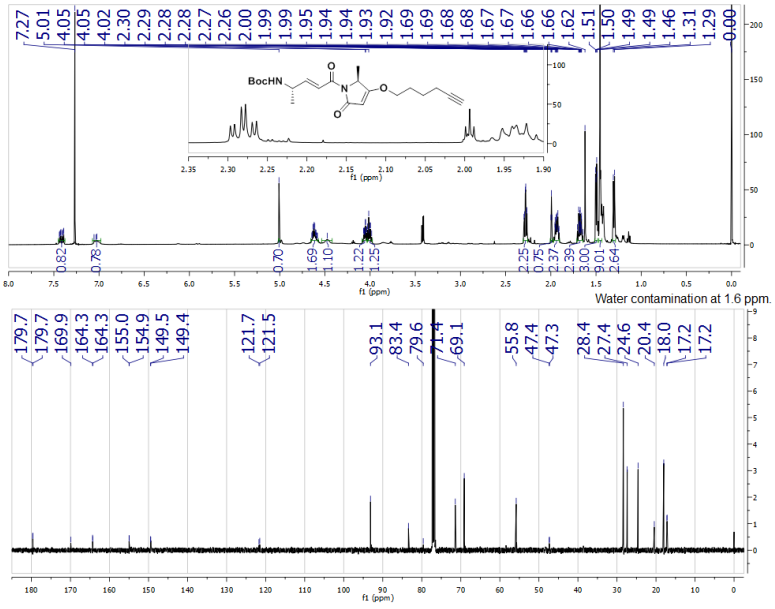


Figure C.25: ^{1}H *and* ^{13}C*-NMR spectra of* ***113j****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

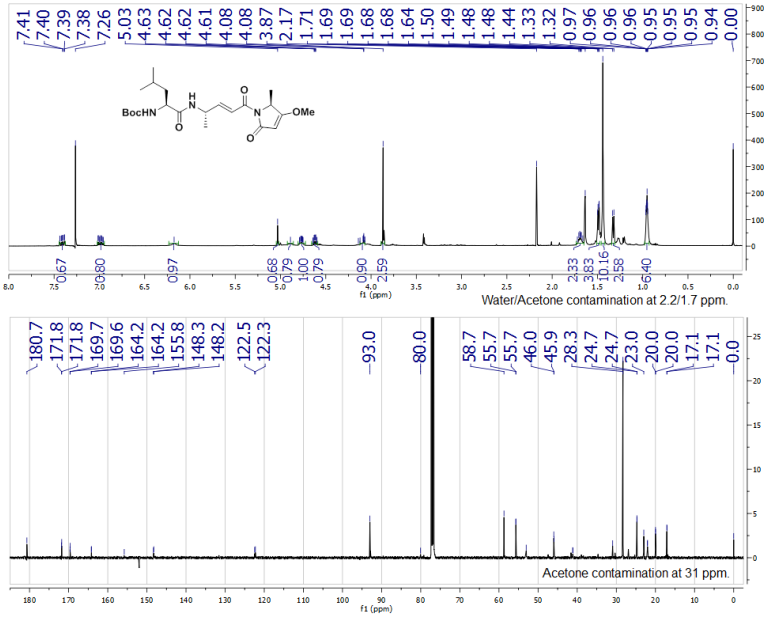


Figure C.26: ^{1}H *and* ^{13}C*-NMR spectra of* ***117a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

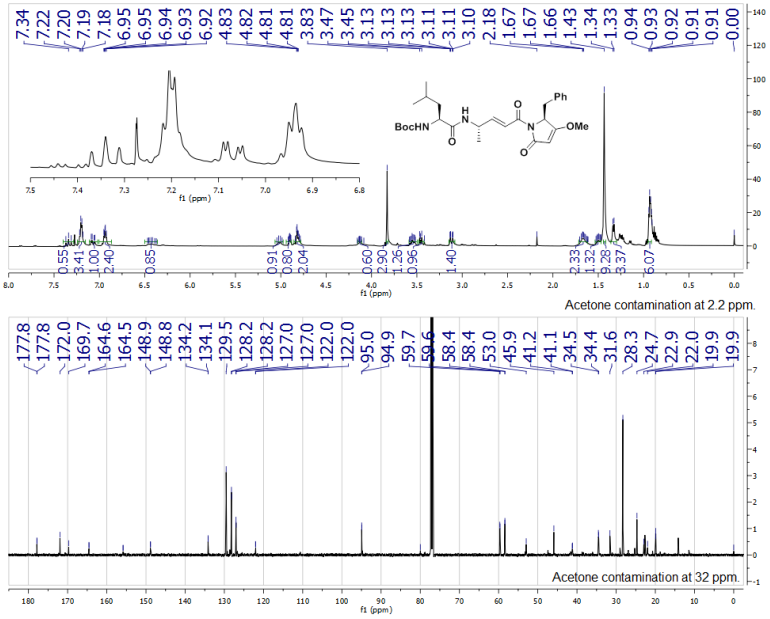


Figure C.27: ^{1}H *and* ^{13}C*-NMR spectra of* ***117b***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

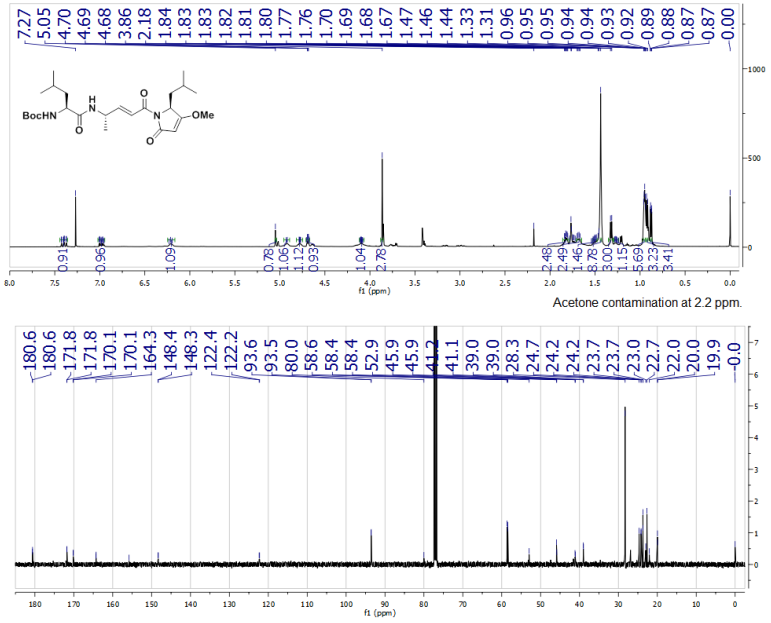


Figure C.28: ^{1}H *and* ^{13}C*-NMR spectra of* ***117c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

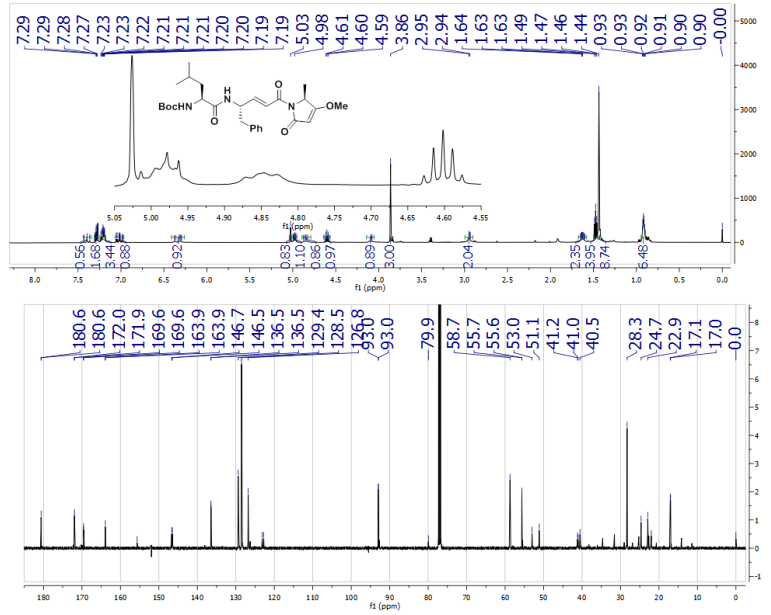


Figure C.29: ^{1}H *and* ^{13}C*-NMR spectra of* ***117d***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

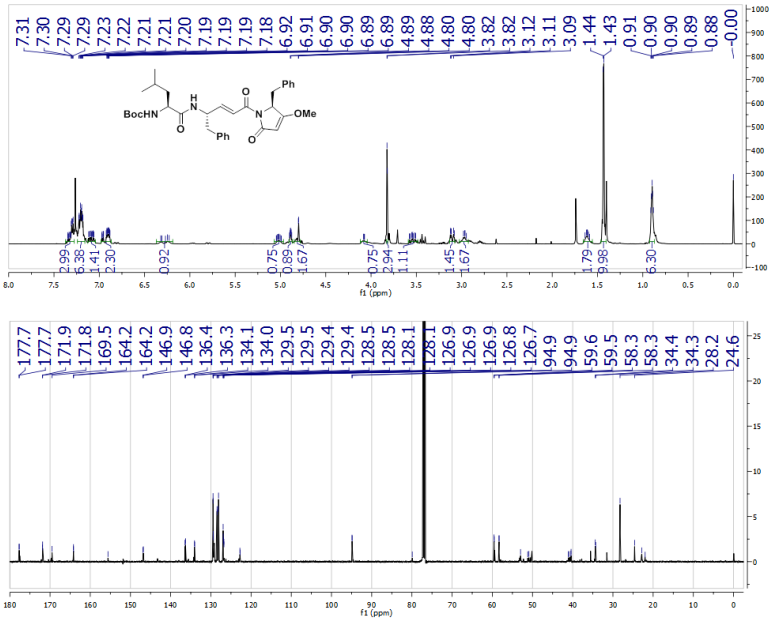


Figure C.30: ^{1}H *and* ^{13}C*-NMR spectra of* ***117e***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

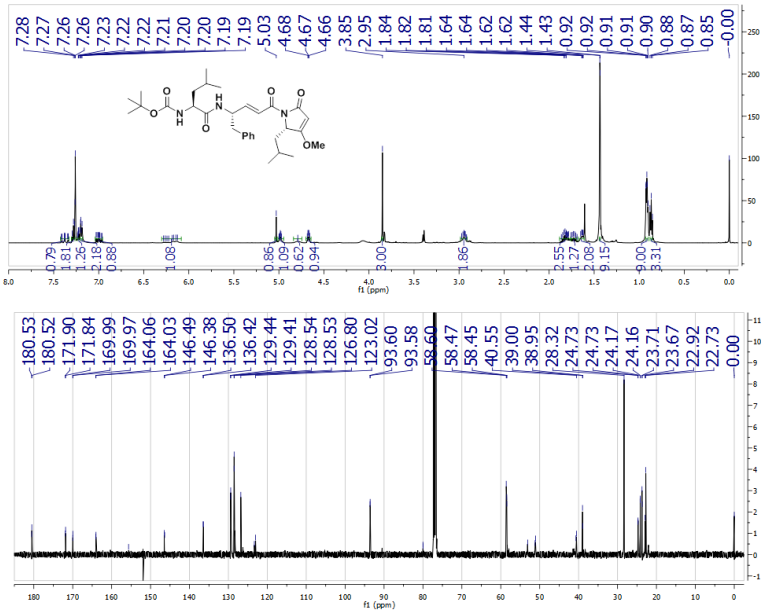


Figure C.31: ^{1}H *and* ^{13}C*-NMR spectra of* ***117f***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

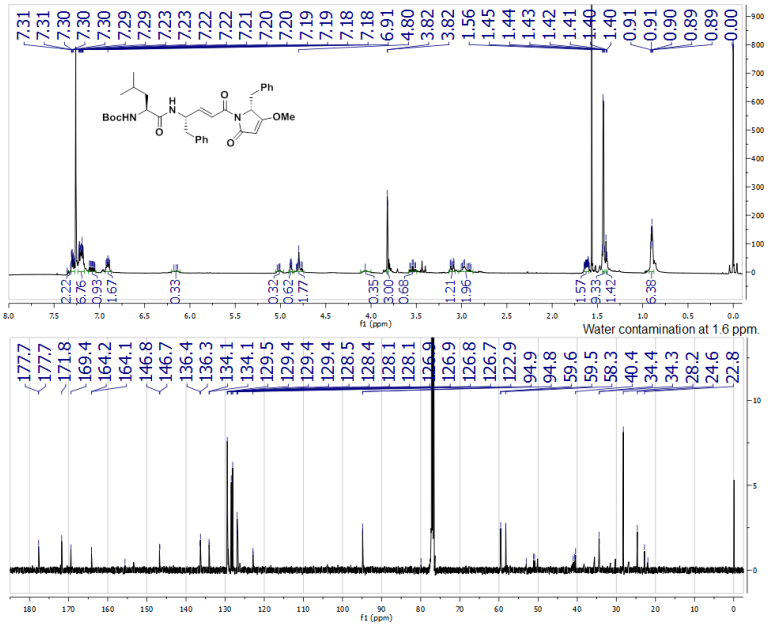


Figure C.32: ^{1}H *and* ^{13}C*-NMR spectra of* ***117g***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

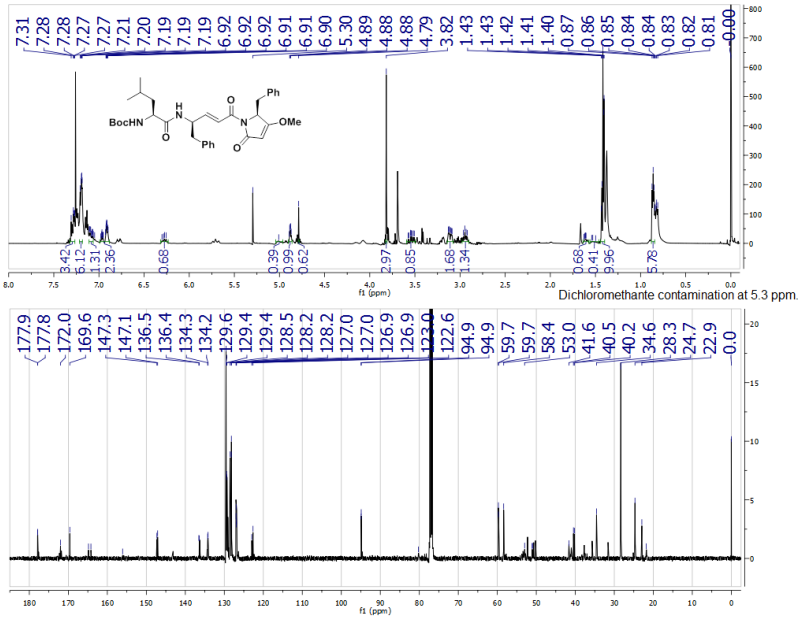


Figure C.33: ^{1}H *and* ^{13}C*-NMR spectra of* ***117h***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

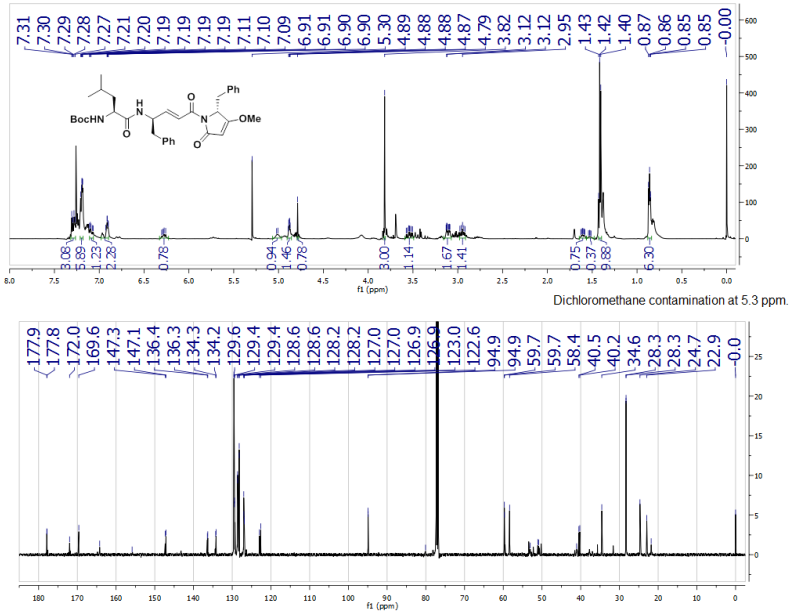


Figure C.34: ^{1}H *and* ^{13}C*-NMR spectra of* ***117i****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

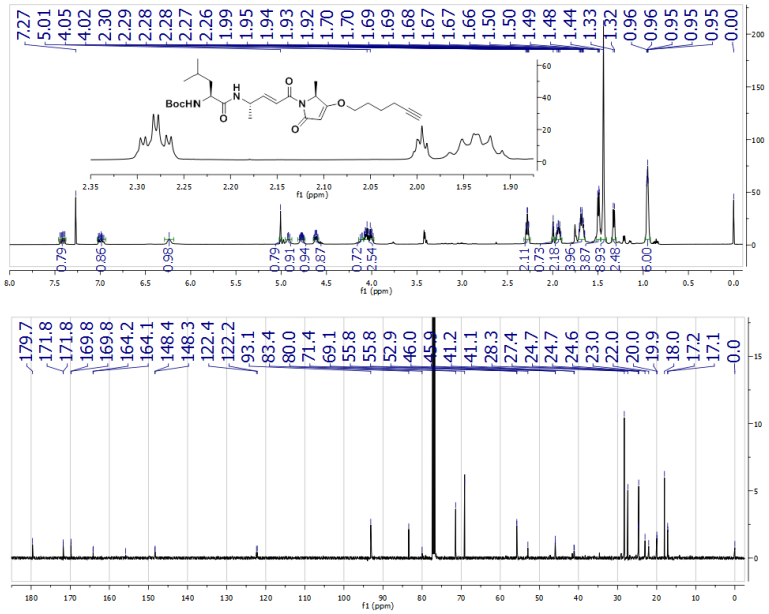


Figure C.35: ^{1}H *and* ^{13}C*-NMR spectra of* ***117j****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

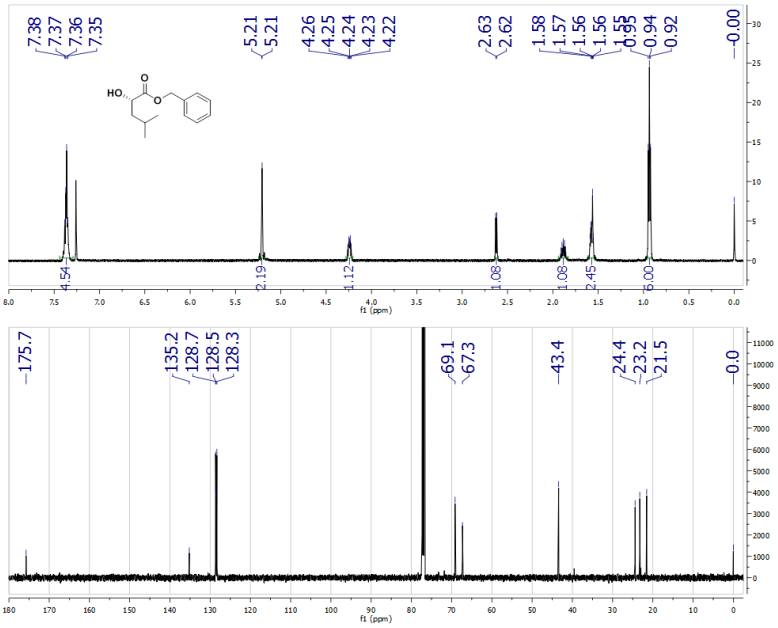


Figure C.36: ^{1}H *and* ^{13}C*-NMR spectra of* ***113***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

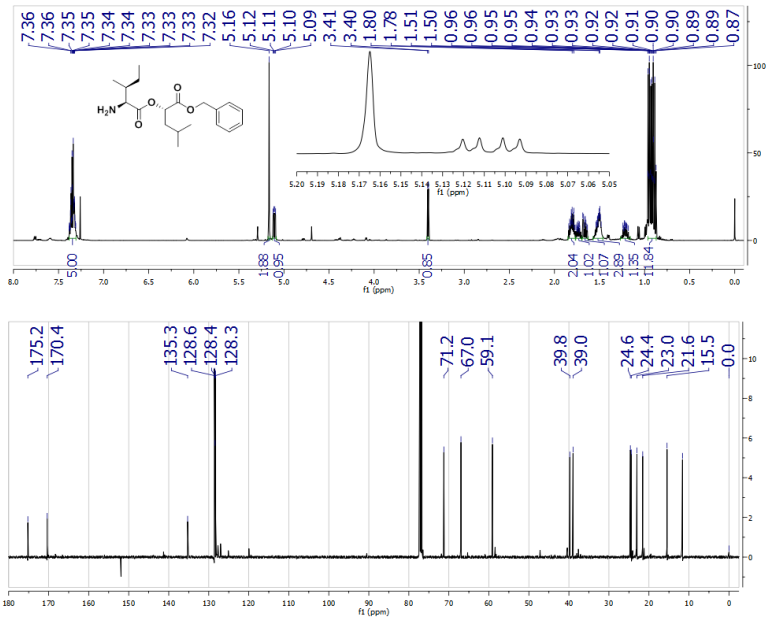


Figure C.37: ^{1}H *and* ^{13}C*-NMR spectra of* ***115a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

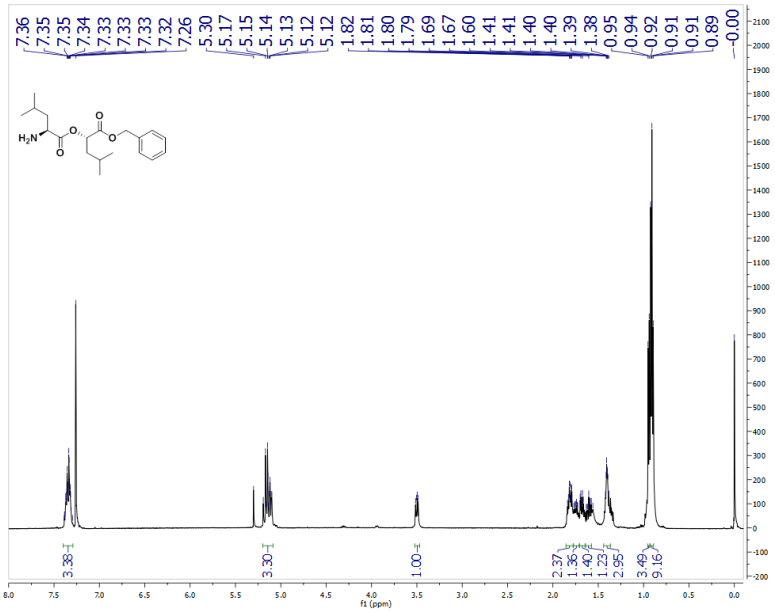


Figure C.38: ^{1}H *spectrum of* ***115b***. Spectrum collected in $CDCl_3$ with 0.03% *v/v* TMS on a 500 MHz probe.

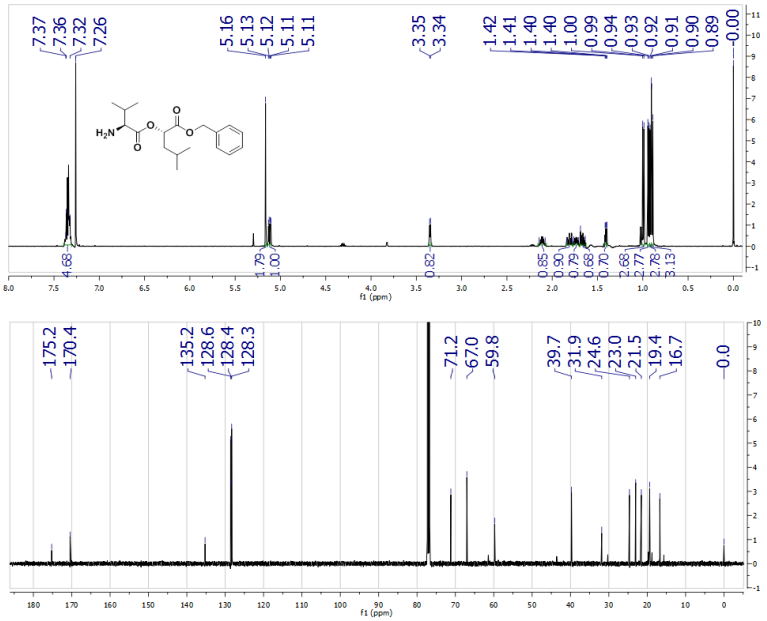


Figure C.39: ^{1}H *and* ^{13}C*-NMR spectra of* ***115c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

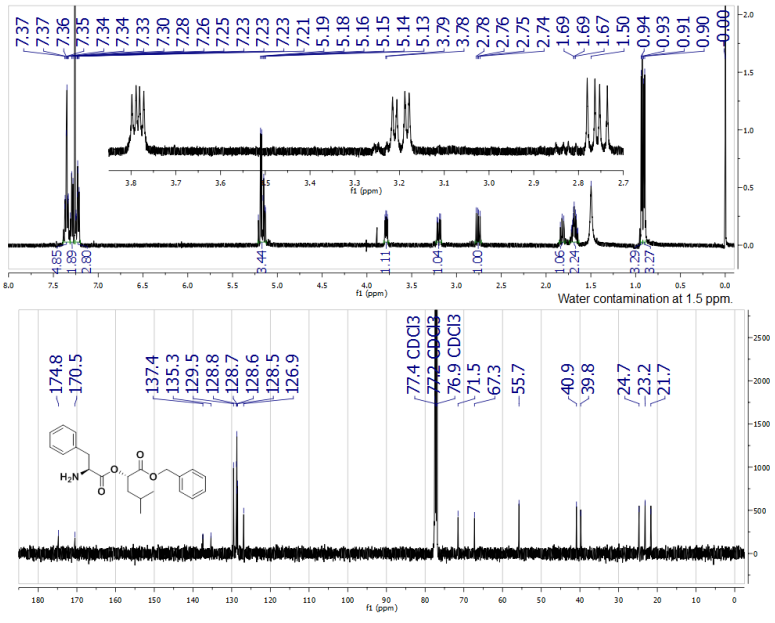


Figure C.40: ^{1}H *and* ^{13}C*-NMR spectra of* ***115d****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

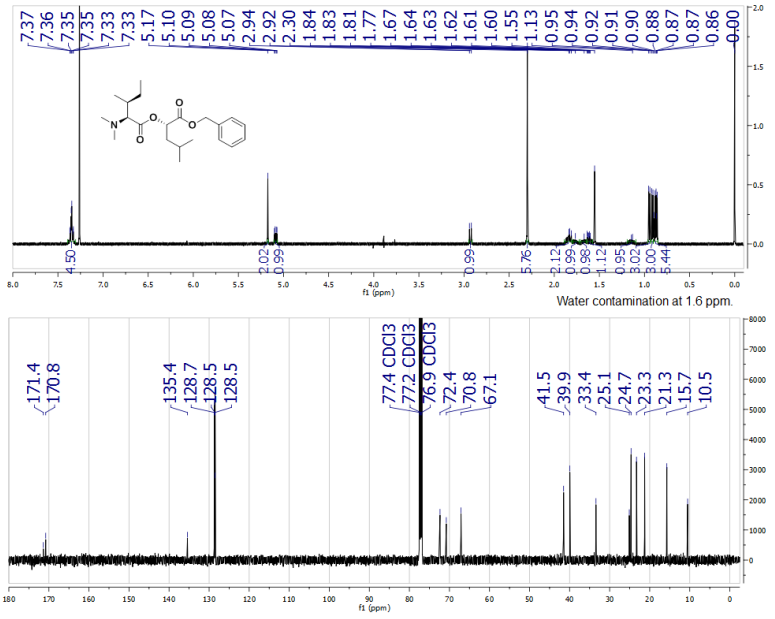


Figure C.41: ^{1}H *and* ^{13}C*-NMR spectra of* ***115a****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

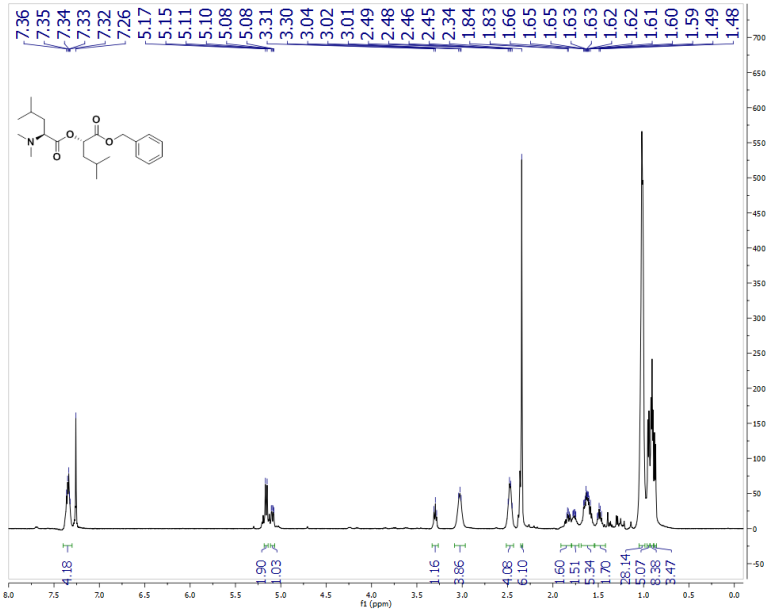


Figure C.42: ^{1}H *spectrum of* ***115b***. Spectrum collected in $CDCl_3$ with 0.03% *v/v* TMS on a 500 MHz probe.

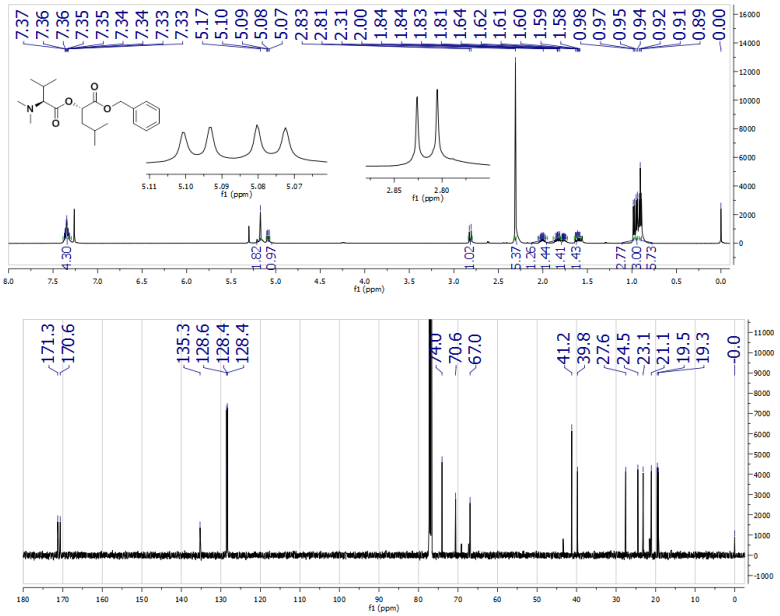


Figure C.43: ^{1}H *and* ^{13}C*-NMR spectra of* ***116c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

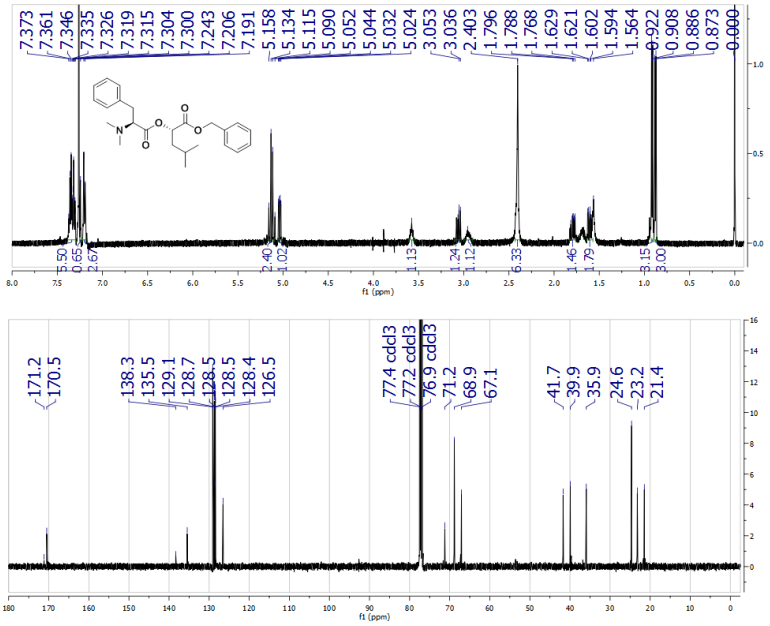


Figure C.44: ^{1}H *and* ^{13}C*-NMR spectra of* ***116d***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe.

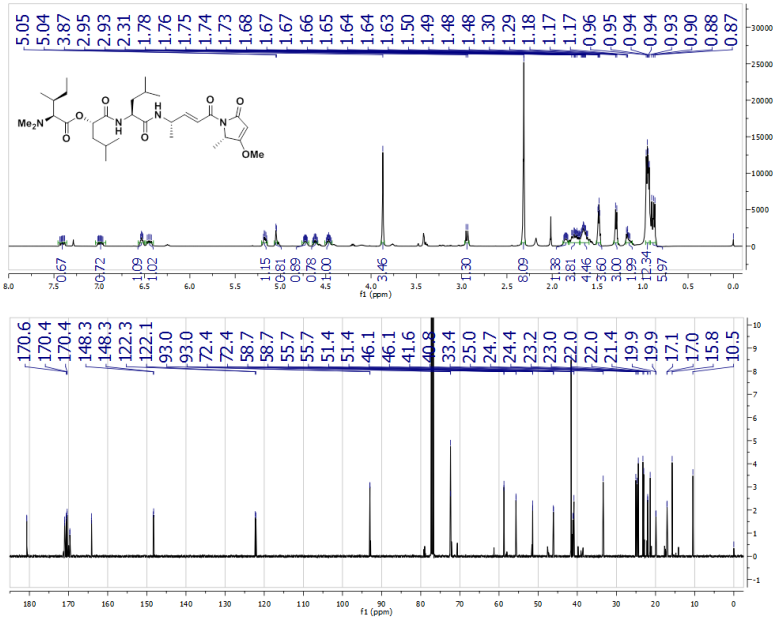


Figure C.45: ^{1}H *and* ^{13}C*-NMR spectra of* ***106***. Spectra collected in $CDCl_3$ on an inverse 500 MHz cryo-probe at 30 °C.

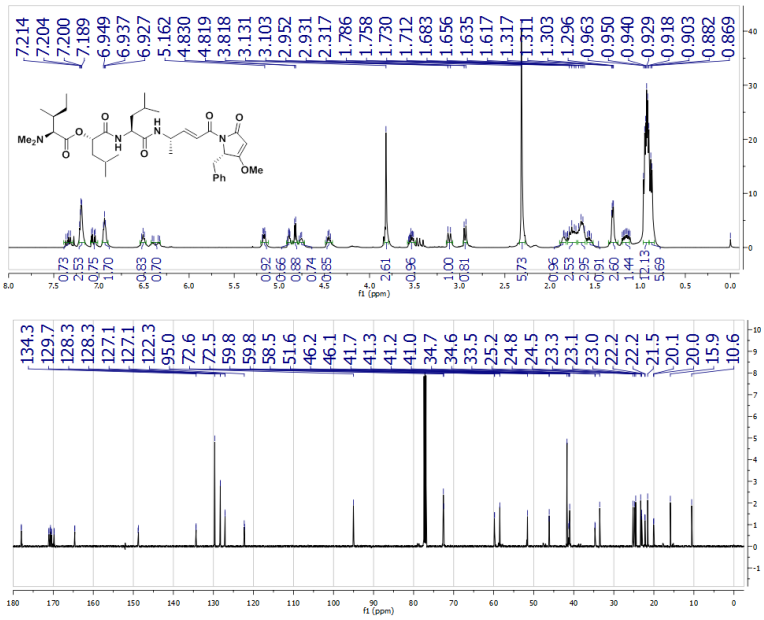


Figure C.46: ^{1}H *and* ^{13}C*-NMR spectra of* ***118a***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on a 500 MHz probe for the proton, and an inverse 500 MHz cryo-probe at 30 °C for the carbon.

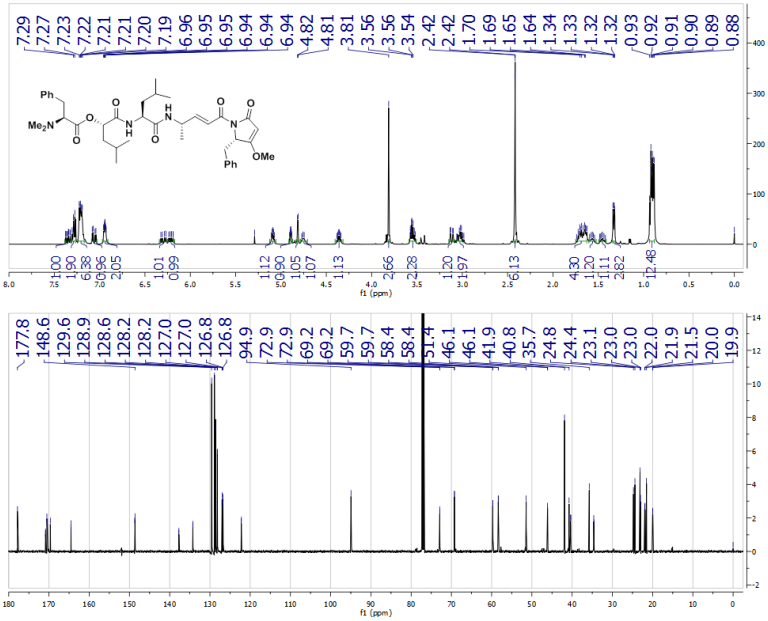


Figure C.47: ^{1}H *and* ^{13}C*-NMR spectra of* ***118b***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

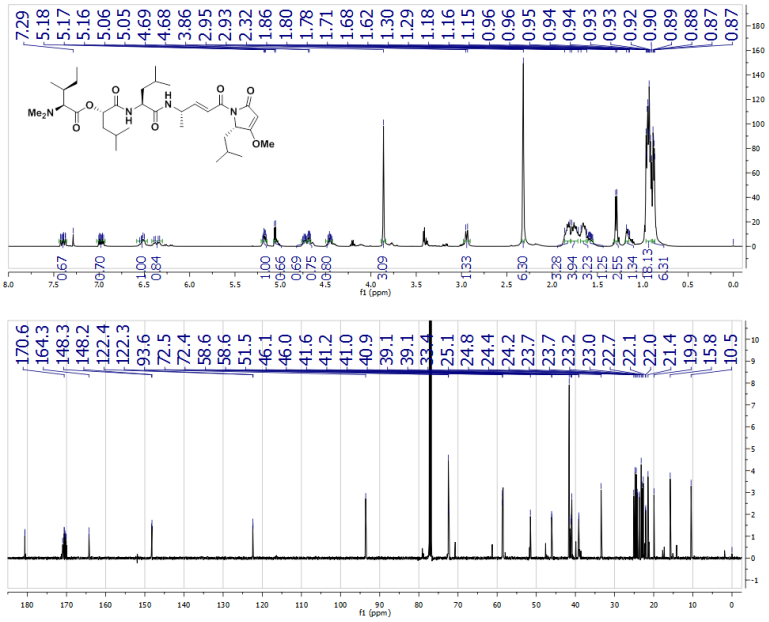


Figure C.48: ^{1}H *and* ^{13}C*-NMR spectra of* ***118c***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

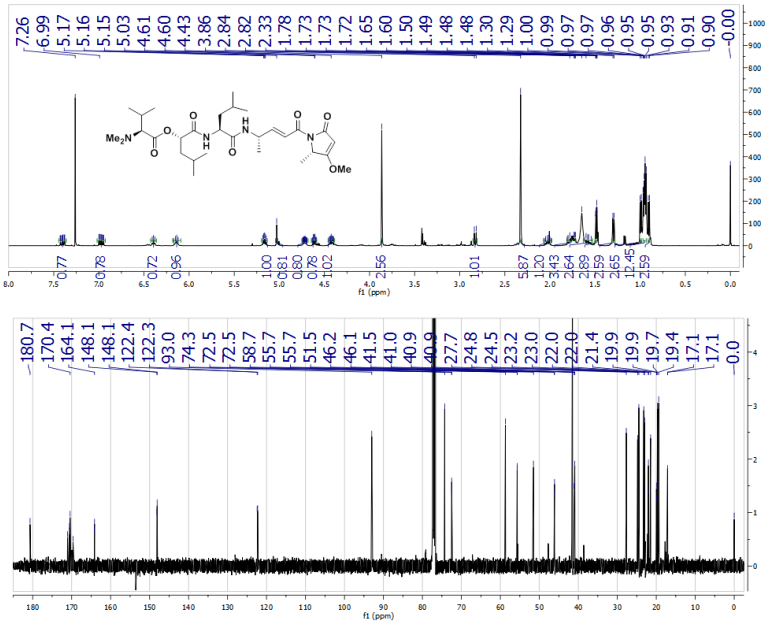


Figure C.49: ^{1}H *and* ^{13}C*-NMR spectra of* ***118d***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

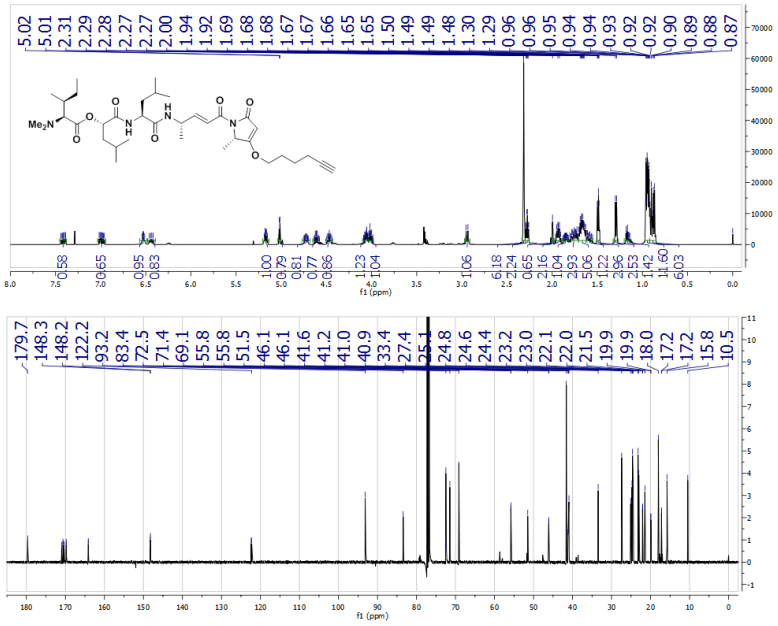


Figure C.50: ^{1}H *and* ^{13}C*-NMR spectra of* ***118e***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS, on a 500 MHz probe for the proton, and on an inverse 500 MHz cryo-probe at 30 °C for the carbon.

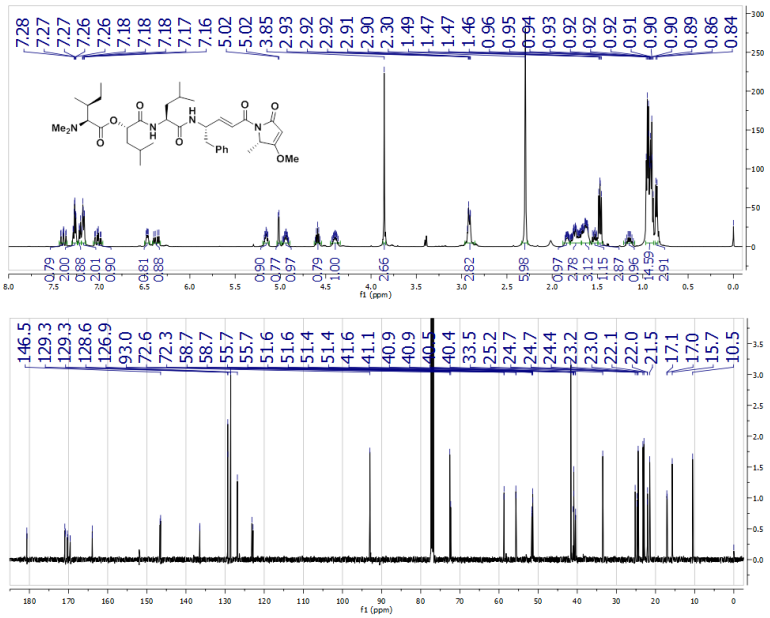


Figure C.51: ^{1}H *and* ^{13}C*-NMR spectra of* ***118f***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

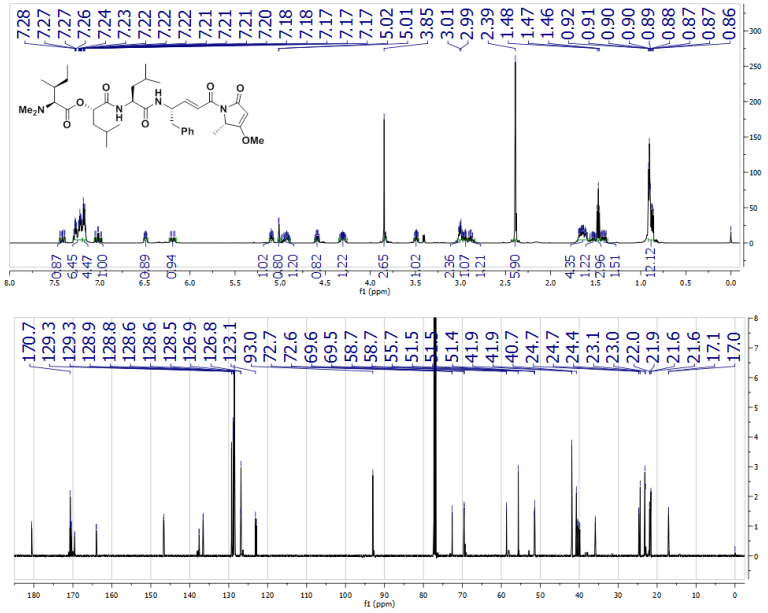


Figure C.52: ^{1}H *and* ^{13}C*-NMR spectra of* ***118g***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

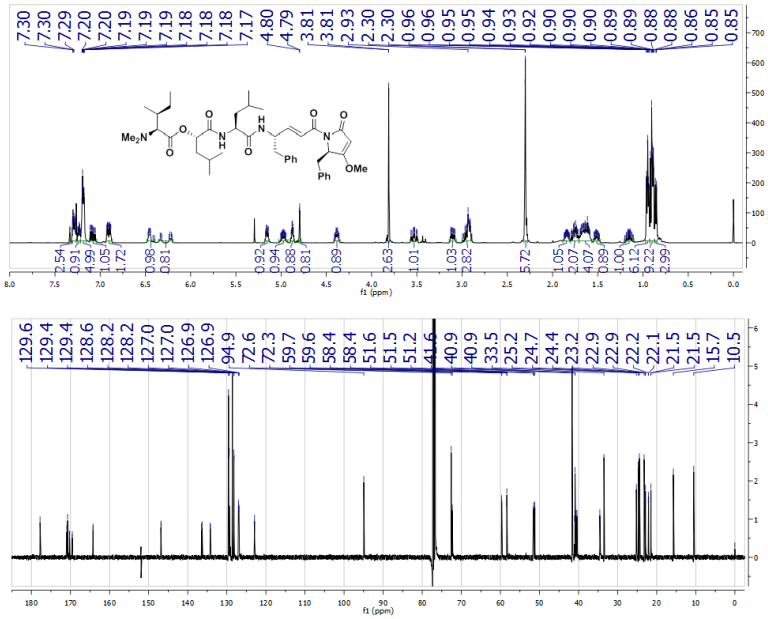


Figure C.53: ^{1}H *and* ^{13}C*-NMR spectra of* ***118h***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

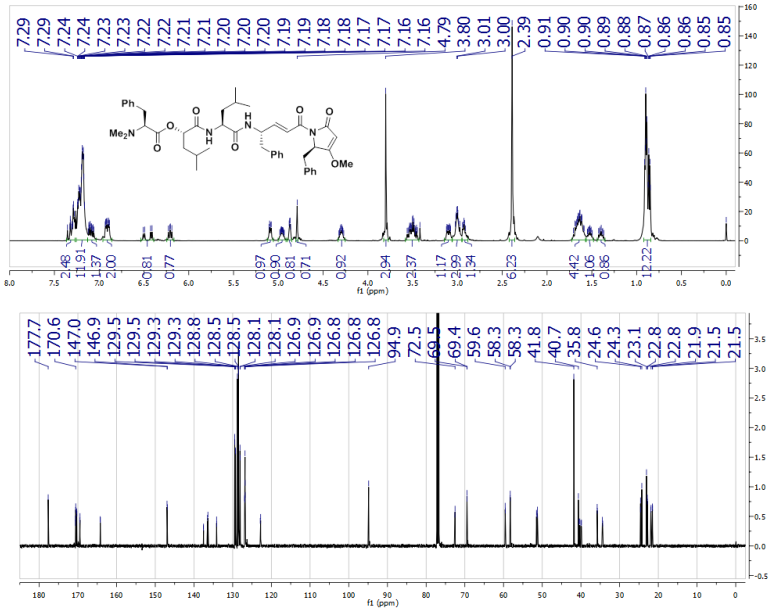


Figure C.54: ^{1}H *and* ^{13}C*-NMR spectra of* ***118i****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

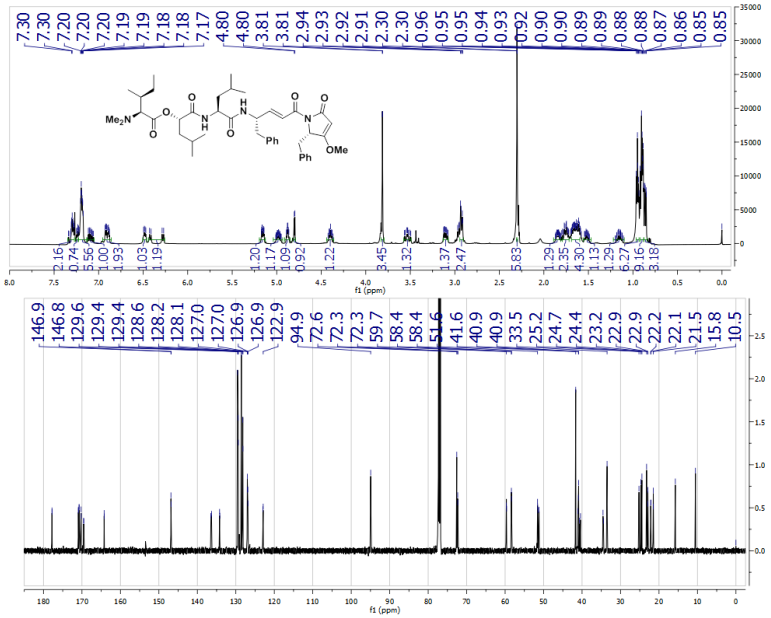


Figure C.55: ^{1}H *and* ^{13}C*-NMR spectra of* ***118j****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

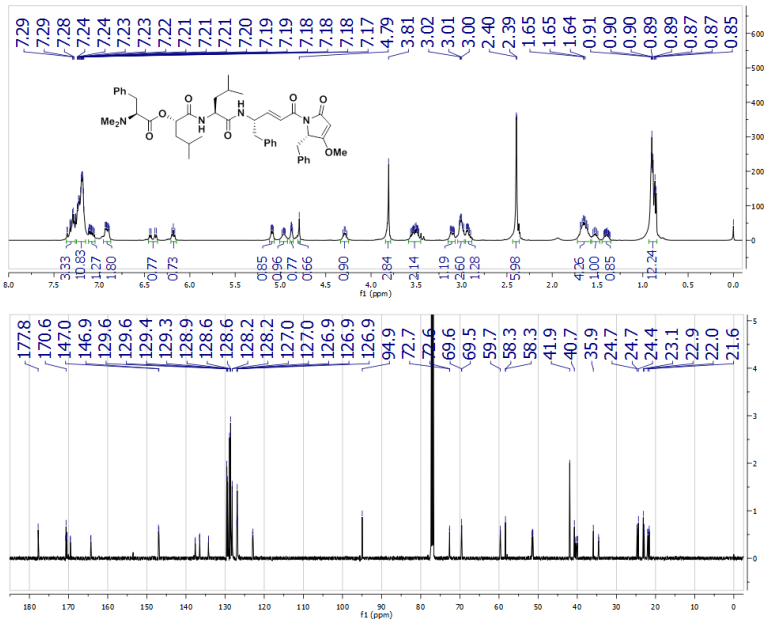


Figure C.56: ^{1}H *and* ^{13}C*-NMR spectra of* ***118k****.* Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

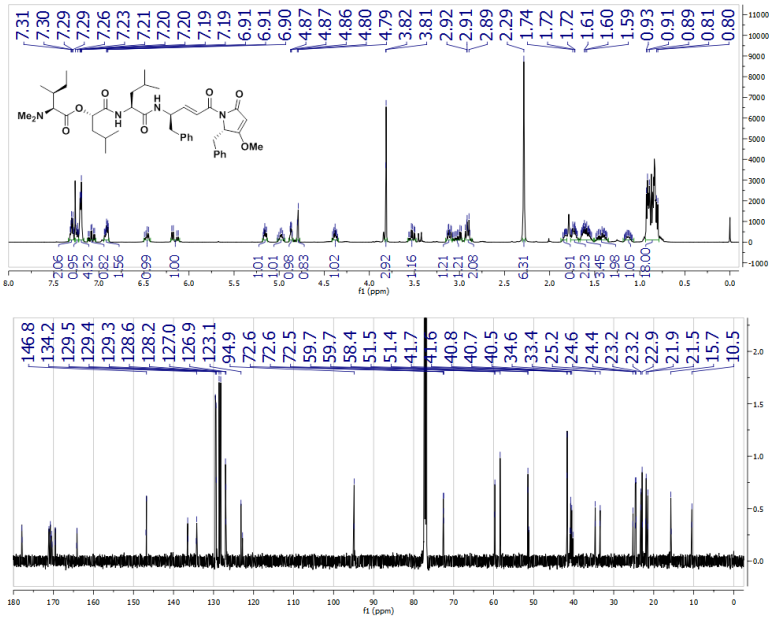


Figure C.57: ^{1}H *and* ^{13}C*-NMR spectra of* ***118l****.* Spectra collected in $CDCl_3$, on an inverse 500 MHz cryo-probe at 30 °C.

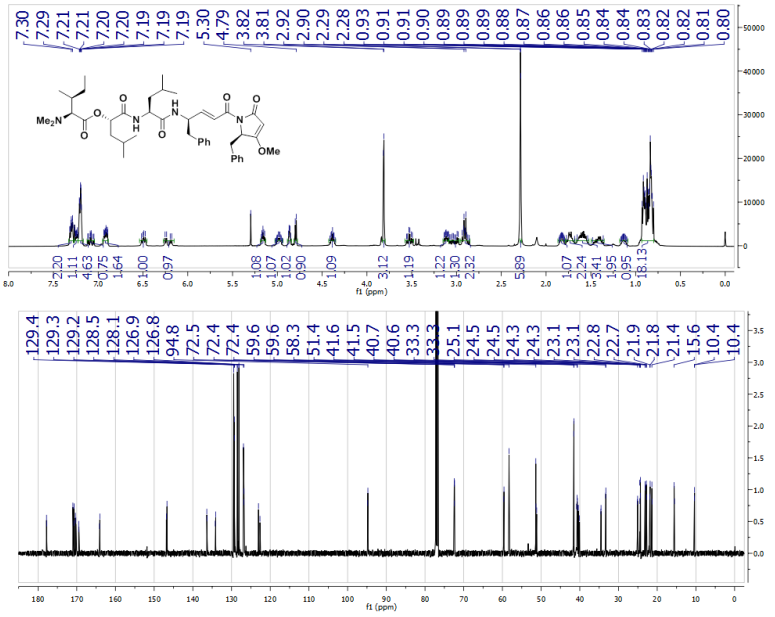


Figure C.58: ^{1}H *and* ^{13}C*-NMR spectra of* ***118m***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS, on a 500 MHz probe for the proton and an inverse 500 MHz cryo-probe at 30 °C for the carbon.

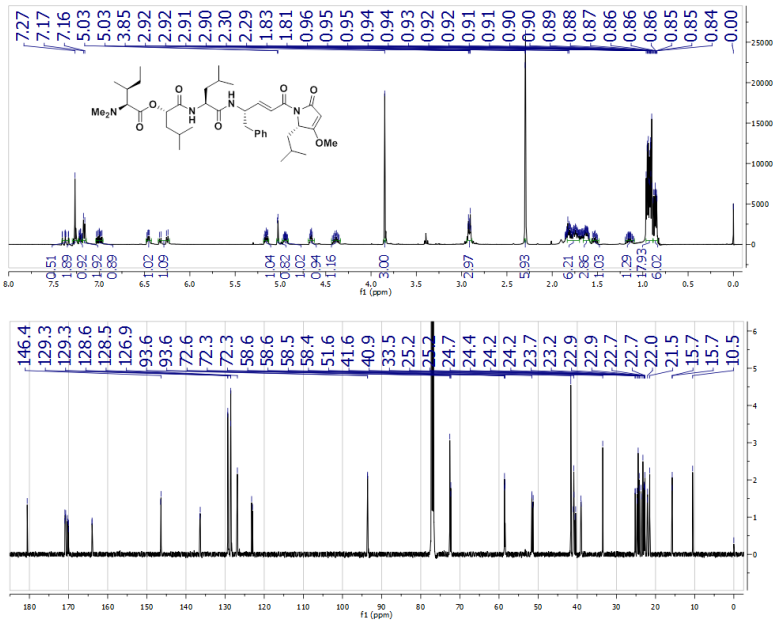


Figure C.59: ^{1}H *and* ^{13}C*-NMR spectra of* ***118n***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS, on a 500 MHz probe for the proton and an inverse 500 MHz cryo-probe at 30 °C for the carbon.

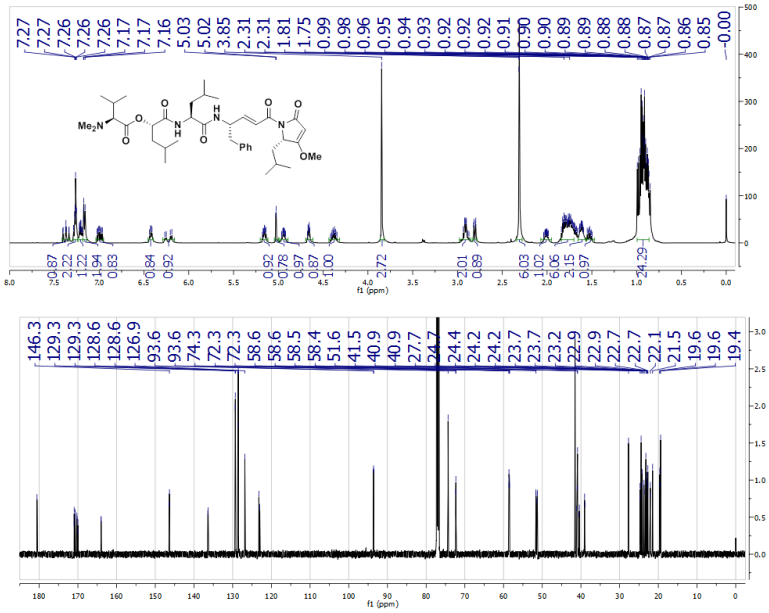


Figure C.60: ^{1}H *and* ^{13}C*-NMR spectra of* ***118o***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

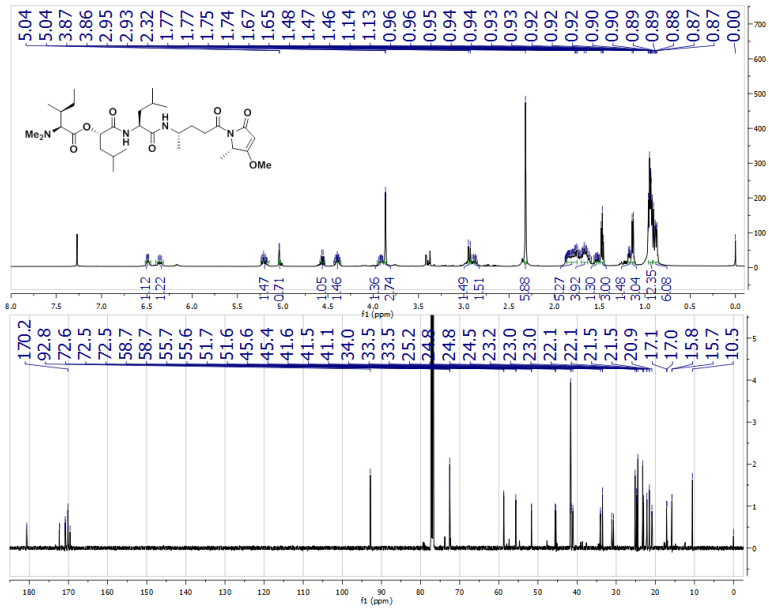


Figure C.61: ^{1}H *and* ^{13}C*-NMR spectra of* ***119***. Spectra collected in $CDCl_3$ with 0.03% *v/v* TMS on an inverse 500 MHz cryo-probe at 30 °C.

Bibliography

[ant, 2015] (2015). Anitbiotics & secondary metabolite analysis shell. http://antismash.secondarymetabolites.org/. Accessed: 2015-01-09.

[Cyt, 2015] (2015). Cytoscape. http://www.cytoscape.org/. Accessed: 2015-01-03.

[GNP, 2015] (2015). Gnps: Global natural products social molecular networking. http://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp. Accessed: 2015-07-23.

[NRP, 2015] (2015). Nrps predictor 2. http://nrps.informatik.uni-tuebingen.de. Accessed: 2015-07-25.

[Ska, 2015] (2015). Uc san diego skaggs school of pharmacy and pharmaceutical sciences nmr facility. http://sopnmr.ucsd.edu/. Accessed: 2015-07-25.

[Aelterman et al., 1997] Aelterman, W., De Kimpe, N., and Kalinin, V. (1997). One-step synthesis of laurencione. *Journal of Natural Products*, 60:385–386.

[Ainslie et al., 1985] Ainslie, R. D., Barchi, J. J., Kuniyoshi, M., Moore, R. E., and Mynderse, J. S. (1985). Structure of malyngamide c. *Journal of Organic Chemistry*, 50:2859–2862.

[Andrianasolo et al., 2011] Andrianasolo, E. H., Haramaty, L., McPhail, K. L., White, E., Vetriani, C., Falkowski, P., and Lutz, R. (2011). Bathymodiolamides a and b, ceramide derivatives from a deep-sea hydrothermal vent inverebrate mussel, *Bathymodiolus thermophilus*. *Journal of Natural Products*, 74:842–846.

[Aue et al., 1976] Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976). Two-dimensional spectroscopy. application to nuclear magnetic resonance. *Journal of Chemical Physics*, 64(5):2229–2246.

[Bernart et al., 1992] Bernart, M. W., Gerwick, W. H., Corcoran, E. E., Lee, A. Y., and Clardy, J. (1992). Laurencione, a heterocycle from the red alga *Laurencia spectabilis*. *Phytochemistry*, 31:1273–1276.

[Berrendero et al., 2011] Berrendero, E., Perona, E., and Mateo, P. (2011). Phenotypic variability and phylogenetic relationships of the genera *Tolypothrix* and *Calothrix* (nostocales, cyanobacteria) from running water. *International Journal of Systematic and Evolutionary Microbiology*, 61:3039–3051.

[Bertin et al., 2015] Bertin, M., Schwartz, S. L., Lee, J., Korobeynikov, A., Dorrestein, P. C., Gerwick, L., and Gerwick, W. H. (2015). Spongosine production by a *Vibrio harveyi* strain associated with the sponge *Tectitethya crypta*. *Journal of Natural Products*, 78:493–499.

[Bhushan and Brückner, 2004] Bhushan, R. and Brückner, H. (2004). Marfey's reagent for chiral amino aci analysis: a review. *Amino Acids*, 27:231–247.

[Blasiak and L., 2009] Blasiak, L. C. and L., D. C. (2009). Structural perspective on enzymatic halogenation. *Accounts of Chemical Research*, 42:147–155.

[Blin et al., 2013] Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013). antismash 2.0a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, 41:W204–W212.

[Blunt et al., 2013] Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., and Prinsep, M. R. (2013). Marine natural products. *Natural Product Reports*, 30:237–323.

[Blunt et al., 2014] Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., and Prinsep, M. R. (2014). Marine natural products. *Natural Product Reports*, 31:160–258.

[Boudreau et al., 2015] Boudreau, P. D., Monroe, E. A., Mehrotra, S., Desfor, S., Korobeynikov, A., Sherman, David H. Gerwick, L., Dorrestein, P. C., and Gerwick, W. H. (2015). Expanding the described metabolome of the marine cyanobacterium moorea producens jhb through orthogonal natural products workflows. *Public Library of Science Online Edition*, 10:e0133297.

[Bruland, 1983] Bruland, K. W. (1983). Chapter 45. In Riley, J. P. and Chester, R., editors, *Chemical Oceanography*, pages 172–173. Academic Press.

[Bunyajetpong et al., 2006] Bunyajetpong, S., Yoshida, W. Y., Sitachitta, N., and Kaya, K. (2006). Trungapeptins a-c, cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 69:1539–1542.

[Butler and Walker, 1993] Butler, A. and Walker, J. V. (1993). Marine haloperoxidases. *Chemistry Reviews*, 93:1937–1944.

[Cao et al., 2008] Cao, Z., George, J., Gerwick, W. H., Baden, D. G., Rainier, J. D., and Murray, T. F. (2008). Influence of lipid-soluble gating modifier toxins on sodium influx in neocortical neurons. *Journal of Pharmacology and Experimental Therapeutics*, 326:1681–1691.

[Cao et al., 2011] Cao, Z., Shafer, T. J., and Murray, T. F. (2011). Mechanisms of pyrthroid insecticide-induced stimulation of calcium influx in neocortical neurons. *Journal of Pharmacology and Experimental Therapeutics*, 336:197–205.

[Carey and Sundberg, 2007] Carey, F. A. and Sundberg, R. J. (2007). *Advanced Organic Chemistry 5th Edition, Part A: Structure and Mechanisms*. Springer.

[Castenholz et al., 2001] Castenholz, R. W., Rippka, R., and Herdman, M. (2001). *Manual of Systematic Bacteriology*. Springer.

[Challis, 2008] Challis, G. L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology*, 154:1555–1569.

[Challis and Ravel, 2000] Challis, G. L. and Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome : structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiology Letters*, 187:111–114.

[Chang et al., 2002] Chang, Z., Flatt, P., Gerwick, W. H., Nguyen, V. A., Willis, C. L., and Sherman, D. H. (2002). The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (pks)-non-ribosomal peptide synthetase (nrps) origin involving an unusual trichloroleucyl starter unit. *Gene*, 296:235–247.

[Conroy et al., 2010] Conroy, T., Guo, J. T., Hunt, N. H., and Payne, R. J. (2010). Total synthesis and antimalarial activity of symplostatin 4. *Organic Letters*, 12:5576–5579.

[Conroy et al., 2011] Conroy, T., Guo, J. T., Linington, R. G., Hunt, N. H., and Payne, R. J. (2011). Total synthesis, stereochemical assignment, and antimalarial activity of gallinamide a. *Chemisty a European Journal*, 17:13544–13552.

[Conway et al., 2001] Conway, G. C., Smole, S. C., Sarracino, D. A., Arbeit, R. D., and Leopold, P. E. (2001). Phyloproteomics: Species identification of *Enterobacteriaceae* using matix-assissted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Molecular Microbiology and Biotechnology*, 3:103–112.

[Dalisay et al., 2009] Dalisay, D. S., Rogers, E. W., Edison, A. S., and Molinski, T. F. (2009). Structure elucidation at the nanomole scale. 1. trisoxazole macrolides and thiazole-containing cyclic peptides from the nudibrach *Hexabranchus sanguineus*. *Journal of Natural Products*, 72:732–738.

[Domínguez-Escobar et al., 2011] Domínguez-Escobar, J., Beltŕan, Y., Bergman, B., Díez, B., Ininbergs, K., Souza, V., and Falcón, L. I. (2011). Phylogenetic and molecular clock inferences of cyanobacterial strains within *Rivulariaceae* from distant environments. *FEMS Microbiology Letters*, 316:90–99.

[Dorrestein et al., 2006] Dorrestein, P. C., Blackhall, J., Straight, P. D., Fischbach, M. A., Garneau-Tsodikova, S., Edwards, D. J., McLaughlin, S., Lin, M., Gerwick, W. H., Kolter,

R., Walsh, C. T., and Kelleher, N. L. (2006). Activity screening of carrier domains within nonribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry. *Biochemistry*, 45:1537–1546.

[Drummond et al., 2011] Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Haras, S., Thierer, T., and Wilson, A. (2011). *Geneious*. Biomatters Limited.

[Edwards et al., 2004] Edwards, D. J., Marquez, B. L., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. (2004). Structure and biosynthesis of the jamaicamides new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chemistry and Biology*, 11:817–833.

[Engene et al., 2012] Engene, N., Rottacker, E. C., Kaštovsky, Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., and Gerwick, W. H. (2012). *Moorea producta* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *International Journal of Systematics and Evolutionary Microbiology*, 62:1171–1178.

[Esquenazi et al., 2011] Esquenazi, E., Jones, A. C., Byrum, T., Dorrestein, P. C., and Gerwick, W. H. (2011). Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proceedings of the National Academy of Science USA*, 108:5226–5231.

[Fattorusso et al., 1971] Fattorusso, E., L., M., and Trivellone (1971). Isolation and structure of nitenin and dihydronitenin, new furanoterpenes from *Spongia nitens*. *Tetrahedron*, 27:3909–3917.

[Feling et al., 2003] Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2003). Salinosporamide a: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angewandte Chemie International Edition*, 42:355–357.

[Fenn, 2003] Fenn, J. B. (2003). Electrospray wings for molecular elephants (nobel lecture). *Angewandte Chemie International Edition*, 42:3871–3894.

[Flatt et al., 2006] Flatt, P. M., O'Connel, S. J., McPhail, K. L., Zeller, G., Willis, C. L., Sherman, D. H., and Gerwick, W. H. (2006). Characterization of the initial enzymatic steps of barbamide biosynthesis. *Journal of Natural Products*, 6:938–944.

[Fleischmann et al., 1995] Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., and Merrick, J. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* rd. *Science*, 269:496–512.

[Freiwald and Sauer, 2009] Freiwald, A. and Sauer, S. (2009). Phylogenetic classification and identification of bacteria by mass spectrometry. *Nature Protocols*, 4:732–742.

[Fujimori and Walsh, 2007] Fujimori, D. G. and Walsh, C. T. (2007). Whats new in enzymatic halogenations. *Current Opinions in Chemical Biology*, 11:553–560.

[Gallagher, 2015] Gallagher, K. A. (2015). *Ecology and Evolution of Hybrid Isoprenoid Secondary Metabolite Production in a Streptomyces Lineage*. University of California, San Diego.

[Galonić et al., 2006] Galonić, D. P., Vaillancourt, F. H., and Walsh, C. T. (2006). Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula*. *Journal of the American Chemical Society*, 128:3900–3901.

[Gao et al., 2011] Gao, S., Sung, W. K., and Nagaraian, N. (2011). Opera: reconstructing optimal genomic scaffolds with high-throughput pair-end sequences. *Journal of Computational Biology*, 18:1681–1691.

[Garrison, 1999] Garrison, T. (1999). *Oceanography: An Invitation to Marine Science*. Wadsworth Publishing Company, 3 edition.

[Gerwick and Moore, 2012] Gerwick, W. H. and Moore, B. S. (2012). Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chemistry and Biology*, 19:85–91.

[Gerwick et al., 1994] Gerwick, W. H., Proteau, P. J., Nagle, D. G., Hamel, E., Blokhin, A., and Slate, D. L. (1994). Structure of curacin a, a novel antimitotic, antiproliferative and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Organic Chemistry*, 59:1243–1245.

[Gross et al., 2007] Gross, H., Stockwell, V. O., Henkels, M. D., Nowak-Thompson, B., Loper, J. E., and Gerwick, W. H. (2007). The genomeisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chemistry and Biology*, 14:53–63.

[Gu et al., 2009] Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., Håkansson, K., Wipf, P., Smith, J. L., Gerwick, W. H., and Sherman, D. H. (2009). Metamorphic enzyme assembly in polyketide diversification. *Journal of Natural Products*, 459:731–735.

[Guindon and Gascuel, 2003] Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biologists*, 53:696–704.

[Gunasekera et al., 2011] Gunasekera, S. P., Owle, C. S., Montaser, R., Luesch, H., and Paul, V. J. (2011). Malyngamide 3 and cocosamides a and b from the marine cyanobacterium *Lyngbya majuscula* from cocos lagoon, guam. *Journal of Natural Products*, 74:871–876.

[Han et al., 2005] Han, B., Goeger, D., Maier, C. S., and Gerwick, W. H. (2005). The wewakpeptins, cyclic depsipeptides from a papua new guinea collection of the marine cyanobacterium *Lyngbya semiplena*. *Journal of Organic Chemistry*, 70:3133–3139.

[Harrigan et al., 1998] Harrigan, G. G., Yoshida, W. Y., Moore, R. E., Nagle, D. G., Park, P. U., Biggs, J., Paul, V. J., Mooberry, S. L., Corbett, T. H., and Valeriote, F. A. (1998). Isolation, structure determination, and biological activity of dolastatin 12 and lyngbyastatin 1 from *Lyngbya majuscula/Schizothrix calcicola* cyanobacterial assemblages. *Journal of Natural Products*, 61:1221–1225.

[Hay, 2009] Hay, M. E. (2009). Marine chemical ecology: Chemical signals and cues structure marine populations, communities, and ecosystems. *Annual Review of Marine Science*, 1:193–212.

[Hayden, 2014] Hayden, E. C. (2014). Technology: The $1,000 genome. *Nature*, 507:294–295.

[Hayes et al., 2014] Hayes, R., Ahmed, A., Edge, T., and Zhang, H. (2014). Core-shell particles: preparation, fundamentals and applications in high performance liquid chromatography. *Journal of Chromatography A*, 1357:36–52.

[Henrikson et al., 2009] Henrikson, J. C., Hoover, A. R., Joyner, M., and Cichewicz, R. H. (2009). A chemical epigenetics approach for engineering the *in situ* biosynthesis of a cryptic natural product from *Aspergillus niger*. *Organic and Biomolecular Chemistry*, 7:435–438.

[Hou et al., 2012] Hou, Y., Braun, D. R., Michel, C. R., and Klassen, J. L. Adnani, N. (2012). Microbial strain prioritization using metabolomics tools for the discovery of natural products. *Analytical Chemistry*, 84:4277–4283.

[Huelsenbeck and Ronquist, 2001] Huelsenbeck, J. P. and Ronquist, F. (2001). Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17:745–755.

[Ianora et al., 1999] Ianora, A., Miralto, A., and Poulet, S. A. (1999). Are diatoms good or toxic for copepods? reply to comment by jónasdóttir et al. *Marine Ecology Progress Series*, 117:305–308.

[Isogai et al., 2012] Isogai, S., Nishiyama, M., and Kuzuyama, T. (2012). Identification of 8-amino-2,5,7-trihydroxynapthalene-1,4-dione, a novel intermediate in the biosynthesis of *Streptomyces* meroterpenoids. *Bioorganic and Medicinal Chemistry Letters*, 22:5823–5826.

[Jabba et al., 2010] Jabba, S. V., Prakash, A., Dravid, S. M., Gerwick, W. H., and Murray, T. F. (2010). Antillatoxin, a novel lipopeptide, enhances neurite outgrowth in immature cerebrocortical neurons through activation of voltage-gated sodium channels. *Journal of Pharmacology and Experimental Therapeutics*, 332:698–709.

[Jensen et al., 2014] Jensen, P. R., Chavarria, K. L., Fenical, W., Moore, B. S., and Ziemert, N. (2014). Challenges and triumphs to genomics-based natural product discovery. *Journal of Interdisciplinary Microbial Biotechnology*, 41:203–209.

[Johnson et al., 2010] Johnson, T. A., Morgan, M. V. C., Aratow, N. A., Estee, S. A., Sashidhara, K. V., Loveridge, S. T., Segraves, N. L., and Crews, P. (2010). Assessing pressurized liquid extraction for the high-throughput extraction of marine-sponge-derived natural products. *Journal of Natural Products*, 73:359–364.

[Jones et al., 2011] Jones, A. C., Monroe, E. A., Podell, S., Hess, W. R., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothman, M., Lasken, R. S., Yates, J. R. I., Reinhardt, R., Kube, M., Burkart, M. D., Allen, E. E., Dorrestein, P. C., Gerwick, W. H., and Gerwick, L. (2011). Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proceeds of the National Academy of Science USA*, 108:8815–8820.

[Katoh et al., 2002] Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002). Biogeochemical cycling and microbial diversity in the thrombolitic microbialites of highborne cay, bahamas. *Nucleic Acids Research*, 30:3059–3066.

[Keck et al., 2011] Keck, G. E., Poudel, Y. B., Cummins, Thomas J. Rudra, A., and Covel, J. A. (2011). Total synthesis of bryostatin 1. *Journal of the American Chemical Society*, 133:744–747.

[Kleigrewe et al., 2015] Kleigrewe, K., Almaliti, J., Yuheng, T. I., Kinnel, R. B., Korobeynikov, A., Monroe, E. A., Duggan, B., Di Marzo, V., Sherman, D. H., Dorrestein, P. C., Gerwick, L., and Gerwick, W. H. (2015). Combining mass spectrometric metabolic profiling with genomic analysis: A powerful approach for discovering natural products from cyanobacteria. *Journal of Natural Products*, 78:1671–1682.

[Koehn and Carter, 2005] Koehn, F. E. and Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery*, 4:206–220.

[Komáek, 2005] Komáek, J.; Anagnostidis, K. (2005). Cyanoprokaryota part 2: Oscillatoriales. In Büdel, B.; Gärtner, G. K. L. S. M., editor, *Süßwasserflora von Mitteleuropa 19/2*, pages 35, 53–55. Elsevier GmbH.

[Komárek and Anagnostidis, 1989] Komárek, J. and Anagnostidis, K. (1989). Modern approach to the classification system of cyanophytes 4 – nostocales. *Archiv für Hydrobiologie Supplements*, 82:247–345.

[Kupče, 2003] Kupče (2003). Two-dimensional hadamard spectroscopy. *Journal of Magnetic Resonance*, 162:310–329.

[Lautru et al., 2005] Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nature Chemical Biology*, 1:265–269.

[Lee et al., 2012] Lee, J., Currano, J. N., Carroll, P. J., and Joulli, M. M. (2012). Didemnins, tamandarins and related natural products. *Natural Product Reports*, 29:404–424.

[Lewtas and Young, 2015] Lewtas, T. and Young, J. (2015). Bruker announces fda clearance for second, expanded claim for the maldi biotyper ca system. www.bruker.com/nc/news-records/single-view/article/bruker-announces-fda-clearance-for-second-expanded-claim-for-the-maldi-biotyper-ca-system.html. Accessed: 2015-07-23.

[Linington et al., 2009] Linington, R. G., Clark, B. R., Trimble, E. E., Almanza, A., Ureña, L.-D., Kyle, D. E., and Gerwick, W. H. (2009). Antimalarial peptides from marine cyanobacteria: Isolation and structural elucidation of gallinamide a. *Journal of Natural Products*, 72:14–17.

[Link, 2015] Link, B. (2015). Customer relations lead aquatics and firstrax divisions, instant ocean, spectrum brands inc. In a personal e-mail. Received: 2015-01-13.

[Liu et al., 2013] Liu, W. T., Lamsa, A., Wong, W. R., Boudreau, P. D., Kersten, R., Peng, Y., Moree, W. J., Duggan, B. M., Moore, B. S., Gerwick, W. H., Linington, R. G., Pogliano, K., and Dorrestein, P. C. (2013). Ms/ms-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *Journal of Antibiotics*, 67:99–104.

[Liu et al., 2009] Liu, W. T., Ng, J., Meluzzi, D., Bandeira, N., Gutierrez, M., Simmons, T. L., Schultz, A. W., Linington, R. G., Moore, B. S., Gerwick, W. H., Pevzner, P. A., and Dorrestein, P. C. (2009). Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Analytical Chemistry*, 81:4200–4209.

[Long et al., 2007] Long, J. D., Smalley, G. W., Barsby, T., Anderson, J. T., and Hay, M. E. (2007). Chemical cues induce consumer-specific defenses in a bloom-forming marine phytoplankton. *Proceedings of the National Academy of Sciences USA*, 104:10512–10517.

[Lowery et al., 2005] Lowery, C. A., McKenzie, K. M., Qi, L., Meijler, M. M., and Janda, K. D. (2005). Quorum sensing in *Vibrio harveyi*: probing the specificity of the luxp binding site. *Bioorganic and Medicinal Chemistry Letters*, 15:2395–2398.

[Luesch et al., 2001a] Luesch, H., Pangilinan, R., Wesley, Y., Moore, R., and Paul, V. J. (2001a). Pitipeptolides a and b, new cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 64:304–307.

[Luesch et al., 2002] Luesch, H., Wesley, Y.and Moore, R., and Paul, V. J. (2002). New apratoxins of marine cyanobacterial origin from guam and palau. *Bioorganic and Medicinal Chemistry*, 30:1973–1978.

[Luesch et al., 2001b] Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J., and Corbett, T. H. (2001b). Total structure determination of apratoxin a, a potent novel cytotoxin

from the marine cyanobacterium *Lyngbya majuscula*. *Journal of the American Chemical Society*, 123:5418–5423.

[Luesch et al., 2000] Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J., and Mooberry, S. L. (2000). Isolation, structure determination, and biological activity of lyngbyabellin a from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 63:611–615.

[Malloy, 2011] Malloy, K. L. (2011). *Structure Elucidation of Biomedically Relevant Marine Cyanobacterial Natural Products*. Oregon State University.

[Mamer, 2000] Mamer, O. A. (2000). Synthesis and gas chromatography/mass spectrometry analysis of stereoisomers of 2-hydroxy-3-methylpentanoic acid. *Methods in Enzymology*, 324:3–10.

[Marfey, 1984] Marfey, P. (1984). Determination of d-amino acids. ii. use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzne. *Carlsberg Research Communications*, 49:591–596.

[Marquez et al., 2002] Marquez, B. L., Watts, K. S., Yokochi, A., Roberts, M. A., Verdier-Pinard, P., Jimenez, J. I., Hamel, E., Scheuer, P. J., and Gerwick, W. H. (2002). Structure and absolute stereochemistry of hectorchlorin, a potent stimulator of actin assembly. *Public Library of Science Online Edition*, 65:866–871.

[Medema et al., 2011] Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011). antismash: rapid identification, annotation and anlysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39:W339–W346.

[Medina, 2009] Medina, R. A. (2009). *Biologically Active Cyclic Depsipeptides from Marine Cyanobacteria*. Oregon State University.

[Mevers et al., 2011] Mevers, E., Liu, W.-T., Engene, N., Mohimani, H., Byrum, T., Pevzner, Pavel A.and Dorrestein, P. C., Spadafora, C., and Gerwick, W. H. (2011). Cytotoxic veraguamides, alkynyl bromide-containing cyclic depsipeptides from the marine cyanobacterium *Oscillatoria margaritifera*. *Journal of Natural Products*, 74:928–936.

[Miller et al., 2014] Miller, B., Friedman, A. J., Choi, H., Hogan, J., McCammon, J. A., Hook, V., and Gerwick, W. H. (2014). The marine cyanobacterial metabolite gallinamide a is a potent and selective inhibitor of human cathepsin l. *Journal of Natural Products*, 77:92–99.

[Miralto et al., 1999] Miralto, A., Barone, G., Romano, G., Poulet, S. A., Ianora, A., Russo, G. L., Buttino, I., Mazzarella, G., Laabir, M., Cabrini, M., and Giacobbe, M. G. (1999). The insidious effect of diatoms on copepod reproduction. *Nature*, 402:173–176.

[Molinski, 2010] Molinski, T. F. (2010). Nmr of natural products at the 'nanomole-scale'. *Natural Products Reports*, 27:321–329.

[Montaser et al., 2011] Montaser, R., Paul, V. J., and Luesch, H. (2011). Pitipeptolides c-f, antimycobacterial cyclodepsipeptides from the cyanobacterium *Lyngbya majuscula* from guam. *Phytochemistry*, 72:2068–2074.

[Murray, 2010] Murray, P. R. (2010). Rapid identification of clinical yeast isolates by mass spectrometry. *Current Fungal Infection Reports*, 4:145–150.

[Nakao et al., 1998] Nakao, Y., Yoshida, Wesley, Y., Szabo, C. M., Baker, B. J., and Scheuer, P. J. (1998). More peptides and other diverse constituents of the marine mollusk *Philinopsis speciosa*. *Journal of Organic Chemistry*, 63:3272–3280.

[Neumann et al., 2008] Neumann, C. A., Fujimori, D. G., and Walsh, C. T. (2008). Halogenation strategies in natrual product biosynthesis. *Chemistry and Biology*, 15:99–109.

[Ng et al., 2009] Ng, J., Bandeira, N., Lie, W.-T., Ghassemian, M., Simmons, T. L., Gerwick, W. H. L. R., Dorrestein, P. C., and Pevzner, P. A. (2009). Dereplication an *de novo* sequencing of nonribosomal peptides. *Nature Methods*, 6:596–600.

[Nikolenko et al., 2013] Nikolenko, S., Korobeynikov, A., and Alekseyev, M. A. (2013). Bayeshammer: Bayesian clustering for error correction in single-cell sequencing. *BCM Genomics*, 14:S7.

[Nogle and Gerwick, 2002] Nogle, L. M. and Gerwick, W. H. (2002). Isolation of four new cyclic depsipeptides, antanapeptins a-d, and dolastatin 16 from a madagascan collection of *Lyngbya majuscula*. *Journal of Natural Products*, 65:21–24.

[Nunnery et al., 2012] Nunnery, J. K., Engene, N., Byrum, T., Cao, Z., Jabba, S. V., Pereira, A. R., Matainaho, T., Murray, T. F., and Gerwick, W. H. (2012). Biosynthetically intriguing chlorinated lipophilic metabolites from geographically distant tropical marine cyanobacteria. *Journal of Organic Chemistry*, 77:4198–4208.

[Nunnery et al., 2011] Nunnery, J. K., Suyama, T. L., Linington, R. G., and Gerwick, W. H. (2011). Expedient synthesis of α,α-dimethyl-β-hydroxy carbonyl scaffolds via evans' aldol reaction with a tertiary enolate. *Tetrahedron Letters*, 52:2929–2932.

[Nurk et al., 2013] Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., Prjibelski, A. D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S. R., Woyke, T., Mclean, J. S., Lasken, R., Tesler, G., Alekeyev, M. A., and Pevzner, P. A. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric mda products. *Journal of Computational Biology*, 20:714–737.

[Paul and Fenical, 1980] Paul, V. J. and Fenical, W. (1980). Toxic acetylene-containing lipids from the red marine algae *Liagora farinosa lamouroux*. *Tetrahedron Letters*, 21(35):3327–3330.

[Pereira et al., 2012] Pereira, A. R., Kale, A. T., Byrum, T., Debonsi, H. M., Gilson, M. K., Valeriote, F. A., Moore, B. S., and Gerwick, W. H. (2012). The carmaphycins: new proteasome inhibitors exhibiting an α,β-epoxyketone warhead from a marine cyanobacterium. *ChemBioChem*, 13:810–817.

[Perez et al., 2005] Perez, E. A., Hillman, D. W., Fishkin, P. A., Krook, J. E., Tan, W. W., Kuriakose, P. A., Alberts, S. R., and Dakhil, S. R. (2005). Phase ii trial of dolastatin-10 in patients with advanced breast cancer. *Investigational New Drugs*, 23:257–261.

[Pettit et al., 1989] Pettit, G. R., Kamano, Y., Herald, C. L., Tuinman, A. A., Boettner, F. E., Kizu, H., Schmidt, J. M., Baczynskyj, L., Tomer, K. B., and Bontems, R. J. (1989). The isolation and structure of a remarkable marine animal antineoplastic constituent: Dolastatin 10. *Journal of the American Chemical Society*, 109:6683–6685.

[Pettit, 2011] Pettit, R. K. (2011). Culturability and secondary metabolite diversity of extreme mircobes: Expanding contribution of deep sea and deep-sea vent microbes to natural product discovery. *Marine Biotechnology*, 13:1–11.

[Piel, 2004] Piel, J. (2004). Metabolites from symbiotic bacteria. *Natural Product Reports*, 21:519–538.

[Piel, 2009] Piel, J. (2009). Metabolites from symbiotic bacteria. *Natural Product Reports*, 26:338–362.

[Pinet, 1992] Pinet, P. R. (1992). *Oceanography, an introduction to planet oceanus*. West Coast Publishing Company.

[Pitot et al., 1999] Pitot, H. C., McElroy, E. A. J., Reid, J. M., Windebank, A. J., Sloan, J. A., Erlichman, C., Bagniewski, P. G., Walker, D. L., Rubin, J. Goldberg, R. M., Adjei, A. A., and Ames, M. M. (1999). Phase i trial of dolastatin-10 (nsc 376128) in patients with advanced solid tumors. *Clinical Cancer Research*, 5:525–531.

[Posada, 2008] Posada, D. (2008). jmodeltest: Phylogenetic model averaging. *Molecular Biology and Evolution*, 25:1253–1256.

[Quilico et al., 1956] Quilico, A., Piozzi, F., and Pavan, M. (1956). Sulla dendrolasina (italian). *La Ricerca Scientifica*, 26:177–180.

[Quilico et al., 1957] Quilico, A., Piozzi, F., and Pavan, M. (1957). The structure of dendrolasin. *Tetrahedron*, 1:177–185.

[Ramaswamy et al., 2007] Ramaswamy, A., Sorrels, C. M., and Gerwick, W. H. (2007). Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 70:1977–1986.

[Rausch et al., 2005] Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D. H. (2005). Specificity prediction of sentlation domains in nonribosomal peptide synthetases (nrps) using transductive support vector machines (tsvms). *Nucleic Acids Research*, 33:5799–5808.

[Reese et al., 1996] Reese, M. T., Gulavita, N. K., Nakao, Y., Hamann, M. T., Yoshida, Wesley, Y., Coval, S. J., and Scheuer, P. J. (1996). Kulolide: A cytotoxic depsipeptide from a cephalaspidean mollusk, *Philinopsis speciosa*. *Journal of the American Chemical Society*, 118:11081–11084.

[Rinehart et al., 1981a] Rinehart, K. L. J., Gloer, J. B., and Cook, J. C. J. (1981a). Structures of the didemnins, antiviral and cytotoxic depsipeptides from a caribbean tunicate. *Journal of the American Chemical Society*, 103:1857–1859.

[Rinehart et al., 1981b] Rinehart, K. L. J., Gloer, J. B., Hughes, R. G., Renis, H. E., McGovern, J. P., Swynenberg, E. B., Stringfellow, D. A., Kuentzel, S. L., and Li, L. H. (1981b). Didemnins: antiviral and antitumor depsipeptides from a caribbean tunicate. *Science*, 212:933–935.

[Röttig et al., 2011] Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). Nrpspredictor2a web server for predicting nrps adentlation domain specificity. *Nucleic Acids Research*, 39:W362–W367.

[Salvador et al., 2011] Salvador, L. A., Biggs, J. S., Paul, V. J., and Luesch, H. (2011). Veraguamides a-g, cyclic hexadepsipeptides from a dolastatin 16-producing cyanobacterium *Symploca* cf. *hydnoides* from guam. *Journal of Natural Products*, 74:917–927.

[Sauer and Kliem, 2010] Sauer, S. and Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology*, 8:74–82.

[Sharma and Burkholder, 1967a] Sharma, G. M. and Burkholder, P. R. (1967a). Studies on antimicrobial substances of sponges. i. isolation, purification, and properties of a new bromine-containing antibacterial substance. *Journal of Antibiotics Series A (Japan)*, 20:200–203.

[Sharma and Burkholder, 1967b] Sharma, G. M. and Burkholder, P. R. (1967b). Studies on the antimicrobial substances of sponges ii. structure and synthesis of a bromine-containing antibacterial compound from a marine sponge. *Tetrahedron Letters*, 42:4147–4150.

[Shiomi et al., 1986] Shiomi, K., Iinuma, H., Hamada, M., Naganawa, H., Manabe, M., Matsuki, C., Takeuchi, T., and Umezawa, H. (1986). Novel antibiotics napyradiomycins production, isolation, physio-chemical properties and biological actvity. *The Journal of Antibiotics*, 39:487–493.

[Sihvonen et al., 2007] Sihvonen, L., Lyra, C., Fewer, D. P., Rajaniemi-Wacklin, P., Lehtimäki, J. M., Wahlsten, M., and Sivonen, K. (2007). Strains of the cyanobacterial genera *Calothrix* and *Rivularia* isolated from the baltic sea display cryptic diversity

and are distantly related to *Gloeotrichia* and *Tolypothrix*. *FEMS Microbiology Ecology*, 61:74–84.

[Sitachitta et al., 2000] Sitachitta, N., Williamson, R. T., and Gerwick, W. H. (2000). Yanucamides a and b, two new depsipeptides from an assemblage of the marine cyanobacteria lyngbya majuscula and schizothrix species. *Journal of Natural Products*, 63:197–200.

[Starr et al., 1999] Starr, M., Runge, J. A., and Therriault, J.-C. (1999). Effects of diatom diets on the reproduction of the planktonic copepod *Calanus finmarchicus*. *Sarisa*, 84:379–389.

[Stolze et al., 2012] Stolze, S. C., Deu, E., Kaschani, F., Li, N., Florea, B. I., Richau, K. H., Colby, T., van der Hoorn, R. A. L., Overkleeft, H. S., Bogyo, M., and Kaiser, M. (2012). The antimalarial natural product symplostatin 4 is a nanomolar inhibitor of the food vacuole falcipains. *Chemistry and Biology*, 19:1546–1555.

[Suntornchashwej et al., 2005] Suntornchashwej, S., Chaichit, N., Isobe, M., and Suwanborirux, K. (2005). Hectochlorin and morpholine derivatives from the the thai sea hare, *Bursatella leachii*. *Journal of Natural Products*, 68:951–955.

[Tan, 2007] Tan, L. T. (2007). Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry*, 68:954–979.

[Tan, 2010] Tan, L. T. (2010). Filamentous tropical marine cyanobacteria: a rich source of natural products for anticancer drug discovery. *Journal of Applied Phycology*, 22:659–676.

[Taniguchi et al., 2010] Taniguchi, M., Nunnery, J. K., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., and Gerwick, W. H. (2010). Palmyramide a, a cyclic depsipeptide from a palmyra atoll collection of the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 73:393–398.

[Taori et al., 2009] Taori, K., Liu, Y., Paul, V. J., and Luesch, H. (2009). Combinatorial strategies by marine cyanobacteria: Symplostatin 4, an antimitotic natural dolastatin 10/15 hybrid that synergizes with the coproduced hdac inhibitor largazole. *ChemBioChem*, 10:1634–1639.

[Taori et al., 2008] Taori, K., Paul, V., and Luesch, H. (2008). Structure and activity of largazole, a potent antiproliferative agent from the floridian marine cyanobacterium *Symploca sp*. *Journal of the American Chemical Society*, 130:1806–1807.

[Terpe, 2006] Terpe, K. (2006). Overview of bacterial expression systems for protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72:211–222.

[Thronburg et al., 2010] Thronburg, C. C., Zabriskie, M., and McPhail, K. L. (2010). Deep-sea hydrothermal vents: Potential hot spots for natural products discovery? *Journal of Natural Products*, 73:489–499.

[Tripathi et al., 2009] Tripathi, A., Puddick, J., Prinsep, M. R., Lee, P. P. F., and Tan, L. T. (2009). Hantupeptin a, a cytotoxic depsipeptide from a singapore collection of *Lyngbya majuscula*. *Journal of Natural Products*, 72:29–32.

[Tripathi et al., 2010] Tripathi, A., Puddick, J., Prinsep, M. R., Lee, P. P. F., and Tan, L. T. (2010). Hantupeptins b and c, cytotoxic cyclodepsipeptides from the marine cyanbacterium *Lyngbya majuscula*. *Phytochemistry*, 71:307–311.

[Udwary et al., 2007] Udwary, D. W., Zeigler, L., N., A. R., Singan, V., Lapidus, A., W., F., Jensen, P. R., and Moore, B. S. (2007). Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proceedings of the National Academy of Sciences USA*, 104:10376–10381.

[Vanderah and Schmitz, 1975] Vanderah, D. J. and Schmitz, F. J. (1975). Marine natural products: Isolation of dendrolasin from the sponge *Oligoceras hemorrhages*. *Lloydia*, 38(3):271–272.

[Wan and Erickson, 2001] Wan, F. and Erickson, K. L. (2001). Georgamide, a new cyclic depsipeptide with an alkynoic acid residue from an australian cyanobacterium. *Journal of Natural Products*, 64:143–146.

[Watrous et al., 2012] Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J. Bandeira, N., and Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences USA*, 109:E1743–E1752.

[Weinheimer and Spraggins, 1969] Weinheimer, A. J. and Spraggins, R. L. (1969). The occurrence of two new prostaglandin derivatives (15-epi-pga_2 and its acetate, methyl ester) in the gorgonian *Plexaura Homomalla*. *Tetrahedron Letters*, 59:5185–5188.

[Xu et al., 2012] Xu, Y., Kersten, R. D., Nam, S.-J., Lu, L., Al-Suwailem, A. M., Zheng, H., Fenical, W., Dorrestein, P. C., Moore, B. S., and Qian, P.-Y. (2012). Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *Journal of the American Chemical Society*, 134:8625–8632.

[Xu et al., 2003] Xu, Z., Peng, Y., and Ye, T. (2003). The total synthesis and stereochemical revision of yanucamide a. *Organic Letters*, 5:355–357.

[Yang et al., 2012] Yang, J. Y., Phelan, V. V., Simkovsky, R., Watrous, J. D., Trial, R. M., Fleming, T. C., Wenter, R., Moore, B. S., Golden, S. S., Pogliano, K., and Dorrestein, P. C. (2012). Primer on agar-based microbial imaging mass spectrometry. *Journal of Bacteriology*, 194:6023–6028.

[Yang et al., 2013] 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., Ruh Wong, W. Linington, R. G., Zhang, L., Debonsi, H. M., Gerwick, W. H., and Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. *Journal of Natural Products*, 76:1686–1689.

[Ziemert et al., 2012] Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012). The natural product domain seeker napdos: A phylogeny based tool to classify secondary metabolite gene diversity. *Public Library of Science Online Edition*, 7:e34064.