Investigation of Marine Invertebrate-Associated Bacteria for Antibacterial Natural Products

by

Jia-Xuan Yan

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Pharmaceutical Sciences)

at the

University of Wisconsin-Madison

Graduation Year: 2020

Date of oral examination: December 18, 2019

The dissertation is approved by the following members of the Final Oral Committee:

Dr. Tim S. Bugni, Professor, Pharmaceutical Sciences

Dr. Weiping Tang, Professor, Pharmaceutical Sciences

Dr. Jennifer E. Golden, Assistant Professor, Pharmaceutical Sciecnes

Dr. Warren E. Rose, Associate Professor, Pharmacy Practice

Dr. Rodney A. Welch, Medical Microbiology & Immunology

ABSTRACT

This dissertation focuses on the discovery and development of antibacterial natural products derived from marine invertebrate-associated bacteria. Natural products have historically played an important role in the development of therapeutics, but several challenges still hinder natural product drug discovery. Methods for improving the efficiency of natural product discovery and structure elucidations of natural products are reviewed in Chapter 1.

The discovery of new antibacterial compounds madurastatin D1 and D2 from a marine-derived bacterium (*Actinomadura* sp. WMMA-1423) is described in Chapter 2. Biosynthetic mechanisms were proposed for these two compounds by analyzing the genome of the producing bacterium.

The discovery of six new antibacterial compounds bacillimidazoles A-F from a marine-derived bacterium (*Bacillus* sp. WMMC-1349) is described in Chapter 3. These compounds were discovered rapidly through a marine natural product library discovery platform. Possible biosynthetic mechanisms were proposed based on isotopic labeling studies.

Unambiguous molecular formula determination method based on Isotopic Fine Structure (IFS) analysis was developed and applied for rapid structure elucidations of three classes of natural products discussed in Chapters 4-6. In Chapter 4, the discovery of a new antibacterial ribosomally synthesized post-translationally modified peptide (RiPP) haliclonamycin from a marine-derived bacterium (*Streptomyces* sp. WMMC-911) is described. IFS analysis was used to provide a possible molecular formula for this compound with larger than 2,000 M.W.

Due to signal overlapping in NMR spectra, the structure elucidation of this compound has not been accomplished yet. The whole genome of the producing organism has been sequenced and will be analyzed to provide structure information of this RiPP.

The discovery of two new non-ribosomal peptides (NRP) ecteinamines A and B from a marine-derived bacterium (*Micromonospora* sp. WMMB-482) is described in Chapter 5. IFS analyses were applied for both compounds and provided their molecular formulae unambiguously. In addition, MS^2 and MS^3 studies on ecteinamine B provided important functional groups information. Combined with NMR studies, including ^{13}C-^{13}C COSY, ^{13}C-^{15}N HMQC correlations on isotopic enriched compounds, the structures of these two novel compounds were elucidated rapidly.

The discovery of one new non-ribosomal peptides (NRP) streptnatamide A from a marine-derived bacterium (*Streptomyces* sp. WMMB-303) is described in Chapter 6. IFS analysis was applied this compound and provided its molecular formula unambiguously. The structure elucidation of streptnatamide A was conducted based on multiple NMR experiments including COSY, TOCSY, ROESY, HMBC. Advanced Marfey's analysis was used to determine the absolute configuration of streptnatamide A.

Acknowledgments

First, I would like to express my appreciation to my advisor Dr. Tim Bugni for his patience, guidance, and great support over the past five years. Without his disciplined training and meticulous guidance, I would never have a chance to learn natural products and do important research in natural products. His passion for seeking the truth, teachning science, and mentoring new students will guide me continuously throughout my own career. I would also like to thank my thesis committee: Prof. Weiping Tang, Prof. Jennifer Golden, Prof. Warren Rose, and Prof. Rodney Welch for their kind advice and academic support.

I would like to say thank you to all members of the Bugni lab, both current and past members. These people are the ones who I have learned from everyday since I joined this happy and diligient lab. I would especially like to thank Dr. Fan Zhang, who helped me throughout my graduate school, as well as my good friends Dr. Navid Adnani, Dr. Chris Thomas, Dr. Rene Ramos, Dr. Deepa Acharya, Mr. Shaurya Chanana, Miss I-Wei Tsai. I also would like to thank Mr. Doug Braun, Dr. Kenneth Barns for helping me with a lot of biological experiments that I am not very good at. Also, thanks to the new members of the Bugni lab, Dr. Qihao Wu, Dr. Changyeol Lee, and Mr. Imraan Alas for bringing in new energies and expertise to this great lab. All these people have helped me through my graduate school and keep my life happy.

The work described in this dissertation would not be possible without collaborations with other great people. Dr. Marc Chevrette from Dr. Cameron Currie’s lab helped me with

assembling the whole genome of some bacteria and his help was crucial in proposing the biosynthetic mechanisms of two classes of new compounds. Dr. Ziqing Lin and Dr. Yanlong Zhu from Dr. Ying Ge's lab trained me and Dr. Fan Zhang on their top tier instrument Bruker MRMS and help us develop the IFS analysis method. Special thanks to the Analytical Instrumentation Facility in the School of Pharmacy and the National Magnetic Resonance Facility at Madison (NMRFAM) for the great resources for me to learn NMR and mass spectrometry. I would especially like to acknowledge Gary Girdaukus, Dr. Cameron Scarlett, Dr. Thomas Stringfellow, Ms. Molly Pellitteri Hahn, Dr. Paulo Cobra, Dr. Marco Tonelli from those facillities. I would also like to thank the UW School of Pharmacy and the NIH for funding my research.

Finally, it would be impossible for me to complete my graduate school without the love and support from my family and friends. I would like to give my greatest appreciations to my parents Mr. Kelin Yan and Mrs. Shiyan Yang for bringing me to this lovely world and giving me selfless care and support throughout my life. Also, I would like to thank my grandparents, Mr. Hairui Yan and Mrs. Xiaoci Zhan for their kind support and great suggestions. Last but not the least, I would like to say thank you to my girlfriend Miss Huiran Peng for bringing me love and happiness in the last few years of my graduate school.

Table of Contents

List of Tables

List of Figures

List of Compounds

Abbreviations

ASW artificial seawater

ATCC American type culture collection

c concentration

˚C Celsius

COSY correlation spectroscopy

˚ degree

DFT density functional theory

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

ESI-MS electrospray ionization mass spectrometry

FDLA 1-fluoro-2,4-dinitrophenyl-5-leucine amide

FT-IR Fourier transform infrared

g gram

h hour

Hz hertz

HMBC heteronuclear multiple bond correlation

HMQC heteronuclear single quantum multiple bond correlation

HPLC high-performance liquid chromatography

HRMS high resolution mass spectrometry

HSQC heteronuclear single quantum coherence

IR infrared

L liter

m/z mass/charge

MHz megahertz

MIC minimum inhibitory concentration

MeCN acetonitrile

MeOH methanol

μg microgram

min minute

mL milliliter

MRSA methicillin-resistant *Staphylococcus aureus*

MS mass spectrometry

nm nanometer

NMR nuclear magnetic resonance

NOE nuclear Overhauser enhancement

PCA principal component analysis

ppm parts per million

ROESY rotating frame Overhauser enhancement spectroscopy

RT retention time

SAR structure-activity relationship

TOCSY total correlation spectroscopy

UPLC ultra-performance liquid chromatrography

UV ultraviolet

WMMA Wisconsin Marine Microbe (collection A)

WMMB Wisconsin Marine Microbe (collection B)

WMMC Wisconsin Marine Microbe (collection C)

Chapter 1:

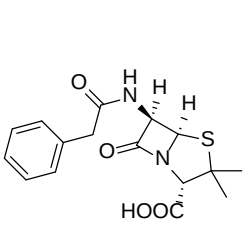
Introduction

1.1. Introduction to Antibiotic Drug Discovery

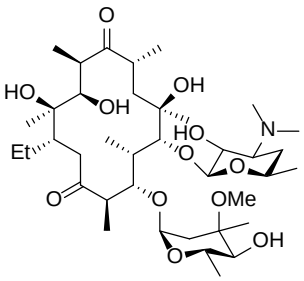
Natural products have served as one of the most abundant and reliable sources of antibiotics historically[1-3]. Starting in 1930s, the antibiotic employments in medical practices have been recognized as one of the most significant milestones in extending the life expectancy of human beings by reducing the risk of mortal microbial infections[4]. Sir Alexander Fleming's serendipitous discovery of the bactericidal fungal metabolite penicillin in 1929 and the discovery of streptomycin in 1943 (Waksman and co-workers) by mining the soil-derived actinomyces secondary metabolites began the golden era of antibiotics discovery[5]. Penicillin was recognized as a miraculous drug during the Second World War for saving millions of lives[6]. For the proceeding decades, natural products continued to provide the pharmaceutical community with numerous antibiotics with a high degree of chemical diversity and antimicrobial biological activities. Meanwhile, the complex and diverse pharmacophores of natural products have also inspired the design and discovery of synthetic as well as semi-synthetic antibiotic compounds[7], which further expanded the diversity of antibiotics and enabled the successful treatment of a broad spectrum of pathogens. Those pharmacophores have also led to the classifications of several antibacterial and antifungal mechanisms, such as interfering with cell wall synthesis (penicillin G, **1**), interfering with protein synthesis (erythromycin, **2**), and interfering with cell membrane permeability (amphotericin B, **3**)[8].

The majority of antibiotics are derived from secondary metabolites of microorganisms

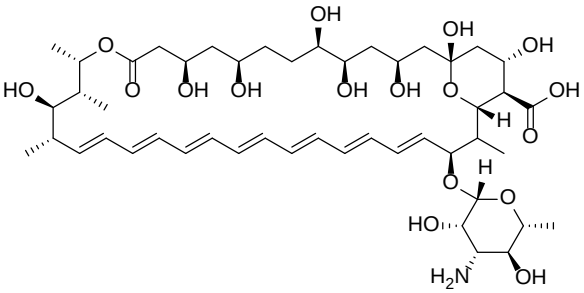
typically discovered via biological activity-guided detection and purification method (Waksman's platform)[9]. The antibiotics produced by microorganisms are in fact the results of billions of years of evolution and natural selection[10]. These compounds could recognize specific cell targets with high efficiency and selectivity, and since they are produced by microorganisms, they tend to have better physiological and chemical properties compared to synthetic compounds that make them more biocompatible and less toxic[11]. It is estimated that around 28,000 antibiotic compounds have been discovered from microorganisms; around 200 of these compounds have been directly developed as drugs in clinical practices[12].



Penicillin G (**1**)



Erythromycin (**2**)



amphotericin B (**3**)

1.2. Challenges for Current Antibiotic Discovery.

However, antibiotic resistance has always been a life-threatening issue for human beings[13]. Resistance towards certain antibiotics, such as penicillin, was first observed in clinical trials starting in 1940, usually a few years after the deployment of newly discovered antibiotic. In a more specific and recent case, the antibiotic resistance towards the lipopeptide antibiotic daptomycin was observed within a year after its first clinical use[13]. Improper antibiotics treatment towards infectious diseases, including but not limited to over-prescription and improper duration of treatments would induce drug-resistant mechanisms developed by microorganisms, which would lead to failure in both treatment as

well as control of infectious diseases[14]. The common antibiotic resistance mechanisms include decreasing permeability of antibiotics, developing efflux pumps, chaning antibiotic target sites, and enzymatic deactivations of antibiotics by modifying their chemical structures[15]. More dangerously, scientists have observed that pathogens can collect different genetic elements to develop broader spectrum resistance to different classes of antibiotics to make them multidrug resistant (MDR)[16]. It is estimated that antibiotic resistance is leading to over 2,000,000 infections and 23,000 deaths each year globally[14], and unfortunately, scientists have no solutions for the rise of antibiotic resistance but trying our best to discover new classes of antibiotics with new antimicrobial mechanisms.

Paradoxically, despite a growing clinical need, the pharmaceutical industry, which was once at the forefront of infectious disease research and development, has largely abandoned antibiotics[17]. Another reason for the severe situation we are currently facing is because the traditional antibiotic discovery platform has been too successful[4]. In fact, a majority of antibiotics discovered back in 1940s-1970s are still playing an active role in contemporary pharmaceutical practices. These old drugs are cheap, effective, broad spectrum and non-toxic, which gave scientists and clinicians an illusion that old drugs could solve all types of infections. These legendary old drugs have set up strict criteria for new antibiotic compounds. In most cases, broad-spectrum antimicrobial activities towards different pathogens and low toxicities towards human cells are required for a successful antibiotic drug candidate. Also, the traditional method, which is based on mining antibiotic secondary metabolites produced by soil-derived actinomyces[18], has been extremely successful and leading to not enough attention from scientists to other microorganism sources, such as marine-derived bacteria[19-22],

both actinomyces and proteobacteria. Nowadays the traditional methods lead to a high level of compound rediscovery – greater than 99.5% rediscovery rate[23]. These challenges need to be addressed in order to speed up the discovery of novel antibiotics to avoid a pre-antibiotic era type situation.

In order to counter the abovementioned problems and challenges, our lab has made several efforts, including modifying the traditional discovery platform, using metabolomics analysis to achieve strain prioritization, and looking for new compounds produced by relatively less studied bacterial genera. These would be discussed in further details in the next section.

1.3. Our Efforts on Compound Discovery from Marine Bacteria.

As mentioned in the previous section, the traditional method of antibiotic discovery from natural products generally follows the Waksman's biological activity guided detection and purification platform. In recent years, it is becoming more and more often for known compounds or simple analogs of known compounds discovery through this platform. Scientists have proposed multiple strategies in order to improve the traditional strategy, and our lab has selected three of them: prioritizing bacteria strains through metabolomics analysis[24, 25]; investigating secondary metabolites from relatively underexplored bacteria genera[26, 27]; and establishing and developing a natural product library based discovery platform for rapid dereplication of known compounds and discovery of new antibiotics.

Metabolomics have proven to be one of the leading trends in 21st century natural product research[28]. The understandings of bacterial metabolomics can lead to the prioritization of bacterial strains that can produce novel chemistry[29]. Recent advances in analytical

instruments (UHPLC-HRMS) and databases have stimulated the natural product community to focus on this research field[30]. LC/MS analysis and Principal Component Analysis (PCA) of bacterial secondary metabolites was utilized by our lab in 2012 aiming to achieve strain prioritization and solve one of the fundamental problems in natural product research, which is detecting new molecules in the abundance of knowns (Figure 1.1)[31]. Secondary metabolites of a group of 40-100 different bacteria were analyzed by UHPLC-HRMS and the acquired LC/MS spectra were analyzed by the Bruker ProfileAnalysis software. Data from each UHPLC/HRMS analysis were condensed into buckets of UHPLC retention time (RT)-mass-to-charge ratio (*m/z*) pairs and their corresponding MS peak intensities. The generated data buckets underwent PCA to reduce the dimensionality of the data through multivariate analysis and plot most of the variance between samples in three dimensions. The variance between different samples (different chemistry produced) can be visualized in 2D plots of the Principal component (PC) axes. Each dot represents the LC/MS spectra of a specific bacteria strain on the Scores plots and a specific RT-*m/z* pair (compound) on the Loadings plots. The Scores plot indicates the overall variance among different samples (secondary metabolites produced by different bacteria) while the Loadings plots show the variance of the different RT-*m/z* pairs. If two bacteria are producing similar natural products, they would be close to each other on the Scores Plot; if they are producing different chemistry, they would be far away from each other on the Scores Plot. Prioritized strains would be cultured in larger scale fermentations and the PCA result could provide guidance for future purification of the target compounds.

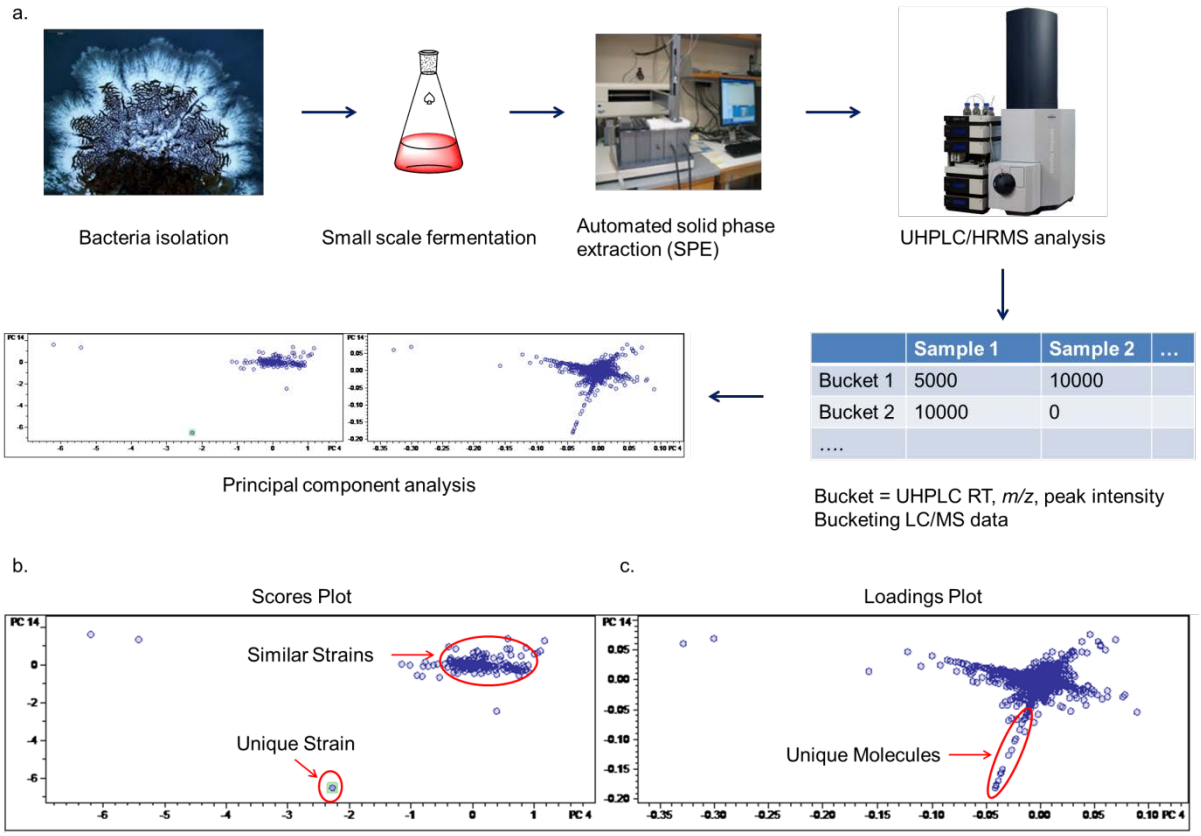
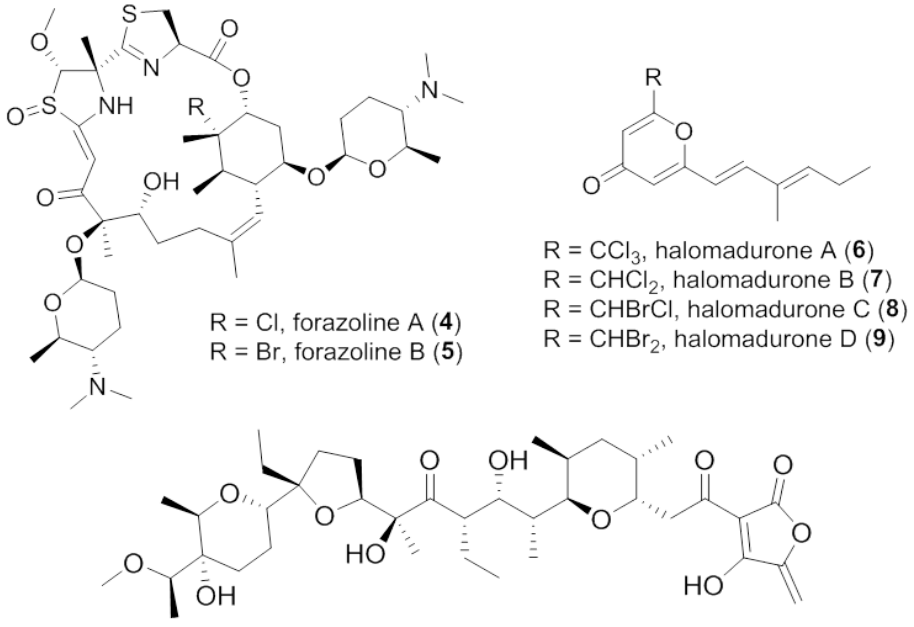


Figure 1.1. Overview of LCMS-PCA analysis of bacterial secondary metabolites. (a). Generation of PCA analysis data starting from isolation and fermentation of bacterial strains under lab conditions. Generally a 25 mL bacteria culture would produce enough material for the following automated solid phase extraction (SPE) and UHPLC/HRMS analysis. The acquired LC/MS spectra of 40-100 strains would be analyzed by a Bruker software ProfileAnalysis. The UHPLC RT, *m/z*, and MS intensities of the secondary metabolites produced by bacteria would be combined together to generate a three-dimensional space. (b). Two-dimensional planes can be constructed in the 3D space by principal component axes (PC planes) to generate Scores Plots. Each dot on the Scores Plot represents a bucket of data (RT, *m/z*, MS intensity) produced by a specific bacteria and the grouped together dots indicate their secondary metabolites are similar. The outliers are producing some secondary metabolites that are different from the other strains analyzed in the same batch. (c). Each dot on the Loadings Plot corresponds to a specific bucket of data (or specific compounds). Geometrically, the Loadings Plot is related to the Scores Plot because the outlier dots are making the most contributions to the variance of buckets, which can be depicted on the Scores Plot. The dots (compounds) in the upper left circle are the result of the outlier dot to the upper left of the Scores Plot.

PCA has proven to be successful for selection of bacteria strains that can produce structural diverse compounds[32-37]. As an example, *Actinomadura sp.* WMMB-499 was prioritized after LC/MS PCA analysis of 36 bacteria strains and resulted in the discovery of three classes of new compounds: forazolines (**4**, **5**)[32], halomadurones (**6**-**9**)[33], and ecteinamycin (**10**)[34]. As a comparison, it is estimated that one novel antibiotic can be discovered per 100,000 to even 1,000,000 strains screened using the traditional antibiotic discovery strategy[23]. The discovery of these three classes of compounds with diverse chemical structures and biological activities highlights the strength of our LC/MS PCA analysis method since it is able to predict and identify novel chemistry production by analyzing a relatively small group of bacteria.



Historically, soil derived actinomyces have been the most abundant source of antibiotics. However in recent years, high known compounds rediscovery rate has been observed from these reliable sources of antibiotics. Our lab has been focusing on novel natural products discovery from marine derived bacteria for the great biological diversity of the marine

environment. Some of the relatively less studied marine bacteria such as *Actinomadura* sp.[38], *Micromonospora* sp.[39], and *Bacillus* sp.[40] are collected and studied as a major part of our natural product library platform. Because the LC/MS PCA analysis on those relatively less studied bacteria genera can not predict the biological activities of the produced novel chemistry, we have designed and developed a marine natural product library based discovery platform in order to streamline the rapid discovery of new antibiotics as well as rapid dereplication of known ones (Figure 1.2). Fractions instead of crude extracts of bacterial secondary metabolites were generated and used for antimicrobial screening because crude extracts do not perform well compared to the fractioned library in modern high throughput screenings due to the complexity of crude extracts[41]. This platform highlights the utilization of advanced analytical instruments and techniques in order to minimize the amount of analyte and the fermentation scale of the selected bacteria. Starting from microscale fermentation (generally 0.5 – 25 mL scale) for LC/MS PCA analysis, prioritized strains were cultured in small scale (100 mL scale using different media components). The extracts of the small scale fermentation were collected and undergo an automated SPE-LC/MS fractionation to generate 80 fractions collected on 96-well plates, which were sent to the Small Molecule Screening and Synthetis Facility (SMSSF) for high-throughput biological activities screening against *E. coli*, Methicillin-resistant *Staphylococcus aureus* (MRSA), *C. albicans*, and *P. aeruginosa*. The active wells, generally containing <100 μg of semi-pure fraction, were analyzed by 1.7 mm NMR and UHPLC/HRMS instruments to provide spectroscopic data for dereplication searches using databases of antibioitcs (AntiBase[42] and SciFinder). If known antibiotics were responsible for the biological activities, we would deprioritize the producing bacteria and

record the singatures of these compounds (locations on the 96-well plates, NMR and MS spectra) to provide more information for future dereplications. If no matches in the databases can be identified, large scale fermentation would be done (1 L to 20 L scale) to produce enough compounds for structural elucidations, *in vivo* biological activity studies and mechanism of action studies. Biosynthetic mechanism studies on selected compounds were also performed in order to fully understand the biosynthetic potential of the producing organisms. Our platform highlights the strength of having close collaborations between biologists and natural product chemists and has led to the discovery of more than 10 classes of new antibiotic compounds as well as more than 30 classes of known compounds.

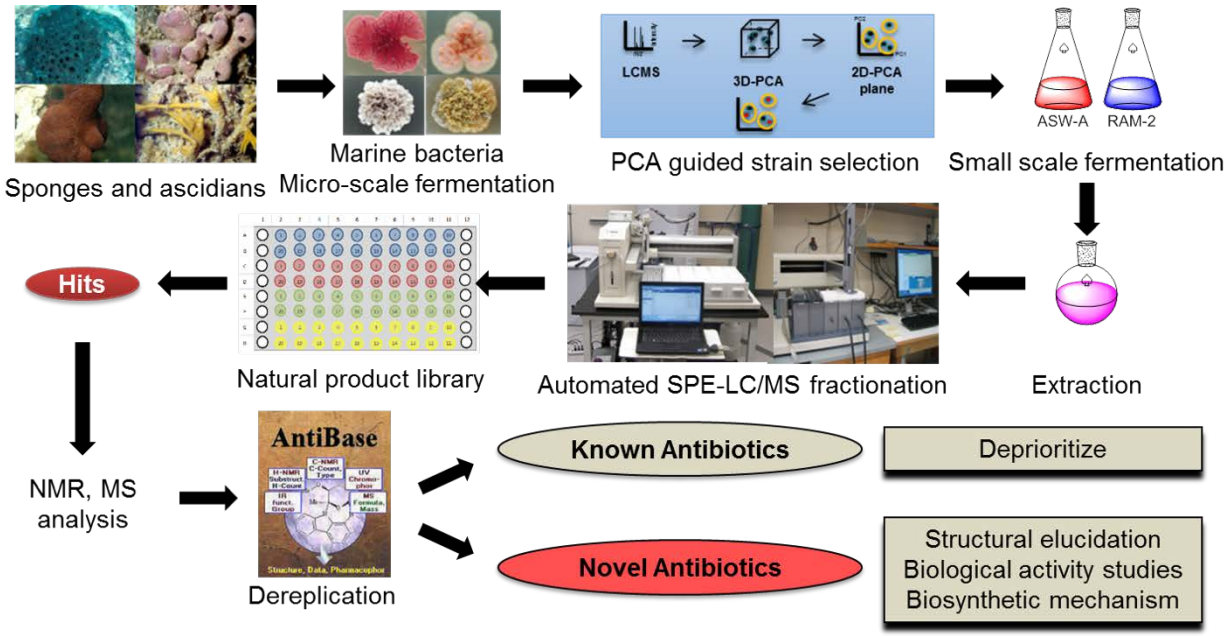
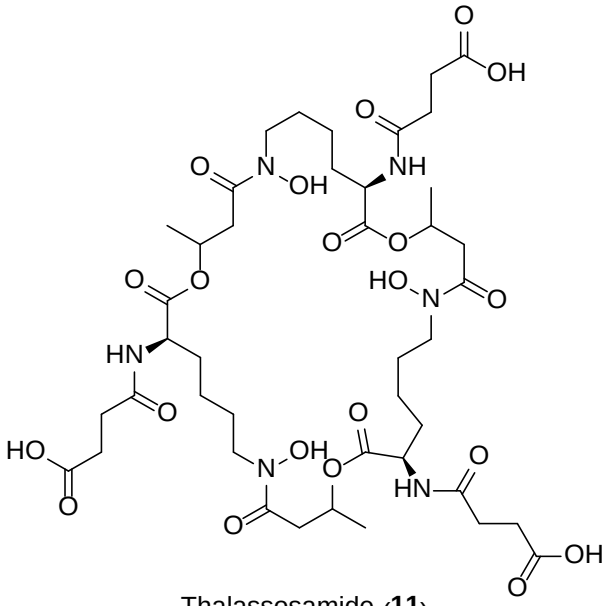


Figure 1.2. Workflow of the marine natural product library based discovery platform. Sponges and ascidians were collected by the professional divers. Bacteria isolation was conducted by Mr. Douglas Braun, the Sr. Research Specialist in Bugni lab. Active hits analyses were done by Dr. Gene Ananiev in the SMSSF and 1.7 mm NMR data were collected Bruker instruments located in the National Magnetic Resonance Facility at Madison (NMRFAM). Biological activity studies were done by Dr. David Andes lab (*in vivo* bioactivity studies using mouse models, antifungal mechanism of action studies), Dr. Bruce Klein lab (antifungal mechanism of action studies), and Dr. Rodney Welch lab (antibacterial mechanism of action studies).

In 2017, our lab published the discovery and biological activity studies of thalassosamide (**11**), an antibacterial siderophore isolated from *Thalassospira profundimaris*[43]. This compound was discovered by first prioritizing the producing bacteria in our PCA analysis and then identifying an active anti-*P. aeruginosa* hit in our natural product library discovery platform. Thalassosamide also showed *in vivo* anti-*P. aeruginosa* activities in the murine thigh infection model. This result indicates the efficiency and effectiveness of our marine natural product discovery platform.



Thalassosamide (**11**)

1.4. Challenges in structure elucidation and new tools to overcome them.

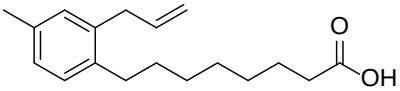
Another challenge in natural product research is the structure elucidation of the newly discovered compounds. Mass spectrometry is one of the most important tools for determining the exact masses and molecular formulae of natural products. Analysis on fragment ions generated from tandem MS techniques can provide more important information for the structure elucidations of natural products[44]. NMR techniques have experienced revolutionary

advances in recent years because of the application of cryoprobes and increased magnetic fields[45]. These advances in NMR instruments have enabled the application of new pulse programs, which have led to the detections of some unseen correlations (for example, long range HMBC correlation) in certain compounds[46]. With the advances in whole genome sequencing techniques, genomic analysis of bacteria tends to give better biosynthetic mechanism proposals of certain bacterial secondary metabolites. The better understanding of biosynthetic mechanism of certatin molecules also benefits the structural elucidations of those challenging structuress[47]. Some recent advances in these three fields (MS, NMR, genomic analysis) will be briefly reviewed in this section.

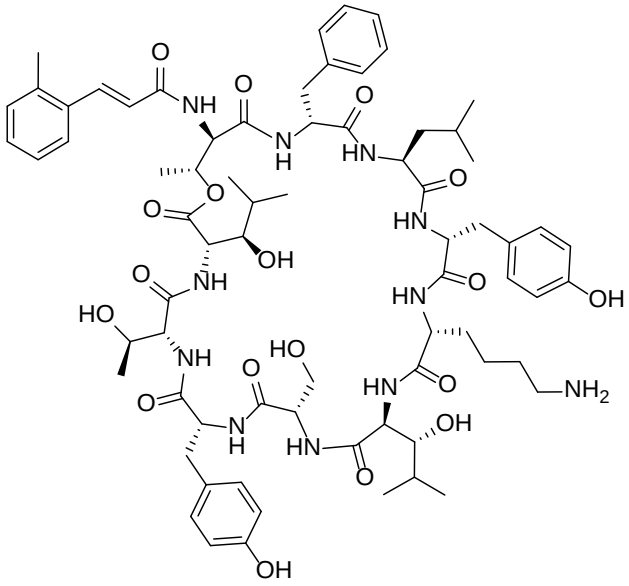
MS instrument with high resolution was one of the revolutionary innovations in the past fifty years. This application of high-resolution MS has enabled the determination of accurate molecular formulae of complicated compounds[48], which can benefit the rapid structure elucidation for those challenging molecules. In recent years, the resolutions of some MS instruments are becoming high enough to study the small differences between ions with different isotopic compositions[49, 50]. Since certain isotopes of different elements have certain natural abundances, the isotopic composition analysis of those compounds can be developed as an unambiguous molecular formula determination method.

Isotopic labeling has always been a powerful tool for natural product research. By labeling certain atoms with isotopes and detecting the locations of those labeled atoms, valuable information about their biosynthetic mechanisms can be collected. When it comes to unknown structures, direct detection of C-C, C-N connectivities can provude important structure information to facilitate structure elucidation of those challenging compounds. The

detection of C-C and C-N connectivities becomes more important when it comes to structure motifs with limited number of hydrogen atoms[51]. However, the direct detection of these correlations usually requires isotopic enrichments of ^{13}C and ^{15}N because of the low natural abundance of these two nuclides (~1% for ^{13}C, ~0.4% for ^{15}N). Our lab has demonstrated the importance of acquiring ^{13}C-^{13}C COSY NMR correlations of ^{13}C labeled compounds to complete the structure elucidations of solwaric acid A (**12**)[52] and forazoline A (**4**)[32]; and ^{13}C-^{15}N HMQC NMR correlations of ^{13}C, ^{15}N labeled compounds to complete the structure elucidation of forazoline A (**4**)[32]. An H-^{13}C-^{15}N 3D NMR experiments were tested on an antibacterial compound eudistamide B (**13**)[53] in order to detect the connectivities among different amino acid residues of complicated peptides.



Solwaric acid A (**12**)



Eudistamide B (**13**)

The advances in whole genome sequencing and better interpretation of functions of biosynthetic clusters of bacteria are making bioinformatics a more and more powerful in predicting the biosynthetic outcomes of bacteria[54]. In recent years, the application of

bioinformatics in the structure elucidation of ribosomally synthesized and post-translationally modified peptides (RiPPs) has been well documented[55]. RiPPs are generally biosynthesized from a ribosomal peptide precursor containing a leader segment and a core segment. The core segment can undergo post-translationally modification catalyzed by the enzymes encoded by the neighboring genes of the RiPPs biosynthetic genes. Further cleavage of the leader peptide segment would release the mature RiPPs with diverse chemical structures and biological activities[56]. By detecting the biosynthetic genes and analyzing the functions of those genes, amino acid residues, amino acid sequences, and the post-translationally modifications of those amino acids in the RiPPs can be determined, leading to the structure determination of those complicated peptides.

Winter and co-workers described the discovery of four RiPPs, the salinipeptins A-D (**14**-**17**)[57], from the Great Salt Lake *Streptomyces* sp. GSL-6C (Figure 1.3). The peptide sequences were suggested by the genomic analysis of the producing organism *Streptomyces* sp. GSL-6C. The post-translationally modifications were determined by different spectroscopic techniques (NMR, MS, etc.) and chemical derivatizations. Salinipeptin A showed moderate anti-bacterial activities against Group A *Streptococcus pyogenes* M1T1 and moderate cytotoxicity against U87 glioblastoma and HCT-116 colon carcinoma cancer cell lines. The successful structural elucidation of these four complicated RiPPs highlights the importance of combining NMR, MS techniques with the bioinformatics analysis.

In 2017, RiPPMiner, an online based bioinformatics program, was developed for the prediction of the amino acid sequences of RiPPs (Figure 1.4)[58]. The algorithm of this program utilized Support Vector Machine to distinguish RiPPs precursors from other small

peptides; and machine learning based classifiers to classify the uploaded sequence into 12 major classes of RiPPs by comparing to a well-documented database of more than 500 experimentally characterized RiPPs. By uploading the predicted precursor peptide sequences of RiPPs, the server can decipher the uploaded sequences by comparing them with the backend database and give predictions of the structures of mature RiPPs by homology search. This interactive online based platform can potentially be the next generation solution of structure elucidations of complicated RiPPs.

Figure 1.3. Structures of salinipeptins A-D. Purple color means the amino acid residues had no post-translationally modifications. Green color means the amino acid residues were post-translationally modified. Yellow means the configurations of the amino acid residues were D-configurations.

1.5. Summary

Therefore, advances in LC/MS based metabolomics analysis, investigations of relatively underexplored bacteria genera, and generation of a streamlined natural product library platform using modern analytical instrumentations are the three major approaches used by our lab to overcome the current high known compound rediscovery problem in new antibiotics discovery from natural products. These methods have led to the discovery of

multiple classes of bioactive compounds, including forazolines (**4**, **5**)[32], halomadurones (**6**-**9**)[33], ecteinamycin (**10**)[34], thalassosamide (**11**)[43], solwaric acids (**12**)[52], and eudistamide B (**13**)[53]. These encouraging results highlight the importance of utilizing new concepts, strategies, and analytical techniques in current natural products research.

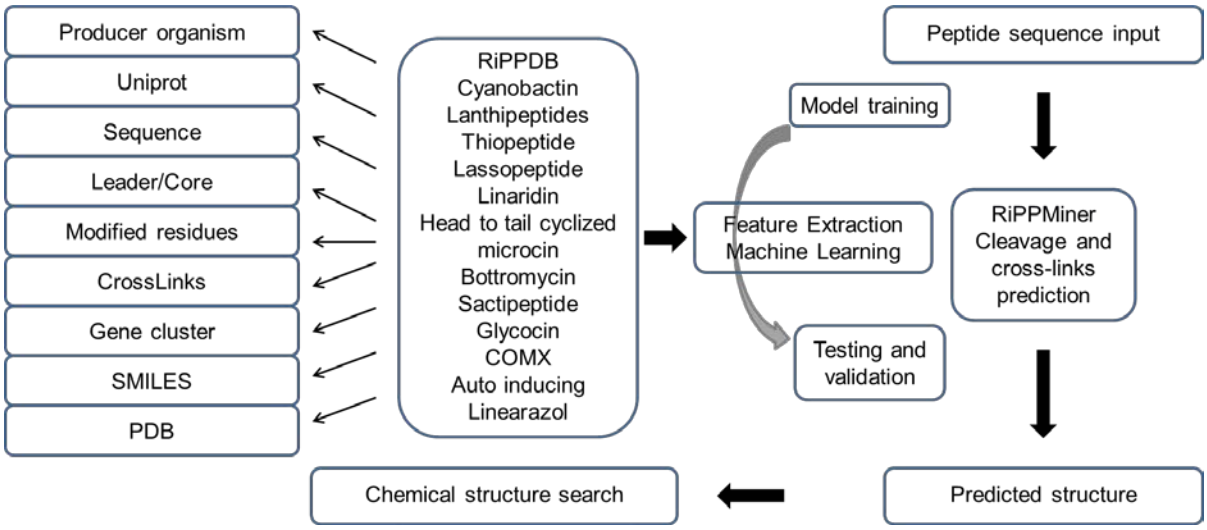


Figure 1.4. Workflow of the RiPPMiner. Starting with the input of RiPPs precursor query sequence, modified core peptide sequence and structures were predicted based on homology search against a backend database.

As for the challenges in the structure elucidations of natural products, scientists have developed new analytical techniques taking advantage of the advances in analytical instruments. Isotopic composition of certain molecule resulted from the mass defects and natural abundance of the nuclides that the molecule consists. The application of high resolution MS instruments has revealed the small differences between ions with different isotopic compositions, which will eventually lead to unambiguous molecular formula determination of natural products. New NMR pulse programs, such as ^{13}C-^{13}C COSY, ^{13}C-^{15}N HMQC on isotopic enriched compounds can provide valuable molecular skeleton information, which can benefit the rapid and accurate strucuture elucidations of some

challenging structural motifs. Last but not the least, advances in bioinformatics is providing another great tool for natural product chemists to complete the structure elucidations of some challenging compounds, such as ribosomally synthesized post-translationally modified peptides (RiPPs). These advances can essentially benefit the whole natural product community and make the structure elucidations of new natural products more smoothly.

1.6. Overview of the Thesis

This thesis will describe the investigations of marine invertebrate-associated bacteria for new antibiotic compounds. LC/MS based metabolomics analysis, investigating relatively underexplored bacteria genera, and generating a marine bacteria natural product library platform are the three major methods used to achieve the discovery of 12 new compounds with different biological activities. The structures of these compounds were analyzed with a variety of analytical techniques, including but not limited to 1D NMR, 2D NMR, qTOF-ESI MS, Magnetic Resonance Mass Spectrometry (MRMS), molecular modeling, DFT calculations, and bioinformatics analysis. A class of antibacterial madurastatin compounds were discovered and the structure elucidations were conducted. The absolute configurations of madurastatins were determined by Marfey's analysis, molecular modeling and DFT calculations. Whole genome sequencing and genomic analysis of the producing organism resulted in a possible biosynthetic mechanism proposal for the madurastatin compounds for the first time. Another class of antibacterial compounds, bacillimidazoles A-F, was discovered from relatively underexplored marine bacterium *Bacillus* sp. WMMC-1349. These compounds feature an uncommon positively charged imidazolium ring structure. Carbon-13 isotopic labeling of bacillimidazoles was performed in order to understand the biosynthetic

mechanisms of these compounds. High resolution MRMS instrument was used to study the isotopic fine structure (IFS) of natural products in order to achieve the unambiguous molecular formulae determinations. This method together with isotopic labeling was applied to the rapid structure elucidation of ecteinamines A and B, two novel compounds isolated from WMMB-482. The IFS analysis has also assisted the structure elucidation of two peptides, one nonribosomal peptide (NRP) streptnatamide A, and one ribosomally synthetized post-translationally modified peptide (RiPP) haliclonamycin.

1.7. References

(1). Cragg, G. M.; Newman, D. J. *Pure Appl. Chem.*, **2005**, *77*, 7-24.

(2). Newman, D. J.; Cragg, G. M. *J. Nat. Prod.*, **2012**, *75*, 311-335.

(3). Newman, D. J.; Cragg, G. M.; In *Functional Molecules from Natural Sources*; Wrigley, S. K; Thomas, R.; Nicholson, N.; Bedford, C., Eds.; RSC Publications: Cambridge, UK, 2010; pp 3-36.

(4). Wright, G. D. *Nat. Prod. Rep.*, **2017**, *34*, 694-701.

(5). Brown, E. D.; Wright, G. D. *Nature*, **2016**, *529*, 336-343.

(6). Fleming, A.; Chain, E. B.; Florey, H.; *Sir Alexander Fleming Nobel Lecture: Penicillin, Nobel Lectures*, 1964.

(7). Barnes, E. C.; Kumar, R.; Davis, R. A. *Nat. Prod. Rep.*, **2016**, *33*, 372-381.

(8). Walsh, C; *Antibiotics: Actions, Origins, Resistance.* ASM Press, Washington, DC, 2003; pp 11-88.

(9). Lewis, K.; *Nature*, **2012**, *485*, 439-440.

(10). Silver, L. L. *Bioorg. Med. Chem.*, **2016**, *24*, 6379-6389.

(11). O'Shea, R.; Moser, H. E. *J. Med. Chem.*, **2008**, *51*, 2871-2878.

(12). Bérdy, J. *J. Antibiot.*, **2012**, *65*, 385-395.

(13). Clatworthy, A. E.; Pierson, E.; Hung, D. T. *Nat. Chem. Biol.*, **2007**, *3*, 541-548.

(14). Shah, N. S.; Auld, S. C.; Brust, J. C. M.; Mathema, B.; Ismail, N.; Moodley, P.; Milsana, K.; Allana, S.; Campbell, A.; Mthiyane, T.; Morris, N.; Mpangase, P.; van der Meulen, H.; Omar, S. V.; Brown, T. S.; Narechania, A.; Shaskina, E.; Kapwata, T.; Kreiswirth, B.; Gandhi,

N. R.; *N. Engl. J. Med.*, **2017**, *376*, 243-253.

(15). Walsh, C; *Antibiotics: Actions, Origins, Resistance.* ASM Press, Washington, DC, 2003; pp 89-135.

(16). Davies, J.; Davies, D.; *Microbiol. Mol. Biol. Rev.*, **2010**, *74*, 417-433.

(17). Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. *Nat. Rev. Drug Discov.*, **2007**, *6*, 29-40.

(18). Watve, M. G.; Tickoo, R.; Jog, M. M.; Bhole, B. D.; *Arch. Microbiol.*, **2001**, *176*, 386-390.

(19). Faulkner, D. J. *Nat. Prod. Rep.*, **2001**, *18*, 1-49.

(20). Lane, A. L.; Moore, B. S. *Nat. Prod. Rep.*, **2011**, *28*, 411-428.

(21). Paul, C.; Pohnert, G. *Nat. Prod. Rep.*, **2011**, *28*, 186-195.

(22). Fusetani, N. *Nat. Prod. Rep.*, **2011**, *28*, 400-410.

(23). Zäehner, H.; Fiedler, H.P. The Need for New Antiobiotics. *In Fifty Years of Antimicrobials*. Hunter, P.A.; Darby, G.K.; Russel, N.J., Eds.; Cambridge University Press: Cambridge, England, 1995; 67-84.

(24). Lang, G.; Mayhudin, N. A.; Mitova, M. I.; Sun, L.; van der Sar, S.; Blunt, J. W.; Cole, A. L.; Ellis, G.; Laattsch, H.; Munro, M. H. *J. Nat. Prod.*, **2008**, *71*, 1595-1599.

(25). Williamson, R. T.; Chapin, E. L.; Carr, A. W.; Gilbert, J. R.; Graupner, P. R.; Lewer, P.; McKamey, P.; Carney, J. R.; Gerwick, W. H. *Org. Lett.*, **2000**, *2*, 289-292.

(26). Earl, A. M.; Losick, R.; Kolter, R. *J. Bacteriol.*, **2007**, *189*, 1163-1170.

(27). Maughan, H.; van der Auwera, G. *Infect., Genet. Evol.*, **2011**, *11*, 789-797.

(28). Krug, D.; Zurek, G.; Schneider, B.; Garcia, R.; Müller, R. *Anal. Chim. Acta.*, **2008**, *624*, 97-106.

(29). Nielsen, K. F.; Månsson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O. *J. Nat. Prod*. **2011**, *74*, 2338-2348.

(30). Hou, Y.; Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W.; Bugni, T. S. *Org. Lett.* **2012**, *14*, 5050-5053.

(31). Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni,

T. S. *Anal. Chem.*, **2012**, *84*, 4277-4283.

(32). Wyche, T. P.; Piotrowski, J. S.; Hou, Y.; Braun, D.; Deshpande, R.; Mcllwain, S.; Ong, I. M.; Myers, C. L.; Guzei, I. A.; Westler, W. M.; Andes, D. R.; Bugni, T. S. *Angew. Chem. Int. Ed.* **2014**, *53*, 11583−11586.

(33). Wyche, T. P.; Standiford, M.; Hou, Y.; Braun, D.; Johnson, D. A.; Johnson, J. A.; Bugni, T. S. *Mar. Drugs* **2013**, *11*, 5089−5099.

(34). Wyche, T. P.; Alvarenga, R. F. R.; Piotrowski, J. S.; Duster, M. N.; Warrack, S. R.; Cornilescu, G.; De Wolfe, T. J.; Hou, Y.; Braun, D. R.; Ellis, G. A.; Simpkins, S. W.; Nelson, J.; Myers, C. L.; Steele, J.; Mori, H.; Safdar, N.; Markley, J. L.; Rajski, S. R.; Bugni, T. S. *ACS Chem. Biol.*, **2017**, *12*, 2287-2295.

(35). Hou, Y.; Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W.; Bugni, T. S. *Org. Lett.*, **2012**, *14*, 5050-5053.

(36). Zhang, Y.; Adnani, N.; Braun, D. R.; Ellis, G. A.; Barns, K. J.; Parker-Nance, S.; Guzei, I. A.; Bugni, T. S. *J. Nat. Prod.*, **2016**, *79*, 2968-2972.

(37). Wyche, T. P.; Hou, Y.; Vazquez-Rivera, E.; Braun, D.; Bugni, T. S. *J. Nat. Prod.*, **2012**, *75*, 735-740.

(38). Shin, B.; Kim, B. –Y.; Cho, E.; Oh, K. –B.; Shin, J.; Goodfellow, M.; Oh, D. –C. *J. Nat. Prod.*, **2016**, *79*, 1886-1890.

(39). Subramani, R., Sipkema, D. *Mar. Drugs*, **2019**, *17*, 249-

(40). Kaspar, F.; Neubauer, P. Gimpel, M. *J. Nat. Prod.*, **2019**, *82*, 2038-2053.

(41). Bugni, T. S.; Harper, M. K.; McCulloch, M. W. B.; Whitson, E. L. *Natural Product Chemistry for Drug Discovery.* Buss, A. D.; Butler, M. S. Eds.; RSC Press, Cambridge, UK, pp 276-277.

(42). Laatsch, H. *AntiBase, A Data Base for Rapid Dereplication and Structure Determination of Microbial Natural Products*; Wiley-VCH: Weinheim, Germany, 2012.

(43). Zhang, F.; Barns, K.; Hoffmann, F. M.; Braun, D. R.; Andes, D. R.; Bugni, T. S. *J. Nat. Prod.* **2017**, *80*, 2551−2555.

(44). Konishi, Y.; Kiyota, T.; Draghici, C.; Gao, J.-M.; Yeboah, F.; Acoca, S.; Jarussophon, S.; Purisima, E. *Anal. Chem.* **2007**, *79*, 1187–97.

(45). Seco, J.M.; Quinoa, E.; Riguera, R. *Chem. Rev.* **2004**, *104*, 17-117.

(46). Williamson, R. T.; Buevich, A. V.; Martin, G. E.; Parella, T. *J. Org. Chem.*, **2014**, *79*, 3887-3894.

(47). Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Research*, **2011**, *39*, W339–W346.

(48). Rose, C. M.; Merrill, A. E.; Bailey, D. J.; Hebert, A. S.;Westphall, M. S.; Coon, J. J. *Anal. Chem.*, **2013**, *85*, 5129-5137.

(49). Yu, X.; Zhong, W. *Anal. Chem.*, **2016**, *88*, 5914-5919.

(50). Liang, Z.; Yu, X.; Zhong, W. *Anal. Chem.*, **2019**, *91*, 4381-4387.

(51). Bifulco, G.; Riccio, R.; Martin, G.E.; Buevich, A.V.; Williamson, R.T. *Org. Lett.* **2013**, *15*, 654-657.

(52). Ellis, G. A.; Wyche, T. P.; Fry, C. G.; Braun, D. R.; Bugni, T. S. *Mar. Drugs*, **2014**, *12*, 1013-1022.

(53). Zhang, F.; Adnani, N.; Vazquez-Rivera, E.; Braun, D. R.; Tonelli, M.; Andes, D. R.; Bugni, T. S. *J. Org. Chem.*, **2015**, *80*, 8713-8719.

(54). Walsh, C. T.; Fischbach, M. A. *J. Am. Chem. Soc.*, **2010**, *132*, 2469-2493.

(55). Ortega, M. A., van der Donk, W. A. *Cell Chem. Biol.*, **2016**, *23*, 31-44.

(56). Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., van der Donk, W. A. *Nat. Prod. Rep.*, **2013**, *30*, 108−160.

(57). Shang, S.; Winter, J. M.; Kauffman, C. A.; Yang, I.; Fenical, W. *ACS Chem. Biol.*, **2019**, *14*, 415-425.

(58). Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. *Nucleic Acids Research*,

2017, *45*, W80-W88.

Chapter 2:

Madurastatin D1 and D2, oxazoline containing siderophores isolated from an *Actinomadura sp.*

Portions of this chapter have been published as:

Yan, J. –X.; Chevrette, M. G.; Braun, D. R.; Harper, M. K.; Currie, C. R.; Bugni, T. S. Madurastatin D1 and D2, oxazoline containing siderophores isolated from an *Actinomadura sp.* Org. Lett. **2019**, *21*, 6275−6279.

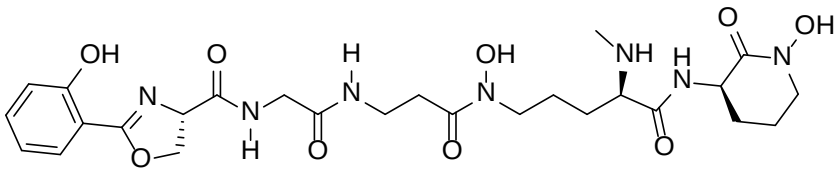
2.1. Introduction

As discussed in Chapter 1, metabolomics analysis guided bacteria strain prioritization coupled with investigations of relatively underexplored bacteria genera can potentially maximize the chance to discover new chemical entities among the abundance of known compounds. In this chapter, the discovery, structural elucidation, and biosynthetic mechanism studies of two new antibacterial siderophores, madurastatin D1 and D2 is described.

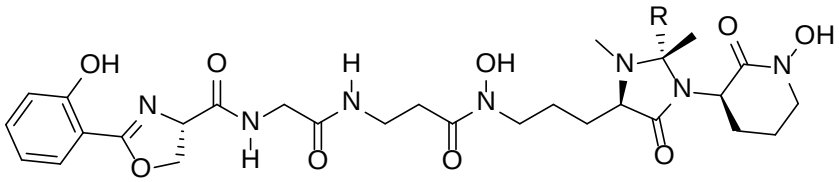
Siderophores represent a crucial class of bacterial secondary metabolites that are employed as iron chelators by microorganisms to facilitate the absorption of poorly soluble environmental iron[1]. Siderophores feature high binding affinities toward Fe(III) and often contain multiple N, O functionalities[1]. They can also be employed by pathogenic bacteria to acquire iron, the fundamental limiting nutrient for life, from their host organisms and influence the cellular iron level of the hosts[2].

The bacterial genus *Actinomadura* from the phylum Actinobacteria has been reported to produce a series of phenolate-hydroxamate siderophores[2], including madurastatins[3-5] and

maduraferrin[6]. The madurastatins were first characterized as aziridine-containing pentapeptide siderophores; however, structural revision on the aziridine ring was suggested by Thorson and Shaaban[7] based on their analyses of siderophores with 2-(2-hydroxyphenyl)oxazoline moieties. NMR data of the isolated madurastatin C1 (also designated as MBJ-0034) from bacteria cell culture and synthetic analogs with aziridine- and 2-(2-hydroxyphenyl)oxazoline moieties were compared by Hall in 2017[8], which confirmed the structural revision proposed by Thorson and Shaaban. The absolute configuration of madurastatin C1 (**18**) (Figure 2.1) was also established by Hall[8].



(−)-madurastatin C1 (**18**)



R = H, madurastatin D1 (**19**)
R = CH_3, madurastatin D2 (**20**)

Figure 2.1. Structures of madurastatin C1 (**18**), D1 (**19**) and D2 (**20**).

As part of our lab's goal in discovering new natural products from marine invertebrate associated bacteria[9-11], our attention was drawn to strain WMMA-1423, a marine *Actinomadura* sp. cultivated from the sponge *Tedania* sp., after its crude extract showed inhibition on the growth of both MRSA and *Bacillus subtillis* on agar plates. We isolated two new madurastatins, madurastatin D1 (**19**) and D2 (**20**) (Figure 2.1), as well as the enantiomer [**18**, (−)-madurastatin C1] of the known compound madurastatin C1 by bioassay guided

isolation using MRSA (ATCC #33591) as the test strain. Madurastatin D1 (**19**) and D2 (**20**) feature the cyclic 4-imidazolidinone in contrast to the linear -amino amide in madurastatin C1. Both **19** and **20** exhibited lower MICs against *M. luteus* than **18**, suggesting that the additional heterocycle increases antibacterial activity. Due to the unusual cyclization bridging two nitrogens, whole genome sequencing of strain WMMA-1423 was conducted and a biosynthetic pathway to madurastatins D1 and D2 was proposed.

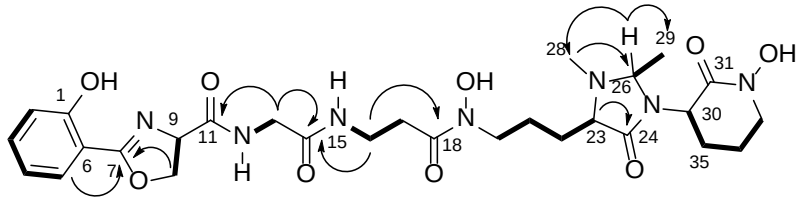
2.2. Results and Discussion

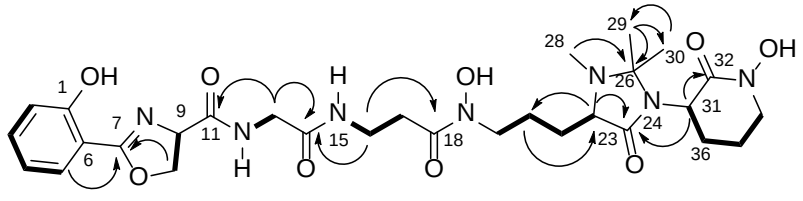
HRMS data suggested the molecular formula of $C_{28}H_{39}N_7O_9$ (*m/z* = 618.2876, $[M+H]^+$) for madurastatin D1 (**19**). Since Fe has a unique isotope distribution (^{54}Fe:^{56}Fe:$^{57}Fe \approx 6:92:2$), the observation of Fe(III) adduct ions $[M-2H+Fe]^+$ (*m/z* = 671.1991, Figure S14) was a clear indicator of compound **19**'s ability of binding iron and further confirmation with regard to the assigned molecular ion. Compared to the known madurastatin C1, **19** had one additional degree of unsaturation. By comparing of ^1H and ^{13}C NMR data (Table 2.1, Figure S3−S7) between the reported madurastatin C1[8] and **19**, The tertiary carbon C-26 (74.7 ppm) and the methyl group carbon C-29 (19.7 ppm) were observed only in **19**, which have HSQC correlation to H-26 (4.01 ppm) and H-29 (1.21 ppm), respectively. H-26 showed COSY correlations to H-29, and it also showed HMBC correlations to C-28 (37.9 ppm, N-Me group attached to N-25) and to C-29, which suggested the connection between N-25 and C-26. Considering all of the above mentioned observations, the structure of **19** was proposed with the 1, 2-dimethyl-4-imidazolidinone cyclic structure.

A molecular formula of $C_{29}H_{41}N_7O_9$ (*m/z* = 632.3025, $[M+H]^+$) was suggested by HRMS data for madurastatin D2 (**20**). Fe(III) adduct ions $[M+Fe+Na-3H]^+$ (*m/z* = 707.1941)

were also observed (Figure S15). The NMR data (Table 2.1, Figure S8−S12) of **20** was compared to **19** and since these two compounds have the same unsaturation degrees, a similar 4-imidazolidinone cyclic structure was also proposed for **20** despite a few differences. The major difference was the tertiary carbon C-26 in **19** was replaced by a tetrasubstituted carbon C-26 (77.9 ppm) in **20**. Two methyl groups (C-29 and C-30) showed HMBC correlations to C-26 and each other as well, which suggested both of the methyl groups were attached to C-26. These two methyl groups also showed different ^{13}C chemical shifts (25.7 ppm for C-29, 19.8 ppm for C-30) and ^{1}H chemical shifts (1.29 for H-29, 1.11 for H-30), which suggested both of them are connected to a carbon that belongs to a cyclic system. Thus, the structure of **20** was proposed with a similar 1, 2, 2-trimethyl-4-imidazolidinone cyclic structure.

Table 2.1. ^{1}H and ^{13}C NMR data for **19** and **20** (600 MHz for ^{1}H, 125 MHz for ^{13}C, d_6-DMSO)





COSY ⟶ HMBC (key)

Position	**19**		**20**	
	δ_C, type	δ_H, (J in Hz)	δ_C, type	δ_H, (J in Hz)
1	159.1		159.1	
2	116.7, CH	7.01, d (8.4)	116.7, CH	6.98, d (7.6)
3	134.1, CH	7.47, ddd (8.4, 7.3, 1.6)	134.1, CH	7.45, t (7.6)
4	119.1, CH	6.95, t (7.5)	119.0, CH	6.91, t (7.6)

5	128.1, CH	7.64, dd (7.6, 1.4)	128.1, CH	7.63, d (7.4)
6	110.0		110.0	
7	165.9		165.9	
9	67.5, CH	5.01, dd (10.4, 7.7)	67.4, CH	5.00, dd (10.2. 8.0)
10	69.5, CH_2	4.65, dd (10.3, 8.5) 4.51, t (8.0)	69.4, CH_2	4.63, dd (10.2, 8.1) 4.51, t (8.3)
11	170.2		170.2	
12		8.61, t (5.9)		8.56, s
13	42.2, CH_2	3.73, dd (16.6, 6.0) 3.67, dd (16.6, 6.0)	42.2, CH_2	3.75, d (16.0) 3.66, d (16.0)
14	168.4		168.4	
15		8.01, t (5.4)		7.97, s
16	34.7, CH_2	3.24, dd (6.6, 6.0)	34.8, CH_2	3.25, dd (7.0, 5.6)
17	31.9, CH_2	2.56-2.48, m	31.7, CH_2	2.50-2.55, m
18	170.8		170.6	
20	47.4, CH_2	3.51-3.42, m	51.1, CH_2	3.41-3.53, m
21	21.6, CH_2	1.70-1.43, m	20.9, CH_2	1.36-1.46, m
22	26.6, CH_2	1.53-1.43, m 1.98-1.82, m	25.0, CH_2	1.85-1.94, m
23	64.8, CH	2.88, m	62.0, CH	2.93, t (5.0)
24	171.6		169.9	
26	74.7, CH	4.01, q (5.0)	77.9	
28	37.9, CH_3	2.29, s	32.5, CH_3	2.24, s
29	19.7, CH_3	1.21, d (5.2)	25.7, CH_3	1.29, s
30	51.8, CH	4.32-4.24, m	19.8, CH_3	1.11, s
31	162.4		51.7, CH	3.91, dd (6.0, 5.4)
32			163.3	
33	51.1, CH_2	3.51-3.57, m 3.41-3.46, m		
34	21.6, CH_2	1.87-1.79, m	47.5, CH_2	3.43-3.47, m
35	21.3, CH_2	2.03-1.95, m	21.3, CH_2	1.67-1.73, m 1.84-1.93, m
36			25.5, CH_2	2.34-2.42, m 1.67-1.73, m

The siderophoric properties of compounds **18**–**20** were assessed on the basis of the

chrome azurol S (CAS) assay (Figure 2.2).[9] Deferoxamine mesylate was also analyzed and served as a positive Fe(III) binding control. The colorimetric CAS assay revealed that **18** and **19** bind iron with efficiencies comparable to deferoxamine mesylate whereas compound **20**, bearing the geminal-dimethyl moiety also bound iron but with noticeably impaired affinity relative to **18** and **19**. This trend in iron binding supports the notion that C-26 dimethylation imparts conformational limitations to **20** that are lacking in compounds **18** and **19**.

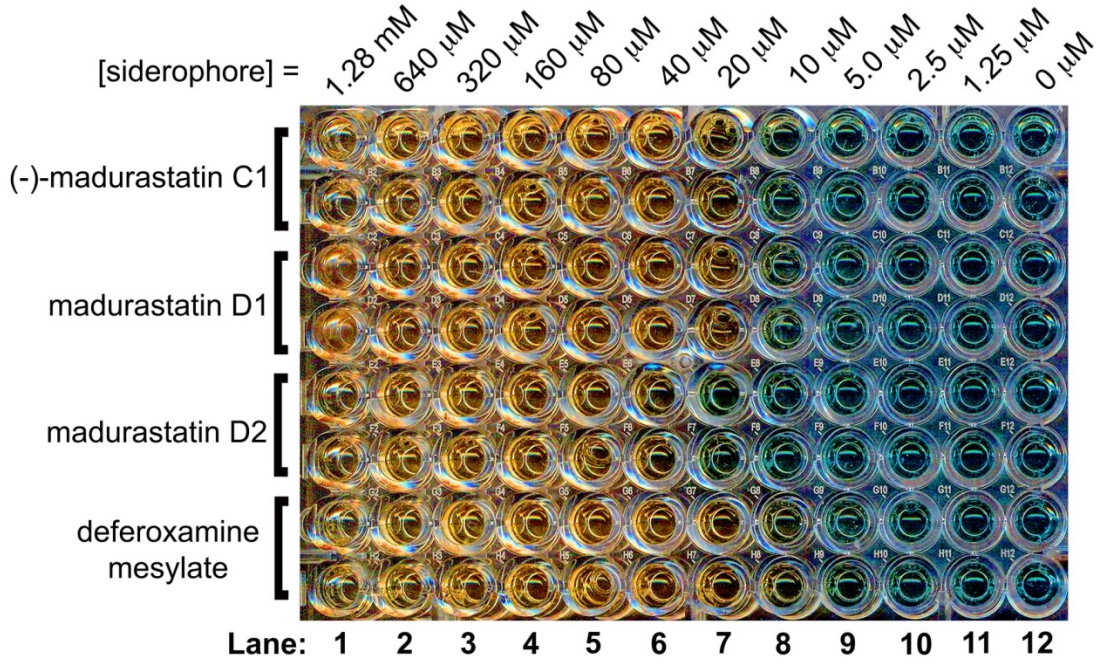
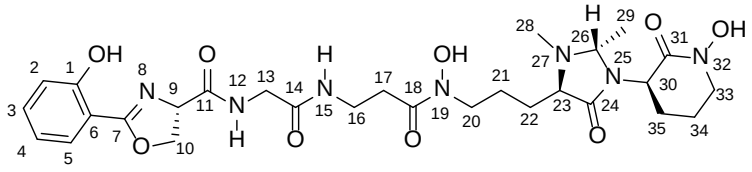


Figure 2.2. Chrome Azurol S (CAS) Assay of Compounds **18**–**20**

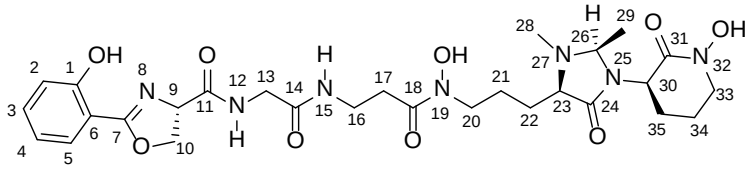
The absolute configuration of madurastatin C1 was determined by Hall and coworkers[8] using the Marfey's method (Figure S16).[12] The ^{13}C chemical shifts (Figure S2) of **18** were identical to those reported for madurastatin C1 although the optical rotation of **18** was found to be inverted from that of madurastatin C1. The C-9 configuration of madurastatin C1 was shown to be *9R* by Marfey's method.[8,12] Thus, we determined that **18** is the enantiomer of madurastatin C1 and has stereocenter assignments as *9S*, *23R*, *28R*. Notably, enantiomeric

natural product pairings are well known and have been rigorously reviewed.[13] Since **2** and **3** were produced by the same organism, both compounds were proposed to have the same configuration as found in **1** [*9S*, *23R*, *30R* (*31R* for **3**)]. Further efforts to determine the configuration of C-26 in **2** employed molecular modeling and density functional theory (DFT) NMR calculations of the two models (*23R*, *26R*, *30R*- and *23R*, *26S*, *30R*-). Spartan 14 (v.1.1.7 Wavefunction Inc. 2014) was used to identify the conformer Boltzmann distribution for each diastereoisomer using molecular mechanics. Gaussian 09 was then used to optimize the local geometry of the low energy conformers based on DFT energy calculation (B3LYP/6-31G(d,p)), and the NMR chemical shifts of different stereoisomers were calculated using GIAO method (Table 2.2). NMR chemical shifts were referenced to TMS and benzene using the multistandard method.[14] The DP4 probability method[15] was used to compare the calculated ^{13}C NMR chemical shifts of the two models with the experimental ^{13}C chemical shifts and yielded a 100% probability for the *R* compared to the *S* configuration (Table 2.3).

Table 2.2. DFT Calculated ^{13}C NMR Chemical Shifts of Madurastatin D1 (**19**) diastereoisomers models.



model 1



model 2

Position	Experimental	Model 1		Model 2	
		Calcd. (ppm)	δ (ppm)	Calcd. (ppm)	δ(ppm)
1	159.1	161.0	-1.9	160.6	-1.5
2	116.7	117.2	-0.5	117.7	-1.0

3	134.1	133.0	1.1	134.1	0.0
4	119.1	118.4	0.7	119.5	-0.4
5	128.1	130.7	-2.6	130.7	-2.6
6	110	114.2	-4.2	113.9	-3.9
7	165.9	160.0	5.9	162.0	3.9
9	67.5	68.7	-1.2	69.5	-2.0
10	69.5	71.3	-1.8	70.1	-0.6
11	170.2	168.1	2.1	172.3	-2.1
13	42.2	43.1	-0.9	47.9	-5.7
14	168.4	172.4	-4.0	168.2	0.2
16	34.7	39.3	-4.6	41.2	-6.5
17	31.9	33.9	-2.0	37.0	-5.1
18	170.8	171.6	-0.8	172.1	-1.3
20	47.4	46.1	1.3	48.8	-1.4
21	21.6	25.8	-4.2	25.5	-3.9
22	26.6	26.2	0.4	31.0	-4.4
23	64.8	65.3	-0.5	67.2	-2.4
24	171.6	176.7	-5.1	173.8	-2.2
26	74.7	78.9	-4.2	79.4	-4.7
28	37.9	35.5	2.4	39.9	-2.0
29	19.7	17.3	2.4	21.4	-1.7
30	51.8	58.9	-7.1	57.4	-5.6
31	162.4	165.3	-2.9	164.3	-1.9
33	51.1	52.8	-1.7	51.5	-0.4
34	21.6	25.1	-3.5	25.0	-3.4
35	20.9	26.6	-5.7	27.1	-6.2

Table 2.3. DP4 Probability calculation of Madurastatin D1 (**19**) diastereoisomers models

Models	Configuration	DP4 Probability
1	*9S, 23R, 26S, 30R*	0%
2	*9S, 23R, 26R, 30R*	100%

Previous reports have shown that madurastatin C1 inhibits the growth of *M. luteus.*[4] Therefore, compounds **18**–**20** were tested for antibacterial activity against *M. luteus*. Notably, compounds **19** and **20** both inhibited *M. luteus* with the MICs of 25.3 and 25.8 μM,

respectively; the MIC of compound **18**, devoid of the imidazolidinone, was 108.1 μM. The presence of the 4-imidazolidinone clearly benefits the antibacterial activity of these agents although efforts to elucidate the specific mechanisms at play are beyond the scope of this report. Noteworthy is that, although MRSA proved important during bioassay-guided fractionation, none of the new madurastatins inhibited MRSA with MICs less than 100 μM.

To identify the biosynthetic cluster for madurastatin biosynthesis, the whole genome of *Actinomadura sp.* WMMA-1423 was sequenced (GenBank accession number CP041244) and analyzed using antiSMASH (version 5.0.0 rc1).[16-17] Two gene clusters (57.9 kb and 46.7 kb, respectively) were identified as the nonribosomal peptide synthetases (NRPSs), transport and regulatory genes for (−)-madurastatin C1 (**18**), madurastatin D1 (**19**) and madurastatin D2 (**20**) production. Although antiSMASH identified these as separate clusters, they were found to be co-localized in the genome (12.1 kb apart from each other) and thus, represent one cluster. The architecture and annotation of the madurastatin (*mad*) biosynthetic gene cluster shown in Figure 2.3 and Table S1. The *mad* cluster consists of two NRPS genes (Figure 2.3a, red, *mad*30, *mad*63), two chain initiation genes (red, *mad*31, *mad*60), two amino acid tailoring genes (red, *mad*28, *mad*61), one *S*-adenosyl-methionine (SAM)-dependent methyltransferase gene (red, *mad*11) as well as additional biosynthetic (orange), transport (teal), regulation-related (green) and other genes (grey); full annotations are provided in Table S1). Most of the proteins encoded in the *mad* cluster are homologous to those involved in cahuitamycin,[18] albachelin,[19] and amychelin[20] biosynthetic pathways. The genome analysis has allowed us to put forth a hypothesis with regard to the possible biosynthesis of madurastatins, where a SAM-dependent mechanism is involved in the biosynthesis of

madurastatin D2 (Figure 2.3b, Table S1). Further biosynthetic studies to more completely understand the assembly of **18**–**20**, with special emphasis on the unique imidazolidinone of **19** and **20**, are clearly warranted. Mad28, Mad61 are likely ornithine N-monooxygenase and aspartate 1-decarboxylase enzymes, responsible for generating the L-*N*-OH-Orn (L-hOrn) and β-alanine moieties, respectively. Mad31 is homologous to salicylate synthase AmcL in amychelin biosynthesis (59% identity) and Mad60 is homologous to the salicylate-AMP ligase CahJ (67% identity) in cahuitamycin biosynthesis. Together, Mad31 and Mad60 likely catalyze the formation of a Mad63-salicylate conjugate starting from chorismate. Mad63 contains one set of C, A, T domains and very likely installs the oxazoline ring since it shows 58% identity to CahA; CahA contains a cyclization domain that creates the oxazoline ring in cahuitamycins. We envision that the growing chain is further extended by Mad30, which contains four sets of C, A, T domains and one N-methylation (nMT) domain; antiSMASH analysis revealed that two of the C domains function as $^{D}C_{L}$ domains (Fig. 2.3c). The A domains in Mad30 are proposed to activate Gly, β-Ala and L-hOrn substrates. We envision that epimerization of each L-hOrn, accounting ultimately for the R stereocenters found in **18**–**20**, likely results from the dual epimerase/condensation activities of the $^{D}C_{L}$ domains[21]. The nMT domain located between the third and fourth A domains is envisioned to catalyze both mono- and dimethylation of the α-amine group of the initially added D-hOrn. Both mono- and dimethylated intermediates would undergo further condensation to add on the final hOrn unit. Intramolecular nucleophilic substitution is then proposed to install the *N*-hydroxy lactam (with liberation from Mad30) to afford (−)-madurastatin C1 (**18**) and the dimethylated variant of **18**; both are possible intermediates *en route* to madurastatin D1 (**19**) although we favor the

dimethylated species based on its complete absence during the course of compound isolations.

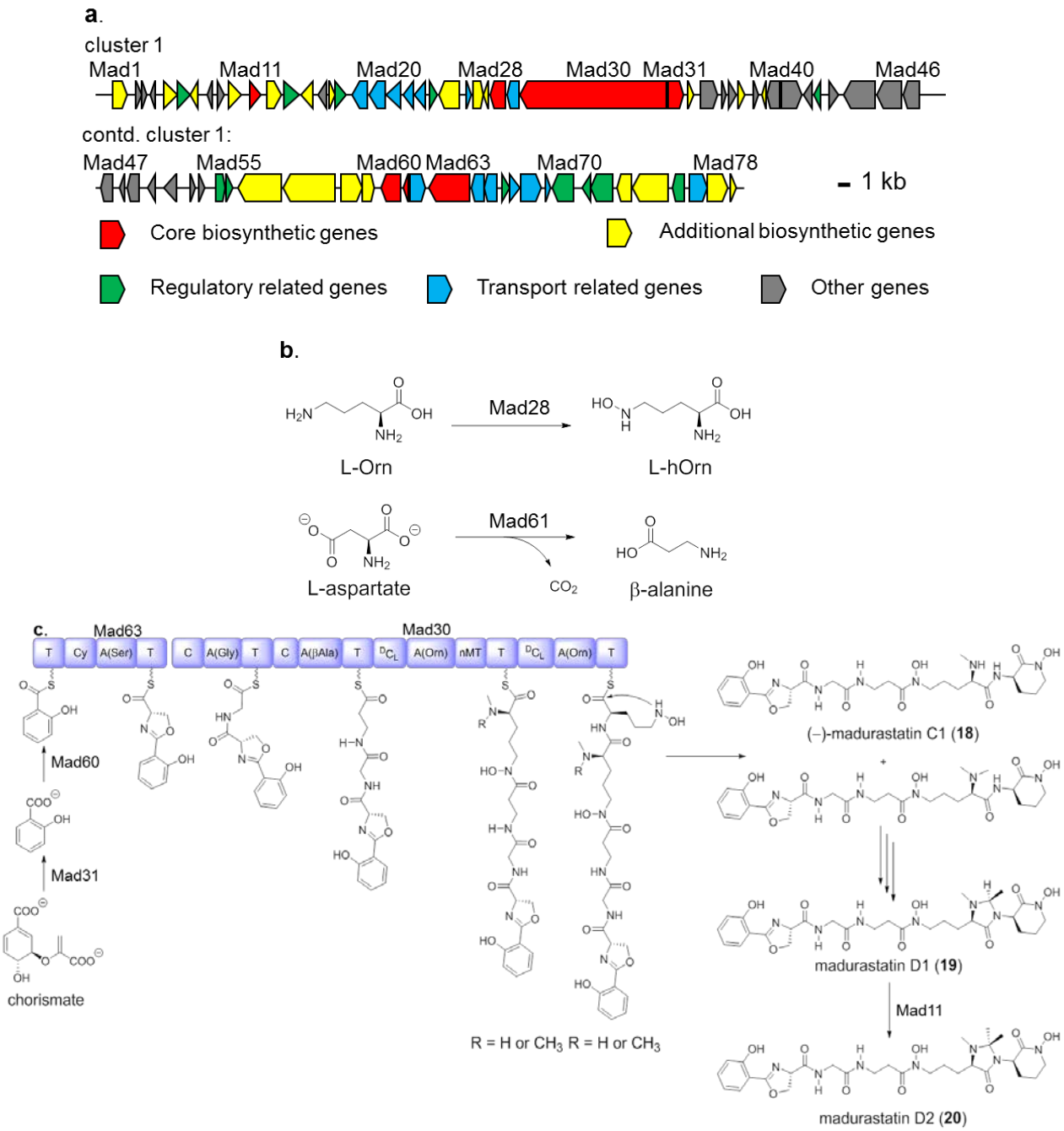


Figure 2.3. (**a**) Organization of *mad* cluster. (**b**) Proposed functions of Mad28 and Mad61 as building block producers, and (**c**), Proposed biosynthetic pathway to **18**–**20**. Domain annotations: A, adenylation; C, condensation; Cy, cyclization; $^{D}C_{L}$, epimerization/condensation; T, thiolation; nMT, N-methylation.

We envision that N-Me oxidation may play a central role in imidazolidinone assembly, from the dimethylated version of **18**, although it is not yet clear which gene/s in the *mad* cluster would code for this chemistry (Figure 2.3c). In proceeding from **19** to **20**, we invoke the SAM-dependent methyltransferase Mad11; SAM-dependent methylation at both sp^2 and sp^3 carbon adjacent to heteroatoms are well known.[22-25] At present, Mad11 appears to be the most likely candidate for converting monomethylated **19** into dimethylated madurastatin D2 (**20**).

2.3. Conclusion

In summary, we report the isolation and structural elucidation of madurastatin D1 (**19**) and D2 (**20**), two new madurastatin siderophores that showed *in vitro* activity against both *M. leuteus* and MRSA. To the best of our knowledge, only five other madurastatin analogs[3-5] have been isolated from *Actinomadura* sp., and none of those five compounds had the 4-imidazolidinone cyclic moiety. A SAM dependent, radical based cyclization is proposed for this unusual cyclization. While scientifically intriguing, further studies will be required to test this biosynthetic hypothesis.

2.4. Material and Methods

General Experimental Procedures

Optical rotations were measured on a Perkin–Elmer 241 Polarimeter. UV spectra were recorded on an Aminco/OLIS UV-Vis spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR spectrophotometer. NMR spectra were obtained in d_6-DMSO (δ_H 2.50 ppm, δ_C 39.50 ppm) with a Bruker Avance 600 III MHz spectrometer equipped with a $^1H\{^{13}C/^{15}N/^{31}P\}$ cryoprobe, a Bruker Avance III 500 MHz spectrometer equipped with a

$^{13}C/^{15}N\{^{1}H\}$ cryoprobe, and a Bruker Avance III HD 400 MHz spectrometer. HRMS data were acquired with a Bruker MaXis™ 4G ESI-QTOF mass spectrometer. RP HPLC was performed using a Shimadzu Prominence HPLC system and a Phenomenex Luna phenyl-hexyl column (250 × 4.6 mm). UHPLC-HRMS was acquired using a Bruker MaXis™ 4G ESI-QTOF mass spectrometer coupled with a Waters Acquity UPLC system operated by Bruker Hystar software and a C18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm).

Biological Material

Sponge specimens were collected on August 7, 2013 near Stan Blum State Park boat launch (27°49′45.7’’N, 80°18’42.8’’W) in Florida. A voucher specimen is housed at the University of Wisconsin-Madison. For cultivation, a sample of sponge (1 cm^3) was ground in 500 μL sterile seawater and dilutions were made using 500 μL sterile seawater. Subsequently, 400 μL of diluted sponge sample was added to 200 μL of sterile seawater and 100 μL was plated using a sterile L-shaped spreader. Diluted sample was plated on R2A[26] media supplemented with artificial seawater. Each medium was supplemented with 50 μg/mL cycloheximide, 25 μg/mL nystatin, and 25 μg/mL nalidixic acid. Plates were incubated at 28˚C and colonies were isolated over the course of two months.

Sequencing

16S rDNA sequencing was conducted as previously described[27]. WMMA-1423 was identified as an *Actinomadura* sp. The 16S sequence for WMMA-1423 was deposited in GenBank (accession number KY014999).

Fermentation, Extraction and Isolation

Two 10 mL seed cultures (25 × 150 mm tubes) in medium DSC (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract per liter of artificial seawater) were inoculated with strain WMMA-1423 and shaken (200 RPM, 28 °C) for seven days. Two liter flask (1 × 500 mL) containing ASW-A (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract, 5 g $CaCO_3$ per liter of artificial seawater) were inoculated with 20 mL seed culture and were incubated (200 RPM, 28 °C) for seven days. Four-liter flasks (6 × 1 L) containing medium RAM2 (4 g corn meal, 10 g glucose, 15 g maltose, 7.5 g pharmamedia, 5 g yeast extract per liter of 1:1 artificial seawater and distilled water) with Diaion HP20 (7% by weight) were inoculated with 50 mL from the 500 mL culture and shaken (200 RPM, 28 °C) for seven days. For making artificial sea water, solutions I (415.2 g NaCl, 69.54 g Na_2SO_4, 11.74 g KCl, 3.40 g $NaHCO_3$, 1.7 g KBr, 0.45 g H_3BO_3, 0.054 g NaF) and II (187.9 g $MgCl_2 \cdot 6H_2O$, 22.72 g $CaCl_2 \cdot 2H_2O$, 0.428 g $SrCl_2 \cdot 6H_2O$) were made up separately using distilled water and combined to give a total volume of 20 L.

Filtered HP20 and cells were washed with H_2O and extracted with acetone. The acetone extract was subjected to liquid-liquid partitioning using 30% aqueous MeOH and $CHCl_3$ (1:1). The $CHCl_3$-soluble partition (1.59 g) was fractionated by Sephadex LH20 column chromatography (column size 500 × 40 mm, $CHCl_3$:MeOH=1:1, 20 mL for each fraction). Fractions containing **2** and **3** (157.2 mg) were further fractionated using a benchtop HP20-ss column with a gradient from 10%/90% MeOH/H_2O to 100% MeOH (column size 500 × 40 mm, 100 mL eluent for each fraction). The 60%/40% MeOH/H_2O fraction (26.3 mg) was subjected to RP HPLC (10%/90% to 40%/60% MeCN/H_2O, 39.5 min, 4 mL/min) using a

Phenomenex Luna phenyl hexyl column (250 × 4.6 mm), yielding **1** (15.2 mg, t_R 22.5 min). The 100% MeOH fraction (46.2 mg) was subjected to RP HPLC (20%/80% to 50%/50% MeCN/H_2O, 39.5 min, 4 mL/min) using a Phenomenex Luna phenyl hexyl column (250 × 4.6 mm), yielding **2** (1.1 mg, t_R 16.2 min) and **3** (3.9 mg, t_R 21.0 min).

(−)-Madurastatin C1 (**18**): light yellow oil, $[\alpha]^{25}_{436}$ = −16.8 (*c* 2.5, MeOH); HRMS $[M+H]^+$ *m/z* = 592.2735 (calcd. for $C_{26}H_{38}N_7O_9^+$ 592.2726), $[M-2H+Fe]^+$ *m/z* = 645.1841 (calcd. for $C_{26}H_{35}N_7O_9Fe^+$ 645.1829). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.49 (t, *J* = 5.8 Hz, 1H), 8.17 (d, *J* = 8.2 Hz, 1H), 7.92 (t, *J* = 5.4 Hz, 1H), 7.64 (dd, *J* = 7.8, 1.6 Hz, 1H), 7.53 – 7.40 (m, 1H), 7.00 (d, *J* = 8.2 Hz, 1H), 6.94 (dd, *J* = 11.5, 4.2 Hz, 1H), 5.01 (dd, *J* = 10.4, 7.8 Hz, 1H), 4.64 (dd, *J* = 10.3, 8.5 Hz, 1H), 4.52 (t, *J* = 8.1 Hz, 1H), 4.37 – 4.27 (m, 1H), 3.81 – 3.62 (m, 4H), 3.26 (d, *J* = 6.1 Hz, 2H), 2.93 (t, *J* = 6.3 Hz, 1H), 2.56 – 2.51 (m, 2H), 2.23 (s, 3H), 1.95 – 1.82 (m, 4H), 1.72 – 1.38 (m, 5H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 173.2, 171.1, 170.3, 168.5, 166.0, 165.0, 159.1, 134.2, 128.2, 119.2, 116.7, 110.0, 69.5, 67.5, 63.5, 51.3, 49.6, 47.1, 42.3, 34.7, 34.0, 32.0, 30.0, 27.8, 22.8, 20.4.

Madurastatin D1 (**19**): light yellow oil, $[\alpha]^{25}_{436}$ = +16.6 (*c* 0.6, MeOH); UV-Vis (MeOH): λ_{max} (log ε) 224 nm (3.03), 243 nm (2.98), 306 nm (2.71); IR (ATR): υ_{max} 3329, 2944, 2832, 1639, 1547, 1530, 1449, 1413, 1330, 1258, 1233, 1200, 1114, 1023, 922, 826, 759, 667 cm^{-1}; HRMS $[M+H]^+$ *m/z* = 618.2876 (calcd. for $C_{28}H_{40}N_7O_9^+$ 618.2882), $[M-2H+Fe]^+$ *m/z* = 671.1988 (calcd. for $C_{28}H_{37}N_7O_9^+$ 671.1986). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.61 (t, *J* = 5.9 Hz, 1H), 8.01 (t, *J* = 5.4 Hz, 1H), 7.64 (dd, *J* = 7.6, 1.4 Hz, 1H), 7.47 (ddd, *J* = 8.4, 7.3, 1.6 Hz, 1H), 7.01 (d, *J* = 8.4 Hz, 1H), 6.95 (t, *J* = 7.5 Hz, 1H), 5.01 ((dd, *J* = 10.4, 7.7 Hz, 1H), 4.65 (dd, *J* = 10.3, 8.5 Hz, 1H), 4.51 (t, *J* = 8.0 Hz, 1H), 4.32-4.24 (m, 1H), 4.01 (q, *J* =

5.0 Hz, 1H), 3.73 (dd, *J* = 16.6, 6.0 Hz, 1H), 3.67 (dd, *J* = 16.6, 6.0 Hz, 1H), 3.57-3.41 (m, 2H), 3.24 (dd, *J* = 6.6, 6.0 Hz, 2H), 2.88 (m, 1H), 2.56-2.48 (m, 2H), 2.29 (s, 3H), 2.03 – 1.76 (m, 6H), 1.70 – 1.43 (m, 2H), 1.21 (d, *J* = 5.2 Hz, 3H). ^{13}C NMR (125 MHz, DMSO-d_6) δ 171.6, 170.8, 170.2, 168.4, 165.9, 162.4, 159.1, 134.1, 128.1, 119.1, 116.7, 110.0, 74.7, 69.5, 67.5, 64.8, 51.8, 51.1, 47.4, 42.2, 37.9, 34.7, 31.9, 26.6, 21.6, 21.6, 20.9, 19.7.

Madurastatin D2 (**20**): light yellow oil, $[\alpha]^{25}_{436}$ = −184.3 (*c* 1.4, MeOH); UV-Vis (MeOH): λ_{max} (log ε) 224 nm (3.79), 243 nm (3.74), 306 nm (3.38); IR (ATR): υ_{max} 3337, 2944, 2831, 1737, 1639, 1547, 1446, 1422, 1329, 1258, 1202, 1157, 1024, 923, 759 cm^{-1}; HRMS $[M+H]^+$ *m/z* = 632.3025 (calcd. for $C_{29}H_{42}N_7O_9^+$ 632.3039), $[M-2H+Fe]^+$ *m/z* = 685.2125 (calcd. for $C_{29}H_{39}N_7O_9^+$ 685.2142). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.56 (s, 1H), 7.97 (s, 1H), 7.63 (d, *J* = 7.4 Hz, 1H), 7.45 (t, *J* = 7.6 Hz, 1H), 6.98 (d, *J* = 7.6 Hz, 1H), 6.91 (t, *J* = 7.6 Hz, 1H), 5.00 ((dd, *J* = 10.2, 8.0 Hz 1H), 4.63 (dd, *J* = 10.2, 8.1 Hz, 1H), 4.51 (t, *J* = 8.3 Hz, 1H), 3.91 (dd, *J* = 6.0, 5.4 Hz, 1H), 3.75 (d, *J* = 16.0 Hz, 1H), 3.66 (d, *J* = 16.0 Hz, 1H), 3.53-3.41 (m, 2H), 3.25 (dd, *J* = 7.0, 5.6 Hz, 2H), 2.93 (t, *J* = 5.0 Hz, 1H), 2.55-2.50 (m, 2H), 2.24 (s, 3H), 2.03 – 1.79 (m, 6H), 1.46 – 1.36 (m, 2H), 1.29 (s, 3H), 1.11 (s, 3H). ^{13}C NMR (125 MHz, DMSO-d_6) δ 170.6, 170.2, 169.9, 168.4, 165.9, 163.3, 159.1, 134.1, 128.1, 119.0, 116.7, 110.0, 77.9, 69.4, 67.4, 62.0, 51.7, 51.1, 47.5, 42.2, 34.8, 32.5, 31.7, 25.7, 25.5, 25.0, 21.3, 20.9, 19.8.

Chrome Azurol S (CAS) assay[28]

Hexadecyltrimethylammonium bromide (CTAB) (21.9 mg) was dissolved in 25 mL of H_2O at 35 °C. To this solution were added 1.5 mL of 1 mM iron(III) chloride solution (prepared by dissolving anhydrous FeCl3 in a 10 mM aqueous HCl solution) and 7.5 mL of a

2 mM aqueous CAS solution at RT. In a separate Erlenmeyer flask, 9.76 g of 2-(N-morpholino)ethanesulfonic acid (MES) was diluted in 50 mL of water, and a 50% KOH solution was used to adjust the pH of this solution to 5.6. Then the premade CTAB−CAS−Fe(III) solution was poured into this MES buffer while stirring, and H_2O was added to make 100 mL to finish the preparation of the modified CAS assay solution. CTAB−CAS−Fe(III) with MES buffer solution (100 μL) was added to each well of the 96-well microplate. Each well was then treated with 100 μL of the dilute solution of madurastatins (**18**-**20**) or deferoxamine mesylate in H_2O to achieve final concentrations ranging from 1.28 mM to 1.25 μM. After incubation at 37 °C for 3 h, the resulting color changes were observed by visual inspection, or the corresponding absorption changes were recorded.

Marfey's Analysis

L-FDLA was synthesized following the experimental procedures in the original report[29]. D-Serine was purchased from TCI America Ltd. L-Serine was purchased from Sigma Aldrich Ltd. Separation of different diastereoisomers was carried out and detected on UHPLC-HRMS system using a linear gradient of 0.3 mL/mL of 10%-100% MeOH/H_2O (0.1 % formic acid) over 13 minutes and a hold at 100% MeOH for 3 minutes.

Five hundred μg of **18** was dissolved in 400 μL of concentrated HI and heated to 110 °C for 4 hours. The solution was evaporated to dryness and the sample dissolved in 200 μL of 1 M $NaHCO_3$. To this was added 170 μL of 1% L-FDLA in acetone and the mixture stirred for 1 hour at 40 °C. The reaction was quenched by the addition of 60 μL of 2 M aqueous HCl. The sample was diluted with LC-MS grade MeOH and analyzed on UHPLC-HRMS.

D-Serine and L-Serine (0.25 mg) were dissolved in 100 μL of distilled water and 20 μL of 1 M $NaHCO_3$ separately. 100 μL of 1% L-FDLA in acetone was added to both solutions, forming DL and LL pairs. The solutions were heated at 40 °C for 1 hour, diluted with LC-MS grade MeOH and analyzed on UHPLC-HRMS.

Antibacterial Testing

Madurastatin C1 (**18**), D1 (**19**) and D2 (**20**) were tested for antibacterial activity against *M. leuteus* (environmental isolate from sponge *Ircinia sp.*) and Methicillin-resistant *Staphylococcus aureus* (MRSA) (ATCC #33591), and MICs were determined using a dilution antimicrobial susceptibility test for aerobic bacteria[30]. Compounds **18**, **19** and **20** were dissolved in DMSO and serially diluted to 10 concentrations (0.25−128 μg/mL) in 96-well plates. Vancomycin was used as a positive control and exhibited an MIC of 0.25 μg/mL against *M. leuteus* and MRSA. Madurastatins and vancomycin were tested in triplicates. On each plate, there were six untreated media controls. The plates were incubated at 37 °C for 18 h. The MIC was determined as the lowest concentration that inhibited visible growth of bacteria.

Whole Genome Sequencing

DNA extraction of WMMA-1423 was conducted as previously described[31]. The complete genome was sequenced at the University of Wisconsin Biotechnology Center (UWBC) using PacBio RSII (Pacific Biosciences) technology. The assembly of the whole genome was conducted using Canu[32]. The biosynthetic gene clusters were characterized using antiSMASH v4.2[17] and the annotations of the gene clusters were made by blastp against NCBI's Non-redundant protein sequences (NR) database.

2.5. References

(1). Hider, R. C.; Kong, X. *Nat. Prod. Rep.* **2010**, *27*, 637−657.

(2). Miethke, M.; Marahiel, M. A. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413−451.

(3). Harada, K.; Tomita, K.; Fujii, K.; Masuda, K.; Mikami, Y.; Yazawa, K.; Komaki, H. *J. Antibiot*. **2004**, *57*, 125−135.

(4). Mazzei, E.; Iorio, M.; Maffioli, S. I.; Sosio, M.; Donadio, S. *J. Antibiot*. **2012**, *65*, 267−269.

(5). Kawahara, T.; Itoh, M.; Izumikawa, M.; Sakata, N.; Tsuchida, T.; Shin-ya, K. *J. Antibiot.* **2014**, *67*, 577−580.

(6). Keller-Schierlein, W.; Hagmann, L.; Zähner, H.; Huhn, W. *Helv. Chim. Acta* **1988**, *71*, 1528−1540.

(7). Shaaban, K. A.; Saunders, M. A.; Zhang, Y.; Tran, T.; Elshahawi, S. I.; Ponomareva, L. V.; Wang, X.; Zhang, J.; Copley, G. C.; Sunkara, M.; Kharel, M. K.; Morris, A. J.; Hower, J. C.; Tremblay, M. S.; Prendergast, M. A.; Thorson, J. S. *J. Nat. Prod*. **2017**, *80*, 2−11.

(8). Tyler, A. R.; Mosaei, H.; Morton, S.; Waddell, P. G.; Wills, C.; McFarlane, W.; Gray, J.; Goodfellow, M.; Errington, J.; Allenby, N.; Zenkin, N.; Hall, M. J. *Nat. Prod.* **2017**, *80*, 1558−1562.

(9). Zhang, F.; Barns, K.; Hoffmann, F. M.; Braun, D. R.; Andes, D. R.; Bugni, T. S. *J. Nat. Prod.* **2017**, *80*, 2551−2555.

(10). Wyche, T. P.; Standiford, M.; Hou, Y.; Braun, D.; Johnson, D. A.; Johnson, J. A.; Bugni, T. S. *Mar. Drugs* **2013**, *11*, 5089−5099.

(11). Wyche, T. P.; Piotrowski, J. S.; Hou, Y.; Braun, D.; Deshpande, R.; Mcllwain, S.; Ong, I. M.; Myers, C. L.; Guzei, I. A.; Westler, W. M.; Andes, D. R.; Bugni, T. S. *Angew. Chem. Int. Ed.* **2014**, *53*, 11583−11586.

(12). Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591−596.

(13). Finefield, J. M.; Sherman, D. H.; Kreitman, M.; Williams, R. M. *Angew. Chem. Int. Ed.* **2012**, *51*, 4802-4836.

(14). Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254−7260.

(15). Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946−12959.

(16). Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, *47*, W81−W87.

(17). Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Duran, H. G. S.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H. *Nucleic Acids Research*, **2017**, *45*, W36–W41.

(18). Park, S. R.; Tripathi, A.; Wu, J.; Schultz, P. J.; Yim, I.; McQuade, T. J.; Yu, F.; Arevang, C. –J.; Mensah, A. Y.; Tamayo-Castillo, G.; Xi, C.; Sherman, D. H. *Nat. Commun.* **2016**, *7:10710*.

(19). Kodani, S.; Komaki, H.; Suzuki, M.; Hemmi, H.; Ohnishi-Kameyama, M. *Biometals*, **2015**, *28*, 381−389.

(20). Seyedsayamdost, M. R.; Traxler, M. F.; Zheng, S. –L.; Kolter, R.; Clardy, J. *J. Am. Chem. Soc.* **2011**, *133*, 11434−11437.

(21). Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78.

(22). Grove, T. L.; Benner, J. S.; Radle, M. I.; Ahlum, J. H.; Landgraf, B. J.; Krebs, C.; Booker, S. J. *Science* **2011**, *332*, 604-607.

(23). Buckel, W.; Thauer, R. K. *Angew. Chem. Int. Ed.* **2011**, *50*, 10492–10494.

(24). Kim, H. J.; McCarty, R. M.; Ogasawara, Y.; Liu, Y. –N. ; Mansoorabadi, S. O.; LeVieux, J. Liu, H. –W. *J. Am. Chem. Soc.* **2013**, *135*, 8093–8096.

(25). Woodyer, R. D.; Li, G.; Zhao, H.; van der Donk, W. A. *Chem. Commun.* **2007**, *28*, 359–361.

(26). Reasoner, D. J.; Geldreich, E. E. *Appl. Environ. Microbiol.* **1985**, *49*, 1−7.

(27). Wyche, T. P.; Hou, Y.; Braun, D.; Cohen, H. C.; Xiong, M. P.; Bugni, T. S. *J. Org. Chem.* **2011**, *76*, 6542−6547.

(28). Alexander, D. B.; Zuberer, D. A. *Biol. Fertil. Soils* **1991**, *12*, 39−45.

(29). Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591−596.

(30). National Committee for Clinical Laboratory Standards. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 7th ed.; NCCLS: Villanova, PA, USA, 2006; Approved standard M7−A7.

(31). Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S. *ACS Chem. Biol.* **2017**, *12*, 3093−3102.

(32). Koren, S.; Walenz, B. P.; Berlin, K.; Miller, J. R.; Bergman, N. H.; Phillippy, A. M. *Genome Res.* **2017**, *27*, 722−736.

Chapter 3:

Discovery, structure elucidation, antibacterial activity studies, and biosynthetic mechanism proposal of bacillimidazoles A−F, imidazolium containing compounds isolated from *Bacillus sp.* WMMC-1349

Portions of this chapter will be submitted as:

Jia-Xuan Yan, Eric J. N. Helfrich, Marc G. Chevrette, Doug R. Braun Gene E. Ananiev, F. Michael Hoffman, Cameron R. Currie, Jon Clardy, and Tim S. Bugni Bacillimidazoles A−F, imidazolium containing antibacterial compounds isolated from *Bacillus sp.* WMMC-1349

3.1. Introduction

Metabolomics analysis guided bacteria strain prioritization coupled with investigations of relatively underexplored bacteria genera has proven to be effective in potentially discovering new chemical entities among the abundance of known compounds in Chapter 2. However, the biological activity information was not obtained after the PCA analysis of *Actinomadura* sp. WMMA-1423, which resulted in the discovery of new madurastatin analogs with moderate antibacterial activities. As an addition, a natural product library based discovery platform followed by high throughput screening was used to guide the discovery, purification, and biological activity studies of bacillimidazoles A-F isolated from *Bacillus* sp. WMMC-1349. Bacillimidazoles feature positively charged imidazolium ring structures that are relatively uncommon for bacteria natural products.

Heterocycles are widely distributed in natural products isolated from both terrestrial or marine organisms[1−4], among which nitrogen containing heterocyclic compounds, such as

pyrroles, imidazoles, thiazoles, oxazoles, pyridines, quinolones, are often found to exhibit a diverse array of biological activities including antibacterial[5], antifungal[6], anticancer[7], *etc*. The vast structural diversity and the biological properties of these heterocycles have motivated natural product chemists to the discovery of new heterocyclic compounds[4].

Historically sponges have been one of the most abundant sources of imidazole alkaloids with diverse structural scaffolds and important biological activities[8, 9]. Major classes of imidazoles produced by sponges include bromopyrrole-imidazoles[10], indole-containing imidazoles[11], and 2-aminoimidazoles[12]. In recent years, marine microorganisms have been proven to be sustainable and productive sources of new imidazole-containing bioactive natural products[9]. Despite the fact of the ubiquity of neutral imidazoles, imidazolium natural products remain relatively rare. A 1,3-dimethyl-5-methylthio structural moiety is featured in a majority of imidazolium natural products discovered to date[13−15]. Several reports also described the discovery of 2-aminoimidazolium containing natural products with structural similarity to guanidinium[10, 11]. Wide applications in ionic liquids[16], important biological activities[17], and the amenability to further structural modifications have made 1-3-difunctionalized imidazolium salts an attractive target of contemporary research[18]. Discovery of more 1,3-difunctionalized imidazoliums and gaining further insight into their biological activities therefore, remain interesting yet challenging tasks for natural product chemists.

As part of our ongoing effort towards discovering new natural products from marine invertebrate associated bacteria[19], we have developed a discovery platform that includes bacterial strain prioritization by principal component analysis (PCA) followed by an

automated, two step high throughput LC-MS fractionation to generate marine natural product screening libraries directly into 96-well plates for further antibacterial and antifungal screening[20, 21]. The strain WMMC-1349 corresponding to a *Bacillus* sp. isolated from the sponge *Cinachyrella apion* displaying activity against methicillin-resistant *Staphylococcus aureus* (MRSA) was identified as potential target for isolation of novel bioactive compounds based on PCA data. Bruker timsTOF™ (trapped ion mobility spectrometry) MS data for the active-well fraction resulting from this strain (WMMC-1349) indicated a series of new molecules with *m/z* ratios of 305.2, 319.2, 344.4, 358.2, 383.2 and 397.2 respectively. Subsequent LC/MS and bioassay guided fractionation led to the discovery of six new 1,3-difunctionalized imidazolium-containing compounds, which we named bacillimidazoles A−F (**21**−**26**, Figure 3.2). All six (**21**-**26**) compounds displayed *in vitro* activity against MRSA and *B. subtilis.* A biosynthetic pathway for these compounds was proposed based on whole genome analysis of the strain WMMC-1349 and isotopic labeling of bacillimidazoles by using isotopic enriched culture media.

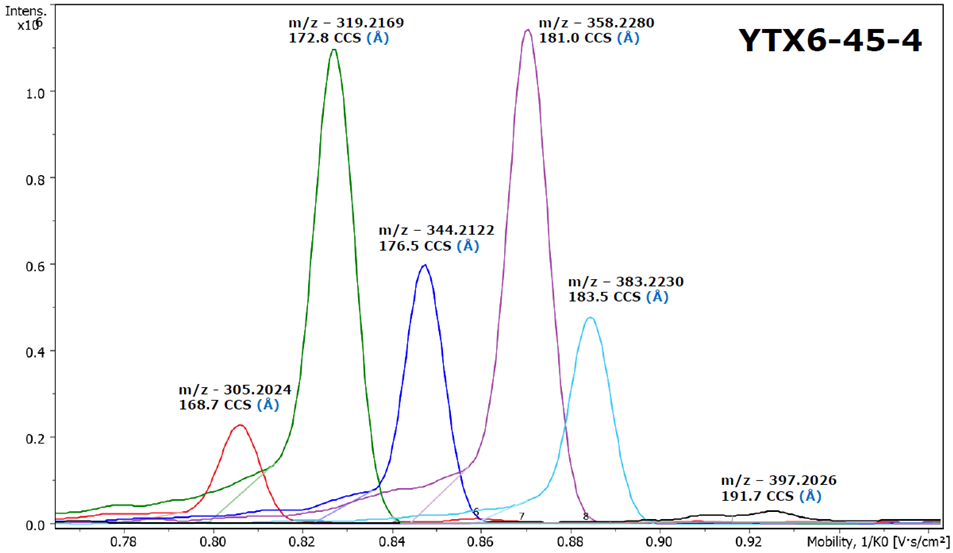
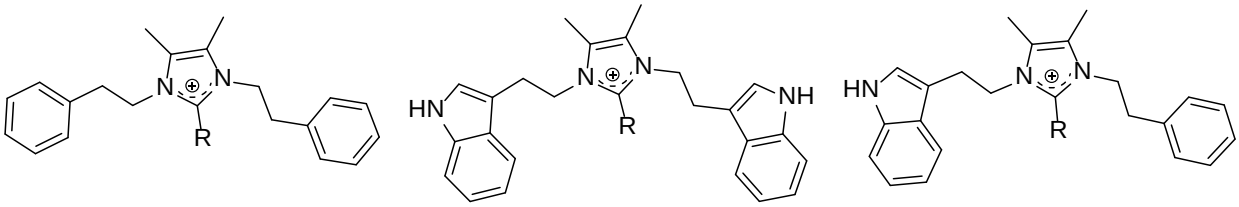


Figure 3.1. Bruker timsTOF MS spectrum of the active well produced by WMMC-1349. Peak assignments from left to right: **21**, **22**, **25**, **26**, **23**, **24**.



R = H, bacillimidazole A (**21**)
R = CH_3, bacillimidazole B (**22**)

R = H, bacillimidazole C (**23**)
R = CH_3, bacillimidazole D (**24**)

R = H, bacillimidazole E (**25**)
R = CH_3, bacillimidazole F (**26**)

Figure 3.2. Structures of bacillimidazoles A−F (**21**−**26**).

3.2. Results and Discussion

Molecular formulae of bacilimidazoles A (**21**) and B (**22**) were determined to be $C_{21}H_{25}N_2^+$ (*m/z* = 305.2029, M^+, Figure S47) and $C_{22}H_{27}N_2^+$ (*m/z* = 319.2171, M^+, Figure S48) respectively based on HRMS data. In the ^{13}C NMR spectra of **21** (Table 3.1, Figure S18) and **22** (Table 3.1, Figure S23), only 11 and 12 C-signals were observed respectively, suggesting that both **21** and **22** possess symmetrical structures. Furthermore, ^1H and ^{13}C NMR data (Table 3.1, Figure S17−S26) suggested high degree of structural similarities to lepidiline A and B, two imidazolium-containing alkaloids isolated from South American plants *Lepidium meyenii* Walp[22]. The presence of two phenyl rings and one 4,5-dimethyl imidazolium ring in **21** (2,4,5-trimethyl imidazolium ring for **22**) was established by comparison of NMR data for **21**, **22** with that reported for lepidilines A and B. ^1H NMR signals at 3.06 (H-7, 7', 4H, t, *J* = 6.6 Hz), 4.35 (H-8, 8', 4H, t, *J* = 6.6 Hz) showed HSQC correlations to C-7 (37.1 ppm) and C-8 (49.3 ppm) respectively. They also showed COSY correlations to each other, suggesting the presence of an *N*-2-phenylethyl moiety. Overall, the structure of **21** was established as 1,3-bis(2-phenylethyl)-4,5-dimethyl-1*H*-imidazolium and by analogy **22** was characterized as 1,3-bis(2-phenylethyl)-2,4,5-trimethyl-1*H*-imidazolium.

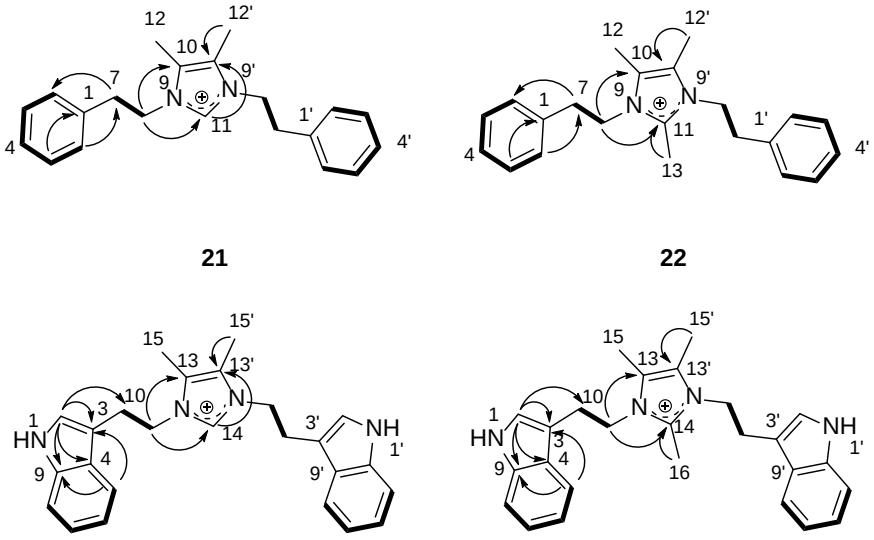
HRMS data indicated molecular formulae of bacilimidazole C (**23**) and bacilimidazole D (**24**) as $C_{25}H_{27}N_4^+$ (*m/z* = 383.2228, M^+, Figure S49) and $C_{26}H_{29}N_4^+$ (*m/z* = 397.2385, M^+,

Figure S50) respectively. Analysis of their ^{1}H and ^{13}C NMR data (Table 3.1, Figure S27−S36) also suggested symmetrical structures for **23** and **24**. The 4,5-dimethyl imidazolium and 2,4,5-trimethyl imidazolium core structures were also found to be present for **23** and **24** respectively. Five sets of unassigned aromatic protons (H-2, 2', 2H, s, 6.96 ppm; H-5, 5', 2H, d, *J* = 7.6 Hz, 7.32 ppm; H-6, 6', 2H, t, *J* = 7.6 Hz, 7.03 ppm; H-7, 7', 2H, t, *J* = 7.6 Hz, 7.15 ppm; H-8, 8', 2H, d, *J* = 7.6 Hz, 7.39 ppm) and 8 unassigned aromatic carbons (C-2, CH, 124.4 ppm; C-3, 110.5 ppm; C-4, 128.3 ppm; C-5, CH, 118.4 ppm; C-6, CH, 120.1 ppm; C-7, CH, 122.8 ppm; C-8, CH, 112.6 ppm; C-9, 138.1 ppm) were observed in **23**, which suggested the presence of two 3-indole rings. COSY correlation was observed between H-10 (4H, t, *J* = 6.8 Hz, 3.07 ppm) and H-11 (4H, t, *J* = 6.8 Hz, 4.25 ppm), which suggested a similar linkage unit between the imidazolium ring and the aromatic ring. Based on aforementioned observations, the structure of **23** was established as 1,3-bis[2-(1*H*-indol-3-yl)ethyl]-4,5-dimethyl-1*H*-imidazolium and the structure of **24** differed from **23** by the presence of a methyl group on C-14 and hence was characterized as 1,3-bis[2-(1*H*-indol-3-yl)ethyl]-2,4,5-trimethyl-1*H*-imidazolium.

The molecular formulae of bacilimidazole E (**25**) and bacilimidazole F (**26**) were determined to be $C_{23}H_{26}N_3^+$ (*m/z* = 344.2134, M^+, Figure S51) for and $C_{24}H_{28}N_3^+$ (*m/z* = 358.2283, M^+, Figure S52) respectively based on HRMS data. Analysis on ^{1}H and ^{13}C NMR data of **23** and **24** (Table 3.2, Figure S37−S46) suggested asymmetrical structures for both **25** and **26** since 23 ^{13}C NMR signals were observed in **25** (24 signals for **26**), which matched the HRMS suggested molecular formulas. Additionally, comparison of NMR data for **25** to that of **21** and **23**, established the presence of both 2-phenylethyl and 2-(1*H*-indol-3-yl)ethyl

moieties, in **25**. Furthermore, compound 5 also featured the same 4,5-dimethyl imidazolium cyclic core structure already established for compounds **21**−**24**. Thus, the structure of **25** was characterized as 1-[2-(1*H*-indol-3-yl)ethyl]-3-(2-phenylethyl)-4,5-dimethyl-1*H*-imidazolium. Similarly, and by analogy to **25**, the structure of **26** was determined as 1-[2-(1*H*-indol-3-yl)ethyl]-3-(2-phenylethyl)-2,4,5-trimethyl-1*H*-imidazolium.

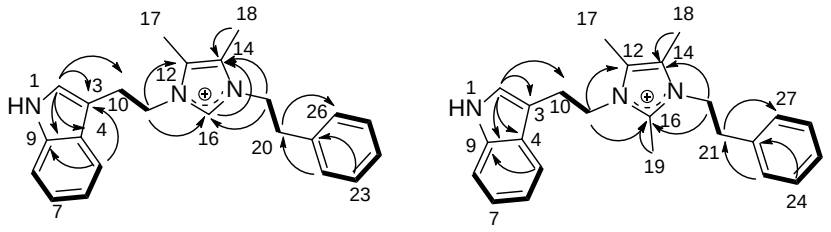
Table 3.1. ^{1}H and ^{13}C NMR data for **21**−**24** (600 MHz for ^{1}H (500 MHz for **24**), 125 MHz for ^{13}C, CD_3OD)



Position	**1**		**2**		**3**		**4**	
	δ_C, type	δ_H, (*J* in Hz)	δ_C, type	δ_H, (*J* in Hz)	δ_C, type	δ_H, (*J* in Hz)	δ_C, type	δ_H, (*J* in Hz)
1, 1′	137.8		138.2					
2, 2′	130.0, CH	7.12, dd (8.0, 1.8)	130.2, CH	7.11, dd (8.0, 1.8)	124.4, CH	6.96	124.5, CH	6.98
3, 3′	130.0, CH	7.34, t (7.4)	130.1, CH	7.34, t (7.8)	110.5		110.8	
4, 4′	128.4, CH	7.30, t (7.4)	128.6, CH	7.32, t (7.8)	128.3		128.4	
5, 5′	130.0, CH	7.34, t (7.4)	130.1, CH	7.34, t (7.8)	118.4, CH	7.32, d (7.6)	118.2, CH	7.24, d (8.0)
6, 6′	130.0, CH	7.12, dd (8.0, 1.8)	130.2, CH	7.11, dd (8.0, 1.8)	120.1, CH	7.03, t (7.6)	120.2, CH	7.03, t (8.0)
7, 7′	37.1, CH_2	3.06, t (6.6)	36.3, CH_2	3.02, t (6.8)	122.8, CH	7.15, t (7.6)	122.9, CH	7.14, t (8.0)
8, 8′	49.3, CH_2	4.35, t (6.6)	47.9, CH_2	4.31, t (6.8)	112.6, CH	7.39, d (7.8)	112.7, CH	7.39, d (8.0)

9, 9′					138.1		138.0	
10, 10′	128.5		127.1		26.8, CH_2	3.07, t (6.8)	26.0, CH_2	3.03, t (6.0)
11, 11′	135.5	8.51	144.0		48.7, CH_2	4.25, t (6.8)	47.5, CH_2	4.19, t (6.0)
12, 12′	7.9, CH_3	2.08	8.2, CH_3	2.08				
13, 13′			9.7, CH_3	2.01	128.2		127.0	
14					135.4, CH	8.12	144.0	
15, 15′					7.9, CH_3	2.09	8.3, CH_3	2.18
16							9.3, CH_3	1.64

Table 3.2. ^{1}H and ^{13}C NMR data for **25**, **26** (600 MHz for ^{1}H (500 MHz for **26**), 125 MHz for ^{13}C, CD_3OD)



25 **26**

COSY HMBC (key)

Position	**25**		**26**	
	δ_C, type	δ_H, (J in Hz)	δ_C, type	δ_H, (J in Hz)
1				
2	124.6, CH	7.03	124.6, CH	7.02
3	110.4		110.8	
4	128.4		128.5	
5	112.7, CH	7.40, dd (8.2, 1.0)	112.7, CH	7.37, d (7.8)
6	120.2, CH	7.04, dt (7.6, 1.2)	120.3, CH	7.00, t (7.8)
7	122.9, CH	7.15, dt (8.0, 1.0)	123.9, CH	7.12, t (7.8)
8	118.4, CH	7.34, dd (8.0, 1.0)	118.2, CH	7.23, d (7.8)
9	138.1		138.1	
10	26.5, CH_2	3.23, t (6.6)	26.0, CH_2	3.17, t (6.2)
11	48.9, CH_2	4.38, t (6.6)	47.7, CH_2	4.30, t (6.2)
12				
13	128.3		127.1	
14	128.3		127.0	
15				
16	135.4, CH	8.27	144.0	
17	8.0, CH_3	2.12	8.3, CH_3	2.15

18	7.9, CH_3	2.04	8.2, CH_3	2.07
19	48.7, CH_2	4.18, t (7.2)	9.4, CH_3	1.77
20	36.7, CH_2	2.84, t (7.2)	47.6, CH_2	4.11, t (6.8)
21	137.8		36.3, CH_2	2.77, t (6.8)
22	130.0, CH	7.04, d (7.2)	138.0	
23	129.9, CH	7.29, m	130.1, CH	7.03, dd (7.3, 1.6)
24	128.3, CH	7.28, m	130.0, CH	7.30, m
25	129.9, CH	7.29,m	128.5, CH	7.26, m
26	130.0, CH	7.04, d (7.2)	130.0, CH	7.30, m
27			130.1, CH	7.03, dd (7.3, 1.6)

Bacillimidazoles A−F were tested for antibacterial activity against *E. coli*, Methicillin-resistant *Staphylococcus aureus* (MRSA) (Table 3.3) and *B. subtilis*. Compounds **21**−**26** displayed better activity against MRSA than *B. subtilis* and no significant activity against *E.coli* (MIC>200 μM). All compounds (**21**-**26**) showed better inhibition activities against MRSA than *B. subtillis*. Bacillimidazole F (**26**) showed the most potent antibacterial activity among all 6 compounds with MIC values of 38.3 μM and 76.6 μM against MRSA and *B. subtilis* respectively. The asymmetric structure and the 2,4,5-trimethylimidazolium ring system might lead to the increased antibacterial activity and this brief structure activity relationship (SAR) study may provide guidance for future drug designs.

Table 3.3. Antibacterial MIC values of **21**−**26**

	21	**22**	**23**	**24**	**25**	**26**
E. coli	>200 μM	>200 μM	>200 μM	>200 μM	>200 μM	>200 μM
MRSA	76.5 μM	74.0 μM	64.4 μM	70.1 μM	70.0 μM	38.3 μM
B. subtilis	152.9 μM	148.0 μM	128.9 μM	140.2 μM	140.0 μM	76.6 μM

To identify the biosynthetic cluster for bacillus biosynthesis in order to understand the biosynthetic mechanism for the uncommon imidazolium cyclic structures in bacillimidazoles,

the whole genome of *Bacillus sp.* WMMC-1349 was sequenced (GenBank accession number CP041244) and analyzed using antiSMASH (version 5.0.0 rc1, results shown in Figure 3.3).[23-24] In total nine biosynthetic gene clusters were identified in the whole genome sequence of WMMC-1349, however, the gene cluster(s) responsible for the bacillimidazole biosynthesis was not identified. Two possible reasons were proposed for this observation: similar imidazolium containing structures have limited reports and studies in bacteria natural products, led to no reliable results from homology search; or bacillimidazoles were not produced by the producing organism *Bacillus* sp. WMMC-1349. Isotopic labeling of the bacillimidazoles using ^{13}C enriched culture media was performed to exclude the second possibility that bacillimidazoles were not produced by the producing organism. By substituting two carbon sources, soluble starch and D-glucose, with $^{13}C_6$-D-glucose, only four carbons on the imidazolium ring were labeled with high level of ^{13}C incorporation for bacillimidazole C (**23**) and E (**25**) (Figure 3.4a, Figure S53-S58), which substantiated that the bacillimidazoles are biologically synthetized by *Bacillus* sp. Only low level of ^{13}C incorporation was observed for the rest of the molecule, suggesting the biological source of those carbons are different from glucose. We propose amino acids are the biological source of those carbons with low level of ^{13}C labeling. Inspired by recent advances in a traditional organic reaction pinacol coupling[25], which happens between two carbonyl groups to form 1,2-diol structures via a single-electron pathway, a similar biosynthetic mechanism for the biosynthesis of bacillimidazoles was proposed (Figure 3.4b).

Overview 9.1 9.2 9.3 9.4 9.5 9.6 9.7 9.8 9.9

Identified secondary metabolite regions

tig00000493

Region	Type	From	To	Most similar known cluster		Similarity
Region 9.1	sactipeptide , head_to_tail	232,876	255,818	Sporulation killing factor skfA	head-to-tail cyclized peptide	71%
Region 9.2	NRPS , arylpolyene	498,279	560,819			
Region 9.3	bacteriocin	979,160	988,900			
Region 9.4	betalactone	1,921,230	1,949,761	Fengycin	other-nrps	53%
Region 9.5	terpene	2,030,262	2,052,136			
Region 9.6	lanthipeptide	2,109,391	2,132,435	Bicereucin	other	18%
Region 9.7	bacteriocin	2,820,535	2,830,912			
Region 9.8	NRPS , CDPS	3,088,212	3,147,023	Zwittermycin A	nrps-t1pks	14%
Region 9.9	NRPS	3,396,219	3,445,944	Bacillibactin	NRPS	53%

Figure 3.3. antiSMASH result of *Bacillus* sp. WMMC-1349. Nine biosynthetic gene clusters were identified based on homology search of the uploaded whole genome and none of them were responsible for bacillimidazole biosynthesis.

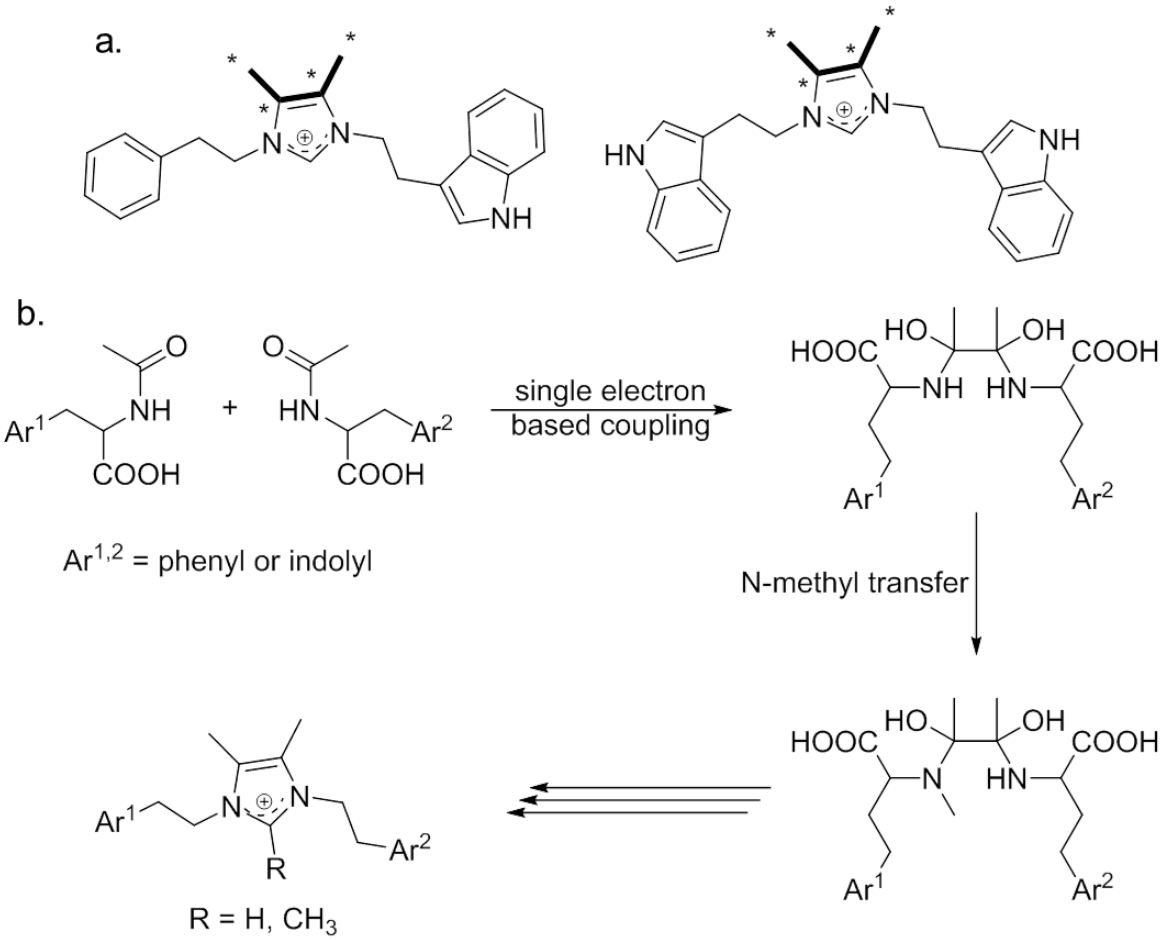


Figure 3.4. (a). Isotopic labeling of bacillimidazoles. Carbons highlighted with bold bonds and asterisks had high level of ^{13}C incorporation. (b). Proposed biosynthetic mechanism of bacillimidazoles.

Acetyl groups were transferred to the amine functional groups of phenylalanines and tryptophans catalyzed by acyltransferases. The biological source of the acetyl groups were proposed to be acetyl-CoA, which is one of the products of glycolysis[26], which supported the high levels of ^{13}C incorporation for those carbons. Single electron based coupling was proposed between two acylated amino acids to form a coupled intermediate. This intermediate would undergo another N-methylation reaction catalyzed by a SAM based N-methyltransferase and the biological source of the methyl group was methionine[27]. Further decarboxylation, dehydration, and aromatization reactions would generate the bacillimidazole compounds.

3.3. Conclusion

In summary, we report the isolation and structural elucidation of bacillimidazoles A−F (**21**-**26**), six imidazolium containing heterocycles displaying antibacterial activity against MRSA and *B. subtilis*. A possible biosynthetic pathway was proposed for all six compounds based on ^{13}C labeling of the bacillimidazoles using ^{13}C enriched culture media. To the best of our knowledge, there are only a very few reports on natural products containing 1,3-difunctionalized imidazolium alkaloids. Therefore, the discovery of these six bacillimidazoles provides useful information on understanding the relatively less studied biosynthetic pathways, which will be conducive towards exploring the full biosynthetic potential of *Bacillus* sp. Additionally, the brief SAR study of these six compounds may lead to more efficient drug designs in future.

3.4. Material and Methods

General Experimental Procedures

UV spectra were recorded on an Aminco/OLIS UV-Vis spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR spectrophotometer. NMR spectra were obtained in CD_3OD (δ_H 3.34 ppm, δ_C 49.0 ppm) with a Bruker Avance 600 III MHz spectrometer equipped with a $^{1}H\{^{13}C/^{15}N/^{31}P\}$ cryoprobe, a Bruker Avance III 500 MHz spectrometer equipped with a $^{13}C/^{15}N\{^{1}H\}$ cryoprobe, and a Bruker Avance III HD 400 MHz spectrometer. HRMS data were acquired with a Bruker MaXis™ 4G ESI-QTOF mass spectrometer. RP HPLC was performed using a Shimadzu Prominence HPLC system and a Phenomenex Gemini C18 column (250 × 30 mm). UHPLC-HRMS was acquired using a Bruker MaXis™ 4G ESI-QTOF mass spectrometer coupled with a Waters Acquity UPLC system operated by Bruker Hystar software and a C18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm). Bruker timsTOF Pro instrument was used for the trapped ion mobility MS analysis using direct infusion with 0.003 mL/min of flow rate and ESI+ ionization source. Nebulizer gas 0.4 bar, dry gas 3.5 L/min, source temperature 220 ^{o}C, ESI voltage 4200V (+). MS scan range : 20-1000 *m/z*.

Biological Material

Sponge specimens were collected on May 27, 2015 near the west shore of Ramrod Key (24°39′38.1’’N, 81°25’25.0’’W) in Florida. A voucher specimen is housed at the University of Wisconsin-Madison. For cultivation, a sample of sponge (1 cm^3) was ground in 500 μL sterile seawater and dilutions were made using 500 μL sterile seawater. Subsequently, 400 μL of diluted sponge sample was added to 200 μL of sterile seawater and 100 μL was plated using a

sterile L-shaped spreader. Diluted sample was plated on Gauze 1[28] media supplemented with artificial seawater. Each medium was supplemented with 50 μg/mL cycloheximide, 25 μg/mL nystatin, and 25 μg/mL nalidixic acid. Plates were incubated at 28˚C and colonies were isolated over the course of two months.

Sequencing

16S rDNA sequencing was conducted as previously described[29]. WMMC-1349 was identified as a *Bacillus* sp. The 16S sequence for WMMC-1349 was deposited in GenBank (accession number MK892477).

Fermentation, Extraction and Isolation

Two 10 mL seed cultures (25 × 150 mm tubes) in medium DSC (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract per liter of artificial seawater) were inoculated with strain WMMC-1349 and shaken (200 RPM, 28 °C) for seven days. Two liter flask (1 × 500 mL) containing ASW-A (20 g soluble starch, 10 g D -glucose, 5 g peptone, 5 g yeast extract, 5 g $CaCO_3$ per liter of artificial seawater. For ^{13}C enriched ASW-A media, 10 g D-glucose (U-$^{13}C_6$, 99%) was used instead of soluble starch and D-glucose) were inoculated with 20 mL seed culture and were incubated (200 RPM, 28 °C) for seven days. Four-liter flasks (10 × 1 L) containing medium ASW-A with Diaion HP20 (7% by weight) were inoculated with 50 mL from the 500 mL culture and shaken (200 RPM, 28 °C) for seven days. For making artificial sea water, solutions I (415.2 g NaCl, 69.54 g Na_2SO_4, 11.74 g KCl, 3.40 g $NaHCO_3$, 1.7 g KBr, 0.45 g H_3BO_3, 0.054 g NaF) and II (187.9 g $MgCl_2 \cdot 6H_2O$, 22.72 g $CaCl_2 \cdot 2H_2O$, 0.428 g $SrCl_2 \cdot 6H_2O$) were made up separately using distilled water and combined to give a total

volume of 20 L.

Filtered HP20 and cells were washed with H_2O and extracted with acetone. The acetone extract was subjected to liquid-liquid partitioning using 30% aqueous MeOH and $CHCl_3$ (1:1). The $CHCl_3$-soluble partition (3.12 g) was fractionated by Sephadex LH20 column chromatography (column size 500 × 40 mm, $CHCl_3$:MeOH=1:1, 20 mL for each fraction). Fractions containing **21**−**26** (1.1 g) were subjected to RP HPLC (20%/80% to 100%/0% MeOH/H_2O (with 0.1% acetic acid), 23.5 min, 20 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm). The fraction collected between 16−18 minutes was further fractionated by RP HPLC (22%/78% to 51%/49% MeCN/H_2O (with 0.05% trifluoroacetic acid), 29.5 min, 20 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm), yielding **21** (50.2 mg, t_R 28.2 min), **22** (35.8 mg, t_R 27.6 min), **23** (2.2 mg, t_R 27.3 min), 2**4** (3.1 mg, t_R 28.9 min), 2**5** (40.5 mg, t_R 28.6 min), 2**6** (32.3 mg, t_R 29.3 min).

Bacillimidazole A (**21**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 211 nm (3.83), 259 nm (2.65), 279 nm (2.55), 291 nm (2.55); IR (ATR): υ_{max} 3385, 3144, 3033, 2935, 2873, 2834, 1781, 1679, 1563, 1498, 1456, 1399, 1357, 1199, 1129, 1083, 1029, 831, 801, 751, 720, 702 cm^{-1}; HRMS M^+ *m/z* = 305.2029 (calcd. for $C_{21}H_{25}N_2^+$ 305.2012). ^1H NMR (600 MHz, MeOD) δ 8.51 (s, 1H), 7.34 (t, *J* = 7.2 Hz, 4H), 7.31 (d, *J* = 7.2 Hz, 2H), 7.12 (d, *J* = 7.0 Hz, 4H), 4.35 (t, *J* = 7.0 Hz, 4H), 3.06 (t, *J* = 6.9 Hz, 4H), 2.08 (s, 6H). ^{13}C NMR (126 MHz, MeOD) δ 137.83, 135.53, 130.02, 129.97, 128.45, 128.44, 49.28, 37.07, 7.94.

Bacillimidazole B (**22**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 217 nm (3.85), 274 nm (3.02), 282 nm (3.04), 290 (2.97); IR (ATR): υ_{max} 3383, 3065, 3034, 2934, 2870, 1681, 1564, 1525, 1497, 1440, 1400, 1355, 1200, 1128, 1030, 934, 840, 801, 749, 722, 703

cm^{-1}; HRMS M^+ *m/z* = 319.2171 (calcd. for $C_{22}H_{27}N_2^+$ 319.2169). ^{1}H NMR (600 MHz, MeOD) δ 7.36 – 7.33 (m, 6H), 7.11 (d, *J* = 6.4 Hz, 4H), 4.31 (t, *J* = 6.7 Hz, 4H), 3.02 (t, *J* = 6.7 Hz, 4H), 2.08 (s, 6H), 2.01 (s, 3H). ^{13}C NMR (126 MHz, MeOD) δ 144.04, 138.15, 130.22, 130.08, 128.56, 127.14, 47.87, 36.33, 9.67, 8.21.

Bacillimidazole C (**23**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 223 nm (4.25), 274 nm (3.67), 282 nm (3.69), 290 nm (3.61); IR (ATR): υ_{max} 3357, 2946, 2835, 1679, 1564, 1449, 1432, 1341, 1203, 1185, 1137, 1024, 838, 802, 746, 722 cm^{-1}; HRMS M^+ *m/z* = 383.2228 (calcd. for $C_{25}H_{27}N_4^+$ 383.2230). ^{1}H NMR (600 MHz, MeOD) δ 8.12 (s, 1H), 7.39 (d, *J* = 8.1 Hz, 2H), 7.32 (d, *J* = 7.9 Hz, 2H), 7.15 (t, *J* = 7.4 Hz, 2H), 7.03 (t, *J* = 7.4 Hz, 2H), 6.96 (s, 2H), 4.25 (t, *J* = 6.7 Hz, 4H), 3.07 (t, *J* = 6.7 Hz, 4H), 2.09 (s, 9H). ^{13}C NMR (126 MHz, MeOD) δ 138.05, 135.37, 128.30, 128.19, 124.42, 122.83, 120.13, 118.43, 112.64, 110.49, 26.81, 7.94.

Bacillimidazole D (**24**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 224 nm (3.98), 281 nm (3.44), 290 nm (3.38); IR (ATR): υ_{max} 3360, 3292, 2925, 2854, 1729, 1648, 1561, 1456, 1411, 1342, 1256, 1235, 1181, 1105, 1073, 1025, 926, 744 cm^{-1}; HRMS M^+ *m/z* = 397.2385 (calcd. for $C_{26}H_{29}N_4^+$ 397.2387). ^{1}H NMR (500 MHz, MeOD) δ 7.39 (d, *J* = 8.0 Hz, 1H), 7.24 (d, *J* = 7.9 Hz, 1H), 7.14 (t, *J* = 7.5 Hz, 1H), 7.03 (t, *J* = 7.5 Hz, 1H), 6.98 (s, 1H), 4.19 (t, *J* = 5.9 Hz, 2H), 3.03 (t, *J* = 5.7 Hz, 2H), 2.18 (s, 3H), 1.64 (s, 1H). ^{13}C NMR (126 MHz, MeOD) δ 143.95, 138.00, 128.40, 126.96, 124.50, 122.88, 120.21, 118.23, 112.67, 110.76, 47.50, 25.97, 9.30, 8.29.

Bacillimidazole E (**25**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 221 nm (3.87), 274 nm (3.17), 282 nm (3.18), 290 (3.11); IR (ATR): υ_{max} 3376, 2991, 2950, 2836, 1677,

1564, 1497, 1456, 1398, 1356, 1341, 1201, 1134, 1078, 1025, 934, 834, 801, 747, 721, 703 cm^{-1}; HRMS M^+ *m/z* = 344.2134 (calcd. for $C_{23}H_{26}N_3^+$ 344.2121). ^{1}H NMR (600 MHz, MeOD) δ 8.27 (s, 1H), 7.40 (d, *J* = 8.2 Hz, 1H), 7.34 (d, *J* = 7.9 Hz, 1H), 7.32 – 7.26 (m, 3H), 7.15 (t, *J* = 7.4 Hz, 1H), 7.07 – 7.02 (m, 4H), 4.38 (t, *J* = 6.5 Hz, 2H), 4.18 (t, *J* = 7.2 Hz, 2H), 3.23 (t, *J* = 6.5 Hz, 2H), 2.84 (t, *J* = 7.2 Hz, 2H), 2.12 (s, 3H), 2.04 (s, 3H). ^{13}C NMR (126 MHz, MeOD) δ 138.08, 137.82, 135.44, 129.95, 129.87, 128.38, 128.34, 128.29, 128.27, 124.55, 122.85, 120.17, 118.37, 112.68, 110.42, 37.03, 26.76, 7.97, 7.89.

Bacillimidazole F (**26**): light yellow solid, UV-Vis (MeOH): λ_{max} (log ε) 225 nm (3.96), 274 nm (3.55), 282 nm (3.56), 290 nm (3.51); IR (ATR): υ_{max} 3356, 3275, 3062, 3001, 2971, 2928, 2830, 1647, 1562, 1497, 1454, 1401, 1353, 1234, 1203, 1179, 1108, 1077, 1028, 924, 745, 703 cm^{-1}; HRMS M^+ *m/z* = 358.2283 (calcd. for $C_{24}H_{28}N_3^+$ 358.2278). ^{1}H NMR (500 MHz, MeOD) δ 7.37 (d, *J* = 8.0 Hz, 1H), 7.28 (m, 3H), 7.23 (d, *J* = 7.9 Hz, 1H), 7.11 (t, *J* = 7.5 Hz, 1H), 7.06 – 6.97 (m, 3H), 4.31 (t, *J* = 5.5 Hz, 2H), 4.11 (t, *J* = 6.6 Hz, 2H), 3.17 (t, *J* = 5.5 Hz, 2H), 2.76 (t, *J* = 6.6 Hz, 2H), 2.15 (s, 2H), 2.07 (s, 3H), 1.77 (s, 3H). ^{13}C NMR (126 MHz, MeOD) δ 143.99, 138.09, 138.03, 130.07, 130.02, 128.47, 128.45, 127.06, 127.00, 124.64, 122.91, 120.28, 118.18, 112.72, 110.76, 47.74, 47.59, 36.27, 26.00, 9.43, 8.32, 8.18.

Antibacterial Testing

Bacillimidazoles A−F (**21**−**26**) were tested for antibacterial activity against *E. coli* (ATCC #25922), *B. subtilis* strain NRS-231 and Methicillin-resistant *Staphylococcus aureus* (MRSA) (ATCC #33591), and MICs were determined using a dilution antimicrobial susceptibility test for aerobic bacteria[30]. Compounds **21**−**26** were dissolved in DMSO and serially diluted to 10 concentrations (0.25−128 μg/mL) in 96-well plates. Vancomycin was

used as a positive control against *M. leuteus* and MRSA and exhibited an MIC of 0.25 μg/mL. Gentamicin was used as a positive control against *E. coli* and exhibited an MIC of 4 μg/mL. Bacillimidazoles, vancomycin and gentamicin were tested in triplicates. On each plate, there were six untreated media controls. The plates were incubated at 37 °C for 18 h. The MIC was determined as the lowest concentration that inhibited visible growth of bacteria.

Whole Genome Sequencing

DNA extraction of WMMC-1349 was conducted as previously described[31]. The complete genome was sequenced at the University of Wisconsin Biotechnology Center (UWBC) using PacBio RSII (Pacific Biosciences) technology. The assembly of the whole genome was conducted using Canu[32]. The biosynthetic gene clusters were characterized using antiSMASH v5.0.0[23] and the annotations of the gene clusters were made by blastp against NCBI's Non-redundant protein sequences (NR) database.

3.5. References

(1). Lewis, J. R. *Nat. Prod. Rep.*, **2001**, *18*, 95−128.

(2). Jin, Z.; Li, Z.; Huang, R.; *Nat. Prod. Rep.*, **2002**, *19*, 454−476.

(3). Blunt, J. W.; Copp, B. R.; Keyzers, R. A.; Munro, M. H. G.; Prinsep, M. R. *Nat. Prod. Rep.*, **2017**, *34*, 235−294.

(4). Jin, Z. *Nat. Prod. Rep.*, **2016**, *33*, 1268−1317.

(5). Ziar, N.; Montalvão, S.; Hodnik, Z.; Nawrot, D. A.; Žula, A.; Ilaš, J.; Kikelj, D.; Tammela, P.; Mašič, L. P. *Mar. Drugs*, **2014**, *12*, 940–963.

(6). Hassan, W.; Edrada, R.; Ebel, R.; Wray, V.; Berg, A.; van Soest, R.; Wiryowidagdo, S.; Proksch, P. *J. Nat. Prod.*, **2004**, *67*, 817–822.

(7). Dyson, L.; Wright, A. D.; Young, K. A.; Sakoff, J. A.; McCluskey, A. *Bioorg. Med. Chem.*, **2014**, *22*, 1690−1699.

(8). Jin, Z. *Nat. Prod. Rep.*, **2013**, *30*, 869−915.

(9). Jin, Z. *Nat. Prod. Rep.*, **2011**, *28*, 1143−1191.

(10). Tanaka, N.; Kusama, T.; Takahashi-Nakaguchi, A.; Gonoi, T.; Fromont, J.; Kobayashi, J. *Org. Lett.*, **2013**, *15*, 3262−3265.

(11). Zhang, F.; Wang, B.; Prasad, P.; Capon, R. J.; Jia, Y. *Org. Lett.*, **2015**, *17*, 1529−1532.

(12). Gong, K. –K.; Tang, X. –L.; Liu, Y. –S.; Li, P. −L.; Li, G. –Q. *Molecules*, **2016**, *21*, 150.

(13). Bourguet-Kondracki, M. L.; Martin, M. T.; Guyot, M. *Tetrahedron Lett.*, **1996**, *37*, 3457−3460.

(14). Pedpradab, S.; Edrada, R.; Ebel, R.; Wray, V.; Proksch, P. *J. Nat. Prod.*, **2004**, *67*, 2113−2116.

(15). Carroll, A. R.; Avery, V. M. *J. Nat. Prod.*, **2009**, *72*, 696−699.

(16). Malhotra, S. V.; Kumar, V.A *Bioorg. Med. Chem. Lett.*, **2010**, *20*, 581−585.

(17). Liu, L. –P.; Zong, M. –H.; Linhardt, R. J.; Lou, W. –Y.; Li, N.; Huang, C.; Wu, H. *Biotechnol. Biofuels*, **2016**, *9*:266.

(18). Wright, B. D.; Deblock, M. C.; Wagers, P. O.; Duah, E.; Robishaw, N. K.; Shelton K. L.; Southerland, M. R.; DeBord, M. A.; Kersten, K. M.; McDonald, L. J.; Stiel, J. A.; Panzner, M. J.; Tessier, C. A.; Paruchuri, S.; Youngs, W. J. *Med. Chem. Res.*, **2015**, *24*, 2838−2861.

(19). Wyche, T. P.; Piotrowski, J. S.; Hou, Y.; Braun, D.; Deshpande, R.; Mcllwain, S.; Ong, I. M.; Myers, C. L.; Guzei, I. A.; Westler, W. M.; Andes, D. R.; Bugni, T. S. *Angew. Chem. Int. Ed.*, **2014**, *53*, 11583−11586.

(20). Zhang, F.; Barns, K.; Hoffmann, F. M.; Braun, D. R.; Andes, D. R.; Bugni, T. S. *J. Nat. Prod.*, **2017**, *80*, 2551−2555.

(21). Zhang, F.; Braun, D. R.; Ananiev, G. E., Hoffmann, F. M.; Tsai, I−W.; Rajski, S. R., Bugni, T. S. *Org. Lett.*, **2018**, *20*, 5529−5532.

(22). Cui, B.; Zheng, B. L.; He, K.; Zheng, Q. Y. *J. Nat. Prod.*, **2003**, *66*, 1101−1103.

(23). Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, *47*, W81–W87.

(24). Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Duran, H. G. S.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H. *Nucleic Acids Research*, **2017**, *45*, W36–W41.

(25). Vacas, T.; Alvarez, E.; Chiara, J. L. *Org. Lett.*, **2007**, *9*, 5445-5448.

(26). Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 16.1, Glycolysis Is an Energy-Conversion Pathway in Many Organisms. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22593/

(27). Struck, A. –W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *Chembiochem*, **2012**, *13*, 2642-2655.

(28). Gauze, G. F.; Preobrazhenskaya, J. P.; Kudrina, E. E.; Blinov, N. O.; Ryabova, I. D.; Sveshnikova, M. A. **1957**. Problems in the classification of antagonistic actinomycetes. State Publishing House for Medical Literature. Medgiz, Moscow.

(29). Wyche, T. P.; Hou, Y.; Braun, D.; Cohen, H. C.; Xiong, M. P.; Bugni, T. S. *J. Org. Chem.* **2011**, *76*, 6542−6547.

(30). National Committee for Clinical Laboratory Standards. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 7th ed.; NCCLS: Villanova, PA, USA, 2006; Approved standard M7-A7.

(31). Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S. *ACS Chem. Biol.* **2017**, *12*, 3093−3102.

(32). Koren, S.; Walenz, B. P.; Berlin, K.; Miller, J. R.; Bergman, N. H.; Phillippy, A. M. *Genome Res.* **2017**, *27*, 722−736.

Chapter 4:

Efforts towards the structure elucidation of antibacterial ribosomally synthetized post-translationally modified peptide (RiPP) using analytical techniques

Portions of this chapter have been submitted as:

Zhang, F.; Wyche, T. P.; Zhu, Y.; Braun, D. R.; Yan, J. –X.; Ge, Y.; Guzei, I. A.; Chevrette, M. G.; Currie, C. R.; Thomas, M. G.; Rajski, S. R.; Bugni, T. S. MS-derived isotopic fine structure reveals forazoline A as a thioketone-containing marine-derived natural product.

Portions of this chapter have been published as:

Yan, J. –X.; Zhang, F.; Adnani, N.; Bugni, T. S. Unambiguous molecular formula identification of natural products by analyzing isotopic fine structures acquired from high resolution mass spectrometer. Bruker application note, http://www.alanabaranov.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/ApplicationNotes/1860647_MRMS_60_Unambiguous_Natural_Product_ID_05_2018.pdf

4.1. Introduction

In this chapter, the discovery, antibacterial activity study, and efforts towards the structure elucidation of a ribosomally synthesized post-translationally modified peptide (RiPP) with more than 2,000 M.W. will be described. Isotopic fine structure (IFS) analysis unambiguous molecular formula determination method was developed and validated as a powerful tool that can assist the structure elucidation of complicated natural products.

RiPPs represent a major class of natural products with diverse chemical structures as well as biological activities[1]. The systematic classification of this class of compound was made possible by recent advances in genome sequencing efforts in the 21st century for the ubiquitous presence of their biosynthetic gene clusters among all three domains of life[2-4]:

bacteria, archaea, and eukaryotes. The extensive post-translationally modifications allow different chemical structures compared to other ribosomal peptides, generally leading to increased metabolic and chemical stabilities, different biological activities, and more specific target identifications and recognitions[5]. The biosynthetic pathway of RiPPs usually starts with the biosynthesis of a longer peptide sequence called precursor peptide, typically 20-110 amino acid residues in length. The precursor peptide could be divided into four parts: signal sequence, usually found near the N-terminus that directs the precursor peptide to the specific cell compartment for post-translationally modifications[6]; leader peptide, also found near the N-terminus that is crucial for the recognition of precursor peptide by post-translationally modifications related enzymes[7]; core peptide that undergoes post-translationally modifications and eventually becomes the mature RiPPs[8, 9]; and recognition sequence, usually near the C-terminus, that is important for some cyclization and excision modifications[10]. Compared to the core peptides with high mutations and variabilities, variances in leader peptides are relatively limited, which makes the leader peptides great targets for bioinformatics tools, such as RiPPMiner to recognize[11]. Once the leader peptides are identified, analysis of the core peptides, post-translationally modification related enzymes, and spectroscopic data can divide the RiPPs into different classes including lanthipeptides[12], linaridins[13], proteusins[14], linear azole containing peptides (LAPs)[15], cyanobactins[16], thiopeptides[17], bottromycins[18], microcins[19], lasso peptides[20] and etc. Among these different classes of RiPPs, sulfur (S) containing functional groups are crucial in some major classes, such as lanthipeptides, linaridins, and thiopeptides. Therefore, having evidence of the presence or absence of sulfur atoms will result in better classifications of the RiPPs, which

will eventually lead to their structure elucidations.

One of the most straightforward methods to examine the presence of sulfur atoms in the molecule is to determine the molecular formula. In fact, identifying the exact molecular formula of natural product is also a challenging problem in natural product research[21]. Knowing the exact formula can assist the classification of RiPP peptide for further amino acid sequence prediction by genomic analysis and the structure elucidation of the post-translationally modified core peptide.

Mass spectrometry is a well-established technique for measuring the accurate mass to charge (*m/z*) ratios of ions. This technique has been widely used in determining the molecular formulas of natural products as well as synthetic compounds. In mass spectrometry, resolution[22] could be defined as:

$$\text{Resolution} = R = \frac{M}{\Delta M}$$

M is the *m/z* ratio of the selected peak and ΔM is usually defined as the peak width at its half-maximum peak height. The resolution of the measurement could be improved by increasing the resolving power of the mass spectrometers. Thus the monoisotopic mass peak could be identified correctly and lead to an accurate molecular weight of the compound. However, the accurate mass alone is not sufficient for determining the unambiguous molecular formula of natural products due to combinatorial explosion.[23] It has been demonstrated that even with 0.1 ppm mass accuracy for MS instrument, a unique molecular formula could not be determined when C, H, N, S, O, P atoms were included in the search list for molecules with molecular weight above 185.9760 Da[23]. Therefore, additional information, such as isotope abundance ratio, would be required for the determination.

The concept of mass defect is the result of different nuclear binding energies of different elements and their nuclides[24]. Conventionally, ^{12}C was defined as the element with zero mass defect while other nuclides have different mass defects depending on their relative nuclear binding energies to ^{12}C. Additionally, each element has a certain ratio of isotopes based on their natural abundance. Therefore, compounds with different isotopic compositions would show unique isotopic fine structure (IFS). Mass defects and natural abundance of common nuclides in natural products are displayed in Table 4.1.

Table 4.1. Mass defects and natural abundance of common nuclides in natural products

Element	Nuclide	Atomic Mass (u)	Mass Defect (u)	Natural abundance (%)
Hydrogen	^{1}H	1.00783	0.00783	99.9885
	^{2}H (D)	2.01410	0.01410	0.0115
Carbon	^{12}C	12.00000	0.00000	98.93
	^{13}C	13.00335	0.00335	1.07
Nitrogen	^{14}N	14.00307	0.00307	99.632
	^{15}N	15.00011	0.00011	0.368
Oxygen	^{16}O	15.99491	-0.00509	99.757
	^{17}O	16.99913	-0.00087	0.038
	^{18}O	17.99916	-0.00084	0.205
Sulfur	^{32}S	31.97207	-0.02793	94.93
	^{33}S	32.97146	-0.02854	0.76
	^{34}S	33.96787	-0.03213	4.29
Chlorine	^{35}Cl	34.96885	-0.03115	75.78
	^{37}Cl	36.96590	-0.03410	24.22
Bromine	^{79}Br	78.91834	-0.08166	50.69
	^{81}Br	80.91629	-0.08371	49.31

Here we demonstrate the IFS analysis based unambiguous molecular formula determination methodology development using a bacterial natural product echinomycin A (**27**)[25] as the standard analyte. The molecular formula of haliclonamycin, an antibacterial linear azole containing RiPP isolated from the marine natural product library based discovery platform, was determined unambiguously using the IFS analysis method despite the >2,000 M.W. of haliclonamycin. Haliclonamycin was produced by *Streptomyces* sp. WMMC-911 isolated from the sponge *Haliclona* sp. and showed anti-MRSA activity in the original screening. Due to high degree of signal overlap in NMR spectra, the NMR based structure elucidation of haliclonamycin remained a challenging problem. Whole genome sequencing of the producing organism WMMC-911 was conducted in order to perform genomic analysis and predict the peptide sequence of haliclonamycin as well as the functions of the post-translationally modifications related enzymes. These data combined with NMR, MS data can lead to the structure elucidation of the challenging compound haliclonamycin.

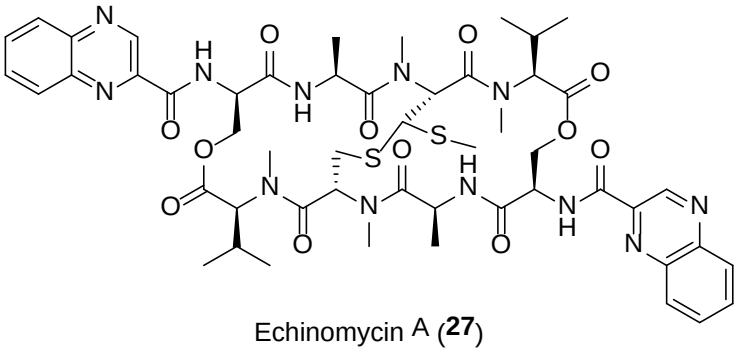
4.2. Results and Discussion

Different molecules have different elemental compositions and since natural abundance of different nuclides are certain numbers, unique isotopic fine structures of different molecules could be observed with the improved resolution of the MS instrument. To demonstrate this, direct infusion MS data of **27** was collected under ESI (electrospray ionization) positive mode on Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry) and displayed an intense $[M+H]^+$ adduct peak with *m/z* ratio of 1101.4275. Isotopologues spectra of **27** ($[MH+1]^+$, $[MH+2]^+$ and $[MH+3]^+$ ions) were also analyzed (Figure 4.1a) to reveal the isotopic fine structures formed by different isotopic compositions.

On the other hand, data acquired on a Bruker qTOF did not show resolved isotopic fine structures of those isotopologues (Figure 4.1b). SmartFormula was used to provide a range of reasonable formulas. The search was set to be $C_{48}S_1N_6O_6$ and the upper formula C_{54} based on preliminary NMR data. The MS error tolerance was set to be 2 ppm, and seven possible molecular formulas were given by SmartFormula analysis (Figure 4.1c): $[C_{51}H_{73}N_8O_{11}S_4]^+$, $[C_{51}H_{65}N_{12}O_{12}S_2]^+$, $[C_{50}H_{69}N_8O_{16}S_2]^+$, $[C_{48}H_{57}N_{22}O_6S_2]^+$, $[C_{52}H_{77}N_8O_6S_6]^+$, $[C_{52}H_{69}N_{12}O_7S_4]^+$ and $[C_{51}H_{61}N_{16}O_8S_2]^+$. For $[MH+1]^+$ ion, the difference between the two peaks was 0.0066 u, which matched the mass defect difference of $^{15}N^{12}C$ and $^{14}N^{13}C$ (0.0063 u). Similar calculations were performed and the different peaks observed in $[MH+2]^+$ and $[MH+3]^+$ ions were assigned to the combinations of $^{34}S^{12}C_2/^{32}S^{13}C_2$ and $^{34}S^{13}C^{12}C_2/^{32}S^{13}C_3$ respectively. In contrast, the spectra collected from qTOF instrument were not able to display similar fine structures (Figure 4.1b). Therefore, the accurate molecular formula could be identified from various possibilities by matching its unique IFS to the actual experimental MS spectra.

The IFS of three possible molecular formulae, $[C_{51}H_{73}N_8O_{11}S_4]^+$, $[C_{48}H_{57}N_{22}O_6S_2]^+$ and $[C_{51}H_{65}N_{12}O_{12}S_2]^+$ were simulated by Simulate Pattern function and the simulated MS peaks were overlaid with the actual spectra. (Figure 4.2) For $[MH+2]^+$ ion, the relative abundance of $[^{12}C_{51}H_{73}N_8O_{11}{}^{32}S_3{}^{34}S]^+$ was significantly higher than the actual molecule (Figure 4.2a), which indicated the tested compound should have less sulfur atoms than $[C_{51}H_{73}N_8O_{11}S_4]^+$. Similar trend was found in the $[MH+3]^+$ ion overlaid spectra. The tested compound was suggested to contain 2 sulfur atoms since the sulfur related peaks in $[MH+2]^+$ and $[MH+3]^+$ overlaid spectra matched the simulated IFS of $[C_{48}H_{57}N_{22}O_6S_2]^+$ (Figure 4.2b). However, the simulated $[C_{48}H_{57}{}^{14}N_{21}{}^{15}NO_6S_2]^+$ ion abundance and $[^{12}C_{47}{}^{13}CH_{57}{}^{14}N_{21}{}^{15}NO_6S_2]^+$ ion

abundance were higher than the corresponding ion abundance in the actual spectra, indicating this adduct formula was a mismatch. Similar evaluations were performed on $[C_{51}H_{65}N_{12}O_2S_2]^+$ and the simulated fine structure could match the actual spectra (Figure 4.2c), indicating the molecular formula of the tested compound was $C_{51}H_{64}N_{12}O_2S_2$, which matched the actual molecular formula of echinomycin A (**27**).



Echinomycin A (**27**)

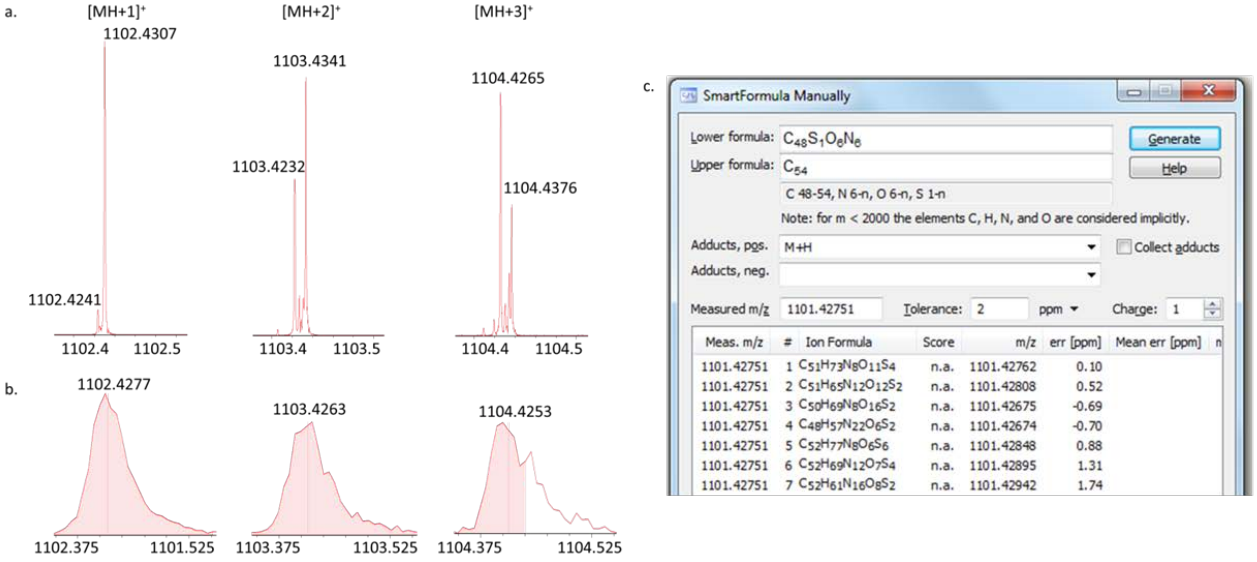


Figure 4.1. a.) Bruker 12T SolariX XR^{TM} MRMS spectra and b.) Bruker qTOF MS spectra of tested compound **27**. $[MH+1]^+$, $[MH+2]^+$, $[MH+3]^+$ isotopologue ions were displayed. c.) SmartFormula analysis of possible adducts formulae.

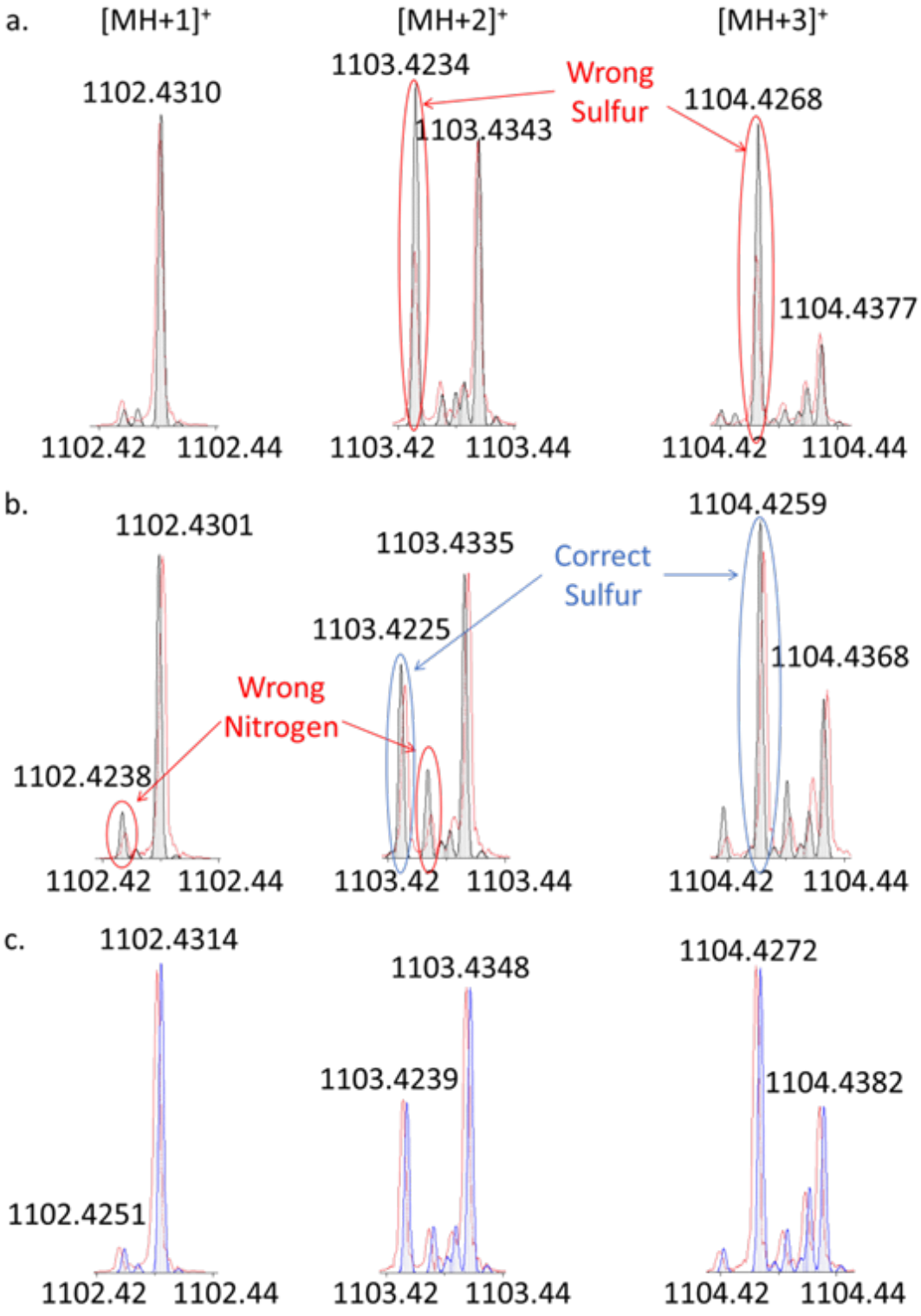


Figure 4.2. Overlaid IFS of the actual tested compound spectra (red) and simulated IFS (grey, blue) of (a.) $[C_{51}H_{73}N_8O_{11}S_4]^+$; (b.) $[C_{48}H_{57}N_{22}O_6S_2]^+$; (c.) $[C_{51}H_{65}N_{12}O_2S_2]^+$.

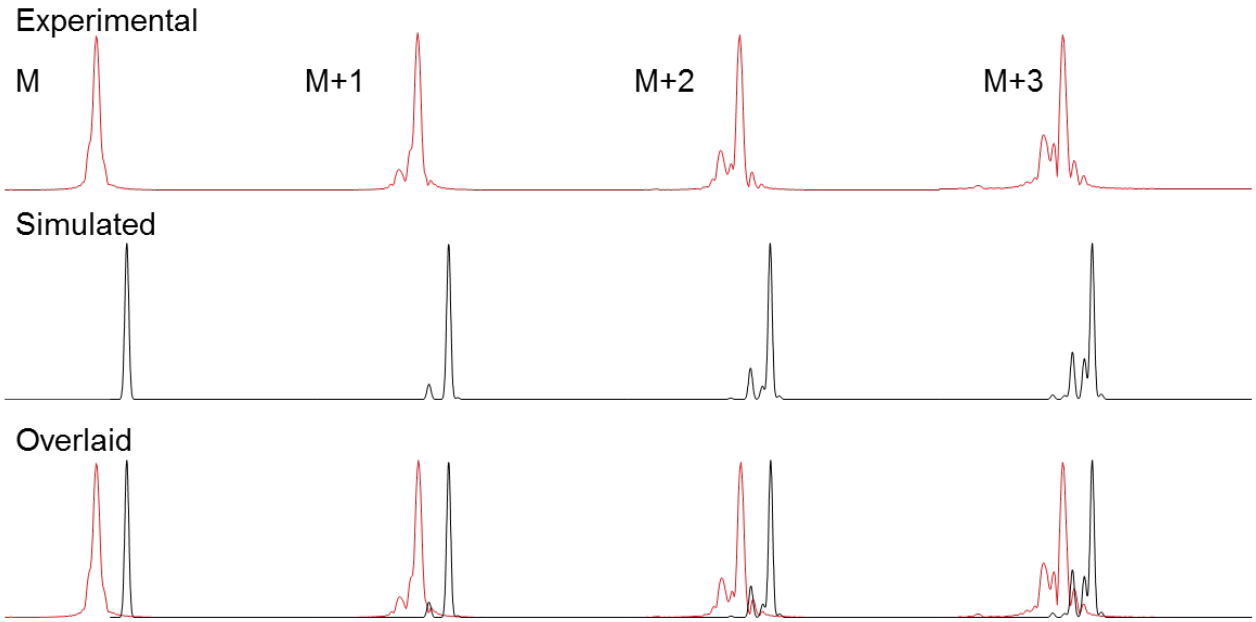


Figure 4.3. Experimental (red trace), simulated (grey), and overlaid IFS of the experimental $(C_{104}H_{112}N_{31}O_{26})^+$ ($[M+H]^+$) and its isotopologues acquired from Bruker MRMS instrument.

The molecular formula of haliclonamycin (*m/z* = 2210.8291, MH^+, Figure S71) was determined to be $C_{104}H_{111}N_{31}O_{26}$ based on HRMS data and the abovementioned IFS analysis based unambiguous molecular formula analysis (Figure 4.3). The predicted molecular formula of haliclonamycin $C_{104}H_{111}N_{31}O_{26}$ was selected among >20 possible choices with less than 2 ppm error by matching the simulated IFS of $C_{104}H_{111}N_{31}O_{26}$ to the experimental IFS collected on Bruker 12T SolariX XR^{TM} MRMS instrument. Preliminary NMR analysis of haliclonamycin (Figure S59-S64) revealed a possible linear azole containing peptide structure, a major class of RiPPs. Due to high level of signal overlap, only part of ^{1}H and ^{13}C chemical shifts are shown in Table 4.2:

Table 4.2. ^{1}H and ^{13}C NMR data for haliclonamycin (600 MHz for ^{1}H, 125 MHz for ^{13}C, $CD_3OD:CDCl_3$ = 1:1 (v:v))

Number	δ_C, type	δ_H, (J in Hz)	Number	δ_C, type	δ_H, (J in Hz)
1	176.7		47	128.9, CH	7.25, m
2	174.8		48	127.2	
3	174.7		49	108.2, CH_2	6.34, 5.56, s
4	174.4		50	107.2, CH_2	6.67, 5.73, s
5	174.2		51	107.1, CH_2	6.44, 5.55, s
6	174.1		52	104.2, CH_2	6.83, 5.79, s
7	174.0		53	65.3, CH	4.91, d (9.30)
8	173.9		54	62.7, CH_2	3.78, m
9	173.8		55	62.4, CH_2	3.93, m
10	173.4		56	59.8, CH	5.15, m
11	171.4		57	57.4	
12	170.9		58	57.3	
13	169.8, CH	8.56, s	59	57.1	
14	164.8		60	56.9	
15	164.6		61	56.8	
16	164.5		62	55.6, CH	6.01, t (5.5)
17	163.5		63	54.0, CH	5.10, t (5.5)
18	163.2		64	43.1, CH_2	m, 5.58, 4.10
19	163.1		65	39.9, CH_2	3.50, m
20	163.1		66	39.7	
21	162.6		67	39.5	
22	162.5		68	39.3	
23	162.3		69	39.2	
24	161.7		70	38.1, CH_2	4.55, m
25	158.9		71	37.6, CH_2	4.50, 4.15, m
26	158.7		72	37.4, CH_2	3.89, m
27	158.2		73	37.1, CH	2.52, m
28	157.5		74	32.3, CH_2	2.34, m
29	157.5		75	30.2, CH_2	1.25, m
30	156.5		76	25.6, CH_2	1.74, 1.32, m
31	155.2		77	22.7, CH_2	2.02, m

32	154.9		78	19.0, CH_3	1.68, d (7.2)
33	154.8		79	18.9, CH_3	1.70, d (7.2)
34	154.6		80	18.2, CH_3	1.51, d (8.0)
35	142.8, CH	8.04, s	81	18.0, CH_3	1.56, d (8.0)
36	142.3, CH	8.11, s	82	17.2, CH_3	1.72, d (7.2)
37	138.1		83	15.6, CH_3	1.18, d (7.0)
38	137.7		84	12.6, CH_3	2.72
39	130.6		85	11.8, CH_3	2.48
40	130.3, CH	7.15, t (7.2)	86	11.5, CH_3	2.36
41	130.1		87	11.5, CH_3	2.32
42	130.0, CH	7.24, t (8.0)	88	11.2, CH_3	2.31
43	129.8		89	11.1, CH_3	0.98, t (7.6)
44	129.4		90	10.9, CH_3	2.10
45	129.3				
46	129.1				

The carbon chemical shifts for six methyl groups, C-84, C-85, C-86, C-87, C-88, and C-90, are around 10-13 ppm, while the proton chemical shifts of those methyl groups are around 2.1-2.8 ppm. Two types of methyl groups can result in these proton and carbon chemical shifts: S-Me groups or methyl groups attached to heterocycles, such as oxazoles or thiazoles. Since the IFS analysis of haliclonamycin did not suggest a sulfur containing molecular formula, the possibilities of S-Me groups or thiazole attached methyl groups were excluded and suggested some oxazole containing structures. Five methyl groups belonging to alanine residues, C-78 to C-82, were also identified. C-83 and C-89 were identified as two methyl groups of the same isoleucine residues. C-64, C-71 were identified as the α-carbons of two glycine residues and C-62 was identified as the α-carbon of a phenylalanine residue. C-65, C-70 were identified as the β-carbons of two serine residues and C-49 to C-52 represented the β-carbons of four dehydroalanine (Dha) residues. Figure 4.4 summarized the

proposed amino acid residues based on current NMR analysis.

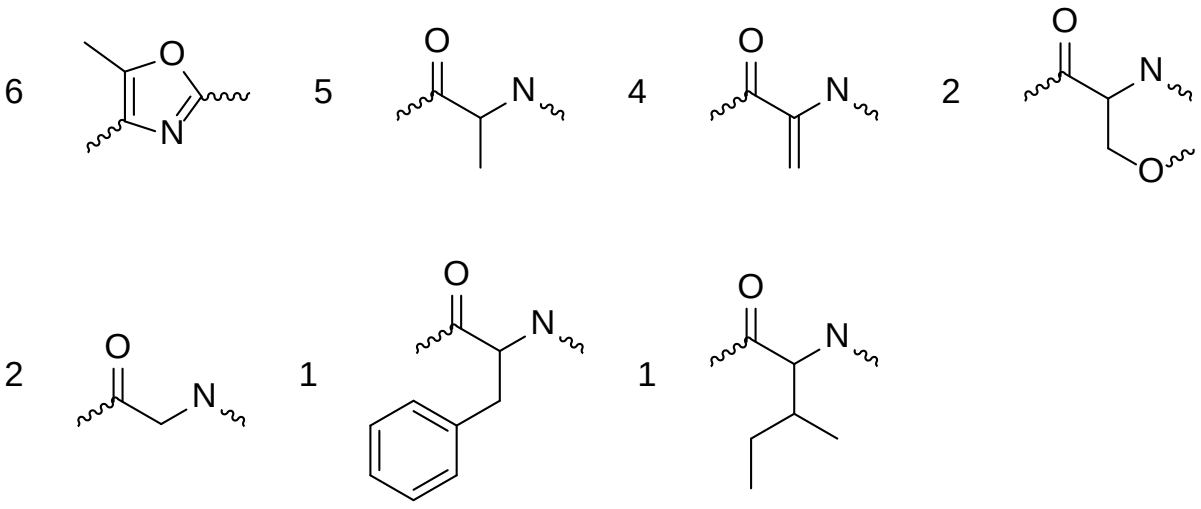


Figure 4.4. Proposed partial structures (spin systems) of haliclonamycin.

However, due to the lack of HMBC, TOCSY, ROESY correlations to link the whole peptide structure together, the complete structure elucidation of haliclonamycin cannot be finished at this time. Several attempts to develop solvent systems that can show the correlations between exchangeable protons (mostly amide N-H protons of amino acids residues) and the rest of the molecules were explored but only $CDCl_3$:CD_3OD = 1:1 (v:v) gave high quality spectra. Therefore, NMR experiments using solvent suppression of CD_3OH:$CDCl_3$ = 1:1 (v:v) were carried out to detect the exchangeable N-H protons (Figure S65-S70). A total number of 17 exchangeable protons were found based on 1D proton NMR (Figure S65) and ^{15}N-HSQC experiments(Figure S70). However, due to high degree of signal overlap, the connectivity cannot be solved even with additional exchangeable proton information. So DNA extraction of the producing organism WMMC-911 was performed and is currently being sequenced and assembled. Once we have those data available, we plan to use genomic analysis based RiPP structure elucidation method to understand the structure of

haliclonamycin. The preliminary NMR analysis we have done will provide valuable information for identifying the biosynthetic gene cluster, the whole precursor peptide, the core peptide sequence, and some of the post-translationally modifications when performing the genomic analysis and will essentially lead to results with higher confidence.

Haliclonamycin was tested for antibacterial activity against *E.coli* and Methicillin-resistant *Staphylococcus aureus* (MRSA) and displayed moderate activity against MRSA (MIC = 57.9 μM) and no significant activity against *E.coli* (MIC>200 μM). Since haliclonamycin belongs to the RiPP class, which owns great potential to undergo structure modifications while maintaining its biological compatibility[26], haliclonamycin can potentially be a good target for structure modifications to meet the optimized antibiotic activities.

4.3. Conclusion

In summary, we report the development of an Isotopic Fine Structure (IFS) analysis based unambiguous molecular formula identification method using a high resolution Bruker Magnetic Resonance Mass Spectrometry (MRMS). This method made significant contributions to discovery and efforts towards partial structural elucidation of haliclonamycin, an antibacterial linear azole containing ribosomally synthesized post-translationally modified peptide (RiPP) isolated from *Streptomyces* sp. WMMC-911. Due to high level of signal overlapping and the lack of 2D NMR correlations, only partial amino acid residues were proposed with the complete structure elucidation undone. DNA extraction of the producing organism has been done in order to study it whole genome in order to gain better understanding about the RiPP biosynthetic gene cluster, which combined with spectroscopic data will lead to the complete structure elucidation of haliclonamycin hopefully in the near

future.

4.4. Material and Methods

General Experimental Procedures

Optical rotations were measured on a Perkin–Elmer 241 Polarimeter. UV spectra were recorded on an Aminco/OLIS UV-Vis spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR spectrophotometer. NMR spectra were obtained in CD_3OD:$CDCl_3$ = 1:1 (v:v) and CD_3OH:$CDCl_3$ = 1:1 (v:v) (δ_H 3.34 ppm, δ_C 49.0 ppm) with a Bruker Avance 600 III MHz spectrometer equipped with a $^{1}H\{^{13}C/^{15}N/^{31}P\}$ cryoprobe, a Bruker Avance III 500 MHz spectrometer equipped with a $^{13}C/^{15}N\{^{1}H\}$ cryoprobe, and a Bruker Avance III HD 400 MHz spectrometer. HRMS data were acquired with a Bruker MaXis™ 4G ESI-QTOF mass spectrometer. RP HPLC was performed using a Shimadzu Prominence HPLC system equipped with a Phenomenex Gemini C18 column (250 × 30 mm) and Phenomenex Luna C18 column (250 × 10 mm). UHPLC-HRMS was acquired using a Bruker MaXis™ 4G ESI-QTOF mass spectrometer coupled with a Waters Acquity UPLC system operated by Bruker Hystar software and a C18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm). MS data for IFS analysis was collected using a Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry).

Biological Material

Sponge specimens were collected on May 11, 2015 near the Florida Keys (24°39′66.3’’N, 81°26’46.3’’W). A voucher specimen is housed at the University of Wisconsin-Madison. For cultivation, a sample of sponge (1 cm^3) was ground in 500 μL sterile seawater and dilutions were made using 500 μL sterile seawater. Subsequently, 400 μL of diluted sponge sample was

added to 200 μL of sterile seawater and 100 μL was plated using a sterile L-shaped spreader. Diluted sample was plated on Gauze 1^{27} media supplemented with artificial seawater. Each medium was supplemented with 50 μg/mL cycloheximide, 25 μg/mL nystatin, and 25 μg/mL nalidixic acid. Plates were incubated at 28˚C and colonies were isolated over the course of two months.

Fermentation, Extraction and Isolation

Two 10 mL seed cultures (25 × 150 mm tubes) in medium DSC (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract per liter of artificial seawater) were inoculated with strain WMMC-911 and shaken (200 RPM, 28 °C) for seven days. Two liter flask (1 × 500 mL) containing ASW-A (20 g soluble starch, 10 g D -glucose, 5 g peptone, 5 g yeast extract, 5 g $CaCO_3$ per liter of artificial seawater. Four-liter flasks (10 × 1 L) containing medium ASW-A with Diaion HP20 (7% by weight) were inoculated with 50 mL from the 500 mL culture and shaken (200 RPM, 28 °C) for seven days. For making artificial sea water, solutions I (415.2 g NaCl, 69.54 g Na_2SO_4, 11.74 g KCl, 3.40 g $NaHCO_3$, 1.7 g KBr, 0.45 g H_3BO_3, 0.054 g NaF) and II (187.9 g $MgCl_2 \cdot 6H_2O$, 22.72 g $CaCl_2 \cdot 2H_2O$, 0.428 g $SrCl_2 \cdot 6H_2O$) were made up separately using distilled water and combined to give a total volume of 20 L.

Filtered HP20 and cells were washed with H_2O and extracted with acetone. The acetone extract was subjected to liquid-liquid partitioning using 30% aqueous MeOH and $CHCl_3$ (1:1). The $CHCl_3$-soluble partition (3.12 g) was fractionated by Sephadex LH20 column chromatography (column size 500 × 40 mm, $CHCl_3$:MeOH=1:1, 20 mL for each fraction). Fractions containing haliclonamycin (248.2 mg) were subjected to preparative RP HPLC

(20%/80% to 100%/0% MeOH/H_2O (with 0.1% acetic acid) over 25.0 min gradient, 20 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm). The fraction collected between 27.0−28.5 minutes was further fractionated by RP HPLC (25%/75% to 53%/47% MeCN/H_2O (with 0.1% acetic acid), 40 min, 4.5 mL/min) using a Phenomenex Luna C18 column (250 × 10 mm), yielding haliclonamycin (8.0 mg, t_R 30.1 min).

Haliclonamycin: light yellow amorphous solid, $[\alpha]^{25}_D$ = +54.4 (*c* 5.3, MeOH:$CHCl_3$ = 1:1 (v:v)); UV-Vis (MeOH): λ_{max} (log ε) 219 nm (4.42), 228 nm (4.44), 232 nm (4.45), 265 nm (3.72), 279 nm (3.42); IR (ATR): υ_{max} 3360, 3344, 3330, 3301, 3266, 3246, 3228, 3204, 2977, 2959, 2948, 2835, 1643, 1595, 1543, 1453, 1406, 1314, 1272, 1214, 1160, 1112, 1022, 950, 926, 878, 851, 827, 754, 720, 702, 688, 672, 659 cm^{-1}; HRMS $[M+H]^+$ *m/z* = 2210.8342 (calcd. for $C_{104}H_{111}N_{31}O_{26}^+$ 2210.8389). ^{1}H NMR (600 MHz, CD_3OD:$CDCl_3$ = 1:1 (v:v)) δ 8.56, (s, 1H), 8.11 (s, 1H), 8.04 (s, 1H), 7.25 (m, 1H), 7.24 (t, *J* = 8.0 Hz, 2H), 7.15 (t, *J* = 7.2 Hz, 2H), 6.83, (s, 1H), 6.67, (s, 1H), 6.44, (s, 1H), 6.34, (s, 1H), 6.01 (t, *J* = 5.5 Hz, 1H), 5.79, (s, 1H), 5.73, (s, 1H), 5.57-5.59, (m, 1H), 5.56, (s, 1H), 5.55, (s, 1H), 5.13-5.17, (m, 3H), 5.10, (t, *J* = 5.5 Hz, 1H), 4.91, (d, *J* = 9.3 Hz, 1H), 4.57-4.49, (m, 4H), 4.15-4.10, (m, 2H), 3.93-3.50, (m, 12H), 2.72, (s, 3H), 2.52-2.30 (m, 6H), 2.48, (s, 3H), 2.36, (s, 3H), 2.32, (s, 3H), 2.31, (s, 3H), 2.10, (s, 3H), 2.04-2.00 (m, 1H), 1.76-1.72 (m, 1H), 1.72, (d, *J* = 7.2 Hz, 3H), 1.70, (d, *J* = 7.2 Hz, 3H), 1.68, (d, *J* = 7.2 Hz, 3H), 1.56 (d, *J* = 8.0 Hz, 3H), 1.51 (d, *J* = 8.0 Hz, 3H), 1.32-1.24 (m, 2H), 1.18, (d, *J* = 7.0 Hz, 3H), 0.98, (t, *J* = 7.6 Hz, 3H). ^{13}C NMR (125 MHz, CD_3OD:$CDCl_3$ = 1:1 (v:v)) δ176.7, 174.8, 174.7, 174.4, 174.2, 174.1, 174, 173.9, 173.8, 173.4, 171.4, 170.9, 169.8, 164.8, 164.6, 164.5, 163.5, 163.2, 163.1, 163.1, 162.6, 162.5, 162.3, 161.7, 158.9, 158.7, 158.2, 157.5, 157.5, 156.5, 155.2, 154.9, 154.8, 154.6,

142.8, 142.3, 138.1, 137.7, 130.6, 130.3, 130.1, 130.0, 129.8, 129.4, 129.3, 129.1, 128.9, 127.2, 108.2, 107.2, 107.1, 104.2, 65.3, 62.7, 62.4, 59.8, 57.4, 57.3, 57.1, 56.9, 56.8, 55.6, CH, 54.0, 43.1, 39.9, 39.7, 39.5, 39.3, 39.2, 38.1, 37.6, 37.4, 37.1, 32.3, 30.2, 25.6, 22.7, 19.0, 18.9, 18.2, 18.0, 17.2, 15.6, 12.6, 11.8, 11.5, 11.5, 11.2, 11.1, 10.9.

Antibacterial Testing

Haliclonamycin was tested for antibacterial activity against *E. coli* (ATCC #25922) and Methicillin-resistant *Staphylococcus aureus* (MRSA) (ATCC #33591), and MICs were determined using a dilution antimicrobial susceptibility test for aerobic bacteria[28]. Haliclonamycin were dissolved in DMSO and serially diluted to 10 concentrations (0.25−128 μg/mL) in 96-well plates. Vancomycin was used as a positive control against *M. leuteus* and MRSA and exhibited an MIC of 0.25 μg/mL. Gentamicin was used as a positive control against *E. coli* and exhibited an MIC of 4 μg/mL. Haliclonamycin, vancomycin and gentamicin were tested in triplicates. On each plate, there were six untreated media controls. The plates were incubated at 37 °C for 18 h. The MIC was determined as the lowest concentration that inhibited visible growth of bacteria.

Whole Genome Sequencing

DNA extraction of WMMC-911 was conducted as previously described[29]. The complete genome was sequenced at the University of Wisconsin Biotechnology Center (UWBC) using PacBio RSII (Pacific Biosciences) technology.

4.5. References

(1). McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.*, **2009**, *26*, 537-559.

(2). Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G.,

Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., van der Donk, W. A. *Nat. Prod. Rep.*, **2013**, *30*, 108−160.

(3). Velásquez, J. E.; van der Donk, W. A. *Curr. Opin. Chem. Biol.*, **2011**, *15*, 11-21.

(4). Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 5879-5884.

(5). Haft, D. H.; Basu, M. K. *J. Bacteriol.*, **2011**, *193*, 2745-2755.

(6). Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Müller, R. *Chem. Biol.*, **2012**, *19*, 1278-1287.

(7). Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.*, **2010**, *6*, 9-18.

(8). Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.*, **2011**, *133*, 5016-5023.

(9). Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.*, **2008**, *4*, 341-343.

(10). Hou, Y.; Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Michel, C. R.; Ellis, G. A.; Vazquez-Rivera, E.; Braun, D. R.; Rose, W. E.; Schmidt, E. W.; Bugni, T. S. *Org. Lett.*, **2012**, *14*, 5050-5053.

(11). Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. *Nucleic Acids Research*, **2017**, *45*, W80-W88.

(12). Knerr, P. J.; van der Donk, W. A. *Annu. Rev. Biochem.*, **2012**, *81*, 479-505.

(13). Claesen, J.; Bibb, M. J. *Proc. Natl. Acad. Sci. U. S. A.*, **2010**, *107*, 16297-16302.

(14). Hamada, T.; Matsunaga, S.; Yano, G.; Fusetani, N. *J. Am. Chem. Soc.*, **2005**, *127*, 110-118.

(15). Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T.; *Science*, **1996**, *274*, 1188-1193.

(16). Ireland, C.; Scheuer, P. J. *J. Am. Chem. Soc.*, **1980**, *102 (17)*, 5688-5691.

(17). Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. *Chem. Rev.*, **2005**, *105*, 685-714.

(18). Waisvisz, J. M.; van der Hoeven, M. G.; van Peppen, J.; Zwennis, W. C. M. *J. Am. Chem. Soc.*, **1957**, *79 (16)*, 4520-4521.

(19). Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. *Mol. Microbiol.*, **2007**, *65*, 1380-1394.

(20). Maksimov, M. O.; Pan, S. J.; Link, A. J. *Nat. Prod. Rep.*, **2012**, *29*, 996-1006.

(21). Dührkop, K.; Hufsky, F.; Böcker, S., *Mass Spectrom (Tokyo)* **2014**, *3*(3): S0037; DOI: 10.5702/massspectrometry.S0037

(22). Guan, S.; Marshall, A. G., *Anal. Chem.* **1996**, *68*(1), 46-71.

(23). Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*:234.

(24). Sleno, L., *J. Mass. Spectrom.* **2012**, *47*, 226-236.

(25). Dell, A.; Williams, D. H.; Morris, H. R.; Smith, G. A.; Feeney, J.; Roberts, G. C., *J. Am. Chem. Soc.* **1975**, *97*(9), 2497-2502.

(26). Murphy, K.; O'Sullivan, O.; Rea, M. C.; Cotter, P. D.; Ross, R. P.; Hill, C. *Plos One*, **2011**, *6*, e20582.

(27). Gauze, G. F.; Preobrazhenskaya, J. P.; Kudrina, E. E.; Blinov, N. O.; Ryabova, I. D.; Sveshnikova, M. A. **1957**. Problems in the classification of antagonistic actinomycetes. State Publishing House for Medical Literature. Medgiz, Moscow.

(28). National Committee for Clinical Laboratory Standards. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 7th ed.; NCCLS: Villanova, PA, USA, 2006; Approved standard M7-A7.

(29). Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S. *ACS Chem. Biol.* **2017**, *12*, 3093−3102.

Chapter 5:

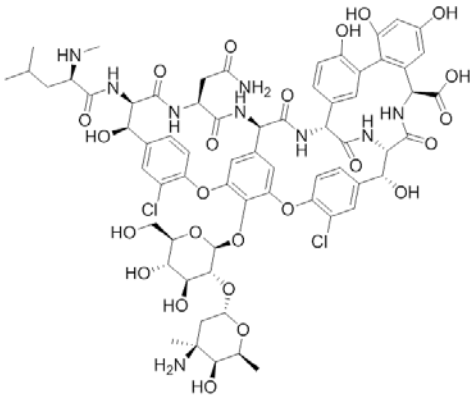
Discovery and structure elucidation of two novel compounds with polyketide-nonribosomal peptide structures, ecteinamines A and B.

5.1. Introduction

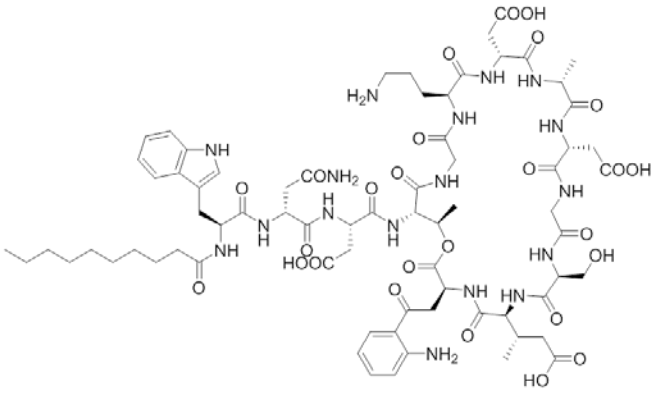
In this chapter, the discovery, planar structure elucidation, and efforts towards understanding the behavior of ^{15}N chemical shifts in complicated natural products of two novel compounds synthesized via polyketide-nonribosomal peptide biosynthetic pathway will be described. Isotopic fine structure (IFS) analysis unambiguous molecular formula determination method was applied in the rapid unambiguous molecular formulae identifications and planar structure elucidations of these two novel compounds.

Nonribosomal peptides represent a major class of natural products with diverse chemical structures as well as a wide range of biological activities[1]. Different from the ribosomal peptides discussed mainly in Chapter 4, nonribosomal peptides usually feature heterocyclic structures, D-amino acids, glycosylated structures, N-methylations, and a large variety of structurally modification amino acid residues that are different from the 20 common amino acids used in the ribosomal peptide biosynthesis[2]. In the 1970s, Lipmann and co-workers reported the nucleic-acid-independent syntheses of two peptide antibiotics, gramicidin S and tyrocidine A[3]. These two peptides were biosynthesized by large enzyme complexes and this report opened up the door for discovery and understanding of novel nonribosomal peptides and their biosynthetic mechanisms. Generally the pharmacophore of nonribosomal peptides

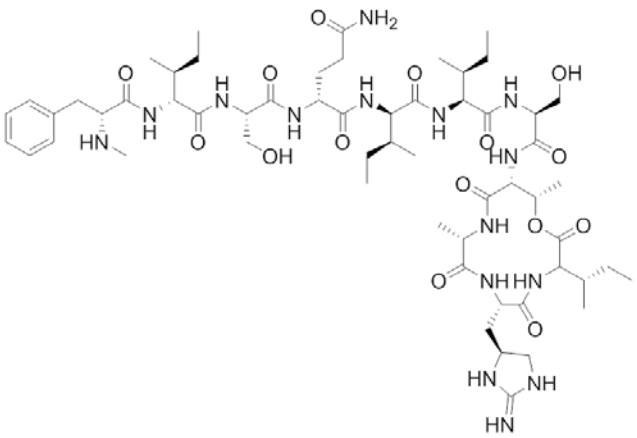
can take advantage of the high biological compatibility the peptide sequence that can help them to be better recognized and bounded to their biological targets[4]. The modified chemical structures, such as glycosylations, heterocyclizations, aliphatic chain attachments, and D-amino acids, can give more variances to these pharmacophores and can potentially lead to better chemical stabilities and biological activities[5]. Some of great examples of nonribosomal peptide antibiotics include vancomycin (**28**)[6], daptomycin (**29**)[7], teixobactin (**30**)[8], which also highlights the importance for nonribosomal peptide discovery since this class of compounds are still proving us with new classes of antibiotics with diverse antibiotic mechanisms in the past 20 years.



Vancomycin (**28**)



Daptomycin (**29**)



Teixobactin (**30**)

As discussed in Chapter 1 earlier, traditional antibiotics discovery was based mainly on

the Waksman platform[9], which is mining the secondary metabolites of soil derived actinomycetes and follow certain compounds guided by antibiotic biological activities. In recent years, this traditional strategy is leading to high known compound rediscovery rate (> 99.5%), which is causing huge costs for traditional antibiotics discovery and forcing large pharmaceutical companies to leave this field[10]. In Chapters 2-4, we have demonstrated new analytical methods and concepts including LC/MS metabolomics analysis (PCA analysis)[11] and generation of natural product library discovery platform for high throughput screening[12] can be good solutions to the challenges antibiotics discovery from natural products is facing. Here we are introducing another possible solution, which is following some new mechanisms of action towards the treatment of microbial infections.

Human immune response activities, also known as host defense activities, are the oldest and most important mechanisms towards the treatments of infectious diseases[13, 14] (Figure 5.1). Inevitably, humans can get infected by the numerous microorganisms living around us; however, even without antibiotic treatment, these infectious will not lead to mortality in most cases and that should be attributed to the functions of our immune response systems. Macrophages can clear the invading pathogens by phagocytosis and produce cytokines, which are a relatively diverse group of proteins with low molecule weight (usually 8,000 – 30,000 daltons) that act as transmitters during the human immunological activities[15]. They can be classified into two major classes: pro-inflammatory cytokines, which will stimulate further immunological activities; or anti-inflammatory cytokines, which will suppress further immunological activities. The effects of the pro-inflammatory and anti-inflammatory cytokines on immunological activities have been studied in recent years[16]. Rose and

co-workers observed elevated ratio between the serum concentrations of anti-inflammatory cytokine IL-10 and pro-inflammatory cytokine TNF-α in dead patients with endovascular *Staphylococcus aureus* inoculums[17, 18] compared to the survival patients. In addition, Minejima and co-workers compared the plasma or serum cytokine concentrations of pro-inflammatory cytokines and anti-inflammatory cytokine IL-10 in patients infected by *Staphylococcus aureus*[19] before and after antibiotic treatments. The resolving group had significant lower IL-10/TNF-α ratios after 72 hours of antibiotic treatments compared to the persistent group. They also mentioned the increased IL-10/TNF-α ratio might result in a state of "immunoparalysis", which thwarts the host's ability to clear the primary infection.

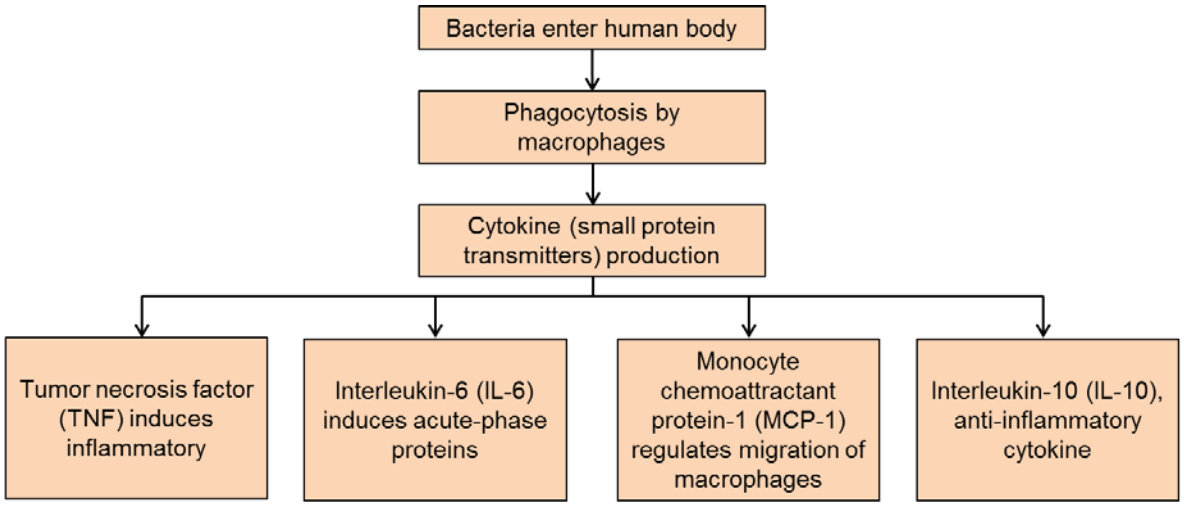


Figure 5.1. Mechanism of human immunological activities and the functions of some classes of pro-inflammatory cytokines (TNF, IL-6, MCP-1) and anti-inflammatory cytokines (IL-10).

Inspired by these immunological studies, we wanted to test if our marine bacteria natural product library can modulate the production levels of different pro-inflammatory and anti-inflammatory cytokines, which could potentially lead to modulated human immunological activities. In a proof of concept study (collaboration with Dr. John-Demian

Sauer's lab), we applied fractions of our natural product library to a mouse macrophage based assay and tested the produced cytokine concentrations (Figure 5.2). We also wanted to test if some of the pure compounds isolated from the initial active fractions could show similar activities. Following the initial hits, three bacteria strains WMMB-272 (*Streptomyces* sp.), WMMB-303 (*Streptomyces* sp.), and WMMB-482 (*Micromonospora* sp.) were prioritized to explore cytokines modulating compounds. Two classes of known compounds, phenelfamycin analogs (**31**)[20] and pyridinopyrone A (**32**)[21] (Figure 5.3a), were isolated from secondary metabolites of WMMB-272. They have shown activities on the concentrations of both IL-10 and TNF-α cytokines compared to the untreated group and the DMSO control group (Figure 5.3b), which reproduced the results of the initial screening using fractions of the natural product library, which are relatively more complicated mixtures.

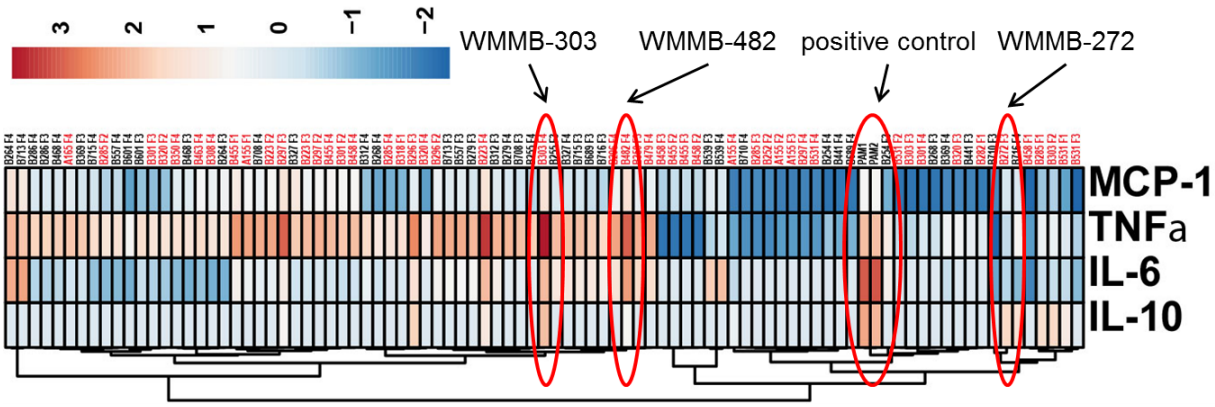


Figure 5.2. Result of the mouse macrophage based assay treated by the marine bacteria natural product fractions. Fluorescence activated cell sorting (FACS) was used to sort macrophages with different cytokine concentrations. Different colors represent log-scale differences in cytokine concentrations between untreated macrophages and macrophages treated with different fractions. Pam_3Cys-Ser-$(Lys)_4$ (PAM) was used as a positive control.

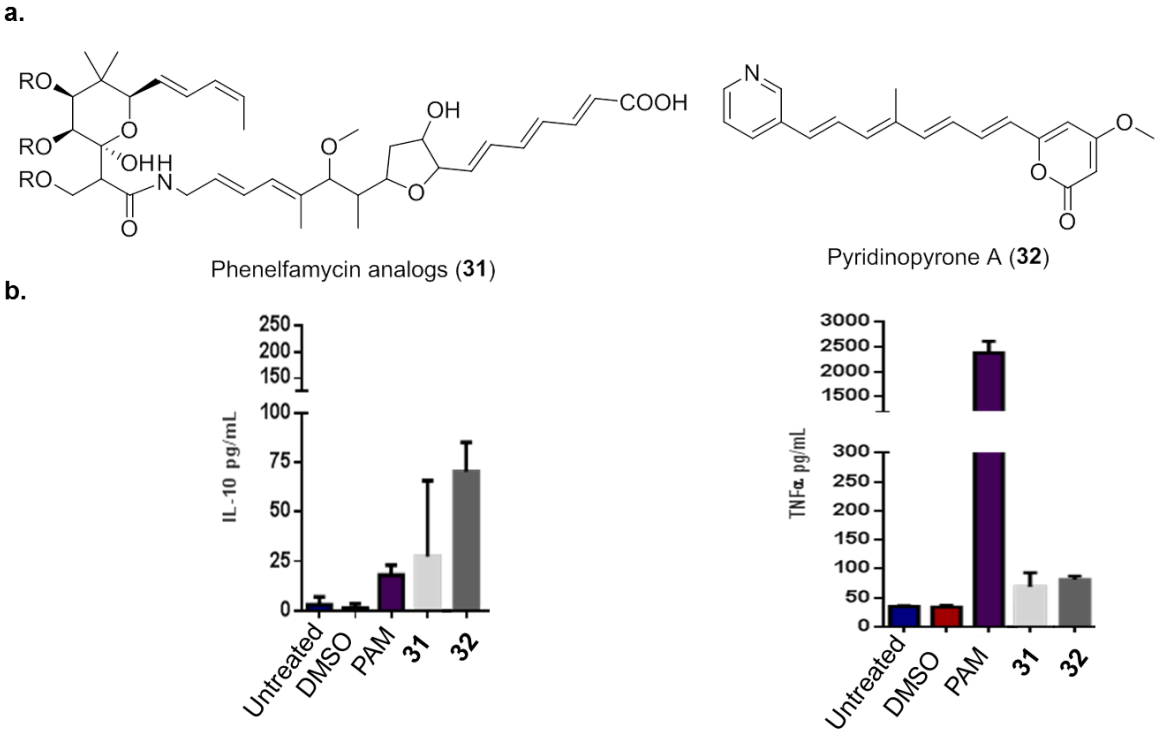
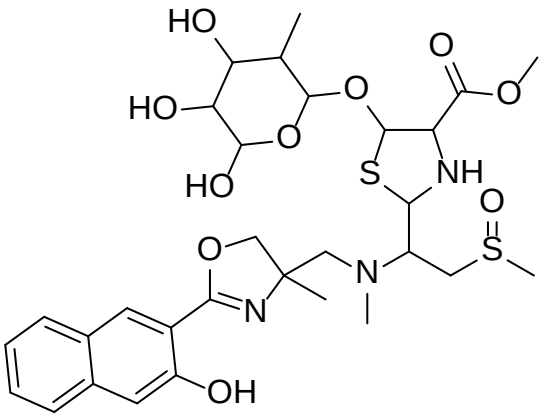


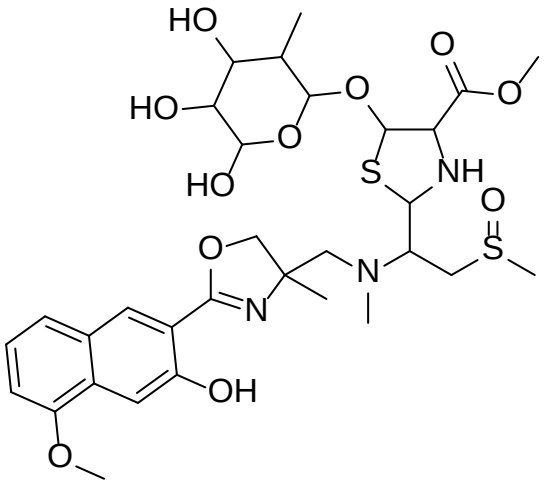
Figure 5.3. (a). Structures of phenelfamycin analogs (**31**) and pyridinopyrone A (**32**). **(b)**. Concentrations of IL-10 and TNF-α cytokines produced by mouse macrophages after treated by **31** and **32**.

Besides the known compounds discovered from *Streptomyces* sp. WMMB-272, we have discovered two classes of new non-ribosomal peptides, ecteinamine A (**33**), ecteinamine B (**34**), and streptnatamide A (**37**), produced by two bacteria, *Micromonospora* sp. WMMB-482 and *Streptomyces* sp. WMMB-303 respectively. Streptnatamide A produced by *Streptomyces* sp. WMMB-303 will be discussed in Chapter 6. Isotopic Fine Structure analysis was applied to the unambiguous molecular formula identification and rapid planar structure elucidations of ecteinamines A (**33**) and B (**34**). Modern NMR experiments including ^{13}C-^{13}C COSY, ^{15}N-coupled vs decoupled 1D ^{13}C experiments were performed in order to confirm the novel ecteinamines. Acid base NMR experiments on ecteinamine B will provide more information

for the natural product community to better understand the behavior of ^{15}N chemical shifts in complicated natural products. The discovery and structural elucidation of these two compounds highlight the strength of applying modern analytical techniques to speed up the structural elucidation of complicated natural products.



Ecteinamine A (**33**)



Ecteinamine B (**34**)

5.2. Results and Discussion

Molecular formula of ecteinamine B (**34**) was determined unambiguously by analyzing the Isotopic Fine Structure of **34** (Figure S91). Direct infusion MS data of **34** was collected under ESI (electrospray ionization) positive mode on Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry) and displayed an intense $[M+H]^+$ adduct peak with *m/z* ratio of 698.2412. Isotopologues spectra of **34** ($[MH+1]^+$, $[MH+2]^+$ and $[MH+3]^+$ ions) were also analyzed (Figure 5.4) to reveal the isotopic fine structures formed by different isotopic compositions. The molecular formula of **34** was determined as $C_{31}H_{43}N_3O_{11}S_2$ as the simulated IFS of this formula matched the acquired MS spectra of all major isotopologues analyzed. The monoisotopic peak showed only -0.1 ppm error compared to the predicted

molecular formula, which highlighted the accuracy of the MRMS instrument when calibrated properly. For $[MH+2]^+$ and $[MH+3]^+$ isotopologue ions, isotope peaks containing ^{34}S became dominant since the natural abundance of ^{34}S is 4.29%, which is significantly higher than that of ^{13}C, ^{15}N, and ^{18}O. The molecular formula of ecteinamine A (**33**) (Figure S90) was determined similarly to be $C_{30}H_{41}N_3O_{10}S_2$ by performing the IFS analysis of **33**.

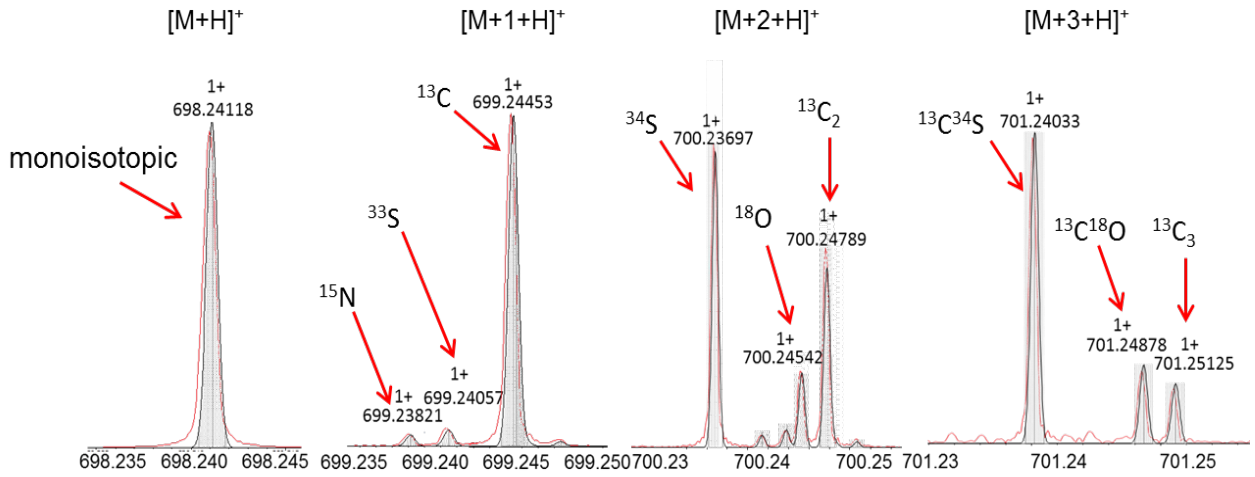


Figure 5.4. Overlaid MS spectra of the experimental (red trace), simulated (grey) IFS of $(C_{31}H_{44}N_3O_{11}S_2)^+$ ($[M+H]^+$) and its isotopologues acquired from the Bruker MRMS instrument.

In order to gain more structure information of **34**, MS^2 (Figure S92) and MS^3 (Figure S93) analysis of **34** was performed to investigate the IFS of the fragment ions. The major MS^2 fragment of **34** was *m/z* = 634.2424 and the IFS analysis revealed a loss of $C_1H_4O_1S_1$, which indicated a $-SOCH_3$ functional group. This MS^2 fragment ion underwent further fragmentation in MS^3 analysis and resulted in a major fragment ion with *m/z* ratio of 488.1846. IFS analysis suggested the loss of a sugar unit with molecular formula of $C_6H_{10}O_4$. The MS^2 and MS^3 analysis of $[MH+2]^+$ isotopologue ion of **34** was summarized in Figure 5.5.

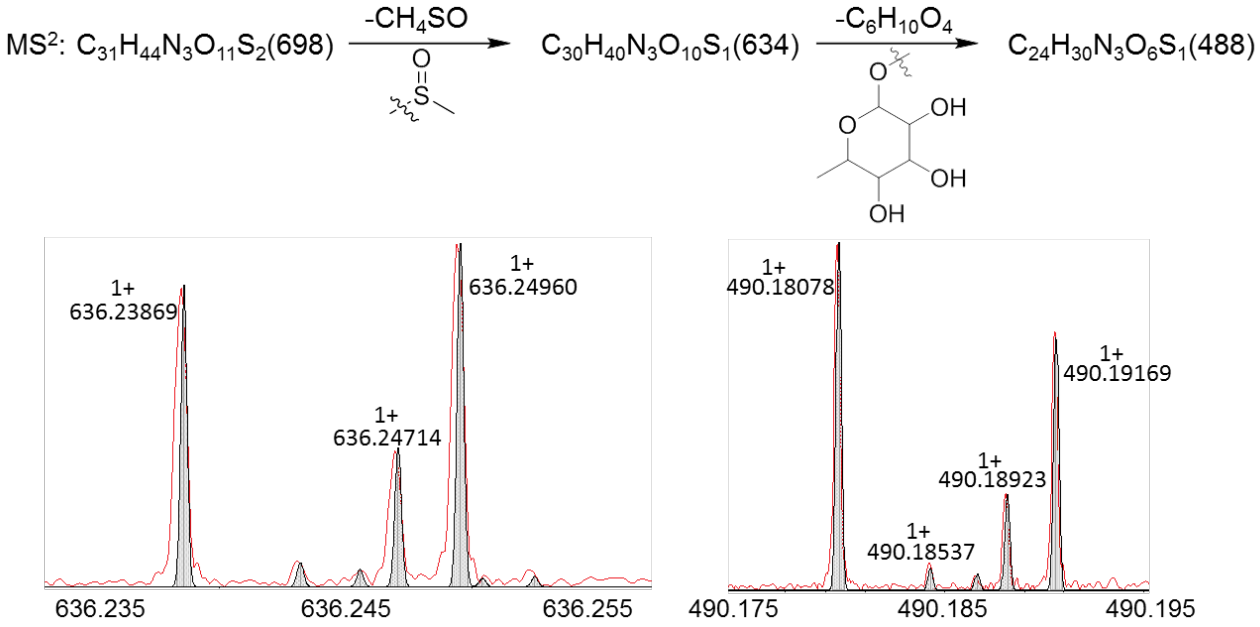
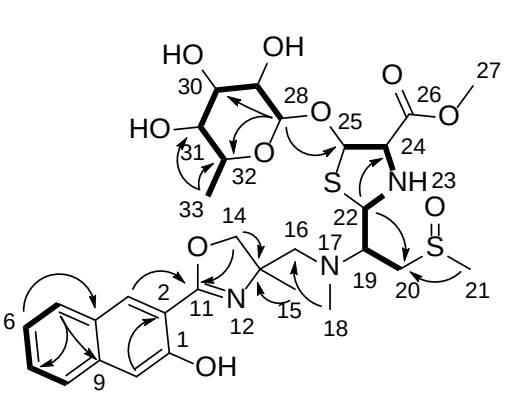


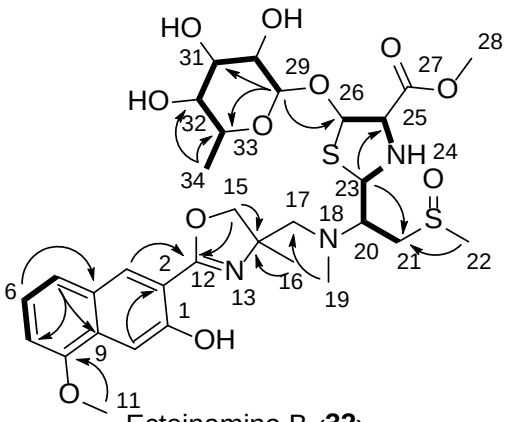
Figure 5.5. Overlaid MS spectra of the experimental (red trace), simulated (grey) IFS of $C_{30}H_{40}N_3O_{10}S_1$ $[MH+2]^+$ fragment ion and $C_{24}H_{30}N_3O_6S_1$ $[MH+2]^+$ fragment ion using Bruker MRMS instrument.

With the molecular formula and some of the key functional groups of **33** and **34** determined unambiguously, rapid structure elucidations of the planar structures of both compounds were completed by analyzing the 1D and 2D NMR spectra (Figure S72-S83). The ^1H and ^{13}C chemical shifts, COSY correlations and some key HMBC correlations of both compounds are summarized in Table 5.1.

Table 5.1. ^{1}H and ^{13}C NMR data for **33** and **34** (600 MHz for ^{1}H, 125 MHz for ^{13}C, MeOD)



Ecteinamine A (**31**)



Ecteinamine B (**32**)

^{1}H-^{1}H COSY ⟶ HMBC (key)

Position	**33**		**34**	
	δ_C (δ_N), type	δ_H, (J in Hz)	δ_C (δ_N), type	δ_H, (J in Hz)
1	155.0		155.9	
2	114.4		115.2	
3	129.9, CH	8.34, s	130.1, CH	8.27, s
4	136.7		129.5	
5	129.0, CH	7.91, d (8.3)	122.1, CH	7.45, d (8.3)
6	124.4, CH	7.38, t (7.9)	124.8, CH	7.27, t (7.9)
7	128.6, CH	7.54, t (7.6)	106.7, CH	6.94, d (7.6)
8	126.7	7.80, d (7.6)	155.4	
9	125.8		130.1	
10	111.6, CH	7.36, s	107.1, CH	7.67, s
11	164.5		56.1, CH_3	4.05, s
12			165.8	
13	71.7		228.7[a]	
14	74.1, CH_2	4.58-4.65, m 4.23, d (7.9)	72.9	
15	25.9, CH_3	1.32, s	75.4, CH_2	4.65, d (7.9) 4.22, d (7.9)

16	62.3, CH_2	3.15-3.21, m 2.40, d (15.2)	26.4, CH_3	1.32, s
17			62.6, CH_2	3.11-3.15, m 2.42, d (14.8)
18	37.8, CH_3	2.33, s	27.5[a]	
19	62.3, CH	3.26-3.33, m	38.5, CH_3	2.34, s
20	56.0, CH_2	3.11-3.17, m 3.01, dd (13.2, 2.7)	63.0, CH	3.24-3.29, m
21	39.3, CH_3	2.73, s	56.3, CH_2	3.11-3.15, m 3.03, dd (13.0, 2.6)
22	71.9, CH	4.79, d (11.4)	39.8, CH_3	2.73, s
23			72.6, CH	4.82, d (11.6)
24	71.0, CH	3.77, d (3.4)	49.3[a], NH	3.41-3.47, m[b]
25	91.9, CH	3.78-3.82, m	71.0, CH	3.62, d (3.8)
26	168.2		91.5, CH	3.92-3.94, m
27	52.0, CH_3	3.69, s	169.6	
28	103.9, CH	3.16-3.20, m	52.2, CH_3	3.59, s
29	73.8, CH	2.82, m	102.7, CH	3.86, s
30	76.8, CH	2.89, t (9.4)	71.6, CH	3.22-3.25, m
31	76.1, CH	2.77, t (9.4)	72.0, CH	3.28-3.33, m
32	72.2, CH	2.56, dq (9.2, 6.4)	73.7, CH	3.24, t (9.4)
33	17.5, CH_3	1.16, d (6.4)	70.6, CH	3.57, dq (9.4, 6.4)
34			17.6, CH_3	1.16, d (6.4)

[a.] ^{15}N chemical shift acquired from ^{15}N-HMBC experiment. [b.] Acquired in d_6-DMSO

In order to gain more evidence to validate the proposed novel planar structures of **33** and **34**, ecteinamine B was labeled with ^{13}C, ^{15}N isotopes by using isotopic enrich bacteria culture media. Direct connectivities between C-C bonds and C-N bonds were detected by running ^{13}C-^{13}C COSY and ^{15}N coupled/decoupled (proton decoupled for both cases) 1D ^{13}C NMR

experiments (Figure S87-S89) on the labeled compound (Figure S94-S95). The ^{13}C-^{13}C COSY correlations, C-N connectivities (labeled as ^{13}C-^{15}N HMQC correlations), and ^{15}N-HMBC correlations are shown in Figure 5.6.

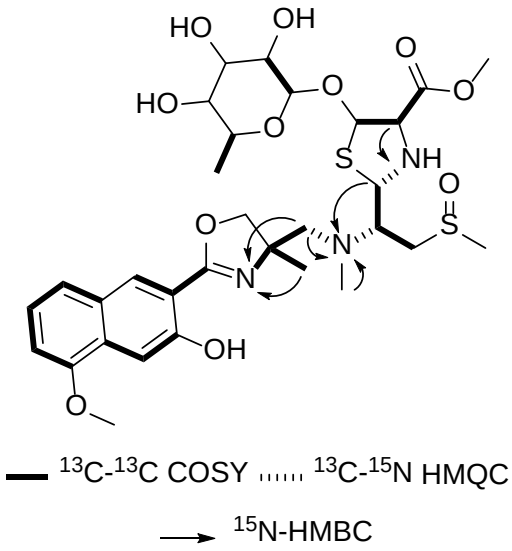


Figure 5.6. Carbon-carbon bond connectivities, carbon-nitrogen bond connectivities, and ^{15}N-HMBC correlations of ^{13}C, ^{15}N isotopic enriched ecteinamine B (**34**).

The structure elucidation of **34** was finished at first because the production of **34** (around 0.3 mg/L bacteria culture) was better than **33** (<0.1 mg/L bacteria culture). Analysis of 1D proton, ^{13}C spectra revealed 10 aromatic carbons (C-1 to C-10), suggesting a naphthalene ring structure. H-6 showed COSY correlations to H-5, H-7 and starting from C-4 all the way to C-8, ^{13}C-^{13}C COSY correlations were observed between the neighboring carbons. HMBC correlation was observed from H-11 to C-8, and the chemical shifts of C-11 and H-11 suggested a –OMe substitution on C-8. H-3 and H-10 were two singlet protons that showed correlations to other important carbons, such as C-12, which is the imine carbon outside the aromatic ring. The chemical shift of C-1 suggested a hydroxyl group substitution and with all

the information above, the aromatic ring structure of **34** was established as a naphthalene structure with three substitutions: one hydroxyl, one methoxyl, and one imine group. The oxazoline ring structure was established by analyzing HMBC correlations from H-15 to C-12, C-14; H-16 to C-14; and ^{15}N-HMBC correlation from H-16 to N-13. Direct C-C bond between C-14 and C-16 by detecting ^{13}C-^{13}C COSY correlation between those two carbons and thus the oxazoline structure was established. Additionally, ^{13}C-^{13}C COSY and ^{13}C-^{15}N HMQC correlations were detected between C-14 and C-17, C-17 and N-18, N-18 and C-20, establishing the direct linkage between those atoms. HMBC and ^{15}N-HMBC correlations from H-19 to C-17 and H-19 to N-18 were observed and combined with the chemical shifts of H-19, C-19, a N-Me group was proposed. COSY correlations were observed between H-20 and H-21, H-20 and H-23, H-23 and H-24, H-24 and H-25, H-25 and H-26, linking those carbons and nitrogens together. Data from ^{13}C-^{13}C COSY and ^{13}C-^{15}N HMQC also suggested those connectivities. HMBC correlation from H-22 to C-21 was observed and the chemical shifts of H-22, C-22 suggested they belong to a heteroatom attached methyl group. Considering a –SOCH$_3$ functional group identified by IFS analysis of MS^2 fragment ions, C-22 was identified as the sulfinyl methyl group. HMBC correlation from H-28 to C-27 and ^{13}C-^{13}C COSY correlation between C-25 and C-27 were observed, suggesting a methyl ester functional group was attached to C-25. With all the information collected so far, the total number of unassigned atoms was 6 carbons, 10 hydrogens, 4 oxygens, and 1 sulfur atom. The sulfur was likely to be attached to C-23 and C-26 to form a tetrahydrothiazole and the sugar functional group identified from the IFS analysis of MS^3 ions can match the leftover atoms. HMBC correlation from H-29, the anomeric proton, to C-26 was observed, suggesting the

sugar unit was attached to C-26. COSY, HMBC analysis of the sugar moiety was performed to connect the last piece of the molecule and with all atoms assigned, the planar structure of ecteinamine B (**34**) was established.

The planar structure of ecteinamine A (**33**) was elucidated similarly since it is only one methoxyl functional group different from its analog **34**. By analyzing the aromatic region, C-8 with no methoxyl group substitution was identified in **33**. Other part of the molecule showed high degree of similarity to **34** in terms of ^{1}H, ^{13}C chemical shifts, COSY and HMBC correlations. Then configuration analysis of both compounds were tried, however, no complete results are available currently. The relative configurations of the sugar moiety and the tetrahydrothiazole moiety were proposed by analyzing the ROESY correlations in ecteinamine B (**34**, Figure S83) as well as coupling constants of some H, H-spin systems in ecteinamine A (**33**, Figure 5.7)[22].

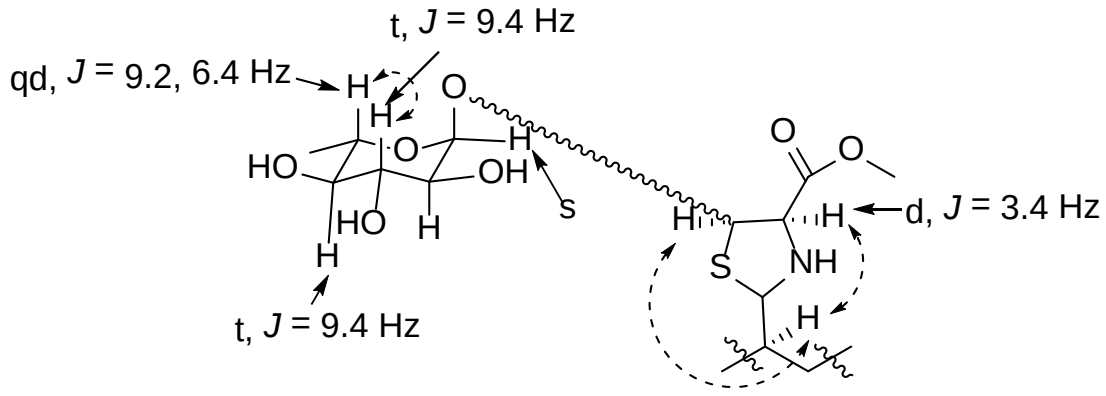


Figure 5.7. ROESY, coupling constant analysis of sugar moiety and tetrahydrothiazole moiety.

The multiplicity of H-32 was doublet of quartets with coupling constants of 9.2 Hz and 6.4 Hz. The $^{3}J_{H\text{-}H}$ coupling constant between H-33 and H-32 was measured to be 6.4 Hz,

suggesting $^{3}J_{H\text{-}H}$ coupling constant between H-32 and H-31 was 9.2 Hz, which was a good evidence of two axial protons in the six-membered ring. ROESY correlation between H-32 and H-30 was observed, suggesting two 1,3-diaxial protons. The coupling constant between H-31 and H-30 was measured to be 9.4 Hz, which also substantiated the axial position of H-30. Coupling constant between H-30 and H-29 was also 9.4 Hz, indicating H-29 was also an axial proton. The small coupling constant between H-28 and H-29 (could not be detected by proton NMR and ROESY) and the lack of ROESY correlation between H-28 and H-32 suggested that H-28 was an equatorial proton and the relative configuration of the sugar moiety was established. For the tetrahydrothiazole moiety, ROESY correlation between H-19 and H-24, H-19 and H-25 was observed, suggesting H-19, H-24 and H-25 were on the same face of the cyclic structure. The small $^{3}J_{H\text{-}H}$ coupling constant between H-24 and H-25 (3.4 Hz) also agreed that proposal. However, for the configuration of the whole molecule, more experiments and analyses will be required.

The influence of acid-base chemistry on ^{15}N chemical shifts of complicated natural products is a relatively understudied field in natural product research. Several reports of both computation NMR studies[23, 24] and model compound NMR studies[25] have revealed an interesting fact that unlike sp^3 type nitrogens, which will have increased ^{15}N chemical shifts (~ 20 ppm) once protonated; sp^2 type unsaturated nitrogens will have significantly decreased ^{15}N chemical shifts (~ 100 ppm) once protonated. This phenomenon has not been well understood by the natural product community and in order to provide good examples for understanding the behavior of ^{15}N chemical shifts of complicated natural products, the ^{15}N chemical shift of ecteinamine B (**34**) was tested under either acidic or basic conditions by

adding either 10% of deuterated trifluoroacetic acid or 20% of potassium deutertoxide to NMR solvent CD_3OD (Figure 5.8, S84-S86). While the ^{15}N chemical shift of two sp^3 nitrogen atoms remained almost no changed, the chemical shift of the double bonded sp^2 nitrogen had significant changes. Starting from the neutral 228 ppm, it shifted upfield to 140 ppm corresponding to the acidic form, and shifted back downfield to 232 ppm, corresponding to the basic form. This observation also met the predicted acid-base behaviors of sp^2 type nitrogens and this example will essentially help us with better understandings of ^{15}N NMR chemical shifts in the future.

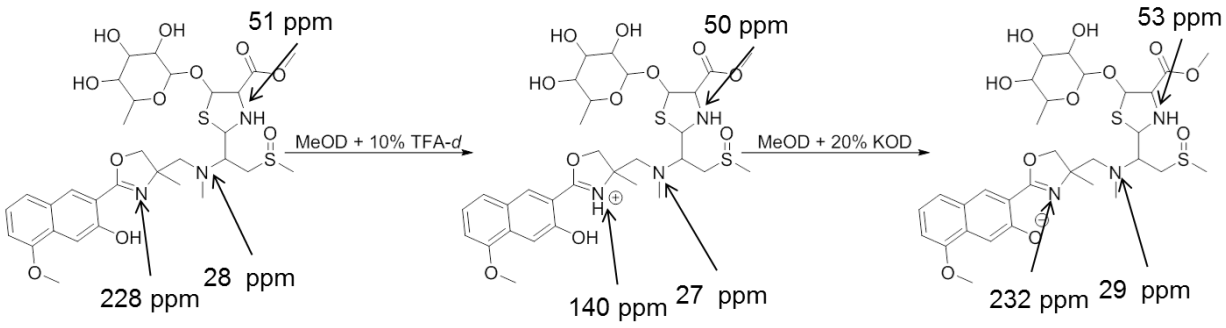


Figure 5.8. Experiment of different ^{15}N chemical shifts of ecteinamine B (**34**) under acidic and basic conditions.

Although ecteinamines A (**33**) and B (**34**) were isolated in a fraction that modulated the level of several cytokines using a macrophage based assay, the biological activities of these two pure compounds have not been studied systematically. The DNA of the producing organism WMMB-482 has been extracted and the whole genome is being sequenced. Hopefully in the near future, genomic information of the producing organism can be analyzed to reveal the possible biosynthetic mechanism of these two novel nonribosomal peptides.

5.3. Conclusion

In summary, we report the discovery and planar structure elucidation of two novel

nonribosomal peptides ecteinamines A and B from a fraction with unusual biological activities. Isotopic Fine Structure (IFS) analysis based unambiguous molecular formula identification method together with analysis of fragment ions in MS^2 and MS^3 mode have made our rapid structure elucidation possible. In addition to common 1D and 2D NMR experiments, ^{13}C-^{13}C COSY and ^{13}C, ^{15}N connectivity experiments were performed using isotopic enriched compounds to reveal direct connections between carbon and nitrogen atoms. Efforts towards configuration analysis have revealed the relative configuration of partial structure. Studies on ^{15}N chemical shifts under different PH conditions will help us understand the behaviors of complicated N-containing natural products better in the future.

5.4. Material and Methods

General Experimental Procedures

Optical rotations were measured on a Perkin–Elmer 241 Polarimeter. UV spectra were recorded on an Aminco/OLIS UV-Vis spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR spectrophotometer. NMR spectra were obtained in CD_3OD (δ_H 3.34 ppm, δ_C 49.0 ppm) and d_6-DMSO (δ_H 2.50 ppm, δ_C 39.50 ppm) with a Bruker Avance 600 III MHz spectrometer equipped with a $^1H\{^{13}C/^{15}N/^{31}P\}$ cryoprobe, a Bruker Avance III 500 MHz spectrometer equipped with a $^{13}C/^{15}N\{^1H\}$ cryoprobe, and a Bruker Avance III HD 400 MHz spectrometer. HRMS data were acquired with a Bruker MaXis™ 4G ESI-QTOF mass spectrometer. RP HPLC was performed using a Shimadzu Prominence HPLC system equipped with a Phenomenex Gemini C18 column (250 × 30 mm) and Phenomenex Luna C18 column (250 × 10 mm). UHPLC-HRMS was acquired using a Bruker MaXis™ 4G ESI-QTOF mass spectrometer coupled with a Waters Acquity UPLC system operated by

Bruker Hystar software and a C18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm). MS data for IFS analysis was collected using a Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry).

Biological Material

Sponge specimens were collected on Aug. 15, 2011 near the Florida Keys (24°41′42.6’’N, 81°26’29.0’’W). A voucher specimen is housed at the University of Wisconsin-Madison. For cultivation, a sample of sponge (1 cm^3) was ground in 500 μL sterile seawater and dilutions were made using 500 μL sterile seawater. Subsequently, 400 μL of diluted sponge sample was added to 200 μL of sterile seawater and 100 μL was plated using a sterile L-shaped spreader. Diluted sample was plated on Gauze 1^{26} media supplemented with artificial seawater. Each medium was supplemented with 50 μg/mL cycloheximide, 25 μg/mL nystatin, and 25 μg/mL nalidixic acid. Plates were incubated at 28˚C and colonies were isolated over the course of two months.

Fermentation, Extraction and Isolation

Two 10 mL seed cultures (25 × 150 mm tubes) in medium DSC (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract per liter of artificial seawater) were inoculated with strain WMMC-911 and shaken (200 RPM, 28 °C) for seven days. Two liter flask (1 × 500 mL) containing ASW-A (20 g soluble starch, 10 g D -glucose, 5 g peptone, 5 g yeast extract, 5 g $CaCO_3$ per liter of artificial seawater. For isotopic enriched media, 10 g D-glucose (U-$^{13}C_6$, 99%) was used instead of D-glucose, 2.5 g $^{15}NH_4Cl$ was used to replace 2.5 g of peptone). Four-liter flasks (6 × 1 L) containing medium ASW-A with Diaion HP20 (7% by weight)

were inoculated with 50 mL from the 500 mL culture and shaken (200 RPM, 28 °C) for seven days. For making artificial sea water, solutions I (415.2 g NaCl, 69.54 g Na_2SO_4, 11.74 g KCl, 3.40 g $NaHCO_3$, 1.7 g KBr, 0.45 g H_3BO_3, 0.054 g NaF) and II (187.9 g $MgCl_2 \cdot 6H_2O$, 22.72 g $CaCl_2 \cdot 2H_2O$, 0.428 g $SrCl_2 \cdot 6H_2O$) were made up separately using distilled water and combined to give a total volume of 20 L.

Filtered HP20 and cells were washed with H_2O and extracted with acetone. The acetone extract was subjected to liquid-liquid partitioning using 30% aqueous MeOH and $CHCl_3$ (1:1). The $CHCl_3$-soluble partition (2.13 g) was fractionated by Sephadex LH20 column chromatography (column size 500 × 40 mm, $CHCl_3$:MeOH=1:1, 20 mL for each fraction). Fractions containing ecteinamines (0.3286 g) were subjected to RP HPLC (20%/80% to 100%/0% MeOH/H_2O (with 0.1% acetic acid), 26.0 min, 20 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm). The fraction collected between 23−25 minutes was further fractionated by RP HPLC (30%/70% to 80%/20% MeOH/H_2O (with 0.1% acetic acid), 30 min, 4.5 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm), yielding **33** (0.5 mg, t_R 27.0 min), **34** (1.3 mg, t_R 29.7 min).

Ecteinamine A (**33**): light yellow amorphous solid, $[\alpha]^{25}_D$ = +16.9 (*c* 0.5, MeOH); UV-Vis (MeOH): λ_{max} (log ε) 205 nm (3.60), 237 nm (3.45), 254 nm (3.32), 287 nm (2.90), 298 nm (2.85), 354 nm (2.48); IR (ATR): υ_{max} 3327, 2945, 2833, 1663, 1651, 1560, 1449, 1410, 1118, 819, 804, 777, 745, 715, 687, 658 cm^{-1}; HRMS $[M+H]^+$ *m/z* = 668.2301 (calcd. for $C_{30}H_{42}N_3O_{10}S_2^+$ 668.2306). ^{1}H NMR (600 MHz, CD_3OD) δ 8.34, (s, 1H), 7.91, (d, *J* = 8.3 Hz, 1H), 7.80, (d, *J* = 7.6 Hz, 1H), 7.54, (t, *J* = 7.6 Hz, 1H), 7.38, (t, *J* = 7.9 Hz, 1H), 7.36, (s,

1H), 4.79, (d, *J* = 11.4 Hz, 1H), 4.65-4.58, (m, 1H), 4.23, (d, *J* = 7.9 Hz, 1H), 3.82-3.78, (m, 1H), 3.77, (d, *J* = 3.4 Hz, 1H), 3.69, (s, 3H), 3.33-3.26, (m, 1H), 3.21-3.15, (m, 1H), 3.20-3.16, (m, 1H), 3.17-3.11, (m, 1H), 3.01, (dd, *J* = 13.2, 2.7 Hz, 1H), 2.89, (t, *J* = 9.4 Hz, 1H), 2.82, (m, 1H), 2.77, (d, *J* = 9.4 Hz, 1H), 2.73, (s, 3H), 2.56, (dq, *J* = 9.4, 6.4 Hz, 1H), 2.40, (d, *J* = 15.2 Hz, 1H), 2.33, (s, 3H), 1.32, (s, 3H), 1.16, (d, *J* = 6.4 Hz, 3H). ^{13}C NMR (125 MHz, CD_3OD:$CDCl_3$ = 1:1 (v:v)) δ168.0, 164.5, 155.0, 136.7, 129.9, 129.0, 128.6, 126.7, 125.8, 124.4, 114.4, 111.6, 103.9, 91.9, 76.8, 76.1, 74.1, 73.8, 72.2, 71.9, 71.7, 71, 62.3, 62.3, 56, 52, 39.3, 37.8, 25.9, 17.5.

Ecteinamine B (**34**): light yellow amorphous solid, $[\alpha]^{25}_{D}$ = +90.0 (*c* 0.25, MeOH); UV-Vis (MeOH): λ_{max} (log ε) 204 nm (3.98), 212 nm (3.97), 241 nm (3.96), 264 nm (4.00), 285 nm (3.37), 297 nm (3.36), 366 nm (2.88); IR (ATR): υ_{max} 3501, 3420, 3390, 3366, 3309, 3288, 3276, 3224, 1653, 1558, 1464, 1416, 1264, 1101, 1016, 838, 734, 716, 699, 691, 682, 671, 659 cm^{-1}; HRMS $[M+H]^+$ *m/z* = 698.2408 (calcd. for $C_{31}H_{44}N_3O_{11}S_2^+$ 698.2412). ^{1}H NMR (600 MHz, CD_3OD) δ^{1}H NMR (600 MHz, MeOD) δ. ^{1}H NMR (600 MHz, MeOD) δ 8.27 (s, 1H), 7.67 (s, 1H), 7.46 (d, *J* = 8.3 Hz, 1H), 7.27 (t, *J* = 7.9 Hz, 1H), 6.94 (d, *J* = 7.6 Hz, 1H), 4.82 (d, *J* = 11.6 Hz, 1H), 4.65 (d, *J* = 7.9 Hz, 1H), 4.22 (d, *J* = 7.9 Hz, 1H), 4.05, (s, 3H), 3.94-3.92 (m, 1H), 3.86 (s, 1H), 3.62 (d, *J* = 3.8 Hz, 1H), 3.59, (s, 3H), 3.57 (dq, *J* = 9.4, 6.4 Hz, 1H), 3.32 – 3.20 (m, 4H), 3.19 – 3.11 (m, 2H), 3.03 (dd, *J* = 13.0, 2.6 Hz, 1H), 2.73 (s, 3H), 2.42 (d, *J* = 14.8 Hz, 1H), 2.34 (s, 3H), 1.32 (s, 4H), 1.16 (d, *J* = 6.4 Hz, 3H). ^{13}C NMR (126 MHz, MeOD) δ 169.6, 165.8, 155.9, 155.4, 130.1, 130.1, 129.5, 124.8, 122.1, 115.3, 107.1, 106.7, 102.7, 91.5, 75.4, 73.7, 72.9, 72.6, 72.0, 71.6, 71.0, 70.6, 63.0, 62.6, 56.3, 56.1, 52.2, 39.8, 38.5, 26.4, 17.6.

NMR experiment under acidic/basic conditions

Ecteinamine B (**34**, 1.2 mg) with ^{15}N labeling was dissolved in 170 μL of CD_3OD. To make the solution acidic, 17 μL of deuterated trifluoroacetic acid was added and the ^{15}N chemical shifts were acquired using ^{15}N-HMBC experiment. To make the solution basic, 34 μL of saturated potassium deutertoxide was added and the ^{15}N chemical shifts were acquired using ^{15}N-HMBC experiment.

Whole Genome Sequencing

DNA extraction of WMMB-482 was conducted as previously described[27]. The complete genome was sequenced at the University of Wisconsin Biotechnology Center (UWBC) using PacBio RSII (Pacific Biosciences) technology.

5.5. References

(1). Newman, D. J.; Cragg, G. M. *J. Nat. Prod.*, **2007**, *70*, 461-477.

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

(3). Lipmann, F.; Gevers, W.; Kleinkauf, H.; Roskoski, R., Jr. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **1971**, *35*, 1-34.

(4). Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.*, **1997**, *97*, 2651-2674.

(5). Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.*, **2003**, *20*, 275-287.

(6). Zasloff, M. *Nature*, **2002**, *415*, 389-395.

(7). Heidary, M.; Khosravi, A. D.; Khoshnood, S.; Nasiri, M. J.; Soleimani, S.; Goudarzi, M. *J. Antimicrob. Chemother.*, **2018**, *73*, 1-11.

(8). Yang, H.; Chen, K. H.; Nowick, J. S. *ACS Chem. Biol.*, **2016**, *11*, 1823-1826.

(9). Lewis, K.; *Nature*, **2012**, *485*, 439-440.

(10). Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. *Nat. Rev. Drug Discov.*, **2007**, *6*, 29-40.

(11). Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S. *Anal. Chem.*, **2012**, *84*, 4277-4283.

(12). Zhang, F.; Barns, K.; Hoffmann, F. M.; Braun, D. R.; Andes, D. R.; Bugni, T. S. *J. Nat. Prod.* **2017**, *80*, 2551−2555.

(13). Chandra, R. K. *Military Strategies for Sustainment of Nutrition and Immune Function in the Field*. National Academies Press: Washington D. C. 1999, pp 205-220.

(14). Host Defenses, Levinson W. *Review of Medical Microbiology and Immunology, 14e;* 2016.

(15). Rossio, J. L. *Military Strategies for Sustainment of Nutrition and Immune Function in the Field*. National Academies Press: Washington D. C. 1999, pp 221-234.

(16). Tanaka, T.; Narazaki, M.; Kishimoto, T. *Cold Spring Harb. Perspect. Biol.*, **2014**, *6*:a016295.

(17). Rose, W. E.; Eickhoff, J. C.; Shukla, S. K.; Pantrangi, M.; Rooijakkers, S.; Cosgrove, S. E.; Nizet, V.; Sakoulas, G. *J. Infect. Dis.*, **2012**, *206*, 1604-1611.

(18). Rose, W. E.; Shukla, S. K.; Berti, A. D.; Hayney, M. S.; Henriquez, K. M.; Ranzoni, A.; Cooper, M. A.; Proctor, R. A.; Nizet, V.; Sakoulas, G. *Clin. Infect. Dis.*, **2017**, *64*, 1406-1412.

(19). Minejima, E.; Bensman, J.; She, R. C.; Mack, W. J.; Tran, M. T.; Ny, P.; Lou, M.; Yamaki, J.; Nieberg, P.; Ho, J.; Wong-Beringer, A. *Crit. Care Med.*, **2016**, *44*, 671-679.

(20). Swanson, R. N.; Hardy, D. J.; Shipkowitz, N. L.; Hanson, C. W.; Ramer, N. R.; Coen, L. J.; Fernandes, P. B. *J. Antibiot.*, **1989**, *42*, 94-101.

(21). Fukuda, T.; Miller, E. D.; Clark, B. R.; Alnauman, A.; Murphy, C. D.; Jensen, P. R.; Fenical, W. *J. Nat. Prod.*, **2011**, *74*, 1773-1778.

(22). Matsumori, N.; Kaneno, D.; Murata, M.; Nakamura, H.; Tachibana, K. *J. Org. Chem.*, **1999**, *64*, 866-876.

(23). Gil, V. M. S.; Murrell, J. N. *Trans. Faraday Soc.*, **1964**, *60*, 248-255.

(24). Solum, M. S.; Altmann, K. L.; Strohmeier, M.; Berges, D. A.; Zhang, Y.; Facelli, J. C.; Pugmire, R. J.; Grant, D. M. *J. Am. Chem. Soc.*, **1997**, *119*, 9804-9809.

(25). Lambert, J. B.; Binsch, G.; Roberts, J. D. *Proc. Natl. Acad. Sci. U. S. A.*, **1964**, *51*, 735-737.

(26). Gauze, G. F.; Preobrazhenskaya, J. P.; Kudrina, E. E.; Blinov, N. O.; Ryabova, I. D.; Sveshnikova, M. A. **1957**. Problems in the classification of antagonistic actinomycetes. State Publishing House for Medical Literature. Medgiz, Moscow.

(27). Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S. *ACS Chem. Biol.* **2017**, *12*, 3093−3102.

Chapter 6:

Discovery and structure elucidation of a new nonribosomal cyclic peptide, streptnatamide A.

6.1. Introduction

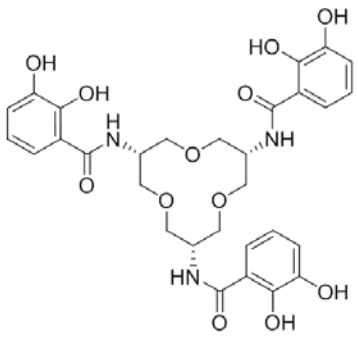
In this chapter, the discovery and structure elucidation of a nonribosomal cyclic peptide streptnatamide A will be described. Isotopic fine structure (IFS) analysis unambiguous molecular formula determination method and detailed 2D NMR analysis were applied in the rapid unambiguous molecular formula identification and structure elucidation of streptnatamide A.

Nonribosomal peptides represent a major class of natural products with diverse chemical structures as well as a wide range of biological activities[1]. As stated in Chapter 5, nonribosomal peptides usually feature heterocyclic structures, D-amino acids, glycosylated structures, N-methylations, and a large variety of structurally modification amino acid residues that are different from the 20 common amino acids used in the ribosomal peptide biosynthesis[2]. Among these modifications, N-methylations and structurally modified amino acids are two significant factors that are leading to different nonribosomal peptides with high level of structural diversities and biological activities. It has been reported that higher levels of N-methylation of the nonribosomal peptides will strengthen up the peptide by preventing them from cellular premature proteolytic breakdown[3]. Also, N-methylations will potentially change the conformation of the peptide, leading to different biological activities and chemical

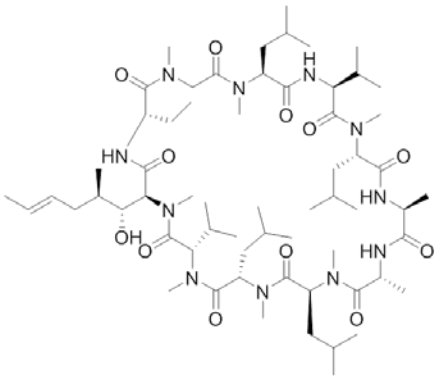
stabilities[4]. Similarly, structurally modified amino acids that are different from the 20 common ones will lead to increased chemical stability and different biological activities. More importantly, some modified amino acids with hydroxyl groups, carboxylic acid groups, and amino groups can undergo cyclization reactions with either the N-terminus or the C-terminus of the premature nonribosomal peptide, leading to macrocyclic structures[5]. These macrocyclic peptides can be divided into two major groups: macrolactams, which are the results of amide bond formations; or macrolactones, which are results of ester bond formations. Because the cyclic structures are extremely different from the linear chain structures, the macrocyclic peptides (or cyclic peptides) usually have diverse biological activities. Some of the very famous cyclic peptides includes cyclosporin A (**35**)[6], the immunosuppressive macrolactam; enterobactin A (**36**)[7], the macrolactone siderophore; and daptomycin (**29**)[8], the antibacterial lactone.

As described in Chapter 5, our lab has applied fractions from our natural product library in a mouse macrophage based assay to test their abilities to modulate the cytokine productions. Streptnatamide A (**37**) was discovered from a active fraction produced by *Streptomyces* sp. WMMB-303 and the rapid structure elucidation streptnatamide A was made possible by a combination of Isotopic Fine Structure analysis and detailed 1D, 2D NMR analysis. Solvent suppression NMR experiments[9] were performed in order to confirm some key ROESY correlations that eventually helped the establishment of the proposed macrolactams structure. Out of 11 amino acid residues identified in streptnatamide A, six of them were N-methylated and one was identified as an uncommon amino acid 3-(methylamino)alanine. The discovery and structural elucidation of streptnatamide A

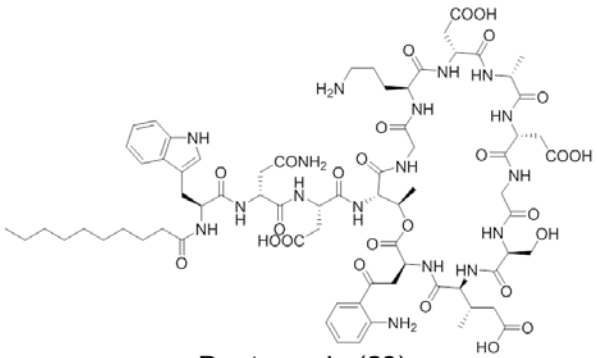
highlights the strength of applying modern analytical techniques to speed up the structural elucidation of complicated natural products.



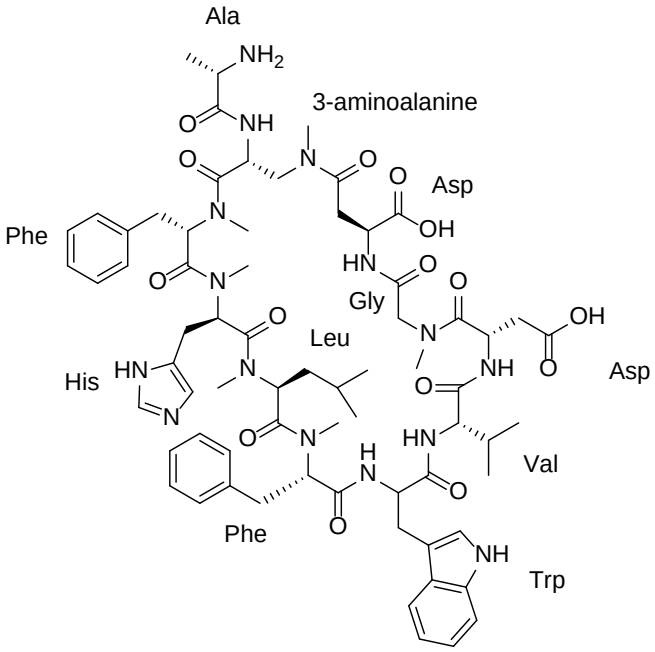
Enterobactin A (**35**)



Cyclosporin A (**36**)



Daptomycin (**29**)



Streptnatamide A (**37**)

6.2. Results and Discussion

Molecular formula of streptnatamide A (**37**) was determined unambiguously by analyzing the Isotopic Fine Structure of **37**. Direct infusion MS data of **37** was collected under ESI (electrospray ionization) positive mode on Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry) and displayed an intense $[M+2H]^{2+}$ adduct peak with *m/z* ratio of 679.8482 (Figure S106). Isotopologues spectra of **37** ($[M+2H+1]^{2+}$, $[M+2H+2]^{2+}$ and $[M+2H+3]^{2+}$ ions) were also analyzed (Figure 6.1) to reveal the isotopic fine structures formed by different isotopic compositions. The molecular formula of **37** was determined as $C_{68}H_{91}N_{15}O_{15}$ as the simulated IFS of this formula matched the acquired MS spectra of all isotopologues analyzed. The monoisotopic peak showed only +0.07 ppm error compared to the predicted molecular formula, which highlighted the accuracy of the MRMS instrument when calibrated properly. Despite having minor disagreements among some peak intensities, for $[M+2H+2]^{2+}$ and $[M+2H+3]^{2+}$ isotopologue ions, the ratio between isotope peaks containing ^{15}N and ^{18}O suggested the number of N, O atoms in **37** are very similar ($\leq$3 atoms difference), which narrowed down the choices of possible molecular formulae within 1 ppm error.

With the molecular formula **37** determined unambiguously, rapid structure elucidations of the planar structure was completed by analyzing the 1D and 2D NMR spectra in details (Figure S96-S105). The ^{1}H and ^{13}C chemical shifts, COSY correlations, some key HMBC correlations, and some key ROESY correlations of **37** are summarized in Table 6.1. Other 2D NMR experiments such as HSQC, TOCSY were also performed in order to study the structure of **37**.

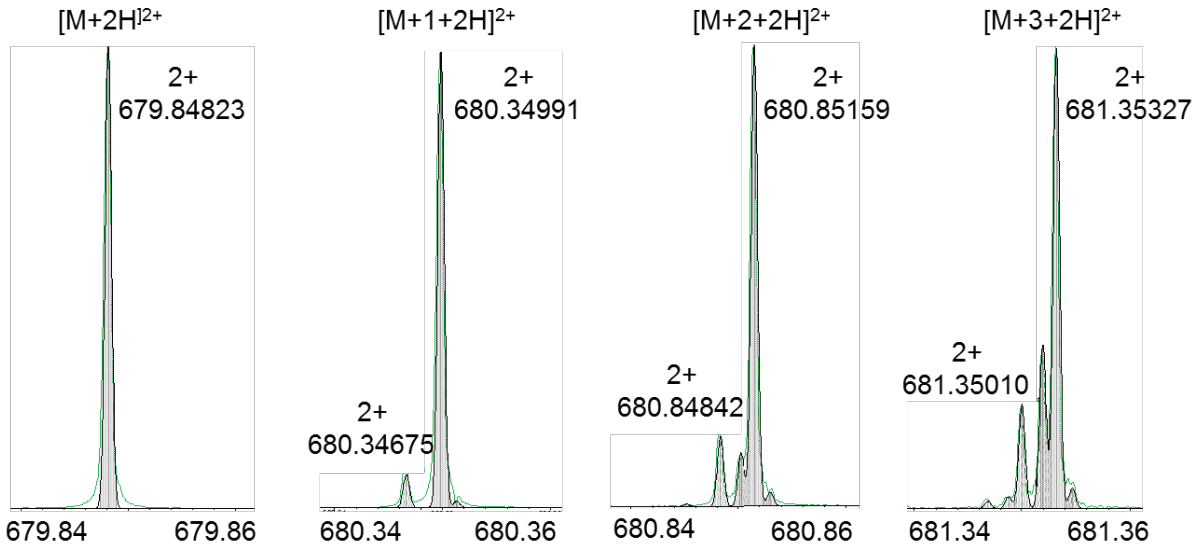
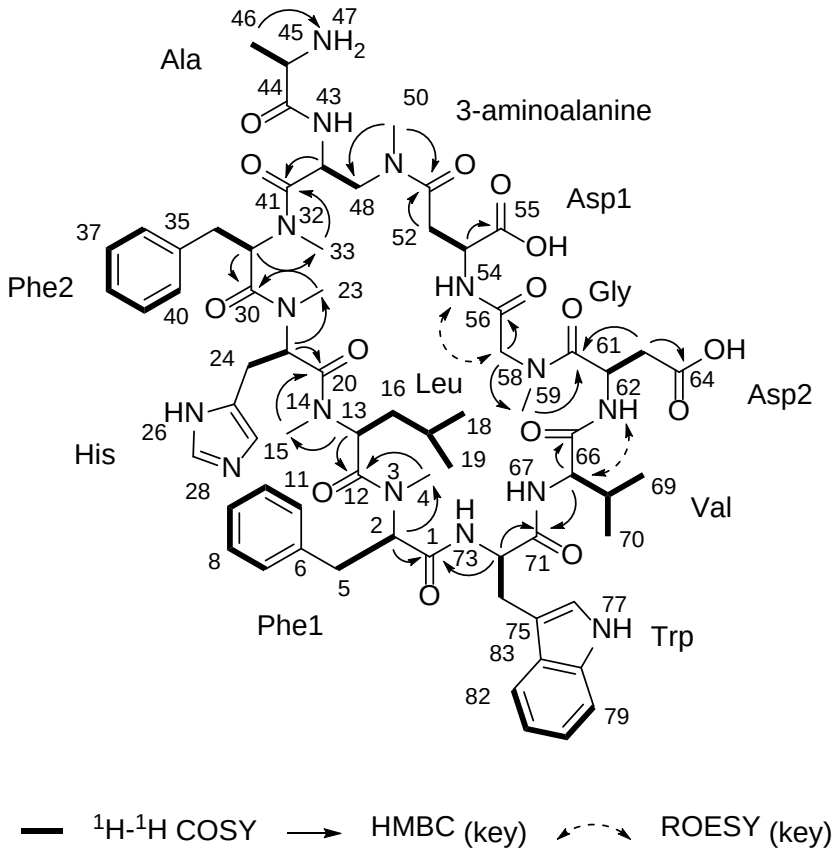


Figure 6.1. Overlaid MS spectra of the experimental (green trace), simulated (grey) IFS of $(C_{68}H_{93}N_{15}O_{15})^{2+}$ ($[M+2H]^{2+}$) and its isotopologues acquired from the Bruker MRMS instrument.

Table 6.1. ^{1}H and ^{13}C NMR data for **37** (600 MHz for ^{1}H, 125 MHz for ^{13}C, MeOD)



Number	Amino Acid	δ_C, type	δ_H, (J in Hz)
1	Phe1	170.2	
2	Phe1	62.8, CH	5.42, dd (9.4, 4.4)
3	Phe1		
4	Phe1	30.0, CH_3	3.14, s
5	Phe1	36.3, CH_2	3.24, m 2.99, dd (14.2, 4.0)
6	Phe1	138.8	
7	Phe1	130.6, CH	7.20, d (7.8)
8	Phe1	130.3, CH	7.32, t (7.2)
9	Phe1	128.1, CH	7.26, t (7.2)
10	Phe1	130.3, CH	7.32, t (7.2)
11	Phe1	130.6, CH	7.20, d (7.8)
12	Leu	172.9	
13	Leu	50.5, CH	4.79, s
14	Leu		
15	Leu	30.9, CH_3	2.45, s
16	Leu	36.9, CH_2	1.41, m -0.36, dt (10.8, 4.0)
17	Leu	25.2, CH	0.61, m
18	Leu	20.6, CH_3	0.40, d (5.4)
19	Leu	23.8, CH_3	0.60, m
20	His	171.7	
21	His	52.2, CH	4.98, d (8.8)
22	His		
23	His	30.3, CH_3	2.45, s
24	His	25.3, CH_2	2.80, m 0.66, d (8.0)
25	His	130.2	
26	His		
27	His	136.9, CH	8.46, s
28	His		
29	His	120.2, CH	7.13, s
30	Phe2	169.9	
31	Phe2	57.0, CH	
32	Phe2		

33	Phe2	30.1, CH_3	2.43, s
34	Phe2	35.5, CH_2	3.23, m 2.56, dd (13.0, 4.8)
35	Phe2	138.6	
36	Phe2	130.8, CH	7.19, d (7.8)
37	Phe2	129.3, CH	7.23, t (8.2)
38	Phe2	127.6, CH	7.14, t (8.8)
39	Phe2	129.3, CH	7.23, t (8.2)
40	Phe2	130.8, CH	7.19, d (7.8)
41	3-aminoalanine	171.5	
42	3-aminoalanine	49.0, CH	5.34, d (9.6)
43	3-aminoalanine		
44	Ala	171.2	
45	Ala	50.2, CH	3.92, q (6.7)
46	Ala	17.6, CH_3	1.44, d (6.7)
47	Ala	36.2[a], NH_2	
48	3-aminoalanine	53.4, CH_2	3.64, dd (14.6, 4.4) 2.66, m
49	3-aminoalanine		
50	3-aminoalanine	40.4, CH_3	3.43, s
51	Asp1	175.0	
52	Asp1	40.3, CH_2	2.73, m 2.55, m
53	Asp1	48.9, CH	4.84, dd(9.6, 4.8)
54	Asp1		8.66[b], s
55	Asp1	177.6	
56	Gly	172.7	
57	Gly	53.1, CH_2	4.45, d (16.2) 3.59, d (16.0)
58	Gly		
59	Gly	38.1, CH_3	3.23, s
60	Asp2	172.3	
61	Asp2	48.9, CH	5.06, dd (18.1, 9.0)
62	Asp2		8.30[b], s
63	Asp2	39.6, CH_2	2.72, m 2.27, dd (18.4, 8.6)
64	Asp2	178.2	

65	Val	173.2	
66	Val	60.7, CH	4.14, d (10.8)
67	Val		
68	Val	31.4, CH	2.15, septet (6.8)
69	Val	19.6, CH_3	1.04, d (7.1)
70	Val	20.1, CH_3	1.03, d (7.1)
71	Trp	171.4	
72	Trp	55.9, CH	4.94, d (12.8)
73	Trp		
74	Trp	29.4, CH_2	3.39, dd (12.4, 3.8) 3.20, dd (13.4, 4.2)
75	Trp	110.3	
76	Trp	125.5, CH	6.75, s
77	Trp		10.64[b], s
78	Trp	137.5	
79	Trp	112.7, CH	7.39, d (8.0)
80	Trp	122.0, CH	7.20, m
81	Trp	119.8, CH	6.88, t (7.5)
82	Trp	119.8, CH	7.13, d (8.4)
83	Trp	129.3	

[a.] ^{15}N chemical shift acquired from ^{15}N-HMBC experiment. [b.] Acquired in CD_3OH

In total, 11 α-carbons of amino acids were identified in **37**, suggesting the total number of amino acid residues that formed **37** was also 11. By analyzing the COSY, TOCSY, HSQC and HMBC correlations of all α-C-H protons, the amino acid residues were identified as following: 2 phenylalanines (C-2 and, C-31, number of α-carbon(s) listed in parentheses), 2 aspartic acids (C-53, C-61), 1 tryptophan (C-72), 1 valine (C-66), 1 leucine (C-13), 1 glycine (C-57), 1 histidine (C-21), 1 alanine (C-45), and one modified uncommon 3-aminoalanine (amino-β-alanine, C-42). Once identified the amino acid residues, the total unsaturation degree of all residues were calculated to be 30 and this number was compared to 31, which was the unsaturation degree of streptnatamide A. We proposed the macrocyclic formation

between some amino acid residues was leading to the 1 unsaturation degree difference between those two values, thus connecting all 11 amino acid residues became our next target.

By analyzing the HMBC correlations in details, especially the six N-Me groups (attached to 2 phenylalanines, 1 histidine, 1 glycine, 1 leucine, and 1 3-aminoalanine respectively) identified by HSQC, dept-90, and dept-135; three major pieces of the whole molecule was identified: one big piece with 8 amino acid residues (N-Asp-3-aminoalanine-Phe-His-Leu-Phe-Trp-Val-C); one piece with two amino acid residues (N-Asp-Gly-C); and one alanine residue (NH_2-Ala-C). The free amino group attached to the alanine residue was identified by comparing the ^{15}N chemical shift of N-47 (36.2 ppm, suggesting a terminal -NH_2 group[10]) to the amide ^{15}N chemical shifts (mostly around 120 ppm). The major HMBC correlations that led to the abovementioned connections were all described in Table 6.1. ROESY correlations between some amide N-H protons and α-C-H protons were essential to connect those three pieces together and complete the planar structure elucidation; however, developing a solvent system that can display the exchangeable protons was challenging for us initially. The solubility issues and molecular motion issues were leading to poor results in common aprotic solvents such as $CDCl_3$, d_6-DMSO, d_6-acetone, d_7-DMF. Since we had good results from the CD_3OD, we tried solvent suppression experiments with CD_3OH NMR solvent[9] and got good results. ROESY correlations between H-54 (NH, Asp, 8.66 ppm), H-57 (α-C-H, Gly, 4.45 ppm), and H-62 (NH, Asp, 8.30 ppm), H-66 (α-C-H, Val, 4.14 ppm) were observed, connecting the two pieces together and formed a cyclic structure with 10 amino acids. The only alanine residue left was proposed to be attached to the modified 3-aminoalanine residue by an amide bond formation.

Therefore, the planar structure of streptnatamide A (**37**) was elucidated. The proposed structure was also supported by MS^2 and MS^3 analysis of **37** (Figure 6.2, S107).

Advanced Marfey's analysis[11, 12] was used to determine the absolute configurations of all amino acid residues (Figure S108). Acid hydrolysis of **37** using 6M HCl was performed to release the free amino acids, which were coupled with either L-FDLA (Marfey's reagent) or DL-FDLA (racemic mixture of Marfey's reagent). The resulting mixtures were separated and analyzed by UHPLC-HRMS to check the eluting order of different diastereoisomers corresponding to different amino acids. The configurations of amino acids were identified as L-Phe, L-Asp, L-Trp, L-Ala, L-Leu, L-Val, L-His, and L-3-aminoalanine. With all configurations of amino acid residues determined, the absolute configuration of **37** was also determined.

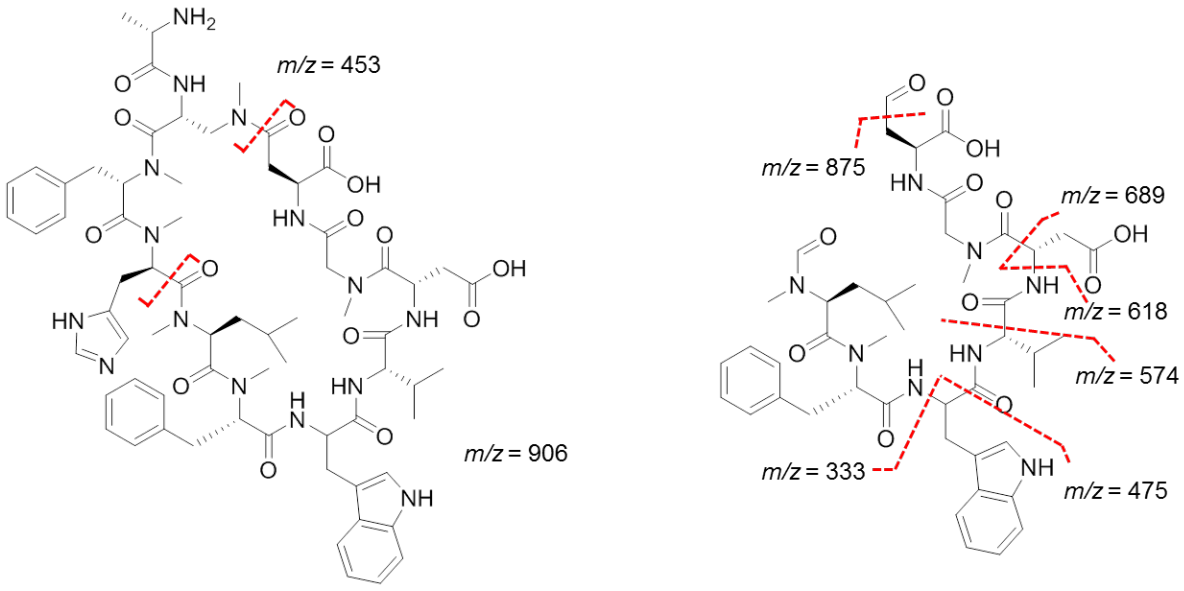


Figure 6.2. Major MS^2 fragment ions (left), and MS^3 ions from *m/z* = 906 (right) of streptnatamide A (**37**).

Although streptnatamide A was isolated in a fraction that modulated the level of several

cytokines using a macrophage based assay, the biological activities of pure streptnatamide A have not been studied systematically. Hopefully in the near future, detailed and systematic studies on its biological activities will unveil its potential as future drug leads.

6.3. Conclusion

In summary, we report the discovery and planar structure elucidation of one new nonribosomal cyclic peptide streptnatamide A from a fraction with unusual biological activities. Isotopic Fine Structure (IFS) analysis based unambiguous molecular formula identification method combined with detailed analysis of 1D and 2D NMR spectra led to the rapid structure elucidation of streptnatamide A. Detection of some ROESY correlations between exchangeable N-H protons and α-C-H protons in CD_3OH NMR solvent using solvent suppression was the key towards the assembly of the macrolactam structure, which was further substantiated by MS^2 and MS^3 analysis. Advanced Marfey's analysis was performed to determine the configurations of amino acid residues, leading to the determination of absolute configuration of streptnatamide A. This study also highlights the importance of combining modern MS and NMR techniques together in order to complete the structure elucidation of challenging compounds rapidly.

6.4. Material and Methods

General Experimental Procedures

Optical rotations were measured on a Perkin–Elmer 241 Polarimeter. UV spectra were recorded on an Aminco/OLIS UV-Vis spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR spectrophotometer. NMR spectra were obtained in CD_3OD (δ_H 3.34 ppm, δ_C 49.0 ppm) and CD_3OH (δ_H 3.34 ppm, δ_C 49.0 ppm) with a Bruker Avance 600

III MHz spectrometer equipped with a $^{1}H\{^{13}C/^{15}N/^{31}P\}$ cryoprobe, a Bruker Avance III 500 MHz spectrometer equipped with a $^{13}C/^{15}N\{^{1}H\}$ cryoprobe, and a Bruker Avance III HD 400 MHz spectrometer. HRMS data were acquired with a Bruker MaXis™ 4G ESI-QTOF mass spectrometer. RP HPLC was performed using a Shimadzu Prominence HPLC system equipped with a Phenomenex Gemini C18 column (250 × 30 mm) and Phenomenex Luna C18 column (250 × 10 mm). UHPLC-HRMS was acquired using a Bruker MaXis™ 4G ESI-QTOF mass spectrometer coupled with a Waters Acquity UPLC system operated by Bruker Hystar software and a C18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm). MS data for IFS analysis was collected using a Bruker 12T SolariX XR^{TM} MRMS (Magnetic Resonance Mass Spectrometry).

Biological Material

Sponge specimens were collected on Nov. 3, 2010 near the Florida Keys (24°47′58.6’’N, 81°28’21.2’’W). A voucher specimen is housed at the University of Wisconsin-Madison. For cultivation, a sample of sponge (1 cm^3) was ground in 500 μL sterile seawater and dilutions were made using 500 μL sterile seawater. Subsequently, 400 μL of diluted sponge sample was added to 200 μL of sterile seawater and 100 μL was plated using a sterile L-shaped spreader. Diluted sample was plated on Gauze 1^{13} media supplemented with artificial seawater. Each medium was supplemented with 50 μg/mL cycloheximide, 25 μg/mL nystatin, and 25 μg/mL nalidixic acid. Plates were incubated at 28˚C and colonies were isolated over the course of two months.

Sequencing

16S rDNA sequencing was conducted as previously described[14]. WMMB-303 was identified as a *Streptomyces* sp.

Fermentation, Extraction and Isolation

Two 10 mL seed cultures (25 × 150 mm tubes) in medium DSC (20 g soluble starch, 10 g glucose, 5 g peptone, 5 g yeast extract per liter of artificial seawater) were inoculated with strain WMMC-911 and shaken (200 RPM, 28 °C) for seven days. Two liter flask (1 × 500 mL) containing ASW-A (20 g soluble starch, 10 g D-glucose, 5 g peptone, 5 g yeast extract, 5 g $CaCO_3$ per liter of artificial seawater. Four-liter flasks (3 × 1 L) containing medium ASW-A with Diaion HP20 (7% by weight) were inoculated with 50 mL from the 500 mL culture and shaken (200 RPM, 28 °C) for seven days. For making artificial sea water, solutions I (415.2 g NaCl, 69.54 g Na_2SO_4, 11.74 g KCl, 3.40 g $NaHCO_3$, 1.7 g KBr, 0.45 g H_3BO_3, 0.054 g NaF) and II (187.9 g $MgCl_2{\cdot}6H_2O$, 22.72 g $CaCl_2{\cdot}2H_2O$, 0.428 g $SrCl_2{\cdot}6H_2O$) were made up separately using distilled water and combined to give a total volume of 20 L.

Filtered HP20 and cells were washed with H_2O and extracted with acetone. The acetone extract was subjected to liquid-liquid partitioning using 30% aqueous MeOH and $CHCl_3$ (1:1). The $CHCl_3$-soluble partition (1.35 g) was fractionated by Sephadex LH20 column chromatography (column size 500 × 40 mm, $CHCl_3$:MeOH=1:1, 20 mL for each fraction). Fractions containing streptnatamide A (0.2542 g) were subjected to RP HPLC (20%/80% to 100%/0% MeOH/H_2O (with 0.1% acetic acid), 26.0 min, 20 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm). The fraction collected between 25.0-25.5 minutes was

further fractionated by RP HPLC (15%/85% to 45%/55% MeCN/H_2O (with 0.1% acetic acid), 30 min, 4.5 mL/min) using a Phenomenex Gemini C18 column (250 × 30 mm), yielding **35** (10.2 mg, t_R 25.6 min).

Streptnatamide A (**37**): white amorphous solid, $[\alpha]^{25}_D$ = -108.1 (*c* 3.2, MeOH); UV-Vis (MeOH): λ_{max} (log ε) 209 nm (4.15), 282 nm (3.05), 291 nm (2.97); IR (ATR): υ_{max} 3409, 3385, 3264, 3090, 3047, 3004, 2976, 2957, 2894, 2879, 2861, 2842, 2825, 1670, 1633, 1561, 1524, 1496, 1458, 1416, 1396, 1097, 1078, 1016, 900, 878, 826, 804, 775, 743, 743, 732, 721, 713, 689, 676 cm^{-1}; HRMS $[M+2H]^{2+}$ *m/z* = 679.8483 (calcd. for $C_{68}H_{93}N_{15}O_{15}^{2+}$ 679.8482) .δ^1H NMR (600 MHz, MeOD) δ 8.46 (s, 1H), 7.39 (d, *J* = 8.0 Hz, 1H), 7.35 – 7.30 (t, *J* = 7.2 Hz, 2H), 7.26 (dd, *J* = 13.1, 5.9 Hz, 1H), 7.23 – 7.17 (m, 7H), 7.17 – 7.10 (m, 3H), 6.88 (t, *J* = 7.5 Hz, 1H), 6.75 (s, 1H), 5.41 (dd, *J* = 11.2, 3.0 Hz, 1H), 5.34 (dd, *J* = 9.4, 4.4 Hz, 2H), 5.06 (m, 1H), 4.85-4.78 (m, 2H), 4.45 (d, *J* = 16.2 Hz, 1H), 4.14 (d, *J* = 10.8 Hz, 1H), 3.92 (q, *J* = 6.7 Hz, 1H), 3.65-3.55 (m, 2H), 3.43 (s, 3H), 3.40 – 3.35 (m, 4H), 3.27 – 3.17 (m, 8H), 3.14 (s, 3H), 3.04 – 2.96 (m, 2H), 2.86 – 2.49 (m, 8H), 2.45 (d, *J* = 5.4 Hz, 9H), 2.34 – 2.08 (m, 2H), 1.48 – 1.37 (m, 4H), 1.04 (dd, *J* = 11.9, 6.7 Hz, 6H), 0.76 – 0.50 (m, 6H), 0.40 (d, *J* = 5.4 Hz, 3H), -0.33 – -0.43 (m, 1H).δ ^{13}C NMR (126 MHz, MeOD) δ 178.2, 177.6, 175.0, 173.2, 172.9, 172.7, 172.3, 171.7, 171.5, 171.4, 171.2, 170.2, 169.9, 138.8, 138.6, 137.5, 136.9, 130.8, 130.6, 130.3, 129.3, 129.2, 128.1, 127.6, 125.4, 122.0, 119.8, 119.8, 112.7, 110.2, 62.8, 62.2, 60.7, 57.0, 55.9, 53.4, 53.1, 52.2, 50.5, 50.2, 40.3, 40.3, 39.6, 38.1, 36.9, 36.3, 35.5, 31.4, 30.9, 30.2, 30.1, 30.0, 29.3, 25.3, 23.8, 20.6, 20.1, 19.5, 17.6.

Advanced Marfey's Analysis

L-FDLA and racemic mixture of DL-FDLA were synthesized following the experimental

procedures in the original report[11]. Separation of different diastereoisomers was carried out and detected on UHPLC-HRMS system using a linear gradient of 0.3 mL/mL of 10%-100% MeOH/H_2O (0.1 % formic acid) over 13 minutes and a hold at 100% MeOH for 3 minutes.

One milligram of **37** was dissolved in 800 μL of 6M HCl and heated to 110 °C for 4 hours (for determination of Trp configuration, 1% phenol was added[15]). The solution was evaporated to dryness and the sample dissolved in 400 μL of 1 M $NaHCO_3$. The solution was divided into two parts evenly and to each part was added 170 μL of either 1% L-FDLA or 1% DL-FDLA in acetone and the mixture stirred for 1 hour at 40 °C. Each reaction was quenched by the addition of 60 μL of 2 M aqueous HCl. The sample was diluted with LC-MS grade MeOH and analyzed on UHPLC-HRMS.

6.5. References

(1). Newman, D. J.; Cragg, G. M. *J. Nat. Prod.*, **2007**, *70*, 461-477.

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

(3). Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.*, **2003**, *20*, 275-287.

(4). Weber, G.; Schörgendorfer, K.; Schneider-Scherzer, E.; Leitner, E. *Curr. Genet.*, **1994**, *26*, 120-125.

(5). de Mattos-Shipley, K. M. J.; Greco, C.; Heard, D.M.; Hough, G.; Mulholland, N. P.; Vincent, J. L.; Micklefield, J.; Simpson, T. J.; Willis, C.L.; Cox, R.J.; Bailey, A. M. *Chem. Sci.*, **2018**, *9*, 4109-4117.

(6). Faulds, D.; Goa, K. L.; Benfield, P. *Drugs*, **1993**, *45*, 953-1040.

(7). Gehring, A. M.; Mori, I.; Walsh, C. T. *Biochemistry*, **1998**, *37*, 2648-2659.

(8). Heidary, M.; Khosravi, A. D.; Khoshnood, S.; Nasiri, M. J.; Soleimani, S.; Goudarzi, M. *J. Antimicrob. Chemother.*, **2018**, *73*, 1-11.

(9). You, W.; Padgett, E.; MacMillan, S. N.; Muller, D. A.; Coates, G. W. *Proc. Natl. Acad.*

Sci. U. S. A., **2019**, *116*, 9729-9734.

(10). Platzer, G.; Okon, M.; McIntosh, L. P. *J. Biomol. NMR*, **2014**, *60*, 109-129.

(11). Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591−596.

(12). Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Anal. Chem.*, **1997**, *69*, 5146-5151.

(13). Gauze, G. F.; Preobrazhenskaya, J. P.; Kudrina, E. E.; Blinov, N. O.; Ryabova, I. D.; Sveshnikova, M. A. **1957**. Problems in the classification of antagonistic actinomycetes. State Publishing House for Medical Literature. Medgiz, Moscow.

(14). Wyche, T. P.; Hou, Y.; Braun, D.; Cohen, H. C.; Xiong, M. P.; Bugni, T. S. *J. Org. Chem.* **2011**, *76*, 6542−6547.

(15). Muramoto, K.; Sunahara, S.; Kamiya, H. *Agric. Biol. Chem.*, **1987**, *51*, 1607-1616.

Chapter 7:

Concluding Remarks and Future Directions

7.1. Concluding remarks

Historically natural products have been one of the most reliable sources of antibiotics with a broad spectrum of antimicrobial activities[1]. However, the pharmaceutical industry had experienced high known compound rediscovery rate[2] when trying to discover new antibiotics from natural products, their traditional sources. Meanwhile, the demand for new antibiotic compounds has never stopped and it is becoming even more urgent as the emergence of multi-drug resistant (MDR) strains is becoming more and more frequent globally[3]. Advances in analytical chemistry, metagenomics, metabolomics, and data science have all contributed to the efficient discovery of new compounds from natural products, but more efforts towards developing and streamlining new dereplication methods are still necessary. Also, completing the structure elucidation of newly discovered compounds rapidly has always been a challenge to chemists. With the recent advancements in analytical instruments and genomic analyses based biosynthetic mechanism predictions, new tools for rapid structure elucidation can be developed for chemists to solve the structures more efficiently.

LCMS based metabolomics profiling and analysis have emerged as a powerful tool for discovering new compounds from a big batch of biological samples. Principal Component Analysis (PCA) of big data set of secondary metabolites has been developed and optimized by our lab aiming to solve one of the fundamental problems in natural product research –

discovering new compounds in the abundance of knowns. This PCA method has helped me to discover three classes of new antibacterial compounds rapidly: madurastatins (Chapter 2), bacillimidazoles (Chapter 3), and haliclonamycin (Chapter 4). The producing organisms of these three classes of compounds were prioritized after PCA analysis of the secondary metabolites of ~120 bacteria strains, which indicated the strength and efficiency of our PCA analysis compared to the traditional Waksman's platform based antibiotics discovery.

A streamlined natural product library based antibiotics discovery platform has also been established and currently, new data are being added to the library every day. PCA based strain selection has guaranteed the high chemical diversity of bacteria strains selected for library generation, which will eventually maximize the chance for discovering new antibiotics. Also, the advanced analytical instruments, such as 1.7 mm cryoprobe NMR, UHPLC/HRMS, have allowed a relative small fermentation scale, leading to a rather high throughput natural product library generation and antimicrobial screening. This library is also an interactive library because more and more information about the LC/MS retention time, *m/z* ratio, NMR patterns, and antimicrobial activities are being added to the library by all of the scientists following the same protocols. I have discovered two classes of antibacterial compounds bacillimidazoles (Chapter 3) and haliclonamycin (Chapter 4) using this natural product library platform so far. The biological activity and spectroscopic data acquired from the initial active hit wells also helped me target the bioactive compounds in larger scale fermentations.

Isotopic Fine Structure (IFS) analysis based unambiguous molecular formula determination method has been developed in order to solve one of the major challenges in natural product research – determining the molecular formula rapidly. By using a high

resolution Bruker 12T SolariX XR^{TM} MRMS instrument, IFS of natural products with different isotopic compositions can be resolved clearly, leading to unambiguous molecular formula determination. Also, by analyzing the IFS of some MS^2, MS^3 fragment ions, information about the functional groups can also be collected, which will further assist the structure elucidation of target compounds. By using this IFS analysis method, I have completed the structure elucidations of two classes of new nonribosomal peptides, ecteinamines (Chapter 5) and streptnatamide A (Chapter 6) rapidly. Ecteinamines A and B feature multiple uncommon modified amino acid residues with oxidations, C-methylation, glycosylation, and N-methylation; streptnatamide A features a macrolactam structure with high level of N-methylations and an uncommon amino acid 3-aminoalanine. Some advanced NMR techniques, such as ^{13}C-^{13}C COSY, ^{13}C-^{15}HMQC experiments on isotopic enriched substrates, solvent suppression experiments for exchangeable proton detections have been applied to the rapid structure elucidation of both compounds. The combinations of modern NMR, MS techniques highlight the strength of such combinations in providing rapid and precise answers to newly discovered compounds.

Recent advances in metagenomics have made genomic analysis based biosynthetic pathway prediction possible. Understanding the biosynthetic pathway of new compounds will provide more information for structure elucidations and mining of other new compounds. I have proposed the biosynthetic pathway of madurastatin compounds for the first time based on genomic analysis of the producing organism WMMA-1423 (Chapter 2), which will provide solutions to biosynthetic pathways of other similar bacteria natural products in future.

7.2. Future directions

The major purpose of this thesis is describing new methods and concepts for rapid new compound discovery from natural product sources, which will contribute to lowering the known compound rediscovery rate problem the community is facing; and developing new tools for natural product chemists to complete the structure elucidation of newly discovered compounds rapidly. We have applied PCA analysis to a marine bacteria natural product library generation in order to maximize our chance to discover new antibiotics. I believe in future, more and more compounds will be put into this library, which will eventually become a good guideline to determine whether a new hit is actually a new compound or a known one. Also, our library can be screened for other activities, such as modulating the concentrations of cytokines (Chapter 5). I think in future, we should test the biological activities of our natural product library on other target, such as cytotoxicity against cancer cell lines.

The Isotopic Fine Structure (IFS) analysis method will potentially be a good solution for unambiguous molecular formula determination of natural products. Advanced NMR techniques, such as ^{13}C-^{13}C COSY, ^{13}C-^{15}HMQC, solvent suppression experiments will lead to faster solutions to natural products. In my opinion, these improvements in MS and NMR techniques will essentially benefit the quick and correct structure elucidations of compounds with uncommon chemical structures.

For the RiPP haliclonamycin (Chapter 4), preliminary NMR studies have revealed several oxazole units in the structure, which classify this compound into the category of azole-containing linear RiPP. Whole genome sequencing of the producing organism is currently underway and once it is available, genomic analysis will be conducted to predict the

biosynthetic mechanism of haliclonamycin. With the help of some genomic analysis tools such as antiSMASH[4], RippMiner[5], the amino acid sequence of haliclonamycin can be solved, leading to the structure determination of haliclonamycin.

In Chapter 5, we have applied fractions of our natural product library to a mouse macrophage based assay. But doing this, we wanted to test if those mixtures could modulate the pro-inflammatory or anti-inflammatory cytokine concentrations produced by the macrophages since cytokines are related to human host defense mechanisms, one of the most important immunological activities. One fraction collected from natural products produced by WMMB-272 (*Streptomyces* sp.) show significant influences on the concentrations of several cytokines and it has guided me to isolate two known compounds, phenelfamycin analogs and pyridinopyrone A, that are likely to be responsible for the initially observed biological activities. More research and experiments need to be done in future to achieve the modulations of human host defenses by natural products. However, our studies have proved the concept that use fractions of natural product library in this cytokine concentration assay can lead to the identifications of pure compounds with the corresponding biological activities.

For ecteinamines (Chapter 5), the absolute configurations need to be analyzed. For now, only the relative configuration of the sugar moiety and part of the tetrahydrothiazole moiety has been proposed. I think *J*-based analysis of all possible conformations needs be conducted in order to predict the relative configurations of C-19, C-23, C-25, C-26, and the sugar unit. Molecular modeling of both *R* and *S* configurations on C-14 will be conducted in order to determine this challenging target. ECD experiments will potentially solve the absolute configuration of ecteinamines. Since the whole genome sequencing of their producing

organism WMMB-482 is also underway, biosynthetic pathway proposal will also be a good direction to follow because this type of highly modified nonribosomal peptide is not common in nature. Biological activities of ecteinamines (and streptnatamide A, Chapter 6) need to be evaluated as well.

7.3. References

(1). Newman, D. J.; Cragg, G. M. *J. Nat. Prod.*, **2012**, *75*, 311-335.

(2). Newman, D. J.; Cragg, G. M.; In *Functional Molecules from Natural Sources*; Wrigley, S. K; Thomas, R.; Nicholson, N.; Bedford, C., Eds.; RSC Publications: Cambridge, UK, 2010; pp 3-36.

(3). Davies, J.; Davies, D.; *Microbiol. Mol. Biol. Rev.*, **2010**, *74*, 417-433.

(4). Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, *47*, W81–W87.

(5). Agrawal, P.; Khater, S.; Gupta, M.; Sain, N.; Mohanty, D. *Nucleic Acids Research*, **2017**, *45*, W80-W88.

Appendix A:

Supplementary Data for Chapter 2

Figure S1. ^{1}H NMR Spectrum of (−)-Madurastatin C1 (**18**, 400 MHz, DMSO-d_6)

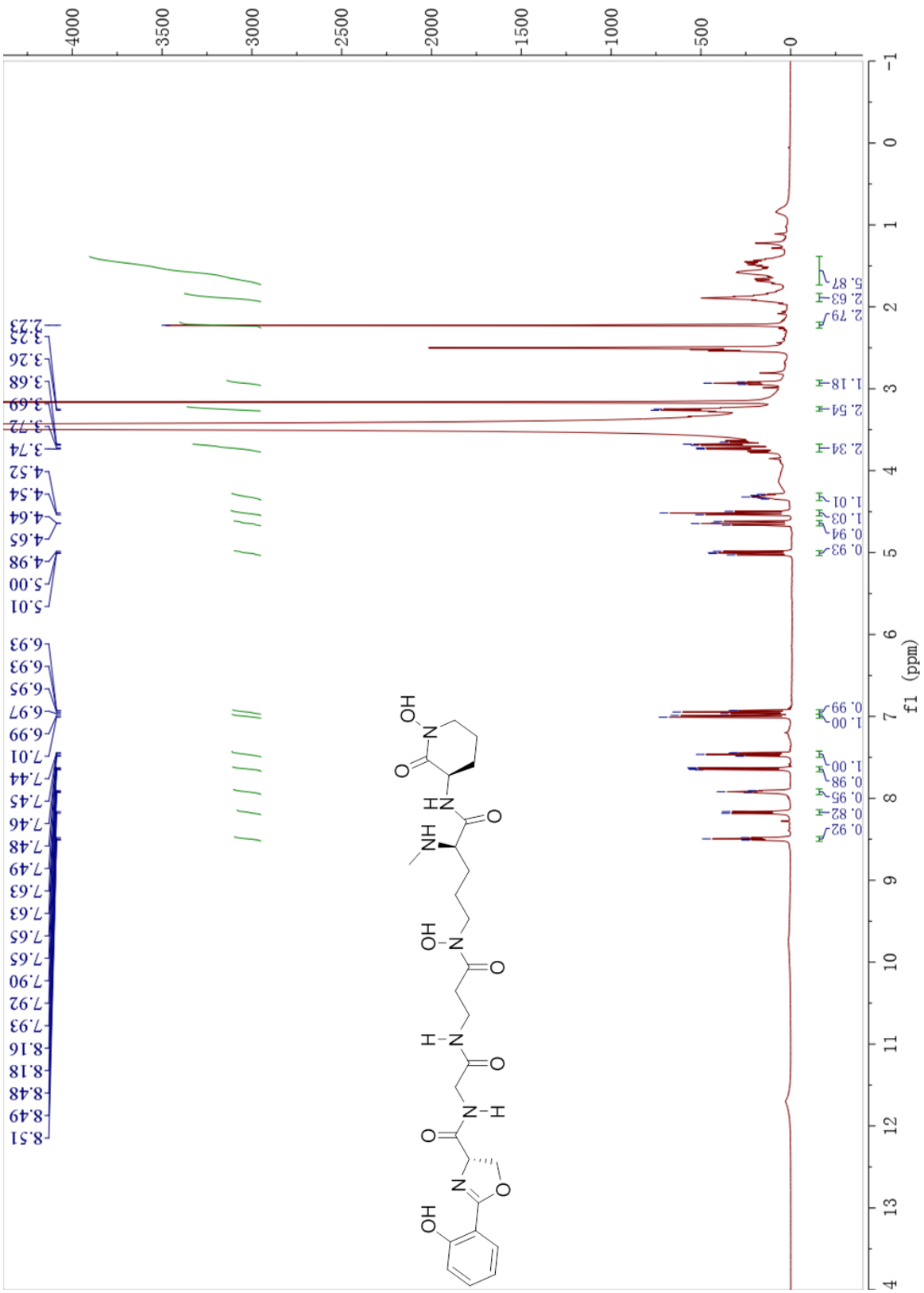


Figure S2. ^{13}C NMR Spectrum of (−)-Madurastatin C1 (**18**, 101 MHz, DMSO-d_6)

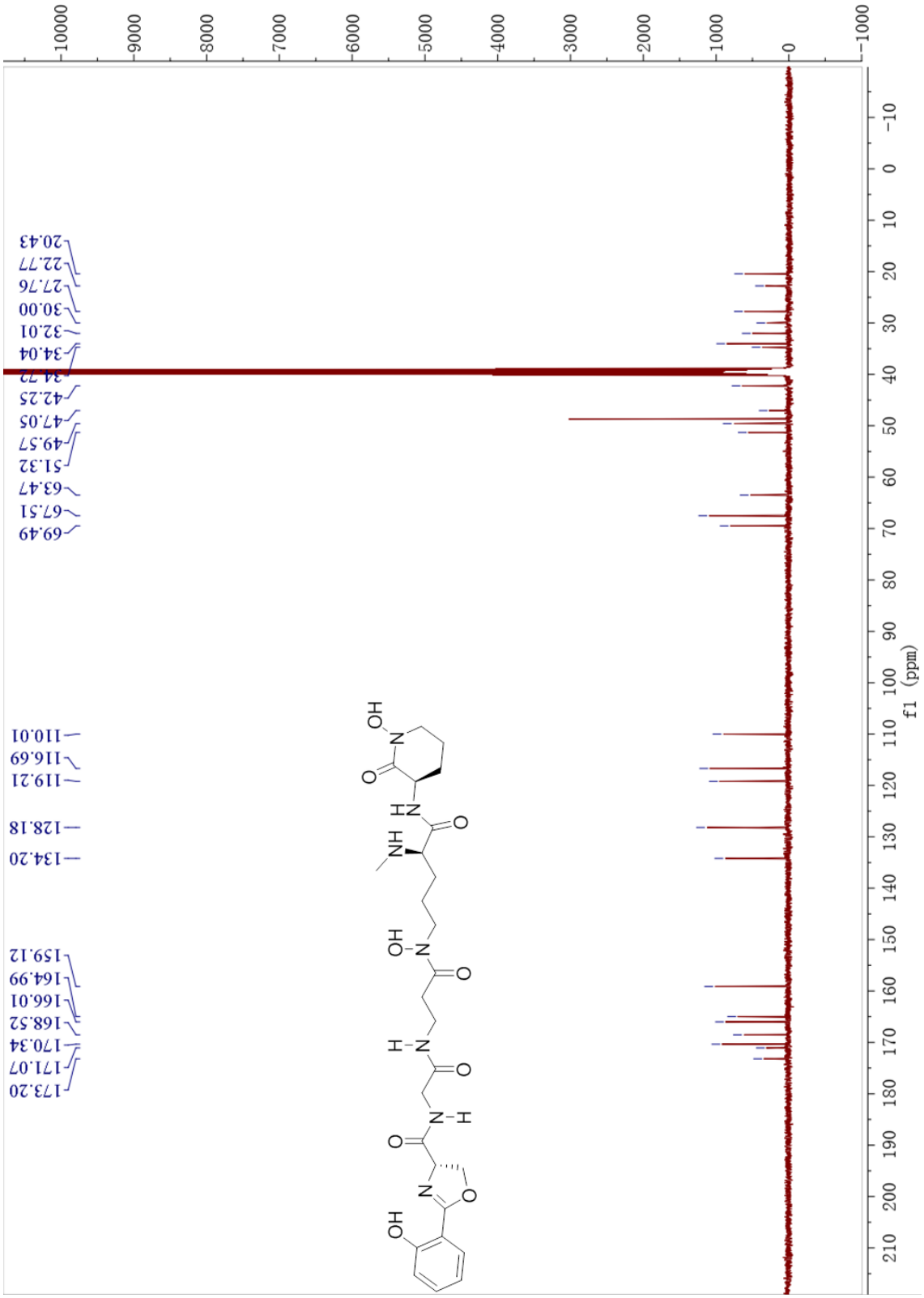


Figure S3. ^{1}H NMR Spectrum of Madurastatin D1 (**19**, 500 MHz, DMSO-d_6)

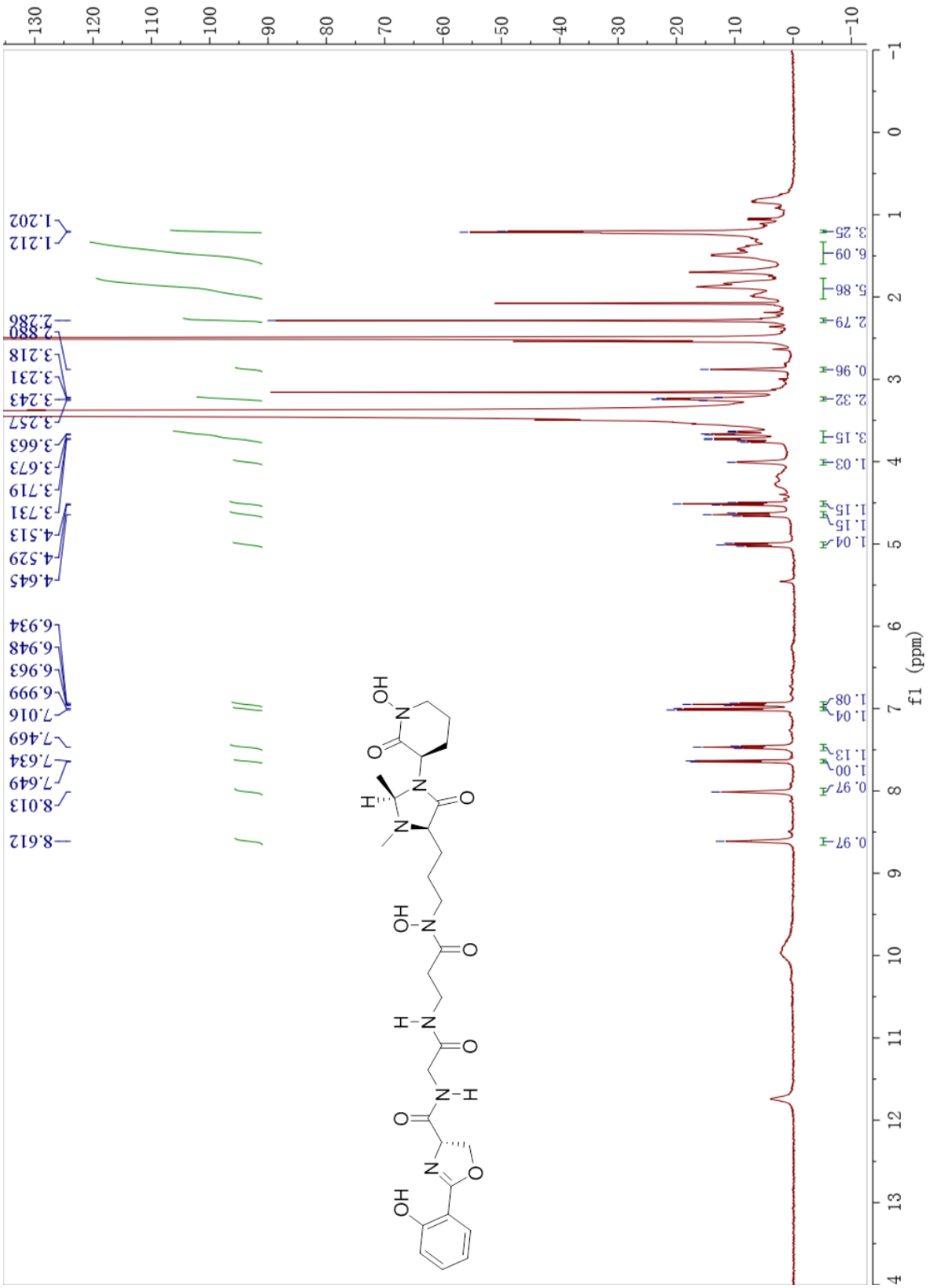


Figure S4. ^{13}C NMR Spectrum of Madurastatin D1 (**19**, 125 MHz, DMSO-d_6)

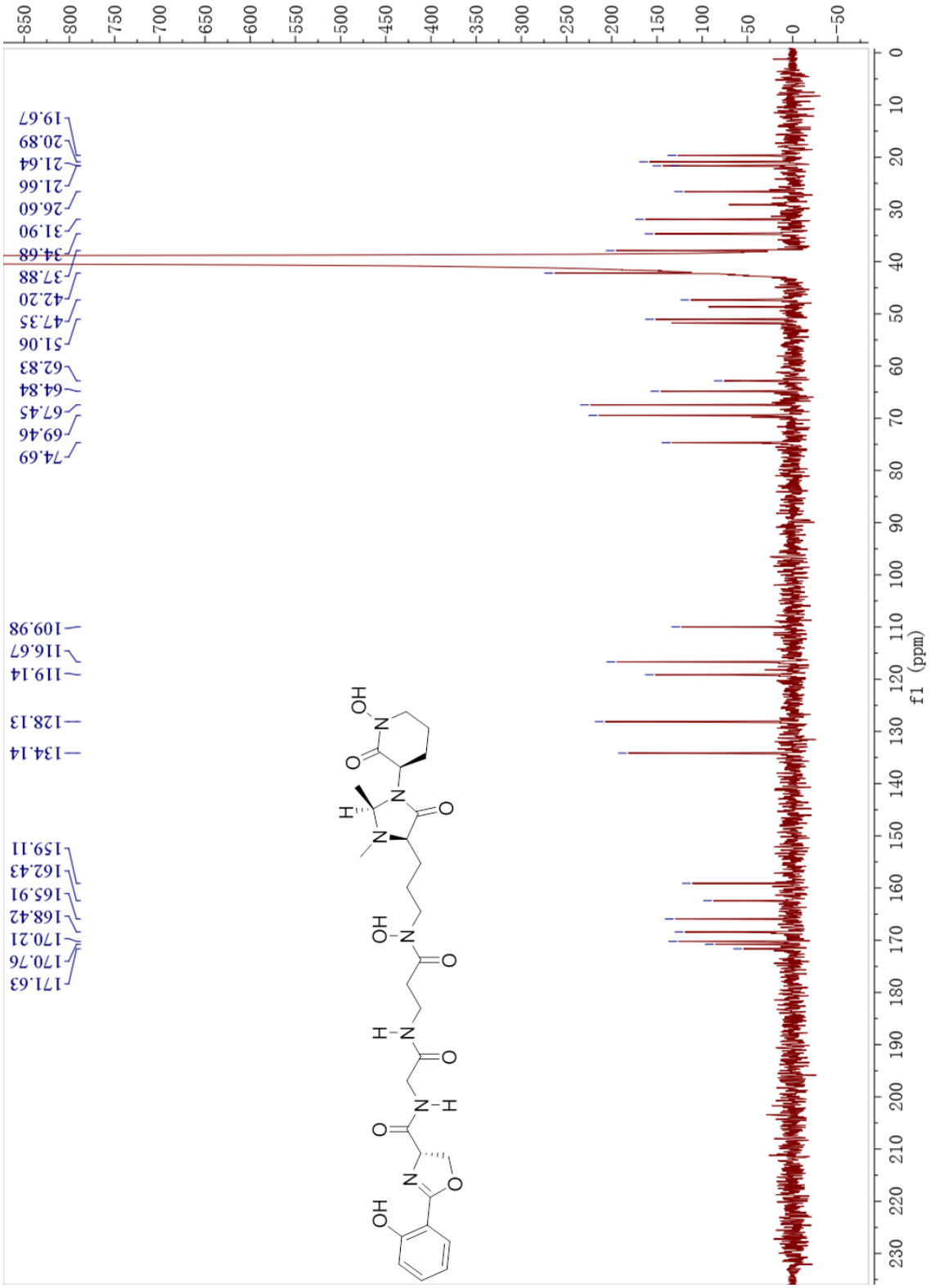


Figure S5. gCOSY Spectrum of Madurastatin D1 (**19**, 600 MHz, DMSO-d_6)

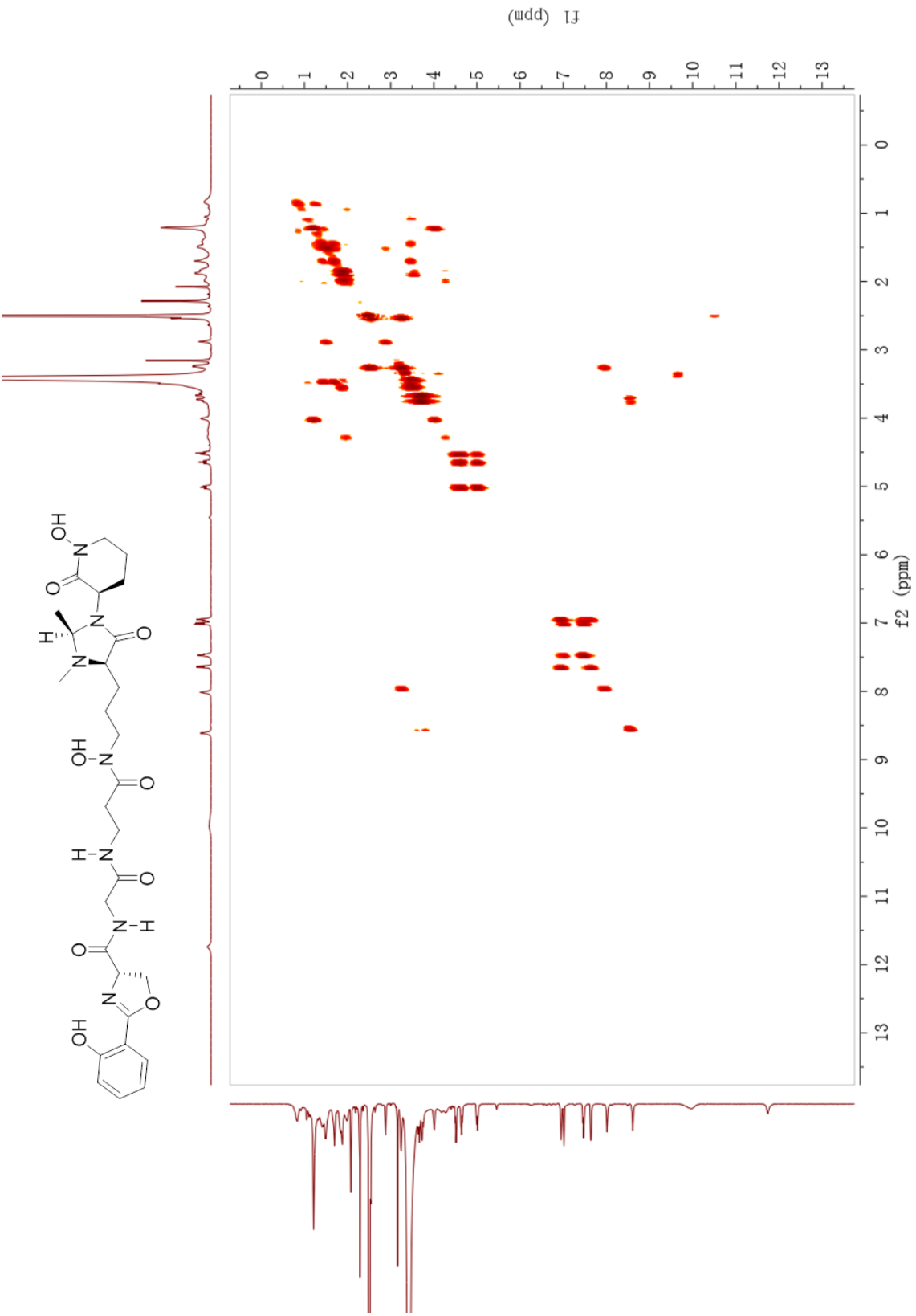


Figure S6. gHSQC Spectrum of Madurastatin D1 (**19**, 600 MHz, DMSO-d_6)

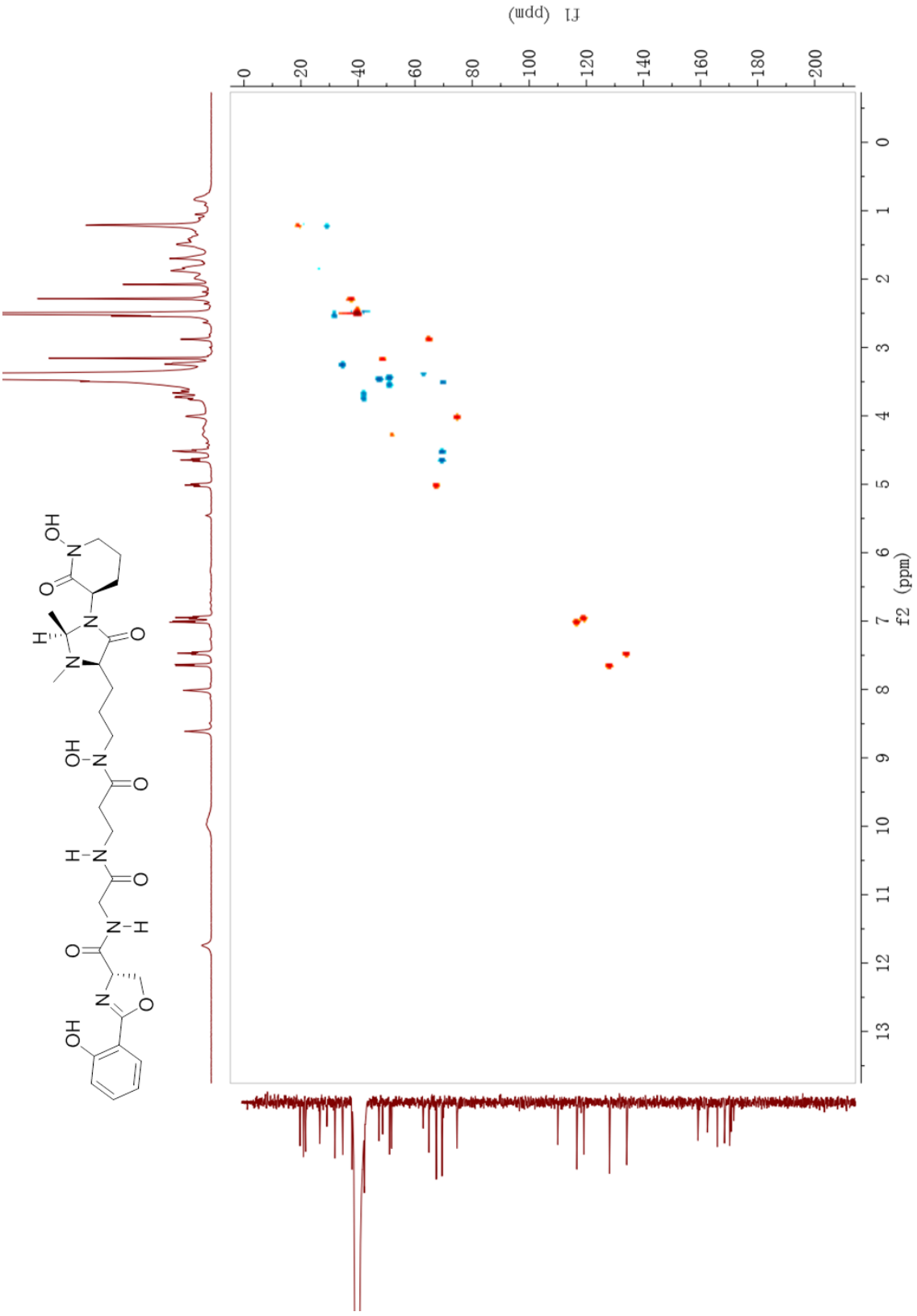


Figure S7. gHMBC Spectrum of Madurastatin D1 (**19**, 600 MHz, DMSO-d_6)

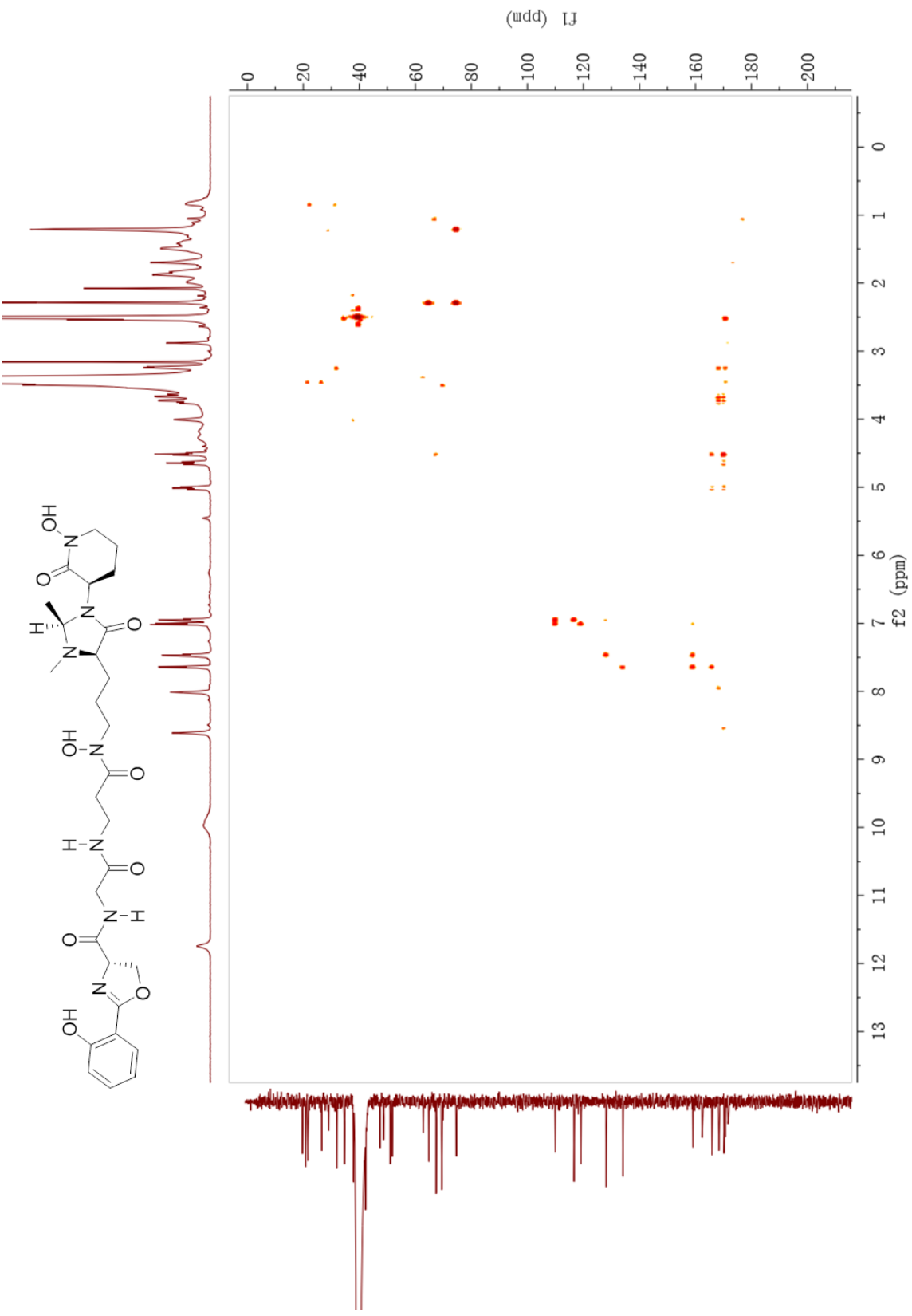


Figure S8. ^{1}H NMR Spectrum of Madurastatin D2 (**20**, 500 MHz, DMSO-d_6)

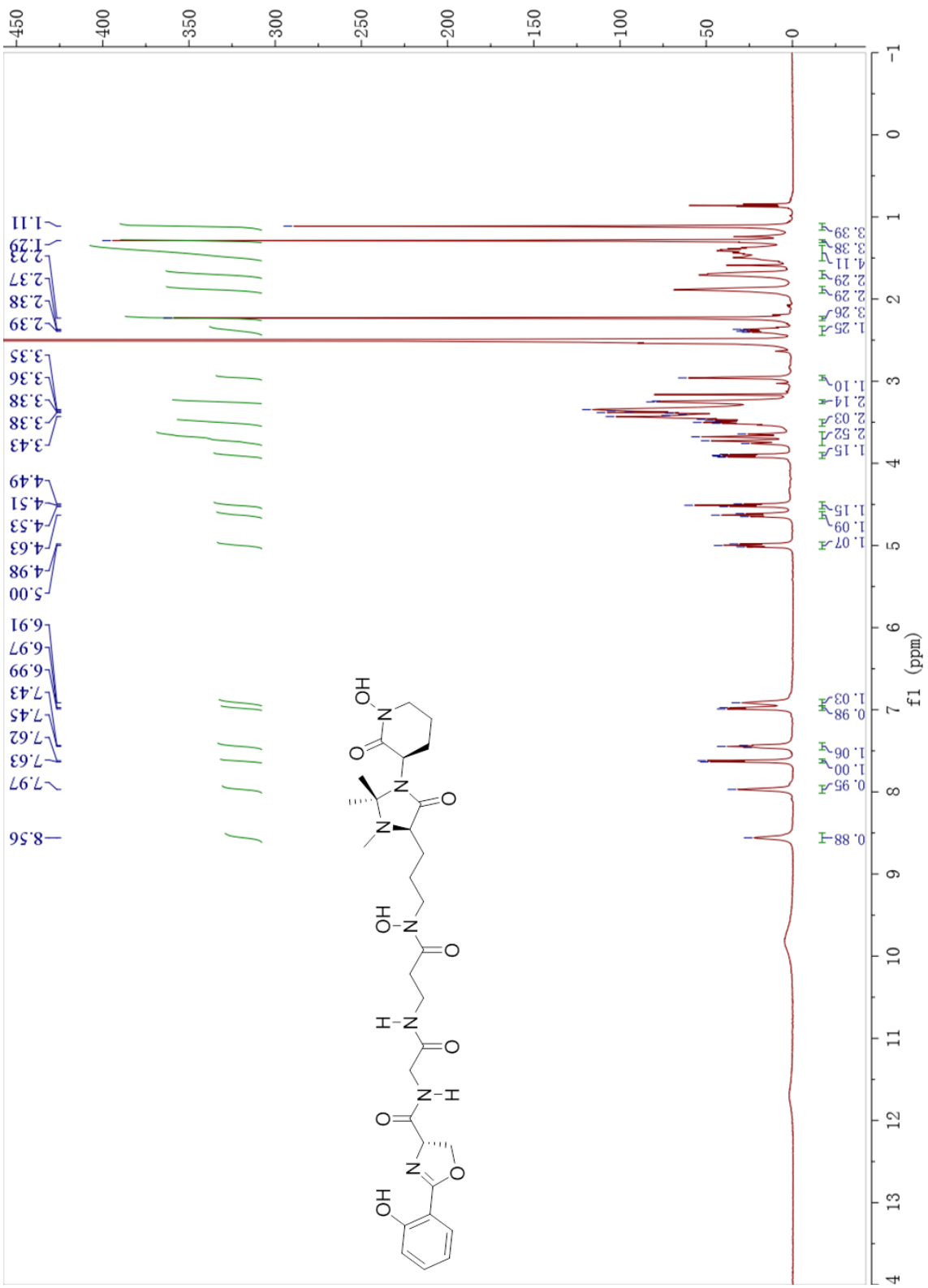


Figure S9. ^{13}C NMR Spectrum of Madurastatin D2 (**20**, 125 MHz, DMSO-d_6)

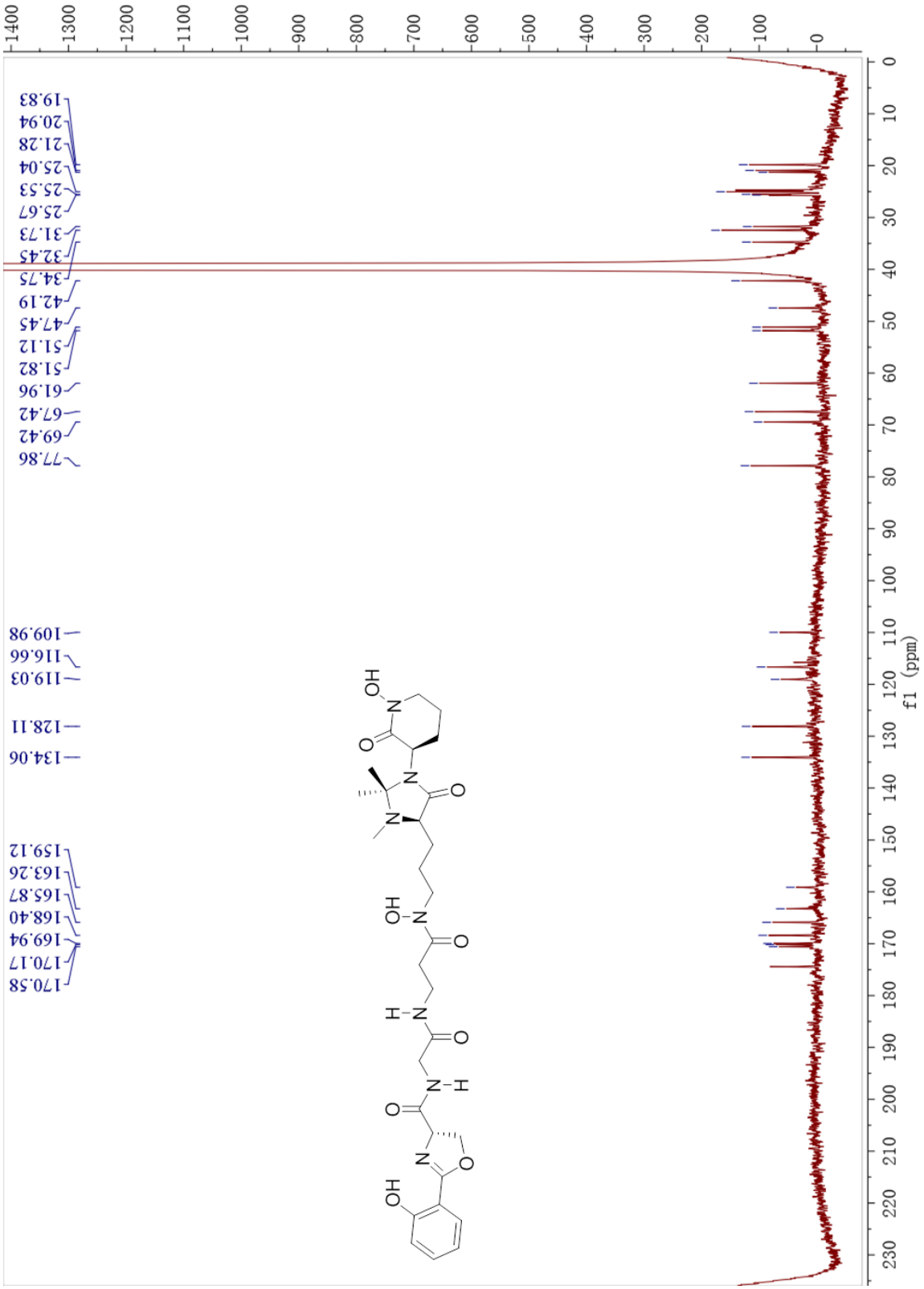


Figure S10. gCOSY Spectrum of Madurastatin D2 (**3**, 600 MHz, DMSO-d_6)

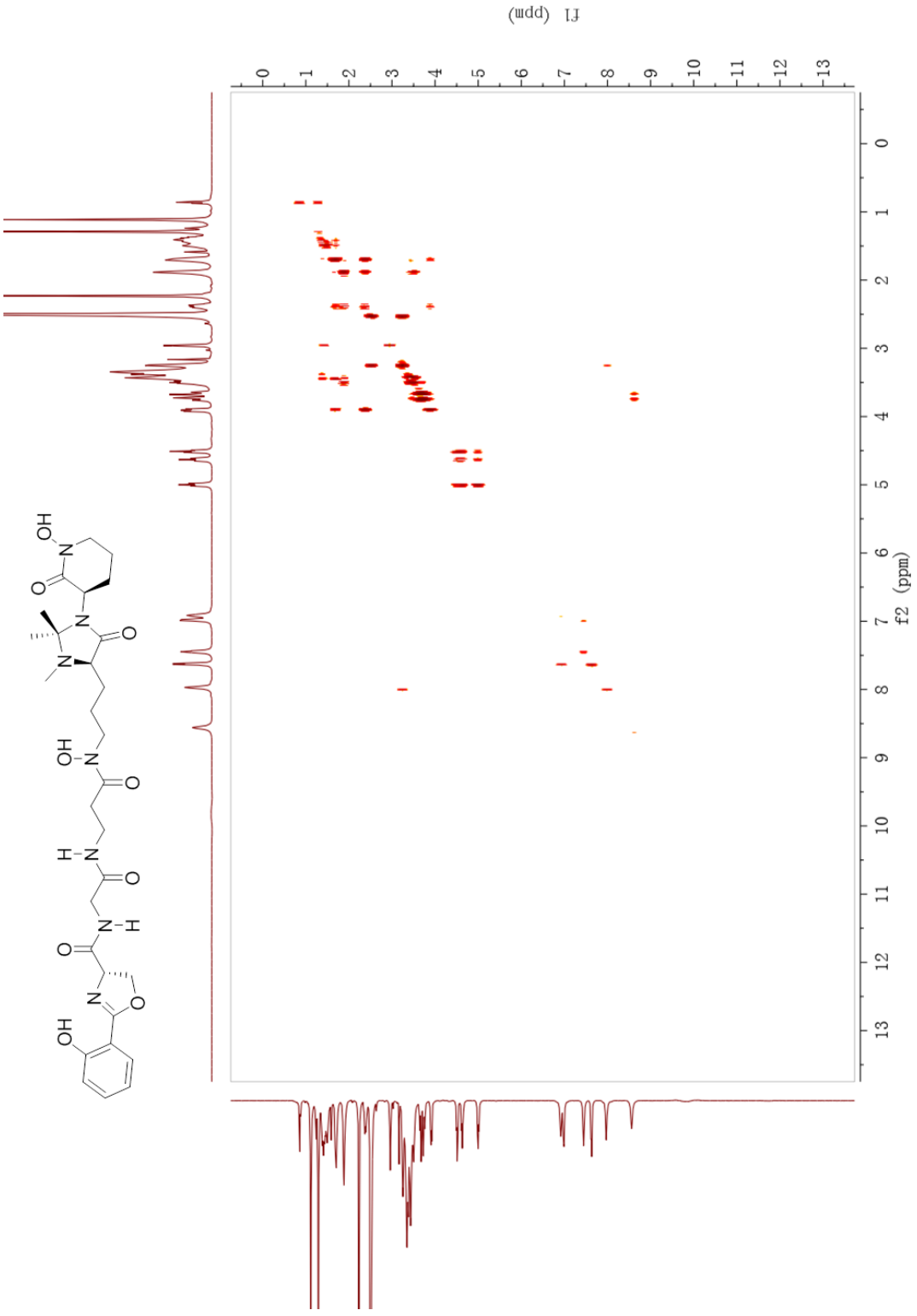


Figure S11. gHSQC Spectrum of Madurastatin D2 (**3**, 600 MHz, DMSO-d_6)

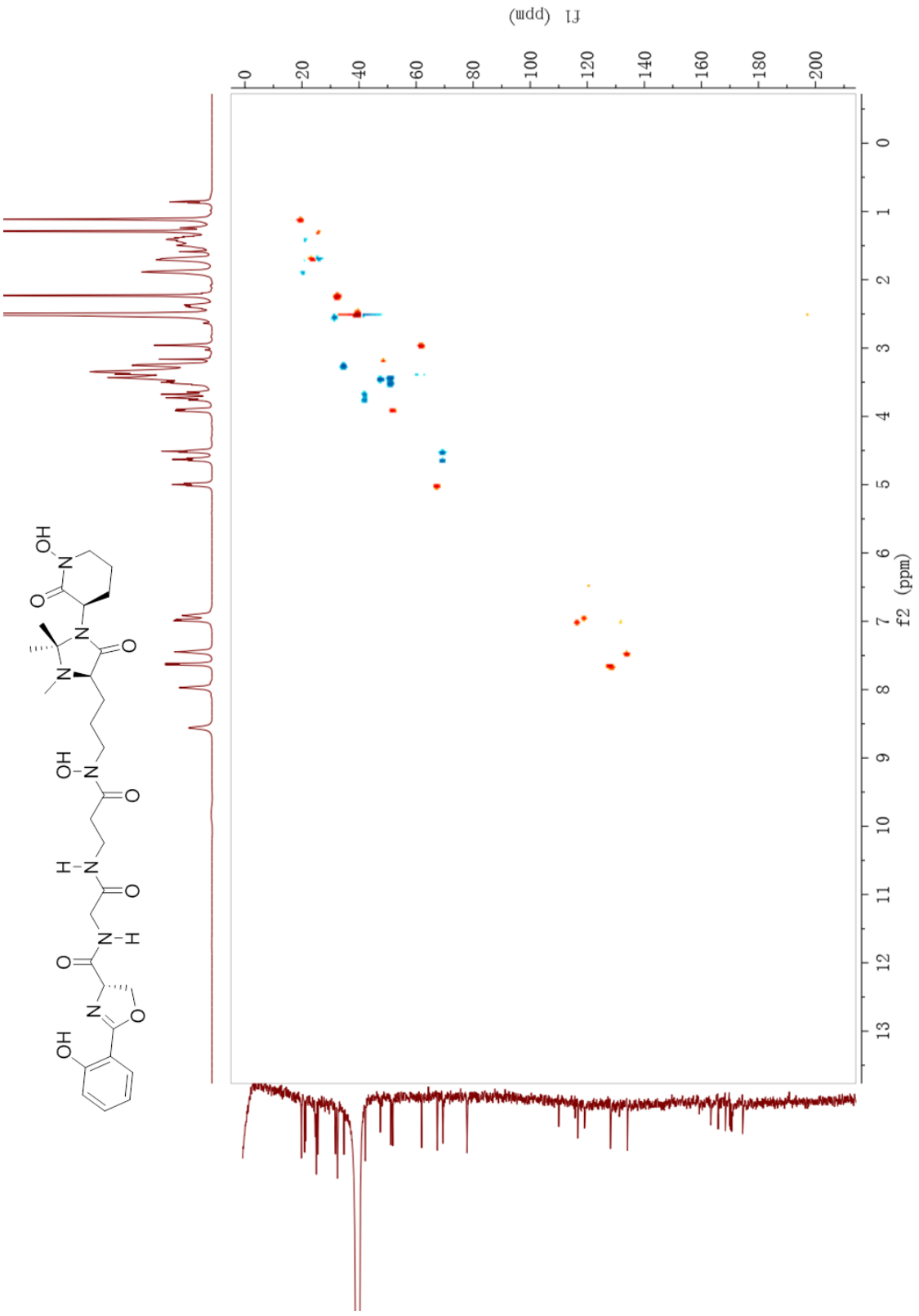


Figure S12. gHMBC Spectrum of Madurastatin D2 (**3**, 600 MHz, DMSO-d_6)

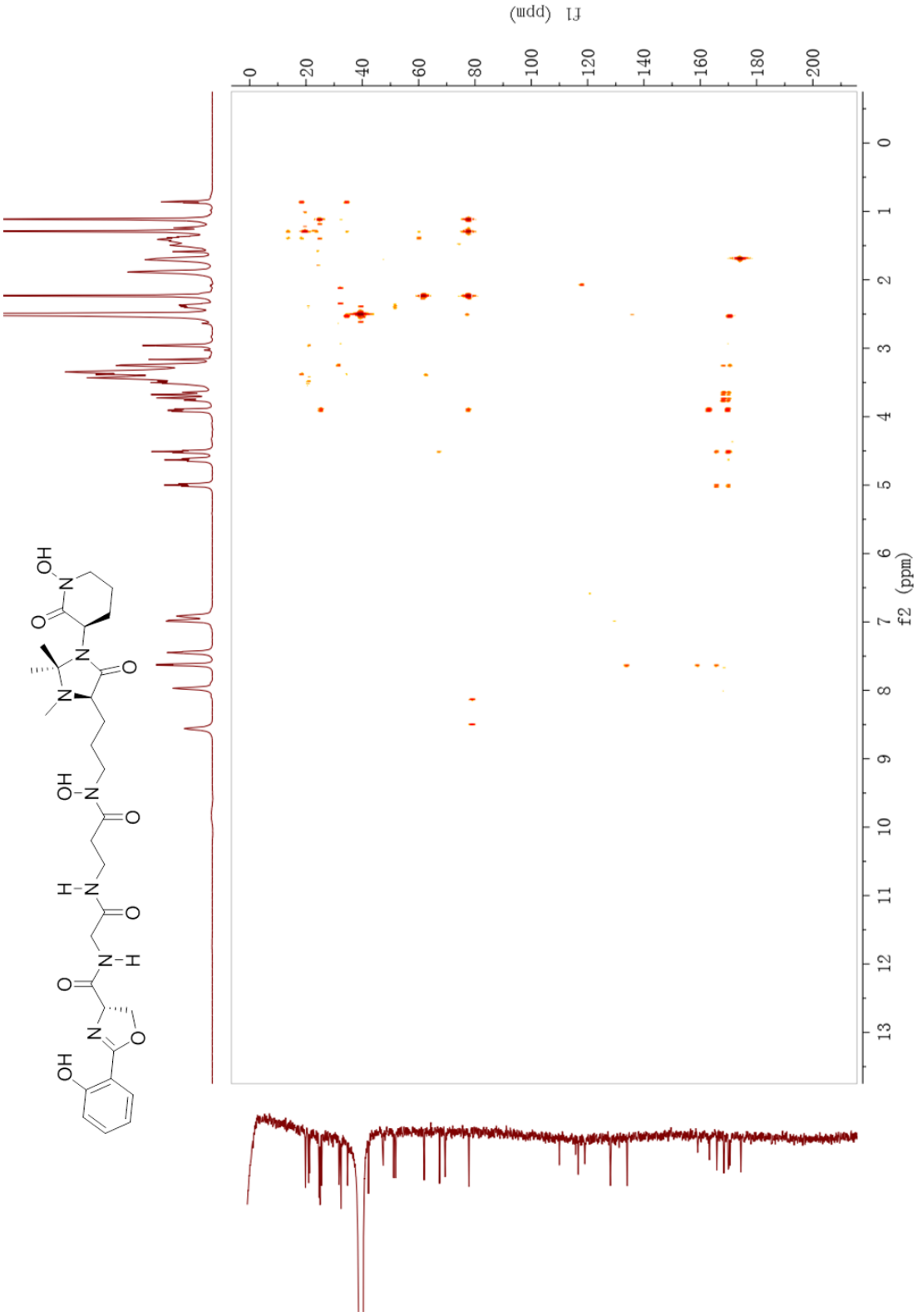


Figure S13. Positive Ion HRESIMS of (−)-Madurastatin C1 (**18**) and the Fe adduct.

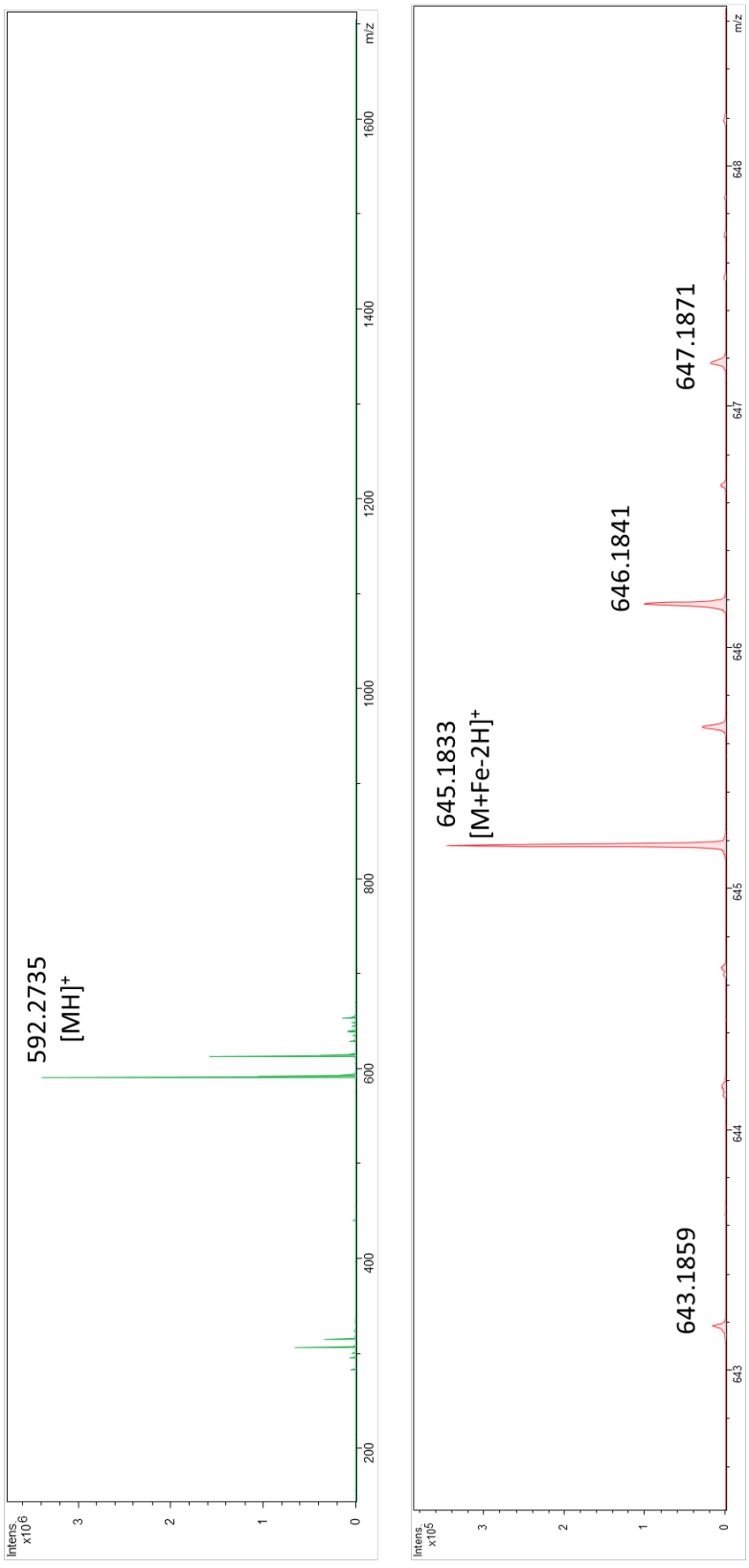


Figure S14. Positive Ion HRESIMS of Madurastatin D1 (**19**) and the Fe adduct.

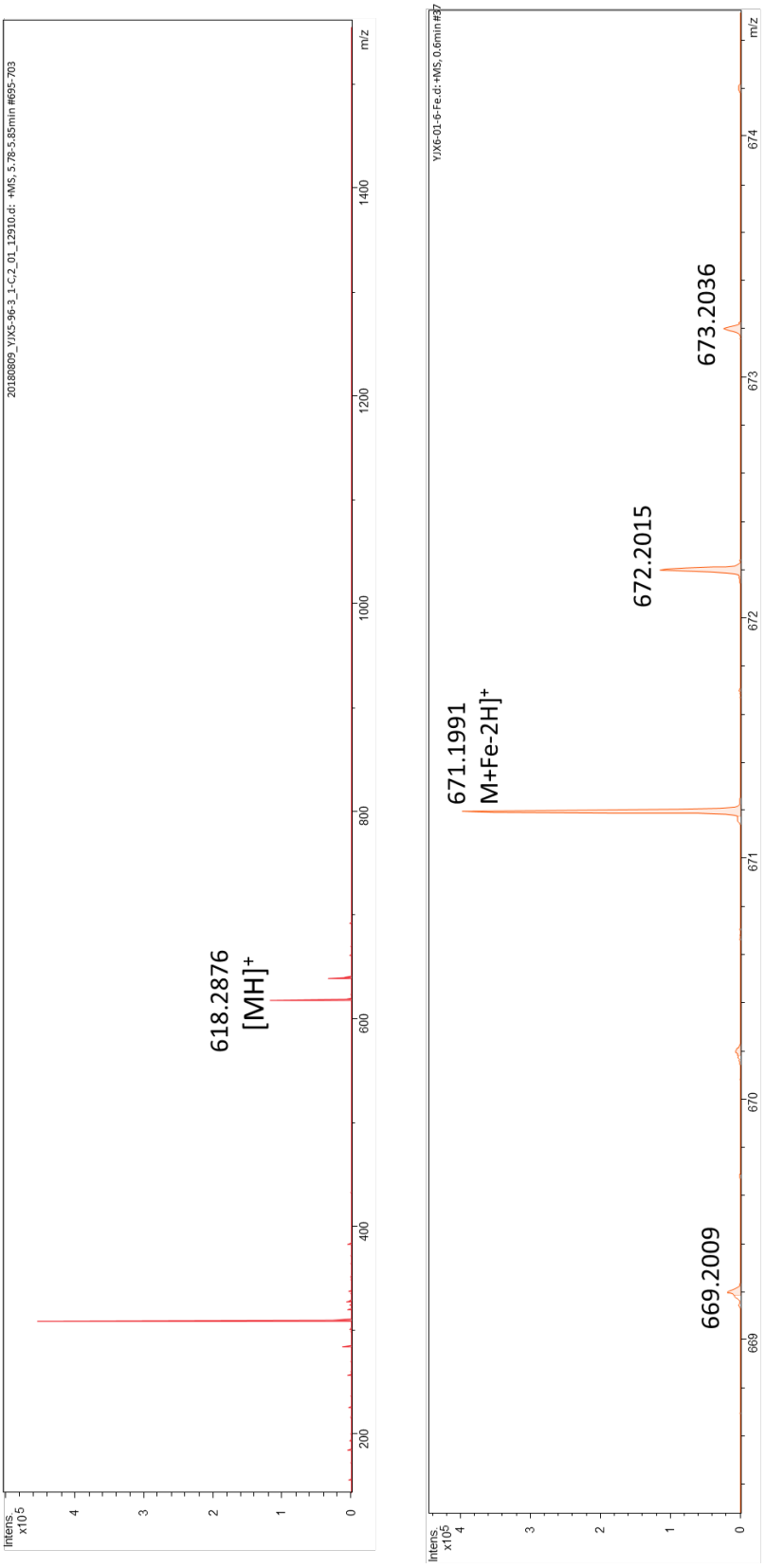


Figure S15. Positive Ion HRESIMS of Madurastatin D2 (**20**) and the Fe adduct.

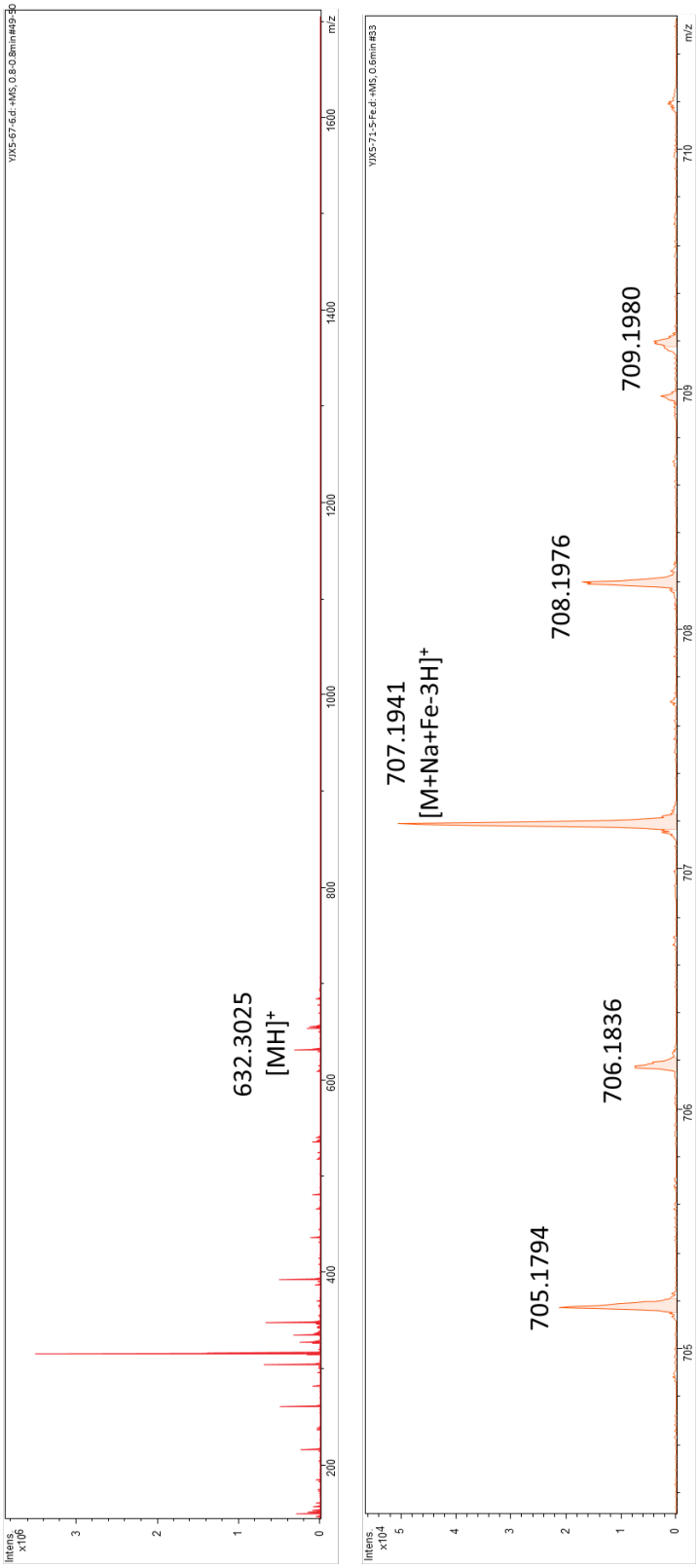
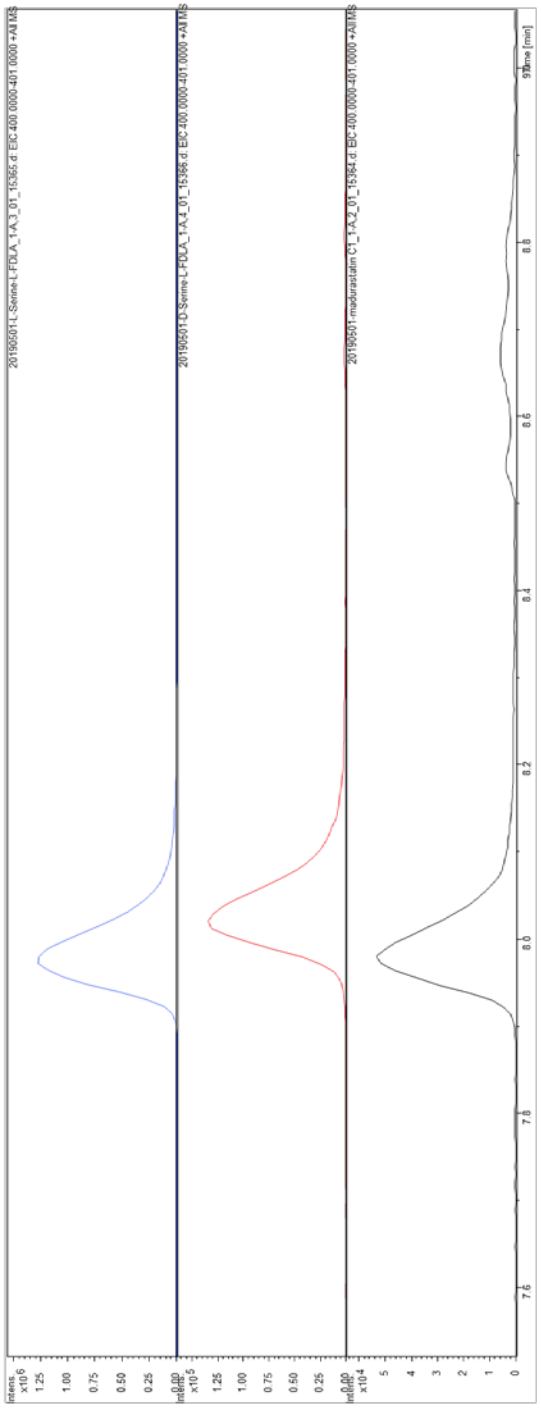


Figure S16. LCMS spectra of Marfey's analysis of (−)-Madurastatin C1 (**18**)



Extracted LC/MS chromatogram spectra of the hydrolyte of L-serine (left trace, blue), D-Serine (middle trace, red) and (−)-Madurastatin C1 (**1,** right trace, black) coupled with L-FDLA.

Table S1. Annotations on biosynthetic gene cluster

Gene	Proposed Annotation	Notable Homologs	Protein % ID	Organism
Mad1	acyl-CoA dehydrogenase	WP_089316159.1	96	Actinomadura mexicana
Mad2	hypothetical protein	WP_067467552.1	90	Actinomadura macra
Mad3	hypothetical protein	WP_021600299.1	95	Actinomadura madurae
Mad4	methyltransferase domain-containing protein	WP_130186440.1	64	Massilia lutea
Mad5	YbdK family carboxylate-amine ligase	WP_089316162.1	92	Actinomadura mexicana
Mad6	sensor histidine kinase	WP_089316163.1	82	Actinomadura mexicana
Mad7	aminoglycoside phosphotransferase	WP_132196454.1	81	Actinomadura darangshiensis
Mad 8	DUF2269 domain-containing protein	WP_131742939.1	73	Actinomadura sp. LMG 30035
Mad 9	hypothetical DoxX membrane protein	WP_089316164.1	86	Actinomadura mexicana
Mad 10	GNAT family N-acetyltransferase	WP_089316165.1	92	Actinomadura

				mexicana
Mad 11	SAM-dependent methyltransferase	WP_132196462.1	80	Actinomadura darangshiensis
Mad 12	acyltransferase	WP_089316167.1	93	Actinomadura mexicana
Mad 13	PucR family transcriptional regulator	WP_033331132.1	94.06	Actinomadura madurae
Mad 14	esterase	WP_132205301.1	78.78	Actinomadura darangshiensis
Mad 15	hypothetical protein	WP_132205297.1	91.35	Actinomadura darangshiensis
Mad 16	1-phosphofructokinase	WP_131941588.1	92.42	Actinomadura bangladeshensis
Mad 17	DeoR/GlpR transcriptional regulator	WP_132203352.1	94.85	Actinomadura darangshiensis
Mad 18	ABC transporter substrate-binding protein	WP_089316172.1	96.46	Actinomadura mexicana
Mad 19	branched-chain amino acid ABC transporter permease	WP_075021706.1	89.06	Actinomadura madurae
Mad 20	branched-chain amino acid ABC transporter permease	WP_089316174.1	96.58	Actinomadura mexicana
Mad 21	ABC transporter ATP-binding protein	WP_089328063.1	90.21	Actinomadura

				meyerae
Mad 22	ABC transporter ATP-binding protein	WP_089316176.1	92.66	Actinomadura mexicana
Mad 23	TetR/AcrR family transcriptional regulator	WP_089316177.1	95.50	Actinomadura mexicana
Mad 24	long-chain fatty acid--CoA ligase	WP_089316224.1	93.05	Actinomadura mexicana
Mad 25	MFS transporter	WP_089328068.1	72.63	Actinomadura meyerae
Mad 26	hypothetical esterase	WP_089316178.1	92.66	Actinomadura mexicana
Mad 27	MbtH family protein	CahE	63	Streptomyces gandocaensis
Mad 28	lysine N(6)-hydroxylase/L-ornithine N(5)-oxygenase family protein	cahMO	57	Streptomyces gandocaensis
Mad 29	iron-siderophore ABC transporter substrate-binding protein	WP_089316225.1	89.85	Actinomadura mexicana
Mad 30	non-ribosomal peptide synthetase	AMYAL_RS0130210	47	Amycolatopsis alba
Mad 31	salicylate synthase	AmcL	59	Amycolatopsis sp. AA4
Mad 32	Fpg/Nei family DNA glycosylase	WP_089328073.1	73.51	Actinomadura

				meyerae
Mad 33	hypothetical protein	WP_089316188.1	60.54	Actinomadura mexicana
Mad 34	hypothetical protein	WP_117402065.1	70.83	Actinomadura sp. LHW52907
Mad 35	exonuclease	WP_117402874.1	75.86	Actinomadura sp. LHW52907
Mad 36	cysteine hydrolase	WP_084521215.1	96.67	Nocardia uniformis
Mad 37	hypothetical protein	WP_112468680.1	87.66	Streptomyces triticisoli
Mad 38	alpha-ketoglutarate-dependent dioxygenase AlkB	WP_075024900.1	83.90	Actinomadura madurae
Mad 39	hypothetical protein	WP_131732831.1	63.64	Actinomadura formosensis
Mad 40	TROVE domain-containing protein	WP_131888826.1	82.28	Actinomadura sp. H3C3
Mad 41	hypothetical protein	WP_089310155.1	84.88	Actinomadura mexicana
Mad 42	PadR family transcriptional regulator	WP_067474735.1	77.20	Actinomadura hibisca
Mad 43	hypothetical protein	WP_131985403.1	59.77	Actinomadura sp.

				6K520
Mad 44	hypothetical protein kinase	WP_067640675.1	74.19	Actinomadura latina
Mad 45	DEAD/DEAH box helicase	WP_067640681.1	93.34	Actinomadura latina
Mad 46	BREX system ATP-binding protein BrxD	WP_067640683.1	94.43	Actinomadura latina
Mad47	hypothetical protein	WP_083947146.1	83.33	Actinomadura latina
Mad48	hypothetical protein	WP_028806279.1	75.00	Streptomyces sp. 303MFCol5.2
Mad49	hypothetical protein	WP_109379674.1	70.15	Streptomyces sp. NWU339
Mad50	TIGR04222 domain-containing membrane protein	WP_131944619.1	89.12	Actinomadura bangladeshensis
Mad51	hypothetical protein	None	NA	NA
Mad52	hypothetical protein	WP_131877791.1	78.52	Actinomadura sp. KC345
Mad53	hypothetical protein	WP_131877789.1	58.28	Actinomadura sp. KC345
Mad54	helix-turn-helix domain-containing protein (putative XRE regulator)	WP_132147513.1	77.78	Actinomadura sp. 7K507

Mad55	DUF397 domain-containing protein (putative abaA regulator)	WP_132147510.1	73.02	Actinomadura sp. 7K507
Mad56	BREX-2 system adenine-specific DNA-methyltransferase PglX	WP_067637150.1	67.55	Actinomadura latina
Mad57	BREX system serine/threonine kinase PglW	WP_067637152.1	95.89	Actinomadura latina
Mad58	DUF262 domain-containing protein (putative nuclease)	WP_067637154.1	78.87	Actinomadura latina
Mad59	CBS domain-containing protein	WP_131900158.1	69.14	Actinomadura sp. H3C3
Mad60	hydroxybenzoate -AMP ligase	CahJ	67	Streptomyces gandocaensis
Mad61	aspartate 1-decarboxylase	CahF	77	Streptomyces gandocaensis
Mad62	siderophore transporter EntS	WP_132199668.1	88.65	Actinomadura darangshiensis
Mad63	non-ribosomal peptide synthetase	CahA	58	Streptomyces gandocaensis
Mad64	Fe(3+)-siderophore ABC transporter permease	CahT3	49	Streptomyces gandocaensis
Mad65	Fe(3+)-siderophore ABC transporter permease	CahT2	49	Streptomyces gandocaensis

Mad66	helix-turn-helix transcriptional regulator (TetR/AcrR family)	WP_089316200.1	87.50	Actinomadura mexicana
Mad67	ABC transporter ATP-binding protein	WP_089316201.1	91.00	Actinomadura mexicana
Mad68	ABC transporter permease	WP_132199663.1	83.06	Actinomadura darangshiensis
Mad69	STAS domain-containing protein	WP_089316203.1	83.97	Actinomadura mexicana
Mad70	hypothetical sensor histidine kinase	WP_089316205.1	89.80	Actinomadura mexicana
Mad71	response regulator transcription factor	WP_089316206.1	95.09	Actinomadura mexicana
Mad72	hypothetical protein	WP_089316207.1	89.33	Actinomadura mexicana
Mad73	DNA polymerase IV	WP_089316208.1	92.37	Actinomadura mexicana
Mad74	cellulose 1,4-beta-cellobiosidase	WP_089316209.1	93.32	Actinomadura mexicana
Mad75	LacI family DNA-binding transcriptional regulator	WP_089316210.1	95.65	Actinomadura mexicana
Mad76	cellulose-binding protein	WP_132204372.1	87.35	Actinomadura darangshiensis

Mad77	cellulose 1,4-beta-cellobiosidase	WP_089316211.1	90.24	Actinomadura mexicana
Mad78	cellulose 1,4-beta-cellobiosidase	WP_089316211.1	96.77	Actinomadura mexicana

Appendix B:

Supplementary Data for Chapter 3

Figure S17. ^{1}H NMR Spectrum of Bacillimidazole A (**21**, 600 MHz, CD_3OD)

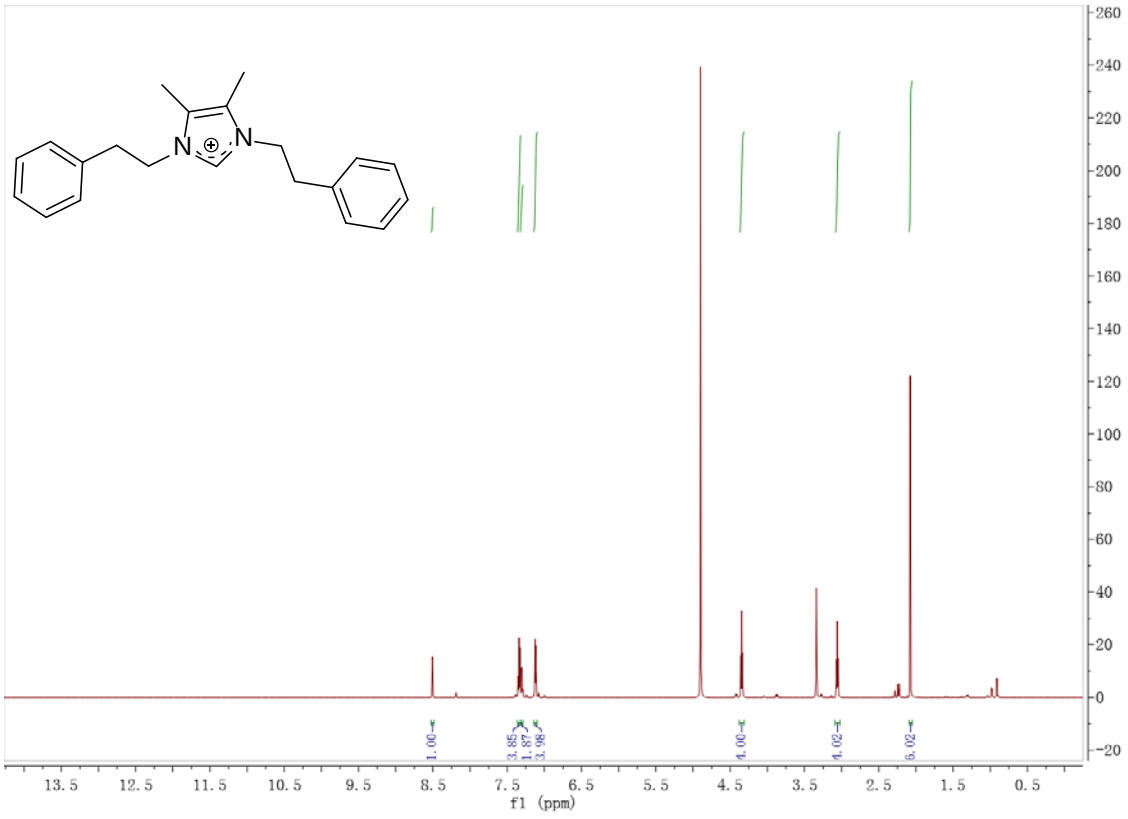


Figure S18. ^{13}C NMR Spectrum of Bacillimidazole A (**21**, 125 MHz, CD_3OD)

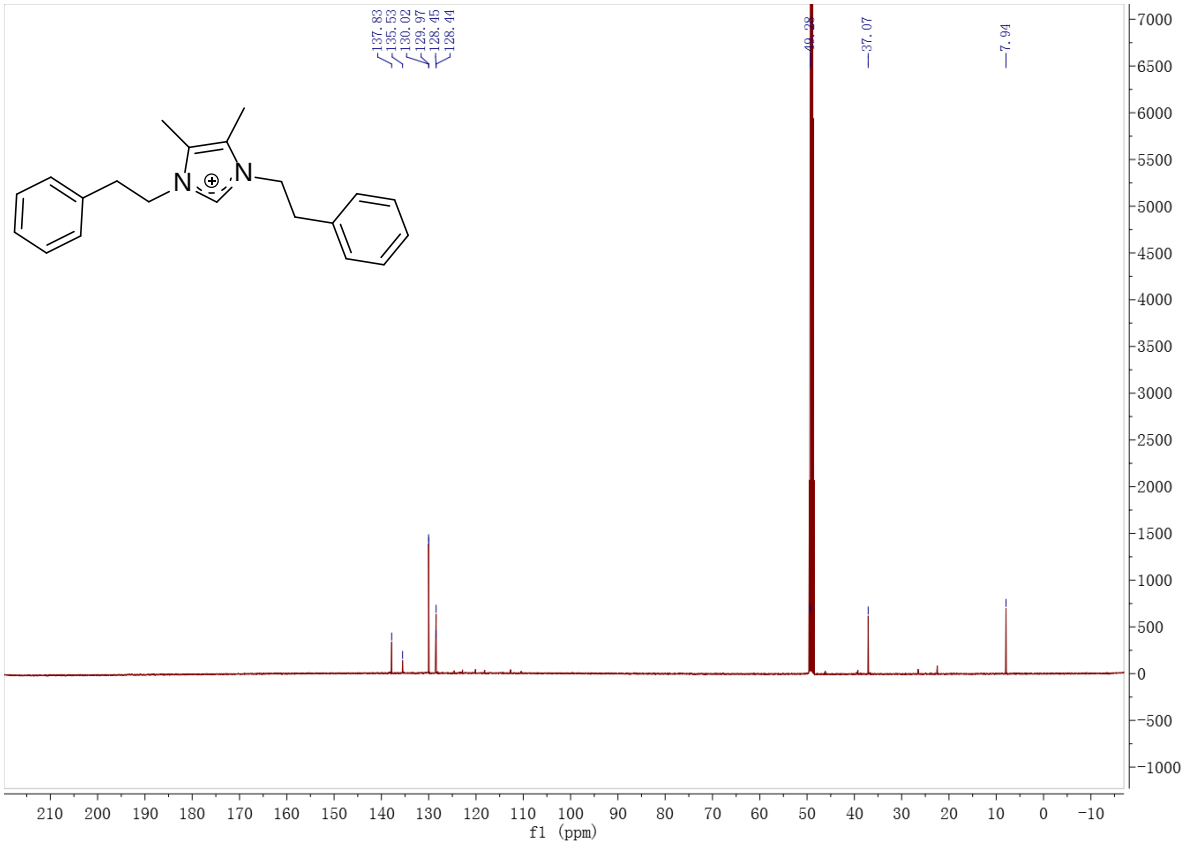


Figure S19. gCOSY Spectrum of Bacillimidazole A (**21**, 600 MHz, CD_3OD)

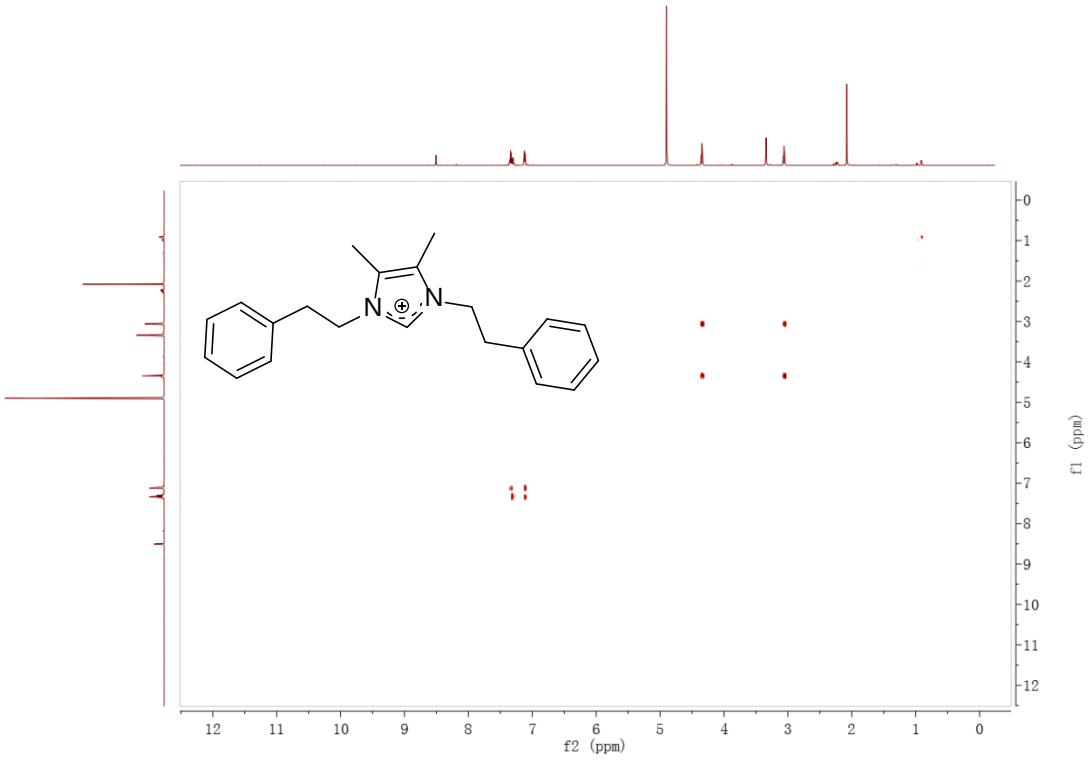


Figure S20. gHSQC Spectrum of Bacillimidazole A (2**1**, 600 MHz, CD_3OD)

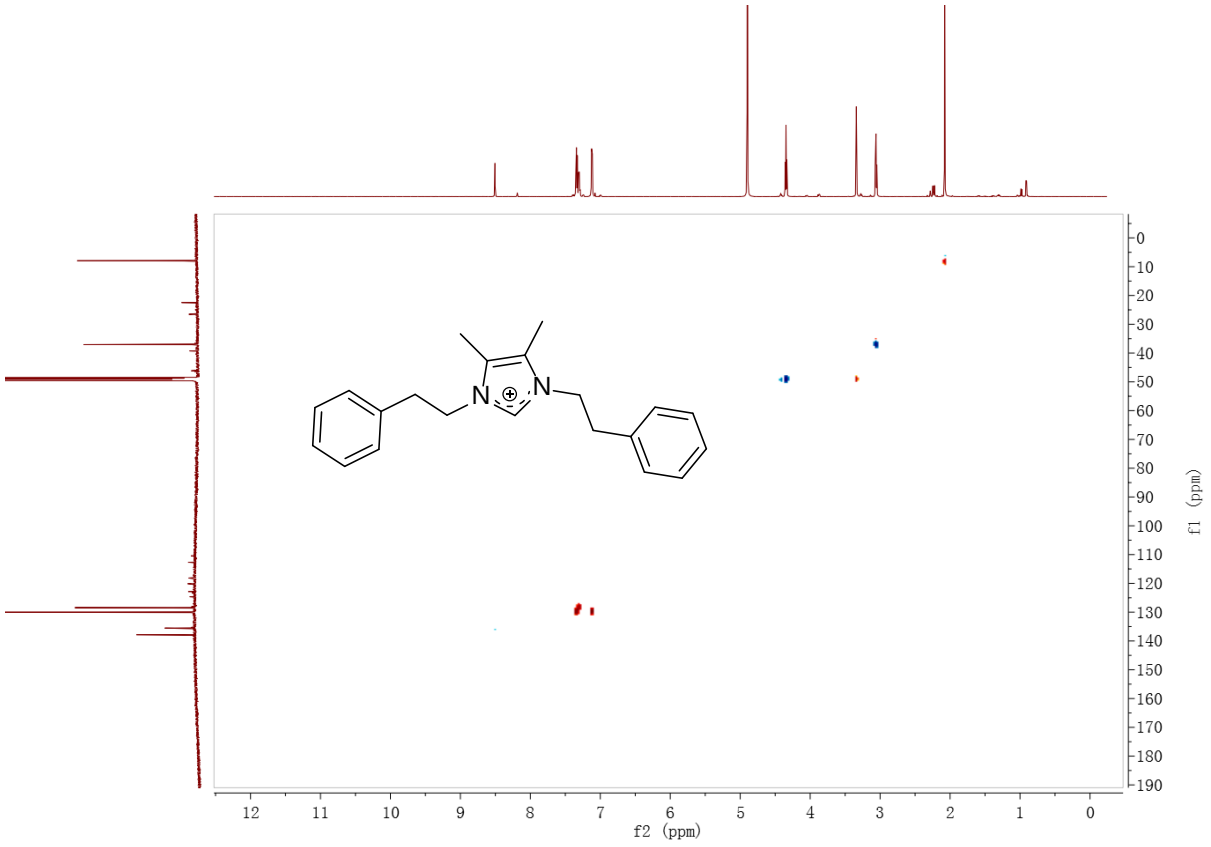


Figure S21. gHMBC Spectrum of Bacillimidazole A (**21**, 600 MHz, CD_3OD)

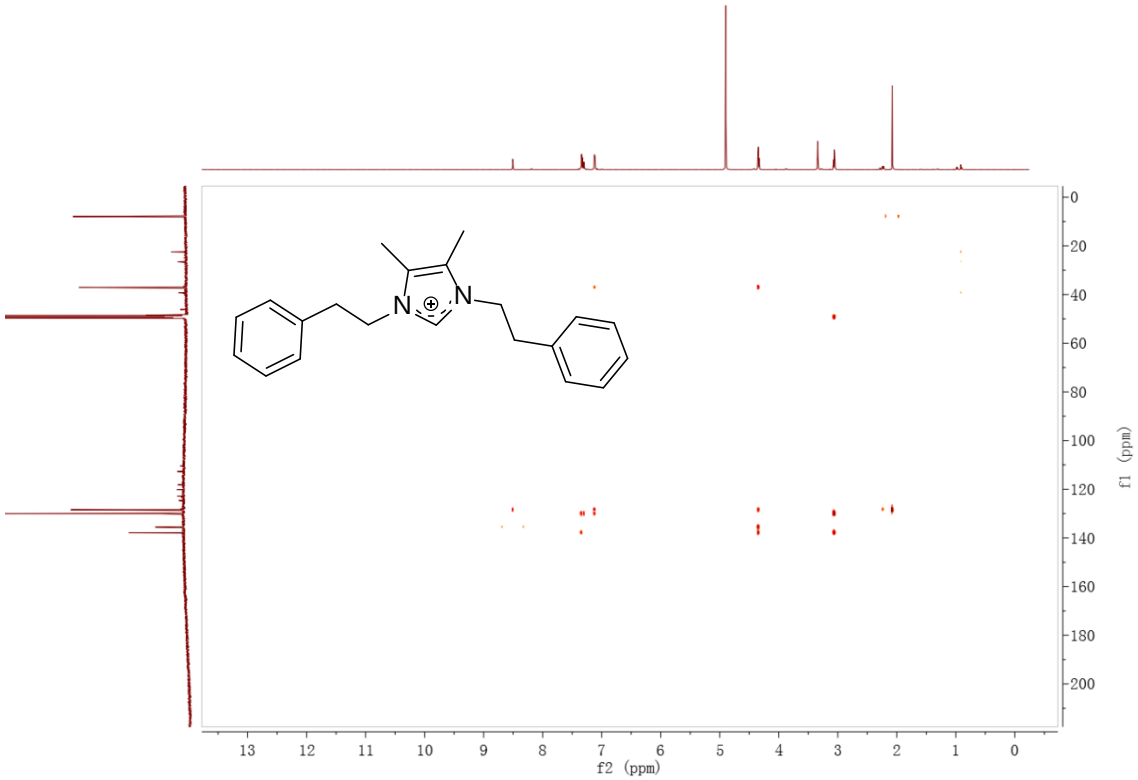


Figure S22. ^1H NMR Spectrum of Bacillimidazole B (**22**, 600 MHz, CD_3OD)

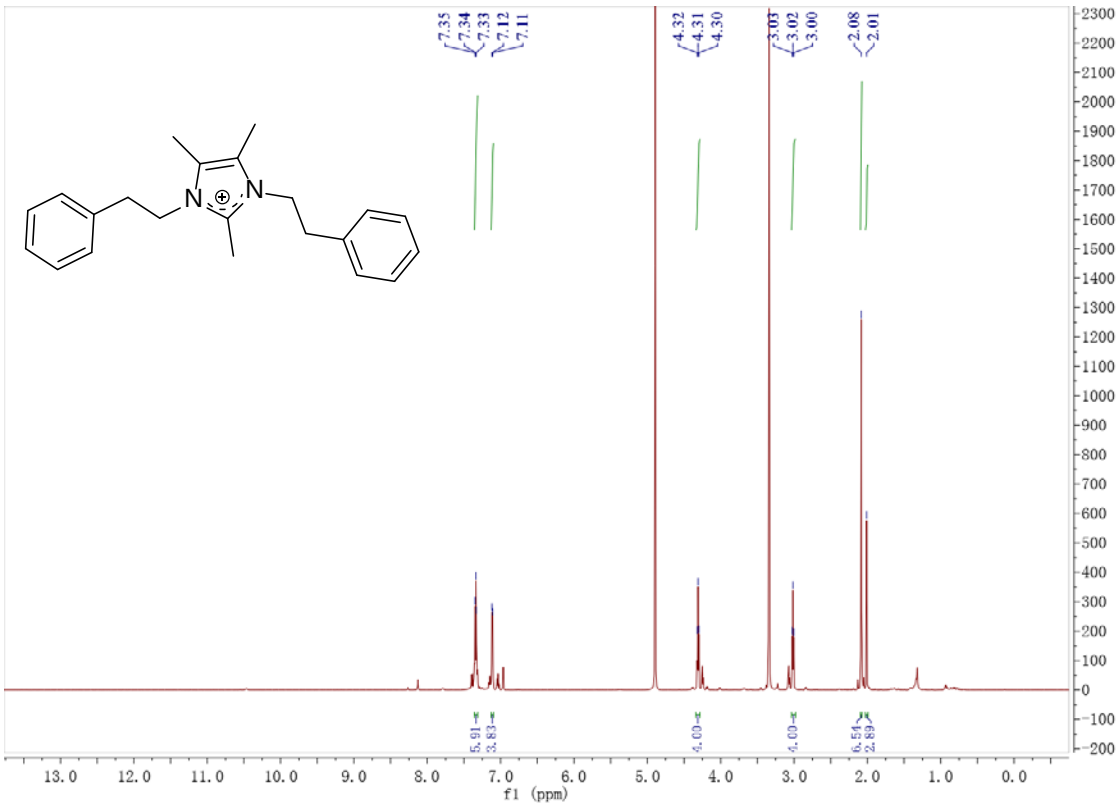


Figure S23. ^{13}C NMR Spectrum of Bacillimidazole B (**22**, 125 MHz, CD_3OD)

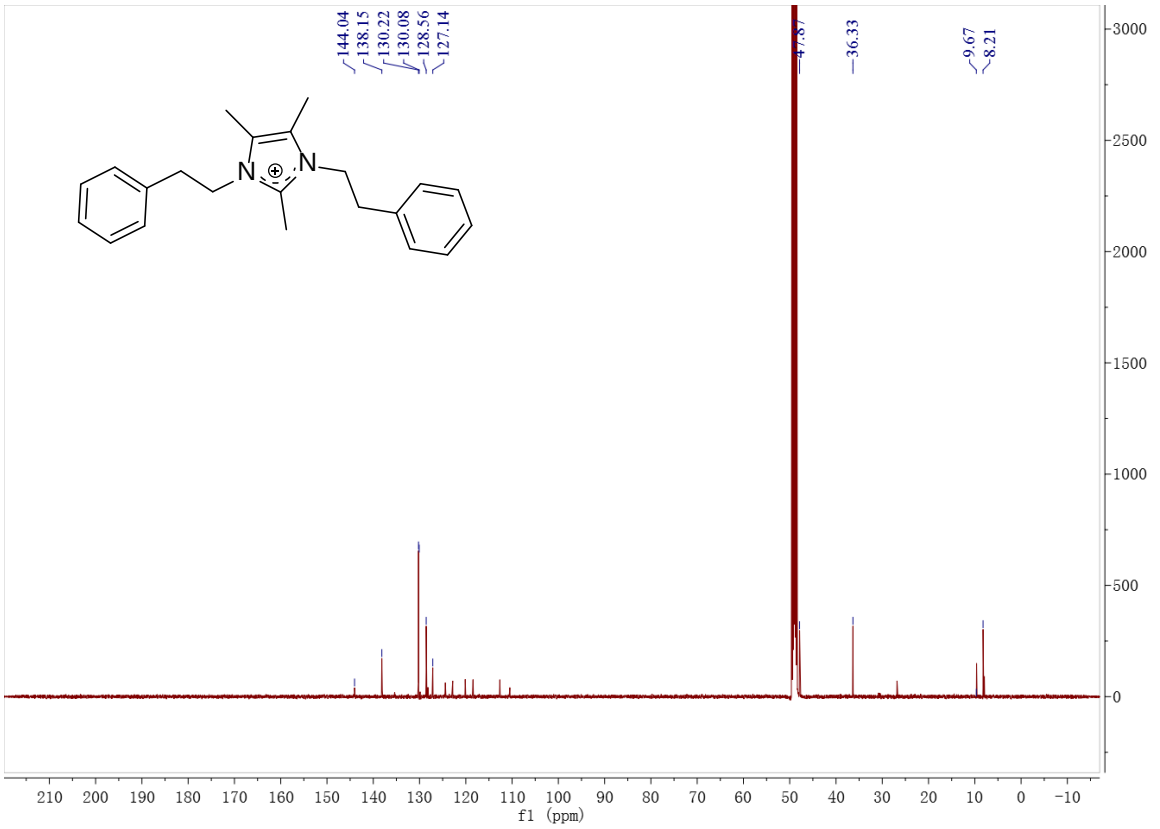


Figure S24. gCOSY Spectrum of Bacillimidazole B (**22**, 600 MHz, CD_3OD)

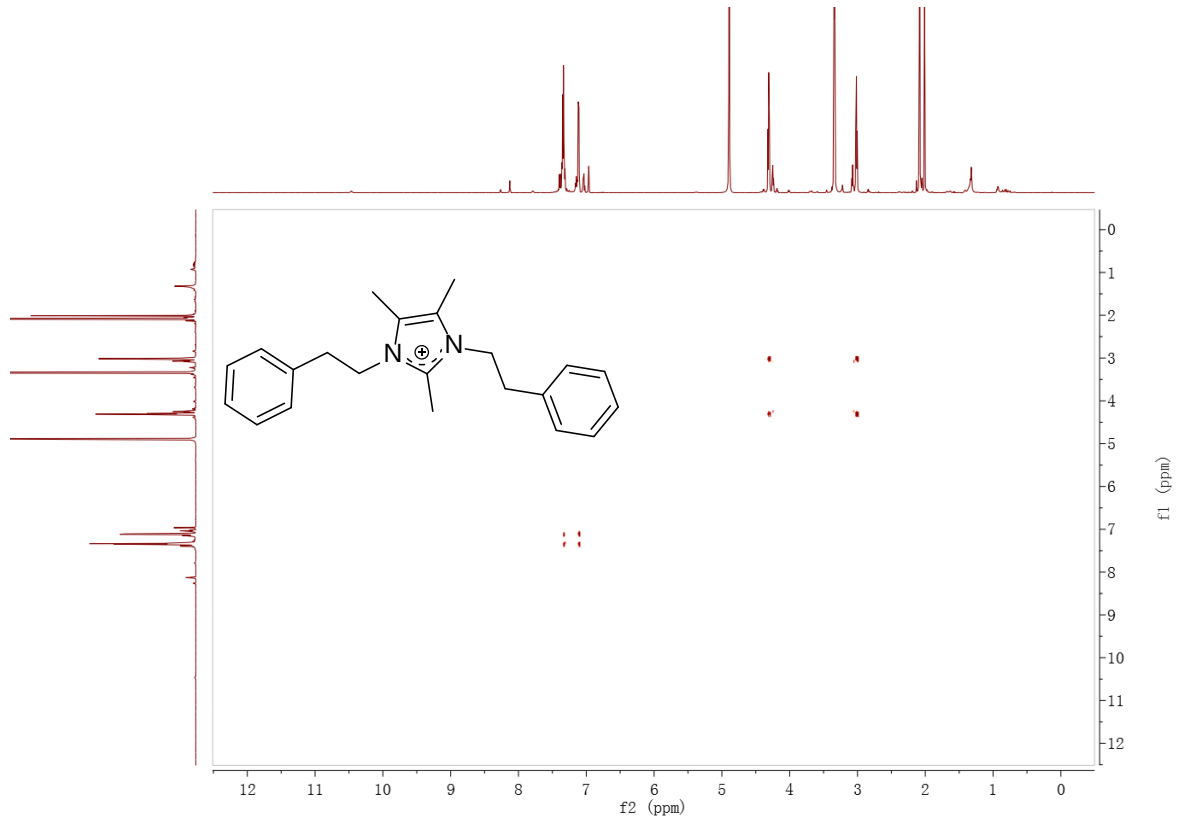


Figure S25. gHSQC Spectrum of Bacillimidazole B (**22**, 600 MHz, CD_3OD)

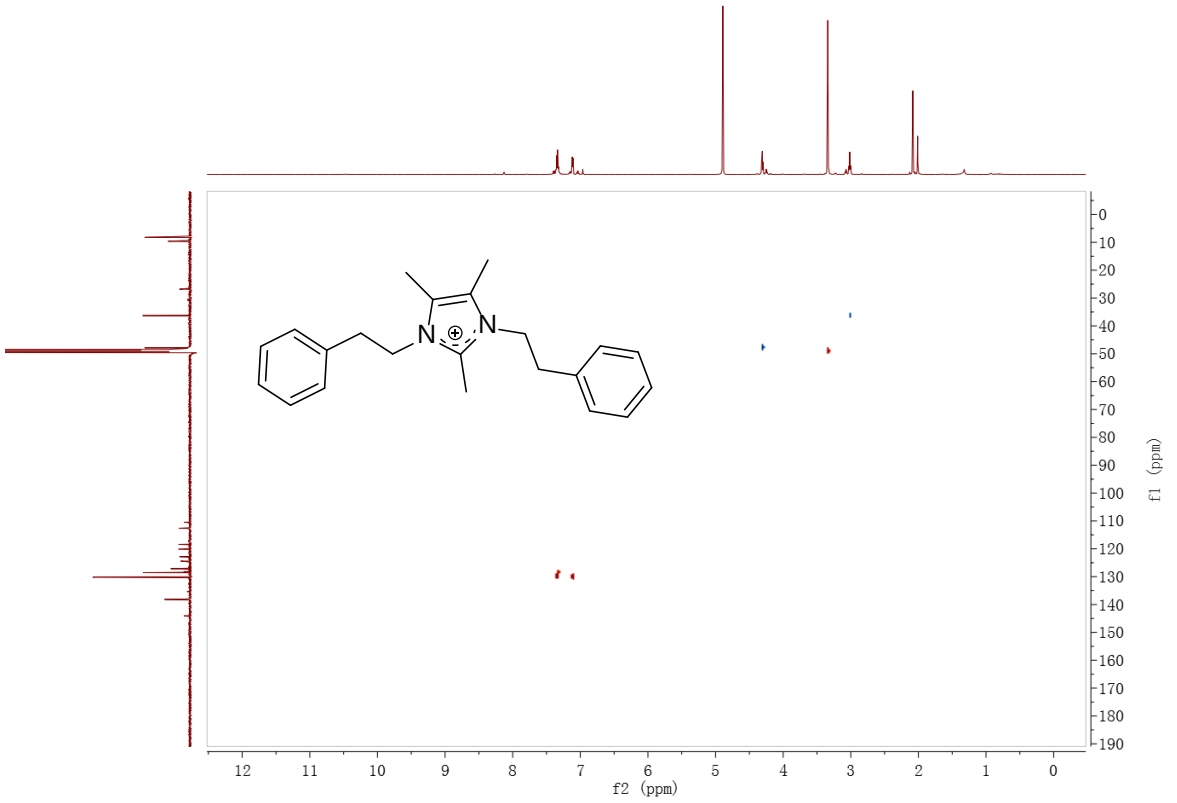


Figure S26. gHMBC Spectrum of Bacillimidazole B (**22**, 600 MHz, CD_3OD)

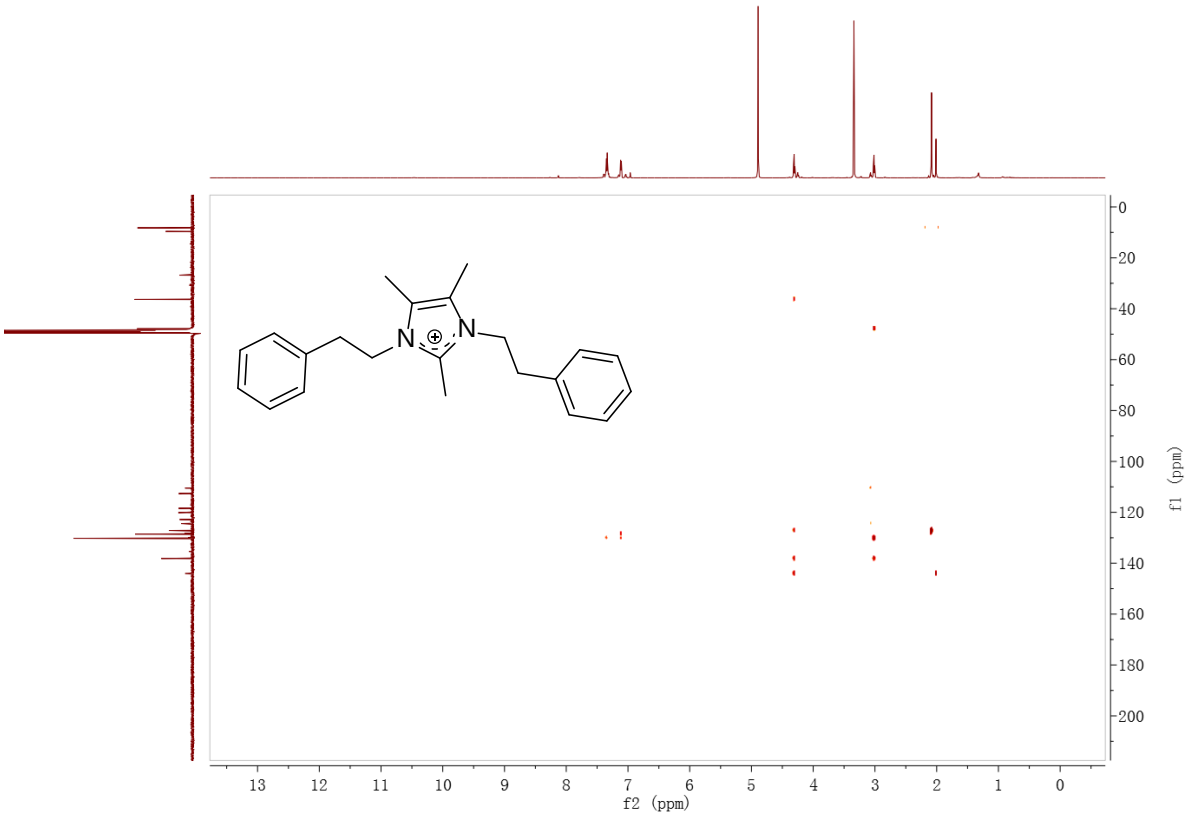


Figure S27. ^{1}H NMR Spectrum of Bacillimidazole C (**23**, 600 MHz, CD_3OD)

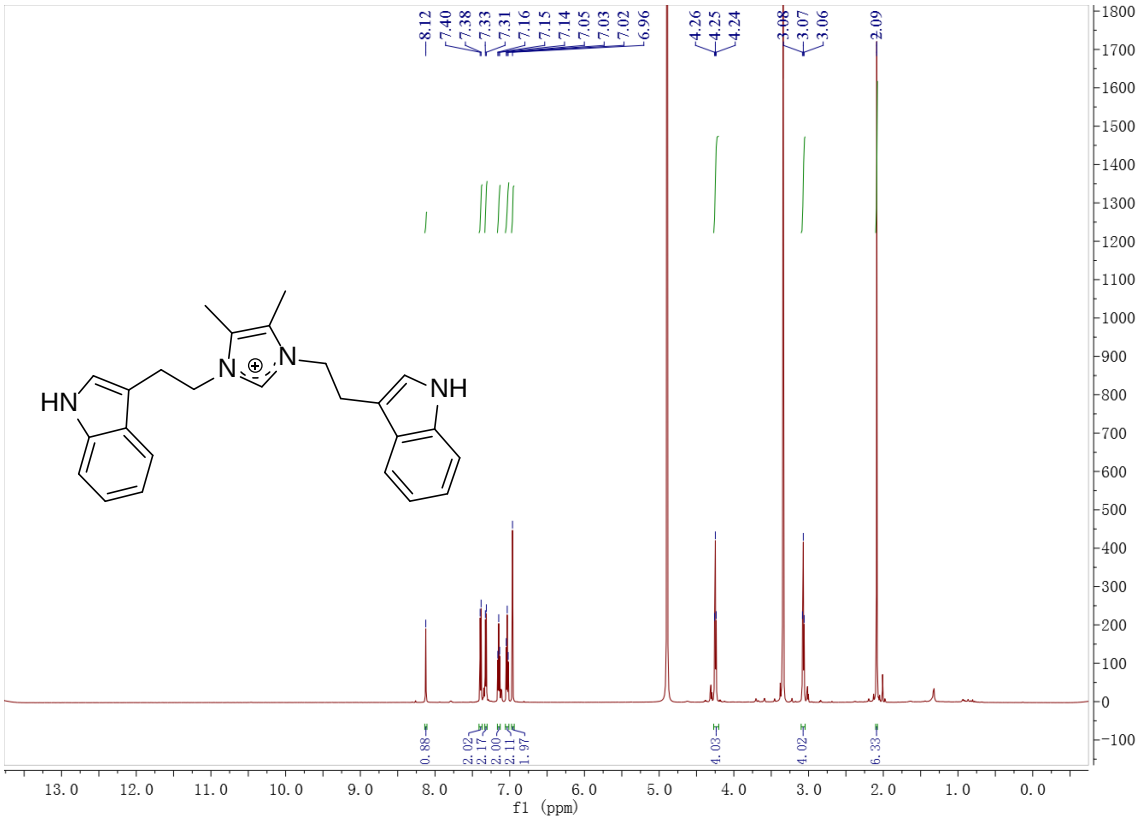


Figure S28. ^{13}C NMR Spectrum of Bacillimidazole C (**23**, 125 MHz, CD_3OD)

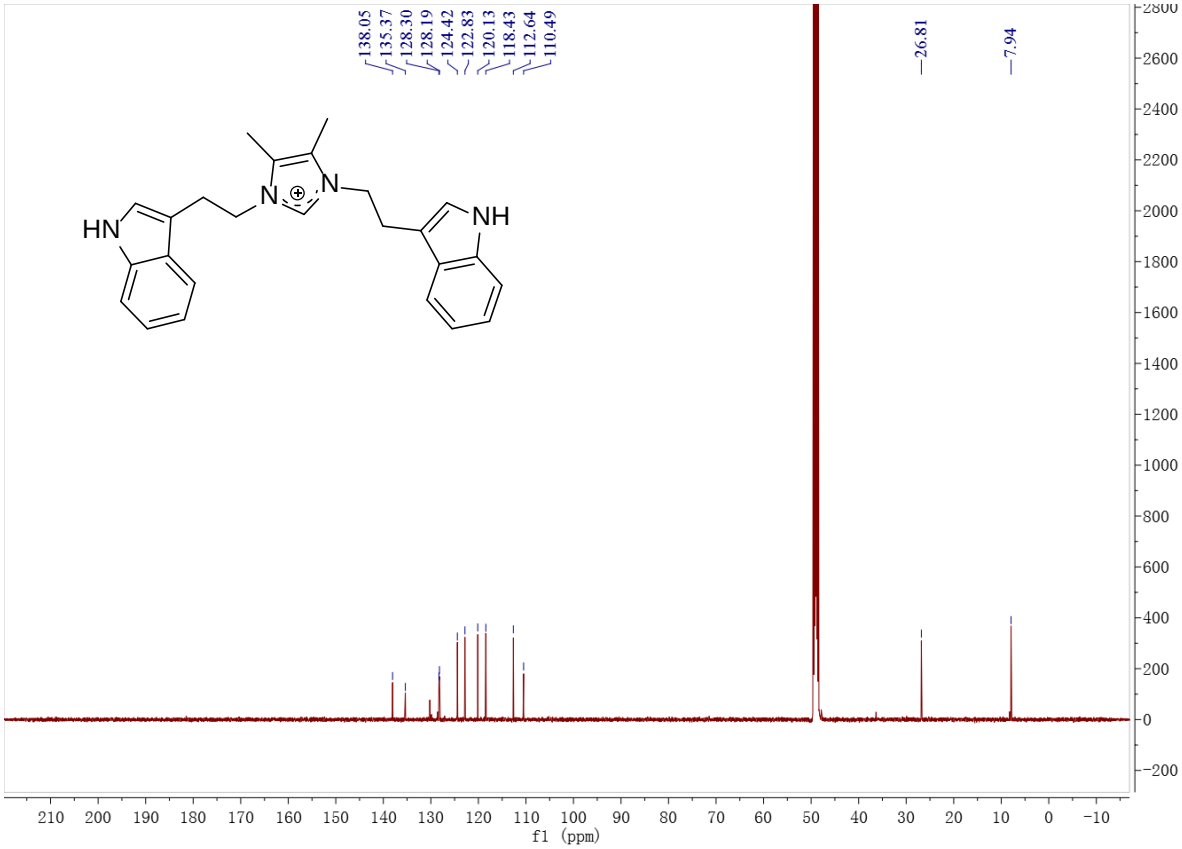


Figure S29. gCOSY Spectrum of Bacillimidazole C (**23**, 600 MHz, CD_3OD)

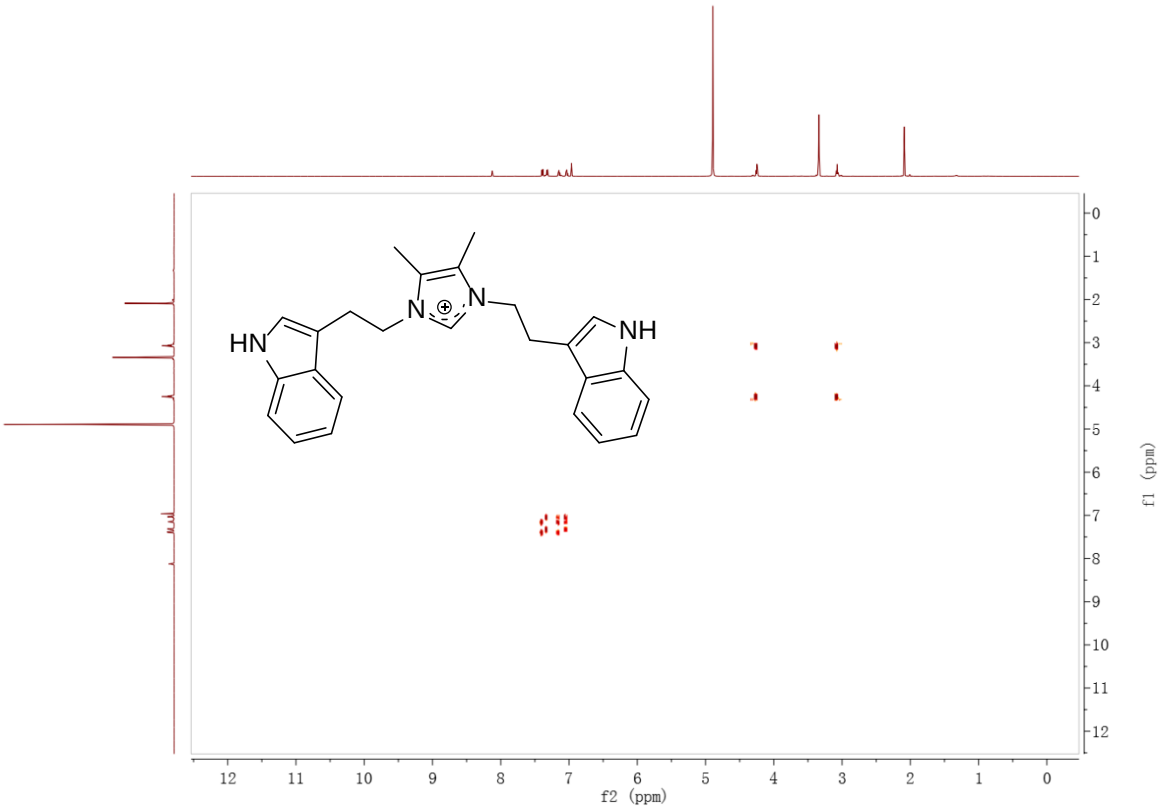


Figure S30. gHSQC Spectrum of Bacillimidazole C (**23**, 600 MHz, CD_3OD)

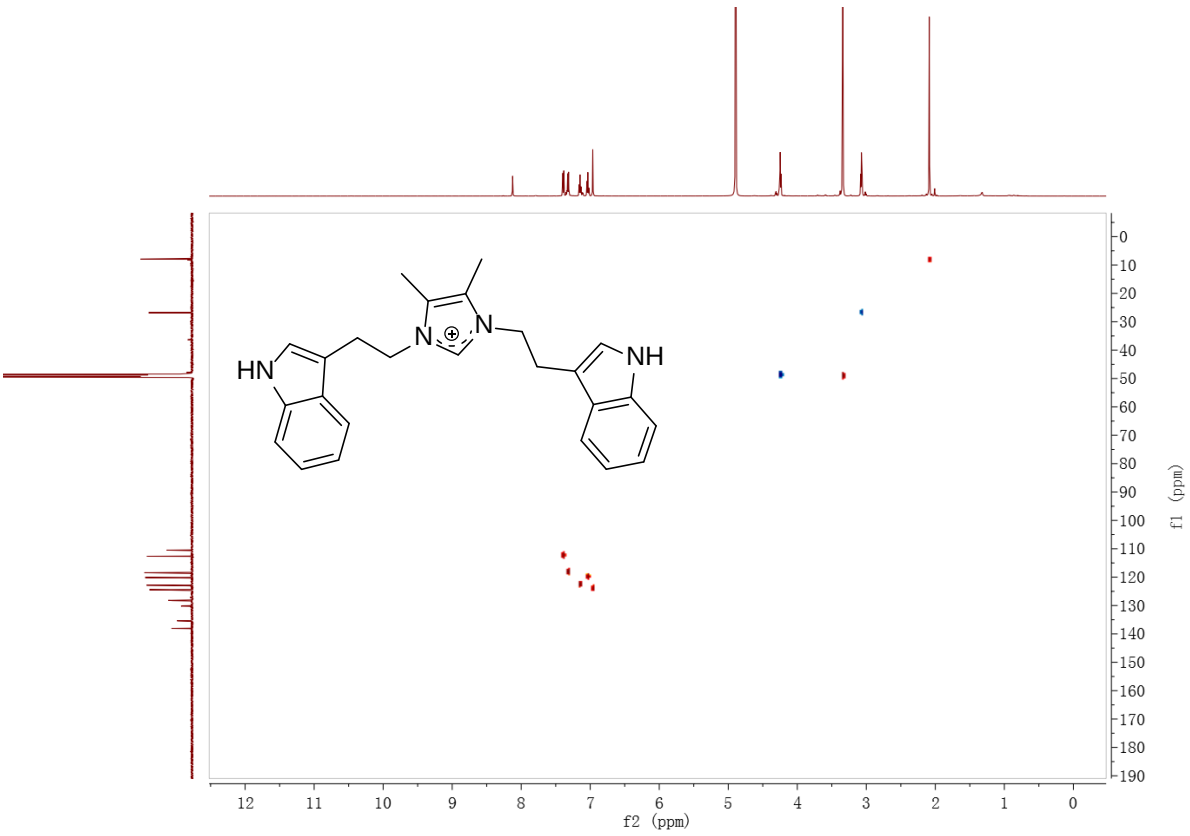


Figure S31. gHMBC Spectrum of Bacillimidazole C (**23**, 600 MHz, CD_3OD)

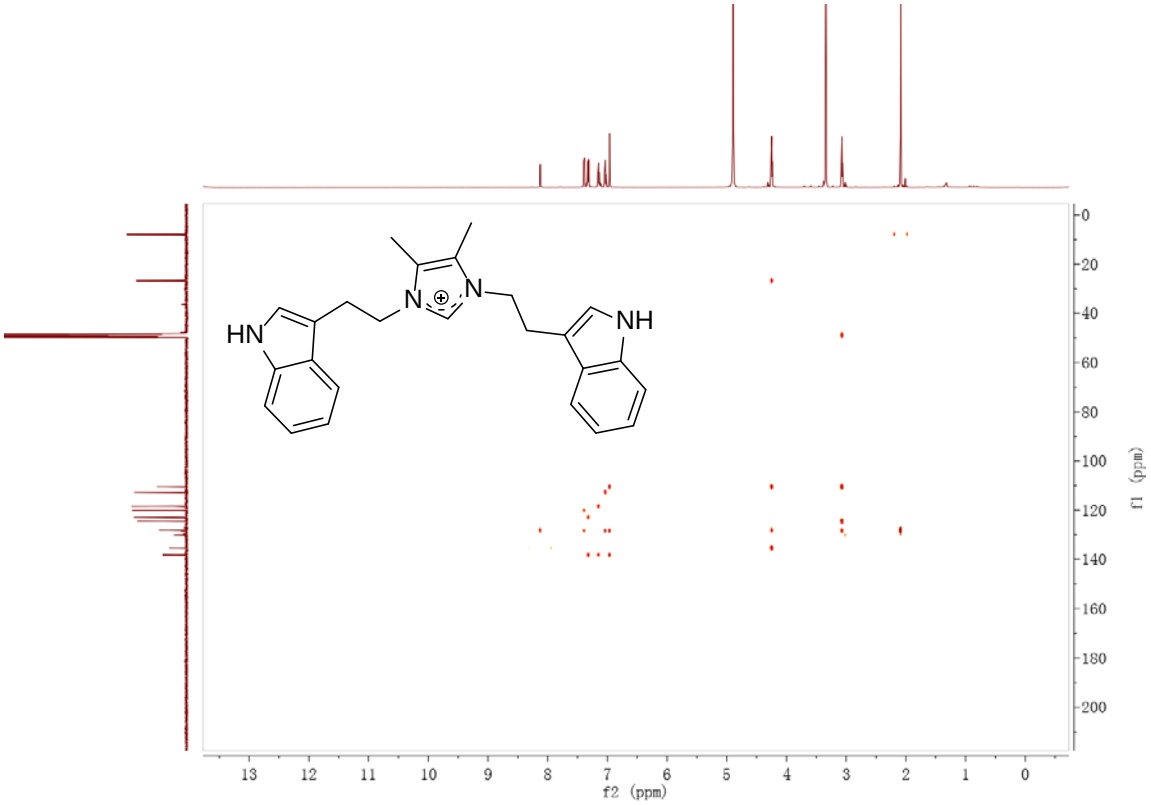


Figure S32. ^1H NMR Spectrum of Bacillimidazole D (**24**, 500 MHz, CD_3OD)

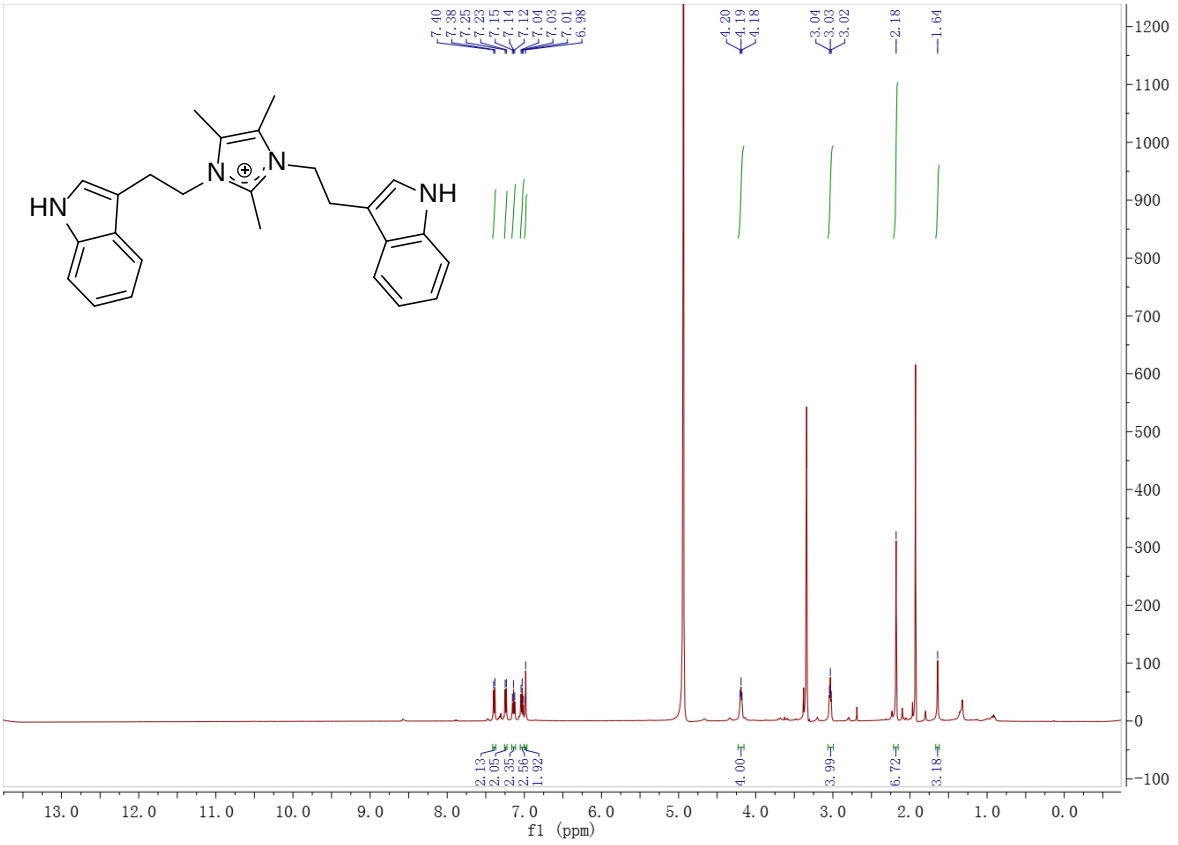


Figure S33. ^{13}C NMR Spectrum of Bacillimidazole D (**24**, 125 MHz, CD_3OD)

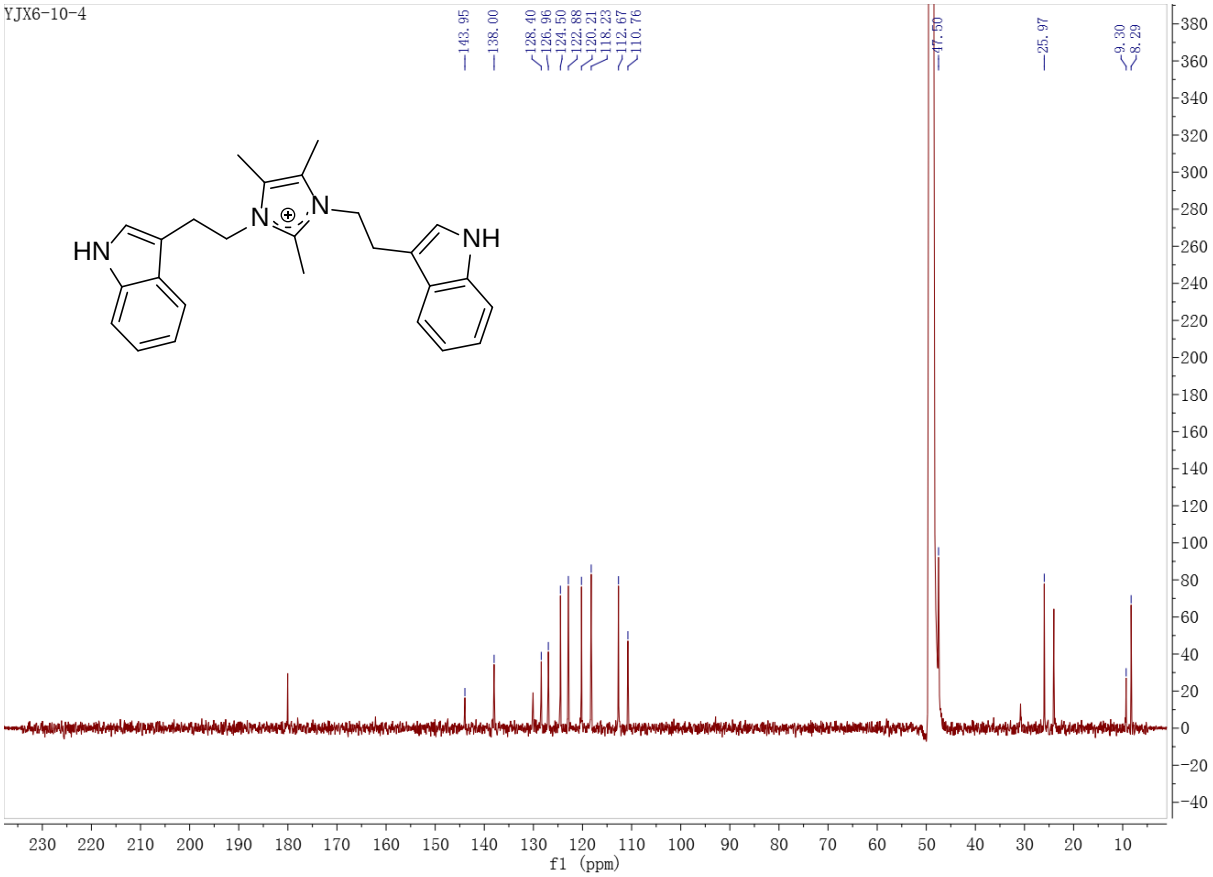


Figure S34. gCOSY Spectrum of Bacillimidazole D (**24**, 500 MHz, CD_3OD)

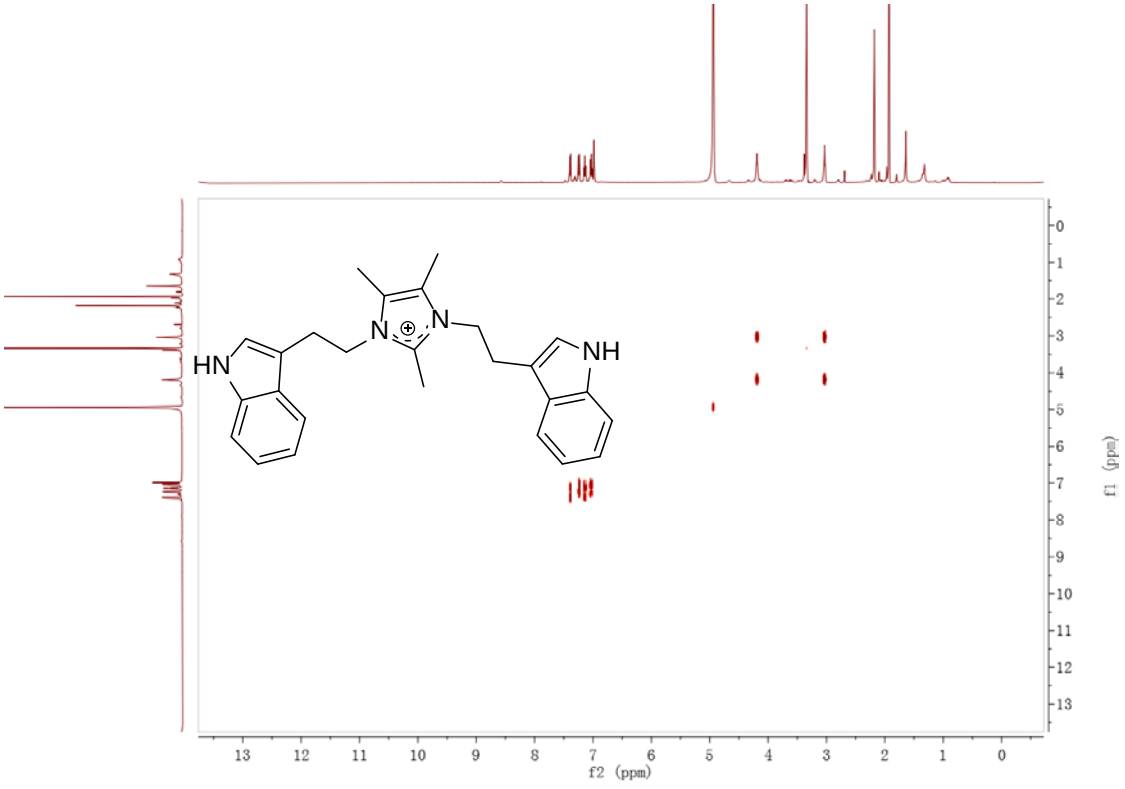


Figure S35. gHSQC Spectrum of Bacillimidazole D (**24**, 500 MHz, CD_3OD)

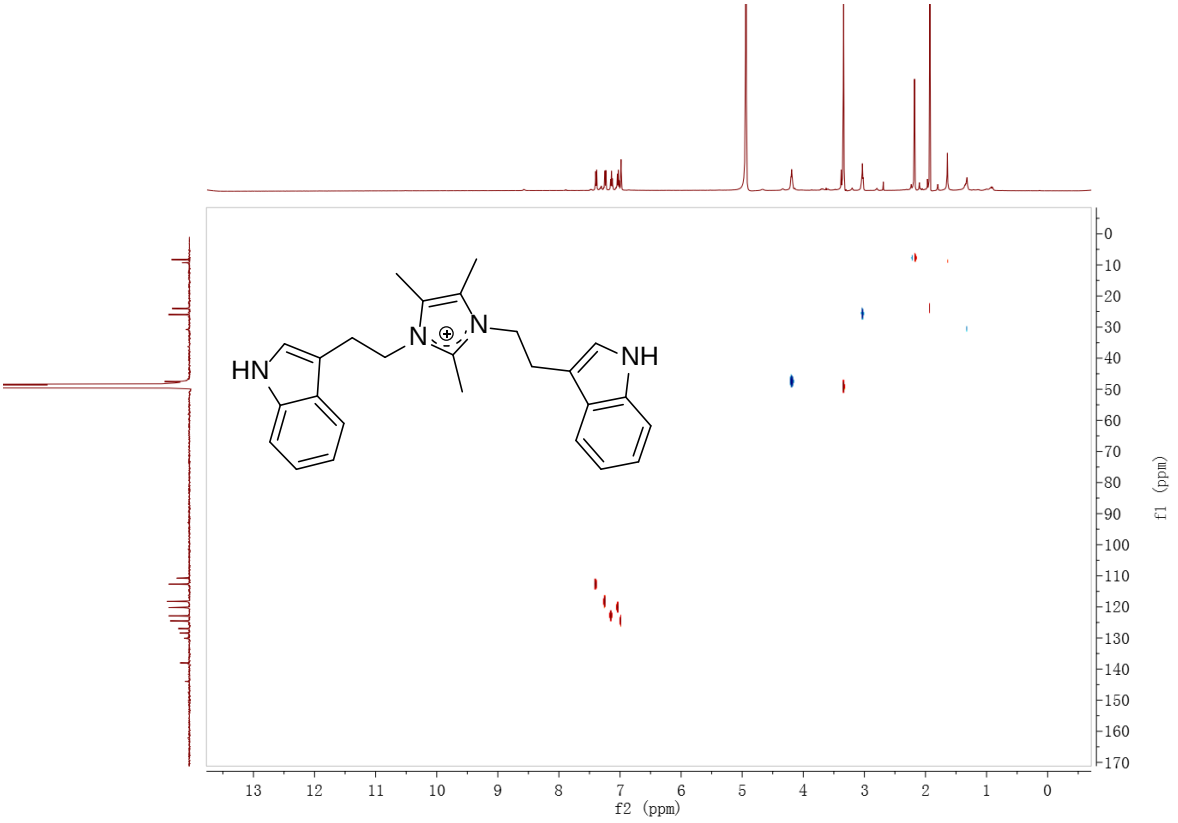


Figure S36. gHMBC Spectrum of Bacillimidazole D (**24**, 500 MHz, CD_3OD)

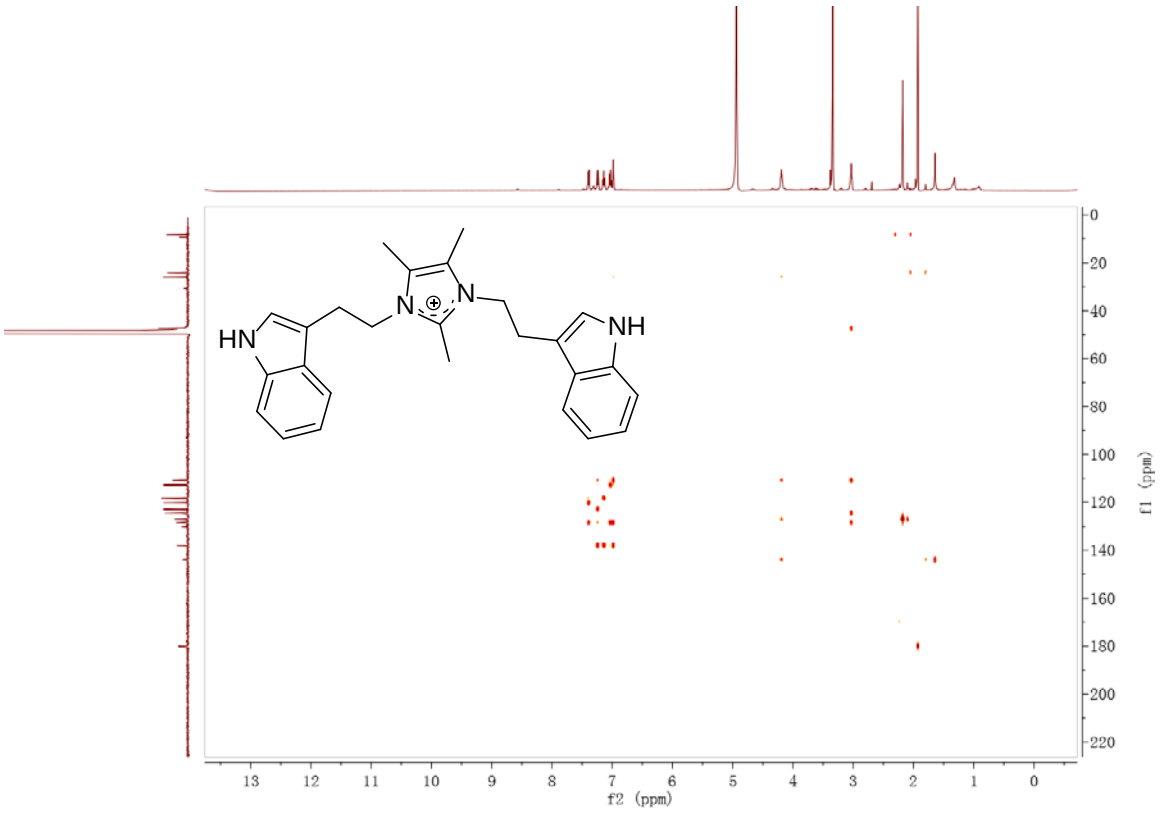


Figure S37. ^{1}H NMR Spectrum of Bacillimidazole E (**25**, 600 MHz, CD_3OD)

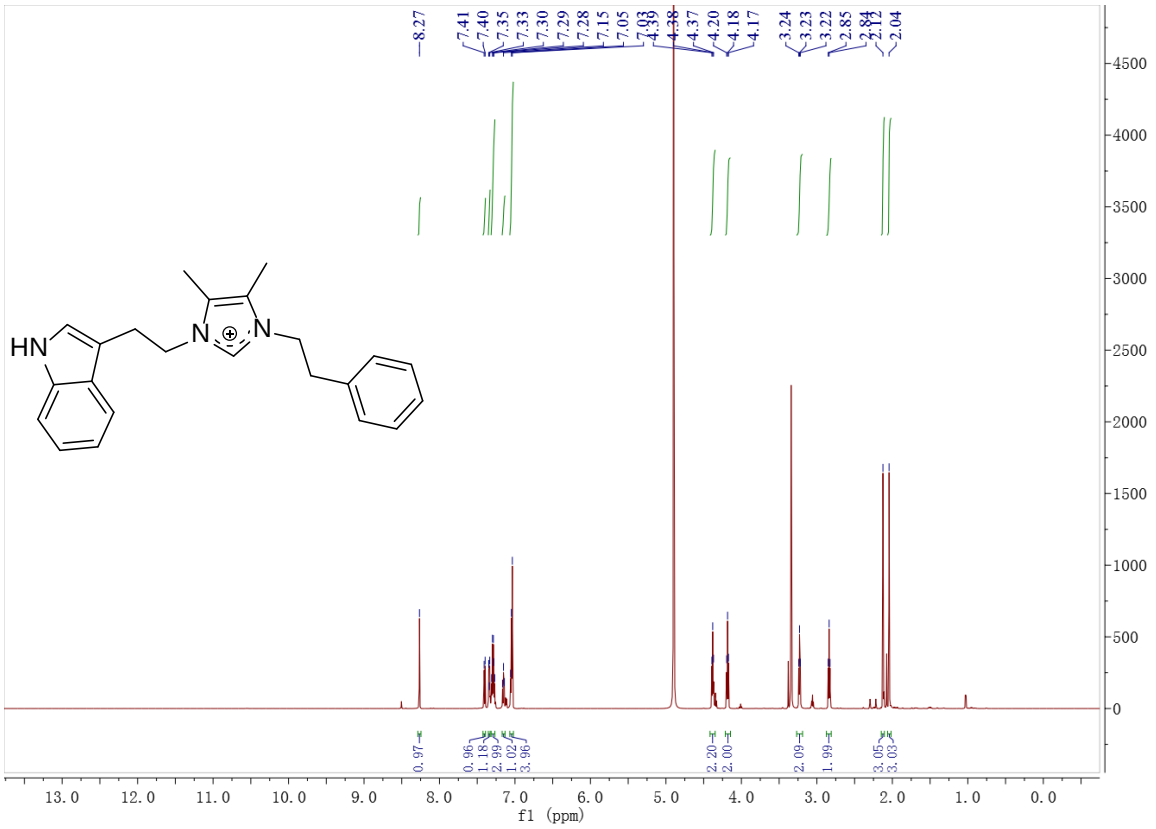


Figure S38. ^{13}C NMR Spectrum of Bacillimidazole E (**25**, 125 MHz, CD_3OD)

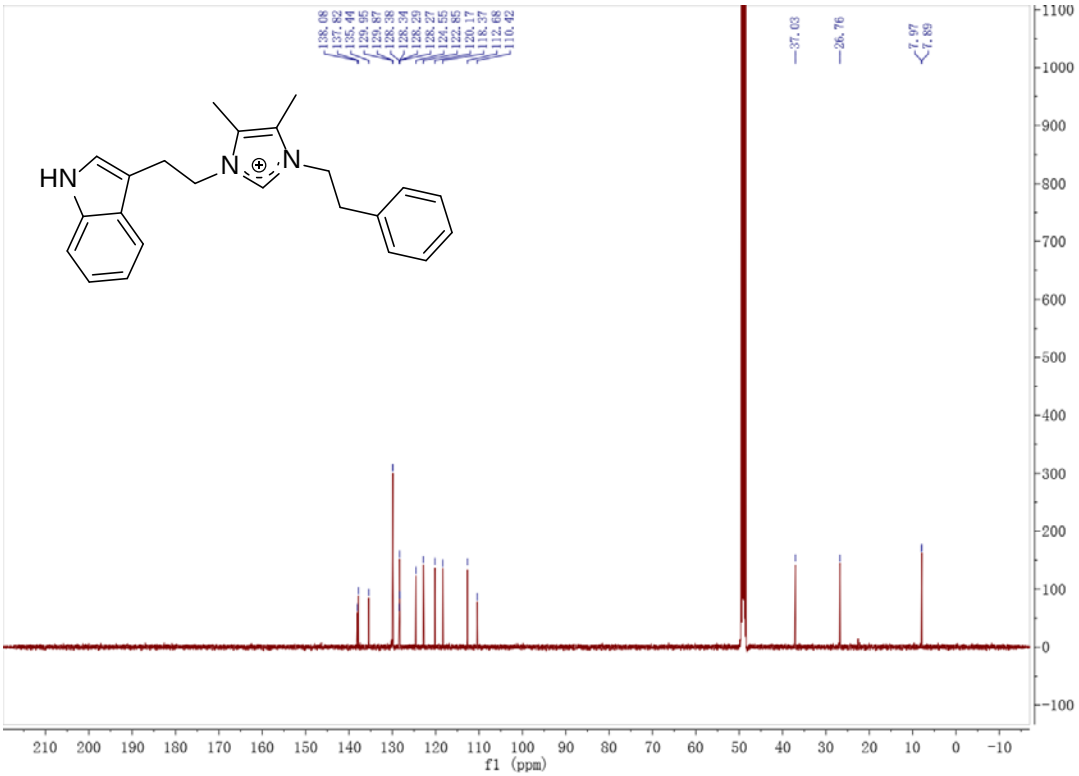


Figure S39. gCOSY Spectrum of Bacillimidazole E (**25**, 600 MHz, CD_3OD)

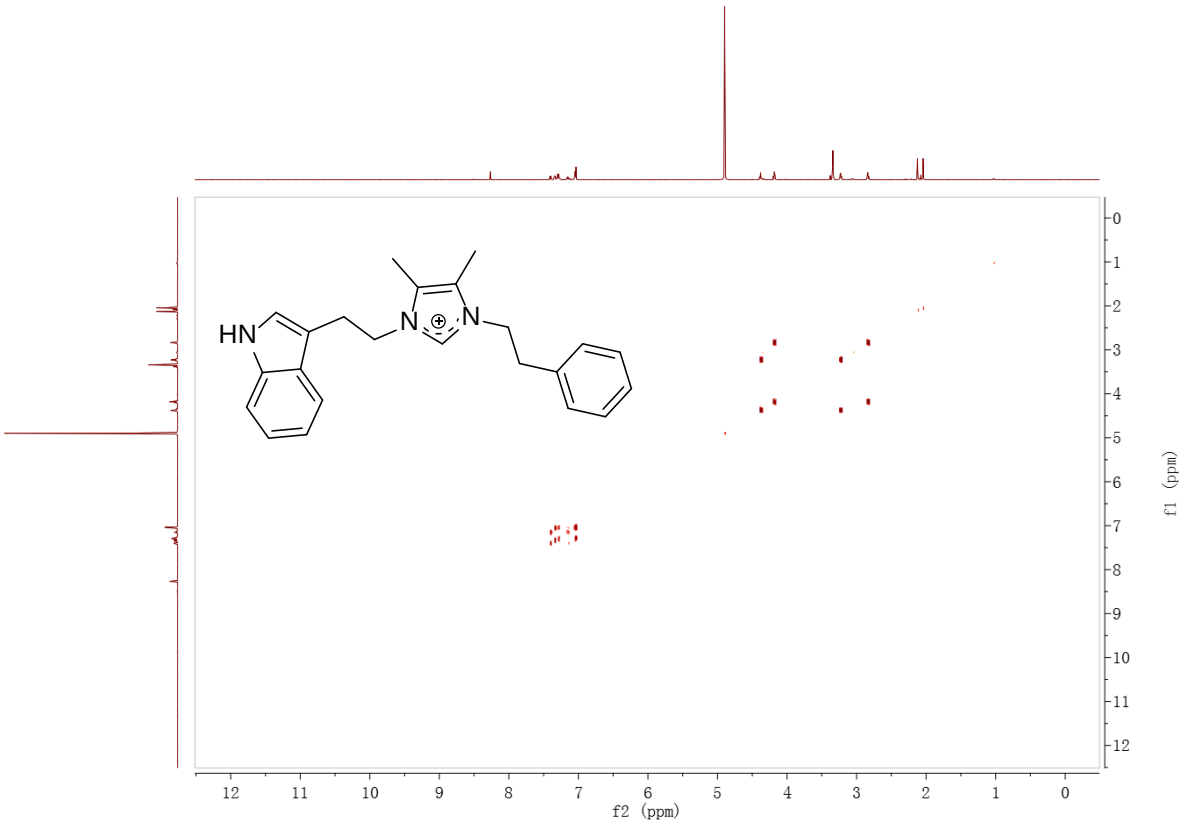


Figure S40. gHSQC Spectrum of Bacillimidazole E (**25**, 600 MHz, CD_3OD)

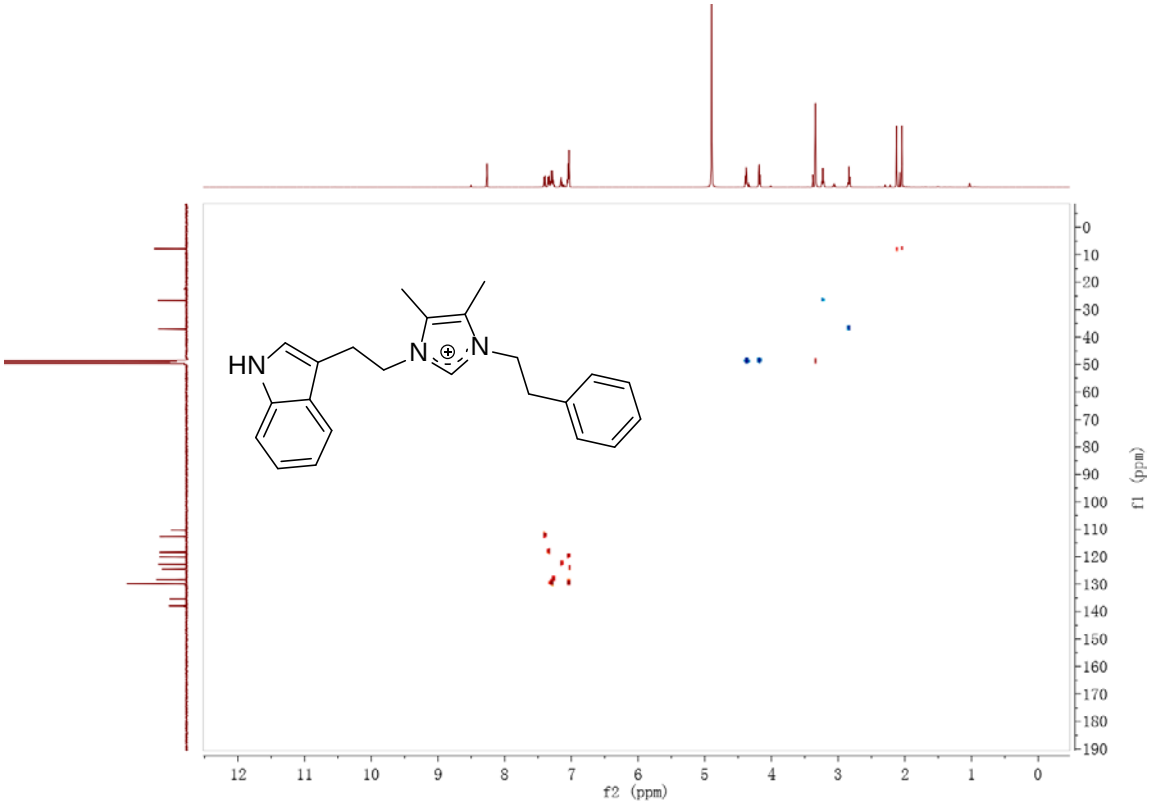


Figure S41. gHMBC Spectrum of Bacillimidazole E (**25**, 600 MHz, CD_3OD)

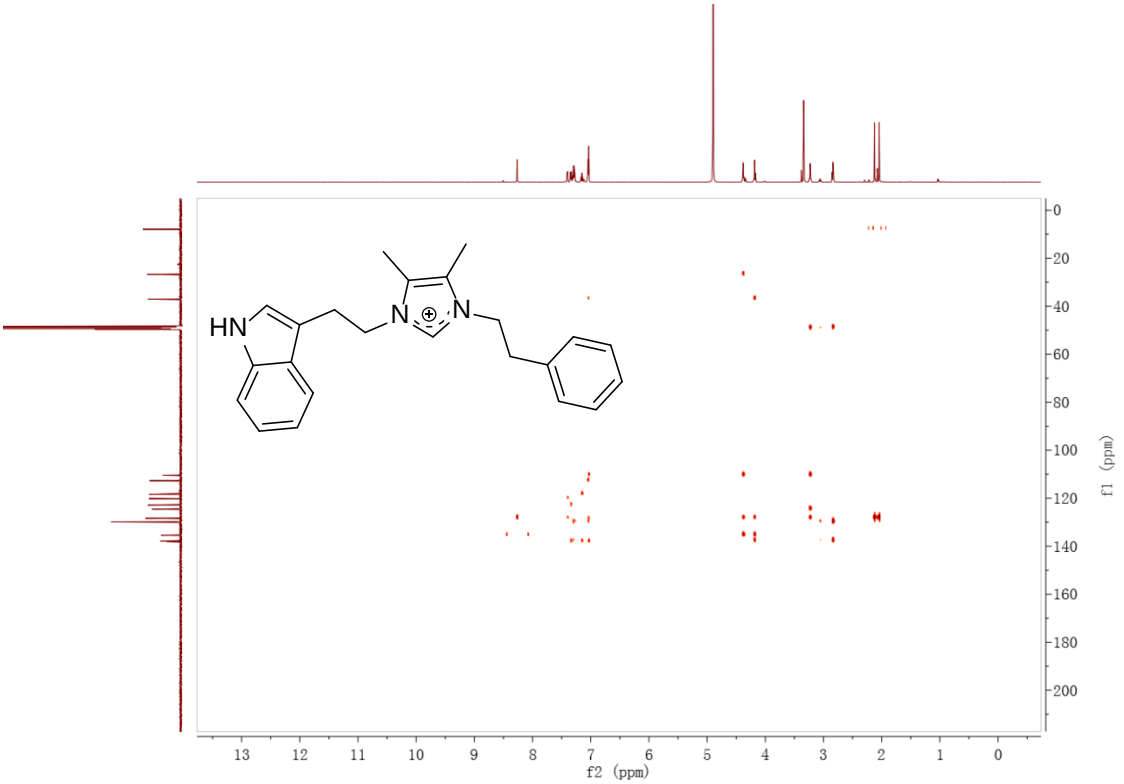


Figure S42. ^1H NMR Spectrum of Bacillimidazole F (**26**, 500 MHz, CD_3OD)

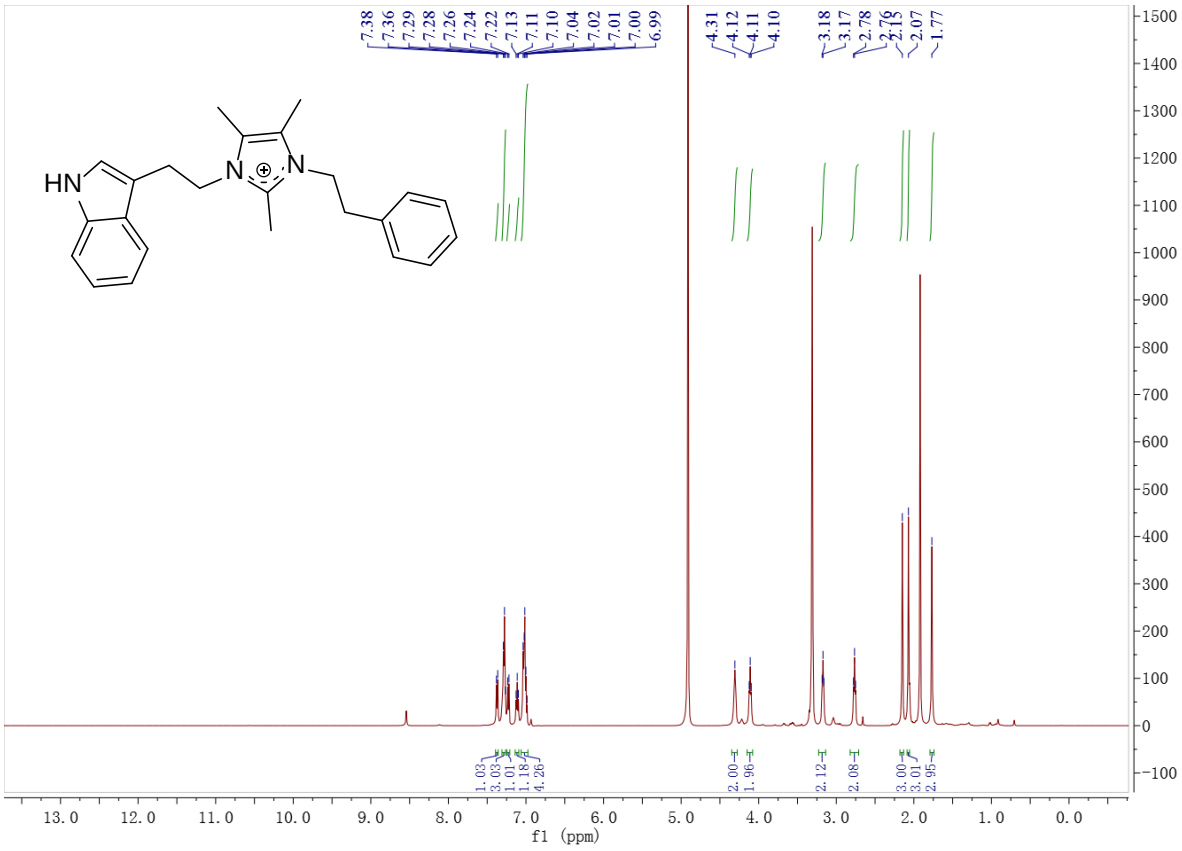


Figure S43. ^{13}C NMR Spectrum of Bacillimidazole F (**26**, 125 MHz, CD_3OD)

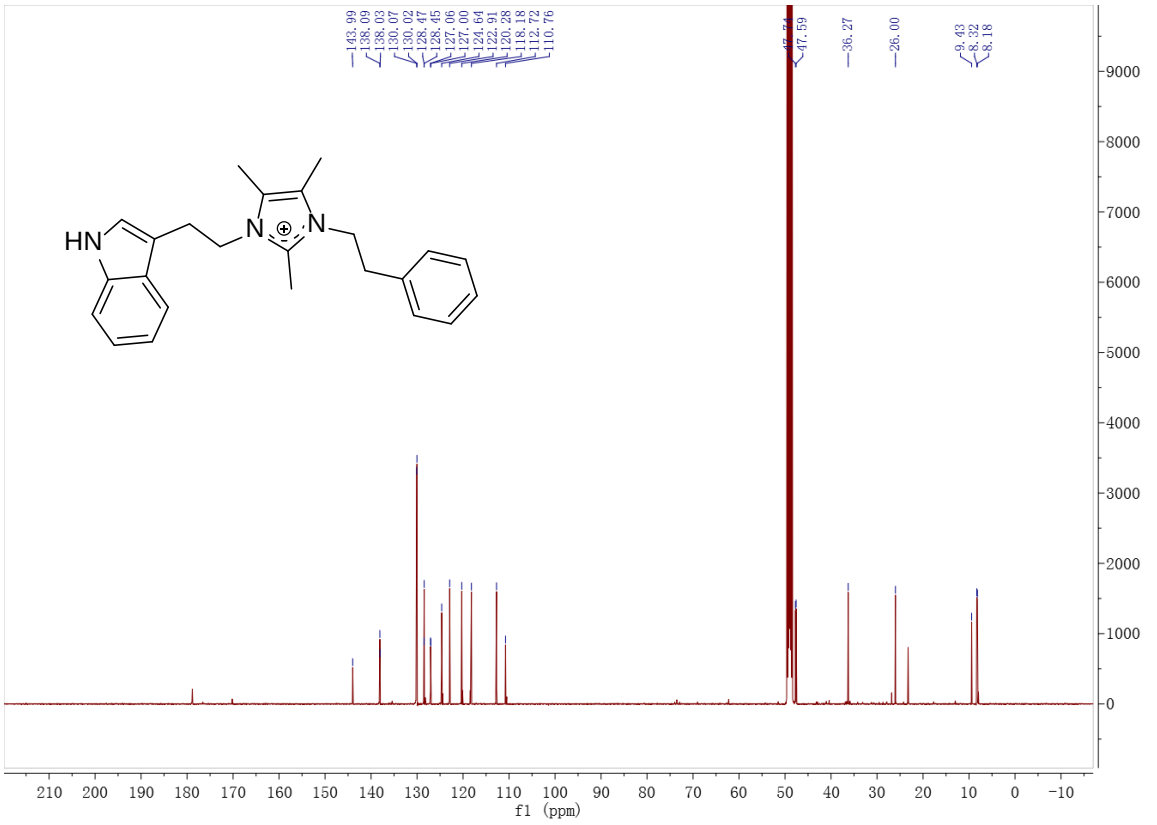


Figure S44. gCOSY Spectrum of Bacillimidazole F (**26**, 500 MHz, CD_3OD)

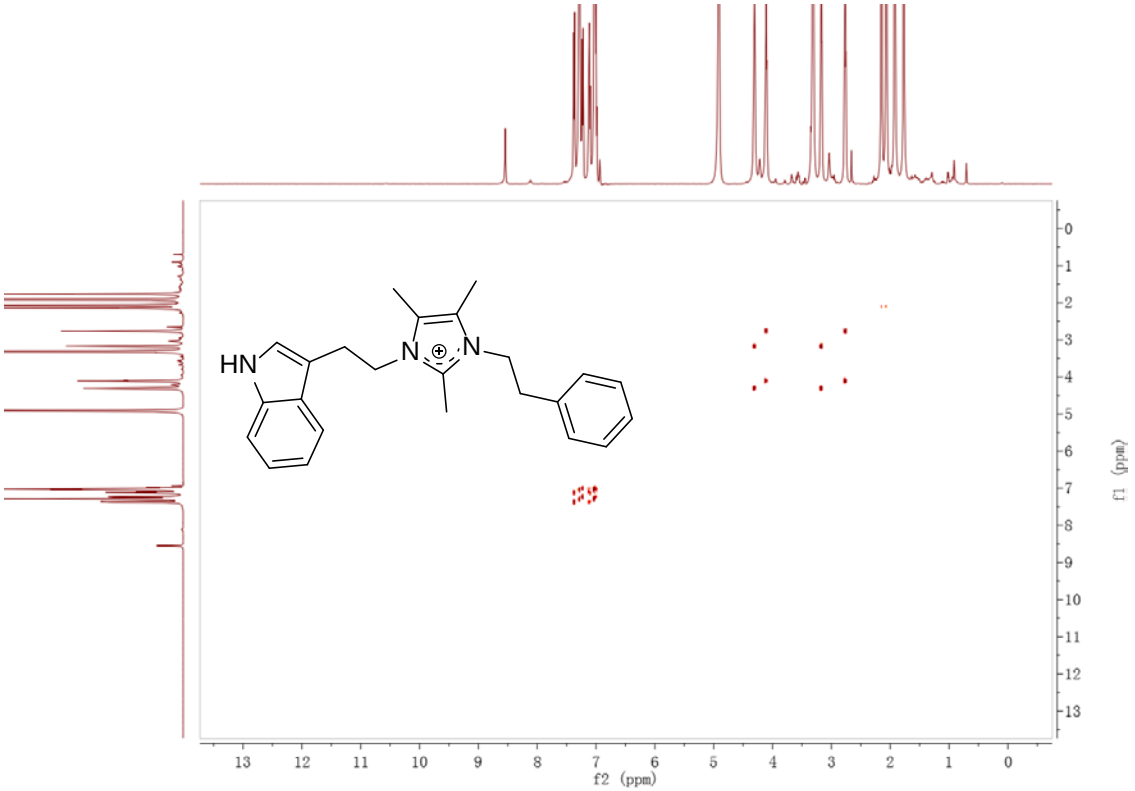


Figure S45. gHSQC Spectrum of Bacillimidazole F (**26**, 500 MHz, CD_3OD)

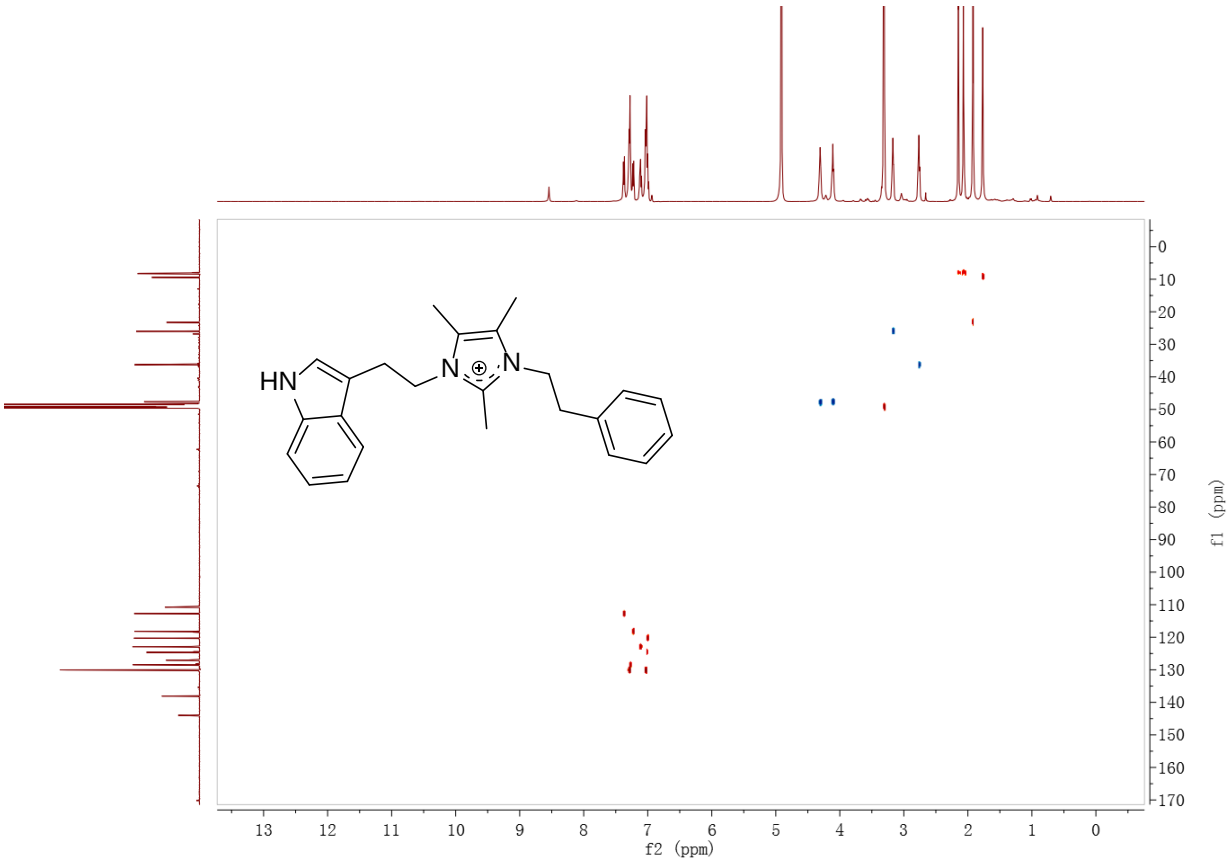


Figure S46. gHMBC Spectrum of Bacillimidazole F (**26**, 500 MHz, CD_3OD)

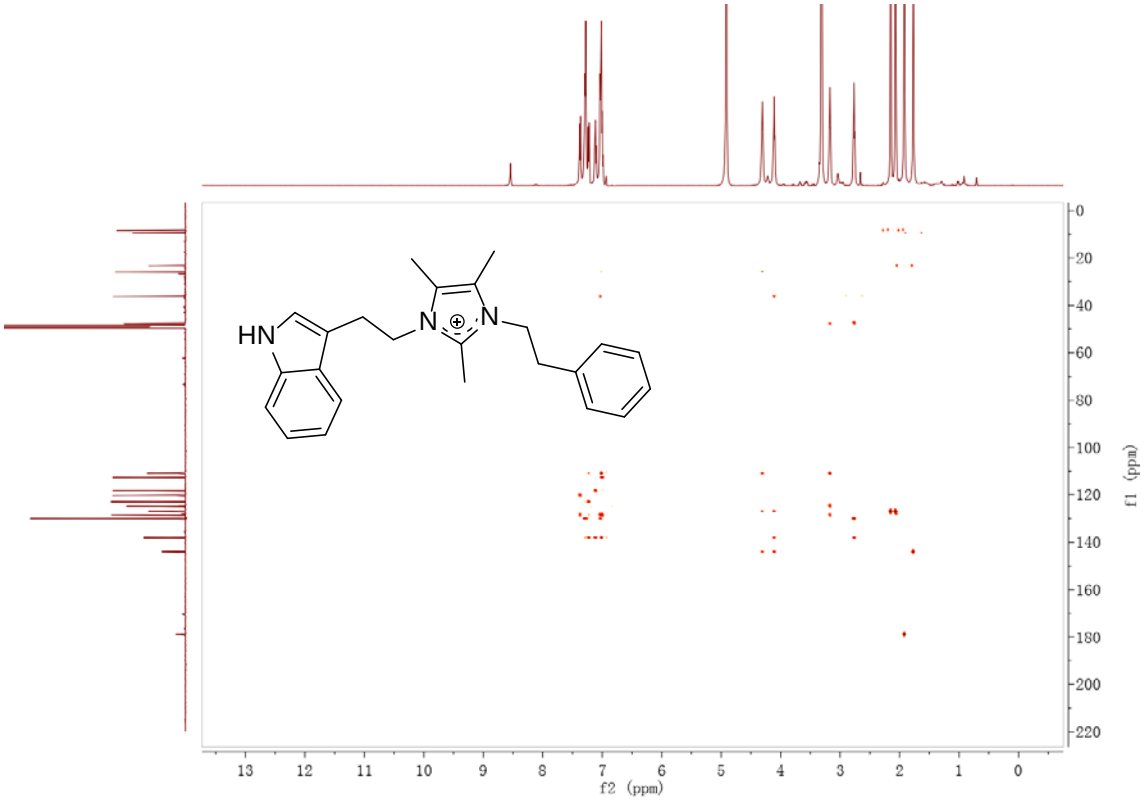


Figure S47. Positive Ion HRESIMS of Bacillimidazole A (**21**)

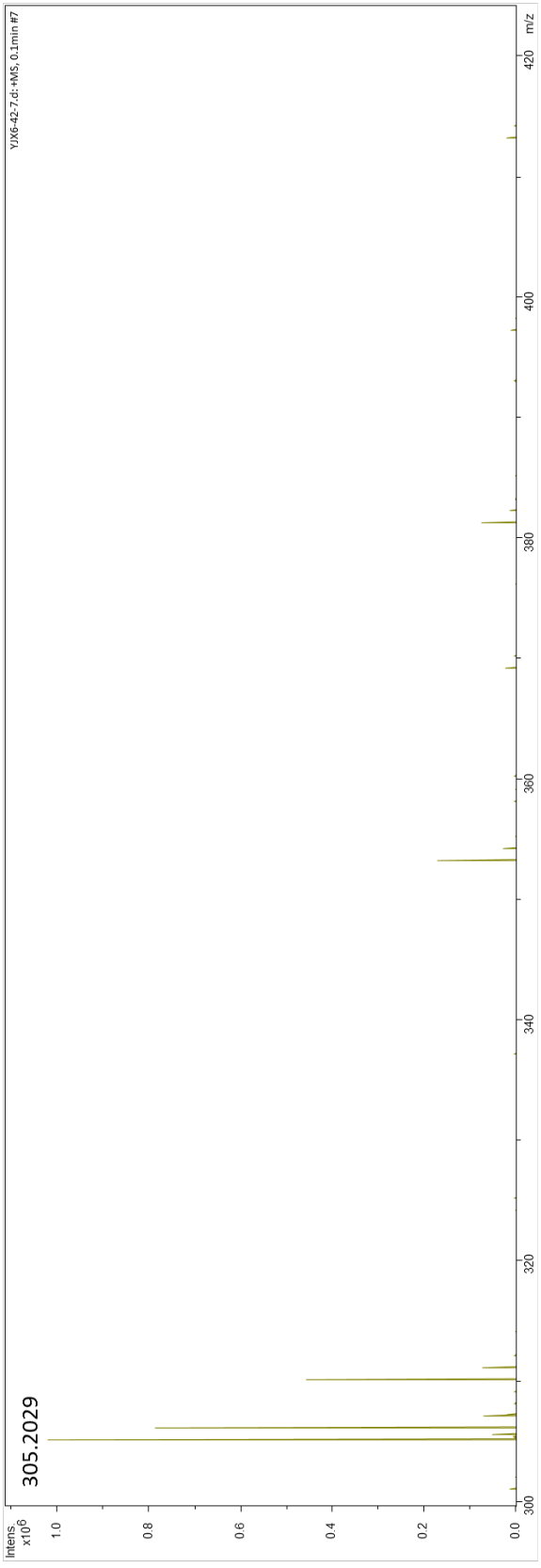


Figure S48. Positive Ion HRESIMS of Bacillimidazole B (**22**)

Figure S49. Positive Ion HRESIMS of Bacillimidazole C (**23**)

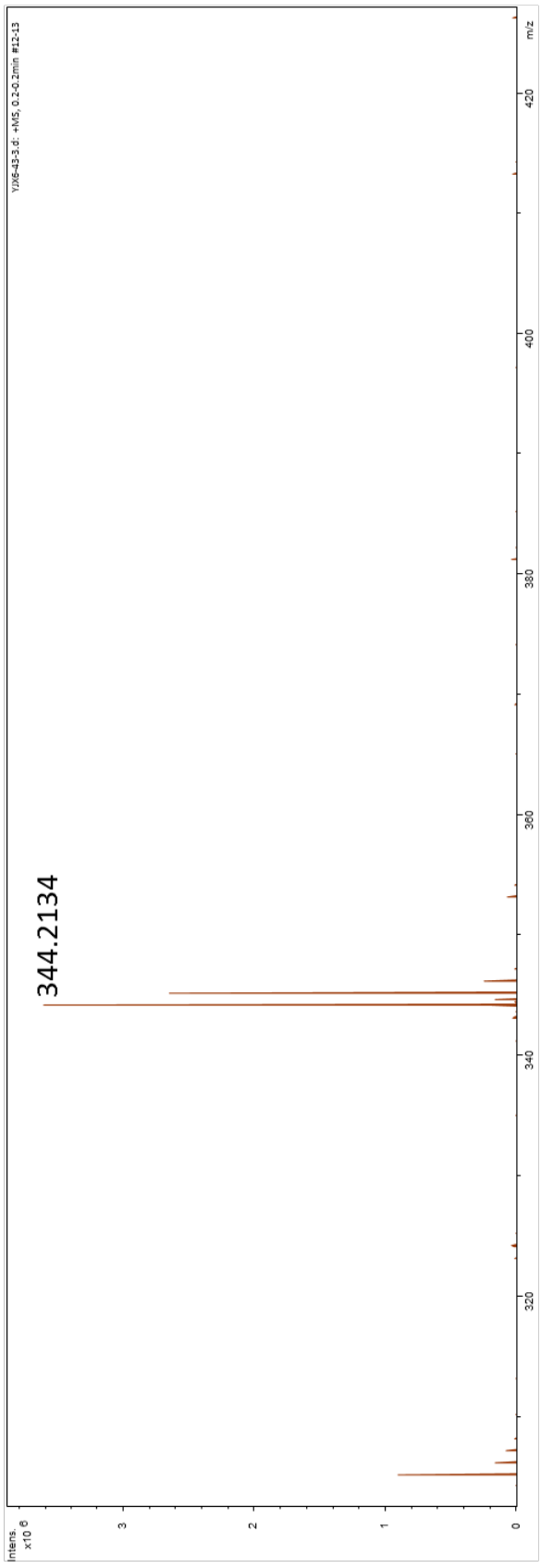


Figure S50. Positive Ion HRESIMS of Bacillimidazole D (**24**)

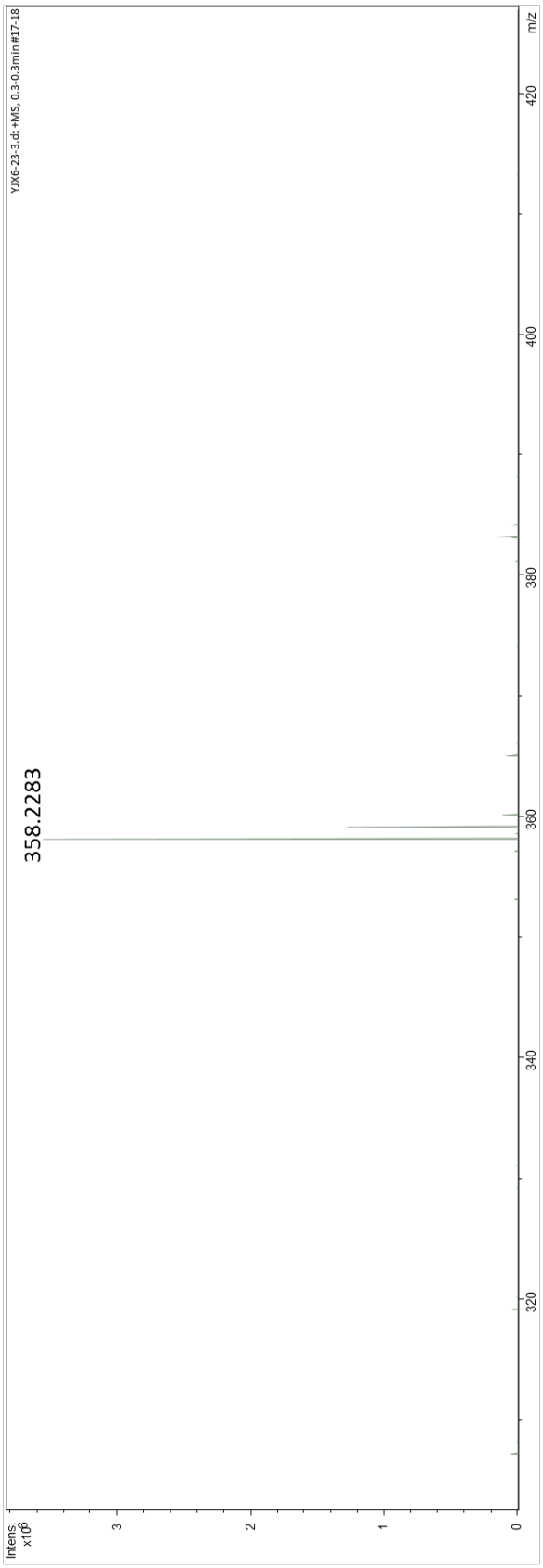


Figure S51. Positive Ion HRESIMS of Bacillimidazole E (**25**)

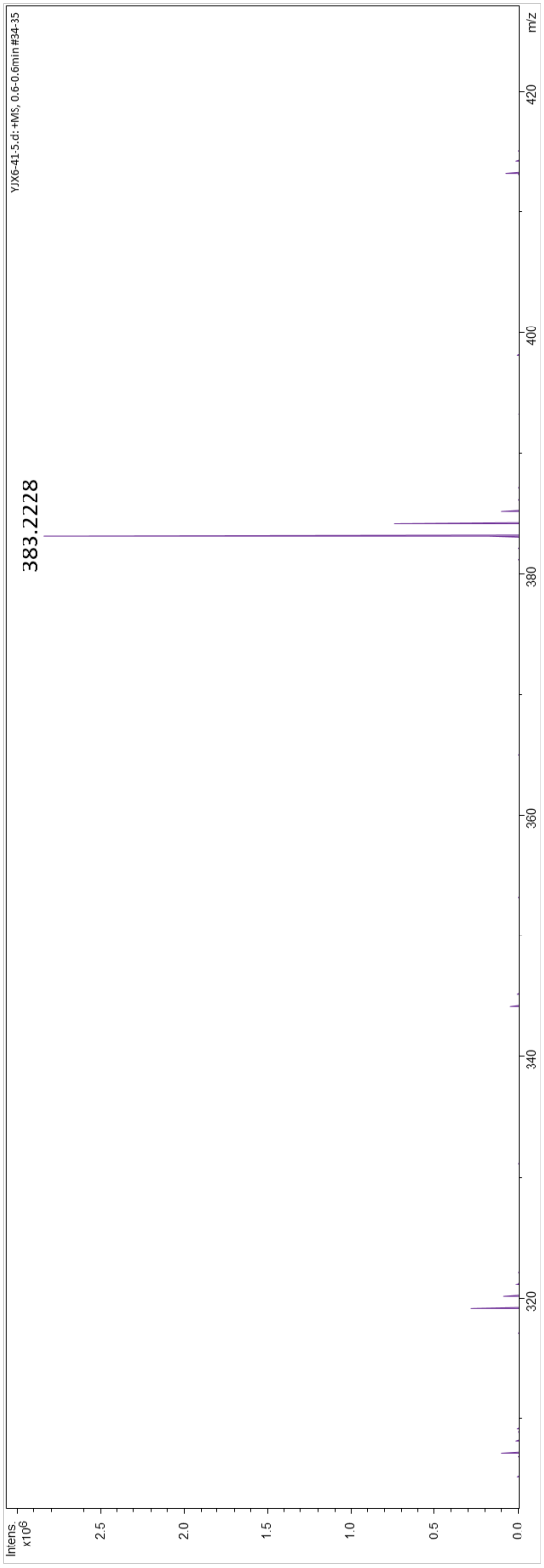


Figure S52. Positive Ion HRESIMS of Bacillimidazole F (**26**)

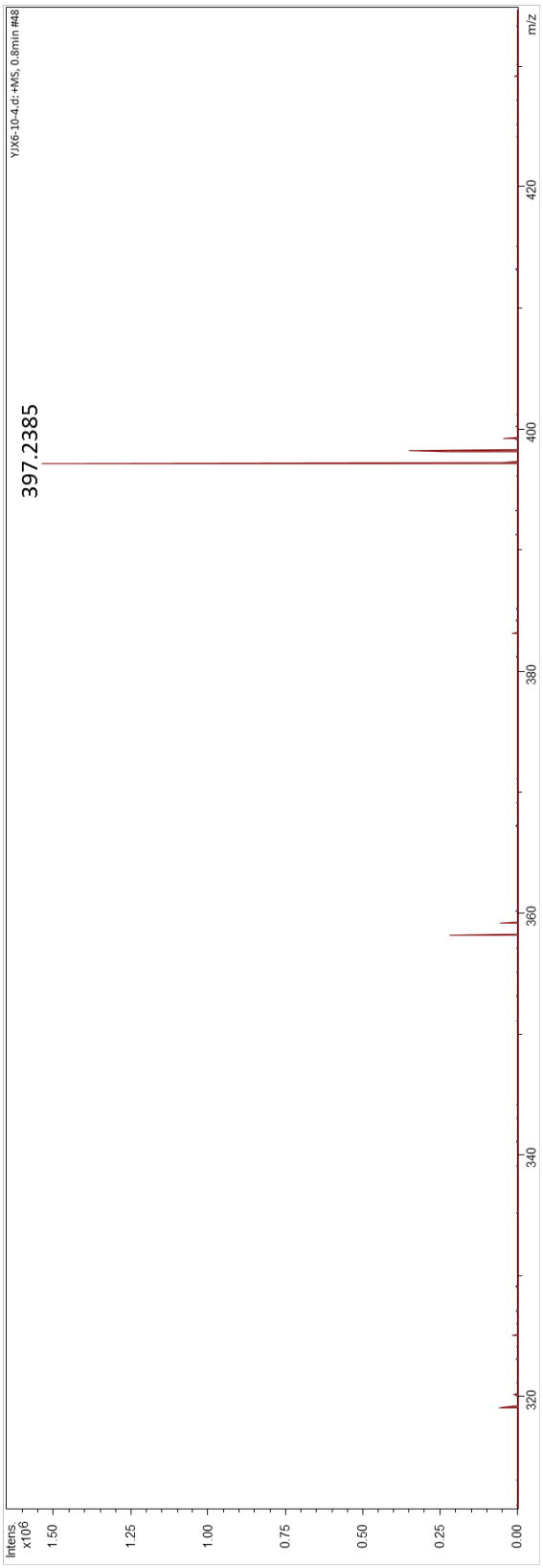


Figure S53. ^{1}H NMR Spectrum of ^{13}C Labeled Bacillimidazole C (**23**, 500 MHz, CD_3OD)

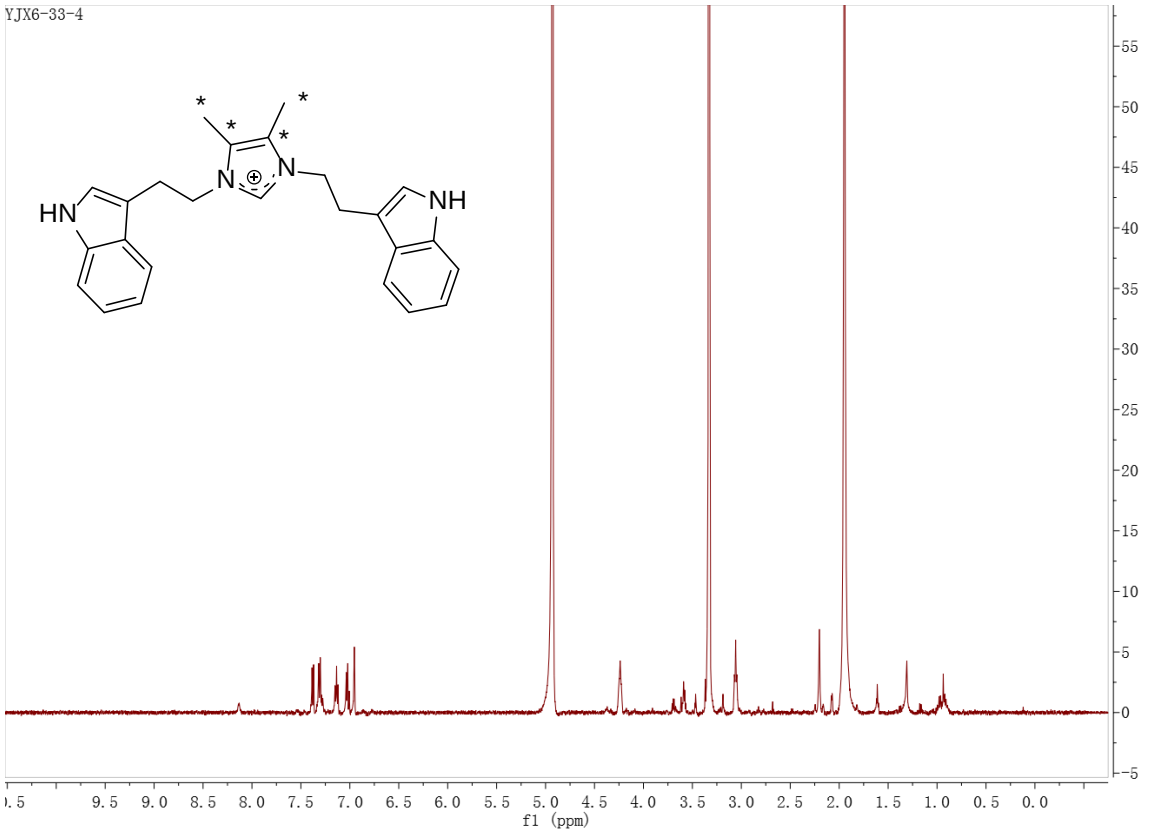


Figure S54. ^{13}C NMR Spectrum of ^{13}C Labeled Bacillimidazole C (**23**, 125 MHz, CD_3OD)

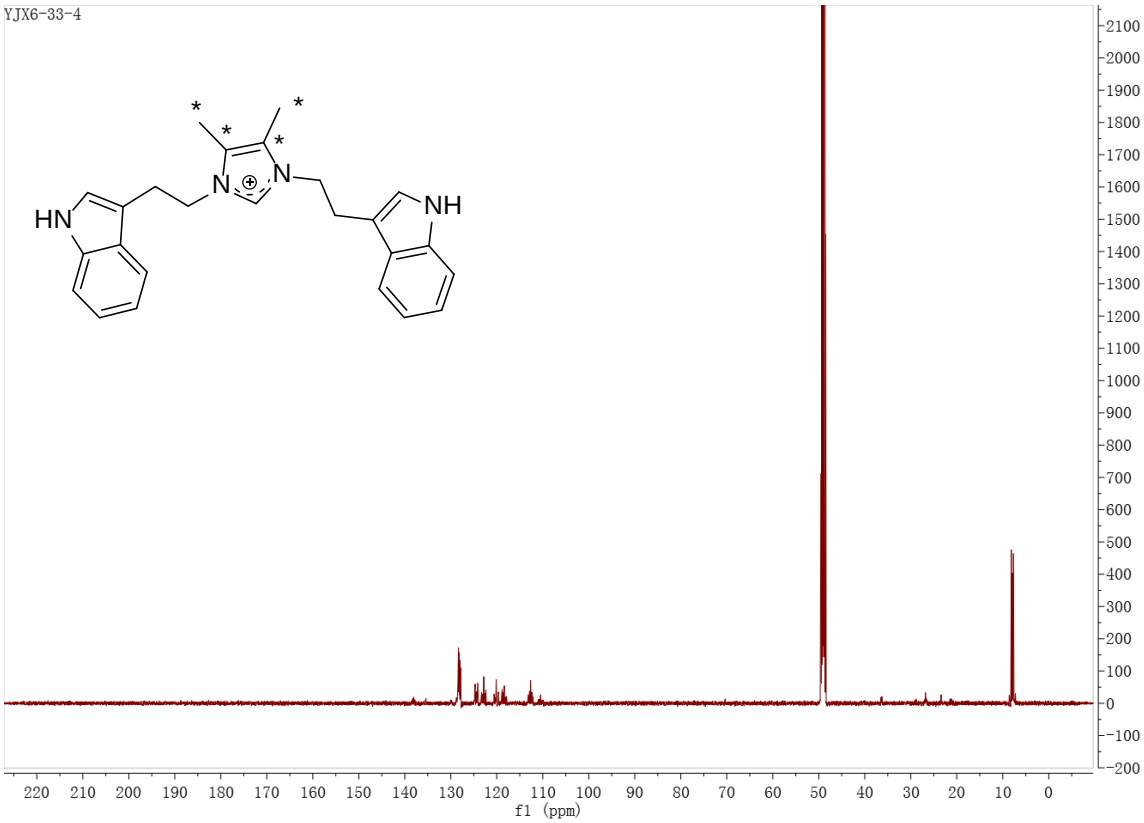


Figure S55. ^{1}H NMR Spectrum of ^{13}C Labeled Bacillimidazole E (**25**, 500 MHz, CD_3OD)

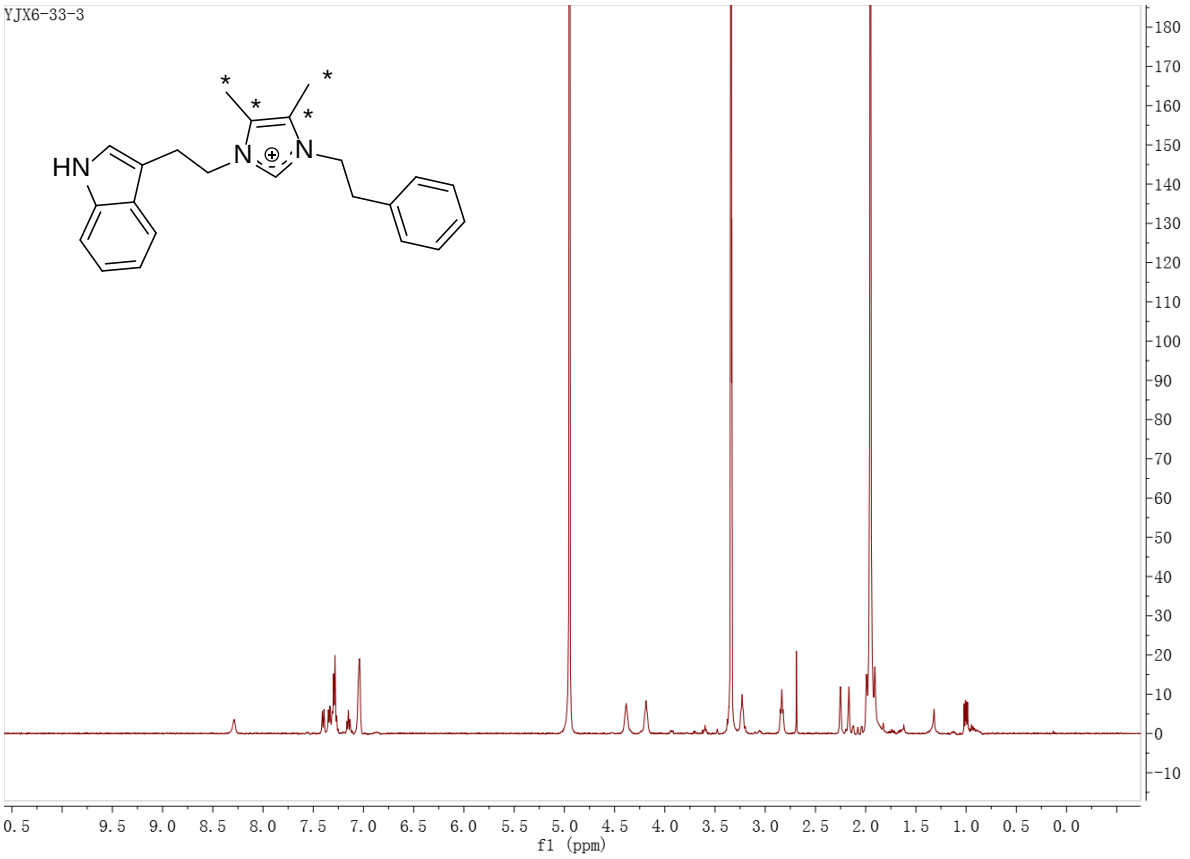


Figure S56. ^{13}C NMR Spectrum of ^{13}C Labeled Bacillimidazole E (**25**, 125 MHz, CD_3OD)

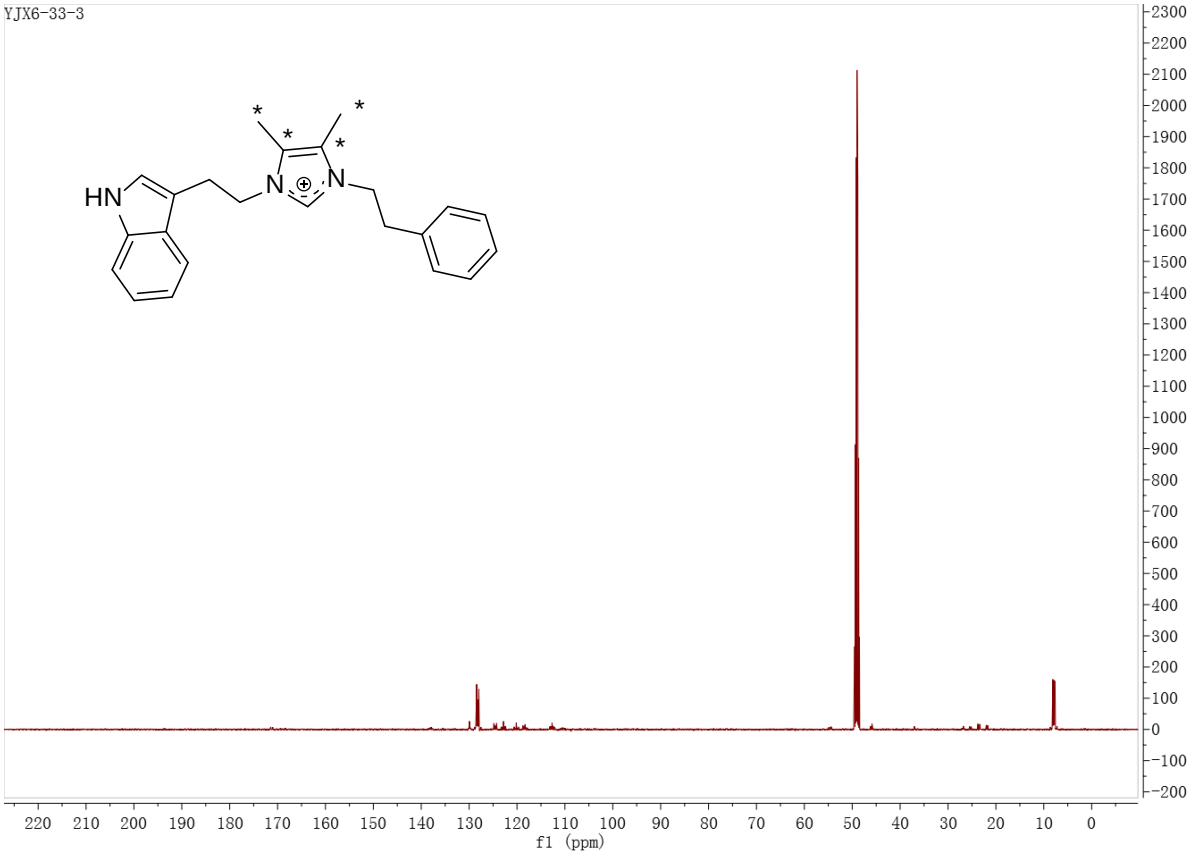


Figure S57. Positive Ion HRESIMS of ^{13}C Enriched Bacillimidazole C (**23**)

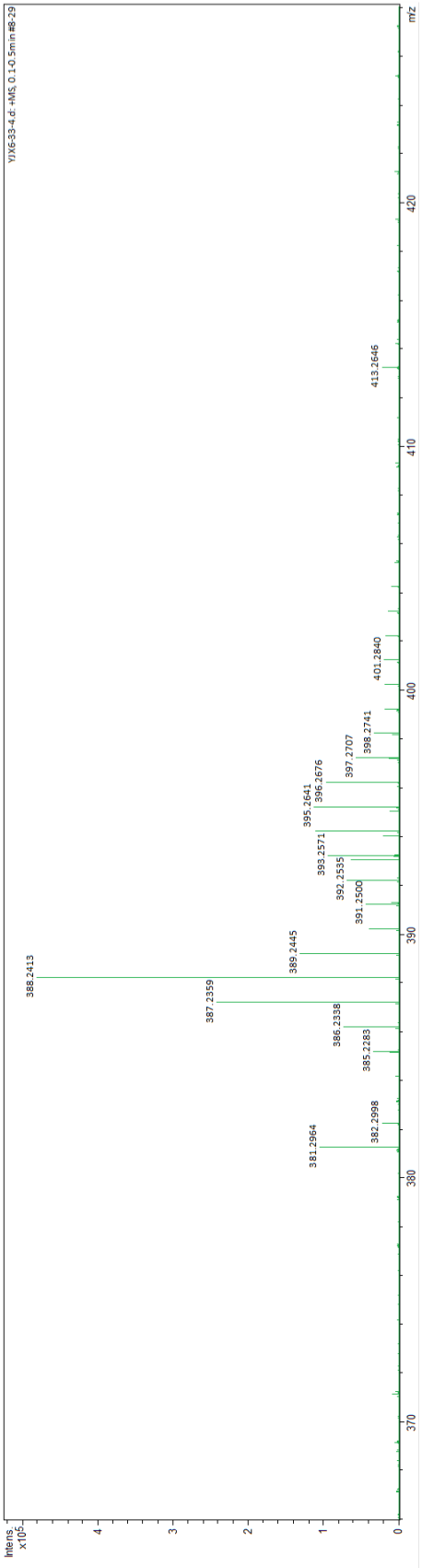
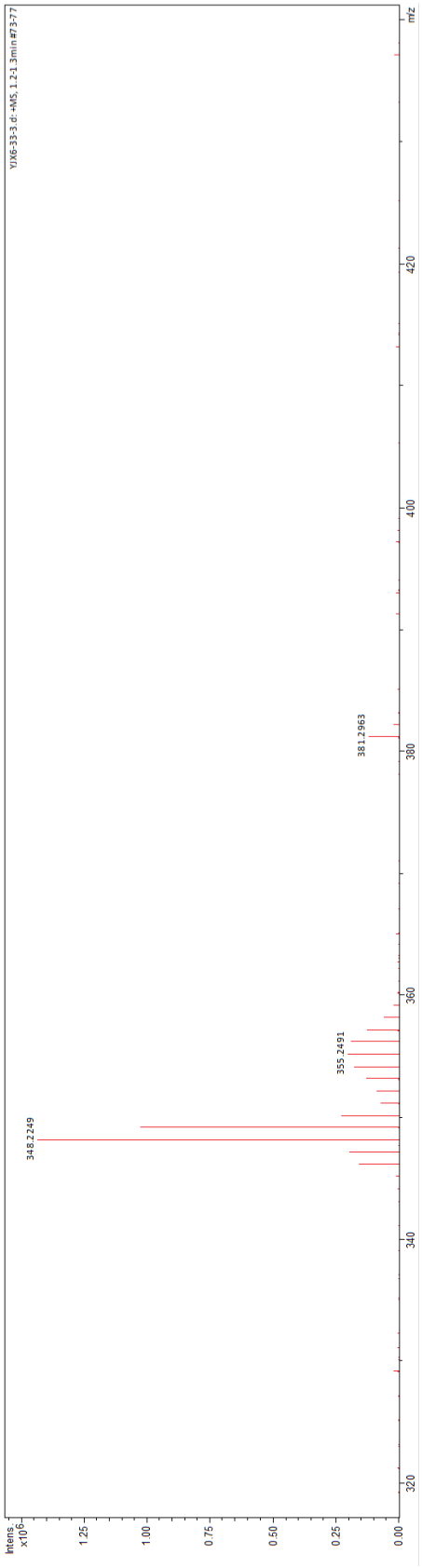


Figure S58. Positive Ion HRESIMS of ^{13}C Enriched Bacillimidazole E (**25**)



Appendix C:

Supplementary Data for Chapter 4

Figure S59. ^{1}H NMR Spectrum of Haliclonamycin (600 MHz, CD_3OD:$CDCl_3$ = 1:1)

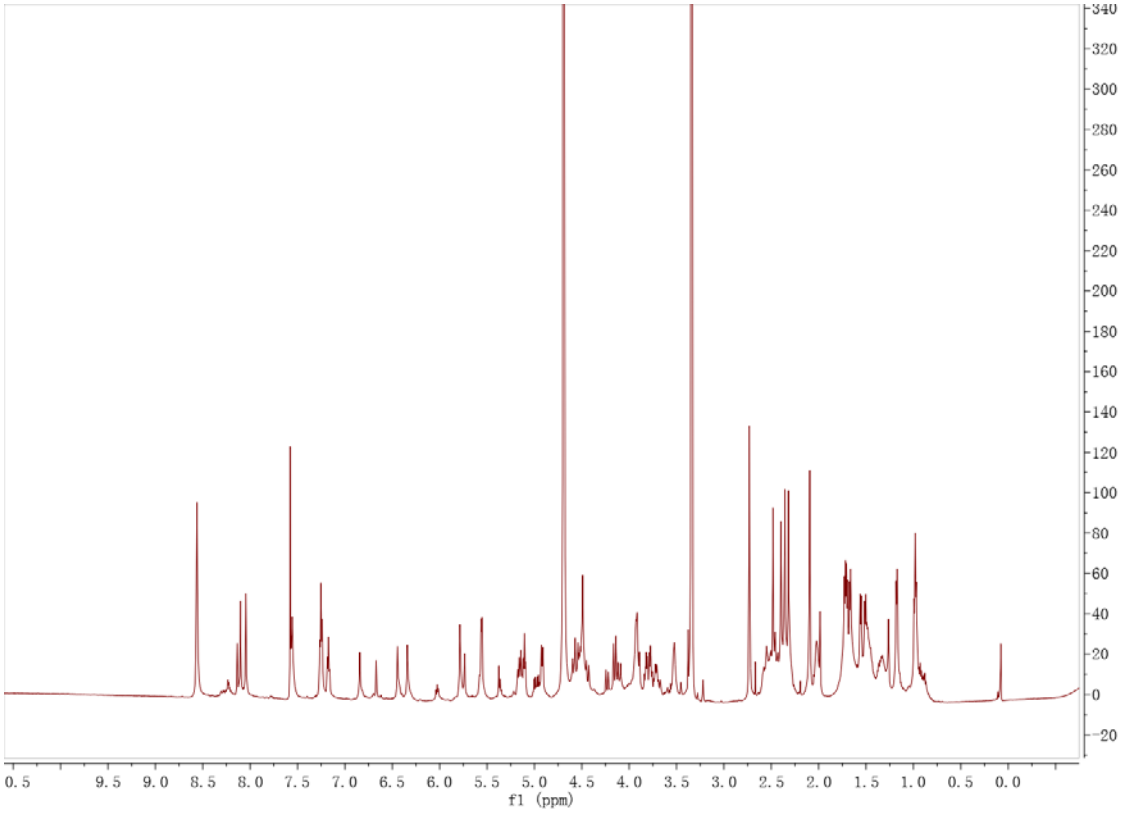


Figure S60. ^{13}C NMR Spectrum of Haliclonamycin (125 MHz, CD_3OD:$CDCl_3$ = 1:1)

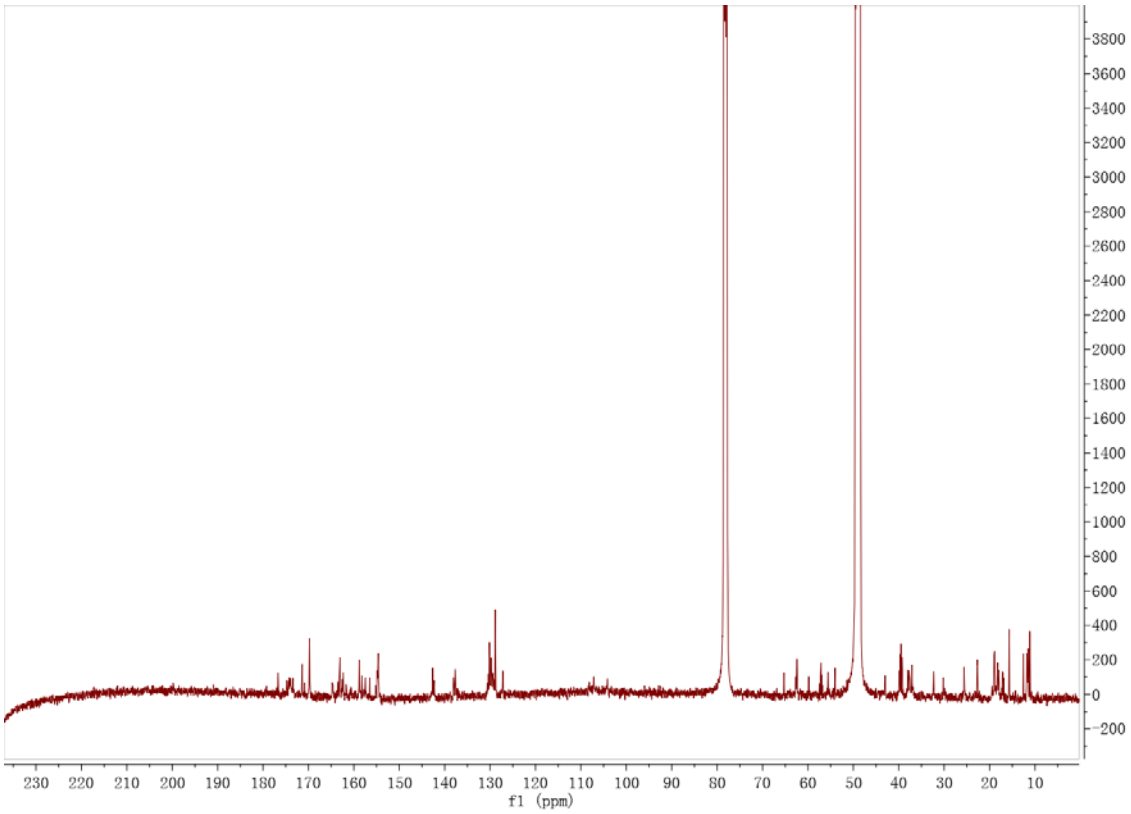


Figure S61. gCOSY NMR Spectrum of Haliclonamycin (600 MHz, CD_3OD:$CDCl_3$ = 1:1)

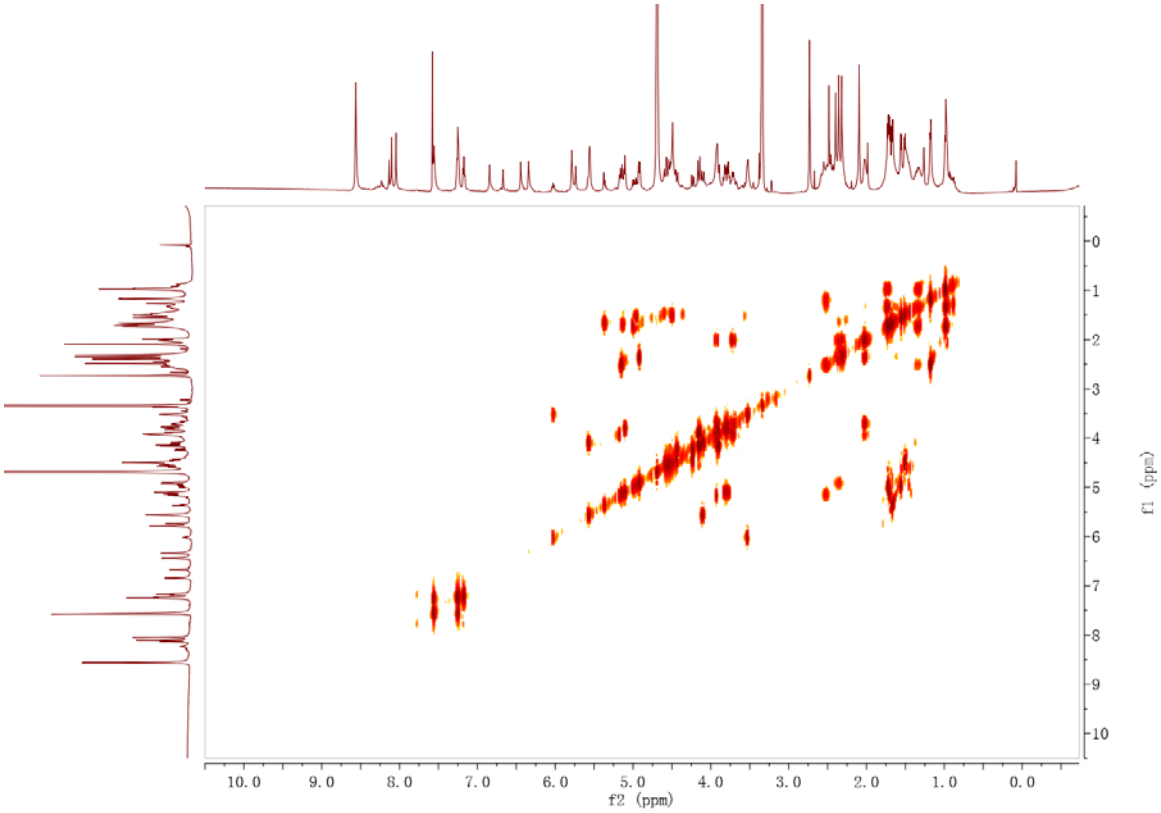


Figure S62. gHSQC NMR Spectrum of Haliclonamycin (600 MHz, CD_3OD:$CDCl_3$ = 1:1)

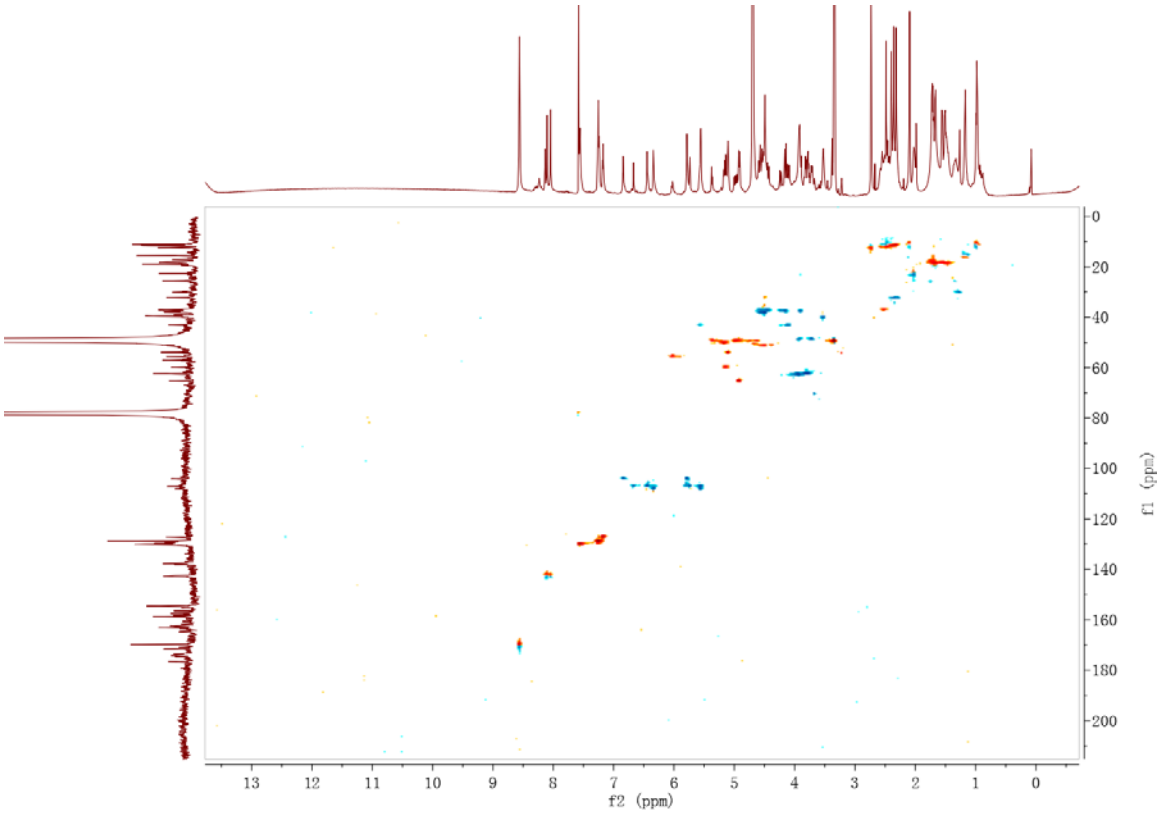


Figure S63. gHMBC NMR Spectrum of Haliclonamycin (600 MHz, CD_3OD:$CDCl_3$ = 1:1)

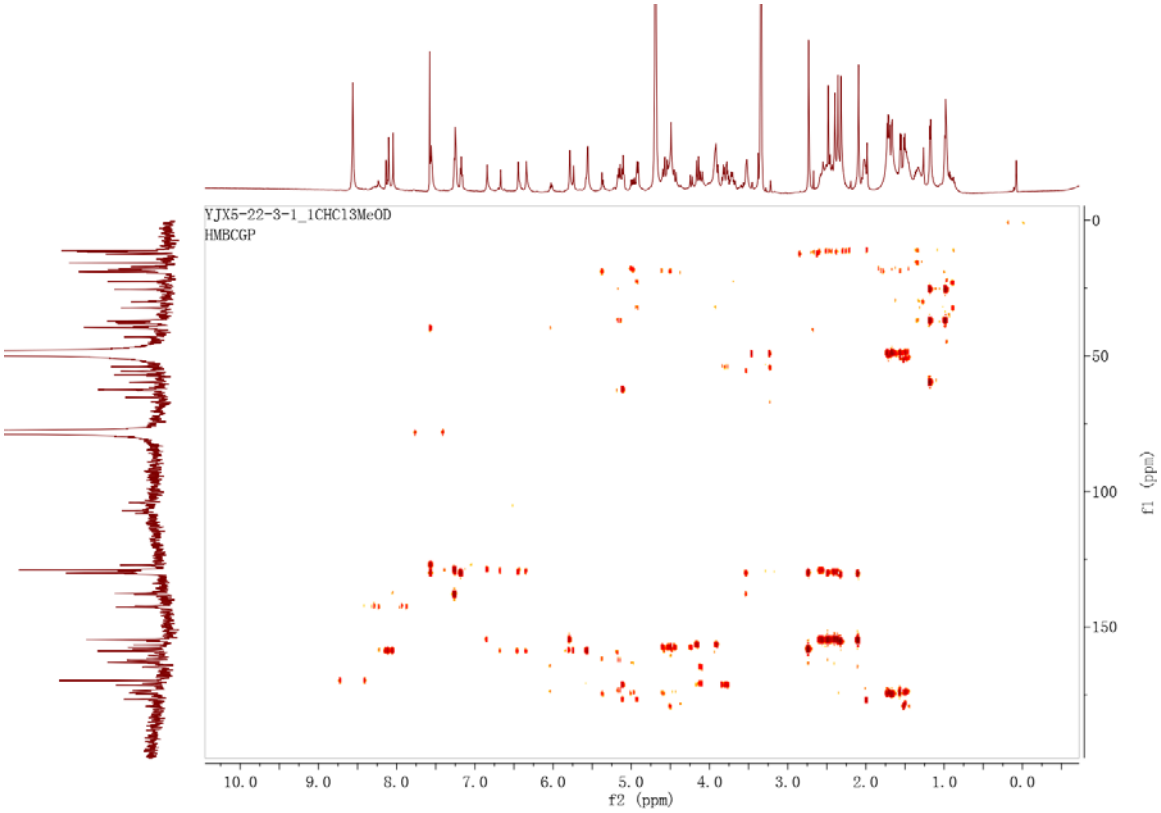


Figure S64. TOCSY Spectrum of Haliclonamycin (600 MHz, CD_3OD:$CDCl_3$ = 1:1)

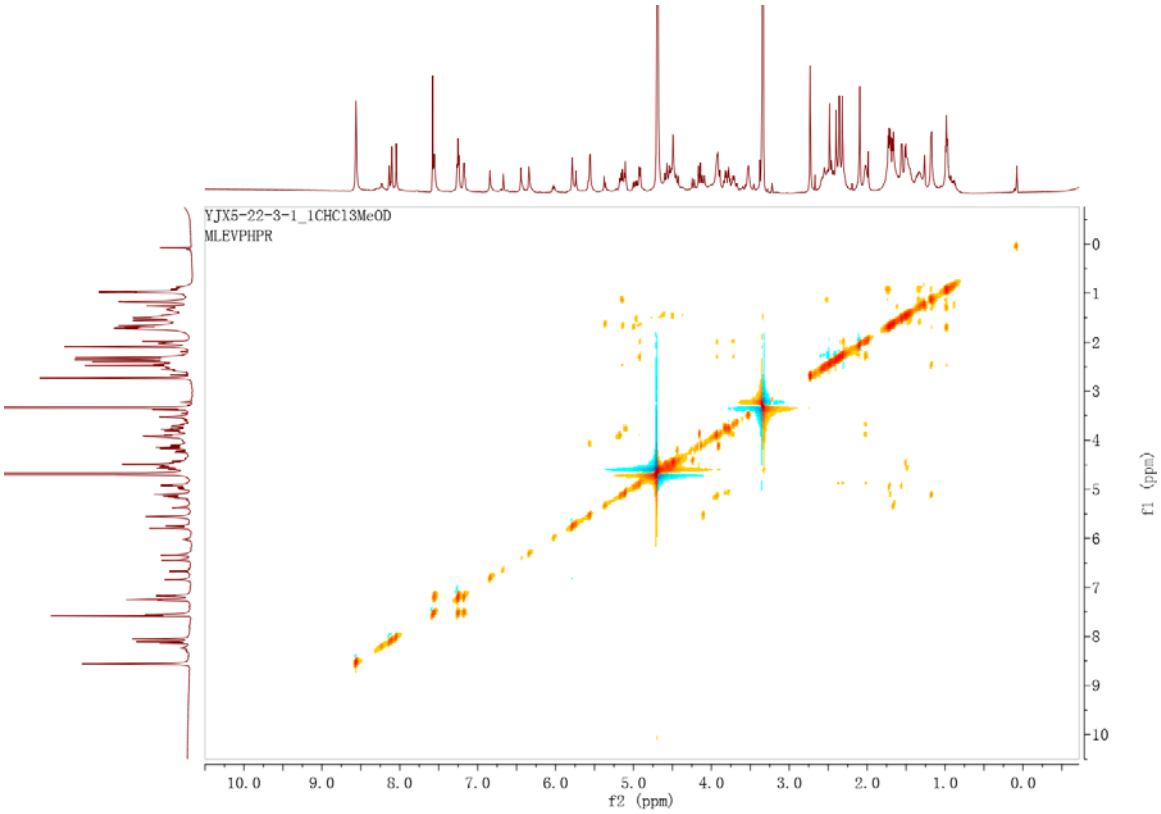


Figure S65. ^{1}H NMR Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

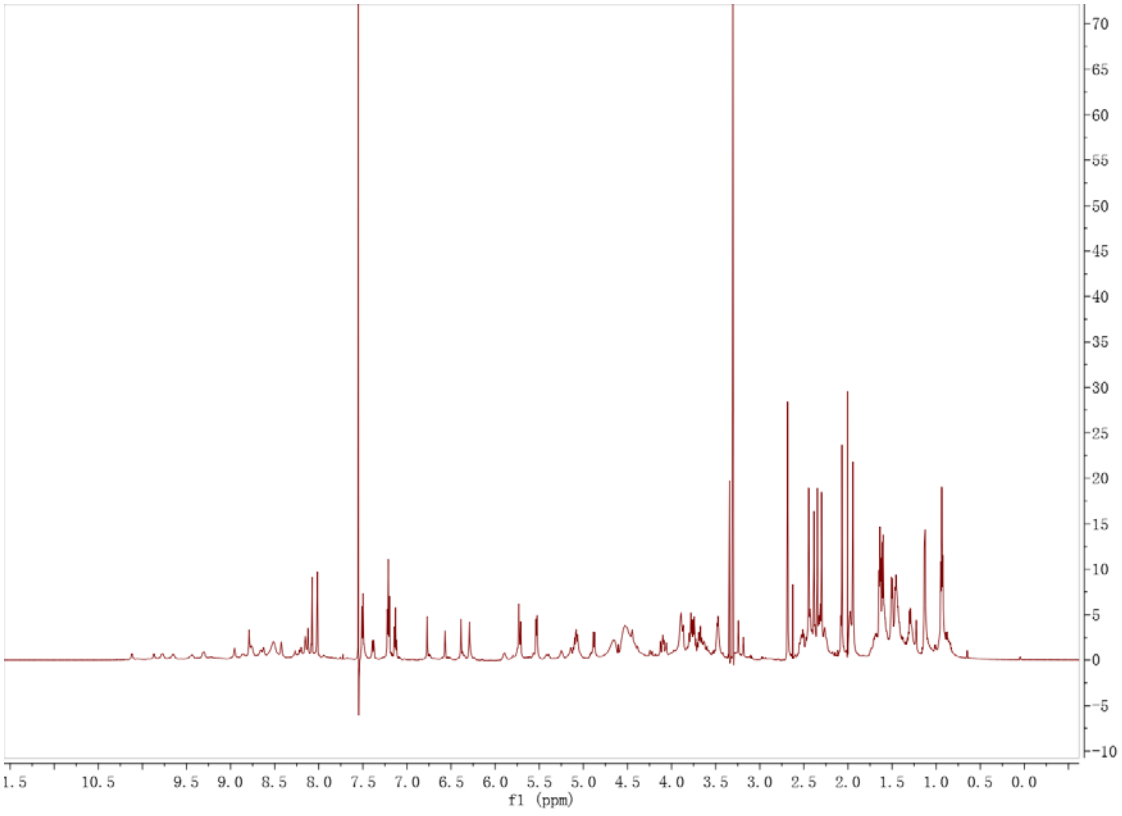


Figure S66. gHSQC NMR Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

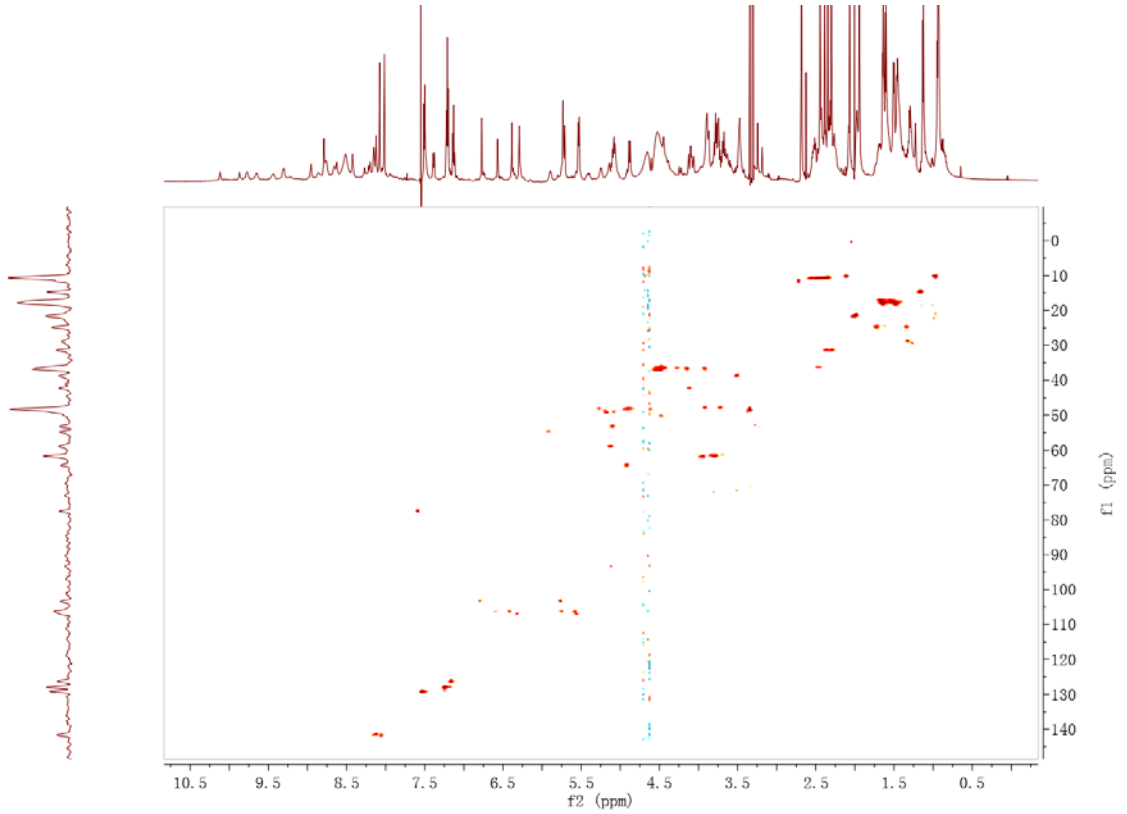


Figure S67. gHMBC NMR Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

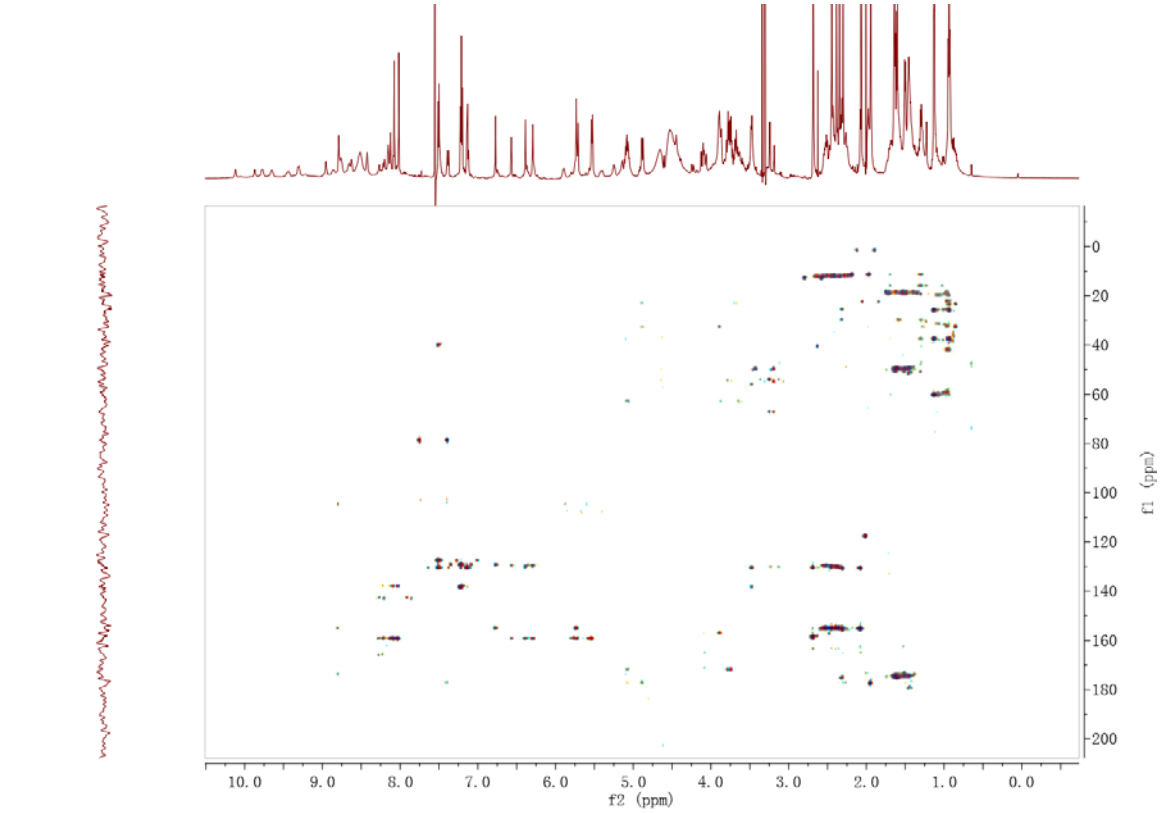


Figure S68. TOCSY Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

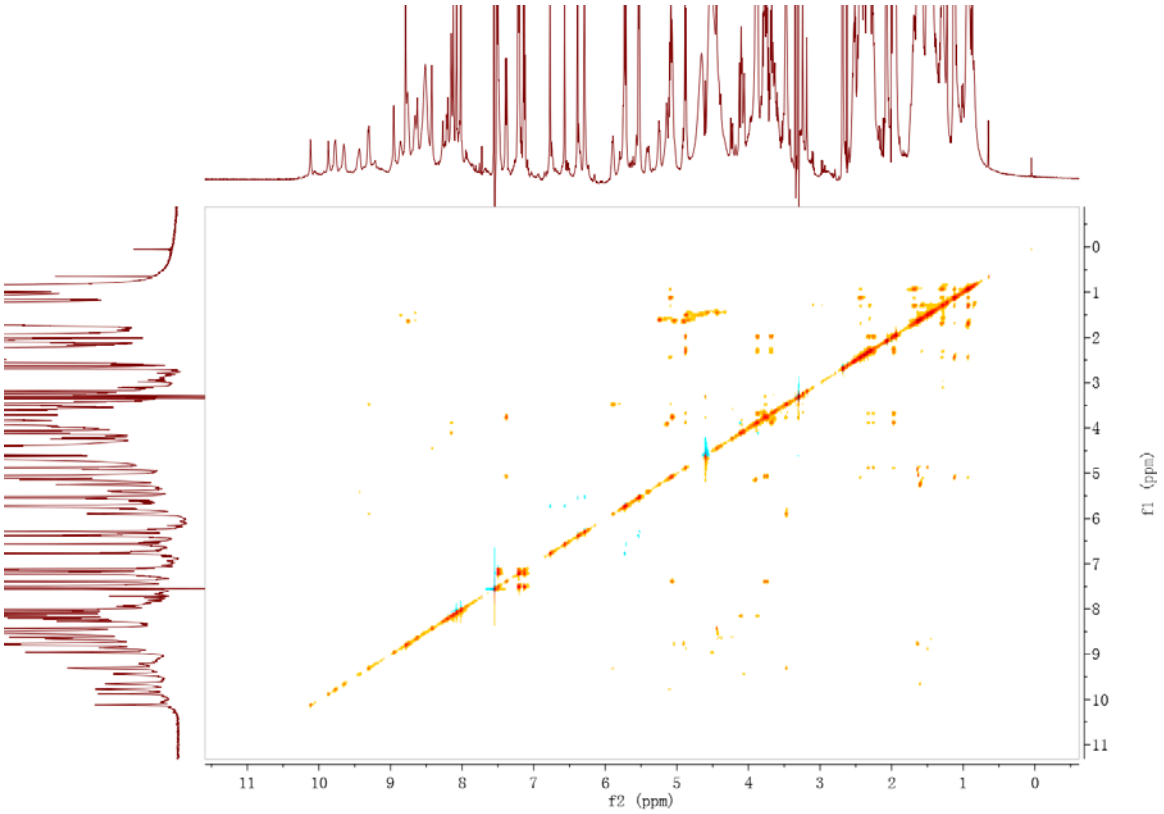


Figure S69. ROESY Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

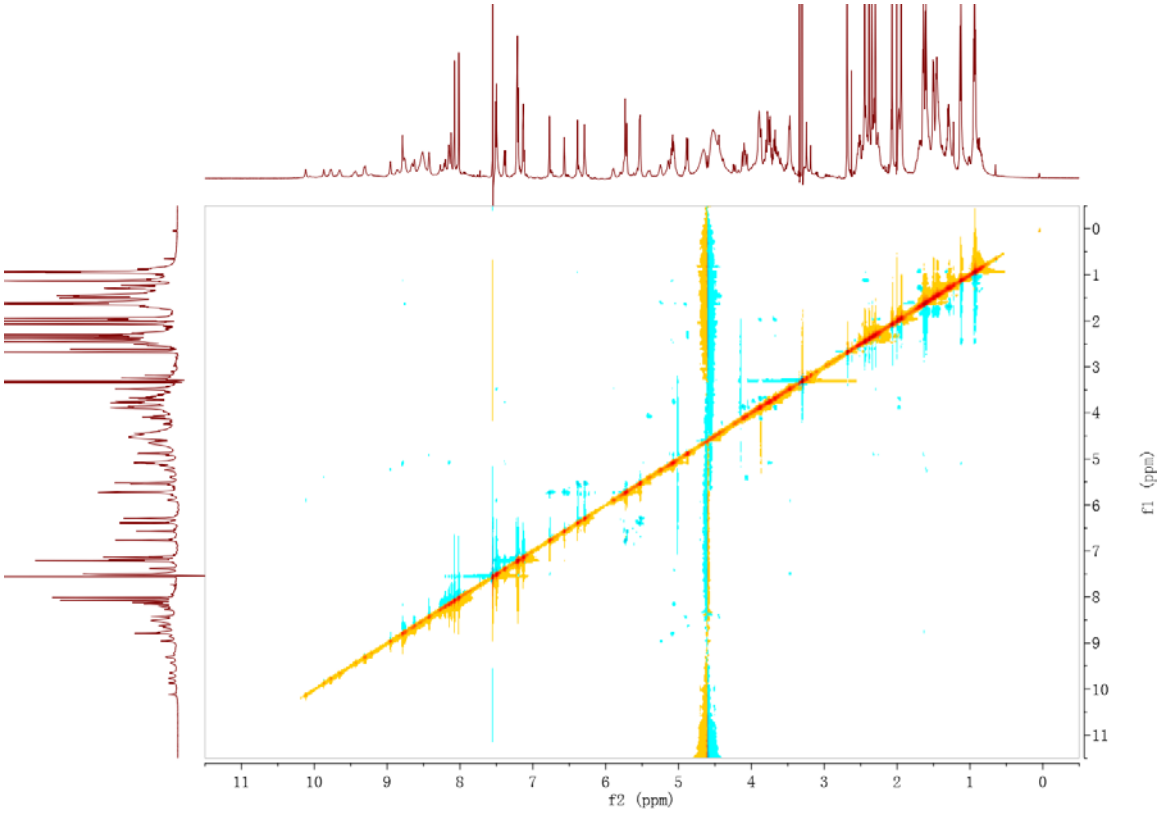


Figure S70. ^{15}N-HSQC Spectrum of Haliclonamycin (600 MHz, CD_3OH:$CDCl_3$ = 1:1)

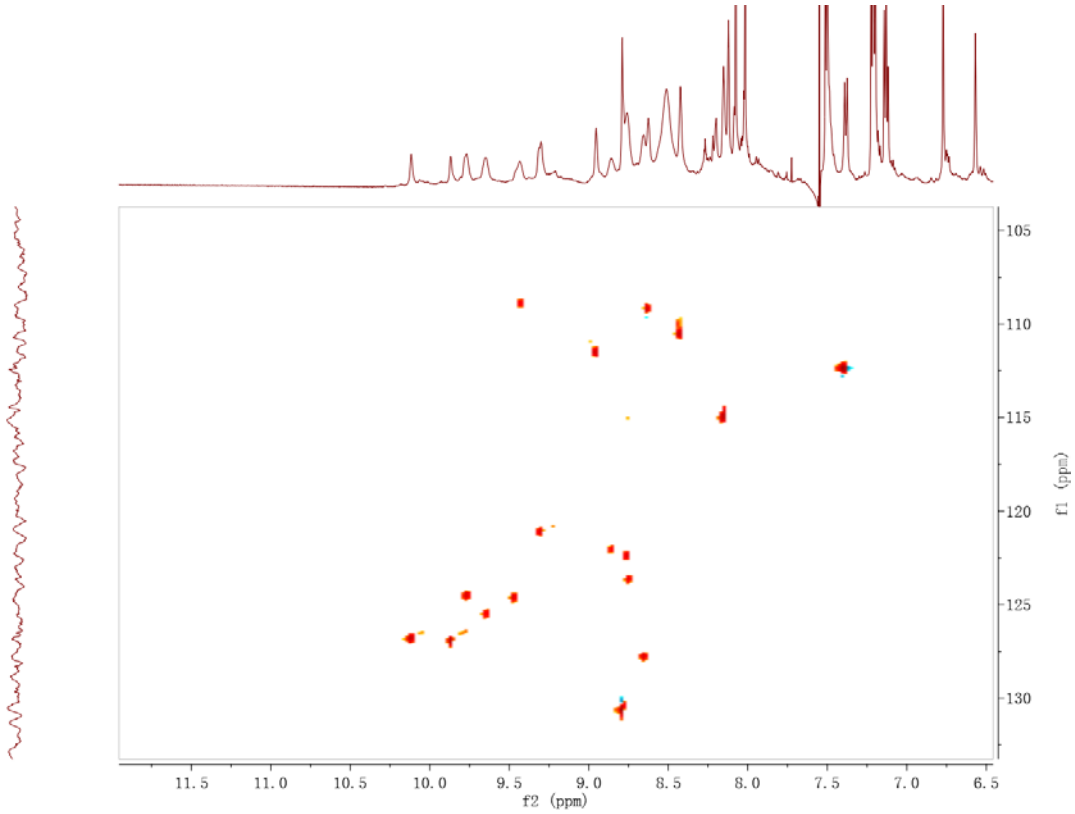
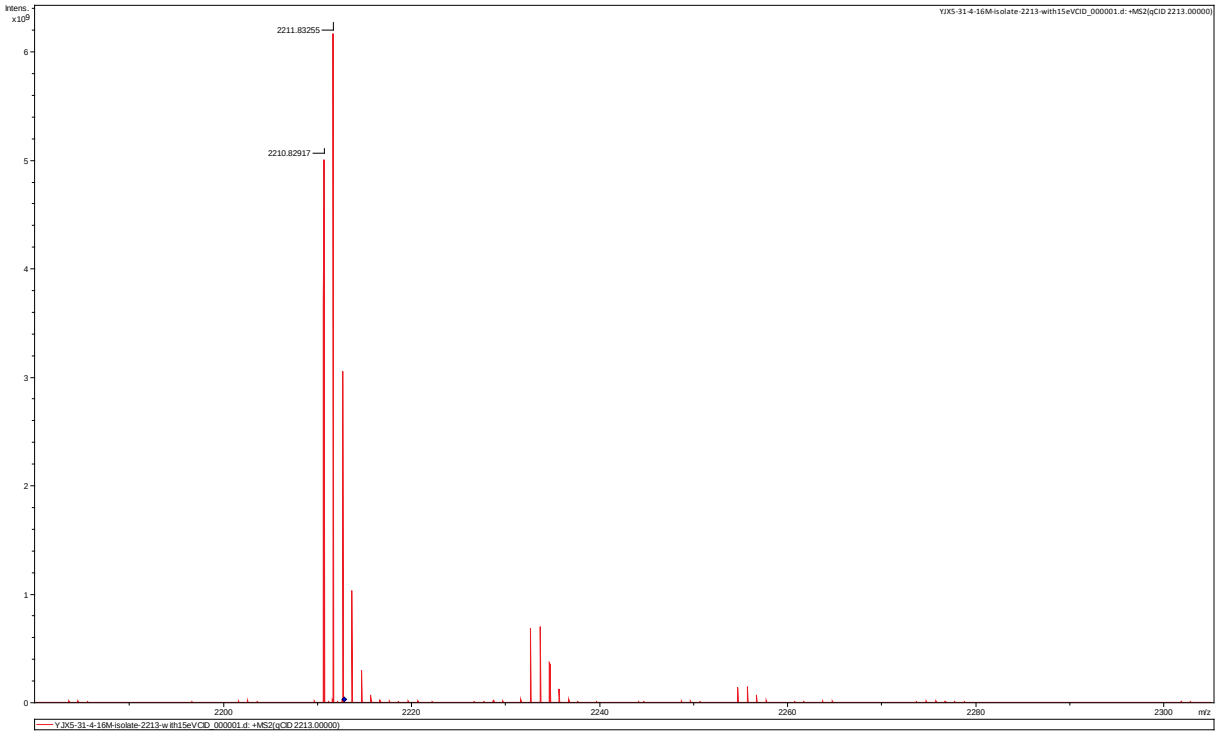


Figure S71. Positive Ion ESI-MRMS of Haliclonamycin



Appendix D:

Supplementary Data for Chapter 5

Figure S72. ^{1}H NMR Spectrum of Ecteinamine A (**33**, 600 MHz, CD_3OD)

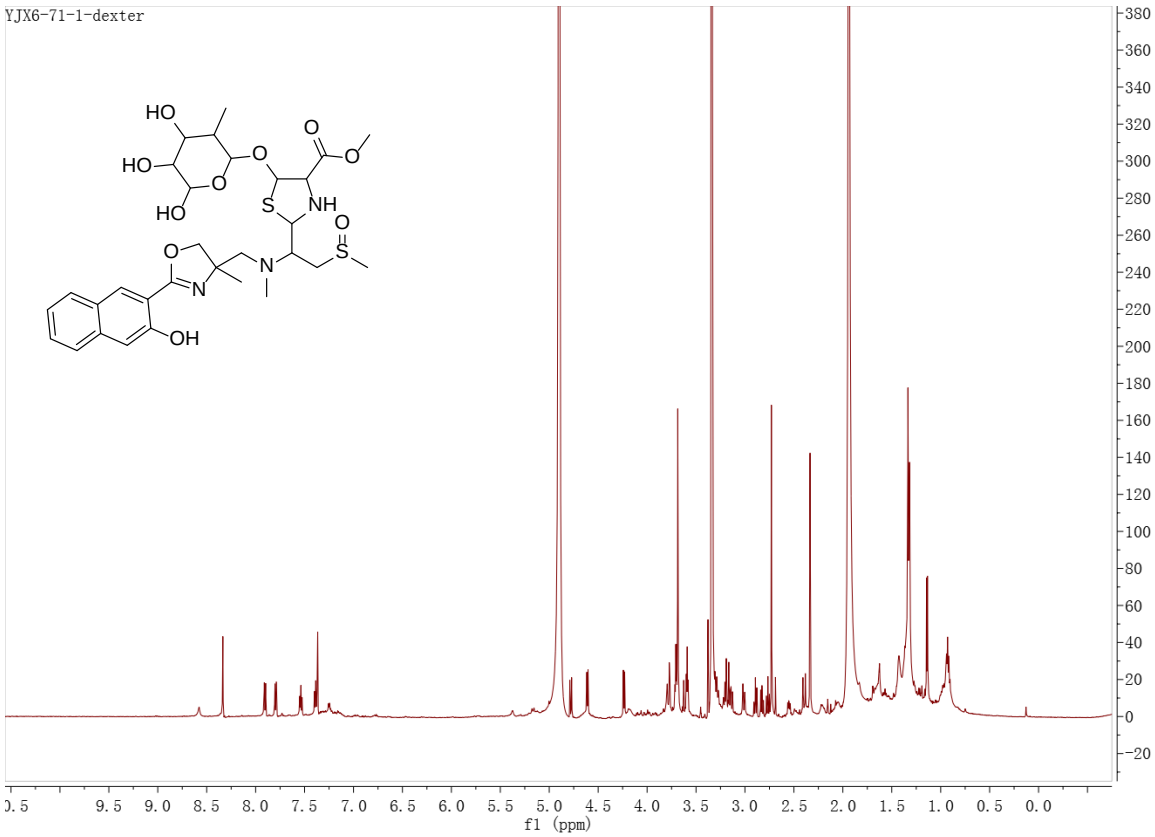


Figure S73. ^{13}C NMR Spectrum of Ecteinamine A (**33**, 125 MHz, CD_3OD)

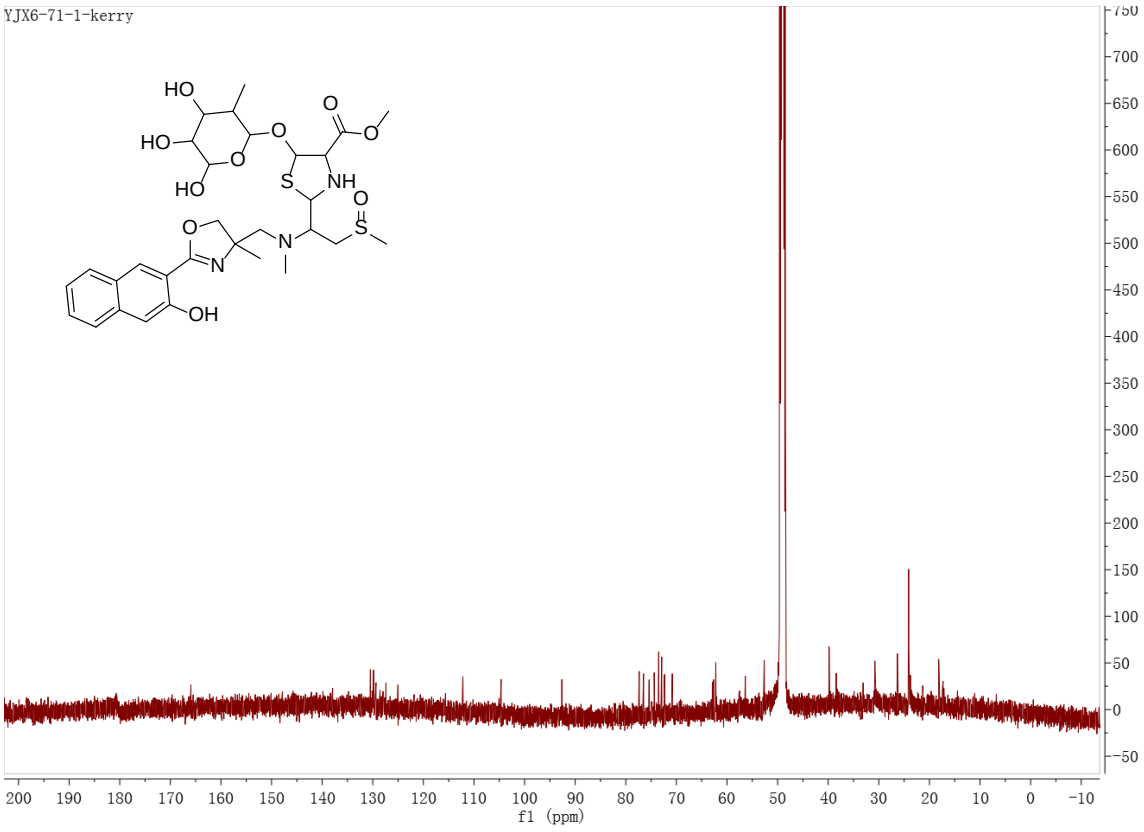


Figure S74. gCOSY NMR Spectrum of Ecteinamine A (**33**, 600 MHz, CD_3OD)

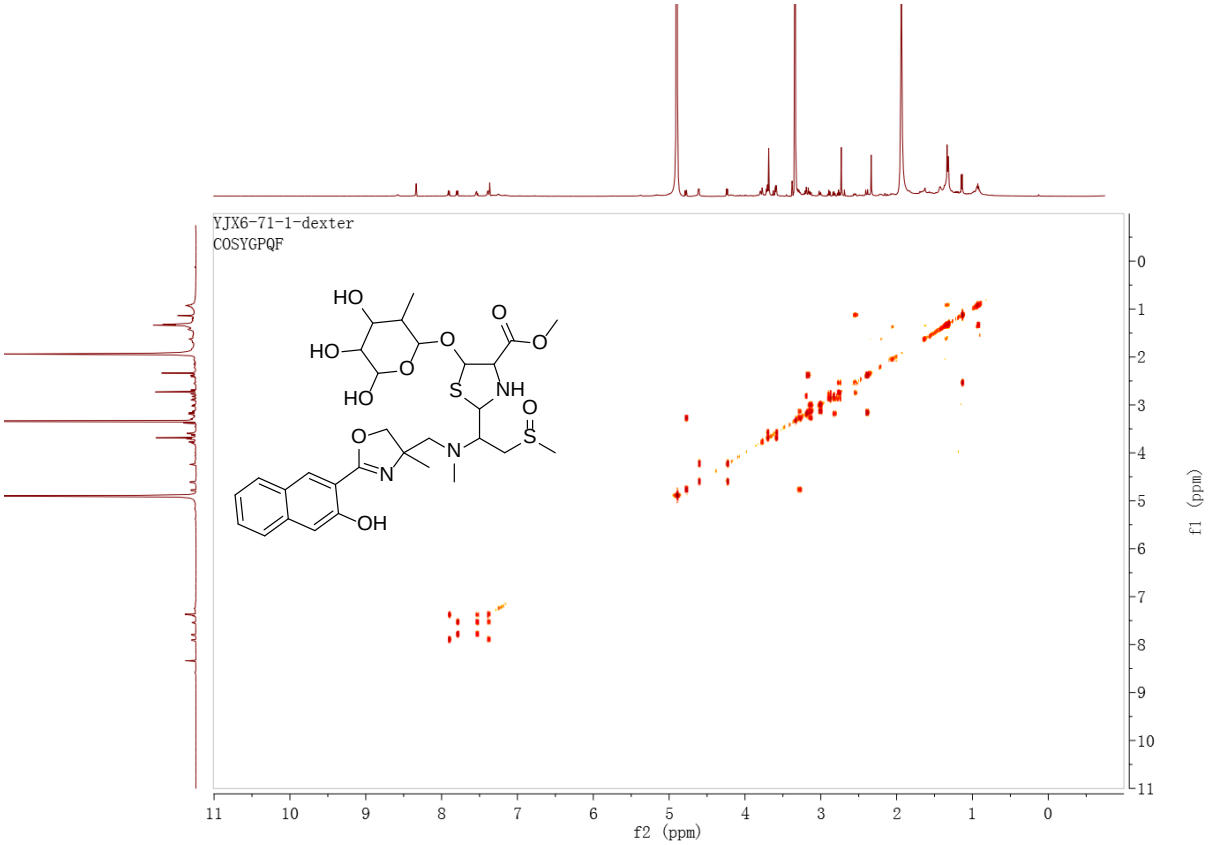


Figure S75. gHSQC NMR Spectrum of Ecteinamine A (**33**, 600 MHz, CD_3OD)

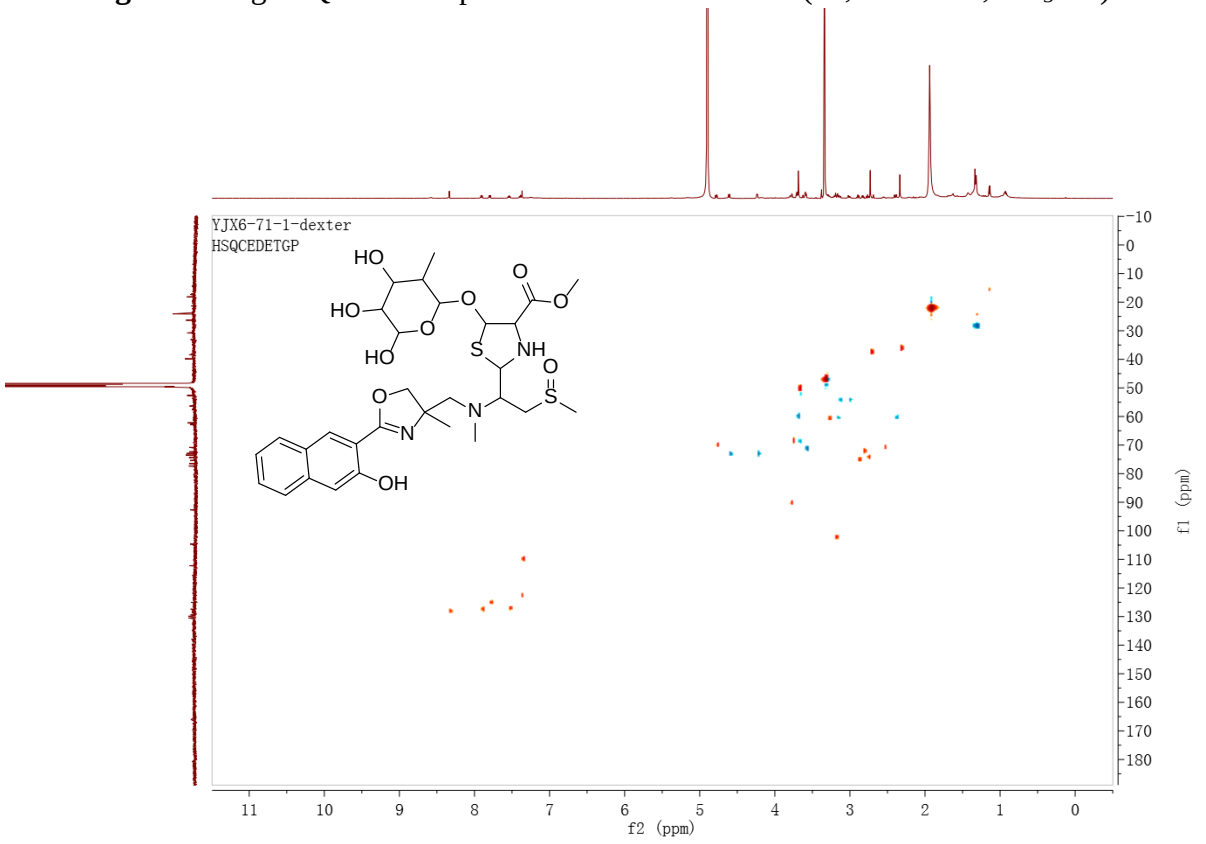


Figure S76. gHMBC NMR Spectrum of Ecteinamine A (**33**, 600 MHz, CD_3OD)

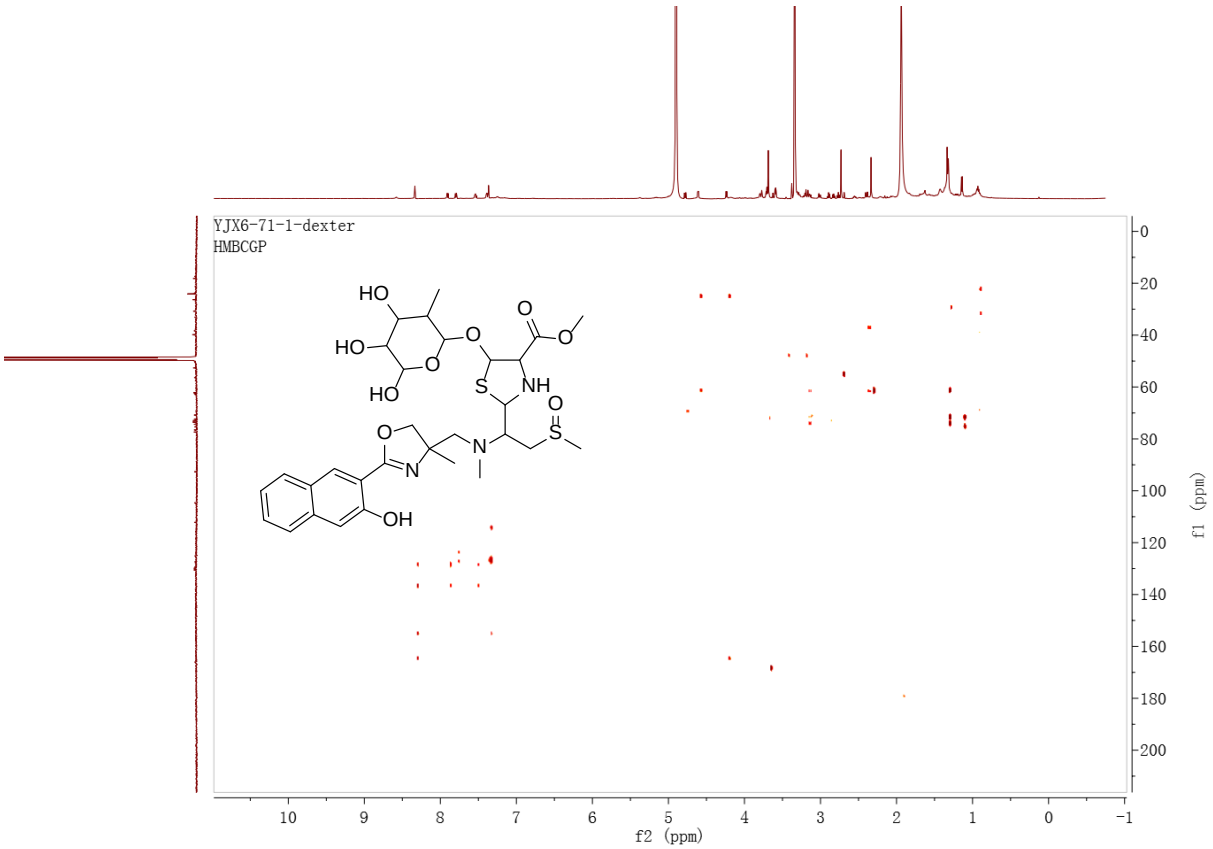


Figure S77. ^1H NMR Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

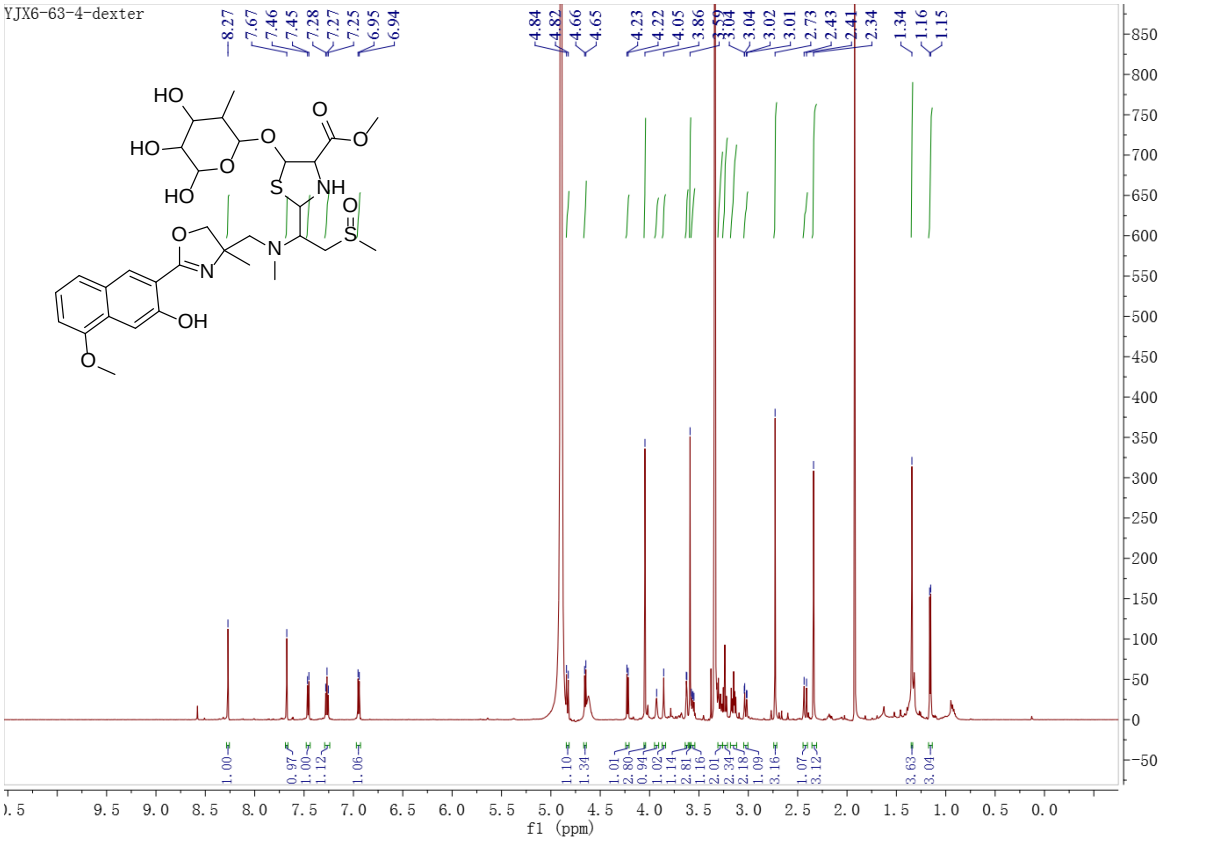


Figure S78. ^{13}C NMR Spectrum of Ecteinamine B (**34**, 125 MHz, CD_3OD)

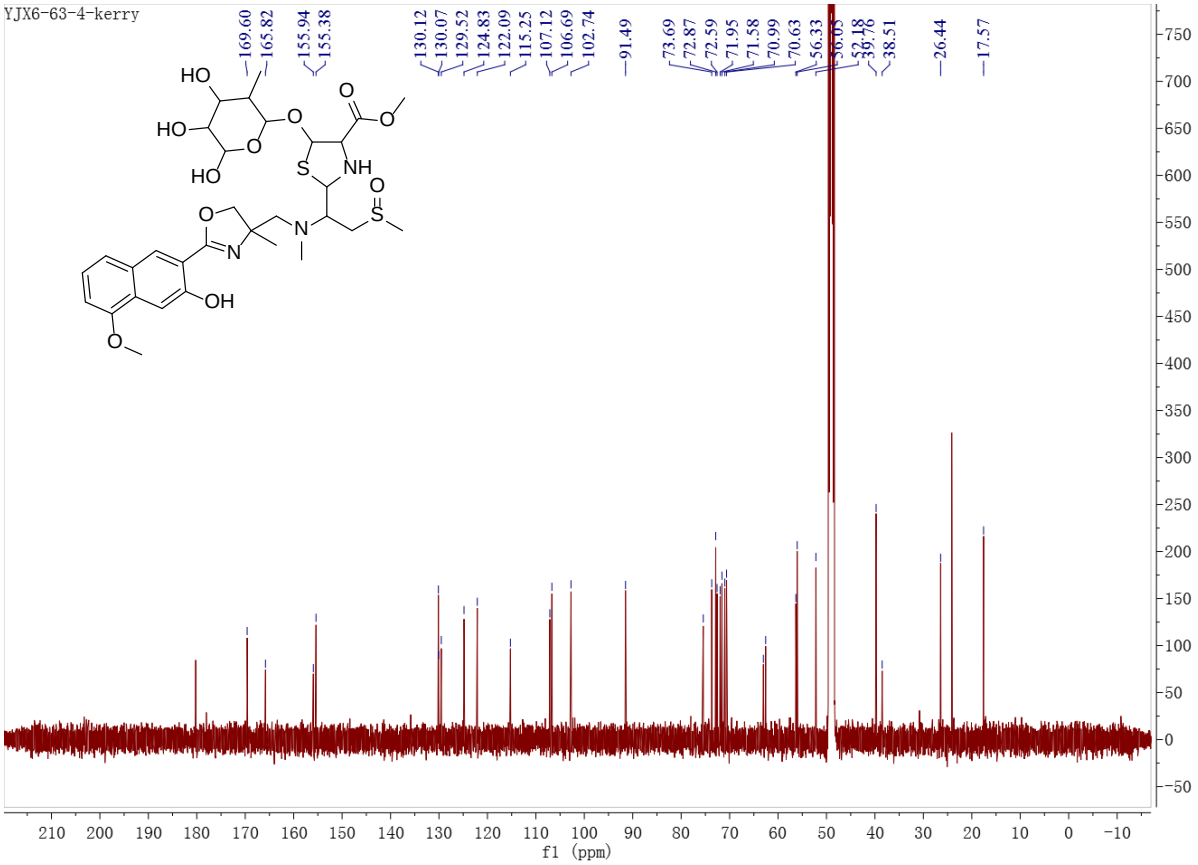


Figure S79. gCOSY NMR Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

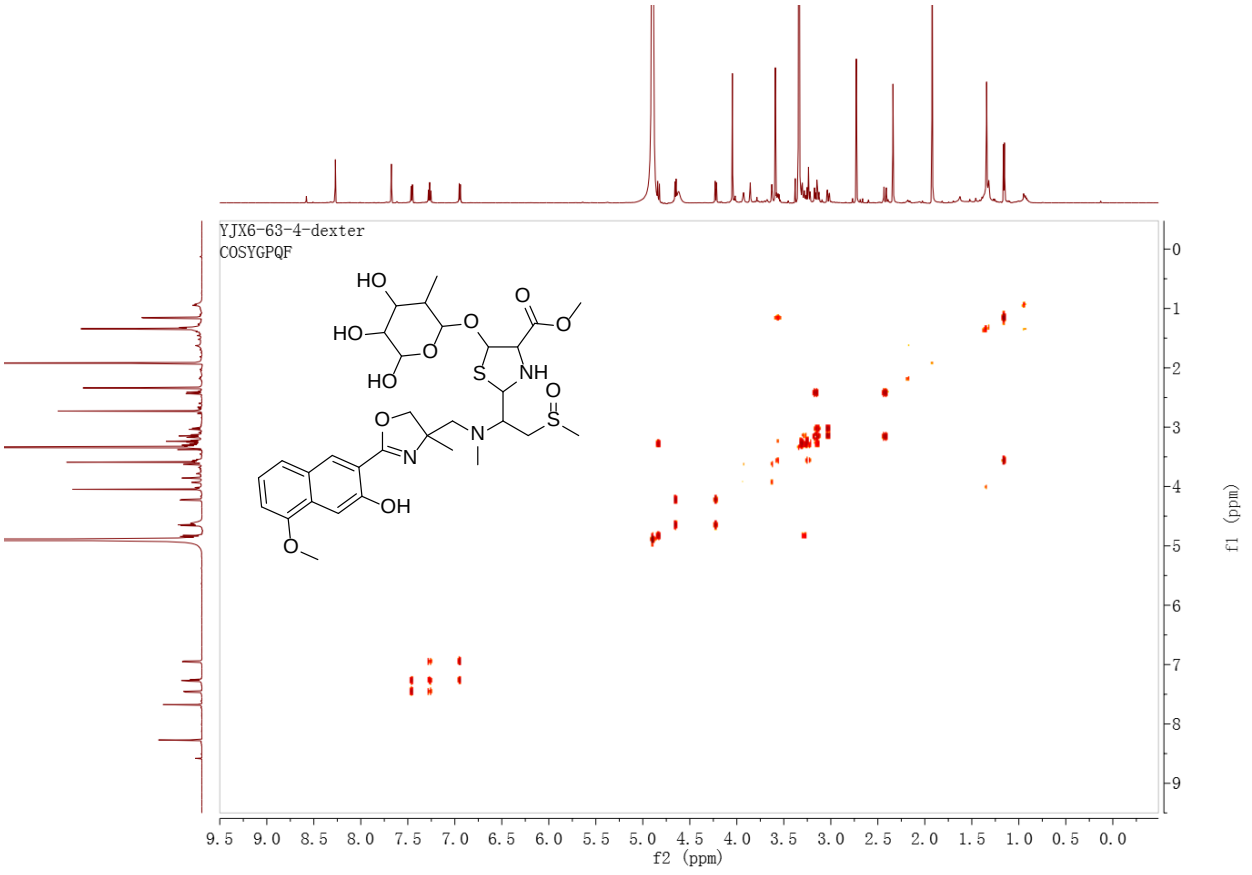


Figure S80. gHSQC NMR Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

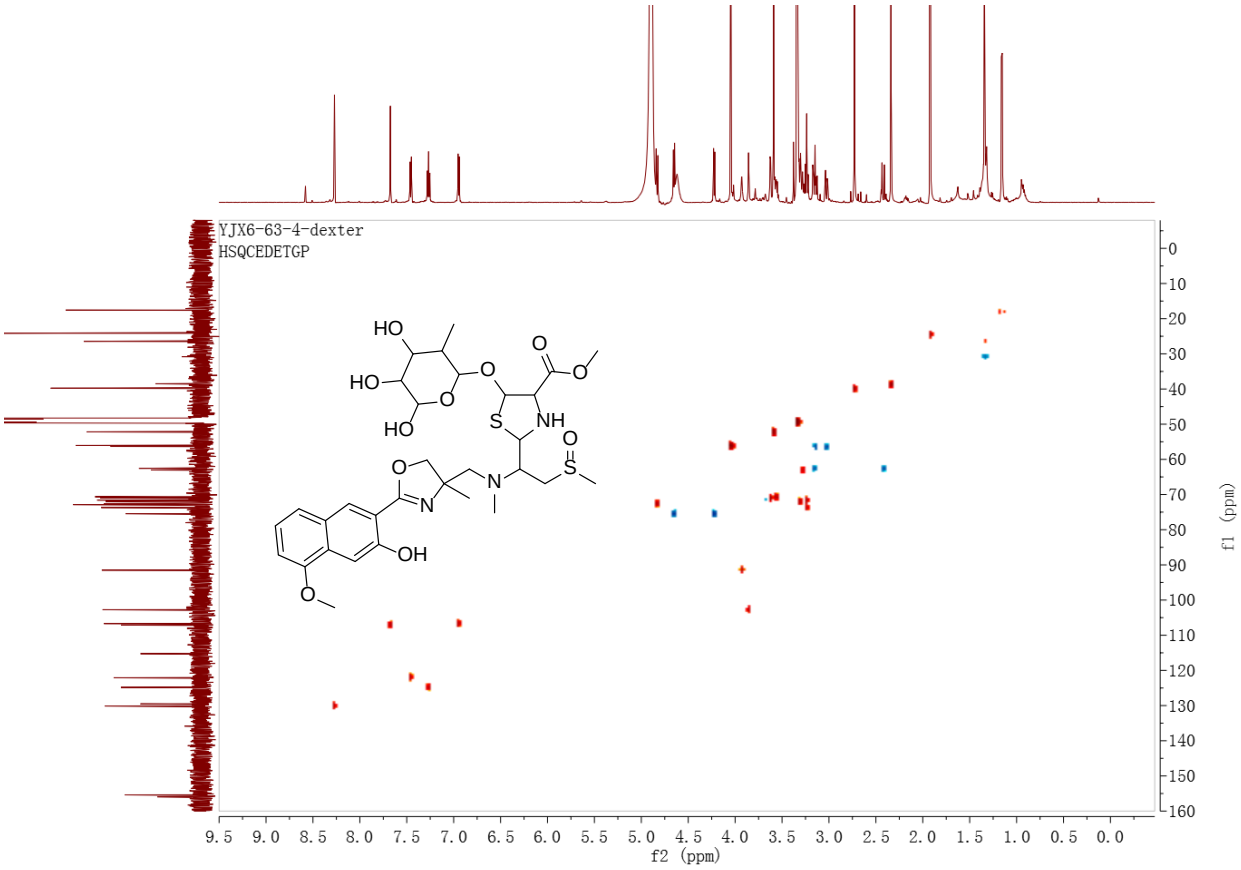


Figure S81. gHMBC NMR Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

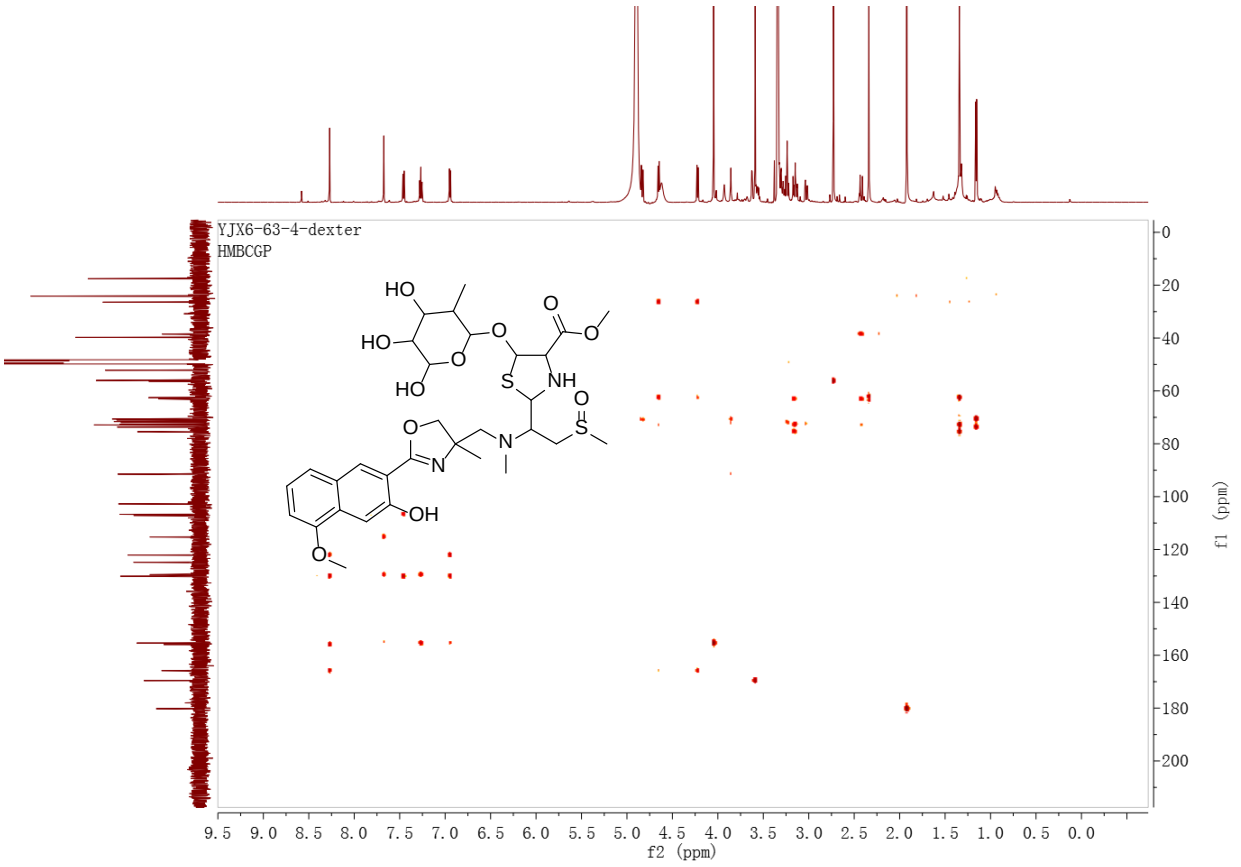


Figure S82. TOCSY Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

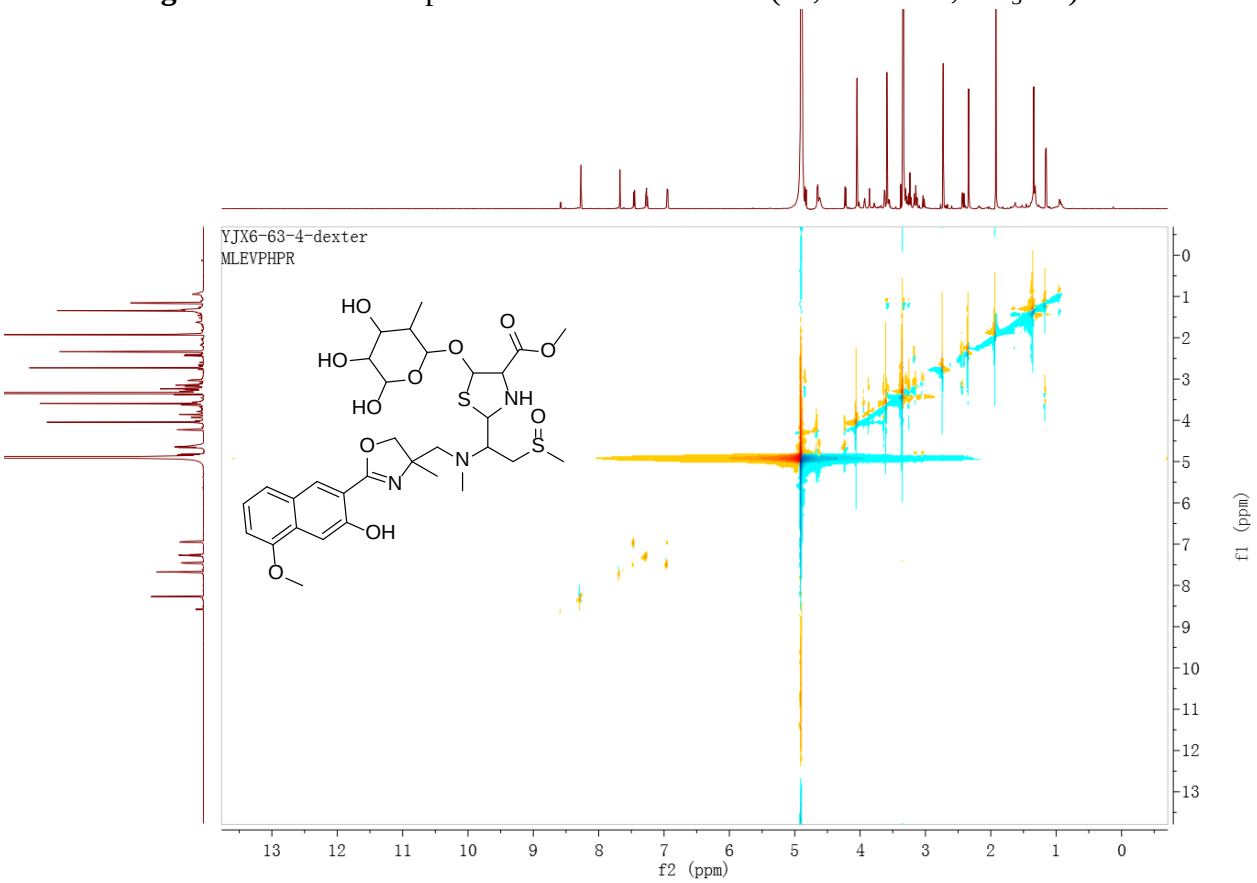


Figure S83. ROESY Spectrum of Ecteinamine B (**34**, 600 MHz, d_6-DMSO)

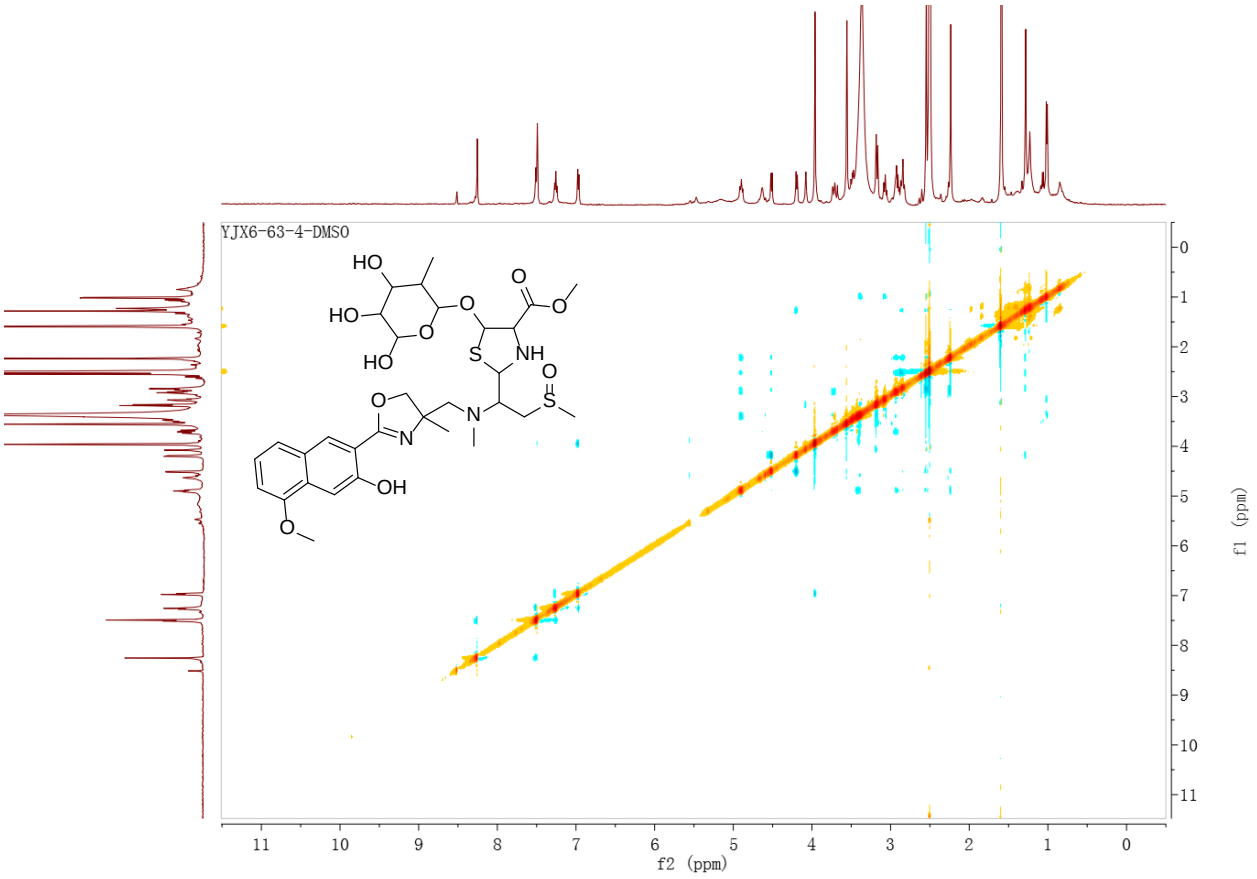


Figure S84. ^{15}N-HMBC Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD)

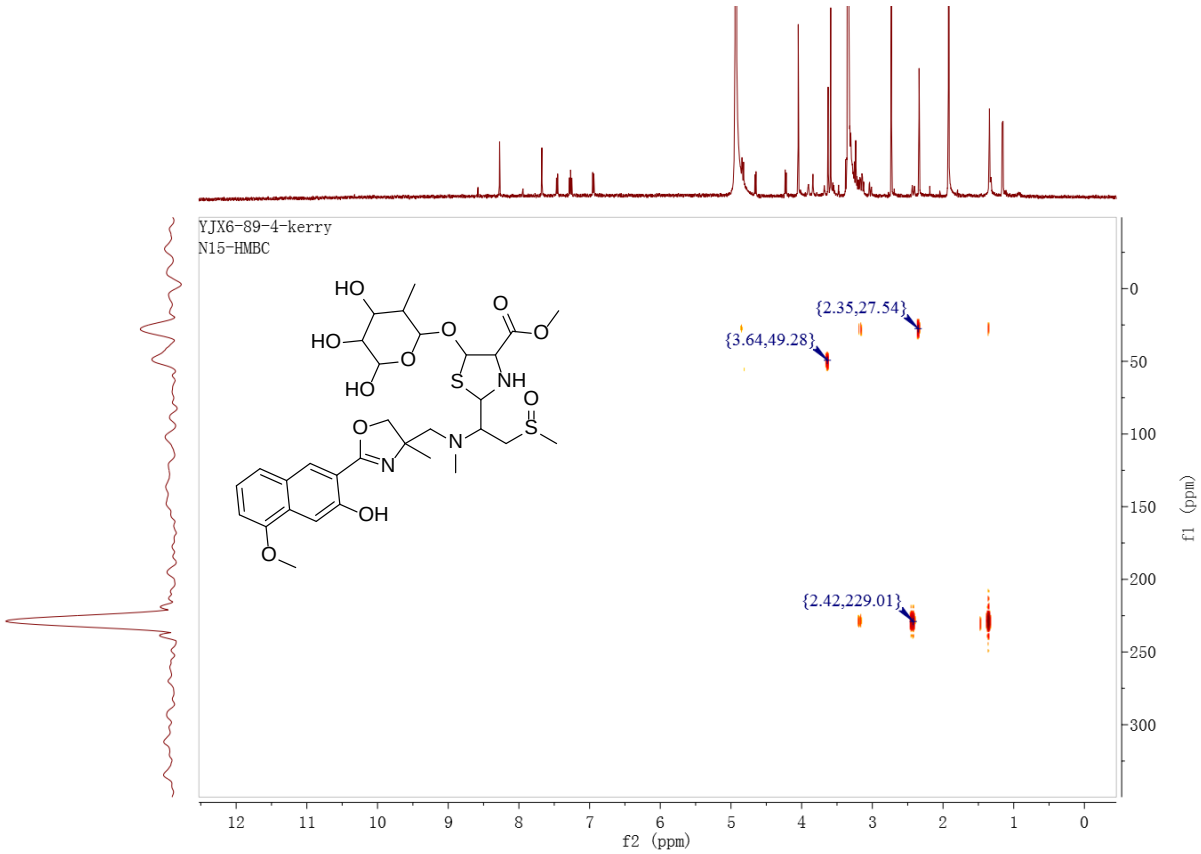


Figure S85. ^{15}N-HMBC Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD + 10% *d*-TFA)

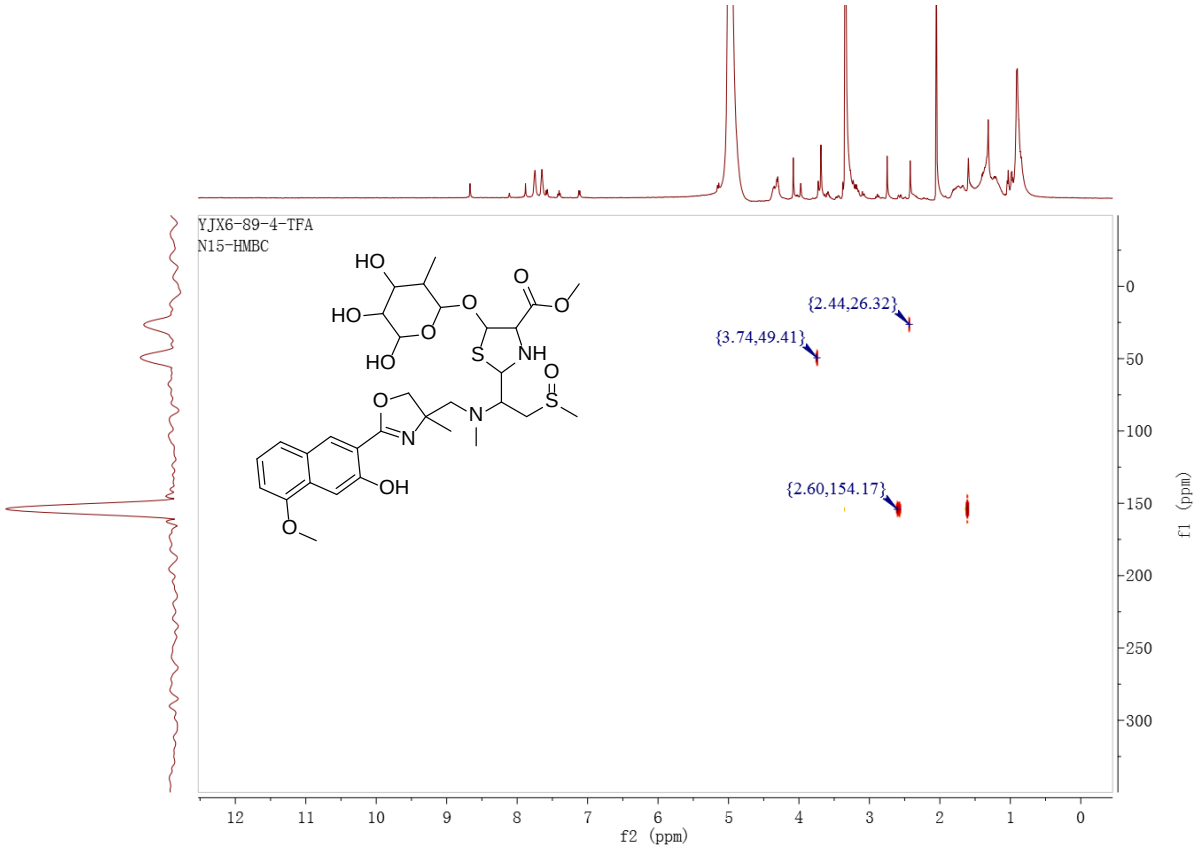


Figure S86. ^{15}N-HMBC Spectrum of Ecteinamine B (**34**, 600 MHz, CD_3OD + 20% KOD)

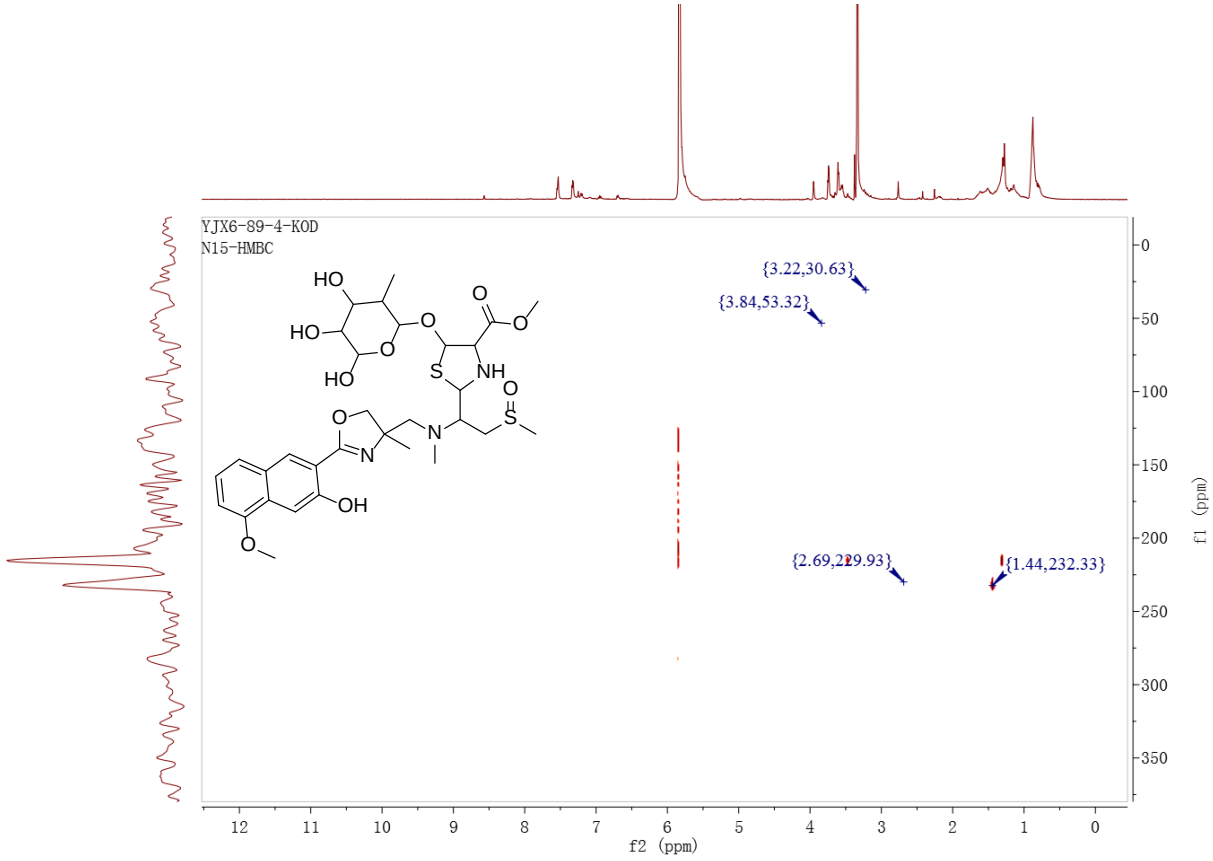


Figure S87. ^{13}C-^{13}C COSY Spectrum of ^{13}C-enriched Ecteinamine B (**34**, 600 MHz, CD_3OD)

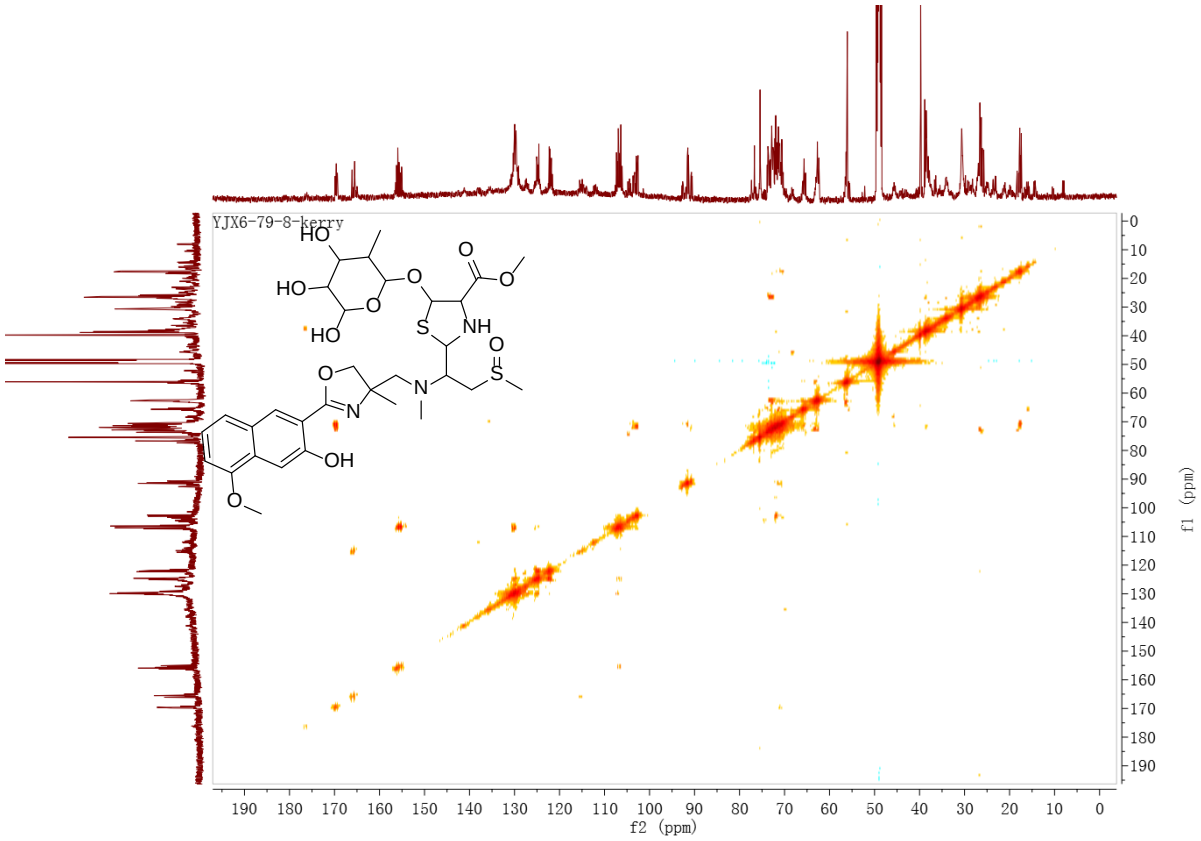


Figure S88. ^{15}N-decoupled ^{13}C Spectrum of ^{15}N-enriched Ecteinamine B (**34**, 600 MHz, CD_3OD)

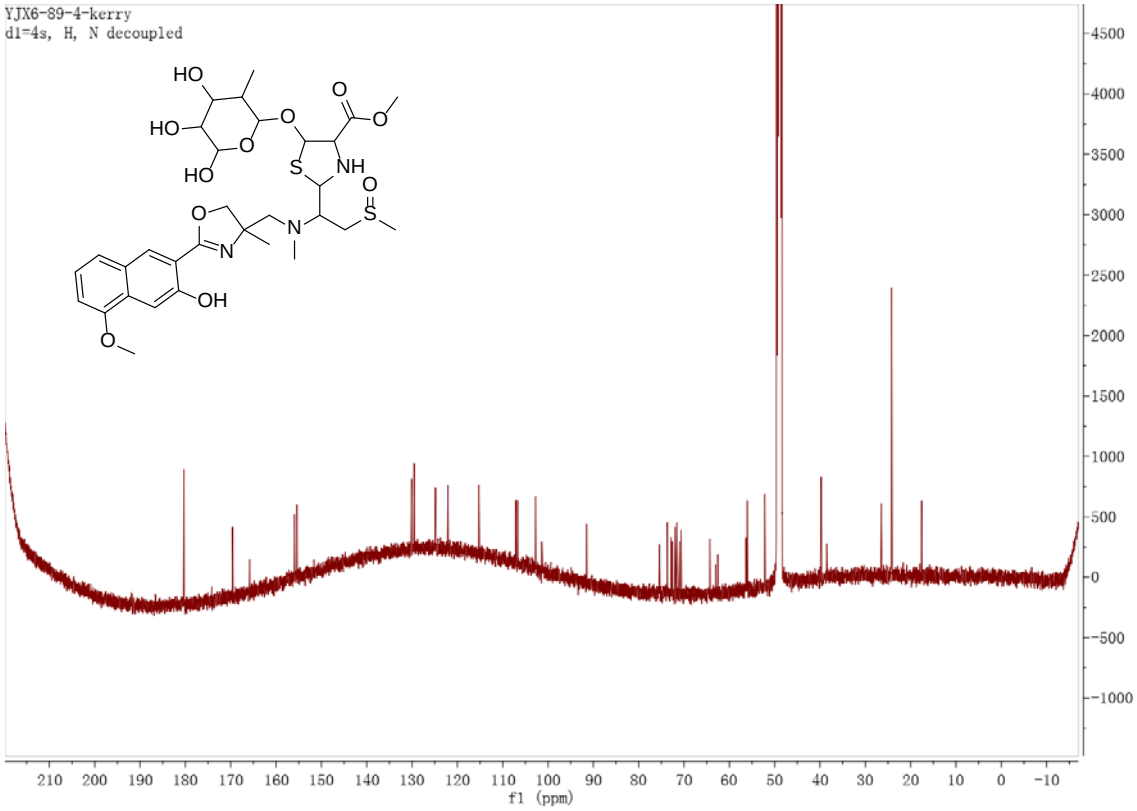


Figure S89. ^{15}N-coupled ^{13}C Spectrum of ^{15}N-enriched Ecteinamine B (**34**, 600 MHz, CD_3OD)

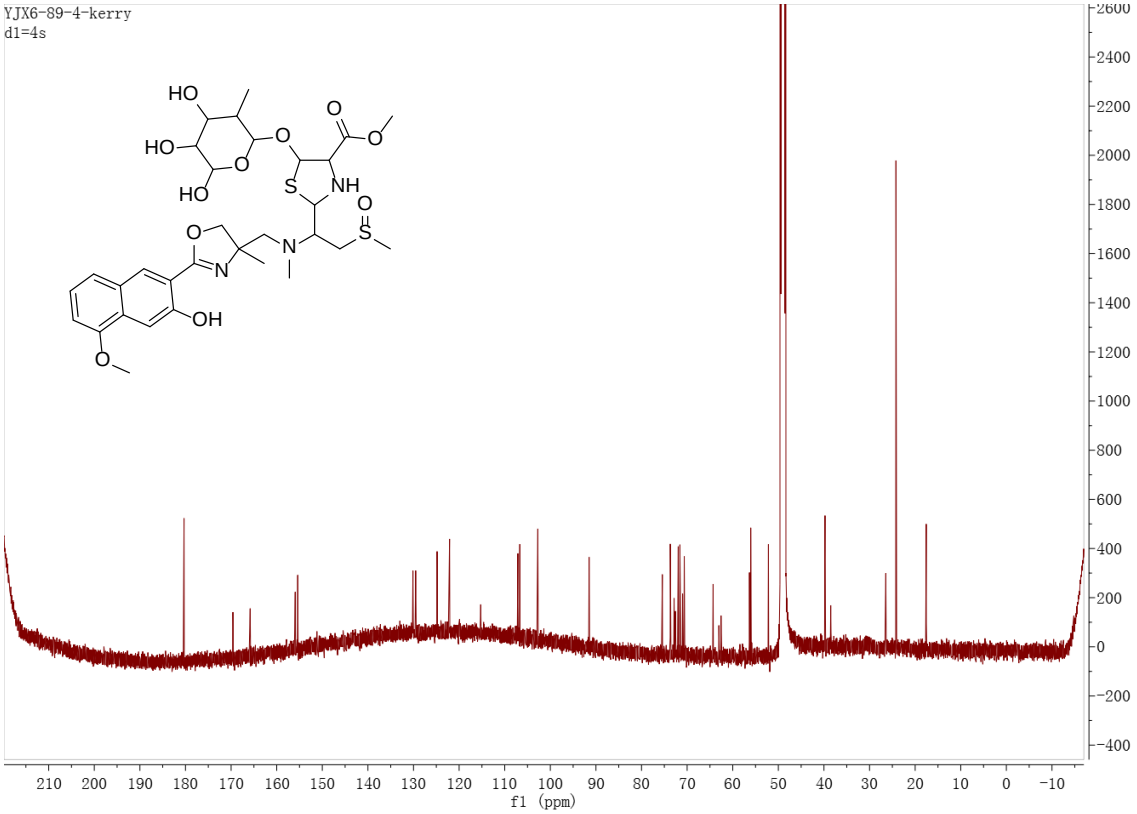


Figure S90. Positive Ion ESI-MRMS of Ecteinamine A (**33**)

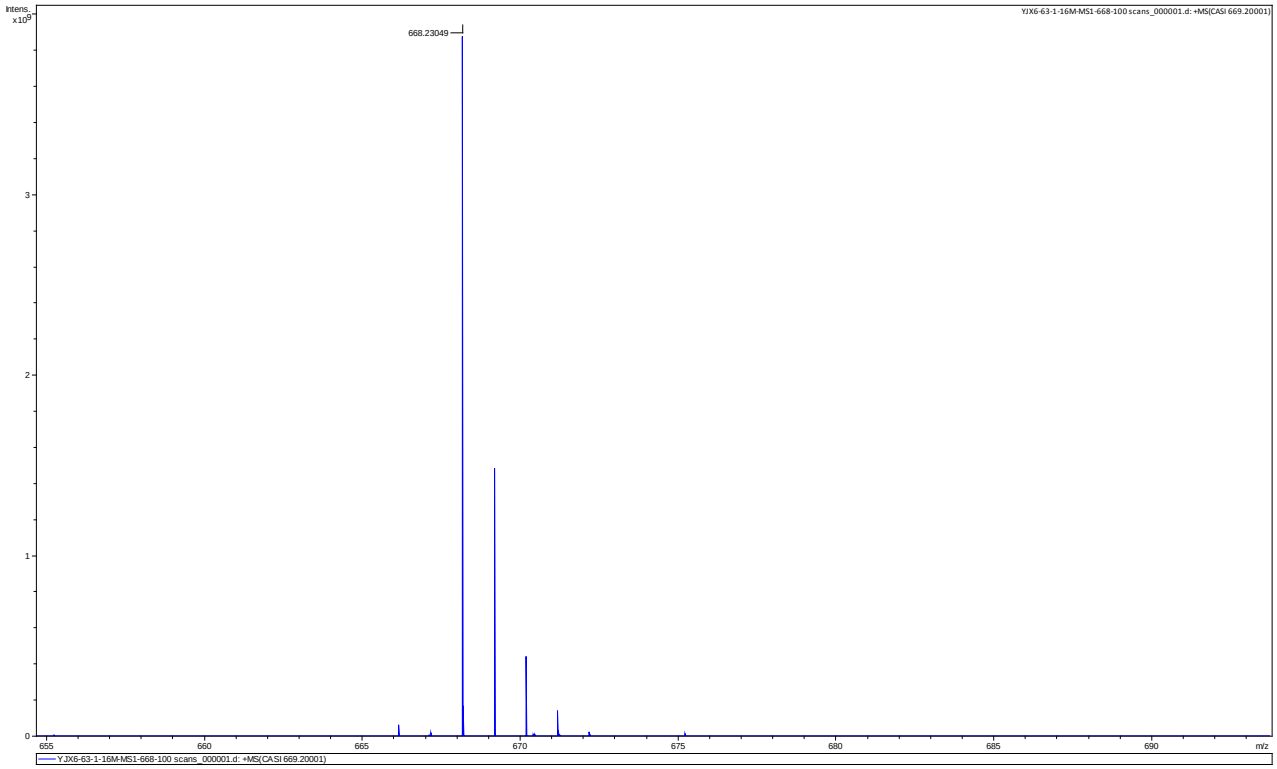


Figure S91. Positive Ion ESI-MRMS of Ecteinamine B (**34**)

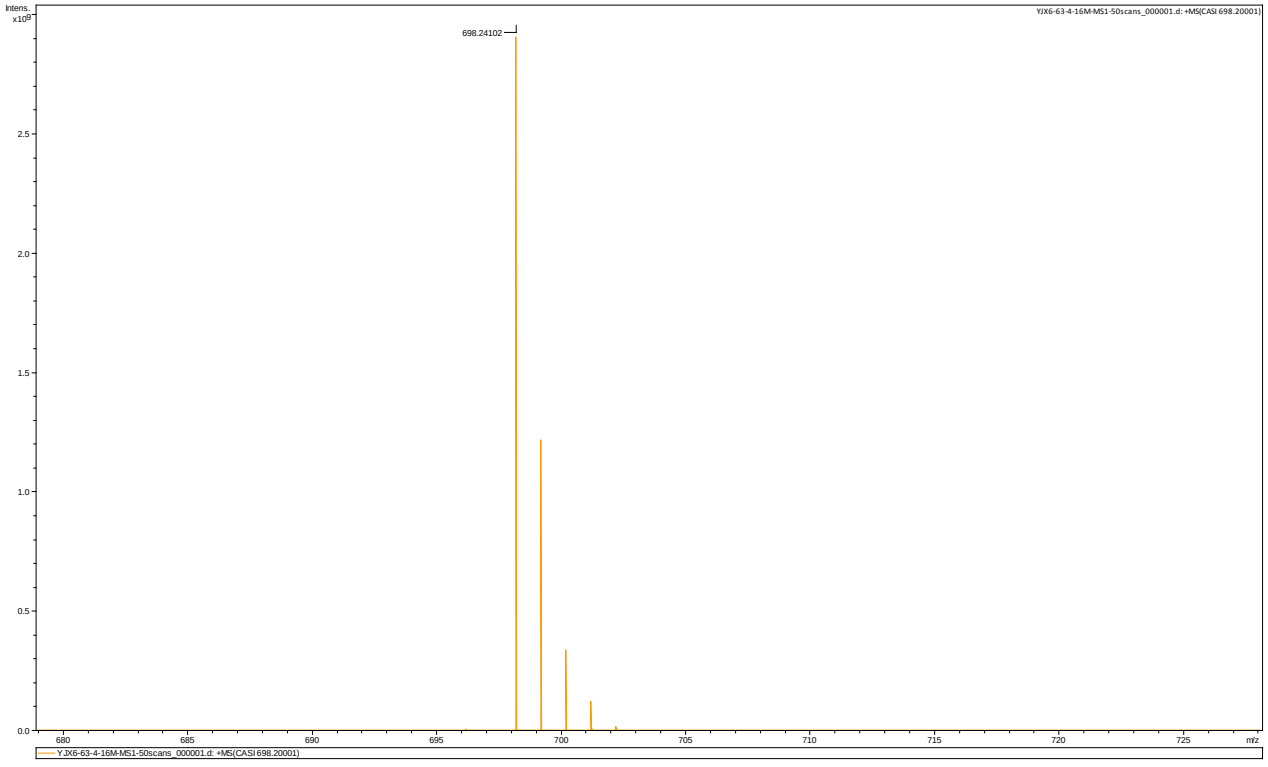


Figure S92. Positive Ion ESI-MRMS MS^2 of Ecteinamine B (**34**)

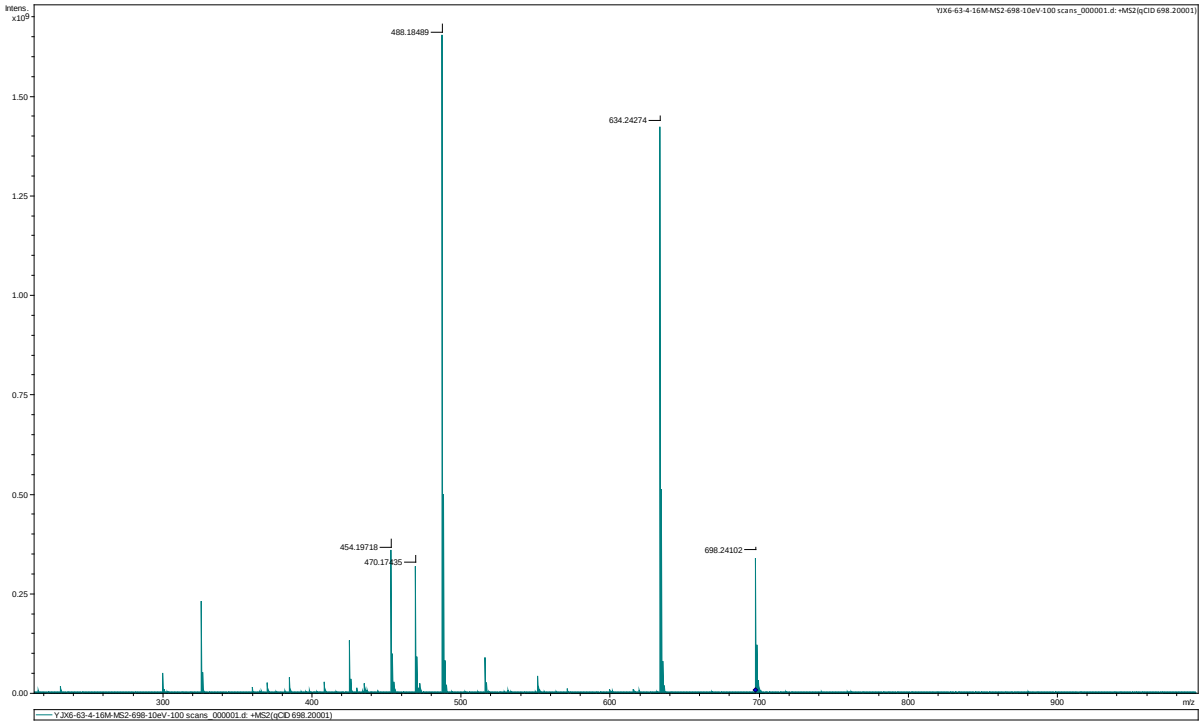


Figure S93. Positive Ion ESI-MRMS MS^3 of Ecteinamine B (**34**)

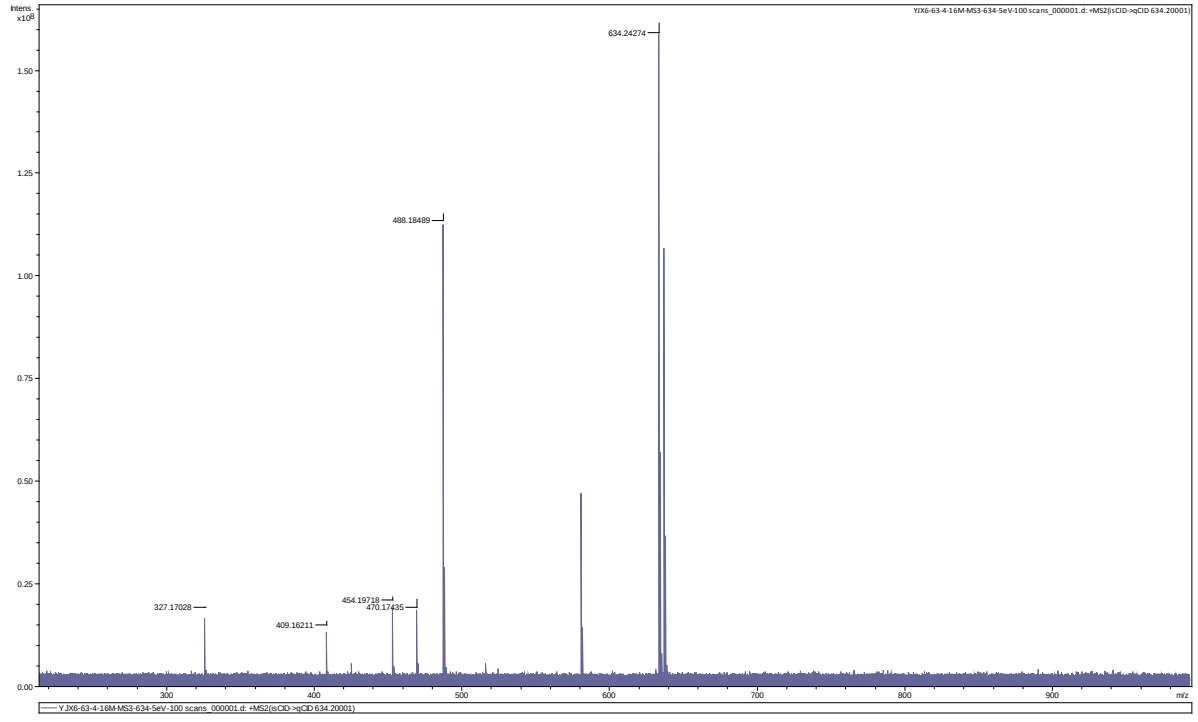


Figure S94. Positive Ion ESI-HRMS of ^{15}N enriched Ecteinamine B (**34**)

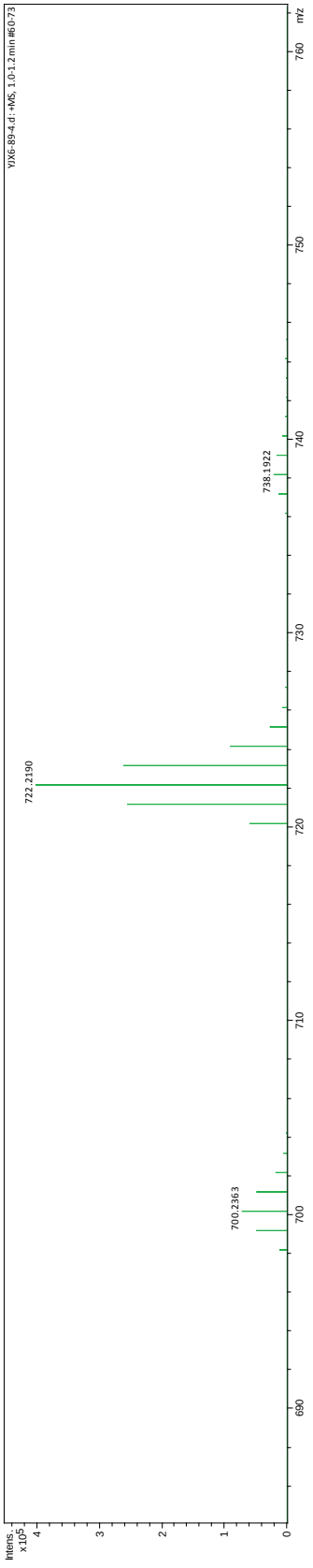
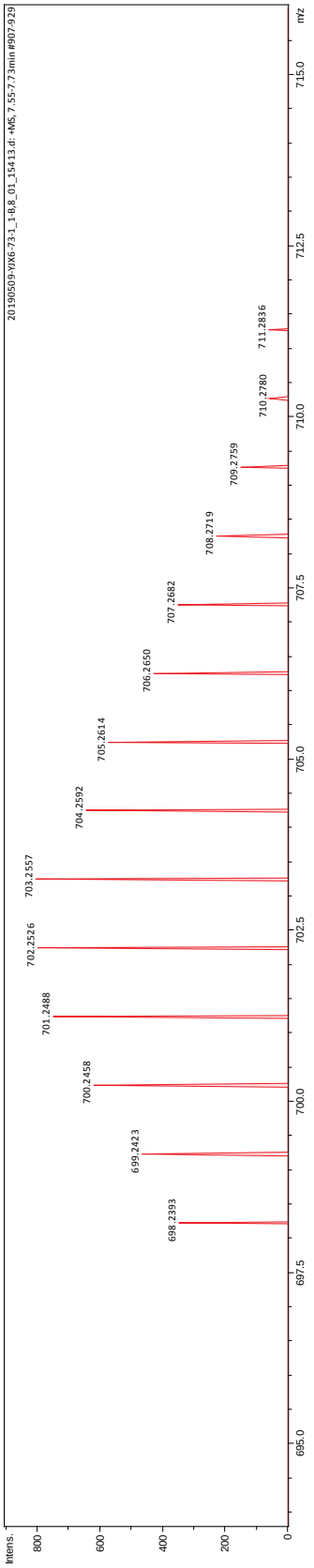


Figure S95. Positive Ion ESI-HRMS of ^{15}C enriched Ecteinamine B (**34**)



Appendix E:

Supplementary Data for Chapter 6

Figure S96. ^{1}H NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

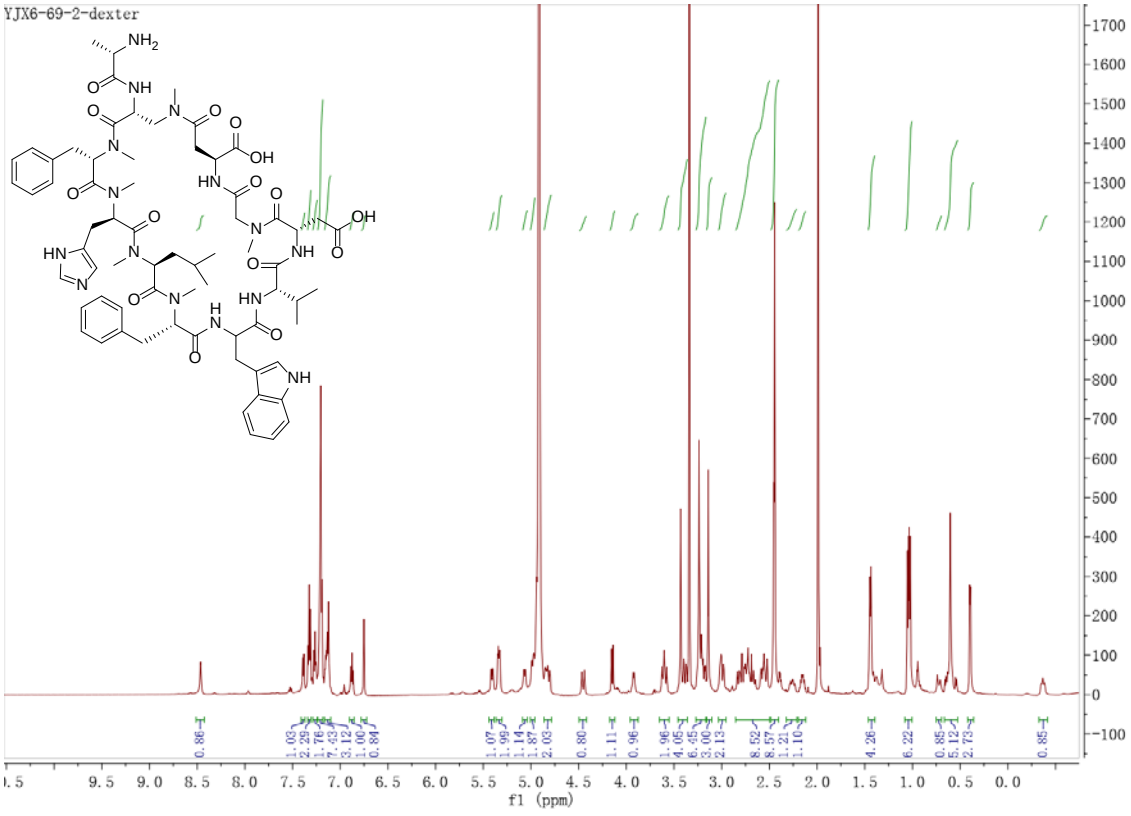


Figure S97. ^{13}C NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

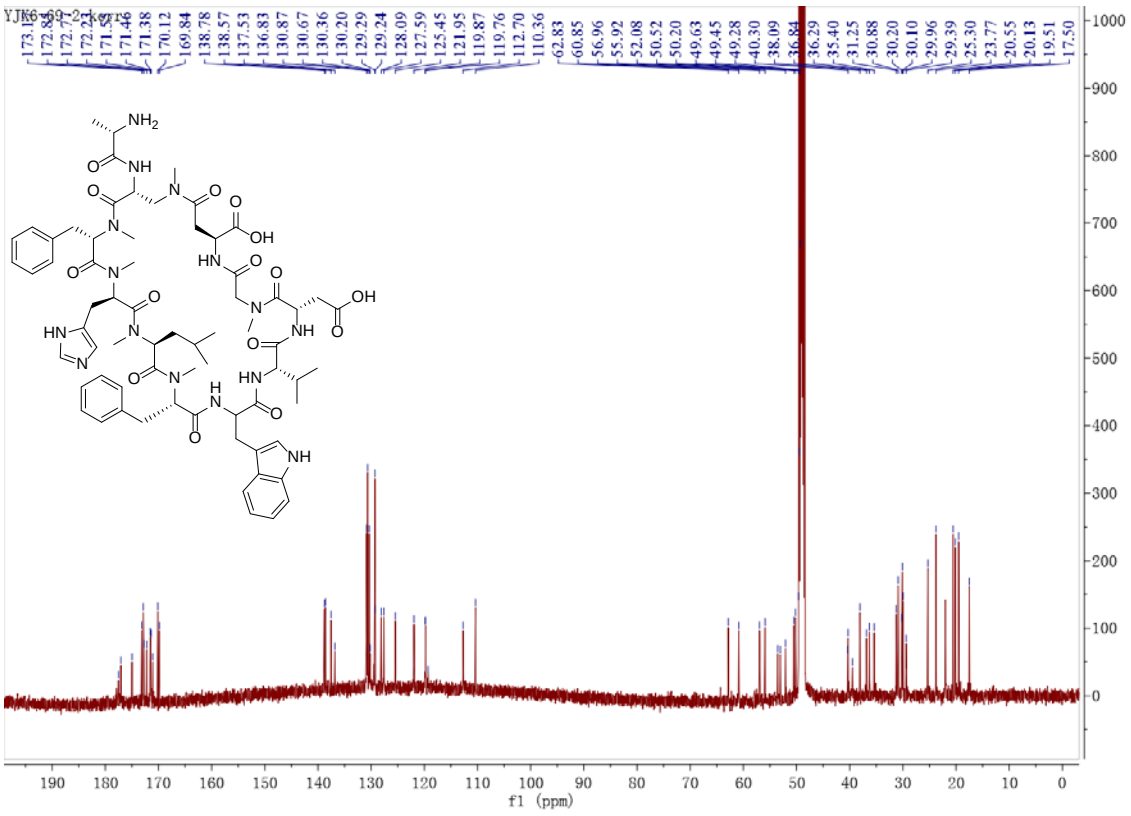


Figure S98. DEPT-90 ^{13}C NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

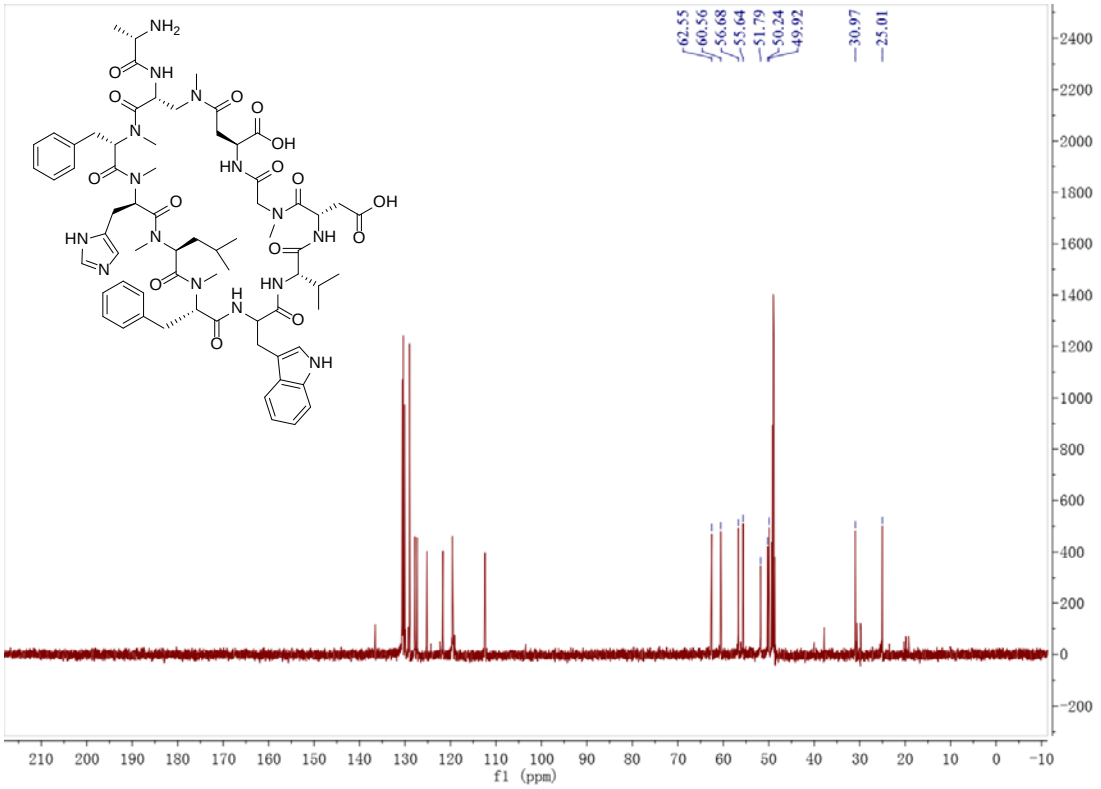


Figure S99. DEPT-135 ^{13}C NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

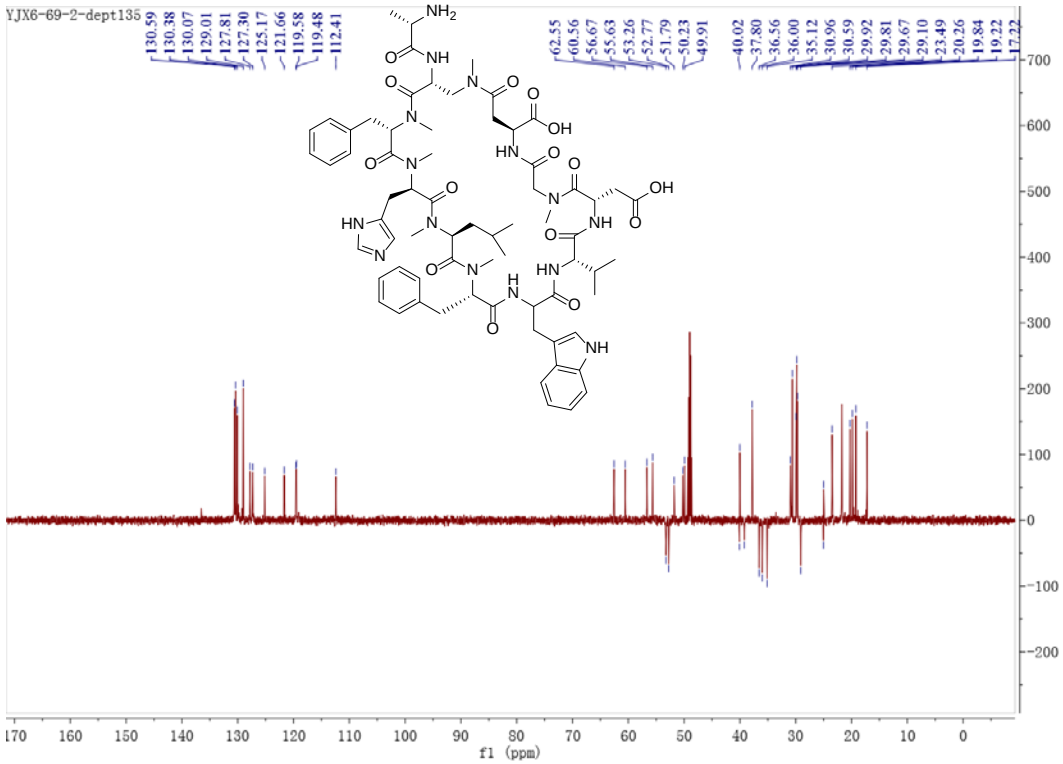


Figure S100. gCOSY NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

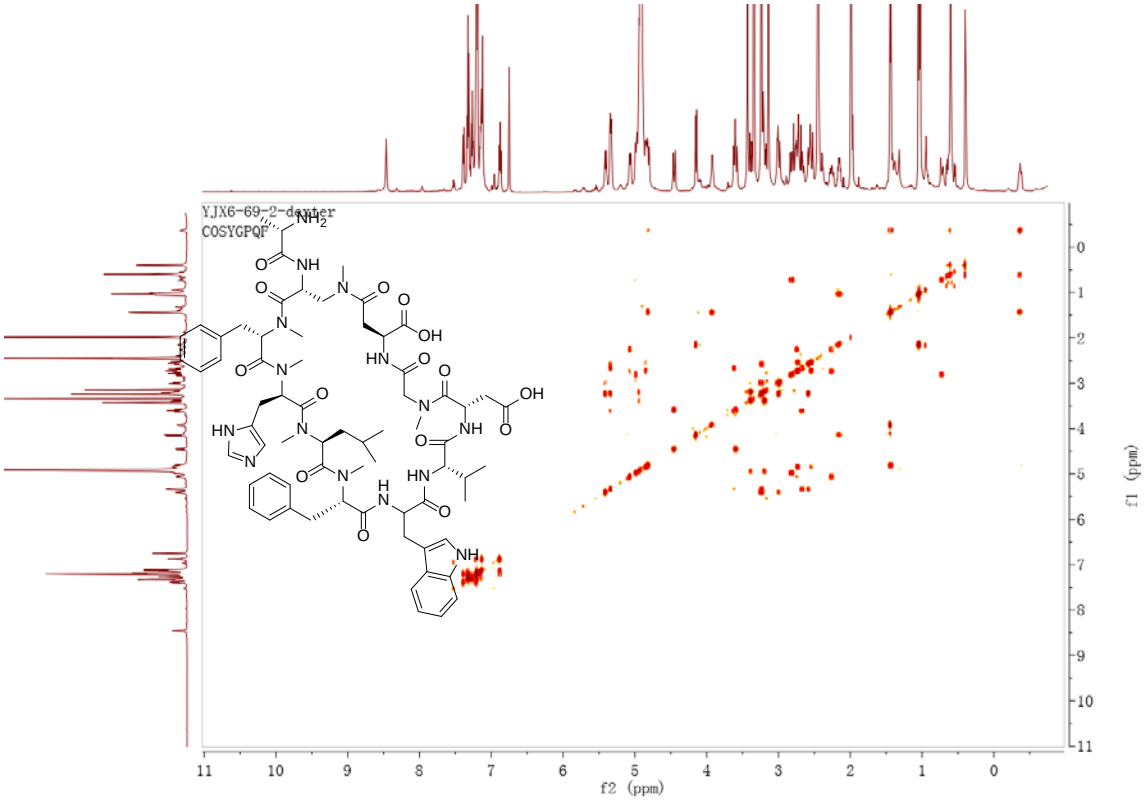


Figure S101. gHSQC NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

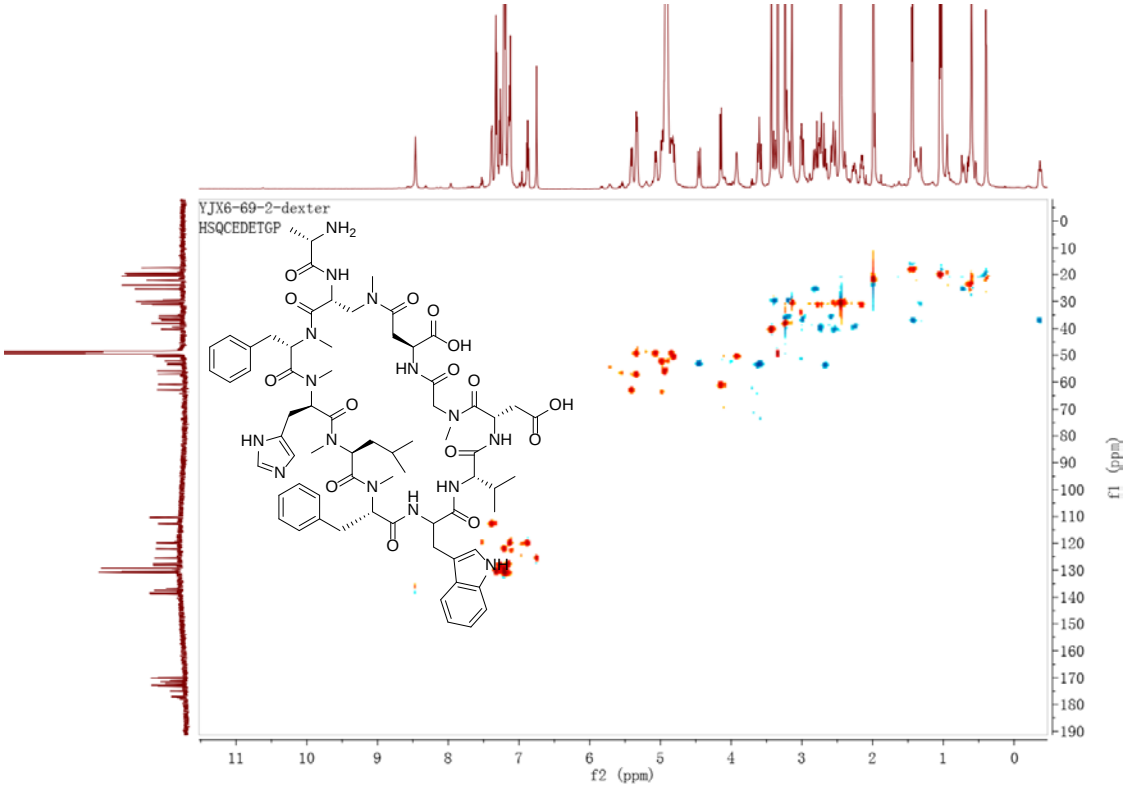


Figure S102. gHMBC NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OD)

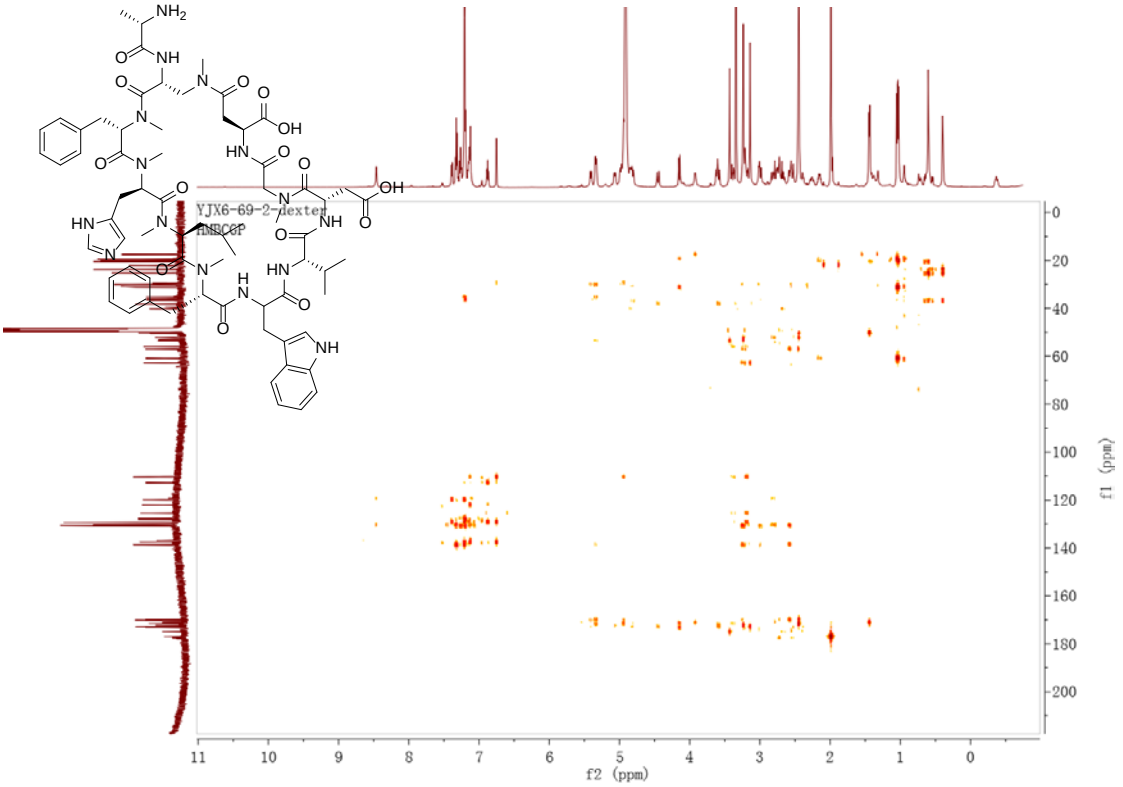


Figure S103. ^1H NMR Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OH)

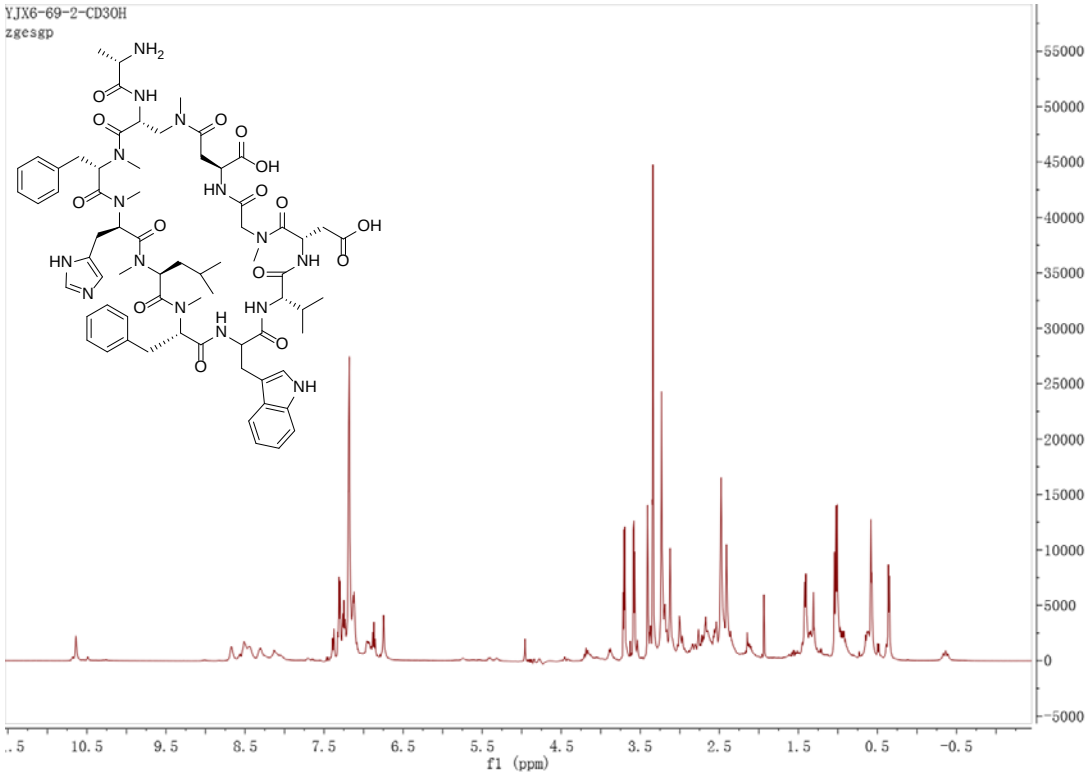


Figure S104. TOCSY Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OH)

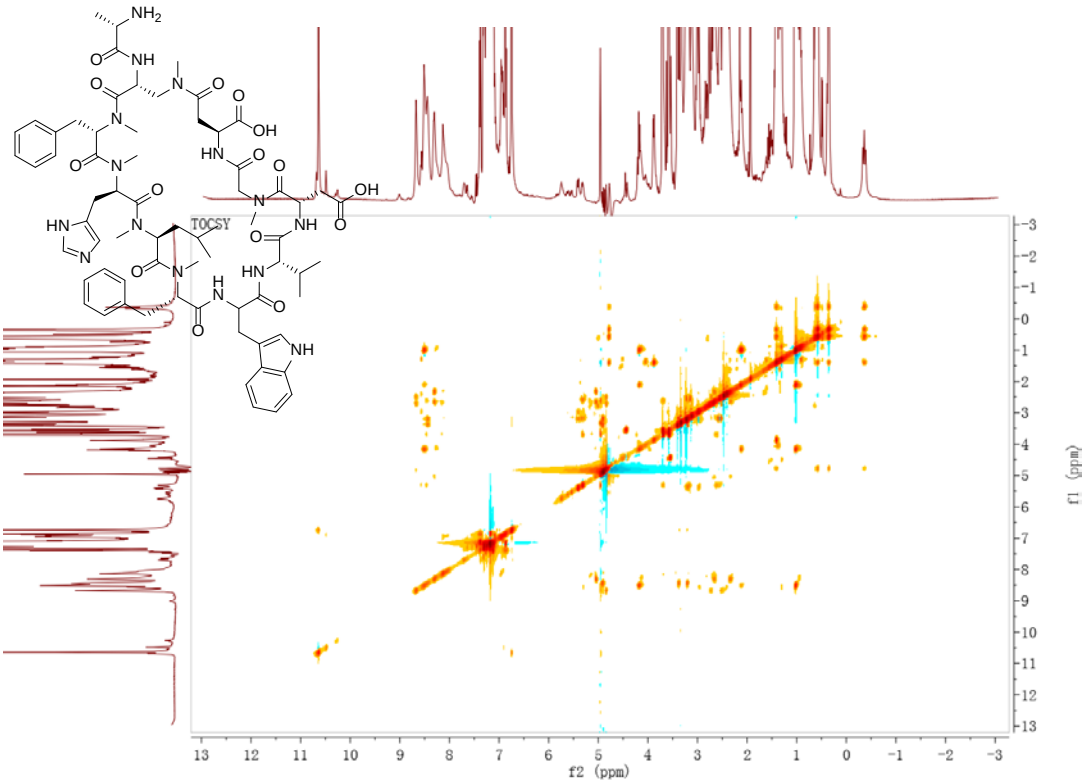


Figure S105. ROESY Spectrum of Streptnatamide A (**37**, 600 MHz, CD_3OH)

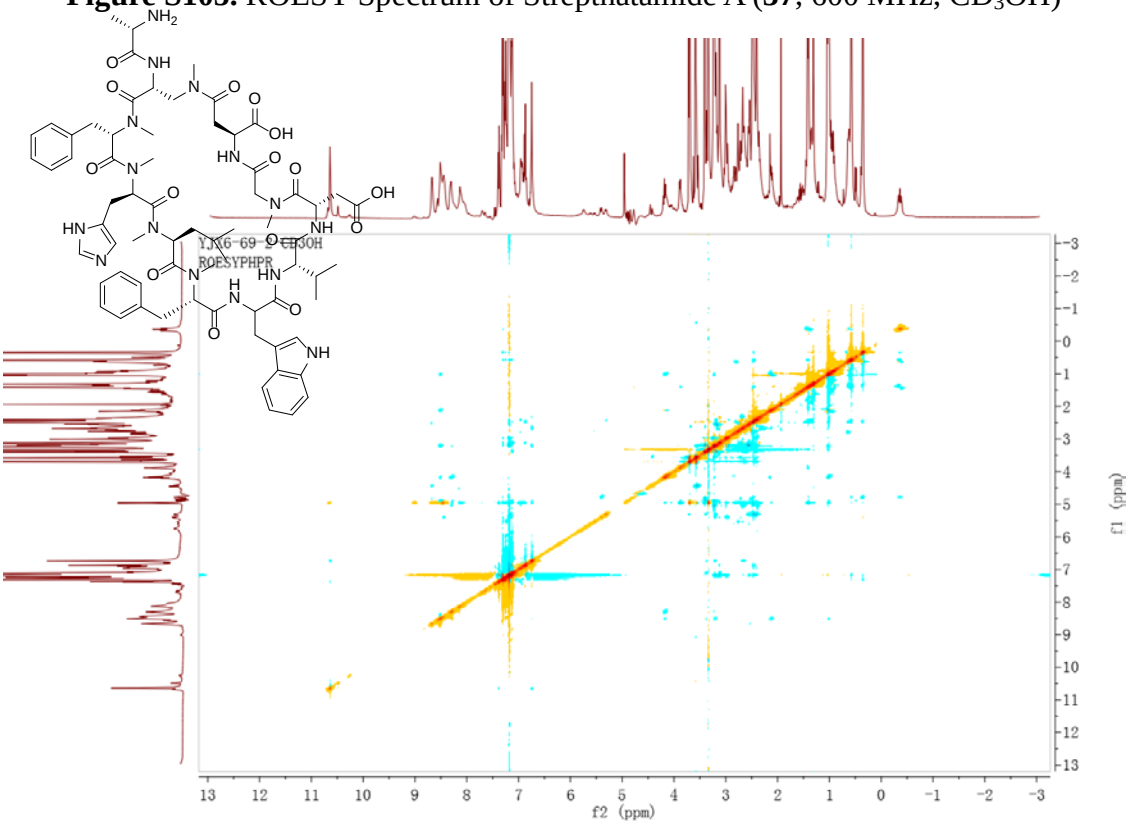


Figure S106. Positive Ion ESI-MRMS of Streptnatamide A (**37**)

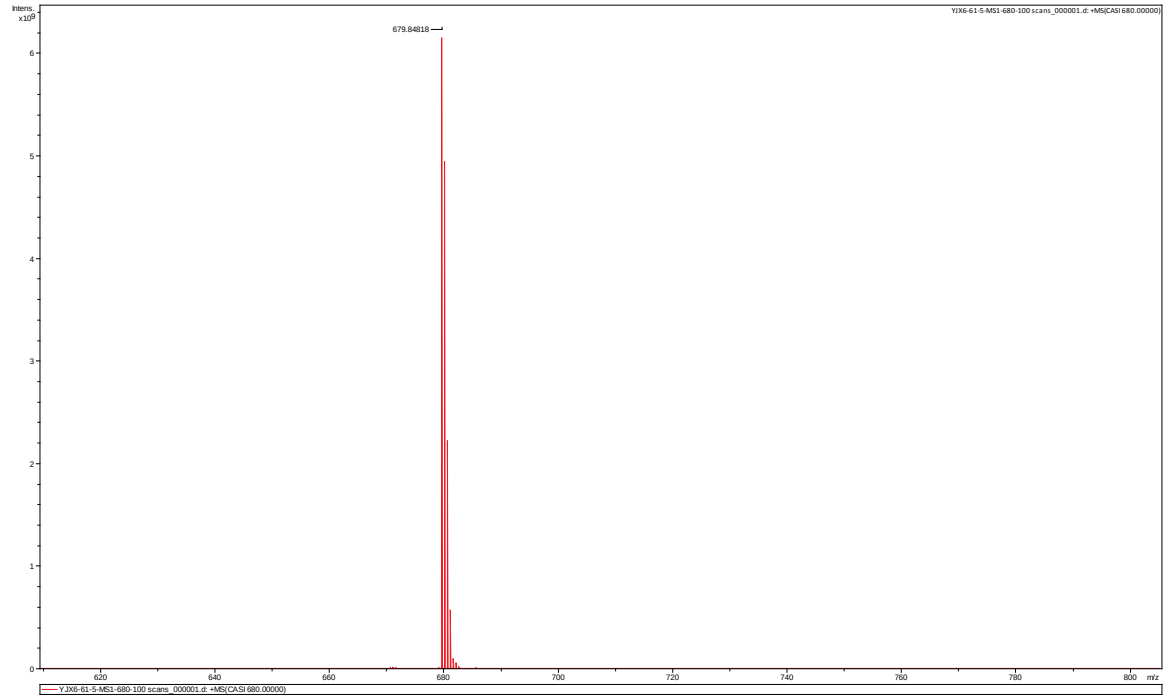


Figure S107. Positive Ion ESI-MRMS MS^2 of Streptnatamide A (**37**)

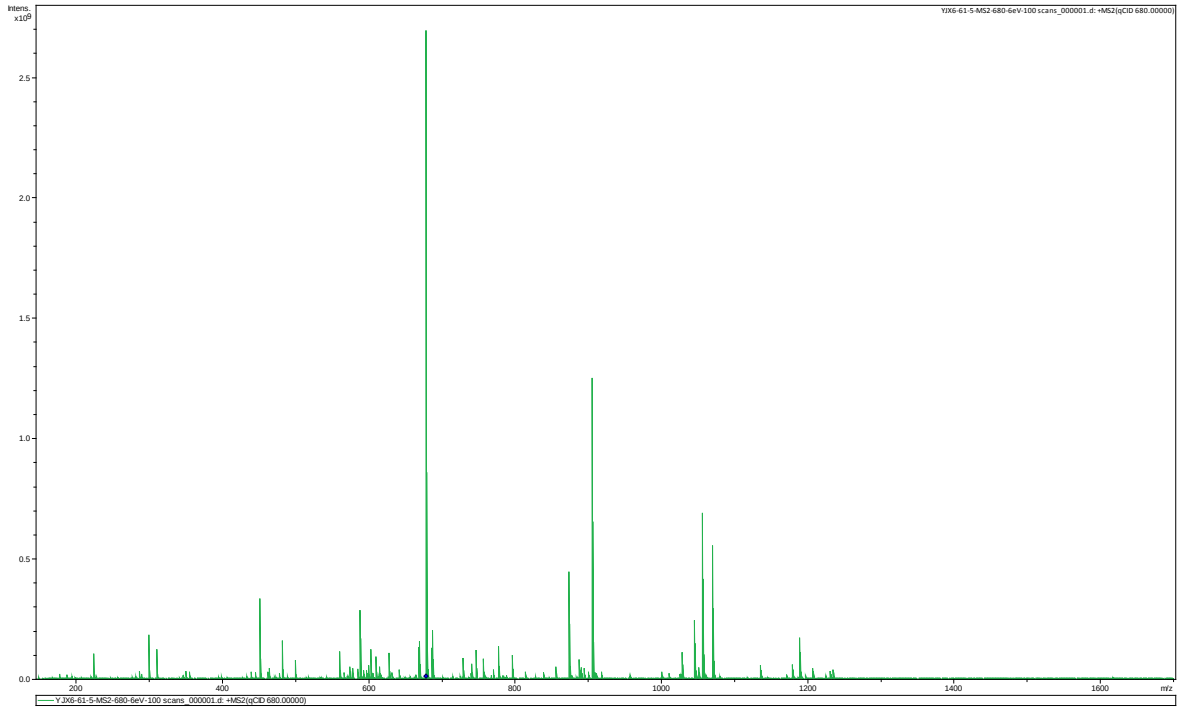


Figure S108. Advanced Marfey's Analysis of Streptnatamide A (**37**) (Red trace: hydrolysate + L-FDLA; Blue trace: hydrolysate + DL-FDLA)