MICROBIAL SECONDARY METABOLOMICS FOR NATURAL PRODUCT DISCOVERY

MICROBIAL SECONDARY METABOLOMICS FOR NATURAL PRODUCT DISCOVERY

By ASHRAF MOHAMED IBRAHIM, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy.

DESCRIPTIVE NOTE

DOCTOR OF PHILOSOPHY (2017) McMaster University

(Chemistry and Chemical Biology) Hamilton, Ontario, Canada

TITLE: Microbial Secondary Metabolomics for Natural Product Discovery: Development of metabolomic tools and strategies for the discovery of specialized metabolites from bacteria and endophytic fungi.

AUTHOR: Ashraf Mohamed Ibrahim, B.Sc. (Concordia University)

SUPERVISOR: Professor Brian E. McCarry, Professor Alfredo Capretta

NUMBER OF PAGES: xiii, 220, Appendix 366.

ABSTRACT

Microbial natural products have been a source for new drugs for many decades and are unrivaled in their capacity to generate not only future therapeutic agents, but also providing key agents for agricultural and industrial use. LC-MS/MS based metabolomic tools and technologies have been developed that can rapidly dereplicate nonribosomal peptides and statistically identify related congeners in an automated nontargeted process from complex natural product extracts with nanogram sensitivity. This data-base search approach is designed to handle linear, cyclic and cyclic-branched nonribosomal peptides from proteinogenic and nonproteinogenic amino acids without genomic data or traditional bioactivity directed fractionation. Chemometric work-flows combined with a comprehensive metabolomic guided discovery strategy were used to profile the chemical space of a diverse collection of understudied fungal endophytes from fruiting plants. This approach allowed for the prioritization of unique isolates and for the focused discovery, isolation and characterization of distinct outlier metabolites by LC-SPE, 1D and 2D NMR, HRMS and single crystal X-ray analysis. These metabolomic tools and strategies have led to the discovery and characterization of 35 new and over 40 known natural products, many of which are biologically active. This thesis with enabling metabolomic tools and novel discoveries has demonstrated the utility of these analytical methodologies as an effective strategy for the untargeted discovery of new natural products from bacteria and endophytic fungi.

ACKNOWLEDGMENTS

Natural products discovery has become one of my great passions since working at Ecopia BioSciences and Thallion Pharmaceuticals. I would first like to thank Dr. James B. McAlpine (Jim) who first introduced me to natural products and Dr. Dan Sørensen, Dr. Arjun H. Banskota and Dr. Faustinus Yeboah (Kobi) who taught me the drug discovery and development process from extracting microbial cultures to isolating and characterizing complex compounds and in pharmacokinetic and drug metabolism profiling by LC-MS/MS analysis.

I would like to thank Dr. Nathan Magarvey who invited me from Thallion Pharmaceuticals to McMaster University and for the development of iSNAP. I would like to give a special thank you to my supervisors Dr. Brian E. McCarry and Dr. Alfredo Capretta (Fred) who gave me the guidance, flexibility, support to purse my passion in analytical and natural products chemistry. Brian was awesome; he was a great mentor and he introduced me to metabolomics and inspired many of the accomplishments presented within this thesis. We had numerous discussions on science, including the development of iSNAP's technology platform and in applying analytical and metabolomic strategies to target novel natural products from microbes. Fred is amazing, he fully supported all my research projects and provided valuable advice, mentorship and his time, always being available. Many of the new discoveries outlined in this thesis would not have been possible otherwise. I would like to also thank Dr. Bin Ma and Lian Yang for their amazing work and support in the development of the iSNAP platform. This was an incredible project and was very successful even though Lian and I operated from two different universities, lots of meetings and discussions. I would also like to thank Dr. Mark W. Sumarah from Agriculture and Agri-Food Canada for collaborating on the endophyte project. Mark provided me with an awesome research project and introduced me to fungal endophytes and the mycology world. He provided endless support and we had numerous and great discussions on all aspects of research and a great collaborator. I would like to thank Dr. Dan Sørensen who taught me how to expertly solve NMR structures. Dan and I had many discussions on research and he helped guide the endophyte project along the way with Dr. Sumarah and Dr. J.D. Miller.

I would like to give a special thanks to Dr. J. David Miller for his critical support, mentorship with Mark, and expertise on fungal endophytes. I would also like to thank Dr. Hilary A. Jenkins for all the crystal X-ray analysis and discussion on crystallization. I would like to thank my committee members Dr. Gerry Wright and Dr. John Brennen who asked the tough questions and provided valuable advice. I would also like give a special thank you to Fan Fei from the McCarry lab for showing me how to perform metabolomic analysis and for always being there to help. I also thank Dr. M. Kirk Green for the many mass spectrometry discussions. Most importantly, I would like to thank the McCarry lab members Fan Fei, Dave Bowman, Sujan Fernando, Jonathan Bloomfield (Junior), Vi Dang, Roger Luckham, and Kenneth Chalcraft for their amazing support and lab spirit.

I would like to thank my family and friends who have been very supportive and for all the out-of-town visits, long-distance phone calls, discussions and trips. Finally, I would like to thank my wife Jennie, who has provided endless love and support. You are my greatest joy and inspiration. It's been an incredible journey and I dedicated this to you.

PREFACE

This chapter thesis contains the published and unpublished research works by Ashraf Ibrahim at McMaster University. The works herein, focus on the development of a novel chemoinformatic technology platform to rapidly and statistically screen microbial natural product extracts for nonribosomal peptides. Secondly, applying a comprehensive chemoinformatic and metabolomic guided strategy for the focused discovery of novel specialized metabolites from endophytic fungi of understudied fruiting plants. Chapter 2 and 3 was in collaboration with Dr. Nathan Magarvey, Chad Johnston, Michael Skinnider (McMaster University) and Dr. Bin Ma, Lian Yang (University of Waterloo) and Dr. Xiaowen Liu (Indiana University-Purdue University). Chapters 4-6 was in collaboration with Dr. Brian E. McCarry, Dr. Alfredo Capretta, Dr. Dan Sørensen, Dr. Fan Fei (McMaster University) and Dr. Mark W. Sumarah, Dr. Keith A. Seifert, Dr. Joey Tanney and Tim McDowell (Agriculture and Agri-Food Canada) and Dr. J.D. Miller (Carleton University).

The following research publication by Ashraf Ibrahim is related to Chapters 4-6 on fungal endophytes, but was not included in the thesis because of length limitations.

Ibrahim, A., Sørensen, D., Jenkins, H.A., McCarry, B.E., Sumarah, M.W. New diplosporin and agistatine derivatives produced by the fungal endophyte *Xylaria sp.* isolated from *Vitis labrusca*, Phytochemistry Letters, 9, 179-183 (2014).

PUBLICATIONS

1. Ibrahim, A., Sørensen, D., Jenkins, H.A., Ejim, L., Capretta, A., Sumarah, M.W. Epoxynemanione A, nemanifuranones A-F, and nemanilactones A-C, from *Nemania serpens*, an endophytic fungus isolated from Riesling grapevines. Phytochemistry 140, 16-26 (2017).

2. Burgess, K. M.N., Ibrahim, A., Sørensen, D., Sumarah, M.W. Trienylfuranol A and trienylfuranone A-B: metabolites isolated from an endophytic fungus, *Hypoxylon submoniticulosom*, in the raspberry *Rubus idaeus*. The Journal of Antibiotics, 70, 721-725 (2017)

3. Yang, L., Ibrahim, A., Johnston, C.W., Skinnider, M.A., Ma, B., Magarvey, N. Exploration of nonribosomal peptide families with an automated informatic search algorithm. Cell Chemical Biology, 22 (9), 1259-1269 (2015).

4. Ibrahim, A., Sørensen, D., Jenkins, H.A., McCarry, B.E., Sumarah, M.W. New diplosporin and agistatine derivatives isolated from *Xylaria* sp., a fungal endophyte of *Vitis labrusca*. Phytochemistry Letters, 9, 179-183 (2014).

5. Richardson, S.N., Walker, A.K., Nsiama, T., McFarlane, J., Sumarah, M.W., Ibrahim, A., Miller, J.D. Griseofulvin-producing *Xylaria* endophytes of *Pinus strobus* and *Vaccinium angustifolium*: evidence for a conifer-understory species endophyte ecology. Fungal Ecology, 11, 107-113 (2014).

6. Johnston, C.W., Wyatt, M.A., Li, X., Ibrahim, A., Shuster, J., Southam, G., Magarvey, N.A. Gold biomineralization by a secondary metabolite from a gold-associated microbe. Nature Chemical Biology, 9, 4, 241-3 (2013).

7. Ibrahim, A., Yang, L., Johnston, C., Liu, X., Ma, B., Magarvey, N.A. Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. Proceedings of the National Academy of Sciences, 109, 47, 19196-19201 (2012).

8. Johnston, C., Ibrahim, A., Magarvey, N. Informatic strategies for the discovery of polyketides and nonribosomal peptides. Medicinal Chemistry Communications, 3, 932-937 (2012).

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 4

CHAPTER 5

CHAPTER 6

List of ABBREVIATIONS

A	Adenylation domanin
AMDIS	Automated mass spectrometry deconvolution and identification
C	Condensation domain
CID	Collision-induced dissociation
CLSI	Clinical Laboratory Standards Institute
COSY	Correlation spectoscopy
DDA	Data dependent acquisition
DESI	Desorption electrospray ionization
DH	Dehydratase
DS	Direct-sequence ions
ER	Enoyl reductase
ESI	Electrospray ionization
FDA	Food and Drug Administration
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum
GNPS	Global Natural Products Social Molecular
HILIC	Hydrophilic interaction liquid chromatography
HMBC	Heteronuclear Multiple Bond Correlation
HRMS	High resolution mass spectrometry
hSFs	Hypothetical spectral fragments
HSQC	Heteronuclear single quantum coherence
HTS	High-throughput screening
IS	Internal standards for peak area normalization
iSNAP	Informatic Search for Natural Products
ITS	Internal transcribed spacer
KR	Keto reductase
KS	Keto-acyl synthase
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
m/z	Mass-to-charge
MALDI	Matrix assisted laser desorption ionization
MIC	Mimimum inhibitory concentration
MRSA	Methicillin resistant Staphylococcus aureus
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS-CPA	MS-Cyclic Peptide Annotation Program
NDS	Non-direct sequence ions
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NRPS	Nonribosomal peptide synthesis
OPLS-DA	Orthogonal partial least-square discriminative analysis
PCA	Principle component analysis
PKS	Polyketide synthesis

Q^2	Prediction statistic
RE	Reductase domain
ROESY	Rotational nuclear Overhauser Effect Spectroscopy
SMILES	Simplified molecular-input line-entry system
SPE-NMR	Solid phase extraction -NMR
TE	Thioesterase domain
TOF	Time of flight
UHPLC	Ultra-high pressure liqid chromatography
UV	Ultra violet
VIP	Variable influence on projection
VRE	Vancomycin resistant Enterococcus

CHAPTER ONE

INTRODUCTION

In Chapter 1, the Introduction contains sections that have been taken from and/or adapted from, the following review, where the author contributed to the section.

Johnston, C., Ibrahim, A., Magarvey, N. Informatic strategies for the discovery of polyketides and nonribosomal peptides. Medicinal Chemistry Communications, 3, 932-937 (2012).

Natural Product Discovery

Over the past century, microbial natural products have been at the forefront of the pharmaceutical world and have been used historically as therapeutic agents against disease and as natural poisons. Traditionally, therapeutic medicines had been primarily sourced from plants with notable examples such as aspirin (**2**), derived from willow bark, and morphine (**3**) derived from opium poppy, both used for pain and fever management.[1] However, with the serendipitous discovery of penicillin G (**1**) by Fleming in 1928, a new golden era for microbial discovery was initiated as some of the most life threatening diseases of our time, such as syphilis and staphylococci infections, could be successfully treated.[2] The discovery of penicillin was transformative as it brought microbes to the frontier of modern medicine and marked the beginning of their industrial exploitation as a source for discovery.[3,4]

In this pre-genomic era, some of the most widely used therapeutic drugs had been discovered from bacterial, fungal and endophytic sources; for example, vancomycin (glycopeptide antibiotic, 1953-54) (**4**), erythromycin (macrolide antibiotic, 1952) (**5**), amphotericin B (antifungal, 1955) (**6**), doxorubicin (anticancer, 1950's) (**7**), cyclosporin (immunosuppressant, 1969-72) (**8**), taxol (anticancer, 1971) (**9**), and mevastatin (hypolipidemic, 1970's) (**10**).[5] The majority of these small molecules were discovered using traditional bioactivity guided isolation methods and were often the most potent and abundant compounds. The discovery of low abundance compounds that were masked or below the detection limits was considerably more difficult and often impossible.

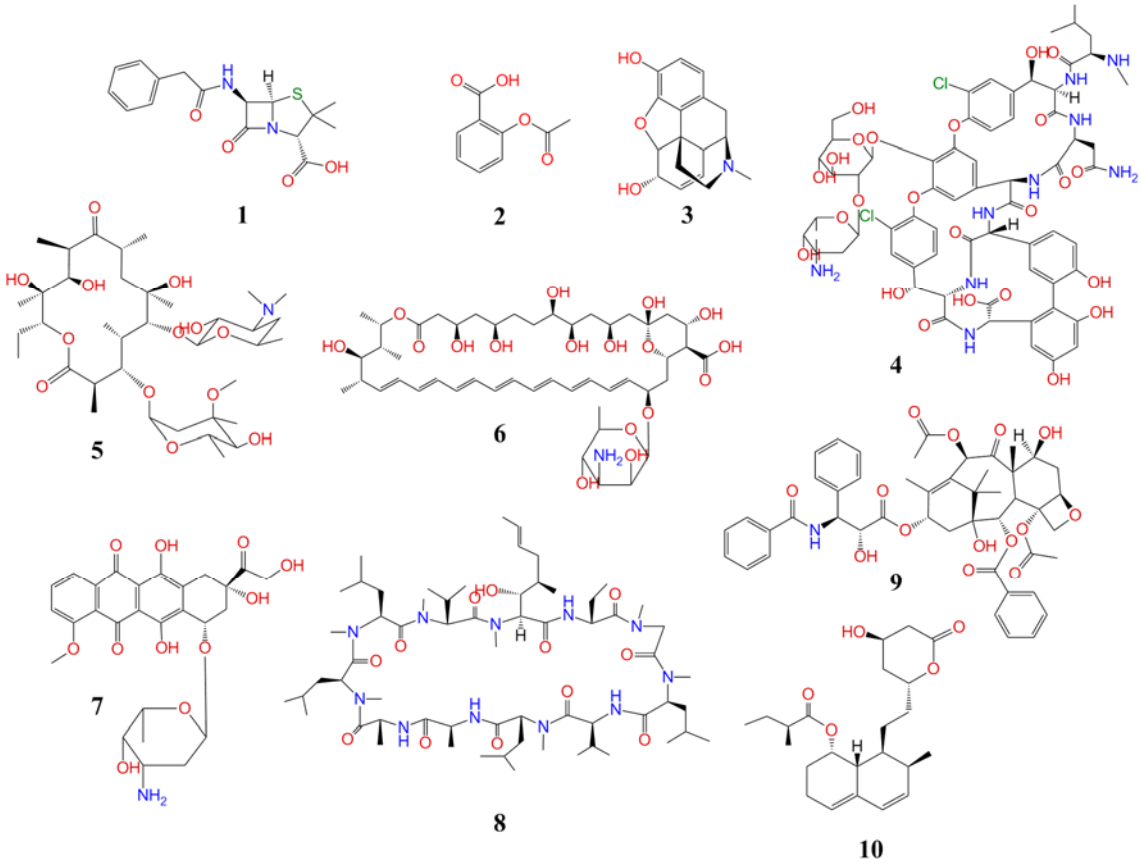


Fig.1. Therapeutic natural products derived from plant and microbial sources discovered during the pre-genomic era; penicillin G (1), aspirin (2), morphine (3), vancomycin (4), erythromycin (5), amphotericin B (6), doxorubicin (7), cyclosporin (8), taxol (9), mevastatin (10).

While combinatorial and synthetic libraries have yielded few or a limited number of novel therapeutic agents, microbial natural products have been an excellent source for new drugs as they have a high affinity to their natural biological target(s) or substrate, and their chemical scaffolds can be more readily exploited for lead optimization or in next-generation development. With the rise of bacterial resistance, the need for new antimicrobial agents has become crucial. For example, effective antibiotic choices for methicillin-resistant Staphylococcus aureus (MRSA) and vancomycin-resistant Enterococcus (VRE) super bug infections are limited to only a handful of treatments.[6] These facts also underscore the need for advancing more natural product

agents as many of our "go-to" drugs are showing themselves ineffective with more virulent strains as bacteria and pathogenic organism are continuously evolving and developing resistance to many commonly used and over prescribed drugs.[6,7]

A constant challenge to the drug discovery paradigm, however, has been the re-isolation of known compounds through traditional bioactivity guided screening approaches. In the case of antimicrobials, Linezolid (synthetic oxazolidinones) (**11**) is active against MRSA and VRE, Daptomycin (nonribosomal lipopeptide) (**12**) and Retapamulin (pleuromutilins) (**13**) are active against Gram-positive bacteria.[6] They were FDA approved in 2000, 2003 and 2007 respectively. The decreased rate of discovery can be attributed primarily to the high rates of re-discovery and re-isolation of known biologically active compounds. The ability to dereplicate these compounds is often a key rate-limiting step to the discovery of novel compounds.[8]

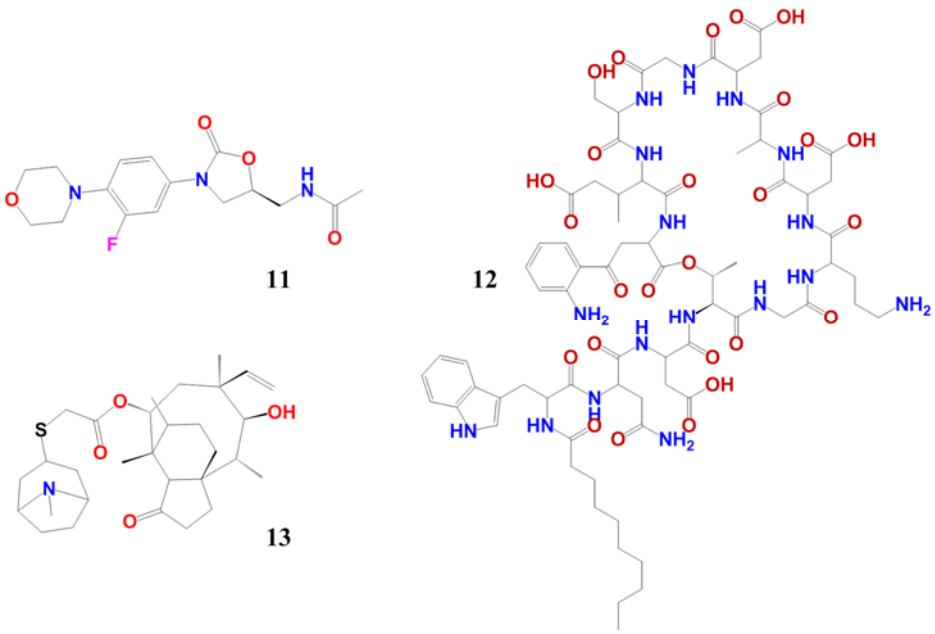


Fig.2. Potent antimicrobial agents active against Gram-positive bacteria

Classically, dereplication approaches often incorporate crude microbial extract screening work-flows based on liquid-chromatography (LC) coupled with UV absorbance detection and

downstream verification with in-house or commercial databases such as Antibase, Dictionary of Natural Products and Metlin. In this way, comparisons between an analyte's eluting retention times can be indexed with known standards.[9]

Modern approaches have advanced the dereplication process by incorporating mass spectrometry with improved accurate mass measurements and isotope analysis at low nanogram levels for high confidence molecular formula determination. Advances in mass spectrometer technologies that allow for an increased level of mass accuracy needed (0.5-5ppm), resolving power (30,000 - 1,000,000 at full width at half maximum (FWHM)) as well as acquisition turn-around times for high-throughput screening (HTS) have further enabled the dereplication process. Indeed the entire natural products isolation field has benefitted from new instrument types (FT-ICR, Orbitrap, q-TOF), new sources for sensitive compounds (APCI, DESI, ESI, MALDI), and new methods for molecular fragmentation (CID, ECD, ETD, HCD) allowing for the elucidation of isolated compounds with diverse chemical architects. Equally important are technology developments in separation sciences and column chemistries (C18, HILIC, Amino) that now allow for low 1-3 μm particles sizes (Phenomenex Kinetex Core-shell) that in turn have afforded more efficient columns (>400,000 plates/m), robust separations of compound peaks, and significantly reduced run times (UPLC).[8,10]

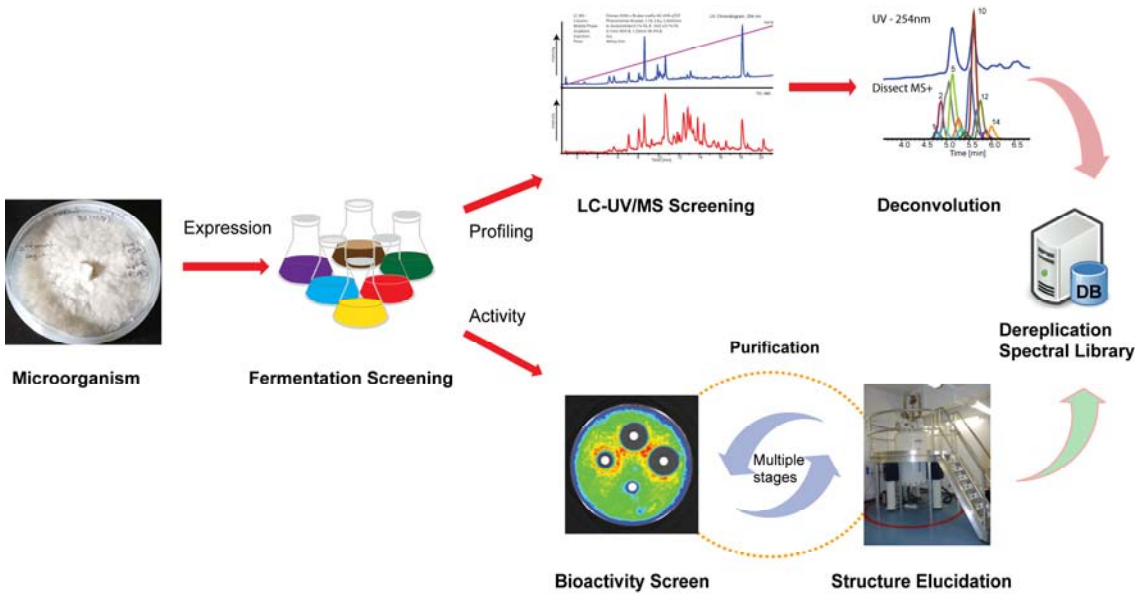


Fig.3. Traditional bioactivity-guided discovery process and work-flow.

Unfortunately, many of the traditional work-flows are only partially-automated, requiring manual verification of LC, UV and MS data, with reliable identifications requiring targeted compounds to be isolated, purified and structural elucidated by 1D and 2D NMR for validation (Figure 3). Improved compound identifications can be achieved and correlated using tandem MS data (MS/MS) and through use of computer generated spectral libraries of *in silico* MS spectra. However, many workflows do not incorporate MS/MS data, as applications lack data mining algorithms and no commercially available MS/MS spectral databases of microbial natural products currently exist. As such, time consuming manual efforts are needed to determine the structures of known and unknown compounds. In addition, downstream isolation and characterization processes, microbial sourcing from primarily prolific taxa such as actinomycetes, and limited targeting tools for selective cell-based screening of lead candidates have contributed to the decline in the discovery of novel therapeutic compounds.

In the post-genomic era, advancements in bio-informatic tools and technologies have expanded our capacity to rapidly sequence whole microbial genomes, with sequencing rates

doubling every 20 months, thus providing new insights into the biosynthetic potential of microbes.[11] DNA sequencing data from whole genomes of well-studied producers (*Streptomyces coelicolor* at 8.7 Mbp, *Streptomycetes avermitilis* at 9 Mbp, for example) have shown that bacteria have a far greater capacity to produce secondary metabolites then what was previously known (at upwards of more than 10 fold of what was previously estimated).[12] The biosynthetic machinery and secondary metabolite production has been underestimated, as more and more proteins and DNA encoding gene clusters are being identified, suggesting a vast potential of novel discoveries from microbial sources can still be achieved. Historically, screening efforts have been primarily focused on the optimization of fermentation media and culture conditions. Although fruitful, some cryptic or "silent" genes are unexpressed. Recent work using knock-out mutants, targeted over-expression of selective biosynthetic genes, heterologous expression in secondary hosts, epigenetic regulation, and ribosomal engineering have resulted in some new discoveries.[13] While these processes are non-trivial their success has been the result of advancements in analytical U/HPLC systems and Ultra-High Resolution mass spectrometers which can discriminate analytes at low detection levels with high sensitivity and capable of resolving metabolites from complex biological mixtures. In tandem, with new dereplication and discovery tools, comprehensive multivariate statistical analysis of under-explored producers, niche microbes, or endophytes of under-explored vascular plants, may also reveal new chemical entities with potential applications in medicine or agriculture.

Endophytes as a Source for Natural Products

It is estimated that of the >300,000 known vascular plant species on earth, from across varied ecosystems and habitats, contain over 1 million endophytic bacteria and fungi, with plants containing at least 1 or more endophytic species.[14] The earliest discovery of a plant endophyte

was by Freeman in 1904,[15] investigating the Darnel plant (*Lolium temulentum*), a ryegrass who's seeds or grains were toxic to rabbits, small carnivores animals and caused vomiting in humans. It was discovered that the Darnel plant was inhabited by an endophytic fungus, *Neotyphodium occultans*, which produced a series of loline alkaloids, that were responsible for the toxicity.[16] Historically, discoveries of endophytic microbes and their specialized metabolites from fescues plants was often a result of toxicity seen in livestock and foraging wild animals.[17]

If new metabolomic tools and strategies can be developed that can rapidly identify known or predicted chemical structures from natural sources then the classical dereplication bottle-neck can be drastically reduced. Furthermore, discovery strategies incorporating multivariate statistical analysis of under explored endophytes may aid dereplication efforts by focusing on unique and diverse endophyte species, which potentially may contain greater chemical diversity and novel specialized metabolites.

Secondary Metabolites: Polyketides and Nonribosomal Peptides

Genomic and biochemical analysis of microbial secondary metabolites has revealed that their construction or assembly can be readily annotated from their DNA gene sequence data, thus allowing for the prediction of potential physiochemical structures and even protein functions. This is possible as nonribosomal peptides and polyketides are constructed or assembled from a series of modular enzymes, which follow 'co-linearity rules' that build on acyl-CoA's in an iterative processes for polyketide synthesis (PKS) or activated amino acids and other carboxylic acids for nonribosomal peptide synthesis (NRPS). [18,19]

In the case of NRPS, amino acids are first activated by an adenylation domain (A), forming an aminoacyl adenylate, which is then transferred to the phospho-pantetheine arm of a thiolation (T) domain, forming a thioester linkage. Through a step-wise condensation reaction

mechanism, the condensation (C) domain catalyzes the nucleophilic attack of a neighboring amino group by forming an amide bond, resulting in chain elongation. The cleavage or cyclization step terminating the assembly-line is often controlled by thioesterase (TE) domain or a reductase (RE) domain. In a step-wise fashion, many thousands of monomer blocks can be independently selected for and assembled, creating numerous combinatorial sequence possibilities, thus leading to the varied structural diversity seen in Nature.[19, 20]

In the case of PKS, they work through a series of protein modular domains which govern catalytic activities; condensation reactions, keto-acyl synthase (KS), acyl transferase (AT) and acyl carrier protein in an iterative cyclic fashion, generating acetate building blocks.[21] These in turn can undergo ketone modifications such as keto reductase (KR), deyhratase (DH), and enoyl reductase (ER).[18]

Classical examples of NRPS and PKS type compounds that have been found through traditional bioactivity-guided screening assays are polymixin, bacitracin, cyclosporin, gramicidin, and amphotericin, epothilone, and rifamycin respectively. Nonribosomal peptides have clearly defined themselves as therapeutic agents, importantly, approximately five percent of the 1100 known nonribosomal peptides (as identified from Norine) are FDA approved drugs.[22] This is highly significant as the chance of finding a novel therapeutic agent far exceeds that of any other drug class. Significantly, FDA approved drugs from synthetic combinatorial approaches have success rate of only <0.001%, while out of the >7000 known polyketides, just over 20 are FDA approved, representing a success rate of 0.3%. [3, 23]

Genomic-guided discovery and dereplication strategies

Though polyketides and nonribosomal peptides are constructed in a predictable manner, the number of different monomers available for their construction provides a large amount of

chemical diversity.[18,24] Given their ability to produce bioactive molecules and the tendency to organize modular PKS and NRPS biosynthetic genes into predictable clusters, bacteria such as the *Streptomyces* sp. became the focus of new informatic-based discovery programs. By the late 90's, several biotechnology companies had begun developing platforms to harness genomic data to guide discovery efforts toward new chemical scaffolds. In 1998, Ecopia Biosciences developed a bioinformatics platform known as DECIPHER that was capable of sifting through partially sequenced genomes and applying biosynthetic logic to generate accurate predictions of genetically encoded natural products.[25–28] While this work had initially focused on enediynes, DECIPHER was later used to target novel PKS and NRPS products. Several of the subsequently discovered compounds included novel chemical scaffolds from known antibiotic producers, including ECO-0501, a novel MRSA (methicillin resistant *Staphylococcus aureus*) and VRE (vancomycin resistant *Enterococcus*) active antibiotic from the vancomycin producer *Amycolatopsis orientalis*, and ECO-4601, a novel benzodiazepenone nonribosomal peptide with potent broad spectrum anticancer activity.[29–31] While other biotechnology firms such as Biotica have since focused on genetic manipulation of PKS genes to produce variants of known compounds,[32] Ecopia served as one of the first examples of a 'front end' discovery program, using genomic information to detect novel chemical scaffolds and dereplicate known bioactive compounds using genomic information.

Historically, developments in structure prediction and dereplication have proceeded in parallel with mainstream technological advancements, including the progression of genome sequencing technologies and improvements in computational capabilities. As such, several academic research groups have followed in the footsteps of early biotechnology firms, and recent years have seen a rapid evolution of publicly available informatic tools for predicting secondary metabolites from microbial DNA sequences. Research into polyketide and nonribosomal peptide

biosynthesis had provided a framework for robust NRPS/ PKS domain specificity prediction programs like NRPS Predictor and NRPS-PKS that could be used for manual structure predictions of polyketides and nonribosomal peptides.[18,21,33,34] As whole genome sequencing has become common place, these powerful, early programs have become incorporated into global, genome-wide, automated NRPS and PKS structure prediction algorithms such as ClustScan, and NPSearcher, which also predicts potential exotic transformations, including oligomerization, heterocyclization, and modifications by tailoring enzymes.[35,36]

Recently, an increasingly holistic focus on the *in silico* detection and prediction of secondary metabolic machinery has lead to the development of antiSMASH – a program designed to identify secondary metabolic pathways and predicting structural archetypes, including aminoglycosides, polyketides, and nonribosomal peptides from bacterial and fungal genome sequences.[37,38] This comprehensive approach, along with an increased focus on prediction accuracy, will feature prominently in the development of improved predictive algorithms. For example, a second edition of the NRPS Predictor program can now apply different scoring parameters for adenylation domain specificity if the sequence originated from fungi or bacteria.[39] This program still lacks an automated predicted structure generation component, but it highlights the emergence of predictions tailored toward specific classes of productive microbes, crucial for producing accurate leads to drive 'front end' discovery programs.

Despite relatively advanced tools to predict modular polyketides and nonribosomal peptides from partial microbial genomes, downstream processing and interrogation of culture extracts was limited to traditional natural products methods like bioactivity guided fractionation.[29] Directed tools are now required for linking these new structural predictions with complex *in situ* analysis, including tandem MS data to identify structures with high confidence.

To complement these 'front end' discovery engines, recent advances have been made in 'back end' processing capabilities with a particular focus on nonribosomal peptides, building on previous advances in proteomic research.

Tandem MS analysis of ribosomally processed peptides and proteins is well understood, and can be carried out using top-down or shotgun methodologies.[40,41] Several proteomic software packages including PepNovo, MASCOT, PEAKS, and Sequest utilize real and hypothetical databases along with *de novo* sequencing to construct linear peptide sequences which can be annotated from tandem MS data.[42–45] Several issues have prevented the application of these approaches to nonribosomal peptides. A core problem is that nonribosomal peptides and NRP/PK hybrids often possess a complex structural architecture, including linear, cyclic, and mixed forms.[46] Backbone fragmentation of cyclic peptides is particularly challenging as random ring opening events can generate a series of ''non-direct'' sequences.[47–49] These linear peptide fragments overlay in the tandem MS spectra and prevent accurate proteomic analysis. In addition, there are over 500 known monomer blocks used in nonribosomal peptide assembly, compared to the ~20 proteinogenic amino acids used in the formation of ribosomal peptides.[18] Such structural and chemical diversity does not translate effectively to traditional linear *de novo* sequencing approaches. This challenge has led to the development of novel and modified approaches to understand the fragmentation, detection, and the dereplication of complex peptides.

Given the prominence of natural products as bioactive compounds with applications as human medicines, several natural product databases have been established that can facilitate a variety of dereplication efforts. Such resources include the nonribosomal peptide database NORINE,[22] the microbial natural compound collection Antibase, the marine natural product library MarineLit, and the Dictionary of Natural Products – a collection of compounds from a diverse spectrum of biological sources. Each of these databases includes specific molecules

grouped by a shared characteristic (e.g. nonribosomal peptides) that can be useful when considering database selection for various identification and dereplication efforts.

In a pioneering study in 2009, Ng *et al.* tackled the dereplication and *de novo* sequencing issues of cyclic NRPS peptides though the development of MS-Cyclic Peptide Annotation program (MS-CPA).[50] MS-CPA was an automated *de novo* sequencing algorithm designed for purely cyclic nonribosomal peptides and could enable their identification and comparative dereplication from a suitable database, such as NORINE via tandem MS data (Figure 4). While this work did not address the full diversity of nonribosomal peptide structures, it provided a directed means to identify purely cyclic compounds. Following this work, the development of Cycloquest was undertaken: an algorithm which demonstrated that the tandem MS spectra of cyclic ribosomal peptides could be used for homology searches from genetic and proteomic databases in a similar fashion to linear peptide database search programs like Sequest or MASCOT.[51] In 2011 a study by Kersten *et al.* had detailed a 'peptidogenomic' approach for matching bioinformatic ribosomal and nonribosomal peptide predictions to tandem MS spectra through iterative *de novo* sequencing.[52] Importantly, this tandem MS based identification approach integrates genetic predictions of amino acid identity, including the nonproteinogenic amino acids frequently found within nonribosomal peptides. However, manual *de novo* sequencing of tandem MS spectra can be time consuming and often requires sample enrichment to obtain ideal product ion spectra. Still, this work has shown again that genetic predictions are powerful tools for guiding the discovery of polyketides and nonribosomal peptides when paired with tandem MS data.

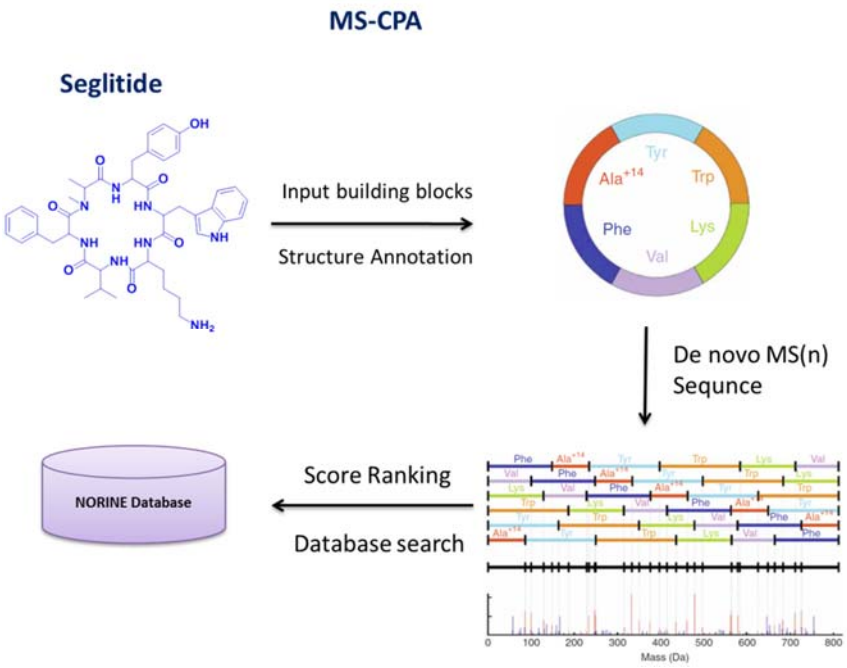


Fig.4. MS-CPA simplified overview. NRP structures are inputted with monomers assigned. Analyses of experimental and theoretical MS/MS spectra. Structure is scored against database for ranking. Adapted from Ng *et al.*2009. [50]

Expanding on genome mining of biosynthetic products, glycogenomics tools were also created for the rapid identification of glycosylated natural products (GNPs).[53] In 2012, Watrous *et al.* developed a "molecular networking" approach in conjunction with nano-spray desorption electrospray ionization (nanoDESI) as an effective strategy to visualize compound groupings (Figure 5).[54] They demonstrated that an MS based workflow can be used to visualize observed molecules as family groupings, based on commonalities within their mass spectral fragmentation patterns. These fragment commonalities were scored using cosine vectors and correlated to represent tandem MS networks of different, but related precursor ions. In a 2013 study by Nguyen *et al.*, an "MS/MS networking" strategy of connecting molecular family networking with gene cluster families was used successful in identifying nonribosomal peptides from over 60 organisms, the majority being un-sequenced.[55]

Molecular networking, although effective in illustrating groupings of like molecules, is not readily amenable to the identification of individual compounds with high specificity. Endogenous matrix effects from media components, artifact MS fragments or differing molecular scaffolds with conserved residues can be difficult to distinguish from non-identical compounds, as the cosine vector analysis is based primarily on partial *de novo* sequences tags, which are limited to peptidogenomics sequencing of purely proteinogenic amino acids. In this fashion, partial alignments within the tandem MS spectra are assessed for similarity. Manual investigation of the networked nodes and associations is then paramount and critical to determine the validity, with added MS network-analysis of reference standard matching, database searches, gene-knock-out examination or isolation, purification and compound characterization is needed to validate the correlations and confirm that structures are true.

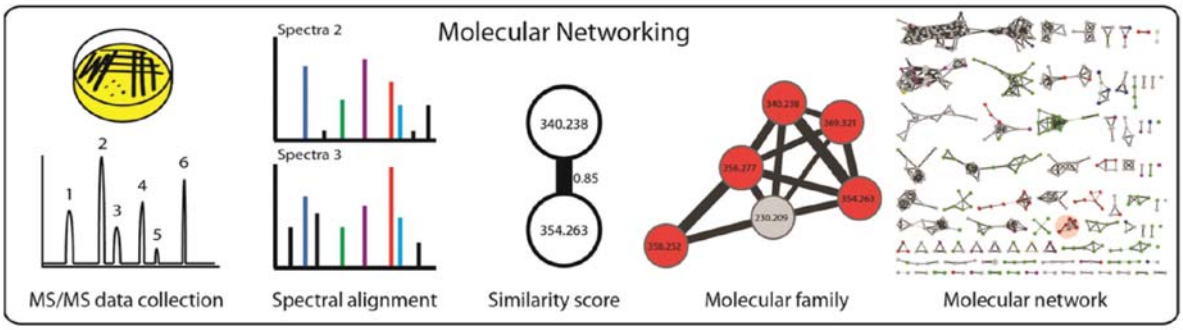


Fig. 5. Molecular networking of mass spectrometry data obtained from a microbial sample as an organizing and visualization approach. MS/MS data are collected from the microbial sample. Spectra are aligned and scored according to fragmentation similarity and neutral losses. Each node represents a consensus spectrum and is labeled by parent mass; the edge width indicates the cosine similarity score, creating a cluster of nodes that share structural similarities as molecular families within the entire network. Figure from Luzzatoo-Knaan and Dorrestein, 2015.[56]

In a 2012 study, we introduced iSNAP (Informatic Search Algorithm for Natural Products), a metabolomic dereplication and discovery platform.[57] iSNAP is an automated platform that uses a custom database search approach to dereplicate LC-MS/MS data through the use of a *in silico* spectral library (Figure 6). The spectral library is composed of hypothetical MS/MS spectra and their fragments, generated by applying mass spectrometry fragmentation

rules to linear SMILES codes,[58] requiring only a real or predicted chemical structure for accurate identification or dereplication of complex nonribosomal structures. The iSNAP informatic discovery platform is not limited by "direct-sequence" or "non-direct sequence" ions and is designed to handle the varied architectures found within nonribosomal peptides: linear, cyclic and branched-cyclic peptidic structures. The automated search algorithm requires only tandem MS data, is effective at identifying nonribosomal peptides at low nanogram levels from complex microbial extracts and is effective with low or high resolution mass data. The algorithms can elaborate the chemical structures of peptidic natural products, handling proteinogenic and non-proteinogenic amino acids; classically dereplicating knowns and identify potential analogous metabolites.

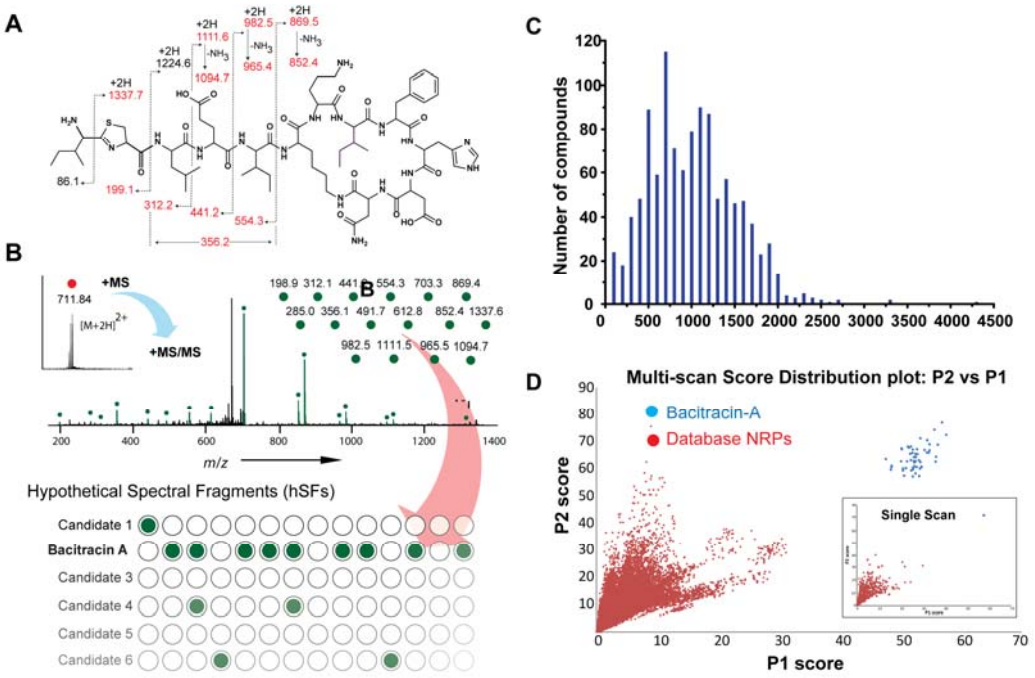


Fig. 6. Chemoinformatic analysis of bacitracin A. (A) Structure of bacitracin A. (B) Raw matching score overview. Hypothetical spectral library fragments composed of mass-to-charge ratios are compared with peaks from real MS/ MS spectra. Peak fragments in green represent matched mass-to-charge ratios within the tandem MS spectra. Matched peaks are then processed through the in-house nonribosomal peptide database and statistically scored to determine a candidate's match significance for dereplication. (C) Histogram representing the hypothetical spectral library of 1,107 compounds. (D) Dereplicating bacitracin A in (B) using doubly protonated (+711.82 m/z) MS/MS spectra. Multiple MS/MS scans are generated from an ~1-min direct infusion of bacitracin A; each blue point indicates a match between an MS/MS spectrum and bacitracin A. The red points show the score distribution of the other matches between the MS/MS spectra and the rest of the 1,106 database NRPs. The score distribution plots indicate the capability of the P1 andP2 scores indistinguishing true and false matches. Adapted from Ibrahim et al.[57]

Building on these efforts, in 2015 study, Lang *et al.* presented a new analog module for the iSNAP discovery platform, capable of elaborating families of peptide natural products and identify their structural analogues with statistical significance (Figure 7-8).[59] Importantly, the method was developed to reveal families of related peptidic natural products from crude extracts and is capable of mapping large numbers of new structural analogues with high accuracy; providing candidate structures, identifying site-specific residue or monomer modifications and providing retention times in an automated data-dependent fashion that is nontargeted. These features make the platform applicable to a wide user base and ideal for natural products discovery without the use of *de novo* sequencing.

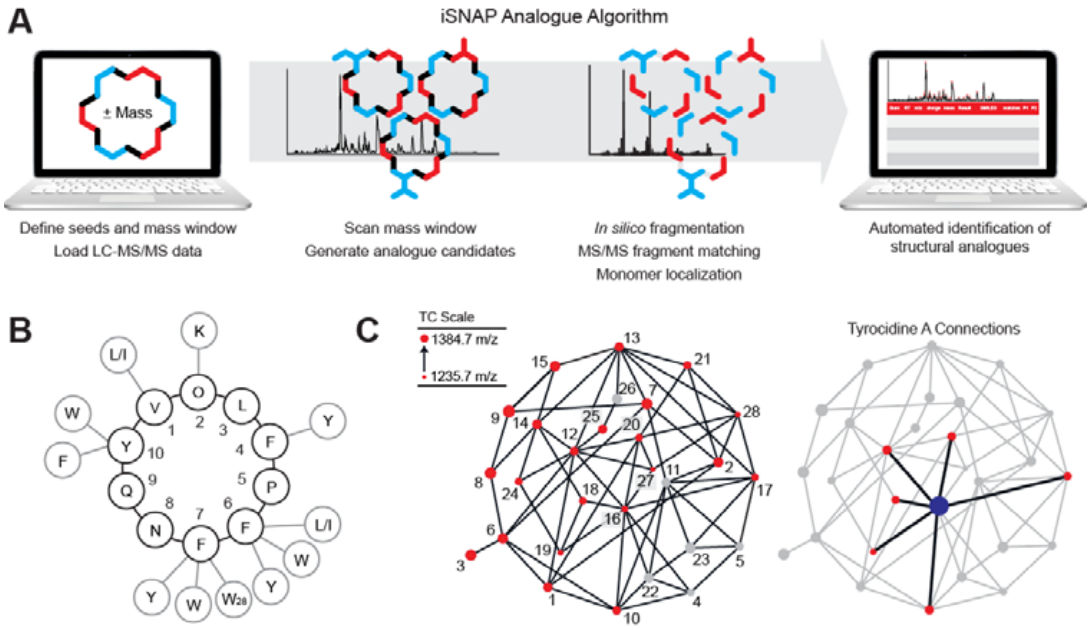


Fig. 7. iSNAP Analog Discovery Algorithm Can Map Families of Natural Peptides. (A) Schematic workflow of the automated iSNAP analog algorithm. After uploading an LC-MS/MS data file (.mzXML), users are required to define seed structures of interest—either predefined or following initial iSNAP dereplication analysis—in addition to a mass window around these seeds. Following dereplication, the algorithm searches the defined mass window and generates candidate analog structures. Hypothetical candidates are fragmented in silico to facilitate MS/MS fragment matching and monomer localization, driving the automated identification of structural analogs from the initial seed. (B) Amino acid composition of tyrocidine A (tyrocidine 16; black) and various substitutions found in the remaining 27 known members of the tyrocidine family of cyclic peptide natural products (gray). (C) Tyrocidine single-monomer substitution networks. Structures of tyrocidines 1–28 dereplicated by iSNAP (red) or known from previous literature (gray) are connected to one another through single amino acid substitutions (left). Through iSNAP analog analysis, single seeds can be used to access related structures, demonstrated using tyrocidine A (right). Figure from Liang et al. 2015.[59]

Similarly, in 2015 Johnston *et al.* incorporated iSNAP technologies with nonribosomal predicted biosynthetic data in a "Genomes-to-Natural Products" discovery approach.[60] Expanding on these technological developments, in 2015 Wang *et al.* presented "Global Natural Products Social Molecular" or GNPS for short, a pioneering tool for an open-access community-sharing based approach to deposit, process and share quality validated MS and tandem MS/MS data of natural product secondary metabolites.[61] Integrated within the GNPS platform are "molecular networking" algorithms, thus providing visual information for correlating related compounds based on MS/MS precursor and/or fragment ion commonalties. In 2016, Mohimani *et al.* described a new high-throughput dereplication algorithm for peptidic natural products called DEREPLICATOR.[62] This recent dereplication algorithm addressed some of the key challenges faced with other database search approaches (MS-CPA and iSNAP) such as variable dereplication and evaluating statistical significance (*P* values) in peptide-spectrum matches.

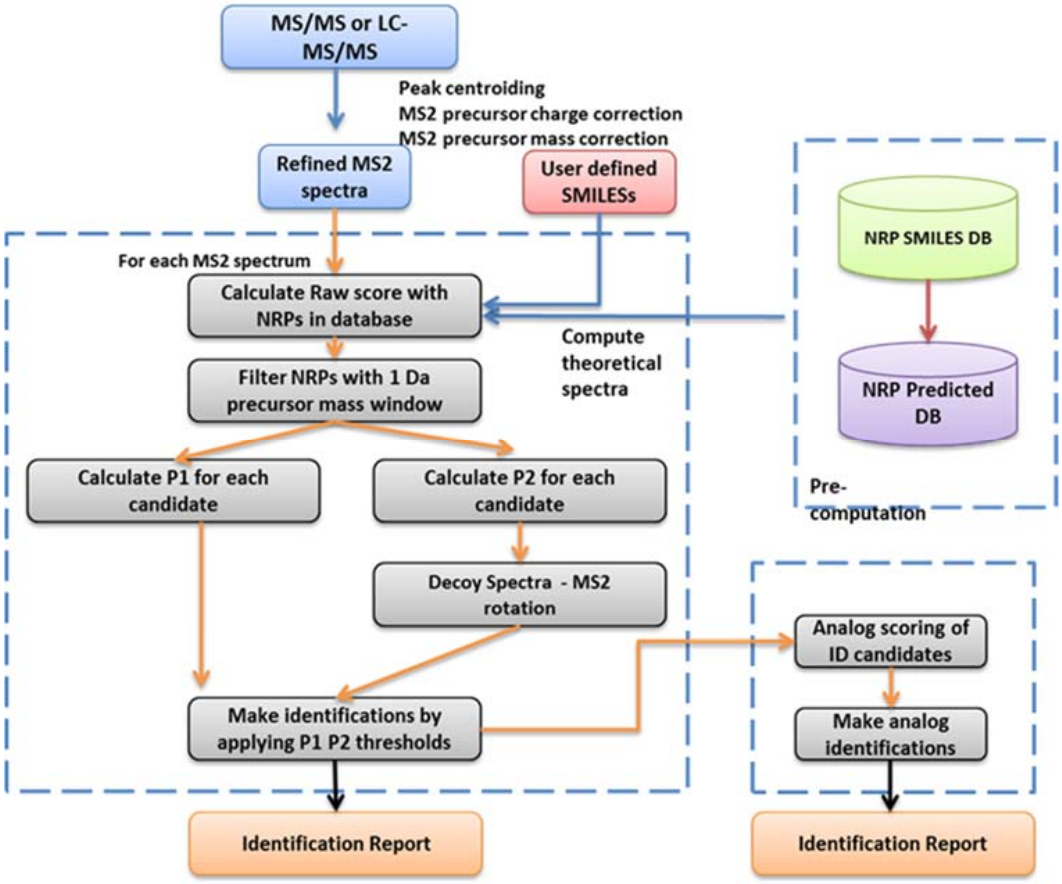


Fig.8. iSNAP metabolomic work-flow processes.[57,59]

LC-MS analysis is likely to remain the detection and isolation method of choice, owing to its high sensitivity, rapid processing speed, and ability to scale from analytical to preparative loads. The new frontier for discovery will involve the development of metabolomic and genomic prediction algorithms that can handle diverse and complex natural product architectures, facilitating the identification and acquisition of new chemical entities in an automated and high throughput fashion with high confidence and statistical significance. Combining tandem MS spectral data with bio- and chemo- informatic search engines capable of predicting and detecting genetically encoded small molecules from complex mixtures will allow a seamless merger of metabolomic capabilities with known genomic potential.

Specialized Metabolite Discovery via Multivariate Statistical Analysis.

While these MS-based work-flows have proven fruitful in addressing some of the challenges of dereplication and genomic-guided discovery, they are still evolving as discovery platforms with only a handful emerging over the last decade. Important challenges in natural product discovery still remain, including the ability to rapidly screen and process large data sets of LC-MS and LC-MS/MS data; the analysis of crude biological extracts from varied culture conditions; extracts of unknown origin or source, including undescribed or un-sequenced species; and a means to target and prioritize select microbial extracts for novel metabolite discovery. Specifically, comprehensive metabolomic strategies that combine multivariate statistical analysis processes like supervised (untargeted) principle component analysis (PCA) or supervised (targeted) orthogonal partial least-squares discriminative S-plot analysis (OPLS-DA),[63] may provide new discovery opportunities by exploring the global chemical space of microbial natural products.

In a useful study by Krug *et al.* (2008) they explored the interspecific species diversity of 98 *Myxococcus xanthus* strains, isolated from 78 global locations, by multivariate statistical analysis.[64] Microbial extracts, in replicate, were first screened by LC-MS, with their data processed using principle component analysis (PCA) to identify trends, unique groups or extract outliers for priority metabolite characterization. This approach afforded a means to investigate the secondary metabolite potential of *Myxococcus xanthus*, a well-studied strain known for its complex life cycles and fruiting bodies, independent of bioactivity or chemical-guided screening. They discovered that the diversity of secondary metabolites was significantly greater than previously thought and identified over 37 putative novel metabolites. Similarly, in 2012, Cortina *et al.* applied a "metabolome mining" approach to discover novel secondary metabolites from *Myxococcus xanthus* DK1662.[65] They first created a unique library of knock-out mutants for all

18 biosynthetic gene clusters found. Principle component analysis (PCA) was then applied to replicate LC-MS data sets of crude microbial extracts from wild-type and knock-out strains, to statistically identify metabolite difference. Using this approach, they discovered extracts containing three novel secondary metabolites (myxcoprincomide, c884 and c329) and five known compound families from targeting outlier extracts and metabolites. Importantly, myxcoprincomide, was isolated at ~ 1.2 μg/L, making traditional bioactivity-guided efforts unfeasible. This case has shown that comprehensive metabolomic profiling via LC-MS/MS can be a very powerful discovery tool in revealing metabolite differences between biological samples.

LC-MS metabolomic analysis does have its challenges: careful sample clean-up by de-salting/filtering is essential to limit adducts and ion suppression, efficient and reproducible chromatography through the use of core-shell columns, sensitive and stable high resolution mass accurate spectrometers; pre-processing LC-MS data with adduct and isotope corrections algorithms; proper bucketing of *m/z* values, compound retention time-windows and tolerances; and co-variance evaluation settings for ion intensities.

Endophytes as source for novel discovery

The term endophyte often refers to any organism (bacteria, fungi or algae) that lives within the vascular tissue of a host plant for all or most of its life-cycle. This relationship can be either a mutualistic symbiotic union or phytopathogenetic union established under stress. These ecological interactions provide a natural survival mechanism as endophytes gain vital nutrients, while providing enhanced protection through the production of secondary metabolites. Secondary metabolites of endophytic fungi are emerging as important chemical agents who have a role in the chemical ecology of host plants and their defense mechanisms. Specifically, in their role against predatory insects, plant fungal pathogens, and even providing increased drought

tolerance. Fungal endophytes are also known to produce a broad spectrum of biologically diverse specialized metabolites such anti-insectants (lolines), antimicrobials (beauvericin, cryptocandins, phomol), anti-cancer agents (cytochalasins, torreyanic acid) antivirals (cytonic acids), immunosuppressant agents (subglutinol A) and plant hormone regulating agents (abscisic acid).[66,67] Endophytes can also be found in almost all environmental eco-systems from tropical rain forest to Arctic tundra, covering a broad spectrum of microbial and chemical biodiversity.[68,69] A classic example of a endophyte sourced drug is Taxol, a potent anticancer diterpene, originally found in the bark of the Pacific yew tree and later isolated from a resident fungal endophyte , *Taxomyces andreanae*.[70]

While historically, microbial exploitation has relied heavily on terrestrial sources such as the Actinomycetes, and microbes of marine origins,[8] very few studies have been carried out on microbial endophytes from fruiting plants.[70–73] While endophytes from fescue grasses have been well documented, limited works have been reported on the endophytes from fruiting plant familes such as Rosaceae (rasberry, genus Rubus), Vitaceae (grapevine, genus Vitis) and Ericaceae (blueberry and cranberry, genus Vaccinium) and their specialized metabolites.[74] The present thesis aims exploit this new, untapped resource and structurally elucidate and characterize their specialized metabolites.

Opportunities for novel discoveries may be realized by combining MS-based work-flows, integrating novel dereplication and discovery algorithms with comprehensive metabolomic strategies focused on under explored niches, such as endophytes from fruiting plants. The targeted discovery of new agents and strategies for their discovery is of great interest, especially in the case of agriculturally important crops, and in our understanding of plant-fungi micro-biomes, and as potential biological agents for use as theraputics. This tandem approach provides a powerful means to realize and examine a global survey of the chemical space and biodiversity

of these endophytic fungi and their secondary metabolites. Furthermore, by rapidly dereplicating diverse extracts and applying a metabolomic-guided mining approach, unique collection isolates can be readily prioritized with a targeted means to investigated specific or key outlier metabolites for a true secondary metabolomic natural products discovery approach.

Thesis Objectives

The overall goal of this thesis is for the rapid and efficient discovery of novel secondary metabolites using an LC-MS based metabolomic guided workflow, incorporating an automated untargeted *in silico* dereplication algorithm. This approach affords high-throughput methods and protocols to addressing some of the key challenges facing microbial natural product discovery efforts. One of the greatest rate-limiting steps in traditional chemical or bioactivity-guided workflows has been the isolation and re-isolation of known bioactive agents, a process known as dereplication. As such, time, resources and energy are wasted screening, identifying and characterizing known compounds. While LC-MS screening methods have focused on MS, UV and retention time indexes; there are no commercial sources of comprehensive MS and UV spectral libraries available for a comparative analysis. Similarly, while HRMS molecular formula measurements and tandem MS data provides additional clarity on structural identification, isobaric species and limited fragmentation pathways are still challenging, with manual interpretation often required. LC-MS-based metabolomic tools can provide a global survey of the secondary metabolome and identify unique variances between large microbial datasets. The secondary goal is the structural elucidation and characterization of secondary metabolites from poorly understood fungal endophytes from fruiting plants by LC-MS and 1D and 2D NMR and single crystal X-ray analysis. These findings may lead to a better understanding of the chemical diversity of Canadian *spermatopsida* (seed-bearing plants) and their ecological significance.

In Chapter 2 of this thesis, a novel automated dereplication algorithm, called iSNAP,[57] has been developed and applied to a diverse set of nonribosomal peptides for an untargeted LC-MS/MS dereplication approach for natural products discovery. The iSNAP algorithm has been designed around an *in silico* spectral library database of hypothetical spectral fragments (hSFs), based on amide cleavage, which are generated from a compounds SMILES codes. The hSFs are calculated estimations of how a protonated peptide may fragment or be generated by collision-induced dissociation (CID) within a mass spectrometer and comprises all of the mass-to-charge values which may be observed in real MS/MS spectra. Unlike modern proteomic efforts which use de novo sequencing methods, this approach does not need to discriminate between direct and non-direct sequence ions and can handle the diverse architects of nonribosomal peptides (liner, cyclic and cyclic-branching). iSNAP's algorithms are designed to process LC-MS/MS and MS/MS data and dereplicating nonribosomal peptides in an automated, non-targeted approach with nano-gram sensitivity and statistically scores MS/MS fragment matches with high confidence. This approach provides a new metabolomic discovery strategy compared to traditional de novo sequencing methods for a high-throughput means of classical dereplication.[57]

In Chapter 3, a novel analog module has been developed for the iSNAP metabolomic platform that combines the robust dereplication processes with a highly accurate analog identification algorithm; capable of discriminating between isobaric species and identifying site specific (monomer) modifications of analogous compounds.[59] The iSNAP analog program is transformative, as under-explored producers can now be re-examined to identify analogous, low abundant metabolites, which may have been obscured through traditional bioactivity-screening methods. This nano-gram sensitive discovery approach does not require genomic sequence data, HRMS data or prior knowledge of the screening samples. iSNAP's automated software can either dereplicate LC-MS/MS data sets and identify their analogs or through up-loadable seed

structures, representing dereplicated knowns. This novel metabolomic discovery tool was tested and validated against a series of nonribosomal peptides and all without the use of traditional bioassay-guided isolation methods.

In Chapter 4, an MS-based workflow combining the iSNAP dereplication and discovery platform with a comprehensive LC-MS/MS metabolomic-guided discovery approach is used to investigate a diverse collection of 184 under-explored endophytic fungi, for the guided mining and isolation, characterization and structure elucidation of novel specialized metabolites by 1D and 2D NMR and HRMS analysis (Figure 10). This strategy allows for the rapid dereplication of knowns using iSNAP informatic search algorithms, prioritization of unique collection outlier extracts and their metabolites via unsupervised PCA multivariate analysis. This approach provides a unique handle in dealing with large data sets of unknown or poorly described microbial extracts. Using this approach, a total of 12 previously undescribed specialized metabolites were discovered, including >30 known compounds from several outlier extracts. The majority of the isolated compounds show bioactivity when tested against panel of Gram-negative and Gram-positive bacteria, including several yeasts, thus highlighting the significance of this approach, independent of traditional bioactivity screening.

In Chapter 5, an untargeted LC-MS based metabolomic-guided discovery approach was applied to screening 15 isolates from a novel endophytic xylaria species, *Xylaria ellisi*, common to low and highbush blueberries. Isolates from different locations and varieties, were grown in different culture media, with extracellular and intracellular extracts screened using supervised OPLS-DA and S-plot metabolomic analysis to target, isolate and characterize new bioactive agents by 1D and 2D NMR and HRMS analysis. Using this supervised metabolomic approach, a total of 8 previously undescribed cyclic nonribosomal peptides were discovered, including 11

known compounds. Importantly, one new cyclic peptide demonstrated Gram-negative activity against *E. coli* bacteria, a first report for this scaffold.

Chapter 6 focuses on a unique fungal endophyte isolate whose extracts showed biological activity in a preliminary bioactivity screen and was also a modest extract outlier in Chapter 3. In total, 10 previously undescribed polyketide specialized metabolites and 4 known compounds were isolated from the culture filtrates of *Nemania serpens* (Pers.) Gray (1821), an endophytic fungus isolated from the grapevines and leaves of a Riesling (Vitis vinifera) grape plant.[74] The previously undescribed compounds have been named epoxynemanione A, nemanifuranones A-C, and nemanilactones A-C. The structures were elucidated based on 1D and 2D NMR, HRMS measurements and single crystal X-ray analysis of nemanifuranone A, a known nordammarane triterpenoid and a related compound 2,3-dihydro-2-hydroxy-2,4-dimethyl-5-trans-propenylfuran-3-one, isolated from the culture filtrates of *Mollisia nigrescens*, and endophytic fungus from lowbush blueberries. The nemanifuranones contain a rare C2 hemiacetal and nemanifuranone A was active against both Gram-negative and Gram-positive bacteria.

The works herein, describes the development of a novel LC-MS based technology platform named iSNAP, which allows for the rapid and statistically significant dereplication of knowns and accurate analog identification of nonribosomal peptidic compounds from complex biological mixtures. A comprehensive metabolomic-guided discovery strategy, in combination with iSNAP technologies, was applied to screen and target unique secondary metabolite outliers from a diverse collection of 188 fungal endophytes isolated from fruiting plants (Figure 10). The prioritization of these specific strains and metabolites using multivariate statistical analysis allowed for the rapid high-throughput screening of an entire microbial collection, comprised of 13649 and 5590 metabolite features by positive and negative LC-MS analysis. Importantly, these combined efforts have demonstrated the effectiveness of this approach, which led to the discovery

of over 30 previously undescribed specialized metabolites and over 30 known compounds. These findings further demonstrate that endophytes from fruiting plants are an excellent source of novel bioactive metabolites. Importantly, new and innovative approaches are needed to realize the potential of new natural products from understudied sources.

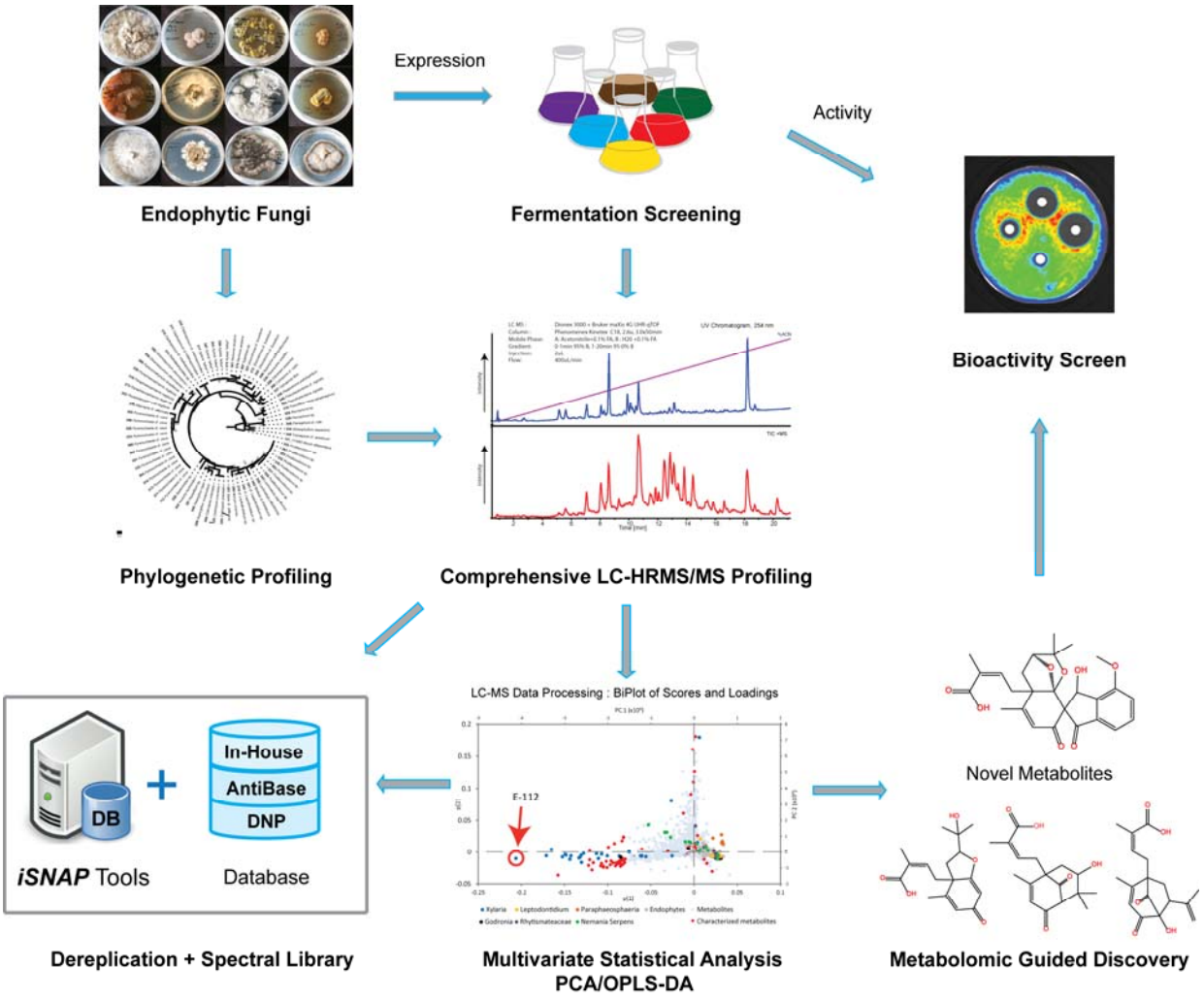


Fig. 10. Schematic of a Metabolomic Guided workflow, incorporating the iSNAP dereplication platform.

References

1. Butler, M. S. The role of natural product chemistry in drug discovery. *J. Nat. Prod.* **67,** 2141–53 (2004).
2. Fleming, A. On the antibacterial action of cultures of a penecillium, with special reference to their use in the isolation of B. influenae. *Br. J. Exp. Pathol.* 226–236 (1929).
3. Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier?*Science* **325,** 161–5 (2009).
4. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661 (2016).
5. Larsen, T. O., Smedsgaard, J., Nielsen, K. F., Hansen, M. E. & Frisvad, J. C. Phenotypic taxonomy and metabolite profiling in microbial drug discovery. *Nat. Prod. Rep.* **22,** 672–695 (2005).
6. Arias, C. A. & Murray, B. E. Antibiotic-Resistant Bugs in the 21st Century — A Clinical Super-Challenge. *N. Engl. J. Med.* **360,** 439–443 (2009).
7. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529,** 336–343 (2016).
8. Gaudêncio, S. P. & Pereira, F. Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **32,** 779–810 (2015).
9. Nielsen, K. F. *et al.* Dereplication of Microbial Natural Products by LC-DAD-TOFMS. *J. Nat. Prod.* **74,** 2338–2348 (2011).
10. Seger, C., Sturm, S. & Stuppner, H. Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques--state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. *Nat. Prod. Rep.* **30,** 970–987 (2013).
11. McAlpine, J. B. Advances in the Understanding and Use of the Genomic Base of Microbial Secondary Metabolite Biosynthesis for the Discovery of New Natural Products Advances in the Understanding and Use of the Genomic Base of Microbial Secondary Metabolite Biosynthesis for. *J. Nat. Prod.* **72,** 566–572 (2009).
12. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26,** 1362 (2009).
13. Winter, J. M., Behnken, S. & Hertweck, C. Genomics-inspired discovery of natural products. *Curr. Opin.*

Chem. Biol. **15,** 22–31 (2011).

14. Strobel, G. & Daisy, B. Bioprospecting for Microbial Endophytes and Their Natural Products Bioprospecting for Microbial Endophytes and Their Natural Products. *Microbiol. Mol. Biol. Rev.* **67,** 491–502 (2003).
15. Freeman, E. M. The Seed-Fungus of Lolium temulentum, L., the Darnel. *Philos. Trans. R. Soc. Lond. [Biol]* **196,** 1–27 (1904).
16. Schardl, C. L., Grossman, R. B., Nagabhyru, P., Faulkner, J. R. & Mallik, U. P. Loline alkaloids: Currencies of mutualism. *Phytochemistry* **68,** 980–996 (2007).
17. Strobel, G. & Daisy, B. Bioprospecting for Microbial Endophytes and Their Natural Products. *Microbiol. Mol. Biol. Rev.* **67,** 491–502 (2003).
18. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem. Rev.* **106,** 3468–3496 (2006).
19. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem*. *Rev*. **97,** 2651-2674**,** (1997).
20. Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides : from genes to products. *Nat. Prod. Rep.* **20,** 275–287 (2003).
21. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18,** 380–416 (2001).
22. Pupin, M., Fontaine, A., Jacques, P. & Kucherov, G. NORINE : a database of nonribosomal peptides. *Nucleic Acids Res.* **36,** 326–331 (2008).
23. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **70,** (2007).
24. Schulze, C. J. *et al.* Genome-Directed Lead Discovery: Biosynthesis, Structure Elucidation, and Biological Evaluation of Two Families of Polyene Macrolactams against Trypanosoma brucei. *ACS Chem. Biol.* **10,** 2373–2381 (2015).
25. Zazopoulos, E. *et al.* A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **21,** 187–90 (2003).
26. McAlpine, J. B. *et al.* Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation:

ECO-02301 , a Novel Antifungal Agent , as an Example. *J. Nat. Prod.* **68,** 493–496 (2005).

27. Banskota, A. H. *et al.* TLN-05220, TLN-05223, new Echinosporamicin-type antibiotics, and proposed revision of the structure of bravomicins. *J. Antibiot.* **62,** 565–570 (2009).

28. Banskota, A. H. *et al.* Isolation and Identification of Three New 5-Alkenyl-3 , 3 (2H) -furanones from Two Streptomyces species using a Genomic Screening Approach. *J. Antibiot.* **59,** 168–176 (2006).

29. Banskota, A. H. *et al.* Genomic analyses lead to novel secondary metabolites. Part 3. ECO-0501, a novel antibacterial of a new class. *J. Antibiot.* **59,** 533–542 (2006).

30. McAlpine, J. B. *et al.* Biosynthesis of Diazepinomicin / ECO-4601 , a Micromonospora Secondary Metabolite with a Novel Ring System. *J. Nat. Prod.* **71,** 1585–1590 (2008).

31. Gourdeau, H. *et al.* Targeted Delivery of ECO-4601 to Brain Tumors in an Orthotopic Model. *3rd Mod. Drug Discov. Dev. Summit* 11195 (2007).

32. Gregory, M. A. *et al.* Mutasynthesis of rapamycin analogues through the manipulation of a gene governing starter unit biosynthesis. *Angew. Chemie - Int. Ed.* **117,** 4757–4760 (2005).

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

34. Ansari, M. Z., Yadav, G., Gokhale, R. S. & Mohanty, D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32,** W405-13 (2004).

35. Starcevic, A. *et al.* ClustScan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* **36,** 6882–6892 (2008).

36. Li, M. H. T., Ung, P. M. U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D. H. Automated genome mining for natural products. *BMC Bioinformatics* **10,** 185 (2009).

37. Medema, M. H. *et al.* AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** 339–346 (2011).

38. Weber, T. *et al.* antiSMASH 3.0--a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** 237-243 (2015).

39. Röttig, M. *et al.* NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39,** 362–367 (2011).

40. Kelleher, N. L. Top-Down Proteomics. *Anal. Chem.* **76,** 196–203 (2004).

41. Liu, H., Sadygov, R. G. & Yates III, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76,** 4193–4201 (2004).

42. Frank, A. & Pevzner, P. PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **77,** 964–973 (2005).

43. Hirosawa, M., Hoshida, M., Ishikawa, M. & Toya, T. Mascot - Multiple Alignment System For Protein Sequences Based On 3- Way Dynamic-Programming. *Comput. Appl. Biosci.* **9,** 161–167 (1993).

44. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17,** 2337–42 (2003).

45. Razumovskaya, J. *et al.* A computational method for assessing peptide- identification reliability in tandem mass spectrometry analysis with SEQUEST. *Proteomics* **4,** 961–969 (2004).

46. Fischbach, M. A., Walsh, C. T., Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics : Logic , Machinery , and Mechanisms Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics : Logic , Machinery , and Mechanisms. *Chem. Rev.* **106,** 3468–3496 (2006).

47. Harrison, A. G., Young, A. B., Bleiholder, C., Suhai, S. & Paizs, B. Scrambling of sequence information in collision-induced dissociation of peptides. *J. Am. Chem. Soc.* **128,** 10364–10365 (2006).

48. Liu, W.-T. *et al.* Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Anal. Chem.* **81,** 4200–9 (2009).

49. Eckart, K. Mass spectrometry of cyclic peptides. *Mass Spectrom. Rev.* **13,** 23–55 (1994).

50. Ng, J. *et al.* Dereplication and de novo sequencing of nonribosomal peptides. *Nat. Methods.* **6,** 596–599 (2009).

51. Mohimani, H. *et al.* Cycloquest: Identification of cyclopeptides via database search of their mass spectra against genome databases. *J. Proteome Res.* **10,** 4505–4512 (2011).

52. Kersten, R. D. *et al.* A mass spectrometry-guided approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7,** 794–802 (2011).

53. Kersten, R. D. *et al.* Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4407-16 (2013).

54. Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci.* **109,** E1743–E1752 (2012).

55. Nguyen, D. D. *et al.* MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E2611-20 (2013).

56. Luzzatto-Knaan, T., Melnik, A. V & Dorrestein, P. C. Mass spectrometry tools and workflows for revealing microbial chemistry. *Analyst* **140,** 4949–66 (2015).

57. Ibrahim, A. *et al.* Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19196–201 (2012).

58. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28,** 31–36 (1988).

59. Yang, L. *et al.* Exploration of Nonribosomal Peptide Families with an Automated Informatic Search Algorithm. *Chem. Biol.* **22,** 1259–1269 (2015).

60. Johnston, C. W. *et al.* An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun.* **6,** 8421 (2015).

61. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34,** 828–837 (2016).

62. Mohimani, H. *et al.* Dereplication of peptidic natural products through database search of mass spectra. *Nat. Chem. Biol.* **13,** 30–37 (2016).

63. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **1,** 92–107 (2013).

64. Krug, D. *et al.* Discovering the hidden secondary metabolome of Myxococcus xanthus: A study of intraspecific diversity. *Appl. Environ. Microbiol.* **74,** 3058–3068 (2008).

65. Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R. Myxoprincomide: A natural product from myxococcus xanthus discovered by comprehensive analysis of the secondary metabolome. *Angew. Chemie - Int. Ed.* **51,** 811–816 (2012).

66. Deshmukh, S. K., Verekar, S. A. & Bhave, S. V. Endophytic fungi: A reservoir of antibacterials. *Front. Microbiol.* **5,** 1–43 (2015).

67. Guo, B., Wang, Y., Sun, X. & Tang, K. Bioactive natural products from endophytes: A review. *Appl. Biochem. Microbiol.* **44,** 136–142 (2008).

68. Strobel, G., Daisy, B., Castillo, U. & Harper, J. Natural Products from Endophytic Microorganisms. *J. Nat. Prod.* **67,** 257–268 (2004).

69. Zhang, H. W., Song, Y. C. & Tan, R. X. Biology and chemistry of endophytes. *Nat. Prod. Rep.* **23,** 753–771 (2006).

70. Strobel, G. A. Endophytes as sources of bioactive products. *Microbes Infect.* **5,** 535–544 (2003).

71. Burgess, K. M. N., Ibrahim, A., Sørensen, D. & Sumarah, M. W. Trienylfuranol A and trienylfuranone A–B: metabolites isolated from an endophytic fungus, Hypoxylon submoniticulosum, in the raspberry Rubus idaeus. *J. Antibiot.* **70**, 721-725 (2017).

72. Kuldau, G. & Bacon, C. Clavicipitaceous endophytes: Their ability to enhance resistance of grasses to multiple stresses. *Biol. Control* **46,** 57–71 (2008).

73. Ibrahim, A. *et al.* New diplosporin and agistatine derivatives produced by the fungal endophyte Xylaria sp. isolated from Vitis labrusca. *Phytochem. Lett.* **9,** 179–183 (2014).

74. Ibrahim, A. *et al.* Epoxynemanione A, nemanifuranones A–F, and nemanilactones A–C, from Nemania serpens , an endophytic fungus isolated from Riesling grapevines. *Phytochemistry* **140,** 16–26 (2017).

CHAPTER TWO

DEREPLICATING NONRIBOSOMAL PEPTIDES USING AN INFORMATIC SEARCH ALGORITHM FOR NATURAL PRODUCTS (iSNAP) DISCOVERY

CHAPTER TWO PREFACE

The following works was previously published in:

Ibrahim, A.*, Yang, L.*, Johnston, C., Liu, X., Ma, B., Magarvey, N.A. Dereplicating Nonribosomal Peptides Using An Informatic Search Algorithm For Natural Products (iSNAP) Discovery. Proceedings of the National Academy of Sciences (109) 47, 19196-19201 (2012).

* The authors contributed equally to this work.

http://www.pnas.org/content/109/47/19196

http://www.pnas.org/content/suppl/2012/11/02/1206376109.DCSupplemental

Author Contributions:

A.I., L.Y., X.L., B. M., and N.A.M. designed research and developed iSNAP – A metabolomic dereplication and discovery tool; A.I., L.Y., and X.L. performed research; A.I., L.Y., C.J., X.L., B.M., N.A.M. analyzed data; and A.I., L.Y., B.M., and N.A.M. wrote the manuscript.

Acknowledgments:

The authors would like to thank Dr. Suzanne Osborne and Dr. Brian Coombes for IVIS imaging as well the Centre for Microbial Chemical Biology at McMaster University for access to state-of-the-art instrumentation.

ABSTRACT

Nonribosomal peptides are highly sought after for their therapeutic applications. Like other natural products, dereplication of known compounds and focused discovery of new agents within this class is a central concern of modern natural product-based drug discovery. Development of a chemoinformatic library-based and informatic search strategy for natural products (iSNAP) has been constructed and applied to nonribosomal peptides and proved useful for true non-targeted dereplication across a spectrum of nonribosomal peptides and within natural product extracts.

INTRODUCTION

Nonribosomal peptides (NRPs) are a group of natural products with diverse biological activities and pharmacophores (1-2). The evolutionarily-selected status of these peptides translates to intrinsic bioactivity, and ~5% of the 205 NRP structural families are used clinically as inhibitors of enzymes, agonists and antagonists of receptors, modulators of eukaryotic signaling cascades, potentiators of epigenetic modification, and perturbants of protein-protein interactions (3). Efforts to discover new NRPs have increasingly resulted in the rediscovery of known compounds, stifling new therapeutic advances and highlighting the need for rapid and efficient methods of dereplication (4). Further, rapid advances in microbial genome sequencing have exposed a wealth of novel gene clusters that encode for NRPs (5-6). New analytical tools are needed to dereplicate NRPs and reveal novel potential therapeutics.

Modern proteomic research has utilized mass spectrometry to achieve efficient and automated peptide dereplication from complex mixtures through *de novo* sequencing and database-derived methods (7). As all peptides share a common amide monomer linkage they should follow similar MS fragmentation patterns. However, two important variations

necessitate the development of divergent informatics tools for NRP dereplication. First, NRPs can be assembled from a much larger range of monomers (>500) and often incorporate polyketide building blocks. Second, nonribosomal peptide architecture is varied between linear, cyclic, and mixed or 'branched' combinations thereof (8).

In linear peptides, the fragmentation pattern proceeds from the termini, providing a series of diagnostic ladder ions or 'direct sequence ions' (DS) with relatively few internal cleavages or rearrangements that generate 'non-direct sequences" (NDS). In contrast, cyclic NRPs are prone to multiple ring opening events, with each linear form producing unique ladder ions and enrichments of other NDS (9, 10). The resulting output is a mix of DS and NDS, and has proven to be a considerable challenge for *de novo* sequencing methods (11). Recently, Dorrestein and Pevzner presented a *de novo* sequencing approach for purely cyclic nonribosomal peptides and demonstrated the utility in 'comparative dereplication'. In their approach, a comparative dereplication (similarity ranking) was illustrated using eighteen pure known cyclic nonribosomal peptides, with four of these being correctly classically dereplicated in a manual fashion (12-14). Similarly, Dorrestein connected chemotypes with microbial genomic data by an iterative *de novo* sequencing approach in peptidogenomics (15).

Use of nonribosomal peptide databases and scoring of fragment matches may provide an alternative strategy to *de novo* approaches and result in classical dereplication of nonribosomal peptides. A structural matching design would not require differentiation of NDS from DS, and may work for the varied architectural forms, backbone modifications, altered connectivities, and non-peptidic building blocks found in NRPs and hybrids thereof (NRP-polyketide, NRP-terpene). Unfortunately, no spectral library of mass-to-charge ratios of known NRPs exists and no scoring matrices are established.

In this work we present a platform for informatic searching of natural products (iSNAP) to detect NRPs using a database-searching algorithm in an automated data-dependent mode that is non-targeted and affords a nanogram sensitive, efficient, and high-throughput means of classical dereplication of NRPs in natural product extracts.

RESULTS

Development of an informatic platform and chemoinformatic database for natural product discovery

Numerous challenges are confronted in constructing NRP natural product databases for automated dereplication. First, there is no compiled spectral database with information on all the known NRPs or a ready supply of compounds to create one. Further there are no mathematical tools available to computationally compare unknown analytes to known nonribosomal peptides and no infrastructure existing to create hypothetical MS/MS spectra of known compounds in a rapid fashion.

Nonribosomal peptides (represented in SMILES format) were taken from the NORINE database (3), PubChem, the Journal of Antibiotics, and other resources (see *SI Appendix*, Section I.*F*). Simplified Molecular Input Line Specification (SMILES) is a linear string code that contains all the structural information of a given small molecule (16). The assembled in-house NRP database contains 1107 NRP structures, and for the initial part of our *i*nformatic *s*earch *a*pproach for *n*atural *p*roducts (iSNAP) we created a script that would identify all amide bonds and generate hypothetical spectral fragments (hSFs) based on amide cleavage. These hSFs are calculated estimations as to how a protonated peptide may fragment or be generated from collision-induced dissociation (CID) within the gas phase of an MS/MS experiment (17). The iSNAP algorithm labels all amide cleavage sites within a compounds SMILES code. The

hSF's are generated by enumerating the cleavage at two amide sites at a time. These fragments arise from the cleavage of N-terminal (b-and a-ions) and C terminal (y-ions) cleavage and the iSNAP program takes these and adds mass offsets of +H and +H+1 to account for protonation and the first isotope ion, respectively. In this way, the initial 1107 NRP structures, resulted in a hypothetical spectral library (HSL) of 100,747 hSFs. Of these 27,036 fragments resulted from amide cleavage, with each having a corresponding fragment bearing values indicative of the sequestration ionization charges (hydrogen and hydrogen plus one species) (81,108 mass-to-charge values) and neutral losses species (water, ammonia, and carbon monoxide) generating 19,639 off-set mass-to-charge values.

The collective of these hSFs comprise all of the mass-to-charge ratio ions that may be observed in real MS/MS spectra of the known NRPs. As such, a direct comparison of the hypothetical versus the experimental spectra for a given NRP should yield a significant number of shared high-intensity peaks.

Comparative analysis of hypothetical mass-to-charge ratios and tandem mass spectra for the detection of nonribosomal peptides

We sought to determine how computational fragmentation of NRPs (described above) would compare with actual NRP fragmentation (Fig. 1*A*-*B*). For this we compared the spectral fragments derived from bacitracin-A, an antimicrobial NRP comprised of both linear and cyclic portions with the hSFs generated by iSNAP. An authentic standard of bacitracin-A was subjected to ESI-MS/MS analysis by direct infusion with the double charged ion (+711.4 m/z) selected and subjected to CID. iSNAP analysis of bacitracin-A generated 102 hSFs and a total of 301 mass-to-charge values from these by +H and +H+1 mass offsets (See Data Set 1), in addition to neutral loss species (H_2O, NH_3, and CO). Of these, 89 mass-to-charge values could

be detected by iSNAP matching algorithm from the doubly charged MS/MS spectrum (Fig. 1*B*).

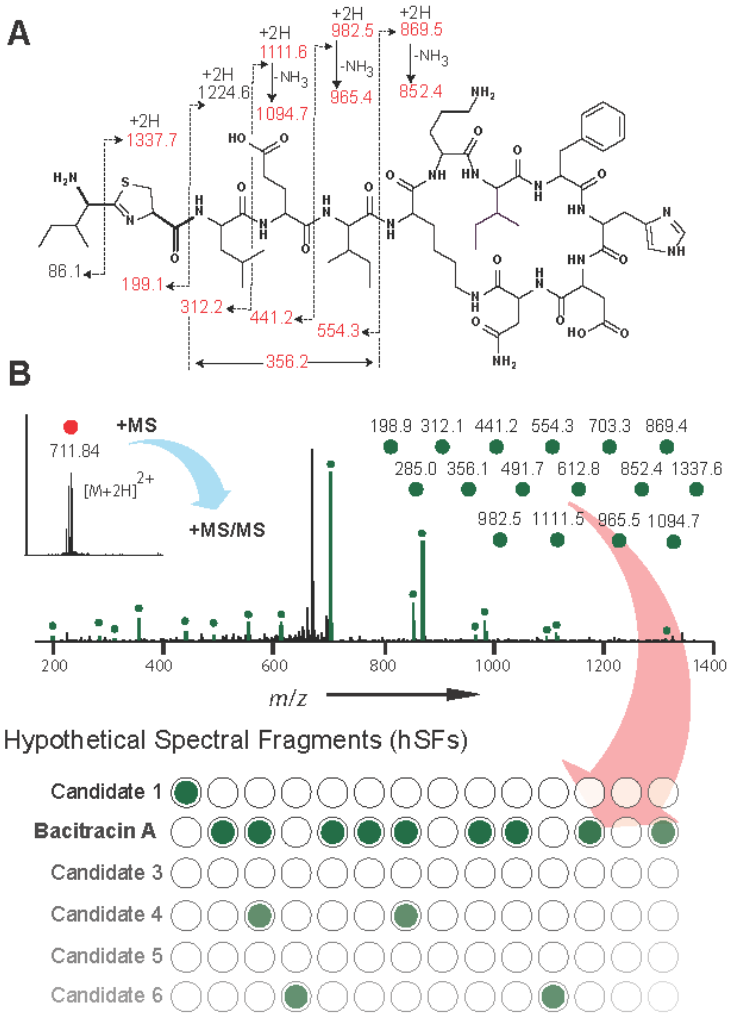


Figure 1. Chemoinformatic analysis of Bacitracin-A. (A) Structure of bacitracin A. (**B**) Raw matching score overview. Hypothetical spectral library fragments comprised of mass-to-charge ratios are compared with peaks from real MS/MS spectra. Peak fragments in green represent matched mass-to-charge ratios within the tandem MS spectra. Matched peaks are then processed through the in-house nonribosomal peptide database and statistically scored to determine a candidate's match significance for dereplication

Creation of a Scoring Scheme (Raw score, P_1 score, and P_2 score)

Having generated an NRP hypothetical spectral library (Fig. 2*A*), we next focused on deriving a scoring mechanism to compare experimentally-generated spectra to the hypothetical spectral library. Three scores are computed for these two purposes: Raw score, P_1 score, and P_2 score. Raw score is an overall spectral match between the MS/MS spectrum of an analyte

and hypothetical spectrum of a known NRP. Raw score alone however, does not remove bias toward larger sized NRPs and spectra with large numbers of fragment peaks. In this way, Raw score is not a comparable measure across different spectra and therefore we derived probability scores denoted P_1 and P_2 that use Raw scoring but derive match significance differently. In general, as NRPs increase in mass, the number of hSFs also increases due to the presence of potentially more amide bonds and cyclic/cyclic-branching connectivity's. With added offsets and neutral losses, the total number of hSFs can rapidly accumulate, thus the chances of falsely matching fragment ions rises, creating and artificial bias.

Raw score calculation

In calculating the Raw score, or spectral-matching score, iSNAP algorithm first conducts noise filtering process to remove low intensity peaks from the input MS/MS spectra. In this process, iSNAP calculates the relative peak intensity for all the ion peaks by comparing them to the highest peak within the spectrum, and filters out peaks of less than 0.5%. This pre-filtering is applied to reduce the likelihood of randomly matched peaks, and such pre-processing in embedded within most proteomic ribosomal peptide algorithms (18-20). iSNAP program collects the remaining peaks and matches only those with the hypothetical spectral library. In the event that an input MS/MS spectrum is from a multiply charged ion, the algorithm correlates and adjusts the protonated hypothetical spectrum to account for difference in charge states. When the parent ion of the MS/MS spectrum bears a charge *k*, the m/z values of hypothetical fragments with charges up to *k* are combined to form the charge-*k* hypothetical spectrum. By using a mass error tolerance of 0.1 *Da*, the algorithm finds all spectrum peaks that have matches and computes the Raw score as

$$\text{Raw score} = \sum_{\text{each matched peak } m_i} \log_{10}(200 \times \text{relative intensity of } m_i)$$

The fraction 1/0.5% (factor 200) in the formula is used to ensure a match to a peak of significant intensity (>=0.5% relatively intensity) will not contribute negatively to the overall score. Within the iSNAP algorithm a mass error tolerance of 0.1 *Da* is set to accommodate errors arising from use of low-resolution mass spectral files. Values set too low will limit matched fragments, and higher ones increase matches, possibly increasing random assignments.

For each MS/MS spectrum, the Raw score is calculated against the database compounds within a mass range of 0 *Da* to [M] +100 *Da*, where [M] represents parent mass. Having a relaxed mass range ensures sufficient Raw scores are calculated for statistical distribution and the upper limit of [M]+100 *Da* avoids a potential bias for large molecules that may score higher due to more fragment matching possibilities. The +100 *Da* value is chosen empirically by experimenting with + 0, 50, 100, 200 and 500 *Da* (see *SI Appendix*, Section I.*G).* Only database compounds within the mass range of [M] ±1 *Da* are considered candidates of known NRPs and ultimately subjected to P_1 and P_2 calculations.

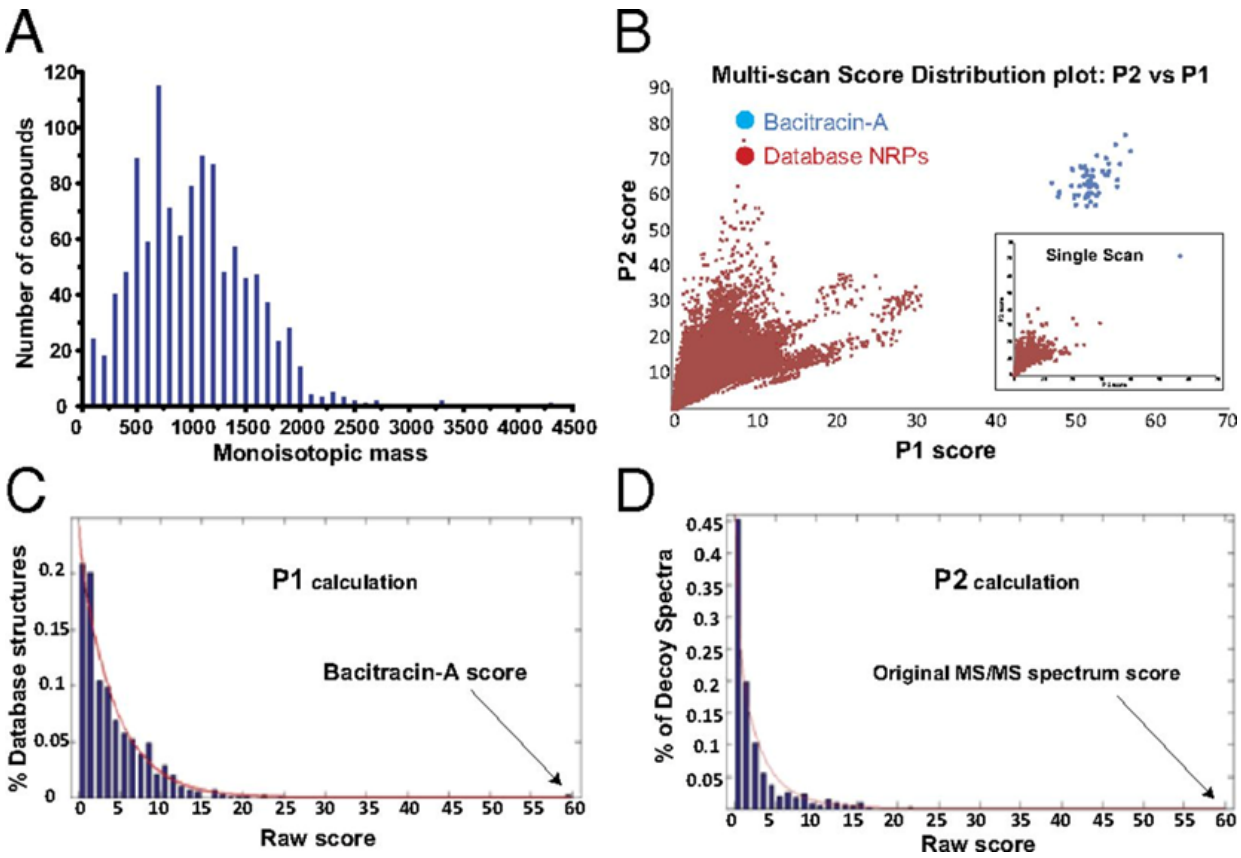


Figure 2. iSNAP scoring scheme. (A) Histogram representing the hypothetical spectral library of 1107 compounds. **(B)** Dereplicating bacitracin-A in 1B using the doubly protonated (+711.82 m/z) MS/MS spectra. Multiple MS/MS scans are generated from ~1 min direct infusion of bacitracin-A; each blue point indicates a match between an MS/MS spectra and bacitracin-A. The red points show the score distribution of the other matches between the MS/MS spectra and rest of the 1106 database NRPs. The score distribution plots indicate the capability of the P_1 and P_2 scores, in distinguishing true and false matches. **(C)** P_1 score calculation of bacitracin-A. The raw score distribution is generated by scoring the MS2 spectrum against database compounds within the 0 Da to [M]+100 Da mass range. The Raw matching score of the bacitracin-A candidate is 59.1, where the *p*-value on the distribution is 1.74e-006. The P_1 score is calculated as -10log10(*p*-value)=57.6. The fitted gamma distribution is shown as the red curve. **(D)** P_2 score calculation of bacitracin-A. The Raw score distribution is generated by scoring each decoy spectrum against bacitracin-A. The original spectrum has a raw score of 59.1, which is greater than that of the decoy spectra. The *p*-value on the distribution is 5.87e-008, with a P_2 score calculated as -10log10(*p*-value)= 72.31.

P_1 score

A P_1 score is introduced as a normalized version of the Raw score in order to add statistical significance. Empirically, when an MS/MS spectrum is scored against all database compounds within the 0 *Da* to [M]+100 *Da* mass range, the statistical distribution of the Raw scores closely fits a gamma distribution (Fig. 2*C*). In figure 2C, the fitted gamma distribution is shown as a red curve. The parameters required for a gamma distribution are estimated with the

maximum-likelihood method. For each compound, the *p-value* is the exceedance frequency at the compound's Raw score, which is the area under the curve and to the right of the raw score. The *p*-value represents the probability of a random structure scoring higher with the MS/MS spectrum than the correct structure. A low *p*-value indicates the match is unlikely random and therefore is likely a correct one. The P_1 score is calculated as -10log10(p-value).

P_2 score

While the P_1 score measures the significance of the candidate structure as compared with other NRP structures in the database, a P_2 score is used to measure the significance of the MS/MS spectrum compared with artificially generated "decoy" spectra. If the MS/MS spectrum S is from an NRP structure, then the structure should be scored significantly higher using S than using the artificially generated decoy spectra. Suppose the spectrum S has a mass range from m_1 to m_2. To generate a decoy spectrum, the m/z value of each peak in S is shifted by an integer Δm. More specifically, an m/z value x is changed to $x + \Delta m$ if $x + \Delta m \leq m_2$; and to $x + \Delta m - m_2 + m_1$ if $x + \Delta m > m_2$. Thus, by trying every integer Δm between 1 and $m_2 - m_1$, many decoy spectra can be obtained. The shifting method is inspired by the calculation of cross-correlation score in the SEQUEST algorithm, which was the first computer algorithm for matching ribosomal peptides in a database with MS/MS spectral data (21). A gamma distribution is then estimated from the Raw scores between the decoy spectra and the candidate structure. The *p*-value is the exceedance frequency at the original MS/MS spectrum's raw score (Fig. 2*D*). The P_2 score is calculated as -10log10(p-value).

Hypothetical Spectral Library Matching studies with known Nonribosomal Peptides

iSNAP is designed to analyze individual spectra and reveal the significance of a match between MS/MS spectra and candidate NRP compounds (those within a mass range of [M] ±1

Da). For each MS/MS spectrum with established candidates, a P_1 score and P_2 score is generated for each candidate. A training experiment using six pure NRPs (bacitracin-A, cyclosporin-A, gramicidin A, polymyxin-B, surfactin, and seglitide) were used to reveal a threshold needed for true positive identification from P_1 and P_2 scores. We rationalized the selection of the six NRPs for the training experiment based on structural complexity, backbone modification (e.g. N-methylated amides, amides replaced by esters, and polyketide extended amino acid building blocks), and variance in chemical architecture (linear, cyclic, branched). The expectation from this test set; a true candidate match will have a distinctively higher P_1 and P_2 scores (additional details within *SI Appendix*, Section I).

An initial test with the branched cyclic NRP, bacitracin-A was conducted to reveal whether the designed scoring strategy would result in the true candidate having a distinctively higher P_1 and P_2 scores than those of other database structures. The resulting spectrum from an infusion experiment consisted of 56 bacitracin-A MS/MS scans and using the scoring scheme, without mass filtering ([M] ±1 *Da*), produced bacitracin-A as the top ranking hit and distinguishably higher than other 1106 database NRPs (see multi-scan score distribution plot of P_2 vs P_1 scores Fig. 2*B*).

Applying the scoring scheme and [M] ±1 *Da* filter, pure standards of the five additional test compounds cyclosporin-A, gramicidin, polymyxin-B, surfactin and seglitide underwent manual MS/MS and automated data dependent acquisitions (DDA). In the case of seglitide, a purely cyclic peptide, a doubly protonated $[M+2H]^{2+}$ species within scan #10, underwent a single stage of tandem MS and scored (P_1 =57.5, and P_2 = 48.2) with 17 out of 30 b-ions and 27 matched mass-to-charge values. Another cyclic peptide, polymyxin-B whose complexity derives from repetitive blocks (six a,g-diaminobutyric acid residues), had the second highest

number of total matched peaks at 59 with 33 b-ions matched yielding a P_1 =35.1, and P_2 = 35.0. Matched peaks comprised of repeat amino acid units were of relative low intensity for 4 of 6 monomers. Fragmentation pattern derived from macrocylic ring-opening, acyl chain loss and a diaminobutyric acid monomer (+963.6, +863.5 and +241 m/z) is consistent as the major pathway of fragmentation (22). In the case of cyclosporin-A, iSNAP dereplicated the structure despite the N-methylated peptide back-bone. N-methylation limits peptide cleavage as the amide bond is unable to be protonated through intramolecular proton transfer, thus additional stability is gained by increasing the basicity of its neighboring carbonyl group, favoring a C-terminal fragmentation pathway and the generation of y-ions (23). The highest scoring MS/MS scan came from acquisition #28, and a total of 27 hSFs were matched to the real MS/MS spectra. Of these, 25 were b-ions, a quarter of all possible b-ion fragments and score values of P_1 =35.1, and P_2 = 41.8. In the case of linear polypeptide gramicidin, only 5 of 85 b-ions were generated in the MS-experiment and identified (within scan #19), and overall 13 matched mass-to-charge values were sufficient for dereplication with scores above threshold cut-offs, P_1 =34.6, and P_2 = 40.7. In the case of another cyclic-branching peptide surfactin, 29 low intensity (<10%) peaks were matched in scan #18 of which 22 were b-ions (P_1 =28.5, and P_2= 31.2).

Establishing iSNAP cut-offs for true and false positive rate identification

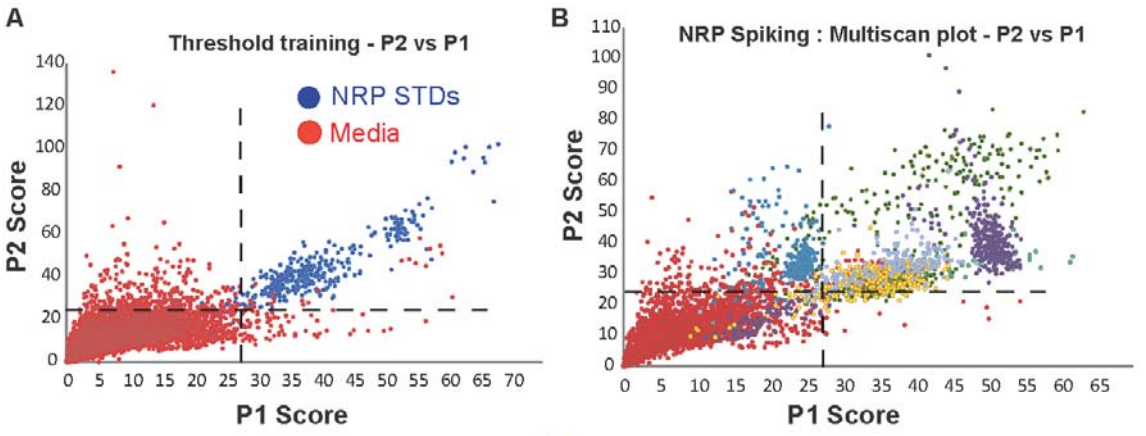
Early stage dereplication of natural product extracts is a key goal of modern natural product screening programs and we probed whether iSNAP enables non-targeted dereplication of known compounds in complex mixtures using low resolution tandem mass spectrometry. Optimized MS/MS and LC-MS/MS settings for optimal P_1 and P_2 scoring and non-targeted dereplication were realized by testing mass resolution (u/sec), activation energy (q), isolation

width (m/z), and data-dependent acquisition (DDA) settings (see *SI Appendix*, Section I.*E*, Data set 2).

DDA acquisitions were performed under the AutoMS/MS setting with the available tuning option active, Smart parameter setting (SPS). A scan range of 100-2000 m/z was selected with precursors over 300 m/z targeted for MS/MS using the active exclusion option set to eight spectra over a release time of 0.25min. The active exclusion feature enables the targeting of lower abundance ions by de-selecting and not fragmenting more abundant ions. Ten precursor ions were selected for MS/MS using the enhanced resolution mode and baseline intensity threshold of 6 x 10^5, with an isolation width of 4 m/z. P_1 and P_2 threshold cut-offs were determined through a combination of two MS/MS experiments. In the first experiment MS/MS spectra were generated from NRP working standards (direct infusion) and the iSNAP scores (P_1 and P_2 scores) used as positive controls in the threshold training (Fig. 3*A*). The second experiment, LC-MS/MS data derived from scanning of eleven common fermentation media (no NRPs added) was used to investigate false matching (see *SI Appendix*, Section I.*D*). As no NRP compounds exist within those matrixes, matches to NRPs within the iSNAP database must be considered as falsely matched and these low P_1 and P_2 scores are used as the negative controls (see *SI Appendix*, Section II. Fig. S3-S4). By combining the true or correct NRP database matches (NRP working standards) with the negative control false matches in a P_2 *vs* P_1 scatter plot, P_1 and P_2 threshold cut-offs were empirically derived (Fig. 3*A*). Candidates with P_1 and P_2 scores above 27 and 24, respectively are considered dereplicated or positively identified. Using the estimated thresholds, 335 of 367 MS/MS scans are identified as true candidates, with a true positive rate of 91.3%, while 24 of 6744 register as false positives, with a false positive rate of 0.0036 %, from the 11 fermentation media. In an effort to further reduce false positive hits, additional filtering is applied to candidate matches with P_1 and P_2 scores

above the empirical threshold. Candidates with less than 4 matched peaks were determined to contribute to false matches, while candidate matches with less than ten matched peaks, of which more than 75 percent are derived from low intensities (<2%) are also excluded.

The output of the iSNAP analysis is a complete report for each MS/MS scan (see *SI Appendix*, Section I.*A)*; showing the scan number, retention time, precursor m/z, charge state, precursor mass and the outputted candidates name, mass and SMILES code and number matched fragments, Raw score, P_1 score and P_2 score (see *SI Appendix*, Section II. Fig. S1-S2).



DDA dereplication of nonribosomal peptides in fermentation media

Compound	Average Rt (min.)	Precursor m/z	Precursor charge	Candidate Mass	MS/MS Peak matches	iSNAP # hSFs	P1 Score*	P2 Score
Bacitracin A	20.92	712.3	2	1421.7	45 - 104	301	27.5 - 66.0	31.9 - 82.8
Cyclosporin A	46.66	602.3	2	1201.8	134 - 163	329	61.5 - 64.4	32.6 - 42.3
Gramicidin A	43.53	942.0	2	1881.1	73 - 117	375	32.2 - 43.6	30.4 - 44.4
Polymyxin B	25.03	402.2	3	1202.8	46 - 58	249	27.1 - 27.9	36.9 - 47.3
Seglitide	23.14	809.8	1	808.4	44 - 47	96	45.6 - 59.2	29.0 - 37.4
Surfactin	50.86	1037.3	1	1035.7	22 - 24	127	32.9 - 36.3	29.1 - 33.6

*Multiple MS/MS scans are analyzed for each NRP. The scores of the MS/MS spectra with the highest P1 score from accross the 11 media panel is presented.

Figure 3. iSNAP threshold determination and complex mixture analysis. (A) MS/MS spectra from the six NRPs standards (in blue) obtained by direct infusion experiments, overlaid with over 6,500 MS/MS spectra from LC-MS/MS analysis of 11 microbial fermentation media, n=3, (in red). The fermentation media represent the blank control. Empirical threshold cut-offs are estimated, P_1=27 and P_2 =24. **(B)** NRPs standards are spiked and extracted from the 11 media and subjected to LC-MS/MS analysis and iSNAP dereplication. (**C**) iSNAP results from B, with the highest scoring MS/MS spectra from across the 11 media panel reported.

Probing iSNAP fidelity in data-dependent acquisition (DDA) within different fermentation conditions and groupings of nonribosomal peptides

To reveal the suitability and fidelity of the iSNAP algorithm for screening extracts a series of liquid media varying in their spectrum of use (differing natural product producers), nutrient and peptide composition were subject to LC-MS/MS and iSNAP analysis to reveal their contributions to potential false-positives. This panel of eleven different microbial fermentation media used for fermentation of NRP producers (myxobacteria, streptomycetes and other actinobacteria, pseudomonads, bacilli and filamentous fungi) used included: YPD (Yeast protein, milk protein), YMPG (yeast, malt, peptone, glucose), GYM (yeast, malt), TSB (soy protein), LB (peptone peptides and yeast protein), nutrient (beef and meat peptides from meat infusion solids), pharmamedia (cotton seed protein), grass seed veg (grass seed extract proteins), fishmeal (Fish meal protein), R2A (proteose peptone, casamino acids, yeast proteins), CY (casitone, yeast). In each of these cases we designed the experiment based on a typical volume of fermentation media used in screening (50 mL cultures) and a final amount of 50 ng of a given NRP analyzed by the mass spectrometer. A panel of NRPs were spiked into each media (final 50 μg/mL), and the mixture extracted with organic solvent and subjected to LC-MS/MS analysis using DDA settings (See SI Appendix, Fig. 3*B*). In these instances, we also determined the true and false positive rate for the study. For this we sought to determine the number of MS/MS spectra acquired for each spiked media, the number of MS/MS spectra matched to the iSNAP database, MS/MS spectra from spiked-NRPs, and false matches (see *SI Appendix*, Section III. Table S1).

Automated LC-MS/MS analysis of the eleven NRP spiked fermentation media revealed as expected a variance in the numbers of product ions, with 485 being the average. In the case

of R2A spiked media, a total of 192 MS/MS spectra were matched to product ion spectra and their m/z off-sets, which are derived from the six NRP candidates, of these, 126 scans were above the P_1 and P_2 cutoff. The false positive rate for R2A is calculated as the total number of MS/MS spectra (minus NRP candidates), divided by the total number of candidates with false positive hits. The false positive rate was determined to be 0.83% for R2A with only one false positive hit (see *SI Appendix*, Section II. Fig. S5). Media's YMPG and Grasseed had zero false positives detected, while the remaining media panel had between 1-4 false positive hits.

In each instance, where an NRP's product ion spectrum is generated from the spiked media extracts, iSNAP made a positive identification (Fig. 3*C*). However, in certain cases, some of the fermentation media had no product ions generated for polymyxin-B (ie. YPD, YMPG, TSB, and Grasseed), and seglitide (ie. YPD, TSB, LB, and CY). Poor extraction efficiency, compound instability, or ion suppression in these matrices is the likely origin. Importantly these studies reveal iSNAP conducts true dereplication in a non-targeted fashion for a series of structural diverse NRPs from various complex matrixes with an average iSNAP processing times of under a minute for each LC-MS/MS data file. The P_1 and P_2 scores of the most representative candidates for each of the six NRP spike-in compounds and media candidates are plotted in Fig. 3*B* with the LC-MS/MS results from the DDA analysis in Fig. 3*C*, highlighting the top scores across the media panels (see *SI Appendix*, Section III. Table S2-S3). As multiple MS/MS scans can be generated for each NRP compound, at least one scan must have an NRP candidate scored above the P_1 and P_2 thresholds for a dereplication to be made.

In the NRP spiking studies, four low scoring false positives were identified with P_1 and P_2 scores from 27-34 and 25-34 respectively. The four false positive hits have been attributed to three compounds; esperin, empedopeptin and tyrocidine C (see *SI Appendix*, Section II. Fig.

S6). Analysis of the detailed iSNAP report revealed that surfactin's MS/MS spectrum was incorrectly matched to that of esperin (as revealed by retention time and fragment analysis). However, the false matching of surfactin to esperin can be rationalized as they are structurally similar cyclic depsipeptides, with C_{13}-C_{15} acyl chains, common monomer building blocks (L-Glu, D-Leu, and L-Asp), and esperin being within a [M] ±1*Da* mass range of surfactin. In comparing the P_1 and P_2 scores, esperin's are lower than that of surfactin. Analysis of surfactin's iSNAP results and matching hits has also revealed that MS/MS spectral data may be useful in revealing analogs. In the case of empedopeptin and tyrocidine C, they were matched to analytes arising from two fermentation media (LB and CY).

Dereplicating complex NRPs by data-dependent acquisition: Kutzneride

Kutznerides are among the most complex NRPs, composed entirely of non-proteinogenic amino acids including several halogenated and oxidized groups (24). We sought to test if iSNAP could dereplicate these from extracts in a non-targeted fashion using DDA and whether halogenated analogs could be detected (see *SI Appendix*, Section I.*C*). Supernatants from *Kutzneria* sp. 744 grown in complex Merlin Norkans medium were extracted with HP20 resin and subjected to solvent partitioning, with organic fractions subjected to LC-MS/MS analysis. Untargeted automated analysis by iSNAP dereplicated kutzneride-1 with matched fragment peaks (+837.3, 836.3, 743.2, and 609.2 m/z). The matched fragment ions can be correlated to cleavage at the lactone ring opening (-17, -18,), and subsequent amide cleavages (-111 m/z and -245 m/z) between the 6,7-dichloro-3a-hydroxy-1,2,3,3a,8,-8a hexahydropyrrolo[2,3-*b*]indole-2-carboxylic acid and the 3-hydroxyglutamine residue (+609.2 m/z). Positive identification of kutzneride-1 was achieved using iSNAP with P_1 and P_2 scores of 31.3 and 33.4 respectively.

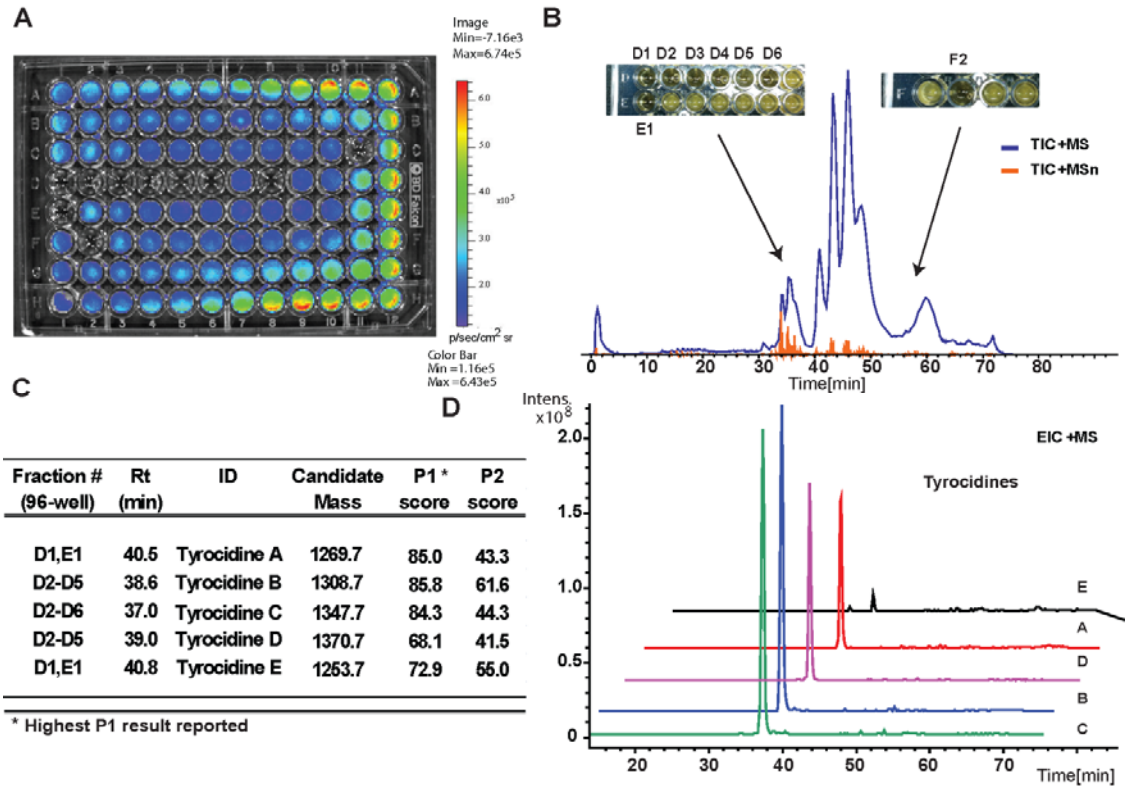
Frequently in modern natural product discovery simple variants of known NRP families are revealed in screening efforts. As such, it would therefore be useful to dereplicate 'probable' variants of knowns (e.g. methylated, hydroxylated or halogenated). We used the kutzneride producer to probe whether hypothetical variants of the known NRP could be detected using iSNAP. To promote the formation of a new kutzneride, we grew the producing strain in a medium containing bromide salts, replacing the original chloride ones. We anticipated that brominated kutznerides would be biosynthesized as halogenases are known to accept either halide. As expected, the LC-MS/MS chromatogram of the resulting extract indicated the presence of the dibromo-kutzneride analog with a molecular weight of +942.1 $[M+H]^+$ and absence of kutzneride-1 (see *SI Appendix*, Section II. Fig. S7). Analyzing this kutzneride fraction with iSNAP did not generate hits (despite a wide candidate window of [M] +/-150 Da), and did not reveal false positives by scoring with the original kutzneride-1. Adding the dibromo-kutzneride SMILES code to the database and rerunning the previous spectra revealed that 4 high intensity fragment peaks were identified from the MS/MS spectra (+ 942.2, + 925.2, + 924.2 and + 830.2 m/z), an analogous fragmentation sequence as seen for kutzneride-1, with P_1 and P_2 score values of 75.9 and 29.3 respectively (see Data set 2). These experiments highlight the utility of the iSNAP upload feature, and how iSNAP can be used to reveal variants of known complex nonribosomal peptides.

Probing the utility of iSNAP to interrogate complex extracts and dereplicate known compounds

Natural product screening campaigns often use bioactivity-guided fractionation to isolate active compounds. To explore how iSNAP may assist in dereplication within a bioactivity-guided fractionation campaign we applied it to a screening of natural products for

anti-staphylococcal agents. One of the natural product extracts derived from an environmental bacillus produced a large zone of inhibition using agar disk diffusion assays. The extract was subjected to LC-MS/MS and coordinate time-dependent fractionation into a 96-well plate. Bioactivity assays were conducted with the resulting 96-well plate with a bioluminescent *Staphylococcus aureus* strain Xen29, and the LC/MS file uploaded onto iSNAP (Fig. 4*A*).

In the analysis of a crude pellet extract, a total of 1964 MS/MS scan were acquired over a 75min LC-MS/MS run, and of these, 45 had $P_1$and P_2 scores above the threshold cut off and 41 were for members of the tyrocidine family (25). Collectively these 41 tyrocidine matches correlated with wells D1-6, D8, E1, which all lacked *S. aureus* growth (see *SI Appendix*, Section II. Fig. S8-S10). iSNAP scoring revealed high P_1 and P_2 scores for tyrocidine A (P_1= 85, P_2 = 43.3), B (P_1= 85.8, P_2= 61.6), C (P_1= 84.3, P_2 = 44.3), D (P_1=68.1, P_2 = 41.5), and E (P_1=72.9, P_2=55.0), from their double protonated precursor masses of + 636.2, + 655.8, + 675.3, + 686.8, and + 628.2 m/z, respectively (Fig. 4*B-D*). High resolution mass determination of the dereplicated candidates using LTQ-Orbitrap HRS-FTMS measurements revealed the candidates were within ~0.6 - 4 ppm of the tyrocidines (see *SI Appendix*, Section III. Table S4). And further comparison of the MS-MS fragmentation pattern of authentic tyrocidines with the candidates laddering b-ions, acylium ions, (see *SI Appendix*, Section II. Fig. S11-S12) provided confirmatory evidence (26). The positive identification of each tyrocidine analog, and distinguishing between them, with increased P_1 and P_2 scores highlights the selectivity of iSNAP and detection of low abundance analogs (i.e. tyrocidine E, relative abundance is 2 %).



Fraction # (96-well)	Rt (min)	ID	Candidate Mass	P1 * score	P2 score
D1,E1	40.5	Tyrocidine A	1269.7	85.0	43.3
D2-D5	38.6	Tyrocidine B	1308.7	85.8	61.6
D2-D6	37.0	Tyrocidine C	1347.7	84.3	44.3
D2-D5	39.0	Tyrocidine D	1370.7	68.1	41.5
D1,E1	40.8	Tyrocidine E	1253.7	72.9	55.0

* Highest P1 result reported

Figure 4. Dereplicating bioactives from *Bacillus sp.* (**A**) IVIS bioluminescence imaging of crude fermentation extracts of *Bacillus sp.* against *Staph. aureus* (Xen29 strain), following HPLC fractionation. **(B)** LC-MS/MS chromatogram of *Bacillus sp.* extract. Total ion chromatograms (TIC) for MS and MS(n) shown, bioactive wells highlighted. (**C**) iSNAP dereplication results identifying a series of tyrocidines from the inputted LC-MS/MS data file in .mzXML format. (**D**) Extracted ion chromatogram (EIC) of the five dereplicated tyrocidines.

The remaining 4 MS/MS spectral matches were identified as belonging to 3 compounds (see *SI Appendix*, Section II. Fig. S13); capreomycin IB (P_1=28, P_2=39.4), emerimicin III (P_1=28.6, P_2 =27.9), and nepadutant (P_1=29.7, P_2=57.9). Of note, however upon further investigation, capreomycin and nepadutant had only 4 matched fragments, with only one high intensity peak contributing significantly to the scoring scheme. Given these findings, we suggest that MS/MS spectra with low matched peaks should be further examined for positive dereplication (see *SI Appendix*, Section II. Fig. S14-S15).

Discussion

Nonribosomal peptides comprise a highly privileged section of chemical space, which is diverse due to varied use of over 500 building blocks and molecular architectures (cyclic, linear, branched) and modifications and fusions with other chemical classes (i.e. polyketides). Critical to new nonribosomal peptide natural product discovery is efficient dereplication within complex extracts in a non-directed fashion. iSNAP is the first strategy to achieve this and we have shown it is applicable to a spectrum of nonribosomal peptide types, linear, cyclic, branched (linear and cyclic portions) and those with highly modified subunits (e.g. halogenation), mixed backbone linkages (e.g. lactones, N-methylated amides) and polyketide extensions. False positive scores were evaluated in a number of matrices and shown to be relatively insignificant in all the media tested. Through this design we have created a platform that is robust enough to tackle a battery of differing media compositions dereplicated the correct NRP at low nanogram levels from complex matrixes in an un-targeted fashion, using a relative low-resolution mass spectrometer. While the current version of iSNAP dereplicates, an enhanced ability may be realized by isotopic labeling. The design of iSNAP and its flexible use of informatic databases of natural product SMILES codes may provide a mechanism to couple needs of dereplication with the discovery potential of novel substances revealed by microbial genomic sequencing.

Methods

Details relating to the materials used, bacterial strains, culture conditions, the isolation and purification of the kutznerides and tyrocidines, fermentation media screening conditions and NRP compound spiking; Mass Spectrometry; MS/MS and LC-MS/MS Experiments, access

to the data files and user guide for iSNAP can be found in the *SI Appendix*, Section I.*A*-*H*. The iSNAP online research tool is available at http://www-novo.cs.uwaterloo.ca:8180/isnap.

Supplemental Information

http://www.pnas.org/content/suppl/2012/11/02/1206376109.DCSupplemental

Appendix (PDF)

Dataset_S01 (XLSX)

Dataset_S02 (XLSX)

References

1. Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. *Nat Prod Rep* 20:275-287.
2. Fischbach M, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, macinery and mechanisms. *Chem Rev* 106:3468-3496.
3. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G (2008) NORINE: database of nonribosomal peptides. *Nucl Acids Res* 36:326-331.
4. Li JW, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier. *Science* 325:161-165.
5. McAlpine JB (2009) Advances in the understanding and use of the genomic base of microbial secondary metabolite biosynthesis for the discovery of new natural products. *J Nat Prod* 72:566-572.
6. Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. *Nat Prod Rep* 26:977-986.
7. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567.
8. Finking R, Marahiel MA(2004) Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol* 58:453-488.
9. Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B (2006) Scrabling of seqiuence information in collision-induced dissociation of peptides. *J Am Chem Soc* 128:10364-5.
10. Eckart K (1994) Mass spectrometry of cyclic peptides. *Mass spectrometry Reviews* 12: 23-55.

11. Bleiholder C, et al. (2008) Sequence-scrambling fragmentation pathways of protonated peptides. *J Am Chem Soc* 130:17774-17789.

12. Liu WT, et al. (2009) Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. *Anal Chem* 81:4200-4209.

13. Mohimani H, et al. (2011) Sequencing cyclic peptides by multistage mass spectrometry. *Proteomics* 11:3642-50.

14. Ng J, et al. (2009) Dereplication and de novo sequencing of nonribosomal peptides. *Nat Methods* 6:596-600.

15. Kersten, et al. (2011) A mass spectrometry-guided genome mining approach for natural products peptidogenomics. Nature Chemical Biology 7:794-802.

16. Weininger D, (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comut Sci* 28:31-36.

17. Paizes B, Suhai S (2004) Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews* 24:508-548.

18. Chamrad DC, et al. (2004) Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics* 4:619-628.

19. Eng J, McCormiack AL,Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein Database. *J Am Soc Mass Spectrom* 5:976-989.

20. Zhang J, et al. (2011) PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell proteomics* 11(4):M111.010587.

21. Razumovskaya J, et al. (2004) A computational method for assessing peptide-identification reliability in tandem mass spectrometry analysis with SEQUEST. *Proteomics* 4:961-969.

22. Govaerts C, at al. (2002) Mass spectrometric fragmentation of cyclic peptides belonging to the polymyxin and colistin antibiotics studied by ion trap and quadrupole/orthogonal-acceleration time-of-flight technology. *Rapid Commun Mass Spectrom* 16(9):823-33.

23. Vaisar T, Urban J (1998) Gas-phase Fragmentation of protonated mono-n-methylated peptides. Analogy with solution-phase acid-catalyzed hyrdolysis. *J Mass Spectrom* 33:505-525.

24. Broerg A, Menkis A, Vasiliauskas R (2006) Kutznerides 1-4, Depsipeptides from the actinomycete *Kutzneria sp. 744* inhabiting mycorrhizal roots of Picea abies seedlings. *J Nat Prod* 69:97-102.

25. Barber M, et al. (1992) An investigation of the tyrothricin complex by tandem mass spectrometry. *Int J Mass Spectrom Ion Processes* 122:143-151

26. Tang X, Thibault P, Boyd R (1992) Characterisation of the tyrocidine and gramicidin fraction of the tyrothricin complex from *Bacillus brevis* using liquid chromatography and mass spectrometry. *Int J Mass Spectrom Ion Processes* 122:153-179.

CHAPTER THREE

EXPLORATION OF NONRIBOSOMAL PEPTIDE FAMILES WITH AN AUTOMATED INFORMATIC SEARCH ALGORITHM

CHAPTER THREE PREFACE

The following works was previously published in:

Yang, L.*, Ibrahim, A.*, Johnston, C. W.*, Skinnider, M.A., Ma, B., Magarvey, N.A. Exploration of Nonribosomal Peptide Families with an Automated Informatic Search Algorithm. Cell Chemistry and Biology 22, 1259-1269 (2015).

* The authors contributed equally to this work.

DOI: http://dx.doi.org/10.1016/j.chembiol.2015.08.008

Author Contributions:

L.Y. and A.I. conceived of and designed the iSNAP analog algorithm. L.Y. wrote the software. M.A.S. designed the user interface. L.Y., A.I., and C.W.J. performed software testing and method validation. A.I. and C.W.J. performed experiments and analyzed data. A.I. cultured strains for tyrocidines and WS9326A. C.W.J. cultured strains for loloatins, LI-F0 antibiotics, and arylomycins, isolated compounds, performed structural analysis, and characterized new compounds. L.Y., A.I., C.W.J., B.M., and N.A.M. contributed to study design. L.Y., A.I., C.W.J., B.M., and N.A.M wrote the manuscript.

ABSTRACT

Microbial natural products are some of the most important pharmaceutical agents and possess unparalleled chemical diversity. Here we present an untargeted metabolomics algorithm that builds on our validated iSNAP platform to rapidly identify families of peptide natural products. By utilizing known or *in silico*-dereplicated seed structures, this algorithm screens tandem MS data to elaborate extensive molecular families within crude microbial culture extracts with high confidence and statistical significance. Analysis of peptide natural product producers revealed an abundance of unreported congeners, revealing one of the largest families of natural products described to date, as well as a novel variant with greater potency. These findings demonstrate the effectiveness of the iSNAP platform as an accurate tool for rapidly profiling large families of nonribosomal peptides.

INTRODUCTION

Small molecule natural products have long served as a valuable source of pharmaceuticals, providing molecular scaffolds useful in treating an ever-expanding range of pathologies. Microbial natural products – particularly the polyketides and nonribosomal peptides – have proven particularly useful as a source of antibacterial drugs and scaffolds, making up roughly two thirds of clinical antibacterials since 1981 (Newman and Cragg, 2007). Polyketides and nonribosomal peptides are produced by modular assembly line-like enzymes (PKSs and NRPSs) that frequently display promiscuity in substrate selection and chemical tailoring reactions, giving rise to molecular families based on a set scaffold. By utilizing this promiscuity, a single assembly-line can deploy a library of molecules that can possess divergent affinities for a given target, or even different activities (Yu et al., 2012). As the forces that drive the evolution of these molecules in the environment are typically not directed towards clinical

needs, minor analogues can often be identified with substantially improved clinical utility, including drugs and leads such as mannopeptimycin (He et al., 2002), rhizoxin (Scherlach et al., 2006), epothilone (Chou et al., 1998), pneumocandin (Balkovec et al., 2014), and burkholdine (Lin et al., 2012). Similarly, minor differences in homologous assembly lines (observable in sequenced genomes and metagenomes) can produce related natural products with superior activity (Wang et al., 2011). In each instance, variations in substrate- or monomer-selection promiscuity by NRPS/PKS enzymes can lead to the generation of natural bio-synthetic libraries with a range of target affinities and biological effects. Traditionally, natural product discovery efforts have led to the most abundant members being targeted through bioassay guided fractionation, as bioactivity is concentration-dependent. Such limited sampling can lead to an underestimation of a scaffolds potential as a therapeutic agent, as superior minor variants may go undiscovered. Reanalysis and re-engagement of over-looked natural product scaffolds will likely facilitate the discovery of analogues with improved pharmaceutical properties, providing new clinically-useful structures and scaffolds for further development.

Traditional natural products chemistry approaches based on bioactivity-guided isolation and rigorous structure elucidation techniques have provided the vast majority of pharmaceutically-relevant natural products. Unfortunately, these techniques are biased towards abundant compounds and necessitate large amounts of material, expertise, and time – often requiring months of work for the identification, isolation, and elucidation of a single natural product. Re-analyzing microbial culture extracts for over-looked molecules will require techniques that can provide systems-level analysis and define both structures and retention times to prioritize and facilitate subsequent isolation efforts. This requirement is complicated by the diversity of natural product families, whose exotic molecular scaffolds confound the

development of automated analyses, and whose frequent isobaric species prevent accurate and automated structure elucidation. While genome sequencing data can provide predictive value in guiding discovery efforts (Kersten et al., 2011), obtaining sufficiently well-assembled biosynthetic gene clusters for accurate structure predictions can be time consuming, impeding discovery and development efforts. Pure, untargeted metabolomic approaches that make use of liquid chromatography-coupled mass spectrometry (LC-MS) remain a powerful means of rapidly assessing chemical potential (Winnikoff et al., 2014; Yang et al., 2013; Hou et al., 2012), and are uniquely positioned to facilitate high-throughput, information-rich analysis of complicated microbial extract libraries with high sensitivity. Mirroring advances in LC-MS-based proteomics, a number of targeted metabolomic approaches have been explored for the detection of peptide natural products (Kersten et al., 2011; Mohimani et al., 2011; Mohimani et al., 2014; Mohimani et al., 2014), which reliably fragment along amide bonds but frequently possess complicated architectures. Following pioneering work in 2009 defining a *de novo* means of sequencing cyclic peptides (Ng et al., 2009), a number of approaches with varying degrees of automation have worked towards rapid, accurate detection of complex peptide sequences first from standards (Mohimani et al., 2011) and then from microbial culture extracts (Kersten et al., 2011; Mohimani et al., 2014; Mohimani et al., 2014). In 2012, we defined iSNAP: a novel algorithm capable of detecting known peptide natural products from complex culture extracts (Ibrahim et al., 2012). Taking advantage of a comprehensive in-house database of known peptide natural product structures, this automated 'dereplication' technology screens LC-MS/MS data and applies a series of statistical processes to identify known compounds based on matching peaks between real and *in silico* MS/MS fragments. Here, we present a new analogue module for the iSNAP platform (available at http://magarveylab.ca/analogue/), facilitating both dereplication and highly accurate analogue identification; capable of discerning

between isobaric species, and automatically providing structures for novel, superior analogues from nanograms of material.

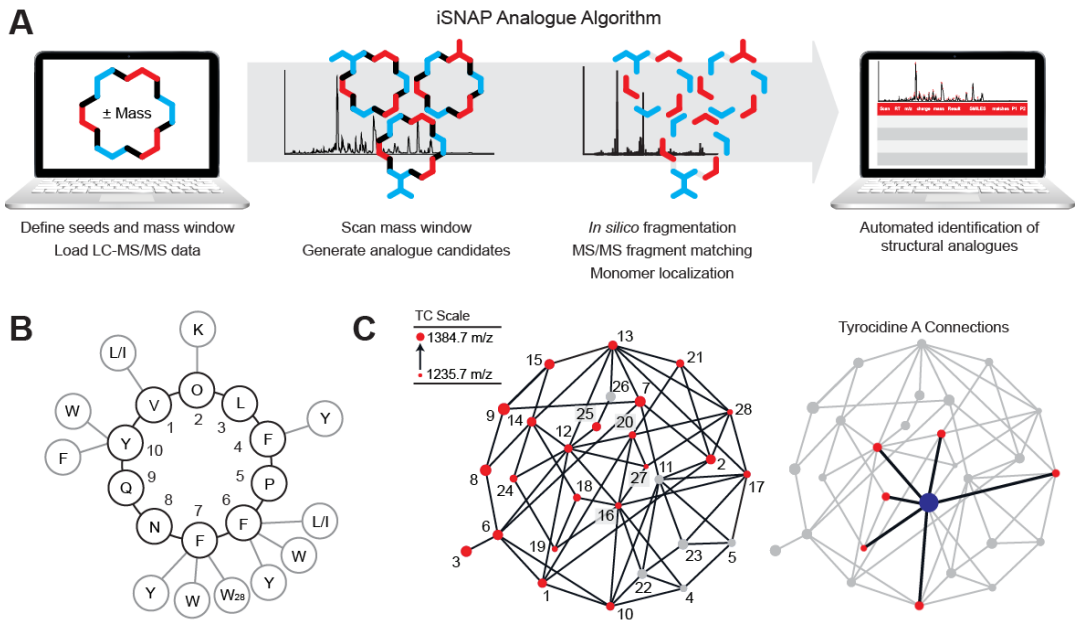


Figure 1. iSNAP analogue discovery algorithm can map families of natural peptides.

(A) Schematic work flow of the automated iSNAP analogue algorithm. After uploading an LC-MS/MS data file (.mzXML) users are required to define seed structures of interest – either predefined or following initial iSNAP dereplication analysis – in addition to a mass window around these seeds. Following dereplication, the algorithm searches the defined mass window and generates candidate analogue structures. Hypothetical candidates are fragmented in silico to facilitate MS/MS fragment matching and monomer localization, driving the automated identification of structural analogues from the initial seed. (B) Amino acid composition of tyrocidine A (tyrocidine 16; black) and various substitutions found in the remaining 27 known members of the tyrocidine family of cyclic peptide natural products (gray). (C) Tyrocidine single monomer substitution networks. Structures of tyrocidines 1-28 dereplicated by iSNAP (red) or known from previous literature (gray) are connected to one another through single amino acid substitutions (left). Through iSNAP analogue analysis, single seeds can be used to access related structures, demonstrated using tyrocidine A (right).

RESULTS

iSNAP Analogue Search Algorithm

iSNAP analogue search is an algorithm designed for the discovery of novel analogues of known peptide natural products from tandem mass spectral data. It is built on the iSNAP platform to extend the platform's capability of elucidating families of peptide natural products. As it is known that families of peptide natural products can co-exist in microbial culture extracts, the dereplication of a peptide is often a good indication to the existence of analogues

that have similar structures. In line with this observation, the algorithm is designed to utilize dereplicated peptides as seed structures to guide the search for analogues with similar structures.

The analogue search algorithm takes tandem mass spectral data and a list of seed structures as inputs. It analyzes each tandem mass spectrum individually in the following steps: (1) construct analogue candidates that are one monomer different from seed structures; (2) match the spectrum to hypothetical spectra of analogue candidates; (3) evaluate the matches by calculating p-value derived scores, P1 and P2; (4) report the identified analogue, if the best match has scores above specified thresholds (Figure 1A, Figure S1). The algorithm reports the seed structure, the monomer site of difference, and the mass difference for each identified analogue.

Construction of Analogue Candidates

Assuming the parent mass of a tandem mass spectrum is M and the mass values of seed S_1, S_2, …, S_n are $m(S_1)$, $m(S_2)$, …, $m(S_n)$, the algorithm first compares M to the mass of each seed. A seed is selected to generate analogue candidates if the mass difference is smaller than a user-specified threshold M_T. We denote the selected seeds as $\hat{S}_1$, $\hat{S}_2$, …, $\hat{S}_m$, and we have

$$\hat{S} \in \{S_i \mid |M - m(S_i)| < M_T\} \quad where\ i = 1 \ldots n$$

For each selected seed $\hat{S}$, the algorithm annotates amide bonds and ester bonds in its structure. Monomer blocks between two adjacent bonds are therefore detected and numbered as R_1, R_2, …, R_r. The algorithm assumes that the true structure for the spectrum is an analogue that can be constructed from the seed by modifying one monomer. The modified monomer accounts for the mass difference between the seed and the spectrum parent mass. As such, analogue candidates are generated by adding the mass difference to each and every monomer block in the seed.

$$\begin{aligned} \hat{S}'_1 &\longleftarrow R'_1, R_2, \dots, R_r \\ \hat{S}'_2 &\longleftarrow R_1, R'_2, \dots, R_r \\ &\vdots \\ \hat{S}'_r &\longleftarrow R_1, R_2, \dots, R'_r \end{aligned} \qquad where\ m(R'_i) = m(R_i) + M_\Delta$$

For each generated analogue candidate, there is one monomer R_i with altered mass value to makes up for the mass difference M_Δ. This ensures the mass of every analogue candidate matches with the spectrum parent mass M. The process only involves mass calculation, and the program does not attempt to make structural interpretation for the modified monomer, as this is challenging to decipher using MS/MS alone. Therefore, the program does not need a monomer database, such as Norine, for generating analogues, providing an unbiased method for identifying substitutions and monomers.

Evaluation of Analogue Candidates

Having generated analogue candidates from selected input seeds, a scoring mechanism is needed to evaluate the significance of matching between a spectrum and an analogue candidate. In this work, we adopted the scoring system which had been validated in the original iSNAP database search algorithm. A match between a spectrum and analogue candidate is evaluated with three scoring metrics: raw score, P1 score, and P2 score.

Raw score is a basic spectral-matching score. The algorithm generates hypothetical spectral fragments of the analogue candidate using the original iSNAP platform. The mass-to-charge ratios of these hypothetical fragments are matched to the spectrum. The raw score is then calculated as the logarithmic sum of the peak intensity of all the matched peaks.

P1 score is a statistical normalization of the raw score. By scoring the spectrum with all compounds in the dereplication database, we acquire a raw score distribution of random matches. For an analogue candidate, a p-value is calculated by its raw score on this

distribution. As P1 score is subjected to the composition of the database, having similar structures in a database would slightly skew the distribution and affect the p-value. A P2 score was introduced to alleviate this issue. By shifting all peaks in the spectrum, a list of false spectra is created to match with the compound, generating a distribution for p-value calculation. The calculation of P2 score is independent of other database structures. Both P1 and P2 score are converted from p-values for better readability using the formula P1=-10log10(p-value). A higher P1 score indicates lower probability to have a random structure matching better with the spectrum. A higher P2 score indicates lower probability to have a random spectrum matching better with the analogue candidate.

For each spectrum, P1 score and P2 score are calculated for every analogue candidate. Analogue candidates are deemed significant if having both P1 score and P2 score higher than the specified thresholds. By default, the thresholds for P1 and P2 are set to 27 and 24, respectively, as empirically determined in our previous work. The thresholds can be adjusted by the user to give more flexibility in result filtering.

Elaborating the Tyrocidine Family

To test whether this new algorithm would be capable of accurate analogue detection we chose to analyze crude extracts of *Bacillus parabrevis*, which produces the tyrocidines – one of the most diverse and well annotated natural product complexes known. The tyrocidine family of cyclic decapeptides is comprised of 28 known structural variants (Tang et al., 1992) made up of amino acid substitutions within the peptide core (Figure 1B), with relative abundances ranging from ~1 to 100%. An examination of the chemical space of the tyrocidines (Figure 1C) reveals ***the*** single-monomer interconnectivities or 'network' of the tyrocidine scaffold, highlighting the correlations between the structures, in terms of monomer position and mass difference. As a first step, we optimized our LC conditions with a focus on the use of high-

efficiency core-shell columns. This proved useful in resolving many of the co-eluting peaks, allowed for greater confidence of low abundance compounds, and avoided excess artifact hits within the ion-trap (Figure S2). Following established fermentation and extraction conditions (Tang et al., 1992), LC-MS/MS analysis was performed followed by automated metabolite dereplication by iSNAP, to validate the number of tyrocidine compounds within the crude extract. Over a 37 minute interval, iSNAP dereplicated 21 of the 28 reported structures (see Figure 1C, Table S1). While several of the low abundance tyrocidines series were not detected, this is likely a result of minor discrepancies in culture conditions affecting metabolite production or analytes being below the intensity threshold of the automated MS/MS settings, as manual investigation also failed to reveal the missing, previously reported structures. Having identified 21 tyrocidines within the fermentation culture, we used these dereplicated spectra as references for evaluating analogue identification and testing the monomer localizations within the tyrocidine single-monomer substitution network (Figure 1C). The LC-MS/MS data file was re-screened 21 times with the iSNAP analogue search algorithm, using each dereplicated structure as the seed. For each seed structure processed, only the analogue identifications corresponding to the reference spectra are evaluated. By comparing the analogue identifications to the dereplicated structures, we can determine if a correct identification is made. As we had expected, the iSNAP analogue program correctly matched each MS/MS spectra with its corresponding analogue seed structure (see Figure 1C, Figure S3). As an example, using tyrocidine A (TC#16) as the seed, the analogue search made correct identifications on the MS/MS spectra of tyrocidines #10, 12, 17, 18, and 19, with mass differences and localization correctly identified; +39 m/z at monomer 7, +39 m/z at monomer 6, +14 m/z at monomer 2, +23 m/z at monomer 10, -16 m/z at monomer 10, respectively. By making use of tandem MS data, the iSNAP analogue program was capable of reliably distinguishing between isobaric

analogues; (1) TC# 10 and #12, difference in one site modification Phe_6-Trp_7 and $Tryp_6$-Phe_7 respectively (2) TC#18 (Phe_6-Phe_7,Trp_{10}) and TC-24 (Trp_6-Phe_7,Phe_{10}). These reliably accessed networks suggests that through iterative dereplication and analogue steps, it is possible to access all known tyrocidines following the dereplication of a single structure (Figure 1C, Figure S3). These examples demonstrate the utility of the iSNAP analogue program, having elaborated the tyrocidine family of cyclic peptides and correctly localized sites of variation within the scaffold.

In Silico Investigation of Analogue Identification and Localization Rate

We next sought to establish a controlled testing scenario in which we could evaluate, with confidence, the effectiveness of iSNAP's ability in making sensitive analogue identifications and also determine the accuracy in which the algorithm can localize structural differences. More specifically, the analog identification and localizations rates can provide a performance indicator or a measure for how the algorithm may perform when evaluating LC-MSMS data sets in a real analogue searching scenario.

In this experiment, we evaluate LC-MS/MS and MS/MS data sets using the iSNAP analogue search algorithm to first search and dereplicate any known NRP structures. We then artificially create, *in silico*, new analogue seed structures of the dereplicated knowns, by making alterations to the NRPs scaffold, one monomer site at a time. In this testing scenario, the iSNAP analogue search algorithm should be capable of correctly identifying all of the artificial "*in silico"* seed structures as being analogues to the known NRP. Secondly, the algorithm should also be capable of accurately localizing the position of the altered sites within the NRP scaffold. By evaluating a series of NRP structures from crude microbial extracts, we can realize with high confidence, iSNAP's potential as being a sensitive and accurate discovery tool within this controlled setting.

For this experiment, we first manually created a series of *in silico* seed structures for 17 of the dereplicated tyrocidines by substituting a new monomer block at each of the 10 residue positions. The selected tyrocidines represent those with low and high abundance (Figure 2, Figure S4, and S5), avoiding any bias towards those with higher quality MS/MS spectra. We next incorporated a range of monomers that are not present within the tyrocidine family scaffold. As the tyrocidine family scaffold is comprised of only 12 amino acid building blocks (Figure 1B), we selected five different amino acids as *in silico* substitutions for the artificial seed structures: glycine, alanine, threonine, histidine and a 5-Br-tryptophan, representing increasing mass-to-charge values. We selected the new monomers to avoid any bias towards common fragmentation losses or conserved sequences as well as any bias associated with small or large mass differences that might be more readily matched within the tandem MS spectra of a tyrocidines.

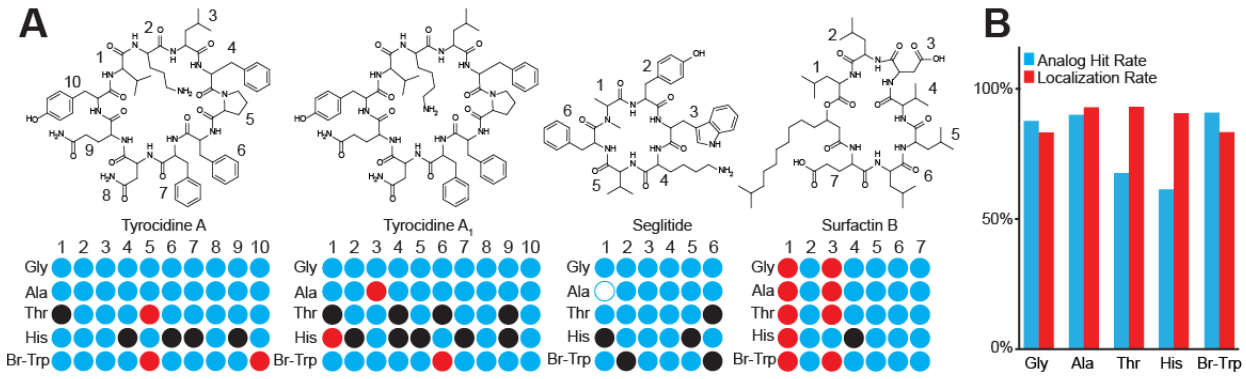


Figure 2. Analogue identification and localization.

(A) *In silico* candidate analogue screening with predefined monomer substitutions. To evaluate analogue identification and correct monomer localization rates, hypothetical seeds of cyclic peptides were substituted with glycine, alanine, threonine, histidine, or 5-Br-tryptophan at each position and used to identify their corresponding standard within LC-MS/MS data. Positively identified analogue candidates with correct localization are shown in cyan, while analogue candidates with incorrect localization are shown in red, and non-matching seed structures (unidentified) are shown in black. (B) Analogue identification and localization rates from *in silico* candidate screening of 17 dereplicated tyrocidines. A total of 17 tyrocidine compounds have been evaluated, representing candidates with relative abundances of ~1 to 100%. A total of 170 candidate seed structures are evaluated for each substituted monomer, representing a total of 850 analogue screens. The analogue hit rate represents the total number of analogue seed structures correctly matched to an MS/MS spectrum (localized and mislocalized), scored above the P1 and P2 cut-offs, and divided by the total number of matches possible. The localization rate is the total number of positive localizations divided by the total number of analogue matches.

For the 17 dereplicated tyrocidines, these newly created and artificial seed structures, represent a total of 850 *in silico* variants being evaluated, with 50 per structure (10 positions x 5 substitutions), and 170 variants globally per substitution (Figure 2, Figure S4, S5, Table S3). Using the constructed seed structures, we could then use the iSNAP analogue search algorithm and examine the analogue identifications made to the dereplicated tyrocidine spectra. From these results, analogue identification rates and monomer localization rates can be established, providing a diagnostic for the algorithm's performance. The analogue identification rate for the *in silico* variants is calculated by the total number of analogue seed structures correctly matched to an MS/MS spectrum (localized and mislocalized), scored above the P1 and P2 cut-offs, and divided by the total number of matches possible. The monomer localization rate is calculated as the total number of positive localizations divided by the total number of analogue matches. As we had expected, variations in the analogue identification rate could be seen across the substituted amino acids resulting in a global analogue identification rate of 79.41% (675/850) for the tyrocidine family (Figure 2, Figure S5, Table S2, Table S3). On an individual basis, the analogue identification rates are 87.65%, 90.00%, 67.65%, 61.18%, and 90.59% for glycine, alanine, threonine, histidine and 5-Br-tryptophan respectively. Rewardingly, of the analogues identified, the global monomer localization rate is over 88% (595/675), with rates of 83.22%, 92.81%, 93.04%, 90.38% and 83.12%, for glycine, alanine, threonine, histidine and 5-Br-tryptophan respectively. To expand on these initial findings, we further probed the molecular scaffolds of several other cyclic peptides (WS9326a, seglitide, surfactin B) to evaluate analogue identification rates and monomer localizations rates, using the five substituted monomers as *in silico* seeds. In the case of WS9326a, seglitide, and surfactin, analogue identification rates were ~73%, 87% and ~98%, with monomer localizations rates of ~59%, 100% and 87% respectively

(Figure 2, Figure S5). These findings demonstrate the effectiveness of the iSNAP analogue algorithm to identify analogue variants and localize structural differences with high accuracy.

Iterative Analogue Analysis Maps the Loloatins - Rare Tyrocidine-like Cyclic Peptides

Building on our tyrocidine network findings, we sought to identify small molecule products of a tyrocidine-like NRPS in the genome of the previously uninvestigated isolate *Brevibacillus laterosporus* (DSM 25). *B. laterosporus* was cultured, extracted, subjected to LC-MS/MS analysis, providing an .mzXML data file that could be analyzed by the iSNAP algorithm to identify peptide natural products. Rather than the well-studied tyrocidines, iSNAP LC-MS/MS analysis revealed the loloatins - related natural products which possess two constitutive substitutions that differentiate them from the tyrocidines (Gerard et al., 1999). These peptides possess considerably improved activity relative to the tyrocidines, against multidrug resistant Gram positive bacteria and Gram negative bacteria (Gerard et al., 1999). In contrast to the well-studied tyrocidines, only four loloatin structures have been previously reported, and of these, we only observed production of loloatin A. To identify novel variants, we screened the culture extract using the four known structures (loloatin A-D) as seeds for iterative analogue profiling. Following each round of analysis, analogue identifications were confirmed by manual MS/MS examination and annotation (Appendix 1), with correctly identified analogues forwarded as seeds for the next round of discovery. After four rounds of iterative iSNAP analogue search, we identified a total of 33 new loloatins (Figure 3), including 22 which have been identified as unique structures, along with 9 structures which are presumably isomeric analogues, and 2 structures that could not be confirmed by manual MS/MS annotation due to poor quality MS2 spectra. Of these 33 identified loloatins, 25 demonstrated correct monomer localizations (Appendix 1). Only 3 MS/MS scans were detected as false positives throughout the analysis, and with low P-scores (avg. FP P1/P2 of 24.5/23.4;

global P1/P2 of 61.4/30.1). By iterating our analogue approach, we successfully identified an entire family of loloatins comprised of 22 distinct, novel variants, revealing an extensive network similar to the tyrocidine family.

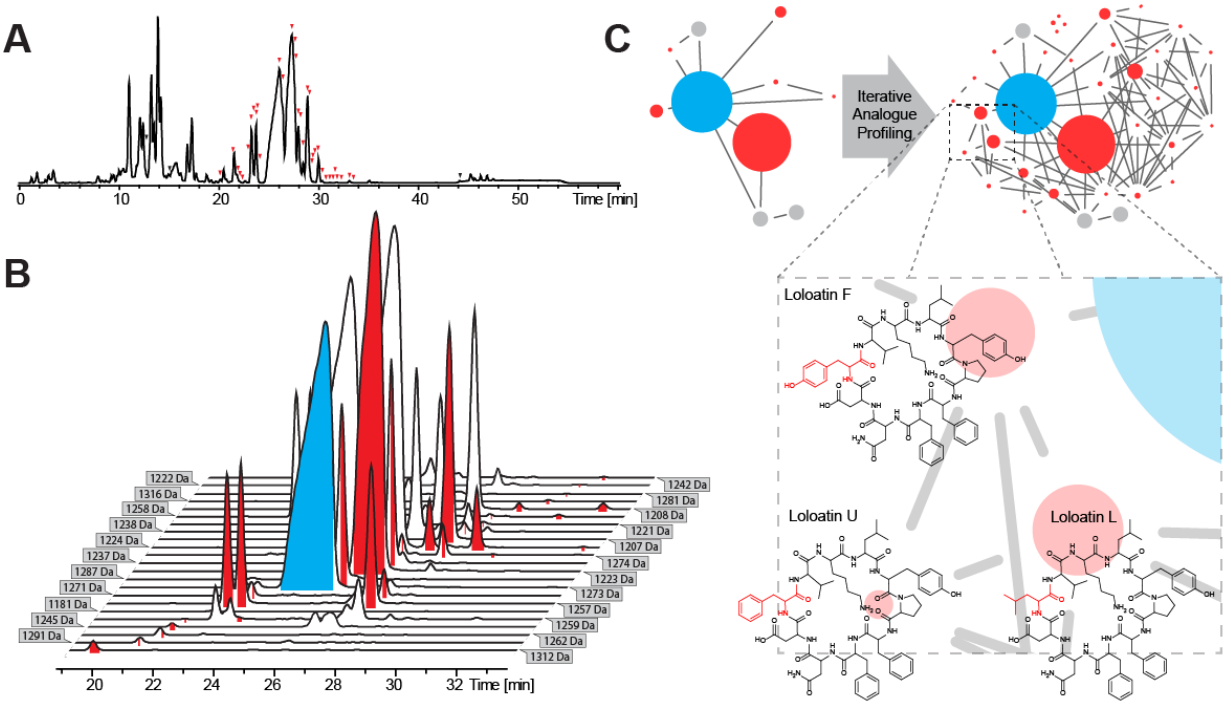


Figure 3. Iterative informatic exploration of loloatin chemical space.

(A) Informatic detection of a large family of loloatin natural products from a crude extract of *B. laterosporus* (DSM 25). Red arrowheads denote unique loloatins identified by iSNAP. Black arrowheads denote natural products that were falsely identified.(B) Extracted ion chromatograms of loloatin natural products. Each MS/MS scan determined by iSNAP to contain new loloatin structures is colored in red, while MS/MS scans containing known loloatins are colored in cyan.(C) Expansion of known loloatin chemical space through four rounds of iterative iSNAP analogue identification. Loloatin species are represented as nodes sized by abundance and connected by single-monomer substitutions. Known loloatins identified in this work are shown in cyan, while new loloatins identified in this work are shown in red. Previously discovered loloatins that were not observed in this work are shown in gray. The magnified section of the analogue network depicts structural alterations linking three new loloatins.

Automated Analoguing Reveals a Massive Family of Lipodepsipeptides

Next, we chose to examine a more modified peptide architecture, one incorporating acyl and ester moieties. We focused our efforts on the LI-F0 series of lipodepsipeptides from *Paenibacillus polymyxa*. Originally described in 1987 (Kurusu et al., 1987), these nonribosomal

hexapeptides possess a rare guanidinylated acyl tail, and demonstrate considerable variability in the incorporation of hydrophobic amino acids (Val/Ile and Val/Ile/Phe/Tyr) at positions 2 and 3. These substitutions, along with variable Asn/Gln incorporation at position 5, has led to the elucidation of 12 reported structures, though 16 distinct molecules are theoretically possible with modifications at these sites. These structures have not yet been discovered through earlier MS/MS works (Kuroda et al., 2001) or synthetic biology efforts (Han et al., 2012), likely a result of their low abundance. To assess whether the iSNAP analogue search algorithm could display sufficient sensitivity to identify these proposed analogues and reveal any unforeseen variants, we cultured 6 L of *Paenibacillus polymyxa* and extracted the supernatant for iSNAP analysis. As we had expected, each of the 12 known LI-F0 structures was reliably dereplicated, with retention times relatively similar to published works (Kurusu et al., 1987) and MS/MS fragmentation patterns consistent with the LI-F0 scaffold (Figure 4, Appendix 2). Surprisingly, iSNAP analogue search also revealed the presence of an entire suite of unreported LI-F0 variants (Figure 4). iSNAP analysis of the *Paenibacillus polymyxa* crude extract revealed 39 novel variants, detected over 42.4 minutes (Figure 4), making this – to our knowledge – the largest natural product complex ever discovered from a single organism, with over 50 structures identified (Figure 4, Appendix 2). During the analysis, no other secondary metabolites were detected, while 10 of 4339 scans (avg. FP P1/P2 of 24.8/23.4; global P1/P2 of 29.0/36.5) were false positives and attributed to artifacts, as revealed by manual inspection (Appendix 2). Further, seven compounds were correctly identified as LI-F0 antibiotics by iSNAP but were incorrectly annotated due to convoluted fragmentation or insufficient abundance; three could be assigned by manual MS/MS annotation. This complex series of structures appears to arise from several recurring variations, including amino acid substitutions at every position (such as Ser/Thr1, Val/Ile/Leu2, Val/Ile/Leu/Phe/Tyr3, Ser/Thr4, Asn/Gln/Glu5, Ala/Gly6),

linearization of the macrocycle ester, loss of the terminal alanine, or addition of a second C-terminal monomer (Ala or Gly). The four analogues which complete the original set arising from combinatorialization of monomers at positions 2, 3, and 5 were also observed, with preliminary quantification efforts demonstrating an isolatable yield of 1 μg/L. The obscurity of these variants is evident in both the LC-MS/MS chromatogram as well as the corresponding LI-F0 family network (Figure 4D), thus underscoring the analytical and discovery value of informatic search approaches in reveal previously undiscovered structures.

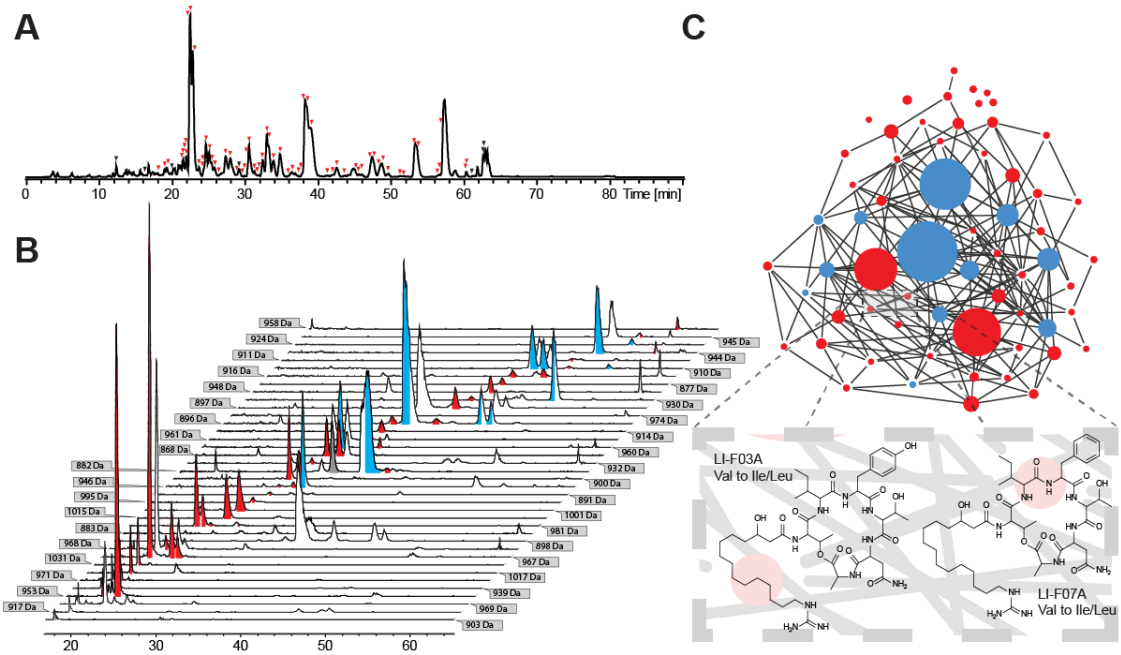


Figure 4. iSNAP-driven discovery of a massive family of LI-F0 series natural products.

(A) Informatic detection of a massive family of LI-F0 series lipodepsipeptides from a crude extract of *P. polymyxa* (ATCC no. 21830). Red arrowheads denote unique LI-F0 series products identified by iSNAP. Black arrowheads denote natural products that were falsely identified.(B) Extracted ion chromatograms of LI-F0 series molecules. Each MS/MS scan determined by iSNAP to contain new LI-F0 series structures is colored in red, while MS/MS scans containing known LI-F0 series molecules are colored in cyan. Scans containing molecules incorrectly identified as LI-F0 series compounds are shown in gray.(C) Expansion of known LI-F0 series chemical space through iSNAP analogue identification. LI-F0 species are represented as nodes sized by abundance and connected by single-monomer substitutions. Known LI-F0 series molecules identified in this work are shown in cyan, while new LI-F0 series molecules identified in this work are shown in red. The magnified section of the analogue network depicts structural alterations linking two new LI-F0 series structures.

Identification and Elucidation of a Minor Arylomycin Variant with Improved Bioactivity

Following our lipodepsipeptide screening, we next chose to investigate glycosylated natural products for new, minor variants that may yield superior bioactivity. We analyzed a series of extracts from environmental Actinomycetes to identify candidate natural products, and detected a suite of glycosylated arylomycins (Figure 5A) which had previously been reported by researchers at Eli Lily as potent inhibitors of type 1 signal peptidase with antibacterial and β-lactam-sensitizing activity (Kulanthaivel et al., 2004). iSNAP analogue analysis of these partially-cyclic, glycosylated lipopeptides identified a total of 6 novel variants, with alterations in acyl tail length and phenylglycine dihydroxylation. With these new findings, we then re-investigated our crude extracts for additional analogues using the novel structures as seeds. Prior to running the iSNAP analogue search algorithm, we first dereplicated the novel structures to further validate and confirm the annotations. Next, we expanded the analogue mass tolerance window to facilitate the discovery of more divergent analogues. During this second round of study, the iSNAP analogue search identified an additional 4 congeners, corresponding to aglycones of the original series (Schimana et al., 2002) which would could be confirmed by multistage MSn analysis (Figure 5, Appendix 3). Previous SAR studies on this antibacterial complex identified that decreasing acyl tail lengths led to increased activity against Gram negative bacteria such as *E. coli* (Kulanthaivel et al., 2004). To assess whether the shorter acyl tails observed in our novel minor variants led to improved activity, we isolated one of these structures (MW 1000 Da; isolated at ~17 μg/L) for bioactivity testing and structure elucidation alongside the previously described parent structure (MW 1014 Da; isolated at ~ 220 μg/L). High resolution mass measurement of the novel variant matched our expectation as glycosylated arylomycin having a shorter acyl tail ($[M+H]^+$: $C_{49}H_{73}N_6O_{16}$; 1.198 ppm error; Appendix 4). However, comparison of the [^{1}H, ^{13}C]-HMBC, [^{1}H, ^{1}H]-COSY and [^{1}H, ^{1}H]-

TOCSY experiments between the 1014 Da parent and the 1000 Da analogue revealed differences in the N-methylated N-terminal amino acid, which is not readily seen within the tandem MS spectra. In contrast to the N-methyl serine observed in the parent structure, the minor novel variant possesses both an N-methylated threonine and shorter acyl tail (C12 vs. C14). With the exception of the chemical shifts associated with the substituted amino acid, our findings are consistent with the fully resolved and elucidated parent structure (Appendix 4). To assess whether this minor variant possessed superior activity to the more abundant parent compound, we selected a sensitive *E. coli* test strain, similar to that used originally by researchers at Eli Lily (Kulanthaivel et al., 2004). Microdilution MIC assays of our novel arylomycin variants against *E. coli* revealed that – consistent with previous findings (Kulanthaivel et al., 2004) – our short-tail variant possessed modestly greater activity as compared to the parent compound, with MIC values of 0.6 ug/mL to 0.4 ug/mL respectively.

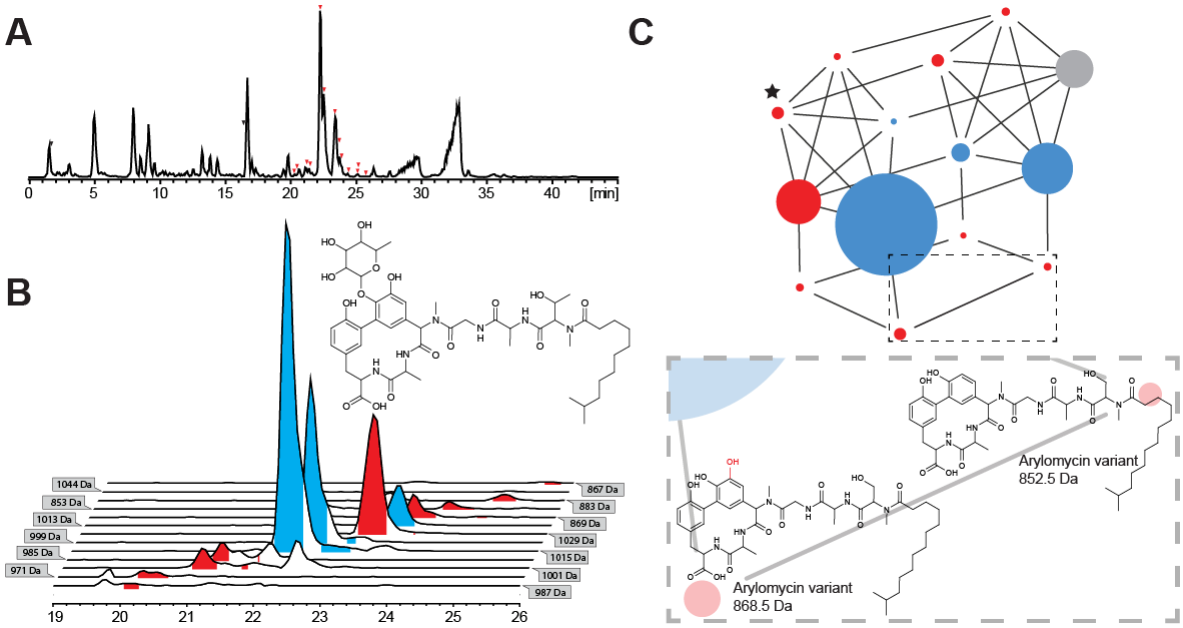


Figure 5. Expansion of glycosylated arylomycin chemical space facilitates the discovery of an analogue with improved bioactivity.

(A) Informatic detection of a family of arylomycin natural products from a crude extract of environmental Actinomycete isolate NAM12. Red arrowheads denote unique arylomycins identified by iSNAP. Black arrowheads denote natural products that were falsely identified.(B) Extracted ion chromatograms of arylomycins. MS/MS scans determined by iSNAP to contain new arylomycin structures based on the dereplicated glycosylated arylomycin scaffold are shown in red, and scans containing known glycosylated arylomycins are shown in cyan. A new analogue (1001 Da [M+H]; *inset*) was isolated and its structure was elucidated by HRMS, 1D, and 2D NMR experiments.(C) Expansion of known arylomycin chemical space through two rounds of iterative iSNAP analogue identification. Arylomycin species are represented as nodes sized by abundance and connected by single-monomer substitutions. Known glycosylated arylomycins identified in this work are shown in cyan, while previously unobserved glycosylated arylomycins and corresponding aglycones identified in this work are shown in red. Previously discovered glycosylated arylomycins that were not observed in this work are shown in gray. The new, more active arylomycin variant is denoted with a star. The magnified section of the analogue network depicts structural alterations linking two arylomycin aglycones.

DISCUSSION

During the golden age of natural product antibiotic discovery, researchers identified a staggering number of unique chemical scaffolds, including several which have endured as modern therapeutic agents (Newman and Cragg, 2007). However, the vast majority of these bioactive molecules were discarded, as apparent chemical diversity provided more efficacious or easily-developed leads, such as glycopeptides and macrolide antibiotics. Daptomycin –

originally LY 146032 – was one such discarded scaffold that went through years of genomic and metabolomic engineering to emerge as an immeasurably useful antibiotic (Eisenstein et al., 2010). As resistance to conventional antibiotic scaffolds continues to rise, reinvestigation of over-looked natural products may prove to be a valuable method for meeting clinical demand.

Microbial natural products are thought to be exceptionally bioactive because they possess evolved chemical scaffolds that provide fitness benefits to the producer organism within their native environment. While these evolved molecules provide excellent leads for drug development, properties that are useful in a natural environment are often at odds with our desires for their use as pharmaceutical agents. Given this disparity, there is not necessarily a direct correlation between the abundance of a given congener and its activity towards a pathogen of interest (He et al., 2002; Balkovec et al., 2014; Lin et al., 2012; Gerard et al., 1999; Kurusu et al., 1987). Given the intrinsic promiscuity of many biosynthetic enzymes and the resultant chemical diversity observed in natural product extracts, metabolomic investigations of undeveloped scaffolds can lead to the discovery of new variants with improved pharmaceutical potential. Fast and reliable automated methods are now sorely needed to mine large extract libraries and expand the chemical space of clinically promising pharmacophores.

LC-MS/MS is currently the benchmark standard for rapidly profiling complex natural product extracts, and with the use of tandem MS, vast quantities of rich data can be mined to identify chemical entities. During the past decade a number of methodologies have been developed that make use of *de novo* MS/MS sequencing techniques (Ng et al., 2009; Medema et al., 2014) and genomic data (Kersten et al., 2011; Mohimani et al., 2014), in efforts to guide the discovery of ribosomal and nonribosomal peptides. In the case of iSNAP, by making use of a comprehensive library of peptide natural product structures, we can utilize our validated statistical algorithms (Ibrahim et al., 2012) to expand the known chemical space of key natural

products identified in an untargeted and automated manner. Because the iSNAP analogue program is supplied as a user-friendly web application, it provides a more straightforward and directed means of identifying the structures and locations of families of peptide natural products using LC-MS/MS data from crude microbial extracts. This pure metabolomics approach capitalizes on the extensive knowledgebase of natural product chemistry and is not necessarily tied to genomic data, which is not always available for extant microbial extract libraries.

We have demonstrated that the iSNAP analogue function can provide consistent and accurate results using low-resolution LC-MS/MS data of crude extracts containing low concentrations of natural products. This is in contrast to previously published *de novo* sequencing approaches that required pure, cyclic peptide standards that were directly infused for high-resolution multistage MS^n analysis (Ng et al., 2009). While *de novo* sequencing represented a pioneering achievement compared to earlier works, the challenges of this technique remain, including issues with mixtures of "direct sequence ions" (DS) and "non-direct sequence ions" (NDS) in cyclic peptides, where multiple ring opening events can occur and complicate sequencing. While early programs could be designed for analysis of pure standards of cyclic peptides (Ng et al., 2009; Mohimani et al., 2011), they were not applicable to complex mixtures containing other architectures of nonribosomal peptides. In addition, while foundational early work demonstrated effective use of informatic search strategies to correctly identify structures from large databases (Mohimani et al., 2011) peptide standards were typically pretreated with chemical reducing agents to increase fragmentation, an approach which is not amenable to metabolomics studies of natural product extracts.

One important application of the iSNAP analogue algorithm is the rapid analysis of peptide natural product mixtures and identification of related molecular families. Visualization

of natural product families can also be performed using molecular networking approaches (Watrous et al., 2012; Nguyen et al., 2013) that cluster observed ions based on similarities in their mass spectral fragmentation patterns. Fragment commonalities are scored using cosine vectors (Frank et al., 2008) that ease the clustering of related spectra, but do not provide a means to identify individual compounds with high specificity or without extraneous analysis. Unlike molecular networking approaches, the iSNAP analogue function provides results with retention times and confidence scores for individual scans, and indicates peptide identities without the need for genetic knockouts (Watrous et al., 2012) or direct comparison with known compounds (Yang et al., 2013) or known producers (Nguyen et al., 2013). In contrast to molecular networking approaches, iSNAP also does not merge or discard seemingly identical spectra (Watrous et al., 2012; Frank et al., 2008), and thus allows for the detection of distinct isobaric species at different retention times, allowing users to take full advantage of optimized chromatography. The iSNAP analogue program also does not require manual annotation of MS/MS data to assign amino acids (Kersten et al., 2011; Medema et al., 2014) or monomer-modifications (Watrous et al., 2012), and has been designed to handle proteinogenic and non-proteinogenic amino acids, allowing it to function as an information-rich and rapid means of detecting families of nonribosomal peptides. These differences allow the iSNAP analogue program to define important information about specific peptides – as well as general trends about peptide families – with superior speed and accuracy to previously published methods. Despite this, there remain limitations to the iSNAP analogue algorithm and mass spectrometry-based approaches in general. First, although mass spectrometry can provide a wealth of information about the nature and identity of molecules or monomers units, more comprehensive spectroscopic techniques like NMR are still required to define exact structural features. Second, sensitivity of LCMS-based detection is often tied to the ability of a given molecule to be

ionized, meaning that some molecules are observed more easily than others. We have conclusively demonstrated that the iSNAP analogue algorithm can map peptide families, but the identification of new analogues is currently limited by alterations to a single monomer site, so low abundance analogues with multiple independent substitutions remain challenging to correctly detect and define. In regard to the scoring scheme, P1 and P2 scores served their purpose as quality indicators for individual matches, but are not corrected for multiple testing. The algorithm can be improved if it also reports an estimated false discovery rate for identification results. In addition, this approach is currently limited to peptidic natural products, which possess reliable and predictable potential fragmentation patterns, providing sufficient data for the highly effective analogue detection and elucidation presented here.

In this work, we investigated the tyrocidine family of cyclic peptides and demonstrated that the iSNAP analogue search algorithm correctly identified all analogue compounds for each dereplicated tyrocidine. Through an in silico screening approach, we have shown that our discovery method is highly sensitive in analogue identification, while highly accurate in site-specific monomer localization. We have also demonstrated iSNAP's discovery potential by identifying over 125 peptide structures, of which 70 are novel variants, from a series of cyclic, lipo-, and glyco-peptides. While the extensive genetic diversity observed in microbial life offers a glimpse of new, promising leads for natural products discovery, we have demonstrated here that the chemical diversity of lone biosynthetic assembly lines can provide new, improved variations of desired scaffolds that parallel this boundless potential.

SIGNIFICANCE

In this work, we present the iSNAP analogue method as an effective strategy for the untargeted discovery of new nonribosomal peptides from crude microbial extracts. This

approach does not require the aid of genomic sequence information, high-resolution MS systems, or prior knowledge of the sample for positive analogue identifications. iSNAP's automated analogue processes have been applied to several potent antimicrobial producers, leading to the discovery of over 70 novel unreported peptide variants including one with improved potency; all without the use of bioassay guided isolation. HRMS, multistage MSn analysis, and 1D/2D NMR measurements of isolated variants further establishes iSNAP's analogue capabilities as a true discovery tool.

EXPERIMENTAL PROCEDURES

General

LC-MS/MS data acquisition was obtained on a Bruker Amazon-X Ion-trap mass spectrometer coupled to a Dionex Ultimate 3000 HPLC running under Hystar 3.2 control with Trap control 7.0 and Chromeleon 6.2. Spectra were generated using electrospray ionization and under collision-induced dissociation (N2 nebulizer gas: 25psi, He dry gas: 7.5 psi, temperature: 250C). Automated MSn acquisitions were performed from 450-1600m/z for tyrocidines, across a scan range of 100-2000 m/z using the Enhanced Resolution setting. MS/MS analysis parameters included: isolation width at n=4, precursor ions at n=10, threshold cut-off from 400-600,000, active exclusion at n=4 spectra over 20 sec., and CID fragmentation set to 1.25 V across a voltage sweep of 25-200%. Bruker raw data files were converted to .mzXML format using Bruker conversion software, CompassXport prior to iSNAP Analogue analysis (http://www.ionsource.com/functional_reviews/CompassXport/CompassXport.htm). For analytical flow rates a UV/MS flow splitter of 10:1 was used. LCMS spectral analysis was performed using Compass DataAnalysis 4.1 (Bruker). High resolution mass spectrometry (HRMS) measurements were performed in positive electrospray ionization using a Bruker maXis 4G UHR-TOF mass spectrometer coupled to a Dionex Ultimate 3000 HPLC system

running Hystar 3.2 control and under standardized LC conditions, with calibrations done using sodium formate. 1D and 2D NMR measurements were acquired using a Bruker Avance III 700 MHz NMR spectrometer equipped with a 5mm QNP cryoprobe, operating at 700.17 MHz for ^{1}H NMR and 176.08 MHz for ^{13}C NMR respectively. Chemical shifts were referenced to the internal solvent peaks: 3.31 ppm (^{1}H) and 49.00 ppm (^{13}C) for CD_3CD. Surfactin (S3523) and seglitide (S-1316) standards were purchased from Sigma Aldrich (USA).

Microbial Strains

Brevibacillus laterosporus and *Bacillus parabrevis* were obtained from the German Resource Centre for Biological Material (DSMZ, DSM no. 25 and 362 respectively). *Paenibacillus polymyxa* were obtained from the American Type Culture Collection (ATCC, ATCC no. 21830). Environmental Actinomycete NAM12 was isolated from an environmental soil sample collection performed at McMaster University. *B. laterosporus*, *B. parabrevis*, and *P. polymyxa* were maintained on LB agar plates at 30°C. NAM12 was maintained on Bennett's agar plates at 30°C.

Cytoscape depiction of tyrocidine, loloatin, LI-F0, and arylomycin chemical space

Cytoscape plots were assembled using Cytoscape 3.0.1, imported from a manually curated network file of for each peptide, connected through single amino acid substitutions. Nodes were automatically distributed by preset preferred layout and manually adjusted to account for subsequent resizing. An exported .pdf file was manipulated using Adobe Illustrator CS6, resizing nodes according to their relative abundance, and in the case of the tyrocidines, by relative size.

In silico Scanning Analysis

Compound structures were modified in ChemBioDraw Ultra 13.0 to substitute a a glycine, alanine, threonine, histidine and Br-Tryptophan at each of the ten monomer positions for the tyrocidines, six positions for seglitide, seven for surfactin B, and 6 for WS9326a. The structures were converted to SMILES codes and compiled into an uploadable text file for analysis. The LC-MS/MS chromatograms of the tyrocidines, WS9326a, seglitide and surfactin were then analyzed by iSNAP analogue using each of the monomer substituted analogue candidates as individually seeds via the uploaded text files. Reports were analyzed manually to identify whether an analogue candidate of the monomer-substituted structures was matched to the MS/MS scans containing the appropriate peptide structure.

Additional experimental procedures can be found with the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information, including additional experimental procedures, four figures, three tables, and four appendixes, can be found with this article online.

DOI: http://dx.doi.org/10.1016/j.chembiol.2015.08.008

REFERENCES

1. Balkovec, J.M., Hughes, D.L., Masurekar, P.S., Sable, C.A., Schwartz, R.E., and Singh, S.B. (2014). Discovery and development of first in class antifungal caspofungin (CANCIDAS®) — a case study. Nat. Prod. Rep. *15*, 15-34.

2. Chou, T.C., Zhang, X.G., Balog, A., Su, D.S., Meng, D., Savin, K., Bertino, J.R., and Danishefsky, S.J. (1998). Desoxyepothilone B: an efficacious microtubule-targeted antitumor agent with a promising in vivo profile relative to epothilone B. Proc. Natl. Acad. Sci. USA *95*, 9642-9647.

3. Eisenstein, B.I., Oleson, F.B.Jr., and Baltz, R.H. (2010). Daptomycin: From the Mountain to the Clinic, with Essential Help from Francis Tally, MD. Clin. Infect. Dis. *50*, S10-S15.

4. Frank, A.M. Bandeira, N., Shen, Z., Tanner, S., Briggs, S.P., Smith, R.D., and Pevzner, P.A. (2008). Clustering millions of tandem mass spectra. J. Proteome. Res. *7*, 113-122.

5. Gerard, J.M., Haden, P., Kelly, M.T., and Andersen, R.J. (1999). Loloatins A-D, cyclic decapeptide antibiotics produced in culture by a tropical marine bacterium. J. Nat. Prod. *62*, 80-85.

6. Han, J.W., Kim, E.Y., Lee, J.M., Kim, Y.S., Bang, E., and Kim, B.S. (2012). Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. Biotechnol. Lett. *34*, 1327-1334.

7. He, H., Williamson, R.T., Shen, B., Graziani, E.I., Yang, H.Y., Sakya, S.M., Petersen, P.J., and Carter, G.T. (2002). Mannopeptimycins, novel antibacterial glycopeptides from *Streptomyces hygroscopicus*, LL-AC98. J. Am. Chem. Soc. *124*, 9729–9736.

8. Hou, Y., Braun, D.R., Michel, C.R., Klassen, J.L., Adnani, N., Wyche, T.P., and Bugni, T.S. (2012). Microbial strain prioritization using metabolomics tools for the discovery of natural products. Anal. Chem. *84*, 4277-4283.

9. Ibrahim, A., Yang, L., Johnston, C.W., Liu, X., Ma, B., and Magarvey, N.A. (2012). Dereplicating nonribosomal peptides using an informatic search algorithm for natural products. Proc. Natl. Acad. Sci. USA. *109*, 19196-19201.

10. Kersten, R.D., Yang, Y.L., Xu Y., Cimermancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., and Dorrestein, P.C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. Nat. Chem. Biol. *7*, 794-802.

11. Kulanthaivel, P., Kreuzman, A.J., Strege, M.A., Belvo, M.D., Smitka, T.A., Clemens, M., Swartling, J.R., Minton, K.L., Zheng, F., Angleton, E.L., Mullen, D., Jungheim, L.N.,

Klimkowski, V.J., Nicas, T.I., Thompson, R.C., and Peng, S.B. (2004). Novel lipoglycopeptides as inhibitors of bacterial signal peptidase I. J. Biol. Chem. *279*, 36250-36258.

12. Kuroda, J., Fukai, T., and Nomura, T. (2001). Collision-induced dissociation of ring-opened cyclic depsipeptides with a guanidino group by electrospray ionization/ion trap mass spectrometry. J. Mass. Spectrom. *36*, 30-37.

13. Kurusu, K., Ohba, K., Arai, T., and Fukushima, K. (1987). New peptide antibiotics LI-F03, F04, F05, F07, and F08, produced by Bacillus polymyxa. I. Isolation and characterization. J. Antibiot. *40*, 1506-1514.

14. Lin, Z., Falkinham, J.O., Tawfik, K.A., Jeffs, P., Bray, B., Dubay, G., Cox, J.E., and Schmidt, E.W. (2012). Burkholdines from *Burkholderia ambifaria*: antifungal agents and possible virulence factors. J. Nat. Prod. *75*, 1518-1523.

15. Medema, M.H., Paalvast, Y., Nguyen, D.D., Melnik, A., Dorrestein, P.C., Takano, E., and Breitling, R. (2014). Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. PLoS Comput. Biol. *10*, e1003822.

16. Mohimani, H., Liu, W.T., Mylne, J.S., Poth, A.G., Colgrave, M.L., Tran, D., Selsted, M.E., Dorrestein, P.C., and Pevzner, P.A. (2011). Cycloquest: Identification of cyclopeptides via database search of their mass spectra against genome databases. J. Proteome. Res. *10*, 4505-4512.

17. Mohimani, H., Kersten, R.D., Liu, W.T., Wang, M., Purvine, S.O., Wu, S., Brewer, H.M., Pasa-Tolic, L., Bandeira, N., Moore, B.S., Pevzner, P.A., and Dorrestein, P.C. (2014). Automated genome mining of ribosomal peptide natural products. ACS Chem. Biol. *9*, 1545-1551.

18. Mohimani, H., Liu, W.T., Kersten, R.D., Moore, B.S., Dorrestein, P.C., and Pevzner, P.A. (2014). NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery. J. Nat. Prod. *77*, 1902-1909.

19. Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of drugs over the last 25 years. J. Nat. Prod. *70*, 461-477.

20. Ng, J., Bandeira, N., Liu, W.T., Ghassemian, M., Simmons, T.L., Gerwick, W.H., Linington, R., Dorrestein, P.C., and Pevzner, P.A. (2009). Dereplication and de novo sequencing of nonribosomal peptides. Nat. Methods. *6*, 596-599.

21. Nguyen, D.D., Wu, C.H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X., Gavilan, R.G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J.D., Phelan, V.V., van de Wiel, C., Kersten, R.D., Mehnaz, S., De Mot, R., Shank, E.A., Charusanti, P., Nagarajan, H., Duggan, B.M., Moore, B.S., Bandeira, N., Palsson, B.Ø., Pogliano, K., Gutiérrez, M.,

and Dorrestein, P.C. (2013). MS/MS networking guided analysis of molecule and gene cluster families. Proc. Natl. Acad. Sci. USA. *110*, E2611-E2620.

22. Scherlach, K., Partida-Martinez, L.P., Dahse, H.M., and Hertweck, C. (2006). Antimitotic rhizoxin derivatives from a cultured bacterial endosymbiont of the rice pathogenic fungus *Rhizopus microsporus*. J. Am. Chem. Soc. *128*, 11529-11536.

23. Schimana, J., Gebhardt, K., Höltzel, A., Schmid, D.G., Süssmuth, R., Müller, J., Pukall, R., and Fiedler, H.P. (2002). Arylomycins A and B, new biaryl-bridged lipopeptide antibiotics produced by *Streptomyces* sp. Tü 6075. I. Taxonomy, fermentation, isolation and biological activities. J. Antibiot. *55*, 565-570.

24. Tang, X.J., Thibault, P., and Boyd, R.K. (1992). Characterisation of the tyrocidine and gramicidin fractions of the tyrothricin complex from *Bacillus brevis* using liquid chromatography and mass spectrometry. Int. J. Mass. Spectrom. Ion. Processes. *122*, 153-179.

25. Wang, C., Henkes, L.M., Doughty, L.B., He, M., Wang, D., Meyer-Almes, F.J., and Cheng, Y.Q. (2011). Thailandepsins: bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. J. Nat. Prod. *74*, 2031-2038.

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

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

28. Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., Wong, W.R., Linington, R.G., Zhang, L., Debonsi, H.M., Gerwick, W.H., and Dorrestein, P.C. (2013). Molecular networking as a dereplication strategy. J. Nat. Prod. *76*, 1686-1699.

29. Yu, Z., Vodanovic-Jankovic, S., Kron, M., and Shen, B. (2012). New WS9326A congeners from *Streptomyces* sp. 9078 Inhibiting *Brugia malayi* asparaginyl-tRNA synthetase. Org. Lett. *14*, 4946-4949.

CHAPTER FOUR

DISCOVERY OF NOVEL ANTIMICROBIALS FROM FUNGAL ENDOPHYTES BY COMPREHENSIVE SECONDARY METABOLOMICS

CHAPTER FOUR PREFACE

Discovery of Novel Antimicrobials from Fungal Endophytes by Comprehensive Secondary Metabolomics

Ashraf Ibrahim, Fan Fei, Tim McDowell, Joey Tanney, Linda Ejim, Dan Sørensen, Keith A. Seifert, Alfredo Capretta, J. David Miller, Brian E. McCarry, Mark W. Sumarah.

The work presented within this chapter is unpublished.

Author Contributions

B.E.M., A.I. and M.W.S., conceived the study.

B.E.M., A.I., M.W.S., J.D.M. and D.S., contributed to the study design.

A.I performed LC-MS analysis, created the in-house dereplication library of LC-MS spectral data, performed metabolomic analysis and analyzed data, performed HPLC and LC-SPE isolation and purifications, compound derivatization, performed structural elucidation and characterization (1D and 2D NMR, HRMS, crystallization) of new and known compounds, and bioactivity screening of crude extracts. F.F performed metabolomic analysis, analyzed data, and discussed results. T.D. cultured endophytes, performed fermentations and extracted the cultures. J. T. and K.A.S. performed ITS DNA sequencing and fungal species identifications. L.E. performed antimicrobial testing of purified compounds. D.S. provided NMR expertise, performed structural validation, and discussed results. J.D.M provided scientific expertise on fungal endophytes, and discussed results. A.I wrote the chapter and A.C. provided NMR expertise, structure elucidation/validation, discussed results and edited the chapter.

INTRODUCTION

Natural products from terrestrial bacteria and fungi have been widely studied since the early 20th century and have shown to be an excellent source for a diverse range of bioactive metabolites with applications in modern medicine, industry and agriculture.[1–4] While these past successes have been fruitful, many of these early discoveries are based on traditional chemical or bioactivity-guided isolations efforts that are inherently biased towards the most abundant or active species. As such, traditional screening approaches have often resulted in the re-isolation and characterization of known bioactive agents. The dereplication of known substances has hindered modern screening campaigns and is a leading road block to novel discoveries.[5–8] Secondary metabolomic based analytical methodologies encompassing LC-HRMS/MS analysis and multivariate statistical methods may offer new opportunities for novel discovery when applied to screening large, diverse collections of underexplored microbes.

In recent years, microbial discovery efforts have focused on understudied and niche ecological environments to identify new sources of novel taxa and to investigate their specialized metabolites for potential utility.[9–12] As an example, fungal needle endophytes from economically important confer tees have been investigated to combat serious fungal diseases and insect herbivory from the white pint blister and spruce budworm.[13–17] Fungal endophytes from white (*Picea glauca*), red (*Picea rubens*), black (*Picea mariana*) spruce and white pine (*Pinus strobus*) trees found in the Acadian Forest of Nova Scotia and New Brunswick, Canada, were screened for biological activity. These early studies had shown that fungal endophytes from conifer trees are capable of producing previously undescribed antifungal and anti-insectant metabolites that were toxic to the eastern spruce budworm larvae (*Choristoneura fumiferana*). It was also demonstrated that conifer trees infected with these fungal endophytes,

though horizontal transmission of new seedlings, reduced the growth rate of developing larvae, exposing them to greater predation times. Similarly, the potent FDA approved drug griseofulvin was isolated from an endophytic Xylaria species common to wild lowbush blueberries and adjacent white pine (*Pinus strobus*).[18]

There is currently limited research on the fungal endophytes from agriculturally important fruit crops (blueberry, cranberry, grapevine, and raspberry) and their specialized metabolites (Table 1). Recent studies have shown that new diplosporin and agistatine derivatives were produced from a *Xylari*a species isolated from concord grapevines (*Vitis labrusca*) located in Canada's Niagara region.[19] Similarly, new nemanifuranone and nemanilactone polyketides were reported from *Nemania serpens*, isolated from Riesling grapevines.[20] The antifungal compound Trienylfuranol A and related trienylfuranone A-B derivatives were also reported from *Hypoxylon submonticulosum*, an endophyte isolated from a raspberry leaf in Ontario, Canada.[21] These preliminary studies are key indicators of the potential chemical and biological diversity that fungal endophytes from fruiting plants may contain.

Canadian Crop	Yearly Production (tonnes)	Crop Distribution	Year Reported
Lowbush Blueberry	96,527	Quebec ~ 43%	2014
Highbush Blueberry	68,830	British Colombia ~ 89%	2014
Cranberry	79,163	British Colombia ~ 54%	2007
Grape	80,561	Ontario ~ 60%	2013
Raspberry	9,691	British Colombia ~ 58%	2013

Table 1. Agriculture and Agri-Food Canada crop profile data.[22–26]

In this study, a comprehensive LC-HRMS based metabolomics-guided discovery approach is used to investigate the secondary metabolome of 184 fungal endophyte strains isolated from underexplored Canadian *spermatopsida* (seed-bearing plants). This screening approach applies principle component analysis (PCA) to investigating the chemical space of a large and diverse collection of endophytic fungal strains comprised of over 30 identified species, including 33 isolates of unidentifiable origin, based on nucleotide BLAST screening of ITS DNA sequences. Pioneering works by Krug *et al.*, (2008) had demonstrated the utility of a LC-HRMS metabolomic approach to the investigation of the secondary metabolite profiles of 98 *Myxococcus xanthus* bacterial strains isolated from 78 locations worldwide.[27] They showed that the interspecies secondary metabolite diversity was greater than previously thought, based on genomic in insights, and identified over 37 putative novel metabolites by LC-HRMS. Building on these efforts, works by Hou *et al.*, (2012) applied LC-MS analysis and principle component analysis for a proof of concept study in prioritizing 47 bacterial strains (32 Verrucosispora sp., 5 Micromononospora sp., 10 Nocardia sp.), for novel natural products discovery.[28] This initial screen was effective in distinguishing the strains by their secondary metabolite profiles and they identified one putative new metabolite and several recently discovered natural products. Importantly, they investigated marine invertebrate associated bacteria in subsets of pre-sorted grouping of 20-50 strains based on gross morphology and in a follow-up study discovered bottromycin D, an alanine substituted derivative of the bottromycin A2, an MRSA active antibiotic.[29]

This metabolomic study distinguishes itself from these earlier works by investigating a significantly larger and more diverse microbial collection in search for novel natural products. A total of over 30 identified fungal endophyte species and 30 undescribed isolates have been

sourced from 5 types of fruiting plants and from 7 locations in Canada (3 Ontario, 4 Nova Scotia). While earlier studies on bacterial derived extracts have found some limitations in efficiently processing several hundred or more strains by multivariate analysis, we found pre-processing LC-HRMS data using CAMERA and XCMS for filtering and alignment, in conjunction with SIMCA-P proved fruitful.[30,31] Variances in data analysis may be attributed to individual isolates, extract complexity, extraction methods (solvent/resin based), pre-processing filters and molecular feature alignment processes (CAMERA, XCMS, MZMine) and differing commercial and/or free-ware processing suites for multivariate analysis (Bruker Profile Analysis 2.0, Sirius version 9.0, MetaboAnalyst and SIMCA-P+), including applied normalization or scaling parameters (pareto). Multivariate statistical analysis provides an efficient tool for mining large LC-MS data sets to rapidly identify metabolite variances between the fungal isolates.[32] However, by using an unsupervised PCA screening approach, we are assuming that ubiquitous or common fungal metabolites, found amongst these diverse species, should cluster or group together while truly unique metabolite entities will show a greater separation and variance relative to the common. These unique metabolite outliers can then be readily dereplicated using: (1) an in-house spectral library of deconvoluted LC-MS and MS/MS spectra; (2) commercial databases such as Antibase and Dictionary of Natural Products; and (3) iSNAP's automated chemo-informatic algorithms for processing LC-MS/MS and MS/MS data in .mzXML format.[7,8] Key outlier metabolites can then be targeted for isolation and characterization by LC-SPE-NMR and semi-preparative HPLC; where as in traditional bioactivity or chemically-guided screening approaches these could be missed or obscured, especially if in low abundance.

This study reports the discovery of twelve new specialized metabolites including >30 known compounds from several outlier extracts as well as clustered species groupings and summarizes the results. Compounds (**1**-**6**) are a previously undescribed series of unusual bicyclic/tricyclic-spiro non-prenylated polyketide acids with some similarity to plant derived acylphloroglucinols and coleophomone A-B, novel bacterial transglycosylase inhibitors discovered by Merck research laboratories.[33] Griseofernaneoside (**7**) is a previously undescribed and rare pentacyclic fernane-type triterpene with an unusual bridged bifuran and methoxy-glucose moiety. Fernane-type trierpenes have been exclusively reported from ferns and isolated from several plants.[34] To our knowledge, Kolokoside B, is the only reported fernane-type triterpene that has been isolated from a microbial source.[35] Compound (**8**) is a previously undescribed methoxy derivative of the terpenoid tricycloalternarene (TCA) **9b**.[36] Compound (**9**) is a previously undescribed polyketide analogue of folic acid. Compound (**10**) is a previously undescribed derivative of the antibiotic canescin A.[37] Compound (**11**) is a previously undescribed 2-(1-hydroxypropan-2-yl) hexanoic acid. Compound (**12**) is a new natural product and a derivative of guignardic acid.[38] The isolated compounds were tested for antimicrobial activity against a panel of pathogenic bacteria and several yeasts.

RESULTS

Endophyte Isolation, Fermentation and Species Identification

The fungal endophyte collection was sourced from the leaves, twigs, and stems of flowering fruit plants located in Ontario and Nova Scotia Canada. These included wild and cultivated grapes (Riesling and Concord varieties), raspberries, high and low bush blueberries and cranberries. In total, 263 fungal strains were isolated and initially grown on solid microbial agar media containing malt extract (30 g/L). The inoculated agar plates were incubated at 28 °C for 6-8 weeks, depending on the presence of filamentous hypha. The endophytic fungi growths are then sorted manually by visible phenotype and phylogeny prior to culturing. From the 263 isolates, only 184 strains (~70%) were meaningfully cultivable in liquid media containing potato dextrose and grown under standardized conditions. Cell-free culture filtrates were extracted with ethyl acetate solvent, facilitating processing efforts and limiting water soluble media components, and dried under reduced pressure by rotary evaporation. The dried extracts were re-suspended in deuterated acetonitrile for NMR fingerprinting and LC-MS analysis.

Canadian Crop	Location	Province
Lowbush Blueberry	Debert, Mt. Thom, Portapique, Rawdon	Nova Scotia
Highbush Blueberry	Simeco Rawdon	Ontario Nova Scotia
Cranberry	Bala, Muskoka lake	Ontario
Grape	Jordan Research Farm	Ontario
Raspberry	Jordan Research Farm	Ontario

Table 2. Sampling locations.

ITS DNA sequencing of the cultivated extracts identified a total of 30 unique taxa representing 151 fungal isolates (80% of the collection) predominantly from the Xylariaceae family of Ascomycota fungi. Of these, the largest subgroupings belonged to *Xylaria sp*. (23 extracts), *Godronia cf. cassandrae* (15 extracts), *Nemania Serpens* (12 extracts), *Rhizospaera sp.,* (11 extracts), *Phialocephala sp.* (11 extracts), and *Leptodontidium sp.* (9 extracts). Surprisingly, 33 of the extracts are undescribed fungal endophytes with no direct species matches based on ITS DNA BLAST screening. Multivariate statistical analysis of these subgroupings may also provide additional insights into the chemical space of these isolates in follow-up studies. Select ITS DNA sequences can be found in Appendix A, S1A.

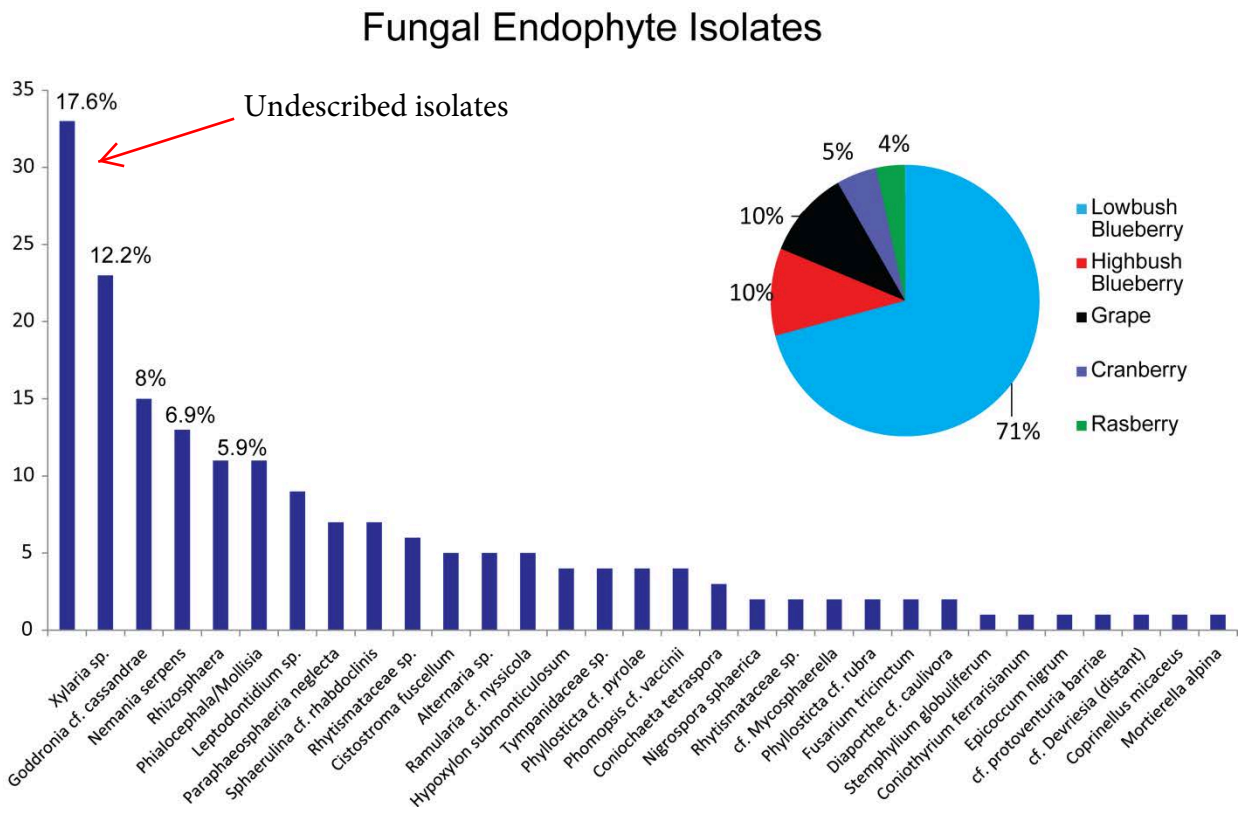


Figure 1. Fungal endophyte diversity in Canadian fruiting plants. The largest species grouping is of undescribed endophytes at 17.6% of the collection. Endophytes were sourced mostly from lowbush blueberry plants at 71% of the collection isolates.

LC-UV/MS Spectral database and MS deconvolution

There are currently no public databases of open-source LC-HRMS and MS/MS spectra of natural products from fungal endophytes. One of the key challenges in dealing with complex mixture analysis is the development of a screening and dereplication strategy to rapidly obtain LC-HRMS data and identify metabolites through accurate mass measurements using their molecular formula, UV chromophore profiles or through retention index values. Using Bruker Data Analysis 4.0 software, we created a unique database of all major UV abundant peaks and their corresponding mass spectra using Bruker's Dissect feature and Compass Library Editor. Dissect algorithms deconvoluted mass spectral components or molecular features to distinguish co-eluting components and create signature mass spectra profiles of each individual metabolite. The Dissect algorithms are similar in nature to pioneering software such as AMDIS, (Automated Mass Spectral Deconvolution and Identification System) common to GC-MS analysis, and used in conjunction with the NIST library for spectral matching.[39] The NIST library is an open source mass spectral database of ~250,000 compounds generated by electron impact (EI). A few notable software algorithms in this field include PyMS, AnalyzerPro, ProMass (Thermo), and XCMS (TSRI).[31,40] The algorithms function by subtracting background interferences or noise filtering, identifying key apex peaks for each individual mass (ion) feature within an MS profile. Profiles with similar elution times and component features are assumed to be unique. Isotope analysis and high resolution accurate mass, further improve accuracy, as related component features can be readily assigned. This process though is not completely without fault, as false positives can arise from low intensity or overly broad peaks.

The sensitivity of these algorithms is dependent on factors such as signal-to-noise ratios, MS scan rate, and on the quality of the chromatographic separations. In the case of endophyte

extracts, the deconvolution software functions fairly well provided that the metabolites are of sufficient intensity (threshold intensity ~ $1e^5$) and with a maximum of ten overlapping molecular features being selected. Instrument scan rates are becoming less of an issue with modern spectrometers, but this should also be mentioned. Assuming a typical acquisition rate is 2 Hz, this translates into ~30 data points over a 15 second elution window, typical for 3-4 mm I.D. columns, while 1Hz generates ~15 data points. Depending on the complexity of the mixture being analyzed, higher rates may be needed.

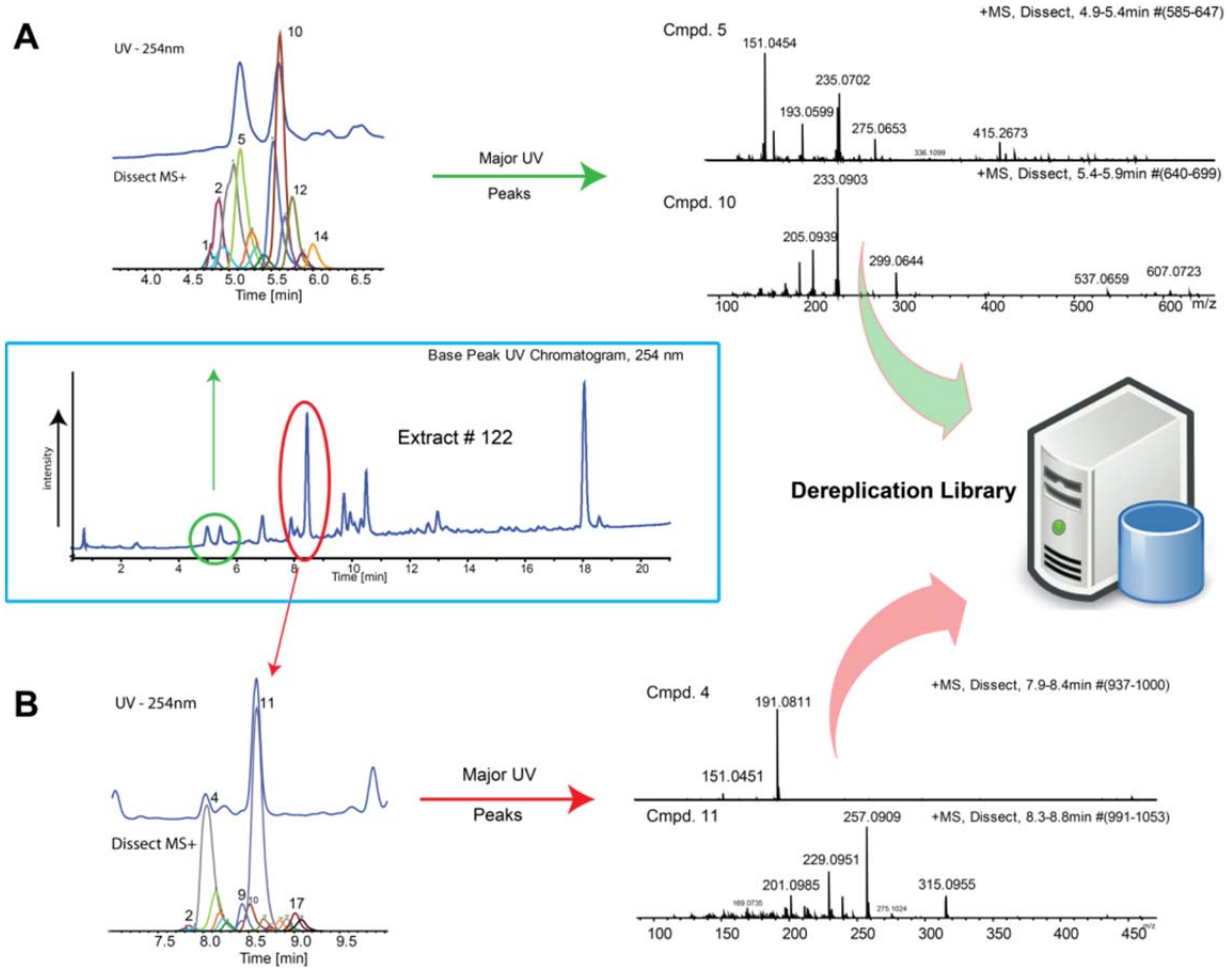


Figure 2. MS Deconvolution and Spectral library construct of endophyte extract E-122. LC-UV/MS profiles for two segments (A) and (B) are shown and deconvoluted spectra added to the in-house database.

Spectral libraries are created by applying Bruker MS Dissect algorithms to every endophyte extract being screened using a standardized LC-HRMS method. Individual Dissect molecular feature profiles are manually inspected and correlated to the corresponding UV metabolite peaks. Spectra with intensities above $1e^5$ are compiled into the in-house dereplication database and assigned an extract name, retention time, and accompanying MS spectral features. With each successive extract being screened and added to the spectral library, endophytes are systematically deconvoluted, dereplicated and categorized according to their chemotaxonomy. Spectra matches are manually verified with scores above 700 considered dereplicated. In total, 1572 MS Dissect spectra have been compiled from the 188 extracts, creating a unique MS spectral library that is specific to Canadian *spermatopsida* endophytes.

The advantage of this approach is that extracts can be rapidly dereplicated and grouped exclusively on their metabolite profiles, independent of ITS DNA sequencing or bioactivity. Using this chemometric strategy, 22 extracts were identified as unique collection isolates. In our early studies, we prioritized two of these unique extracts (E-75 and E-82) for LC-SPE-NMR isolation and characterization. This dereplication and strain prioritization approach led to the discovery of new diplosporin and agistatine derivatives isolated from an undescribed *Xylaria sp*, and a novel series of antifungal compounds from *Hypoxylon submoniticulosom*, which possess an unusual triene moiety; trienylfuranol A and trienylfuranone A-B (Figure 3).[19,21] In a similar fashion, 19 extracts from high and lowbush blueberries where dereplicated using the in-house library and found to produce the potent antifungal compound griseofulvin and de-chlorogriseofulvin.[18] This large grouping demonstrated the effectiveness our chemometric dereplication process as ITS DNA sequencing identifying all extracts as belonging to the same undescribed *Xylaria sp*. with a homology of 99%-100%. In an earlier

study, we found that these specific endophyte isolates were common to lowbush blueberries and adjacent eastern white pine trees (*Pinus strobus*), a first-report for the existence of an interesting pine-blueberry ecotype.[18]

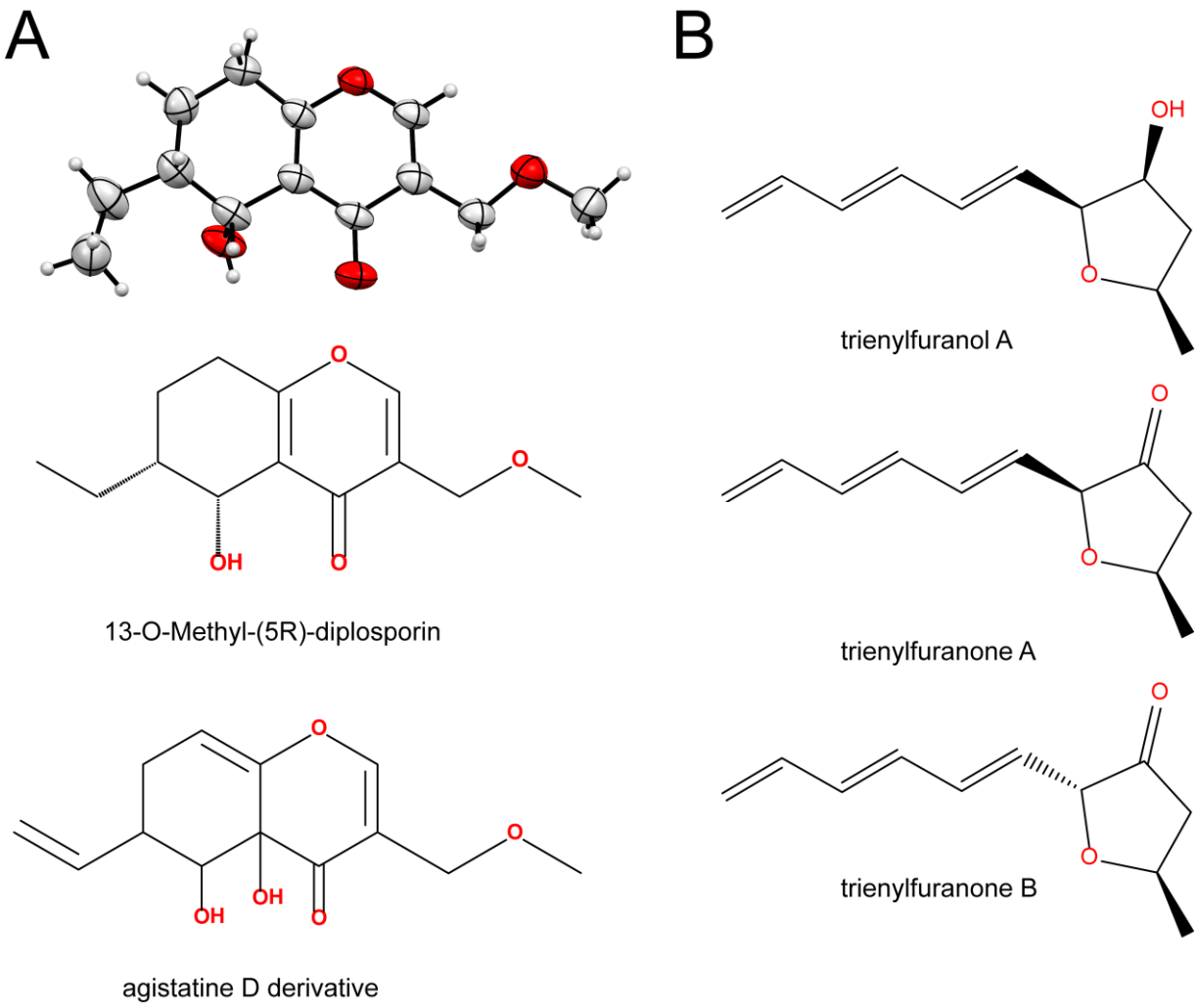


Figure 3. Reported metabolites from fungal endophyte extract E-75 and E-82. (A) ORTEP single crystal X-ray structure of 13-O-Methyl-(5R)-Diplosporin and agistatine D derivative, new metabolites isolated from an undescribed *Xylaria sp*. endophyte (B) Trienylfuranol A and trienylfuranone A-B, novel series of metabolites with an unusual triene moiety isolated from the endophyte *Hypoxylon submoniticulosom.*

LC-MS Screening and Data Pre-Processing

In an effort to identify and differentiate unique secondary metabolites between extracts, we applied unsupervised principle component analysis (PCA) to visualize any grouping patterns and variances within the collection.[11,32] The crude fungal extracts (184) were screened using a two-step process under standardized conditions on a Bruker MaXis 4G ultra-high resolution quadrupole time-of-flight (UHR-qTOF) mass spectrometer coupled to a Dionex ultimate LC system. Extracts were first run in positive and then negative electrospray ionization modes covering a mass window from 150-1200 *m/z*, over a 20 min. chromatographic separation time using a core-shell column. Each extract was spiked with sodium formate prior to analysis for internal calibration and blank samples. Pooled extracts were introduced throughout the runs, intermittently, for greater confidence. An LC-MS system suitability check (six replicate runs) was performed using two different endophyte extracts (E-104 and E-122) and no significant deviations were noted in terms of reproducibility (peak retention times, peak areas, peak tailing). The raw LC-MS data files were re-calibrated post-acquisition using the internally spiked sodium formate ion clusters and converted to mzXML format using Bruker's CompassXport software. Each of the mzXML data files were then processed using CAMERA for extracting compound peaks or molecular features, and removing of adduct peaks and annotation of isotopes. This was followed by XCMS processing for nonlinear retention time alignments for deconvoluted molecular features, peak detection and filtering. We applied a filter of 0.1% and a minimum intensity threshold of $1e^4$ to reduce false positives and improve peak alignments.

Principle component analysis (PCA) of crude fungal endophyte extracts

Multivariate statistical analysis was used to investigate the relationship between 184 diverse fungal endophyte extracts and their secondary metabolites. Principle component analysis (PCA) can reduce the dimensionality (retention times, m/z values, peak intensity) of large LC-HRMS data sets to reveal trends, observations and unique outlier extracts that can be prioritized and targeted for metabolite isolation and characterization.[10,32,41] In total 13649 (ESI+) and 5590 (ESI-) extracellular metabolite features were identified from the post-processed LC-HRMS datasets using CAMERA and XCMS software. The extracellular data sets were subjected to pareto scaling, to reduce the significance of high intensity peaks and enhance the contribution of weaker peaks, using SIMCA-P + 12.0.1 prior to PCA modeling. PCA models are represented visually as two-dimensional Scatter plots of Scores and Loadings (Figure 4A and 5A). The Scores plot represents each individual observation of the fungal extract that has been processed, while the colour class labels associated with each extract represents the identified fungal species (Appendix A, S2-S3A). The separations or clustering between the observations (extracts) in a Scores plot is representative of the covariance between the individual observations. In essense, the smaller the covariance between each observation, the tighter the clustering or grouping will be, based on common metabolite features present within each extract. The Loadings plot can be viewed as representing the individual variables or metabolite features (compounds) that are associated with a corresponding observation in the Scores plot. Each molecular feature is defined as being unique, with a specific mass-to-charge value and retention time. Similarly, the location or positioning (X and Y axis) of the variables in the Loadings plot, are generally those that are contributing or belonging to observations seen in a similar locality within the Scores plot (Figure 4B and 5B, Appendix A, S2-S3A).

In the PCA (ESI+) scatter plot of the Observation Scores (Figure 4A), some of the fungal extracts can be seen clustering or grouping together, near the X/Y-axis, while others are clearly divergent, separating away from the bulk of the collection. Interestingly, the grouping of low and highbush blueberry isolates from *Xylaria sp.* (blue), identified earlier by dereplication and database screening, are separated and clustering to the left along the X-axis. Clustering can also be seen with other fungal species such as *Papraphaeosphaeria sp.* (red) and *Nemania serpens* (green). The clustering of the various species classes is not unexpected as many of the isolates may produce common metabolites. The separations within species can be attributed to varying yields and production of different or unique metabolites. Highlighted in red circles are two highly divergent outlier extracts, E-112 and E-184, which have been prioritized for metabolomic-guided discovery efforts. In the PCA (ESI-) scatter plot of the Observation Scores (Figure 5A), a similar separation and clustering pattern of the fungal extracts can be seen. Highlighted in red is E-184, which is also a unique outlier extract in negative ionization mode. In Figure 4B and 5B, a scatter plot of the Scores and Loadings is combined into a single BiPlot, revealing the covariance between the extracts and their associated metabolite features. Highlighted in light grey are the Loading variables or metabolite features that have not been characterized, while in red are metabolites that have been characterized by LC-HRMS, MS/MS and/or 1D and 2D NMR analysis. Based on the Scores and Loading plots of the outlier extracts and clustered species groupings we prioritized a total of 17 extracts (E-006, E-035, E-038, E-046, E-051, E-101, E-112, E-140,E1-69, E-182, E-184, E-195, E-223, E-225, E-252, E-260 and E-261), spanning 14 fungal species (12 identified and 2 unidentified) (Appendix A, S1A).

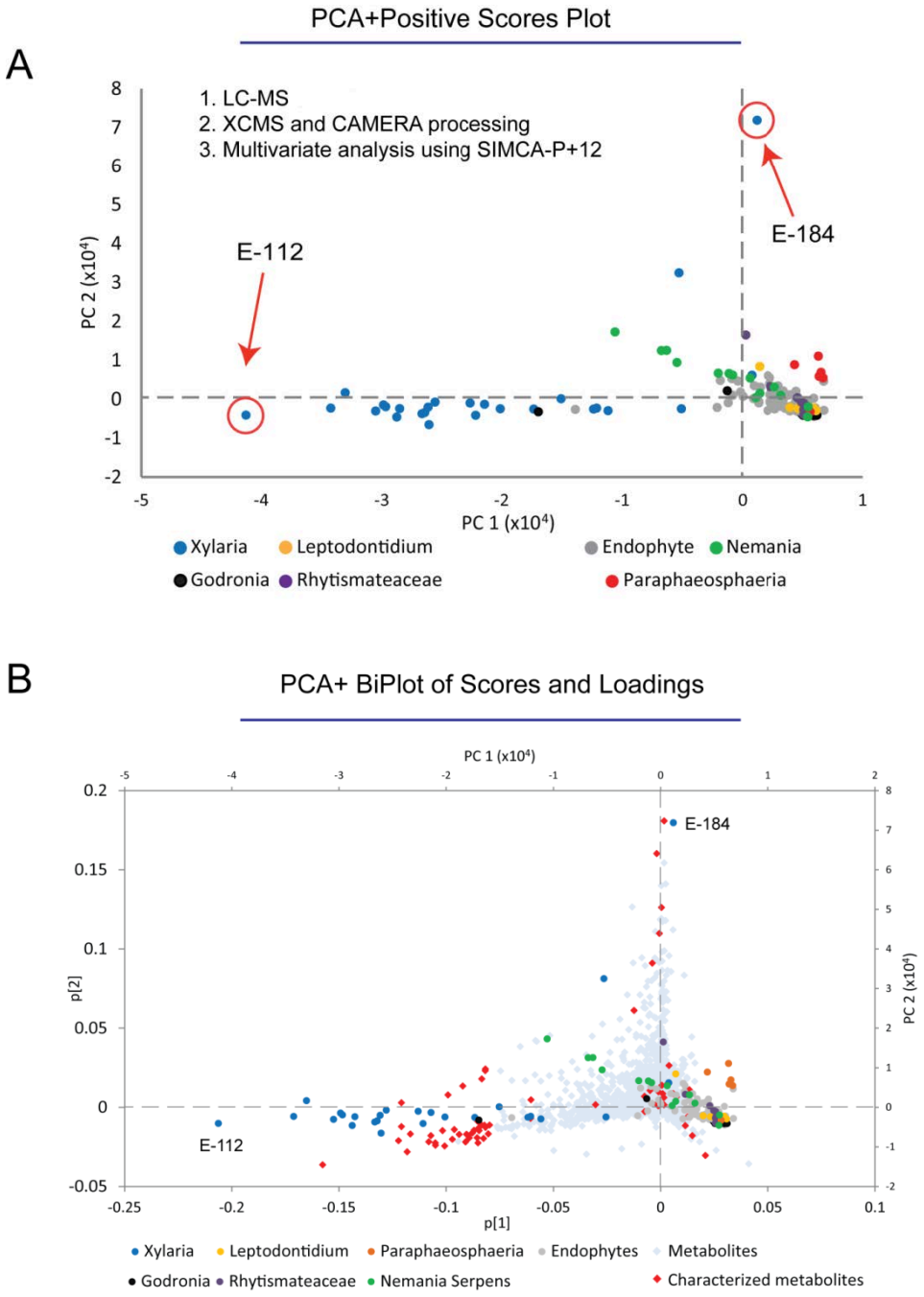


Figure 4. PCA (ESI+) Scatter plots of (A) Observation Scores and (B) BiPlot of Observation Scores and Variable Loadings. (A) PCA+ Score plot of 184 fungal endophyte extracts analyzed by LC-HRMS and pre-processed using CAMERA and XCMS. The color labels represent the species identification based on ITS DNA sequencing. Highlighted in red circles are two extracts, E-112 and E-184, which are divergent outlier extracts in the X and Y dimension. (B) A BiPlot of the Scores in (A) and Loadings representing the variables or metabolite features associated with the scores (total of 13649 variables). Metabolite features shown in Gray are not characterized. Characterized metabolites are represented in Red.

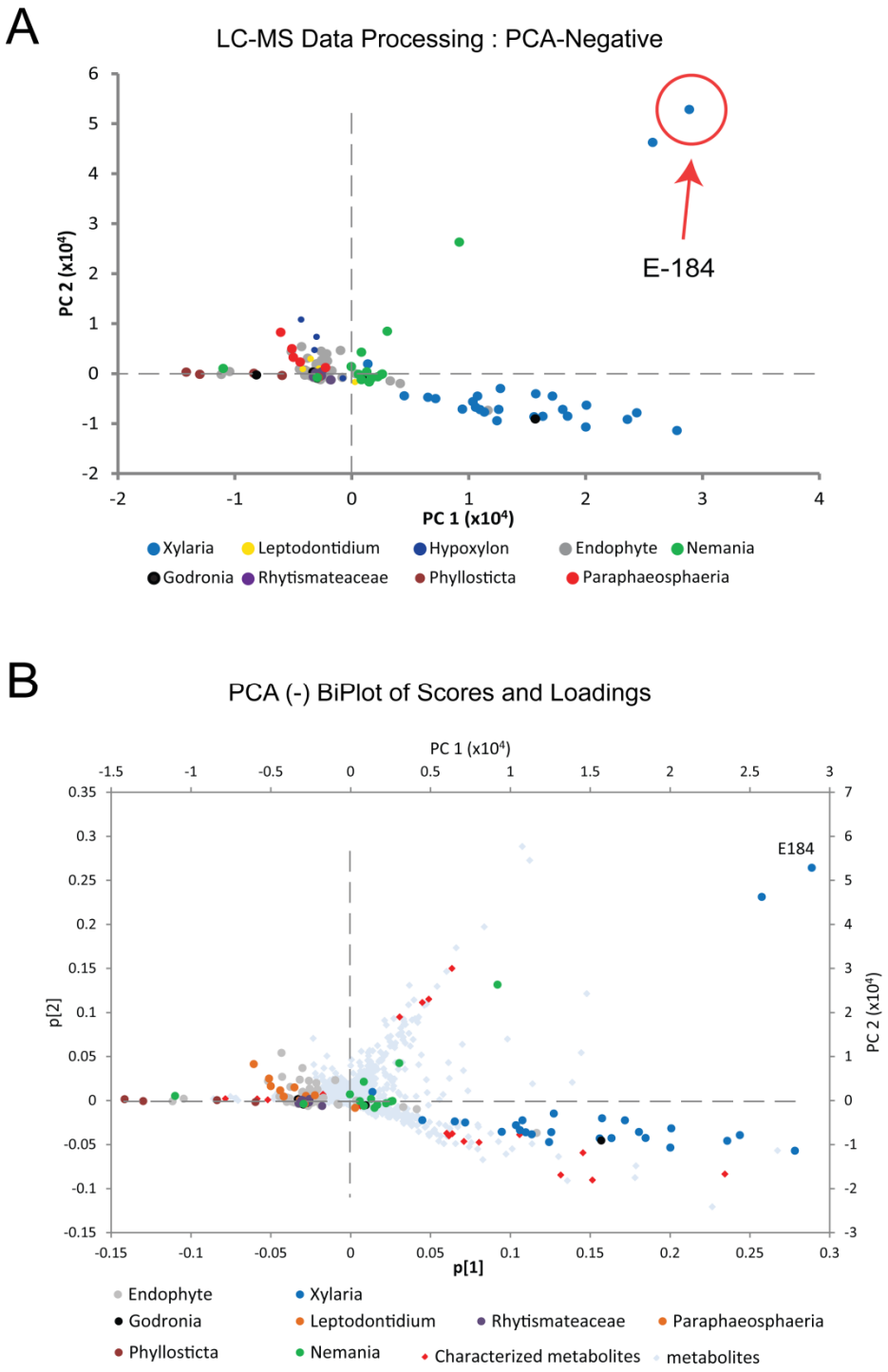


Figure 5. PCA (ESI-) Scatter plots of (A) Observation Scores and (B) BiPlot of Observation Scores and Variable Loadings. (A) PCA- Score plot of 184 fungal endophyte extracts analyzed by LC-HRMS and pre-processed using CAMERA and XCMS. The color labels represent the species identification based on ITS DNA sequencing. Highlighted in a red circle is E-184, a divergent outlier extracts in the X and Y dimension. (B) A BiPlot of the Scores in (A) and Loadings representing the variables or metabolite features associated with the Scores (total of 5590 variables). Metabolite features shown in light Gray are not characterized. Characterized metabolites are represented in Red.

Metabolomic-guided discovery of novel secondary metabolites

Having observed a number of outlier extracts and clustered groupings by PCA, we next sought to examine each extract in a step wise process by: (1) manually examining and correlating the major molecular features (retention times and mass-to-charge value) from Score Contribution plots to LC-UV/HRMS chromatograms, validating true or false positive hits; (2) LC-HRMS dereplication screening of putative compounds using their molecular formula and LC-MS/MS dereplication screening using iSNAP metabolomic tools; (3) LC-SPE-NMR and/or semi-preparative reversed-phase HPLC isolation and purification; (4) structure elucidation and characterization; and (5) preliminary assessment of bioactivity against a panel of pathogenic microorganism. Using this approach, a total of 12 new and 37 knowns compounds were discovered (Figure 6 and 7).

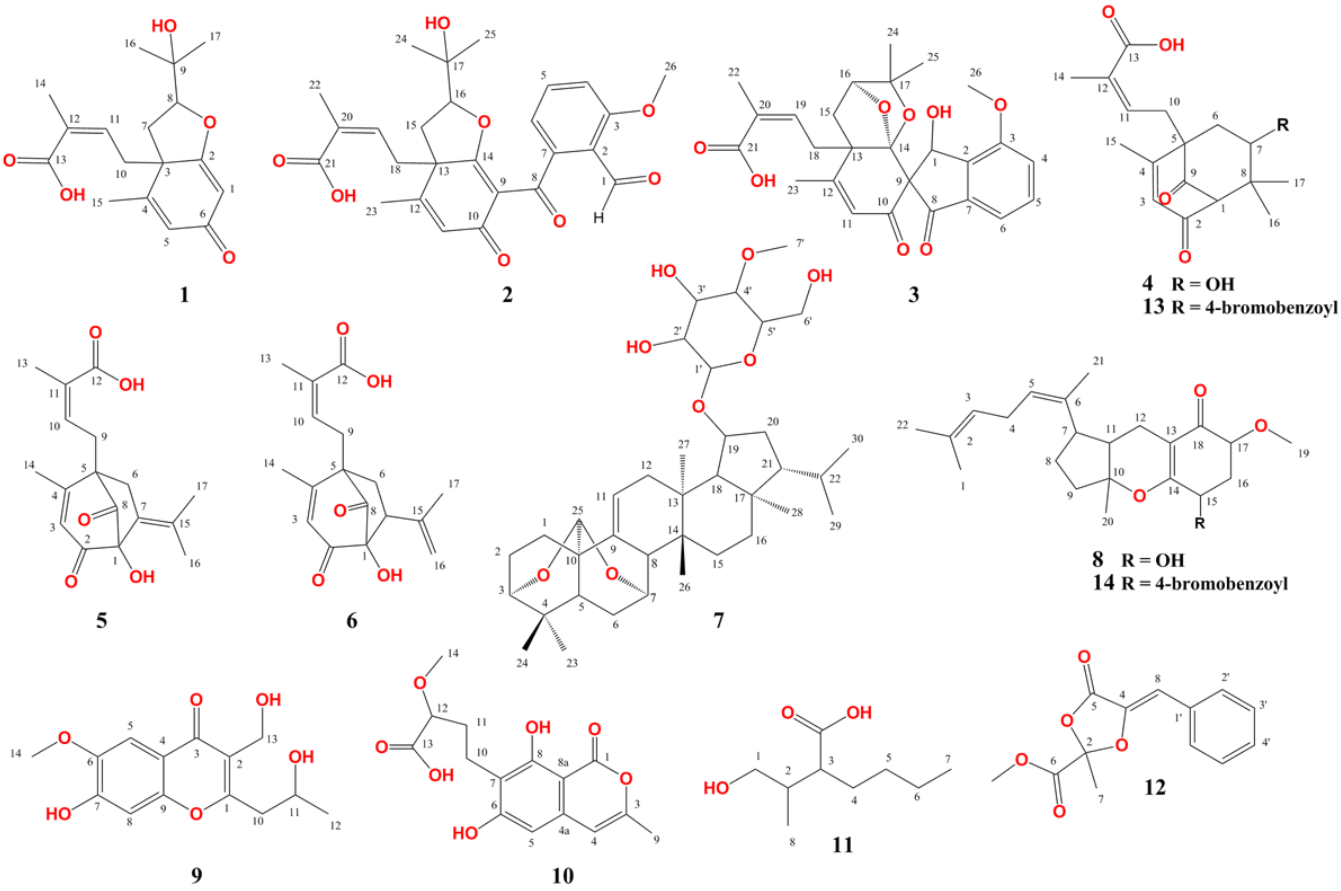


Figure 6. New compounds **1-12** and semisynthetic derivatives **13-14**.

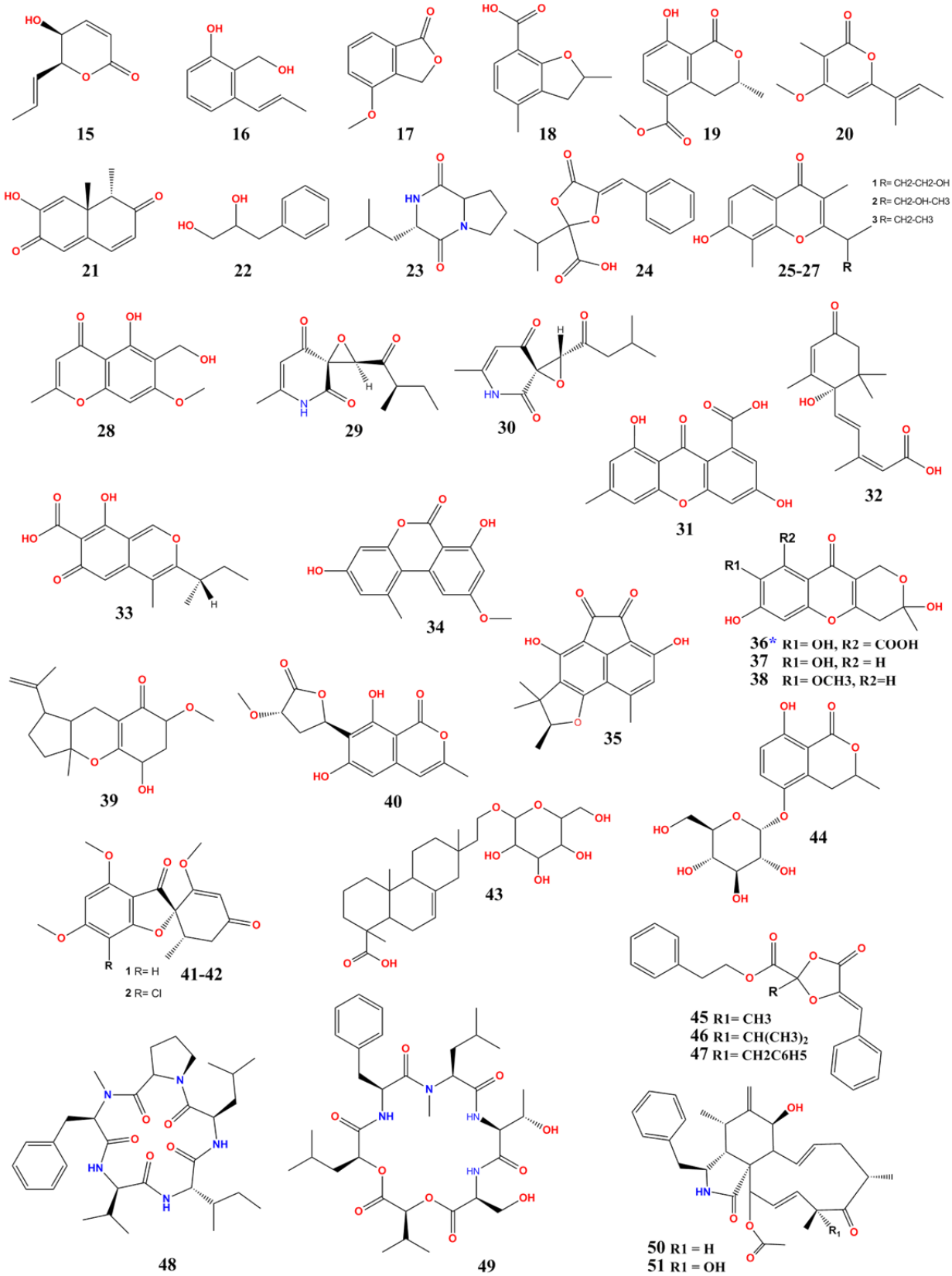


Figure 7. Known compounds 15-51.

Fungal Endophyte extract E-184

A total of 6 previously undescribed and 4 known compounds were discovered from the culture filtrates of *Xylaria cubensis,* an endophytic fungus from the leaves and stems of a Riesling grape located at Jordan Research Farm (Agriculture and Agri-food Canada) in Ontario, Canada. The cell-free filtrate extract was assigned the name E-184 and identified as a distinct isolate within the fungal collections. Examination of E-184's positive (ESI+) Score Contribution plot revealed a number of significant contributing compounds, of these, two priority candidates were selected; +451.1 *m/z* at R_t=17.2min, and +165.1 *m/z* at R_t=12.4min (Figure 8A, Appendix A, S4A). In the negative (ESI-) Score Contribution plot, two candidate compounds were selected having identical m/z values but unique retention times: +305.1 *m/z*, R_t=10.2 min and +305.1 *m/z*, R_t = 11.6 min (Figure 8 B, Appendix A, S5A). LC-UV/MS analysis of the putative candidates' extracted ion chromatograms (EIC) revealed them as true hits (Figure 8C). We next fermented 4 L (20 x 200 mL Roux bottles) of strain E-184, to obtain sufficient material for compound isolation, structure elucidation and characterization. Extracted with ethyl acetate solvent (Figure 8C), the E-184 structures appear to be biosynthetically related, with all structures having a similar carboxylic acid moiety.

Compound **1** was isolated as golden oil with a HRESIMS affording a protonated molecular ion at *m/z* = 307.1545, indicating a molecular formula of $C_{17}H_{22}O_5$ with 7 double bond equivalents. The dereplication screening using Antibase revealed 15 fungal derived candidates, however, none correlated with NMR data. Similarly, iSNAP dereplication screening identified no candidates. Compound **1** is a novel specialized metabolite with sub-structure similarity to plant derived acylphloroglucinol compounds and hyperforatin E.[42,43] However,

compound **1** is non-prenylated and possesses a carboxylic acid moiety. The structure was solved with extensive 1D and 2D NMR spectroscopy (Table 3, Figure 9, Appendix B, S1-8B).

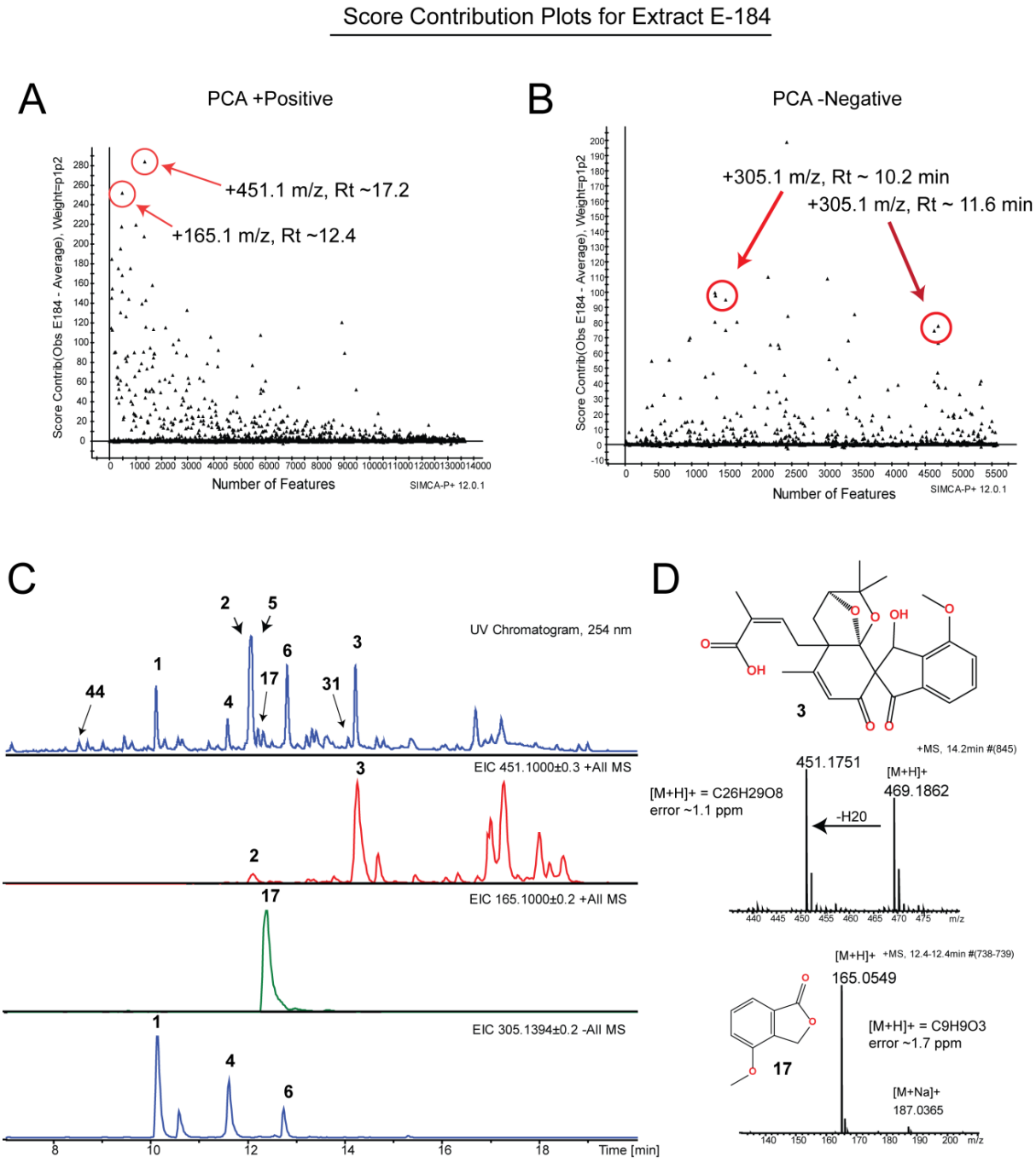


Figure 8. Metabolomic guided discovery. (A) PCA (ESI+) Score Contribution plot of extract E-184, with major candidate compounds highlighted. (B) PCA (ESI-) Score Contribution plot of extract E-184, with major candidate compounds highlighted. (C) LC-UV/MS chromatogram and extracted ion chromatograms of the putative candidates in (A) and (B) with compound **1**-**6** highlighted. (D) Compounds 3 and 15 structures with corresponding LC-HRESIMS spectra.

Table 3. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of (1) in CD_3OD.

CD_3OD				
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	NOESY
1	101.3, CH	5.53, s	2, 5,	
2	185.4, C			
3	55.3, C			
4	157.7, C			
5	128.3, CH	5.97, d (1.6)	1, 3, 15	15
6	191.9, C			
7a	34.7, CH2	2.14, dd (12.2, 10.3)	2, 3, 4, 8, 9, 10	
7b		2.31, dd (12.2, 5.4)		
8	91.4, CH	4.73, dd (10.2, 5.4)	17	7a, 10, 16, 17
9	71.7, C			
10	36.9, CH2	3.10, qd (14.6, 7.7)	2, 3, 4, 7, 11, 12	7a, 9
11	133.8, CH	5.52, m	13, 14	14
12	133.4, C			
13	171.3, C			
14	21.3, CH3	1.81, d (1.5)	12, 13	11
15	19.0, CH3	2.03, d (1.4)	3, 4, 5	5, 10
16	25.0, CH3	1.16, s	8, 9, 17	7a/b, 9
17	26.2, CH3	1.31, s	8, 9, 16	9

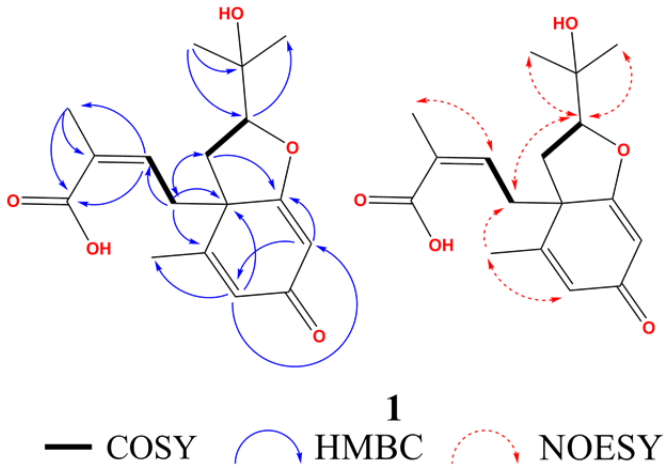


Figure 9. HMBC, COSY and NOEY correlations for compound 1

Compounds **2** and **3** were isolated as light-golden oils with similar molecular weights. We discovered that the putative compound at *m/z* of +451.17, is likely arising from the loss of *m/z* 18 (-H_2O) from a parent molecule at m/z +469.18 (Figure 8D). As such, we isolated three *m/z* +469.18 targets at R_t = 12.1, 14.2 and 17.2 min. The HRESIMS for **2** and **3** afforded

protonated molecular ions at *m/z* 469.1847 and *m/z* 469.1862 respectively, indicating a molecular formula of $C_{26}H_{28}O_5$ with 13 double bond equivalents. No dereplication matches were found using iSNAP (LC-MS/MS data) or in Antibase. Extensive 1D and 2D NMR spectroscopy identified the compounds as novel with some similarity to the coleophomones A and B (Table 4, Figure 10-11, Appendix B, S9-26B). Coleophomones A-B were isolated from the fungus Coleophomoa sp. (Merck Frost 6338), isolated from unidentified plant litter, and are a novel series of bacterial transglycosylase inhibitors (IC_{50} = 62 μM) with weak antibacterial activity.[33] We speculate that compound **1** is a biosynthetic intermediate of compounds **2-3.**

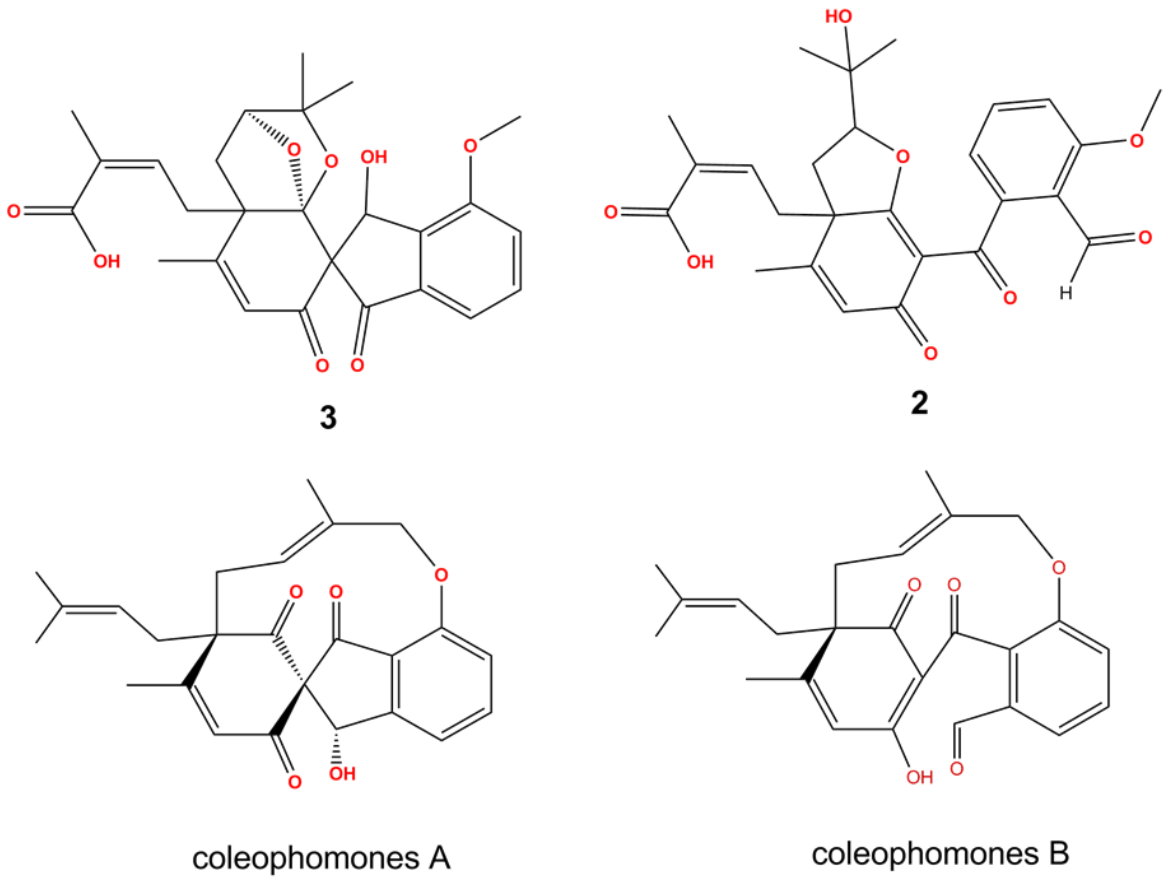


Figure 10. Structures of compounds 2-3 and coleophomones A-B.

Table 4. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR data of (**2**) and (**3**) in C_6D_6.

	(**2**)			(**3**)		
Position	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC
1	190.9, C	10.16, s	2, 7	72.0, CH	6.16, s	2, 3, 7, 8, 9, 14
2	122.3, C			141.7, C		
3	161.3, C			157.5, C		
4	111.9, CH	6.17, d (8.3)	1, 3, 6, 7,	116.0, CH	6.24, d (7.9)	1, 2, 3, 6
5	135.7, CH	7.08, t (7.9)	3, 4, 6, 7,	131.1, CH	6.89, t (7.8)	3, 6, 7 ,8
6	121.8, CH	7.24, d (7.5)	2, 4, 5, 8S	115.2, CH	7.48, d (7.6)	2, 3, 4, 8
7	145.9, C			140.1, C		
8	194.1, C			197.8, C		
9	115.9, C			67.5, C		
10	185.1, C			194.8, C		
11	128.8, CH	5.87, d (1.5)	9, 10, 13, 23	127.9, CH	6.01, d (1.43)	9, 13, 23
12	152.2, C			164.3, C		
13	55.4, C			56.7, C		
14	183.8, C			111.7, C		
15a	32.7, CH2	2.29, dd (12.0, 9.7)	12,13,16,17,18	33.6, CH2	1.58, m	13, 16, 18
15b		1.50, dd (12.1, 5.7)				
16	92.2, CH	4.52, dd (9.8, 5.6)	17, 25	82.6, CH	3.73, dd (3.7, 2.2)	13, 14, 17
17	70.8, C			81.9, C		
18a	36.6, CH2	3.08, dd (14.7, 7.3)	12,13,14,19,20	34.6, CH2	3.80, dd (6.1, 8.7)	12,13,14,15,19,20
18b		2.75, dd (14.5, 8.3)			3.13, m	
19	136.3, CH	5.63, ddd (8.7,7.1,1.7)	13, 21, 22	141.5, CH	6.11, t (7.3)	13, 21, 22
20	131.3, C			129.2, C		
21	170.8, C			171.6, C		
22	20.8, CH3	1.76, d (1.5)	19, 20, 21	21.1, CH3	1.94, d (1.8)	19, 20, 21
23	18.0, CH3	1.35, d (1.4)	11, 12, 13,14	20.4, CH3	1.47, d (1.3)	11, 12, 13
24	24.2, CH3	0.99, s	16, 17, 25	21.4, CH3	0.88, s	17, 25
25	27.7, CH3	1.57, s	16, 17, 24	27.2, CH3	0.93, s	17, 24
26	55.1, CH3	2.91, s	3	55.2, CH3	3.11, s	3

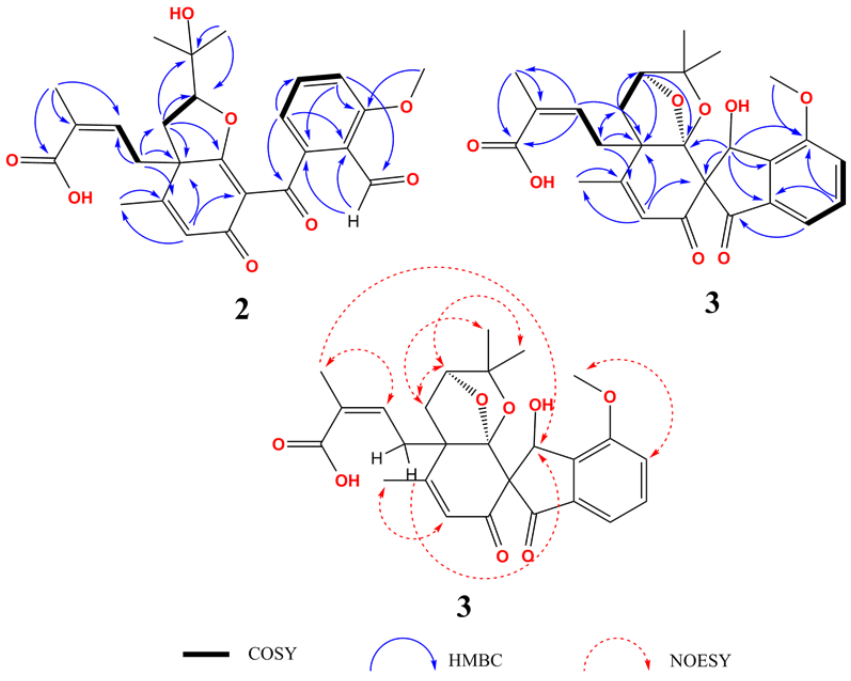


Figure 11. HMBC, COSY and NOEY correlations for compound 2-3

Compound **4** was isolated as golden oil with a HRESIMS affording protonated molecular ion at *m/z* 307.1555, indicating a molecular formula of $C_{17}H_{22}O_5$ with 7 double bond equivalents. Compound **4** is bicyclic non-prenylated metabolite possessing a similar carboxylic acid moiety as seen in compounds **1-3**. The structure was solved with extensive 1D and 2D NMR spectroscopy and is structurally similar to plant derived polycyclic prenylated acylphloroglucinols (Table 5, Figure 12, Appendix B, S27-44B).[43] Synthesis of a 4-bromobenzoate derivative of **4** resulted in compound **13** (Appendix B, S100-106B). Derivatization was performed to generate crystals suitable for X-ray analysis and to determine the absolute stereochemical assignments. The derivatization method was adapted from Burgess *et al.*, which was recently reported.[21] Compound **13** was successfully synthesized however, no suitable crystal was obtained.

Table 5. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR data of (**4**) in C_6D_6 and (**13**) in CD_2Cl_2.

	(**4**)			(**13**)		
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
1	73.8, CH	3.12, d (0.9)	2, 3, 4, 5, 7, 8, 9, 16, 17	74.3, CH	3.08, dd (8.8,1.0)	2, 3, 7, 8, 16
2	194.4, C			194.4,C		
3	131.7, C	5.96, t (1.4)	1, 5,15	132.4, CH	6.37, m	4, 5, 15
4	161.0, C			161.4, C		
5	55.2, C			55.1,C		
6a	38.4, CH2	1.32, dd (13.5, 11.3)	4, 5, 7, 8, 9, 10	33.5, CH2	1.89, m	1, 4, 5, 7, 8, 9
6b		1.42, dd (13.6, 5.3)			2.36, m	
7	70.8, CH	3.16, dd (11.3, 5.3)	6, 8 , 16, 17	74.31,CH	5.16, dd (11.2, 5.4)	6, 8, 16, 17, 18
8	43.2, C			43.0, C		
9	205.5, C			205.0, C		
10a	32.7, CH2	2.77, ddd (17.5, 7.0, 1.6)	4, 5, 6, 9, 11, 12	31.8, CH2	2.49, dd (17.1, 7.0)	5, 9, 10, 11, 12
10b		3.03, ddd (17.4, 5.7, 2.1)			2.76, m	
11	142.3, CH	5.84, ddd (7.2, 5.6, 1.6)	5, 13, 14	139.7, CH	6.60, ddd (7.2, 5.7, 1.6)	5, 10, 12,13
12	127.7, C			129.5, C		
13	171.9, C			171.3, C		
14	20.4, CH3	1.73, q (1.7)	11, 12, 13	20.1, CH3	2.08, d (1.4)	3, 4, 5
15	19.2, CH3	1.38, d (1.4)	2, 3, 5, 6,	20.8, CH3	1.89, d (1.4)	12, 13
16	18.7, CH3	0.72, s	1, 7, 8, 17	20.8, CH3	1.15, s	1, 7, 8, 17,
17	25.1, CH3	0.96, s	1, 7, 8, 16	25.1, CH3	1.06, s	1, 7, 8, 16
18				165.5, C		
19				128.9, C		
20				131.6, CH	7.86, m	18, 19, 21
21				132.4, CH	7.61, m	19, 20
22				129.3, CBr		

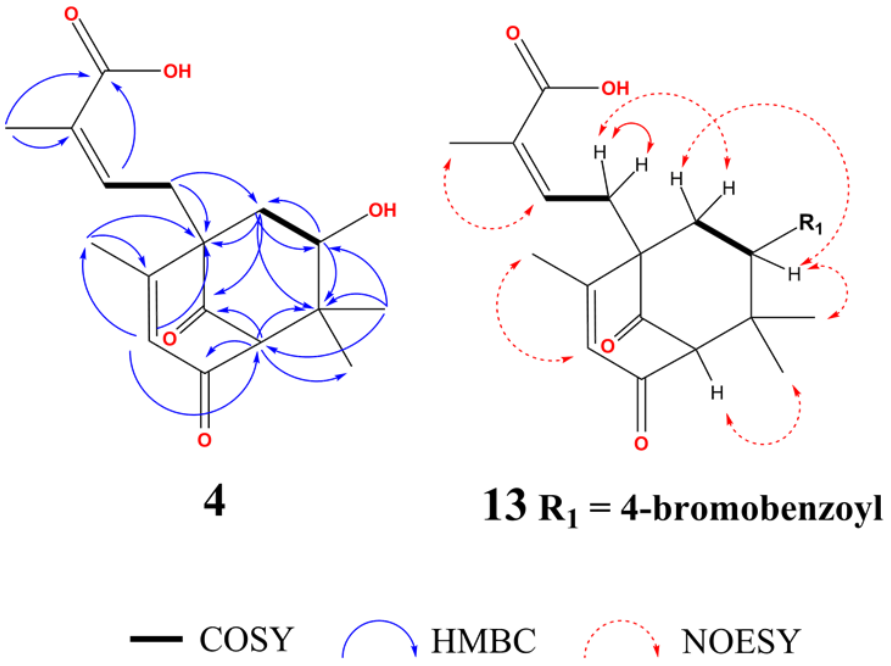


Figure 12. HMBC, COSY and NOESY correlations for **4** and **13**.

Compounds **5** and **6** were isolated as golden oils with similar HRESIMS at Rt = 12.2 min and 12.9 min respectively. Their HRESIMS affording protonated molecular ions of *m/z* +305.1390 and +305.1385, indicating a molecular formula of $C_{17}H_{20}O_5$ with 8 double bond equivalents. Compound **5** and **6** are unusual bicyclic non-prenylated metabolites possessing a similar carboxylic acid moiety as seen in compounds **1-3**. Their structures were solved with extensive 1D and 2D NMR spectroscopy (Table 6, Figure 13, Appendix B, S34-49B).

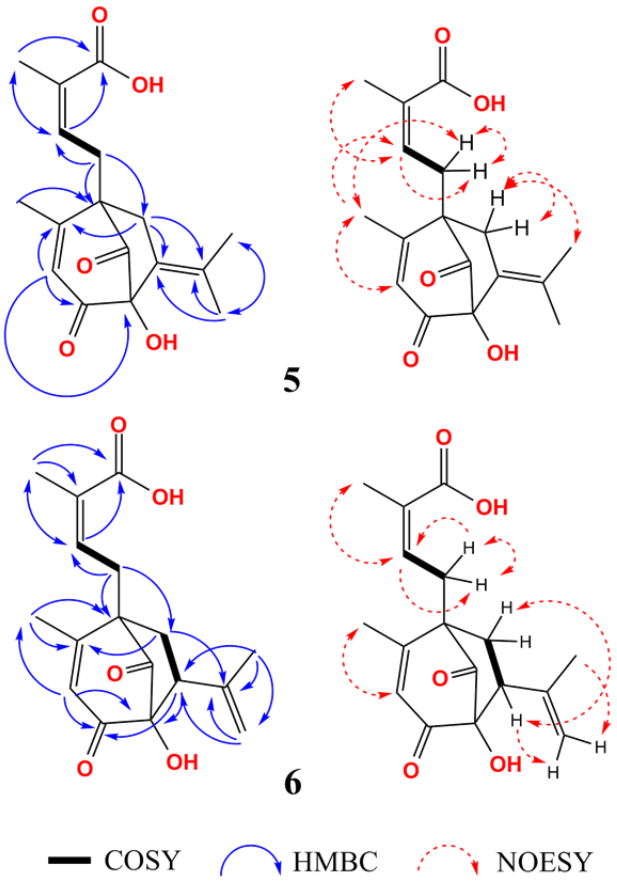


Figure 13. HMBC, COSY and NOESY correlations for **5** and **6**.

Table 6. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of (**5**) in DMSO-*d*6 and (**6**) in C_6D_6.

	(**5**)			(**6**)		
Position	δ_C	δ_H (*J* in Hz)	HMBC	δ_C	δ_H (*J* in Hz)	HMBC
1	88.8, C			90.9, C		
2	200.9, C			196.5, C		
3	130.7, CH	5.99, s	1, 2, 4, 5, 14	124.6, CH	5.74, s	1, 2, 4, 5, 14
4	177.1, C			170.2, C		
5	55.6, C			55.1, C		
6a	34.0, CH2	2.35, d (16.5)	4, 5, 7, 8, 9, 15	35.0, CH2	1.47, dd (13.8, 10.4)	1, 4, 5, 7, 8, 9, 15
6b		2.87, m			1.81, dd (13.8, 6.0)	
7	128.1, C			46.6, CH	2.38, dd (10.4, 6.0)	1, 2, 8, 15,16, 17,
8	199.5, C			204.6, C		
9a	32.1, CH2	2.45, dd (14.2, 6.5)	5, 6, 10, 11	29.7, CH2	2.88, ddd (16.0, 7.8, 1.4)	4, 5, 6, 8, 10, 11
9b		2.85, m			3.18, ddd (16.0, 6.5, 1.7)	
10	124.4, CH	4.88, t (8.0)	11, 12	138.9, CH	6.07, tt (8.0, 1.6)	5, 11, 12
11	137.6, C			130.7, C		
12	171.8, C			170.7, C		
13	21.8, CH3	1.62, s	5,10, 11, 12	20.9, CH3	1.84, d (1.6)	10, 11, 12
14	14.9, CH3	2.02, s	2, 3, 4, 5	19.7, CH3	1.43, d (1.3)	2, 3 ,4 ,5
15	151.6, C			143.4, C		
16a	20.6, CH3	2.08, d (2.4)	7, 8, 15, 17,	115.5, CH2	4.62, d (1.74)	7, 15, 17
16b					4.73, t (1.6)	
17	24.4, CH3	1.87, s	7, 8, 15, 16	18.8, CH3	1.68, d (1.2)	7, 15, 16

Compound **17** was identified as 4-methoxy-1(3H)-isobenzofuranone and was the second greatest contributor to the Score Contribution plot for E-184 (Appendix A, S4A). The HRESIMS afforded a protonated molecular ion at *m/z* 165.0549, giving a molecular formula of $C_9H_8O_3$, with 6 double bond equivalents. Compound **17** was isolated from semi-purified HPLC fractions and crystallized out of solution in cold methanol at -20 °C, producing colorless needles. The structure was confirmed by proton and carbon NMR analysis (Appendix C, S5-6C).[44]

Compound **31** was identified as a known naphthalene derivative co-eluting with **3** and was crystalized in methanol (Appendix C, S25C).[45] Compound **44** was identified as 5-*O*-α-D-glucopyranosyl-5-hydroxymellein and present within the top 120 contributing features (Appendix A, S4A; Appendix C, S47-48C).[46] Similarly, compound **19** was identified as a known mellein derivative, common to fungal endophytes (Appendix C, S7-8C). [19]

Fungal Endophyte extract E-112

In a similar PCA mining approach to E-184, one previously undescribed compound and 8 known compounds were discovered from the culture filtrates of *Xylaria sp.*, an endophytic fungus from the leaves and stems of lowbush blueberries isolated in the Acadian forest of Nova Scotia, Canada. Examination of E-112's positive (ESI+) Score Contribution plot revealed a number of significant contributing compounds, of these, two priority candidates with similar retention times were selected for isolation; *m/z* +437.3 at Rt=17.2 min, and *m/z* +631.4 at Rt = 17.2 min (Figure 14, Appendix A, S6A). In a second pass screen, a number of significant candidates were investigated that were common to E-112 and the related neighboring *Xylaria sp.* isolates. These were later identified as known compounds **18**, **41**-**42**, and **48**-**51**.[18,47–50] Dereplication and analog screening using iSNAP metabolomic tools dereplicated the seed

structure for compound **48** and discovered 3 new analogues, as putative cyclic nonribosomal peptides.

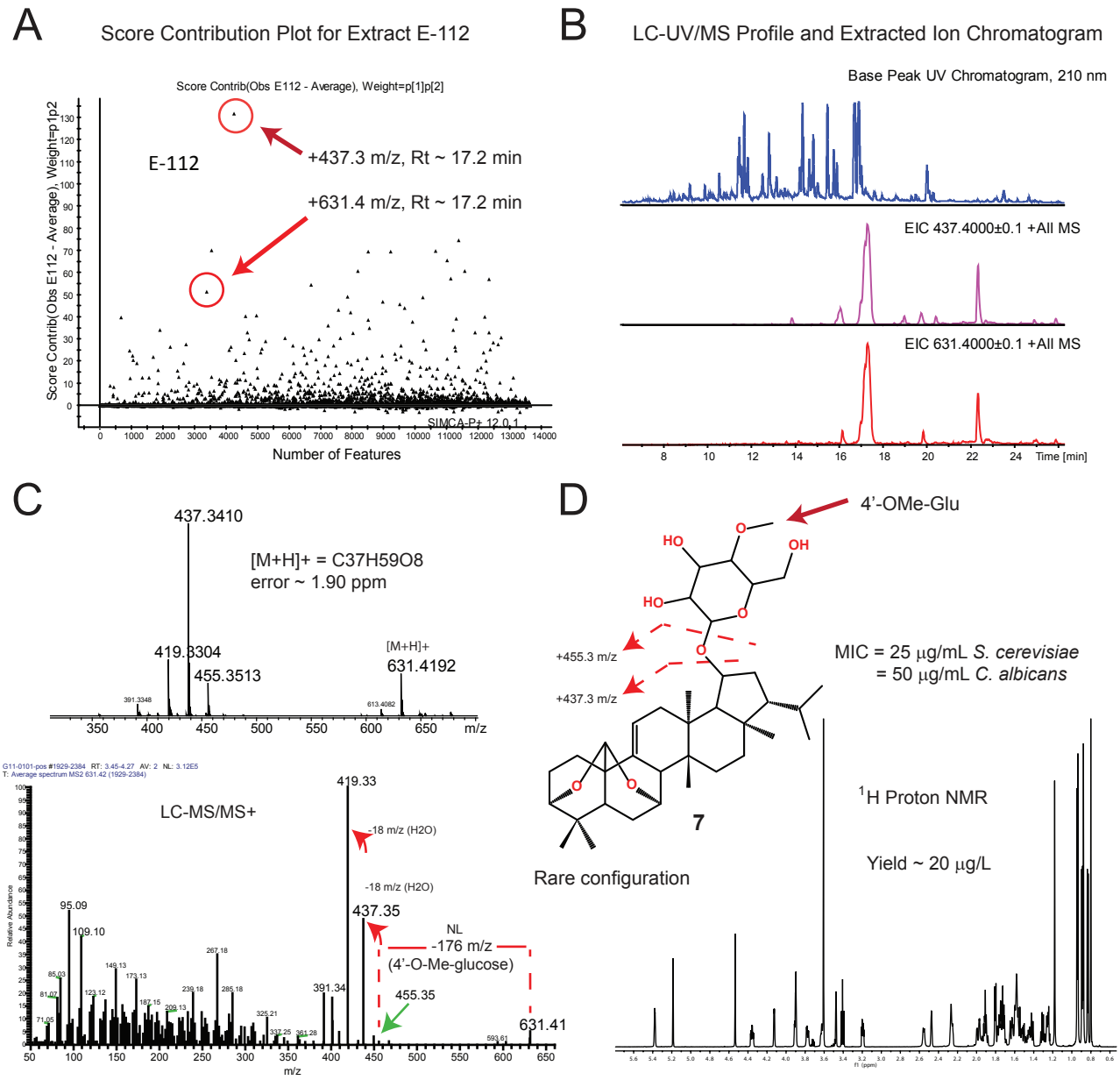


Figure 14. Metabolomic guided discovery of griseofernaneoside A. (A) PCA (ESI+) Score Contribution plot of extract E-112, with major candidate compounds highlighted. (B) LC-UV/MS chromatogram and extracted ion chromatograms of the putative candidates in (A). (C) LC-HRMS and MS/MS spectra of griseofernanoside A. (D) ^{1}H proton NMR spectra.

Griseofernaneoside A (**7**) was isolated as white powder with HRESIMS affording a protonated molecular ion at m/z 631.4192, indicating a molecular formula of $C_{37}H_{58}O_8$ with 9

double bond equivalents. Examination of the LC-MS and MS/MS spectra in Figure 14C, shows a potential neutral loss of a aglycone moiety (m/z 176) from m/z +631.45 to 455.35, and an additional H_2O loss, giving the major contributing molecular feature at m/z = +437.34 (Figure 14 B-C). Griseofernaneoside A (**7**) was isolated by semi-preparative HPLC and crystalized from methanol. Although crystals were generated, they were of poor quality for complete structural determination. The structure was solved with extensive 1D and 2D NMR spectroscopy and identified as a novel fernane-type triterpenoid (Table 7, Figure 17, Appendix B, S50-59B). The fernane-type structure belongs to the hopane class of pentacyclic triterpenoids, that have been isolated exclusively from ferns and several plants.[34,51] To our knowledge, kolokoside B is the only reported fernane-type triterpenoid that has been isolated from a microbial source.[35] Griseofernaneoside A is interesting as it contains an unusual bridged furan ring system, with only a few reported *ent*-kaurane diterpenoids having a similar moiety.[52] Similarly, a SciFinder and Antibase substructure screening of pentacyclic scaffolds containing a 4-O-methyl glucoside moiety revealed no matches, making griseofernaneoside A a unique metabolomic-guided discovery. However, a lanostane-type triterpenoid has been reported that contains a α–4-O-methyl glucoside; the antifungal compound ascosteroside isolated from *Ascotricha amphitricha*.[53]

Examination of the LC-MS/MS data revealed a new putative analogue of griseofernaneoside A (**7**) with a HRESIMS affording a protonated molecular ion at m/z +647.4216, indicating a molecular formula of $C_{37}H_{58}O_9$ with 9 double bond equivalents. Examination of the LC-MS/MS spectra showed a similar pattern to (**7**) with a neutral loss of the aglycone moiety from m/z +647.4216 to 471.3451 and with similar fragmentation for the

pentacyclic core (Appendix B, S60B). The putative compound has been named griseofernaneoside B and is a new hydroxylated analogue of (**7**).

Interestingly, LC-MS/MS dereplication screening of extract E-112 using iSNAP algorithms revealed no dereplication hits for the two outlier compounds that had been isolated; the nonribosomal peptides cyclic pentapeptide 1 (**48**) and hirsutatin A (**49**), a cyclohexadepsipeptide. Upon closer examination, the structures were not present in the nonribosomal peptide database. However, using iSNAP's database and structure upload feature, the SMILES code of compound **48** was uploaded and screened for analogue structures using LC-MS/MS data acquired on a Thermo Scientific Q-Exactive Quadrupole Orbitrap mass spectrometer. Compound **48** was correctly identified and dereplicated by iSNAP with P1 and P2 scores 54.3 and 31.2 respectively. iSNAP analog screening using compound **48** as seed structure revealed 3 new putative analog matches; (1) m/z +550.50 at P1=38.5, P2=25.9; (2) +556.35 at P1=54.1, P2=39.7; (3) +600.38 at P1=38.2,P2=31.4 (Figure 16). These new putative analogue compounds are investigated in Chapter 5 as a follow-up study.

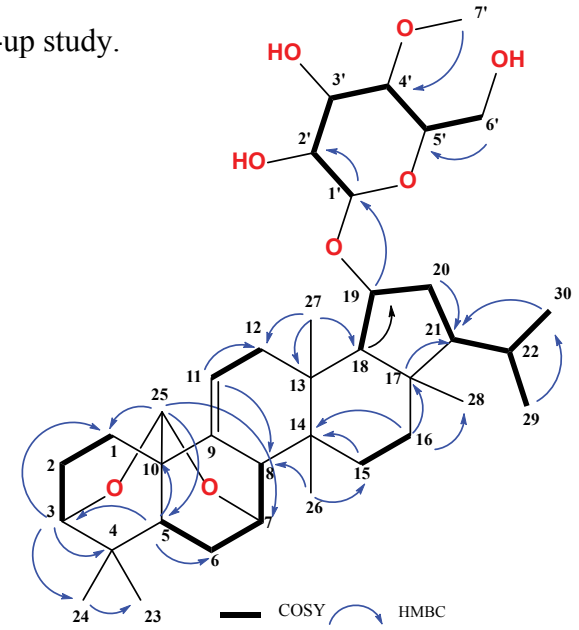


Figure 15. Grisefernaneoside A (**7**) HMBC and COSY connectivities.

Table 7. ^{1}H and ^{13}C NMR Spectroscopic data (700 MHz, $CDCl_3$) for griseofernaneoside A (**7**).

Position	$\delta_{C\ type}$	δ_{H} (*J* in Hz)	HMBC	NOESY
1a	22.3, CH2	1.54, overlap	2, 5, 10, 25	25
1b		1.91, overlap		
2a	23.1, CH2	1.75, overlap	1, 4	25
2b		1.90, overlap		
3	78.0, CH	3.47, m	1, 4, 5, 24	2a/b, 23, 24
4	35.6, C			
5	43.5, CH	1.41, dt (11.5, 2.7)	1, 4, 6, 10, 23, 25	
6a	27.2, CH2	1.74, overlap	5, 7	
6b		2.26, overlap		5', 6a,16, 23
7	71.3, CH	4.12, dd (4.7, 2.0)	5, 8, 9, 10, 14, 25	6a/b,8, 15a, 26
8	49.1, CH	2.47, brd	7, 9, 10, 11, 26	7, 25, 26, 27
9	139.8, C			
10	37.0, C			
11	117.0, CH	5.37, dt (5.0, 2.5)	8, 12	1a/b,12b, ,27
12a	37.5, CH2	1.79, overlap	9, 11, 13, 27	
12b		1.99, overlap		
13	38.0, C			
14	37.0, C			
15a	29.7, CH2	1.25, m	13, 14, 16, 17	7
15b		1.60, m		11
16a	35.8, CH2	1.50, m	15, 17, 18, 28	
16b		1.63, m		
17	43.9, C			
18	55.1, CH	1.80, overlap	13,14,17,19,20,21,27,28	12b
19	75.8, CH	4.35, ddd (10.1, 8.2, 3.9)	1', 13, 18, 21	27
20	36.6, CH2	1.72, overlap	17, 19, 21, 27, 28	
21	57.4, CH	1.30, q (9.7)	17, 19, 20, 22	28
22	30.5, CH	1.44, m	20, 21, 29, 30	
23	26.4, CH3	1.18, s	3, 4, 5, 6, 24	3, 6b, 24
24	30.7, CH3	0.94,s	3, 4, 5, 23	3, 23
25	97.0, CH	5.19, d (1.9)	1, 3, 5, 7, 9, 10	24
26	16.9, CH3	0.88, s	8, 13, 14, 15	8
27	18.0, CH3	0.95, s	12, 13, 14, 18	19
28	16.0, CH3	0.80, s	16, 17, 18, 21	19, 22
29	22.1, CH3	0.89, d (6.5)	21, 22, 30,	
30	23.0,CH3	0.83, d (6.5)	21, 22, 29	18, 19, 21
4'-OMe-glucose				
1'	97.7, CH	4.53, d (1.0)	2', 19	2', 20
2'	71.7, CH	3.89, overlap	1', 3', 4'	
3'	74.6, CH	3.62, dd (10.0, 6.6)	4'	
4'	78.1, CH	3.41, t (9.4)	3', 5', 6', 7'	
5'	75.0, CH	3.19, ddd (9.7, 4.5,3.0)	3', 4', 6'	
6'a	62.5, CH2	3.77, dd (11.9, 4.5)	4', 5'	1'
6'b		3.90, overlap	4',5'	
7'	61.1, CH3	3.61, s	4'	

Click labels to sort

Scan No.	RT	Precursor M/Z	Precursor Charge	Precursor Mass	Result	SMILES	Mass	Matched Peaks	P1	P2
2168	3.9	584.38	1	583.37	user_Cyclic pentapeptide 1	O=C(NC(C(NC(C(NC(C(N1C2CCC1)=O)CC(C)C)=O)C(C)CC)=O)C(C)C)C(CC3=CC=CC=C3)N(C)C2=O	583.37	14	54.3	31.2
2052	3.69	556.35	1	555.34	user_Cyclic pentapeptide 1_analog_1	O=C1NC(C(=O)N3CCCC3C(=O)N(C)C(C(=O)NC(C(=O)NC1C(C)CC)C(C)C)CC2=CC=CC=C2))CC(C)C	583.37	11	54.1	39.7
2169	3.9	550.4	1	549.39	user_Cyclic pentapeptide 1_analog_1	O=C1NC(C(=O)N3CCCC3C(=O)N(C)C(C(=O)NC(C(=O)NC1C(C)CC)C(C)C)CC2=CC=CC=C2))CC(C)C	583.37	14	38.5	25.9
1966	3.53	600.38	1	599.37	user_Cyclic pentapeptide 1_analog_1	O=C1NC(C(=O)N3CCCC3C(=O)N(C)C(C(=O)NC(C(=O)NC1C(C)CC)C(C)C)CC2=CC=CC=C2))CC(C)C	583.37	14	38.2	31.4

Figure 16. Snap-shot of iSNAP's dereplication and analog screening results for compound **48**.

Fungal Endophyte extracts E-038 and E-225

In total, 1 previously undescribed compound, 1 new natural product and 7 known compounds were discovered from the culture filtrates of *Rhizosphaera.sp*. (E-038) and *Phyllosticta cf. pyrolae* (E-225), endophytic fungi from the leaves and stems of a lowbush blueberries isolated in the Acadian forest of Nova Scotia, Canada. Extracts E-038 and E-225 were investigated as they were outlier extracts in the PCA (ESI-) Score Plots, located on the far left of the X-axis (Appendix A, S7A). Compound **12** eluted next to compound **8**, and was identified as new natural product, a derivative of the phytotoxic agent guignardic acid **24**.[38] Additionally, known compounds **22-24, 39** and related guignardic acid derivatives **45-47** were discovered from these three extracts (Appendix C, S13-18, 39-40, 49-54C).[38,54–56]

Compound **8** was isolated as a golden oil. Its HRESIMS afforded a protonated molecular ion at m/z +361.2369, giving a molecular formula of $C_{22}H_{32}O_4$ with 7 double bond equivalents. The structure was elucidated by 1D and 2D NMR as a new 19-O-Methyl derivative of tricycloalternarene **9b** (Table 8, Figure 17, Appendix B, S61-69B).[36] In order to determine the relative stereochemical assignments of the 19-O-Methyl moiety, preparation of its 4-bromobenzoate derivative was undertaken and successfully resulted in compound **14** (Appendix B, S107-111B). The derivatization method was adapted from Burgess *et al.*, which was recently reported.[21] HMBC and NOEY correlations further confirmed the 19-O-methyl position and protons H15 and H17 are in a *syn* orientation relative to each other.

Compound **12** was isolated as a golden oil. The HRESIMS afforded a protonated molecular ion at m/z +247.0618, giving a molecular formula of $C_{13}H_{10}O_5$ with 9 double bond equivalents. The structure was elucidated by 1D and 2D NMR and was previously reported as a

semisynthetic derivative (Table 10, Figure 19, Appendix B, S94-99B).[56] This is the first report of the compound as a new natural product.

Table 8. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of (**8**) and (**14**) in C_6D_6.

	(**8**)			(**14**)		
Position	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC
1	25.9, CH3	1.65, d (1.5)	2, 3, 4, 22	25.7, CH3	1.67, d (1.4)	
2	131.4, C			131.5, C		
3	123.6, CH	5.13, tp (7.2, 1.5)	2, 4, 5	123.8, CH	5.15,ddt (8.7, 5.7, 1.4)	
4	26.9, CH2	2.67, q (7.9)	2, 3, 5, 6	27.1, CH2	2.69, ddq (22.9, 15.0, 7.6)	
5	127.5, CH	5.33, td (7.4, 1.5)	3, 4, 7, 21	127.7, CH	5.34, td (7.4, 1.6)	
6	134.1, C			134.0, C		
7	42.0, CH	2.88, ddd (12.0, 10.5, 6.5)	5, 6, 8, 11,	42.3, CH	2.89, dt (11.8, 9.0)	
8a	25.6, CH2	1.41, overlap	7, 9, 10, 11	25.5, CH2	1.34, overlap	
8b		1.71, m			1.67, overlap	
9a	37.7,CH2	1.42, overlap	7, 8, 10, 11	37.8, CH2	1.27, m	
9b		2.01, m			1.84, ddd (13.6, 9.2, 3.8)	
10	87.0, C			87.3, C		
11	43.6, CH	1.59, ddd (11.8, 6.7, 1.1)	6, 7, 12, 13	43.4, CH	1.54, overlap	
12a	16.2, CH2	2.11, m	7, 11, 13, 17, 18	17.0, CH2	2.16, ddd (17.8, 6.7, 1.7)	7, 11,10,13,14,18
12b		2.54, dt (17.6, 1.3)			2.52, dt (17.8, 1.5)	
13	106.1, C			108.9, C		
14	167.8, C			163.1, C		
15	66.2, CH	4.12, t (4.7)	13, 14,16, 17	67.1, CH	5.96, ddd (7.5, 4.6, 2.9)	13,14,16,17,23,24
16a	34.9, CH2	1.96, ddd (14.0, 4.6, 3.5)	14, 15, 17, 18	33.6, CH2	2.22, m	14, 15, 17, 18
16b		2.13, m				
17	79.9, CH	3.47, dd (5.8, 3.4)	13, 15, 16, 18,19	77.8, CH	3.42, dd (8.7, 5.3)	15, 16, 18, 19
18	193.9, C			194.8, C		
19	58.4, CH3	3.25, s	17	58.9, CH3	3.48, s	17
20	22.3, CH3	1.04, s	9, 10, 11	22.5, CH3	0.92, s	
21	17.3, CH3	1.50, overlap	5, 6, 7	17.8, CH3	1.55, d (1.3)	
22	18.2, CH3	1.50, overlap	1, 2, 3	18.8, CH3	1.49, q (1.2)	
23				165.0, C		
24				128.4, C		
25				131.6, CH	7.83, m	
26				132.0, CH	7.16, m	
27				129.5, CBr		
28				132.0, CH	7.16, m	
29				131.6, CH	7.83,m	

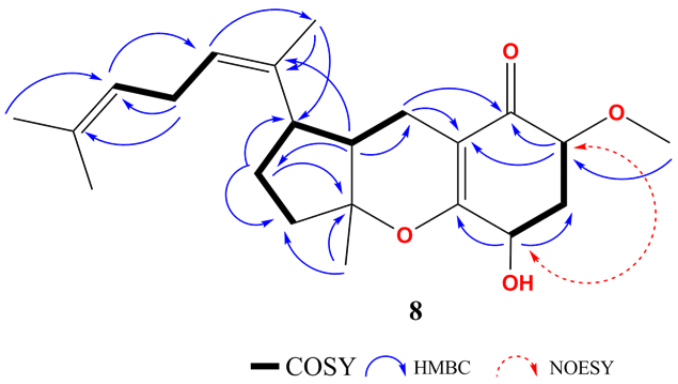


Figure 17. Compound **8** HMBC and COSY connectivities with NOESY from **14**.

Fungal Endophyte extracts E-260 and E-252

In total, one previously undescribed compound and 3 known compounds were discovered from the culture filtrates of the endophyte *Sphaerulina sp.*, in addition to one previously undescribed compound and 1 known compound from an undescribed isolate. The endophytic fungi were isolated from the leaves and stems of lowbush blueberries isolated in the Acadian forest of Nova Scotia, Canada. Extract E-260 was identified as a distinct isolate and was the most divergent observation on the left X-axis in the PCA (ESI+) Score plot (Appendix A, S8A). The most significant molecular feature from an analysis of the Scores Contribution plot data was m/z +309.05 at R_t = 9.8 min. LC-MS analysis and targeted HPLC isolation and down-stream methanol crystallization, identified the contributor as fulvic acid, compound **36**, by single crystal X-ray analysis and 1D and 2D NMR analysis. Dereplication screening and comparative analysis of the deconvoluted LC-MS spectra of extract E-260 identified fulvic acid derivatives **37**-**38** and compound **9** (Appendix C, S37-40C).[57] Similarly, the antibiotic canescin A, compound **40**, was identified with compound **10** (Appendix C, S41-42C).[37,57]

Compound **9** was isolated as a light beige solid. The HRESIMS afforded the protonated molecular ion of m/z 281.1021, giving a molecular formula of $C_{14}H_{16}O_6$, with 7 degrees of unsaturation. The structure was elucidated by 1D and 2D NMR as a previously undescribed compound, 7-hydroxy-3-(hydroxymethyl)-2-(2-hydroxypropyl)-6-methoxy-4H-chromen-4-one, a derivative of fulvic acid (Table 9, Figure 18, Appendix B, S70-77B) .

Compound **10** was isolated by LC-SPE-NMR with/without formic acid. The HRESIMS afforded a protonated molecular at m/z = 309.0965, giving a molecular formula of $C_{15}H_{16}O_7$, with 8 degrees of unsaturation. The structure was elucidated by 1D and 2D NMR and identified

as a previously undescribed carboxylic acid derivative compound **40** (Table 9, Figure 18, S78-84B).

Table 9. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of (**9**) in CD_3OD and (**10**) in CD_3CN.

	(**9**)			(**10**)		
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
1	167.6, C			167.7, C		
2	121.4, C					
3	178.6, C			155.0, C		
4	116.7, C			104.9, CH	6.29, s	3, 4a, 5, 8a, 9
4a				138.5, C		
5	105.3, CH	7.49, s	3,4,6,7,8,9	102.7, CH	6.36, s	1, 4, 7, 8, 8a
6	148.4, C			161.8, C		
7	154.9, C			114.5, C		
8	103.8, CH	6.89, s	3,4,5,6,7,9	163.7, C		
8a				99.8, C		
9	154.1, C			19.3, CH3	2.21, s	3, 4, 4a
10a	42.2, CH2	3.01, dd (14.2, 8.3)	1, 2, 11, 12	19.4, CH2	2.72, m	6, 7, 8, 11a/b, 12
10b		2.90, dd (14.2, 4.5)				
11	66.8, CH	4.24, dqd (8.3, 6.3, 4.4)				
11a				32.0, CH2	1.93, overlap	7, 9, 12, 13
11b					1.82, dtd (13.9, 8.7, 5.6)	
12	23.5, CH3	1.31, d (6.3)	10, 11	80.8, CH	3.75, dd (7.9, 4.1)	9, 11a/b, 13, 14
13	55.3, CH2	4.64, m	1, 2, 3	174.1, C		
14	56.6, OCH3	3.95, s	6	58.3, OCH3	3.36, s	12

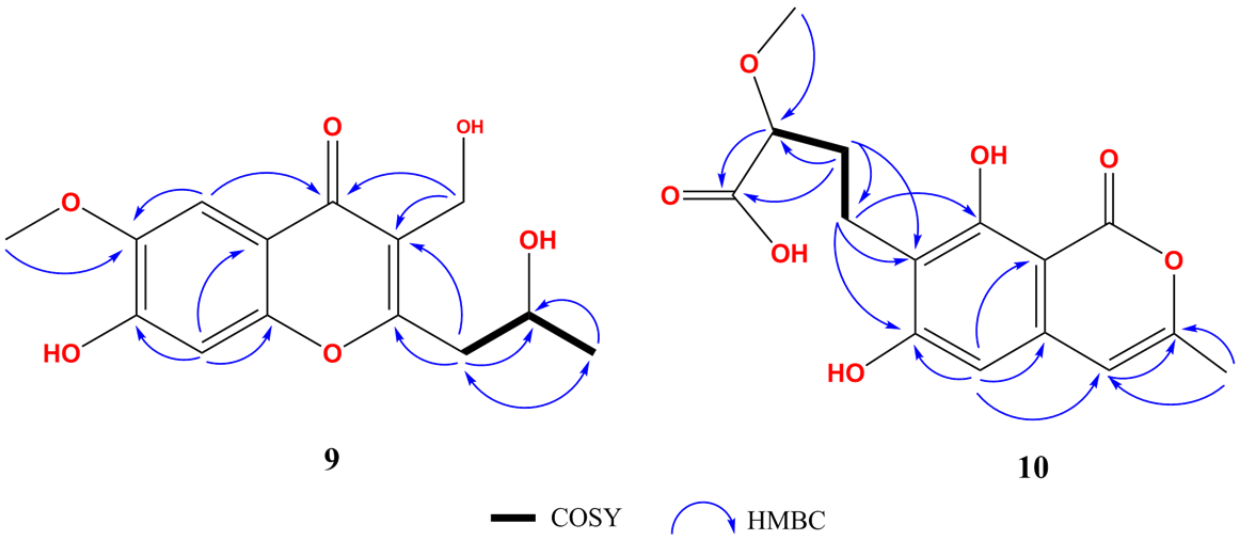


Figure 18. Compound **9** and **10** HMBC and COSY connectivities

Fungal Endophyte extracts E-195

In total, 1 previously undescribed compound and 1 known compound were discovered from the culture filtrates of *Nemania serpens,* endophytic fungi from the leaves and stems of potato plant located at Jordan Research Farm (Agruculture and Agr-food Canada) in Ontario, Canada. Extract E-195 was a divergent outlier in the PCA (ESI+) Score plot, to the center left in X and Y-axis (Appendix A, S9A). The top two significant molecular features were targeted from the Scores Contribution plot for isolation; m/z +195.1 at R=9.2 min, m/z +155.1 at Rt=9.8 min. Unfortunately, the yields from several fermentations were not fruitful. However, one significant feature, m/z +457.2 at Rt = 11.4, overlapped with a major UV abundant peak in the LC-MS chromatogram. We investigated this further by semi-preparative HPLC isolation of the major UV peak and discovered that compound **11** and known compound **28**, 6-hydroxymethyleugenin, co-eluted with the target mass.[58] Compounds **11** and **28** were isolated as a co-eluting mixture of relatively high purity, but distinguishable by 1D and 2D NMR, indicating that the target compound at m/z = +457.2 was likely a highly ionizable species or an MS artifact. Compounds **11** and **28** were separated using a second HPLC purification step, with **28** being fully resolved. However, **11** was not recovered. HRESIMS of **11** afforded a protonated molecular ion of 175.1329, giving a molecular formula of $C_9H_{18}O_3$, with 1 degree of unsaturation. Examination of the high quality 1D and 2D NMR spectra of co-eluting **11** and **28**, and resolved **28,** identified compound **11** as a previously unreported 2-(1-hydroxypropan-2-yl) hexanoic acid (Table 10, Figure 19, Appendix S85-93B).

Table 10. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of (**11**) in CD_3OD and (**12**) in CD_3CN.

	(**11**)			(**12**)		
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
1a	66.2, CH2	3.59, dd (10.9, 5.3)	2, 3, 8			
1b		3.43, dd (10.9, 6.4)				
2	39.3, CH	1.79, qd (6.7, 5.2)	1, 3, 4, 8, 9	105.1, C		
3	50.4, CH	2.30, ddd (10.7, 6.8, 4.2)	1, 2, 4, 8, 9			
4a	30.9, CH2	1.61, dtd (12.9, 10.3, 4.4)	2, 3, 5, 6, 9	136.4, C		
4b		1.49, m				
5a	31.4, CH2	1.31, overlap	3, 4, 6, 7	162.7, C		
5b		1.26, overlap				
6	23.8, CH2	1.32, overlap	4, 5, 7	166.1, C		
7	14.4, CH3	0.91, t (7.0)	5, 6	21.4, CH3	1.48, s	2, 6
8	15.2, CH3	0.96, d (6.9)	1, 2, 3	109.9, CH	6.61, s	4, 5, 2'
9	180.8, C			52.9, OCH3	2.99, s	6
1'				132.8, C		
2'				130.2, CH	7.51, d (7.0)	8, 3', 4'
3'				128.9, CH	7.06, t (7.7)	1', 2'
4'				129.1, CH	6.98, t (7.4)	3'

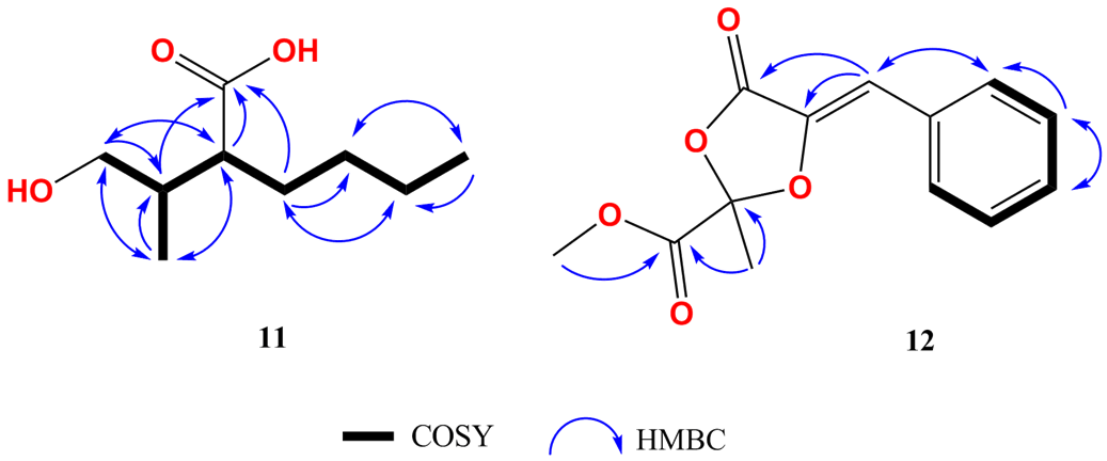


Figure 19. Compound **11** and **12** HMBC and COSY connectivities

Fungal Endophyte Outliers – Known compounds

The extracts listed in Figure 20 were investigated as modest outliers within the ESI (+/-) PCA Scores plots. In total 14 known compounds where discovered, many of which are known bioactive agents.[59–68] Compounds were isolated by LC-SPE-NMR, semi-preparative HPLC and/or by crystallization. Their structures were elucidated by a combination of 1D and 2D NMR, HRESIMS measurements and by single crystal X-ray analysis. Supporting information can be found in Appendix C.

#	Extract	ID #	Name	[M+H]+	Rt (min)	Species	Isolation
1	E-261	15	5,6-E-Dihydro-5-hydroxy-6-propenyl-2H-pyran-2-on	155.07	7.2	*Nigrospora sphaerica*	HPLC
2	E-035	16	RKB 3564S	165.09	7.4	*Creosphaeria sassafras*	LC-SPE-NMR
3	E-101	20	Nectriapyrone	195.23	15.1	*cf. Leptodontidium*	LC-SPE-NMR
4	E-182	21	dihydronaphthalene-2,6-dione	205.09	10	*cf. Paraphaeosphaeria neglecta*	HPLC
5	E-006	25	Lachnochromonin A	263.13	11.9	*Nigrospora cf. sphaerica*	LC-SPE-NMR
6	E-006	26	Lachnochromonin B	277.14	16.4	*Nigrospora cf. sphaerica*	LC-SPE-NMR
7	E-006	27	Lachnochromonin C	247.13	10.6	*Nigrospora cf. sphaerica*	LC-SPE-NMR
8	E-140	29	Fruit rot toxin A, FRT-A	238.11	11.5	*Nemania serpens*	HPLC
9	E-140	30	Fruit rot toxin A, FRT-B	238.11	11.6	*Nemania serpens*	HPLC
10	E-261	32	Abscisic Acid	265.14	12.1	*Nigrospora sphaerica*	HPLC
11	E-051	33	Ascochitine	277.11	18.5	*Coniochaeta tetraspora*	HPLC/Crystalization
12	E-169	34	Alternariol monomethylether	273.08	12.9	*Alternaria sp.*	LC-SPE-NMR
13	E-046	35	Sclerodine	313.11	17.7	*Godronia cf. cassandrae*	Crystalization
14	E-223	43	16-α-D-Glucopyranosyloxyisopimar-7-en-19-oic acid	483.29	14.9	*Nemania serpens*	HPLC

Figure 20. Extract and compound identification results for isolated knowns.

Preliminary Bioassay Screening

Compounds **1** and **3-7** were screened for biological activity against six species of microorganisms in accordance with the Clinical Laboratory Standards Institute (CLSI) protocols.[69] The microorganism included *E. coli* BW25113 ΔbamBΔtolC, *Staphylococcus aureus* ATCC#29213, *Bacillus subtilis* 1A1, *Micrococcus luteus*, *Saccharomyces cerevisiae* B4741, and *Candida albicans* ATCC# 90028) (Appendix B, S112B). [70,71]

Compound **1** showed weak inhibition against *C. albicans* (>25% inhibition) at a concentration of 200 μg/mL, but demonstrated no biological activity against the tested strains. Compound **3** showed weak inhibition only against *S. cerevisiae* (>25% inhibition) at a concentration of 100 μg/mL. Griseofernaneoside A (**7**) showed strong activity against *S. cerevisiae* with a minimum inhibitory concentration (MIC) of 25 μg/mL, and was modestly active against *C. albicans* with a MIC of 50 μg/mL. Triterpenoids are well understood to exhibit a diverse spectrum of biological activities, including cytotoxicity against tumor cell lines.[51] However to our knowledge, there is no cytotoxicity data reported for fernane-type triterpenoids. Griseofernaneoside A (**7**) was further tested in the NCI-60 human tumor cell panel screen and showed weak inhibition (>30% inhibition) against leukemia MOLT-4, renal UO-31and prostate PC-3 cell lines at a concentration of 10 μg/mL (Appendix B, S113B).[72] Compounds **4-6** showed no activity against the test organism.

CONCLUSION

In this study, we have shown that secondary metabolomic based analytical methodologies encompassing LC-HRMS/MS analysis and multivariate statistical methods was effective in profiling the chemical space of a diverse collection of understudied fungal endophytes from fruiting plants and revealing distinct chemical entities. An in-house library of deconvoluted LC-HRMS spectra was established, allowing for the rapid dereplication, chemometric sorting and prioritization of undescribed endophyte extracts, independent of ITS DNA sequencing or bioactivity screening. Importantly, there are no commercially available databases of LC-MS spectra from fungal endophytes, especially those from fruiting plants. This approach proved fruitful in guiding the discovery of our previously reported new diplosporin and agistatine derivatives, including the novel antifungal compound trienylfuranol A and trienylfuranones A-B.[19,21] Comprehensive metabolomic-guided profiling of the fungal endophyte collection using principal component analysis (PCA) identified a number of unique outlier extracts and also revealed clustered groupings which were prioritized for analysis. This untargeted metabolomic strategy guided the discovery of 12 new and 37 known specialized metabolites independent of genomic data or bioactivity-guided fractionation. Significantly, the majority of these compounds are bioactive agents. Dereplication screening of the outlier extracts and cluster groupings using iSNAP metabolomic tools identified a known nonribosomal peptide and several new putative analogues. This approach has demonstrated the utility of combining iSNAP informatic search algorithms with multivariate statistical methods for the untargeted discover of new natural products from a diverse fungal endophyte collection.

EXPERIMENTAL

General Procedures

NMR experiments were performed using 3 mm NMR tubes (Wilmad 335-PP-7) on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm QNP cryoprobe (700.17 MHz for ^{1}H and 176.08 MHz for ^{13}C), with chemical shifts referenced to the residual solvent signals.[73] NMR data processing was done using MNOVA v11.0.1 by Mestrelab Research. LC-MS analysis and HRESIMS measurements were performed using two LC-MS/MS systems; (1) Bruker MaXis 4G UHR-qTOF mass spectrometer, operated under Hystar 3.2 control, coupled to a Dionex Ultimate 3000 HPLC system, operating in the positive and negative electrospray ionization mode (ESI) with calibrations done using HCO_2Na (m/z range 100-1200); (2) Thermo Scientific Q-Exactive Quadrupole Orbitrap mass spectrometer coupled to a Agilent 1290 UHPLC system, operating in positive ESI (m/z range 100-1000) with automated MS/MS acquisition. Bruker LC-MS chromatography was carried out using a Supleco Ascentis Express C18 reversed-phase core-shell column (150 × 4.6 mm, 2.7 μm) with a mobile phase consisting of a linear gradient of mobile phase A (H_2O with 0.1% (v/v) HCO_2H) in B (CH_3CN with 0.1% (v/v) HCO_2H); 95% (v/v) A in B from 0 to 2.5 min, 95–5% A in B from 2.5 to 25 min, 5-0% A in B from 25 to 25.1 min, 100% B from 25.1 to 30 min, 0-95% A in B from 30 to 32 min, and 95% A in B from 32 to 40 min at a flow rate of 0.75 mL/min. Thermo Scientific Q-Exactive LC-MS/MS chromatography was carried out on Agilent Eclipse Plus C18 column (2.1 × 50 mm, 1.8 μm) using an identical mobile phase composition with a gradient from 0 to 100% B over 6 min. LC-MS spectral libraries of deconvoluted LC–HRMS data were manually compiled using Bruker Compass Data Analysis 4.0 and Library Editors software, with spectra having intensity values above 1×10^5 included in the in-house dereplication library. Principle component analysis (PCA) was performed with SIMCA-P+ (v12.0, Umetrics, Sweden). LC-

SPE isolations was adapted from Ibrahim *et al.* (2014) and performed on an Agilent 1100 series HPLC system coupled to a Bruker/Spark Prospekt II LC-SPE-NMR system running under Hystar control.[19] Semi-preparative HPLC isolations were performed on a Agilent 1100 series HPLC system with separations performed using a Phenomenex Synergi-Max RP C-12 column (250 × 10 mm,4 μm) or a Columnex Chromenta KAQ C-18 column (250 × 10 mm, 5 μm) maintained at 40 °C and operating at 5 mL/min.

Sampling and Fungal Strains

Plant materials' including leafs, twigs, stems, flowering fruit and soil samples from wild and cultivated grapes, raspberries, high and low bush blueberries and cranberries were sourced form Ontario and Nova Scotia Canada. Low and highbush blueberry leafs and stems were isolated from fields located in Simcoe and Jordan Research Farm (Agriculture and Agri-Food Canada), Jordan Station, Ontario, Canada (GPS =43.17534° N/79.35905°W), and from fields in Debert, Portapique, Rawdon and Mt. Thom, Nova Scotia, Canada (GPS = 45.49169° N/62.98954°W). Cranberry, grape (v*itis vinifera and labrusca),* and raspberry leaves, stems and vines were isolated from Jordan Research Farm (Agriculture and Agri-Food Canada). Samples were bagged or placed in 50 mL conical vials and stored on ice during transport. All materials were surface sterilized using a two-step process; (1) ethanol (90%), (2) bleach (65%), and then short segments were cut and excised prior to transferring to malt (MLT) and potato dextrose agar plates (PDB). Identification and DNA extraction was done using a UltraClean Microbial DNA Kit (MoBio, Carlsbad, CA) with ITS1 and ITS4 PCR primers used for amplification as previously described by Sumarah *et al.*, (2008).[14]

Fermentation and extraction

The collection of 263 fungal endophyte strains were first isolated and grown on MLT and PDB media and then fermented in 1 L Roux bottles containing 200 mL of water and potato dextrose media (24 g/L). The Roux bottles were fermented stationary for 6-8 weeks at 25 °C, at which point they were harvested by vacuum filtration using Whatman # 4 filters and the cell-free filtrates extracted with equal volumes (1:1) of ethyl acetate solvent. The extracts were dried under reduced pressure and stored at -20 °C in 2 mL vials. The fungal filtrates were first processed by re-suspended each sample in 600 μL of 100% deuterated CD_3CN and sonicated for 10 min, then transferred to 1.5 mL eppendorf tubes and centrifuged for 15 min at 13,000 RPM. The centrifuged samples were filtered using a 13 mm Acro-disk (0.45 μm, GHP) and transferred to 2 mL vials (with fused-conical insert) prior to LC-MS and NMR analysis.

Metabolite isolation: Extract E-184

Metabolites **1**-**6**, **31** and **44** were isolated from the cell-free filtrates of *Xylaria cubensis* isolated from Riesling grapevines using LC-SPE-NMR and semi-preparative HPLC purification. In total 4 L was fermented in 2 batches of 10×1 L Roux bottles containing 200 mL of PDB broth and extracted with ethyl acetate. LC-SPE-NMR was used to isolate metabolites **1**, **3** and **6** using a method adapted from Ibrahim *et al.* (2014).[19] However, for bioactivity screening more material was needed and two-step semi-preparative preparative HPLC purification processed was performed. In a first-pass, crude HPLC fractions were first obtained using a Phenomenex Synergi-Max reversed-phase C-12 column (250 × 10 mm, 4 μm) operating at 5 mL/min and 40 °C. Mobile phase composition was a linear gradient of A (H2O with 0.1% (v/v) HCO2H) in B (CH3CN with 0.1% (v/v) HCO2H): 95% (v/v) of A in B from 0 to 3 min, 95–80% of A in B from 3 to 10 min, 80% of A in B from 12 to 15 min, 80-60% A

in B from 15 to 20 min, 60% of A in B from 15 to 23 min, 60-25% A in B from 23 to 34 min, 100% B from 34 to 37 min, 0-95% A in B from 37 to 38 min and 95% A in B from 38 to 45 min. Compound **1** eluted at 15.1min, compound **2** eluted at 22.3 min, compound **3** eluted at 27.1 min, compound **4** (4 mg, 1 mg/L) eluted at 21.3 min, compound **5** eluted at 21.9 min, compound **6** eluted at 23.6 min, compound **31** eluted at 26.8 min and compound **34** (4 mg, 0.5 mg/ L) eluted at 12.8 min. Compound **31** was crystalized out of methanol by slow evaporation in a 3mm NMR tube. Single crystal X-ray analysis confirmed the structure (Appendix C, S25C).

Second-pass semi-preparative HPLC purifications for compounds **1-3** and **5-6** were done using a Columnex Chromenta KAQ C-18 column (250 × 10 mm, 5 μm) operating at 5 mL/min and 40 °C. Compound **1** (5.2 mg, 1.3 mg/L) eluted at 15.6 min ; 95% (v/v) of A in B from 0 to 3 min, 95–80% of A in B from 3 to 12 min, 80% of A in B from 12 to 18 min, 80-0% A in B from 18 to 20 min. Compound **2** (1.5 mg, 0.38 mg/L) eluted at 19.2 min ; 95% (v/v) of A in B from 0 to 3 min, 95–74% of A in B from 3 to 10 min, 74% of A in B from 10 to 20 min, 74-0% A in B from 20 to 21 min. Compound **3** (16 mg, 4.0 mg/L) eluted at 20.6 min; 95% (v/v) of A in B from 0 to 3 min, 95–65% of A in B from 3 to 8 min,65% of A in B from 8 to 12 min, 65-60% A in B from 12 to 18 min, 60% of A in B from 18 to 25 min. Compound **5** (2.5 mg, 0.63 mg/L) eluted at 18.8 min; 95% (v/v) of A in B from 0 to 3 min, 95–80% of A in B from 3 to 8 min, 80% of A in B from 8 to 22 min. Compound **6** (9 mg, 1.13 mg/L) eluted at 20.8 min; 95% (v/v) of A in B from 0 to 3 min, 95–75% of A in B from 3 to 11 min, 75% of A in B from 11 to 22 min.

Metabolite isolation: Extract E-112

Metabolites **7**, **41**, **48-51** were isolated from the cell-free filtrates of and undescribed *Xylaria sp.* isolated from low and highbush blueberries using LC-SPE-NMR and semi-preparative HPLC purification. LC-SPE-NMR was used to isolate metabolites **41** using a method adapted from Ibrahim et al. (2014). Compounds **48-51** were isolated from a 1 L PDB culture, while compound **7** was isolated from several 2-4 L pooled fermentation batches totaling 40 L grown in PDB media. Metabolites were isolated using a Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm). Mobile phase composition as above: 95% (v/v) of A in B from 0 to 3 min, 95–70% of A in B from 3 to 16 min, 70% of A in B from 16 to 20 min, and 70-15% A in B from 20 to 37 min with fractions collected every 20s. Griseofernaneoside A (**7**) (5 mg, 0.13 mg/L) eluted at 36.8 min, compound **48** (4 mg/L) eluted at 34.9 min, compound **49** (2 mg/L) eluted at 33.9 min, compound **50** (2 mg/L) eluted at 32.5 min and compound **51** (2.5 mg/L) eluted at 30.2 min. Griseofernaneoside A was crystalized in methanol by slow evaporation in a 3mm NMR tube, yielding very fine colorless needles.

Metabolite isolation: Extract E-038 and E-225

Metabolites **8**, **12**, **22-24, 39 and 45-47** were isolated from cell-free filtrates by semi-preparative HPLC purification. The filtrates from 1 L fermentations for E-038 and E-225 were pooled (2 L total) and isolated using a Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm). Mobile phase composition as above: 95% (v/v) of A in B from 0 to 3 min, 95–5% of A in B from 3 to 29 min, 5 % of A in B from 29 to 30 min, 0% A in B from 30 to 34 min. Compound **8** (9 mg, 4.5 mg/L) eluted at 24.4 min, compound **12** (1.5 mg, 0.75 mg/L) eluted at 23.4 min, compound **22** (1.3 mg, 0.65 mg/L) eluted at 10.4 min, compound **23** (3.5 mg, 1.75 mg/L) eluted at 14.6 min, compound **24** (24 mg, 12 mg/L) eluted at 22.3 min,

compound **39** (5.5 mg, 2.75 mg/L) eluted at 18.7 min, compound **45** (1 mg, 0.5 mg/L) eluted 25.2 min, compound **46** (4 mg, 2 mg/L) eluted at 26 min, compound **47** (3 mg, 1.5 mg/L) eluted at 27 min.

Metabolite isolation: Extract E-260 and E-252

Compounds **9** and **36-38** were isolated from the cell-free filtrates by semi-preparative HPLC purification while compound **10** was isolated by LC-SPE-NMR from the original screening cultures. E-260 was re-cultured in a 2 L fermentation batch consisting of 10 × 1L Roux bottles grown in PDB broth media. HPLC isolations were performed using a Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm). Mobile phase composition and method as above: 95% (v/v) of A in B from 0 to 3 min, 95–5% of A in B from 3 to 29 min, 5 % of A in B from 29 to 30 min, 0% A in B from 30 to 34 min. Compound **9** (3.5 mg, 1.75 mg/L) eluted at 11.8 min, compound **36** (11 mg, 5.5 mg/L) eluted at 14.5 min, compound **37** (1.5 mg, 0.75 mg/L) eluted at 10.5 min, compound **38** (1.5 mg, 0.75 mg/L) eluted 12.9 min. Compound **36** was crystalized from methanol by slow evaporation in a 20 mL scintillation vial.

Metabolite isolation: Extract E-195

Compounds **11** and **28** were isolated from a 1 L filtrate culture by a two-step semi-preparative HPLC purification. First –pass isolations were performed using a Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm). Mobile phase composition and method as above: 95% (v/v) of A in B from 0 to 3 min, 95–5% of A in B from 3 to 29 min, 5 % of A in B from 29 to 30 min, 0% A in B from 30 to 34 min. Compounds **11** and **28** co-eluted at 15.2 min. A second-pass isolation was performed a Columnex Chromenta KAQ C-18 column (250 × 10 mm, 5 μm) operating at 5 mL/min and 40 °C. Mobile phase composition and method as above:

95% (v/v) of A in B from 0 to 3 min, 95–40% of A in B from 3 to 25 min, with compound **28** (2.5 mg) eluting at 13.4 min.

Metabolite isolation of known compounds

Metabolites **15**, **21**, **29-30**, **32, 35, 43** and **51** were isolated by semi-preparative HPLC from the 1 L fermentation cultures grown in PDB broth media (5 x1 L Roux bottles). The isolations were performed on Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm) with methods adapted using the their LC-MS retention times and mobile phase composition. Metabolites **16**, **20**, **25-27**, and **34** and were isolated by LC-SPE-NMR with isolation methods adapted from Ibrahim et al. (2014) in a similar fashion (Figure 20). Compounds **33** and **35** were also crystalized from menthol in 20 mL scintillation vials, with structures validated by single crystal X-ray analysis.

Bioassays

Minimum inhibitory concentrations (MIC) were carried out accordance with the Clinical Laboratory Standards Institute (CLSI, formerly NCCLS) protocols M7-A5 and M27-A. Compounds **1**, **3-7** were tested at a maximum concentration of 50-200 μg/mL in 96-well liquid culture format. Stock solutions were made to 5, 10 and 20 mg/mL in DMSO. Test strains included *E. coli* BW25113 ΔbamBΔtolC, a membrane and efflux pump compromised strain, *Staphylococcus aureus* ATCC#29213, *Bacillus subtilis* 1A1, *Micrococcus luteus*, *Saccharomyces cerevisia*e B4741, and *Candida albicans* ATCC# 90028. A cut off of 75% growth was used for inhibition, with the trends across the dilutions considered (Appendix B, S112B).

References

1. Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325,** 161–5 (2009).
2. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661 (2016).
3. Cragg, G. M. & Newman, D. J. Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* **1830,** 3670–3695 (2013).
4. Strobel, G. & Daisy, B. Bioprospecting for microbial endophytes and their natural products. *Microbiol. Mol. Biol. Rev.* **67,** 491–502 (2003).
5. Nielsen, K. F. *et al.* Dereplication of microbial natural products by LC-DAD-TOFMS. *J. Nat. Prod.* **74,** 2338–2348 (2011).
6. Gaudêncio, S. P. & Pereira, F. Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **32,** 779–810 (2015).
7. Ibrahim, A. *et al.* Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19196–201 (2012).
8. Yang, L. *et al.* Exploration of Nonibosomal peptide families with an automated informatic search algorithm. *Chem. Biol.* **22,** 1259–1269 (2015).
9. Strobel, G. A. Endophytes as sources of bioactive products. *Microbes Infect.* **5,** 535–544 (2003).
10. Floros, D. J., Jensen, P. R., Dorrestein, P. C. & Koyama, N. A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* **12,** 145 (2016).
11. Krug, D., Müller, R. & Rolf, M. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **31,** 768–83 (2014).
12. Zhang, H. W., Song, Y. C. & Tan, R. X. Biology and chemistry of endophytes. *Nat. Prod. Rep.* **23,** 753–771 (2006).
13. Miller, J. D., Cherid, H., Sumarah, M. W. & Adams, G. W. Horizontal transmission of the Picea glauca foliar endophyte Phialocephala scopiformis CBS 120377. *Fungal Ecol.* **2,** 98–101 (2009).
14. Sumarah, M. W., Puniani, E., Blackwell, B. A. & Miller, J. D. Characterization of polyketide metabolites from foliar endophytes of Picea glauca. *J. Nat. Prod.* **71,** 1393–8 (2008).
15. Sumarah, M. W., Puniani, E., Sørensen, D., Blackwell, B. a. & Miller, J. D. Secondary metabolites from anti-insect extracts of endophytic fungi isolated from Picea rubens. *Phytochemistry* **71,** 760–5 (2010).
16. Sumarah, M. W., Kesting, J. R., Sørensen, D. & Miller, J. D. Antifungal metabolites from fungal endophytes of Pinus strobus. *Phytochemistry* **72,** 1833–1837 (2011).
17. McMullin, D. R., Green, B. D., Prince, N. C., Tanney, J. B. & Miller, J. D. Natural Products of *Picea*

Endophytes from the Acadian Forest. *J. Nat. Prod.* **80,** 1475-1483 (2017).

18. Richardson S, . N. *et al.* Griseofulvin-producing Xylaria endophytes of *pinus strobus* and *vaccinium angustifolium*: evidence for a conifer-understory species endophyte ecology. *Fungal Ecol.* **11,** 107–113 (2014).
19. Ibrahim, A. *et al.* New diplosporin and agistatine derivatives produced by the fungal endophyte Xylaria sp. isolated from *Vitis labrusca*. *Phytochem. Lett.* **9,** 179–183 (2014).
20. Ibrahim, A. *et al.* Epoxynemanione A, nemanifuranones A–F, and nemanilactones A–C, from *Nemania serpens*, an endophytic fungus isolated from Riesling grapevines. *Phytochemistry* **140,** 16–26 (2017).
21. Burgess, K. M. N., Ibrahim, A., Sørensen, D. & Sumarah, M. W. Trienylfuranol A and trienylfuranone A–B: metabolites isolated from an endophytic fungus, *Hypoxylon submoniticulosum*, in the raspberry Rubus idaeus. *J. Antibiot.* **70,** 721–725 (2017).
22. Agriculture and Agri-Food Canada & Pest Management centre. *Crop Profile for Cranberry in Canada*. (2007).
23. Agriculture and Agri-Food Canada & Pest Management centre. *Crop Profile for Grape in Canada, 2013*. (2013).
24. Agriculture and Agri-Food Canada & Pest Management centre. *Crop Profile for Highbush Blueberry in Canada, 2014*. (2016).
25. Agriculture and Agri-Food Canada & Pest Management centre. *Crop Profile for Lowbush Blueberry in Canada, 2014*. (2017).
26. Agriculture and Agri-Food Canada & Pest Management centre. *Crop Profile for Rasberry in Canada, 2013*. (2016).
27. Krug, D. *et al.* Discovering the hidden secondary metabolome of Myxococcus xanthus: A study of intraspecific diversity. *Appl. Environ. Microbiol.* **74,** 3058–3068 (2008).
28. Hou, Y. *et al.* Microbial strain prioritization using metabolomics tools for the discovery of natural products. *Anal. Chem.* **84,** 4277–83 (2012).
29. Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org. Lett.* **14,** 5050–5053 (2012).
30. Kuhl, C., Tautenhahn, R. & Neumann, S. LC-MS peak annotation and identification with CAMERA. (Http://www.bioconductor.org/packages/bioc/vignettes/CAMERA), *October* 1–14 , (2010).
31. Smith, C. a *et al.* XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching,and identification. *ACS Publ.* **78,** 779–87 (2006).
32. Worley, B. & Powers, R. Multivariate analysis in netabolomics. *Curr. Metabolomics* **1,** 92–107 (2012).
33. Wilson, K. E. *et al.* Isolation and structure elucidation of coleophomones A and B, novel inhibitors of bacterial cell wall transglycosylase. *Tetrahedron Lett.* **41,** 8705–8709 (2000).
34. Hamed, A. I. *et al.* Unusual fernane and gammacerane glycosides from the aerial parts of *Spergula fallax*. *J. Nat. Prod.* **77,** 657–662 (2014).

35. Deyrup, S. T., Gloer, J. B., O'Donnell, K. & Wicklow, D. T. Kolokosides A-D: Triterpenoid glycosides from a Hawaiian isolate of Xylaria sp. *J. Nat. Prod.* **70,** 378–382 (2007).

36. Nussbaum, R. P., Günther, W., Heinze, S. & Liebermann, B. New tricycloalternarenes produced by the phytopathogenic fungus Alternaria alternata. *Phytochemistry* **52,** 593–599 (1999).

37. Lewis, C. N., Staunton, J. & Sunter, D. C. Biosynthesis of Canescin, a metabolite of *Aspergillus malignus*: Incorporation of a methionine, acetate, succinate, and isocoumarin precursors, labelled with deuterium and carbon-13. *J. Chem. Soc. Perkin Trans* **1,** 747–754 (1988).

38. Molitor, D. *et al.* Phenguignardic acid and guignardic acid, phytotoxic secondary metabolites from *Guignardia bidwellii*. *J. Nat. Prod.* **75,** 1265–1269 (2012).

39. Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* **10,** 770–781 (1999).

40. O'Callaghan, S. *et al.* PyMS: a Python toolkit for processing of gas chromatography-mass spectrometry (GC-MS) data. Application and comparative study of selected tools. *BMC Bioinformatics* **13,** 115 (2012).

41. Vaidyanathan, S. Profiling microbial metabolomes: what do we stand to gain? *Metabolomics* **1,** 17–28 (2005).

42. Guo, Y. *et al.* Tricyclic polyprenylated acylphloroglucinols from St John ' s Wort, *Hypericum perforatum*. *J. Nat. Prod.* **80,** 1493-1504 (2016).

43. Ciochina, R. & Grossman, R. B. Polycyclic polyprenylated acylphloroglucinols. *Chem. Rev.* **106,** 3963–3986 (2006).

44. Wang, J. *et al.* Isolation and identification of an endophytic fungus Pezicula sp. in *Forsythia viridissima* and its secondary metabolites. *World J. Microbiol. Biotechnol.* **30,** 2639–2644 (2014).

45. Hanumaiah, T. *et al.* Naphthalenes and naphthoquinones from Ventilago species. *Phytochemistry* **24,** 1811–1815 (1985).

46. Lu, C., Lin, X. & Shen, Y. A new dihydroisocoumarin from the strain Aspergillus sp. CMM, an endophytic fungus of Cephalotaxus mannii. *Chem. Nat. Compd.* **44,** 569–571 (2008).

47. Wu, W. *et al.* Isolation and structural elucidation of proline-containing cyclopentapeptides from an endolichenic Xylaria sp. *J. Nat. Prod.* **74,** 1303–1308 (2011).

48. Isaka, M., Palasarn, S., Sriklung, K. & Kocharin, K. Cyclohexadepsipeptides from the insect pathogenic fungus Hirsutella nivea BCC 2594. *J. Nat. Prod.* **68,** 1680–1682 (2005).

49. Edwards, R. L., Maitland, D. J. & Whalley, A. J. S. Metabolites of the higher fungi. Part 24. Cytochalasin N, O, P, Q and R. *J. Chem. Soc., Perkin Trans.* **1,** 57–65 (1989).

50. Wagenaar, M. M., Corwin, J., Strobel, G. & Clardy, J. Three new cytochalasins produced by an endophytic fungus in the genus Rhinocladiella. *J. Nat. Prod.* **63,** 1692–1695 (2000).

51. Hill, R. A. & Connolly, J. D. Triterpenoids. *Nat. Prod. Rep.* **30,** 1028–1065 (2013).

52. Luo, X. *et al.* Cytotoxic ent-kaurane diterpenoids from *Isodon rubescens var. lushiensis*. *J. Nat. Prod.* **73,** 1112–1116 (2010).

53. Gorman, J. a *et al.* Ascosteroside, a new antifungal agent from Ascotricha amphitricha. I. Taxonomy, fermentation and biological activities. *J. Antibiot.* **49,** 547–552 (1996).

54. Ayer, W., Browne, L., Feng, M., Orszanska, H. & Saeedi-Ghomi, H. The chemistry of the blue stain fungi. Part 1. Some metabolites of *Ceratocystis* species associated with mountain pine beetle infected lodgepole pine. *Can. J. Chem.* **64,** 904–909 (1986).

55. Martins, M. B. & Carvalho, I. Diketopiperazines: biological activity and synthesis. **63,** 9923–9932 (2007).

56. Andernach, L. *et al.* Assignment of configuration in a series of dioxolanone-type secondary metabolites from *Guignardia bidwellii* - A comparison of VCD and ECD spectroscopy. *European J. Org. Chem.* 5946–5951 (2013).

57. Nielsen, K. F. & Smedsgaard, J. Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography – UV – mass spectrometry methodology. *J. Chromatogr. A* **1002,** 111–136 (2003).

58. Feng, Y., Blunt, J. W., Cole, A. L. J. & Munro, M. H. G. The isolation of two new chromone derivatives from the New Zealand fungus *Tolypocladium extinguens*. *J. Nat. Prod.* **65,** 1681–1682 (2002).

59. Evans, R. H., Ellestad, G. A. & Kunstmann, M. P. Two new metabolites from an unidentified Nigrospora species. *Tetrahedron Lett.* **22,** 1791–1794 (1969).

60. Nair, M. S. R. & Carey, S. T. Metabolites of Pyrenomycetes II: Nectriapyrone, an antibiotic monoterpenoid. *Tetrahedron Lett.* **19,** 1655–1658 (1975).

61. Rukachaisirikul, V., Arunpanichlert, J., Sukpondma, Y., Phongpaichit, S. & Sakayaroj, J. Metabolites from the endophytic fungi *Botryosphaeria rhodina* PSU-M35 and PSU-M114. *Tetrahedron* **65,** 10590–10595 (2009).

62. Sassa, T. & Onuma, Y. Isolation and identification of fruit rot toxins from the fungus-caused Macrophoma fruit rot of Apple and Identification. *Agri.Biol.Chem.* **47,** 1155–1157 (1983).

63. Kikuzaki, H. *et al.* Abscisic acid related compounds and lignans in prunes (*Prunus domestica L.*) and their oxygen radical absorbance capacity (ORAC). *J. Agric. Food Chem.* **52,** 344–349 (2004).

64. Colombo, L., Gennari, C., Ricca, G. S., Scolastico, C. & U. Biosynthetic origin and revised structure of Ascochitine, a phytotoxic fungal metabolite. Incorporation of [1-13C]- and [1,2-13C2]-acetates and [Me-13C] methionine. *J. Chem. Soc. Perkin Trans.* **1,** 675–676 (1980).

65. Kelman, M. J. *et al.* Identification of six new Alternaria sulfoconjugated metabolites by high-resolution neutral loss filtering. *Rapid Commun. Mass Spectrom.* **29,** 1805–1810 (2015).

66. Ayer, W. & Hoyano, Y. Metabolites produced by the Scleroderris canker fungus, *Gremmeniella abietina*. Part 4. *Can. J. Chem.* **65,** 760–764 (1987).

67. Shiono, Y. *et al.* Isopimarane diterpene glycosides, apoptosis inducers, obtained from fruiting bodies of the ascomycete Xylaria polymorpha. *Phytochemistry* **70,** 935–9 (2009).

68. Chang, H. S. *et al.* Secondary metabolites of the endophytic fungus *Lachnum abnorme* from Ardisia cornudentata. *Int. J. Mol. Sci.* **17,** 1–12 (2016).

69. Martínez-Lozano Sinues, P. *et al.* Mass spectrometry fingerprinting coupled to National Institute of Standards and Technology Mass Spectral search algorithm for pattern recognition. *Anal. Chim. Acta* **755,** 28–36 (2012).

70. National Committee for Clinical Laboratory Standards. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically ; approved Standard — 5th Edition. *Wayne, PA Natl. Comm. Clin. Lab. Stand.* **NCCLS,** M7-A5 (2000).

71. National Committee for Clinical Laboratory Standards. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts ; Approved Standard. *Wayne, PA Natl. Comm. Clin. Lab. Stand.* **NCCLS,** M27–A (1997).

72. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6,** 813–823 (2006).

73. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.* **62,** 7512–7515 (1997).

CHAPTER FIVE

Metabolomic Guided Discovery of Cyclic Nonribosomal Peptides from *Xylaria elissi* - a new Griseofulvin-producing endophyte species from *Vaccinium angustifolium*.

CHAPTER FIVE PREFACE

Metabolomic Guided Discovery of Cyclic Nonribosomal Peptides from *Xylaria elissi* - a new Griseofulvin-producing endophyte species from *Vaccinium angustifolium*.

Ashraf Ibrahim, Fan Fei, Joey Tanney, Tim McDowell, Linda Ejim, Keith A. Seifert, Dan Sørensen, Alfredo Capretta, J.D. Miller, Brian E. McCarry, Mark W. Sumarah.

The work presented within this chapter is unpublished.

Author Contributions

B.E.M., A.I. and M.W.S., conceived the study.

B.E.M., A.I., M.W.S., J.D.M. and D.S., contributed to the study design.

A.I performed LC-MS/MS analysis, performed metabolomic analysis and analyzed data, performed HPLC isolation and purifications, performed structural elucidation (1D and 2D NMR, and HRMS). F.F performed metabolomic analysis, analyzed data, created figures and discussed results. J. T. and K.A.S. performed ITS DNA sequencing and fungal species identifications. T.D. cultured endophytes, performed fermentations and extracted the cultures L.E. performed antimicrobial testing of purified compounds. D.S. provided NMR expertise, performed structural validation, and discussed results. J.D.M provided scientific expertise on fungal endophytes, and discussed the results. A.C. provided NMR expertise, and discussed the results. A.I wrote the chapter and A.C., and M.W.S. edited the chapter.

ABSTRACT

Introduction: Fungal endophytes are emerging as new sources for novel bioactive compounds. New agents may play a role in the chemical ecology of host plants or in their defense mechanism. A new griseofulvin-producing *Xylaria* species has been discovered from high and lowbush blueberry plants. **Objective:** Identify secondary metabolite differences between high and lowbush griseofulvin-producing blueberry endophytes, using a focused metabolomics-guided discovery approach to target and subsequently isolate and characterize new bioactive agents. **Methods:** An untargeted LC-MS based metabolomics discovery approach to screen endophytic blueberry isolates (different locations, culture media, and varieties) was used to identify unique metabolite outliers using multivariate statistical analysis. Metabolite outliers were isolated, characterized by 1D and 2D NMR, HRMS, HRMS/MS and subjected to bioactivity screening. **Results:** Targeting of key outlier features within different culture media and blueberry varieties resulted in the discovery of eight new cyclic nonribosomal peptides, three of which have been isolated, with structures elucidated and characterized by 1D and 2D NMR and HRMS/MS analysis. New cyclic pentapeptide #3 demonstrated Gram-negative activity, the first reported for this scaffold. Additionally, potent natural product agents such as griseofulvin, dechlorogriseofulvin, epoxy/cytochalasin D, zygosporin E, hirsutatin A, and cyclic pentapeptides #1-2 and xylariotide A were also discovered. **Conclusion:** Multivariate statistical analysis of a newly described endophytic *Xylaria* species from blueberry plants guided the discovery of new bioactive cyclic nonribosomal peptides without the use of traditional bioactivity guided approaches.

Keywords: Metabolomics, Cyclic nonribosomal peptides, Fungal endophytes, Mass spectrometry, *Xylaria griseofulvin*, *Vaccinium angustifolium*

1 Introduction

Microbial natural products are unparalleled in their capacity to generate a diverse spectrum of specialized metabolites with applications in medicine and agriculture. Between 1981 and 2014 natural products and their derivatives accounted for over 50% of all new small-molecule anti-infective (antibacterial, fungal, parasitic, and viral) approved drugs (n=221) (Newman and Cragg, 2016). In efforts to discover novel natural products, we have applied a LC-MS metabolomic-guided discovery approach to understudied endophytes from wild and cultivated blueberries. Metabolomics allows for a global survey of small molecule metabolites from a biological system and visually representing metabolite variances between groupings or extracts to identify distinct features that can be prioritized for investigation (Hou et al., 2012; Worley and Powers, 2012).

In this study, we have discovered that a previously undescribed endophytic *Xylaria sp.*, common to eastern white pine needles (*Pinus stobus*) and lowbush blueberry stems (*Vaccinium angustifolium*), is a novel *Xylaria* strain that produces the potent antifungal drug griseofulvin and its de-halogenated analogue (Richardson et al., 2014) (Figure 1). Griseofulvin was first discovered from the fungus *Penicllium griseofulvum* in 1939, and has been shown to be a fungistatic agent against plant pathogens, and effective against animal and human fungal infections (Corvis et al., 2006; Dekker, 1963; Oxford et al., 1939). Recently, griseofulvin has also demonstrated new anticancer properties, decades after its commercialization in 1975 as antifungal drug, by inducing an apoptotic response in a number of human and murine cell lines such as colorectal, lymphoma, and myeloma cells (Kim et al., 2011; Sica et al., 2016; Yuan-Soon Ho et al., 2001). These findings are of interest as limited information is known about endophytic fungal secondary metabolites from agriculturally important fruiting plants such as

Ericaceae (blueberries, genus *Vaccinium*). These metabolites may play a role in host plant defense mechanisms by reducing plant pathogens, insect predation and may ultimately prove useful as new therapeutic agents for animal or human diseases (Strobel, 2003; Strobel and Daisy, 2003; Sumarah et al., 2011; Tan and Zou, 2001). The new endophytic species has thus been named *Xylaria elissi* and were present in both cultivated highbush and wild lowbush blueberries, isolated from three different locations in the Acadian forest regions of Nova Scotia, Canada (Figure 1).

In total, 15 endophyte isolates of *Xylaria elissi* were recovered (4 highbush and 11 lowbush), and grown in two fermentation media (PDB and Malt) in efforts to maximize secondary metabolite production and identify metabolite differences between the *Xylaria* isolates (Appendix D, 1S). Filtrate extracts from solvent extracted supernatant (extracellular) cultures and mycelium (intracellular) were screened under standardized LC-UV/MS conditions using a core-shell column for improved separations. Multivariate OPLS-DA S-plot analysis of the intracellular and extracellular secondary metabolomes between the fermentation media and low *vs.* highbush isolates revealed a number of VIP metabolite features (Figure 2-3). Annotation, isolation and structure elucidation of key outlier metabolites lead to the discovery of a family of cyclic nonribosomal peptides. Nonribosomal peptides (NRPS) are of great interest as they represent a unique class of natural products with diverse therapeutic applications such as antimicrobial agents (caspofungin, penicillin, vancomycin), anticancer compounds (bleomycin, daptomycin), immunosuppressant's (cyclosporine, rapamycin) and as mycotoxin's (beauvercin, enniatin, rhizonin) (Fischbach et al., 2006; Gallo et al., 2013; Schwarzer et al., 2003; Walsh, 2015). This complex structural diversity of linear, cyclic, and cyclic branched architectures is synthesized through a modular enzymatic assembly line

process, capable of incorporating >500 proteinogenic and nonproteinogenic building blocks, including polyketide and terpene hybrid moieties. In total, eight new and three known cyclic nonribosomal pentapeptides were discovered (Wu et al., 2011). Of these, a new valine (Val) substituted variant, cyclic pentapeptide # 3 (**12**), demonstrated Gram-negative activity, a first report for this scaffold. Additionally, nine known metabolites were also discovered from *Xylaria elissi*, including the cyclic depsipeptide hirsutain A, cytotoxic agents cytochalasin and epoxy cytochalasin D, zygosporin E, pilformic acid, a benzofuran metabolite Xylarotide A and common hydroxy mellein derivatives (Table 1).

2 Materials and Methods

2.1 Plant Material and Culture extraction and ITS DNA Sequencing

Plant materials including leaves, twigs, stems, flowering fruit and soil samples from high and lowbush blueberries were collected from three different locations within the Acadian forest regions of Nova Scotia, Canada. Highbush blueberry endophyte isolates were obtained from Rawdon, Nova Scotia Lowbush blueberry endophytes isolates were collected from Mount Thom, Debert and Portapique Nova Scotia. Collected specimens were either bagged or stored in labeled conical tubes and stored at -20 C prior to culture screening. Plant tissue specimens are first washed with sterile deionized water to remove any loose debris and surface contaminants, followed by a chemical surface-sterilization process using sodium hypochlorite bleach (6%) and ethanol (70%), ensuring the cultivated organism is an endophytic agent. Small segments were then cut and/or incised, and placed onto microbial agar plates containing Malt extract agar (30 g/L). The inoculated agar plates were incubated at 25°C for 4-8 weeks, depending on the presence of filamentous hypha. Endophytic fungi that grew were then transferred to Potato dextrose agar (PDA) plates for an additional 2 weeks and identified using

a UltraClean Microbial DNA Kit (MoBio, Carlsbad, Ca) of ITS DNA sequenced with ITS1 and ITS4 PCR primers as previously described (Sumarah et al., 2008).

High and lowbush *Xylaria* strains were cultured in PDB (24 g/L potato dextrose broth) and ML (30g/L malt) fermentation media. Each strain was grown in 1L Roux bottles containing 200 mL of media and grown statically for 4-6 weeks at 25°C. The culture broth was then separated from the mycelium by vacuum filtration using a Whatman #4 filter paper. The filtrate was extracted with equal volumes of ethyl acetate, while the mycelium was first lyophilized for 24 h and then extracted with equivalent volumes of methanol and acetone (1:1). The organic fractions were then dried under reduced pressure by rotary vacuum. The extracts were then re-suspended in 600 μL of acetonitrile with minimal amounts of DMSO added for solubility. The filtrates were then centrifuged at 13,000 rpm for 15min and Acro-disk (13 mm, 0.45 μm GHP) filtered prior to LC-MS analysis.

ITS sequences were aligned using MAFFT v7 and visually inspected and manually aligned when necessary in Geneious R8 v8.1.5 (Biomatters, Auckland, New Zealand)(Katoh and Standley, 2013). The most suitable sequence evolution model (GTR+I+G) was determined based on the optimal Akaike information criterion scores in MrModeltest v2.2.6. The ex-type of Mucor ellipsoideus (ATCC MYA-4767; NR_111683) was selected as outgroup on account of its basal position (Zygomycota). Maximum likelihood (ML) analysis was performed using RAxML v8.2.4 in PAUP v4.0b10 starting from a random starting tree with 1000 bootstrap replicates (Stamatakis, 2014; Swofford, 2002). Consensus trees were visualized in FigTree 1.4.2 (available at http://tree.bio.ed.ac.uk/software/figtree/) and exported as SVG vector graphics for assembly in Adobe Illustrator v10 (Adobe Systems, San Jose, CA, USA).

2.2 LC-MS and LC-MS/MS

Endophytic samples were screened using a Dionex Ultimate 3000 HPLC system coupled to a Bruker maXis 4G ultra-high resolution-qTOF mass spectrometer operated in positive and negative electrospray ionization (ESI) with calibrations done using sodium formate ion clusters. LC-MS data was collected using a scan range from 150 to 1100 *m/z*, with the nebulizer gas (nitrogen) at 3 Bar, dry gas flow at 8L/min, dry gas temperature at 240^{O}C, and capillary voltage at 4500 V. Chromatographic separations were performed using a standardized HPLC method with a Supelco Ascentis Express C18 reverse-phase core-shell column (150 × 4.6 mm, 2.7 μm, Sigma Aldrich, USA) operating at 750 μ l/min and at 40^{O}C. Mobile phase composition was linear with a gradient of 5% organic from 0 to 1 min, 5–95% from 1 to 24 min, 95-100% from 24 to 25 min, and 100% 25-31 min. Solvent A was H_2O + 0.1% formic acid and solvent B was acetonitrile with 0.1% formic acid (v/ v). HRS-MS/ MS analysis was performed on a Thermo Q-Exactive Orbitrap mass spectrometer operated in positive electrospray ionization (ESI+) and coupled to an Agilent 1290 HPLC system.

2.3 Data processing and Multivariate Statistical analysis

2.3.1 Data processing

The data processing and analysis were modified from a previously published protocol (Fei et al., 2014). Post-acquisition internal calibration using sodium formate clusters in both ESI+ and ESI- were performed with Bruker's Data Analysis 4.0 SP4. The LC-MS data files were converted to .mzXML format using Bruker CompassXport. The metabolic features were extracted and aligned using open source XCMS with centWave algorithm (Smith et al., 2006); adducts, isotopic ions, and in-source fragments were identified using CAMERA (Kuhl et al. 2010; Kuhlisch and Pohnert 2015). To acquire the final metabolite feature list, isotopic ions and

features with integrated peak area under 10,000 were removed. For intracellular metabolome, metabolite features eluted after 25 min. were eliminated.

2.3.2. *Multivariate Statistical Analysis*

Both intracellular and extracellular metabolomic data were analyzed using supervised multivariate orthogonal partial least-squares discriminative analysis (OPLS-DA) after pareto scaling by SIMCA-P+ 12.0.1 (Umetrics, Kinnelon, NJ). The statistical parameters R^2X(cum), R^2Y(cum), and Q^2(cum) of OPLS-DA were used to assess the fitness of the model. R^2X (R^2Y) indicated the fraction in which metabolite features (X) and group (Y) matrix was were explained by the model. A prediction statistic (Q^2) above 0.4 was indicative of a robust model, i.e. true differences between the comparing groups, and Q^2 between 0.7-1.0 indicated the model was highly robust (Jones et al., 2008). Both R^2 and Q^2 followed an upward trend from 0 to 1. For an over fit model, R^2 approached 1, and Q^2 fell toward 0 (Broadhurst and Kell, 2006). Significant features between classes were identified based on OPLS-DA S-plot and their variable importance in projection (VIP) score. As recommended by Broadhurst and Kell (Broadhurst and Kell, 2006), to ensure the identified metabolites are the sole important biomarkers, the two OPLS-DA analyses were built in parallel by only including the significant features or by remove the significant features from the raw data. A robust metabolite subset was produced if the first model was successful and the later model failed. If both models succeeded, then there were more significant features not included in the initial subset.

2.4 Metabolite Isolation and characterization.

Purification of metabolomic targeted metabolites was performed on a semi-preparative HPLC system consisting of an Agilent 1100 series HPLC with a G1311A Quaternary Pump, a G1379A Degasser, a G1367A Wellplate Autosampler, a G1316A Column Thermostat, a

G1315B Diode Array Detector (DAD), and a G1364C Automatic Fraction Collector controlled by Agilent ChemStation software (Rev. B.03.02-SR2). Metabolites were isolated using a Phenomenex Synergi-Max reverse-phase C-12 column (250 × 10 mm, 4 μm) (Torrence, CA, USA) operating at 5 mL/min and 40^{O}C. Mobile phase composition was a linear gradient of 5% organic from 0 to 3 min, 5–30% from 3 to 16 min, 30% from 16 to 20 min, and 30-85% from 20-37 min with fractions collected every 20 s. Known isolated compounds (mg/L): dechlorogriseofulvin eluted at 27.1 min (4 mg); griseofulvin eluted at 29.1 min (2.8 mg); cytochalasin D eluted at 30.2 min (2.5 mg); 5-hydroxymellein eluted at 30.5 min (1.5 mg); 5-hydroxy-8-O-methylmellein eluted at 31.5min (1 mg); zygosporin E eluted at 32.5 min (2 mg); hirsutatin A eluted at 33.9 min (2 mg); and cyclic pentapeptide #1 eluted at 34.9 min (4 mg). New isolated compounds (mg/L): cyclic pentapeptide # 6 eluted at 31.6 min (0.3 mg); cyclic pentapeptide # 3 eluted at 32.8 min (2.0 mg); cyclic pentapeptide # 4 eluted at 33.4 min (1.3 mg); and cyclic pentapeptide # 5 eluted at 35.6 min (2.3 mg). Compound fractions, from multiple HPLC runs, were pooled together and dried under N_2 gas in pre-weighed vials prior to NMR and optical rotation measurements.

NMR experiments for 1D and 2D measurements were performed on a Bruker Advance III 700 MHz NMR spectrometer equipped with a 5 mm QNP cryoprobe, operating at 700.17 MHz for ^{1}H NMR and 176.08 MHz for ^{13}C NMR, with chemical shifts referenced to the residual solvent signal (Gottlieb et al.,1997). Nitrogen dried compounds were re-suspended in 200 μL of deuterated solvent (C_6D_6, CD_3OD, or DMSO-*d6*) and transferred to 3 mm NMR tubes (Wilmad 335-pp-7) for NMR measurements. NMR data processing was done using MNOVA NMR software ver. 10.0.1 by Mestrelab Research. Optical rotation measurements were done using an Autopol IV Polarimeter (Rudolph Research Analytical).

2.5 Biological activity screening.

Compounds were tested for their minimum inhibitory concentration (MIC) according to the Clinical Laboratory Standards Institute (CLSI) protocols M7-A5 and M27-A (National Committee for Clinical Laboratory Standards, 2000, 1997). Stock working solutions were made to 5, 10 and 20 mg/mL and tested at a maximum concentration of 200 μg/mL in 96-well liquid culture (National Committee for Clinical Laboratory Standards, 1997, 2003). Preliminary evaluation of biological activity was against *E. coli* BW25113 ΔbamBΔtolC, a membrane and efflux pump compromised strain, *Staphylococcus aureus* ATCC#29213, *Bacillus subtilis* 1A1, *Micrococcus luteus, Saccharomyces cerevisiae* B4741, and *Candida albicans* ATCC# 90028. A cut off of 75% growth was used for inhibition, with the trend across dilutions also considered.

3 Results and Discussion

3.1 Multivariate data analysis and validation

In an effort to identify and differentiate unique secondary metabolite differences between the *Xylaria* isolates of high and lowbush blueberries, we evaluated the intercellular and extracellular extracts using three different analysis models: (1) ML vs PDB media cultures; (2) ML media cultures of high vs Lowbush varieties; (3) PDB media cultures of high vs lowbush. The supervised OPLS-DA models are then further examined to include and excluded the top VIP (variable importance of projections) scores as part of the validation process to assess the fitness of the models (R^2X, R^2Y, and Q^2 parameters) (Figure 2-3 and Appendix D, 2S). In general, a robust and valid model can be defined as having a prediction statistic of $Q^2>0.4$, with values above 0.7 being highly robust (Jones et al., 2008). The VIP scores provide a measure of the influence of each correlated metabolite feature, contributing to the separations as seen in the OPLS-DA S-plot (Figure 2), with VIP scores above 0.7 being significant in separations

(Eriksson et al. 2006). A removal of the top VIP scores from each model should thus reduce the prediction statistic Q^2 value, while inclusion of only the top VIP scores should result in a more robust model. In our discovery efforts, VIP scores above 1.7 are evaluated as potential key target outliers for the identification and structural elucidation of novel metabolites. The OPLS-DA validation parameters for each of the intracellular and extracellular metabolite models tested is summarized in Figure 3 and in Appendix Table 1S, respectively. In total, 3856 extracellular metabolite features were identified, with Q^2 values for ML *vs.* PDB at 0.615, and a Q^2 value of 0.778 for both ML and PDB as well as high *vs* lowbush varieties, indicating fairly robust models have been achieved. The total intracellular metabolite features was 2645, with Q^2 values comparable to the intracellular models with (1) ML *vs.* PDB at 0.689; (2) ML high *vs.* lowbush at 0.6, and (3) PDB high *vs.* lowbush at 0.676 (Figure 3). Analysis of the Q^2 values, in relation to the top VIP scores, revealed as expected, greater Q^2 values across all models, when only the top VIP scores (30, 50, 100) are included for both the intracellular and extracellular data sets, while a notable decrease is observed when the top VIP (30, 50, 100, 700, 900) scores are excluded (Figure 3, Appendix D, Table S1-S7).

3.2 Metabolite identification of known compounds 1-11

We first sought to evaluate the top VIP (30, 50,100) scoring metabolites within the extracellular and intracellular models for a targeted isolation and discovery strategy, focusing mostly on UV-MS correlated and deconvoluted peaks of significance (Table 1-2, Appendix D Table S1-7). Specifically, MS correlated metabolite features with observable UV absorptions at λ_{max} values at ~ 210, 254, 275 or 350 nm. Using this approach, HPLC isolated metabolites with yields as low as ~100-200 μg can be effectively characterized by 1D and 2D NMR (cryoprobe equipped), thus avoiding discovery efforts on highly ionizable, low yielding metabolites

(Appendix D 1S). Characterization by 1D and 2D NMR for all the identified metabolite features is neither realistic nor feasible in a discovery process. However, a focused approach aimed at UV abundant VIP targets provides a confident validation metric for the identifications. The UV correlated VIP metabolites were then screened for dereplicated matches against natural product databases such as Antibase and NORINE using their HRMS molecular formulas, including a comparative analysis with reported literature data on known fungal metabolites and mycotoxins (Nielsen et al., 2011; Nielsen and Smedsgaard, 2003; Pupin et al., 2008).

Compounds (**1**-**2**) showed high VIP values (4.6 - 11.42 and 3.2 - 20.43) across the extracellular and intracellular OPLS-DA models. Examination of the HRMS and MS/MS spectra revealed protonated molecular ions at *m/z* 353 and *m/z* 319, with diagnostic daughter ions (at *m/z* 285, 215, 165, 69 and 251, 181, 165, 69 respectively), indicative of the potent antifungal agent griseofulvin and dechlorogriseofulvin (Figure 4, Appendix D, S3 and S38-43). Comparison of the 1D and 2D NMR measurements of compound (**2**) to the literature further validated the identifications (Kimura et al., 1992).

Compound (**3**) showed high VIP values between 4.98 – 7.18 and was predominantly produced from lowbush isolates grown in ML media. The HRMS spectrum showed a protonated molecular ion at *m/z* 508, giving a molecular formula of $C_{30}H_{38}NO_6$ with thirteen degrees of unsaturation. Dereplication screening of the molecular formula *via* Antibase revealed a total of eight potential cytochalasin compounds. Cytochalasins are a unique class of toxic secondary metabolites that are common to the *Xylariaceae* family (Whalley and Edwards, 1995), with over 70 reported analogues in Antibase alone. They are known inhibitors of actin polymerization, affecting cell division, and as plant growth regulators (Casella et al., 1981; Cox et al., 1983; Evidente et al., 1990). Key MS/MS fragment ions at *m/z* 490, 430, 265, 120 and

supporting 1D and 2D NMR (^{1}H, ^{13}C, COSY, HSQC, and HMBC) assignments confirmed the structure as cytochalasin D (Figure 4, Appendix D, S3 and S46-47.).

Examination of the molecular formulas of compounds (**4**) (m/z 492) and (**5**) (m/z 524), indicating possible analogues compounds differing only in one oxygen atom (m/z 16) relative to Cytochalasin D. The 1D and 2D NMR and MS/MS analysis of compounds (**4**) and (**5**) identified them as Zygosporin E (*m/z* 492) and epoxycytochalasin D *(m/z* 524) (Appendix D, S3 and S48-49).

Compound (**6**) had protonated molecular ion of m/z 677 with VIP values from 3.35-7.42, and was predominantly found in the intercellular metabolome from both ML and PDB media. Database screening returned no matches from Norine, and only one potential candidate from Antibase, that of hirsutatin A, a cyclic hexadepsipeptide isolated from the insect pathogenic fungus *Hirsutella nivee* (Isaka et al., 2005). The ^{1}H spectrum of compound (**6**) in C_6D_6 revealed the presence of key diagnostic amide protons (δ 8.50 [d, *J* =6.5 Hz], δ 7.93 [d, *J* =8.8 Hz], δ 6.75 [d, *J* =9.0 Hz]) and one N-methyl-amide signal (δ 2.58 [3H,s]), including six α-protons, which supported the hirsutatin peptide scaffold (Figure 4, Appendix D, S3 and S44-45). Furthermore, the ^{13}C NMR and multiplicity edited HSQC spectrums revealed a total of 34 carbon atoms, with six resonances attributed to carbonyls (δ 172.7, 171.9, 171.8, 170.5, 169.7 and 168.0 ppm), which is consistent with the literature assignments (Isaka et al., 2005).

Compound (**7**) was established to be piliformic acid and based on the protonated molecular ion of *m/z* 215 and key diagnostic MS/MS fragments of *m/z* 197, 169, 241, 120, that

are in agreement with the literature (Mangaleswaran and Argade, 2000) (Figure 4, Appendix D, S3).

Compound (**8**) was identified as 2, 3-dihydro-2,4-dimethylbenzofuran-7-carboxylic acid, a furanobenzene metabolite recently reported from *Xylaria* along with compounds (**1**) and (**7**) (Richardson et al., 2014). The furanobenzene metabolite had a protonated molecular ion at *m/z* 193 and diagnostic MS/MS fragments at *m/z* 175, 165, 147 and 91(Appendix D, S3).

The HRSMS spectra for Compound (**9**) showed a protonated molecular ion at *m/z* 584, affording a molecular formula of $C_{32}H_{49}N_5O_5$, with VIP values of 5.36-16.24. Compound (**9**) is found in both the extracellular and intracellular metabolomes. Database screening of the HRMS molecular formula identified only one candidate, cyclic pentapeptide 1, *cyclo*-(-NMePhe-Pro-Leu-Ile-Val-) (Wu et al., 2011). Isolation and characterization of compound (**9**) by 1D and 2D NMR as well as MS/MS analysis confirmed the identify of cyclic pentapeptide 1 (Appendix D, S4 and S32-S37). Examination of the ^{1}H and ^{13}C NMR spectra revealed the presence of five α protons (δ 5.24/61.0, 5.04/58.8, 4.86/49.0, 4.25/57.8 and 4.01/59.9 ppm) and their associated carbonyl resonances (174.4, 173.2, 172.5, 172.4, and 171.5 ppm). Analysis of the MS/MS spectra of (**9**) further validated the identification (Figure 4, Appendix D, S4). Cyclic peptides in general, can generate a number of direct-sequence and non-direct sequence fragment ions when subjected to collision induced dissociation (CID) within a mass spectrometer. These fragment ions can arise from multiple ring opening events along the peptide back-bone, with cleavage sites at the amide bonds (Eckart, 1994; Liu et al., 2009). In the case of cyclic pentapeptide 1, the presence of the N-methylated phenylalanine (N-MePhe) and the presence of a basic residue (Pro), created two main fragmentation pathways with ring opening events at the (1) N-MePhe-Pro and (2) Pro-Leu amide bonds (Figure 6). Diagnostic MS/MS fragments for (**9**) at *m/z* 487.3,

374.2, 261.2 and 162.1 support the main N-MePhe-Pro fragmentation pathway, while the minor Pro-Leu pathway is evident with diagnostic fragments at *m/z* 471.3, 358.2, 261.2 and 162.1.

Compound (**10**) was identified as xylarotide A based on the protonated molecular ion at *m/z* 550 m/z and diagnostic MS/MS fragments (*m/z* 453.3, 340.3, 227.3, 128.1). Xylarotide is a related cyclic pentapeptide with an N-MeLeu substituted for N-MePhe (Li et al., 2011). Xylarotide was present in all extracellular and intracellular models, and found in both low and high bush blueberry extracts with VIP scores of 6.51-11.53 (Appendix D, S4).

Compounds (**11**) was identified as cyclic pentapeptide 2 with a protonated molecular ion of *m/z* 536 and diagnostic MS/MS fragments in accordance with literature values (*m/z* 423.3, 326.2, 213.2, 100.1) (Appendix D, S4). Cyclic pentapeptide 2 had VIP scores of 2.26-2.81 and is structurally similar to cyclic pentapeptide 1 and xylariotide A with Leu substituted for N-MePhe and N-MeLeu respectively (Wu et al., 2011).

3.3 Structure elucidation of new cyclic pentapeptides

Compounds (**12-15)** were identified by metabolomic analysis of the extracellular and intercellular metabolome models as unique outliers with VIP scores of 2.6 - 11.59 (Figure 2, and Table 1-2). The new cyclic pentapeptides are structurally similar to cyclic pentapeptide 1, with amino acid differences at positions 2 (Ala / IsoLeu *vs.* Val) and 3 (Val *vs.* IsoLeu) within the peptide scaffold (Figure 5-6).

Compound (**12**) was isolated as a colorless powder and afforded a protonated molecular ion at *m/z* 556 ($C_{30}H_{45}N_5O_5$ with 11 double bond equivalents). Examination of the ^{1}H and ^{13}C NMR data revealed the presence of five α protons (δ 5.08/55.8, 4.49/46.0, 4.17/55.7, 4.74/46.5,

5.10/58.9 ppm), the three key amide N-H protons for Ala (δ 8.52), IsoLeu (δ 6.94) and Leu (δ 8.49) and the N-Methyl group at (δ 3.04/30.2 ppm). Examination of the ^{1}H-^{1}H COSY, multiplicity edited ^{1}H-^{13}C HSQC and HMBC NMR data revealed the individual amino acid spin systems based on α proton correlations to the individual carbonyl carbon, amide protons to neighboring amino acid carbonyl, and α, *β* and γ proton correlations (Figure 5, Appendix D, S5-13, Table S8-9) and supported the amino acid sequence of cyclo-(NMePhe-Ala-IsoLeu-Leu-Pro). NOESY through-space correlations of αH(N–MePhe)/NH (Ala), NH(Ala)/*β*H (Ala), NH(IsoLeu)/ αH (Ala), NH(Leu) / αH (IsoLeu) and H_3-NMe (N-MePhe)/ *β*H (Pro) further confirmed the amino acid sequence. Analysis of the MS/MS spectra of compound (**12**) revealed key diagnostic fragments of *m/z* 459.3 (-Pro), 346.2 (-Leu), 233.1 (-IsoLeu), 162.1 (-Ala) and the presence of two fragmentation pathways as seen in cyclic pentapeptide 1 with ring-opening cleavage at the N-MePhe-Pro and Pro-Leu.

Compound (**13**) was isolated as a colorless powder with a protonated molecular ion at *m/z* 570 affording a molecular formula of $C_{31}H_{47}N_5O_5$ with 11 double bond equivalents. Examination of the ^{1}H and ^{13}C NMR data revealed the presence of five α protons (δ 5.10/56.0, 3.95/57.6, 4.10/56.8, 4.72/46.6, 5.08/58.7 ppm), key amide N-H protons for Val (δ 8.18), Val_2 (δ 6.98) and Leu (δ 8.43) and the N-Methyl group at (δ 3.03/30.2 ppm). Compound (**13**) differs from (**9**) with Val substituted for IsoLeuc at position # 3 (Figure 5, Appendix D, S14-21, Table S8 and S10). Examination of the MS/MS spectra revealed a similar fragmentation pattern as in (**9**) and (**12**), with key diagnostic fragment ions at *m/z* 471.3 (-Pro), 360.2 (-Leu), 261.2 (-Val), and 162.1 (-Val). The cyclo-(NMePhe-Val_1-Val_2-Leu-Pro) amino acid sequence was confirmed with key HMBC correlations of αH(N–MePhe)/CO (N-MePhe), H_3-NMe (N-MePhe)/ CO (Pro), αH(Val_1)/CO (Val_1), αH(Val_2)/CO (Val_2), αH (Leu)/CO

(Leu), αH (Pro)/CO (Pro) and key NOESY correlations of αH(N–MePhe)/NH (Val_1), NH(Val_1)/ αH (Val_1), NH(Val_2)/ αH (Val_1) , NH(Val_2)/ αH (Val_2), NH(Val_2)/ βH (Val_2), NH(Leu)/ NH (Val_2) and H_3-NMe (N-MePhe)/ βH (Pro).

Compound (**14**) was isolated as a colorless powder with a protonated molecular ion at *m/z* 598 affording a molecular formula of $C_{33}H_{51}N_5O_5$ with 11 double bond equivalents. Examination of the ^{1}H and ^{13}C NMR data revealed the presence of five α protons (δ 5.10/56.0, 4.05/55.8, 4.25/55.1, 4.72/ 46.5, 5.09/58.7 ppm), key amide N-H protons for IsoLeu1 (δ 8.12), IsoLeu2 (δ 6.94) and Leu (δ 8.45) and the N-Methyl group at (δ 3.04/30.2 ppm). Compound (**14**) differs from (**9**) with IsoLeu substituted for Val at position # 2 (Figure 5, Appendix D, S22-32, Table S8 and S11). Examination of the ^{1}H-^{1}H COSY, multiplicity edited ^{1}H-^{13}C HSQC and HMBC NMR data revealed the individual spin system for the new IsoLeuc group with correlations of β–3H(IsoLeu)/)/ αH (IsoLeu) and δH(IsoLeu)/)/ βH (IsoLeu). Correlations of the remaining α protons to the individual carbonyl carbons, amide protons to neighboring amino acid carbonyl, and HMBC α, β and γ proton correlations for Leu, Pro and N-MePhe is consistent with the cyclic peptide scaffold (Figure 5, SI Table 11S.). NOESY through-space correlations of αH(N–MePhe)/NH ($IsoLeu_1$), NH ($IsoLeu_1$)/NH ($IsoLeu_2$), NH (Leu)/ αH ($IsoLeu_2$)) further supported the amino acid sequence. Analysis of the MS/MS spectra of compound (14) revealed key diagnostic fragments of *m/z* 501.3 (-Pro), 388.3 (-Leu), 275.2 (-$IsoLeu_2$), and 162.1 (-$IsoLeu_1$) further confirming the amino acid sequence of cyclo-(N-MePhe-$IsoLeu_1$-$IsoLeu_2$-Leu-Pro).

3.4 MS/MS analysis of new putative cyclic pentapeptides

Compounds (**15-19)** were identified by metabolomic analysis of the extracellular and intercellular metabolome models as unique outliers with VIP scores of 1.92-6.46. Evaluation of the HRESIMS derived molecular formulas and MS/MS fragmentation patterns of (**15-19**) indicated the fragmentation sequence and ring-opening events are consistent to the cyclic pentapeptides, as such, are reported as new cyclic pentapeptides of low abundance with ppm errors within 1.72-2.41 ppm. (Appendix D, S4 and Table S8).

Compound **15** (Rt ~ 14.47min) was identified in the PDB extracellular metabolome of highbush blueberry extracts. The molecular formula was determined to be $C_{27}H_{47}N_5O_5$ with 7 double bond equivalents based on the protonated molecular ion of *m/z* 522. Analysis of the MS/MS spectra revealed the two main fragmentation pathways with diagnostic fragments of *m/z* 409.3 (-Leu), 312.2 (-Pro), 199.1 (-Leu), 100.1 (-Val) indicating ring-opening events at Val-Leu, and fragments at *m/z* 425.3 (-Pro), 312.2 (-Leu) indicating a ring-opening event at Leu-Val (SI Figure 2S. Compound **15** is analogous to cyclic pentapeptide 2 with Val substituting for IsoLeu at position #3, confirming a sequence of cyclo-(Leu-Val-Val-Leu-Pro).

Compound **16** (Rt ~16.89) was identified in the extracellular and intracellular metabolome and primarily in lowbush blueberry extracts. The protonated molecular ion was *m/z* 564 and the molecular formula was established to be $C_{30}H_{54}N_5O_5$ with 7 double bond equivalents. MS/MS analysis of the tandem MS spectra revealed key fragment ions at m/z 467.4 (-Pro), 354.3 (-Leu), 241.2 (-IsoLeu), 128.1 (-Isoleu/Leu) and at *m/z* 451.3 (-Leu), 338.2 (-IsoLeu) and 225.2 (-IsoLeu) indicating ring-opening events at N-MeLeu-Pro and Pro-Leu, consistent with the cyclic pentapeptide scaffold. Compound (**16**) is most similar to cyclic

pentapeptide 2, with an N-MeLeu substituted for Leu at position 1 and Iso/Leu substituted at position 2.

Compound (**17**) was identified based on the protonated molecular ion at *m/z* 586 affording a molecular formula of $C_{31}H_{47}N_5O_6$ with 11 double bond equivalents and comparative MS/MS analysis to compound (**13**). Examination of the molecular formula of (**13**) and (**17**) indicated a difference of a single oxygen atom, with no change in the degrees of unsaturation. Comparative analysis of the MS/MS tandem spectra of (**17)** revealed diagnostic fragments for ring-opening at N-MePhe-Pro+16 of *m/z* 473.3 (-Pro+16), 360.2 (-Leu), 261.2 (-Val) and 162.1 (-Val) and ring-opening at Pro+16-Leu at *m/z* 473.3 (-Leu), 374.2 (-Val), and 275.1 (-Val). This fragmentation pattern is consistent with the cyclic pentapeptide scaffold, and similar to (**13**), indicative of a hydroxyl-proline moiety at position 5. This is further supported by the earlier, more polar elution time of (**17**) vs (**13**) , Rt ~14.11 min vs ~15.19 min on the reverse-phase C-18 column (SI Figure 2S and SI Table 8A).

Similar to (**17**), compound (**18**) had a protonated molecular ion of *m/z* 600, affording a molecular formula of $C_{32}H_{51}N_5O_6$ with 11 double bond equivalents and differing in a single oxygen atom as compared to (**9**). MS/MS analysis of compound (**18**) revealed a similar fragmentation pattern to (**9**) with diagnostic fragments at *m/z* 487.3 (-Pro+16), 374.2 (-Leu), 261(-IsoLeu), 162.1 (-Val) indicating a hydroxyl proline moiety. Examination of the elution times of (**18**) and (**9**) reveals Rt~14.00 vs ~16.2min respectively, supporting the presence of a hydroxyl addition at position #5.

Compound (**19**) was identified as another new hydroxy proline substituted cyclic pentapeptide. The protonated molecular ion for (**19**) was m/z 614, affording a molecular formula of $C_{33}H_{51}N_5O_6$ with 11 double bond equivalents and differing in a single oxygen atom

as compared to (**14**). MS/MS analysis of the tandem MS spectra showed diagnostic ions at *m/z* 501.3 (-Pro+16), 388.2 (-Leu), 275.2 (-IsoLeu), 162.1 (-IsoLeu) with only one main fragmentation pathway visible, indicating ring-opening at N-MePhe-Pro+16. Similar to above, compound (**19**) eluted earlier than (**14**) at Rt~14.89 vs ~17.04 respectively.

4.1 Bioactivity

Compounds **9** and **12-14** were screened for biological activity against three species of microorganisms in accordance with the Clinical Laboratory Standards Institute (CLSI) protocols. The microorganism included *E. coli* BW25113 ΔbamBΔtolC, *Saccharomyces cerevisiae* B4741, and *Candida albicans* ATCC# 90028).

Compound **12**, cyclic pentapeptide #3, showed modest activity against *E. coli* with a minimum inhibitory concentration (MIC) of 100 μg/mL. This is a first report for the scaffold. However, no antifungal activity was observed at 100 μg/mL. Compounds 9 and 12-14 showed no activity when tested at 200 μg/mL and 50 μg/mL respectively.

5.1 Conclusion

In this study, we have applied an LC-MS metabolomic guided discovery approach to profile the chemical space of endophytic Xylaria isolates from low and highbush blueberries, grown in different culture conditions. These isolates have similar ITS DNA sequences, yet differ in their LC-MS metabolite profiles. OPLS-DA and S-plot analysis identified a number of VIP features from the extracellular and intercellular extracts which were targeted for HPLC semi-preparative isolation and characterization. This approach led to the discovery and structure elucidation of 3 new cyclic pentapeptides and putative identification of 5 additional new cyclic pentapeptides by LC-HRMS and LC-MS/MS analysis. Additionally, 11 known

compounds were discovered. Cyclic pentapeptide #3 (**12**) was active against Gram-negative bacteria and is a first report for the scaffold. These findings are of interest as the isolates have also been reported from eastern white pine needles (*Pinus strobus*) in a pine-blueberry ecotype. Endophytes from fruiting plants are an understudied source for new natural products. These findings have shown that this understudied niche environment may provide new opportunities for the discovery of novel bioactive compounds and provide a better understanding of the diverse chemical space and chemical ecology of plant-fungi microbiomes.

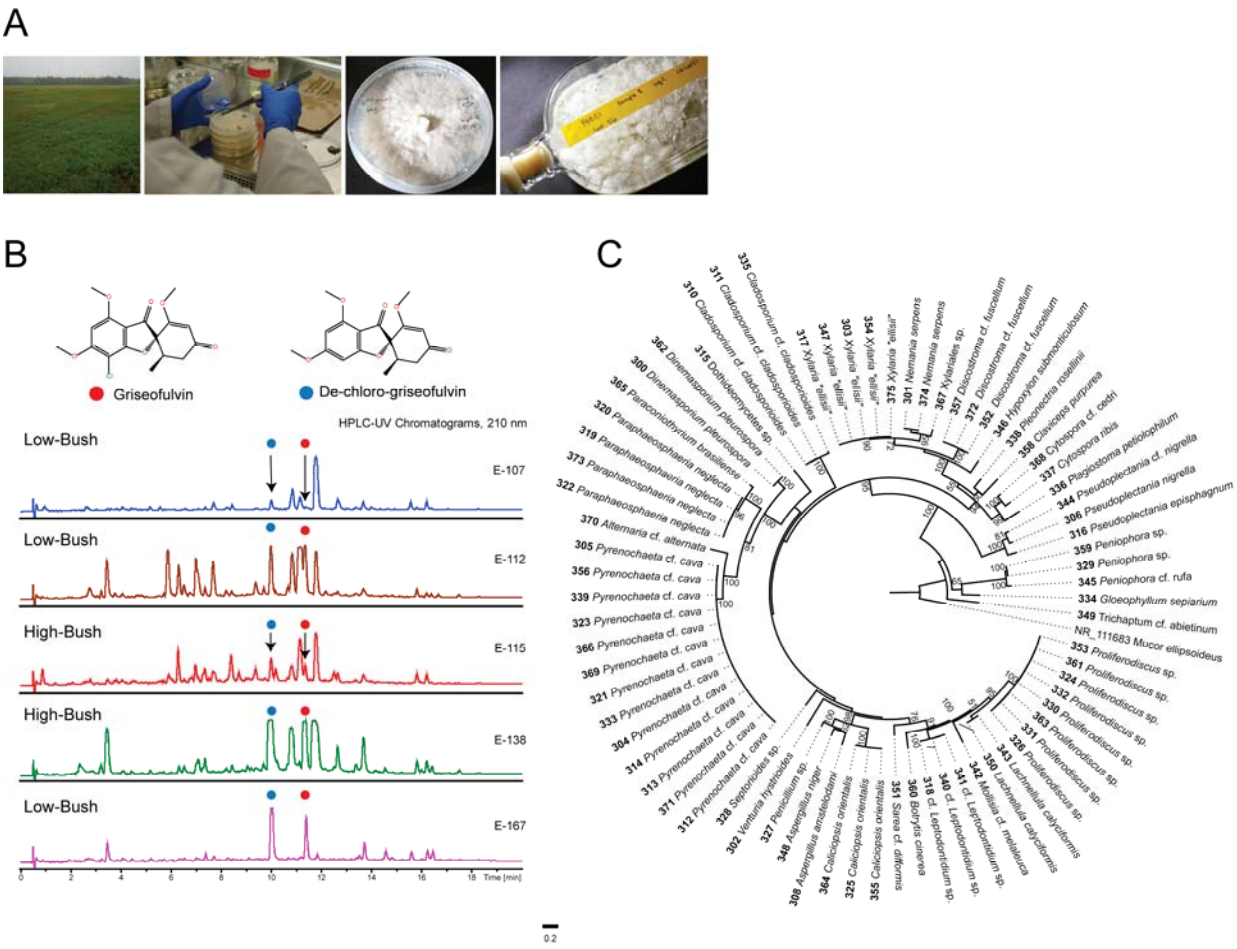


Figure. 1 Discovery of new griseofulvin-producing fungal endophyte species, *Xylaria elissi* isolated from high and lowbush blueberry leaves and stems. **a** Isolation and culturing of fungal endophyte, **b** LC-UV comparative profile analysis of crude filtrate extracts at ƛ 210 nm, revealing differences in metabolite production, **c** Most likely tree from a RAxML analysis of ITS dataset containing representative endophytes. Culture numbers precede the species name and RAxML bootstrap support percentages ≥50 from a summary of 1000 replicates are presented at the branch nodes. The tree was rooted with *Mucor ellipsoideus* (ATCC MYA- 4767; NR_111683) and the scale bar represents the number of substitutions per site.

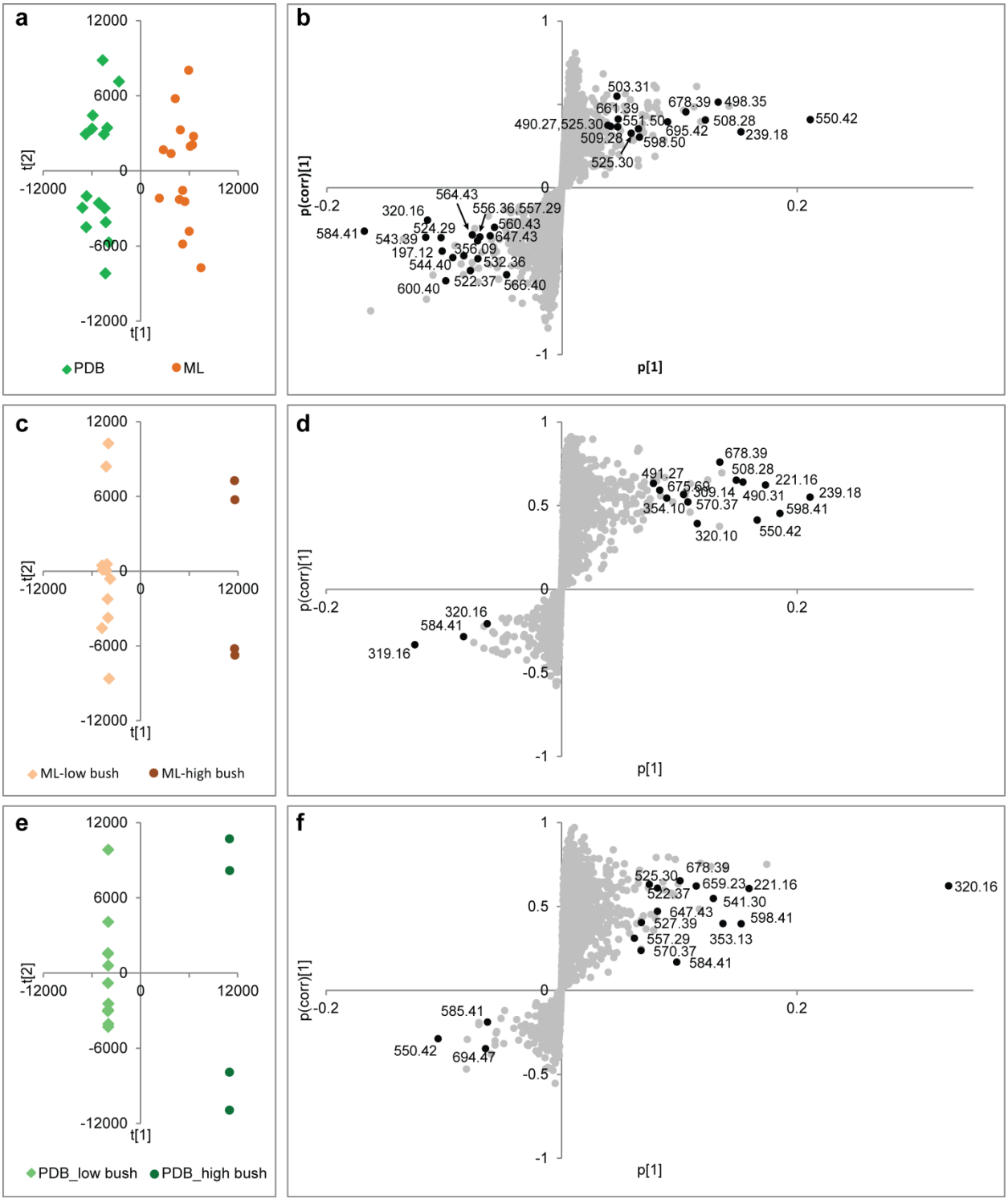


Figure 2. Supervised multivariate analyses of extracellular metabolome of Griseofulvin endophytes. The OPLS-DA score plot (a) and S-plot (b) for comparison between Griseofulvin endophytes cultured in ML or PDB media. The OPLS-DA score plots and S- plots compared the Griseofulvin endophytes isolates from high or low bush blueberries cultured in ML (c, d) or PDB (e, f) medium, respectively.

Model	Variables*	R^2X(cum)	R^2Y(cum)	Q^2(cum)	Conditions
1a	3856	0.15	0.939	0.615	ML, PDB
1b	100	0.313	0.935	0.737	ML, PDB (include top 100 VIP)
1c	3756	0.299	0.982	0.434	ML, PDB (exclude top 100 VIP)
1d	3556	0.207	0.954	0.394	ML, PDB (exclude top 300 VIP)
2a	3856	0.477	0.998	0.778	ML-H, ML-L
2b	30	0.861	0.984	0.864	ML-H, ML-L (include top 30 VIP)
2c	3826	0.544	1	0.718	ML-H, ML-L (exclude top 30 VIP)
2d	3156	0.121	0.73	-0.187	ML-H, ML-L (exclude top 700 VIP)
3a	3856	0.648	1	0.778	PDB-H, PDB-L
3b	50	0.589	0.995	0.885	PDB-H, PDB-L (include top 50 VIP)
3c	3806	0.651	1	0.668	PDB-H, PDB-L (exclude top 50 VIP)
3d	2956	0.0851	0.894	-0.175	PDB-H, PDB-L (exclude top 900 VIP)

* Number of metabolic features included in the OPLS-DA analyses

Figure 3. A summary of validation parameters (R^2X, R^2Y, Q^2) of all calculated OPLS-DA models for the extracellular extracts of Griseofulvin endophytes isolates from low and high bush cultured in ML and PDB media. ML-H, endophyte isolates from high bush cultured in ML medium; ML-L, endophyte isolates from low bush cultured in ML medium; PDB-H, endophyte isolates from high bush cultured in PDB medium; PDB-L, endophyte isolates from low bush cultured in PDB medium.

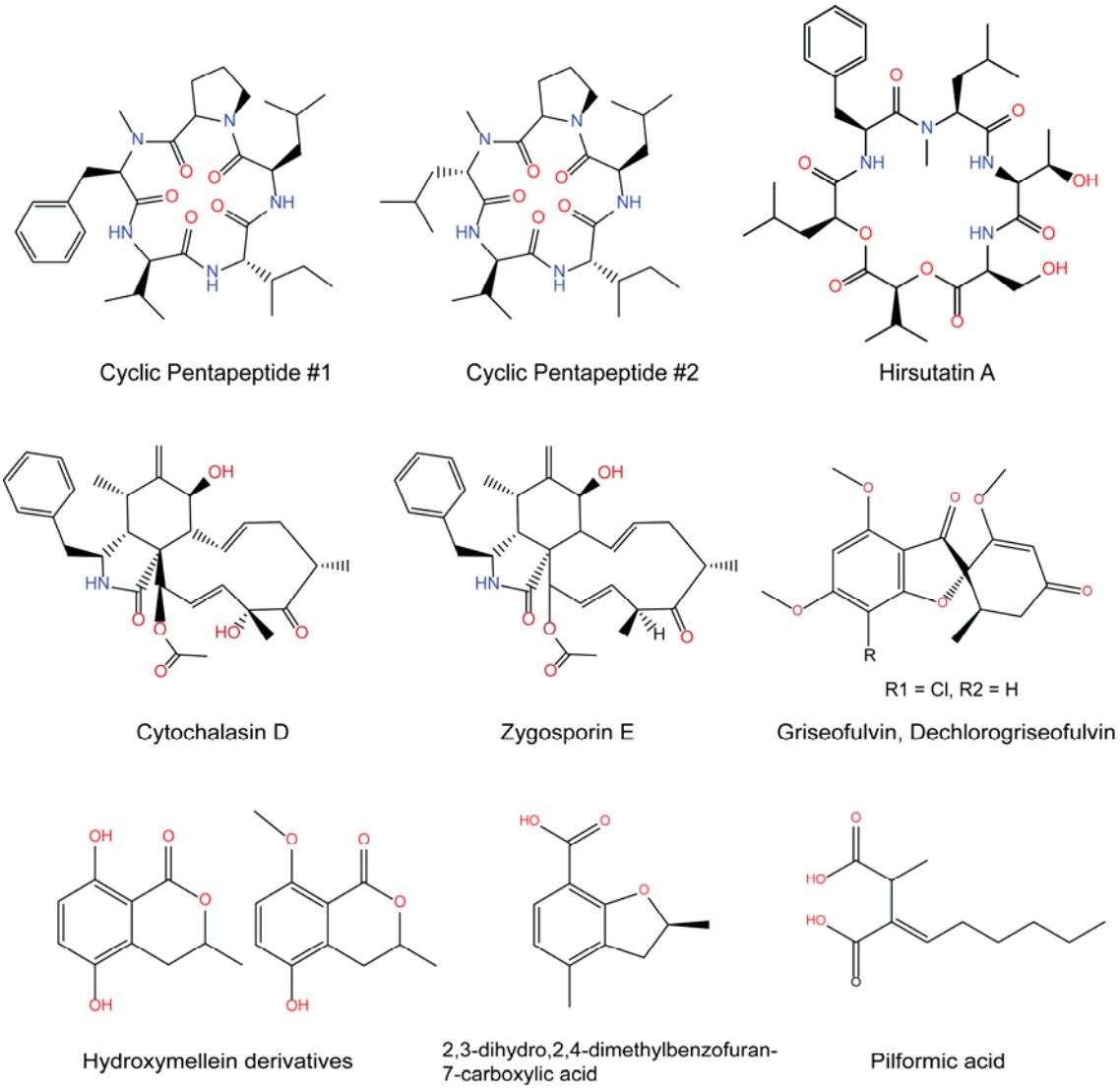


Figure 4. Structures of selected known compounds.

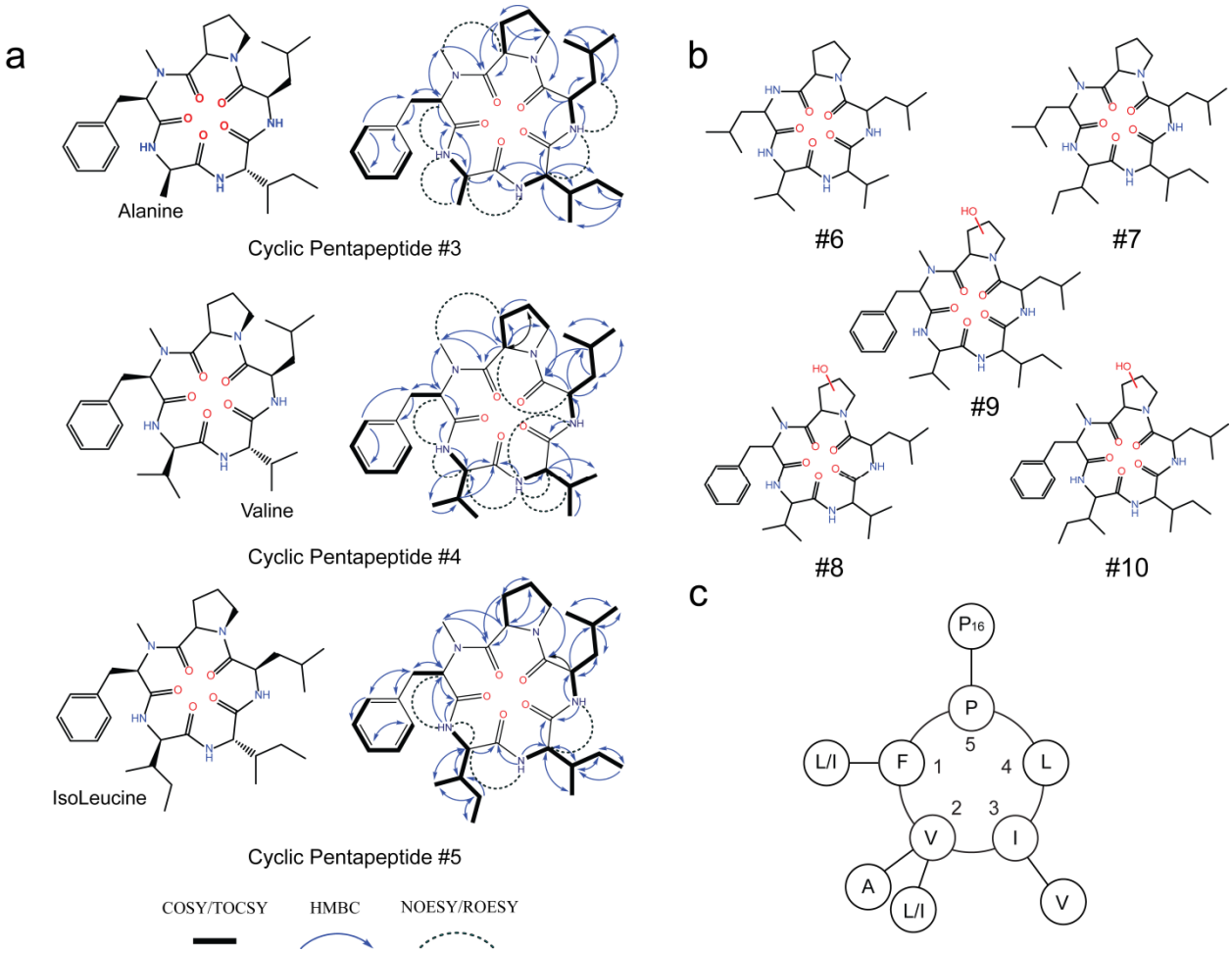


Figure 5. Metabolomic guided discovery of new cyclic nonribosomal peptides from *Xylaria Griseofulvin*. **a** New cyclic pentapeptides #3-5 isolated and characterized by 1D and 2D NMR, HR-MS and MS/MS analysis with new amino acid substituent highlighted with corresponding COSY/TOCSY (^{1}H -^{1}H), HMBC (^{1}H - ^{13}C) and long-range NOESY/ROESY correlations shown. **b** Structures of new cyclic Pentapeptides #6-10 based on HR-MS measurements and comparative HR-MS/MS analysis to cyclic pentapeptides #1-5. **c** Amino acid scaffold of the cyclic pentapeptide family compounds. Cyclic pentapeptide #1 shown with established amino acid substituents.

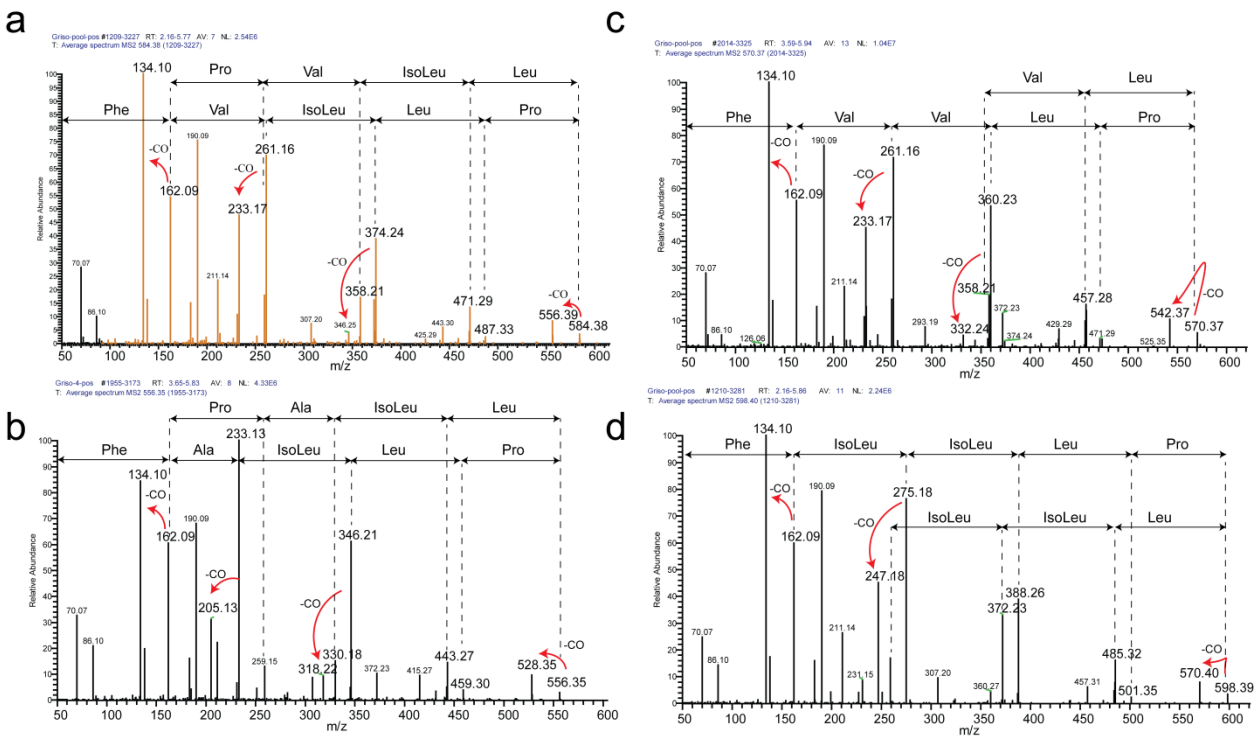


Figure 6. MS/MS spectra of cyclic pentapeptides; (A) cyclic pentapeptide #1 (**9**) , (B) cyclic pentapeptide #3 (**12**), (C) cyclic pentapeptide 4 (#**13**), (D) cyclic pentapeptide #5 (**14**).

Table 1. Identification of known and new secondary metabolites from *Xylaria elissi* via LC-UV/HRMS and MS/MS analysis

Compound	Name	Class	Rt	Molecular Formula	Measured and Calculated $[M+H]^+$		ppm error
Known							
2	De-chlorogriseofulvin *	PKS	10.01	$C_{17}H_{19}O_6$	319.1173	319.1176	0.94
7	Piliformic acid	PKS	10.84	$C_{11}H_{18}O_4Na$	237.1094	237.1097	1.27
5	Epoxycytochalasin D	PKS-NRPS	10.87	$C_{30}H_{38}N0_7$	524.2651	524.2661	1.91
8	2,3-dihydro,2,4-dimethylbenzofuran -7-carboxylic acid	PKS	11.15	$C_{11}H_{13}O_3$	193.0857	193.0859	1.04
1	Griseofulvin	PKS	11.42	$C_{17}H_{18}ClO_6$	353.0793	353.0786	-1.98
3	Cytochalasin D *	PKS-NRPS	11.81	$C_{30}H_{38}NO_6$	508.2687	508.2694	1.38
4	Zygosporin E *	PKS-NRPS	13.94	$C_{30}H_{38}NO_5$	492.2742	492.2744	0.41
11	Cyclic Pentapeptide #2 *	NRPS	14.28	$C_{28}H_{50}N_5O_5$	536.3819	536.3806	-2.42
6	Hirsutain A *	NRPS	15.85	$C_{34}H_{53}N_4O_{10}$	677.3741	677.3756	2.21
10	Xylarotide A	NRPS	15.97	$C_{29}H_{52}N_5O_5$	550.3973	550.3963	-1.82
9	Cyclic Pentapeptide #1 *	NRPS	16.20	$C_{32}H_{50}N_5O_5$	584.3816	584.3806	-1.71
New							
18	Cyclic Pentapeptide #9	NRPS	14.00	$C_{32}H_{50}N_5O_6$	600.3768	600.3756	-2.00
17	Cyclic Pentapeptide #8	NRPS	14.11	$C_{31}H_{48}N_5O_6$	586.3616	586.3599	-1.72
15	Cyclic Pentapeptide #6	NRPS	14.47	$C_{27}H_{48}N_5O_5$	522.3662	522.3650	-2.30
12	Cyclic Pentapeptide #3 *	NRPS	14.76	$C_{30}H_{46}N_5O_5$	556.3501	556.3493	-1.44
19	Cyclic Pentapeptide #10	NRPS	14.89	$C_{33}H_{52}N_5O_6$	614.3936	614.3912	-2.41
13	Cyclic Pentapeptide #4 *	NRPS	15.19	$C_{31}H_{48}N_5O_5$	570.3656	570.3650	-1.05
16	Cyclic Pentapeptide #7	NRPS	16.89	$C_{30}H_{54}N_5O_5$	564.4132	564.4119	-2.30
14	Cyclic Pentapeptide #5 *	NRPS	17.04	$C_{33}H_{52}N_5O_5$	598.3968	598.3963	-0.84

* Structures elucidated by 1D and 2D NMR, HRMS and MS/MS analysis.

Table 2. Known and new compound TOP VIP scores from OPLS-DA models of extracellular and intracellular metabolomes of griseofulvin producing Xylaria endophytes.

Compound	Name	Class	VIP Scores - Extracellular*			VIP Scores - Intracellular*		
			1b	2b-ML Low vs High Bush	3b-PDB	1b	2b- ML Low vs High Bush	3b-PDB
	Known							
2	De-chlorogriseofulvin	PKS	8.07	8.12	20.43	7.78	21.82	3.20
7	Piliformic acid	PKS	6.53	/	3.70	2.75	1.80	
5	Epoxycytochalasin D	PKS-NRPS	5.82	/	4.23	/	9.43	4.86
8	2,3-dihydro,2,4-dimethylbenzofuran - 7- carboxylic acid	PKS	/	/	/	2.48	/	/
1	Griseofulvin	PKS	4.81	5.66	8.53	10.77	5.72	17.22
3	Cytochalasin D	PKS-NRPS	7.18	9.45	7.28	3.94	4.55	/
4	Zygosporin E	PKS-NRPS	/	/	/	2.36	2.81	/
11	Cyclic Pentapeptide #2	NRPS	2.81	/	/	2.60	2.26	/
6	Hirsutain A	NRPS	/	/	4.01	3.90	8.70	5.79
10	Xylarotide A	NRPS	11.37	10.52	6.51	11.98	6.44	6.14
9	Cyclic Pentapeptide #1	NRPS	11.10	5.36	6.09	5.60	16.24	/
	New							
18	Cyclic Pentapeptide #9	NRPS	6.46	/	/	/	/	/
17	Cyclic Pentapeptide #8	NRPS	/	/	/	/	/	/
15	Cyclic Pentapeptide #6	NRPS	4.60	/	5.07	/	/	/
12	Cyclic Pentapeptide #3	NRPS	3.72	/	/	2.60	/	/
19	Cyclic Pentapeptide #10	NRPS	/	/	/	/	/	/
13	Cyclic Pentapeptide #4	NRPS	/	6.87	4.21	/	/	/
16	Cyclic Pentapeptide #7	NRPS	5.59	/	/	2.83	1.92	/
19	Cyclic Pentapeptide #5	NRPS	3.23	11.59	/	7.17	5.45	4.84

* Assignments based on the parent molecular ion [M+H]+, double charged species [2M+H]+, including Na and NH4 adducts or precursor ion fragments.

References

1. Broadhurst, D.I., Kell, D.B., 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2, 171–196.
2. Casella, J.F., Flanagan, M.D., Lin, S., 1981. Cytochalasin D inhibits actin polymerization and induces depolymerization of actin filaments formed during platelet shape change. Nature 293, 302–305.
3. Corvis, Y., Barzyk, W., Brezesinski, G., Mrabet, N., Badis, M., Hecht, S., Rogalska, E., 2006. Interactions of a fungistatic antibiotic, griseofulvin, with phospholipid monolayers used as models of biological membranes. Langmuir 22, 7701–7711.
4. Cox, R.H., Cutler, H.G., Hurd, R.E., Cole, R.J., 1983. Proton and Carbon-13 nuclear magnetic resonance studies of the conformation of cytochalasin H derivatives and plant growth regulating effects of cytochalasins. J. Agric. Food Chem. 31, 405–408.
5. Dekker, J., 1963. Antibiotics in the control of plant diseases. Annu. Rev. Microbiol. 17, 243–262.
6. Eckart, K., 1994. Mass spectrometry of cyclic peptides. Mass Spectrom. Rev. 13, 23–55.
7. Evidente, A., Randazzo, G., Vurro, M., 1990. Cytochalasins: structure-activity relationships. Phytochemistry 29, 93–96.
8. Fei, F., Bowdish, D.M.E., McCarry, B.E., 2014. Comprehensive and simultaneous coverage of lipid and polar metabolites for endogenous cellular metabolomics using HILIC-TOF-MS. Anal. Bioanal. Chem. 406, 3723–3733.
9. Fischbach, M.A., Walsh, C.T., Fischbach, M.A., Walsh, C.T., 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic , machinery , and mechanisms. Chem Rev 106, 3468–3496.
10. Gallo, A., Ferrara, M., Perrone, G., 2013. Phylogenetic study of polyketide synthases and nonribosomal peptide synthetases involved in the biosynthesis of mycotoxins. Toxins 5, 717–742.
11. Hou, Y., Braun, D.R., Michel, C.R., Klassen, J.L., Adnani, N., Wyche, T.P., Bugni, T.S., 2012. Microbial strain prioritization using metabolomics tools for the discovery of natural products. Anal. Chem. 84, 4277–83.
12. Isaka, M., Palasarn, S., Sriklung, K., Kocharin, K., 2005. Cyclohexadepsipeptides from the insect pathogenic fungus Hirsutella nivea BCC 2594. J. Nat. Prod. 68, 1680–1682.
13. Jones, O.A.H., Spurgeon, D.J., Svendsen, C., Griffin, J.L., 2008. A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm Lumbricus rubellus. Chemosphere 71, 601–609.
14. Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
15. Kim, Y., Alpmann, P., Blaum-Feder, S., Krämer, S., Endo, T., Lu, D., Carson, D., Schmidt-Wolf, I.G.H.,

2011. In vivo efficacy of griseofulvin against multiple myeloma. Leuk. Res. 35, 1070–1073.

16. Kimura, Y., Shiojima, K., Nakajima, H., Hamasaki, T., 1992. Structure and biological activity of plant growth regulators produced by *Penicillium* sp. No. 31f. Biosci. Biotechnol. Biochem. 56, 1138–1139.

17. Li, Y.-Y., Hu, Z.-Y., Yue-Mao, Shen, 2011. Two new cyclopeptides and one new nonenolide from *Xylaria* sp. 101. Nat. Prod. Commun. 6, 1843–18464.
18. Liu, W.-T., Ng, J., Meluzzi, D., Bandeira, N., Gutierrez, M., Simmons, T.L., Schultz, A.W., Linington, R.G., Moore, B.S., Gerwick, W.H., Pevzner, P. a, Dorrestein, P.C., 2009. Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. Anal. Chem. 81, 4200–9.
19. Mangaleswaran, S., Argade, N.P., 2000. An efficient synthesis of (±)-piliformic acid. J. Chem. Soc. Perkin Trans. 1, 3290–3291.
20. National Committee for Clinical Laboratory Standards, 2000. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically ; approved Standard — 5th Edition. Wayne, PA Natl. Comm. Clin. Lab. Stand. NCCLS, M7-A5.
21. National Committee for Clinical Laboratory Standards, 1997. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts ; Approved Standard. Wayne, PA Natl. Comm. Clin. Lab. Stand. NCCLS, M27–A.
22. Newman, D.J., Cragg, G.M., 2016. Natural products as sources of new drugs from 1981 to 2014. J. Nat. Prod. 79, 629–661.
23. Nielsen, K.F., Månsson, M., Rank, C., Frisvad, J.C., Larsen, T.O., Ma, M., Rank, C., Frisvad, J.C., Larsen, T.O., 2011. Dereplication of microbial natural products by LC-DAD-TOFMS. J. Nat. Prod. 74, 2338–2348.
24. Nielsen, K.F., Smedsgaard, J., 2003. Fungal metabolite screening : database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography – UV – mass spectrometry methodology. J. Chromatogr. A 1002, 111–136.
25. Oxford, A.E., Raistrick, H., Simonart, P., 1939. Studies in the biochemistry of micro-organisms: griseofulvin, C17H17O6Cl, a metabolic product of penicillium griseo-fulvum. Biochem. J. 33, 240–248.
26. Pupin, M., Fontaine, A., Jacques, P., Kucherov, G., 2008. NORINE : a database of nonribosomal peptides. Nucleic Acids Research 36, 326–331. doi:10.1093/nar/gkm792
27. Richardson, S.N., Walker, A.K., Nsiama, T.K., Mcfarlane, J., Sumarah, M.W., Ibrahim, A., Miller, J.D., 2014. Griseofulvin-producing Xylaria endophytes of *Pinus strobus* and *Vaccinium angustifolium*: evidence for a conifer-understory species endophyte ecology. Fungal Ecol. 11, 107–113. doi:10.1016/j.funeco.2014.05.004
28. Schwarzer, D., Finking, R., Marahiel, M.A., 2003. Nonribosomal peptides : from genes to products. Nat. Prod. Rep. 20, 275–287.
29. Sica, V.P., Rees, E.R., Tchegnon, E., Bardsley, R.H., Raja, H.A., Oberlies, N.H., 2016. Spatial and temporal profiling of Griseofulvin production in *Xylaria cubensis* using mass spectrometry mapping.

Front. Microbiol. 7, 1–14.

30. Smith, C. a, Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., CA Smith, J Elizabeth, G O'Maille, Ruben Abagyan, and G.S., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. ACS Publ. 78, 779–87.
31. Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
32. Strobel, G.A., 2003. Endophytes as sources of bioactive products. Microbes Infect. 5, 535–544.
33. Strobel, G., Daisy, B., 2003. Bioprospecting for microbial endophytes and their natural products Microbiol. Mol. Biol. Rev. 67, 491–502. doi:10.1128/MMBR.67.4.491
34. Sumarah, M.W., Kesting, J.R., Sørensen, D., Miller, J.D., 2011. Antifungal metabolites from fungal endophytes of *Pinus strobus*. Phytochemistry 72, 1833–1837.
35. Swofford, D.L., 2002. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4., Sinauer Associates, Sunderland, Massachusetts. doi:10.1007/BF02198856
36. Tan, R.X., Zou, W.X., 2001. Endophytes: a rich source of functional metabolites. Nat. Prod. Rep. 18, 448–459.
37. Walsh, C.T., 2015. A chemocentric view of the natural product inventory. Nat. Chem. Biol. 11, 620–624.
38. Whalley, a. J.S., Edwards, R.L., 1995. Secondary metabolites and systematic arrangement within the Xylariaceae. Can. J. Bot. 73, 802–810.
39. Worley, B., Powers, R., 2012. Multivariate analysis in metabolomics. Curr. Metabolomics 1, 92–107.
40. Wu, W., Dai, H., Bao, L., Ren, B., Lu, J., Luo, Y., Guo, L., Zhang, L., Liu, H., 2011. Isolation and structural elucidation of proline-containing cyclopentapeptides from an endolichenic Xylaria sp. J. Nat. Prod. 74, 1303–1308.
41. Yuan-Soon Ho, J.-S.D., Jeng, J.-H., , Ying-Jan Wang , Yu-Chih Liang , Chien-Huang Lin, C.-J.T., YU, C.-F., CHEN, R.-J., Lin, and J.-K., 2001. Griseofulvin potentiates antitumorigenesis effects of nocodazole through induction of apoptosis and G2 / M Cell. Int. J. Cancer 91, 393–401.

CHAPTER SIX

EPOXYNEMANIONE A, NEMANIFURANONES A-F, AND NEMANILACTONES A-C FROM *NEMANIA SERPENS*, AN ENDOPHYTIC FUNGUS ISOLATED FROM RIESLING GRAPEVINES

CHAPTER SIX PREFACE

The following works was previously published in to Phytochemistry.

Ibrahim, A., Sørensen, D., Jenkins, H. A., Ejim, L., Capretta, A., Sumarah, M.W. Epoxynemanione A, Nemanifuranones A-F, and Nemanilactones A-C, from *Nemania serpens*, an Endophytic Fungus Isolated from Riesling Grapevines. Phytochemistry 140, 16-26.

http://www.sciencedirect.com/science/article/pii/S0031942217301589

Author Contributions:

A.I. performed bioactivity screening of the crude fungal endophyte extracts, performed LC-MS analysis, performed semi-preparative HPLC isolations, crystalized compound and performed structural elucidation (1D and 2D NMR) for all compounds. D.S. performed structural validation. H.A.J. performed single crystal X-ray analysis. L.E. performed antimicrobial testing of the isolated and purified compounds. A.I wrote the manuscript and M.W.S, A.C and D.S. edited the manuscript.

Acknowledgments:

This work is dedicated to the late Professor Brian E. McCarry (1946–2013). The authors would like to thank Professor J.D. Miller and Dr. D. McMullin (Carleton University) and Dr. K. Burgess (AAFC, London) for helpful suggestions for this manuscript; Drs. K. Seifert and Joey Tanney (AAFC, Ottawa) for taxonomic expertise; T. McDowell and M. Kelman (AAFC, London) provided critical technical support; Dr. J. Britten (McMaster University) for X-ray expertise. We thank The Center for Microbial Chemical Biology (CMCB), and the Biointerfaces Institutes (BI) at McMaster University for access to state-of-the-art instrumentation. A.I was funded through an Ontario Graduate Scholarship (OGS) Doctoral Research Award. The work was funded by an AAFC A-base grant to MWS.

ABSTRACT

Ten polyketide specialized metabolites, epoxynemanione A, nemanifuranones A–F, and nemanilactones A–C, were isolated from the culture filtrate of *Nemania serpens* (Pers.) Gray (1821), an endophytic fungus from a Riesling grapevine (*Vitis vinifera*) found in Canada's Niagara region. Additionally, four known metabolites 2-(hydroxymethyl)-3-methoxy-benzoic acid, phyllostine, 5-methylmellein and a nordammarane triterpenoid were isolated. A related known metabolite 2,3-dihydro-2-hydroxy-2,4-dimethyl-5-*trans*-propenylfuran-3-one has also been included for structural and biological comparison to the nemanifuranones. The latter was isolated from the culture filtrates of *Mollisia nigrescens,* an endophytic fungus from the leaves and stems of lowbush blueberry (*Vaccinium angustifolium*) found in the Acadian forest of Nova Scotia, Canada. Their structures were elucidated based on 1D and 2D NMR, HRESIMS measurements, X-ray crystallographic analysis of nemanifuranone A, the nordammarane triterpenoid and 2,3-dihydro-2-hydroxy-2,4-dimethyl-5-*trans*-propenylfuran-3-one compounds, and comparison of NOE and vicinal ^{1}H-^{1}H coupling constants to literature data for relative stereochemical assignments. Nemanifuranone A possesses a rare C2 hemiacetal and was active against both Gram-negative and Gram-positive bacteria.

Keywords: Vitis vinifera, Vitaceae, Nemania serpens, Xylariaceae, Fungal endophyte, Specialized metabolites, Epoxynemanione, Nemanifuranone, Nemanilactone, Antibiotic

1. INTRODUCTION

Specialized metabolites from fungal endophytes have emerged as potential new sources for previously undescribed bioactive compounds. In some cases, it is known that such compounds play an important role in the chemical ecology of host plants and in their defense mechanisms against pathogens and insect herbivory (Strobel et al., 2003; Strobel, 2003). While endophytes from fescue grasses have been well studied, there is limited information on the endophytes of agriculturally important flowering plants, including their specialized metabolites. A few examples of these plant families include the Rosaceae (raspberry, genus *Rubus*), Vitaceae (grapevine, genus *Vitis*), and Ericaceae (blueberry and cranberry, genus *Vaccinium*) (Kuldau and Bacon, 2008; Zhang et al., 2006).

In earlier studies, it was shown that conifer endophytes produce a number of specialized metabolites that are toxic agents to the spruce budworm and provide protection to seedlings against deforestation pests (Sumarah et al., 2011, 2010). Insights into the chemical ecology of a pine-blueberry forest ecotype were also gained. As reported, the potent antifungal agent griseofulvin was produced by endophytic *Xylaria* species, isolated from white pine (*Pinus strobus*) needles and low bush blueberries (*Vaccinium angustifolium*), demonstrating that the lifecycles of fungi and plant microbiomes are more complex than previously thought (Richardson et al., 2014).

Expanding on these efforts, previously undescribed diplosporin and agistatine derivatives produced from an endophytic *Xylaria* sp., isolated from a Concord grape leaf (*Vitis labrusca*) were reported (Ibrahim et al., 2014). The yearly production of grapes in Canada is approximately 81,000 metric tonnes, of which 95% are from *Vitis vinifera* and 5% from *Vitis labrusca* (Agriculture and Agri-Food Canada, 2013). The production of organic grapes has also increased significantly over the past decade in Canada and globally, with increasing concern for

crop protection against disease and insect pests without the use of synthetic pesticides (Hopeross, 2006; Pedneault and Provost, 2016; Provost and Pedneault, 2016). These findings spurred further investigation into the specialized metabolites of endophytic fungi from *Vitis vinifera* and their biological activity. In total, seven endophytes were recovered from the grapevines and leaves of Riesling grape, maintained in a fungicide free area located on Jordan Research Farm (Agriculture and Agri-food Canada). ITS DNA sequencing identified a number of taxa including *Anthostomella, Coniothyrium ferrarisianum, Creosphaeria sassafra, Hypoxylon submonticulosum, Nemania serpens* (Pers.) Grey (1821)*, Xylaria* sp., and unidentified *Sordariomycetes*. We further investigated *Nemania serpens* (Pers.) Grey (1821), in the Xylariaceae family, as the crude filtrate and mycelia extracts showed biological activity against *B. subtilis*, *M. luteus* and *S. aureus* bacteria in preliminary screening.

This study reports herein the isolation and characterization of ten previously undescribed (**1**–**10**) and four known (**11**–**14**) specialized metabolites from the endophytic fungus *Nemania serpens* and (**15**) from *Mollisia nigrescens.* Epoxynemanione A (**1**) is a previously undescribed polyketide metabolite with a bicyclic ring system, containing two oxymethine groups, a pyrone ring and a unique epoxide moiety, structurally similar to the antifungal compound epoxyroussoeone (Honmura et al., 2015). Nemanifuranones A–F (**2**–**7**) are a previously undescribed series of specialized metabolites containing a rare C2 hemiacetal within the 4-methyl-5-alkenyl- and alkane -3(2H)-furanone ring system, exhibiting shortened alkyl chains and furanone-lactone coupled derivatives. Nemanilactones A–C (**8**–**10**) are previously undescribed lactone polyketides, including lactone-lactone derivatives (Fig. 1). The isolated compounds were tested for preliminary biological activity against a panel of Gram-negative and Gram-positive bacteria, including several yeasts.

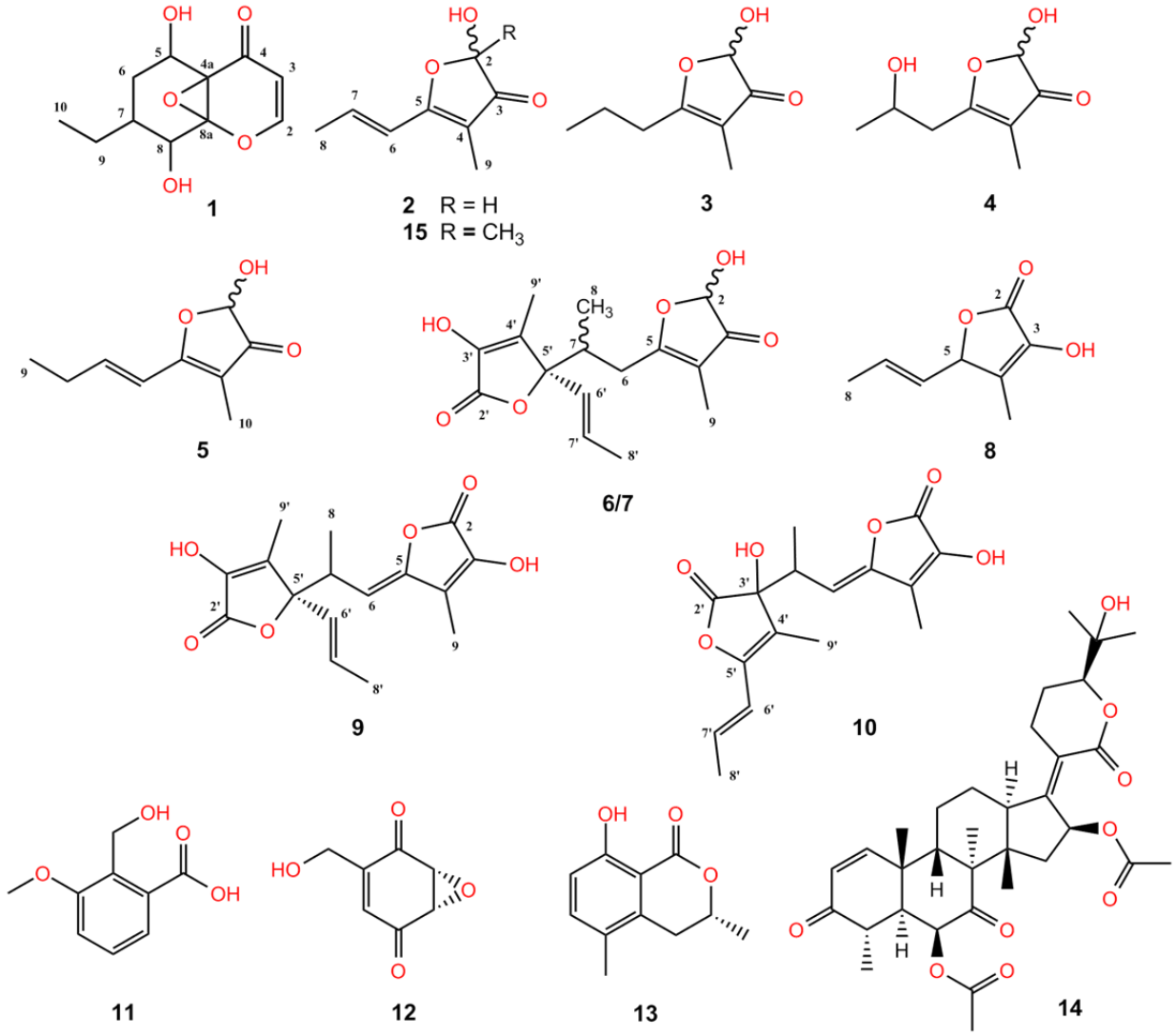


Fig.1. Structures of compounds 1-15

RESULTS AND DISCUSSION

2. Results and discussion

2.1 Isolation and structure determination

The endophytic fungus *Nemania serpens* (Pers.) Grey (1821) was isolated from the surface sterilized grapevines and leaves of Riesling variety grape plants located at Jordon Research Farms in Ontario, Canada. A voucher specimen was deposited in the Canadian Collection of Fungal Cultures (CCFC) as DAOMC 251481 and the sequence data was deposited in GenBank (accession # KY713249). Epoxynemanione A (**1**), nemanifuranones A–F (**2**–**7**) and nemanilactones A–C (**8**–**10**) were isolated from the culture filtrates by ethyl acetate extraction. The ethyl acetate extract was subjected to semi-preparative reversed-phase HPLC fractionation yielding metabolites **1**-**2**, **4**, and **8**. Metabolites **3**, **5**-**7**, and **9**-**10** required a second semi-preparative reversed-phase HPLC purification to resolve co-eluting peaks.

Epoxynemanione A (**1**) was isolated as a golden-brown solid. Its HRESIMS afforded the molecular ion at *m/z* 227.0810, and a more abundant sodium adduct ion at *m/z* 249.0736 ($C_{11}H_{14}O_5Na$), thus establishing its molecular formula as $C_{11}H_{14}O_5$ indicative of 5 double bond equivalents (SI Fig. 1). The ^{1}H NMR spectrum (Table 1) indicated the presence of two olefinic methine protons at δ 7.48 (d, J = 6.4 Hz) and δ 5.60 (d, J = 6.4 Hz), two oxymethine protons at δ 4.10 (dd, J = 9.4, 5.9 Hz) and δ 3.93 (d, J = 7.9 Hz), a methine at δ 1.14 (m), two methylene groups at δ 1.63a, 1.13b (dq, J = 12.5, 7.4 Hz, m) and δ 1.84a, 1.07b (ddd, J = 13.2, 5.9, 2.0 Hz, ddd, J = 13.2, 11.5, 9.4 Hz) and a methyl group at δ 0.83 (t, J = 7.3 Hz). Examination of the ^{13}C DEPTq and multiplicity edited HSQC indicated compound (**1**) possessed 11 carbon atoms. Resonances were attributed to a carbonyl group at C4 (189.3 ppm), a ring epoxide system at C4a/C8a (68.4 and 89.8 ppm), two oxymethine protons at C5 and C8 (64.7, 66.4 ppm), two

methylenes at C6 and C9 (32.7, 23.5 ppm), three methines at C2, C3 and C7 (157.5, 106.8, 36.3 ppm) and a methyl group at C10 (10.9 ppm). A pyrone ring system was evident from α/β carbon assignments of the two olefinic methines at H2 [δ 7.48 (d, J = 6.4 Hz)] and H3 [δ 5.60 (d, J = 6.4 Hz)] being conjugated to the carbonyl group and both showed COSY connectivity. An HMBC correlation further supported the pyrone ring systems with connectivities of H2 to C8a, H3 to C4a and to the C4 carbonyl, thus accounting for three degrees of unsaturation (Fig. 2). The presence of only two other quaternary carbons at C4a and C8a indicated an epoxide ring system neighboring the two oxymethines, with a 1,4-substituted configuration. A second ring system accounted for the two remaining degrees of unsaturation and was established from the key HMBC correlations of the oxymethine proton at H5 [δ 4.10 (dd, J = 9.4, 5.9 Hz)] to C4a, C6 and C7, and the second oxymethine proton H8 [δ 3.93 (d, J = 7.9 Hz)] to the methine at C7 and methylene at C9. The methyl triplet group at H10 [δ 0.83 (t, J = 7.3 Hz)] showed a COSY connectivity to the methylene at H9 [δ 1.63a, 1.13b (dq, J = 12.5, 7.4 Hz), m] and showed HMBC correlations to C7 and C9, thus indicating the alkyl group position on the second ring system (Fig. 2).

TOCSY connectivities further supported the second ring system showing H5 to H7, and subsequently H7 to H9-H10 and H8 connectivities (SI Fig. S2-S14). The relative stereo-configuration was established based on NOESY and ROESY correlations and by comparison to the structurally similar epoxyroussoeone (Honmura et al., 2015). Epoxynemanione A (**1**) and epoxyroussoeone both contain pyrone and epoxide ring substructures, with comparable epoxide quaternary carbons at δ 68.4 *vs.* 61.8 and 89.8 *vs.* 88.2 ppm, and two oxymethines at 64.7 *vs.* 65.1 ppm and 66.4 *vs.* 68.5 ppm, respectively. The structures differ in that epoxyroussoeone has a benzene ring and a methyl group at C2. NOESY and ROESY correlations for epoxynemanione A (**1**) could be observed between H5 to H6a/b, H8 to H6b, H8 to H9a, but no

observable correlations between H7 and H5 or H7 to H8. Based on these findings, it is speculated that the oxymethine protons at H5 and H8 are in a "syn" orientation, similar to that of epoxyroussoeone, and that H7 to H8 are "anti" to each other.

Table 1. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data for epoxynemanione A (**1**).

DMSO-d_6				CD_3OD		
Position	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC
2	157.5, CH	7.48, d (6.4)	3, 4, 8a	158.7, CH	7.33, d (6.4)	3, 4, 8a
3	106.8, CH	5.64, d (6.4)	2, 4a	107.8, CH	5.63, d (6.4)	2, 4a
4	189.3, C			191.6, C		
4a	68.4, C			69.3, C		
5	64.7, CH	4.1, dd (9.4, 5.9)	4a, 6, 7,8a	66.6, CH	4.23, dd (9.8, 6.2)	4, 4a, 6, 7,8a
6	32.7, CH2	1.84, ddd (13.2, 5.9, 2.0) 1.07, ddd (13.2, 11.5, 9.4)	4a, 5, 7, 8, 8a, 9	34.2, CH2	2.01, ddd (13.4, 6.1,2.5) 1.11, ddd (13.4, 12.2, 9.8)	4a, 5, 7, 8, 8a, 9
7	36.3, CH	1.14, m	5, 6 , 8, 9	37.3, CH	1.32 dddd (17.9, 8.8, 3.5, 2.5)	5, 6, 8, 8a, 9
8	66.4, CH	3.93, d (7.9)	4a, 7, 9	69.2, CH	4.07, d (9.1)	4a, 7, 9
8a	89. 8, C			91.3, C		
9	23.5, CH2	1.63, dq (12.50, 7.4) 1.13, m	6, 7, 8, 10	25.0, CH2	1.79, dtd, (17.5, 7.5, 3.7) 1.24, ddq (13.3, 8.4, 7.2)	6, 7, 8, 10
10	10.9, CH3	0.83, t (7.3)	7, 9	11.2, CH3	0.93, t (7.5)	7, 9

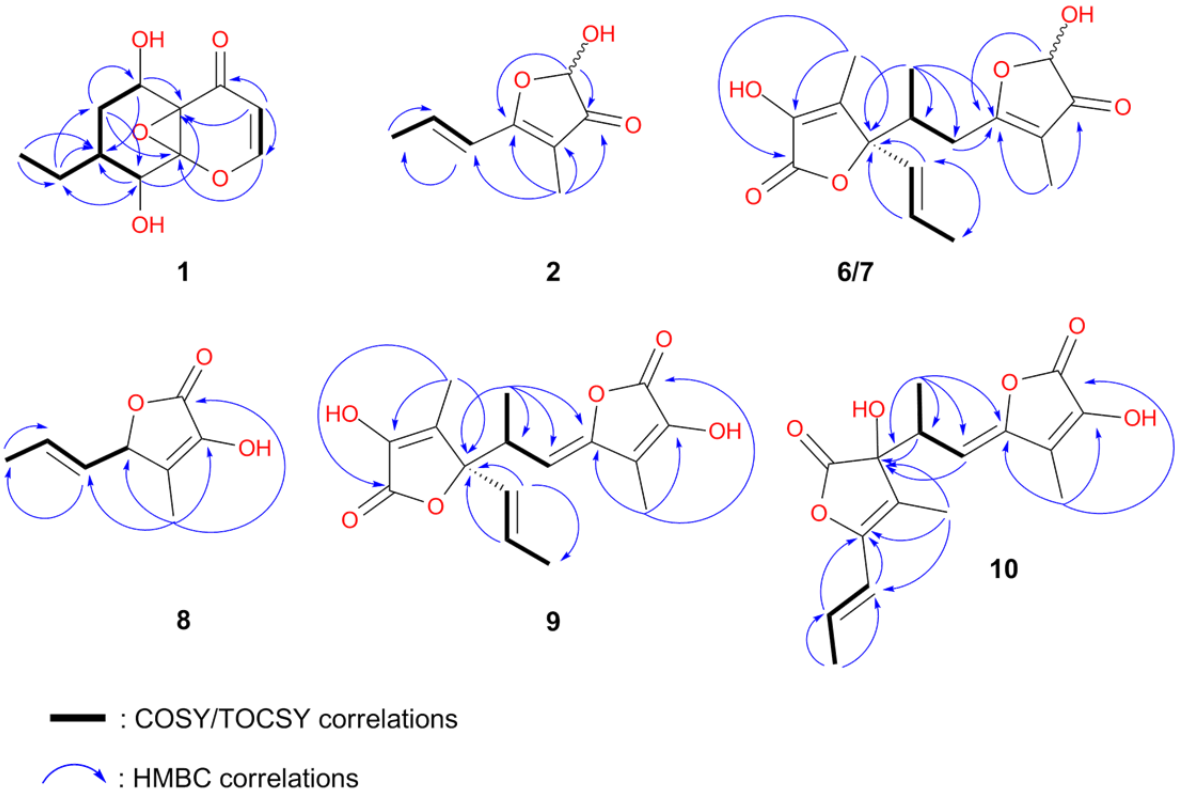


Figure 2. Select HMBC and COSY/TOCSY correlations for structures 1-10

Nemanifuranone A (**2**) was the most abundant metabolite produced by *N. serpens* and was isolated as colorless crystals from the HPLC fractions containing a mixture of acetonitrile and H_2O with 0.1% formic acid (v/v). Its HRESIMS spectra showed a protonated molecular ion at *m/z* 155.0705, affording a formula of $C_8H_{10}O_3$ with four double bond equivalents. The ^1H NMR spectrum indicated the presence of two olefinic protons at δ 6.82 (dq, J = 15.6, 6.9 Hz) and δ 6.51 (dq, J = 15.5, 1.7 Hz), one oxymethine at δ 5.32 (s), and two methyl groups at δ 1.97 (dd, J = 6.9, 1.7 Hz) and δ 1.67 (s) (Table 2, SI Fig. S15–S20). The core structure of (**2**) became evident as a 5-alkenyl-3(2H)-furanone based on the ^{13}C DEPTq and HSQC assignments, with quaternary carbons at C3, C4 and C5 (201.4, 109.1 and 177.7 ppm), and a hemiacetal group at C2 (96.4 ppm). To our knowledge, the hemiacetal moiety at the C2 position is the first report from an endophytic fungi containing the 5-alkenyl-3(2H)-furanone core. A hemiketal moiety has been reported from bacterial-derived compounds, such as the actinofuranones A/B, aurafuron A/B, E-492, E837, and E-975, hyafurones and linfuranone A, with few notable exceptions such as AS-183 isolated from the fungus *Scedosporium* sp. SPC-15549 and polypropionates 1–3 from marine mollusks (Banskota et al., 2006; Bromley et al., 2012; Cho et al., 2006; Indananda et al., 2013; Kunze et al., 2005; Kuroda et al., 1993; Okanya et al., 2014). HMBC correlations of the oxymethine proton at C2 to the carbonyl at C3 and carbon at C5 supported the presence of a hemiacetal moiety, with the methyl singlet at C9 (5.4 ppm) having correlation to C3, C4 and C5, and long-range ^4J coupling to C6 (SI Fig. S18). The olefinic protons at C6 (119.9 ppm) and C7 (139.7 ppm) exhibited a COSY connectivity and in a trans (*E*) configuration based on the ^3J vicinal ^1H-^1H coupling constant of 15.5-15.6 Hz for the double bond. The methyl doublet of doublets at C8 (19.0 ppm) showed a COSY connectivity with C7 and HMBC correlations to C6 and C7, confirming the 5-alkenyl moiety (Fig. 2). The

final structure and relative stereochemistry of nemanifuranone A (**2**) was confirmed based on single crystal X-ray data and found to be a 1:1 racemic mixture at the C2 hemiacetal (Fig. 3).

Nemanifuranone B (**3**) was isolated as a beige solid and was a minor product relative to the other compounds. HRESIMS for the protonated species afforded an *m/z* at 157.0857, providing a molecular formula of $C_8H_{12}O_3$, with three double bond equivalents. The ^{1}H NMR spectrum revealed the absence of the olefinic protons as seen in (**2**), yet the oxymethine proton was clearly present at δ 5.30 (s), with two methyl groups at δ 1.55 (s) and δ 0.93 (t, J = 7.4 Hz). The major difference between (**2**) and (**3**) was the presence of two methylene groups at δ 2.48 (dd, J = 7.5, 1.8 Hz) and δ 1.60 (h, J = 7.4 Hz), indicating that the double bond had been reduced (Table 2). The ^{13}C DEPTq NMR and multiplicity-edited HSQC spectra showed the presence of eight carbon atoms, with three quaternary carbons at C3, C4, and C5 (200.9, 108.2, 186.8 ppm), a hemiacetal at C2 (96.3 ppm) to complete the furanone moiety, with two methylenes at C6 and C7 (29.9, 19.0 ppm), and two methyl groups at C8 and C9 (13.4, 5.3 ppm). The alkyl chain moiety was evident from the COSY connectivities of the C8 methyl triplet to the methylene C7, and C7 to both C6 and C8, including HMBC long-range coupling of C6 to C5 and C9 (Fig. 2, SI Fig. S21–S27).

Nemanifuranone C (**4**) was isolated as a golden oil and was the most polar compound of the nemanifuranone family. Its HRESIMS showed a protonated species at *m/z* 173.0812, affording a molecular formula of $C_8H_{12}O_4$, indicating three degrees of unsaturation. Similar to (**3**), no olefinic protons were present in the ^{1}H NMR spectrum, but an additional oxymethine proton was observed at δ 3.98 (ddd, J = 13.5, 6.6, 3.4 Hz), with the C2 hemiacetal at δ 5.28 (s) (Table 2). Additionally, one methylene group was observed at δ 2.63 (td, J = 13.9, 6.9 Hz) with two methyl groups at δ 1.56 (s) and δ 1.13 (d, J = 6.2 Hz). Examination of the ^{13}C NMR and multiplicity-edited HSQC spectra showed the presence of two sets of signals for the 8 carbon

atoms, indicating the presence of an isomer or a second diastereomer (SI Fig. S27–S34). The furanone moiety was intact with quaternary carbons at C3, C4, and C5 (201.0/200.9 ppm, 109.3/109.2 ppm, 184.9/184.9 ppm), and the hemiacetal at C2 (96.3/96.3 ppm). The second oxymethine proton can be attributed to the C7 (64.2/64.1 ppm) position based on the HMBC correlations and COSY connectivities to the methylene at C6 (38.7 ppm) and the methyl doublet at C8 (23.4/23.4 ppm). The presence of a fixed chiral center at C6 and epimers at C2 results in two diasteriomers, and hence the equivalent intensity of the two sets of carbon signals. This has also been observed with several compounds containing a 5-alkenyl-3, 3(2H)-furanone moiety (Banskota et al., 2006; Bromley et al., 2012; Okanya et al., 2014).

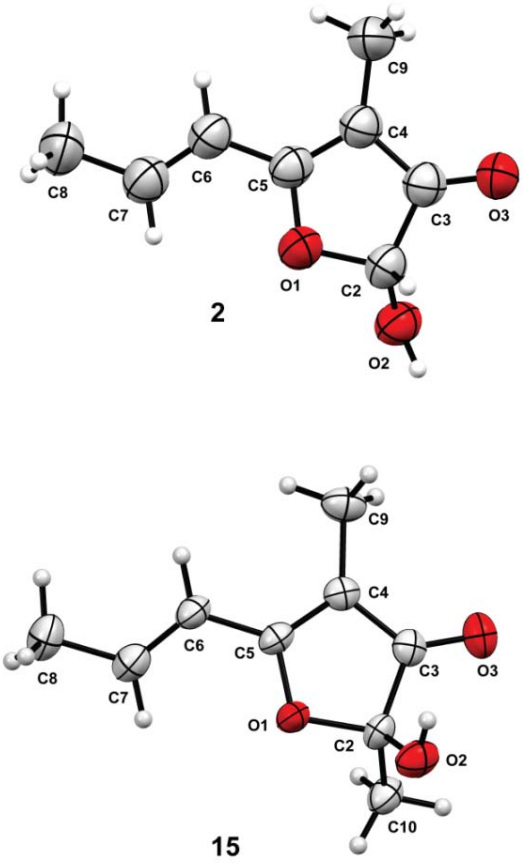


Figure 3. ORTEP single crystal X-Ray structures of **2** and **15.**

Table 2. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of nemanifuranones A–C (**2–4**)

	Nemanifuranone A (**2**)			Nemanifuranone B (**3**)			Nemanifuranone C (**4**)		
	CD_3CN			DMSO-d_6			DMSO-d_6		
Position	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC
2	96.4, CH	5.32, s	3, 5	96.3, CH	5.30, s	5	96.3, 96.3, CH	5.28, s	5
3	201.4, C			200.9, C			201.0, 201.0, C		
4	109.1, C			108.28, C			109.3, 109.2, C		
5	177.7, C			186.8, C			185.0, 184.9, C		
6	119.9, CH	6.51, dq (15.5, 1.7)	5, 7, 8	30.0, CH_2	2.48, dd (7.5, 1.8)	4, 5, 7, 8	38.7, 38.7, CH_2	2.63, td (13.9, 6.9)	4, 5, 7, 8
7	139.7, CH	6.82, dq (15.6, 6.9)	5, 8	19.03, CH_2	1.60, h (7.4)	5, 6 , 8	64.2, 64.1, CH	3.98, ddd (13.5, 6.6, 3.4)	5, 6 , 8
8	19.0, CH_3	1.97, dd (6.9, 1.7)	5, 6, 7	13.43, CH_3	0.93, t (7.4)	6, 7	23.4, 23.4, CH_3	1.13, d (6.2)	6, 7
9	5.4, CH_3	1.67, s	3, 4, 5, 6	5.3, CH_3	1.55, s	3, 4, 5	5.47,CH_3	1.56, s	3, 4, 5

Nemanifuranone D (**5**) was isolated as a beige solid with the HRESIMS spectrum giving the protonated ion at *m/z* 169.0864, and a sodium adduct ion at *m/z* 191.0681 affording a molecular formula of $C_9H_{12}O_3$ with four degrees of unsaturation (SI Fig. 35). Its ^{1}H NMR spectrum showed the presence of two olefinic protons at δ 6.64 (dt, J = 15.6, 6.7 Hz) and δ 5.90 (dt, J = 15.6, 1.7 Hz), one oxymethine proton at δ 5.14 (s), a methylene at δ 1.79 (pd, J = 7.4, 1.7 Hz), and two methyl groups at δ 1.52 (s) and δ 0.74 (t, J = 7.5 Hz) (Table 3). This indicated a butyl chain at the 5 position in (**5**) as compared to the propenyl evident in (**2**). Its ^{13}C NMR spectra showed the furanone moiety at C2, C3, C4, C5 and C9 (96.2, 200.8, 108.8, 177.2, 5.w ppm). HMBC correlations between the methyl singlet at C10 to quaternary carbons at C3, C4, and C5; and the C2 hemiacetal to C5 were also evident. In the HMBC, the olefinic protons at C6 (116.9 ppm) showed a correlation with C5, while C7 (144.5 ppm) showed a correlation to C5 and olefinic C6 (SI Fig. 27). Similarly, COSY correlations were found between C7 to H6 and the methylene at C8 (Fig. 2, SI Fig. S36-S40). The double bond configuration was established to be trans (*E*) based on the ^{3}J vicinal ^{1}H-^{1}H coupling constant of 15.5-15.6 Hz, also observed in (**2**) and (**15**). The methyl triplet at C9 (12.5 ppm) had a COSY and HMBC correlation to the methylene at C8, thus confirming the extended alkyl chain.

Table 3. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of nemanifuranone D (**5**) and nemanilactone A (**8**) in C_6D_6.

	Nemanifuranone D (**5**)			Nemanilactone A (**8**)		
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
2	96.2, CH	5.14, s	5	170.5, C		
3	200.8, C			138.0, C		
4	108.8, C			129.7, C		
5	177.2, C			82.5, CH	4.35, dt (8.5, 1.2)	3, 4, 6, 7
6	116.9, CH	5.90, dt (15.6, 1.7)	5, 8	126.7, CH	4.66, ddq (15.3, 8.5, 1.7)	5, 8
7	144.5, CH	6.64, dt (15.6, 6.7)	5	132.9, CH	5.29, dqd (14.0, 6.6 , 0.8)	5, 8
8	26.3, CH_2	1.79, pd (7.4, 1.7)	6, 7, 9	17.5, CH_3	1.29, dd (6.6, 1.7)	6, 7
9	12.5, CH_3	0.74, t (7.5)	7, 8	9.0, CH_3	1.34, d (1.3)	2, 3, 4, 5
10	5.4, CH_3	1.52, s	2, 4, 5			

Nemanifuranone E–F (**6–7**) were minor products isolated separately as beige solids, with HPLC elution times differing by ~ 1.2 min. Their HRESIMS spectra gave identical molecular ions at *m/z* 309.1334, and a sodiated ion adduct at *m/z* 331.1157, both affording the same molecular formula of $C_{16}H_{20}O_6$, with seven degrees of unsaturation. Given the identical formulae, but differing retention times, this indicated two diasteriomers. The ^{1}H NMR spectrum of (**6**) and (**7**) showed two olefinic protons at δ 5.61/5.62 and δ 5.74/5.81, one oxymethine proton at δ 5.31/5.33, one methylene at δ 2.34, 2.27 /2.85, 2.50, one methine at δ 2.50/2.51 and three methyl groups at δ 1.86/1.84, δ 1.74/1.73, and δ 1.08/0.76 ppm (Table 4). Analysis of the ^{13}C NMR and multiplicity-edited HSQC spectra indicated 16 carbon resonances, with seven quaternary carbons (203.9/204.0, 188.3/189.1, 171.1/171.2, 139.0/138.8, 135.04/135.1, 111.3/111.4, and 90.1/89.9). The carbon resonances at 98.1/98.0 ppm were indicative of a hemiacetal at C2. Long-range HMBC correlations of C2 to C5 (188.3/189.1 ppm) and the methyl singlet at C9 (5.6/5.6 ppm) to the C3 carbonyl and C4, C5 quaternary carbons confirmed a furanone moiety, accounting for three double bond equivalents (SI Fig. S41-S53).

One of the remaining double bond equivalents could be assigned to a short acyl chain based on the olefinic protons, while the three remaining degrees of unsaturation indicated the presence of a second ring system. HMBC correlations of the methyl singlet C9′ (9.2/9.0 ppm) to the four quaternary carbons at C2′ (171.1 ppm), C3′ (135.0/135.1 ppm), C4′ (139.0/138.8 ppm), and C5′ (90.1/89.9 ppm) supported the presence of a lactone ring (Fig. 2). The coupling of the furanone and lactone ring was established based on the key HMBC correlations of the methyl doublet at C8 (15.0/13.3 ppm) to the methylene at C6, the methine proton at C7 (36.5/36.5 ppm), and the lactone quaternary at C5′ (90.1/89.9 ppm), and was further corroborated by COSY and TOCSY connectivities. The alkyl chain was unambiguously determined to be attached to the lactone ring at C5′, with key HMBC and COSY/TOCSY correlations of the olefinic C6′ (129.6/129.8 ppm) and C7′ (127.6/128.1 ppm) to C5′, and the methyl doublet of doublets at C8′ (17.9/17.9 ppm) to C7′ and C6′. The alkyl double bond was consistent with (**2**) and (**5**) having an *E* configuration based the ^{3}J vicinal ^{1}H-^{1}H coupling constant of 15.3-15.7 Hz and accounted for the final degree of unsaturation.

The difference between (**6**) and (**7**) can be seen by the chemical shift differences at the C6 methylene δ 2.34a, 2.27b vs. δ 2.85a, 2.50b and the C8 methyl doublet at δ 1.08 vs δ 0.76. Analysis of the ROESY spectrum revealed the key connectives of H9′ to H7, H7 to H6a, H8 to H6b, and H7 to H6′, which are indicative of a fixed stereochemistry at C5′ and a relative stereochemical orientation of “anti” for compound (**6**) and “syn” for compound (**7**) relative to C5′. Examination of the lactone ring system shows similarity to the known compound serpenone, a 3-methoxy-4-methyl-5-(prop-1-enyl) furan-2(5H)-one isolated from the soil fungus *Hypoxylon serpens* in Ontario, Canada (Anderson et al., 1982).

Table 4. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of Nemanifuranone E-F (**6**-**7**) in CD_3OD.

Nemanifuranone E (**6**)				Nemanifuranone F (**7**)		
Position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
2	98.0, CH	5.31, s	5	98.1, CH	5.33, s	5
3	204.0, C			204.0, C		
4	111.3, C			111.4, C		
5	188.3, C			189.1, C		
6	30.80, CH_2	2.34, dd (14.2, 4.5)	4, 5, 7, 5'	30.8, CH_2	2.85, t (14.5)	
		2.27, dd (14.3, 9.4)			2.50, m	
7	36.5, CH	2.50, ddd (9.3,7.0,4.5)		36.5, CH	2.51, m	
8	15.0, CH_3	1.08, d (6.9)	6, 7, 5'	13.3, CH_3	0.76, d (6.4)	6, 7, 5'
9	5.6, CH_3	1.60, s	3, 4, 5	5.6, CH_3	1.63, s	3, 4, 5
2'	171.1, C			171.2, C		
3'	135.0, C			135.1, C		
4'	139.0, C			138.8, C		
5'	90.1, C			90.0, C		
6'	129.6, CH	5.61, dq (15.3, 1.6)	5', 8'	129.8, CH	5.62, dq (15.7, 1.7)	5', 8'
7'	127.6, CH	5.74, dq (15.5, 6.6)	5', 8'	128.1, CH	5.81, dq (15.5, 6.6)	5', 8'
8'	17.9, CH_3	1.74, dd (6.6, 1.6)	6', 7'	17.9, CH_3	1.73, dd (6.6, 1.7)	6', 7'
9'	9.2, CH_3	1.86, s	2', 3', 4', 5'	9.0, CH_3	1.84, s	2', 3', 4', 5'

Nemanilactone A (**8**) was isolated as a golden oil. Its HRESIMS gave a molecular ion at *m/z* 155.0703, affording a molecular formula of $C_8H_{10}O_3$, with four degrees of unsaturation (SI Fig. 54). Its ^{1}H NMR spectrum showed olefinic resonances at δ 5.29 (dqd, J = 14.0, 6.6, 0.8 Hz) and δ 4.66 (ddq, J = 15.3, 8.5, 1.7 Hz), with an oxymethine at δ 4.35 (dt, J = 8.5, 1.2 Hz) and two methyl groups at δ 1.34 (d, J = 1.3 Hz) and δ 1.29 (dd, J = 6.6, 1.7 Hz). The ^{13}C NMR and multiplicity-edited HSQC spectra revealed 9 carbon atoms with a carbonyl at 170.5 ppm, two additional quaternary carbons at 138.0, and 129.7 ppm, olefinic methines at 132.9 and 126.7 ppm, an oxymethine proton at 82.5 ppm, and two methyl groups at 17.5 and 9.0 ppm (Table 3). Examination of the multiplicity-edited HSQC-TOCSY and COSY connectivities showed the methyl doublet at C8 (17.5 ppm) connected to C7 (132.9 ppm) and subsequently to C6 (129.9 ppm) and the oxymethine proton at C5 (82.5 ppm). Based on the ^{3}J vicinal ^{1}H-^{1}H coupling constant of 14.0–15.3 Hz, the alkenyl group was established to be a *trans* configuration. Key HMBC correlations between C9 to C2, C3, C4 and the oxymethine proton at C5, supported the

presence of a lactone core (Fig. 2, SI Fig. S54-S60). Nemanifuranones E–F (**6/7**) identified earlier are likely products of compounds (**2**) and (**8**). Nemanilactone A (**8**) is a previously undescribed structural analogue of the known compound serpenone, which possesses a hydroxyl moiety at C3 instead of a methoxy group (Anderson et al., 1982).

Nemanilactone B (**9**) was a minor product that was isolated as a beige solid. Its molecular formula of $C_{16}H_{18}O_6$ was established from the sodiated adduct in the HRESIMS spectrum at *m/z* 329.0998, and indicated eight double bond equivalents. Its ^{13}C NMR and multiplicity-edited HSQC revealed 16 carbon atoms, with assignments similar to the lactone core (171.4, 136.6, 138.5, 89.95 ppm) and an alkyl chain with the olefinic methines (129.2, 128.1, 18.0 ppm), as seen in (**6/7**) and in (**8**) (Table 5). Key HMBC correlations confirmed the lactone group with the methyl singlet at δ1.79/9.0 ppm correlating as expected to the carbonyl at 171.4 ppm and the two quaternary carbons at 136.8 (hydroxyl) and 138.5 ppm (Fig. 2, SI Fig. S61–S68). The alkenyl moiety was assigned an *E* configuration based on the ^{3}J vicinal ^{1}H-^{1}H coupling constant of 15.5-15.5 Hz, and was connected to the quaternary at C5′ (89.9 ppm). The presence of additional quaternary carbons at 166.6, 142.4, 121.9 and 150.2 ppm, suggested the presence of a second lactone group, and accounted for the remaining four double bond equivalents. The methyl singlet at δ1.84/7.1 ppm showed HMBC correlations to the C2 carbonyl at 166.6 ppm and to C3, C4 and C5. The final methyl doublet at δ1.15/16.0 ppm had a COSY coupling to the C7 methine at δ3.34/36.5 ppm, and HMBC correlations to C6, C7 and C5′. The C6 methine at δ4.64/106.9 ppm correlated to C5, C7 and C8, thus connecting the two lactone systems. ROESY correlations show coupling between the methyl at H8 to the methine at H6, and H6′ to H9′, indicating the orientation is similar to (**6/7**). Efforts to crystallize compound (**9**) proved unsuccessful in acetone, acetonitrile, benzene, methanol and chloroform, with gradual degradation over time.

Nemanilactone C (**10**) was isolated as a beige solid, eluting ~1 min after (**9**). Its HRESIMS spectrum gave a protonated parent ion at *m/z* 307.1180, giving a molecular formula of $C_{16}H_{18}O_6$ with eight double bond equivalents, identical to (**9**). Its ^{1}H NMR spectrum showed resonances at δ 6.23 (dq, J = 15.6, 1.4 Hz) and δ 6.18 (dq, J = 15.6, 6.3 Hz) for an olefinic group, two additional methines at δ 5.41 (d, J = 10.0 Hz) and δ 3.30 (m), and four methyl groups at δ 1.95 (s), δ 1.89 (d, J = 6.9 Hz), δ 1.80 (s), and δ 0.96 (d, J = 7.2 Hz) ppm (Table 5). The ^{13}C NMR and multiplicity-edited HSQC spectra revealed 16 carbon atoms, with similar assignments to those displayed by (**9**). The two lactone systems are also evident in compound (**10**). Thus the structures likely differ at the coupling site between the two lactone moieties. The first lactone ring was nearly identical in assignments with the C2, C3, C4 and C5 quaternary carbons (167.1, 142.1, 122.51, 150.8 ppm). The key differences arise in the HMBC correlation of methyl singlet at C9′ (δ 1.80/7.2 ppm) to C3′ (81.1 ppm), C4′ (114.5 ppm), importantly to C5′ (148.2 ppm) and the olefinic methine at C6′ (117.6 ppm). Similarly, C6′ has an HMBC correlation to C5′ (148.2 ppm), which was part of a double bond with C4′ (Fig. 2). The change in the double bond placement removes the conjugation with the carbonyl, thus shifting C4′ from 138.5 ppm in (**9**) to 114.5 ppm, and C5′ from 89.9 ppm to 148.2 ppm (SI Fig. S61, S69-S74). The *E* configuration of the alkenyl moiety was based on the ^3J vicinal ^{1}H-^{1}H coupling constant of 15.6-15.60 Hz. The two systems are similarly connected through the COSY coupling and HMBC correlations of the C6 (107.5 ppm) and C7 (37.1 ppm) methine and C8 (15.6 ppm) methyl doublet to the hydroxyl quaternary carbon at C3′ (81.1 ppm). Crystallization efforts with compound (**10**) proved unsuccessful.

Table 5. ^{1}H (700 MHz) and ^{13}C (176 MHz) NMR spectroscopic data of nemanilactone B-C (**9-10**) in CD_3OD.

Nemanilactone B (**9**)				Nemanilactone C (**10**)		
Position	δ_C	δ_H (J in Hz)	HMBC	δ_C	δ_H (J in Hz)	HMBC
2	166.6, C			167.1, C		
3	142.4, C			142.1, C		
4	121.9, C			122.5, C		
5	150.2, C			150.6, C		
6	106.9, CH	4.64 , d (10.5)	4, 5, 8, 5'	107.5, CH	5.41, d (10.0)	4, 5, 8, 3'
7	36.5, CH	3.34, dq (10.9,7.1)	6, 8 , 5'	37.1, CH	3.30, m	6, 2', 3'
8	16.0, CH3	1.15, d (6.9)	6, 7, 5'	15.6, CH3	0.96, d (7.2)	6, 7, 3'
9	7.1, CH3	1.84, s	2, 3, 4, 5	7.3, CH3	1.95, s	2, 3, 4, 5, 6
2'	171.4, C			178.4, C		
3'	136.6, C			81.1, C		
4'	138.5, C			114.5, C		
5'	89.9, C			148.2, C		
6'	129.2, CH	5.62, dq (15.5, 1.6)	5', 8'	117.6, CH	6.23, dq (15.6,1.4)	5', 7', 8'
7'	128.1, CH	5.78, dq (15.5, 6.6)	5', 8'	132.2, CH	6.18, dq (15.6,6.3)	5', 6', 8'
8'	18.0, CH3	1.75, dd (6.6, 1.7)	6', 7'	18.6, CH3	1.89, d (6.9)	5', 6', 7'
9'	9.0, CH3	1.79, s	2', 3', 4', 5'	7.2, CH3	1.80, s	3', 4', 5', 6'

The known metabolites **11-13** were isolated as minor products from the filtrate extracts, with varying yields across batches grown in PDB media (SI Fig. S100). Although previously synthesized, isolation of **(11)** herein**,** is the first report from a natural source. To our knowledge, (**11**) has only been reported as a synthetic intermediate for non-steroidal ligands used for inducing or suppressing gene expression in animal/plant cells (Carlson et al., 2001). The metabolite was isolated as a golden-yellow solid, eluting after compounds **5-7**. Its HRESIMS afforded a molecular ion at *m/z* 183.0649, thus establishing the molecular formula as $C_9H_{10}O_4$ with 5 double bond equivalents. The ^{1}H NMR and multiplicity-edited HSQC spectra revealed the presence of three aromatic protons at δ 7.57/131.0 ppm, δ 7.41/116.5 ppm, and δ 7.36/115.9 ppm, one oxymethylene proton at δ 5.36/68.0 ppm, and a methoxy group at 3.91/55.8 ppm. Analysis of the ^{13}C NMR spectrum revealed the expected nine carbon resonances with a carboxylic acid moiety at 170.5 ppm and the quaternary carbons at 154.0, 134.9 and 126.5 ppm. Key HMBC correlations of the methoxy group to the quaternary carbon at 154.0 ppm, and the

oxymethylene to the quaternary carbons at 154.0, 126.5 and 134.9 ppm, unambiguously confirmed the known structure (SI Fig. S75–S78).

Compound **12** was identified as the known phytotoxic agent phyllostine (SI Fig. S79–S82), which has been reported to cause wilting and dead spots in clover leaf tests (Sakamura et al., 1970, 1971). The discovery of phyllostine from a grape endophyte is interesting, as it has previously been isolated with the known agricultural mycotoxin patulin, which was isolated from a marine endophytic fungi (*Penicillium* sp.) (Nicoletti and Trincone, 2016).

Compound **13** was identified as 5-methylmellein (SI Fig. S83-S88), a common fungal metabolite, and one that we previously isolated from foliar endophytes of *Picea glauca* (white spruce) (Sumarah et al., 2008).

Compound **14** was the only metabolite isolated from the mycelium. Single crystal X-ray analysis and 1D and 2D NMR assignments confirmed the metabolite as a recently reported nordammarane triterpenoid (Fig. 4, SI Fig. S89–S93) (Afiyatullov et al., 2012).

Compound **15** was isolated from the culture filtrates of *Mollisa nigrescens*, an endophytic fungus from the leaves and stems of lowbush blueberry (*Vaccinium angustifolium*). Single crystal X-ray analysis and 1D and 2D NMR assignments identified the metabolite as the known tyrosine kinase inhibition 2,3-dihydro-2-hydroxy-2,4-dimethyl-5-*trans*-propenylfuran-3-one (Fig. 3, SI Fig.94-99)(Grove, 1971; Nagle et al., 2004).

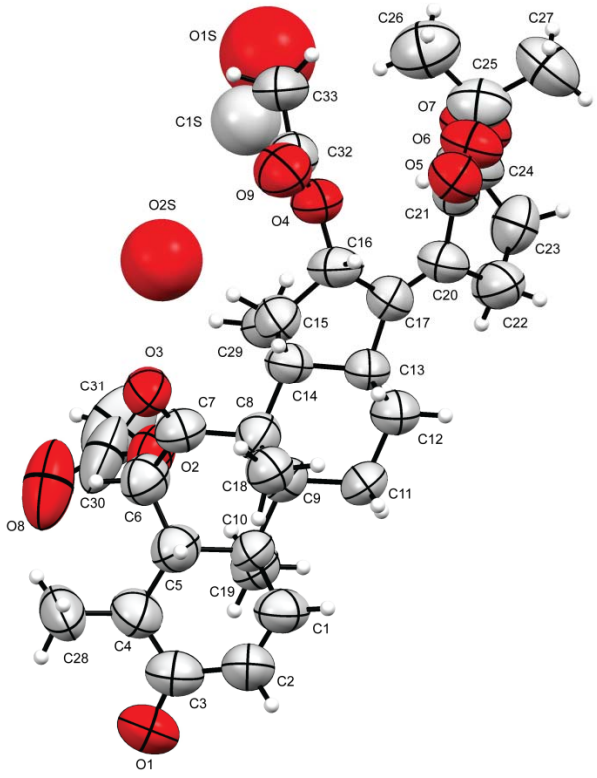


Figure 4. ORTEP single crystal X-Ray structure of **14** with partial H_2O (OS2) and methanol molecule shown (C1S,O1S) **.**

2.1 Biological activity

Compounds **1**–**2**, **4**, **8**, and **14**–**15** were evaluated for preliminary biological activity against six species of microorganisms comprised of bacteria and yeasts, in accordance with the Clinical Laboratory Standards Institute (CLSI) protocols M7-A5 and M27-A (National Committee for Clinical Laboratory Standards, 2000, 1997), (*E. coli* BW25113 Δ*bamB*Δ*tolC*, *Staphylococcus aureus* ATCC#29213, *Bacillus subtilis* 1A1, *Micrococcus luteus, Saccharomyces cerevisiae* B4741, and *Candida albicans* ATCC# 90028) (SI Fig. 101). Compounds **3**, **5**–**7** and **9**–**10** were not tested because of limited amounts of material.

Nemanifuranone A (**2**) showed modest activity against *E. coli* bacteria with a minimum inhibitory concentration (MIC) of 200 μg/mL, and showed significant inhibition (>75 % inhibition) against several gram-positive bacteria (*S. aureus*, *B. subtilis* and *M. luteus*) at a

concentration of 100-200 μg/mL (SI Fig. 101). Nemanifuranone A (**2**) also showed no activity against *C. albicans* but demonstrated some inhibition (>25% inhibition) against *S. cerevisiae* at 200 μg/mL. Compounds **1**, **4**, and **8** showed no biological activity against the tested microorganism at concentrations of 50-200 μg/mL (SI Fig. 101). The nordammarane triterpenoid (**14**) had previously been reported to have no activity against *S. aureus*, *B. subtilis* and *E. coli,* which is in accordance with our findings (Afiyatullov et al., 2012). However, (**14**) showed significant inhibition (>75% inhibition) of *M. luteus* at a concentration of 100 μg/mL, which has not previously been reported.

In a study of fungal endophytes from blueberry plants, a related known compound, 2,3-dihydro-2-hydroxy-2,4-dimethyl-5-*trans*-propenylfuran-3-one (**15**) was isolated and crystallized from *Mollisia nigrescens* (Fig. 3, SI Fig.94-99)(Grove, 1971). The single crystal X-ray structure had not been previously reported. Structures (**2**) and (**15**) differ at the C2 position, where (**2**) has the hemiacetal within the 5-alkenyl-3(2H) furanone moiety, while (**15**) possesses a methyl group, forming a hemiketal. Single crystal X-ray analysis confirmed (**15**) was a 1:1 racemic mixture at the C2 hemiketal (Fig.3). Compounds (**2**) and (**15**) were tested as racemic compounds.

Compound (**15**) is of interest as it was first isolated from the fungi *Stemphylium radicinum,* and *Ascohyta Salicorniae*, and it showed tyrosine kinase inhibition against the lymphocyte protein kinase $TKp56^{LCK}$, while also demonstrating weak antifungal activity against *Eurotium repens* and *Microbotryum violaceum* (Grove, 1971; Nagle et al., 2004; Osterhage et al., 2000). Compound (**15**), however, showed no activity against *S. aureus*, *S. cerevisiae* and *C. albicans* at a concentration of 200 μg/mL. However, some inhibition (>25% inhibition) was observed against *B. subtilis* and *E. coli* at 200 μg/mL (SI Fig. 101). This is interesting as (**2**) showed inhibition against *S. aureus* unlike (**15**). These results indicate that the

hemiacetal derivatives of the 5-alkenyl-3(2H)-furanone core may be more biologically inhibitory or active relative to similar structures with a hemiketal moiety. Similarly, nemanifuranone C (**4**) demonstrated no activity against the tested strains, indicating that the olefinic group was essential to the inhibition.

Examination of the literature revealed that few of the reported 5-alkenyl-3(2H)-furanone compounds show significant activity against Gram-positive or Gram-negative bacteria; however, several are cytotoxic. Aurafuron A and B showed moderate activity against filamentous fungi, but no Gram-positive or Gram-negative activities were noted (Kunze et al., 2005). Ecopia Biosciences' E-837 and E-975 compounds showed activity against *S. cerevisiae* (deleted of *pdr1, pdr3, erg6*) at 64 μg/mL and having electron transport inhibition, with IC_{50} values of 1-4 μg/mL against *Ascaris suum* NADH-fumarate reductase complex (Banskota et al., 2006). In the case of linfuranone A, no antimicrobial or cytotoxic activity was reported (Indananda et al., 2013). Hyafurones A_1 and A_2, are reported to have no antibacterial activity, but hyafurones A_1 was cytotoxicity against HUVEC mammalian cells and hyafurone B had an MIC of 8.3 μg/mL against *Nocardia flava* (Okanya et al., 2014). Interestingly, while not demonstrating significant antimicrobial activity, aurafuron A and B showed cytotoxicity against mouse fibroblast, while actinofuranone A and B showing activity against mouse splenocyte T-cells and macrophages.

3. Conclusion

The discovery of epoxynemanione A (**1**), nemanifuranones A-F (**2-7**) and nemanilactones A-C (**8-10**), including four known compounds (**11-14**), has demonstrated the potential for fungal endophytes from fruiting plants to be an excellent source for previously undescribed specialized metabolites, that may play a role in the chemical ecology of host plant defense mechanisms. Nemanifuranones A-F (**2-7**) are a previously undescribed series of 5-alkenyl-3(2H)-furanones possessing a rare C2 hemiacetal, with shortened alkyl chains, and furanone-lactone derivatives with few structurally similar compounds. These discoveries are of interest as nemanifuranone A (**2**) was active against both Gram-negative and Gram-positive bacteria, warranting further study into plant-microbe microbiomes and agriculturally important flowering plants.

EXPERIMENTAL SECTION

4. Experimental

4.1. General experimental procedures

NMR experiments were performed using 3 mm NMR tubes (Wilmad 335-PP-7) on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm QNP cryoprobe (700.17 MHz for ^{1}H and 176.08 MHz for ^{13}C), with chemical shifts referenced to the residual solvent signals (Gottlieb et al., 1997). NMR data processing was done using MNOVA v10.0.1 by Mestrelab Research. LC-MS analysis and HRESIMS measurements were performed using a Bruker MaXis 4G UHR-qTOF mass spectrometer coupled to a Dionex Ultimate 3000 HPLC system, operating in the positive electrospray ionization mode with calibrations done using HCO_2Na. LC-MS chromatography was carried out using a Supleco Ascentis Express C18 reversed-phase core-shell column (150 × 4.6 mm, 2.7 μm) with a mobile phase consisting of a linear gradient of mobile phase A (H_2O with 0.1% (v/v) HCO_2H) in B (CH_3CN with 0.1% (v/v) HCO_2H): 95% (v/v) A in B from 0 to 2.5 min, 95–5% A in B from 2.5 to 25 min, 5-0% A in B from 25 to 25.1 min, 100% B from 25.1 to 30 min, 0-95% A in B from 30 to 32 min, and 95% A in B from 32 to 40 min at a flow rate of 0.75 mL/min. Optical rotations were measured using an Autopol IV polarimeter (Rudolph Research Analytical).

4.2. Endophyte Sampling and Identification

The endophytic fungus *Nemania serpens* (Pers.) Grey (1821), in the Xylariaceae family, was isolated from the grapevine and leaves of a Riesling plant, maintained in a fungicide free area located at Jordon Research Farm (Agriculture and Agri-Food Canada), Jordan Station, Ontario, Canada (GPS = 43.17534° N /79.35905°W). In the case of (**15**), the endophytic fungus

Mollisia nigrescens, was isolated from the stems and leaves of a lowbush blueberry plant (*Vaccinium angustifolium*) located in a blueberry field at Mt. Thom, Nova Scotia, Canada (GPS = 45.49169° N /62.98954°W) (SI Fig. S94-99). Plant material was first surface sterilized using EtOH (90%) and bleach (65%) under standardized conditions (Carroll and Carroll, 1978), then cut and excised prior to plating on MEA (2% malt agar) plates for 28 days at 25 °C. Endophytic fungal colonies that grew on the MEA media were transferred to PDA (potato dextrose agar, Fluka) plates for an additional 15 days. Identification was done using a UltraClean Microbial DNA Kit (MoBio, Carlsbad, CA) with ITS1 and ITS4 PCR primers used for amplification as previously described (Sumarah et al., 2008).

4.3. Fermentation and Metabolite Isolation

Metabolites (**1**-**14**) were isolated from several 2 L pooled fermentation batches of *N. serpens* grown in potato dextrose media (SI Fig. 100). Each fermentation batch consisted of 10 × 1 L Roux bottles containing PDB media (200 mL, 24 g/L potato dextrose) and fermented in stationary culture for 40 days at 25 °C. In total, 12 L of fermentation culture was harvested, with some differences in metabolite production seen across each batch (SI Figure 100). Culture filtrates were vacuum filtered using Whatman #4 filters to separate the broth from the mycelium. The filtrate was extracted with equal volumes of EtOAc (1:1 ratio) and dried under reduced pressure (2.68 g/ 12 L). The mycelium was extracted in two stages, first with MeOH and secondly with acetone, each at 1 L per 2 L of fermentation. The organic mycelia fractions were then pooled and dried under reduced pressure. The crude filtrate and mycelia extracts were next re-suspended in minimal amounts of MeOH. Filtrates were then centrifuged at 13,226 g for 10 min and Acrodisc (13 mm, 0.45 mm GHP) filtered prior to HPLC purification. Semi-preparative HPLC purification was performed on an Agilent 1100 series HPLC equipped

with a G1315B diode array detector and monitored at four wavelengths of 210, 254, 275 and 350 nm. Metabolite (**15**) was isolated from a 2 L pooled fermentation batch of *Mollisia nigrescens* grown in potato dextrose media and harvested in a similar fashion as for *N. serpens.*

Metabolites **1-15** were first isolated directly from the crude filtrates or mycelia using a single semi-preparative HPLC purification process. In this first pass purification, metabolites **1-2**,**4**, **8,** and **11-14** were sufficiently resolved to obtain quality data using a Phenomenex Synergi-Max reversed-phase C-12 column (250 × 10 mm, 4 μm). However, for metabolites **3**, **5-7**, **9-10** a second semi-preparative reversed-phase HPLC fractionation process was need to further resolve co-eluting peaks.

First pass semi-preparative HPLC purifications were done using a Phenomenex Synergi-Max reversed-phase C-12 column (250 × 10 mm, 4 μm) operating at 5 mL/min and 40 ^{o}C. Mobile phase composition was a linear gradient of A (H_2O with 0.1% (v/v) HCO_2H) in B (CH_3CN with 0.1% (v/v) HCO_2H): 90% (v/v) of A in B from 0 to 4 min, 90–70% of A in B from 4 to 15 min, 70% of A in B from 15 to 20 min, 70-10% A in B from 20 to 30 min, 10-0% A in B from 30 to 31 min, 100% B from 31 to 36 min, 0-90% A in B from 36 to 38 min, and 95% A in B from 38 to 45 min with fractions collected every 20 s. Epoxynemanione A (**1**) (2.8 mg, 0.23 mg/L) eluted at 11 min, nemanifuranone A (**2**) (99 mg, 8.26 mg/L) eluted at 14.5 min, nemanifuranone C (**4**) (1.9 mg, 0.16 mg/L) eluted at 5.8 min, and nemanilactone A (**8**) (8.4 mg, 0.7 mg/L) eluted at 18.6 min. Metabolites **3**, **5-7**, and **9-10** eluted at 15.1 min, 19.8 min, 19.3 min, 21.5 min, 26.2 min and 27.2 min respectively, as semi-pure compounds. Known metabolites **11** (4 mg, 0.33 mg/L) eluted at 19.7 min, **12** (4 mg, 0.33 mg/L) eluted at 22.2 min, and compound **13** (2 mg, 0.17 mg/ L) eluted at 3.7 min. In the case of compound **15** (22 mg, 11 mg/L), it eluted at 25.4 min using the same HPLC method. Metabolite **14** was isolated from the mycelium using a modified gradient from the first-pass HPLC method. The composition was a

linear gradient of 95% (v/v) A in B from 0 to 4 min, 95–75% A in B from 4 to 17 min, 75% A in B from 17 to 23 min, and 75-15% A in B from 23 to 33 min with **14** (4 mg, 0.33 mg/L) eluting at 31.6 min.

Second pass semi-preparative HPLC purifications were done using a Columnex Chromenta KAQ C-18 column (250 × 10 mm, 5 μm) operating at 5 mL/min and 40 ^{o}C. Metabolites **3, 5-6** were isolated using a mobile phase composition of A (H_2O with 0.1% (v/v) HCO_2H) in B (CH_3CN with 0.1% (v/v) HCO_2H): 90% (v/v) A in B from 0 to 4 min, 90–80% A in B from 4 to 15 min, 80% A in B from 15 to 25 min. Nemanifuranone B (**3**) (1.32 mg, 0.11 mg/L) eluted at 16.3 min, nemanifuranone D (**5**) (2 mg, 0.16 mg/L) eluted at 22.5 min, and nemanifuranone E (**6**) (2 mg, 0.16 mg/L) eluted at 19.7 min. Nemanifuranone F (**7**) (2.6 mg, 0.22 mg/L) was purified using a modified gradient composition of 90–77% (v/v) A in B from 4 to 15 min, 77% A in B from 15 to 24 min, and eluting at 22.7 min. Nemanilactone B (**9**) (1.5 mg, 0.13 mg/L) and C (**10**) (1.5 mg, 0.13 mg/L) were isolated using a modified gradient composition of 90% (v/v) A in B from 0 to 3 min, 90–63% A in B from 3 to 13 min, 63% A in B from 13 to 22 min, with elution times at 17.4 and 20.6 min respectively.

4.3.1. Epoxynemanione A (**1**): golden-brown solid; $[\alpha]^{23}_{D}$ - 21.2 (0.03, MeOH); For ^{1}H and ^{13}C NMR (DMSO-d_6, CD_3OD) spectroscopic data see Table 1: HRESIMS *m/z* 249.0736 [M + Na]+ (calcd for $C_{11}H_{14}O_5Na$, 249.0739).

4.3.2. Nemanifuranone A (**2**): colorless crystals from CH_3CN; $[\alpha]^{23}_{D}$ 0.0 (0.92, MeOH); For ^{1}H and ^{13}C NMR (CD_3CN) spectroscopic data see Table 2: HRESIMS *m/z* 155.0705 [M + H]+ (calcd for $C_8H_{11}O_3$, 155.0703).

4.3.3. Nemanifuranone B (**3**): beige solid; $[\alpha]^{22}_{D}$ 0.0 (0.02, DMSO); For ^{1}H and ^{13}C NMR (DMSO-d_6) spectroscopic data see Table 2: HRESIMS *m/z* 157.0857 [M + H]+ (calcd for $C_8H_{13}O_3$, 157.0859).

4.3.4. Nemanifuranone C (**4**): golden oil; $[\alpha]^{25}_{D}$ 0.0 (0.03, MeOH); For ^{1}H and ^{13}C NMR (DMSO-d_6) spectroscopic data see Table 2: HRESIMS *m/z* 173.0812 [M + H]+ (calcd for $C_8H_{13}O_4$, 173.0808)

4.3.5. Nemanifuranone D (**5**): golden oil; $[\alpha]^{22}_{D}$ 0.0 (0.03, DMSO); For ^{1}H and ^{13}C NMR (C_6D_6) spectroscopic data see Table 3: HRESIMS *m/z* 191.0681 [M + Na]+ (calcd for $C_9H_{12}O_3Na$, 191.0684).

4.3.6. Nemanifuranone E (**6**): beige solid; $[\alpha]^{25}_{D}$ +8.0 (0.03, MeOH); For ^{1}H and ^{13}C NMR (CD_3OD) spectroscopic data see Table 4: HRESIMS m/z 309.1334 [M + H]+ (calcd for $C_{16}H_{21}O_6$, 309.1333)

4.3.7. Nemanifuranone F (**7**): beige solid; $[\alpha]^{25}_{D}$ -41.7 (0.01, MeOH); For ^{1}H and ^{13}C NMR (CD_3OD) spectroscopic data see Table 4: HRESIMS *m/z* 331.1157 [M +Na]+ (calcd for $C_{16}H_{21}O_6Na$, 331.1158)

4.3.8. Nemanilactone A (**8**): golden oil; $[\alpha]^{25}_{D}$ 0.0 (0.16, MeOH); For ^{1}H and ^{13}C NMR (C_6D_6) spectroscopic data see Table 3: HRESIMS *m/z* 155.0703 [M + H]+ (calcd for $C_8H_{11}O_3$, 155.0703).

4.3.9. Nemanilactone B (**9**): beige solid; $[\alpha]^{25}_{D}$ - 10.9 (0.02, MeOH); For ^{1}H and ^{13}C NMR (CD_3OD) spectroscopic data see Table 5: HRESIMS *m/z* 329.0998 [M +Na]+ (calcd for $C_{16}H_{18}O_6Na$, 329.1001).

4.3.10. Nemanilactone C (**10**): beige solid; $[\alpha]^{25}_{D}$ - 6.3 (0.04, MeOH); For ^{1}H and ^{13}C NMR (CD_3OD) spectroscopic data see Table 5: HRESIMS *m/z* 307.1180 [M + H]+ (calcd for $C_{16}H_{19}O_6$, 307.1176).

4.4. Bioassays

Minimum inhibitory concentrations (MIC) were carried out accordance with the Clinical Laboratory Standards Institute (CLSI) protocols M7-A5 and M27-A with compounds tested at a maximum concentration of 50-200 μg/mL in 96-well liquid culture format (National Committee for Clinical Laboratory Standards, 1997, 2000). Stock solutions were made to 5, 10 and 20 mg/mL in DMSO. Test strains included *E. coli* BW25113 ΔbamBΔtolC, a membrane and efflux pump compromised strain, *Staphylococcus aureus* ATCC#29213, *Bacillus subtilis* 1A1, *Micrococcus luteus, Saccharomyces cerevisiae* B4741, and *Candida albicans* ATCC# 90028. A cut off of 75% growth was used for inhibition, with the trend across dilutions also considered (SI Fig. 101).

4.5. X-ray Crystallographic Analysis

All X-ray intensity data were measured at -100^{o} C with a rotating anode diffractometer using CuKα radiation (λ = 1.54178 Å) equipped with a Bruker 6K SMART CCD detector. Supplementary crystallographic data for compounds **2**, **14** and **15** has been deposited under CCDC 1499080-1499082 at the Cambridge Crystallographic Data Center. The data can be obtained free of charge at http:// www.ccdc.cam.ac.uk/conts/retrieving.html (or from the CCDC, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033; deposit@ccdc.cam.ac.uk).

Nemanifuranone A (**2**). The integration of the data in an orthorhombic unit cell yielded a total of 1777 reflections to a maximum θ angle of 68.25° (0.83 Å resolution), of which 1777 were independent (average redundancy 1.000, completeness = 99.2%, Rint = 5.29%, Rsig = 4.32%) and 1293 (72.76%) were greater than 2σ(F2). The final cell constants of a = 7.2115(5) Å, b = 13.9477(8) Å, c = 17.6013(10) Å, volume = 1770.41(19) Å3, are based upon the refinement of the XYZ-centroids of reflections above 20 σ(I). The calculated minimum and maximum transmission coefficients (based on crystal size) are 0.6081 and 0.7531. The structure was solved and refined using the Bruker SHELXTL Software Package, in the centrosymmetric space group P b c a, with Z = 8 for the formula unit, $C_9H_{12}O_3$. The final anisotropic full-matrix least-squares refinement on F2 with 116 variables converged at R1 = 5.68%, for the observed data and wR2 = 16.52% for all data. The goodness-of-fit was 1.091. The largest peak in the final difference electron density synthesis was 0.254 e-/$Å^3$ and the largest hole was -0.201 e-/$Å^3$ with an RMS deviation of 0.057 e-/$Å^3$. On the basis of the final model, the calculated density was 1.262 g/cm^3 and F(000), 720 e-.

Compound (**14**). The integration of the data using a monoclinic unit cell yielded a total of 4567 reflections to a maximum θ angle of 44.03° (1.11 Å resolution), of which 2408 data were independent (average redundancy 1.897, completeness = 97.2%, Rint = 3.75%, Rsig = 5.48%) and 1899 (78.86%) were greater than 2σ(F2). The final cell constants of a = 31.782(9) Å, b = 8.142(2) Å, c = 14.070(4) Å, β = 112.305(6)°, volume = 3368.5(17) Å3, are based upon the refinement of the XYZ-centroids of reflections above 20 σ(I). The calculated minimum and maximum transmission coefficients (based on crystal size) are 0.4816 and 0.7486. The structure was solved and refined using the Bruker SHELXTL Software Package, in the chiral space group C2, with Z = 4 for the formula unit $C_{33}H_{44}O_9$ (with partial water and methanol solvent molecules). The final anisotropic full-matrix least-squares refinement on F2 with 404 variables

converged at R1 = 5.44%, for the observed data and wR2 = 14.91% for all data. The goodness-of-fit was 1.009. The largest peak in the final difference electron density synthesis was 0.150 e-/Å^3 and the largest hole was -0.120 e-/Å^3 with an RMS deviation of 0.036 e-/Å^3. On the basis of the final model, the calculated density was 1.175 g/cm^3 and F(000), 1278 e-.

Compound (**15**). The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using a monoclinic unit cell yielded a total of 7468 reflections to a maximum θ angle of 68.30° (0.83 Å resolution), of which 1457 were independent (average redundancy 5.126, completeness = 98.4%, Rint = 7.26%, Rsig = 6.43%) and 1183 (81.19%) were greater than 2σ(F2). The final cell constants of a = 13.8375(8) Å, b = 10.0481(6) Å, c = 12.9977(8) Å, β = 117.075(4)°, volume = 1609.16(17) Å3, are based upon the refinement of the XYZ-centroids of 3091 reflections above 20 σ(I) with 11.36° < 2θ < 136.3°. Data were corrected for absorption effects using the numerical method (SADABS). The ratio of minimum to maximum apparent transmission was 0.699. The calculated minimum and maximum transmission coefficients (based on crystal size) are 0.6986 and 1.0000. The structure was solved and refined using the Bruker SHELXTL Software Package, in the centrosymmetric space group C2/c, with Z = 8 for the formula unit, $C_8H_{10}O_3$. The crystal was twinned with BASF = 26.1(5)%. The final anisotropic full-matrix least-squares refinement on F2 with 105 variables converged at R1 = 5.26%, for the observed data and wR2 = 15.21% for all data. The goodness-of-fit was 1.031. The largest peak in the final difference electron density synthesis was 0.198 e-/Å^3 and the largest hole was -0.179 e-/Å^3 with an RMS deviation of 0.056 e-/Å^3. On the basis of the final model, the calculated density was 1.273 g/cm^3 and F(000), 656 e-.

Acknowledgments

This work is dedicated to the late Professor Brian E. McCarry (1946–2013). The authors would like to thank Professor J.D. Miller and Dr. D. McMullin (Carleton University) and Dr. K. Burgess (AAFC, London) for helpful suggestions for this manuscript; Drs. K. Seifert and Dr. J. Tanney (AAFC, Ottawa) for taxonomic expertise; T. McDowell and M. Kelman (AAFC, London) provided critical technical support; Dr. J. Britten (McMaster University) for X-ray expertise. We thank The Center for Microbial Chemical Biology (CMCB), and the Biointerfaces Institutes (BI) at McMaster University for access to state-of-the-art instrumentation. A.I was funded through an Ontario Graduate Scholarship (OGS) Doctoral Research Award. The work was funded by an AAFC A-base grant #1173 to MWS.

Appendix A. Supplementary Information

Characterization data for compounds **1–15**, including 1D and 2D NMR, HRESIMS, UV and HPLC (PDF).

Appendix B. Supplementary Information

X-ray crystallographic data for compounds **2**, **14** and **15** (ZIP).

http://www.sciencedirect.com/science/article/pii/S0031942217301589

References

1. Afiyatullov, S.S., Zhuravleva, O.I., Antonov, A.S., Kalinovsky, A.I., Pivkin, M. V, Menchinskaya, E.S., Aminin, D.L., 2012. New metabolites from the marine-derived fungus *Aspergillus fumigatus*. Nat. Prod. Commun. 7, 497–500.
2. Agricutlture and Agri-Food Canada, P.M.C., 2013. Crop Profile for Grape in Canada, 2013.
3. Anderson, J.R., Edwards, R.L., Whalley, A.J.S., 1982. Metabolites of the higher fungi. Part 19. Serpenone, 3-methoxy-4-methyl-5-prop-1-enylfuran-2(5*H*)-one, a new- γ-butyrolactone from the fungus *Hypoxylon serpens* (Barrons strain) (Persoon ex Fries)Kickx. J. Chem. Soc. Perkin Trans 1, 215–221.
4. Banskota, A.H., Mcalpine, J.B., Sørensen, D., Aouidate, M., Piraee, M., Alarco, A., Omura, S., Shiomi, K., Farnet, C.M., Zazopoulos, E., 2006. Isolation and identification of three new 5-alkenyl-3,3 (2*H*) -furanones from two streptomyces species using a genomic screening approach. J. Antibiot. 59, 168–176.
5. Bromley, C.L., Popplewell, W.L., Pinchuck, S.C., Hodgson, A.N., Davies-coleman, M.T., 2012. Polypropionates from the South African marine mollusk *Siphonaria oculus*. J. Nat. Prod. 75, 497–501.
6. Carroll, G.C., Carroll, F.E., 1978. Studies on the incidence of coniferous needle endophytes in the Pacific Northwest. Can. J. Bot. 56, 3034–3043.
7. Cho, J.Y., Kwon, H.C., Williams, P.G., Kauffman, C.A., Jensen, P.R., Fenical, W., 2006. Actinofuranones A and B , polyketides from a marine-derived bacterium related to the genus. J. Nat. Prod. 61, 425–428.
8. Gottlieb, H.E., Kotlyar, V., Nudelman, A., 1997. NMR chemical shifts of common laboratory solvents as trace impurities. J. Org. Chem. 62, 7512–7515.
9. Grove, J.F., 1971. Metabolic products of *Stemphylium radicinum*. Part IV. Minor products. J. Chem. Soc. C, 2261–2263.

10. Honmura, Y., Takekawa, H., Tanaka, K., Maeda, H., Nehira, T., Hehre, W., Hashimoto, M., 2015. Computation-Assisted structural elucidation of epoxyroussoeone and epoxyroussoedione isolated from *Roussoella japanensis* KT1651. J. Nat. Prod. 78, 1505–1510.
11. Hope-ross, P., 2006. Analysis in brief from the vine to the glass : Canada ' s grape and wine industry from the vine to the glass : Canada ' s Grape and Stat. Canada (No. 11-621-MIE2006).
12. Ibrahim, A., Sørensen, D., Jenkins, H. A., McCarry, B.E., Sumarah, M.W., 2014. New diplosporin and agistatine derivatives produced by the fungal endophyte Xylaria sp. isolated from *Vitis labrusca*. Phytochem Lett 9, 179–183.
13. Indananda, C., Igarashi, Y., Ikeda, M., Oikawa, T., Thamchaipenet, A., 2013. Linfuranone A, a new polyketide from plant-derived Microbispora sp. GMKU 363. J. Antibiot. (Tokyo). 66, 675–7.
14. Kuldau, G., Bacon, C., 2008. Clavicipitaceous endophytes: Their ability to enhance resistance of grasses to multiple stresses. Biol. Control. 46, 57–71.
15. Kunze, B., Reichenbach, H., Muller, R., Hofle, G., Müller, R., Höfle, G., 2005. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (Myxobacteria). Fermentation, isolation, physico-chemical properties, structure and biological activity. J. Antibiot. 58, 244–251.
16. Kuroda, K., Yoshida, M., Uosaki, Y., Ando, K., Kawamoto, I., Oishin, E., Matsuda, Y., 1993. AS-183, A novel inhibitor of acyl-CoA : cholesterol acyltransferase produced by Scedosporium sp. SPC-15549. J. Antibiot. (Tokyo). 46, 1196–1202.
17. Nagle, D.G., Zhou, Y.D., Mora, F.D., Mohammed, K.A., Kim, Y.P., 2004. Mechanism targeted discovery of antitumor marine natural products. Curr. Med. Chem. 11, 1725–56.
18. National Committee for Clinical Laboratory Standards, 2000. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically ; approved Standard — 5th Edition. Wayne, PA Natl. Comm Clin Lab Stand, NCCLS M7-A5.
19. National Committee for Clinical Laboratory Standards, 1997. Reference Method for Broth

Dilution Antifungal Susceptibility Testing of Yeasts ; Approved Standard. Wayne, PA Natl. Comm Clin Lab Stand, NCCLS M27–A.

20. Nicoletti, R., Trincone, A., 2016. Bioactive compounds produced by strains of Penicillium and Talaromyces of marine origin. Marine Drugs, 14, 37, 1-35.
21. Okanya, P.W., Mohr, K.I., Gerth, K., Kessler, W., Jansen, R., Stadler, M., Muller, R., 2014. Hyafurones, hyapyrrolines, and hyapyrones: Polyketides from *Hyalangium minutum*. J. Nat. Prod. 77, 1420–1429.
22. Osterhage, C., Kaminsky, R., Ko, G.M., Wright, A.D., 2000. Ascosalipyrrolidinone A , an antimicrobial alkaloid , from the obligate marine fungus *Ascochyta salicorniae*. J. Org. Chem. 65, 6412–6417.
23. Pedneault, K., Provost, C., 2016. Fungus resistant grape varieties as a suitable alternative for organic wine production: benefits, limits, and challenges. Sci. Hortic. (Amsterdam). 208, 57–77.
24. Provost, C., Pedneault, K., 2016. The organic vineyard as a balanced ecosystem: Improved organic grape management and impacts on wine quality. Sci. Hortic. (Amsterdam). 208, 43–56.
25. Richardson, S.N., Walker, A.K., Nsiama, T.K., Mcfarlane, J., Sumarah, M.W., Ibrahim, A., Miller, J.D., 2014. Griseofulvin-producing Xylaria endophytes of *Pinus strobus* and *Vaccinium angustifolium*: evidence for a conifer-understory species endophyte ecology. Fungal Ecol. 11, 107–113.
26. Sakamura, S., Ito, J., Sakai, R., 1971. Phytotoxic metabolites of Phyllosticta. Agri. Biol. Chem. 35, 105–110.
27. Sakamura, S., Ito, J., Sakai, R., 1970. Phyllostine, a new phytotoxic compound produced by Phyllosticta sp. Agric. Biol. Chem. 34, 153–155.
28. Strobel, G.A., 2003. Endophytes as sources of bioactive products. Microbes Infect. 5, 535–544.
29. Strobel, G., Daisy, B., Strobel, G., Daisy, B., 2003. Bioprospecting for microbial endophytes and their natural products bioprospecting for microbial endophytes and their natural products .

Microbiol. Mol. Biol. Rev 67, 491-502.

30. Sumarah, M.W., Kesting, J.R., Sørensen, D., Miller, J.D., 2011. Antifungal metabolites from fungal endophytes of *Pinus strobus*. Phytochemistry 72, 1833–1837.
31. Sumarah, M.W., Puniani, E., Blackwell, B.A., Miller, J.D., 2008. Characterization of polyketide metabolites from foliar endophytes of *Picea glauca*. J. Nat. Prod. 71, 1393–1398.
32. Sumarah, M.W., Puniani, E., Sørensen, D., Blackwell, B. a., Miller, J.D., 2010. Secondary metabolites from anti-insect extracts of endophytic fungi isolated from *Picea rubens.* Phytochemistry 71, 760–5.
33. Zhang, H.W., Song, Y.C., Tan, R.X., 2006. Biology and chemistry of endophytes. Nat. Prod. Rep. 23, 753–71.

CHAPTER SEVEN

CONCLUDING REMARKS

CONCLUDING REMARKS

Natural product discovery efforts have historically been hindered by the re-isolation of known bioactive agents, predominantly from prolific producers. Critical to new discovery efforts are new methods, protocols, technologies and strategies that mitigate the dereplication process and allow for a focused prioritized approach to profiling microbes from understudied niche environments in an untargeted fashion.

In Chapter 2 and 3, an informatic search algorithm for natural products discovery (iSNAP) was developed for the nontargeted dereplication of nonribosomal peptides from complex natural product extracts. This database search approach applies MS-based fragmentation rules to the SMILES code of a nonribosomal peptide, generating a series of fragment ions. These hypothetical spectral fragments are statistically evaluated against real LC-MS/MS spectra for true dereplication matches. The algorithm was tested against a series of linear, cyclic and cyclic-branching peptides by direct-infusion MS and LC-MS/MS and proved effect with a low false positive rate when extracted from 11 different media conditions. The utility of the algorithm was also demonstrated by effectively dereplicating known tyrocidines and complex halogenated kutznerides from crude extracts. This including brominated derivatives kutznerides which were created from feeding experiments. To address new discovery efforts, a new module to iSNAP was developed call iSNAP analogue. The new algorithm builds on the dereplicate capabilities by accurately identifying analogue congeners with statistical significance. Importantly, site-specific monomer localizations within a structures scaffold are determined between dereplicated known and its variant can be identified high confidence. This approach greatly simplifies and accelerates the discovery of new variants and is also applicable to screening peptidic derivatives created synthetically, through feeding

experiments, for post-translation modifications and in pharmacokinetic profiling. The iSNAP analog technology platform was applied to a series of architecturally diverse peptidic natural product producers and was effective in elaborating over 70 previously undescribed variants. The iSNAP platform also incorporates a user-upload interface allowing for customizable libraries for tailored screening. Future works may encompass new computational algorithms that expand iSNAP by integrating a dynamic *in silico* database with multiple fragmentation pathways. This can then be further expanded to include LC-HRMS/MS and MSn fragmentation and isotope patterns to give diagnostic sub-structure information on unknowns with statistical confidence. This is essential as it's not realistic to isolate and characterize every metabolite within an extract.

In Chapter 4 and 5 understudied fungal endophytes from fruiting plants were investigated for novel discovery using an LC-MS/MS metabolomic guided discovery approch in combination with the iSNAP platform. Multivariate statistical analysis such as PCA was first applied to a diverse and large collection of crude extracts and identified several distinct outlier isolates and clustered groupings. In total 17 endophytic fungal species were investigated and 12 new and 37 known specialized metabolites were discovered. Importantly, this approach highlights the need to pursue understudied niche environments and applying strategies that can rapidly profile large microbial collections and identify the distinct needles within the hay stack. Interestingly, iSNAP was unable to identify some of the outlier peptide metabolites as knowns. This discovery highlights several important findings from the study; (1) the known compound was not present within the in-house database of hypothetical structures; (2) the need for a continual advancement of dereplication technologies and for the creation of dynamic *in silico* libraries; (3) metabolomics guided the discovery of a known peptide and upon re-screening

with iSNAP, correctly dereplicated the known and identified several new putative analogues; (4) two of these putative analogues were identified and characterized as outliers in Chapter 5 ; (6) ~14% of the new compounds discovered were peptidic or peptide containing. Potato dextrose media was used for the fermentation. Fermentation screening with nitrogen enhanced media may provide greater opportunities for peptide expression. In Chapter 5 a metabolomic guided discovery approach using supervised OPLS-DA S-plot analysis was performed on the largest sub-set of a novel endophyte species, *Xylaria elissi*, from low and highbush blueberries. Metabolite isolations were guided by VIP scores from OPLS-DA S-plot analysis. Characterization of the targeted compounds identified 3 new cyclic pentapeptides, 5 additional new putative cyclic peptides, and 11 known compounds. Discovery efforts may be further enhanced by integrating microbial species identification (genomic data) and multivariate statistical analysis protocols with iSNAP technologies as a new module. This would result in the creation of a minable depository of species specific metabolites. Similarly, integrating molecular-networking features that can greatly increase the confidence of analogue identifications and visually elaborating related peptidic families by showing the interconnectivities of matched fragments. Expanding on these initiatives, would be the design of computational algorithms for specific scaffolds such as large polyketide macrolides, terpenes and glycoside derivatives.

In Chapter 6 a total of 10 new specialized metabolites are reported from an endophytic *Nemania serpens*, isolated from Riesling grapevines. The extract was earlier identified as an outlier in the comprehensive metabolomic screen however no significant metabolite features were correlated with the LC-UV-abundant peaks. Bioactivity guided fractionation led to the discovery of nemanifuranone A and isolation of the reported compounds. This discovery

highlights the potential for combined a metabolomic and bioactivity guided discovery approach for novel discovery. Future works may be fruitful by mining the endophyte collection in a supervised fashion using OPLS-DA S-plot analysis with bioactivity and species groupings

APPENDIX A

S1A. Internal transcribed spacer (ITS) DNA sequences (5'-3') of selected endophytic fungal strains. Compounds **1-51** were isolated from these endophytes.

Extract: E-006 **Species**: *Nigrospora cf. sphaerica* **Source**: Cranberry Leaf

ACCCATGTGNCTTATCTCTTTGTTGCCTCGGCGCAAGCTACCCGGGACCTCGCGCCCCGGGCGGCCCG
CCGGCGGACAAACCAAAACTCTTGTTATCTTAGTTGATTATCTGAGTGTCTTATTTAATAAGTCAAAA
CTTTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTG
AATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCATTAGTATTCTAGTGGGCA
TGCCTGTTCGAGCGTCATTTCAACCCCTAAGCACAGCTTATTGTTGGGAACCTACGGCTTCGTAGTTC
CTCAAAGACATTGGCGGAGTGGCAGTGGTCCTCTGAGCGTAGTAATCTTTTATCTCGCTTCTGTTAGG
TGCTGCCCCCCCGGCCGTAAAACCCCCAAATTTTTCTGGTTGACCTCGGATCAGGTAGGAATACCCGC
TGAACTTAAGCATATCA

Extract: E-008 **Species**: *Nemania Serpens* **Source**: Cranberry Leaf

GCGTCTCGCCCCGTAAGAACCTACCCTGTAGGACCTTACCCGGTAGACGACCCTGCCGACGGCCCCC
GAAACTCTGTTTTATAGCATTAAACTTCTGAAAATATAACTAAATAAGTTAAAACTTTCAACAACGG
ATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTC
AGTGAATCATCGAATCTTTGAACGCACATTGCGCCCACTAGTATTCTGGTGGGCATGCCTGTTCGAGC
GTCATTTCAACCCTTAAGCCCCTGTTGCTTAGCGTTAGGAGCCTACCGGAACTCTCTGGTAGCTCCCC
AAAGTCAGTGGCGGAGCCGGTTCGCACTCCAGACGTAGTAGCTTTTACACGTCGCCTGTAGCGCGGG
CCGGTCCCCTGCCGTAAAACACCCCAATTTTTATAGGTTGACCTCGGATCAGGTAGGAATACCCGCT
GAA

Extract: E-035 **Species**: *cf. Leptodontidium* **Source**: Lowbush blueberry

GAGGGATCATTACAGAGTTCATGCCCTTACGGGTAGATCTCCCACCCTGTGTTATTATAACCTTTGTT
GCTTTGGCGGGCCGCCAGGCTTCGGCCAGGCTACCGGCTCCGGCTGGTGAGCGCCCGCCAGAGGAAA
TCAAACCCTGAGTATTAGTGTCGTCTGAGTACTATATAATAGTTAAAACTTTCAACAACGGATCTCTT
GGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAAT
CATCGAATCTTTGAACGCACATTGCGCCCCTTGGTATTCCGAGGGGCATGCCTGTTCGAGCGTCATTA
CAACCCTCAAGCATTGCTTGGTCTTGGGCTGCGCTGCTAACCCAGCGGGCCTTAAAATCAGTGGCGG
TGCCGTCGGGCTCTGAGCGTAGTAATTCTTCTCGCTACAGGGACCCGGTGGTGACTGGCCAGCAACC
CCCAATTTTCTATGGTTGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAAGCATAT

Extract: E-038 **Species**: *cf. Rhizosphaera* **Source**: Cranberry Leaf

TCCGTAGGGTGACCTGCGGAAGGATCATTAAAGAGTAAGGGTCTCTGGCCCGAACCTCCAACCCTCT
GTTGTTAAAACTACCTTGTTGCTTTGGCGGGACCGCTTGGTCCTCCGAGCGCGTCGGTCTTCGGATTG
ACGAGCGCCCGCCAGAGTCCAACCAAACTCTTGTATTAAACCAGTCGTCTGAGTATAAAATTTTAAT
TTAATTAAAACTTTCAACAACGGATCTCTTGGTTCTCGCATCGATGAAGAACGCAGCGAAATGCGAT
AAGTAATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCTTGGTATT
CCGAGGGGCATGCCTGTTCGAGCGTCATTACACCACTCAAGCATTGCTTGGTATTGGGCACTCGTCCG
CCGCAAGGCGGGCGTGCCTCGAAGACCTCGGCGTGGCCTAMCCGGCTTCGGGCGTAGTAGAGTTAA
ATCGAACGTCTTATAAGTCYGGATAGGTTCCACTTTGCCGTAAAACCTTTATATTTGGGAGGTGACCT
GGGATCGGGTAGGGATACCCGCTGAACTTAAGCATATCAATAAGCGGAGGA

Extract: E-046 **Species**: *Godronia cf. cassandrae* **Source**: Highbush blueberry

CGGAAGGATCATTAAGGAGTATTTGCGGGGAATCGAAAGaAAGTACCGCTCTCCCACCCGTGACTAT
ATACTATGTTGCTTTCCGGGCTTAAACCCCCGGAGAGGACCAAACTCTTGAATTTATTACTGTCTGAG
TACTATATAATAGTTAAAACTTTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCG
AAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCC
CCTTGGTATTCCGGGGGGCATGCCTGTTCGAGCGTCATTAATACCAATCCCTTCGGGGGTCTTGGGGC
TTGGGATCTCCCAGCTCTTAAAATCAGTGGCGGTGCCTCTCGGCTCTAAGCGTAGTAATTCTTCTCGC
TATAGTCCCCGGGAGAACACTTGCCATAACCCCCACACTTTCAAGGTTGACCTCGGATCAGGTAGGG
ATACCCGCTGAACTTAAGCATATCATTTGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAAGCA
TAGACAATCGACCTTCGGTCTATTACTTCCAAGGGTTGACCTCGGATCAGGTAGGGATACCCGCTGA
ACTTAAG

Extract: E-051 **Species**: *Coniochaeta tetraspora* **Source**: Lowbush blueberry

CGTACCCGTCAGCGTTGCCTCGGCGGGCGGCCCCTCCCTGGGGCCGCTGCCTCCCTCGGGGGGTGCC
CGCCGGCGTACCAAAACTCTTTTGTATTTTAGTGGCCTCTCTGAGAAAACAAGCAAATAAGTTAAAA
CTTTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTG
AATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCGCCAGTACTCTGGCGGGC
ATGCCTGTTCGAGCGTCATTTCAACCCTCAAGCCCTGCTTGGTGTTGGGGTCCTACGGCTGCCGTAGG
CCCTGAAAGCTAGTGGCGGGCTCGCTATAACTCCGAGCGTAGTAGTAAAATATCTCGCTAGGGAGGT
GTCGCGGGTTCCGGCCGTGAAAGCCCATCTTTTACACAAGGTTGACCTCGGATCANGTAGGAATACC
CGCTGAACTTAAGCATATNTCTTCTGTCCCCTAAGCGTTGTGGAAACTATTCGCTAAAGGGTGTTCGG
GAGGCTACGCCGTAAAACAACCCCATTTCTAAGGTTGACCTCGGATCAGGTAGGGATACCCGCTGAA
CTTAAGCATA

Extract: E-101 **Species**: *Creosphaeria sassafras* **Source**: Concord grape leaf

AGGGATCATTACAGAGTTATCTAACTCCCAAACCCAATGTGAACCTACCTATGTTGCCTCGGCGGGA
GAAGCCTACCCGGTACCTACCCTGTAGCTACCCGGGAACTACCCTGTAGTTGATGGCGGACCCGCCG
GTGGACATTTTAAAAACTCTTGTTTTATTGTGACACTCTGAGTAACAAACTAAATAAGTTAAAACTTT
CAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATT
GCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCATTAGTATTCTAGTGGGCATGCC
TGTTCGAGCGTCATTTCAACCCTTAAGCCTTAGTTGCTTAGCGTTGGGAGGTTGCCTCTCGCAACTCC
CTAAAATCAGTGGCAAGGACCACGACACCTTGAAGTGTAGTAGTTTAATTCTCACTTTGGAGGATCG
TAGCCCAGCCATAAACCAACTATTTTTAATTATTGGTTGACCTCGGATCAGGTAGGAATACCCGCTGA
ACTTAA

Extract: E-132 **Species**: *Mollisia nigrescens* **Source**: Lowbush blueberry

AAGGATCATTAATAAGGATACCGGGAAACCGGTAGACCCACCCGTGTCTATCTACTCTTGTTGCTTTG
GCAGGCCGTGGCCTCCACCGTGGGCTCTGCTTGCGTGTGCCTGCCAGAGGACCAAACTCTGAATTTT
AGTGATGTCTGAGTACTATATAATAGTTAAAACTTTCAACAACGGATCTCTTGGTTCTGGCATCGATG
AAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAAC
GCACATTGCGCCCGGTGGTATTCCGCCGGGCATGCCTGTTCGAGCGTCATTATAACCACTCAAGCCTA
GCTTGGTATTGGAGTTCGCGGTCCCGCGGCTCCTAAAATCAGTGGCGGTGCCGGTGGGCTCTAAGCG
TAGTAA

Extract: E-138 **Species**: *Xylaria ellisii*. **Source**: Highbush blueberry

AGGGATCATTAAAGAGTTCTATAACTCCCAAACCCATGTGAACATACCTTACGTTGCCTCGGCAGGT
CGCGCCTACCCCGTAACGTCCTACCCTGTAGGACCTACCCGGTAGACGCGGGTAAGCCTGCCGGCGG
CCCACGAAACTCTGTTTAATATTGAATTCTGAACCTAAAACTAAATAAGTTAAAACTTTCAACAACG
GATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATT
CAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCATTAGTATTCTAGTGGGCATGCCTGTTCGAG
CGTCATTTCAACCCTTAAGCCTCTGTTGCTTAGTGTTGGGAGCCTACGGCACCCGTAGCTCCTCAAAG
TTAGTGGCGGAGTCGGTTCACACTCTAGACGTAGTAATTCTTATCTCGCCTATCAGTTGGACCGGTCC
CCTGCCGTAAAACCCCCCATTTTTAAAGGTTGACCTCGGATCAGGTAGGAATACCCGCTGAACTTAA

Extract: E-140 **Species**: *Nemania serpens* **Source**: Concord grape leaf

AGGGATCATTACAGAGTTATCAAAACTCCCAAACCCATGTGAACATACCTCGCGTTGCCTCGGCAGG
TGGCGCCCCGCCCTGTGAGGGCCTACCCTGTAGGAGCTCACCCGGTGGTCGCCCCTGCCGACGGCCC
ACGAAACTCTGTTTTTTATAGCATTAGACCTCTGAAAAGATAACTAAATAAGTTAAAACTTTCAACA
ACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGA
ATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCACCAGTATTCTGGTGGGCATGCCTGTTC
GAGCGTCATTTCAACCCTTAAGCCCCTGTTGCTTAGCGTTGGGAGCCTACCGGAAACCGTCCGGTAG
CTCCCCAAAGTCAGTGGCGGAGCCGGTTCGCACTCCAGACGTAGTAGCTTTTACACGTCGCCTGTAG
CGCGGGCCGGTCCCCTGCCGTAAAACCCCCAAATTCTTATAGGTTGACCTCGGATCAGGTAGGAATA
CCCGCTGAACTTAA

Extract: E-169 **Species**: *Alternaria sp.* **Source**: Raspberry leaf

AGGGATCATTACACAAATATGAAGGCGGGCTGGAACCTCTCGGGGGTTACAGCCTTGCTGAATTATT
CACCCTTGTCTTTTGCGTACTTCTTGTTTCCTTGGTGGGTTCGCCCACCACTAGGACAAACATAAACCT
TTTGTAATTGCAATCAGCGTCAGTAACAAATTAATAATTACAACTTTCAACAACGGATCTCTTGGTTC
TGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAGTGTGAATTGCAGAATTCAGTGAATCATC
GAATCTTTGAACGCACATTGCGCCCTTTGGTATTCCAAAGGGCATGCCTGTTCGAGCGTCATTTGTAC
CCTCAAGCTTTGCTTGGTGTTGGGCGTCTTGTCTCTAGCTTTGCTGGAGACTCGCCTTAAAGTAATTG
GCAGCCGGCCTACTGGTTTCGGAGCGCAGCACAAGTCGCACTCTCTATCAGCAAAGGTCTAGCATCC
ATTAAGCCTTTTTTCAACTTTTGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAA

Extract: E-182 **Species**: *cf. Paraphaeosphaeria neglecta* **Source:** Lowbush blueberry Leaf

NNNNNNNNNNNNNNCCNGNNNNNAGGTCAAGANGGTAATGTTGCTTCGTGGACGCGGGCCACGCC
CCCCCgcagacgcAATTGtGCTGCGCGAGAGGAGGCAAAGACCGCTGCCAATGAATTTGGGGCGAGTCCG
CGCGCAGAGGCGGGACAGACGCCCAACACCAAGCAGAGCTTGAGGGTGTAGATGACGCTCGAACAG
GCATGCCCCATGGAATACCAAGGGGCGCAATGTGCGTTCAAAGATTCGATGATTCACTGAATTCTGC
AATTCACACTACTTATCGCATTTCGCTGCGTTCTTCATCGATGCCAGAGCCAAGAGATCCATTGTTGA
AAGTTGTAACGATTATTTGTATCAGAACAGGTAATGCTAGATGCAAAAAAGGTTTTGTTTTGGTTCCA
GCGGCAGGTTGCCCCGCCGAAGGAGAACGAAAGGTGCTCGTAAAAAGAGGGTACAGGCAGGCGGC
GCGTGAGGGTGTTACCCCTACCACCCGGGAGCAAGCCCCCGGGGGCCGCGACCGCACCTGGTTTGAG
ATGGATAATGATCCTTCCGCAGGTTCACCTACGGAAGGATCATTATCCATCTCAAACNNGGTGNGGT
CGCGGCNNNNGGGGGCTTGCTCNNGGGTGGNNGGGGTAACACCNTCACGCGCNGCCTGNNTGTACN
NTCTTTTTACGAGCACCTTTNGTTCNCNTTNGGNGGGGNNACNTGNNGNTGGAACCAAAACAAAANN
TTTTNNGNATNNAGCATTACNNGNNNNNGATACAAATAATNGNTACANCTTTNANNANNGGATNNC
ATGGNTCNNNNNTNNATGANGNANGNNGNGNANTGNNGNTAAGNNNGNGTGAATTGCNGAATTNN
GNGNNTCNATNGNNTCCNNTGNANNGNNNNNNGGGNNNNNNNGNNNTTNCNNGGGGGCATGNNNN
GCTNGNGNGNNNNTNNNNTNNNGNGGNNNGNNCCNNCNNNNNNNNNGNNNGNNNNNNNNNNNNT
GNGGNNCNNNTNNNNNNNNNNC

Extract: E-184 **Species**: *Xylaria cubensis* **Source**: Riesling grape leaf

CATTAAAGAGTTCTATAACTCCCAAACCCATGTGAACATACTTAACGTTGCCTCGGCAGGTCGCGCCT
ACCTCGTAGCACCCTACCCTGTAGGGCCTACCCGGGAGACGCGGGTCAGCCTGCCGGCGGCCTACCA
AACTCTGTTTGATATTGAATTCTGAACCTATAACTAAATAAGTTAAAACTTTCAACAACGGATCTCTT
GGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAAT
CATCGAATCTTTGAACGCATATTGCGCCCATTAGTATTCTAGTGGGCAGGACCGTTCGAGCGTCATTT
CGACCCTTAAGCCTTGGTTGCTTAGTGTTGGGAGCCTACGGCAACGTAGCTCCTCAAAGTTAGTGGC
GGAGTTGGTTCACACTCTAGACGTAGTAGATTTTTATCTCGCCTATCAGTTGGACCGGTCCCCTGCCG
TAAAACCCCCCAACTTCTTAAAGTTGACCTCGAATCGGTTCAGACAAACTCGCTAAATTGAAGCATC
TTAGTCAGCGAAGGAAAAGAAACCAACAGGGATTGCCCTAGTAACGGCGAGTGAAGCGGCAACAGC
TC

Extract: E-195 **Species**: *Nemania Serpens* **Source**: Potato Leaf

CAACCAGGTATGTTCACATGGGTTNGGAGTTTTGATAACTCNGTAATGATNNTTCCGTAGGGNNNNC
CTGCGGAGGGATCATTACAGAGTTATCAAAACTCCCaAACCCATGTGAACATACCTCGCGTTGCCTCG
GCAGGTGGCGCCCCGCCCTGTGAGGGCCTACCCTGTAGGAGCTCACCCGGTGGTCGCCCCTGCCGAC
GGCCCACGAAACTCTGTTTTTTATAGCATTAGACCTCTGAAAAGATAACTAAATAAGTTAAAACTTTC
AACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTG
CAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCACCAGTATTCTGGTGGGCATGCCT
GTTCGAGCGTCATTTCAACCCTTAAGCCCCTGTTGCTTAGCGTTGGGAGCCTACCGGAAACCGTCCGG
TAGCTCCCCAAAGTCAGTGGCGGAGCCGGTTCGCACTCCAGACGTAGTAGCTTTTACACGTCGCCTG
TAGCGCGGGCCGGTCCCCTGCCGTAAAACCCCCAAATTCTTATAGGTtGACCTCGGATCAGGTAGGAA
TACCCGCTGAACTTAAGCATATCAATAANNNGGAGGAAAGGATCATTACAGAGTTNTCAAAACTCCC
AAACCNTGTGAACATACCT

Extract: E-223 **Species**: *Nemania Serpens* **Source**: Lowbush blueberry leaf

NNACCTGCGGAGGGATCATTACAGAGTTACCAAAACTCCCAAACCCATGTGAACATACCTCGCGTTG
CCTCGGCAGGTGGCGTCCTACCCCGTGAGACCTACCCTGTAGGACCTACCCGGTAGGCGACCCTGCC
GACGGCCCCCGAAACTCTGTTTTTATAGCATTGGACTTCTGAAAAGATAACTAAATAAGTTAAAACT
TTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAA
TTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCACTAGTATTCTGGTGGGCATG
CCTGTTCGAGCGTCATTTCAACCCTTAAGCCCCTGTTGCTTAGCGTTGGGAGCCTACTGGAACTCTCC
TGTAGCTCTCCAAAGTCAGTGGCGGAGCCGGTTCGCACTCCAGACGTAGTAGCTTTTACATATCGCCT
GTAGCTTGGACCCGGTCCCCTGCCGTAAAACACCCCAAATCTTCTAGGTtGACCTCGGATCAGGTAGG
AATACCCGCTGAACTTAAGCATATCAATAAGNCGGAGGAAN

Extract: E-252 **Species**: *Ramularia cf. bellunensis* **Source**: Lowbush blueberry leaf

NTGCGGAGGGATCATTACTGAGTTTAGGTGGAATCCACCCAACTCCAACCCTTTGTGAACACATCTTGTTGCTTCGGGGGCG
ACCCTGCCATTCGTGGCATTCCCCCCGGAGGTCATCAAAACACTGCATTCTTACGTCGGAGTAAAAAGTTAATTTAATAAAAC
TTTCAACAACGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGT
GAATCATCGAATCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCATGCCTGTTCGAGCGTCATTACACCACTCAAG
CCTCGCTTGGTATTGGGCGTCGCGAGTCTCTCGCGCGCCTCAAAGTCTCCGGCTAGGCAGTTCGTCTCCCAGCGTTGTGGCA
ACTATTTCGCAGTGGAGTTCGAGTCGTCGCGGCCGTTAAATCTTtCAAAGGtTGACCTCGGATCAGGTAGRGATACCCGCTGA
ACTTAAGCATATCAATAAGNCGGAGGAAANGAT

Extract: E-260 **Species**: *Sphaerulina cf. rhabdoclinis* **Source**: Lowbush blueberry leaf

CTCGGTAAANCTTCCGTAGGNGAACCTGCGGAGGGATCATTACCGAGCGAGGGCCTTCGGGCTCGAC
CTCCAACCCTTTGTGAACACAACTTGTTGCTCCGGGGGCGACCCTGCCGTTCCGACGGCGAGCGCCC
CCGGAGGCCTTCCAACACTGCATCTTTGCGTCGGAGTTTAAGTAAATTTAAACAAAACTTTCAACAA
CGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAA
TTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCTCTGGTATTCCGGAGGGCATGCCTGTTCG
AGCGTCATTTCACCACTCAAGCCTGGCTTGGTATTGGGCGCCGCGGTGTTCCGCGCGCCTCAAAGTCT
CCGGCTGAGCTGTCCGTCTCCAAGCGTTGTGATTTCATTAATCGCTTCGGGGTGCGGGCGGCCGCGGC
CGTTAAATcCTTTCACAAGGTtGACCTCGGATCAGGTAGGGATACCCGCTGAACTTAAGCATATCAAT
AAGNCGGAGGAAN

Extract: E-261 **Species**: *Nigrospora sphaerica* **Source**: Lowbush blueberry leaf

NANTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCaACCAGA
AAAAATTGGGGGTTTTACGGCCGGGGGGGCAGCACCTAACAGAAGCGAGATAAAAGATTACTACGC
TCAGAGGACCACTGCCACTCCGCCAATGTCTTTGAGGAACTACGAAGCCGTAGGTTCCCAACAATAA
GCTGTGCTTAGGGGTTGAAATGACGCTCGAACAGGCATGCCCACTAGAATACTAATGGGCGCAATGT
GCGTTCAAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
CATCGATGCCAGAACCAAGAGATCCGTTGTTGAAAGTTTTGACTTATTAAATAAGACACTCAGATAA
TCAACGAAGATAACAAGAGTTTTGGTTTGTCCGCCGGCGGGCCGCCCGGGGCGCGAGGTCCCGGGTA
GCTTGCGCCGAGGCAACAAAGAGATAAGTTCACATGGGTTTGGGAGTTGGATAACTCTGTAATGATC
CCTCCGCAGGTNN

S2A. Observation coordinates from the PCA (ESI+) Scores plot.

Obs ID (Primary)	M2.t[1]	M2.t[2]	Obs ID (Primary)	M2.t[1]	M2.t[2]	Obs ID (Primary)	M2.t[1]	M2.t[2]
E001	1334.41	-939.088	E124	4937.4	-2350.73	E204	5665.57	-2731.37
E006	2122.29	6053.36	E125	4978.25	-1436.64	E206	-22148.6	-4114.96
E008	5497.49	-1939.44	E126	5072.96	-4107.17	E207	3012.91	-397.913
E010	5218.89	-3417.53	E128	5437.3	-2963.84	E208	-5086.22	-2456.59
E017	5281.74	-2610.37	E129	-25584.5	-720.523	E209	5947.97	-3999.85
E019	5498.63	-3364.21	E130	2307.78	3262.25	E210	6163.76	-3894.71
E021	2805.6	-796.838	E131	4815.14	-3109.91	E211	5876.1	-3721.14
E022	5726.57	-3662.62	E132	2539.13	-372.767	E212	5189.52	-2197.79
E023	5987.24	-4445.42	E133	4566.31	-3068.94	E213	5754.57	-3738.03
E024	6153.06	-2946.73	E134	4660.13	-2361.89	E214	5131.08	-1015.47
E025	-836.987	1019.84	E136	4798.83	-2070.73	E215	4470.12	-554.701
E026	3096.55	-1134.08	E138	-30506.2	-3031.16	E216	-12117.3	-2273.63
E031	3385.13	-452.95	E139	2279.73	5505.17	E217	4460.2	-2731.74
E032	5573.39	-3133.19	E140	-10582.1	17284.4	E218	4938.95	-2259.1
E033	5507.59	-3273.11	E142	-17328.9	-2536.19	E219	2221.97	3949.64
E034	4948.1	-2933.44	E143	-34221.2	-2302.47	E220	4670.92	-1537.62
E035	5226.23	-2534.42	E144	-394.907	4644.43	E221	5923.49	-2246.71
E037	5126.45	-3958.88	E145	3274.36	963.402	E222	5376.91	-2767.8
E038	4762.37	-652.503	E148	-6750.8	12520.1	E223	-6313.43	12550.1
E042	-904.735	5268.7	E149	2721.16	-2478.63	E224	6398.21	5868.79
E043	4888.63	-2794.4	E150	-20108.3	-2443.42	E225	6086.77	-2441.56
E044	-13879.1	-2610.86	E151	-1288.29	2159.4	E226	-15057	110.85
E046	5870.45	-3829.3	E152	3895.48	-1954.32	E227	4595.36	380.903
E050	1091.38	431.893	E153	-21407.6	-1317.5	E228	6729.2	5468.56
E051	4971.74	-956.251	E154	4272.59	-2020.98	E229	4877.12	-1331.41
E052	5261.11	-2801.89	E155	4899.7	-2460.37	E230	5780.41	-3447.44
E053	1413.72	8435.53	E156	1409.26	1509.78	E232	3665.8	865.205
E062	4212.21	-1866.9	E157	-28512.2	-2391.17	E233	2263.58	-724.025
E065	-191.055	1841.04	E158	4611.82	-2486.17	E234	905.881	174.311
E071	5951.21	-4151.81	E159	5451.15	-4544.75	E235	1333.18	-606.029
E073	5441.93	-1023.47	E160	2261.63	-564.602	E237	5092.92	-2829.94
E074	-1134.53	6630.72	E162	-26439.1	-3437.32	E239	4957.6	-785.317
E075	-5281.07	32489.4	E164	-26653.1	-3728	E240	5711.18	-3631.31
E076	3418.96	2140.54	E166	5737.72	-2685.81	E241	-12370.7	-2529.18
E079	-1868.3	4831.16	E169	6070.56	-2211.47	E242	-1205.56	-889.094
E080	-2036.28	6722.35	E170	-22604.1	-988.026	E243	-28756.5	-4581.97
E082	1462.09	3099.84	E171	960.975	3487.25	E244	-11166	-2956.95
E083	5122.71	-2733.53	E172	2702.89	1861.12	E245	4933.76	-3612.28
E084	27.9849	745.727	E173	-841.39	6223.88	E246	5361.59	-3848.45
E087	4746.12	-2518.31	E174	4450.22	-2507.69	E247	4963.63	-2349.58
E089	-16950.7	-3236.63	E175	5731.6	-2578.25	E248	3582.68	-2326.54
E091	2991.87	3440.4	E176	5057.56	-2589.29	E249	767.091	6173.72
E097	4166.93	-1726.55	E177	6559.77	6926.89	E250	2938.59	1160.61
E100	4085.76	-2217.61	E179	5636.93	-3575.74	E251	2249.57	1172.12
E101	2097.47	-2891.49	E180	3427.42	1536.33	E252	5357.44	-2995.4
E102	5117.96	-2683.9	E181	5956.4	233.063	E253	3747.67	976.262
E103	2697.41	3168.01	E186	3217.19	-1816.27	E254	2740.87	-4.19276
E104	-33023.5	1671.76	E187	5171.67	-3286.94	E255	4596.29	-2890.53
E105	5144.64	-3278.2	E188	5290.56	-1927.1	E256	6038.93	-3068.03
E106	4683.86	-2465.12	E189	6340.21	11069.3	E257	3350.61	-1599.39
E107	-29674	-1924.56	E190	5544.63	-3409.15	E258	6217.9	-4107.29
E111	261.026	16506.9	E191	5543.41	-3230.64	E259	4817.52	-1848.04
E112	-41248.9	-4066.48	E192	5630.89	-3469.1	E260	6804.36	-2837.27
E182	4363.61	8882.51	E193	5365.06	-3268.66	E261	4477.05	2044.45
E183	5300.23	-3374.96	E194	617.678	5430.72	E262	6773.89	4626.55
E184	1194.05	71875.1	E195	-5457.5	9446.57	E263	5655.65	-3263.75
E185	4276.69	1348.81	E196	4593.93	-3101.01			
E113	3981.28	-2138.78	E197	4612.72	547.481			
E114	-26073.3	-6527.01	E198	4785.54	-2390.67			
E115	-29853	-1449.81	E199	-2120.72	-2133.51			
E116	-26158.9	-2038.81	E200	4964.37	-1976.26			
E119	3962.78	-3019.36	E201	4419.31	-2309.72			
E121	3991.18	-1858.07	E202	3213.08	1034.59			
E122	3948.3	1127.89	E203	5370.64	-2661.22			

S3A. Observation coordinates from the PCA (ESI-) Scores plot.

Obs ID (Primary)	M2.t[1]	M2.t[2]	Obs ID (Primary)	M2.t[1]	M2.t[2]	Obs ID (Primary)	M2.t[1]	M2.t[2]
E006	-811.974	-757.678	E128	-3080.54	-139.154	E203	-2831.3	-501.478
E008	-10995.3	1025.01	E129	24375.5	-7855.52	E204	-5954.72	-373.392
E010	-2839.68	-964.622	E130	-1777.7	-1247.18	E206	20073.6	-6337.8
E011	-3026.73	-785.214	E131	-2736.02	-378.71	E207	-2605.44	4471.92
E017	-2802.68	1128.29	E132	-2363.81	432.809	E208	7170.06	-5000.02
E019	-2724.55	-81.3474	E133	-2874.21	-863.647	E209	-2770.26	-577.616
E021	-3987.07	696.591	E134	-4190.06	897.643	E210	-2889.61	-561.115
E022	-2777.92	-663.344	E136	-2950.1	-859.638	E211	-2877.52	-669.617
E023	-2117.16	407.466	E138	18468.8	-8527.51	E212	-3266.03	-632.346
E024	-3137.23	669.866	E139	-4286.98	5386.19	E213	-2976.2	-900.87
E025	-2552.22	-765.895	E140	583.499	-148.501	E214	-2599.15	-431.119
E026	-2617.25	-225.119	E142	11345	-7692.69	E215	-2575.63	-609.433
E031	-2074.32	2506.04	E143	16337.9	-8552.88	E216	10753.7	-4509.81
E032	-3301.62	286.256	E144	-3005.62	1953.73	E217	-3602.73	-367.963
E033	-3046.57	-816.847	E148	2525.32	-200.418	E218	-3044.2	-167.038
E034	-2613.79	-822.311	E149	-3261.15	-803.717	E219	-2143.35	3963.68
E035	-2795	1024.72	E150	10598.7	-6730.88	E220	-4267.62	844.011
E037	-2915.3	-660.999	E151	942.346	-1066.08	E221	-3523.13	2965.35
E038	-10453.8	401.465	E152	-3708.76	-195.873	E222	-3558.71	-360.187
E040	-2887.3	-300.494	E153	10355.8	-5617.57	E223	831.957	-1221.2
E042	-2381.32	1255.61	E154	-2769.07	-775.228	E224	-4996.99	3263.74
E043	-2729.75	-722.338	E155	-3077.68	-88.3082	E225	-14153.1	293.74
E044	11651.4	-7362.2	E156	1279.21	342.695	E226	17176.2	-4501.28
E046	-2968.35	-921.266	E157	10966.6	-7209.18	E227	-2575.12	316.67
E050	-42.5249	1411.39	E158	-3414.9	65.025	E228	-4405.66	2306.58
E051	-2663.31	-921.207	E159	-2942.71	-778.965	E229	-2812.05	-894.359
E052	-2945.48	1162.04	E160	-2053.42	-274.404	E230	-3066.72	-832.98
E053	281.642	-1693.18	E162	12567.6	-7172.23	E232	-2392.51	182.193
E062	-2719.41	-273.392	E164	9469.12	-7117.64	E233	-2584.85	340.744
E065	-1966.21	363.124	E166	-3824.4	494.859	E234	-1702.95	600.914
E071	-3038.53	-179.682	E169	-2564.41	2694.02	E235	-2615.35	-365.453
E073	-3359.07	-566.027	E170	15594.1	-8626.57	E237	-2776.23	-971.263
E074	1492.32	-1663.35	E171	-2878.23	-535.026	E239	-2950.93	2.54455
E075	6521.62	-4752.82	E172	-1864.25	-1186.81	E240	-2713.61	141.56
E076	-2371.37	-466.415	E173	3047.4	8466.02	E241	12723.8	-2978.67
E079	4141.79	-1978.08	E174	-2809.39	-849.065	E242	1474.41	-1591.37
E080	1668.85	-989.044	E175	25746.3	46247.3	E243	15742.2	-4033.66
E082	-4329.52	10798.6	E176	-8366.54	55.7683	E245	-2729.81	-566.125
E083	-3074.88	127.795	E177	-5112.26	4970.47	E246	-2866.76	-698.797
E084	-755.882	-949.6	E179	-3154.18	-782.439	E247	-3726.14	-223.619
E087	-2756.33	-725.132	E180	-2299.52	437.343	E248	-2671.18	-1235.47
E089	15699.3	-9092.14	E181	-3771.07	3071.71	E249	1370.58	1919.01
E091	-963.323	4629.66	E182	-2237.43	1198.42	E250	-3010.4	7357.64
E093	-11148.2	-179.739	E183	-3208.47	-544.383	E251	-3179.33	4728.02
E097	-2699.74	-617.809	E184	28874.1	52843	E252	-3065.66	-802.128
E100	-3572.85	-266.486	E185	-2713.74	-485.602	E253	-2670.13	1471.47
E101	-3110.26	-668.669	E186	-2020.38	-104.004	E254	-2262.18	-293.463
E102	-3184.47	68.4905	E187	-2836.88	-631.337	E255	-3155.19	-734.848
E103	2222.34	-629.898	E188	-3931.42	281.688	E256	-12990.8	-139.831
E104	23586.2	-9170.52	E189	-6070.4	8267.65	E257	-2652.6	-597.226
E107	18052.5	-7164.68	E190	-2969.84	-526.392	E258	-2860.53	-492.513
E111	-3191.09	-671.25	E191	-2683.42	-631.972	E259	-2912.87	-731.752
E112	12427.2	-9442.17	E192	-2839.77	-531.15	E260	-4004.46	-335.724
E113	-2761.42	462.259	E193	-2785.06	-797.499	E261	-2028.11	215.098
E114	4493.12	-4462.52	E194	9194.88	26294.4	E262	-5166.33	4417.23
E115	20019.7	-10681.3	E195	2618.27	-93.4451	E263	-2883.03	-774.31
E116	27815.9	-11393.9	E196	-3511.76	-477.87			
E119	-3021.26	-836.946	E197	-3032.8	1858.61			
E121	-4456.76	896.717	E198	-2848.82	-895.819			
E122	341.749	-15.9367	E199	3298.43	-1461.41			
E124	-2815.88	-1012.08	E200	-2829.21	-956.08			
E125	-2587.23	-410.934	E201	-3988.22	216.449			
E126	-2947.42	-713.67	E202	830.208	4267.5			

S4A. Top 120 variable contributions from the PCA (ESI+) Scores Contribution plot of E-184.

#	Var ID (Primary)	Var ID (RT(min))	M6.Score Contrib(Obs E184 - Average), Weight=p1p2	#	Var ID (Primary)	Var ID (RT(min))	M6.Score Contrib(Obs E184 - Average), Weight=p1p2
1	**451.175**	17.2388	283.7	61	**651.286**	7.57154	61.4697
2	**165.055**	12.4385	252.265	62	**359.139**	16.7017	61.0076
3	**529.313**	13.1604	219.69	63	**497.309**	12.7912	60.9119
4	**433.24**	18.679	217.983	64	**166.094**	11.9687	60.7514
5	**466.28**	19.0682	208.121	65	**674.251**	14.9223	59.7294
6	**520.298**	12.5389	195.693	66	**701.391**	9.90344	58.5174
7	**918.368**	17.255	184.717	67	**937.358**	16.114	58.0247
8	**435.144**	17.2381	175.432	68	**654.238**	16.3476	57.7684
9	**496.302**	13.3449	175.257	69	**361.155**	11.1461	54.4777
10	**759.411**	16.9444	168.407	70	**925.332**	17.0959	53.0546
11	**775.432**	15.1157	158.319	71	**593.314**	12.522	52.7624
12	**954.382**	16.315	154.613	72	**348.68**	12.8748	52.7046
13	**466.229**	14.654	151.799	73	**764.325**	9.50024	52.4436
14	**954.358**	14.2592	145.53	74	**741.291**	17.7419	52.1319
15	**494.272**	15.4596	143.991	75	**961.352**	12.1367	52.0144
16	**619.332**	15.6187	142.877	76	**565.289**	12.1868	51.7707
17	**454.28**	17.6602	139.183	77	**530.33**	13.7305	51.3587
18	**305.125**	10.6092	133.104	78	**287.045**	14.1338	50.9693
19	**758.433**	18.9503	130.878	79	**433.222**	14.285	50.9525
20	**449.133**	14.7211	126.521	80	**407.111**	17.3557	50.7157
21	**959.346**	12.296	120.556	81	**742.429**	14.7449	50.3852
22	**742.42**	19.5714	115.617	82	**792.44**	13.7808	49.4716
23	**755.331**	17.1374	115.157	83	**335.137**	8.88777	48.4538
24	**901.35**	19.5525	114.828	84	**671.35**	10.6427	46.9567
25	**241.054**	8.26776	113.658	85	**512.235**	10.2816	45.4058
26	**163.039**	16.6681	112.628	86	**952.337**	13.4371	44.8248
27	**434.232**	17.4902	107.34	87	**986.317**	19.142	44.0894
28	**287.131**	12.6874	105.787	88	**556.295**	14.1339	44.0132
29	**775.515**	12.8747	103.177	89	**486.287**	14.3348	43.8204
30	**205.143**	15.208	97.8994	90	**554.271**	12.0441	43.3602
31	**436.21**	17.2387	94.5629	91	**386.207**	14.1166	40.943
32	**291.121**	12.9501	92.1893	92	**742.282**	17.9516	40.8711
33	**594.336**	14.9221	90.6885	93	**956.841**	12.1366	40.1371
34	**935.347**	12.3545	90.3387	94	**956.337**	15.7529	38.4189
35	**901.387**	17.3306	90.052	95	**179.034**	10.2987	38.263
36	**552.342**	17.6149	90.0161	96	**745.243**	19.7054	37.5997
37	**383.176**	16.3657	89.8809	97	**384.152**	17.188	37.2763
38	**960.35**	12.12	89.3579	98	**953.473**	14.3683	37.0399
39	**776.463**	12.8581	88.9135	99	**519.236**	17.6574	34.9853
40	**612.313**	13.7314	88.732	100	**722.25**	12.1505	34.8724
41	**527.244**	15.7952	87.112	101	**636.318**	10.1731	34.1608
42	**763.322**	8.99773	86.3891	102	**929.343**	17.9266	34.0703
43	**905.396**	14.2676	84.5701	103	**550.31**	17.2047	33.9862
44	**615.314**	9.48513	84.2318	104	**621.311**	14.8551	32.1801
45	**756.255**	16.9361	81.2007	105	**484.254**	12.3299	32.0538
46	**337.125**	8.98144	78.7336	106	**601.349**	19.3358	31.9211
47	**791.435**	15.1239	78.5425	107	**822.347**	15.1573	30.436
48	**570.306**	12.1367	76.9635	108	**704.221**	12.7572	30.3364
49	**344.192**	9.53395	75.8326	109	**399.191**	17.9287	30.1046
50	**191.085**	8.41032	75.5547	110	**475.254**	18.0189	29.2395
51	**458.095**	16.2652	75.1881	111	**940.477**	17.7252	28.4222
52	**776.441**	15.5518	73.3084	112	**984.327**	14.7209	27.7948
53	**620.349**	15.1576	71.4233	113	**750.457**	19.823	27.7453
54	**653.356**	15.6525	67.0244	114	**485.225**	12.472	27.0571
55	**906.377**	14.6623	66.037	115	**957.344**	16.7347	26.9285
56	**936.328**	12.6569	64.94	116	**949.336**	15.4756	26.6074
57	**241.106**	8.53503	63.9554	117	**980.579**	18.1018	26.0356
58	**952.437**	15.0985	63.8956	118	**950.597**	17.9771	25.9469
59	**578.308**	15.8945	63.8372	119	**357.118**	8.53668	25.9148
60	**672.235**	10.5925	62.8639	120	**307.63**	10.0382	25.709

S5A. Top 120 variable contributions from the PCA (ESI-) Scores Contribution plot of E-184.

#	Var ID (Primary)	Var ID (RT)	M2.Score Contrib(Obs E184 - Average), Weight=p1p2
1	**398.146**	14.3431	198.805
2	**372.175**	16.2229	109.947
3	**452.244**	**18.3295**	**109.157**
4	**287.082**	10.245	**99.795**
5	**288.096**	13.9896	97.9105
6	**305.148**	10.1625	**95.0939**
7	**494.252**	13.4942	85.6929
8	**399.647**	16.8779	84.2097
9	**287.056**	13.8054	80.5058
10	**324.158**	10.9002	80.367
11	**612.289**	10.1612	77.8196
12	**305.14**	11.6397	**75.3176**
13	**607.255**	12.5794	74.5788
14	**245.119**	16.1224	69.9729
15	**243.151**	13.7847	68.7066
16	**485.128**	12.6293	68.1374
17	**612.289**	11.7562	66.738
18	**375.141**	11.9751	65.7093
19	**387.096**	6.63516	65.3792
20	**389.121**	11.1017	63.3308
21	**455.192**	11.2526	56.1093
22	**209.046**	6.9629	55.6501
23	**175.061**	8.07922	54.7582
24	**338.097**	12.1089	51.9405
25	**467.227**	13.284	49.6127
26	**353.101**	9.5319	47.3696
27	**611.285**	10.1449	46.9872
28	**285.041**	14.1246	46.6395
29	**395.093**	15.2493	45.5864
30	**495.244**	12.8976	44.5685
31	**271.097**	10.6987	44.5337
32	**675.31**	11.0506	41.678
33	**486.244**	11.4879	41.4939
34	**607.096**	9.50668	41.1506
35	**673.306**	11.3618	39.9828
36	**611.285**	11.6388	39.2041
37	**557.223**	8.7502	38.8658
38	**613.257**	10.1954	37.5089
39	**623.287**	8.83438	37.2942
40	**543.261**	12.5789	35.8708
41	**385.108**	13.8475	34.9158
42	**373.186**	17.4147	34.5972
43	**643.298**	9.13702	33.0104
44	**201.088**	8.70061	32.4277
45	**501.247**	14.3761	31.5025
46	**289.108**	13.4524	30.9946
47	**229.109**	13.1838	30.5331
48	**397.633**	15.1903	29.7946
49	**390.127**	10.6551	27.5161
50	**596.286**	12.1594	25.4874
51	**177.056**	8.38183	25.4579
52	**352.976**	17.197	24.8143
53	**676.307**	11.0258	24.4463
54	**163.033**	14.2925	24.2451
55	**371.171**	15.3649	24.0658
56	**482.262**	14.813	23.6146
57	**653.312**	10.7318	22.7727
58	**453.205**	19.1695	22.7301
59	**617.237**	14.8207	22.1553
60	**397.128**	14.6449	21.8303

#	Var ID (Primary)	Var ID (RT)	M2.Score Contrib(Obs E184 - Average),
61	**659.112**	9.38048	20.5992
62	**324.156**	7.8775	20.4786
63	**484.169**	12.2266	20.3103
64	**661.346**	10.6985	19.9041
65	**635.3**	9.47257	19.8814
66	**407.265**	11.6219	19.0436
67	**522.252**	14.2586	18.9773
68	**633.194**	10.4813	18.7573
69	**395.233**	14.7281	18.659
70	**330.083**	7.64257	18.5853
71	**544.262**	12.7305	18.4439
72	**626.315**	17.5494	18.3284
73	**644.331**	9.35525	18.041
74	**412.98**	13.637	17.8632
75	**304.055**	11.4379	17.8096
76	**635.078**	8.80924	17.3943
77	**433.075**	7.18913	17.3884
78	**496.254**	12.9996	17.3208
79	**313.105**	13.67	15.944
80	**201.15**	9.06973	15.5583
81	**520.254**	12.5791	15.5009
82	**502.176**	15.1322	15.4452
83	**691.138**	8.91008	15.3664
84	**249.15**	12.2517	15.1608
85	**266.065**	8.66775	14.7941
86	**240.005**	13.9234	14.6371
87	**416.181**	14.4606	14.5537
88	**237.041**	7.13898	14.2575
89	**265.072**	8.38994	14.0528
90	**194.054**	9.4898	13.743
91	**401.109**	8.71776	13.7116
92	**676.182**	15.3496	13.6392
93	**419.121**	7.05483	13.4791
94	**369.166**	15.3324	13.4435
95	**369.145**	9.55713	13.27
96	**505.191**	8.48181	12.3537
97	**575.147**	9.1441	12.2145
98	**421.073**	7.57517	12.149
99	**230.112**	12.9732	12.1248
100	**307.168**	9.28812	11.402
101	**688.233**	8.63288	11.2472
102	**535.088**	7.20637	11.2266
103	**402.184**	16.8448	10.9
104	**614.297**	10.9334	10.4841
105	**149.025**	8.86829	10.3789
106	**317.133**	9.52303	10.2816
107	**339.144**	8.95194	10.159
108	**314.005**	13.066	10.1524
109	**272.069**	10.3215	10.041
110	**535.219**	6.8703	10.0104
111	**674.31**	11.9834	9.96395
112	**386.172**	12.4279	9.88915
113	**640.23**	14.1581	9.87102
114	**641.366**	19.3955	9.85953
115	**147.059**	7.79397	9.73527
116	**221.082**	10.5146	9.57192
117	**337.165**	9.92667	9.48261
118	**243.087**	10.2935	9.48052
119	**391.288**	12.2935	9.26233
120	**301.119**	7.87797	9.22127

S6A. Top 120 variable contributions from the PCA (ESI+) Scores Contribution plot of E-112.

#	Var ID (Primary)	Var ID (RT(min))	M6.Score Contrib(Obs E112 - Average), Weight=p1p2
1	**437.342**	17.188	131.785
2	**524.266**	14.083	74.5229
3	**438.345**	16.802	70.0669
4	**223.133**	18.6145	69.7242
5	**355.176**	15.5093	69.3874
6	**407.28**	17.4059	69.3114
7	**423.29**	14.7715	65.4178
8	**707.148**	15.4934	60.7347
9	**320.149**	14.2428	60.7171
10	**638.304**	14.8693	57.3476
11	**699.358**	19.5206	56.9172
12	**469.331**	14.9059	54.2724
13	**631.348**	17.2381	51.2808
14	**453.337**	17.0285	48.8869
15	**441.3**	13.7976	48.7738
16	**354.145**	15.51	46.6762
17	**705.152**	15.5092	45.6153
18	**453.336**	15.1073	44.0455
19	**659.285**	14.3689	43.06
20	**951.615**	19.2856	42.0633
21	**505.353**	17.5407	40.5136
22	**706.154**	15.493	40.1665
23	**1069.5**	14.8693	39.8292
24	**1085.76**	17.2692	39.5786
25	**487.304**	19.3359	39.067
26	**550.395**	20.1248	38.648
27	**1048.53**	14.8555	37.5935
28	**419.287**	17.0698	36.0578
29	**503.274**	18.4967	34.8525
30	**405.222**	17.271	34.1338
31	**490.26**	15.7922	34.0022
32	**239.128**	13.0764	32.9618
33	**525.287**	15.6104	32.8301
34	**708.151**	15.4936	32.7717
35	**952.62**	19.3196	32.5238
36	**637.296**	14.3687	30.5745
37	**819.231**	9.92118	30.4136
38	**1049.53**	14.8558	30.3206
39	**456.259**	14.8549	29.9006
40	**420.256**	17.1712	29.8412
41	**469.332**	17.1547	29.2711
42	**456.344**	14.5872	28.8581
43	**508.269**	15.7951	28.3847
44	**1015.53**	15.7952	27.4886
45	**632.342**	16.7099	27.4142
46	**240.137**	16.9867	27.0236
47	**490.259**	17.9934	26.9911
48	**467.226**	18.5982	26.8339
49	**315.055**	13.6804	26.4371
50	**429.295**	18.8326	26.3096
51	**469.332**	18.9139	26.2774
52	**408.347**	17.1549	25.9884
53	**267.107**	11.5158	25.6229
54	**679.403**	19.5042	25.2813
55	**551.357**	19.6385	25.1339
56	**727.131**	15.4934	25.0273
57	**981.688**	13.8815	24.7802
58	**203.179**	18.7322	24.637
59	**1047.52**	14.872	24.5448
60	**1037.51**	15.7952	23.8271

#	Var ID (Primary)	Var ID (RT(min))	M6.Score Contrib(Obs E112 - Average),
61	**442.231**	14.9389	23.7166
62	**463.286**	18.3792	23.587
63	**647.386**	16.7321	23.1257
64	**407.19**	11.8822	22.5804
65	**660.277**	14.3825	22.4943
66	**1087.76**	17.3644	22.3562
67	**239.108**	7.2693	22.2092
68	**573.381**	19.8567	21.8316
69	**453.351**	12.2282	21.696
70	**408.28**	17.1628	21.3285
71	**488.306**	19.4539	20.9562
72	**307.118**	9.9378	20.7578
73	**215.081**	11.8848	20.6102
74	**729.13**	15.4928	20.5196
75	**504.279**	18.4972	20.5167
76	**982.742**	13.8817	20.4126
77	**527.393**	18.4466	20.2622
78	**430.238**	15.7951	20.0652
79	**207.138**	19.2859	19.4003
80	**435.327**	15.2149	19.3829
81	**224.136**	18.6148	19.0476
82	**909.696**	17.2554	19.043
83	**470.368**	19.0512	19.0426
84	**317.126**	12.422	18.9195
85	**953.623**	19.3193	18.3975
86	**639.309**	13.58	18.2207
87	**341.162**	14.7719	17.91
88	**727.302**	11.2298	17.4023
89	**459.309**	17.9434	17.2998
90	**491.244**	15.7446	17.1253
91	**453.279**	19.0847	16.9934
92	**1064.55**	14.8556	16.9786
93	**351.063**	15.6948	16.8652
94	**1035.04**	19.521	16.5292
95	**694.402**	19.518	16.5292
96	**470.335**	17.1038	16.3727
97	**910.645**	17.2717	15.9698
98	**1121.77**	20.1249	15.816
99	**1138.7**	18.0007	15.5654
100	**1034.54**	19.521	15.4941
101	**1010.58**	17.9771	15.4535
102	**459.273**	12.5222	15.2906
103	**487.342**	16.0807	15.1359
104	**979.519**	14.8625	14.9118
105	**728.135**	15.4932	14.7109
106	**1067**	14.8724	14.4631
107	**470.335**	13.412	14.4199
108	**455.276**	12.7406	14.4168
109	**1011.57**	18.0089	14.3583
110	**464.735**	18.631	14.2286
111	**1065.55**	14.8555	13.9556
112	**457.358**	13.8813	13.8915
113	**583.178**	10.592	13.7375
114	**569.872**	20.1247	13.3346
115	**728.298**	11.566	13.3165
116	**514.337**	17.6409	13.014
117	**422.349**	17.2216	13.0096
118	**983.745**	13.8791	12.836
119	**439.284**	14.469	12.8302
120	**287.15**	10.7015	12.7329

S7A. Top 120 variable contributions from the PCA (ESI+) Scores Contribution plot of E-225.

#	Var ID (Primary)	Var ID (RT(min))	M2.Score Contrib(Obs E225 - Average), Weight=p1p2
1	**293.172**	15.0564	10.9175
2	**607.326**	14.3155	10.7613
3	**245.128**	9.91124	10.6258
4	**775.432**	15.1157	10.5766
5	**361.203**	19.403	10.3388
6	**608.335**	14.3011	8.83095
7	**384.152**	17.188	7.92701
8	**285.106**	15.5098	7.86694
9	**211.144**	8.98087	7.29287
10	**261.124**	7.11802	7.23871
11	**294.168**	14.4105	6.22701
12	**594.349**	18.1947	6.1574
13	**197.128**	7.4538	5.81427
14	**377.204**	14.7211	5.52031
15	**776.441**	15.5518	4.97094
16	**359.139**	16.7017	4.84925
17	**544.15**	16.5671	4.6877
18	**383.22**	19.6213	4.5731
19	**390.192**	18.6984	4.30195
20	**458.308**	16.7684	4.0688
21	**544.652**	16.7095	3.85309
22	**286.132**	16.6677	3.74317
23	**548.129**	15.51	3.68951
24	**275.162**	14.2004	3.58124
25	**315.157**	17.3972	3.49733
26	**687.301**	11.0454	3.45693
27	**604.33**	13.5295	3.43253
28	**438.197**	19.3193	3.22455
29	**825.207**	16.5505	3.14859
30	**475.118**	11.3224	3.14164
31	**294.185**	16.8353	3.10969
32	**249.084**	2.26828	3.0688
33	**526.068**	16.7347	3.04631
34	**675.776**	16.5339	3.04097
35	**205.047**	7.03412	2.97011
36	**743.448**	19.6214	2.88637
37	**263.128**	16.2647	2.88179
38	**779.467**	19.5206	2.7691
39	**351.166**	9.75297	2.76396
40	**628.136**	16.5505	2.68993
41	**719.23**	19.1516	2.65427
42	**597.285**	18.5308	2.48493
43	**563.122**	16.5153	2.45286
44	**521.239**	7.10159	2.37697
45	**676.203**	16.4396	2.3686
46	**698.39**	19.6119	2.35436
47	**316.156**	16.1143	2.26699
48	**826.21**	16.5839	2.25712
49	**268.081**	8.69588	2.21592
50	**201.139**	16.3491	2.05164
51	**599.063**	18.363	2.04128
52	**695.449**	14.7714	2.03776
53	**362.183**	17.0204	2.03458
54	**384.223**	19.6357	1.98713
55	**425.227**	10.5925	1.86302
56	**246.121**	15.9124	1.85208
57	**801.35**	18.4297	1.8316
58	**615.314**	9.48513	1.79131
59	**321.141**	11.3474	1.75568
60	**655.221**	9.88773	1.71774
61	**413.654**	16.4998	1.70468
62	**243.111**	7.96591	1.69704
63	**351.165**	7.45418	1.66967
64	**450.718**	17.0039	1.63678
65	**315.189**	19.6721	1.61281
66	**300.051**	9.38372	1.60899
67	**277.118**	8.39375	1.59573
68	**404.232**	19.8731	1.56204
69	**745.244**	17.8841	1.55028
70	**663.36**	18.1275	1.54648
71	**255.056**	11.6159	1.53858
72	**419.112**	19.6968	1.52038
73	**483.276**	12.1533	1.51934
74	**600.253**	18.9338	1.50344
75	**206.081**	9.43301	1.49189
76	**359.233**	14.5869	1.4195
77	**470.723**	14.3183	1.39107
78	**412.212**	18.9163	1.38676
79	**172.069**	9.2666	1.37984
80	**744.452**	18.883	1.37629
81	**461.14**	19.3787	1.37602
82	**676.696**	16.5507	1.36894
83	**260.112**	10.4078	1.35598
84	**339.056**	12.8582	1.35302
85	**591.127**	18.3966	1.31749
86	**542.135**	8.10797	1.30611
87	**280.157**	16.8276	1.28896
88	**511.128**	2.28438	1.28546
89	**182.079**	4.75217	1.27312
90	**267.077**	2.2676	1.27009
91	**264.159**	16.3491	1.26242
92	**299.14**	11.4065	1.24531
93	**742.282**	17.9516	1.24432
94	**710.811**	18.1102	1.24248
95	**542.31**	14.0649	1.22741
96	**756.378**	11.3814	1.22487
97	**646.848**	14.3658	1.21518
98	**518.127**	9.26687	1.20406
99	**271.146**	19.8063	1.20349
100	**362.237**	19.6216	1.19818
101	**609.327**	13.6469	1.18884
102	**331.152**	11.5485	1.17054
103	**961.56**	15.96	1.16243
104	**663.16**	17.7587	1.14175
105	**282.124**	11.3478	1.12001
106	**604.818**	16.0972	1.11071
107	**827.212**	15.8944	1.10973
108	**777.466**	12.9676	1.09049
109	**977.461**	10.2899	1.08035
110	**632.18**	18.2272	1.07512
111	**243.106**	15.7611	1.06648
112	**816.233**	19.6724	1.0631
113	**607.347**	19.6554	1.0555
114	**568.171**	11.868	1.04432
115	**530.618**	16.6681	1.0391
116	**466.229**	14.654	1.02899
117	**1087.29**	16.2987	1.02254
118	**506.296**	19.6722	1.01944
119	**659.122**	17.5401	1.00947
120	**329.197**	9.98815	1.00919

S8A. Top 60 variable contributions from the PCA (ESI+) Scores Contribution plot of E-260.

#	Var ID (Primary)	Var ID (RT(min))	M2.Score Contrib(Obs E260 - Average), Weight=p1p2
1	**309.054**	9.88773	15.8678
2	**291.12**	10.24	15.1161
3	**639.096**	9.88767	14.9072
4	**491.231**	12.7235	13.3769
5	**305.125**	10.6092	12.8945
6	**331.043**	10.0306	9.19464
7	**291.122**	15.2914	8.81133
8	**279.121**	9.38413	8.55312
9	**640.1**	9.88745	7.72324
10	**543.127**	12.8251	7.10094
11	**569.092**	12.9919	6.78349
12	**603.075**	14.6119	6.39121
13	**313.032**	14.6709	5.93525
14	**310.128**	10.2735	5.3597
15	**339.127**	8.36068	5.16981
16	**583.178**	8.67751	5.13137
17	**659.367**	17.296	4.92798
18	**358.368**	19.907	4.8237
19	**572.112**	9.71705	4.19013
20	**589.347**	13.6131	4.02183
21	**551.094**	15.7113	4.01172
22	**641.118**	9.9127	3.87842
23	**288.121**	11.1129	3.78223
24	**553.135**	15.8791	3.68214
25	**535.102**	14.3349	3.64495
26	**253.118**	9.04858	3.63569
27	**283.027**	12.8585	3.57336
28	**303.108**	8.76288	3.52489
29	**477.228**	8.69472	3.50754
30	**570.096**	12.9923	3.48553
31	**655.208**	7.35352	3.30016
32	**604.078**	14.889	3.21289
33	**544.133**	13.58	3.15616
34	**492.237**	12.2373	3.09905
35	**636.591**	9.92127	3.0383
36	**317.094**	7.30337	2.99596
37	**250.177**	15.5936	2.97464
38	**509.129**	12.7407	2.9426
39	**510.136**	10.9617	2.66027
40	**282.111**	8.3103	2.65533
41	**533.111**	10.735	2.59795
42	**631.143**	11.381	2.59431
43	**571.112**	12.9926	2.50784
44	**584.181**	8.6789	2.38458
45	**621.097**	10.5673	2.37303
46	**482.062**	9.92127	2.27411
47	**1009.25**	13.7976	2.20526
48	**552.119**	16.5339	2.13627
49	**304.116**	8.95612	2.10606
50	**829.232**	15.3923	2.1015
51	**573.202**	9.86782	2.06711
52	**323.124**	8.32693	2.01632
53	**676.278**	9.60229	2.01531
54	**577.154**	10.3574	1.96712
55	**512.578**	13.3449	1.96375
56	**600.11**	14.7546	1.94997
57	**287.045**	14.1338	1.9397
58	**667.135**	9.11555	1.90295
59	**601.091**	14.7713	1.85624
60	**287.05**	8.5451	1.83827

S9A. Top 60 variable contributions from the PCA (ESI+) Scores Contribution plot of E-195.

#	Var ID (Primary)	Var ID (RT(min))	M2.Score Contrib(Obs E195 - Average), Weight=p1p2
1	**195.099**	9.24995	60.1561
2	**155.088**	9.78682	48.0271
3	**317.212**	20.0747	38.4007
4	**457.233**	11.398	34.3492
5	**158.126**	16.1627	29.698
6	**950.597**	17.9771	28.3021
7	**953.637**	14.2512	27.9056
8	**480.267**	18.1614	24.6906
9	**953.625**	16.7349	24.4242
10	**954.641**	14.2343	23.6806
11	**425.304**	17.0373	23.5751
12	**155.107**	15.2245	23.2584
13	**367.132**	8.29292	23.226
14	**426.293**	17.0368	22.8016
15	**441.3**	13.7976	20.7412
16	**458.202**	11.3644	20.3431
17	**442.338**	14.3347	20.0884
18	**407.28**	17.4059	19.5921
19	**935.59**	14.2654	19.5619
20	**221.153**	18.5138	19.4583
21	**273.154**	16.1811	19.2508
22	**954.628**	16.4668	17.7049
23	**655.332**	19.2695	17.0476
24	**491.231**	12.7235	16.877
25	**317.214**	14.2651	15.6286
26	**936.55**	14.2841	14.9286
27	**332.675**	17.0035	14.6802
28	**173.097**	11.3478	14.3085
29	**475.296**	12.2862	12.9958
30	**179.105**	17.322	12.8149
31	**223.206**	16.9528	12.7129
32	**173.117**	9.20773	12.68
33	**156.102**	9.25792	11.9816
34	**955.543**	14.2678	11.4628
35	**949.667**	19.1249	11.1949
36	**269.222**	15.5176	11.0997
37	**535.344**	13.1268	10.4926
38	**167.107**	12.0102	10.1863
39	**955.592**	16.4495	9.9025
40	**311.237**	10.9616	9.87683
41	**274.163**	16.1977	9.81707
42	**1103.52**	19.5794	9.81173
43	**560.38**	17.5654	9.64014
44	**935.626**	16.3826	9.62692
45	**386.218**	17.3389	9.51163
46	**458.351**	11.5993	9.27589
47	**221.19**	13.6302	9.24416
48	**404.339**	17.2955	9.091
49	**460.263**	16.5159	9.03193
50	**273.16**	14.4856	8.97168
51	**439.254**	11.8514	8.92868
52	**1101.77**	20.091	8.85225
53	**444.332**	17.7251	8.61902
54	**457.266**	14.8553	8.05383
55	**203.179**	13.6298	7.89409
56	**443.351**	11.9853	7.81121
57	**333.206**	17.0018	7.79214
58	**492.237**	12.2373	7.40841
59	**289.127**	15.6774	7.39738
60	**565.347**	19.638	7.18977

APPENDEX B

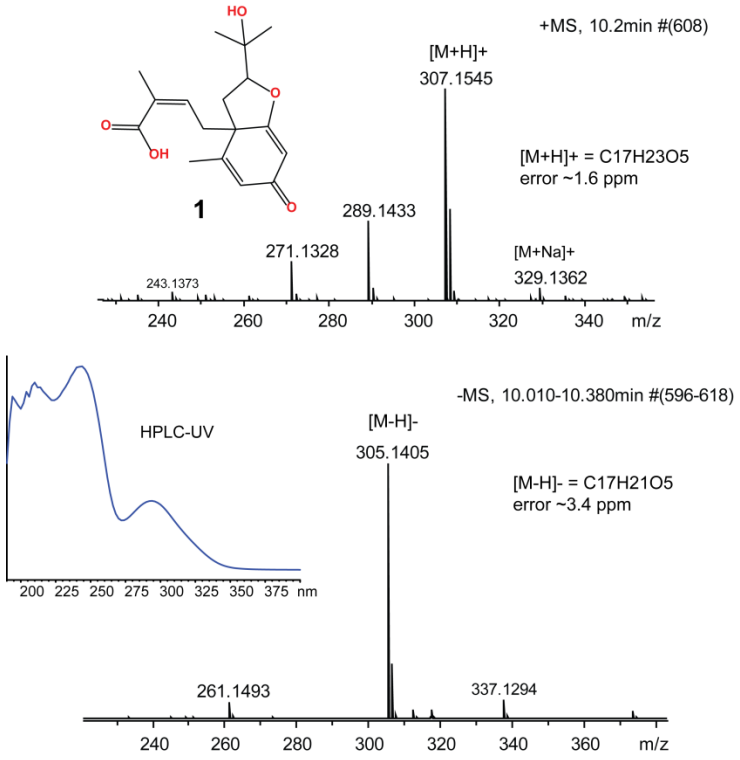


Figure S1 B. LC-HRMS and LC-UV spectra of compound **1.**

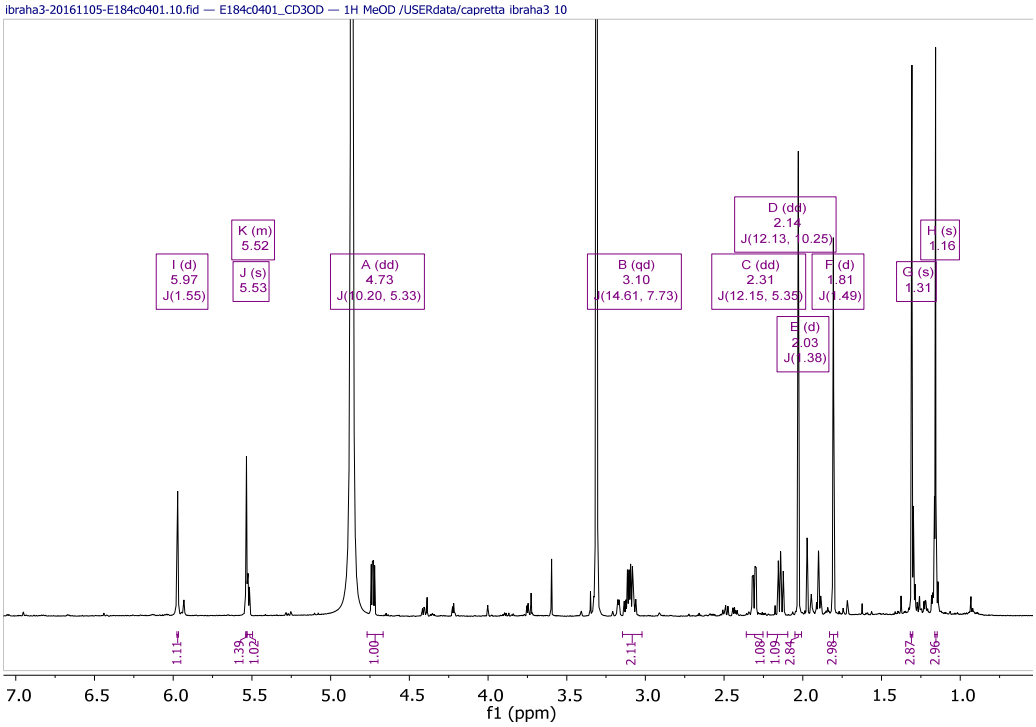


Figure S2 B. ^{1}H NMR spectrum (700 MHz,CD_3OD) of compound **1.**

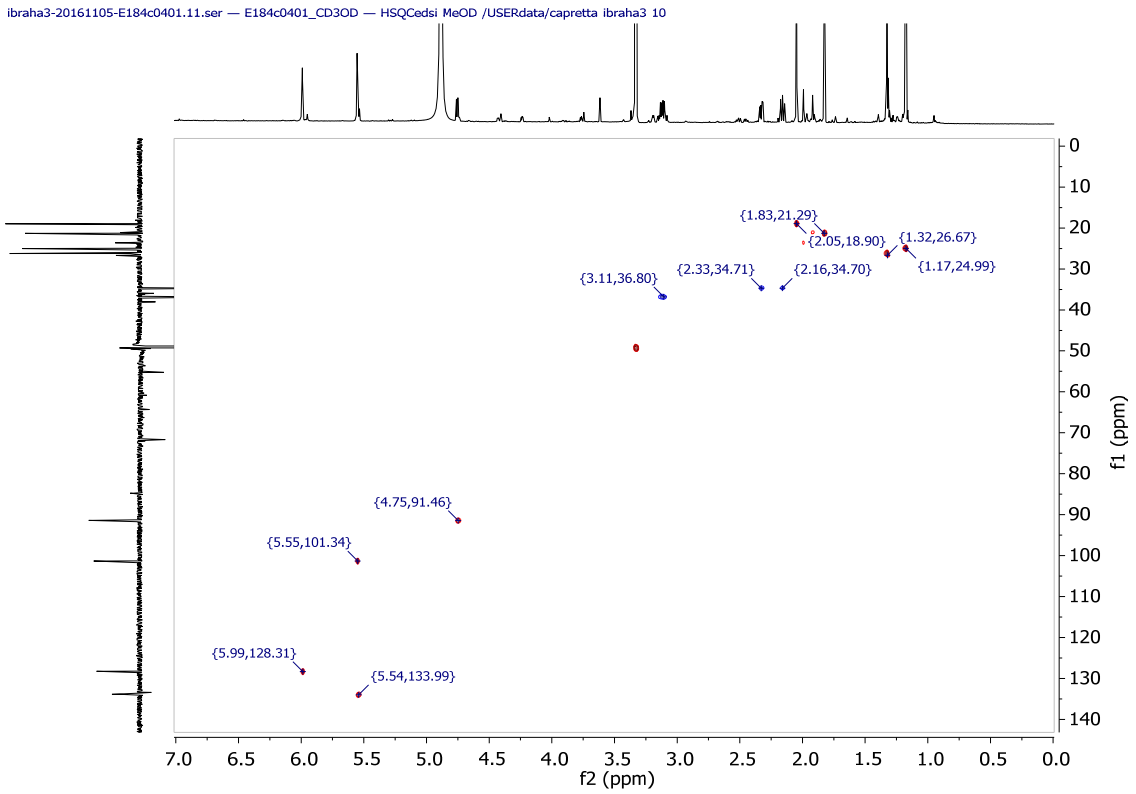


Figure S3 B. Multiplicity-edited HSQC NMR spectrum (CD_3OD) of compound **1.**

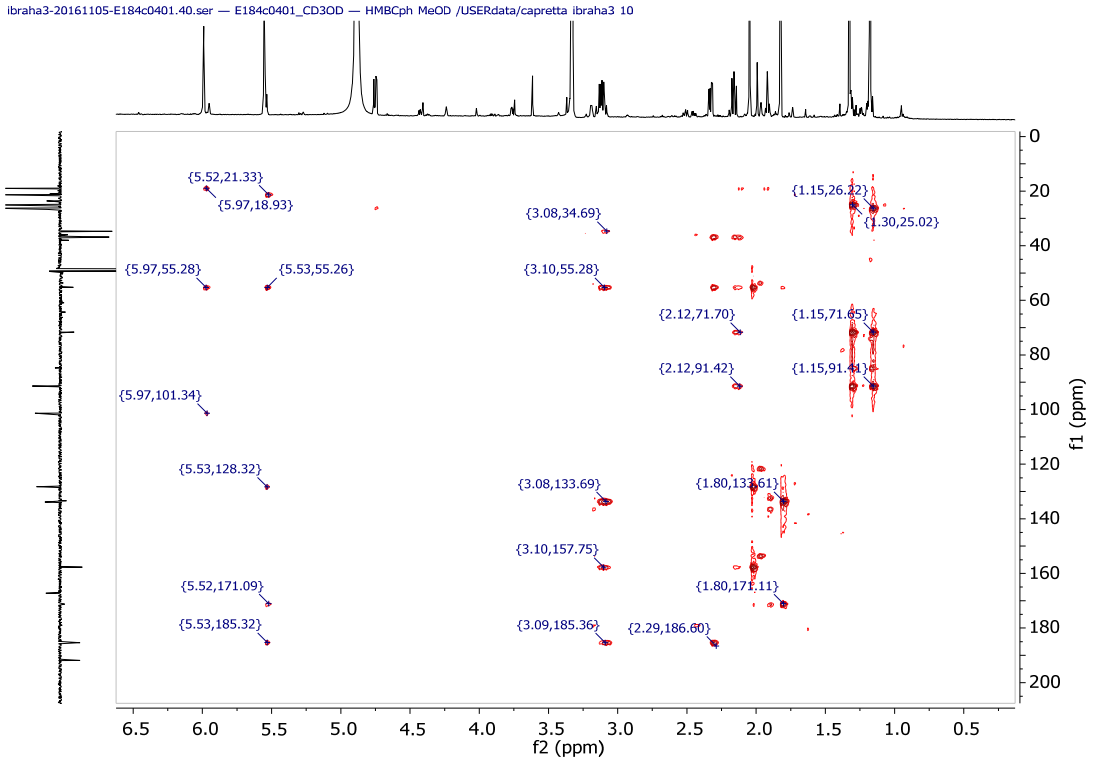


Figure S4 B. HMBC NMR spectrum (CD_3OD) of compound **1.**

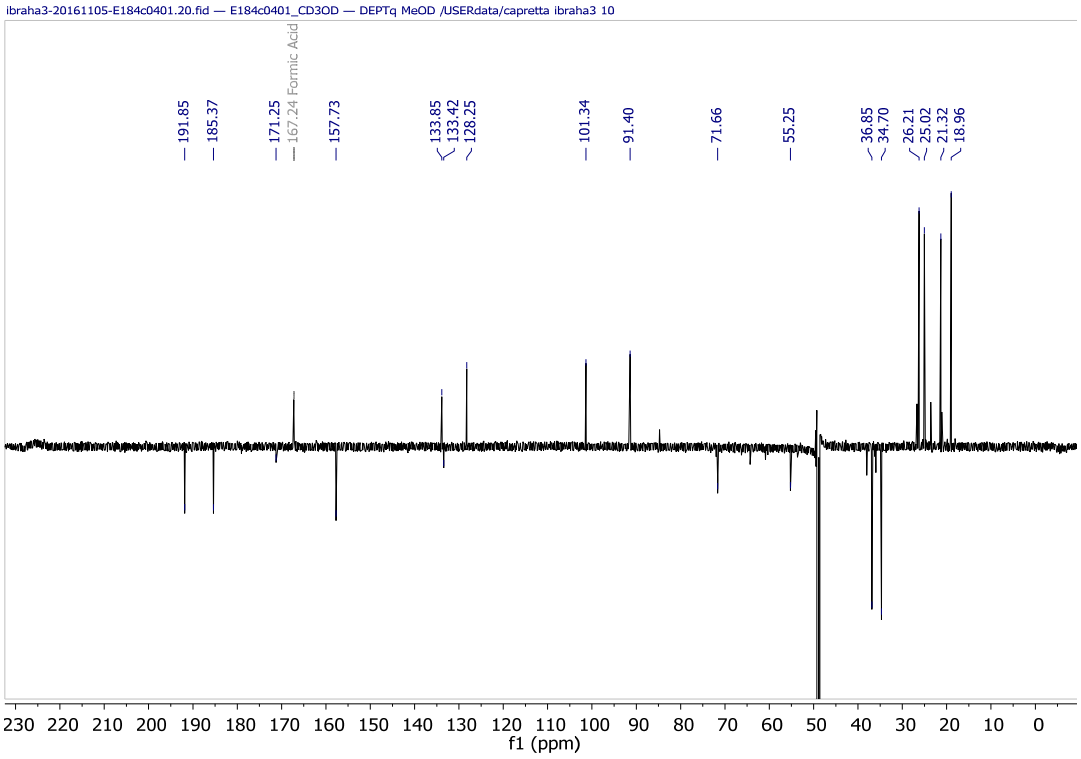


Figure S5 B. ^{13}C DEPTq NMR spectrum (176 MHz,CD_3OD) of compound **1.**

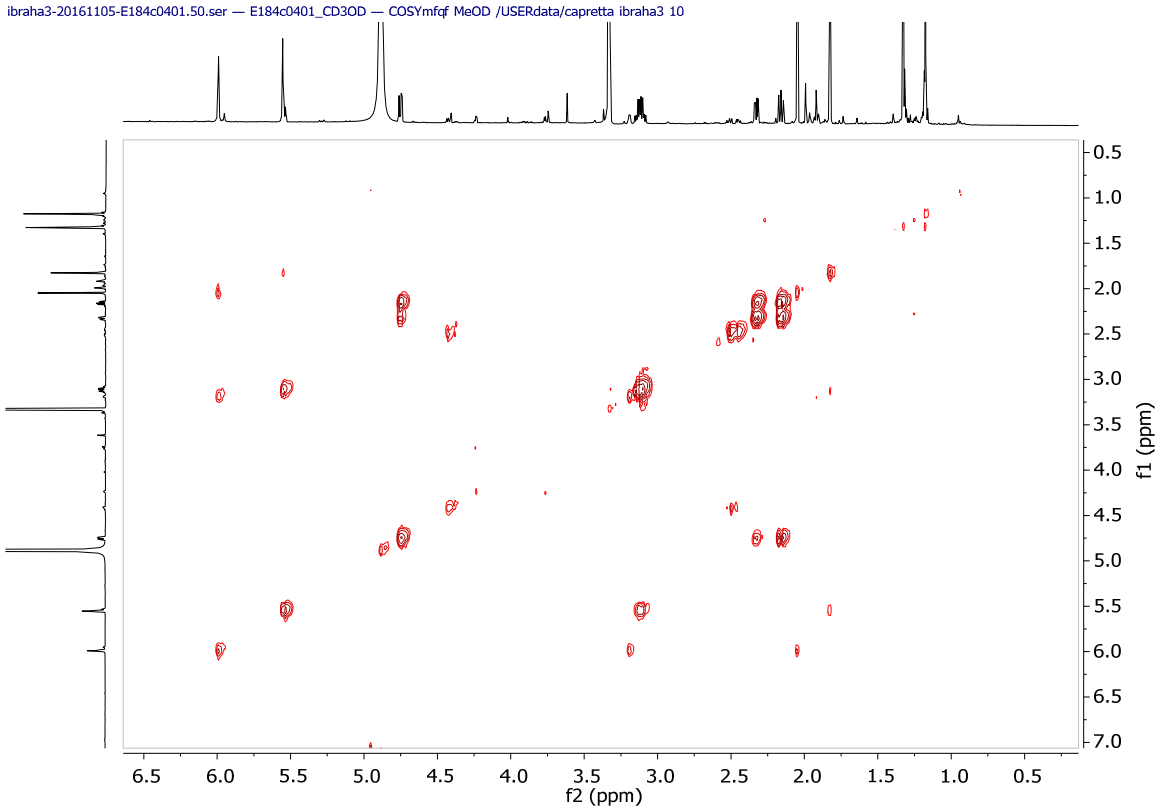


Figure S6 B. COSY NMR spectrum (700 MHz, CD_3OD) of compound **1.**

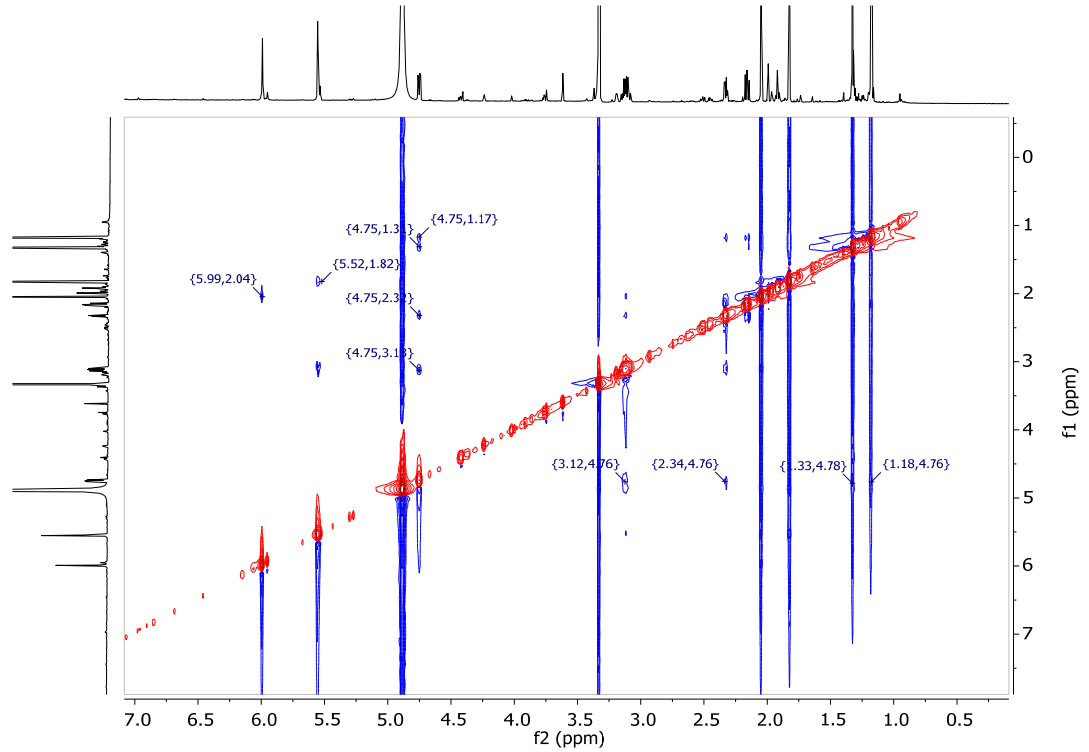


Figure S7 B. ROESY 300ms NMR spectrum (700 MHz, CD_3OD) of compound **1.**

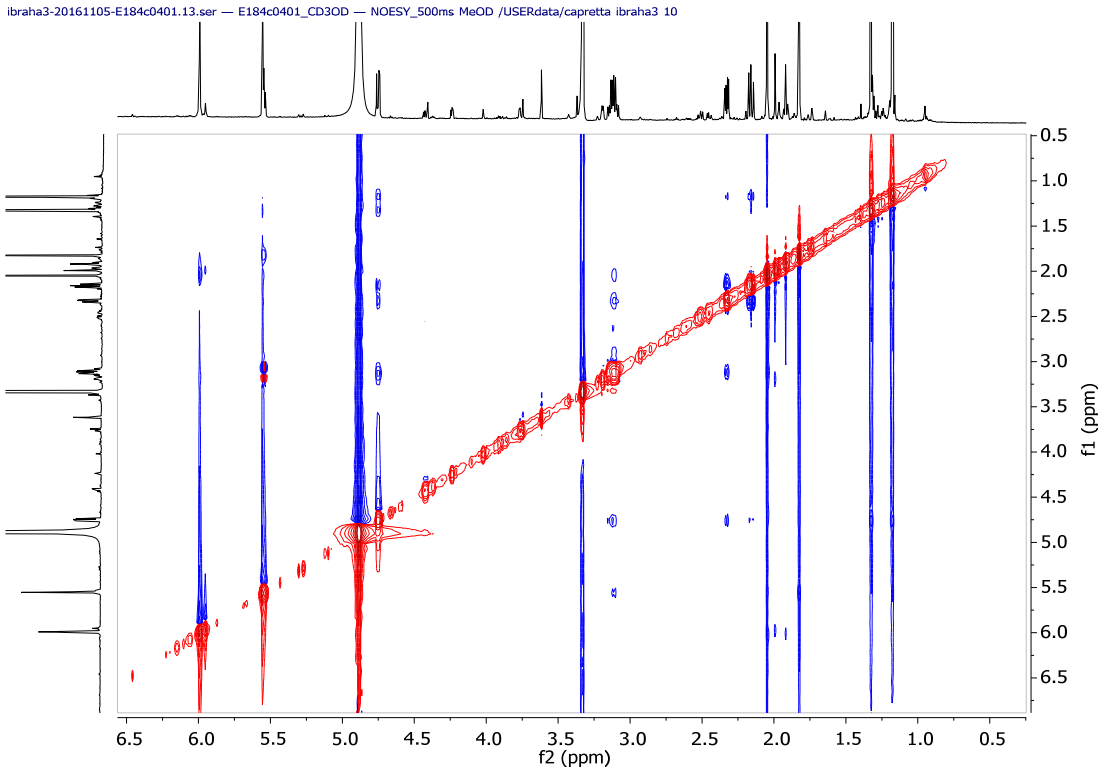


Figure S8 B. NOESY 500ms NMR spectrum (CD_3OD) of compound **1.**

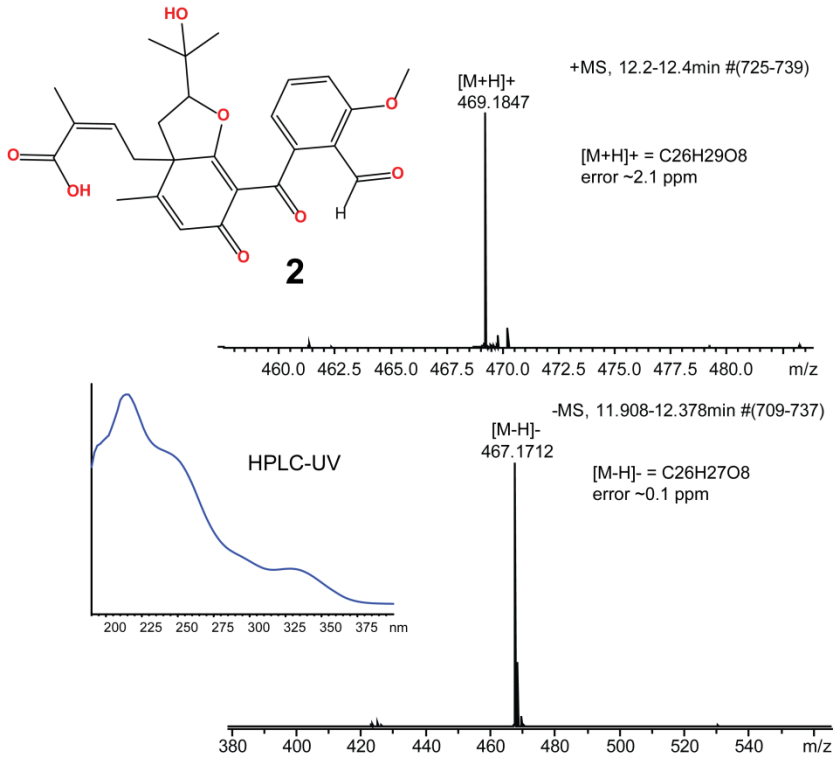


Figure S9 B. LC-HRMS and LC-UV spectra of compound **2.**

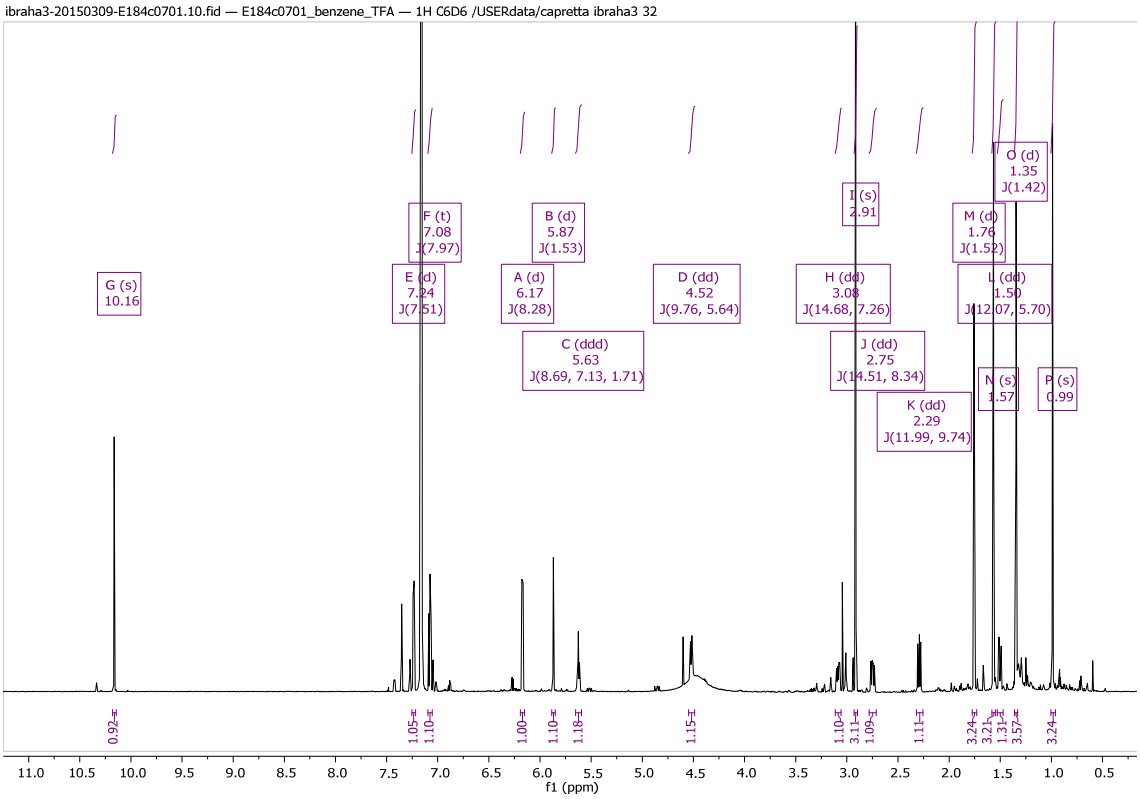


Figure S10 B. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **2.**

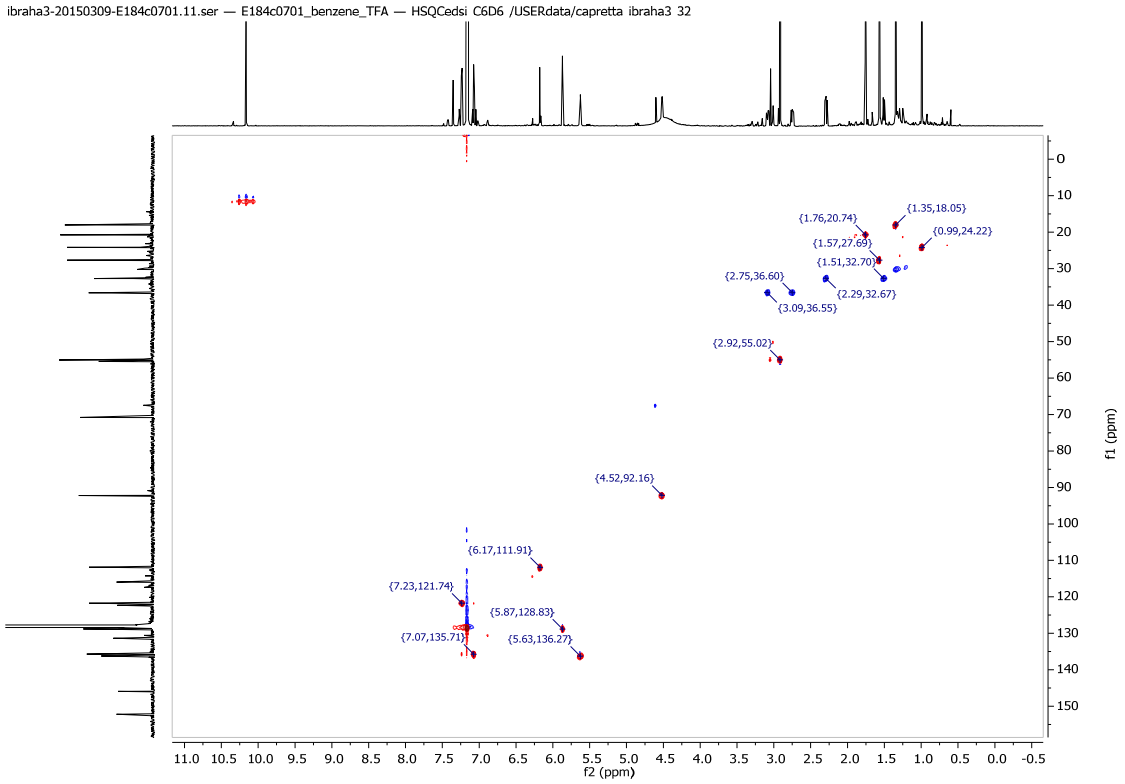


Figure S11 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **2.**

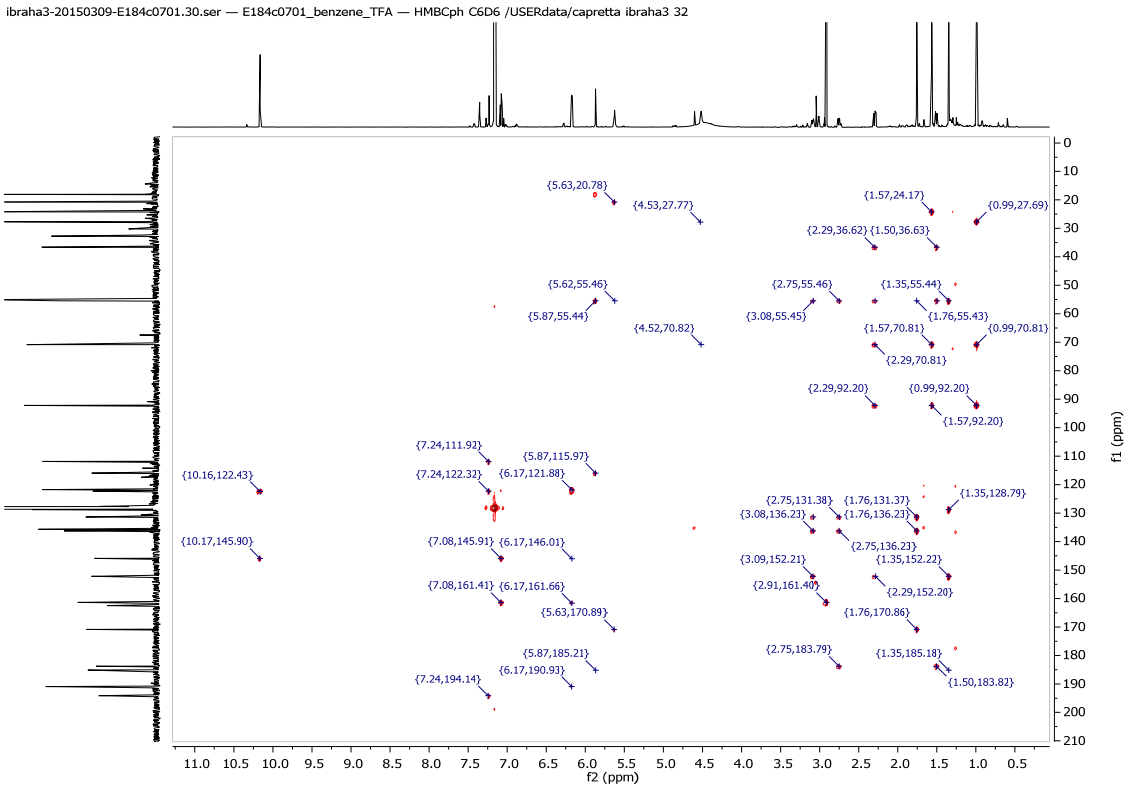


Figure S12 B. HMBC NMR spectrum (C_6D_6) of compound **2.**

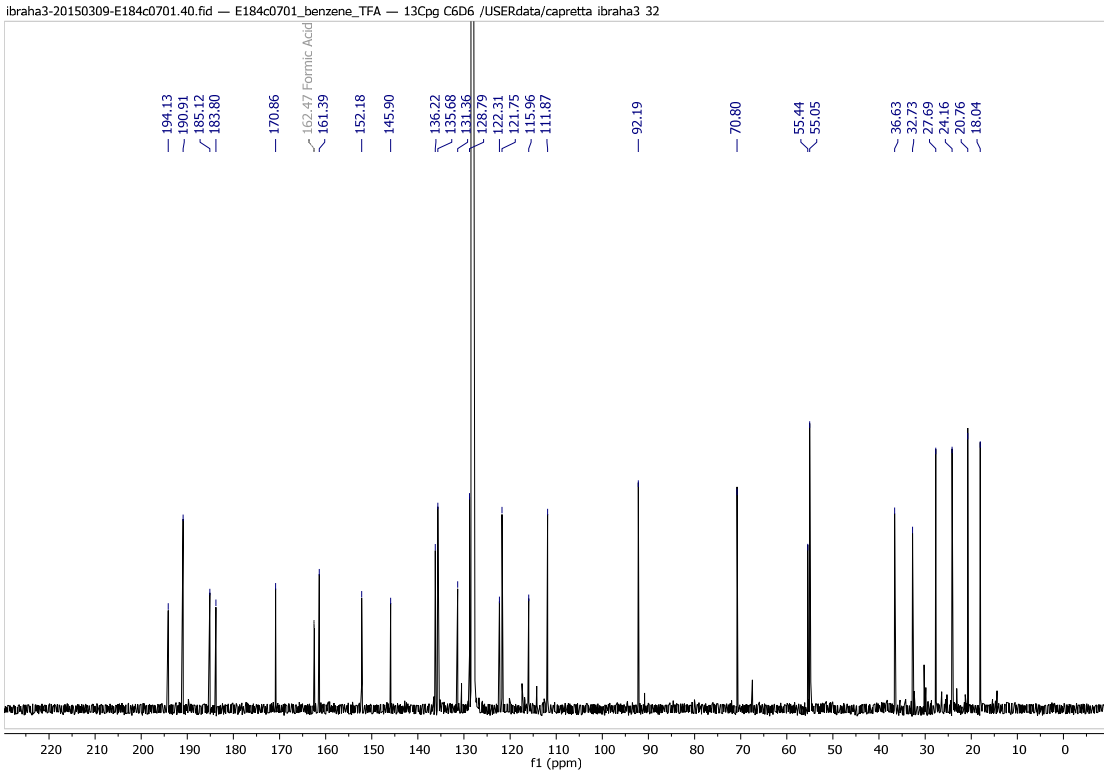


Figure S13 B. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **2.**

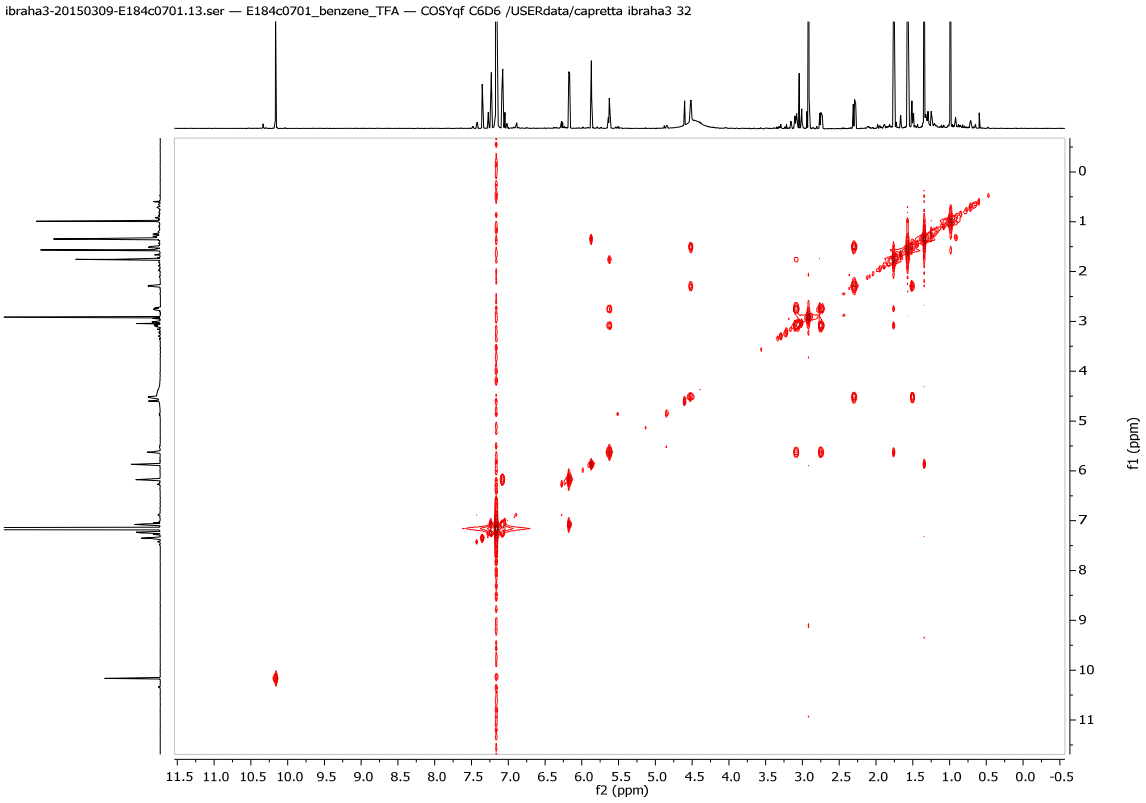


Figure S14 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **2.**

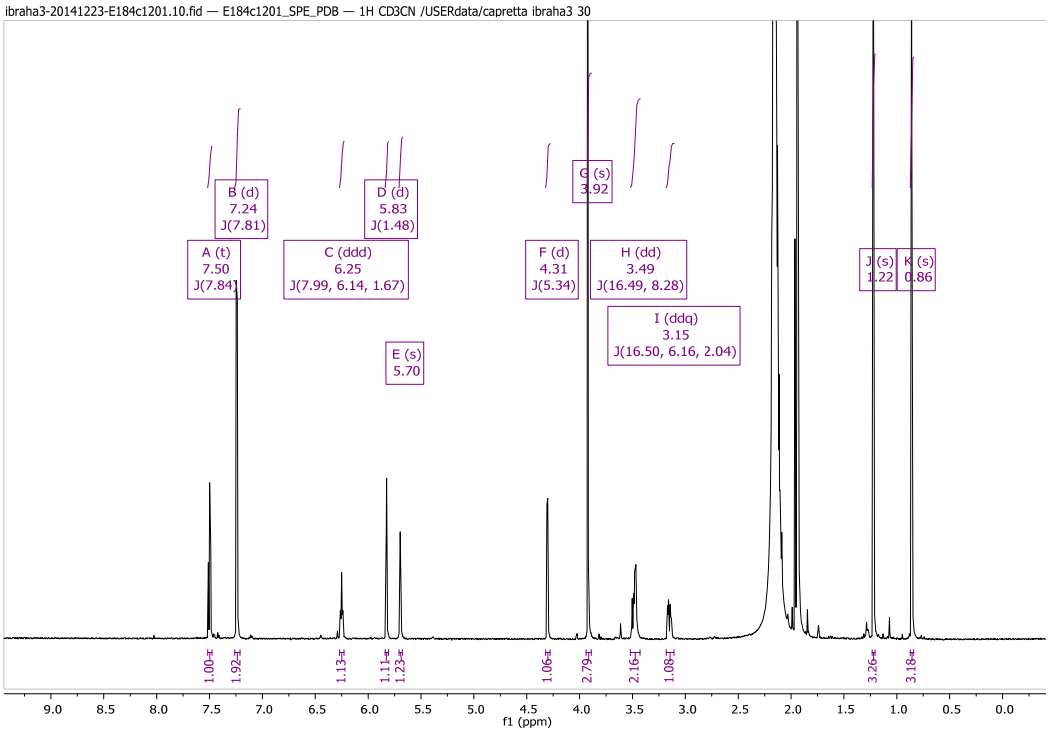


Figure S15 B. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **3.**

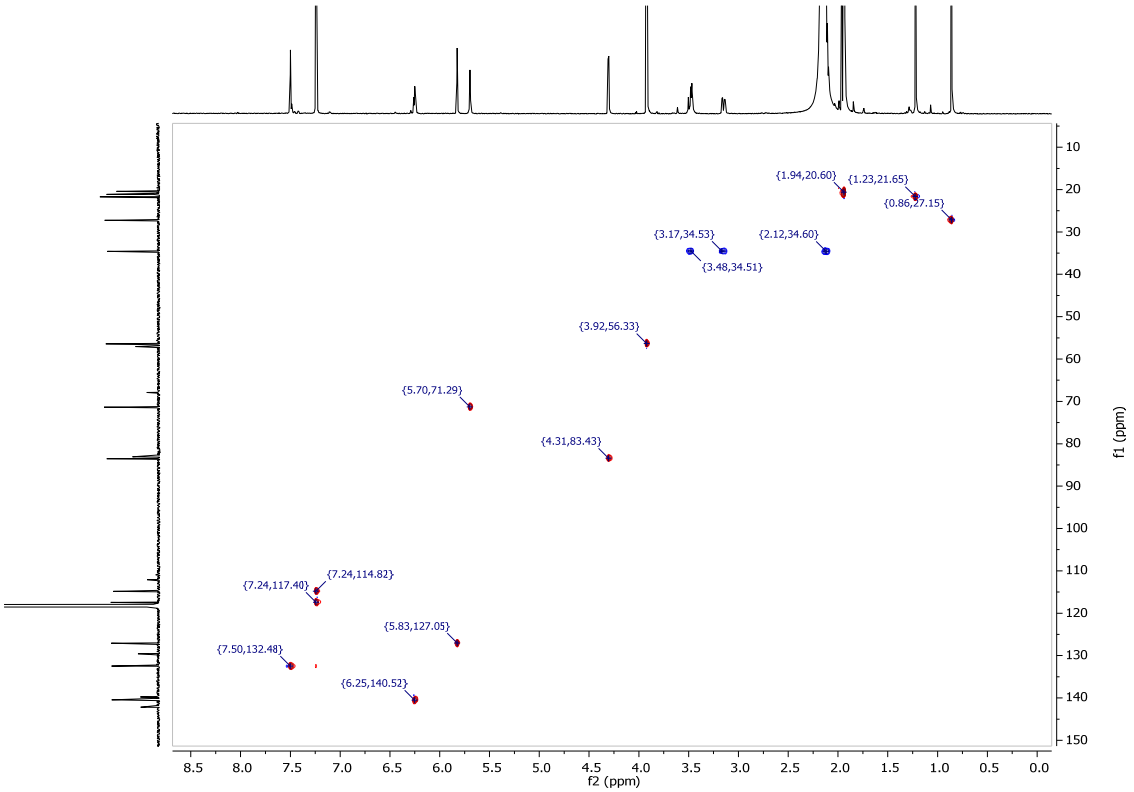


Figure S16 B. Multiplicity-edited HSQC NMR spectrum (CD_3CN) of compound **3.**

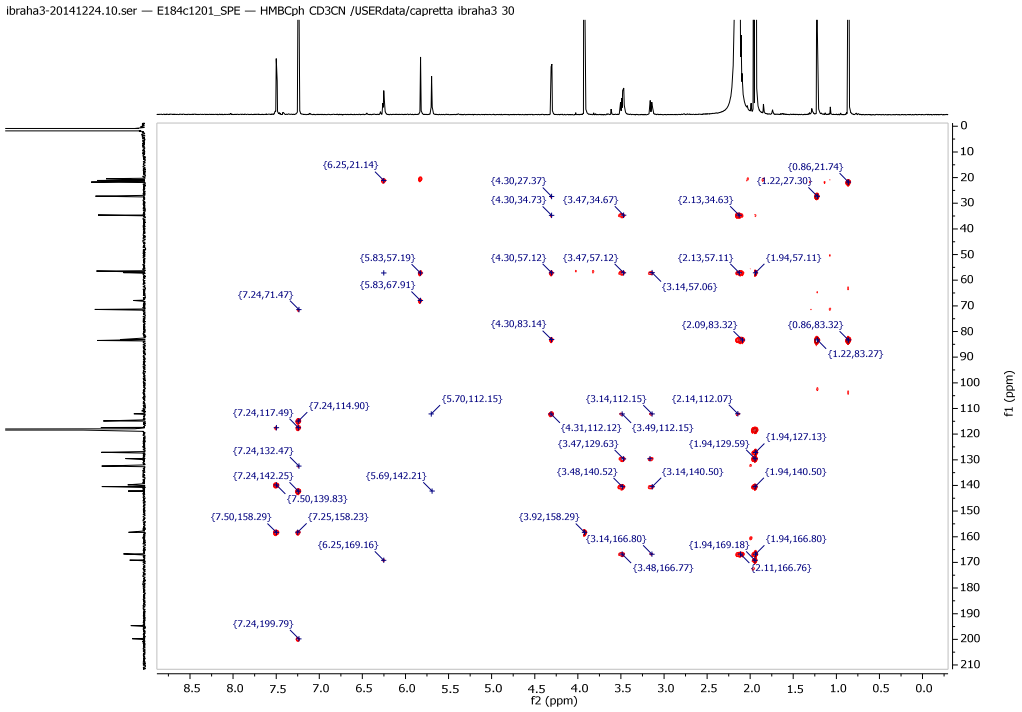


Figure S17 B. HMBC NMR spectrum (CD_3CN) of compound **3.**

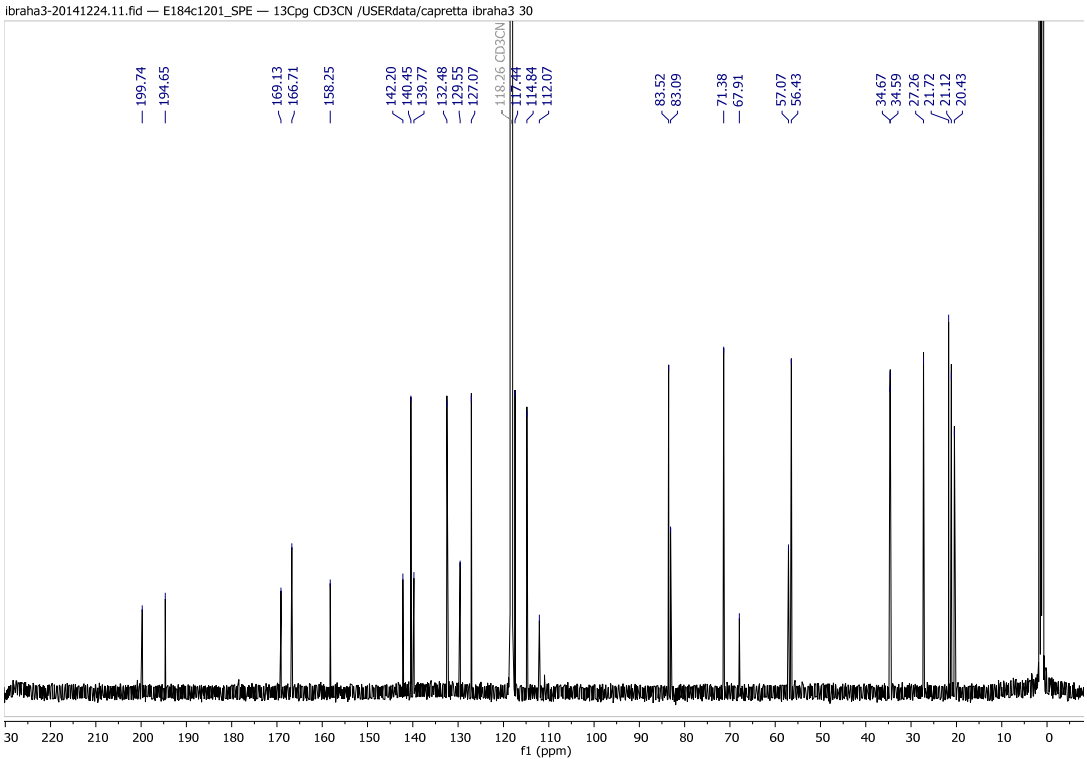


Figure S18 B. ^{13}C NMR spectrum (176 MHz, CD_3CN) of compound **3.**

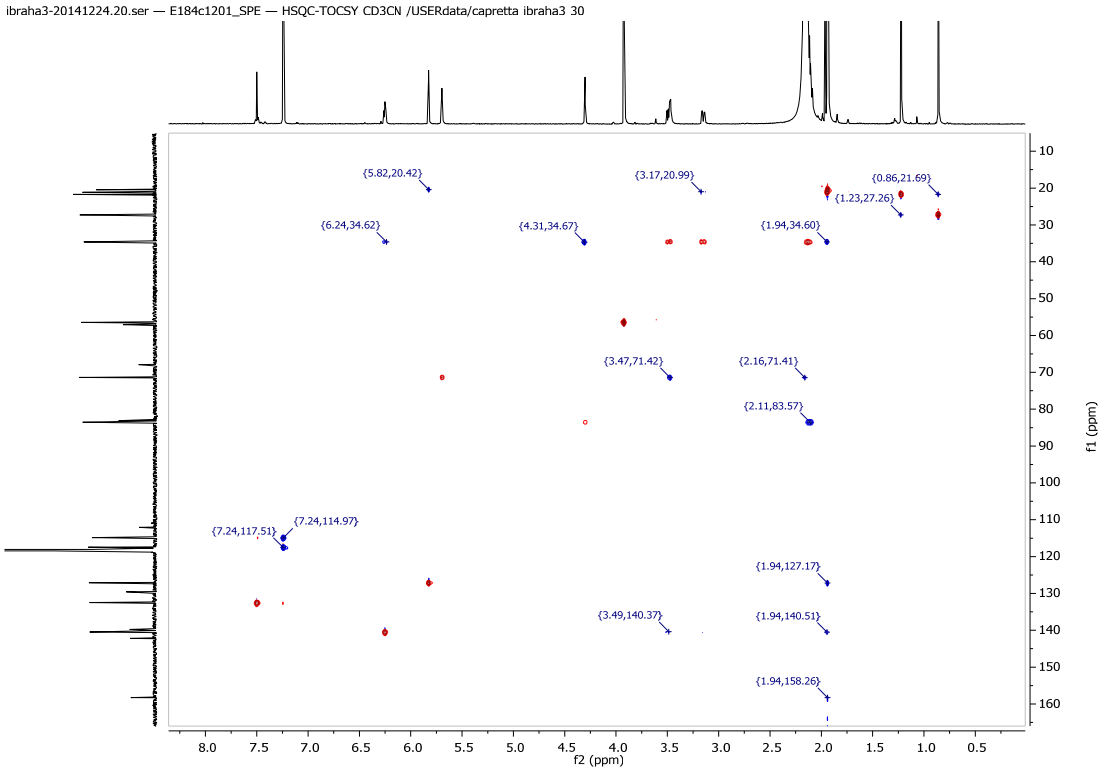


Figure S19 B. HSQC-TOCSY spectrum (CD_3CN) of compound **3.**

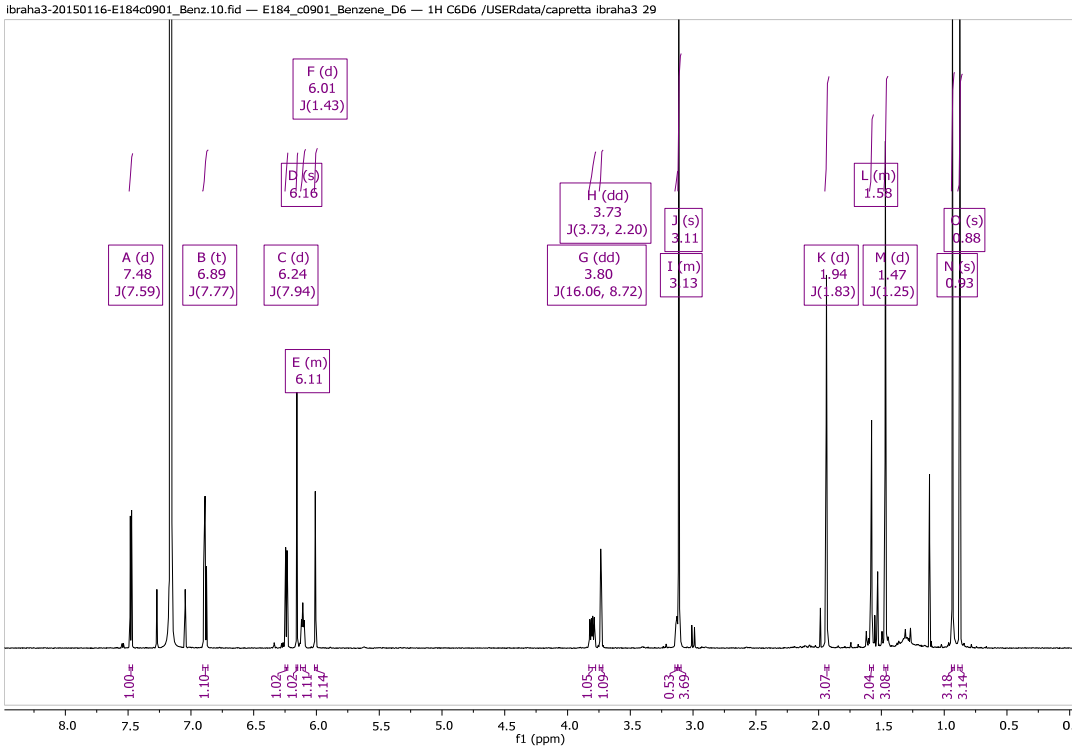


Figure S20 B. ^1H NMR spectrum (700 MHz, C_6D_6) of compound **3.**

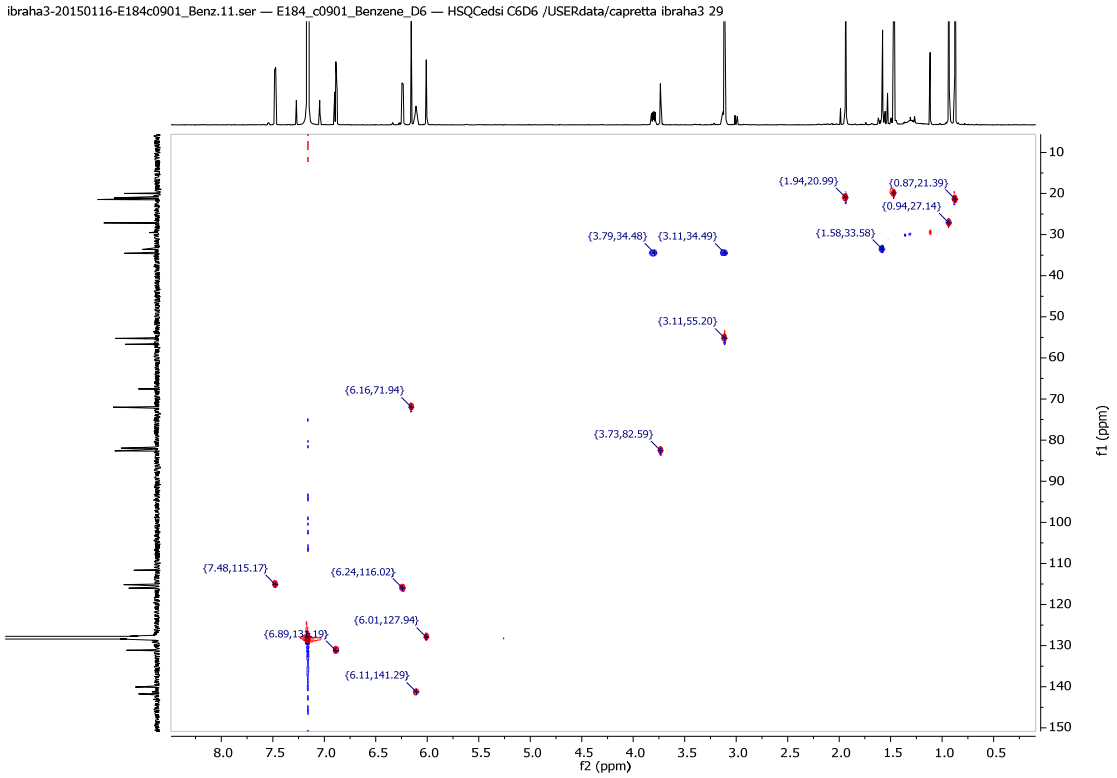


Figure S21 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **3.**

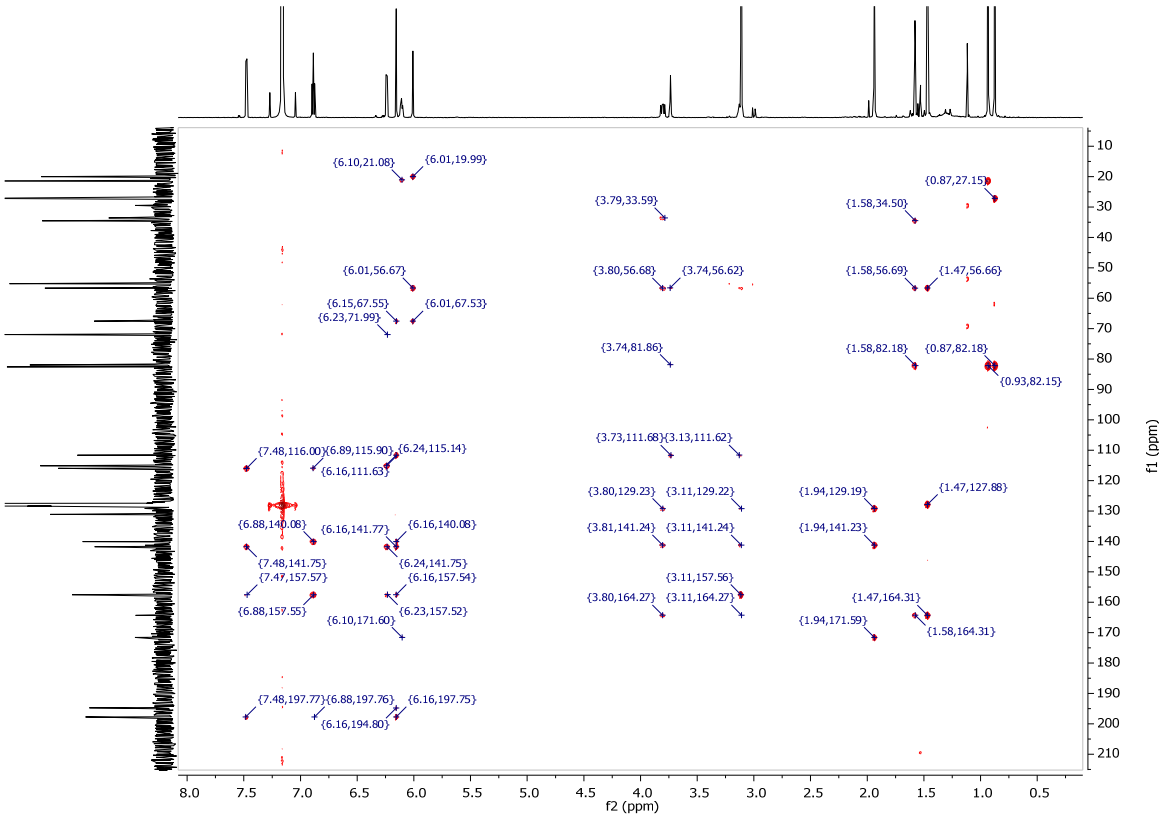


Figure S22 B. HMBC NMR spectrum (C_6D_6) of compound **3.**

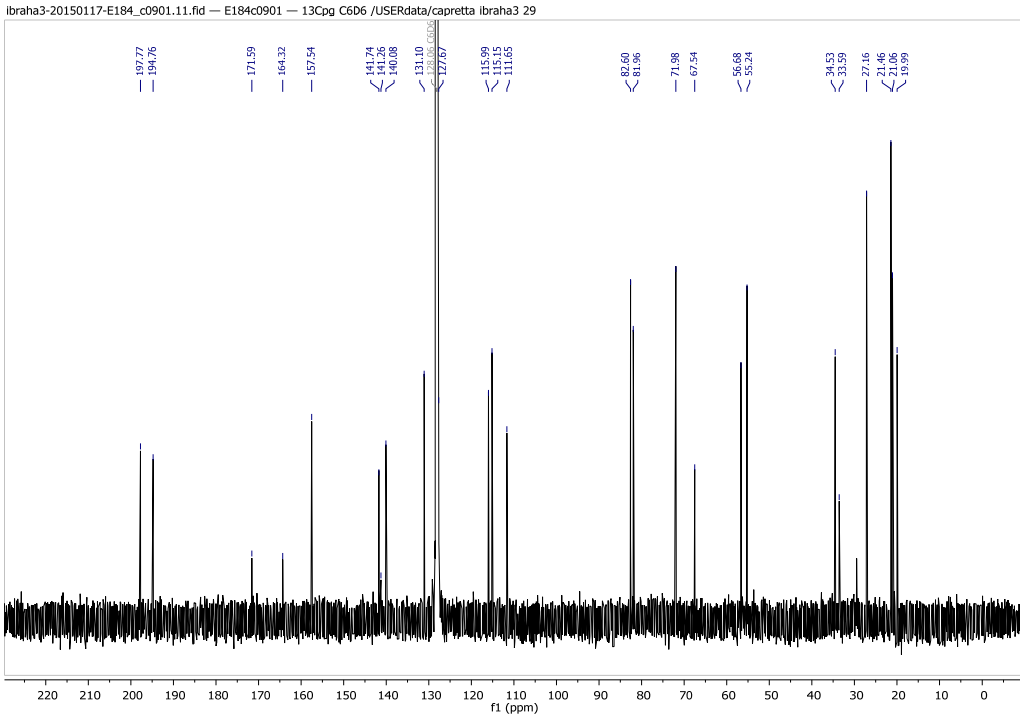


Figure S23 B. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **3.**

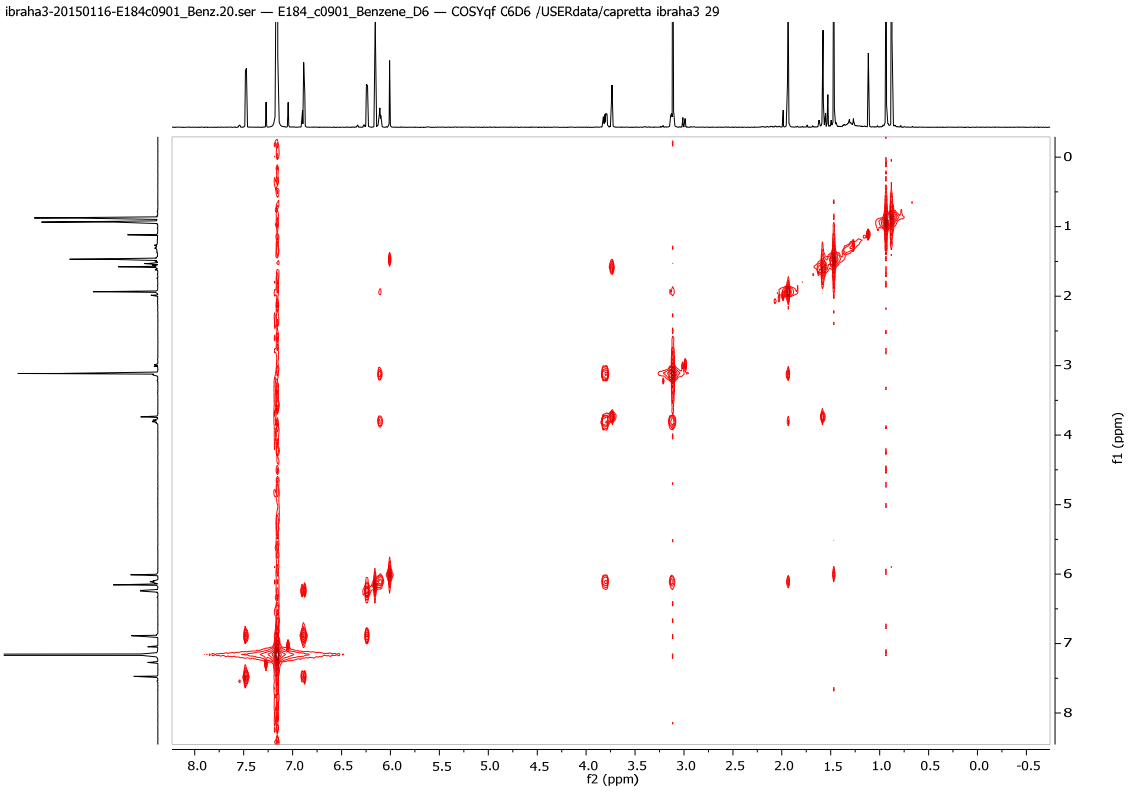


Figure S24 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **3.**

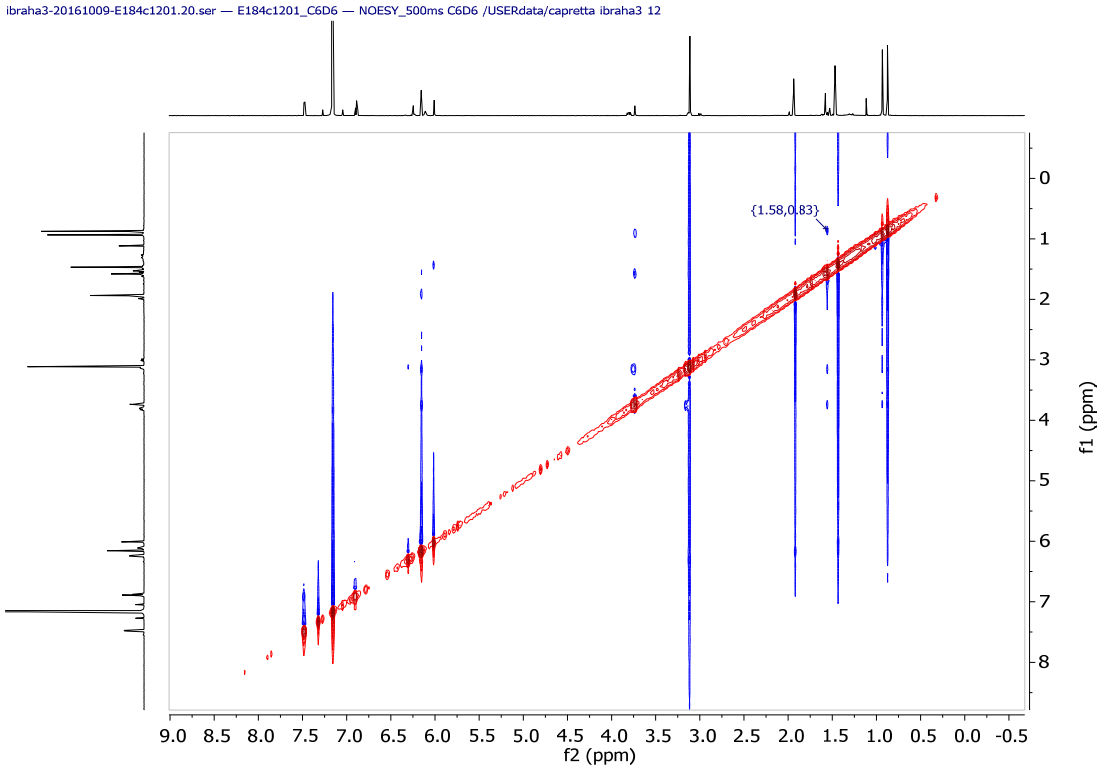


Figure S25 B. NOESY 500ms NMR spectrum (C_6D_6) of compound **3.**

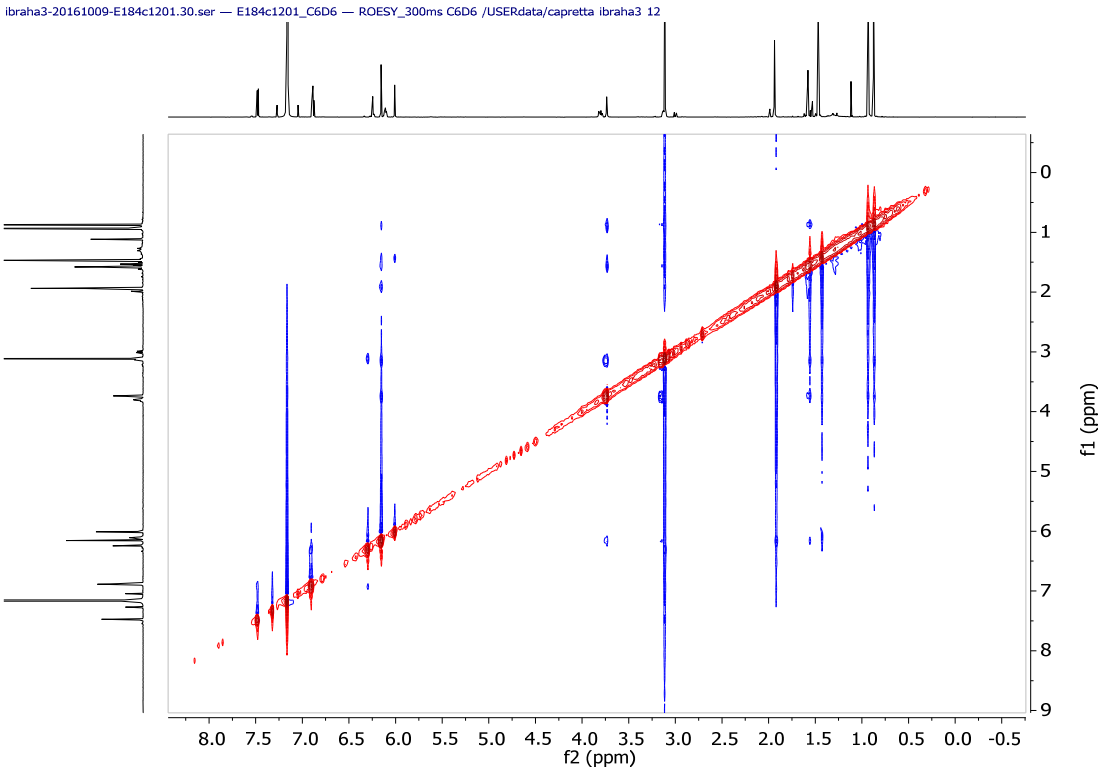


Figure S26 B. ROESY 300ms NMR spectrum (C_6D_6) of compound **3.**

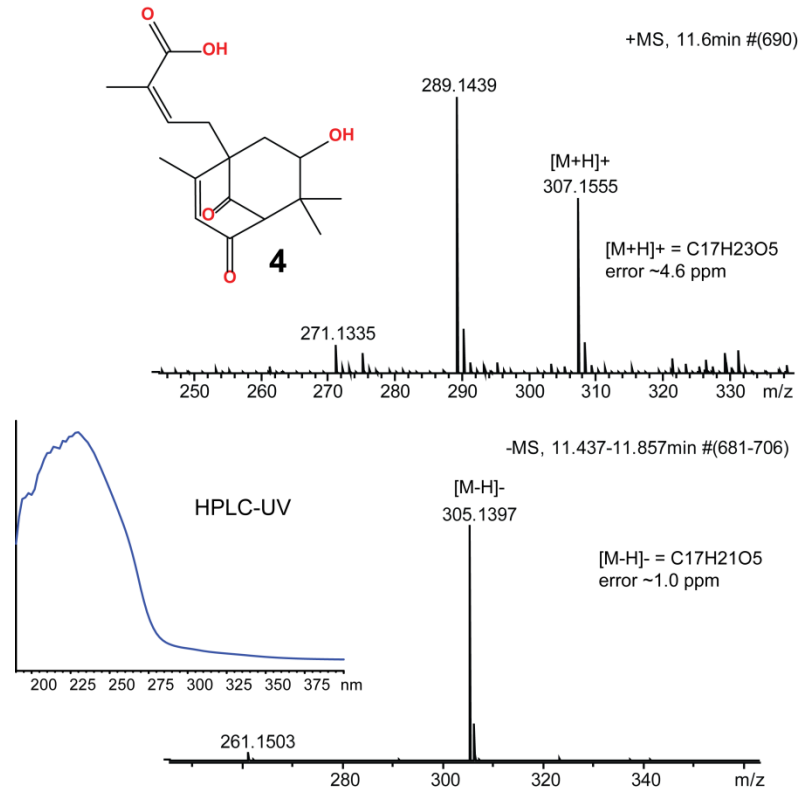


Figure S27 B. LC-HRMS and LC-UV spectra of compound **4.**

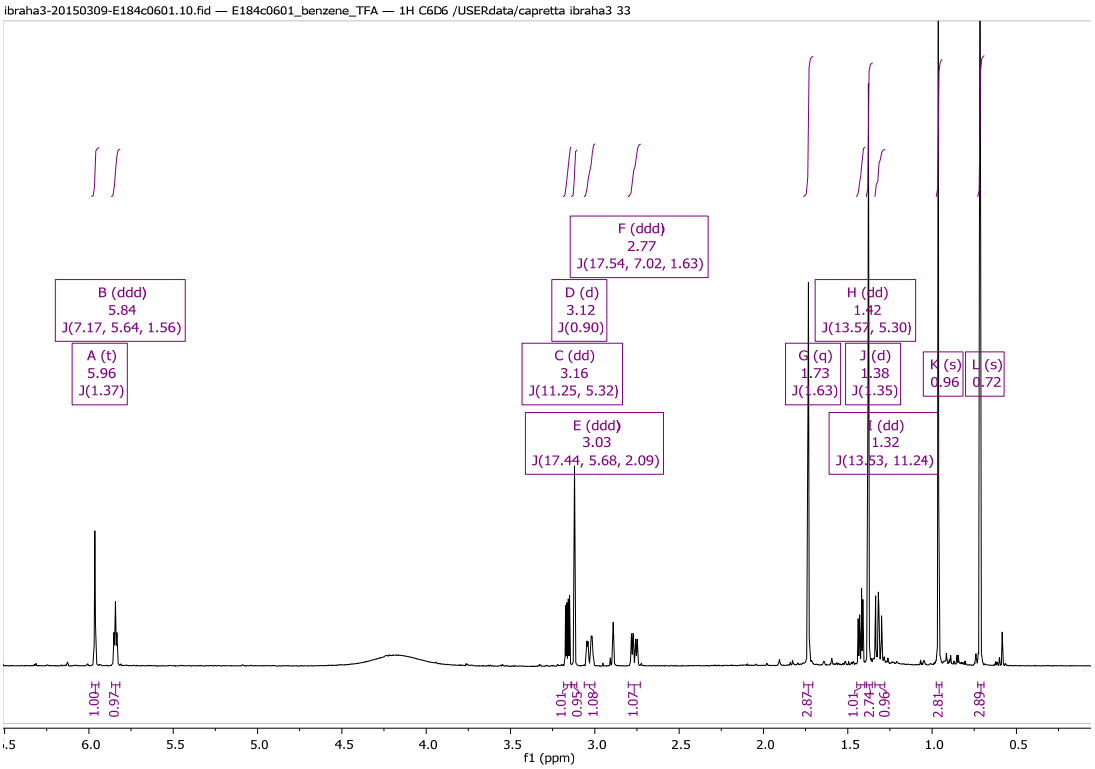


Figure S28 B. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **4.**

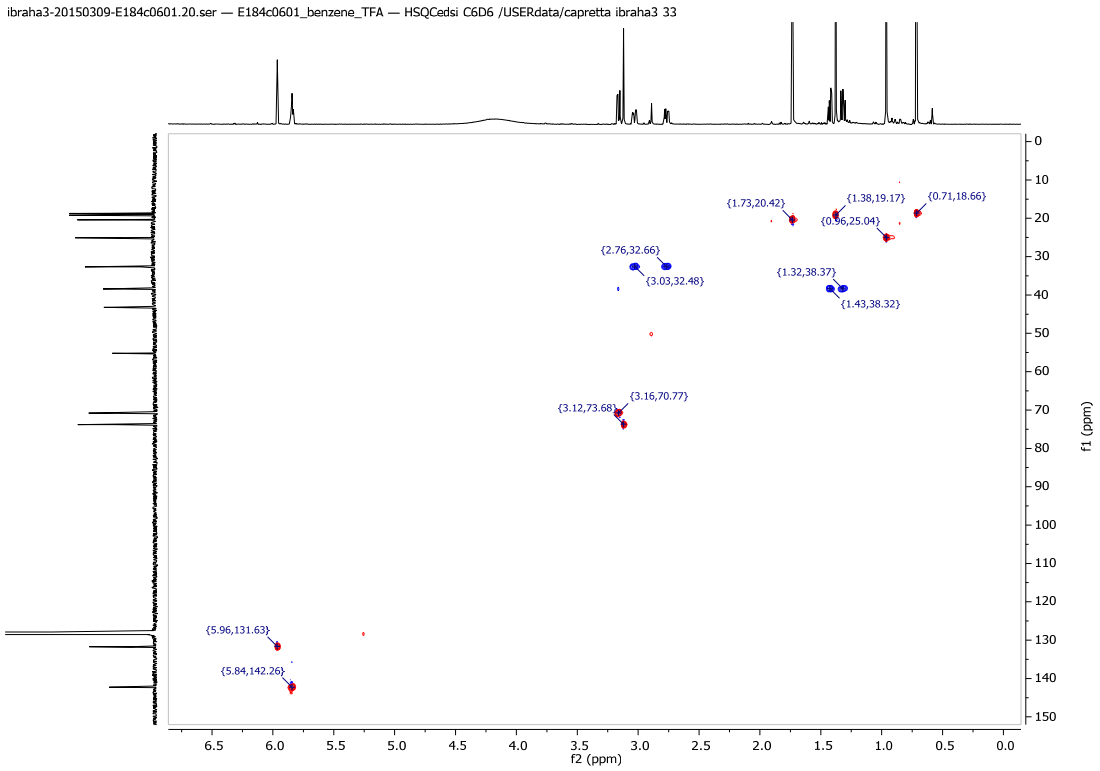


Figure S29 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **4.**

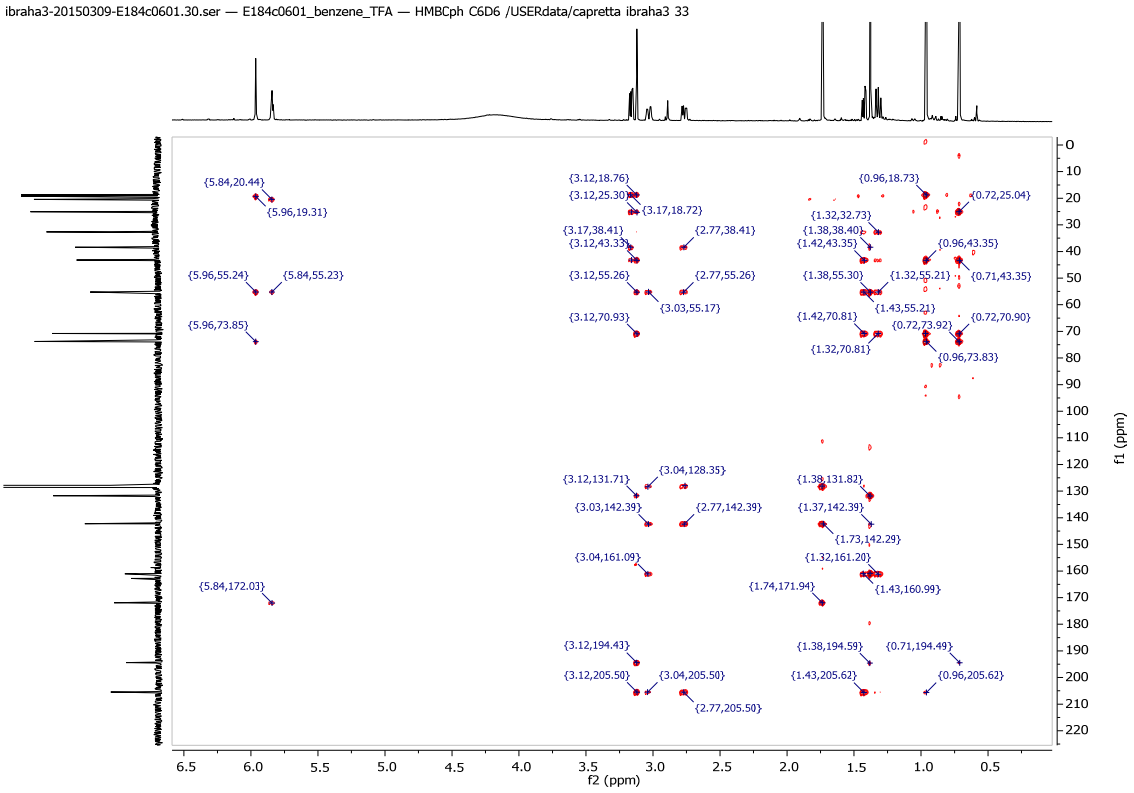


Figure S30 B. HMBC NMR spectrum (C_6D_6) of compound **4.**

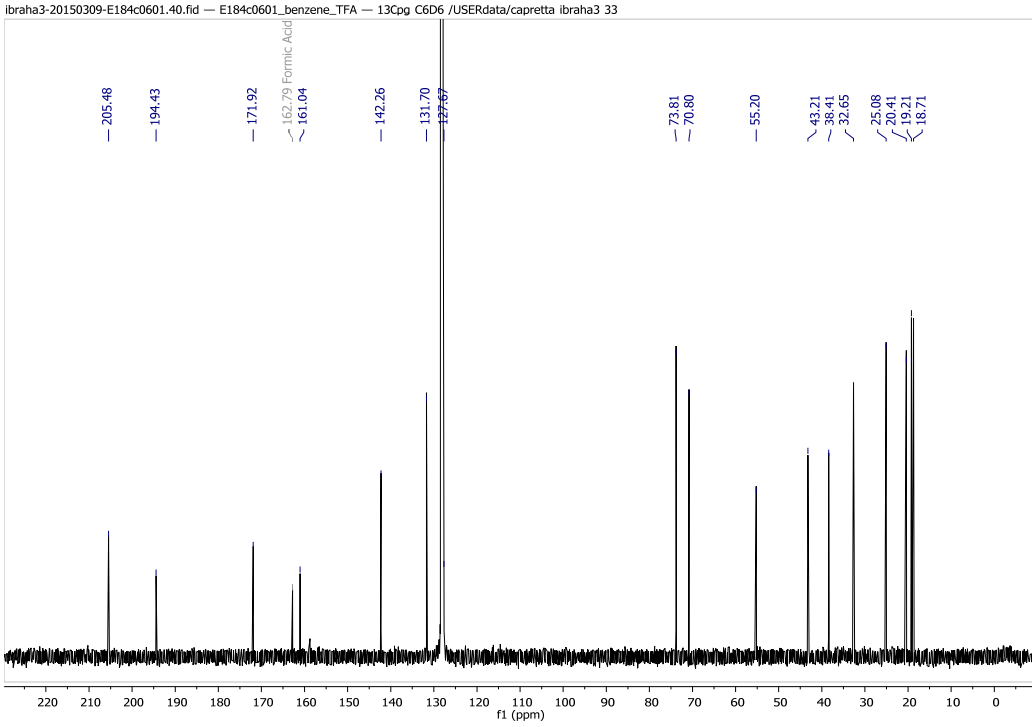


Figure S31 B. ^{13}C DEPTq NMR spectrum (176 MHz, C_6D_6) of compound **4.**

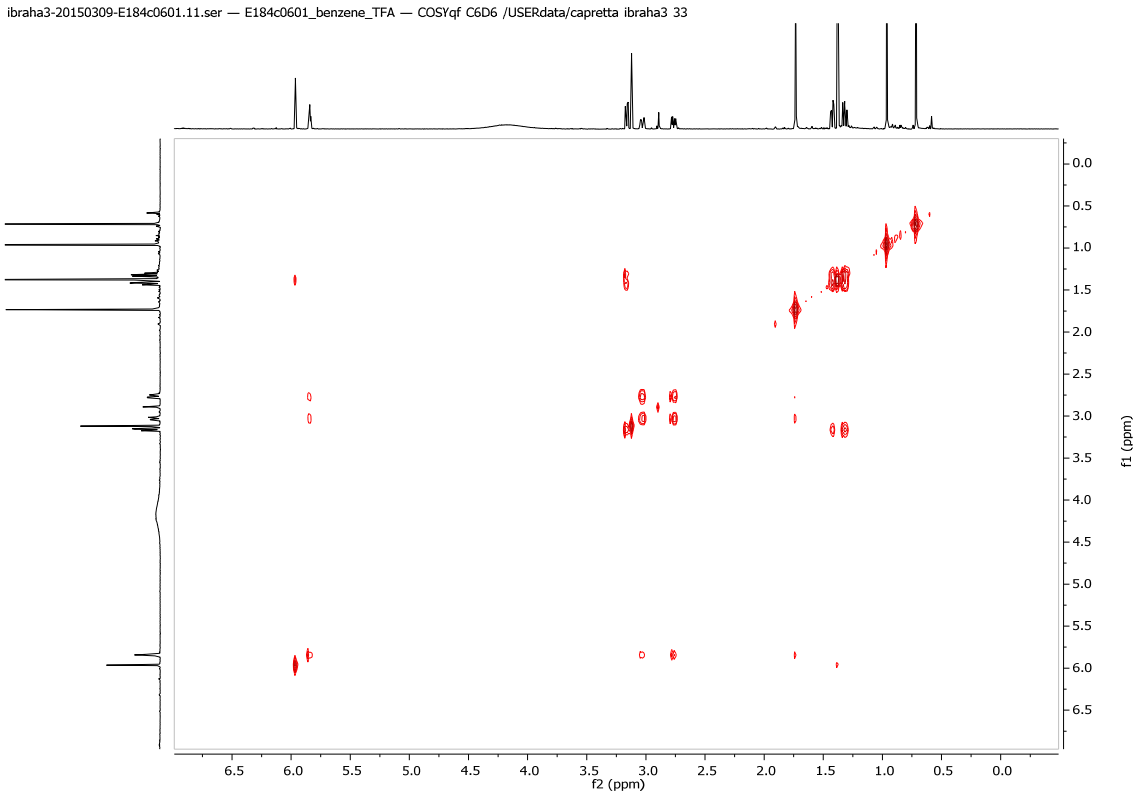


Figure S32 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **4.**

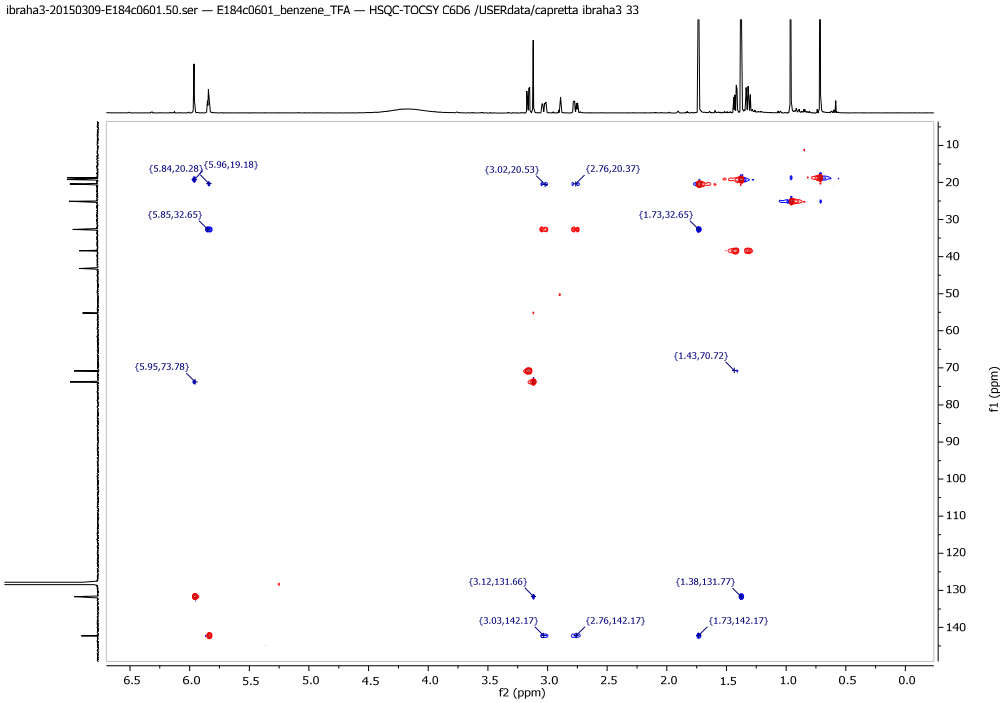


Figure S33 B. HSQC-TOCSY NMR spectrum (C_6D_6) of compound **4.**

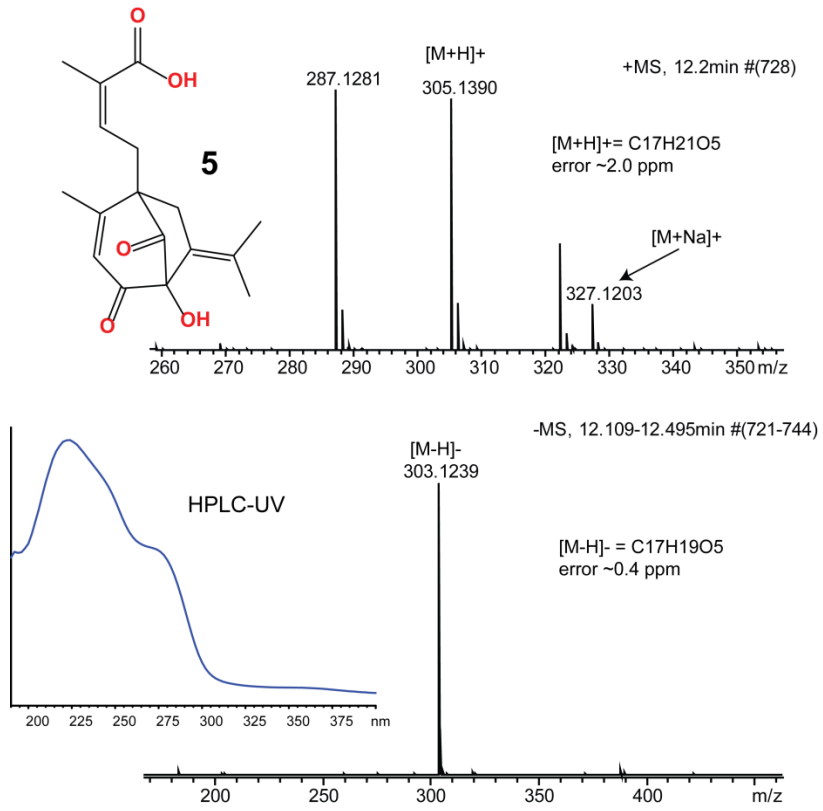


Figure S34 B. LC-HRMS and LC-UV spectra of compound **5.**

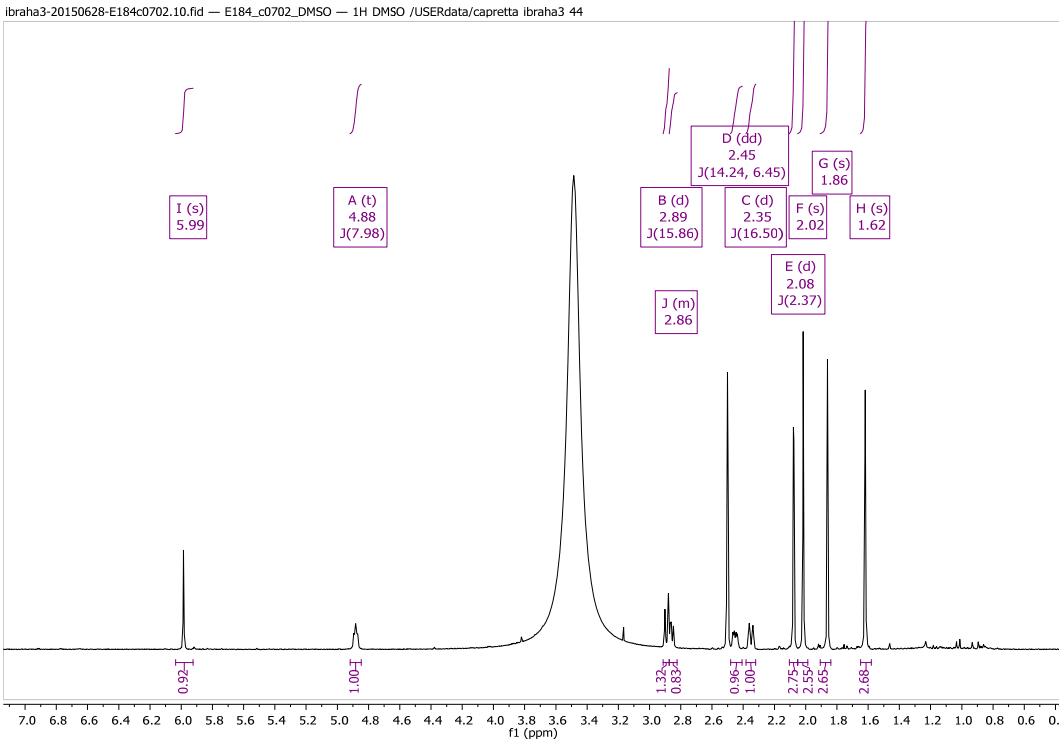


Figure S35 B. ^{1}H NMR spectrum (700 MHz, DMSO-d_6) of compound **5.**

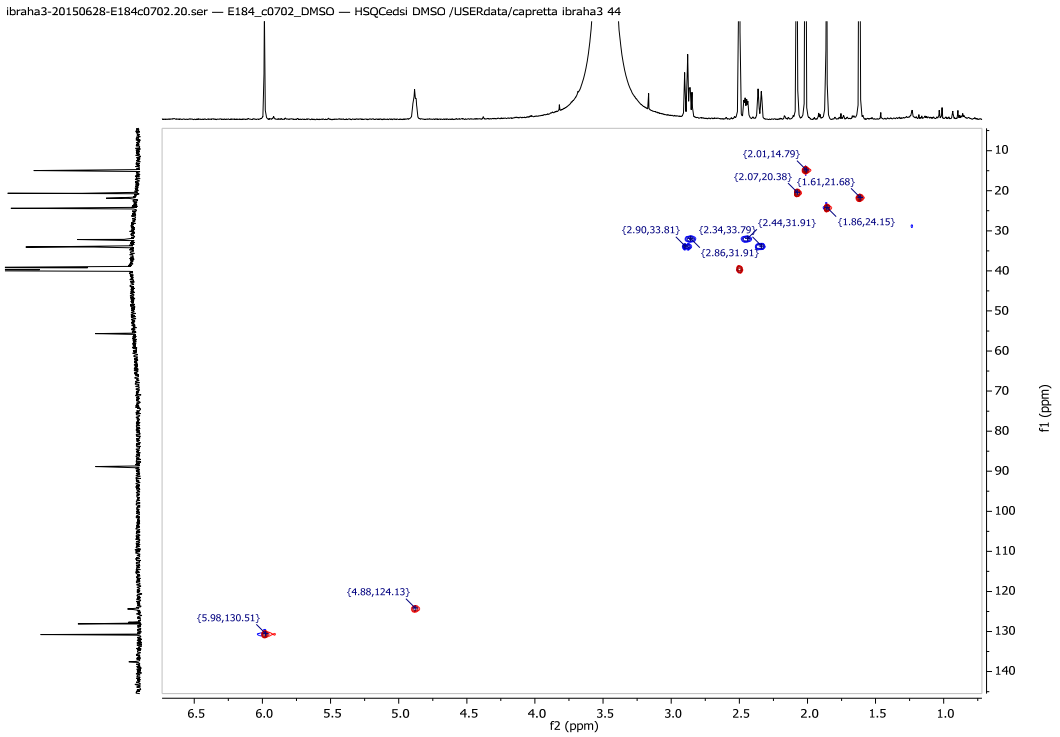


Figure S36 B. Multiplicity-edited HSQC NMR spectrum (DMSO-d_6) of compound **5.**

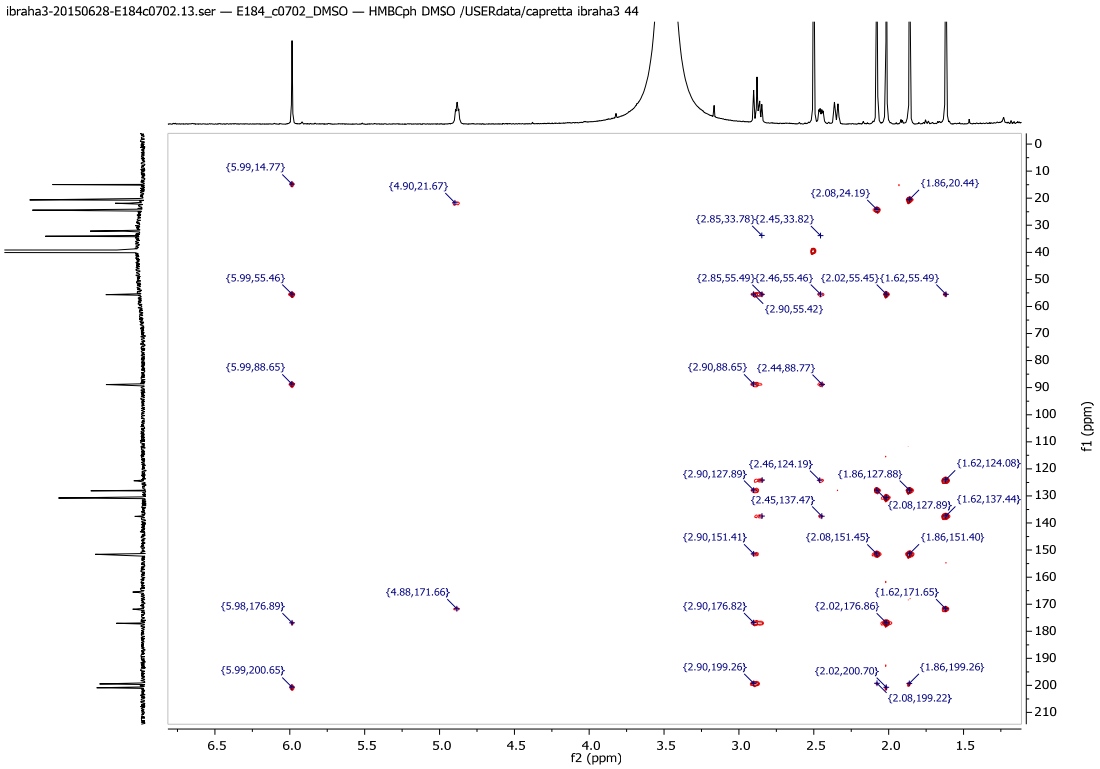


Figure S37 B. HMBC NMR spectrum (DMSO-d_6) of compound **5.**

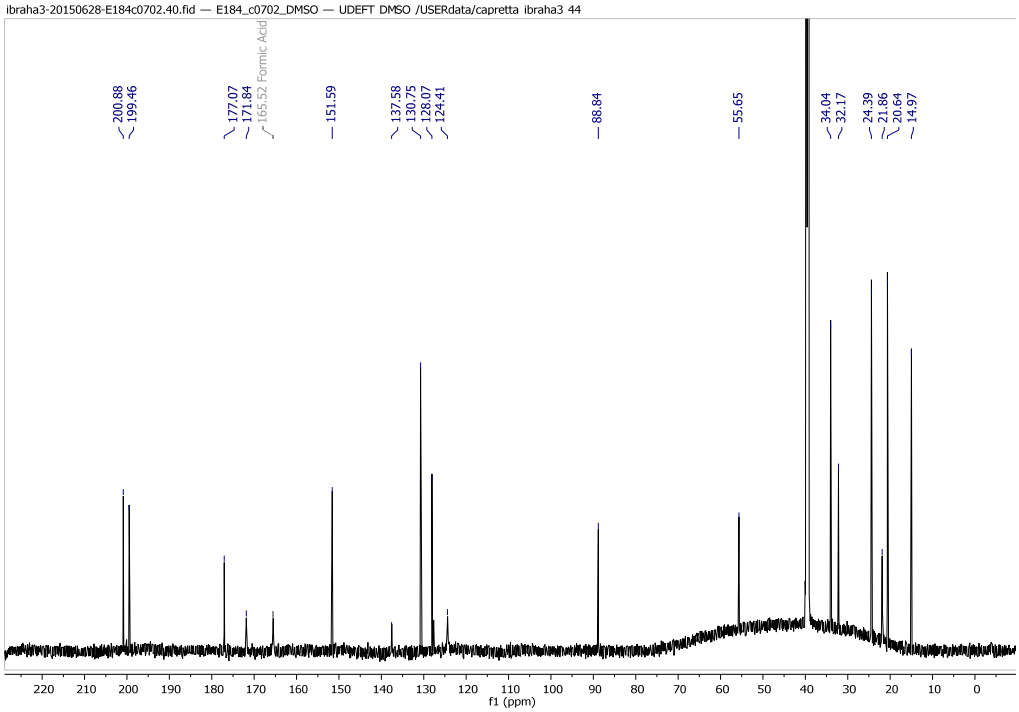


Figure S38 B. ^{13}C NMR spectrum (176 MHz, DMSO-d_6) of compound **5.**

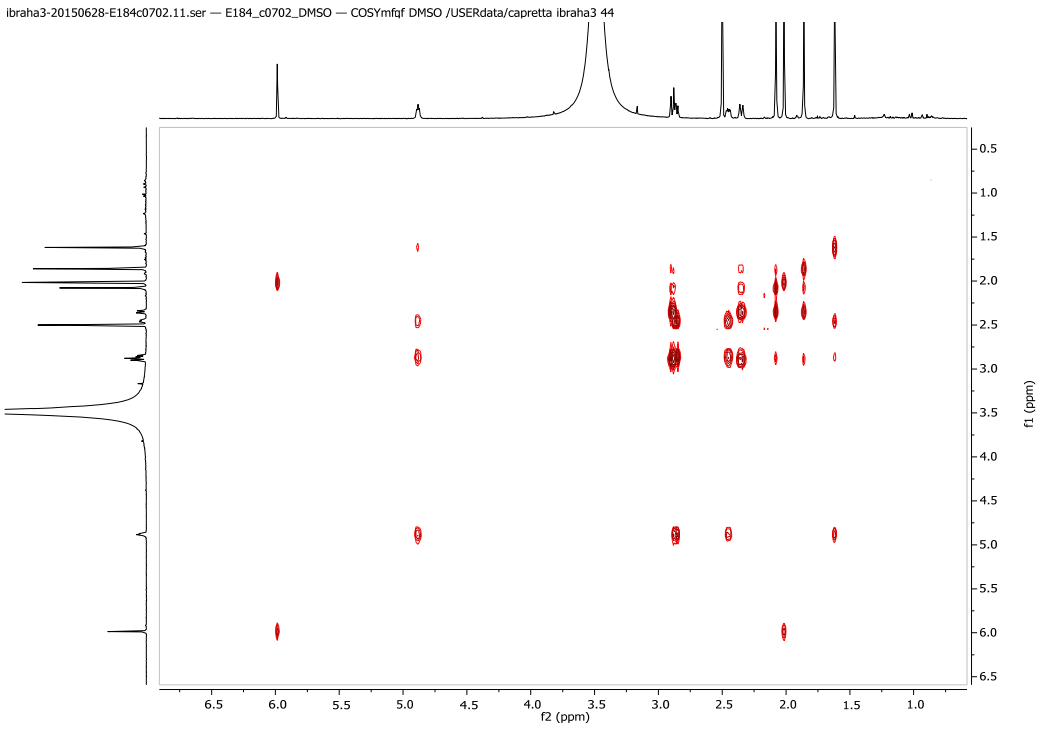


Figure S39 B. COSY NMR spectrum (700 MHz, DMSO-d_6) of compound **5.**

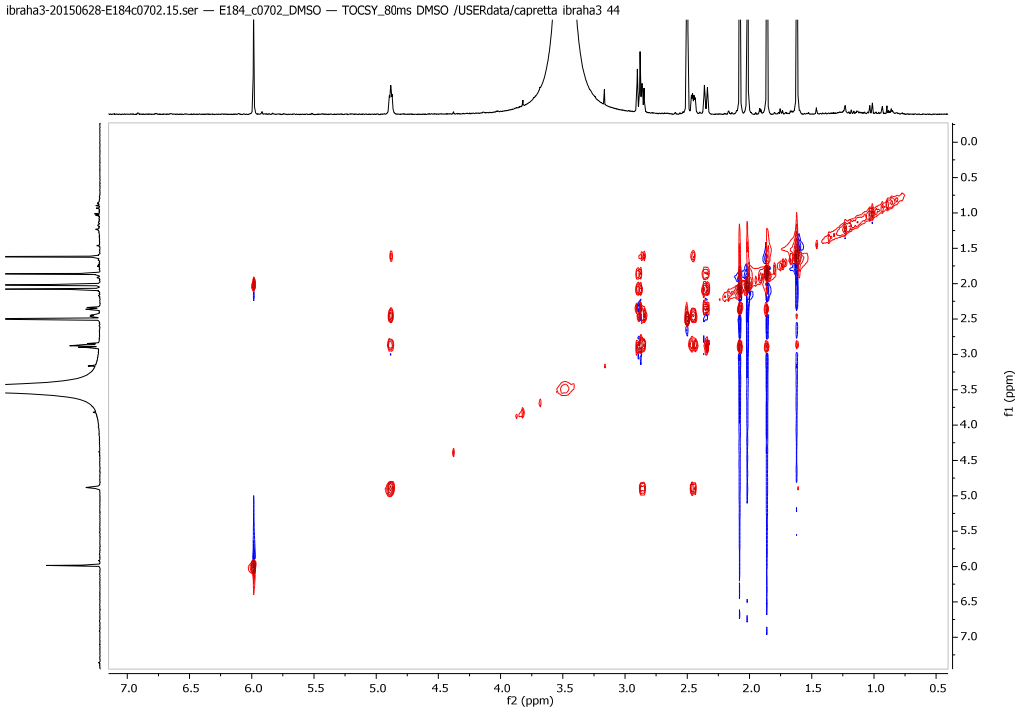


Figure S40 B. TOCSY NMR spectrum (700 MHz, DMSO-d_6) of compound **5.**

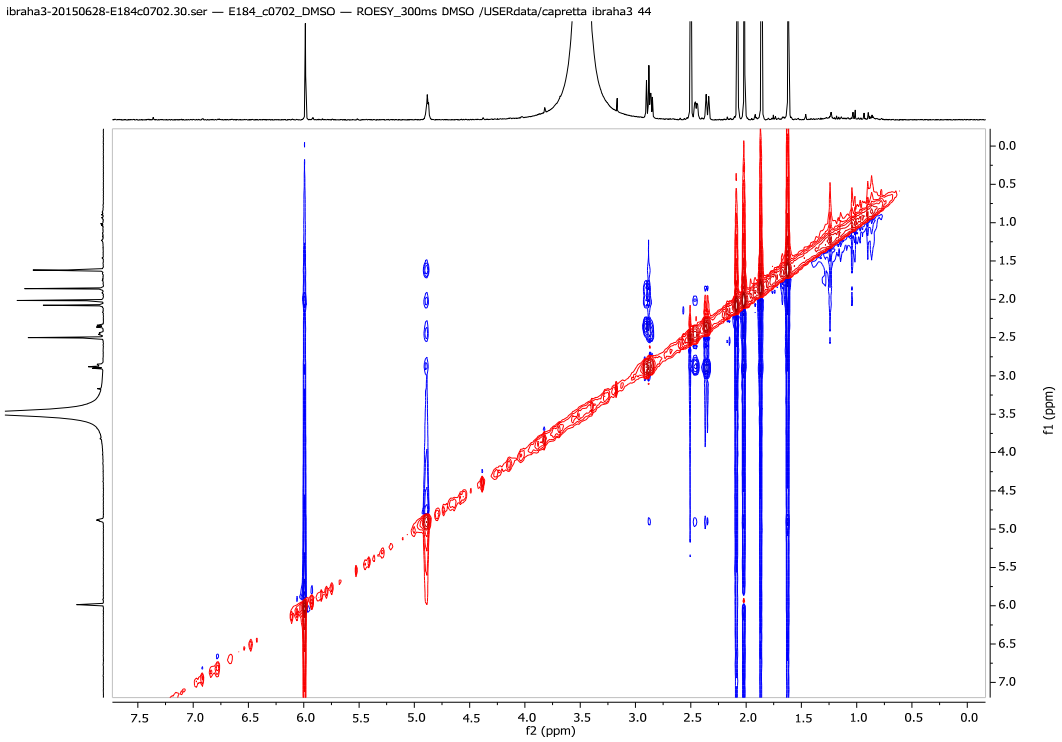


Figure S41 B. ROESY 300ms NMR spectrum (700 MHz, DMSO-d_6) of compound **5.**

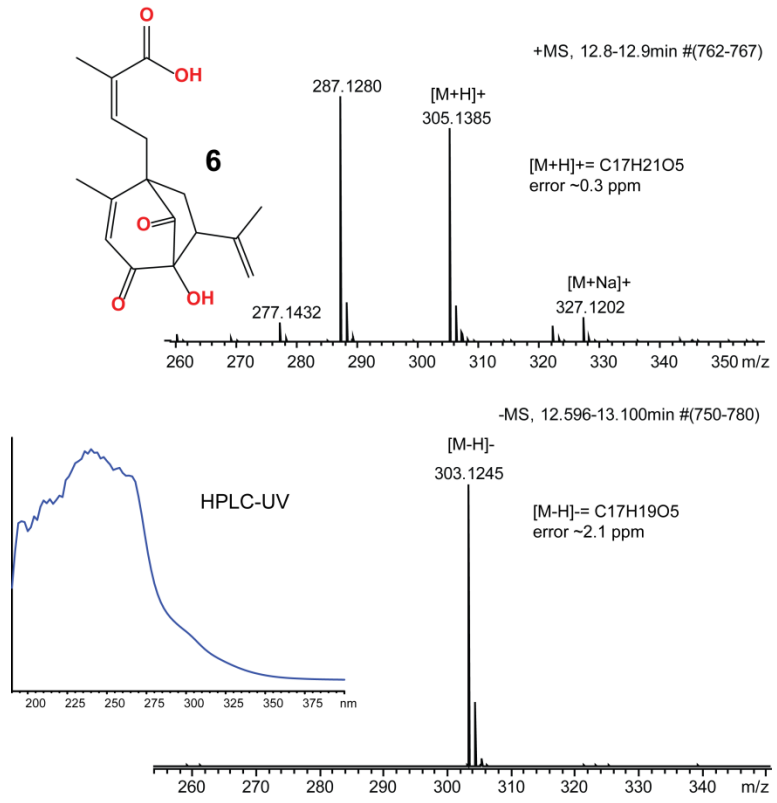


Figure S42 B. LC-HRMS and LC-UV spectra of compound **6.**

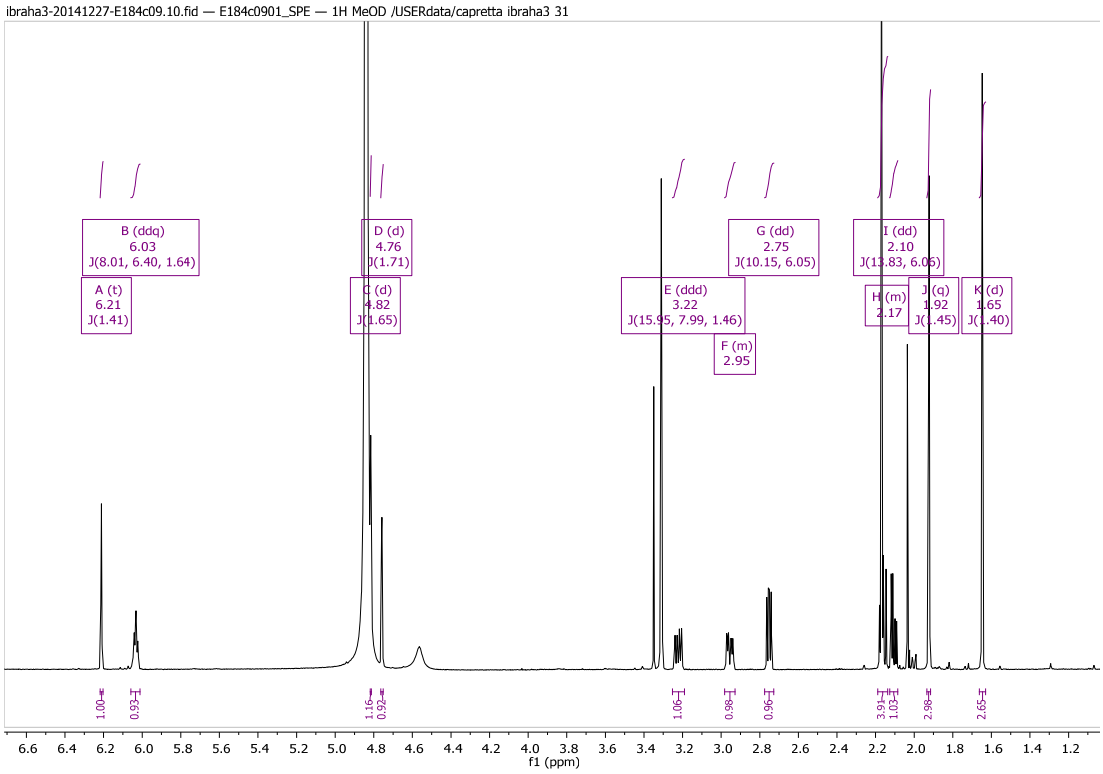


Figure S43 B. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **6.**

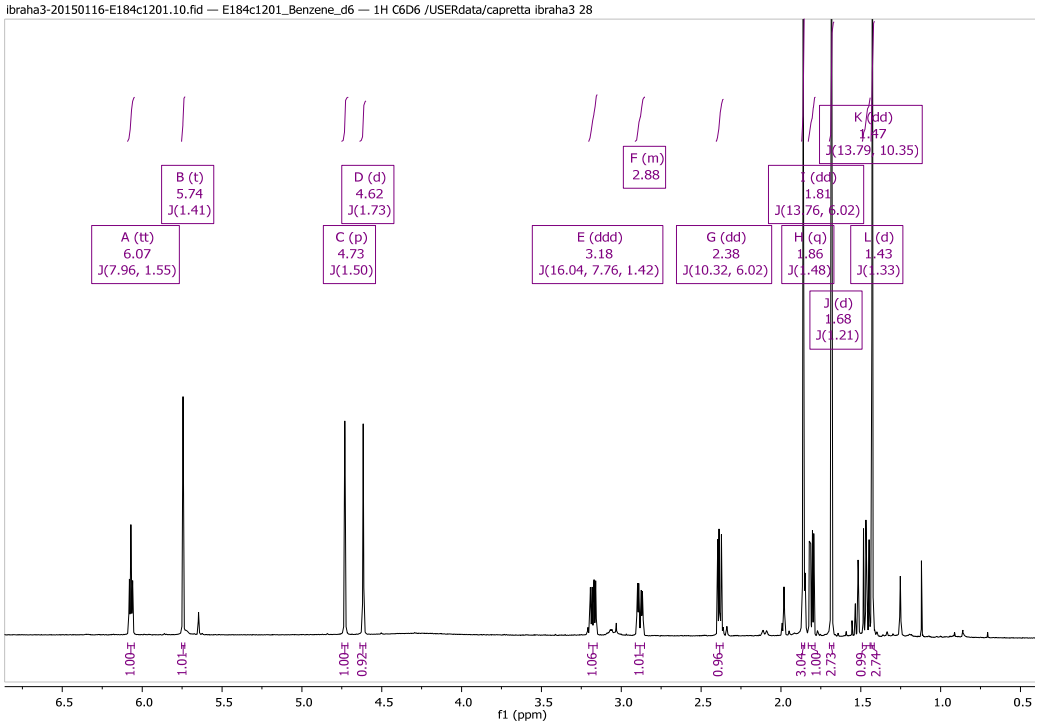


Figure S44 B. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **6.**

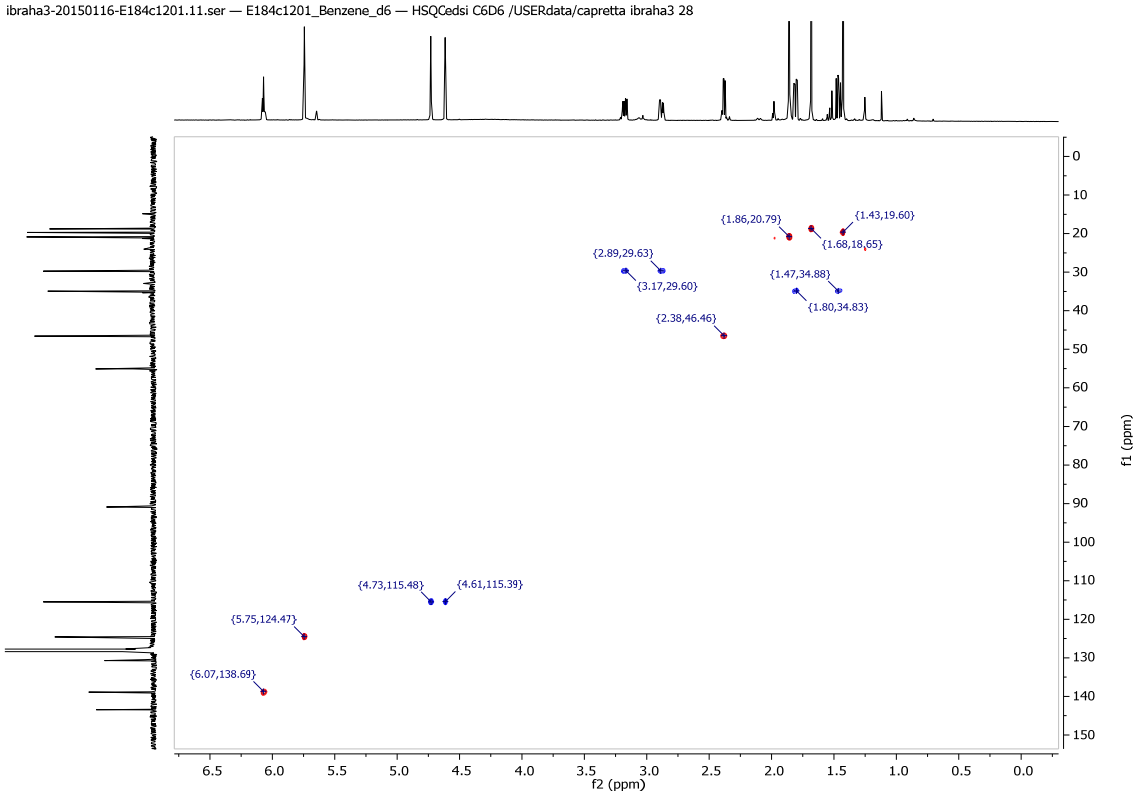


Figure S45 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **6.**

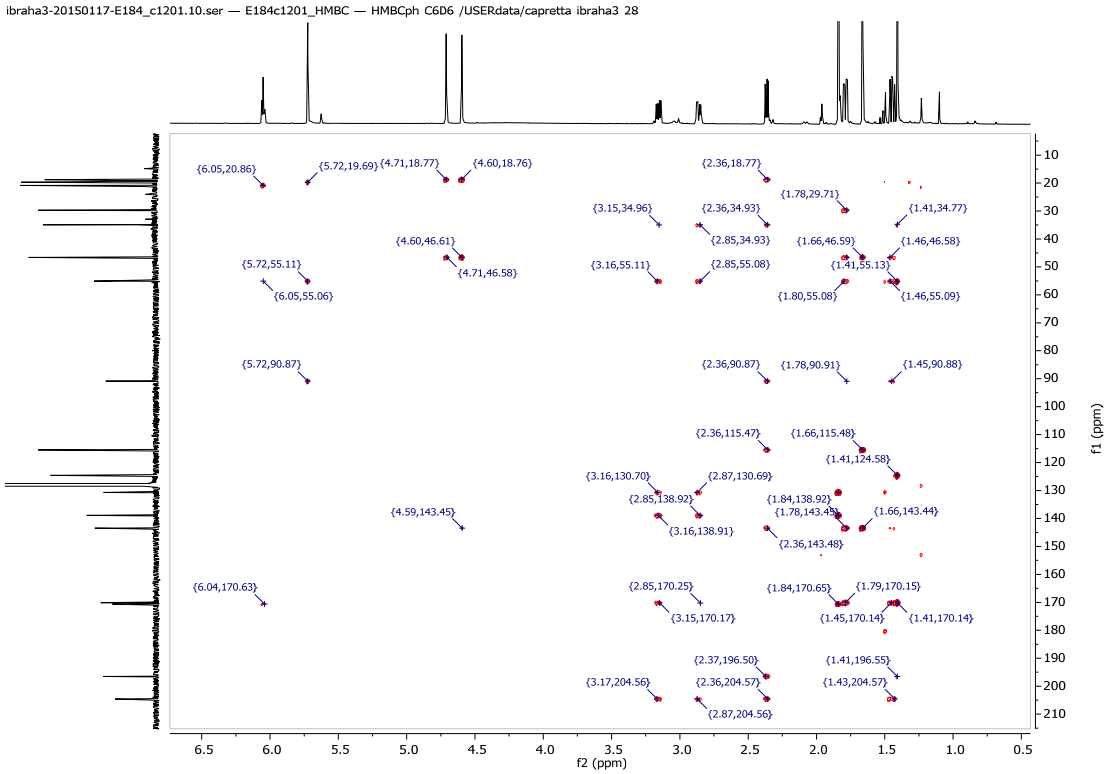


Figure S46 B. HMBC NMR spectrum (C_6D_6) of compound **6.**

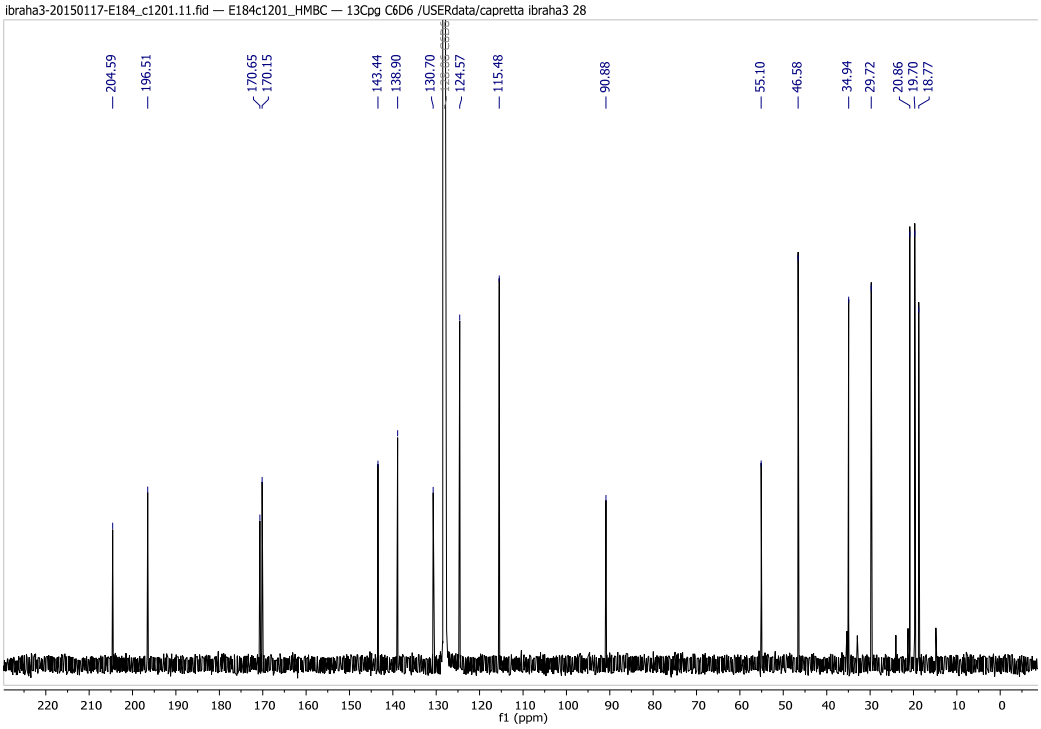


Figure S47 B. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **6.**

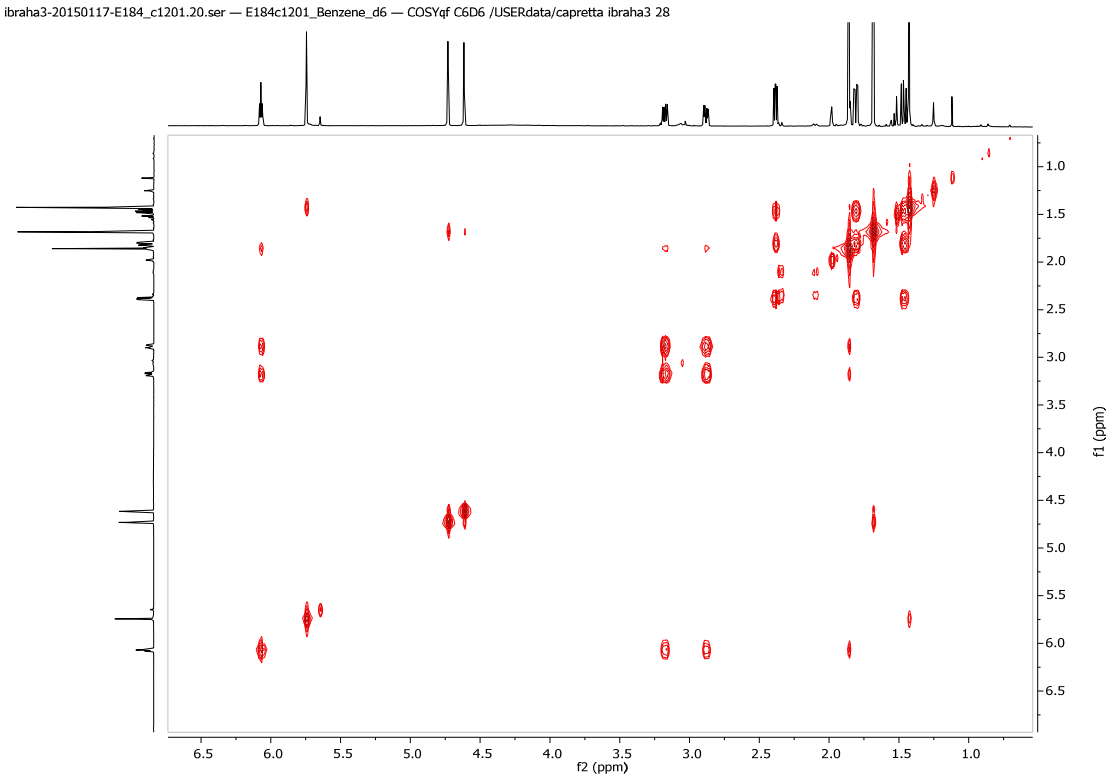


Figure S48 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **6.**

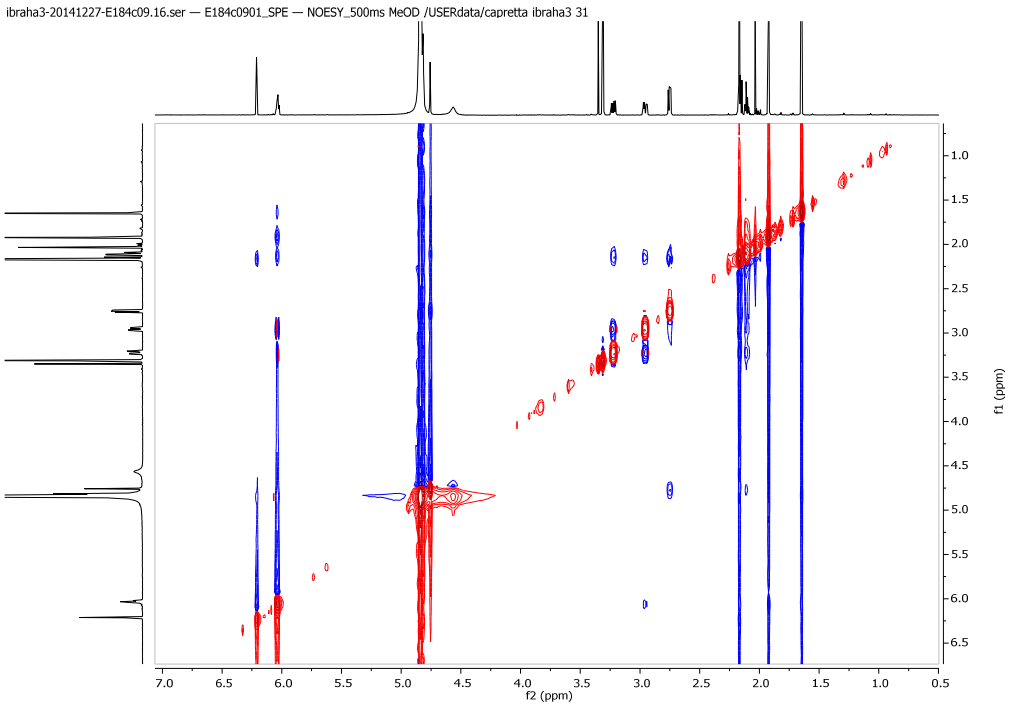


Figure S49 B. NOESY 500ms NMR spectrum (CD_3OD) of compound **6.**

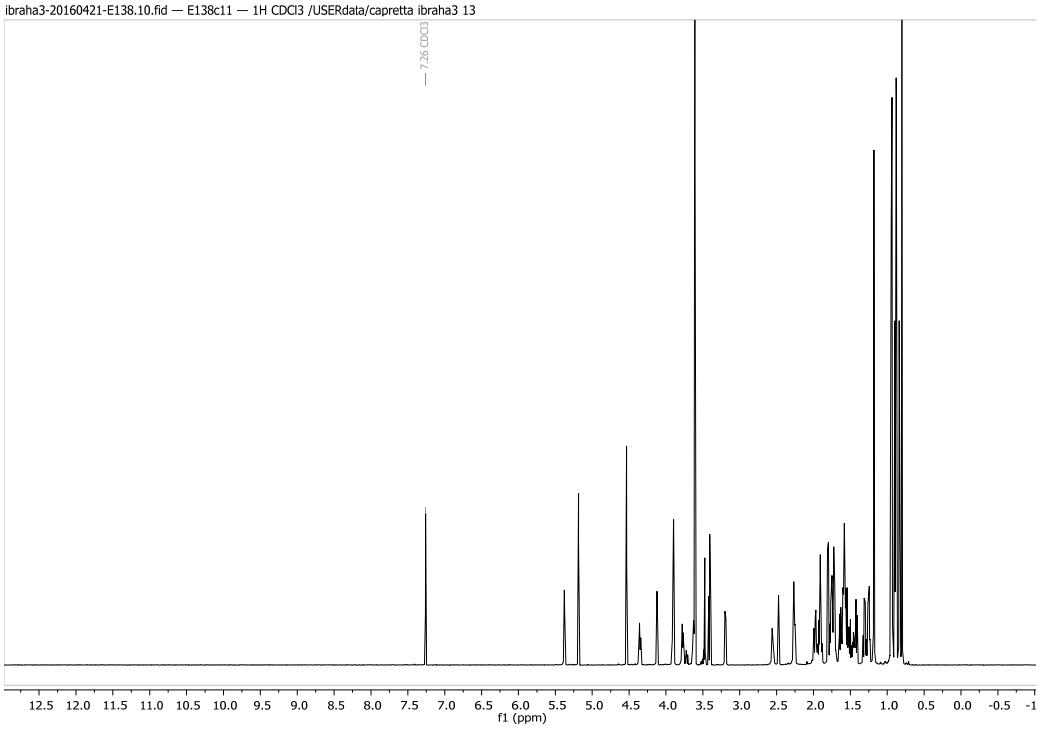


Figure S50. ^{1}H NMR spectrum (700 MHz, $CDCl_3$) of compound **7.**

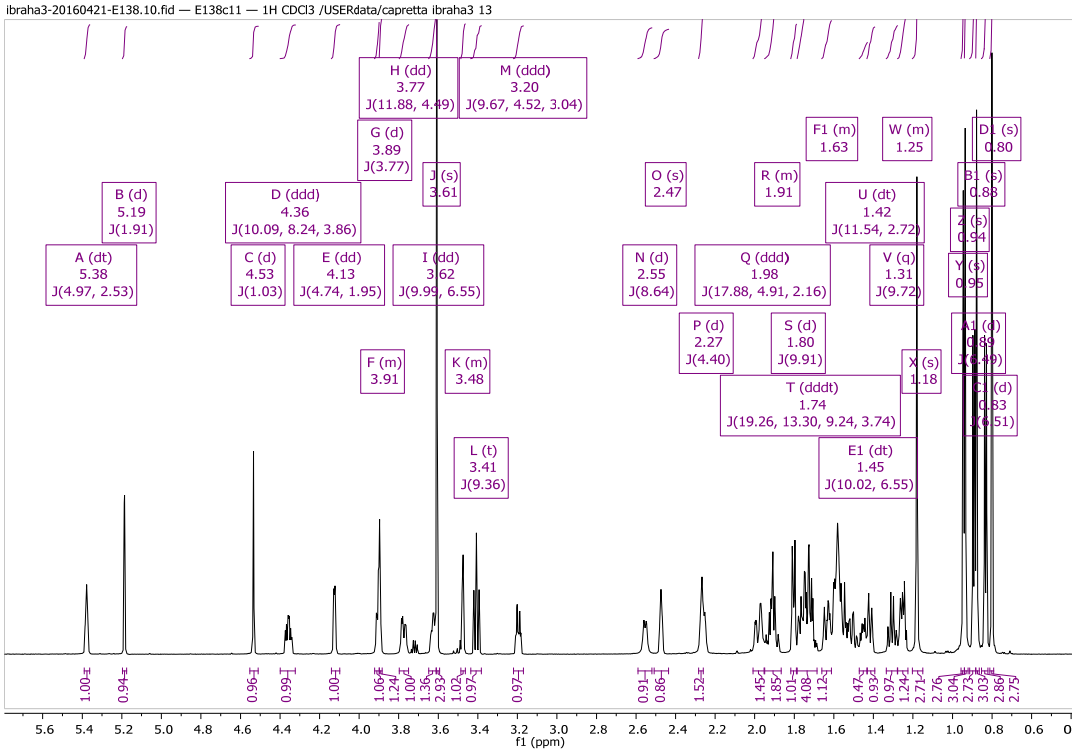


Figure S51 B. ^{1}H NMR spectrum from 0.4-5.8 ppm (700 MHz, $CDCl_3$) of compound **7.**

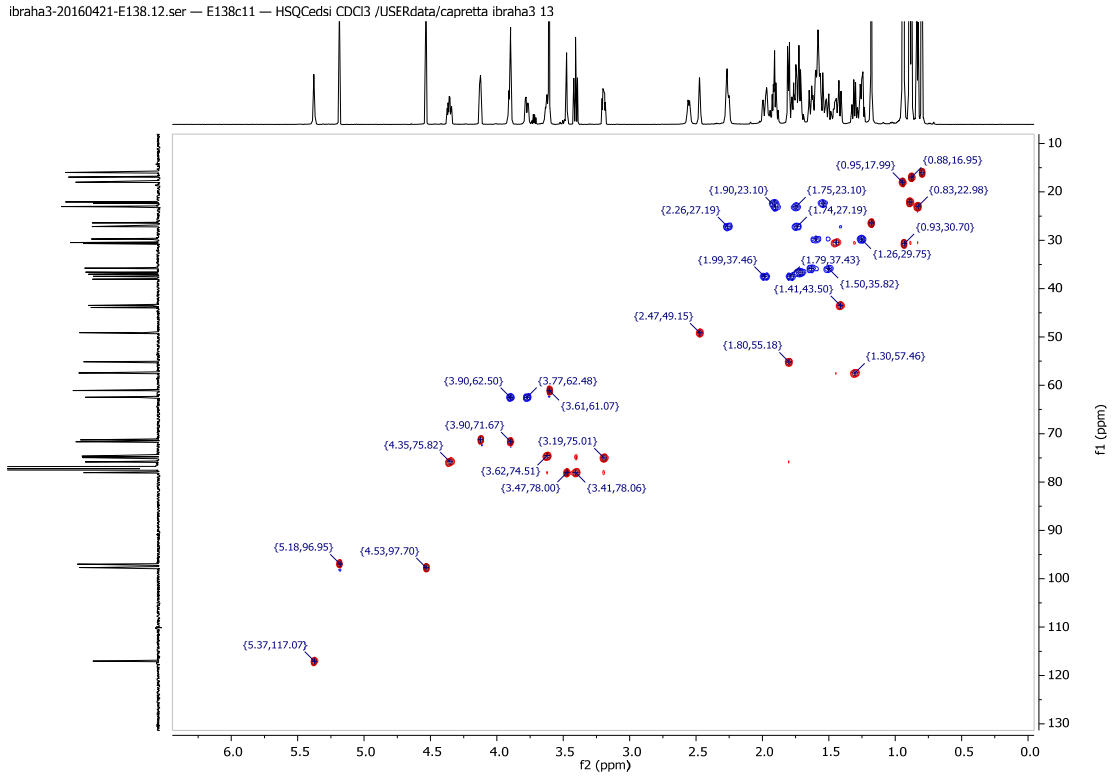


Figure S52 B. Multiplicity-edited HSQC NMR spectrum ($CDCl_3$) of compound **7.**

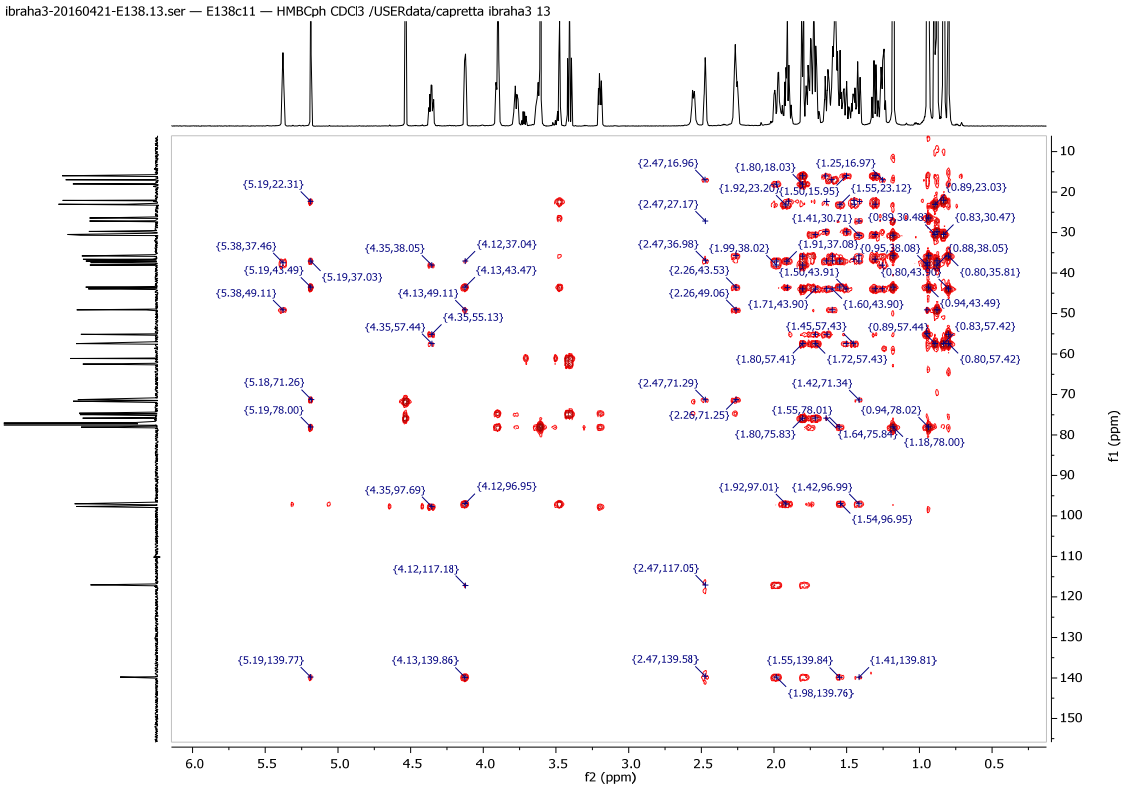


Figure S53 B. HMBC NMR spectrum ($CDCl_3$) of compound **7.**

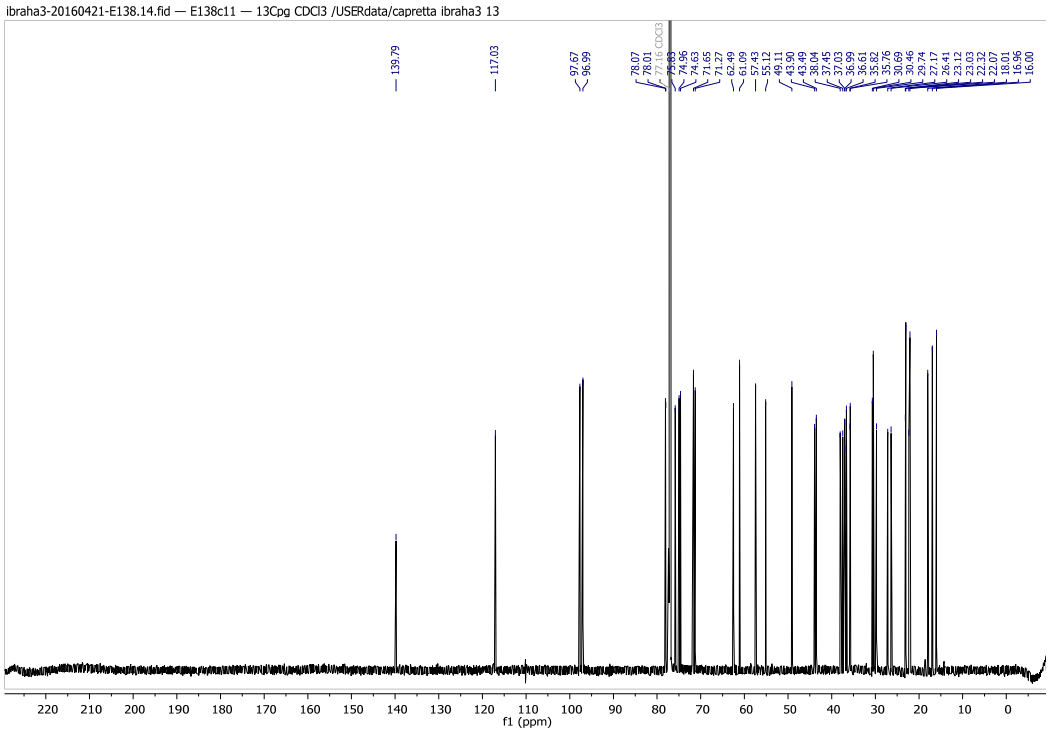


Figure S54 B. ^{13}C NMR spectrum (176 MHz, $CDCl_3$) of compound **7.**

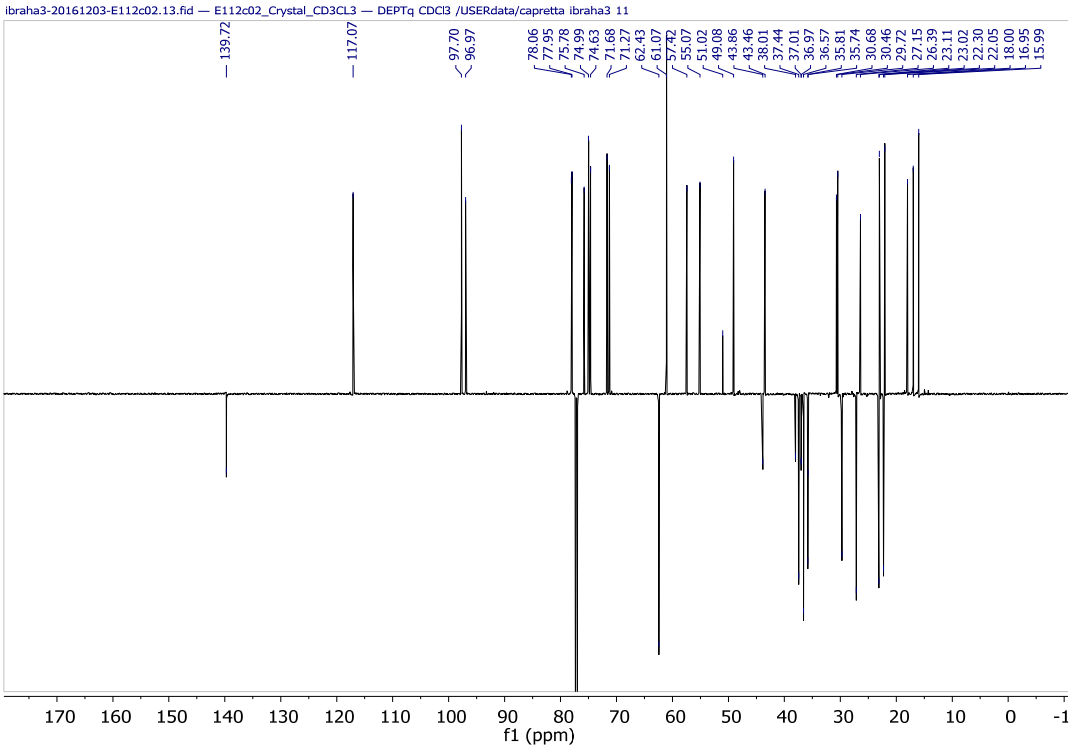


Figure S55 B. ^{13}C DEPTq NMR spectrum (176 MHz, $CDCl_3$) of compound **7.**

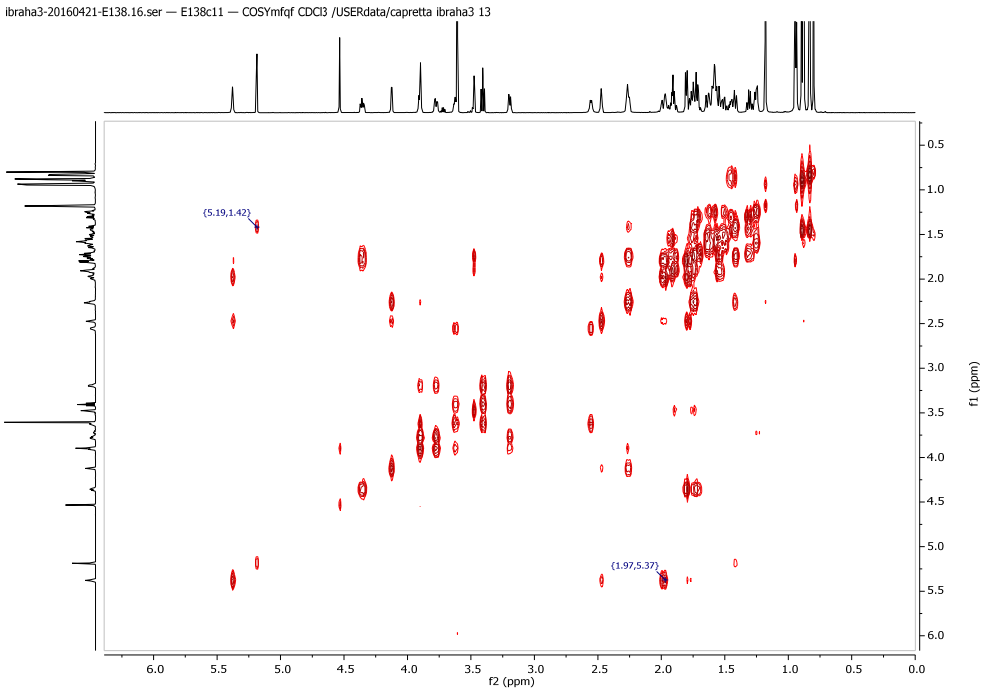


Figure S56 B. COSY NMR spectrum (700 MHz, $CDCl_3$) of compound **7.**

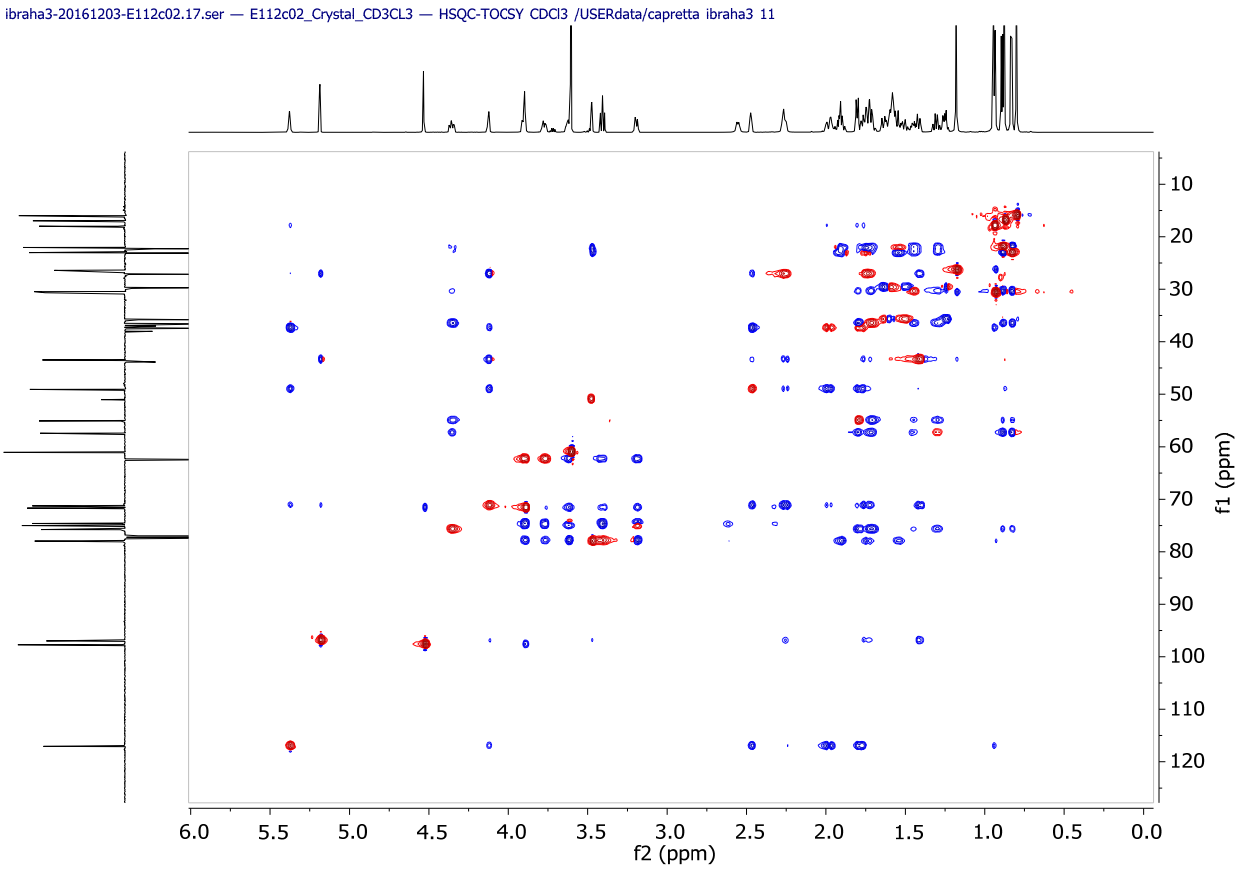


Figure S57 B. HSQC-TOCSY NMR spectrum (700 MHz, $CDCl_3$) of compound **7.**

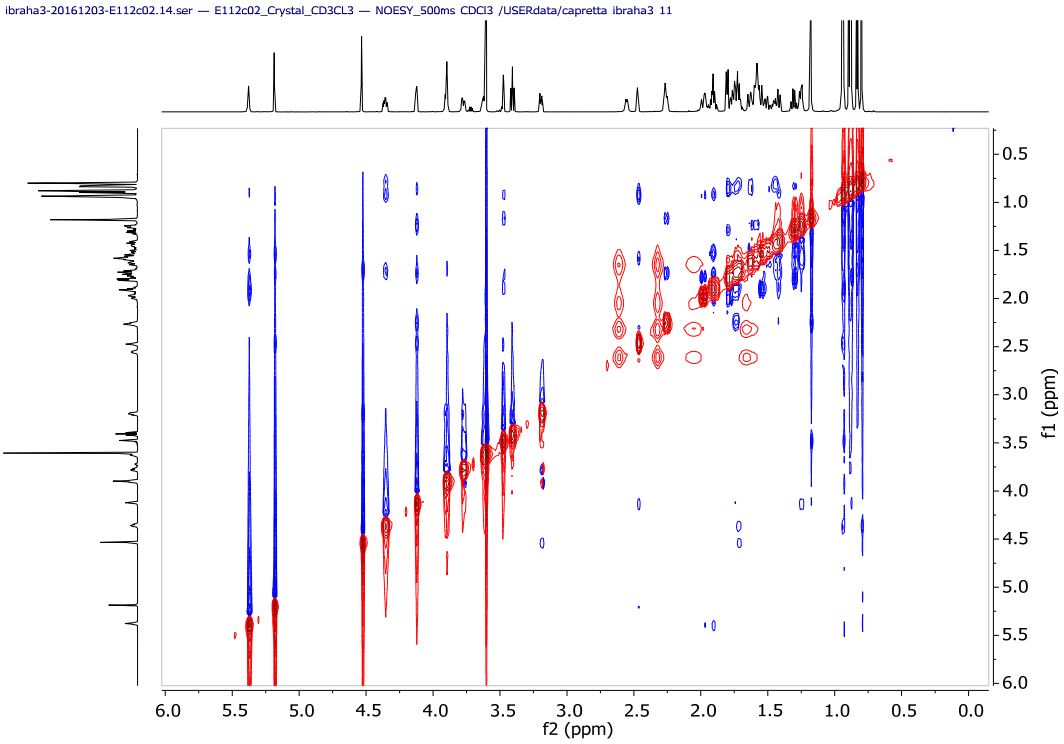


Figure S58 B. NOESY 500ms NMR spectrum (700 MHz, $CDCl_3$) of compound **7.**

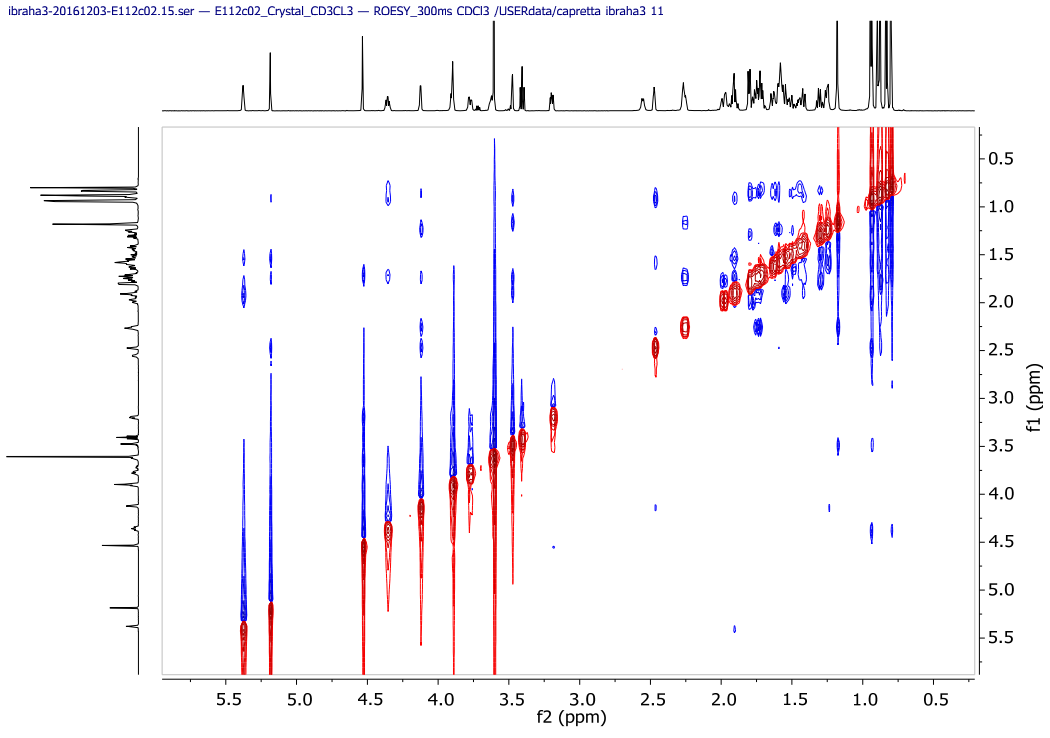


Figure S59 B. ROESY 300ms NMR spectrum (700 MHz, $CDCl_3$) of compound **7.**

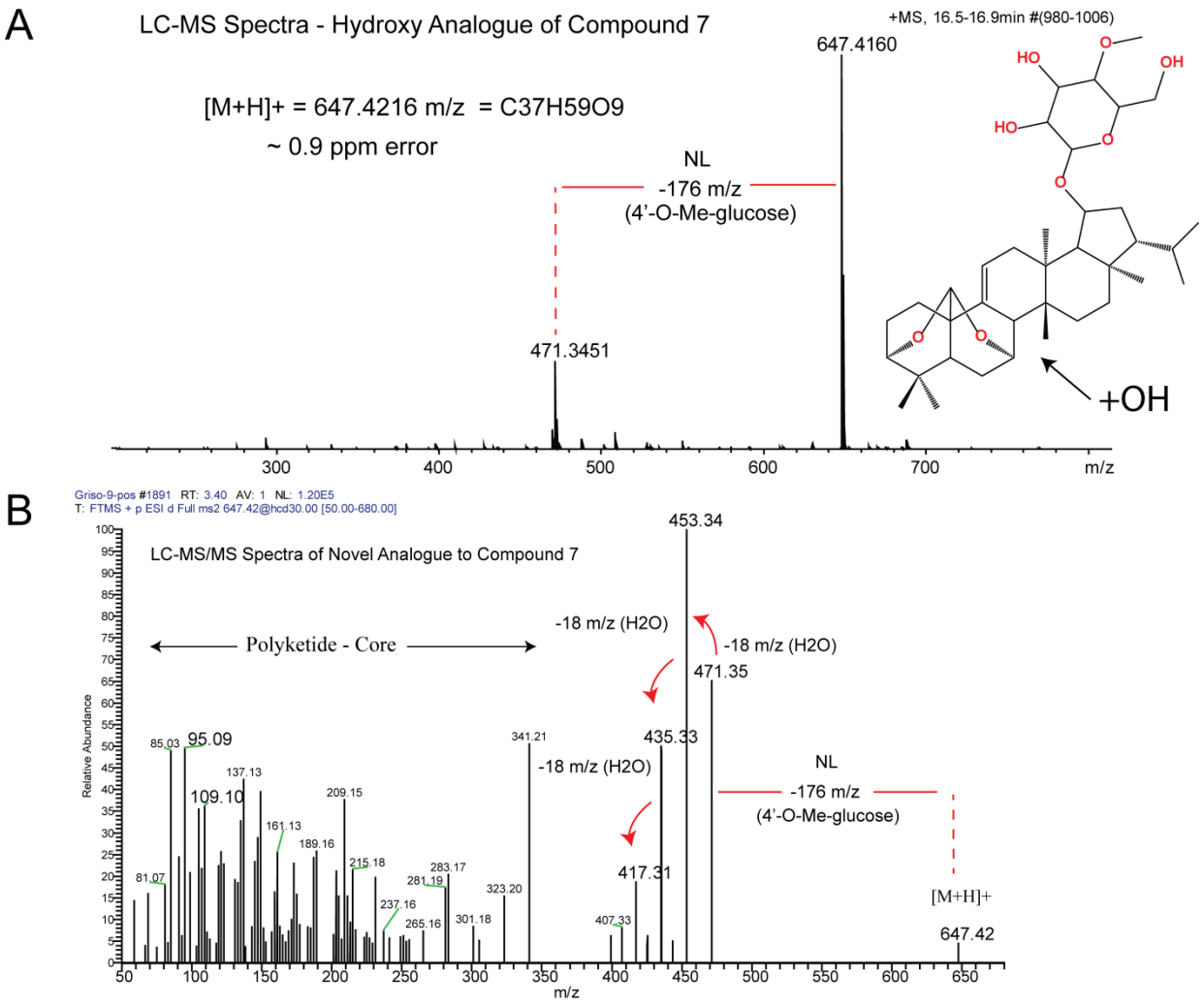


Figure S60. (A) LC-HRMS and (B) MS/MS spectra of compound 7 B, a new hydroxyl analogue of compound 7.

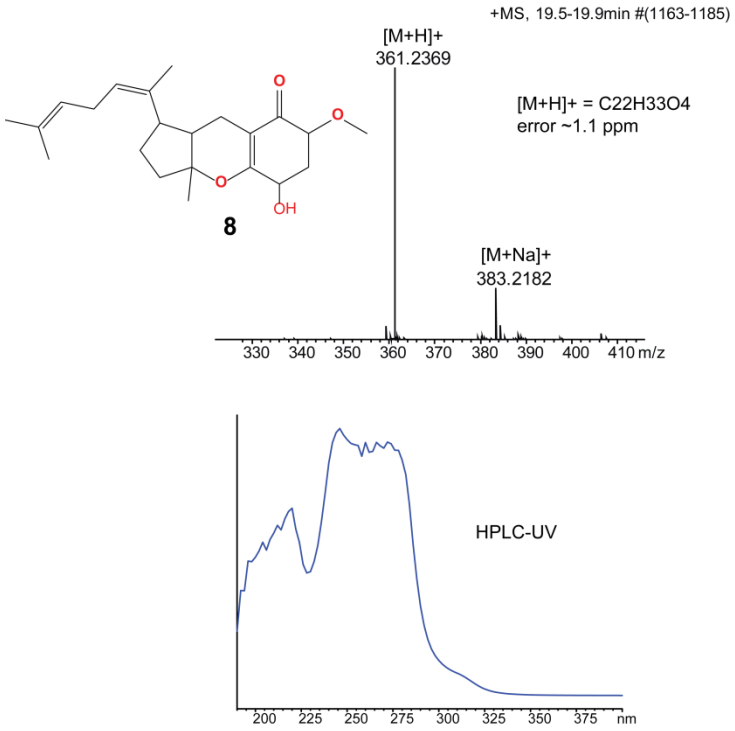


Figure S61 B. LC-HRMS and LC-UV spectra of compound **8.**

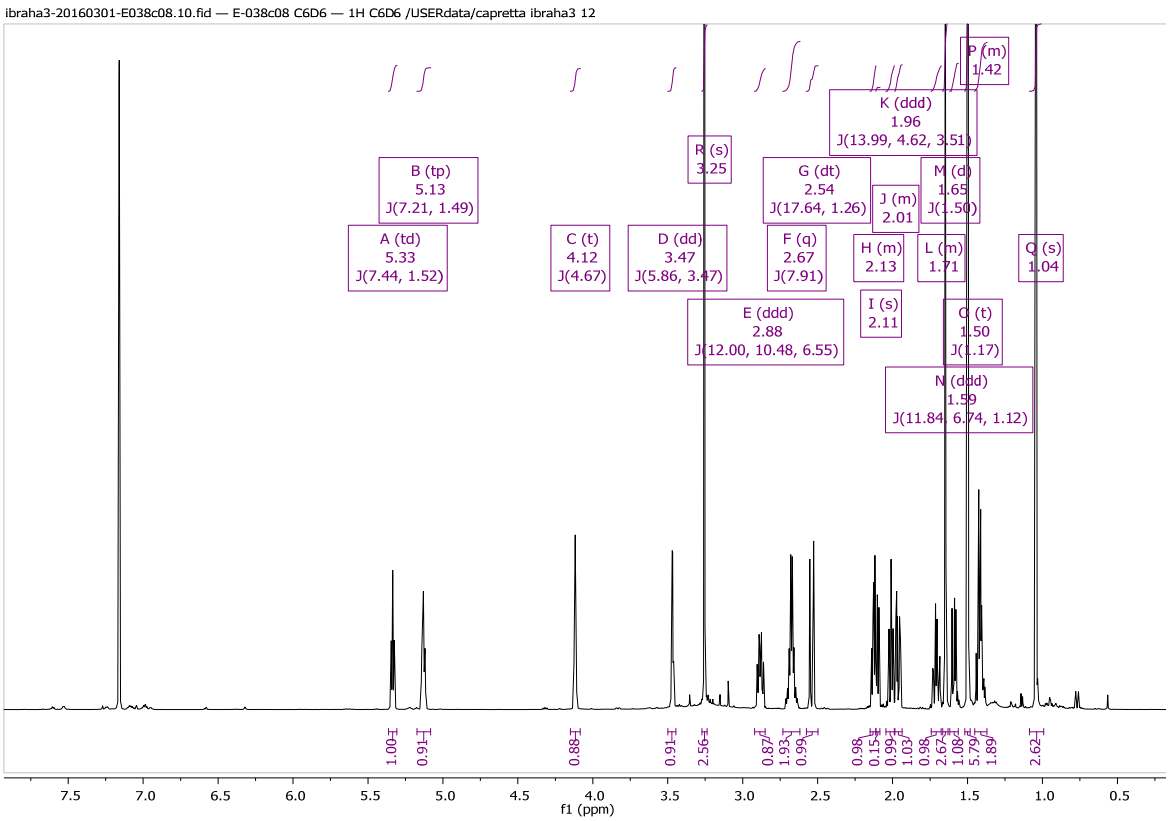


Figure S62 B. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **8.**

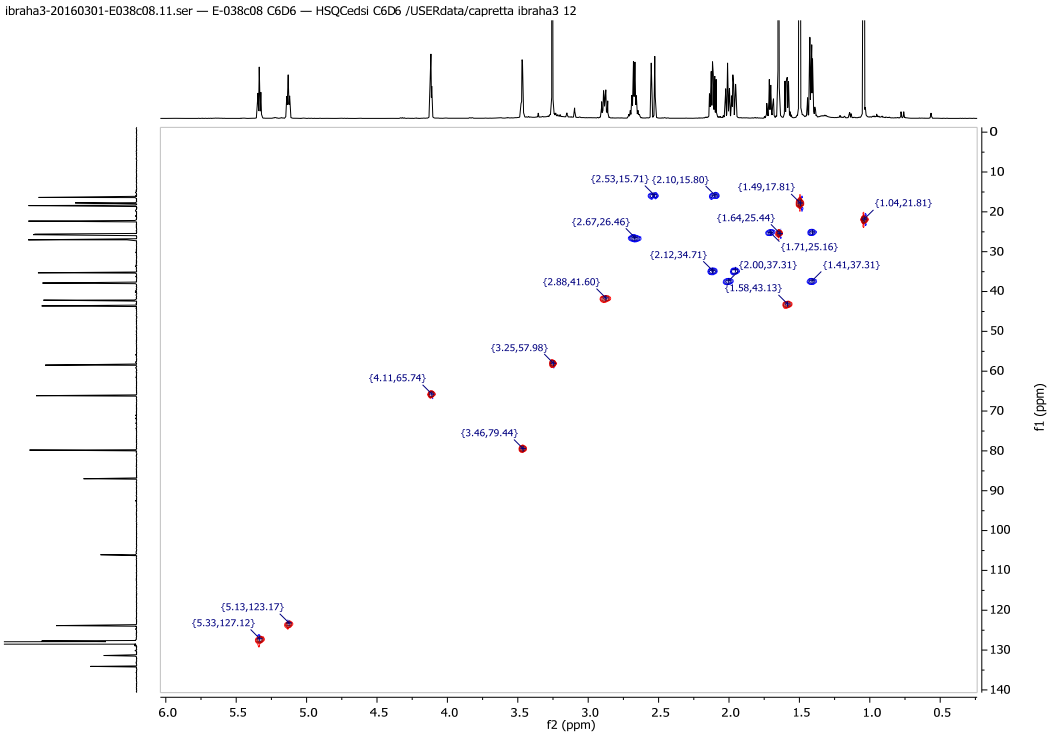


Figure S63 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **8.**

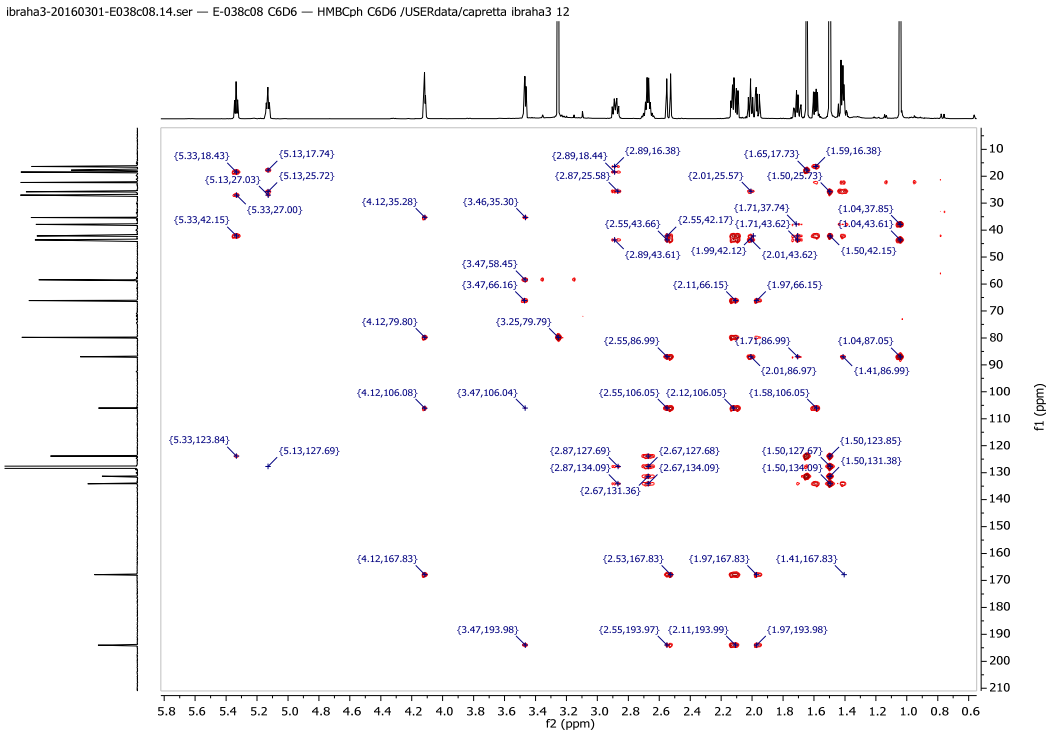


Figure S64 B. HMBC NMR spectrum (C_6D_6) of compound **8.**

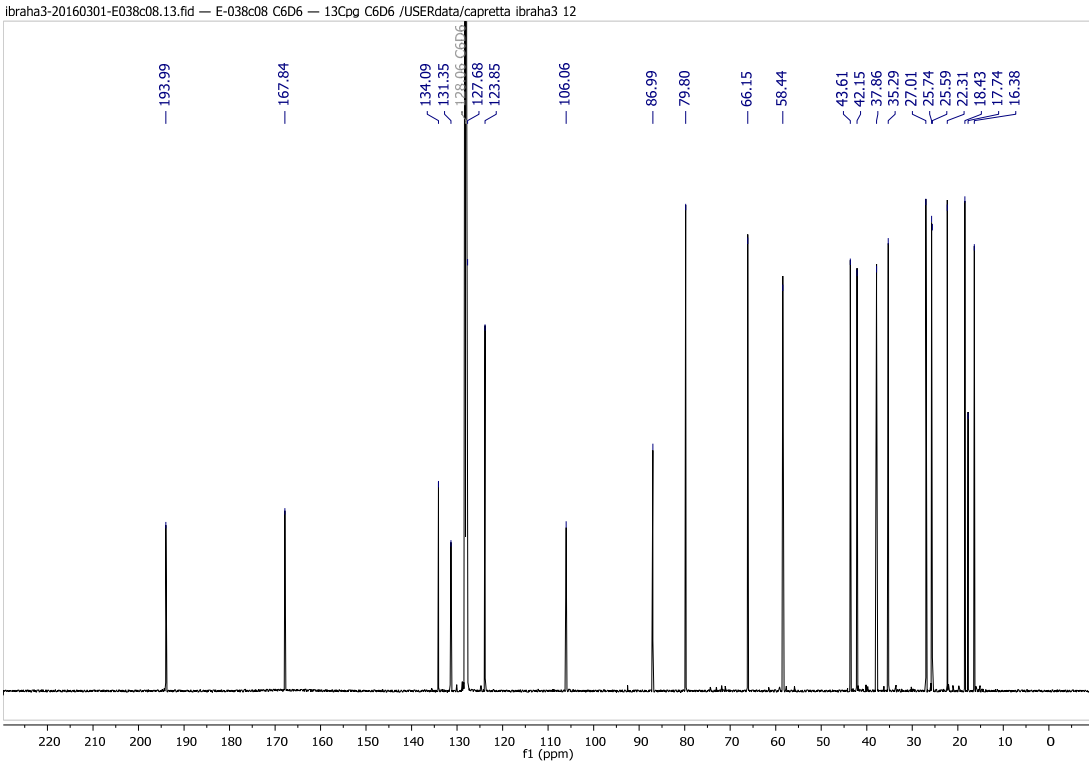


Figure S65 B. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **8.**

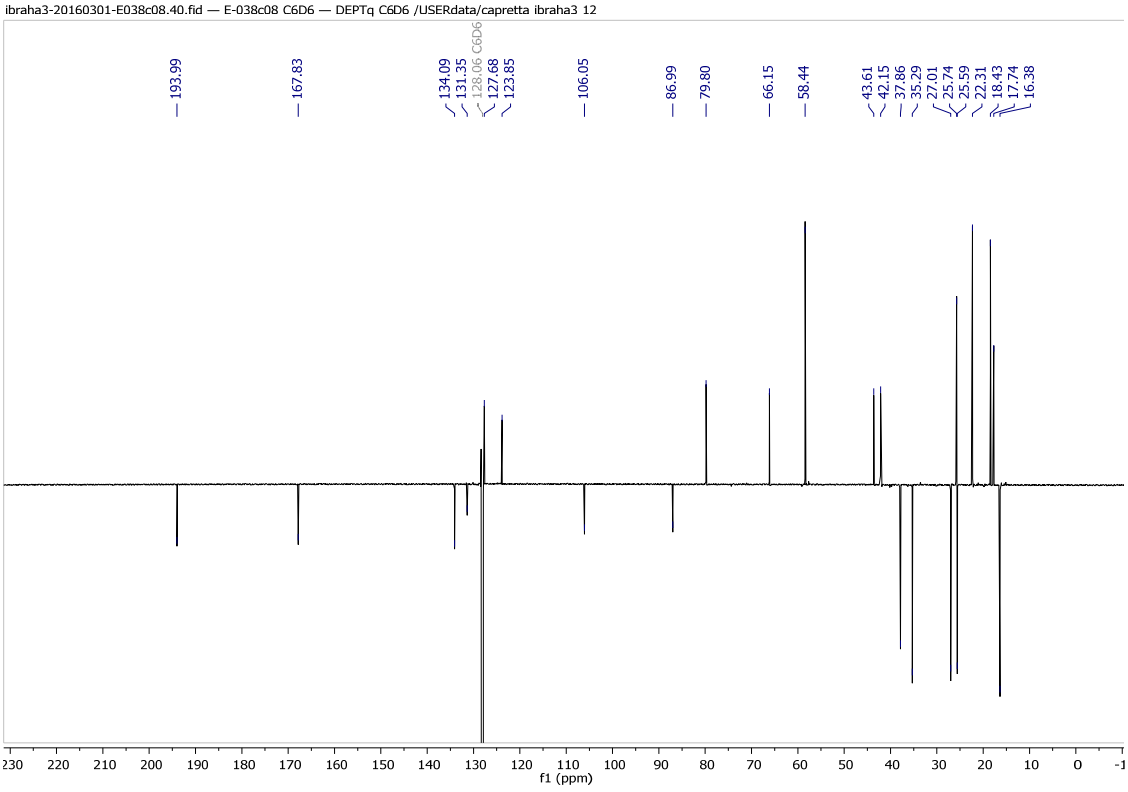


Figure S66 B. ^{13}C DEPTq NMR spectrum (176 MHz, C_6D_6) of compound **8.**

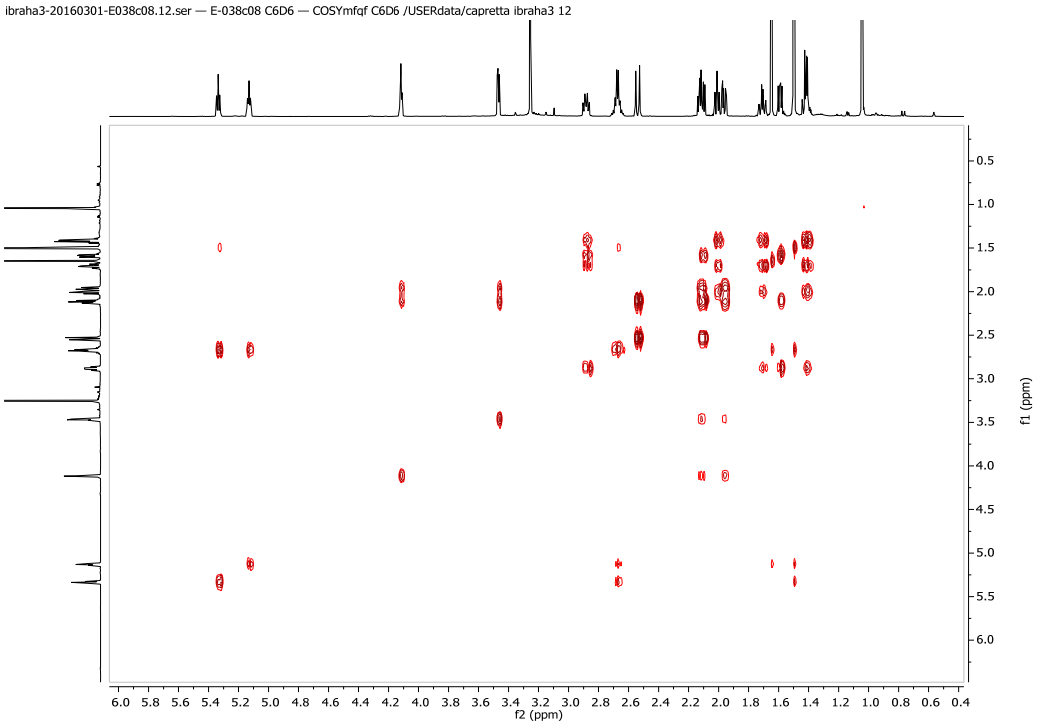


Figure S67 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **8.**

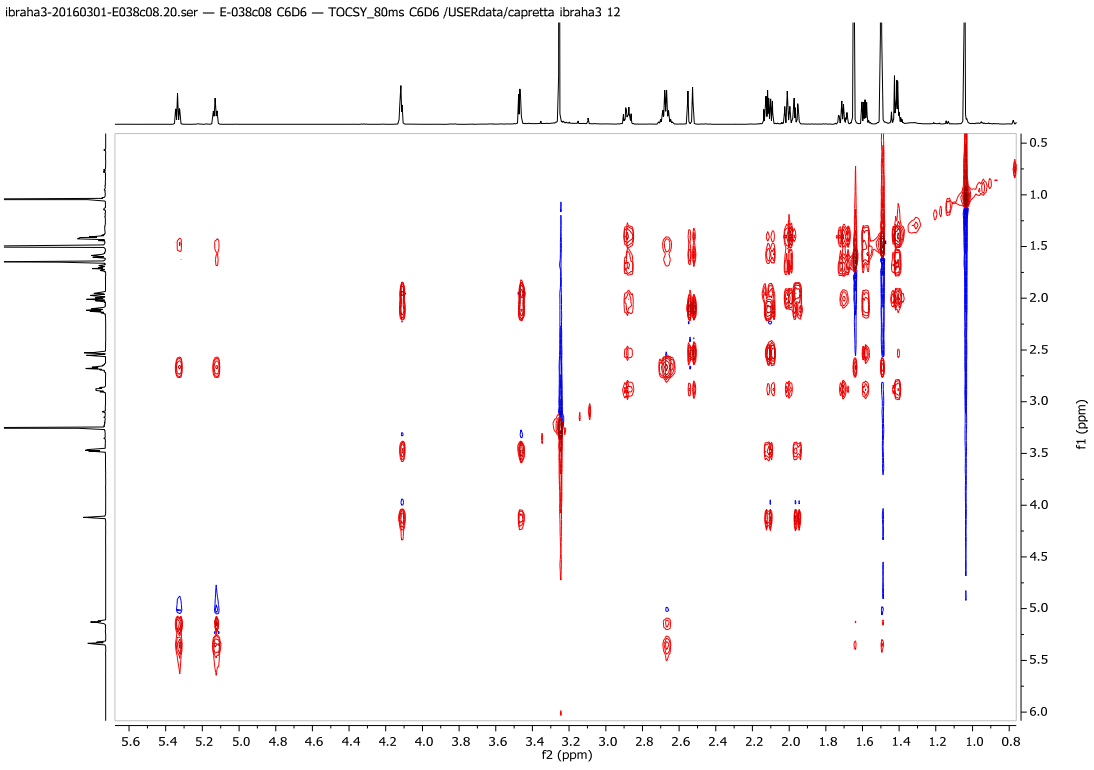


Figure S68 B. TOCSY NMR spectrum (C_6D_6) of compound **8.**

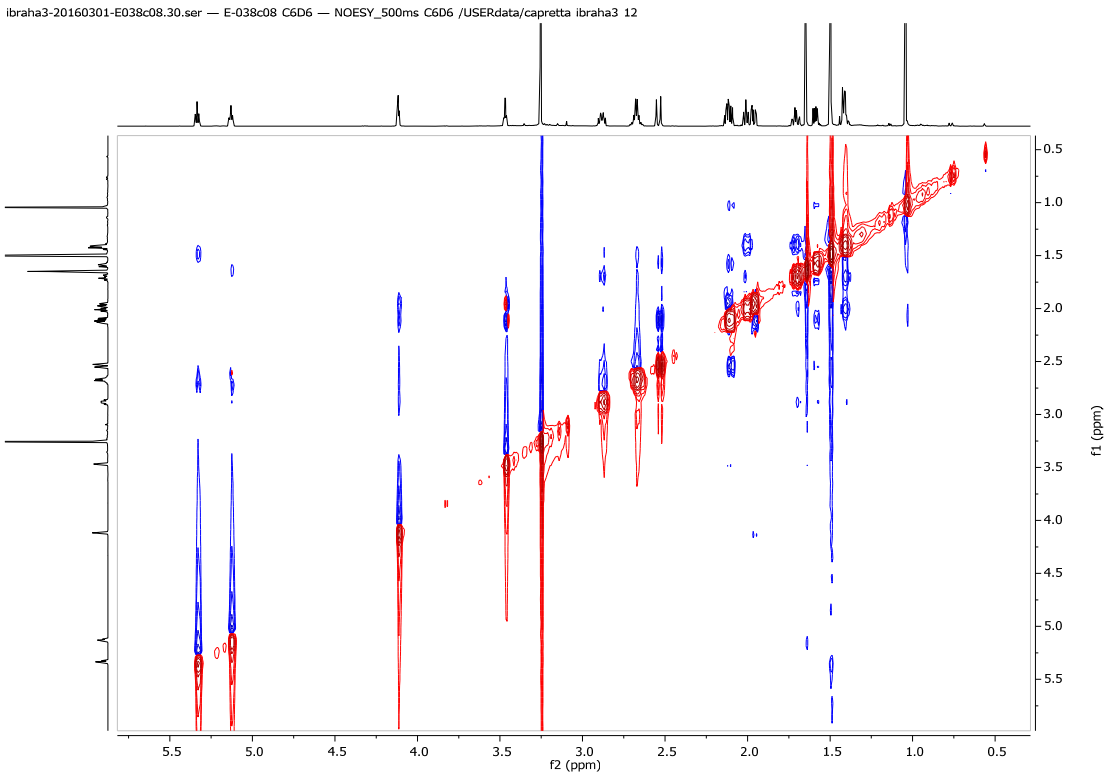


Figure S69 B. NOESY 500ms NMR spectrum (C_6D_6) of compound **8.**

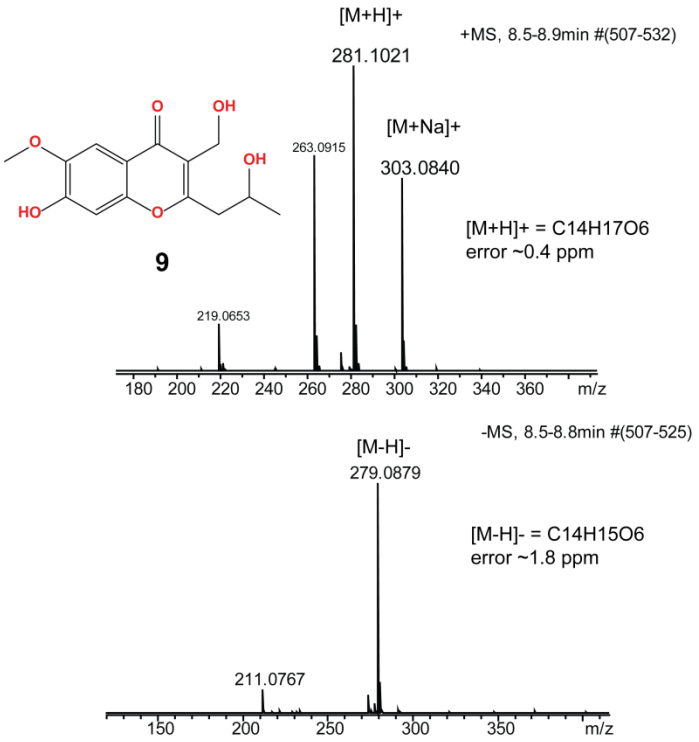


Figure S70 B. LC-HRMS positive and negative ion spectra of compound **9.**

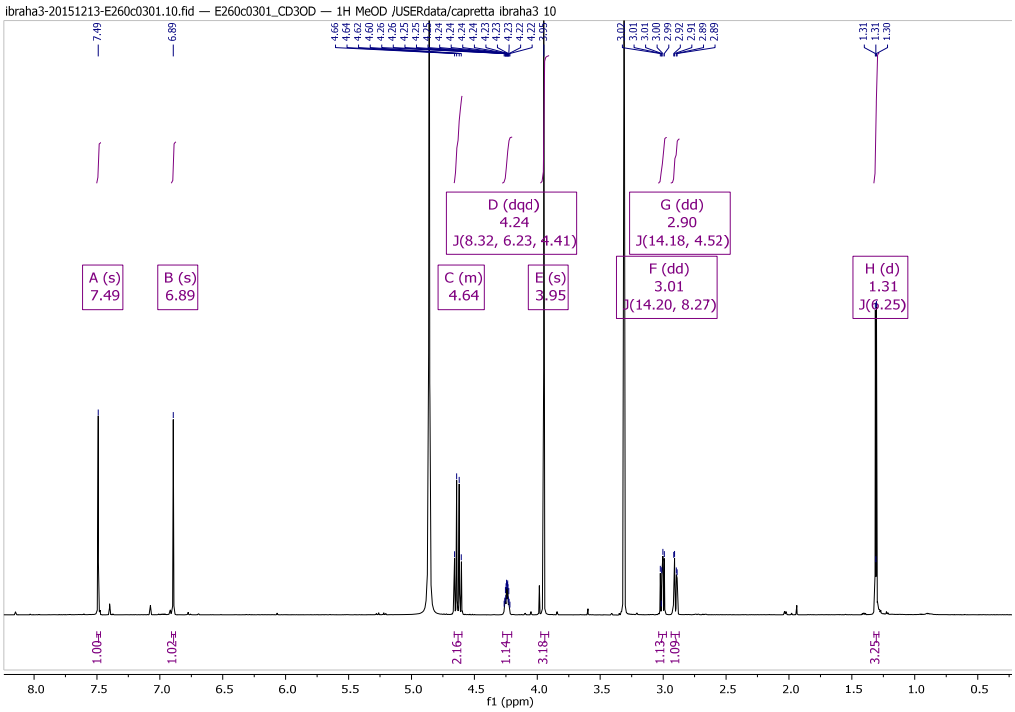


Figure S71. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **9.**

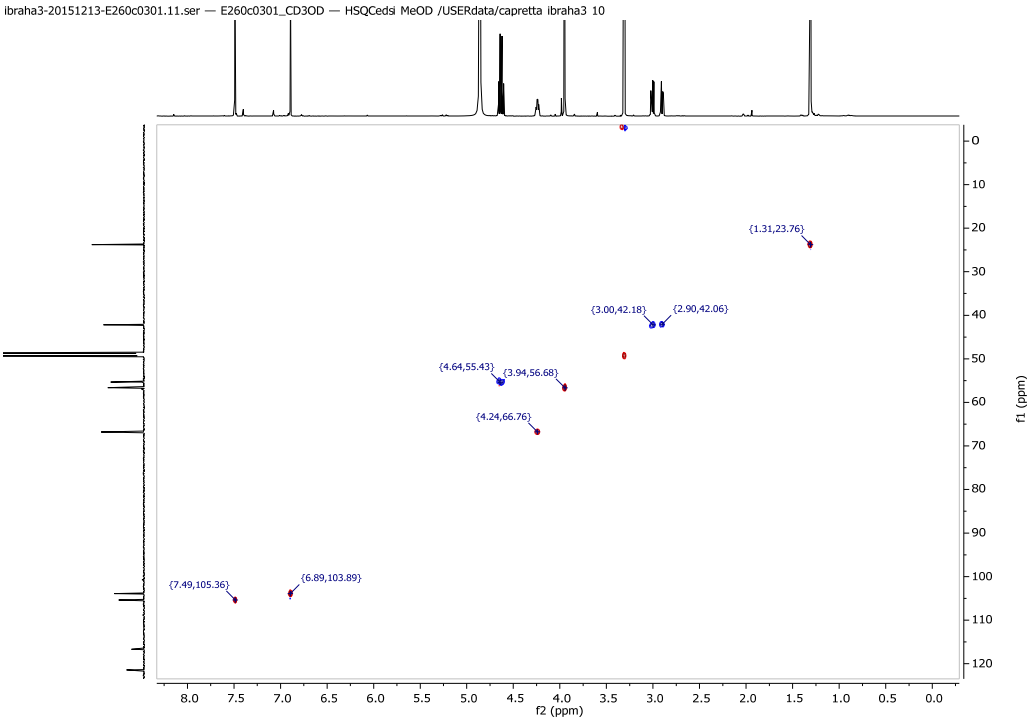


Figure S72 B. Multiplicity-edited HSQC NMR spectrum (CD_3OD) of compound **9.**

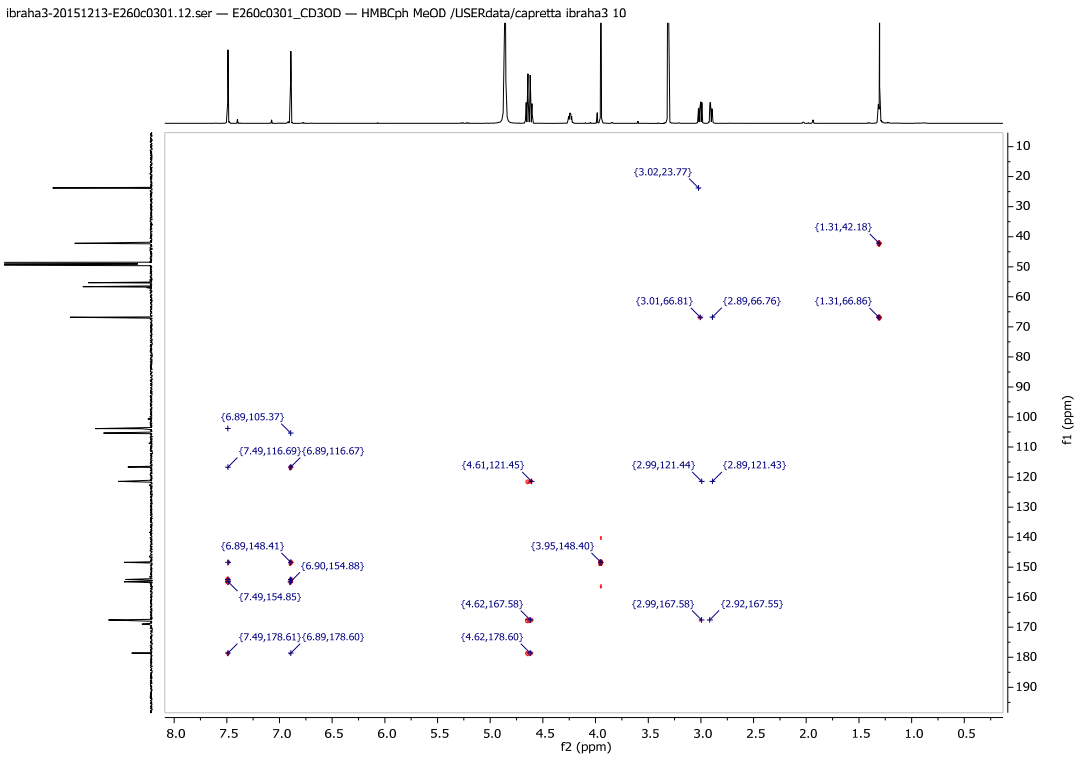


Figure S73 B. HMBC NMR spectrum (CD_3OD) of compound **9.**

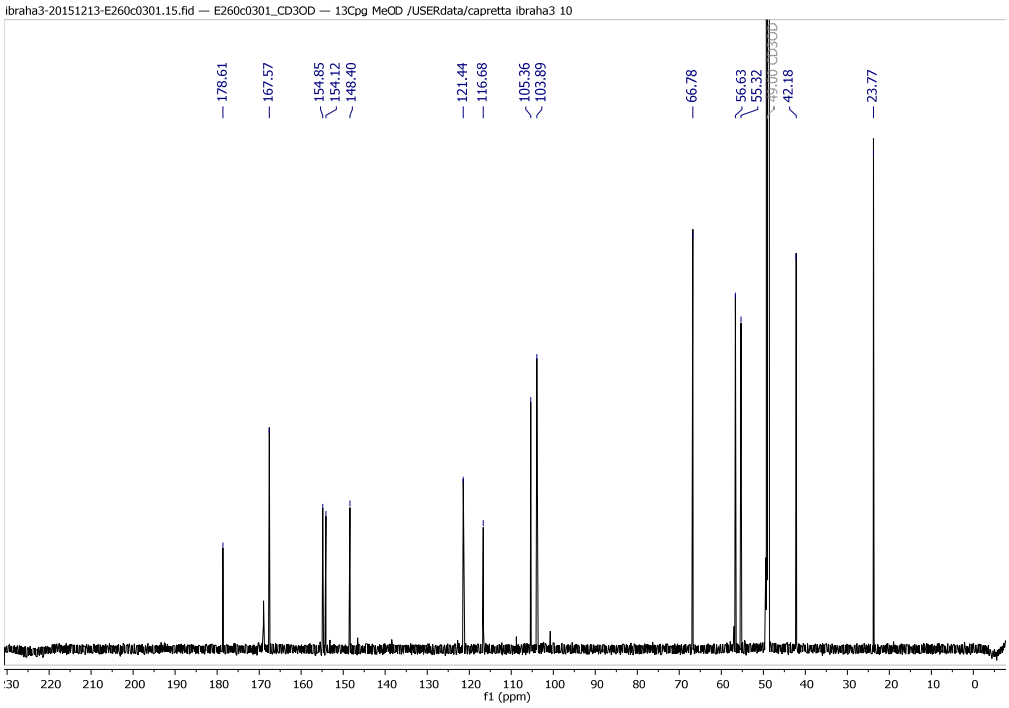


Figure S74 B. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **9.**

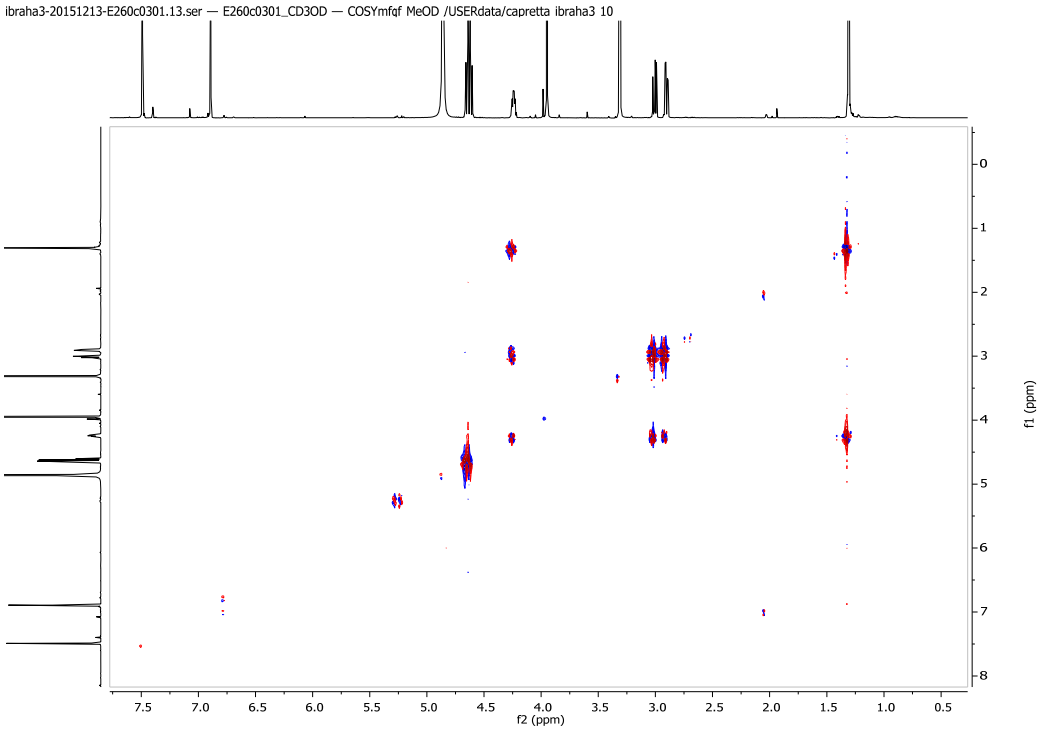


Figure S75 B. COSY NMR spectrum (700 MHz, CD_3OD) of compound **9.**

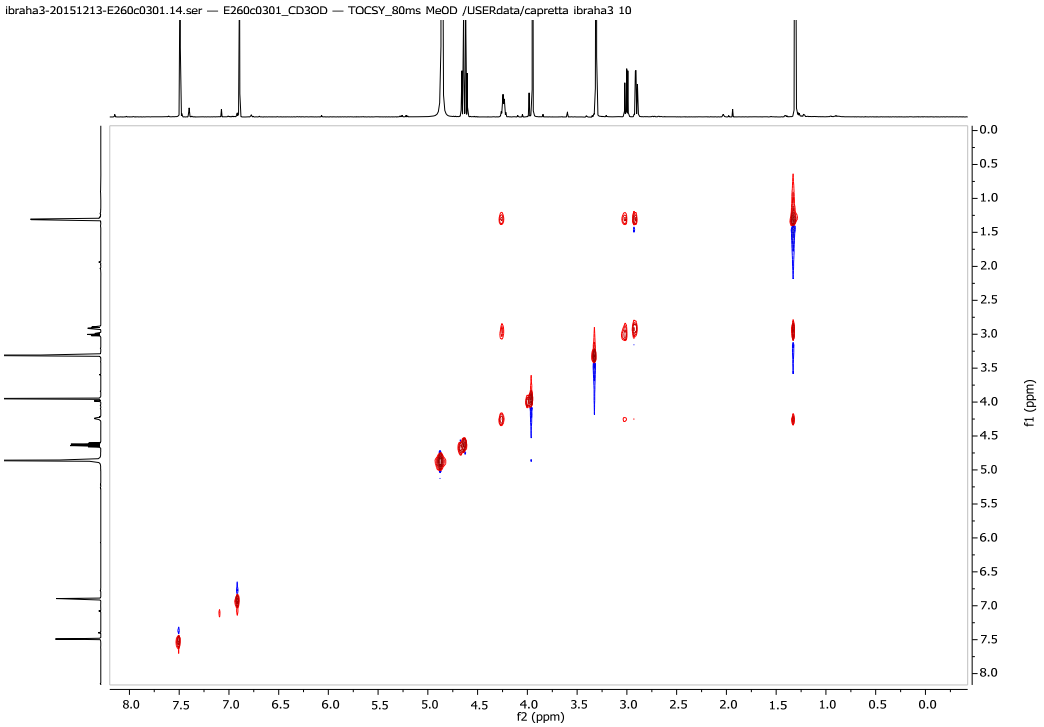


Figure S76 B. TOCSY NMR spectrum (CD_3OD) of compound **9.**

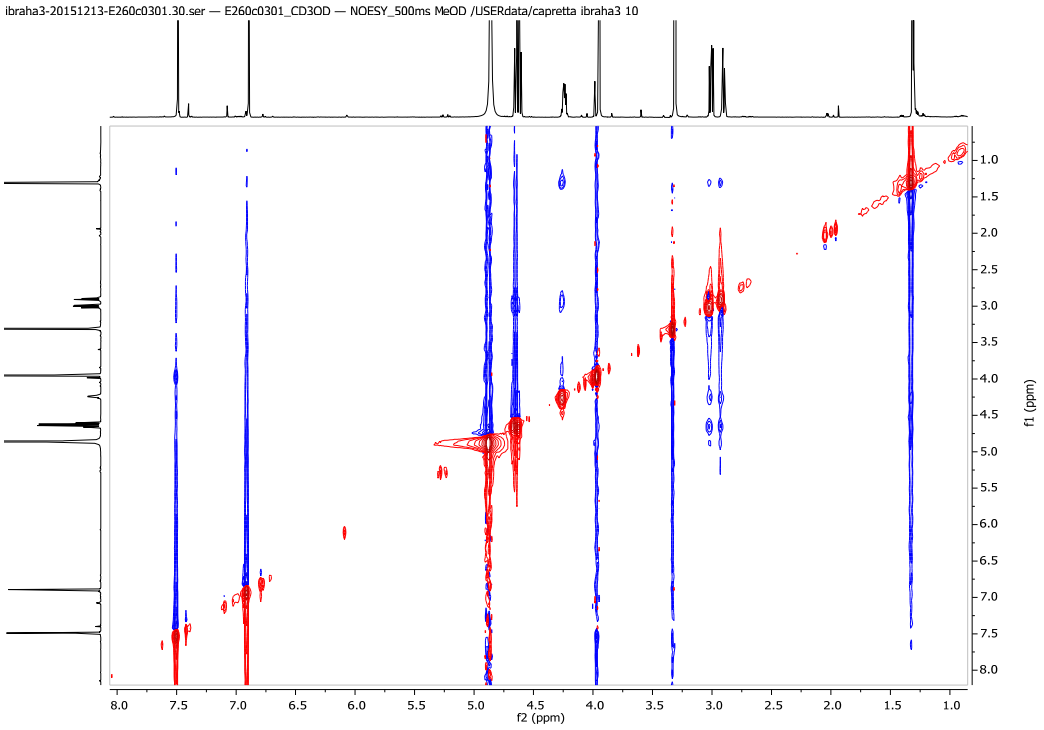


Figure S77 B. NOESY 500ms NMR spectrum (CD_3OD) of compound **9.**

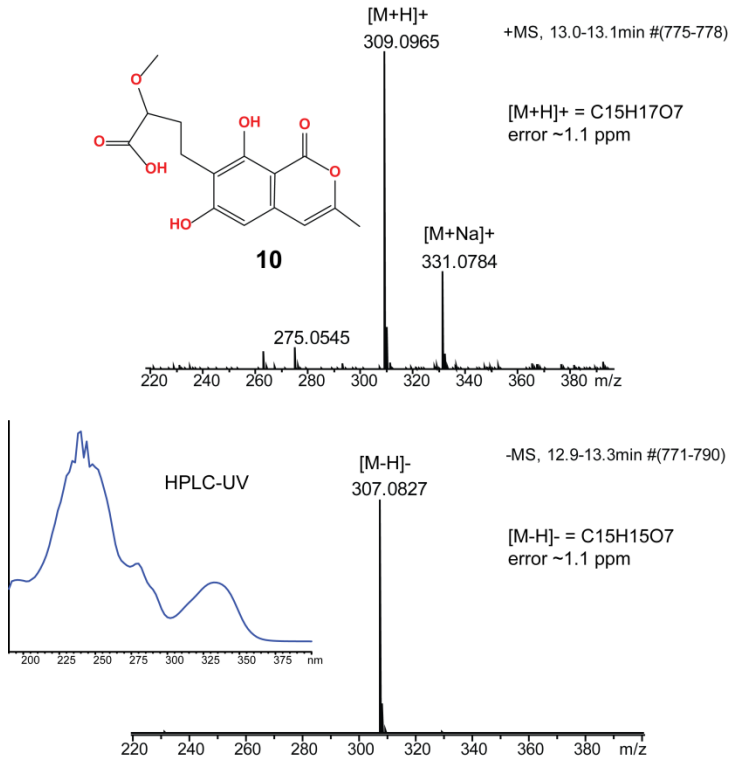


Figure S78 B. LC-HRMS positive and negative ion spectra of compound **10.**

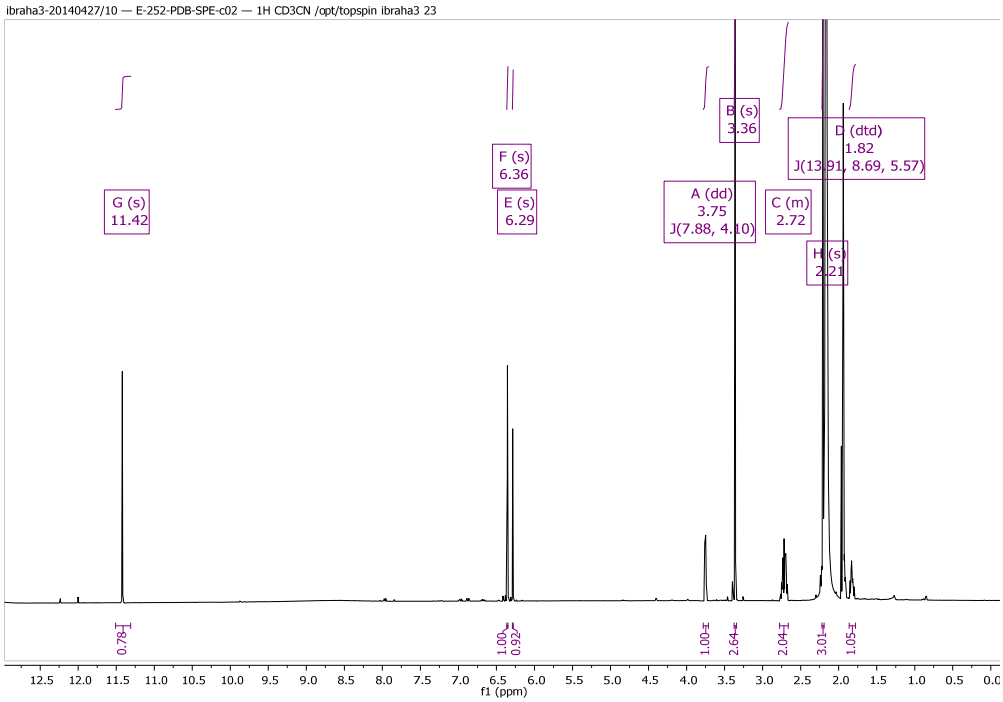


Figure S79 B. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **10.**

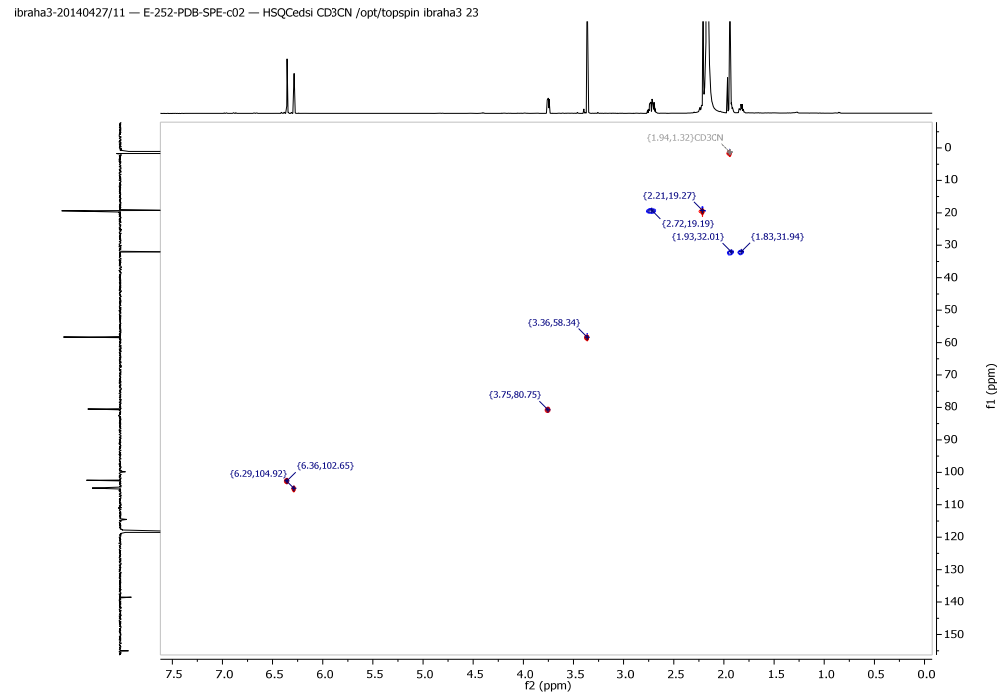


Figure S80 B. Multiplicity-edited HSQC NMR spectrum (CD_3CN) of compound **10.**

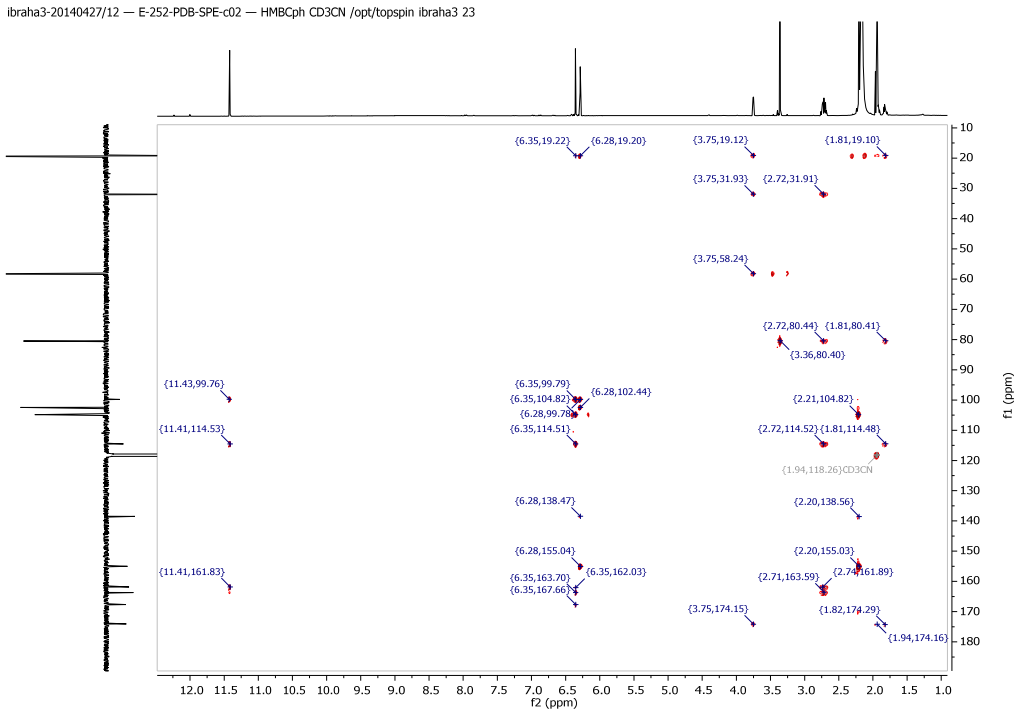


Figure S81 B. HMBC NMR spectrum (CD_3CN) of compound **10.**

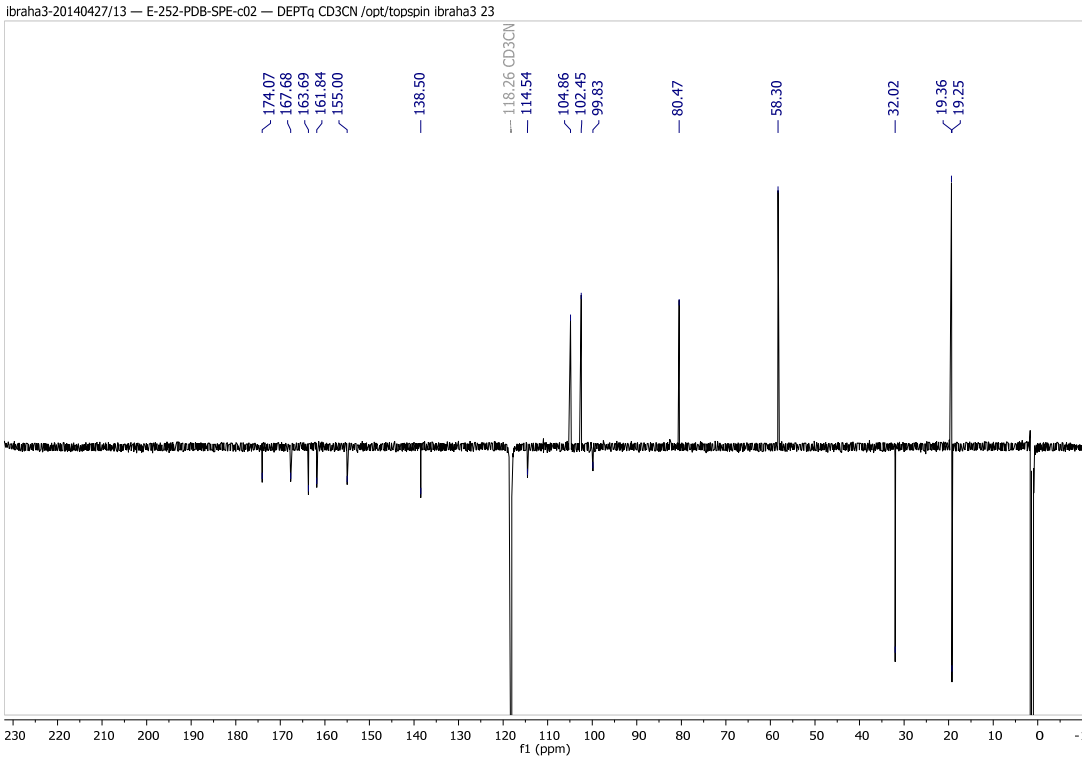


Figure S82 B. ^{13}C DEPTq NMR spectrum (176 MHz, CD_3CN) of compound **10.**

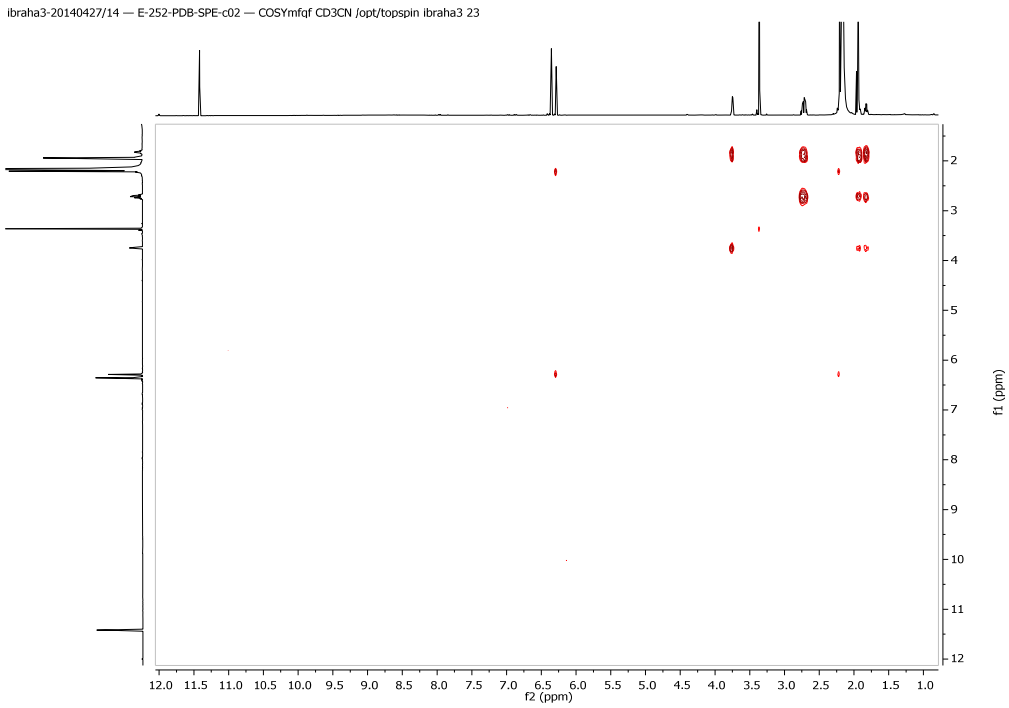


Figure S83 B. COSY NMR spectrum (700 MHz, CD_3CN) of compound **10.**

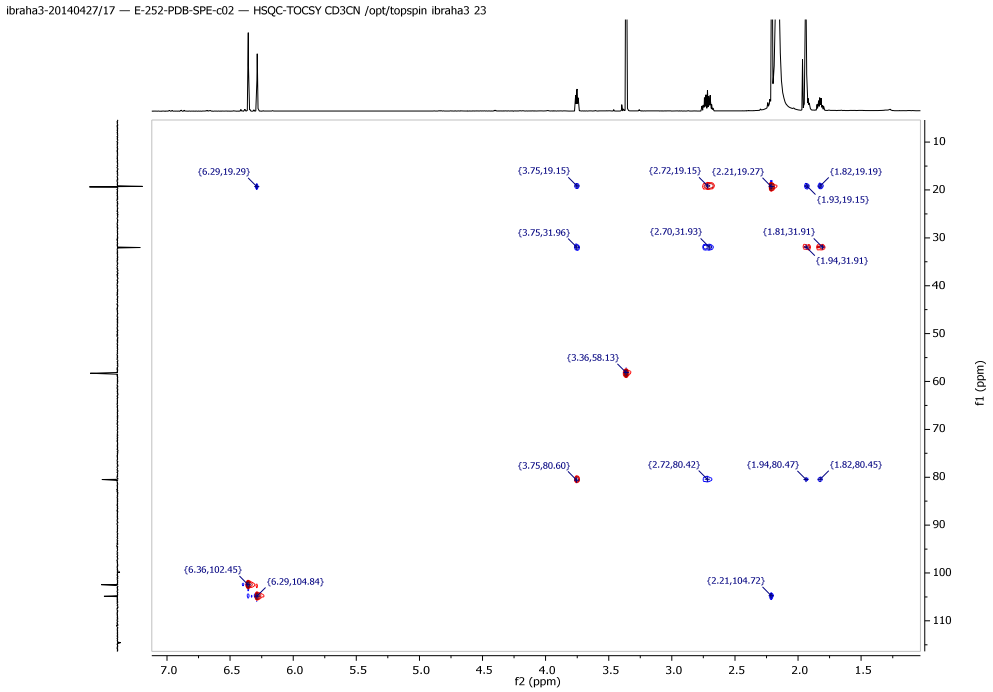


Figure S84 B. HSQC-TOCSY NMR spectrum (CD_3CN) of compound **10.**

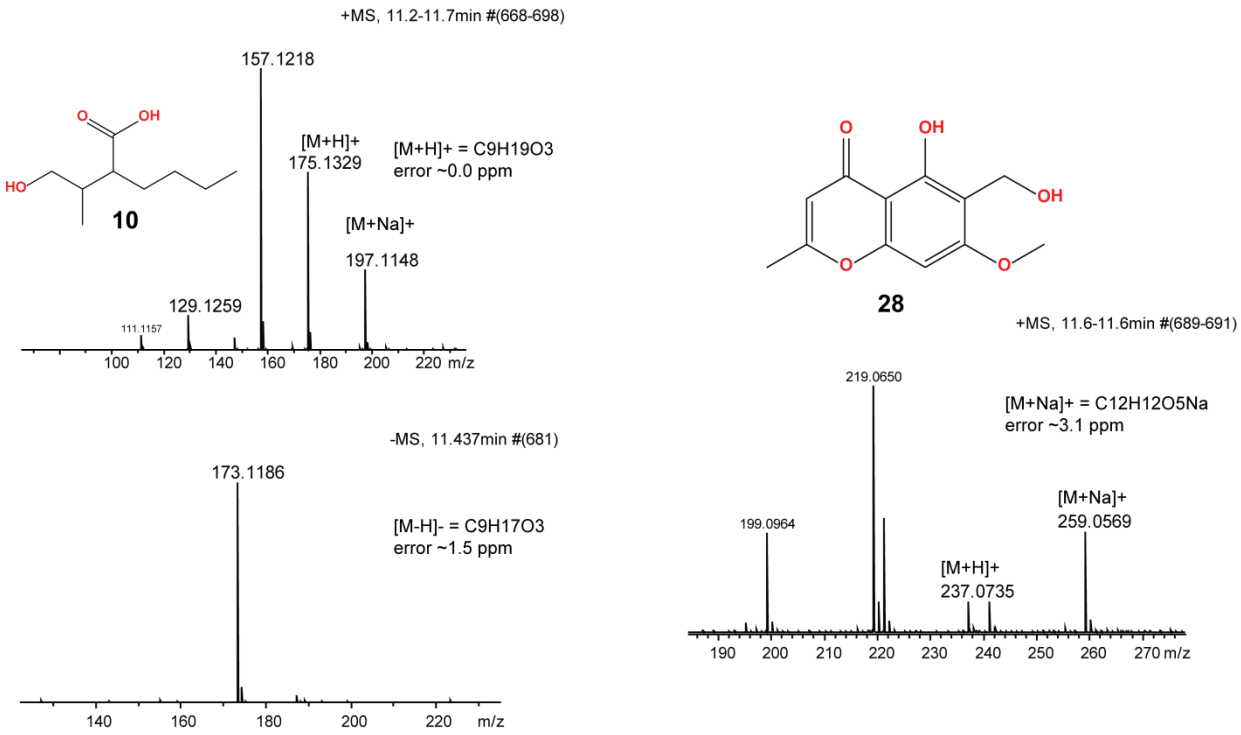


Figure S85. LC-HRMS spectra of compound **11** and compound **28,** 6-hydroxymethyleugenin. Compounds **11** and **28** were isolated as a co-eluting mixture.

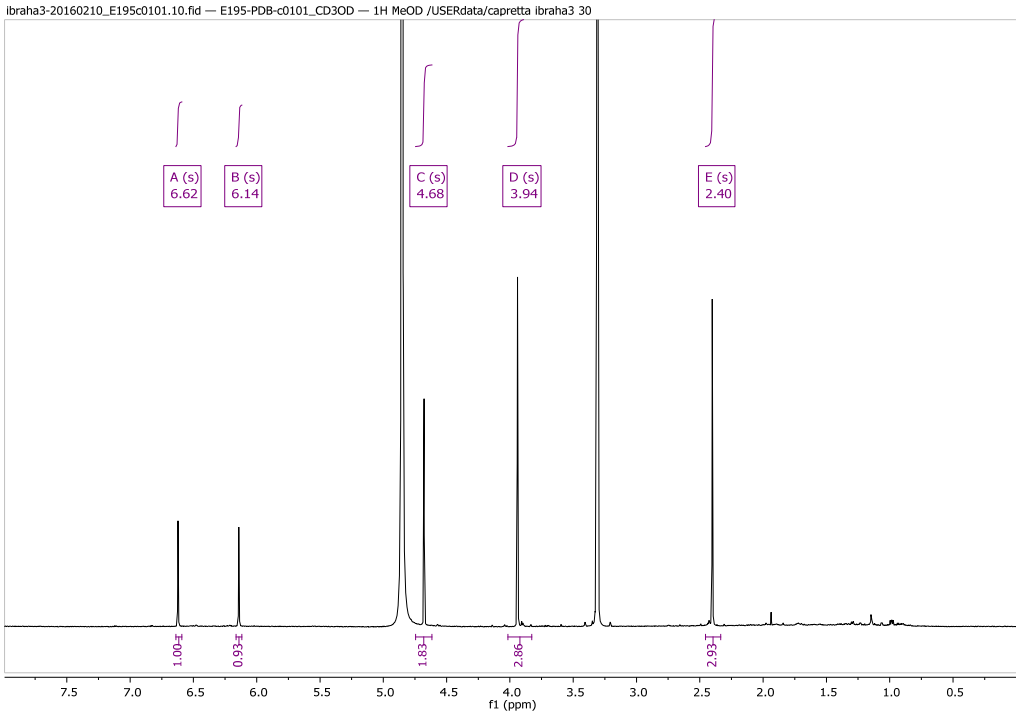


Figure S86. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **28**, after a second HPLC purification step to resolve co-elution with compound **10.**

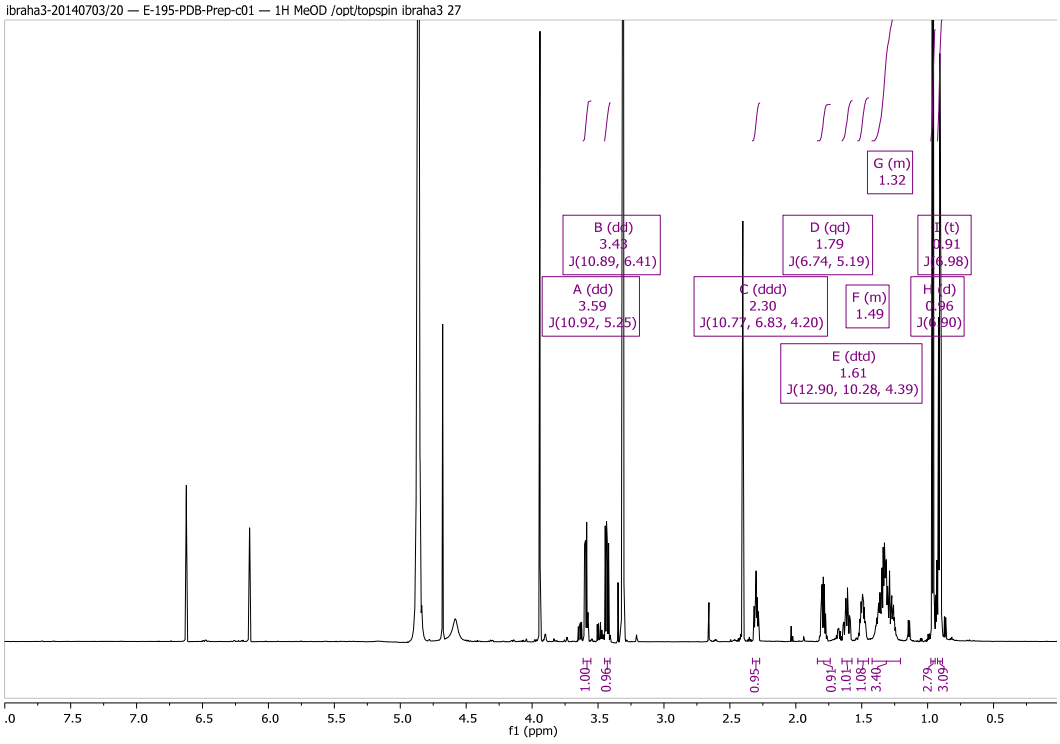


Figure S87 B. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **11** and **28.**

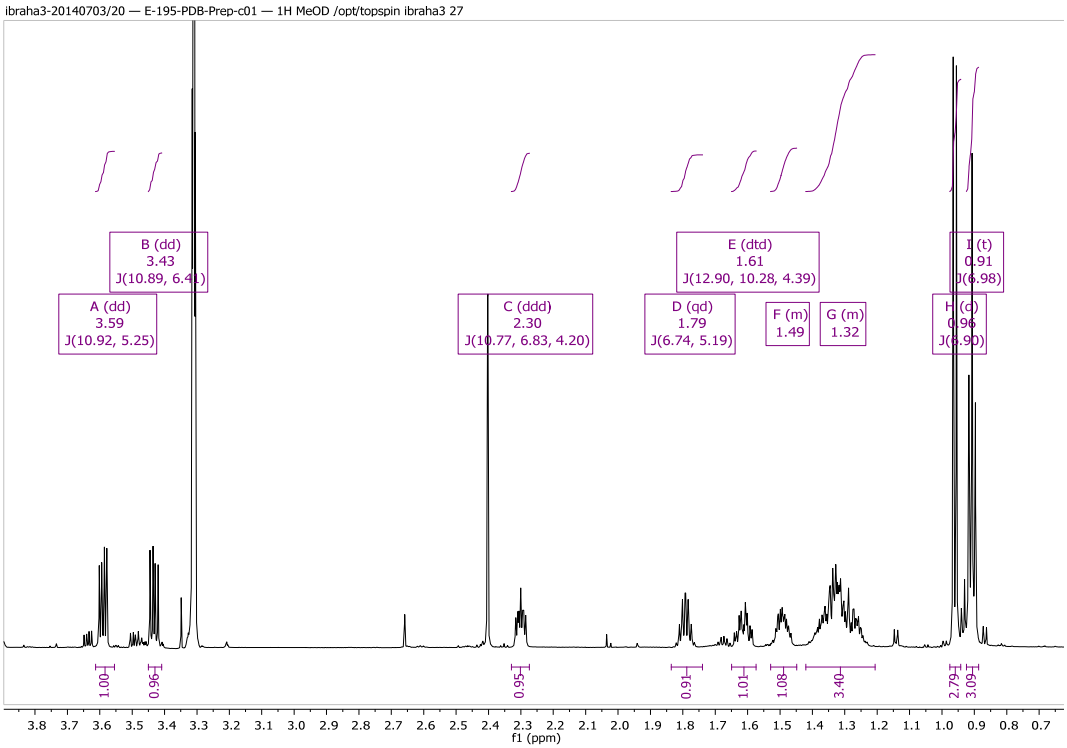


Figure S88 B. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **11** and **28** from 0.6-4 ppm.

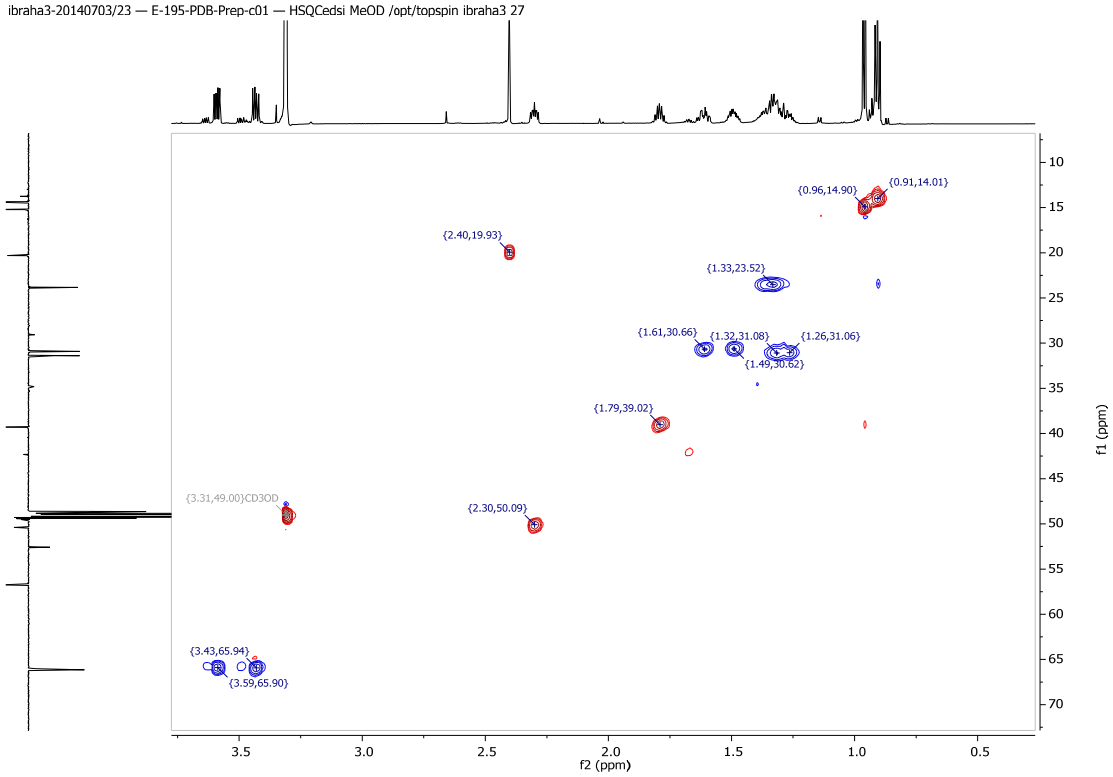


Figure S89 B. Multiplicity-edited HSQC NMR spectrum (CD_3OD) of compound **11** and **28** from 0.6-4 ppm.

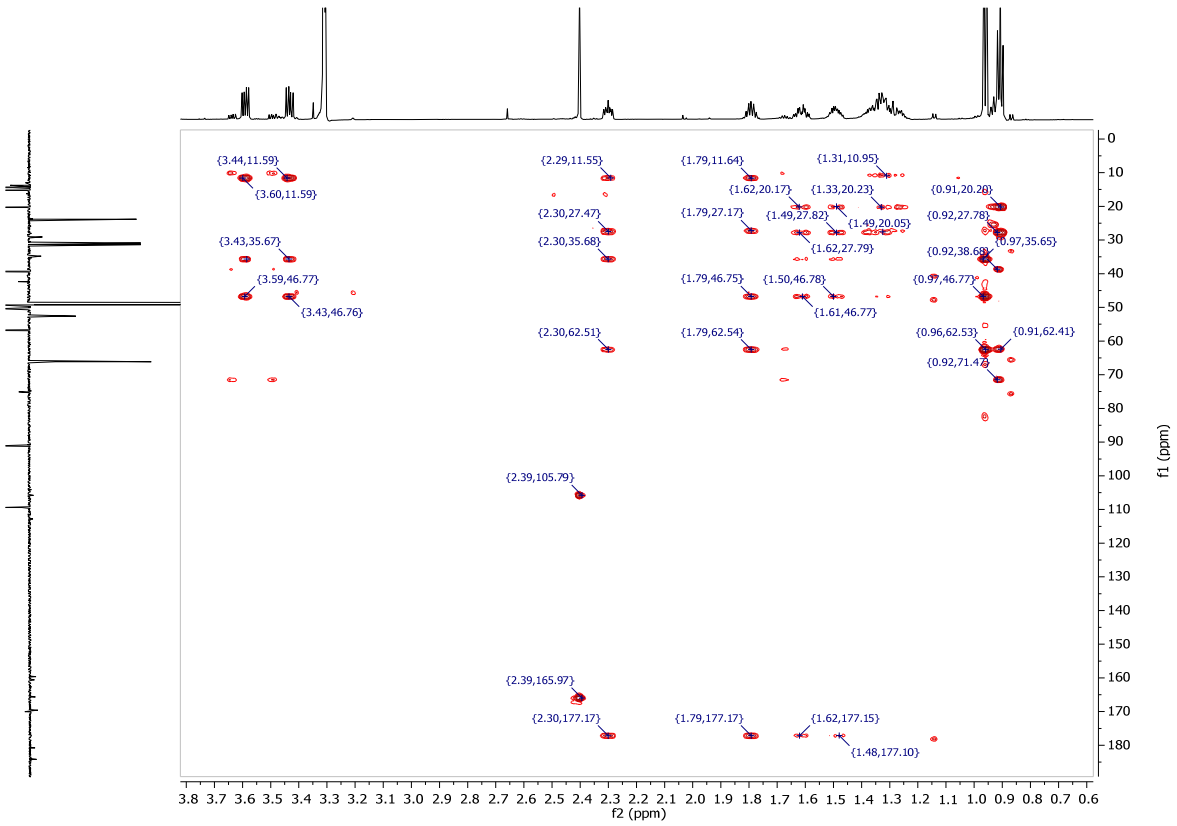


Figure S90 B. HMBC NMR spectrum (CD_3OD) of compound **11** and **28** from 0.6-4 ppm.

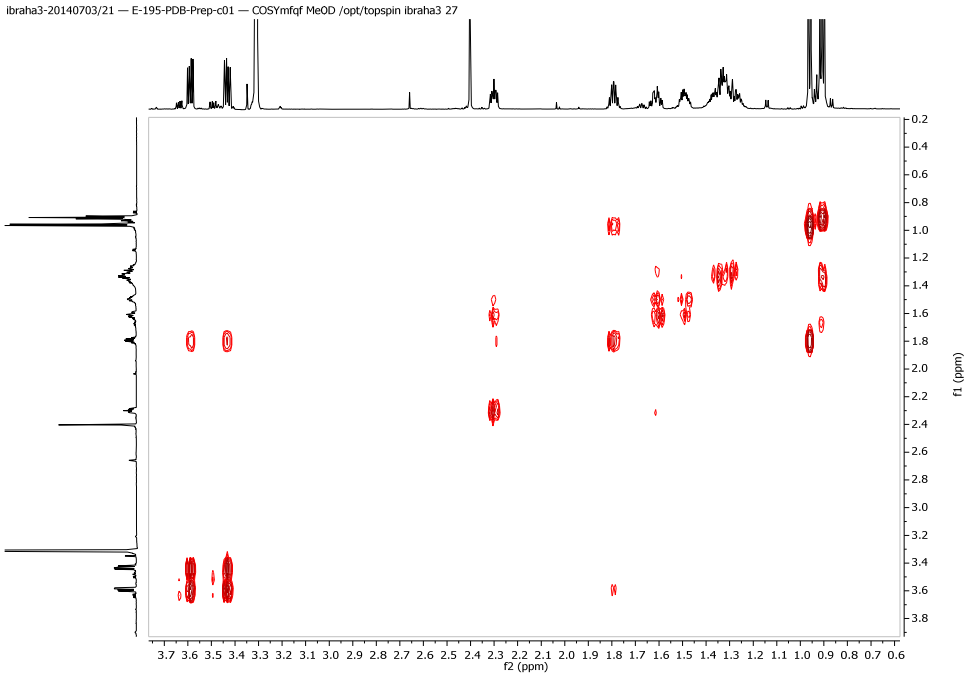


Figure S91 B. COSY NMR spectrum (CD_3OD) of compound **11** and **28** from 0.6-4 ppm.

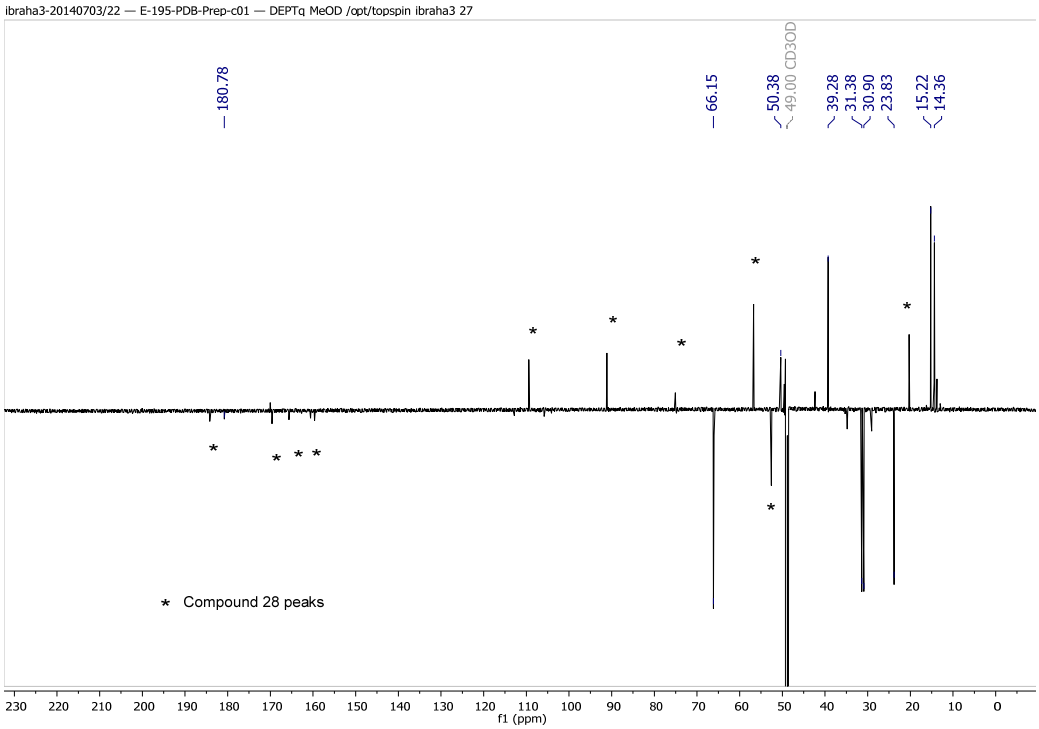


Figure S92 B. ^{13}C DEPTq NMR spectrum (176 MHz, CD_3OD) of compound **11** and **28.**

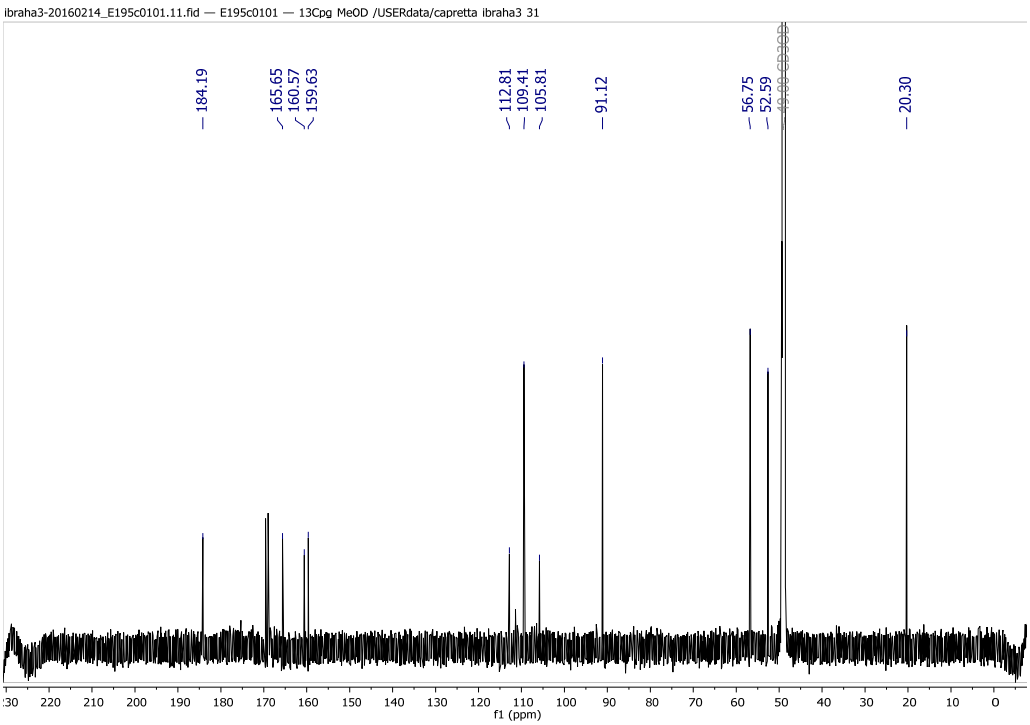


Figure S93 B. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **28.**

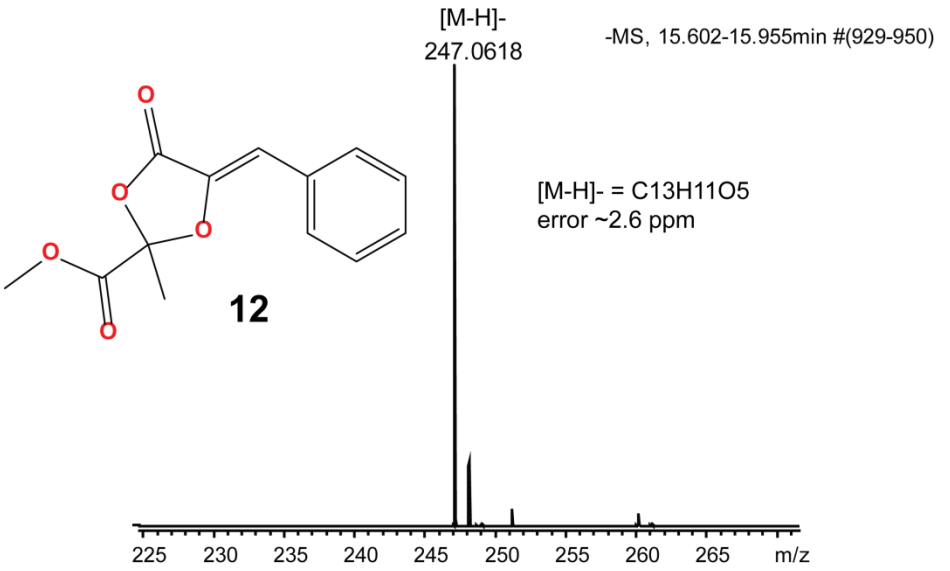


Figure S94 B. LC-HRMS negative ion spectra of compound **12.**

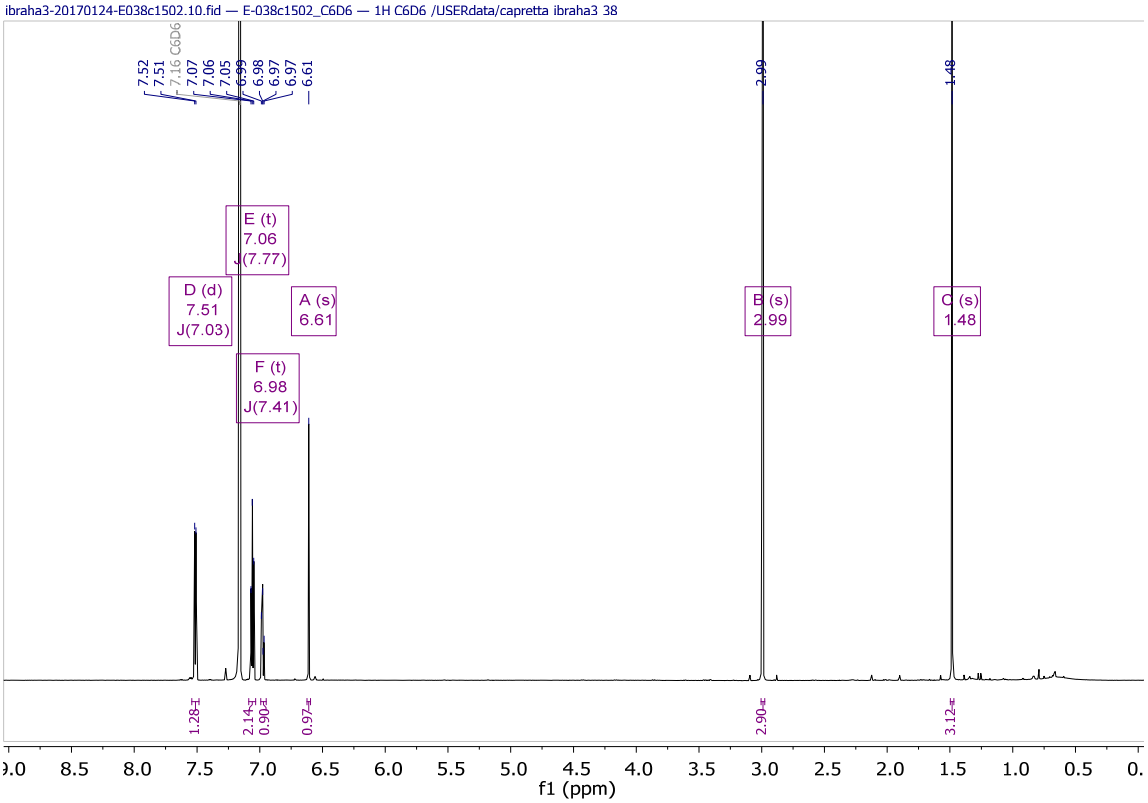


Figure S95 B. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **12.**

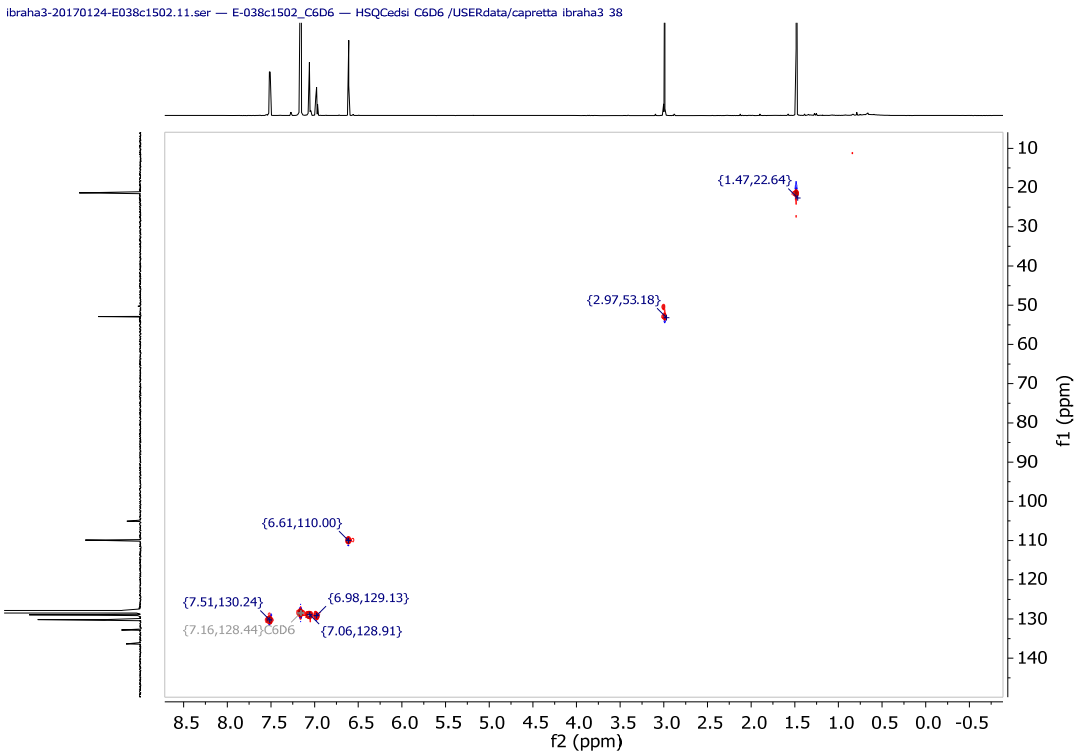


Figure S96 B. Multiplicity-edited HSQC NMR spectrum (C_6D_6) of compound **12.**

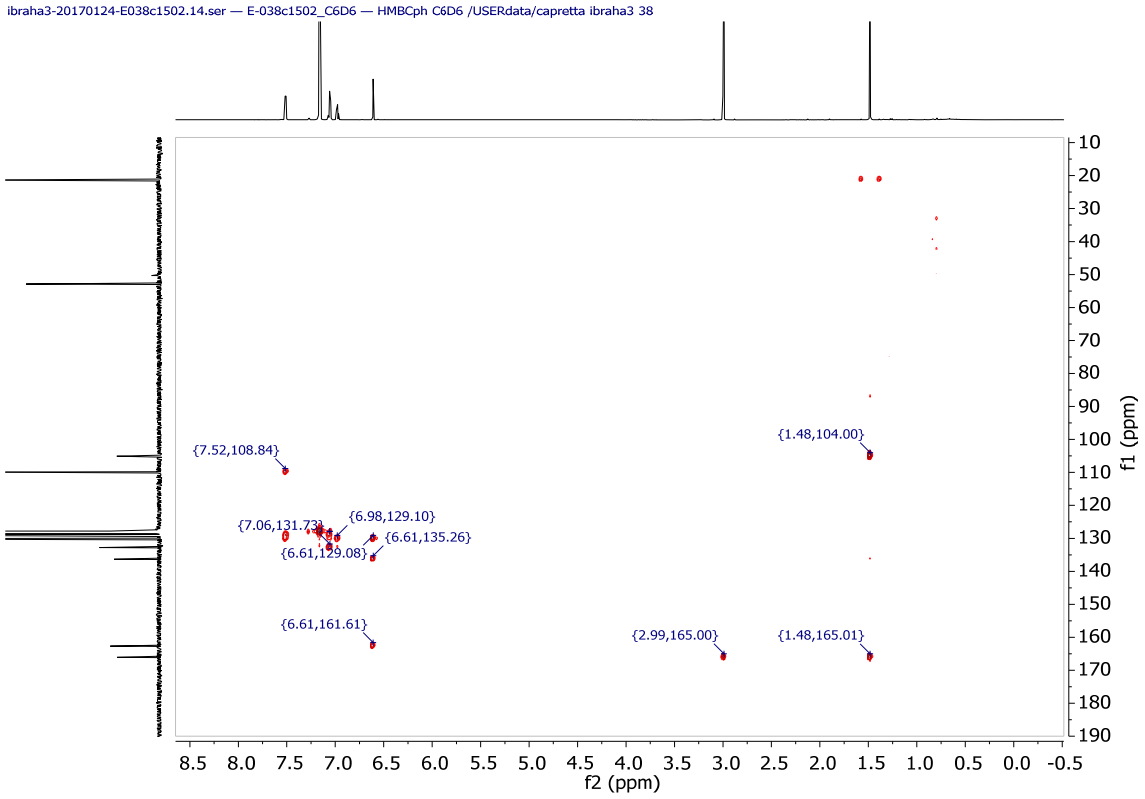


Figure S97 B. HMBC NMR spectrum (C_6D_6) of compound **12.**

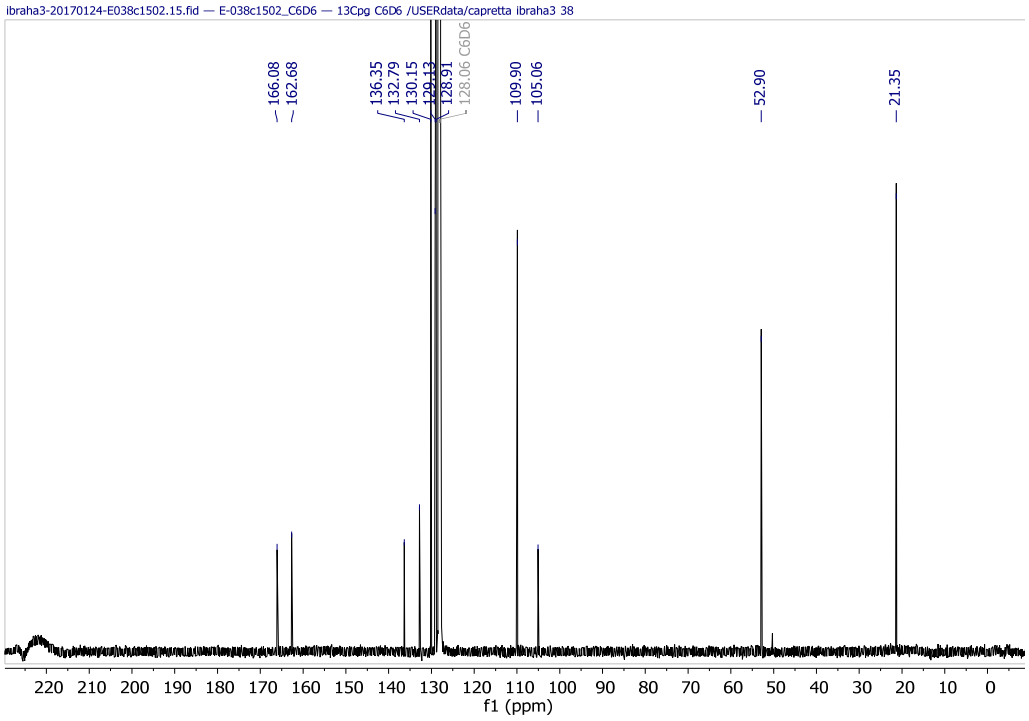


Figure S98 B. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **12.**

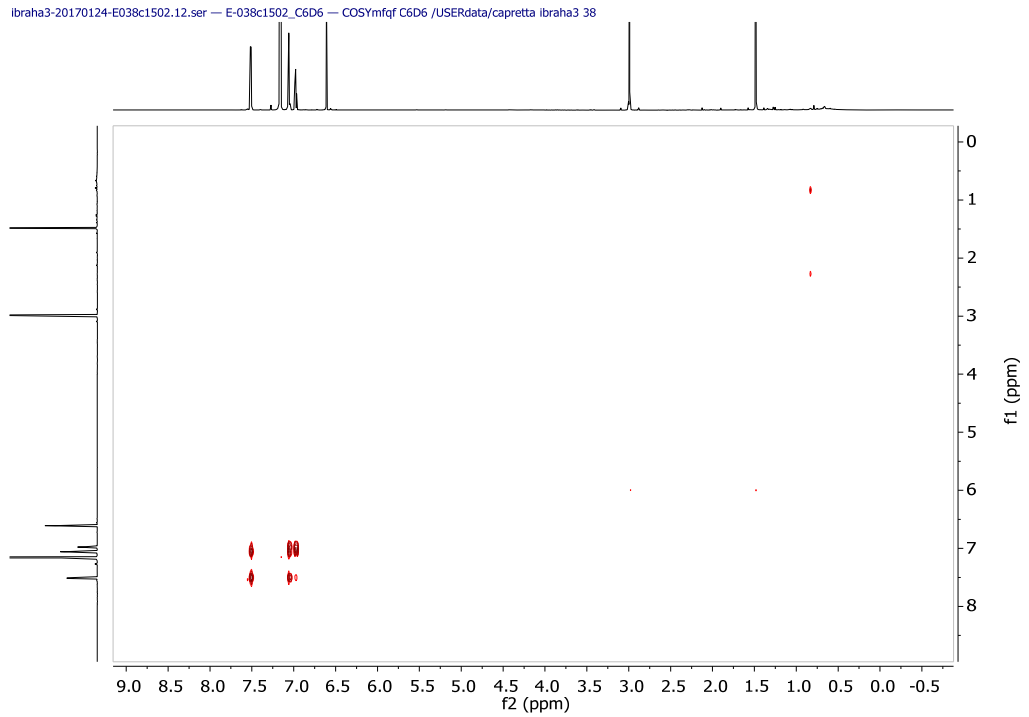
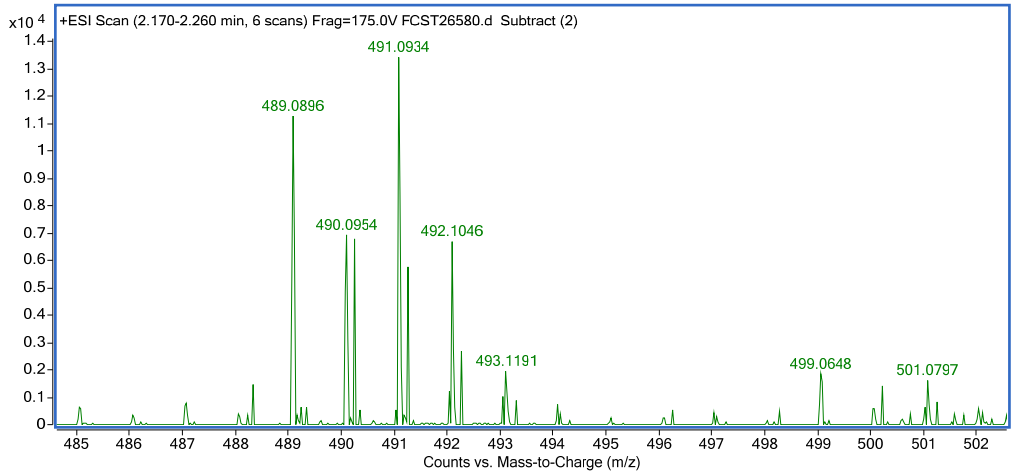


Figure S99 B. COSY NMR spectrum (700 MHz, C_6D_6) of compound **12.**



Formula (M)	Score	Mass	Calc Mass	Calc m/z	Diff (ppm)	DBE	m/z
C24 H25 Br O6	97.37	488.0823	488.0835	489.0907	2.31	12	489.0896

Figure S100 B. HRMS positive ion spectra of compound **13** by direct infusion**.**

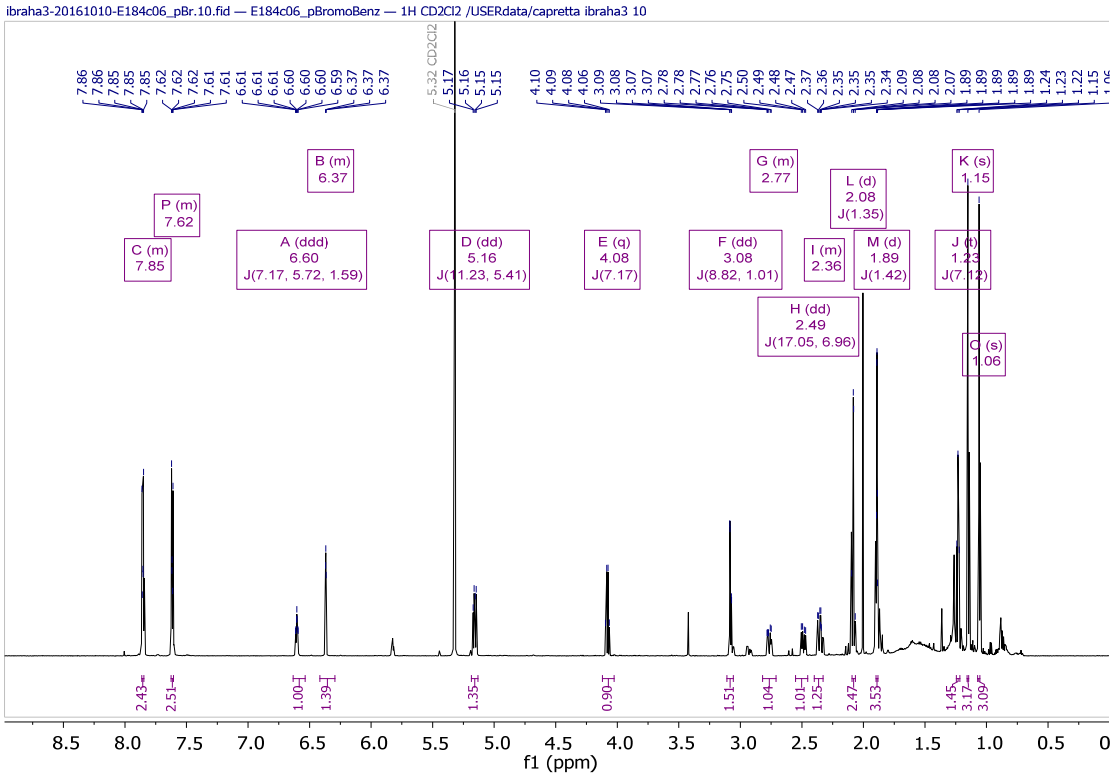


Figure S101 B. ^{1}H NMR spectrum (700 MHz, CD_2Cl_2) of compound **13.**

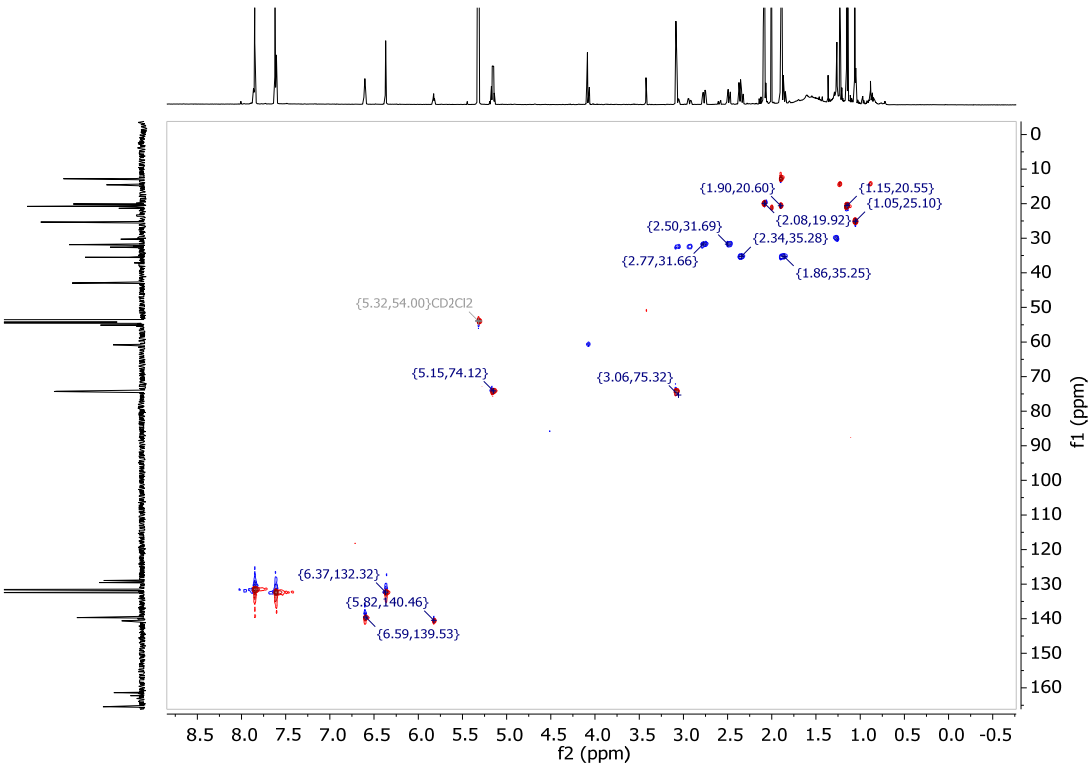


Figure S102 B. Multiplicity-edited HSQC NMR spectrum (CD_2Cl_2) of compound **13.**

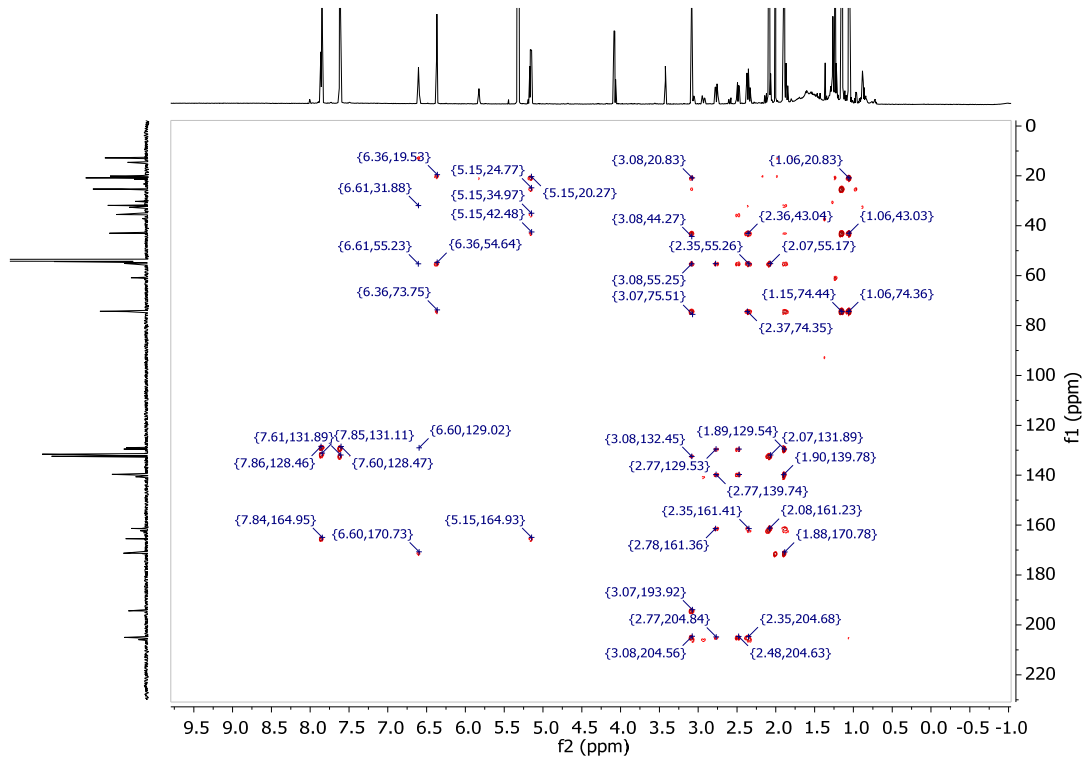


Figure S103 B. HMBC NMR spectrum (CD_2Cl_2) of compound **13.**

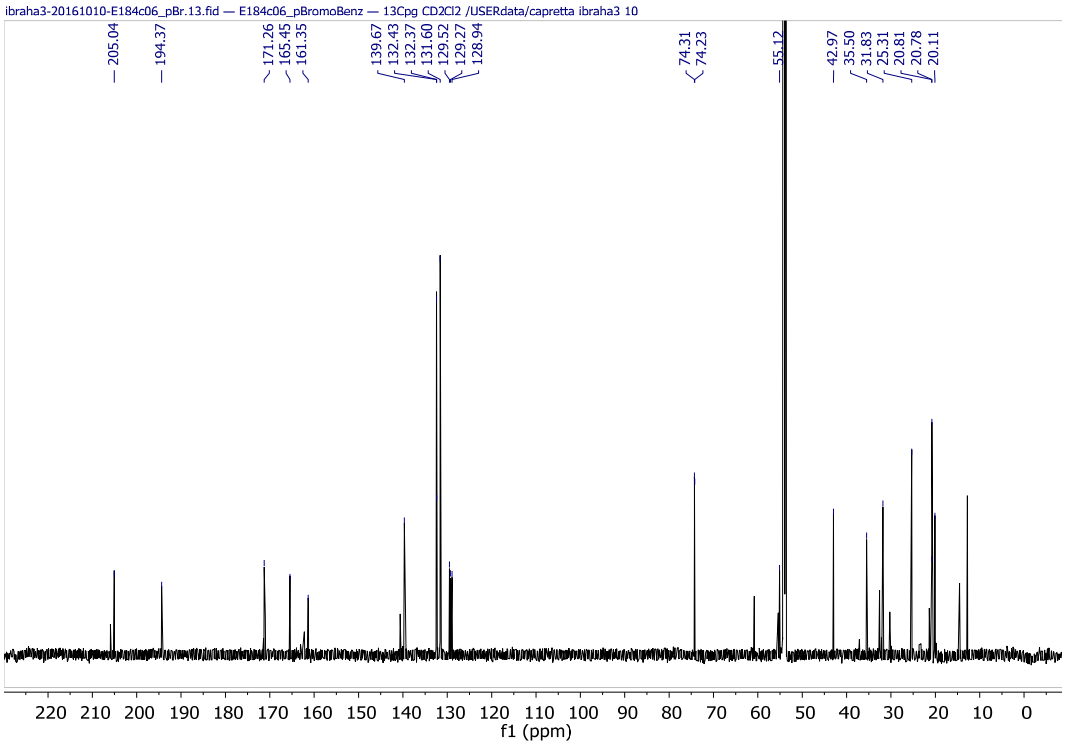


Figure S104 B. ^{13}C NMR spectrum (176 MHz, CD_2Cl_2) of compound **13.**

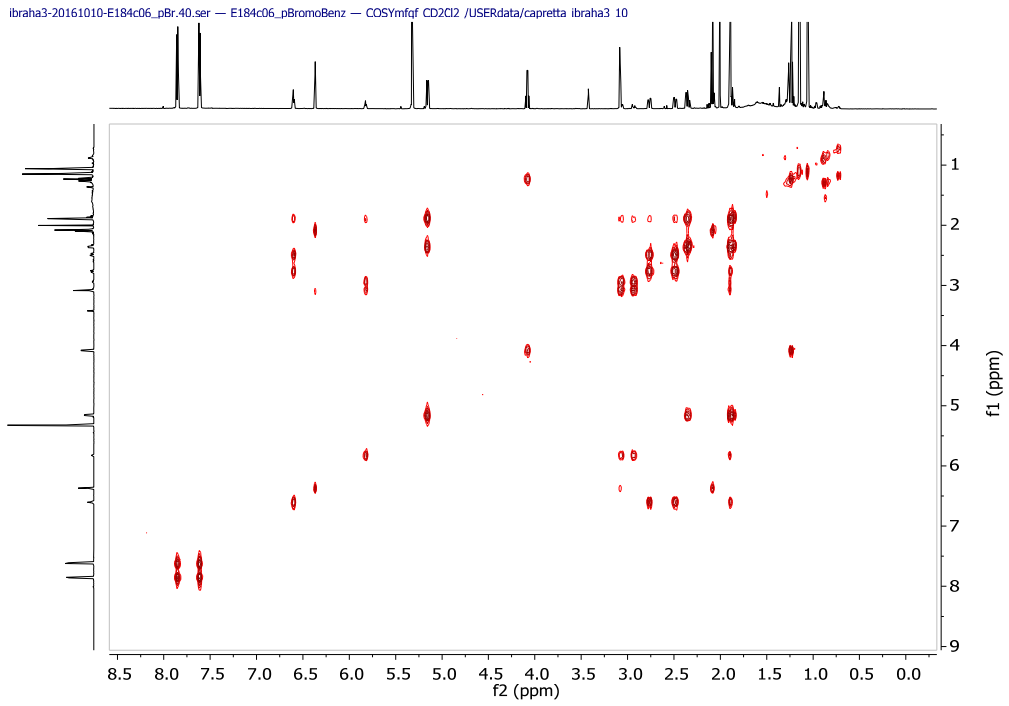


Figure S105 B. COSY NMR spectrum (700 MHz, CD_2Cl_2) of compound **13.**

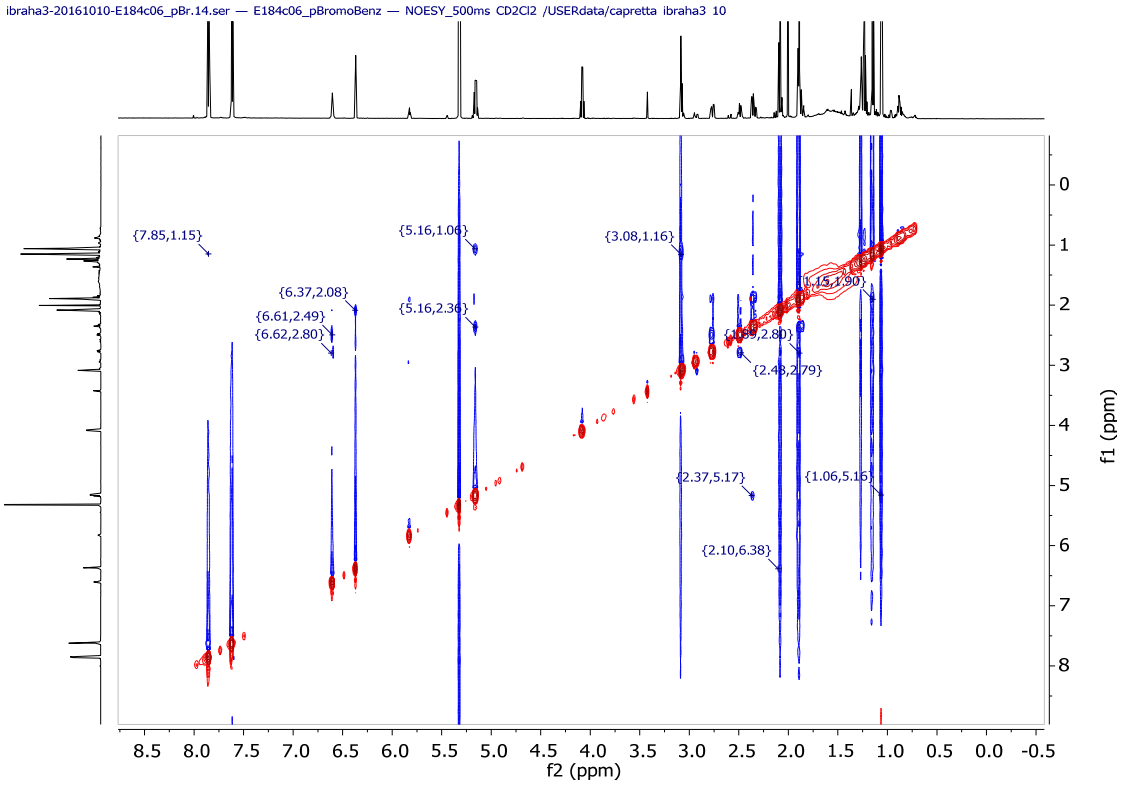


Figure S106 B. NOESY 500ms NMR spectrum (700 MHz, CD_2Cl_2) of compound **13.**

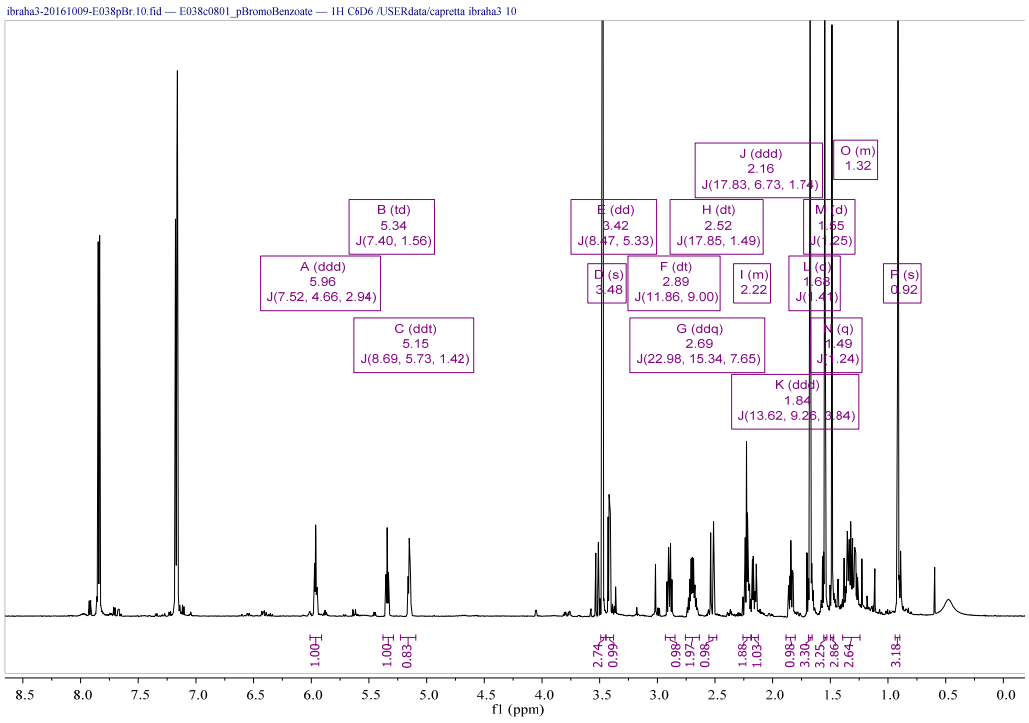


Figure S107 B. ^{1}H NMR spectrum (700 MHz, C6D6) of compound **14.**

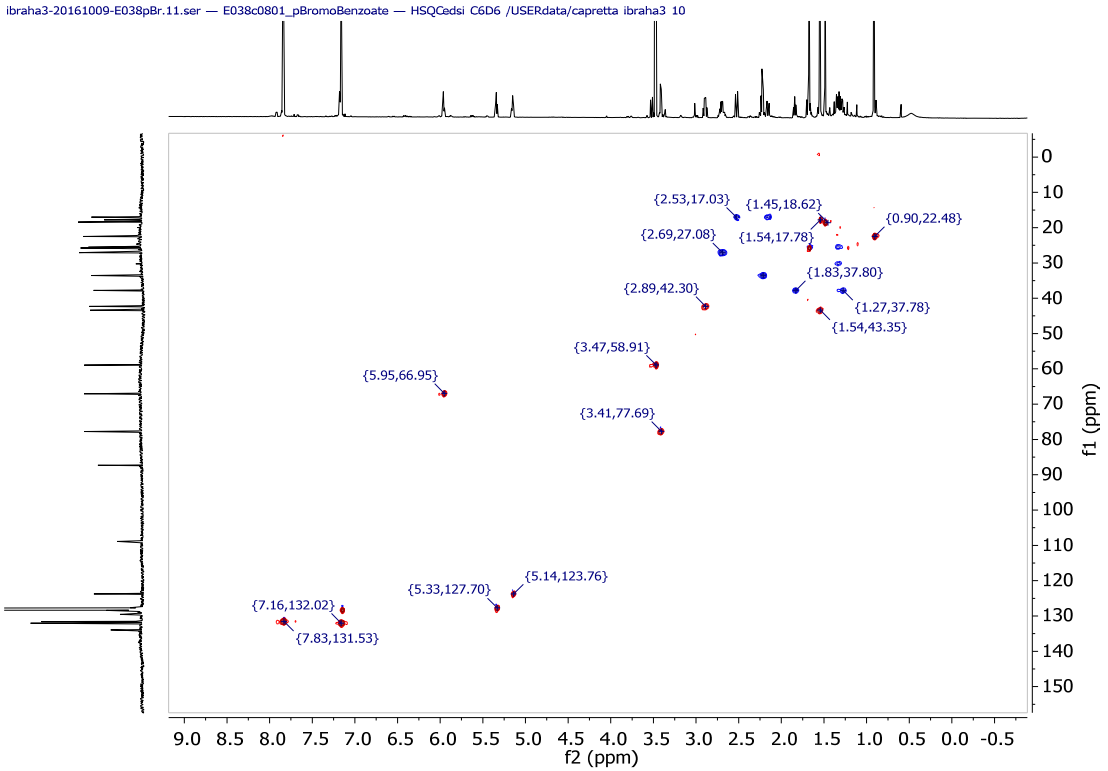


Figure S108 B. Multiplicity-edited HSQC NMR spectrum (C6D6) of compound **14.**

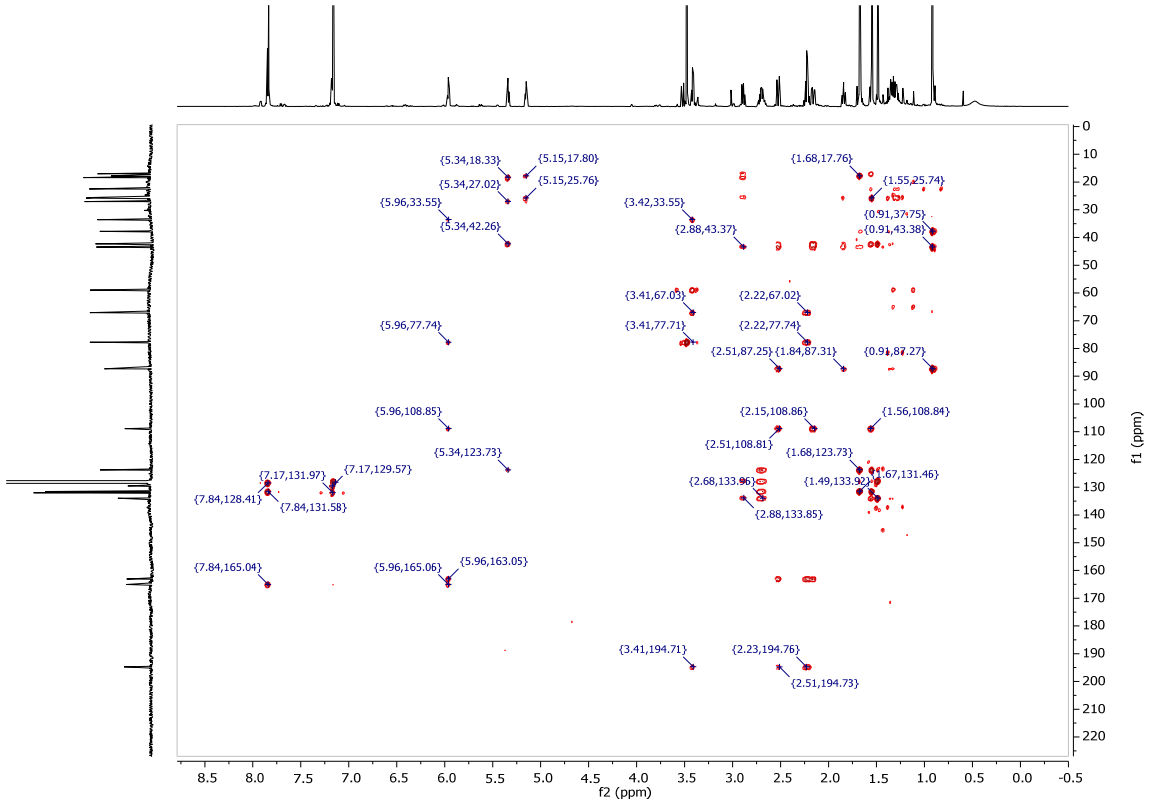


Figure S109 B. HMBC NMR spectrum (C6D6) of compound **14.**

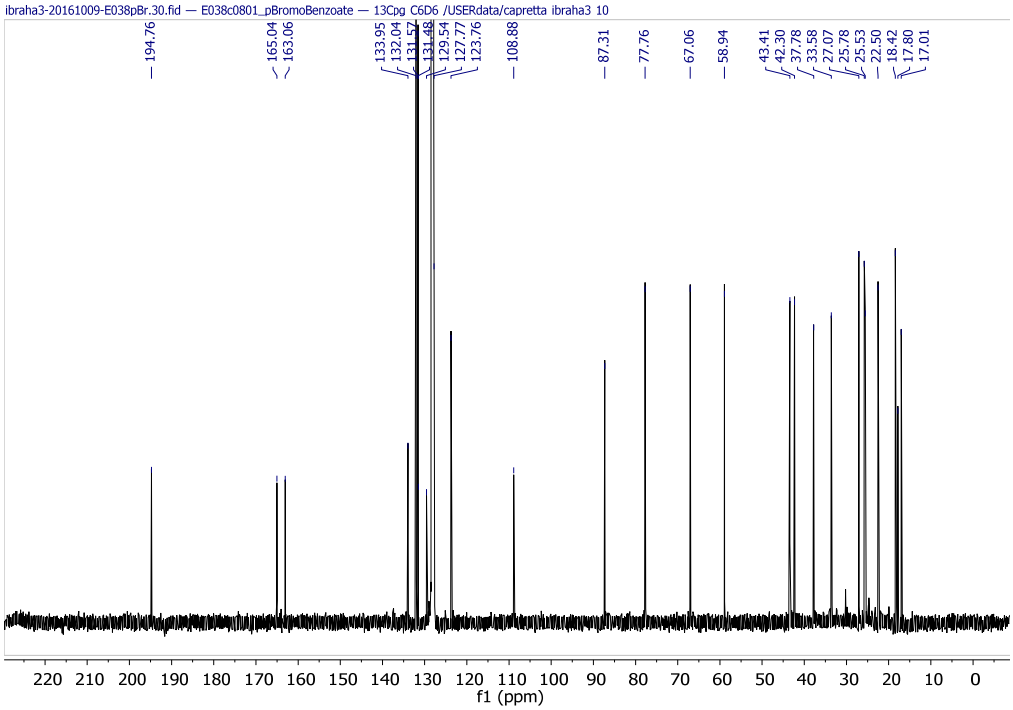


Figure S110 B. ^{13}C NMR spectrum (176 MHz, C6D6) of compound **14.**

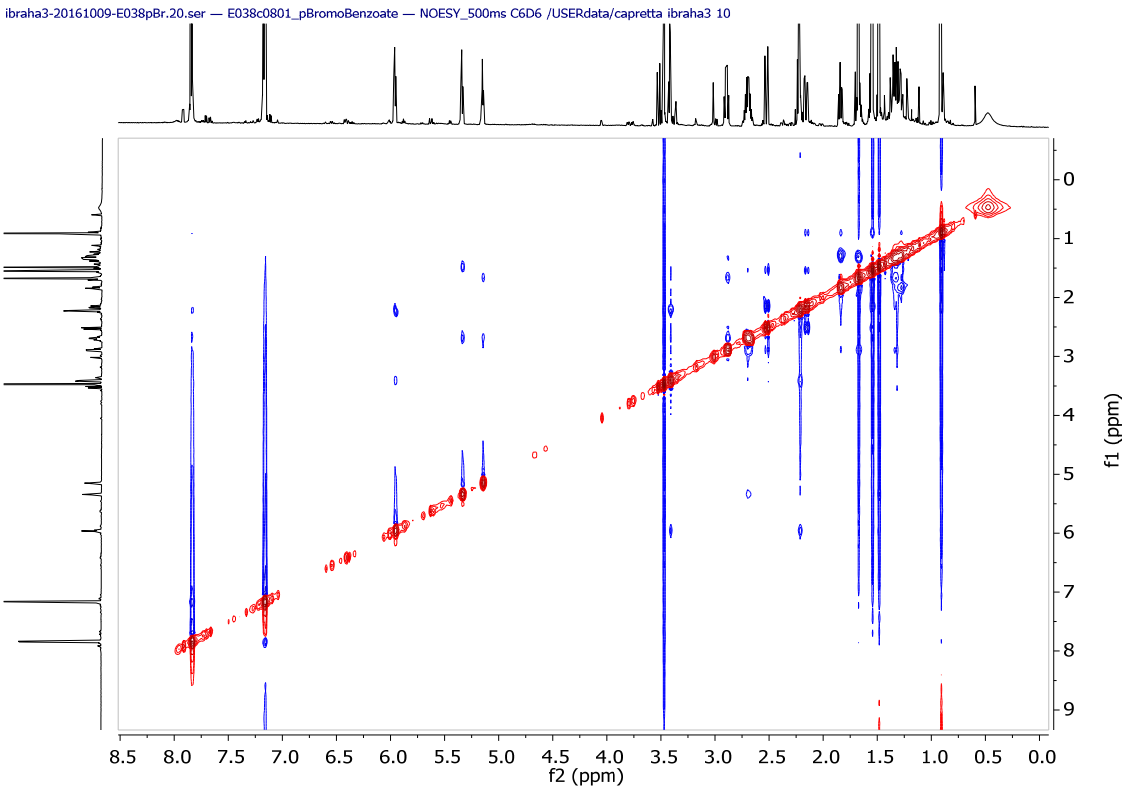


Figure S111 B. NOESY 500ms NMR spectrum (700 MHz, C6D6) of compound **14.**

Bioactivity testing (MICs) were carried out according to CLSI protocols (M7-A5 & M27A)*

#	[highest] in µg/ml	*E. coli ΔbamBΔtolC BW25113*					*Staphylococcus aureus ATCC# 29213*					*Bacillus subtilis 1A1*					*Micrococcus luteus*					*Saccharomyces cerevisiae BY4741*					*Candida albicans ATCC# 90028*				
1	200	112	94	92	91	94	109	102	93	99	101	114	106	108	102	104	315	272	244	222	184	120	114	109	105	104	73	88	100	103	106
3	200	116	102	105	103	106	137	101	102	105	102	99	97	100	102	104	167	138	116	97	98	62	65	88	95	94	199	213	199	137	154
4	200	109	102	107	105	109	132	118	117	112	117	112	111	99	93	98	248	256	158	195	143	118	113	103	107	108	85	91	99	77	117
5	50	125	101	101	100	102	125	116	113	98	108	133	116	114	115	112	149	122	115	120	121	107	110	106	103	105	119	148	100	108	130
6	200	108	91	92	90	85	116	89	101	109	98	134	120	117	119	88	111	110	87	87	78	110	108	102	100	102	135	114	129	113	88
7	200	110	99	96	108	105	\	\	\	\	\	\	\	\	\	\	\	\	\	\	\	18	3	2	3	71	11	7	3	104	104

<25% growth
25-75% growth

Working stock solutions prepared in DMSO to 5, 10 or 20 mg/mL.

1/2 serial dilutions of each compound in DMSO were used
Each series shows the % growth for the 5 highest concentrations tested
(highest concentration on left)
Values shown are the mean of 2 replicates

*National Committee for Clinical Laboratory Standards. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically: approved standard. 5th Ed. Wayne, PA: National Committee for Clinical Laboratory Standards. **2000**. NCCLS document M7-A5.

*National Committee for Clinical Laboratory Standards. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts: approved standard. Wayne, PA: National Committee for Clinical Laboratory Standards. **1997**. NCCLS document M27-A

Figure S112 B. Bioactivity testing of compounds **1** and **3-7.**

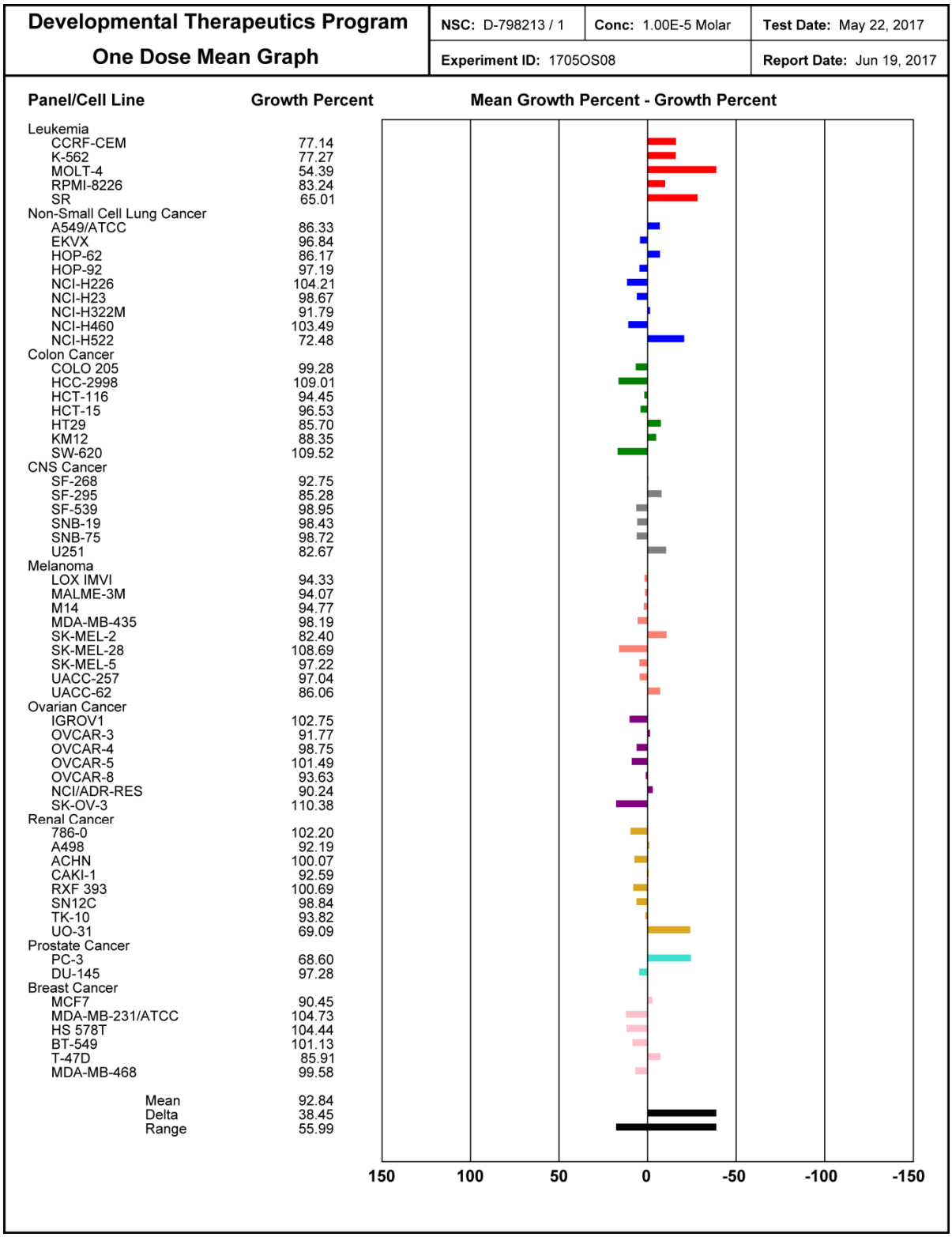


Figure S113 B. Compound 7 NCI-60 one dose mean graph**.**

APPENDEX C

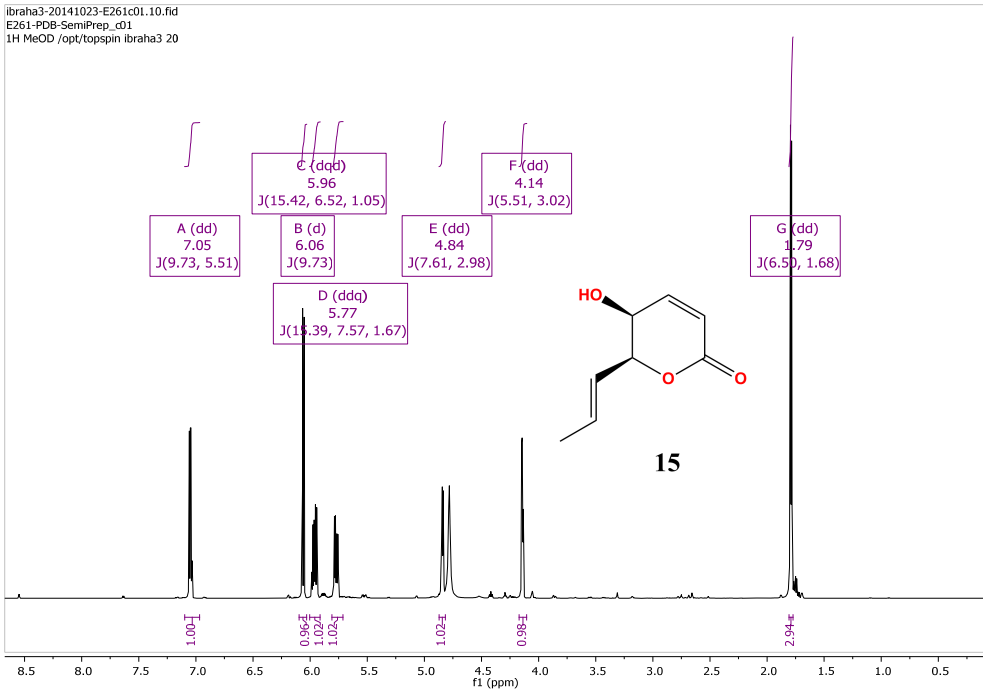


Figure S1 C. ^{1}H NMR spectrum (700 MHz,CD_3OD) of compound **15,** 5(S),6(S)-E-Dihydro-5-hydroxy-6-propenyl-2H-pyran-2-on.

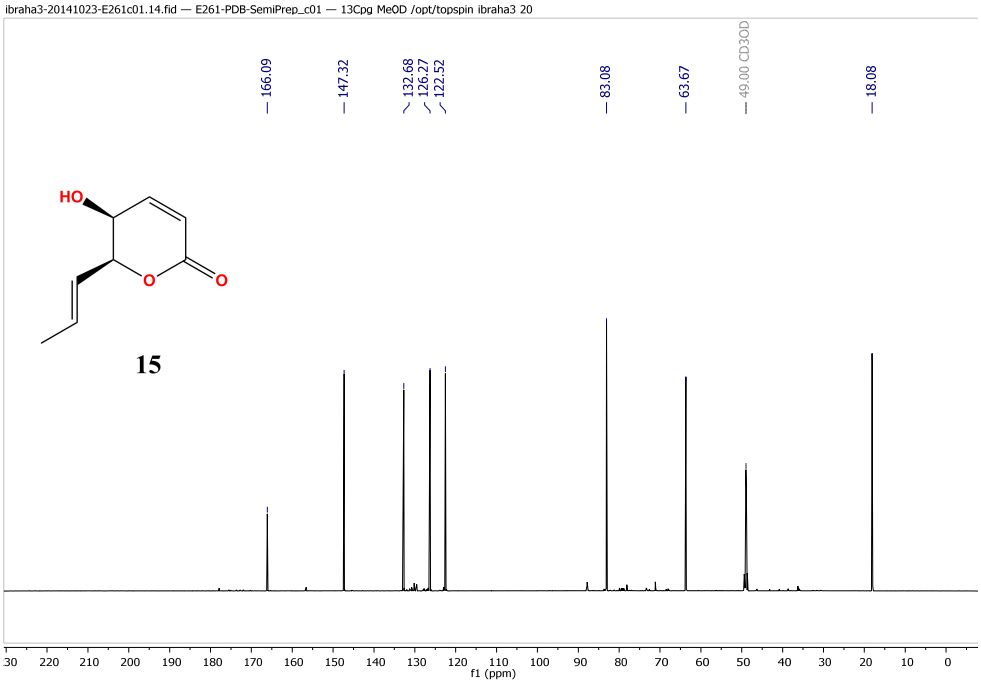


Figure S2 C. ^{13}C NMR spectrum (176 MHz,CD_3OD) of compound **15,** 5(S),6(S)-E-Dihydro-5-hydroxy-6-propenyl-2H-pyran-2-on.

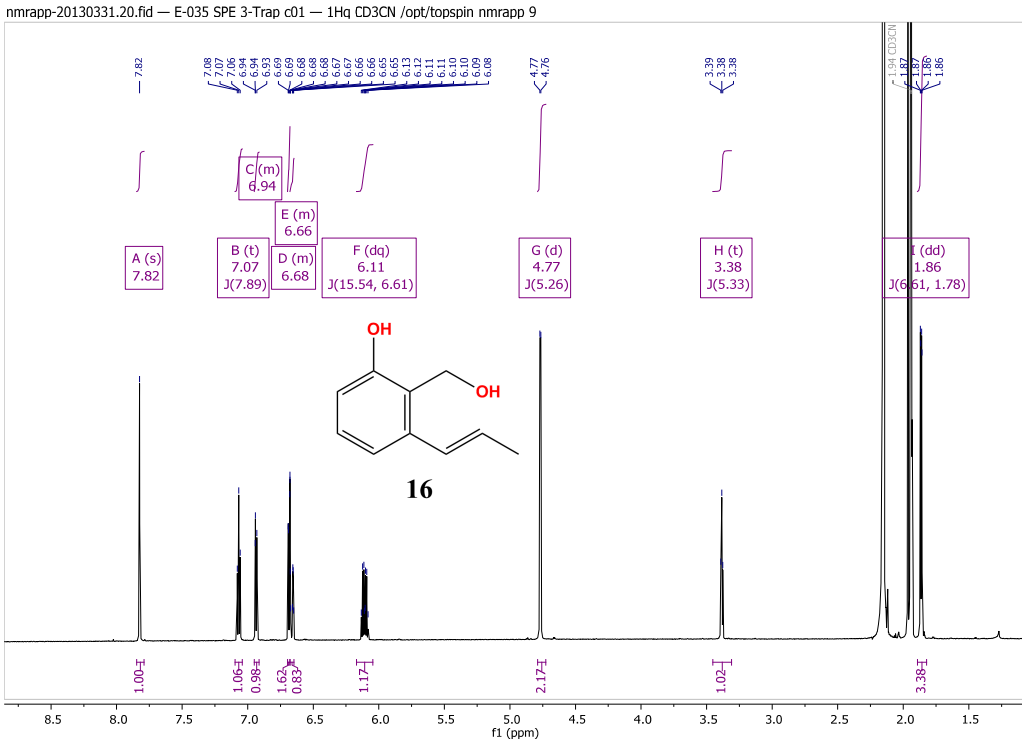


Figure S3 C . ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **16,** RKB-35646.

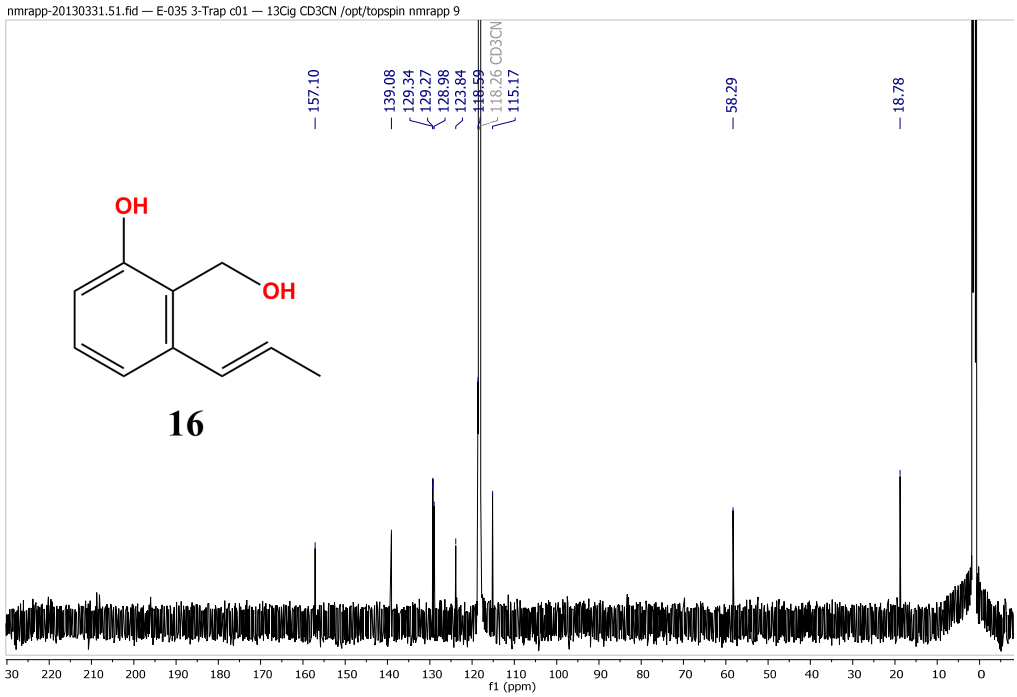


Figure S4 C. ^{13}C NMR spectrum (176 MHz, CD_3CN) of compound **16,** RKB-35646.

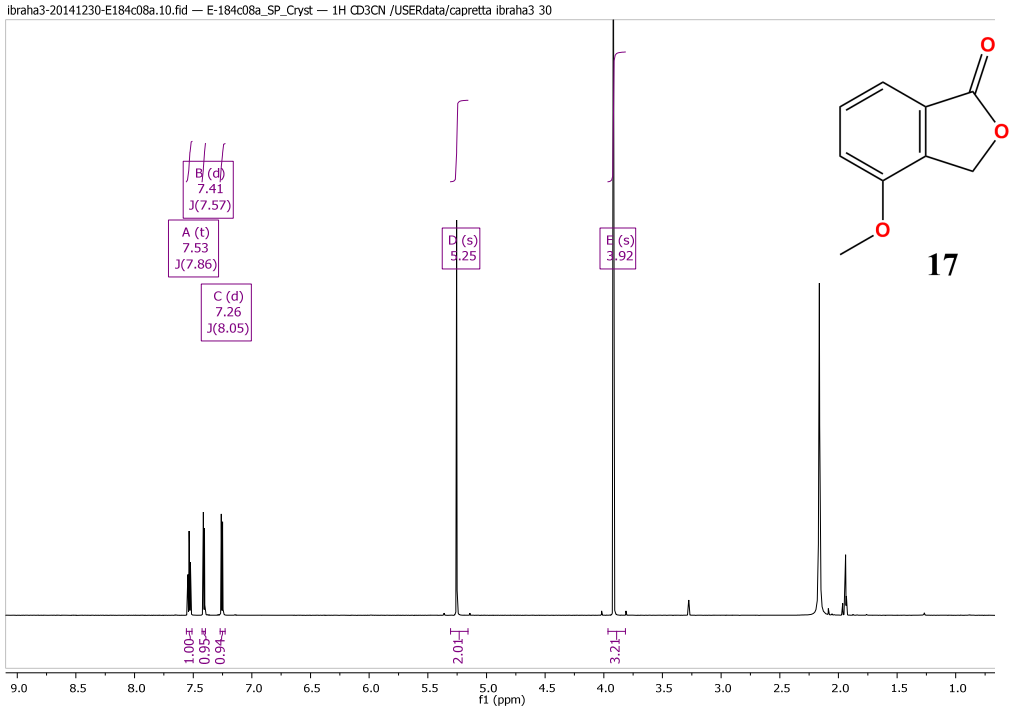


Figure S5 C. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **17**, 4-methoxy-isobenzofuran-1(3H)-one.

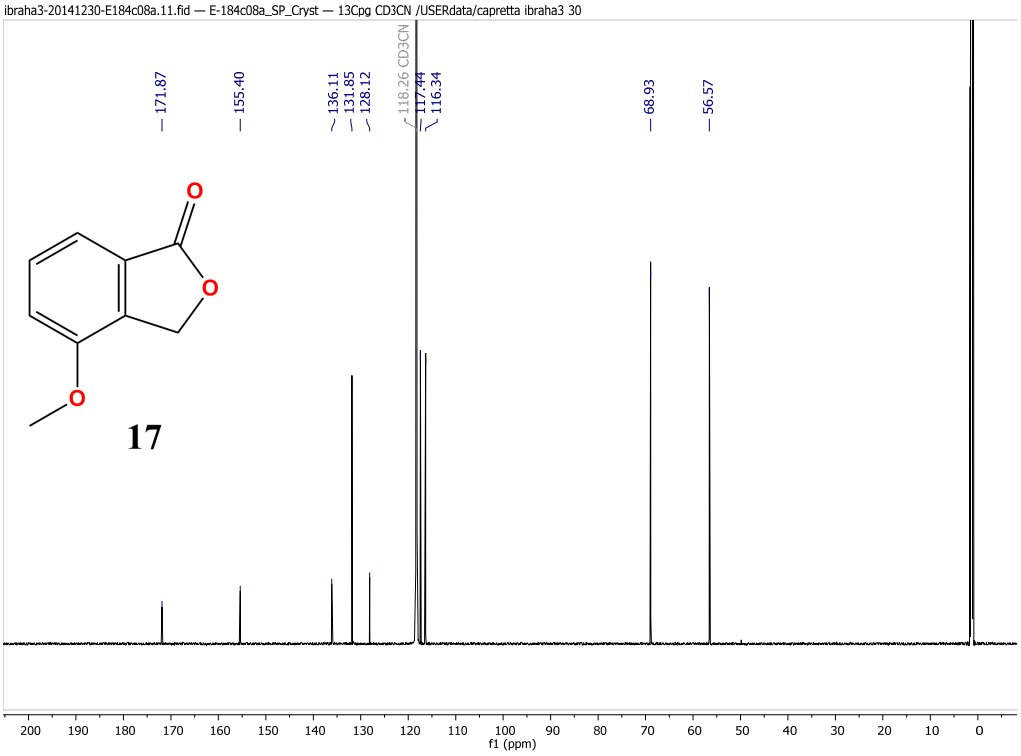


Figure S6 C. ^{13}C NMR spectrum (176 MHz, CD_3CN) of compound **17**, 4-methoxy-isobenzofuran-1(3H)-one.

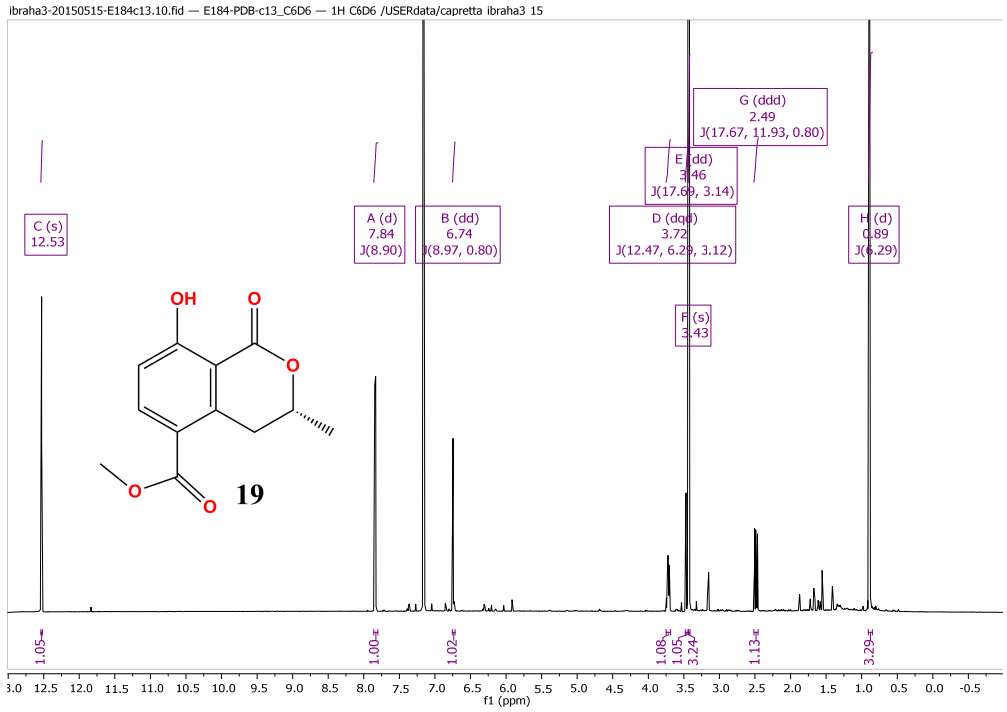


Figure S7 C**.** ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **19**, 5-methoxy-carbonylmellein.

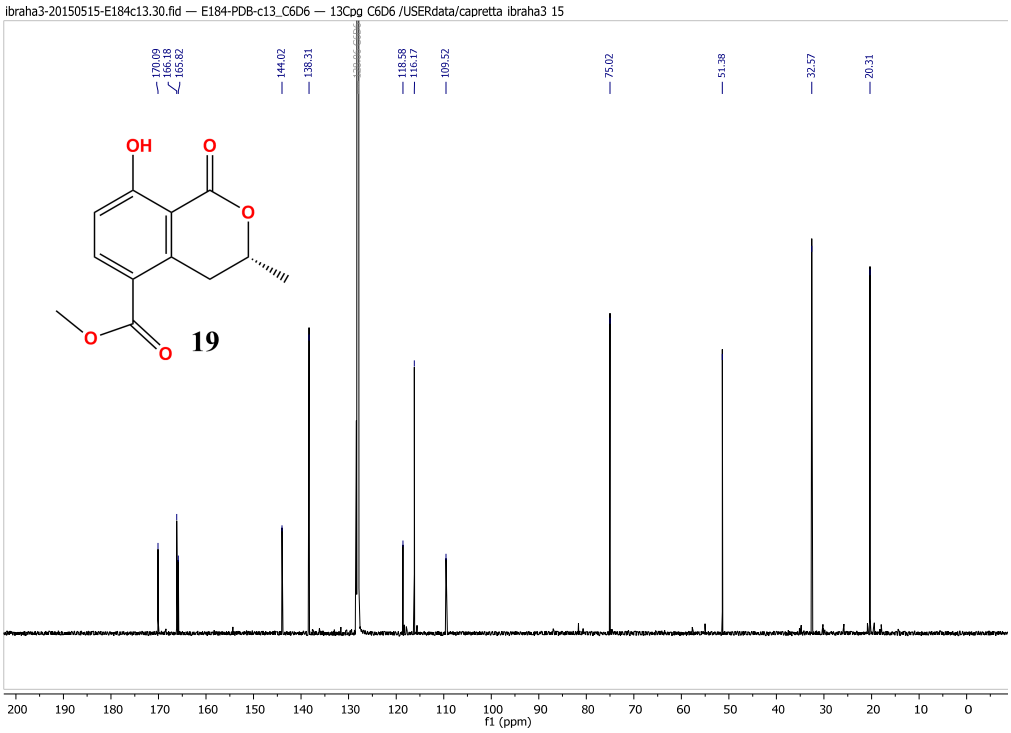


Figure S8 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **19,** 5-methoxy-carbonylmellein.

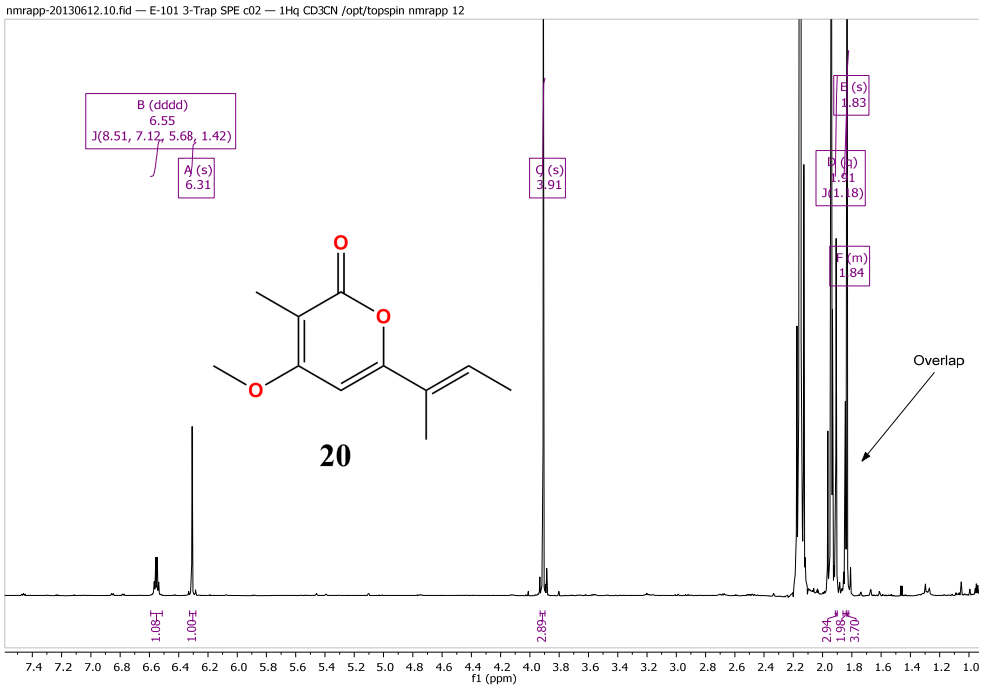


Figure S9 C**.** ^{1}H NMR spectrum (700 MHz) of compound **20**, nectriapyrone.

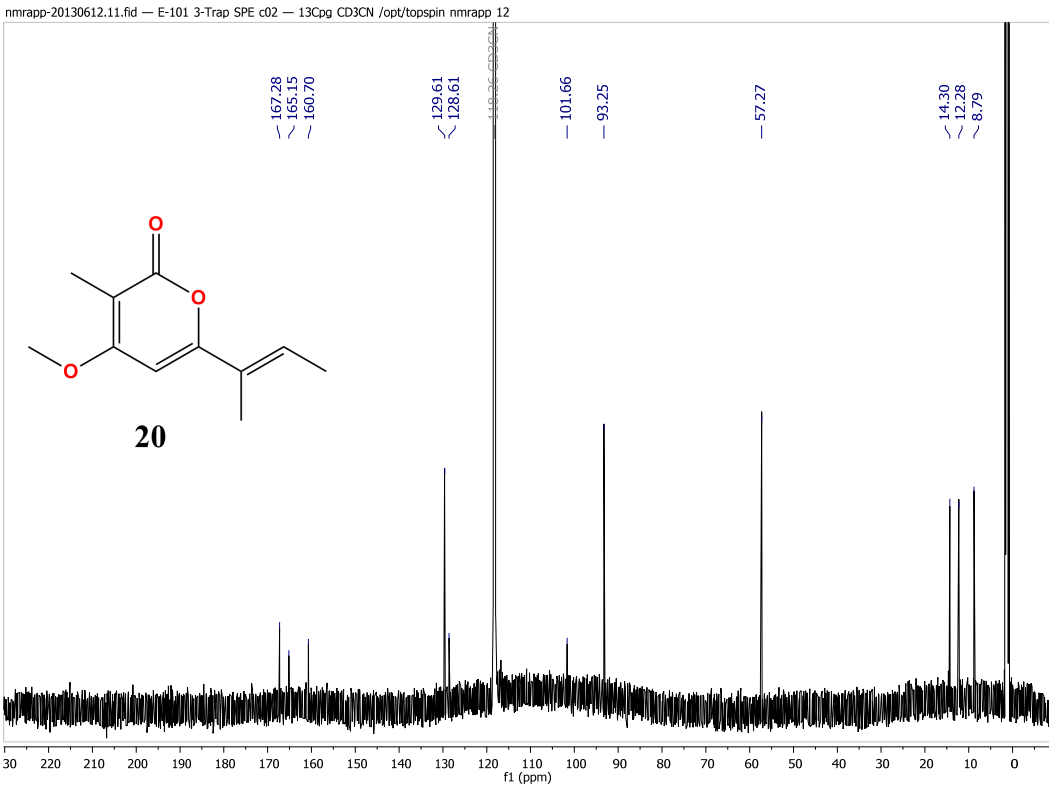


Figure S10 C. ^{13}C NMR spectrum (176 MHz) of compound **20,** nectriapyrone.

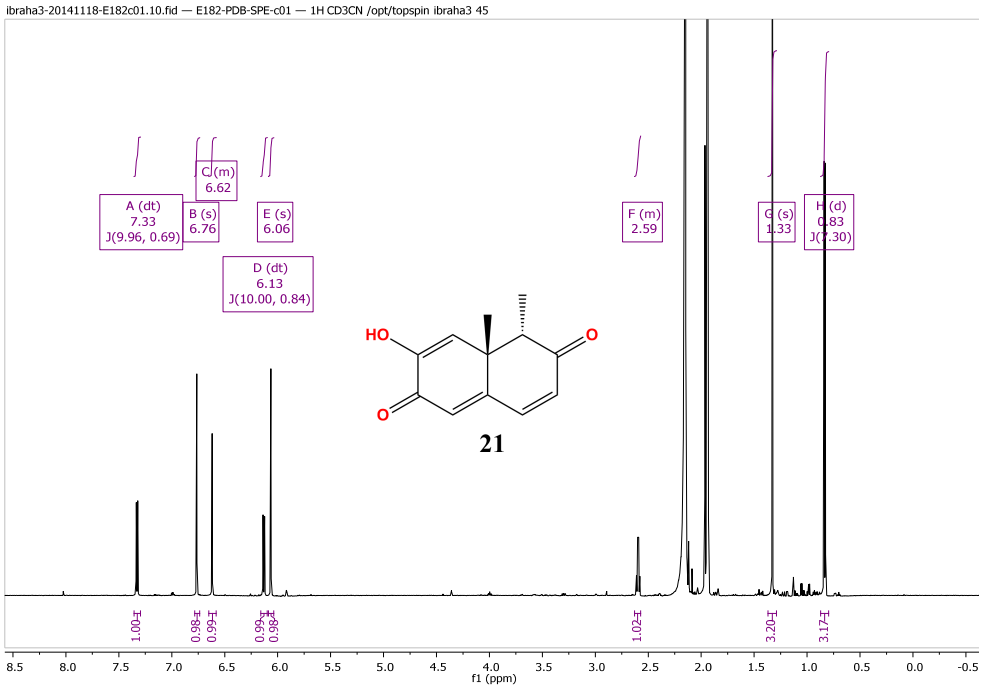


Figure S11 C. ^{1}H NMR spectrum (700 MHz,CD_3CN) of compound **21**, botryosphaeridione.

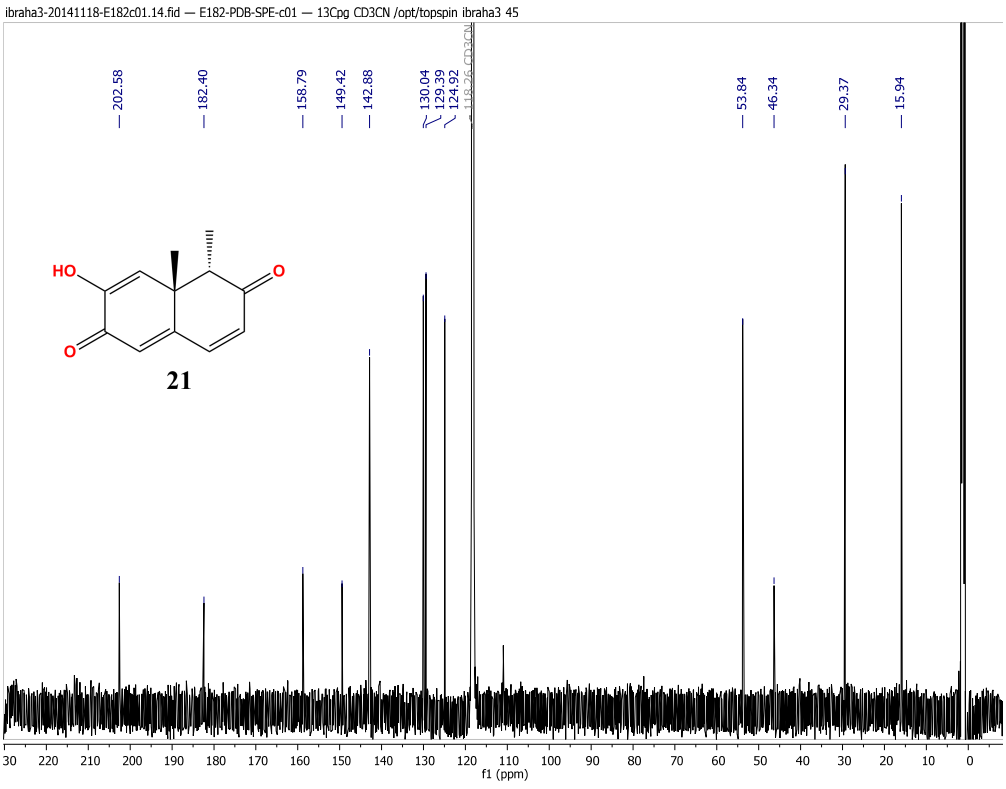


Figure S12 C. ^{13}C NMR spectrum (176 MHz,CD_3CN) of compound **21,** botryosphaeridione.

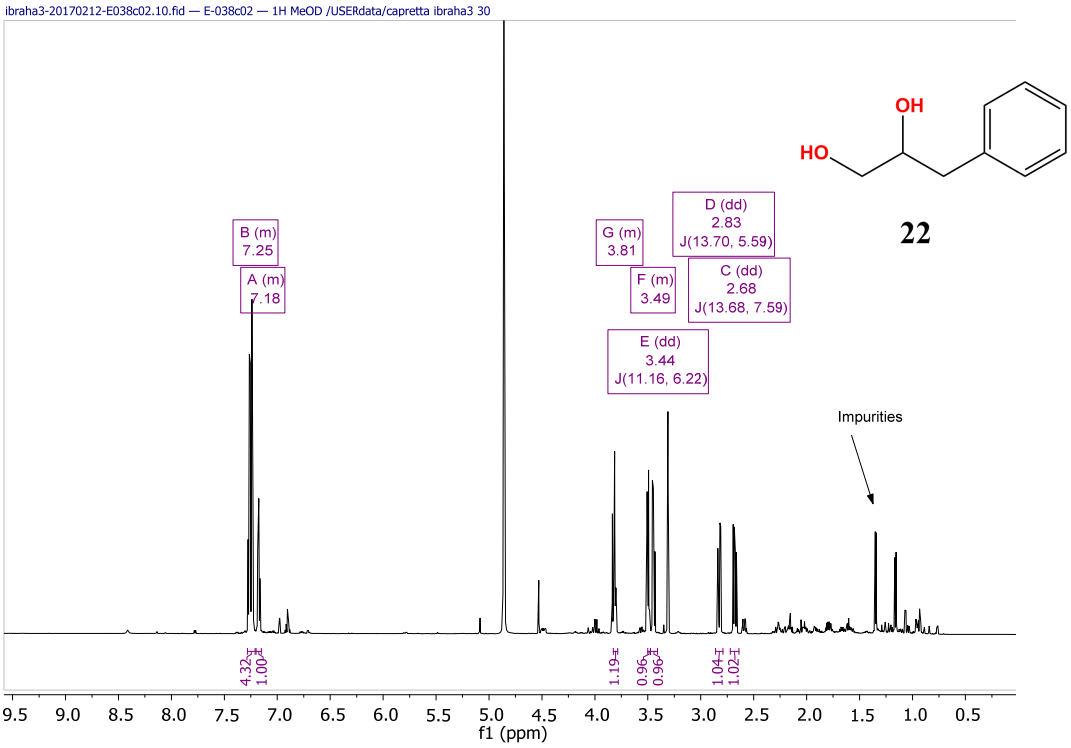


Figure S13 C**.** ^{1}H NMR spectrum (700 MHz,CD_3OD) of compound **22**, 3-phenylpropane-1,2-diol.

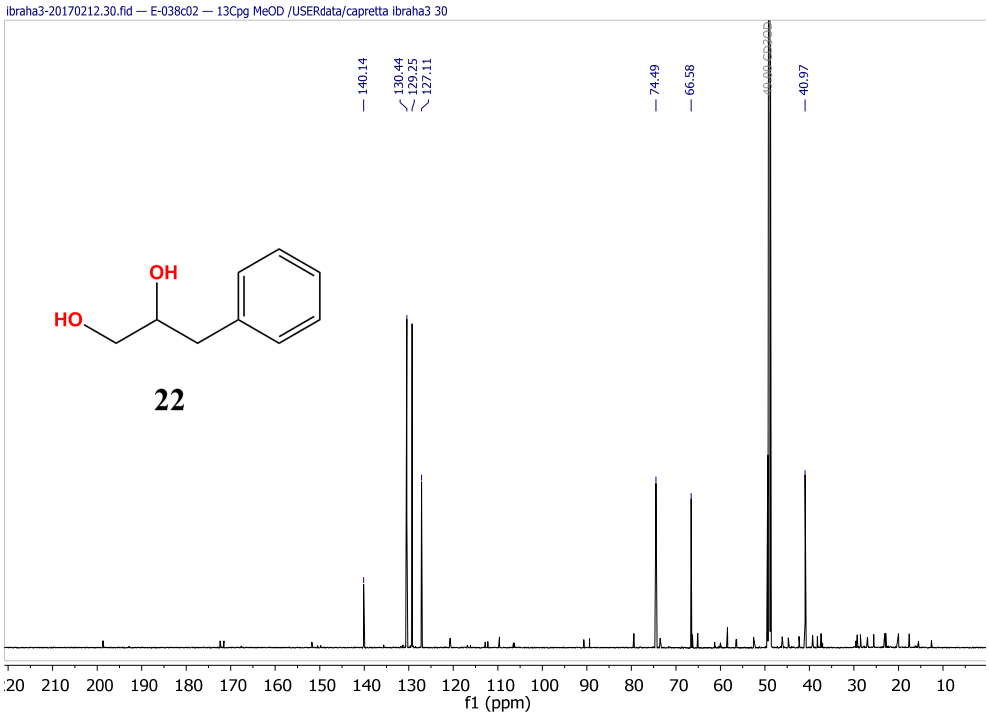


Figure S14 C. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **22,** 3-phenylpropane-1,2-diol.

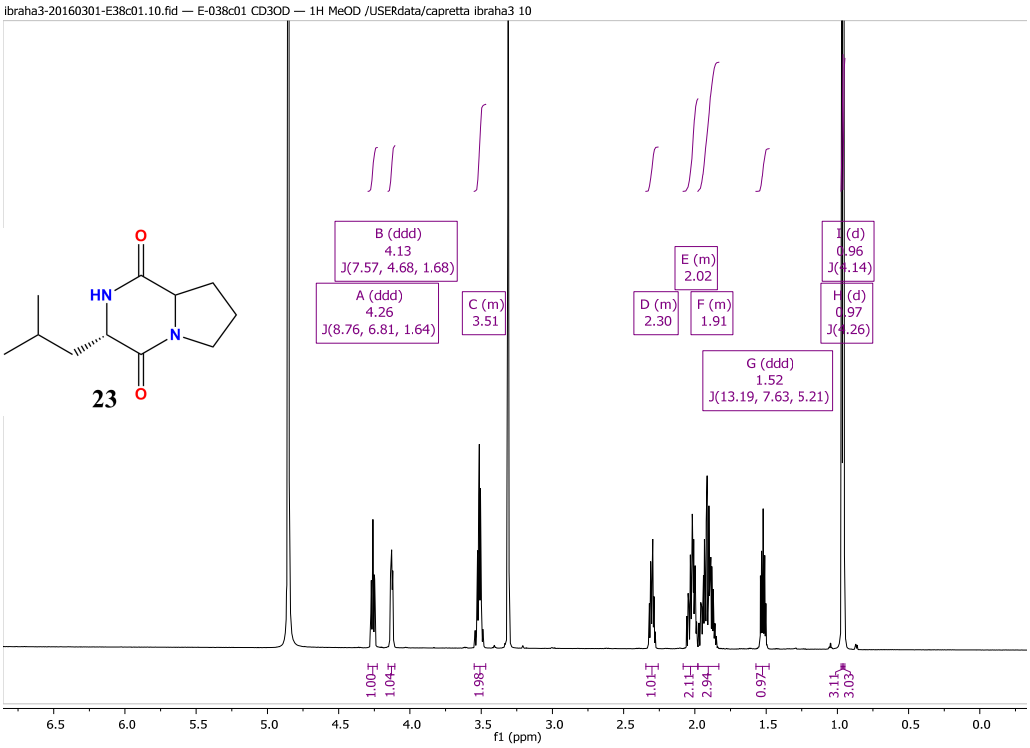


Figure S15 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **23**, gancidin W.

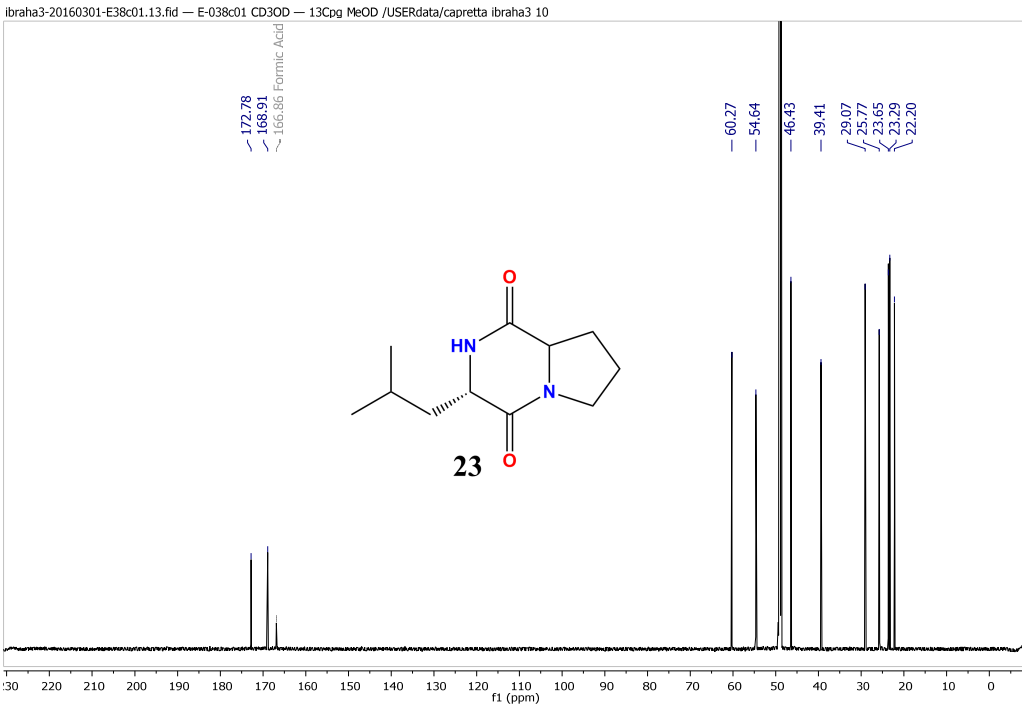


Figure S16 C. ^{13}C NMR spectrum (176 MHz,CD_3OD) of compound **23,** gancidin W.

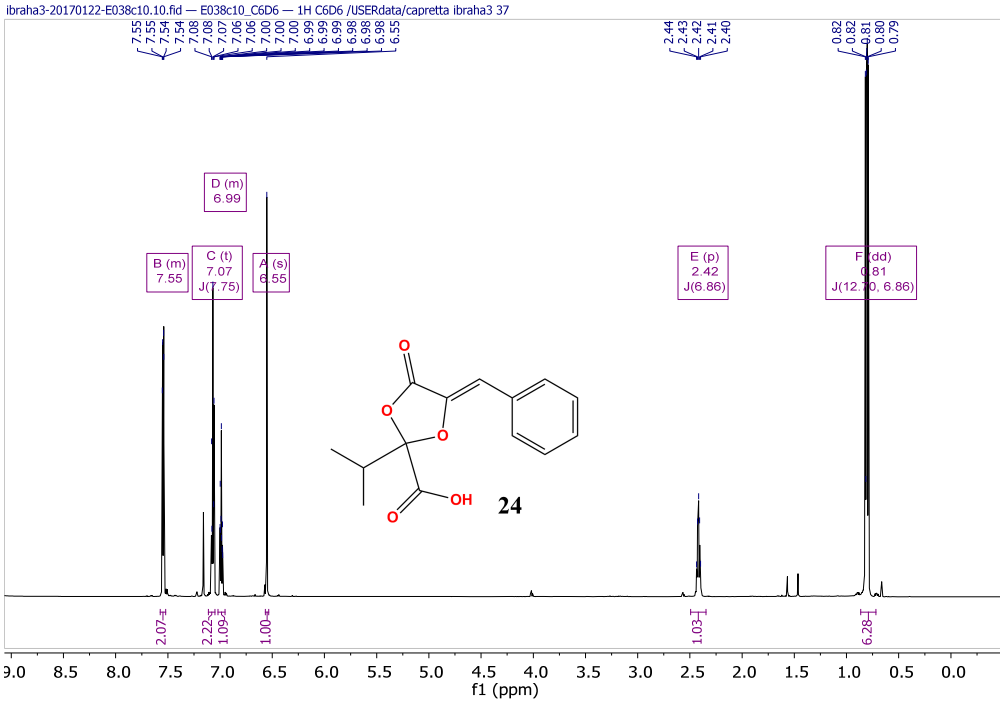


Figure S17 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **24**, guignardic acid.

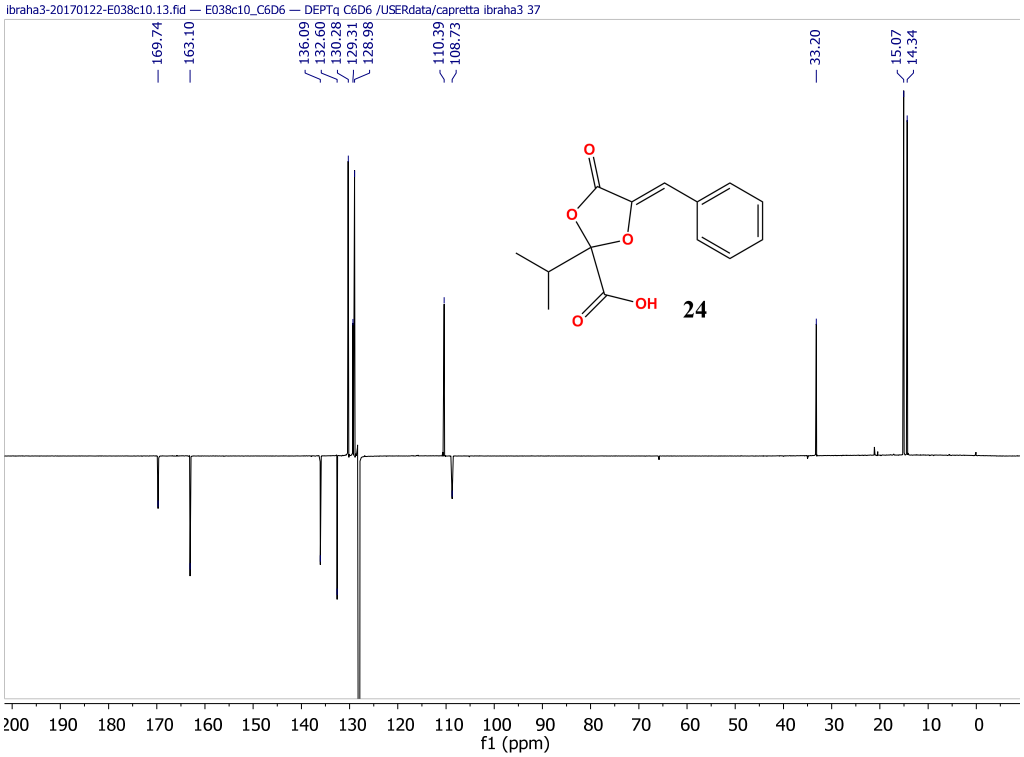


Figure S18 C. ^{13}C DEPTq NMR spectrum (176 MHz, C_6D_6) of compound **24,** guignardic acid.

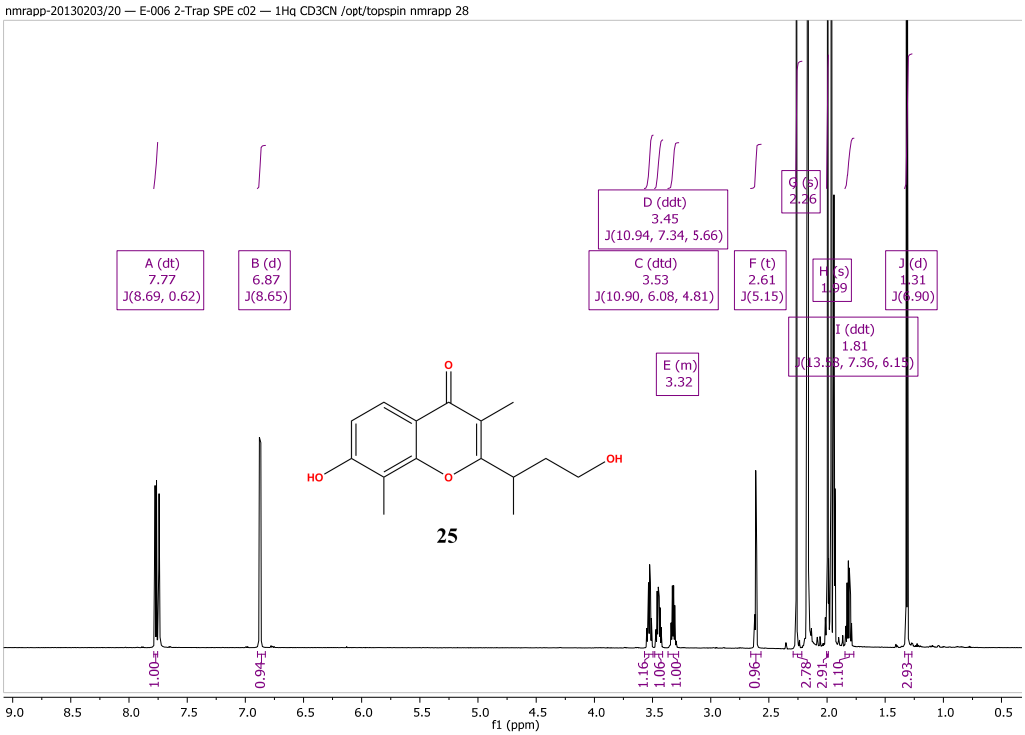


Figure S19 C. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **25**, lachnochromonin A

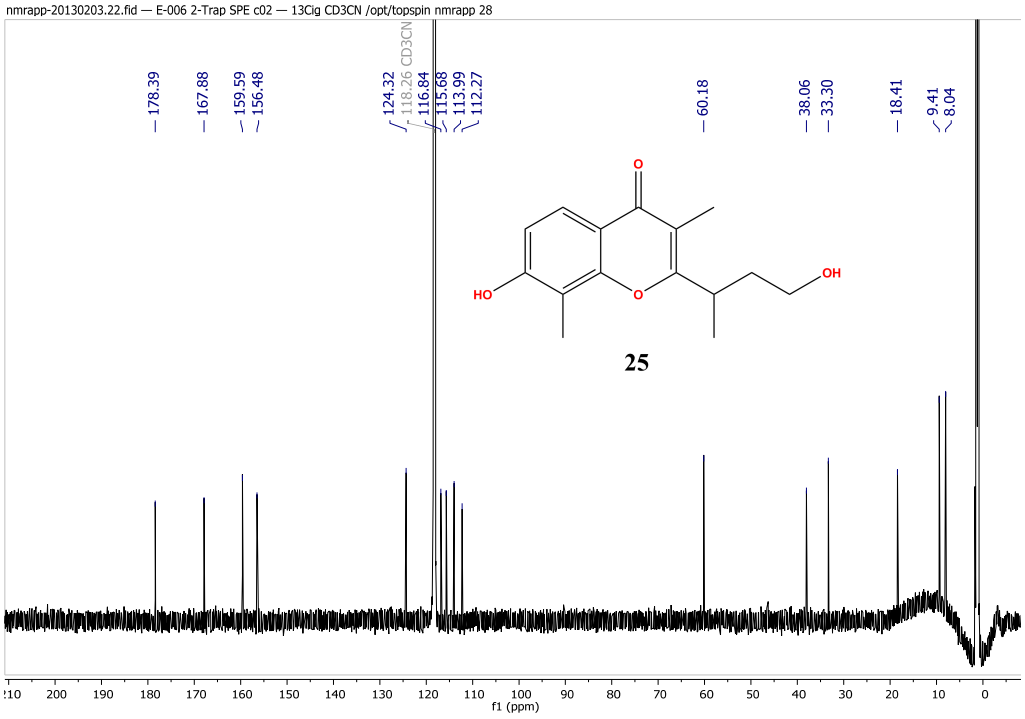
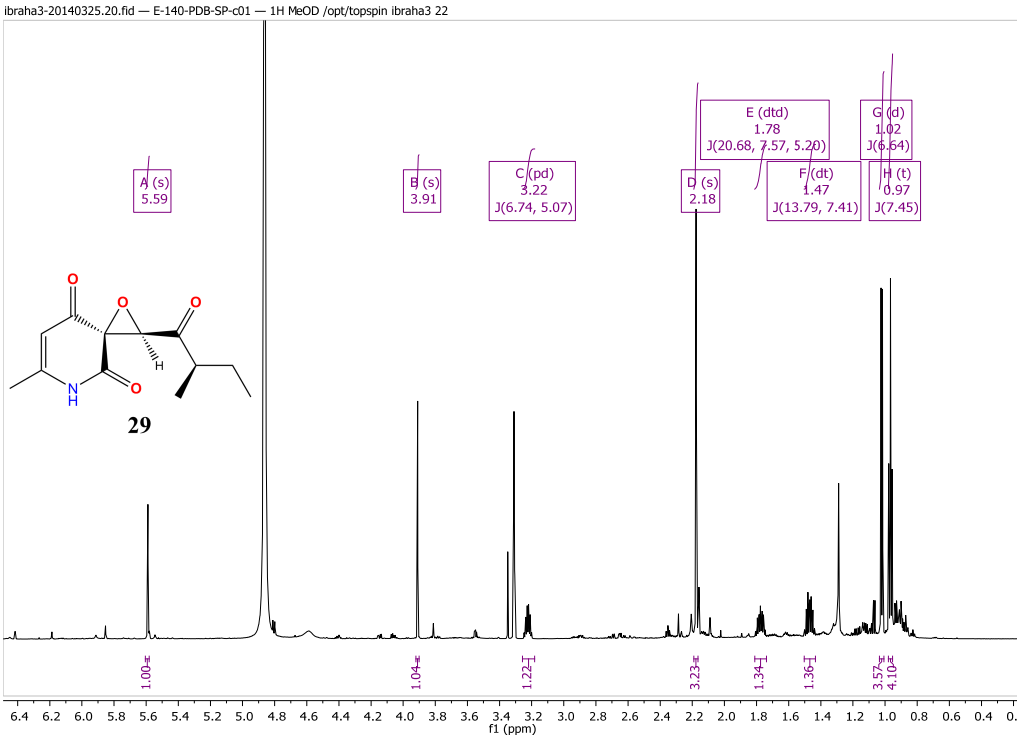
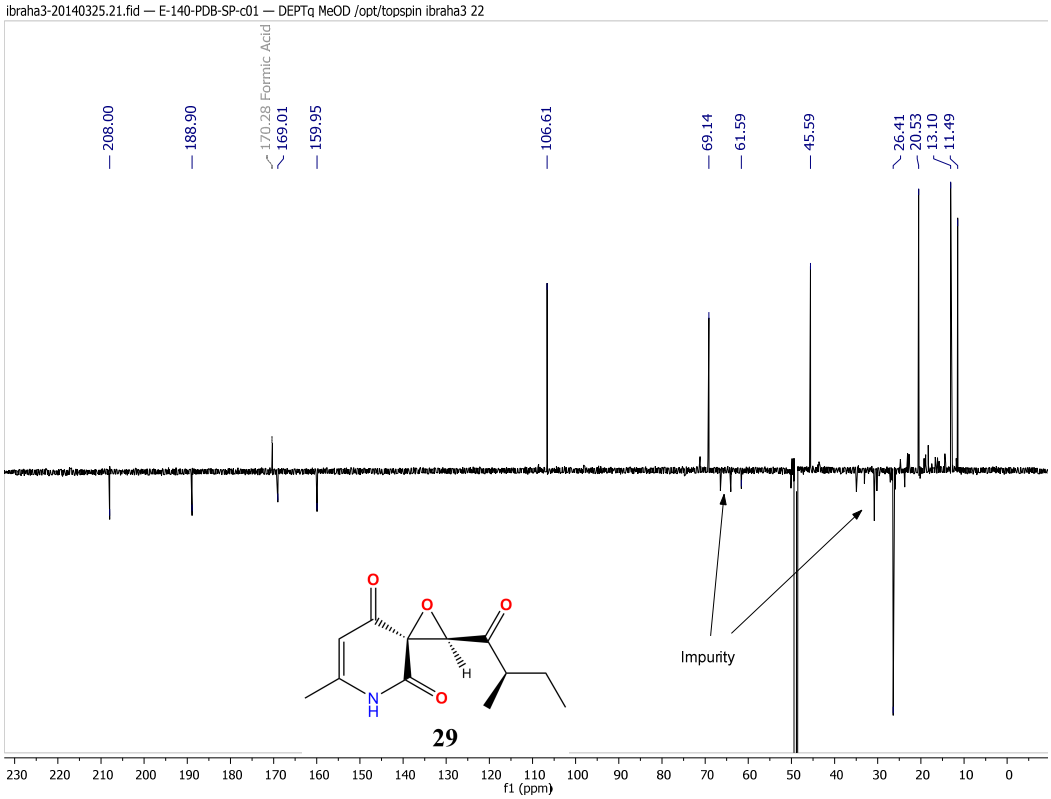


Figure S20 C. ^{13}C NMR spectrum (176 MHz, CD_3CN) of compound **25,** lachnochromonin A.



29

Figure S21 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **29**, fruit rot toxin A.



29

Figure S22 C. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **29,** fruit rot toxin A.

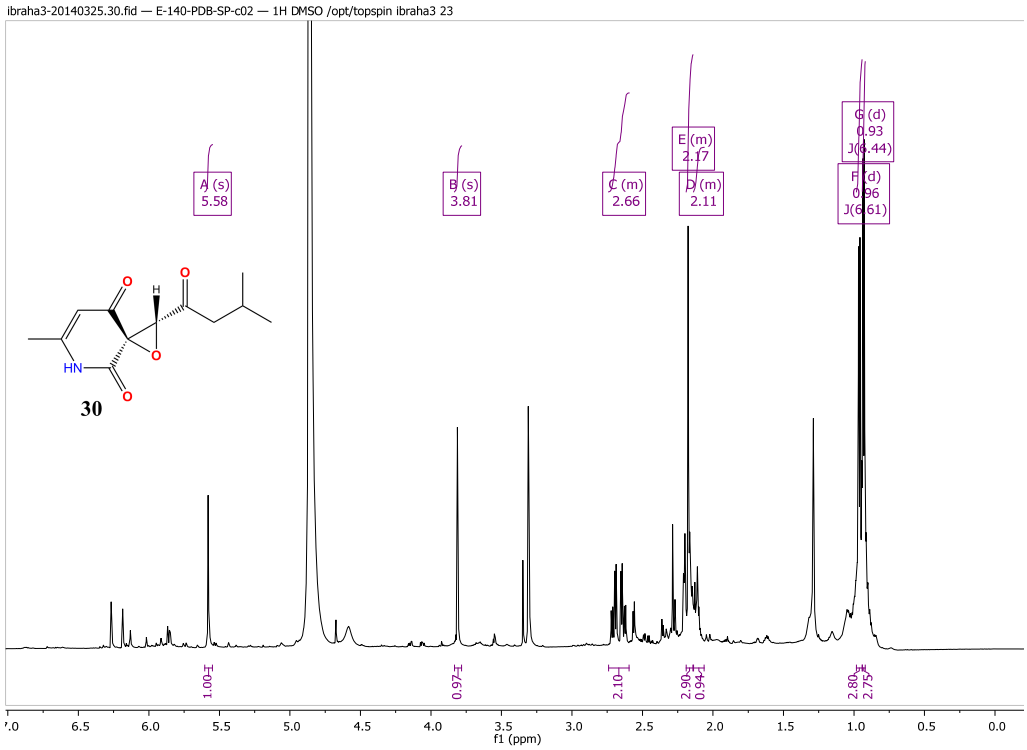


Figure S23 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **30**, fruit rot toxin B.

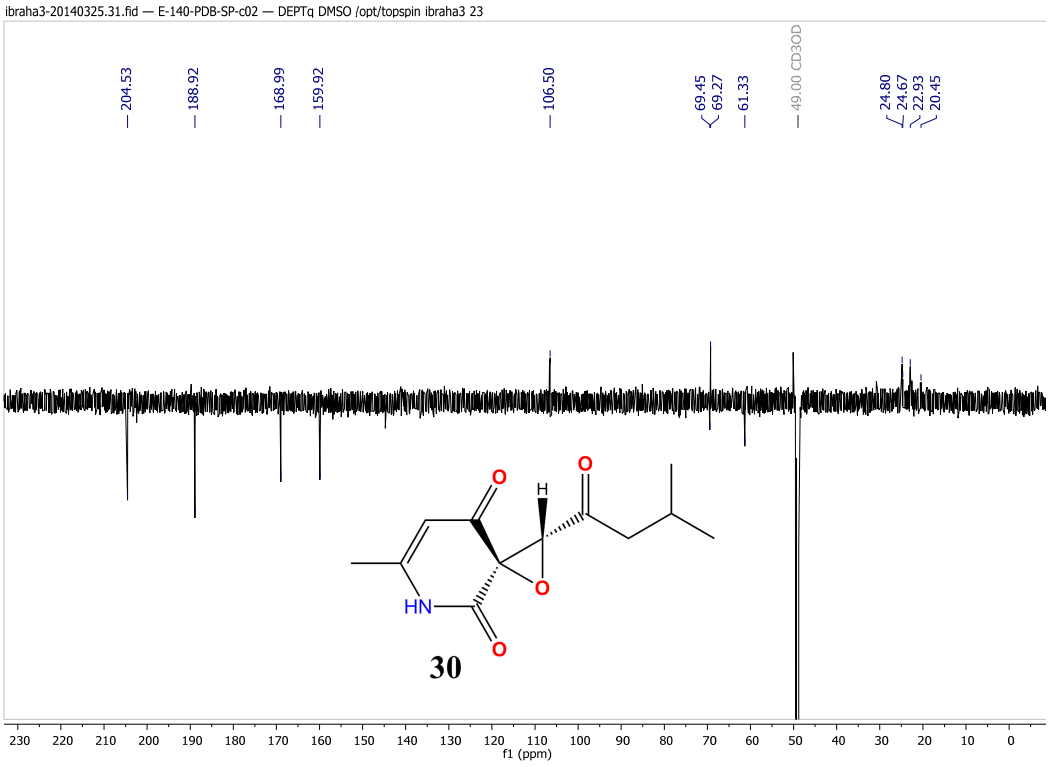


Figure S24 C. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **30,** fruit rot toxin B.

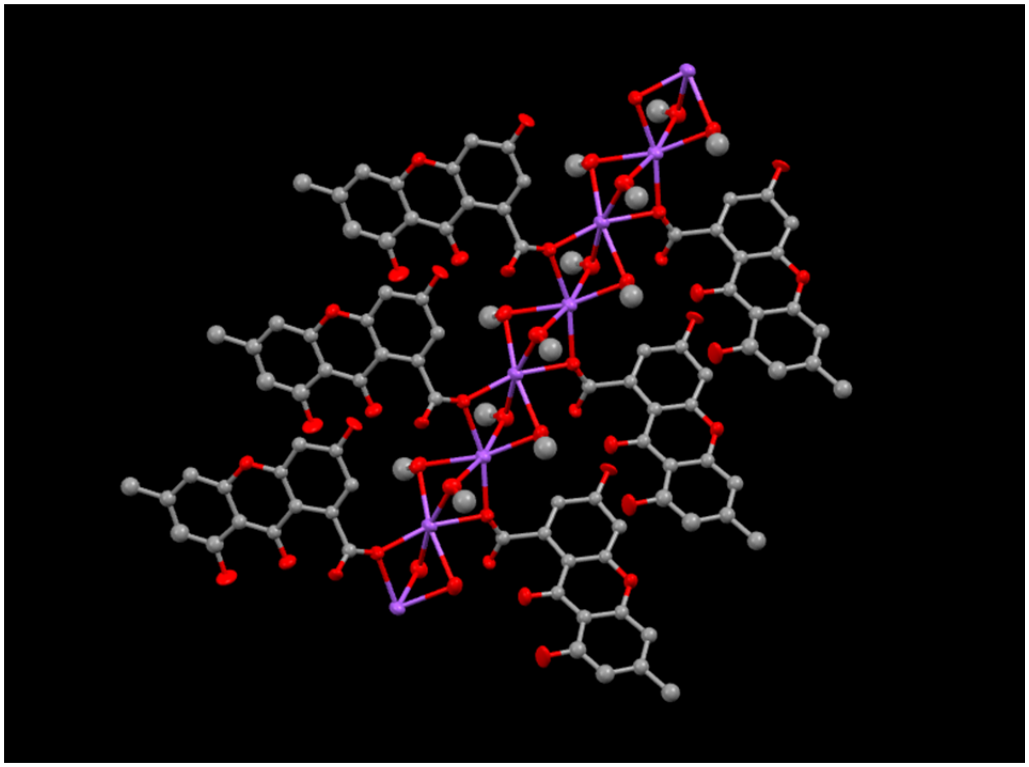
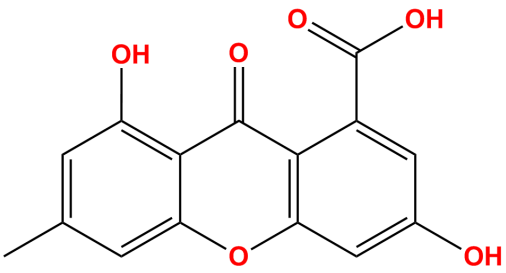


Figure S25 C. Network X-ray crystal structure of compound **31**, networked through a sodium/methanol backbone; 3,8-dihydroxy-6-methyl-9-oxo-9H-xanthene-1-carboxylic acid. Compound 31 was crystalized from methanol (300 uL) by slow evaporation in a 3mm NMR tube and was co-eluting with compound **3** as a minor impurity, front-shoulder peak.



31

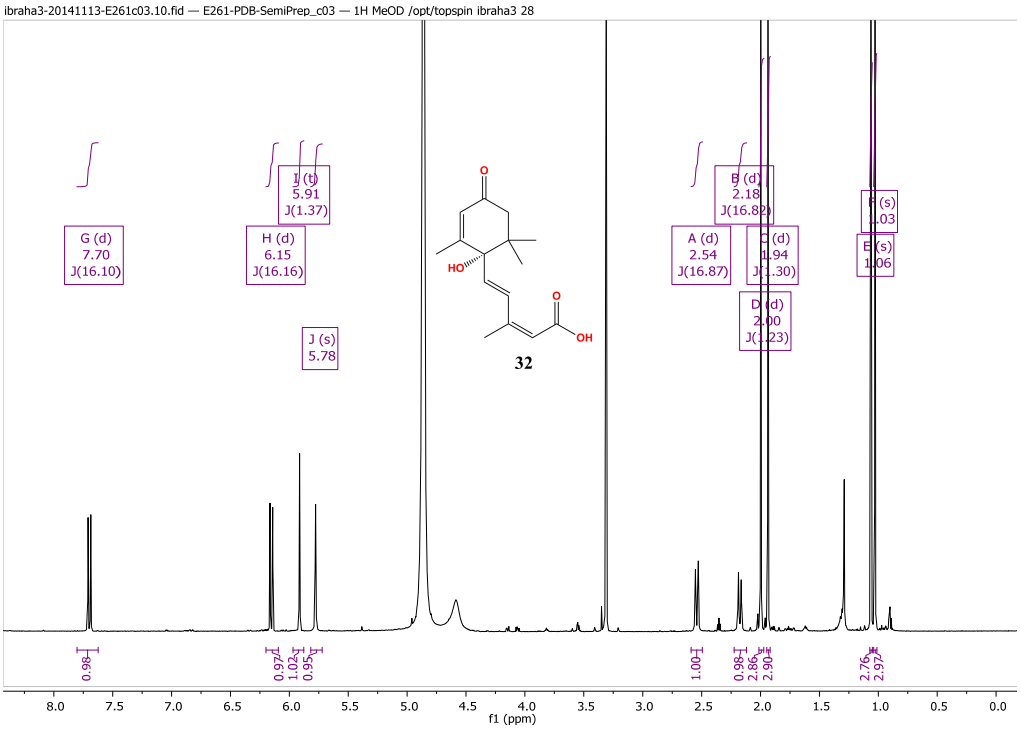


Figure S26 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **32**, abscisic acid.

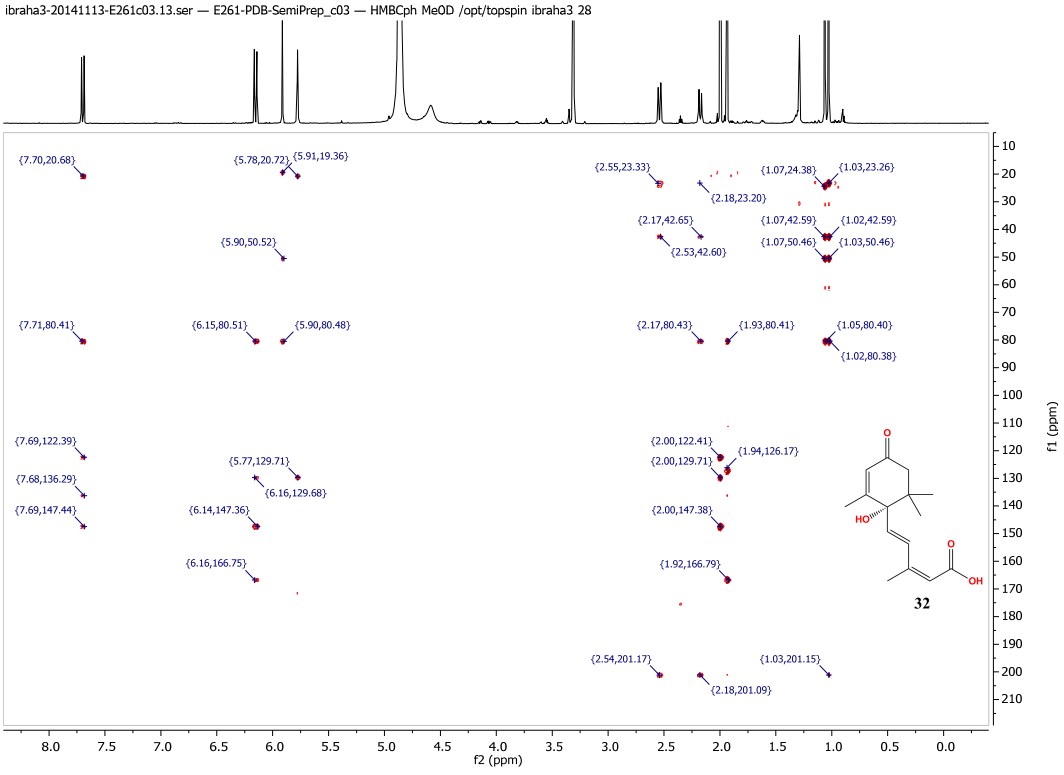


Figure S27 C. HMBC NMR spectrum (700 MHz, CD_3OD) of compound **32,** abscisic acid.

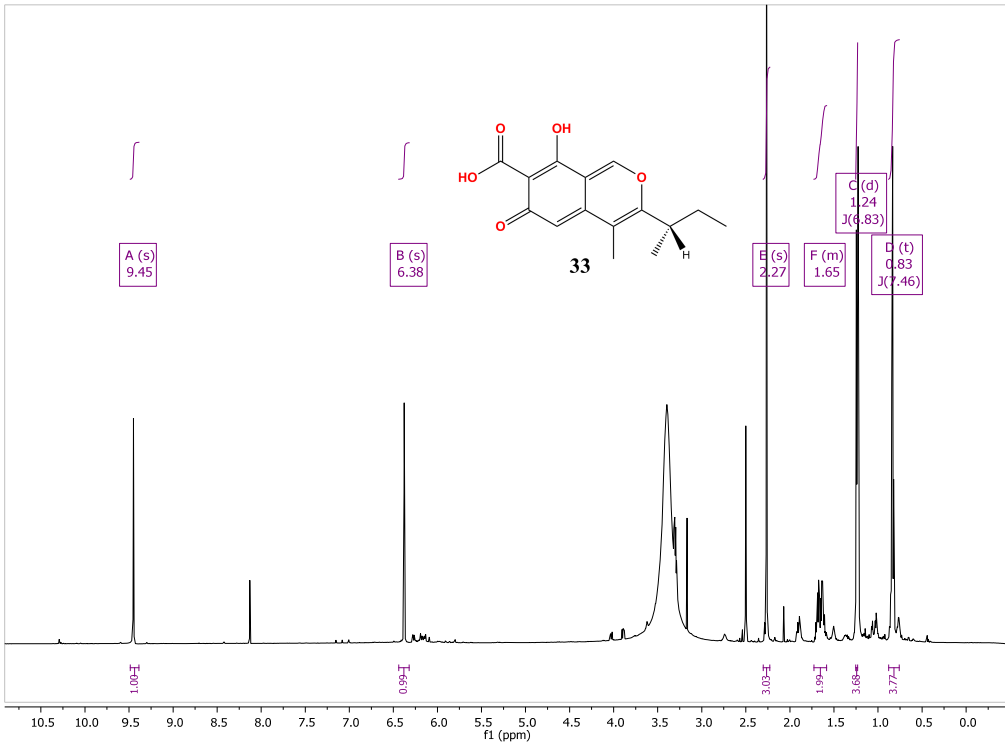


Figure S28 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **33**, ascochitine.

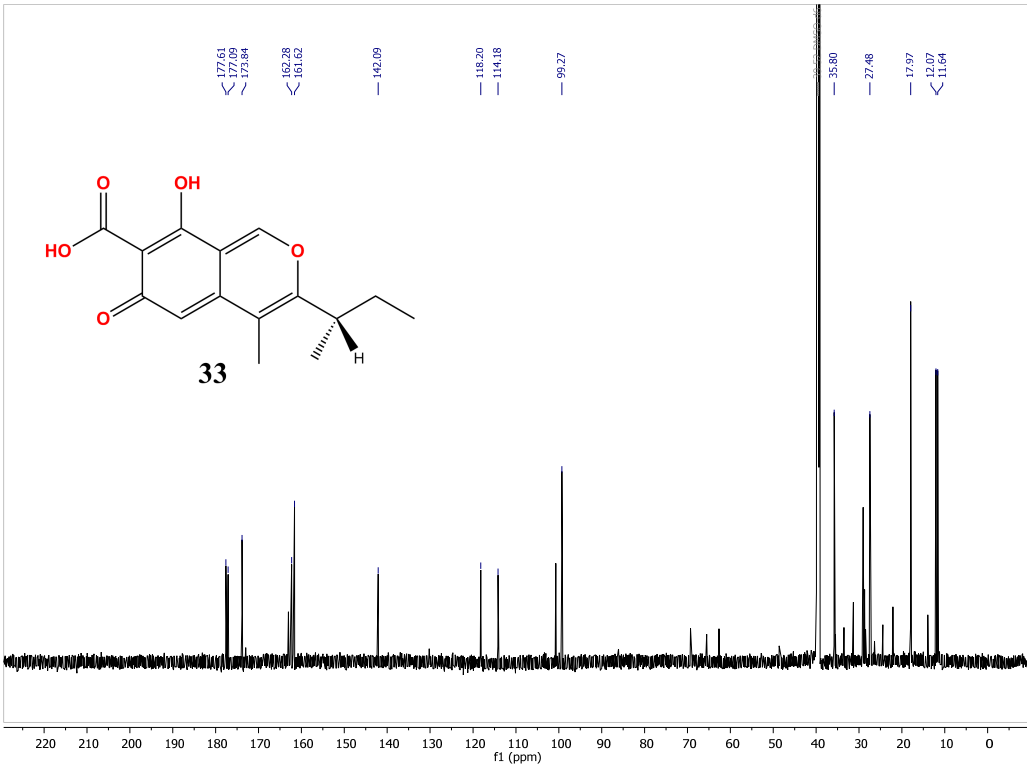


Figure S29 C. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **33,** ascochitine.

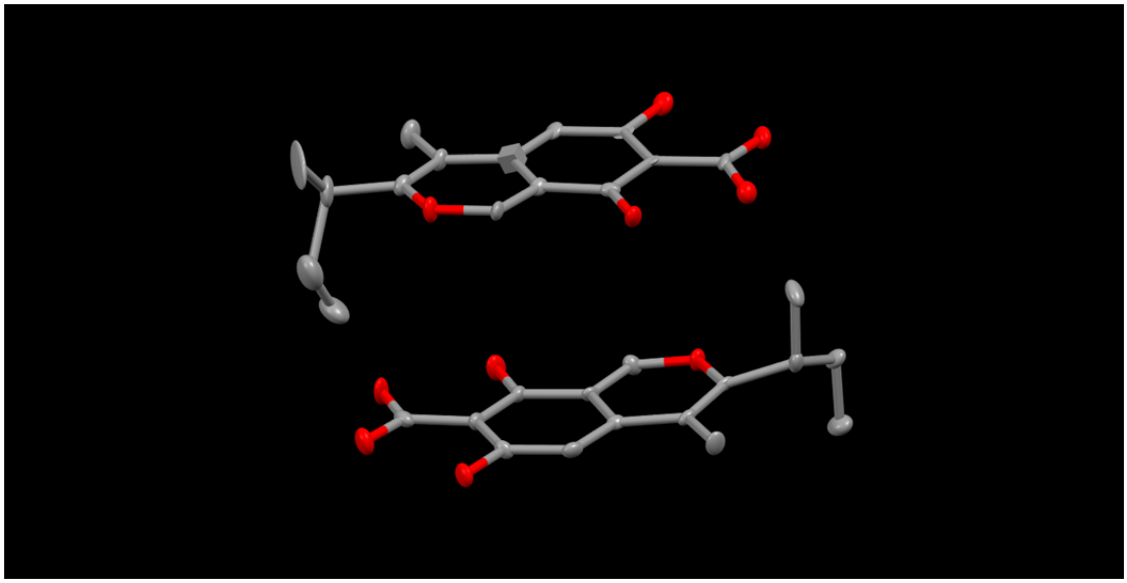
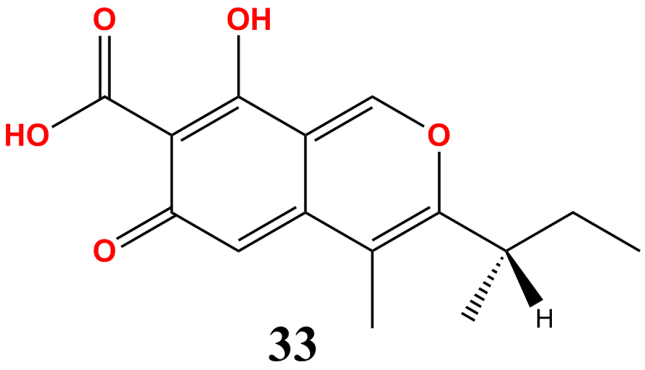


Figure S30 C. Single crystal X-ray structure of compound **33,** ascochitine, racemic mixture (1:1).



33

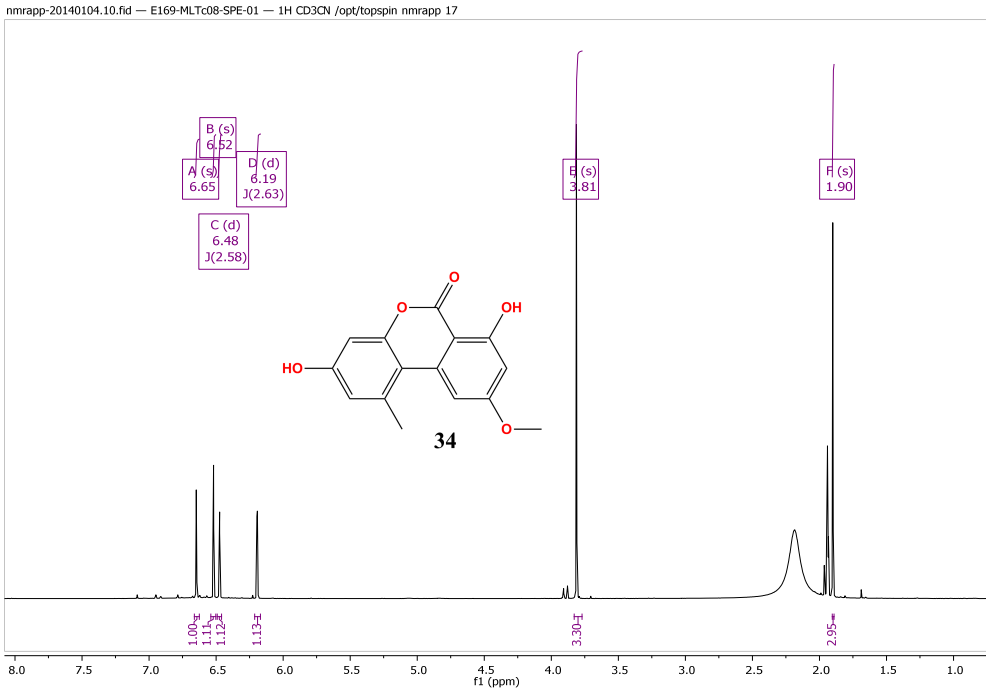


Figure S31 C. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **34,** alternariol monomethylether.

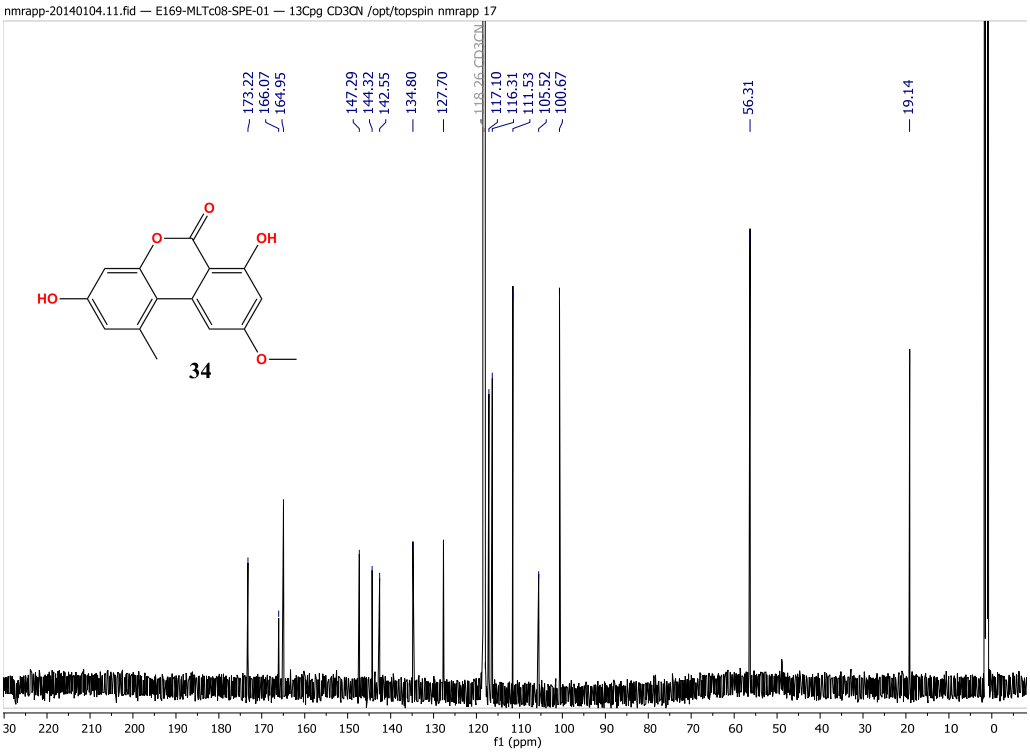


Figure S32 C. ^{13}C NMR spectrum (176 MHz, CD_3CN) of compound **34,** alternariol monomethylether.

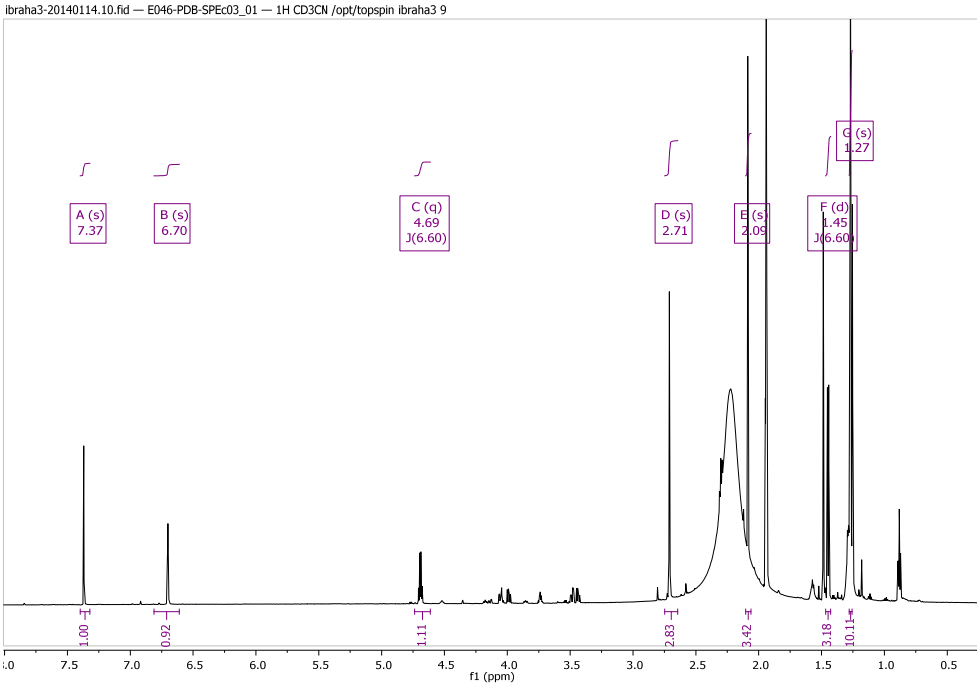
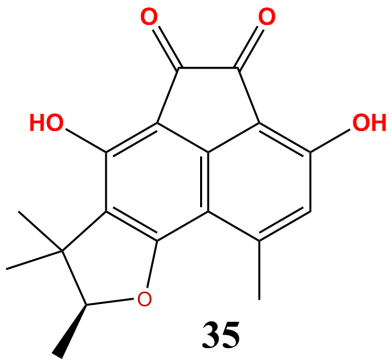


Figure S33 C. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **35,** sclerodine.



35

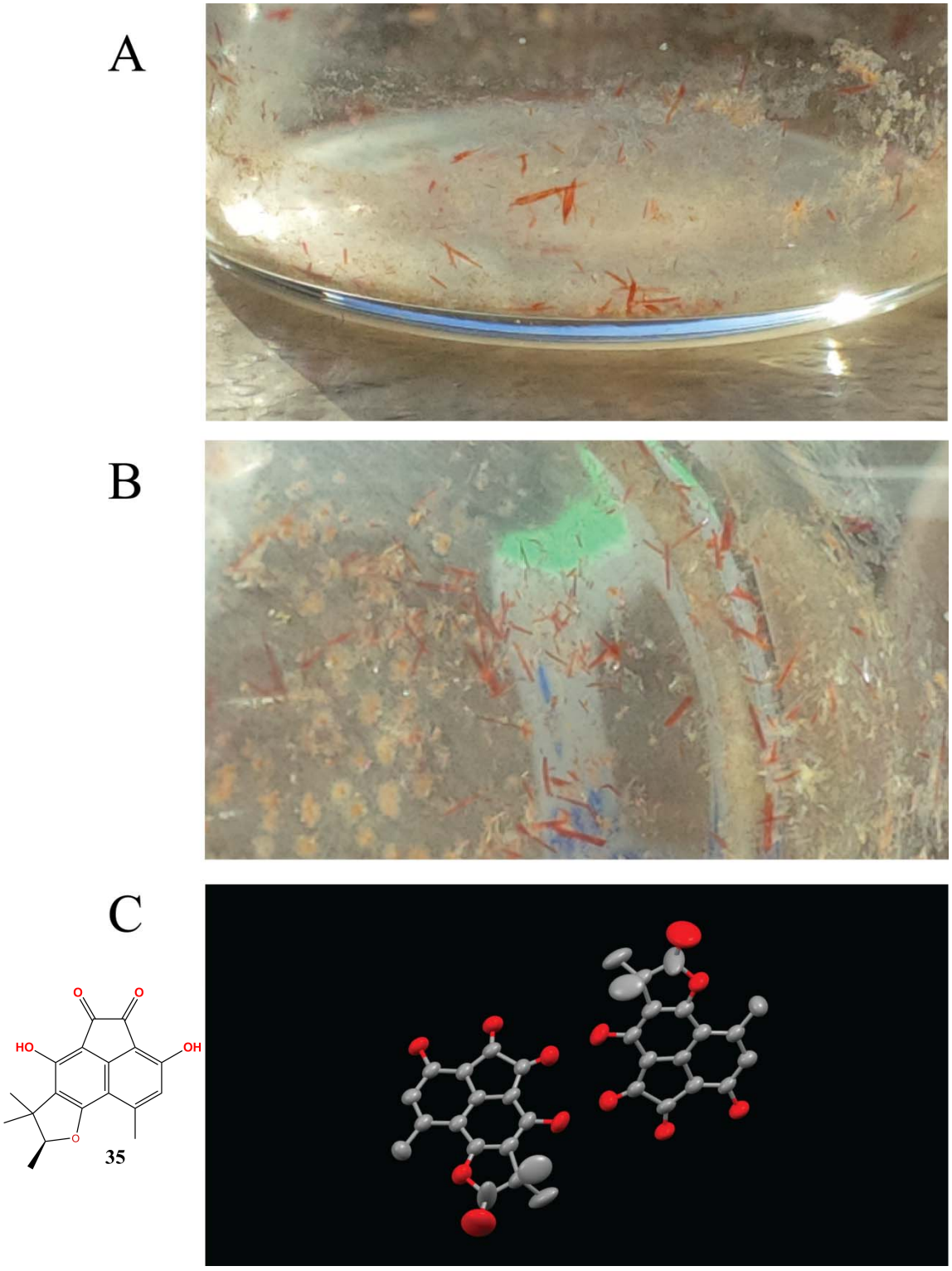


Figure S34 C**.** (A-B) Crystallization of compound **35,** sclerodine, from methanol via slow evaporation in a 20 mL scintillation vial; red crystals are visible. (C) Single crystal X-ray structure of sclerodine, with one of the methyl groups on the furan ring miss-identified as an oxygen moiety.

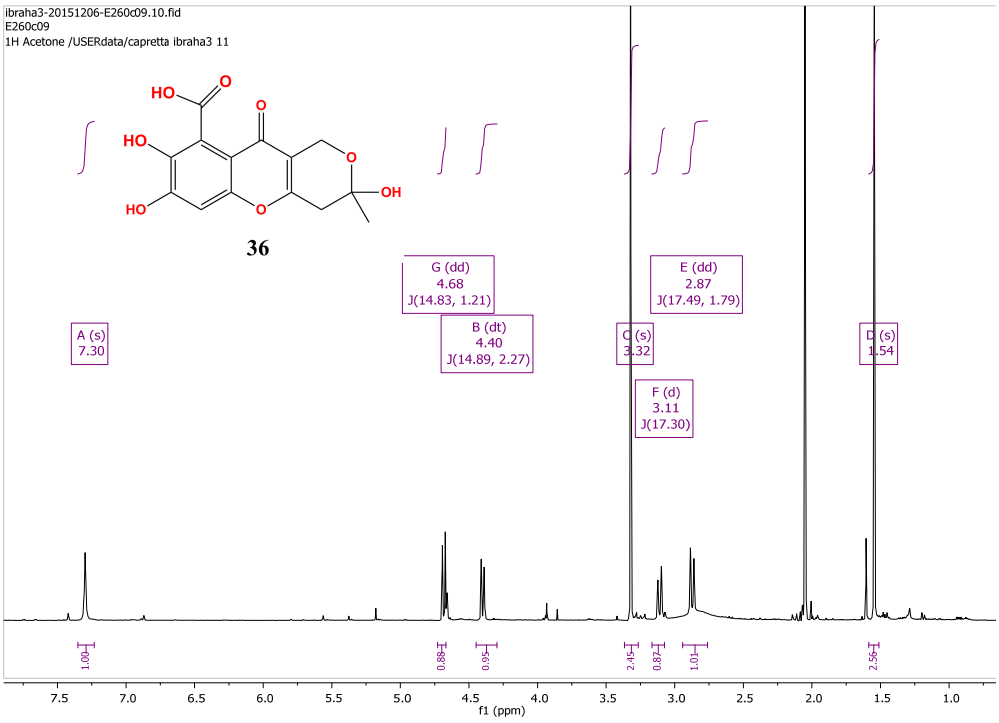


Figure S35 C. ^{1}H NMR spectrum (700 MHz, Acetone) of compound **36**, fulvic acid**.**

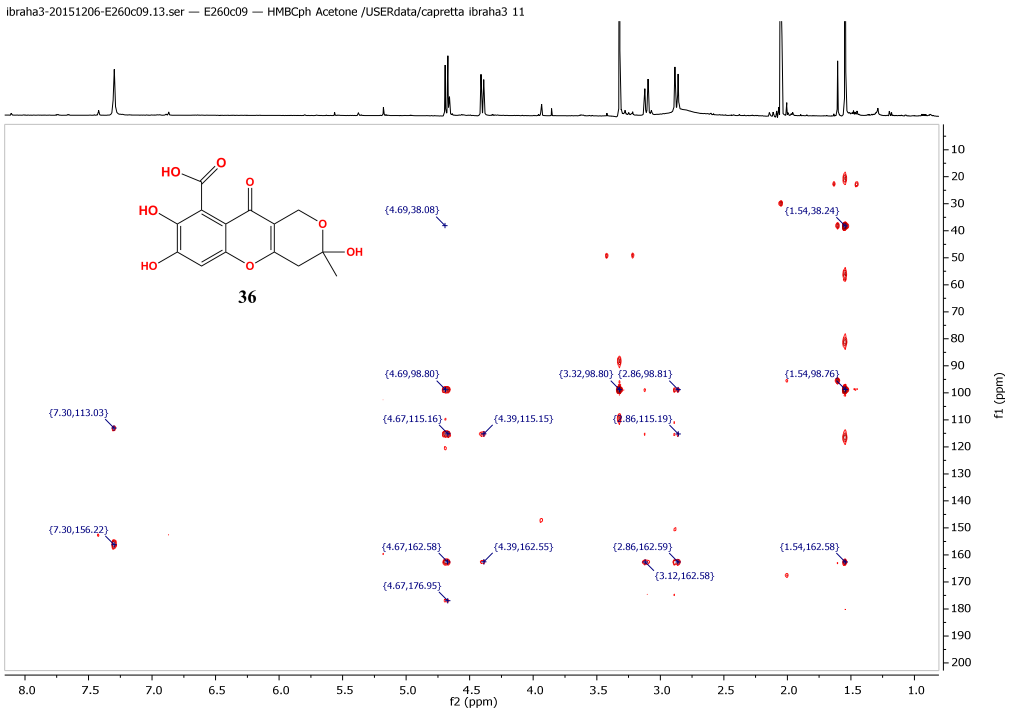


Figure S36 C. HMBC NMR spectrum (176 MHz, Acetone) of compound **36**, fulvic acid.

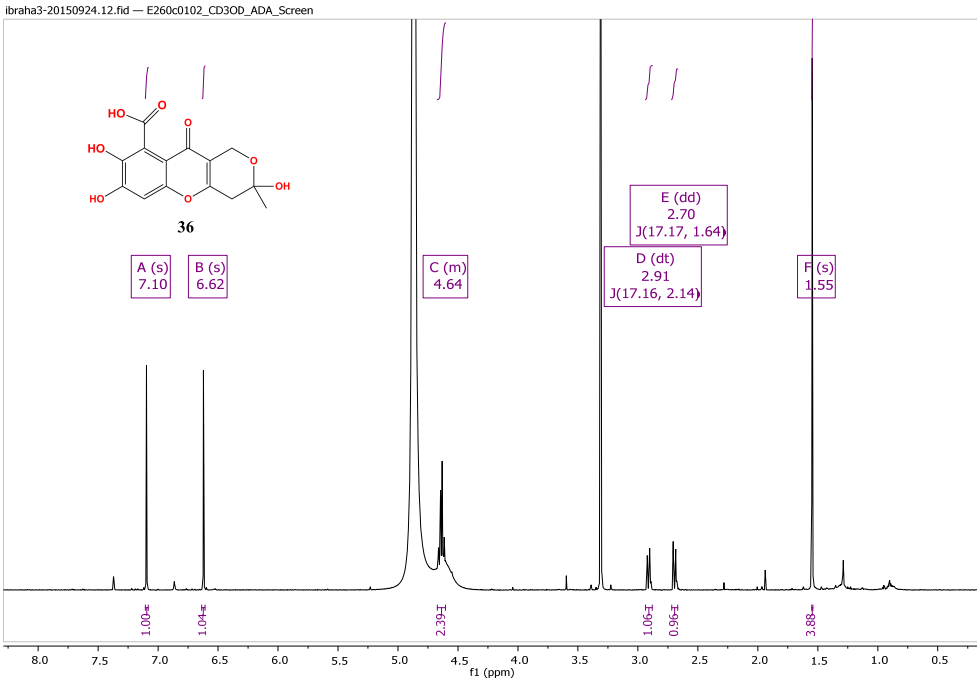


Figure S37 C. ^{1}H NMR spectrum (700 MHz, Acetone) of compound **37**, fulvic acid analogue**.**

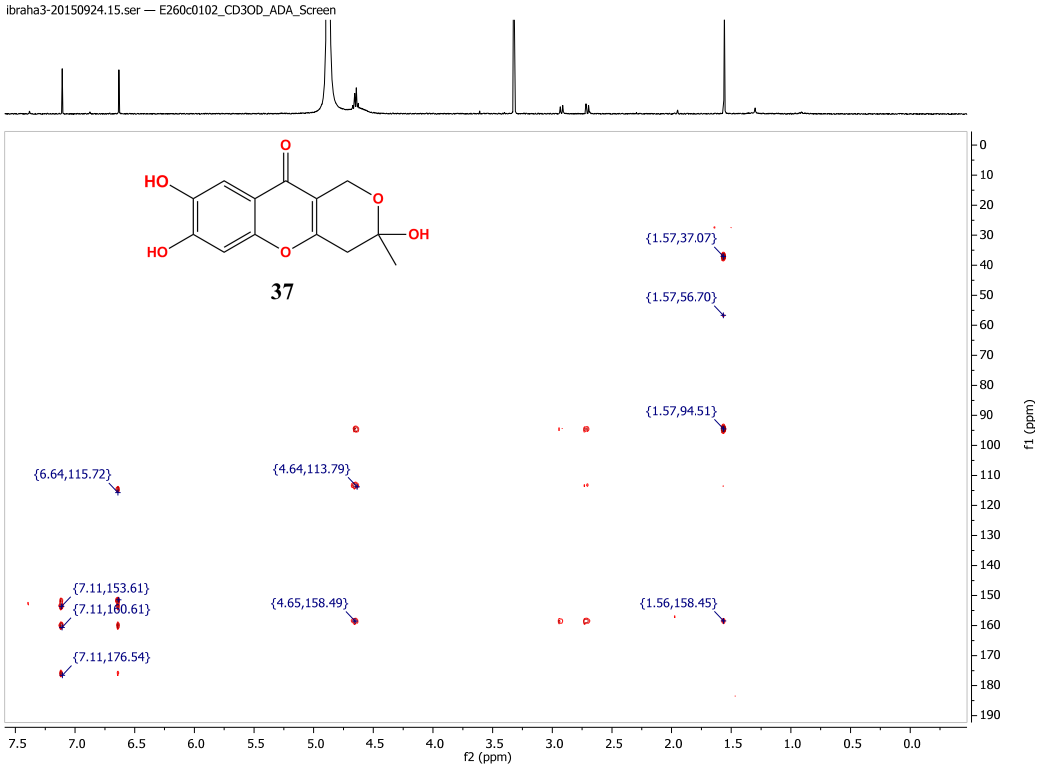


Figure S38 C. HMBC NMR spectrum (176 MHz, Acetone) of compound **37**, fulvic acid analogue.

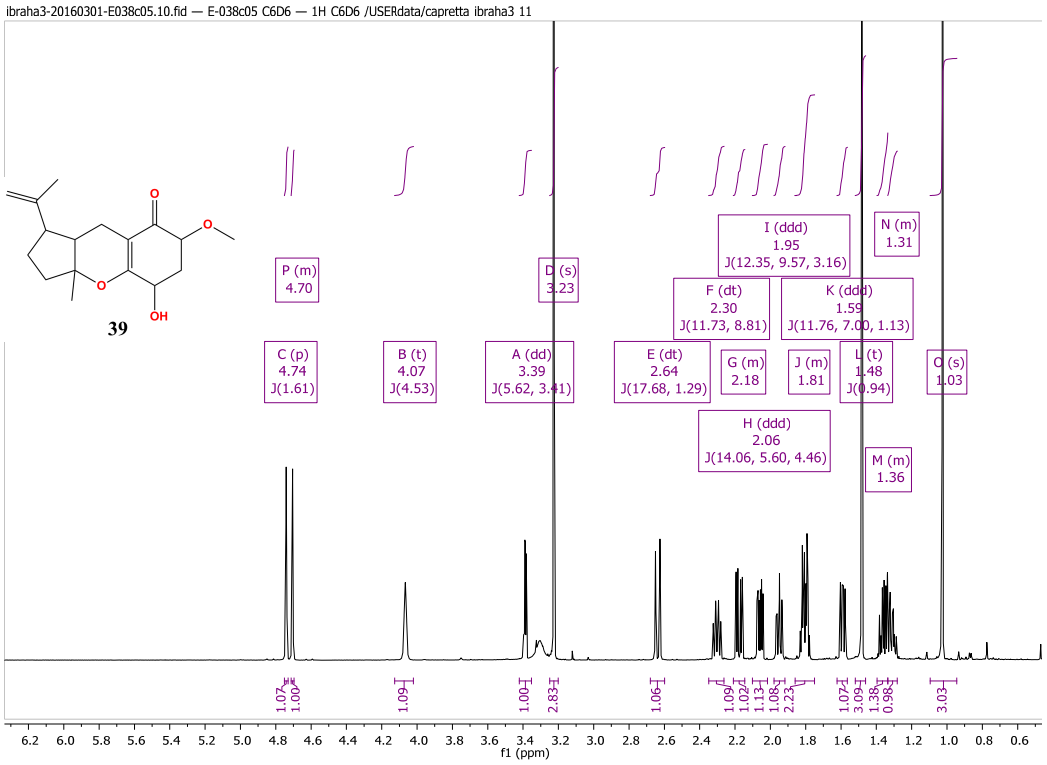


Figure S39 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **39**, coibanol B**.**

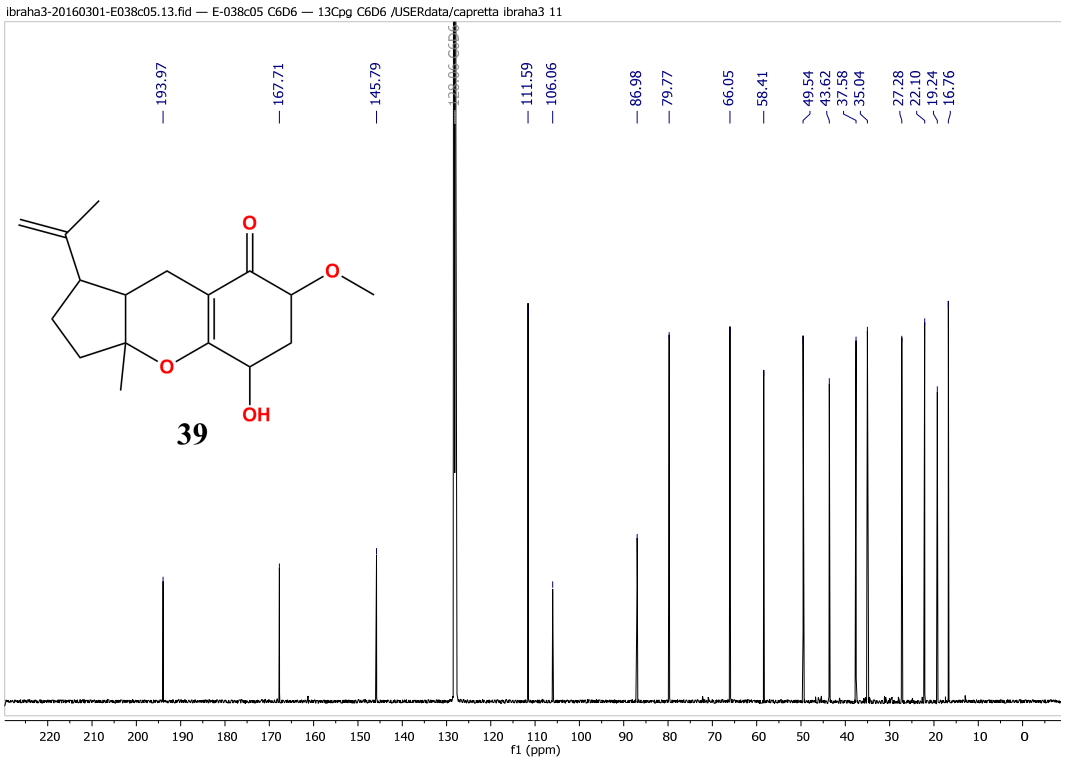


Figure S40 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **39**, coibanol B.

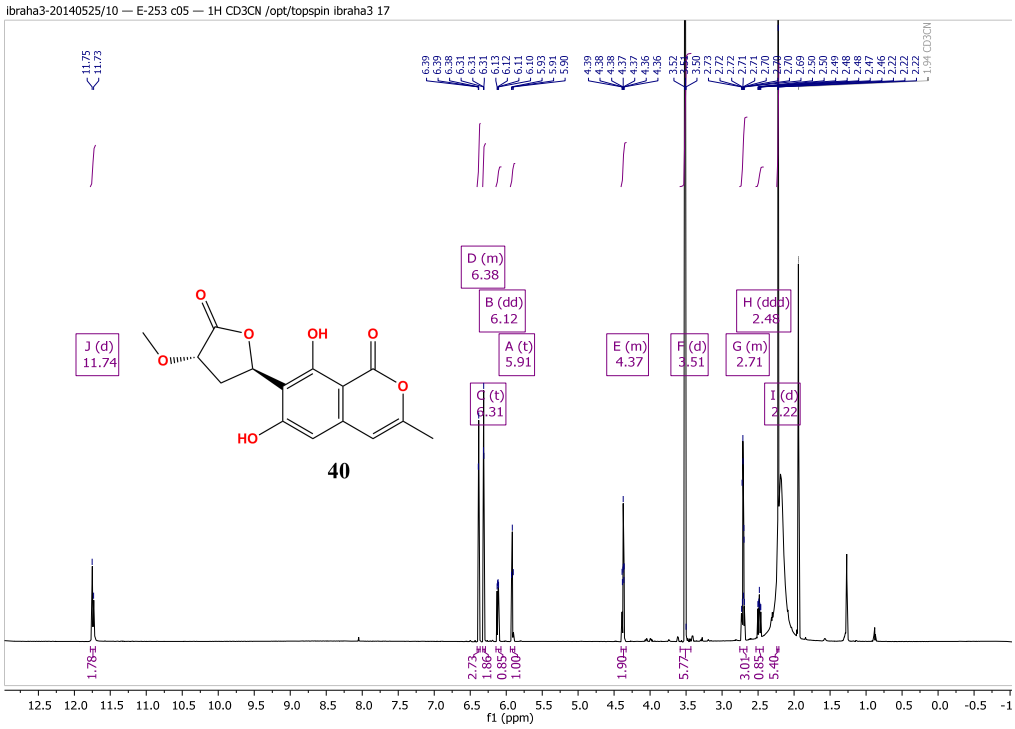


Figure S41 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **40,** mixture of canescin A isomers**.**

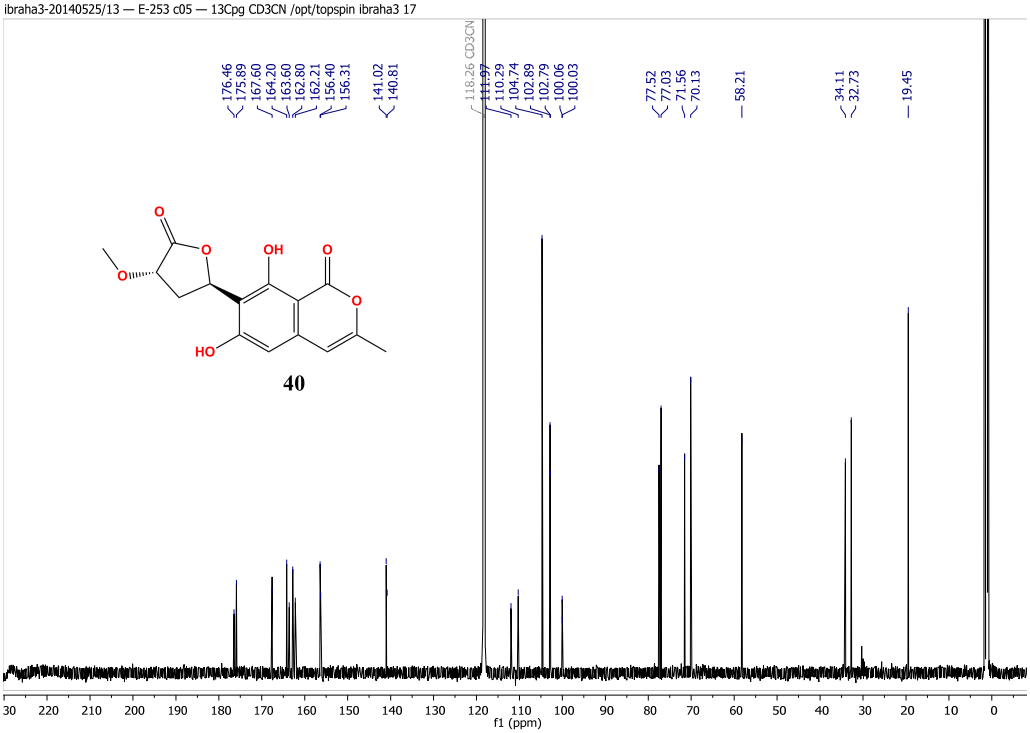


Figure S42 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **40**, mixture of canescin A isomers**.**

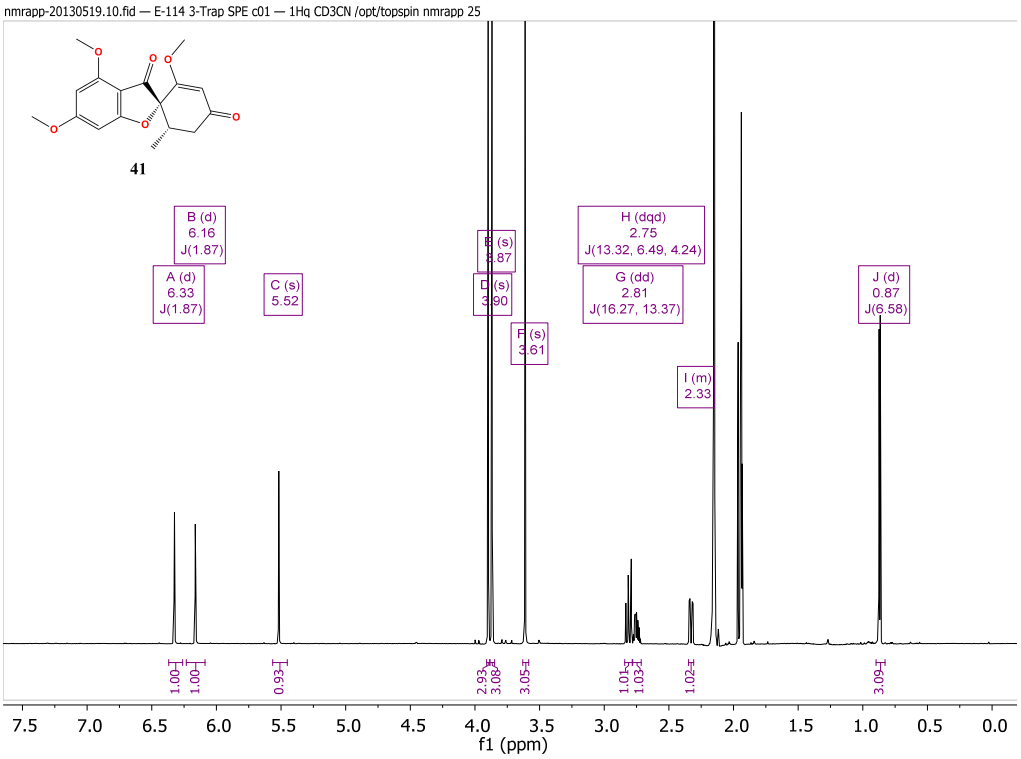


Figure S43 C. ^{1}H NMR spectrum (700 MHz, CD_3CN) of compound **41,** dechlorogriseofulvin**.**

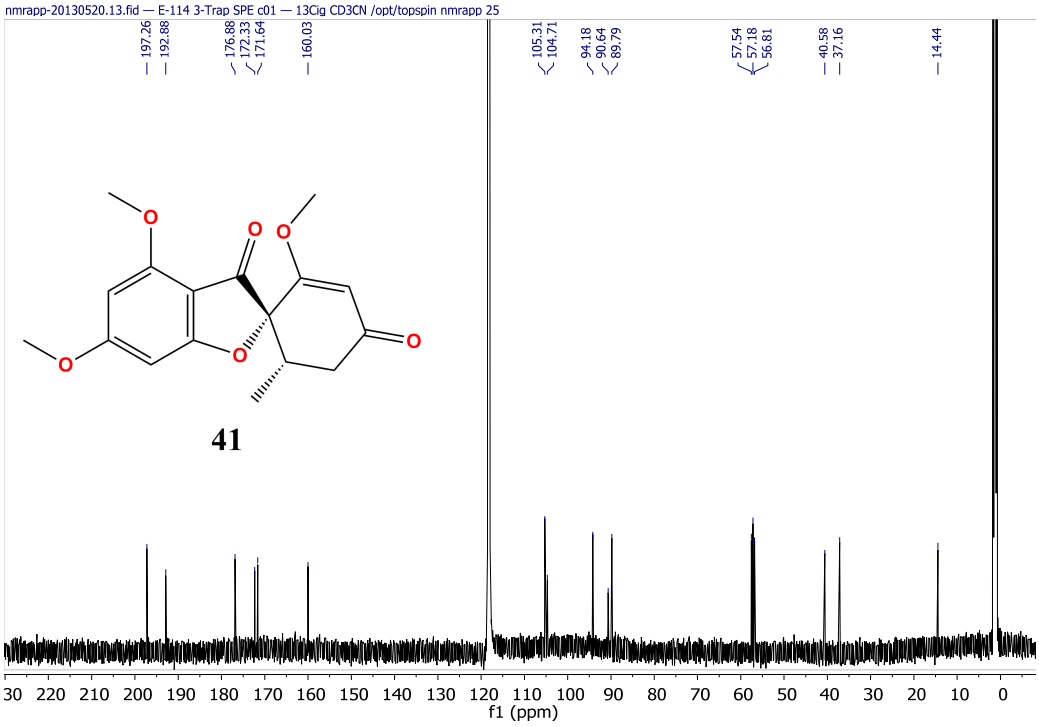


Figure S44 C. ^{13}C NMR spectrum (176 MHz, CD_3CN)of compound **41**, dechlorogriseofulvin.

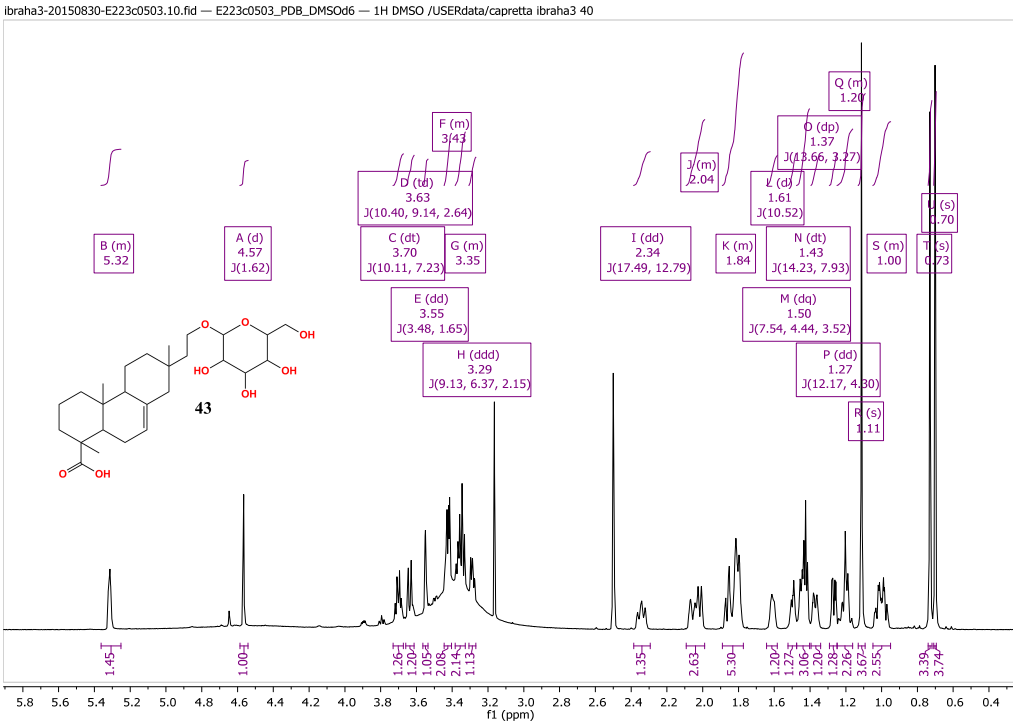


Figure S45 C. ^{1}H NMR spectrum (700 MHz, DMSO-d_6) of compound **43,** 16-α-D-mannopyranosyloxyisopimar-7-en-19-oic acid**.**

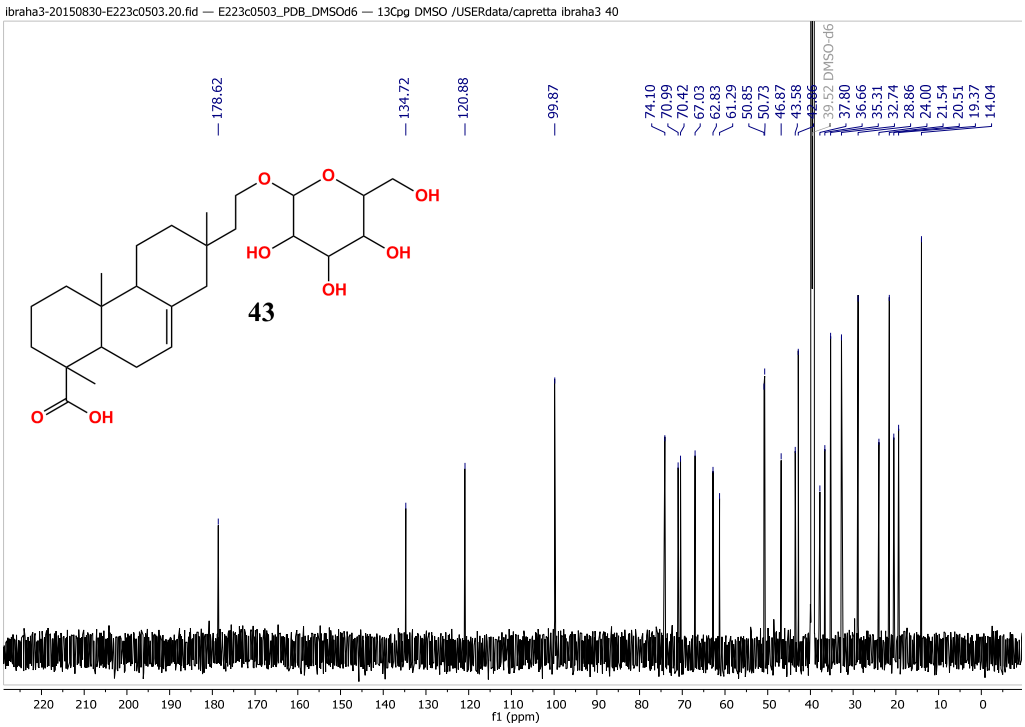


Figure S46 C. ^{13}C NMR spectrum (176 MHz, DMSO-d_6) of compound **43**, 16-α-D-mannopyranosyloxyisopimar-7-en-19-oic acid.

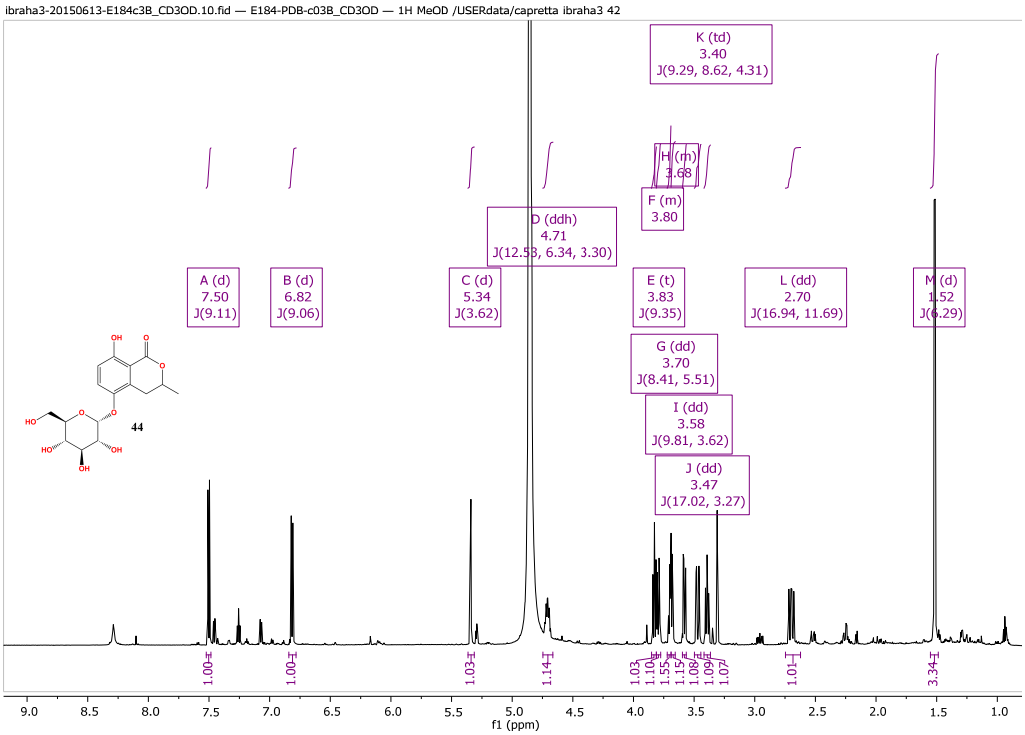


Figure S47 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **44,** 5-O-α-D-glucopyranosyl-5-hydroxymellein.

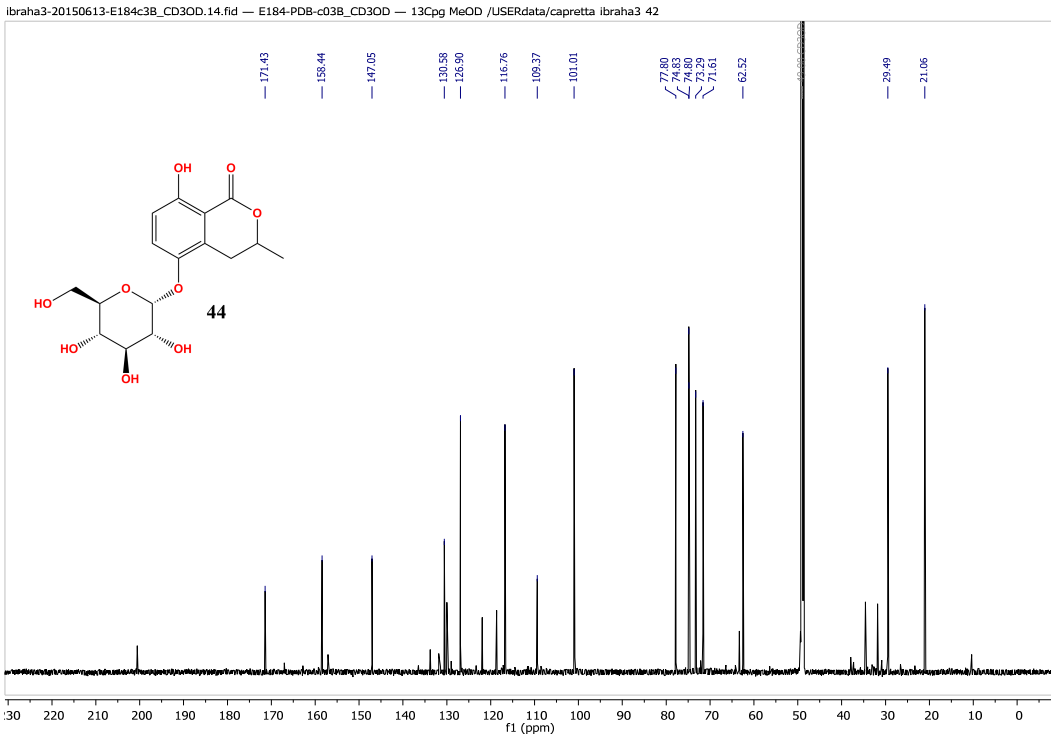


Figure S48 C. ^{13}C NMR spectrum (176 MHz, CD_3OD) of compound **44**, 5-O-α-D-glucopyranosyl-5-hydroxymellein.

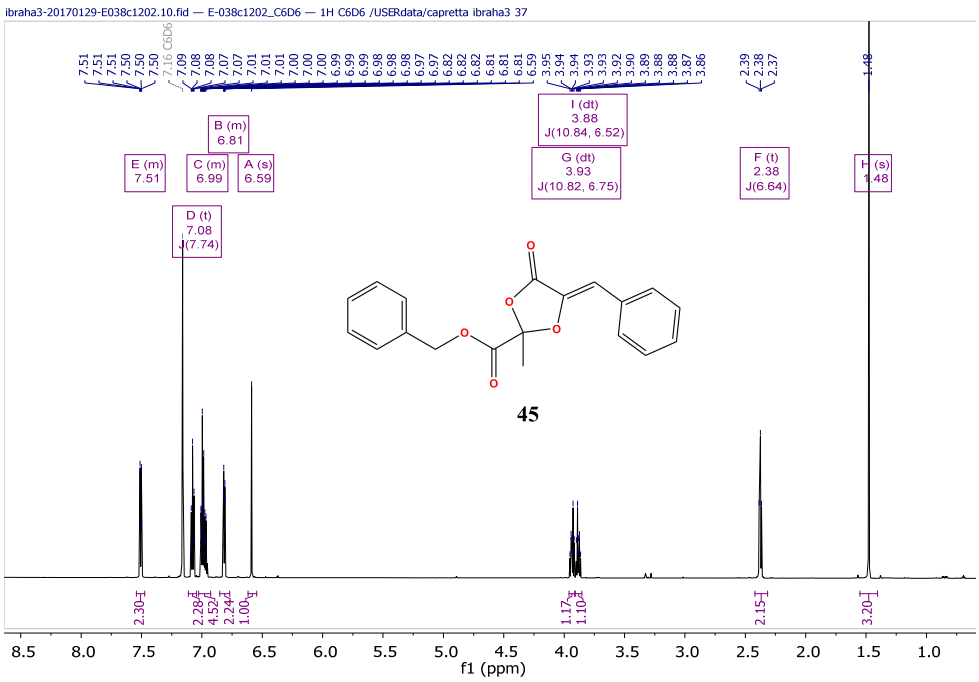


Figure S49 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **45,** guignardianone B.

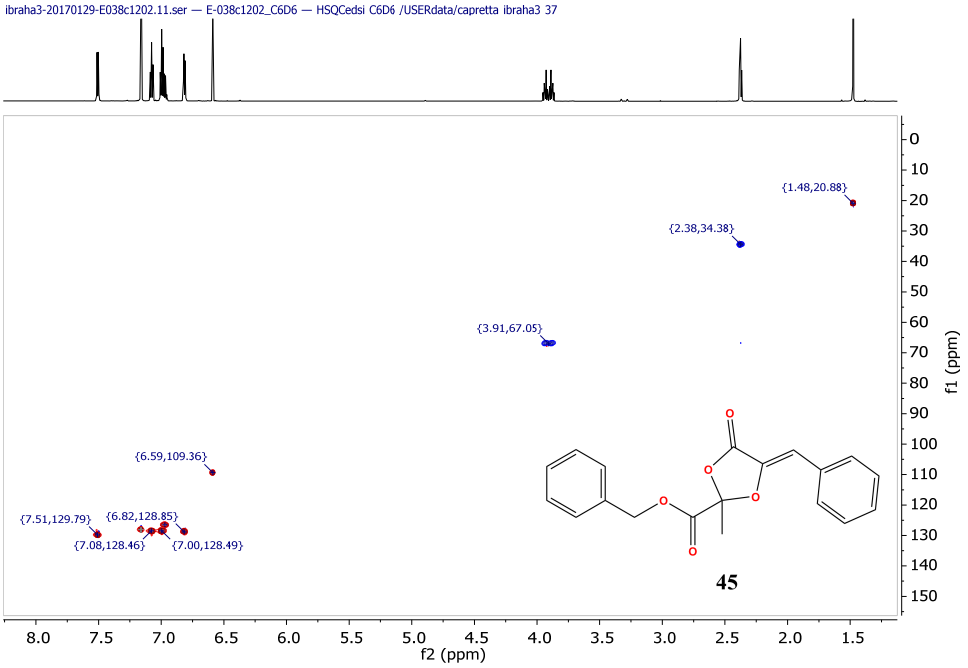


Figure S50 C. Multiplicity-edited HSQC spectrum (700 MHz, 176 MHz, C_6D_6) of compound **45,** guignardianone B.

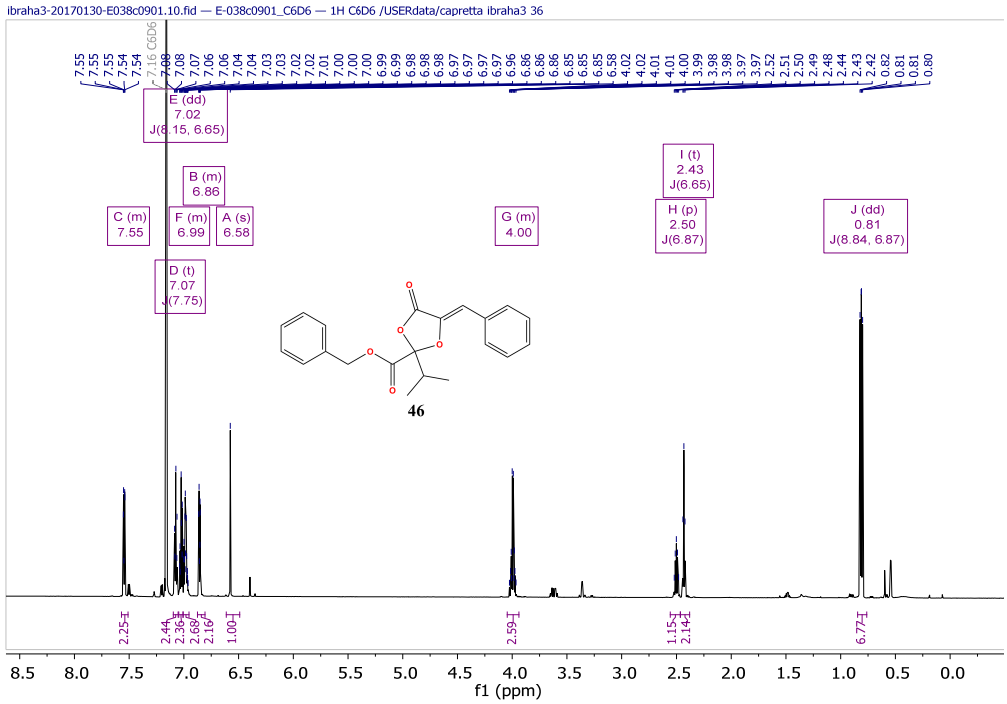


Figure S51 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **46,** guignardianone A.

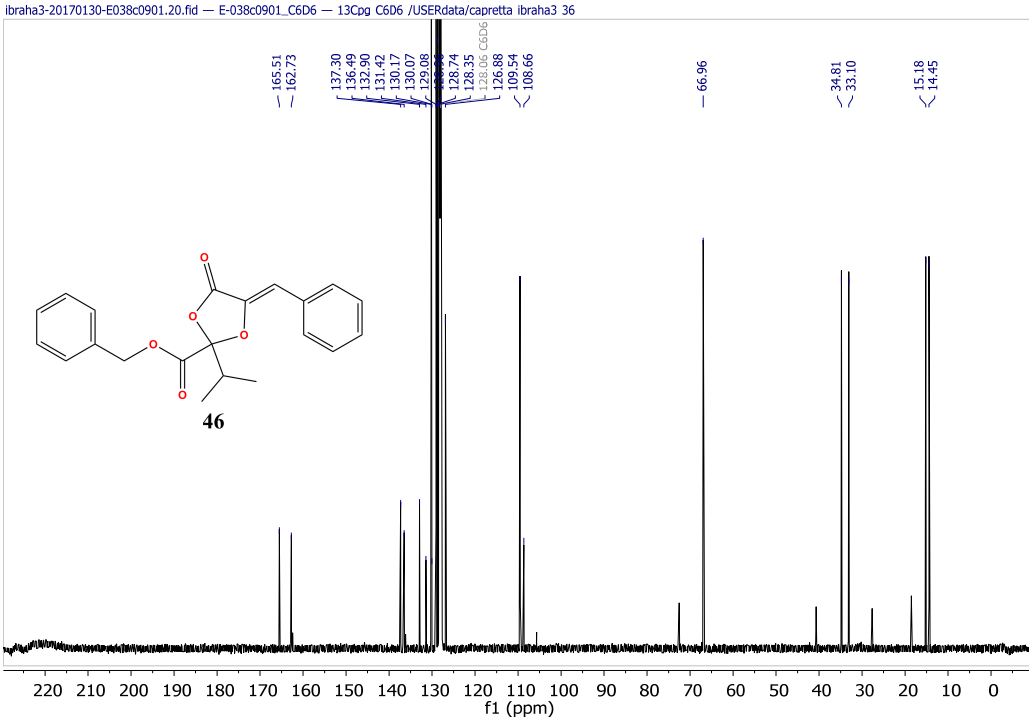


Figure S52 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **46,** guignardianone A.

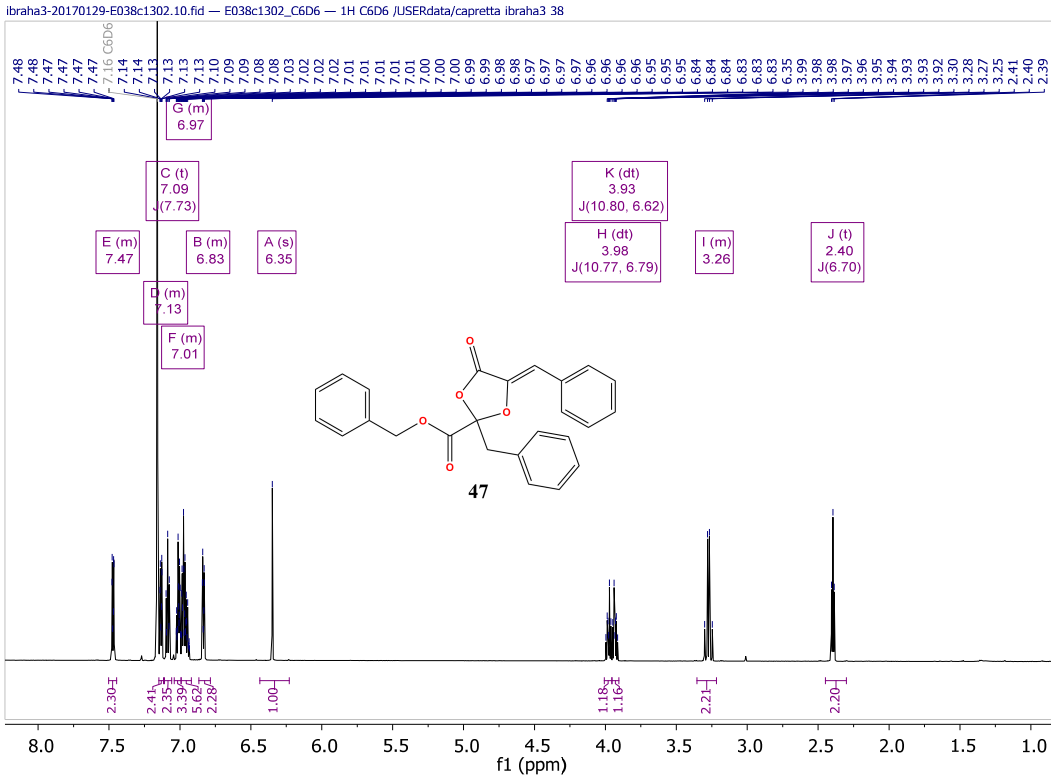


Figure S53 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **47,** guignardianone D

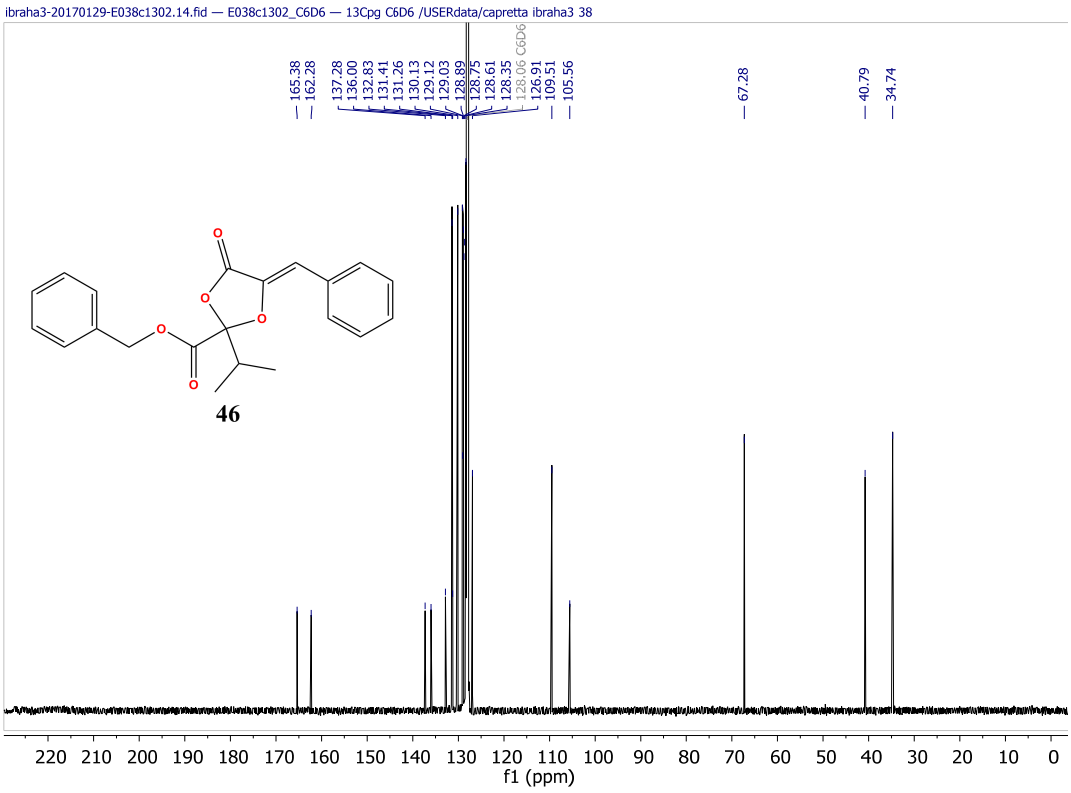


Figure S54 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **47,** guignardianone D

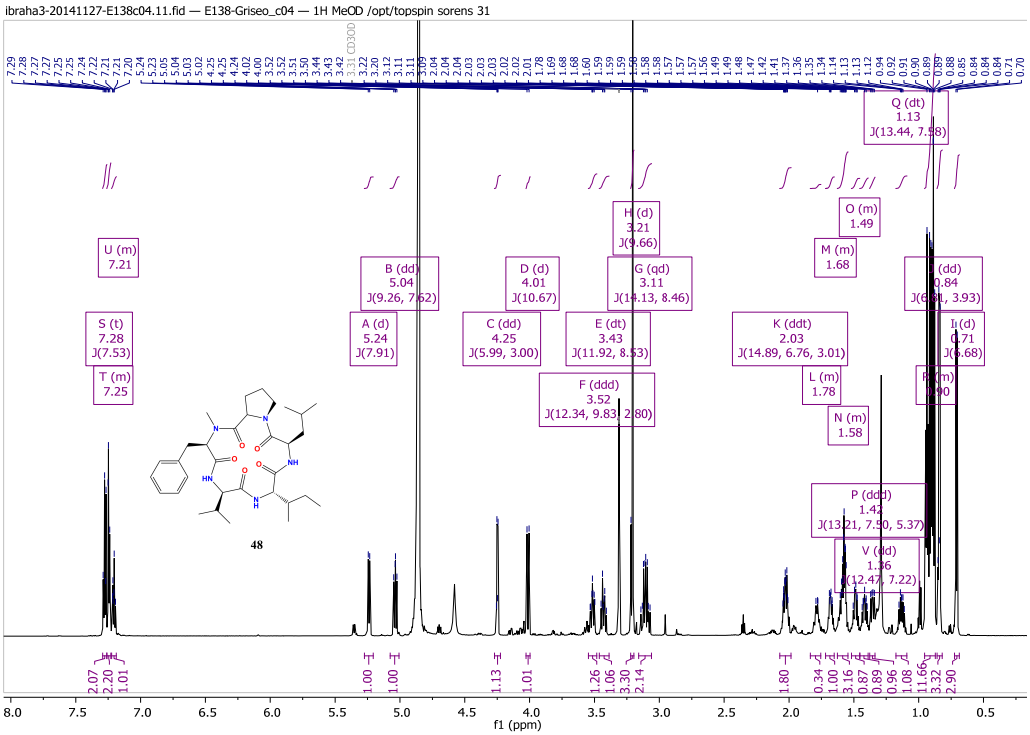


Figure S55 C. ^{1}H NMR spectrum (700 MHz, CD_3OD) of compound **48,** cyclic pentapeptide 1.

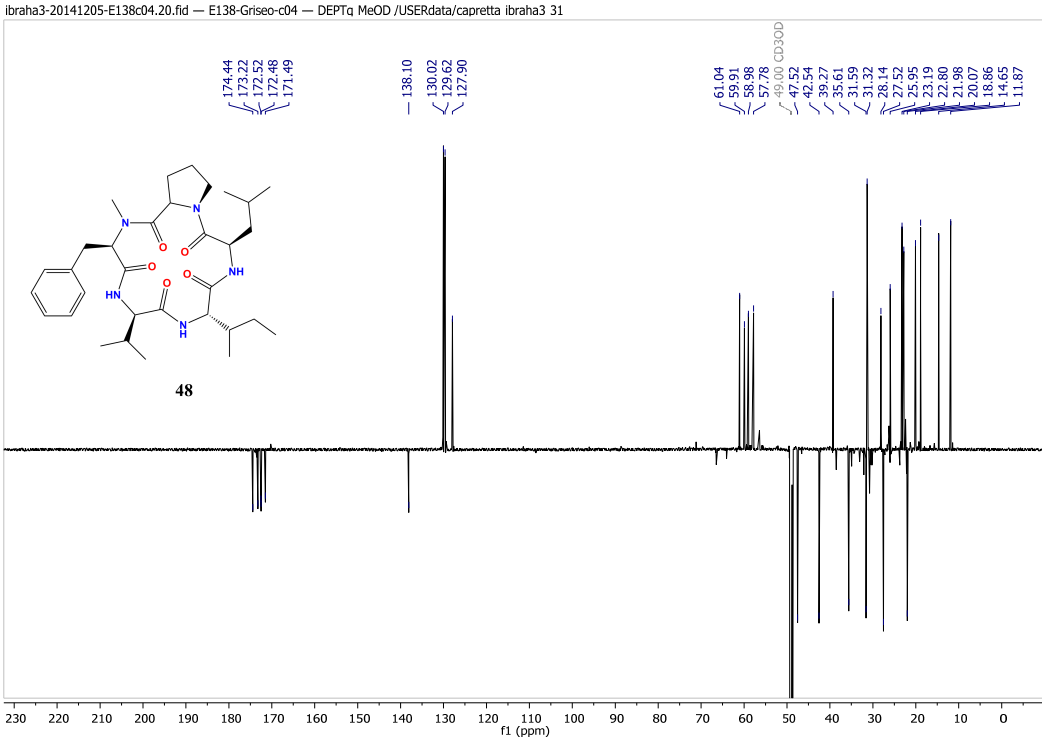


Figure S56 C. ^{13}C DEPTq NMR spectrum (176 MHz, CD_3OD) of compound **48,** cyclic pentapeptide 1.

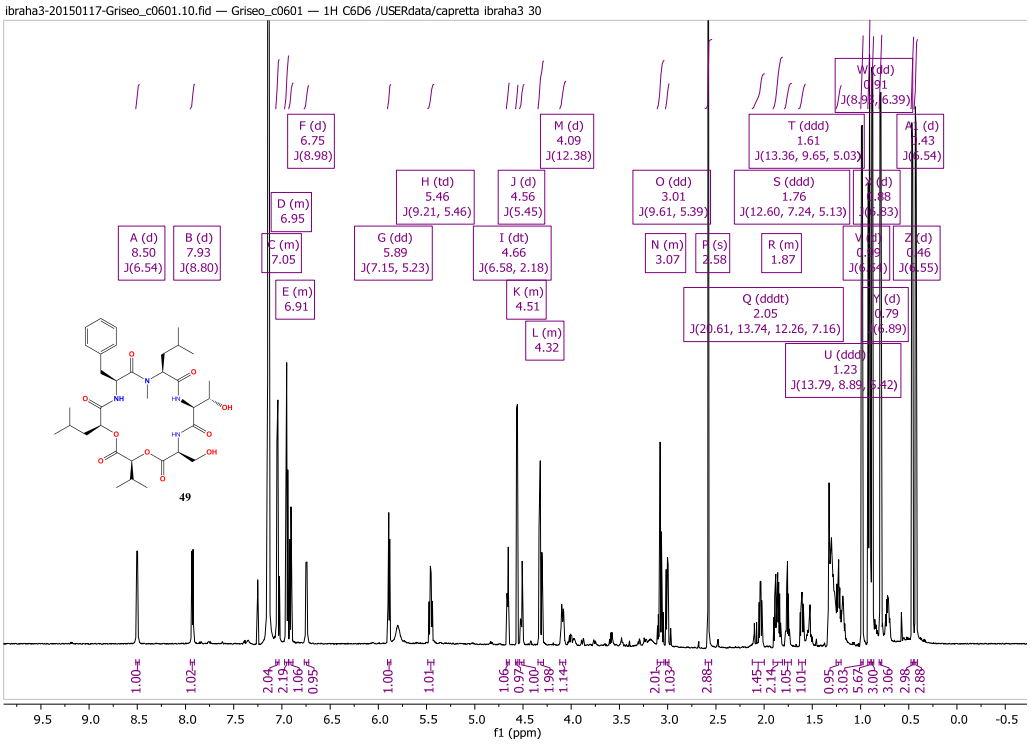


Figure S57 C. ^{1}H NMR spectrum (700 MHz, C_6D_6) of compound **49,** hirsutatin A.

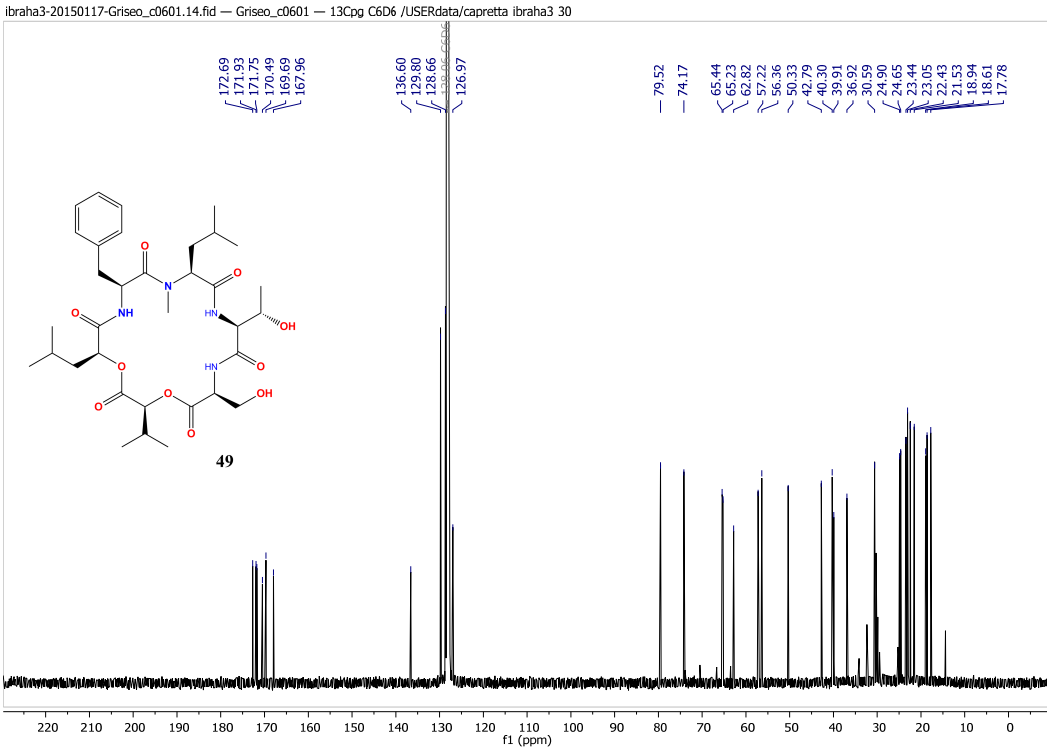


Figure S58 C. ^{13}C NMR spectrum (176 MHz, C_6D_6) of compound **49,** hirsutatin A.

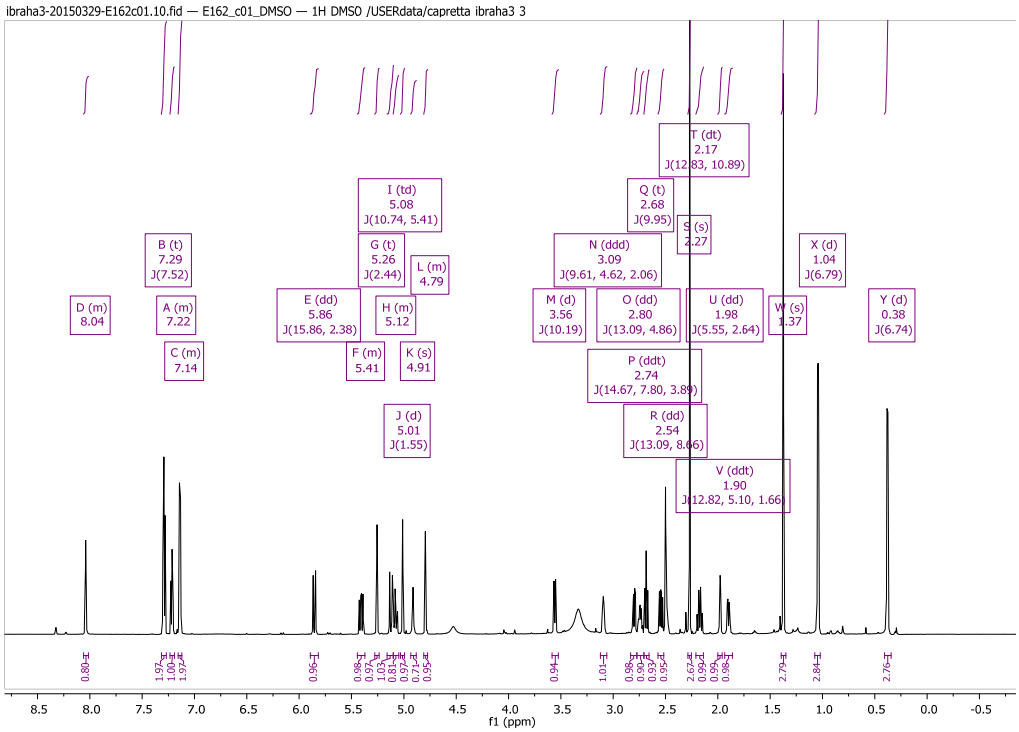


Figure S59 C. ^{1}H NMR spectrum (700 MHz, DMSO-d_6) of compound **51,** cytochalasin D.

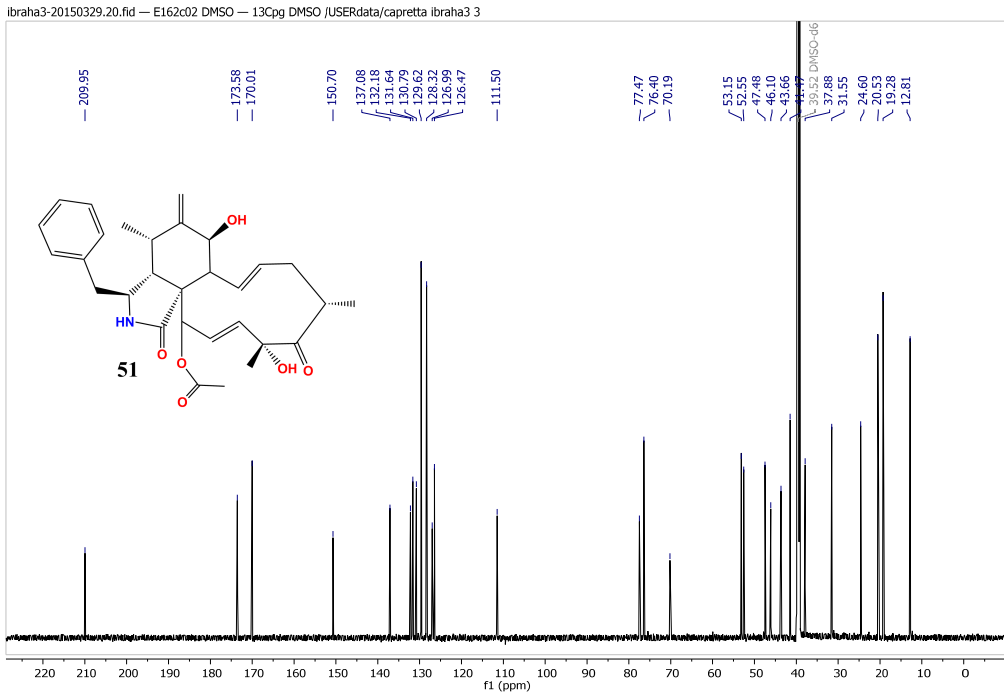


Figure S60 C. ^{13}C NMR spectrum (176 MHz, DMSO-d_6) of compound **51,** cytochalasin D.

APPENDIX D

LC-UV/MS Chromatogram of E-107 from Malt and PDB media at 254 and 210 nm

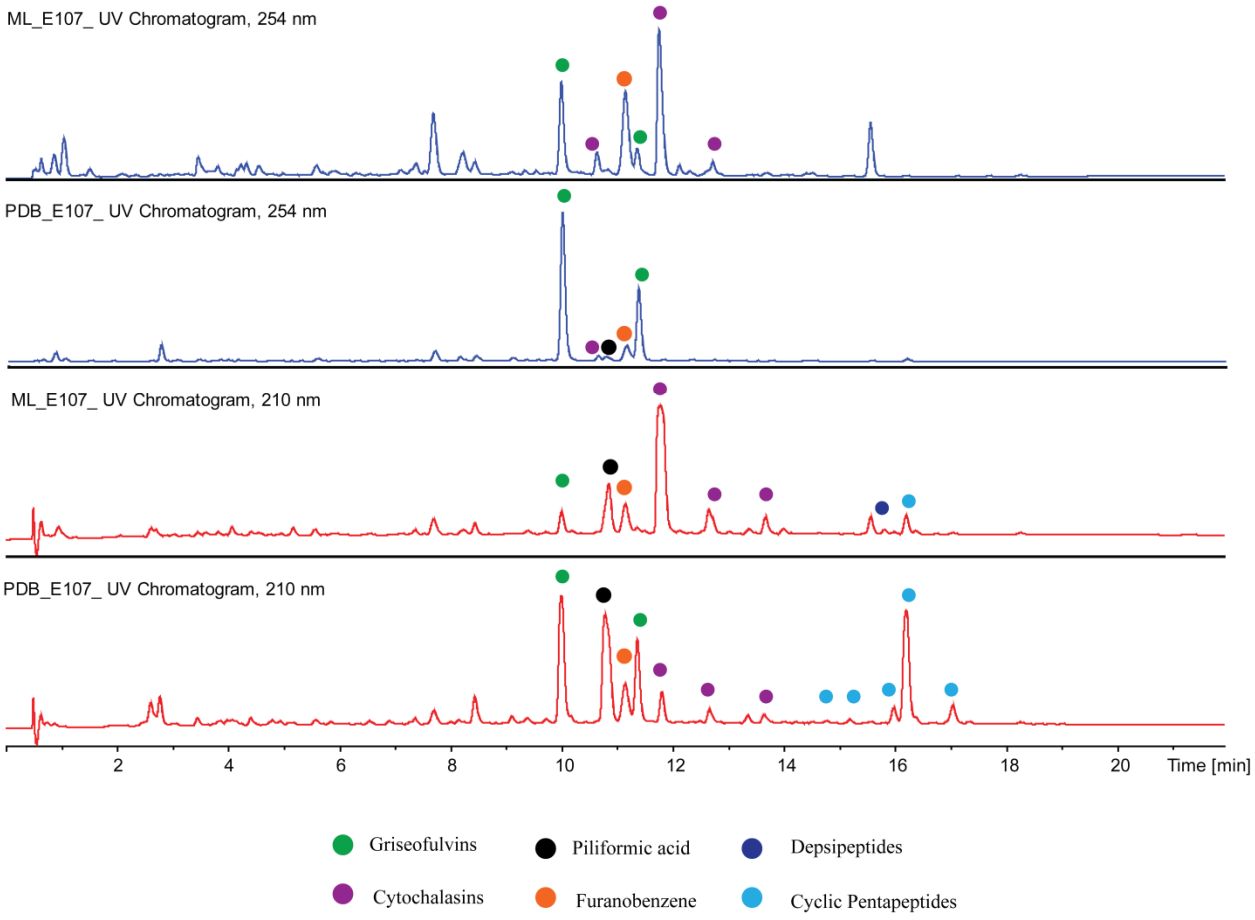


Figure 1S . LC-UV/MS Chromatogram of extract E-107 from Malt and PDB media at 254 and 210 nm respectively, with key metabolites highlighted.

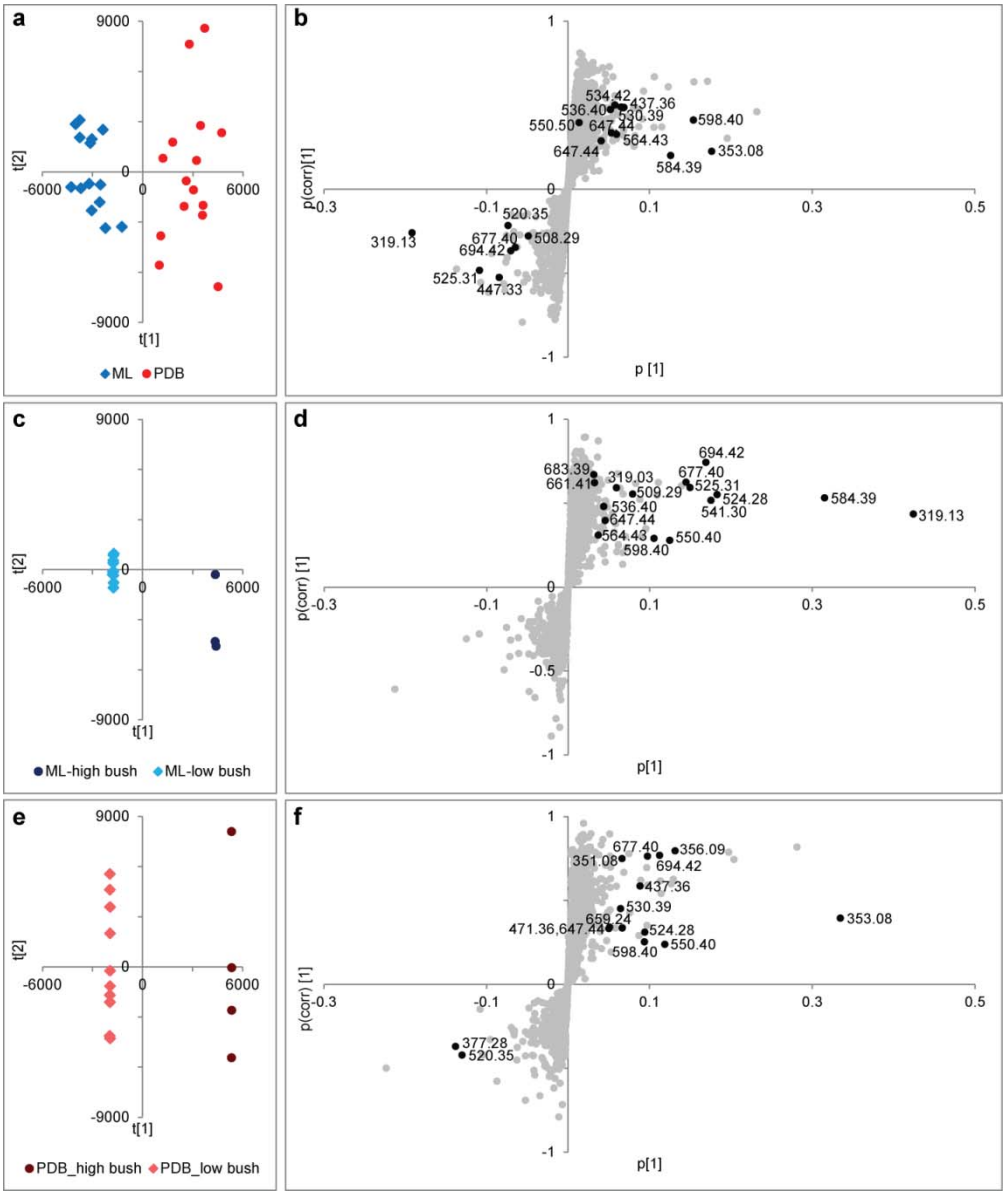


Figure 2S. Supervised multivariate analyses of intracellular metabolome of griesofulvin producing *xylaria* endophytes. The OPLS-DA score plot (a) and S-plot (b) is a comparison between endophytes cultured in ML or PDB media. The OPLS-DA score plots and S- plots compares the endophytes isolates from high or low bush blueberries cultured in ML (c, d) or PDB (e, f) medium, respectively.

Table 1S. A summary of validation parameters (R^2X, R^2Y, Q^2) of all calculated OPLS-DA models for the intracellular extracts from the griseofulvin *Xylaria* endophytes isolates from low and high bush blueberries cultured in ML and PDB media. ML-H, endophytes isolated from high bush blueberries cultured in ML medium; ML-L, endophytes isolated from low bush blueberries cultured in ML medium; PDB-H, endophytes isolated from high bush blueberries cultured in PDB medium; PDB-L, endophytes isolated from low bush blueberries cultured in PDB medium.

Model	Variables*	R^2X(cum)	R^2Y(cum)	Q^2(cum)	Conditions
1a	2645	0.239	0.9	0.689	ML, PDB
1b	100	0.292	0.847	0.697	ML, PDB (include top 100 VIP)
1c	2545	0.359	0.995	0.743	ML, PDB (exclude top 100 VIP)
1d	1945	0.249	0.951	0.347	ML, PDB (exclude top 700 VIP)
2a	2645	0.691	1	0.6	ML-H, ML-L
2b	100	0.865	1	0.616	ML-H, ML-L (include top 100 VIP)
2c	2545	0.0682	0.895	-0.446	ML-H, ML-L (exclude top 100 VIP)
3a	2645	0.659	1	0.676	PDB-H, PDB-L
3b	50	0.768	0.999	0.836	PDB-H, PDB-L (include top 50 VIP)
3c	2595	0.0715	0.89	-0.348	PDB-H, PDB-L (exclude top 50 VIP)

* Number of metabolic features included in the OPLS-DA analyses

Table 2S. The top 100 metabolic features ranked by VIP scores for OPLS-DA analysis of extracellular metabolome of griseofulvin producing *Xylaria* endophytes cultured in ML or PDB media. Features highlighted in yellow, indicate metabolites that were identified during this study. Identifications are based on structural characterization of Semi-preparative HPLC or LC-SPE isolated compounds, which were subjected to HRMS measurements, comprehensive high-field 1-2D NMR experiments and/or MS/MS analysis.

Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score
1	550.42	15.97	11.37	34	481.32	14.42	4.74	67	1047.62	10.87	3.65
2	584.41	16.2	11.1	35	487.35	8.92	4.74	68	263.11	6.43	3.62
3	261.13	2.62	10.64	36	557.29	10	4.68	69	487.35	13.17	3.61
4	239.18	8.46	9.78	37	1032.58	11.81	4.66	70	551.5	15.97	3.6
5	320.16	10.01	8.07	38	522.37	14.47	4.6	71	437.35	15.42	3.57
6	508.28	12.56	8.02	39	267.12	7	4.55	72	560.43	12.58	3.55
7	498.35	14.44	7.91	40	982.77	13.9	4.49	73	445.3	14.72	3.41
8	167.11	3.41	7.71	41	441.31	11.21	4.44	74	353.13	11.37	3.41
9	320.1	9.98	7.69	42	333.21	8.81	4.42	75	253.11	5.04	3.4
10	543.39	13.03	7.25	43	978.69	14.42	4.41	76	245.13	4.83	3.31
11	430.24	11.81	7.18	44	491.27	11.81	4.41	77	405.19	5.6	3.25
12	490.31	11.81	7.1	45	482.36	15.57	4.35	78	598.5	17.04	3.23
13	169.13	10.77	7.09	46	455.36	12.33	4.28	79	509.28	11.83	3.18
14	678.39	15.82	6.98	47	532.36	10.42	4.25	80	929.67	15.58	3.18
15	463.31	14.45	6.91	48	1000.62	13.69	4.23	81	305.11	7.28	3.17
16	197.13	2.72	6.74	49	647.43	13.11	4.22	82	661.39	17.24	3.16
17	207.16	15.59	6.69	50	317.22	14.71	4.2	83	439.36	13.35	3.09
18	197.12	10.77	6.53	51	447.32	15.57	4.18	84	169.11	3.51	2.99
19	600.4	14	6.46	52	706.2	11.37	4.15	85	237.15	7.13	2.96
20	524.29	10.87	5.82	53	946.77	15.59	4.14	86	561.41	9.59	2.93
21	208.11	4.25	5.61	54	309.14	6.82	4.07	87	240.16	8.46	2.92
22	564.43	16.89	5.59	55	207.1	3.47	4	88	1064.59	10.87	2.91
23	151.12	10.77	5.47	56	413.28	14.54	3.98	89	525.3	13.97	2.91
24	356.09	11.37	5.28	57	933.74	16.39	3.92	90	566.4	13.65	2.9
25	544.4	12.56	5.27	58	193.09	6.33	3.92	91	540.27	10	2.88
26	695.42	15.82	5.24	59	535.41	8.87	3.87	92	1014.69	11.89	2.86
27	485.33	9.33	5.16	60	355.12	11.36	3.86	93	321.13	10.01	2.85
28	1015.59	11.86	5.11	61	489.37	12.5	3.83	94	536.39	14.22	2.81
29	947.7	15.58	5.1	62	525.3	11.8	3.76	95	469.34	13.24	2.79
30	429.31	15.57	4.88	63	647.43	15.03	3.76	96	487.35	11.61	2.78
31	423.3	11.21	4.87	64	585.53	16.2	3.75	97	431.32	16.39	2.77
32	209.1	3.69	4.86	65	675.69	24.17	3.73	98	456.28	11.2	2.73
33	705.23	11.37	4.81	66	556.36	14.76	3.72	99	239.13	2.78	2.72
								100	203.18	8.92	2.71

Table 3S. The top 30 metabolic features ranked by VIP scores for OPLS-DA analysis of extracellular metabolome of griesofulvin *Xylaria* endophyte isolate from low or high bush blueberries cultured in ML medium Features highlighted in yellow, indicate metabolites that were identified during this study. Identifications are based on structural characterization of Semi-preparative HPLC or LC-SPE isolated compounds, which were subjected to HRMS measurements, comprehensive high-field 1-2D NMR experiments and/or MS/MS analysis

Rank	m/z	Rt. (min)	VIP score
1	239.18	8.46	13.02
2	598.41	17.04	11.59
3	221.16	8.84	10.74
4	550.42	15.97	10.42
5	490.31	11.81	9.45
6	508.28	12.56	9.21
7	1032.58	11.81	8.26
8	678.39	15.82	8.24
9	319.16	10.01	8.12
10	1015.59	11.86	8.05
11	473.29	9.71	7.73
12	320.1	9.98	7.01
13	570.37	15.19	6.87
14	279.2	8.84	6.81
15	1016.62	11.81	6.69
16	203.18	8.92	6.53
17	309.14	6.82	6.51
18	430.24	11.81	6.43
19	267.12	7	5.9
20	354.1	11.37	5.66
21	240.1	8.45	5.45
22	584.41	16.2	5.36
23	305.11	7.28	5.26
24	295.17	7.03	5.22
25	675.69	24.17	5.13
26	219.18	6.96	5.05
27	173.1	7.38	4.93
28	491.27	11.81	4.75
29	237.12	5.52	4.73
30	263.11	6.43	4.63

Table 4S. The top 50 metabolic features ranked by VIP scores for OPLS-DA analysis of extracellular metabolome of griesofulvin *Xylaria* endophytes isolated from low or high bush blueberries cultured in PDB medium. Features highlighted in yellow, indicate metabolites that were identified during this study. Identifications are based on structural characterization of Semi-preparative HPLC or LC-SPE isolated compounds, which were subjected to HRMS measurements, comprehensive high-field 1-2D NMR experiments and/or MS/MS analysis.

Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score
1	320.16	10.01	20.43	26	490.31	11.81	4.98
2	706.2	11.37	10.84	27	707.17	11.37	4.97
3	221.16	8.84	9.89	28	1048.6	10.89	4.88
4	598.41	17.04	9.49	29	203.18	8.92	4.72
5	356.09	11.37	8.66	30	525.3	11.8	4.65
6	353.13	11.37	8.53	31	265.19	13.26	4.58
7	541.3	10.87	8.02	32	393.3	24.07	4.47
8	321.13	10.01	7.9	33	705.23	11.37	4.36
9	181.05	10.01	7.35	34	320.1	9.98	4.33
10	1015.59	11.86	7.28	35	209.12	10.92	4.32
11	659.23	10.01	7.13	36	1047.62	10.87	4.23
12	550.42	15.97	6.51	37	527.39	14.24	4.22
13	678.39	15.82	6.26	38	763.66	24.04	4.22
14	223.17	13.79	6.16	39	570.37	15.19	4.21
15	584.41	16.2	6.09	40	456.28	11.2	4.13
16	251.11	10	6.02	41	317.11	9.56	4.07
17	473.29	9.71	5.95	42	1088.82	14.21	4.06
18	219.18	6.96	5.79	43	694.47	15.82	4.01
19	351.07	11.6	5.65	44	391.29	24.1	3.98
20	675.69	24.17	5.44	45	585.41	16.2	3.89
21	279.2	8.84	5.36	46	293.18	6.41	3.88
22	647.43	13.11	5.08	47	557.29	10	3.87
23	522.37	14.47	5.07	48	247.17	13.36	3.86
24	165.08	10	5	49	237.11	10.84	3.7
25	193.09	6.33	5	50	459.32	12.04	3.67

Table 5S. The 100 metabolic features ranked by VIP scores for OPLS-DA analysis of intracellular metabolome of griesofulvin *Xylaria* endophytes cultured in ML or PDB media. Features highlighted in yellow, indicate metabolites that were identified during this study. Identifications are based on structural characterization of Semi-preparative HPLC or LC-SPE isolated compounds, which were subjected to HRMS measurements, comprehensive high-field 1-2D NMR experiments and/or MS/MS analysis

Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score
1	550.4	16.03	11.98	34	520.35	16.46	3.45	67	205.07	0.52	2.79
2	353.08	11.42	10.77	35	758.6	23.31	3.38	68	169.13	10.83	2.75
3	666.54	24.7	8.06	36	677.4	15.85	3.35	69	754.65	28.53	2.72
4	761.64	28.66	7.91	37	294.16	0.76	3.32	70	251.16	0.84	2.71
5	319.13	10.05	7.78	38	764.67	25.16	3.31	71	441.31	12.71	2.69
6	598.4	17.08	7.17	39	1032.59	11.86	3.29	72	489.38	12.55	2.69
7	682.54	23.84	7.07	40	217.1	2.63	3.26	73	509.29	11.85	2.68
8	933.69	16.41	6.65	41	500.41	18.31	3.25	74	734.62	28.84	2.63
9	205.1	1.77	5.69	42	490.28	12.25	3.23	75	514.36	14.43	2.62
10	584.39	16.24	5.6	43	768.63	24.15	3.23	76	708.61	28.54	2.62
11	698.55	22.63	5.54	44	705.26	11.42	3.22	77	536.4	14.28	2.6
12	223.17	13.82	5.44	45	782.59	22.33	3.22	78	489.32	9.25	2.58
13	760.63	28.95	5.35	46	687.51	23.86	3.19	79	437.36	17.53	2.57
14	235.13	0.65	5.24	47	647.44	13.16	3.17	80	768.59	23.74	2.57
15	353.28	17.13	4.82	48	437.36	15.36	3.14	81	437.36	20.84	2.55
16	525.31	11.86	4.71	49	229.16	1.88	3.14	82	231.24	2.01	2.54
17	413.31	16.41	4.57	50	317.22	14.75	3.12	83	407.3	13.71	2.52
18	355.14	11.42	4.54	51	478.31	16.25	3.09	84	647.44	16.1	2.52
19	245.19	3.15	4.5	52	274.1	0.67	3.05	85	193.09	11.19	2.48
20	707.17	11.42	4.42	53	455.37	15.35	3.04	86	508.29	12.68	2.47
21	488.21	2.11	4.39	54	377.28	20.33	3.02	87	752.63	27.38	2.47
22	671.51	24.74	4.33	55	455.36	12.4	3.02	88	665.53	23.86	2.46
23	425.31	13.72	4.33	56	740.55	22.81	2.99	89	169.13	3.55	2.45
24	548.39	10.36	4.28	57	471.36	14.26	2.96	90	193.09	6.58	2.38
25	457.38	14.79	4.17	58	344.29	18.88	2.95	91	983.66	13.93	2.36
26	320.14	10.05	4.08	59	530.39	12.27	2.95	92	736.61	25.13	2.35
27	430.25	11.85	3.94	60	783.6	21.79	2.89	93	474.4	18.15	2.34
28	694.42	15.85	3.9	61	782.62	29.01	2.86	94	705.52	21.97	2.32
29	217.09	0.63	3.88	62	166.13	1.01	2.86	95	151.12	10.83	2.31
30	447.33	13.93	3.83	63	184.14	0.73	2.85	96	549.4	10.04	2.3
31	784.61	22.69	3.75	64	762.64	28.02	2.84	97	279.14	4.13	2.3
32	431.33	16.42	3.67	65	564.43	16.96	2.83	98	338.35	24.21	2.29
33	439.37	13.39	3.55	66	534.42	13.07	2.81	99	356.09	11.42	2.29
								100	642.56	25.17	2.28

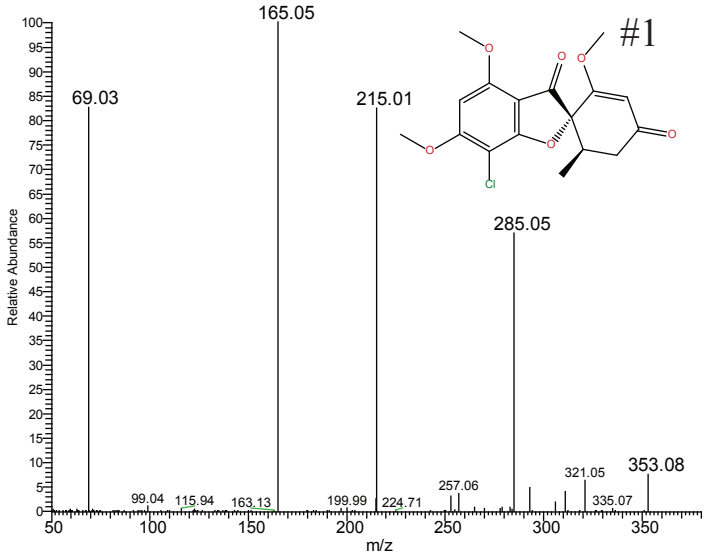
Table 6S. The 100 metabolic features ranked by VIP scores for OPLS-DA analysis of intracellular metabolome of griesofulvin *Xylaria* endophytes isolated from low or high bush blueberries cultured in ML medium. Features highlighted in yellow, indicate metabolites that were identified during this study. Identifications are based on structural characterization of Semi-preparative HPLC or LC-SPE isolated compounds, which were subjected to HRMS measurements, comprehensive high-field 1-2D NMR experiments and/or MS/MS analysis.

Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score
1	319.13	10.05	21.82	34	493.3	13.72	3.14	67	279.2	8.9	2.12
2	584.39	16.24	16.24	35	144.99	27.6	3.14	68	146.99	25.81	2.1
3	933.69	16.41	10.93	36	637.25	10.06	3.08	69	249.23	16.9	2.09
4	524.28	10.92	9.43	37	319.03	12.57	3.05	70	395.34	20.65	2.07
5	541.3	10.91	9.05	38	706.19	11.42	2.99	71	204.13	0.71	2.07
6	694.42	15.85	8.7	39	740.55	22.81	2.98	72	317.07	9.61	2.06
7	525.31	11.86	7.7	40	206.08	0.55	2.92	73	727.16	11.42	2.01
8	677.4	15.85	7.42	41	982.72	13.94	2.81	74	1049.58	11.36	2
9	1064.58	10.92	7.18	42	223.17	13.82	2.54	75	551.42	20.18	1.97
10	550.4	16.03	6.44	43	671.51	24.74	2.53	76	172.1	0.88	1.96
11	783.6	21.79	6.4	44	282.28	21.55	2.48	77	321.02	12.56	1.93
12	355.14	11.42	5.72	45	740.65	22.69	2.47	78	564.43	16.96	1.92
13	166.13	1.01	5.65	46	528.25	3.72	2.46	79	263.1	6.35	1.91
14	1032.59	11.86	5.63	47	760.63	28.95	2.46	80	358.18	15.85	1.91
15	598.4	17.08	5.45	48	146.98	28.37	2.44	81	254.16	0.91	1.89
16	490.28	12.25	4.93	49	224.14	13.81	2.38	82	184.14	0.73	1.89
17	705.26	11.42	4.61	50	335.04	10.02	2.36	83	166.06	1.02	1.88
18	430.25	11.85	4.55	51	320.14	10.05	2.35	84	431.33	16.42	1.87
19	1047.56	10.92	4.28	52	647.44	16.1	2.34	85	295.2	7.07	1.86
20	509.29	11.85	4.08	53	377.33	25.07	2.33	86	175.06	0.54	1.86
21	339.42	24.21	4.05	54	413.31	16.41	2.29	87	1005.57	13.72	1.84
22	758.6	23.31	3.87	55	1000.61	13.71	2.28	88	786.63	23.76	1.84
23	318.31	14.52	3.72	56	277.22	12	2.27	89	184.07	0.59	1.84
24	785.61	22.96	3.66	57	536.4	14.28	2.26	90	699.39	15.87	1.83
25	205.1	1.77	3.64	58	437.36	17.53	2.25	91	207.1	8.26	1.83
26	707.17	11.42	3.62	59	666.54	24.7	2.24	92	237.12	10.85	1.8
27	782.59	22.33	3.5	60	356.09	11.42	2.24	93	203.19	9.36	1.78
28	267.1	7.03	3.46	61	169.13	3.55	2.21	94	268.11	0.72	1.75
29	377.28	20.33	3.42	62	478.31	16.25	2.19	95	355.3	20.12	1.74
30	526.31	11.58	3.28	63	457.31	10.22	2.19	96	1069.54	10.92	1.74
31	239.14	8.49	3.18	64	1015.57	11.86	2.17	97	687.51	23.86	1.7
32	659.24	10.06	3.17	65	804.6	22.12	2.17	98	491.32	12.25	1.7
33	716.55	24.94	3.15	66	771.61	24.08	2.15	99	231.24	2.01	1.7
								100	151.12	3.16	1.69

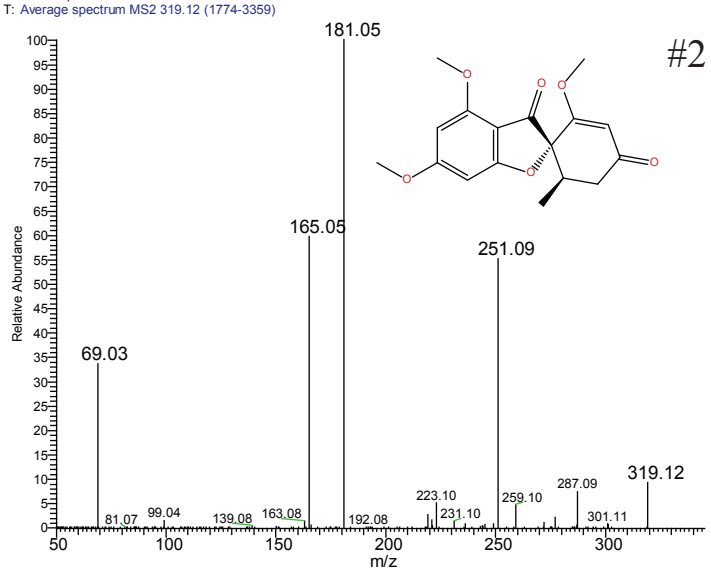
Table 7S. The top 50 metabolic features ranked by VIP scores for OPLS-DA analysis of intracellular metabolome of griesofulvin *Xylaria* endophytes isolated from low or high bush blueberries cultured in PDB medium.

Rank	m/z	Rt. (min)	VIP score	Rank	m/z	Rt. (min)	VIP score
1	353.08	11.42	17.22	26	184.07	0.59	4.5
2	707.17	11.42	14.48	27	760.63	28.95	4.49
3	782.59	22.33	11.48	28	1064.58	10.92	3.92
4	666.54	24.7	10.49	29	727.16	11.42	3.86
5	682.54	23.84	10.16	30	338.35	24.21	3.58
6	377.28	20.33	7.1	31	181.05	10.05	3.53
7	356.09	11.42	6.78	32	457.38	14.79	3.46
8	520.35	16.46	6.68	33	353.28	17.13	3.44
9	706.19	11.42	6.66	34	351.08	11.65	3.42
10	768.59	23.74	6.54	35	659.24	10.06	3.42
11	550.4	16.03	6.14	36	254.16	0.91	3.39
12	698.55	22.63	5.89	37	764.67	25.16	3.38
13	489.32	9.25	5.85	38	530.39	12.27	3.33
14	694.42	15.85	5.79	39	804.6	22.12	3.22
15	761.64	28.66	5.55	40	320.14	10.05	3.2
16	758.6	23.31	5.38	41	642.56	25.17	3.07
17	671.51	24.74	5.04	42	639.33	10.06	3
18	677.4	15.85	5.03	43	514.36	14.43	2.85
19	541.3	10.91	5	44	488.21	2.11	2.74
20	490.28	12.25	4.98	45	925.75	28.42	2.73
21	783.6	21.79	4.9	46	375.07	11.42	2.72
22	524.28	10.92	4.86	47	353.09	12.9	2.71
23	598.4	17.08	4.84	48	251.1	10.05	2.69
24	341.11	10.07	4.61	49	339.42	24.21	2.69
25	437.36	15.36	4.56	50	217.09	0.63	2.68

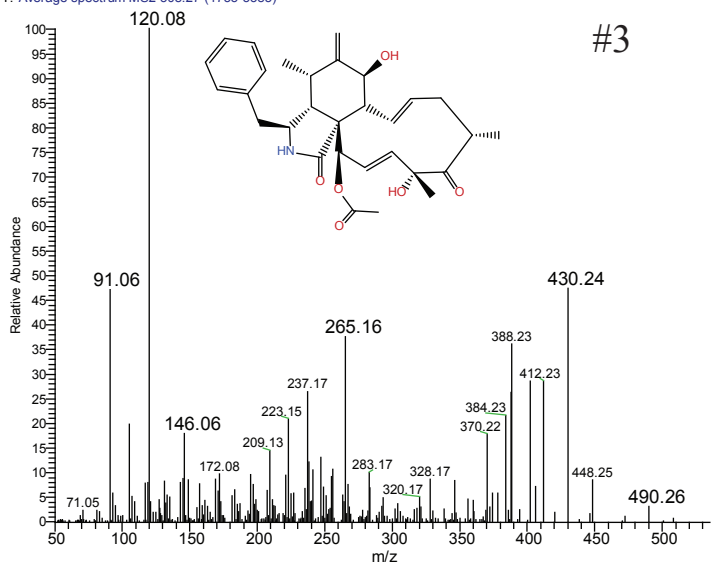
Griso-NL2-pos #1903-3308 RT: 3.37-5.85 AV: 8 NL: 3.85E7
T: Average spectrum MS2 353.08 (1903-3308)



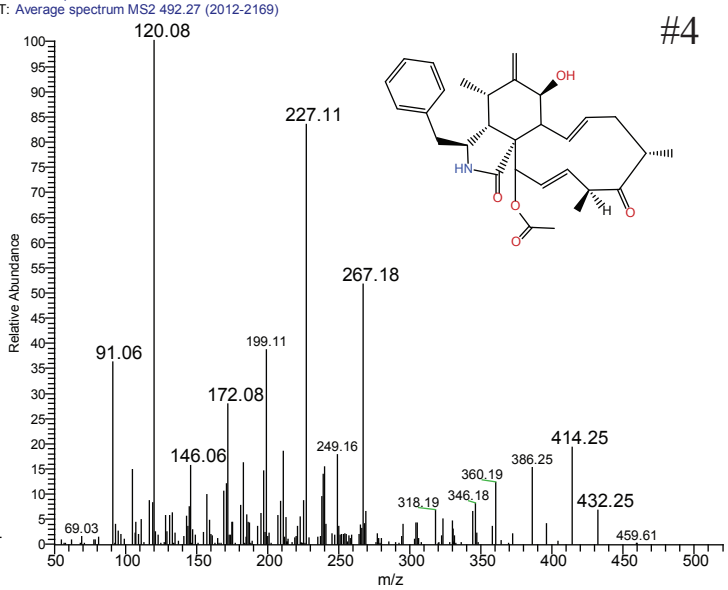
Griso-NL2-pos #1774-3359 RT: 3.14-5.94 AV: 16 NL: 2.33E8
T: Average spectrum MS2 319.12 (1774-3359)



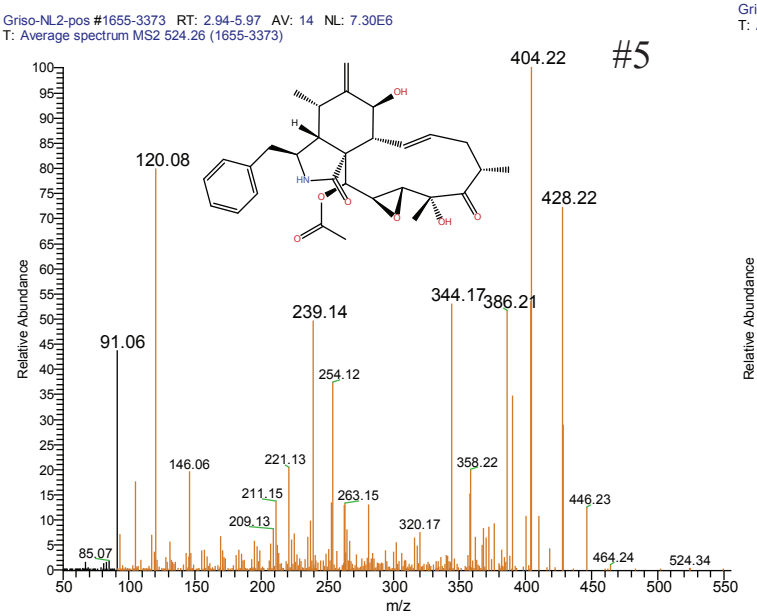
Griso-NL2-pos #1733-3385 RT: 3.07-5.99 AV: 7 NL: 2.67E6
T: Average spectrum MS2 508.27 (1733-3385)



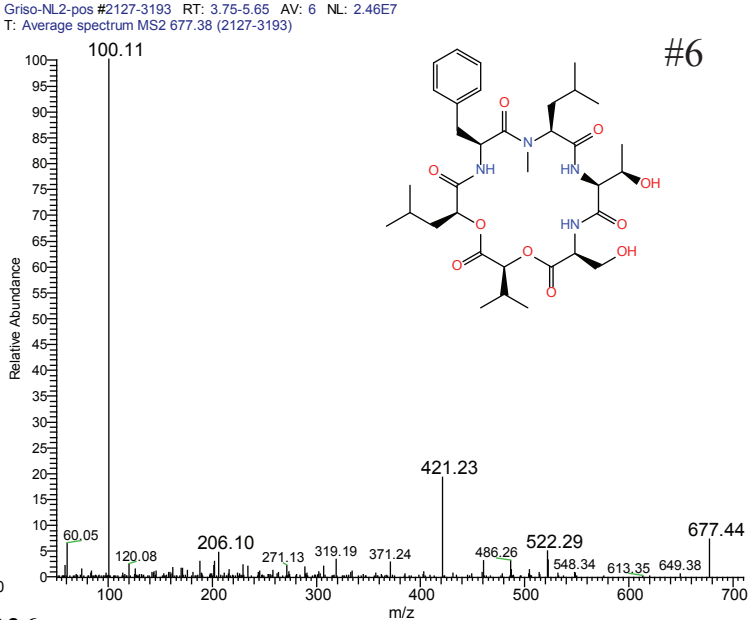
Griso-NL2-pos #2012-2169 RT: 3.55-3.83 AV: 2 NL: 1.57E7
T: Average spectrum MS2 492.27 (2012-2169)



Griso-NL2-pos #1655-3373 RT: 2.94-5.97 AV: 14 NL: 7.30E6
T: Average spectrum MS2 524.26 (1655-3373)



Griso-NL2-pos #2127-3193 RT: 3.75-5.65 AV: 6 NL: 2.46E7
T: Average spectrum MS2 677.38 (2127-3193)



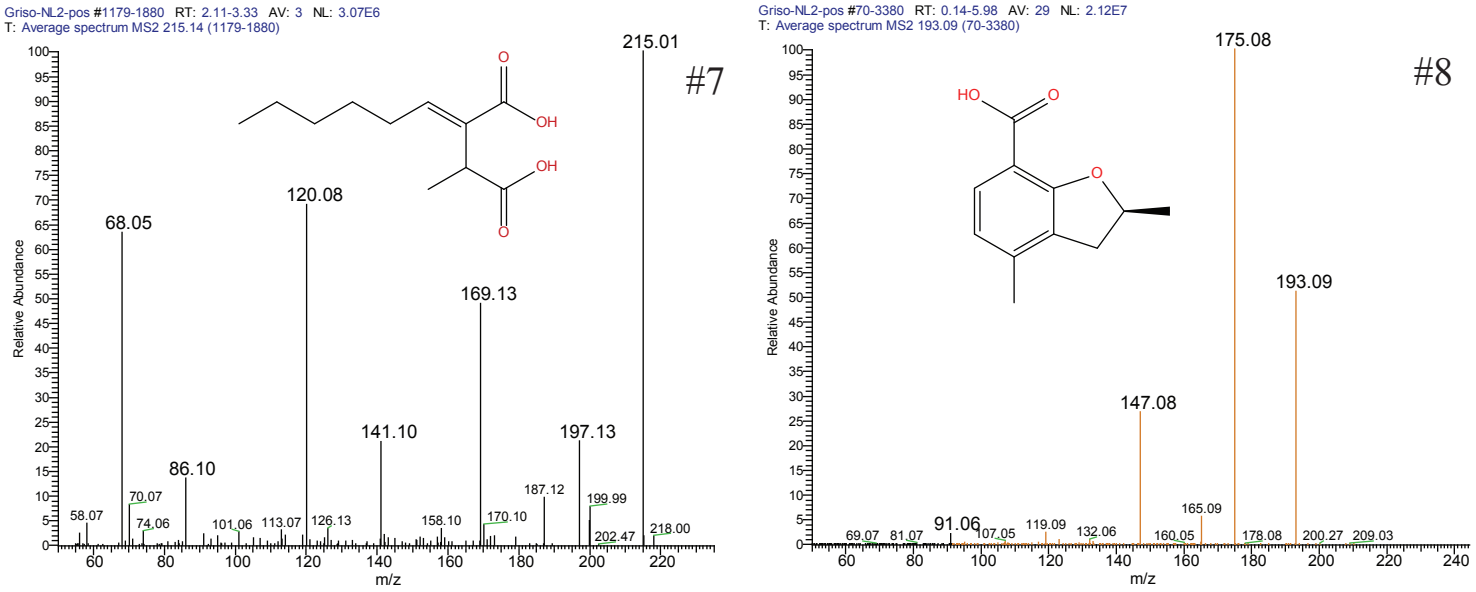


Figure 3S LC-MS/MS spectra of compounds 1-8.

Table 8S A. Cyclic nonribosomal peptide HRMS and peptide monomer sequences. Total of nine peptides have been identified, of these, seven are new secondary metabolites.

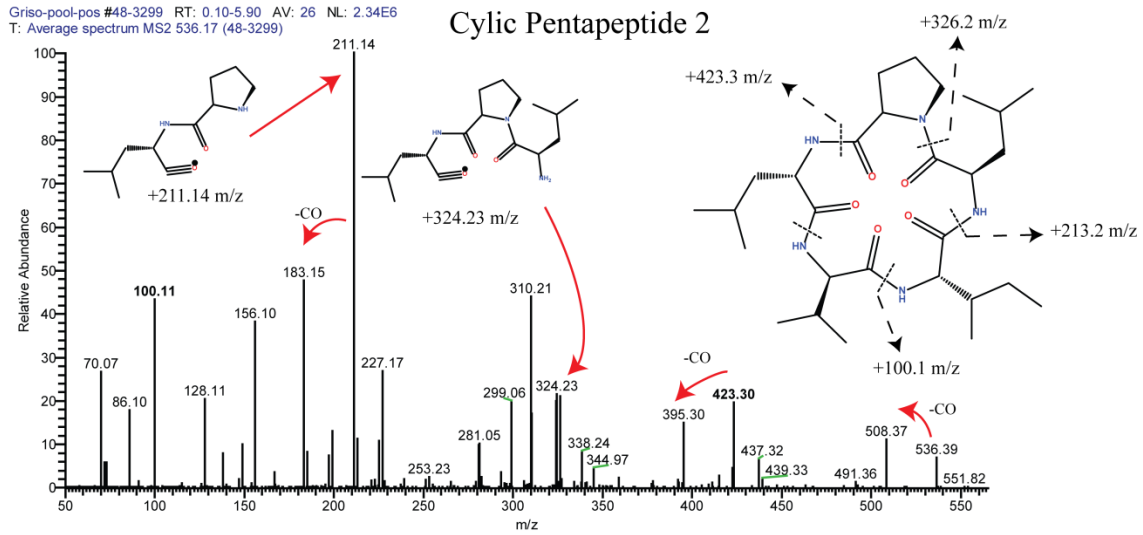
Metabolite	Rt (min)	Formula	Measured * [M+H]+	Calculated [M+H]+	ppm error	Peptide Sequence (MS/MS) 1	2	3	4	5
**Cyclic Pentapeptide 1	16.20	$C_{32}H_{50}N_5O_5$	584.3816	584.3806	-1.71	N-Me-Phe	Val	Ile	Leu	Pro
Cyclic Pentapeptide 2	14.28	$C_{28}H_{50}N_5O_5$	536.3819	536.3806	-2.42	Leu	Val	Ile	Leu	Pro
Xylarotide A	15.97	$C_{29}H_{52}N_5O_5$	550.3973	550.3963	-1.82	N-Me-Leu	Val	Ile	Leu	Pro
**Cyclic Pentapeptide 3	14.76	$C_{30}H_{46}N_5O_5$	556.3501	556.3493	-1.44	N-Me-Phe	Ala	Ile	Leu	Pro
**Cyclic Pentapeptide 4	15.19	$C_{31}H_{48}N_5O_5$	570.3656	570.3650	-1.05	N-Me-Phe	Val	Val	Leu	Pro
**Cyclic Pentapeptide 5	17.04	$C_{33}H_{52}N_5O_5$	598.3968	598.3963	-0.84	N-Me-Phe	Ile	Ile	Leu	Pro
Cyclic Pentapeptide 6	14.47	$C_{27}H_{48}N_5O_5$	522.3662	522.3650	-2.30	Leu	Val	Val	Leu	Pro
Cyclic Pentapeptide 7	16.89	$C_{30}H_{54}N_5O_5$	564.4132	564.4119	-2.30	N-Me-Leu	Ile	Ile	Leu	Pro
Cyclic Pentapeptide 8	14.11	$C_{31}H_{48}N_5O_6$	586.3616	586.3599	-1.72	N-Me-Phe	Val	Val	Leu	Pro+16
Cyclic Pentapeptide 9	14.00	$C_{32}H_{50}N_5O_6$	600.3768	600.3756	-2.00	N-Me-Phe	Val	Ile	Leu	Pro+16
Cyclic Pentapeptide 10	14.89	$C_{33}H_{52}N_5O_6$	614.3936	614.3912	-2.41	N-Me-Phe	Ile	Ile	Leu	Pro+16

* HRMS and MSMS measurements performed on modified HPLC gradient for improved metabolite resolution.
** Structural characterization by comprehensive 1D and 2D NMR and MS/MS analysis

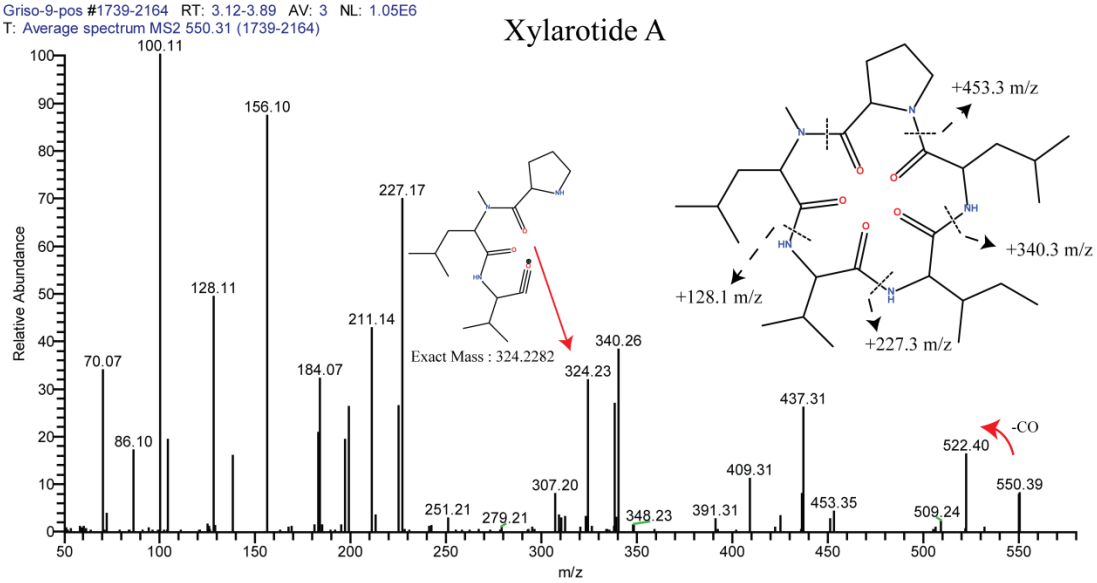
Table 8S B. Optical rotation measurments for the isolated cyclic pentapeptides 1, 3-5.

Compound	Wt. meas. (mg)	Wt. g/100ml	Solvent	Vol. (mL)	Spec. Rot. ^{o}	StDev	Temp ^{o}C
Cyclic Pentapeptide 1	4.79	0.177	MeOH	2.7	-63.356	0.1932	20.9
Cyclic Pentapeptide 3	1.59	0.059	MeOH	2.7	-86.138	0	21.1
Cyclic Pentapeptide 4	1.03	0.038	MeOH	2.7	-43.101	0.8832	20.2
Cyclic Pentapeptide 5	1.5	0.056	MeOH	2.7	-47.802	0	21.5

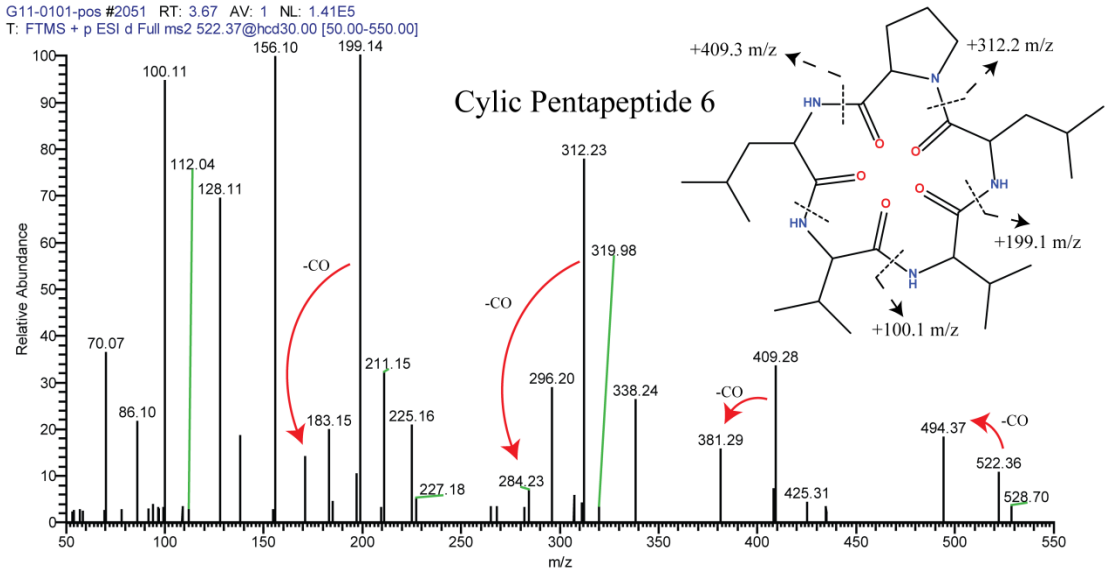
A


Griso-pool-pos #48-3299 RT: 0.10-5.90 AV: 26 NL: 2.34E6
T: Average spectrum MS2 536.17 (48-3299)
Cylic Pentapeptide 2
+326.2 m/z
+423.3 m/z
+213.2 m/z
+100.1 m/z
+211.14 m/z
+324.23 m/z
-CO
Relative Abundance
m/z

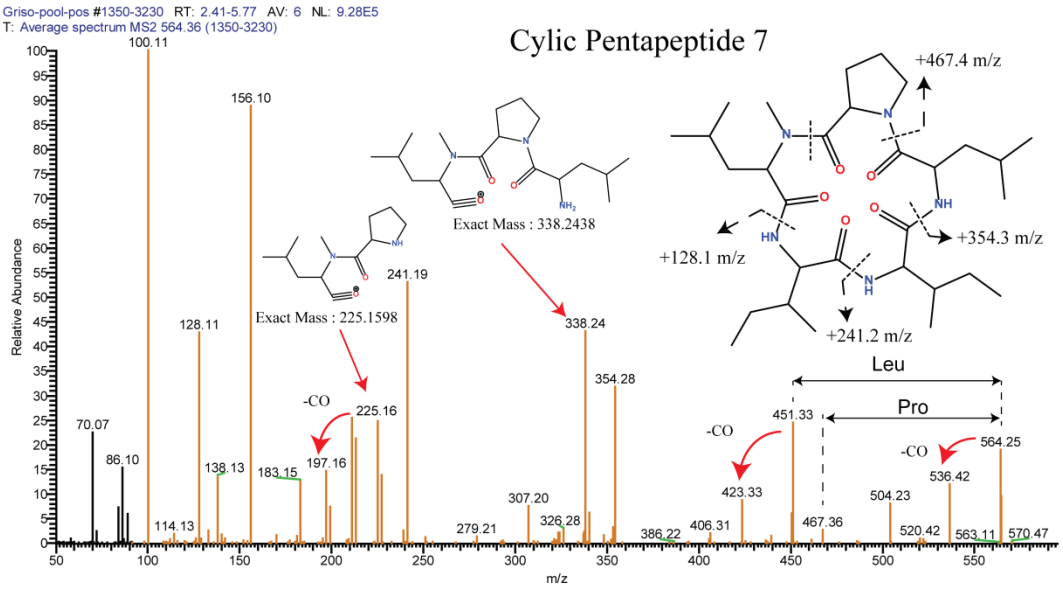
B


Griso-9-pos #1739-2164 RT: 3.12-3.89 AV: 3 NL: 1.05E6
T: Average spectrum MS2 550.31 (1739-2164)
Xylarotide A
+453.3 m/z
+340.3 m/z
+128.1 m/z
+227.3 m/z
Exact Mass : 324.2282
-CO
Relative Abundance
m/z

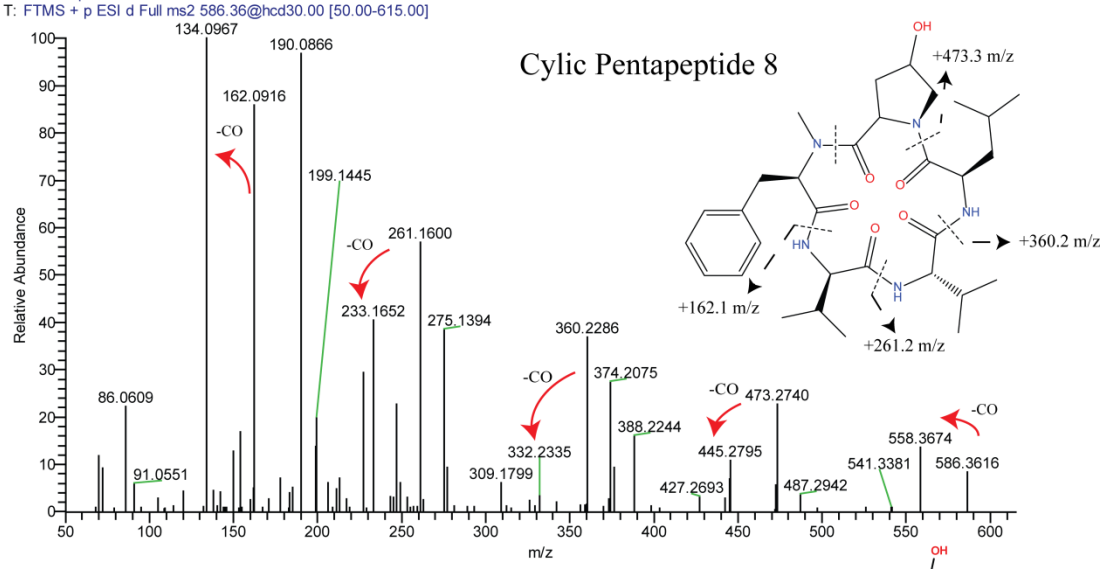
C


G11-0101-pos #2051 RT: 3.67 AV: 1 NL: 1.41E5
T: FTMS + p ESI d Full ms2 522.37@hcd30.00 [50.00-550.00]
Cylic Pentapeptide 6
+409.3 m/z
+312.2 m/z
+199.1 m/z
+100.1 m/z
-CO
Relative Abundance
m/z

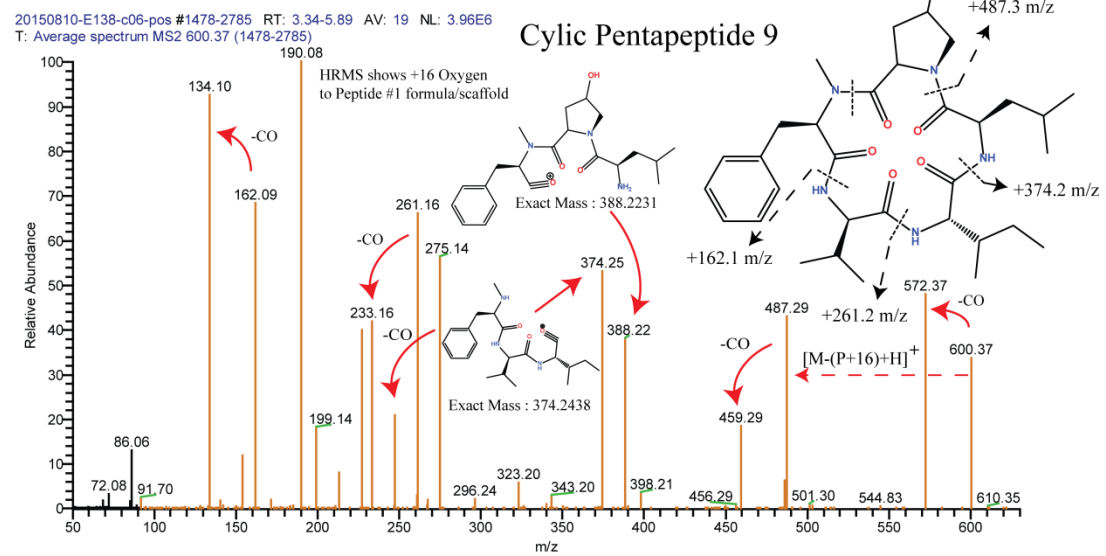
D


Griso-pool-pos #1350-3230 RT: 2.41-5.77 AV: 6 NL: 9.28E5
T: Average spectrum MS2 564.36 (1350-3230)
Cylic Pentapeptide 7
Relative Abundance
m/z
100.11
156.10
241.19
128.11
70.07
86.10
138.13
183.15
197.16
225.16
114.13
279.21
307.20
326.28
338.24
354.28
386.22
406.31
423.33
451.33
467.36
504.23
520.42
536.42
563.11
564.25
570.47
Exact Mass : 338.2438
Exact Mass : 225.1598
-CO
+467.4 m/z
+354.3 m/z
+128.1 m/z
+241.2 m/z
Leu
Pro

E


Griso-NL2-pos #1908 RT: 3.37 AV: 1 NL: 1.03E6
T: FTMS + p ESI d Full ms2 586.36@hcd30.00 [50.00-615.00]
Cylic Pentapeptide 8
Relative Abundance
m/z
134.0967
190.0866
162.0916
199.1445
261.1600
233.1652
275.1394
86.0609
91.0551
309.1799
332.2335
360.2286
374.2075
388.2244
427.2693
445.2795
473.2740
487.2942
541.3381
558.3674
586.3616
-CO
+473.3 m/z
+360.2 m/z
+162.1 m/z
+261.2 m/z

F


20150810-E138-c06-pos #1478-2785 RT: 3.34-5.89 AV: 19 NL: 3.96E6
T: Average spectrum MS2 600.37 (1478-2785)
Cylic Pentapeptide 9
HRMS shows +16 Oxygen
to Peptide #1 formula/scaffold
Relative Abundance
m/z
190.08
134.10
162.09
261.16
275.14
233.16
199.14
374.25
388.22
86.06
72.08
91.70
296.24
323.20
343.20
398.21
456.29
459.29
487.29
501.30
544.83
572.37
600.37
610.35
Exact Mass : 388.2231
Exact Mass : 374.2438
-CO
+487.3 m/z
+374.2 m/z
+162.1 m/z
+261.2 m/z
[M-(P+16)+H]+

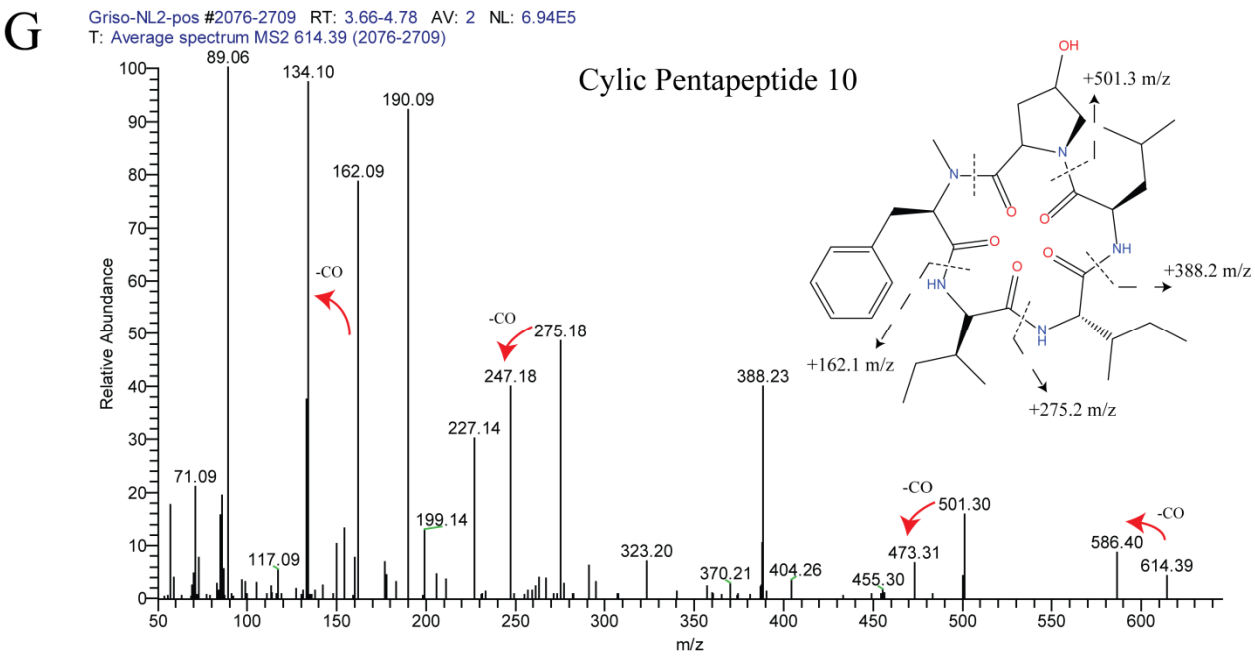


Figure 4S. Q-Exactive MS/MS spectra of the cyclic pentapeptides with key diagnostic ions shown. (A) Cyclic pentapeptide 2, (B) Xylarotide A, (C) Cyclic pentapeptide 6, (D) Cyclic pentapeptide 7, (E) Cyclic pentapeptide 8, (F) Cyclic pentapeptide 9, (G) Cyclic pentapeptide 10.

Table 9S. 1H (700 MHz) and 13C (176 MHz) NMR data and HMBC correlations for cyclic pentapeptide 3 ($DMSO_{d6}$)

Amio acid	position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
L-*N*-methyl phenylalanine	CO	170.6		
	α	55.8	5.08 (dd, 11.9, 5.1)	*β, N*–CH3*, γ,* CO
	β	33.5	3.08 (m)	
			2.9 (dd, 14.8, 5.1)	
	γ	137.4		
	ortho	128.4	7.21 (m)	*meta, para, β*
	meta	128.1	7.27 (m)	*ortho, γ*
	para	126.4	7.20 (m)	*ortho*
	N-CH3	30.2	3.04 (s)	α, CO-L-Pro
L-alanine	CO	171.9		
	α	46.0	4.49 (dq, 8.9, 6.9)	*β,* CO, CO-N-MePhe
	β	14.2	1.1 (d, 7.0)	α, CO
	NH		8.52 (d, overlap)	α, CO-N-MePhe
D-isoleucine	CO	170.1		
	α	55.7	4.17 (dd, 8.3, 5.2)	*β, β*–CH3*, γ,* CO, CO-L-Ala
	β	38.5	1.50 (ddt, 14.1, 7.2, 3.5)	*α, β*–CH3*, δ,* CO
	γ	25.8	1.31 (m)	*α, β, β*–CH3*, δ*
			0.99 (dt, 13.4, 7.4)	
	β - CH3	14.4	0.72 (d, 6.8)	*α, β, γ,*
	δ	11.6	0.85 (t, 7.4)	*β, γ,*
	NH		6.94 (d,8.3)	α, CO, CO-L-Ala
L-leucine	CO	168.8		
	α	46.5	4.74 (m)	*β, γ,* CO, CO-L-Pro
	β	41.1	1.45 (m)	*α, γ, δ,* CO
			1.40 (m)	
	γ	24.2	1.39 (m)	*α, β, δ*
	δ	22.2	0.79 (d, 5.9)	*β, γ, δ*
		22.9	0.82 (d, 6.1)	
	NH		8.49 (d, overlap)	α, CO-D-Ile
L-proline	CO	172.2		
	α	58.9	5.10	*β, γ, δ,* CO, N-MePhe
	β	30.0	1.83 (tt, 12.1, 8.0)	*α, γ,* CO
			0.90 (dd, 12.1, 7.4)	
	γ	20.32	1.6 (m)	*α, β,*
			1.26 (m)	
	δ	45.7	3.27 (m)	*α, β, γ,* CO-Leu

Table 10S. 1H (850 MHz) and 13C (176 MHz) NMR data and HMBC correlations cyclic pentapeptide 4 ($DMSO_{d6}$)

Amio acid	position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
L-*N*-methyl phenylalanine	CO	170.5		
	α	56.0	5.10 (dd, 10.9, 6.0)	*β, N*–CH3*, γ,* CO
	β	33.8	3.05 (m)	α, γ, *ortho*, CO
			2.90 (dd, 14.5, 6.0)	
	γ	137.3		
	ortho	128.6	7.21 (m)	*meta, para, β*
	meta	128.1	7.26 (m)	*ortho, γ*
	para	126.4	7.19 (td, 7.1, 1.4)	*ortho*
	N-CH3	30.2	3.03 (s)	α, CO-L-Pro
L-valine	CO	170.8		
	α	57.6	3.95 (m)	*β, γ,* CO
	β	26.4	1.97 (dt, 10.5, 6.7)	α, γ, CO
	γ	18.5	0.73 (d, 6.7)	*α, β,*
		19.7	0.84 (d, 6.5)	
	NH		8.18 (d, 9.6)	α, CO-N-MePhe
D-valine	CO	170.1		
	α	56.8	4.10 (dd, 8.7, 6.7)	*β, γ,* CO
	β	31.3	1.74 (h, 6.8)	α, γ, CO
	γ	17.9	0.74 (d, 6.8)	*α, β,*
		19.2	0.78 (m)	
	NH		6.98 (d, 8.8)	α, CO-L-Val
L-leucine	CO	168.7		
	α	46.6	4.72 (q, 4.7, 7.2)	*β, γ,* CO
	β	41.4	1.46 (m)	*α, γ, δ,* CO
			1.38 (m)	
	γ	24.2	1.39 (m)	*α, β, δ,* CO
	δ	22.8	0.82 (d, 6.4)	*β, γ, δ*
		22.4	0.79 (m)	
	NH		8.43 (d, 9.4)	α, CO-D-Val
L-proline	CO	171.9		
	α	58.7	5.08 (d, 7.9)	*β, γ, δ,* CO, N-MePhe
	β	30.1	1.86 (tt, 12.1, 8.1)	*α, γ, δ,* CO, N-MePhe
			0.98 (dd, 12.1, 7.4)	
	γ	20.4	1.62 (dq, 13.1, 7.0, 6.6)	*α, β,*
			1.31 (m)	
	δ	45.71	3.30 (dd, 9.1, 5.9)	*α, β, γ,* CO-Leu

Table 11S. 1H (700 MHz) and 13C (176 MHz) NMR data and HMBC correlations for cyclic pentapeptide 5 ($DMSO_{d6}$)

Amio acid	position	δ_C	$\delta_{H\ (J\ in\ Hz)}$	HMBC
L-*N*-methyl phenylalanine	CO	170.3		
	α	56.0	5.1 (m)	*β, N*–CH3*, γ,* CO
	β	33.8	3.03 (m)	α, γ, *ortho*, CO
			2.90 (dd, 14.4, 6.1)	
	γ	137.3		
	ortho	128.6	7.21 (m)	*meta, para, β*
	meta	128.1	7.26 (m)	*ortho, γ*
	para	126.3	7.19 (m)	*ortho*
	N-CH3	30.2	3.04 (s)	α, CO-L-Pro
L- isoleucine	CO	171.0		
	α	55.8	4.05 (dd, 10.8, 9.6)	*β, β*–CH3*, γ,* CO
	β	32.0	1.82 (m)	α
	γ	24.0	1.30 (m)	*β,β*–CH3*, δ*
			0.92 (m)	
	β - CH3	15.7	0.79 (m)	*α, β, γ,*
	δ	10.1	0.74 (t, 7.4)	*β, γ,*
	NH		8.12 (d, 9.6)	α, CO-N-MePhe
D-isoleucine	CO	170.3		
	α	55.1	4.25 (dd, 8.8, 5.4)	*β, β*–CH3*, γ,* CO
	β	37.7	1.55 (p, 6.5)	*α, β*–CH3*, δ,* CO
	γ	25.9	1.27 (m)	*α, β, β*–CH3*, δ*
			1.01 (m)	
	β - CH3	14.3	0.7 (6.8)	*α, β, γ,*
	δ	11.5	0.85 (t, 7.4)	*β, γ,*
	NH		6.94 (d, 8.8)	α, CO-L-Ile
L-leucine	CO	168.8		
	α	46.5	4.72 (td, 9.1, 8.5, 6.5)	*β, γ,* CO
	β	41.2	1.44 (m)	*α, γ, δ,* CO
			1.41 (m)	
	γ	24.2	1.4 (m)	*α, β, δ*
	δ	22.2	0.80 (m)	*β, γ, δ*
		22.9	0.83 (d, 6.1)	
	NH		8.45 (d, 9.4)	α, CO-D-Ile
L-proline	CO	171.9		
	α	58.7	5.09 (m)	*β, γ, δ,* CO, N-MePhe
	β	30.0	1.85 (m)	*α, γ, δ,* CO, N-MePhe
			0.98 (dd, 12.6, 7.6)	
	γ	20.4	1.62 (m)	*α, β,*
			1.29 (m)	
	δ	45.7	3.29 (td, 7.8, 7.4, 3.3)	*α, β, γ,* CO-Leu

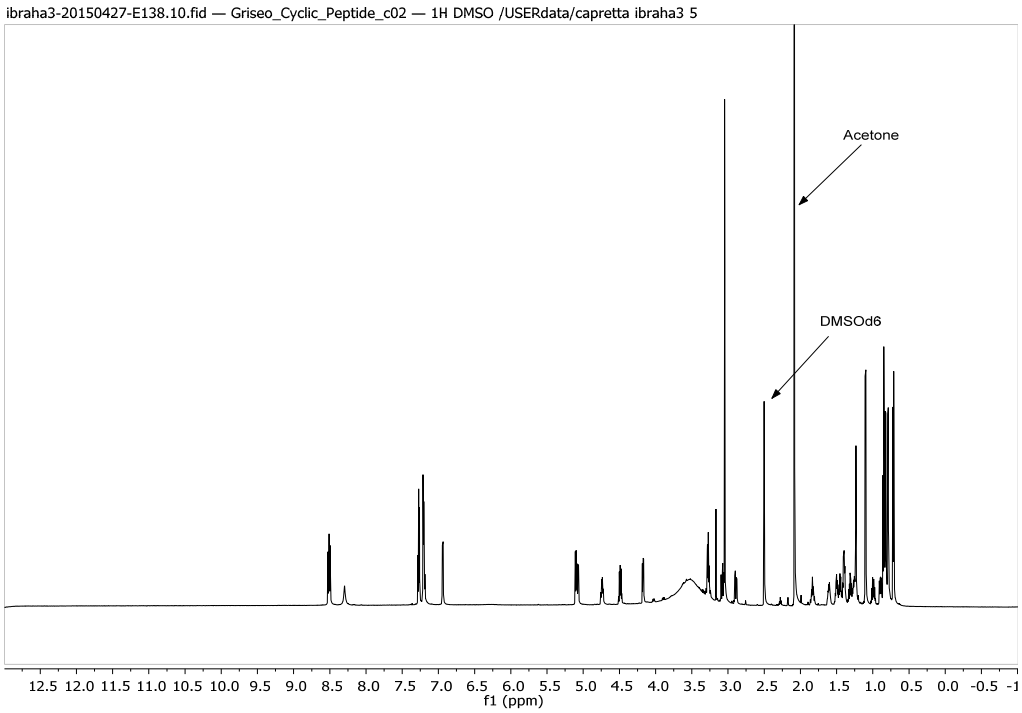


Figure 5S. ^{1}H spectrum of cyclic pentapeptide 3 (700 MHz, $DMSO_{d6}$)

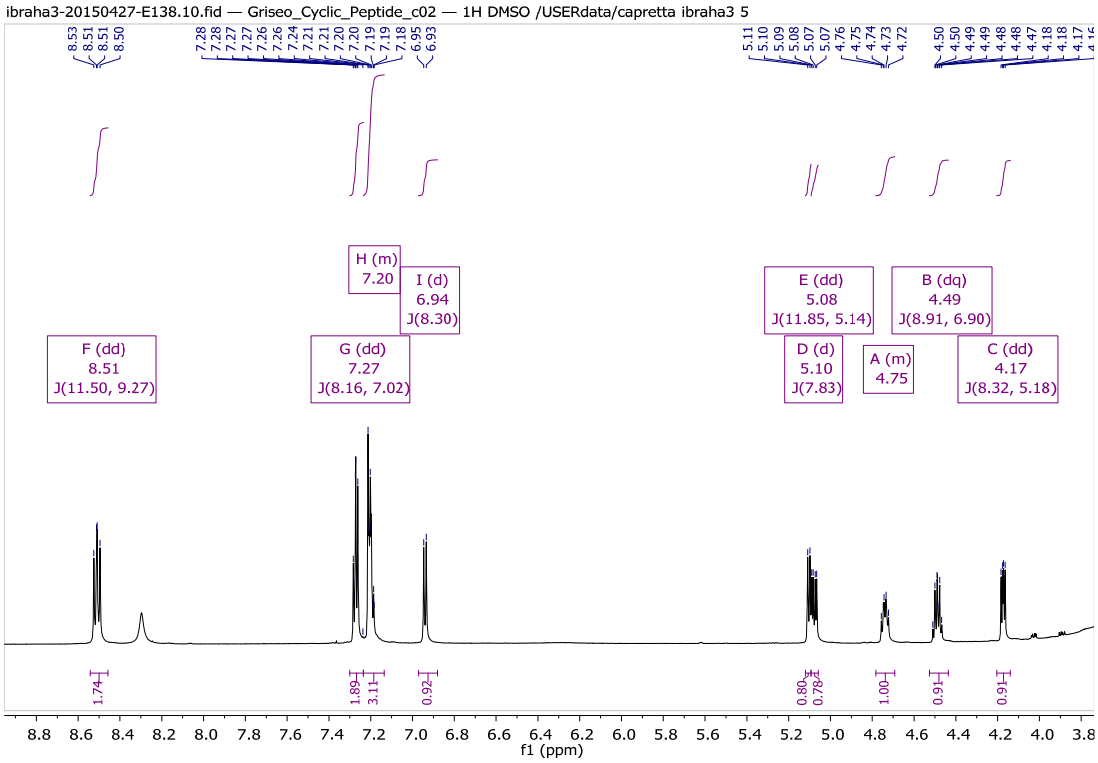


Figure 6S. Expanded ^{1}H spectrum of cyclic pentapeptide 3 (700 MHz, $DMSO_{d6}$).

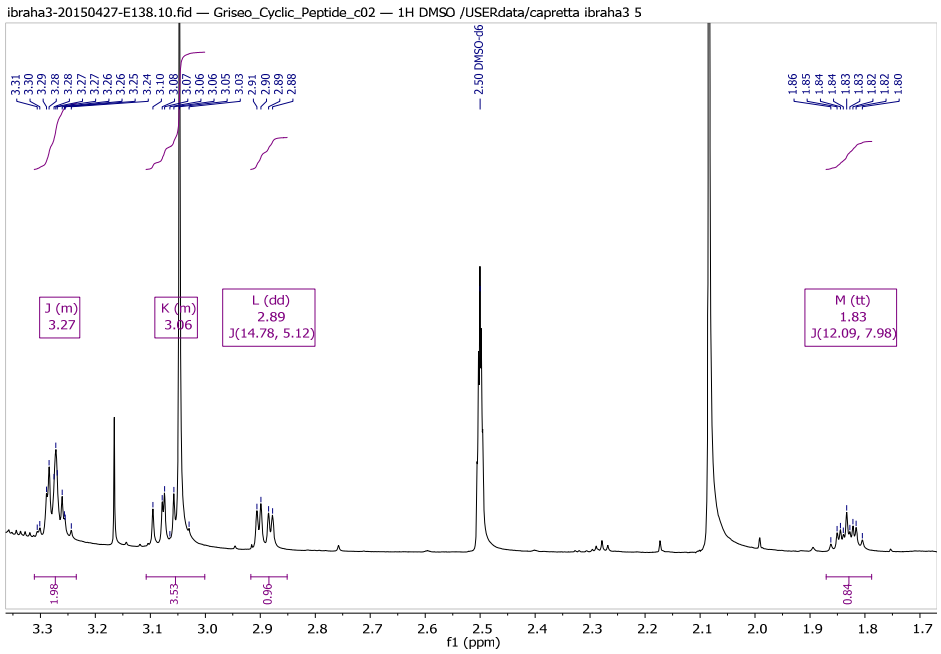


Figure 7S. Expanded ^{1}H spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$).

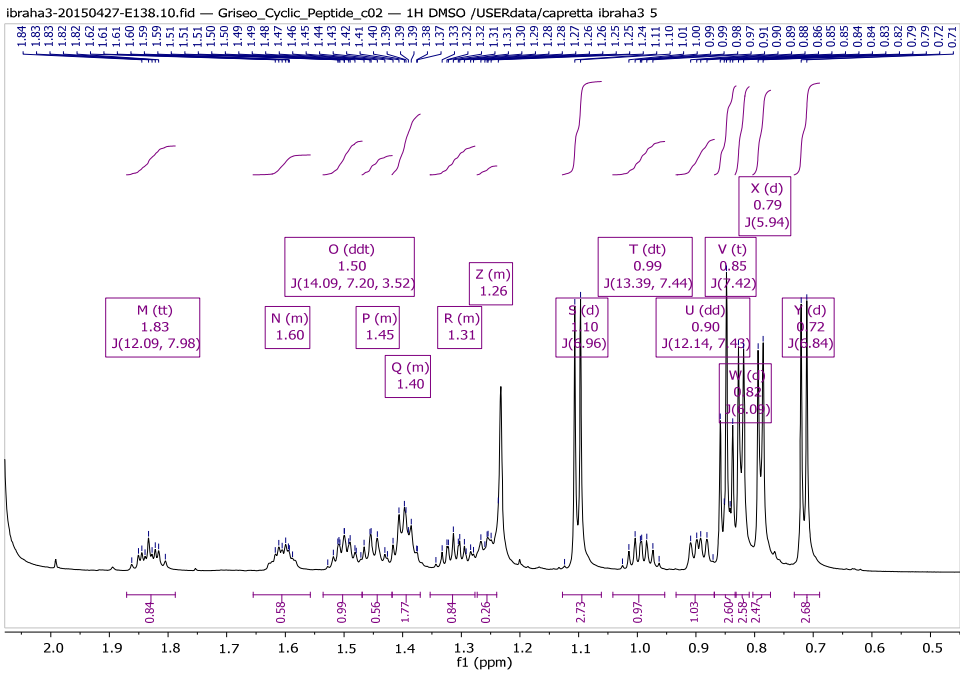


Figure 8S. Expanded ^{1}H spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$).

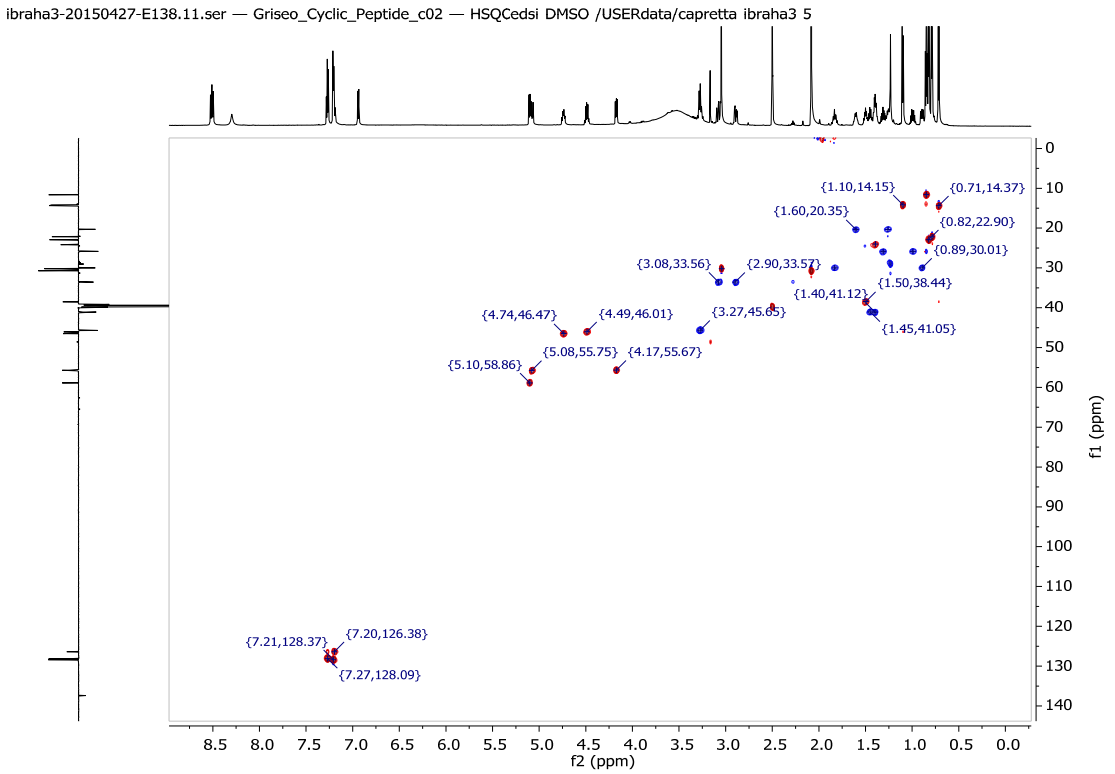


Figure 9S. HSQC spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$), multiplicity edited.

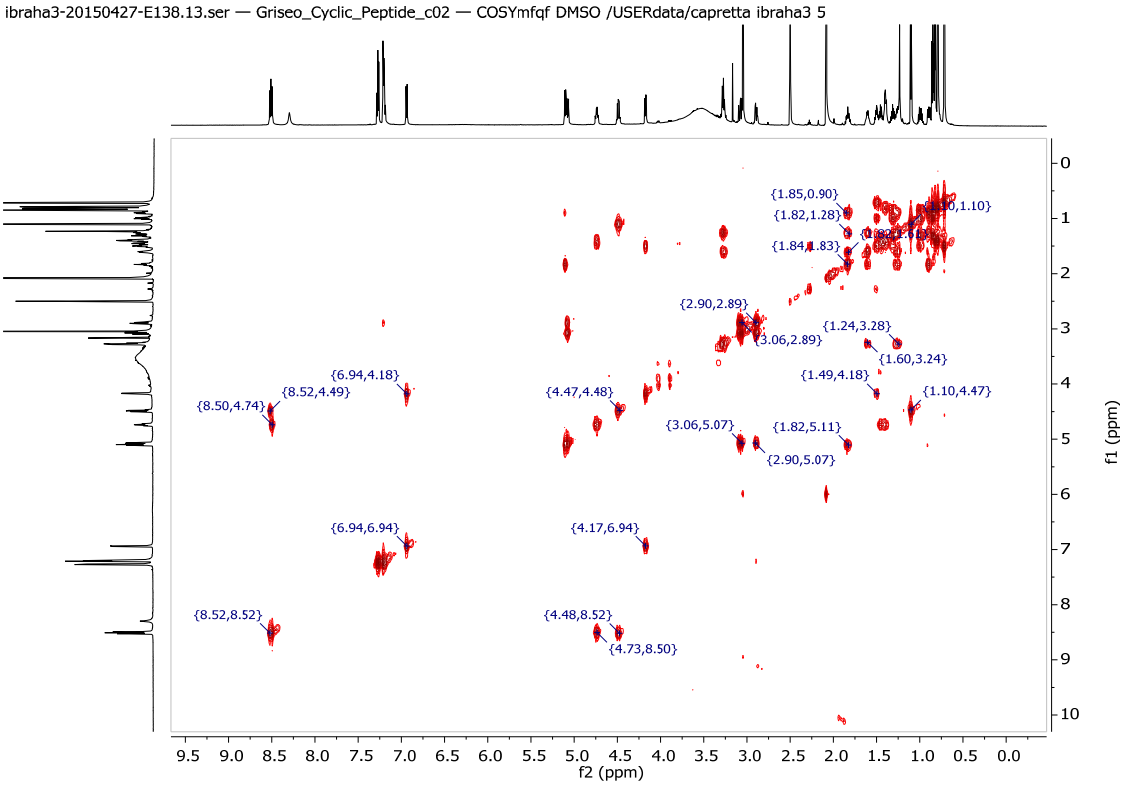


Figure 10S. COSY spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$).

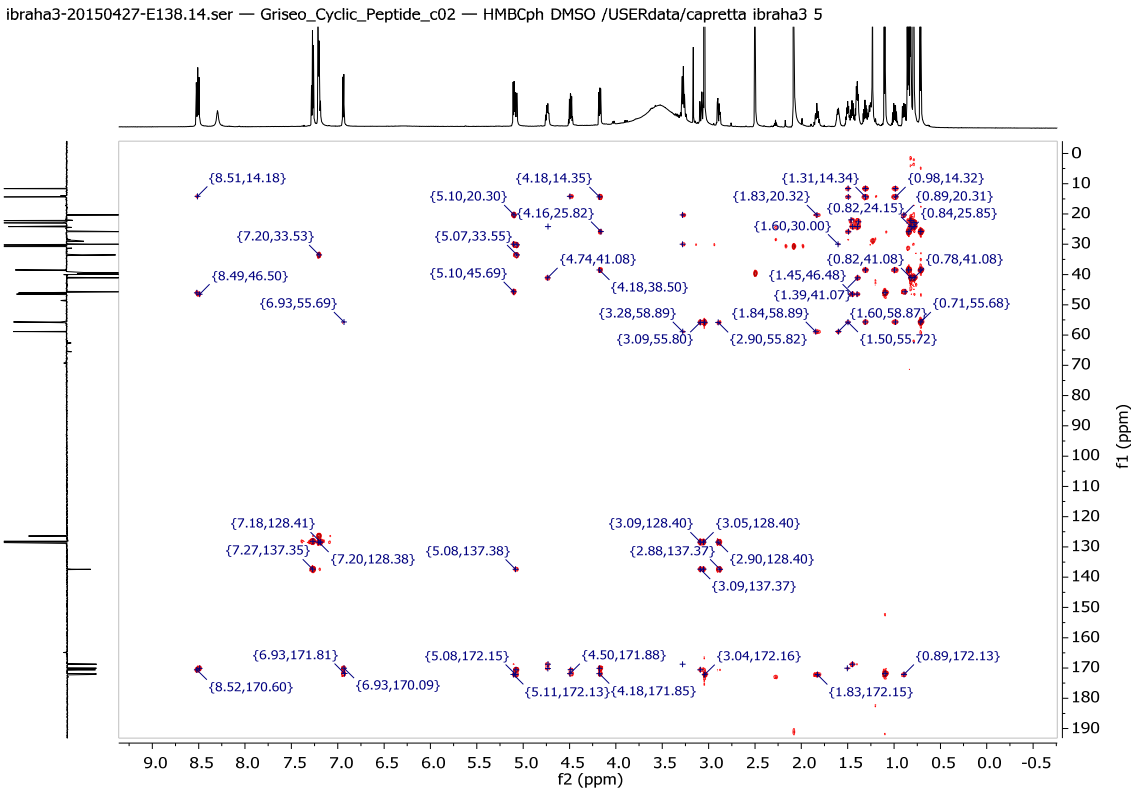


Figure 11S. HMBC spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$).

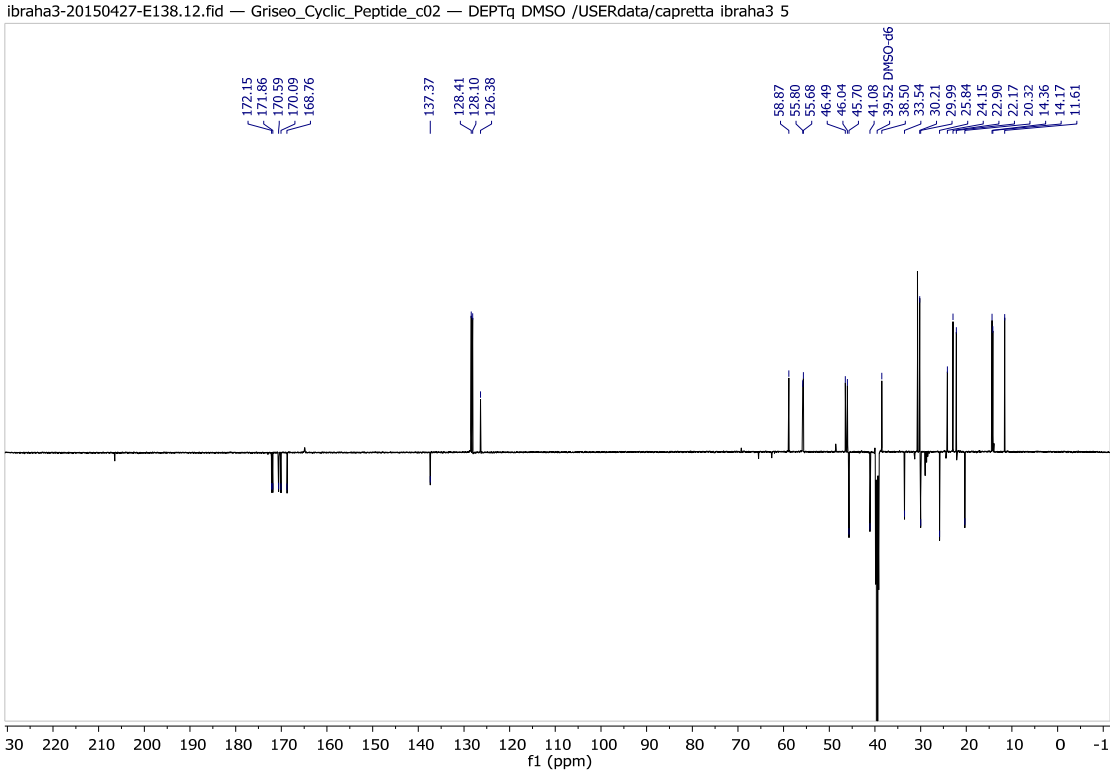


Figure 12S. ^{13}C -DEPTq spectrum of cyclic pentapeptide 3 (176 MHz, DMSO $_{d6}$).

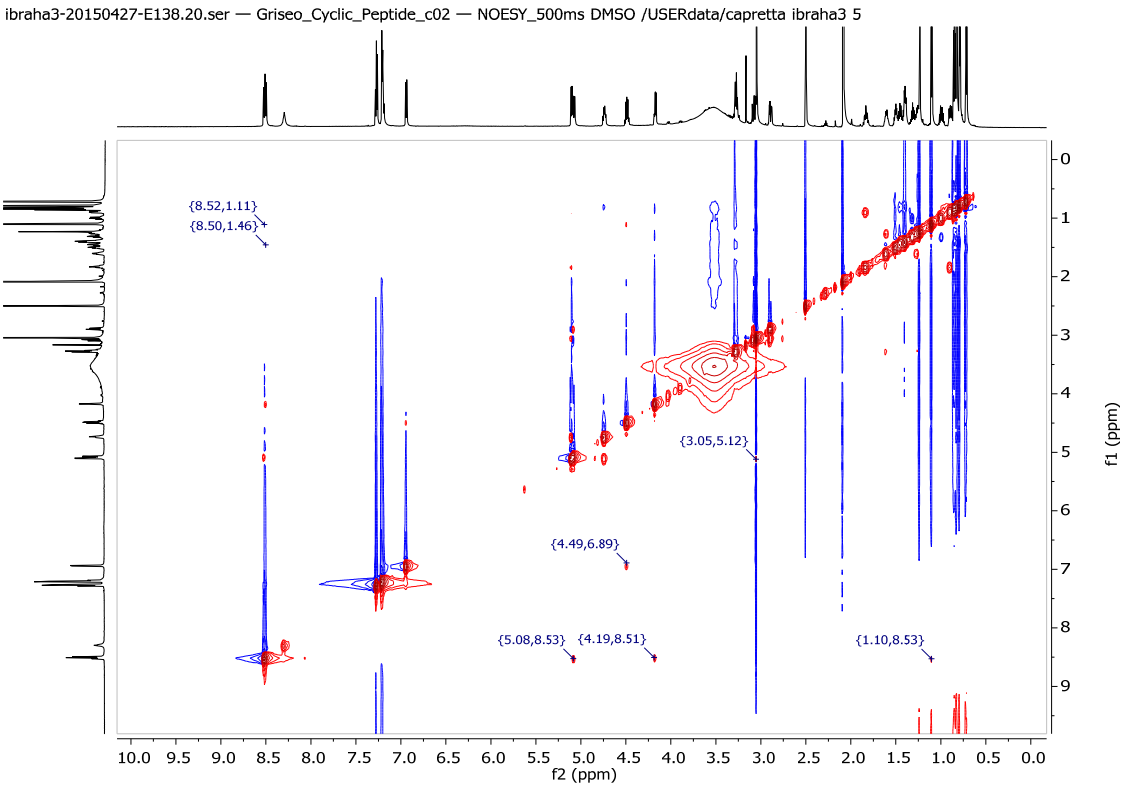


Figure 13S. NOESY spectrum of cyclic pentapeptide 3 (700 MHz, DMSO $_{d6}$).

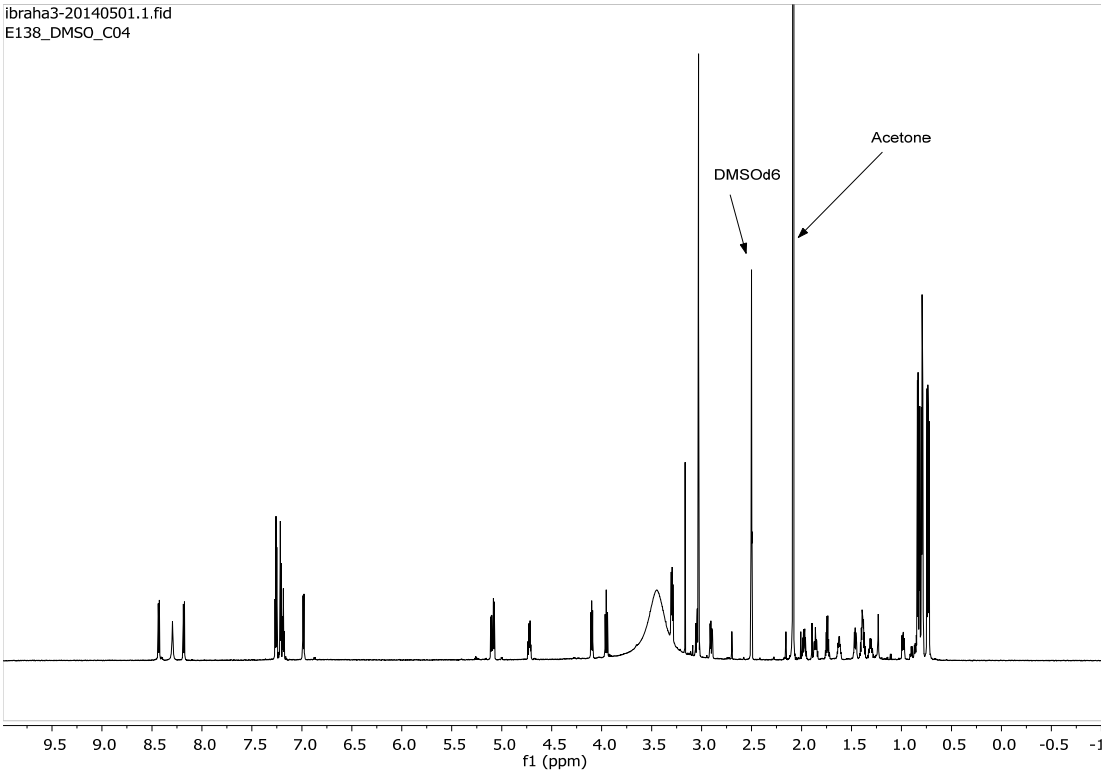


Figure 14S. ^{1}H spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$)

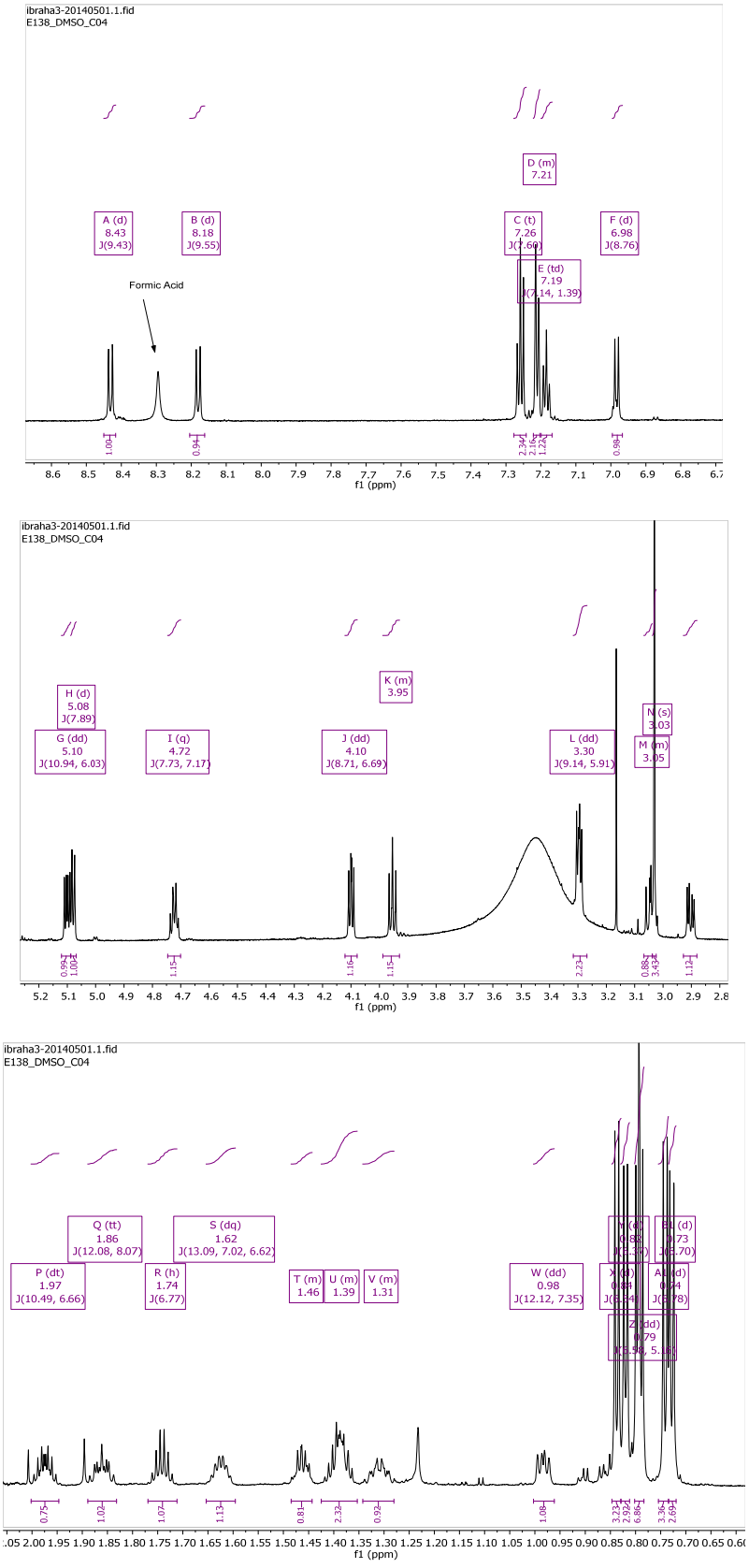


Figure 15S. Expanded ^{1}H spectrums of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$)

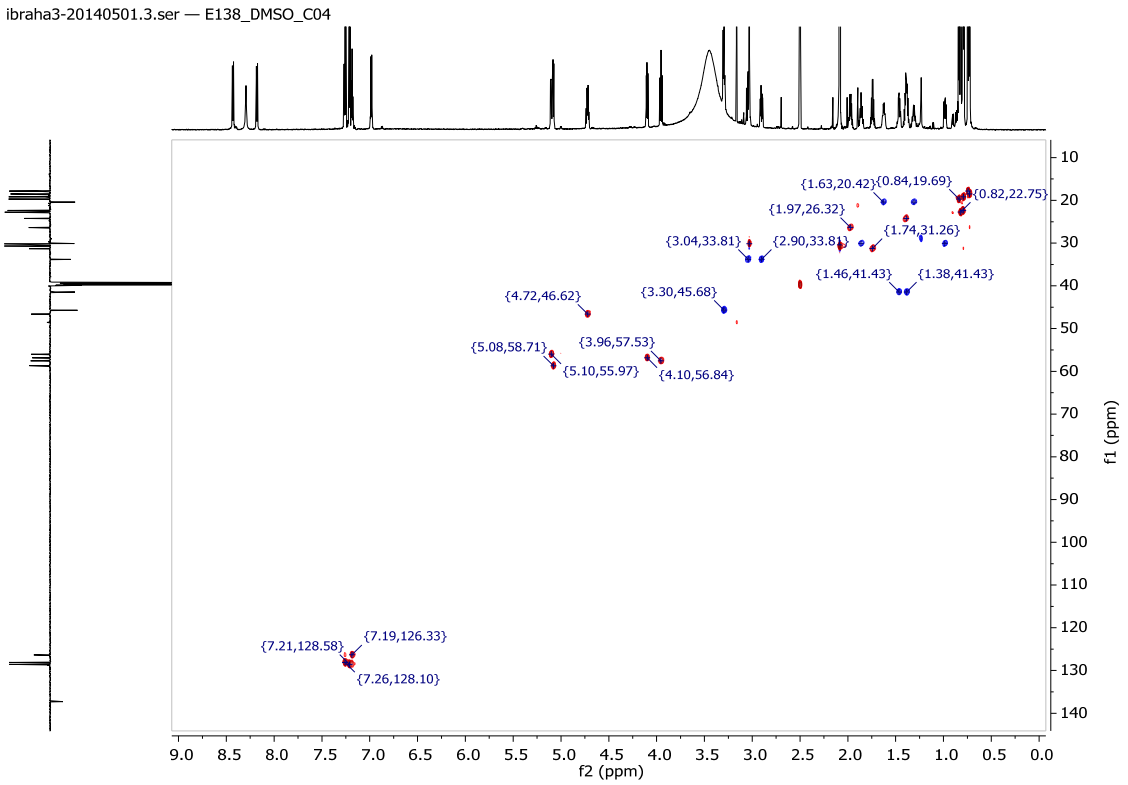


Figure 16S. HSQC spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$), multiplicity edited.

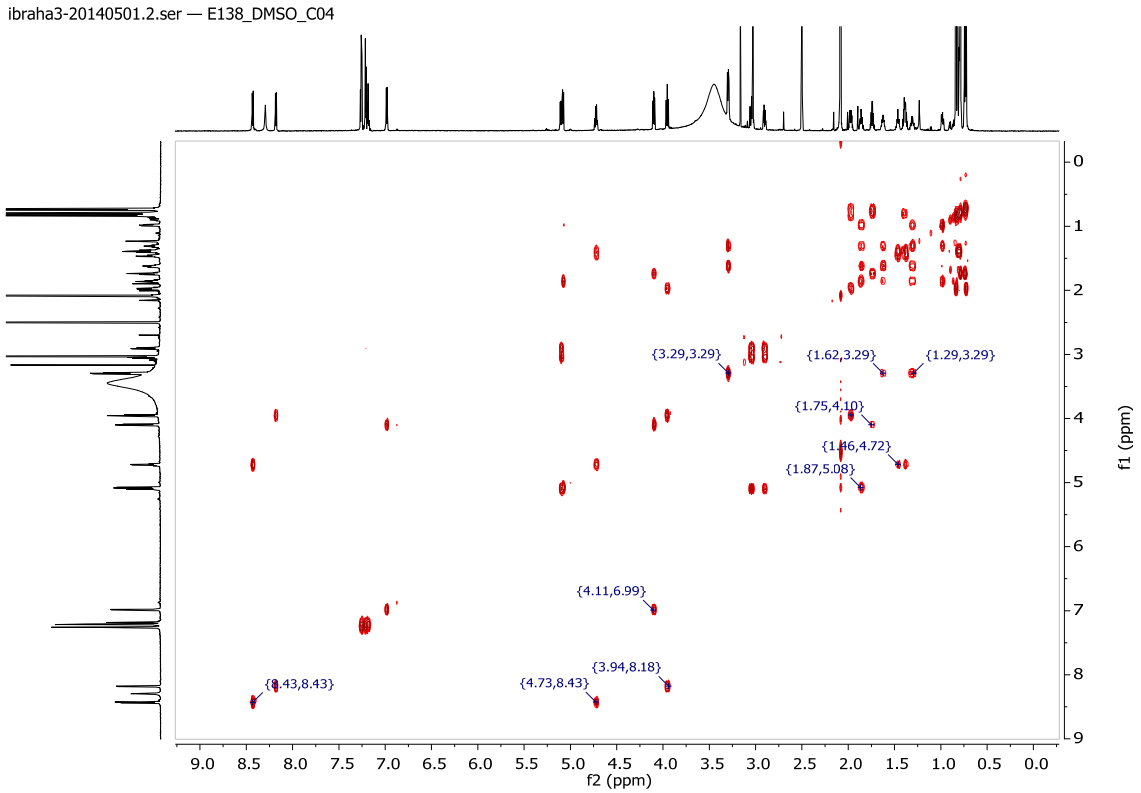


Figure 17S. COSY spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$).

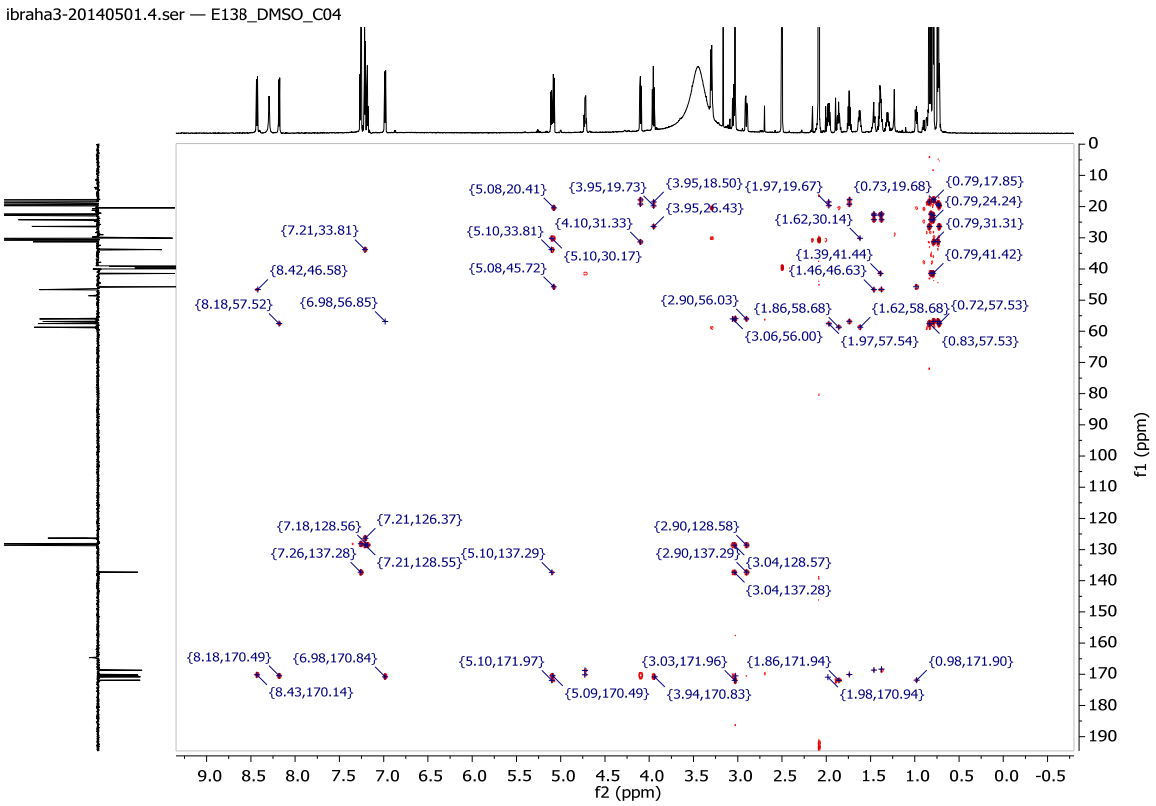


Figure 18S. HMBC spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$).

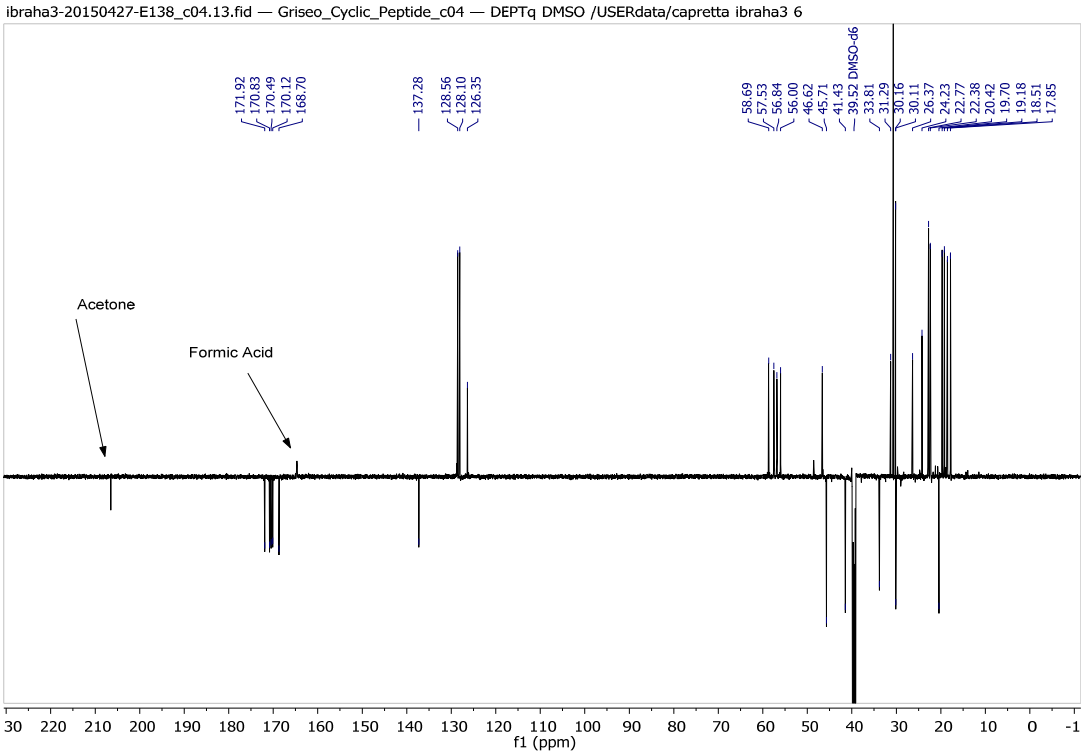


Figure 19S. ^{13}C -DEPTq spectrum of cyclic pentapeptide 4 (176 MHz, DMSO $_{d6}$).

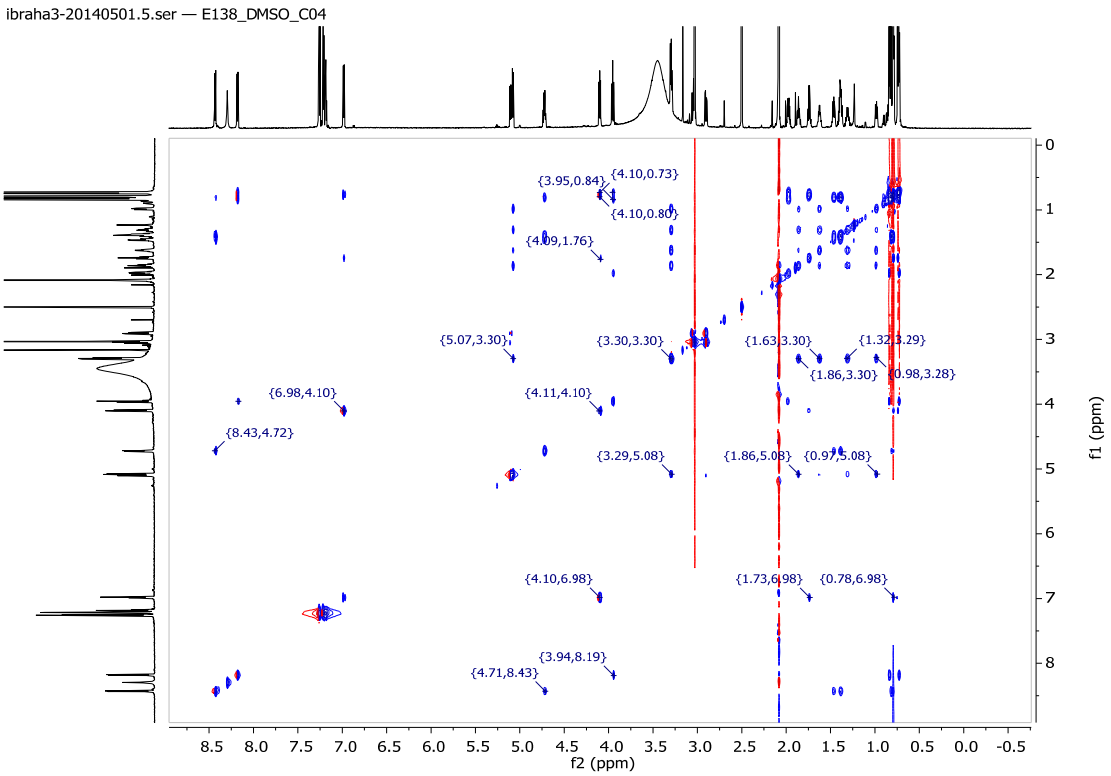


Figure 20S. TOCSY spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$).

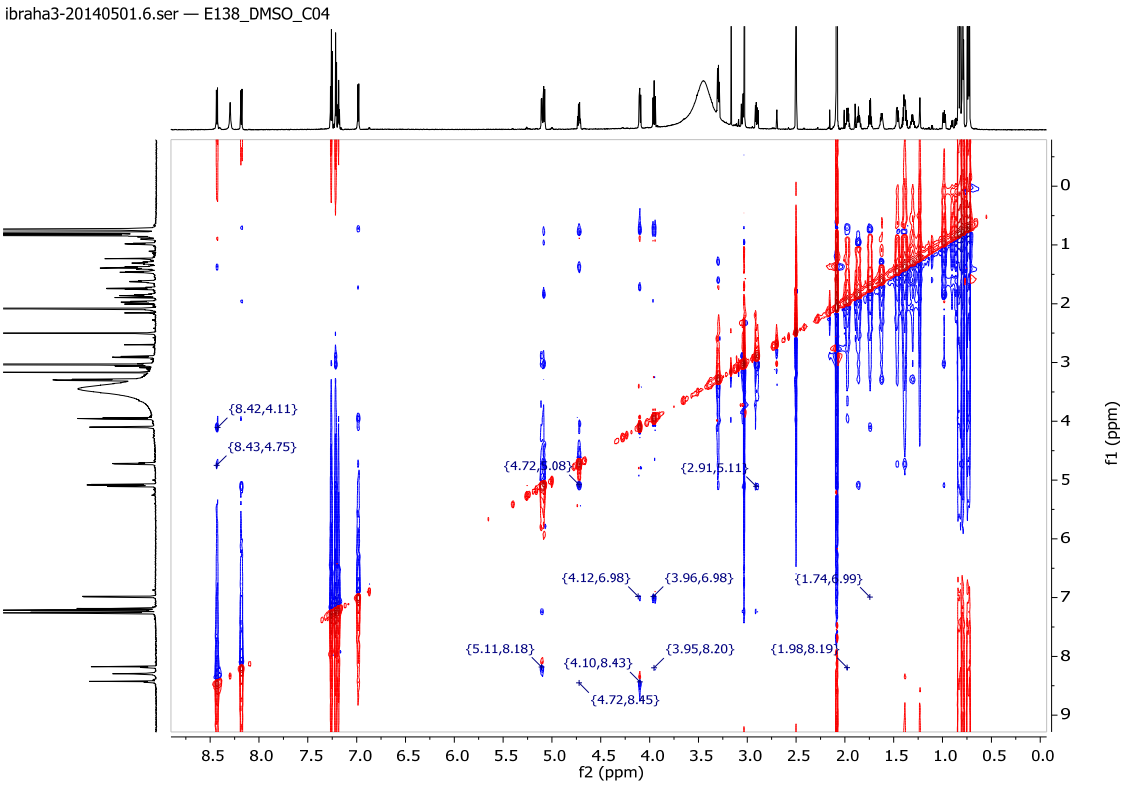


Figure 21S. ROESY spectrum of cyclic pentapeptide 4 (850 MHz, DMSO $_{d6}$).

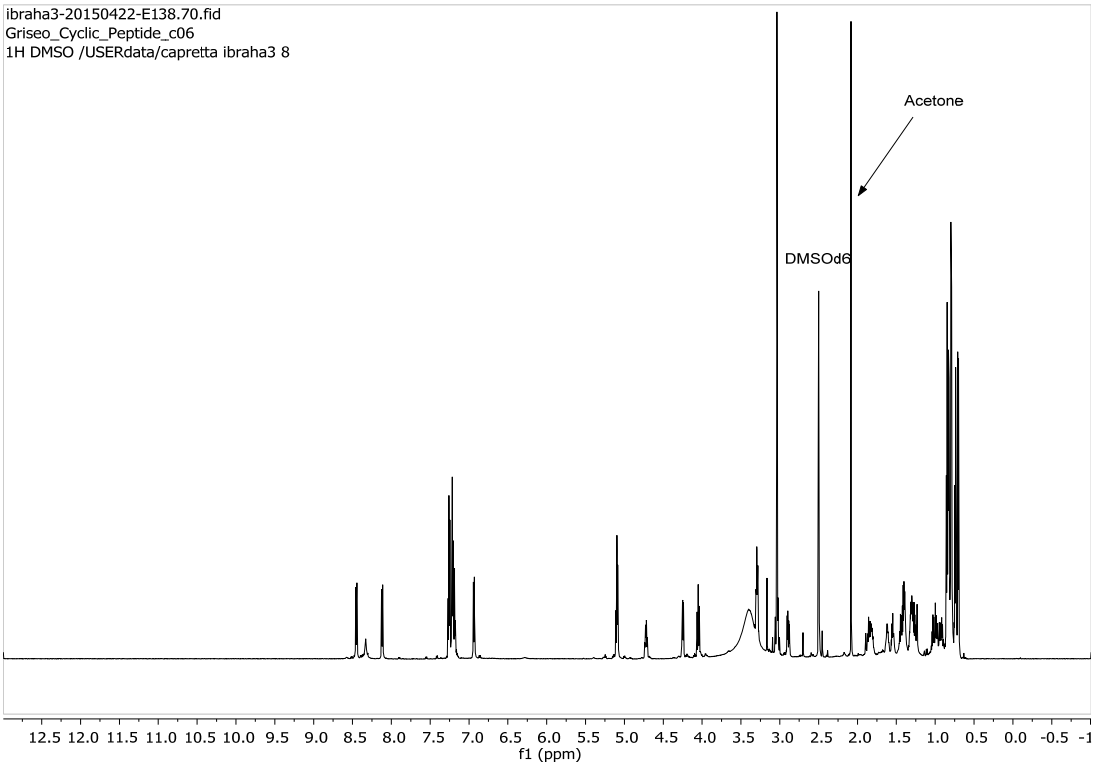


Figure 22S. ^{1}H spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

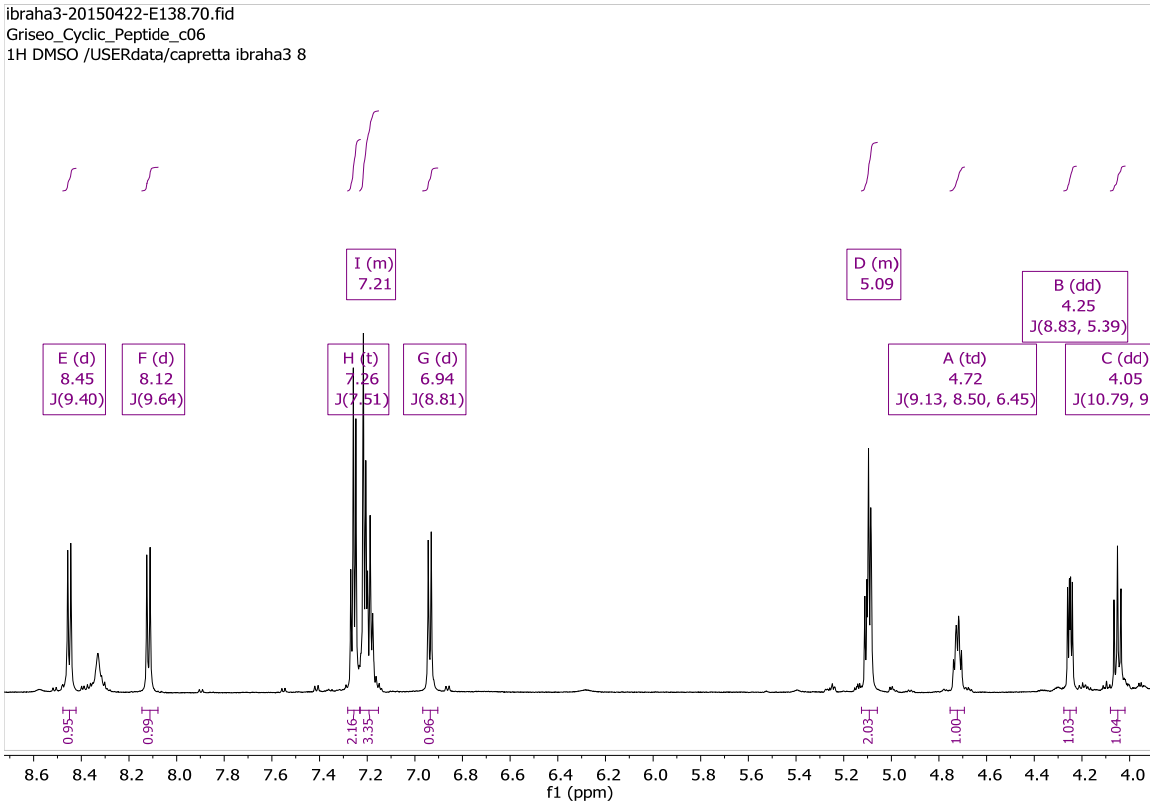


Figure 23S. Expanded ^{1}H spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

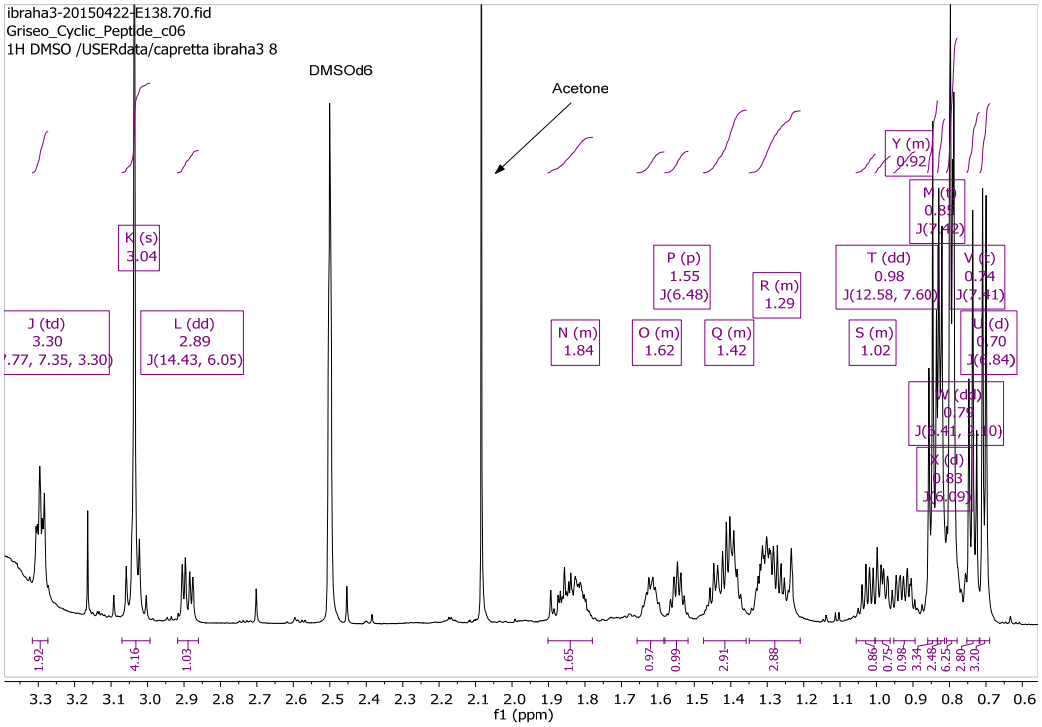


Figure 24S. Expanded ^{1}H spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

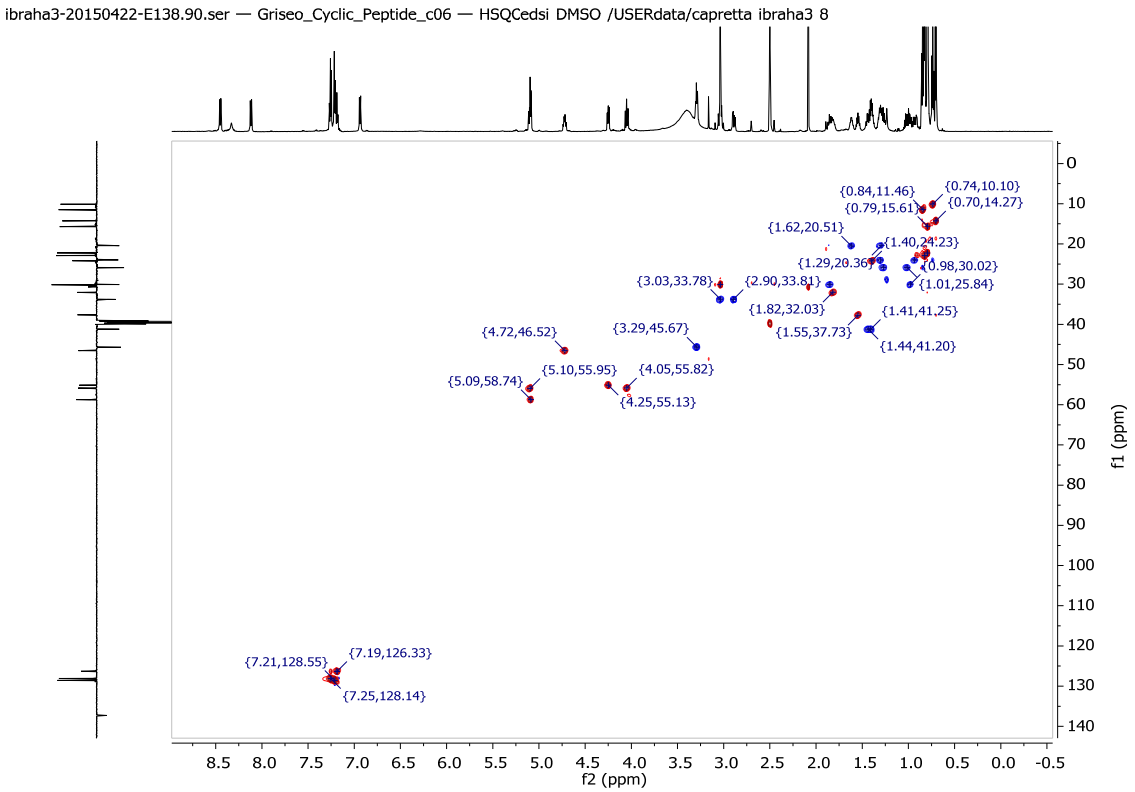


Figure 25S. HSQC spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$), multiplicity edited.

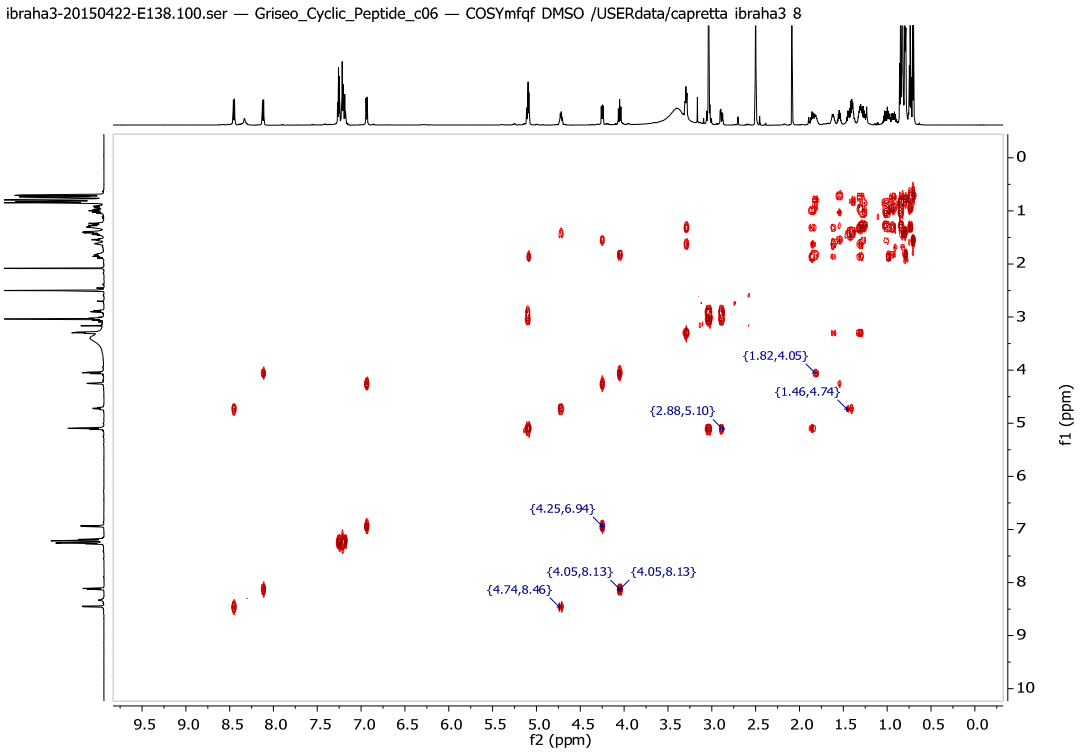


Figure 26S. COSY spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

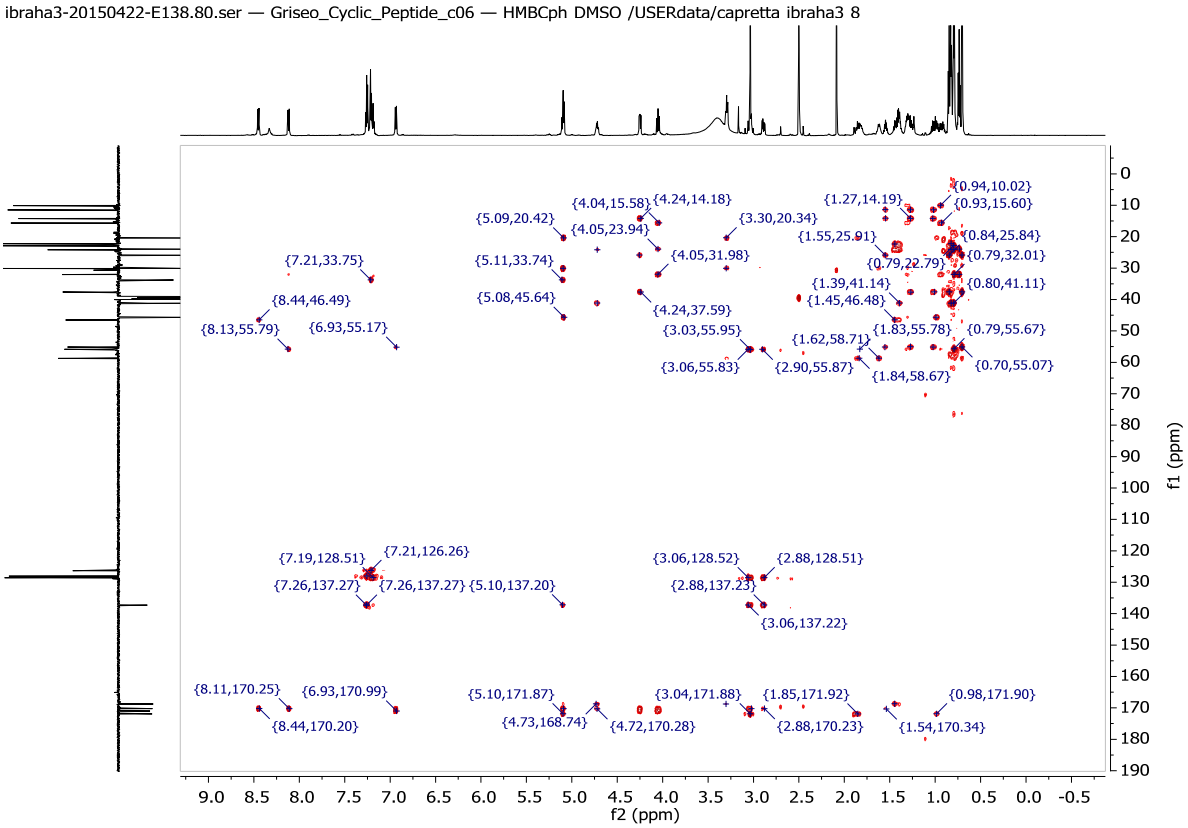


Figure 27S. HMBC spectrum of cyclic penta-peptide 5 (700 MHz, DMSO $_{d6}$).

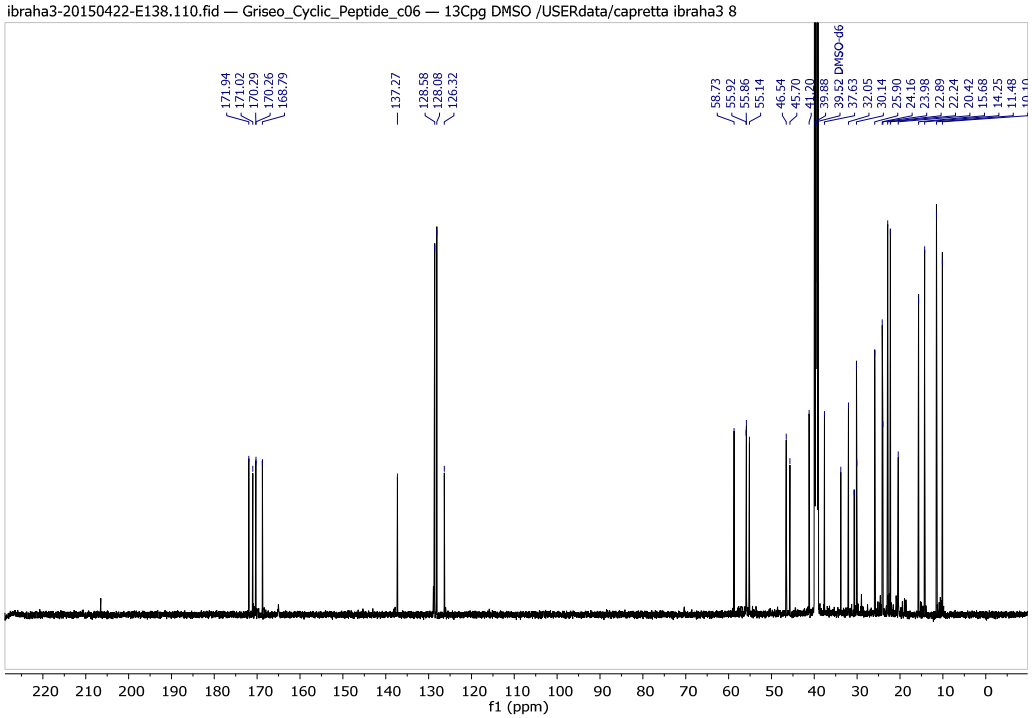


Figure 28S. ^{13}C spectrum of cyclic pentapeptide 5 (176 MHz, DMSO $_{d6}$).

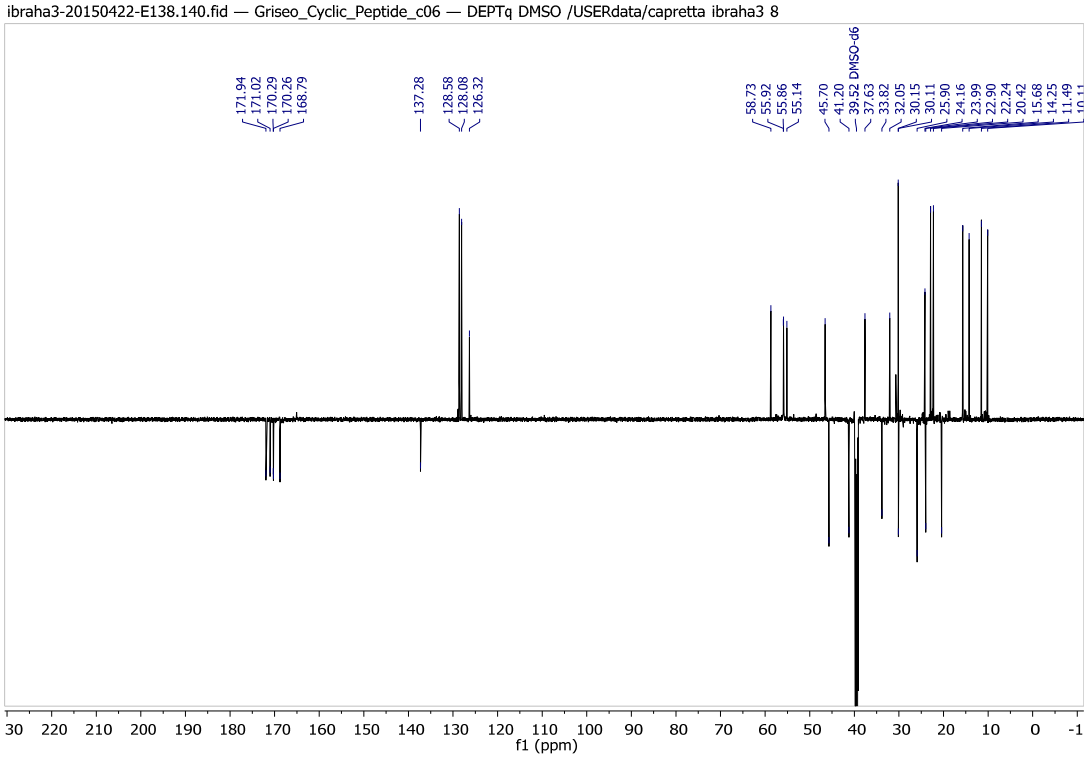


Figure 29S. ^{13}C -DEPTq spectrum of cyclic pentapeptide 5 (176 MHz, DMSO $_{d6}$).

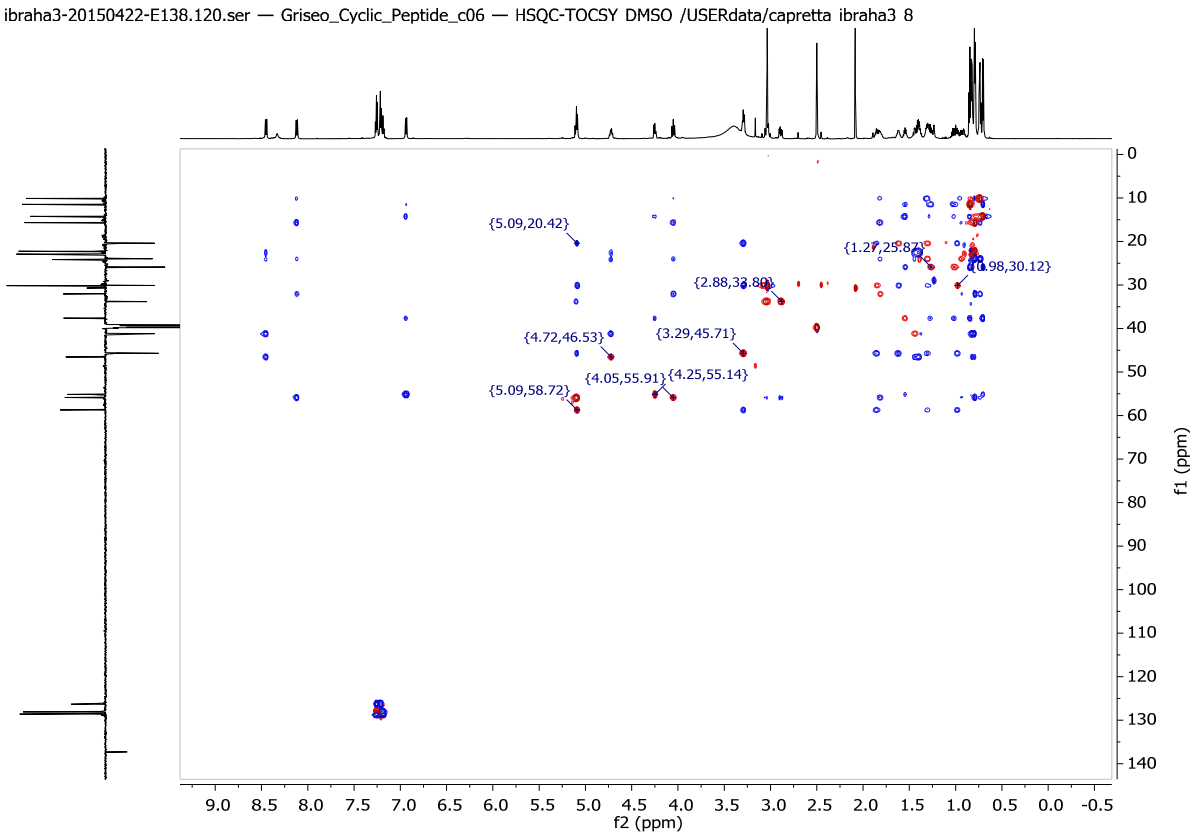


Figure 30S. HSQC-TOCSY spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

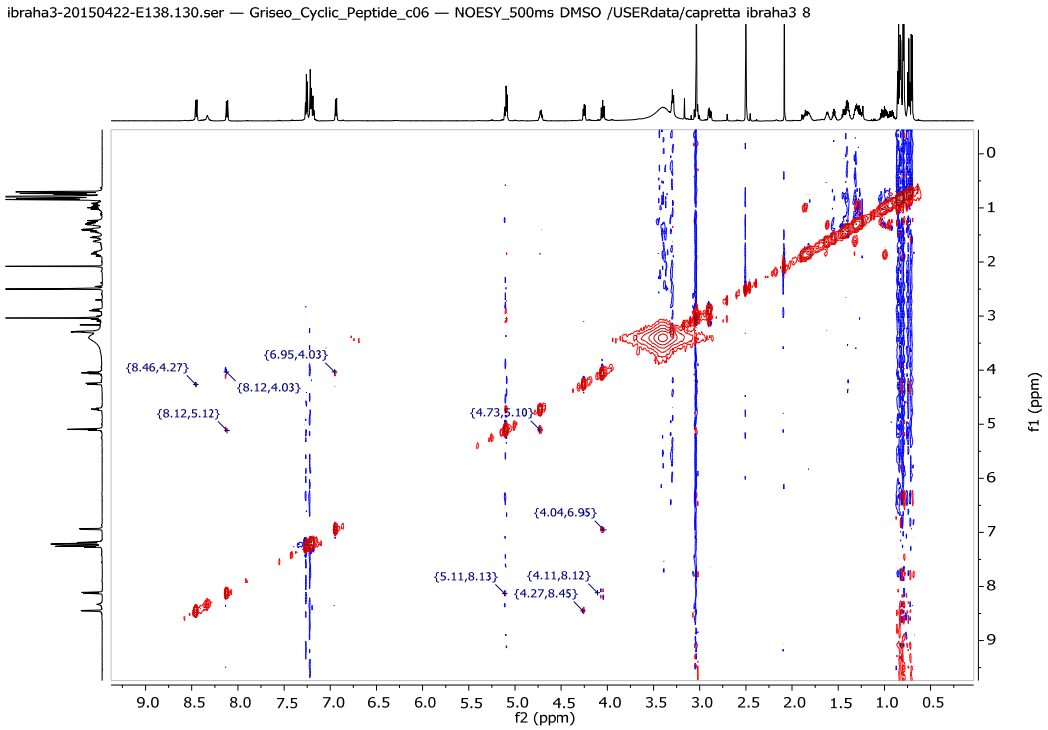


Figure 31S. NOESY spectrum of cyclic pentapeptide 5 (700 MHz, DMSO $_{d6}$).

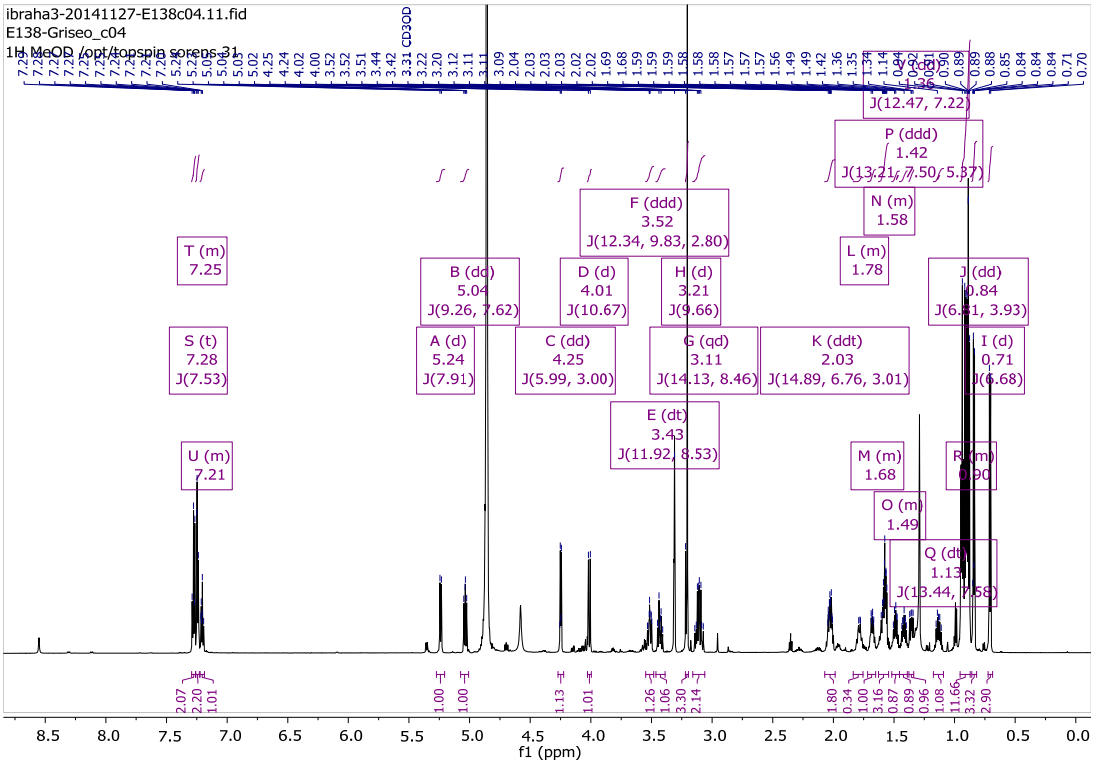


Figure 32S. ^{1}H spectrum of cyclic pentapeptide 1 (700 MHz, CD_3OD).

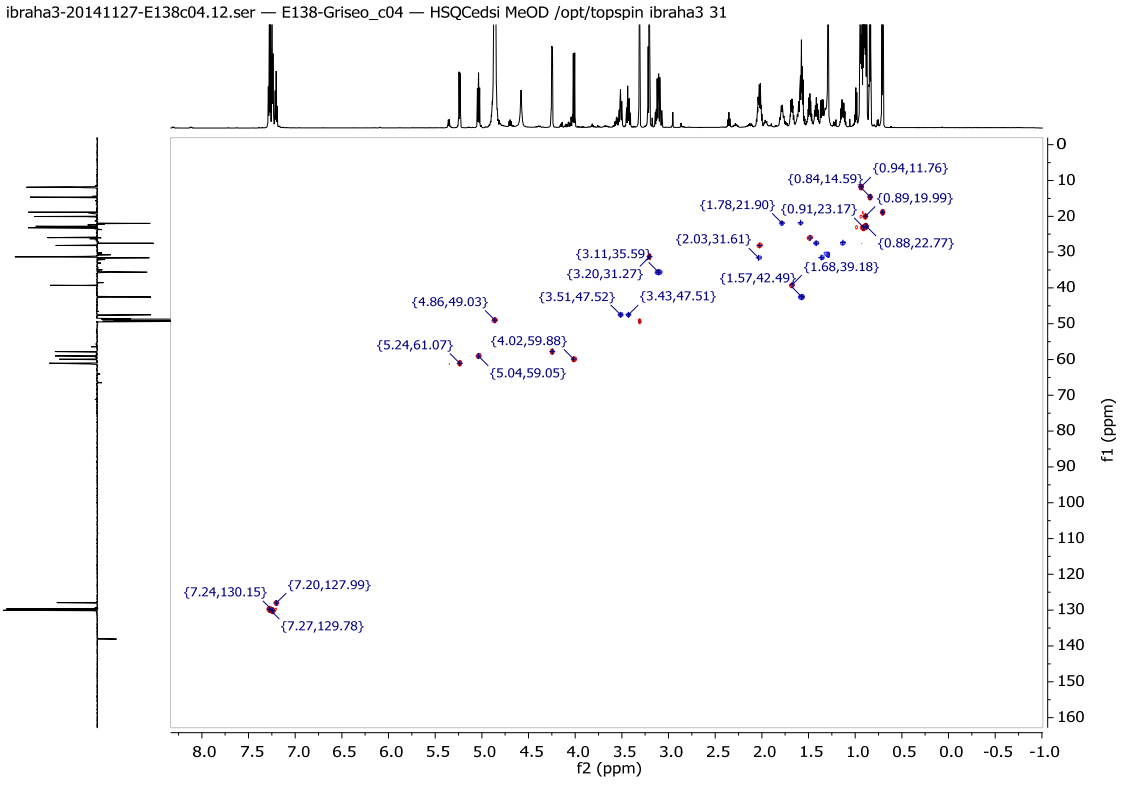


Figure 33S. HSQC spectrum of cyclic pentapeptide 1 (700 MHz, CD_3OD).

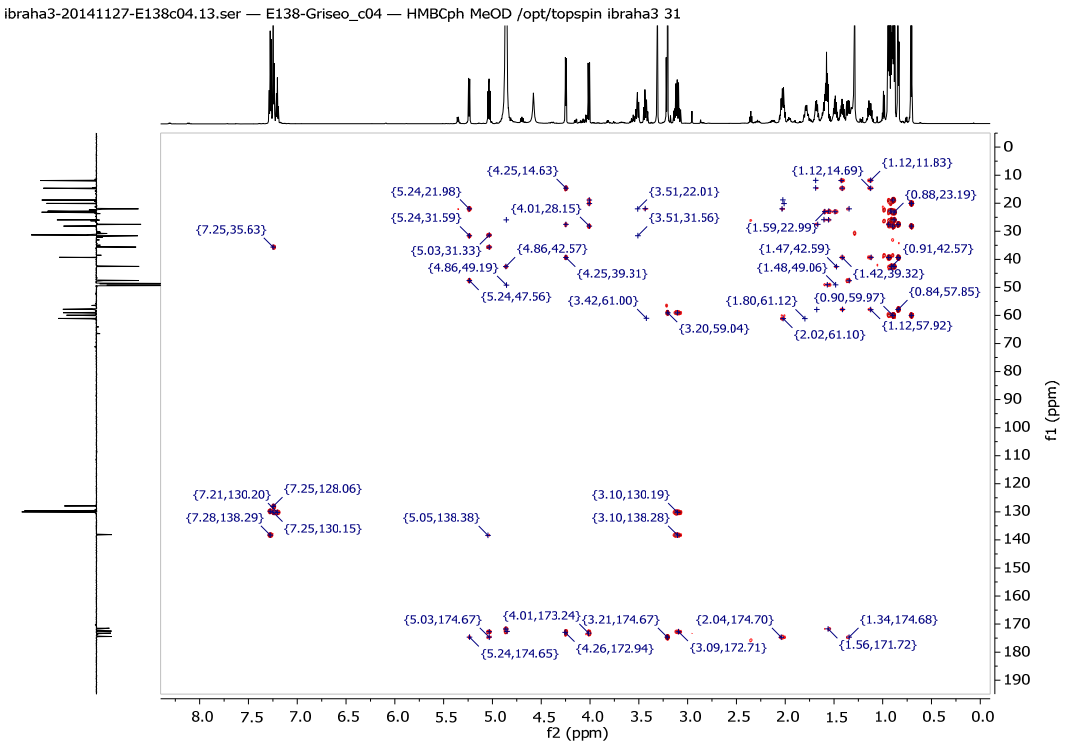


Figure 34S. HMBC spectrum of cyclic pentapeptide 1 (700 MHz, CD_3OD).

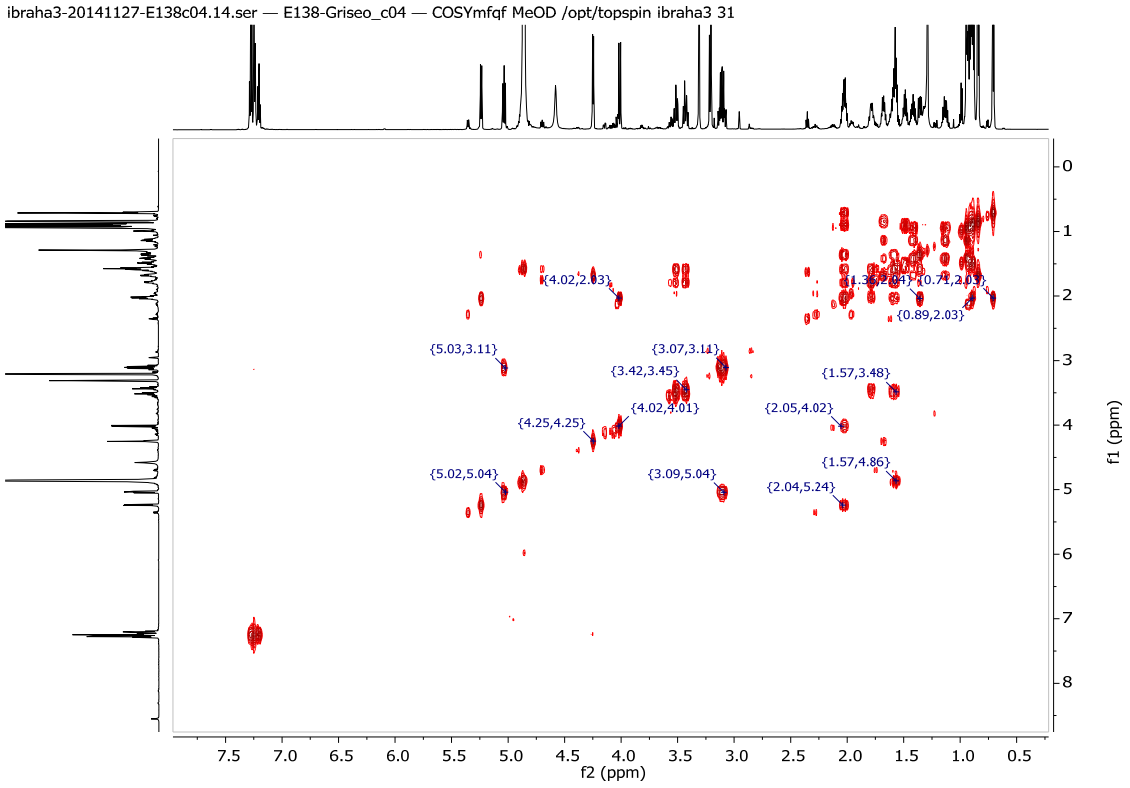


Figure 35S. COSY spectrum of cyclic pentapeptide 1 (700 MHz, CD_3OD).

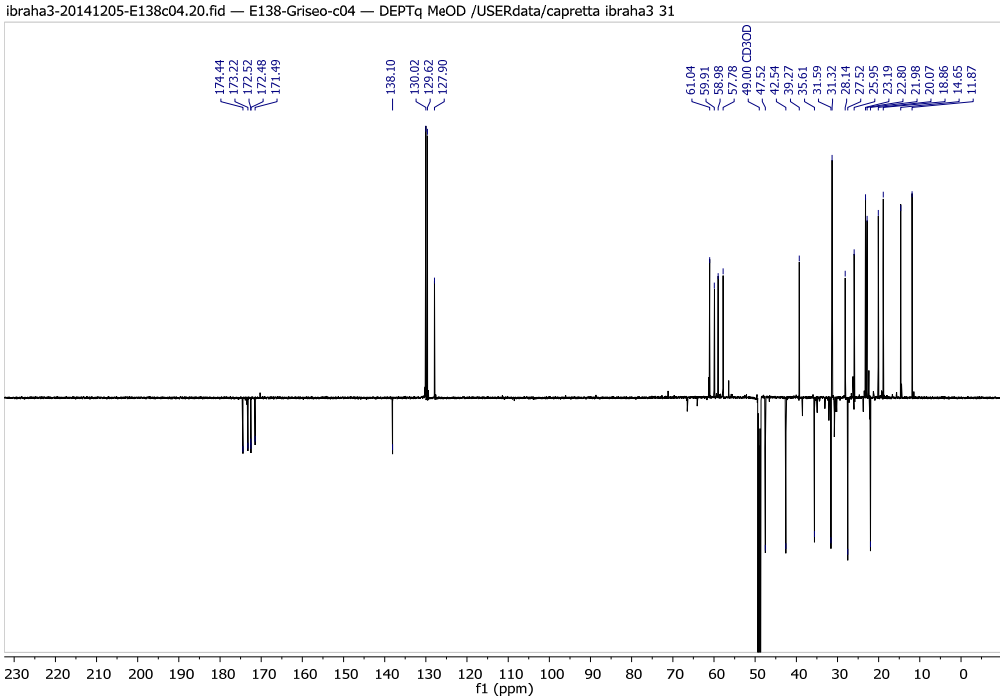


Figure 36S. ^{13}C -DEPTq spectrum of cyclic pentapeptide 1 (176 MHz, CD_3OD).

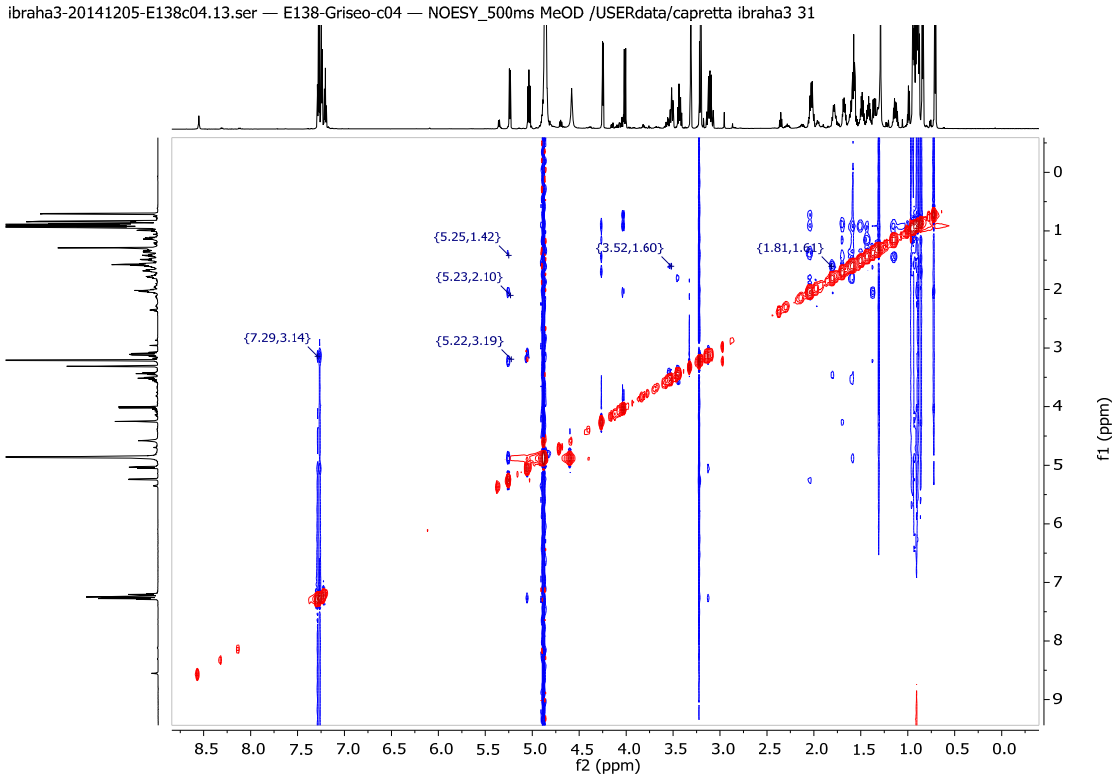


Figure 37S. NOESY spectrum of cyclic pentapeptide 1 (700 MHz, $DMSO_{d6}$).

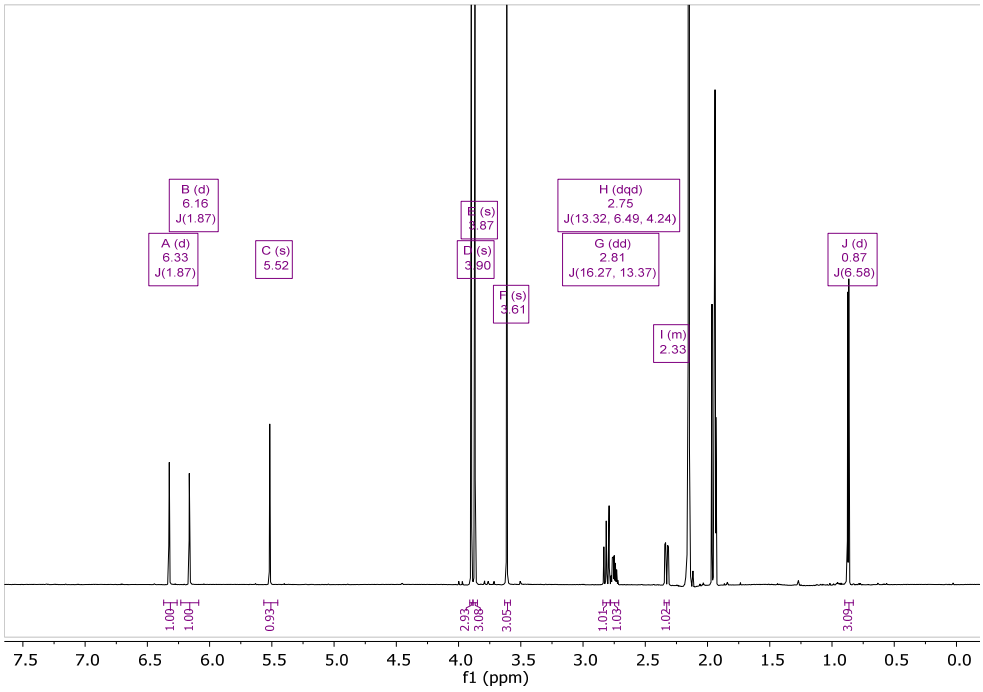


Figure 38S. ^{1}H spectrum of dechlorogriseofulvin (700 MHz, CD_3CN).

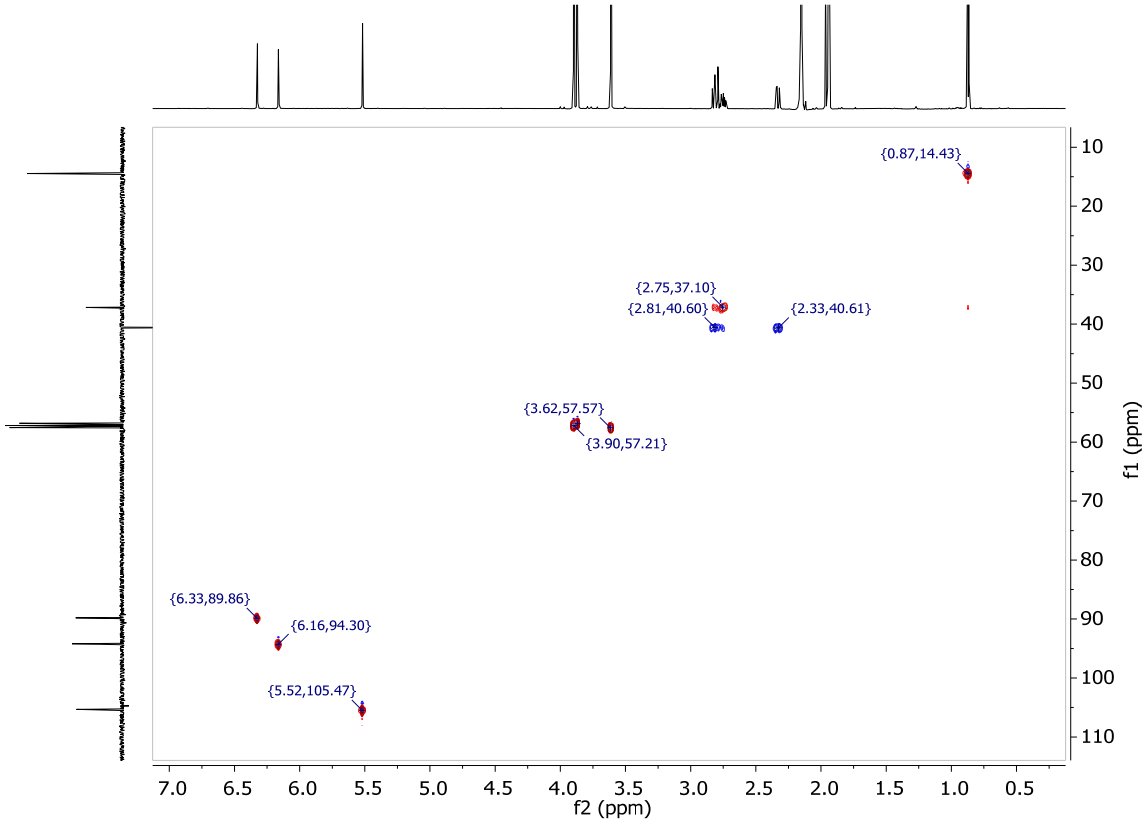


Figure 39S. HSQC spectrum of dechlorogriseofulvin (700 MHz, CD_3CN).

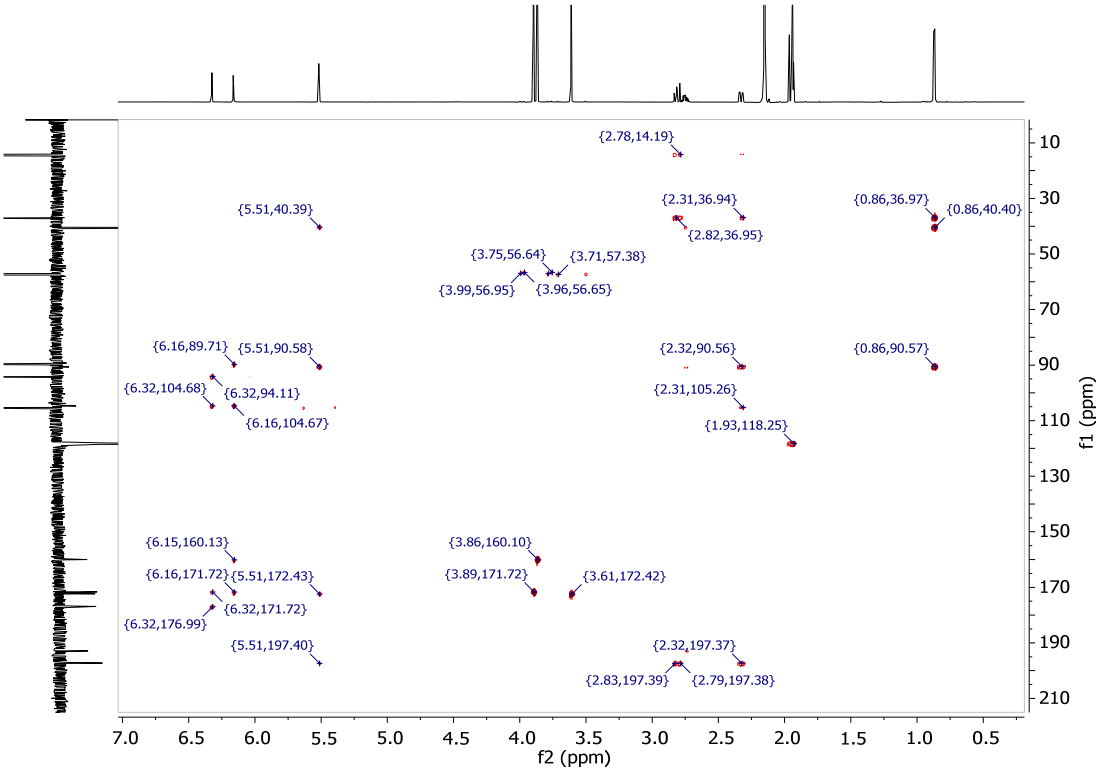


Figure 40S. HMBC spectrum of dechlorogriseofulvin (700 MHz, CD_3CN).

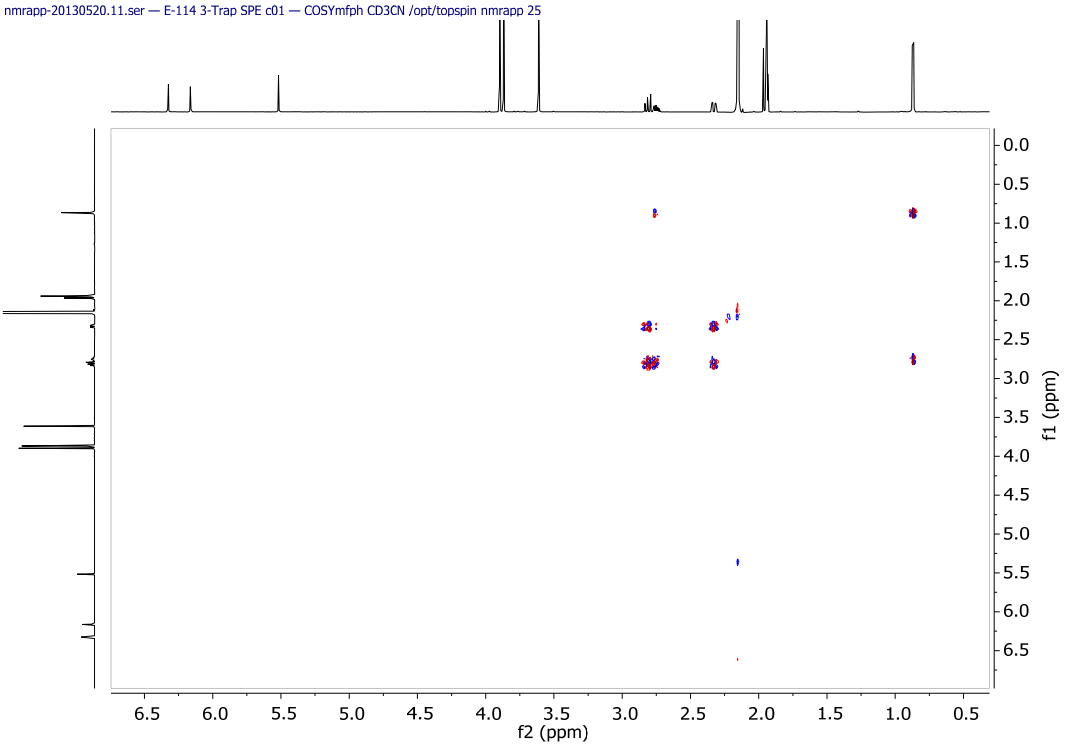


Figure 41S. COSY spectrum of dechlorogriseofulvin (700 MHz, CD_3CN).

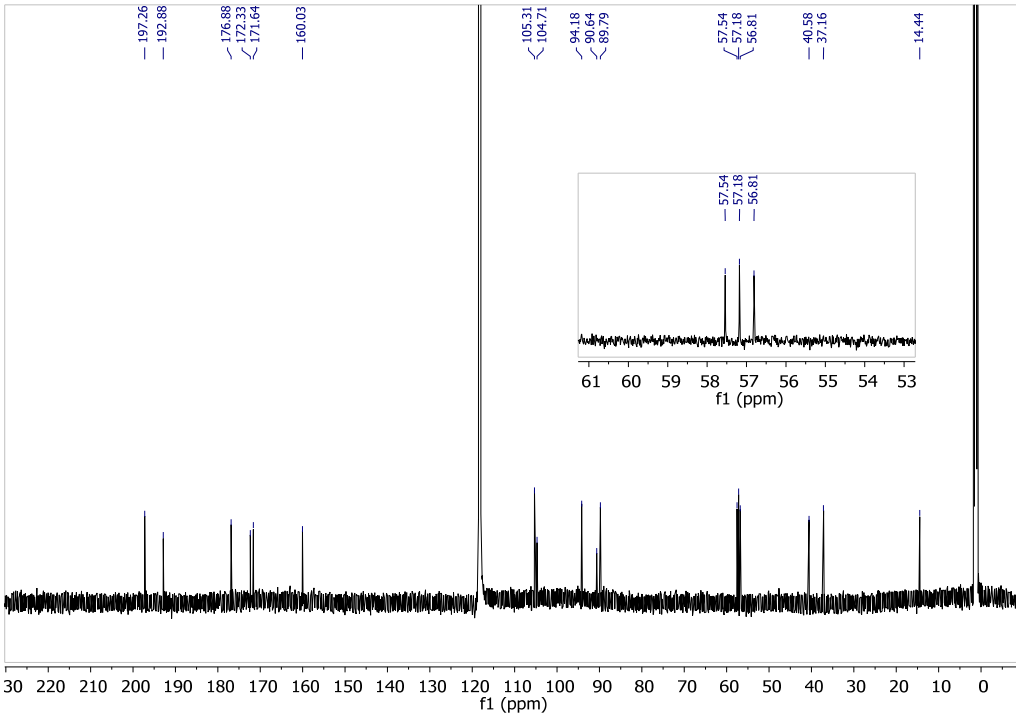


Figure 42S. ^{13}C spectrum of dechlorogriseofulvin (176 MHz,CD_3CN).

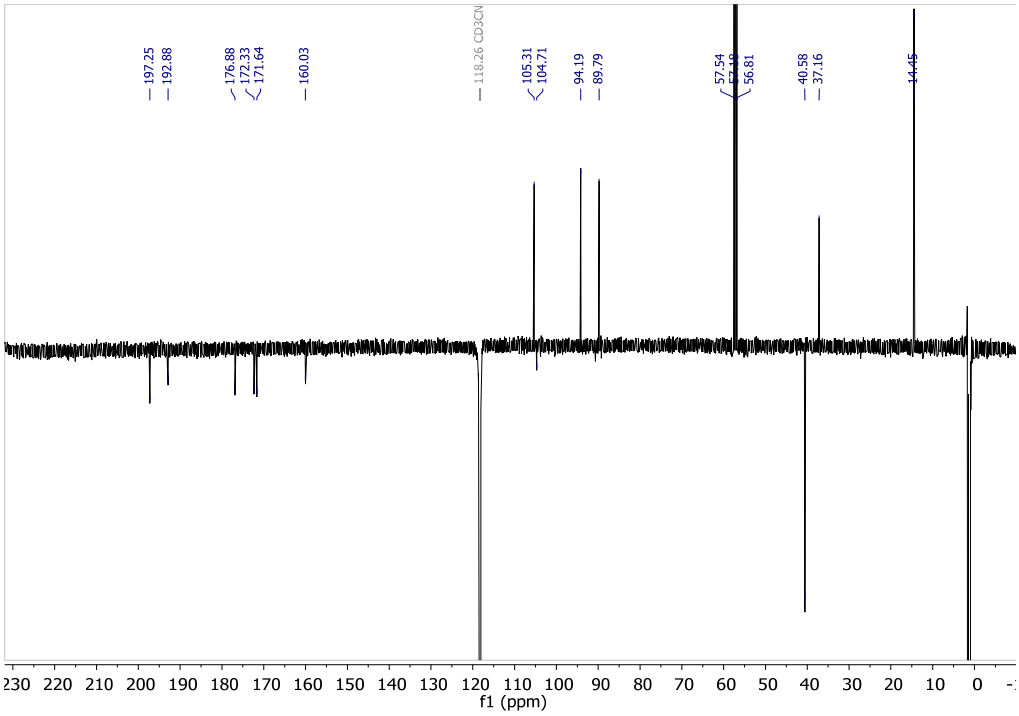


Figure 43S. ^{13}C -DEPTq spectrum o of dechlorogriseofulvin (176 MHz, CD_3CN).

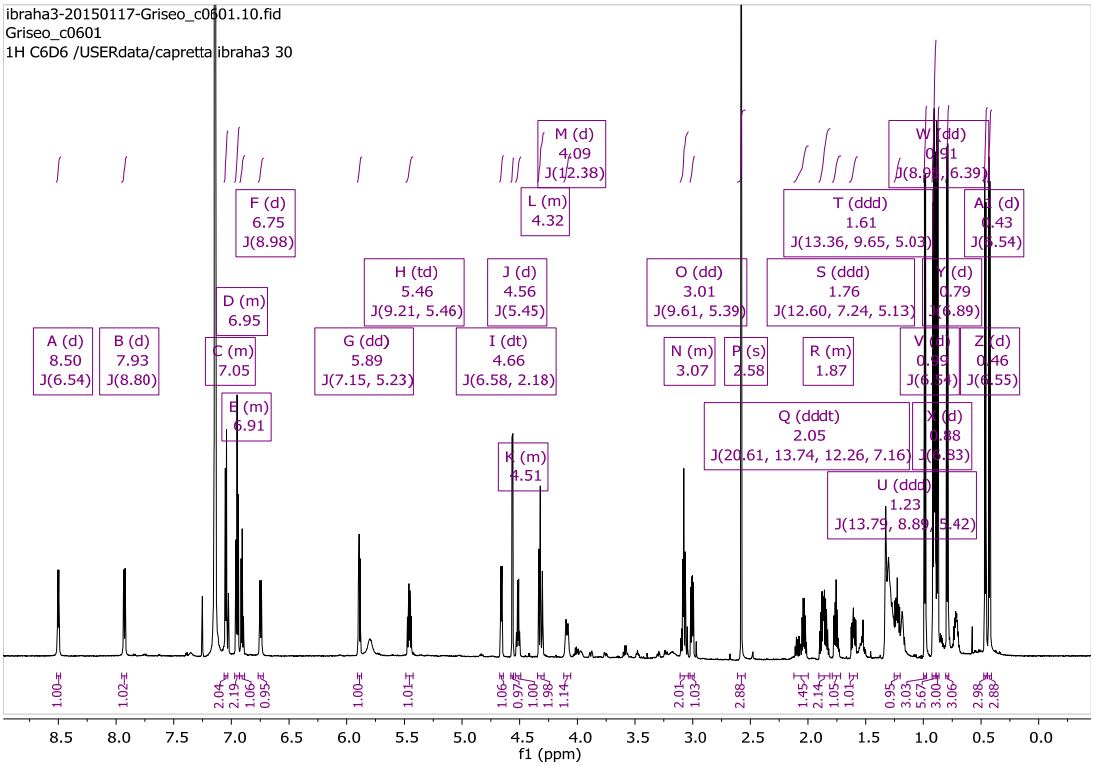


Figure 44S. [1]H spectrum of Hirsutatin A (700 MHz, C6D6).

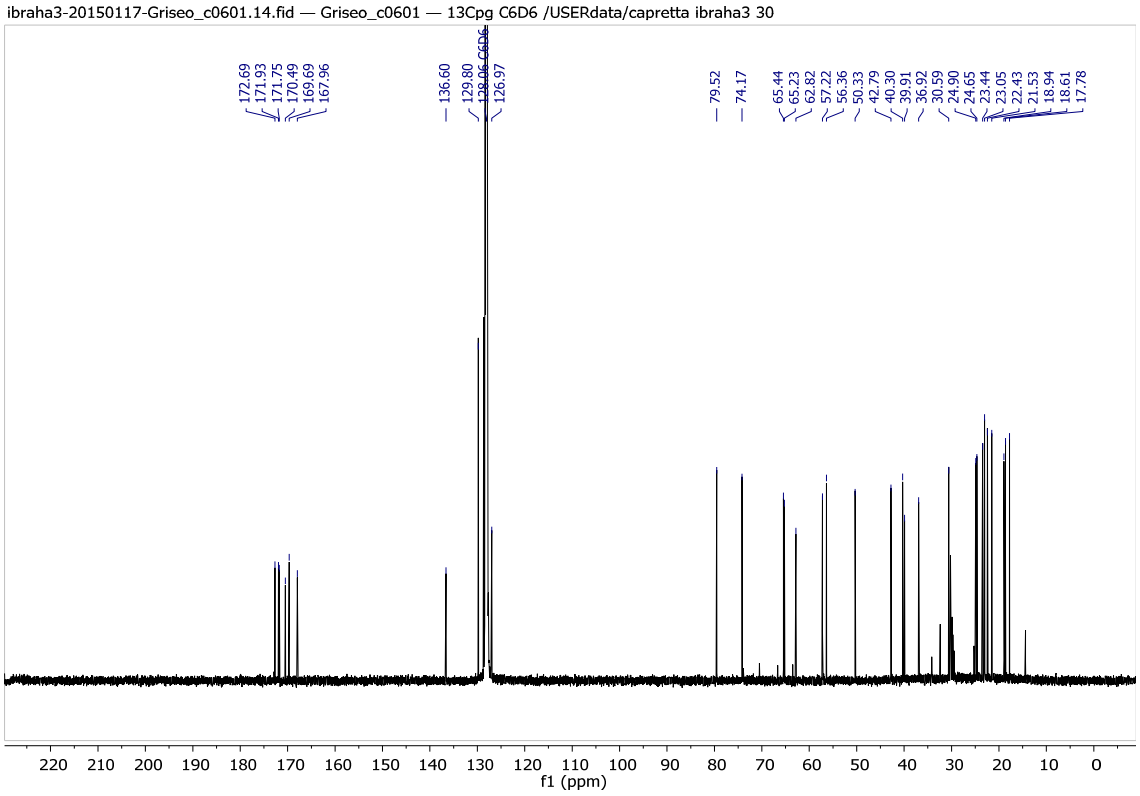


Figure 45S. [13]C spectrum of Hirsutatin A (176 MHz, C6D6).

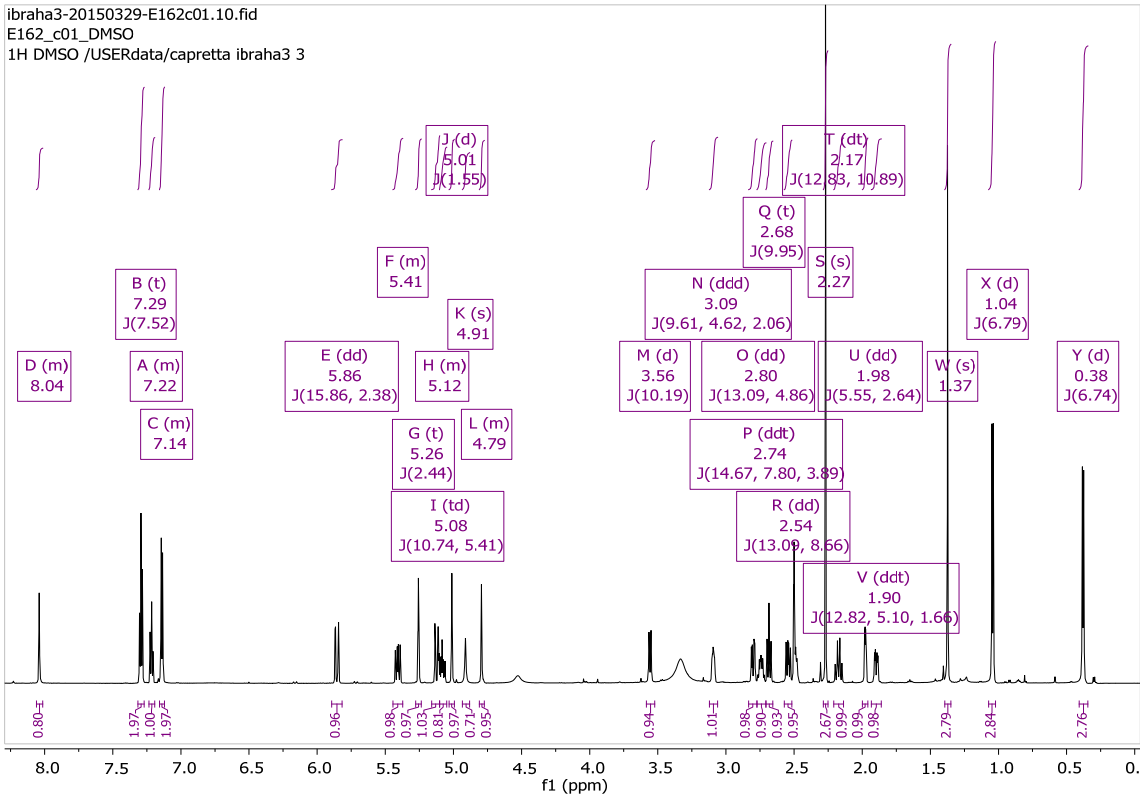


Figure 46S. ^{1}H spectrum of Cytochalasin D (700 MHz, $DMSO_{d6}$).

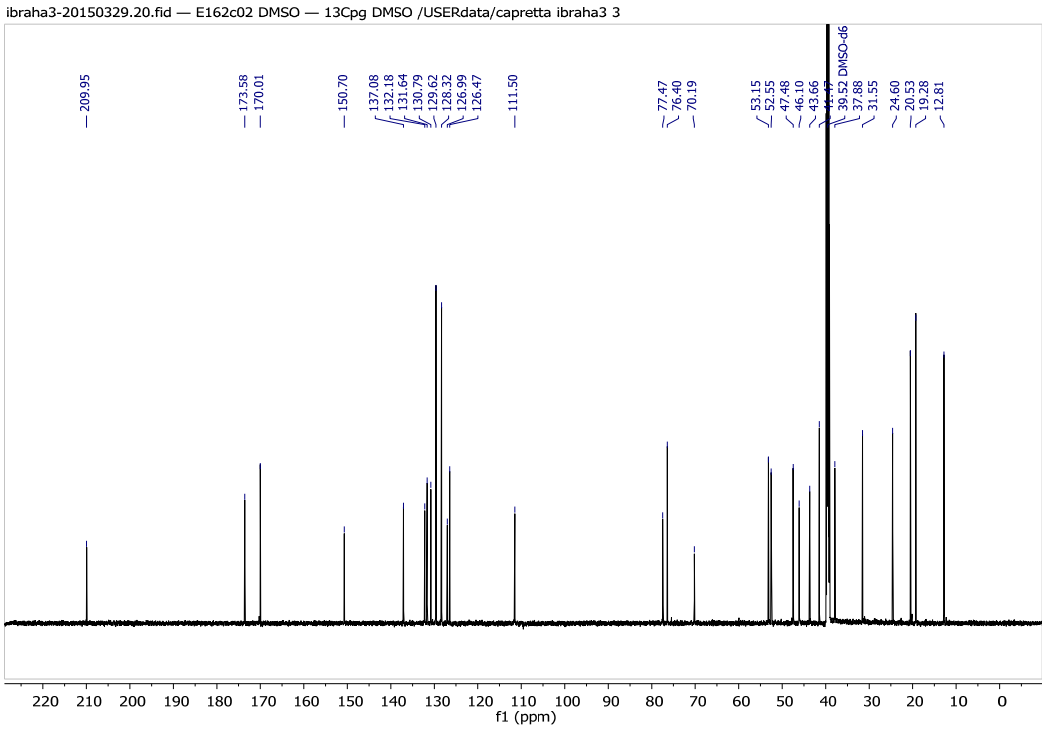


Figure 47S. ^{13}C spectrum of Cytochalasin D (176 MHz, $DMSO_{d6}$).

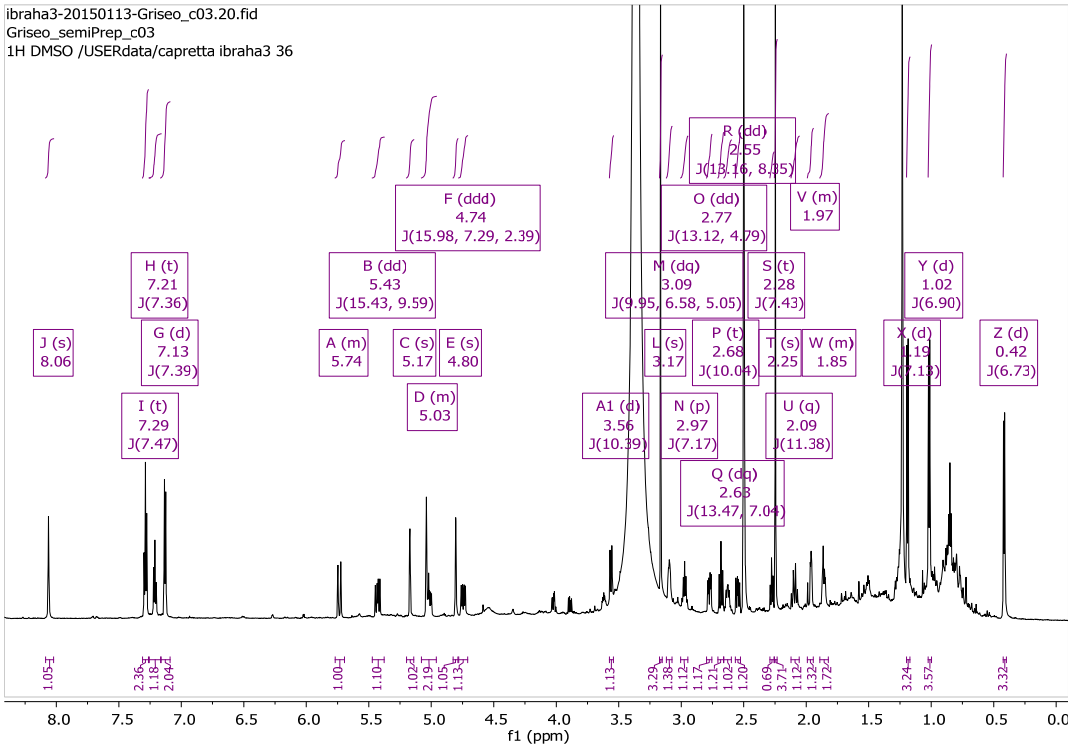


Figure 48S. ^{1}H spectrum of Zygosporin E (700 MHz, $DMSO_{d6}$).

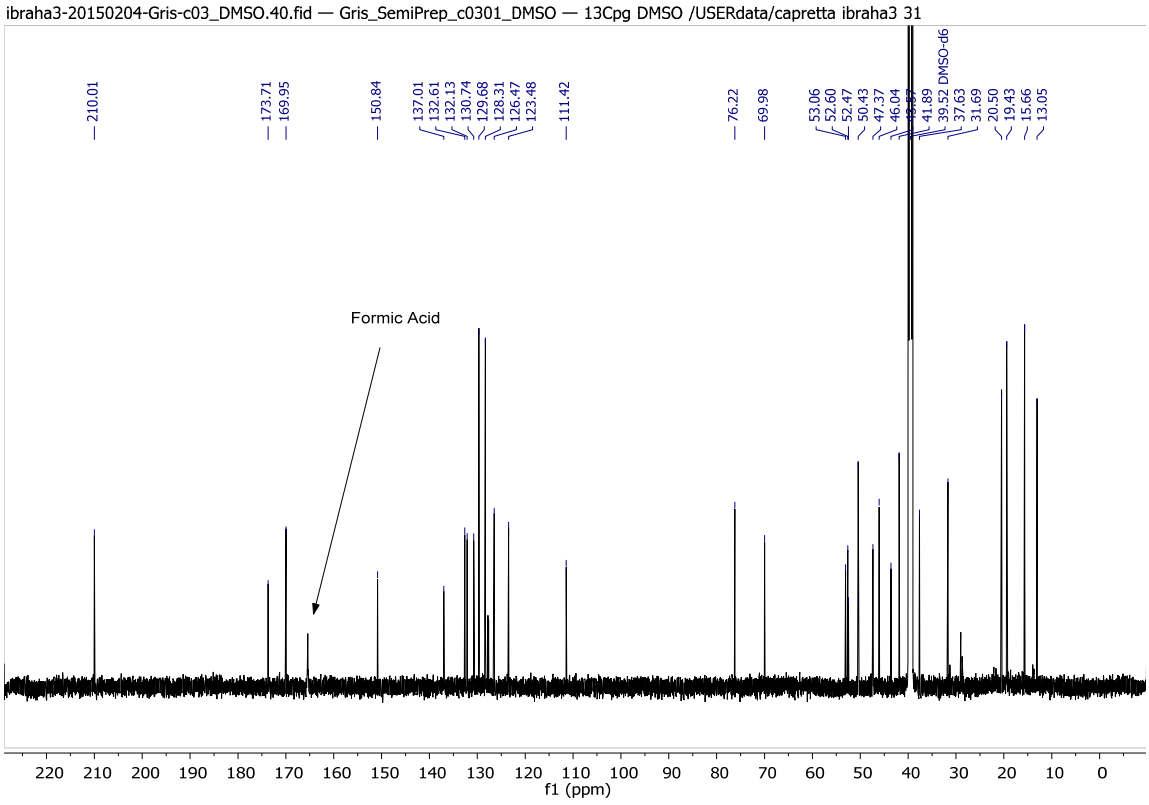


Figure 49S. ^{13}C spectrum of Zygosporin E (176 MHz, $DMSO_{d6}$).