Expediting Natural Product Discovery with Mass-Spectrometry based Metabolomic Analyses

By

Brett Covington

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

August 31, 2018

Nashville, Tennessee

Approved:

Brian O. Bachmann, Ph.D.

John A. McLean, Ph.D.

David W. Wright, Ph.D.

Eric P. Skaar, Ph.D.

To my Darlin’

Acknowledgment

I have never considered myself particularly smart or clever, but over the years I have been fortunate to receive tremendous support from family, friends, and mentors. My grandfather, Donald Pfeffer, has always been a source of inspiration. He challenged me to focus on solutions rather than problems. I think it was this positive mindset that allowed him to build his own small airplane, though it must have been some form of insanity that prompted him to actually fly it around. I was first drawn into science by my high school chemistry teacher, Cathy Petty. She helped inspire me to pursue chemistry in college, and there I was convinced to pursue a doctoral degree in chemistry by a good mentor and friend, Meagan Mann. Without her encouragement, I most certainly would not be writing this document.

Once at Vanderbilt I was initially drawn to Dr. Brian Bachmann's lab for the chance to go on exciting cave adventures, and I must say I will never forget the cave trips that I have experienced with the group. From rappelling down into the dark pit of Cagle's chasm to hiking through the crystalline cathedral rooms at Blue Springs cave, these trips were always a blast. Though even before my first cave trip, I became convinced that I wanted to continue working in the group from the wonderful environment and exciting research. Many times I have thought that natural product discovery, if not a career, could even be a fun hobby. The thrill of chasing down potentially novel complex molecules can be almost addicting. I received lots of support from Dr. Bachmann, fellow lab-mates (notably Dagmara (Kasia) Derewacz, and Dr. C. Ruth McNees), and collaborators like Dr. Cody Goodwin. I have learned so much through the years thanks to their freely given help.

I would like to thank my committee members, Dr.Bachmann, Dr. John McLean, Dr. David Wright, and Dr. Eric Skaar for working with me through the painful process of scheduling all the committee meetings throughout my graduate career. Looking back on how bad my early presentations were, I am

even more thankful for the time and effort my committee surrendered to further my development. I sincerely appreciate all your advice and support.

Lastly, I would like to thank all of my family for their love and support through the years. My parents John and Joni Covington, my older brothers Drs. Cody and Kyle Covington, and my younger sisters Julie and Brittney Covington have always been quick to offer advice and support. I couldn't image going through this without them.

Table of contents

Acknowledgment iii

List of Tables viii

List of Figures x

List of Abbreviations xv

CHAPTER
I Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites 1

Introduction 1
Natural products for antibiotic discovery. 1
Strategies to access novel natural products 3

Methods of Generating Inventories of Microbial Metabolites 5
Mass measurement accuracy 6
Isotopic modeling 7
Chromatographic retention time 8
Size and shape by ion mobility 9
Ion fragmentation for structural information 10
Leveraging spatiotemporal metabolomics inventories to capture inter-organism interactions 11

Preparation of High Content Mass Spectral Data for Metabolomics Studies 12
Strategies for formatting data for effective comparative analysis 12
Methods and considerations for metabolite peak detection and alignment 13

Analysis of metabolomics data in the context of secondary metabolites 16
Multivariate statistical analysis for identification of abundant covarying metabolites 17
Discovering molecular inventories of microbial responses via self-organizing map analytics 21
Molecular networking to reveal structurally related ions 24
Investigations of secondary metabolite bioactivity 28

Conclusions 30

Acknowledgements 31

Dissertation Statement and Chapter Contributions 31
Dissertation Statement 31
My contributions to works presented in each chapter 32

References 33

II Prioritization of secondary metabolites in Streptomyces coelicolor A3(2) with self-organizing maps 61

Introduction 61

Results 63
Multiplexing stimuli of secondary metabolism 63
Identifying products of multiplexed stimulation through MVSA 64
Identifying products of multiplexed stimulation by self-organizing maps 67

Measuring and structuring metabolic perturbations 70
Response profile analysis 71

Discussion 74

Materials and Methods 75
Eliciting antibiotic resistance and fermentations 76
Rare earth element fermentations 76
Extraction of liquid fermentations 76
Co-culture generation 77
UPLC-IM-MS Data Acquisition and Processing 78
Mass spectrometry data processing workflow 79

Acknowledgments 81

References 81

III Discovery of the ciromicins from *Nocardiopsis FU40* 85

Introduction 85

Results 86
Self-organizing maps to reveal consequences of mixed culture 86
Mixed culture induces large metabolomic changes 88
Response features identified by comparative metabolomics. 89
A polyene prioritized for structure elucidation 90
Ciromicin biosynthesis 94

Discussion 97

Materials and Methods 99
Fermentation conditions 99
Culture extractions 100
Ciromicin purification 100
Photochemical conversion of ciromicin 100
UPLC/IM-MS Data Acquisition 101
Metabolomic analysis 102

Acknowledgments 102

References 103

IV Prioritization of natural products from hypogean actinomycetes 107

Introduction 107

Results 109
Phylogenetically diverse actinomycetes isolated from hypogean environments. 109
Stimuli increase natural product biosynthesis across actinomycete genera. 111
Comparative metabolomics prioritizes secondary metabolites from stimuli exposure. 113
Novel polyketide produced from interactions between rare Streptosporangium and microbial co-culture 117
Investigating funisamine biosynthesis 120

Discussion 122

Materials and methods 125
Reagents and strains 125

Cave strain isolation and identification 125
Microbial stimulation and culture conditions.....126
Extraction of liquid fermentations.....127
LCMS Data Acquisition and Processing.....127
Funisamine Isolation 128
High Resolution Mass Spectrometry.....129
Genome Sequencing of Streptosporangium caverna 129

Acknowledgements.....130

References.....130

V Dissertation summary and future directions.....137

Dissertation Summary 137

Future directions 139

References.....141

APPENDIX
A Supporting data for Chapter 3 145

B Supporting data for Chapter 4 206

C Curriculum vitae.....264

List of Tables

Table 1-1: Overview of analytical descriptors for metabolomics-based NP discovery..... 6

Table 1-2: Overview of methods for metabolomic data analysis 16

Table 2-1: Maximum resultant metabolite production abundances compared to matched control cultures 71

Table 2-2: Gene clusters correlated to SM production 72

Table A-1 NMR correlation table of ciromicin A (1) in DMSO-d_6 145

Table A-2 NMR correlation table of ciromicin A (1) in CD_3OD. 146

Table A-3 NMR correlation table of ciromicin B (2) in CD_3OD. 147

Table A-4 Features within ROI 1 from MEDI Analysis 160

Table A-5 Features within ROI 2 from MEDI Analysis 161

Table A-6: Features within ROI 3 from MEDI Analysis 162

Table A-7: Features within ROI 3 from MEDI Analysis 163

Table A-8: Features within ROI 3 from MEDI Analysis 164

Table A-9: Features within ROI 4 from MEDI Analysis 165

Table A-10: Features within ROI 5 from MEDI Analysis 166

Table A-11: Features within ROI 6 from MEDI Analysis 166

Table A-12: Features within ROI 6 from MEDI Analysis 167

Table A-13: Tentatively identified features from MEDI regions of interest 168

Table B-1: Table of identified actinomycete genera isolated from hypogean samples 206

Table B-2: Table of identified NP responses across stimuli conditions 211

Table B-3:: Table of putative NP responses across stimuli conditions 212

Table B-4: NMR data for funisamine..... 236

Table B-5: Table of ions detected through fragmentation analysis of purified funisamine 250

Table B-6: Table of putative BGCs identified by antiSMASH in the genome of *Streptosporangium* KDCAGE35 254

Table B-7: Genetic organization of putative funisamine BGC in *Streptosporangium* KDCAGE35. 255

List of Figures

Figure 1-1: General metabolomics workflow. 13

Figure 1-2: Using a PCA scores plot to prioritize microbial producers. 18

Figure 1-3: MVSA S-, Loadings, and Volcano plots to identify induced features. 20

Figure 1-4: Feature organization within a SOM. 22

Figure 1-5: Three example comparisons of prioritized features through PCA and SOM analyses 23

Figure 1-6: Applications of molecular networking to explore data. 25

Figure 2-1: Solid agar mixed culture apparatus 64

Figure 2-2: (A) Loadings plot for Figure 2-3A. 65

Figure 2-3: PCA of metabolomic inventories. 66

Figure 2-4: General self-organizing map (SOM)-based approach to feature prioritization 68

Figure 2-5: Differential metabolic phenotype heat maps representing increased production/decreased consumption of molecules using a single growth medium. 69

Figure 3-1 Metabolomic analysis of mono- and co-cultures. 87

Figure 3-2 Distribution of metabolomics features between cocultures. 89

Figure 3-3 Comparison of PCA and MEDI analyses for feature prioritization 90

Figure 3-4 Structures of ciromicins A (1) and B (2). 91

Figure 3-5 Genetic organization of ciromicin gene cluster in NF 95

Figure 4-1 Phylogenetic tree of cave organisms. 110

Figure 4-2 Prioritization of natural products from comparative metabolomics. 114

Figure 4-3 Activation of funisamine biosynthesis through mixed culture. 118

Figure 4-4 Proposed biosynthesis of funisamine. 121

Figure A-1 ^{1}H proton NMR spectrum of ciromicin A (1) in DMSO-d6 148

Figure A-2 ^{13}C carbon NMR spectrum of ciromicin A (1) in DMSO-d6. 148

Figure A-3 2D COSY NMR spectrum of ciromicin A (1) in DMSO-d6. 149

Figure A-4 2D HSQC NMR spectrum of ciromicin A (1) in DMSO-d6. 149

Figure A-5 2D HMBC NMR spectrum of ciromicin A (1) in DMSO-d6. 150

Figure A-6 2D TOCSY NMR spectrum of ciromicin A (1) in DMSO-d6. 150

Figure A-7 2D NOESY NMR spectrum of ciromicin A (1) in DMSO-d6. 151

Figure A-8 ^{1}H proton NMR spectrum of ciromicin A (1) in CD_3OD. 151

Figure A-9 ^{13}C carbon NMR spectrum of ciromicin A (1) in CD_3OD. 152

Figure A-10 2D COSY NMR spectrum of ciromicin A (1) in CD_3OD. 152

Figure A-11 2D HSQC NMR spectrum of ciromicin A (1) in CD_3OD. 153

Figure A-12 2D HMBC NMR spectrum of ciromicin A (1) in CD_3OD. 153

Figure A-13 2D TOCSY NMR spectrum of ciromicin A (1) in CD_3OD. 154

Figure A-14 2D NOESY NMR spectrum of ciromicin A (1) in CD_3OD. 154

Figure A-15 ^{1}H proton NMR spectrum of ciromicin B (2) in CD_3OD. 155

Figure A-16 ^{13}C carbon NMR spectrum of ciromicin B (2) in CD_3OD. 155

Figure A-17 2D COSY NMR spectrum of ciromicin B (2) in CD_3OD. 156

Figure A-18 2D HSQC NMR spectrum of ciromicin B (2) in CD_3OD. 156

Figure A-19 2D HMBC NMR spectrum of ciromicin B (2) in CD_3OD. 157

Figure A-20 2D TOCSY NMR spectrum of ciromicin B (2) in CD_3OD. 157

Figure A-21 2D NOESY NMR spectrum of ciromicin B (2) in CD_3OD. 158

Figure A-22: Heatmaps and density map for *Nocardiopsis* mixed culture screen. 159

Figure A-23: MetaXCMS screenshot of imported files 170

Figure A-24: MetaXCMS filtering parameters. 171

Figure A-25: MetaXCMS alignment parameters 172

Figure A-26: MEDI advanced parameters 173

Figure B-1 Metabolomic response to environmental stimuli. 206

Figure B-2: Metabolomic response from antibiotic exposure. 208

Figure B-3: Metabolomic response from rare earth metal exposure. 209

Figure B-4: Metabolomic responses from mixed culture exposures. 210

Figure B-5: Distribution of isolated and putative NPs from Tables S2-3 215

Figure B-6: Analysis of volcano plot-prioritized features with chromatogram overlays. 216

Figure B-7: EICs of (A) okicenone and (C) aloesaponarin II 217

Figure B-8: EICs of (A) hypogeamicin A, (C) hypogeamicin B, and (E) hypogeamicin C 218

Figure B-9: EICs of (A) actinomycin C2 219

Figure B-10: EICs of (A) propeptin 1, (C) propeptin 2, and (D) tetarimycin 220

Figure B-11: EICs of (A) funisamine 221

Figure B-12: Mass spectral data for *Micromonospora* BBHARD22 features 222

Figure B-13: UV spectrum of *Micromonospora* BBHARD22 features 223

Figure B-14: UV spectrum of *Nonomuraea* BBHARD23 features 224

Figure B-15: UV spectrum of features from *Streptomyces* BCCAGE06 225

Figure B-16: UV spectrum of features from *Streptomyces* BCCAGE31 226

Figure B-17: UV spectrum of features from *Nonomuraea* BCCAGE42 227

Figure B-18: UV spectrum of features from *Streptomyces* BCCAGE18 228

Figure B-19: UV spectrum of features from *Saccharothrix* BBHARD27 229

Figure B-20: UV spectrum of features from *Kribella* BBSNAI08 230

Figure B-21: UV spectrum of features from *Microbispora* BCCAGE54 231

Figure B-22: UV spectrum of funisamine from *Streptosporangium* KDCAGE35 232

Figure B-23: H-NMR spectrum of aloesaponarin II 233

Figure B-24: Expanded H-NMR spectrum of aloesaponarin II 233

Figure B-25: H-NMR spectrum partially purified okicenone 234

Figure B-26: H-NMR spectrum of tetarimycin B 234

Figure B-27: H-NMR spectrum of hypogeamicin B 235

Figure B-28: H-NMR spectrum of actinomycin C 235

Figure B-29: H-NMR (600 MHz) spectrum of funisamine 238

Figure B-30: C-NMR (600 MHz) spectrum of funisamine 238

Figure B-31: COSY (600 MHz) spectrum of funisamine 239

Figure B-32: Expanded COSY (600 MHz) spectrum of funisamine 240

Figure B-33: Expanded COSY (600 MHz) spectrum of funisamine 241

Figure B-34: TOCSY (600 MHz) spectrum of funisamine 242

Figure B-35: TOCSY (600 MHz) spectrum of funisamine 243

Figure B-36: TOCSY (600 MHz) spectrum of funisamine 244

Figure B-37: HSQC (600 MHz) spectrum of funisamine 245

Figure B-38: HSQC (600 MHz) spectrum of funisamine 246

Figure B-39: HSQC (600 MHz) spectrum of funisamine 247

Figure B-40: HMBC (600 MHz) spectrum of funisamine 248

Figure B-41: NOESY (600 MHz) spectrum of funisamine 248

Figure B-42**:** High-Res mass spectrum for purified funisamine 249

Figure B-43: Fragmentation spectra of isolated funisamine 251

Figure B-44: Fragmentation assignments for funisamine 251

Figure B-45**:** (A) S-plot prioritization of amicoumacin B and funisamine from mixed culture 252

Figure B-46**:** Mixed culture competitor growth curves 253

Figure B-47: Multiple sequence alignment of ketoreductase domains within the funisamine BGC 259

Figure B-48: Predicted stereochemical configurations of products for each polyketide synthase module in funisamine biosynthesis. 260

Figure B-49: Expanded biosynthesis of funisamine from FunP1. 261

Figure B-50: Expanded biosynthesis of funisamine from FunP2 - FunP4. 262

Figure B-51: Expanded biosynthesis of funisamine from FunP5 - FunP8. 263

List of Abbreviations

Abbreviation	Meaning
NP	Natural Product
SM	Secondary metabolite
BGC	Biosynthetic Gene Cluster
UV	Ultra-violet
UV/Vis	Ultraviolet-visible spectroscopy
DNA	Deoxyribonucleic acid
rDNA	Ribosomal deoxyribonucleic acid

Mass spectrometry

UPLC-MS	Ultra-performance liquid chromatography mass spectrometry
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
HPLC-MS	High-performance liquid chromatography mass spectrometry
LC-MS	Liquid chromatography mass spectrometry
MS^E	Tandem mass spectrometry acquisition on Waters SYNAPT G2
m/z	Mass / charge ratio
MALDI	Matri-assisted laser ablation/desorption ionization
IM-MS	Ion mobility mass spectrometry
IMS	Imaging mass spectrometry
TOF	Time of flight
SILAC	Stable isotope labeling using amino acids in cell culture
SID	Surface induced dissociation
CID	Collision induced dissociation
RPLC	Reversed phase liquid chromatography
HILIC	Hydrophobic interaction chromatography
EIC	Extracted ion chromatogram

Nuclear magnetic resonance spectroscopy

COSY	Correlation spectroscopy
NOESY	Nuclear overhouser effect spectroscopy
NMR	Nuclear magnetic resonance
HMBC	Heteronuclear multi-bond correlation spectroscopy
HSQC	Hereronuclear single quantum correlation
TOCSY	Total correlation spectroscopy

Metabolomics methods

MEDI	Molecular expression dynamics inspector
SOM	Self-organizing map
MVSA	Multivariate statistical analyses
PCA	Principal component analysis
PLS	Partial least squares
ROI	Region of interest
GNPS	Global natural products social molecular networking
XCMS	Processing tool for various forms (X) of chromatography mass spectrometry
MAM	Multiplexed activity metabolomics

Organisms

BK	*Bacillus sp* KDCAGE13 (99 % 16S similarity with *Bacillus simplex*)
BS	*Bacillus subtilis* (ATCC23857)
EC	*Escherichia coli* (ATCC10536)
ML	*Micrococcus luteus*
NF	*Nocardiopsis sp FU40*
RW	*Rhodococcus sp.* BBSNAI13 (100 % 16S similarity with *R. wratislaviensis*)
SC	*Streptomyces coelicolor A3(2)*
TP	*Tsukamurella pulmonis* (ATCC700081)

Stimuli

La	Lanthanum
Sc	Scandium
Rif	Rifampicin
Str	Streptomycin
OSMAC	One strain many compounds

Secondary metabolite biosynthesis

ACP	Acyl carrier protein
AT	Acyltransferase domain
BGC	Biosynthetic gene cluster
DH	Dehydratase domain
ER	Enoylreductase domain
KR	Ketoreductase domain
KS	Ketosynthase domain
PKS	Polyketide synthase

Chapter 1 Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites

This chapter was adapted with permission from the Royal Society of Chemistry from a review written by Brett C. Covington, John A. McLean, and Brian O. Bachmann, first published in *Nat. Prod. Rep*. **2017,** 34, 6-24

Introduction

Natural products for antibiotic discovery.

Since their discovery and implementation, antibiotics have become a central component of modern medicine and changed the landscape of the pharmaceutical industry. Their importance is apparent by the 250 million antibiotic prescriptions that are given in the United States each year, enough for over 80 % of the population. However, over the past 30 years there has been a rapid increase in the incidence of antibiotic resistant infections, and at the same time it is becoming more and more difficult to discover new antimicrobial agents. To make matters worse, despite the increasing demand for new compounds with novel targets, many drug companies have scaled back or eliminated their antibiotic development programs.[18] The combination of these events is leading us towards what many experts are calling a post-antibiotic era in which antibiotics will no longer be effective at treating bacterial infection, and widespread resistance of this nature will significantly impact the death rate, period of infection, and the cost of health care. However, by unlocking hidden bioactive compounds from nature we may be able to expedite antibiotic discovery and effectively postpone or even prevent this crisis.

Fully 75 % of clinically relevant antibiotics are derived from natural products (NP)s, compounds produced by living organisms, and the majority of these are produced by microorganisms.[19] These

metabolites are seemingly not involved directly with normal growth, development, or reproductive processes within their producer and are commonly called secondary metabolites (SM)s. Microbially derived SMs, particularly those from actinobacteria, have historically been rich sources of antibiotics,[10, 20-21] and over 60 %, around 14000, of identified bioactive microbial compounds have shown antimicrobial activity . Actinobacteria have been known to produce a diverse array SMs clinically relevant antibiotics from many classes including β-lactams, macrolides, polyketides, aminoglycosides, and peptides SM production is often facilitated by large enzyme complexes which are encoded by readily identifiable biosynthetic gene clusters (BGC)s within the genomes of producing organisms. SM biosynthetic potential can be rapidly estimated from genomic sequence data via automated bioinformatics platforms capable of comparing sequenced BGCs to previously sequenced microorganisms and inferring putative structures of SMs by biosynthetic inference.[22-23] Recent reviews from Ziemert, Moammad and Weber,[29-34] and Medema and Fischbach have described several computational tools for BGC analysis. Though these actinobacteria and other natural sources have been extensively mined over the past several decades, genomic sequencing of both cultivated microorganisms and uncultivated microbiomes has revealed that most of the biosynthetic potential of microorganisms remains inaccessible to date. Even genomes of extensively studied microorganisms contain a large fraction BGCs for which NPs have not been identified. It is estimated that the products of greater than 90% of NP BGCs are either not sufficiently produced under standard laboratory growth conditions and/or their products are difficult to identify within extracted metabolomes. Recent studies indicate that in many cases these presumedly 'silent' BGCs are actually transcribed in levels that should be sufficient to observe produced SMs. In these cases, the main factor impeding NP discovery may be the ability to identify produced NPs from complex biological extracts. NPs play crucial roles in the chemical ecology of their producing organisms, and these roles often correlate translationally into applications in human medicine. Therefore, solving the linked problems of NP

biosynthesis and the identification of NPs within metabolomic inventories have become central efforts in the field of NP discovery.

Strategies to access novel natural products

Increasingly sophisticated methodologies have been developed to activate production of NPs from orphan BGCs with no known products. These methodologies can be subdivided into heterologous and native approaches. Heterologous strategies endeavour to recapitulate functional secondary metabolic BGCs in surrogate producers. Gene clusters may be cloned, and/or synthesized and refactored into alternate organisms with the aim of detecting newly produced metabolites in comparison to a clean background.[39] The success of heterologous expression is dependent upon functional expression within the host organism, which is a function of successful transcription, translation, and precursor availability as discussed in other reviews.[43-44] Given the phylogenetic diversity of microbial SM producers, a number of hurdles must be addressed to successfully identify constructs for functional expression. In addition to optimizing genetic regulatory elements for heterologous expression, differences in protein stability, post-translational modification of biosynthetic enzymes, and precursor availability must be addressed. Alternatively, native expression methods endeavour to activate SM production from within the native producer. In addition to refactoring BGCs via genome editing, several chemical and biological stimuli have been reported over the past few decades that activate SM expression. The concept of exposing a producing organism to an array of growth conditions, One Strain MAny Compounds (OSMAC), is relatively old and several activation stimuli have been reviewed. For example, microbial SM producers have shown responses to subinhibitory antibiotic exposure, vertically acquired antibiotic resistance mutations, rare earth metal exposure,[51] and mixed culture with competing organisms. The central hypothesis of the works presented in this dissertation is that Taken together these phenomenon suggest that SMs are produced by microorganisms to respond to environmental stimuli[68] and this is supported by the apparent BGC

activation selectivity of various stimulatory methods as well as the intrinsically complex nature of SM gene cluster regulation.[69-70]

Regardless, all categories of genome-prioritized NP discovery require a means of measuring the modulation of SM production within the extracted metabolome of native or heterologous producers, and metabolomics methods have been continually adapted to this task. Metabolomics is often defined as the comprehensive study of small molecules within a biological system and provides a direct measure of detectable SM production within an organism of interest. Currently there are two analytical platforms to facilitate metabolome profiling for NP discovery. Nuclear magnetic resonance (NMR) based metabolomic analyses, reviewed elsewhere,[71] are not biased by molecular class and provide enhanced structural information for metabolites but are limited by the inherently low sensitivity of NMR. In contrast, the metabolomic analyses through mass spectrometry (MS), which are most relevant to the work presented in this dissertation, are exceptionally sensitive but are exclusively biased towards ionisable metabolites. The structural diversity of SMs, which span a broad range of functionality, molecular weight, and ionization efficiency, renders comprehensive detection of all metabolites through MS a challenging endeavour, and there is no universal approach for bioanalytical detection. For this reason, the development of metabolomics methods with MS strategies necessitates a discussion of contemporary practices and advances in analytical instrumentation.

As the product of the central dogma, the metabolome also contains information regarding a wide variety of cellular processes unrelated, or indirectly related to secondary metabolism. Correspondingly, metabolomics information may encode insights into how SM producing organisms respond to chemical and biological stimuli and may also provide a means of investigating the biological mechanisms of newly isolated NPs, antibiotics, and chemotherapeutics, from the metabolomic changes engendered within treated organisms. SMs are generally biosynthetic end-products and unlike primary metabolites, they accumulate at higher levels than the fluxes observed in central metabolism. Hence, comparatively

abundant SMs are well suited for comparative metabolomics work-flows. The central hypothesis of the works presented in this dissertation is that SM are used to respond to environmental stimuli and can be prioritized from comparative metabolomic analyses of stimulated culture extracts. The remainder of this chapter will describe the analytical tools and methods used in comparative metabolomics analyses for NP discovery, many of which are employed in Chapters 2-4.

Methods of Generating Inventories of Microbial Metabolites

A variety of MS techniques are available to acquire metabolomics data with corresponding advantages and challenges depending on the analytical descriptor(s) that is/are desired. In each method, the end result of the analysis is a set of metabolomic 'features', ions with a determined mass-to-charge ratio (*m/z*) and potentially additional descriptive information. This additional information may include descriptors such as mass accuracy, chromatographic retention time, isotopic envelope, size and shape information, fragmentation data, and topological distribution, among others. A summary of key descriptors and the information they provide in SM characterization is provided in Table 1-1. As the dimensionality of feature characterization has an impact upon subsequent effectiveness of comparative metabolomics analyses, we will briefly discuss in this section commonly utilized methods for MS acquisition and highlight several of these key molecular characteristics that can be obtained with MS.

Table 1-1: Overview of analytical descriptors for metabolomics-based NP discovery Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[1] with permission from the Royal Society of Chemistry

Analytical Descriptor	Description	Analytical technique
Mass accuracy	Deviation of the experimentally determined *m/z* from the true *m/z*. Expressed as the mass error (*e.g.* ppm), with sufficiently small error an exact chemical formula can be determined.	Mass analyzer • Space-dispersive (*e.g.* ion trap, quadrupole) • Time-dispersive (*e.g.* Time of flight)
Isotopic modeling	Comparison of the abundances of specific isotopes in the molecular isotopic envelope. Can provide rapid indication of amount and identity of heteroatoms.	Chemometrics • Theoretical isotope calculators • Mass defect analysis • Quantitation
Chromatographic retention time	Time required for fluid-solid phase partitioning across a column. Provides separation on the basis of a differentiating characteristic orthogonal to mass.	Chromatography • Hydropathy (e.g. Liquid chromatography) • Volatility (e.g. Gas Chromatography) • Size and charge (e.g. Size exclusion, Charge exclusion, Ion capture)
Ion mobility drift time	Gas-phase electrophoretic separation based on size and shape of the metabolite as ions pass through a gas filled drift tube.	Ion drift tube • Time-dispersive (e.g. Drift time ion mobility, Traveling-wave ion mobility) • Space-dispersive (e.g. Field-asymmetric ion mobility)
Fragmentation	Tandem MS using ion activation to provide characteristic fragment species. Provides metabolite structural information to prioritize which of multiple isomers are the likely identity for a given elemental formula.	Ion activation • Collisional (e.g. Collision induced dissociation, and Surface induced dissociation) • Electron (e.g. Electron transfer dissociation and Electron capture dissociation) • Photoactivation (e.g. Infrared multiphoton dissociation and Wavelength-Tunable Ultraviolet Photodissociation)

Mass measurement accuracy

The mass-to-charge ratio (*m/z*) of detected metabolites is the most useful property used to initiate the process of dereplication. For more than a decade mass analyzers have been able to determine mass accuracy with an error of under 1 ppm.[72] This level of mass accuracy allows for the determination of elemental composition boundaries for compounds under 600 Da[75] when coupled with isotopic mass ratios.[77] While advances in Fourier Transform ion cyclotron resonance MS can now routinely perform at sub-ppm mass errors, typical instrumentation provides mass errors in the range of 1 to 10 ppm (e.g. time-of-flight MS). Unfortunately, this alone is insufficient to confidently dereplicate features, because of the

extensive number of potential isomers for a given elemental composition.[78-80] Early compound dereplication is thereby often dependent on obtaining additional distinguishing characteristics such as those listed in Table 1-1, or via additional characteristic such as UV/Vis spectrum and biological activity. It is also noteworthy that MS analysis is predicated on the ability to generate ions of the species of interest. Neutral or poorly ionizing species are transparent to MS, and because of this the number of detectable compounds from a metabolomic extract will vary depending on the analytical methods used during acquisition, in particular the specific ionization source and ionization conditions that are used.

Isotopic modeling

The isotopic envelope, comprised of both the major and minor isotopic contributions to the elemental formula, provides several opportunities for enhanced characterization information,[81] including: (i) the presence of heteroatoms,[82] and (ii) isotopic enrichment strategies for relative and absolute quantitation of the abundance of the SM.[83] The MS analysis of most biological molecules is typically concerning elemental formula comprising C, H, O, and N. The shared characteristic of these elements is that the monoisotopic peak also corresponds to the lowest mass isotope and thus, the lowest mass peak in the envelope is also the highest abundance. However, the vast majority of the periodic table is characterized by isotopic abundances that are somewhat varied from lightest to heaviest mass isotope and their isotopic signatures are oftentimes used in MS-based atomic analyses for identification purposes. The presence of heteroatoms, such as chlorine or bromine, are readily discernable in their contribution to the isotopic abundance observed for SMs and their stoichiometric contribution can be quickly verified through the use of isotopic calculator algorithms.[84-85] Furthermore, these approaches are equally well suited by the addition of non-natural isotopic enrichment or depletion for determining the relative or absolute abundance of the SM that is expressed. One such approach termed stable isotope labeling in cell culture (SILAC) has been demonstrated as a facile tool for incorporating enrichment or depletion in experimental

protocols. Finally, genomic-based structural predictions, implying biosynthetic precursors, can be combined with stable isotope studies to identify targeted metabolites within organisms. [86]

Chromatographic retention time

Liquid chromatography (LC) is one of the most commonly used approaches to separate individual constituents of complex NP extracts, and various LC methods and their applications have been previously reviewed. For NP separations, reversed phase LC, and hydrophobic interaction chromatography are most commonly employed with a water-acetonitrile, or water-methanol gradient. This is typically performed on the basis of hydropathy, where reversed phase LC (RPLC) and hydrophobic interaction chromatography (HILIC) are most commonly utilized, and column retention will be affected by the ionization of these groups. Mobile phase pH can thereby significantly affect the separation efficiency for NP extracts. Due to the dependence of compound retention on pH, and to assist ionization, mobile phases are commonly buffered with either acetic acid, trifluoroacetic acid, or formic acid to protonate acidic sites and facilitate retention. However, as low pH may suppress detection of negatively charged species in switched scanning modalities, neutral volatile buffers are often preferred.

Liquid chromatography-mass spectrometry (LC-MS) acquisition can take minutes to hours per chromatographic separation, and environmental changes throughout the course of the sample set (column conditioning, instrumental sensitivity and accuracy drift, etc.) can affect the quality of the data. Consequently, for multiple extract samples analysed in a sequential fashion, conditional changes between the start and end of analysis could lead to significant artefactual differences in group metabolomes, which complicate interpretation of subsequent comparative analyses. While challenging, recent reviews have outlined metabolomic experimental design strategies to accommodate these technical problems.

Size and shape by ion mobility

Additional metabolomic feature information can be obtained by using gas-phase ion mobility-mass spectrometry (IM-MS), without significantly increasing analysis time over MS-alone.[92] The mechanism and utility of IM-MS has been the topic of several recent reviews.[93] Briefly, in time-dispersive IM-MS, a uniform weak electric field is applied to a post-ionization ion drift tube containing an inert gas, where the ion velocity through the chamber is dependent upon thermal collisions with the background gas and its charge state. The number of collisions ions make as they traverse the drift cell are proportional to their collision cross-sectional area, providing distinguishing information regarding an ion's shape and/or conformation in the gas phase.[96-100] The separations in IM are very low energy in comparison with collisions used for fragmentation analysis, where in IM the ions experience approximately 10^4 to 10^6 collisions across a size separation versus 1 or several high energy collisions in collision induced dissociation (CID), respectively. Typical drift tube resolving power of IM-MS is sufficient such that conformationally restricted or extended metabolites, such as cyclic peptides, polycyclic polyketides, and polyenes often possess distinct ion mobility profiles that are obtained over the course of micro to milliseconds. IM-MS is often coupled with time-of-flight (TOF) MS that can rapidly acquire the *m/z* ratios for ions eluting from the IM-MS cell in a few microseconds. The frequency of data collection allows for sufficient time sampling across chromatographic peaks, which occur over the course of minutes, to be coupled to IM-TOFMS. When applied to microbial metabolomics the enhanced separation and sensitivity provided by IM-MS has been beneficial for identifying known SMs, dereplication, and prioritizing features. Our laboratory has previously used IM-MS to help obtain high quality fragmentation data (IM-MS/MS) for all detected ions from crude extracts while comparing the differences between antibiotic resistant and wild-type NF. This facilitated the putative identification of several metabolites over-produced in mutant strains. IM-MS has been applied to differentiate halogenated NPs in cyanobacteria as well as peptide NPs from cave actinomycetes. IM-MS has also been applied to investigate the 3-dimensional structures of lasso peptides,

interlocked microbial peptides with a range of bioactivities, and this will likely find other useful applications to NP discovery as the technology becomes more widely available.

Ion fragmentation for structural information

Both time-dispersive (e.g. TOFMS) and scanning mass spectrometers (e.g. liner quadrupole MS and ion trap MS) can be used to acquire both precursor and fragment ion information (i.e. tandem MS), which can provide a wealth of highly specific structural information that can be used to help identify and dereplicate metabolites.[102] In metabolomics-driven NP discovery workflows, fragmentation data is commonly collected via an automated data-dependent acquisition method in which the most abundant ions within a scanning cycle are automatically selected for fragmentation. Fragmentation data analysis facilitates NP dereplication which, as will be discussed in more detail below, is a critical step in the process of NP discovery. There are a variety of methods applied to activate and induce dissociation of target ions, primarily categorized on the basis of how the ion is activated, collisionally, electron attachment, or through photon absorption, where the observed fragment ions will vary based on the method and parameters selected for fragmentation. For small molecules, collision induced dissociation (CID), and surface induced dissociation (SID) are commonly utilized. The degree of fragmentation observed using these methods depends on the number and degree of scissile bonds within a given molecule as well as the resulting internal ion energies used for analysis. In automated data-dependent tandem mass spectrometric fragmentation analysis, a given single set of dissociation parameters may not be appropriate for every feature of interest within a sample, requiring multiple experiments to determine optimal fragmentation parameters, and to effectively capture fragmentation data for a broad cross section of molecular classes. Ultimately, these methods provide characteristic fragmentation spectra that can be compared to established libraries of SM fragmentation data to identify known SMs within the experimental sample. Additionally, tandem mass spectrometric data are useful for elucidating the

structures of peptide NPs and have been used in 'peptidogenomics' strategies to link ribosomal and non-ribosomal peptide NPs to their cognate BGCs.[110-116]

Leveraging spatiotemporal metabolomics inventories to capture inter-organism interactions

SM producing microorganisms can be cultivated on agar medium or in planktonic liquid culture medium, and several methods have been developed to extract and chromatographically separate resulting metabolomes.[123] However, microorganisms cultivated as monocultures or mixed cultures on agar may display planar metabolite distributions containing valuable information about chemical pleiotropism, nutrient dependence, and chemical ecology, and bulk liquid extractions discard the spatial metabolomic feature differentiation that could otherwise be observed.[51, 127] Correspondingly, imaging mass spectrometry (IMS) methods have been developed for agar cultivated[126, 130] and environmental[124, 131-136] microbial samples to provide a second distinguishing ion characteristic, spatial localization. In IMS experiments, the area of a sample is divided into pixels which are individually analysed by the mass spectrometer. Matrix assisted laser desorption ionization (MALDI) is a commonly used ionization technique for IMS. MALDI requires the application of an ionization matrix to facilitate ionization of cell and agar embedded metabolites. This ablative technique has been applied to visualize spatial temporal distributions of SM production in marine cyanobacteria and to elucidate microbial producers responsible for observed SM biosynthesis. MALDI-IMS has also be used to visualize metabolic exchange between interacting organisms and identify novel antibiotic production in *Streptomyces*.[143] The efficiency of MALDI ionization is matrix dependent, and varies across metabolite classes. Correspondingly, desorption electrospray ionization and secondary ion mass spectrometry, which do not require the addition of an ionizing matrix, have also been applied to visualize NP distribution through IMS[144] among others.[149] Determining the spatial distribution of produced SMs can be useful in NP research, and as these IMS technologies continue to develop they are expected to become an integral component of metabolomic

investigations into microbial SMs.[150] Metabolomic features generated via IMS consist of *m/z* and its corresponding Cartesian coordinate in agar culture.

Preparation of High Content Mass Spectral Data for Metabolomics Studies

Strategies for formatting data for effective comparative analysis

In a general metabolomics experiment (Figure 1-1) LC-MS analysis of extracted metabolites results in thousands of detectable features characterized by *m*/z, and retention time, as well as potentially ion mobility and fragmentation. Unbiased manual comparison of features between samples is challenging, especially when analysing a large number of samples. Therefore, it is necessary to automate feature collection from the acquired data in several processing steps that will facilitate data analysis. There are a variety of non-compatible vendor-specific data file formats for mass spectrometric data which originally impeded the development of universal processing software. To address this issue the Protein Standards Initiative group developed a standardized format, mzData, to facilitate data exchange.[155-156] An additional format, mzXML, was developed by Pedrioli et.al.to serve as a standard format for MS and MS/MS data processing.[157-159] While both of these formats were popular, the scientific community pushed for a unified standard format to simplify software development. A new format mzML was released to replace both mzXML and mzData formats, however, all of these are still commonly used for metabolomics data. Correspondingly, one of the first steps in metabolomics data processing is to convert the data files from a vendor-specific format into one of the appropriate standard formats listed above for the processing software. One common utility for this is ProteoWizard's MSconvert,[161] which also has the ability to pre-filter the data with user defined parameters.

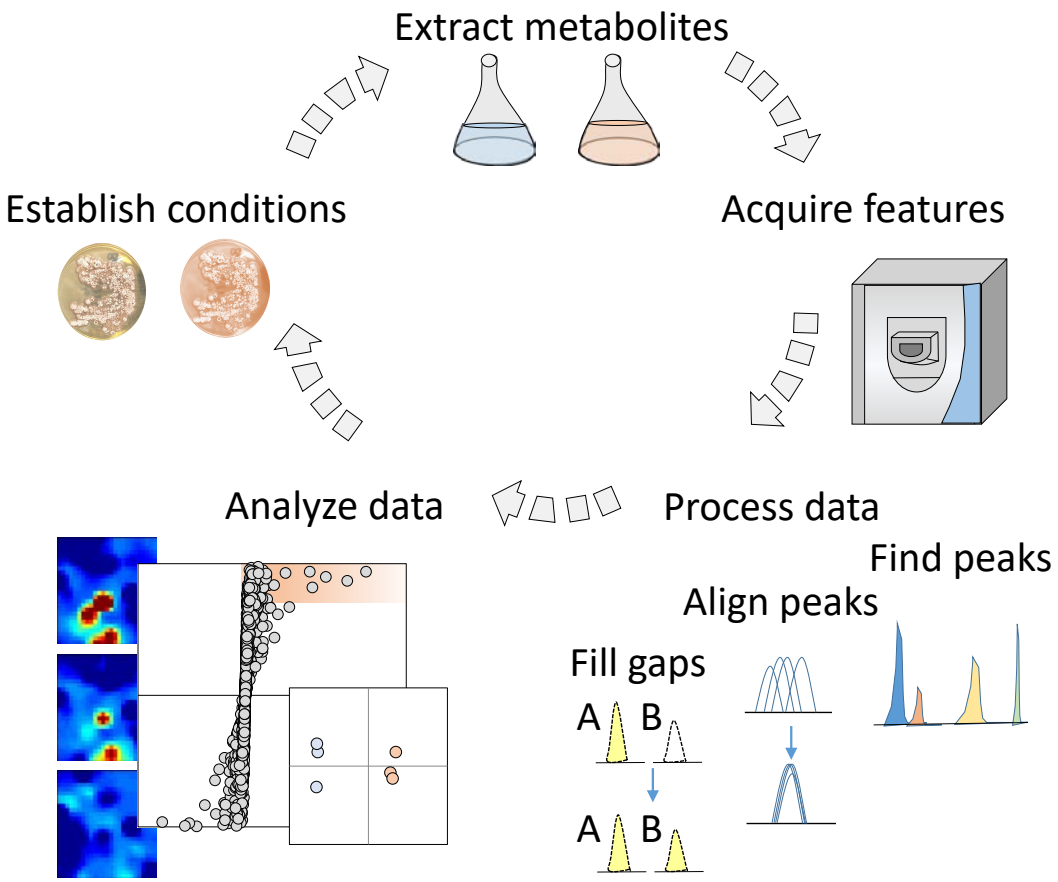


Figure 1-1: General metabolomics workflow. Metabolites are extracted from experimental conditions and detected through MS analysis. MS data is then formatted and processed before undergoing statistical analyses to determine important metabolomic changes between the sample groups. These results may then be used to direct new experiments to optimize SM production or test biological hypotheses generated from the initial experiment. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[1] with permission from the Royal Society of Chemistry

Methods and considerations for metabolite peak detection and alignment

After data format conversion, metabolite peaks must be identified and extracted from the data and aligned for all samples. A number of reviews have covered and compared the various processing packages and their algorithms.[162] In this section we highlight a few of the common computational methods employed NP discovery-based metabolomics. The initial peak identification can be fairly challenging, as LC-MS ionization methods typically generate high levels of background chemical noise largely from mobile

phases and buffers. Therefore, the automated processing methods must be able to identify genuine sample features while omitting detected chemical background and instrumental noise, and there have been several algorithms developed to accomplish this task Vectorized peak detection algorithms identify data points above a set intensity threshold in both the *m/z* and retention time dimensions.[166-167] There are also a number of 1-dimensional LC-MS processing algorithms commonly used for peptide analysis which detect peaks by using the isotope patterns in the *m/z* dimension.[161, 168] Another of the more common methods involves separating the LC-MS data into extracted ion chromatograms (EIC), each covering a very narrow *m/z* range. This process is called binning and, while fast and generally effective, this can lead to problems if the bin size is too large or too small. A matched filter[170-176] is commonly applied to EICs to select for *m/z* peak shapes in the chromatographic time domain, and if features are split between multiple bins due to inappropriate sizing, they can be excluded by the algorithm resulting in false negatives. The traditional XCMS peak detection algorithm, a widely used LC-MS processing software package, sections off 0.1 Dalton wide EICs and then applies a second derivative Gaussian filter that aids in the discrimination of authentic peaks from noise along with a 10:1 signal to noise intensity threshold. An alternative to the binning approach for high resolution MS data is the centWave algorithm which identifies ion dense regions of interest in centroid data. Peaks are detected along these regions using a continuous wavelet transform, which allows for a much more dynamic range of peak shapes. The quality and validation of peak detection from increasingly complex datasets remains an area of intense research efforts.

Another consideration in data processing is the tendency for retention times of features to vary between multiple injections due to changes in chromatographic conditions discussed previously. It is therefore necessary to match mass features between samples of an experiment and align the retention times of matched peaks to generate a discrete feature list. Originally, internal reference standards were used to adjust retention times of each sample.[177] However, retention time drifts throughout an acquisition

are often not linear,[179-180] and this also required additional sample preparation steps to incorporate the standards. A variety of algorithms have been developed to align features between sample runs without the use of internal standards. The original XCMS alignment algorithm identified hundreds to thousands of peak groups that are present in a large number of samples. These "well behaved" groups are used as markers to align the remaining detected features. Typically, the number of these markers identified from metabolome extracts is sufficient to cover the chromatographic profile of samples and correctly align the nonlinear retention time drifts. Local regression, LOESS,[181-182] is then used to approximate drifts for regions without sufficient peak markers. Several alignment algorithms have been developed to process LC-MS data, and in a comparative study of six freely available retention time alignment methods the XCMS algorithm was shown to be the best for processing metabolomics data.[189] However, it was noted that the appropriate selection of parameters used for the methods could have a large impact on the data output, such that the apparent success of any particular method is dependent upon the user's experience. A software package, Isotopologue Parameter Optimization (IPO), was recently released to automatically optimize XCMS parameter settings using natural C13 isotopic peaks.[39-40, 190-191] This software applies to a variety of different sample types, chromatographic strategies, and instrument methods and aids to simplify and systematize method development while optimizing metabolomics processing for non-experts.

After peak alignment it is common for several mass/retention time features to possess few or even no matches between samples. This may be because some peaks are entirely unique to a subset of experimental samples but can also stem from errors in peak detection due to inappropriate parameter settings, noisy data, etc. Gap-filling is commonly used to ensure these are not false negatives and provide a non-zero value for subsequent statistical analyses. In the absence of a detectable peak, the values obtained through gap filling reflect noise within the region peaks that were detected in other samples. For low abundance features, the integrated noise level over the peak region may be similar to the value determined for the feature, and this can lead to the observation of a metabolite ion that statistically

correlates with a single condition while its lower abundance isotopes show no correlations in subsequent statistical analyses.

Analysis of metabolomics data in the context of secondary metabolites

Table 1-2: Overview of methods for metabolomic data analysis Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[1] with permission from the Royal Society of Chemistry.

Method	Description	Applications	Disadvantages
Principle component analysis and Projections to latent structures	MVSA to identify significant covariance within data	• Identifying data outliers • Strain prioritization • Grouping samples • Compound prioritization	• Less effective with large datasets
Self-organizing Maps	Organizes features into a 2-dimensional map based on feature response trends across a variety of experimental conditions	• Grouping samples • Compound prioritization • Comparing large numbers of experimental conditions	• Less effective with small numbers of conditions
Molecular networking	Organizes features into a connectivity network based on similarities in molecular	• Compound prioritization • Compound dereplication	• Fragmentation can vary with instrument parameters

The next stage of metabolomics analysis consists of applying one or more methods to compare metabolomics datasets, some of which are outlined in . Depending on the objectives of a given study, several complementary methods may be applied. In the following discussion we review extant methods for comparative metabolomics analysis. To illustrate the application of these methods, we apply them to the analysis of a metabolomics dataset focused on the cytotoxic macrolide producing organism, *Nocardiopsis sp FU40* (NF), and its exposure to multiple competing organism in mixed culture. In selected mixed culture conditions, this organism increased production of the SMs called ciromicins, which we highlight throughout data analyses.

Multivariate statistical analysis for identification of abundant covarying metabolites

Subsequent to pre-processing, metabolomics data can be analysed through multivariate statistical analyses (MVSA) which simplify and identify significant correlations within the data. Two common methods for metabolomics data analysis are partial least squares (PLS), or projections to latent structures, regression methods and principal component analyses (PCA) reviewed in more detail elsewhere. Briefly, PLS methods assume that changes within the data are largely driven by a subset of latent variables, which are not themselves measured within the data but are more abstract, such as experimental treatments/stimulants or conditions. With this assumption, a PLS analysis will identify latent vectors within the data that describe the maximal covariance between user defined groups. Alternatively, PCA makes no assumptions about the data and identifies the sources of the highest variance across the samples to distinguish the samples from one another. The fundamental difference between these two analyses is that PLS are supervised with user defined groups while PCA are unsupervised variable reduction methods. Orthogonal signal corrections can be applied to PLS regressions to improve separation between predictive and non-predictive variation. The product of these analyses are scoresplots, or projections of samples onto a hyperplane within the data describing sample covariance, from PCA and PLS analyses. Interpretation of scoresplots show the separation of samples based on feature variance to determine which samples are similar (nearby in Cartesian space) and dissimilar (far away) with regards to their most significantly varying features. Replicate analyses of the same sample should cluster within the scoresplot, and in this way scoresplots are a useful means of identifying errors in sample acquisition or data pre-processing. Additionally, a control comprised of pooled samples should locate close to the origin of a PCA plot. Another useful product of the PCA analyses are loadings plots, which show correlations between variables in the data and summarize these variables' impacts on the scoresplot. Nearby features

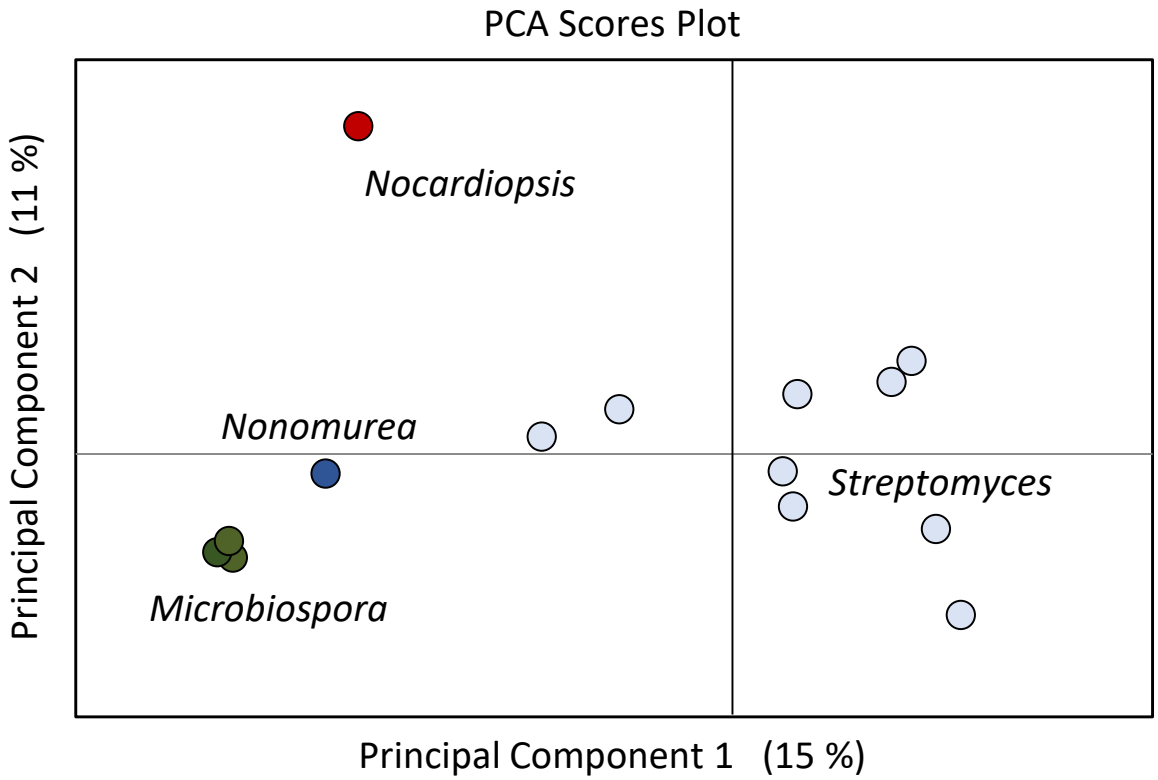


Figure 1-2: Using a PCA scores plot to prioritize microbial producers. A panel of actinomycetes including Microbiospora, Streptomyces, and Nonomurea genera. In this analysis, 14 strains grown under identical conditions were compared and PCA was used to display metabolomic feature variance between the strains. Principal component 1 primarily groups Streptomyces from other strains, and component 2 further distinguishes Nocardiopsis sp. FU40 as metabolomically unique compared to other tested strains. Percentages shown in parentheses correspond to the variance between the samples contained within the specific component. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24 with permission from the Royal Society of Chemistry.

are positively correlated, while distant features are negatively correlated, and features in the same region as samples in the scoresplot will be more abundantly or uniquely present in those samples.

One approach to the discovery of new NPs has been to prioritize organisms distinguished as metabolically unique through a PCA analysis. There is often a great deal of redundancy in the compounds identified through microbial NP screening endeavours, and this redundancy can be reduced through the selection of metabolomically diverse microbial strains. Under the hypothesis that organisms with similar secondary metabolic potential would cluster in PCA space, Hou et al. analysed 47 microbial strains to

demonstrate how MVSA could prioritize strains with diverse secondary metabolic potential.[50, 192-194] Similarly, PCA has been used to prioritize marine microbial symbionts as well as phylogenetically similar *Streptomyces* for NP isolation. Figure 1-2 demonstrates how metabolically unique organisms are distinguishable along the principal component vectors of the PCA scoresplot. This method may also be useful for identifying new classes of bioactive microbial compounds as has been done for plant extracts. However, caveats to this approach include (1) that the correlating features responsible for PCA prioritization of a subset of organisms from a library may not be SMs, which are generally present in relatively low abundance within crude extracts, and (2) that low abundant SMs, will not be emphasized by these methods.

One important application of PCA and PLS metabolomics for NP discovery is to prioritize induced SMs in comparative analyses between chemically and/or biologically stimulated and control conditions. SM production can be activated in microorganisms through a variety of chemical and environmental stimulation,[195] and PCA and PLS are commonly applied to identify abundantly produced features in these conditions.[201-202] Binary comparisons using S-plots can be a used to identify group specific features of a PLS model. These graphs separate features by their covariance along the x-axis and their correlation to user defined groups on the y-axis. More simply, more abundant features are farther from the origin on the x-axis, and features with correlations closer to 1 or -1 are likely to be unique or specific to one group or the other. Volcano plots have also recently been used to identify significantly covarying metabolites in binary comparisons of NP extracts. Volcano plots show each features' statistical significance, p-value, on the y-axis and fold change along the x-axis. Similarly, PCA loadings plots can be used to visualize significant feature differences between sample sets. Figure 1-3 demonstrates how S-plots, volcano plots, and loadings plots can distinguish induced metabolomic features in a case study with the ciromicin producer,

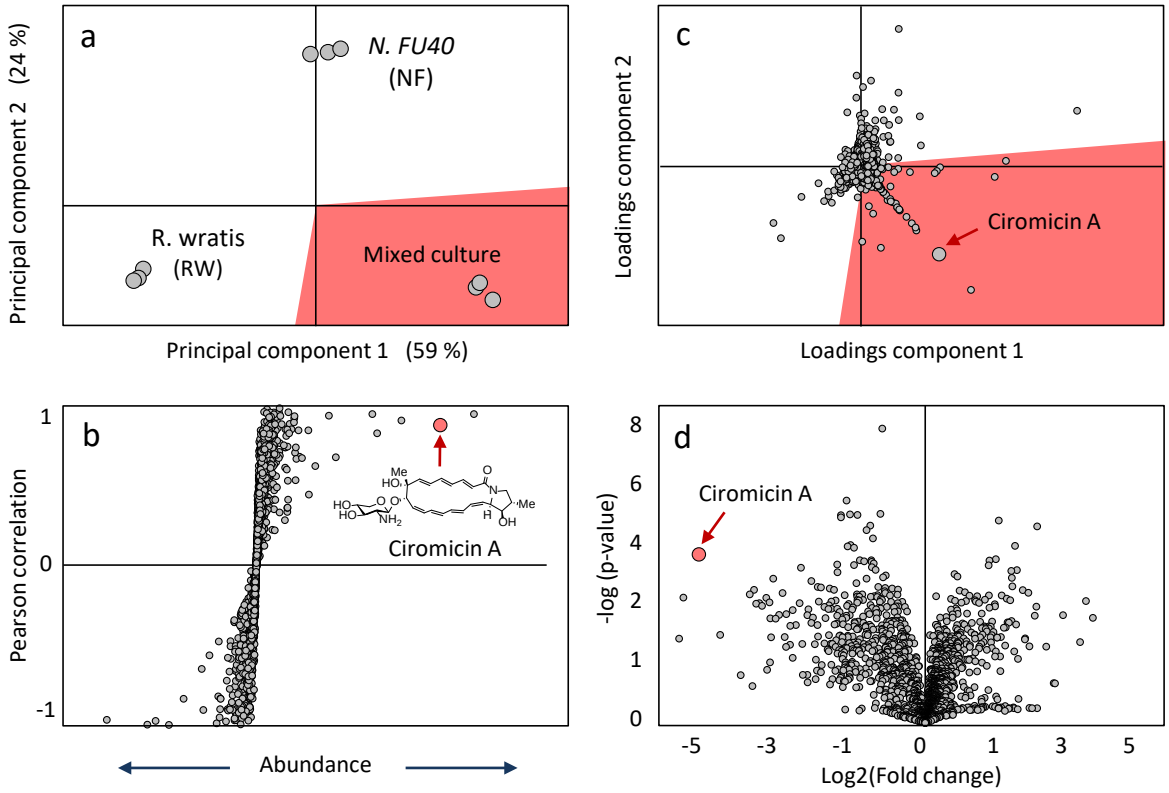


Figure 1-3: MVSA S-, Loadings, and Volcano plots to identify induced features. (a) The scoresplot reveals group separation between the Nocardiopsis monoculture (NF), the Rhodococcus wratis competitor monoculture (RW), and the mixed culture (RW&NF). (b) S-plot shows ciromicin significantly correlates ($p < 0.1$) to the mixed culture group in a binary comparison vs the Nocardiopsis monoculture. (c) Loadings plot of features shows ciromicin contributes significantly to group differentiation on the PCA scoresplot. (d) Volcano plot also prioritizes ciromicin ions which have a high correlation (low p-value) on the y-axis and high fold change on the x-axis. Shading in panels a and c are used to highlight the data corresponding to the different sample subtypes. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[9] with permission from the Royal Society of Chemistry.

Nocardiopsis FU40 (NF), described in more detail in Chapter 3 . The loadings plot tripartite comparison identifies features that correlate with either the NF or *Rhodococcus sp.* BBSNAI13 (RW) monocultures or a mixed culture where the two compete for nutrients. In this plot the induced cytotoxic macrolactam ciromicin is clearly distinguishable as positively correlated with the mixed culture extract. Similarly, ciromicin was clearly identified through the S-plot and volcano plot comparisons between the NF monoculture and the mixed culture. These methods can be very powerful, and freely available online metabolomics packages, such as XCMS Online and

Metaboanalyst can perform some routine MVSA data analyses in addition to data pre-processing. An alternative and fairly unique comparative analysis available through XCMS Online is the cloud plot.[206] These plots convey feature fold changes, *m/z*, retention time, and statistical distribution in the same Figure, and can perform both binary and multigroup comparisons.

Discovering molecular inventories of microbial responses via self-organizing map analytics

A strength in MVSA analysis of metabolomics datasets is the identification of the most unique and abundant features between small numbers of treatment conditions. However, these methods are limited to displaying data in two or three dimensions and are biased towards the largest differences within the entire dataset. Therefore, the utility of MVSA to represent an experiment diminishes as the number and diversity of samples increases. For instance, it is common to screen a target organism under dozens of stimulus conditions to optimize compound production, or to induce silent BGCs, and in these cases we have previously demonstrated that an alternative method utilizing Kohonen self-organizing map (SOM) analytics can be more effective at representing multiplexed stimuli data than PCA. As discussed above, metabolomic acquisition via LC-MS results in the acquisition of thousands of detectable features. Through SOM analyses these features are organized using an artificial neural network into a 2-dimensional grid based on feature response patterns across all experimental conditions. Features that share similar trend patterns are grouped in nearby nodes of the map as shown in Figure 1-5. Through multiple iterations, typically several hundred, this organization is improved ultimately resulting in a feature map where features in this case correspond to clusters of similar response trends. Unlike MVSA, SOM analyses improve with increasing amount of data and response conditions (e.g. stimuli), as this leads to more varied response trends which in turn enhances feature organization. A metabolomics workflow – molecular expression dynamics inspector (MEDI), provides an open

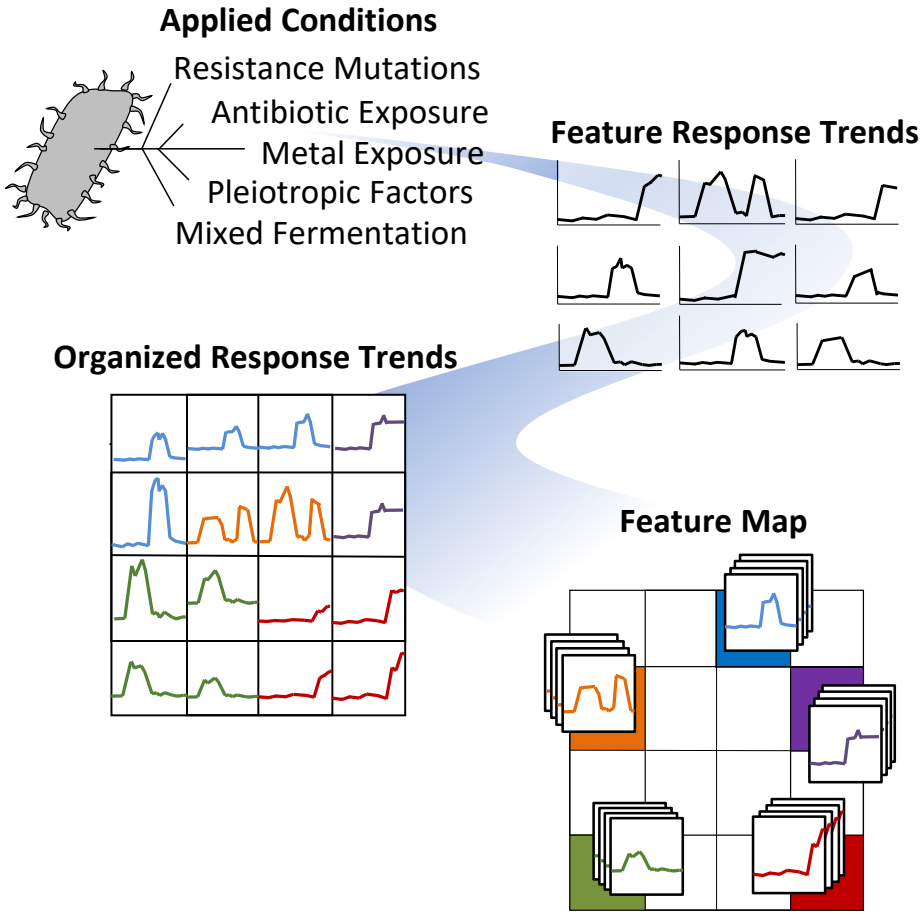


Figure 1-4: Feature organization within a SOM. Feature abundance profiles are illustrated for each feature as a response trend across all experimental conditions shown in the upper right. These trends are organized for similarity as shown on the bottom left. These organized data serve as the basis for visual heatmap representations of the observed metabolomic content of experimental cultures. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[1] with permission from the Royal Society of Chemistry.

access methodology for SOM analysis from MS data and is readily applicable to microbial metabolomics.[208] In MEDI, each tile, or node, of the grid is colored based on the centroid intensity of its features to generate heatmaps. Difference maps can be generated by subtractive analysis (e.g. control and stimulus conditions) to readily prioritize abundant and treatment-specific metabolomic features into regions of interest. For our first applications of these SOM analytics, described in Chapter 2 we mapped stimuli-induced metabolomic responses from 23 distinct

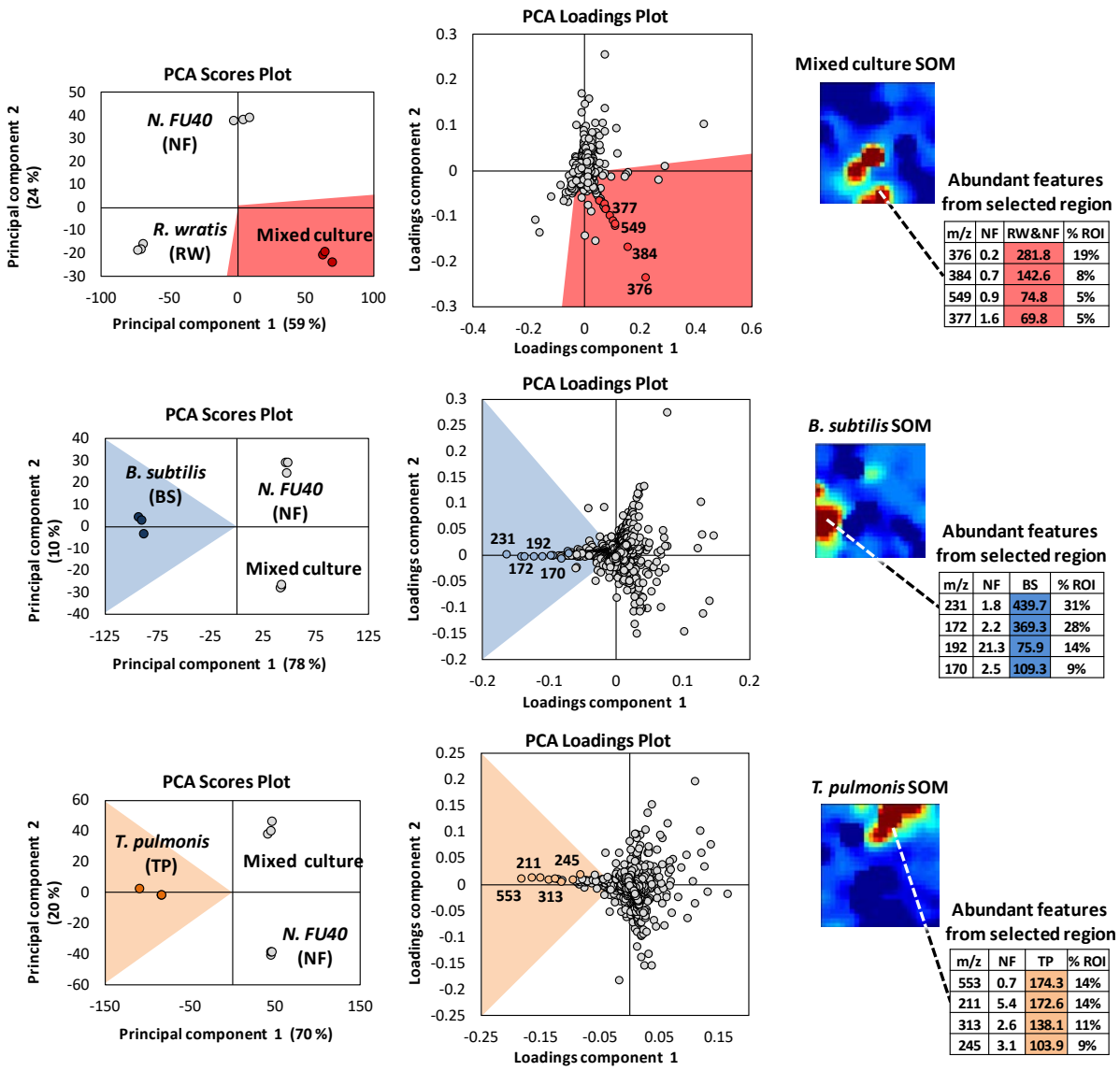


Figure 1-5: Three example comparisons of prioritized features through PCA and SOM analyses on mixed cultures with *Nocardiopsis FU40* (NF), *Rhodococcus wratis* (RW), *Tsukamurella pulmonis* (TP), and *Bacillus subtilis* (BS). Features prioritized within SOM regions of interest recapitulate PCA tripartite analyses when sorted by abundance, or percentage of the region of interest (% ROI). Shading in scores and loadings plots used to highlight the data corresponding to the different sample subtypes. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[6] with permission from the Royal Society of Chemistry.

conditions in *Streptomyces coelicolor A(3).*[209] In our NF case study several metabolites, including the ciromicins, are readily prioritized as unique to mixed culture conditions (Figure 1-5). Additionally, the single SOM analysis recapitulates the results of multiple MVSA analyses. In Figure 1-5 three PCA loadings plots are compared with three SOM heatmaps from the NF mixed culture

example study. When the features held within regions of interest on the SOM maps are sorted by abundance they are highly consistent with the loadings plots from PCA analyses. Indeed, there is correspondence between PCA and the neural networks used for MEDI analysis. As the data and/or conditions become sparser, the SOM heatmap begins to decompose into a similar functional form as PCA.

Molecular networking to reveal structurally related ions

Microorganisms have been extensively mined for NPs throughout much of the past century in the search for new pharmaceuticals, and the rediscovery of known compounds or known families of compounds is quite common. Identifying and removing these rediscovered NPs, a process known as dereplication, is both critical and challenging.[98, 210] Typically accurate masses or determined molecular formula of extracted compounds are used to search databases of known NPs. However, the large number of isobaric compounds complicate dereplication. UV/Vis absorbance[205] spectra and chromatographic retention times can be used to further match extracted features to database compounds, and as technologies and databases improve, it is likely that ion mobility will play a role in NP dereplication as well. Fragmentation spectra acquired through tandem MS is another useful property for dereplication. Metabolite fragmentation patterns observed through MS/MS analysis can be matched to those in databases like PubChem, METLIN[213] and MassBank[214] to putatively identify MS features. Kernel based machine learning algorithms have recently been applied to dereplicate metabolites using multiple levels of tandem MS, and while this works well for primary metabolites, public databases for microbial SMs with fragmentation spectra encompass only a small fraction of known NPs. For example, GNPS: Global

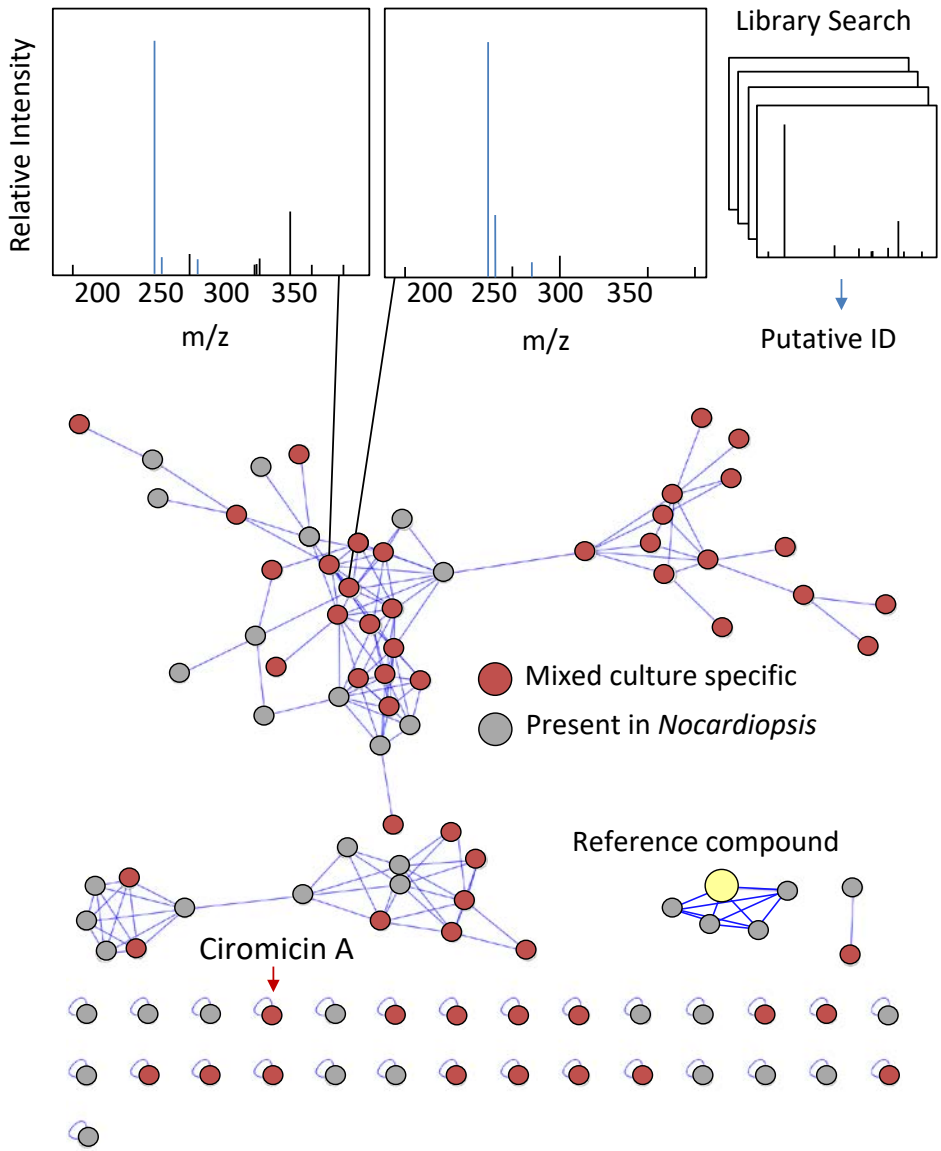


Figure 1-6: Applications of molecular networking to explore data. Comparisons of acquired fragmentation spectra to established databases facilitates putative feature identification. Connectivity between features shown with blue lines relates structural similarities. Reference compounds seeded into the network can identify structural analogs. Feature distributions between experimental conditions are indicated by node coloring, red for mixed culture specific, and grey for features detected within the monoculture. Reproduced from *Nat. Prod. Rep*. **2017,** 34, 6-24[3-5] with permission from the Royal Society of Chemistry.

Natural Products Social Molecular Networking, the largest NP public database with MS/MS spectra, contains more than 140,000 NPs, and there are an estimated 600,000 published natural compounds.

Computational methods to generate theoretical fragmentation spectra have been employed to compensate for the lack of experimental data on NPs. These in-silico MS/MS spectral databases can further facilitate NP dereplication when coupled with molecular networking, and as both experimental and in-silico database coverage improves, comparisons of fragmentation spectra may become the most useful method of NP dereplication. In addition to matching fragmentation spectra with database compounds, fragmentation data can be used to cluster related classes of molecules by fragment similarity. Molecular networking analyses cluster families of molecules through vector correlations between fragment ions. Yang et al. demonstrated the utility of this approach for NP discovery by dereplicating 58 NPs from marine and terrestrial microorganisms.[220-222] Molecular networking in this study also identified a number of novel analogs to known compounds, which are more difficult to obtain through other dereplication methods. In Figure 6 we have applied molecular networking to our NF example dataset. Using the network visualizer in the GNPS: Global NPs Social Molecular Networking website, fragmentation spectra for each object in the network can be easily viewed and compared to matched reference spectra from the GNPS library. The features in the network can also be colored by the user-defined group or condition to which they are correlated. In Figure 6 we have highlighted features unique to mixed culture conditions in red. As shown, molecular networking identifies unique features which is unbiased by compound abundance. Several features of this dataset share no significant fragment similarity with the network and are isolated as “self-loops”. In fact, ciromicin A is among these uniquely fragmenting features, and this in itself may be another useful means to prioritize leads, as outlying features may be more structurally unique. Molecular networking analysis can be enhanced by combining additional metabolomics techniques. Klitgaard et al. used a combination of molecular networking and stable isotope labelling to identify novel analogs of nidulanin A and fungisporin in the well-studied fungus *Aspergillus nidulans.* Fragment based clustering in this manner can also

be used to identify modified NPs stemming from interactions between organisms. Moree et al. used molecular networking with IMS to investigate the interkingdom interactions between *Pseudomonas* and *Aspergillus* and observed a variety of biotransformed metabolites arising from this microbial competition. Similarly, Briand et al. applied molecular networking to identify new compounds and analogs arising from intraspecific interactions between algae.[226] The application of molecular networking for the NF mixed culture data shown in Figure 6 links a number features found in the NF monoculture with similarly fragmenting features only detectable in mixed culture. These may represent compounds made by NF that are stimulated or modified in some way by the competitor TP.

Molecular networking can also prioritize features by linking observed NPs to their cognate BGCs and gene cluster families[227-228] when used in conjunction with genomic sequence analysis. This can be an advantageous means of prioritizing metabolite leads as demonstrated by the work from Kleigrewe et al. where molecular networking was combined with genomic sequence analysis to identify a novel group of acyl amides, termed columbamides, from marine cyanobacteria.[229-232] The Crawford lab has recently employed 'pathway-targeted' molecular network analyses to identify metabolites from the colibactin gene cluster, which had been linked to increased virulence in *E. coli.* As previously discussed, heterologous hosts are often used for the production of microbial SMs, and molecular networking is a useful tool for comparative metabolomics to visualize the output of these heterologous hosts. Schorn et al. used molecular networking to identify novel eponemycin congeners produced through heterologous expression in *Streptomyces albus* J1046.[49, 233] Molecular networking has even been applied to identify virulence factors in pathogenic organisms,[234-237] and this method will become more beneficial for NP discovery as databases and technologies improve.

Investigations of secondary metabolite bioactivity

NPs are intrinsically biologically active; however, the clinical relevance of this activity may not always be discernible. Typically, NP structure and mode of action are determined fairly late in the NP discovery pipeline, which contributes to high rediscovery rates. Therefore, prioritizing NP leads by deep profiling of pharmacologically relevant biological activities would expedite NP-based drug discovery. NP extracts are commonly divided into multiple fractions which are then screened to identify the components underlying the desired biological activity. However, low abundance compounds can often be overlooked in complex extracts, and recently MVSA have been utilized to help link observed fraction bioactivity to detectable features from metabolomic analyses.[238] Even after correlating metabolites with biological activity, determining the mode of action for active compounds can be difficult and expensive.[239-240] One approach has been developed that uses the antibiotic spectrum of activities across different organisms, mode of action profiles (BioMAP), to group similar antibiotics. This method was effectively able to cluster antibiotics of the same compound class and led to the identification of a novel naphthoquinone antibiotic, arromycin. Gene expression profiling with either the entire transcriptome or a subset of reporter genes has also been used to predict modes of action for NPs. However, because these transcriptomic screens are still relatively costly, there is a great interest in applying metabolomics analyses to predict NP modes of action using either NP extracts or purified compounds.[250] Vincent et al. have recently shown untargeted metabolomics can effectively identify compound modes of action when specific metabolic pathways are the primary drug target.[254-255] Metabolomic consequences of drug combinations may additionally be able to identify synergism, or antagonism between coadministered drug therapies.[256] In a study with *M. smegmati*s, Halouska et al. observed that antibiotics which share similar biological targets engender similar metabolomic changes and are grouped together through MVSA.[257] The group additionally applied their metabolomic

methods to investigate antibiotics with unknown biological targets and found them to group with membrane disrupting antibiotics, ampicillin, D-cycloserine, and vancomycin.[257] This methodology could prove very useful to prioritize compounds for isolation. Antimicrobial extracts which separate themselves metabolomically through MVSA or other analyses may exert their activity through a novel biological target or mechanism. In this way, pharmaceutically relevant NPs could be prioritized for isolation. These metabolomic analyses have even been applied to investigate the underlying methods by which known antibiotics kill pathogens.[242] Another approach, cytological profiling, uses automated image and microscopy analyses to identify phenotypic changes induced from bioactive compounds,[14, 258-262] and this method has been used to classify biologically active compounds by their respective modes of action even within more complex marine derived bacterial extracts. A combined approach integrating these phenotypic screens with untargeted metabolomics has recently been developed to predict the modes of action for complex libraries of NPs and prioritize unique bioactive components. Applying this method, Compound Activity Mapping, on data from 234 NP extracts led to the discovery of the quinocinnolinomycins, a new family of NPs implicated to induce endoplasmic reticulum stress based on further cytological profile clustering. Ultimately, these multi-omic combinatorial methods may become the preferred means of predicting molecular modes of action. Integrating the phenotypic data from cytological profiling and the transcriptomic functional signature ontologies with metabolomics data using one or combinations of the powerful analytical platforms discussed previously, SOMs, molecular networking, MVSA, etc., could provide new insights into the modes of action of bioactive compounds and greatly facilitate novel drug discovery.

Conclusions

Metabolomic analyses are powerful tools for NP discovery. However, while metabolomics can provide a wealth of information regarding the activity and responses of microorganisms, with current technologies it is practically impossible to analyze the entire metabolome of an organism comprehensively due to variations in ionization efficiency and limitations in detection across wide concentration dynamic range. Instead, only detectable metabolites, which make up a fraction of the total metabolites present, are used to draw conclusions from current studies. While the full transcriptomic and proteomic potential of an organism can be determined through modern genome sequencing, there is no readily discernible limit to the number of metabolites present within organisms, so it is difficult to predict the number of metabolites omitted by current analyses. Due to these limitations, extra care must be taken when drawing conclusions from metabolomics datasets. Nonetheless, metabolomics analyses benefit microbial NP discovery pipelines in a variety of ways as described in this report. These can be used to prioritize organisms, identify activated compounds from stimuli exposure, prioritize features through bioactivity spectrums or molecular class, and even dereplicate prioritized SMs. The metabolomics methods described herein may also facilitate investigations into the fundamental purpose behind SM production within microbial communities. It is largely unclear how the production of SMs is regulated *in situ* as well as which ecological stimuli trigger secondary metabolic production. Such studies would benefit NP discovery endeavors by facilitating predictions of stimuli to induce SM production within the endogenous producer and could additionally provide insight into human health and wellness. Microbial SMs have a significant impact on human health by means of both isolated pharmaceuticals and compounds produced *in situ* from within human microbiomes, and metabolomics may be able to offer insights into how these organisms modulate their secondary metabolism in response to diet, medicine, and endogenous host factors.

Comparative metabolomics methods are aiding in unleashing the repressed and/or hidden wealth of microbial secondary metabolism predicted by whole genome sequencing. The combination of complimentary methods (e.g. SOM and molecular networking) has the potential provide new tools to accelerate discovery by comparison. Ultimately, the purpose of these efforts is to identify biological roles for SMs, be they biochemical, chemical ecological, or translational in human medicine. We feel that the next era in SM discovery and application will be facilitated by methods that combine high content biological activity data measurements for metabolites within metabolomes with corresponding multidimensional metabolomic data to illuminate effectors of natural small molecule interactions, and their roles in biological systems.

Acknowledgements

This work was supported by the National institutes of Health (no. R01GM092218 awarded to B. O. B. and J. A. M., and T32 no. GM0650086 awarded to B. C. C.), the Vanderbilt Institute of Chemical Biology, the Vanderbilt Institute for Integrative Biosystems Research and Education, and the Vanderbilt University College of Arts and Sciences.

Dissertation Statement and Chapter Contributions

Dissertation Statement

There is an urgent need for novel antibiotic and chemotherapeutic compounds, and traditionally NPs have been important sources for these therapeutics. At present most clinically relevant antibiotics and chemotherapeutics are either NPs themselves or NP derivatives,[1] with many of these produced by microorganisms from soil environments. Through advances in genome sequencing we recognize only a small fraction of the total arsenal of microbial NPs has been

accessed in a large part due to poor productivity in lab cultivated strains. Several methods exist to boost NP production within cultured strains,[3-4] but detection of NPs from complex culture extracts remains very challenging. In this dissertation work we endeavored to apply (1) established methods to boost NP production within actinomycetes and (2) comparative metabolomics to prioritize NPs from their response to stimuli. In the works presented in this dissertation we demonstrate that the combination of stimuli and comparative metabolomic analyses can increase production of NPs and prioritize NPs by their response to applied stimuli.

My contributions to works presented in each chapter

For Chapter 2, I worked with Dr. Ruth McNees and Dagmara Derewacz to culture and extract *Streptomyces coelicolor A3(2)* grown under a battery of stimuli conditions including mixed culture, and rare earth metal exposure with Sc and La. I also worked with Dr. Cody Goodwin to analyze the metabolomic changes engendered through stimulation conditions using self-organizing map analytics. For Chapter 3, I acquired the data using UPLC-IM-MS on a Waters SYNAPT G2 and performed all of the metabolomic analyses comparing mixed cultures of *Nocardiopsis FU40* with *Rhodococcus* sp BBSNAI13, *Tsukamurella pulmonis*, *Bacillus subtilis*, and *Escherichia coli*, to monoculture conditions. For Chapter 4 I participated in the isolation and 16S rDNA sequencing of several of the cave derived actinomycetes. I cultured and extracted the 20 selected cave actinomycetes under 6 different stimuli conditions with assistance from a student worker, Zac Hilton. I acquired data for all extracts on a TSQ Thermo Quantum Access Max and performed comparative metabolomics analyses. I isolated and identified all described NPs including the novel compound funisamine. I determined the structure of funisamine using accurate mass measurements and multidimensional NMR analyses. I additionally performed bioinformatic

analyses to identify the funisamine BGC from the genome of the producer, *Streptosporangium* sp. KDCAGE35.

References

1. Covington, B. C.; McLean, J. A.; Bachmann, B. O., Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Nat. Prod. Rep.* **2016,** *34*, 6-24

2. Hicks, L. A.; Jr, T. H.; Hunkler, R. J., US outpatient antibiotic prescribing, 2010. *New Engl J Med* **2013,** *368* (15), 1461-1462.

3. Nathan, C., Antibiotics at the crossroads. *Nature* **2004,** *431* (7011), 899-902.

4. Franz von, N.; Michael, B.; Berthold, H.; Stefan, W.; Dieter, H., Antibacterial Natural Products in Medicinal Chemistry—Exodus or Revival? *Angew Chem Int Edit* **2006,** *45*.

5. Fischbach, M. A.; Walsh, C. T., Antibiotics for emerging pathogens. *Science* **2009,** *325* (5944), 1089-93.

6. Clatworthy, A. E.; Pierson, E.; Hung, D. T., Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol.* **2007**.

7. Newman, D. J.; Cragg, G. M.; Snader, K. M., Natural products as sources of new drugs over the period 1981-2002. *J Nat Prod* **2003**.

8. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **2007,** *70* (3), 461-477.

9. Bérdy, J., Bioactive microbial metabolites. *J Antibiot (Tokyo).* **2004,** *58* (1), 1-26.

10. Zarins-Tutt, J. S.; Barberi, T. T.; Gao, H.; Mearns-Spragg, A.; Zhang, L.; Newman, D. J.; Goss, R. J. M., Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep.* **2016,** *33* (1), 54-72.

11. Zakrzewski, P.; Fischbach, M. A.; Weber, T., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucl. Acids Res.* **2011,** *39*, 339-46.

12. Li, M. H. T.; Ung, P. M. U.; Zajkowski, J., Automated genome mining for natural products. *BMC bioinformatics* **2009,** *10* (185).

13. Ju, K. S.; Gao, J.; Doroghazi, J. R.; Wang, K. A.; Thibodeaux, C. J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J. K.; Lee, J.; Cioni, J. P.; Evans, B. S.; Hirota, R.; Labeda, D. P.; van der Donk, W. A.; Metcalf, W. W., Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (39), 12175-12180.

14. Donia, M. S.; Cimermancic, P.; Schulze, C. J.; Wieland Brown, L. C.; Martin, J.; Mitreva, M.; Clardy, J.; Linington, R. G.; Fischbach, M. A., A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **2014,** *158* (6), 1402-1414.

15. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucl. Acids Res.* **2015,** *43* (1), 237-43.

16. Jensen, P. R.; Chavarria, K. L.; Fenical, W.; Moore, B. S., Challenges and triumphs to genomics-based natural product discovery. *J. Ind. Microbiol. Biotechnol.* **2014,** *41* (2), 203-9.

17. Farnet, C. M.; Zazopoulos, E., Improving Drug Discovery From Microorganisms. In *Natural Products: Drug Discovery and Therapeutic Medicine*, Zhang, L.; Demain, A. L., Eds. Humana Press: Totowa, NJ, 2005; pp 95-106.

18. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016**.

19. Medema, M. H.; Fischbach, M. A., Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015,** *11* (9), 639-648.

20. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A. M. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002,** *417* (6885), 141-147.

21. Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S., Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat. Biotechnol.* **2003,** *21* (5), 526-531.

22. Scherlach, K.; Hertweck, C., Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009,** *7* (9), 1753-60.

23. Walsh, C. T.; Fischbach, M. A., Natural products version 2.0: connecting genes to molecules. *J. AM. CHEM. SOC.* **2010,** *132* (8), 2469-2493.

24. Amos, G. C. A.; Awakawa, T.; Tuttle, R. N.; Letzel, A.-C.; Kim, M. C.; Kudo, Y.; Fenical, W.; S. Moore, B.; Jensen, P. R., Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. U S A.* **2017,** *114* (52), E11121-E11130.

25. Shabuer, G.; Ishida, K.; Pidot, S. J.; Roth, M.; Dahse, H.-M. M.; Hertweck, C., Plant pathogenic anaerobic bacteria use aromatic polyketides to access aerobic territory. *Science (New York, N.Y.)* **2015,** *350* (6261), 670-674.

26. Findlay, B. L., The Chemical Ecology of Predatory Soil Bacteria. *ACS Chem. Biol* **2016,** *11* (6), 1502-1510.

27. Vizcaino, M. I.; Guo, X.; Crawford, J. M., Merging chemical ecology with bacterial genome mining for secondary metabolite discovery. *J Ind Microbiol Biotechnol.* **2014,** *41* (2), 285-99.

28. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016,** *79* (3), 629-661.

29. Juan Pablo, G.-E.; Mervyn, J. B., Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **2013,** *41* (2), 425-31.

30. Komatsu, M.; Uchiyama, T.; mura, S.; Cane, D. E.; Ikeda, H., Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (6), 2646-2651.

31. Kim, E.; Moore, B. S.; Yoon, Y. J., Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **2015,** *11* (9), 649-659.

32. Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J. J., Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **2015,** *10* (12), 2841-9.

33. Ross, A. C.; Gulland, L. E. S.; Dorrestein, P. C., Targeted capture and heterologous expression of the Pseudoalteromonas alterochromide gene cluster in Escherichia coli represents a promising natural product exploratory platform. *ACS Synth. Biol.* **2014,** *4* (4), 414-20.

34. Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W., Accessing the hidden majority of marine natural products through metagenomics. *Chembiochem.* **2011,** *12* (8), 1230-6.

35. Ongley, S. E.; Bian, X.; Neilan, B. A.; Müller, R., Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2013,** *30* (8), 1121-1138.

36. Luo, Y.; Enghiad, B.; Zhao, H., New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* **2016,** *33*, 174-182.

37. Bibb, M. J., Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **2005,** *8* (2), 208-15.

38. Seyedsayamdost, M. R.; Chandler, J. R.; Blodgett, J. A. V., Quorum-sensing-regulated bactobolin production by Burkholderia thailandensis E264. *Org. Lett.* **2010,** *12* (4), 716-719.

39. Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A., Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *Chembiochem.* **2002,** *3* (7), 619-627.

40. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015,** *13*, 509-23.

41. Ochi, K.; Hosaka, T., New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl. Microbiol. Biot.* **2013,** *97* (1), 87-98.

42. Reen, F. J.; Romano, S.; Dobson, A. D. W.; O'Gara, F., The sound of silence: Activating silent biosynthetic gene clusters in marine microorganisms. *Mar. Drugs* **2015,** *13* (8), 4754-4783.

43. Imai, Y.; Sato, S.; Tanaka, Y.; Ochi, K.; Hosaka, T., Lincomycin at subinhibitory concentrations potentiates secondary metabolite production by Streptomyces spp. *Appl. Environ. Microbiol.* **2015,** *81* (11), 3869-3879.

44. Wang, W.; Ji, J.; Li, X.; Wang, J.; Li, S.; Pan, G.; Fan, K.; Yang, K., Angucyclines as signals modulate the behaviors of Streptomyces coelicolor. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (15), 5688-5693.

45. Tokuyama, S.; Kaji, A.; Ikeda, H.; Ochi, K., Antibiotic overproduction by rpsL and rsmG mutants of various actinomycetes. *Appl. Environ. Microbiol.* **2009,** *75* (14), 4919-22.

46. Hosaka, T.; Ohnishi-Kameyama, M.; Muramatsu, H.; Murakami, K.; Tsurumi, Y.; Kodani, S.; Yoshida, M.; Fujie, A.; Ochi, K., Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **2009,** *27* (5), 462-4.

47. Wang, G.; Hosaka, T.; Ochi, K., Dramatic activation of antibiotic production in Streptomyces coelicolor by cumulative drug resistance mutations. *Appl. Environ. Microbiol.* **2008,** *74* (9), 2834-2840.

48. Ochi, K.; Okamoto, S.; Tozawa, Y.; Inaoka, T.; Hosaka, T.; Xu, J.; Kurosawa, K., Ribosome engineering and secondary metabolite production. *Adv. Appl. Microbiol.* **2003,** *56*, 155-184.

49. Wu, C.; Du, C.; Gubbens, J.; Choi, Y. H., Metabolomics-Driven Discovery of a Prenylated Isatin Antibiotic Produced by Streptomyces Species MBT28. *J. Nat. Prod.* **2015,** *78* (10), 2355-63.

50. Derewacz, D. K.; Goodwin, C. R.; McNees, R. C.; McLean, J. A.; Bachmann, B. O., Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (6), 2336-41.

51. Esquenazi, E.; Jones, A. C.; Byrum, T.; Dorrestein, P. C.; Gerwick, W. H., Temporal dynamics of natural product biosynthesis in marine cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (13), 5226-5231.

52. Kawai, K.; Wang, G.; Okamoto, S.; Ochi, K., The rare earth, scandium, causes antibiotic overproduction in Streptomyces spp. *FEMS Microbiol. Lett.* **2007,** *274* (2), 311-315.

53. Haferburg, G.; Kothe, E., Microbes and metals: interactions in the environment. *J. Basic Microbiol.* **2007,** *47* (6), 453-67.

54. Abdelmohsen, U. R.; Grkovic, T.; Balasubramanian, S.; Kamel, M. S.; Quinn, R. J.; Hentschel, U., Elicitation of secondary metabolism in actinomycetes. *Biotechnol. Adv.* **2015,** *33* (6 Pt 1), 798-811.

55. Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O., Mapping microbial response metabolomes for induced natural product discovery. *ACS Chem. Biol.* **2015,** *10* (9), 1998-2006.

56. Angell, S.; Bench, B. J.; Williams, H.; Watanabe, C. M. H., Pyocyanin isolated from a marine microbial population: synergistic production between two distinct bacterial species and mode of action. *Chem. Biol.* **2006,** *13* (12), 1349-1359.

57. Oh, D.; Kauffman, C. A.; Jensen, P. R.; Fenical, W., Induced Production of Emericellamides A and B from the Marine-Derived FungusEmericellasp. in Competing Co-culture. *J. Nat. Prod.* **2007,** *70* (4), 515-20.

58. Cueto, M.; Jensen, P. R.; Kauffman, C.; Fenical, W.; Lobkovsky, E.; Clardy, J., Pestalone, a New Antibiotic Produced by a Marine Fungus in Response to Bacterial Challenge. *J. Nat. Prod.* **2001,** *64* (11), 1444-6.

59. Brakhage, A. A., Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **2012,** *11* (1), 21-32.

60. Forseth, R. R.; Schroeder, F. C., NMR-spectroscopic analysis of mixtures: from structure to function. *Curr. Opin. Chem. Biol.* **2011,** *15* (1), 38-47.

61. Bereman, M. S.; Lyndon, M. M.; Dixon, R. B., Mass measurement accuracy comparisons between a double-focusing magnetic sector and a time-of-flight mass analyzer. *Rapid Commun. Mass Spectrom.* **2008,** *22* (10), 1563-6.

62. Stroh, J. G.; Petucci, C. J.; Brecker, S. J.; Huang, N., Automated sub-ppm mass accuracy on an ESI-TOF for use with drug discovery compound libraries. *J. Am. Soc. Mass Spectrom.* **2007,** *18* (9), 1612-6.

63. Sleno, L.; Volmer, D. A.; Marshall, A. G., Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (2), 183-198.

64. Bristow, A. W. T.; Webb, K. S., Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2003,** *14* (10), 1086-1098.

65. Kind, T.; Fiehn, O., Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2010,** *2* (1-4), 23-60.

66. Nielsen, K. F.; Månsson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O., Dereplication of microbial natural products by LC-DAD-TOFMS. *J. Nat. Prod.* **2011,** *74* (11), 2338-2348.

67. Kind, T.; Fiehn, O., Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC bioinformatics* **2007,** *8* (105).

68. May, J. C.; McLean, J. A., Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annu. Rev. Anal. Chem.* **2016,** *9*, 387-409.

69. Kildgaard, S.; Mansson, M.; Dosen, I.; Klitgaard, A., Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC-DAD-QTOFMS and a MS/HRMS library. *Mar. Drugs* **2014,** *12* (6), 3681-3705.

70. El-Elimat, T.; Figueroa, M.; Ehrmann, B. M.; Cech, N. B.; Pearce, C. J.; Oberlies, N. H., High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. *J. Nat. Prod.* **2013,** *76* (9), 1709-1716.

71. Neumann, S.; Böcker, S., Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. *Anal. Bioanal. Chem.* **2010,** *398* (7-8), 2779-88.

72. Hernandes M.Z., C. S. M., Moreira D.R., de Azevedo Junior W.F., Leite A.C., Halogen atoms in the modern medicinal chemistry: hints for the drug design. *Curr. Drug Targets* **2010,** *11* (3), 303-314.

73. Kim, J. K.; Harada, K.; Bamba, T.; Fukusaki, E., Stable Isotope Dilution-Based Accurate Comparative Quantification of Nitrogen-Containing Metabolites in Arabidopsis thaliana T87 Cells Using in Vivo 15N-Isotope Enrichment. *Biosci. Biotechnol. Biochem.* **2005,** *69* (7), 1331-40.

74. Böcker, S.; Letzel, M. C.; Lipták, Z.; Pervukhin, A., SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* **2009,** *25* (2), 218-224.

75. S. Haglund, P.; Löfstrand, K.; Siek, K.; Asplund, L., Powerful GC-TOF-MS Techniques for Screening, Identification and Quantification of Halogenated Natural Products. *Mass Spectrom.* **2013,** *2* (Spec Iss), S0018.

76. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell Proteomics* **2002,** *1* (5), 376-386.

77. Klitgaard, A.; Nielsen, J. B.; Frandsen, R. J. N.; Andersen, M. R.; Nielsen, K. F., Combining Stable Isotope Labeling and Molecular Networking for Biosynthetic Pathway Characterization. *Anal. Chem.* **2015,** *87* (13), 6520-6526.

78. Bucar, F.; Wube, A.; Schmid, M., Natural product isolation - how to get from biological material to pure compounds. *Nat. Prod. Rep.* **2013,** *30* (4), 525-545.

79. Ebada, S. S.; Edrada, R. A.; Lin, W.; Proksch, P., Methods for isolation, purification and structural elucidation of bioactive secondary metabolites from marine invertebrates. *Nat. Protoc.* **2008,** *3* (12), 1820-1831.

80. Sticher, O., Natural product isolation. *Nat. Prod. Rep.* **2008,** *25* (3), 517-554.

81. Wolfender, J. L.; Marti, G.; Thomas, A.; Bertrand, S., Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatogr. A.* **2015,** *1382*, 136-64.

82. Månsson, M.; Phipps, R. K.; Gram, L.; Munro, M. H.; Larsen, T. O.; Nielsen, K. F., Explorative solid-phase extraction (E-SPE) for accelerated microbial natural product discovery, dereplication, and purification. *J. Nat. Prod.* **2010,** *73* (6), 1126-1132.

83. Nielsen, K. F.; Larsen, T. O., The importance of mass spectrometric dereplication in fungal secondary metabolite analysis. *Front. Microbiol.* **2015,** *6* (71).

84. Hounoum, B. M.; Blasco, H.; Emond, P.; Mavel, S., Liquid chromatography–high-resolution mass spectrometry-based cell metabolomics: Experimental design, recommendations, and applications. *Trac-Trend. Anal. Chem.* **2016,** *75*, 118-128.

85. Beisken, S.; Eiden, M.; Salek, R. M., Getting the right answers: understanding metabolomics challenges. *Expert Rev. Mol. Diagn.* **2015,** *15* (1), 97-109.

86. May, J. C.; McLean, J. A., Ion mobility-mass spectrometry: time-dispersive instrumentation. *Anal. Chem.* **2015,** *87* (3), 1422-1436.

87. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014,** *6* (4), 281-294.

88. Cumeras, R.; Figueras, E.; Davis, C. E.; Baumbach, J. I., Review on ion mobility spectrometry. Part 1: current instrumentation. *Analyst* **2015,** *140* (5), 1476-90.

89. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H., Ion mobility–mass spectrometry: a new paradigm for proteomics. *Int. J. Mass Spectrom.* **2005,** *240* (3), 301-315.

90. Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S., Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *Analyst* **2004,** *129*, 984-994.

91. Kulchania, M.; Barnes, C. A. S.; Clemmer, D. E., Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. *Int. J. Mass Spectrom.* **2001,** *212* (1-3), 97-109.

92. Esquenazi, E.; Daly, M.; Bahrainwala, T., Ion mobility mass spectrometry enables the efficient detection and identification of halogenated natural products from cyanobacteria with minimal sample preparation. *Bioorg. Med. Chem.* **2011,** *19* (22), 6639-44.

93. Goodwin, C. R.; Fenn, L. S.; Derewacz, D. K., Structural mass spectrometry: rapid methods for separation and analysis of peptide natural products. *J. Nat. Prod.* **2012,** *75* (1), 48-53.

94. Jeanne Dit Fouque, K.; Afonso, C.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Lavanant, H., Ion mobility-mass spectrometry of lasso peptides: signature of a rotaxane topology. *Anal. Chem.* **2015,** *87* (2), 1166-1172.

95. Payne, A. H.; Glish, G. L., Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Methods Enzymol.* **2005,** *402*, 109-48.

96. Zhu, Z.-J. J.; Schultz, A. W.; Wang, J.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G., Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat. Protoc.* **2013,** *8* (3), 451-460.

97. Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G., XCMS 2 : Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. *Anal. Chem.* **2008,** *80* (16), 6382-6389.

98. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S., Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (41), 12580-5.

99. Hufsky, F.; Böcker, S., Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrom. Rev.* **2016,** *9999*, 1-10.

100. Vaniya, A.; Fiehn, O., Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *Trac-Trend Anal. Chem.* **2015,** *69*, 52-61.

101. Tawfike, A. F.; Viegelmann, C.; Edrada-Ebel, R., Metabolomics and dereplication strategies in natural products. *Methods Mol. Biol.* **2012,** *1055*, 227-244.

102. Kertesz, T. M.; Hall, L. H.; Hill, D. W.; Grant, D. F., CE 50: Quantifying collision induced dissociation energy for small molecule characterization and identification. *J. Am. Soc. Mass Spectrom.* **2009,** *20* (9), 1759-67.

103. Wysocki, V. H.; Joyce, K. E.; Jones, C. M., Surface-induced dissociation of small molecules, peptides, and non-covalent protein complexes. *J. Am. Soc. Mass. Spectrom.* **2008,** *19* (2), 190-208.

104. Kersten, R. D.; Yang, Y.-L. L.; Xu, Y.; Cimermancic, P.; Nam, S.-J. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C., A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **2011,** *7* (11), 794-802.

105. Kersten, R. D.; Ziemert, N.; Gonzalez, D. J.; Duggan, B. M.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (47), 4407-16.

106. Nguyen, D. D.; Wu, C.-H. 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.; Dorrestein, P. C., MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (28), 2611–2620.

107. Reeta Rani, S.; Anil Kumar, P.; Carlos, R. S.; Ashok, P., Recent advances in solid-state fermentation. *Biochem. Eng. J.* **2009,** *44* (1), 13-18.

108. Ing-Lung, S.; Chia-Yu, K.; Feng-Chia, H.; Suey-Sheng, K.; Chienyan, H., Use of surface response methodology to optimize culture conditions for iturin A production by Bacillus subtilis in solid-state fermentation. *J. Chin. Inst. Chem. Eng.* **2008,** *39* (6), 635-643.

109. Robinson, T.; Singh, D.; Nigam, P., Solid-state fermentation: a promising microbial technology for secondary metabolite production. *Appl. Microbiol. Biotechnol.* **2001,** *55* (3), 284-9.

110. Kunze, B.; Bohlendorf, B.; Reichenbach, H., Pedein A and B: Production, isolation, structure elucidation and biological properties of new antifungal cyclopeptides from Chondromyces pediculatus (…. *J. Antibiot.* **2008,** *61* (1), 18-26.

111. Abdelfattah, M. S.; Kharel, M. K.; Hitron, J. A., Moromycins A and B, isolation and structure elucidation of C-glycosylangucycline-type antibiotics from Streptomyces sp. KY002. *J. Nat. Prod.* **2008,** *71* (9), 1569-73.

112. Surup, F.; Wagner, O.; von Frieling, J., The iromycins, a new family of pyridone metabolites from Streptomyces sp. I. Structure, NOS inhibitory activity, and biosynthesis. *J. Org. Chem.* **2007,** *72* (14), 5085-90.

113. Rančić, A.; Soković, M.; Karioti, A.; Vukojević, J., Isolation and structural elucidation of two secondary metabolites from the filamentous fungus Penicillium ochrochloron with antimicrobial activity. *Environ. Toxicol. Pharmacol.* **2006,** *22* (1), 80-4.

114. Yoon, T. M.; Kim, J. W.; Kim, J. G.; Kim, W. G.; Suh, J. W., Talosins A and B: new isoflavonol glycosides with potent antifungal activity from Kitasatospora kifunensis MJM341. *J. Antibiot. (Tokyo)* **2006,** *59* (10), 640-5.

115. Umezawa, K.; Ikeda, Y.; Kawase, O., Biosynthesis of polyoxypeptin A: novel amino acid 3-hydroxy-3-methylproline derived from isoleucine. *J. Chem. Soc.* **2001,** *1* (13), 1550-1553.

116. Murphy, B.; Anderson, K.; Borissow, C.; Caffrey, P., Isolation and characterisation of amphotericin B analogues and truncated polyketide intermediates produced by genetic engineering of Streptomyces nodosus. *Org. Biomol. Chem.* **2010,** *8* (16), 3758-70.

117. Zhou, B.; Xiao, J.; Tuli, L.; Ressom, H. W., LC-MS-based metabolomics. *Mol. BioSyst.* **2012,** *8* (2), 470-481.

118. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D., Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *J. Chromatogr. A.* **2013,** *1292*, 51-65.

119. Kuehnbaum, N. L.; Britz-McKibbin, P., New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* **2013,** *113* (4), 2437-2468.

120. Straight, P. D.; Kolter, R., Interspecies chemical communication in bacterial development. *Annu. Rev. Microbiol.* **2009,** *63*, 99-118.

121. Little, A. E.; Robinson, C. J.; Peterson, S. B.; Raffa, K. F.; Handelsman, J., Rules of engagement: interspecies interactions that regulate microbial communities. *Annu. Rev. Microbiol.* **2008,** *62*, 375-401.

122. Ryan, R. P.; Dow, J. M., Diffusible signals and interspecies communication in bacteria. *Microbiology* **2008,** *154* (Pt 7), 1845-1858.

123. Watrous, J. D.; Dorrestein, P. C., Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **2011,** *9* (9), 683-694.

124. Yang, J. Y.; Phelan, V. V.; Simkovsky, R.; Watrous, J. D.; Trial, R. M.; Fleming, T. C.; Wenter, R.; Moore, B. S.; Golden, S. S.; Pogliano, K.; Dorrestein, P. C., Primer on Agar-Based Microbial Imaging Mass Spectrometry. *J. Bacteriol.* **2012,** *194* (22), 6023-6028.

125. Lane, A. L.; Nyadong, L.; Galhena, A. S., Desorption electrospray ionization mass spectrometry reveals surface-mediated antifungal chemical defense of a tropical seaweed. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (18), 7314-7319.

126. Simmons, T. L.; Coates, R. C.; Clark, B. R.; Engene, N.; Gonzalez, D.; Esquenazi, E.; Dorrestein, P. C.; Gerwick, W. H., Biosynthetic origin of natural products isolated from marine microorganism–invertebrate assemblages. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (12), 4587-4594.

127. Esquenazi, E.; Coates, C.; Simmons, L.; Gonzalez, D.; Gerwick, W. H.; Dorrestein, P. C., Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. *Mol. BioSyst.* **2008,** *4* (6), 562-70.

128. Eriksson, C.; Masaki, N.; Yao, I.; Hayasaka, T., MALDI Imaging Mass Spectrometry—A Mini Review of Methods and Recent Developments. *Mass Spectrom.* **2013,** *2*, S0022.

129. Gessel, M. M.; Norris, J. L.; Caprioli, R. M., MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery. *J. Proteomics.* **2014,** *107*, 71-82.

130. Masatoshi, T.; Joshawna, K. N.; Niclas, E.; Eduardo, E.; Tara, B.; Pieter, C. D.; William, H. G., Palmyramide A, a Cyclic Depsipeptide from a Palmyra Atoll Collection of the Marine CyanobacteriumLyngbya majuscula†. *J. Nat. Prod.* **2010,** *73* (3), 393-398.

131. Yu-Liang, Y.; Yuquan, X.; Paul, S.; Pieter, C. D., Translating metabolic exchange with imaging mass spectrometry. *Nat. Chem. Biol.* **2009,** *5* (12), 885-7.

132. Bleich, R.; Watrous, J. D.; Dorrestein, P. C.; Bowers, A. A.; Shank, E. A., Thiopeptide antibiotics stimulate biofilm formation in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (10), 3086-3091.

133. Johannes, K.; Martin, K.; Bernd, S.; Maria-Gabriele, S.; Christian, H.; Ravi Kumar, M.; Erhard, S.; Aleš, S., Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat. Chem. Biol.* **2010,** *6* (4), 261-263.

134. Gonzalez, D. J.; Haste, N. M.; Hollands, A.; Fleming, T. C.; Hamby, M.; Pogliano, K.; Nizet, V.; Dorrestein, P. C., Microbial competition between Bacillus subtilis and Staphylococcus aureus monitored by imaging mass spectrometry. *Microbiology* **2011,** *157* (9), 2485-2492.

135. Shih, C.-J. J.; Chen, P.-Y. Y.; Liaw, C.-C. C.; Lai, Y.-M. M.; Yang, Y.-L. L., Bringing microbial interactions to light using imaging mass spectrometry. *Nat. Prod. Rep.* **2014,** *31* (6), 739-755.

136. Harn, Y. C.; Powers, M. J.; Shank, E. A.; Jojic, V., Deconvolving molecular signatures of interactions between microbial colonies. *Bioinformatics* **2015,** *31* (12), 42-50.

137. Wei-Ting, L.; Roland, D. K.; Yu-Liang, Y.; Bradley, S. M.; Pieter, C. D., Imaging Mass Spectrometry and Genome Mining via Short Sequence Tagging Identified the Anti-Infective Agent Arylomycin in Streptomyces roseosporus. *J. AM. CHEM. SOC.* **2011,** *133* (45), 18010-18013.

138. Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein, P. C., Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome. *MBio.* **2013,** *4* (4), e00459-13.

139. Vaidyanathan, S.; Fletcher, J.; Goodacre, R.; Lockyer, N.; Micklefield, J.; Vickerman, J., Subsurface Biomolecular Imaging of Streptomyces coelicolor Using Secondary Ion Mass Spectrometry. *Anal. Chem.* **2008,** *80* (6), 1942-1951.

140. Vaidyanathan, S.; Fletcher, J.; Lockyer, N.; Vickerman, J., TOF-SIMS investigation of Streptomyces coelicolor, a mycelial bacterium. *Appl. Surf. Sci.* **2008,** *255* (4), 922-925.

141. Esquenazi, E.; Dorrestein, P. C.; Gerwick, W. H., Probing marine natural product defenses with DESI-imaging mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (18), 7269-7270.

142. Bhardwaj, C.; Hanley, L., Ion sources for mass spectrometric identification and imaging of molecular species. *Nat. Prod. Rep.* **2014,** *31*, 756-767.

143. Bouslimani, A.; Sanchez, L. M.; Garg, N.; Dorrestein, P. C., Mass spectrometry of natural products: current, emerging and future technologies. *Nat. Prod. Rep.* **2014,** *31* (6), 718-729.

144. Patti, G. J.; Yanes, O.; Siuzdak, G., Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (4), 263-269.

145. Orchard, S.; Hermjakob, H.; Apweiler, R., The proteomics standards initiative. *Proteomics* **2003,** *3* (7), 1374-6.

146. Hermjakob, H., The HUPO proteomics standards initiative–overcoming the fragmentation of proteomics data. *Proteomics* **2006,** *6* (2), 34-8.

147. Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M., A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* **2004,** *22* (11), 1459-66.

148. Lin, S. M.; Zhu, L.; Winter, A. Q.; Sasinowski, M., What is mzXML good for? *Expert Rev. Proteomics* **2005,** *2* (6), 839-45.

149. Martens, L.; Chambers, M.; Sturm, M.; Kessner, D., mzML—a community standard for mass spectrometry data. *Mol. Cell Proteomics* **2011,** *10* (1).

150. Kessner, D.; Chambers, M.; Burke, R.; Agus, D., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008,** *24* (21), 2534-6.

151. Hendriks, M.; Eeuwijk, F. A.; Jellema, R. H.; Westerhuis, J. A.; Reijmers, T. H.; Hoefsloot, H. C. J.; Smilde, A. K., Data-processing strategies for metabolomics studies. *Trac-Trend. Anal. Chem.* **2011,** *30* (10), 1685-1698.

152. Mahieu, N. G.; Genenbacher, J. L.; Patti, G. J., A roadmap for the XCMS family of software solutions in metabolomics. *Curr. Opin. Chem. Biol.* **2016,** *30*, 87-93.

153. Liland, K. H., Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *Trac-Trend. Anal. Chem.* **2011,** *30* (6), 827-841.

154. Willem, W.; Phalp, J. M.; Alan, W. P., A Noise and Background Reduction Method for Component Detection in Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* **1996,** *68* (20), 3602-3606.

155. Hastings, C. A.; Norton, S. M.; Roy, S., New algorithms for processing and peak detection in liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom.* **2002,** *16* (5), 462-467.

156. Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P., Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **2003,** *75* (18), 4818-4826.

157. Monroe, M. E.; Tolić, N.; Jaitly, N.; Shaw, J. L.; Adkins, J. N., VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **2007,** *23* (15), 2021-3.

158. Mueller, L. N.; Rinner, O.; Schmidt, A.; Letarte, S., SuperHirn–a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **2007,** *7* (19), 3470-80.

159. Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A., OpenMS–an open-source software framework for mass spectrometry. *BMC bioinformatics* **2008,** *9* (163).

160. Rolf, D.; Dan, B.; Karin, E. M., Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography–mass spectrometry. *Anal. Chim. Acta.* **2002,** *454* (2), 167-184.

161. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006,** *78* (3), 779-787.

162. Tautenhahn, R.; Böttcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2007,** *9* (504).

163. Du, P.; Kibbe, W. A.; Lin, S. M., Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **2006,** *22* (17), 2059-2065.

164. French, W. R.; Zimmerman, L. J.; Schilling, B.; Gibson, B. W.; Miller, C. A.; Townsend, R. R.; Sherrod, S. D.; Goodwin, C. R.; McLean, J. A.; Tabb, D. L., Wavelet-based peak detection and a new charge inference procedure for MS/MS implemented in ProteoWizard's msConvert. *J. Proteome Res.* **2014,** *14* (2), 1299-1307.

165. Aoshima, K.; Takahashi, K.; Ikawa, M.; Kimura, T.; Fukuda, M.; Tanaka, S.; Parry, H. E.; Fujita, Y.; Yoshizawa, A. C.; Utsunomiya, S.-i.; Kajihara, S.; Tanaka, K.; Oda, Y., A simple peak detection and label-free quantitation algorithm for chromatography-mass spectrometry. *BMC bioinformatics* **2014,** *15* (1), 1-14.

166. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry based metabolomics. *Mass Spectrom. Rev.* **2007,** *26* (1), 51-78.

167. Frenzel, T.; Miller, A.; Engel, K. H., A methodology for automated comparative analysis of metabolite profiling data. *Eur. Food Res. Technol.* **2003,** *216* (4), 335-342.

168. Podwojski, K.; Fritsch, A.; Chamrad, D. C.; Paul, W., Retention time alignment algorithms for LC/MS data must consider nonlinear shifts. *Bioinformatics* **2009,** *25* (6), 758-64.

169. Cleveland, W. S.; Grosse, E., Computational methods for local regression. *Stat. Comput.* **1991,** *1* (1), 47-62.

170. Perera, V.; Zabala, M. D. T.; Florance, H.; Smirnoff, N.; Grant, M., Aligning extracted LC-MS peak lists via density maximization. *Metabolomics* **2012,** *8* (1), 175-85.

171. Lange, E.; Gröpl, C.; Schulz-Trieglaff, O.; Leinenbach, A.; Huber, C.; Reinert, K., A geometric approach for the alignment of liquid chromatography-mass spectrometry data. *Bioinformatics* **2007,** *23* (13), 273-281.

172. Lommen, A., MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.* **2009,** *81* (8), 3079-3086.

173. Bork, C.; Ng, K.; Liu, Y.; Yee, A., Chromatographic peak alignment using derivative dynamic time warping. *Biotechnol Prog.* **2013,** *29* (2), 394-402.

174. Voss, B.; Hanselmann, M.; Renard, B. Y.; Lindner, M. S.; Köthe, U.; Kirchner, M.; Hamprecht, F. A., SIMA: simultaneous multiple alignment of LC/MS peak lists. *Bioinformatics* **2011,** *27* (7), 987-993.

175. Katajamaa, M.; Oresic, M., Processing methods for differential analysis of LC/MS profile data. *BMC bioinformatics.* **2004,** *6* (179).

176. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics* **2010,** *11* (395).

177. Lange, E.; Tautenhahn, R.; Neumann, S., Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. *BMC bioinformatics* **2008,** *9* (375).

178. Libiseller, G.; Dvorzak, M.; Kleb, U.; Gander, E.; Eisenberg, T.; Madeo, F.; Neumann, S.; Trausinger, G.; Sinner, F.; Pieber, T.; Magnes, C., IPO: a tool for automated optimization of XCMS parameters. *BMC bioinformatics* **2014,** *16* (118).

179. Worley B, P. R., Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2013,** *1* (1), 92-107.

180. Fonville, J. M.; Richards, S. E.; Barton, R. H., The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping. *J. Chemometr.* **2010,** *24* (11-12), 636-649.

181. H, A., Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip. Rev. Comput. Stat.* **2010,** *2* (1), 97-106.

182. Rosipal, R.; Krämer, N., Overview and recent advances in partial least squares. *Lect. Notes Comput. Sc.* **2006,** *3940*, 34-51.

183. Bro, R.; Smilde, A. K., Principal component analysis. *Anal. Methods* **2014,** *6*, 2812-2831.

184. Bylesjö, M.; Rantalainen, M.; Cloarec, O., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometr.* **2006,** *20* (8-10), 341-351.

185. Genilloud, O.; González, I.; Salazar, O.; Martín, J.; Tormo, J. R. R.; Vicente, F., Current approaches to exploit actinomycetes as a source of novel natural products. *J. Ind. Microbiol. Biotechnol.* **2011,** *38* (3), 375-389.

186. Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S., Microbial strain prioritization using metabolomics tools for the discovery of natural products. *Anal. Chem.* **2012,** *84* (10), 4277-4283.

187. Macintyre, L.; Zhang, T.; Viegelmann, C.; Martinez, I.; Cheng, C.; Dowdells, C.; Abdelmohsen, U.; Gernert, C.; Hentschel, U.; Edrada-Ebel, R., Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Mar. Drugs* **2014,** *12* (6), 3416-3448.

188. Forner, D.; Berrué, F.; Correa, H.; Duncan, K.; Kerr, R. G., Chemical dereplication of marine actinomycetes by liquid chromatography-high resolution mass spectrometry profiling and statistical analysis. *Anal. Chim. Acta.* **2013,** *805*, 70-79.

189. Norazwana, S.; Pei Jean, T.; Khozirah, S.; Faridah, A.; Hong Boon, L., Prioritization of Natural Extracts by LC–MS-PCA for the Identification of New Photosensitizers for Photodynamic Therapy. *Anal. Chem.* **2014,** *86* (3), 1324-1331.

190. Grond, S.; Papastavrou, I.; Zeeck, A., Novel α-L-Rhamnopyranosides from a Single Strain of Streptomyces by Supplement-Induced Biosynthetic Steps. *Eur. J. Org. Chem.* **2002,** *2002* (19), 3237-3242.

191. Goodwin, C. R.; Covington, B. C.; Derewacz, D. K.; McNees, C. R.; Wikswo, J. P.; McLean, J. A.; Bachmann, B. O., Structuring Microbial Metabolic Responses to Multiplexed Stimuli via Self-Organizing Metabolomics Maps. *Chem. Biol.* **2015,** *22* (5), 661-670.

192. Krug, D.; Zurek, G.; Schneider, B.; Garcia, R.; Müller, R., Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography–electrospray ionisation-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta.* **2008,** *624* (1), 97-106.

193. 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. Chem. Int. Ed. Engl.* **2012,** *51* (3), 811-6.

194. Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Muller, R., Discovering the Hidden Secondary Metabolome of Myxococcus xanthus: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **2008,** *74* (10), 3058-3068.

195. Wiklund, S.; Johansson, E.; Sjöström, L., Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **2008,** *80* (1), 115-22.

196. González-Menéndez, V.; Pérez-Bonilla, M.; Pérez-Victoria, I.; Martín, J.; Muñoz, F.; Reyes, F.; Tormo, J.; Genilloud, O., Multicomponent Analysis of the Differential Induction of Secondary Metabolite Profiles in Fungal Endophytes. *Molecules* **2016,** *21* (2), 234.

197. Hur, M.; Campbell, A. A.; Almeida-de-Macedo, M.; Li, L.; Ransom, N.; Jose, A.; Crispin, M.; Nikolau, B. J.; Wurtele, E. S., A global approach to analysis and interpretation of metabolic data for plant natural product discovery. *Nat. Prod. Rep.* **2013,** *30* (4), 565-583.

198. Albright, J. C.; Henke, M. T.; Soukup, A. A.; McClure, R. A.; Thomson, R. J.; Keller, N. P.; Kelleher, N. L., Large-Scale Metabolomics Reveals a Complex Response of Aspergillus nidulans to Epigenetic Perturbation. *ACS Chem. Biol.* **2015,** *10* (6), 1535-1541.

199. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G., XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **2012,** *84* (11), 5035-5039.

200. Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G., Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* **2014,** *86* (14), 6931-6939.

201. Xia, J.; Wishart, D. S., Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Curr. Protoc. Bioinformatics* **2011,** *Chapter14: Unit14*.

202. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucl. Acids Res.* **2015,** *43* (W1), W251-7.

203. Patti, G. J.; Tautenhahn, R.; Rinehart, D.; Cho, K.; Shriver, L.; Manchester, M.; Nikolskiy, I.; Johnson, C.; Mahieu, N.; Siuzdak, G., A View from Above: Cloud Plots to Visualize Global Metabolomic Data. *Anal. Chem.* **2013,** *85* (2), 798-804.

204. Goodwin, C. R.; Sherrod, S. D.; Marasco, C. C.; Bachmann, B. O.; Schramm-Sapyta, N.; Wikswo, J. P.; McLean, J. A., Phenotypic mapping of metabolic profiles using self-organizing maps of high-dimensional mass spectrometry data. *Anal Chem* **2014,** *86* (13), 6563-6571.

205. Gaudêncio, S. P.; Pereira, F., Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **2015,** *32*, 779-810.

206. Eugster, P. J.; Boccard, J.; Debrus, B.; Bréant, L.; Wolfender, J.-L.; Martel, S.; Carrupt, P.-A., Retention time prediction for dereplication of natural products (CxHyOz) in LC–MS metabolite profiling. *Phytochemistry* **2014,** *108*, 196-207.

207. Stow, S. M.; Goodwin, C. R.; Kliman, M., Distance Geometry Protocol to Generate Conformations of Natural Products to Structurally Interpret Ion Mobility-Mass Spectrometry Collision Cross Sections. *J. Phys. Chem. B.* **2014,** *118* (48), 13812-20.

208. Colin, A. S.; Grace, O. M.; Elizabeth, J. W.; Chuan, Q.; Sunia, A. T.; Theodore, R. B.; Darlene, E. C.; Ruben, A.; Gary, S., METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

209. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010,** *45* (7), 703-714.

210. Shen, H.; Dührkop, K.; Böcker, S.; Rousu, J., Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics* **2014,** *30* (12).

211. Bérdy, J., Thoughts and facts about antibiotics: Where we are now and where we are heading. *J. Antibiot. (Tokyo)* **2012,** *65* (8), 385-95.

212. Hufsky, F.; Scheubert, K.; Böcker, S., New kids on the block: novel informatics methods for natural product discovery. *Nat. Prod. Rep.* **2014,** *31* (6), 807-817.

213. Allard, P.-M. M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. L., Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal. Chem.* **2016,** *88* (6), 3317-3323.

214. 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.; Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (26), 1743-1752.

215. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C., Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013,** *76* (9), 1686-1699.

216. Moree, W. J.; Phelan, V. V.; Wu, C.-H.; Bandeira, N.; Cornett, D. S.; Duggan, B. M.; Dorrestein, P. C., Interkingdom metabolic transformations captured by microbial imaging mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (34), 13811-13816.

217. Briand, E.; Bormans, M.; Gugger, M.; Dorrestein, P. C.; Gerwick, W. H., Changes in secondary metabolic profiles of Microcystis aeruginosa strains in response to intraspecific interactions. *Environ Microbiol.* **2015,** *18* (2), 384-400.

218. Duncan, K. R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B. S.; Dorrestein, P. C.; Jensen, P. R., Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from Salinispora species. *Chem. Biol.* **2015,** *22* (4), 460-471.

219. Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B., Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod.* **2015,** *78* (7), 1671-1682.

220. Vizcaino, M. I.; Crawford, J. M., The colibactin warhead crosslinks DNA. *Nat. Chem.* **2015,** *7* (5), 411-7.

221. Trautman EP, C. J., Linking Biosynthetic Gene Clusters to their Metabolites via Pathway-Targeted Molecular Networking. . *Curr. Top. Med. Chem.* **2015,** *16* (15), 1705-16.

222. Vizcaino, M. I.; Engel, P.; Trautman, E.; Crawford, J. M., Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *J. AM. CHEM. SOC.* **2014,** *136* (26), 9244-9247.

223. Luo, Y.; Li, B.-Z.; Liu, D.; Zhang, L.; Chen, Y.; Jia, B.; Zeng, B.-X.; Zhao, H.; Yuan, Y.-J., Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev.* **2015,** *44* (15), 5265-5290.

224. Zhang, H.; Boghigian, B. A.; Armando, J., Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.* **2011,** *28* (1), 125-51.

225. Donia; Ruffner, D. E.; Cao, S.; Schmidt, E. W., Accessing the hidden majority of marine natural products through metagenomics. *Chembiochem.* **2011,** *12* (8), 1230-6.

226. Schorn, M.; Zettler, J.; Noel, J. P.; Dorrestein, P. C., Genetic basis for the biosynthesis of the pharmaceutically important class of epoxyketone proteasome inhibitors. *ACS Chem. Biol.* **2013,** *9* (1), 301-309.

227. Gonzalez, D. J.; Corriden, R.; Akong-Moore, K.; Olson, J., N-Terminal ArgD Peptides from the Classical Staphylococcus aureus Agr System Have Cytotoxic and Proinflammatory Activities. *Chem. Biol.* **2014,** *21* (11), 1457-62.

228. Gonzalez, D. J.; Vuong, L.; Gonzalez, I. S.; Keller, N.; McGrosso, D.; Hwang, J. H.; Hung, J.; Zinkernagel, A.; Dixon, J. E.; Dorrestein, P. C., Phenol soluble modulin (PSM) variants of community-

associated methicillin-resistant Staphylococcus aureus (MRSA) captured using mass spectrometry-based molecular networking. *Mol. Cell Proteomics* **2014,** *13* (5), 1262-1272.

229. Macherla, V. R.; Liu, J.; Sunga, M.; White, D. J., Lipoxazolidinones A, B, and C: antibacterial 4-oxazolidinones from a marine actinomycete isolated from a Guam marine sediment. *J. Nat. Prod.* **2007,** *70* (9), 1454-1457.

230. Desjardine, K.; Pereira, A.; Wright, H., Tauramamide, a lipopeptide antibiotic produced in culture by Brevibacillus laterosporus isolated from a marine habitat: structure elucidation and synthesis. *J. Nat. Prod.* **2007,** *70* (12), 1850-3.

231. McArthur, K. A.; Mitchell, S. S.; Tsueng, G., Lynamicins A− E, Chlorinated Bisindole Pyrrole Antibiotics from a Novel Marine Actinomycete†. *J. Nat. Prod.* **2008,** *71* (10), 1732-7.

232. Teasdale, M. E.; Shearer, T. L.; Engel, S., Bromophycoic acids: bioactive natural products from a Fijian red alga Callophycus sp. *J. Org. Chem.* **2012,** *77* (18), 8000-6.

233. Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A., Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. *J. Nat. Prod.* **2016,** *79* (2), 376-386.

234. Salvador-Reyes, L. A.; Luesch, H., Biological targets and mechanisms of action of natural products from marine cyanobacteria. *Nat. Prod. Rep.* **2015,** *32* (3), 478-503.

235. Johnston, C. W.; Skinnider, M. A.; Dejong, C. A.; Rees, P. N.; Chen, G. M.; Walker, C. G.; French, S.; Brown, E. D.; Bérdy, J.; Liu, D. Y.; Magarvey, N. A., Assembly and clustering of natural antibiotics guides target identification. *Nat. Chem. Biol.* **2016**.

236. Rees, M. G.; Seashore-Ludlow, B.; Cheah, J. H.; Adams, D. J.; Price, E. V.; Gill, S.; Javaid, S.; Coletti, M. E.; Jones, V. L.; Bodycombe, N. E.; Soule, C. K.; Alexander, B.; Li, A.; Montgomery, P.; Kotz, J. D.; Hon, C. S.; Munoz, B.; Liefeld, T.; Dančík, V.; Haber, D. A.; Clish, C. B.; Bittker, J. A.; Palmer, M.; Wagner, B. K.;

Clemons, P. A.; Shamji, A. F.; Schreiber, S. L., Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **2015,** *12*, 109-116.

237. Farha, M. A.; Brown, E. D., Strategies for target identification of antimicrobial natural products. *Nat. Prod. Rep.* **2016**.

238. Wong, W. R.; Oliver, A. G.; Linington, R. G., Development of antibiotic activity profile screening for the classification and discovery of natural product antibiotics. *Chem. Biol.* **2012,** *19* (11), 1483-1495.

239. Lamb, J.; Crawford, E. D.; Peck, D.; Modell, J. W.; Blat, I. C., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006,** *313* (5795), 1929-35.

240. Hutter, B.; Schaab, C.; Albrecht, S.; Borgmann, M.; Brunner, N. A.; Freiberg, C.; Ziegelbauer, K.; Rock, C. O.; Ivanov, I.; Loferer, H., Prediction of Mechanisms of Action of Antibacterial Compounds by Gene Expression Profiling. *Antimicrob. Agents Chemother.* **2004,** *48* (8), 2838-2844.

241. Hu, Y.; Potts, M. B.; Colosimo, D.; Herrera-Herrera, M. L.; Legako, A. G.; Yousufuddin, M.; White, M. A.; MacMillan, J. B., Discoipyrroles A–D: Isolation, Structure Determination, and Synthesis of Potent Migration Inhibitors from Bacillus hunanensis. *J. AM. CHEM. SOC.* **2013,** *135* (36), 13387-13392.

242. Potts, M. B.; Kim, H. S.; Fisher, K. W.; Hu, Y.; Carrasco, Y. P., Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. *Sci. Signal* **2013,** *6* (297).

243. Yu, Y.; Yi, Z.; Liang, Y. Z., Main antimicrobial components of Tinospora capillipes, and their mode of action against Staphylococcus aureus. *FEBS Lett.* **2007,** *581* (22), 4179-83.

244. Liu, Y.; Wen, J.; Wang, Y.; Li, Y.; Xu, W., Postulating modes of action of compounds with antimicrobial activities through metabolomics analysis. *Chromatographia* **2010,** *71* (3), 253-528.

245. Halouska, S.; Chacon, O.; Fenton, R. J.; Zinniel, D. K.; Barletta, R. G.; Powers, R., Use of NMR metabolomics to analyze the targets of D-cycloserine in mycobacteria: role of D-alanine racemase. *J. Proteome Res.* **2007,** *6* (12), 4608-4614.

246. Halouska, S.; Fenton, R. J.; Barletta, R. G.; Powers, R., Predicting the in vivo mechanism of action for drug leads using NMR metabolomics. *ACS Chem. Biol.* **2012,** *7* (1), 166-171.

247. Vincent, I. M.; Weidt, S.; Rivas, L.; Burgess, K.; Smith, T. K.; Ouellette, M., Untargeted metabolomic analysis of miltefosine action in Leishmania infantum reveals changes to the internal lipid metabolism. *Int. J. Parasitol. Drugs Drug Resist.* **2014,** *4* (1), 20-27.

248. Vincent, I. M.; Creek, D. J.; Burgess, K.; Woods, D. J., Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against Trypanosoma brucei. *PLoS Negl. Trop. Dis.* **2012,** *6* (5), e1618.

249. Trochine, A.; Creek, D. J.; Faral-Tello, P.; Barrett, M. P., Benznidazole biotransformation and multiple targets in Trypanosoma cruzi revealed by metabolomics. *PLoS Negl. Trop. Dis.* **2014,** *8* (5), e2844.

250. Vincent, I. M.; Ehmann, D. E.; Mills, S.; Perros, M., Untargeted metabolomics to ascertain antibiotic modes of action. *Antimicrob. Agents Ch.* **2016,** *4* (1), 20-27.

251. Lobritz, M. A.; Belenky, P.; Porter, C. B.; Gutierrez, A.; Yang, J. H.; Schwarz, E. G.; Dwyer, D. J.; Khalil, A. S.; Collins, J. J., Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (27), 8173-8180.

252. Perlman, Z. E.; Slack, M. D.; Feng, Y.; Mitchison, T. J., Multidimensional drug profiling by automated microscopy. *Science* **2004,** *306* (5699), 1194-8.

253. Woehrmann, M. H.; Bray, W. M.; Durbin, J. K.; Nisam, S. C., Large-scale cytological profiling for functional analysis of bioactive compounds. *Mol. BioSyst.* **2013,** *9* (11), 2604-17.

254. Peach, K. C.; Bray, W. M.; Winslow, D.; Linington, P. F., Mechanism of action-based classification of antibiotics using high-content bacterial image analysis. *Mol. BioSyst.* **2013,** *9* (7), 1837-48.

255. Navarro, G.; Cheng, A. T.; Peach, K. C., Image-based 384-well high-throughput screening method for the discovery of skyllamycins A to C as biofilm inhibitors and inducers of biofilm detachment in Pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **2014,** *58* (2), 1092-9.

256. Schulze, C. J.; Bray, W. M.; Woerhmann, M. H.; Stuart, J., “Function-first” lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem. Biol.* **2013,** *20* (2), 285-95.

257. Kurita, K. L.; Glassey, E.; Linington, R. G., Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (39), 11999-12004.

258. Schmidt, E. W., Natural products: Hunting microbial metabolites. *Nat. Chem.* **2015,** *7* (5), 375-376.

259. Sharon, G.; Garg, N.; Debelius, J.; Knight, R.; Dorrestein, P. C.; Mazmanian, S. K., Specialized metabolites from the microbiome in health and disease. *Cell Metab.* **2014,** *20* (5), 719-730.

260. Hooper, L. V.; Littman, D. R.; Macpherson, A. J., Interactions between the microbiota and the immune system. *Science* **2012,** *336* (6086), 1268-1273.

261. Clarke, T. B.; Davis, K. M.; Lysenko, E. S.; Zhou, A. Y.; Yu, Y., Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. *Nat. Med.* **2010,** *16* (2), 228-31.

262. Koppel, N.; Balskus, E. P., Exploring and Understanding the Biochemical Diversity of the Human Microbiota. *Cell Chem. Biol.* **2016,** *23* (1), 18-30.

Chapter 2 Prioritization of secondary metabolites in Streptomyces coelicolor A3(2) with self-organizing maps

The work presented in this chapter is adapted from a journal article written by Cody R. Goodwin, Brett C. Covington, Dagmara K. Derewacz, C. Ruth McNees, John P. Wikswo, John A. McLean, and Brian O. Bachmann, first published in *Chem. Biol.* **2015,** 22, 661-670[6-9]

Introduction

Chapter 1 introduced the concepts of using MS-based comparative metabolomics to identify NPs from complex microbial extracts by their responses to stimuli. In this chapter we discuss our applications of these comparative metabolomics methods to analyze SM inventories from extracts of stimulated cultures of genomically characterized organisms, *Streptomyces coelicolor A3(2)* (SC). SC is a model *Streptomycete* with more than 20 putative SM producing gene clusters and a significant number of regulatory genes likely used to govern responses to stimuli. With this model strain we demonstrated for the first time the utility of SOM analytics to prioritize SMs by their abundance profiles across a range of stimuli conditions including metal exposure, antibiotic resistance, and mixed culture. The observed overproduction of NPs from stimuli conditions is consistent with SMs governing adaptive organismal responses to environmental stimuli. Identifying SMs and associating them to gene clusters that are linked to discrete chemical and biological stimuli can provide insight into the chemical ecological role of SMs. Moreover, the ability to selectively stimulate native expression of secondary metabolic gene clusters via chemical or biological stimuli and detect their corresponding products without resorting to genetic recombinant methods, would greatly expedite microbial SM discovery.

The work presented in this chapter was centred on the hypothesis that both SM regulation has adapted to selectively respond to chemical and biological stimuli may be identifiable within metabolomes

by possessing characteristic abundance trends across multiplexed stimulus conditions. To investigate this hypothesis and enable SM discovery, we assessed the potential for stimulus-mediated production of SMs in the native microbe by multiplexed chemical and biological stimulation. To access a broad spectrum of responses, a battery of 23 perturbations in a single growth medium was utilized from three reported categories of activating conditions for SC. The resulting collected sum of detectable metabolomic response inventories was analysed by ultra-performance liquid chromatography-ion mobility-mass spectrometric analysis (UPLC-IM-MS). To structure and categorize the response-specificity of metabolic features within these data, we developed and implemented a SOM analysis, described in Chapter 1, for the identification and prioritization of increased metabolite production resulting from the multiplexed perturbations. SOM analysis converted the collected metabolomes into a navigable topological response phenotype map and efficiently identified specific primary and SMs that are produced at increased levels in response to stimuli. For example, in primary metabolism, we identified discrete changes in guanosine and phenylalanine pools on lanthanide exposure and evidence of unique adaptive cell wall remodelling in several conditions. Notably, a large fraction (16 total SMs) of detected SMs was prioritized via this workflow as the most intense response-specific features, providing insight into the roles secondary metabolism plays in adapting to chemical stimuli and microbial interactions. The combination of multiplexed stimulation of native expression and structuring of the resulting metabolomic responses comprise a generalizable method for activating and detecting products of natively regulated primary and secondary metabolism.

Results

Multiplexing stimuli of secondary metabolism

SC was cultivated under a battery of processes known to potentiate secondary metabolism. SC was selected as a model microorganism because it has been extensively mined for SMs,[5, 10-12] methods for native gene cluster activation have been most commonly developed for this organism, and the majority of SMs isolated from this strain have been correlated to a gene cluster.[14]

We selected three known categories of activating stimuli: eliciting spontaneous resistance to transcription or translation-targeting antibiotics, exposure to rare earth elements, and cultivation in the presence of competing microorganisms. Using a single growth medium (International Streptomyces Protocol 2, ISP2), we cultivated (1) liquid cultures in the presence and absence of five separate scandium or five lanthanum concentrations, (2) liquid cultures of ten different spontaneous rifampicin or streptomycin resistance mutants, and (3) mixed cultures with three different challenge organisms, *Micrococcus luteus* (ML), (RW), or *Tsukamurella pulmonis* (TP). These mixed cultures were prepared using a custom designed apparatus (Figure 2-1) which allowed SC and competitors to grow in a 'checkerboard' pattern on the agar plate to maximize contact between competing colonies. Hence, we generated a total of 25 conditions, including controls, spanning these three methodologies.

Total cellular extracts were generated from fermentations via methanol extraction, concentrated, and processed for reverse phase UPLC analysis. Technical triplicates of extracts were analysed in a randomized sequence using UPLC-IM-MS (Waters Synapt G2, Milford, MA) with lock mass correction to provide accurate mass measurements. During each spectral acquisition, an intact and fragmentation spectrum was taken for all ions present (herein referred to as MS^E analysis[15]). Fragmentation was performed

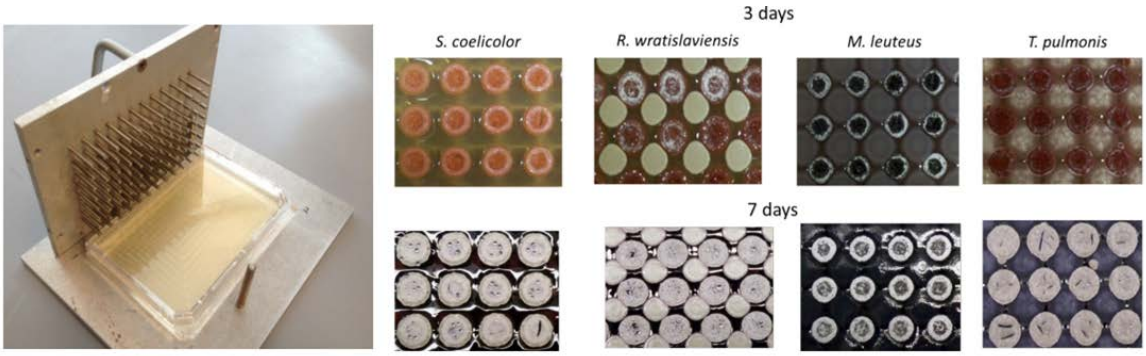


Figure 2-1: Solid agar mixed culture apparatus Left, apparatus used for co-culture on agar plates generating heat maps for co-culture panel in Figure 3. A guide rail permits the 'checkerboard' application of two organisms. Right, *Streptomyces coelicolor A3(2*), grown in monoculture on the left, was co-cultured with *Rhodococcus wratislaviensis*, *Micrococcus leuteus*, and *Tsukamurella pulmonis*. Upper and lower sets show growth and phenotype after incubation for 3 and 7 days respectively. Reproduced from *Chem. Biol.* **2015,** 22, 661-670[1] with permission from Cell Press.

subsequent to IM separation, which allowed for the correlation of product ions to precursor origins through matched mobility.

Raw data were converted to distinct *m/z* and retention time pairs termed 'features', and aligned across all samples.[4] The resultant data matrix of discrete features, or ions, and associated intensities for each condition were averaged across technical replicates and subjected to MVSA and SOM.

Identifying products of multiplexed stimulation through MVSA

The identification of new metabolites with characteristic responses from multiplexed microbial stimuli requires methods for comparing and classifying covarying ions in the response inventories. Previously, we[17] and others[12] have presented MVSA approaches for identifying the most abundant new ions resulting from individual stimulating microbial metabolic perturbations. MVSA methods for data analysis are powerful tools for identifying distinguishing features of small data sets (2 – 3 conditions) or extracting information regarding global sample grouping and which metabolites contribute to coarse trends. However, MVSA methods are not ideal for similar prioritization of metabolites in multiplexed

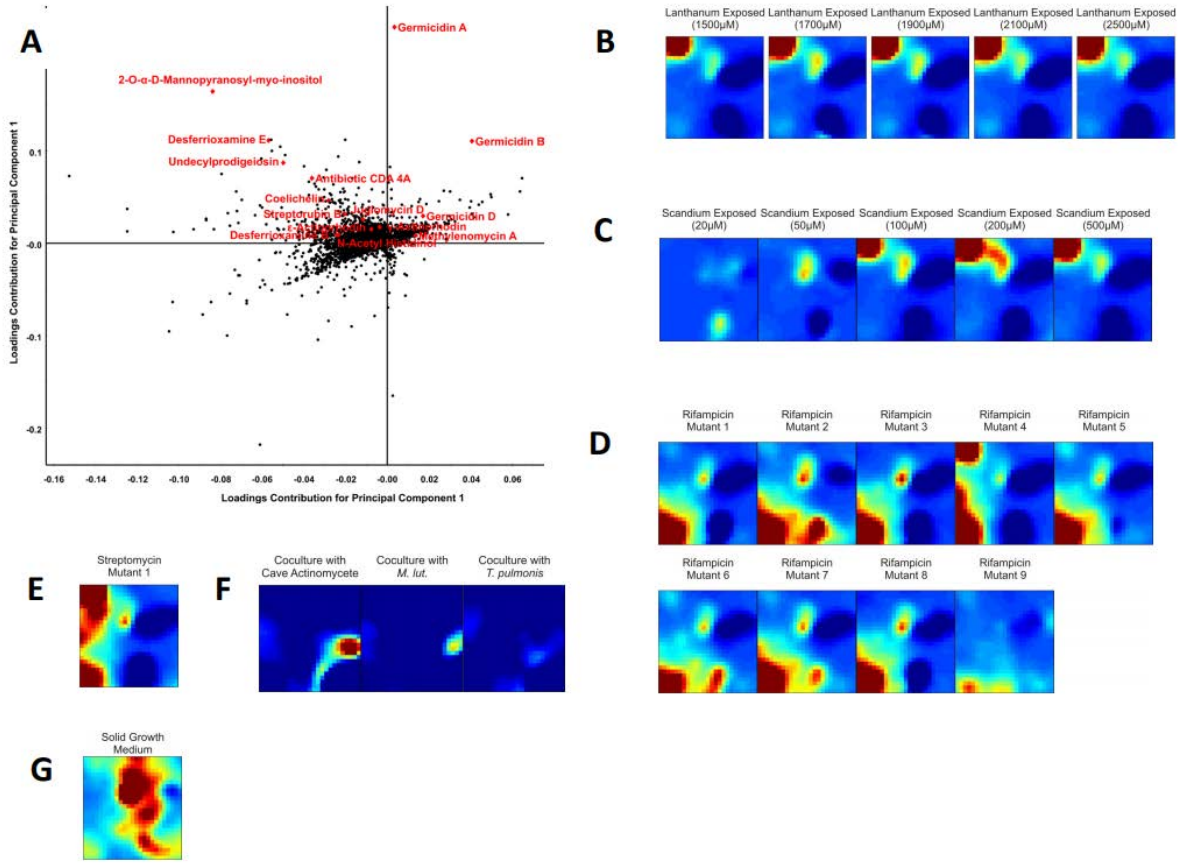


Figure 2-2: (A) Loadings plot for Figure 2-3A. The locations of identified SMs are annotate in Figure 3. Differential self-organizing map depictions of (B) lanthanum, (C) scandium, (D) rifampicin-resistant, (E) streptomycin resistant, (F) mixed fermentation, and (G) Streptomyces coelicolor A3 grown on ISP2 agar. Reproduced from *Chem. Biol.* **2015,** 22, 661-670 with permission from Cell Press.

perturbations, as MVSA is inherently biased for the largest differences amongst perturbations and is limited in the ability to reflect multiple stimuli in two- or three-dimensional space. Therefore, it omits the lower abundance or minorly covarying, yet still unique, metabolic reflexes (*i.e.,* changes in metabolism resulting from a stimulus). In applying our previously described MSVA workflow[5] to the 25 conditions, as shown in Figure 2-3A, we can visualize the gross distinctions of metabolomic profiles that exist amongst different stimuli. When each perturbation is analysed in isolation (Figure 2-3B-D), the distinct differences in global metabolism shifts are seen. A loading plot analysis can be used to determine which ions

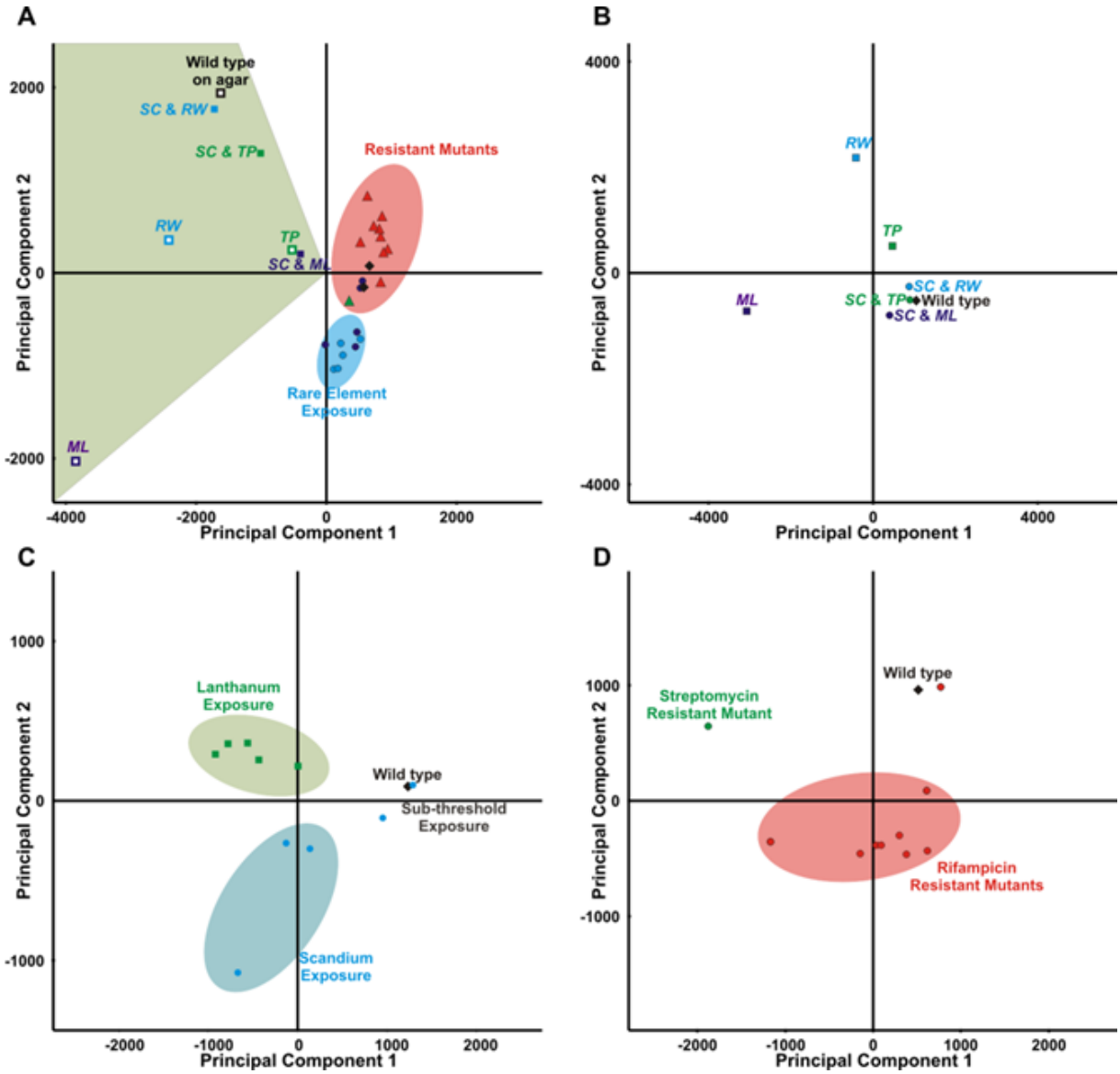


Figure 2-3: PCA of metabolomic inventories. **A**. PCA of all cultures. **B.** Co-culture of *S. coelicolor* with ML, RW, and TP. **C.** Metabolomic profiles in response to rare earth metals. **D.** Comparison of antibiotic resistant mutants selected by plating on rifampicin and streptomycin. Reproduced from *Chem. Biol.* **2015,** 22, 661-670[1] with permission from Cell Press.

contribute to sample distinction, as seen in Figure 2-2. For comparison, detected SMs are annotated, which highlights a significant shortcoming of MVSA-based prioritization for a large number of culturing conditions: the largest contributors to sample differences are highlighted, but feature selection in MVSA can be subjective and requires a cumbersome process of filtering through adducts and in-source fragments. To garner conditionally distinct differences, experimental subsets (see Figure 2-3B-D), or even

smaller subsets (*e.g.,* a single co-culture versus monocultures, or PLS approaches) can be analysed. However, for 25 conditions, pairwise analysis of multiple iterations of stimulus conditions for prioritization purposes becomes a time-intensive and subjective method of prioritizing SMs from extracts. As a result, we have developed and applied SOM-based methods to ion association and filtering.

Identifying products of multiplexed stimulation by self-organizing maps

Assessing trends in a large number of biochemical or biological conditions requires methods for the rapid visualization and organization of distinct differences in metabolic profiles across many perturbations to sort ions in a response-dependent manner. To address this, we developed a SOM-based approach to sort the complete inventory of ions across all growth conditions into regions based upon similarities in abundance profiles across experiments. This method is particularly well-suited for SM prioritization, as SMs are the end products of biochemical pathways and accumulate during fermentation. Differentially expressed, high abundance ions may be ranked subsequent to SOM analysis using the percent contribution of an ion to a region of interest (ROI) on the MEDI map. These ROIs exclusively contain features that respond specifically to a particular perturbation, while other loosely regulated metabolites will cluster outside of these prioritized regions. This method then prioritizes metabolites in a response-specific manner.

Figure 2-4 demonstrates the general workflow of the SOM-based approach, as applied to multiple perturbing conditions. Experimental and control conditions are processed by UPLC-IM-MS (**Step 1)**, significant m/z retention time features are identified, and integrated intensity trends lines are generated for each feature. SOM analysis of these feature trends is performed using the open-source Gene

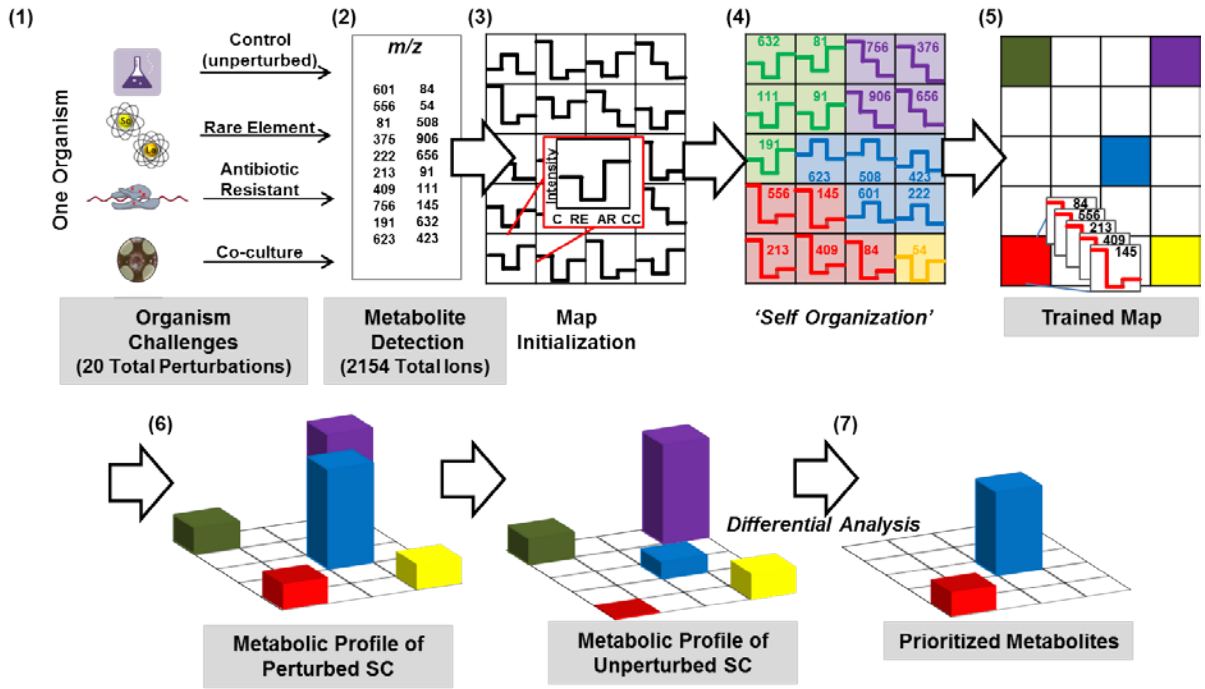


Figure 2-4: General self-organizing map (SOM)-based approach to feature prioritization. (1) This method begins with extracts from cultures of an organism cultivated under a battery of perturbing conditions. (2) Extracts are analyzed using UPLC-IM-MS^{E} (or other feature-producing methodology) and converted into a matrix of discrete, aligned peaks with associated intensities for each culturing condition. (3-5) These features are then organized based upon intensity trends across culturing conditions. (6) Subsequently, extracts are represented as heat maps based upon the sum abundance of each organized metabolite in a region. (7) Differential analysis comparing data from perturbed cultures to controls allows generation of regions of interest. Reproduced from *Chem. Biol.* **2015,** 22, 661-670[10] with permission from Cell Press.

Expression Dynamics Investigator (GEDI) software (**Step 2)** [19]. Detected ion abundance trends are first randomly seeded into a user-defined asymmetric grid (**Step 3**). The coordinates of the grid are only meaningful in relation to other grid locations and have no associated dimensions. Feature intensity trends are then iteratively organized based upon intensity similarities across experiments in a competitive-cooperative process in a process analogous to a tile puzzle (**Step 4**) [20]. As a result, metabolites that are produced as similar responses to the experimental conditions occupy the same or close coordinates in the grid. This sorts ions in a data-driven manner into regions of interconnected ions (**Step 5**).

The presumed correspondence of SM expression profiles to responses is premised on the hypothesis

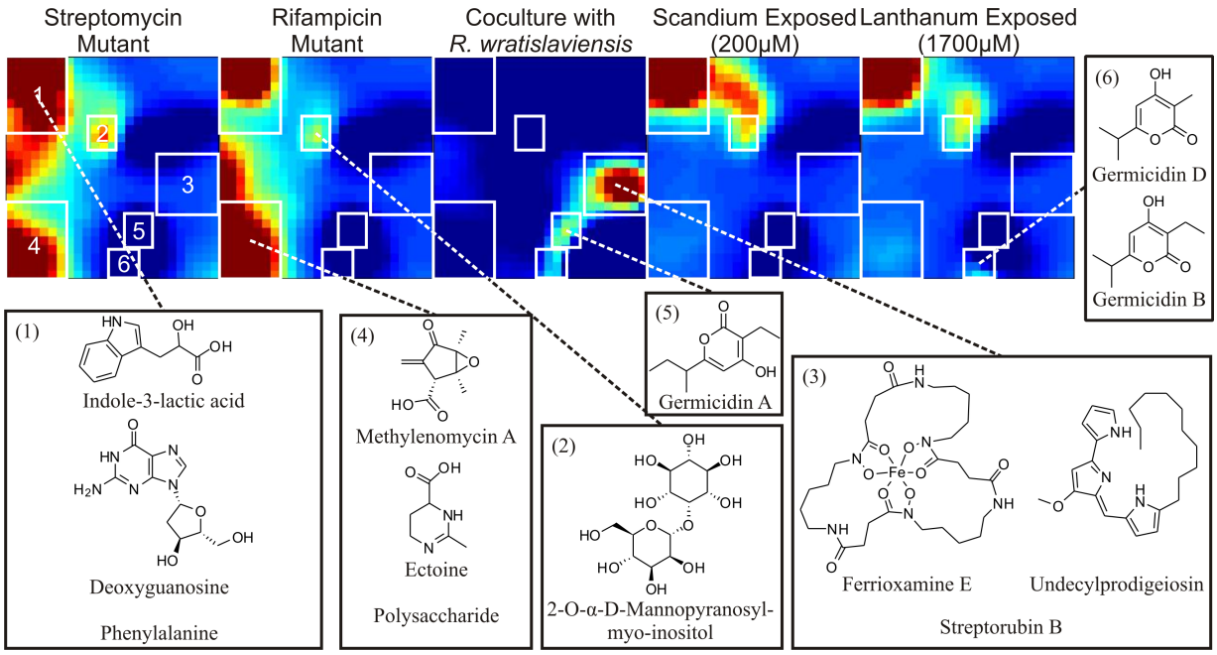


Figure 2-5: Differential metabolic phenotype heat maps representing increased production/decreased consumption of molecules using a single growth medium. Representative extracts from each culturing condition are shown above, with regions of interest boxed and labelled. Corresponding putative identifications and structures for each region of interest (ROI) are labelled, and comprehensive catalogues of inhabiting features for each region, including relative abundance and percent contribution to total ROI intensity, are presented in the ESI. Reproduced from *Chem. Biol.* **2015,** 22, 661-670[8] with permission from Cell Press.

that microorganisms use SMs to respond to discrete external stimuli (e.g., antibiotic challenge, competition, and metal exposure). The metabolic profile of each sample or experimental condition is then depicted as a topological heat map, which is a function of the intensity of each ion in that sample (**Step 6**). Ions occupying the same coordinates in the SOM are summed. These heat maps (Figure 2-5), or metabolic profiles, are then differentially compared with unperturbed metabolic profiles, resulting in heat maps with prioritized ROIs, indicating metabolic responses to experimental conditions (**Step 7**). Each pixel or node within the heat maps contains *m/z*-retention time feature lists, which are used for subsequent feature identifications. We selected six ROIs based on visual comparison of the differential phenotype heat maps, generated tables of covarying features via summing islands of high intensity within the heat maps and ranked features by percentage. The ions occupying these regions of interest are then prioritized

for further identification using accurate mass and fragmentation data acquired using MS^E technology. A given ROI may comprise only one or numerous ions. To rank within a given ROI, percent contributions of each ion to the total ROI intensity may be determined. For example, if 100 ions are grouped into an ROI, the summed relative intensity of all these ions can be used to determine the percent contribution of a given small molecule. Through self-organization, features corresponding to fragment ions, adducts, and isotopes are also all clustered for rapid triage. The determination of molecular identity of features is facilitated by the untargeted fragmentation acquisition, accurate mass measurements, retention time, and other fragmentation interpretation afforded by the ion mobility separation dimension.

Figure 2-5 demonstrates the utility of the SOM-based approach for molecular prioritization using this workflow across the multiplexed inducing conditions reported for enhanced SM production. Each heat map is representative of 2154 detected features (including detected isotopologs) observed in SC grown with a unique perturbation or condition, following subtraction of the unperturbed culture extract. In the case of monoclonal cultures (*i.e.,* streptomycin- and rifampicin-selected point mutations, rare element exposure), this baseline subtraction is simply subtraction of the ISP2 unperturbed culture metabolic profile. In co-culturing conditions, metabolic profiles from both wild type SC and competing organism monoculture were subtracted. Six dominant ROIs are indicated as boxed regions in Figure 2-5, and identified ions that occupy these regions are annotated. A majority of the annotated ions correspond to SMs SC is known to produce.[21] However, we gain additional biological insight into the microbial response to the various stimuli by observing the other biochemical results which are sorted with these SMs (*e.g.,* deoxyguanosine, phenylalanine).

Measuring and structuring metabolic perturbations

Each inducing condition provoked unique metabolic responses, as observed in the differential profiles in Figure 2-5. In total, of the 2154 significant features detected, 1318 were found to be either undetected

Table 2-1: Maximum resultant metabolite production abundances compared to matched control cultures in ISP2 medium. Color scale: 0% (red); 100% (yellow); 200% (green). Heavy metal and antibiotic resistance performed in liquid cultures and co-culture performed on agar medium. Reproduced from *Chem. Biol.* **2015,** 22, 661-670[2] with permission from Cell Press.

Metabolite	Heavy Metals	Antibiotic Resistance	Co-culture
Undecylprodigiosin	98%	470%	2200%
Germidicidin A	120%	180%	130%
Germicidin B	140%	250%	140%
Germicidin D	250%	480%	84%
Methylenomycin A	110%	170%	98%
N-Acetylhistidinol	130%	220%	18%
Juglomycin D	160%	200%	20%
2-O-α-D-Mannopyranosyl -myo-inositol	300%	200%	62%
Streptorubin B	250%	110%	1300%
Coelichelin	100%	150%	16%
γ-Actinorhodin	170%	500%	29%
ε-Actinorhodin	100%	120%	61%
Antibiotic CDA 4A	130%	330%	87%
Ferrioxamine E	50%	150%	120%
Desferrioxamine B	280%	130%	220%
Indole-3-lactic acid	4600%	3400%	84%
Ectoine	250%	290%	240%

or produced in at least two-fold abundance in at least one perturbed system relative to control. This corresponds to induced overproduction of ~61 % of all detected ions. For the subset of known SMs, 16 were observed in at least one SC expression condition (<20 ppm mass accuracy), and increased production was observed for all of these in at least one stimulated culture relative to the unperturbed control. The magnitude of amplification is shown in Table 2-1. In certain cases, a nearly 22-fold increase in production was found (*i.e.* undecylprodigiosin).

Response profile analysis

SM BGCs in microorganisms are often organized into operons, and within a given organism gene

Table 2-2: Gene clusters correlated to SM production (shown in Figure 2-5) in Streptomyces coelicolor (A3). Reproduced from *Chem. Biol.* **2015,** 22, 661-670[1] with permission from Cell Press.

Detected Metabolite(s)	Correlated Gene Cluster
Desferrioxamine	SCO2782-2785
Coelichelin	SCO0489-0499
Actinorhodin	SCO5071-5092
Prodiginine	SCO5877-5898
Calcium dependent antibiotic	SCO3210-3249
Germicidin	SCO7221
Ectoine	SCO1861-1871
Methylenomycin	SCP1

clusters share common regulatory elements programmed to respond to specific cellular states (e.g. pleiotropic signals).[18] Correspondingly, we hypothesize that SM features that are differentially produced as a result of multiplexed chemical and biological stimuli will structure into grouped regions in a SOM of metabolites based on similarity of production response profiles. In this chapter, we describe and demonstrate the application of this approach, which ameliorates the limitations of an MVSA-based analytics for multiplexed stimuli data interpretation. One practical advantage of the SOM approach is that dozens of chemical, biochemical, or genetic perturbations may be analysed using a single computation (in this case 25 x 3 analyses, comprising >780,000 spectra, in excess of 58 gigabytes of data, spanning three classes of stimuli), resulting in the generation of sets of simple and easily navigable metabolic phenotype graphical representations. Additionally, though features within an ROI can be ranked by abundance, SOM organizes features by intensity trends, so low intensity features can also be identified.

At least 22 gene clusters within the SC genome have been assigned involvement in SM production.[10] Of these 22 clusters, SOM maps prioritized metabolites associated with 8 of the 22 gene clusters, which are listed in Table 2-2, of which all display elevated SM production in some capacity as a result of the introduction of challenges (Table 2-1). The analytical strategy presented prioritizes SMs generated from these gene clusters from within the metabolomic pool, yet this begs the question as to the biological rationale of these lower abundance, yet overproduced ions. Significantly, an increase in the production of

germicidins (Figure 2-5, ROIs 5 & 6) was observed in both mixed-fermentation conditions and select antibiotic resistant strains. Metabolomic analysis indicated that the process of culturing SC on agar versus the liquid cultures affected the production of germicidins. However, germicidins A and B were present in higher concentrations in the mixed fermentation cultures versus the monoclonal cultures grown on agar, in addition to all germicidins observed in increased abundances in many of the antibiotic resistant cultures. The production of germicidins inhibits spore germination and is a self-regulatory mechanism in the production response to high population densities. Additionally, mixed fermentation resulted in the enhanced production of undecylprodigiosin (22-fold increase when co-cultured) and streptorubin B (13-fold increase when co-cultured) (Figure 2-5**,** ROI 3), known SMs of SC with antimicrobial and other clinically relevant properties.[22] This is consistent with previous studies, which linked undecylprodigiosin and streptorubin B production to external factors, including mixed fermentation with *Bacillus subtilis* (BS)[23] and salt-stress.[15] Within this same ROI, we observe the enhanced production of siderophores functioning as an iron scavenger for nutrient acquisition in all perturbed conditions.[24] This response is likely a concerted rebuttal to the microbial competition encountered in the mixed fermentation environment. A variant of calcium-dependent antibiotic production was observed to be upregulated (3.3-fold increase) specifically in agar culturing and co-culturing conditions with TP. This Gram-positive targeting metabolite may be attributed to TP production of mycolic acid, which activates SM production in once assumedly silent clusters[22, 26-27] and underpins the necessity of multi-conditional culturing. We also observed altered production of potentially exo-polysaccharides (Figure 2-5, ROI 4) as a result of these persistent resistant mutations. Furthermore, 2-*O*-α-D-mannopyranosyl-myo-inositol (3-fold increase in rare element exposure) was observed in increased abundance in mutant- and rare element-exposed cultures (Figure 2-5, ROI 2), the production of which has been demonstrated previously in liquid culture, supported by the absence in mixed fermentation conditions. Elevated production of ectoine (2.8-fold increase) was observed as a general response to perturbations and is consistent with previous results we have observed

within rifampicin- and streptomycin-resistant mutants in NF. This osmoprotectant has been shown to provide enzyme activity stabilizing effects,[28-29] and stimulate growth in osmotically inhibitory environments.

Discussion

Microbial genome sequencing has revealed a vast reservoir of SM-encoding gene clusters, suggesting largely untapped molecular diversity with potential biomedical application. Advances in sequencing have outpaced the developments of the requisite steps to produce, study, and ultimately purify the encoded metabolites: gene cluster expression and translation, identification and purification of the resulting produced metabolites, and structure elucidation. Two complementary strategies for addressing the expression component of these processes consist of refactoring targeted gene clusters for increased expression, typically in heterologous hosts,[15, 30-31] or expression of gene clusters in their native hosts using non-recombinant chemical or biochemical methods to stimulate native expression [1]. In either case, the analysis of the resulting metabolomes for upregulated or otherwise perturbed metabolites potentially becomes the next rate-limiting step. Rapid unbiased identification and prioritization of newly produced metabolites is an essential prerequisite for what remain the most labour-intensive steps of SM discovery: purification, isolation, and structure elucidation.

In this chapter we analyse three categories of microbial stimulus (antibiotic induced resistance, heavy metal exposure, and co-culture) on a single metabolomic platform. To convert the microbial metabolomic responses from 25 distinct conditions spanning these three perturbations into navigable phenotypic maps, we developed and implemented, for the first time to our knowledge, SOM analytics for multiplexed responses to microbial metabolomics. This approach localizes metabolomic features that covary across conditions into ROIs that can be used to identify metabolic features that are similarly regulated, or that respond similarly to challenge. SMs are the end products of metabolic pathways, accumulate, and are

slowly degraded. As a result, they are well suited for the application of SOM analytics that not only prioritize features, but also illuminate trends in similarly responding metabolites. Previous studies of biological, biochemical, and chemical microbial challenge are consistent with our central hypothesis that the upregulation of secondary metabolism may be an adaptive response to challenge stimuli.[1]. Moreover, the work presented in this chapter provides additional support for the broad reaching effects of chemical and biological stimulus and a new means for identification of important microbial response chemicals.

From a "genome mining" perspective, we observed substantial metabolomic expansion of the biomolecular inventory of SC grown in a single medium using multiplexed chemical and biochemical induction methods. Of the nearly 2200 total detected molecular features, 61% were found to be either undetected in control cultures or produced in at least two-fold greater amounts, relative to control, in at least one culture challenge. Indeed, using these methodologically simple and rapid non-recombinant techniques, we have observed the increased production of all 16 of the detected SMs, comprising products of up to 8 out of 22 annotated gene clusters in at least one unique culturing condition and prioritizing eight NPs within ROIs. These results challenge the notion of "silent" gene clusters in native hosts and support the potential of systematic induction of native secondary metabolism as a method of accessing the hidden reservoirs of secondary metabolic diversity in microorganisms. Indeed, with a comparatively small set of stimuli in simple culture media, which can be generated and analysed in less than a month, the majority of known secondary metabolism was activated. This work indicates much of the unknown NP potential may be accessible from the native producer through the applications of chemical or biological elicitors and comparative metabolomics analyses.

Materials and Methods

All reagents were obtained from the Sigma-Aldrich Chemical company unless otherwise specified. SC was obtained from the John Innes Center, TP from the American Type Culture Collection (ATCC700081),

and RW was obtained via dilution plating from hypogean sediments.

Eliciting antibiotic resistance and fermentations

To generate antibiotic resistant mutants, the spore inoculum of SC was uniformly spread on GYM (glucose 0.4 %, yeast extract 0.4 %, malt extract 1 %, peptone 0.1 %, sodium chloride 0.2 %, agar 2 %) agar plates containing streptomycin at one of two concentrations (100 µg/mL, 300 µg/mL) or rifampicin at either 200 µg/mL or 400 µg/mL (concentrations of antibiotics were chosen so they exceed the minimum inhibitory concentration for SC on GYM medium). After two weeks of incubation at 30 °C, the agar plates were inspected for the presence of resistant colonies, which were then aseptically transferred to antibiotic-free ISP2 (glucose 0.4%, yeast extract 0.4 %, malt extract 1 %, agar 2 %) plates. Each *S. coelicolor* mutant was then inoculated to 20 mL ISP2 liquid seed culture, incubated for 7 days, and from seed culture to 50 mL liquid ISP2 fermentation culture for 7 days of incubation at 30 °C. Progenitor SC was incubated under the same conditions to generate the control culture.

Rare earth element fermentations

For rare element additives, the spore suspension of *S. coelicolor* was inoculated on ISP2 agar plates for incubation at 30 °C for 7 days, then inoculated from plates into 20 mL liquid seed culture and from seed culture to 50 mL liquid ISP2 production cultures containing various concentrations of scandium chloride (20 µM, 50µM, 100 µM, 200 µM, 500 µM) or lanthanum chloride (1500 µM, 1700 µM, 1900 µM, 2100 µM, 2500 µM) for 7-day incubations at 30°C. To generate a control, *S. coelicolor* was incubated in 50 mL additives-free ISP2 medium under the same conditions.

Extraction of liquid fermentations

Total culture metabolite extracts from liquid cultures were generated by adding 50mL of methanol to

each fermentation flask and shaking the flasks on a rotary shaker for 1 h. Mycelia were then separated on a centrifuge, and supernatants were dried in vacuo to yield crude extracts.

Co-culture generation

Co-culture plates were prepared by addition of 40 mL of sterile ISP2 medium to a one well OmniTray plate. Cryogenic spore suspensions of *S. coelicolor* were cultivated on agar plates (100 x 15 mm) containing 30mL of ISP2 medium and incubated at 30°C until the production of spores occurred. The spores were removed from the surface of the plate using a sterile loop and suspended in 25 mL of ISP2 liquid medium at a concentration of approximately 108 spores/mL as determined via hemocytometer. This suspension was homogenized and decanted into a one well plate as a reservoir. The pins of a 96 well replicator were submerged into the spore solution and applied to the surface of the solid support within the previously prepared one well OmniTray plate without puncturing the surface (Figure 2-1). The plates were incubated for 24 hours at 30°C. Cryogenically stored M. luteus was inoculated into 5 mL of sterile ISP2 medium 8 h prior to application to the co-culture plate. *Rhodococcus wratislaviensis* stock was inoculated into 5 mL of sterile ISP2medium 24 h prior to application to the co-culture plate. Cryogenically stored *TP* stock was inoculated into 5 mL of sterile heart infusion medium 24 h prior to application to the co-culture plate. For all competing organisms, once an OD 600 of ~1 was achieved, the 5 mL sample was diluted into 30 mL of medium in separate one well plate reservoirs. The pins of a 96 well replicator were submerged into the solution and applied to the surface of the solid support within the one well OmniTray plate without puncturing the surface in an offset manner relative to the previously inoculated actinomycete. After7 days, co-culture plates were cut into 1 x 1 cm segments and extracted with equal volumes of methanol by shaking for 3 h at 170 rpm and 30 °C. The apparatus shown in Figure 2-1 was used to make solid agar mixed culture plates. A guiderail permits the 'checkerboard' application of two organisms.

UPLC-IM-MS Data Acquisition and Processing

Extract samples were resuspended in methanol at a concentration of200 mg/mL. UPLC-IM-MS^E data acquisition was performed analysis on a Synapt G2 HDMS platform (Waters Corporation, Milford, MA) with a 25 min gradient. Mobile phase A consisted of H2O with 0.1 % formic acid, and mobile phase B consisted of acetonitrile with 0.1 % formic acid. A 1x100 mm 1.7 µm particle BEH-T3 C18 column (Waters Corporation, Milford, MA) was used for chromatographic separations with a flow rate of 75 µL/min and a column temperature of 40 °C. An autosampler with a loop size of 5 µL held at 4 °C was used for sample injection. The initial solvent composition was100 % A, which was held for 1 min and ramped to 0% A over the next 11 min, held at 0 % A for 2 min, and returned to 100% A over a 0.1 min period. The gradient was held at 100 % A for the next 10.9 min for equilibration. Prior to analysis of the sample queue, ten sequential column-load injections were performed with 5 µL of the quality control. This protocol increases retention time stability and is critical to reproducible analyses. Quality control injections were then performed after every 10 sample injections to ensure instrument stability. Quality controls were comprised of pooled equal aliquots of all samples analysed.

IM-MS^E spectra were acquired at a rate of 2 Hz from 50-2000 Da in positive ion mode for the duration of each sample. The instrument was calibrated to less than 1 ppm mass accuracy using sodium formate clusters prior to analysis. A two-point internal standard of leucine enkephalin was infused in parallel to the sample at a flow rate of 7 µL/min, and data were acquired every 10 s. This was used for post-processing determination of accurate mass for prioritized ions. The source capillary was held at 110 °C and 3.0 kV, with a desolvation gas flow of 400 L/h and a temperature of 150 °C. The sampling cone was held at a setting of 35.0, with the extraction cone at a setting of 5.0. In the MS^E configuration, low and high energy spectra were acquired for each scan. High energy data provided a collision energy profile from 10-30 eV in the transfer region, providing post-mobility fragmentation. Ion mobility separations were performed with a wave velocity of 550 m/s, a wave height of 40.0 V, and a nitrogen gas flow of 90 mL/min, with the

helium cell flow rate at 180 mL/min. Internal calibrant correction was performed in real time.

Data were converted to mzXML format using the msconvert tool from the ProteoWizard package.1 Peak picking and alignment were performed using XCMS in R.2. The resulting data matrix contained 2154 detected features and was formatted for analysis using both GEDI and Umetrics. Formatting for GEDI is outlined below, formatting for Umetrics was performed by extracting and transposing the sample-feature intensity matrix generated from XCMS and importing it into Umetrics software. Prior to GEDI and MVSA, analytical triplicates were averaged. For GEDI analysis, a grid of 25 x 26 was generated. Software specific parameters include: 100 first phase training iterations with an initial training radius of 10.0, a learning factor of 0.5, a neighborhood block size of 20, and a conscience of 5.0, and 160 second phase training iterations with a neighborhood radius of 1.0, learning factor of 0.05, neighbourhood block size of 2, and conscience of 2.0. A random seed of 10 with a Pearson's correlation distance metric and random selection initialization was used.

Metabolite identifications were performed using accurate mass measurements and fragmentation spectra extracted from IM-MS^E data. Utilizing drift time correlations, product ions were correlated appropriately to precursors for extraction of high energy spectra.

Mass spectrometry data processing workflow

(1) To begin, data are converted from vendor-specific formats to open source mzXML formats. For Waters data, msconvert was used for conversion purposes through the command prompt:

> msconvert [input file] --mzXML –filter "sortByScanTime" –o [output location]

(2) The msconvert utility is included in the ProteoWizard package1, which can be download at: http://proteowizard.sourceforge.net/. To perform feature detection, xcms was used within R. The download and a fantastic tutorial are available on the METLIN website:(http://metlin.scripps.edu/xcms/)

(3) The resulting output from xcms analysis (report.tsv) was then used for further analysis. Technical replicates were averaged within any available spreadsheet software and formatted for GEDI analysis. Formatting and download can be found at: (http://apps.childrenshospital.org/clinical/research/ingber/GEDI/GEDI_Help.htm)

(4) Regions of interest were first generated through differential analysis within GEDI. These regions were then given defined boundaries. This was done manually, though boundary recognition algorithms are being investigated.

(5) Within GEDI, metabolite locations were exported through:

>File>Export Results>Save Gene Assignments List

This generated a list of the SOM locations of all metabolites, which were then compiled into lists corresponding to ROIs. Using the original data matrix, intensities for all features within an ROI were summed, and the percent contribution of each determined for prioritization purposes.

(6) Metabolite identification then began, with the only notable caveat being the extraction of fragmentation spectra using post-mobility fragmentation, which allowed untargeted fragmentation with relatively contaminant-free spectra. Retention time and drift time occurrences were also used to determine if certain features were in-source fragments. Retention times were similar, and the product ion appeared both at a drift time unique to the in-source fragment within the intact spectrum and as a product ion with the same drift time as the precursor within the fragmentation spectrum.

(7) Monoisotopic peaks were used for database searching, with candidate molecules compared to

fragmentation data and retention time occurrence.

Acknowledgments

We acknowledge support from NIH grant 1R01GM92218, the Defence Threat Reduction Agency grant HDTRA1-09-1-0013, the Vanderbilt Institute of Chemical Biology, the Vanderbilt Institute for Integrative Biosystems Research and Education, Vanderbilt University, and the Systems Biology and Bioengineering Undergraduate Research Experience (SyBBURE) funded by Gideon Searle at Vanderbilt. We thank Dr. Carol A. Rouzer for her gracious and significant editorial and intellectual contributions and guidance. We also thank Ms. Allison Price for her editorial assistance.

References

1. Goodwin, C. R.; Covington, B. C.; Derewacz, D. K.; McNees, C. R.; Wikswo, J. P.; McLean, J. A.; Bachmann, B. O., Structuring Microbial Metabolic Responses to Multiplexed Stimuli via Self-Organizing Metabolomics Maps. *Chem. Biol.* **2015,** *22* (5), 661-670.

2. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A. M. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002,** *417* (6885), 141-147.

3. Goodwin, C. R.; Sherrod, S. D.; Marasco, C. C.; Bachmann, B. O.; Schramm-Sapyta, N.; Wikswo, J. P.; McLean, J. A., Phenotypic Mapping of Metabolic Profiles Using Self-Organizing Maps of High-Dimensional Mass Spectrometry Data. *Anal Chem* **2014,** *86* (13), 6563-6571.

4. Eichler, G. S.; Huang, S.; Ingber, D. E., Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* **2003,** *19* (17), 2321-2322.

5. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3 (2). *Nature* **2002,** *417* (6885), 141-147.

6. Tanaka, Y.; Hosaka, T.; Ochi, K., Rare earth elements activate the secondary metabolite biosynthetic gene clusters in Streptomyces coelicolor A3 (2). *J antibiot (Tokyo)* **2010,** *63* (8), 477-481.

7. Xu, J.; Tozawa, Y.; Lai, C.; Hayashi, H.; Ochi, K., A rifampicin resistance mutation in the rpoB gene confers ppGpp-independent antibiotic production in Streptomyces coelicolor A3 (2). *Mol. Genet. Genomics* **2002,** *268* (2), 179-189.

8. Luti, K. J. K.; Mavituna, F., Streptomyces coelicolor increases the production of undecylprodigiosin when interacted with Bacillus subtilis. *Biotechnol. Lett.* **2011,** *33* (1), 113-118.

9. Hosaka, T.; Ohnishi-Kameyama, M.; Muramatsu, H.; Murakami, K.; Tsurumi, Y.; Kodani, S.; Yoshida, M.; Fujie, A.; Ochi, K., Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **2009,** *27* (5), 462-4.

10. Barona-Gomez, F.; Lautru, S.; Francou, F.-X.; Leblond, P.; Pernodet, J.-L.; Challis, G. L., Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in Streptomyces coelicolor A3(2) and Streptomyces ambofaciens ATCC 23877. *Microbiology* **2006,** *152* (11), 3355-3366.

11. Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L., Type III Polyketide Synthase Β-Ketoacyl-ACP Starter Unit and Ethylmalonyl-CoA Extender Unit Selectivity Discovered by Streptomyces coelicolor Genome Mining. *J. AM. CHEM. SOC.* **2006,** *128* (46), 14754-14755.

12. Challis, G. L., Exploitation of the Streptomyces coelicolor A3 (2) genome sequence for discovery of new natural products and biosynthetic pathways. *J Ind Microbiol Biot* **2013**, 1-14.

13. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K., UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Sp* **2006,** *20* (13), 1989-1994.

14. Smith, C. A.; Elizabeth, J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,** *78* (3), 779-787.

15. Derewacz, D. K.; Goodwin, C. R.; McNees, C. R.; McLean, J. A.; Bachmann, B. O., Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. *Proc Natl Acad Sci U S A* **2013,** *110* (6), 2336-2341.

16. Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C.; Edison, A. S., NMR in Metabolomics and Natural Products Research: Two Sides of the Same Coin. *Accounts Chem Res* **2012,** *45* (2), 288-297.

17. Kohonen, T.; Niklasson, L.; Bod\'en, M.; Ziemke, T., Self-Organization of Very Large Document Collections: State of the Art. In *Proceedings of ICANN98, the 8th International Conference on Artificial Neural Networks*, Springer: 1998; Vol. 1, pp 65-74.

18. Sevcikova, B.; Kormanec, J., Differential production of two antibiotics of Streptomyces coelicolor A3(2), actinorhodin and undecylprodigiosin, upon salt stress conditions. *Arch. Microbiol* **2004,** *181* (5), 384-389.

19. Aoki, Y.; Matsumoto, D.; Kawaide, H.; Natsume, M., Physiological role of germicidins in spore germination and hyphal elongation in Streptomyces coelicolor A3 (2). *J antibiot (Tokyo)* **2011,** *64* (9), 607-611.

20. Williamson, N. R.; Fineran, P. C.; Leeper, F. J.; Salmond, G. P. C., The biosynthesis and regulation of bacterial prodiginines. *Nat Rev* **2006,** *4* (12), 887-899.

21. Luti, K.; Mavituna, F., Streptomyces coelicolor increases the production of undecylprodigiosin when interacted with Bacillus subtilis. *Biotechnol. Lett.* **2011,** *33* (1), 113-118.

22. Medema, M. H.; Breitling, R.; Bovenberg, R.; Takano, E., Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev* **2011,** *9* (2), 131-137.

23. Pospisil, S.; Sedmera, P.; Halada, P.; Petricek, M., Extracellular Carbohydrate Metabolites from Streptomyces coelicolor A3(2). *J Nat Prod* **2007,** *70* (5), 768-771.

24. Lippert, K.; Galinski, E., Enzyme stabilization be ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl Microbiol Biot* **1992,** *37* (1), 61-65.

25. Jebbar, M.; Talibart, R.; Gloux, K.; Bernard, T.; Blanco, C., Osmoprotection of Escherichia coli by ectoine: uptake and accumulation characteristics. *J Bacteriol* **1992,** *174* (15), 5027-5035.

26. Wilkinson, B.; Micklefield, J., Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol.* **2007,** *3* (7), 379-386.

27. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A* **2014**, 201319584.

28. Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O., Mapping microbial response metabolomes for induced natural product discovery. *ACS chemical biology* **2015**.

29. Zerikly, M.; Challis, G. L., Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* **2009,** *10* (4), 625-633.

30. Tanaka, Y.; Komatsu, M.; Okamoto, S.; Tokuyama, S.; Kaji, A.; Ikeda, H.; Ochi, K., Antibiotic overproduction by rpsL and rsmG mutants of various actinomycetes. *Appl Environ Microbiol* **2009,** *75* (14), 4919-4922.

31. Ochi, K.; Tanaka, Y.; Tojo, S., Activating the expression of bacterial cryptic genes by rpoB mutations in RNA polymerase or by rare earth elements. *J Ind Microbiol Biot* **2014,** *41* (2), 403-414.

Chapter 3 Discovery of the ciromicins from *Nocardiopsis FU40*

This section was adapted with permission from the American Chemical Society from an article written by Dagmara K. Derewacz, Brett C. Covington, John A. McLean, and Brian O. Bachmann first published in *ACS Chem Biol*, **2015,** 10, 1998-2006.[1]

Introduction

In this chapter we continue our discussion on the utility and in some cases redundancies of SOM analytics for comparative metabolomics analyses relative to other MVSA. After establishing SOM analytics with the model strain, SC, we then applied comparative metabolomics to prioritize metabolite responses from a genetically modified NF, exposed to mixed culture. NF is known for production of apoptosis inducing macrolides, apoptolidins, and genome sequencing of this strain identified a minimum of 20 putative gene clusters related to secondary metabolism including six putative polyketide synthase encoding gene clusters.[2] Two of these, encoding type-I (reduced) polyketides, are relevant to this work. One encodes the cytotoxic macrolide apoptolidin, and the other an orphan polyene macrolactam with predicted structural relatedness to vicenistatin[3] and incednine,[4] via translated sequence similarity across the clusters. In order to obtain clean background for genome mining of new SMs from this strain, we have previously deactivated the BGC for the production of apoptolidins by replacing the terminal polyketide synthase *ApoS8* with an apramycin resistance cassette.[2] Exposure of NF *ΔApoS8* to low inoculum competing *Escherichia coli* (EC), *Bacillus subtilis* (BS), TP, and RW strains elicits significant metabolomic responses. A primary observation is that the metabolic inventory of co-culture is far greater than the sum of its monocultures. Exposure of NF to challenger strains stimulated the production of approximately 314 / 2288 (14%) detected metabolomic features not present in monocultures and revealed the production of complex photochemically reactive macrolactam polyenes, ciromicin A (**1**) and its product ciromicin B

(**2**). In addition to facilitating discovery, comprehensively analyzing the consequences of bipartite interactions provide a basis for understanding the effects of multipartite interactions present in more complex microbiomes relevant to human biology and medicine[5].

Results

Self-organizing maps to reveal consequences of mixed culture

In the previous chapter we discussed our adaption of SOM analytics for the study of microbial secondary metabolism and demonstrated how this method can prioritize features and reveal trends in metabolite expression. The product of these SOM analyses are visually interpretable heat maps in which features are localized with regards to their intensity profile across experimental conditions, and tiles are colored by the centroid integrated intensity of features they contain.[6] When applied to NP discovery, MEDI has the added benefit of allowing users to rapidly screen sample conditions and identify stimuli which elicited the most profound metabolomic responses, even when the total abundance of the response may not be the most contributing factors to differences in variance as often highlighted in PCA. In this work we use both complementary tools to answer two distinct questions. Firstly, what is the broad scale metabolic response to challenge for all components, irrespective of their abundances that is identifiable using MEDI. Secondly, we use MVSA strategies (PCA) to rapidly identify the most abundant response changes for isolation. These two approaches were found to have some degree of redundancy and cross validate when MEDI results are ranked by intensity.

The results of the entire UPLC/IM-MS dataset of mono- and co-cultures, comprising 22.5 gigabytes of raw data in total, were converted into features via XCMS and processed in a single SOM analysis, and are

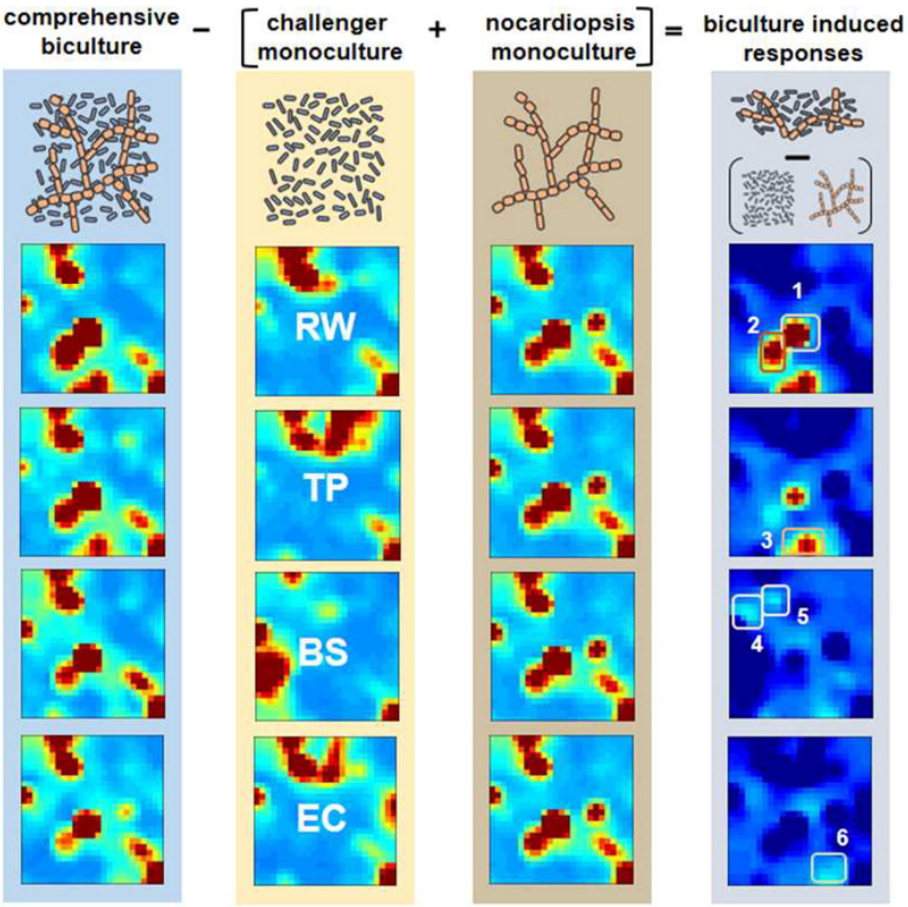


Figure 3-1 Metabolomic analysis of mono- and co-cultures. Self-organizing maps of features (m/z, retention time pairs) were generated from averaged UPLC/IM-MS chromatograms from four co-cultures and five monocultures and analyzed together using the MEDI algorithm. From left to right: column highlighted in blue contains metabolic heat maps from co-cultures of *Nocardiopsis* with a competitor; monocultures of competing organisms RW, TP, BS, and EC are highlighted in yellow; heat maps from NF monocultures are highlighted in brown. Difference maps show co-cultures after the subtraction of respective monocultures to identify unique and upregulated features. Hot spots are designated as regions of interest 1-6. Reproduced from *ACS Chem Biol*, **2015,** 10, 1998-2006[11-17] with permissions from the American Chemistry Society.

displayed in Figure 3-1. Visual inspection of heat maps reveals the differences and similarities in organized metabolomic phenotypes within microbial genera. Of interest, mixed culture plots reveal that the mixed metabolomic phenotype is dominated by the NF culture, as expected due to low inoculum concentration of challenger organism. However, examination of the difference maps, which are generated by subtracting monoculture feature maps from co-cultures, reveals that mixed fermentations with TP and RW lead to significant metabolic activation, while competition with BS and EC elicits a lower new feature response

from NF. Apparent 'hot spots' within difference maps were designated as ROIs and contained several features upregulated in co-cultures relative to monocultures. Ranking features within these ROI by intensity prioritized many new and abundant features with high agreement to prioritization via PCA (*vide supra*).

Mixed culture induces large metabolomic changes

To assess the magnitude of changes in extracted microbial metabolomes, UPLC/IM-MS data were processed via XCMS and Meta-XCMS to align chromatograms and identify feature commonalities and differences across samples. This comparative analysis revealed significant changes between the metabolomes of monoculture of NF and low inoculum challenger co-cultures. A total of 469 features out of 2288 detected (ca. 20%) were upregulated 2-fold or greater in at least one mixed fermentation in comparison to monocultures, and a great majority of these co-culture specific features (nearly 14% of the entire dataset) were not detectable in significant quantities in the monocultures, representing new metabolic features elicited by bipartite interactions. Venn diagrams of tripartite comparisons illustrate that for each co-culture, a large fraction of features appears in higher concentrations (Figure 3-2 a-d). While some features are generally upregulated across mixed culture there are a large number of features unique to a singular co-culture interaction (Figure 3-2 e). For instance, of the 469 newly produced or upregulated features 52% were specific to a single co-culture condition, while 21% were shared by 3 or more of the co-cultures. Taken together, the demonstrated generic selectivity of microbial responses suggests that features accumulating in response to co-culture are not due solely to factors such as nutrient availability, but also to biochemical cues arising from discrete intergeneric interactions.

Response features identified by comparative metabolomics.

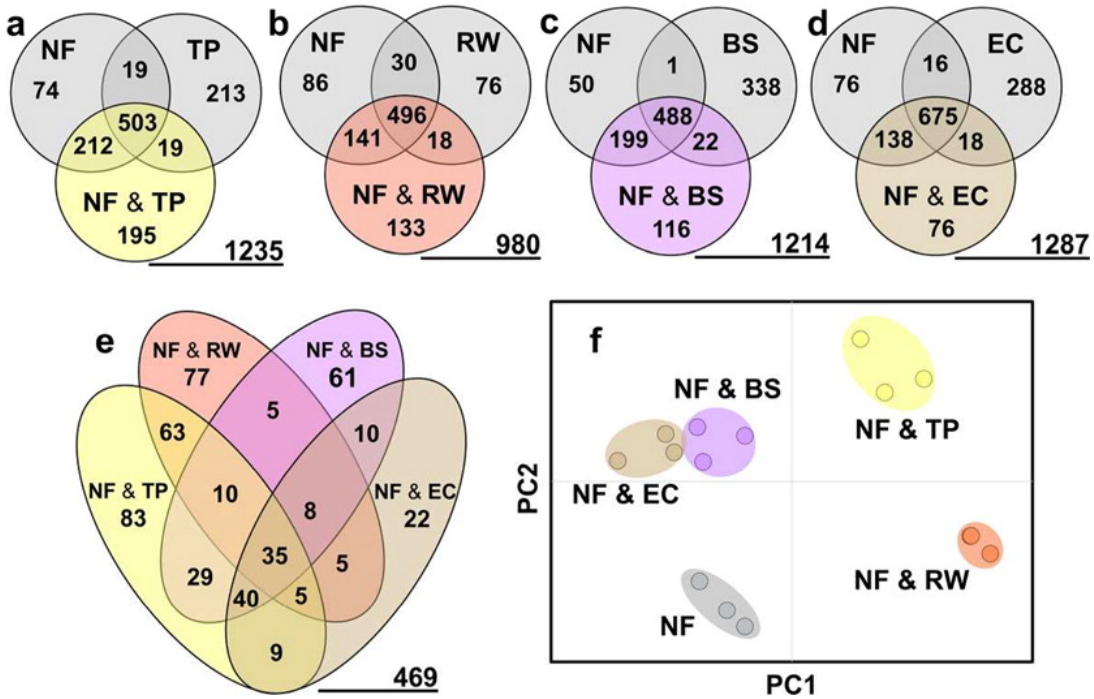


Figure 3-2 Distribution of metabolomics features between cocultures. (a–d) Venn diagrams depicting the distribution of features separated by 2-fold intensity differences with the underlined number representing the total number of detected features from the multipartite analysis. (e) Composite Venn diagram showing the distributions between different cocultures of the 469 total 2-fold or greater upregulated features. (f) PCA plot shows separation of cocultures relative to monoculture *Nocardiopsis* (NF) along PC1 and PC2, which reflect 68% of data variance (51% for PC 1 and 17% for PC 2). Reproduced from *ACS Chem Biol*, **2015,** 10, 1998-2006[7] with permissions from the American Chemistry Society.

PCA, described in Chapter 1 , functions to identify a subset of linear combinations of variables which summarize the entire dataset by identifying principal components that capture the majority of data variance. A PCA comparison of all features (Figure 3-2 f) shows that each generic condition elicits a unique set of extracted metabolomes clearly separable via unsupervised distribution along the first and second principal components. The analysis of co-culture data via PCA (Figure 3-3 a) demonstrates qualitatively how the extracted metabolomes of stationary phase co-cultures differ from their constituent organisms in monoculture. Mixed culture with RW demonstrated the largest degree of differences via MEDI analysis and we next evaluated these differences via exploring features contained within ROI (Figure 3-3 b).

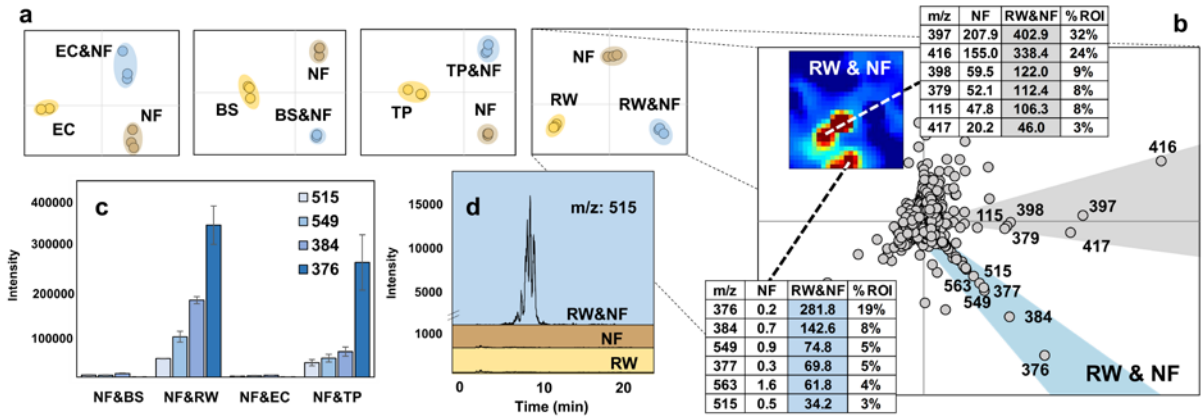


Figure 3-3 Comparison of PCA and MEDI analyses for feature prioritization. **(a)** PCA plots reveal group separation of co-cultures shown in blue from their respective monocultures with EC, BS, TP, and RW shown in yellow, and NF shown in brown. The first and second principal components are aligned along the x-axis and y-axis respectively. (b) Comparison of principal component loadings plot vectors and their corresponding two regions of interest from MEDI analysis show similar results. Features aligned on the blue vector of the loadings plot are uniquely produced in mixed culture, while the grey vector holds features upregulated in the co-culture, but also present in NF monoculture. The distance from the origin reflects relative ion abundance as can be seen in the bar graphs in figure (c), which shows averaged integrated intensities of multiple co-culture specific features across all co-culture conditions to indicate competitor selectivity for metabolic activation in NF. (d) Extracted ion chromatograms of m/z: 515 from RW and *Nocardiopsis* co and monocultures colored in the manner of the PCA plots. Reproduced from *ACS Chem Biol*, **2015,** 10, 1998-2006 with permissions from the American Chemistry Society.

Ranking features within ROI by intensity reveals metabolites that are most detectably increasing as a result of co-culture. These rankings were cross-validated by principal component loadings analysis, and confirmed via inspection of extracted ion chromatograms of those abundant features which were predicted to be most unique to the co-cultures (Figure 3-3 c-d)

A polyene prioritized for structure elucidation

Considering the strong genomic evidence for a polyene macrolactam type polyketide, intensity ranked lead features prioritized in ROI were searched for expected extended chromophores. One of the features identified, a new apparent polyene (with a strong λ_{max} of 290 nm) with an accurate mass of 515.275 Da) was upregulated in all mixed cultures and most highly upregulated in co-cultures with RW and TP. Also noted was an isobaric species with a different retention time but lacking the characteristic polyene

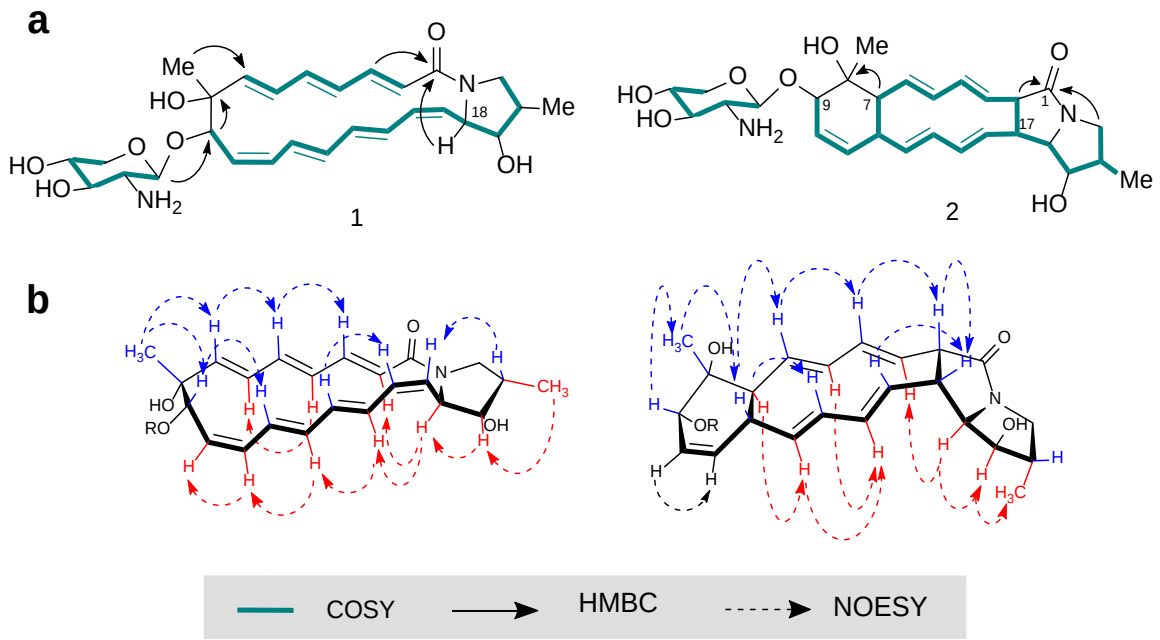


Figure 3-4 Structures of ciromicins A (1) and B (2). (a) Stuctures were generated from analysis of COSY, HSQC and HMBC spectral data. Ciromicin A (**1**) undergoes cyclization when exposed to ambient light yielding ciromicin B (**2**). (b) Major NOESY correlations and relative stereochemistry of ciromicins. Reproduced from *ACS Chem Biol*, **2015,** 10, 1998-2006[1] with permissions from the American Chemistry Society.

structure. The NF & RW co-culture fermentation was scaled 20-fold (20 x 50 mL), and combined extracts were pre-fractionated by size exclusion chromatography in methanol followed by isolation using a water/acetonitrile gradient on C18 HPLC. Compound **1**, which we named ciromicin A (after the Latin irregular verb for war), was isolated as a pale-yellow solid with UV spectrum showing maxima at 207 and 290 nm. High resolution MS identified the *m/z* of 515.275 [M+H] and indicated, along with NMR spectral data, a molecular formula of $C_{28}H_{38}N_2O_7$. Three partial structures of compound **1** were derived from COSY, TOCSY, HSQC, and HMBC spectral data (Appendix A) with two poly-unsaturated chains with three and four double bonds respectively and the residual spin system an unusual five carbon amino sugar. HMBC correlations from proton H-9 to quaternary carbon C-8 and from methylene protons H-21 to carbon C-1 merged the linear polyene substructures to form a cyclic polyene (Figure 3-4a, compound 1). Interestingly, the HMBC correlation from proton H-18 to carbon C-1, which was observed only in elevated temperature

(50 °C), confirmed the presence of an unusual pyrrolidinol moiety which we previously proposed on the basis of COSY and HSQC data. Finally, the HMBC correlation from proton H-1' to carbon C-9 linked the amino sugar to aglycone yielding the complete structure of ciromicin A (**1**) (Figure 3-4a). Ciromicin B (**2**), an isobar of ciromicin A (**1**) with a different retention time was originally observed in crude extract along with compound **1** however, due to a low level of abundance we were unable to obtain 2D NMR spectral data. Surprisingly, we later observed that pure ciromicin A (**1**) in methanol solution under ambient conditions, converted to a chromatographically distinct compound, which overlapped in retention time, UV and mass spectral data with ciromicin B (**2**). Isolating environmental variables, we observed that this isomerization was entirely dependent upon exposure to ambient visible light, and when ciromicin A (**1**) was exposed to ambient sunlight in a borosilicate glass tube for 2 hours, we observed almost full conversion to ciromicin B (**2**). Subsequently, we determined the wavelength dependence of the conversion of ciromicin A (**1**) via monochromatic UV/VIS exposure of varying wavelengths and analyzing for conversion via LC-MS. This further confirmed photochemical dependency and demonstrated wavelength dependency on product formation. Visible light (400 nm) yielded ciromicin B (**2**) as the major product. Maximum conversion to ciromicin B (**2**) was observed at 300 nm but lead to the formation of additional isomers. Isolation, structure elucidation and analysis of these ciromicins will be described in a subsequent study. Ciromicin B (**2**) was isolated as pale-yellow solid with UV spectrum showing maxima at 209 and 230 nm. The high-resolution MS yielded an m/z of 515.275 [M+H], identical to m/z of compound **1**, indicating the same molecular formula of $C_{28}H_{38}N_2O_7$. By comparing the HSQC spectral data of compound **2** to compound **1**, we noticed the presence of four additional methines and the absence of two double bonds. On the basis of COSY, TOCSY and HMBC correlations, one big encompassing spin system was assembled and linked to the quaternary carbons C-8 and C-1 yielding a structure of ciromicin B (**2**) which possess an unusual pyrrolizidinone moiety (Figure 3-4a, compound 2).

The relative stereochemistry of ciromicins was determined on the basis of proton coupling constants combined with careful evaluation of 2D NOESY NMR (Figure 3-4b). Stereochemical analysis was facilitated by the conformationally restrained tetracycle of ciromicin B (**2**). The geometry of double bonds was elucidated by analysis of proton coupling constants and further confirmed by NOESY experiment. The large couplings of $J_{3,4}$ = 15 Hz, $J_{5,6}$ = 15 Hz, $J_{13,14}$ = 15.7 Hz and $J_{15,16}$ = 16.2 Hz reveal a likely 3*E*, 5*E*, 13*E* and 15*E* geometry. The remaining double bond showed overlap in proton resonances but the presence of six-membered ring system strongly suggests a *Z* geometry, which was further confirmed by the NOE correlation between protons H-10 and H-11. The relative stereochemistry of the 9 chiral centers of aglycone was established on the basis of NOESY correlations. Starting with cyclohexenyl ring, we observed strong NOE from the 23-CH_3 to H-6, H-9 and H-12 but not H-7 and NOE from H-7 to H-5 and H-13. This suggested a *trans* relative orientation of H-7 and H-12. For the hexahydro-3H-pyrrolizin-3-one moiety, we observed NOE correlations from H-2 to H-4 and H-17, and from H-17 to H-16 indicating that they were on the same face of the cyclic system with H-2 and H-17 in *cis* relative orientation (Figure 3-4b). NOE correlations from the methine proton H-18 to H-3, H-15, H-19 and 22-methyl placed them all on the opposite face of the cyclic system. We then attempted to relate the stereocenters in one end of the molecule to the stereocenters in the other end. Starting with proton H-2 and knowing that protons H-2 and H-4 were on the same face of the cyclic system and that the geometry of the double bonds was determined to be 3*E*, 5*E*, we deduced that protons H-2 and H-7 were on the opposite face of the cyclic system. The same thing was done to relate the stereocenters C-12 and C-17. Starting with proton H-12 and knowing that protons H-12 and H-14 were on the same face of the cyclic system and double bonds showed 13*E* and 15*E* geometry, we determined that protons H-12 and H-17 were on the same face of the cyclic system. Taken together these data suggest the relative stereochemistry of ciromicin B (**2**) to be 2*S*, 7*S*, 8*R*, 9*S*, 12*R*, 17*R*, 18*R*, 19*R*, 20*S*. Since ciromicin B (**2**) is the product of chemical conversion of ciromicin A (**1**), the stereochemistries of five chiral methine centers of ciromicin A (**1**) were assumed to be the same:

8*R*, 9*S*, 18*R*, 19*R*, 20*S*. This was further confirmed by NOESY correlations, in which we observed NOE interactions from the 23-CH_3 to the H-9 methine and H-7 double bond, placing them on the same face of the macrocyclic system. Furthermore, NOE from H-18 to H-19 and 22-methyl and from H-20 to H-17 supported previously proposed stereochemistry in these positions. The geometry of the double bonds for ciromicin A (**1**) was determined by the analysis of proton coupling constants and 2D NOESY NMR. The large couplings of $J_{2,3}$ = 15 Hz, $J_{4,5}$ = 15 Hz, $J_{6,7}$ = 15 Hz, $J_{12,13}$ = 15 Hz and $J_{14,15}$ = 15 Hz revealed 2*E*, 4*E*, 6*E*, 12*E*, and 14*E* geometry for corresponding double bonds. For the remaining double bonds, due to partial overlap of proton resonances, we were unable to establish definite geometry, therefore we performed an additional NMR experiment in CD_3OD, where the desired proton resonances fully separated. The coupling constants of $J_{10,11}$ = 10 Hz and $J_{16,17}$ = 10.5 Hz indicated 10*Z* and 16*Z* geometry in these positions. The stereochemistry of the amino sugar present in both ciromicins was deduced from the couplings extracted from the proton NMR of compound **2** due to less overlap in proton resonances. Large coupling constants of $J_{1',2'}$ = 8.5 Hz, $J_{2',3'}$ = 9 Hz, $J_{3',4'}$ = 9 Hz indicated that the connected carbon, nitrogen and oxygen were all likely equatorial. Moreover, NOE from H-1' to H-9 of the aglycon suggested 1'*R* relative stereochemistry. Due to unsuccessful attempts to generate crystals of ciromicins for the x-ray analysis, the absolute stereochemistry of C-9 was putatively proposed by inference from analysis of the polyketide synthase (*vide infra*).

Ciromicin biosynthesis

Bioinformatics has been widely used by the scientific community to identify SM gene clusters and connect them to their corresponding compounds.[7-8] We applied these methods[9] to the whole genome sequence of NF and proposed the gene cluster predicted to encode ciromicin. The putative gene cluster encoding ciromicin biosynthesis was refined by closing sequencing gaps via PCR amplification (Figure 3-5). The analysis of putative open reading frames reveals a cluster similar to biosynthetic relative vicenistatin

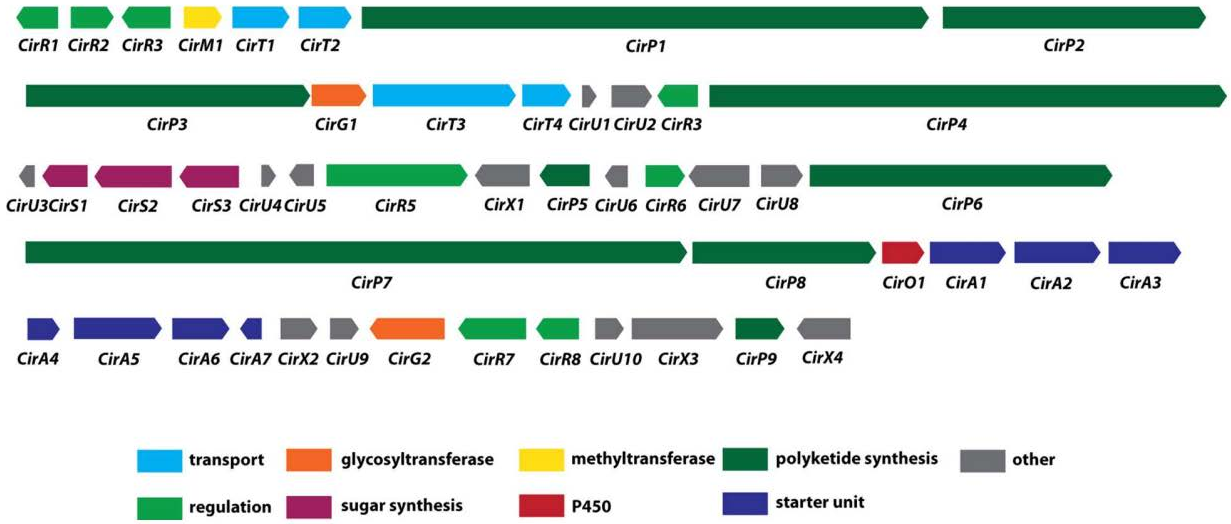


Figure 3-5 Genetic organization of ciromicin gene cluster in NF Reproduced from *ACS Chem Biol*, **2015,** 10, 1998-2006[1] with permissions from the American Chemistry Society.

and wholly consistent with the proposed biosynthesis of ciromicin. A complete set of genes for the biosynthesis of starter unit 3-amino-2-methylpropionate activated onto peptidyl carrier protein (CirA7) and coupling to the characteristic L-alanine, which is unique to polyene macrolactams and serves as a biosynthetic 'protective group' that is cleaved prior to macrocyclization.[10] CirX2 possesses 62% identity with VinJ, the analogous peptidase in vicenistatin. The translated polyketide sequence differs substantially and predicts nine active homologations, predominantly resulting in polyene extension, with only module 6 (CirP2) lacking a dehydratase domain and terminating catalytically with a ketoreductase (KR6), and thereby encoding the hydroxyl group at C-9, with a predicted 'B-type' (*S*, in this case) stereochemistry based on analysis of conserved amino acids in the ketoreductase catalytic site.[11-12] Of the acyltranferase domains, only AT6 is predicted to activate methylmalonate, and is responsible for methyl substitution at C-8. An additional module directly upstream of the thioesterase is predicted to be nonfunctional as it lacks conserved active site residues in both AT and DH domains and is missing a KR domain. The resulting heptaene polyketide sequence predicts a structure bearing comparison to incednine[13] and vicenistatin,[14] and is entirely consistent with the 22-membered polyene macrolactam scaffold observed herein. Coupling

of the macrolactam amide to C-18 may occur via addition to an epoxide precursor, which is speculative, but partially supported by the polyketide domain sequence, which predicts a double bond at C-18/19 in the biosynthetic precursor. An additional oxidation is required at C-8, likely mediated by a cytochrome P450 (CirO1). Finally, a cassette of genes, CirS1/2/3, with high sequence similarity to IdnS1/2/3, encodes the biosynthesis of the appending sugar UDP-xylosamine from UDP-*N*-acetyl-D-glucosamine. CirG2, with 35% identity to VinC from vicenistatin, encodes the likely glycosyltranferase.

Ciromicins A (**1**) and B (**2**) were tested for their *in vitro* cytotoxicity and showed activity against MV-4-11 human leukemia cell line with IC_{50} of 8.1 μM for compound **1** and 9.3 μM for compound **2**. No antibacterial or antifungal activity was detected when tested against *Bacillus*, *E. coli* or *Saccharomyces*. To the best of our knowledge, the pyrrolidinol substructure found in ciromicin A (**1**) has not been previously reported in the peer reviewed literature. However, a similar truncated tetracyclic cyclohexene/hexahydro-3H-pyrrolizin-3-one scaffold in ciromicin B (**2**) is found in one other reported SM, heronamide A, which is co-purified with a putative biosynthetic progenitor heronamide C,[15] a 20-membered polyene macrolactam likely biosynthetically divergent from incednine and vicenistain.[16] Raju and coworkers propose a biosynthetic relationship between heronamide C and A via a conrotatory 4 π + 6 π electrocyclic rearrangement. The observation of wavelength-dependent chemical conversion of ciromicin A (**1**) to B (**2**), suggests that the mechanism of rearrangement of heronamides may be purely photochemical via a previously unreported labile pyrrolidinol-functional biosynthetic intermediate. Photochemical rearrangements of up to 8- π electrons in NPs are rare but not unknown.[17] Notable examples include the 8 π [4 + 4] conversion of alteramide A[18] and the 8 π – 6 π electrocyclization cascade implicated in enandrianic acid biosynthesis.[19-20] However, while the overall mechanism of photochemical conversion of ciromicin A (**1**) remains to be determined, the intriguing diastereoselective visible light-triggered 12-π rearrangement of ciromicin A (**1**) to B (**2**) appears to be unprecedented at this time.

Discussion

Extrapolation of the costs of *de novo* genome sequencing suggests that sequence data for all potential SM producers will become available soon,[21-22] and genome mining efforts will be fully potentiated from a genomic supply data perspective. However, generalizable tools to convert gene clusters of SMs of interest into molecules remain underdeveloped. Exciting progress is being made in expressing cryptic secondary metabolic gene clusters in heterologous hosts and/or endogenous refactoring of native regulatory elements.[23-25] These powerful techniques use recombinant genetic manipulation, cloning, and/or gene synthesis to facilitate NP discovery and will continue to be important methods to unlock repressed gene clusters of interest as technology continues to advance. In this chapter we demonstrate that utilization of native expression mechanisms and enabling them with tools in comparative metabolomics such as SOM response analytics has the potential to release a large fraction of native biosynthesis and fuel new molecular diversity generation for drug discovery efforts. Notably, while this study used genomic prescience of a polyene macrolactam to guide prioritization of feature response data for isolation, a wide variety of analytical techniques can be joined with stimulus response mapping to improve prioritization including molecular networking,[26-27] which is used to identify expected and precedented structural subclasses, in addition to other tandem mass spectrometric methods such as peptidogenomics,[28-29] which identifies predicted peptidic SMs via molecular networking and MS/MS analysis.

The complex problem of decoupling the chemical basis for microbial interactions within ecological contexts is now of increasing interest, and progress in this area is dependent upon the development of innovative new analytical techniques.[30-31] Herein, the analysis of a panel of biological stimuli arising from defined coculture was facilitated by SOM analysis, which efficiently identified unique patterns in microbial metabolomics responses engendered from specific intergeneric interactions (Figure 3-1). The identification of RW as a productive coculture strain was cross-validated via comparative metabolomics (Figure 3-2). Isolation and elucidation campaigns are labor-intensive and narrowing to RW resulted in

identification of a productive system from which several new and upregulated response products were isolated. The output of SOM analysis is distinct from MVSA and molecular networking in that difference maps capture all responding peaks, irrespective of intensity, and organize them not by molecular similarity but by intensity profiles across multiple stimuli and provide a quick assessment and organization of metabolites resulting from stimulus. Whereas PCA analysis of each condition is difficult to distinguish (Figure 3-3a), MEDI shows discernible phenotypes. However, as demonstrated here, when SOM prioritized peaks within ROI are further ranked by intensity (Figure 3-3b), they recapitulate PCA loadings analysis, and it is expected that MS/MS generated fragmentation data should cluster with parent ions into nodes if they are included in the SOM matrix. Notable in this study, the challenger organisms were introduced at low inoculum levels after the NF was well established and were present at undetectable levels after 6 days, highlighting how minor constituents of a microbial ecosystem can have large effects on its observable chemical ecology and how those changes can be mapped and explored by SOM analytics.

There are several categories of potential intergeneric interactions that ultimately arise from the interaction of colocalized microorganisms. Chemical interactions may result from the diffusion of small molecules including organic and inorganic nutrients and byproducts or toxins, signaling molecules, SMs (e.g., pleiotropic factors or antibiotics), pH, or other chemical exchange. Biochemical or biological interactions may result from the action of extracellular enzymes (e.g., lipases) or biochemical recognition of cell wall or membranes (e.g., mycolic acids).[32] These stimuli trigger an array of potential responses including via regulatory networks and/or biochemical pathways. Previously, we have demonstrated and characterized the degree of metabolomic expansion in NF that is induced by developing resistance to rifampicin and streptomycin and determined the structure of a new family of aromatic polyketides induced by adaptive mutations in RNA polymerase.[33] Notably, the single nucleotide changes engendered by this vertically acquired antibiotic resistance resulted in large global changes in measured metabolites. The acquisition of resistance to antibiotic SMs may be considered an intergeneric interaction response.

Similarly, the current study reveals how a small change in the biology of a system, in this case via a low inoculum of a challenger organism, can have large effects on the metabolomic output, of which SMs again play a featuring role. The SOM stimulus-response identification workflow provides a tool to begin examining how these interactions excite changes in microbiological communities intergenerically and can potentially be used to provide insight into more complex interkingdom interactions in eukaryotic systems.

Materials and Methods

Fermentation conditions

Seed culture (1mL) of NF in ISP2 broth was inoculated into 25 mL of R4 fermentation medium and incubated in a rotary shaker at 30 °C. At the same time seed cultures of EC and BS, TP and RW were prepared and incubated for 24 hours prior to addition to NF fermentation flasks. A 200 µL inoculum of each competing organism was added to already established NF cultures and also to flasks containing sterile R4 medium. A total of 4 different co-cultures and a total of 5 monocultures (one for NF and one for each competitor) were then incubated in a rotary shaker at 30 °C for another 6 days. At the time of inoculation, cultures contained 7×10^7 cfu/mL of NF and 3×10^6, 3×10^5, 2×10^6, 7×10^6 cfu/mL of RW, TP, BS, and EC respectively, corresponding to a NF/challenger ratio ranging from 10:1 (EC) to 200:1 (TP) at the beginning of mixed culture. At six days, dilution plating of co-cultures only resulted in NF colonies, further emphasizing the excess population of this strain in culture conditions. For scaled fermentation, the number of 25 mL culture flasks was multiplied up to a total of 500 mL of fermentation volume. Also, only one competing organism (RW) was picked as an activator since it seemed to be the most effective in stimulating ciromicin production in an initial experiment.

Culture extractions

For extraction, 25 mL of methanol was added to each mixed and monoculture and shaken for 1 hour. Mycelia were separated from broth by centrifugation and supernatants were dried *in vacuo* and subjected to UPLC/IM-MS analysis. For scaled fermentation, the same methanol extraction of each 25 mL culture was performed but supernatants were first combined and then dried *in vacuo* for the isolation of ciromicins.

Ciromicin purification

Crude extract containing predominantly ciromicin A (**1**) and some ciromicin B (**2**) was first separated on a preparative RP-HPLC using linear gradient of water/acetonitrile containing 0.1 % of formic acid. Fractions with UV indicative of ciromicins were then combined and applied on size exclusion Sephadex LH-20 column for a gravity run in methanol. The final yields of purified ciromicin A (**1**) and B (**2**) were 5 mg and 0.5 mg respectively.

Photochemical conversion of ciromicin

For photochemical reaction experiment, we dissolved 6 mg of purified ciromicin A (**1**) in 5 mL of methanol and split it into 10 samples, 500 µL each. We then exposed each sample to a different wavelength light for 30 min starting with visible light of 700 nm, through the long and medium wave UV and ending with short wave UV of 150 nm. The experiment was performed in the dark room using ORIEL Illuminator equipped with 1000 W Hg(Xe) ozone free arc lamp and 450-1000W universal arc lamp power supply. We monitored the progress of conversion via LC-MS analysis.

UPLC/IM-MS Data Acquisition

Sample extracts were resuspended at a concentration of 200 mg/mL in mobile phase A (see below). UPLC/IM-MS-MS (MS^E) data acquisition was performed on a SYNAPT G2 HDMS (Waters Corp., Milford, MA) with a 30 min gradient. Mobile phase A consisted of 95 % H2O and 5 % acetonitrile with 10 mM ammonium acetate, and mobile phase B consisted of 95 % acetonitrile and 5 % water with 10 mM ammonium acetate. A 1x100 mm 1.7 µm particle BEH-T3 C18 column (Waters Corp.) was used for chromatographic separations with a flow rate of 75 µL/min and a column temperature of 40°C. An autosampler with a loop size of 5 µL held at 4 °C was used for sample injection. The initial solvent composition was 100 % A, which was held for 1 min and ramped to 0 % A over the next 15 min, held at 0 % A for 2 min, and returned to 100 % A over a 0.1 min period. The gradient was held at 100 % A for the next 10.9 min for equilibration. Prior to analysis of the sample queue, ten sequential column-load injections were performed with 5 µL of the quality control. Samples were analyzed in triplicate with the order of injection randomized.

IM-MS^E spectra were acquired at a rate of 2 Hz from 50-2000 Da in positive ion mode for the duration of each sample analysis. The instrument was calibrated to less than 1 ppm mass accuracy using sodium formate clusters prior to analysis. A two-point internal standard of leucine enkephalin was infused in parallel to the sample at a flow rate of 7 µL/min, and data were acquired every 10 s. The source capillary was held at 110 °C and 3.0 kV, with a desolvation gas flow of 400 L/h and a temperature of 150 °C. The sampling cone was held at a setting of 35.0, with the extraction cone at a setting of 5.0. In the MS^E configuration, low and high energy spectra were acquired for each scan. High energy data provided a collision energy profile from 10-30 eV in the transfer region, providing post-mobility fragmentation. Ion mobility separations were performed with a wave velocity of 550 m/s, a wave height of 40.0 V, and a nitrogen gas flow of 90 mL/min, with the helium cell flow rate at 180 mL/min. Internal calibrant correction was performed in real time.

Metabolomic analysis

Data files were converted from .raw to mzXML using the msconvert tool from ProteoWizard 3.0.5759. Each mixed fermentation was treated as a separate experiment, and samples were grouped into 4 sets, each containing monoculture NF, monoculture competing organism, and NF & competing organism. Peak picking and alignment were then performed using XCMS in R10 for each group, and the 4 resultant XCMS datasets were combined using meta-XCMS. The combined dataset was then manually inspected. This method was found to give better results than performing a single XCMS analysis on the entire dataset. Processed data were then normalized by total ion count, and low intensity features were checked against the raw data to ensure legitimacy. Features not discernable from noise and features after 15 minutes were removed from the dataset. Feature intensities from each co-culture were compared to their respective monocultures and sorted by fold change to select for up and down regulated features.

Multivariate statistical analyses were performed using Umetrics extended statistics software EZinfo version 2.0.0.0 (Waters, Milford, MA). For MEDI analyses, triplicate injections from the XCMS readout were averaged, and data were formatted for GEDI input. GEDI software allows users to adjust sorting parameters. A grid of 25 x 26 nodes was selected, with 1st and 2nd phase training iterations of 80 and 100 respectively.

Acknowledgments

We thank Dr. Cody Goodwin and Dr. C. Ruth McNees for contributions to developing comparative metabolomics concepts described in this work. This work was supported by the National Institutes of Health (grant to B.O.B. and J.A.M. GM092218, and training grant for B.C.C. T32 GM 0650086), and the Vanderbilt Institute of Chemical Biology and Vanderbilt Institute for Integrated Biosystems Research and Education.

References

1. Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O., Mapping microbial response metabolomes for induced natural product discovery. *ACS chemical biology* **2015**.

2. Du, Y.; Derewacz, D. K.; Deguire, S. M.; Teske, J.; Ravel, J.; Sulikowski, G. A.; Bachmann, B. O., Biosynthesis of the apoptolidins in Nocardiopsis sp FU 40. *Tetrahedron* **2011,** *67* (35), 6568-6575.

3. Shindo, K.; Kamishohara, M.; Odagawa, A.; Matsuoka, M.; Kawai, H., Vicenistatin, a Novel 20-Membered Macrocyclic Lactam Antitumor Antibiotic. *J. Antibiot.* **1993,** *46* (7), 1076-1081.

4. Futamura, Y.; Sawa, R.; Umezawa, Y.; Igarashi, M.; Nakamura, H.; Hasegawa, K.; Yarnasaki, M.; Tashiro, E.; Takahashi, Y.; Akarnatsu, Y.; Imoto, M., Discovery of incednine as a potent modulator of the anti-apoptotic function of Bcl-xL from microbial origin. *J. AM. CHEM. SOC.* **2008,** *130* (6), 1822-+.

5. Turnbaugh, P. J.; Ley, R. E.; Hamady, M.; Fraser-Liggett, C. M.; Knight, R.; Gordon, J. I., The Human Microbiome Project. *Nature* **2007,** *449* (7164), 804-810.

6. Eichler, G. S.; Huang, S.; Ingber, D. E., Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* **2003,** *19* (17), 2321-2322.

7. Giessen, T. W.; Franke, K. B.; Knappe, T. A.; Kraas, F. I.; Bosello, M.; Xie, X.; Linne, U.; Marahiel, M. A., Isolation, Structure Elucidation, and Biosynthesis of an Unusual Hydroxamic Acid Ester-Containing Siderophore from Actinosynnema mirum. *J Nat Prod* **2012,** *75* (5), 905-914.

8. Plaza, A.; Viehrig, K.; Garcia, R.; Müller, R., Jahnellamides, α-keto-β-methionine-containing peptides from the terrestrial myxobacterium Jahnella sp.: structure and biosynthesis. *Org. Lett.* **2013,** *15* (22), 5882-5.

9. Medema, M. H.; Blin, K.; Cimermancic, P., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nuc. Acids Res.* **2011,** *39* (39), 339-46.

10. Shinohara, Y.; Kudo, F.; Eguchit, T., A Natural Protecting Group Strategy To Carry an Amino Acid Starter Unit in the Biosynthesis of Macrolactam Polyketide Antibiotics. *J. AM. CHEM. SOC.* **2011,** *133* (45), 18134-18137.

11. Caffrey, P., Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketicle synthases. *ChemBioChem* **2003,** *4* (7), 654-657.

12. Kwan, D. H.; Sun, Y. H.; Schulz, F.; Hong, H.; Popovic, B.; Sim-Stark, J. C. C.; Haydock, S. F.; Leadlay, P. F., Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **2008,** *15* (11), 1231-1240.

13. Takaishi, M.; Kudo, F.; Eguchi, T., Biosynthetic pathway of 24-membered macrolactam glycoside incednine. *Tetrahedron* **2008,** *64* (28), 6651-6656.

14. Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K., Cloning, sequencing, and functional analysis of the biosynthetic gene cluster of macrolactam antibiotic vicenistatin in Streptomyces halstedii. *Chem. Biol.* **2004,** *11* (1), 79-86.

15. Raju, R.; Piggott, A. M.; Conte, M. M.; Capon, R. J., Heronamides A-C, new polyketide macrolactams from an Australian marine-derived Streptomyces sp A biosynthetic case for synchronized tandem electrocyclization. *Org. Biomol. Chem.* **2010,** *8* (20), 4682-4689.

16. Jorgensen, H.; Degnes, K. F.; Sletta, H.; Fjaervik, E.; Dikiy, A.; Herfindal, L.; Bruheim, P.; Klinkenberg, G.; Bredholt, H.; Nygard, G.; Doskeland, S. O.; Ellingsen, T. E.; Zotchev, S. B., Biosynthesis of Macrolactam BE-14106 Involves Two Distinct PKS Systems and Amino Acid Processing Enzymes for Generation of the Aminoacyl Starter Unit. *Chem. Biol.* **2009,** *16* (10), 1109-1121.

17. Beaudry, C. M.; Malerich, J. P.; Trauner, D., Biosynthetic and biomimetic electrocyclizations. *Chem. Rev.* **2005,** *105* (12), 4757-78.

18. Shigemori, H.; Bae, M. A.; Yazawa, K.; Sasaki, T.; Kobayashi, J., Alteramide-a, a New Tetracyclic Alkaloid from a Bacterium-Alteromonas Sp Associated with the Marine Sponge Halichondria-Okadai. *J. Org. Chem.* **1992,** *57* (15), 4317-4320.

19. Beaudry, C. M.; Trauner, D., Synthetic studies toward SNF4435 C and SNF4435 D. *Org. Lett.* **2002,** *4* (13), 2221-2224.

20. Kurosawa, K.; Takahashi, K.; Tsuda, E., SNF4435C and D, novel immunosuppressants produced by a strain of Streptomyces spectabilis - I. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot.* **2001,** *54* (7), 541-547.

21. Metzker, M. L., Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation. *Nat. Rev. Genet.* **2010,** *11* (1), 31-46.

22. Shendure, J.; Ji, H. L., Next-generation DNA sequencing. *Nat. Biotechnol.* **2008,** *26* (10), 1135-1145.

23. Kallifidas, D.; Brady, S. F., Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. *Methods Enzymol.* **2012,** *517*, 225-39.

24. Shao, Z.; Zhao, H., DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Methods Enzymol.* **2012,** *517*, 203-24.

25. Yaegashi, J.; Oakley, B. R.; Wang, C. C., Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in Aspergillus nidulans. *J. Ind. Microbiol. Biotechnol.* **2014,** *41* (2), 433-42.

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.; Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (26), 1743-1752.

27. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C., Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013,** *76* (9), 1686-1699.

28. Kersten, R. D.; Yang, Y.-L. L.; Xu, Y.; Cimermancic, P.; Nam, S.-J. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C., A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **2011,** *7* (11), 794-802.

29. Liu, W. T.; Lamsa, A.; Wong, W. R.; Boudreau, P. D.; Kersten, R.; Peng, Y.; Moree, W. J.; Duggan, B. M.; Moore, B. S.; Gerwick, W. H.; Linington, R. G.; Pogliano, K.; Dorrestein, P. C., MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in Streptomyces roseosporus. *J antibiot (Tokyo)* **2014,** *67* (1), 99-104.

30. VerBerkmoes, N. C.; Denef, V. J.; Hettich, R. L.; Banfield, J. F., Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol* **2009,** *7* (3), 196-205.

31. Nicholson, J. K.; Holmes, E.; Kinross, J. M.; Darzi, A. W.; Takats, Z.; Lindon, J. C., Metabolic phenotyping in clinical and surgical environments. *Nature* **2012,** *491*, 384.

32. Onaka, H.; Mori, Y.; Igarashi, Y.; Furumai, T., Mycolic Acid-Containing Bacteria Induce Natural-Product Biosynthesis in Streptomyces Species. *Appl. Environ. Microbiol.* **2011,** *77* (2), 400-406.

33. Derewacz, D. K.; Goodwin, C. R.; McNees, R. C.; McLean, J. A.; Bachmann, B. O., Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (6), 2336-41.

Chapter 4 Prioritization of natural products from hypogean actinomycetes

This chapter was adapted with permission from the American Society of Microbiology from an article written by Brett C. Covington, Jeffery M. Spraggins , Audrey E. Ynigez-Gutierrez , Zachary B Hylton , and Brian O. Bachmann, first published in Applied and Environmental Microbiology.

Introduction

Chapters 2 and 3 described our applications of comparative metabolomics and MVSA to prioritize NPs from stimulated, genomically characterized organisms. In this chapter we describe further applications of environmental stimuli to *Streptomyces* and rare actinobacteria isolated from cave sediments. As previously mentioned, inspection of microbial genomes reveals that the products of a large portion of the genetically encoded SMs remain unidentified in culture extracts.[1] This discrepancy between the potential and isolable SM production may stem from low expression and/or translation of SM biosynthetic genes and limitations in our methods of detection via abundance or biological activity.[2] So called 'brothological' methods (such as OSMAC)[3] that vary cultivation parameters, such as media composition, pH, aeration, and temperature, can effectively increase the observable SM output from microbial producers. These general approaches may increase SM production by altering precursor levels, biosynthetic enzyme activity/stability, or through changes in gene regulation, microbial metabolism, and growth phase timing. In the previous chapters we have described our alternative strategy, which employs defined chemical or biological agents in a single growth medium that model ecological and environmental stimuli. The finding that these variations in discrete chemical and biological stimuli can alter SM production are consistent with our central hypothesis that most NPs are used to respond to environmental interactions.

Regardless of the means of stimulating SM production, the identification of the often low-abundant SMs of interest within the observable metabolome remains a substantial challenge. The application of

comparative metabolomics methods described in Chapter 1 such as MVSA and/or molecular networking[4-6] has superseded the 'stare and compare' approach, providing unbiased assessments of metabolite covariance across samples. In previous chapters we have demonstrated that combining comparative metabolomics with the application of discrete stimuli can effectively prioritize SMs for isolation. However, while the combination of chemical/biological stimuli and comparative metabolomics has led to the activation, production, and identification of NPs from actinomycetes of the genus *Streptomyces*[7-10] and less commonly other actinomycetes,[5, 11-13] these strategies have not been thoroughly evaluated across other actinobacterial genera. Cave environments are in many ways unique compared to surface habitats. The lack of sunlight, temperature stability, high humidity, and limited nutrients likely foster specialized communities of microorganisms within these isolated environments. Hypogean bacteria may be an important source of chemical novelty, but it was unclear how SM production from these microorganisms will respond to current activation methods. In this chapter we discuss the isolation of bacteria from cave sediments and the stimulation of cultured hypogean actinomycetes to induce NP production. We demonstrate the applicability of subinhibitory antibiotic exposure, rare earth metal exposure, and microbial competition activation strategies to activate secondary metabolism across a phylogenetically diverse selection of cave-derived actinomycetes. We also show that comparative metabolomics analyses of stimulated cultures prioritize SMs from crude extracts. This work revealed large changes in the observed metabolomes and overproduction of SMs in response to environmental stimuli. We isolated a subset of metabolites prioritized through comparative metabolomics analyses and identified members of the actinomycin, hypogeamycin, tetarimycin, aloesaponarin, and propeptin classes as well as a novel linear polyketide (funisamine) produced by a rare *Streptosporangium* strain under mixed culture stimuli conditions.

Results

Phylogenetically diverse actinomycetes isolated from hypogean environments.

Phylogenetically diverse bacteria from unique unexplored environments have a strong track record of revealing useful chemical diversity[14-15]. Hypogean sediments are unique environments that have been shown to be replete with bacterial diversity[16-18] but not extensively explored for SM producing organisms. Caves maintain microbial habitats distinct from epigean (surface) ecosystems. Limestone caves in the central United States are generally isothermal year round (~13 °C), highly oligotrophic,[19] and completely aphotic. Formed millions of years ago by chemical and biological[20] dissolution of limestone, they are often highly humid due to hydrological activity, and therefore present vast, wet, highly oxygenated surface areas for maintaining microbial biofilms competing for allochthonous dissolved organic matter. Based on these unique environmental parameters, we selected actinomycetes sourced from cave environs for stimulus-response mapping.

To investigate the secondary metabolic responses of cave-derived actinomycetes to environmental stimuli, environmental samples were acquired from Hardin's (Ashland City), Snail Shell (Rockvale), and Cagle's Chasm (South Pittsburg) caves in Tennessee by aseptically collecting cave sediments and swabbing cave formations. These formations (speleothems) included flowstone, stalactites, stalagmites, wall coralloids, and columns. To provide a basis for comparison of cultivatable diversity to studies of actinomycetes from surface environments we used established actinobacterial isolation procedures.[21] After collection, cave sediment samples were dried and used 100-fold diluted International *Streptomyces* Protocol 2 (ISP2) agar through dilution plating, while swabs were suspended and directly applied to agar plates without prior desiccation. Hundreds of individual bacterial colonies were isolated using this

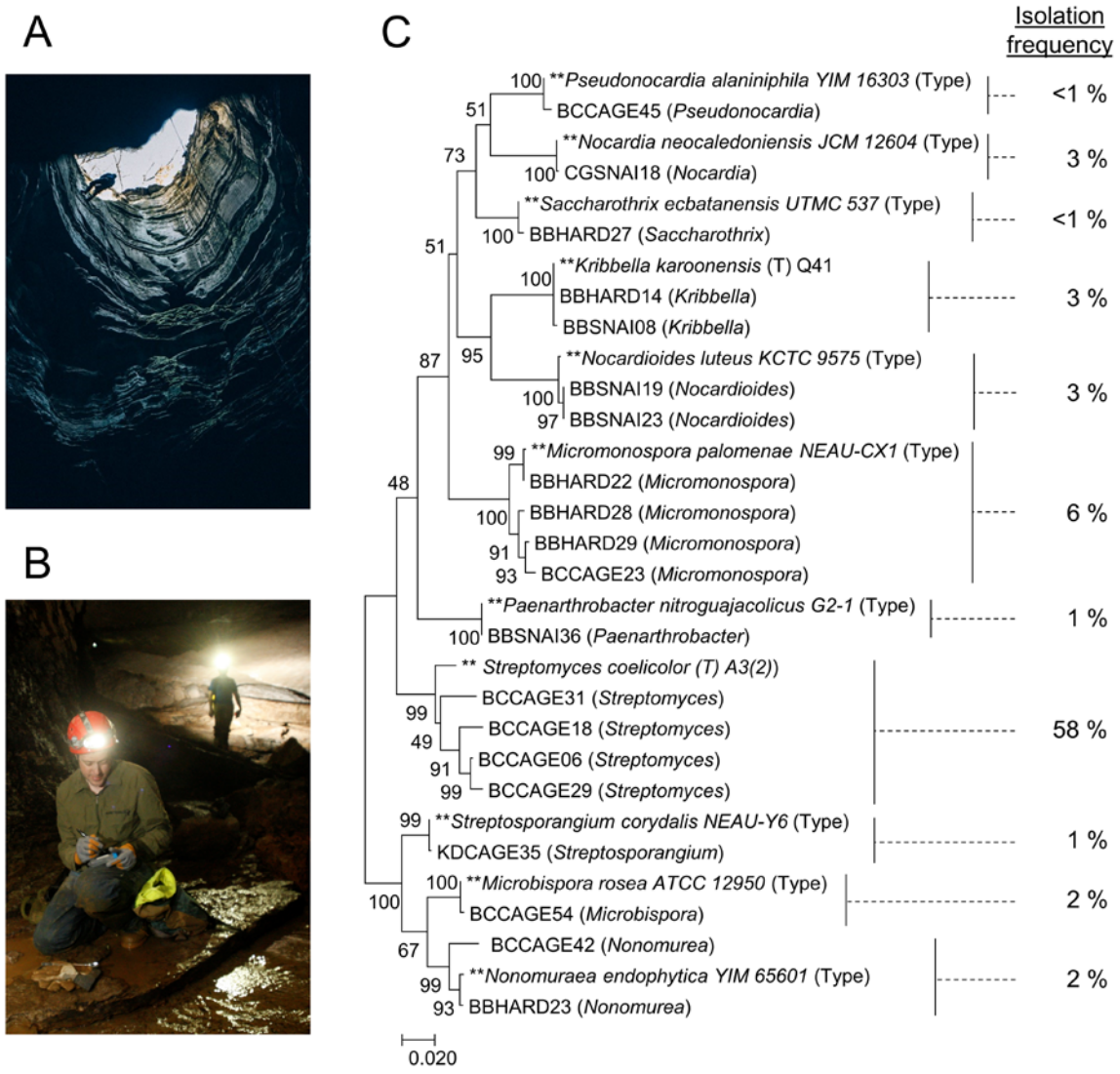


Figure 4-1 Phylogenetic tree of cave organisms. Images from (A) Cagle's chasm and (B) Snail shell caves. (C) Phylogenetic tree constructed using MEGA7 with maximum likelihood analysis from 16S sequence alignment of the 20 strains selected for stimulus exposure studies and 11 type strains for reference. Bootstrap values based on 1000 pseudoreplicates are shown next to the branch points. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Isolation frequency of genera in our study is calculated as the number of strains of a given genus identified divided by the total number (155) of sequenced isolates. Resubmitted from *ASM* with permissions from the American Society for Microbiology

approach, and 16S rDNA sequences were acquired for 155 strains. The majority of these (58 %) were most

closely related to actinomycetes of the genus *Streptomyces*; however, a number of rare actinomycetes were isolated (Table B-1), which are important sources of novel SMs [22]. A group of 20 cave strains from this sequenced collection was selected for this study, consisting of four *Streptomyces* species and 16 less commonly described actinomycetes with high 16S rDNA sequence similarity to the genera: *Microbispora*, *Micromonospora*, *Streptosporangium*, *Saccharothrix*, *Nonomuraea*, *Pseudonocardia*, *Nocardioides*, *Nocardia*, *Kribella*, and *Paenarthrobacter* (Figure 4-1).

Stimuli increase natural product biosynthesis across actinomycete genera.

Selected strains were exposed to six different stimuli previously reported to activate SM production in actinomycetes: sub-inhibitory antibiotic concentrations (1/10 MIC) of either rifampicin or streptomycin [23], rare earth metal exposure with lanthanum or scandium,[24] and microbial competition with mycolic acid containing bacteria TP or RW.[10] All organisms were cultured in ISP2 media, a benchmark defined media, and metabolites were extracted after seven days. These total cell extracts were then analyzed through HPLC-ESI-MS in both positive and negative ionization modes. In this study, combined positive and negative ESI data for all extracts were prepared for initial comparative analysis. The metabolomic platform XCMS[25] was used to process datasets, providing peak detection, retention-time correction, and chromatogram alignment to generate a combined list of ~19,000 features - ions with distinct *m/z* and retention times. Feature ion abundances from technical replicates were averaged, and pairwise comparisons were performed between each stimuli condition and respective controls to determine the number of features with increased production through stimuli exposure.

The overall responses to stimuli were measured by the number of features with ≥10-fold increased ion abundance compared to control conditions. Responses varied between genera (Figure B-1-Figure B-4), and even phylogenetically similar isolates displayed markedly differing metabolomic response patterns. For instance, of the two *Kribella* isolates, one (BBHARD14) responded most strongly to metal exposure

while the other (BBSNAI08) responded most strongly to mixed culture. Of the four *Micromonospora* isolates, two responded most strongly to mixed culture (BCCAGE23 and BBHARD28), while two responded most strongly to metal exposure (BBHARD22 and BBHARD29). Interestingly, across strains there was relatively little overlap of overproduced features between stimuli of the same type. For example, within the antibiotic category, rifampicin and streptomycin treatment generally led to accumulation of different metabolites (Figure B-2). Similar results were observed for metal and mixed culture exposure (Figure B-3, B-4), with the exception of two *Nocardioides* isolates (BBSNAI19 and BBSNAI23) which exhibited a high degree of overlap between responses to lanthanum and scandium exposure conditions. The four *Streptomyces* isolates demonstrated relatively fewer responses, suggesting that rare actinomycetes are particularly sensitive to applications of stimuli.

At least eight families of NPs were identified within the extracts (*vide infra*), and increased production was observed for seven of these in at least one stimulus condition. Two of these NP families, an unidentified polyene NP and a novel aminopolyol funisamine, were detectable only under stimuli conditions, and others, like tetarimycin, were barely detectable in control conditions and would have been overlooked if not for enhanced production under stimuli conditions (Table B-2). The unidentified polyene was activated in a cave isolate *Nonomuraea* sp. BCCAGE42 under antibiotic exposure conditions, and funisamine was produced by a *Streptosporangium* sp. KDCAGE35 under mixed culture conditions. Interestingly, while exposure to either rifampicin or streptomycin elicited comparatively few bulk metabolomic responses across strains relative to mixed culture and metal stimuli, antibiotic exposure demonstrated a substantial effect on the accumulation of individual NP features, which were conspicuous in pairwise analysis. Of 23 putative NP features with abundances ≥10-fold in stimuli conditions vs controls, 17 were observed in greater abundance under sub-inhibitory antibiotic exposure conditions (Figure B5).

Comparative metabolomics prioritizes secondary metabolites from stimuli exposure.

Two multivariate statistical analysis methods were used to prioritize metabolomic features for isolation and dereplication, volcano plots and S-plots (Figure 4-2). Volcano plots organize features on the y-axis by the probability of observing such a fold change under the null hypothesis of no-change (p-value) and by relative ion abundance fold change on the x-axis, but not by feature ion abundance. By comparison, S-plots organize features on the y-axis by their Pearson correlation coefficient, a measure of their linear correlation, and on the x-axis by their component coefficients, a measure of a feature's contribution to the total variance between conditions, which is influenced by feature ion abundance. It is also important to note that the relative intensities of ions from a mixture of small molecules as detected by MS do not necessarily correlate with the relative concentration of those compounds, due to varying ionization efficiencies. While these analyses aid in prioritizing a number of features, the practical isolability of compounds from an extract is limited by many factors, including abundance, chromatographic properties, and chemical stability. Both MS and UV/Vis absorption spectroscopy were utilized to characterize culture extracts, and the absorption spectra aided in determining the abundance and isolability of prioritized features.

Comparative metabolomics analyses of extracts from the cave isolate *Micromonospora* sp. BBHARD22 revealed a number of abundant features increased under scandium supplemented conditions (Figure 4-2, Table B-2, B-3). Overlaying extracted ion chromatograms for S-plot prioritized features onto the total ion and UV/Vis chromatograms indicated one feature was readily isolable based on abundance in both MS and UV/Vis detection as well elution in a non-complex region of the chromatogram. The production of this feature was increased six-fold under scandium metal exposure conditions and was ultimately identified by NMR spectroscopy to be the known anthraquinone NP, aloesaponarin II, originally isolated

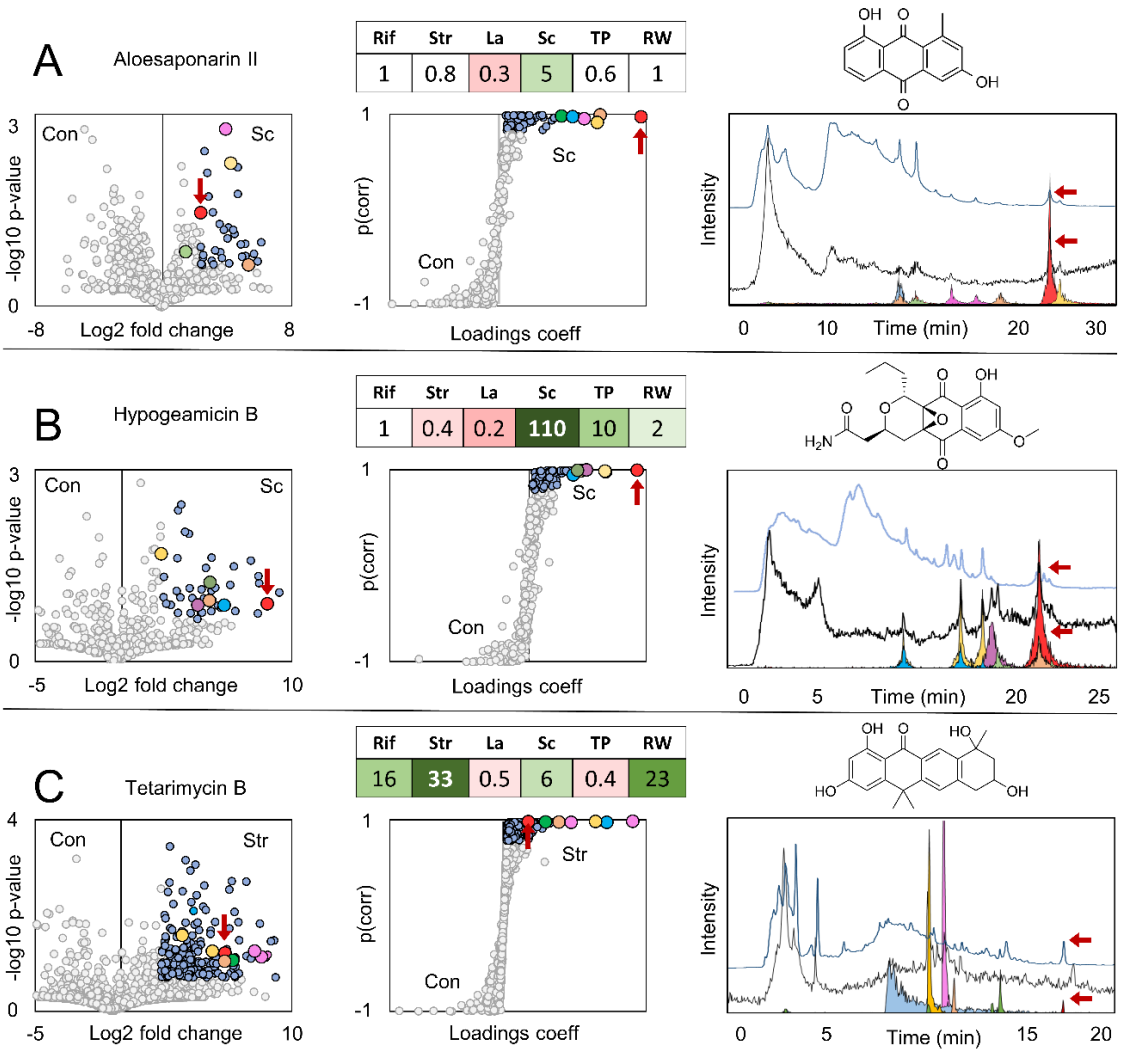


Figure 4-2 Prioritization of natural products from comparative metabolomics. Volcano and S-plot prioritization of (A) aloesaponarins (ESI -) from *Micromonospora* sp. BBHARD22 exposed to scandium (Sc), (B) hypogeamicins (ESI +) from *Nonomuraea* sp. BBHARD 23 exposed to scandium (Sc), and (C) tetarimycins (ESI +) from *Microbispora* sp. BCCAGE54 exposed to streptomycin (Str). Circles are colored for features outside (gray) and within (blue) 0.8 Pearson correlation coefficient thresholds for S-plots and 0.8 probability value, 5-fold change thresholds for volcano plots. Features corresponding to (A) aloesaponarin, (B) hypogeamicin, and (C) tetarimycin are colored red in volcano and S-plots and are additionally distinguished by red arrows in volcano and S-plots as well as the UV/Vis and MS chromatograms. Overlays of TIC chromatograms, total absorption spectrums, and extracted ion chromatograms for abundant features prioritized through S-plot analyses. Within each row the color of the peak corresponds to the color of the circle in the volcano and S-plots. The fold change of the highlighted NP in stimuli conditions relative to the controls are shown tables over the S-plots for rifampicin (Rif), streptomycin (Str), lanthanum (La), scandium (Sc), TP, and RW conditions. Resubmitted from *ASM* with permissions from the American Society for Microbiology

from the plant *Aloe saponaria*[26] but also previously observed in extracts of both marine[27] and terrestrial[28]

Streptomyces sp. Another abundant feature prioritized from scandium exposure culture extracts was additionally identified as the anthraquinone okicenone.[29] In addition to these confirmed anthraquinones, seven additional features were identified across multiply applied stimuli that were assigned as likely SMs based on their MS, MS/MS, and UV (Figure B-12) spectra in comparison to compounds isolated and characterized from this strain. Notably, these putative SMs, likely congeners of aloesaponarin and okicenone, vary in concentration across stimuli, with discrete stimuli facilitating increased production of individual congeners.

Analysis of the cave isolate *Nonomuraea* sp. BBHARD23 under stimuli conditions prioritized several features with enhanced production under scandium metal exposure conditions (Figure 4-2B). Under these conditions the production of the pyronapthoquinones hypogeamicins A, B, and C, were increased and these SMs were prioritized through comparative metabolomics; the most abundant of these was identified as hypogeamicin B.[30] Additionally, several unidentified features were prioritized with similar *m/z* and UV absorbance spectra as the identified hypogeamicins, but with distinct chromatographic retention times. These putative hypogeamicin analogs were increased under mixed culture conditions, with production increased almost 600-fold under exposure to competition with *Rhodococcus sp*. In both of these examples, the highlighted SM classes had very high ion abundance relative to other prioritized metabolites. Again, individual characterized and putative hypogeamicin congeners vary across stimuli (Table B-2, B-3). For example, the production of hypogeamicin C and four other putative analogs could be elicited by administration of sub-inhibitory streptomycin, whereas hypogeamicin A could not. Scandium exposure led to a ~2, 110, and 6-fold increase in hypogeamycin A, B, and C respectively, but resulted in a decrease in the abundance of some putative hypogeamycin analogs.

Comparative analysis of resulting stimuli extracts from *Microbispora* sp. BCCAGE54 revealed increased production of a large number of metabolites. Volcano and S-plot analyses of sub-inhibitory streptomycin exposure prioritized 295 features, but the features with the highest ion intensity co-eluted in a crowded

region of the chromatogram and eluded isolation attempts (Figure 4-2C). By comparison, one prioritized feature with relatively low ion abundance, yet a 33-fold increased production under antibiotic exposure conditions eluted in a chromatographically tractable region and was correlated with a relatively large UV/Vis absorption peak. This induced feature was identified as tetarimycin B, a polyketide SM recently discovered by Kallifidas et al. via the heterologous expression of an environmentally-derived type II polyketide synthase (PKS) gene cluster.[31] Sequence homology searches have indicated that tetarimycin encoding gene clusters may be widespread in the environment, and the lack of identified tetarimycins from culture based screens indicates the difficult to detect nature of this SM family.[31] Indeed, while tetarimycin B was produced in detectable levels within our unstimulated *Microbispora* cultures, the low abundance ions detected in these conditions would likely have been overlooked entirely without prioritization through comparative metabolomics. As in previous examples, production of several putative new tetarimycin analogs was elicited in multiple stimuli, along with several features displaying properties of SMs with high molecular weight, extended UV/VIS chromophores, and responses profiles matching tetarimycins (Figure B20). During the isolation of tetarimycin, an additional pair of metabolites were isolated from large scale culture extracts and identified as propeptin 1 and 2. The relative abundances of these two congeners varied subtly but consistently across stimuli classes with mixed culture and antibiotic exposure conditions resulting in higher ratios of propeptin 1 and metal exposure conditions resulting in higher ratios of propeptin 2 (Table B-2).

As noted, volcano and S-plots emphasize different *m/z* features within the data. Features with significant fold-changes are prioritized by volcano plots whereas those with significant differences in abundance are prioritized by S-plots. The chromatogram overlays in Figure 4-2 show features emphasized through S-plot analyses, which were found to be the most reliable method for prioritizing features for isolation. Similar overlays for features prioritized by volcano plot thresholds are available in Appendix B (Figure B-6). Tables of observed SM fold change across stimuli conditions (Table B-2, B3) along with

extracted ion (Figure B-7 - Figure B-11) and UV (Figure B-12 - Figure B-22) chromatograms for prioritized metabolites and NMR spectra of identified metabolites (Figure B-23 - Figure B-29) are provided in Appendix B.

Novel polyketide produced from interactions between rare Streptosporangium and microbial co-culture

A cave isolate, *Streptosporangium* sp. KDCAGE35 with close 16S sequence similarity to *Streptosporangium corydalis* (99.6 % identity), hereafter referred to as *Streptosporangium caverna* (from the Latin noun for cave, *caverna*), was shown to be highly responsive to mixed culture stimulus with ~16 % of the total detected features displaying increased ion intensity at 10-fold or higher levels under these conditions. While the production of many features under mixed culture conditions were increased relative to unstimulated cultures, most remained insufficiently abundant for isolation. Further mixed culture screens were performed with *S. caverna* using EC, *Bacillus* sp. KDCAGE13 (BK) (99.9 % 16S similarity to *Bacillus simplex*), and BS in addition to the previously screened RW and *TP*. Volcano and S-plot analyses comparing mixed cultures with *Bacillus* strains and *S. caverna* monocultures identified a metabolite with a mass of 1176.6 Da that eluted in a tractable region of the chromatogram. This mass was correlated with a UV spectrum (λ-max: 370, 389, and 413 nm) characteristic of conjugated polyene systems (e.g. heptaenes). The production of this feature was also induced in screens with TP, but the significantly enhanced production under BS competition stimulus conditions (~70-fold increase) enabled isolation and identification. Comparisons of production yields of this prioritized feature from co-cultivation of different species with *S. caverna* are shown in Figure 4-3F. The target compound was isolated as a bright yellow solid from several small scale mixed cultures (totaling 2 L) of the *S. caverna* and BS mixed culture using a combination of adsorbent resin, size exclusion chromatography, and reverse phase chromatography. Electrospray Fourier-transform ion cyclotron resonance MS of the isolated compound yielded an accurate

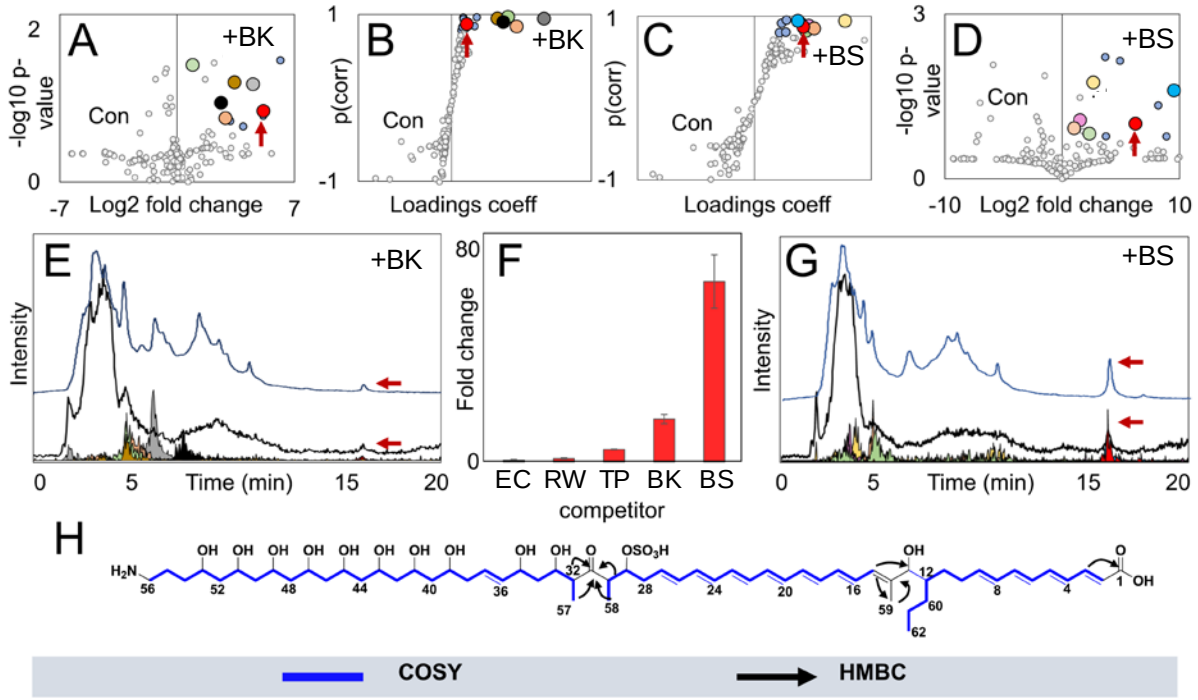


Figure 4-3 Activation of funisamine biosynthesis through mixed culture. (A, B, C, D) Volcano and S-plot comparisons between *S. caverna* & BK mixed (+BK) and mono (Con) cultures and *S. caverna* & BS mixed (+BS) and mono (Con) cultures (ESI -). Circles are colored for features outside (gray) and within (blue) 0.8 Pearson correlation coefficient thresholds for S-plots and 0.8 probability value, 5-fold change thresholds for volcano plots. Feature corresponding to funisamine is enlarged colored red. (E, G) Overlays of total ion chromatograms, total absorption spectrums, and extracted ion chromatograms for funisamine (*m/z*: 1176.6) and other features prioritized from *S. caverna* mixed cultures with BS and BK. (F) Fold change of funisamine production in mixed cultures with RW, EC, TP, BK, and BS vs control quantified by UV absorption at 413 nm. (H) Proposed structure of funisamine generated from HSQC, COSY, TOCSY, HMBC and NOESY spectra. Major COSY and HMBC correlations shown. Stereochemical configurations of substituent groups and double bonds inferred from bioinformatic analysis. Resubmitted from *ASM* with permissions from the American Society for Microbiology

mass of 1176.6507 $[M-H]^-$ (calcd. for $C_{62}H_{98}NO_{18}S^-$, 1176.6510, 0.25 ppm), which in combination with formula constraints from NMR data and isotope analyses suggested a molecular formula of $C_{62}H_{99}NO_{18}S$. The full structure was derived from COSY, TOCSY, HSQC, HMBC, and NOESY NMR spectroscopy (Figure B-29 - Figure B-41), which identified two conjugated systems assignable to four and seven double bonds, respectively (Figure 4-3H). The structure was separated into three spin systems by two quaternary carbons at C-14 and C-31. Fig. 3H shows important HMBC correlations which facilitated connections across these

spin systems. HMBC correlations show interactions from the proton at position CH-13 and methyl protons CH_3-59 to the C-14 quaternary carbon and to CH-15 of the next spin system. Methyl protons at CH_3-59 also showed HMBC correlations with CH-13 and long-range coupling in the COSY spectra with the proton at CH-15. The second quaternary carbon, a ketone at C-31, was bridged by HMBC correlations from protons at CH-30, CH-32, CH_3-57, and CH_3-58 to the C-31 carbon. NOESY correlations between methine protons CH-29, CH-30, CH-32, and CH-33 to methyl protons CH_3-57, and CH_3-58 confirmed connectivity in these spin systems. A shift of 4.61 ppm for the proton at CH-29 was consistent with a sulfate group at C29. This is additionally supported by the identification of a fragment consistent with the loss of a sulfate group by MS (Figure B-42 - Figure B-44, Table B-5). We named this compound funisamine, (from the Latin noun for rope, *funis*).

Co-cultivated elicitor strains *E. coli* and *B. subtilis* had similar growth rates and doubling times under the culture conditions (Figure B-46) yet had differing impacts on funisamine compound production (Figure 4-3). The mechanism and factors governing the apparent *Bacillus* species selectivity of this activation remain unknown, though there appears to be no correlation to elicitor strain growth rate. While some related aminopolyols do have reported antimicrobial properties against *Bacillus* strains,[32] funisamine exhibited antimicrobial activity only at very high concentrations, ~25 times the levels produced under mixed culture conditions. Microtiter plate antibacterial assays with *S. aureus, E. coli*, and *C. albicans* yielded MIC values of ~1 mM. Funisamine (and perhaps other long chain aminopolyols) may also serve a role other than as an antibiotic in the producer's natural environment.

An issue with mixed culture stimulus approaches is the uncertainty of which organism(s) within the culture are ultimately responsible for producing compounds observed within an extract. Funisamine was produced in mixed culture conditions with the *S. caverna* and three separate co-cultivated species, strongly suggesting that *S. caverna* was likely responsible for the production of funisamine. Another prioritized and isolable metabolite was observed to be entirely unique to BS and *S. caverna* mixed cultures

(Figure B-45). This feature was isolated and identified through NMR spectroscopy to be the antibiotic, amicoumacin B, which is known to be produced by *Bacillus* strains and appears to be produced in response to *S. caverna*. To test if amicoumacin induced funisamine production, we added purified amicoumacin to monocultures of *S. caverna* and did not detect funisamine production (data not shown).

Investigating funisamine biosynthesis

To understand the biosynthesis of funisamine, we endeavored to identify the cognate BGC within *S. caverna*. The genome of *S. caverna* was determined using a combination of Pacific Biosciences RSII and Illumina sequencing to reveal a chromosome of 11,340,955 bps. Potential SM gene clusters were then identified using the antibiotics and SM analysis shell algorithm (antiSMASH, v. 4).[33] This analysis identified 22 putative SM producing gene clusters (Table B-6), including: 6 terpenes, 5 nonribosomal peptide synthetases, 3 lantipeptides, 2 bacteriocins, 1 lasso peptide, and 3 PKS gene clusters. One of these PKS gene clusters, a type I PKS consisting of 26 modules (GenBank accession number: MH128143), contained modular organization of catalytic domains consistent with the determined structure of funisamine. The modular PKS diagram of this cluster is shown in Figure 4-4, and gene descriptions as well as a genetic organization are provided in the supplementary material (Table B-7). While the stereochemistry of substituent hydroxyl and methyl groups have not been experimentally validated, analysis of the ketoreductase (KR) domains within each module (Figure B-47) facilitated the estimation of the stereochemistry of the hydroxyl and methyl substituents in the structure based on the amino acids present in six distinguishing positions along the loop and catalytic regions of the KR as described by Keatinge-Clay.[34] Predicted stereochemical configurations of products for each polyketide synthase module in funisamine biosynthesis are shown in Figure B-48. Methyl groups at positions C14, C30, and C32 were consistent with antiSMASH methylmalonyl extender unit predictions based on acyltransferase (AT) sequences for modules at those positions. The AT for the module predicted to activate the precursor

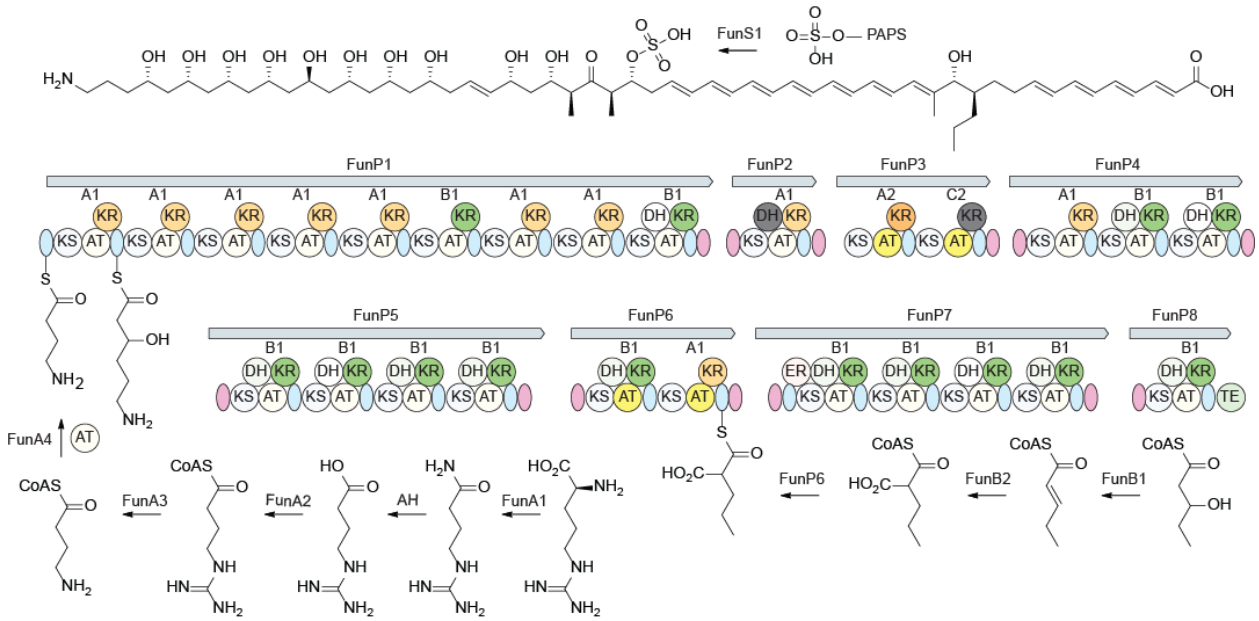


Figure 4-4 Proposed biosynthesis of funisamine. Biosynthesis of the 4-aminobutanoyl-CoA starter unit from arginine and loading of the polyketide synthase shown on the bottom left. Propylmalonyl-CoA biosynthesis via beta-oxidation of odd-chain fatty acid pentanoate shown on the bottom right. Stereochemical assignment of funisamine structure shown at the top was inferred based on ketoreductase type and geometry indicated as A1, A2, B1, B2, C1, or C2, as previously described. Sulfation by putative sulfotransferase FunS1 and PAPS cofactor shown at the top. The modular organization of funisamine polyketide synthase genes (FunP1-FunP8) in *S. caverna* is shown in the center. The layout of catalytic domains: ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) present within the polyketide synthases are shown with inactive domains colored grey. A-type and B-type KR domains are colored orange and green respectively. Acyltransferase domains predicted to use methyl or propylmalonyl extender units are colored yellow. Acyl carrier proteins are colored light blue, and docking domains are purple. Stereochemical configurations of substituent groups and double bonds inferred from bioinformatic analysis. Resubmitted from *ASM* with permissions from the American Society for Microbiology

at position C12 was predicted to use a modified malonyl (ethylmalonyl) CoA unit for polyketide extension, consistent with the propyl group at this position.

The proposed 4-aminobutyryl-CoA starter precursor unit is likely derived from arginine through the unusual pathway described by Leadlay and coworkers (Figure 4-4).[35] Genes encoding the essential arginine monooxygenase (*funA1*), 4-guanidinylbutanoate-CoA ligase (*funA2*), agmatinase (*funA3*), and 4-aminobutyryl-CoA-ACP acyltransferase (*funA4*) were all evident within the gene cluster. A putative sulfotransferase, encoded by (*funS1*), likely catalyzes the sulfation at C29. FunS1 is similar yet distinct in

sequence to other observed sulfotransferase enzymes involved in the biosynthesis of aminopolyol polyketides, such as MedB (BAW35627.1, 47 % identical, 60 % similar) and SMALA2697 (WP_099013767.1, 44 % identical, 59 % similar) which facilitate the sulfation of mediomycin and clethramycin. Production of the evident propyl-malonyl-CoA unit added at this position most likely derives from a 3-hydroxypentyl-CoA by dehydrogenase and carboxylase enzymes encoded by *funB1* and *funB2* observed within the cluster. Expanded versions of Figure 4-4 showing all thioester intermediates are included in Appendix B (Figure B-49 - Figure B-51).

Discussion

Despite the oligotrophic, relatively cold, and aphotic nature of hypogean ecosystems, caves have been demonstrated to harbor rich, stable, and diverse communities of microorganisms. Metagenomic[36-37] and metatranscriptomic[38] studies indicate that hypogean bacteria are primarily heterotrophic but include some autotrophic bacteria, which may be more represented in regions more remote from allochthonous energy inputs. Actinobacteria have been noted as one of the most abundant phyla associated with hypogean ecosystems,[39] have been found associated with speleothems,[40] and constitute some of the first studied hypogean microbial families due to their association with degradation of prehistoric cave paintings.[41] Herein we confirm that actinomycetes are abundant in the three oligotrophic cave locations we sampled using standard actinomycete protocols, and that the actinobacterial generic distribution we observed mirrors what is found on epigean ecosystems.

We selected 20 diverse hypogean actinobacteria for analysis of their secondary metabolism using a stimulus-response paradigm in a single growth medium. The organism set was biased for less commonly described actinobacterial genera including several from which few or no SMs have been previously reported. Overall response data from stimulated culture extracts indicate that stimuli exposure, even for stimuli within the same class, leads to distinct changes in the metabolic inventories of exposed strains.

This global trend is also evident within the response of identified and putative NP features, which largely demonstrate responsiveness to a specific stimulus rather than to a class of stimuli. Previous studies have indicated that the production of some classes of SMs may be induced by general bacterial stressors such as oxidative stress[42] and ppGpp[43-44] or by master regulators such as ScmR in *Burkholderia*.[45] However, the trends observed within our dataset, particularly the variations in congener abundance across stimuli, suggest that the increased abundance of NPs under stimulated conditions stem largely from results of specific interactions with a given stimulus rather than a general response to stress.

Analysis of abundant and covarying metabolites along with manual inspections of HPLC-MS chromatograms identified ~58 putative SMs from an estimated 13 classes based upon MS and UV analysis, spectroscopic similarity to validated congeners isolated in producing strains, and the matching of productive response of putative metabolites to multiplexed stimuli conditions. The isolation of a subset of SMs from these candidates for structural identification was biased towards abundant compounds found in more isolable chromatographic regions with regard to reversed phase HPLC techniques. In all, the subset of isolated compounds revealed 12 identified SMs within eight classes of compounds, seven of which have been previously isolated from epigean ecosystems. One, funisamine, represents a new SM based on its carbon framework, though it bears similarity to a class of giant linear polyenes isolated from other actinobacteria.[46-48] This subset of isolated structurally validated compounds revealed a moderate NP rediscovery rate, with one of eight compound families representing a new carbon scaffold. However, we note that focusing on abundant chromatographically tractable compounds may have biased our campaign towards rediscovery, and that the products of a significant portion of putative SMs identified using this workflow remain to be isolated and characterized.

The conservation of core structural elements within the aminopolyol class of SMs suggests they serve important functions for producing strains within their natural environment and the evolution[49] and ecological role[50] of aminopolyol NPs have recently been investigated. This class of NPs commonly exhibit

antibiotic and/or antifungal activity. The linearmycins were found to inhibit the growth of several gram-positive bacteria and lyse *Bacillus* species by directly targeting the cytoplasmic membrane.[51] A role in vesical biogenesis within producing *Streptomyces* has also been suggested for linearmycins.[50] To our knowledge, funisamine is the only known long chain aminopolyol produced by a *Streptosporangium*, and phylogenetic analyses reveal that many of the ketosynthase (KS) and acyltransferase (AT) sequences within the funisamine PKS are divergent from domains observed in *Streptomyces*-derived aminopolyol gene clusters. Thus, while funisamine is structurally similar to other known aminopolyols, low sequence similarity of the funisamine gene cluster from *S. caverna* to *Streptomyces* gene clusters suggests a substantial amount of evolutionary distance.

Genome mining methods have empowered the development of a variety of complementary methods to expedite the labor-intensive process of NP discovery. For example, native and heterologous expression technologies, using BGCs obtained from genetically 'interesting' producers or environmental DNA, identify SMs via comparative analysis between an expressing strain and the non-expressing host strain. Herein, we demonstrate that comparisons of metabolite responses to multiple stimuli conditions can highlight low abundance SMs, in this case across a wide variety of actinobacterial genera. In comparison to bioactivity-based approaches, the advantage of comparative metabolomics approaches is that they are not constrained by specificity of a given bioassay and permit the discovery of compounds active against a broader array of targets. Indeed, nearly all of the compounds of this study identified solely via comparative metabolomics are reported to possess a variety of useful biological activities.

However, one disadvantage of prioritization of metabolites by comparative metabolomics is that it is a 'shoot first, ask questions later' proposition. There is no guarantee that compounds identified possess therapeutically relevant biological activity. This study highlights a critical need for prioritization in discovery. In addition to the discovery of a new scaffold, within the known compound classes we identified herein, we observed stimulus-dependent amplification of multiple individual congeners of compounds

that were previously not described. These were not prioritized for isolation based on comparative metabolomics covariance ranking as they represented analogs of known compounds, and we had no knowledge of beneficial activity. Thus, a potential future avenue for these techniques is to combine bioactivity measurements with comparative metabolomics data sets. Examples of this combined approach include overlaying bioactivity to molecular networking [52] and our own multiplexed activity metabolomics (MAM) approach.[53] In both cases, these methods enable estimation of activity of unknown compounds and structure-activity relationships within compound families prior to compound isolation using activity metabolomics techniques.[52-53]

Materials and methods

Reagents and strains

All reagents were obtained from the Sigma-Aldrich chemical company unless otherwise specified. Mixed culture strains TP, BS, and EC were obtained from the American Type Culture Collection (ATCC 700081, ATCC 23857, and ATCC 10536), and RW and BK were obtained via dilution plating from hypogean sediments (as described below) and taxonomically assigned by 16S rDNA analysis.

Cave strain isolation and identification

Desiccated cave sediments from Hardin's (Ashland City), Snail Shell (Rockvale), and Cagle's Chasm (South Pittsburg) caves in Tennessee collected between 2010 and 2015 were vortexed in sterile water (100 mg/mL), and supernatants were serially diluted (10-100-1000 fold) and plated on minimal media agar plates (ISP2 medium /100-fold diluted, 1.5 % agar). Plates were then incubated at 30 °C, and colonies were picked over a 3-week period. DNA isolations for purified colonies were performed with a commercial kit (Wizard DNA isolation kit, Promega Inc.). The 16S rRNA genes for these were then amplified with universal

primers 27F (AGA GTT TGA TCC TGG CTC AG) and 1525R (AAG GAG GTG ATC CAG CCG CA) and the use of high fidelity DNA polymerase (Phusion, Thermo Inc). The PCR thermocycler conditions were: (1) initial incubation at 98 °C for 1 min and 30 cycles of (2) denaturation at 98 °C for 0.5 min, (3) annealing at 59 °C for 1 min, and (4) extension at 72 °C for 1.5 min. The final extension step was conducted at 72 °C for 10 min. Target 16S rRNA amplicons were purified using a gel extraction kit (QIAquick gel extraction kit). Purified PCR products were directly sequenced from both the 5′ and 3′ ends, and these sequences were combined to give the near full-length 16S rDNA sequences. Preliminary genus identifications were determined by comparison of 16S rDNA sequences to type strains in the EZBioCloud database (30).

Microbial stimulation and culture conditions

For all conditions, a spore suspension of the selected strain was inoculated on ISP2 agar plates from frozen stocks and cultured for 7 days at 30 °C. These plates were then used to inoculate 25 mL (in 250 mL Erlenmeyer flasks) ISP2 liquid seed cultures which were also cultured for 7 days at 30 °C on a rotary shaker at 180 rpm. For sub-inhibitory antibiotic exposure fermentations, 1.25 mL of the seed culture was transferred into 25 mL (5 % inoculum) ISP2 media experimental cultures containing either rifampicin (120 nM) or streptomycin (170 nM). For rare earth metal fermentations, 1.25 mL of the seed culture was transferred into 25 mL (5 % inoculum) ISP2 media experimental cultures containing either scandium chloride (200 µM) or lanthanum chloride (1500 µM). For mixed culture fermentations, 1.25 mL of the seed culture was transferred into 25 mL (5 % inoculum) ISP2 media experimental cultures. After 24 hrs, 250 µL (1 %) of either TP or RW overnight cultures were added to the actinomycete culture. The controls were prepared by inoculating 1.25 mL (5 %) of seed cultures into 25 mL of ISP2 without additives. All cultures were allowed to ferment at 30 °C, 180 rpm for 7 days before extraction.

Extraction of liquid fermentations

Whole culture metabolite extracts were generated by adding equal volume of methanol (25 mL) to each culture. These were then shaken on a rotary shaker for 1 h before the extractions were centrifuged at 3750 rpm. Supernatants were dried *in vacuo* (Genevac) to yield crude extracts.

LCMS Data Acquisition and Processing

Extract samples were resuspended in 50:50 methanol:water at a concentration of 200 mg/mL and centrifuged at 13,000 rpm and the supernatant was transferred to a fresh vial to remove insoluble extract components. LC-MS data acquisition was performed with a 30-min gradient. Mobile phase A consisted of 95 % water and 5 % acetonitrile with 10 mM ammonium acetate, and mobile phase B consisted of 95 % acetonitrile and 5 % water with 10 mM ammonium acetate. A Luna 4.6 × 250 mm 5 µm particle C18 100 A column (Phenomenex) was used for chromatographic separations with a flow rate of 1 mL/min and a column temperature of 25 °C. An autosampler with a loop size of 20 µL was used for sample injection. The solvent composition began at 100 % A, which was held for 1 min and ramped linearly to 100 % B over the next 29 min, held at 100 % B for 15 min, and returned to 100 % A over a 1 min period. The gradient was held at 100 % A for the next 10 min for equilibration. The flow was split 3:1 using a flow splitter, with 750 µL/min directed through a Surveyor PDA Plus detector and 250 µL/min diverted to the mass spectrometer. Mass spectra were acquired at a rate of 1 Hz from 150–2000 Da in both positive and negative ion mode for the duration of each sample analysis on a TSQ H-ESI mass spectrometer (Thermo scientific). The source capillary was held at 350 °C and 3.0 kV, with a desolvation gas flow of 35 L/h with a vaporizer temperature of 300 °C. Data were converted to mzXML format using the msconvert tool from the ProteoWizard package (31). Peak picking and alignment were performed using XCMS in R (21). See supplementary section A.1 for details and package locations. The resulting data matrices were formatted in Microsoft Excel. Data were filtered by abundance to remove features with a maximal intensity of less than 1 x 10^5

and features eluting after the 30 min gradient (i.e. wash) were removed. TIC for each injection was normalized to a total value of 10,000 such that feature abundance values of 100 represent 1 % of the total abundance for that injection. Principal component and S-plot analyses were performed using Umetrics (Waters Corp.), and formatting for Umetrics was performed by transposing the XCMS generated feature intensity matrix into the Umetrics software. Volcano plots were generated in Excel using p-values determined by the pairwise XCMS analysis of LC-MS data of selected stimulus conditions versus runs with ISP2 medium without stimulus. Fold change calculations for each feature were performed in Excel after data normalization.

Funisamine Isolation

A spore suspension of *S. caverna* was inoculated on ISP2 agar plates from frozen stocks and cultured for 7 days at 30 °C. Plate cultures were then used to inoculate 25 mL ISP2 liquid seed cultures which were also cultured for 7 days (30 °C, 180 rpm). These were then used to inoculate 50 mL (250 mL Erlenmeyer flasks) production cultures (5 % inoculum) in ISP2 media totaling 2 L. After 24 hrs, 500 µL (1 % inoculum) of overnight *B. subtilis* culture were added to the production culture. After 7 days the cultures were combined and extracted with Diaion HP-20 resin (50 g/L) at 30 °C for 3 h to allow for compound partition within the resin. Resin and mycelia were then isolated via centrifugation and extracted with methanol for 1 hr at room temperature. Methanol extract was separated from resin and mycelia via centrifugation and concentrated in vacuo. Funisamine was then purified from this concentrated extract by following its m/z and UV elution through (1) size exclusion chromatography on Sephadex LH-20 column (2 cm dia, x 100 cm length) in methanol (2) preparative C-18 HPLC with a linear gradient of water/acetonitrile buffered with 10 mM ammonium acetate (3) Sephadex LH-20 column (2 cm dia, 50 cm length) in methanol to yield 3 mg of pure funisamine.

High Resolution Mass Spectrometry

Mass accuracy and fragmentation measurements were performed using an electrospray 15T solariX Fourier Transform Ion Cyclotron Resonance mass spectrometer (Bruker Daltonics, Billerica, MA, USA). External mass calibration was performed prior to analysis (ESI-L Tuning Mix, Agilent Technologies, Santa Clara, CA). Funisamine was detected in negative mode as the doubly ($[M-2H]^{2-}$: *m/z* 587.8220, Resolving Power: 497,566) and singly charged ion ($[M-H]^{-}$: *m/z* 1176.6507, Resolving Power: 250,639) by ESI. The calculated *m/z* for $C_{62}H_{98}NO_{18}S^{-}$ was 1176.6510, yielding an error of 0.25 ppm. For fragmentation experiments, the singly charged ion was isolated in the source region of the instrument (quadrupole isolation window: 5.0 Da), accumulated in the collisional hexapole (20 s), and fragmented by sustained off-resonance irradiation collision induced dissociation (SORI-CID) in the ICR cell using pulsed argon (pulse length: 0.25 s, frequency offset: 500 Hz, SORI Power: 2.5%).

Genome Sequencing of Streptosporangium caverna

High molecular weight genomic DNA was isolated from *Streptosporangium caverna* using cetyltrimethylammonium bromide.[54] Whole genome sequencing was performed using a PacBio RS II platform with >40X coverage by Novogene Inc. followed by short-read Illumina sequencing by the Vanderbilt Vantage sequencing core facility. Illumina reads were trimmed using FaQCs to remove Illumina adapters and trim at minimum quality.[55] PacBio sequences and trimmed Illumina reads were then assembled using SPAdes genome assembler v3.11.1[56] using default settings and the careful flag. Resulting assembled scaffolds were polished by using the same trimmed Illumina reads using Pilon version 1.22[57] with default settings. The Bam file required for pilon was generated using BWA mem.[58]

Acknowledgements

This work was supported by the National Institutes of Health (RO1 GM092218, T32 GM 0650086, and 065086-13), the National Science Foundation (NSF-GRFP No. 1445197), the Vanderbilt Department of Chemistry, the Vanderbilt Institute of Chemical Biology. J.M.S. is supported by the National Institutes of Health (P41 GM103391-07, PI - Caprioli, RM) and the Defense Advanced Research Projects Agency (W911NF-14-2-0022, PI-Caprioli, RM). The 15T MALDI FT-ICR mass spectrometer was acquired through the National Institutes of Health Shared Instrumentation Grant Program (1S10OD012359-01). The Vanderbilt VANTAGE Core provided technical assistance for this work. VANTAGE is supported in part by CTSA Grant (5UL1 RR024975-03), the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126), and NIH/NCRR (G20 RR030956).

References

1. Scherlach, K.; Hertweck, C., Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009,** *7* (9), 1753-60.

2. Nett, M.; Ikeda, H.; Moore, B. S., Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **2009,** *26* (11), 1362-1384.

3. Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A., Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002,** *3* (7), 619-627.

4. Chanana, S.; Thomas, C.; Braun, D.; Hou, Y.; Wyche, T.; Bugni, T., Natural Product Discovery Using Planes of Principal Component Analysis in R (PoPCAR). *Metabolites* **2017,** *7* (3), 34.

5. Derewacz, D. K.; Goodwin, C. R.; McNees, R. C.; McLean, J. A.; Bachmann, B. O., Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (6), 2336-41.

6. Goodwin, C. R.; Covington, B. C.; Derewacz, D. K.; McNees, C. R.; Wikswo, J. P.; McLean, J. A.; Bachmann, B. O., Structuring Microbial Metabolic Responses to Multiplexed Stimuli via Self-Organizing Metabolomics Maps. *Chem. Biol.* **2015,** *22* (5), 661-670.

7. Hosaka, T.; Ohnishi-Kameyama, M.; Muramatsu, H.; Murakami, K.; Tsurumi, Y.; Kodani, S.; Yoshida, M.; Fujie, A.; Ochi, K., Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **2009,** *27* (5), 462-4.

8. Wu, C.; Du, C.; Gubbens, J.; Choi, Y. H., Metabolomics-Driven Discovery of a Prenylated Isatin Antibiotic Produced by Streptomyces Species MBT28. *J. Nat. Prod.* **2015,** *78* (10), 2355-63.

9. Xu, F.; Nazari, B.; Moon, K.; Bushin, L. B.; Seyedsayamdost, M. R., Discovery of a Cryptic Antifungal Compound from Streptomyces albus J1074 Using High-Throughput Elicitor Screens. *J. AM. CHEM. SOC.* **2017,** *139* (27), 9203-9212.

10. Onaka, H.; Mori, Y.; Igarashi, Y.; Furumai, T., Mycolic Acid-Containing Bacteria Induce Natural-Product Biosynthesis in Streptomyces Species. *Appl. Environ. Microbiol.* **2011,** *77* (2), 400-406.

11. Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O., Mapping microbial response metabolomes for induced natural product discovery. *ACS Chem. Biol.* **2015,** *10* (9), 1998-2006.

12. Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S., Coculture of Marine Invertebrate-Associated Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. *ACS Chem Biol* **2017,** *12* (12), 3093-3102.

13. Adnani, N.; Vazquez-Rivera, E.; Adibhatla, S. N.; Ellis, G. A.; Braun, D. R.; Bugni, T. S., Investigation of Interspecies Interactions within Marine Micromonosporaceae Using an Improved Co-Culture Approach. *Mar. Drugs* **2015,** *13* (10), 6082-6098.

14. Kamjam, M.; Sivalingam, P.; Deng, Z.; Hong, K., Deep Sea Actinomycetes and Their Secondary Metabolites. *Front Microbiol.* **2017,** *8* (760).

15. Chávez, R.; Fierro, F.; García-Rico, R. O.; Vaca, I., Filamentous fungi from extreme environments as a promising source of novel bioactive secondary metabolites. *Front Microbiol.* **2015,** *6* (903).

16. Tomczyk-Żak, K.; Zielenkiewicz, U., Microbial Diversity in Caves. *Geomicrobiol J* **2016,** *33* (1), 20-38.

17. Adetutu, E.; Thorpe, K.; Shahsavari, E.; Bourne, S.; Cao, X.; Mazaheri, R.; Kirby, G.; Ball, S., Bacterial community survey of sediments at Naracoorte Caves, Australia. *Int J Speleol* **2012,** *41*, 137-147.

18. Wu, Y.; Tan, L.; Liu, W.; Wang, B.; Wang, J.; Cai, Y.; Lin, X., Profiling bacterial diversity in a limestone cave of the western Loess Plateau of China. *Front. Microbiol.* **2015,** *6* (244).

19. Barton, H. A.; Taylor, M. R.; Pace, N. R., Molecular Phylogenetic Analysis of a Bacterial Community in an Oligotrophic Cave Environment. *Geomicrobiol J* **2004,** *21* (1), 11-20.

20. Barton, H. A.; Luiszer, F., Microbial metabolic structure in a sulfidic cave hot spring: Potential mechanisms of biospeleogenesis. *J Cave Karst Stud* **2005,** *67*, 28-38.

21. Hamaki, T.; Suzuki, M.; Fudou, R.; Jojima, Y.; Kajiura, T.; Tabuchi, A.; Sen, K.; Shibai, H., Isolation of novel bacteria and actinomycetes using soil-extract agar medium. *J BioSci BioEng* **2005,** *99* (5), 485-492.

22. Tiwari, K.; Gupta, R. K., Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.* **2012,** *32* (2), 108-132.

23. Goh, E.-B.; Yim, G.; Tsui, W.; McClure, J.; Surette, M. G.; Davies, J., Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *P. Natl. Acad. Sci. USA* **2002,** *99* (26), 17025-17030.

24. Tanaka, Y.; Hosaka, T.; Ochi, K., Rare earth elements activate the secondary metabolite–biosynthetic gene clusters in Streptomyces coelicolor A3(2). *J. Antibiot.* **2010,** *63* (8), 477-481.

25. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006,** *78* (3), 779-787.

26. Yagi, A.; Makino, K.; Nishioka, I., Studies on the Constituents of Aloe sapnaria HAW. I. The Structures of Tetrahydroanthracene Derivatives and the Related Anthraquinones. *Chem. Pharm. Bull.* **1974,** *22* (5), 1159-1166.

27. Cui, H.; Shaaban, K. A.; Qin, S., Two Anthraquinone Compounds from a Marine Actinomycete Isolate M097 Isolated from Jiaozhou Bay. *World. J. Microb. Biot.* **2006,** *22* (12), 1377-1379.

28. Abdalla, M. A.; Win, H. Y.; Islam, M. T.; von Tiedemann, A.; Schuffler, A.; Laatsch, H., Khatmiamycin, a motility inhibitor and zoosporicide against the grapevine downy mildew pathogen Plasmopara viticola from Streptomyces sp. ANK313. *J antibiot (Tokyo)* **2011,** *64* (10), 655-659.

29. Funayama, S.; Ishibashi, M.; Komiyama, K.; Omura, S., A new antibiotic, okicenone. *J Antibiot (Tokyo).* **1991,** *44* (8), 819-823.

30. Derewacz, D. K.; McNees, C. R.; Scalmani, G.; Covington, C. L.; Shanmugam, G.; Marnett, L. J.; Polavarapu, P. L.; Bachmann, B. O., Structure and Stereochemical Determination of Hypogeamicins from a Cave-Derived Actinomycete. *J. Nat. Prod.* **2014,** *77* (8), 1759-1763.

31. Kallifidas, D.; Kang, H.; Brady, S. F., Tetarimycin A, an MRSA-Active Antibiotic Identified through Induced Expression of Environmental DNA Gene Clusters. *J. AM. CHEM. SOC.* **2012,** *134* (48), 19552-19555.

32. Stubbendieck, R. M.; Straight, P. D., Linearmycins Activate a Two-Component Signaling System Involved in Bacterial Competition and Biofilm Morphology. *J Bacteriol* **2017,** *199* (18).

33. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* **2017,** *45* (W1), W36-w41.

34. Keatinge-Clay, A. T., A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **2007,** *14* (8), 898-908.

35. Hong, H.; Fill, T.; Leadlay, P. F., A Common Origin for Guanidinobutanoate Starter Units in Antifungal Natural Products. *Angew Chem Int Ed.* **2013,** *52* (49), 13096-13099.

36. De Mandal, S.; Chatterjee, R.; Kumar, N. S., Dominant bacterial phyla in caves and their predicted functional roles in C and N cycle. *BMC Microbiology* **2017,** *17* (1), 90.

37. D'Auria, G.; Artacho, A.; Rojas, R. A.; Bautista, J. S.; Méndez, R.; Gamboa, M. T.; Gamboa, J. R.; Gómez-Cruz, R., Metagenomics of Bacterial Diversity in Villa Luz Caves with Sulfur Water Springs. *Genes* **2018,** *9* (1), 55.

38. Portillo, M. C.; Gonzalez, J. M.; Saiz-Jimenez, C., Metabolically active microbial communities of yellow and grey colonizations on the walls of Altamira Cave, Spain. *J Appl Microbiol* **2008,** *104* (3), 681-91.

39. Oliveira, C.; Gunderman, L.; Coles, C. A.; Lochmann, J.; Parks, M.; Ballard, E.; Glazko, G.; Rahmatallah, Y.; Tackett, A. J.; Thomas, D. J., 16S rRNA Gene-Based Metagenomic Analysis of Ozark Cave Bacteria. *Diversity* **2017,** *9* (3), 31.

40. Adam, D.; Maciejewska, M.; Naômé, A.; Martinet, L.; Coppieters, W.; Karim, L.; Baurain, D.; Rigali, S., Isolation, Characterization, and Antibacterial Activity of Hard-to-Culture Actinobacteria from Cave Moonmilk Deposits. *Antibiotics* **2018,** *7* (2), 28.

41. Gonzalez, J.; Portillo, M.; Saiz-Jimenez, C., Microbes Pose a Risk to Prehistoric Cave Paintings. *Microbe* **2008,** *3*.

42. Hong, S. Y.; Roze, L. V.; Linz, J. E., Oxidative stress-related transcription factors in the regulation of secondary metabolism. *Toxins (Basel)* **2013,** *5* (4), 683-702.

43. Ochi, K., Metabolic initiation of differentiation and secondary metabolism by Streptomyces griseus: significance of the stringent response (ppGpp) and GTP content in relation to A factor. *J Bacteriol* **1987,** *169* (8), 3608-3616.

44. Kawai, K.; Wang, G.; Okamoto, S.; Ochi, K., The rare earth, scandium, causes antibiotic overproduction in Streptomyces spp. *FEMS Microbiol. Lett.* **2007,** *274* (2), 311-315.

45. Mao, D.; Bushin, L. B.; Moon, K.; Wu, Y.; Seyedsayamdost, M. R., Discovery of scmR as a global regulator of secondary metabolism and virulence in Burkholderia thailandensis E264. *Proc Natl Acad Sci U S A* **2017,** *114* (14), E2920-E2928.

46. Sakuda, S.; Guce-Bigol, U.; Itoh, M.; Nishimura, T.; Yamada, Y., Linearmycin A, a novel linear polyene antibiotic. *Tetrahedron Lett.* **1995,** *36* (16), 2777-2780.

47. Igarashi, Y.; Iwashita, T.; Fujita, T.; Naoki, H.; Yamakawa, T.; Yoshida, R.; Furuma, T., Clethramycin, a new inhibitor of pollen tube growth with antifungal activity from Streptomyces hygroscopicus TP-A0623. II. Physico-chemical properties and structure determination. *J antibiot (Tokyo)* **2003,** *56* (8), 705-8.

48. McAlpine, J. B.; Bachmann, B. O.; Piraee, M.; Tremblay, S.; Alarco, A. M.; Zazopoulos, E.; Farnet, C. M., Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J Nat Prod* **2005,** *68* (4), 493-6.

49. Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin-ya, K.; Abe, I., Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem. Int. Edit.* **2017,** *56* (7), 1740-1745.

50. Hoefler, B. C.; Stubbendieck, R. M.; Josyula, N. K.; Moisan, S. M.; Schulze, E. M.; Straight, P. D., A Link between Linearmycin Biosynthesis and Extracellular Vesicle Genesis Connects Specialized Metabolism and Bacterial Membrane Physiology. *Cell Chem Biol* **2017,** *24* (10), 1238-1249.e7.

51. Stubbendieck, R. M.; Brock, D. J.; Pellois, J.-P.; Gill, J. J.; Straight, P. D., Linearmycins are lytic membrane-targeting antibiotics. *J. Antibiot.* **2018,** *71* (3), 372-381.

52. Nothias, L.-F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C., Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J Nat Prod* **2018,** *81* (4), 758-767.

53. Earl, D. C.; Ferrell, P. B.; Leelatian, N.; Froese, J. T.; Reisman, B. J.; Irish, J. M.; Bachmann, B. O., Discovery of human cell selective effector molecules using single cell multiplexed activity metabolomics. *Nat. Commun.* **2018,** *9* (1), 39.

54. Kieser, T.; Buttner, M.; Hopwood, K. C. D., Practical Streptomyces Genetics. **2000**, 162-170.

55. Lo, C.-C.; Chain, P. S. G., Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC bioinformatics* **2014,** *15* (1), 366.

56. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012,** *19* (5), 455-77.

57. Walker, B. J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C. A.; Zeng, Q.; Wortman, J.; Young, S. K.; Earl, A. M., Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **2014,** *9* (11), e112963.

58. Li, H.; Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009,** *25* (14), 1754-60.

Chapter 5 Dissertation summary and future directions

Dissertation Summary

Microbial SMs are often structurally elaborate molecules that have been in a sense 'engineered' over billions of years by ever evolving microorganisms through competition in complex communities. While the ecological roles of most SM classes are uncertain and sometimes debated, the enhanced fitness many of these compounds provide likely stem from direct interactions with macromolecules (DNA, RNA, protein, etc.) from other environmental co-inhabitants. These highly selective interactions predispose SMs for therapeutic use, and their therapeutic potential is evident in the history of successful SM, or SM derived drug compounds.[1-4] Despite these successes, many drug companies moved away from NPs in favor of synthetic molecules due in part to poor scalability, challenging molecular modifications, and increasing difficulties with novel SM discovery. While synthetic small molecules are certainly more easily produced and modified, they lack much of the structural diversity commonly seen in SM libraries as well as any predisposition for enhanced binding to important molecular targets, and these synthetic libraries have been much less successful at turning out approved therapeutics.[5] Much of the current research in the NP field endeavors to improve methods of SM modification and discovery, and the continuing advancements in these areas, including the works discussed in this dissertation, may ultimately put microbial SMs back in the central focus of drug discovery efforts.

The field of microbial SM discovery has significantly changed since the genome analysis of SC revealed many SM producing BGCs with uncorrelated SMs.[6] This and other analyses of microbial genome sequences led to the realization that the products for a majority of SM producing BGCs are unidentified, and the discrepancies between observed SM potential and realized output are increasing as BGC prediction methods improve.[7] Many classes of BGCs responsible for SM production can now be readily observed within genomic sequence data through automated bioinformatics analyses which can also

provide predictions of the structural framework of emanant SMs.[8-13] Several strategies have been developed to obtain SMs from observed but unproductive (or undetectably productive) BGCs. As described previously, these include stimulating production within the native producer [4, 14], cloning and forcibly expressing the BGCs in a heterologous host,[15-16] and even bypassing the culturing step altogether and chemically synthesizing structures predicted through bioinformatic sequence analyses.[17] In recent years, advancements and applications of these methodologies have enabled the discovery of several novel SMs,[18-20] including some described in this dissertation. The works described herein center on prioritization and discovery of SMs through applications of rationally selected environmental stimuli and comparative metabolomics analyses. We hypothesized that a significant portion of SMs are used to respond to environmental cues and can be prioritized from comparative analyses of stimulated culture extracts. Our works applying established activation stimuli on cultured actinomycetes and analyzing resultant culture extracts through LC-MC based metabolomics analyses support this hypothesis. The majority of readily isolable metabolite features prioritized through these applications have been identified as SMs. We demonstrated that these methods were effective at highlighting SMs from complex culture extracts from a model organism, *Streptomyces coelicolor A3(2)*, through a novel implementation of SOM analytics (Chapter 2). We further demonstrated the effectiveness of comparative analyses of stimulated culture extracts to prioritize SM through our identifications of the ciromicins from NF (Chapter 3), funisamine from *Streptosporangium caverna*, and several known SMs from isolated cave actinomycetes (Chapter 4). Not only have these contributions demonstrated the utility of comparative metabolomics for SM discovery but have also revealed significant metabolomic changes accompanying applications of stimuli. We have shown that stimuli exposure results in largely distinct metabolomic and SM responses even within stimuli of the same apparent class (metals, antibiotics, or mixed culture). These trends suggest that the increased abundance of SMs within stimulated cultures result from specific interactions with a given stimulus and the cultured organism rather than a general response from a stressor as previous studies

have indicated.[21-23] This is an important observation that advocates future research to better understand these interactions between stimuli and SM producing organisms. The apparent selectivity of SM responses suggests that orphan BGCs within microorganisms lacked required small molecule component(s) and that larger arsenals of SM elicitors will be needed to fully access the SM potential of microbial producers. The works presented in this dissertation demonstrate the ability of comparative metabolomics analyses to prioritize SMs from stimulated microbial cultures. Once large arrays of established elicitors are available, these applications should be very effective at accessing much of the currently unavailable SM arsenal in observed microorganisms.

Future directions

Despite the numerous recent advances to potentiate microbial SM discovery there remain many challenges that need to be addressed. In much of the work presented here the genomic sequences of stimulated microorganisms were unknown, such that we are unable to discuss the degree to which our elicitors enabled observation of SMs from otherwise orphan BGCs. We also do not know the mechanisms through which stimuli exposure ultimately resulted in the enhanced SM production. While it is often assumed that many 'orphan' BGCs with no known products are transcriptionally inactive,[24-25] recent studies indicate more downstream regulation or inhibition may be involved. Comparative transcriptomics analyses revealed that the majority of presumed silent BGCs within a marine actinomycete *Salinispora* were transcribed at levels which should have enabled detection of cognate SMs.[26] The lack of detected products from a large portion of microbial BGCs may then stem from more downstream regulation, precursor/cofactor availability, or even instrument limitations (MS ionization efficiencies, chromatographic resolvability, etc.). Going forward, it will be important to understand how the various SM elicitors function to increase SM production. Some of these may enhance BGC transcription or translation of cognate biosynthetic genes, while others may alter the activity of functionally translated

SM biosynthetic complexes. A better understanding of these processes may enable synergistic applications of stimuli that enhance SM production at multiple regulatory levels. More potent SM elicitors or synergistic combinations of applied elicitors may be important for many of the low abundance SMs which may otherwise be only barely detectable against the complex background of a typical culture extract. In the work described in Chapter 4 many prioritized features were not isolable due to their low abundance, and efforts to understand and improve methods of SM elicitation could greatly increase the success of NP discovery from both the endogenous efforts, such as those described herein, as well as efforts involving heterologous expression of targeted BGCs.

Another practical consideration for future work is the therapeutic value of targeted SMs. The NP discovery approach described herein prioritizes SMs by their responses to stimuli and not by biological activity. NP discovery efforts focused on biological activity likely would not have identified funisamine which exhibited little observed activity in human, fungal, or bacterial cell lines. A large portion of NP discovery endeavors have been activity-driven, so activity independent screens may have better chances of identifying novel SMs. In our works identifying compounds of therapeutic values has been secondary to establishing methods to prioritize SMs from culture extracts. However, future endeavors will benefit from the combination of both stimuli-response and bioactivity data to prioritize potential SM leads. The Bachmann lab has recently developed a multiplexed activity metabolomics (MAM) approach which can screen a large number of compounds for multiple markers of differing bioeffectors classes within several distinct cell lineages all in the same experiment.[27] This could potentially be multiplexed further such that all stimulated culture extracts for a given producing organism could be screened at the same time. Other metabolomic methods of NP prioritization have already been shown to benefit from the addition of biological activity data, such as bioactivity-based molecular networking.[28-29] Coupling MAM with our established stimuli-response analysis methods should also potentiate the discovery of novel and therapeutically relevant SMs.

We have demonstrated the power of comparative metabolomics analyses to prioritize SMs from complex extracts of stimulated microbial producers. Future work to (1) increase the effectiveness of SM eliciting stimuli either through the discovery of novel and potent SM elicitors or through the combination of synergistic stimuli and (2) correlate MAM bioeffector and stimuli-response data could result in a powerful pipeline for novel and active SMs. Additionally, these works would even be useful for repurposing known SMs by identifying new biological activities for old NPs. This future research could make significant contributions to reinvigorate NP drug discovery efforts and result in the discovery of much needed, novel therapeutic compounds.

References

1. Bérdy, J., Bioactive microbial metabolites. *J Antibiot (Tokyo).* **2004,** *58* (1), 1-26.

2. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **2007,** *70* (3), 461-477.

3. Jon, C.; Christopher, W., Lessons from natural molecules. *Nature* **2004,** *432*, 829-837.

4. Scherlach, K.; Hertweck, C., Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009,** *7* (9), 1753-60.

5. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016,** *79* (3), 629-661.

6. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.;

Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002,** *417* (6885), 141-7.

7. Cimermancic, P.; Medema, M.; Claesen, J.; Kurita, K.; Wieland, B.; Laura, C.; Mavrommatis, K.; Pati, A.; Godfrey, P.; Koehrsen, M.; Clardy, J.; Birren, B.; Takano, E.; Sali, A.; Linington, R.; Fischbach, M., Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014,** *158* (2), 412-421.

8. Zakrzewski, P.; Fischbach, M. A.; Weber, T., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucl. Acids Res.* **2011,** *39*, 339-46.

9. Li, M. H. T.; Ung, P. M. U.; Zajkowski, J., Automated genome mining for natural products. *BMC bioinformatics* **2009,** *10* (185).

10. Ju, K. S.; Gao, J.; Doroghazi, J. R.; Wang, K. A.; Thibodeaux, C. J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J. K.; Lee, J.; Cioni, J. P.; Evans, B. S.; Hirota, R.; Labeda, D. P.; van der Donk, W. A.; Metcalf, W. W., Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (39), 12175-12180.

11. Donia, M. S.; Cimermancic, P.; Schulze, C. J.; Wieland Brown, L. C.; Martin, J.; Mitreva, M.; Clardy, J.; Linington, R. G.; Fischbach, M. A., A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **2014,** *158* (6), 1402-1414.

12. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucl. Acids Res.* **2015,** *43* (1), 237-43.

13. Jensen, P. R.; Chavarria, K. L.; Fenical, W.; Moore, B. S., Challenges and triumphs to genomics-based natural product discovery. *J. Ind. Microbiol. Biotechnol.* **2014,** *41* (2), 203-9.

14. Bibb, M. J., Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **2005,** *8* (2), 208-15.

15. Komatsu, M.; Uchiyama, T.; mura, S.; Cane, D. E.; Ikeda, H., Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (6), 2646-2651.

16. Kim, E.; Moore, B. S.; Yoon, Y. J., Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **2015,** *11* (9), 649-659.

17. Vila-Farres, X.; Chu, J.; Inoyama, D.; Ternei, M. A.; Lemetre, C.; Cohen, L. J.; Cho, W.; Reddy, B. V. B.; Zebroski, H. A.; Freundlich, J. S.; Perlin, D. S.; Brady, S. F., Antimicrobials Inspired by Nonribosomal Peptide Synthetase Gene Clusters. *J. AM. CHEM. SOC.* **2017,** *139* (4), 1404-1407.

18. Kallifidas, D.; Kang, H.; Brady, S. F., Tetarimycin A, an MRSA-Active Antibiotic Identified through Induced Expression of Environmental DNA Gene Clusters. *J. AM. CHEM. SOC.* **2012,** *134* (48), 19552-19555.

19. Clevenger, K. D.; Bok, J. W.; Ye, R.; Miley, G. P.; Verdan, M. H.; Velk, T.; Chen, C.; Yang, K.; Robey, M. T.; Gao, P.; Lamprecht, M.; Thomas, P. M.; Islam, M. N.; Palmer, J. M.; Wu, C. C.; Keller, N. P.; Kelleher, N. L., A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat Chem Biol* **2017,** *13* (8), 895-901.

20. Henke, M. T.; Soukup, A. A.; Goering, A. W.; McClure, R. A.; Thomson, R. J.; Keller, N. P.; Kelleher, N. L., New Aspercryptins, Lipopeptide Natural Products, Revealed by HDAC Inhibition in Aspergillus nidulans. *ACS Chem Biol* **2016,** *11* (8), 2117-23.

21. Hong, S. Y.; Roze, L. V.; Linz, J. E., Oxidative stress-related transcription factors in the regulation of secondary metabolism. *Toxins (Basel)* **2013,** *5* (4), 683-702.

22. Ochi, K., Metabolic initiation of differentiation and secondary metabolism by Streptomyces griseus: significance of the stringent response (ppGpp) and GTP content in relation to A factor. *J Bacteriol* **1987,** *169* (8), 3608-3616.

23. Kawai, K.; Wang, G.; Okamoto, S.; Ochi, K., The rare earth, scandium, causes antibiotic overproduction in Streptomyces spp. *FEMS Microbiol. Lett.* **2007,** *274* (2), 311-315.

24. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015,** *13*, 509-23.

25. Chiang, Y. M.; Chang, S. L.; Oakley, B. R.; Wang, C. C., Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol* **2011,** *15* (1), 137-43.

26. Amos, G. C. A.; Awakawa, T.; Tuttle, R. N.; Letzel, A.-C.; Kim, M. C.; Kudo, Y.; Fenical, W.; S. Moore, B.; Jensen, P. R., Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. U S A.* **2017,** *114* (52), E11121-E11130.

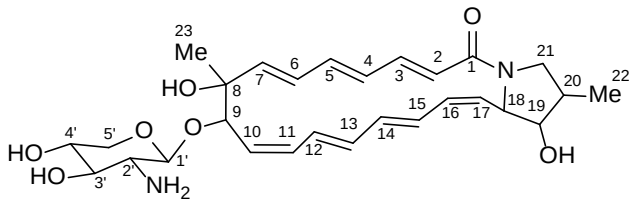
27. Earl, D. C.; Ferrell, P. B.; Leelatian, N.; Froese, J. T.; Reisman, B. J.; Irish, J. M.; Bachmann, B. O., Discovery of human cell selective effector molecules using single cell multiplexed activity metabolomics. *Nat. Commun.* **2018,** *9* (1), 39.

28. Naman, C. B.; Rattan, R.; Nikoulina, S. E.; Lee, J.; Miller, B. W.; Moss, N. A.; Armstrong, L.; Boudreau, P. D.; Debonsi, H. M.; Valeriote, F. A.; Dorrestein, P. C.; Gerwick, W. H., Integrating Molecular Networking and Biological Assays To Target the Isolation of a Cytotoxic Cyclic Octapeptide, Samoamide A, from an American Samoan Marine Cyanobacterium. *J Nat Prod* **2017,** *80* (3), 625-633.

29. Nothias, L.-F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C., Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J Nat Prod* **2018,** *81* (4), 758-767.

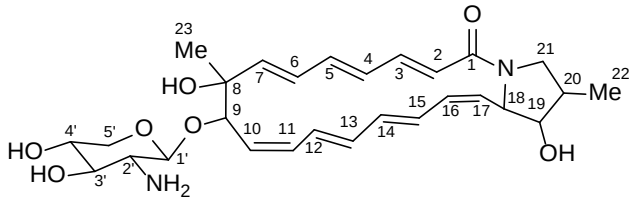
Appendix A Supporting data for Chapter 3

Table A-1 NMR correlation table of ciromicin A (1) in DMSO-d_6



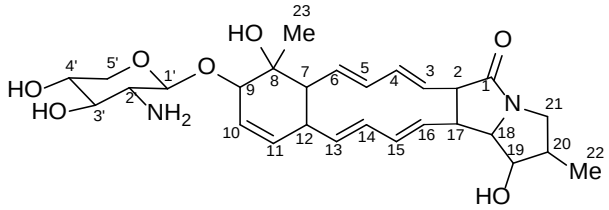
Pos.	^{13}C	^{1}H	HMBC	NOESY
1	165.4			
2	123.9	6.09, 1H, d, (15 Hz)	1, 4	15, 18
3	138.9	6.83, 1H, dd (11, 15 Hz)	1, 5	5
4	128.6	6.04, 1H, dd (10.5, 15 Hz)	6	5
5	135.8	6.26, 1H, overlap	7	3, 4
6	125.1	6.01, 1H, overlap	4, 8	7
7	140.2	5.81, 1H, d (15 Hz)	5, 8, 23	5, 6, 9, 12, 23
8	77			
9	81.2	4.40, 1H, d (9.5 Hz)	10, 11, 1’	7,12
10	133.2	5.41, 1H, dd (9.5, 10 Hz)		11
11	127.7	5.99, 1H, dd (11 Hz)	9	10, 13
12	130.8	6.60, 1H, dd (11, 15 Hz)	14	6 (tiny), 7, 9, 13, 14
13	132.6	6.32, 1H, dd (11, 15 Hz)	11, 14	11, 12 (small), 15
14	135.5	6.51, 1H, dd (11, 15 Hz)	12	12, 16
15	128.4	6.82, 1H, dd (11, 15 Hz)	13	2, 13, 18
16	130.2	6.28, 1H, overlap	14, 18	14, 17
17	129.5	5.43, 1H, dd (10.5, 11 Hz)	15	16, 20
18	57.6	4.86, 1H, dd (7.5, 11 Hz)	1, 16, 19, 20, 21	2, 15, 19
19	77.3	3.77, 1H, dd (7.5, 9 Hz)	17, 20, 22	18, 22
20	36.2	2.12, 1H, m	19, 21, 22	17, 21, 22
21	50.2	2.89, 1H, dd (11, 12 Hz)	20, 22	21, 22
		3.55, 1H, dd (8, 12 Hz)	19, 20	20, 21
22	15.4	1.02, 3H, d (6 Hz)	19, 20, 21	19, 20, 21
23	26.2	1.36, 3H, s	7, 8, 9	7, 9, 1’
1’	106.2	4.14, 1H, d (7.5 Hz)	9, 2’, 3’, 5’	9, 12, 5’
2’	58.2	2.49, 1H, under solvent	1’, 3’	
3’	76.7	2.99, 1H, overlap	2’, 4’, 5’	1’
4’	69.9	3.26, 1H, ddd (5.5, 9, 10.5 Hz)	3’, 5’	3’, 5’
5’	66.4	2.98, 1H, overlap	3’	1’, 5’
		3.66, 1H, dd (6, 11 Hz)	1’, 3’, 4’	4’, 5’

Table A-2 NMR correlation table of ciromicin A (1) in CD_3OD.



Pos.	^{13}C	^{1}H	HMBC	NOESY
1	167.0			
2	122.8	6.21, 1H, d, (15 Hz)	1, 4	3, 18
3	139.6	6.95, 1H, dd (10, 15.2 Hz)	1	5
4	128.9	6.19, 1H, overlap	2, 5, 6	5
5	135.4	6.33, 1H, dd (8, 15 Hz)	3	3, 4
6	126.2	6.14, 1H, overlap	4, 8	7
7	138.0	5.86, 1H, d (15 Hz)	5, 8	6, 9
8	76.8			
9	80.7	4.61, 1H, d (9.5 Hz)	10, 11, 1’	7,12
10	130.4	5.53, 1H, dd (9.5, 10 Hz)		11
11	128.6	6.15, 1H, overlap	9	10, 13
12	130.1	6.58, 1H, dd (11, 15 Hz)	14	9, 11
13	133.0	6.39, 1H, dd (11, 15 Hz)	11, 14	11
14	135.5	6.54, 1H, dd (11, 15 Hz)	12	12, 16
15	127.7	6.84, 1H, dd (11, 15 Hz)	13	13, 14, 18
16	131.0	6.41, 1H, dd (10.5, 11 Hz)	18	15, 17
17	126.8	5.47, 1H, dd (10.5, 11 Hz)	15	16, 18, 20
18	58.3	5.01, 1H, dd (7.5, 11 Hz)	1, 16, 19, 20, 21	2, 15, 17, 19
19	77.1	3.91, 1H, dd (7.5, 9 Hz)	17, 20, 22	18, 22
20	35.7	2.27, 1H, m	19, 21, 22	17, 21, 22
21	49.7	3.03, 1H, dd (11, 12 Hz)	20, 22	20, 21
		3.74, 1H, dd (8, 12 Hz)	19, 20	21
22	13.9	1.15, 3H, d (6 Hz)	19, 20, 21	19, 20
23	24.4	1.44, 3H, s	7, 8, 9	
1’	101.8	4.59, 1H, d (7.5 Hz)	9, 2’, 3’, 5’	
2’	56.6	2.84, 1H, broad	1’, 3’	
3’	73.5	3.42, 1H, broad	2’, 4’, 5’	
4’	69.9	3.53, 1H, broad	3’, 5’	
5’	65.6	3.25, 1H, overlap	3’	
		3.92, 1H, overlap	1’, 3’, 4’	

Table A-3 NMR correlation table of ciromicin B (2) in CD_3OD.



Pos.	^{13}C	^{1}H	HMBC	NOESY
1	175.2			
2	58.4	3.39, 1H, dd, (7.5, 10.5 Hz)	1, 3, 4, 18	4, 17
3	129.1	5.25, 1H, dd (11 Hz)		5, 15
4	136.6	5.90, 1H, dd (10.5, 15 Hz)	2	2
5	138.9	5.88, 1H, dd (10.5 Hz)	7	7
6	129.3	5.32, 1H, dd (10.5, 15 Hz)	4	4, 12, 23
7	53.2	2.54, 1H, br dd (9.6, 11 Hz)	5, 6, 8, 12, 23	5, 13
8	72.4			
9	81.1	3.75, 1H, d (6 Hz)	10, 11, 1’	10, 23, 1’
10	126.7	5.83, 1H, overlap	9	9
11	131.9	5.82, 1H, overlap	6, 7, 12	12, 13
12	49.8	2.71, 1H, br dd (8, 10 Hz)	7, 13, 14	11, 23
13	131.3	5.08, 1H, dd (10, 15.7 Hz)	12, 15	7, 11, 15
14	138.1	5.80, 1H, dd (10.5, 15.7 Hz)	12,15, 16	12, 16
15	132.5	6.04, 1H, dd (10.5, 16.2 Hz)	13, 14, 17	3, 18
16	129.5	5.26, 1H, dd (6, 16.2 Hz)	14, 17, 18	2, 14, 17
17	43.3	3.68, 1H, br ddd	15, 16, 18	16
18	62.7	4.41, 1H, dd (3.7, 9.5 Hz)		3,15,19, 22,
19	73.4	3.77, 1H, dd (1.4, 3.7 Hz)		18, 22
20	43.9	2.49, 1H, m	19, 22	21, 22
21	46.1	2.79, 1H, dd (2.4, 11.7 Hz)	22	21, 22
		3.74, 1H, dd (7.5, 11.7 Hz)	1, 22	20, 21
22	17.1	1.15, 3H, d (7.5 Hz)	19, 20, 21	18, 19, 20, 21
23	18.5	1.22, 3H, s	7, 8, 9	6, 9, 12
1’	105.5	4.47, 1H, d (8.5 Hz)	9	9, 5’
2’	57.2	2.57, 1H, br dd, (8.5, 9 Hz)	1’, 3’	
3’	76.2	3.22, 1H, dd, (9 Hz)	2’, 4’	1’
4’	69.9	3.47, 1H, ddd (5.4, 9, 10 Hz)		2’, 5’
5’	65.8	3.23, 1H, dd (10, 11.5 Hz)		5’
		3.89, 1H, dd (5.4, 11.5 Hz)	3’, 4’	4’, 5’

Figure A-1 ^{1}H proton NMR spectrum of ciromicin A (1) in DMSO-d6

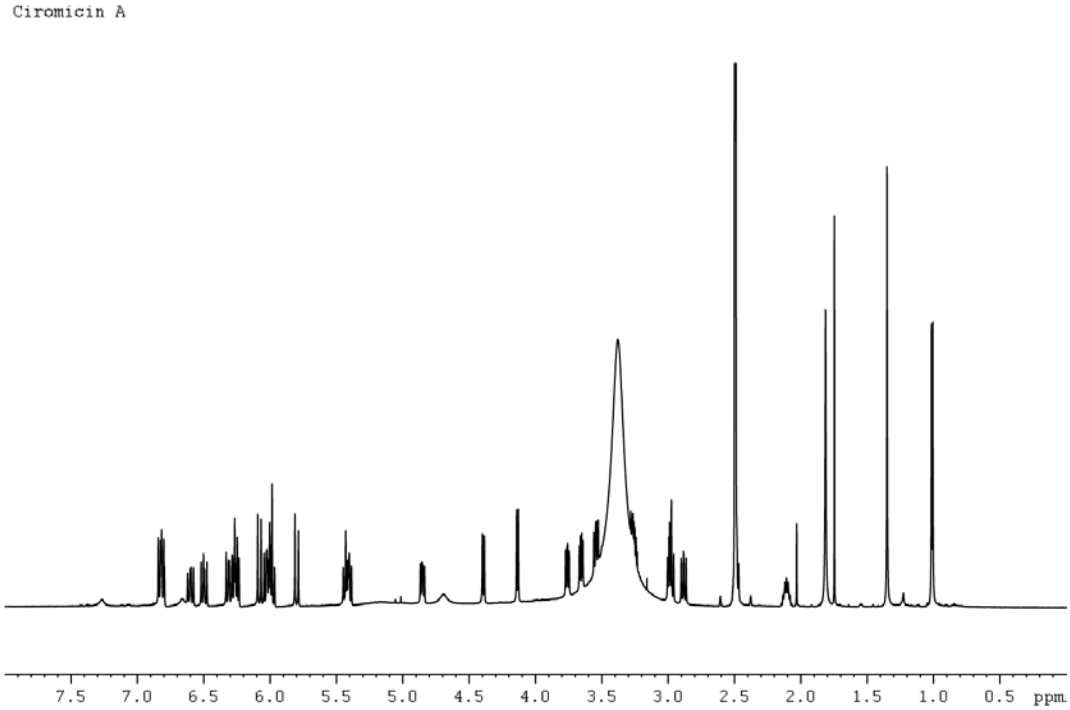


Figure A-2 ^{13}C carbon NMR spectrum of ciromicin A (1) in DMSO-d6.

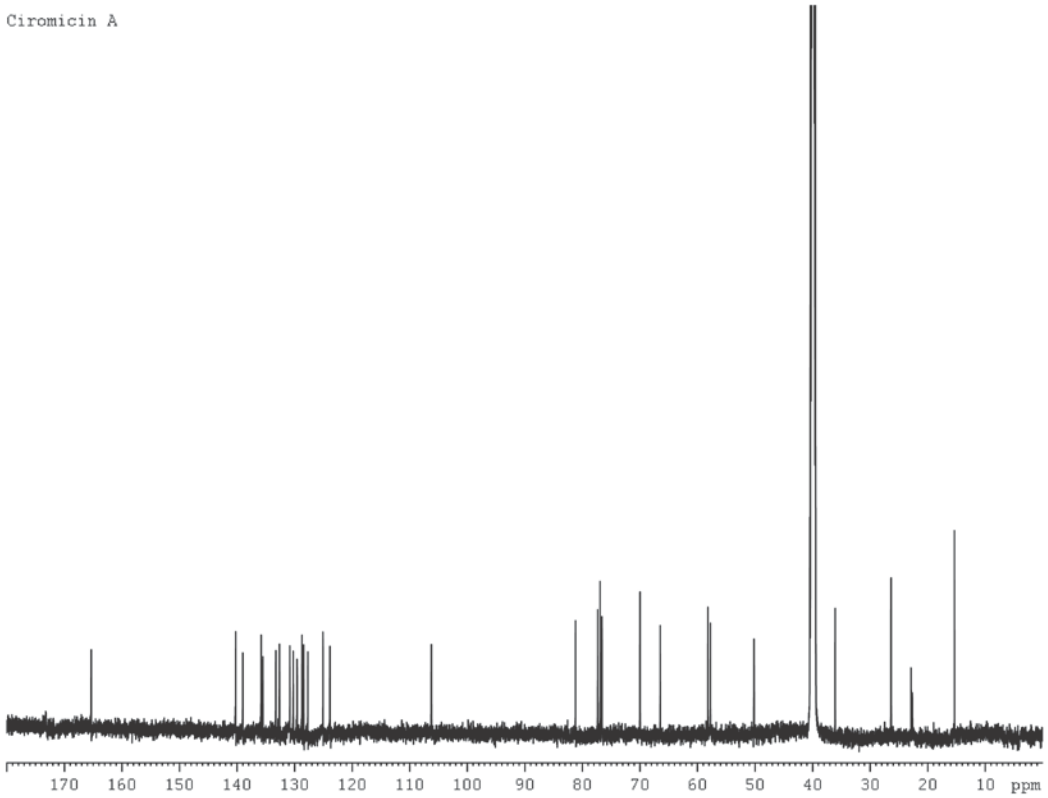


Figure A-3 2D COSY NMR spectrum of ciromicin A (1) in DMSO-d6.

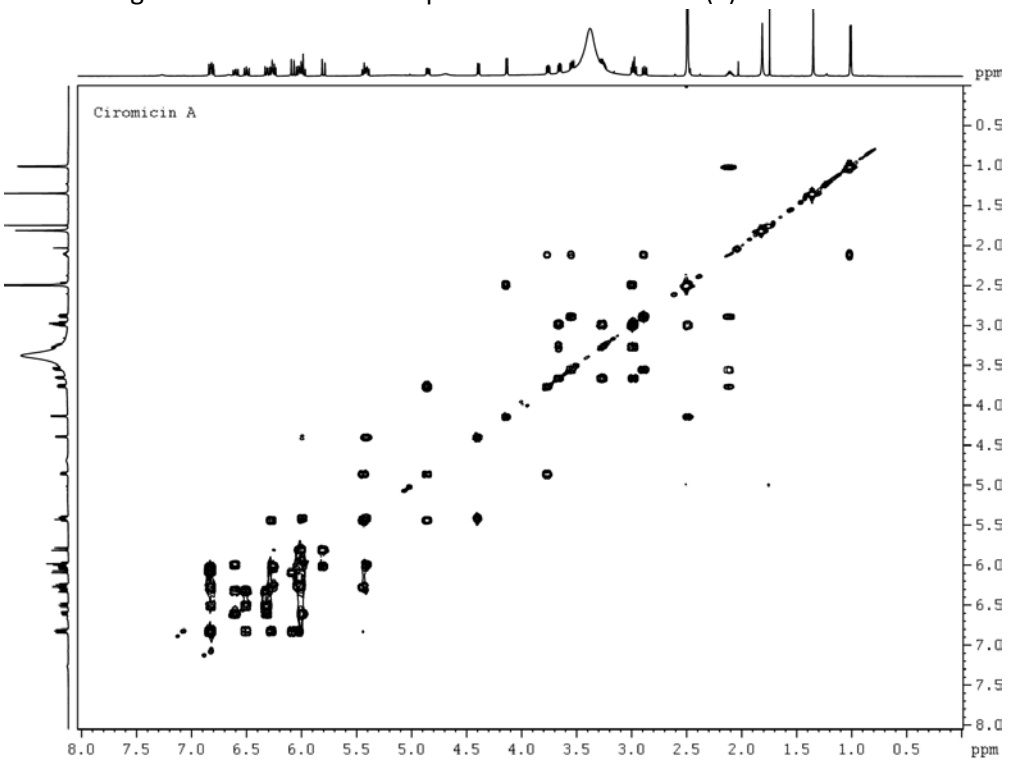


Figure A-4 2D HSQC NMR spectrum of ciromicin A (1) in DMSO-d6.

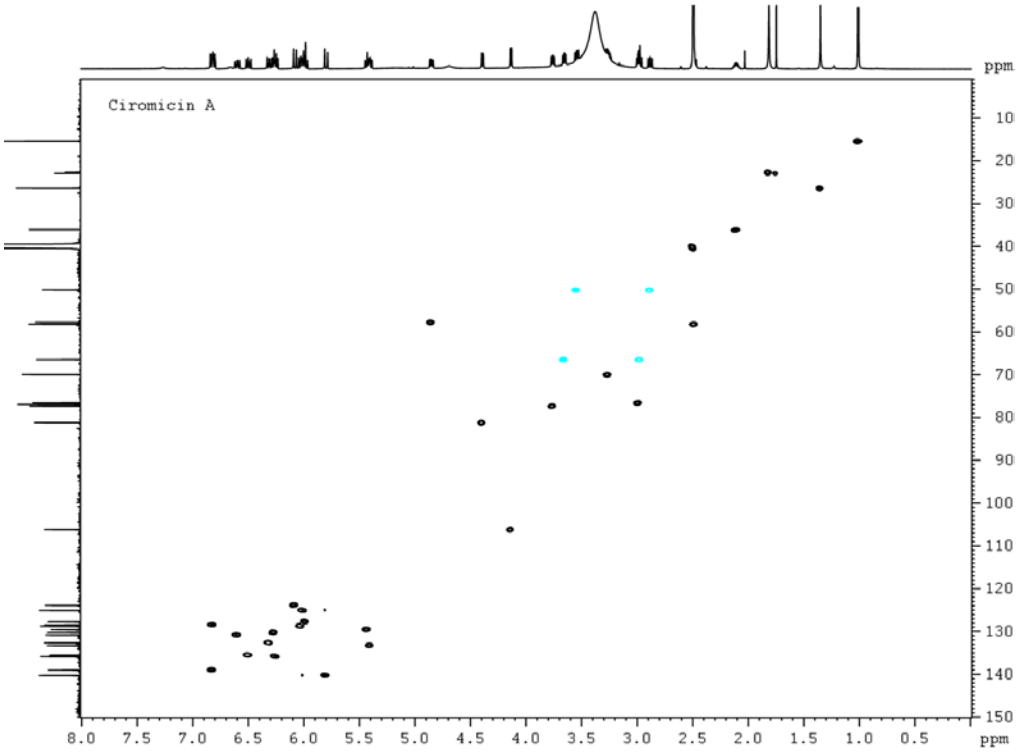


Figure A-5 2D HMBC NMR spectrum of ciromicin A (1) in DMSO-d6.

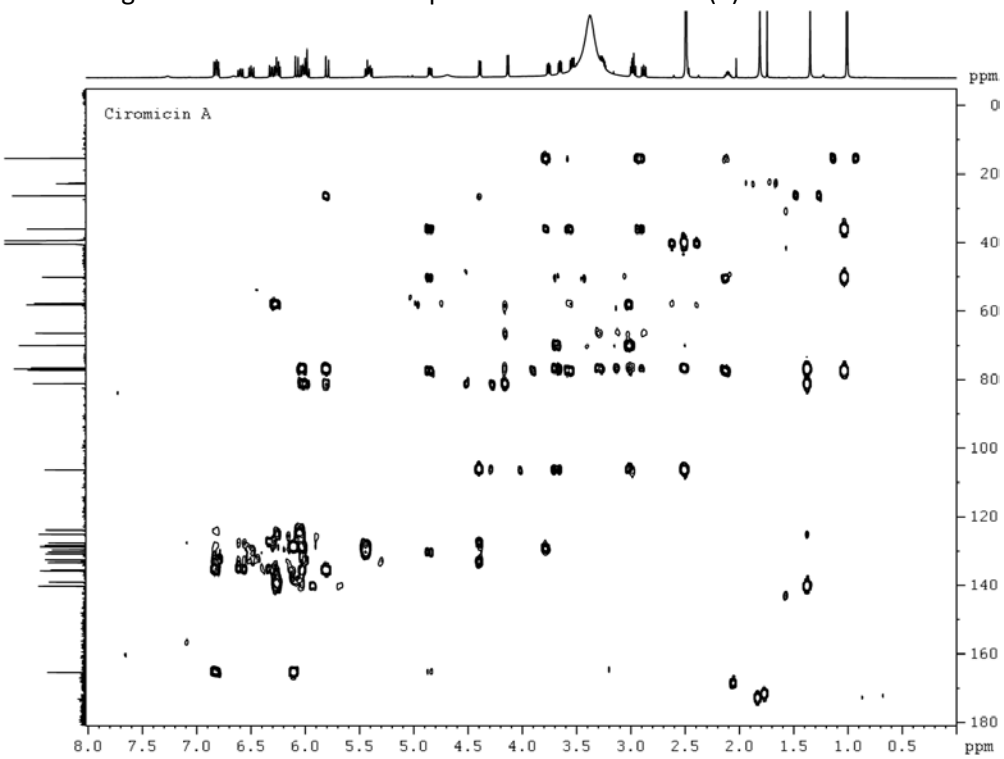


Figure A-6 2D TOCSY NMR spectrum of ciromicin A (1) in DMSO-d6.

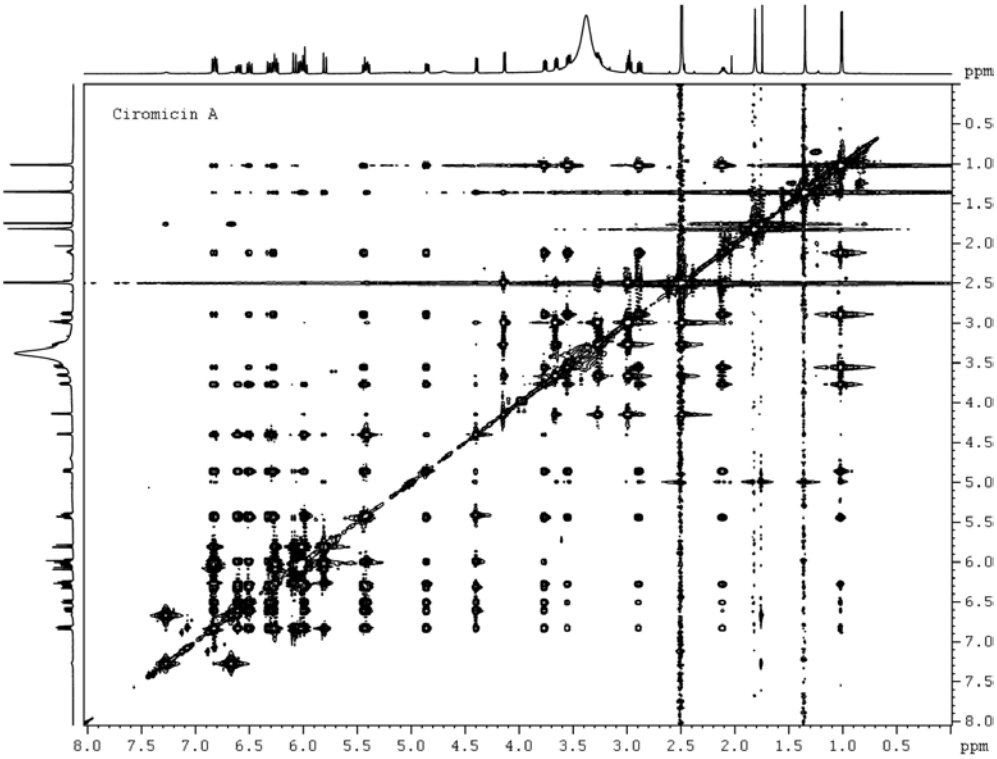


Figure A-7 2D NOESY NMR spectrum of ciromicin A (1) in DMSO-d6.

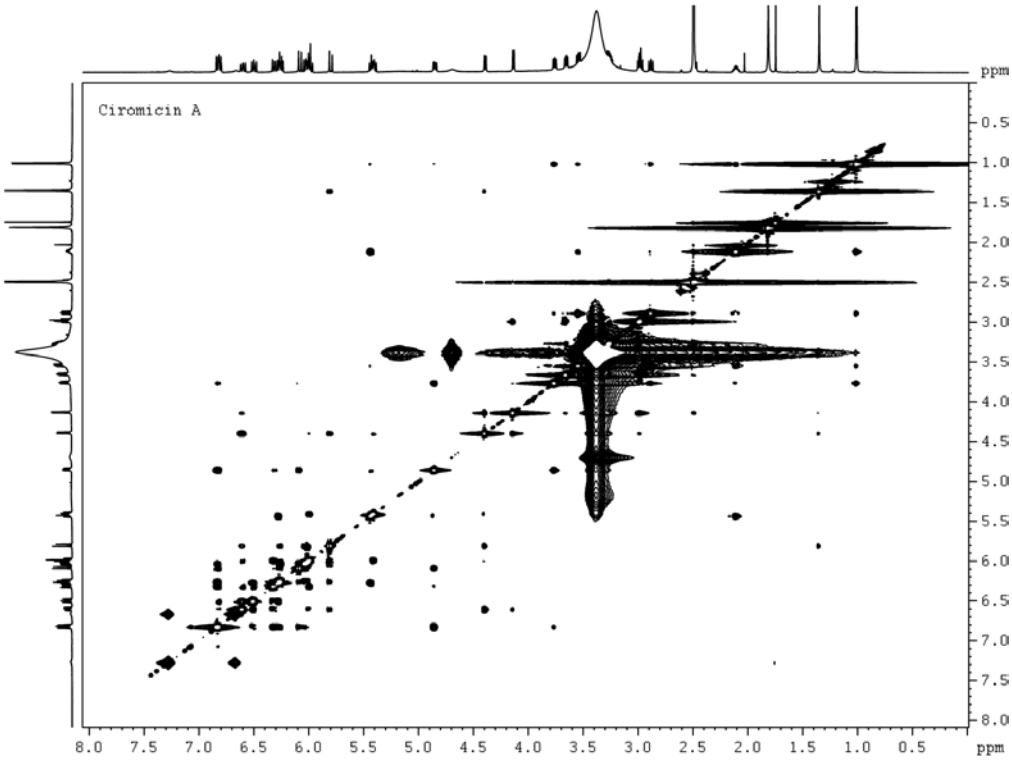


Figure A-8 ^{1}H proton NMR spectrum of ciromicin A (1) in CD_3OD.

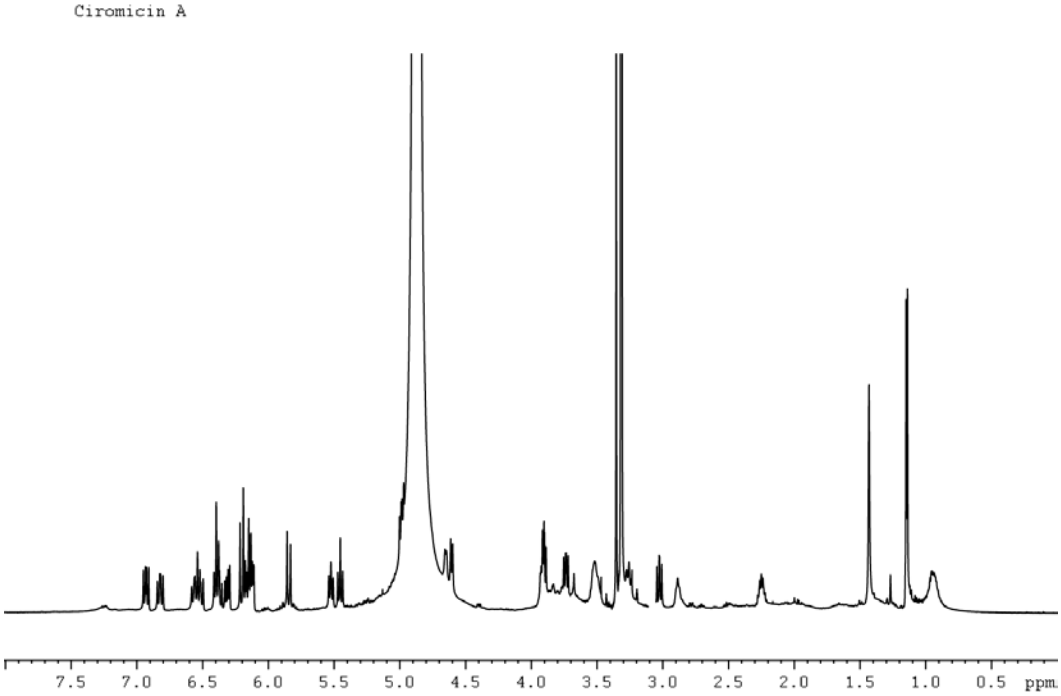


Figure A-9 ^{13}C carbon NMR spectrum of ciromicin A (1) in CD_3OD.

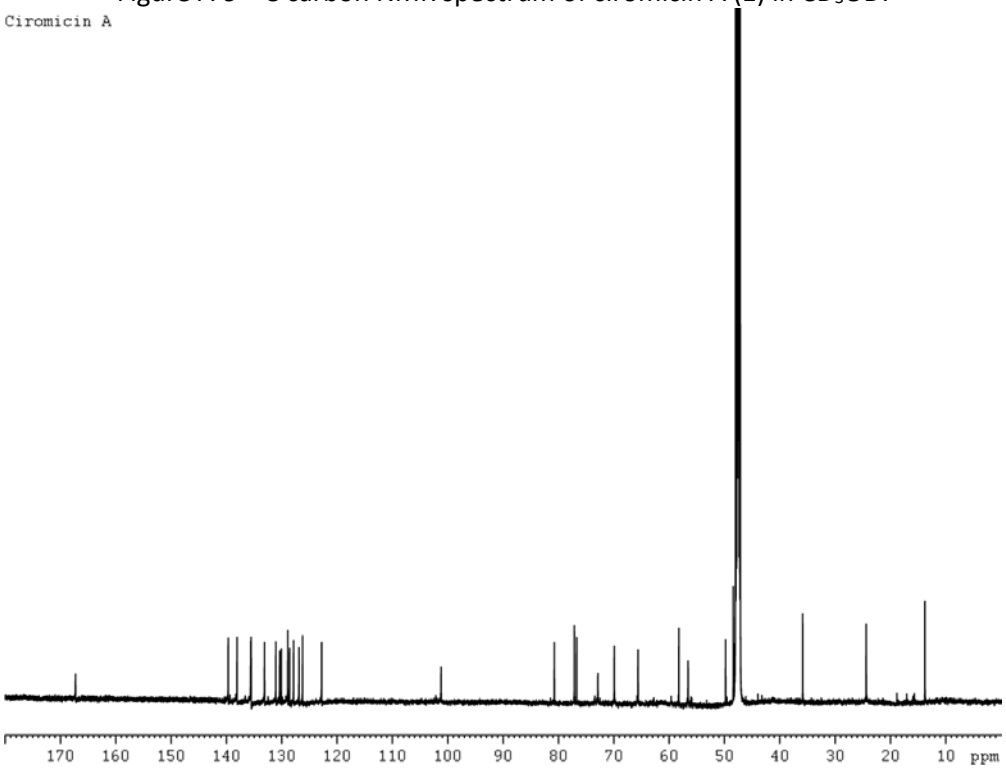


Figure A-10 2D COSY NMR spectrum of ciromicin A (1) in CD_3OD.

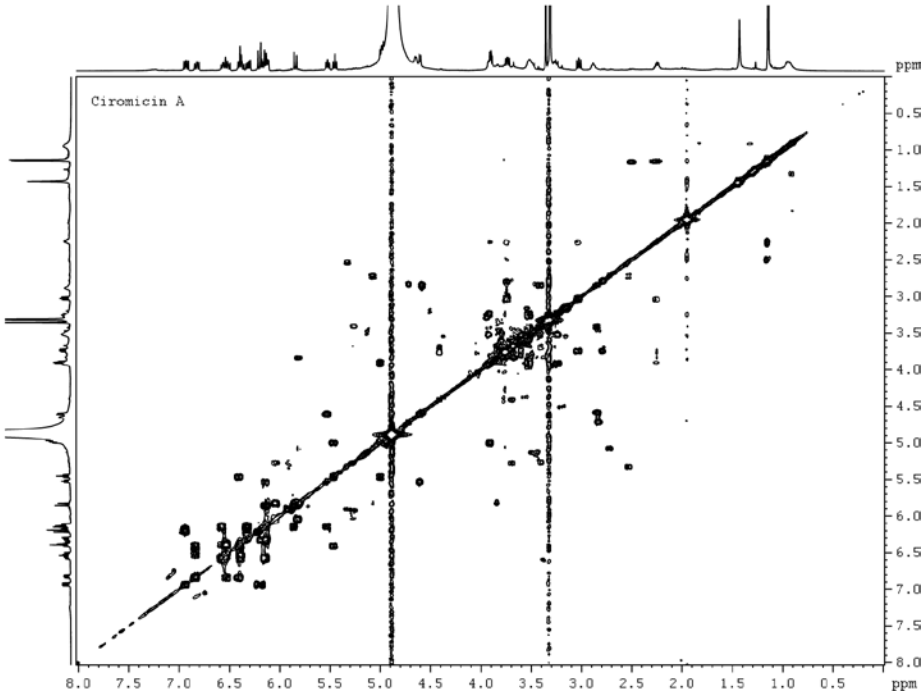


Figure A-11 2D HSQC NMR spectrum of ciromicin A (1) in CD_3OD.

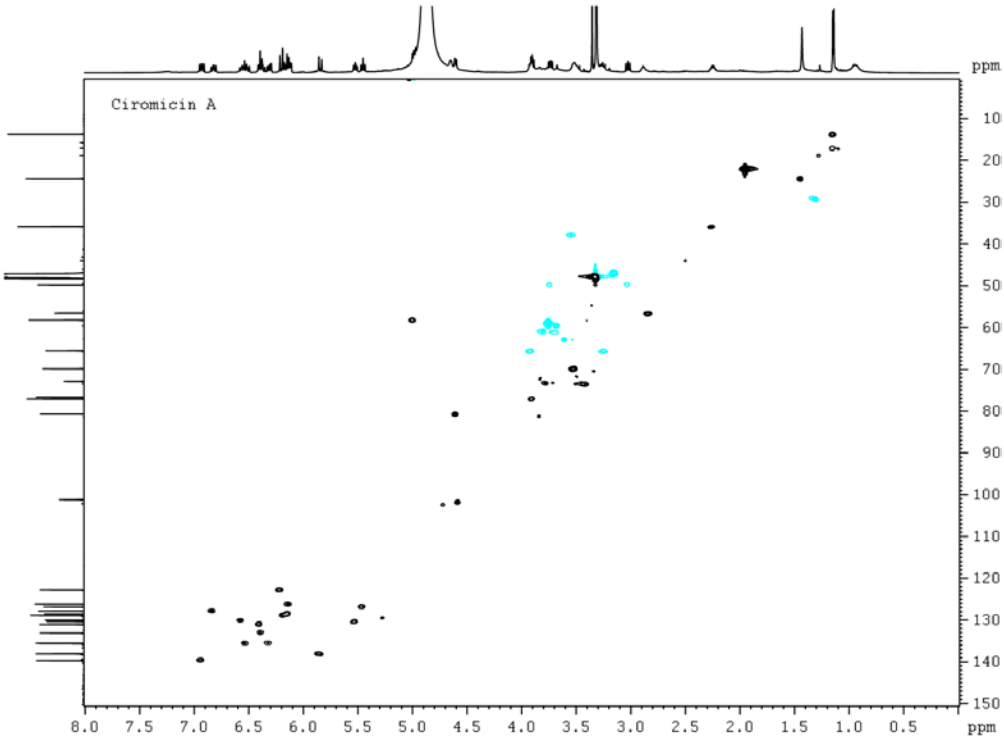


Figure A-12 2D HMBC NMR spectrum of ciromicin A (1) in CD_3OD.

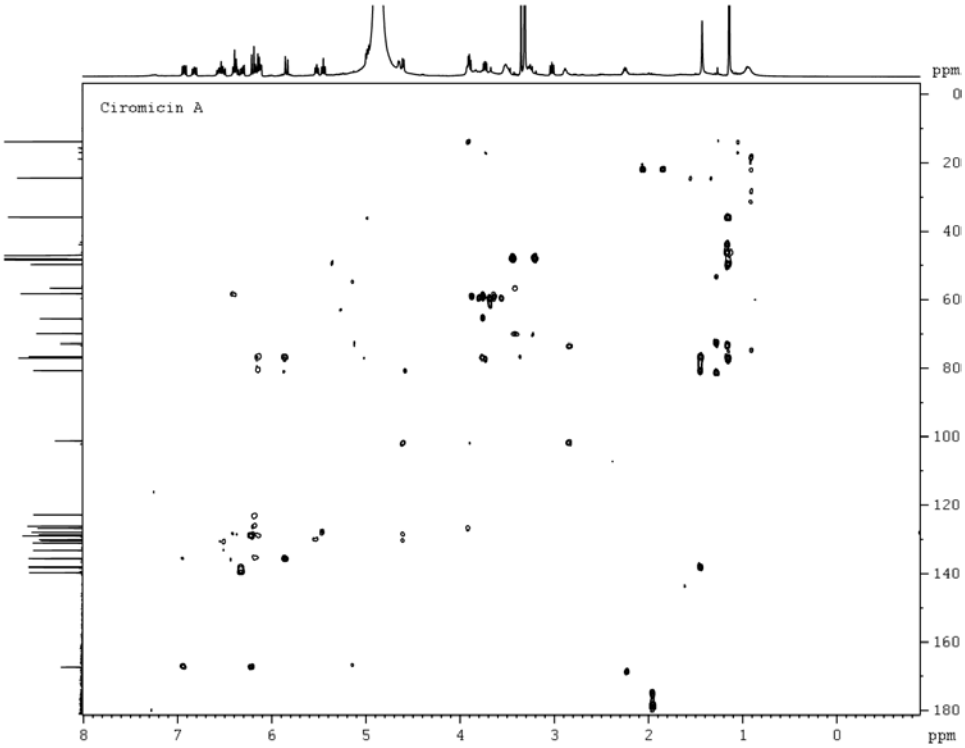


Figure A-13 2D TOCSY NMR spectrum of ciromicin A (1) in CD_3OD.

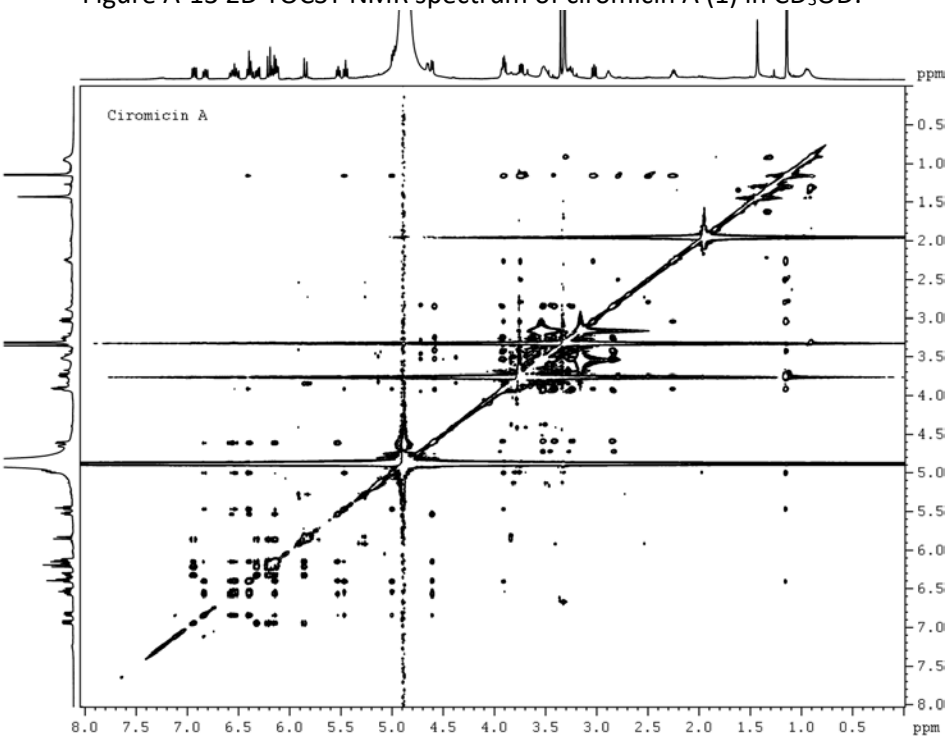


Figure A-14 2D NOESY NMR spectrum of ciromicin A (1) in CD_3OD.

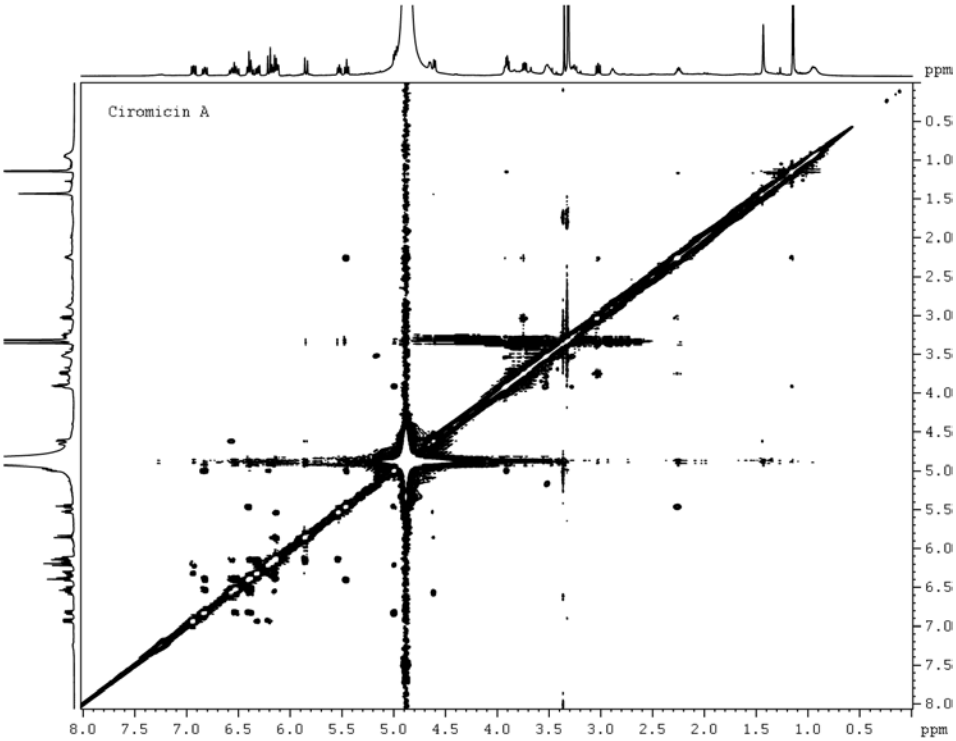


Figure A-15 ^{1}H proton NMR spectrum of ciromicin B (2) in CD_3OD.

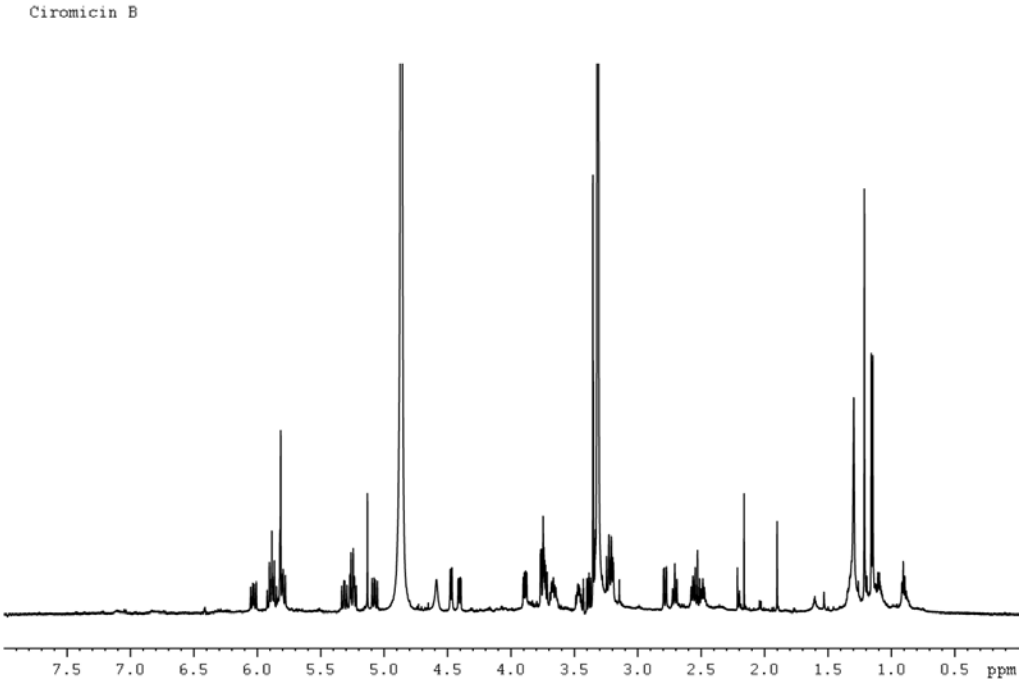


Figure A-16 ^{13}C carbon NMR spectrum of ciromicin B (2) in CD_3OD.

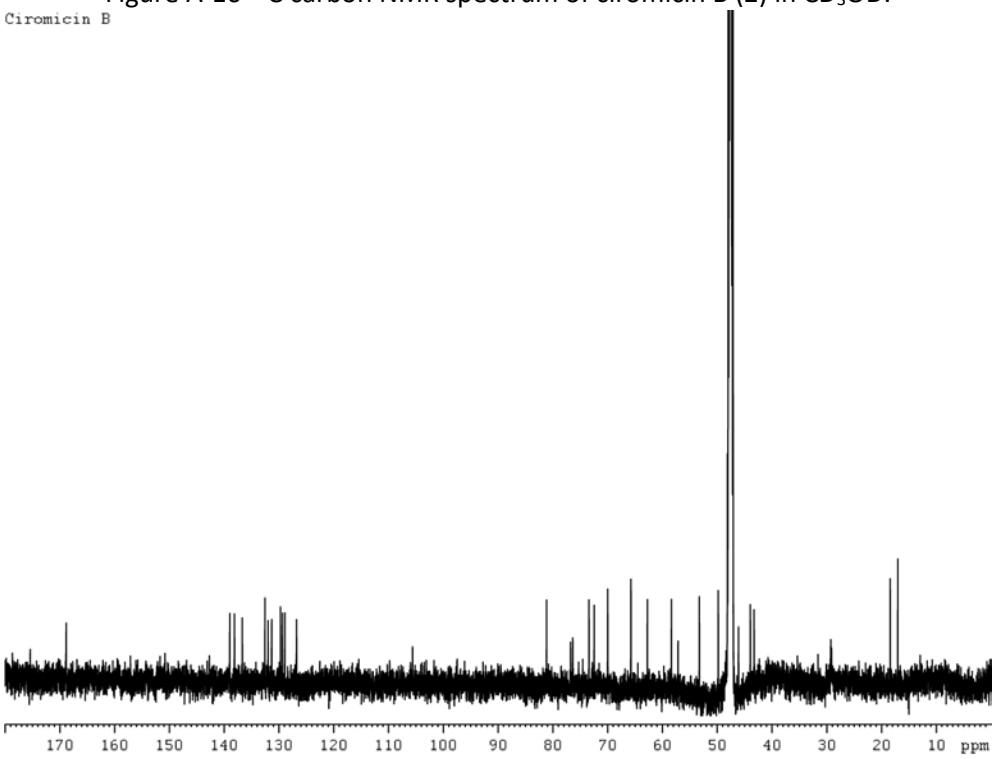


Figure A-17 2D COSY NMR spectrum of ciromicin B (2) in CD_3OD.

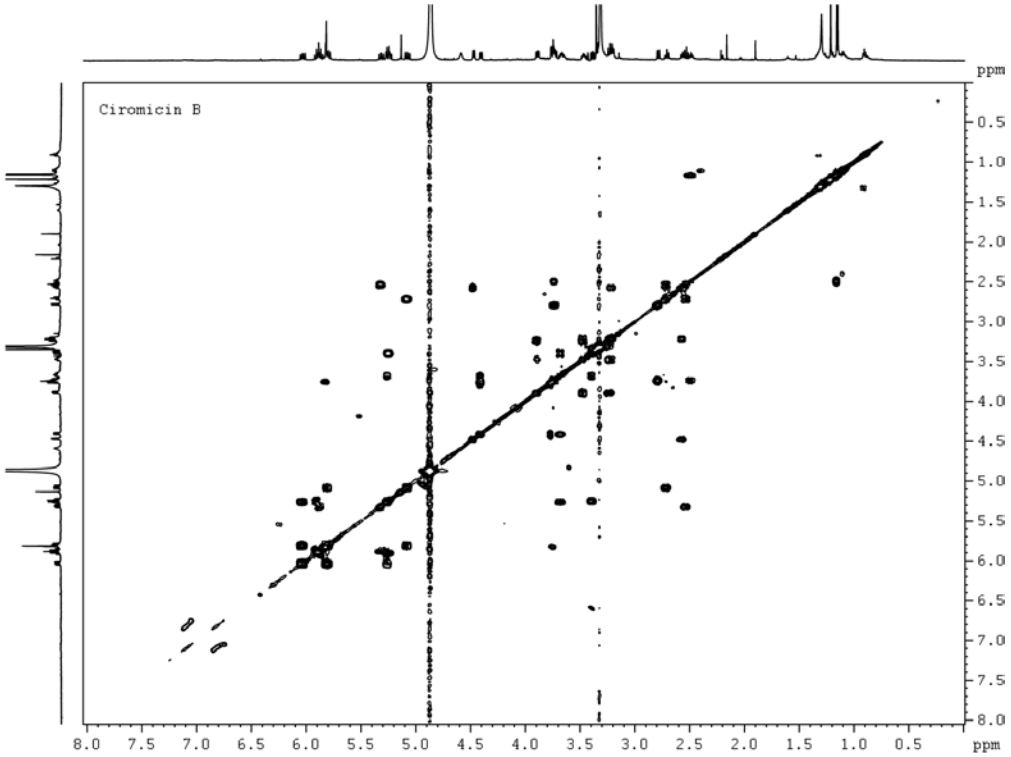


Figure A-18 2D HSQC NMR spectrum of ciromicin B (2) in CD_3OD.

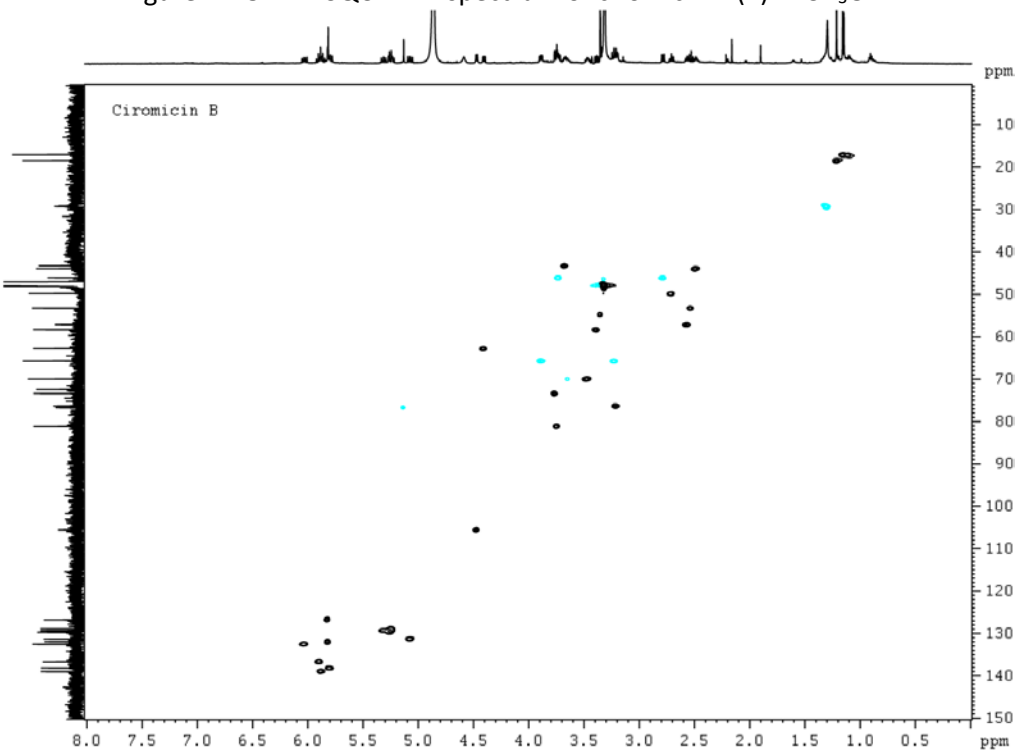


Figure A-19 2D HMBC NMR spectrum of ciromicin B (2) in CD_3OD.

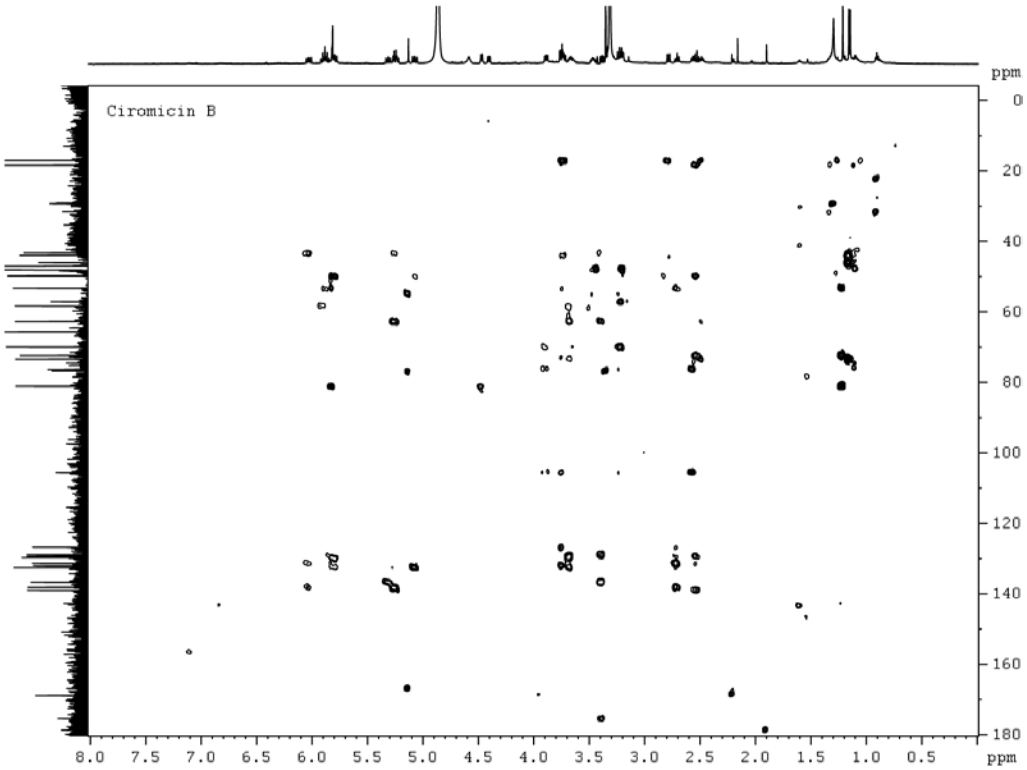


Figure A-20 2D TOCSY NMR spectrum of ciromicin B (2) in CD_3OD.

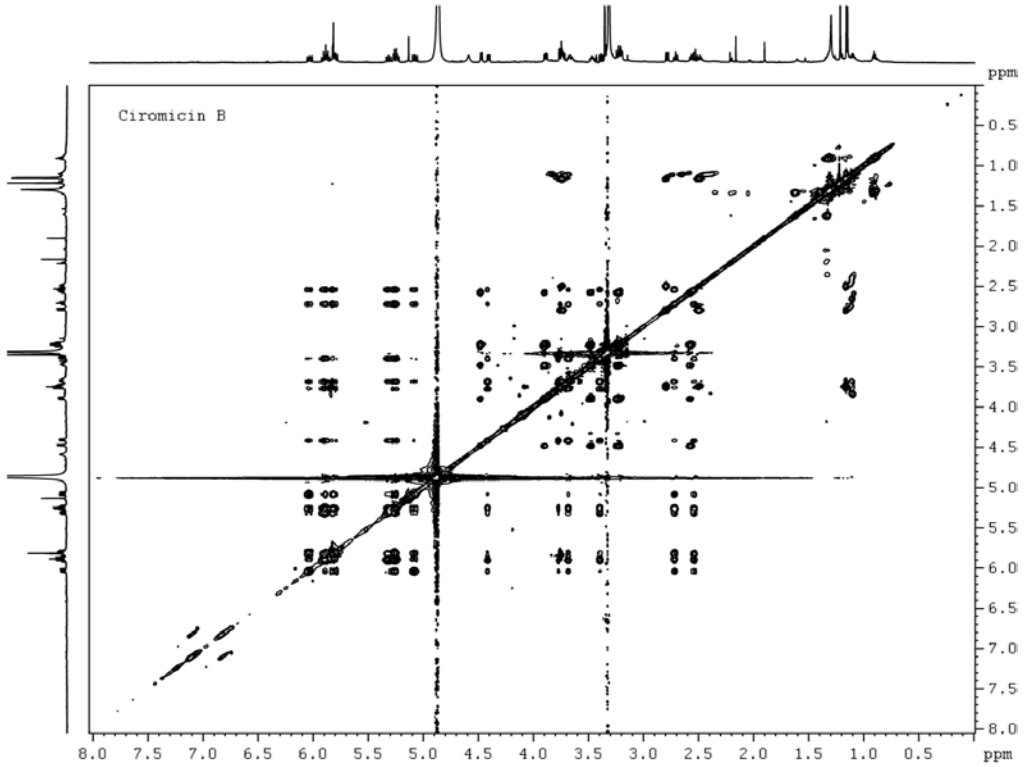


Figure A-21 2D NOESY NMR spectrum of ciromicin B (2) in CD_3OD.

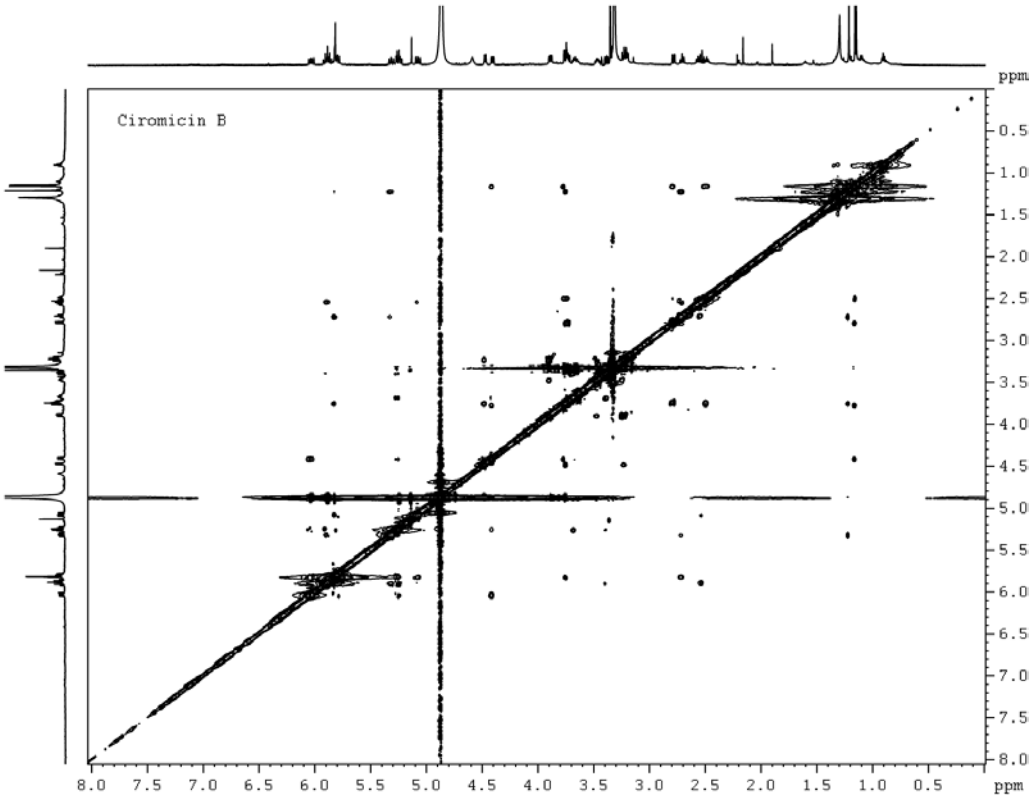
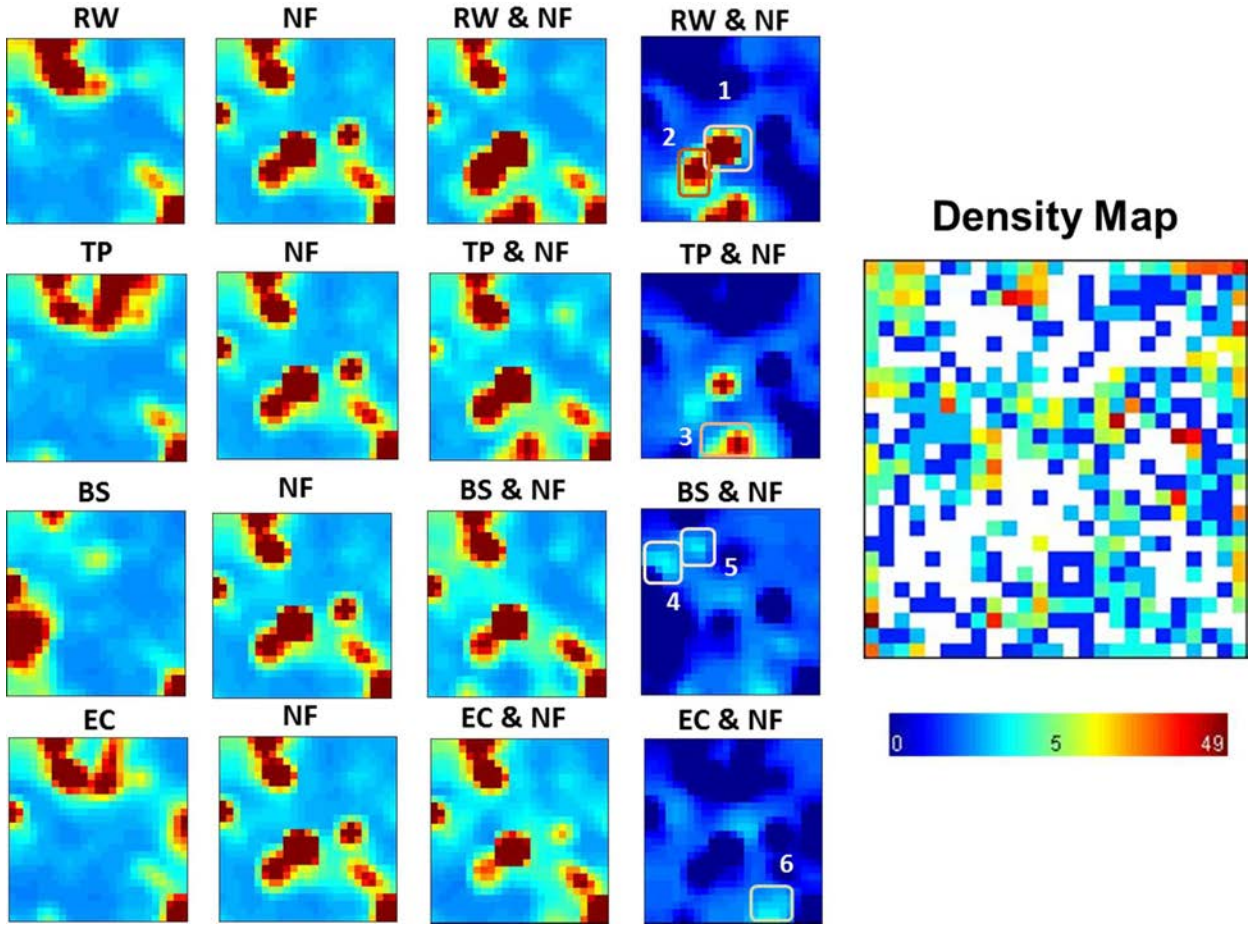


Figure A-22: Heatmaps and density map for *Nocardiopsis* mixed culture screen The density map shows the distribution of features across heat map. Here the number of features in each node is indicated by the color of the node in accordance with the scale. Uncolored areas of the density map represent nodes with no features. For clarification, this distribution is the same across all experimental conditions (every map in the analysis is consistent with respect to feature distribution).



Tables A4-A12 show the averaged and normalized intensity values for features MxxxTxxx, where the numbers following M and T are mass to charge and retention time respectively, in the description (DESC) column across the various experimental conditions. These feature descriptions are sorted by their contribution to the region of interest (shown as a percentage in % ROI). Highlighted experimental condition columns have features significantly more abundant than in the control (Noc mono) condition.

Table A-4 Features within ROI 1 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M415T959	536.8	508.2	7.0	7.5	881.7	4.4	385.5	2.4	643.4	82%
M437T959	44.5	41.5	0.7	0.8	68.7	0.7	31.2	0.5	48.5	7%
M214T259	12.6	19.9	6.6	2.2	19.9	5.4	20.8	0.9	17.4	3%
M301T259	5.4	9.5	5.7	1.6	15.1	3.7	15.2	1.1	8.0	2%
M438T959	9.4	9.1	0.3	0.3	14.6	0.3	6.8	0.2	10.3	1%
M204T345	3.2	5.2	1.0	0.8	5.6	1.4	6.6	0.7	4.9	1%
M103T347	3.1	4.7	0.5	0.8	5.1	1.8	6.1	0.9	4.5	1%
M248T345	2.7	4.9	2.6	0.8	5.1	1.2	5.2	0.6	4.7	1%
M369T346	1.1	2.5	0.7	0.6	3.7	0.6	4.0	0.4	2.5	0%
M170T259	1.9	3.0	0.7	0.9	3.7	2.6	4.2	0.7	2.7	1%
M141T958	2.2	1.8	0.2	0.3	3.8	0.4	1.6	0.3	2.4	0%
M453T959	1.9	2.0	0.2	0.4	2.8	1.2	1.5	0.4	2.1	0%
M439T960	1.8	1.9	0.2	0.4	2.1	0.4	1.5	1.1	1.7	0%
M158T958	1.2	1.0	0.1	0.3	2.1	0.3	1.1	0.2	1.3	0%
M77T346	0.7	1.2	0.4	0.3	1.2	0.4	1.7	0.2	1.2	0%

Table A-5 Features within ROI 2 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M397T958	207.9	178.0	6.4	12.3	402.9	12.6	136.7	14.6	244.3	32%
M416T958	155.0	134.4	1.6	1.9	338.4	1.2	96.8	1.0	187.3	24%
M398T958	59.5	48.7	0.8	3.1	122.0	3.5	36.4	3.7	69.0	9%
M379T958	52.1	43.3	0.7	1.5	112.4	1.0	25.3	0.6	60.8	8%
M115T958	47.8	41.1	0.7	1.0	106.3	1.9	33.4	0.5	62.0	8%
M417T958	20.2	17.0	0.3	0.5	46.0	0.4	12.3	0.4	24.5	3%
M312T209	5.5	6.4	46.1	6.5	5.8	4.1	6.9	4.1	11.0	3%
M380T958	12.6	10.4	0.5	0.7	26.9	0.7	8.1	0.7	14.6	2%
M296T206	3.1	4.2	37.5	3.0	2.9	3.4	4.9	1.8	6.9	2%
M70T958	9.1	7.7	0.3	0.3	21.6	0.3	6.6	0.2	11.9	2%
M399T958	7.3	5.7	1.1	0.8	14.9	0.8	5.2	5.2	8.8	1%
M234T821	6.3	6.3	12.4	1.2	7.0	0.9	4.2	0.8	9.8	1%
M387T875	7.0	3.9	1.0	1.7	11.4	1.9	4.2	1.1	7.4	1%
M361T958	4.2	3.8	0.2	0.5	9.3	0.5	2.9	0.3	5.0	1%
M565T436	2.8	4.3	1.1	1.5	6.5	1.4	1.4	1.3	4.6	1%
M116T958	3.8	3.3	0.1	0.3	8.3	0.2	2.8	0.1	4.9	1%
M205T638	2.2	2.1	3.2	1.6	2.1	2.2	2.0	0.7	3.5	1%
M441T650	0.7	0.9	11.6	0.7	2.1	0.0	0.0	0.6	2.3	0%
M369T875	2.9	1.8	0.4	0.8	4.6	0.6	1.9	0.6	3.0	0%
M298T687	0.5	0.8	7.2	0.6	2.6	0.0	0.0	0.5	3.3	0%
M418T958	2.0	1.7	0.2	0.3	3.8	0.2	1.5	0.3	2.1	0%
M133T958	2.0	1.7	0.3	0.5	3.8	0.0	0.0	0.4	2.1	0%
M388T875	1.6	1.0	0.3	0.5	2.6	1.3	1.1	0.4	1.7	0%
M504T753	1.7	0.9	0.2	0.8	2.6	0.5	1.4	0.5	1.7	0%
M157T958	1.4	1.3	0.2	0.3	2.7	0.4	1.4	0.3	1.5	0%
M362T958	1.2	1.2	0.2	0.3	2.4	0.3	0.9	0.2	1.4	0%

Table A-6: Features within ROI 3 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M376T896	0.2	0.0	0.0	0.2	281.8	0.0	0.0	0.3	253.6	19%
M384T609	0.7	9.4	16.9	1.0	142.6	0.8	5.4	0.9	56.2	8%
M549T472	0.9	4.9	1.2	1.8	74.8	1.4	4.2	1.3	41.7	5%
M377T896	0.3	0.0	0.0	0.4	69.8	0.0	0.0	0.4	57.6	5%
M563T514	1.6	2.0	0.7	1.4	61.8	0.6	1.9	1.2	47.6	4%
M515T602	0.5	5.1	1.2	0.4	34.2	0.8	2.8	0.8	31.7	3%
M225T851	3.8	0.0	0.0	4.3	65.6	0.0	0.0	0.0	0.0	3%
M843T649	4.0	12.6	0.3	0.7	18.6	0.8	11.0	0.8	22.5	3%
M400T480	1.2	3.4	2.0	1.5	32.3	0.9	2.7	2.4	21.5	2%
M432T536	1.0	0.0	0.0	1.3	35.1	0.0	0.0	1.2	25.6	2%
M418T475	0.9	0.0	0.0	1.3	36.5	0.0	0.0	1.1	22.6	2%
M366T545	1.6	2.3	1.2	1.0	28.8	0.4	1.5	0.8	15.3	2%
M564T545	9.9	0.0	0.0	1.5	22.3	0.0	0.0	1.2	13.4	2%
M382T512	1.1	0.0	0.0	1.3	24.9	0.0	0.0	1.1	17.9	2%
M531T481	1.3	0.0	0.0	2.6	19.9	0.0	0.0	1.8	12.8	1%
M550T472	0.8	0.0	0.0	1.6	20.8	0.0	0.0	1.2	12.1	1%
M567T607	0.6	0.0	0.0	0.7	20.6	0.0	0.0	1.1	12.2	1%
M844T649	2.1	5.5	0.3	0.7	11.9	0.5	5.3	0.8	10.0	1%
M385T608	0.8	0.0	0.0	1.1	31.4	0.0	0.0	0.0	0.0	1%
M414T538	1.0	0.0	0.0	1.2	16.0	0.0	0.0	1.2	11.5	1%
M564T589	6.8	2.8	1.8	2.9	7.8	3.2	2.0	1.5	2.2	1%
M712T436	0.5	7.4	0.4	1.0	4.4	0.0	0.0	0.6	16.1	1%
M377T506	1.4	3.9	1.2	1.1	8.6	1.5	3.7	1.2	7.2	1%
M487T396	1.6	2.5	1.0	1.5	8.6	1.0	2.7	0.9	7.9	1%
M419T477	1.5	0.0	0.0	1.3	9.2	3.1	3.0	1.7	6.9	1%
M318T531	2.1	3.5	1.9	1.1	4.3	2.8	2.3	1.0	6.5	1%
M516T655	0.3	0.0	0.0	0.4	9.9	0.0	0.0	0.4	13.4	1%
M569T589	1.0	2.0	1.8	1.3	7.2	2.9	1.6	2.2	4.1	1%
M234T544	3.2	5.5	2.5	2.3	5.1	0.0	0.0	3.0	1.8	1%
M348T649	0.5	0.0	0.0	0.8	12.5	0.0	0.0	0.5	8.8	1%
M433T535	1.2	0.0	0.0	1.5	9.1	0.0	0.0	1.4	6.9	1%
M434T439	0.8	1.6	2.2	1.5	6.6	0.0	0.0	1.0	4.6	1%
M436T616	0.5	0.0	0.0	0.5	10.0	0.0	0.0	0.6	6.7	1%
M565T546	3.0	1.1	2.0	1.0	4.0	2.4	0.8	0.9	2.4	1%
M434T897	0.2	0.0	0.0	0.2	7.0	0.0	0.0	0.2	7.9	1%
M378T896	0.3	0.0	0.0	0.4	7.7	0.0	0.0	0.4	6.7	1%
M393T896	1.3	0.0	0.0	1.0	6.3	0.0	0.0	1.2	5.5	1%

Table A-7: Features within ROI 3 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M567T535	0.8	0.0	0.0	1.1	13.0	0.0	0.0	0.0	0.0	1%
M452T409	0.8	0.0	0.0	1.3	7.3	0.0	0.0	1.0	4.1	1%
M499T673	0.4	0.0	0.0	0.5	10.4	0.0	0.0	0.8	2.3	1%
M865T649	0.7	2.1	0.2	0.7	4.2	0.5	1.5	0.6	3.6	1%
M568T601	0.5	0.0	0.0	0.6	5.7	0.0	0.0	3.3	3.7	0%
M497T661	0.4	0.0	0.0	0.4	8.3	0.0	0.0	0.4	4.3	0%
M435T437	0.9	2.9	1.1	1.4	2.4	0.0	0.0	2.8	1.7	0%
M367T561	0.6	0.0	0.0	0.9	6.8	0.0	0.0	0.8	3.8	0%
M806T499	1.5	1.1	0.7	1.7	4.1	0.9	0.5	1.3	1.0	0%
M537T664	0.3	0.0	0.0	0.3	5.6	0.0	0.0	0.4	5.5	0%
M713T437	1.1	2.4	0.4	1.0	1.6	0.0	0.0	0.7	4.7	0%
M415T542	1.7	2.3	1.3	1.5	4.7	0.0	0.0	0.0	0.0	0%
M350T654	1.7	2.7	0.9	0.7	5.3	0.0	0.0	0.0	0.0	0%
M436T544	0.8	0.0	0.0	0.9	4.7	0.0	0.0	0.8	4.1	0%
M585T548	1.4	0.0	0.0	2.0	7.7	0.0	0.0	0.0	0.0	0%
M368T744	0.4	0.0	0.0	0.6	5.8	0.0	0.0	0.4	3.8	0%
M107T890	0.7	1.9	0.3	0.5	2.7	0.0	0.0	1.3	2.8	0%
M728T421	0.5	0.0	0.0	0.9	4.0	0.0	0.0	0.9	2.7	0%
M513T692	0.4	0.0	0.0	0.5	3.4	0.0	0.0	0.3	4.3	0%
M293T897	1.2	0.0	0.0	0.8	2.7	0.0	0.0	1.0	2.5	0%
M678T478	0.6	0.0	0.0	1.2	2.7	0.0	0.0	0.8	2.7	0%
M348T561	0.7	0.0	0.0	2.2	5.1	0.0	0.0	0.0	0.0	0%
M438T616	0.7	0.0	0.0	0.6	3.5	0.0	0.0	0.6	2.5	0%
M521T663	0.4	0.0	0.0	1.8	5.5	0.0	0.0	0.0	0.0	0%
M97T959	1.5	0.0	0.0	0.4	2.8	0.0	0.0	0.6	2.0	0%
M304T258	1.0	1.8	0.6	1.3	2.5	0.0	0.0	0.0	0.0	0%
M675T474	1.5	0.0	0.0	2.2	3.5	0.0	0.0	0.0	0.0	0%
M302T897	0.3	0.0	0.0	0.3	3.1	0.0	0.0	0.4	2.9	0%
M569T535	0.8	0.0	0.0	1.3	4.9	0.0	0.0	0.0	0.0	0%
M329T676	0.5	0.0	0.0	0.8	3.3	0.0	0.0	0.6	1.5	0%
M499T744	0.3	0.0	0.0	0.4	3.4	0.0	0.0	0.2	2.2	0%
M448T450	0.9	0.0	0.0	1.6	4.0	0.0	0.0	0.0	0.0	0%
M551T473	0.7	0.0	0.0	1.7	4.0	0.0	0.0	0.0	0.0	0%
M866T648	0.5	0.0	0.0	0.7	2.2	0.0	0.0	0.8	2.0	0%
M526T674	0.4	0.0	0.0	0.5	3.4	0.0	0.0	0.4	1.6	0%
M349T632	0.7	0.0	0.0	1.1	4.4	0.0	0.0	0.0	0.0	0%
M396T540	0.9	0.0	0.0	1.0	4.2	0.0	0.0	0.0	0.0	0%

Table A-8: Features within ROI 3 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M511T780	1.1	0.0	0.0	0.9	2.2	0.0	0.0	0.5	1.3	0%
M292T896	0.4	0.0	0.0	0.4	2.5	0.0	0.0	0.4	2.2	0%
M589T616	0.5	0.0	0.0	0.6	2.6	0.0	0.0	0.7	1.6	0%
M568T535	0.7	0.0	0.0	1.0	3.9	0.0	0.0	0.0	0.0	0%
M338T560	1.9	0.0	0.0	0.9	2.7	0.0	0.0	0.0	0.0	0%
M1177T678	0.1	0.0	0.0	0.8	2.0	0.0	0.0	0.6	2.0	0%
M314T531	0.9	0.0	0.0	1.2	3.4	0.0	0.0	0.0	0.0	0%
M317T940	0.3	0.0	0.0	0.6	3.2	0.0	0.0	0.3	1.1	0%
M441T996	0.3	0.0	0.0	1.0	2.5	0.0	0.0	0.4	1.3	0%
M414T959	1.3	0.0	0.0	0.2	2.2	0.0	0.0	0.4	1.3	0%
M1137T677	0.2	0.0	0.0	0.9	1.8	0.0	0.0	0.6	1.8	0%
M435T897	0.3	0.0	0.0	0.4	1.8	0.0	0.0	0.3	2.1	0%
M545T525	0.7	0.0	0.0	1.2	3.0	0.0	0.0	0.0	0.0	0%
M351T876	1.1	0.0	0.0	0.4	1.6	0.0	0.0	0.5	1.2	0%
M376T648	0.5	1.1	0.1	0.1	1.3	0.0	0.0	0.1	1.5	0%
M160T471	0.8	0.0	0.0	1.0	2.9	0.0	0.0	0.0	0.0	0%
M361T347	1.0	0.0	0.0	0.5	1.4	0.0	0.0	0.3	1.2	0%
M570T602	1.7	0.0	0.0	0.6	2.1	0.0	0.0	0.0	0.0	0%
M520T897	0.2	0.0	0.0	0.2	1.6	0.0	0.0	0.5	1.7	0%
M400T947	0.7	0.0	0.0	0.2	1.6	0.0	0.0	1.1	0.7	0%
M241T767	0.5	0.0	0.0	1.6	2.2	0.0	0.0	0.0	0.0	0%
M233T821	0.8	0.0	0.0	0.5	1.1	0.0	0.0	0.3	1.4	0%
M153T392	1.1	0.0	0.0	0.8	2.1	0.0	0.0	0.0	0.0	0%
M460T896	0.2	0.0	0.0	0.2	1.6	0.0	0.0	0.1	1.6	0%
M402T365	1.4	0.0	0.0	0.7	1.6	0.0	0.0	0.0	0.0	0%
M730T421	0.5	0.0	0.0	1.3	1.9	0.0	0.0	0.0	0.0	0%
M743T440	0.7	0.0	0.0	1.0	1.8	0.0	0.0	0.0	0.0	0%
M913T679	0.7	0.0	0.0	1.1	1.6	0.0	0.0	0.0	0.0	0%
M504T336	0.2	0.0	0.0	0.4	1.0	0.0	0.0	0.3	1.5	0%
M893T681	0.4	0.0	0.0	1.1	1.7	0.0	0.0	0.0	0.0	0%
M541T750	1.3	0.0	0.0	0.4	1.5	0.0	0.0	0.0	0.0	0%
M392T896	0.2	0.0	0.0	0.4	1.2	0.0	0.0	0.2	1.2	0%
M1012T678	0.3	0.0	0.0	1.2	1.7	0.0	0.0	0.0	0.0	0%
M1181T678	0.2	0.0	0.0	1.0	1.5	0.0	0.0	0.0	0.0	0%

Table A-9: Features within ROI 4 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M273T644	19.6	61.9	23.9	33.1	16.8	12.4	21.0	15.3	36.3	16%
M331T710	13.7	38.4	20.1	25.3	13.3	7.9	27.7	20.1	33.4	14%
M270T446	20.0	26.8	11.0	26.0	17.2	18.7	17.8	20.1	31.7	13%
M229T494	7.4	38.0	11.2	4.3	9.6	11.1	18.7	4.2	17.8	8%
M272T513	4.8	5.9	4.3	7.1	4.6	4.4	4.9	4.7	6.4	3%
M526T200	4.3	9.3	3.0	6.8	3.4	5.0	6.0	5.1	3.8	3%
M254T522	4.7	5.9	6.1	6.3	3.8	4.1	4.3	4.6	5.8	3%
M662T440	1.4	17.3	0.8	1.6	7.0	1.4	5.4	1.6	6.0	3%
M447T821	3.3	7.5	2.4	2.7	2.2	1.9	7.3	2.9	7.5	3%
M818T535	1.5	19.2	0.3	0.8	3.1	0.0	0.0	0.6	8.3	2%
M285T598	3.8	4.7	2.7	4.6	3.2	3.7	3.9	3.2	3.6	2%
M274T643	2.3	6.9	4.2	4.0	2.6	2.1	3.2	1.9	4.5	2%
M332T710	2.0	5.5	3.1	3.8	1.9	1.4	4.1	2.9	4.7	2%
M739T475	0.7	9.7	0.7	1.3	2.8	1.0	5.4	1.3	5.8	2%
M152T623	2.4	3.1	2.6	3.2	2.4	2.5	2.3	2.4	3.3	2%
M249T468	2.5	8.1	1.4	5.1	2.1	0.0	0.0	2.0	2.9	2%
M547T490	1.0	17.4	0.6	1.1	3.9	0.0	0.0	0.0	0.0	2%
M350T493	2.1	7.6	1.2	0.0	0.0	1.4	4.3	1.5	3.9	1%
M307T514	2.0	2.8	1.5	4.1	2.2	1.4	2.3	1.6	3.1	1%
M390T769	1.5	3.2	1.4	1.7	1.3	0.9	3.3	1.9	3.4	1%
M425T465	1.6	8.8	1.1	1.2	0.6	0.7	0.5	0.7	1.9	1%
M298T603	1.5	2.2	1.9	1.9	1.4	1.6	1.7	1.3	1.7	1%
M506T430	2.5	4.4	2.8	2.7	2.1	0.0	0.0	0.0	0.0	1%
M461T645	1.4	3.6	0.8	0.7	0.7	0.6	2.0	0.7	2.8	1%
M632T428	1.3	2.1	0.9	1.4	0.9	1.2	1.6	1.3	1.7	1%
M674T998	0.9	1.8	0.7	1.2	0.7	0.7	2.0	1.2	1.8	1%
M592T484	0.9	5.3	1.2	0.0	0.0	0.0	0.0	2.4	1.1	1%
M361T387	1.1	4.9	0.9	0.4	1.5	0.0	0.0	0.3	1.6	1%
M423T427	1.2	3.9	1.5	1.5	2.6	0.0	0.0	0.0	0.0	1%
M819T519	1.0	5.3	0.4	0.0	0.0	0.0	0.0	0.6	3.0	1%
M575T488	0.5	5.8	0.5	0.0	0.0	0.0	0.0	0.7	2.6	1%
M649T422	0.8	4.5	0.9	0.0	0.0	0.0	0.0	1.1	1.9	1%
M417T374	0.5	2.5	1.1	2.1	0.6	0.0	0.0	1.5	0.7	1%
M769T509	1.0	3.8	0.8	1.4	1.4	0.0	0.0	0.0	0.0	1%
M675T411	0.5	4.2	0.9	0.0	0.0	0.0	0.0	0.7	1.8	1%
M624T501	0.5	3.4	1.1	0.0	0.0	0.0	0.0	0.4	1.3	0%
M759T417	0.4	2.4	0.7	1.3	1.6	0.0	0.0	0.0	0.0	0%
M645T451	1.4	3.8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0%
M521T777	0.5	1.4	0.4	0.0	0.0	0.4	1.2	0.4	1.8	0%
M798T533	0.9	3.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0%
M447T620	0.9	3.3	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0%
M757T518	1.0	2.5	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0%
M674T778	0.2	1.5	0.1	0.0	0.0	0.0	0.0	0.2	1.6	0%

Table A-10: Features within ROI 5 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M239T542	269.2	295.9	9.5	459.2	284.4	371.8	270.2	187.5	365.0	71%
M225T463	98.3	131.8	8.9	181.8	96.4	103.7	104.8	71.3	128.2	26%
M226T464	11.1	15.6	1.3	22.1	11.2	11.9	11.6	9.0	16.0	3%

Table A-11: Features within ROI 6 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M244T394	7.4	49.4	2.5	37.8	58.2	19.7	50.0	5.6	69.7	14%
M284T394	5.3	36.1	1.6	37.3	38.9	1.3	57.8	1.1	39.0	10%
M268T336	14.2	27.7	1.9	16.8	19.4	6.7	21.7	11.2	37.1	7%
M152T244	22.3	6.4	14.0	17.6	3.5	4.0	31.7	2.4	29.9	6%
M208T394	3.1	19.5	1.4	13.5	21.0	2.7	26.2	1.2	25.0	5%
M529T711	7.1	0.0	0.0	22.4	8.3	7.1	42.5	0.0	0.0	4%
M484T500	1.3	14.1	2.9	1.5	14.5	0.8	19.2	0.6	25.1	4%
M226T394	2.3	11.5	1.3	10.1	14.0	1.2	17.0	1.3	16.5	4%
M180T388	11.3	7.9	2.4	2.2	9.1	2.0	13.2	2.5	18.8	3%
M512T509	0.9	12.0	3.4	1.4	5.5	1.2	8.3	2.7	22.6	3%
M321T203	3.4	9.1	2.8	6.3	6.0	5.3	9.0	2.7	9.8	3%
M337T203	4.0	8.0	2.8	5.8	5.9	5.6	8.6	1.7	8.9	2%
M304T444	5.0	6.5	2.0	4.1	4.5	3.2	5.4	4.4	9.5	2%
M268T439	3.9	3.8	3.2	3.3	4.7	1.4	5.6	1.7	8.5	2%
M448T200	5.6	5.6	2.2	3.5	1.6	2.5	7.8	2.6	1.6	2%
M209T439	3.2	5.5	2.7	3.0	4.0	2.3	4.3	1.4	6.3	2%
M371T383	2.0	0.0	0.0	4.9	6.3	5.1	8.5	0.7	4.0	1%
M532T202	2.2	4.2	1.9	5.1	2.9	3.2	4.9	1.9	4.0	1%
M285T395	1.4	4.4	1.2	4.4	4.3	1.1	6.3	1.0	4.4	1%
M419T204	2.2	3.8	1.9	4.3	3.0	2.9	4.4	1.9	3.2	1%
M325T641	2.1	4.7	0.7	0.4	2.8	4.0	4.2	0.6	6.3	1%
M288T333	2.1	0.0	0.0	2.6	4.7	1.1	6.2	1.2	7.5	1%
M534T483	1.1	0.0	0.0	0.0	0.0	1.6	12.2	1.3	9.0	1%
M485T500	1.0	3.6	1.4	1.5	3.4	1.1	4.7	1.1	5.7	1%

Table A-12: Features within ROI 6 from MEDI Analysis

DESC	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	% ROI
M290T438	2.7	2.5	1.2	3.1	2.9	1.1	3.6	1.6	4.4	1%
M100T395	1.2	3.1	0.6	2.9	3.1	1.3	4.4	1.8	4.6	1%
M168T260	2.4	3.6	0.7	3.4	3.5	1.2	3.8	1.0	3.3	1%
M387T204	2.3	2.1	1.6	3.1	1.9	2.7	4.0	2.2	2.9	1%
M442T478	1.0	3.0	1.4	1.3	3.9	1.1	3.0	1.2	6.4	1%
M170T205	2.4	2.7	1.1	2.1	2.4	1.8	3.5	1.6	4.1	1%
M556T489	1.2	0.0	0.0	0.0	0.0	1.6	8.9	3.1	6.9	1%
M760T493	0.9	4.0	0.8	1.4	5.0	0.9	3.4	1.1	3.4	1%
M84T396	1.6	3.3	1.8	2.3	2.6	1.6	2.8	1.1	3.3	1%
M513T509	0.7	3.5	1.2	0.9	1.3	1.2	2.6	1.0	5.7	1%
M518T648	0.9	2.3	0.5	2.2	2.5	0.2	1.9	0.5	3.2	1%
M307T641	1.5	2.4	0.9	0.7	1.6	0.7	2.4	0.6	3.1	1%
M817T201	1.2	1.9	0.8	1.7	1.4	1.5	2.2	1.1	1.7	1%
M405T678	1.2	2.2	1.0	0.9	1.3	0.8	2.1	0.7	2.5	1%
M653T424	0.6	0.0	0.0	1.5	1.7	1.1	7.5	0.0	0.0	1%
M631T648	0.7	1.8	0.5	0.2	2.1	1.2	1.6	0.2	2.5	1%
M542T436	0.8	0.0	0.0	1.3	1.5	1.2	2.1	1.2	2.0	0%
M483T403	0.7	0.0	0.0	1.3	1.5	1.2	3.7	0.0	0.0	0%
M500T381	0.5	0.0	0.0	0.8	2.0	0.7	2.2	0.6	1.5	0%
M353T346	0.7	0.0	0.0	0.6	2.2	0.4	2.2	0.3	1.7	0%
M406T387	0.5	0.0	0.0	1.2	1.6	0.7	2.2	0.0	0.0	0%
M752T432	0.4	0.0	0.0	1.1	1.1	0.8	2.3	0.0	0.0	0%
M625T343	0.5	0.0	0.0	0.0	0.0	0.2	1.8	0.2	1.9	0%
M762T984	0.2	0.0	0.0	0.0	0.0	0.3	1.4	0.4	1.7	0%

Table A-13: Tentatively identified features from MEDI regions of interest**.** Criteria for identification were as follows. Mass difference was less than 30 ppm, and compound was of microbial origin. When multiple possible identifications were possible the lower ppm match was selected. All matched were made using the Metlin database with the exception of the ciromicins which were experimentally identified. Only features > 1% of the ROI were selected for tentative identification.

Accurate mass	Noc Mono	BS&Noc	BS Mono	RW Mono	RW&Noc	EC Mono	EC&Noc	TP Mono	TP&Noc	Tentative ID	Δppm	ROI
415.3553	536.8	508.2	7.0	7.5	881.7	4.4	385.5	2.4	643.4	hexacosadienoic acid	1	1
214.1111	12.6	19.9	6.6	2.2	19.9	5.4	20.8	0.9	17.4	N-Acyl homoserine lactone	17	1
397.3489	207.9	178.0	6.4	12.3	402.9	12.6	136.7	14.6	244.3	hopanoid	6	2
376.2832	0.2	0.0	0.0	0.2	281.8	0.0	0.0	0.3	253.6	N-acyl amide	2	3
384.2170	0.7	9.4	16.9	1.0	142.6	0.8	5.4	0.9	56.2	ciromicin aglycone	na	3
515.2740	0.5	5.1	1.2	0.4	34.2	0.8	2.8	0.8	31.7	Ciromicins	na	3
563.2980	1.6	2.0	0.7	1.4	61.8	0.6	1.9	1.2	47.6	His-Ile-Phe-Phe	0	3
273.1669	19.6	61.9	23.9	33.1	16.8	12.4	21.0	15.3	36.3	Dimethylallyltryptophan	26	4
331.2116	13.7	38.4	20.1	25.3	13.3	7.9	27.7	20.1	33.4	Gly Val Arg tripeptide	8	4
229.1582	7.4	38.0	11.2	4.3	9.6	11.1	18.7	4.2	17.8	Pro Leu	15	4
662.3517	1.4	17.3	0.8	1.6	7.0	1.4	5.4	1.6	6.0	Moxidectin	22	4
447.2904	3.3	7.5	2.4	2.7	2.2	1.9	7.3	2.9	7.5	Lys Lys Ala Thr	4	4
239.1760	269.2	295.9	9.5	459.2	284.4	371.8	270.2	187.5	365.0	sesquiterpenoid	4	5
225.1600	98.3	131.8	8.9	181.8	96.4	103.7	104.8	71.3	128.2	sesqueterpenoid	6	5
152.0610	22.3	6.4	14.0	17.6	3.5	4.0	31.7	2.4	29.9	Guanine	28	6
268.1053	14.2	27.7	1.9	16.8	19.4	6.7	21.7	11.2	37.1	Deoxyguanosine	4	6
208.0979	3.1	19.5	1.4	13.5	21.0	2.7	26.2	1.2	25.0	N-Acetyl-phenylalanine	5	6
284.1093	5.3	36.1	1.6	37.3	38.9	1.3	57.8	1.1	39.0	Glutamyl-Histidine	7	6

XCMS, meta XCMS and MEDI processing workflow

XCMS Processing

ProteoWizard converted mzXML files were divided into two folders, “control” and “test”. Monoculture NF extracts were placed in the control, while bicultures and competitor monocultures were placed in the test folder. These two folders were placed inside another folder which was designated as the directory for XCMS processing in R.

The following commands were used in the XCMS processing and alignment:

```
> library (xcms)
> xset<-xcmsSet()
> xset<-group(xset)
> xset2<-retcor(xset, family=”s”, plottype=”m”)
> xset2<-group(xset2, bw=30)
> xset3<-fillPeaks(xset2)
> reporttab<-diffreport(xset3, “control”, “test”, 10, metlin=0.15)
```

Meta XCMS Processing

Output files from XCMS for each co-culture experiment were imported into MetaXCMS 0.1.20

Figure A-23: MetaXCMS screenshot of imported files. Practically all features were retained through the selection of highly lenient filter criteria.

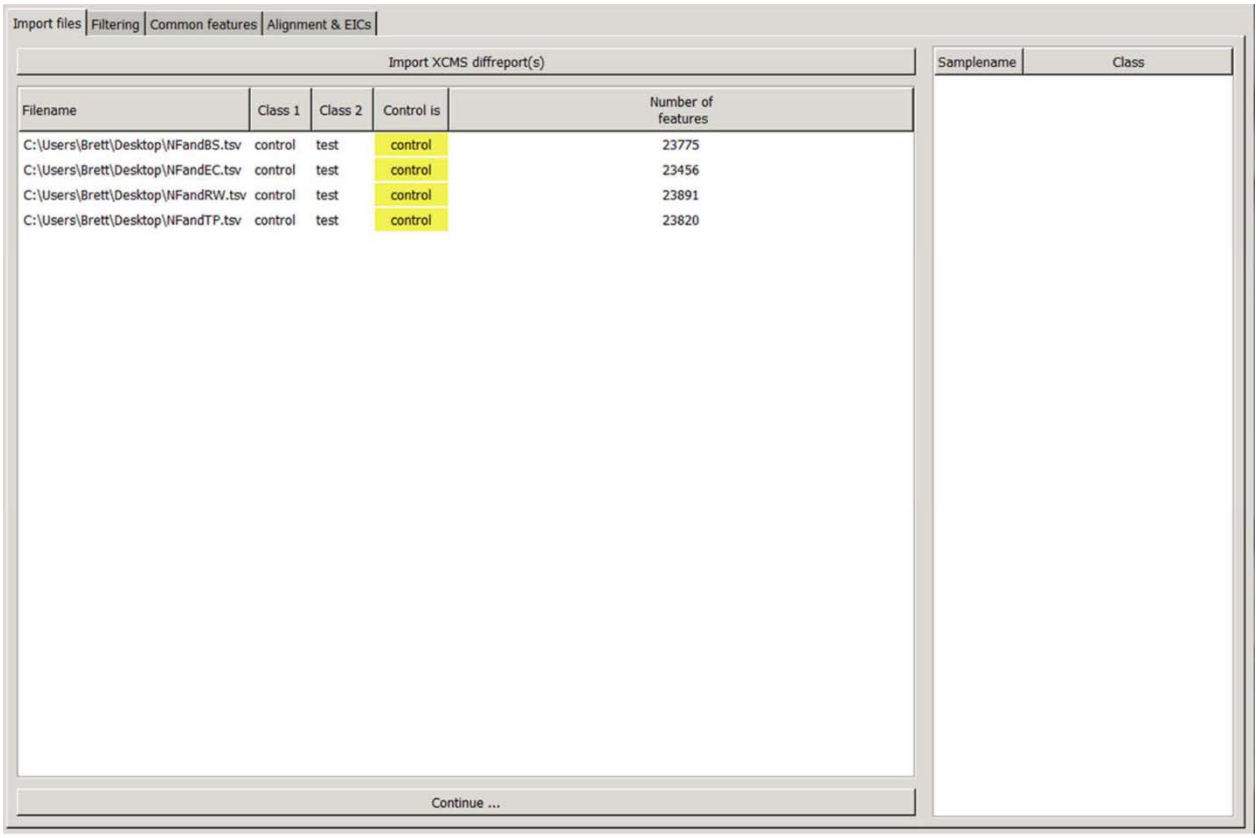


Figure A-24: MetaXCMS filtering parameters. *m/z* tolerance of 0.1 and retention time tolerance of 60 sec was imposed for combining feature lists.

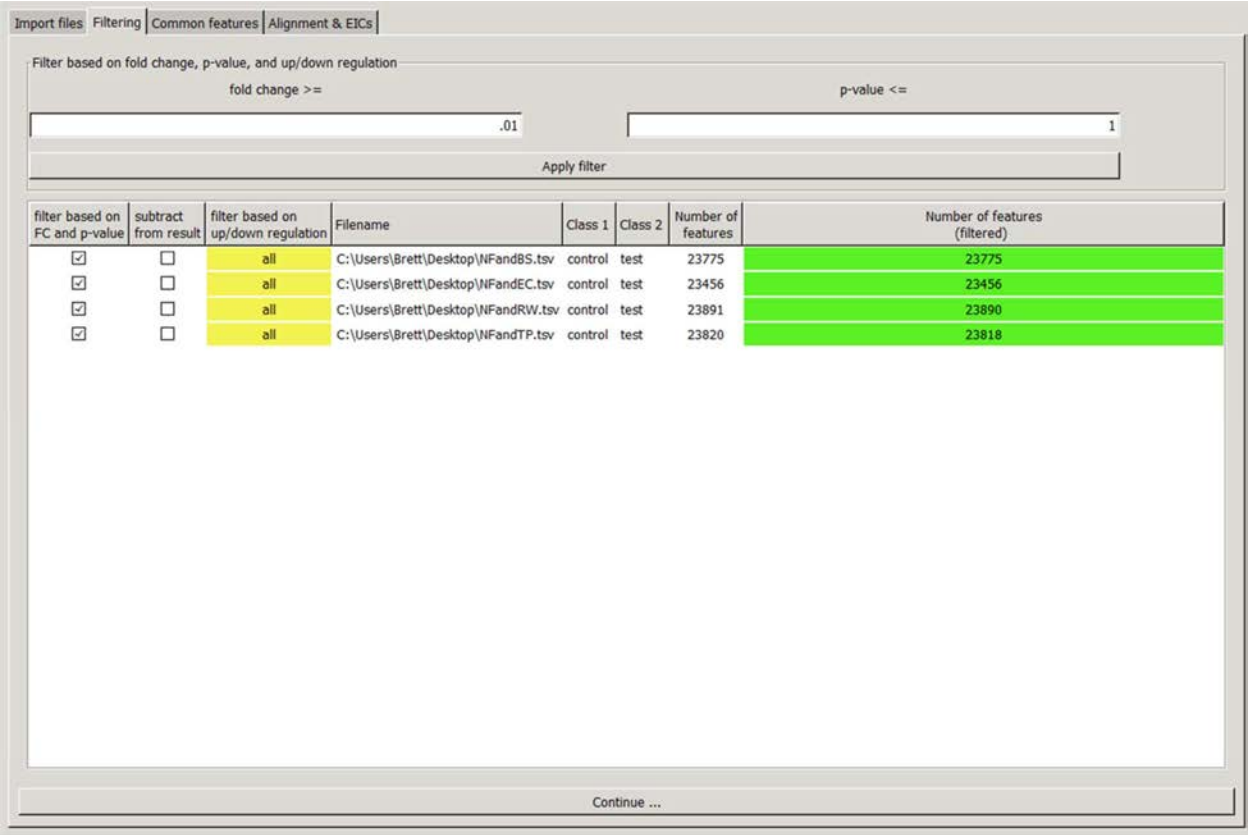
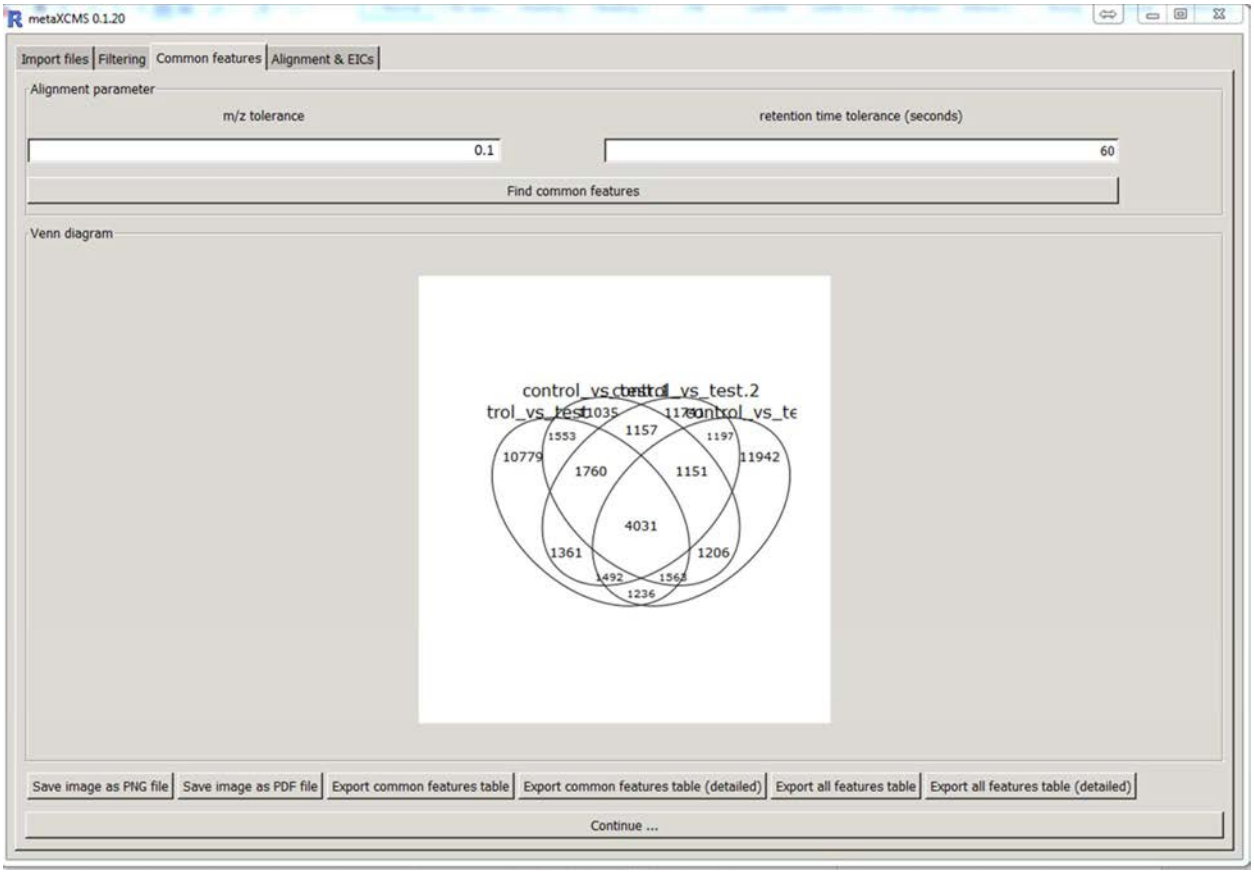


Figure A-25: MetaXCMS alignment parameters

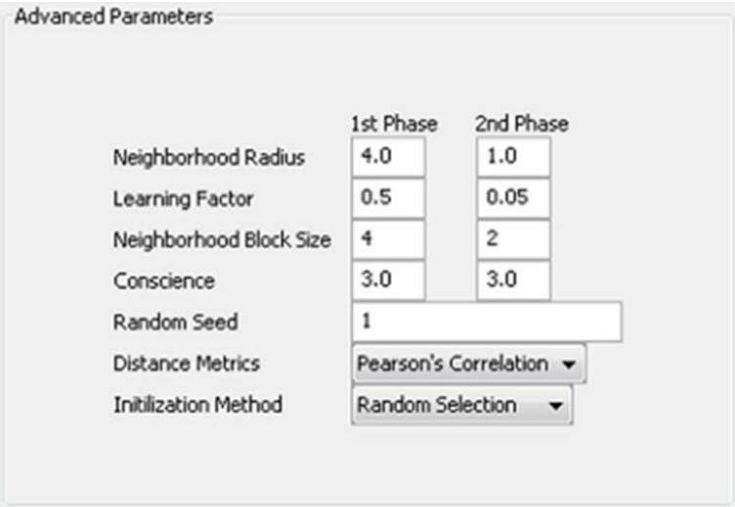


A detailed table was then exported and edited in excel. Features not present in samples are designated as n/a after exporting. These were changed to 1 such that they would not cause errors during mathematical, statistical processing steps. It should be noted that this value becomes very small after normalization. Features were normalized by a total ion count on 10,000, and low intensity features were checked against the raw data to ensure legitimacy. Features not discernible from noise and features after 1000 seconds (during column wash and re-equilibration) were removed from the dataset. Feature intensities from each biculture were compared to their respective monocultures and sorted by fold change to select for up and down regulated features.

MEDI Analysis

The combined dataset contained 1693 features after filtering. The dataset was formatted for MEDI analysis as shown in supporting data tables: MEDI input format. This was then saved as a Text (Tab delineated) file and uploaded into the GEDI software. A grid of 25 x 26 nodes was selected, with 1st and 2nd phase training iterations of 80 and 100 respectively. Advanced GEDI parameters are shown in figure below.

Figure A-26: MEDI advanced parameters



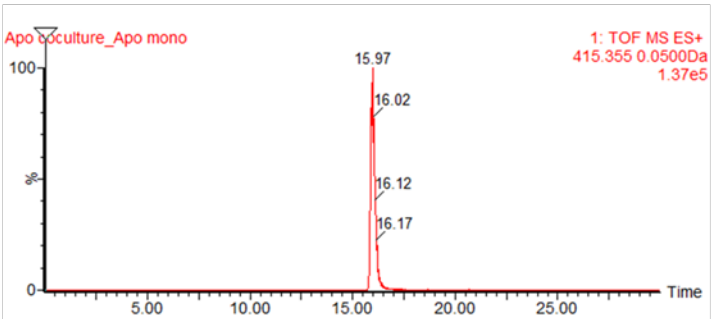
Using GEDI software the monoculture conditions were subtracted out from the bicultures to yield the difference maps shown in the manuscript. Regions of interest were selected from the difference maps as indicated in Figure 1 of the manuscript and section A.1. of the supporting. Features from these regions were extracted into excel, and each features contribution to the ROI was determined by calculating the total intensity of the ROI and dividing the total intensity of a feature across all conditions by the total feature intensity of the ROI. Features were then sorted by their contributions to the ROI, or % ROI to give Tables S1 to S9 shown above.

Extracted ion chromatograms

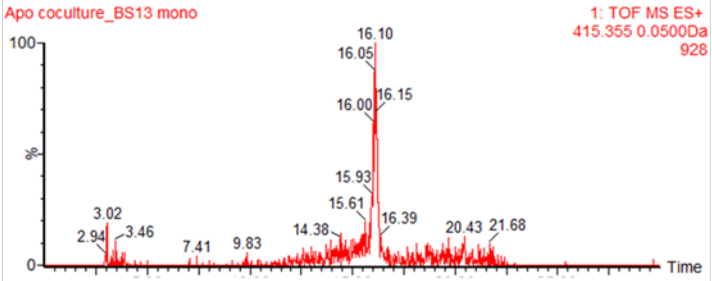
The following are extracted ion chromatograms from the raw data for selected abundant features prioritized in regions of interest from the MEDI analysis. These reveal that the features are indeed more abundant in the conditions specified. This is especially apparent in features from ROI 3 where we see that several prioritized m/z peaks (including the ciromicin peaks with m/z 515 and 384) are only detectable in the coculture conditions and not in any of the monocultures.

ROI: 1
M415T959

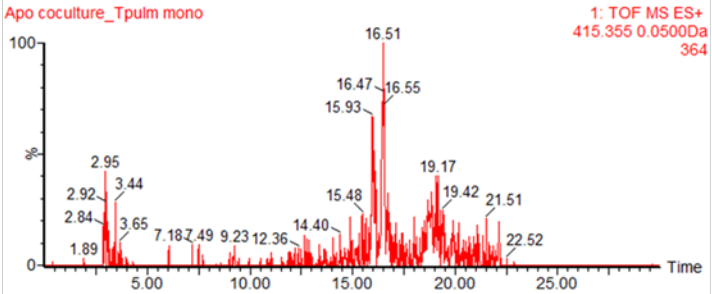
NF



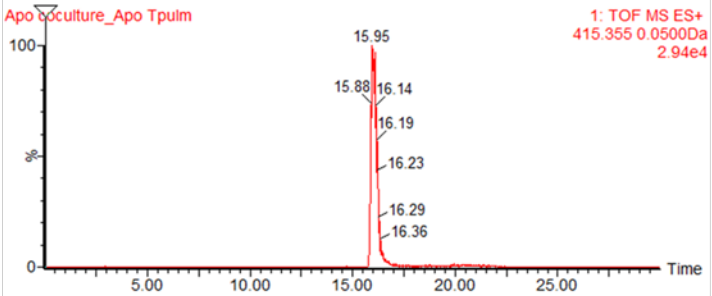
TP



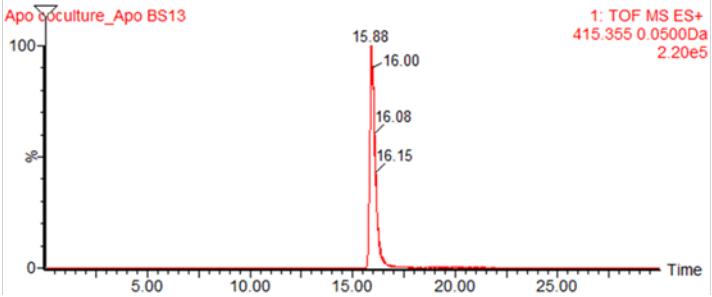
RW



NF&TP

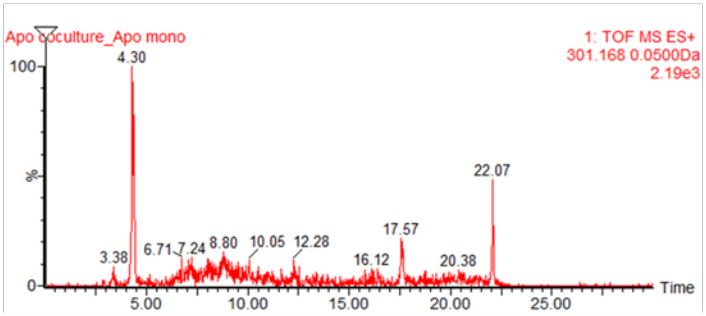


NF&RW

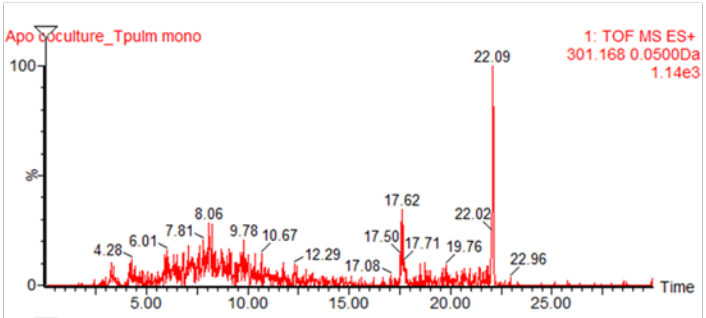


ROI: 1
M301T259

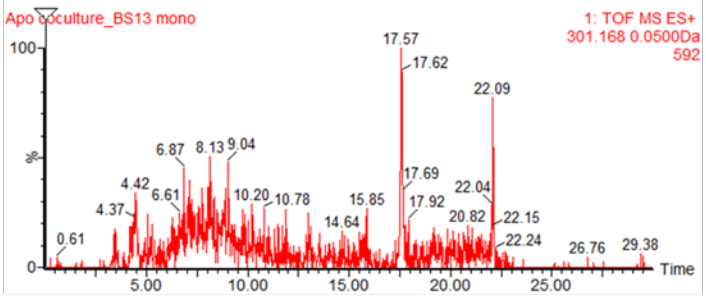
NF



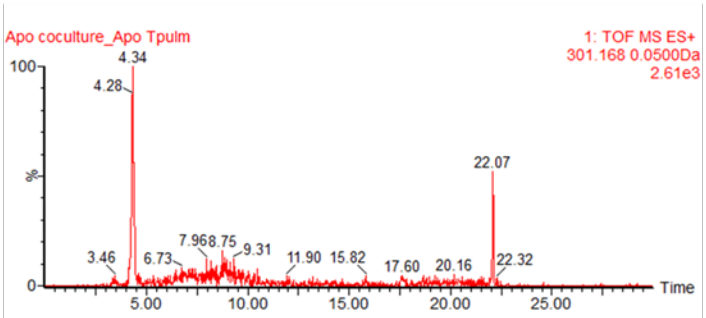
TP



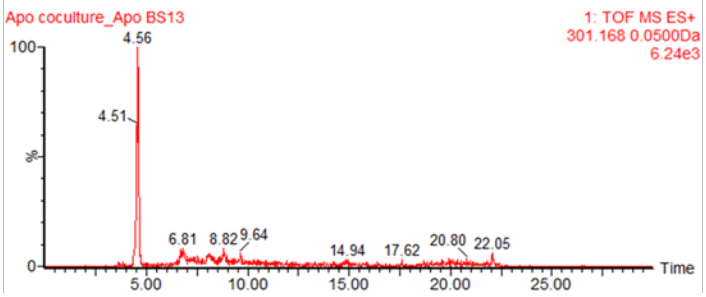
RW



NF&TP

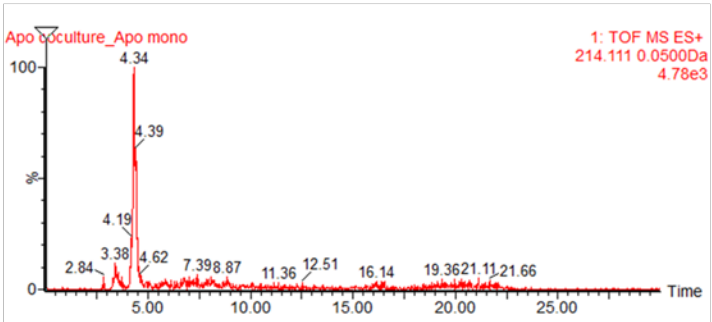


NF&RW

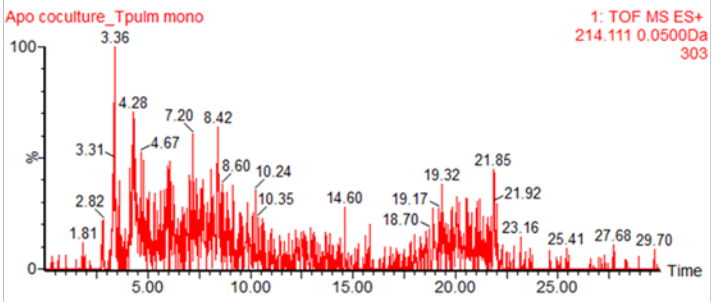


ROI: 1
M214T295

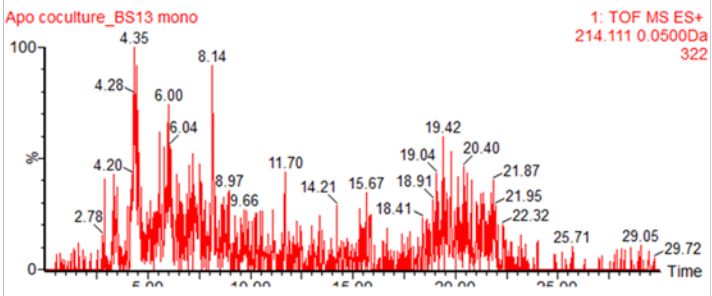
NF



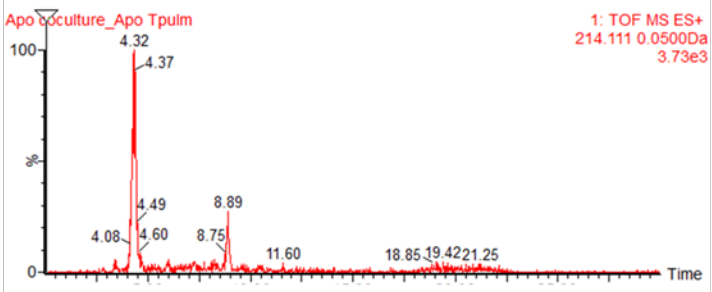
TP



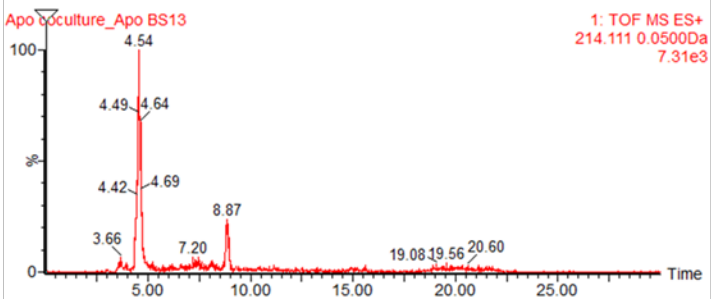
RW



NF&TP

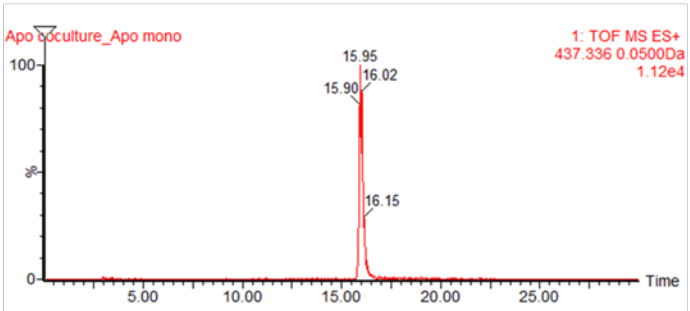


NF&RW

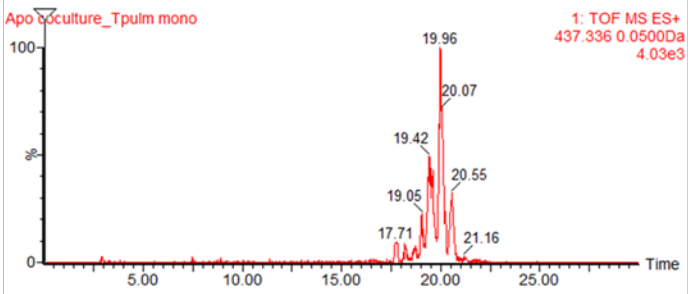


ROI: 1
M437T959

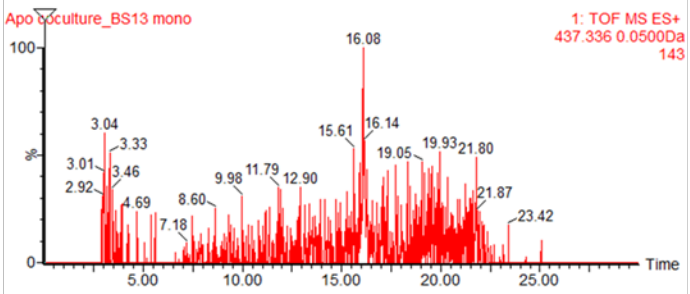
NF



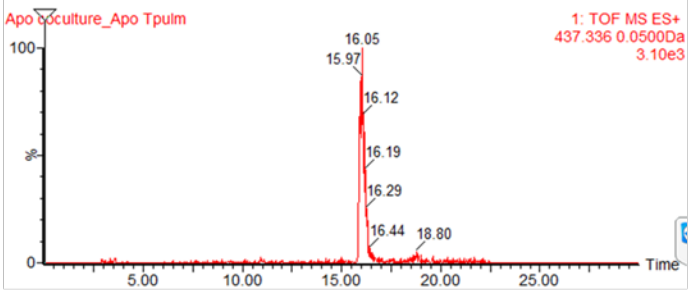
TP



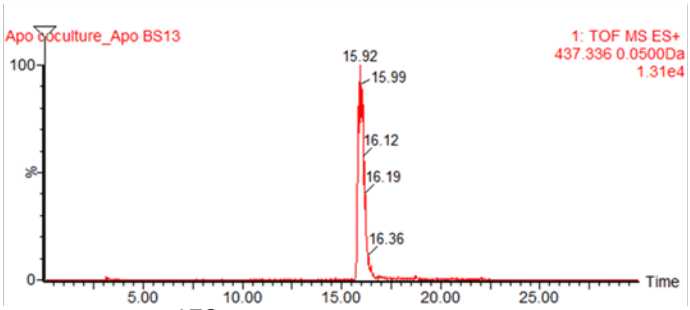
RW



NF&TP

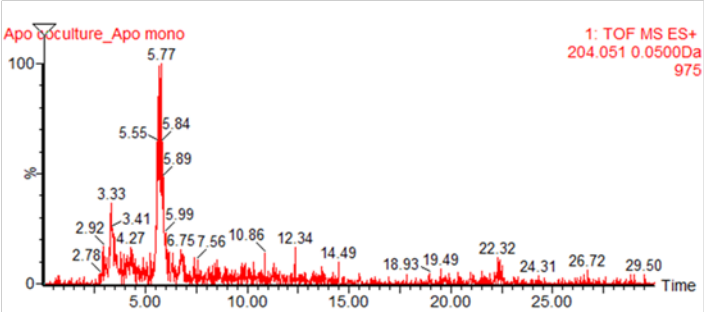


NF&RW

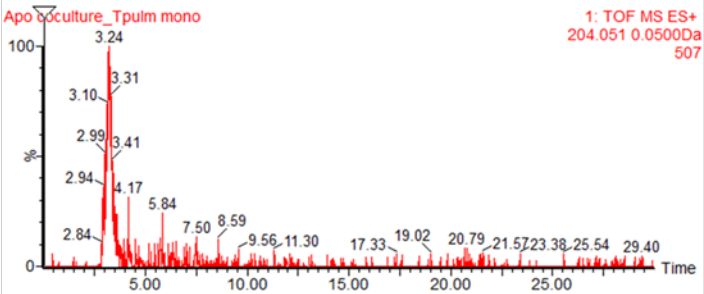


ROI: 1
M204T345

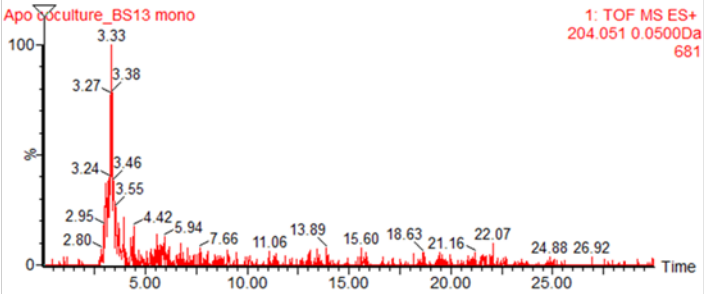
NF



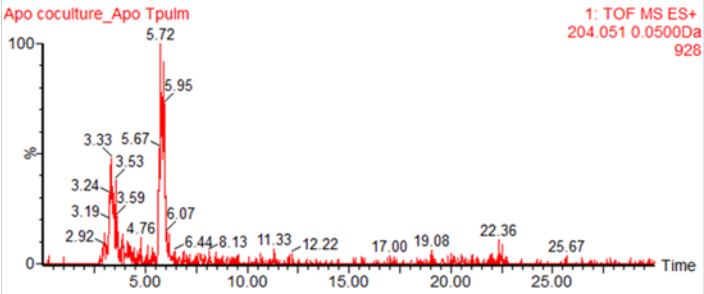
TP



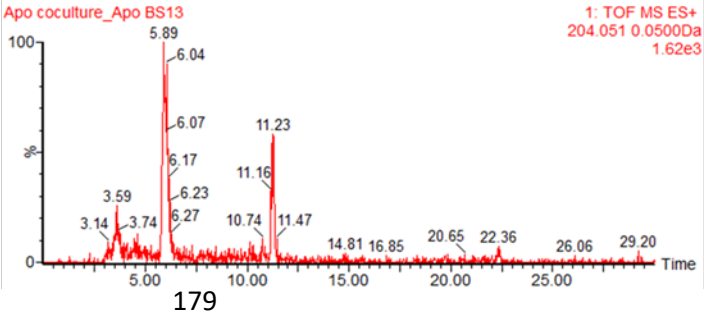
RW



NF&TP

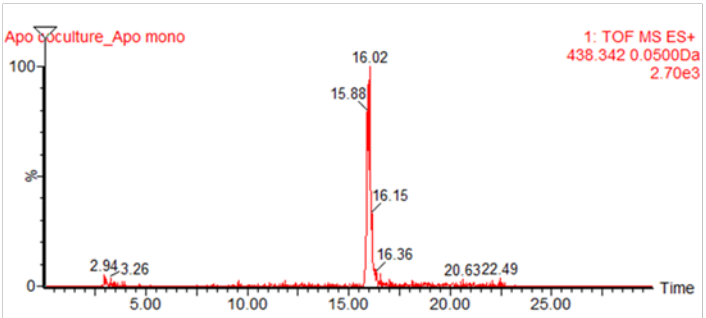


NF&RW

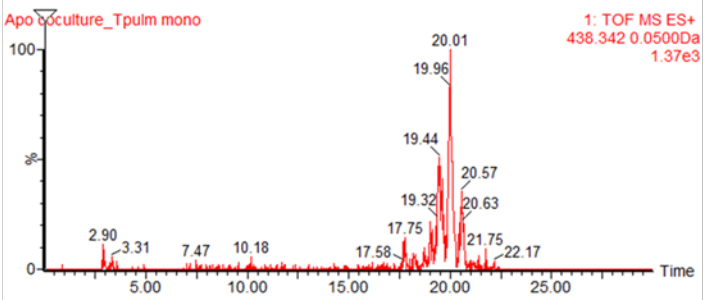


ROI: 1
M438T959

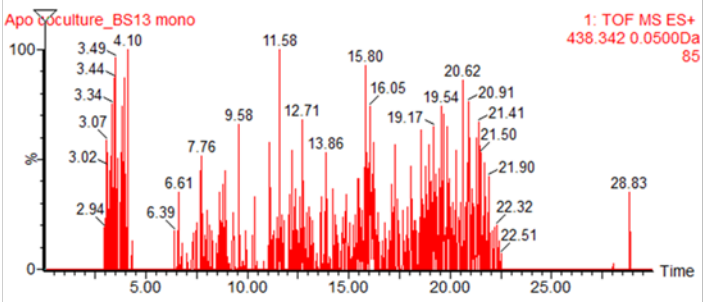
NF



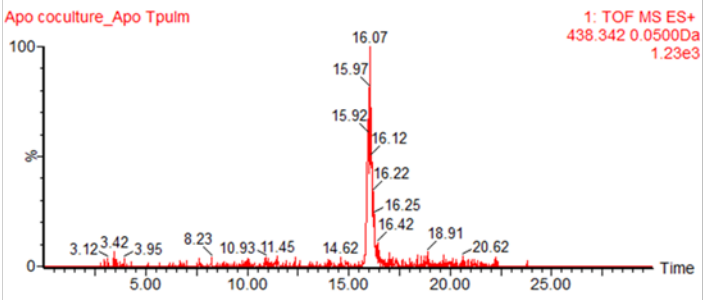
TP



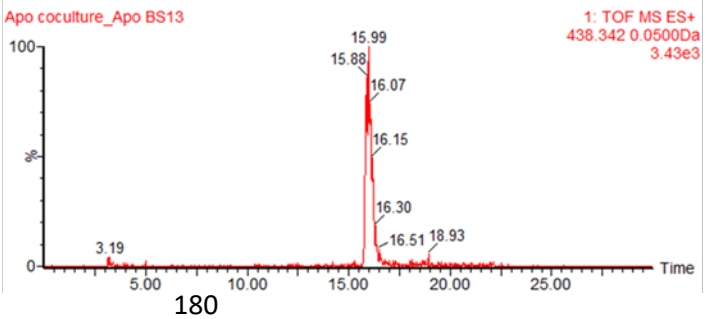
RW



NF&TP

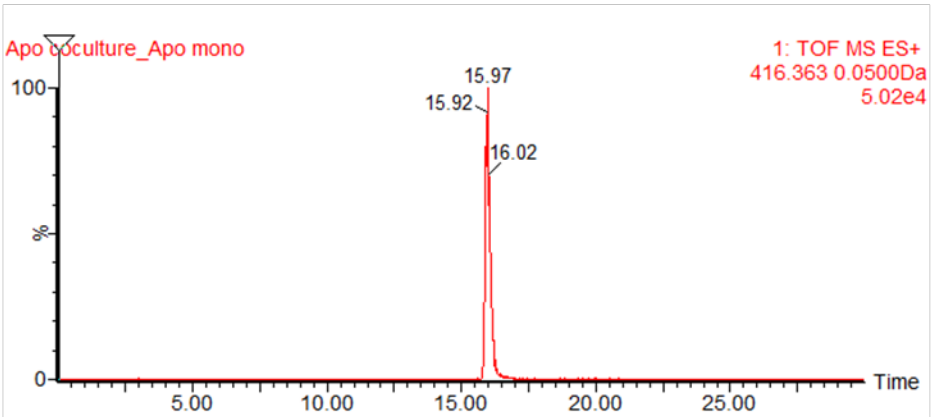


NF&RW

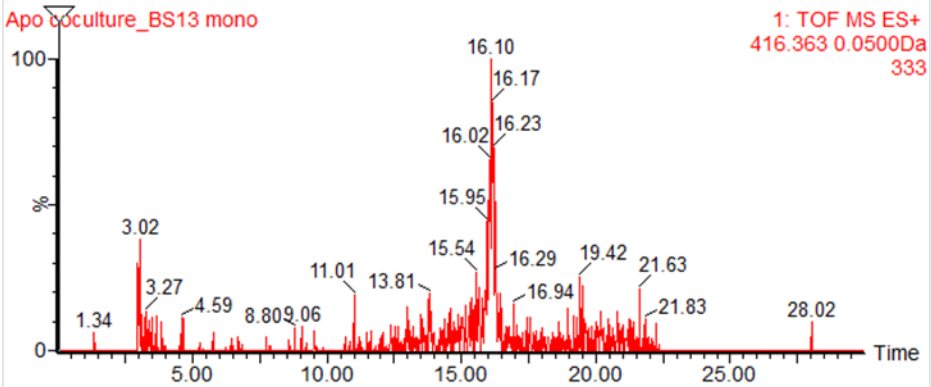


ROI: 2
M416T958

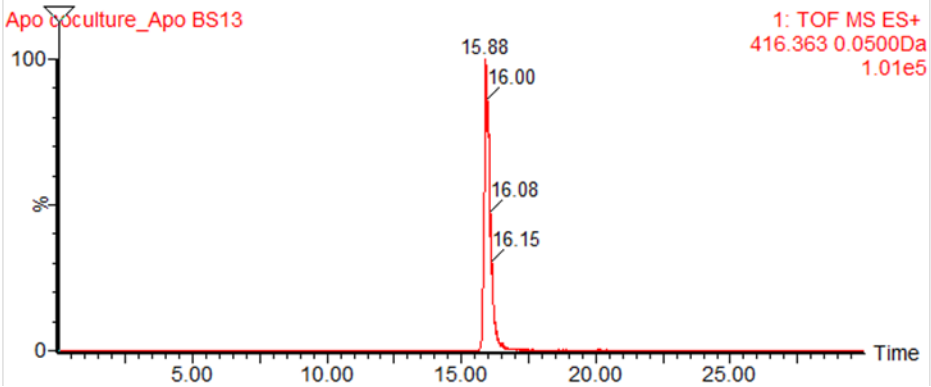
NF



RW

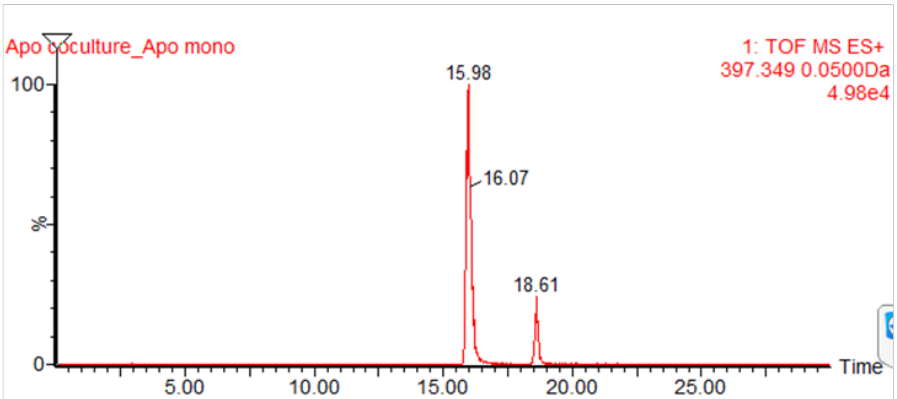


NF&RW

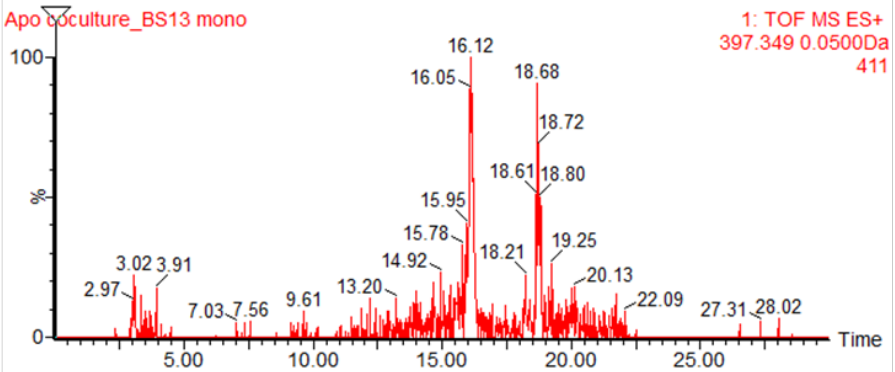


ROI: 2
M397T958

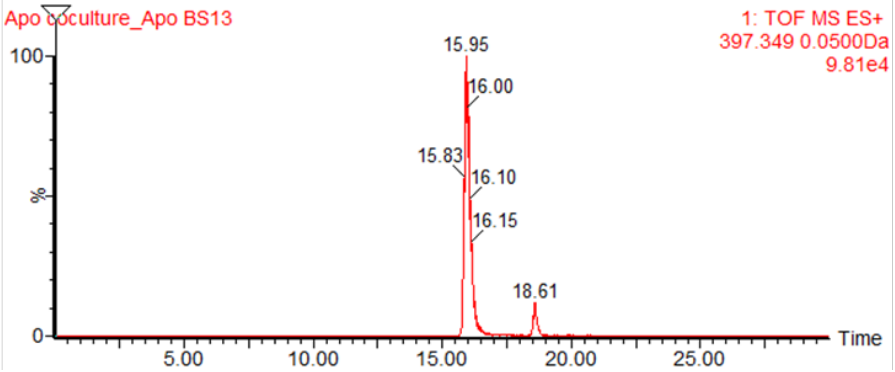
NF



RW

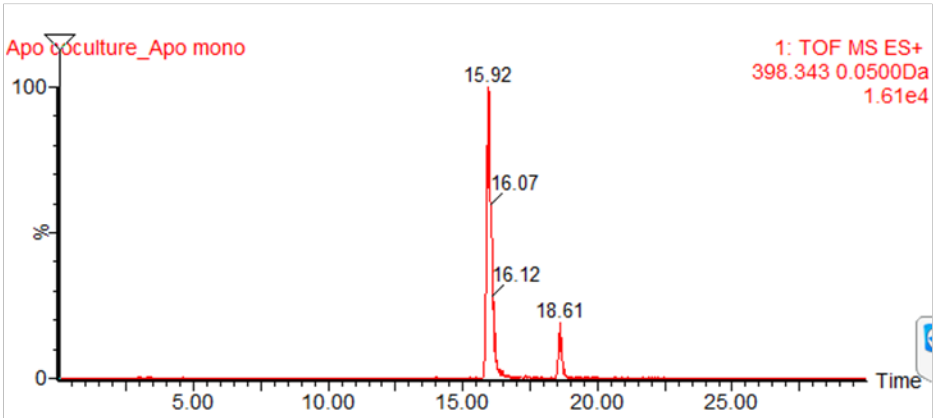


NF&RW

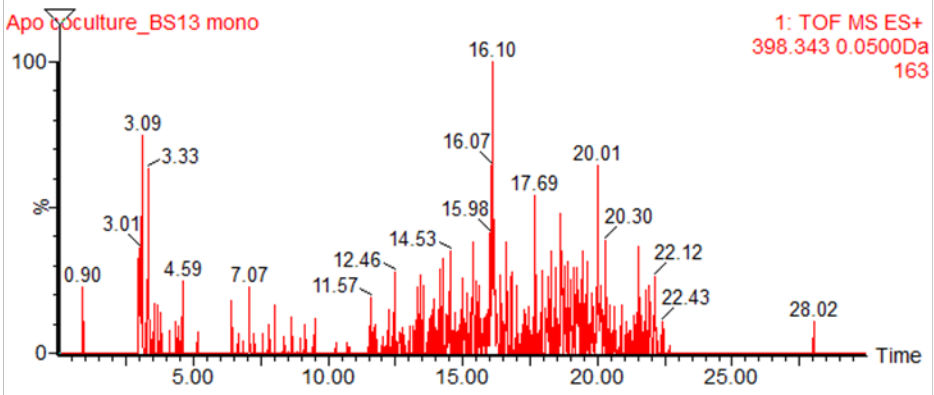


ROI: 2
M398T958

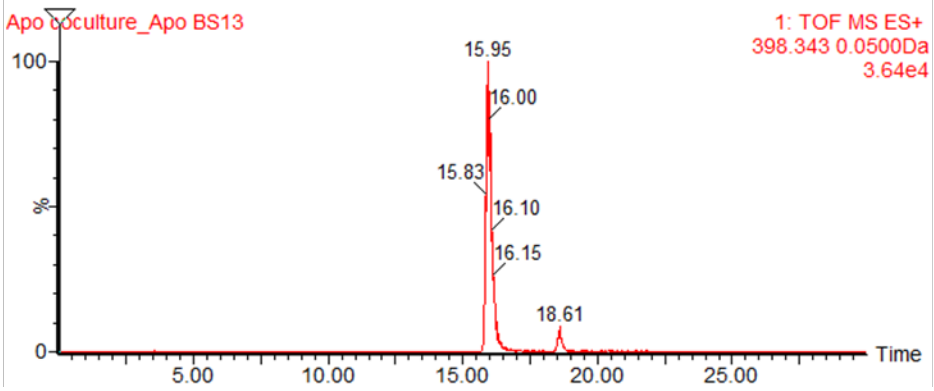
NF



RW

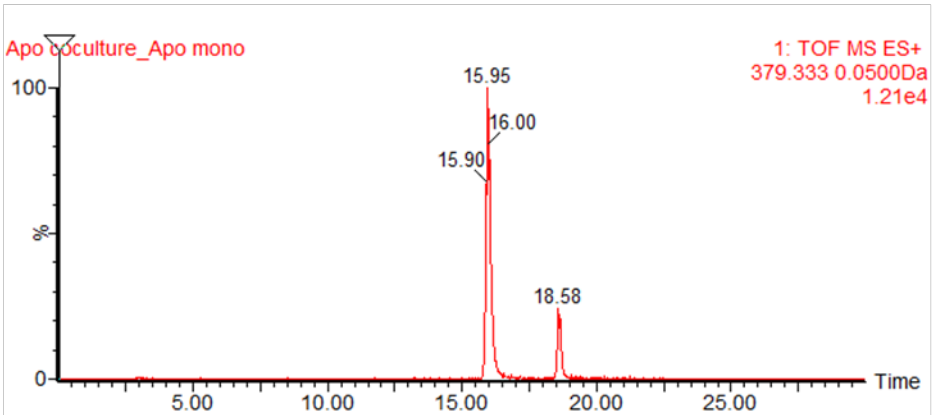


NF&RW

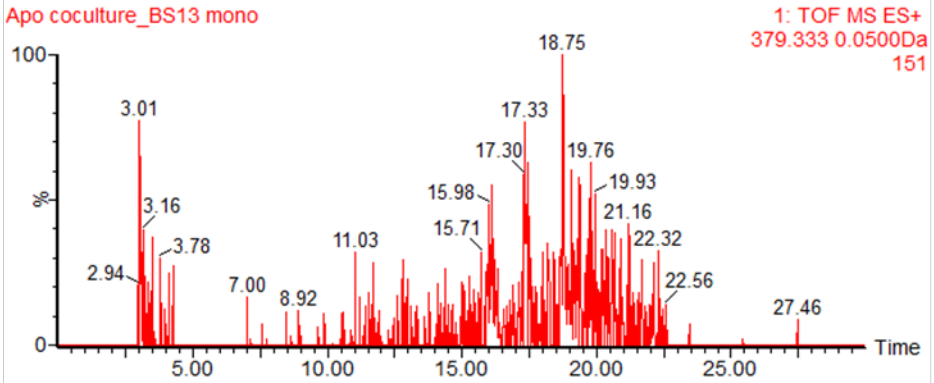


ROI: 2
M379T958

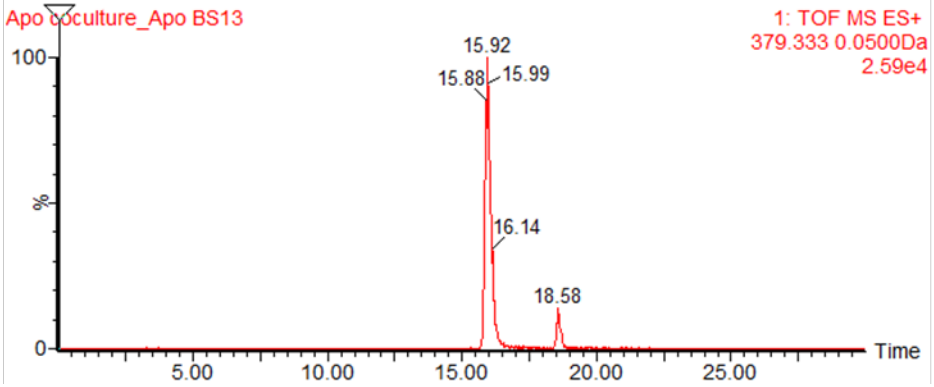
NF



RW

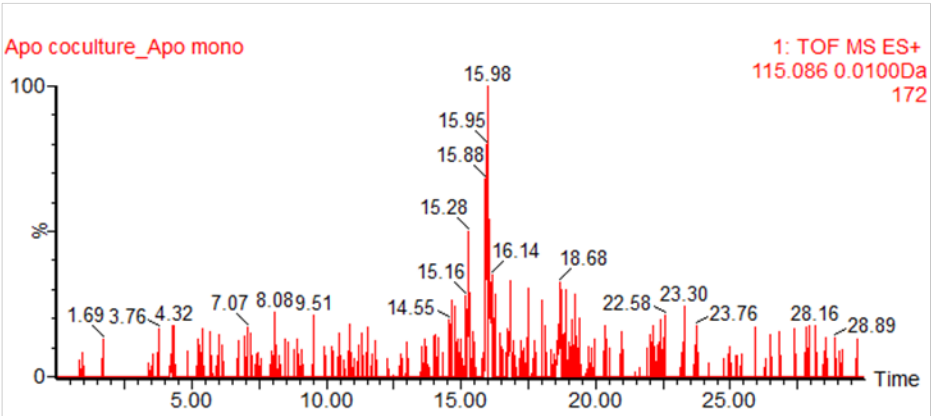


NF&RW

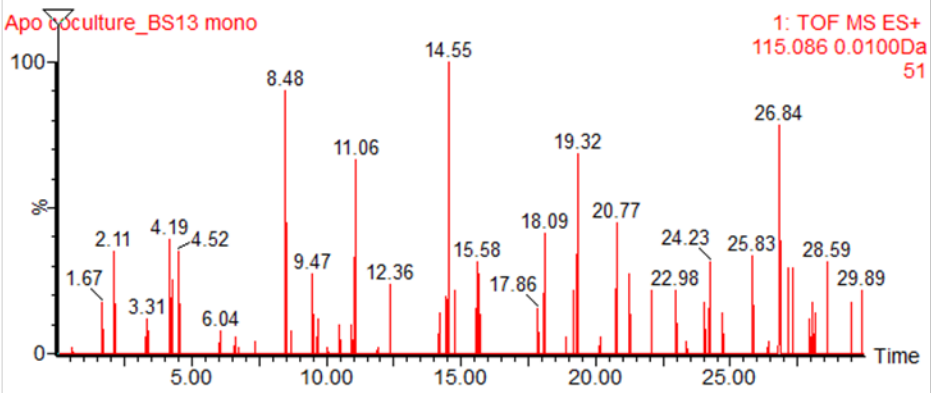


ROI: 2
M115T958

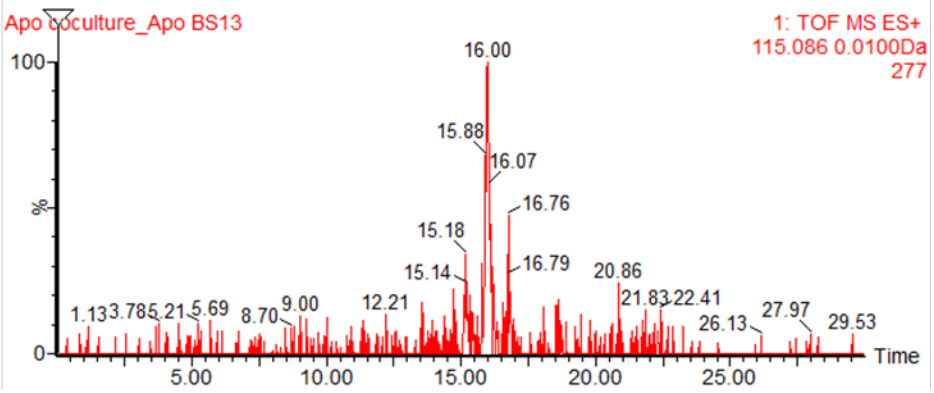
NF



RW

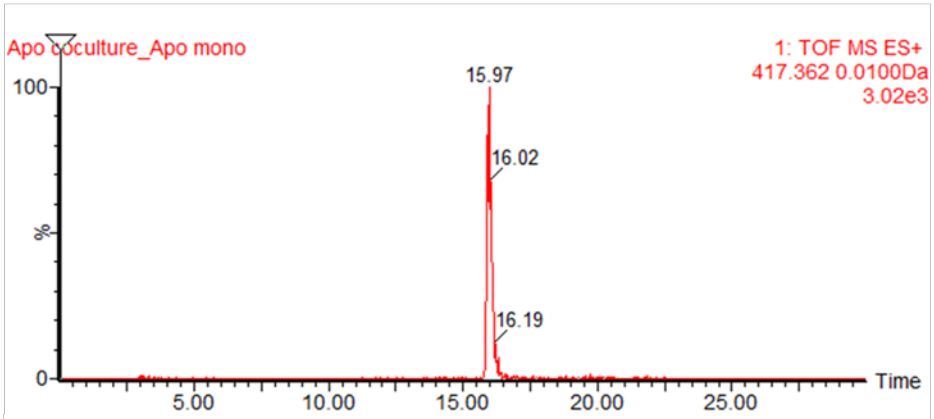


NF&RW

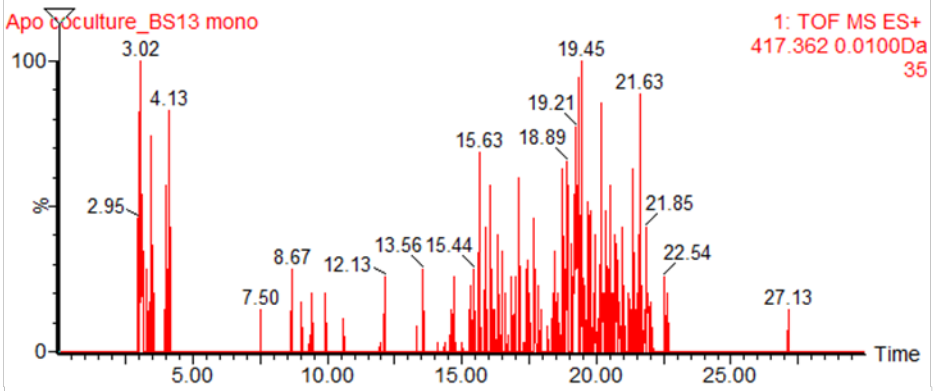


ROI: 2
M417T958

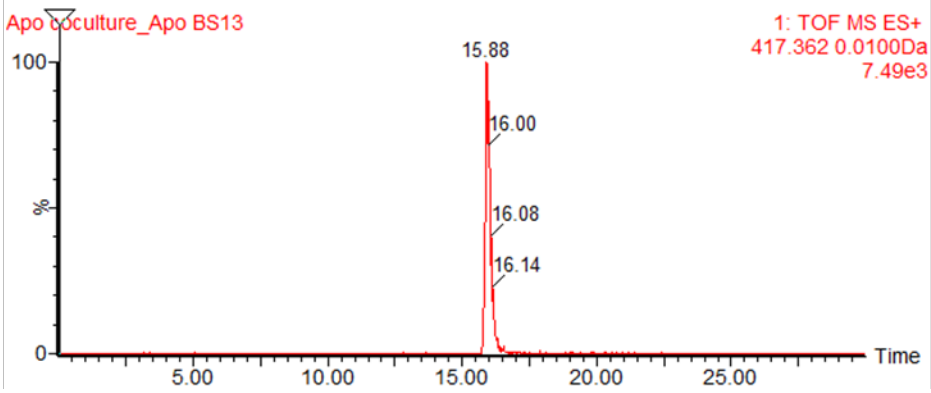
NF



RW

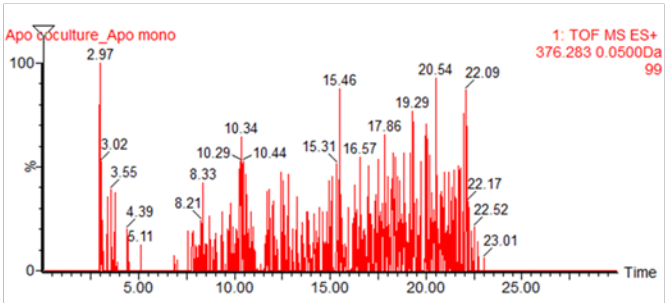


NF&RW

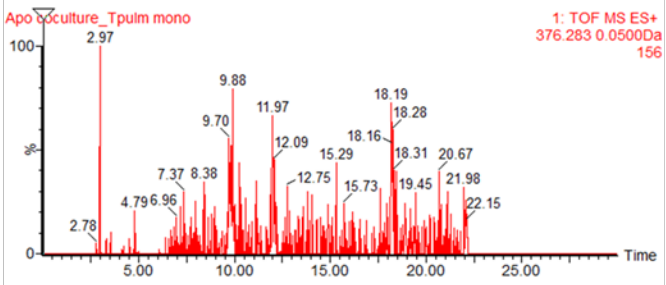


ROI: 3
M376T897

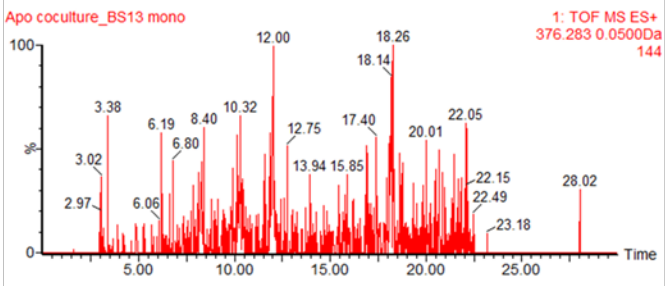
NF



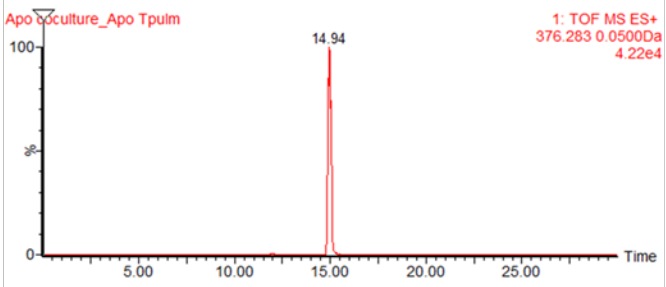
TP



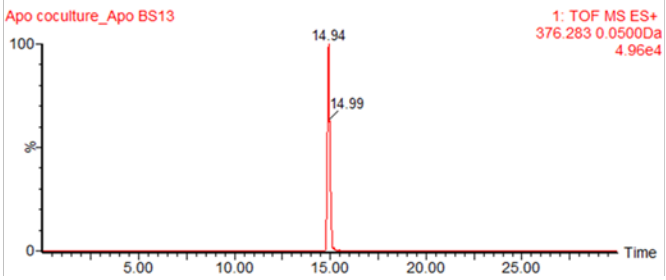
RW



NF&TP

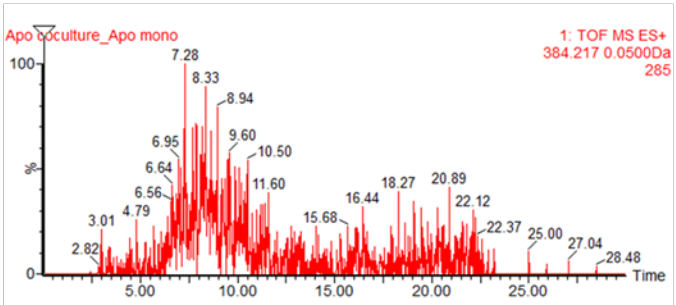


NF&RW

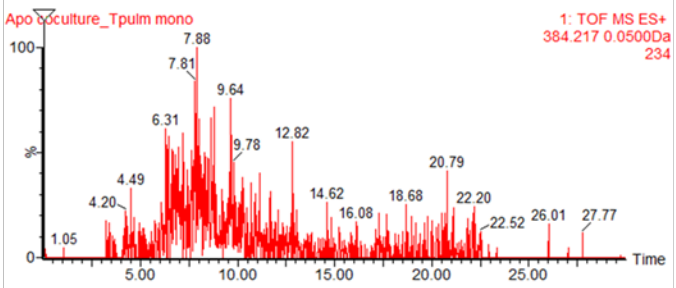


ROI: 3
M384T609

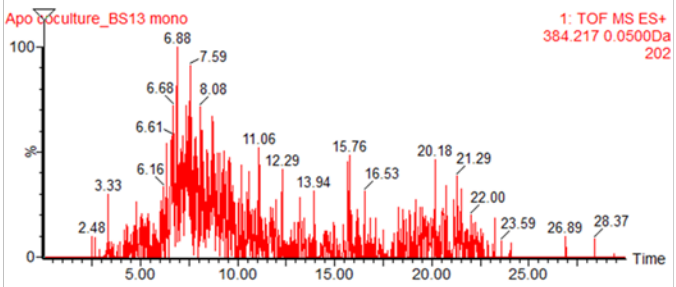
NF



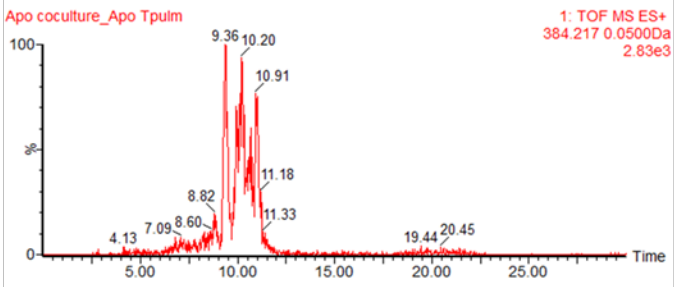
TP



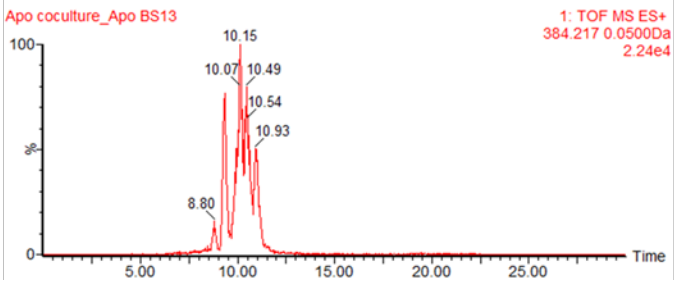
RW



NF&TP

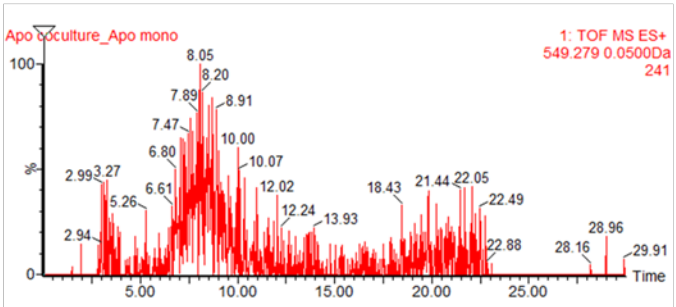


NF&RW

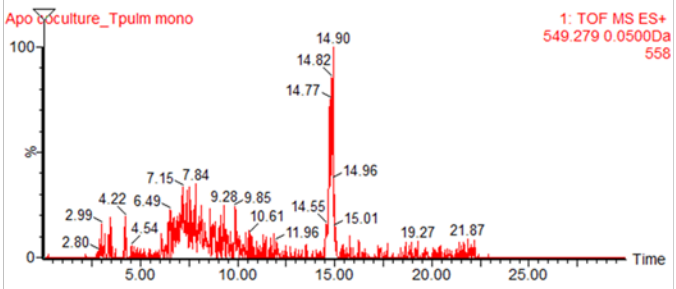


ROI: 3
M549T472

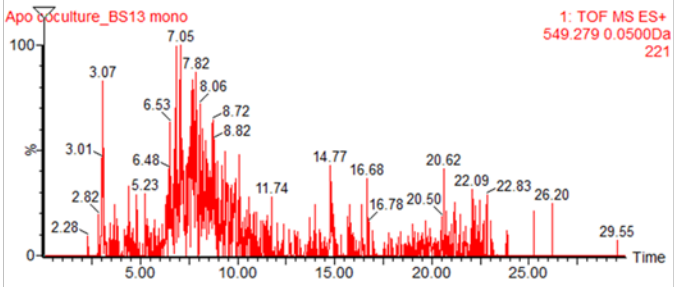
NF



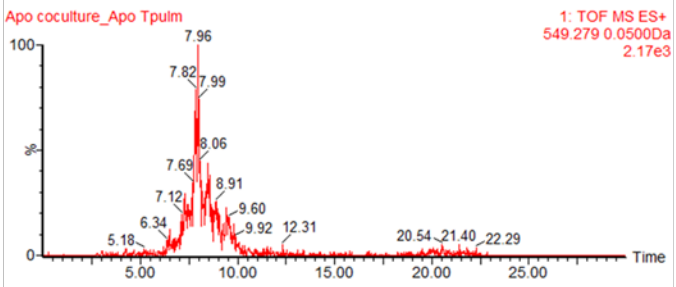
TP



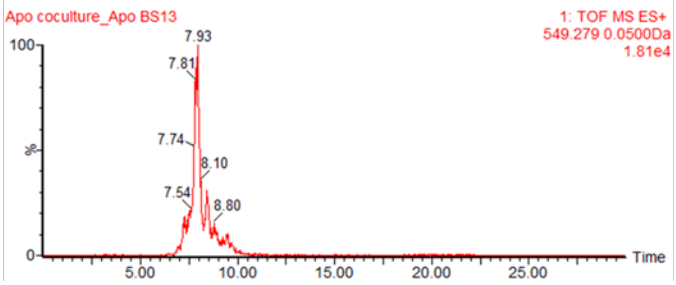
RW



NF&TP

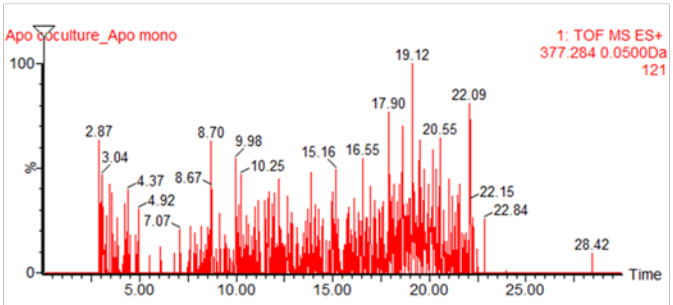


NF&RW

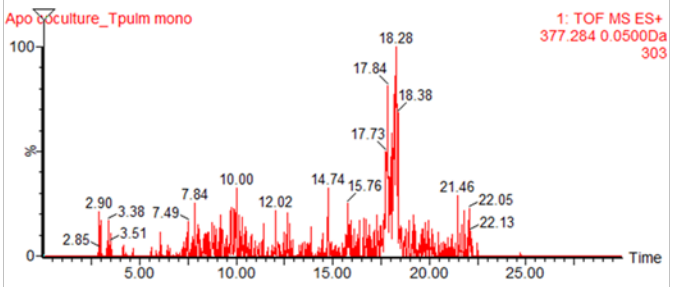


ROI: 3
M377T896

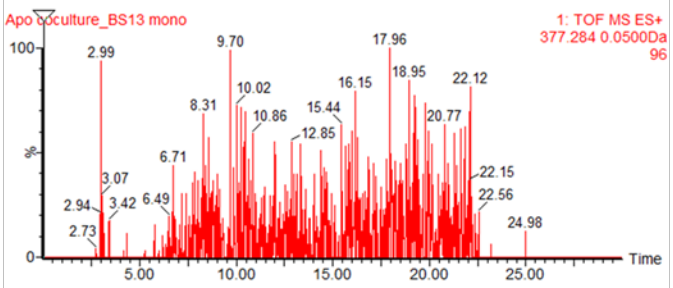
NF



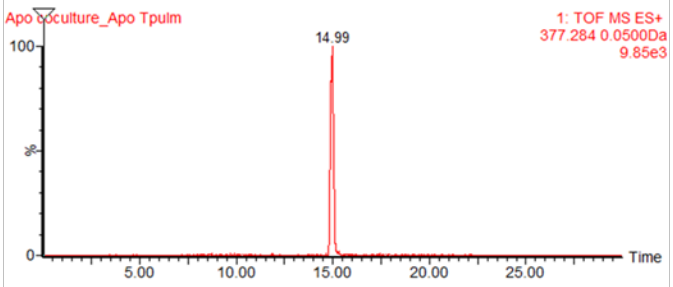
TP



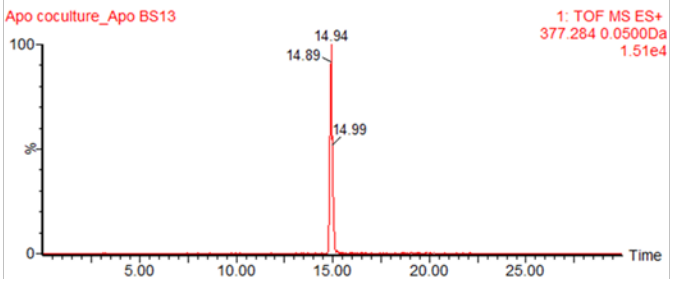
RW



NF&TP

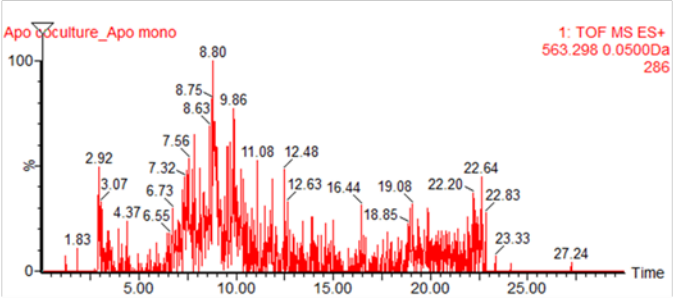


NF&RW

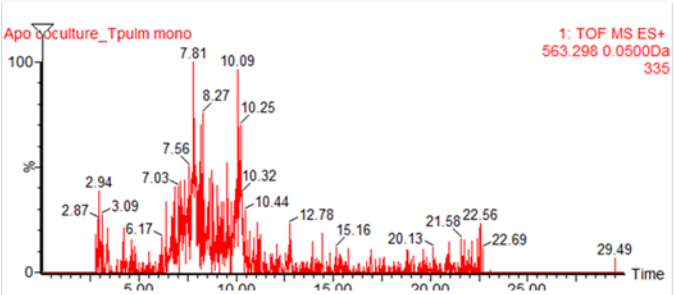


ROI: 3
M563T514

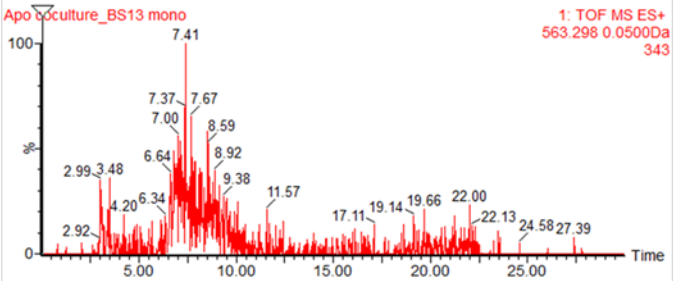
NF



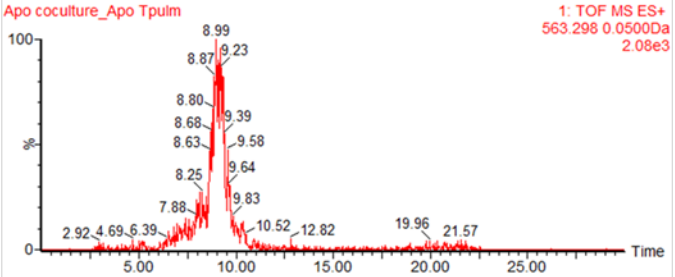
TP



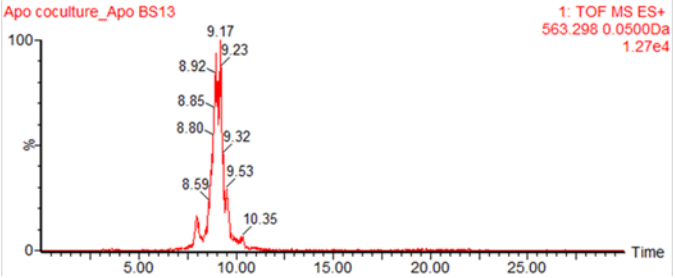
RW



NF&TP

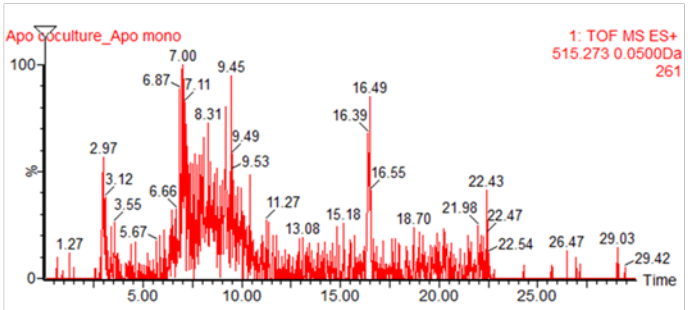


NF&RW

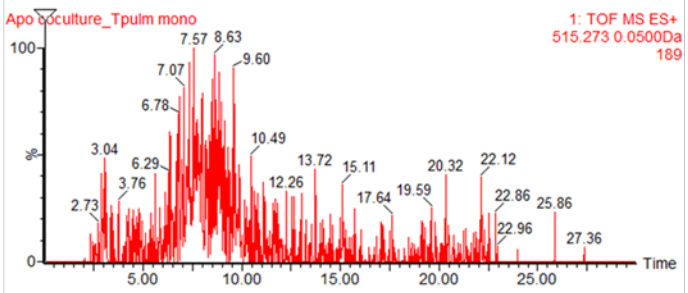


ROI: 3
M515T602

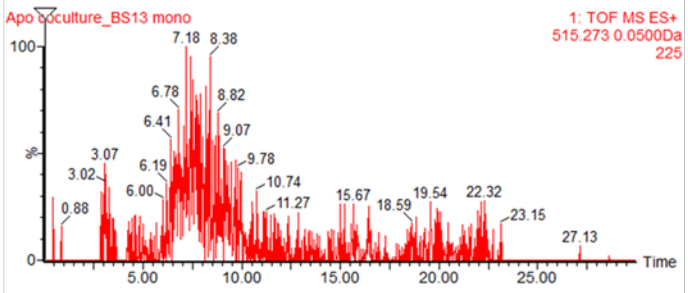
NF



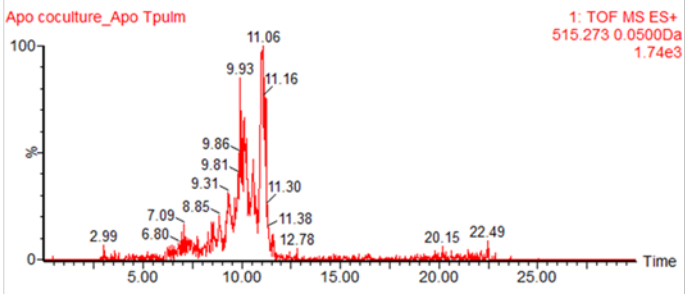
TP



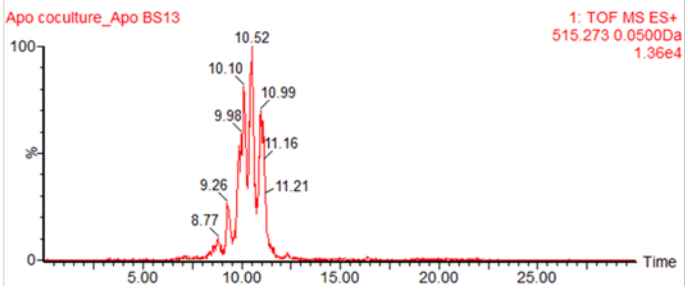
RW



NF&TP

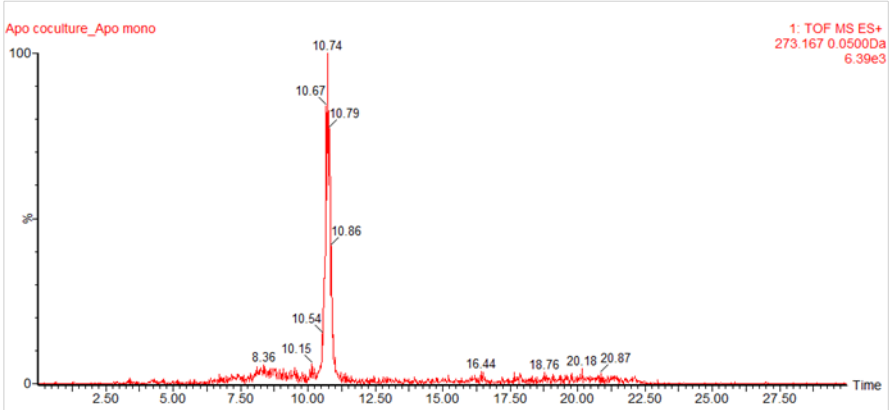


NF&RW

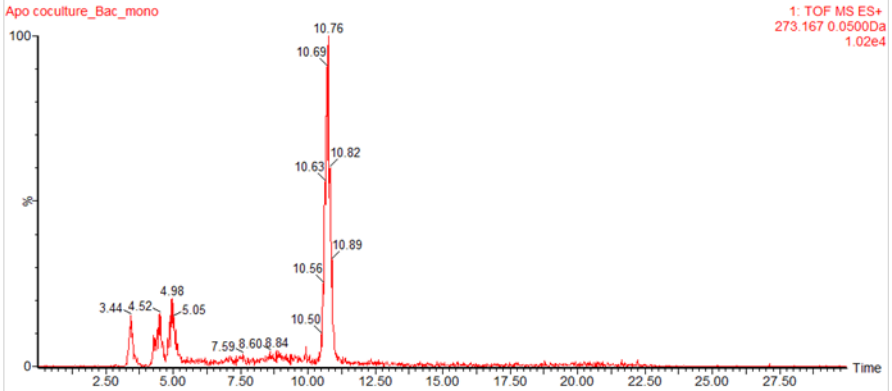


ROI: 4
M273T644

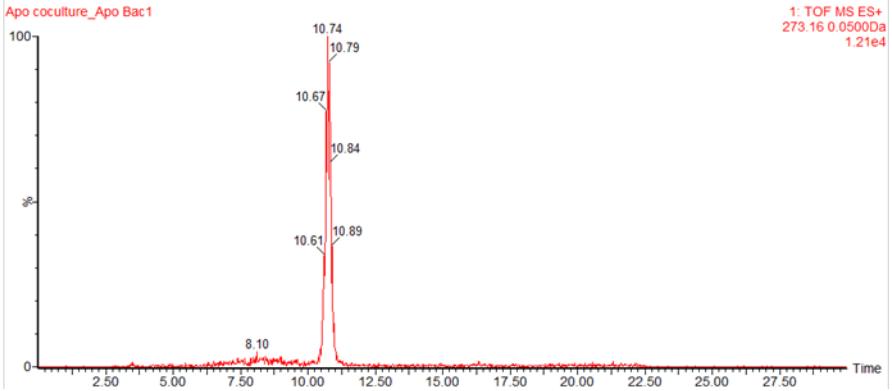
NF



BS

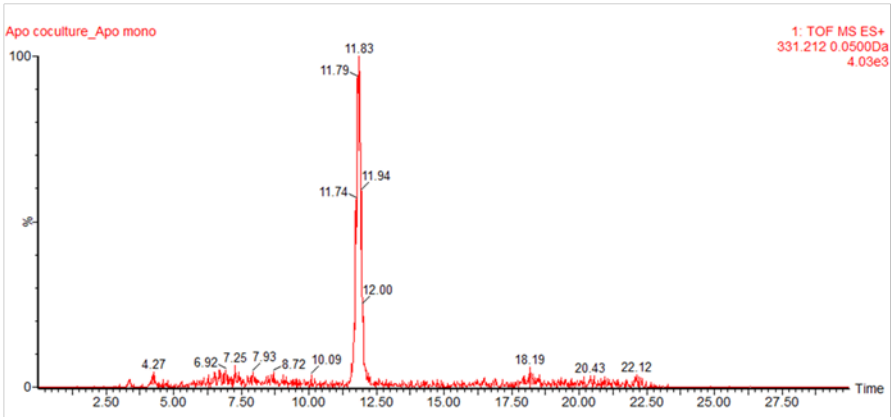


NF&BS

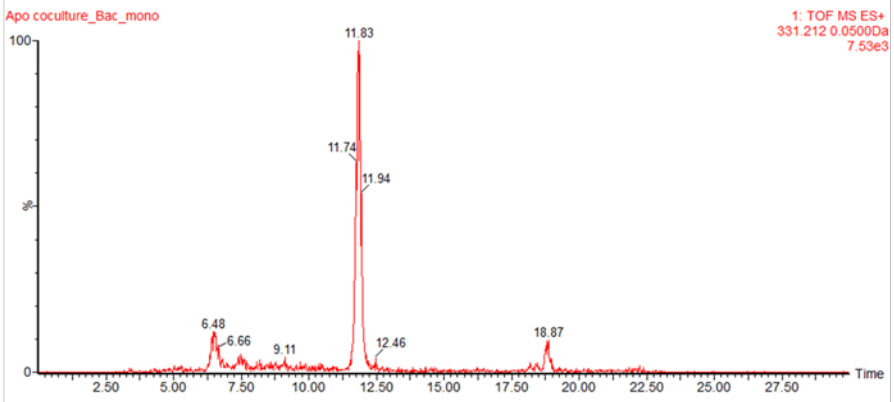


ROI: 4
M331T710

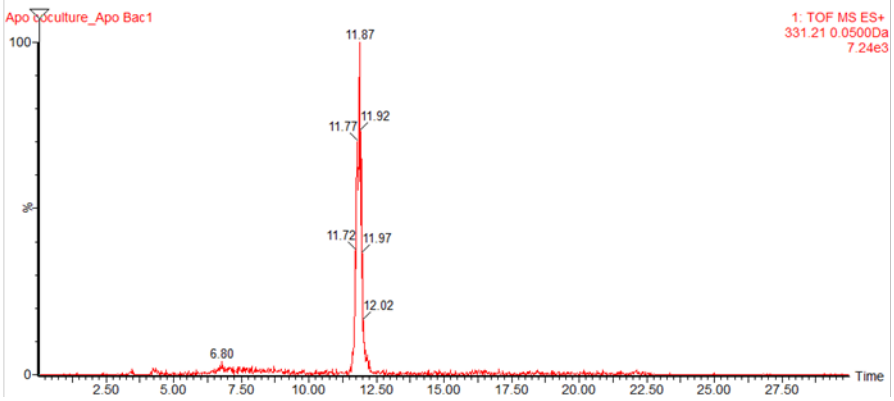
NF



BS

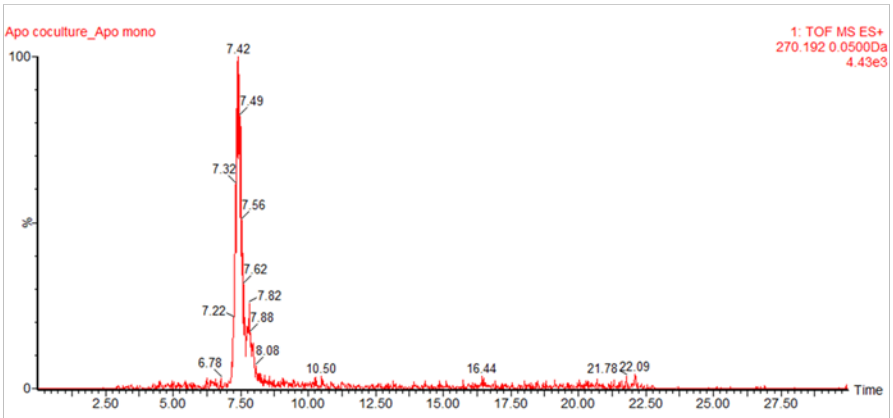


NF&BS

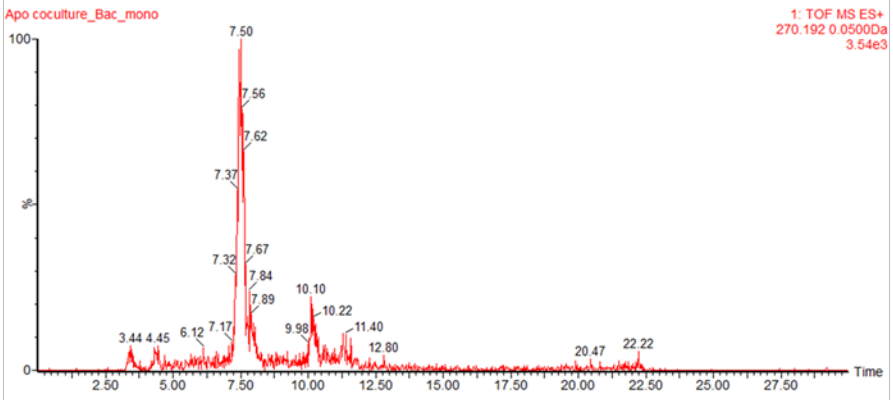


ROI: 4
M270T446

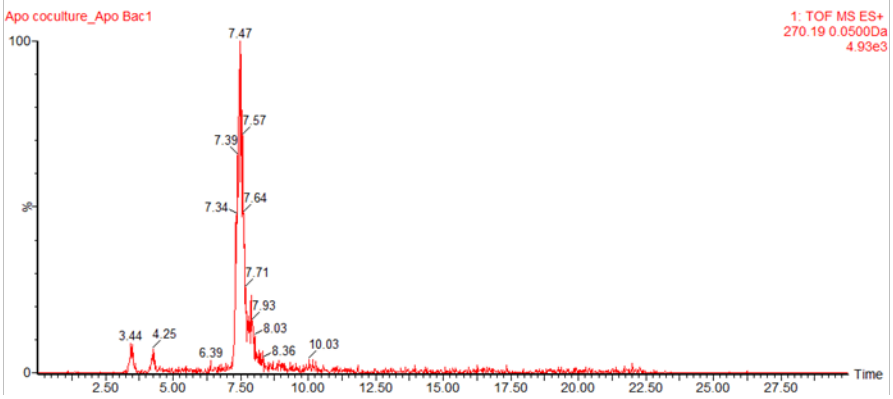
NF



BS

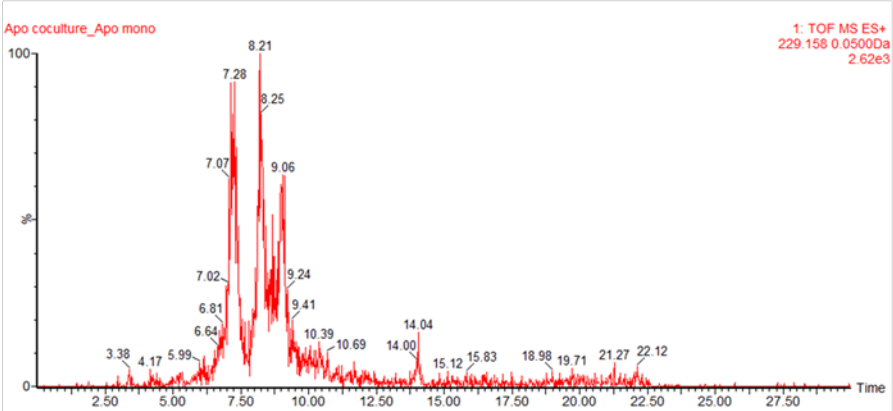


NF&BS

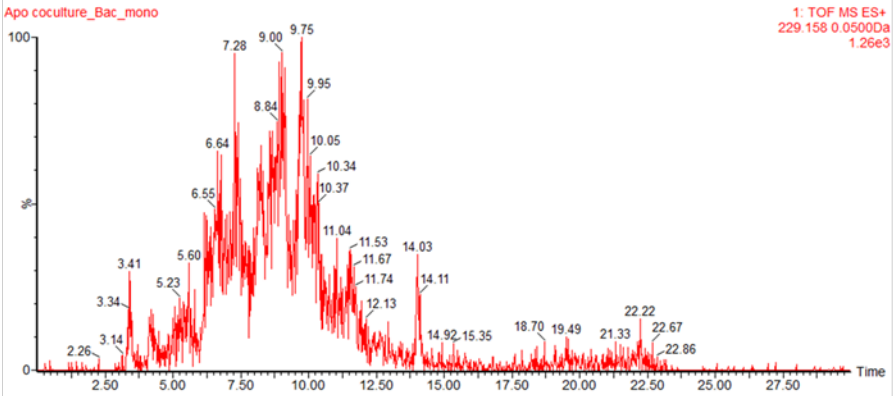


ROI: 4
M29T494

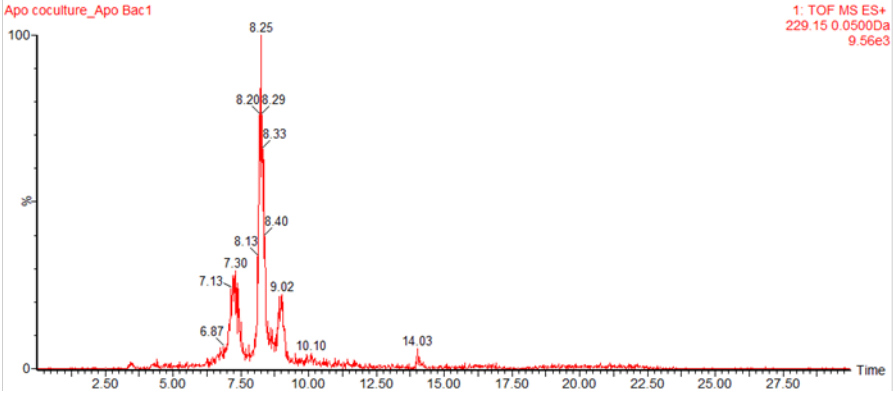
NF



BS

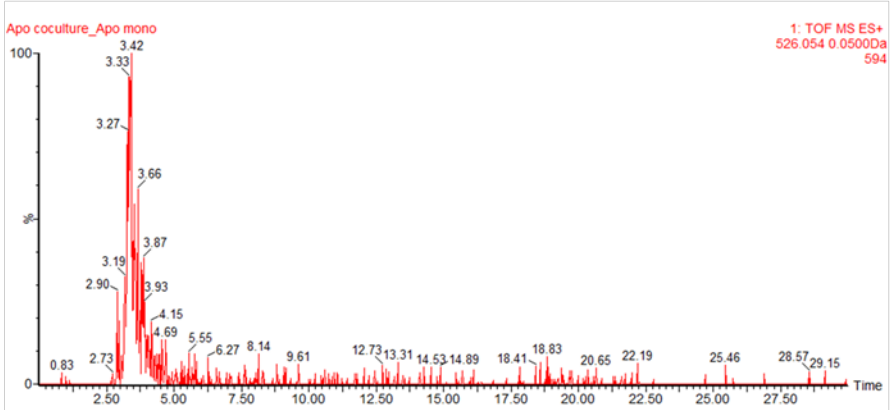


NF&BS

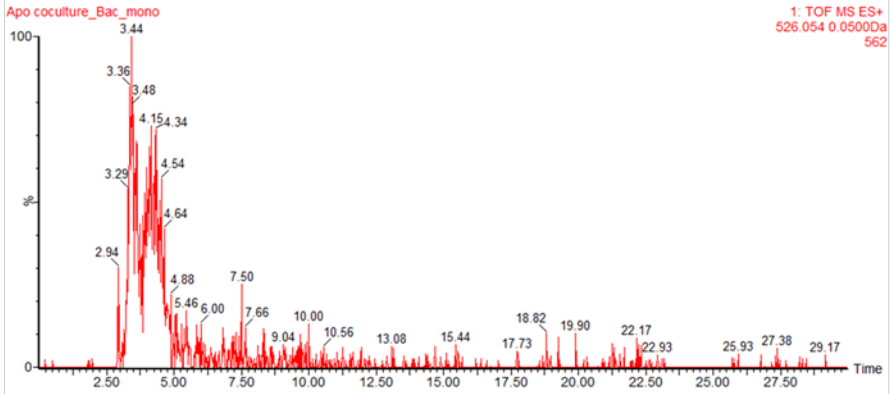


ROI: 4
M526T200

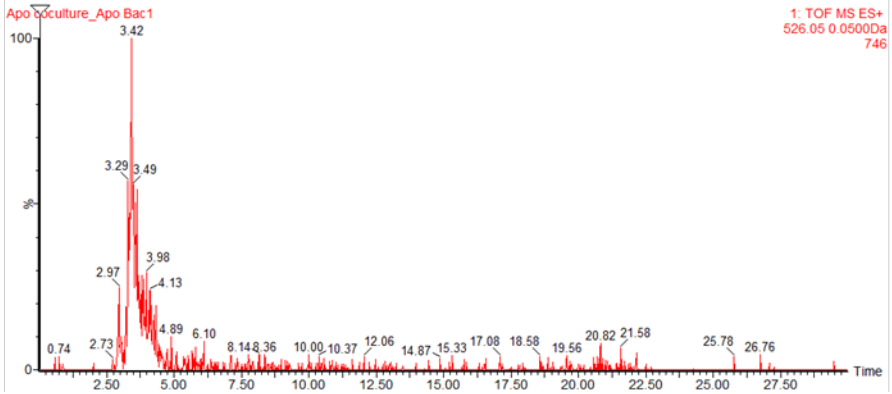
NF



BS

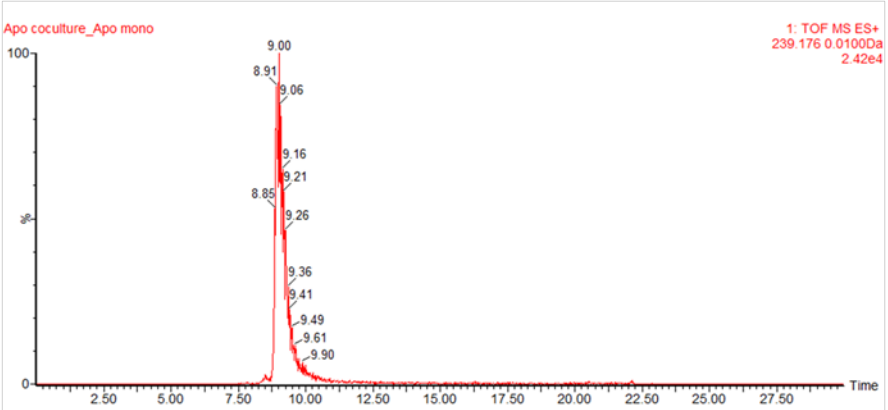


NF&BS

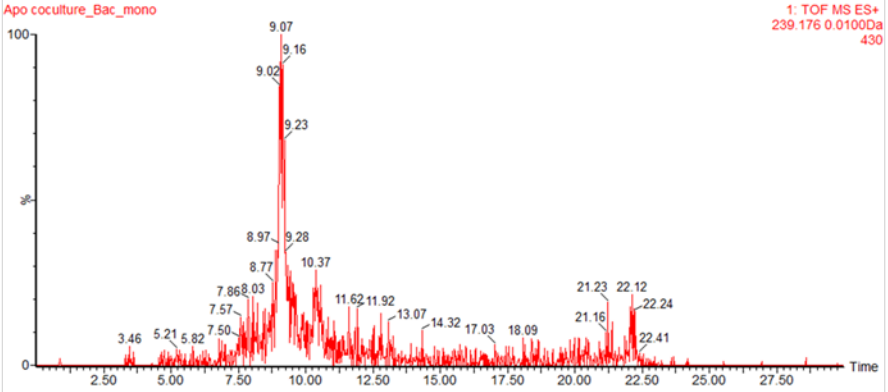


ROI: 5
M239T542

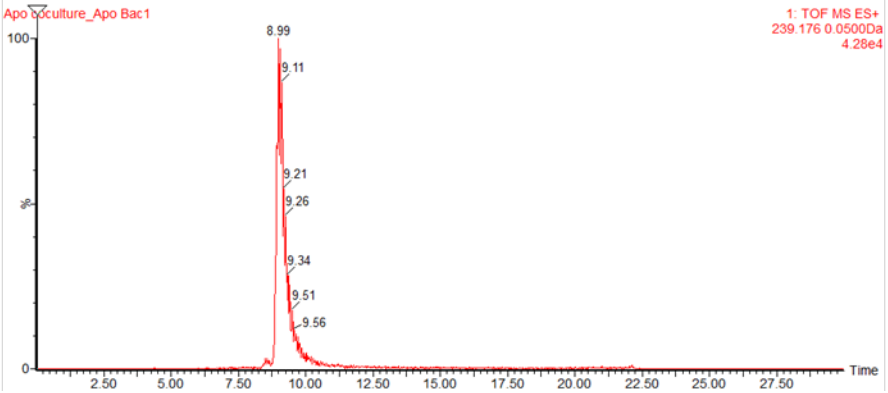
NF



BS

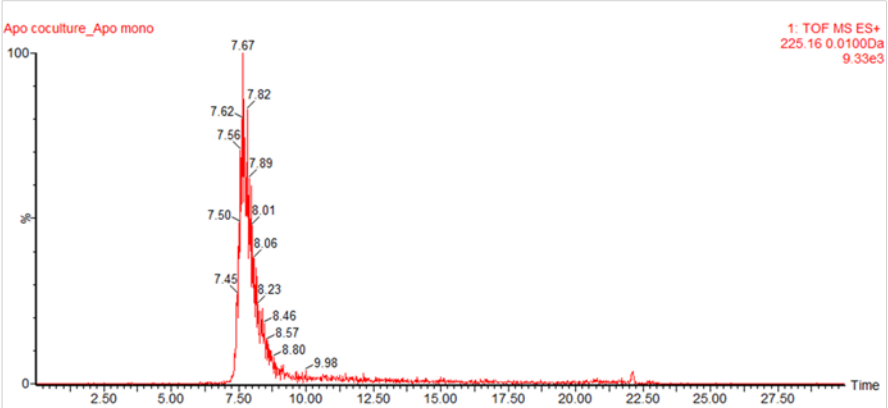


NF&BS

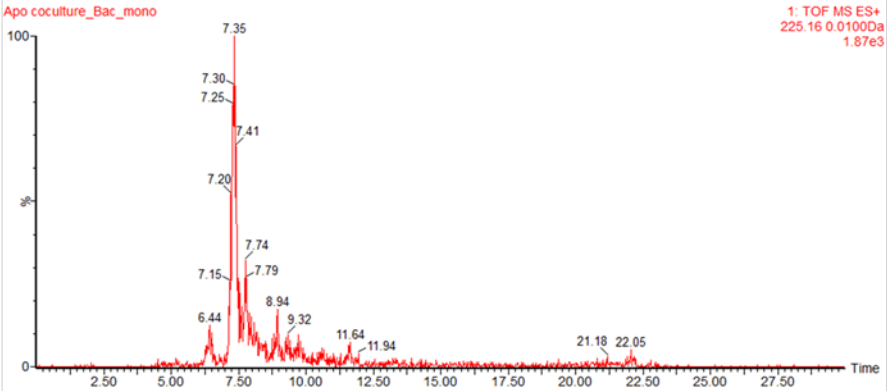


ROI: 5
M225T463

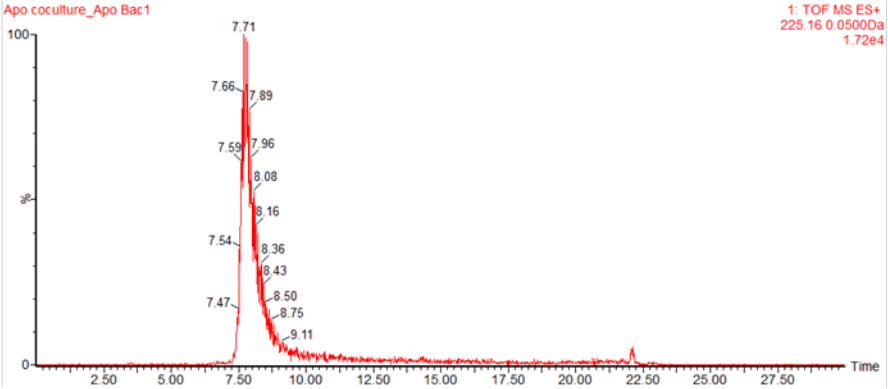
NF



BS

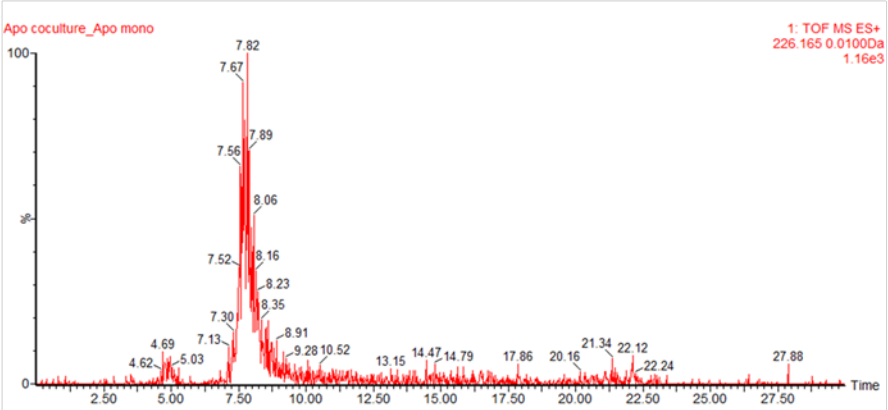


NF&BS

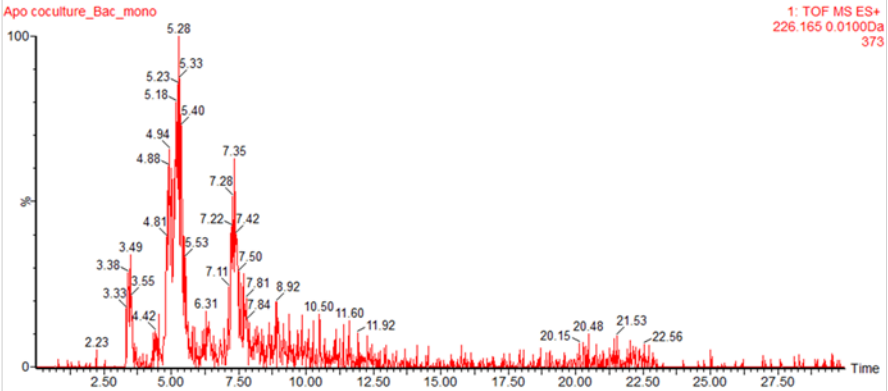


ROI: 5
M226T463

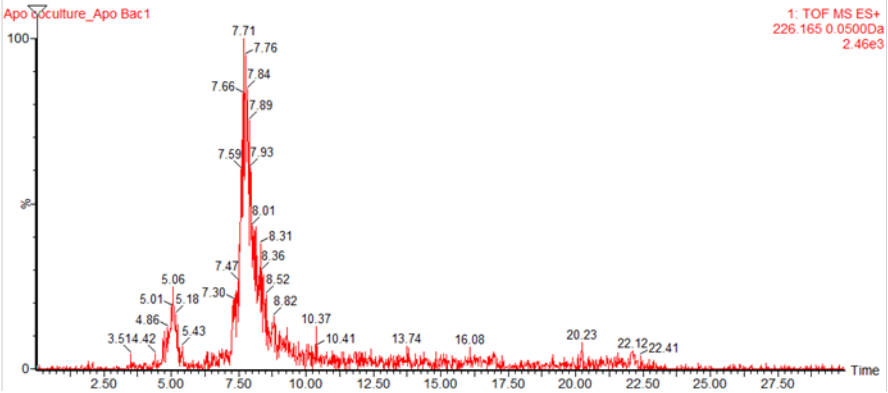
NF



BS

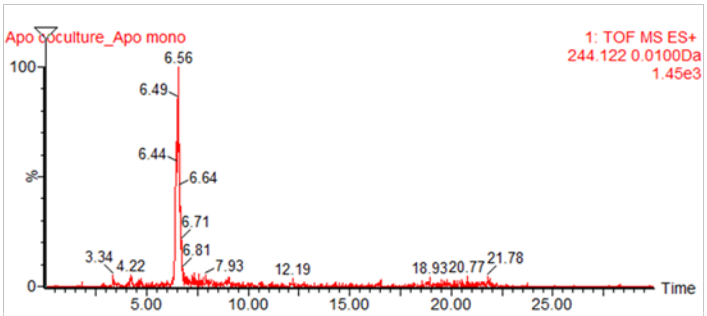


NF&BS

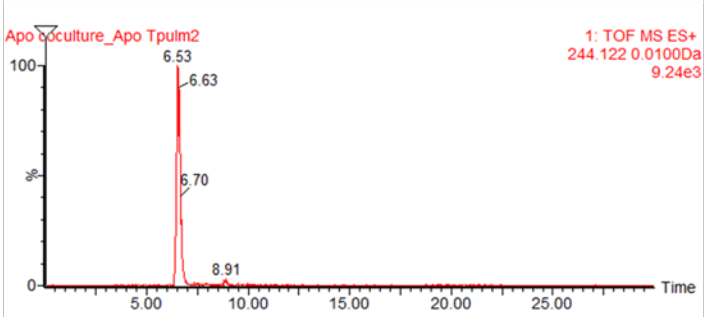


ROI: 6
M244T394

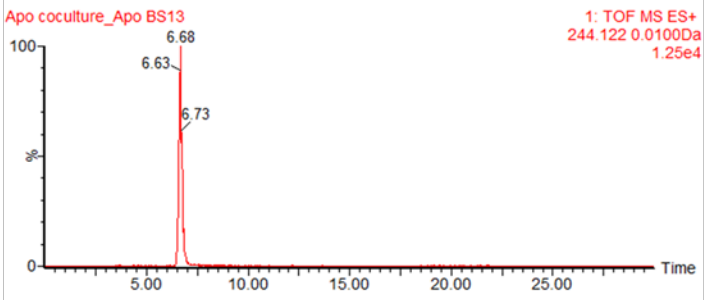
NF



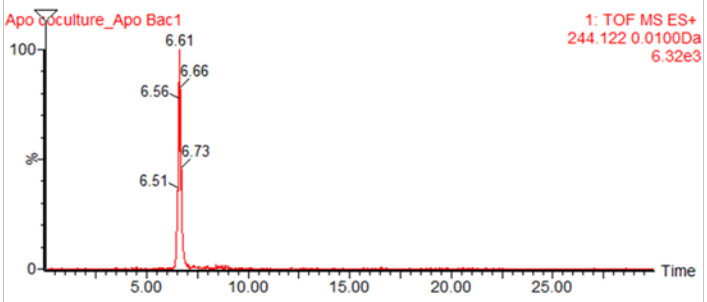
NF&TP



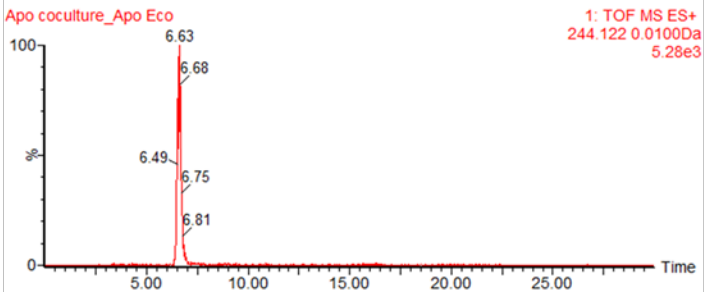
NF&RW



NF&BS

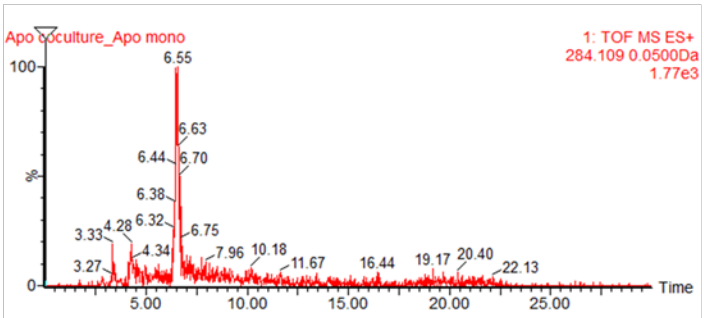


NF&EC

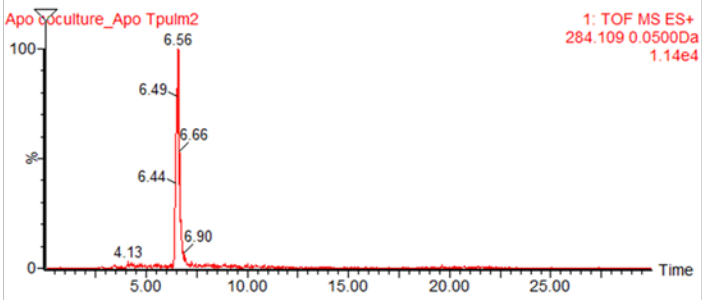


ROI: 6
M284T394

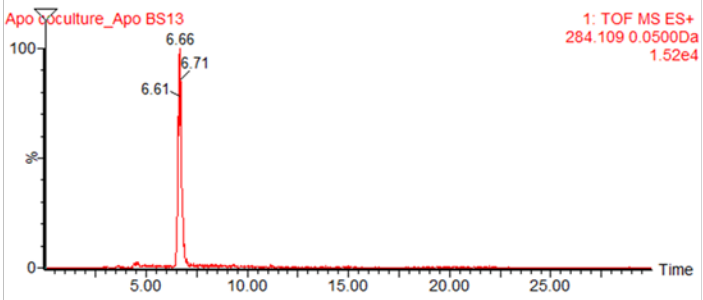
NF



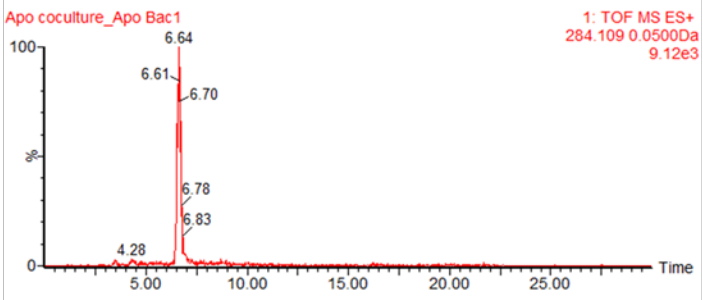
NF&TP



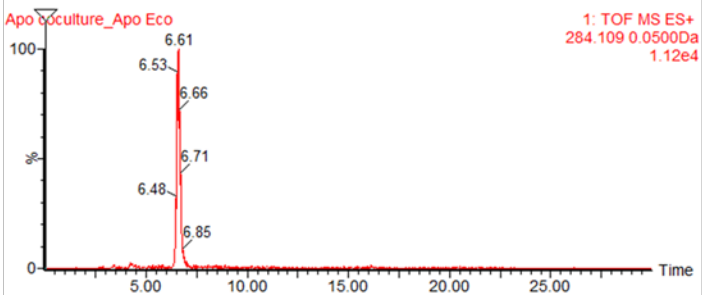
NF&RW



NF&BS

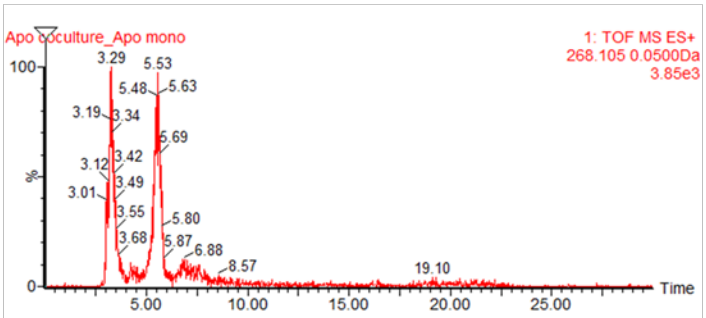


NF&EC

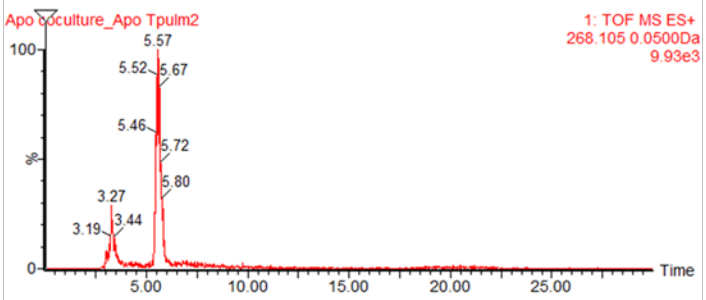


ROI: 6
M268T336

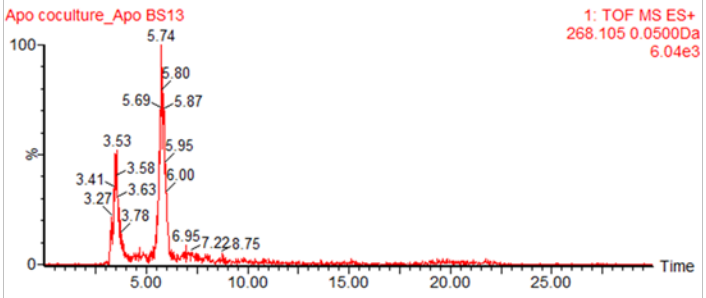
NF



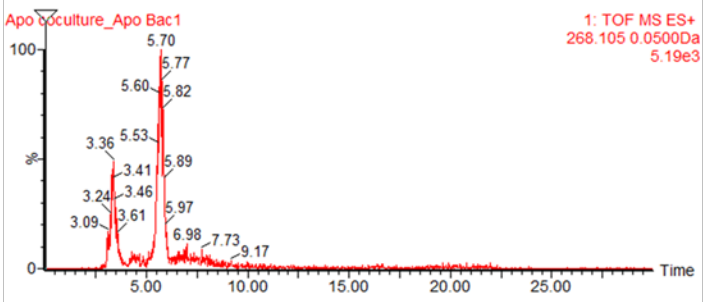
NF&TP



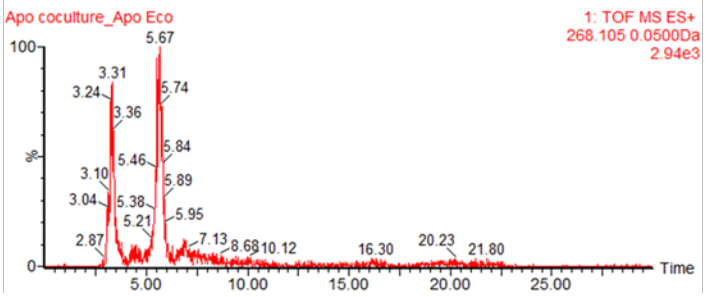
NF&RW



NF&BS

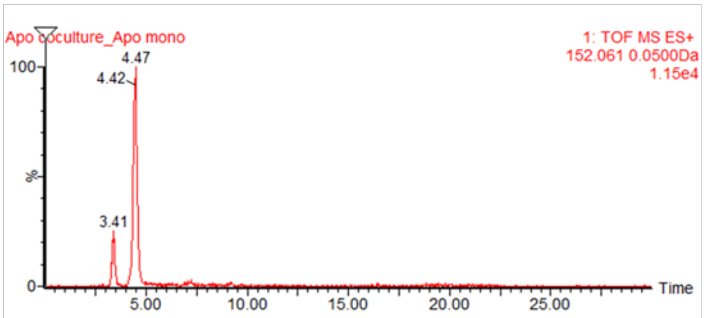


NF&EC

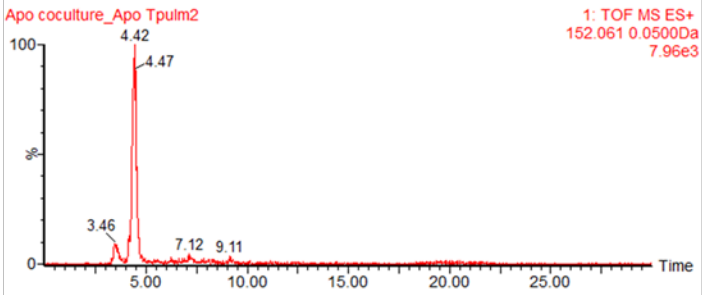


ROI: 6
M152T244

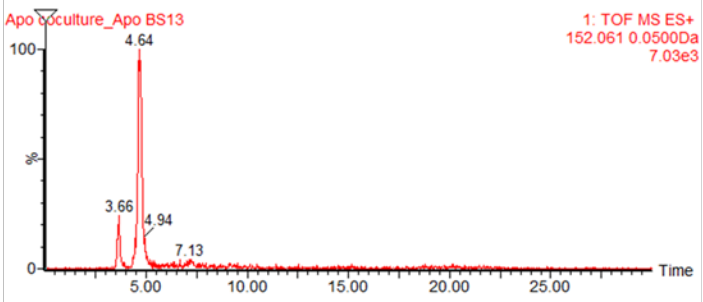
NF



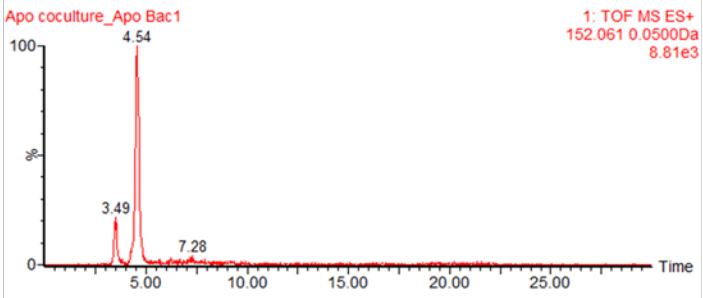
NF&TP



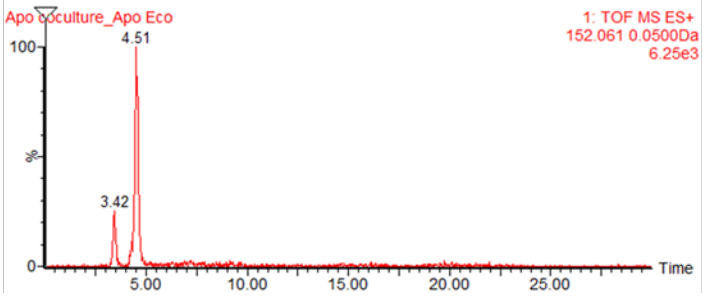
NF&RW



NF&BS

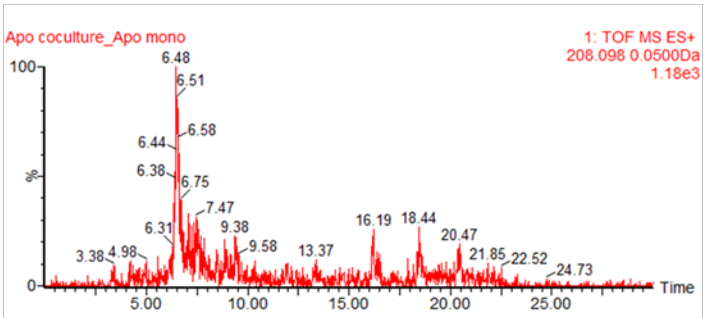


NF&EC

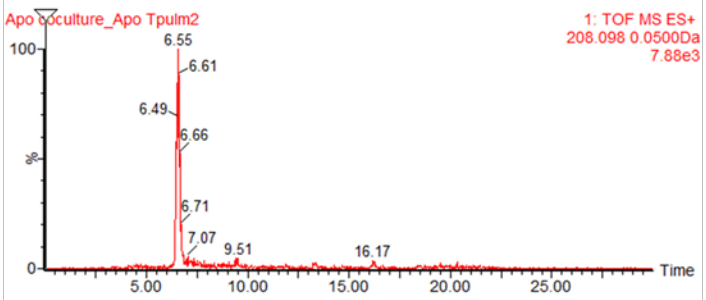


ROI: 6
M208T394

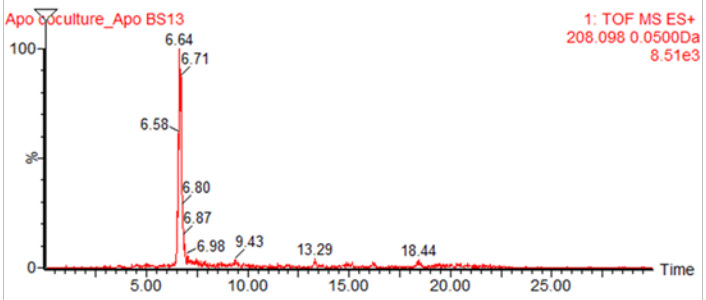
NF



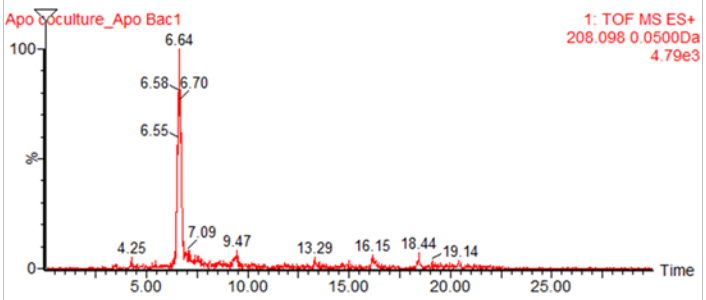
NF&TP



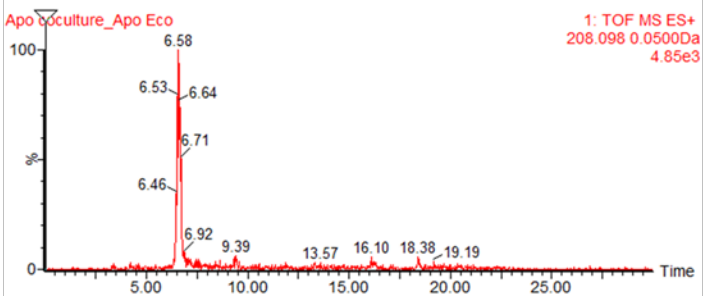
NF&RW



NF&BS



NF&EC



Appendix B Supporting data for Chapter 4

Figure B-1 Metabolomic response to environmental stimuli. Venn diagrams show the distribution of features with at least 10-fold higher abundance in stimuli vs control conditions with a minimum feature value set to 0.01 % of the total ion count. Distribution shown as a percentage of the total number of features detected within that strain. The genus, based on the nearest 16S relative, is shown above, and the isolate ID is given in parenthesis. Sections ≥ 5% are colored red.

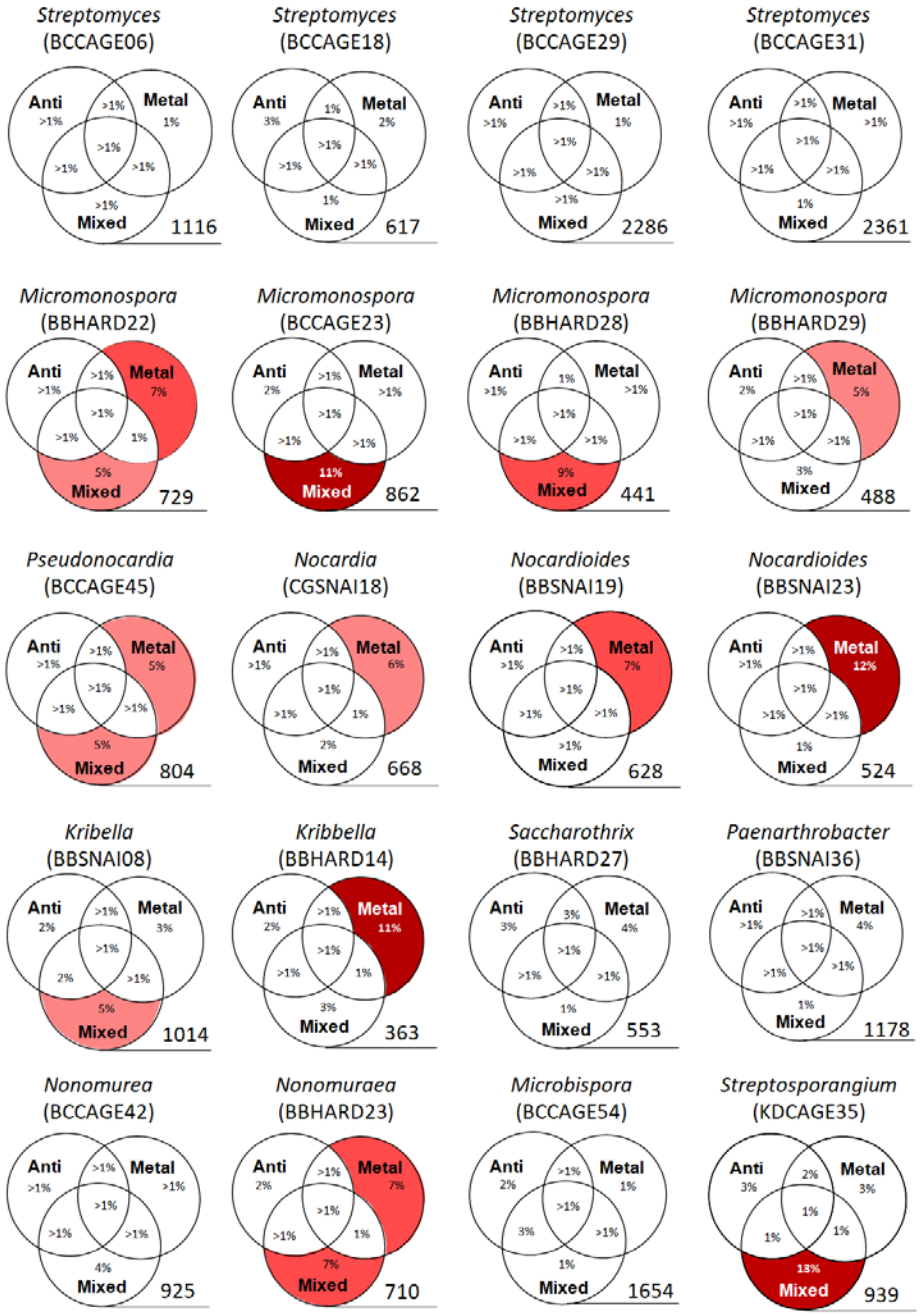


Table B-1: Table of identified actinomycete genera isolated from hypogean samples

Genus List	Number isolated
Streptomyces	91
Micromomospora	9
Pseudomonas	6
Kribella	5
Nocardiodes	5
Bacillus	4
Micrococcus	4
Microbispora	3
Nonomuraea	3
Rhodococcus	3
Arthrobacter	2
Flavobacterium	2
Pseudoduganella	2
Streptosporangium	2
Variovorax	2
Agromyces	1
Azobacter	1
Catellatospora	1
Dactylosporangium	1
Massilia	1
Methylobacterium	1
Nocardia	1
Oerskovia	1
Pseudonocardia	1
Stenotrophomonas	1
Williamsia	1
Saccharothrix	1

Figure B-2**:** Metabolomic response from antibiotic exposure. Stacked bar graph shows the percentage of detected features per strain with increased production of at least 10-fold in rifampicin (red), streptomycin (yellow), and both (gray) stimuli vs control conditions. Below each bar is shown the isolate identifier and the genus based on the nearest 16S rDNA relative.

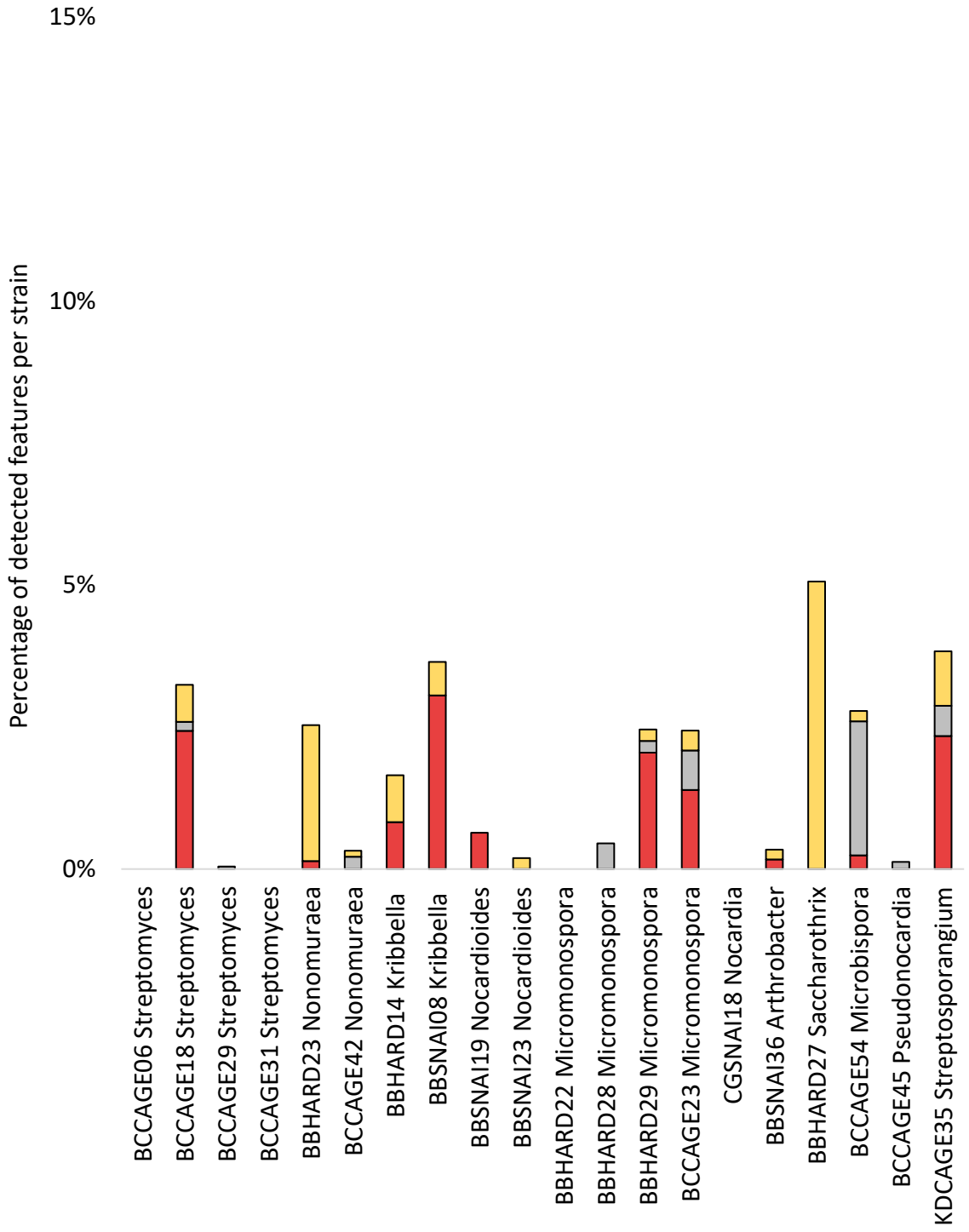


Figure B-3**:** Metabolomic response from rare earth metal exposure. Stacked bar graph shows the percentage of detected features per strain with increased production of at least 10-fold in lanthanum (blue), scandium (black), and both (gray) stimuli vs control conditions. Below each bar is shown the isolate identifier and the genus based on the nearest 16S rDNA relative.

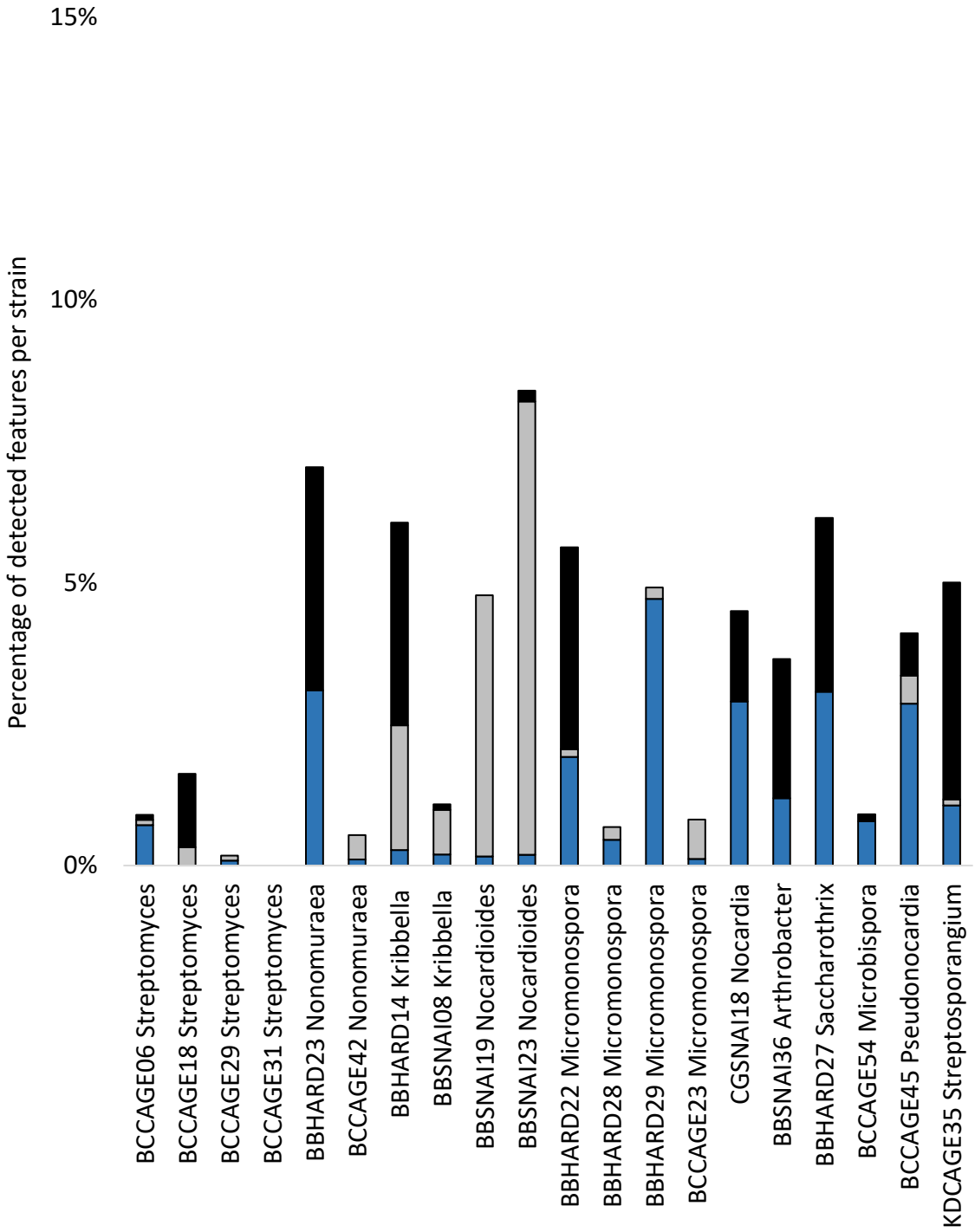


Figure B-4: Metabolomic responses from mixed culture exposures. Stacked bar graph shows the percentage of detected features per strain with increased production of at least 10-fold in mixed culture with TP (orange), RW (green), and both (gray) stimuli vs control conditions. Below each bar is shown the isolate identifier and the genus based on the nearest 16S rDNA relative.

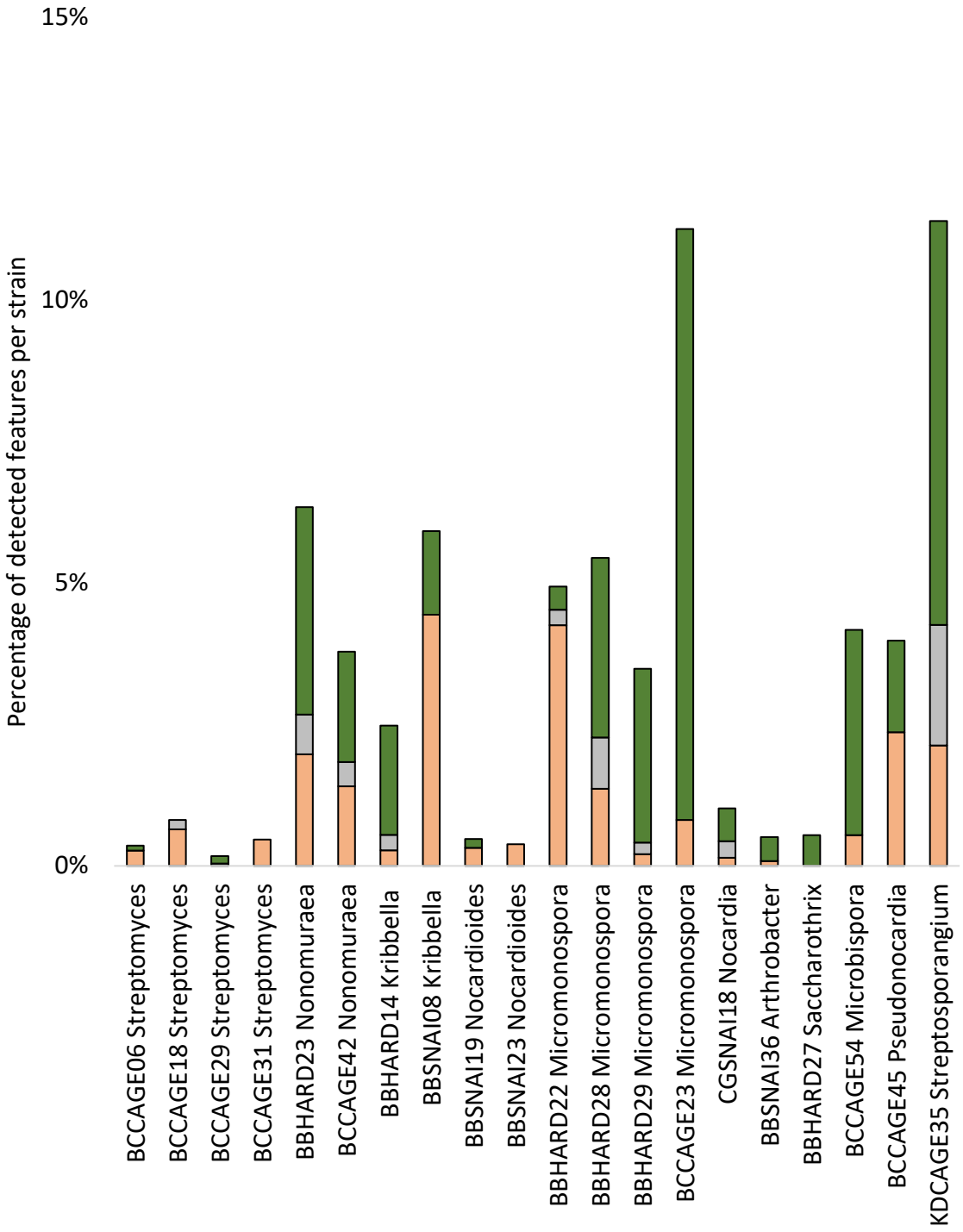


Table B-2: Table of identified NP responses across stimuli conditions of subinhibitory concentrations of rifampicin (Rif) and streptomycin (Str), rare earth metal exposure of lanthanum (La) and scandium (Sc), and mixed culture conditions with TP and RW. "Con" values represent averaged, manually integrated peak intensities for features in the control conditions. Values for stimuli columns represent fold change vs control calculated by dividing the integrated abundance for features of each stimuli condition by the Con value. The heatmap was colored using excels conditional formatting with fold change values of 0.1 set as red, 1 set as white, and 10 as green. Speculative ID provided in parenthesis under description when available.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Micromonospora (BBHARD22)	M257.7T17.2	Okicenone	4E+06	1.3	0.7	0.8	6.0	1.1	2.1
	M253.6T24.6	Aloesaponarin II	2E+07	1.4	0.8	0.3	5.4	0.6	1.2
Nonomuraea (BBHARD23)	M785.1T21.5	Hypogeamicin A	7E+05	0.5	0.2	0.0	2.3	1.1	2.8
	M376T21.1	Hypogeamicin B	6E+05	1.4	0.4	0.2	110.2	10.2	2.0
	M394T15.8	Hypogeamicin C	5E+06	1.4	3.5	0.0	6.2	5.8	5.2
Streptomyces (BCCAGE06)	M1269.6T25.4	Actinomycin C2	2E+07	0.8	1.0	1.5	0.7	3.4	2.0
Microbispora (BCCAGE54)	M1056.4T14	Propeptin 1	7E+05	3.2	1.3	0.1	0.7	1.9	2.9
	M1146.9T14.6	Propeptin 2	2E+06	0.8	0.0	0.6	1.9	0.5	0.9
	M369.3T17	Tetarimycin B	1E+06	16.1	32.8	0.5	6.2	0.4	23.0
Streptosporangium (KDCAGE35)	M1178.3T15.6	Funisamine A	7E+03	4.3	6.4	1.3	1.4	31.2	1.4

Table B-3:: Table of putative NP responses across stimuli conditions of subinhibitory concentrations of rifampicin (Rif) and streptomycin (Str), rare earth metal exposure of lanthanum (La) and scandium (Sc), and mixed culture conditions with TP and RW. Features were assigned as putative NPs based on MS, and MS/MS changes in comparison to known compounds, UV spectrum, and response to stimuli. "Con" values represent averaged, manually integrated peak intensities for features in the control conditions. Values for stimuli columns represent fold change vs control calculated by dividing the integrated abundance for features of each stimuli condition by the Con value. The heatmap was colored using excels conditional formatting with fold change values of 0.1 set as red, 1 set as white, and 10 as green. Speculative ID provided in parenthesis under description when available.

Strain	DESC	Description (*speculative ID*)	Con	Rif	Str	La	Sc	TP	Rw
Micromonospora (BBHARD22)	M256.8T9.9	Unident. NP 1	3E+05	0.8	0.5	0.9	0.4	0.3	0.6
	M233.8T11.1	Unident. NP 2	6E+06	2.4	1.0	0.4	2.4	0.9	2.5
	M283.6T13.1	Unident. Anthraquinone 1	3E+05	12.5	1.0	0.9	19.5	6.3	7.6
	M297.5T14.3	Unident. Anthraquinone 2 (*DMAC*)	4E+07	1.0	1.0	0.4	2.6	1.3	1.3
	M238.7T20.8	Unident. Anthraquinone 3 *(Dihydroxyanthraquinone)*	4E+05	0.6	0.6	3.4	41.7	0.1	0.3
	M241.7T22	Unident. Anthraquinone 4	6E+05	2.9	0.8	0.5	6.8	1.9	4.0
	M269.6T25.4	Unident. Anthraquinone 5 *(Hydroxyaloesaponarin II*)	4E+06	1.5	0.6	0.2	9.0	0.1	1.5
Nonomuraea (BBHARD23)	M605.1T14.7	Unident. Hygogeamicin 1	3E+06	0.8	0.3	0.0	0.5	0.3	1.9
	M410T15.1	Unident. Hygogeamicin 2	1E+06	0.4	0.6	0.0	0.0	0.7	3.9
	M833.2T21	Unident. Hygogeamicin 3	3E+04	0.9	1.4	0.2	1.0	8.1	590.3
	M818.1T21.2	Unident. Hygogeamicin 4	1E+06	1.4	5.1	0.0	1.3	4.7	3.5
Streptomyces (BCCAGE06)	M1271.6T22.6	Unident. Actinomycin 1 (*Y5*)	5E+06	0.8	0.3	1.1	0.6	3.6	2.7
	M1287.5T25.4	Unident. Actinomycin 2 (*Y3*)	5E+05	0.3	0.3	1.5	0.4	2.3	11.9
	M1293.6T25.4	Unident. Actinomycin 3 (*G2*)	5E+05	4.0	1.6	3.7	3.2	4.6	4.4
	M1257.7T25.4	Unident. Actinomycin 4 (*F4*)	9E+05	8.6	1.5	1.0	3.4	9.4	1.8
	M1307.5T25.4	Unident. Actinomycin 5 (*Y2*)	1E+06	0.3	1.3	0.7	0.7	3.3	3.8
	M1255.6T25.4	Unident. Actinomycin 6 (*D*)	7E+06	5.4	0.8	0.8	2.7	5.2	2.5
	M1327.7T25.4	Unident. Actinomycin 7	4E+05	1.7	1.3	2.5	1.1	11.1	5.0
	M1313.6T25.4	Unident. Actinomycin 8	1E+05	29.4	3.9	2.9	7.4	28.7	15.1

Strain	DESC	Description (*speculative ID*)	Con	Rif	Str	La	Sc	TP	Rw
Streptomyces glauciniger (BCCAGE31)	M311.4T8.3	Unident. NP 3	2E+07	0.3	0.7	1.2	0.6	1.1	1.0
	M287.5T9.5	Unident. NP 4	2E+07	0.8	1.1	1.2	0.7	0.3	0.0
	M634.9T12.5	Unident. NP 5	3E+06	0.1	0.5	0.9	0.3	1.0	0.5
	M295.5T12.9	Unident. NP 6	4E+06	1.2	1.0	1.4	0.5	1.3	1.0
	M707.1T16.2	Unident. NP 7	1E+05	0.9	0.7	0.7	0.4	0.0	0.5
	M311.4T16.6	Unident. NP 8	6E+06	0.8	0.8	1.2	0.6	1.5	0.9
	M253.6T26.3	Unident. NP 9 (*Aloesaponarin II*)	3E+06	1.1	0.4	1.3	0.2	0.3	0.3
Nonomuraea (BCCAGE42)	M411.4T12	Unident. NP 10	5E+05	3.5	0.8	0.1	0.6	0.0	0.0
	M628.1T12.7	Unident. Polyene 1	1E+05	14.5	0.1	1.5	0.6	0.0	0.2
	M628.1T13.3	Unident. Polyene 2	2E+04	49.9	0.6	3.8	1.6	0.1	0.0
	548.3T15.4	Unident. Polyene 3	6E+03	140.6	0.0	4.6	2.6	4.0	2.9
	548.3T16.2	Unident. Polyene 4	1E+04	80.9	0.0	1.1	1.0	1.0	1.7
Streptomyces (BCCAGE18)	M900.2T17.1	Unident. NP. 11	1E+06	1.9	0.7	0.7	1.4	0.3	0.3
	M895.2T17.9	Unident. NP. 12	6E+06	2.7	0.7	1.1	2.2	1.0	0.9
Saccharothrix (BBHARD27)	M398.4T13.5	Unident. NP. 13	1E+07	1.3	0.0	0.0	1.9	0.8	1.1
	M247.8T15.2	Unident. NP. 14	1E+07	1.1	0.0	0.1	1.0	0.6	1.0
	M365.6T24.5	Unident. NP. 15	2E+07	1.0	0.0	0.0	1.1	0.7	1.6
Kribella (BBSNAI08)	M732.4T20	Unident. Polyene 5	2E+07	0.2	0.9	0.0	0.0	0.4	0.0
Microbispora (BCCAGE54)	M689.2T10.5	Unident. NP. 16	1E+05	80.5	79.4	0.2	1.1	0.6	48.8
	M561.1T10.7	Unident. NP. 17	1E+06	17.0	21.1	0.5	0.5	0.6	52.9
	M575.2T11.5	Unident. NP. 18	4E+05	120.1	142.2	0.2	0.8	1.0	268.0

Strain	DESC	Description (*speculative ID*)	Con	Rif	Str	La	Sc	TP	Rw
Microbispora (BCCAGE54)	M717.2T11.6	Unident. NP. 19	5E+05	16.4	26.4	0.0	0.6	0.3	15.6
	M731.3T12.1	Unident. NP. 20	6E+05	13.9	27.5	0.1	0.6	0.0	14.9
	M355.4T13	Unident. NP. 21 (*Arromycin*)	8E+04	25.6	52.5	0.0	1.0	2.2	38.8
	M393.2T13.7	Unident. NP. 22 (*Linfuranone A*)	6E+05	4.8	6.0	0.2	0.4	0.0	4.7
	M385.2T14.7	Unident. Tetarimycin 1	3E+04	72.3	98.0	1.3	15.8	3.9	80.7
	M399.2T16.1	Unident. Tetarimycin 2	2E+05	23.0	39.5	0.5	4.4	0.7	20.2
	M406.3T17.5	Unident. Tetarimycin 3	2E+04	8.8	92.5	0.0	3.5	2.4	19.6
	M395.2T19.5	Unident. Tetarimycin 4	3E+04	53.9	49.0	0.8	3.1	1.0	49.8

Figure B-5: Distribution of isolated and putative NPs from Tables S2-3 with ≥10-fold changes in antibiotic exposure (anti), metal exposure (metal), or mixed culture (mixed) stimuli conditions relative to the control.

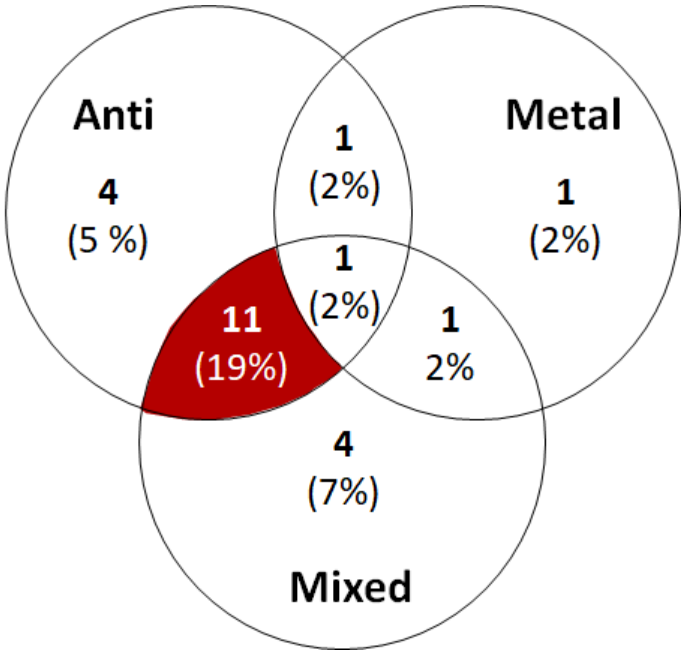


Figure B-6**:** Analysis of volcano plot-prioritized features with chromatogram overlays. (A) Binary volcano plot comparison between *Microbispora* BCCAGE54 control and streptomycin antibiotic exposure conditions. Features below 0.8 significance 5-fold change thresholds colored grey. Features over thresholds colored by retention time brackets: <10 min (purple), between 10-15 min (yellow), and >15 min (green). (B) Volcano plot of features above thresholds and eluting before 10 min. (C) Volcano plot of features above thresholds and eluting between 10 and 15 min. (D) Volcano plot of features above thresholds and eluting after 15 min. Overlays of extracted ion, total ion, and UV/Vis chromatograms for features eluting after 15 min are shown in (E) for streptomycin exposed and (F) control *Microbispora* BCCAGE54 extracts. Red arrows highlight tetarimycin B peak in the EICs and PDA.

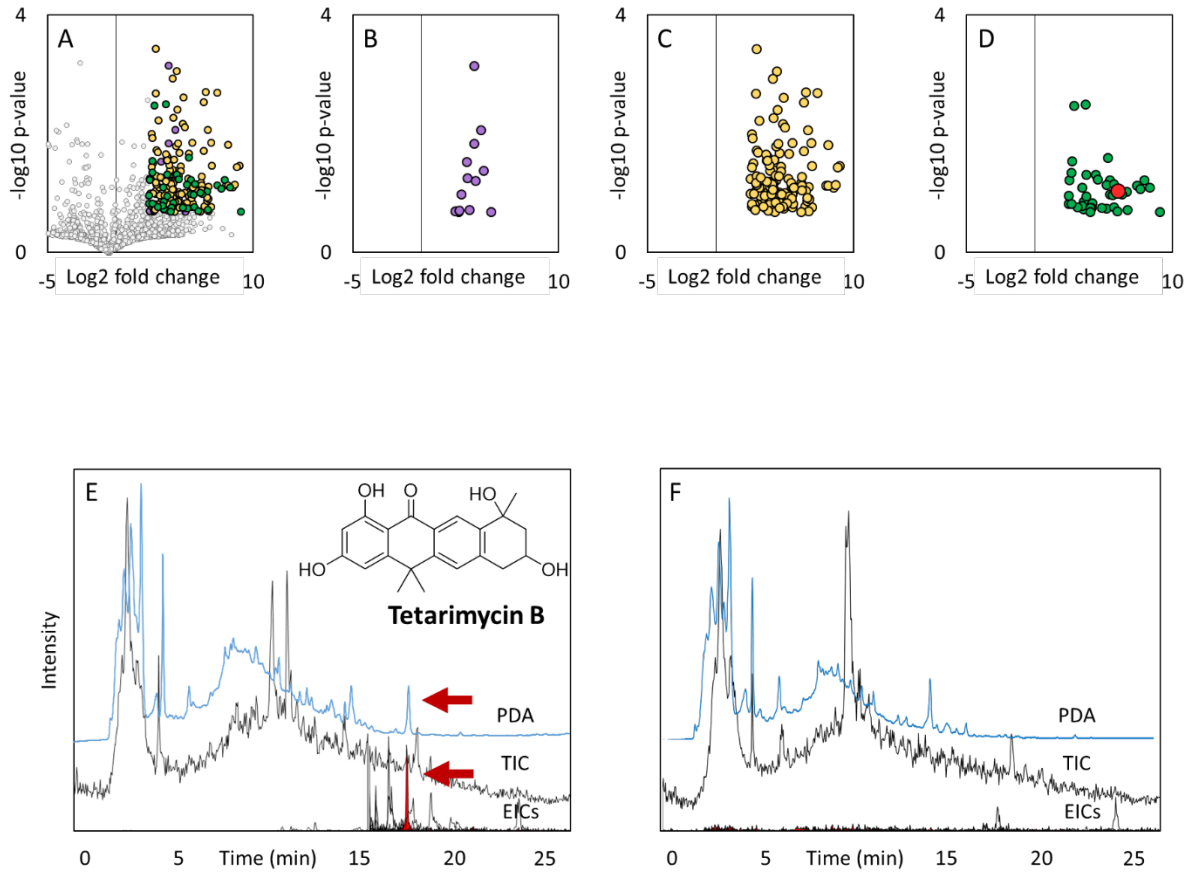


Figure B-7**:** EICs of (A) okicenone and (C) aloesaponarin II from control (black) and Sc treated (red) *Micromonospora* BBHARD22 extracts and mass spectral data for (B) okicenone and (D) aloesaponarin II. Structures shown on right.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Micromonospora (BBHARD22)	M257.7T17.2	okicenone	4E+06	1.3	0.7	0.8	6.0	1.1	2.1
	M253.6T24.6	aloesaponarin II	2E+07	1.4	0.8	0.3	5.4	0.6	1.2

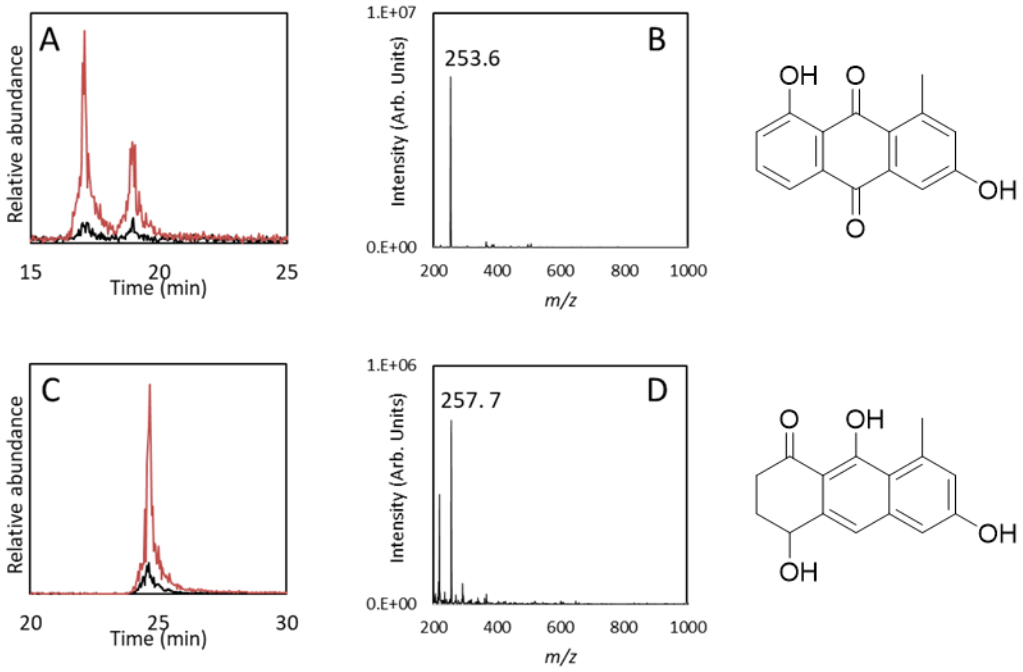


Figure B-8: EICs of (A) hypogeamicin A, (C) hypogeamicin B, and (E) hypogeamicin C from control (black) and Sc treated (red) *Nonomuraea* BBHARD23 extracts and mass spectral data for (B) hypogeamicin A, (D) hypogeamicin B, and (F) hypogeamicin C. Structures shown on right.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Nonomuraea (BBHARD23)	M785.1T21.5	hypogeamicin A	7E+05	0.5	0.2	0.0	2.3	1.1	2.8
	M376T21.1	hypogeamicin B	6E+05	1.4	0.4	0.2	110.2	10.2	2.0
	M394T15.8	hypogeamicin C	5E+06	1.4	3.5	0.0	6.2	5.8	5.2

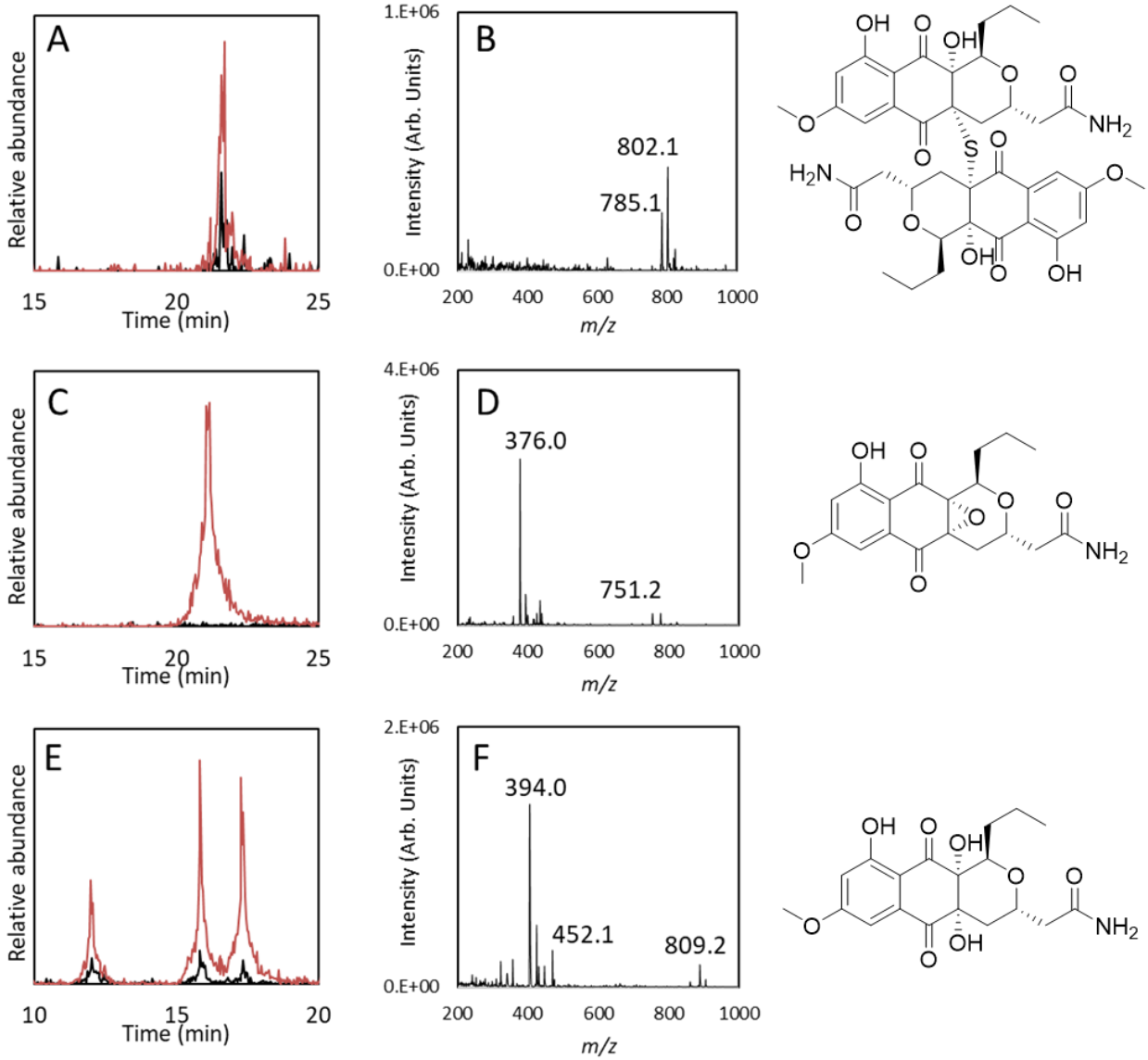
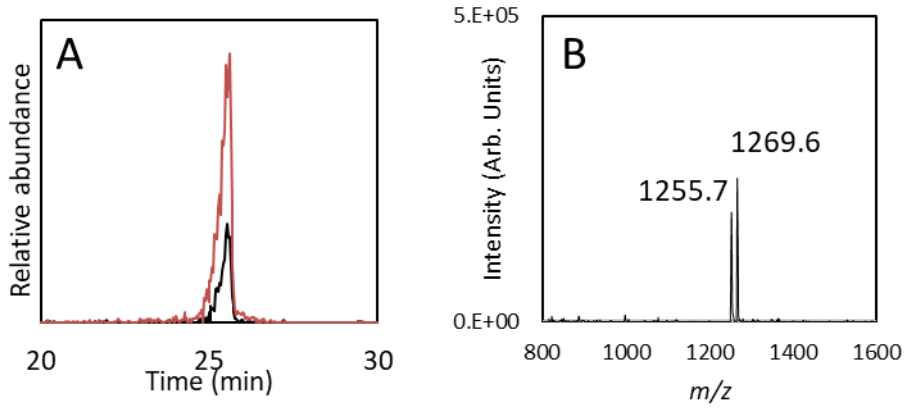


Figure B-9**:** EICs of (A) actinomycin C2 from control (black) and TP treated (red) *Streptomyces* BCCAGE06 extracts and mass spectral data for (B) actinomycin C2. The structure of actinomycin C2 is shown below.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Streptomyces (BCCAGE06)	M1269.6T25.4	actinomycin C2	2E+07	0.8	1.0	1.5	0.7	3.4	2.0



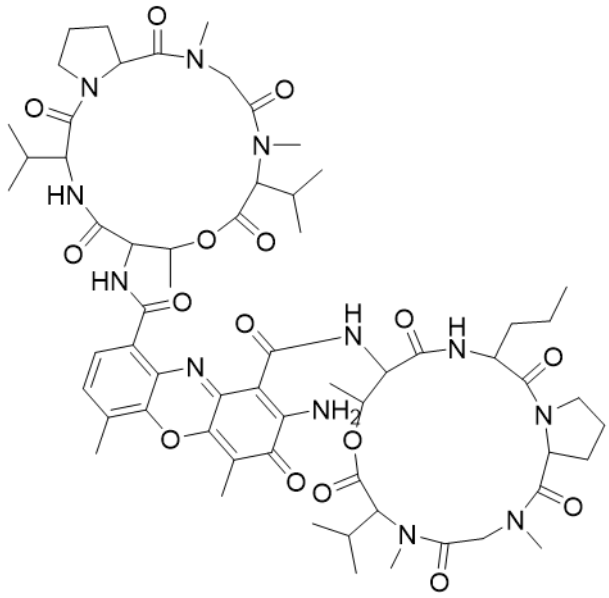


Figure B-10: EICs of (A) propeptin 1, (C) propeptin 2, and (D) tetarimycin B from control (black) and Rif (a), La (b), and Str (c) treated (red) *Microbispora* BCCAGE54 extracts and mass spectral data for (B) propeptin 1, (D) propeptin 2, and (F) tetarimycin B. Structures shown on right.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Microbispora (BCCAGE54)	M1056.4T14	propeptin 1	7E+05	3.2	1.3	0.1	0.7	1.9	2.9
	M1146.9T14.6	propeptin 2	5E+05	0.6	0.0	2.0	1.5	0.5	0.7
	M369.3T17	tetarimycin B	1E+06	16.1	32.8	0.5	6.2	0.4	23.0

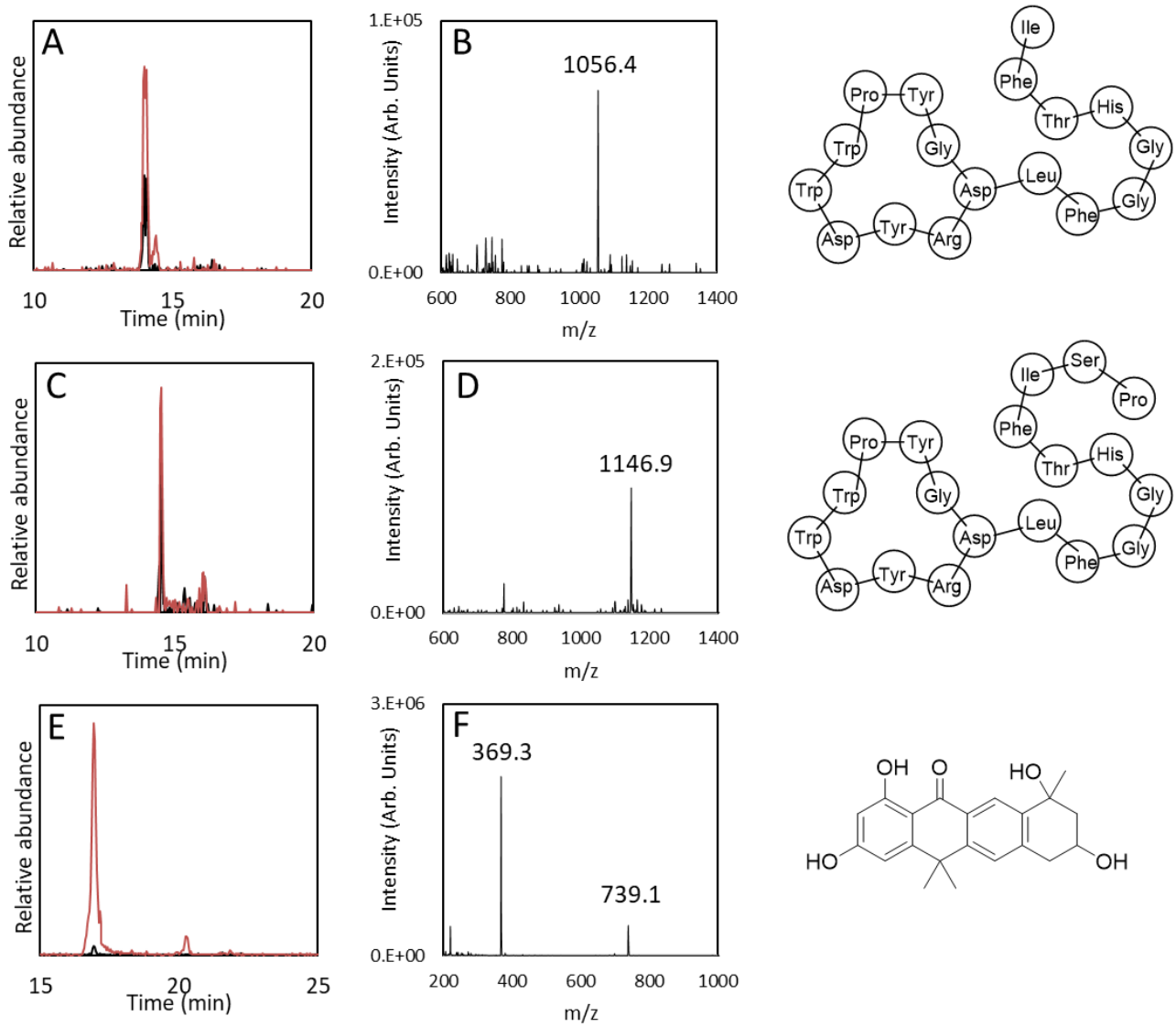
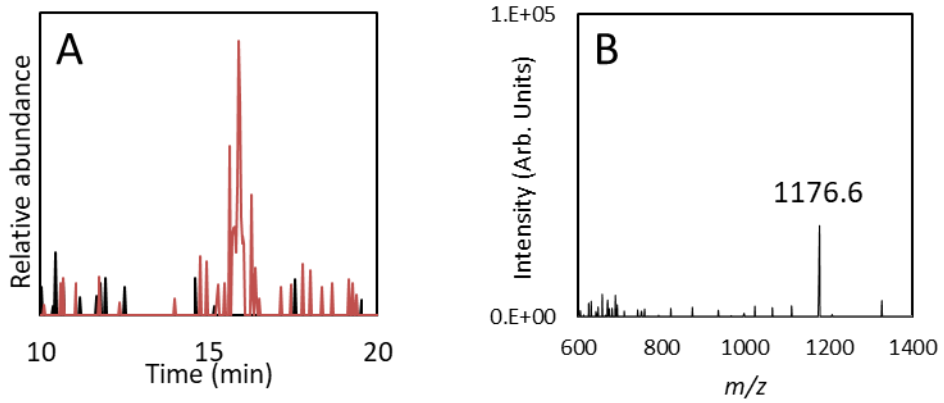


Figure B-11**:** EICs of (A) funisamine from control (black) and TP treated (red) *Microbispora* BCCAGE54 extracts and mass spectral data for (B) funisamine. The structure of funisamine is shown below.

Strain (Isolate)	DESC	ID	Con	Rif	Str	La	Sc	TP	Rw
Streptosporangium caverna (KDCAGE35)	M1178.3T15.6	funisamine A	7E+03	4.3	6.4	1.3	1.4	31.2	1.4



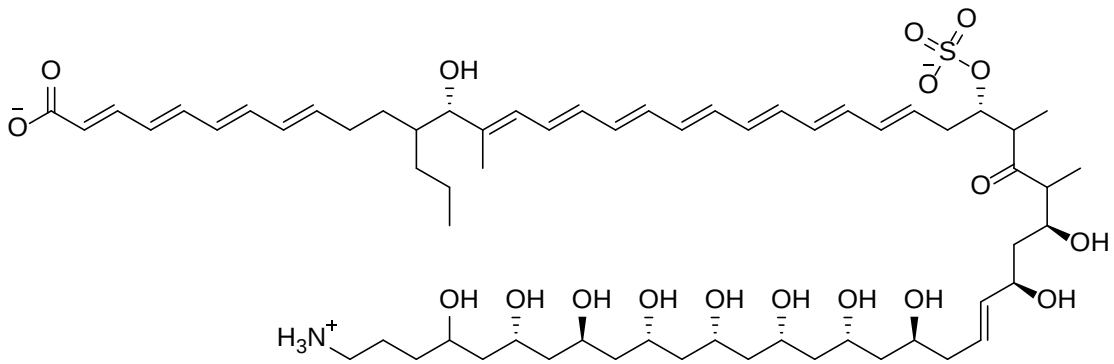


Figure B-12**:** Mass spectral data for *Micromonospora* BBHARD22 features (A) unident. NP 1 (M256.8T9.9), (B) unident. NP 2 (M233.8T11.1), (C) unident. anthraquinone 1 (M283.6T13.1), (D) unident. anthraquinone 2 (M297.5T14.3, *DMAC*), (E) okicenone (M257.7T17.2), (F) unident. anthraquinone 3 (M238.7T20.8, *dihydroxyanthraquinone*), (G) unident. anthraquinone 4 (M241.7T22), (H) aloesaponarin II (M253.6T24.6), and (I) unident. anthraquinone 5 (M269.6T25.4, *Hydroxyaloesaponarin II*) in *Micromonospora* BBHARD22

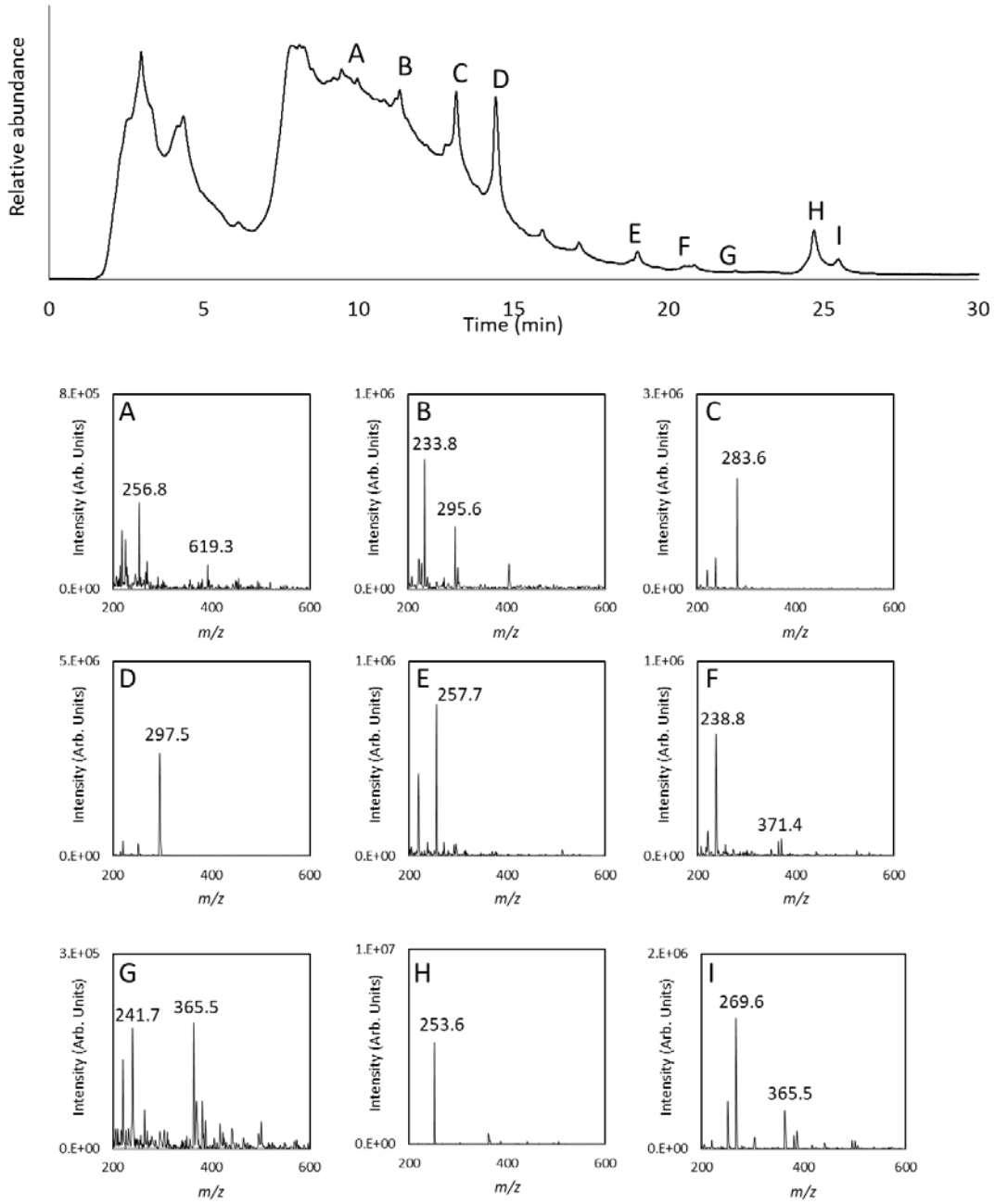


Figure B-13**:** UV spectrum of *Micromonospora* BBHARD22 features (A) unident. NP 1 (M256.8T9.9), (B) unident. NP 2 (M233.8T11.1), (C) unident. anthraquinone 1 (M283.6T13.1), (D) unident. anthraquinone 2 (M297.5T14.3, *DMAC*), (E) okicenone (M257.7T17.2), (F) unident. anthraquinone 3 (M238.7T20.8, *dihydroxyanthraquinone*), (G) unident. anthraquinone 4 (M241.7T22), (H) aloesaponarin II (M253.6T24.6), and (I) unident. anthraquinone 5 (M269.6T25.4, *Hydroxyaloesaponarin II*) in *Micromonospora* BBHARD22

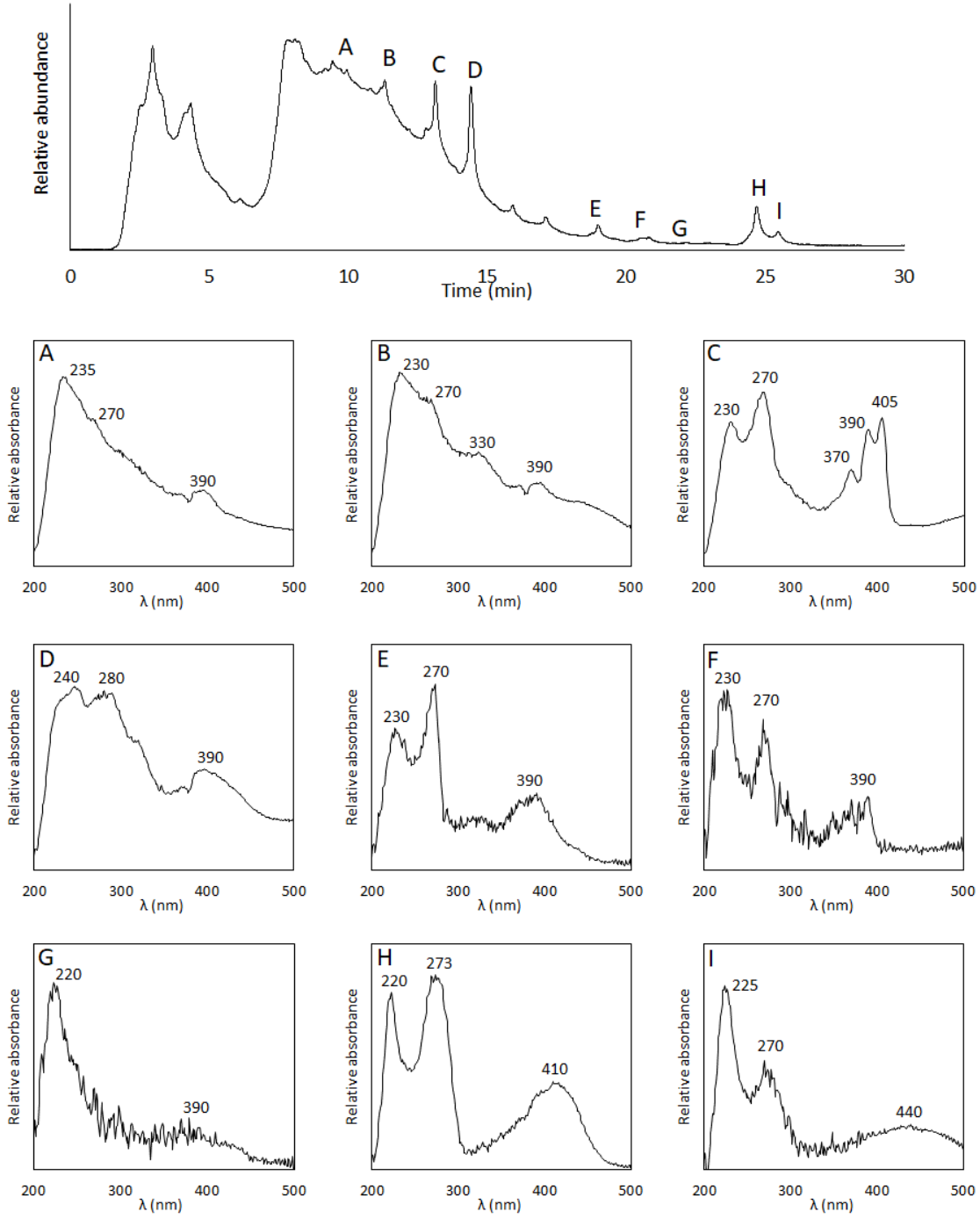


Figure B-14**:** UV spectrum of *Nonomuraea* BBHARD23 features (A) unident. hypogeamicin 4 (M605.1T14.7), (B) unident. hypogeamicin 3 (M410T15.1) in *Nonomuraea* (BBHARD23) unident. hypogeamicin 3 (M809.1T15.8), (C) hypogeamicin C (M394T15.8), (D) hypogeamicin B (M376T21.1), (E) unident. hypogeamicin 1 (M833.2T21), (F) unident. hypogeamicin 2 (M818.1T21.2), and (G) hypogeamicin A (M785.1T21.5) in *Nonomuraea* BBHARD23 extracts.

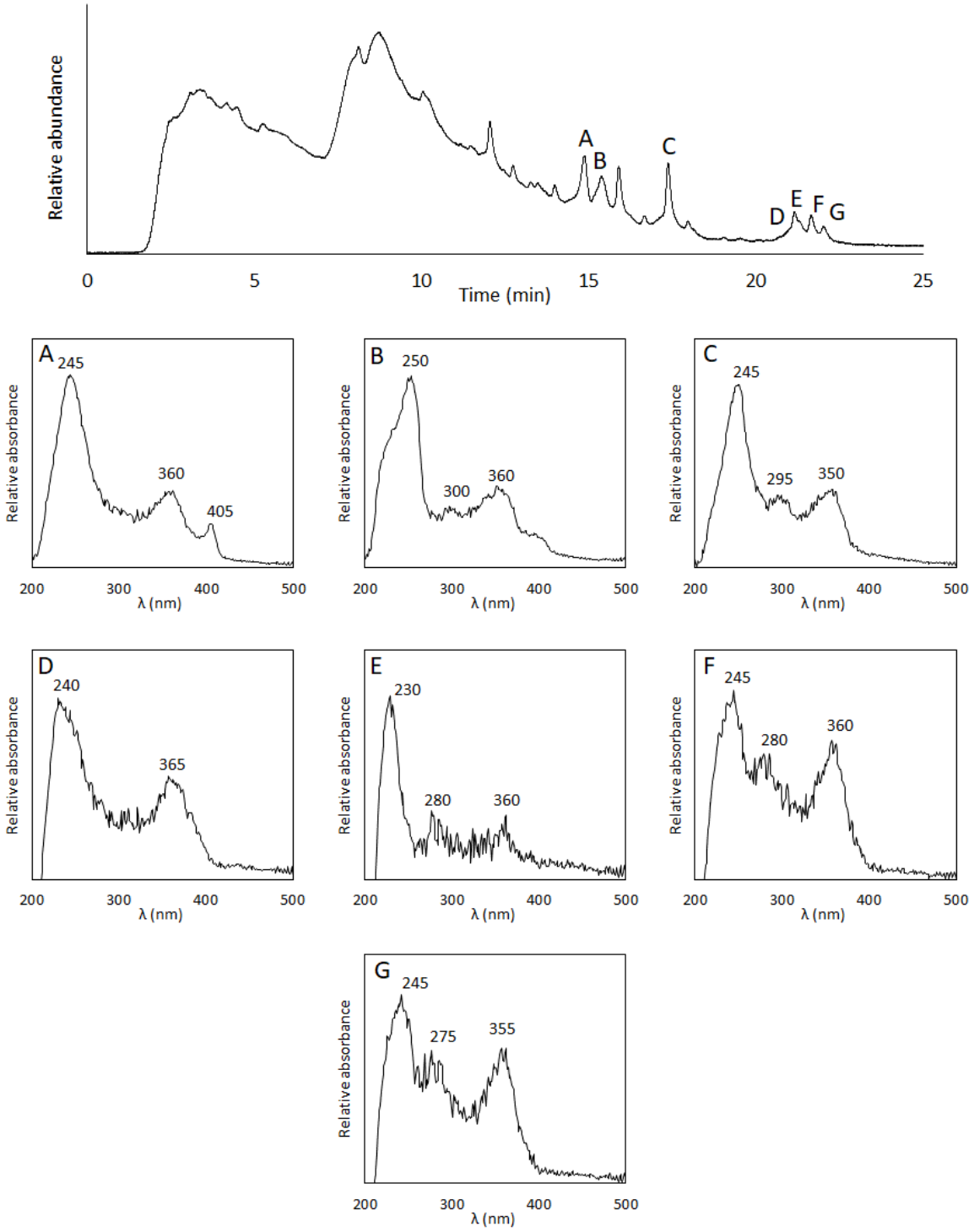


Figure B-15**:** UV spectrum of features from *Streptomyces* BCCAGE06 (A) unident. actinomycin 1 (*Y5*) (M1271.6T22.6), and (B) actinomycin C2 (M1270.6T25.4), unident. actinomycin 2 (M1287.5T25.4, *actinomycin Y3*), unident. actinomycin 3 (M1293.6T25.4, *actinomycin G2*), unident. actinomycin 4 (M1257.7T25.4, *actinomycin F4*), unident. actinomycin 5 (M1307.5T25.4, *actinomycin Y2*), unident. actinomycin 6 (M1255.6T25.4, *actinomycin D*), unident. actinomycin 7 (M1327.7T25.4), and unident. actinomycin 8 (M1313.7T25.4) in *Streptomyces* BCCAGE06

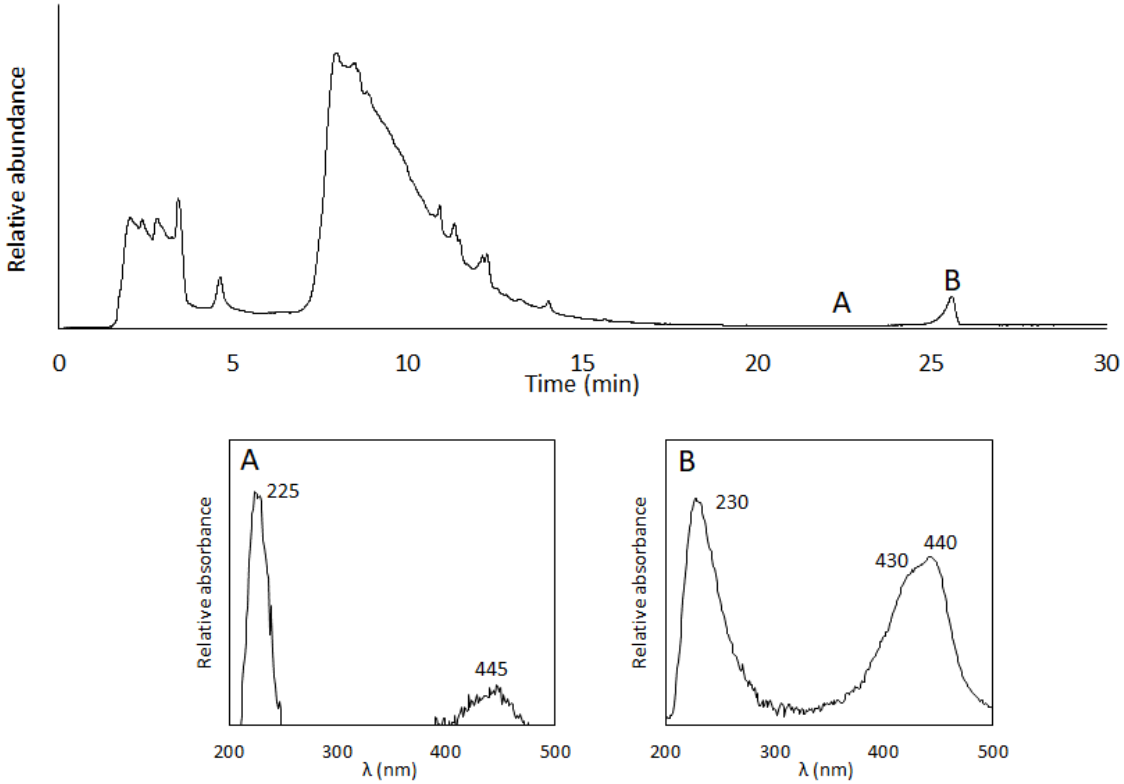


Figure B-16**:** UV spectrum of features from *Streptomyces* BCCAGE31 (A) unident. NP 3 (M311.4T8.3), (B) unident. NP 4 (M287.5T9.5), (C) unident. NP 5 (M634.9T12.5), (D) unident. NP 6 (M295.5T12.9), (E) unident. NP 7-8 (M707.1T16.2, M311.4T16.6), and (F) and unident. NP 9 (M253.6T26.3, aloesaponarin II) in *Streptomyces* BCCAGE31.

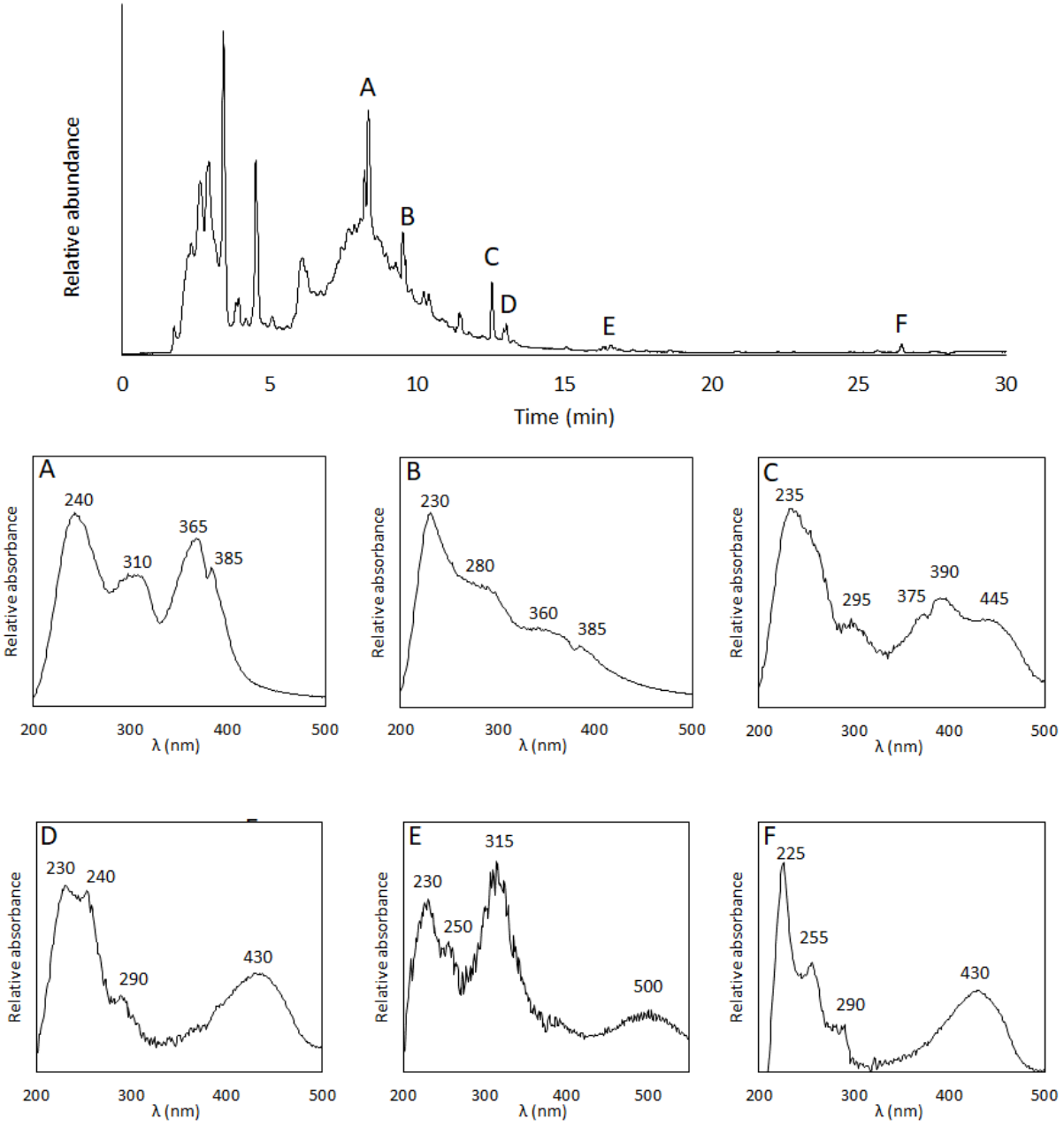


Figure B-17**:** UV spectrum of features from *Nonomuraea* BCCAGE42 (A) unident. NP 10 (M411.4T12), (B) unident. Polyene 1 (M628.1T12.7), (C) unident. Polyene 2 (M628.1T13.3), (D) unident. Polyene 3 (548.3T15.4), and (E) unident. Polyene 4 (548.3T16.2) in *Nonomuraea* BCCAGE42 extracts.

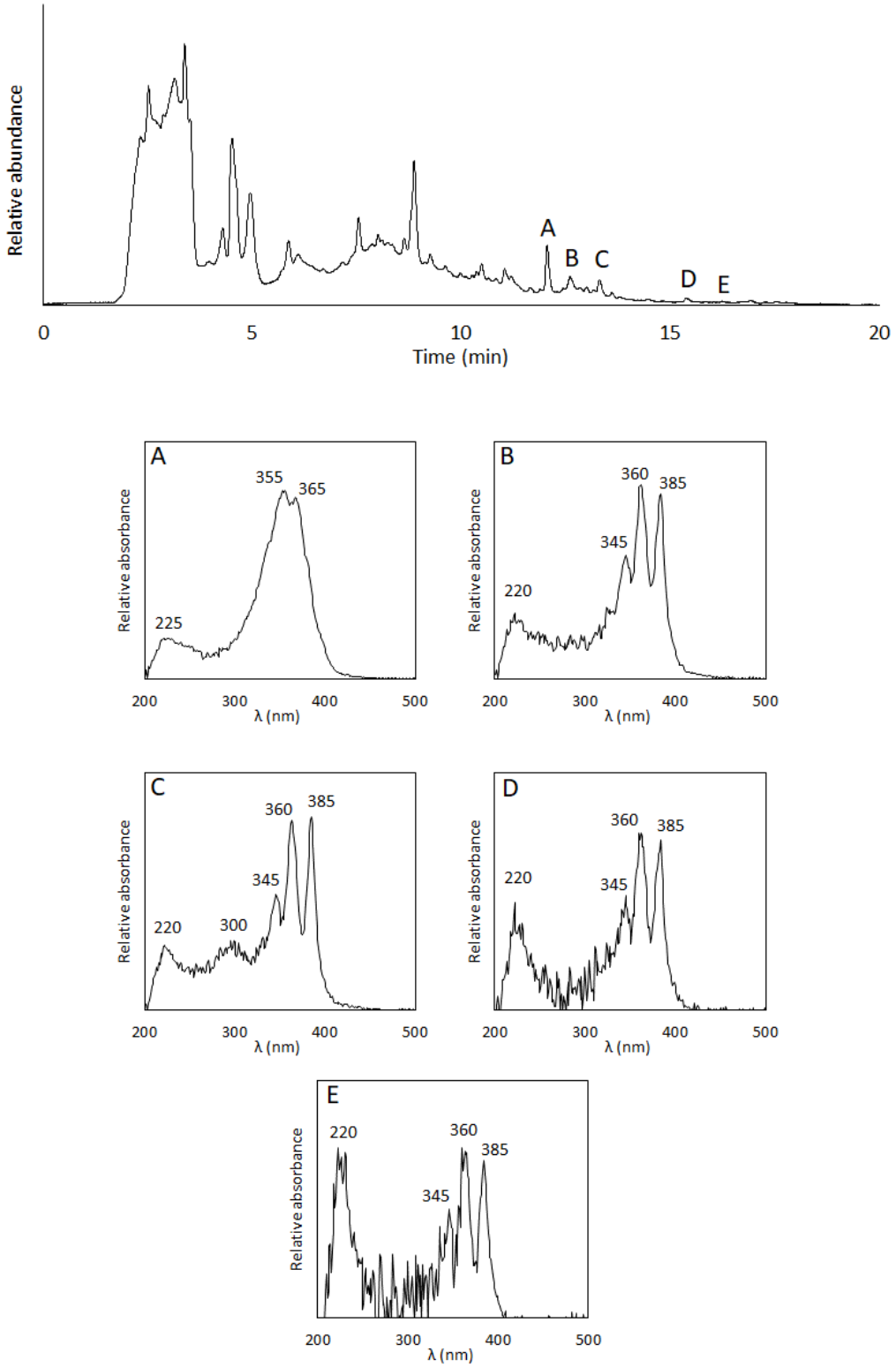


Figure B-18**:** UV spectrum of features from *Streptomyces* BCCAGE18 (A) unident. NP. 11-12 (M900.2T17.1, M895.2T17.9) in *Streptomyces* BCCAGE18 extracts.

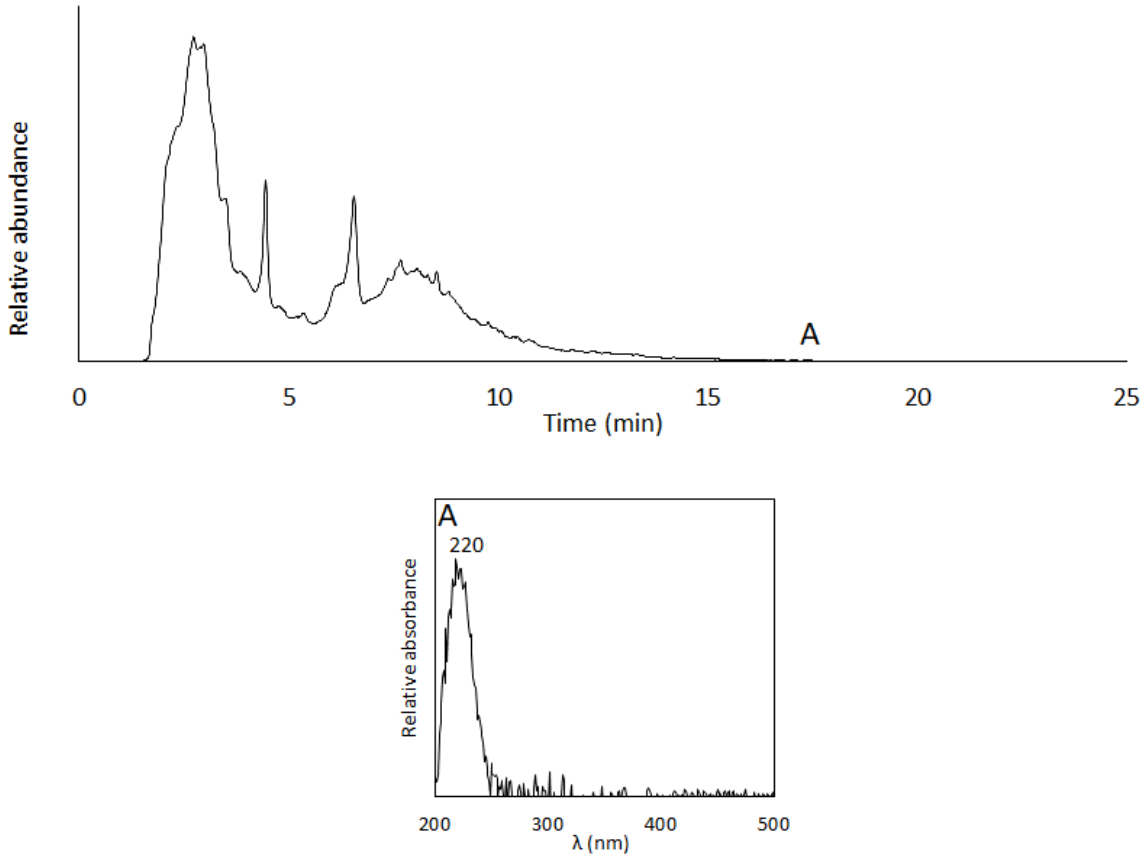


Figure B-19**:** UV spectrum of features from *Saccharothrix* BBHARD27 (A) unident. NP 13 (M398.4T13.5), (B) unident. NP. 14 (M247.8T15.2), and (C) unident. NP. 15 (M257.8T24.2) from *Saccharothrix* BBHARD27 extracts.

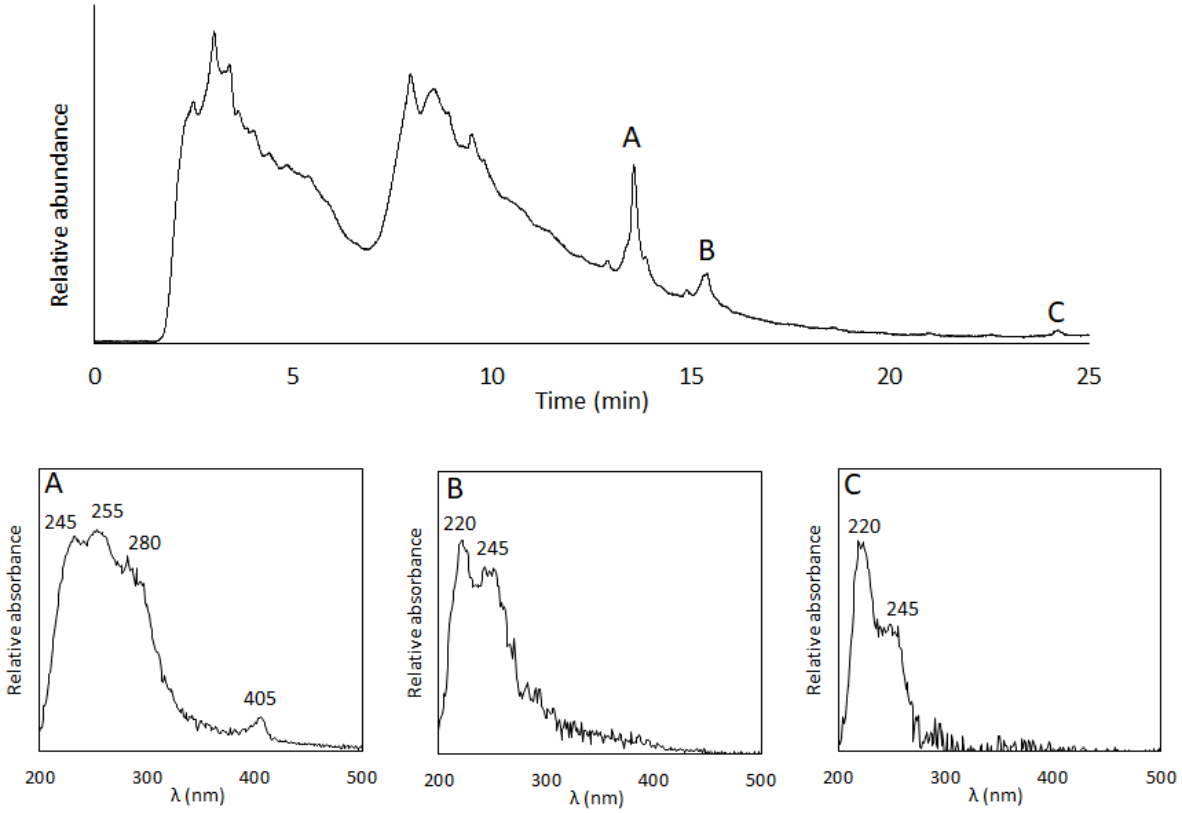


Figure B-20**:** UV spectrum of features from *Kribella* BBSNAI08 (A) unident. Polyene 5 (M730.2T20.3) from *Kribella* BBSNAI08 extracts.

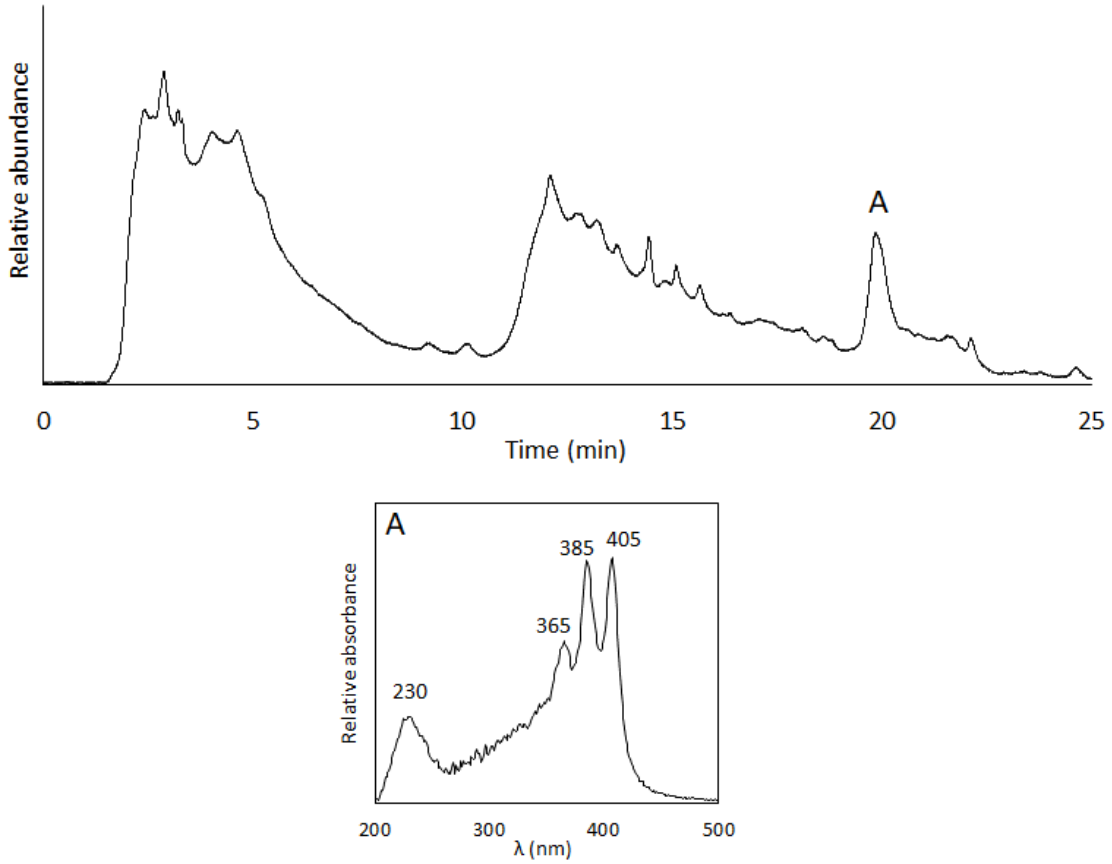


Figure B-21: UV spectrum of features from *Microbispora* BCCAGE54 (A) unident. NP. 16 and 17 (M689.2T10.5, M561.1T10.7), (B) unident. NP. 18 and 19 (M575.2T11.5, M717.2T11.6), (C) unident. NP. 20 (M731.3T12.1), (D) unident. NP. 21 (M355.4T13, *Arromycin*), (E) unident. NP. 22 (M393.2T13.7, *Linfuranone A*), (F) propeptin 1 (M1056.4T14), (G) propeptin 2 (M1146.9T14.6), (H) unident. tetarimycin 1 (M385.2T14.7), (I) unident. tetarimycin 2 (M399.2T16.1), (J) tetarimycin B (M369.3T17), (K), unident. tetarimycin 3 (M406.3T17.5) and (L) unident. tetarimycin 4 (M395.2T19.5) from *Microbispora* BCCAGE54 extracts.

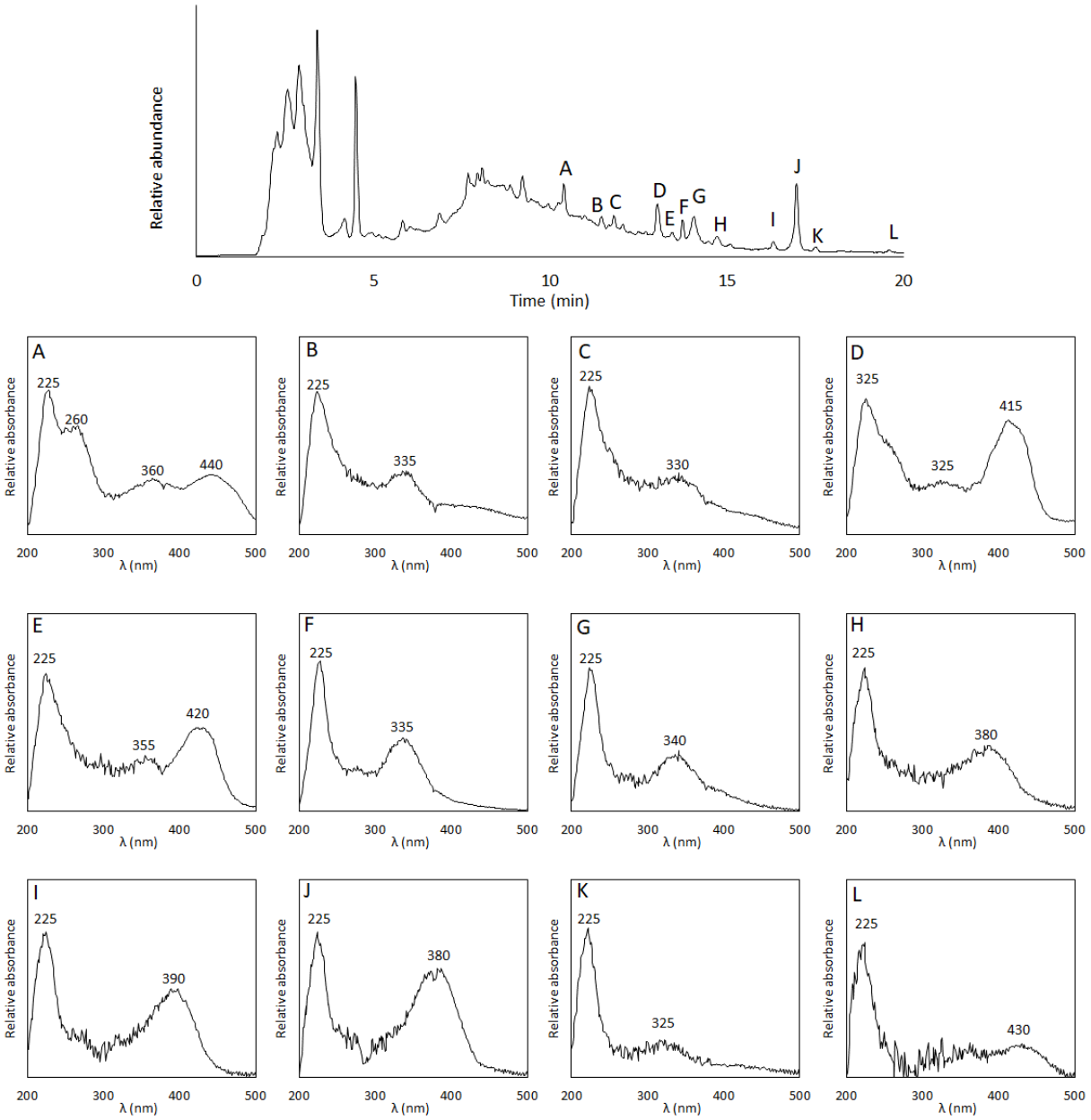


Figure B-22**:** UV spectrum of funisamine from *Streptosporangium* KDCAGE35

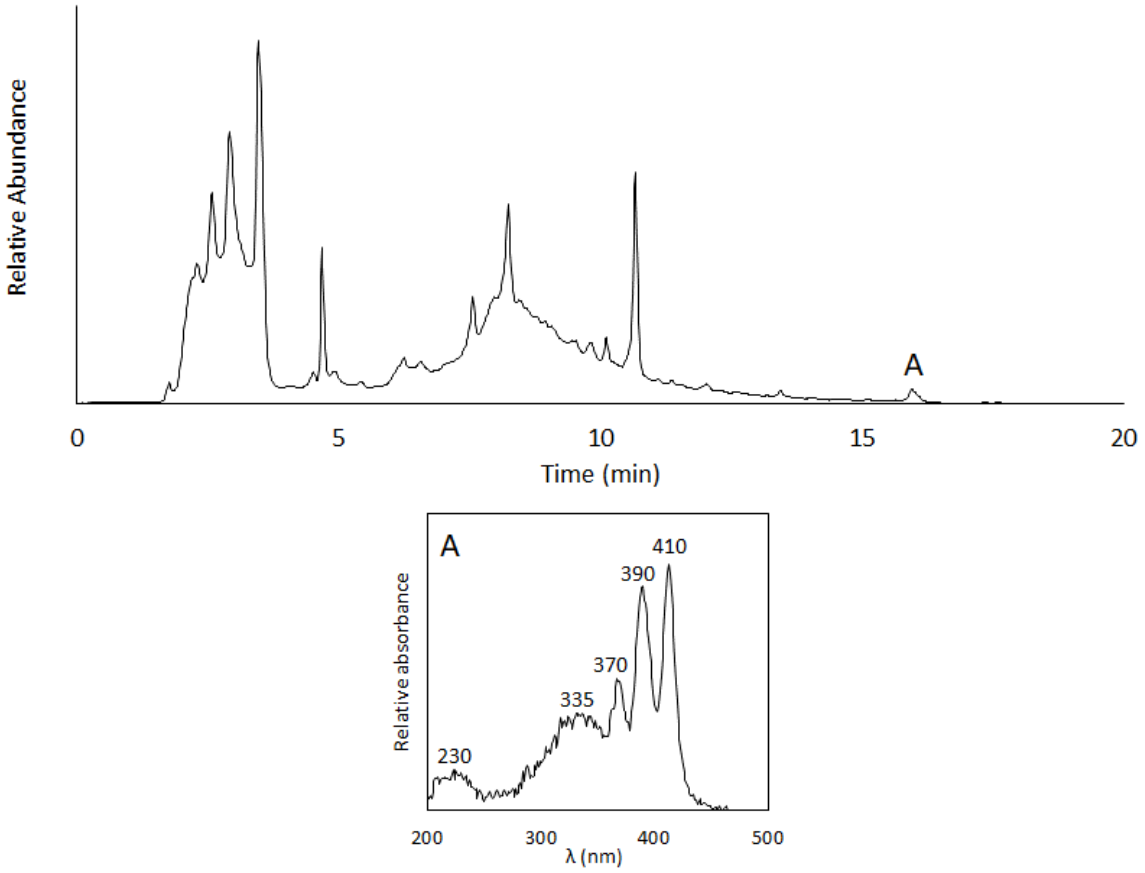


Figure B-23: H-NMR spectrum of aloesaponarin II from *Nonomuraea* BBHARD22 in Acetone-d6

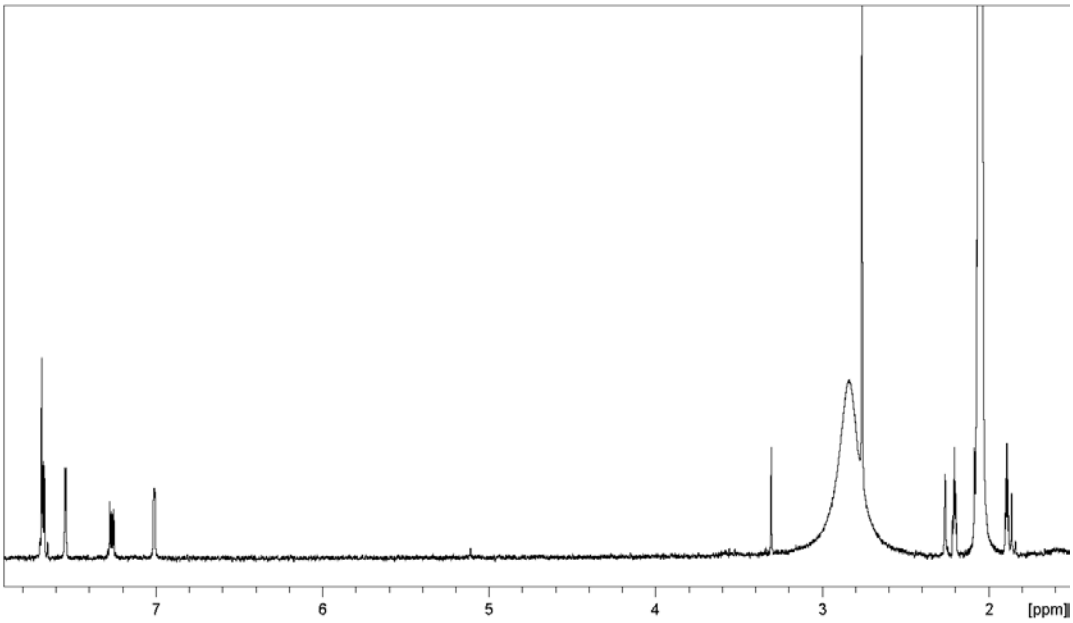


Figure B-24: Expanded H-NMR spectrum of aloesaponarin II from *Nonomuraea* BBHARD22 in Acetone-d6 showing the aromatic region.

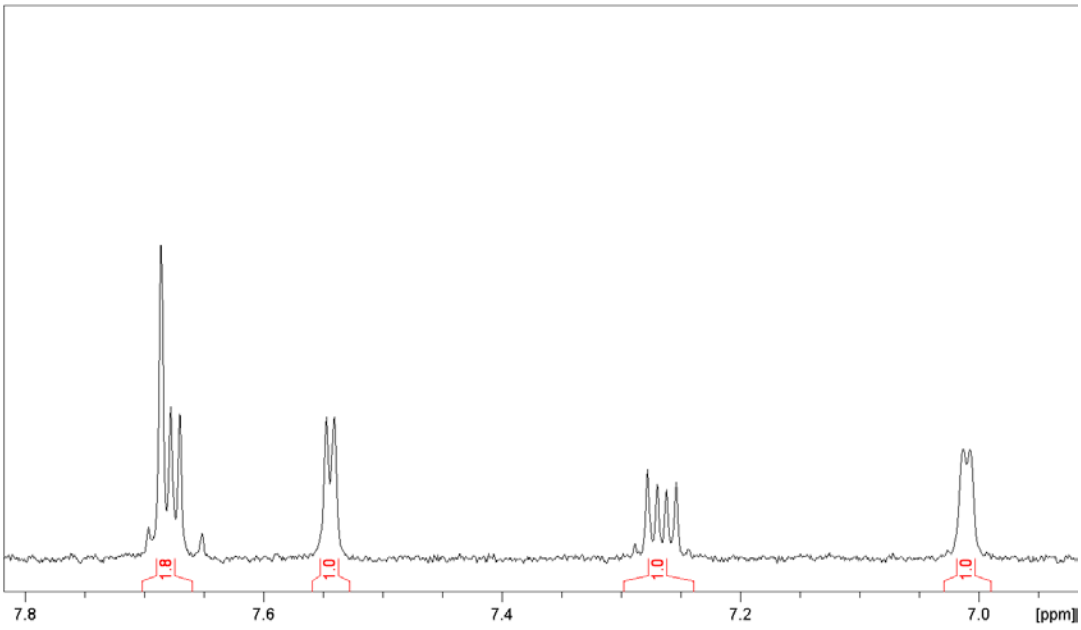


Figure B-25: H-NMR spectrum partially purified okicenone from *Nonomuraea* BBHARD22 in Methanol-d4. Peaks corresponding to okicenone indicated by (*).

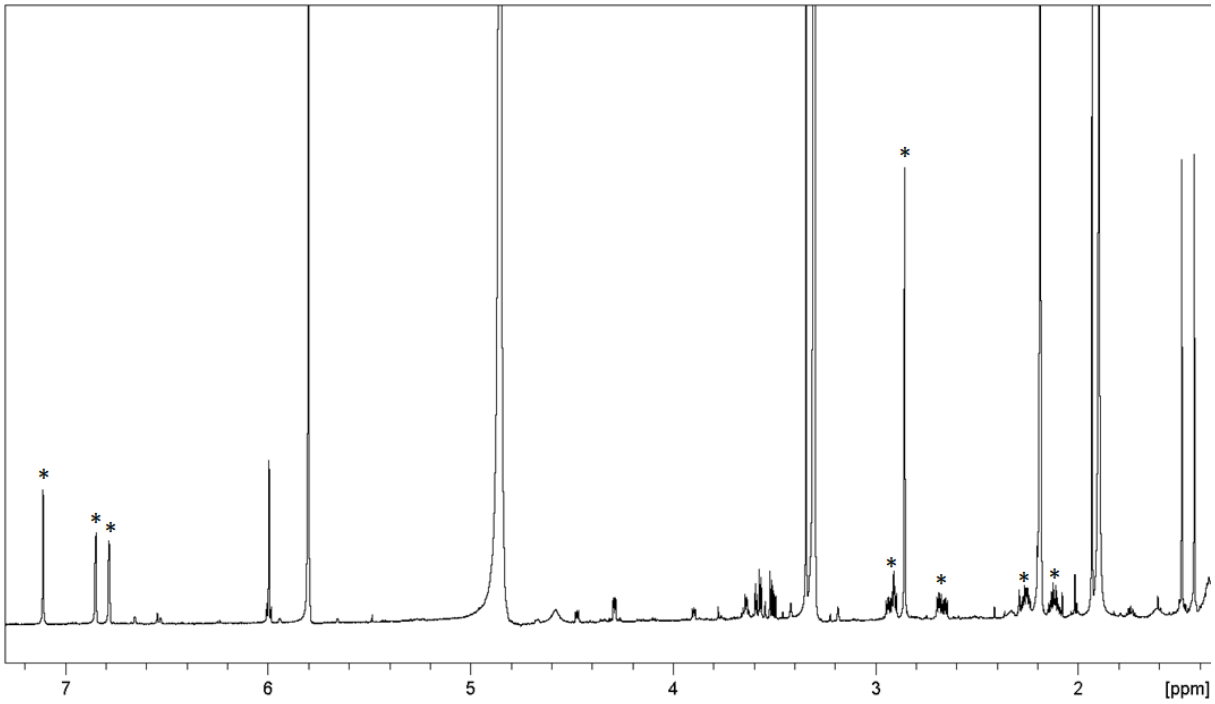


Figure B-26: H-NMR spectrum of tetarimycin B from *Microbispora* BCCAGE54 in Methanol-d4

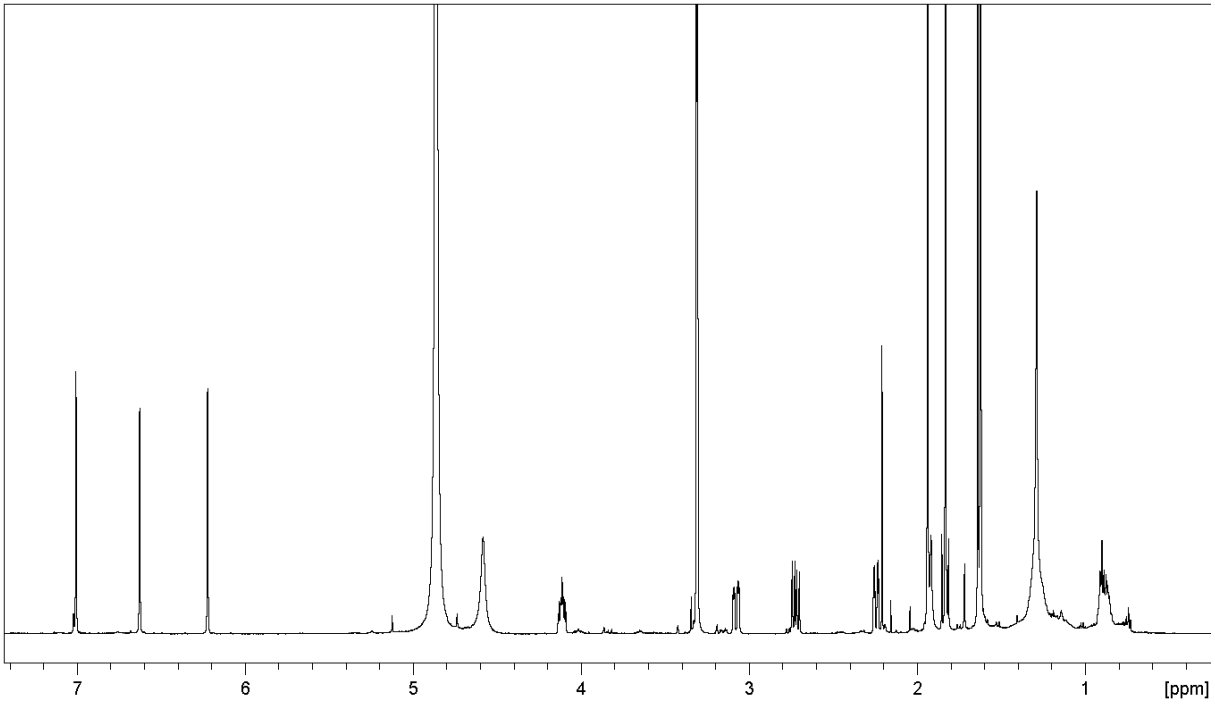


Figure B-27: H-NMR spectrum of hypogeamicin B from *Nonomuraea* BBHARD23 in Methanol-d4

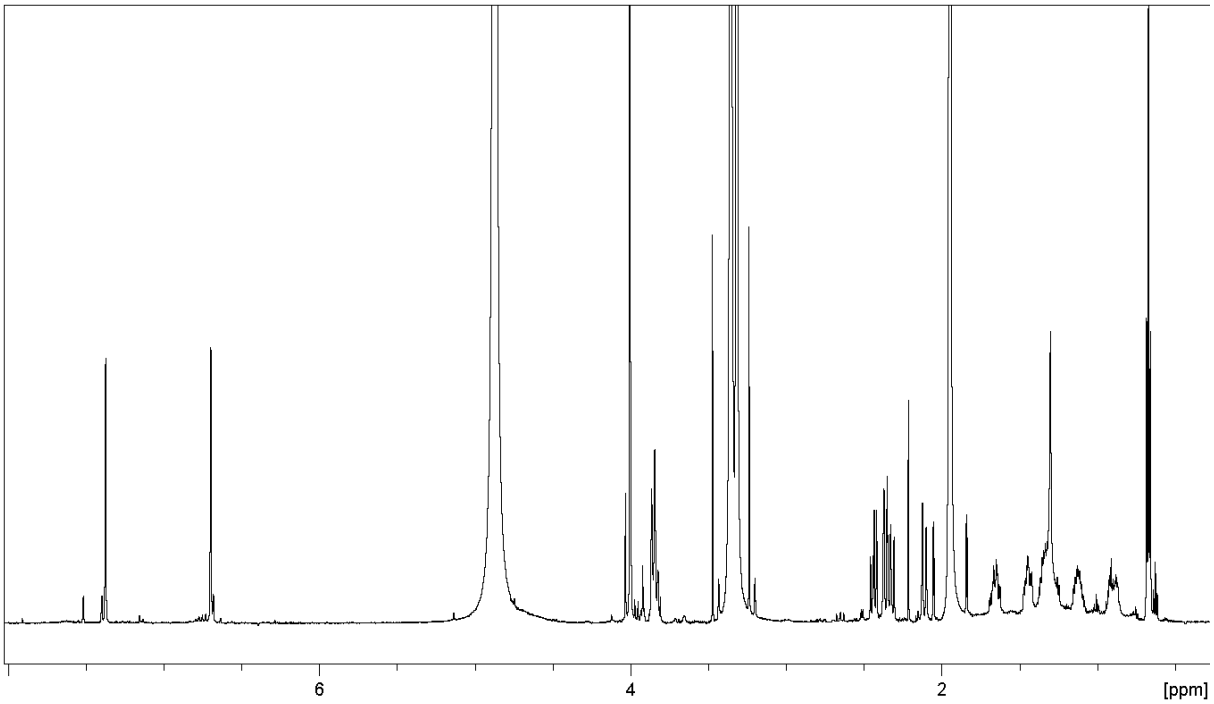


Figure B-28: H-NMR spectrum of actinomycin C from *Streptomyces* BCCAGE06 in Chloroform-d1

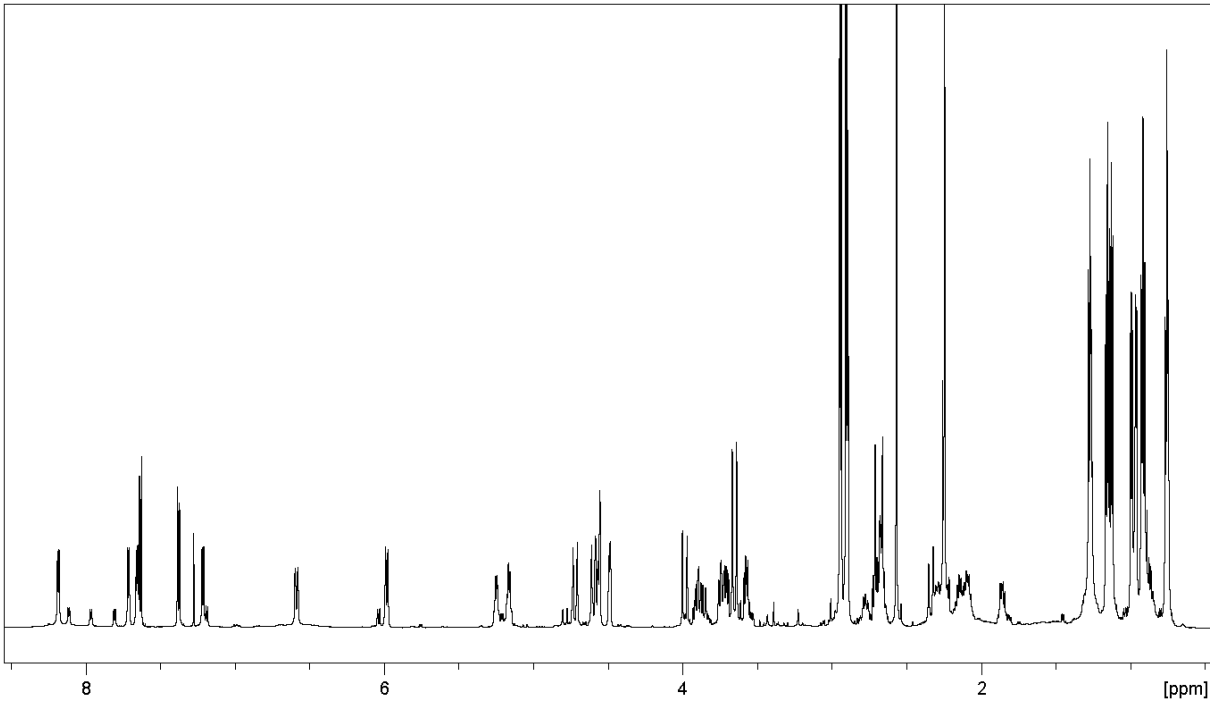


Table B-4: NMR data for funisamine

Pos.	δ_C	δ_H	HMBC	NOESY
1	172.6			
2	125.1	5.88, 1H, d (15.2 Hz)	4	
3	141.7	7.13, 1H, dd (11.7, 15.2 Hz)	1, 4, 5	5
4	129.9	6.32, 1H, dd (11.7, 14.5 Hz)		
5	138.9	6.53, 1H, dd (11, 14.5 Hz)	3	3, 7
6	130.0	6.22, 1H, dd (11, 14.8 Hz)		
7	136.2	6.35, 1H, dd (10.1, 14.8 Hz)	5, 6, 9	5
8	130.7	6.12, 1H, dd (10.1, 14.9 Hz)	10	
9	137.0	5.78, 1H, (dt (7.5, 14.9 Hz)	7, 10	
10	29.7	2.13, 2H, dd (7.2 Hz)	8, 9, 10, 12	8, 11, 12, 13
11	29.5	1.30, 1.41, 2H	8, 9, 10, 12, 13	
12	39.9	1.57, 1H		
13	78.3	3.93, 1H, d	10, 12, 59, 60	15
14	140.2			
15	126.1	6.13, 1H, d (10.5 Hz)	13, 59	17
16	128.7	6.52, 1H, dd (11.6, 14.5 Hz)	18	
17	133.2	6.26, 1H, overlap		
18	133.5	6.37, 1H, overlap		
19	133.5	6.37, 1H, overlap		
20	133.5	6.37, 1H, overlap		
21	133.5	6.37, 1H, overlap		
22	133.5	6.37, 1H, overlap		
23	133.0	6.36, 1H, overlap		
24	132.9	6.3, 1H, overlap		
25	131.8	6.25, 1H, overlap		
26	133.5	6.2, 1H, overlap		
27	129.8	5.82, 1H, dt (7.4, 14.7 Hz)	28, 29	
28	35.1	2.45, 2.62, 2H, overlap	26, 27, 29, 30	
29	79.0	4.61, 1H, dd (5.6 Hz)	27, 30, 31, 58	
30	47.5	3.2, 1H	28, 29, 31, 58	
31	215.5			
32	50.8	2.91, 1H,	31, 57	
33	66.7	4.01, 1H		
34	41.3	1.56, 1.71 2H	39	
35	70.9	4.25, 1H, q (6.8 Hz)	34, 36, 37	
36	135.0	5.53, 1H, dd (7.3, 15.3 Hz)	35, 38	
37	127.7	5.72, 1H, dt (7.2, 15.3 Hz)	35, 38	
38	40.1	2.24, 2H, multiplet	36, 37, 39, 40,	

Pos.	δ_C	δ_H	HMBC	NOESY
39	69.7	3.84, 1H		
40	43.1	1.56, 1.67, 2H		
41	68.6	3.99, 1H		38, 39
42	43.0	1.6, 2H		
43	66.8	4.04, 1H	45	
44	43.0	1.6, 2H		
45	64.5	4.1, 1H		
46	43.0	1.6. 2H		
47	69.0	4, 1H		
48	43.2	1.6, 2H		
49	69.0	4, 1H		
50	44.3	1.6, 2H		
51	68.6	3.96, 1H		
52	44.7	1.53, 2H		
53	68.9	3.81, 1H		
54	33.5	1.47, 1.62, 2H	53	
55	23.7	1.76, 1.81, 2H	53, 54, 56	
56	39.4	2.95, 2H	54, 55	
57	10.1	1.1, 3H, d (7 Hz)	31, 32	29, 30, 32, 33, 35
58	12.2	1.14, 3H, d (7 Hz)	29, 30, 31	28, 29, 30, 32, 33
59	11.8	1.74, 3H, s	13, 14, 15, 16	
60	30.7	1.26, 1.47 2H		
61	19.5	1.25, 1.42, 2H	9	8, 10, 11, 13, 60, 62
62	13.4	0.9, 3H, t (7.1 Hz)		61

Figure B-29: H-NMR (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

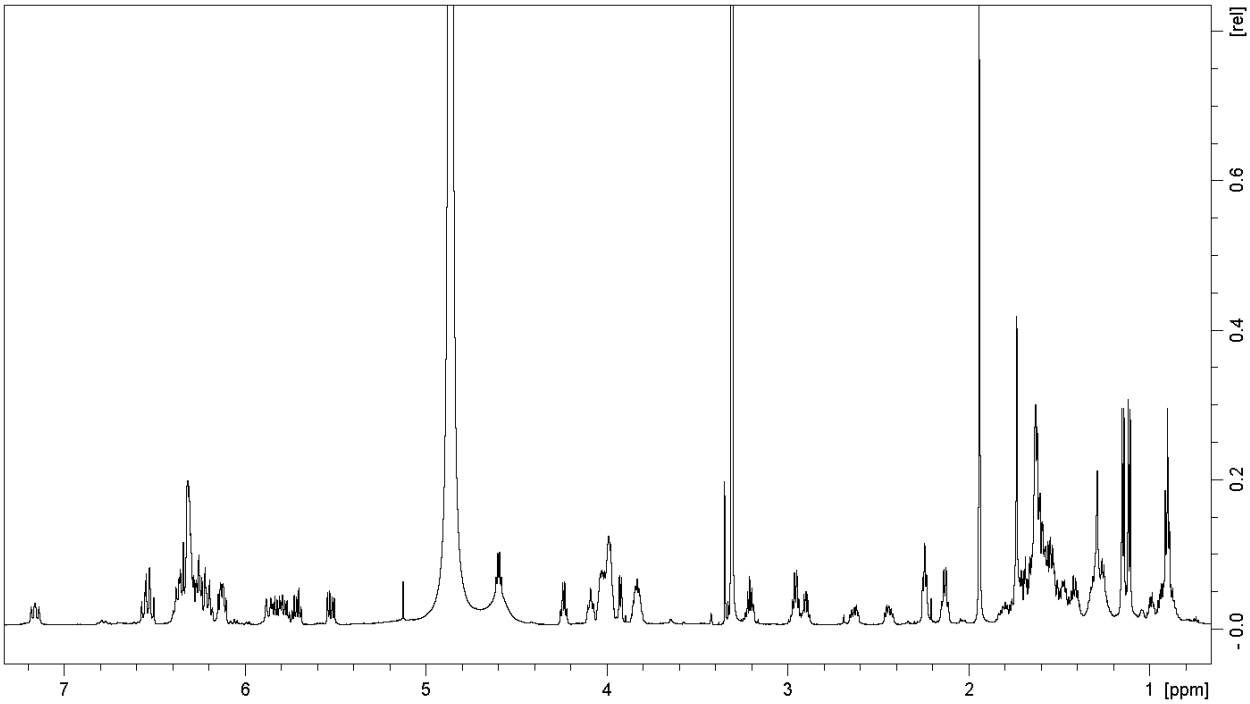


Figure B-30: C-NMR (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

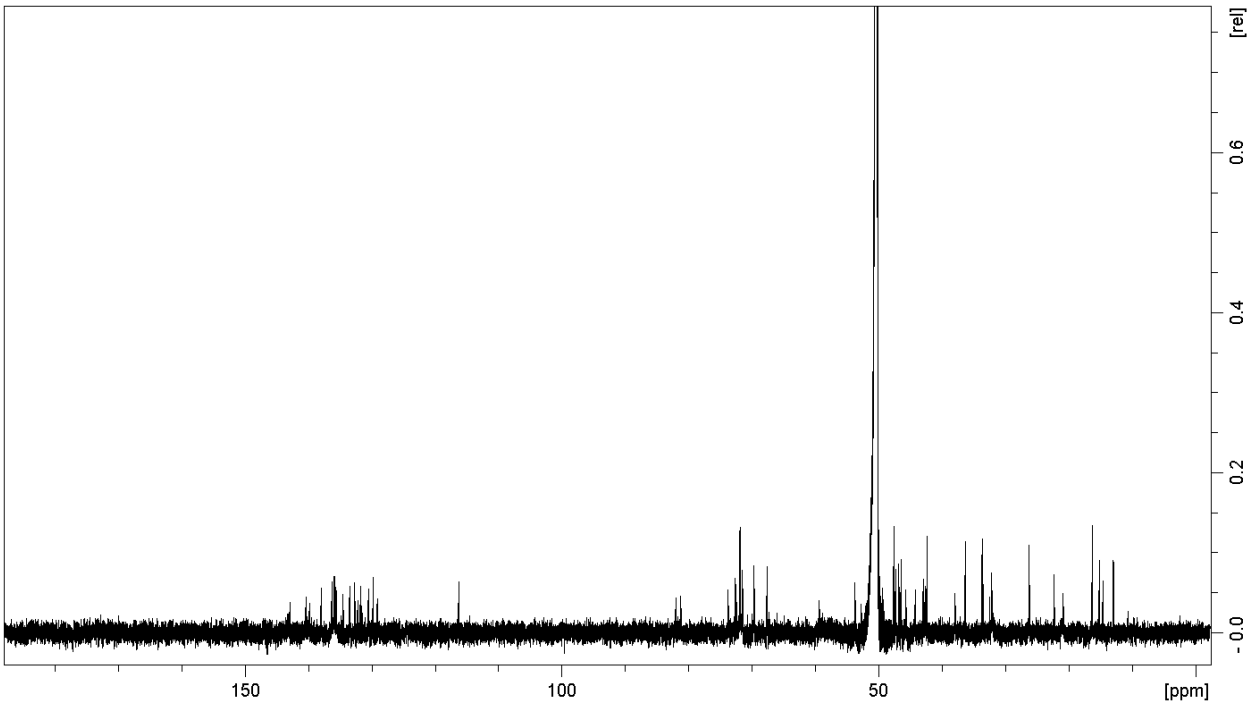


Figure B-31: COSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

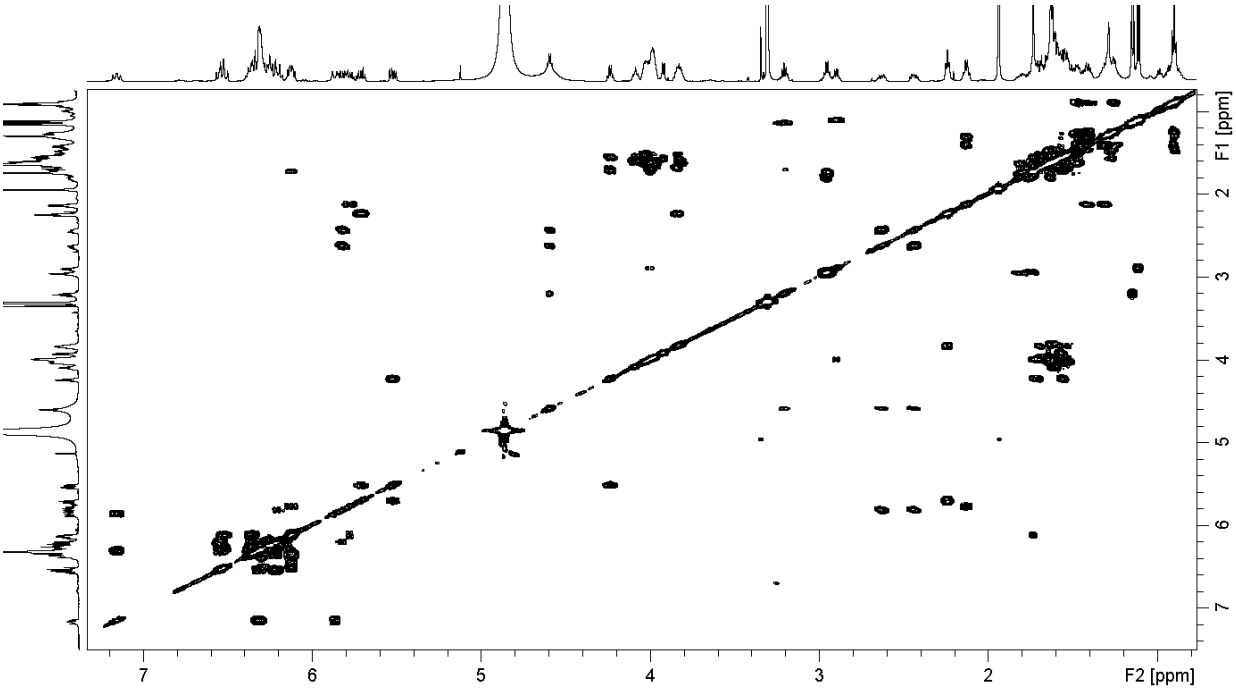


Figure B-32: Expanded COSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to show aromatic couplings.

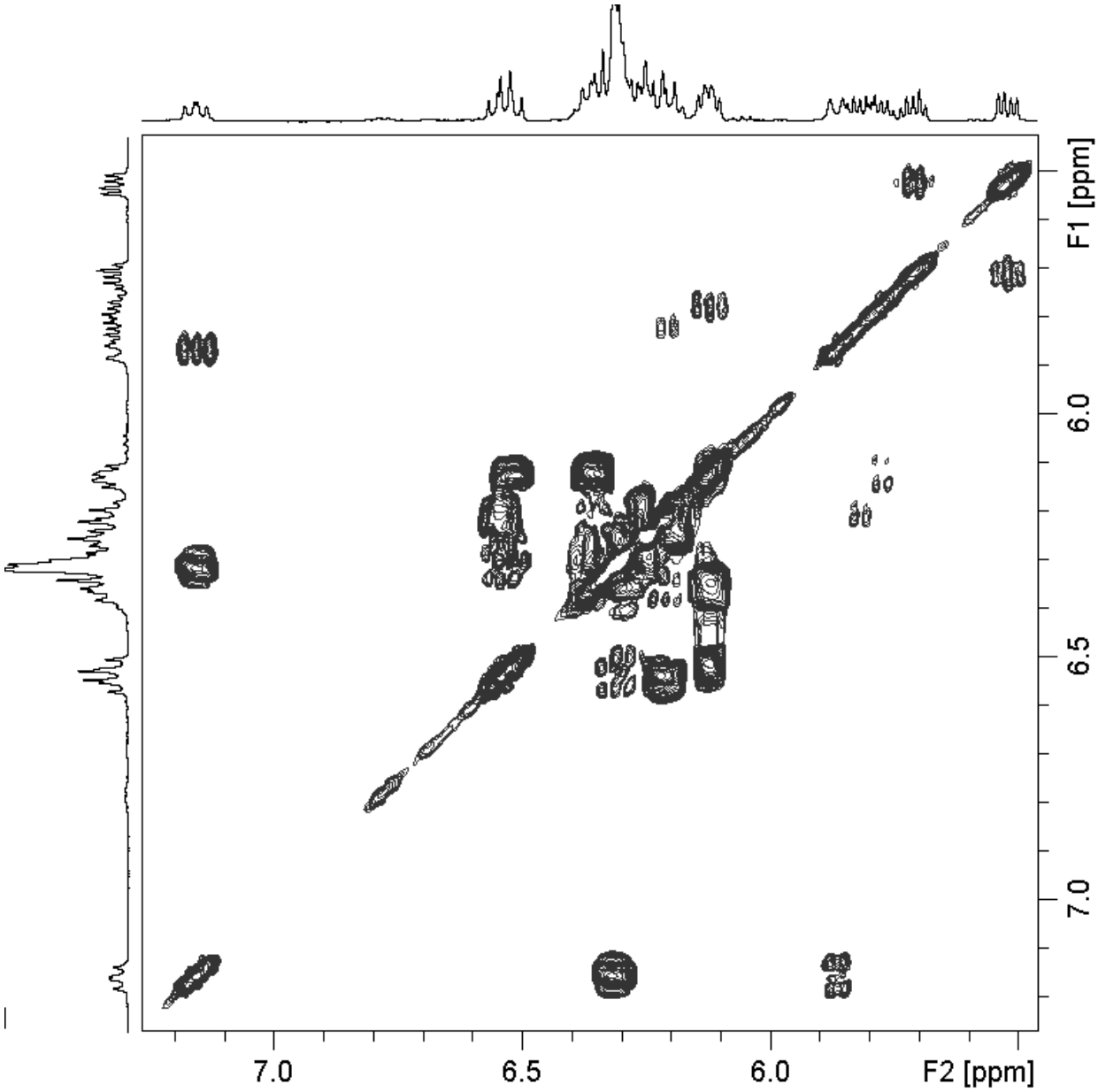


Figure B-33: Expanded COSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to methylene and hydroxy-methine couplings.

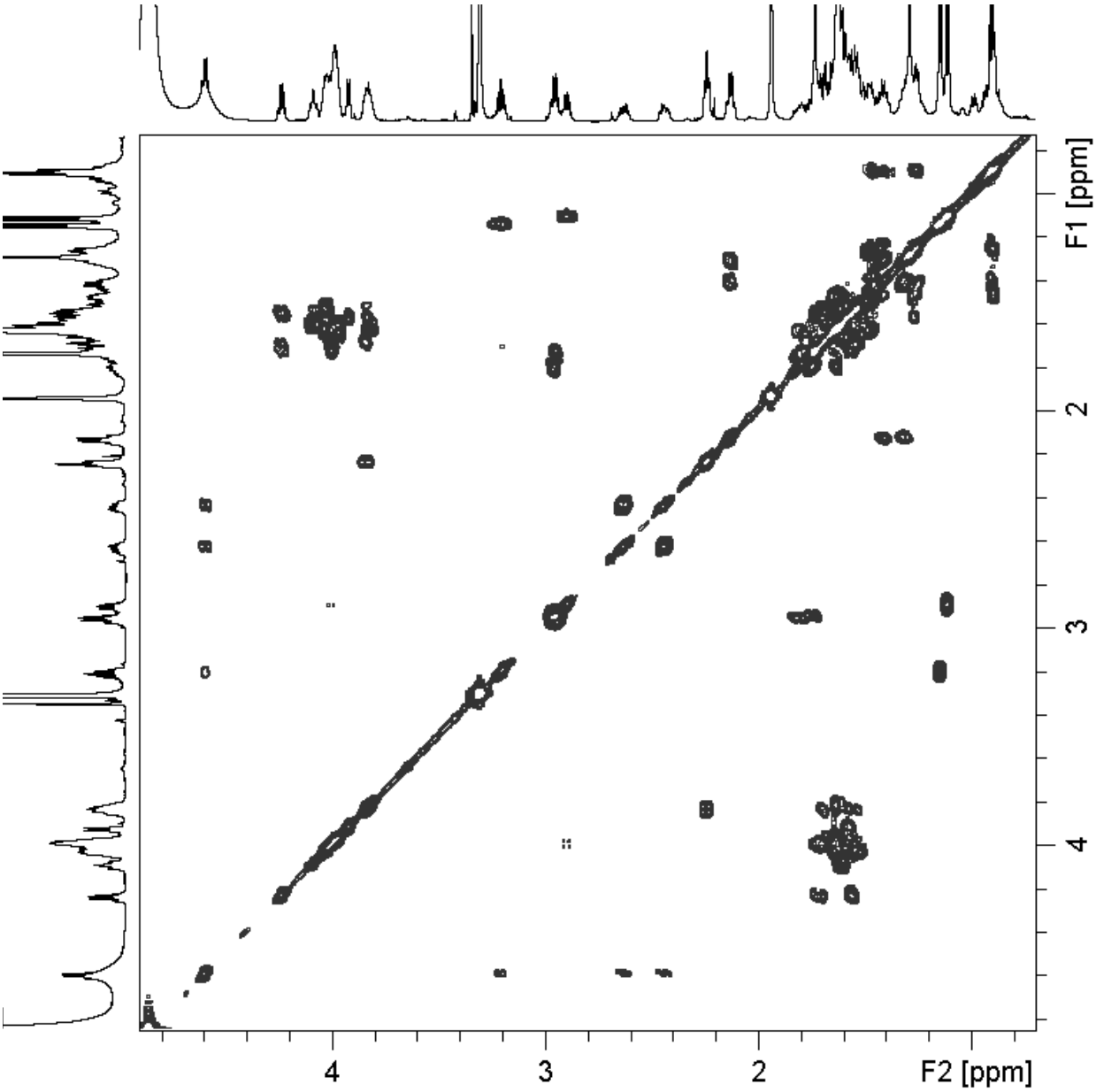


Figure B-34: TOCSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

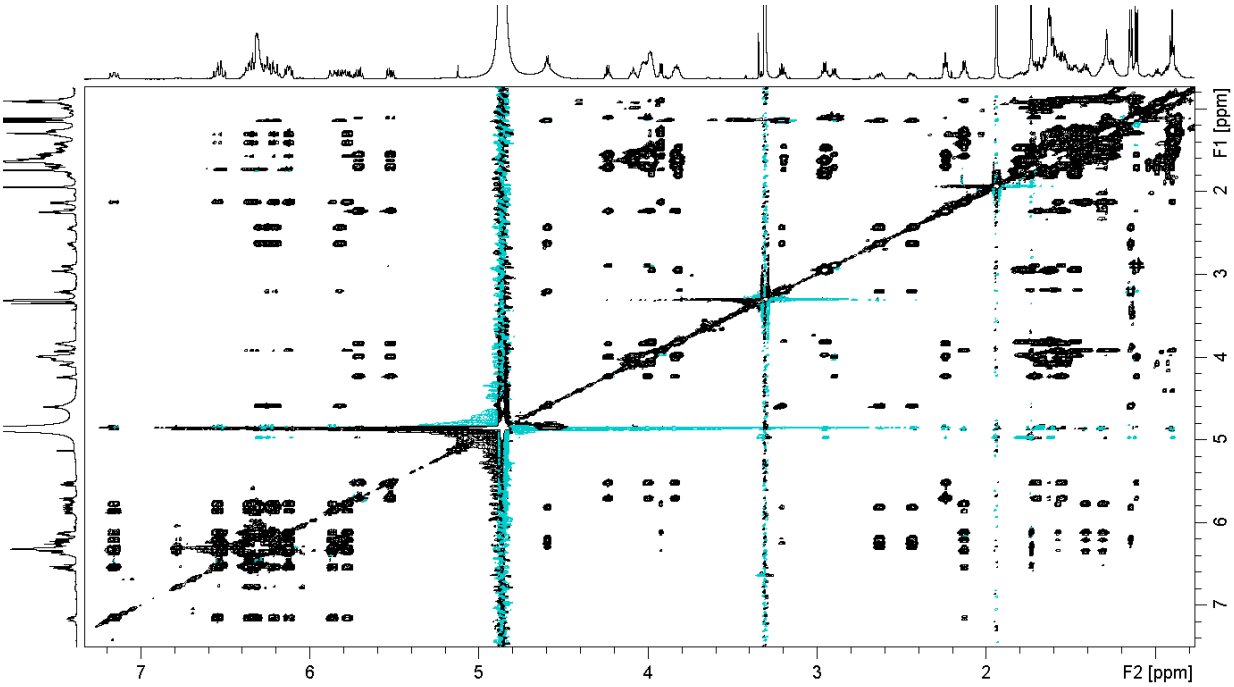


Figure B-35: TOCSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to show upfield couplings.

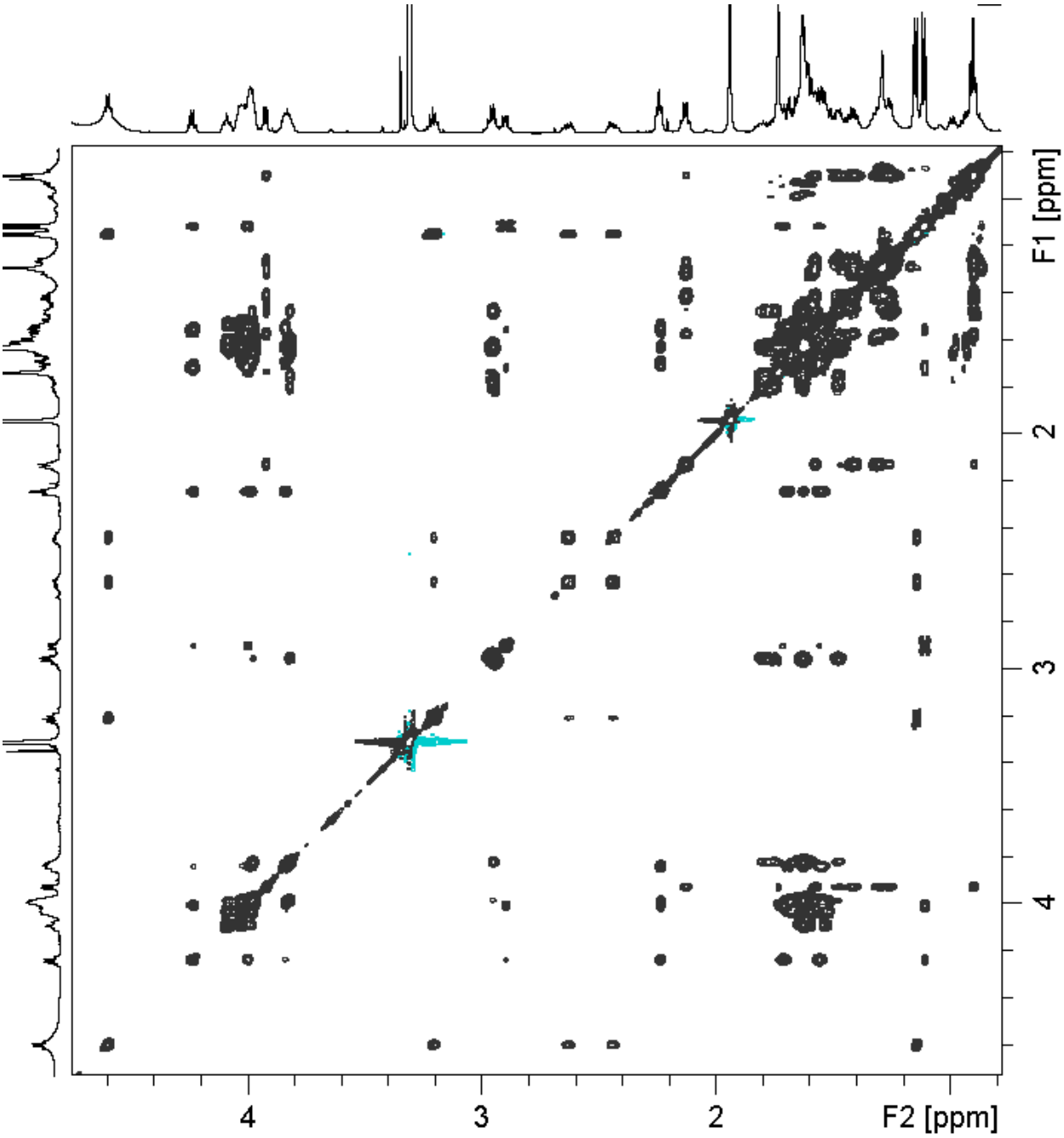


Figure B-36: TOCSY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to show downfield couplings.

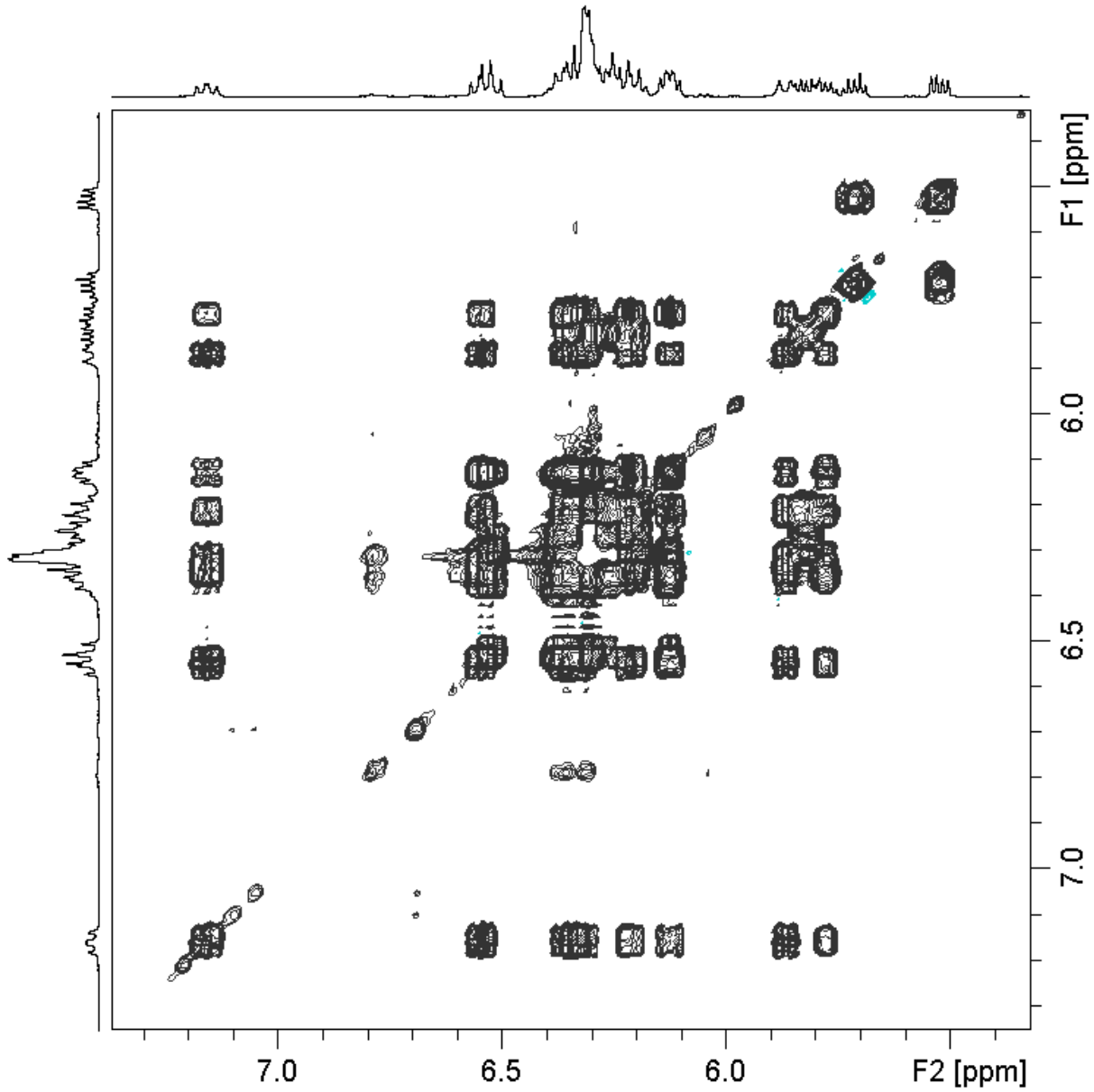


Figure B-37: HSQC (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

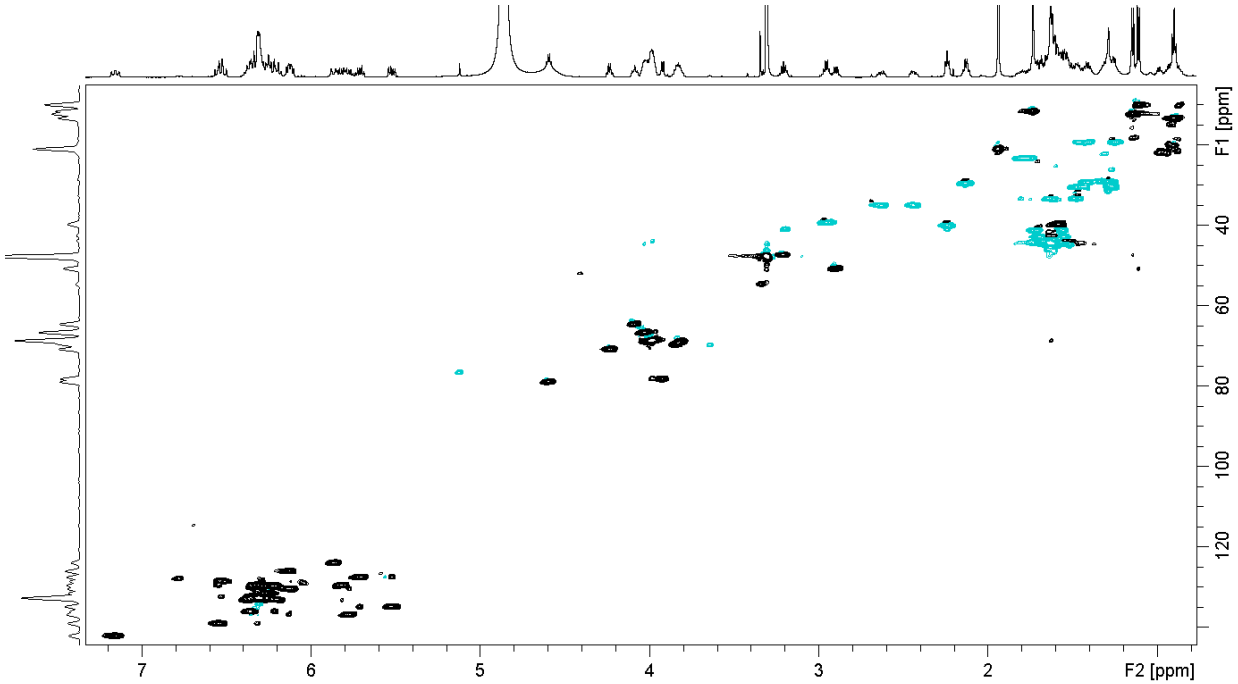


Figure B-38: HSQC (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to show downfield couplings.

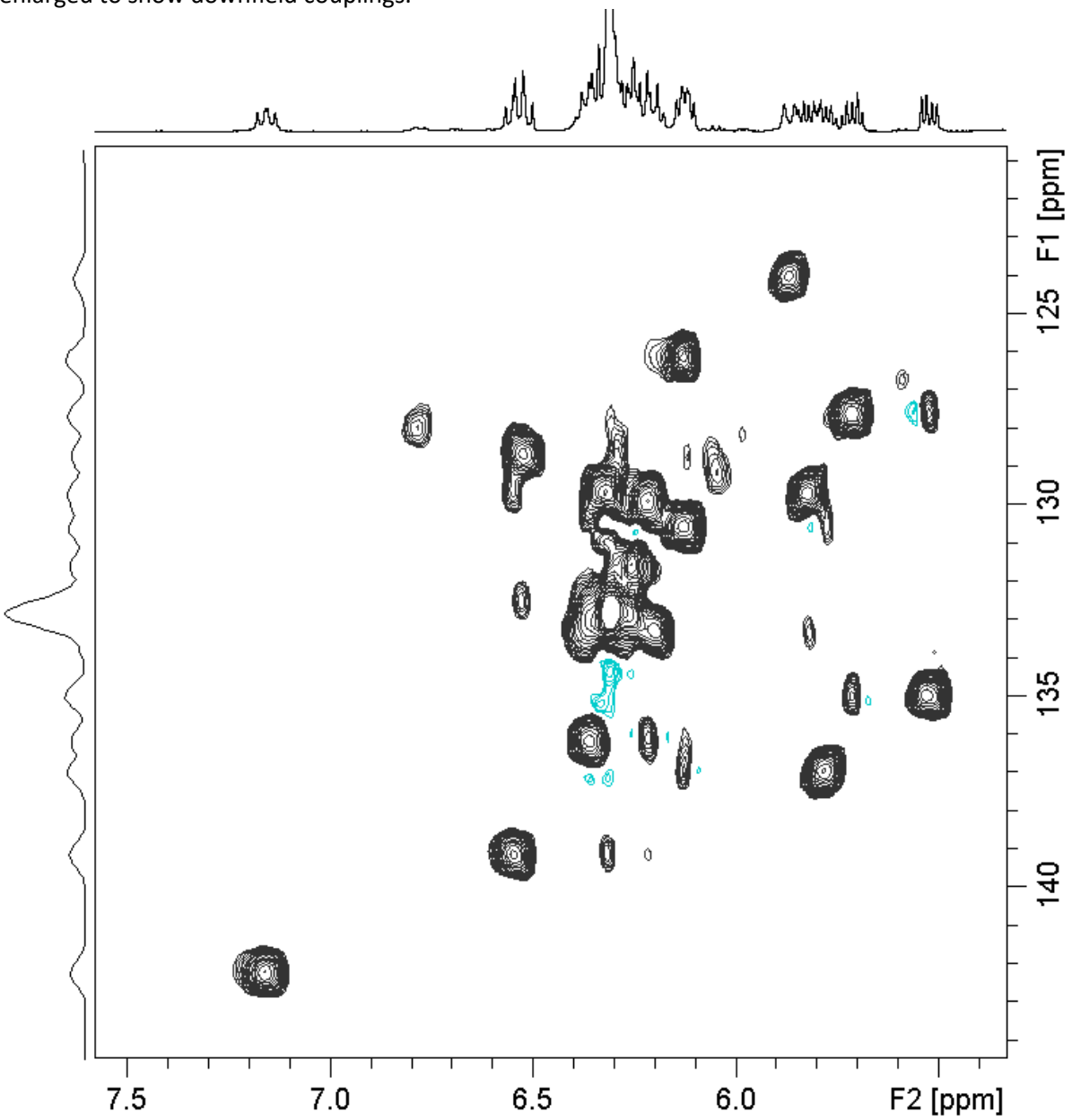


Figure B-39: HSQC (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4 enlarged to show upfield couplings.

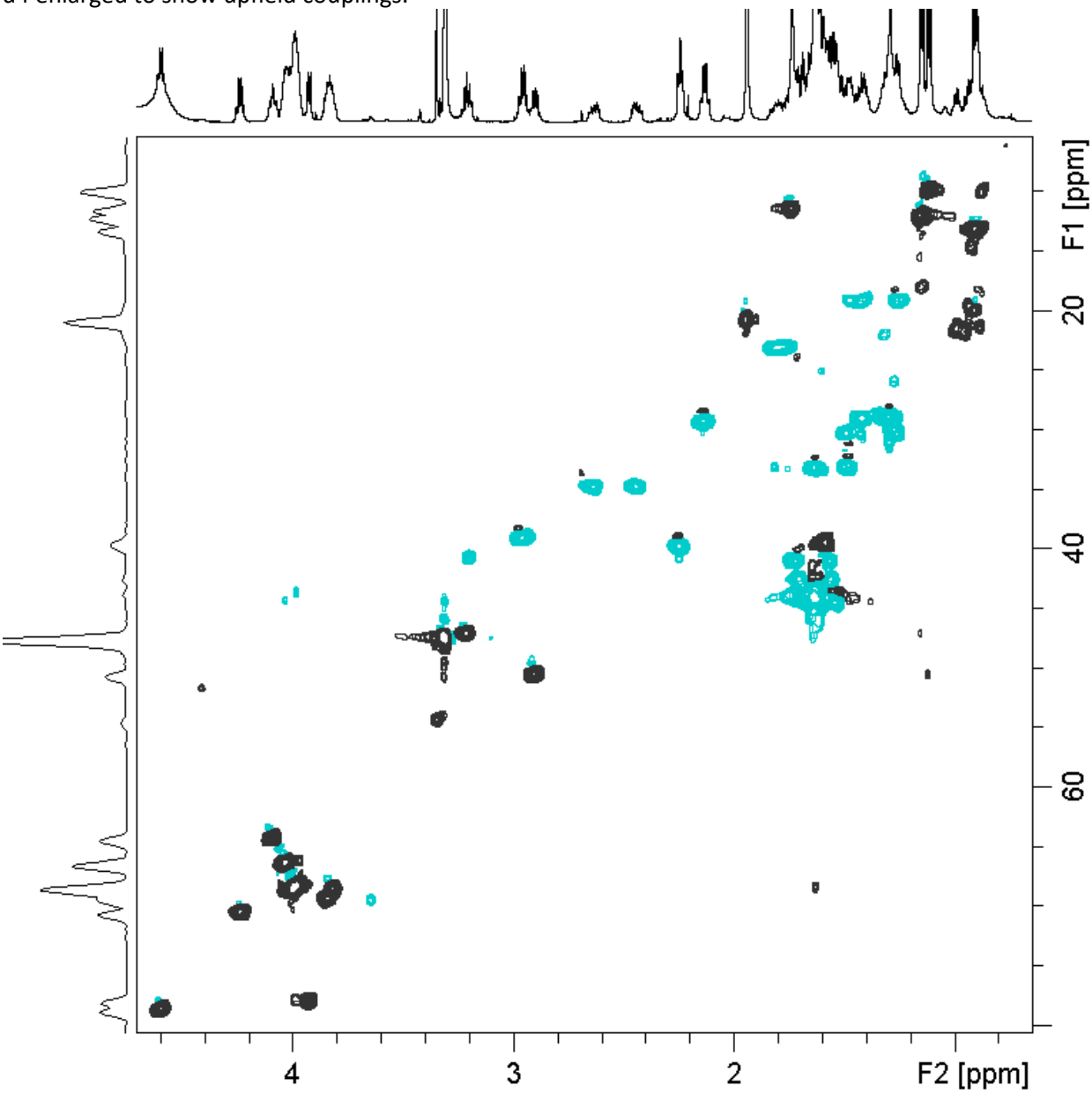


Figure B-40: HMBC (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

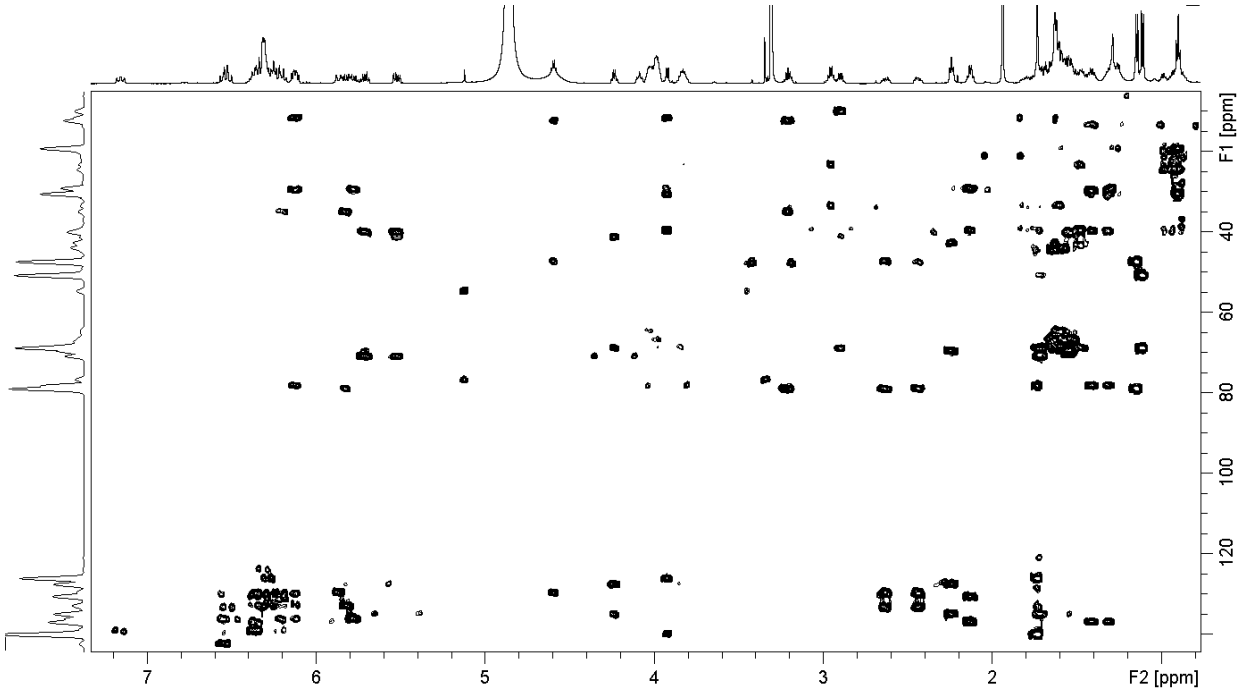


Figure B-41: NOESY (600 MHz) spectrum of funisamine from *Streptosporangium* KDCAGE35 in Methanol-d4

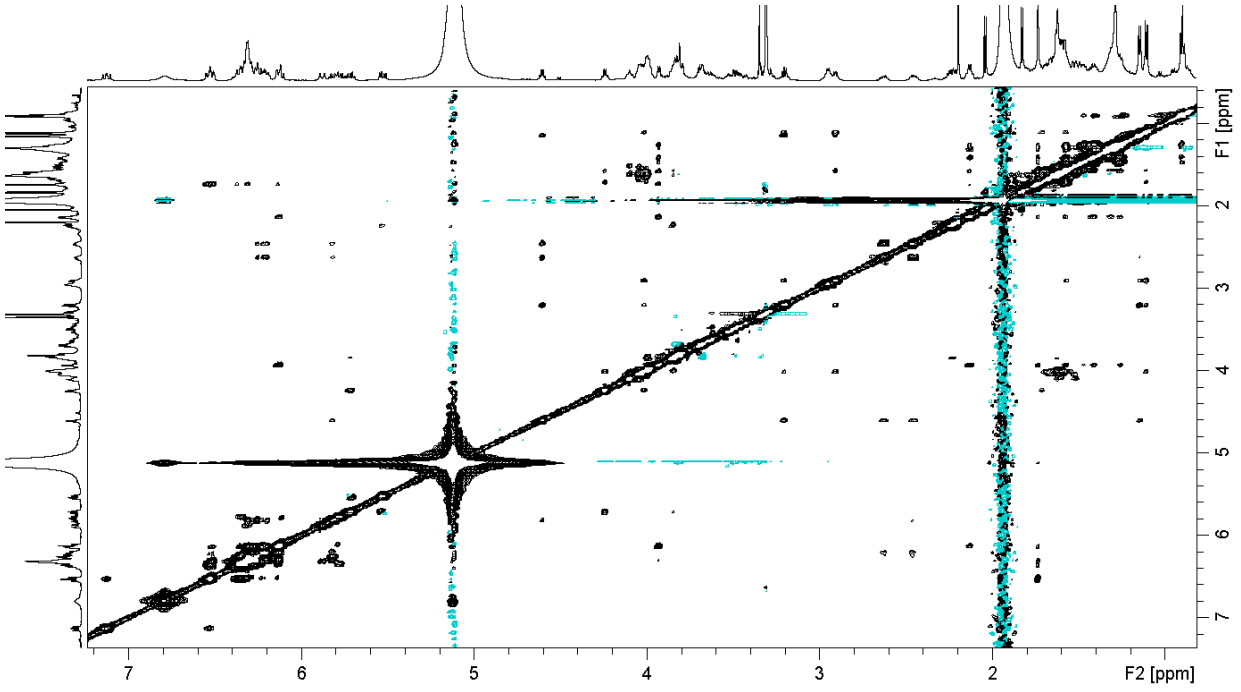


Figure B-42**:** High-Res mass spectrum for purified funisamine acquired in negative mode on a 15T solariX Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer (Bruker Daltonics, Billerica, MA, USA)

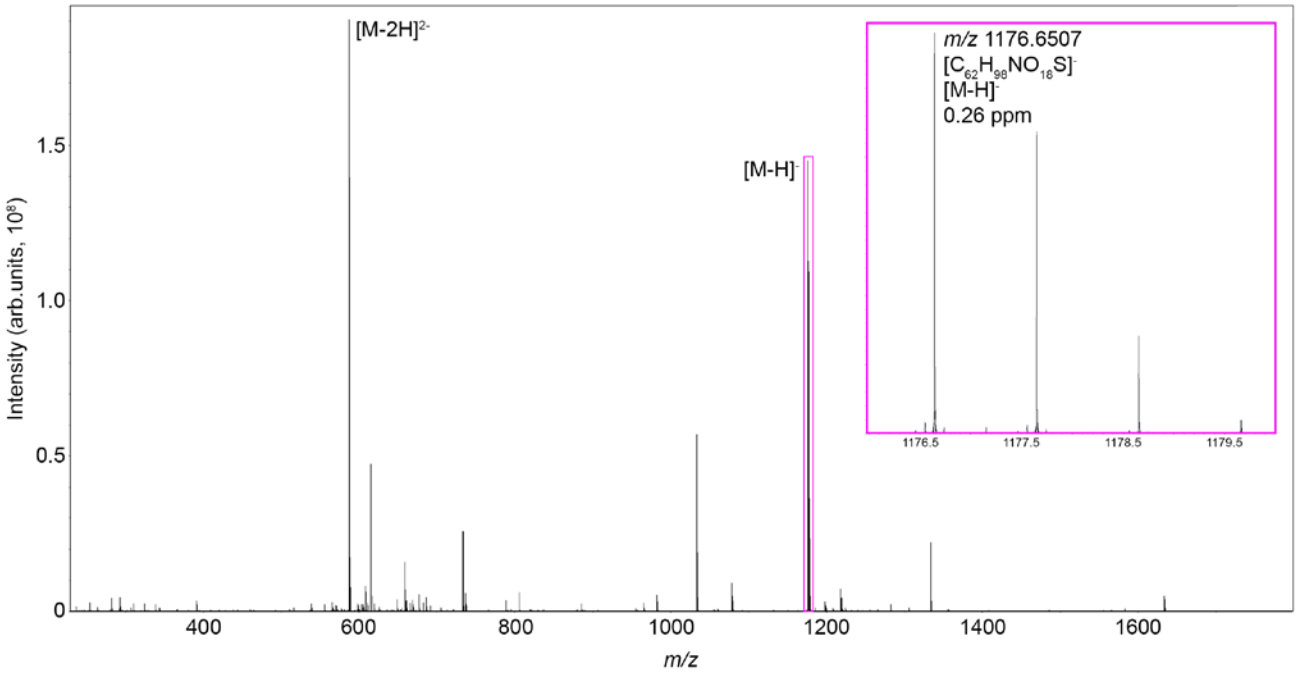


Table B-5: Table of ions detected through fragmentation analysis of purified funisamine

Observed m/z	Intensity	Molecular Forumula	Theoretical m/z	Mass Accuracy (ppm)
1176.65095	6.84E+07	C62H98NO18S-	1176.6510	0.05
1158.64094	1.52E+06	C62H96NO17S-	1158.6404	-0.47
1098.65309	5.97E+05	UNK		
1098.60508	9.70E+06	C58H92NO18S-	1098.6041	-0.93
1080.59553	7.35E+05	C58H90NO17S-	1080.5935	-1.88
1078.68607	7.53E+05	C62H96NO14-	1078.6836	-2.29
697.34181	6.43E+06	C39H53O9S-	697.3416	-0.30
679.33228	4.08E+05	C39H51O8S-	679.3310	-1.88
674.34418	3.56E+05	C29H56NO14S-	674.3427	-2.19
668.33516	1.03E+06	C41H48O8-	668.3355	0.46
667.33181	9.67E+05	C38H51O8S-	667.3310	-1.21
639.29996	1.12E+06	C36H47O8S-	639.2997	-0.41
619.29486	3.81E+06	C33H47O9S-	619.2946	-0.42
589.34925	5.33E+05	UNK		
589.28417	4.51E+06	C32H45O8S-	589.2841	-0.19
588.34589	8.49E+05	C37H48O6-	588.3456	-0.42
588.26933	3.74E+05	C24H46NO13S-	588.2695	0.36
570.25987	4.56E+05	C24H44NO12S-	570.2590	-1.58
561.25307	5.41E+05	C30H41O8S-	561.2528	-0.48
558.2599	2.53E+05	C23H44NO12S-	558.2590	-1.67
549.30256	3.05E+05	C37H41O4-	549.3010	-2.84
544.2437	6.34E+05	C22H42NO12S-	544.2433	-0.70
501.30079	3.55E+05	C33H41O4-	501.3010	0.42
464.28649	2.41E+05	C22H42NO9-	464.2865	0.02
457.31085	4.60E+05	C32H41O2-	457.3112	0.77
446.2759	2.68E+05	C22H40NO8-	446.2759	0.00
428.26553	2.59E+05	C22H38NO7-	428.2654	-0.30
394.24449	4.36E+05	C18H36NO8-	394.2446	0.28
350.21822	7.26E+05	C16H32NO7-	350.2184	0.51
317.20884	6.37E+05	UNK		
306.19217	1.24E+06	C14H28NO6-	306.1922	0.13
262.16589	7.44E+05	C12H24NO5-	262.1660	0.42
226.1448	2.21E+05	C12H20NO3-	226.1449	0.44
209.04884	2.28E+05	C7H13O5S-	209.0489	0.29
173.09714	3.45E+05	C12H13O-	173.0972	0.35
147.08152	4.34E+05	C10H11O-	147.0815	-0.14
121.06587	4.40E+05	C8H9O-	121.0659	0.25
96.9601	1.36E+07	HSO4-	96.9601	0.00

Figure B-43: Fragmentation spectra of isolated funisamine

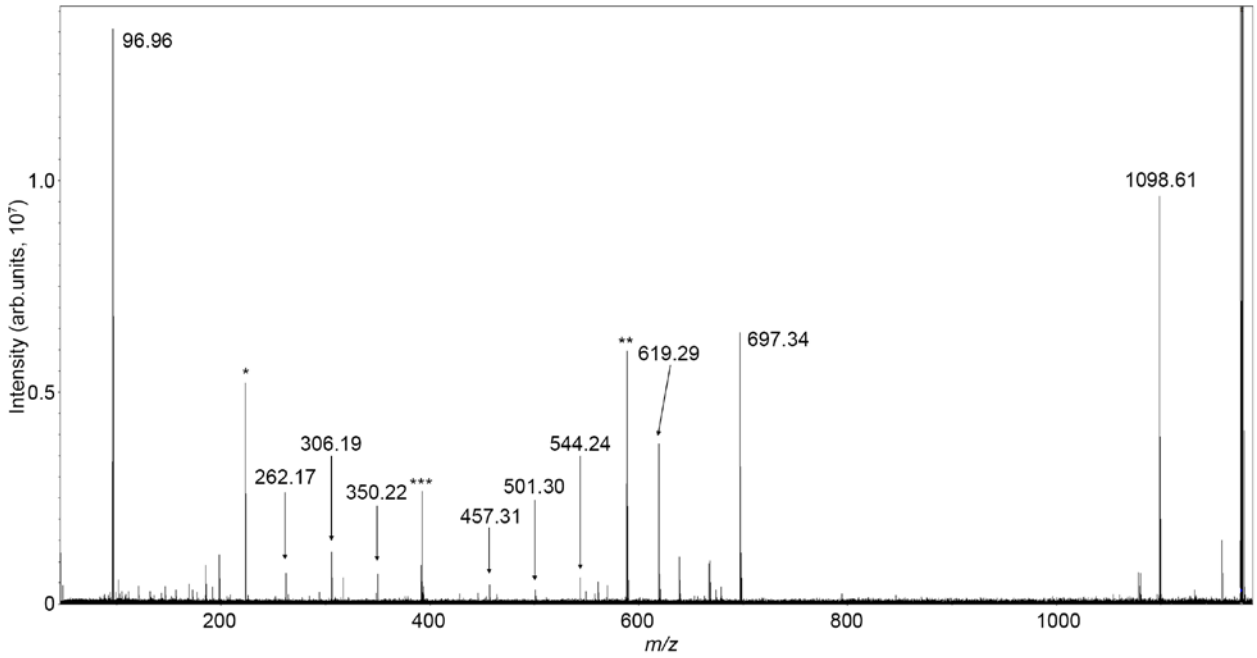


Figure B-44: Fragmentation assignments for funisamine

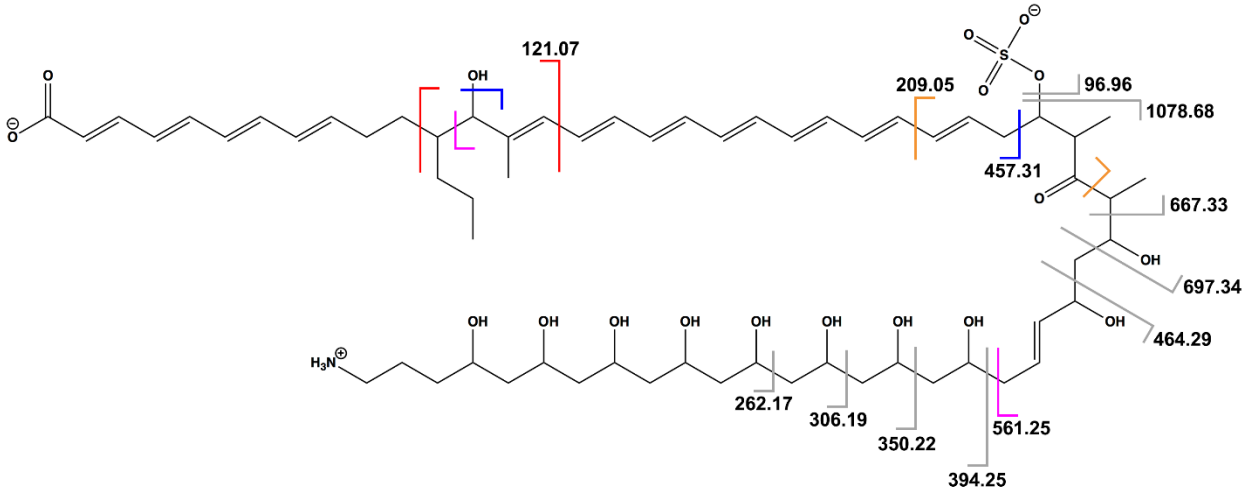


Figure B-45**:** (A) S-plot prioritization of amicoumacin B and funisamine from mixed culture extract with *Streptosporangium* KDCAGE35 and *B. subtilis*. Extracted ion chromatograms for labeled features are shown overlaid on the total ion chromatogram.

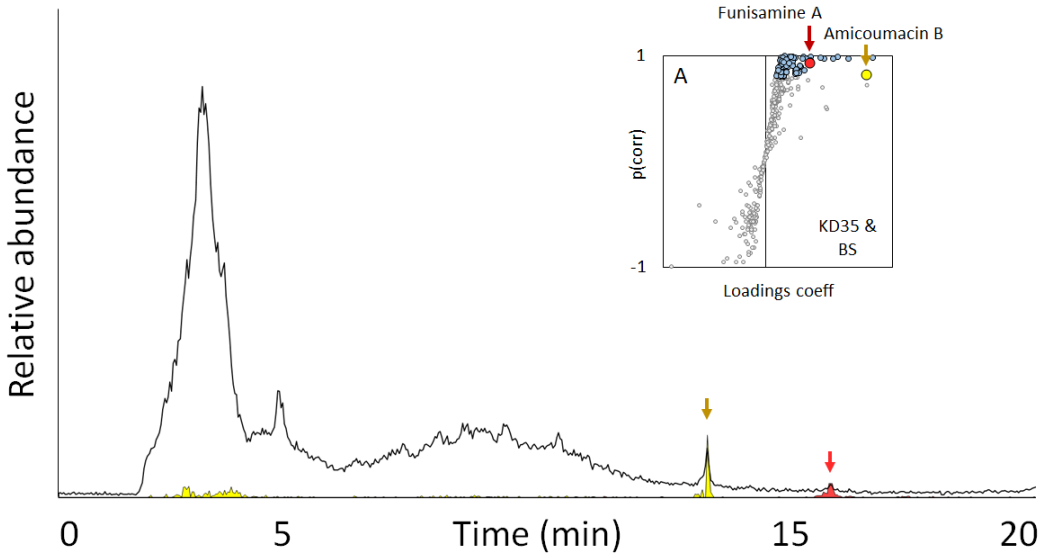


Figure B-46**:** Mixed culture competitor growth curves for TP, RW, EC, BS, and BK in international Streptomyces protocol 2 (ISP2) broth. Doubling times in hours were calculated to be: 1.8 (TP), 2.8 (RW), 1.5 (EC), 1.5 (BS), and 1.0 (BK).

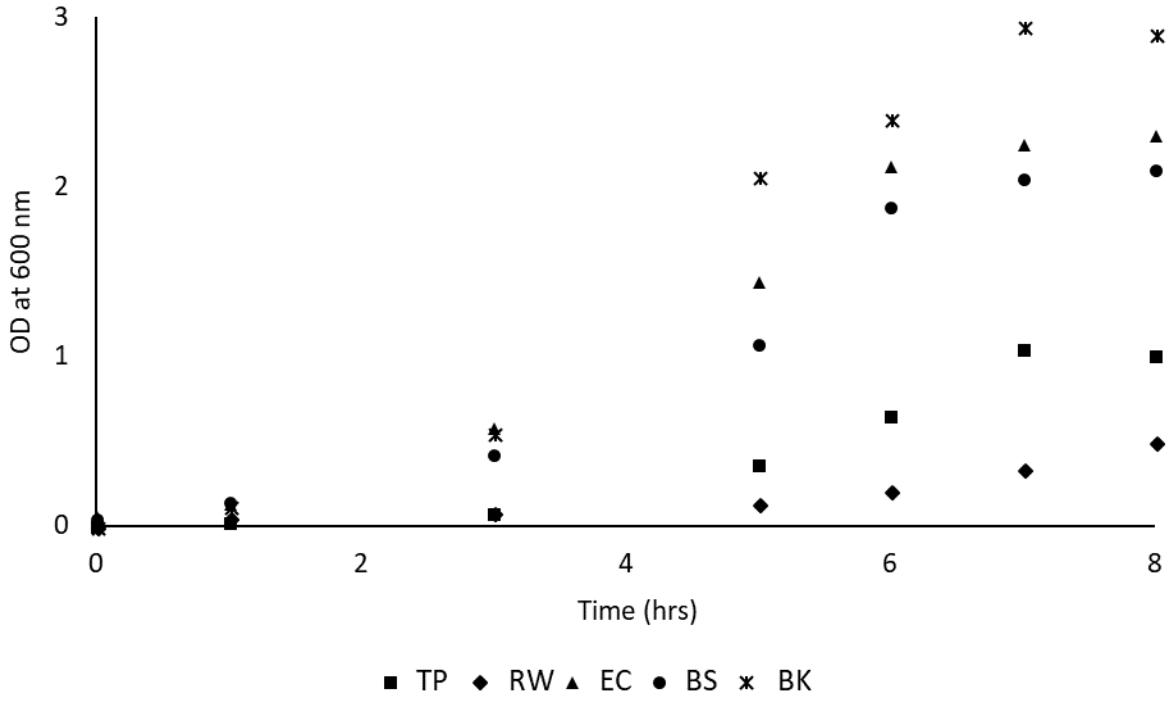
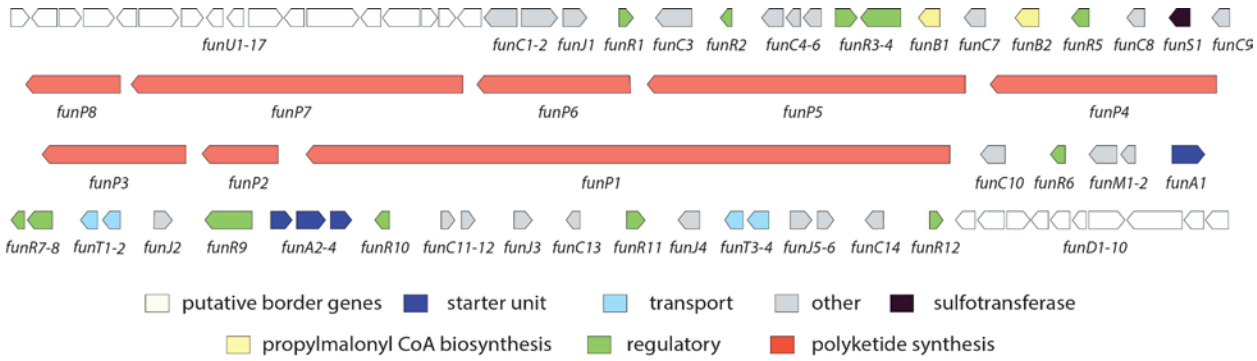


Table B-6: Table of putative BGCs identified by antiSMASH in the genome of *Streptosporangium* KDCAGE35

Cluster	Type	Most similar known cluster from antiSMASH
Cluster 1	Lassopeptide	-
Cluster 2	Nrps	Daptomycin_biosynthetic_gene_cluster (4% of genes show similarity)
Cluster 3	Nrps	Landomycin_biosynthetic_gene_cluster (9% of genes show similarity)
Cluster 4	Nrps	Naphthyridinomycin_biosynthetic_gene_cluster (17% of genes show similarity)
Cluster 5	Terpene	Chlortetracycline_biosynthetic_gene_cluster (5% of genes show similarity)
Cluster 6	T3pks	Alkylresorcinol_biosynthetic_gene_cluster (100% of genes show similarity)
Cluster 7	Nrps	Scabichelin_biosynthetic_gene_cluster (100% of genes show similarity)
Cluster 8	Bacteriocin	Hedamycin_biosynthetic_gene_cluster (6% of genes show similarity)
Cluster 9	Terpene	Spinosad_biosynthetic_gene_cluster (8% of genes show similarity)
Cluster 10	T3pks-T1pks	Kendomycin_biosynthetic_gene_cluster (55% of genes show similarity)
Cluster 11	Terpene	Sioxanthin_biosynthetic_gene_cluster (40% of genes show similarity)
Cluster 12	Lantipeptide	-
Cluster 13	Terpene	Griseobactin_biosynthetic_gene_cluster (11% of genes show similarity)
Cluster 14	Other	Monensin_biosynthetic_gene_cluster (5% of genes show similarity)
Cluster 15	Lantipeptide	Catenulipeptin_biosynthetic_gene_cluster (60% of genes show similarity)
Cluster 16	Terpene-Nrps	-
Cluster 17	T1pks	Concanamycin_A_biosynthetic_gene_cluster (42% of genes show similarity)
Cluster 18	Terpene	9-methylstreptimidone_biosynthetic_gene_cluster (9% of genes show similarity)
Cluster 19	Terpene	Hopene_biosynthetic_gene_cluster (46% of genes show similarity)
Cluster 20	Siderophore	-
Cluster 21	Bacteriocin	-

Table B-7: Genetic organization of putative funisamine BGC in *Streptosporangium* KDCAGE35.



GenBank accession #	Start/Stop (bp)	Protein	Size (AAs)	Proposed function	Protein Homolog	I / S (%)	Accession Number
	2-679	FunU1	225	unknown	Nucleoprotein TPR (*Xenopus laevis*)	38 / 44	Q5EE04.1
	1239-40	FunU2	399	hypothetical protein	Hypothetical protein (*Streptomyces sp. PCS3-D2*)	58 / 63	EYU65272.1
	250-1275	FunU3	343	hydrolase	Epoxide hydrolase B (*Mycobacterium tuberculosis CDC1551*)	39 / 51	P95276.2
	219-1817	FunU4	532	hypothetical protein	---	----	---
	2356-1277	FunU5	359	transporter	Uncharacterized transporter YcbK (*Bacillus subtilis subsp. subtilis str. 168*)	27 / 43	P42243.1
	1295-3595	FunU6	766	hypothetical protein	---	----	---
	2418-3335	FunU7	305	regulator	HTH-type transcriptional regulator GltC (Bacillus subtilis subsp. subtilis str. 168)	27 / 46	P20668.3
	3599-2832	FunU8	255	regulator	LysR family transcriptional regulator (*Nonomuraea jiangxiensis*)	41 / 52	WP_090940584.1
	4339-3608	FunU9	243	hypothetical protein	Hypothetical protein (*Ralstonia mannitolilytica*)	40 / 58	AJW47704.1
	3702-5567	FunU10	621	phosphatase	Alkaline phosphatase D (*Bacillus subtilis subsp. subtilis str. 168*)	45 / 60	P42251.3
	4817-3750	FunU11	355	hypothetical protein	---	----	---
	4823-7591	FunU12	922	galactosidase	alpha-galactosidase (Streptomyces alboflavus)	45 / 58	WP_087883000.1
	5675-4848	FunU13	275	unknown	hypothetical protein STAWA0001_0347 (*Staphylococcus warneri L37603*)	42 / 47	EEQ79474.1
	7555-5483	FunU14	690	hypothetical protein	Hypothetical protein SCLAV_2787 [*Streptomyces clavuligerus ATCC 27064*]	39 / 45	EFG07859.1
	5698-6303	FunU15	201	kinase	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (*Photorhabdus luminescens subsp. laumondii TTO1*)	32 / 50	Q7N589.1
	8036-8707	FunU16	223	monooxygenase	Pyrimidine monooxygenase RutA (*Bradyrhizobium sp. BTAi1*)	41 / 59	A5EB33.1
	10043-8868	FunU17	391	hypothetical protein	---	----	---

10707-9010	FunC1	565	tRNA ligase	Cysteine--tRNA ligase (*Salmonella enterica subsp. arizonae serovar*)	29 / 43	A9MLA7.1
9112-11118	FunC2	668	4'-phosphopantetheinyl transferase	4'-phosphopantetheinyl transferase AcpS (*Dechloromonas aromatica RCB*)	52 / 62	Q47EF3.1
9285-10616	FunJ1	443	hypothetical protein	---	----	---
12099 - 12629	FunR1	176	RNA polymerase sigma-F factor	RNA polymerase sigma-F factor (*Bacillus megaterium*)	54 / 75	P35145.1
14530-12644	FunC3	628	Phospholipase	Non-hemolytic phospholipase C (*Burkholderia pseudomallei K96243*)	37 / 48	Q9RGS8.2
15321-15016	FunR2	101	regulator	Outer membrane protein assembly factor BamA (*Helicobacter pylori 26695*)	43 / 60	O25369.1
16441-15506	FunC4	311	dehydrogenase	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating (*Dictyostelium discoideum*)	29 / 44	Q54L85.1
17220-16597	FunC5	207	reductase			
18172-17438	FunC6	244	oxidoreductase	oxidoreductase YkvO (*Bacillus subtilis subsp. subtilis str. 168*)	51 / 70	O31680.1
18436 - 19335	FunR3	299	regulator	Hca operon transcriptional activator HcaR (*Escherichia coli K-12*)	40 / 55	Q47141.2
21794-19701	FunR4	697	elongation factor G	Elongation factor G (*Streptomyces coelicolor A3(2)*)	81 / 87	O87844.1
23211-22279	FunB1	310	3-hydroxypentyl-CoA dehydrogenase	3-hydroxybutyryl-CoA dehydrogenase (*Mycobacterium tuberculosis CDC1551*)	56 / 72	P9WNP6.1
24197-23208	FunC7	329	ketoacyl-ACP synthase III	Acetoacetyl CoA synthase NphT7 (*Streptomyces sp. CL190*)	52 / 65	D7URV0.1
25546-24200	FunB2	448	2-pentenyl-CoA carboxylase/reductase	Crotonyl-CoA reductase (*Streptomyces collinus*)	52 / 65	Q53865.1
26373-25579	FunR5	264	regulator	Sugar fermentation stimulation protein homolog (*Pyrobaculum calidifontis JCM 11548*)	60 / 72	A3MSH9.1
27129-26458	FunC8	223	4'-phosphopantetheinyl transferase	4'-phosphopantetheinyl transferase Npt (*Nocardia iowensis*)	49 /62	A1YCA5.1
28178-27129	FunS1	349	sulfotransferase	sulfotransferase associated with clethramycin BGC (Streptomyces malaysiensis)	44 / 59	ATL82931.1
28945-28175	FunC9	256	Thioesterase	Thioesterase PikA5 (*Streptomyces venezuelae*)	52 / 66	Q9ZGI1.1
34875-28942	FunP8	1977	funisamine PKS 8	---	----	---
56330-34905	FunP7	7181	funisamine PKS 7	---	----	---
66246-56476	FunP6	3256	funisamine PKS 6	---	----	---
86928-66277	FunP5	6883	funisamine PKS 5	---	----	---
101612-86958	FunP4	4884	funisamine PKS 4	---	----	---
110910-101770	FunP3	3046	funisamine PKS 3	---	----	---
115796-111198	FunP2	1532	funisamine PKS 2	---	----	---

158396-116280	FunP1	14038	funisamine PKS 1	---	----	---
160153-158831	FunC10	440	FAD-dependent monooxygenase	FAD-dependent monooxygenase cctM (*Talaromyces islandicus*)	27 / 40	A0A0U1LQD9.1
160696-160202	FunR6	164	PadR family transcriptional regulator	Transcriptional regulator PadR-like family protein (*Nonomuraea jiangxiensis*)	89 / 96	SDK84512.1
162531-160984	FunM1	515	membrane protein	UPF0699 transmembrane protein YdbT (*Bacillus subtilis subsp. subtilis str. 168*)	16 / 39	A7RHG8.1
162953-162528	FunM2	141	membrane protein	UPF0699 transmembrane protein YdbS (*Bacillus subtilis subsp. subtilis str. 168*)	32 / 52	P96615.1
163265-164947	FunA1	560	arginine 2-monooxygenase	arginine 2-monooxygenase (*Streptomyces malaysiensis*)	67 / 75	ATL82933.1
165743-165120	FunR7	207	regulator	Response regulator protein VraR (*Staphylococcus aureus subsp. aureus MSSA476*)	44 / 65	Q6G850.1
166951-165740	FunR8	403	signal transduction histidine kinase	Sensor histidine kinase LiaS (*Bacillus subtilis subsp. subtilis str. 168*)	30 / 49	O32198.1
167778-166909	FunT1	289	transport permease	transport permease ycf38 (*Porphyra purpurea*)	27 / 47	P51321.1
168620-167775	FunT2	281	ABC transporter ATP-binding protein	ABC transporter ATP-binding protein YvfR (*Bacillus subtilis subsp. subtilis str. 168*)	38 / 56	O07016.1
168756-169583	FunJ2	275	hypothetical protein	hypothetical protein STAWA0001_0347 (*Staphylococcus warneri L37603*)	42 / 47	EEQ79474.1
172413-169660	FunR9	917	transcriptional regulator	HTH-type transcriptional regulator MalT (*Vibrio campbellii ATCC BAA-1116*)	25 / 41	A7N5N6.1
173031-174149	FunA3	372	agmatinase	agmatinase (*Streptomyces sp. Mg1*)	34 / 52	AKL64821.1
174191-175633	FunA2	480	4-guanidinobutanoate CoA ligase	4-guanidinobutanoate:CoA ligase (*Streptomyces malaysiensis*)	61 / 73	ATL87716.1
175630-176556	FunA4	308	4-guanidinobutanoyl-CoA:ACP acyltransferase	4-guanidinobutanoyl-CoA:ACP acyltransferase (*Streptomyces malaysiensis*)	48 / 65	ATL87717.1
177375-176761	FunR10	204	TetR family transcriptional regulator	Tetracycline repressor protein class E (*Escherichia coli*)	25 / 44	P21337.1
177445-178020	FunC11	191	alkyl hydroperoxide reductase	alkyl hydroperoxide reductase (*Streptomyces violaceusniger Tu 4113*)	70 / 79	AEM81606.1
178215-178826	FunC12	203	dehydrogenase	Malate dehydrogenase (*Bacillus cytotoxicus NVH 391-98*)	30 / 44	A7GQI9.1
178877-178080	FunJ3	265	hypothetical protein	---	----	---
178197-178826	FunC13	209	dehydrogenase	Malate dehydrogenase (*Geobacillus sp. WCH70*)	29 / 51	C5DAE0.1
179363-180136	FunR11	257	regulator	HTH-type transcriptional regulator CueR (*Escherichia coli O157:H7*)	33 / 55	Q8XD09.1
181146-180160	FunJ4	328	hypothetical protein	---	----	---
181118-180276	FunT3	280	transporter	ABC-2 family transporter protein (*Streptosporangium subroseum*)	94 / 96	SNT45499.1
182104-181115	FunT4	329	transporter	ABC transporter ATP-binding protein YhcH (*Bacillus subtilis subsp. subtilis str. 168*)	35 / 55	P54592.1
181118-182110	FunJ5	330	hypothetical protein	hypothetical protein STEPF1_06867 (*Streptomyces sp. F-1*)	39 / 49	SFY53584.1
181188-181880	FunJ6	230	hypothetical protein	hypothetical protein BJF79_17010 (*Actinomadura sp. CNU-125*)	57 / 65	OLT19263.1

183592-182798	FunC14	264	ketoreductase	Beta-ketoacyl-ACP reductase (*Vibrio cholerae O1 biovar El Tor str. N16961*)	36 / 52	Q9KQH7.2
183770-184417	FunR12	215	regulator	HTH-type transcriptional regulator YhgD (*Bacillus subtilis subsp. subtilis str. 168*)	38 / 59	P32398.1
184483-183782	FunD1	233	hypothetical protein	---	----	---
185811-184630	FunD2	393	methylase	Serine methylase (*Thermus thermophilus HB27*)	26 / 45	Q72IH2.2
184870-185931	FunD3	353	phosphoribosyltransferase	anthranilate phosphoribosyltransferase (*Streptomyces sp. M1013*)	29 / 47	WP_076976541.1
187265-186549	FunD4	238	peptide hydrolase	ATP-dependent Clp protease proteolytic subunit 1 (*Nocardia farcinica IFM 10152*)	73/ 86	Q5Z0M4.1
187822-187136	FunD5	228	peptide hydrolase	ATP-dependent Clp protease proteolytic subunit 3 (*Streptomyces coelicolor A3(2)*)	48 / 66	Q9X7R9.1
188692-188066	FunD6	208	hypothetical protein	---	----	---
188726-190639	FunD7	637	hypothetical protein	---	----	---
191317-188747	FunD8	856	Membrane protein	Outer membrane protein assembly factor BamB (*Vibrio fischeri ES114*)	31 / 45	Q5E769.1
189953-189270	FunD9	227	hypothetical protein	---	----	---
191234-190353	FunD10	293	Iron permease	putative iron permease FTR1 [*Mycobacterium abscessus 47J26*]	40 / 49	EHB98808.1

Figure B-47: Multiple sequence alignment of ketoreductase domains within the funisamine BGC to designate KR-type for stereochemistry predictions. The KR-types are distinguished in accordance with the five highlighted columns showing the 'fingerprint' amino acids. Group A1 is distinguished by no LDD in column 1, W in column 2, and no H in column 3. Group A2 is distinguished by no LDD in 1, W in 2, and H in 3. Group B1 is distinguished by LDD in 1 and no P in 5. Group B2 is distinguished by LDD in 1 and P in 5. Group C1 is distinguished by no Y in 4, and group C2 is distinguished by N in 6.

```
                                                1
Fun1e   CDLTDRSAVATLVREFDGEP---DLAIVHAAGVLDDGVIAGLDPARLDAVLASKAGAAWQ     112
Fun3a   CDLADRAAVEALLASV-----GTVDAVVHAAGVGANVPFDQTDVALVERLLAGKVAGAVN     114
Fun2a   CDLSDRSAVEALLADV-----GEVDAVVHAAGVSQDVPLAAEDADHLRAVAAGKVDGAAH     110
Fun1f   CDLADRSAVEALLDTV-----GDVDAVVHAAGIVEDVPLADADQAHLDRVIRGKVDGALH     110
Fun1c   CDLADRSAVEALLATI-----GQVNAVVHAAGVGEDVALVDADEEHLRRVVGGKVDGALH     110
Fun1g   CDLADRSAVEGLLATI-----GQVNAVVHAAGVSEHAALTDVDEEHLRRVVVGKVDGALH     110
Fun1d   CDLADPSAVEGLLATI-----GQVDAVVHAAGVAEDAELVDADAAHLNRVLSGKVDGALY     110
Fun1h   CDLADRSAVEGLLATI-----GQVNAVVHAAGVAEDVELVDADAAHLNRVLSGKVDGALH     110
Fun1a   CDLADRSAVEALLAVV-----GAVDAVVHAAGVGEDAELVEADAAHLNRVLSGKVDGASH     110
Fun12   CDLADRSAVEALLAVV-----GAVDAVVHAAGVAADVPLRDADEAHFRTVLSGKVDGALH     110
Fun6b   CDTADRAQVAALLAGL----PEPVTAVVHAAGTLTPVLLADSTPEELADVRSGKVEGAVH     115
Fun3b   CDLADRDQVAALVADL----PADLTAVVHAAGVAQDTPIADLTTAETAAVTGARVTGTLL     115
Fun4a   TDLADRDAVAALLKEATADPEAPLTAVVHAAGIAHSAPLADLDAAGLASVLAGKTTGALH     119
Fun1i   ADVADRAALASVLAEIPA--EHPLTAVVHTAGVLADGIVERMTPDQLDRVMRPKVDGALH     114
Fun7a   CDVADRDDLDRVLD------GVDVRAVVHVAGVLDDTVLTGLTPDRLDAVLRAKVDAVVN     114
Fun5d   CDLADPAAVQRLIG------PVEVGAVLHAAGSTDDAMLTSLTPDRLASVLAAKVDAAVN     114
Fun6a   CDVTDPEAVEAALR------GVTVSAVFHTAGVLDDGLLADLTPDRLDAVLRPKADAVWN     104
Fun7b   CDLADAGAVAGALR------DEPVTAVIHAAGVLDDALLTDLTPERLRAVFRAKVDAAVN     110
Fun4b   CDLSDAGAVMAALR------DEPVSAVVHAAGVIDDGLLTDLTPERLDTVFRAKVDAARN     104
Fun4c   CDVSDADALTAALR------DEPVTAVIHAAGVLDDGTLESLTPERLDAVFRAKVDAARN     108
Fun8a   CDVADAAALTAALR------DEPVTAVIHVAGVLDDGLLTDLTPARLDTVFRAKVDAARN     108
Fun5c   CDVADAGAVTEALR------GESVSAVIHAAGVLDDGMIESLTPERLDTVFRAKIDAVRA     104
Fun7c   CDVADAGAVAEALR------GESVSAVVHAAGVLDDALLADLTPERLDTVFRAKVDAARN     104
Fun7d   CDVADAGAVAEALR------GEPVSAVIHAAGVLDDALLADLTPERLDTVFRAKIDAARN     108
Fun5a   CDVADAGAVAAALR------DEPVTAVIHAAGVLDDGTLESLTPERLDTVFRAKVDAARN     108
Fun5b   CDVSDAVAVTAALR------DESVSAVIHAAGVLDDALLADLTPERLRTVFRAKVDAARN     108


                                2    3   4 5 6
Fun1e   LHELTEHRPLSAFVLFSSTAGVFGNPGQANYAAANAALDALAEYRKVLGLPATSIAWGPW     172
Fun3a   LDALVGD--VDAFVTFSSLSGVWGSQSHAAYAVANAALDALAEQRRARGGAMTAIAWGSW     172
Fun2a   LDALLPD--VP-LIVFSSIAGVWGSAEQAAYAAANAALDALIARRRARGRPGTAVAWGPW     167
Fun1f   LDALVGD--VDAFVVFSSISATWGSGRQAAYGAANTALDGLILRRRAAGLPGTSIAWGPW     168
Fun1c   LDALVGD--VDAFVVFSSISGIWGSAEQTAYGAANAALDALIARRRASGLPGTAVAWGPW     168
Fun1g   LDALVGD--VDAFVVFSSISGIWGSAEQAAYGAANAALDALIARRRASGLPGTAVAWGPW     168
Fun1d   LDALVGD--VDAFVVFSSISGIWGSAEQAAYGAANAALDALIARRRASGLPGTAVAWGPW     168
Fun1h   LDALVGD--VDAFVVFSSISGIWGSAEQTAYGAANAALDALIARRRASGLPGTAVAWGPW     168
Fun1a   LDALVGD--VDAFVVFSSISGVWGSRGQAAYGAANAALDALVERRRAAGRPGTAIAWGPW     168
Fun12   LDALVGD--VDAFVVFSSISGVWGSGEQAAYGAANAALDGLVARRRAHGLPGTAVAWGPW     168
Fun6b   LLDLLDPAHLEQVVLFSSNAGVWGSARQGTYGAANAALDALAEQARERGLPVTSVAWGLW     175
Fun3b   LDELLADVELDAFVVFSSVAGVWGTARHPAYAAADAFLDAFAGWRRAQGRPATAIAWSPW     175
Fun4a   LDELLGDTDLDAFVLFSSIAATWGSGWGGAYAAANAGLDALAQRRRARGLAGTSLAWGPW     179
Fun1i   LHELTRDLDLSAFVLFSSASAIFGTPGQANYAAANAFLDALAQHRRALGLPGQALAWGPW     174
Fun7a   LHEATAGADLDAFVLYSSVAGLFGTPGQGNYAAANAFLDAFAAARRSRGLPGTSLAWGAW     174
Fun5d   LRAATADRPLSAFVLFSSVAGLLGSAGQANYAAANTFLDAYAARLRAEGVPATSLAWGLW     174
Fun6a   LHAATADRPLAAFVLFSSAAGLLGNAGQANYAAANTFTDAFAAFRRAQGLPATSLAWGLW     164
Fun7b   LRAATADHRLSAFVLYSSAAGLFGNAGQANYAAANAFLDAYAAQLRAEGVPATSLAWGLW     170
Fun4b   LAAATEDRPLSAFVLYSSASGLFGSAGQANYAAANAFLDAYATQLRAQGVPATSLAWGLW     164
Fun4c   LAAATKDRPLRAFVLYSSASGLFGNAGQANYAAANTFLDAYATRLRGEGVPATSLAWGLW     168
Fun8a   LAAATKDRPLRAFVLYSSVAGIFGNPGQANYAAANAFLDAYATQLRAQGVPATSLAWGLW     168
Fun5c   LRAATADQPLTAFVLYSSVAGLFGNAGQANYAAANAFLDAYATQLRGQGVPATSLAWGLW     164
Fun7c   LAAATADQPLTAFVLYSSAAGVFGNAGQANYAAANAFLDAYATQLRGQGVPATSLAWGLW     164
```

```
Fun7d    LAAATADQPLAAFVLYSSAAGLFGNAGQANYAAANAFLDAYATQLREQGVPATSLAWGLW    168
Fun5a    LVAATKDQPLTAFVLYSSAAGVFGNAGQANYAAANAFLDAYATQLRGQGVPATSLAWGLW    168
Fun5b    LVAATKDQPLTAFVLYSSAAGLFGNAGQANYAAANAFLDAYATQLHAQGIPATSLAWGLW    168
```

Figure B-48**:** Predicted stereochemical configurations of products for each polyketide synthase module in funisamine biosynthesis.

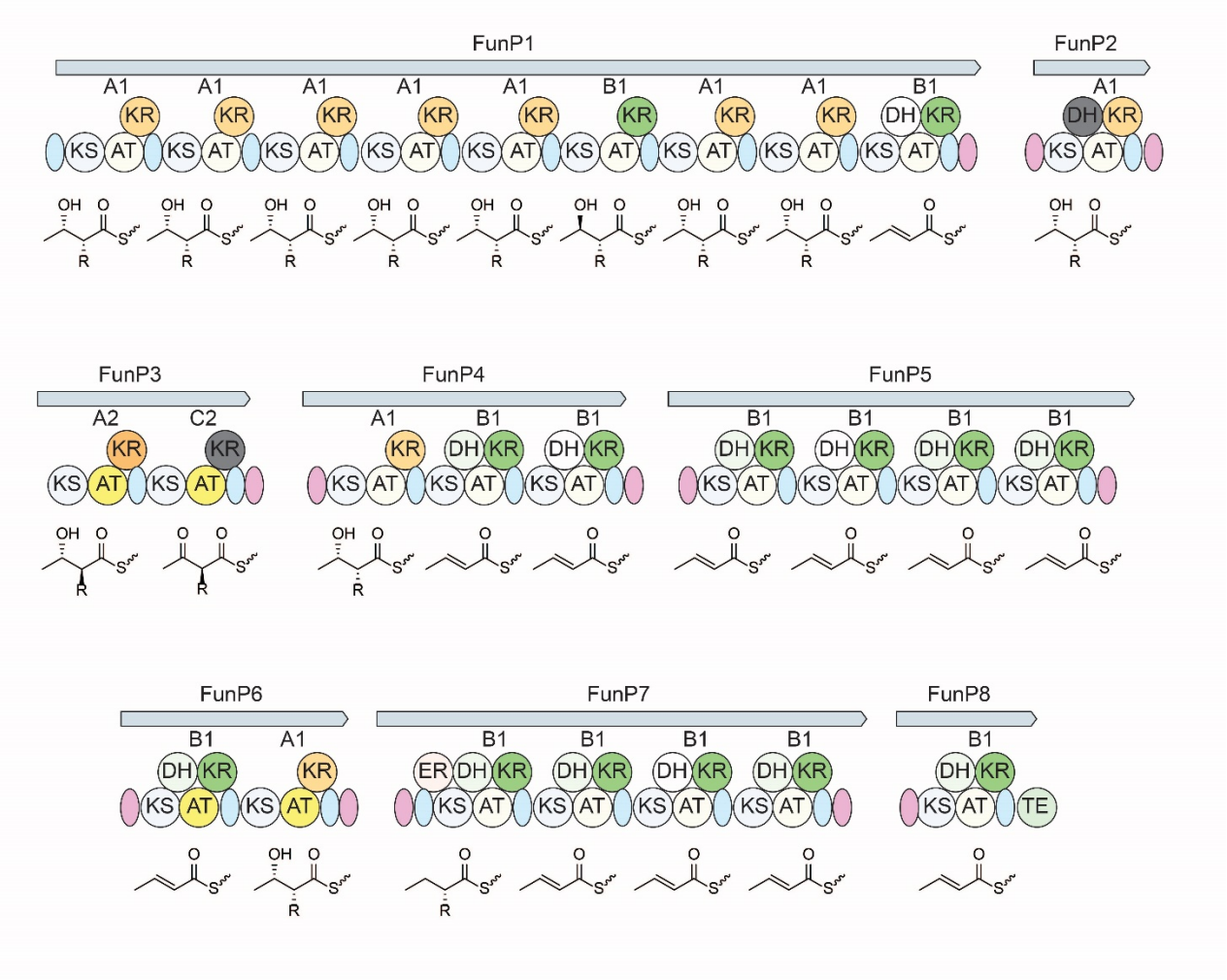


Figure B-49**:** Expanded biosynthesis of funisamine from FunP1. The layout of catalytic domains: ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) present within the polyketide synthases are shown with inactive domains colored grey. A-type and B-type KR domains are colored orange and green respectively. Acyltransferase domains predicted to use methyl or propylmalonyl extender units are colored yellow. Acyl carrier proteins are colored light blue, and docking domains are purple.

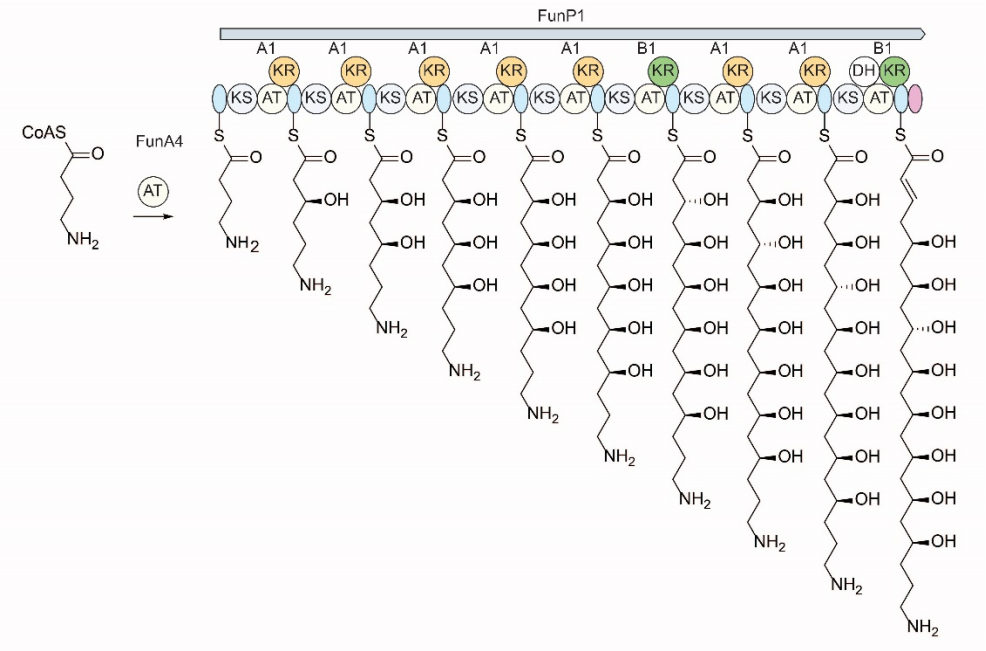


Figure B-50**:** Expanded biosynthesis of funisamine from FunP2 - FunP4. The layout of catalytic domains: ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) present within the polyketide synthases are shown with inactive domains colored grey. A-type and B-type KR domains are colored orange and green respectively. Acyltransferase domains predicted to use methyl or propylmalonyl extender units are colored yellow. Acyl carrier proteins are colored light blue, and docking domains are purple.

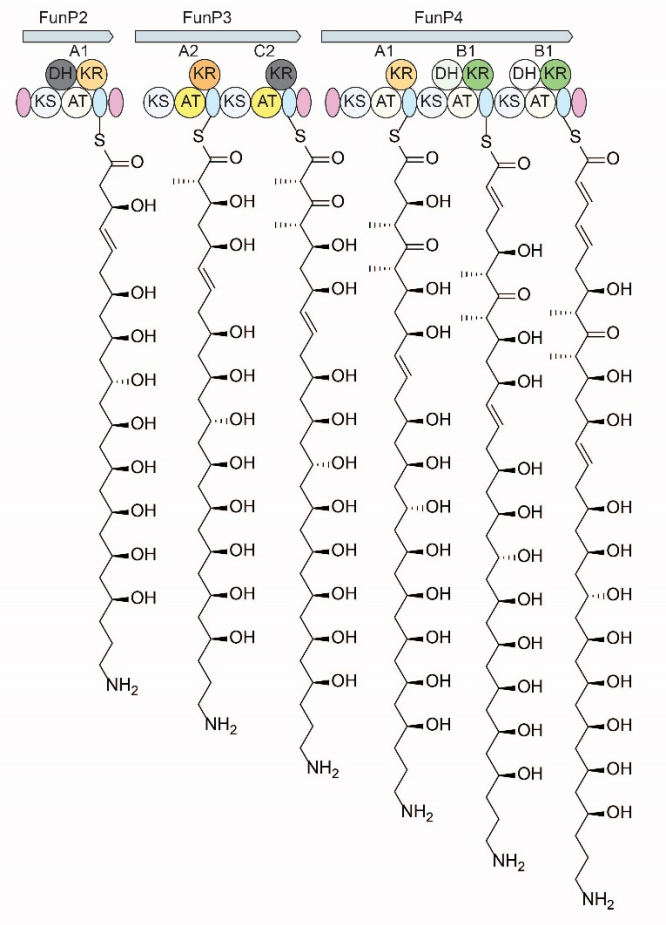
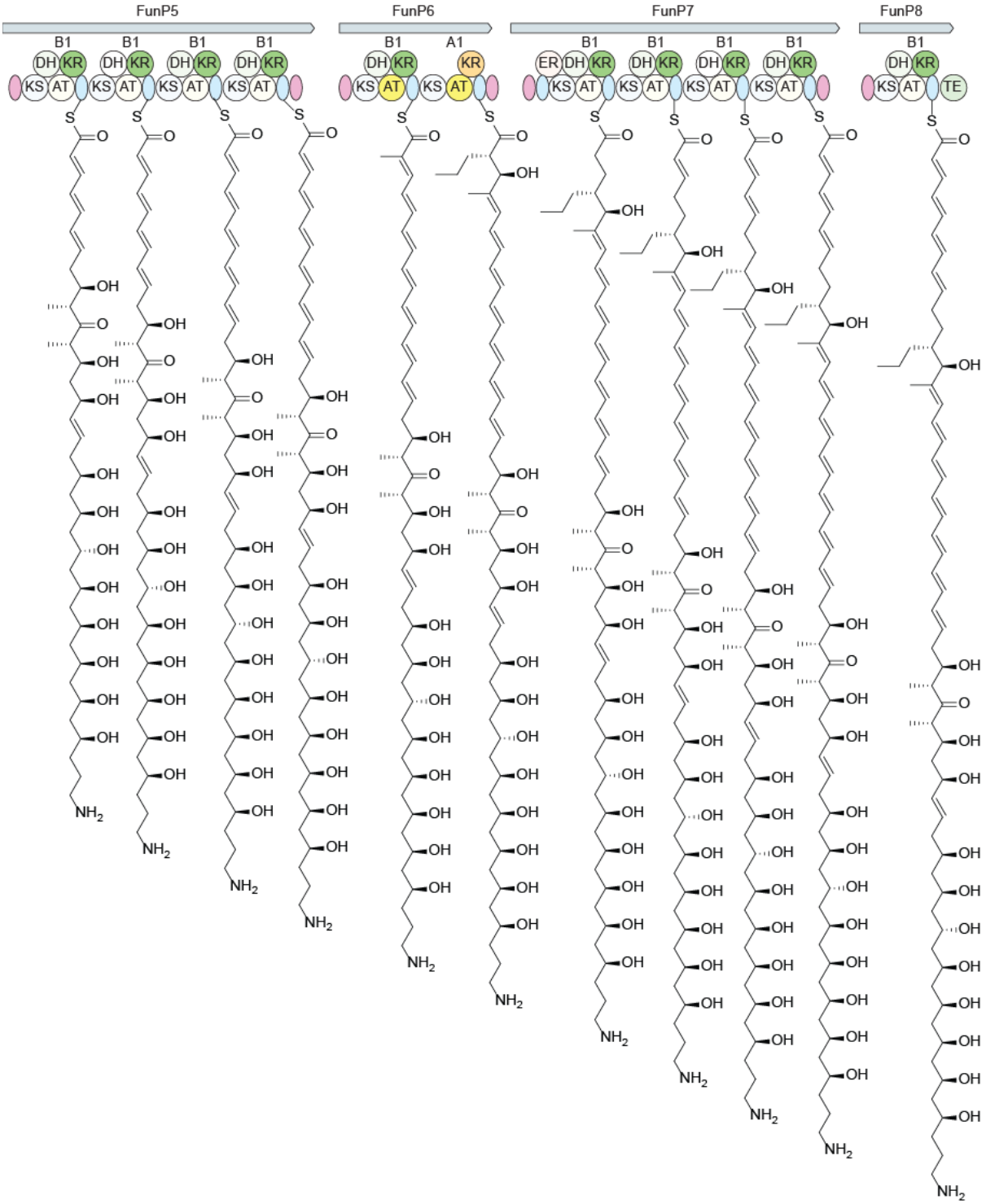


Figure B-51**:** Expanded biosynthesis of funisamine from FunP5 - FunP8. The layout of catalytic domains: ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) present within the polyketide synthases are shown with inactive domains colored grey. A-type and B-type KR domains are colored orange and green respectively. Acyltransferase domains predicted to use methyl or propylmalonyl extender units are colored yellow. Acyl carrier proteins are colored light blue, and docking domains are purple.



Appendix C Curriculum vitae

Brett Casey Covington
1407 S. Callywood ct. | Clarksville, TN 37040
931-561-4534 | brett.c.covington@vanderbilt.edu

EDUCATION

Ph.D. Chemistry — Expected Aug 2018
Advisor: Brian Bachmann, PhD
Vanderbilt University, Nashville, TN

B.S. Chemistry, *cum laude* — May 2012
Minor: Biology
Austin Peay State University, Clarksville, TN

PUBLICATIONS

Brett C. Covington, Jeffery M. Spraggins, Audrey E. Ynigez-Gutierrez, Zachary B. Hylton, Brian O. Bachmann (**2018**) Response of hypogean actinobacterial genera secondary metabolism to chemical and biological stimuli. *Applied & Environmental Microb*iology (in review, manuscript available upon request).

Brett C. Covington, John A. McLean, Brian O. Bachmann (**2017**) Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Natural Product Reports* 34 (1), 6-24. doi: 10.1039/C6NP00048G (Artwork and article featured on journal front cover)

Ma. Diarey Tianero, Elizabeth Pierce, Shrinivasan Raghuraman, Debosmita Sardar, John A. McIntosh, John R. Heemstra, Zachary Schonrock, **Brett C. Covington**, J. Alan Maschek, James E. Cox, Brian O. Bachmann, Baldomero M. Olivera, Duane E. Ruffner, Eric W. Schmidt (**2016**) Metabolic model for diversity-generating biosynthesis. *Proceedings of the National Academy of Science*, 113(7), 1772-7. doi: 10.1073/pnas.1525438113

Dagmara K Derewacz, **Brett C. Covington**, John A McLean, Brian O Bachmann (**2015**) Mapping microbial response metabolomes for induced natural product discovery. *ACS Chemical Biology* 10(9), 1998-06. doi: 10.1021/acschembio.5b00001

Cody R Goodwin, **Brett C. Covington**, Dagmara K Derewacz, C Ruth McNees, John P Wikswo, John A McLean, Brian O Bachmann (**2015**) Structuring microbial metabolic responses to multiplexed stimuli via self-organizing metabolomics maps. *Chemistry & Biology*, 22(5), 661-70. doi: 10.1016/j.chembiol.2015.03.020

FELLOWSHIPS & PROPOSALS

Unlocking nature's chemical arsenal for antibiotic discovery — 2015
NIH F31 predoctoral fellowship proposal (FOA: PA14-147)
Impact score: 47
Percentile: 37
Not funded (reviewer comments available upon request)

Chemistry Biology Interface Training Program — 2013-15
NIH NRSA predoctoral fellowship (FOA: PAR-17-341)

RESEARCH EXPERIENCE

Dissertation Research — 2012 - present
Principal Investigator: Brian O. Bachmann, PhD

- Cultured and identified actinomycetes from hypogean (cave) sediments
- Developed metabolomic methods to prioritize natural products from extracts of stimulated microbial cultures
- Purified and structurally characterized secondary metabolites from small scale microbial cultures
- Analyzed microbial genome sequences for genes involved in secondary metabolite biosynthesis
- Performed routine maintenance on mass spectrometer (Thermo, TSQ Quantum Access MAX) and high-performance liquid chromatography devices.
- Techniques include: Microbial (mixed) culture, liquid chromatography systems, mass spectrometry, nuclear magnetic resonance spectroscopy, metabolomics analyses, natural product isolation/discovery, and genome mining.

Undergraduate Research — 2010-12
Principal Investigator: Meagan K. Mann, PhD

- Purified components of lavender oils and quantified their affinity to bind to the estrogen receptor through fluorescence polarization and radiometric binding assays.
- Tested fractionated lavender extract in proliferation assays with estrogen-responsive MCF-7 cell lines.

TEACHING EXPERIENCE

Vanderbilt University, *Organic Chemistry Recitation TA* — Spring 2018

- Presented brief lectures reviewing 2nd semester organic chemistry coursework and answered student questions on lecture topics

Vanderbilt University, *Organic Chemistry Recitation TA* — Fall 2017

- Presented brief lectures reviewing 1st semester organic chemistry coursework and answered student questions on lecture topics

Vanderbilt University, *Foundations of Chemical Biology* (graduate student course) — Fall 2015

- Presented lectures on topics of chemical biology
- Developed student semester projects
- Assisted instructor in grading coursework

Vanderbilt University, *Organic Chemistry lab TA* — Spring 2013

- Presented brief lectures on organic theory relevant to laboratory activities
- Assigned and graded coursework

Vanderbilt University, *Organic Chemistry lab TA* — Fall 2012

- Presented brief lectures on organic theory relevant to laboratory activities
- Assigned and graded coursework

COMMUNITY OUTREACH

Middle TN Science and Engineering Fair (MTSEF) Judge; Science, Technology, Engineering, and Mathematics (STEM) Expo Judge	2011-14

HONORS & AWARDS

Best poster award, VICB annual research symposium	Aug 2014
VICB certificate in chemical biology, Vanderbilt University	Aug 2014
Biology departmental award, Austin Peay State University	Apr 2012
German departmental award, Austin Peay State University	Apr 2012
Biochemistry departmental award, Austin Peay State University	Apr 2011
TN HOPE Scholarship Award	2007-12

PRESENTATIONS

Oral

Mapping microbial natural product responses to activation stimuli. — Mar 2016
Marine Natural Products, Gordon Research Seminar
Ventura, CA

Unlocking nature's chemical arsenal — Mar 2015
Vanderbilt 3 Minute Thesis competition (finalist)
Nashville, TN

Metabolomic methods to unlock nature's chemical arsenal. — May 2014
Vanderbilt Chemical Biology Association of Students (CBAS) Seminar
Nashville, TN

Unlocking nature's chemical arsenal — Sep 2014
Vanderbilt Chemistry Departmental Seminar
Nashville, TN

Poster

Unlocking hidden reservoirs of microbial natural products — Jan 2018
Natural Product Discovery and Development in the Genomic Era, Society for Industrial Microbiology & Biotechnology (SIMB) Conference
Clearwater beach, FL

Metabolomic methods to unlock nature's chemical arsenal. — Aug 2017
Vanderbilt Institute of Chemical Biology (VICB) Student Research Symposium
Nashville, TN

Mapping microbial natural product responses to activation stimuli — Jun 2016
Marine Natural Products, Gordon Research Conference
Andover, NH

Mapping microbial natural product responses to activation stimuli — Mar 2016
Marine Natural Products, Gordon Research Conference
Ventura, CA

PRESENTATIONS (continued)

Poster

Mapping microbial natural product responses to activation stimuli — Mar 2016
Marine Natural Products, Gordon Research Seminar
Ventura, CA

Unlocking Nature's Chemical Arsenal — Sep 2014
VICB Student Research Symposium (Won best poster award)
Nashville, TN

Unlocking Nature's Chemical Arsenal — Jun 2014
37th Steenbock Symposium
Madison, WI

Unlocking Nature's Chemical Arsenal — Jun 2014
Chemistry-Biology Interface career development conference
Madison, WI

Efforts towards understanding the estrogenicity of lavender oil — Apr 2012
7th Annual APSU Research Forum
Clarksville, TN

Efforts towards understanding the estrogenicity of lavender oil — Apr 2011
6th Annual APSU Research Forum
Clarksville, TN

Insights into the estrogenicity of lavender oil: A review of the literature and implications for further research — Aug 2010
240th ACS National Convention
Boston, MA

Insights into the estrogenicity of lavender oil: A review of the literature and implications for further research — Apr 2010
5th Annual APSU Research Forum.
Clarksville, TN

REFERENCES

Thesis Advisor
Brian Bachmann, Ph.D.
Professor of Chemistry
Vanderbilt University
(615) 322-8865
brian.bachmann@vanderbilt.edu

Undergraduate Research Advisor
Meagan K. Mann, Ph.D.
Associate Professor of Chemistry
Austin Peay State University
931-221-6175
mannm@apsu.edu

Dissertation Committee Member
John A. McLean, Ph.D.
Stevenson Professor of **Chemistry**
Vanderbilt University
(615) 322-1195
john.a.mclean@vanderbilt.edu