Integrated Chemical and Biological Profiling for the Discovery of Antibiotics Against Resistant Pathogens

by

Hannah Cavanagh

B.Sc., Carleton University, 2021

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Chemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2024

Declaration of Committee

Name:	**Hannah Cavanagh**
Degree:	**Master of Science (Chemistry)**
Title:	**Integrated Chemical and Biological Profiling for the Discovery of Antibiotics Against Resistant Pathogens**
Committee:	**Chair:** **Neil Branda** Professor, Chemistry
	Roger Linington Supervisor Professor, Chemistry
	Erika Plettner Committee Member Professor, Chemistry
	Robert Britton Committee Member Professor, Chemistry
	Bingyun Sun Examiner Associate Professor, Chemistry

Abstract

Natural products (NPs) are prolific sources of therapeutics, particularly antibiotics. Numerous NP antibiotics were discovered from 1940-1960 however, the discovery rate declined as known compounds were repeatedly discovered. This led to a decrease in this research and contributed to the antimicrobial resistance (AMR) crisis. This project uses computational platforms to discover NP antibiotics to combat the AMR crisis without constant rediscovery. A profiling platform, ResistoMAP, screened a NP library and antimicrobial standards against a panel of 29 drug-resistant strains of *Escherichia coli*. A software, NP Analyst, combined these profiles with mass spectrometry (MS) to identify bioactive metabolites. Focusing on unique MS and ResistoMAP profiles, three candidates were selected for characterization. Two were identified as streptorubin B and undecylprodigiosin, while the third demonstrates activity and is being studied further. This suggests that the technique outlined throughout this project is effective in identifying compounds effective against resistant pathogens while minimizing rediscovery.

Keywords: Natural Products; Antimicrobial Resistance; Rediscovery; ResistoMAP; NP Analyst

Dedication

To my parents,

Your endless love, sacrifices, and belief in my dreams have been my guiding light.
Thank you for everything.

Acknowledgements

I would like to express my deepest appreciation to my thesis supervisor, Roger Linington, whose passion for the field is not only inspiring but also instrumental in my journey to complete this thesis. His guidance, expertise, and support have been invaluable in shaping my understanding and approach throughout this research.

I extend my sincere appreciation to Erika Plettner and Rob Britton, members of my thesis committee, for their insightful feedback and contributions that have strengthened this work.

To the past and present members of the Linington Lab, your help and camaraderie have been a true blessing. Special thanks to Dr. Trevor Clark for his unwavering aid and encouragement at every step of this thesis. In moments of doubt, his positivity and support were a constant source of strength. Dr. Michael Recchia, my thesis buddy, our shared journey these last few months has made this process more enjoyable. Our Google Map adventures will always be cherished memories. Thank you both for letting me spend valuable time with your doggos, Nala and Carlie. I owe a debt of gratitude to Dr. Dennis Liu for his foundational work on ResistoMAP, his mentorship, and his patience in expanding my knowledge of this field. Dr. Tim Baumeister, your enthusiasm for coding and guidance on MVPA were crucial for this project. Thank you for always being up for a good meal and chat. Liana Zaroubi your ongoing support and advice has been a true source of comfort and love away from home. Julia Saulog and Emily McMann, your laughter and good times brought light to my days. Thank you all for your contributions, support, and friendship throughout this experience.

I owe immense gratitude to my friends and family, whose unwavering love and support have made this achievement possible. To Saeid, your constant belief in me and steadfast presence have been my rock during this time. To my forever friend Taz, I am incredibly grateful for your support and ability to listen and empathize. Meeting you during this journey has been a blessing I cherish deeply. Lastly, to my family, thank you. Though you may not understand this thesis, your endless encouragement and bright spirits have illuminated my days and fueled my determination.

Table of Contents

Declaration of Committee ii
Abstract iii
Dedication iv
Acknowledgements v
Table of Contents vi
List of Tables viii
List of Figures ix
List of Acronyms xi

Chapter 1. Introduction 1
1.1. Natural Product Applications 1
1.1.1. Origins 1
1.1.2. Classifications 2
1.1.3. Marine Natural Products (MNP) 3
1.2. Natural Product Drug Discovery 4
1.2.1. The Decline in Antibiotic Discovery 4
1.3. Antimicrobial Drug Resistance 5
1.4. Techniques for Antimicrobial Drug Discovery 7
1.4.1. Computational Methodologies 7
1.4.2. Bioassay Screening Methodologies 9
1.5. Summary and Proposal 10

Chapter 2. ResistoMAP Data Acquisition and Analysis 12
2.1. Introduction 12
2.1.1. ResistoMAP 12
2.1.2. NP Analyst 13
2.1.3. Combined Applications of ResistoMAP and NP Analyst 16
2.1.4. Selectivity Ratio Profile Network (SRPNT) and Multivariate Pattern Analysis (MVPA) 17
2.2. Results 18
2.2.1. ResistoMAP Selections 18
2.2.2. NP Analyst Network Analysis 18
2.2.3. Secondary ResistoMAP 21
2.2.4. Secondary NP Analyst and SRPNT 23
Candidate Selection Using SRPNT 23
Candidate Selection Using NP Analyst Network 24
2.3. Discussion 27
2.4. Materials and Methods 34
2.4.1. ResistoMAP 34
2.4.2. Mass Spectrometry-Based Metabolomics Data Acquisition 34
2.4.3. NP Analyst Network Production and Community Visualization 35
Communities Category 1 – Only Antimicrobial Standards 36

Communities Category 2 – NP Prefractions and Antimicrobial Standards 36
Communities Category 3 – NP Prefractions.....37
2.4.4. Fermentation Process for the 18 Extracts.....37
2.4.5. Peak Library Separation for 15 Extracts 38
2.4.6. Secondary ResistoMAP Screening and Data Analysis 38
2.4.7. Secondary NP Analyst Network 39
SRPNT/MVPA Visualization 40
Community 1 – RLUS-2085E-3 41
Community 2 – RLUS-2028A-2 41
Community 3 – RLUS-2204D-3 42

Chapter 3. Isolation and Structure Elucidation 43
3.1. Introduction.....43
3.2. Results 45
3.2.1. Isolation of features 392.27 *m/z* and 394.30 *m/z* from RLUS-2085E 45
3.2.2. Structure Elucidation of Streptorubin B (2) 46
3.2.3. Structure Elucidation of Undecylprodigiosin (4) 49
3.2.4. Isolation of Candidate 552.40 *m/z* (X3) from RLUS-2204D 52
3.2.5. Structure Comparison for X3 53
3.2.6. ResistoMAP Screening for Isolated Candidates 54
3.3. Discussion 55
3.4. Materials and Methods 59
3.4.1. Fermentation, Isolation, and Discovery of RLUS-2085E Prodiginines 59
Fermentation and Subfraction Separation..... 59
Bioactive Analog Isolation 60
3.4.2. Fermentation, Isolation, and Analysis of RLUS-2028 61
3.4.3. Purification, Isolation, and Discovery of RLUS-2204D 62
Bioactive Analog Isolation 63
3.4.4. Isolated Candidates ResistoMAP 63
Minimum Inhibitory Concentration (MIC)..... 63

Chapter 4. Conclusion and Outlook.....64

References.....66

Appendix A. Supplemental Figures and Tables 73

Appendix B. Supplemental Mass Spectrometry and NMR Spectra 76

List of Tables

Table 1.1. Natural Product Antibiotics Discovered from Genus *Streptomyces*.....3

Table 2.1. List of extracts grown up through fermentation processes for novel mass candidates for further research.....21

Table 2.2. Masses identified from secondary NP Analyst network from RLUS-2204. 27

Table 2.3. List of extracts that underwent prefraction separation into subfractions..38

Table 3.1. Tabulated NMR data for 2 isolated from RLUS-2085E. All spectra were acquired in chloroform-d4 at 600 MHz (^{1}H), 150 MHz (^{13}C), 60.80 MHz (^{15}N). 46

Table 3.2. Tabulated NMR data for 4 isolated from RLUS-2085E. All spectra were acquired in chloroform-d at 600 MHz (^{1}H), 150 MHz (^{13}C), 60.80 MHz (^{15}N). *Peak values were estimated from spectra provided by Papireddy *et al.* 50

Table 3.3. Customized gradients for the separation of RLUS-2085 prefractions into subfraction peak libraries. 60

Table 3.4. Customized gradients for the separation of RLUS-2028 prefractions into subfraction peak libraries. 61

Table 3.5. Customized gradients for the separation of RLUS-2204 prefractions into subfraction peak libraries. 62

List of Figures

Figure 1.1. Distribution of sources of 161 antibacterial drugs introduced from January 1981 to September 2019. 2

Figure 1.2. Timeline of antibiotic class discovery on the top and events associated with AMR on the bottom. 7

Figure 2.1. ResistoMAP screening showing hierarchical clustering of prefractions RLUS-1505C and RLUS-1530C clustering with rifamycin-like standards between 8 to 128μM.[46] 13

Figure 2.2. NP Analyst default output network showing clusters of NP extracts sharing similar bioactive MS features. Each square node represents a Sample while the circular nodes correspond to MS features. 14

Figure 2.3. NP Analyst community viewing including network view (A), plot of retention time vs. precursor *m/z* for bioactive MS features (B), plot of Cluster Score vs Activity Score for bioactive MS features (C), and activity phenotypic fingerprints for extracts within the community (D). Rows are the NP extracts and columns contain the bioassay values. 15

Figure 2.4. Full NP Analyst output consisting of the 384 selected NP extracts and 68 commercial antimicrobials from the ResistoMAP dataset. 19

Figure 2.5. Community 4 used for dereplication as it contains the NP prefractions and commercial antimicrobials labeled. Specifically, RLUS-1878 prefractions are directly connected to Spiramycin. 20

Figure 2.6. Community 10 (A) and 27 (B) belong to the third category of communities. Each prefraction label corresponds to the closest square node. 20

Figure 2.7. Peak Library ResistoMAP profiles against 9 drug-resistant strains of *E. coli* and wildtype MG1655 for extracts RLUS-2024, RLUS-2085, RLUS-2028, and RLUS-2204 22

Figure 2.8. Secondary ResistoMAP-NP Analyst network made using the extract subfraction ResistoMAP output (A). Secondary NP Analyst network coloured by community (B). Legend outlines what extracts are found in which community. * Extract only has one node found in the community. 23

Figure 2.9. SRPNT Cytoscape output of 392.27 *m/z* showing the highest selectivity ratios against each drug-resistant and wildtype strain of *E. coli*. This mass was most prominent in RLUS-2085E-3 subfraction. 24

Figure 2.10. Candidate mass feature 392.27 *m/z* for isolation from RLUS-2085E-3. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2085E-3 (C). 25

Figure 2.11. Candidate mass feature 394.30 *m/z* for isolation from RLUS-2085E-6. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2085E-6 (C). 25

Figure 2.12. Candidate mass feature 438.16 *m/z* for isolation from RLUS-2028A-2. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2028A-2 (C). 26

Figure 2.13. Candidate mass feature 552.40 *m/z* for isolation from RLUS-2204D-3. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2204D-3 (C). 27

Figure 2.14. RLUS-2028 ResistoMAP phenotypic profile (A). Secondary ResistoMAP screening on peak library subfractions (B).....32
Figure 2.15. ResistoMAP output for RLUS-2204.....33
Figure 3.1. Structures of natural prodiginine compounds.....44
Figure 3.2. Key 2D NMR correlations of 2 in $CDCl_3$ at 600 MHz for ^{1}H and ^{13}C. COSY (one-headed arrows), HMBC (double-headed arrows), and N^{15} HMBC (blue double-headed arrows) correlations are shown.....49
Figure 3.3. Key 2D NMR correlations of 4 in $CDCl_3$ at 600 MHz for ^{1}H and ^{13}C. COSY (one-headed arrows), HMBC (double-headed arrows), and N^{15} HMBC (blue double-headed arrows) correlations are shown.....52
Figure 3.4. Suggested structure of candidate X3 from comparison with Borisova *et al*.[65].....54
Figure 3.5. ResistoMAP profiles displaying percent growth of Isolated Candidates. Labeling is as follows Candidate_Concentration(µg/mL). If there was 100% growth then no activity was observed.....54
Figure 3.6. Protonation of prodiginines that can occur in an acidic environment.....56

List of Acronyms

ARP	Antibiotic Resistance Platform
AMR	Antimicrobial Resistance
BGC	Biosynthetic Gene Cluster
DCM	Dichloromethane
DMSO	Dimethyl Sulfoxide
ESI	Electrospray Ionization
FA	Formic Acid
HRMS	High Resolution Mass Spectrometry
HTS	High Throughput Screening
H/UPLC	High/Ultra Performance Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LB	Luria-Bertani
MB	Marine Broth
MNP	Marine Natural Product
MS	Mass Spectrometry
MIC	Minimum Inhibitory Concentration
MVPA	Multivariate Pattern Analysis
NP	Natural Products
NMR	Nuclear Magnetic Resonance
PLS	Partial Least Square
qTOF	Quadrupole Time of Flight
ResistoMAP	Resistance Mode of Action Profiling
Rt	Retention Time
RT	Room Temperature
SR	Selectivity Ratio
SRPNT	Selectivity Ratio Profile Networks
SFU	Simon Fraser University
SYP	Starch-Yeast-Peptone
TWIM	Traveling Wave Ion Mobility

WHO	World Health Organization

Chapter 1.

Introduction

1.1. Natural Product Applications

1.1.1. Origins

The chemical diversity of nature provides a vast array of natural products (NPs). Throughout history, NPs have been utilized as traditional treatments for ailments due to their biological properties.[1] These historic applications instigated the scientific interest of NPs for their medicinal applications. NPs are known to have complex structures that cover a large range of molecular weights, sometimes with multiple chiral centers, and many H-bond acceptors and donors that translate into unique biological activities for these compounds.[2–4] Biological activity suggests a potential for a compounds medical applications. Only 10% of the world's biodiversity has been explored for its biological properties leaving a promising pool of structurally diverse NPs that could be applied to human therapeutics as potential drug treatments.[2,3,5] 61% of anticancer compounds and 29% of anti-infectives, accumulated over the past 30 years, are from NPs or have been chemically derived from them.[6] **Figure 1.1** is a representation of the antibacterial drugs produced from different sources from January 1981 to September 2019 reproduced from Newman and Cragg. Of the 161 drugs, over 50% were identified or derived from NPs proving their impact on antibacterial treatments.[4] NPs continue to remain an integral source and inspiration for potential drug candidates.

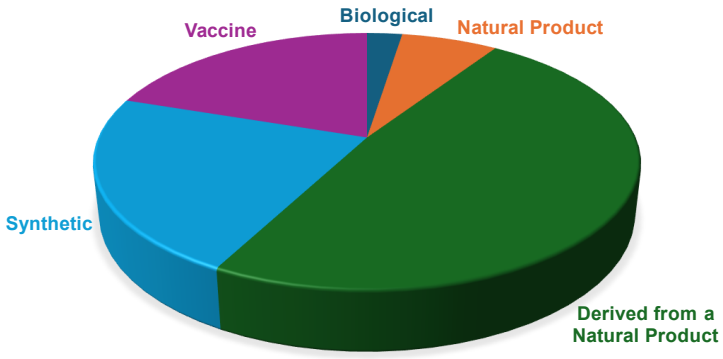


Figure 1.1. Distribution of sources of 161 antibacterial drugs introduced from January 1981 to September 2019.

This figure is an unofficial adaptation from "Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019" by David J. Newman and Gordon M. Cragg.[4]

1.1.2. Classifications

NPs are produced from living organisms derived from microorganisms, plants, and animals. Traditional medicine identified plants as an essential source for NPs while the discovery of penicillin highlighted fungal NP sources.[7,8] Following these, soil and water were investigated for their components leading to bacteria being discovered as a third prolific source of antibacterial NPs.[9] Actinobacteria is a large bacterial phylum in which the majority are saprophytic, soil-dwelling microbials with extensive secondary metabolism and produce biologically active NPs.[10] Their role in human therapeutics is impactful, being the producer of two-thirds of all naturally derived antibiotics in current clinical use as well as a variety of other therapeutic agents.[10] They make up 64% of known antibiotic classes followed by other bacteria and fungi.[11] *Streptomyces* is the largest genus within Actinobacteria accounting for 80% of the phylum.[12] A genome sequence on *Streptomyces coelicolor* revealed that the full potential of these NP producers was massively underestimated. Over 20 biosynthetic gene clusters for secondary metabolites were identified.[10] Consequently, *Streptomyces* are of great interest in the drug discovery field and are important in the fight against multidrug-resistant pathogens. Currently, they are producers of the majority of clinical antibiotics derived from NPs and makeup 50% of the antimicrobial agents used in human medicine

today, **Table 1.1** highlights a variety of essential antibiotics discovered from this genus.[12–14]

Table 1.1. Natural Product Antibiotics Discovered from Genus *Streptomyces.*

Antibiotic	*Streptomyces* Species
Chloramphenicol	*S. venezuelae*
Erythromycin	*S. erythraeus*
Kanamycin	*S. kanamyceticus*
Neomycin	*S. noursei*
Tetracycline	*S. aureofaciens* and *S. rimosus*
Rifamycin	*Amycilatopsis mediterranei*
Cycloserine	*S. garyphalus*
Vancomycin	*S. orientalis*
Novobiocin	*S. niveus*
Streptomycin	*S. griseus*
Daptomycin	*S. roseosprorus*

This table was produced using combined data from the papers: "Antibiotics produced by Streptomyces" by Procópio et al.[13], and "Streptomyces from traditional medicine: sources of new innovations in antibiotic discovery" by Quinn et al.[14]

1.1.3. Marine Natural Products (MNP)

While soil-dwelling Actinobacteria are popular sources for drug discovery the development of drug resistance among pathogenic microorganisms attracted studies to alternative ecosystems.[15] The diverse expanse and alternative environmental stress adaptations of marine habitats provide a unique opportunity for alternative NP sources of drug-relevant bacteria.[16] For example, *Salinispora* was found as a discrete genus of Actinobacteria that flourish in oceanic sediments.[17,18] Its cultivation was the first evidence of endemic actinomycetes within marine habitats. *Salinispora* has been identified as producers of bioactive NPs for use against cancer and infectious diseases.[17,18] A review written in 2018 stated that three years prior, the discovery of new marine natural products (MNPs) from marine bacteria increased by 22% highlighting this exciting source.[19] To continue this growth of MNPs, computational methodologies are crucial in the systematic design of researching this expansive territory.[20] The combination of the under-explored marine environment and the production of potent bioactive compounds by marine actinomycetes has made the marine environment an exciting avenue that the Linington lab focuses on with a NP library partially consisting of marine microbial NPs.

1.2. Natural Product Drug Discovery

With their strong bioactive profiles, NPs were targeted as a main source of research for the treatment of infections and diseases. The discovery of penicillin by Alexander Fleming in 1928 followed by its re-isolation and clinical studies in the 1940s revolutionized the NP drug discovery field.[7,21] Following the discovery of penicillin, Selman Waksman discovered soil-dwelling actinomycetes that produced biologically active compounds able to inhibit pathogenic growth.[11] He developed several culture techniques and strategies, one of which was deemed the 'Waksman platform' which instigated the golden age of antibiotic discovery from 1940-1960.[11,12,22] Waksman and his graduate students discovered several NP antibiotics and antifungals using this platform including actinomycin, streptomycin, neomycin, and clavacin.[14,23] Another breakthrough for NP drug discovery was high-throughput screening (HTS) which allowed mechanism-based screening for bioassay-guided fractionation.[21] Pharmaceutical companies and research groups adopted these methods and began curating large microorganism libraries resulting in rapid and abundant antibiotic discovery. More than 20 classes of antibiotics were discovered during the golden age however, the last classes were discovered in the 1980s.[11]

1.2.1. The Decline in Antibiotic Discovery

Large pharmaceutical companies began withdrawing from this avenue due to the challenges involved in NP discovery. A significant difficulty is the commitment and effort in identifying and isolating target NP compounds with high structural diversity, large molecular masses and complicated stereochemistry.[24] Once a target has been isolated it may only be present in a single fraction at a limited quantity hidden within a mixture. This leads to low concentrations of the target compound that is potentially too low to be detected by HTS.[25] It is also common that the compound may be unstable, have poor solubility, or contain contaminants.[26] These complex mixtures are unfavourable for HTS leaving the pharmaceutical discovery efforts favouring synthetic library HTS instead.[25] Considerable time and precision are required to identify active targets for purification from the complex NP mixtures. To ensure the production of active compounds, individual colonies are isolated from agar plates and grown through small, medium, and large-scale fermentations.[2] High-speed shakers are used to provide each growth stage with

adequate aeration for the actinomycetes to reach full growth potential.[2] This growth process of at least a month is often followed by the misfortune of rediscovery. The common approach of purifying individual compounds often leads to the isolation of the abundant components that are commonly already identified. Mainstream antibiotics occur in ~1% of soil actinomycetes and display characteristic bioactivities that often mask the interesting new antimicrobials.[26] Genome sequencing of *Streptomyces coelicolor* showed that some *Streptomyces* can harbor more than 50 different secondary metabolite gene clusters showing that the 'concealment' of these important bioactive components is habitual for these microorganisms therefore, making it harder to find novel candidates.[10] Considering these ongoing difficulties, new improvements are needed to avoid replication and profit from the lengthy isolation and identification processes.

1.3. Antimicrobial Drug Resistance

Antibiotics identified during the golden age are still in clinical use, but their effectiveness has decreased due to the rise of the antimicrobial resistance (AMR) crisis. Microorganisms including bacteria, viruses, fungi, and parasites are adapting to withstand the effects of their drug treatments and are continuing to flourish in their presence.[27] NPs produce antimicrobials through natural evolution however, resistance mechanisms to other antimicrobials develop to avoid self-toxicity and other unknown implications.[12] Specifically, bacterial AMR has emerged as a leading public health threat throughout the world as being associated with 4.95 million deaths in 2019.[28] They tested six leading pathogens for death associated with resistance; *Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Streptococcus pneumoniae, Acinetobacter baumannii, and Pseudomonas aeruginosa*, with *E. coli* responsible for the most deaths.[28] This analysis determined the percentage of fluoroquinolone-resistant *E. coli*, one of the eight leading pathogen-drug combinations responsible for the most deaths in 2019, within each country and territory. According to the results, the majority of regions have at least 10% of isolates showing resistance, therefore, highlighting the global emergence of AMR.[28]

The crisis of AMR has occurred as a natural consequence of antimicrobial use.[29] A decade after the discovery of penicillin the first sign of AMR was scientifically recorded in the form of penicillinase.[30] This trend became repetitive, discover an antimicrobial and

AMR was sure to follow. The rise of AMR began during the golden age as pharmaceutical companies were continuously producing new antibiotics leading to the widespread misuse of these drugs.[31] The indiscriminate use of antibiotics leads to selective pressure that induces microorganisms' resistance mechanisms to survive making AMR an adaptive response.[25,32] The simple solution to combat this issue during the golden age was to discover more compounds. However, the decline in drug discovery has coincided with the rise of AMR seen in **Figure 1.2**, making it a global issue today. This issue continues to grow due to the continuous misuse of antimicrobial drugs in both the medical and agricultural industries. The prescription of antibiotics has become an unregulated treatment for many infections including those that are not bacterial.[22] Their use as prophylactic agricultural supplements to promote livestock growth, prevent diseases and other implications utilizes 80% of antibiotics produced.[25] This abuse of antibiotic treatments has led to an increase in concentration of antibiotic consumption for humans as they are present in our diets and medical treatments. Therefore, this increases the selective pressure on resistance mutations for the pathogens. More than 70% of pathogenic bacteria have derived a resistant strain against most antibiotics on the market meaning a re-emergence of pre-antibiotic pathogens could occur.[31] The World Health Organization (WHO) has expressed that to control this crisis there must be rational use of antibiotics in all industries.[12] Reducing antibiotic use for treatments unrelated to human disease and bacterial infections is important in reducing the spread of AMR. Additionally, the development of the drug discovery pipeline is essential, innovative ideas are needed to revolutionize this field of research.

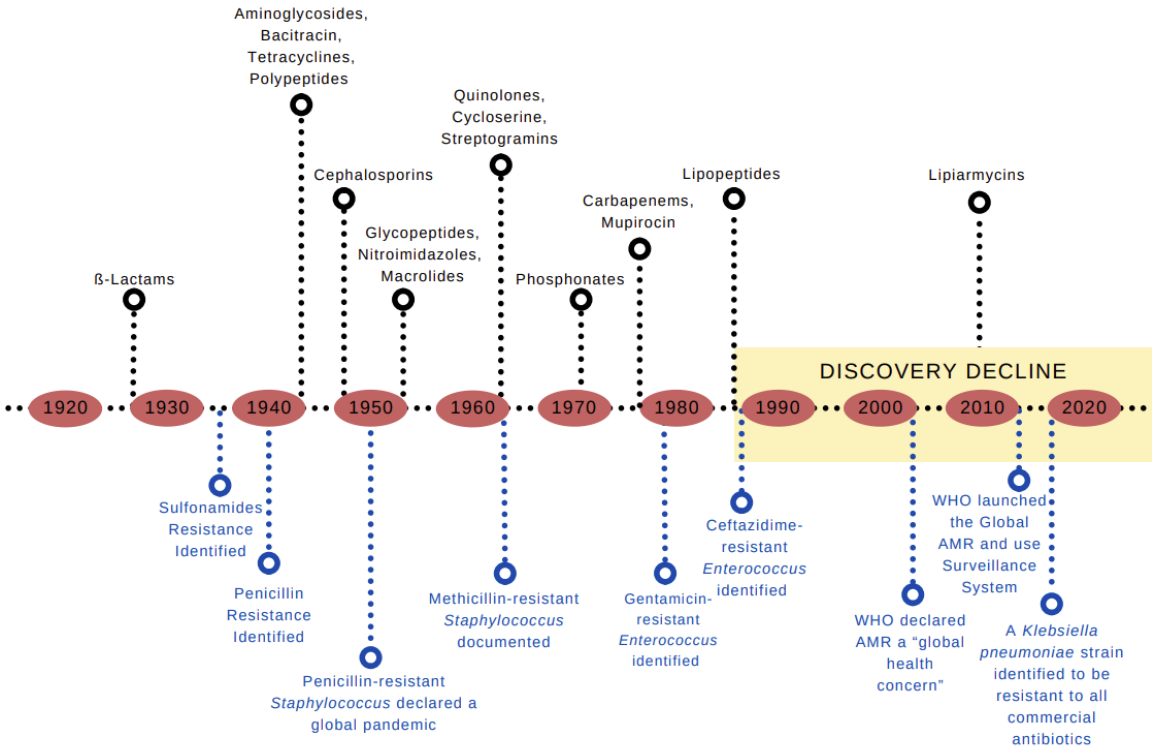


Figure 1.2. Timeline of antibiotic class discovery on the top and events associated with AMR on the bottom.

1.4. Techniques for Antimicrobial Drug Discovery

1.4.1. Computational Methodologies

Two common approaches to combat the bacterial AMR crisis are (1) develop entirely new antibiotics or (2) modify existing treatments to be effective against mutated strains. The first step in investigating NPs is bioassay-guided fractionation, a common technique used to simplify the complexity of NP mixtures. Each stage of fractionation reduces the complexity of the mixture and eventually, the desired bioactive component can be isolated and characterized from the separated prefractions. The development of diverse NP libraries containing these crude extract prefractions is a common high throughput technique for NP drug discovery as it allows bulk analysis of an expansive dataset. Often the first analysis performed on these libraries is the determination of the metabolite composition of each extract obtained through metabolomics.[3] Collecting mass spectrometry (MS) data on a complete NP library yields large volumes of data that make prioritization of compounds a significant challenge. However, the development of

computational platforms makes this process easier by filtering these expansive libraries depending on their identification and classification. These platforms allow rapid dereplication and prioritization of extracts saving valuable work time and aiding in the avoidance of rediscovery of known compounds.[33,34] A common approach is to obtain metabolite fingerprints of extracts through MS to allow the differentiation between extracts containing novel or known masses.[35] Molecular networking is a computational method that produces a visual representation of clusters consisting of structurally similar metabolites that are predicted to share similar MS/MS fragmentation patterns. It has become a beneficial approach to visualize MS datasets and search for metabolite identification within complex NP mixtures.[34,36,37] With the comparison to databases consisting of known analogs these networks allow the rapid prioritization of potential novel candidates that do not share characteristics with these databases.

Traditional techniques for NP drug discovery revolve around the isolation of active NP compounds. However, these approaches have decreased in the identification of new lead compounds. An alternative approach in NP discovery recently is genome mining. It is used to detect the biosynthetic pathway of a bioactive NP and its possible chemical interactions. Genome mining identifies uncharacterized biosynthetic gene clusters (BGCs), predicted to produce NPs, within the genome of a sequenced organism and uses known information to guide the identification and isolation of the target it encodes.[38,39] There is a large amount of data on DNA sequences and their annotations within public databases therefore, genome mining is dependent on computational technology and bioinformatics tools to compare the unknown BGCs with these informative databases. AntiSMASH is a platform used to identify BGCs, determine the enzymatic components, and generate structure predictions. It then compares the results with a database to identify possible relations with known chemistry leading to the rapid deconvolution of unknown NPs.[40] PRISM is an alternative tool for genome mining that predicts genetically encoded NP structures based on microbial genomes. It identifies BGCs and generates a library of structural predictions.[41] Computational tools such as these represent an essential route for the growth of NP discovery. Genome mining has become a popular technique for NP identification and characterization due to the rapid analysis performed by these platforms.

Computational approaches aim to significantly contribute to the efficiency of NP-based drug discovery. The identification and prioritization capabilities of these platforms

help guide research to more promising results. NP databases are computational tools that can aid in identifying the chemical, biological, and structural information of NPs.[42,43] NP Atlas is an open online based platform that contains all known microbial NP structures. The search tool includes both structural and spectral options when looking for matches with an unknown NP target. It can be used for the rapid identification of an isolated compound.[44] Computational platforms can also be used to avoid rediscovery of known compounds. Combined with analytical data, such as MS or nuclear magnetic resonance (NMR), known or undesirable NPs can be deprioritized.[42] These methods can identify spectra from the same compounds while also from structurally related compounds leading to hints on the chemical classes and functional groups of the compound. An example of this kind of platform is METLIN; a high-resolution MS database that allows the identification and characterization of metabolites. It contains over 1 million molecules including primary metabolites, toxins, small peptides, and NPs that are used to identify components of unknown compounds using MS data.[43] There are structural and spectral dereplication platforms however, there are few that apply dereplication using bioactivity profiles. There is a need in NP research for the combined efforts of biological and chemical datasets. With bioactivity being the traditional guide for NP isolation, a platform that could identify what component is contributing this activity would be useful. This project combines biological activity and MS data to predict the bioactive metabolites within a NP mixture.

1.4.2. Bioassay Screening Methodologies

With the importance of biological activity for medicinal applications, bioactivity guided isolation for NP discovery is essential. Bioassays allow the measurement of activity of a substance on a living material. Common applications include toxicity, antiviral, anticancer etc. assays that measure the activity of compounds against different factors such as pathogens, animal cells, human cells etc. Recently, strategies have begun turning to alternative bioassay panels to focus on combatting the AMR crisis. Resistance-based phenotypic screening has been attempted for the identification of novel antibiotics.[45] An antibiotic resistance platform that uses a cell-based library of drug-resistant strains of *E.coli.* was developed to measure the antibacterial activity of NP extracts to prioritize dereplication in NP drug discovery.[45] If the extract contained a known antibiotic then its antibacterial activity would be reduced against the resistant

strains of *E. coli*. It was successful in the dereplication of *Amycolatopsis mediterranei* (rifamycin) and *Saccharopolyspora erythraea* (erythromycin).[45] This method allows the discovery of potential drug candidates that would be effective against resistant bacterial diseases specifically. If used in combination with molecular networking, it could be a new approach to NP discovery that would avoid replication while focusing on finding targets effective against AMR.

1.5. Summary and Proposal

NPs are an essential source of medical therapeutics and instigated a rush in antibiotic discovery between 1940-1960.[11,22] This massive increase in antibiotic discovery improved life expectancy from 40 to 77 years during the 20th century however, it quickly dissipated as pharmaceutical companies began withdrawing from this avenue of research.[24,26] NP discovery is known to be a laborious, inefficient process due to the occurrence of rediscovery that pharmaceutical companies prefer to avoid. A natural form of protection for microorganisms is to essentially harbour important secondary metabolites leaving only a few to be easily isolated resulting in continuous identification of the same compounds.[10,25] Computational methods are being used for the rapid dereplication of NPs through database comparison applications and metabolomic analysis.[32–34] These methodologies are useful dereplication processes but prioritization of unique bioactive compounds is additionally essential for the discovery of novel NP antibiotics. The decline in drug discovery coincided with the rise of AMR, the global health crisis. The indiscriminate use of antimicrobials allows an increase in the selective pressure for resistance mutations of these microorganisms.[24,31] Specifically, 4.95 million deaths in 2019 were associated with bacterial AMR, with *E. coli* being responsible for the majority.[28] A simple solution to combat this crisis would be to discover completely new antibiotics however this is an ongoing difficulty. This project uses two analysis tools, ResistoMAP and NP Analyst, developed in the Linington lab that observe bioassay and metabolomic data in a unique way to assist in the selection of target compounds for isolation. These tools offer a novel approach for NP discovery that will minimize the rediscovery of known compounds while identifying candidates effective against drug-resistant pathogens. ResistoMAP is an antibiotic profiling platform that utilizes a panel of drug-resistant strains of *E. coli* and the wildtype. The Linington NP extract library along with 80 commercial antimicrobial standards were screened against the ResistoMAP

panel. The phenotypic profiles outlining the overall activity profiles produced from ResistoMAP are compared against those of the standards as a dereplication method while simultaneously providing insight into the mechanism and possible structure of the bioactive species.[46,47] The initial ResistoMAP dataset consisting of ~6,000 NP prefractions is filtered to only include prefractions that display 50% or more growth inhibition against 5 or more drug-resistant strains of *E. coli*. Mass Spectrometry (MS) data is obtained on those extracts that pass the filtering along with the 80 antimicrobial standards. The ResistoMAP data along with the MS profiles are inputted into NP Analyst, an online platform designed to identify bioactive metabolites within complex NP extracts. With the combination of these platforms, multiple dereplication techniques are provided that increase the outcome of selecting an extract that contains novel chemistry. Following the initial application of these programs, 18 extracts are selected for further research and analyzed a second time using these platforms. The secondary application of ResistoMAP and NP Analyst allows further dereplication and the highly specified selection of 3 mass candidates for complete isolation and characterization with the aim of discovering novel chemistry that will be effective in combating the AMR crisis.

Chapter 2.

ResistoMAP Data Acquisition and Analysis

2.1. Introduction

This project represents a novel approach to natural product (NP) discovery that minimizes the outcome of rediscovering a known compound while focusing on finding candidates that will be effective against resistant bacterial pathogens contributing to the antimicrobial resistance (AMR) crisis. Rediscovery is an ongoing issue that has strongly influenced the decline in NP drug discovery. Alternative approaches for dereplication are essential for the continued research in this avenue. Although AMR is a rising crisis it does offer new opportunities for NP drug discovery and applications. Screening against a panel of drug-resistant bacterial strains allows for the specified identification of candidates that could assist in combating the AMR crisis. ResistoMAP and NP Analyst are two applications this project uses to address these two current concerns in the NP research field.

2.1.1. ResistoMAP

Resistance **M**ode of **A**ction **P**rofiling (ResistoMAP) is a recent development for antibiotic profiling designed by the Linington research group. This platform is designed for the dereplication of known classes through the mechanism of action while highlighting NP with unique phenotypic profiles.[46] It involves the screening of NP extracts against a target panel of 29 drug-resistant strains of *E. coli* along with the wildtype, MG1655, to produce a unique phenotypic profile of bioactivity.[46,47] This panel may sound similar to that reported earlier by Cox *et al* but instead, ResistoMAP uses a panel of single mutants that connect cell survival with drug resistance.[45,46] Other variables remain unchanged meaning that the resistance observed is a direct result of a singular mutant gene. It is designed to provide a more informative prediction of a possible drug-resistance mechanism for the potential candidate being investigated.

Dr. Dennis Liu screened both the Linington extract library, consisting of 6,195 NP extract prefractions, and a positive control library of 80 commercially available

antimicrobial compounds, against the panel of drug-resistant *E. coli* and wildtype strains. The positive controls covered a range of 30 drug classes and were used to dereplicate the phenotypic outputs.[46,47] Hierarchical clustering is used to group similar phenotypic fingerprints. This platform will dereplicate those extracts depicting similar fingerprints to antimicrobial standards and therefore, prioritize those displaying unique bioactivity. For example, **Figure 2.1** shows the hierarchical clustering of phenotypic profiles of rifamycin-like standards with two prefractions RLUS-1505C and RLUS-1530C.[46] The standards are showing activity against all drug-resistant strains of *E. coli* except the rifamycin-mutated strains, the extract profiles are reflecting this pattern. Therefore, these extracts can be dereplicated as containing rifamycin-like compounds. Using this analysis technique, these fingerprints will provide a guide to the mechanisms and possible structural characteristics of the bioactive species within the NP extract. The clustering application of this platform allows the elimination of extracts clustering with antibiotic standards while prioritizing those with exclusive phenotypic signatures.

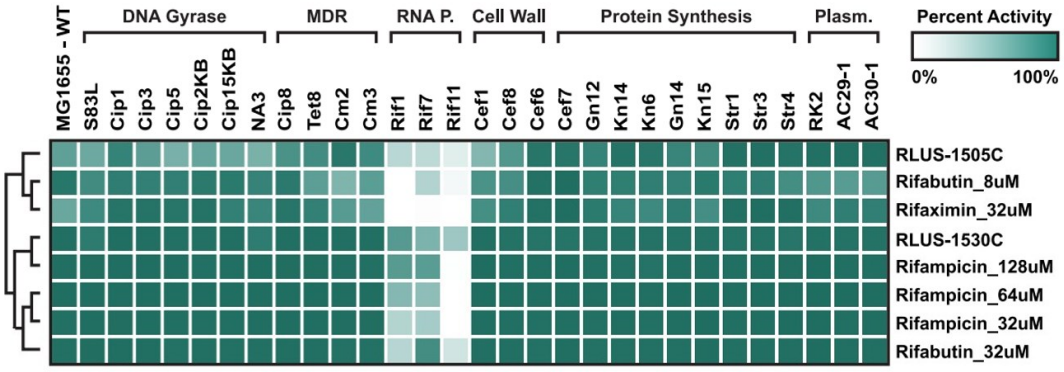


Figure 2.1. ResistoMAP screening showing hierarchical clustering of prefractions RLUS-1505C and RLUS-1530C clustering with rifamycin-like standards between 8 to 128µM.[46]

2.1.2. NP Analyst

NP Analyst is an open-access online metabolomics software designed to identify bioactive metabolites within complex NP mixtures.[48] It relates the presence/absence of mass spectrometry (MS) features (detected *m/z* with its corresponding retention time (Rt)) to bioactivity profiles. Its compatibility with both bioassay data and MS from all major instrument manufacturers provides scientists with a discovery platform that allows extensive dereplication and highlights significant metabolites of interest. It does this by

comparing the distribution of MS features within a mixture against its biological signatures.[48]

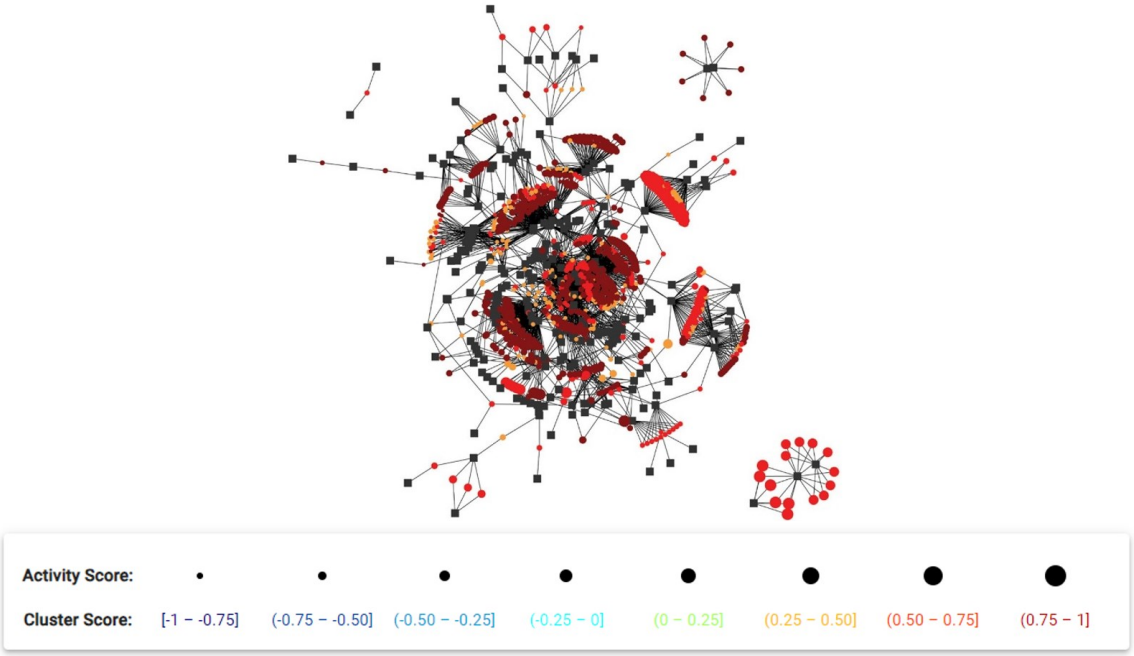


Figure 2.2. NP Analyst default output network showing clusters of NP extracts sharing similar bioactive MS features. Each square node represents a Sample while the circular nodes correspond to MS features.

This figure is taken from the default output on the NP Analyst online platform.[48]

NP Analyst includes two scores that are used to select priority MS features: Activity and Cluster score. The Activity score takes a biological profile of a given MS feature and measures its intensity. It takes an average of the bioactivity observed against each strain for samples showing similar profiles and then takes the sum of these average values. The maximum Activity score is equal to the total tested strains within the bioassay and is represented by the size of MS feature circular nodes seen in **Figure 2.2**. The Cluster score is a representation of the biological fingerprint consistency between all the extracts containing the specific MS feature. The average Pearson similarity score of all the extracts biological fingerprints containing the MS feature is equal to the Cluster score, these values cover a range of -1 to +1 which corresponds to the colour of MS feature nodes in a network (**Figure 2.2**). These scores are calculated for every MS feature. NP Analyst allows cut-offs for both scores to retain only significant data in turn reducing the complexity of the resulting network. These filters allow the removal of MS

features with weak bioactivity profiles that may not correlate with desirable biological phenotypes.

NP Analyst provides a visualization option known as community viewing which divides the full network into smaller interconnected networks called communities (**Figure 2.3**). The Louvain method is used to filter the data into these distinct individual networks. Each one contains various nodes containing the same biologically active MS features allowing the prioritization of interesting extracts. This visualization produces an individual network view, two scatter plots (Rt vs. precursor *m/z* and Cluster score vs. Activity score), and a bioactivity heatmap of each community (**Figure 2.3**). This output allows users to easily identify biological similarities between extracts within a community and distinguish between the bioactive MS features interconnected between the extracts. Investigating the data through this visualization allows a specified view of MS features permitting the dereplication and prioritization for compound isolations.

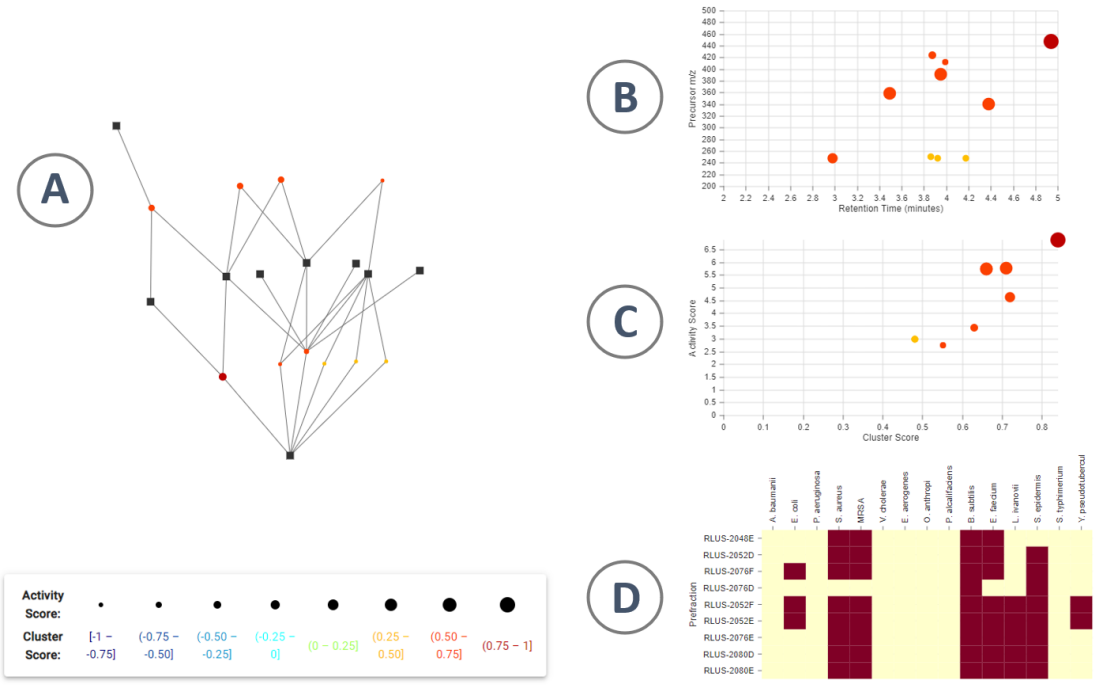


Figure 2.3. NP Analyst community viewing including network view (A), plot of retention time vs. precursor *m/z* for bioactive MS features (B), plot of Cluster Score vs Activity Score for bioactive MS features (C), and activity phenotypic fingerprints for extracts within the community (D). Rows are the NP extracts and columns contain the bioassay values.

This figure is an unofficial adaptation from "NP Analyst: An Open Online Platform for Compound Activity Mapping" by Lee *et al.*[48]

2.1.3. Combined Applications of ResistoMAP and NP Analyst

This project combines both ResistoMAP and NP Analyst to provide multiple dereplication routes and improve prediction confidence. There are three main routes for dereplication: the clustering of standards with extract profiles in ResistoMAP, prior research of NP extract chemistry, and the metabolomic applications NP Analyst provides. The first route was noted earlier: the clustering analysis performed in ResistoMAP allows the prediction of bioactive chemistry based on fingerprint similarities. This prediction can then be confirmed with a comparison of metabolomics data using NP Analyst. These comparisons reduce the number of possible known candidates leaving a smaller, more promising dataset to investigate further. Alternatively, if clustering is not used then prior knowledge of extract chemistry can be applied to find other extracts that share similar chemistry. Lastly, NP Analyst allows a metabolomic perspective that can be applied to the candidates of interest from ResistoMAP. MS data is acquired and produces mass and Rt data that can be used in computational databases to further dereplicate these candidates. The clustering format in NP Analyst and ResistoMAP groups those extracts showing similar chemistry therefore, comparison of metabolomic data allows the rapid dereplication for those that cluster together.

NP extracts are complex mixtures, often containing multiple components of interest. A challenge of biological screening is the overlap of multiple bioactive species that can cause a misleading profile showing increased activity or the lack of diagnostic patterns. To reduce these effects biological profiles that display potent activity across the entire panel are removed from the dataset prior to using NP Analyst. With its application of identifying bioactive metabolites, NP Analyst will identify the MS feature responsible for the bioactive component of interest. Therefore, these misleading profiles become easier to analyze as the bioactive MS feature can be identified individually. Once the MS feature is identified the specific MS profile can offer insight into the components within the suggested compound. Additionally, if these NP extract mixtures are separated from prefractions into subfractions it can further isolate the bioactive component and lead to a more knowledgeable selection of candidates.

2.1.4. Selectivity Ratio Profile Network (SRPNT) and Multivariate Pattern Analysis (MVPA)

ResistoMAP and NP Analyst are the two main applications used in this project however, the Selectivity Ratio Profile Network (SRPNT) is an in-house tool that was tested as an additional application. Metabolomics are used to profile complex NP mixtures. To analyze this data efficiently, strategies are needed to identify the meaningful information from these complex datasets.[49] Additionally, as NP Analyst does, data analysis methods are needed that integrate both biological activity and metabolomics. Partial least-squares (PLS) is a method that can break down spectral data into uncorrelated latent variables to maximize the variance of independent variables (MS features) with a dependent variable (biological activity)[49]. It is used to find the relations between these variables, such as what about the independent variable explains the dependent variable. Using a technique called Target Projection, the feature that explains most of the activity observed can be determined.[50] Target projection allows the calculation of a selectivity ratio (SR) for each analyte. SRs is a ratio between explained predictive and residual variance that represents a measure of the ability of a variable to discriminate different groups.[51] Explained and residual variables are calculated for each x-variable, in this case x-variables represent MS analytes, full equations for these can be found in a study by Kvalheim *et al*.[51]

Multivariate Pattern Analysis (MVPA) is a full R-package of techniques that analyze multivariate patterns while SRPNT is a wrapper program that can be applied to identify active metabolites within a complex NP mixture using SRs as a fingerprint of the independent variable.[52,53] These software are used for the unique interpretations and visualization of collinear data in terms of predictive association patterns.[52] Both SRPNT and MVPA use PLS and Target Projection to model relationships between cause and effect, independent (MS features, x-variables), and dependent variables (activity, y-variables) by calculating SRs for the x-variables. However, the key difference between the two is that MVPA works with a singular y-variable while SRPNT can use multiple. With SRPNT each x-variable will have as many SR values as there are y-variables therefore, SRPNT can take every x-variable and calculate how responsible the feature is for the antibacterial activity depicted. This creates a SR profile for each feature creating an alternative approach for analyzing NP chemistry. With this SR profile, one can determine which analytes/molecules have similar chemistry when looking at the

antibacterial screens, and which are unique. Molecules may display different chemistry but exhibit the same SR profiles, in turn clustering together within the network SRPNT produces. SRPNT requires a MS feature table and bioactivity data to produce these results. A high SR means the corresponding x-variable has a strong association with the predicted y-variable.[51] The SR profiles can be interpreted in this way to rank x-variables associations to y-variables which can lead to a candidate molecule being selected for isolation.

2.2. Results

2.2.1. ResistoMAP Selections

The initial ResistoMAP dataset was reduced from 6,195 NP extract prefractions to 384 by filtering the data to only include extracts that showed an activity of 50% or more against 5 or more drug-resistant strains of *E. coli*. The original ResistoMAP study performed by Dr. Dennis Liu identified six extracts displaying unique phenotypic fingerprints (**Table 2.1**). Extracts RLUS-1505 and RLUS-1530 profiles clustered together with rifamycin-like standards and were selected for full dereplication of this active component. RLUS-2028, RLUS-2045, RLUS-2100, and RLUS-2204 all showed unique activity against specific drug-resistant strains therefore, the profiles were flagged for further investigation. Some extracts selected strictly based on ResistoMAP output were later identified within the NP Analyst network.

2.2.2. NP Analyst Network Analysis

MS data was obtained on these 384 NP prefractions and 73 of the 80 commercial standards (due to availability) to produce the first NP Analyst network, **Figure 2.4**. When observing **Figure 2.4** the square nodes represent an NP prefraction or one of the commercial antimicrobial standards while each circle node corresponds to a MS feature (a *m/z* with its corresponding Rt). If a square connects to a circular node it means the prefraction/standard contains that feature therefore, if two samples connect to one MS feature they both contain that compound. Lastly, the colour and size of a circular node correspond to the legend seen in **Figure 2.4**.

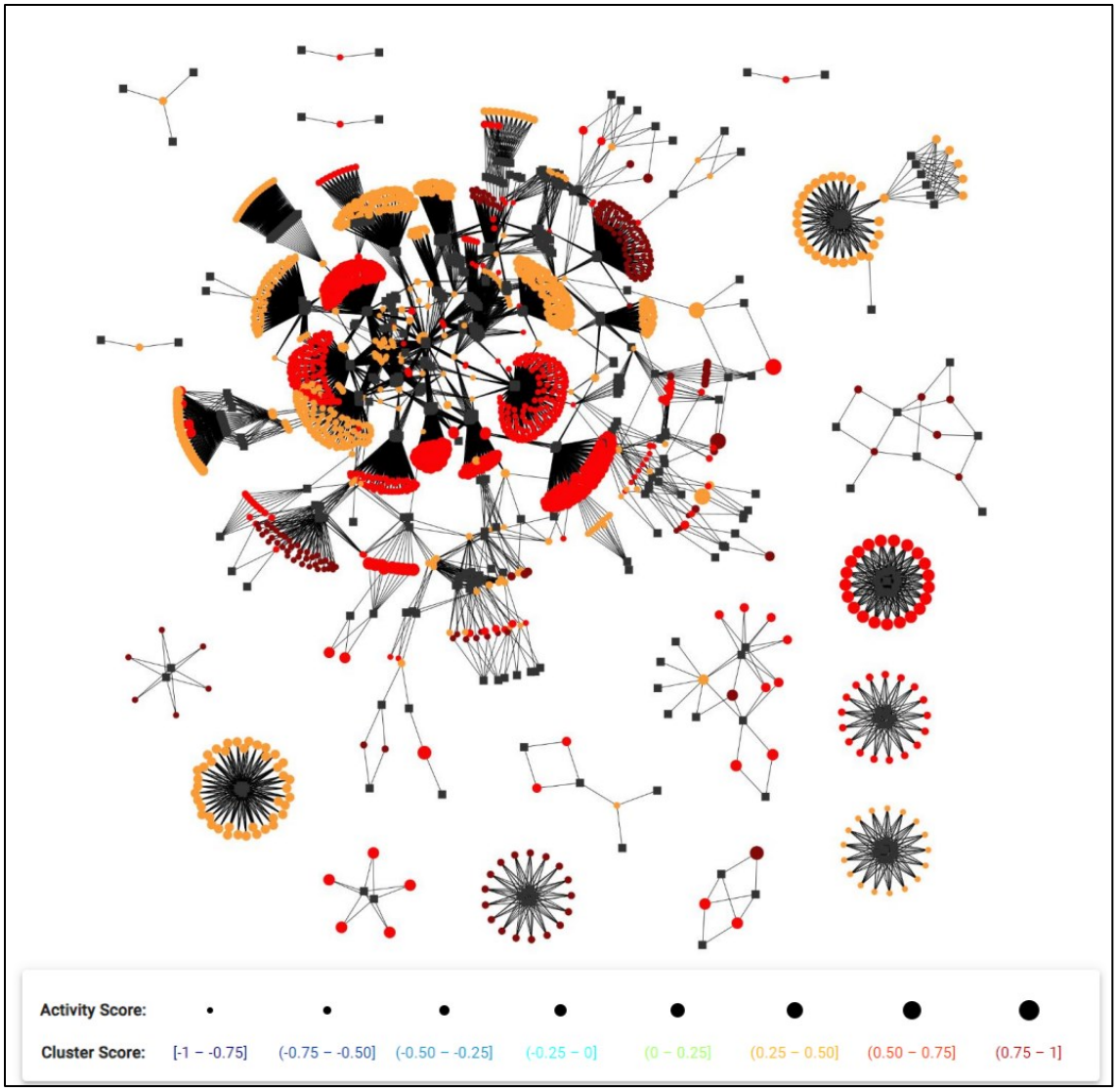


Figure 2.4. Full NP Analyst output consisting of the 384 selected NP extracts and 68 commercial antimicrobials from the ResistoMAP dataset.

Using the community viewing option this network was split into 36 communities that could be divided into 3 different categories. The first were communities that only contained commercial antimicrobials. The second category was communities that contained both NP prefractions and antimicrobial standards. Community 4 consists of several extracts, specifically RLUS-1878C-D prefractions near Spiramycin. NP prefractions found within this community were dereplicated for associations with these standards (**Figure 2.5**).

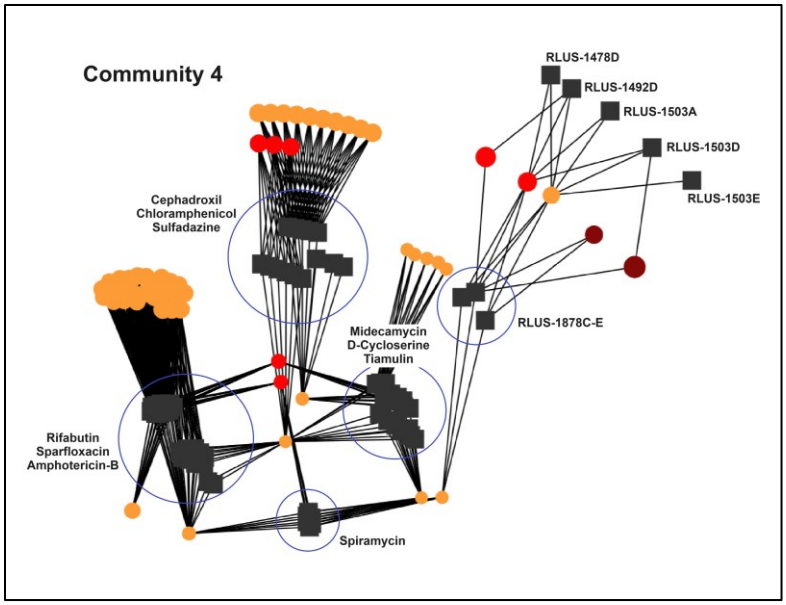


Figure 2.5. Community 4 used for dereplication as it contains the NP prefractions and commercial antimicrobials labeled. Specifically, RLUS-1878 prefractions are directly connected to Spiramycin.

The final category was communities that contained only extracts – these were useful for candidate selections. Community 10 consisted of a variety of extracts displaying an array of activity and cluster scores. Prefractions RLUS-1758B&C, RLUS-1806B, and RLUS-1885B&C were selected for strong correlations within the network (**Figure 2.6A**). Community 27 displayed a selective cluster between 7 extract prefractions all showing features with very strong activity scores. From this community prefractions RLUS-2085E and RLUS-2024E were selected to research further (**Figure 2.6B**). **Appendix Table A.1** lists MS features identified within specific communities that led to the selection of NP prefractions for further steps.

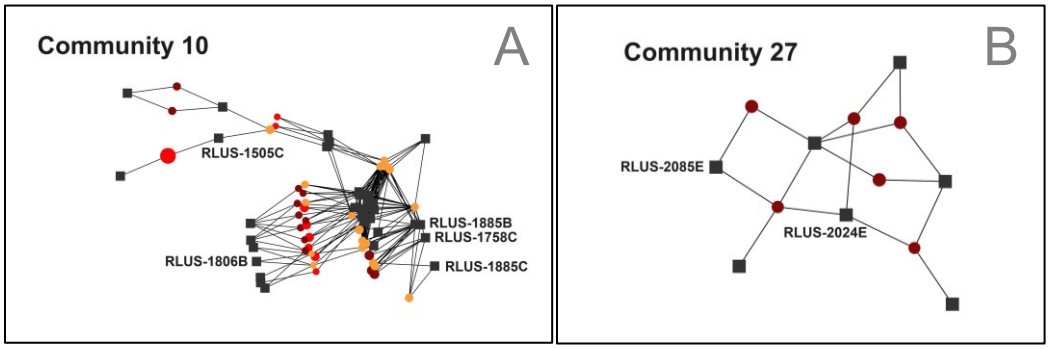


Figure 2.6. Community 10 (A) and 27 (B) belong to the third category of communities. Each prefraction label corresponds to the closest square node.

A final dereplication process was completed using knowledge of compounds previously isolated from the Linington library. Community 26 contained only prefractions RLUS-2108C, D&E. Collismycin A (in-source mass fragment = 244.06 *m/z*, Rt =2.64 min) and related analogue SF2738D ($[M+H]^+$ = 258.07, Rt = 3.43 min) have been previously identified in these RLUS-2108 prefractions. These were both identified within this community. Using these methods 18 extracts (**Table 2.1**) were chosen to continue the project.

Table 2.1. List of extracts grown up through fermentation processes for novel mass candidates for further research.

Extract	Selection Process
RLUS-2028	Chosen from ResistoMAP profiles
RLUS-2045	
RLUS-2100	
RLUS-2204	
RLUS-1505	
RLUS-1530	
RLUS-1459	Selected from NP Analyst Community Visualization
RLUS-1520	
RLUS-1595	
RLUS-1597	
RLUS-1623	
RLUS-1726	
RLUS-1758	
RLUS-1806	
RLUS-1885	
RLUS-2024	
RLUS-2064	
RLUS-2085	

2.2.3. Secondary ResistoMAP

These 18 NP extracts were grown from frozen stocks of the bacterial isolates through a fermentation process explained in **Section 2.4.4** and separated into 5 prefractions, A-E, through column flash chromatography. 15 of the 18 extracts underwent secondary separation from prefractions into peak libraries each consisting of 10 subfractions using peak-guided separation via HPLC. Each prefraction was

separated into its peak libraries using a customized gradient. Secondary ResistoMAP screening was performed on each of these peak libraries. This panel consisted of 9 of the original 29 drug-resistant strains of *E. coli* and the wildtype, commercial antimicrobials were not included. Output data was normalized to indicate the activity at a certain concentration. If the subfraction was active at the lowest concentration (3.75 µg/mL) it got a score of 1, at the highest concentration (30 µg/mL) then a score of 0.25. RLUS-2024 presented consistent profiles in the E2-4 subfractions as well as lower detection in the D2-4 subfractions. RLUS-2085 showed distinct activity in the D2-4 and strong profiles for E subfractions. Consistently strong activity was seen against Cef6 and Cef7 for the subfractions of RLUS-2028. For the A-2, C-2, and D-2 subfractions, there was activity seen against the full panel for this extract. RLUS-2204 had a unique profile in the original ResistoMAP dataset and shows a distinct activity profile in this secondary dataset. The fingerprints are consistent in both the D2 and E2 subfractions with D2 showing stronger signals (**Figure 2.7**).

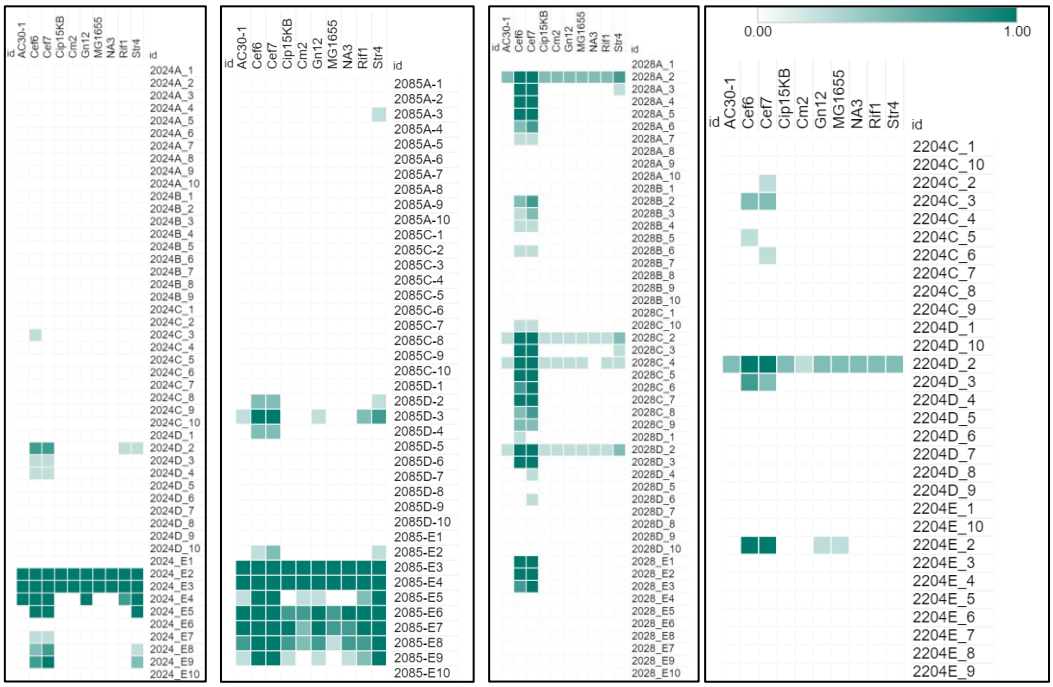


Figure 2.7. **Peak Library ResistoMAP profiles against 9 drug-resistant strains of *E. coli* and wildtype MG1655 for extracts RLUS-2024, RLUS-2085, RLUS-2028, and RLUS-2204.**

2.2.4. Secondary NP Analyst and SRPNT

MS data was then obtained on the subfractions to create a secondary NP Analyst network for the selection of mass candidates for isolation (**Figure 2.8A**).

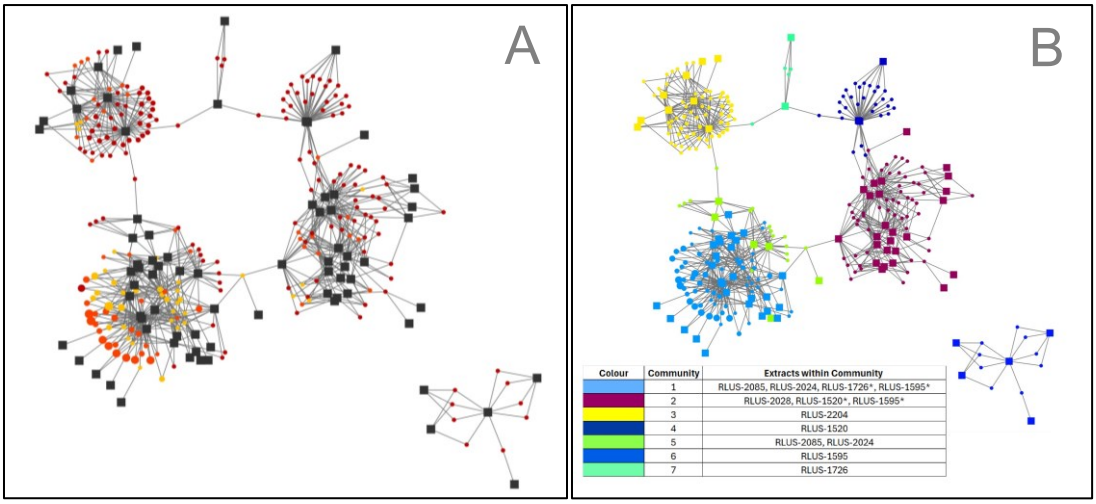


Figure 2.8. Secondary ResistoMAP-NP Analyst network made using the extract subfraction ResistoMAP output (A). Secondary NP Analyst network coloured by community (B). Legend outlines what extracts are found in which community. * Extract only has one node found in the community.

Candidate Selection Using SRPNT

Using the MS and ResistoMAP data, SRPNT was able to predict which MS feature was showing the strongest activity relative to every other feature via selectivity ratios. This output pinpointed one mass: 392.27 *m/z* (Rt 3.90 min) from RLUS-2085E and RLUS-2024E subfractions (**Figure 2.9**). It showed strong selectivity ratios against the majority of the tested strains and overpowered the rest of the data, 392.27 *m/z* was the only candidate selected using this method.

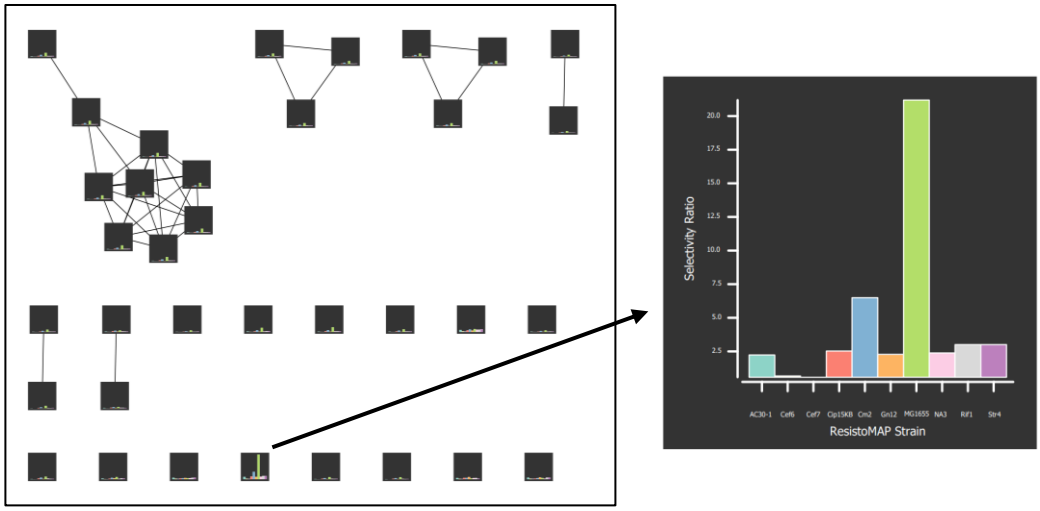


Figure 2.9. SRPNT Cytoscape output of 392.27 *m/z* showing the highest selectivity ratios against each drug-resistant and wildtype strain of *E. coli*. This mass was most prominent in RLUS-2085E-3 subfraction.

Candidate Selection Using NP Analyst Network

Figure 2.8B displays the secondary network separated into 7 communities, each one containing the extracts of interest. The same candidate found in MVPA was identified in Community 1. It is found in both RLUS-2085E and RLUS-2024E subfractions. When looking at the MS data (**Figure 2.10**) shows an intensity of 3.3×10^6, and a clean MS^1 profile in RLUS-2085E-3, making this subfraction the best candidate for isolation due to the intensity and separation of this feature. Another feature, 394.30 *m/z* (Rt 4.29 min) showed a strong activity and was found in RLUS-2085E-6 and -9 subfractions. MS data identified RLUS-2085E-6 with an intensity of about 3.9×10^6, the MS^1 profile looked relatively clean (**Figure 2.11**). Using NP Atlas both these masses were identified as similar to those of the prodiginine family and were selected for isolation.

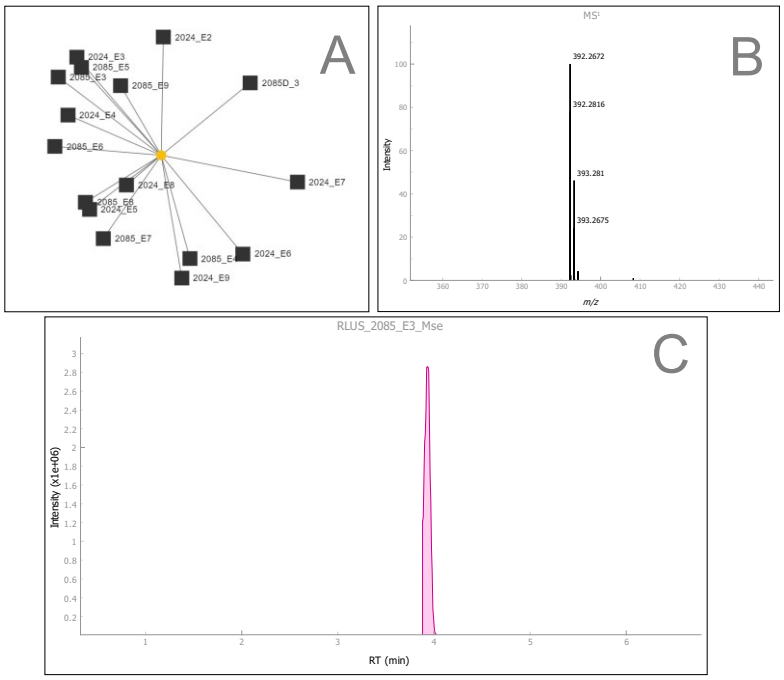


Figure 2.10. Candidate mass feature 392.27 *m/z* for isolation from RLUS-2085E-3. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2085E-3 (C).

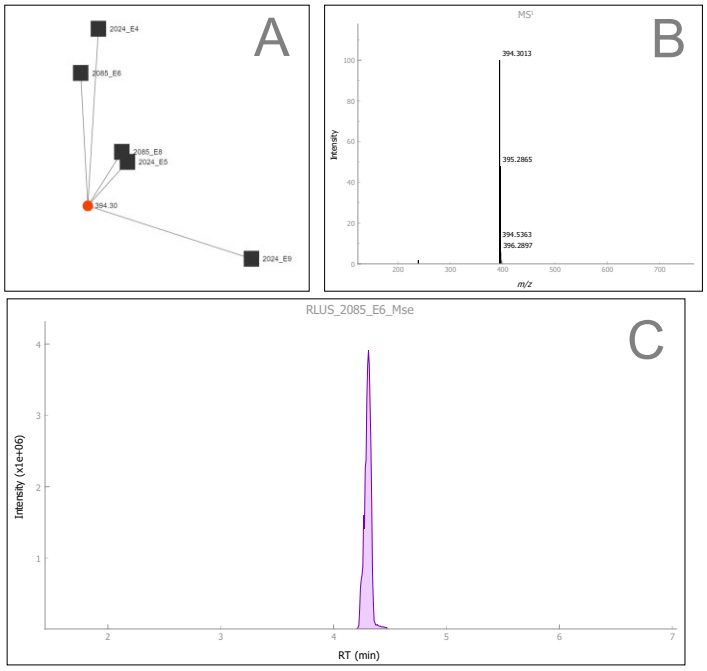


Figure 2.11. Candidate mass feature 394.30 *m/z* for isolation from RLUS-2085E-6. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2085E-6 (C).

Community 2 in **Figure 2.8B** consists of the majority RLUS-2028, an extract initially selected for its ResistoMAP profile. Features 438.16 *m/z* (Rt 1.58 min) and 452.18 *m/z* (Rt 1.64 min) were identified to display strong activity and cluster scores relative to all other features identified in this community. The MS data identified both masses in the second subfraction of RLUS-2028A-C, with the intensity being the highest in RLUS-2028A-2 for both. MS data for 438.16 *m/z* can be seen in **Figure 2.12** intensity was 1.4×10^6 and MS profile is strong with a few potential contaminants. The MS data for 452.18 *m/z* was very similar with the same intensity. NP Atlas had no known matches for these masses and isolation from subfraction RLUS-2028A-2 was chosen.

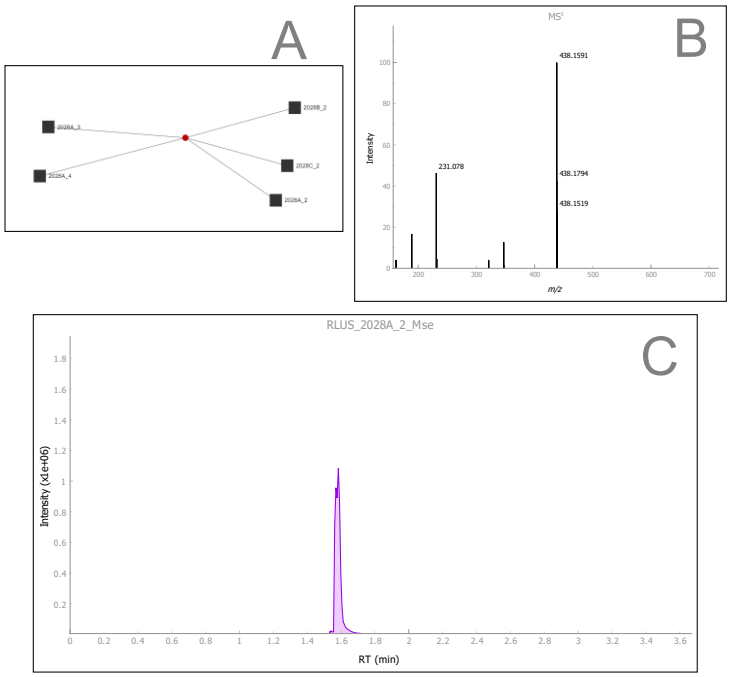


Figure 2.12. Candidate mass feature 438.16 *m/z* for isolation from RLUS-2028A-2. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2028A-2 (C).

RLUS-2204 showed unique phenotypic profiles in both the original and secondary ResistoMAP data set (**Figure 2.7**). A group of masses within a few mass units of one another (**Table 2.2**) was found in Community 3 which consists of extract RLUS-2204.

Table 2.2. Masses identified from secondary NP Analyst network from RLUS-2204.

MS Feature (*m/z*)	Retention Time (min)
538.38	3.10
552.40	3.19
568.39	2.94
582.36	3.06
598.36	2.27

After looking at the MS profiles, RLUS-2204D-3 was selected for isolation as it contained multiple of these masses, specifically 552.40 *m/z* (Rt 3.19 min) with high intensity (1.6×10^6) and a strong MS profile, making it the main target for isolation. NP Atlas matched this mass with that of a macrolide.

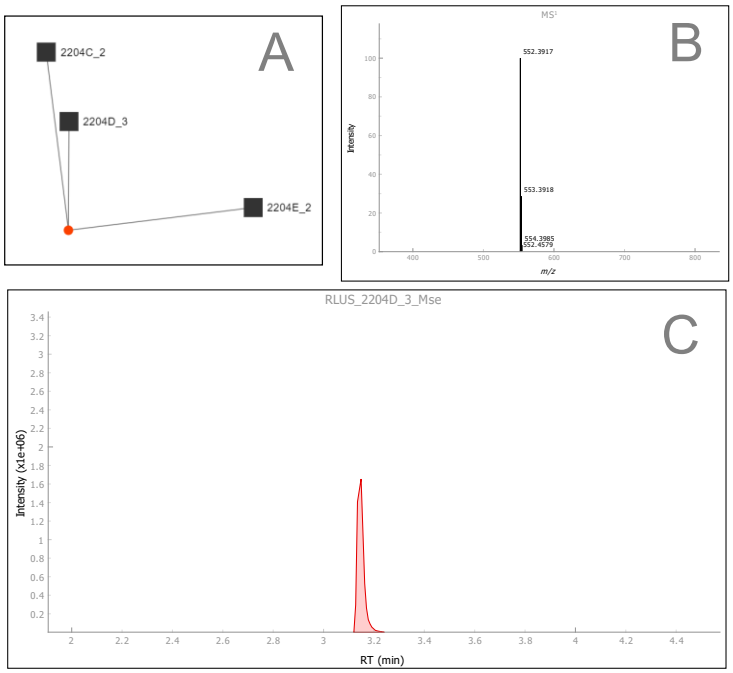


Figure 2.13. Candidate mass feature 552.40 *m/z* for isolation from RLUS-2204D-3. NP Analyst network view (A), MS^1 data of analyte from MS2Analyte (B), and analyte intensity in RLUS-2204D-3 (C).

2.3. Discussion

In this chapter, we outlined the process by which the individual and combined applications of ResistoMAP and NP Analyst can assist in the prioritization of novel candidates while avoiding rediscovery.

ResistoMAP uses the target panel of 29 drug-resistant strains of *E. coli* and wildtype, covering a range of drug classes each consisting of several genotypes of resistance, to produce phenotypes that predict the possible drug mechanism and/or structure of an unknown compound.[46,47] Decreased activity against a specific set of mutants within the panel provides insight into the mechanisms of action and structural characteristics of the bioactive species. Additionally, the hierarchical clustering with antimicrobial standards allows a dereplication process whereby if the profiles differ significantly then the NP extract may contain a novel and/or unrepresented bioactive species. These processes outline ResistoMAP's ability to aid in the prioritization of extracts for future discovery developments. A filter was applied to reduce the volume of the original ResistoMAP dataset for the extent of this project. We focused on relevant ResistoMAP profiles that were showing distinct activity profiles to ensure a possible bioactive candidate was prioritized. An activity of 50% or higher filtered out all the NP mixtures that showed unimpactful profiles, leaving 384 NP prefractions for a workable dataset. The original ResistoMAP study performed by Dr. Dennis Liu flagged extracts as showing unique phenotypic profiles. Using this bioactivity-guided selection, 6 extracts (**Table 2.1**) were chosen, 4 of which were included in the ResistoMAP filtering, and the remaining 2 (RLUS-2204 and RLUS-2100) were added in the secondary steps of this project. To further facilitate dereplication and future prioritization efforts, MS-based metabolomics data were obtained on the 384 prefractions along with 73 (of the original 80) readily available antimicrobial standards.

NP Analyst uses bioassay data and MS-based metabolomics to predict bioactive metabolites. It is built on the rationale that if compounds share the same bioactive metabolites, then these compounds will exhibit similar chemistry. With rediscovery being a major hurdle in NP discovery, NP Analyst applies this rationale to NP mixtures to produce a network that groups compounds with shared chemistry together. It analyses the phenotypic profiles to identify which NP mixtures are displaying similar profiles. If these profiles are being produced by the same source, then the Rt and mass-to-charge ratios of this active component will be identified in both compounds. In summary, if NP mixtures contain the same bioactive compounds, then the metabolomic data and biological screening output will reflect that. NP Analyst provides a visual representation of this comparison by producing a network that connects NP prefractions that show similar bioactive metabolites. It scores the biological screening profiles (Activity score

and Cluster score) of the prefractions that share MS features. Inputting the ResistoMAP data into NP Analyst was a new technique for dereplication as it included both unknown NP prefractions and known antimicrobial standards. When analyzing the full network, the prefractions that were shown to cluster with the standards could be readily dereplicated. This process was made easier when using the community visualization tool available for NP Analyst.

The community visualization format of NP Analyst separates the full network into smaller communities that contain highly interconnected nodes indicative of shared biologically active MS features. These communities allow a clear and in-depth analysis of the NP Analyst network as there are fewer nodes to visualize. These individual communities provided this project with a strong basis for NP candidate selection as it aided in the elimination of several options. The full network was divided into 36 communities which were separated into three categories. The first category consisted of 18 separate communities that only contained antimicrobial standards. With the NP Analyst rationale that similar data groups together, these communities indicate that the standards within them do not share attributes with the unknown NP mixtures. The second category was communities that had antimicrobial standards and NP prefractions clustering together. Prefractions and standards will only cluster together if there is shared biologically active metabolites therefore, this category is optimal for the dereplication and deprioritizing of unknown NP prefractions. Three communities within this category deemed 14 NP prefractions as less likely to contain novel components. Specifically, community 4 has RLUS-1878C-E prefractions directly sharing an MS feature with spiramycin as well as 5 other prefractions near several other antimicrobial standards (**Figure 2.5**). Therefore, these NP prefractions were immediately deprioritized for the selection for further research. In theory, if more antimicrobial standards were included in a dataset such as this, this category of communities could lead to the rapid dereplication of numerous NP extracts.

The final category consisting of 14 communities was those that only displayed NP prefractions. These were ideal for candidate selection as they represented NP prefractions that did not share bioactive metabolites with the antimicrobial standards. Before selecting prefractions for further investigation these communities underwent final dereplication with the previously isolated compounds from the Linington library. This addition of previous knowledge led to the dereplication of collismycin A in community 26.

It is predicted if collismycin A standard was added to this dataset, this would have been a category 2 community. Following these final dereplication processes the prefraction candidates for the project's next steps were selected with three requirements in mind: consistent strong activity and/or cluster score, unique connections within the community, and a MS profile with identifiable fragments and limited contaminations with a Rt between 1.5-5.0 minutes to ensure a purified target could be isolated. With these in mind, 18 extracts were selected to be grown for secondary steps (**Table 2.1**).

An ongoing difficulty with NP discovery is the complexity of the original mixtures. To isolate a NP compound, you must start with a mixture containing a mixture of compounds from the raw natural source.[3] NPs are known to exhibit potent biological activities that can display misleading strong bioactivity when present in a mixture. Determining where the specific product of interest within the mixture is extremely difficult therefore, the common technique is to separate NP extracts into 5-6 prefractions (A-F). Instead of having one potent complex mixture now there are separate smaller mixtures that may or may not contain the product of interest. To take it further this project separated these prefractions into peak libraries consisting of 10 subfractions. These peak libraries were separated based on visualized UV peaks to isolate the bioactive components identified in the ResistoMAP profiles. 15 of the 18 extracts in **Table 2.1** were separated into subfractions and underwent a smaller secondary ResistoMAP screening consisting of only 9 drug-resistant strains of *E. coli* and the wildtype. The output profiles proved that the process was successful as the displayed activity was isolated to specific subfractions. RLUS-2085 and RLUS-2024 were potently active in the E subfractions and active in the first few D subfractions making these targets for mass candidate selection. Originally chosen from bioactivity-guided selection and having passed the ResistoMAP filtering, RLUS-2028 was chosen as a candidate for isolation and identification. The subfraction phenotypic profiles displayed consistent activity against Cef6 and Cef7, and strong profiles for the second subfractions of prefractions A, C, and D. RLUS-2204 from the bioactivity-guided selection showed strong activity against Cef6 and Cef7 strains within the D2 and E2 subfractions, making them targets of interest as well. With this secondary screening, another NP Analyst network was constructed to aid in mass candidate selection for isolation. Of the 11 extracts, 7 displayed unique activity and MS data was collected. The NP Analyst output (**Figure 2.8**) separated the extracts into individual communities making visualization of the data

relatively easy. Prior to analyzing this network, another approach was taken to visualize the data using a tool called Selectivity Ratio Profile Network (SRPNT), a wrapper program of the full R-package: Multivariate Pattern Analysis (MVPA). SRPNT was used for the interpretation and visualization of collinear data in terms of predictive association patterns. It was used in this project to identify which MS feature was showing the strongest activity against the secondary ResistoMAP panel relative to all other MS features. It does this by taking the selectivity ratio of the activity of a bioactive MS feature against each drug-resistant *E. coli* strain and produces a bar graph. When run on this project's dataset one variable stood out: 392.27 *m/z* (Rt 3.90 min) found in both the RLUS-2024E and RLUS-2085 E subfractions (**Figure 2.9**). It displayed the highest selectivity ratios against the whole panel relative to all other variables therefore, it was the only candidate selected using this method. Compared to NP Analyst, SRPNT allows a highly specified selection of candidates strictly based on bioactivity.

Returning to NP Analyst as the main selection source, 5 masses were selected for isolation and identification. As mentioned earlier, **Figure 2.8B** outlines the communities the network is divided into. The focus is on mass candidates showing strong activity and cluster scores as well as a reasonable Rt. To confirm the candidate is a strong choice the MS data is checked to deem the mass is present. Lastly, the masses were cross-checked against the NP Atlas as a final dereplication process.[44] With SRPNT highlighting a mass from the E subfractions of both RLUS-2024 and RLUS-2085, the next step was to look at these extracts within the network. Community 1 contained both extracts and the same mass candidate identified from SRPNT, 392.27 *m/z*. Another mass found in some E subfractions of both these extracts, 394.30 *m/z* (Rt 4.30 min), showed a stronger activity and cluster score and was flagged as another candidate for isolation. The MS data for 392.27 *m/z* showed high intensity in subfraction RLUS-2085E-3 making this the subfraction to attempt isolation on (**Figure 2.10)**. MS data identified RLUS-2085E-6 with the highest intensity of 394.30 *m/z* (**Figure 2.11**). NP Atlas matched masses within the range of these two candidates with a family known as prodiginines. These are known to have strong bioactivity and are red in colour, similar to the crude samples of RLUS-2024 and RLUS-2085. Prodiginine standards were not included within the positive control library of ResistoMAP therefore, full isolation and identification were performed on these candidates to confirm these belong to this family of compounds.

The phenotypic profile of RLUS-2028 in the original ResistoMAP output (**Figure 2.14A**) showed strong activity against Cef6 and Cef7 in all prefractions and had consistent activity against all strains in RLUS-2028D. With this display of activity, it was also included in the 384 prefractions that passed the beginning filtering of ResistoMAP data.

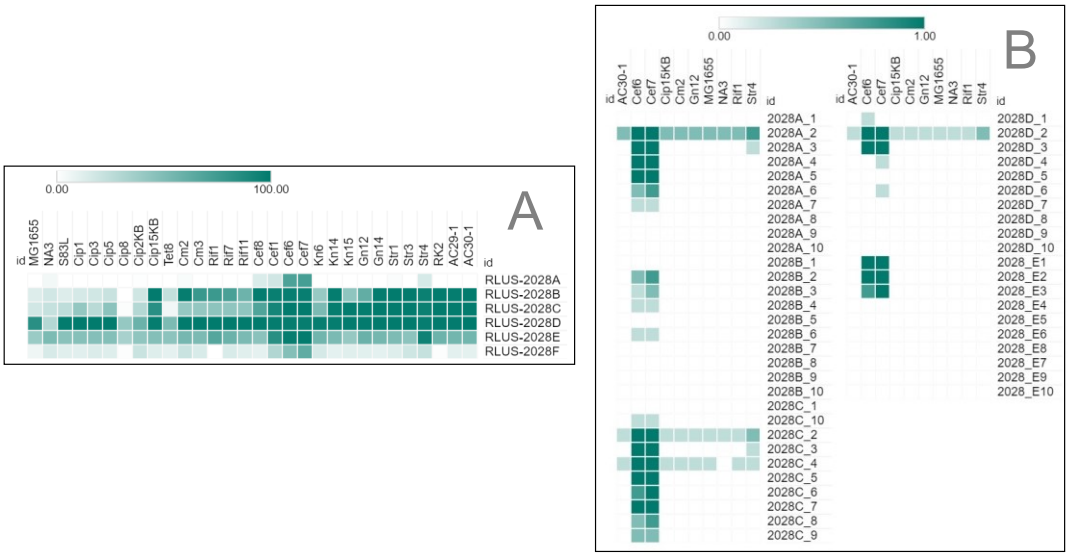


Figure 2.14. RLUS-2028 ResistoMAP phenotypic profile (A). Secondary ResistoMAP screening on peak library subfractions (B).

The secondary ResistoMAP screening on RLUS-2028 subfractions (**Figure 2.14B**) again shows consistently strong activity against Cef6 and Cef7 in the first 3 subfractions of each prefraction as well as an overall strong profile for the second subfractions of RLUS-2028A, C, and D. This same profile is not seen for the second subfractions of RLUS-2028B and E because the separation gradient may have removed the component active against the full panel. Community 2 of the secondary NP Analyst network contains this extract producing a variety of MS features with consistent activity and cluster scores making it difficult to select isolation candidates. When focusing on the second subfractions of RLUS-2028 the features seemed to be similar between the second subfractions of RLUS-2028A-C but not RLUS-2028D and E. Two masses: 438.16 *m/z* (Rt 1.58 min) and 452.16 *m/z* (Rt 1.64 min) were found to have strong activity and cluster scores relative to all of the features seen and were all present in the second subfractions of RLUS-2028A-C. NP Atlas did not match these masses to any known microbial NPs. The MS data for these were of high intensity in RLUS-2028A-2 and RLUS-2028C-2, the two subfractions that showed an overall strong activity profile. The

intensity for both masses was highest in RLUS-2028A-2, making this the source for the isolation of these molecules.

Two extracts, RLUS-2100 and RLUS-2204, did not pass the filtering of the ResistoMAP data but were included in secondary steps as bioactivity-guided selections. RLUS-2204 specifically showed a unique ResistoMAP phenotypic output with potent activity against Cef6 and Cef7 in all prefractions and low activity in the D prefraction.

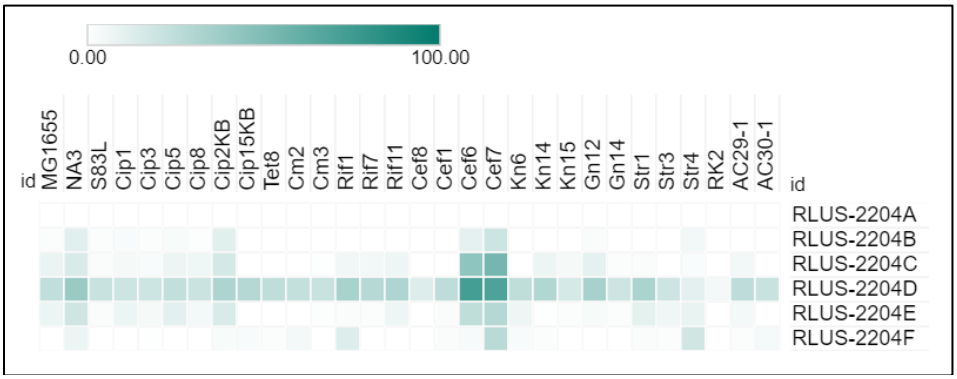


Figure 2.15. ResistoMAP output for RLUS-2204.

Examining the communities of the secondary NP Analyst network, community 3 consists of extract RLUS-2204 with a set of features with high cluster scores and similar activity scores. A group of masses all within 14-18 mass units of one another (**Table 2.2**) were identified to have consistent activity scores and strong cluster scores. Within this group, the predominant molecule was 552.40 *m/z* (Rt 3.19), which displayed an MS profile of high intensity within RLUS-2204D-3 (**Figure 2.13**). NP Atlas identified this mass as being similar to that of a macrolide identified by Kinumaki *et al.* published in 1977.[54] This study had identified 6 fragment ions in the M-4365 component all within a few mass units of one another. With 552.40 *m/z* being the main target these masses were selected for isolation and identification.

The processes outlined within this chapter assisted in the dereplication and prioritization of NP prefractions for the identification of a novel mass candidate. The combination of ResistoMAP and NP Analyst offers opportunities that can be applied to NP discovery with alternative datasets than those used in this project. It utilizes a selection technique that targets potential novel candidates by dereplicating with two processes: comparing phenotypic profiles with a positive control library of commercial standards in ResistoMAP and grouping similar chemistry based on shared bioactive

metabolites with NP Analyst. These processes can be applied with different positive controls or bioassay data for an alternative targeted approach to NP discovery.

2.4. Materials and Methods

2.4.1. ResistoMAP

ResistoMAP design and analysis were performed according to the procedure outlined by Liu *et al.*[47] Raw plate reader values (via optical density of 600 nm) of 80 commercial antimicrobial compounds and 6,195 NP extract prefractions against wildtype and 29 drug-resistant strains of *E. coli* were obtained. In brief, t_0 values (absorbance reading of growth when bacteria is initially added to assay plate) were subtracted from corresponding t_{20} values (absorbance reading of growth 20 hours after t_0 was measured) for each sample, then divided by the difference between the positive and negative controls for normalization (**Equation 2.1**). These resulting percent growth values were inversed by subtracting from 100 to produce percent activity values. These values for each extract were standardized across all extracts for each *E. coli* strain to minimize plate-to-plate variation. For the commercial antimicrobial dataset, percent activity values at each one of 16 concentrations, against each strain, were averaged across three replicates.

Equation 2.1.

$$Percent\ Growth\ (\%) = \frac{x - \mu_{blank\ control}}{\mu_{growth\ control} - \mu_{blank\ control}} \times 100$$

For this project, the complete ResistoMAP dataset underwent filtering to remove all instances whereby the percent activity values were greater than 50% against 5 or more resistant strains of *E. coli*. This trimmed the dataset from 6,195 to 384 NP prefraction profiles.

2.4.2. Mass Spectrometry-Based Metabolomics Data Acquisition

These 384 NP prefractions were collected from Linington Lab marine storage plates stored in a -70 ˚C freezer. Four 96-well conical polypropylene storage plates had every well filled with 195 µL of 1:1 MeOH: H_2O (Waters Optima-Grade MeOH; MilliQ

Filtered dH_2O) using INTEGRA VIAFLO 96 semi-automated 96-channel liquid handler. For each NP prefraction, 5 µL was dispensed from its source deep-well plate to this dilution plate. The dilution plates were then sealed using a plate sealer and aluminum seal (Agilent PlateLoc Thermal Microplate Sealer) and sonicated for 5 minutes to ensure homogeneity. Once sonicated the plates were unsealed and 1 µL of every prefraction was transferred to the assay plate for a final dilution of 1:20,000 with dimethyl sulfoxide (DMSO). These plates were then sealed, sonicated, and centrifuged. A similar process was used to obtain MS data on 73 antimicrobial standards. Four assay plates were prepared, each at a different concentration: 10, 100, 10,000, and 100,000 mg/mL.

Four NP extract assay plates were analyzed using UPLC-ESI-qTOF MS according to the following parameters. Columns 1 and 12 of every plate contained solvent blanks and columns 2 to 11 contained up to 96 diluted extract prefractions. Solvent blanks were included after every 15 extracts. Chromatographic separation was performed with an Acquity I-Class UPLC (Waters) using an Acquity HSS T3 C_{18} column (100 mm × 2.1 mm, 1.8 µm, Waters). Separation used mobile phase (A) as H_2O + 0.1% formic acid (FA) to (B) MeCN + 0.1% FA at a flow rate of 500 µL/min and for 12.8 min (5% MeCN, 0-0.3 min; 5-90% MeCN, 0.3-9.1 min, 90-98% MeCN, 9.1-10.7 min, 98% MeCN, 10.7-11 min, 5% MeCN, 11.01-12.8 min). MS data were acquired on a SYNAPT G2-Si hybrid quadrupole-traveling wave ion mobility (TWIM) time-of-flight (TOF) (Waters) mass spectrometer with an electrospray ionization (ESI) source. The instrument was operated in positive ion mode and conducted using data-independent acquisition (DIA) mode. Settings were as follows: voltage 3.0 kV; cone voltage 40 V; source offset 50V; source temperature 150 ˚C; desolvation temperature 300 ˚C; cone gas flow 30 L/hr; desolvation gas flow 600 L/hr. Detection was acquired in the *m/z* range 50 – 1500 with a scan rate of 0.25 Hz in both MS and MS^2 modes. Leucine enkephalin (Waters) lockspray solution was enabled at a concentration of 200 pg/µL at 0.10 Hz.

2.4.3. NP Analyst Network Production and Community Visualization

Raw MS files produced by the mass spectrometer were converted to .mzml format using the open-source tool MSConvert. Both the MS files in .mzml format and the trimmed ResistoMAP screening data in .csv format were simultaneously uploaded onto the NP Analyst offline web portal. NP Analyst networking was performed according to the following parameters: ppm error 30 *m/z*; Retention Time Window: 0.03 min; Minimum

Intensity: 20,000; Minimum Frequency in Replicates for Replicate Comparison: 2; Minimum Frequency in Samples for Feature Alignment: 1; Activity Score Threshold: 2.03; Cluster Score Threshold: 0.3. For explanation of these parameters, see the corresponding publication by Lee *et al.*[48] Output NP Analyst network file was imported into Cytoscape 3.8.2 for visualization and analysis.

Output NP Analyst community network files were separately examined in Cytoscape 3.8.2. The ResistoMAP assay .csv output was visualized using Morpheus (https://software.broadinstitute.org/morpheus/). The table .csv output was visualized using Tableau. The overall network was divided into 36 communities which were divided into 3 different categories upon observations: communities containing only antimicrobial standards, communities containing NP prefractions and antimicrobial standards, and communities only containing NP prefractions.

Communities Category 1 – Only Antimicrobial Standards

Of the 36 communities, 18 were placed into this category. Filtering through these communities identified numerous antimicrobial standards that do not share characteristics with the NP prefractions within the dataset.

Communities Category 2 – NP Prefractions and Antimicrobial Standards

Only 3 communities were in this category for theoretical dereplication. Community 3 dereplicated the prefractions RLUS-1512B, RLUS-1520B, and RLUS-1520C with standards clarithromycin, erythromycin, puromycin, and roxithromycin. Prefraction RLUS-1512B shared mass 559.38 *m/z* (Rt 2.80 min) directly with Erythromycin indicating shared chemistry between the two. Community 4 dereplicated RLUS-1478D, RLUS-1492D, RLUS-1503A, D, & E, and RLUS-1878C-E with Amphotericin-B, Cephadroxil, Chloramphenicol, D-Cycloserine, Midecamycin, Rifabutin, Sparfloxacin, Spiramycin, Sulfadazine, and Tiamulin. The RLUS-1878 prefractions all shared mass 430.25 *m/z* (Rt 2.14 min) directly with Spiramycin. Community 13 dereplicated RLUS-2064C-E with Clindamycin, Mithramycin, and Minocycline. RLUS-2064D shares 425.1926 *m/z* (Rt 2.22 min) with Clindamycin. Source microorganisms of

extracts RLUS-1520 and RLUS-2064 were selected for full fermentation and grow-up, the other extracts were not selected for future steps.

Communities Category 3 – NP Prefractions

The remaining 15 communities belong to this category. Each community was analyzed and compared to a list of previously isolated compounds identified within the Linington library. Community 26 containing RLUS-2108 was dereplicated with collismycin A (in-source mass fragment = 244.06 *m/z*, Rt =2.64 min) and related analogue SF2738D ($[M+H]^+$ = 258.07, Rt = 3.43 min). Some remaining communities were loosely dereplicated with previously identified compounds, and extracts displaced from those that were dereplicated were selected for full fermentation and grow up. Communities 29, 32, and 34 were not included in the NP prefractions selection process as they contained extracts that had been previously researched. **Table 2.1** contains all extracts selected from the NP Analyst network with **Table A.1** identifying the communities and the masses that differentiated each one.

2.4.4. Fermentation Process for the 18 Extracts

Frozen stocks (1:1 glycerol: starch-yeast-peptone SYP media at -70 ˚C) of the bacterial isolates were struck onto marine broth agar (MB agar; per 1L dH_2O: DIF-COTM Marine Broth, 37.4g; 15.0g) at room temperature for up to 5 days. Individual colonies were used to inoculate 7mL of SYP liquid media (SYP; per 1L dH_2O: Instant Ocean, 31.2g; Soluble Starch, 10g; Yeast Extract, 4g; Peptone, 2g) with 3 sterile glass beads, shaking at room temperature (RT) at 200 RPM for 4-7days. Upon sufficient growth 3mL of this liquid culture was used to inoculate 60mL of SYP media within a sterile Erlenmeyer flask with a coiled stainless-steel spring, agitated at RT, 200 RPM for 5-7 days. 45mL of this culture was then used to inoculate 1L of SYP media within a sterile 2.8L Fernbach flask containing a larger coiled stainless-steel spring and 20g of XAD-17 resin (Thermo Fisher), shaking at RT, 200 RPM for 5-7days. After this period of fermentation, the large-scale culture was filtered with Whatman paper inside a ceramic funnel. The remaining residue was extracted with DCM/MeOH (1:1, 250mL) three times while stirring. The resulting suspension was filtered again to remove cellular debris and torn filter paper. Celite resin (20g) was added to the filtrate and evaporated to dryness in vacuo to ensure binding of extract to resin. The extract-adhered resin was packed into a

reusable CombiFlash prep column with a resin filter on either end and compacted. An elutropic series of MeOH/H_2O via CombiFlash separated each extract into 5 prefractions. Prefractions A (20% MeOH/H_2O), B (40% MeOH/H_2O), C (60% MeOH/H_2O), D (80% MeOH/H_2O), and E (100% MeOH/H_2O) were evaporated to dryness in vacuo and resuspended in minimal MeOH/H_2O (1:1) for optimal collection. Each prefraction was stored at -70 ˚C and prepared for further separation.

2.4.5. Peak Library Separation for 15 Extracts

For 15 of the 18 extracts (**Table 2.3**), each prefraction was diluted to a concentration of 50 µg/mL and centrifuged. 40 µL of the supernatant was stored while the remaining volume was transferred to 1.5 mL LC-MS total recovery vials and separated via HPLC (Agilent 1200 series; Phenomonex Synergi™ 10 µm Fusion-RP 80 A 250x10 mm) with customized gradients into peak libraries containing 10 subfractions. Subfractions were evaporated to dryness in vacuo and resuspended in minimal MeOH/H_2O (1:1) for optimal collection. Fully dried peak libraries were stored at -70 ˚C.

Table 2.3. List of extracts that underwent prefraction separation into subfractions.

Extracts	
RLUS-2045	RLUS-2024
RLUS-2100	RLUS-1806
RLUS-2028	RLUS-1885
RLUS-2085	RLUS-1530
RLUS-1758	RLUS-1520
RLUS-1597	RLUS-1726
RLUS-1595	RLUS-1623
RLUS-2204 *only C-E prefractions	

2.4.6. Secondary ResistoMAP Screening and Data Analysis

A secondary smaller assay was derived from the original ResistoMAP screening containing 9 of the 29 drug-resistant strains and wildtype of *E. coli* and excluding the positive control library of antibiotics. The same procedure outlined by Liu *et al* was performed.[47] In brief, these 10 strains; MG1655, Rif1, NA3, Cef6, Cef7, Cip15KB, Cm2, Gn12, Str4, and AC30-1 were inoculated from glycerol stocks into 3 mL overnight

cultures in Luria-Bertani (LB) media (Fisher), grown at 37 ˚C, 200 RPM. Saturated overnight cultures were diluted in LB media according to turbidity to achieve - 5.0 x 10^5 CFU/mL of final inoculum density and dispensed (Matrix WellMate) into sterile polystyrene 384-well microplates (Thermo ScientificTM 265202) with a final screening volume of 30 µL. Depending on the mass of subfraction they were prepared in a 4-step 1:1 dilution series from 30 µg/mL. If there was not sufficient mass, then the highest concentration possible was used. These solutions were then pinned into each respective assay plate (200 nL) using a high throughput pinning robot (Tecan Freedom EVO 100; V&P Scientific pin tool). In each 384-well assay plate, columns 1 and 24 were reserved for blank control (DMSO vehicle, LB media) while columns 2 and 23 were reserved for growth control (DMSO vehicle; LB media; target bacteria). After compound pinning, assay plates were read using a plate reader (BioTek Synergy Neo2 running GEN5 software) to obtain OD_{600} absorbance values at t_0, sealed with lids, and placed in a humidity-controlled incubator (Thermo Cytomat) at 37 ˚C, 5 % CO_2 for 20h. Post-incubation OD_{600} readings were taken at t_{20}. Post-data acquisition, absorbance values were normalized using the same process outlined in section **2.4.1**. After initial normalization, the data was organized to give a value of 1 if the inhibition is greater than 50% at the lowest concentration and down to zero for the highest concentration (0.75 at 7.5, 0.5 at 15, and 0.25 at 30 µg/mL).

2.4.7. Secondary NP Analyst Network

A total of 53 subfractions showed distinct activity and were profiled by UPLC-HRMS using a Waters SYNAPT G2Si ESI-qTOF-MS system running MassLynx software. Also, LC-MS analyses using an Agilent ESI-quadrupole-MS instrument running ChemStation were completed. An additional 150 inactive subfractions were profiled by UPLC-HRMS. Subfractions were dissolved in DMSO at a concentration of 7.5 µg/mL and placed in three parent 96-well conical polypropylene plates. Using INTEGRA VIAFLO 96 semi-automated 96-channel liquid handler 49 µL of 1:1 MeOH: H_2O (Waters Optima-Grade MeOH; MilliQ Filtered dH_2O) was transferred to designated wells in 384-well pyramidal well polypropylene plate. Again, using the ViaFLO 96 with a fresh set of tips, 1 µL of each NP subfraction was pipetted from its parent 96-well plate to its destination dilution plate. The dilution plates were then sealed using a plate sealer and aluminum seal (Agilent PlateLoc Thermal Microplate Sealer) and sonicated for 5 minutes

to ensure homogeneity of the samples. The plates were then centrifuged using a Thermo SpeedVac Vacuum Concentrator with Microplate attachments. The same parameters outlined in section **2.4.2** were used except IMS mode was not employed.

Raw MS files were converted to .mzml format using the open-source tool MSConvert. MS2Analyte, an in-house MS processing tool, was used for peak-picking of MS data. The following parameters: Manufacturer: Waters; File Type: mzML; Experiment Details: Blanks, Replicate Count: 3, MS2 data, and DIA; MS1 Threshold: 2000; Mass Error (Da): 0.01; Max Peak Count: 10,000; RT Error (min): 0.1. The corresponding .csv assay file and MZmine output .csv file produced from MS2Analyte were used to produce a second NP Analyst network using the online platform (https://www.npanalyst.org/) using an Activity Score of 0.6 and Cluster Score of 0.32. Using the community visualization the network was divided into 7 communities (**Figure 2.8B**). Of the 53 active subfractions, 4 were chosen for isolating and identification processes.

SRPNT/MVPA Visualization

Additionally, this activity and HRMS data were inputted into SRPNT to produce a network identifying the strength of the candidate's bioactive profiles. MS2Analyte MZmine output file was edited to follow SRPNT format: *m/z*mass_retentiontime. The same assay .csv file and this edited MS2Analyte .csv file produced a graphml. output. The procedure followed to run SRPNT is outlined by Baumeister *et al.* (manuscript in preparation) and instructions can be found here: https://github.com/liningtonlab/SRPNT. The graphml. output file was opened in Cytoscape 3.8.2 for visualization. These were the following parameters used: feature transformation method: seed: 42; introduce compound interaction terms: true; feature count limit: 50; number repetitions: 50; percent calibration samples: 50; validation threshold: 0.5; number components: 10; use monte carlo resamples: true; standardize: false; score filter method: f-value; alpha: 0.05; score threshold: 1.0; direction: both; distance metric: Euclidean; edge threshold: 0.85; perform community detection: true. In Cytoscape, using Image/Chart 1, bar graph was selected for all selectivity ratios producing a bar graph visualization for these ratios. The output was strongest for feature 392.27 *m/z* (Rt 3.90 min), showing selectivity ratios, in order from left to right seen in the bar graph: AC30-1: 2.23; Cef6: 0.67; Cef7: 0.58; Cip15KB: 2.53; Cm2: 6.50; Gn12: 2.28; MG1655: 21.20; NA3: 2.38; Rif1 3.01; Str4: 3.01.

Community 1 – RLUS-2085E-3

This community was identified to contain subfractions from extracts RLUS-2024, and RLUS-2085, and only one node from both RLUS-1726 and RLUS-1595. RLUS-2085 was the major presence, with RLUS-2085E being flagged by MVPA, this subfraction was the main focus for this community. The mass 392.27 *m/z* found in MVPA was present in this community, with the same mass and retention time (3.90 mins.), an Activity score of 3.59, and a Cluster score of 0.32. It was present in RLUS-2024E-2 to -9, RLUS-2085D-3, and RLUS-2085E-3 to -9 subfractions. MS data from MS2Analyte identified this feature as an intensity of 3.3×10^6 in subfraction RLUS-2085E-3.

Another feature 394.30 *m/z* with a Rt of 4.30 minutes had a higher Activity score of 3.88 and Cluster score of 0.66. The network identified its presence in RLUS-2024E-4, 5, & 9 and RLUS-2085E-6 and -8. MS data identified this feature as an intensity of 3.9×10^6 in subfraction RLUS-2085E-6.

Using the Basic Search page in NP Atlas (https://www.npatlas.org/search/basic), the mass ranges and origin type of bacteria were searched against the database. Feature 392.27 *m/z* matched with butylcyclohexylprodigiosin (exact mass: 391.2624 Da) and metacycloprodigiosin (exact mass: 391.2624) both with formulas $C_{25}H_{33}N_3O$. Mass 394.30 *m/z* matched with prodigiosin 25-C (exact mass: 393.2780 Da) and undecylprodigiosin (exact mass: 393.2780 Da) both with formulas $C_{25}H_{35}N_3O$.

Community 2 – RLUS-2028A-2

This community was identified to contain subfractions from extracts RLUS-2028 and only one node from both RLUS-1520 and RLUS-1595. Feature 438.16 *m/z* with a Rt of 1.58 minutes has an Activity Score of 1.56 and Cluster Score of 0.95, it is present in RLUS-2028A-2 to -4, RLUS-2028B-2 and -3, and RLUS-2028C-2. Feature 452.18 *m/z* with a Rt of 1.64 minutes has an Activity Score of 2.03 and Cluster Score of 0.96, it is present in RLUS-2028A-2 and -3, RLUS-2028B-2, and RLUS-2028C-2. Checking the second subfractions of these prefractions the MS2Analyte data identified both these features with an intensity of 1.4×10^6 in RLUS-2028A-2. Using the same process outlined above, NP Atlas did not display any suitable matches.

Community 3 – RLUS-2204D-3

This community was identified to contain subfractions from extracts RLUS-2204. These features 538.38 (Rt 3.10 min), 552.40 (Rt 3.19 min), 568.39 (Rt 3.35 min), 582.36 (Rt 3.06 min), 598.36 (Rt 2.27 min) *m/z* were identified to have consistent activity scores and strong cluster scores. MS2Analyte data identified 552.40 *m/z* as having the highest intensity within RLUS-2204D-3 which also contained multiple other masses at a reasonable intensity. Using the same process outlined above, NP Atlas matched 552.40 *m/z* with a compound known as M-4365 G1 (exact mass: 551.3822) with formula $C_{31}H_{53}NO_7$.

Chapter 3.

Isolation and Structure Elucidation

3.1. Introduction

This chapter describes the isolation, identification, and bioactive characterization of three candidates identified from the ResistoMAP and NP Analyst selection pipeline discussed in **Chapter 2**.

Extract prefraction RLUS-2085E was selected for its hit subfractions RLUS-2085E-3 and RLUS-2085E-6 which displayed activity against a range of drug-resistant *E. coli* strains. Features prioritized by MVPA and NP Analyst led to the selection of two molecules, 392.27 *m/z* and 394.30 *m/z* for isolation and full structure elucidation. NP Atlas suggested these masses as members of the prodiginine family. The prodiginine family is comprised of bacterially produced natural products (NP) that share a characteristic tripyrrolic core (**Figure 3.1**).[55–57] These red-pigmented compounds display an array of attractive biological properties including antibacterial, anticancer, and immunosuppressive activity.[58–60] Various *Serratia* and *Streptomyces* strains have been responsible for the production of these compounds and recently, known prodiginines have been isolated from marine-sourced actinomycetota.[55,57,61] Two compounds of the prodiginine family, streptorubin B (**2**) and undecylprodigiosin (**4**) have been identified from extract RLUS-2085E and will be fully elucidated within this chapter.

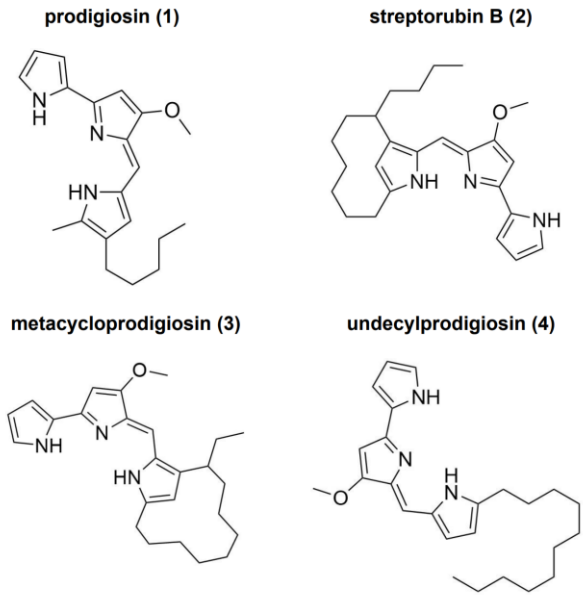


Figure 3.1. Structures of natural prodiginine compounds.

Additionally, extract prefraction RLUS-2028A was selected for its distinct activity against the ceftazidime-resistant *E. coli* identified in RLUS-2028A-2. Analogues 438.2 *m/z* (**X1**) and 452.2 *m/z* (**X2**) were flagged by NP Analyst however, isolation difficulties prevented full purification and elucidation. Lastly, RLUS-2204D displayed weak activity against many drug-resistant *E. coli* strains except for the ceftazidime-resistant strains (Cef-6 and Cef-7). NP Analyst identified five features within a few mass units of one another as displaying consistent strong Activity and Cluster scores. MS data identified feature, 552.40 *m/z* (**X3**) as having the highest intensity. Using this mass as guidance, NP Atlas suggested the macrolide antibiotic, M-4365, as this molecule. This macrolide was identified by Kinumaki and co- from *Micromonospora* along with six other macrolide components within a few mass units of one another, similar to those identified in NP Analyst.[54] Macrolides have been previously identified as useful antibiotics for the treatment of bacterial infections.[62] These compounds are known to inhibit bacterial protein synthesis and are characterized by a large macrocyclic lactone ring attached to one or more amino-deoxy sugars.[62] Popular macrolide antibiotics include erythromycin, azithromycin, and clarithromycin. Based on prominence, analogue 552.40 *m/z* (**X3**) was selected as the main target for isolation and is undergoing further studies.

Concluding screening was performed on each of these isolated candidates to assess the biological profiles of the purified molecules and compare to the observed activity of the secondary ResistoMAP profiles.

3.2. Results

3.2.1. Isolation of features 392.27 *m/z* and 394.30 *m/z* from RLUS-2085E

Both the original and secondary ResistoMAP screening, along with MVPA and NP Analyst procedures highlighted prefraction RLUS-2085E for its potent activity profiles. This prefraction showed activity against all drug-resistant strains, with the strongest against ceftazidime and streptomycin-resistant *E. coli*. Secondary screening revealed that subfractions RLUS-2085E-3 and RLUS-2085E-6 had similar profiles to the original ResistoMAP dataset. Activity was observed for RLUS-2085E-3 against all 10 drug-resistant strains at the concentration of 3.75 µg/mL. Similarly, RLUS-2085E-6 showed activity at this concentration against all strains excluding ciprofloxacin, chloramphenicol-resistant strains, and the wildtype, MG1655. Activity against these three strains was observed at 7.5 µg/mL. As stated in the previous chapter, MS features 392.27 *m/z* (Rt 3.90 min) from RLUS-2085E-3 and 394.30 *m/z* (Rt 4.29 min) from RLUS-2085E-6 were selected for isolation.

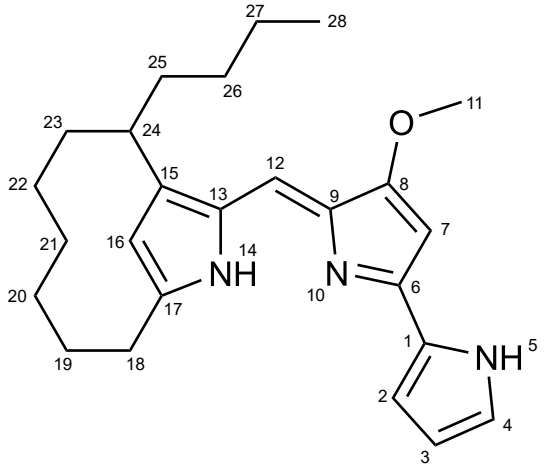
Isolation attempts were made on RLUS-2085E-3 for 392.27 *m/z*, 1.42 mg was obtained. Purity was checked by mass spectrometry (MS) and ^{1}H NMR. Both showed signs of impurity therefore, further purification was performed on the full 1.42 mg. However, not enough pure sample (below 1.0mg) was collected to obtain workable nuclear magnetic resonance (NMR) spectra for elucidation processes. The extract was re-grown through fermentation and isolation attempts were performed on the prefraction RLUS-2085E rather than the subfraction. Through this process, both 392.27 *m/z* and 394.30 *m/z* were observed and isolated simultaneously using distinct UV signatures known to be associated with prodiginines (392.27 *m/z,* 7.81 mg; 394.30 *m/z*, 7.29 mg). Purities were confirmed using ^{1}H NMR and full 2D NMR datasets were obtained for elucidation (**Appendix Figures B.2-16**)

3.2.2. Structure Elucidation of Streptorubin B (2)

Compound **2** was isolated as an amorphous red solid. The molecular formula of the metabolite was determined through HRMS as $C_{25}H_{34}N_3O$ with the monoisotopic mass of the $[M+H]^+$ adduct: obs 392.2705 *m/z*. The UV spectrum of this compound displayed an absorbance of 545 nm, characteristic of a prodiginine's red colour.[63] The planar structure of **2** was elucidated using this information and through NMR analysis using a combination of ^1H, ^{13}C, gCOSY, gHSQC, gHMBC, ^{15}N HSQC, and ^{15}N HMBC spectra in chloroform-d (**Table 3.1**; **Appendix Figures B.2-8**).

The experimental ^1H NMR spectrum contains a characteristic signal at a chemical shift of δ_H -1.55, which matches the literature published by Weyland and coworkers.[64] With a combined comparison of chemical shifts reported by Furstner *et al* and Weyland, the experimental ^1H and ^{13}C NMR were reflective of the values corresponding to those of **2**.[59,64] Carbon signals ranging from δ_C 40.0-20.0 correspond to ten methylene carbons, while furthest upfield, δ_C 14.3 represents a methyl group. The characteristic prodiginine methoxy can be seen at δ_C 58.7 (**Table 3.1**). Six signals represent sp^2 hybridized quaternary carbons leaving the remaining signals as five aromatic CHs, one olefinic CH, and one tertiary CH. Examination of the ^1H NMR spectrum confirms the presence of ~35 protons: 3 NHs, six aromatic and olefinic, and 26 aliphatic protons. Full 2D spectra were used to confirm structural connectivity.

Table 3.1. Tabulated NMR data for 2 isolated from RLUS-2085E. All spectra were acquired in chloroform-d4 at 600 MHz (^1H), 150 MHz (^{13}C), 60.80 MHz (^{15}N).



Position	Experimental Data			Literature Data[64]	
	δ_H (*J* in Hz)	δ_C	δ_N	δ_H	δ_C
1		122.7			122.4
2	6.91, s	116.7		6.90	116.6
3	6.34, d	111.6		6.34	111.6
4	7.25, overlap	127.2		7.22	126.7
5	12.72, s		160.4	12.71	
6		145.3			147.1
7	6.09, s	93.3		6.10	92.7
8		165.9			165.5
9		121.0			120.3
10	12.60, s		139.3	12.58	
11	4.02, s	58.7		4.02	58.6
12	7.11, d	112.7		7.12	112.5
13		125.4			125.1
14	12.70, s		162.3	N/A	
15		150.8			150.6
16	6.50, s	116.8		6.52	116.8
17		154.8			154.7
18	3.33, 2.56, overlap	30.2		3.34, 2.54	29.9
19	1.91,1.25, overlap	31.8		1.98~1.89, 1.42~1.32	31.6
20	1.77, 0.80, overlap	29.4		1.82~1.76, 0.83~0.78	29.1
21	1.15, -1.56, overlap	27.7		1.89~1.11, -1.55	27.6
22	1.56, 0.92, overlap	25.8		1.70~1.50, 0.97~0.90	25.4
23	1.85, 1.56, overlap	31.2		1.89~1.82, 1.70~1.50	30.9
24	3.11, s	37.5		3.10	37.3
25	1.72, 1.15, overlap	39.0		1.85~1.11, 1.76~1.69	38.9
26	~1.26	29.9		1.42~1.32	30.6
27	~1.38, overlap	23.1		1.42~1.32	22.8
28	0.92	14.3		0.97~0.90	14.1

There are two NH groups within **2** however, protonation results in three separate NH correlations within the ^{15}N HMBC due to presence of formic acid seen at δ_H 8.90. The ^{1}H NMR contains two broad signals in the δ_H 12.0 region which could be three overlapping singlets. These can be differentiated using the combined efforts of ^{15}N HMBC and literature references (**Figure 3.1** (blue arrows)). ^{15}N HMBC spectrum contains a nitrogen signal at δ_N 139.3 showing correlations to δ_H 7.11 and 6.09. Literature states the proton that corresponds to this nitrogen would have an upfield signal around 12.0 therefore, N_{10} δ_N 139.3 corresponds to the broad singlet at δ_H

12.60.[57,64] Indeed, a faint ^{15}N HSQC signal can be seen between these signals. Two other nitrogen signals can be seen in the ^{15}N HMBC, δ_N 160.4 to δ_H 6.91 and δ_N 162.3 to δ_H 7.11. This identifies δ_N 162.3 as closest to N_{10} as both display correlations to δ_H 7.11. Based on the literature, the corresponding proton for signal δ_N 160.4 will be the furthest downfield signal at δ_H 12.72, leaving δ_H 12.70 with N_{14} δ_N 162.3. With this information, the three pyrrole rings can be identified with gCOSY and gHMBC correlations.

The methyl groups are easy to identify with gHSQC correlations seen between C_{28}, δ_C 14.3 and proton δ_H 0.92, and C_{11} methoxy at δ_C 58.7 with the strong singlet at δ_H 4.02. The gCOSY spectrum was hard to analyze due to weak signals. However, proton δ_H 6.34 shows coupling to δ_H 6.91 and 7.25 to which gHMBC correlations to the corresponding carbons could be observed as well (**Figure 3.1**). As stated above with ^{15}N HMBC signals N_5 was identified as δ_N 160.4 with δ_H 6.91. Its final connection, the quaternary carbon, is also identified from gHMBC between δ_H 6.91 and δ_C 122.7 completing this first spin system (**Figure 3.1**). No gCOSY coupling was observed for the methoxy at C_{11} however, a gHMBC correlation is seen between its methyl protons and C_8, δ_C 165.9. To complete the second pyrrole spin system as stated above N_{10}, δ_N 139.3 shows ^{15}N HMBC correlations to δ_H 6.09 and δ_H 7.11.

Several gHMBC correlations are observed from δ_H 6.50 to δ_C 125.4, 150.8, and 154.8 (**Figure 3.1**). Correlations from gHMBC signals can be seen from δ_H 7.11 to δ_C 150.8 as well as to N_{10}, δ_N 162.3 indicating this is the second pyrrole connected to this olefin. Very faint gHMBC correlations can be seen between C_{24} δ_H 3.11 and C_{16}, δ_C 116.8, and C_{25} δ_H 1.72 to C_{15}, δ_C 150.8 indicating the aliphatic chain connection. Continual gHMBC and gCOSY correlations can be seen between proton to carbon allowing the construction and differentiation of the diastereotopic protons within the aliphatic chain and ring (**Figure 3.1**).

This full elucidation was compared to literature of Furstner *et al* and Weyland.[59,64] The experimental and literature ^1H and ^{13}C signals were reflective of one another. HSQC and HMBC correlations were consistent with the structure of streptorubin B thus, the complete planar structure of **2** was established.

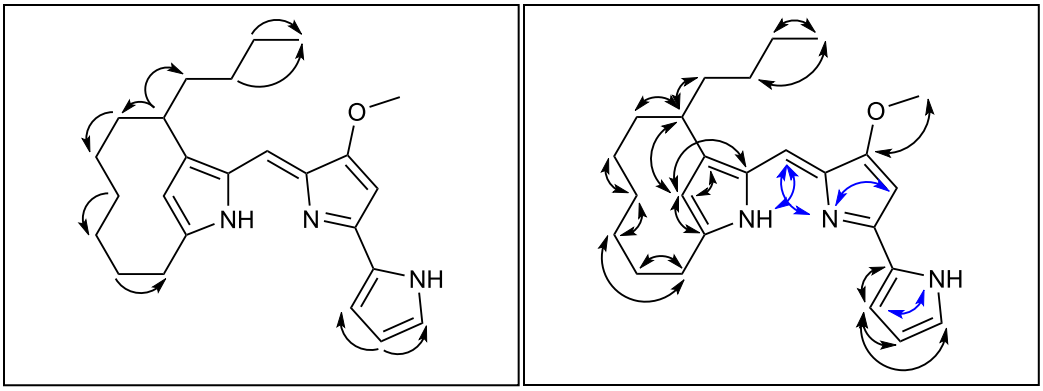


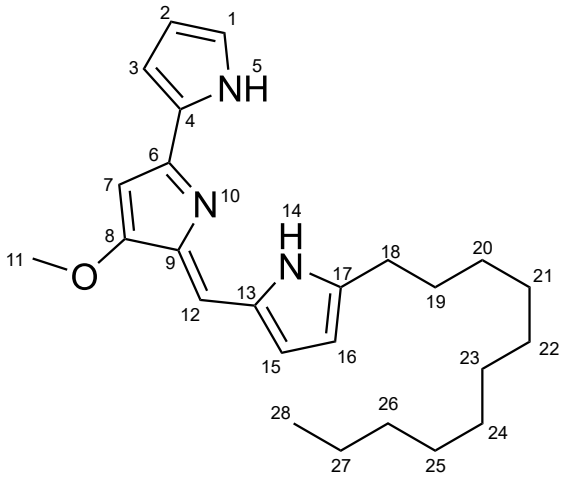
Figure 3.2. Key 2D NMR correlations of 2 in $CDCl_3$ at 600 MHz for ^{1}H and ^{13}C. COSY (one-headed arrows), HMBC (double-headed arrows), and N^{15} HMBC (blue double-headed arrows) correlations are shown.

3.2.3. Structure Elucidation of Undecylprodigiosin (4)

Compound **4** was isolated as an amorphous pink solid. The molecular formula of the metabolite was determined through HRMS as $C_{25}H_{35}N_3O$ with the monoisotopic mass for the $[M+H]^+$ adduct: obs 394.2881 *m/z*. The UV spectrum of this compound displayed an absorbance of 525 nm, which is characteristic of a prodiginine's red colour.[63] The planar structure of **4** was confirmed using this information and through NMR analysis using a combination of ^{1}H, ^{13}C, gCOSY, gHSQC, gHMBC, ^{15}N HSQC, and ^{15}N HMBC spectra in chloroform-d (**Table 3.2**; **Appendix Figures B.10-16**).

The comparison of ^{1}H and ^{13}C NMR between literature published by Papireddy *et al.* with experimental spectra obtained showed similar correlations in the relative shifts for key signals.[58] Examination of the experimental spectra confirmed the presence of 26 carbon signals, six aromatic CHs, one olefinic CH, and six sp^2 hybridized quaternary carbons: one methoxy, one methyl, and ten methylene carbons. Signals from δ_C 32.1-22.8 correspond to the ten-member aliphatic chain of C_{18}-C_{27}. The methoxy group of C_{11} is distinctly seen at δ_C 59.0, while the methyl group of C_{28} is found furthest downfield at δ_C 14.3 (**Table 3.2**). The ^{1}H NMR spectrum confirmed the presence of ~36 protons: 3 NHs, seven aromatic and olefinic, and 26 aliphatic protons. Determining the full assignment of ^{1}H NMR resonances was completed using 2D NMR.

Table 3.2. Tabulated NMR data for 4 isolated from RLUS-2085E. All spectra were acquired in chloroform-d at 600 MHz (^{1}H), 150 MHz (^{13}C), 60.80 MHz (^{15}N). *Peak values were estimated from spectra provided by Papireddy *et al.*



Positions	Experimental Data			Literature Data[58]	
	δ_H (*J* in Hz)	δ_C	δ_N	δ_H (*J* in Hz)*	δ_C*
1	7.25, overlap	127.8		7.30	127.6
2	6.37, q	112.2		6.40	112.1
3	6.96, m	118.0		6.90	117.8
4		122.3			122.4
5	12.69, s		161.0	12.62	
6		149.0			148.8
7	6.09, d	93.2		6.10	93.0
8		166.4			166.3
9		121.6			121.2
10	12.93, s		139.5	12.90	
11	4.03, s	59.0		4.00	58.8
12	7.02, s	116.6		7.00	117.7
13		126.2			126.0
14	12.74, s		164.7	12.70	
15	6.85, t	129.5		6.85	129.3
16	6.21, dd	112.7		6.30	112.6
17		153.4			153.3
18	2.95, t	28.6		2.90	28.0~29.0
19	1.77, qu	29.5		1.20~1.90	28.0~29.0
20	1.40, qu	29.6		1.20~1.90	28.0~29.0
21	1.26, overlap	29.5		1.20~1.90	28.0~29.0
22	1.25, overlap	29.8		1.20~1.90	28.0~29.0

23	1.25, overlap	29.4	1.20~1.90	28.0~29.0
24	1.25, overlap	29.8	1.20~1.90	28.0~29.0
25	1.33, overlap	29.7	1.20~1.90	28.0~29.0
26	1.24, overlap	32.1	1.20~1.90	31.0
27	1.28, overlap	22.8	1.20~1.90	22.7
28	0.87, t	14.3	0.85	14.1

The ^{1}H NMR spectrum contains three prominent signals in the δ_H 12.0 region indicating the presence of three NH groups. There should be only two NH groups present therefore, this is indicative of protonation by formic acid seen at δ_H 8.75. Each of these signals is equally strong but the combination of gCOSY, ^{15}N HSQC, and HMBC can be used to identify which belongs to each pyrrole (**Figure 3.1**). There are three separate spin systems, each for a specific pyrrole. One can be seen between these protons: δ_H 7.25, 6.96, and 6.37, each showing gCOSY signals to one another and proton δ_H 12.69, an NH proton with ^{15}N HSQC correlation with δ_N 161.0. This is determined to be the first pyrrole ring system; its final quaternary carbon is identified with the gHMBC correlations each of these protons shows to C_4, δ_C 122.3.

Using gHSQC signals, the methoxy group shows distinct correlations for C_{11} with the strong singlet signal at δ_H 4.03. A strong gHMBC signal can be observed between this proton and the quaternary carbon at δ_C 166.4 which in turn has gHMBC correlations to proton δ_H 7.02 (δ_C 116.6), the olefin signal connecting two pyrrole rings confirmed by ^{15}N HMBC signals seen from this proton to δ_N 139.2 and 164.7 (**Figure 3.2**). An aromatic doublet at δ_H 6.09 shows strong gHMBC correlations to C_6 at δ_C 149.0 and C_9 121.6, two quaternary carbons. This proton also shows ^{15}N HMBC correlations to δ_N 139.5 which corresponds to proton δ_H 12.93 determined with ^{15}N HSQC and gCOSY coupling with proton δ_H 6.09 (**Figure 3.2**). There are two quaternary carbons within this spin system therefore, it is determined that this pyrrole contains the methoxy substituent at C_{11}.

The final pyrrole ring starts with gHSQC correlations confirming that C_{15} (δ_C 129.5) and C_{16} (δ_C 112.7) correspond to aromatic protons δ_H 6.85 and δ_H 6.21. gCOSY coupling can be seen to one another and both to δ_H 12.74, the final NH group with ^{15}N HSQC signal at δ_N 164.7 and ^{15}N HMBC correlations to δ_H 6.21. There are two quaternary carbons, C_{13} and C_{17}, that are identified from gHMBC correlations between δ_H 6.85 with δ_C 126.2 and δ_H 6.21 with δ_C 153.43 completing the pyrrole ring. Lastly, the olefinic proton at δ_H 7.11 shows gHMBC correlations to C_{13} confirming it is the second

pyrrole connected by this bond. Further gHMBC correlations can be seen between δ_C 153.43 and proton signals at δ_H 2.95 and 1.77 indicating the attachment of the aliphatic chain. Due to overlap determining the signals that correspond to this chain was difficult. Starting from the end, C_{28} (δ_C 14.3) correlates with the methyl protons at δ_H 0.87 that appear as a triplet in the ^{1}H NMR. gHMBC signals can be seen between the methyl protons δ_H 0.87 and C_{27} (δ_C 22.8) and C_{26} (δ_C 32.1). Overlapping gHMBC correlations are roughly seen from δ_H 1.33 to C_{24} (δ_C 29.8) and δ_H 1.25 to C_{23} (δ_C 29.4). The remaining aliphatic chain connections are undetermined due to signal overlap.

This full elucidation was then compared to literature of Papireddy *et al.* The experimental and literature ^{1}H and ^{13}C signals were reflective of one another except the NH signals, this occurred due to protonation of the purified compound. Both HSQC and HMBC correlations were consistent with the structure of undecylprodigiosin, specifically the presence of an aliphatic chain and not a ring. Considering these factors, the complete planar structure of **4** was established.

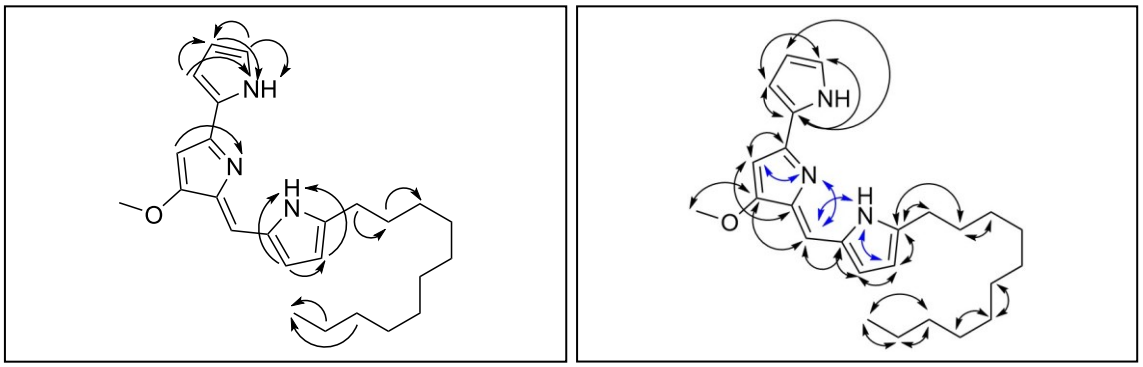


Figure 3.3. Key 2D NMR correlations of 4 in $CDCl_3$ at 600 MHz for ^{1}H and ^{13}C. COSY (one-headed arrows), HMBC (double-headed arrows), and N^{15} HMBC (blue double-headed arrows) correlations are shown.

3.2.4. Isolation of Candidate 552.40 *m/z* (X3) from RLUS-2204D

The first step that filtered the ResistoMAP dataset did not include RLUS-2204 because its ResistoMAP profile only showed strong activity against the ceftazidime-resistant *E. coli* strains Cef-6 and Cef-7. This fraction was added to the peak library separation into subfractions because of its unique profile. The secondary screening displayed activity against all 10 strains for subfraction RLUS-2204D-2, with activity at the lowest concentration of 3.75 µg/mL against the ceftazidime-resistant *E. coli*. As stated in the previous chapter, several mass features were seen within a few mass units of one

another, but **X3** had the strongest MS profile and was selected for prioritized isolation. The highest intensity of **X3** was in RLUS-2204D-3.

During the isolation attempts for **X3** from RLUS-2204D-3 it was determined that several masses could be separated simultaneously. These are the masses separated and the mass obtained; 568.56 (0.07 mg; **X4**), 538.00 (0.10 mg; **X5**), 598.52 (0.40 mg; **X6**), and **X3** (0.39 mg) *m/z*. Purities of **X3** and **X6** were checked via ^{1}H NMR, however, more mass was needed to determine structural characteristics. To obtain additional material the extract was regrown, and isolation was performed on the RLUS-2204D prefraction. This separation revealed four other masses identified in the NP Analyst network to be isolated simultaneously: 582.53 (**X7**), 377.37 (**X8**), 1061.60 (**X9**), and 1054.11 (**10**) *m/z* (**X3**, 1.86 mg; **X4**, 2.22 mg; **X6** *m/z*, 0.40 mg; **X7** *m/z*, 1.23 mg; **X8** *m/z*, 1.89 mg; **X9**, 1.67 mg; **X10**, 1.18 mg). Purities of **X3**, **X4**, **X6**, **X7**, and **X8** were determined using ^{1}H NMR however, only **X3** displayed semi-purity therefore, a full 2D NMR dataset was obtained for elucidation.

3.2.5. Structure Comparison for X3

Compound **X3** was isolated as a colourless solid. The molecular formula of the metabolite was determined through HRMS as $C_{31}H_{53}NO_7$ with the monoisotopic mass for the $[M+H]^+$ adduct: obs 552.3904 *m/z*. Using this information and through NMR analysis using a combination of ^{1}H, ^{13}C, gCOSY, gHSQC, and gHMBC in methanol-d_4 the spectra were compared with literature from Borisova *et al*.[65] (**Appendix Figures B.17-21**). These few key peaks were reflective of one another in the comparison: ketone at δ_C 206.8 and ester at δ_C 175.0, three olefinic protons at δ_H 7.24, 6.43, and 5.67 with corresponding carbons at δ_C 149.8, 119.7, and 147.8, and six methyl groups ranging from δ_C 22.2-9.0 with overlapping proton signals from δ_H 1.85-0.85. This comparison suggests the structure of **X3** is close to **Figure 3.4.**

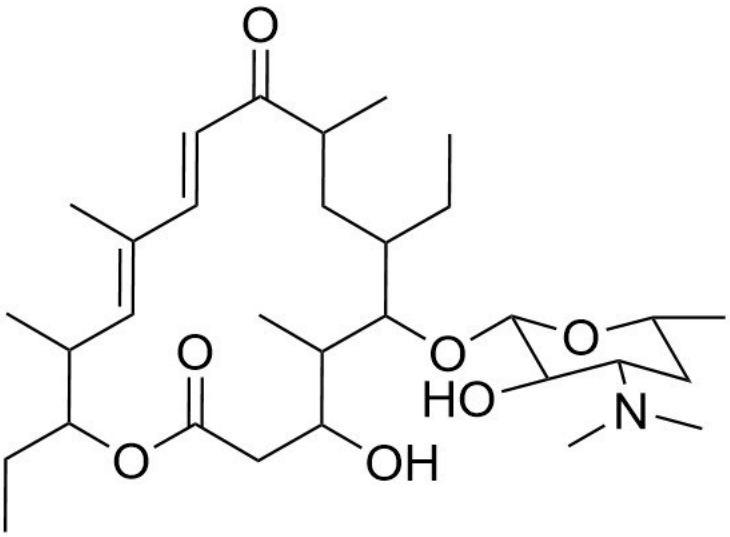


Figure 3.4. Suggested structure of candidate X3 from comparison with Borisova *et al*.[65]

3.2.6. ResistoMAP Screening for Isolated Candidates

Following purification, isolated compounds **2**, **4**, **X3**, **X6**, **X8**, and **X9** were serially diluted (two-fold dilutions; 30 µg/mL to 3.75 µg/mL) and rescreened against the ResistoMAP panel of drug-resistant *E. coli* to determine the minimum inhibitory concentration (MIC) to confirm if bioactive compounds were isolated. Drug-resistant *E. coli* strains can be seen in **Appendix Table A.2**. Due to contamination results of Rif-7 are not included.

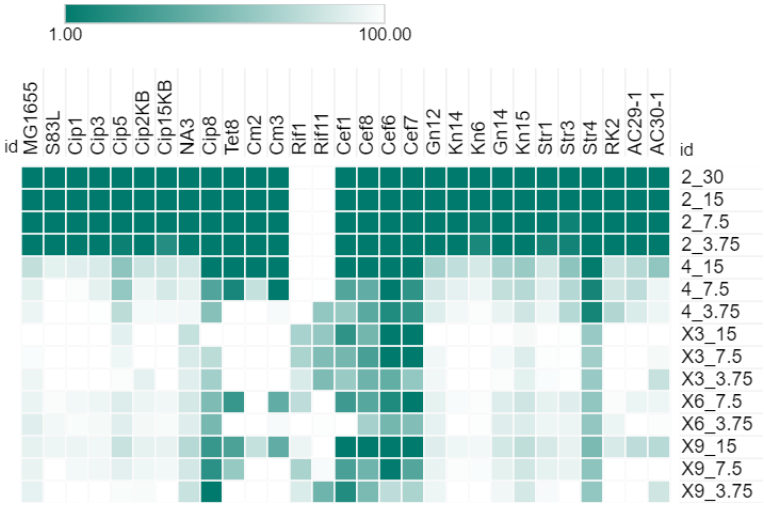


Figure 3.5. ResistoMAP profiles displaying percent growth of Isolated Candidates. Labeling is as follows Candidate_Concentration(µg/mL). If there was 100% growth then no activity was observed.

Compound **2** demonstrated consistent activity with a MIC of 3.75 µg/mL against all drug-resistant strains of *E. coli* except against rifamycin-resistant strains Rif1 and Rif11. Compound **4** displayed a MIC of 15 µg/mL against tetracycline, ciprofloxacin, chloramphenicol, and all ceftazidime-resistant *E. coli*, and an MIC of 3.75 µg/mL against Str4 streptomycin-resistant strain (**Figure 3.5**). These results are reflective of potent bioactivity often displayed by prodiginines.

Isolated components from RLUS-2204 reflected the activity seen in the secondary ResistoMAP screening where the majority activity was observed against the ceftazidime-resistant *E. coli*. Compound **X3** demonstrated consistent activity against all ceftazidime-resistant *E. coli*, with a MIC of 7.5 µg/mL against Cef6 and Cef7. **X6** was similar with an MIC of 7.5 µg/mL against Cef7 and activity at 7.5 µg/mL against the remaining three strains. Additionally, activity was seen against the ciprofloxacin-resistant at 7.5 µg/mL. Lastly, compound **X9** demonstrated a 7.5 µg/mL MIC against Cef6 and 15 µg/mL against the remaining three mutant strains. It also displayed activity against ciprofloxacin and tetracycline-resistant strains at 7.5 µg/mL (**Figure 3.5**).

3.3. Discussion

This work led to the isolation of three bioactive molecules, two of which have been fully elucidated and identified. Prefraction RLUS-2085E displayed activity against all drug-resistant strains, with the strongest being against ceftazidime and streptomycin-resistant *E. coli* strains. Hierarchical clustering of the original ResistoMAP profiles did not group this prefraction with any antimicrobial standards improving confidence in the potential novelty of mass candidates. When divided into a subfraction library, the secondary screening showed consistent activity profiles for the majority of E subfractions. Specifically, RLUS-2085E-3 and RLUS-2085E-6 were of interest due to the observed potent activity at a low concentration of 3.75 µg/mL. These consistent activity profiles led to the selection of mass candidates for isolation from these prefractions. With the guidance of SRPNT and NP Analyst, the first candidate isolated was 392.27 *m/z* (Rt 3.90 min). Initially, NP Atlas matched this candidate with the prodiginine family specifically, **2** and metacycloprodigiosin (**3**). Upon isolation, a characteristic UV/Vis peak was observed at 545 nm (**Appendix Figure B.1**), this is known to correlate to the prodiginine family. Members of the prodiginine family have been isolated from various microorganisms including *Streptomyces coelicolor* and recently, marine bacteria *Hahella*

chejuensis and *Pseudoalteromonas*.[66,67] This family of compounds is known to demonstrate several biological functions including antibacterial, antifungal and anti-inflammatory activity. Antibacterial activity has been observed against *Bacillus subtilis*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *E. coli*.[67–69] Although antibacterial activity has been observed against this bacteria there is limited information on the prodiginine effect on resistant pathogens. Prodiginine standards were not included in the positive control library of ResistoMAP and have not been documented as isolated from the Linington library therefore, extract RLUS-2085 was not dereplicated within the first steps of the project.

Following purification, full 2D NMR was obtained for elucidation. Deconvolution with a study by Weyland and co-workers (peak comparison seen in **Table 3.1**) confirmed the isolated compound as **2**.[54,59] There has been a debate on the assignment of **2** compared to butylcycloheptylprodigiosin however, the signal of δ_H -1.55 is characteristic to **2** therefore, this compound is not butylcycloheptylprodiosin.[59] When in an acidic environment prodiginines become protonated resulting in three NH signals within the NMR spectra (**Figure 3.6**). Prodiginines can also undergo tautomerization producing three alternative structures, where the protonated nitrogen can switch positions.[70] When observing the experimental spectra there is no doubling of proton signals or carbons therefore, multiple tautomers are not present within the purified compound. The three NH signals are a result of protonation from the formic acid present in solution during purification.[55]

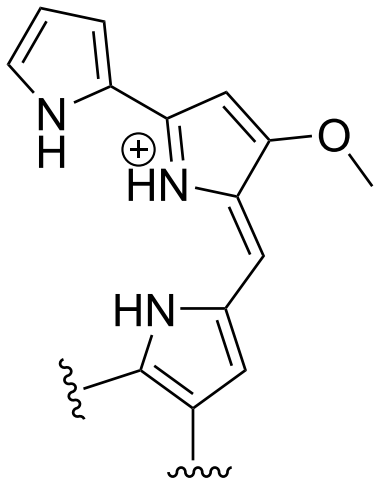


Figure 3.6. Protonation of prodiginines that can occur in an acidic environment.

Additionally, NP Analyst associated feature 394.30 *m/z* (Rt 4.29 min) with strong Activity score within subfraction RLUS-2085E-6. With its proximity to the first isolated mass, NP Atlas also identified this mass as a part of the prodiginine family, **4** or prodigiosin 25-C. The isolation for this candidate was made easy as its separation coincided with the first candidate from prefraction RLUS-2085E. Again, the isolation identified the characteristic UV/Vis peak at 525 nm (**Appendix Figure B.9)**. Following purification, a full 2D NMR set was obtained for elucidation. Upon elucidation, the ^{1}H NMR displayed a unique splitting in the region of δ_H 12.9-12.5. There should be two singlets representing the NH groups within the characteristic prodiginine pyrrole rings however, the experimental spectra displayed three distinct signals. These results were like that of **2** therefore, this additional NH single is attributed to protonation of the purified compound by formic acid. After partial elucidation and deconvolution with the literature, this compound was identified as **4**.

The prodiginine family is of interest for the biological properties the compounds display.[58] Both the prefraction and subfractions of RLUS-2085E displayed a consistent phenotypic profile against the ResistoMAP panel of drug-resistant *E. coli*. Concluding screening determined that both isolated molecules; **2** and **4**, reflected these strong phenotypic profiles identified within the original screens. **2** displayed strong activity against most of the drug-resistant *E. coli* strains while **4** was selectively potent against the tetracycline, ciprofloxacin, chloramphenicol, and all ceftazidime-resistant strains. These activity profiles reflect the known biological properties of prodiginines against *E. coli* while additionally, highlighting compounds **2** and **4** antibacterial activity against drug-resistant pathogens.[68,69] Prodigiosin (**1**) has been identified as active against *E. coli* and against drug-resistant *Staphylococcus aureus* where it exhibited inhibitory activity against biofilm formation however, there is limited information on the impacts prodiginines have on other resistant pathogens.[68,71] The combined efforts of ResistoMAP, NP Analyst, and SRPNT allowed the identification of these antibacterial compounds within the Linington library and revealed the activity against the resistant pathogens. With the observed activity of compounds **2** and **4** the applications of prodiginines to combat the AMR crisis are of importance. Further research on the pharmacodynamics and toxicity are needed to determine the drug applications of **2** and **4**.

An alternative technique that this project used was to select extracts strictly based on their activity profiles. The initial filtering of the ResistoMAP dataset removed extracts that showed less than 50% activity against 5 or more strains meaning it could have removed extracts that displayed unique phenotypic profiles. One extract that did not pass the filtering was RLUS-2204. Its original ResitoMAP profile was relatively weak against the drug-resistant strains however, it showed potent activity against the ceftazidime-resistant *E. coli*. For this reason, it was selected to add to the secondary steps of this project. The secondary screening performed on the subfractions displayed a similar pattern was seen in RLUS-2204D-2, it displayed activity against the 10 drug-resistant strains but was active at the lowest concentration of 3.75 µg/mL against the ceftazidime-resistant strains. After examination of the NP Analyst network a group of masses within a few mass units of one another were reviewed on NP Atlas which matched them with a group of macrolides isolated from *Micromonospora*. After analyzing the intensity of the MS data, candidate 552.40 *m/z* (Rt 3.19, **X3**) was the primary target for isolation from RLUS-2204D-3. Upon isolation it was discovered that several of the candidates could be isolated in a single chromatographic run; therefore, **X4**-**X10** were isolated as well. The UV/Vis profile for the candidates was difficult to identify so a mass-guided separation was performed on the prefraction RLUS-2204D. After purification, a full 2D NMR set for **X3** was obtained for elucidation. Using the suggested structure from NP Atlas, the experimental spectra were compared to that of a NP hybrid produced by biosynthetic engineering and fully elucidated by Borisova *et al* seen in **Figure 3.4**.[65] Literature ^{13}C signals for the ketone and ester were reflective of the experimental signals at δ_C 206.8 and 175.0, while the olefinic and aliphatic chain signals were in the same range but not exact matches. This strongly suggests the planar structure of **X3** is that identified by this study (**Figure 3.4**) however, stereochemistry is needed to confirm the complete configuration of this isolated compound.

Along with this mass candidate to ensure a bioactive candidate was isolated a few of the other analogues were screened as well. **X3** displayed an activity profile that was reflective of those seen in the original screen. It demonstrated consistent activity against ceftazidime-resistant strains specifically, an MIC of 7.5 µg/mL against Cef6 and Cef7. Additionally, **X6** and **X9** demonstrated similar activity profiles with the strongest being against the ceftazidime-resistant strains. Candidate **X9** also revealed activity at 7.5

μg/mL against ciprofloxacin and tetracycline-resistant *E. coli*. **X6** and **X9** are future targets for further purification and elucidation.

This project resulted in the isolation of three molecules, two of which have been identified as members of the bioactive family of compounds, prodiginines. All three main targets displayed activity that was reflective of the original screen. With these conclusions, the selection process outlined using ResistoMAP bioactivity and NP Analyst for metabolomic analysis is successful in identifying bioactive compounds effective against AMR. The dereplication application of the tool was proven to succeed as these isolated compounds did not match those of the antimicrobial standards used in the initial dataset. It is theorized that if a prodiginine standard were included in the initial steps of this project it would have dereplicated these extracts rather quickly. In the future, these isolated compounds will be applied to the Linington library to aid in the dereplication process using previously identified compounds. With the rising crisis of AMR, new strategies are needed to efficiently identify new candidates that can effectively combat AMR. This project outlines a new NP discovery pipeline that has succeeded in the isolation of bioactive compounds that inhibit the growth of drug-resistant *E. coli*. Therefore, if applied to alternative datasets the use of ResistoMAP and NP Analyst may potentially lead to the identification of NP candidates that can contend with AMR while avoiding rediscovery.

3.4. Materials and Methods

3.4.1. Fermentation, Isolation, and Discovery of RLUS-2085E Prodiginines

Fermentation and Subfraction Separation

Frozen stock of *Streptomyces spp.* RL12-007-NTS-A (1:1 glycerol: SYP media; -70 ˚C) were plated on solid media (MB agar; 37.4 g/L, 10 g/L soluble starch, 4 g/L yeast extract, 2 g/L peptone) at room temperature (RT) for 6 days. Individual colonies were used to inoculate 6 × 7 mL SYP liquid media in sterile capped 40 mL culture tubes with 3 glass beads, shaking at RT, 200 RPM for 4 days. 3 mL of liquid culture solution was then used to inoculate 8 × 1 L of SYP media inside sterile 2.8 L Fernbach flasks containing a large stainless-steel spring and 20 g of XAD-17 resin capped with autoclaved milk filter, shaking at RT, 200 RPM for 7 days. Following fermentation, the

large-scale culture (8 × 1 L) was filtered then extracted using DCM/MeOH then separated in 5 prefractions using the CombiFlash system via an eluotropic series of MeOH/H_2O (1:1) and dried in 20 mL borosilicate glass scintillation vials. Prefractions A (99.02 mg), B (133.44 mg), C (86.90 mg), D (88.52 mg), and E (117.63 mg) were diluted to a concentration of 50 µg/mL and centrifuged. 40 µL of each prefraction supernatant was stored while the remaining volume was transferred to 1.5 mL LC-MS total recovery vials and separated via HPLC (Agilent 1200 series; Phenomonex Synergi™ 10 µm Fusion-RP 80 A 250 × 10 mm) using a customized gradient of MeCN/H_2O + 0.02% formic acid at a flow rate of 4.0 mL min^{-1} for each prefraction (**Table 3.3**).

Table 3.3. Customized gradients for the separation of RLUS-2085 prefractions into subfraction peak libraries.

Prefraction	Gradient
RLUS-2085A	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 25 % to 45 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min.
RLUS-2085B	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 30 % to 60 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min.
RLUS-2085C	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 45 % to 65 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min.
RLUS-2085D	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 60 % to 85 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min.
RLUS-2085E	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 60 % to 80 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min.

Regrow of *Streptomyces spp.* RL12-007-NTS-A was performed using the same procedure outlined above. In brief, it was plated for 7 days, at a small scale for 6 days, at a medium scale for 5 days, and lastly was extracted after 5 days at a large scale. Prefractions were separated via CombiFlash and dried in 20 mL scintillation vials: A (240.34 mg), B (169.97 mg), C (148.35 mg), D (265.23 mg), and E (161.85 mg). These prefractions were not separated into subfractions.

Bioactive Analog Isolation

RLUS-2085E-3 starting mass of 7.43 mg was diluted to 3.5 µg/mL and transferred to a 1.5 mL LC-MS total recovery vial and purified via HPLC-MS (Agilent 6130 MS; Phenomonex Kinetix 2.6 µm XB-C18 150 × 4.6 mm) using a gradient of MeCN/H_2O + 0.02% formic acid (30% to 70% MeCN over 30 min, 95% MeCN for 5 min) at a flow rate of 1.0 mL min^{-1}. One analog was separated with this method, collected via

fraction collector and dried down under vacuum, then weighed ($[M+H]^+$ 392.1, Rt 21.2 min, 1.42 mg). The resulting compound was profiled by UPLC-HRMS (Waters SYNAPT G2Si ESI-qTOF-MS) to obtain hi-res mass and MS/MS data to check purity. ^{1}H NMR (600 MHz) was performed to check purity.

Analogues were also separated from re-grow prefraction RLUS-2085E (161.85 mg). It was diluted to 10 µg/mL, transferred to a 1.5 mL LC-MS total recovery vial, and separated via HPLC-MS using a gradient of MeCN/H_2O + 0.02% formic acid (45% to 70% MeCN over 30 min, 95% MeCN for 5 min) at a flow rate of 1.0 mL min^{-1}. Two analogues displayed prominent absorbances within the UV-Vis region (**2**: $[M+H]^+$ 392.3 *m/z*, Rt 7.75 min, 7.81 mg; **4**: $[M+H]^+$ 394.3, Rt 16.1 min, 7.29 mg). Structures were verified using 2D NMR (600 MHz) (**Appendix Figures B.2-16**).

3.4.2. Fermentation, Isolation, and Analysis of RLUS-2028

The fermentation procedure outlined in section **3.4.1** was repeated with a frozen stock of *Streptomyces spp.* RL12-042-HVF-A (1:1 glycerol: SYP media; -70 ˚C). In brief, it was plated for 7 days, at a small scale for 6 days, medium scale for 7 days, and lastly was extracted after 7 days at a large scale. Prefractions were separated via CombiFlash and dried in 20 mL scintillation vials: A (227.16 mg), B (382.52 mg), C (404.17 mg), D (1041.04 mg), and E (306.87 mg). The separation into the subfraction peak library was performed using customized gradients in **Table 3.4**.

Table 3.4. Customized gradients for the separation of RLUS-2028 prefractions into subfraction peak libraries.

Prefraction	Gradient
RLUS-2028A	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 20 % to 60 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2028B	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 25 % to 50 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2028C	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 25 % to 40 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2028D	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 40 % to 75 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2028E	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 60 % to 85 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min

With a mass of 6.70 mg, RLUS-2028A-2 was diluted to 7 µg/mL and transferred to a 1.5 mL LC-MS total recovery vial and purified via HPLC-MS using a gradient of MeCN/H_2O + 0.02% formic acid (5% to 10% MeCN over 10 min, 10% MeCN for 20 min, 95% MeCN for 5 min) at a flow rate of 1.0 mL min^{-1}. Two analogues (**X1**: $[M+H]^+$ 438.2 *m/z,* 12.8 min, 0.25 mg; **X2**: 452.2 *m/z*, Rt 14.61 min, 0.29 mg) were attempted at separation with this method and collected via a fraction collector. Due to lack of mass and difficulty of separation, these isolations were not pursued to completion.

3.4.3. Purification, Isolation, and Discovery of RLUS-2204D

The fermentation procedure outlined in section **3.4.1** was repeated with a frozen stock of *Streptomyces spp.* RL12-082-NTF-B (1:1 glycerol: SYP media; -70 ˚C). In brief, it was plated for 13 days, at a small scale for 15 days, at a medium scale for 14 days, and lastly was extracted after 10 days at a large scale. Prefractions were separated via CombiFlash and dried in 20 mL scintillation vials: A (112.55 mg), B (255.40 mg), C (318.96 mg), D (182.09 mg), and E (220.38 mg). The separation into the subfraction peak library was performed on prefractions C-E using customized gradients in **Table 3.5**.

Table 3.5. Customized gradients for the separation of RLUS-2204 prefractions into subfraction peak libraries.

Prefraction	Gradient
RLUS-2204C	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 20 % to 60 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2204D	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 40 % to 65 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min
RLUS-2204E	MeCN/H_2O + 0.02 % formic acid (5 % MeCN for 2 min, 65 % to 85 % MeCN over 18 min, 100 % MeCN for 2 min) at a flow rate of 4.0 mL/min

Regrow of *Streptomyces spp.* RL12-082-NTF-B was performed using the same procedure outlined previously. In brief, it was plated for 13 days, at a small scale for 15 days, at a medium scale for 10 days, and lastly was extracted after 10 days at a large scale. Prefractions were separated via CombiFlash and dried in 20 mL scintillation vials: A (177.11 mg), B (228.96 mg), C (370.24 mg), D (234.30 mg), and E (430.30 mg). These prefractions were not separated into subfractions.

Bioactive Analog Isolation

RLUS-2204D-3 starting mass of 9.86 mg was diluted to 10 µg/mL and transferred to a 1.5 mL LC-MS total recovery vial and purified via HPLC-MS using a gradient of MeCN/H_2O + 0.02% formic acid (30% to 40% MeCN over 30 min, 95% MeCN for 5 min) at a flow rate of 1.0 mL min^{-1}. Four analogues were separated with this method, collected via fraction collector and dried down under vacuum, then weighed ($[M+H]^+$ 598.0 *m/z*, Rt 4.5 min, 0.4 mg; $[M+H]^+$ 568.0 *m/z*, Rt 13.1 min, 0.07 mg; $[M+H]^+$ 538.0 *m/z*, Rt 15.6 min, 0.10 mg; $[M+H]^+$ 552.0 *m/z*, Rt 19.2 min, 0.39 mg).

Re-grow prefraction RLUS-2204D (234.36 mg) was diluted to 20 µg/mL and transferred to 1.5 mL LC-MS total recovery vials. Analogues listed above were separated via HPLC-MS (Waters 2767 SQ Detector 2; Synergi 10 µm Fusion-RP 80Å 250 × 10 mm) using a gradient of MeCN/H_2O + 0.02% formic acid (25% to 40% MeCN over 40 min, 95% MeCN for 8 min) at a flow rate of 8.0 mL min^{-1}. Seven analogues were separated with this method, collected via fraction collector and dried down under vacuum, then weighed ($[M+H]^+$ 568.56 *m/z*, Rt 6.0 min, 2.22 mg (**X4**); $[M+H]^+$ 598.52 *m/z*, Rt 9.0 min, 0.40 mg (**X6**); $[M+H]^+$ 582.53 *m/z*, Rt 7.0 min, 1.23 mg (**X7**); $[M+H]^+$ 552.57 *m/z*, Rt 18.0 min, 1.86 mg (**X3**); $[M+H]^+$ 377.37 *m/z*, Rt 28.0 min, 1.89 mg (**X8**); $[M+H]^+$ 1061.60 *m/z*, Rt 38.5 min, 1.67 mg (**X9**); $[M+H]^+$ 1054.11 *m/z*, Rt 39.0 min, 1.18 mg (**X10**)). Full 2D NMR was obtained on **X3** (**Appendix Figures B.17-21)**.

3.4.4. Isolated Candidates ResistoMAP

ResistoMAP design and analyses were performed according to the procedure outlined by Liu *et al.*[47] Screening included the full panel of drug-resistant *E. coli* while commercial standards were excluded from this assay. Isolated candidates; **2**, **4**, **X3**, **X6**, **X8**, and **X9** were diluted four-fold from 30 µg/mL. All processing and normalization were completed as explained in **Section 2.4.1.**

Minimum Inhibitory Concentration (MIC)

For each compound and drug-resistant strain combination, the minimum inhibitory concentration (MIC) was determined as the lowest concentration of the compound to inhibit less than 10% growth.

Chapter 4.

Conclusion and Outlook

To address the need for new techniques that expand the research field of natural product (NP) discovery a new method for the identification of NP targets that will be effective against resistant pathogens was developed. Using ResistoMAP, a profiling platform that uses a panel of drug-resistant *E. coli* and a positive control library of commercially available antimicrobial standards, the phenotypic profiles of 384 NP prefractions were dereplicated with those of the standards. To prioritize the extracts further analysis of mass spectrometry (MS) of all prefractions and standards was performed. The activity and MS data were then inputted into NP Analyst, an online metabolomics software designed to identify bioactive metabolites within complex NP mixtures, for further dereplication and prioritization of candidates for investigation.

The combined efforts of these platforms allowed the selection of 18 NP prefractions for full experimentation. Each prefraction was separated into a peak library of 10 subfractions to further isolate the bioactive component observed in the ResistoMAP profile. Screening of the subfractions determined the location of the active components that were selected to obtain MS data. These data were then used to produce a secondary NP Analyst network that allowed differentiation between prefractions and led to the selection of 3 candidates for isolation and identification: 392.27 (**2**) and 394.30 (**4**) *m/z* from RLUS-2085E and 552.40 *m/z* (**X3**) from RLUS-2204D.

Comparison of the metabolomics data with the Natural Product Atlas database, guided the characterization of these candidates. Combined efforts of a full 2D NMR dataset and comparison with literature led to the identification of compound **2** as streptorubin B and compound **4** as undecylprodigiosin. Both compounds are known to display antibacterial activity. Screening of pure candidates demonstrated strong activity against the drug-resistant *E. coli* panel of ResistoMAP. The final candidate, **X3** planar structure has been predicted to be that of a macrolide antibiotic using 2D NMR comparison. Final screening demonstrated activity against ceftazidime-resistant *E. coli*.

In conclusion, the use of ResistoMAP and NP Analyst led to the successful identification of three compounds that are effective against drug-resistant pathogens. The techniques demonstrated by this project enable a selection process that avoids the rediscovery of known chemistry within a control library of antimicrobial standards while identifying active compounds that can combat the AMR crisis.

References

(1) Dias, D. A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2* (2), 303–336. https://doi.org/10.3390/metabo2020303.

(2) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *Journal of Industrial Microbiology and Biotechnology* **2016**, *43* (2–3), 155–176. https://doi.org/10.1007/S10295-015-1723-5.

(3) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nature Reviews Drug Discovery* **2021**, *20* (3), 200–216. https://doi.org/10.1038/s41573-020-00114-z.

(4) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285.

(5) Jensen, P. R.; Mincer, T. J.; Williams, P. G.; Fenical, W. Marine Actinomycete Diversity and Natural Product Discovery. *Antonie van Leeuwenhoek* **2005**, *87* (1), 43–48. https://doi.org/10.1007/s10482-004-6540-1.

(6) Luo, Y.; Cobb, R. E.; Zhao, H. Recent Advances in Natural Product Discovery. *Current Opinion in Biotechnology* **2014**, *30*, 230–237. https://doi.org/10.1016/J.COPBIO.2014.09.002.

(7) Kardos, N.; Demain, A. L. Penicillin: The Medicine with the Greatest Impact on Therapeutic Outcomes. *Applied Microbiology and Biotechnology* **2011**, *92* (4), 677–687. https://doi.org/10.1007/s00253-011-3587-6.

(8) Yuan, H.; Ma, Q.; Ye, L.; Piao, G. The Traditional Medicine and Modern Medicine from Natural Products. *Molecules* **2016**, *21* (5), 559–559. https://doi.org/10.3390/molecules21050559.

(9) Baltz, R. H. Renaissance in Antibacterial Discovery from Actinomycetes. *Current Opinion in Pharmacology* **2008**, *8* (5), 557–563. https://doi.org/10.1016/J.COPH.2008.04.008.

(10) Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; Van Wezel, G. P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev* **2016**, *80* (1), 1–43. https://doi.org/10.1128/MMBR.00019-15.

(11) Hutchings, M.; Truman, A.; Wilkinson, B. Antibiotics: Past, Present and Future. *Current Opinion in Microbiology* **2019**, *51*, 72–80. https://doi.org/10.1016/J.MIB.2019.10.008.

(12) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic Discovery: History, Methods and Perspectives. *International Journal of Antimicrobial Agents* **2019**, *53* (4), 371–382. https://doi.org/10.1016/J.IJANTIMICAG.2018.11.010.

(13) De Lima Procópio, R. E.; Da Silva, I. R.; Martins, M. K.; De Azevedo, J. L.; De Araújo, J. M. Antibiotics Produced by Streptomyces. *The Brazilian Journal of Infectious Diseases* **2012**, *16* (5), 466–471. https://doi.org/10.1016/j.bjid.2012.08.014.

(14) Quinn, G. A.; Banat, A. M.; Abdelhameed, A. M.; Banat, I. M. Streptomyces from Traditional Medicine: Sources of New Innovations in Antibiotic Discovery. *Journal of Medical Microbiology* **2020**, *69* (8), 1040–1048. https://doi.org/10.1099/jmm.0.001232.

(15) Fenical, W.; Jensen, P. R. Developing a New Resource for Drug Discovery: Marine Actinomycete Bacteria. *Nat Chem Biol* **2006**, *2* (12), 666–673. https://doi.org/10.1038/nchembio841.

(16) Kasanah, N.; Triyanto, T. Bioactivities of Halometabolites from Marine Actinobacteria. *Biomolecules* **2019**, *9* (6), 225. https://doi.org/10.3390/biom9060225.

(17) Jensen, P. R. Microbe Profile: Salinispora Tropica: Natural Products and the Evolution of a Unique Marine Bacterium. *Microbiology* **2022**, *168* (4). https://doi.org/10.1099/mic.0.001163.

(18) Nett, M.; Ikeda, H.; Moore, B. S. Genomic Basis for Natural Product Biosynthetic Diversity in the Actinomycetes. *Natural Product Reports* **2009**, *26* (11), 1362–1362. https://doi.org/10.1039/b817069j.

(19) Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2022**, *39* (6), 1122–1171. https://doi.org/10.1039/D1NP00076D.

(20) Pereira, F.; Aires-de-Sousa, J. Computational Methodologies in the Exploration of Marine Natural Product Leads. *Marine Drugs* **2018**, *16* (7), 236. https://doi.org/10.3390/md16070236.

(21) Butler, M. S. The Role of Natural Product Chemistry in Drug Discovery. *J. Nat. Prod.* **2004**, *67* (12), 2141–2153. https://doi.org/10.1021/np040106y.

(22) Genilloud, O. Natural Products Discovery and Potential for New Antibiotics. *Current Opinion in Microbiology* **2019**, *51*, 81–87. https://doi.org/10.1016/J.MIB.2019.10.012.

(23) Lewis, K. The Science of Antibiotic Discovery. *Cell* **2020**, *181* (1), 29–45. https://doi.org/10.1016/j.cell.2020.02.056.

(24) Maitra, U.; Stephen, C.; Ciesla, L. M. Drug Discovery from Natural Products – Old Problems and Novel Solutions for the Treatment of Neurodegenerative Diseases. *Journal of Pharmaceutical and Biomedical Analysis* **2022**, *210*, 114553–114553. https://doi.org/10.1016/J.JPBA.2021.114553.

(25) Rossiter, S. E.; Fletcher, M. H.; Wuest, W. M. Natural Products as Platforms To Overcome Antibiotic Resistance. *Chemical Reviews* **2017**, *117* (19), 12415–12474. https://doi.org/10.1021/acs.chemrev.7b00283.

(26) Li, J. W.-H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **2009**, *325* (5937), 161–165. https://doi.org/10.1126/science.1168243.

(27) Dadgostar, P. Antimicrobial Resistance: Implications and Costs. *IDR* **2019**, *Volume 12*, 3903–3910. https://doi.org/10.2147/IDR.S234610.

(28) Murray, C. J.; Shunji Ikuta, K.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Give Chipeta, M.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Kashef Hamadani, B. H.; P Kumaran, E. A.; McManigal, B.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi, J.; Andrews, J.; Aravkin, A.; Ashley, E.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carvalheiro, C.; CastaÃ, C.; Chansamouth, V.; Chaurasia, S.; ChiurchiÃ, S.; Chowdhury, F.; Cook, A. J.; Cooper, B.; Cressey, T. R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; J Day, N. P.; De Luca, M.; Dokova, K.; Dramowski, A.; Dunachie, S. J.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garrett, D.; Gastmeier, P.; Zergaw Giref, A.; Claire Greer, R.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S. I.; Holm, M.; Hopkins, S.; Iregbu, K. C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Khorana, M.; Kissoon, N.; Kobeissi, E.; Kostyanev, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Hmwe Kyu, H.; Lim, C.; Limmathurotsakul, D.; James Loftus, M.; Lunn, M.; Ma, J.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Marcia Mussi-Pinhata, M.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.; Obiero, C. W.; Olivas-Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Yariv Peleg, A.; Perrone, C.; Plakkal, N.; Ponce-de-Leon, A.; Raad, M.; Ramdin, T.; Riddell, A.; Roberts, T.; Victoria Robotham, J.; Roca, A.; Rudd, K. E.; Russell, N.; Schnall, J.; Anthony Gerard Scott, J.; Shivamallappa, M.; Sifuentes-Osornio, J.; Steenkeste, N.; James Stewardson, A.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Turner, C.; Turner, P.; Rogier van Doorn, H.; Velaphi, S.; Vongpradith, A.; Vu, H.; Walsh, T.; Waner, S.; Wangrangsimakul, T.; Wozniak, T.; Zheng, P.; Sartorius, B.; Lopez, A. D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M.; Resistance Collaborators, A. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. **2022**. https://doi.org/10.1016/S0140-6736(21)02724-0.

(29) Barbosa, F.; Pinto, E.; Kijjoa, A.; Pinto, M.; Sousa, E. Targeting Antimicrobial Drug Resistance with Marine Natural Products. *International Journal of Antimicrobial Agents* **2020**, *56* (1), 106005–106005. https://doi.org/10.1016/J.IJANTIMICAG.2020.106005.

(30) Lobanovska, M.; Pilla, G. *Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future?*; 2017; Vol. 90, pp 135–145.

(31) Bérdy, J. Thoughts and Facts about Antibiotics: Where We Are Now and Where We Are Heading. *Journal of Antibiotics* **2012**, *65* (8), 385–395. https://doi.org/10.1038/ja.2012.27.

(32) Bueno, J. Antimicrobial Adjuvants Drug Discovery, the Challenge of Avoid the Resistance and Recover the Susceptibility of Multidrug-Resistant Strains. *J Microb Biochem Technol* **2016**, *8* (3). https://doi.org/10.4172/1948-5948.1000281.

(33) Medema, M. H.; Fischbach, M. A. Computational Approaches to Natural Product Discovery. *Nature Chemical Biology* **2015**, *11* (9), 639–648. https://doi.org/10.1038/nchembio.1884.

(34) Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. Dereplication Strategies in Natural Product Research: How Many Tools and Methodologies behind the Same Concept? *Phytochem Rev* **2017**, *16* (1), 55–95. https://doi.org/10.1007/s11101-015-9448-7.

(35) Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S. Current Approaches and Challenges for the Metabolite Profiling of Complex Natural Extracts. *Journal of Chromatography A* **2015**, *1382*, 136–164. https://doi.org/10.1016/j.chroma.2014.10.091.

(36) Garg, N.; Kapono, C. A.; Lim, Y. W.; Koyama, N.; Vermeij, M. J. A.; Conrad, D.; Rohwer, F.; Dorrestein, P. C. Mass Spectral Similarity for Untargeted Metabolomics Data Analysis of Complex Mixtures. *International Journal of Mass Spectrometry* **2015**, *377*, 719–727. https://doi.org/10.1016/j.ijms.2014.06.005.

(37) Winnikoff, J. R.; Glukhov, E.; Watrous, J.; Dorrestein, P. C.; Gerwick, W. H. Quantitative Molecular Networking to Profile Marine Cyanobacterial Metabolomes. *J Antibiot* **2014**, *67* (1), 105–112. https://doi.org/10.1038/ja.2013.120.

(38) Butler, M. S.; Blaskovich, M. A.; Owen, J. G.; Cooper, M. A. Old Dogs and New Tricks in Antimicrobial Discovery. *Current Opinion in Microbiology* **2016**, *33*, 25–34. https://doi.org/10.1016/j.mib.2016.05.011.

(39) Kealey, C.; Creaven, C. A.; Murphy, C. D.; Brady, C. B. New Approaches to Antibiotic Discovery. *Biotechnol Lett* **2017**, *39* (6), 805–817. https://doi.org/10.1007/s10529-017-2311-8.

(40) Medema, M. H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Research* **2011**, *39* (suppl_2), W339–W346. https://doi.org/10.1093/nar/gkr466.

(41) Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A. Comprehensive Prediction of Secondary Metabolite Structure and Biological Activity from Microbial Genome Sequences. *Nat Commun* **2020**, *11* (1), 6058. https://doi.org/10.1038/s41467-020-19986-1.

(42) Chen, Y.; Kirchmair, J. Cheminformatics in Natural Product-based Drug Discovery. *Molecular Informatics* **2020**, *39* (12), 2000171. https://doi.org/10.1002/minf.202000171.

(43) Sorokina, M.; Steinbeck, C. Review on Natural Products Databases: Where to Find Data in 2020. *J Cheminform* **2020**, *12* (1), 20. https://doi.org/10.1186/s13321-020-00424-9.

(44) van Santen, J. A.; Poynton, E. F.; Iskakova, D.; McMann, E.; Alsup, T. A.; Clark, T. N.; Fergusson, C. H.; Fewer, D. P.; Hughes, A. H.; McCadden, C. A.; Parra, J.; Soldatou, S.; Rudolf, J. D.; Janssen, E. M.-L.; Duncan, K. R.; Linington, R. G. The Natural Products Atlas 2.0: A Database of Microbially-Derived Natural Products. *Nucleic Acids Research* **2022**, *50* (D1), D1317–D1323. https://doi.org/10.1093/nar/gkab941.

(45) Cox, G.; Sieron, A.; King, A. M.; De Pascale, G.; Pawlowski, A. C.; Koteva, K.; Wright, G. D. A Common Platform for Antibiotic Dereplication and Adjuvant Discovery. *Cell Chemical Biology* **2017**, *24* (1), 98–109. https://doi.org/10.1016/j.chembiol.2016.11.011.

(46) Liu, D. Y. Multidimensional High-Throughput Screening Approaches to Antimicrobial Drug Discovery, Simon Fraser University, 2023. https://summit.sfu.ca/_flysystem/fedora/2023-02/etd22189.pdf.

(47) Liu, D. Y.; Phillips, L.; Wilson, D. M.; Fulton, K. M.; Twine, S. M.; Wong, A.; Linington, R. G. Collateral Sensitivity Profiling in Drug-Resistant Escherichia Coli Identifies Natural Products Suppressing Cephalosporin Resistance. *Nat Commun* **2023**, *14* (1), 1976. https://doi.org/10.1038/s41467-023-37624-4.

(48) Lee, S.; van Santen, J. A.; Farzaneh, N.; Liu, D. Y.; Pye, C. R.; Baumeister, T. U. H.; Wong, W. R.; Linington, R. G. NP Analyst: An Open Online Platform for Compound Activity Mapping. *ACS Central Science* **2022**, *8* (2), 223–234. https://doi.org/10.1021/acscentsci.1c01108.

(49) Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B. 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. https://doi.org/10.1021/acs.jnatprod.5b01014.

(50) Kvalheim, O. M. Interpretation of Partial Least Squares Regression Models by Means of Target Projection and Selectivity Ratio Plots. *Journal of Chemometrics* **2010**, *24* (7–8), 496–504. https://doi.org/10.1002/cem.1289.

(51) Kvalheim, O. M. Variable Importance: Comparison of Selectivity Ratio and Significance Multivariate Correlation for Interpretation of Latent-variable Regression Models. *Journal of Chemometrics* **2020**, *34* (4), e3211. https://doi.org/10.1002/cem.3211.

(52) Baumeister, T. U. H.; Aadland, E.; Linington, R. G.; Kvalheim, O. M. Multivariate Pattern Analysis: A Method and Software to Reveal, Quantify, and Visualize Predictive Association Patterns in Multicollinear Data. *BMC Bioinformatics* **2024**, *25* (1), 51. https://doi.org/10.1186/s12859-024-05660-6.

(53) Baumeister, T. U. H.; Kvalheim, O. M.; Linington, R. G. Selectivity Ratio Profile Networks (SRPNT). *Manuscript in preparation 2024*.

(54) Kinumaki, A.; Harada, K.-I.; Suzuki, T.; Suzuki, M.; Okuda, T. Macrolide Antibiotics M-4365 Produced by Micromonospora. II. Chemical Structures. *J. Antibiot.* **1977**, *30* (6), 450–454. https://doi.org/10.7164/antibiotics.30.450.

(55) Hu, D. X.; Withall, D. M.; Challis, G. L.; Thomson, R. J. Structure, Chemical Synthesis, and Biosynthesis of Prodiginine Natural Products. *Chem. Rev.* **2016**, *116* (14), 7818–7853. https://doi.org/10.1021/acs.chemrev.6b00024.

(56) Jones, B. T.; Hu, D. X.; Savoie, B. M.; Thomson, R. J. Elimination of Butylcycloheptylprodigiosin as a Known Natural Product Inspired by an Evolutionary Hypothesis for Cyclic Prodigiosin Biosynthesis. *J. Nat. Prod.* **2013**, *76* (10), 1937–1945. https://doi.org/10.1021/np400531b.

(57) Reeves, J. T. A Concise Synthesis of Butylcycloheptylprodigiosin. *Org. Lett.* **2007**, *9* (10), 1879–1881. https://doi.org/10.1021/ol070341i.

(58) Papireddy, K.; Smilkstein, M.; Kelly, J. X.; Shweta; Salem, S. M.; Alhamadsheh, M.; Haynes, S. W.; Challis, G. L.; Reynolds, K. A. Antimalarial Activity of Natural and Synthetic Prodiginines. *J. Med. Chem.* **2011**, *54* (15), 5296–5306. https://doi.org/10.1021/jm200543y.

(59) Fürstner, A.; Radkowski, K.; Peters, H.; Seidel, G.; Wirtz, C.; Mynott, R.; Lehmann, C. W. Total Synthesis, Molecular Editing and Evaluation of a Tripyrrolic Natural Product: The Case of “Butylcycloheptylprodigiosin.” *Chemistry A European J* **2007**, *13* (7), 1929–1945. https://doi.org/10.1002/chem.200601639.

(60) Vijay, D.; Alshamsi, N. S.; Moussa, Z.; Akhtar, M. K. Extraction of the Anticancer and Antimicrobial Agent, Prodigiosin, from Vibrio Gazogenes PB1 and Its Identification by 1D and 2D NMR. *Molecules* **2022**, *27* (18), 6030. https://doi.org/10.3390/molecules27186030.

(61) Girão, M.; Freitas, S.; Martins, T. P.; Urbatzka, R.; Carvalho, M. F.; Leão, P. N. Decylprodigiosin: A New Member of the Prodigiosin Family Isolated from a Seaweed-Associated Streptomyces. *Front. Pharmacol.* **2024**, *15*, 1347485. https://doi.org/10.3389/fphar.2024.1347485.

(62) Hifnawy, M. S.; Fouda, M. M.; Sayed, A. M.; Mohammed, R.; Hassan, H. M.; AbouZid, S. F.; Rateb, M. E.; Keller, A.; Adamek, M.; Ziemert, N.; Abdelmohsen, U. R. The Genus *Micromonospora* as a Model Microorganism for Bioactive Natural Product Discovery. *RSC Adv.* **2020**, *10* (35), 20939–20959. https://doi.org/10.1039/D0RA04025H.

(63) Bhagwat, A.; Padalia, U. Optimization of Prodigiosin Biosynthesis by Serratia Marcescens Using Unconventional Bioresources. *Journal of Genetic Engineering and Biotechnology* **2020**, *18* (1), 26. https://doi.org/10.1186/s43141-020-00045-7.

(64) Laatsch, H.; Kellner, M.; Weyland, H. Butyl-Meta-Cycloheptylprodiginine-A Revision of the Structure of the Former Ortho-Isomer. *J. Antibiot.* **1991**, *44* (2), 187–191. https://doi.org/10.7164/antibiotics.44.187.

(65) Borisova, S. A.; Zhang, C.; Takahashi, H.; Zhang, H.; Wong, A. W.; Thorson, J. S.; Liu, H. Substrate Specificity of the Macrolide-Glycosylating Enzyme Pair DesVII/DesVIII: Opportunities, Limitations, and Mechanistic Hypotheses. *Angew Chem Int Ed* **2006**, *45* (17), 2748–2753. https://doi.org/10.1002/anie.200503195.

(66) Sakai-Kawada, F. E.; Ip, C. G.; Hagiwara, K. A.; Awaya, J. D. Biosynthesis and Bioactivity of Prodiginine Analogs in Marine Bacteria, Pseudoalteromonas: A Mini Review. *Front. Microbiol.* **2019**, *10*, 1715. https://doi.org/10.3389/fmicb.2019.01715.

(67) Kim, H. J.; Lee, M.-S.; Jeong, S. K.; Lee, S. J. Transcriptomic Analysis of the Antimicrobial Activity of Prodigiosin against Cutibacterium Acnes. *Sci Rep* **2023**, *13* (1), 17412. https://doi.org/10.1038/s41598-023-44612-7.

(68) Danevčič, T.; Borić Vezjak, M.; Zorec, M.; Stopar, D. Prodigiosin - A Multifaceted Escherichia Coli Antimicrobial Agent. *PLoS ONE* **2016**, *11* (9), e0162412. https://doi.org/10.1371/journal.pone.0162412.

(69) Setiyono, E.; Adhiwibawa, M. A. S.; Indrawati, R.; Prihastyanti, M. N. U.; Shioi, Y.; Brotosudarmo, T. H. P. An Indonesian Marine Bacterium, *Pseudoalteromonas Rubra* , Produces Antimicrobial Prodiginine Pigments. *ACS Omega* **2020**, *5* (9), 4626–4635. https://doi.org/10.1021/acsomega.9b04322.

(70) García-Valverde, M.; Alfonso, I.; Quiñonero, D.; Quesada, R. Conformational Analysis of a Model Synthetic Prodiginine. *J. Org. Chem.* **2012**, *77* (15), 6538–6544. https://doi.org/10.1021/jo301008c.

(71) Yan, J.; Yin, Q.; Nie, H.; Liang, J.; Liu, X.-R.; Li, Y.; Xiao, H. Prodigiosin as an Antibiofilm Agent against Multidrug-Resistant *Staphylococcus Aureus*. *Biofouling* **2023**, *39* (4), 444–458. https://doi.org/10.1080/08927014.2023.2226613.

Appendix A.

Supplemental Figures and Tables

Table A.1. MS features of interest in communities consisting of only NP prefractions. Bolded prefractions include those selected for further investigation of these highlighted MS features.

Community	MS Feature	Retention Time (Rt)	Activity Score	Cluster Score	Prefractions
7	543.3633	4.9543	26.35	0.78	**RLUS-2028D** RLUS-2067E
	913.6287	3.3085	4.39	0.32	**RLUS-1885D&E**
	400.2739	3.307	4.39	0.32	
	282.1838	4.0282	27.08	0.48	**RLUS-2028D** RLUS-2078D
	432.3089	3.7923	5.32	0.54	RLUS-1852D **RLUS-2024D** **RLUS-2064D**
10	411.1057	2.412	8.98	0.97	**RLUS-1758B&C**
	430.1208	2.2722	7.19	0.37	**RLUS-1758B&C** **RLUS-1885B**
	430.1212	2.4126	6.52	0.32	**RLUS-1758B&C** **RLUS-1885B, C&D**
	445.1131	2.089	6.44	0.30	**RLUS-1758C** **RLUS-1885B&C**
	659.8565	1.8996	3.06	0.31	RLUS-1628B RLUS-1734B **RLUS-1806B** RLUS-1837B RLUS-2058B RLUS-2173B
	682.2828	4.2817	21.28	0.72	**RLUS-1505C** **RLUS-1530C**
	622.2613	3.8606	7.33	0.47	**RLUS-1505C** RLUS-1819B&D
13	467.2759	4.1273	2.4	0.57	**RLUS-2064C&D**
	426.1954	2.3904	2.4	0.57	
24	945.4529	2.3911	14.02	0.33	RLUS-1498E **RLUS-1520E** RLUS-1545D&E RLUS-1568E RLUS-1597D&E

					RLUS-1608E
27	424.2579	5.138	8.13	0.82	RLUS-2052E **RLUS-2085E**
	248.2012	2.993	6.97	0.87	**RLUS-2024E** RLUS-2035E RLUS-2052E **RLUS-2085E**
	447.2239	4.8755	7.68	0.86	**RLUS-2024E** RLUS-2035F RLUS-2052E
28	639.3849	3.7372	4.0	0.97	**RLUS-1806D** **RLUS-1806E**
	509.2976	4,289	4.0	0.97	**RLUS-1806D** **RLUS-1806E**
	495.281	3.887	4.0	0.97	**RLUS-1806D** **RLUS-1806E**
30	402.2602	5.0917	16.14	0.61	RLUS-1703E
	398.4397	5.3528	16.14	0.61	**RLUS-1726D**
31	931.1907	3.6968	21.48	0.99	**RLUS-1623C** **RLUS-1623D**
	353.1005	3.573	16.15	0.71	**RLUS-1623C** **RLUS-1623D** **RLUS-1623E**
33	552.3862	3.1925	5.85	0.7	**RLUS-1459D** **RLUS-1459E**
35	745.4002	2.9828	4.31	0.7	**RLUS-1595C** **RLUS-1595D**

Table A.2. ResistoMAP Profiling target panel consisting of 29 drug-resistant *E. coli* strains.

Strain	Gene	Mutation	Protein	Selection Antibiotic	Drug Mechanism
S83L	*gyrA*	S83L	DNA gyrase subunit A	Ciprofloxacin	DNA gyrase
Cip1		S83A			
Cip3		D87Y			
Cip5		D87G			
Cip2KB	*gyrB*	L509G	DNA gyrase subunit B		
Cip15KB		S464Y			
NA3		D426N		Nalidixic Acid	
Cip8	*marR*	R77H	Multiple antibiotic resistance protein	Ciprofloxacin	Multidrug resistance
Tet8		H120fs		Tetracycline	
Cm2	*acrR*	E118*		Chloramphenicol	

Cm3		A151fs	HTH-type transcriptional regulator AcrR		
Rif1	*rpoB*	I572L	DNA-directed RNA polymerase subunit beta	Rifampicin	RNA polymerase
Rif7		I572S			
Rif11		D516N			
Cef1	*evnZ*	T402M	Sensor histidine kinase EnvZ	Ceftazidime	Cell wall synthesis
Cef8		P248S			
Cef6	*rfaH*	W4*	Transcription antitermination protein RfaH		
Cef7	*rfaG*	E289fs	Lipopolysaccharide core biosynthesis protein RfaG		
Gn12	*cyoA*	I127fs	Cytochrome bo(3) ubiquinol oxidase subunit 2	Gentamicin	Protein synthesis
Kn14		W82*		Kanamycin	
Kn6	*ubiB*	Y176*	Probable protein kinase UbiB		
Gn14	*ubiF*	D342H	3-demethoxyubiquinol 3-hydroxylase	Gentamicin	
Kn15		Q120*		Kanamycin	
Str1	*rpsL*	K43R	30S ribosomal protein S12	Streptomycin	
Str3		K43N			
Str4		P91Q			
RK2	N/A	N/A	N/A	Kanamycin	Plasmidborne multidrug resistance
AC29-1	N/A	N/A	N/A	Ampicillin	
AC30-1	N/A	N/A	N/A		

Appendix B.

Supplemental Mass Spectrometry and NMR Spectra

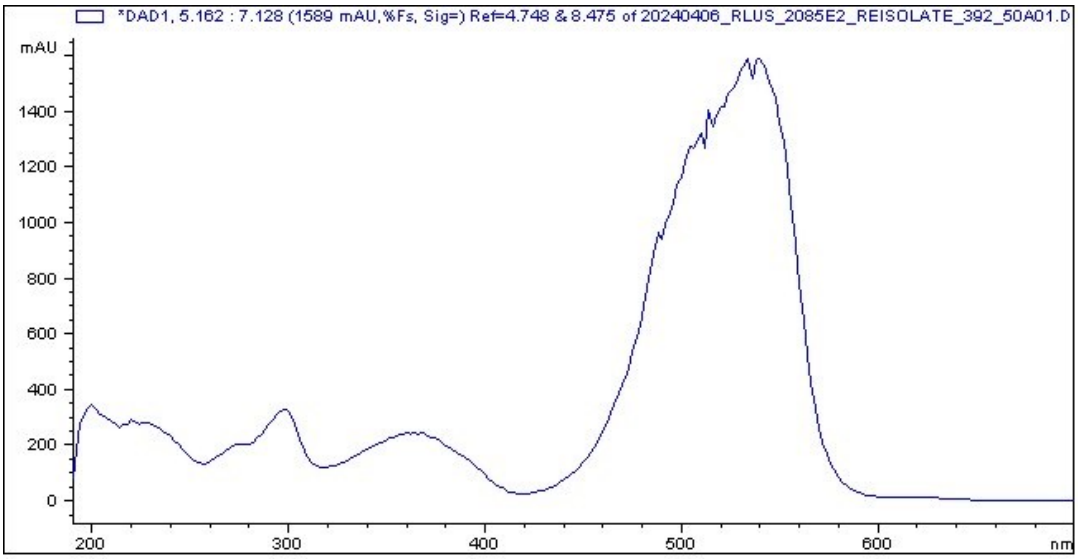


Figure B.1. UV Absorption of 2 used as the guide for isolation.

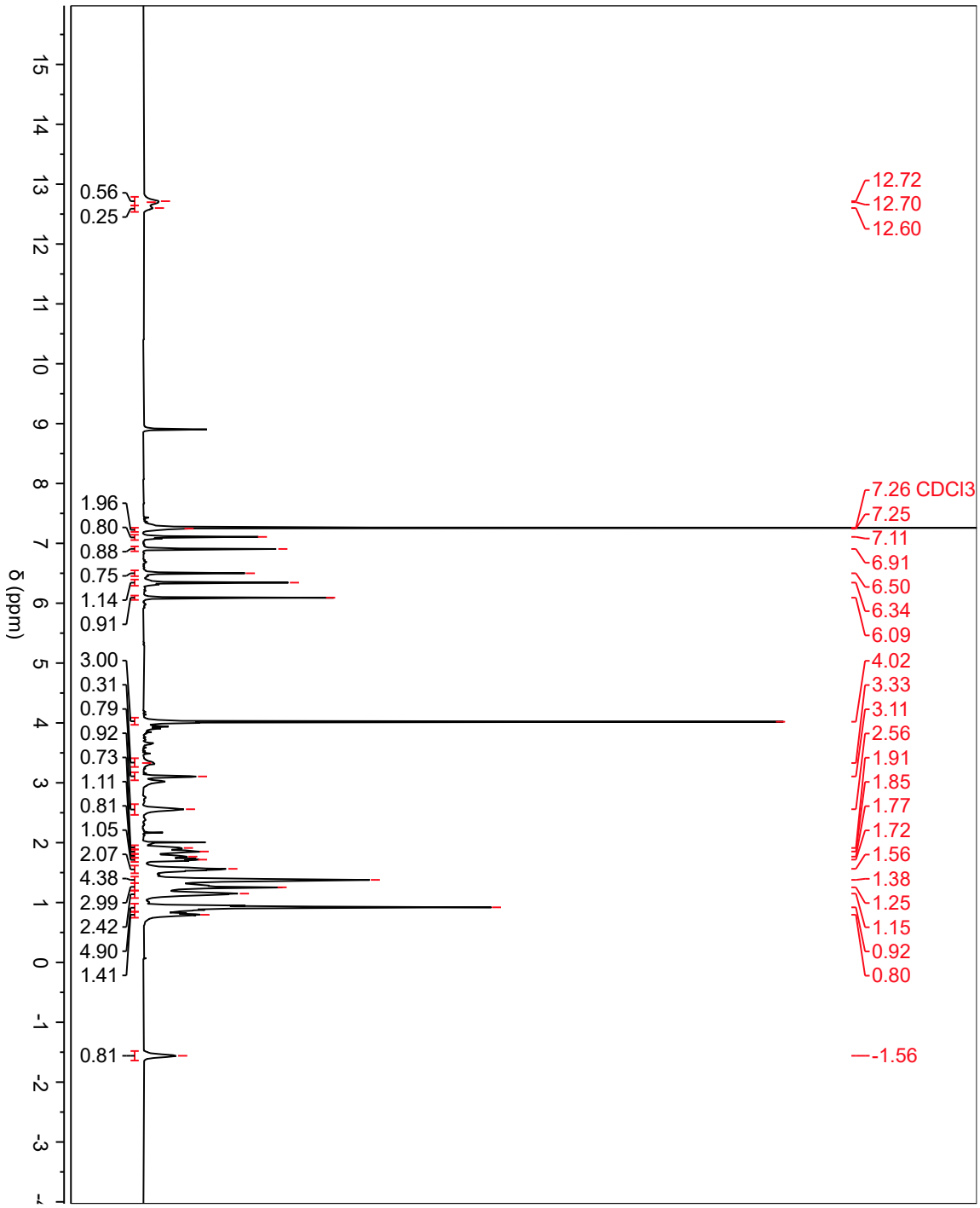


Figure B.2. ^{1}H-NMR spectrum of streptorubin B (2) at 600 MHz in $CDCl_3$-d.

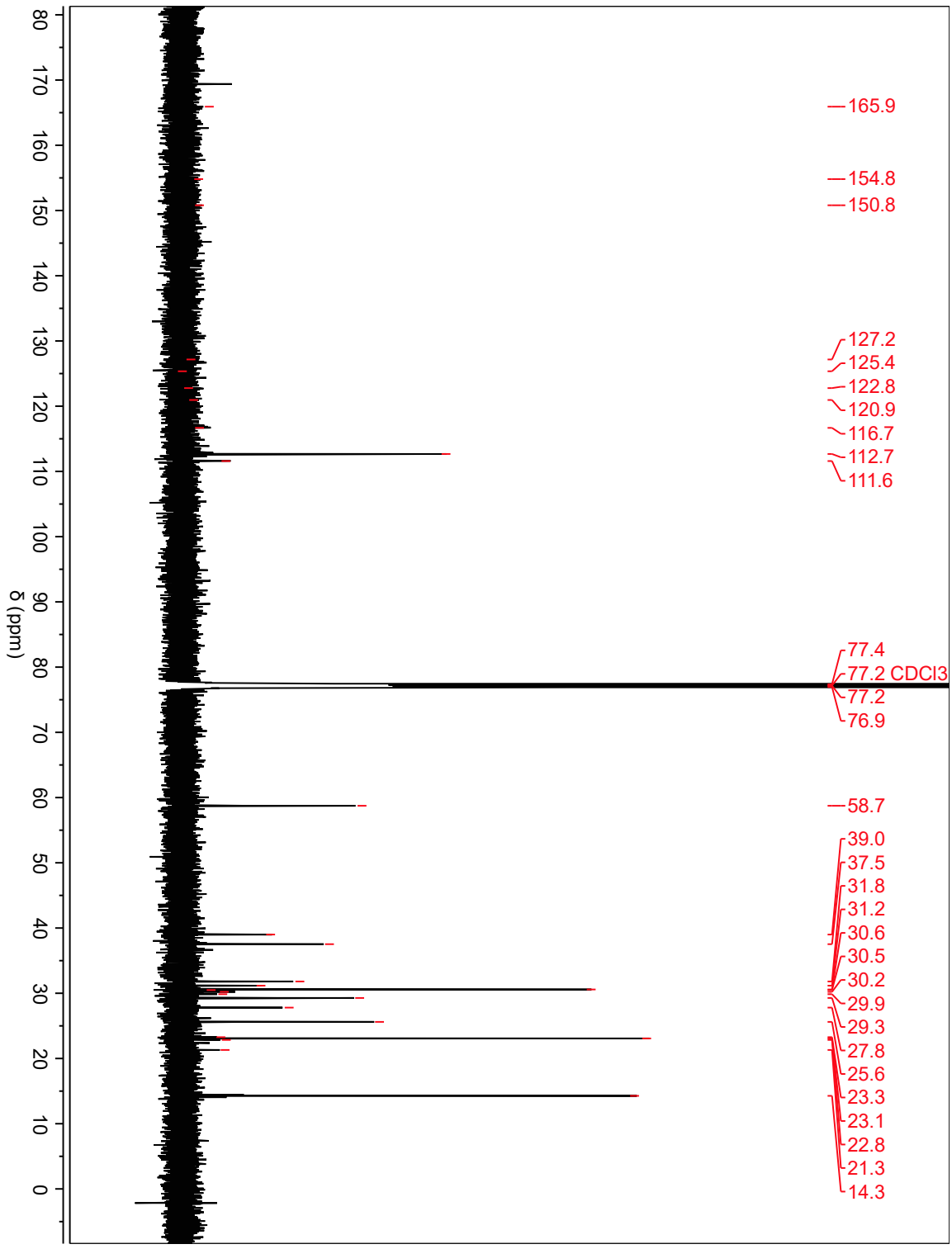


Figure B.3. ^{13}C-NMR spectrum of 2 at 600 MHz in $CDCl_3$-d.

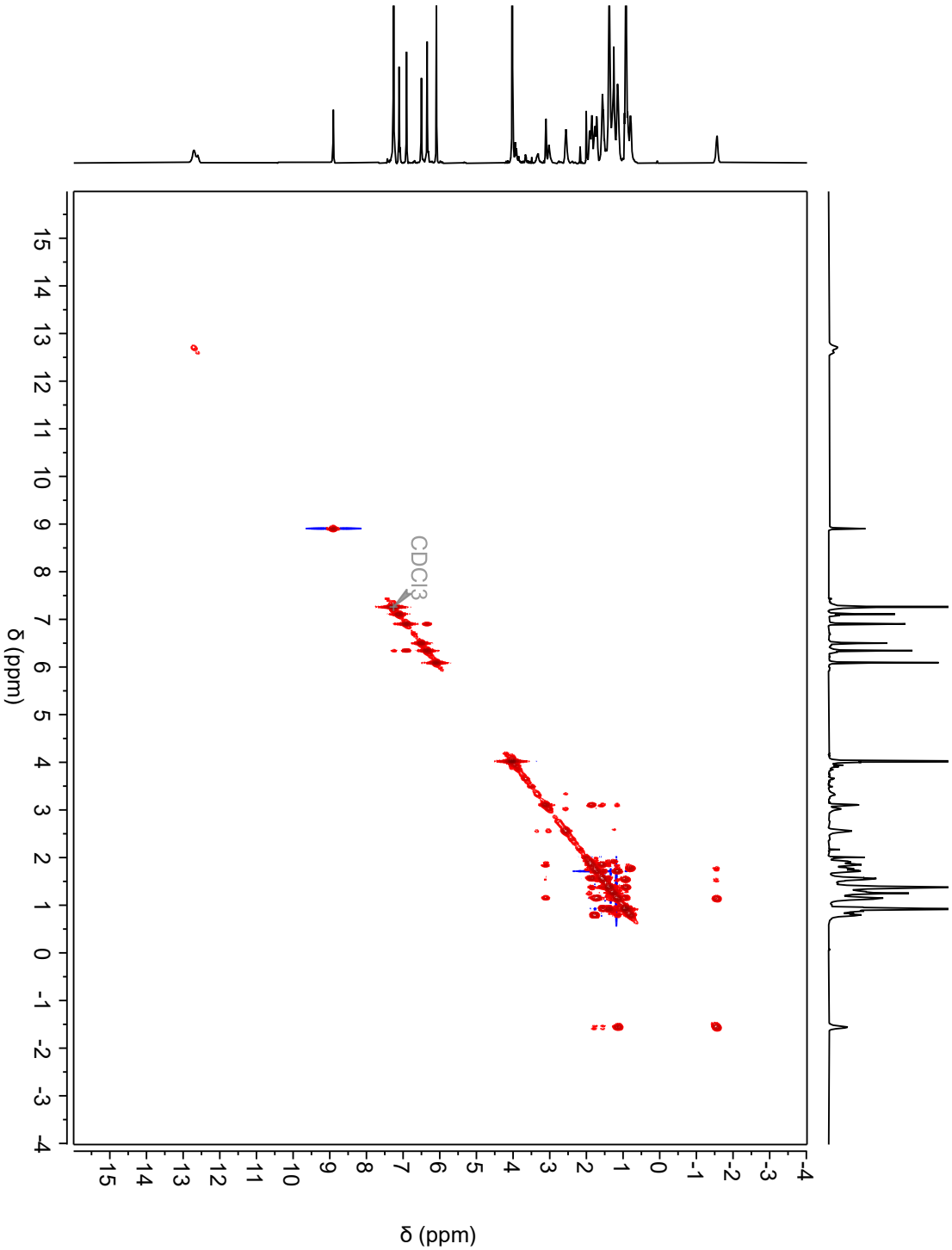


Figure B.4. gCOSY spectrum of 2 at 600 MHz in $CDCl_3$-d

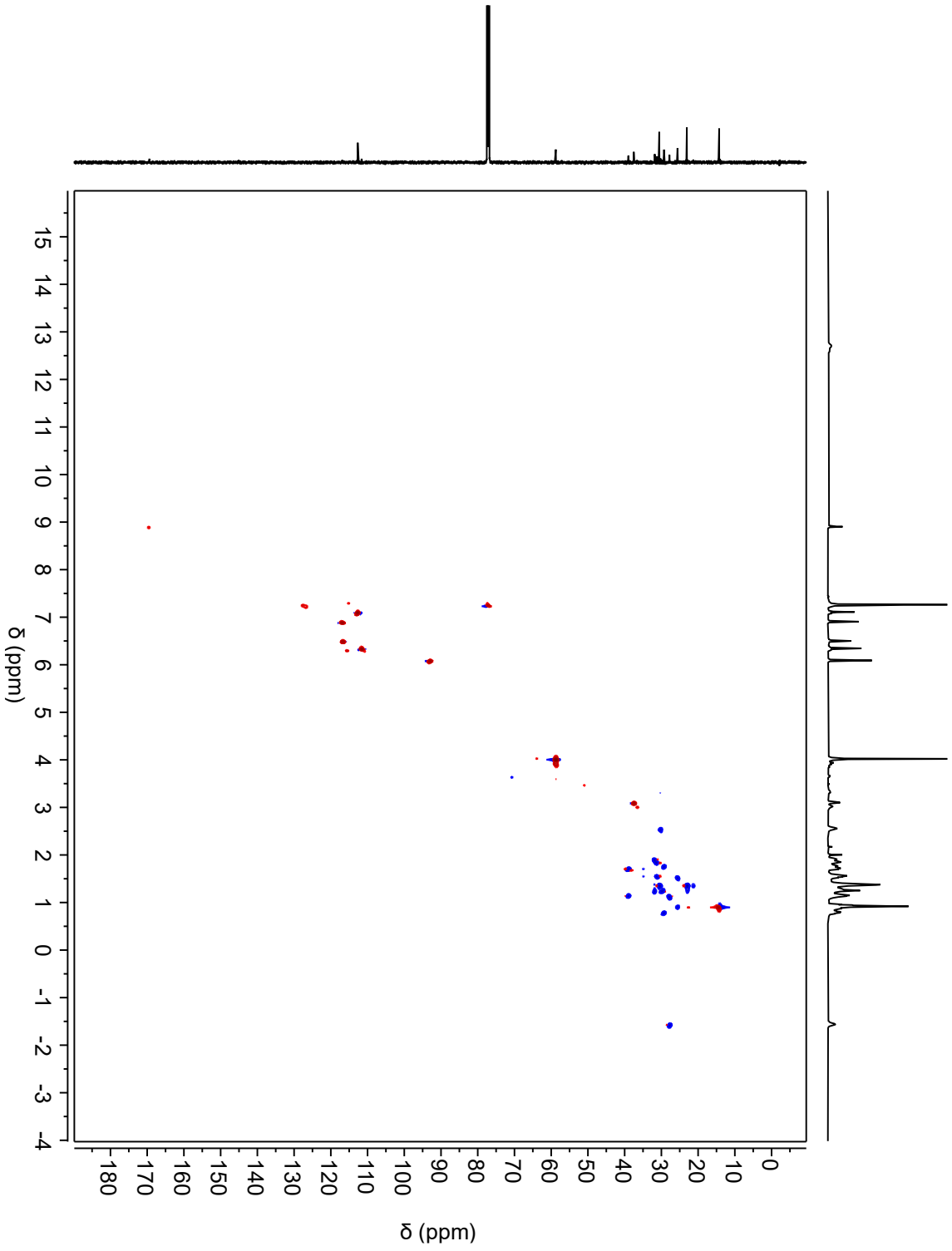


Figure B.5. gHSQC spectrum of 2 at 600 MHz in $CDCl_3$-d.

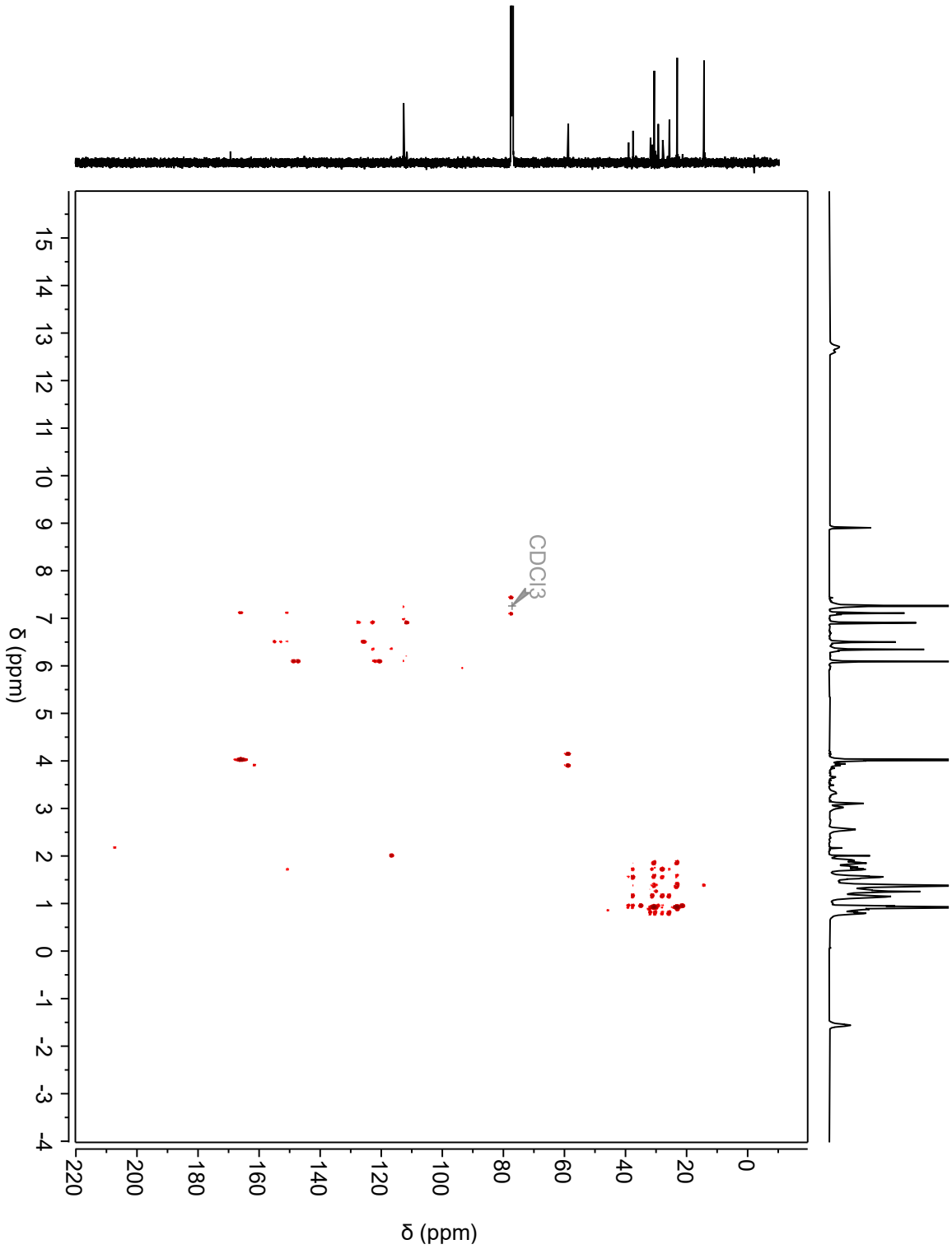


Figure B.6. gHMBC spectrum of 2 at 600 MHz in $CDCl_3$-d

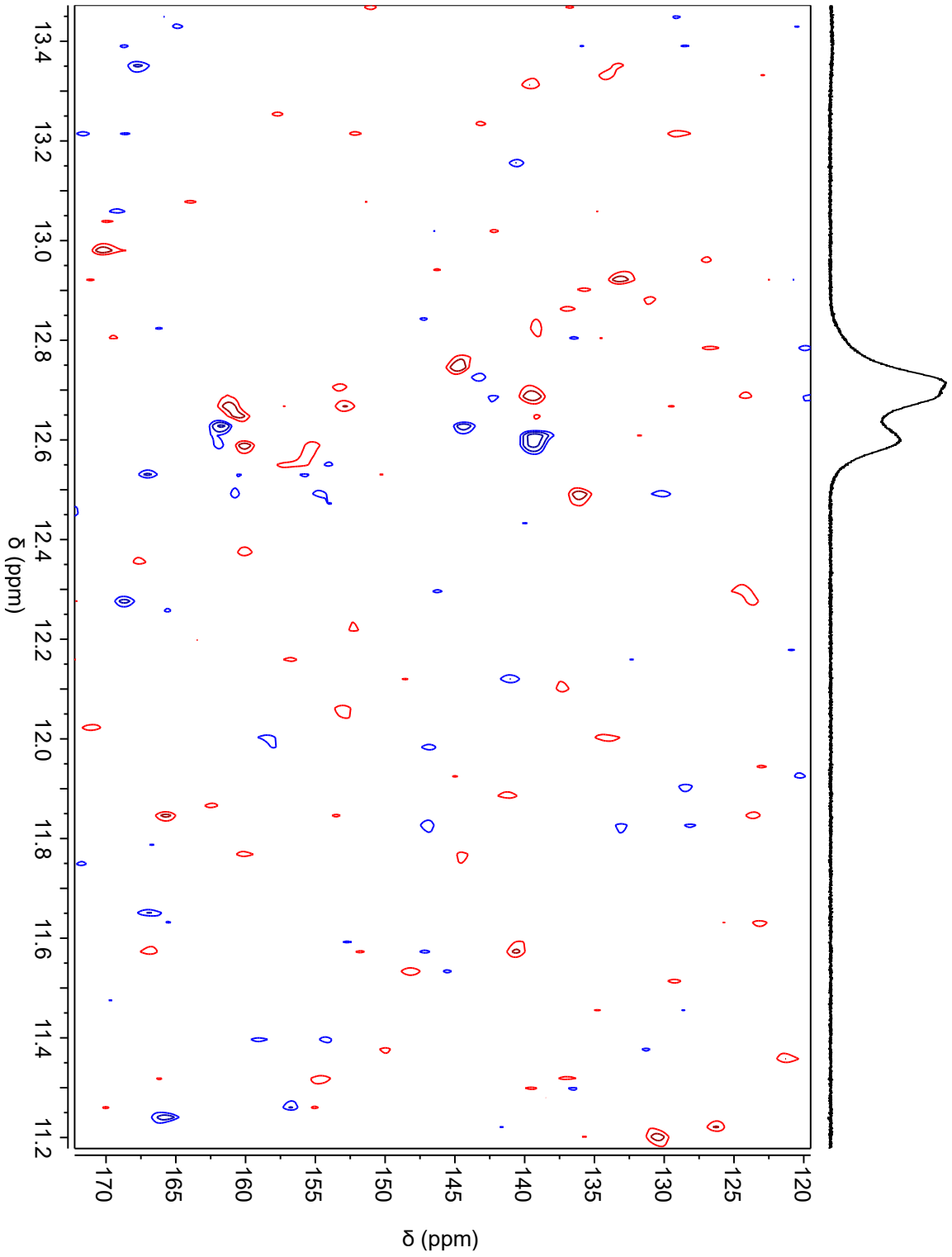


Figure B.7. ^{15}N-HSQC spectrum of 2 at 600 MHz in $CDCl_3$-d

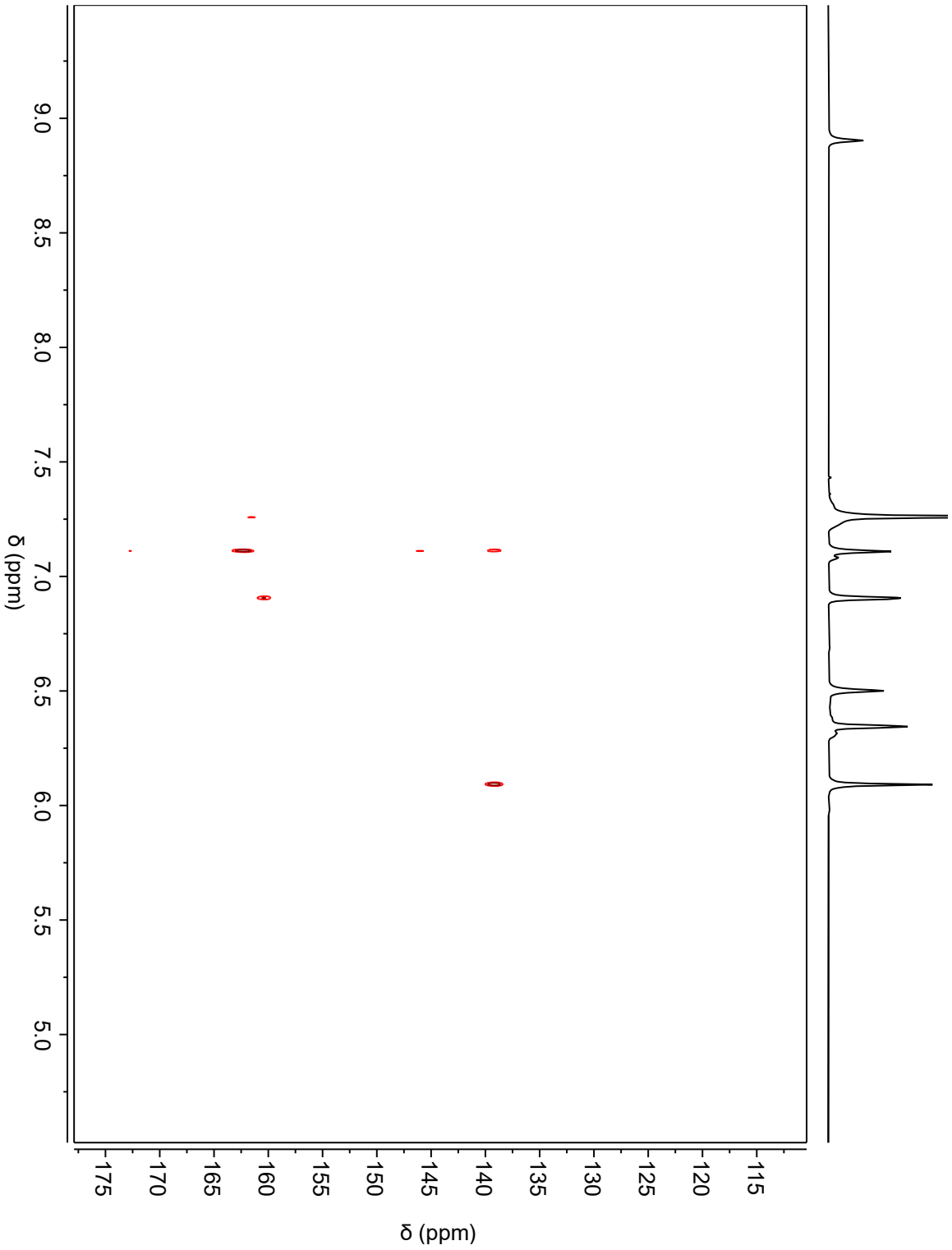


Figure B.8. **^{15}N-HMBC spectrum of 2 at 600 MHz in $CDCl_3$-d.**

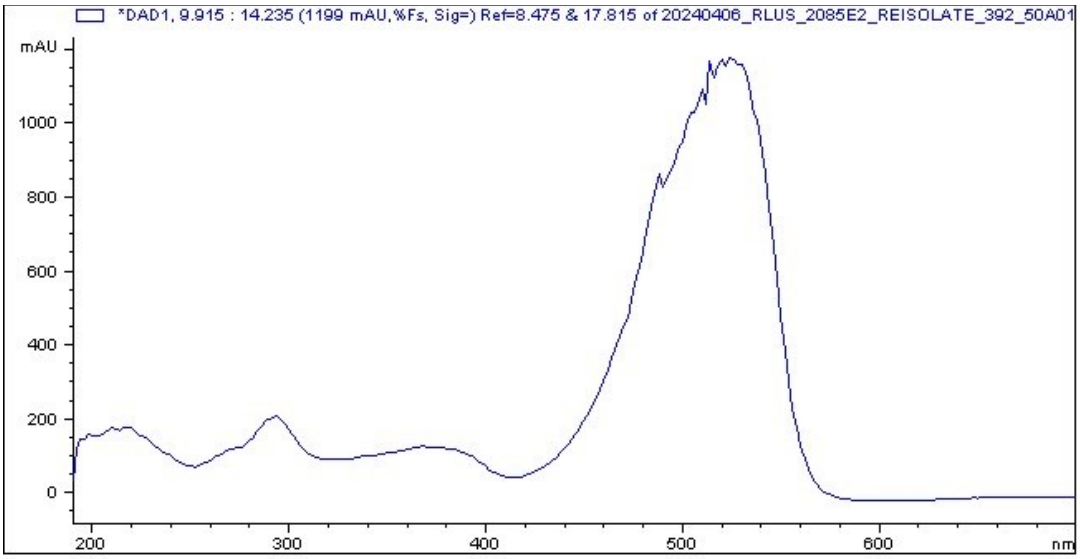


Figure B.9. UV Absorption of 4 used as the guide for isolation.

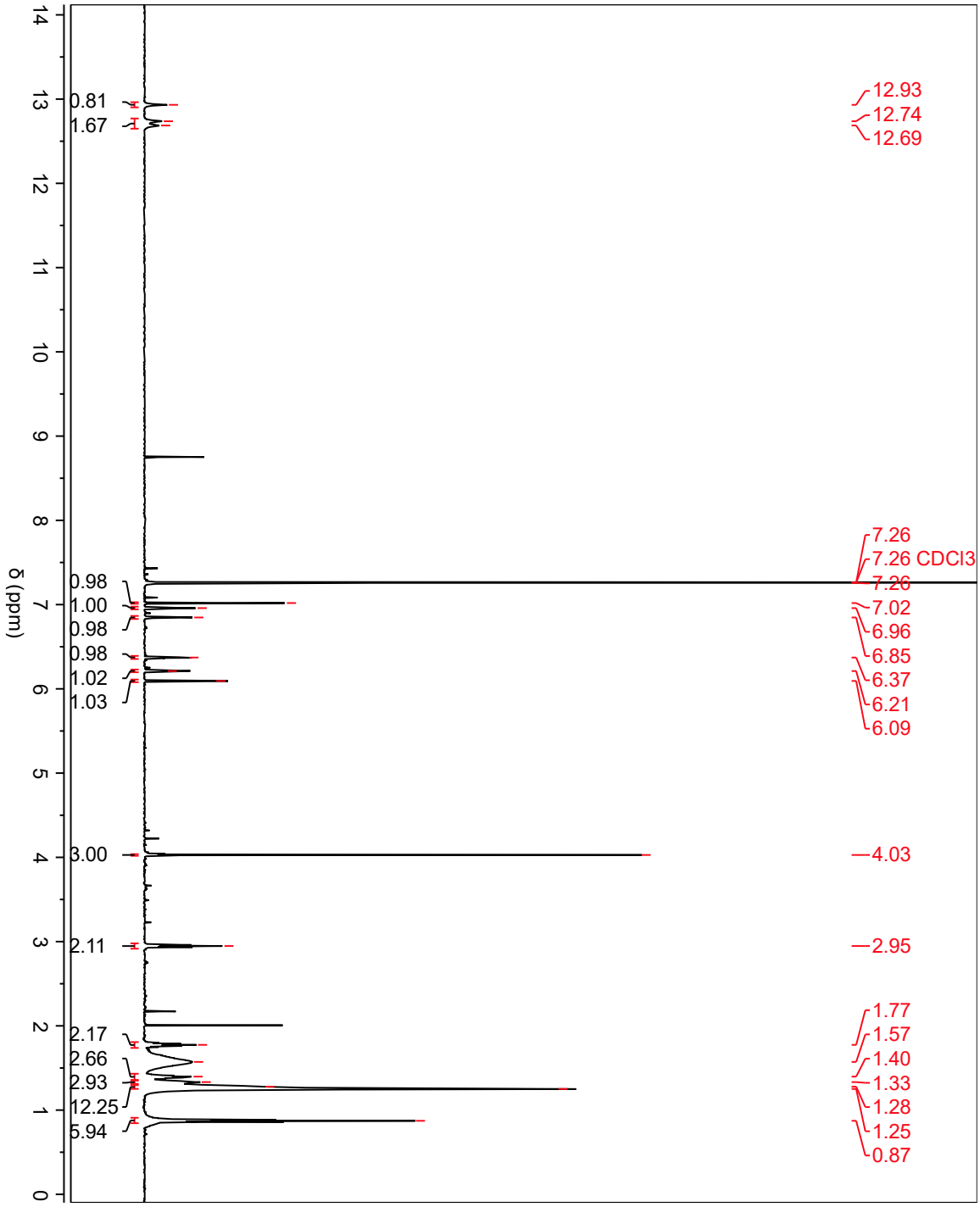


Figure B.10. ^{1}H-NMR spectrum of undecylprodigiosin (4) at 600 MHz in $CDCl_3$-d.

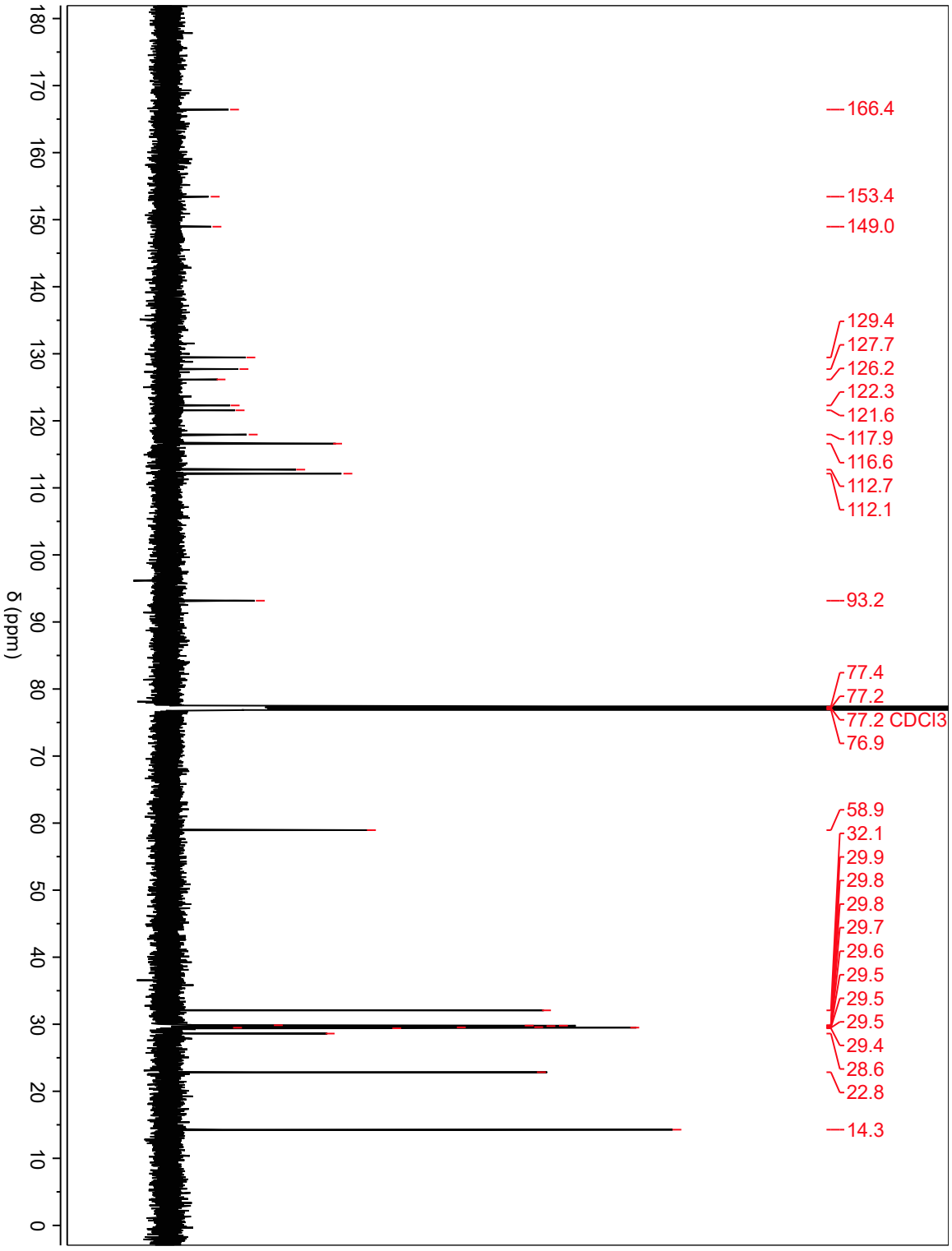


Figure B.11. ^{13}C-NMR spectrum of 4 at 600 MHz in $CDCl_3$-d.

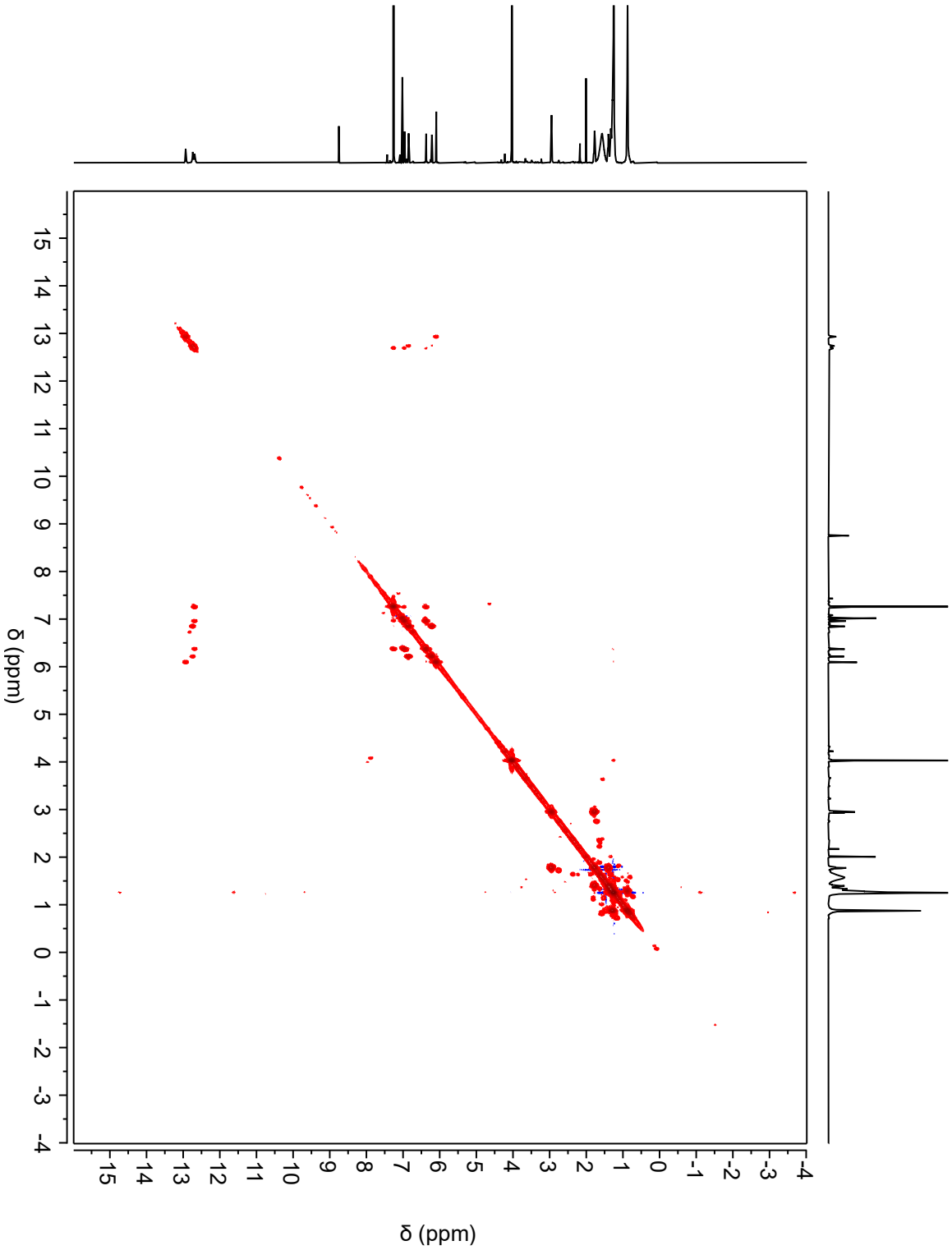


Figure B.12. gCOSY spectrum of 4 at 600 MHz in $CDCl_3$-d.

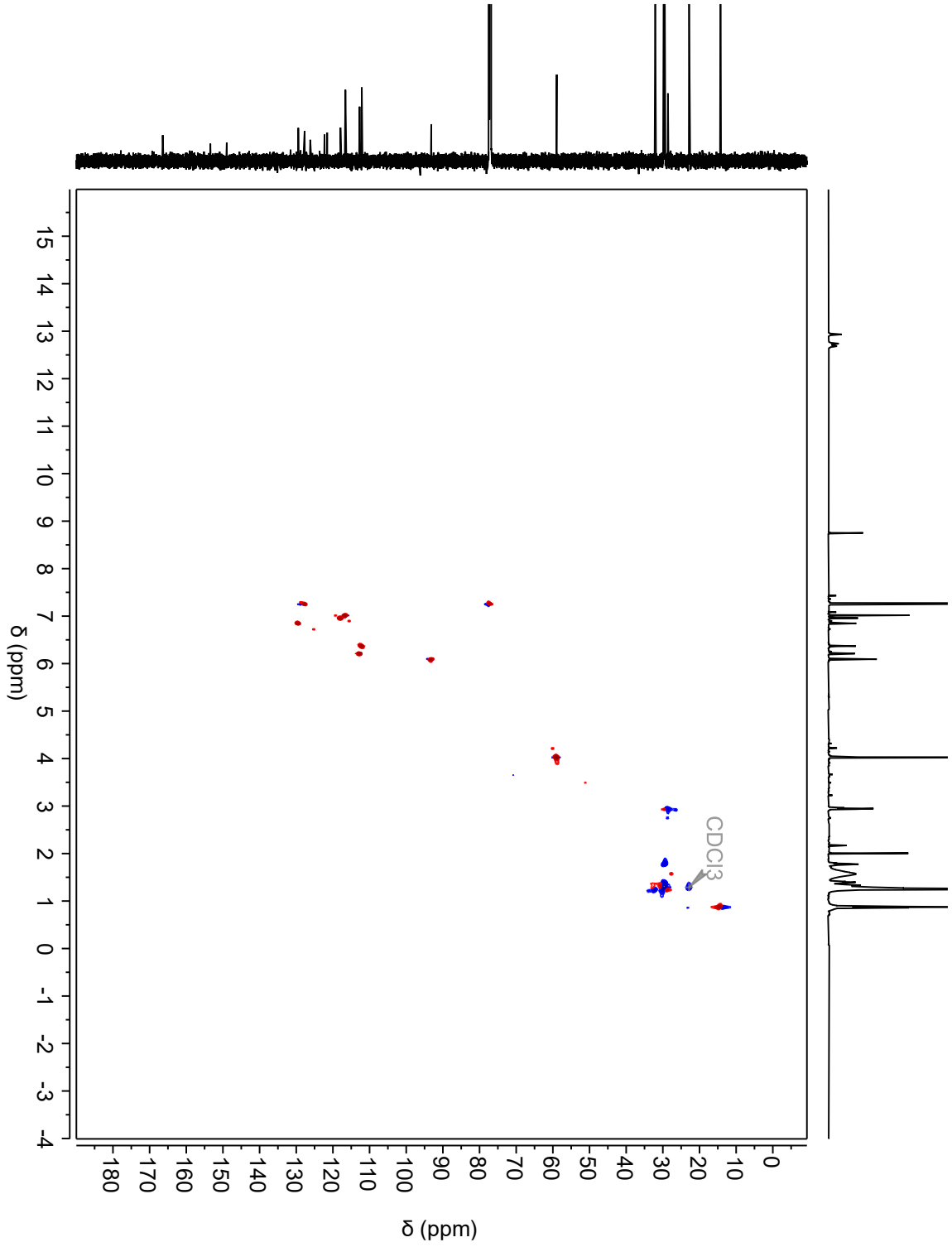


Figure B.13. gHSQC spectrum of 4 at 600 MHz in $CDCl_3$-d.

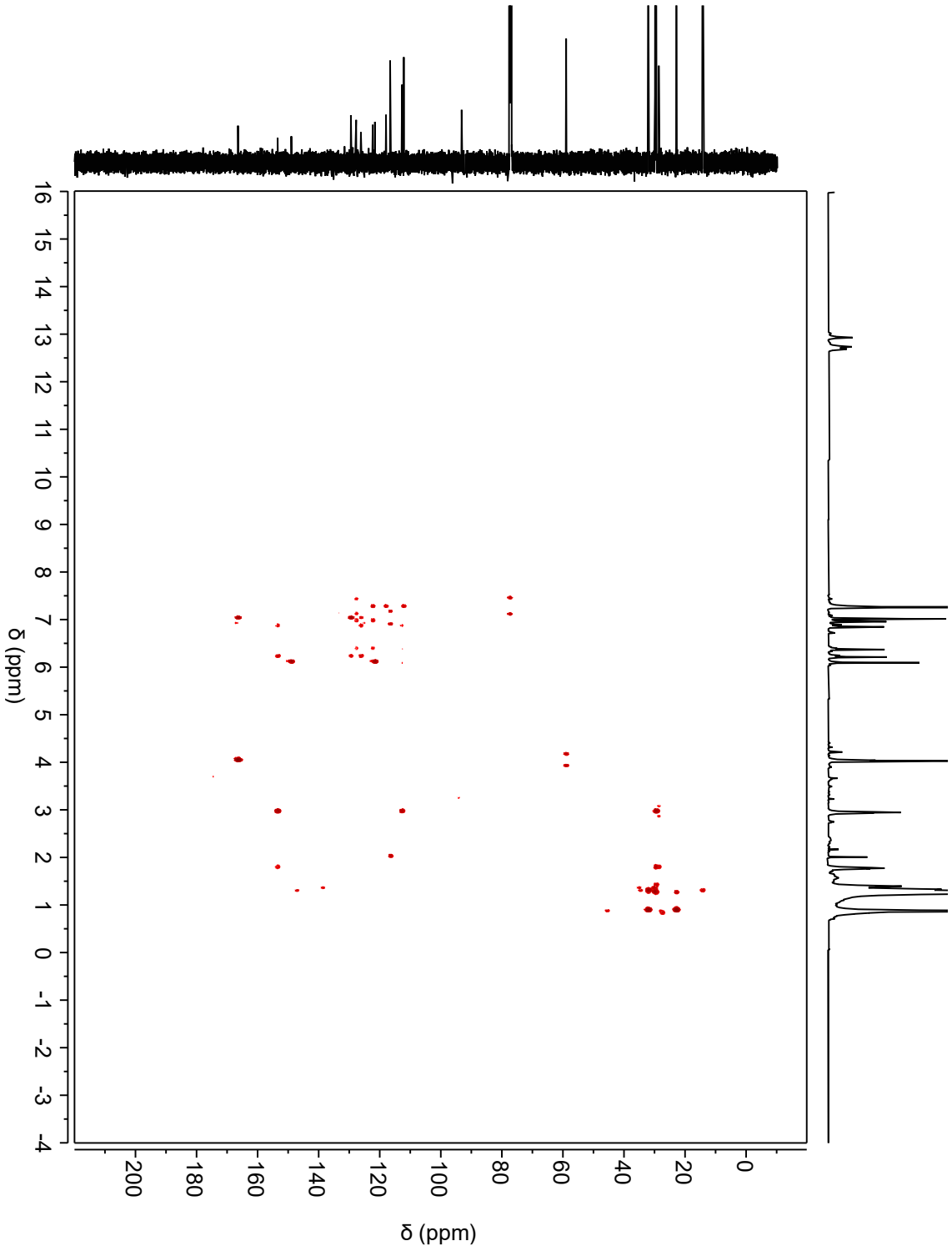


Figure B.14. gHMBC spectrum of 4 at 600 MHz in $CDCl_3$-d.

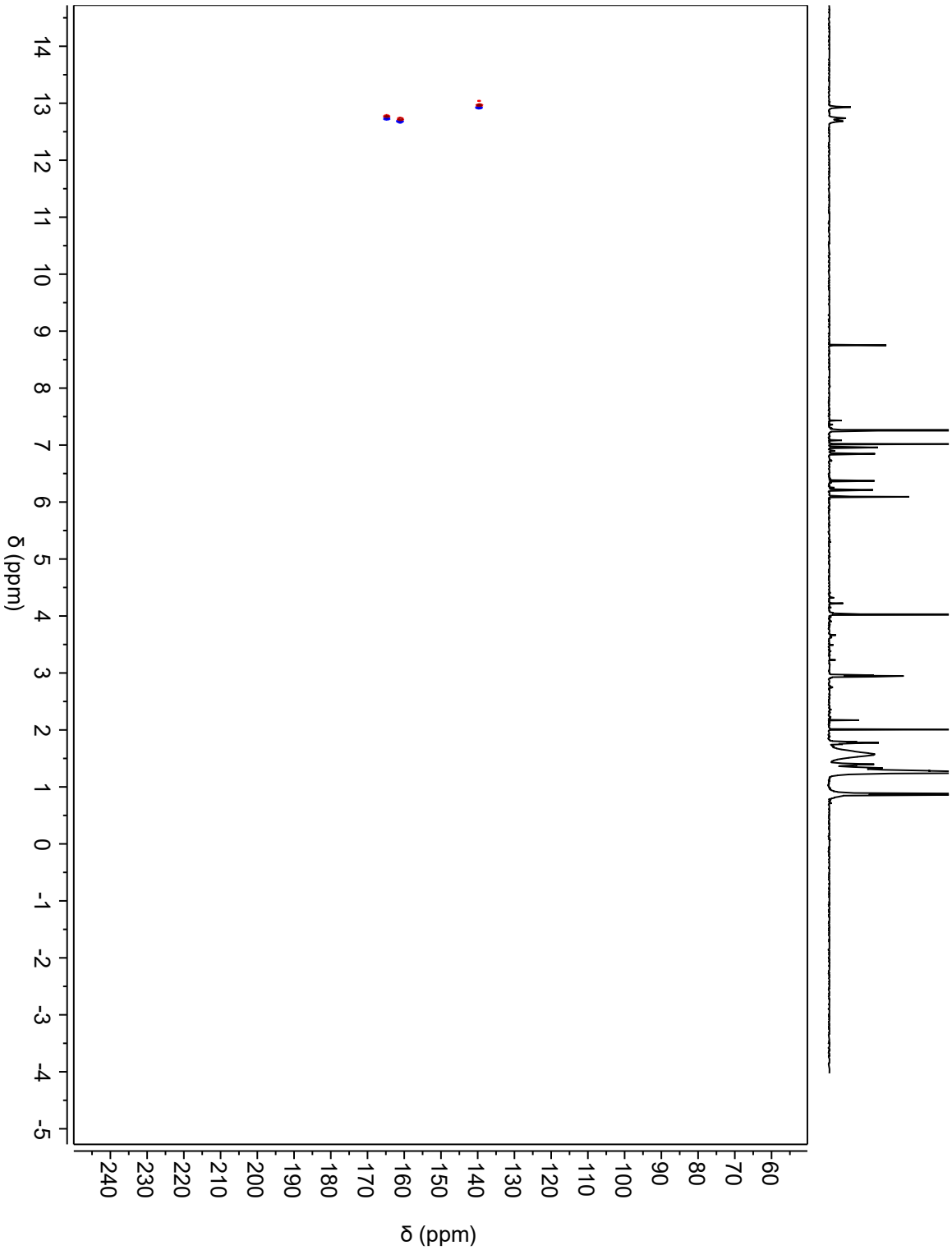


Figure B.15. ^{15}N-HSQC spectrum of 4 at 600 MHz in $CDCl_3$-d.

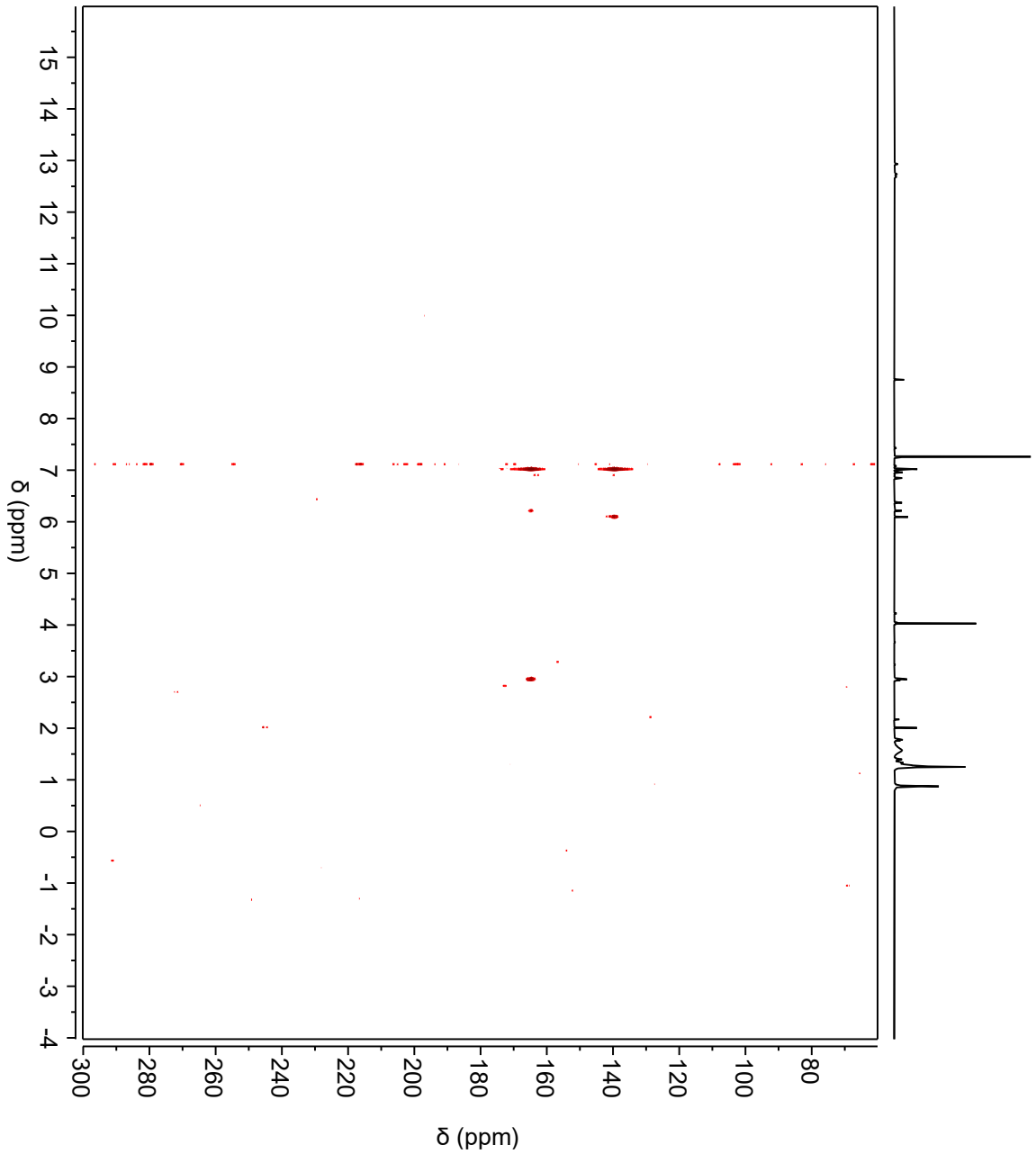


Figure B.16. ^{15}**N-HMBC spectrum of 4 at 600 MHz in** $CDCl_3$**-d.**

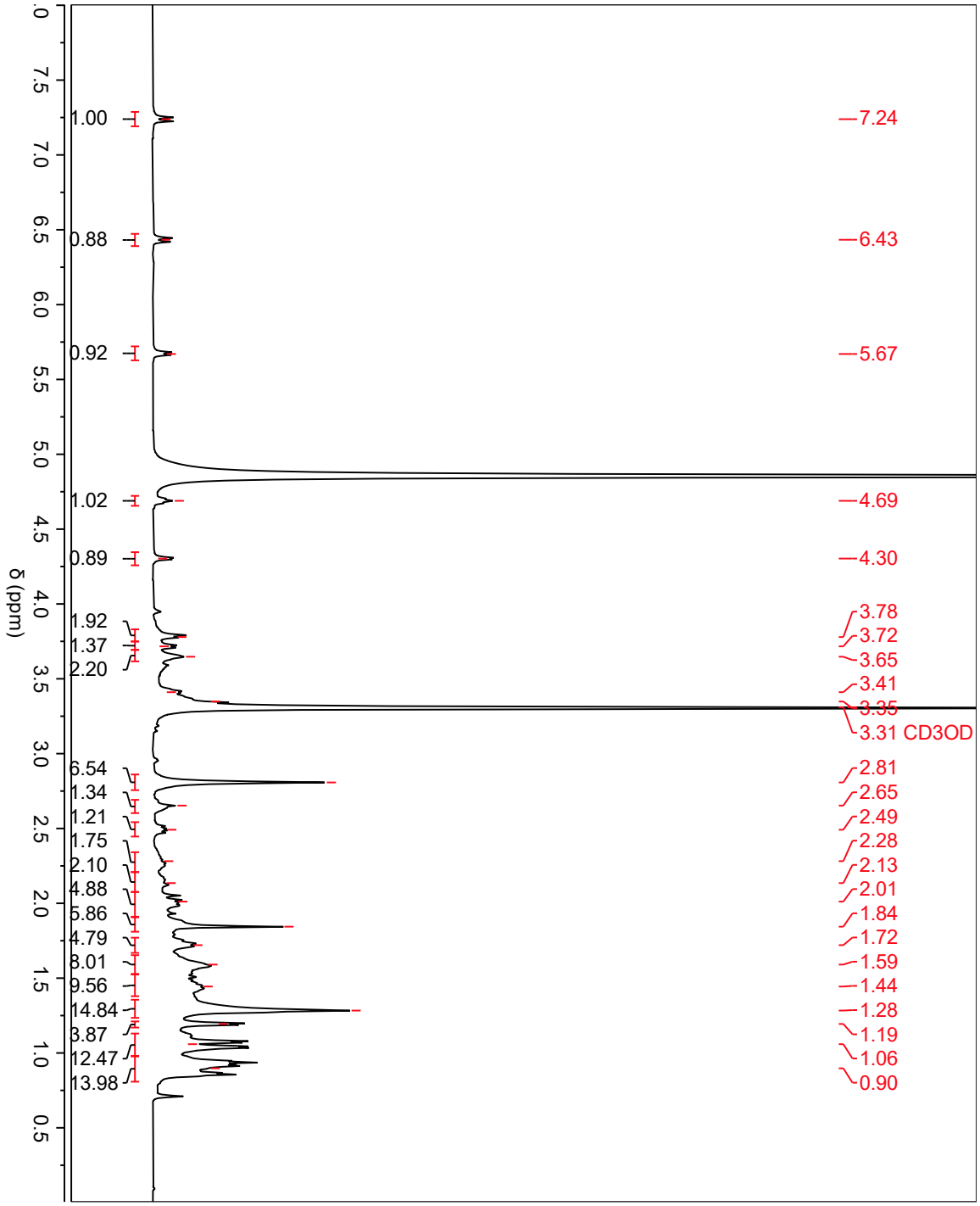


Figure B.17 **^{1}H-NMR spectrum of RLUS-2204D 552.40 *m/z* (X3) at 600 MHz in MeOD.**

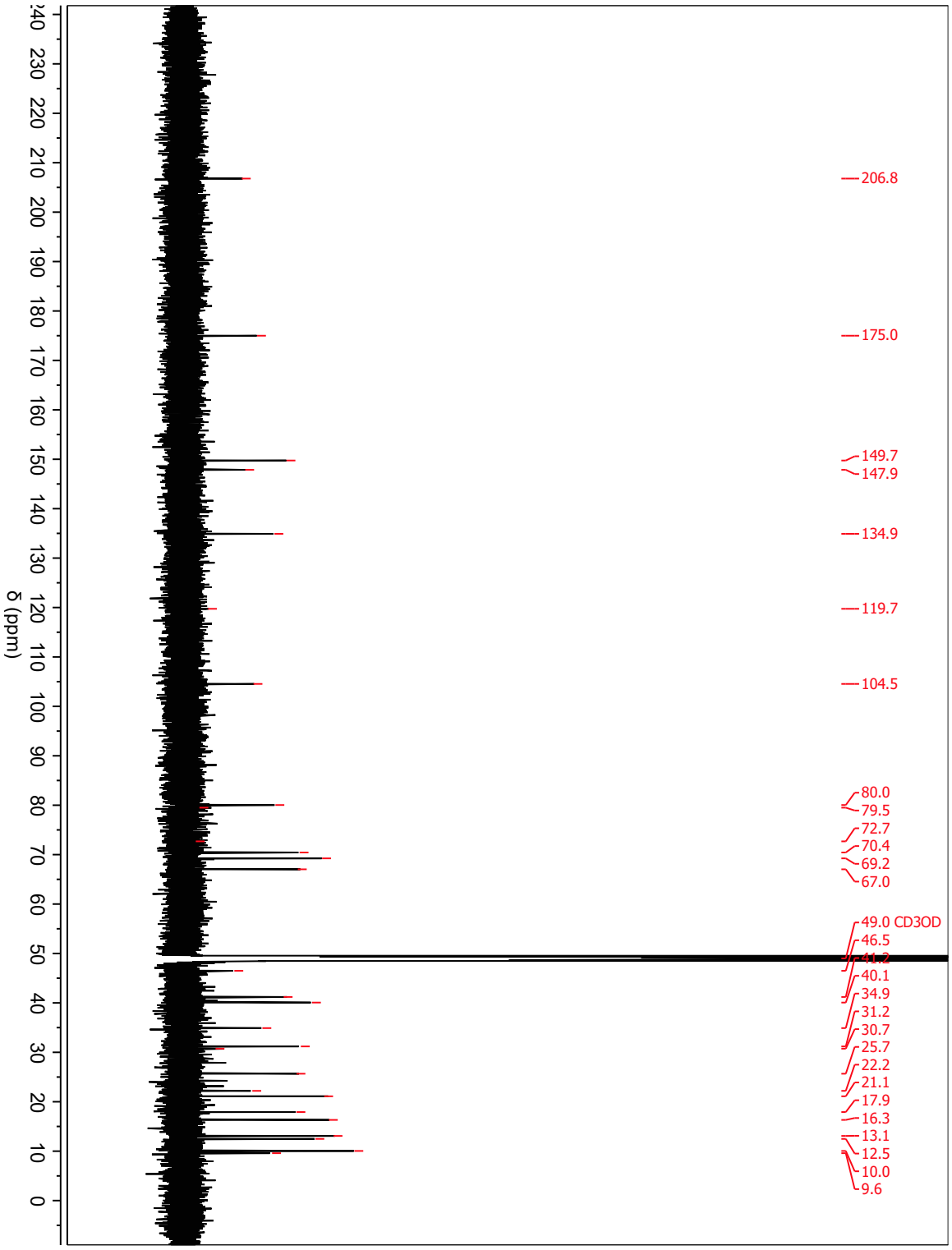


Figure B.18. ^{13}C-NMR spectrum of X3 at 600 MHz in MeOD.

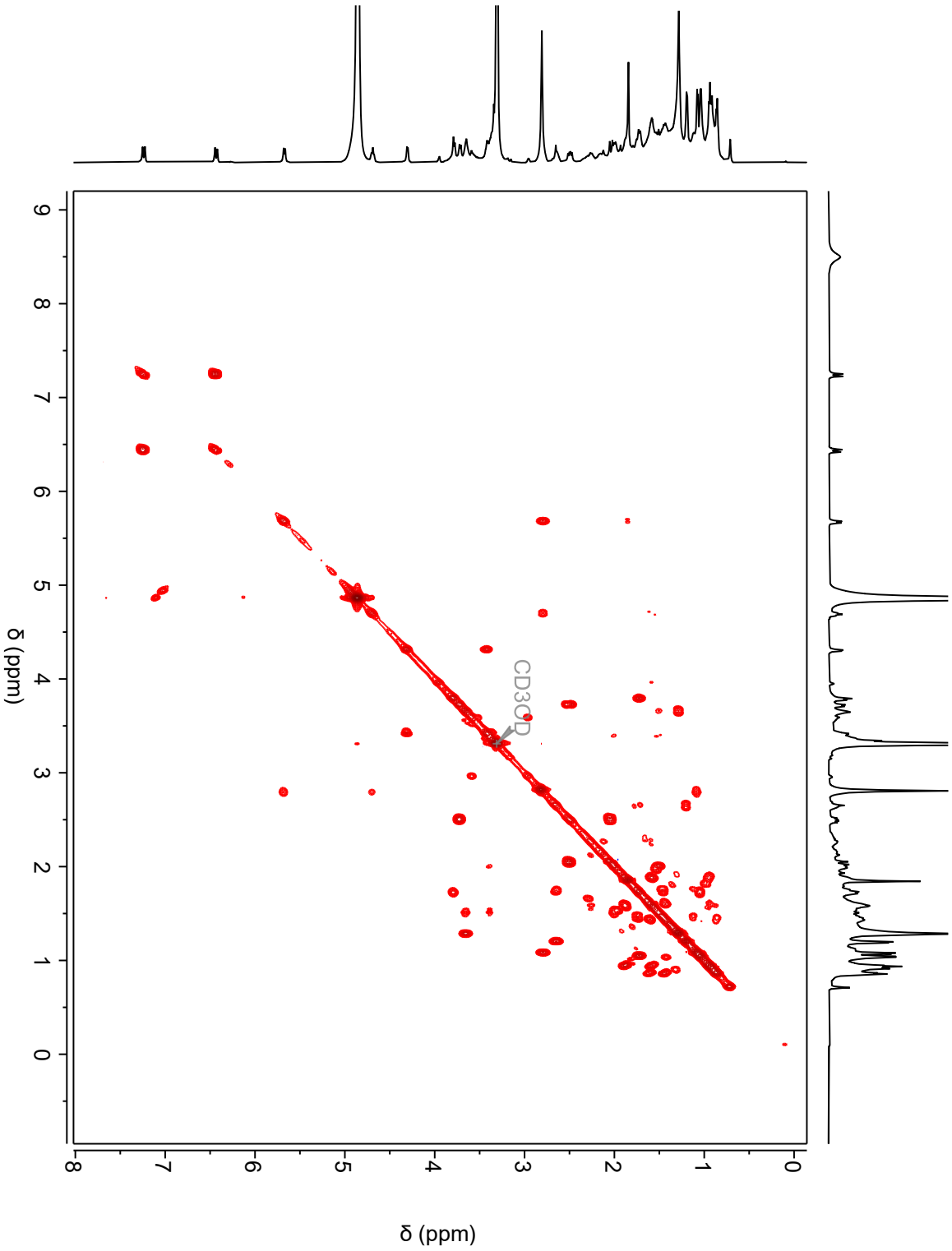


Figure B.19. gCOSY spectrum of X3 at 600 MHz in MeOD.

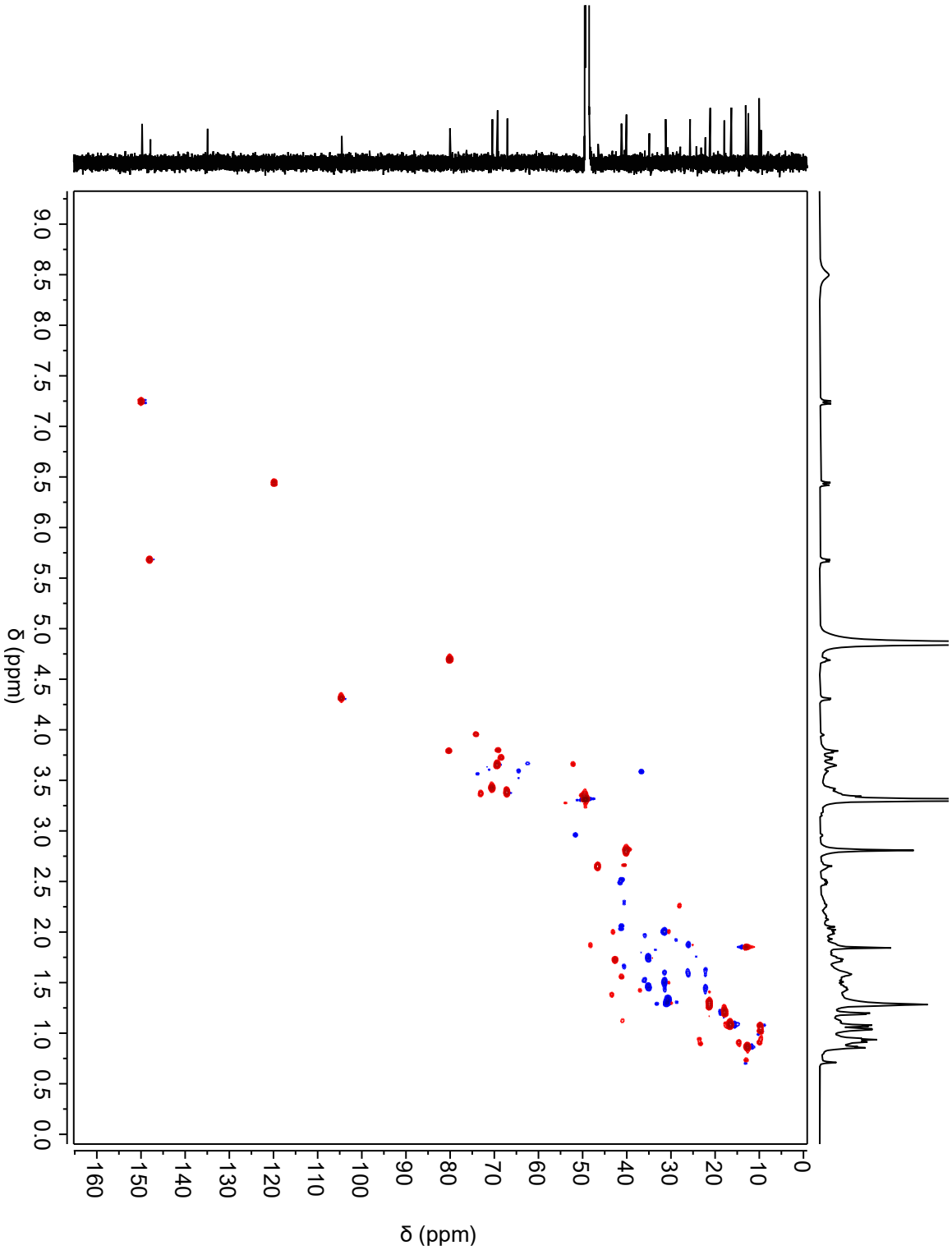


Figure B.20. gHSQC spectrum of X3 at 600 MHz in MeOD.

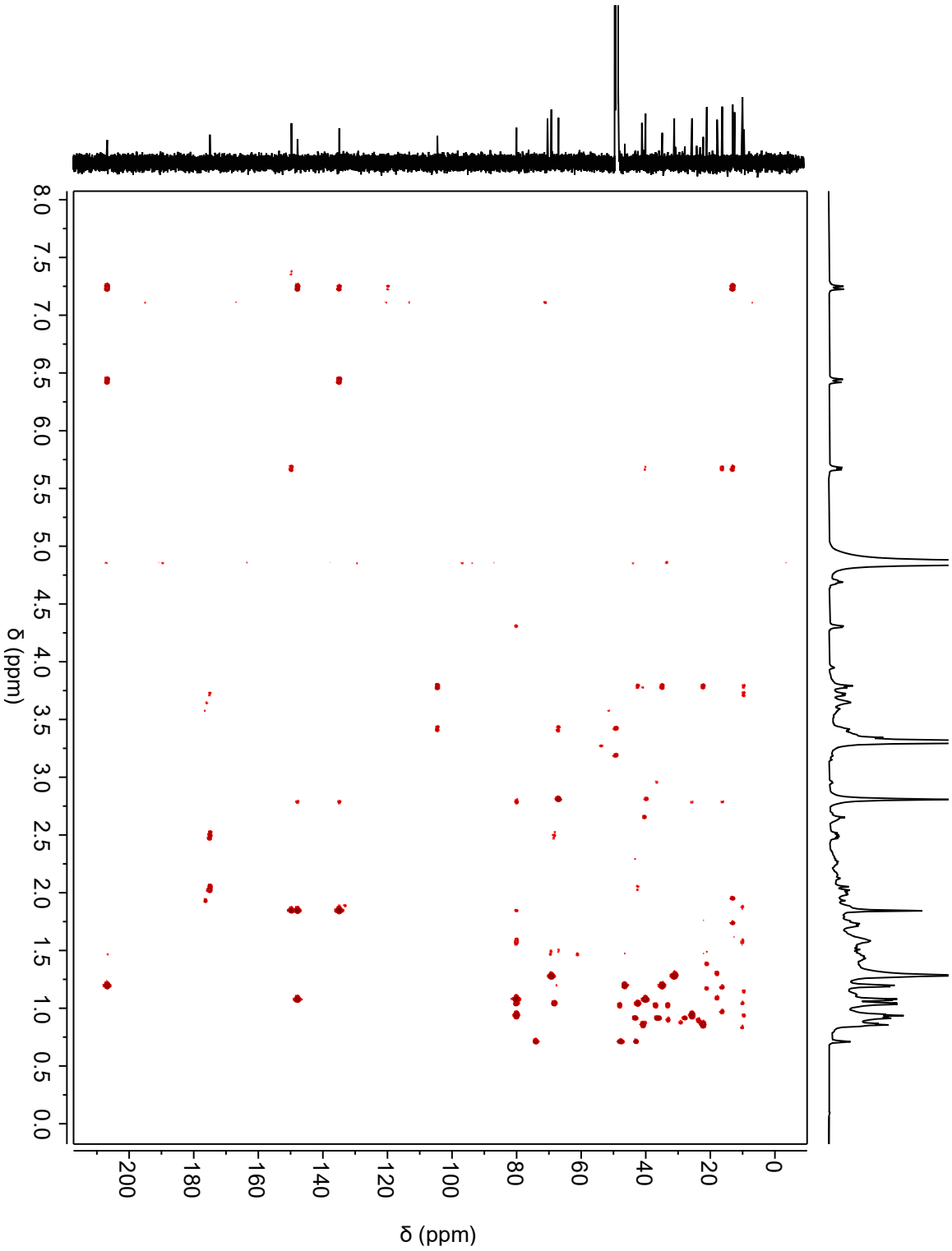


Figure B.21. gHMBC spectrum of X3 at 600 MHz in MeOD.