Chemical and Genetic Investigations of Marine and Terrestrial Bacteria

towards Bioactive Natural Product Discovery

by

Kalindi Dasi Morgan

B.Sc., The University of British Columbia, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Chemical and Genetic Investigations of Marine and Terrestrial Bacteria towards Bioactive Natural Product Discovery

submitted by   Kalindi Dasi Morgan   in partial fulfillment of the requirements for

the degree of   Doctor of Philosophy

in   Chemistry

Examining Committee:

Raymond Andersen, Professor, Joint appointment with Chemistry & Earth, Ocean and Atmospheric Sciences, UBC
Co-supervisor

Katherine Ryan, Associate Professor, Chemistry, UBC
Co-supervisor

Martin Tanner, Professor, Chemistry, UBC
Supervisory Committee Member

Andrew MacFarlane, Associate Professor, Chemistry, UBC
University Examiner

Reinhard Jetter, Professor, Joint appointment with Botany & Chemistry, UBC
University Examiner

Additional Supervisory Committee Members:

Suzana Straus, Professor, Chemistry, UBC
Supervisory Committee Member

Harry Brumer, Professor, Joint appointment with Chemistry & Michael Smith Laboratories, UBC
Supervisory Committee Member

Abstract

The discovery of novel natural products continues to be critical for the development of new pharmaceuticals. Innovative methods to discover novel natural products can reveal previously overlooked chemical diversity. One such method is genome mining, where sequenced bacterial genomes are assessed for the presence of biosynthetic gene clusters. Another such method is bioassay-guided fractionation. Following either of these approaches, the bacteria must be grown and harvested, and novel natural products must be isolated and characterized.

In the first part of this thesis, a nitrogen-NMR guided approach was developed to retrieve genetically predicted natural products from bacterial cultures. Piperazic acid (Piz)-containing natural products were targeted because this unique amino acid is often found in peptidic natural products with biological activity and impressive chemical structures. Piz contains a unique nitrogen-proton NMR correlation targeted through ^{1}H-^{15}N HSQC. The unique N-H correlation also gave access to Piz's diagnostic spin system through ^{1}H-^{15}N HSQC-TOCSY NMR experiments. These two ^{15}N NMR experiments were used to monitor for the presence of peptides containing Piz in culture extracts of genome-mined bacteria. Through the application of these ^{15}N NMR experiments to guide isolation of Piz natural products, four novel compounds were discovered from *Streptomyces incarnatus* NRRL 8089. Three of these were isolated and structure's elucidated as part of this thesis work: dentigerumycin F (**4.2**), dentigerumycin G (**4.1**) and incarnatapeptin A (**4.3**). **4.3** demonstrated a unique bicyclic moiety not previously seen in chemical structures, and a fourth compound, incarnatapeptin B (**4.4**), has *in vitro* cytotoxicity.

In the second part of the thesis, bioassay-guided fractionation is used to screen a small library of marine bacteria in various assays. Using this method, known natural product molecules were uncovered, along with the discovery of two novel natural products from the marine

bacterium *Salinispora arenicola* RJA3005. These two compounds, 6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-hydroxy-3-methyl-2H-pyran-2-one (**6.1**) and N-(3-hydroxy-5-(1-hydroxy-2-(4-hydroxy-3-methyl-2-oxo-2H-pyran-6-yl)propyl)phenyl)acetamide (**6.4**), were isolated from extracts of wild-type bacteria for the first time. Feeding studies and analysis of ^{13}C splitting patterns suggest that these compounds were biosynthesized from bacterium through phosphoenolpyruvate and erythrose precursors.

Altogether, this thesis's work develops and harnesses various natural product discovery methods to uncover diverse natural products.

Lay Summary

Some living organisms produce specialized organic compounds termed natural products. Natural products are an important resource for the development of new pharmaceuticals. Common medicines derived from natural products include penicillin for bacterial infections, statins for lowering blood cholesterol levels, and cyclosporin A to prevent rejection of organ transplants. As we face antibiotic resistance and increased rates of cancers, the search for new medicines from natural sources continues. In many cases, we now have a detailed understanding of how these organisms construct natural products. This knowledge, in turn, enables the discovery of undetected natural products.

In this thesis, I isolated natural products from crude extracts of marine and terrestrial bacteria to search for new potential lead compounds for drug development. Towards that end, I also developed a new method for discovering undetected natural products that combines bacterial genomic information and spectroscopy experiments.

Preface

Chapter 2: A large portion of this material has been modified from Morgan, K. D.; Andersen, R. J.; Ryan, K. S. Piperazic Acid-Containing Natural Products: Structures and Biosynthesis. *Nat. Prod. Rep.* **2019**, *36* (12), 1628–1653. I modified the figures from the published paper to be more compact. Dr. Katherine Ryan and I co-wrote the manuscript for the paper. Initial BLASTP searches were completed in partnership with Dr. Pamoda Ratoneweera. All other bioinformatic analysis was completed on my own.

Chapter 3: A portion of this chapter has been modified from Morgan, K. D.; Williams, D. E.; Patrick, B. O.; Remigy, M.; Banuelos, C. A.; Sadar, M. D.; Ryan, K. S.; Andersen, R. J. Incarnatapeptins A and B, Nonribosomal Peptides Discovered Using Genome Mining and ^{1}H/^{15}N HSQC-TOCSY. *Org. Lett.* **2020**, *22* (11), 4053–4057. I completed all the work involved in developing the method including growing the bacteria and running the NMR experiments. My contributions to the writing of the manuscript include completion of the figures and the supporting information reflected in Chapter 3 method development, as well as small aspects of the manuscript. I am responsible for the text presented in this thesis.

Chapter 4: A portion of this chapter has been modified from Morgan, K. D.; Williams, D. E.; Patrick, B. O.; Remigy, M.; Banuelos, C. A.; Sadar, M. D.; Ryan, K. S.; Andersen, R. J. Incarnatapeptins A and B, Nonribosomal Peptides Discovered Using Genome Mining and ^{1}H/^{15}N HSQC-TOCSY. *Org. Lett.* **2020**, *22* (11), 4053–4057. Michael Leblanc and I worked on the scale-up solid culturing of *Streptomyces incarnatus* NRRL 8089. I isolated and elucidated the constitution of dentigerumycin F (**4.2**) and dentigerumycin G (**4.1**). I isolated and elucidated the constitution of incarnatapeptin A (**4.3**) with incomplete evidence for the bicyclic moiety. Dr.

David Williams methylated incarnatapeptin A and grew the crystal of the methyl ester of incarnatapeptin A. Dr Brian O. Patrick ran and analyzed the x-ray crystallography data. Dr. David Williams re-isolated **4.1, 4.2**, and **4.3** to obtain better data for publication and isolated and elucidated the structure of incarnatapeptin B (**4.4**). The anticancer biological assay was conducted by Carmen A. Banuelos from the Sadar laboratory at the BC Cancer Agency. The disc diffusion bacterial assay was completed by Michael Leblanc, and the MIC assay was completed by Jessie Chen of UBC Bioservices. I co-wrote, with Dr. David Williams, the experimental section and the supporting information for the portions of the publication concerning the isolation and structure elucidation of **4.3** and **4.4**. I co-authored the figures pertaining to the structures of **4.3** and **4.4**. My supervisors contributed to the body of the manuscript. Although some of the material presented is based on a published article, I am the author of chapter 4.

Chapter 5: My contributions for chapter 5 include the work of growing the bacteria, isolating metabolites, dereplicating the known compounds and running the TLC-bioautography assay.

Chapter 6: My contributions include the isolation and structure elucidation of (**6.1**) and (**6.2**). I carried out some of the scale-up bacterial culturing, with assistance for some of the scale-up bacterial culturing from Michael Leblanc. I completed the bacterial culturing for feeding labelled precursor, followed by isolation and interpreted the resulting carbon-13 spectra.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Symbols

°C   degree Celsius

$[\alpha]_D^{25}$ specific rotation at sodium D-line recorded at 25 °C

^{13}C carbon-13

^{15}N nitrogen-15

^{1}H hydrogen/proton

C18   linear octadecane groups

δ chemical shift in parts per million from tetramethylsilane

J coupling constant (in hertz)

m/z mass-to-charge ratio

List of Abbreviations

1D	one dimensional
2D	two dimensional
Ac	acetyl
BC	British Columbia
BGC	biosynthetic gene cluster
br	broad
CCR	crotonyl-CoA carboxylase/reductase
CD_3OD	deuterated methanol
COSY	correlation spectroscopy
d	doublet
# d.	days
DCM	dichloromethane
dd	doublet of doublets
dm	doublet of multiplets
DMSO	dimethyl sulfoxide
DMSO-d_6	deuterated dimethyl sulfoxide
D_2O	deuterated water
dt	doublet of triplets
EtOAc	ethyl acetate
g	gram
HMBC	heteronuclear multiple bond coherence

HPLC	high performance liquid chromatography
HRESIMS	high resolution electrospray ionization mass spectrometry
HSQC	heteronuclear single bond coherence
Hz	hertz
IC50	half maximal inhibitory concentration
LC-MS	liquid chromatography-mass spectrometry hybrid
LC-MS/MS	liquid chromatography-tandem-mass spectrometry hybrid
LRESIMS	low resolution electrospray ionization mass spectrometry
μg	microgram
μM	micromolar
m	multiplet
Me	methyl
MeCN/ACN	acetonitrile
MeOH	methanol
mg	milligram
MHz	megahertz
min	minute
mL	millilitre
mRNA	messenger ribonucleic acid
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry
NMR	nuclear magnetic resonance

NOESY	nuclear Overhauser effect spectroscopy
NRP	nonribosomal peptide
NRPS	nonribosomal peptide syntethase
ORF	open reading frame
ORTEP	oak ridge thermal ellipsoid plot
Piz	piperazic acid
PK	polyketide
PKS	polyketide synthase
ppm	parts per million
ppt	precipitate
pyridine-d_5	deuterated pyridine
tROESY	transverse rotating frame nuclear Overhauser effect spectroscopy
s	singlet
smBGC	secondary metabolite biosynthetic gene cluster
sp.	species
t	triplet
TLC	thin layer chromatography
TOF	time of flight

Acknowledgements

Dedication

I dedicate this thesis to my sons: Leif and Forest

May the love of science and inquiry be a foundation of your lives

Chapter 1: General Introduction

1.1   Natural Products as Drug Leads

1.1.1   Natural Products in History

Natural products have been used for millennia in human cultures as medicines,[1–3] fabric dyes,[4] building materials (resins and waxes),[5] and fragrances.[5,6] Although plant crude extracts have long been used as medicines,[2] the active components of those medicines (if present) began to be deciphered less than two hundred years ago. Beginning in 1828, chemists began to isolate pure natural products such as salicin (**1.1**), quinone, and morphine (**1.4**).[2] Next, they used simple chemistry to transform these isolated natural products into analogues with greater

a
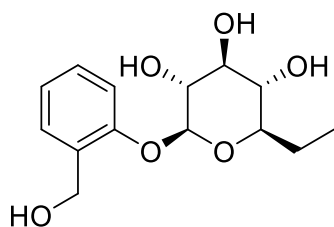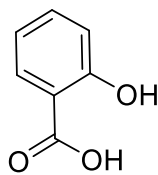
salicin (**1.1**)   salicylic acid (**1.2**)   acetylsalicylic acid (**1.3**)
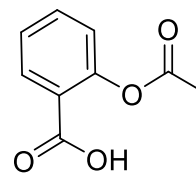
b
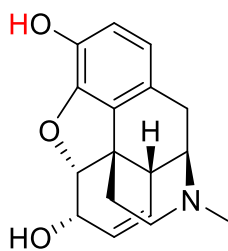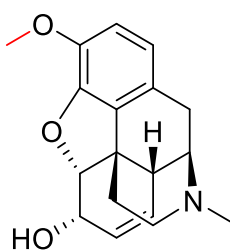
morphine (**1.4**)   codeine (**1.5**)

Figure 1-1. Historical natural products.

1

bioavailability and activity. Salicin, isolated from Salix sp., was converted to acetylsalicylic acid (**1.3**)[7] following the understanding that it was transformed into salicylic acid in the blood[8] (**Figure 1.1a**). Morphine, initially isolated as a pain-killing compound along with many other inactive poppy alkaloids, was transformed into codeine (**1.5**) (**Figure 1.1b**).[2] From these beginnings, natural products have been harnessed to develop world-transforming pharmaceuticals that allow for: successful organ transplants to occur (cyclosporin A[9–11]), bacterial and fungal infections to be defeated (actinomycin and neomycin[12,13]), treatment of tuberculosis (streptomycin[12]), and varying cancers to be battled (taxol (**1.9**), cytarabine (**1.10**), trabectedin (**1.11**), eribulin).[14] Since the mid-1980s, natural products and their derivatives have accounted for 58% percent of antibiotics and 32% of anticancer compounds. Currently, 87% of all known human diseases have a treatment that is a natural product or derived from natural products.[15] These numbers speak to the continued research potential of natural products as a source of pharmaceutical medicines.

1.1.2 Microbial Natural Products

Microbes became an important source for bioactive natural products with Alexander Fleming's discovery of penicillin in 1929 from *Penicillium notatum*. By the 1940s, research efforts transformed penicillin into a potent antibiotic drug that changed the face of World War II.[16] From this outset, 20,000 secondary metabolites have been discovered from bacteria, and fungi[11] with multiple important antibiotics (**1.8**) discovered from microbes. In addition to antibiotics, microbes have provided antidiabetic medicines, immunosuppressants (**1.6**), cholesterol-lowering (lovastatin) and antitumor medicines (**1.7**).[16]
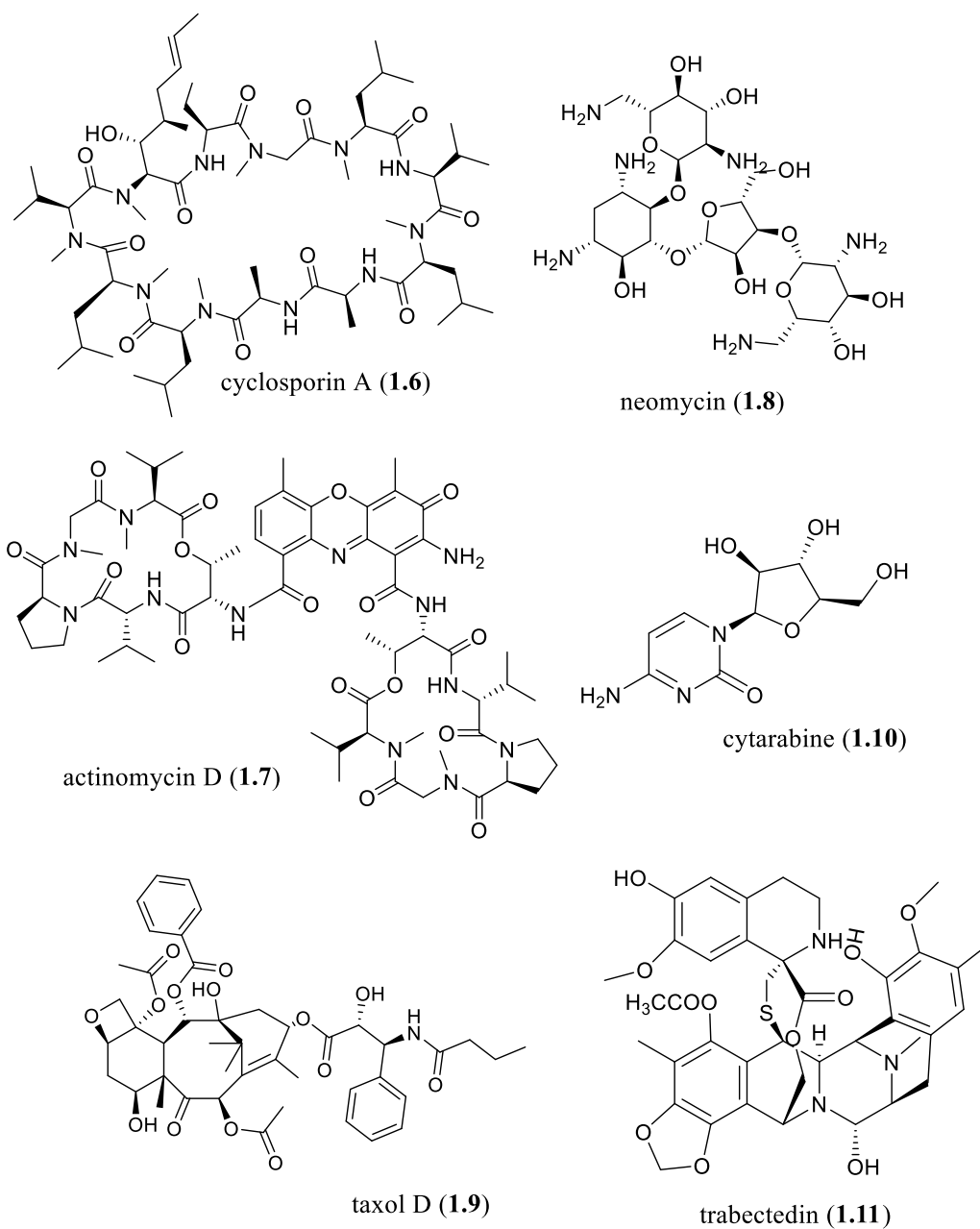
Figure 1-2. Select natural products that are in current use as pharmaceutical compounds.

Microbes is an umbrella term that encompasses both fungal and bacterial kingdoms. Given the scope of this thesis, bacterial natural products will be the focus. Bacteria are prokaryotic, lacking internal membrane-bound structures.[17] Bacteria reproduce, without recombination, via means of single-cell division into daughter cells. Instead, bacteria rearrange

their genes and genomes[17] through multiple reproduction-independent[18] DNA recombination mechanisms. A significant result of these dynamic DNA processes is that microbial communities can be constructed of multiple strains of the same bacteria species.[19] These strains will be unique in their genomic and phenotypic variability in a manner loosely similar to how one might think of individual humans: of the same species but with significant variation.[19] Although bacteria are single-celled organisms, they can function ecologically as multicell units when they form colonies from daughter cells that do not separate after division.[18]

Within the Bacteria Kingdom itself, the Phylum Actinobacteria has dominated bacterial natural product discovery efforts. Actinobacteria is one of the most extensive taxonomic lineages composed mainly of Gram-positive staining bacteria with high G + C DNA content;[20] however, diversity marks this lineage more than any common feature. Several orders, such as Micromonosporineae, Streptomycineae, and Actinomycetaceae,[20] have been well-studied since they grow well in laboratory conditions and have large genomes showing prodigious capacities for natural product production.[21] Many species of this order produce spores that can resist dehydration, mildly hot temperatures and produce the same chemistry when revived. Actinobacteria contain both terrestrial and marine strains,[20] and are numerous in most soils, including marine sediments. Actinomycetes are the source of many bioactive natural products,[22–24] including daptomycin (**1.12**, isolated from *Streptomyces roseosporus* NRRL 11379),[25] a powerful antibiotic.
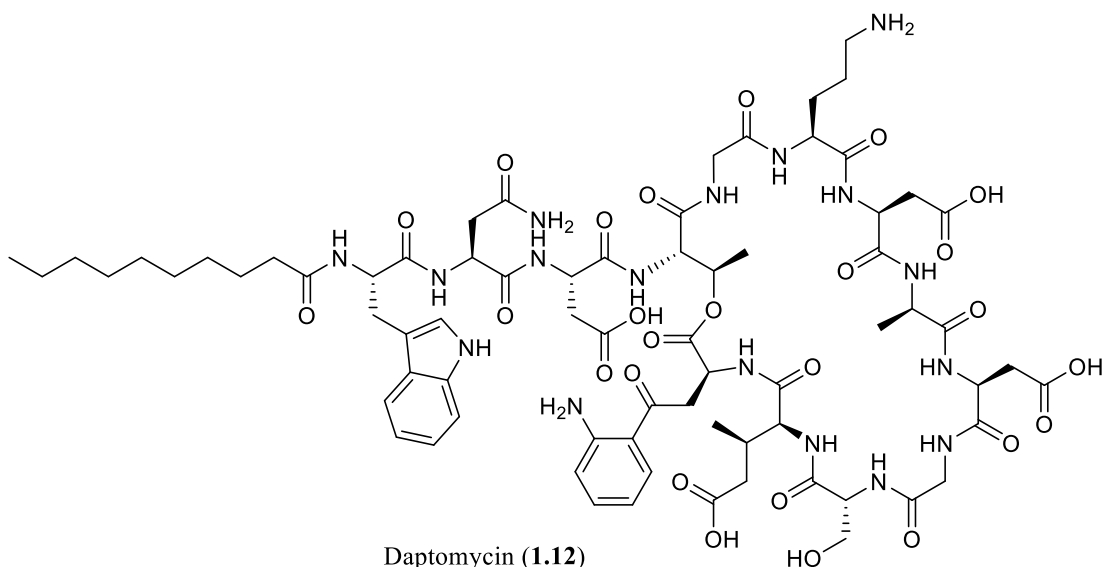
Daptomycin (**1.12**)

Figure 1-3. Daptomycin, an antibiotic isolated from a terrestrial Actinobacteria.

1.1.2.1 Marine Bacteria

The heterogeneous marine environment is a treasure house of diverse bacterial strains

that can be collected from sediments and the microbiomes of invertebrates.[26] Altogether, the

marine environment, which covers 70% of this planet's surface, supports an enormous range of

biodiversity.[27] In fact, early samples taken on marine sediment found higher microorganism

diversity than that found in terrestrial soils.[28] Some bacterial genera, such as the *Salinispora*

genus discussed in **Chapter 6,** require a certain sodium concentration in their environments for

the proper functioning of basic metabolic activity and have truly become marine-dependent

bacteria.[28] Many marine strains, however, are recent adaptations from a terrestrial origin.[29]

Marine microorganisms experience a complex heterogeneous environment,[30] which has been

shown to select for microorganisms with a wide range of specialized metabolites.[29]

This diversity has allowed chemists to isolate and identify numerous bioactive and novel

marine natural products.[31] Historical examples include the pyrrole antibiotic 2,3,4-tribromo-5(1-

hydroxy,2,4-dibromophenyl)pyrrole[32] isolated from seawater bacteria and the marine sponge

5

sourced cytarabine (a chemotherapeutic drug)[33] and vidarabine (an anti-viral).[34] A recent example is salinosporamide A (**1.14**), a proteasome inhibitor in advanced stages of clinical trials for its anticancer properties.[35] Another recent example is the peptide cyclomarin A (**1.13**),[36,37] a potent anti-inflammatory, anti-malarial and anti-viral compound[38] isolated from a *Streptomyces* sp. gathered from marine sediments near San Diego, California. Cyclomarin's structure and bioactivity have also driven interest in characterizing its biosynthetic pathway and several synthetic efforts.[36] Marine organisms continue to yield diverse natural products, with hundreds of pharmaceutically useful novel natural products reported yearly.[39]
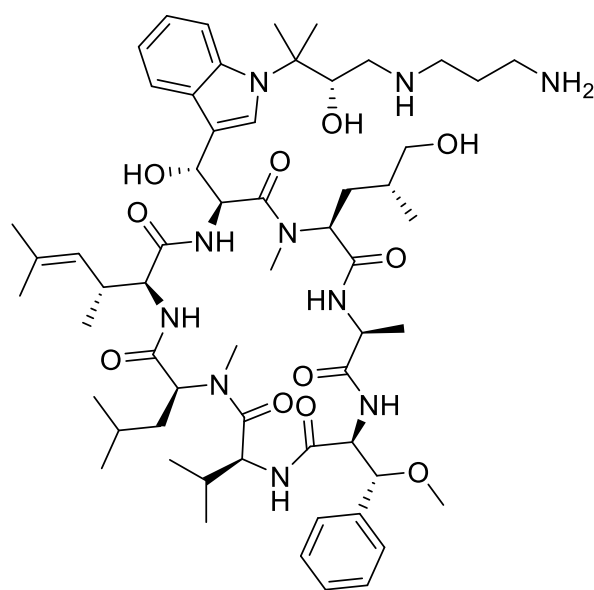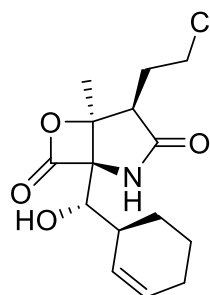
Cyclomarin A (**1.13**)   Salinisporamide A (**1.14**)

Figure 1-4. Potent bioactive compounds isolated from marine bacteria.

1.1.3   Structure Elucidation of Novel Natural Products

An essential aspect of natural product chemistry is characterizing the constitution and configuration of the isolated biological molecules that have utility either as drug leads, insecticides, dye compounds, or other applications. Without knowledge of the natural product structure, there is no capacity for rational synthesis, industrial fermentation, mechanism-of-

action studies, or biosynthetic pathway characterization. In the early years of natural product chemistry, structure elucidation was a painstaking task involving chemical degradation experiments,[40] taking decades of work,[40,41] and sometimes producing erroneous results. Even when supplemented with spectroscopic techniques, the chemical degradation experiments necessitated several hundred milligrams (at minimum) of a natural product and were thus biased towards major metabolites.[40]

Since the 1970s, mass spectrometry[42] and nuclear magnetic resonance[40,43,44] have been amongst the basic techniques for elucidating the constitution of natural products. Mass spectrometry (MS) relies on an ionization source to ionize molecules (as well as fragment them), a mass analyzer that separates those ions by their mass-to-charge ratio (m/z), and a detector that translates that m/z signals into a mass spectrum.[42] A variety of MS instruments are available for analysis, dependent on the target molecules' size and polarity. A reliable method for natural product structure elucidation is Electrospray Ionization (ESI), where a soft ionization source tends not to fragment molecules, generally giving only the parent molecule ion.[42] With high-resolution ESI-MS, the parent ion's mass is known to four to five decimal places, which allows calculation of possible molecular formulae with a high degree of certainty. That said, molecular fragmentation in MS follows chemical principles, and, as such, it provides a molecular fingerprint that can, for certain classes of natural products, be useful in structure elucidation and discovery.[42] In brief, MS is indispensable in providing the molecular formula for novel natural products, particularly if there are multiple heteroatoms or a high degree of unsaturation.

Nuclear magnetic resonance (NMR) is also an indispensable spectroscopic technique that provides predictable information about the chemical environment for each of the atoms in a molecule. In an applied magnetic field, certain isotopes such as ^{13}C, ^{1}H, ^{14}N and ^{15}N have a

magnetic moment that causes them to precess about the direction of the magnetic field at a frequency (Larmor) proportional to the strength of the magnetic field and the gyromagnetic ratio unique to the isotope. This property can be manipulated to give us information about the identity and number of neighbours surrounding an NMR active nucleus. Since the 1950s, NMR has been used for organic structure determination and biosynthetic pathway characterization.[44] From ^{1}H experiments, chemical shift (δ) data can be collected depending on the slight changes in Larmor frequency influenced by the surrounding electron density, and coupling data can be gathered, which is dependent on the number of proton neighbours each proton has. With instrument evolution, ^{13}C experiments allowed the determination of the carbon chemical shift. Between 1950 and 1990, 2-dimensional (2D) experiments were developed that allowed observation of multi-atom correlations between homonuclear atoms (COSY:1H-1H correlation), slowly into 1-bond correlations between heteronuclear atoms (HETeronuclearCORrelation, HETCOR 1H-13C correlations) and then 2,3-bond correlations between heteronuclear atoms (Heteronuclear Multiple Bond Correlation Spectroscopy, HMBC).[44,45] Presently, there is a standard suite of experiments[43,44] used as evidence of atom connectivity as outlined in **Figure 1-5.** Through-space experiments take advantage of nuclear Overhauser enhancements (NOEs) to help identify the relative configuration. As a chemical tool, NMR's strengths are its almost universal detection, the detailed information it delivers on molecular structure (especially for novel compounds), and its quantitative abilities.[46] Additionally, instrumental advances have increased signal/noise proportions at the same concentration while also giving improved resolution, which permits analysis of even microgram amounts of material.[44] This capacity to analyze minute material amounts has increased the discovery of minor metabolites, which are natural products produced at low levels.
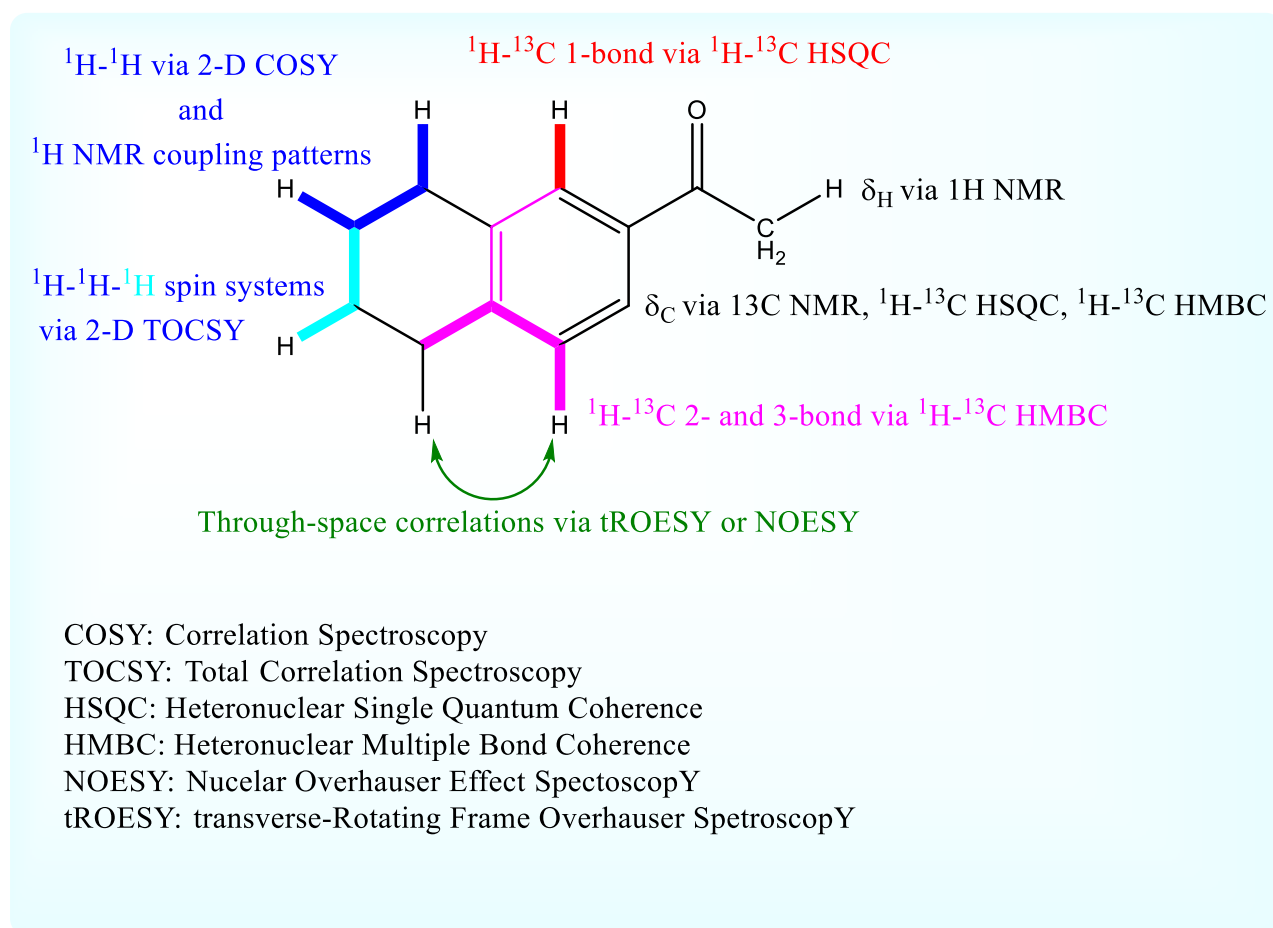
Figure 1-5. The standard suite of NMR experiments utilized to construct the constitution of a novel compound.

1.1.4   Natural Product Discovery

The mining of natural products for human use has successfully yielded many life-saving pharmaceuticals, yet continual discovery efforts are needed to improve on that success.[12–14,47–49] The ongoing need for new medicines drives the search for novel products with drug lead promise.[47,49] In addition to these bioactivities, natural products have utility as dyes, insecticides, herbicides and more. The chemical complexity and bioactivity often found with natural products inspire and inform synthetic organic chemists' work.[50,51] The biochemical transformations leading to natural products have motivated chemical biologists to study the associated enzymatic

pathways.[52–54] Finally, due to the biological interactions mediated by natural products, chemists, biologists and microbiologists study these molecules' chemical ecology.[55,56] As such, the drive to uncover and discover novel natural products is propelled by the need for new medicines and the need to grow a basic scientific understanding of the chemical world around us.
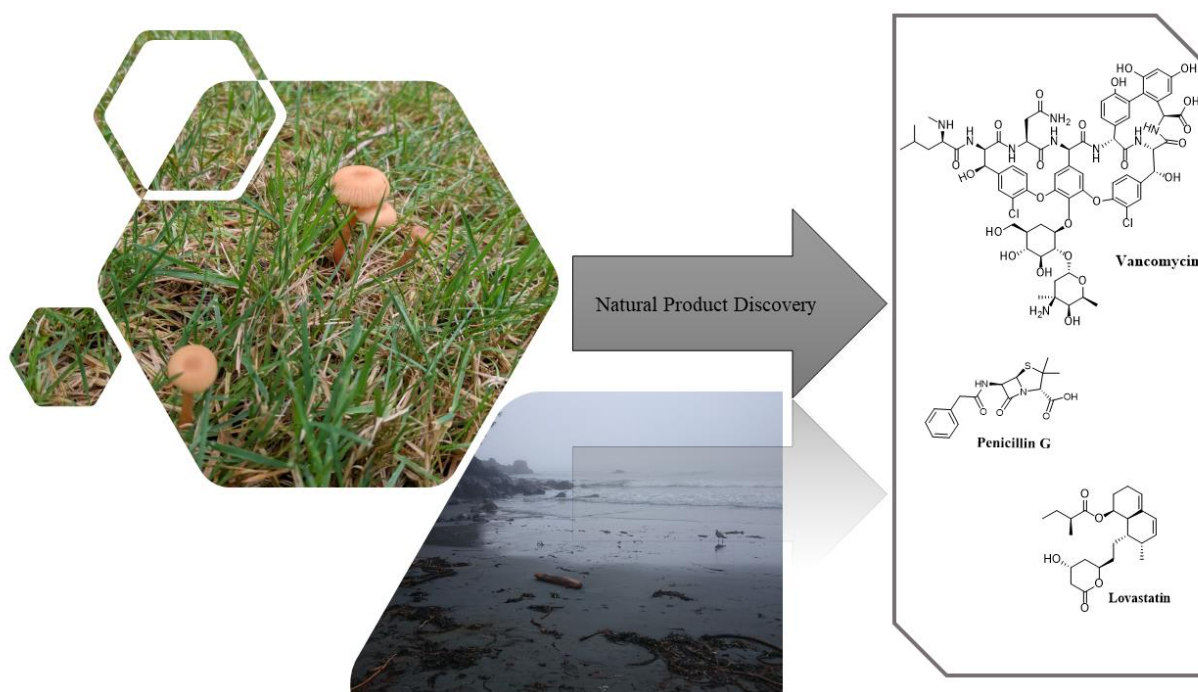
Figure 1-6. Natural product discovery mines the chemical diversity of living organisms for useful bioactive molecules.

Although there are various discovery strategies, the general workflow is roughly the same (**Figure 1-7**). Most discovery efforts begin by determining an organism of interest, then extracting metabolites from that organism in some manner, purifying a compound, and elucidating that compound's structure. Despite a recent plateau in the rate of natural product discovery, analysis of discovery patterns still demonstrates the enormous potential for novel chemistry from diverse living creatures.[47] The development of new discovery methods is a useful way of driving discovery, and there has also been work to match genotype with chemotype for faster discovery.[2,15,46,57,58] Given the centrality of the specific method to optimized natural

product discovery, an in-depth examination of some of these methods follows in **Sections 1.1.4.2-1.1.4.4**, with genome mining[59] explored in detail in **Chapter 2-1** and **2-2**.

Which Strain ?
- Bioassays
- High-throughput screen of crude fractions using bioassays
- Unusual strains
- Mass networks
- Genome mining

Isolation
- Separation methodology
  - Liquid-liquid extraction
  - Solid-liquid extraction
  - Chromatography
    - Size-Exclusion
    - Flash

Purification
- Crystallization
- High Performance Liquid Chromatography

Structure Elucidation
- Nuclear Magnetic Resonance
- Mass Spectrometry
- LC/MS/MS
- X-ray Crystallography

Figure 1-7. General discovery workflow.

1.1.4.1   The Biological Basis of Natural Product Discovery

The process of discovering novel natural products is rooted in the biological diversity of organisms. Given that natural product production has been tied to an organism's environmental interactions,[60,61] and that evolution and ecology are interdependant,[62] one could hypothesize that

natural product variability may be constrained by slower evolutionary processes.[60] However, genomic analysis of bacteria, fungi and plants demonstrates a greater capacity for secondary metabolite production than currently represented in discovered chemistry.[63]

Although primary metabolic pathways are widely conserved across kingdoms, genera and species, specialized metabolites' adaptive significance results in distinct metabolic profiles for strains of even the same microbial species.[64] As a result, spatially and temporally separated environmental samples of marine sediment, invertebrates, or terrestrial soil (as examples) have the potential to yield distinct chemistry.[65–67] Conversely, the effects of horizontal gene transfer (HGT) are such that, for example, rifamycin, a potent antibiotic, can be found in varying bacterial genera from *Streptomyces* to *Salinispora*.[68] As a result, the challenge of drug discovery from bacteria is the tendency to find known compounds. One solution to this challenge is screening bacterial strain extracts, often seeking minor metabolites with the assistance of their bioactivity, in multiple assays and dereplicating using NMR data, MS data, and refined literature searches.

1.1.4.2   Bioassay-Guided Fractionation

Bioassay-guided fractionation is a robust method used in varying capacities since the 1940's[58] for systematic screening of organisms to discover natural products.[69–71] Bioassay-guided fractionation continues to yield novel chemistry in natural product discovery.[69,70,72–74] Over the last 30 years or so, methods to improve bioassay-guided fractionation include high-throughput screening of hundreds of samples,[58,59,75] as well as the refinement of the bioassays towards target-based approaches (such as enzyme assays),[58] or faster turn-around time (TLC bioautography method).[76]

Bioassay-guided fractionation focuses on biological activities, usually with a medicinal utility, utilizing a bioassay designed to test for the desired function. This robust method can identify organisms producing bioactive natural products and guide the isolation scheme of a pure natural product. Antimicrobial bioassays test for natural products that will inhibit the growth of specific human pathogens. Cytotoxic bioassays test for natural products with inhibition of specific cancer cell lines. Enzyme assays target natural products that will inhibit enzymes from specific, medically relevant biochemical pathways in the human body. Bioassays can also be used to test for novel insecticides and herbicides.[11,77] Utilizing a bioassay that tests for the desired functionality, metabolic extracts from organisms such as bacteria, fungi, marine invertebrate, and plant materials are examined for the desired biological activity. After a positive bioassay result, the "crude" extract can be sequentially and iteratively separated chemically via such tools as solvent-solvent partitioning and chromatographic separation. After each round of separation, the resulting fractions are tested in the selected bioassay. The fraction displaying the most significant bioactivity is chosen for further separation rounds. Ideally, since the active compound concentration should be increasing in ratio to the non-active components of the fractions, an increase in bioactivity metrics should be observable in the bioassay results. Bioassay-guided fractionation points through the complex mix of molecules produced by living things towards the compounds of interest. After this retrieval, data collected from nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) allow structure elucidation. The robust technique of bioactivity guided fractionation has yielded many vital pharmaceuticals such as vancomycin (**1.15**),[78] salinosporamide (**1.14**),[35] and lead compounds such as kutzneride (**1.16**),[72] to name but a few. A drawback is that bioassay-guided fractionation

can reveal the same chemistry even from different bacterial species due to the presence of

identical specialized metabolites, often because of horizontal gene transfer.
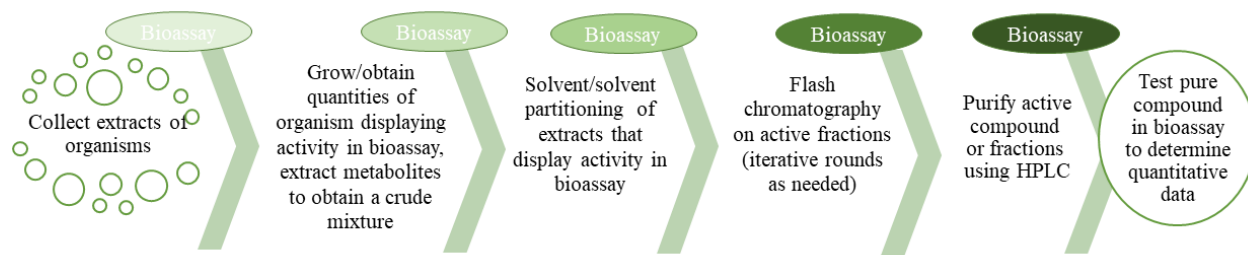
Figure 1-8. The bioassay-guided fractionation process, demonstrating sequential separation guided by the results in bioassays of choice. Represented is the expected increase in activity as the active compound is concentrated and isolated from a complex mixture.

Vancomycin (**1.15**)

Kutzneride (**1.16**)
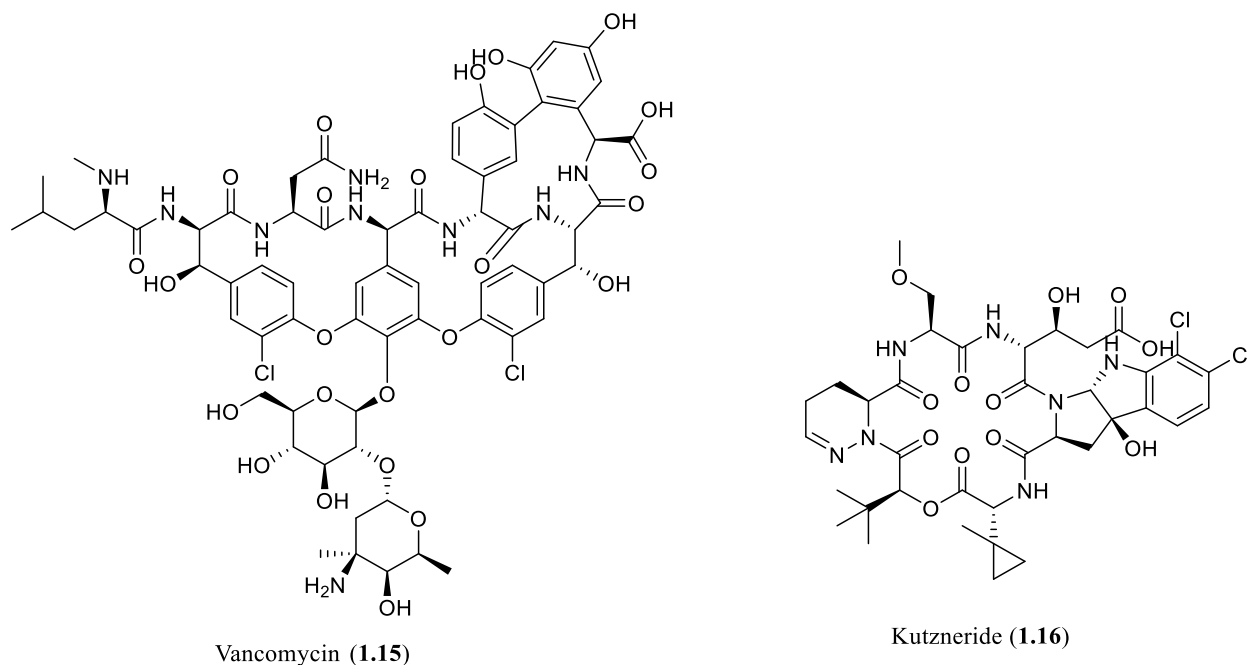
Figure 1-9. Natural products isolated using bioassay-guided fractionation.

1.1.4.3   Mass Spectrometry as a Natural Product Discovery Methodology

Mass spectrometry (MS) can be used directly as a discovery tool. The highlight of MS for discovery is that it can be used both for high-throughput identification of known molecules[79] and investigations into new compounds. When coupled with metabolomic data analysis methods such as multivariate statistical analyses, self-organizing map analytics (SOM), or molecular networking, strains or extracts with potentially novel components can be prioritized for investigation.[80,81] This type of approach has seen the discovery of tambromycin,[82] columbamides,[83] and other molecules. However, there are still some significant limitations when using mass spectrometry as a discovery tool, such as varying tendencies of molecules to ionize, media interference in crude samples,[84] unpredictable fragmentation patterns of some natural product classes,[85–87] and the need to use NMR in order to determine the final structure.

1.1.4.4   NMR as a Natural Product Discovery Methodology

NMR methods for natural product discovery have been more recently explored. NMR data collected on crude biological extracts shows every correlation in the 2D data essential for structure elucidation; this makes it difficult to determine the structure of the components without extensive isolation efforts. One solution to that problem was the development of the differential analysis of 2D NMR spectra (DANS). DANS was initially developed to focus on identifying phenotypically unique molecules from target organisms with ecological or other differences via NMR without complete isolation. In this way, DANS allowed elucidation of the indole alkaloids TC-705A and TC-705B.[88] Indeed, in the chemical ecology field, DANS has proven to be a powerful method for studying signaling molecules of the insect *Caenorhabditis elegans,*[89] with a variant of this strategy used in many metabolomic studies of plant and insect species.[90,91] However, as pure compounds are not isolated with this method, it is difficult to test the natural

product's bioactivity. Attempts have been made to target bioactive compounds utilizing variants

of the DANS strategy.[92,93] These method variants have had limited success, although one attempt

identified elaganolone (**1.18**), previously isolated in traditional bio-activity screening protocols.[94]

In the search for methods to combine with genome mining, DANS has been used to match

orphan gene clusters with the produced molecule by statistical comparison of the NMR spectra

from a targeted deletion mutant compared to a wild-type strain. This method identified bacillaene

(**1.17**), an easily degraded molecule that resisted traditional isolation protocols.[95] In the end,

identification of molecules from within a complex NMR spectrum eases dereplication efforts,

although the databases that would strengthen this are fragmented at best.[96]

Bacillaene (**1.17**)

Elaganone (**1.18**)
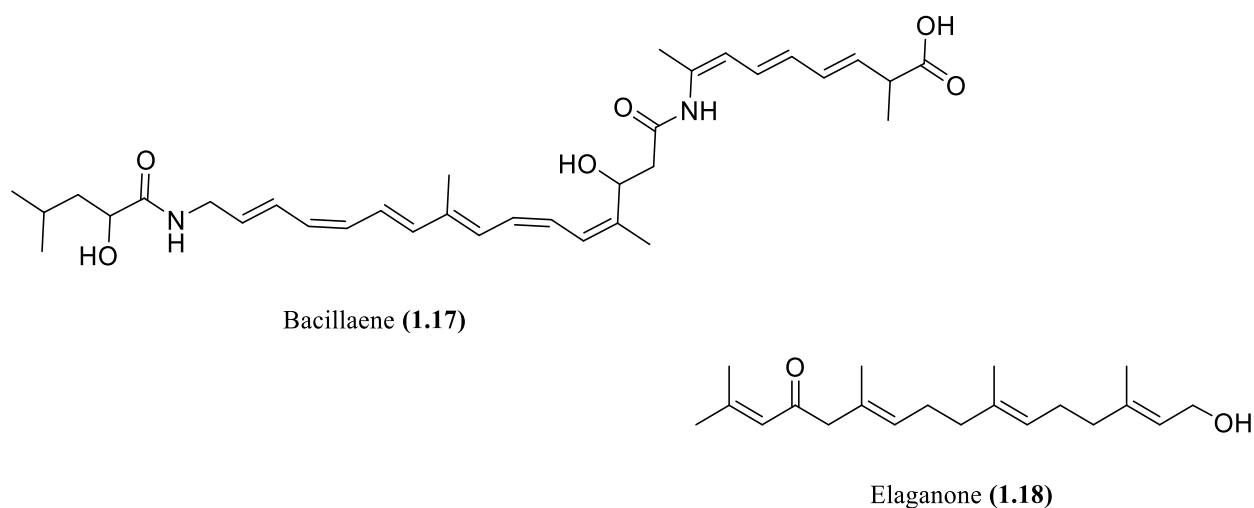
Figure 1-10. Structures solved but not isolated with DANS.

The van Wezel group reported a strategy closely related to DANS in which they utilized

extensive multivariate data analysis of ^{1}H NMR spectra to link bioactive fractions with

differences in proton spectra for selection of the desired fraction. For isolation of the simple

antibiotic 7-prenylisatin, they utilized a common aromatic proton fingerprint. Assessing

differences in proton spectra led to the discovery of iotrochotazine A and the venulosides.[97] The

common challenge with these proton-based techniques is that proton spectra have a limited

chemical shift range, generally 0-15 ppm. This limited range leads to many overlapping signals in crude samples, requiring time-consuming chromatographic steps before confirmation of uniqueness. In 2015, the Wang group employed the larger spectral range of ^{13}C to identify clusters of ^{13}C resonances in fractionation efforts. This strategy led to the discovery of six podoverines (**1.19**) with considerably more structural diversity than previous NMR-discovered molecules.[98] Other methods for NMR-based discovery, heavily reliant on metabolomic methodologies, include hybrid NMR-MS methods combined with MS principal component analysis methods to select a strain or fraction of interest. These hybrid methods have led to the discovery of a new piperamide, ottonoidenamide (**1.20**)[99] and a host of other plant compounds.[100]

Podoverine C (**1.19**)

Ottonoidenamide (**1.20**)
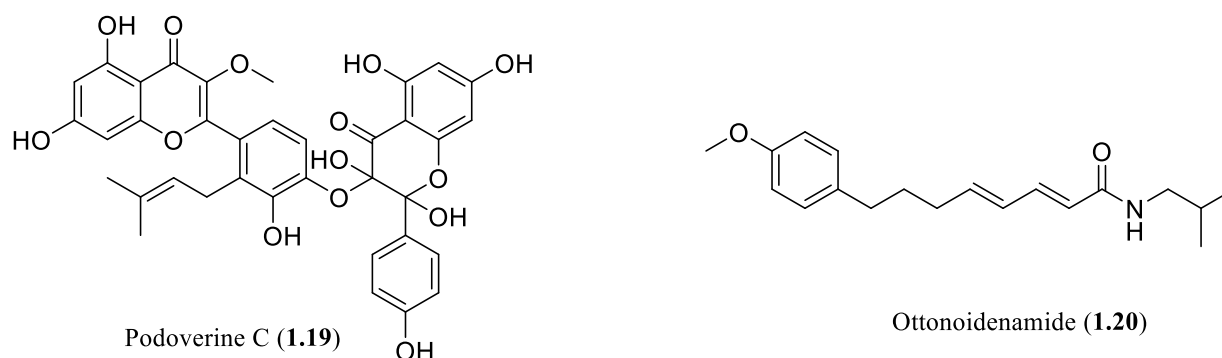
Figure 1-11. Natural Products isolated using NMR in various aspects of the discovery process.

Two additional methods have combined targeted NMR methods with genome mining. These are the genomisotopic approach and a phosphate NMR approach. The genomisotopic approach developed by Gerwick *et al.* uses isotopically labelled precursors as predicted by *in silico* genetic information to link orphan secondary metabolite gene clusters to the putatively produced natural product.[101] In this method, the use of the ^{1}H-^{15}N HMBC experiment to identify a labelled isoleucine (Ile) led to the successful isolation of orfamide A (**1.21**), representative of a new peptide class. As a drawback, this method required extensive labelling due to dependence on the increase of N-H signal of a labelled proteogenic amino acid with signals in a crowded region

17

of the N-NMR spectrum.[101] The complexity of signals in this region biases this method towards isolating major metabolites. In the phosphate NMR method, Metcalf and associates target a specific phosphonate encoding gene, *pepM*, to select microorganisms for culturing that contained that specific gene.[102] They then utilized the diagnostic chemical shift of a C-P in ^{31}P NMR to assess the presence or absence of phosphonates in their crude mixture, which led to the isolation of (2-acetamidoethyl)phosphonic acid, (2-acetamido-1-hydroxyethyl) phosphonic acid (**1.22**), and (cyano (hydroxy)methyl)phosphonic acid (**1.23**).[102] This approach links a diagnostic ^{31}P NMR signal to a specific type of phosphorous metabolite for novel natural product discovery. It demonstrates the utility of targeted genome mining coupled with specific NMR features for rapid natural product discovery efforts. However, phosphonic acid natural products are a relatively small group of natural products with limited structural diversity. This approach was also not able to complete the structure elucidation of all the isolated products and synthesis was required. Overall, NMR has infrequently been used as a tool for natural product discovery. Given the power of NMR spectroscopy, I wanted to expand the approach of using NMR to match genotype with chemotype, which is the focus of **Chapters 2 to 4**.

Orfamide A **(1.21)**

(2-acetamido-1-hydroxyethyl) phosphonic acid (**1.22**)   (cyano (hydroxy)methyl)phosphonic acid (**1.23**)
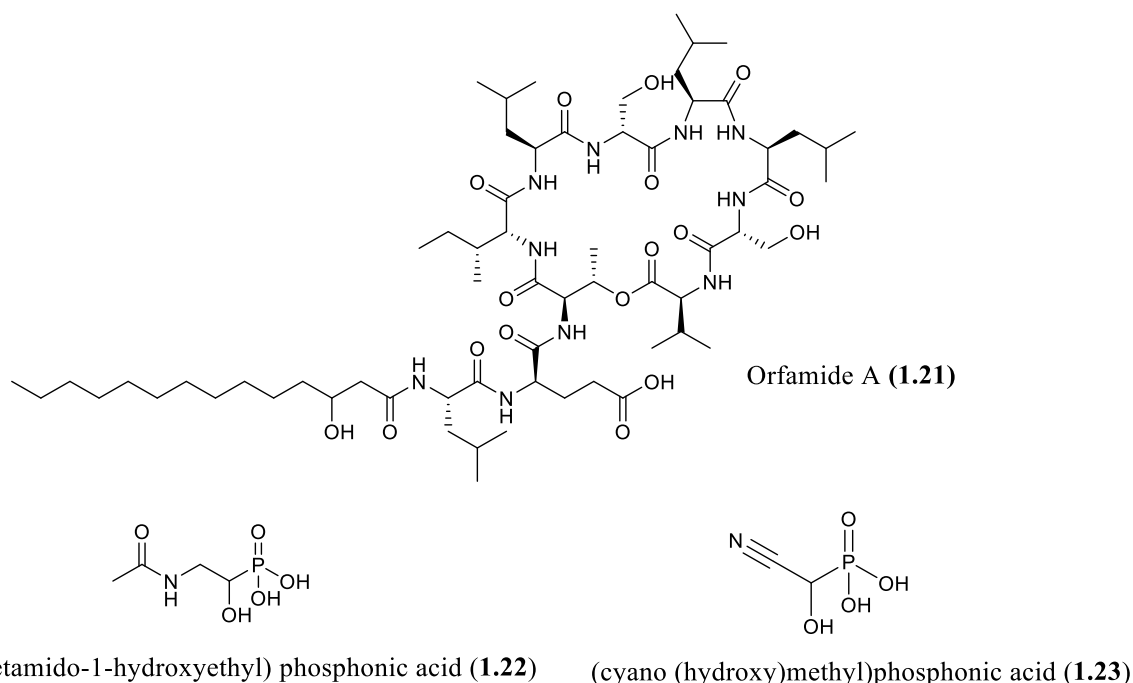
Figure 1-12. Natural Products isolated using variations of NMR-Genome mining approaches.

1.2   Natural Product Biosynthesis

Natural products are often broadly defined as both the primary and secondary metabolites produced by living organisms.[52,103–105] Medicinal and organic chemists are usually referring to secondary metabolites when they discuss natural products.[103,106] Secondary metabolites are generally small molecules (smaller than 2000 Daltons) that are often species-specific.[107] A large body of evidence supports that organisms produce at least some of their secondary metabolites for defense and inter- and intra-species communication purposes.[64,105,108–111] Given their specialized purpose, secondary metabolite production in microbes is tightly regulated.[110] In fact, the expression of secondary metabolite enzymes generally occurs when the bacterial growth rate slows upon the heterogenous exhaustion of carbon, nitrogen and phosphorous nutrient sources.[52,112] At this point, some primary metabolite building blocks,[109–111,113,114] previously stored as pools, are carefully shunted into the biochemical pathways leading to secondary

assemble natural products.[109]

1.2.1   Natural Product Classes

Although different sources will vary, there are six classes of natural products generally

considered to have a secondary metabolite origin.[103,106] These six classes are broadly defined as

polyketides, nonribosomal peptides, shikimate derived, alkaloids, purines and pyrimidines, and

isoprenoids. Molecules from each of these classes can be glycosylated for additional diversity.[103]

These six classes all have their varying precursors rerouted from primary metabolite pathways.[103]

Table 1.1 details a simplistic overview of the connections between primary and secondary

metabolites. Of these natural products classes, bacteria tend to produce polyketides,

nonribosomal peptides, and shikimic acid metabolites,[106] and as such, a more in-depth focus will

spotlight each of those classes. Following this will be a brief discussion about alkaloids. See

Figure 1-13 for representatives of these classes.

Nutrient Source	Central Precursor	Branch Precursor	Primary Metabolites	Secondary Metabolites
Glucose	Acetyl-CoA	Malonyl-CoA	Fatty acids	Polyketides
				Nonribosomal amino Acids
		Mevalonic Acid	Steroids	Terpenoids
		Krebs Cycle	Ribosomal amino acids	Nonribosomal amino acids
				Terpenoids
	Shikimic acid	Ribosomal aromatic amino acids		Flavonoids
				Lignans
				Alkaloids
				Nonribosomal amino Acids

Table 1-1. Overview of the connection between shared precursors of primary and secondary metabolites.
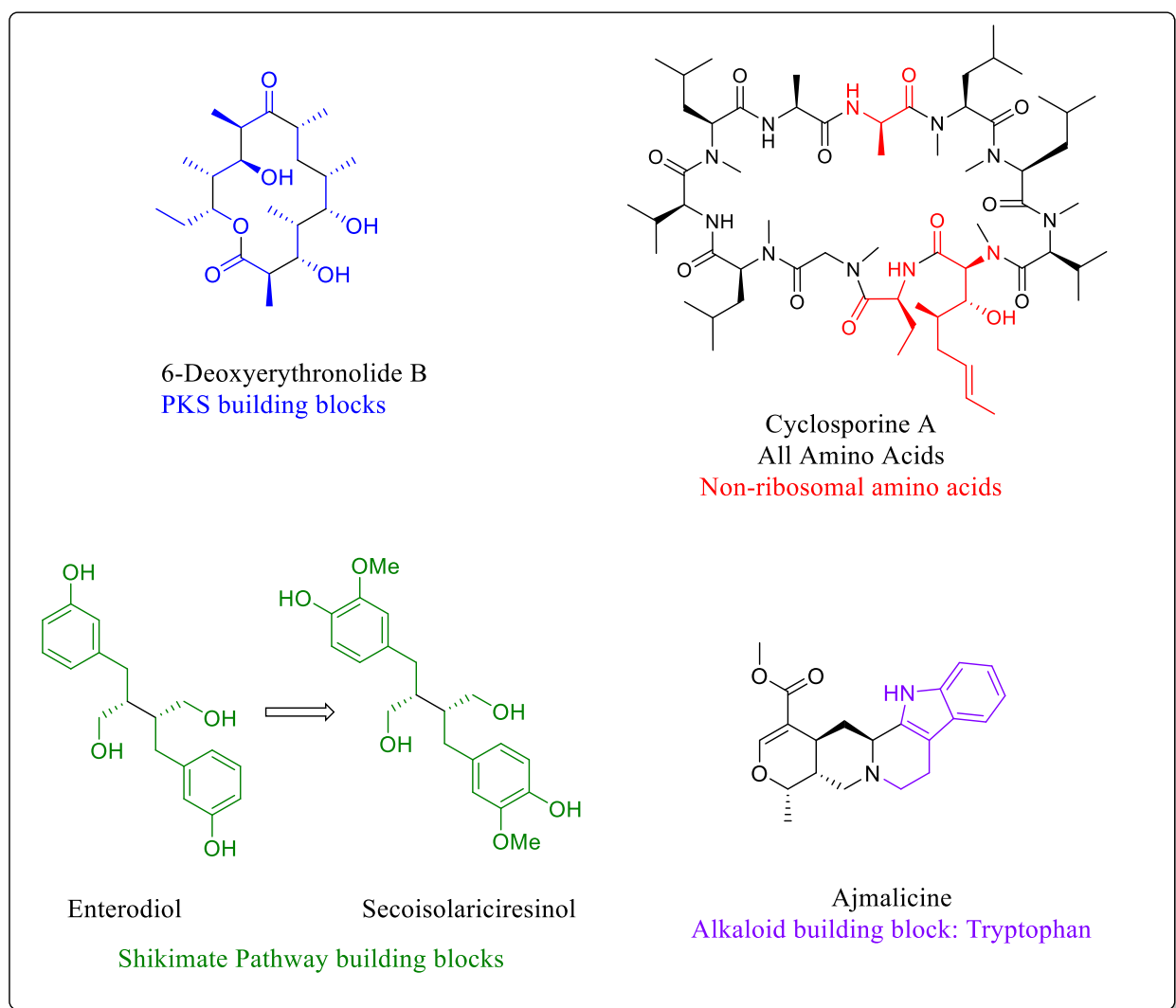
Figure 1-13. Representative natural products of the select natural product classes.

1.2.1.1 Modular Natural Products: Polyketides and Nonribosomal Peptides

Although sourced from different building blocks, both polyketides and nonribosomal peptide natural products frequently possess biological activity and are constructed utilizing assembly-line production of the larger molecules from small monomers. To do so, enzymes work together in module units that are selective for monomer variants and build them together, "lego style", utilizing repetitive chemical logic specific to the natural product class. Additionally, both

can have pre- and post-assembly line tailoring. Together, polyketides and nonribosomal peptides make up roughly thirty percent of known natural products.[115,116]

1.2.1.1.1 Polyketides

Polyketides (PK) are produced from acetyl-CoA in two carbon fragments (**Figure 1-14a**). Owing to the reactive diversity of the carbonyl functional group, those fragments can be arranged into diverse skeletons of natural products (aromatics, pyrones, alkenes, alkanes etc.). Producing organisms use large modular enzymes called polyketide synthases (PKS) constructed of repeating modules of smaller enzyme domains with specific catalytic functions that work together in an
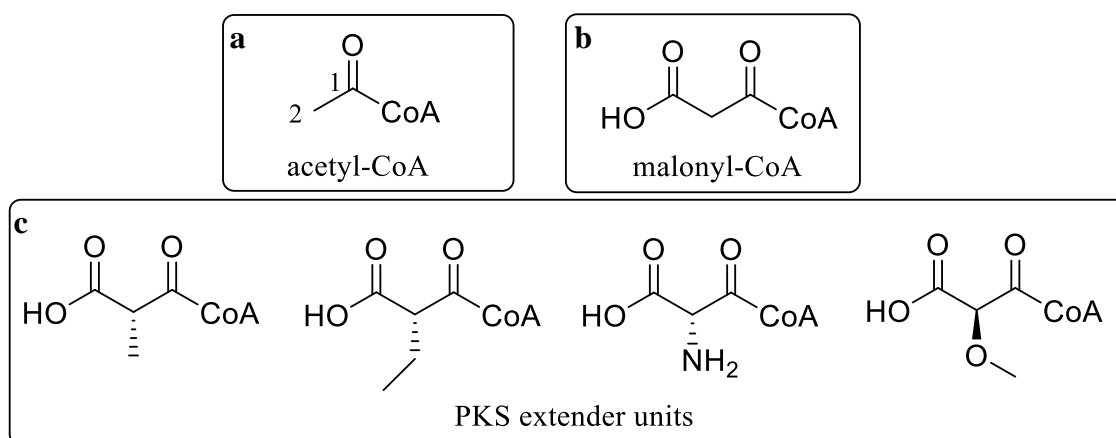
Figure 1-14. Polyketides are built from basic two-carbon building blocks derived from either a) acetyl-CoA or b) malonyl-CoA. c) A large diversity of extender units add novelty to polyketides.

assembly-line fashion to craft the diverse PK skeletons. The typical module contains a β-ketoacylsynthase (KS) domain, an acyl transferase (AT) domain, and a phosphopantethienylated acyl carrier protein (ACP, or T for thiolation).[103,117] The KS domain is the site where malonyl-CoA (**Figure 1-14b**) undergoes *decarboxylative* Claisen condensation chemistry (**Figure 1-15**) to connect the acetyl-CoA monomers.[118] The AT domain confers specificity towards a particular acetyl-CoA extender unit (**Figure 1-14c** shows the diversity of PK extender units). The ACP (T)

domain holds new acetyl units as thioesters, whereby the C-2 carbanion nucleophile attacks the electrophilic carbonyl C-1 held as a thioester on the KS domain of the growing chain (**Figure 1-15a**). These three domains produce chains of ketones or poly-β-ketos.[118] However, these polyketides are often reduced either to a hydroxyl, an olefin, or a saturated alkyl, respectively, using first ketoreductase (KS) domains, next dehydratase (DH) domains, and lastly enoyl reductase (ER) domains (**Figure 1-15b**).[117] The termination module contains a final domain called a thioesterase (TE), where the polyketide is offloaded either through hydrolytic cleavage of the thioester bond or a cyclization reaction.
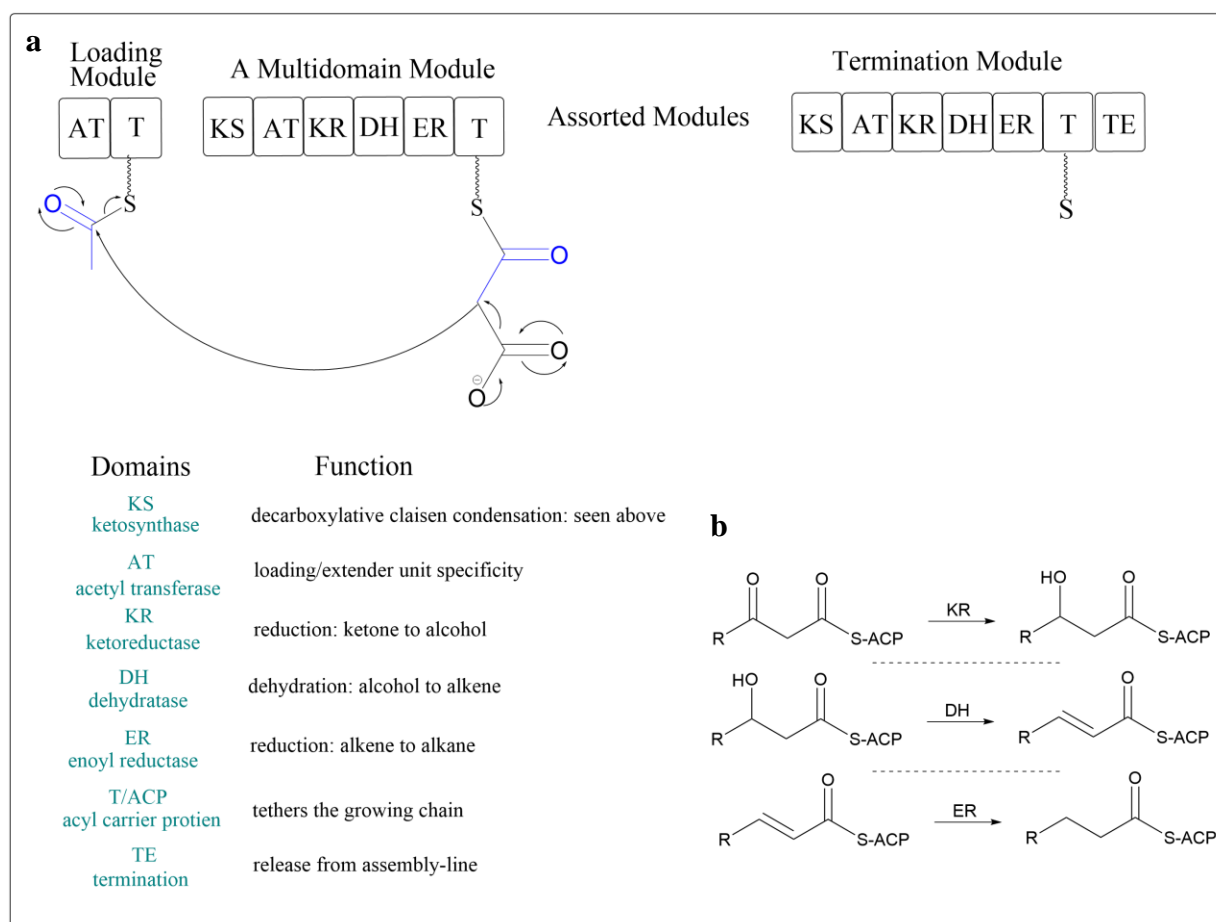
Figure 1-15. a) PKS module and domain organization. Two carbon units shown in blue in the diagram b) PKS reducing domains.

Polyketide modules themselves are often arranged into large genes or subunits containing multiple modules. Within this gene organization, there is usually a starting module (which loads starting units), multiple extending modules (which load extender units), and generally a termination module. Dependent on the type of PKS, the acetyl-CoA building blocks can vary or predictably rearrange malonyl-CoA, the basic extender unit utilized for decarboxylative Claisen condensation. There are four general types of PKS, with many exceptions reported. The first, modular type 1 (non-iterative), can use a variety of PKS extender units and utilize gene clusters composed of multiple modules with unique domain organizations and varying AT specificities. In modular type 1 PKS, most common in bacteria, unusual acetyl-CoA units diversify polyketides with various branching substituents at C-2 (**Figure 14-C**).[118] Next, iterative type 1 and iterative type 2 (II) use only malonyl-CoA extender units and, as the name implies, iteratively re-use the same modules to manufacture polyketides from this simple building block. Type II PKS produce aromatic structures. PKS type III are unique because they are composed only of a dimer of KS domains. Originally thought to be only plant enzymes, it is

Erythromycin
PKS type I modular assembly
post-assembly tailoring

Tetracenomycin
PKS type II assembly
tailoring reactions

Flaviolin
PKS type III assembly
monooxygenase tailoring reaction
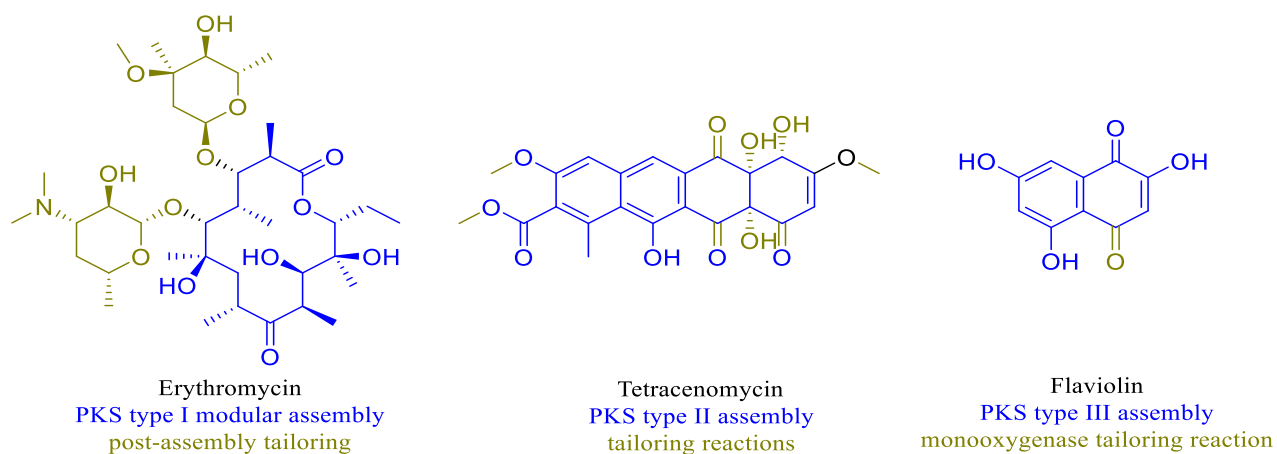
Figure 1-16. Various types of polyketide natural products produced by variations of enzymatic machinery.

now understood that there are bacterial PKS III natural products. PKS III enzymes typically make simple and often aromatic molecules, predominantly using malonyl-CoA as an extender unit, although some bacteria have been shown to use both methylmalonyl-CoA and ethylmalonyl-CoA extender substrates.[119] See **Figure 1-16** for examples of natural products made by each type of enzymatic machinery.

1.2.1.1.2   Nonribosomal Peptides

Nonribosomal peptides (NRP) are constructed from multiple amino acids linked via peptide bonds. They differ from ribosomal peptides in the amino acid building blocks. Twenty to twenty-two amino acids are used to construct ribosomal peptides, whereas over 200 amino acids are known to be utilized to construct NRPs. NRPs are constructed from large modular proteins termed nonribosomal polyketide synthetases (NRPS) in a parallel manner to polyketides. Modules are constructed of three repeated domains: an adenylation (A) domain, a peptidyl carrier protein (PCP, T for thiolation), and a peptide bond condensation domain (C) (**Figure 1-17**).[120,121] The A domain confers specificity towards a particular amino acid. The PCP domain holds new amino acid units as thioesters, whereby the nitrogen nucleophile attacks the electrophilic carbonyl C-1 held as a thioester on the C domain of the growing chain. Modules are organized into initiation modules with only A-CP domains and extender modules with C-A-PCP domains. Other common domains include MT, or methyltransferase, and E or epimerase. Genes will often be composed of multiple modules, with the biosynthetic machinery for large clusters composed of multiple genes with multiple modules. There is also a thioesterase domain (TE) found in termination modules, which determines the mechanism of release from the assembly and determines if the molecule is a cyclic peptide, a depsipeptide, or linear peptide, amongst other possibilities.[103,106]
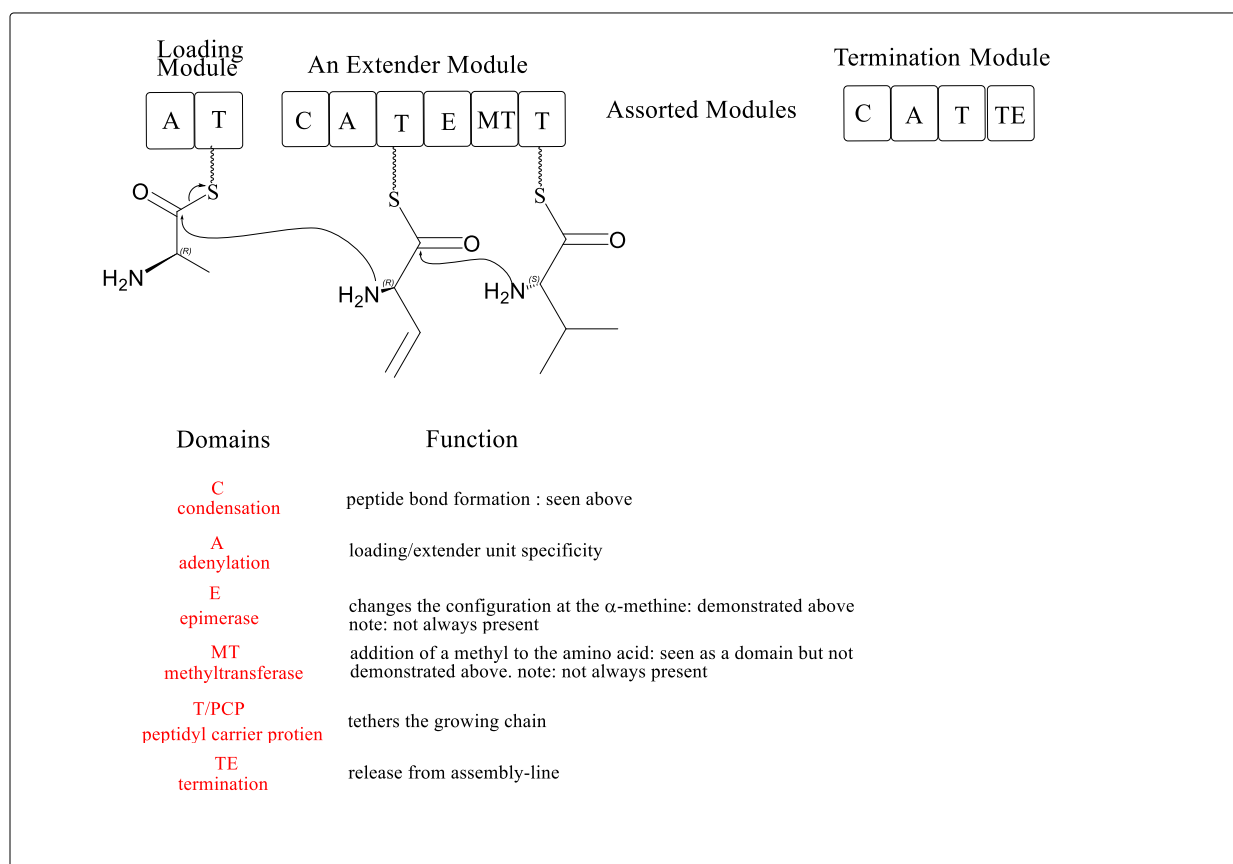
Figure 1-17. NRPS module organization and domain functions.

There are different types of NRPS gene clusters whose gene organization influences the type of nonribosomal peptides that can be biosynthesized. These types are linear, iterative, and non-linear. Module organization of linear NRPS accurately reflects the predicted primary sequence of the peptide, whereas iterative NRPS reuse modules to create peptides of repeated amino acids. Non-linear NRPS, as the name suggests, are non-predictive, with assembly of the natural product often involving a mixture of modules and free-standing domains.[120]

1.2.1.1.3 NRPS-PKS and PKS-NRPS Hybrids

PKS-NRPS and NRPS-PKS are hybrids that connect acetyl-CoA fragments with amino acids.[122,123,124] Nonribosomal and polyketide biosynthetic gene clusters use remarkably similar strategies to build a diverse set of molecules through the repetitive formation of C-N bonds and

26

C-C bonds, respectively. Combination of the two biosynthetic classes results in hybrid molecules incorporating both acetyl units and amino acids to craft an incredible array of complex molecules. NRPS/PKS hybrids with the PKS downstream of the NRPS module build C-C bonds through nucleophilic attack of the PK carbanion onto the carbonyl carbon on the growing molecular tethered to an NRPS thiolation domain.[122] Conversely, PKS/NRPS hybrids utilize the nitrogen on the downstream NRPS module to attack the carbonyl carbon on the PK T domain, building the molecule with the addition of a C-N bond.[122] Examples of these hybrids are abundant in bacteria and will be seen in **Chapter 2**.

1.2.1.1.4   Tailoring Reactions

In addition to the diverse construction strategies mentioned, both NRPS and PKS introduce diversity through tailoring enzymes. These tailoring enzymes can act on substrates tethered to the thiolation carrier proteins or post-assembly on the free substrate.[125] The known tailoring reactions that can occur to tethered substrates include cyclization, halogenation, methylation, oxidation, reduction.[125] The tailoring reactions that can occur post-assembly include oxidation, reduction, cyclization, methylation and halogenation along with the installation of prenyl groups, acyl groups, carbamoyl groups, and aminotransferases.[126] Examples of the results of tailoring reactions are seen in **Figure 1-16**.

1.2.1.2   Shikimic Acid Pathway

While glucose is the nutrient leading to both acetyl-CoA and shikimic acid, the shikimic acid pathway produces six-carbon precursor units directly produced from glucose through a four-carbon precursor, D-erythrose-4-phosphate, and a three-carbon precursor, phosphoenolpyruvate.[127,128] Shikimic acid was first isolated in the late 1800s from a fruit of a plant with the Japanese name *shikimi-no-ki*, and its structure was completely characterized in the

1930's.[129] Its role in the biosynthesis of the aromatic amino acids, L-phenylalanine (L-Phe), L-tyrosine (L-Tyr), and L-tryptophan (L-Trp) was elucidated in the 1950s. Additionally, polyketides can incorporate shikimate pathway genes for the production of starting units such as in cahuitamycins,[130] ascomycin and rapamycin (**1.24**).[103,106] At times, shikimate acid pathway metabolites will be incorporated in the post-assembly process as seen in ansatrienin (**1.25**).[131] Ansatrienin also incorporates 3-amino-5-hydroxy benzoic acid (AHBA) as its starter unit. AHBA is closely related to shikimic acid being also biosynthesized from D-erthyrose-4-phosphate and phosphoenolpyruvate. The AHBA synthase gene cluster utilizes the same enzymes and building blocks as the shikimate acid pathway, save for one enzyme that incorporates nitrogen and catalyzes aromatization as the final step in constructing AHBA.[132] AHBA itself is the starting unit for multiple polyketide natural products such as rifamycin and will be discussed in more detail in **Chapter 6**. Gallic acid (**1.26**) is a metabolite that has been completely produced from the shikimate pathway, with feeding studies analyzing ^{13}C splitting patterns demonstrating that it derives from D-erythrose-4-phosphate and phosphoenolpyruvate.[133] Shikimate pathway metabolites are present in a large variety of natural products produced by plants and microbes such as phenolics, flavonoids, stilbenes, chalcones, quinones, and the side-chain of the terpene-derived taxol.[128]

1.2.1.3   Alkaloids

Alkaloids are perhaps the most diverse of all the natural products with over 20,000 variations of alkaloids produced by a number of biosynthetic precursors.[134,135] Their common feature is a basic nitrogen atom incorporated into the molecule in diverse ways as long as it is not a peptide bond. The nitrogen's may originate from amino acids such as L-ornithine (L-Orn), L- arginine (L-Arg),
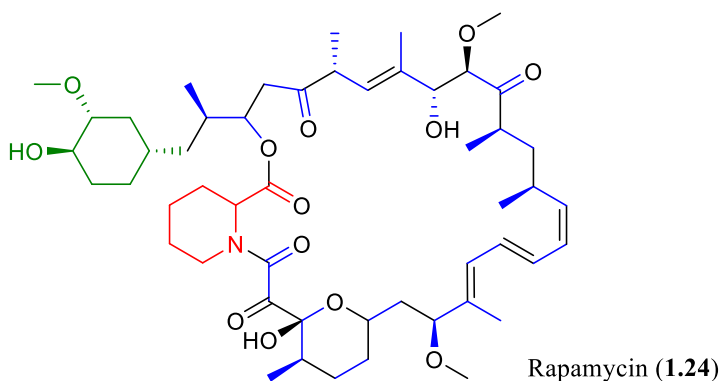
Rapamycin (**1.24**)

Shikimate pathway building block
PKS pathway building blocks
Amino Acid building block

Figure 1-18. Rapamycin, a mixed natural product utilizing a shikimate derived starting unit.
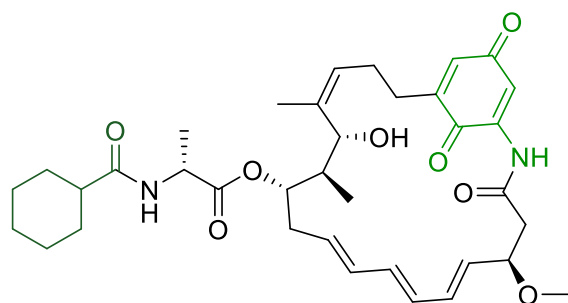
Ansatrienins (**1.25**)
shikimate derived
AHBA derived
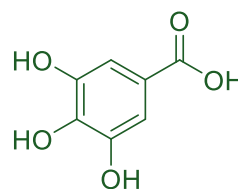
Gallic acid (**1.26**)
shikimate derived

Figure 1-19. Shikimate pathway and analog natural products.

L-tryptophan (L-Trp), or others.[135–137] Alkaloids are often organized by the name of the

heterocycle forming the core of their scaffold and encompass such classes as indolocarbazole

alkaloids (**1.27**) and pyrrole alkaloids (**1.28**). Additionally, they are often bioactive (**1.28** is a

powerful antibiotic) with fascinating, complex scaffolds (**Figure 1-20**).[106,138,139]

Citronamine A
dopamine derived
marine invertebrate origin

Staurosporine (**1.27**)
tryptophan derived
bacterial origin

Streptorubin B (**1.28**)
proline derived
bacterial origin
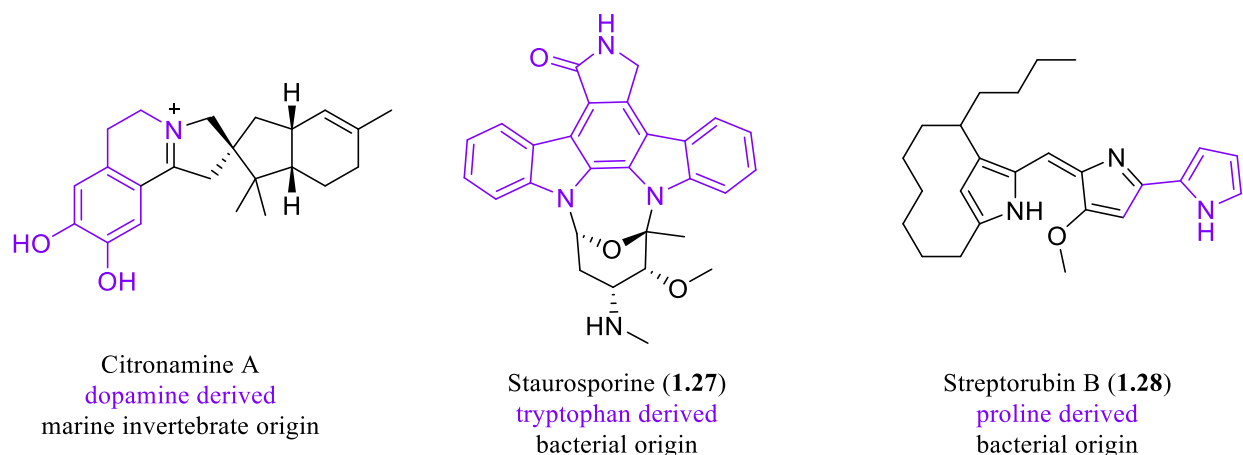
Figure 1-20. The diversity of alkaloid structures.

1.2.2 Genetic Foundation of Bacterial Natural Product Biosynthesis

Secondary metabolite genes are often clustered into co-regulated genes that contain all required biosynthesis, regulation and resistance genes for the successful production of a secondary metabolite.[111,140] These gene clusters often appear in "genomic islands," discrete regions found on bacterial chromosomes or plasmids. Actinomycetes, the focus of this thesis, tend to have genome sizes in the range of 3 Mb to 12 Mb. An important feature of these genomes is that they are G+C rich bacteria because they tend to use codons with a higher percentage of G+C (**Table 1-2**).[141] In secondary metabolism, open-reading frames (ORFs) consisting of G+C rich codons are often present in operons: multiple genes that are transcribed together onto one long mRNA.[142,143] Operons themselves can be clustered together into polyoperons: multiple operons under control of the same regulator for the purpose of a metabolic process.[144]

Amino Acid	Codon	Streptomyces Usage (%)
Alanine	GCC	60
	GCG	33
	GCA	5
	GCT	2

Table 1-2. Example of *Streptomyces* codon bias for alanine because of its high G+C content. [141]

On a practical level, this genome structure enables the identification and manipulation of relevant bacterial biosynthetic genes using bioinformatics and molecular biology tools. Another consequence of this co-regulation is that there is often an exact relationship between the order of the genes in a gene cluster and the order in which biosynthetic enzymes transform building blocks into complex secondary metabolites.[111] This association is called the colinearity rule, whereby the end product can be predicted by the gene sequence in a gene cluster. Once thought to be the norm, it is now understood that the colinearity rule is often broken.[111]

These differences in genes house the information for the production of biosynthetic enzymes[145] capable of catalyzing the multitude of chemical transformations leading to the diversity of natural products that we see. These biosynthetic enzymes often catalyze stereo- and regioselective chemical reactions. By characterizing these enzymes, they can be harnessed as tools for synthetic chemistry, they can be used in the rational design of synthetic biology and can lead to ever better predictions of gene cluster functions.

1.3   Research Goals

The work presented in this thesis was conducted in an effort to discover novel natural products with therapeutic properties. One strategy that was undertaken to accomplish this goal was to isolate molecules with a specific functional group that had previously been shown to be associated with bioactive properties. In Chapters 2, 3 and 4, I targeted the isolation of molecules containing a piperazic acid (Piz) residue. In order to do so, I developed a new method, combining genome mining with ^{15}N NMR. The premise was that nitrogen NMR would allow targeted access to bioinformatically predicted natural products containing Piz. This hypothesis was constructed because Piz has a unique nitrogen-proton correlation that also allows observation of Piz's spin system in an uncrowded region of the nitrogen NMR spectral window.

In Chapter 2, I describe the diversity of Piz-containing natural products and the bioinformatics work I carried out in order to genome mine Piz-containing strains for analysis. In Chapter 3, I describe the development of the ^{15}N NMR-guided method for targeting isolation of Piz-containing compounds from genome-mined strains. In Chapter 4, I describe the structure elucidation and bioactivity of the natural products I isolated using a ^{15}N NMR-guided method for targeting Piz-containing compounds. A second discovery strategy, explored in Chapters 5 and 6, was to use bioassay-guided fractionation to uncover molecules of interest in a small library of marine bacteria. In Chapter 5, I describe the construction of a mini-library and the outcome of bioassay-guided fractionations executed on mini-library strains with best activity. In Chapter 6, I describe how the bioassay results from the mini-library led to the structure elucidation of two natural products not previously found in wild-type bacteria. Based on my findings using this method, the goal was then to utilize stable isotope feeding studies to investigate the biosynthetic transformations leading to the 3,5-dihydroxyphenyl moiety present in one of the novel natural products. Finally, in Chapter 7, I conclude the thesis and detail possible future directions.

Chapter 2: Mining Bacterial Genomes for Piz-containing Natural Products

This chapter first describes the bioinformatic tools employed to identify biosynthetic gene clusters of interest. Known piperazic acid-containing molecules, the focus of my study, are then described, followed by a description of the biosynthesis of known molecules. Finally, the work done to identify gene clusters of interest is then described. This background sets the stage for the experimental work described in Chapter 3.

2.1   Bioinformatic Processes

From the first report of a sequenced bacterial genome in 1995,[146] continual advances in DNA sequencing technologies have enabled the sequencing of thousands of microbial genomes, with the current list of sequenced prokaryote genomes in the National Center for Biotechnology Information standing at just about two hundred and fifty thousand (https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/). Sequence, protein, and genomic databases have been constructed and are maintained by various governments around the world: EBI (Europe and the UK), NIG (Japan), and NCBI (USA). These databases store genomic information for multiple species, including humans, animals, plants, and microbes.[147,148] In terms of the microbial data gathered, sequenced genomic information can be used to deduce evolutionary relationships,[145,149,150] understand bacterial metabolic community pathways,[151] explore the metabolites of uncultivable microbes,[152] and has been increasingly used to access secondary metabolites through genome mining.[153]

The genome sequencing process begins with DNA samples copied into read lengths of contiguous nucleotide sequences dependent on the sequencing technology used. Sanger sequencing was the first developed sequencing technology, it is a slow and expensive procedure,

but it produces copies of DNA in fragments or "reads" of about 900 base pairs (bp). Multiple Next Generation Sequencing (NGS) technology platforms were developed to increase the speed of sequencing, but this decrease in time and expense came with the tradeoff of short read length (50-800 bp in length).[154,155] This tendency towards short reads is one of the drawbacks of NGS and has led to the "third revolution" in sequencing technology.[154,156] These new technologies produce much longer reads, with PacBio capable of sequencing an entire genome into one contig, but they are much slower than NGS and rely on quantified, high-quality DNA.[156] Nevertheless, Sanger and NGS sequencing are presently the most common; genomes are generally sequenced in strings of 800-10,000 base pairs. In this process, many repeated fragments are produced. Since genes and gene clusters of co-expressed genes can be larger[143] than the reads produced by sequencing, assembly is the method for "stitching" or assembling these fragments together. The fragments are assembled either by aligning to a previously analyzed reference genome or *de novo*. *De novo* assemblies use software built on algorithms that can string bigger and bigger reads together based on a statistical likelihood that they are from the same area of the genome.[155] Regardless of the method used for assembly (alignment or *de novo*), when the software can no longer find links between algorithmically stitched fragments, the areas that have now been linked together are termed contigs. With the proliferation of cheap NGS technology, there has also been a proliferation of draft genomes, having up to hundreds of contigs.[146] After contigs are assembled, nucleotide information is categorized into the genes through structural annotation,[157] and many genes can then be assigned putative functions through functional annotation, which can set the stage for identification of biosynthetic gene clusters.

2.2   Bioinformatics for Strain Selection Through Genome Mining

Characterization of gene clusters involved in the production of secondary metabolites has enabled the construction of various software programs or web-based interfaces (antiSMASH 4, BLAST, 2nd find, NRPS/PKS analysis https://www.secondarymetabolites.org) that allow for *in silico* inference of biosynthetic function for genes and biosynthetic gene clusters (BGCs). This inference is based on searching for homologous sequences that possess either nucleotide or amino acid sequence similarity between the characterized enzyme's gene sequence and an unknown open reading frame (ORF). Homologous sequences share common ancestry – either from gene-duplication events (paralogs) or speciation events (orthologs).[158] It has been shown that homologous sequence with a high percentage of sequence similarity likely also share similar functions. BLAST or Basic Local Alignment Search Tool is the foundation of most functional annotation tools as it is linked to the GenBank database. There are various BLAST platforms, including BLASTN that will search nucleotide databases using a nucleotide query, and BLASTP that searches protein databases using a protein query.[159] BLAST finds places of local similarity between a query and sequences in its database; termed pairwise alignment. From there, BLAST then displays a percentage of the two sequences that matched (identity %) and how much of the query and the database sequence matched (similarity %). In addition, it gives a statistical probability of how much the identity match was due to random chance – the smaller this statistical probability, the greater the chance that the query and the database sequence are homologous sequences.[159] With the above stated, there is a caution that although a certain percentage of sequence similarity generally indicates common ancestry,[145] the link between sequence and function is less straight forward.[158] This is particularly true of nucleotide sequence homology so the use of BLASTN can be less of informative. Due to this, genome mining efforts

which compare the amino acids sequence of a query against amino acid sequences in the database (using the BLASTP algorithm) homology utilizing BLASTP will have a higher likelihood of uncovering homologs with both sequence and function similarity.

The other drawback from this sequence similarity approach is that proteins may share a functional similarity, even if evolutionary events have introduced a great deal of sequence changes. In this case then, two sequences may not share a high degree of sequence similarity but may have identical functions.[158] PSI-BLAST is a BLAST algorithm that works to overcome this potential limitation – searching for distantly related sequences sharing function.[158] PSI-BLAST uses an initial BLASTP search then constructs a multiple sequence alignment of those amino acid sequences to create a scoring matrix to compare sequences on the next iteration.[160] During this process it takes into account characterized protein models and conserved protein domains. However, PSI-BLAST is an iterative algorithm, it must be run upwards of 10 to 20 times for the same search query, each time ideally favoring functionally similar proteins. However, if a non-functionally similar protein is introduced due to protein domain similarity, the subsequent iterations may distort the outcome towards non-functionally similar sequences, leading to false positives.[158] A way to account for the weaknesses of both pairwise alignment and multiple sequence alignment-based algorithms is to run both for a complete analysis.

With the above bioinformatic tools, it can be generally predicted if a BGC encodes for nonribosomal peptide synthetases (NRPS) and/or polyketide synthases (PKS) as these modules and domains have predictable sequence motifs across multiple bacterial species.[161,162] The program antiSMASH, which is an acronym for "antibiotics and secondary metabolites analysis shell," allows users to specifically search finished and draft genomes for gene clusters that code for secondary metabolites like those of NRPS or PKS origin (to name but a few).[163]

antiSMASH will find clusters within sequenced data, provide the functional annotation, list the closest homologous sequences from within the closest characterized secondary metabolite biosynthetic pathways, give a breakdown of A and AT domain specificities in NRPS and PKS modules, and provide a likely linear structure based on colinearity rules. 2ndFind is a web-based tool that will find secondary metabolite gene clusters and list putative functions for the genes within the gene clusters (biosyn.nih.go.jp/2ndfind/). It does not give predictions for A(NRPS) or AT(PK) domain specificity, and as such, must be coupled with a secondary bioinformatic tool such as the web served NRPSpredictor[164] or NRPS/PKS Analysis[165] for that analysis. These were the main tools used in the genome mining research work discussed in this thesis.

Due to the colinear rule, the constitution of bacterial natural product can often be predicted[162] by analyzing the gene organization, the A domain (NRPS) specificity, and the AT domains (PKS) specificity of a gene cluster. This is the application of the bioinformatic processes to the genome mining of natural products. This approach has led to an array of discoveries of novel molecules including salinilactam[21] (**2.1**), tambromycin[166] (**2.2**) and avermipeptin B[167], thailandamide A[168] (**2.3**), leptolyngbyalide A (**2.4**), and bacillaene[169] (See **Table 2.1**). Genome mining efforts allow us to explore the genetic machinery of sequenced bacteria for gene clusters of interest. In this thesis, I used genome mining tools to sift through the thousands of sequenced bacteria in public databases to find bacteria predicted to produce Piz-containing natural products.
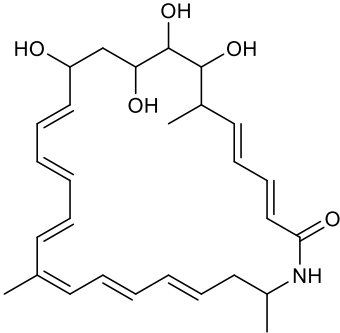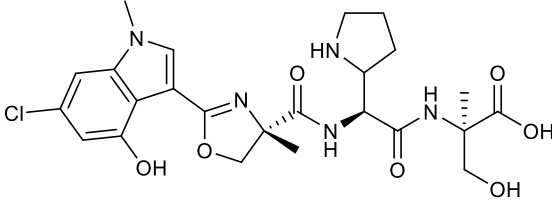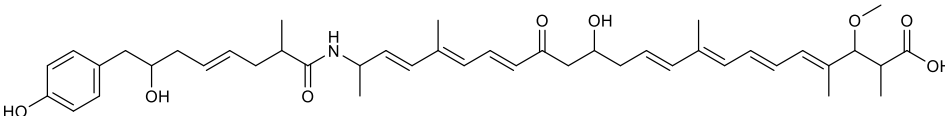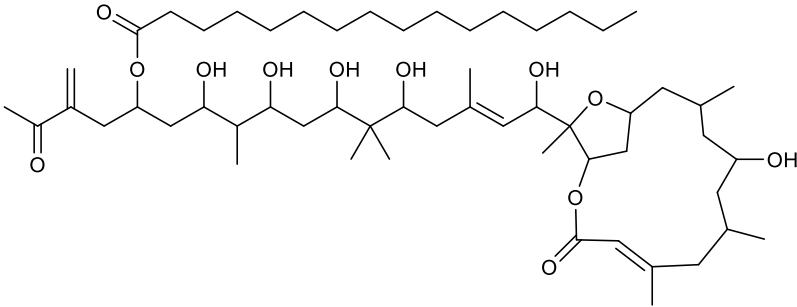
A Small Subset of Natural Products Discovered Through Genome Mining	
Obtained from canonical gene clusters	
Salinilactam (**2.1**)	
Tambromycin (**2.2**)	
Obtained from non-canonical gene clusters	
Thailandamide A (**2.3**)	
Leptolyngbyalide A (**2.4**)	

Table 2-1. Selected natural products discovered using genome mining techniques.

2.3 Piperazic Acid-Containing Natural Products

As natural products are produced by biosynthetic enzymes in biological systems, novel

chemical scaffolds are often linked with enzymes using novel catalytic strategies.[170] Pursuing

chemical diversity can lead us to useful drug leads and pharmaceuticals while expanding our
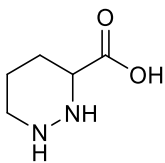
Piperazic acid (2.5)

chemical and biological knowledge; this can inform rational efforts at drug discovery and drug design. Piperazic acid is one such chemically diverse natural product moiety. Piperazic acid (**2.5**) is a cyclic hydrazine amino acid that has been found in diverse natural products, many of which have potent biological activities. Piperazic acid is a unique non-proteinogenic amino acid in several ways. First, it is a rare amino acid with a nitrogen-nitrogen bond and the only known example of a natural cyclic hydrazine amino acid.[171] Second, as an amino acid residue, particularly as the unsaturated derivative, it has a constrained conformation that locks the amino acid in place and can thereby induce β-turns in peptides.[172] This unique feature makes piperazic acid and its congener dehydropiperazic acid potentially useful for engineering peptides with β-turns. Third, natural products with piperazic acid residues often have multiple rare structural features and unique biological activities,[173] and as such, methods for specifically isolating piperazic acid-containing molecules have the potential to lead to particularly potent natural products. To highlight the utility of finding new piperazic acid-containing peptides, I will catalogue the known piperazic acid-containing natural products to demonstrate the breadth of structural diversity.

2.3.1 Linear Peptides

Some of the most exciting and unusual piperazic acid-containing molecules are linear peptides. As a group, the linear nonribosomal piperazic acid peptides are marked by their structural diversity and incorporation of additional unusual amino acids and functional groups. These molecules can be sorted into a few smaller families: antrimycins (**2.6**), Sch 382582.3 (**2.7**), matlystatins/YM.24074 (**2.8, 2.9**), padanamides/actinoramides (**2.10, 2.11, 2.12, 2.13**), and cahuitamycins (**2.14, 2.15**). Even within each of these families, of which representatives of each are shown here, structural diversity can be found.[173]
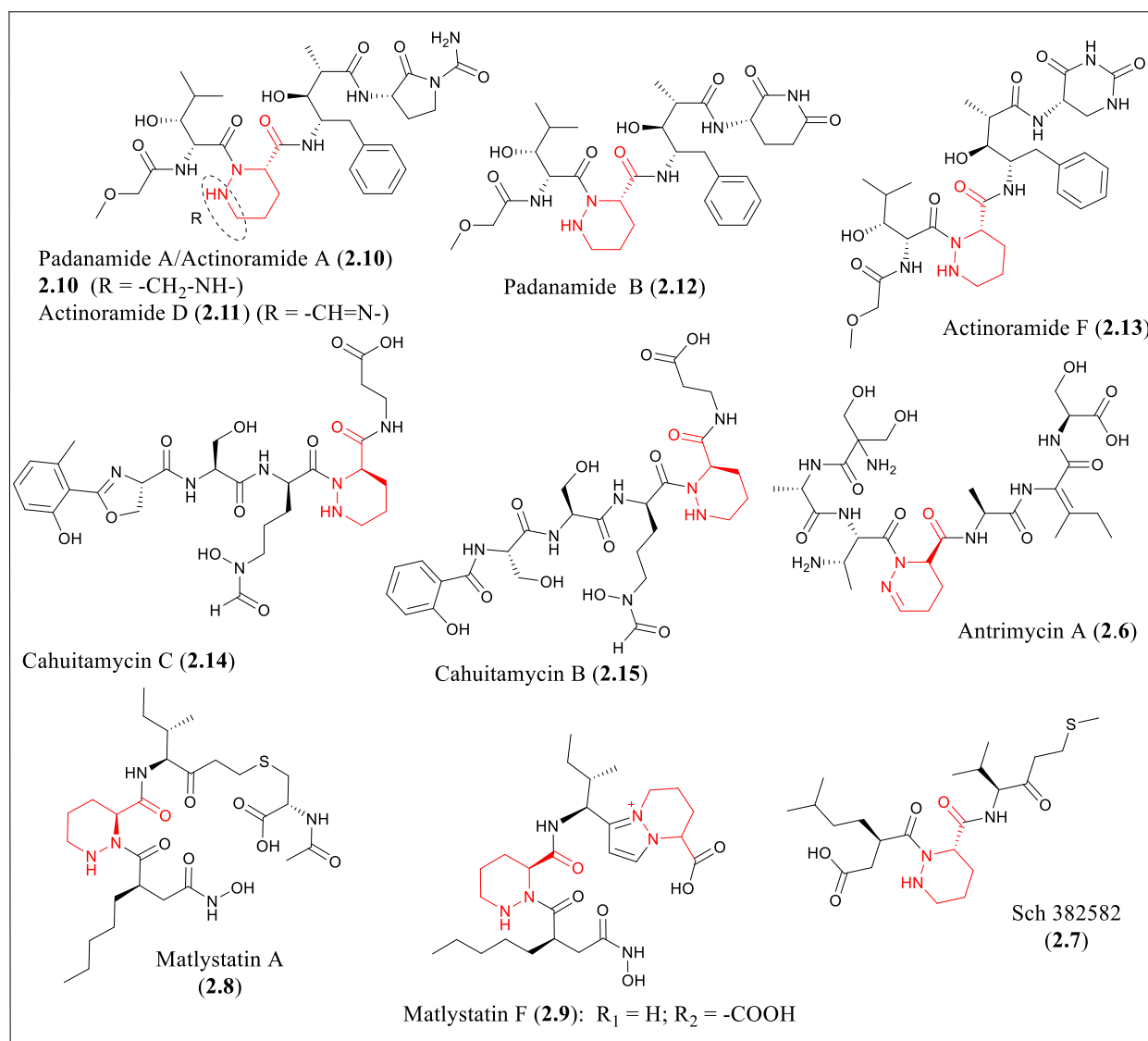
Figure 2-1. Linear Piz-containing peptides.

2.3.2 Cyclic Peptides

Composed entirely of amino acids and cyclized by the formation of a peptide bond this piperazic acid-containing family contains one known cyclic tetrapeptide, glomecidin (**2.17**),[174] and multiple hexapeptides (**2.18 to 2.22**), of which representatives are pictured in **Figure 2-2.**[173]

40
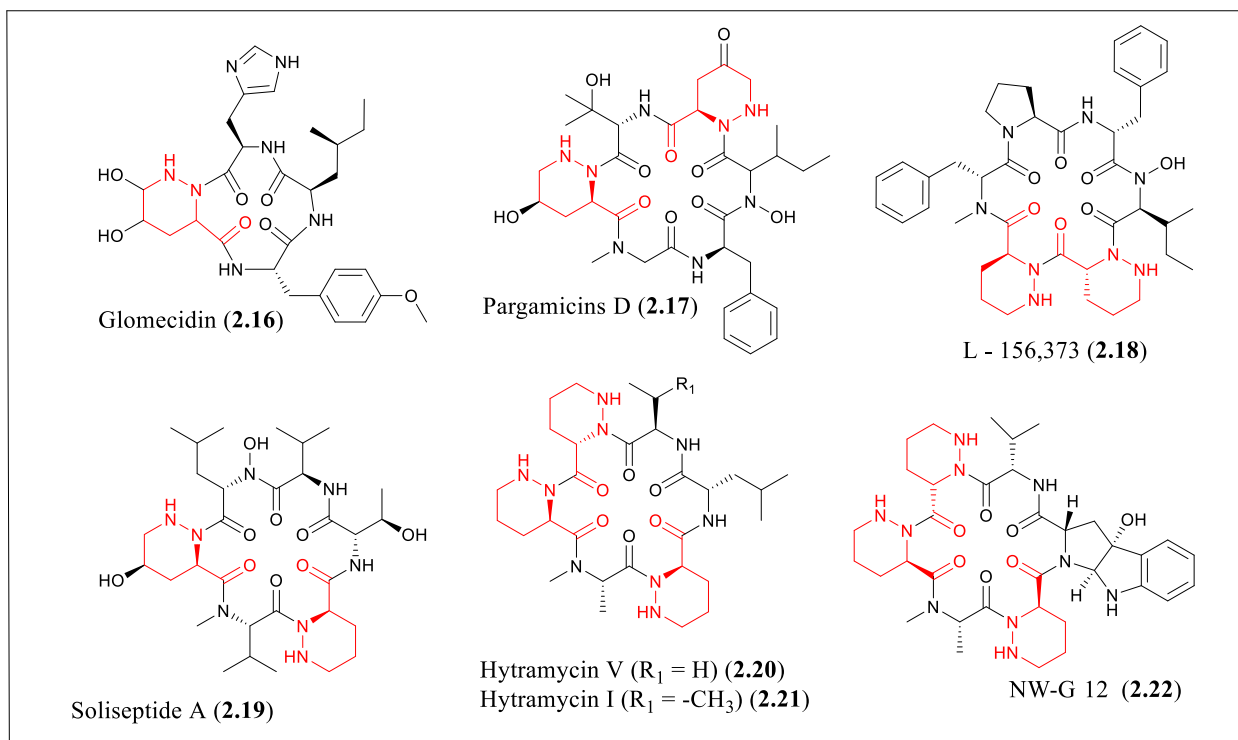
Figure 2-2. Cyclic Piz-containing peptides.

2.3.3 Depsipeptides

Depsipeptides, peptides where one of the amide linkages is replaced with an ester, are the largest group of nonribosomal peptides containing piperazic acid. The majority of these depsipeptides are cyclodepsipeptides, with several exceptions described below. This group can be further subdivided into 13-membered, 18-membered, and 19-membered cyclodepsipeptides.[173]

2.3.3.1 13- and 18-membered Depsipeptides

The antimycobacterial lydiamycins (**2.23**) are the sole examples of piperazate-containing 13-member cyclodepsipeptides. The first family of piperazate-containing molecules on record is the 18-membered depsipeptide family of monamycins (**2.24**) isolated from *Streptomyces jamaicensis* during bioassay-guided screening efforts against *Fusarium oxysporum cubense* by Hassel and coworkers in 1959 and whose structures were elucidated in the early 1970's. Additional

18-membered cyclic depsipeptides, of which a subset is shown in **Figure 2-3**, include the piperazimycins (**2.25**), gerumycin A (**2.26**), the svetamycins (**2.27**), the piperidamycins A (**2.28**),[175] along with the related linear peptides JBIR.39 (**2.29**) and JBIR.40 (**2.30**)[176]. The kutznerides (**2.31**) also belongs to this class.[173]
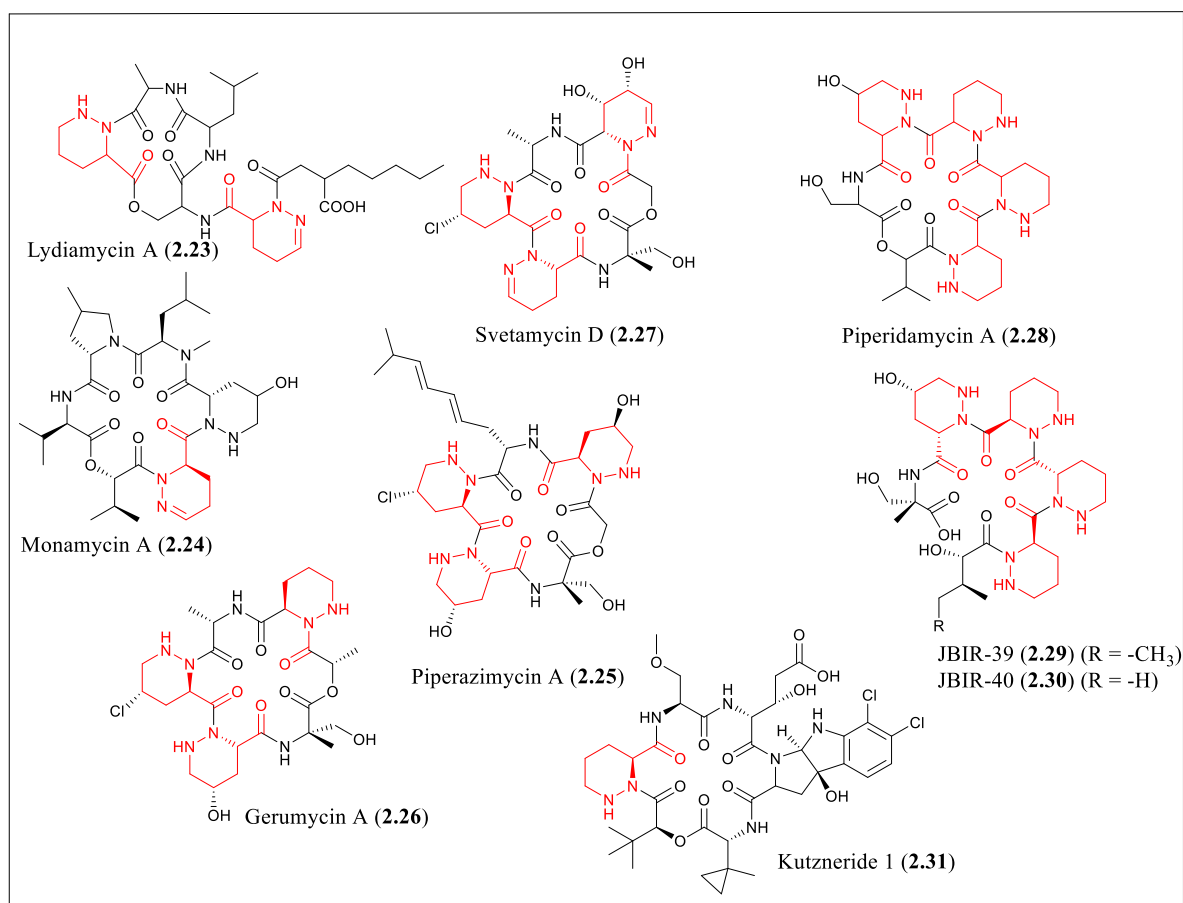
Figure 2-3. Structures of 18-membered cyclodepsipeptides and their linear congeners.

2.3.3.2   Azinothricins: 19-Membered Cyclodepsipeptides

The largest group of piperazic acid-containing molecules is the azinothricins named for the azino sub-structure of piperazic acid. The structures in this family are 19-membered cyclodepsipeptides with acyl side chains, isolated from various *Streptomyces* and related strains. Many of them appear to be hybrid PKS-NRPS derived molecules. The structures of the molecules are remarkably similar

across the family, often with similar residues at similar positions, suggesting conserved biosynthetic pathways. Most of these members contain an oxane moiety linking the amino acid sequences of the molecule with the acetyl derived portions of the molecule. Close examination of the substituents of the oxane ring, along with feeding studies in verucopeptin's and polyoxypeptin's producers suggest a divergence in biosynthetic strategies with a large portion of the acyl tail containing a strictly polyketide origin vs those whose acyl tail itself has origins from both amino acids and acyl-CoA precursors.[173]

2.3.3.2.1   19-Membered Cyclodepsipeptides With Long Acyl Tails

Azinothricin (**2.32**), A83586C (**2.33**), citropeptin (**2.34**),[177] GE3 (**2.35**),[178] kettapeptin (**2.36**)[179], verucopeptin (**2.37**), and diperamycin (**2.38**), are all 19-membered cyclodepsipeptides whose long acyl chains are derived from polyketide acetyl units.[180]

	R₁	R₂	R₃	R₄
Azinothricin (**2.32**)	-CH₂CH₃	-CH₂OCH₃	-CH₃	-CH₃
A83586C (**2.33**)	-CH₂CH₃	-CH₃	-CH₃	-H
Citropeptin (**2.34**)	-CH₃	-CH₂OCH₃	(isobutyl)	-H
GE3 (**2.35**)	-CH₃	-CH₃	-CH₃	-H
Kettapeptin (**2.36**)	-CH₂CH₃	-CH₂OCH₃	-CH₃	-H
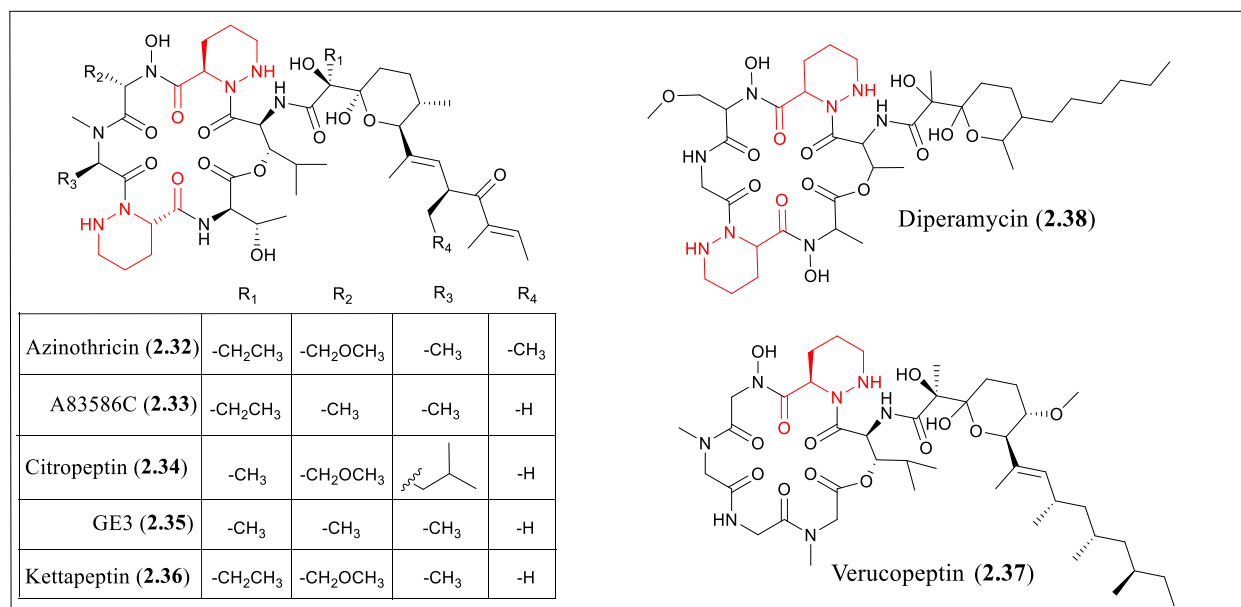
Figure 2-4. 19-membered cyclodepsipeptides with long acyl tails.

2.3.3.2.2   19-Membered Cyclodepsipeptides With Short Acyl Tails

L-156,602 (**2.39**) variapeptin (**2.40**),[177] the aurantimycins A.D (**2.41 to 2.44**), the polyoxypeptins A and B (**2.45, 2.46**), the oleamycins (**2.47, 2.48**), IC101 (**2.49**), the closely related pipalamycin (**2.50**) all have distinctly shorter substituents on the oxane sub-structure of the azinothricins. A subsection is demonstrated here with the rest discussed in **Chapter 4**. Feeding studies completed on polyoxypeptin suggest that the branched acyl chain attached to the oxane ring is produced from an intact isoleucine (Ile) unit.[181] However, additional biosynthetic characterizations build evidence that many branched PKS extender units are constructed through a crotonyl-CoA pathway.[118,182]
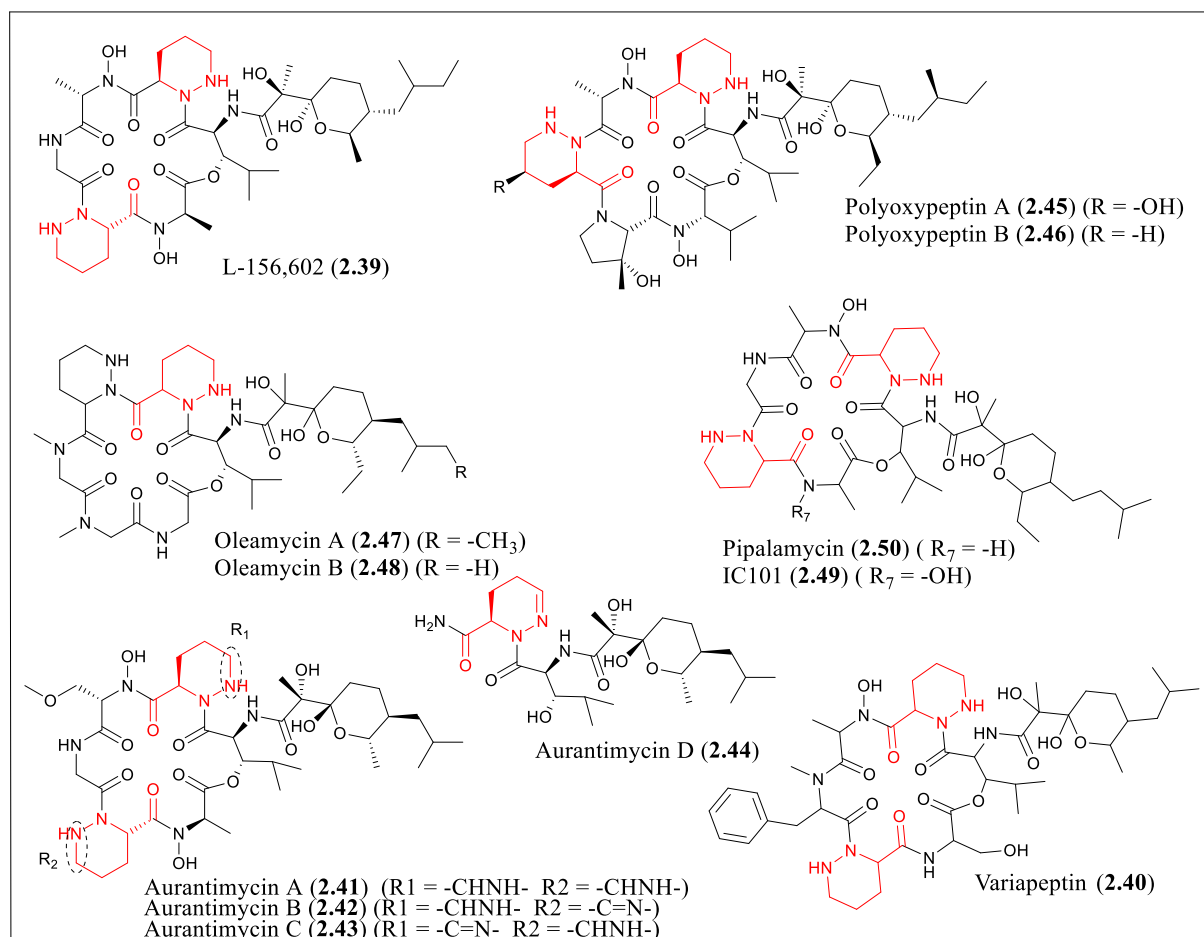
Figure 2-5. 19-membered depsipeptides with short acyl tails.

44

2.3.3.2.3   19-Membered Cyclodepsipeptides With NRPS Derived Acyl Tails

This subset does not have an oxane-based PKS tail. Marformycins A-F (**2.51**) were isolated from a *Streptomyces* strain isolated from sediment at 1396 m depth in the South China Sea (**Figure 2-6**).[173] They feature a unique *N*-formylated side chain while containing only a single piperazic acid and an *O*-methylated tyrosine. Related natural products are the depsidomycins (**2.52**), and the piperastatins (**2.53**).[173]

Marformycin D (**2.51**)   Depsidomycin B (**2.52**)   Piperastatin B (**2.53**)
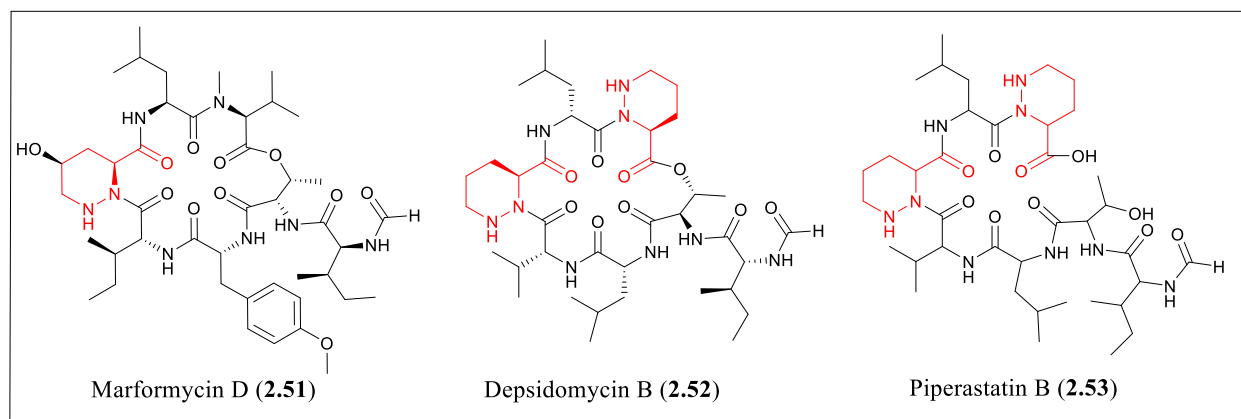
Figure 2-6. 19-membered depsipeptides with formamide tails.

2.3.4   Sanglifehrin, Mollemycin, Dimeric Compounds and the Bisintercalators

Increasing the structural diversity of Piz-containing compounds, sanglifehrin (**2.54**), mollemycin (**2.55**), himastatin (**2.56**), chlopostatin, luzopeptin (**2.57**) and quinoxapeptin uniquely add to known natural products. As of now, each stands in a family on its own. Altogether it can be seen that piperazic acid-containing natural products are rich with diversity and, as such, are excellent leads for natural product discovery efforts.[173]
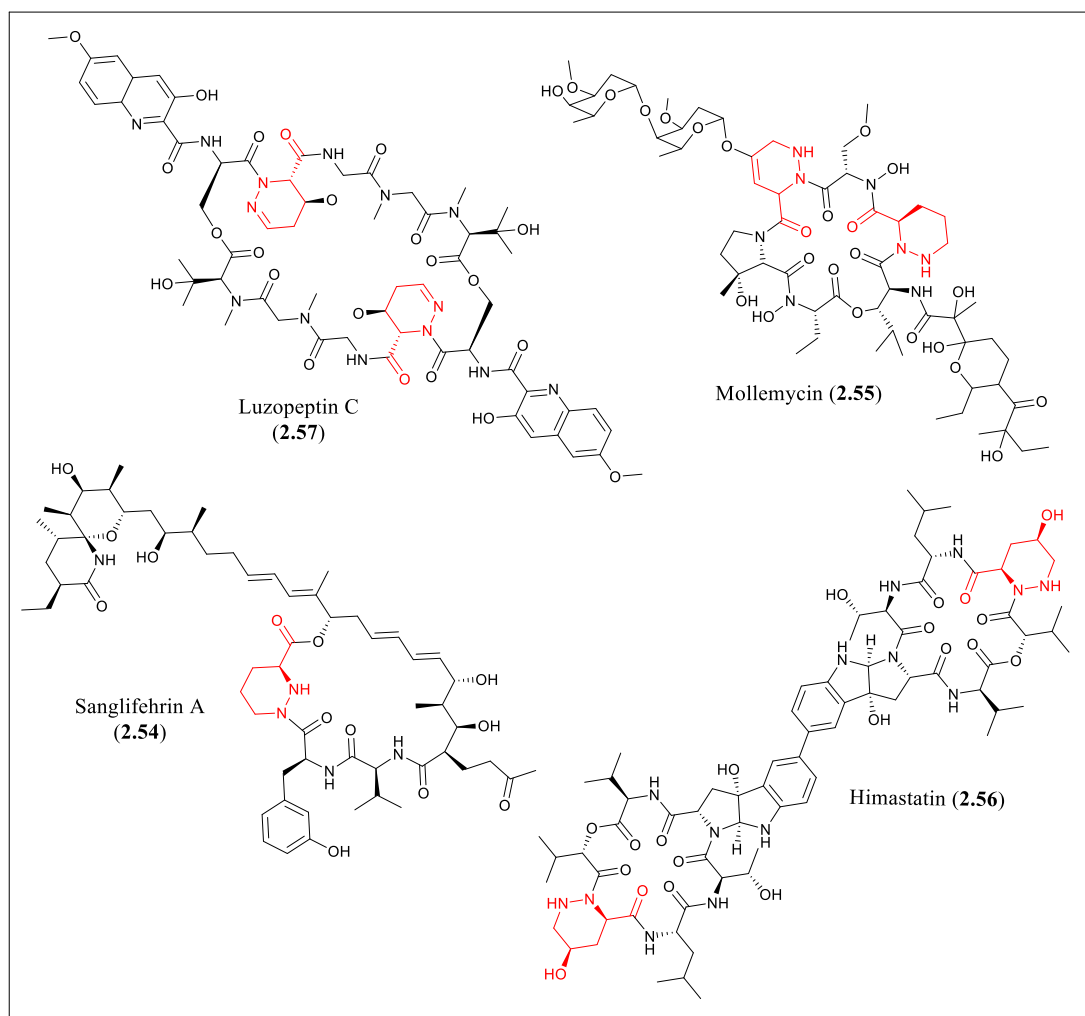
Figure 2-7. Other Piz-containing natural products.

2.4   Biosynthesis of Piperazic Acid-Containing Molecules

All known piperazic acid-containing molecules come from actinobacteria, although recent work has predicted their presence in more widely flung bacteria.[183] The first published gene cluster of a Piz-containing natural product was for kutznerides (**2.31**) (*ktz*), an 18-membered depsipeptide showing it to be produced by a nonribosomal peptide synthase with three major NRPS modules and the enzymes required for pre and post-assembly modifications (**Fig 2.8**).[184] This gene cluster

encodes three multidomain nonribosomal peptide synthetase proteins, KtzE, KtzG, and KtzH, collectively encoding six adenylation domains, with one of the A-domains in KtzH split into two pieces by a methyltransferase domain, consistent with the structures of the kutznerides. The KtzE module contains five domains (A,T,E,C,T) and was postulated to load in the unusual amino acid 2-(1-methylcyclopropyl)- D-glycine. The KtzG module, which contains three domains (A, KR, T), with a valine-for-aspartate substitution in the A domain, was shown to have substrate specificity for a ketoacid - specifically α-ketoisovaleric acid. Other standalone adenylation domains KtzB and KtzN, were shown to be involved in the assembly of unusual amino acids or as replacements for inoperable A domains in the larger modules.[184] Subsequent studies by the Walsh and Garneau-Tsodikova groups interrogated the biosynthesis of these amino acids. Among the biosynthetic arsenal leading to the kutznerides are enzymes KtzQ and KtzR, flavin-dependent halogenases;[185] KtzD and KthP, Fe(II), α-ketoglutarate dependent halogenases;[186,187] KtzA, the acyl-CoA dehydrogenase-like flavoprotein that forms a cyclopropyl ring;[187] KtzP and KtzO, non-heme iron oxygenases that catalyze opposite stereospecific hydroxylations;[188] and a methyltransferase domain in KtzH that catalyzes *O*-methylation of a thioester-tethered serine.[189] Additionally, the enzyme KthP, encoded outside of the immediate boundaries of the cluster, catalyzes stereospecific chlorination of a thioester-tethered L-piperazic acid, giving the (*S*)-5-chloro-L-piperazic acid congener. Since the publication of the *ktz* cluster in 2007, thirteen additional gene clusters for piperazic acid-containing molecules have been published covering a majority of the major structural classes for these molecules (**Table 2-2**).[173]
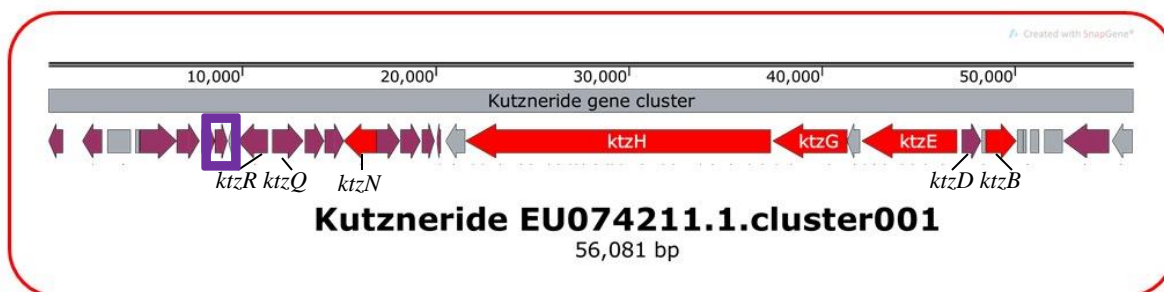
Figure 2-8. Kutzneride gene cluster map with key genes and modules named. Red genes indicate nonribosomal peptide origins. Purple square indicates *ktzT*, the piperazate synthase.

Compound family	Compound	Natural product class	Year	Accession no
Linear peptides	Padanamides	NRP/PK	2013	KC915040
	Cahuitamycins	NRP/ Shikimic	2016	KU363800
	Matlystatin	NRP/PK/ terpene	2017	NZ_KB907724
Cyclic peptides	NW-G (AFN)	NRP	2018	MH497044
18-membered depsipeptides	Gerumycin A & B	NRP	2015	CP010990
	Gerumycin C	NRP	2015	CP011868
	Kutzneride	NRP	2007	EU074211
19-membered depsipeptides	Polyoxypeptin	NRP/PK	2014	KF386858
	Aurantimycin	NRP/PK	2016	NZ_AOPZ01000147.1
	Dentigerumycin A	NRP/PK	2015	CP011862
	Dentigerumycin B	NRP/PK	2017	NZ_MWFK01000002
	Marformycin	NRP	2015	KP715145
Dimeric	Himastatin	NRP	2011	FR823394
Sanglifehrin	Sanglifehrin	PK/NRP	2011	FJ809786

Table 2-2. List of characterized Piz-containing gene clusters.[173]

Additional nonribosomal peptide synthase gene clusters were characterized for the gerumycins (**2.26**, **4.12-4.13**), an 18-membered depsipeptides family. Gerumycin A and B were shown to be produced by *Pseudonocardia sp*. EC080625-04, interestingly by a gene cluster with two loci separated by ~92 kilobase pairs, while gerumycin C was produced by *Pseudonocardia sp*.

HH130629-09 (**Fig 2.9**). The dimeric depsipeptide himastatin was also shown to be produced by nonribosomal peptide machinery in *Streptomyces himastatinicus* ATCC 53653 genome (**Fig 2.10**). The cyclic peptide family, the alboflavusins (formerly the NW-G series), were most recently shown, somewhat expectedly, to be produced in *Streptomyces alboflavus* strain 313 by a nonribosomal peptide synthetase (**Fig 2.10**). Biosynthesis is proposed to involve the construction of piperazic acid as a free amino acid from L-ornithine, then loading of amino acids onto a nonribosomal peptide synthetase assembly line.[190]

Gene clusters characterized for the linear peptides (**Figure 2-10**) all show distinct construction strategies, perhaps reflective of their structural diversity. Padanamides and actinoramides are linear tetrapeptides marked by unusual and distinct heterocycles at the C-termini of the molecules. Characterization of the padanamide gene cluster demonstrated that biosynthetic production involves a hybrid non-canonical NRPS/PKS gene cluster with the N-carbomylation of an intermediate nonribosomal amino acid deciding the formation of the final heterocycle.[191] Cahuitamycins are the only the shikimic acid-NRPS hybrid produced Piz-containing natural product with a shikimic acid pathway-derived starter unit while the rest of the scaffold is produced by NRPS modules. Specifically, cahuitamycins A and B's aromatic starter unit is produced from the salicylate synthase enzymes, CahI and CahJ while cahuitmycin C, on the other hand, has a salicylate derived starter unit – 6-methylsalicylate starter unit, - whose enzymatic machinery is encoded by genes in a second distinct locus. Matlystatins, yet another fascinating linear peptide family, are produced from a gene cluster that utilizes precursors from the PKS, the NRP, and the terpene pathways. This multi-hybrid biosynthetic strategy may help explain the fascinating structures that make up this family. Not a linear peptide, but also utilizing a unique hybrid

biosynthetic strategy, sanglifehrin's (**2.53**) gene cluster loads in Piz as its starting unit and then

completes the rest of its biosynthesis with polyketide machinery.
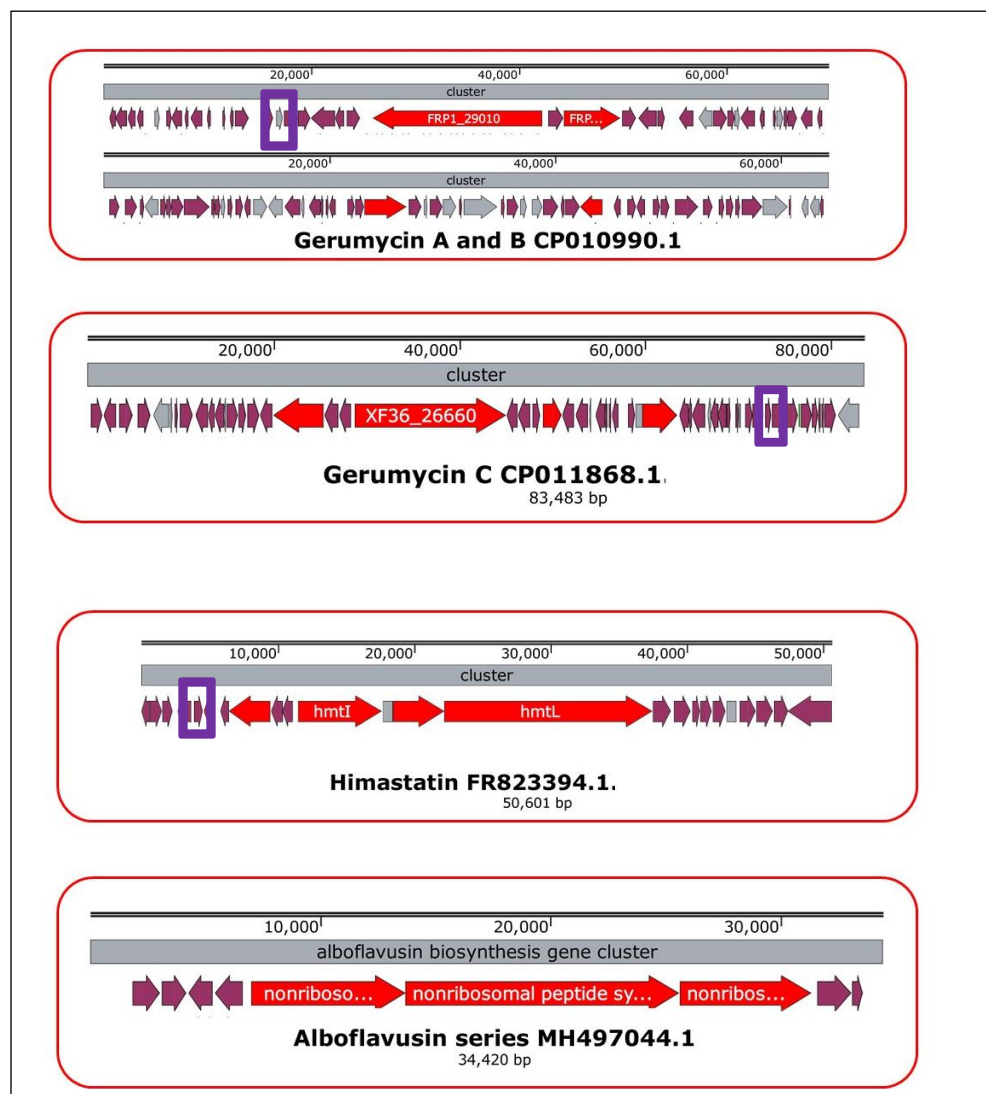
Figure 2-9. Piz-containing natural products produced by nonribosomal peptides synthases. Red genes indicate non-ribosomal peptide origins. Purple square indicates *ktzT* homolog - the piperazate synthase.

Finally, the gene clusters have been characterized for the 19-membered depsipeptides.

Gene clusters for polyoxypeptin (**2-45**) and aurantimycin (**2-41**) have been published (**Figure 2-11 and Figure 2-14**).[192,193] Each of these gene clusters is large (~75 kb for polyoxypeptin, and ~60 kb for aurantimycin) and contain both Type I polyketide synthase and nonribosomal peptide

synthetase genes, as expected from the structures of the corresponding molecules. For polyoxypeptin, deletion of any of the NRPS genes *plyC*, *plyD*, *plyQ*, *plyI*, and *plyS* abolishes the
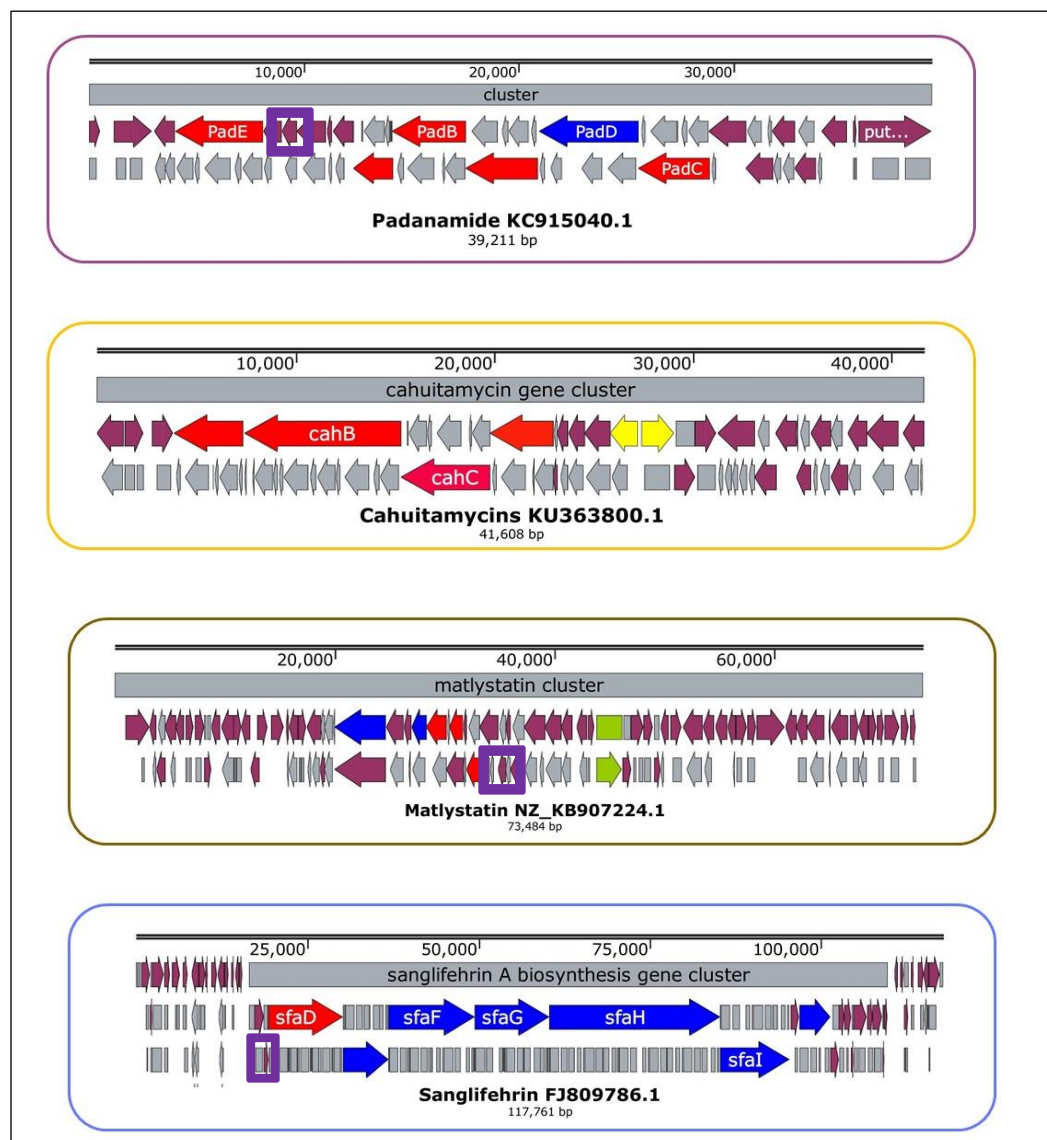
Figure 2-10. Piz-containing natural products produced by hybrid gene clusters. Red genes indicate non-ribosomal peptide origins. Blue genes indicate polyketide origins. Yellow genes indicate shikimate pathway origins. Green genes indicate terpene origins. Purple square indicates *ktzT* homolog - the piperazate synthase.

production of polyoxypeptins.[192] Additionally, deletion of the P450 gene *plyM* abolishes

polyoxypeptin A production and accumulates polyoxypeptin B, suggesting that PlyM catalyzes

hydroxylation of polyoxypeptin B to give polyoxypeptin A. The structure of the C_{15}-acyl chain of

51

polyoxypeptin suggests that the unusual extender unit 2-(2-methylbutyl)malonyl-CoA is required for its assembly. Prior labelling studies have shown that this 2-(2-methylbutyl)malonyl could derive from isoleucine,[181] possibly via a transamination reaction. However, deletion of a candidate aminotransferase gene (*plyN*) did not abolish production of polyoxypeptins, suggesting that the aminotransferase might be elsewhere in the genome.[192] In the case of the related compound aurantimycin, the extender unit should instead be 2-(2-methylpropyl)malonyl, which could possibly arise from a similar transamination reaction, instead from valine. For the aurantimycin gene cluster (*art*), deletion of the *artG*, a candidate NRPS gene, abolishes production of aurantimycins, experimentally linking the gene cluster to production of aurantimycins.[193]

Dentigerumycin A is another example of an actinobacteria-derived azinothricin depsipeptide,[194] produced by a NRPS-PKS hybrid gene cluster in a *Pseudonocardia* strain isolated from an ant (*Apterostigma dentigerum*). It has four main multidomain NRPS modules whose organization is: (C,A,T,TE); (E,C,A,T); (C,A,T); (T,E,C,A,T) and multiple standalone condensation, adenylation, and thioesterase domains (**Figure 2-11**). Three structurally similar molecules, named dentigerumycins B-D (**4.5-4.7**), were isolated from another strain, *Streptomyces* sp. M41 also from an ~100 kb NRPS-PKS hybrid gene cluster, albeit with five main multi-domain NRPS modules (one of which is immediately adjacent to the PKS modules) and multiple standalone NRP domains (**Figure 2-11**). The linear structures of dentigerumycins C-D are suggestive of a premature off-loading of a hybrid PK-NRP intermediate prior to incorporation of the final L-threonine residue. In 2018, another congener, dentigerumycin E (**4.8**), was isolated from the co-culture of a marine-derived *Streptomyces* sp. JB5 with a *Bacillus* sp., and a putative gene
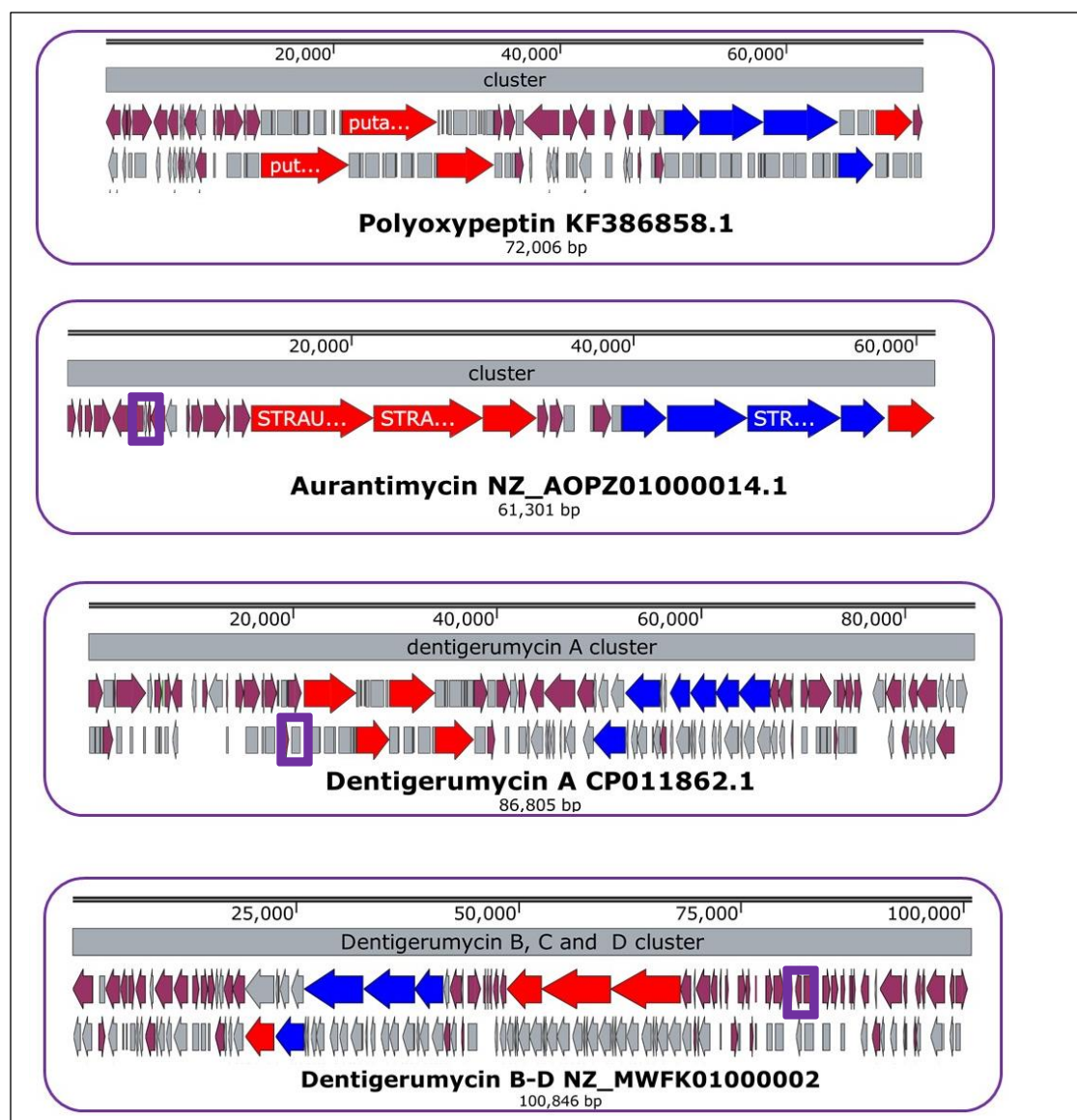
Figure 2-11. Piz-containing natural products produced by nonribosomal peptide/polyketide hybrids. Red genes indicate nonribosomal peptide origin. Blue genes indicate polyketide origin. Purple square indicates *ktzT* **homolog - the piperazate synthase.**

cluster for **4.8** was identified from the genome of *Streptomyces* sp. JB5 however, the gene sequence for this gene cluster has not been deposited with an open-access database.[195]

Ju and co-workers isolated the marformycin gene cluster (*mfn*), which encodes three multidomain NRPS modules (MfnC, MfnD, MfnE), three standalone NRP domains, along with a series of tailoring enzymes (**Figure 2-12 and Figure 2-13**).[196] They established that the tailoring

53

enzyme, MfnG, *O*-methylates free D/L-tyrosine, before loading this amino acid onto the NRPS. Additionally, they provided evidence that MfnA catalyzes *N*-formylation and that the cytochrome P450 MfnN catalyzes hydroxylation of the piperazic acid residue as the terminal step in the biosynthesis. Finally, they demonstrated that MfnO (a Type IV aminotransferase) and MfnH (a 116-amino acid protein of unknown function) work in concert to catalyze the interconversion of diastereomers L-Ile and L-*allo*-Ile[197], giving a precursor to the marformycins.
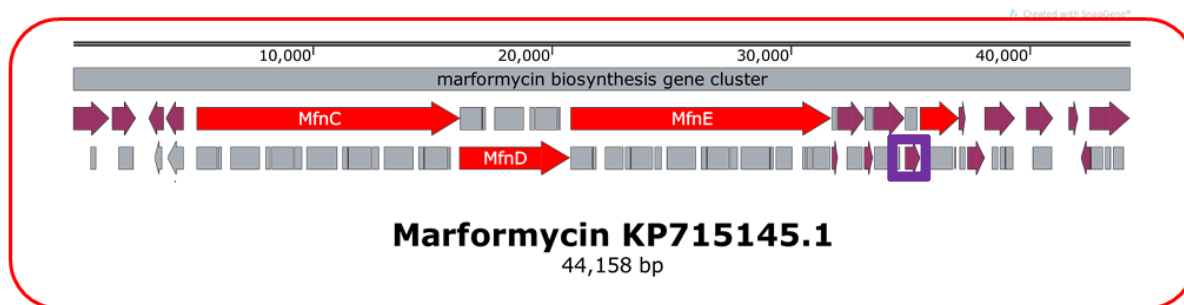
Figure 2-12. The marformycin gene cluster. Red genes indicate nonribosomal peptide origin. Purple square indicates *ktzT* homolog - the piperazate synthase.
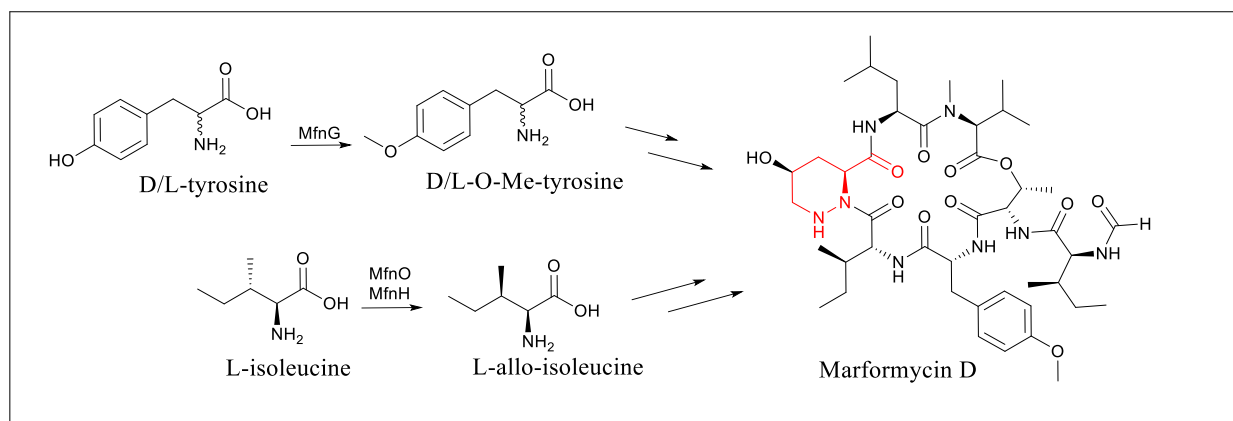
Figure 2-13. Characterized transformations leading to marformycin biosynthesis.[173]

54
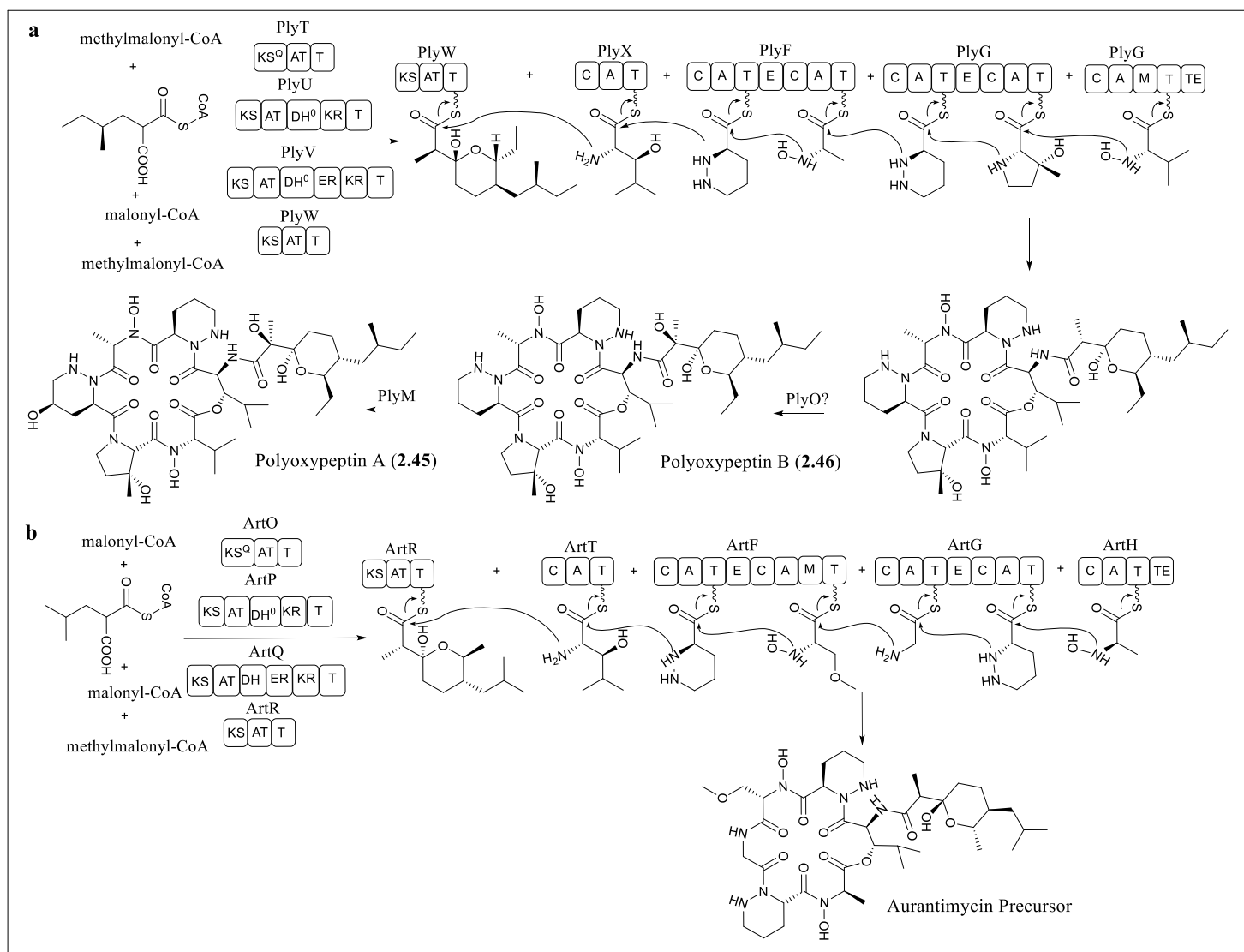
Figure 2-14. Characterized and hypothesized transformations leading to: a) polyoxypeptin biosynthesis b) aurantimycin biosynthesis.[173]

All these gene clusters have a common feature in the hydrazine piperazic acid amino acid that has driven much of their interest. The biosynthetic origins of Piz were of curiosity since the first elucidation of monamycins in the early 1970s. Piz's biosynthetic route was probed using radiolabelled ornithine in the 1970s, research whose result erroneously led to the conclusion that L-ornithine was not a precursor. In the 1990's, the biosynthetic pathway of valanimycin, a nitroso-containing natural product, was investigated, which suggested that a hydroxylamine was involved in the formation of its nitrogen-nitrogen bond.[198] When the Walsh group was exploring kutzneride's (**2.31**) biosynthesis in 2012,[199] they noticed a L-Orn N-hydroxylase homolog gene (*ktz*I) in **2.31**'s gene cluster. This observation, coupled with the work on the biosynthesis of valanimycin, led them to re-investigate L-Orn as a Piz precursor. Walsh and coworkers identified KtzI, encoded in the gene cluster corresponding for the piperazic acid-containing kutznerides, as a flavin-dependent L-ornithine *N*-hydroxylase and demonstrated that free *N*-OH-L-ornithine (**2.57**) is an intermediate on the pathway to piperazic acid (**2.5**).[199] This crucial result established that the δ-nitrogen of ornithine is activated as a hydroxylamine. The question of how the N-N bond was formed was still a mystery until Yiling Du of the Ryan lab recognized that a small open reading frame *-orf4* in the kutzneride gene cluster was misannotated as a regulator. Additionally, homologous open reading frames were found in many characterized Piz-containing gene clusters with either adjacent or nearby *ktzI* homolog genes. Subsequent work by the Ryan lab identified this enzyme as KtzT, a heme-dependent, but not oxygen-dependent, piperazate synthase (**Figure 2-15**), with both *ktzI* and *ktzT* genes generally present in every gene cluster linked to a piperazic acid-containing molecule.[173]
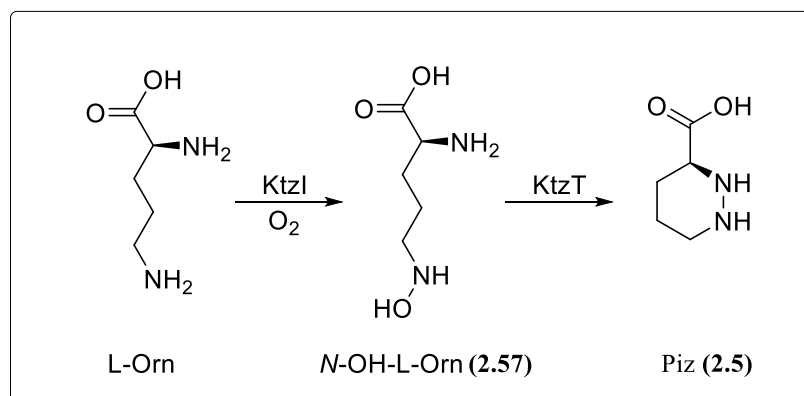
Figure 2-15. Biosynthetic pathway to piperazic acid involves N-hydroxylation of L-ornithine by KtzI and N-N bond formation by KtzT.[173]

2.5 Bioinformatic Analysis Using the Piperazate Cyclizing Gene *ktzT*

From the characterized biosynthetic pathways of piperazic acid-containing natural products, it is evident that production of this family involves many non-canonical pathways. For one, cahuitamycins employ two different starting units, the enzyme machinery for one of them usually found at a great distance from the rest of the producing cluster.[130] Secondly, the machinery for characterized 19-membered depsipeptide gene clusters show non-colinearity with multiple unattached TE domains and other anomalies.[173] Thirdly, the piperazate synthase itself has been shown to be an unusual heme-dependent enzyme that does not rely on oxygen for N-N bond formation.[200] In addition to the interesting biochemistry that accompanies the production of piperazic acid-containing natural products, many of these natural products possess a wide variety of potent bioactivities[173] and, as such, there is the correlation that the discovery of new piperazic acid-containing natural products will have a high likelihood of revealing novel bioactive drug leads. With this logic in hand, I chose to target Piz-containing natural products through genome-mining.

In vitro biochemical characterization had confirmed that the KtzT (Accession no. ABV56600.1) enzyme, found within the kutzneride gene cluster (**Figure 2-8**), was the piperazate synthase involved in the production of the N-N bond found in piperazic acids.[200] This same work by the Ryan lab confirmed that the N-hydroxylated L-Orn (**2.57**) product of KtzI was the piperazate synthases' substrate. Finally, as all known characterized gene clusters for Piz-containing natural products contain the NRPS modules necessary for incorporation of a nonribosomal amino acid, it can be expected that any Piz-producing gene cluster contains an NRPS module. KtzH is the NRPS module within the *Kutzneria* sp. 744 (**Figure 2-8**) whose A domains have been predicted to specify for piperazic acid.[199] Thus, in short, the minimal enzymatic machinery so far understood to be involved in the production of piperazic acid-containing NRP's, is a piperazate cyclase, a hydroxylase, and an NRPS module or modules, which have A-domains with specificity for Piz amino acids. Both hydroxylases and NRPS modules are common across many types of associated natural products. Thus, to target the discovery of the Piz amino acid, I hypothesized that the piperazate synthase should be unique to Piz-containing natural products. With this in mind, I hypothesized that the KtzT (Accession no ABV56600.1) amino acid sequence could be utilized as a genome mining probe in BLASTP[201] searches of sequenced bacterial genome assemblies contained within the NCBI database to find bacterial strains with Piz-producing potential. However, the gene cluster associated with the production of the kutznerides, and thus containing KtzT, which belongs to *Kutzneria* sp. 744, is distantly related to the more ubiquitous *Streptomyces* genus within the Actinobacteria phylum.[20,157] Due to this taxonomical distance, I accepted 50% sequence homology to KtzT to indicate a putative piperazate synthase homolog. I expected that true piperazate synthases would be found near or adjacent to an ORF homologous to a lysine/ornithine hydroxylase as well as an

NRPS module homolog. These requirements, then, added secondary and tertiary pieces of information to build evidence towards a strain's ability to produce Piz-containing natural products. The sequenced genomes found through BLASTP containing an ORF with high sequence homology to KtzT were visually inspected using the NCBI in-house genome browser to determine if they contained proximal hydroxylases (KtzI homologs) and NRPS machinery (KtzH homologs). On occasion, a draft genome would contain KtzT and KtzI homologs on a short contig that displayed no NRPS machinery. In these instances, BLASTP searches were completed against the entire draft genome sequence of the respective strain using KtzH (Accession no. ABV56588.1) as a query. In this manner, I filtered down the approximately 200 KtzT homologs to a shorter list of about 30 strains of sequenced bacteria that had all aspects of genetic coding for the necessary machinery. The majority of these (a subset is shown in **Table 2-3**) were from the *Streptomyces* genus and all of them from the Actinobacteria phylum.

Strain	KtzT homolog Accession no.	Accession no. of contig containing KtzT homolog
Streptomyces sp. NRRL F-6131	WP_030310216.1	NZ_JOHN01000049.1
Streptomyces sp. NRRL B-1347	WP_030688266.1	NZ_JOJM01000099.1
Streptomyces tsukubensis NRRL18488	EIF94522.1	AJSZ01000008.1
Streptomyces sp. NRRL F-2747	WP_051805955.1	NZ_JOIS01000036.1
Streptomyces sp. NRRL F-5053	WP_030885178.1	NZ_JOHT01000002.1
Streptomyces sp. NRRL S-920	WP_051819333.1	NZ_JODF01000006.1
Streptomyces katrae NRRL B-16271	WP_030301364.1	NZ_JNZY01000034.1
Streptomyces phaeochromogenes NRRL B-1248	WP_055612424.1	NZ_LIQZ01000085.1
Streptomyces griseus subsp. griseus NRRL WC-3645	WP_037672886.1	NZ_JOJE01000006.1
Streptomyces griseus NRRL F-5144	WP_030722408.1	NZ_JOGA01000032.1
Streptomyces sp. NRRL F-2664	WP_030773938.1	NZ_JOFX01000040.1
Streptomyces sp. NRRL B-1140	WP_053672929.1	NZ_LGEA01000110.1
Streptomyces incarnatus NRRL 8089	AKJ15811.1	CP011498.1

Table 2-3. Names of selected strains containing KtzT homologs.

2.6   Bioinformatic Prediction of Selected Strains

From within the smaller subset of strains that contained homologs to the three (presumed)

necessary enzymes - KtzT, KtzI, and KtzH - the appropriate contigs were analyzed using both

the bioinformatics software pair 2[nd]Find (http://biosyn.nih.go.jp/2ndfind/) with NRPS/PKS

finder[165] and the standalone antiSMASH.[163,202,203] This effort 1) confirmed that a contig or series

of contigs within a (draft) genome contained homologs to KtzT, KtzI and KtzH; 2) allowed

analysis of the A domain prediction; and 3) produced a linear sequence prediction. Through this

analysis, I was ideally hoping to confirm that the clusters of interest had A domains specific to

piperazic acid. Initial searches through NRPS/PKS finder and antiSMASH3 did not find any Piz

specific A domains. There were, however, multiple A domains for which the software programs

could not predict the loaded amino acid – termed X - and the assumption was made that some of

these X's were, in fact, Piz. With the release of antiSMASH4,[163] the included SANDPUMA ensemble algorithm[204] proved capable of predicting most of the A-domains with Piz specificity. The gene clusters of some of the chosen strains were run back through antiSMASH4 to confirm initial assumptions.

Further analysis of characterized gene clusters, as described in **Section 2.4**, suggested that the larger NRP multidomain modules appear to be involved in the factory-like assembly of the coded natural product. In contrast, the smaller stand-alone NRP modules or domains appear to be involved in the production of pre-assembly transformations for unusual amino acids. Given the non-linear nature of many of the gene clusters, the A domains of multidomain NRPS modules were compared against characterized gene clusters (**Figures 2-9 to 2-12**). The putative amino acid sequence predictions and known compounds' structures were also compared as a dereplication process. These comparisons filtered the hundreds of possible hits to a list of 13 strains (**Table 2-3 and Table 2-4**) that were obtainable from the US government's Agricultural Research Service Culture Collection (NRRL) collection.

Strain	Predicted NP class	Predicted NRPS A- domain colinear sequence for the A domains within main NRPS modules.	Closest known Cluster/
Streptomyces sp. NRRL F-6131	NRP/PK	CATECAT + CATECAT + CAnMTTTE D-ala ala X X X	Sanglifehrin
Streptomyces sp. NRRL B-1347	NRP	CATECAnMTTCATECATTE + CATECAT + CAT D-ala ala D-leu ala D-val ala X	Kutzneride
Streptomyces tsukubensis NRRL18488	NRP/PK	Contig 7 CATECAT CATECAT CAnMTTTE D-pip ala D-pip ala nOH-ala	Polyoxypeptin
Streptomyces sp. NRRL F-2747	NRP	CATCATCATE CATE CATCATCATTE val thr aromaticX D-goh-val pip leu pip	Marformycins
Streptomyces sp. NRRL F-5053	NRP	AT AT TE AT CAT C TE TC X val pro leu	Vazabitide A
Streptomyces sp. NRRL S-920	NRP	Contig 92 Contig 107 CATEC — C CATECAT X D-ala ala	Himastatin
Streptomyces katrae NRRL B-16271	NRP/PK	CATECAT CATECAT CATTE pip ser ala ala thr	Aurantimycins
Streptomyces phaeochromogenes NRRL B-1248	NRP	CATCATECATCATECAnMTTTE X ala ala ala ser	Himastatin
Streptomyces griseus subsp. griseus NRRL WC-3645	NRP	C ATCAT A cys X pip	Kutzneride
Streptomyces griseus strain NRRL F-5144	NRP/PK	Contig 40 Contig 41 CATEC — T CATECAT C X ala ala	Polyoxypeptin
Streptomyces sp. NRRL F-2664	NRP/PK	Contig 12 CAT CATECAnMTTCATECATTE ATECAT ile D-X NMe-val D-X Orn D-phe X	Himastatin
Streptomyces sp. NRRL B-1140	NRP/PK	CATEC A AT C ala X X	Polyoxypeptin
Streptomyces incarnatus NRRL 8089	NRP/PK	CATECAT CATECAT CATTE CATTE D-pip nOH-ala ala pip thr gly	Polyoxypeptin

Table 2-4. Selected strains with the encoding organization of the multidomain NRPS modules with predicted A-domain specificity.

2.7 Conclusion

In conclusion, the ability to cheaply sequence bacterial genomes has led to large databases containing tens of thousands of sequenced bacterial genomes. The concurrent characterization of secondary metabolite enzymes has led to the ability to predict which classes and structures of natural product a bacterial strain can putatively produce. Bioinformatic software, based on the biological characterization of the strings of nucleotides, can aid in the analysis of the prodigious amounts of data.

Within natural products, piperazic acid-containing molecules are an exciting class of natural products due to their structural diversity and often potent biological activities. Using the amino acid sequence for the highly specific KtzT piperazate synthase as a probe to sort through the thousands of sequenced bacteria contained within open-access biological databases, I was able to select a small group of thirteen bacteria. These bacteria possessed genomes that were predicted to code for the enzymatic machinery necessary to produce piperazic acid-containing natural products.

After this effort, the next question became how to find suitable culture conditions for Piz-containing metabolite production and isolate the pure metabolites from crude culture extracts in sufficient quantities to allow structure elucidation of their chemical structures. The first step in this complex process was to observe the Piz amino acid from within the complex of bacterial metabolite to determine if the bacteria was even producing the genomically-predicted Piz-containing natural product under specific laboratory conditions.

2.8   Experimental Section

2.8.1 General Experimental Procedures

Genome mining efforts were completed using browser based BLASTP[201] followed by utilization of the NCBI website gene viewer for manual curation of contigs/genomes. Obtained contigs or complete genomes were further analyzed using the web based 2ndFind analytical software (http://biosyn.nih.go.jp/2ndfind/) in conjunction with NRPS/PKS Website analysis[165]and antiSMASH 3, 4 and 5[202,203,205] with antiSMASH 4 using the SANDPUMA[204] algorithm generally used for identification of Piz-linked A-domains.

Chapter 3: Nitrogen NMR-guided Isolation of Piz-containing NRPs

3.1 Chapter Overview

Methods for the focused isolation of low-abundance natural products with specific chemical sub-structures could expand known bioactive chemical diversity for drug discovery. As discussed, Piz-containing molecules have structural novelty and bioactivity that often accompanies the incorporation of this nonribosomal amino acid. To target Piz-containing molecules, I envisaged developing a method for selecting strains by genome mining for piperazic acid biosynthetic genes, as discussed in **Chapter 2**, followed by NMR-based tracking of fractions for chemical shifts uniquely associated with the piperazic acid residue itself. Such an approach would allow me to target low-abundant metabolites. The first step in using nitrogen NMR experiments to discover Piz-containing compounds was an analysis of the Piz NMR signature. To validate this method, I employed the known Piz-containing natural product producer *Streptomyces* sp. RJA2928. This strain was previously shown to produce the padanamides[206] through a gene cluster[191] containing a KtzT homolog.[200] It was then necessary to determine if a piperazic acid signal was visible in *Streptomyces* sp. RJA2928's crude extract. As a point of comparison, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC-TOCSY experiments were conducted on the crude extract, but the spectra obtained proved too complex to assign a Piz spin system. Simultaneously, ^{1}H-^{15}N HSQC NMR experiments were conducted on *Streptomyces* sp. RJA2928 crude extract and the spectra demonstrated that the Piz N-H correlation (δ_H 4.52, δ_N -299.8) was visible if Piz was present. However, given the crude NMR spectra's low signal-to-noise ratio, it was not possible to unequivocally rule out a false positive based on this N-H correlation. Two additional steps were developed: 1) feeding the labelled ^{15}N$_2$-L-Orn precursor to the natural product producers to increase the Piz N-H ^{15}N NMR signal and 2) carrying out a second

experiment, the ^{1}H-^{15}N HSQC-TOCSY NMR experiment, to access the diagnostic spin system (**Table 3.2**) in a less crowded spectral window. Additionally, as the ^{1}H-^{15}N HSQC-TOCSY NMR experiment is less sensitive than the ^{1}H-^{15}N HSQC NMR experiment, incorporation of the ^{15}N-labelled Piz precursor ^{15}N$_2$-L-Orn would allow access to the ^{1}H-^{15}N HSQC-TOCSY NMR experiment in a crude extract and when smaller quantities of the natural product were present.

Upon confirming that the ^{15}N NMR Piz signal and the spin system (with labelling) were visible in *Streptomyces* sp. RJA2928's crude extract, bacteria previously selected through genome-mining (see **Chapter 2.5 and 2.6**) were grown on various agar media. The presence of the ^{15}N NMR Piz signal was then used to select a bacterium for large-scale growth in a medium in which the microbe was producing Piz-compounds. The species *Streptomyces incarnatus* NRRL 8089 showed a clear ^{15}N NMR Piz signal when grown on ISP4 agar medium and had a gene cluster bioinformatically predicted to have an appreciable amount of novelty. Then, *Streptomyces incarnatus* NRRL 8089 was grown with ^{15}N-labelled Piz precursor ^{15}N$_2$-L-Orn added to ISP4 agar media. The crude extract containing labelled material was assessed for the N-H correlation and then a Piz spin system utilizing the ^{1}H-^{15}N HSQC-TOCSY NMR experiment. These NMR experiments on the labelled crude extract allowed unequivocal observation of the presence of a Piz-containing natural product from the extract. To obtain enough material for final isolation and structure elucidation, a larger amount of this bacterium was grown without a labelled precursor. The labelled and unlabelled extracted material were combined. The ^{15}N NMR Piz signal and spin system were then used to guide isolation of Piz-containing metabolites from the *Streptomyces incarnatus* NRRL 8089 bacterium resulting in four novel compounds, three of which I isolated and then elucidated the constitutional framework as part of this thesis work.

3.2 Nitrogen NMR

Nuclear Magnetic Resonance (NMR) is a standard method for structure elucidation of organic molecules. With steady instrument improvements and new techniques, useful NMR data can be increasingly gathered with less material and more complex molecules.[44] Of the commonly found elements in natural products – hydrogens (i.e. protons), carbon, oxygen, nitrogen, and phosphorous– experiments can be run to directly probe small variations in structural environments for ^{1}H, ^{13}C,^{14}N, ^{15}N, and ^{31}P nuclei (**Table 3.1**). Indirectly, chemical shift and carbon multiplicity can help to infer the presence of electronegative oxygen.[207,208] The primary nuclei investigated by NMR in routine natural product structure elucidation are ^{1}H and ^{13}C. In particular, ^{1}H, with its large natural abundance and large gyromagnetic ratio, is a nucleus for which it is relatively easy to obtain NMR data with even nanogram-scale material.[209] The ^{13}C nucleus is far less abundant with a smaller gyromagnetic ratio; collecting data for a natural product's carbon skeleton requires more material and longer collection times. However, with indirect detection and high field strength, it can be obtainable at the microgram scale.[209,210] Nitrogen has two NMR visible isotopes, ^{14}N and ^{15}N. ^{14}N is the more abundant isotope; however, ^{14}N signals suffer from significant line broadening due to quadrupolar relaxation.[211] Thus, ^{15}N is the more commonly observed nitrogen isotope.[211]

Nuclide	Spin I	Natural Abundance (%)	Relative Sensitivity	Gyromagnetic ratio γ	NMR frequency [MHz]
^{1}H	1/2	99.985	1.00	26.7519	100
^{13}C	1/2	1.108	1.59 x 10-2	6.7283	25.144
^{14}N	1	99.63	1.01 x 10-3	1.9338	7.224
^{15}N	1/2	0.37	1.04 x 10-3	-2.7126	10.133

Table 3-1. NMR properties for relevant nuclei.

However, the ^{15}N isotope suffers from super-low natural abundance and a small gyromagnetic ratio (**Table 3-1**).[212] Methods for overcoming both ^{15}N and ^{13}C signal

insensitivities incorporate a suite of two-dimensional NMR experiments where the sensitivity of the proton is transferred to the more insensitive nuclei increasing the signal-to-noise (S/N) ratio and permitting data analysis of lower abundance metabolites.[212,213] A good example of this is the difference in sensitivity between the ^{13}C signal in ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC spectra and one-dimensional ^{13}C carbon spectra. In two-dimensional experiments, carbon signals can be collected (given enough protons in the molecule) through inverse detection even when material limitations prevent the collection of high quality one-dimensional ^{13}C spectra. In general, there are two phenomena that are exploited to transfer sensitivity from the proton to insensitive nuclei like ^{13}C and ^{15}N: one is the nuclear Overhauser effect, and the other is polarization transfer for scalar coupled nuclei.[214,215] The base experiment for non-selective sensitivity enhancement for scalar coupled nuclei is called the Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) experiment.[214] The maximum sensitivity transferable by the nuclear Overhauser effect is determined by the gyromagnetic ratio of a neighboring nuclei as a dividend of twice the gyromagnetic ratio of the observed nuclei. Due to this relationship, the nuclear Overhauser effect is not of value for sensitivity enhancement of the ^{15}N nuclei with its negative gyromagnetic ratio.[215] However, polarization transfer in scalar coupled nuclei is independent of the sign of the gyromagnetic ratio and, as such, is very useful for ^{15}N experiments.[214] The one caveat is that since this polarization transfer relies on scalar coupling there must be through-bond connectivity for this enhancement to occur. Because of the larger differences in the gyromagnetic ratio between ^{1}H and ^{15}N and thus, a larger difference in populations of the ^{1}H low energy and high energy spin states, the inverse detection of ^{15}N in these 2D experiments, where polarization is transferred from ^{1}H to ^{15}N, results in increased detection sensitivity for ^{15}N.[215] With experiments built off the INEPT foundation,[214] ^{15}N is routinely used in NMR methods for elucidating amino

acids for protein structure determination.[216,217] The low natural abundance can be circumvented by incorporating stable isotope (or labelled) precursors into natural products to increase the ^{15}N signal.[218–221] The use of ^{15}N in small molecule structure determination has had varied use in the literature.[212,213] One barrier to its use has been the tendency to report nitrogen referenced to various internal and external references,[211,222] which has made it difficult to look to the literature for typical chemical shifts that are widely applicable. Fortunately, over the last several decades, there has been a shift to reference ^{15}N NMR spectra to either nitromethane (IUPAC recommendation) or liquid ammonia set equal to zero.[212] Additionally, experimental work has been done to compare the shifts of nitromethane and liquid ammonia, meaning chemical shift values can be easily converted between these two ^{15}N references.[208,211,212,223]

One advantage of nitrogen NMR is the large spectral range (~900 ppm) in different chemical environments such that different nitrogen-containing functional groups will have widely dispersed ^{15}N chemical shifts.[224] Within both proteins and secondary metabolites, there are also usually fewer nitrogen atoms (compared to protons or carbons), which means that the spectra are not as crowded as proton and carbon spectra. Specific experiments have been designed to make use of this larger spectral dispersion to provide a means of separating crowded proton and carbon signals. One such experiment is the ^{1}H-^{15}N-HSQC-TOCSY.[224] This experiment allows observations of spin systems, i.e. the set of protons that are all scalar coupled to an adjacent proton (**Figure 3.1**), connected through a nitrogen's proton. For clarity, this experiment provides TOCSY transfer from the $^{1}H^{N}$ spin to the $^{1}H^{\alpha}$ spin and on through adjacent J-coupled protons. Because the TOCSY transfer begins with the $^{1}H^{N}$ spin, the trace of the spin system begins with the ^{15}N resonance at which the $^{1}H^{N}$ is attached. ^{1}H-^{15}N HSQC experiments can be used to link a primary or secondary nitrogen with its attached proton, which due to the

large dispersion of nitrogen signals, allows observation of spin systems outside the crowded proton spectral window.[224] The TOCSY portion of the ^{1}H-^{15}N HSQC-TOCSY then allows observation of the magnetization transfer through scalar-coupled protons in order to see a complete amino acid ^{1}H spin systems in peptide natural products (**Figure 3.1**).
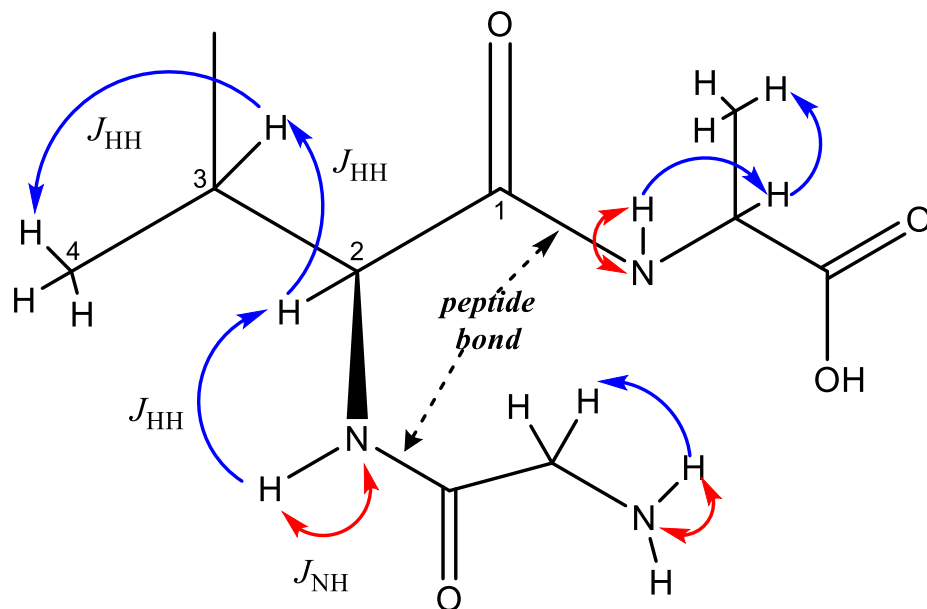
Figure 3-1. ^{1}H-^{15}N HSQC-TOCSY is an NMR experiment that allows us to observe the proton spin system of a single amino acid embedded in a peptide, beginning with the N-H correlation. The carbonyl of the peptide bond provides an insulator functionality that constrains observation to single amino acid spin systems in the TOCSY spectrum.

3.3 Method Development: ^{15}N NMR-Guided Fractionation of Piz-containing NRPs

3.3.1 Piperazic NMR for Targeted Isolation of Piz Natural Products

I was able to use genomic mining to predict which bacterial strains could be capable of producing novel Piz-containing NRPs; however, a robust method for isolating the metabolites was required. While liquid chromatography-mass spectrometry (LC-MS) and liquid chromatography-tandem-mass spectrometry (LC-MS/MS) hybrid systems are commonly coupled with genome mining, depsipeptides (a common class of Piz-containing natural products)

usually have unexpected fragmentation patterns. Furthermore, Piz-containing peptides can often

have the same mass but different structures (**Figure 2-21**), and I lacked high-resolution LC-

MS/MS capacities in the department when the project first began. Finally, NMR is required for

the final structure elucidation of complex peptides. Thus, LC-MS did not seem the best tool for

analysis of bacterial extracts for novel Piz-containing natural products. Considering this, I noted

that proteinogenic amino acids have characteristic NMR spin systems visible in ^{1}H-^{1}H TOCSY

NMR experiments, which are routinely used for protein structure elucidation.[224] Although

piperazic acid is a nonribosomal amino acid, I hypothesized that it, like proteogenic amino acids,

would likely have a characteristic spin system that was distinguishable in such experiments like

^{1}H-^{1}H TOCSY or ^{1}H-^{13}C HSQC-TOCSY.

To substantiate the hypothesis that Piz would have a characteristic spin system, I

analyzed experimental data on Piz-containing natural products reported in the literature (**Table

3-2**) and observed that Piz-containing natural products have predictable and unique NMR proton

spin systems. Furthermore, in several reported literature values, the shielded secondary nitrogen
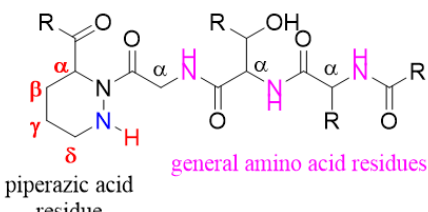
Piz Diagnostic NMR Feature		General Amino Acid	Screening Experiment
	δ (in DMSO-d_6)	δ (in DMSO-d_6)	
N-H	-297-300 ppm	-250 - 270 ppm	^{15}N-HSQC
N-**H**	4.1 – 5.2 ppm	6.7 - 8.7 ppm	^{15}N-HSQC
δ	2.5 - 3.0 ppm	varies	^{15}N-HSQC-TOCSY
γ	1.3 – 1.6 ppm	varies	^{15}N-HSQC-TOCSY
β	1.8 – 2.0 ppm	varies	^{15}N-HSQC-TOCSY
α	4.9 – 5.7 ppm	3.5-5.0 ppm	^{15}N-HSQC-TOCSY

Table 3-2. Piperazic acid has unique NMR signatures for its nitrogen, nitrogen-attached proton, and complete amino acid spin system.

experimentally shown to possess a chemical shift of ~ δ_N −300 (referenced to CH_3NO_2) in a

unique area of the nitrogen spectral field (**Table 3.3**). Given the partial negative charge predicted

to reside on the N-6 of piperazic acid, this increased shielding due to increased electron density

is predictable.[225] Importantly, this unique nitrogen shift is coupled to a distinctly shielded proton compared to amide protons in peptide-based natural products. This nitrogen-attached proton has a chemical shift shown experimentally to fall between δ_H [4.1 to 5.47] in DMSO-d_6 (**Table 3.2 and 3.3**), with even greater shielding in other deuterated solvents such as CDCl$_3$ and ACN-d_3. This shift contrasts with the normal NMR chemical shift range of amide protons seen as predictable markers of peptides, which tend to fall between δ_H [6.7 to 8.7] in DMSO-d_6.[211] This analysis suggested that the ^{1}H-^{15}N HSQC NMR experiment could potentially identify these compounds from within a crude mixture of bacterial metabolites that commonly also contain amino acids, simple peptides and diketopiperazines. The potential of observing the presence or absence of piperazic acids from within a crude bacterial extract would allow for: 1) selecting laboratory conditions in which the genome-mined bacteria were producing the Piz-containing metabolite, and 2) directing a pathway for the isolation workflow to streamline isolation work. Thus, if the N-H signal was visible in the crude extract, the ^{15}N NMR-guided fractionation workflow could mimic the bioassay-guided fractionation workflow. Bioassay-guided fractionation has been the gold standard for decades because it allows the isolation chemist to focus on both the organisms and fractions of interest, leading (ideally) to bioactive compounds (**Figures 1.8 and 1.9**).[226] Using nitrogen NMR experiments to select for the presence of piperazic-acid compounds would not lead directly to bioactive compounds; however, piperazic acid-containing compounds are overwhelmingly associated with bioactivity[173] and often with structural novelty.

Natural Product	Type of Piz	α	β	γ	δ	N-N**H**	N-N**H**	N-N**H**	
		δ_H (δ_C)	δ_H	δ_H	δ_H (δ_C)	δ_H	δ_N	δ_N	
dentigerumycin C*	L-Piz	4.55 (52.6)	1.8	1.3	2.7 (46.4)	5.47			
	D-Piz	5.58 (47.7)	1.8	1.42	2.7 (46.8)	5.2			
	D-Piz	5.44 (47.7)	1.8	1.43	2.7	5.18			
dentigerumycin D*	L-Piz	4.84 (50)	1.8	1.48	2.7 (46.4)	5			
	D-Piz	5.58 (47.8)	1.8	1.48	2.7 (46.4)	5.04			
	D-Piz	5.44 (47.6)	1.8	1.53	2.89 (45.6)	5.11			
matlystatin*	L-Piz	5.00 (49.7)	1.8	1.36	2.7(46.3)	4.74			
Sch 382582*	L-Piz	5.12 (49.52)	1.8	1.49	2.8 (46.38)	4.92			
hytramycins*	D-Piz	4.94(51.91)	1.8	1.5	2.8 (48.48)	4.96			
	D-Piz	5.14(49.03)	1.8	1.47	2.8 (48.64)	4.99			
	L-Piz	5.7 (48.58)	1.8	1.45	2.8	5.07			
padanamides*	L-Piz	4.97 (50.9)	1.8	1.35	2.79 (46.8)	4.5	-299.9	-232	
piperidamycin A	Ave x 3	5.81 (47.62)	1.8	1.5	3.00/2.58 (46.48)	5.2			
	γ-OH	4.81 (48.77)	2.3, 1.8	3.6	2.8 (52.38)	4.97			
sanglifehrins	peptide bond to N5	1.92 (58.0)	1.45	1.65/1.30	4.20/2.65 (40.5)	4.82	-297.5	-239.7	
soliseptide A	L-Piz	5.06 (50.4)	2.06/1.54	1.54/1,46	2.97/2.59	5.43			
	γ-OH	4.97 (46.7)	2.03/1.82	3.65	2.9/2.85 (52.6)	5.49			
svetamycins	γ-chl	5.73(49.8)	2.41/1.90	4.36	2.66/3.38 (54)	5.23			
	δ-Me	4.88 (49.7)	2.17/1.76	1.56/1.51	2.69 (52.3)	4.14			
	δ,δ-dMe	4.88 (49.7)	2.07/1.96	1.42/1.34	2.69 (50.9)	4.57			
himastatin	γ-OH	5.08	2.44/1.92	3.77	3.02/2.79	5.37			
*α, β, and γ positions are averages of the diastereotopic δ_H									

Table 3-3. Experimental NMR chemical shifts in DMSO-d_6 for Piz within natural products as obtained from literature sources.[227]

3.3.2 Analysis of the Piperazic Acid NMR Signature in a Known Producer

To assess the viability of a ^{15}N NMR-based method for screening for piperazic acid-containing peptides, I first evaluated the crude metabolite mixture of *Streptomyces sp.* RJA2928. This species was previously shown to produce the Piz-containing natural product padanamide[206] (**Figure 3-2**) and had a known gene cluster containing a *ktzT* homolog[200]. Initially, this species was grown on 2 L of agar medium to provide 600 milligrams (mg) of a crude metabolite mixture (hereby referred to as "crude") from which an aliquot was evaluated for the Piz NMR N-H correlation using the ^{1}H-^{15}N HSQC experiment as well as the Piz spin system using ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC-TOCSY NMR experiments. The spectra obtained from ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC-TOCSY proved sufficiently complex to prevent unequivocal confirmation of the production of padanamides within the crude mixtures, while the ^{1}H-^{15}N HSQC experiment did not show the Piz N-H correlation. As I was uncertain if this lack of signal was simply because the N-H correlation was not visible in the crude mixture, fractionation of the

Padanamide's Piz NMR Chemical Shifts	
	δ (in DMSO-d_6)
N-H	-299.8 ppm
N-H	4.52 ppm
δ	2.79 ppm
γ	1.35 ppm
β	1.63/2.08 ppm
α	4.97 ppm
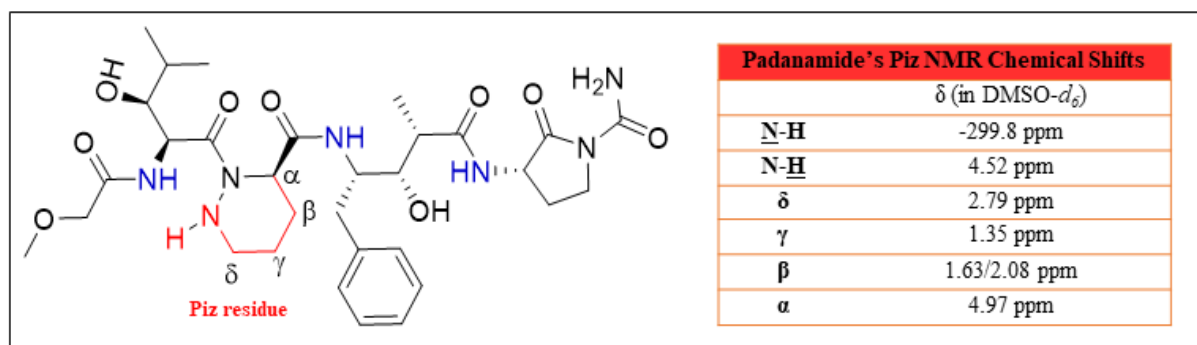
Figure 3-2. Padanamide A's structure and Piz NMR chemical shifts.

crude was performed. *Streptomyces* sp. RJA 2928's crude metabolite mixture was passed through a Sephadex^{TM} LH-20 column to create three fractions which were each evaluated using ^{1}H-^{15}N HSQC experiment, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC-TOCSY NMR experiments. Analysis of the three spectra revealed no change from the crude mixture, and it seemed apparent

that the nitrogen signal likely needed to be enhanced with labelled $^{15}N_2$-L-Orn. I also considered that the frozen *Streptomyces* sp. RJA2928 used for inoculation and growth was a late-stage generation that might have degenerated into being a non-producer of padanamides. Unpublished work in the Ryan and Andersen labs had demonstrated that padanamides could change through colony generations. To resolve these questions and to investigate the detection limit of the experiment, first-generation *Streptomyces* sp. RJA2928 was revived from the -80 °C cryostorage and grown in small quantities both with labelled $^{15}N_2$-L-Orn (**L**) and with unlabelled L-Orn (**UL**) (**Figure 3-3**).

Streptomyces sp. RJA2928	Media	Volume	Additives	Crude EtOAc Extract
L	MM1	30 mL	20 mg $^{15}N_2$-L-Orn	11.5 mg
UL	MM1	30 mL	20 mg L-Orn	10.5 mg

Figure 3-3. Prepared Petri plates of known Piz producer *Streptomyces* sp. RJA2928.

After extraction, the crude's **L** and **UL** were evaluated for Piz NMR signals (**Table 3.4, Figure 3.4**). A signal in the region of ~ δ_N -300 ppm and δ_H 4.52 ppm was visible in both samples as seen in **Figure 3-4**, but it was difficult (to impossible) to distinguish from noise in the unlabelled sample. These results showed that if the producer made the Piz-containing natural product, the correlation was visible, albeit very weakly, even in small amounts of material and with natural abundance ^{15}N. However, adding labelled $^{15}N_2$-L-Orn to the medium provided strongly enhanced S/N and diagnostic N-H Piz NMR signature observation. The ^1H-^{15}N HSQC-TOCSY NMR experiment was then performed to access the diagnostic Piz amino acid spin system. However, it is an insensitive experiment compared to the ^1H-^{15}N HSQC experiment, requiring both increased compound concentration and increased time (up to 7-9 hours vs 48 mins), and attempts to collect this data on samples **L** and **UL** failed.
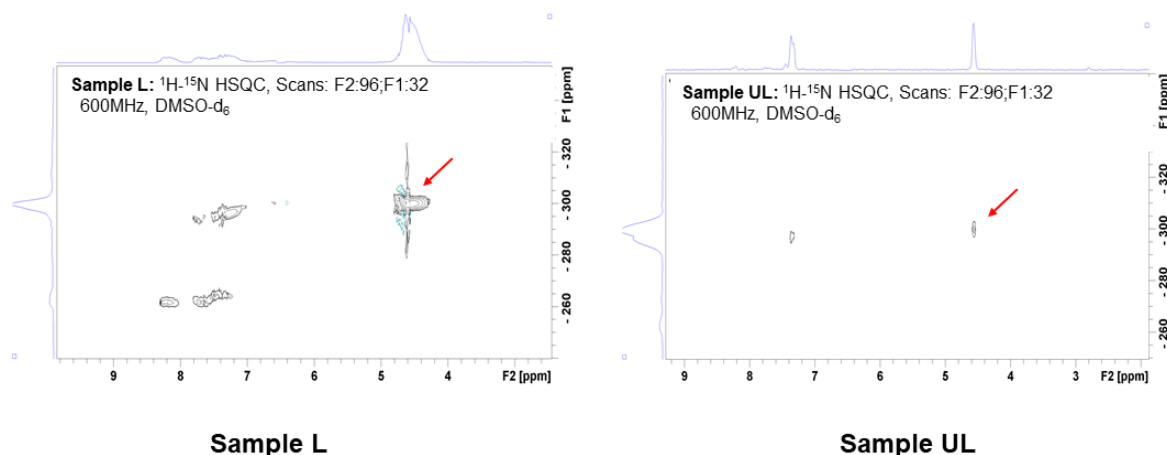
Figure 3-4. ^{1}H-^{15}N HSQC NMR for Samples L and UL of known Piz producer *Streptomyces* sp. RJA2928.

In order to obtain more material to verify that the Piz spin system from padanamides was observable in the crude, the quantity of bacteria grown was increased. A tray (37 cm X 20 cm) of solid MM1 agar was prepared with 100 mg ^{15}N$_2$-L-Orn added to 400 mL of MM1 marine media. After growth and extraction, a small aliquot of this sample **ScL**, (~20 mg) was first subjected to the ^{1}H-^{15}N HSQC experiment followed by the ^{1}H-^{15}N HSQC-TOCSY experiment. The crude obtained from this experiment using medium containing ^{15}N$_2$-L-Orn gave strong enough NMR signals to allow observation of both the N-H correlation and the Piz spin system (**Figure 3.5**) in the ^{1}H-^{15}N HSQC-TOCSY NMR experiment. The reduction of NMR acquisition time in the ^{1}H-^{15}N HSQC was also a valuable benefit, as shown in **Table 3-4**.

Sample	Agar Medium	Amount of Media	L-Orn (mg)	^{15}N$_2$-L-Orn (mg)	^{1}H-^{15}N HSQC	NMR Time for ^{15}N experiment	^{1}H-^{15}N HSQC-TOCSY
L	MM1	30 mL	0	20	Yes	96 min	Not Observable
UL	MM1	30 mL	20	0	Very weak	96 min	Not Observable
ScL	MM1	400 mL	0	100	Yes	48 min*	Yes

Table 3-4. Analysis of Piz production in a known producer. *Less intense signal available in ~15-20min.
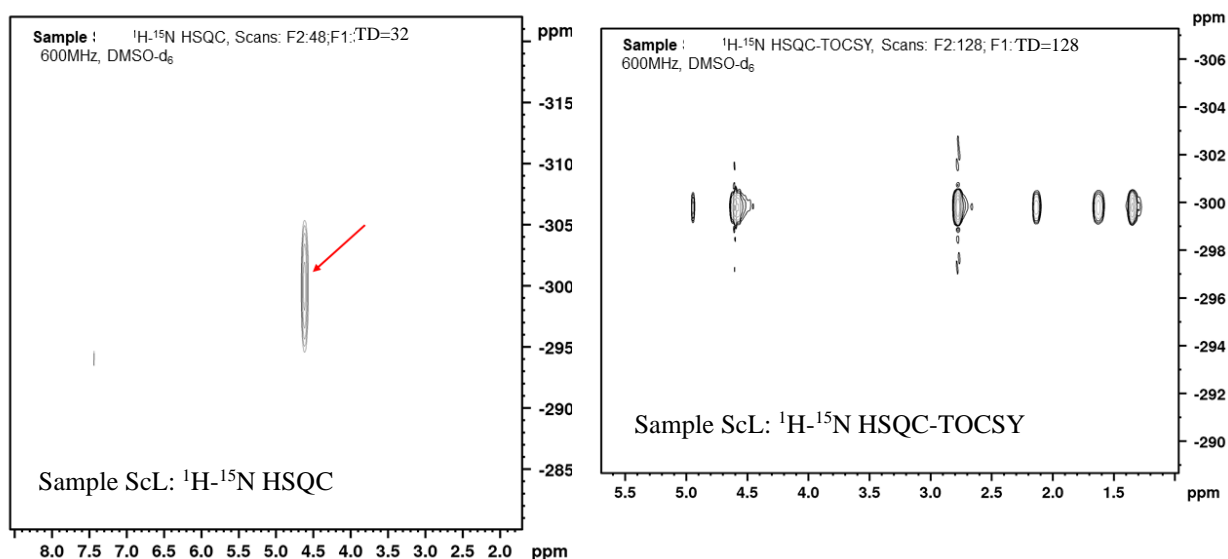
Figure 3-5. NMR spectra of crude metabolite extract of *Streptomyces* sp. RJA2928 grown with labelled $^{15}N_2$-L-Orn. a) ^{1}H-^{15}N HSQC spectrum demonstrating a Piz N-H correlation b) ^{1}H-^{15}N HSQC-TOCSY spectrum demonstrating padanamides' Piz spin system.

3.3.3 Strain and Media Selection Through Diagnostic NMR Resonances

Initially, several strains predicted to have the genomic capacity to produce Piz-containing natural products, as seen in **Table 3-5**, were ordered from the Agricultural Research Service (NRRL) Culture Collection. NRRL collection strains were specifically chosen as they are: 1) characterized Biosafety Level-One non-pathogenic bacteria that create minimal hazards;[228] 2) readily available to any research program; and 3) free-of-charge if under a certain number were ordered. Next, it was considered that secondary metabolite production varies depending on environmental conditions.[229] In the laboratory, this results in a variance of secondary metabolites produced when growth media varies, so there is no certainty that genotype will result in the desired metabolite phenotype. Therefore, several growth media were selected for each NRRL strain (**Table 3.5**). The four strains listed in **Table 3.5** were obtained from the Agricultural Research Service Culture Collection, revived and grown on large Petri plates with added labelled precursor, harvested, and extracted to obtain 15-30 mg of crude material. ^{1}H-^{15}N HSQC spectra

were obtained for each crude sample. Several species were selected based on these results for scale-up.

Strain	L-Orn	Appearance of Bacteria on Agar Media	
		MM1 F	MS
Streptomyces phaeochromogenes NRRL B-1248	Labelled	white spores on brown stain	white spores, yellow stain
Streptomyces griseus subsp. griseus strain NRRL WC-3645	Labelled	white spores, leathery	pinkish white, leathery
Streptomyces griseus NRRL F-5144	Labelled	yellow spores on deep yellow stain	yellow/green spores w. lemon yellow stain
Streptomyces sp. NRRL F-2664	Labelled	pink spores, dark brown stain	white spores, golden yellow stain

Table 3-5. Initial genome mined strains with selected media.

Based on the spectrum seen in **Figure 3-6**, the strain *Streptomyces phaeochromogenes* NRRL B-1248 was grown on a tray of MM1 F agar with 100 mg of $^{15}N_2$-L-Orn added to the media. After growth and extraction, it was fractionated utilizing the Piz N-H correlation at δ_H 5.55; δ_N −304 as the guide. At the end of the isolation workflow, the N-H signal was present, but the compound was not present in amounts that allowed collection of a 2-D data set towards structure elucidation. From this result, it was understood that although a strong signal from labelled material could act as a guide in the isolation process, it would not guarantee that there was enough material for isolation and structure elucidation. The decision was made to focus on other strains rather than scale up this strain.

Subsequently, *Streptomyces griseus* subsp. griseus strain NRRL WC-3645 (**Figure 3-8**) was grown with a small amount of labelled material and a larger amount of unlabelled material. The extracts obtained from both media were combined together and separated using the N-H correlation, δ_H 5.88; δ_N −309.2, as a guide without the use of the ^1H-^{15}N HSQC-TOCSY experiment. This N-H correlation was slightly outside the boundaries observed from examining literature values in **Table 3-2** yet was close enough to these values to investigate if this N-H

correlation belonged to a Piz-containing compound. Upon separation through bench-top

chromatographic steps and purification on reversed-phase HPLC, the compound possessing this

signal was determined to be para-aminobenzoic acid (PABA), a common primary metabolite.

This result demonstrated that 1) the secondary piece of data provided by ^{1}H-^{15}N HSQC-TOCSY

experiment was invaluable for increasing isolation efficiency; and 2) combining labelled and

unlabelled material allowed a nitrogen-NMR guided fractionation to isolate at least a primary

metabolite.

Strain	Nitrogen NMR		
	Media Extract	$\delta_H. \delta_N$	
Streptomyces phaeochromogenes NRRL B-1248	MM1 F	5.55 ; -304	**Figure 3-6**
Streptomyces griseus subsp. griseus strain NRRL WC-3645	MM1 F	5.88; -309.2	**Figure 3-7**
		4.17, -299	
	MS	5.88; -309.2	**Figure 3-8**
Streptomyces griseus strain NRRL F-5144	ISP4	5.49, -302 (v weak)	**Figure 3-10**
		5.12, -300 (v weak)	
	TSB	5.45, -302	**Figure 3-9**
Streptomyces sp. NRRL F-2664	MS	5.41, -302	**Figure 3-11**

Table 3-6. Initial genome mined strains with initial nitrogen NMR experiments per media grown.

Of these strains, initial attempts to grow *Streptomyces griseus* strain NRRL F-5144 did

not result in observations of the N-H Piz correlation. *Streptomyces griseus* strain NRRL F-5144

was grown on several more media including ISP4 and TSB (**Figure 3-9 and 3-10**). The

bacterium was cultured in small amounts but this did not provide enough material for isolation of

the suspected polyoxypeptin congener. Larger amounts of this bacterium would need to be

grown for successful isolation efforts.
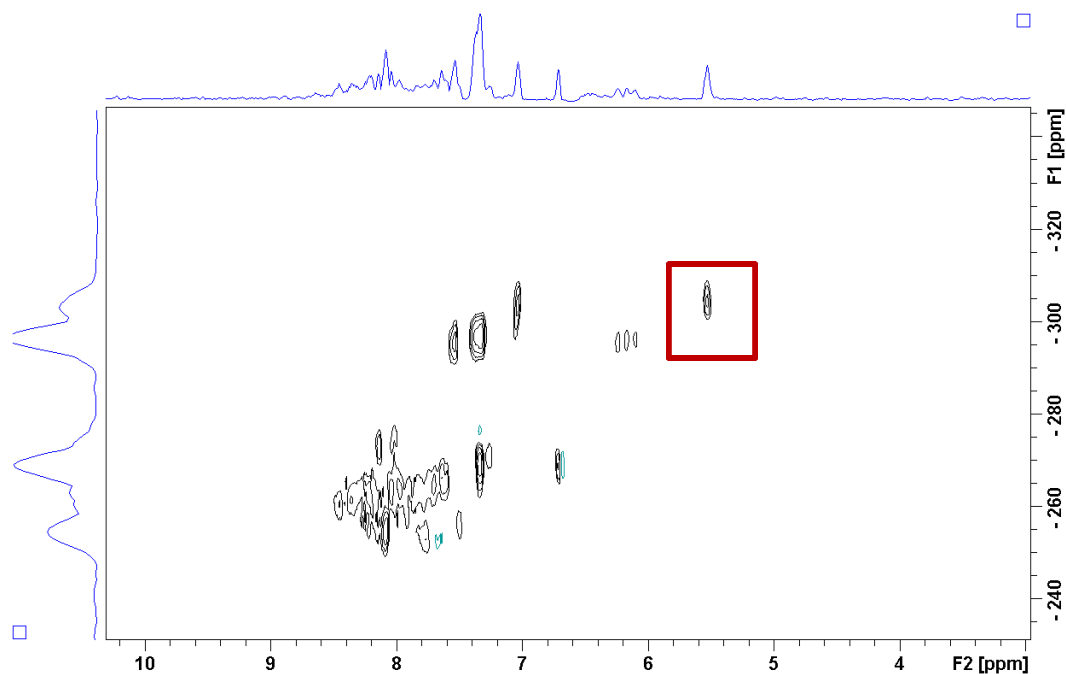
Figure 3-6. ^{1}H-^{15}N HSQC of *Streptomyces phaeochromogenes* **NRRL B-1248 extract from MM1 F agar, recorded at 600 MHz, in DMSO-d_6. Red box indicates presence of Piz-like N-H correlations.**
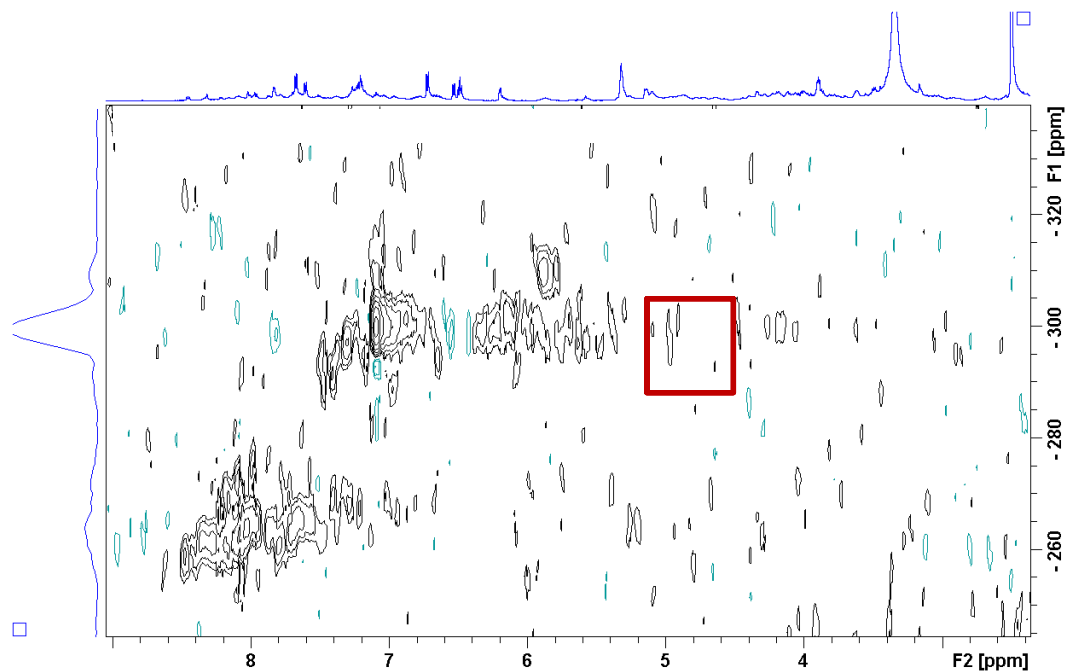
Figure 3-7. ^{1}H-^{15}N HSQC of *Streptomyces griseus* **subsp. griseus NRRL WC-3645 extract from MM1 F agar, recorded at 600 MHz, in DMSO-d_6. Red box indicates presence of Piz-like N-H correlations.**
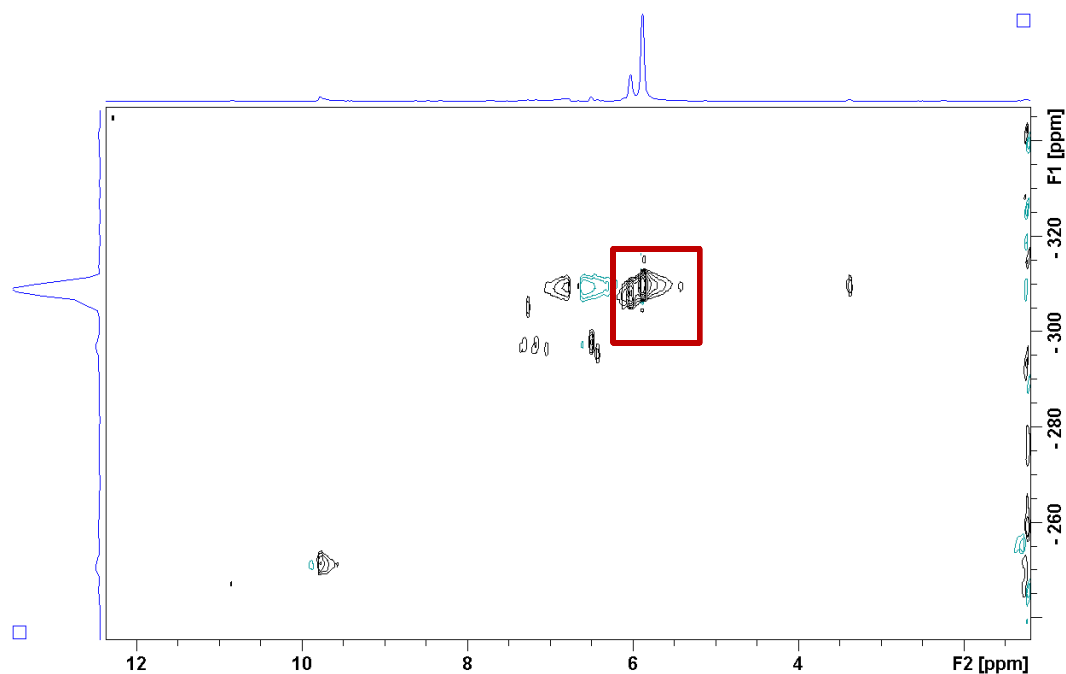
Figure 3-8. ¹H-¹⁵N HSQC of *Streptomyces griseus* subsp. griseus NRRL WC-3645 extract from Mannitol Soy agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.
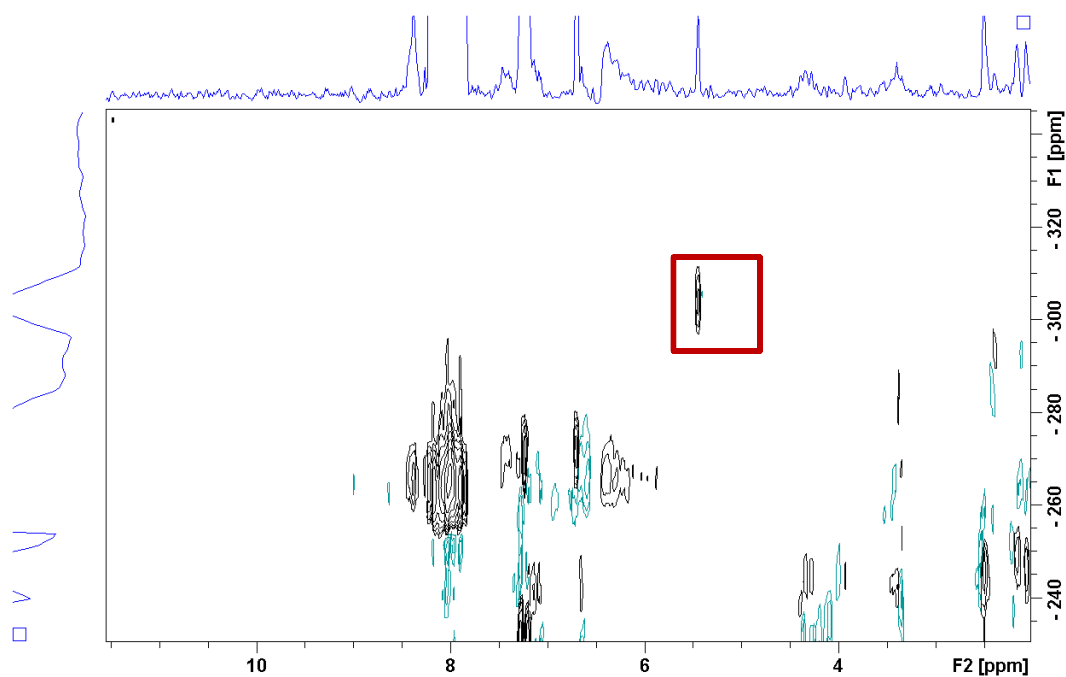
Figure 3-9. ¹H-¹⁵N HSQC of *Streptomyces griseus* NRRL F-5144 extract from TSB agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.
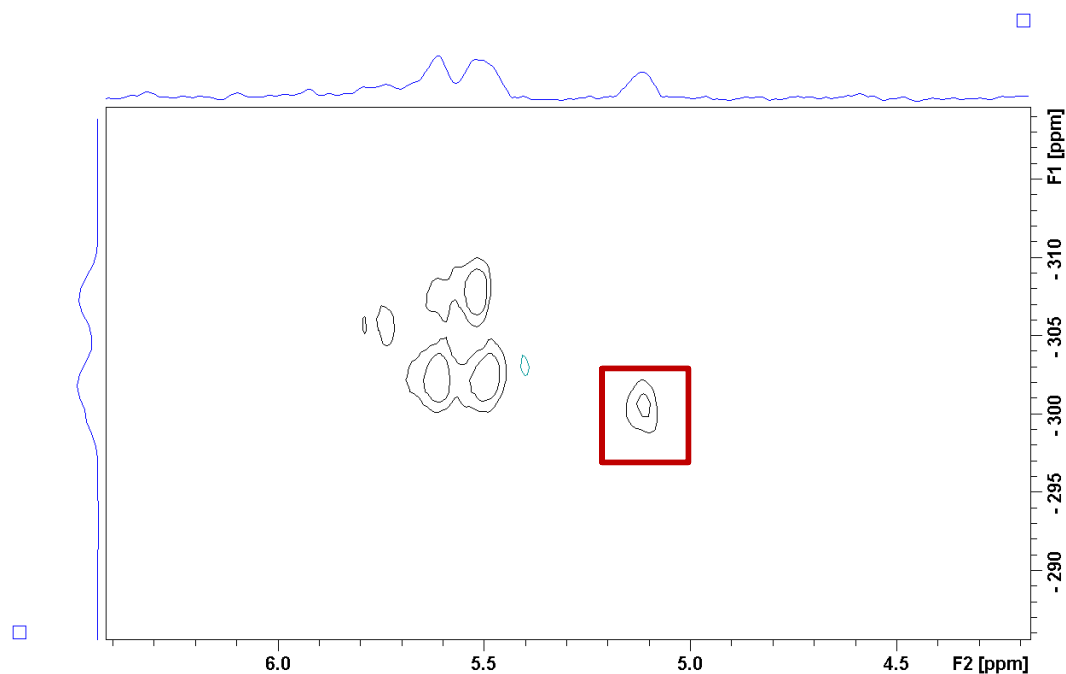
Figure 3-10. ¹H-¹⁵N HSQC of *Streptomyces griseus* **NRRL F-5144 extract from ISP4 agar, recorded at 600 MHz, in DMSO-***d₆***. Red box indicates presence of Piz-like N-H correlations.**
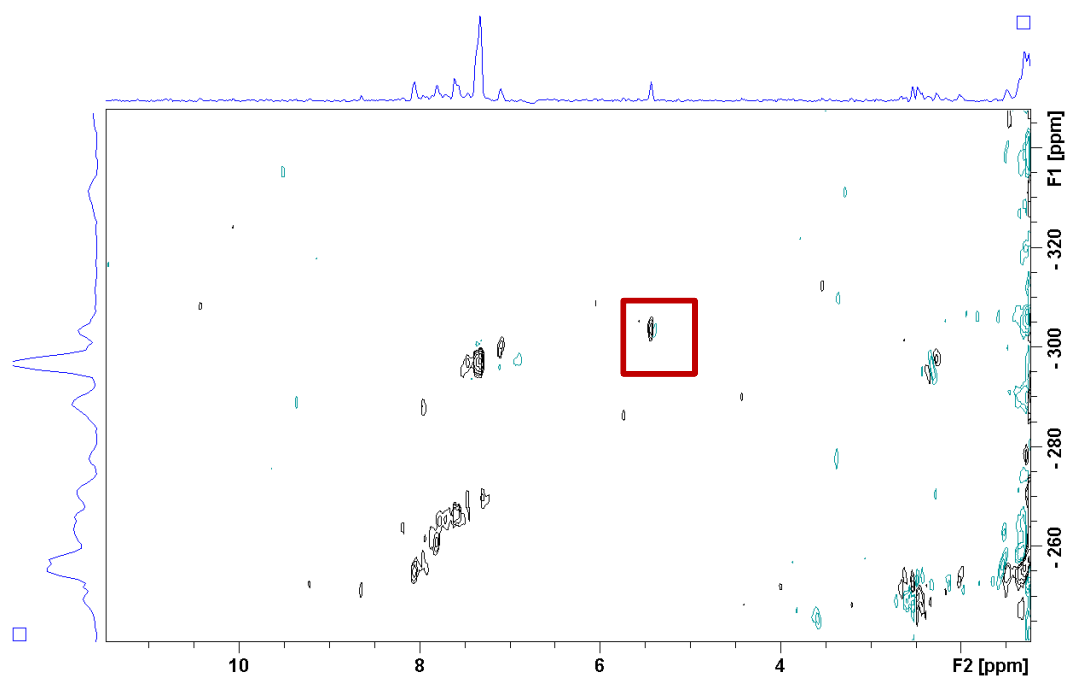
Figure 3-11. ¹H-¹⁵N HSQC of *Streptomyces* **sp. NRRL F-2664 extract from Mannitol Soy agar, recorded at 600 MHz, in DMSO-***d₆***. Red box indicates presence of Piz-like N-H correlations.**

Nine additional strains (**Table 3-7**) were obtained from the Agricultural Research Service Culture Collection and revived in tryptic soy broth to continue the investigation into genome-mined strains. After successful growth, each bacterium was grown on four different agar media prepared as large Petri plates. The EtOAc extracts of each medium's trial were dried under reduced pressure to yield 20-30 mg of crude material. ^{1}H-^{15}N HSQC spectra were obtained for each crude sample after dissolving in DMSO-d_6 with TD=32 scans in the F1 dimension and 48 scans in the F2 dimension (**Table 3-8**, **Figures 3-12 to 3-21**). The *in silico* prediction of KtzT homolog-containing gene clusters were reexamined for extracts in which a putative Piz N-H correlation was observed. Strains were selected for growth with labelled ^{15}N$_2$-L-Orn based on their Piz N-H NMR signal and potential novelty of the cluster.

Strain	L-Orn	Appearance of Bacteria on Agar Media			
		MM1 F	MS	ISP4	TSB
Streptomyces katrae NRRL B-16271	Unlabelled	small white pink spores	cream, few white spores	cream, some white spores, light pink stain	Off-white spores
Streptomyces tsukubensis NRRL18488	Unlabelled	Did not grow			
Streptomyces sp. NRRL F-2747	Unlabelled	large white colonies turn pink with time. Pink/brown stain	small white colonies	large pink spores	small white colonies
Streptomyces sp. NRRL F-5053	Unlabelled	cream small	cream, small	cream, white spores	lots of white spores
Streptomyces sp. NRRL F-6131	Unlabelled	mauve stain, white spores	light brown stain, no spores	pink stain, lots of white spores	orange stain, wrinkled
Streptomyces sp. NRRL B-1140	Unlabelled	Did not grow			
Streptomyces incarnatus NRRL 8089	Unlabelled	brown stain	tiny, clear	white spores	small, clear tiny clear spores
Streptomyces sp. NRRL B-1347	Unlabelled	Off-white colored, leathery	Off-white spores	green stain, yellow spores	wrinkled, cream
Streptomyces sp. S920	Unlabelled	thin white film	large white wrinkled	pinkish colonies	very few white wrinkled

Table 3-7. Additional genome-mined bacterial strains listed with the various growing media.

Strain	Nitrogen NMR		
	Media Extract	δ_H. δ_N	
Streptomyces katrae NRRL B-16271	MS	5.43, -303	**Figure 3-12**
	TSB	5.43, -303	**Figure 3-13**
Streptomyces sp. NRRL F-2747	TSB	5.48, -301.0	**Figure 3-14**
Streptomyces sp. NRRL F-5053	None	No signal	**N/A**
Streptomyces sp. NRRL F-6131	MM1 F	5.45; -301	**Figure 3-15**
	TSB	5.45; -301	**Figure 3-16**
Streptomyces incarnatus NRRL 8089	ISP4	5.43, -301	**Figure 3-17**
		5.11, -300 (v.v weak)	**Figure 3-17**
	TSB	5.43, -302	**Figure 3-18**
Streptomyces sp. NRRL B-1347	MM1 F	5.45; -303.6 (v weak)	**Figure 3-19**
	MS	5.40, -302.0	**Figure 3-20**
Streptomyces sp S920	MS	5.58, -302.6	**Figure 3-21**

Table 3-8. Remainder of the genome-mined strains and their nitrogen NMR N-H correlations.
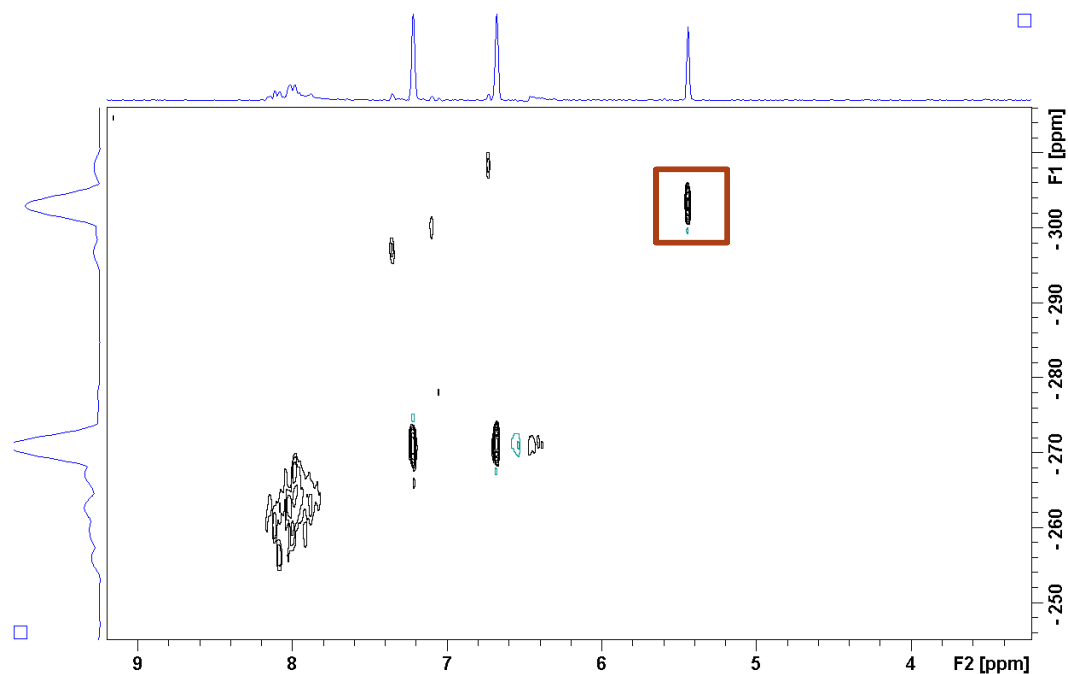
Figure 3-12. ^{1}H-^{15}N HSQC of *Streptomyces katrae* **NRRL F-16271 extract from Mannitol Soy agar, recorded at 600 MHz, in DMSO-d_6. Red box indicates presence of Piz-like N-H correlations.**
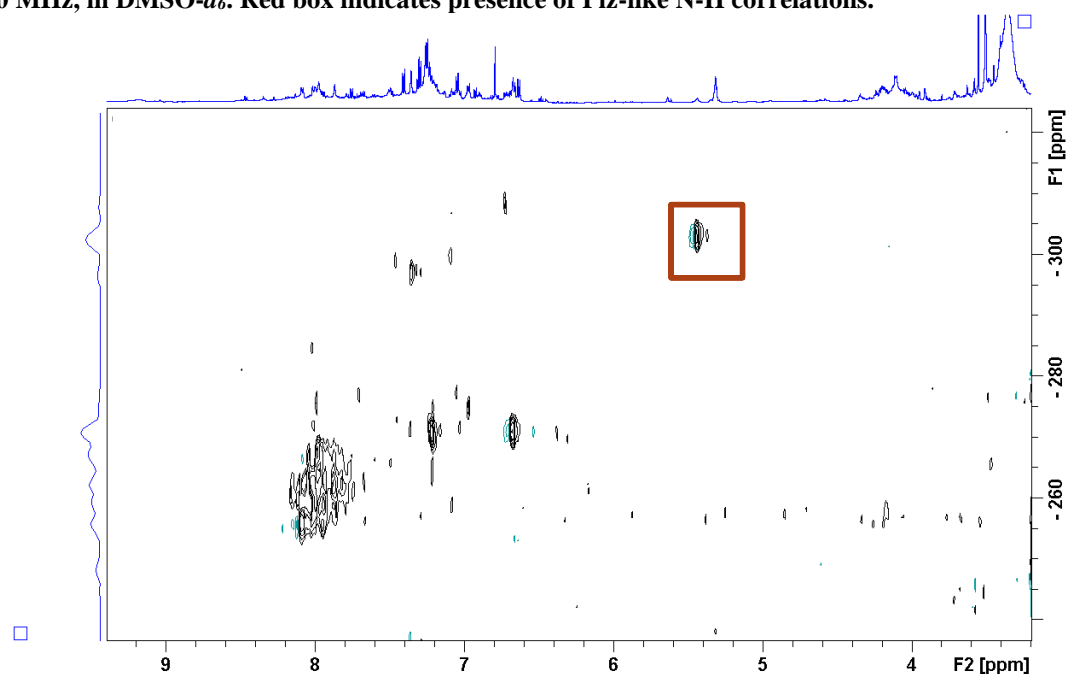
Figure 3-13. ^{1}H-^{15}N HSQC of *Streptomyces katrae* **NRRL F-16271 extract from TSB agar, recorded at 600 MHz, in DMSO-d_6. Red box indicates presence of Piz-like N-H correlations.**
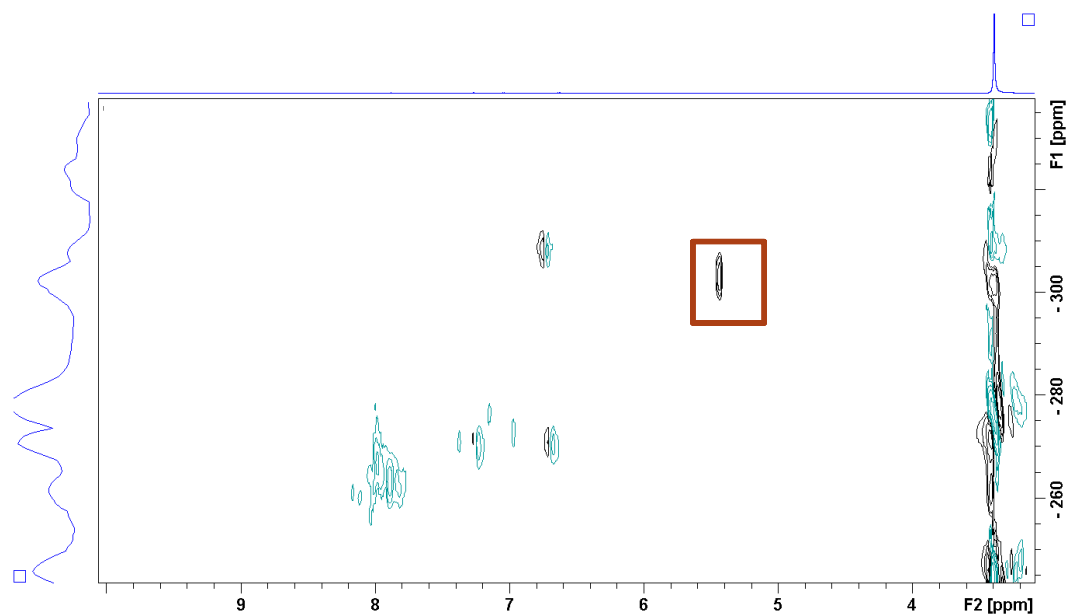
Figure 3-14. ^{1}H-^{15}N HSQC of *Streptomyces* sp. NRRL F-2747 extract from TSB agar, recorded at 600 MHz, in DMSO-*d*$_6$. Red box indicates presence of Piz-like N-H correlations.
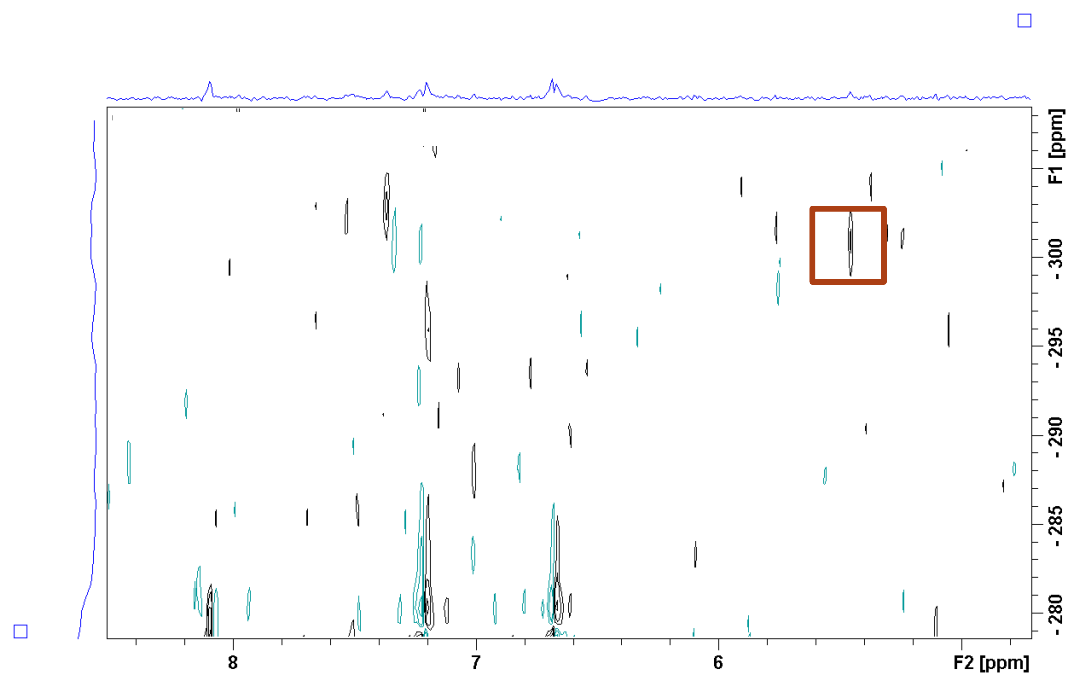
Figure 3-15. ^{1}H-^{15}N HSQC of *Streptomyces* sp. NRRL F-6131 extract from MM1 F agar, recorded at 600 MHz, in DMSO-*d*$_6$. Red box indicates presence of Piz-like N-H correlations.
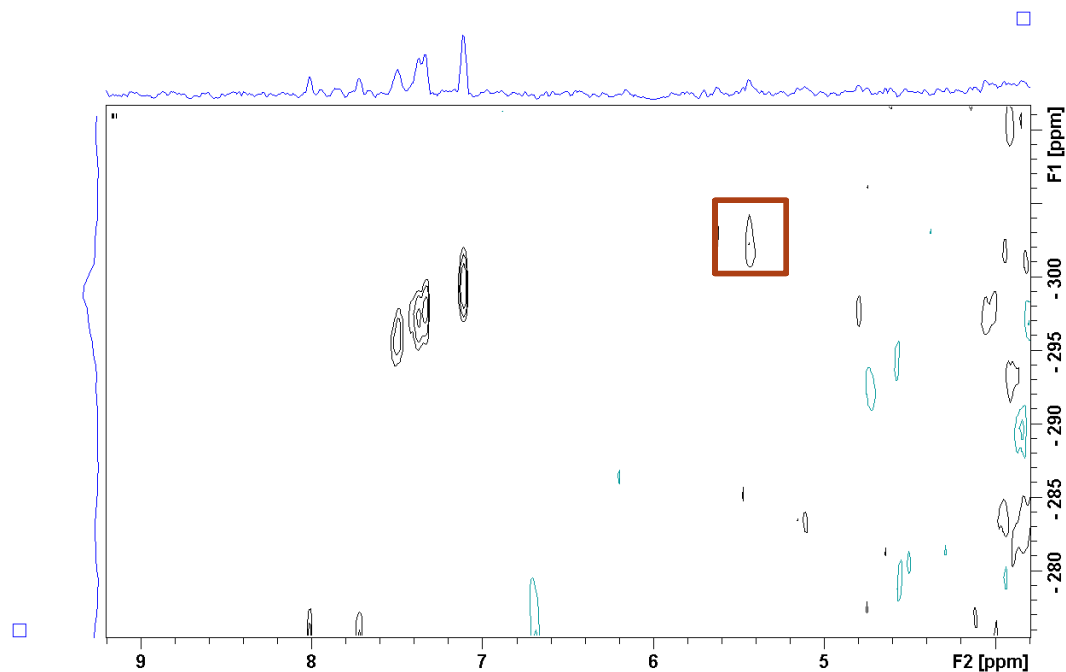
Figure 3-16. ¹H-¹⁵N HSQC of *Streptomyces* sp. NRRL F-6131 extract from TSB agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.
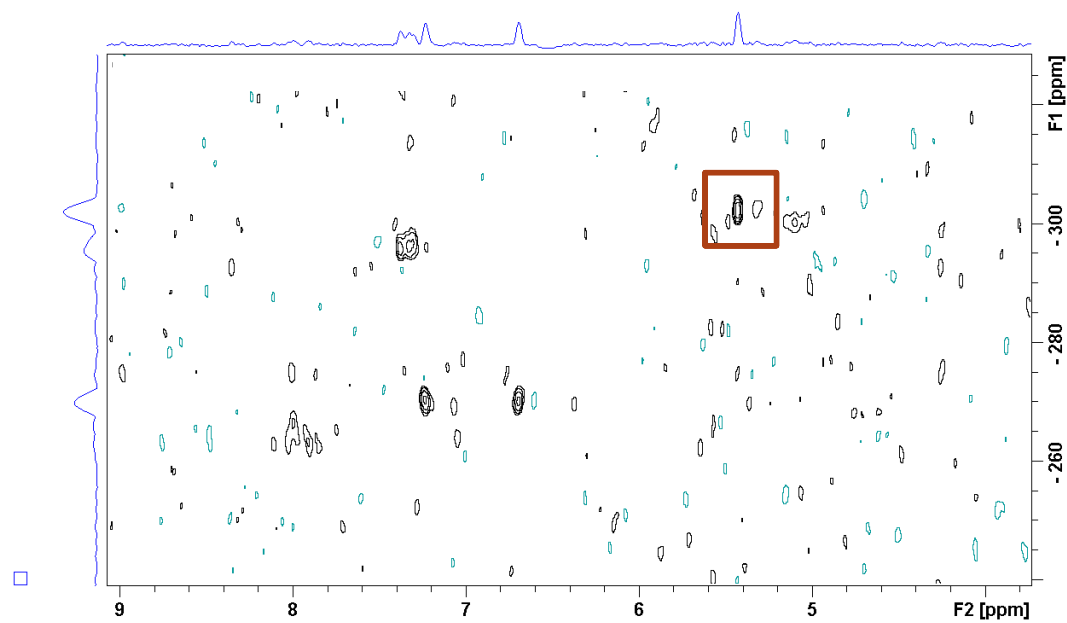
Figure 3-17. ¹H-¹⁵N HSQC of *Streptomyces incarnatus* NRRL 8089 extract from ISP4 agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.
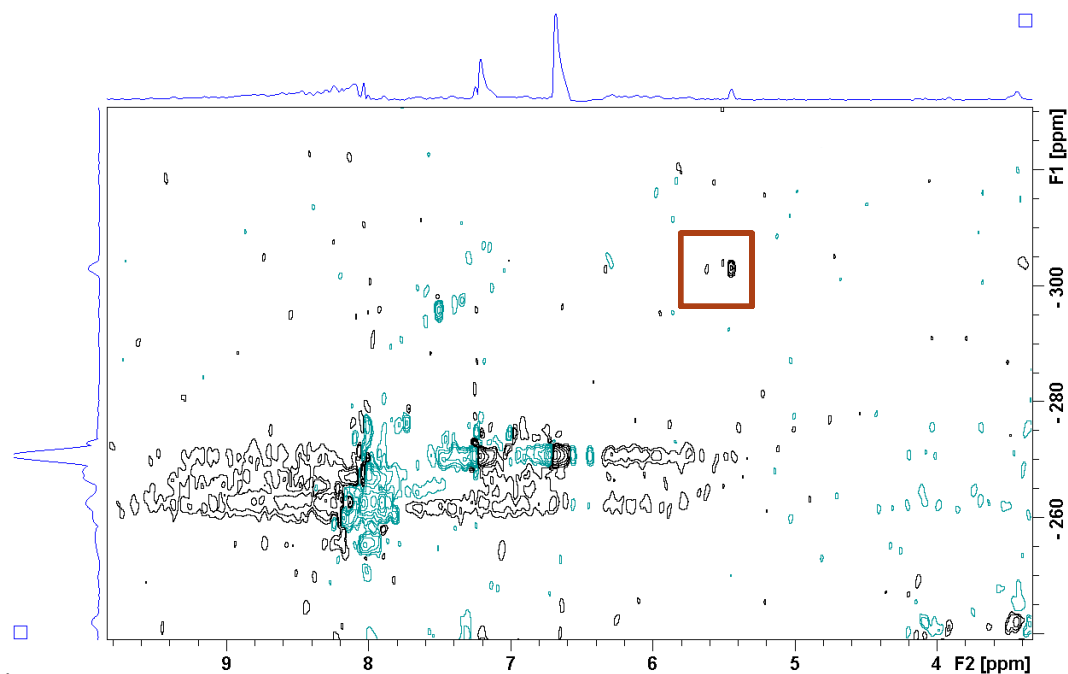
Figure 3-18. ^{1}H-^{15}N HSQC of *Streptomyces incarnatus* NRRL 8089 extract from TSB agar, recorded at 600 MHz, in DMSO-*d_6*. Red box indicates presence of Piz-like N-H correlations.
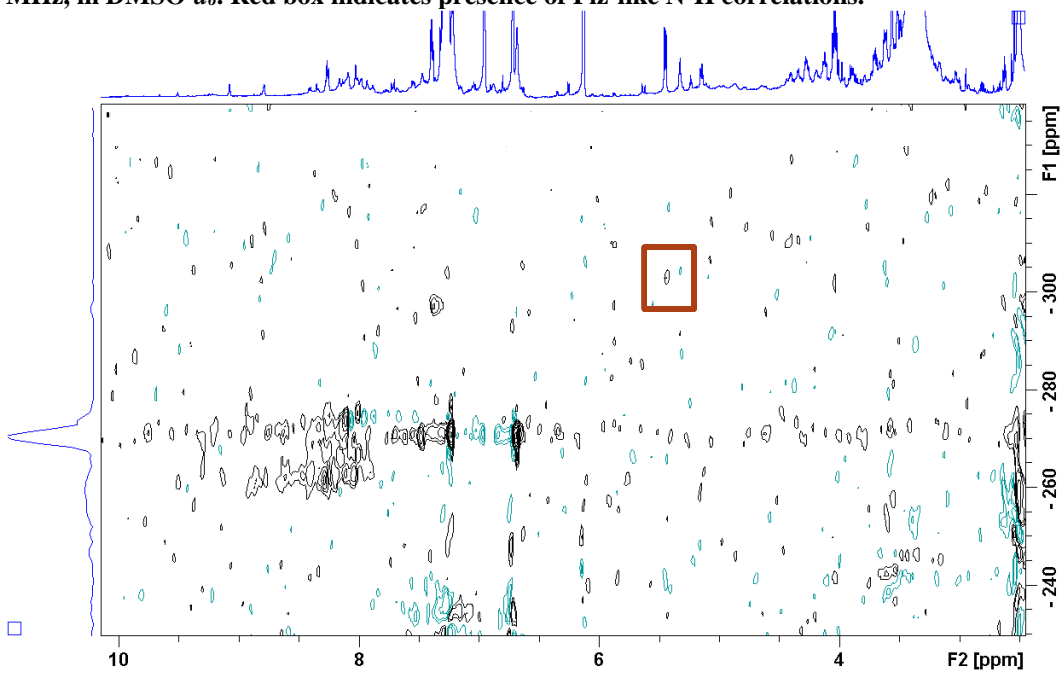
Figure 3-19. ^{1}H-^{15}N HSQC of *Streptomyces* sp. NRRL B-1347 extract from MM1 F agar, recorded at 600 MHz, in DMSO-*d_6*. Red box indicates presence of Piz-like N-H correlations.
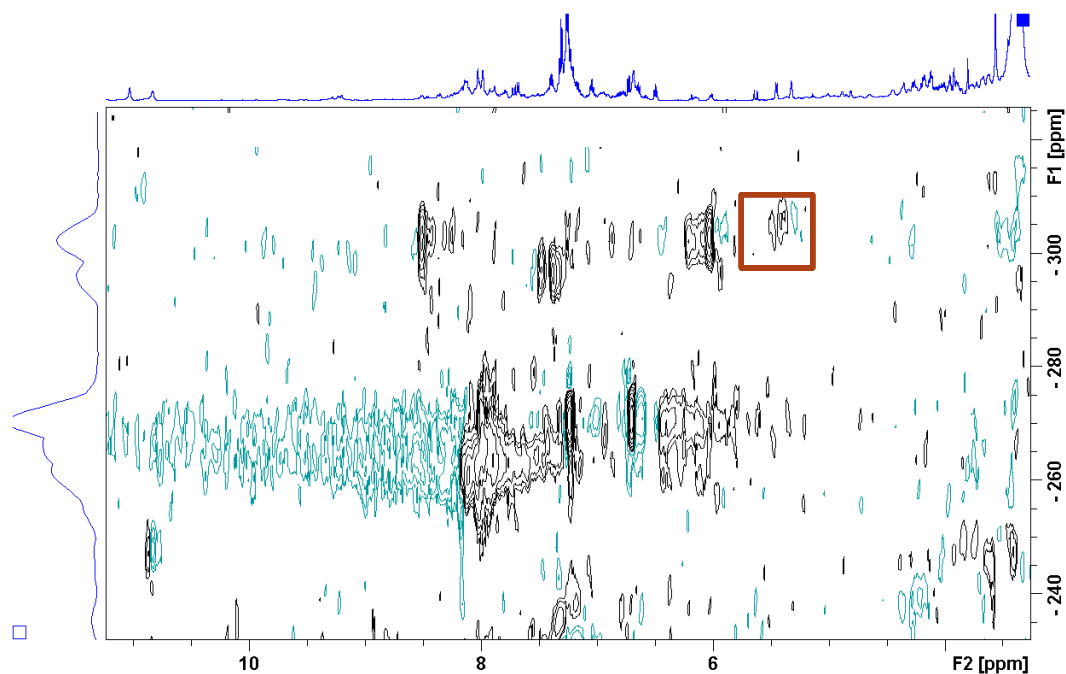
Figure 3-20. ^{1}H-^{15}N HSQC of *Streptomyces* sp. NRRL B-1347 extract from Mannitol Soy agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.
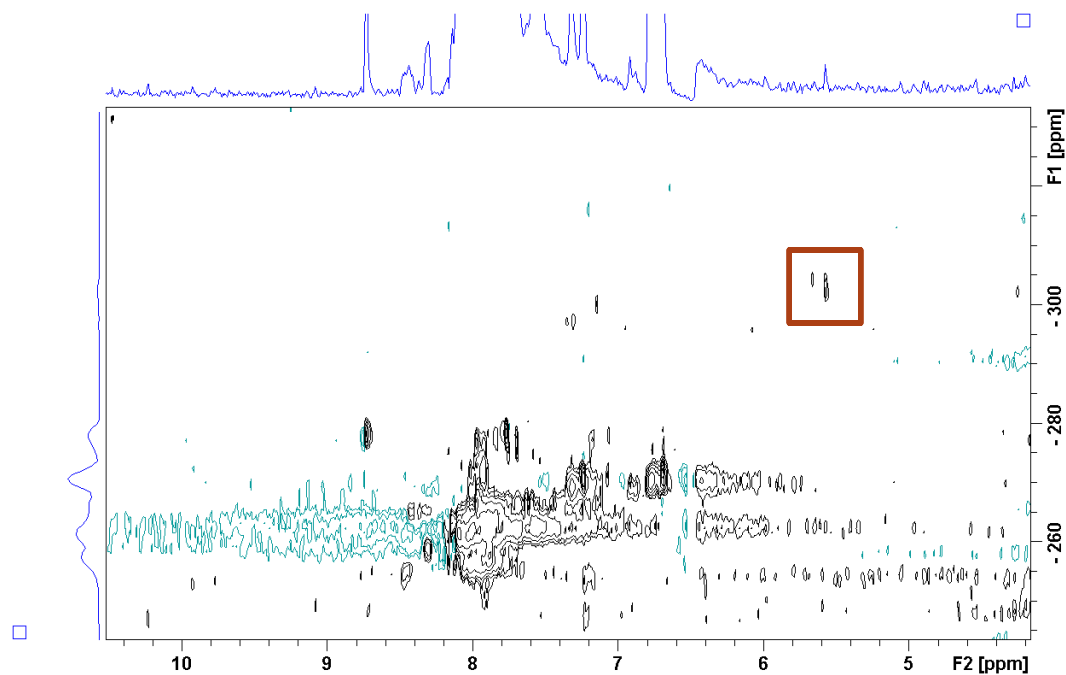
Figure 3-21. ^{1}H-^{15}N HSQC of *Streptomyces* sp. S920 extract from Mannitol Soy agar, recorded at 600 MHz, in DMSO-*d₆*. Red box indicates presence of Piz-like N-H correlations.

3.3.4 Gene Cluster Analysis of *Streptomyces incarnatus* NRRL 8089

As discussed in **Section 2.6**, BLASTP searches for KtzT homologs returned a KtzT homolog contained within a contig with Accession no. CP011498.1. This contig was assembled from a draft sequencing of the actinomycetes species *Streptomyces incarnatus* NRRL 8089, which resulted in 76 contigs containing 8,266 predicted genes for this 8.9 Mb genome.[230] *Streptomyces incarnatus* NRRL 8089 was isolated from Indian soil in 1977 and identified as a novel bacterial species.[231] Interestingly, it has been shown to produce analogs of the potent antifungal natural product sinefungin[231] as well as contain the biosynthetic machinery capable of producing an analog of the unusual natural product foxicin.[232] *Streptomyces incarnatus* NRRL 8089 appears to be 'a biosynthetically talented' species.[233] antiSMASH analysis of contig CP011498.1 revealed that the entire 177 kilobase pair (kb) contig was a hybrid NRPS-PKS gene cluster. This is an unusually large size for such a gene cluster. Biosynthetic studies have demonstrated that hybrid NRPS-PKS gene clusters are larger than single origin clusters; with an average of 71kb for hybrid, 51 kb for PKS, and 45 kb for NRPS clusters. The largest predicted hybrid has been bioinformatically shown to be 185 kb.[234] Still, gene clusters for characterized Piz-containing 19-membered depsipeptides have ranged in size from 61-100 kb. However, as of now, there has not been a full biosynthetic characterization of all the transformations that lead to any one Piz-containing 19-membered depsipeptide. It is thus possible that there has been some premature shortening of predicted secondary metabolite biosynthetic gene clusters (smBGC). In this case, a size of 177 kb for the *Streptomyces incarnatus* NRRL 8089 gene cluster could be accurate but with current knowledge this is unlikely.
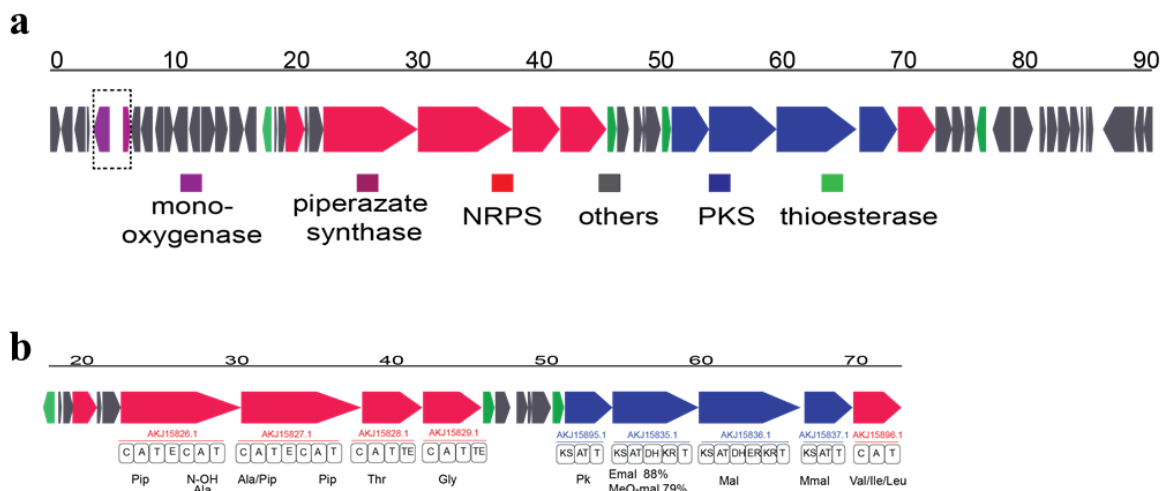
Scheme 3-1. a) Gene cluster from *Streptomyces incarnatus* NRRL 8089 that contains a homolog to ktzT, represented here from genes AKJ15806.1 to AKJ15860.1 b) A close-up on the gene cluster emphasizing the NRP and PK modules, predicted domains and A/AT domain specificities.

After the initial antiSMASH analysis, each gene of the contig was analyzed utilizing the browser-based program FramePlot to confirm that each assigned open reading frame had typical Streptomyces G/C codon usage and contained ribosomal binding sites. After analysis by FramePlot, each gene was then analyzed using BLASTP to confirm the functional annotation assigned by antiSMASH (**Table 3-10**). Next, *in silico* analyses were performed to obtain a clearer picture of the beginning and end of the *S. incarnatus* NRRL 8089 putatively Piz-containing smBGC. A MultiGeneBlast analysis[159] was performed to compare against similar gene clusters in anticipation that one of those gene clusters would contain clear cluster boundaries. Such a gene cluster was not observed. However, through BLASTP searches, it was evident that the individual genes routinely returned the highest sequence homology to genes in the strain *Streptomyces* sp. 75 (**Table 3.10**). *Streptomyces* sp. 75's predicted gene cluster had flanking transposases over an ~90 kb region. The NRRL 8089 contained a transposase on one

end (Accession no. AKJ15852.1) with limited sequence homology to one of the *Streptomyces* sp. 75's transposase. In between the two *Streptomyces* sp. 75's transposases, NRRL 8089 genes demonstrated high sequence homology to the vast majority of *Streptomyces* sp. 75's gene. On the other side of the second transposase, the identity and sequence homology between the genes of *Streptomyces* sp. 75's and *S. incarnatus* NRRL 8089 significantly decreased. Thus, the two *Streptomyces* sp. 75 transposases were used to roughly estimate the size of *Streptomyces incarnatus* NRRL 8089's putative Piz-containing NRP-PK producing gene cluster as a highly speculative 90 kb (**Scheme 3-1-a**).

Initial *in silico* analysis of the *Streptomyces incarnatus* NRRL 8089 gene cluster predicted that it was producing an azinothricin-type molecule (**Scheme 3-1-b**). antiSMASH, PKS/NRPS analysis, nor NRPSpredictor2 were able to give a complete A-domain specificity prediction on the number of Piz amino acids present in the produced molecules. Initially, the SANDPUMA algorithm predicted the presence of two Piz's within the molecule, and the other two algorithms did not predict the presence of Piz-specific A domains. In seeking novel compounds, it was noted that there are numerous azinothricin-type compounds with two Piz's (**Figures 2-4 and 2-5**). Further comparison (**Figure 3-22**) with the module and domain organization of biosynthetically-characterized 19-membered depsipeptide compounds that contain both NRP and PK modules [polyoxypeptin (**2.45**), aurantimycin (**2.41**) (**seen in Figure 2.5**) dentigerumycin A (**4.5**), and dentigerumycin B (**4.6**) with **4.7 and 4.8** being linear peptide congeners) (**seen in Figure 4.1**)] was then completed. This led to the observation that the *S. incarnatus* NRRL 8089 gene cluster contained a fifth main NRPS module leading to a total of seven A domains. When contrasted with the six A domains found within the main NRPS modules in all known 19-membered depsipeptides (**Figure 3-22**) this difference led to the

92

speculation that the produced compound could have some novelty. However, there was uncertainty as to why the two adjacent NRPS modules both contained TE domains. Several hypotheses were: one of the TE domains was non-operational although that was not indicated via the *in silico* analysis; the bacteria was producing a new variant of the Piz-containing NRP-PKS depsipeptides; or that the bacteria was somehow funneling the assembly into one (AKJ15828.1) or the other (AKJ15829.1) of the NRPS TE domain-containing modules. BLASTP analysis of the major NRP and PK domains (**Table 3.9**) against the all five characterized 19-membered depsipeptide compounds demonstrated clear relation to this type of compound but with lower sequence homology in NRPS modules AKJ15828.1 and AKJ15829.1, suggesting possible divergence.
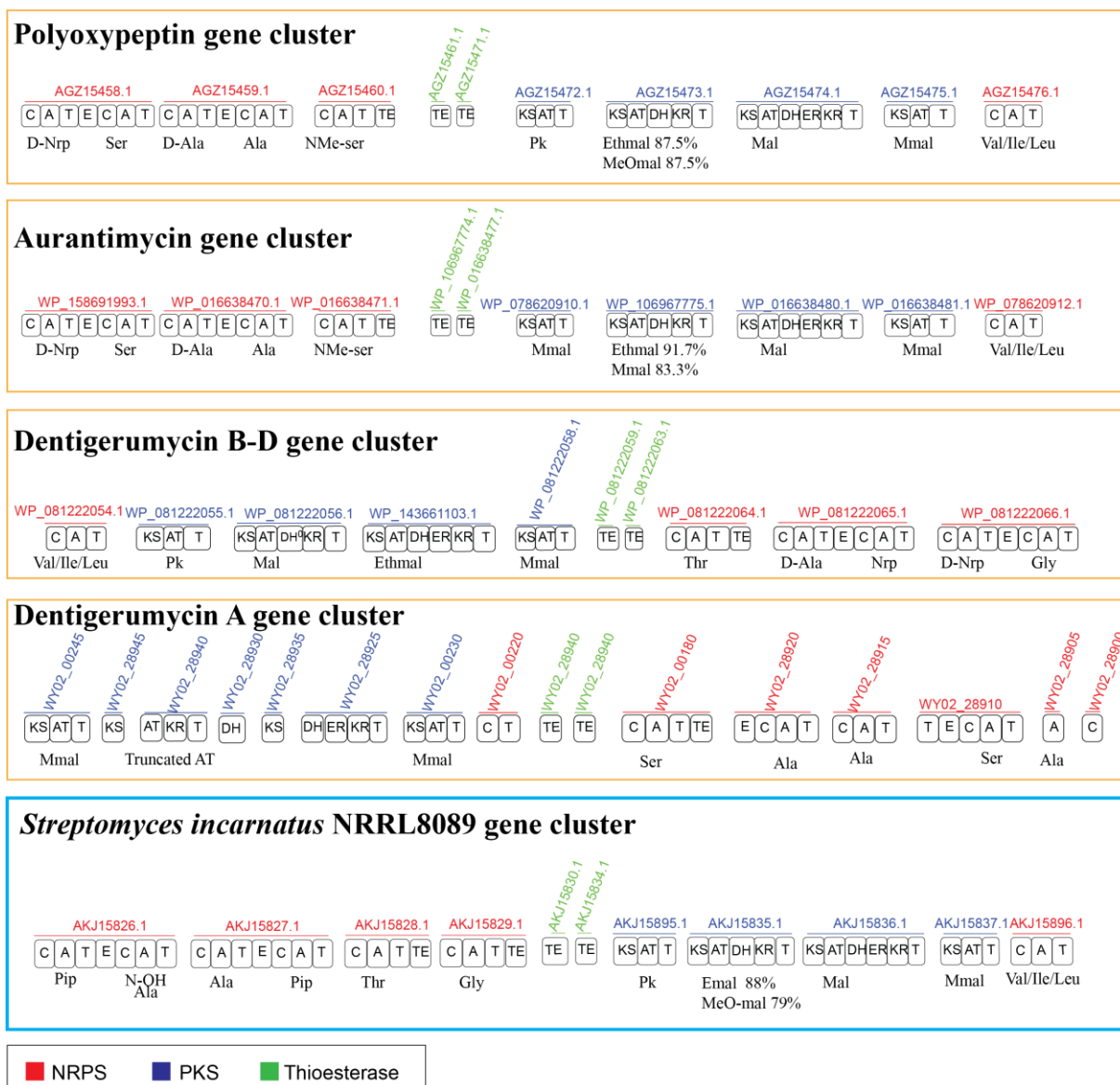
Figure 3-22. Comparison of *Streptomyces incarnatus* NRRL 8089 predicted gene organization with the gene organization of partially characterized 19-membered azinothricin-like gene clusters.

Table 3-9 BLASTP comparison of the genome mined *Streptomyces incarnatus* NRRL 8089 with characterized 19-membered depsipeptides.						
Natural Product	Genome-mined Strain	Polyoxypeptin (**2.45**)	Aurantimycin (**2.41**)	Dentigerumycin A (**4.5**)	Dentigerumycin B/C/D (**4.6, 4.7**, 4.8****)	Marformycins (**2.51**)
Bacterial Strains	*Streptomyces incarnatus* NRRL 8089	*Streptomyces* **sp. MK498-98F14** taxid: 1414447	*Streptomyces aurantiacus* **JA 4570** taxid: 1286094 taxid: 47760 2 taxid id's for different sequenced genomes of the same strain	*Pseudonocardia* **sp. AL041005-10** taxid : 445576 2 reference genomes deposited under this taxid id	*Streptomyces* sp. M41 taxid 1955065	*Streptomyces drozdowiczii* strain SCSIO 10141 taxid: 202862
Proposed function	Accession no.	Identity/Similarity to the *Streptomyces incarnatus* NRRL 8089 protein sequence for a gene. Accession no. or Enzyme name of gene protein sequence with highest identity.				
L-lysine 6-monooxygenase	AKJ15810.1	45/5 AGZ15458.1	72/95 EPH43660.1	59/95 ALE77168.1	84/99 WP_081222078.1	MfnI 50/94
transcriptional regulator (**KtzT homolog**)	AKJ15811.1	No significant similarity (nss)	75/99 EPH43659.1	old: 53/100 WP_083474028.1 new: 52/97 WP_145983356.1	87/100 WP_081222077.1	MfnJ 54/99
thioesterase	AKJ15821.1	34/95 AGZ15477.1	EPH39677.1 62/96 ; EPH43674.1 35/99	33/87 WP_062394358.1	74/94 WP_081222075.1	nss
FAD-dependent oxidoreductase	AKJ15825.1	87/99 AGZ15457.1	WP_016638468.1 79/99	69/99 WP_062394347.1	87/99 WP_081222067.1	nss
peptide synthetase	AKJ15826.1	82/99 AGZ15458.1	80/95 WP_158691993.1	65/98 ALE77164.1	72/99 WP_081222066.1	MfnC 45/98
peptide synthetase	AKJ15827.1	74/100 AGZ15459.1	69/100 WP_016638470.1	59/96 ALE77166.1	70/99 WP_081222065.1	MfnC 47/99
peptide synthetase	AKJ15828.1	57/99 AGZ15460.1	59/100 WP_016638471.1	43/80 ALE77165.1	78/100 WP_081222064.1	MfnE 45/99
peptide synthetase	AKJ15829.1	60/99 AGZ15460.1	75/100 WP_016638471.1	44/65 ALE77165.1	57.24/99 WP_081222064.1	MfnE 44/99
thioesterase	AKJ15830.1	84/100 AGZ15461.1	WP_106967774.1 82/100	35/95 ALE77171.1	85/98 WP_081222063.1	nss

cytochrome P450	AKJ15833.1	81/100 AGZ15470.1	86.54/100 WP_016638746.1	65.06/99 WP_062394378.1	85/100 WP_081222060.1	MfnN 31/91
thioesterase	AKJ15834.1	80/99 AGZ15471.1 (thioesterase II)	76/99 WP_016638477.1	31/82 ALE77169.1	76/96 WP_081222059.1	
modular polyketide synthase	AKJ15895.1	75/100 AGZ15472.1	77/100 WP_078620910.1	53/99 ALE77181.1	70/99 WP_081222058.1	
Type I polyketide synthase	AKJ15835.1	74/99 AGZ15473.1	71/99 WP_106967775.1	49/93 WP_062394429.1 ; [KR-T-KS-AT-DH-KR-T-KS]* 58/35 ALE77178.1	71/100 WP_143661103	
Type I polyketide synthase	AKJ15836.1	76/98 AGZ15474.1	75/100 WP_016638480.1	50/84 WP_062394412.1; [KS-AT-DH-ER-KR]* 67/29 ALE77178.1	73/99 WP_081222056.1	
acyl transferase	AKJ15837.1	82/99 AGZ15475.1	79/99 WP_016638481.1	62/99 ALE77175.1	76/99 WP_081222055.1	
non ribosomal peptide synthetase	AKJ15896.1	73/100 AGZ15476.1	70/100 WP_078620912.1	47/99 ALE77173.1	69/100 WP_081222054.1	

*NRPS/PKS Analysis was used to determine domains of select modules which were highly truncated in the reference genome reported for dentigerumycin A
** **4.7** and **4.8** are linear peptide congeners of the cyclic depsipeptide **4.6;** they are all produced from the same gene cluster

Table 3-9. BLASTP comparison of the NRP and PK domains of the genome mined *Streptomyces incarnatus* NRRL 8089 with characterized 19-membered depsipeptides.

Size (a.a)	Accession #	Proposed Function	Closest Homolog	Source	Identity/ Similarity
		Table 3-10. *Streptomyces incarnatus* NRRL 8089: list of bioinformatically predicted genes from the contig with accession no. CP011498, pg. 1 of 3			
849	AKJ15806.1	amidinotransferase	WP_115914648.1	Streptomyces sp. 75	91/100
942	AKJ15807.1	hypothetical protein	WP_115914649.1	Streptomyces sp. 75	74/56
909	AKJ15808.1	acetylglutamate kinase	WP_115914654.1	Streptomyces sp. 75	70/97
237	AKJ15809.1	MbtH family protein	WP_115914744.1	Streptomyces sp. 75	88/85
1353	AKJ15810.1	L-lysine 6-monooxygenase	WP_115914650.1	Streptomyces sp. 75	86/100
645	AKJ15811.1	FMN-binding negative transcriptional regulator	WP_115914651.1	Streptomyces sp. 75	90/100
777	AKJ15812.1	ABC transporter permease	WP_115907869.1	Streptomyces sp. 75	83/99
975	AKJ15813.1	ATP-binding cassette domain-containing protein	WP_093931193.1	Streptomyces sp. 75	84/97
693	AKJ15814.1	hypothetical protein	WP_115914652.1	Streptomyces sp. 75	86/100
741	AKJ15815.1	MarR family transcriptional regulator	WP_115907872.1	Streptomyces sp. 75	68/99
1215	AKJ15816.1	acetylornithine transaminase	WP_115914653.1	Streptomyces sp. 75	89/99
906	AKJ15892.1	acetylglutamate kinase	WP_115914654.1	Streptomyces sp. 75	87/100
1152	AKJ15817.1	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	WP_115914655.1	Streptomyces sp. 75	91/100
1029	AKJ15818.1	N-acetyl-gamma-glutamyl-phosphate reductase	WP_115914656.1	Streptomyces sp. 75	93/100
1182	AKJ15819.1	class A beta-lactamase-related serine hydrolase	WP_115914657.1	Streptomyces sp. 75	75/93
984	AKJ15820.1	hypothetical protein	WP_115914658.1	Streptomyces sp. 75	85/100
744	AKJ15821.1	thioesterase	WP_115914659.1	Streptomyces sp. 75	87/99
216	AKJ15822.1	MbtH family protein	WP_115914660.1	Streptomyces sp. 75	93/98
645	AKJ15893.1	regulator	WP_115914745.1	Streptomyces sp. 75	92/100
1581	AKJ15823.1	amino acid adenylation domain-containing protein	WP_115914661.1	Streptomyces sp. 75	88/100
303	AKJ15824.1	acyl carrier protein	WP_115914662.1	Streptomyces sp. 75	97/76
1188	AKJ15825.1	NAD(P)/FAD-dependent oxidoreductase	WP_115914663.1	Streptomyces sp. 75	98/100
7794	AKJ15826.1	amino acid adenylation domain-containing protein	WP_115914664.1	Streptomyces sp. 75	93/100
7767	AKJ15827.1	amino acid adenylation domain-containing protein	WP_115914665.1	Streptomyces sp. 75	89/100
3888	AKJ15828.1	peptide synthase	WP_115914666.1	Streptomyces sp. 75	92/100
3789	AKJ15829.1	nonribosomal peptide synthetase [Streptomyces]	WP_115914666.1	Streptomyces sp. 75	60/99
735	AKJ15830.1	thioesterase	WP_115914746.1	Streptomyces sp. 75	93/100

Table 3-10. continued pg. 2 of 3

972	AKJ15831.1	ATP-binding cassette domain-containing protein	WP_115914667.1	Streptomyces sp. 75	94/100
726	AKJ15894.1	ABC transporter permease	WP_115914668.1	Streptomyces sp. 75	95/100
261	AKJ15832.1	acyl carrier protein	WP_115914669.1	Streptomyces sp. 75	91/100
1251	AKJ15833.1	cytochrome P450	WP_115914670.1	Streptomyces sp. 75	92/100
738	AKJ15834.1	thioesterase	WP_115914671.1	Streptomyces sp. 75	90/99
3084	AKJ15895.1	type I polyketide synthase	WP_115914672.1	Streptomyces sp. 75	87/99
5577	AKJ15835.1	SDR family NAD(P)-dependent oxidoreductase	WP_115914673.1	Streptomyces sp. 75	87/100
6585	AKJ15836.1	SDR family NAD(P)-dependent oxidoreductase	WP_115914674.1	Streptomyces sp. 75	87/100
3111	AKJ15837.1	acyltransferase domain-containing protein	WP_115914675.1	Streptomyces sp. 75	92/99
3132	AKJ15896.1	nonribosomal peptide synthetase	WP_115914676.1	Streptomyces sp. 75	88/100
1359	AKJ15838.1	crotonyl-CoA carboxylase/reductase	WP_115914677.1	Streptomyces sp. 75	92/100
990	AKJ15839.1	ketoacyl-ACP synthase III	wp 115914678.1	Streptomyces sp. 75	94/100
885	AKJ15840.1	3-hydroxybutyryl-CoAdehydrogenase	WP_115914679.1	Streptomyces sp. 75	92/99
759	AKJ15841.1	thioesterase	WP_115914680.1	Streptomyces sp. 75	84/100
1500	AKJ15842.1	3-(3-hydroxy-phenyl)propionate hydroxylase	REE26142.1	Streptomyces sp. 75	83/99
1578	AKJ15843.1	acyl-CoA carboxylase subunit beta	WP 115914681.1	Streptomyces sp. 75	94/100
519	AKJ15844.1	hypothetical protein	WP_115914682.1	Streptomyces sp. 75	92/100
843	AKJ15845.1	hypothetical protein	WP_115914683.1	Streptomyces sp. 75	93/99
1008	AKJ15846.1	ATP-binding cassette domain-containing protein	WP_115907870.1	Streptomyces sp. 75	82/94
777	AKJ15847.1	ABC transporter permease	WP_115907869.1	Streptomyces sp. 75	83/99
231	AKJ15848.1	[protein-PII] uridylyltransferase	WP_115905648.1	Streptomyces sp. 75	39/36
198	AKJ15849.1	hypothetical protein	WP_115911074.1	Streptomyces sp. 75	50/26
396	AKJ15897.1	dipeptidase	WP_115906062.1	Streptomyces sp. 75	38/29
2589	AKJ15850.1	FtsX-like permease family protein	WP_115907726.1	Streptomyces sp. 75	31/28
723	AKJ15851.1	ABC transporter ATP-binding protein	WP_115907746.1	Streptomyces sp. 75	42/99
303	AKJ15852.1	IS630 family transposase	WP_115914754.1	Streptomyces sp. 75	31/35
753	AKJ15853.1	Thioesterase	WP 115914629.1	Streptomyces sp. 75	51/96
324	AKJ15854.1	Acyl carrier protein	WP 115914627	Streptomyces sp. 75	32/69
525	AKJ15855.1	putative flippase GtrA	REE26146.1	Streptomyces sp. 75	84/98
189	AKJ15856.1	beta-galactosidase	WP_115907185.1	Streptomyces sp. 75	29/83
1227	AKJ15857.1	cytochrome P450	WP_115914684.1	Streptomyces sp. 75	88/100

903	AKJ15898.1	1-deoxy-D-xylulose-5-phosphate synthase	WP_115910440.1	Streptomyces sp. 75	32/66
693	AKJ15858.1	Transketolase	WP 115910169.1	Streptomyces sp. 75	36/86
1128	no annotation in 8089	MFS transporter	WP_115914646.1	Streptomyces sp. 75	46/95
1011	AKJ15859.1	Aldo/keto reductase	WP 115910951	Streptomyces sp. 75	40/96
561	AKJ15860.1	ribosomal-protein-alanine N-acetyltransferase	REE32307.1	Streptomyces sp. 75	35/63

Table 3-10. *Streptomyces incarnatus* **NRRL 8089: list of bioinformatically predicted genes from the contig with accession no. CP011498, pg. 3 of 3.**

3.3.5 ^{15}N NMR Guided Fractionation of Piperazic Acid-Containing Metabolites in *Streptomyces incarnatus* NRRL 8089

A significant challenge in any natural discovery program is to separate the desired compounds from the complex mixture of primary and secondary metabolites in an organism's crude metabolite extract. *Streptomyces incarnatus* NRRL 8089 was chosen as a strain to further investigate due to its clear and stable ^{15}N NMR signature as well as the potential novelty indicated by its putative gene cluster. In the proton spectrum of *S. incarnatus* NRRL 8089 crude mixture shown in **Figure 3-24,** a complicated crude extract is observed. From this proton spectrum of the crude extract alone, it would be challenging, if not impossible, to make an informed decision on whether the extract contained any Piz-containing natural products. Even with initial fractionation (**Figure 3-25),** it would be near impossible to make an informed decision on where to focus isolation efforts based on proton spectra alone. Given this chemical complexity of the extraction of an organism's metabolic products, bioassays have long been harnessed as a powerful tool to guide natural product isolation (**Figures 1-8 and 1-9**). In the search for novel Piz-containing natural products, I had hypothesized that the Piz ^{15}N NMR signature N-H correlation and resulting N-H centred spin system could be used as a guide to isolate the *in silico* predicted molecules (**Figure 3-23**). This research aimed to isolate Piz-containing natural products from *Streptomyces incarnatus* NRRL 8089 utilizing the observed nitrogen-proton resonance, complete structure elucidation of those compounds, and then test them for bioactivity across a range of bioassays.
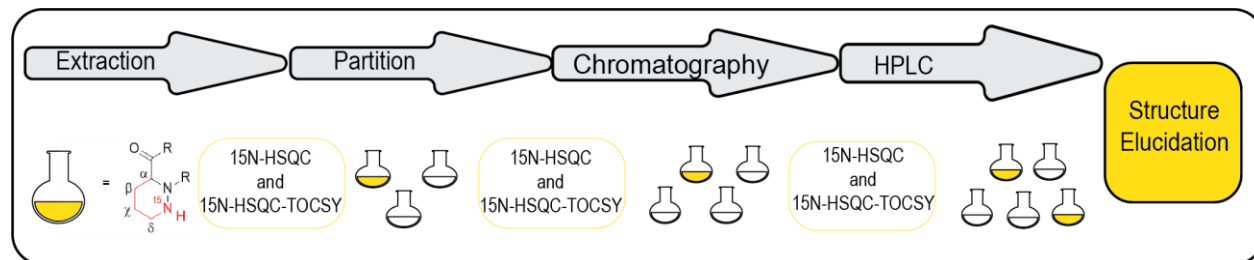
Figure 3-23. Workflow for Piz ^{15}N NMR-guided fractionation.

"8089-C-15N2 L-Orn" 1 1 C:\Bruker\TopSpin4.0.6\data
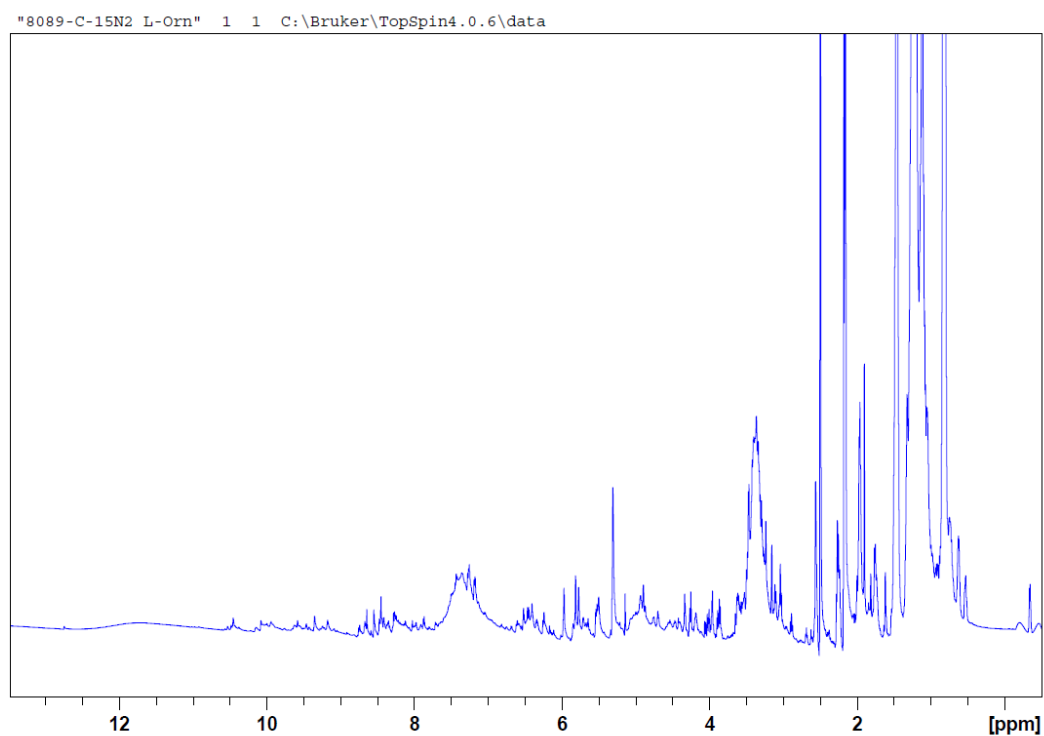
Figure 3-24. Full proton spectrum of labelled crude *S. incarnatus* NRRL 8089 metabolites recorded at 600 MHz in DMSO-d_6.
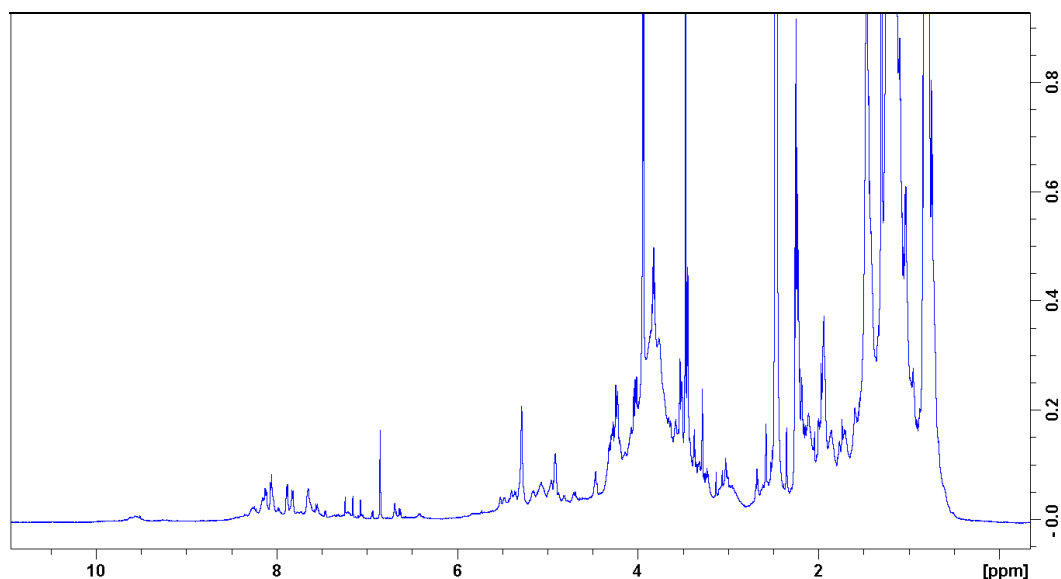
Figure 3-25. Full ^{1}H spectrum of labelled/unlabelled mixture of Fraction B of *S. incarnatus* NRRL 8089 metabolites recorded at 600 MHz in DMSO-*d₆*.

In order to confirm the observed Piz N-H signal (**Figure 3-17**), one tray (37 cm X 20 cm; 400 mL) of *Streptomyces incarnatus* NRRL 8089 was grown on ISP4 agar medium with the addition of labelled ^{15}N$_2$-L-Orn. The resulting crude extract was then subjected first to ^{1}H-^{15}N HSQC (**Figure 3-26 & 3-27)** and then ^{1}H-^{15}N HSQC-TOCSY (**Figure 3-29)** NMR experiments. These experiments confirmed the previously seen Piz N-H correlation at δ_H 5.43; δ_N -301, as well as a Piz-like spin system (**Tables 3-2 and 3-3**). The ^{1}H-^{15}N HSQC spectrum in **Figure 3-28** was taken with half the number of increments in the F1 dimension as in the spectrum in **Figure 3-27** so it required only half the time. Nevertheless, the information obtained is comparable in both spectra.

HSQC15N referenced to CH3NO2 at 0ppm
hsqcetf3gpsi for H1/N15 correlation;J(HN)=90Hz
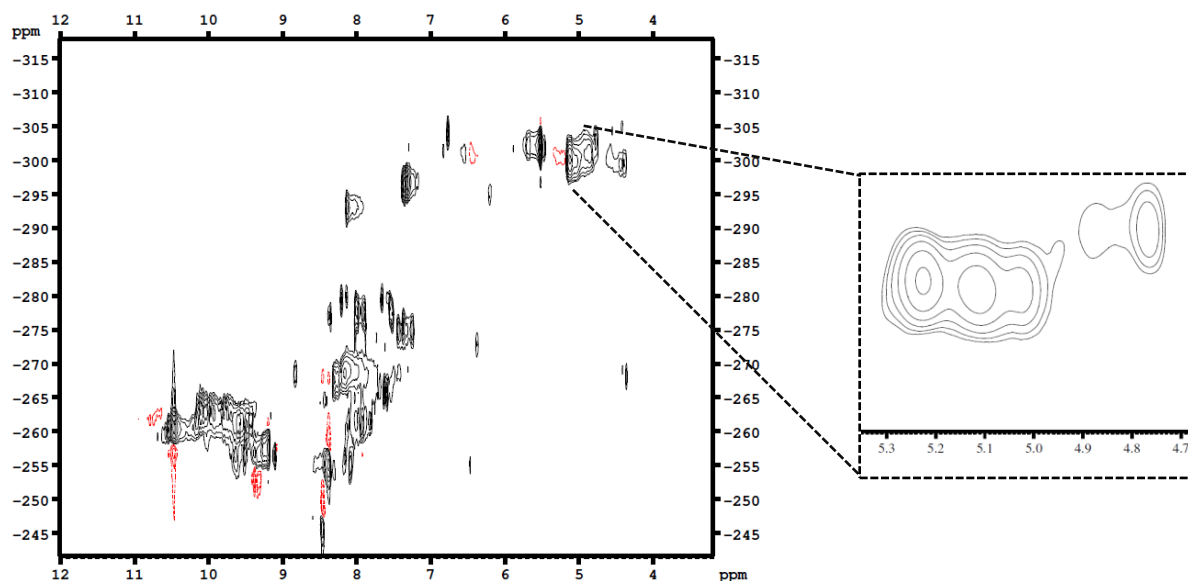
Figure 3-26. Scale up: Full ^{1}H-^{15}N HSQC spectrum of labelled crude _S. incarnatus_ NRRL 8089 metabolites recorded at 600 MHz in DMSO-d_6. Inset: Piz-like N-H correlations.
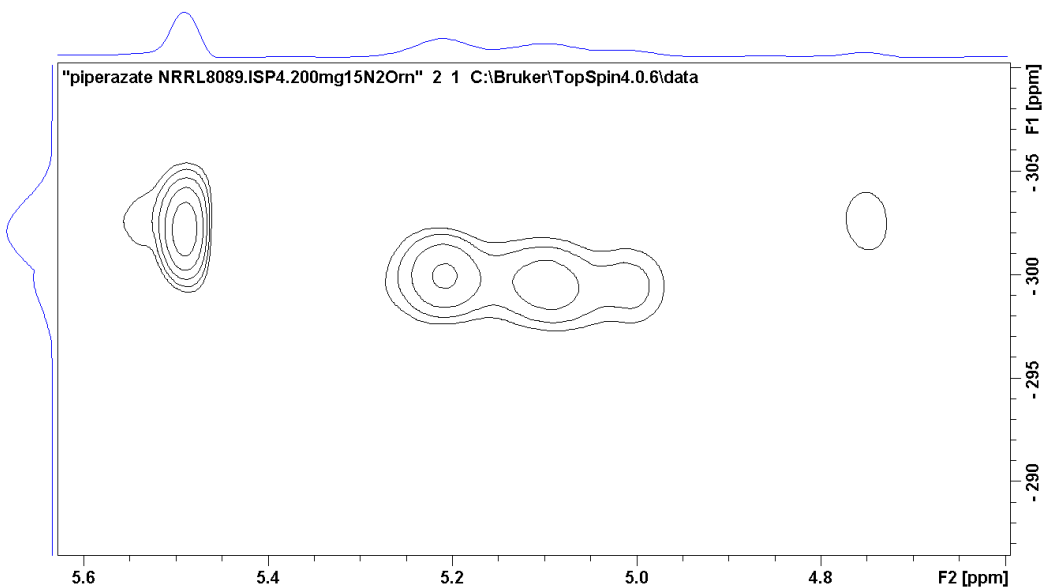
Figure 3-27. Scale-up: ^{1}H-^{15}N HSQC of _Streptomyces incarnatus_ NRRL 8089 crude extract obtained from 400 mL of ISP4 agar with addition of 200 mg of ^{15}N$_2$-L-Orn. 600 MHz, in DMSO-d_6; F2:96 scans F1: TD = 64.

103

"piperazate NRRL8089.ISP4.200mg15N2Orn" 4 1 C:\Bruker\TopSpin4.0.6\data
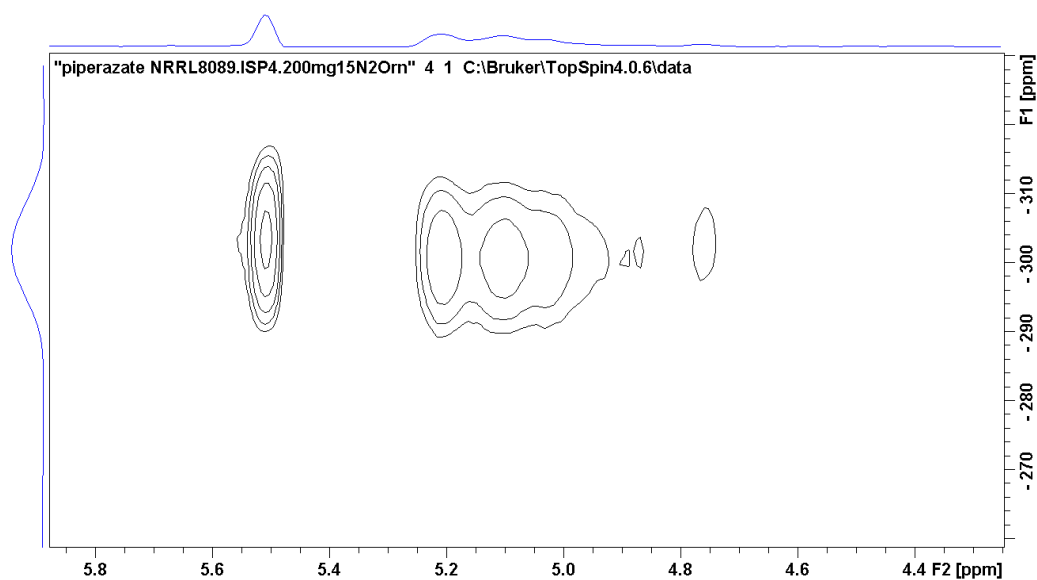
Figure 3-28. Scale-up: ^{1}H-^{15}N HSQC of *Streptomyces incarnatus* NRRL 8089 crude extract obtained from 400 mL of ISP4 agar with addition of 200 mg of ^{15}N$_2$-L-Orn. 600 MHz, in DMSO-d_6; F2:96 scans F1: TD = 32.
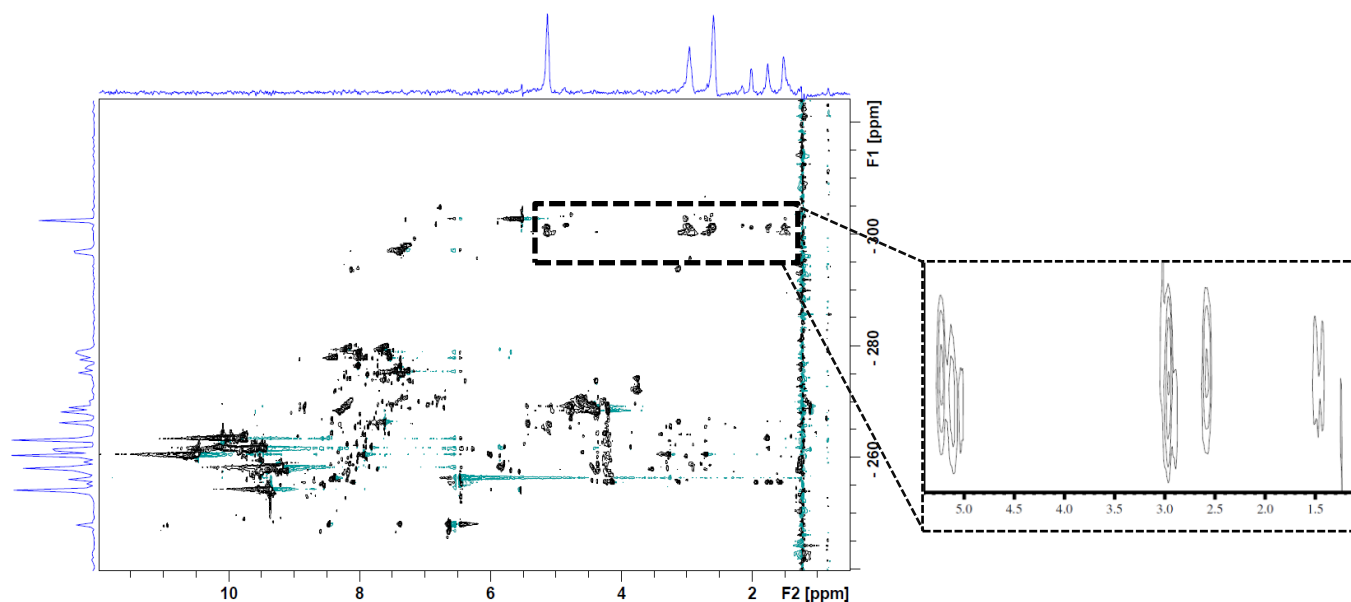
Figure 3-29. Scale-up: ^{1}H-^{15}N HSQC-TOCSY spectrum of labelled crude *S. incarnatus* NRRL 8089, recorded at 600 MHz in in DMSO-d_6. Piz-spin system projected on to the F2 axis and seen within the inset.

104

"piperazate NRRL8089.ISP4.200mg15N2Orn" 3 1 C:\Bruker\TopSpin4.0.6\data
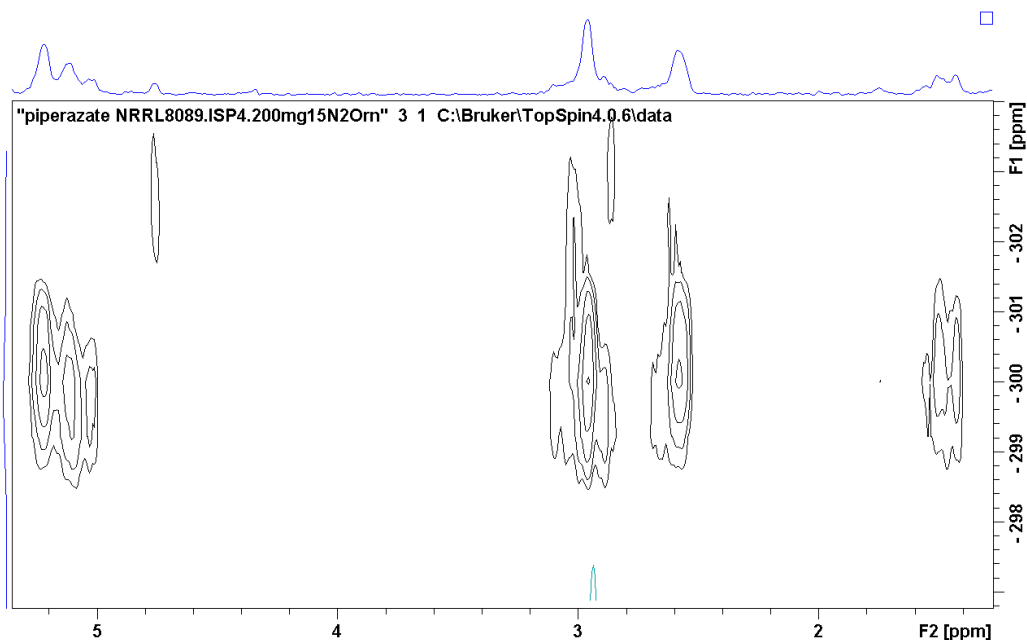
Figure 3-30. ^{1}H-^{15}N HSQC-TOCSY of *Streptomyces incarnatus* NRRL 8089 crude extract obtained from 400 mL of ISP4 agar with addition of 200 mg of ^{15}N$_2$-L-Orn. 600 MHz, in DMSO-d_6; F2:128 scans F1: TD=128.

Earlier work on these genome-mined strains had demonstrated that incorporation of a stable isotopically labelled precursor into a bacterial metabolite could enhance the ^{15}N NMR Piz signal such that it was visible in a crude extract yet could also mask a low compound concentration. Thus, signal strength cannot always be correlated with enough material to complete the discovery process. This result can be compared to bioassay-guided fractionation when a potent bioactive minor metabolite demonstrates profound bioassay results. To obtain enough compound for structure elucidation, the strategy (when working with bacteria) is to scale-up or grow larger quantities of the producing bacterium. Therefore, to obtain more material for isolation and structure elucidation, *Streptomyces incarnatus* NRRL 8089 was grown on 8 L of ISP4 agar medium with the addition of unlabelled L-Orn. The resulting pale yellow waxy crude extract was combined with the ^{15}N labelled crude extract obtained from one tray of bacteria culture as discussed above. The combined crude material was subsequently partitioned between EtOAc and water, with a significant portion of this EtOAc-soluble fraction insoluble in

MeOH and soluble in hexanes. Both hexane and MeOH soluble fractions were tested for the Piz N-H signal using the ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HQSC-TOCSY NMR experiments, with the signal found in the MeOH soluble fraction. The MeOH soluble fraction was then subjected to a Sephadex^{TM} LH-20 column with each subsequent fraction then tested for the Piz N-H signal using the ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HQSC-TOCSY NMR experiments. The signals were found in an earlier eluting fraction, which was then subjected to a step-gradient normal-phase Sep-Pak with the MeOH wash containing the Piz NMR signal of interest. This MeOH-soluble fraction was then fractionated on a C_{18} Sep-Pak with the fractions C and D containing the Piz NMR signal of interest. These fractions were moved to final purification through reversed-phase HPLC. During HPLC purification, however, there proved to be too little sample for adequate method development.

Increased scale-up was thus necessary and the *S. incarnatus* NRRL 8089 was grown as 58 L of ISP4 agar lawns supplemented with unlabelled L-Orn. This was combined with the material obtained from growing *S. incarnatus* NRRL 8089 on 800 mL of ISP4 agar lawns supplemented with ^{15}N$_2$-L-Orn. Given the large amount of hydrophobic material previously seen in the crude metabolite extract, the obtained crude was first partitioned between hexanes and distilled water with the aqueous layer being further partitioned with EtOAc. The hexanes, EtOAc, and water soluble partitions were subjected to nitrogen NMR experiments with the Piz-signal found in the EtOAc soluble partitions. The EtOAc-soluble fraction was then triturated with cold MeOH and gravity filtered to remove the fatty residues. The MeOH soluble fraction was then fractionated on a Sephadex^{TM} LH-20 column, combined via TLC spots, and the resulting fractions subjected to a ^{1}H-^{15}N HSQC NMR experiment, and those fractions with the nitrogen correlation of interest were then subjected to the ^{1}H-^{15}N HQSC-TOCSY experiment. Again, the early eluting fraction

contained the Piz-signal of interest. To optimize the C_{18} flash chromatography, a small portion of the EtOAc soluble material was fractionated using a C_{18} Sep-Pak and the obtained fractions analyzed via nitrogen NMR. This separation was replicated on a larger scale on a C_{18} flash column (30 cm x 20 cm) with fractions C and D demonstrating the Piz N-H correlation and spin system in the ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HQSC-TOCSY NMR experiments. Fraction D, with less material, was set aside. A portion of the fraction C was then subjected to reversed-phase HPLC where-in the HPLC separation was divided into 10 min time signatures since the eluent contained no distinguishing UV absorption pattern. Subsequent ^{15}N NMR analysis with the ^{1}H-^{15}N HSQC experiment demonstrated loss of the Piz signal until all the fractions were recombined. Thus, the separation was shown to require a low PH buffer to prevent significant tailing and compound drag across the column. With the addition of 0.01%TFA to the water solvent, clear HPLC peaks were obtained with the peaks labelled 8089.Et.B.C.11, 8089.Et.B.C.13, 8089.Et.B.C.14, 8089.Et.B.C.15, 8089.Et.B.C.16 and 8089.Et.B.C.17 demonstrating the Piz signal in the ^{1}H-^{15}N HSQC experiment (**Scheme 3.2**). Over the time required to separate all of fraction C, the peak labelled 8089.Et.B.C.15 demonstrated shifts in intensity and the appearance of two peaks, with NMR spectra demonstrating a large mix of compounds, and this was set aside. The peak labelled, 8089.Et.B.C.17, took over 20 minutes to fully elute from the column despite the acidic conditions; subsequent NMR analysis demonstrated that it was not a clean compound and this too was set aside. Thus, structure elucidation efforts, to be described in **Chapter 4**, were focused on the pure compounds initially named via HPLC peak: 8089.Et.B.C.11, 8089.Et.B.C.13, 8089.Et.B.C.14, and 8089.Et.B.C.16. Each of those compounds isolated as HPLC peaks, displayed the Piz N-H correlation in the ^{1}H-^{15}N HSQC experiment. 8089.Et.B.C.14 appeared to be a very interesting compound with the appearance of NMR correlations unique from the rest

and a mass of 796.1 as indicated by LRESIMS. However, it was produced in very low titers, was difficult to purify as it was adjacent to the abundant 8089.Et.B.C.13, and the structure was changing over the NMR collection time. With the minimal quantities collected and the changing spectra, structure elucidation proved especially challenging (**Data Set in Appendix A**) and could not be completed. 8089.Et.B.C.11 (**4.3**) was also produced in minor titers, however, the NMR spectra collected were not changing over NMR collection time – thus, as described in **Chapter 4**, structure elucidation was completed. 8089.Et.B.C.13 (**4.2**) and 8089.Et.B.C.16 (**4.1**) also demonstrated changing spectra over the NMR collection time. However, these compounds were the relatively abundant members from this family of minor compounds. As such, the structure elucidation, described in **Chapter 4**, was completed through careful analysis of the spectra along with collection in various deuterated solvents and varying the temperature of the samples.
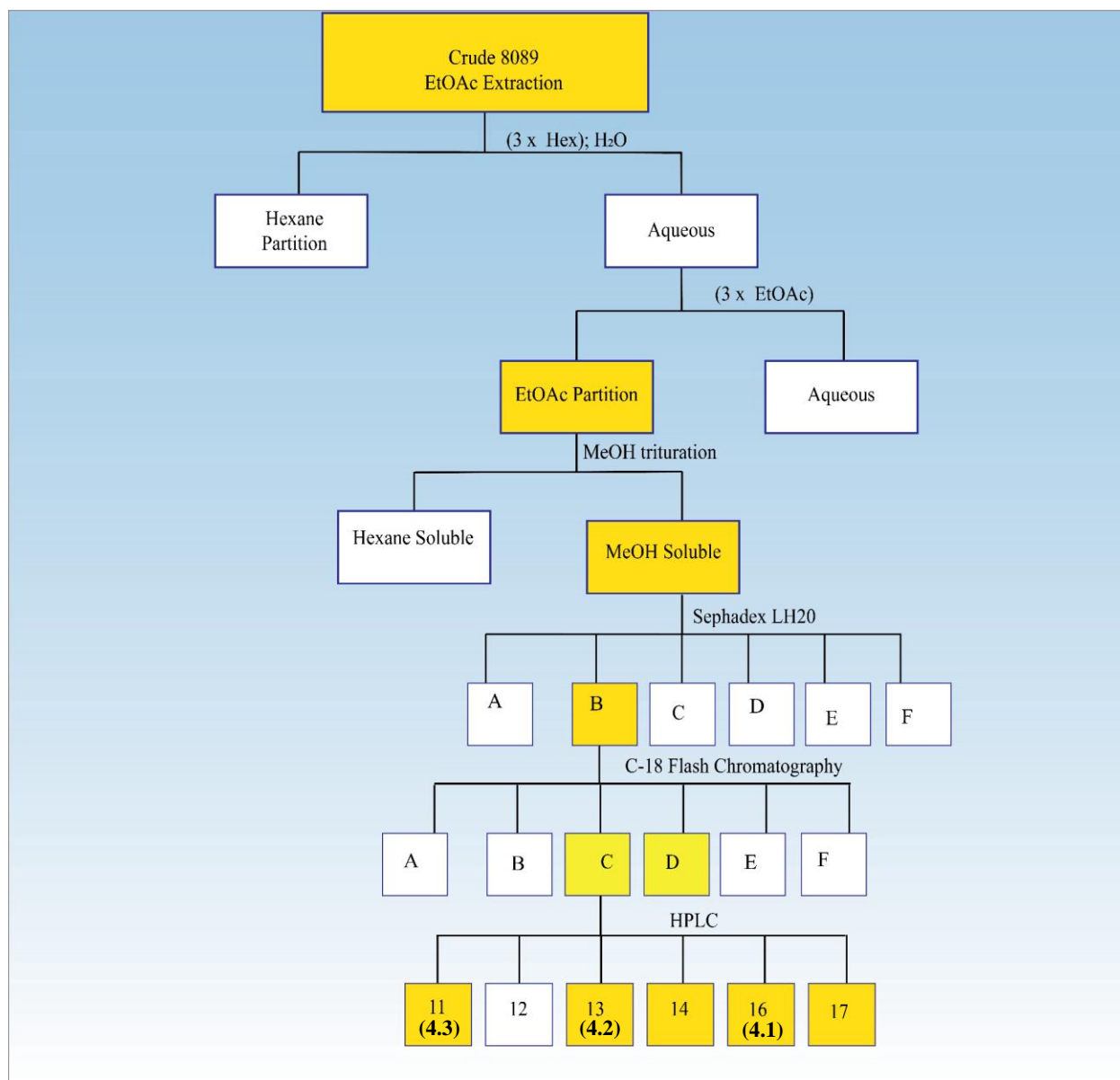
Scheme 3-2. ^{15}**N NMR-guided isolation scheme of** *Streptomyces incarnatus* **NRRL 8089. Yellow indicates the presence of the Piz N-H signal and connected spin system in nitrogen NMR experiments.**

The bacterium was producing very minor amounts of 8089.Et.B.C.11, and portions of this small amount was lost as the isolation scheme was developed. As such, there was not enough material for unequivocal structure elucidation of incarnatapeptin A at this stage. The crude bacterial extract obtained from the growth of 87 L of ISP4 with added was given to David E. Williams, a research associate in the Andersen Lab, who isolated more of the compound with the previously isolated compound utilized as an HPLC standard. Additionally, the methyl ester of

109

incarnatapeptin A was produced and gave a crystal whose crystal structure was used for stereochemical assignment (**Appendix E**). The NMR data sets from this separation and derivatization are in **Appendix D.** Through these efforts incarnatapeptin B (**4.4**) was also found from *Streptomyces incarnatus* NRRL 8089's extract; this data and work are not reported here.[227] David Williams re-isolated 8089.Et.B.C.13 (**4.2**) and 8089.Et.B.C.16 (**4.1**) with formic acid modifier instead of trifluoroacetic acid, which prevented the previously observed equilibrium (to be discussed in **Chapter 4**). The NMR data sets from this separation are also of superior quality, served to support the previous structure assignments of the hemiketal form of the compounds and are in **Appendix C**.

3.4   Conclusion

Here I have shown that piperazic acid is likely to have predictable NMR features and then showed that ^{15}N NMR experiments could be used to observe the diagnostic aza nitrogen chemical shift and attendant proton chemical shifts, even from a crude bacterial metabolite extract. Furthermore, with stable isotopic feeding of the biosynthetic precursor ^{15}N$_2$-L-ornithine, the N-H correlation was visible even in extracts from small amounts of bacterial culture. Furthermore, with feeding ^{15}N$_2$-L-ornithine, I observed the entire Piz spin system from a crude bacterial metabolite extract using a ^{1}H-^{15}N HSQC-TOCSY NMR experiment. The application of these NMR experiments allowed me to select optimal growing conditions for isolation of Piz metabolites from strains of bacteria selected through genome mining. Altogether, this work demonstrated that ^{15}N NMR experiments could be used to guide the isolation of novel Piz compounds from a genome-mined strain predicted to contain Piz-producing machinery.

3.5 Experimental Section

3.5.1 General Experimental Procedures

All solvents used for partitioning and flash chromatography were Sigma-Aldrich HPLC grade, used without further treatment. For HPLC work, Sigma-Aldrich HPLC grade solvents were filtered through Milli-Q filter paper prior to use. Distilled water was used to prepare all growth media for terrestrial bacteria. Seawater collected by M. Leblanc was used to prepare growth media for all marine bacteria. Milli-Q H_2O was used for partitioning, reversed-phase Sep-Pak and HPLC. The ^1H, ^{13}C and ^{15}N NMR spectra were recorded on a Bruker AV-600 spectrometer with a 5 mm CPTCI cryoprobe. ^1H chemical shifts are referenced to the residual DMSO-d_6 peak (chemical shift δ 2.50) and ^{13}C chemical shifts are referenced to the DMSO-d_6 solvent peak (δ 39.51). The ^{15}N NMR spectra are in reference to CH_3NO_2 (0 ppm) with a variability of 1 ppm. Low resolution electrospray-ionization mass spectrometer (LRESIMS) were recorded on Bruker HCT Ultra PTM Discovery System and high-resolution electrospray-ionization mass spectrometer (HRESIMS) were recorded on a Waters / Micromass LCT time-of-flight (TOF) mass spectrometer equipped with an electrospray ion source. Merck Type 5554 silica gel plates and Whatman MKC18F plates were used for analytical thin layer chromatography (TLC). Sephadex™ LH-20 purchased from Sigma was used for gel filtration chromatography. *para*-Anisaldehyde stain with subsequent heat application was used to visualize TLC spots. Sep-Pak SPE Silica and C18 cartridges (2 g, 5 g, and 10 g) were purchased from Waters. C18 Sep-Pak's with free-flowing solid phase were dismantled and repacked into larger columns for large-scale C18 flash chromatography. Labelled $^{15}N_2$-L-Orn (99%) was obtained from Cambridge Isotope Laboratories, Inc and unlabelled L-Orn was obtained from either

Sigma-Aldrich or Fisher Scientific. All other media components were obtained from Fisher Scientific. Reversed phase HPLC purifications were performed on a Waters 1525 Binary HPLC pump attached to a Waters 2998 Photodiode Array Detector and equipped with Empower Software.

3.5.2   Bacterial Cultivation

3.5.2.1   *Streptomyces* sp. RJA2928, A Known Producer

First growth : Five trays with 400 mL of autoclave-sterilized MM1 agar were prepared by (1) autoclaving the medium combined with seawater; (2) allowing the agar to cool; and (3) spreading each tray with 5 mL of a liquid culture of the ~fifth-generation of the padanamide producer *Streptomyces* sp. RJA2928.[206] The five trays were grown for 14 d (22 °C), then the combined agar and mycelia were cut into small squares, transferred to a plastic container, and extracted three times with EtOAc. The three EtOAc extractions were combined and dried *in vacuo* to provide ~600 mg of a crude metabolite mixture. An aliquot of ~20 mg was dissolved in DMSO-d_6 and the sample was subjected to ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC-TOCSY and ^{1}H-^{1}H TOCSY with no Piz correlations observed. Next, the crude metabolite mixture was fractionated utilizing Sephadex™ LH-20 (MeOH) leading to three fractions A (brown oily solid, 126 mg), B (brown oily solid, 237 mg), and C (brown solid, 105 mg). An aliquot from each fraction was subjected to ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC-TOCSY and ^{1}H-^{1}H TOCSY, again with no Piz correlations observed.

Second growth and limit detection: 25 mL of MM1 broth was inoculated with an aliquot of partially thawed first-generation *Streptomyces* sp. RJA2928. Small Petri plates were prepared with 30 mL agar media containing 20 mg ^{15}N$_2$-L-Orn (**L**) and 30 mL agar media/20 mg L-Orn (**UL**) that had been added via syringe sterilization to the warm liquid agar before pouring. The

prepared Petri plates were inoculated with ~1 mL of the padanamide producer RJA2928, grown

for 14 days, and then harvested separately. Each sample was extracted with EtOAc (x3), the

extractions were combined and dried *in vacuo* to provide crude mixtures of **L** (11.5 mg) and **UL**

(10.8 mg) (hereby referred to as "crude"). The ^{1}H-^{15}N HSQC experiment was collected with

F2:96 scans; F1: TD = 32 for samples **L** and **UL**.

<u>Increasing amount of labelled media for bacterial growth</u>: A tray of solid MM1 agar was

prepared by first autoclaving 350 mL of MM1 agar with seawater and then adding 100 mg ^{15}N$_2$-

L-Orn. The labelled amino acid was incorporated by first dissolving it in previously autoclaved

sea water and then filter sterilizing it into the 350 mL of MM1 agar at around 70 °C. After the

agar cooled, this tray was spread with 5 mL of a liquid culture of *Streptomyces* sp. RJA2928,

grown for 14 days and then harvested into EtOAc. The EtOAc extract was gravity filtered to

separate solvent from agar and mycelium, then fresh portions of solvent were added to the

residual agar and mycelia (2x). The resulting EtOAc portions were combined and dried under

reduced pressure to yield 206 mg of a gummy brown crude extract called sample **ScL**. A small

aliquot of ~20 mg was dissolved in DMSO-d_6; the ^{1}H-^{15}N HSQC experiment was collected with

F2:48 scans; F1: TD = 32 and the ^{1}H-^{15}N HSQC-TOCSY was collected F2:128 scans; F1: TD =

128. As the ^{1}H-^{15}N HSQC-TOCSY is traditionally collected with a higher F1:TD value than the

number of scans in the F2 direction, various ratios of F2 scans to F1:TD values were sampled. In

all cases, running the ^{1}H-^{15}N HSQC-TOCSY with equivalent or greater F2 scans than F1:TD

values yielded observation of the more complete Piz spin system; F2:128 scans; F1: TD = 128

appeared to be ideal with a F1:TD = 96 also yielding decent observations

3.5.2.2 Growing Genome-Mined Strains

<u>Growth of genome-mined bacteria</u>: The four strains listed in **Table 3.4** were obtained from the Agricultural Research Service Culture Collection and revived in tryptic soy broth (TSB, purchased from Fisher Scientific) prepared with distilled water and grown at 28 °C for 3-7 d with shaking at 220 RPM. After 3-7 d, the inoculum was spread on two large Petri plates each containing 50 mL of either MM1 fresh water agar (0.5 g of glucose, 0.5 g of yeast extract, 1.0 g of peptone, 0.01 g of $FeSO_4 \cdot 7H_2O$, 0.02 g of Na_2HPO_4, 15.0 g of agar, 1 L of distilled water) or Mannitol Soya Flour agar (MS, 20 g of mannitol, 20 g soya flour, 100 mM calcium carbonate, 20.0 g of agar, 1 L of distilled water). Each of these media had 20-30 mg of $^{15}N_2$-L-Orn diluted in minimal sterilized water and then filter sterilized into the autoclaved agar media. Each Petri plate was grown for 14 d, and then the agar and mycelium were harvested together as small squares of agar lawn and extracted with EtOAc (3 x 150 mL). The EtOAc extracts were combined and dried under reduced pressure to yield 15-30 mg of crude material. ^{1}H-^{15}N HSQC spectra were obtained for each crude sample after dissolving in DMSO-d_6 with TD=32 in the F1 dimension and 48 scans in the F2 dimension. If no signal was obtained, more scans were added to the file using the command go.

***Streptomyces phaeochromogenes* NRRL B-1248 small scale-up:** The strain *Streptomyces phaeochromogenes* NRRL B-1248 was grown on 400mL of MM1 F agar which had had 100 mg of $^{15}N_2$ LOrn added as described above for RJA2928. After 14 days, the agar medium was extracted with EtOAc (3x) with the combined layers dried under reduced pressure to reveal 43 mg of crude material. This crude material was assessed for the Piz N-H correlation at δ_H 5.55; δ_N -304 and then separated via Sephadex^{TM} LH-20 (MeOH), leading to three fractions A (22mg, yellow oil), B (16.5 mg, yellow oily solid), and C (3.5 mg, brown solid) which were assessed for

the Piz N-H correlation. The fraction A (22 mg) containing the Piz signal at δ_H 5.55; δ_N -304 was further separated via step gradient C_{18} flash chromatography (Sep-Pak 2g). The fraction eluting with 3:1 H_2O:MeOH (v/v, 7 mg) contained the Piz N-H correlation at δ_H 5.55; δ_N -304 and was further separated via reversed-phase HPLC using an InertSustain C_{18}, 5µm, 10 x 250 mm column. Samples obtained were not sufficient for structure elucidation.

***Streptomyces griseus* subsp. griseus strain NRRL WC-3645 small scale-up:** *Streptomyces griseus* subsp. griseus strain NRRL WC-3645 (**Figure 3-8**) was grown on mannitol soy agar as 16 L of autoclaved agar medium, with 250 mg unlabelled ʟ-Orn/400mL added via filter-sterilization when at ~70 ℃, poured as 44 sterilized trays. After 14 days of growth, mycelia and agar were harvested into EtOAc, soaked for a minimum of 24 hours (x3), then gravity filtered and dried *in vacuo* to obtain a crude extract. This unlabelled extract was combined with the extract obtained from the growth of 200mL of mannitol soy agar, which had had 40 mg of $^{15}N_2$ ʟOrn added as described above. The ^1H-^{15}N HSQC experiment was performed on the labelled extract before combining. The combined crude extract (1 g) was separated via Sephadex^{TM} LH-20 (MeOH), whereby fraction C (28 mg) contained the N-H correlation of interest via NMR monitoring. This fraction was further separated via a C18 Sep-Pak (5 g) whereby the fraction eluting with 9:1 H_2O:ACN (v/v) demonstrated the N-H correlation of interest. Further separation via reversed-phase HPLC using an InertSustain C_{18}, 5µm, 10 x 250 mm column, gradient solvent system (9:1 H_2O:ACN to 2:3 H_2O:ACN over 60 minutes) led to the isolation of para-aminobenzoic acid (PABA) which demonstrated the N-H correlation of interest.

Growth of additional genome-mined bacteria: Continuing the investigation into genome-mined strains, nine strains were obtained from the Agricultural Research Service Culture Collection as listed in **Table 3.7** and revived in tryptic soy broth (TSB, purchased from Fisher

Scientific) prepared with distilled water and grown at 28 °C for 3-7 d with shaking at 220 RPM. After 3-7 d the inoculum was spread on 4 large Petri plates each containing 60 mL of either MM1 fresh water agar (0.5 g of glucose, 0.5 g of yeast extract, 1.0 g of peptone, 0.01 g of $FeSO_4 \cdot 7H_2O$, 0.02 g of Na_2HPO_4, 15.0 g of agar, 1 L of distilled water); ISP4 (10 g of soluble starch, 1 g of dipotassium phosphate, 1.0 g of magnesium sulfate USP, 1.0 g of sodium chloride, 2.0 g of ammonium sulfate, 2.0 g of calcium carbonate, 0.001 g of $FeSO_4 \cdot 7H_2O$, 0.001 g of $ZnSO_4 \cdot 7H_2O$, 0.001 g of $Mn(II)Cl_2$, 20.0 g of agar, 1 L of distilled water); tryptic soy agar (TSA, purchased from Fisher Scientific), or Mannitol Soya Flour agar (MS, 20 g of mannitol, 20 g soya flour, 100 mM calcium carbonate, 20.0 g of agar, 1 L of distilled water) with the addition of 60 mg of L-Orn diluted in minimal sterilized water and then filter sterilized into the autoclaved agar media. Each Petri plate was grown for 14 d, and then the agar and mycelium were harvested together as small squares of agar lawn and extracted with EtOAc (3 x 150 mL). The EtOAc extracts were combined and dried under reduced pressure to yield 20-30 mg of crude material. ^1H-^{15}N HSQC spectra were obtained for each crude sample after dissolving in DMSO-d_6 with TD=32 scans in the F1 dimension and 48 scans in the F2 dimension. If no signal was obtained, more scans were added to the file using the command go on the Bruker TopSpin software.

***Streptomyces incarnatus* NRRL 8089 Media Selection:** To identify the optimum lab conditions, *Streptomyces incarnatus* NRRL 8089 was revived and inoculated on four different agar media each containing unlabelled L-Orn. One large Petri plate prepared from 60 mL of each media was inoculated with the desired bacteria. After 2 weeks of growth the bacteria was harvested into enough ethyl acetate to cover the cut agar medium. The ethyl acetate extracts were filtered through cotton wool to remove the mycelia and the agar and dried *in vacuo* to give crude extracts; ISP4 (6.6 mg); MM1 F (7.9 mg); TSB (19.8 mg) and Soy Mannitol (10.1 mg). The

116

resulting crude extracts were dissolved in ~ 0.5 mL of deuterated DMSO-d_6, placed in standard 5mm NMR tubes, and were then subjected to ^{15}N-HSQC NMR experiments, where a weak ^{1}H-^{15}N (δ_H 5.01, δ_N −300) correlation was seen in the crude extract from bacteria grown on ISP4 (**Figure 3-5**), possibly consistent with a Piz residue(s).

Streptomyces incarnatus NRRL 8089 small scale-up with labelled material:

Subsequently, *Streptomyces incarnatus* NRRL 8089 was inoculated on a 400 mL lawn of solid ISP4 agar which had 200 mg of labelled ^{15}N$_2$-L-Orn dissolved in 50 mL of previously autoclaved distilled water that was added via syringe-filter to 350 mL of autoclaved ISP4 agar cooled to ~70°C, with the combined portions poured into an autoclaved aluminum foil-covered stainless steel pan. The decision was made to use 200 mg labelled ^{15}N$_2$-L-Orn per 400 mL media when growing *Streptomyces incarnatus* NRRL 8089 versus the 100 mg labelled ^{15}N$_2$-L-Orn per 400 mL medium used when growing padanamide producer *Streptomyces* sp. RJA2928 because it had been shown that NRRL 8089 was a producer of a sinefungin analog[231] which utilized L-Orn in its biosynthetic pathway.[235] As a precaution before scale-up, the structure and available NMR data for sinefungin[236] was analyzed to ensure there were no overlapping N-H signals with Piz. 5 mL of five day liquid growth of *Streptomyces incarnatus* NRRL 8089 was spread on the cooled and solidified ISP4 agar as inoculum with the pan recovered with foil and then left undisturbed for 14 days. After 14 days of growth, the lawns of agar were cut into small pieces and submerged in EtOAc for a minimum of 24 hours for solid-liquid extraction of the bacterial metabolites. Following the minimum period of 24 hours, the EtOAc portion was gravity filtered through cotton wool to separate the agar and bacterial mycelia from the saturated solvent and a fresh EtOAc portion was applied (2x). The EtOAc (3x) portions were then combined and dried *in vacuo* and the crude extract was subjected first to ^{1}H-^{15}N HSQC (**Figure 3-6**) and then to ^{1}H-^{15}N

HSQC-TOCSY (**Figure 3-7**) NMR experiments, giving the proton-nitrogen correlation and Piz-like spin systems seen in, consistent with the Piz system observed in *Streptomyces* sp. RJA2928.

3.5.3 ^{15}N NMR Guided Fractionation

<u>Initial large scale-up of *Streptomyces incarnatus* NRRL 8089:</u> To obtain more material for isolation and structure elucidation, *Streptomyces incarnatus* NRRL 8089 was inoculated on agar lawns prepared on large stainless steels pans through pouring 8 L of ISP4 agar medium to which 0.9 g/400 mL of unlabelled L-Orn was added. After 14 days of growth, the combined cells and solid agar medium were extracted with EtOAc which was dried under reduced pressure to form a crude metabolite mixture as previously described. This pale yellow waxy crude extract was combined with previously grown labelled extract for a total of 266.1 mg. This crude material was then partitioned between EtOAc and water and then dissolved in a minimal amount of MeOH to place on a Sephadex LH-20 column. A significant portion of this partition precipitated out of the MeOH solvent and the precipitate was removed by gravity filtration and rinsed with cold MeOH. The precipitate proved to be most soluble in hexanes leaving a waxy yellow mass, while the MeOH soluble partition proved to be a brown oily solid. Both hexane and MeOH soluble fractions were tested for the Piz signal using the ^{1}H-^{15}N HSQC NMR experiment and the signal was found in the MeOH soluble fraction. The MeOH soluble fraction was fractionated using on a Sephadex LH-20 column and eluted with MeOH:DCM (4:1 v/v) into multiple small fractions. Fractions were spotted using thin layer chromatography (Alumina backed Silica, DCM:MeOH, 9:1, UV 254 nm and *p*-anisaldehyde stain) and combined based on common appearance in the UV and stain visualization methods (A, B, C, D). Each fraction obtained from Sephadex LH-20 was subjected to ^{1}H-^{15}N HSQC NMR Experiment, those fractions with the signal were then subjected to the [1]H-^{15}N HQSC-TOCSY experiment. The fraction B (53.6 mg, brown oil) was then

subjected to Si gel flash chromatography using a step-gradient whereby the piperazic-acid diagnostic NMR signature was seen in the final MeOH wash. This MeOH soluble fraction (18 mg) was then fractionated via C18 Sep-Pak (2 g) with step-gradients of 1:3, 1:1, 3:1, and 1:1 ACN/H_2O with the fractions C (1:1 H_2O:ACN v/v; 1.2 mg; yellowish oil) and D (1:3 H_2O:ACN v/v; 3.1 mg; yellowish-brown oily solid) containing the Piz NMR signal of interest. These fractions were moved to final purification through reversed-phase HPLC using an InertSustain C_{18}, 5 µm, 10 x 250 mm column. After HPLC purification, however, there proved to be too little sample for structure elucidation.

Scale-up of *Streptomyces incarnatus* NRRL 8089 and isolation of dentigerumycins F (4.2) and G (4.1) and incarnatapeptins A (4.3): To obtain more material for isolation of the Piz compounds, *Streptomyces incarnatus* NRRL 8089 was cultured on 58 L volume of ISP4 medium with added filter sterilized unlabelled L-Orn (4.5 g / 2 L) prepared as 160 pans of solid agar and grown at rt for 14 d. When the majority of spores turned a light grayish-pink, the agar media, bacterial mycelia and spores were cut into small squares and extracted with EtOAc (3x). The resulting crude was subjected first to ^{1}H-^{15}NHSQC and then ^{1}H-^{15}NHSQC-TOCSY experiments to obtain weak nitrogen-proton correlations. Concurrently, *S. incarnatus* NRRL 8089 was cultured on an additional two pans of ISP4 agar with added filter-sterilized ^{15}N$_2$-L-Orn prepared as above (final volume/tray 400 mL), harvested and extracted with EtOAc (3x). The extracts were combined and concentrated under reduced pressure. The resulting crude extract was analyzed by ^{1}H-^{15}N HSQC and ^{1}H-^{15}N HSQC-TOCSY NMR experiments to obtain the nitrogen-proton correlations. This material contained the nitrogen signal required for NMR observation throughout subsequent isolation steps. To obtain enough material for isolation, this crude material containing labelled piperazic-acid residues was then combined with the crude material

of the larger growth of *Streptomyces incarnatus* NRRL 8089 grown with unlabelled L-Orn. The combined EtOAc extracts were partitioned between H_2O (750 mL) and hexanes (250 mL x 3). The resulting aqueous extract was partitioned with EtOAc (4 x 300 mL). No nitrogen-proton correlations of interest were observed in the waxy yellow hexanes soluble material and it was set aside. The EtOAc soluble material was triturated with cold MeOH (~5 mL) to give a white precipitate and a brown liquid which, when concentrated *in vacuo* gave a brown oil. The white ppt and the brown oil were subjected to the piperazic-acid diagnostic nitrogen experiments with the nitrogen-proton correlations of interest observable only in the brown oil. This MeOH soluble material (900 mg; brown oil) was then chromatographed on SephadexTM LH-20 with 4:1 (v:v) MeOH/CH_2Cl_2 as eluent. The resulting test tube fractions were spotted using thin layer chromatography (Alumina backed Silica, DCM:MeOH, 9:1, UV 254 nm and *p*-anisaldehyde stain) and combined based on common appearance in the UV and stain visualization methods to give five fractions (A, B, C, D, E) which were subjected to the piperazic-acid diagnostic nitrogen experiments. The fraction labelled B (246.5 mg ; brown oily solid) contained the nitrogen-proton correlations and spin-system of interest. This fraction was then fractionated via C_{18} flash chromatography, where the NMR signature was seen in the fractions eluting with 2:3 and 5:6 ACN/H_2O. To further optimize the C_{18} flash chromatography, a small portion of fraction B was then fractionated using a C_{18} Sep-Pak (2 g) column with step-gradients of 1:4, 21:39, 2:3, 5:6 and 1:1 ACN/H_2O where it was determined via ^{1}H-^{15}N HSQC experiments that the Piz-Signal was eluting in H_2O/ACN (2:3 and 5:6) with no signal present in H_2O/ACN (1:1). The remaining EtOAc portion was then dry-loaded onto a large C18 flash column (30 mm x 20 cm) containing 40 g of C_{18} gel and slowly eluted using increasing amounts of ACN (1:4, 21:39, 2:3, 5:6 and 1:1 ACN/H_2O v/v). The fractions C (42.2 mg; 2:3 H_2O:ACN v/v) and D (22 mg; 5:6 H_2O:ACN v/v)

demonstrated the Piz signal in ^{1}H-^{15}N HSQC. A portion of Fraction D was subjected to

LRESIMS and the rest was set aside. A portion of the fraction C was then subjected to reversed-

phase HPLC. Fractions were collected divided by 10 min time signatures as it was known

through TLC that these compounds had no significant UV chromophore. Subsequent ^{15}N NMR

analysis with the ^{1}H-^{15}N HSQC experiment demonstrated loss of the Piz signal until all the

fractions were recombined. The separation was shown to require a low PH buffer to prevent

significant tailing and dragging across the column. With the addition of 0.01% TFA to the water

solvent with a solvent system of 1:1 ACN/H$_2$O, clear HPLC peaks were obtained with the peaks

labelled 11, 13, 14, and 16 demonstrating the Piz signal in the ^{1}H-^{15}N HSQC experiment

(**Scheme 3.2**). Thus, pure samples of dentigerumycin F (**4.2**) and G (**4.1**) and incarnatapeptin A

(**4.3**) were obtained from the fraction eluting with 2:3 ACN/H$_2$O via purification with reversed-

phase HPLC using an InertSustain C$_{18}$, 5 µm, 10 x 250 mm column with 1:1 ACN/H$_2$O with

0.05 % (v/v) TFA as eluent.

Chapter 4: Piperazic acid-containing Metabolites from *Streptomyces Incarnatus* NRRL 8089

4.1   Chapter Overview

This chapter will describe the structure elucidation of the novel Piz-containing natural products dentigerumycin G (**4.1**), dentigerumycin F (**4.2**), and incarnatapeptin A (**4.3**) isolated from *Streptomyces incarnatus* NRRL 8089 utilizing the Piz-targeted Nitrogen-NMR guided methodology described in **Chapter 3**. Incarnatapeptin A is a linear acid that contains a unique bicyclic ring that has not previously been seen in natural products. Dentigerumcyin F and G are linear peptides that are related to dentigerumycin C and D. Since these natural products all contain three piperazic acids, a brief review of other peptides containing three piperazic acid residues will be presented. Additionally, an overview of the unusual equilibrium observed in the oxane rings of dentigerumycin F and G will be presented. A biosynthetic proposal for the pathway that leads to both the incarnatapeptins and dentigerumycin F and G will be introduced. Finally, the antimicrobial assay results for incarnatapeptin A, dentigerumycin F and G will be discussed, while the bioactivity of incarnatapeptin A and incarnatapeptin B (**4.4**) in prostate cancer cytotoxicity assays will be examined.

4.2   19-Membered Depsipeptides with Three Piperazic Acids

Genome mining identified only two Piz residues in the Piz-containing natural products produced by *Streptomyces incarnatus* NRRL 8089, but ^{1}H-^{15}N-HSQC nitrogen NMR data indicated the presence of three piperazic acids. Several natural products (**Figure 4-1**) that contain three piperazic acids are known, and a subset of them will be discussed below to provide context for the discovery research described in this chapter. The dentigerumycins are 19-membered cyclic depsipeptides (**4.5, 4.6 and 4.9**) or their linear peptide congeners (**4.7 and 4.8**). Dentigerumycin A (**4.5**), isolated from a *Pseudonocardia* strain, contains the amino acid

sequence L-Ala, 3-OH-L-Piz, D-Piz, *N*-OH- L-Ala, D-Piz, β-OH-Leu; dentigerumycins B (**4.6**), C (**4.7**) and D (**4.8**), all isolated from a *Streptomyces* strain, contain the core amino acid sequence, L-Piz, D-Piz, *N*-OH-Gly, D-Piz, β-OH-Leu. **4.6** has an additional amino acid, *N*-OH-L-*allo*-Thr, which is used to form the cyclic molecule. Dentigerumycin E (**4.9**), isolated from another *Streptomyces* strain, shares the amino acid sequence of dentigerumycin B save for the cyclizing *N*-OH-L-Thr, but it has a uniquely oxidized substituent on its oxane ring.[173]

Other classes of natural products that contain three Piz moieties are peptides (**4.10 and 4.11**) and 18-membered depsipeptides (**2.26, 4.12, 4.13**). The hytramycins (**4.10 and 4.11**), peptides isolated from a *Streptomyces* species, have an amino acid sequence of L-Piz, D-Val/Ile, L-Leu, D-Piz, *N*-Me-Ala, and D-Piz. Hytramycin's gene cluster has not been identified. The gerumycins (**2.26, 4.12, 4.13**), 18-membered depsipeptides produced by a *Streptomyces* species, have an amino acid sequence of α-Me-L-Ser, 3-OH L-Piz, 3-Cl-D-Piz, L-Ala, (3-OH) D-Piz. In 18-membered depsipeptides (**Figure 2-3**), it is presumed that the ring closure occurs through the loading of a final hydroxy acid monomer in place of an amino acid; in the gerumycins, the final loading of lactic acid leads to **2.26** and **4.12**, whereas the loading of 2-hydroxybutanoic acid leads to **4.13**.[173]

Figure 4-1. Natural products that contain three piperazic acids.[173]

Dentigerumycin A (**4.5**)

Dentigerumycin C (**4.7**) $R_1 = OH$
Dentigerumycin D (**4.8**) $R_1 = NH_2$

Dentigerumycin B (**4.6**) $(R_1 = $⋮⋯OH, $R_2 = CH_3)$

Dentigerumycin E (**4.9**) $(R_1 = $⋮⋯OH, $R_2 = COOH)$

Hytramycin V (**4.10**) $R_1 = CH_3$
Hytramycin I (**4.11**) $R_1 = CH_2CH_3$

Gerumycin A (**2.26**) $R_1 = R_2 = H$
Gerumycin B (**4.12**) $R_1 = CH_3, R_2 = H$
Gerumycin C (**4.13**) $R_1 = H, R_2 = OH$

4.3   Structure Elucidation of Piperazic Acid-Containing Metabolites from *Streptomyces Incarnatus* NRRL 8089

4.3.1 Structure Elucidation of Dentigerumycin G (4.1)

The structure elucidation of dentigerumycin G (**4.1**) will be described first, as it was the

first structure from the isolated piperazic acid-containing nonribosomal peptides to be solved due

to its greater abundance. Although its structure elucidation was complicated due to the presence

of hemiketal/dehydrated forms, the abundance of the material (~1 mg) that was available for

obtaining a dataset gave strong enough NMR signals to assess which resonances belonged to

which form. Analyzing the samples in various deuterated solvents aided the characterization, as

did assessing both high-res and low-res mass spectra in varying conditions.

Figure 4-2. ¹H spectrum for dentigerumycin G (4.1) recorded after sample was in the fridge for ~1 week, in DMSO-d_6, recorded at 600 MHz.
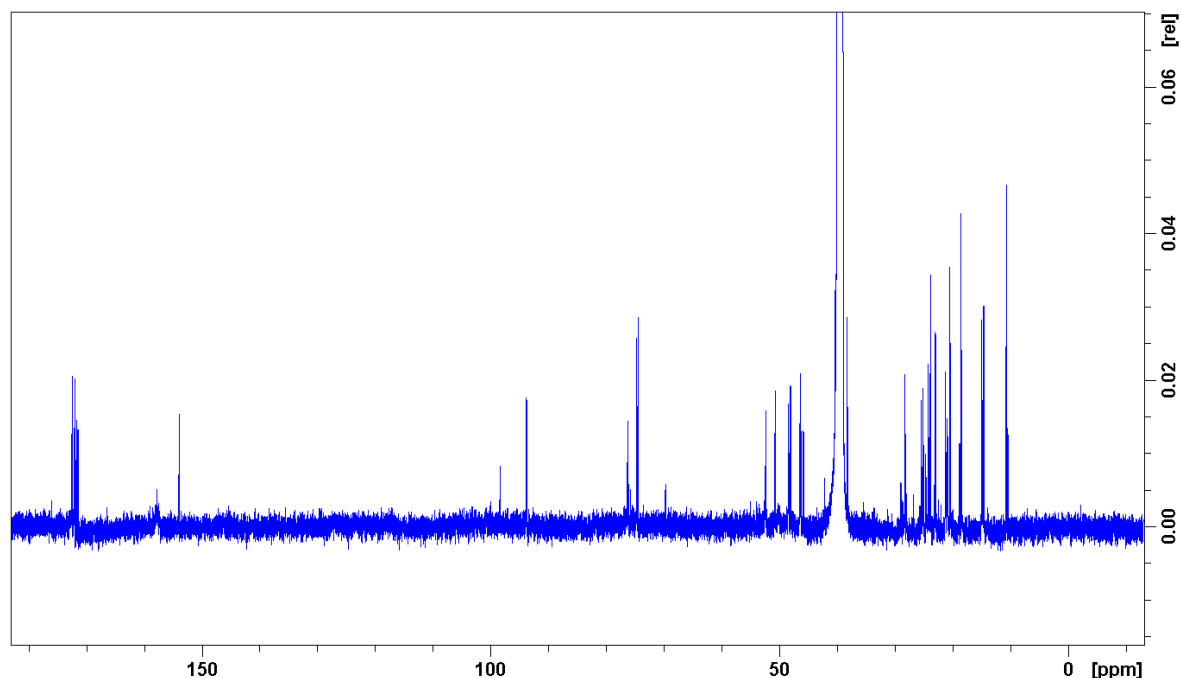
Figure 4-3. ^{13}C **spectrum for dentigerumycin G (4.1) demonstrating the large numbers of carbons in the compound, recorded in DMSO-d_6 at 150 MHz.**

Dentigerumycin G (**4.1**) gave a $[M + Na]^+$ ion in the HRESIMS at *m/z* 807.4237 appropriate for a molecular formula of $C_{35}H_{60}N_8O_{12}$ that also requires 10 sites of unsaturation. Analysis of the ^1H, ^1H-^1H gCOSY60, ^1H-^1H gTOCSY/^1H-^{13}C gHSQC/^1H-^{15}N gHSQC/^1H-^{13}C gHMBC/ ^1H-^{15}N gHSQC-TOCSY NMR data recorded in DMSO-d_6 revealed **4.1** as an acid that contains three Piz residues, an N-hydroxy alanine residue, a β-OH-leucine, along with an acyl side chain containing a tetrahydropyran ring bearing vicinal methyl and ethyl substituents in a trans-configuration. The ^1H NMR data showed a broad ^1H resonance at δ_H 13.26 (**Figure 4-2**) typical of a carboxylic acid, which led to the working hypothesis that this molecule was a linear peptide with a free carboxylic acid at its C terminus. Three exchangeable proton resonances at δ_H 4.72 (δ_N -301.1), 5.08 (δ_N -300.0), and 5.16 (δ_N -300.5) showed one-bond correlations to nitrogen resonances in the ^1H-^{15}N HSQC spectrum, and these were the signals that guided the isolation of

126

the compound (**Figure 4-4**). These Piz-like correlations were subsequently shown to be saturated

piperazic acid spin systems (**Figure 4-5**) in both the ^{1}H-^{15}N HSQC-TOCSY (**Figure 4-6**) and ^{1}H-

^{1}H TOCSY (**Figure 4-7 and 4-8**) spectra. In particular, δ_H 4.72 (NH-1) demonstrated

correlations into H-5a/b, H-4a/b, H-3a/b, and H-2a/b ; δ_H 5.16; (NH-4) showed correlations into

H-10a/b, H-9a/b, H-8a/b, and H-7a/b; while δ_H 5.08 (NH-6) had correlations into the methylene

protons H-18a/b, H-17a/b, H-16a/b, and H-15. Further evidence was seen in the HSQC spectrum

which showed that all methylene carbons adjacent to the Piz N-H's possessed the typical

deshielded carbon resonances of this position (see **Table 3-2**) in these moieties: C-5 (δ_C 46.7), C-

10 (δ_C 46.2) and C-18 (δ_C 45.3).
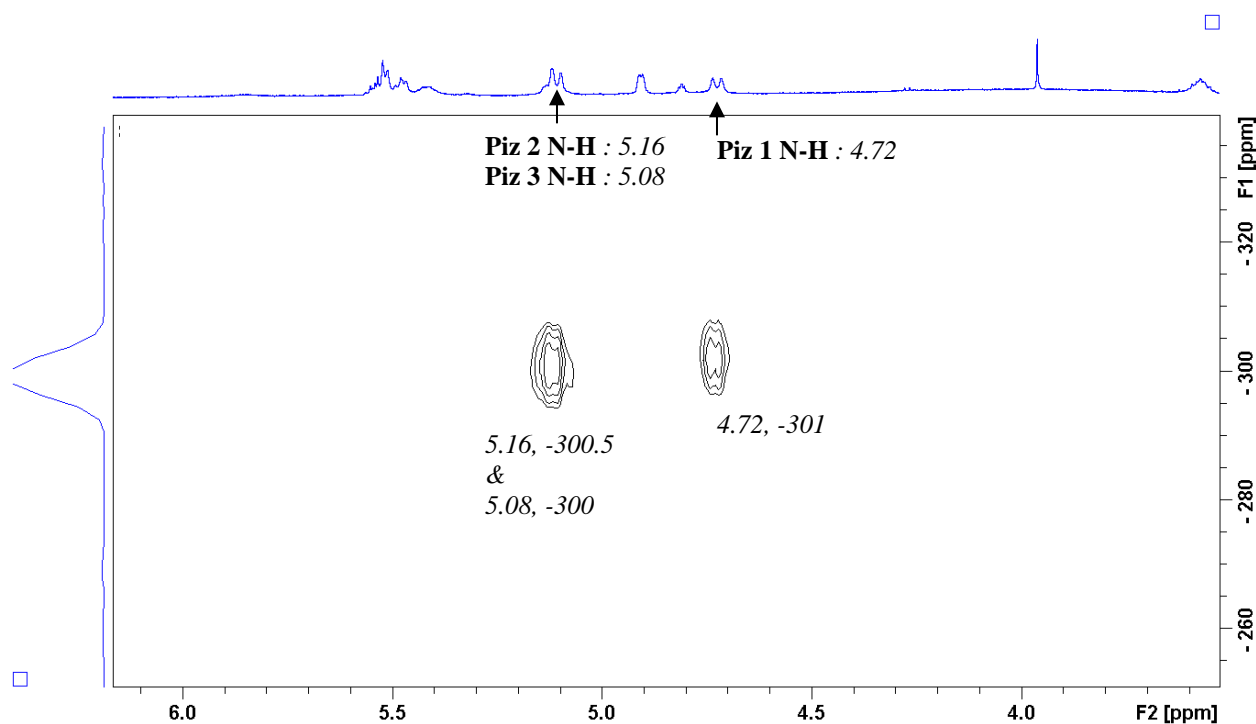
Figure 4-4. ^{1}H-^{15}N HSQC NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-d_6. N-H correlations which identified 4.1 as a Piz-containing molecule.

127
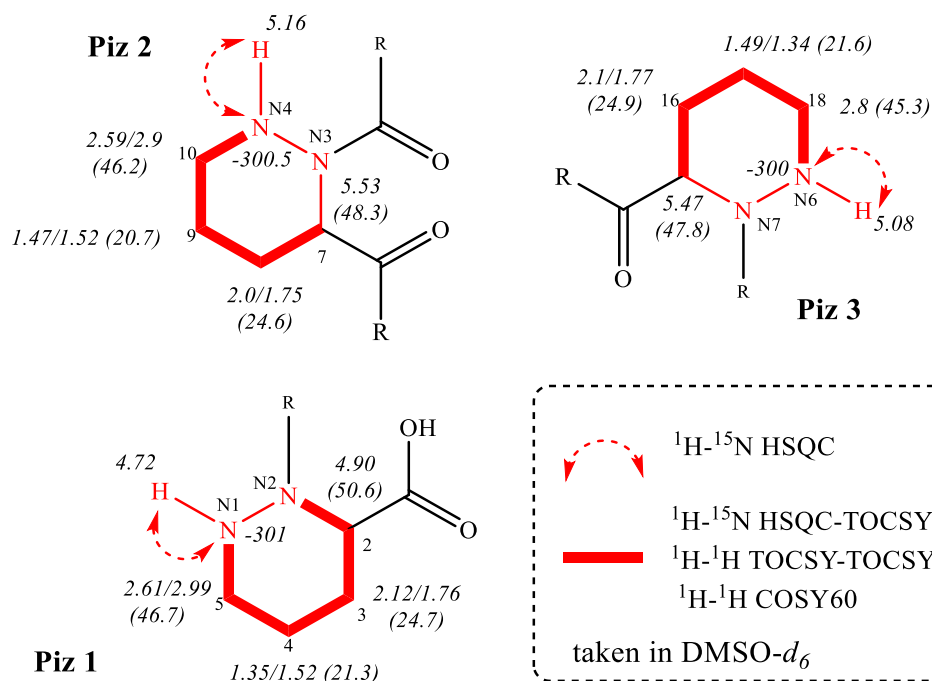
Figure 4-5. Piz NMR spin systems for dentigerumycin G (4.1). These spin systems, seen in the ^{1}H-^{15}N HSQC-TOCSY NMR experiment were the second data point to confirm 4.1 as a Piz containing molecule while also providing a handle for structure elucidation.
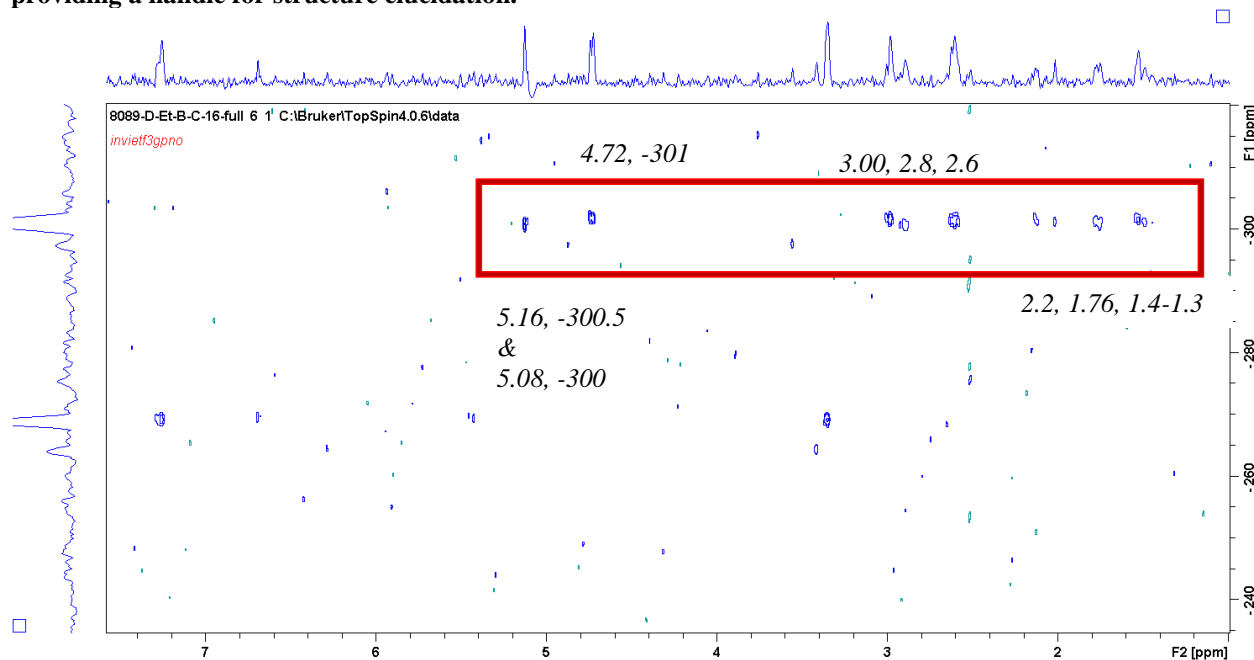
Figure 4-6. ^{1}H-^{15}N HSQC-TOCSY NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-d_6. Piz spin systems, which roughly matched experimental values for a typical Piz, outlined in red.
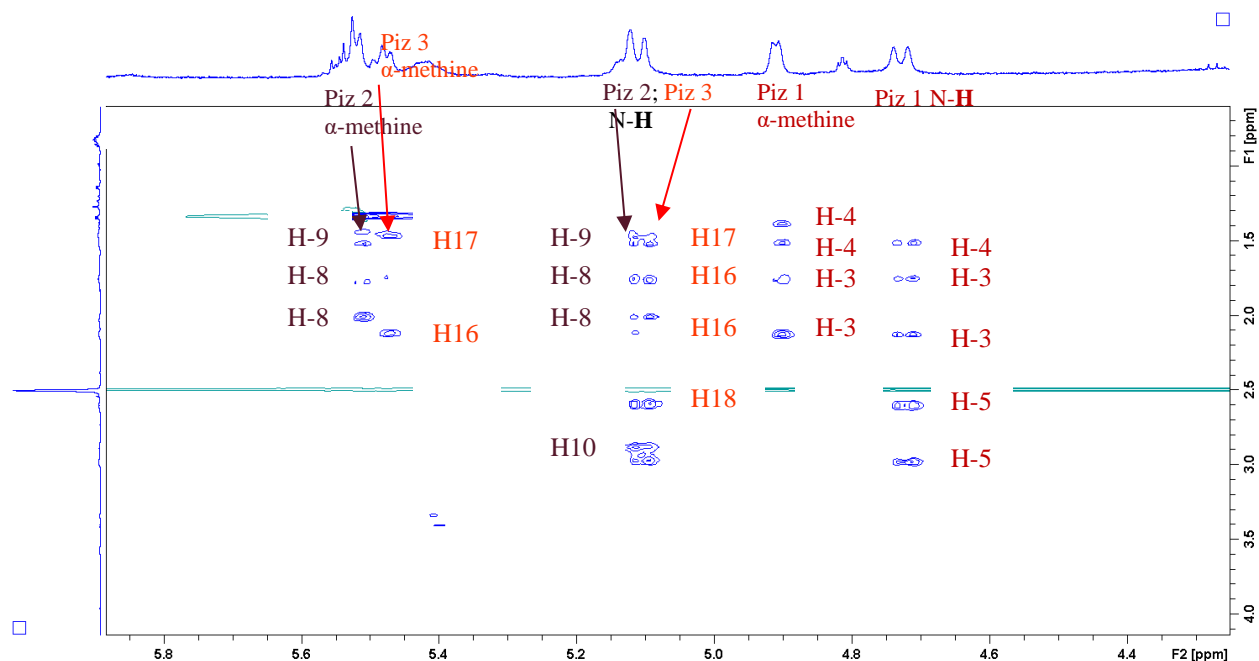
Figure 4-7. Expansion of ^{1}H-^{1}H TOCSY NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-d_6. Demonstrated is the Piz 1, 2 and 5 spin systems (Figure 4-5), from the N-H proton on one end of the spin system and from the α-methine on the other end of the spin system.
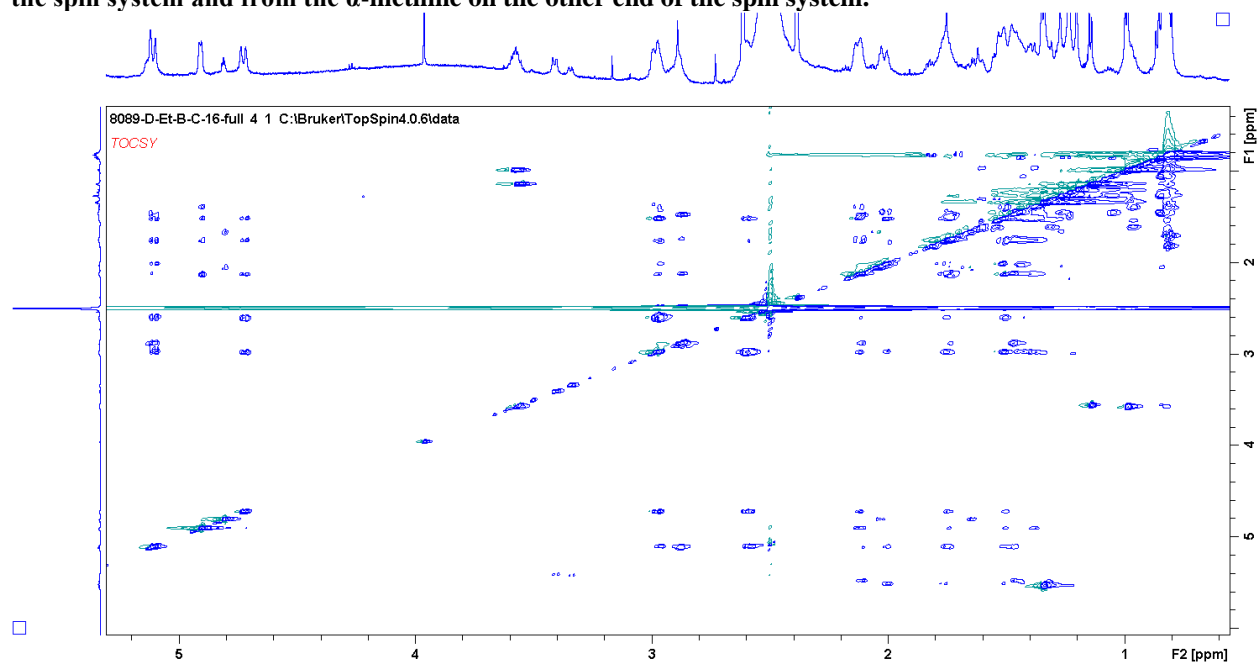
Figure 4-8. Full TOCSY NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-d_6.
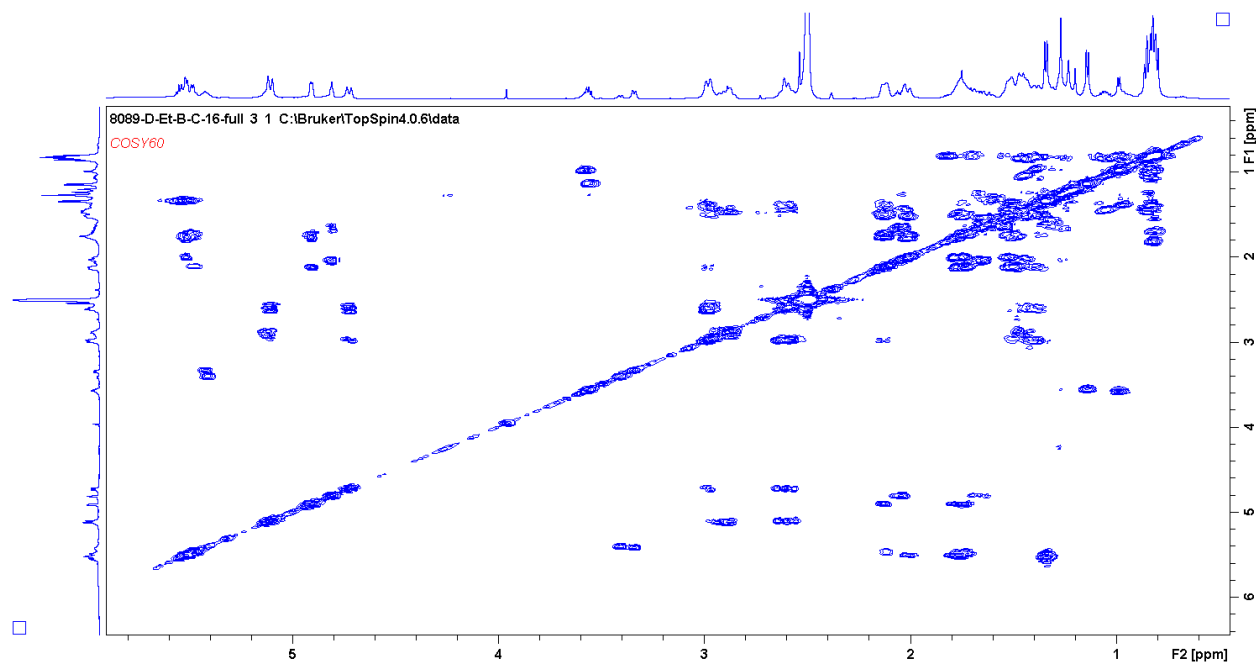
Figure 4-9. COSY60 NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.
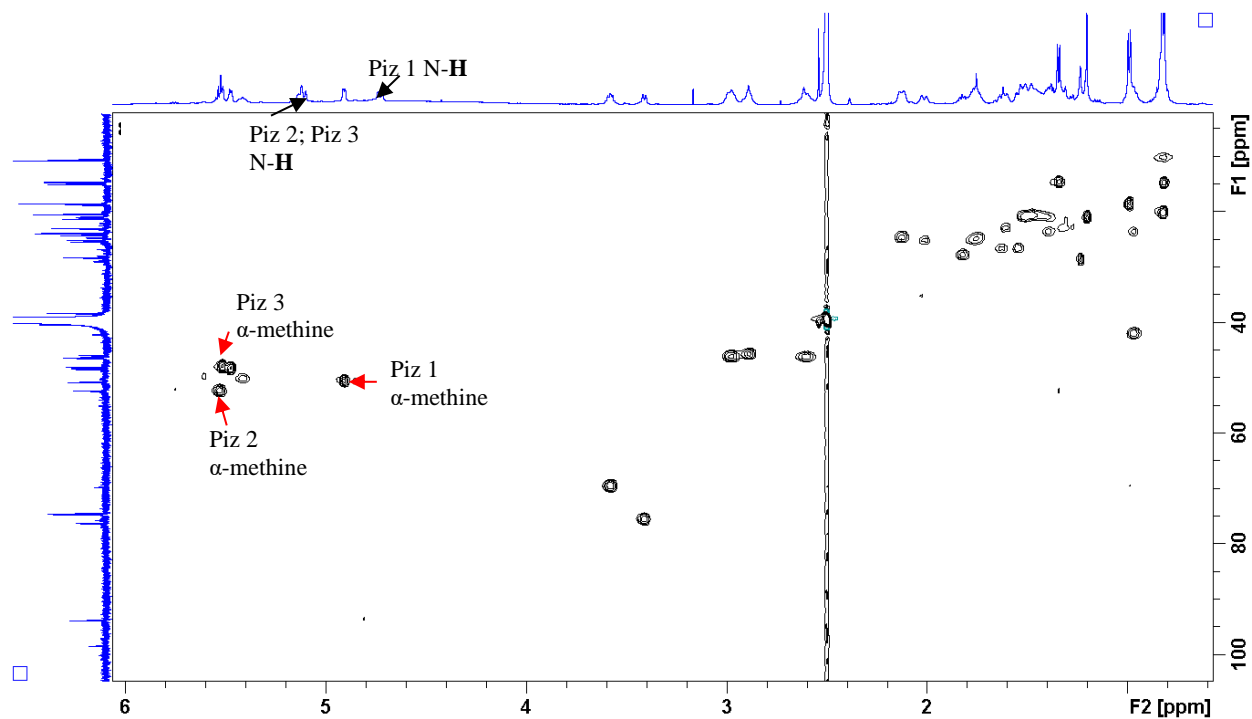
Figure 4-10. ¹H-¹³C HSQC NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*. This is demonstrating the unusually deshielded methines of Piz 2 and Piz 3. Also seen is the lack of correlation of the N-H protons into a carbon.
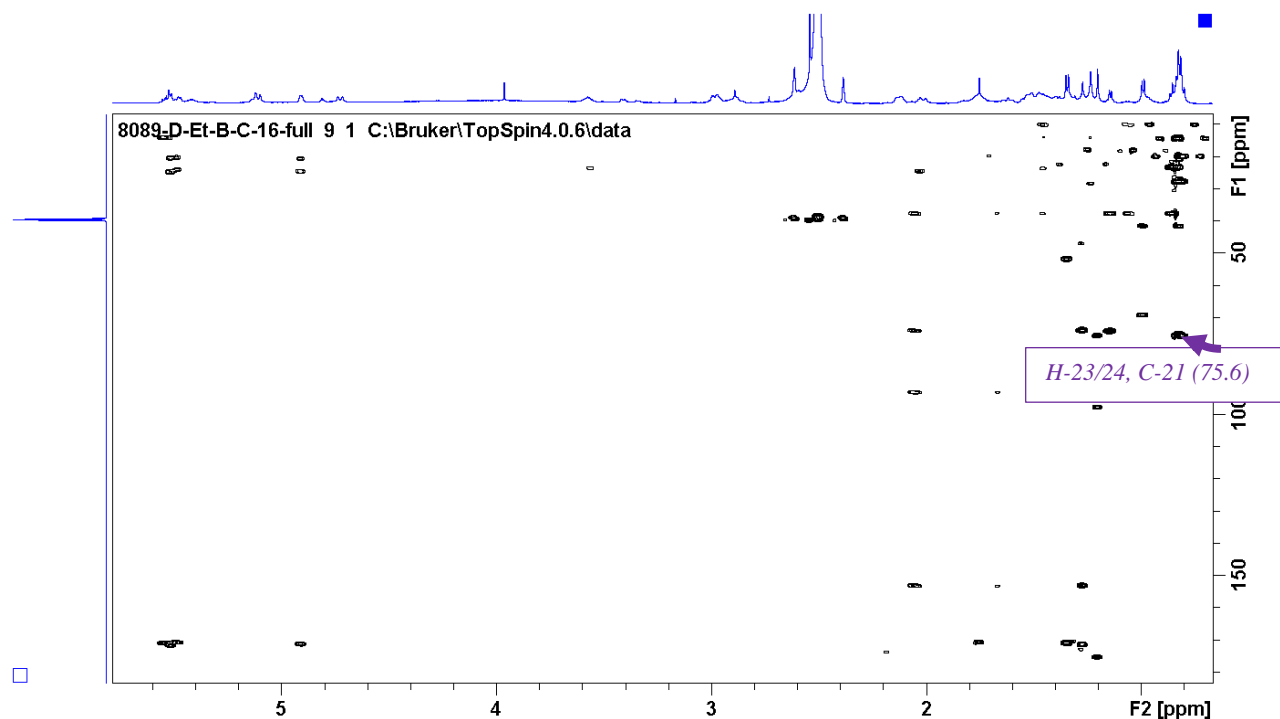
Figure 4-11. ¹H-¹³C HMBC NMR spectrum of dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.

In silico analysis (antiSMASH predictions, **Table 2-4**) predicted an N-hydroxylated alanine, which was confirmed through ¹H-¹H gCOSY and ¹H-¹H TOCSY correlations between an α-methine proton at δ_H 5.53 (H-12) and a methyl doublet at δ_H 1.34 (H-13) both of which showed HMBC correlations into a carbon resonance at δ_C 172.01 (C-11), as seen in NMR spectra expansions in **Figure 4-13**. The nitrogen-attached hydroxyl was suspected due to a broad proton resonance at δ_H 9.55, which is characteristic of a hydroxymate proton resonance. This was corroborated via ¹H-¹⁵N LRHMQC cross peaks between H-13 (δ_C 1.34) and N-5 (δ_N -199.9) (**Figures 4-14**). The amino acid sequence of this free acid was confirmed as Piz, Piz, *N*-OH-Ala, Piz (**Figure 4-12**) through a combination of HMBC correlations between protons and the peptide carbonyl resonances and tROESY resonances between nitrogen-attached Piz protons and the adjacent α-methine protons (**Figure 4-15 & 4-16**).

131
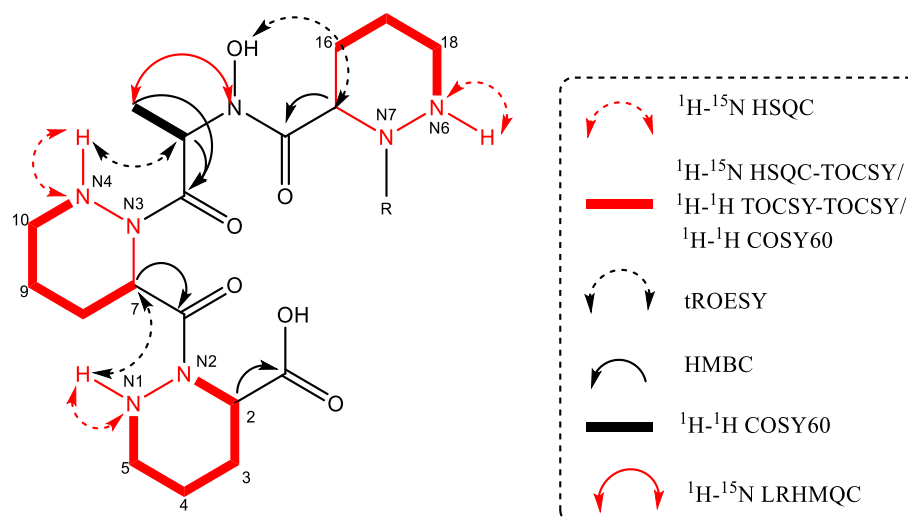
Figure 4-12. Connecting the core amino acids for dentigerumycin G.
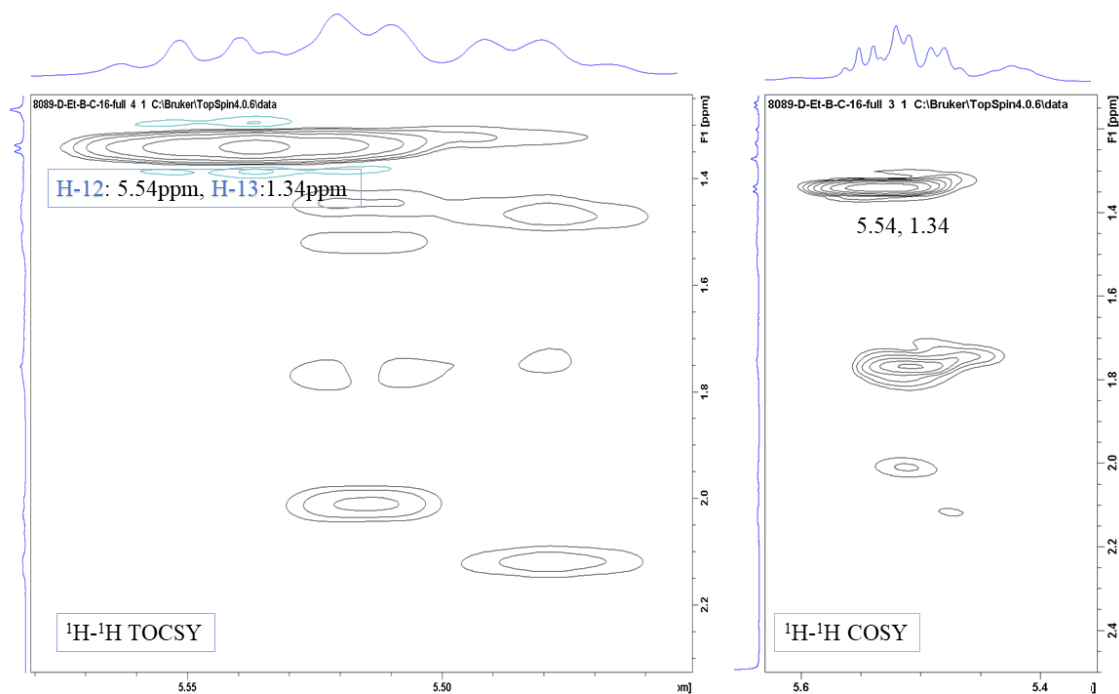
Figure 4-13. TOCSY and COSY60 spectra demonstrating the correlations between *N*-OH-Ala's methyl and α-methine protons leading to the assigned structure of 4.1. Recorded at 600 MHz in DMSO-*d₆*.
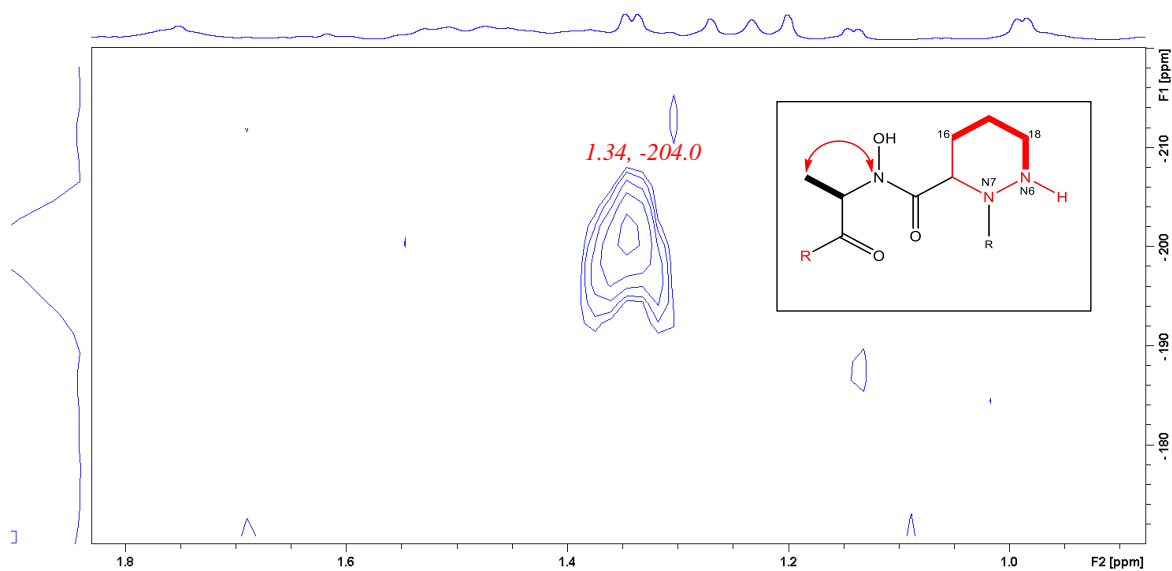
Figure 4-14. ^{1}H-^{15}N LRHMQC spectrum demonstrating the correlations between *N*-OH-Ala's methyl protons and the hydroxymate nitrogen leading to the assigned structure of 4.1. Recorded at 600 MHz in DMSO-*d$_6$*.
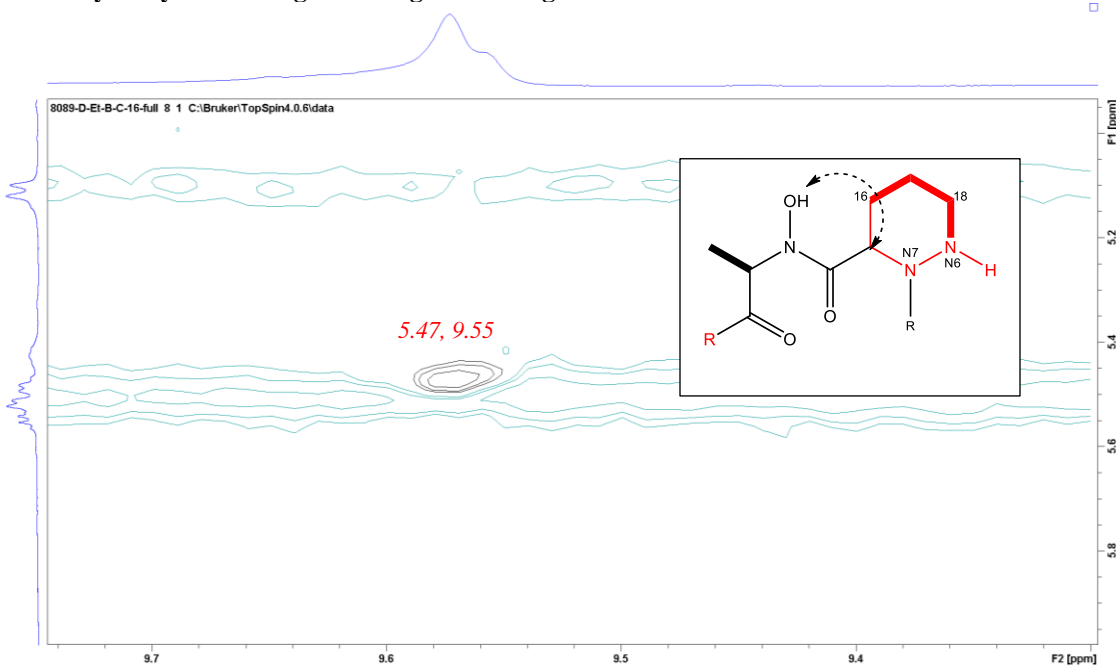
Figure 4-15. tROESY correlations between the proton of the nitrogen-attached hydroxyl proton and Piz 3's α-methine proton leading to the assigned structure of 4.1. Recorded at 600 MHz in DMSO-*d$_6$*.
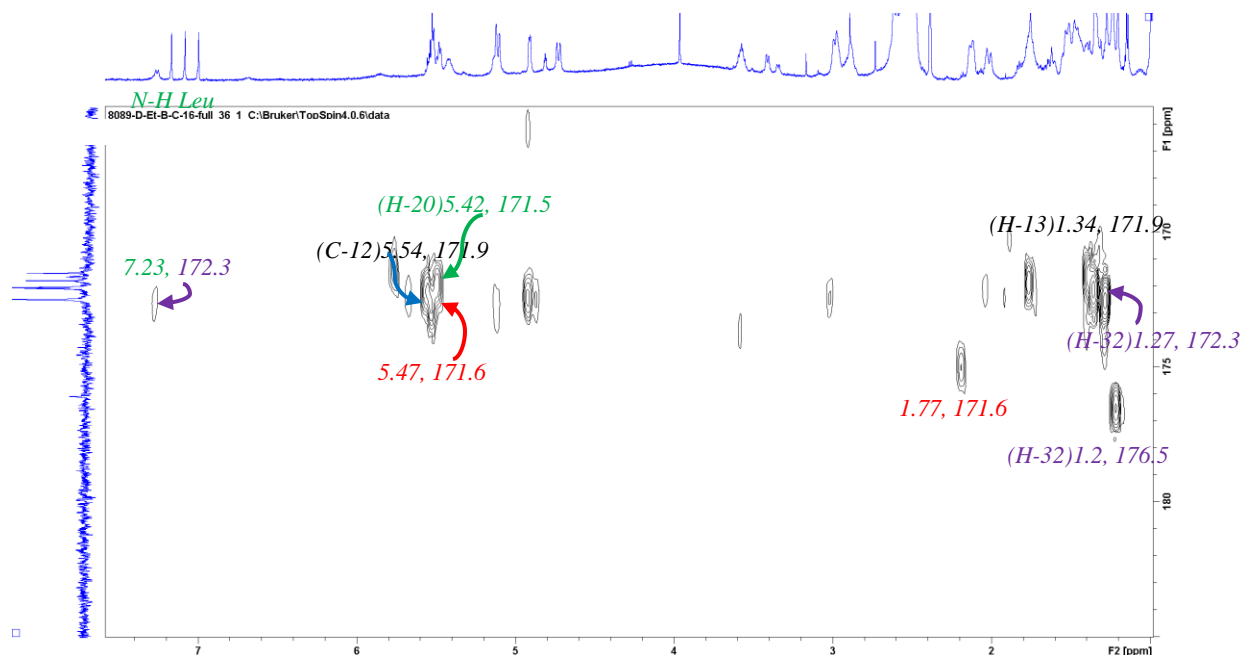
Figure 4-16. ^{1}H-^{13}C HMBC spectrum demonstrating proton-carbonyl correlations that helped connect the amino acids thus leading to the assigned structure of 4.1. Recorded at 600 MHz in DMSO-d_6.
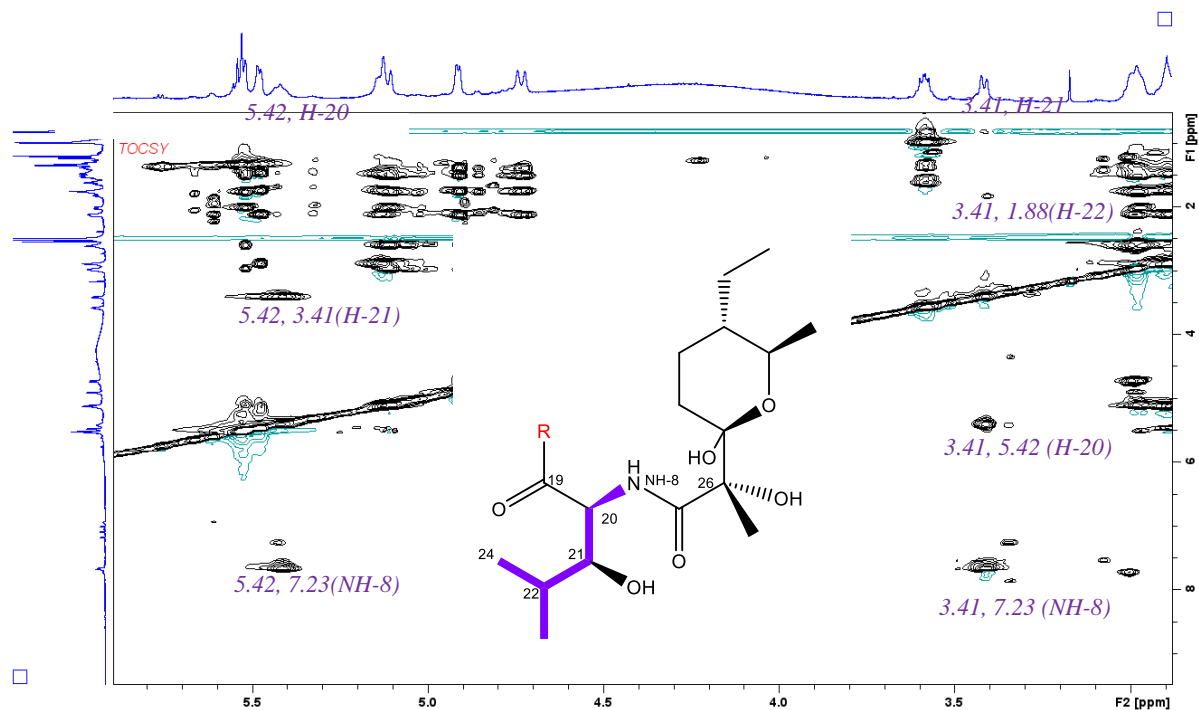
Figure 4-17. ^{1}H-^{1}H TOCSY correlations between the γ-methine (H-22), β-methine (H-21), α-methine (H-20) and nitrogen proton (NH-8) of β-OH-Leu which helped lead to the assigned structure for 4.1. The H-21/H-22 signal was very weak, so the spectrum is shown deep into the contours. Recorded at 600 MHz in DMSO-d_6.

Further analysis of the 2-D data revealed a β-OH-Leucine with characteristic methyl doublets at C-23 (δ_C 14.7, δ_H 0.82) and C-24 (δ_C 20.2, δ_H 0.82) showing HMBC correlations from $\delta_{H(C-23/24)}$ 0.82 into C-21(δ_C 75.6) (**Figure 4-11**). The ^{1}H-^{1}H TOCSY spectrum of a cold sample (**Figure 4-17**) revealed 3 adjacent methines at H-20 (δ_H 5.42), H-21(δ_H 3.41), and H-22(δ_H 1.69) as well as a cross peak from H-20 to an amide proton at δ_H 7.29. Although no hydroxyl proton was observed in this data, the deshielded methine carbon resonance at δ_C 75.6 for C-21 supported this assignment as did later spectra obtained from **4.1** isolated without TFA (**Appendix**). Such a β-OH-Leucine is conserved across the majority of azinothricin type molecules.[173] Thus, the presence of this amino acid, along with the three Piz amino acids discussed earlier, suggests that **4.1** was the product of the *Streptomyces incarnatus* NRRL 8089 NRPS-PKS hybrid gene cluster found using the KtzT probe. Furthermore, despite key gene organization differences between *Streptomyces incarnatus* NRRL 8089 's gene cluster and dentigerumycin A's *Pseudonocardia* producer, (**Figure 3-22 and Table 3-9**) **4.1** was seen to be structurally related to dentigerumycin A (**4.5**) except for a missing cyclizing alanine. However, as the ^{1}H NMR data demonstrated evidence that **4.1** was a free acid, this molecule was presumed to be a linear molecule in a similar fashion to dentigerumycin C (**4.7**)and D (**4.8**).[237] In fact, this molecule proved to be a constitutional isomer of dentigerumycin C[237] (**Figure 4-1**), sharing the molecular formula $C_{35}H_{60}N_8O_{12}$, and the calculated molecular mass of 807.4228.
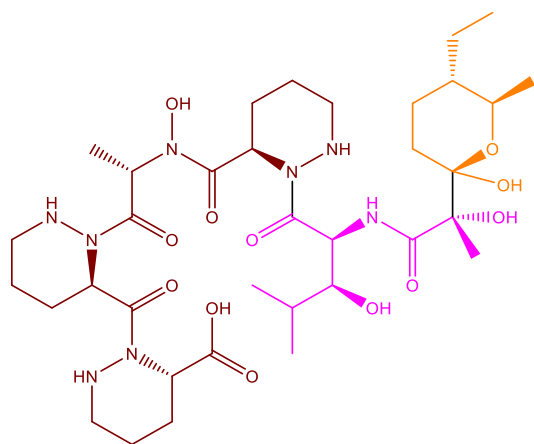
Fragment A : $C_{18}H_{30}N_7O_6$ Calculated mass: 440.2258
Fragment B : $C_9H_{16}NO_4$ Calculated mass: 202.1079
Fragment C : $C_8H_{15}O_2$ Calculated mass: 143.1072

Figure 4-18. Building up fragments of dentigerumycin G (4.1) to help confirm the structure of the PKS portion.

Given the close association to the linear dentigerumycins, it was essential to establish the

structural differences. After assigning Fragment A via NMR correlations, it became apparent that

the first point of difference from dentigerumycin C (**4.7**) was in the core amino acids for which

the molecular formula of the fragment is seen above in **Figure 4-18.** Fragment A was calculated

to have the molecular formula of $C_{18}H_{30}N_7O_6$, while the comparable fragment in dentigerumycin

C has a molecular formula of $C_{17}H_{28}N_7O_6$ (**Figure 4.1**).[237] This left seventeen carbons, thirty

protons, one nitrogen, and six oxygens that were yet to be assigned within the structure. Analysis

of the acyl side chain was complicated by the hemiketal/dehydro equilibrium (discussed in

Subsection 4-3-3), which developed within the data collection time frame. However, the

structure of the molecule was built by analysis of the NMR data that described the standard β-

OH-Leu and was extended by evidence of a hydroxyl and methyl substituent attached to a

quaternary carbon as seen in **Figure 4-19**. The methyl singlet at H-32 (δ_C 21.2, δ_H 1.20) showed

clear HMBC correlations into deshielded quaternary carbons at δ_C 76.9 (C-26) and δ_C 98.0 (C-

27), both of whose resonances indicated the likely attachment to electronegative oxygen[217,238] (**Figure 4-20**).

Additionally, H-32 demonstrated HMBC correlations into a carbonyl carbon at δ_C 172.3 (C-25) (**Figure 4-20**), which was also seeing ^2J correlations from the N-H (δ_H 7.23) of the β-OH-Leucine thus helping to connect the NRPS and PKS portions of the molecule. In the first set of data recorded, this H-32 proton singlet was captured as two peaks (A: δ_H 1.20 and B : δ_H 1.27). For the methyl singlet peak at δ_H 1.20 there appeared to be very low relative integration, despite strong HMBC resonances into multiple quaternary carbons. By capturing the proton spectra at varying times throughout the data collection and at different initial temperatures, it became clear that the relative integration for the proton at δ_H 1.20 was changing. This will be discussed in greater detail in **Subsection 4-3-3**. Furthermore, although the hydroxyl proton was not visible in the data, the deshielded carbon resonances of C-26 and C-32 strongly suggested the presence of a hydroxyl substituent. The singlet coupling pattern of the H-32 proton also provided evidence that there were no coupled protons attached to C-26. Fragment B seen in **Figure 4-18** was thus constructed as a partial molecular formula of $C_9H_{16}NO_4$. This left eight carbons, fifteen protons, and two oxygens that were yet to be assigned within the structure and as such likely contained within the remainder of the molecule. With this information in hand, it was possible to elucidate the remainder of **4.1**.

Through analysis of the 2D data, it became evident that the remaining $C_8H_{15}O_2$ were forming a disubstituted oxane ring. The oxane ring was revealed through gCOSY60 and ^{1}H-^{1}H TOCSY correlations from the H-31 methine (δ_H3.58) to the shielded methine proton at H-30 (δ_H0.97) and extended methylene and methyl protons of H-33 (δ_H1.39, 0.96) and H-34 (δ_H 0.82). The methine proton at H-30 was distinctly attached to a deshielded carbon resonance (δ_C 41.9)

anchoring the structure elucidation efforts despite the challenges associated with solving closely related spin systems appearing in the NMR spectra due to an apparent equilibrium between two forms of the molecule.
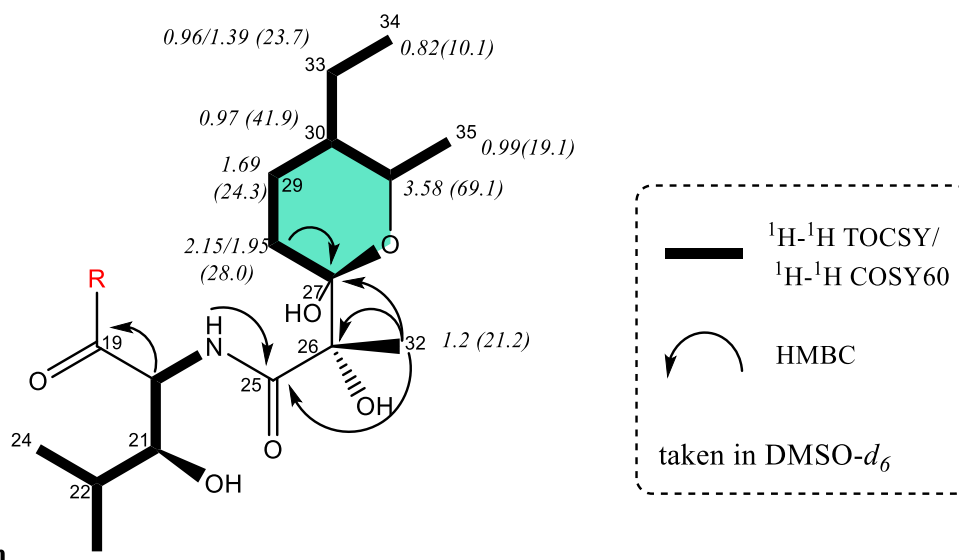
Figure 4-19. Correlations confirming the hydroxy-leucine amino acid and the PKS portion of 4.1.

The methyl doublet at H-35 (δ_H 0.99) was distinctive, demonstrating clear COSY60 correlations into H-31 as well as HMBC correlations into C-31 (δ_C 69.1). More challenging to confirm was the methyl group of the suspected ethyl substituent chain of H-33 and H-34 because the oxane ring substituent resonances were overlapped with the β-OH-Leu methyl doublets (H24/23 δ_H 0.82). The overlapping resonances made NMR assignment challenging. Further adding to this challenge was that these upfield resonances were also present for both a hydrated and dehydrated oxane ring, resulting in an exceedingly complex region of the NMR spectra. An initial TOCSY spectrum recorded on the sample demonstrated the complicating equilibrium (**Figure 4-21**). Subsequently, a TOCSY spectrum recorded immediately after the sample of **4.1** had been stored in the fridge for one week demonstrated that H-34, H-33, H-30, H-31 and H-29 were in a spin system (**Figure 4-22**). A TOCSY spectrum recorded after the sample sat at room

temperature for two weeks (**Figure 4-23**) demonstrates the re-equilibration of the molecule between a hemiketal and olefin form at C-27/C-28, leading, once again, to a complicated spectrum. HMBC correlations (**Figure 4-20**) between the deshielded C-30 (δ_C 41.92) and both the methine of H-31 (δ_H 3.58) and the methyl at δ_H 0.99 (H-35), along with the correlations between the deshielded C-31 (δ_C 69.00) and the methine multiplet of H-30 (δ_H 0.97) supported an adjacent nature of the oxane substitutions. 2D *J*-resolved spectra were recorded to confirm the coupling patterns (**Figure 4-24**) of H-34 and H-35. The H-35 methyl substituent was confirmed as a doublet in both forms, coupling to the methine proton of H-31. The H-34 methyl substituent was confirmed as a triplet coupling to the methylene protons of H-33. Thus, gCOSY60 and HMBC correlations, along with coupling patterns established through a *J*-resolved spectrum, demonstrated that the oxane ring had vicinal methyl and ethyl substituents.

This unique ring pattern differentiated **4.1** as a novel linear piperazic acid-containing natural product. 2D t-ROESY correlations between the H-31 methine proton (δ_H 3.58) and the methylene protons of H-33a/b (δ_H 1.39, 0.99) gave evidence for a trans relative configuration of the two sidechains. 2D t-ROESY correlations taken from data showing the hydroxyl chemical shifts (**Appendix B**) confirmed the hydroxyl at C-27 was on the same face of the six-membered ring as the methine proton at H-31 (δ_H 3.58). The absolute configuration of the molecule was assigned using the crystal structure of incarnatapeptin A (**4.3**) given that they are produced from the same biosynthetic pathway and have closely related NMR resonances for the amino acids and PKS-NRPS link, this will be discussed in **Section 4.4**.
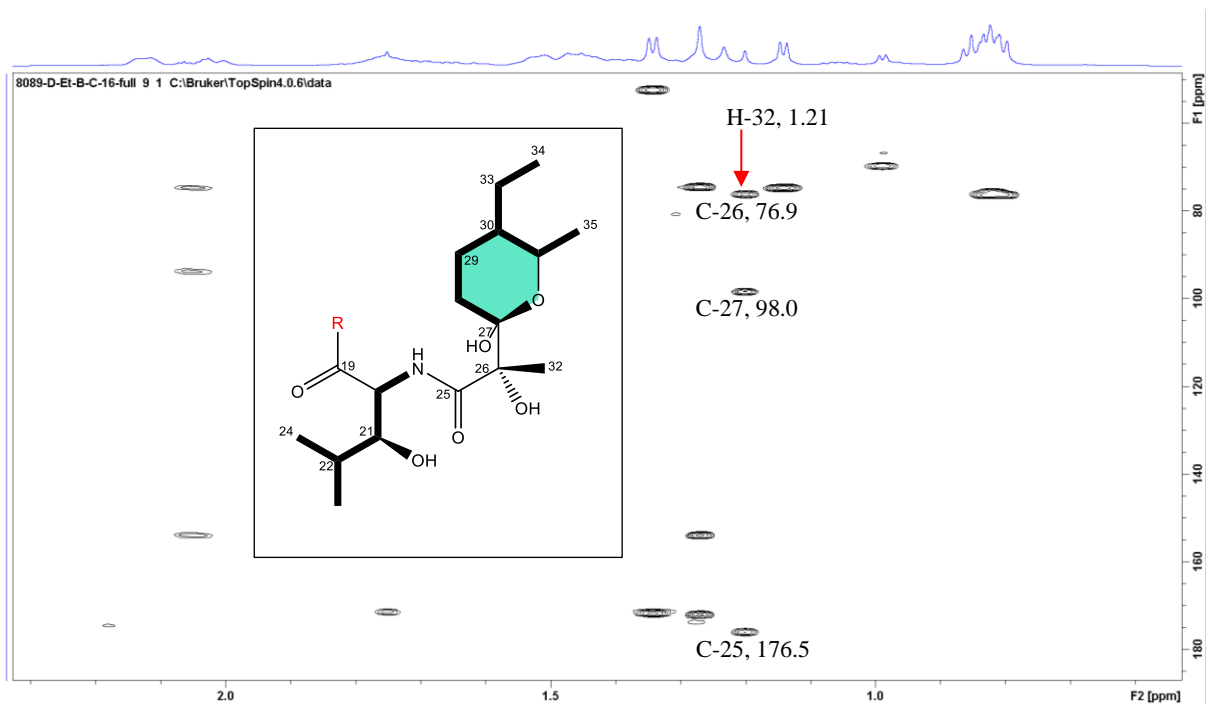
Figure 4-20. ^{1}H-^{13}C HMBC detailing correlations for the PKS linkage portion of 4.1. Recorded in DMSO-d_6, at 600 MHz.
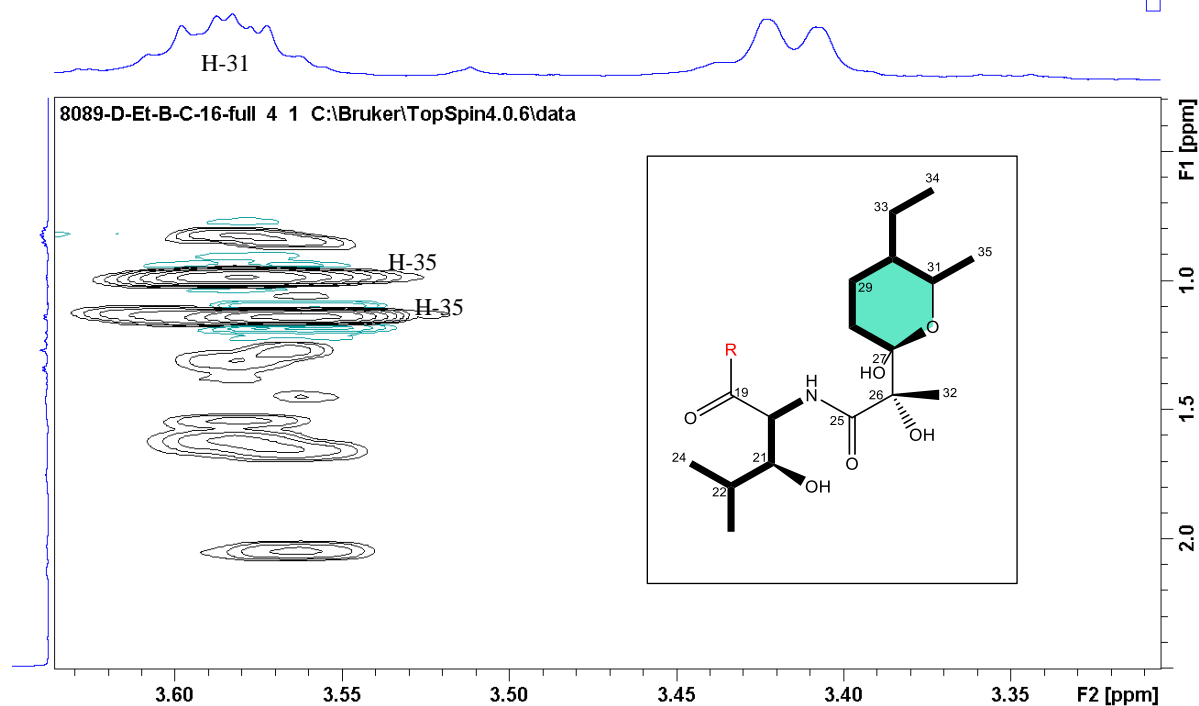
Figure 4-21. Initial ^{1}H-^{1}H TOCSY spectrum showing multiple signals due to an equilibrium in the PKS portion of 4.1, recorded in DMSO-d_6 at 600 MHz.
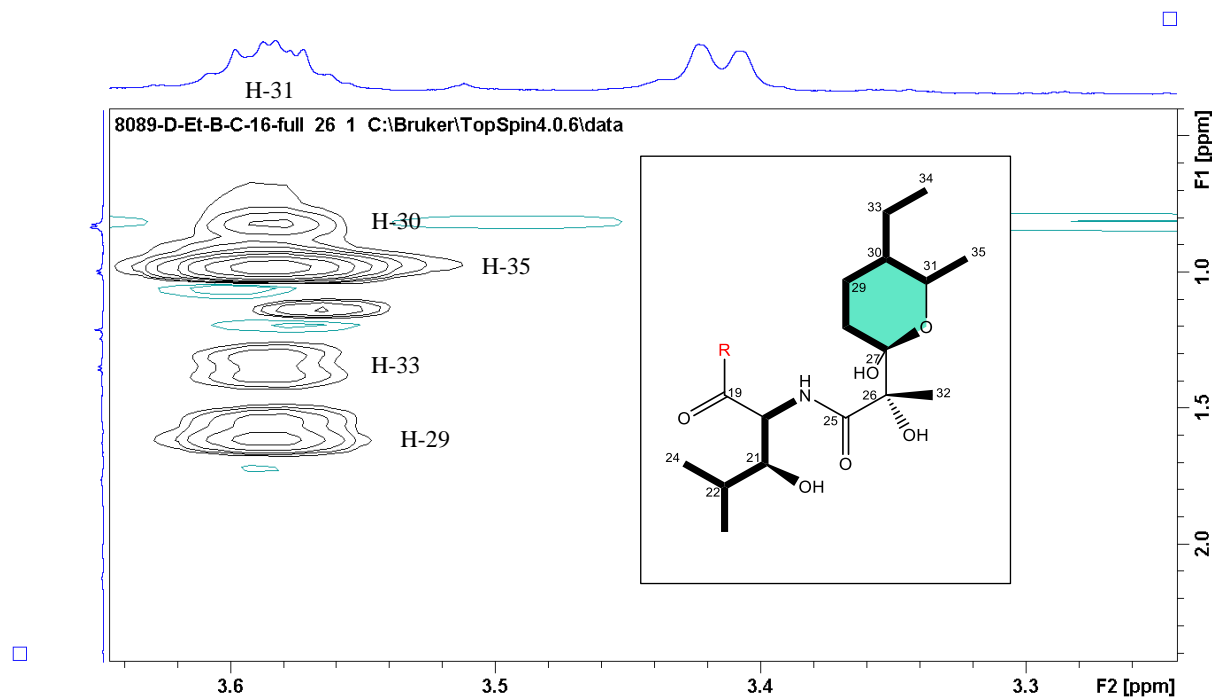
Figure 4-22. ^{1}H-^{1}H TOCSY demonstrating 4.1's disubstituted oxane ring, recorded on a sample sitting in the fridge for one week, recorded in DMSO-d_6, at 600 MHz.
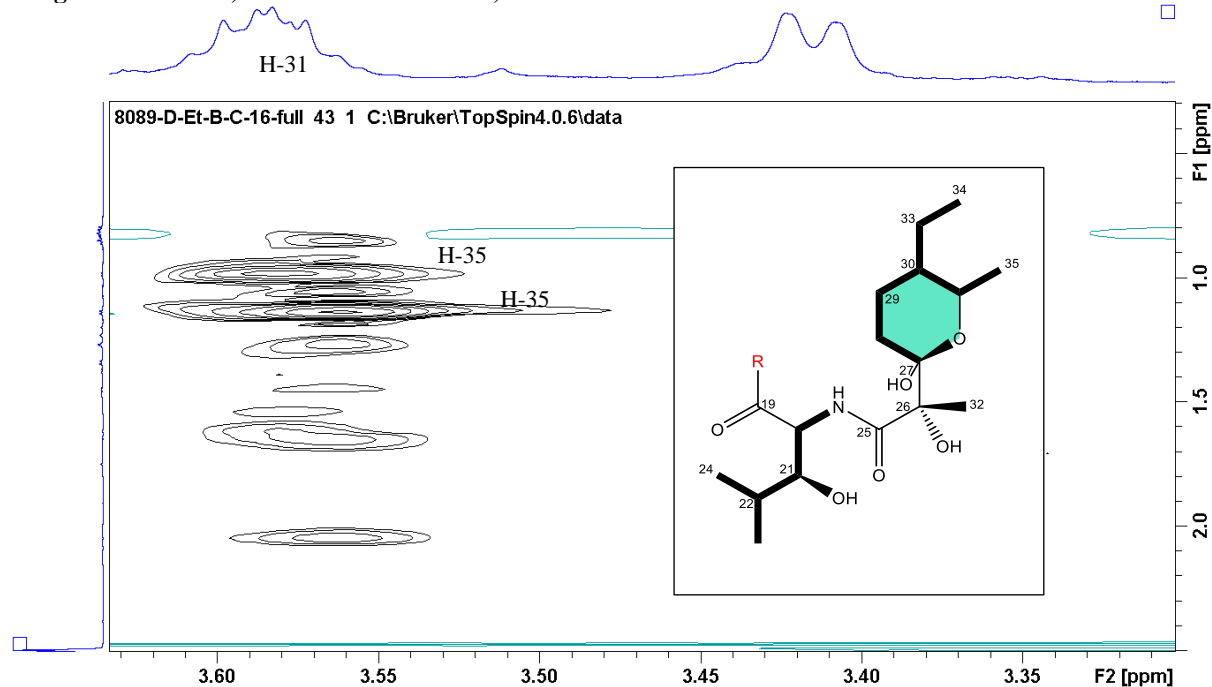
Figure 4-23. ^{1}H-^{1}H TOCSY demonstrating 4.1's disubstituted oxane ring, recorded on a sample sitting at room temperature for two weeks demonstrating a return to intense duplicated resonances, particularly for H-35. Recorded in DMSO-d_6, at 600 MHz.
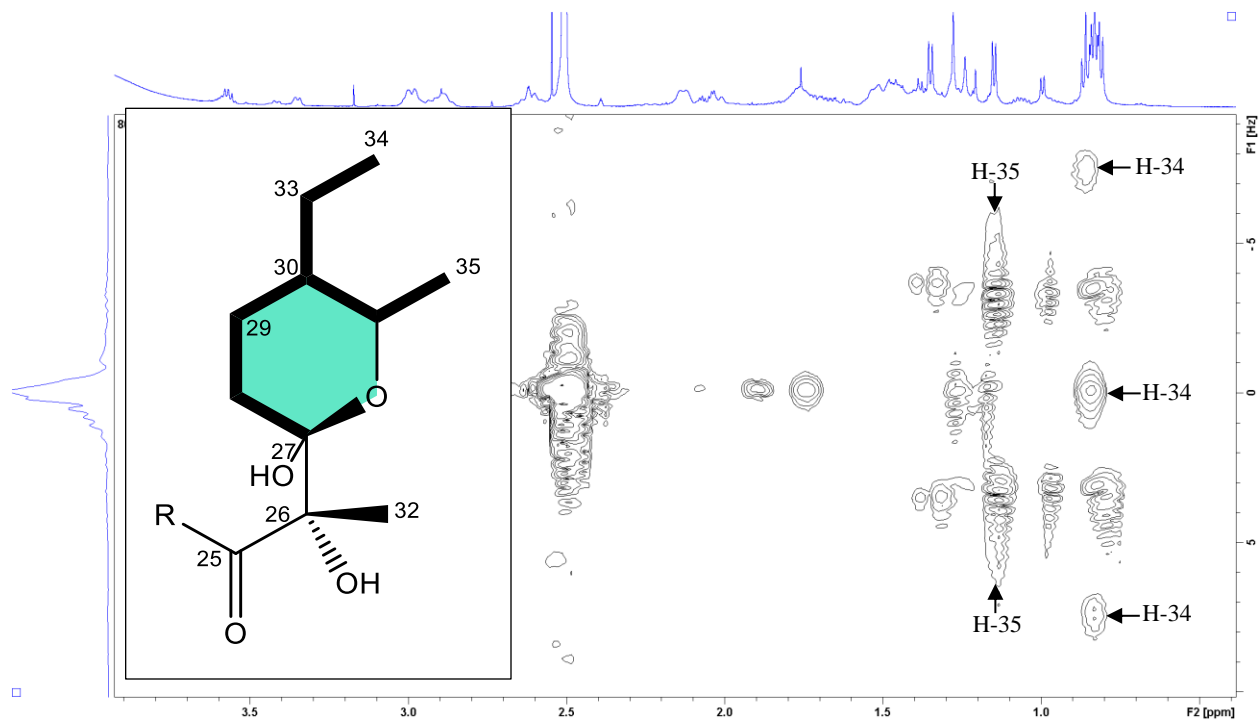
Figure 4-24. 2D *J*-resolved spectrum demonstrating the coupling patterns for the various methyl groups in 4.1 recorded from a sample that had been sitting at room temp so demonstrating multiple resonances for H-35. Recorded in DMSO-*d₆*, at 600 MHz.

| | | | DMSO-d_6 | | | | CD₃OD | | | |
| | | | Hemiketal Form | | Olefinic Form | | Olefinic Form | | Hemiketal Form | |
Residue	Position		δ_C/δ_N	δ_H (J = Hz)	δ_C/δ_N	δ_H (J = Hz)	δ_C/δ_N	δ_H (J = Hz)	δ_C/δ_N	δ_H (J = Hz)
8089-16: Dentigerumycin G (4.1)										
L-Piz 1	1	C	172.4		172.4		171.9			
	2	CH	50.6	4.9 m	50.6	4.9 m	52.4	5.05	52.4	5.05
	3 a/b	CH₂	24.7	1.76, 2.12 m	24.7	1.76, 2.12 m	26.7	1.86, 2.27 m	26.7	1.86, 2.27
	4 a/b	CH₂	21.3	1.35, 1.52 m	21.3	1.35, 1.52 m	22.3	1.64, 1.52 m	22.3	1.64, 1.52
	5 a/b	CH₂	46.7	2.61, 2.99 m	46.7	2.61/2.99 m	48.4	2.74, 3.0 m	48.4	2.74, 3.0
	NH-1	NH	-301.0	4.72	-301.0	4.72	n/a	n/a	n/a	n/a
	COO<u>H</u>	OH		13.09 br.s		13.09 br.s	n/a	n/a	n/a	n/a
D-Piz 2	6	C	172.0		172.0		174.8			
	7	CH	48.3	5.53 m	48.3	5.53 m	50.5	5.68	50.5	5.68
	8 a/b	CH₂	24.6	2.0, 1.75 m	24.6	2.0, 1.75 m	26.7	2.13, 1.88	26.7	2.13, 1.88
	9 a/b	CH₂	20.7	1.47, 1.52 m	20.7	1.47, 1.52 m	22.3	1.61	22.3	1.61
	10 a/b	CH₂	46.2	2.59, 2.9 m	46.2	2.59, 2.9 m	48.4	2.76, 3.02	48.4	2.76, 3.02
	NH-4	NH	-300.5	5.16	-300.5	5.16	n/a	n/a	n/a	n/a
N-OH-L-Ala	C11	C	172.0		172.0		174.3		173.20	
	12	CH	51.6	5.53 m	53.6	5.55	54.8	5.63	54.4	5.76
	13	CH3	14.6	1.34 d (7.4)	14.8	1.35	14.5	1.41	14.4	1.41
	N-5		-204.0				n/a	n/a	n/a	n/a
	N5-O<u>H</u>			9.55 br.s			n/a	n/a	n/a	n/a
D-Piz 3	14	C	171.6		171.6		172.9			
	15	CH	47.8	5.47 m	47.8	5.47 m	50.2	5.5	50.2	5.5
	16 a/b	CH₂	24.9	2.1, 1.77	24.9	2.1, 1.77	26.7	2.24, 1.85	26.7	2.24, 1.85
	17 a/b	CH₂	21.6	1.49, 1.34	21.6	1.49, 1.34	22.3	1.61	22.3	1.61
	18 a/b	CH₂	45.3	2.8 m	45.3	2.8 m	47.4	2.96, 3.0	47.4	2.96, 3.0
	NH-6	NH	-300.0	5.08	-300.0	5.08	n/a	n/a	n/a	n/a
β-S-OH-L-Leu	19	C	171.5		171.5		171.2		172.4	

	20	CH	49.8	5.42 m	49.8	5.42 m	50.5	5.55	50.5	5.56
	21	CH	75.6	3.34 br.d	75.6	3.41 m	77.1	3.7	77.1	3.67
	22	CH	28.8	1.88 m	28.8	1.88 m	28.4	1.87	28.4	1.87
	23	CH$_3$	14.7	0.82 d (6.8)	14.7	0.82 d (6.8)	15.5	0.92	15.5	0.91
	24	CH$_3$	20.2	0.82 d (6.8)	20.2	0.82 d (6.8)	20.3	0.97	20.8	0.94
	NH-8	NH	-267.0	7.66 d (10.0)	-267.0	7.23 d(10.0)	n/a	n/a	n/a	n/a
	21-OH			not visible			n/a	n/a	n/a	n/a
PKS	25	C	176.5		172.3		174.2		177.9	
	26	C	76.9		75.2		76.7		77.0	
	27	C	98.0		154.1		154.5		101.0	
	28 a/b	CH$_2$	28.0	2.15, 1.95	93.1	4.81	97.3	4.95	26.7	2.15, 1.60
	29 a/b	CH$_2$	24.3	1.69	24.1	2.02, 1.68	25.5	2.14, 1.75	23.4	1.74, 1.39
	30	CH	41.9	0.97 m	38.6	1.26 m	40.5	1.38 m	43.8	1.08 m
	31	CH	69.1	3.58 m	74.1	3.35 m	73.5	3.41m	72.2	3.69 m
	32	CH$_3$	21.2	1.2 s	23.7	1.27 s	30	1.29 s	21.2	1.35 s
	33 a/b	CH$_2$	23.7	1.39, 0.96 m	24.1	1.06, 0.98 m	25.2	1.54, 1.12 m	24.9	1.02, 0.88 m
	34	CH$_3$	10.1	0.82 t (7.5)	11.3	0.87 t (7.5)	11.0	0.92 t (7.5)	11.0	0.89 t (7.5)
	35	CH$_3$	19.1	0.99 d (6.5)	19.1	1.14 d (6.5)	19.0	1.23 d (6.5)	19.0	1.08 d (6.5)
	26-OH			not visible		not visible	n/a	n/a		
	27-OH			not visible		not visible	n/a	n/a		

Table 4-1. NMR shifts of dentigerumycin G's hemiketal and dehydrated forms recorded in either DMSO-d_6 or CD$_3$OD at 600 MHz.

4.3.2 Structure Elucidation of Dentigerumycin F (4.2)

The structure elucidation of dentigerumycin F (**4.2**) will be described next, which is the order in which the isolated piperazic acid-containing nonribosomal peptides structures were solved, despite dentigerumycin F eluting from the column first. The data set for **4.2** was obtained in pyridine-d_5 due to earlier challenges solving the structure of dentigerumycin G (**4.1**) using spectra recorded in DMSO-d_6. The complexity of the data in DMSO-d_6 was later understood to be due to an equilibrium between hemiketal and dehydrated forms of the molecule. Interestingly, when **4.2** was placed in pyridine-d5, the rate of equilibrium appeared significantly reduced with the hemiketal form favoured —this greatly aided structure elucidation of **4.2**. After the initial datasets were recorded, pyridine-d_5 was removed, and DMSO-d_6 was added, leading to more complex spectra but with the ability to assign resonances (**Table 4-2**) due to work already completed on **4.1.**

Dentigerumycin F (**4.2**)
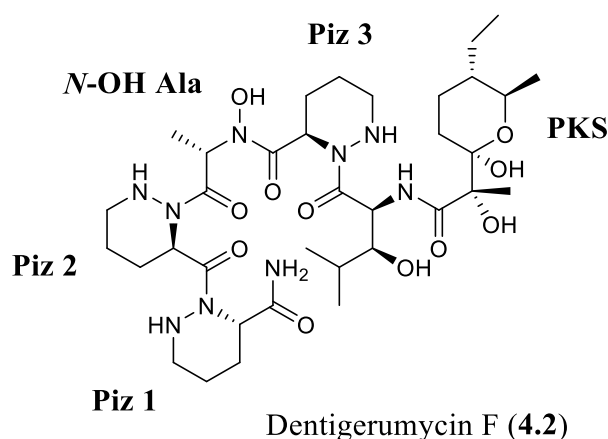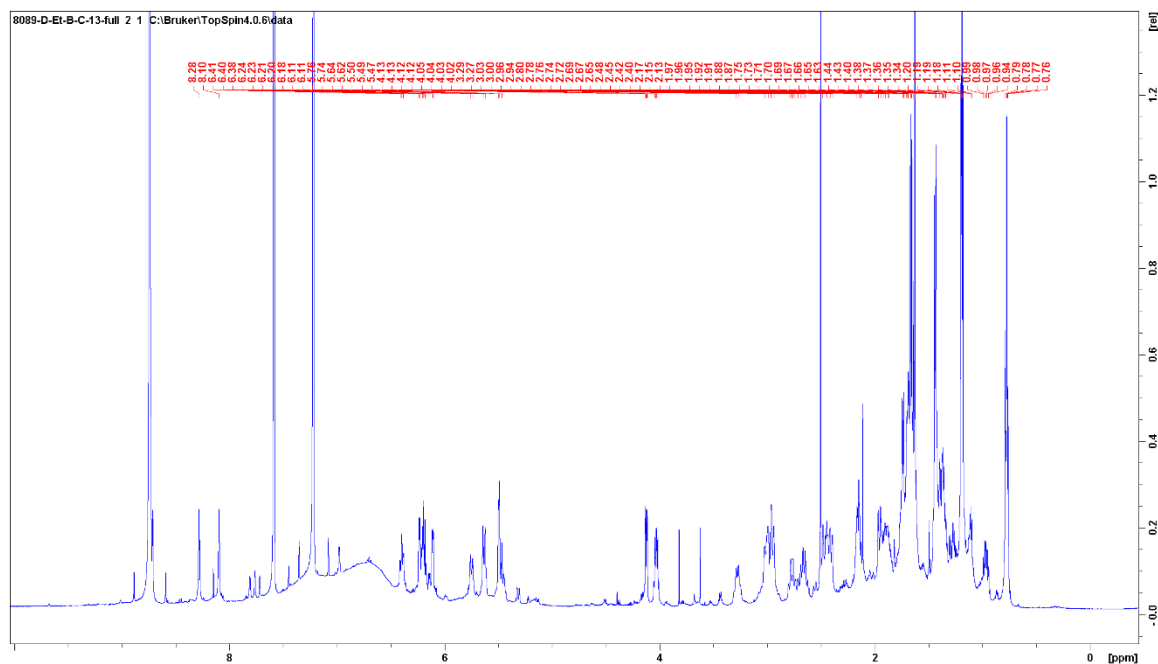
Figure 4-25. ^{1}H spectrum for dentigerumycin F recorded in pyridine-d_5 at 600 MHz.
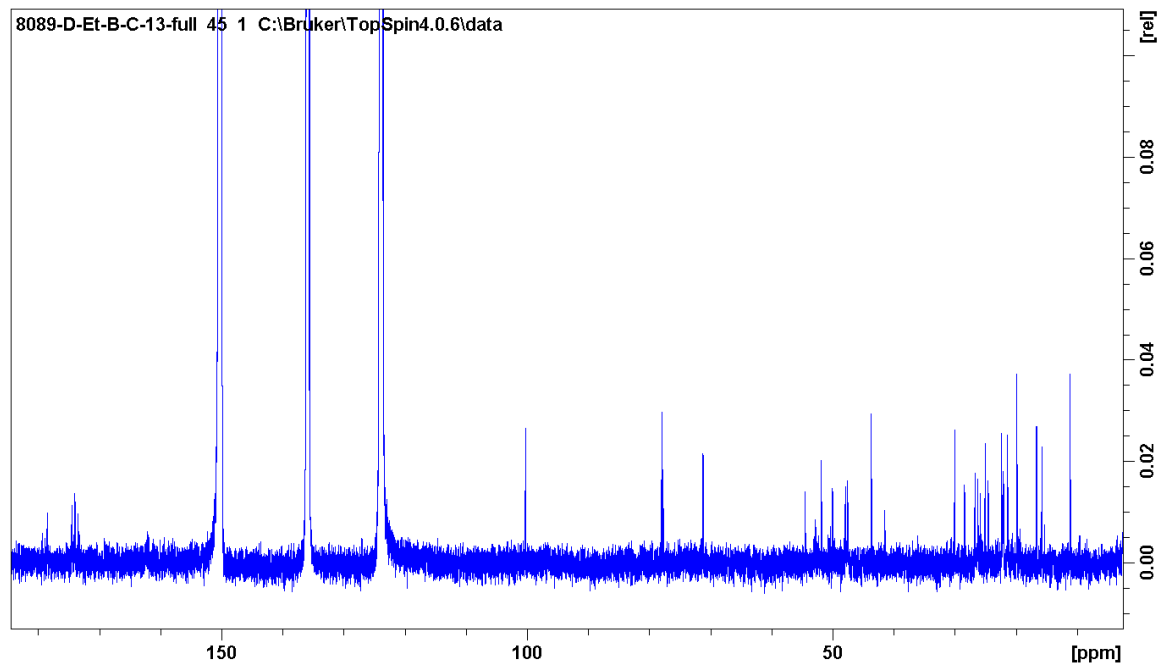
Figure 4-26. ^{13}C spectrum for dentigerumycin F recorded in pyridine-d_5 at 150 MHz.

Dentigerumyin F (**4.2**) was isolated as a clear oil that gave a $[M + Na]^+$ ion in the

HRESIMS at *m/z* 806.4369 appropriate for a molecular formula of $C_{35}H_{61}N_9O_{11}$ that requires ten

146

sites of unsaturation. Analysis of the ^{1}H, ^{1}H-^{1}H gCOSY60, ^{1}H-^{1}H gTOCSY/^{1}H-^{13}C gHSQC/^{1}H-^{15}N HSQC/^{1}H-^{13}C and gHMBC NMR data recorded in pyridine-d_5 for **4.2** revealed that **4.2** contains three Piz residues (one of which contained a terminal amide), an N-hydroxy alanine residue (*N*-OH-Ala), a 3-OH-leucine, along with an acyl side chain containing an oxane ring bearing vicinal methyl and ethyl substituents in a trans relative configuration. The ^{1}H NMR data showed two amide protons at δ_H 8.27 and 8.09 (**Figure 4-25**), demonstrating ^{1}H-^{15}N HSQC correlations into a nitrogen at δ_N -275.8, which led to describing this molecule as a likely linear free-amide peptide-containing molecule. The three exchangeable proton resonances at δ_H 5.50 (δ_N -300.3), 5.75 (δ_N -300.0), and 5.67 (δ_N -300.9) showed one-bond correlations to nitrogen resonances in the ^{1}H-^{15}N HSQC spectrum and were the signals followed throughout the isolation of this compound (**Figure 4-29**). These Piz-like correlations were subsequently shown to be saturated piperazic acid spin systems (**Figure 4-28**) in the ^{1}H-^{1}H TOCSY (**Figure 4-30 and 4-31**) spectrum. In particular, δ_H 5.50 (NH-1) demonstrated correlations into H-5a/b, H-4a/b, H-3a/b, and H-2a/b ; δ_H 5.67 (NH-4) showed correlations into H-10a/b, H-9a/b, H-8a/b, and H-7a/b; while δ_H 5.75 (NH-6) had correlations into the methylene protons of H-18a/b, H-17a/b, H-16a/b, and H-15. Further evidence was seen in the HSQC spectrum, which showed that all methylene carbons adjacent to the Piz N-H's possessed the deshielded carbon resonances typical of these moieties: C-5 (δ_C 47.7), C-10 (δ_C 47.4) and C-18(δ_C 46.9).
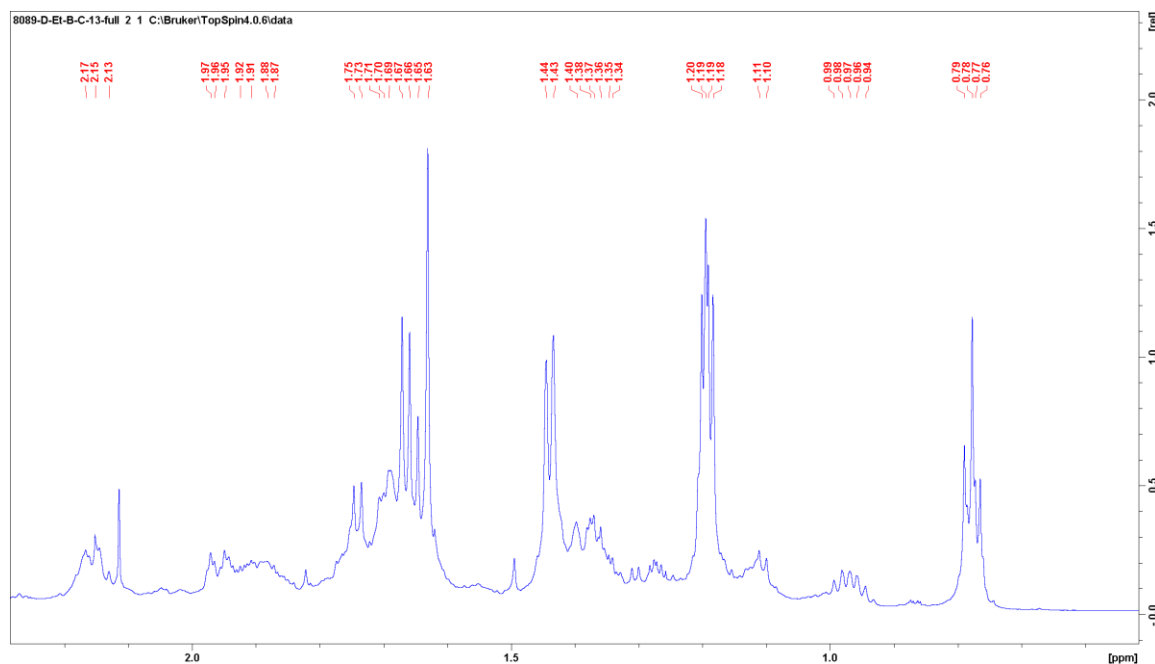
Figure 4-27. Expansion of the ^{1}H NMR spectrum for 4.2, demonstrating a clear triplet coupling pattern at the upfield resonance δ$_H$ 0.77.
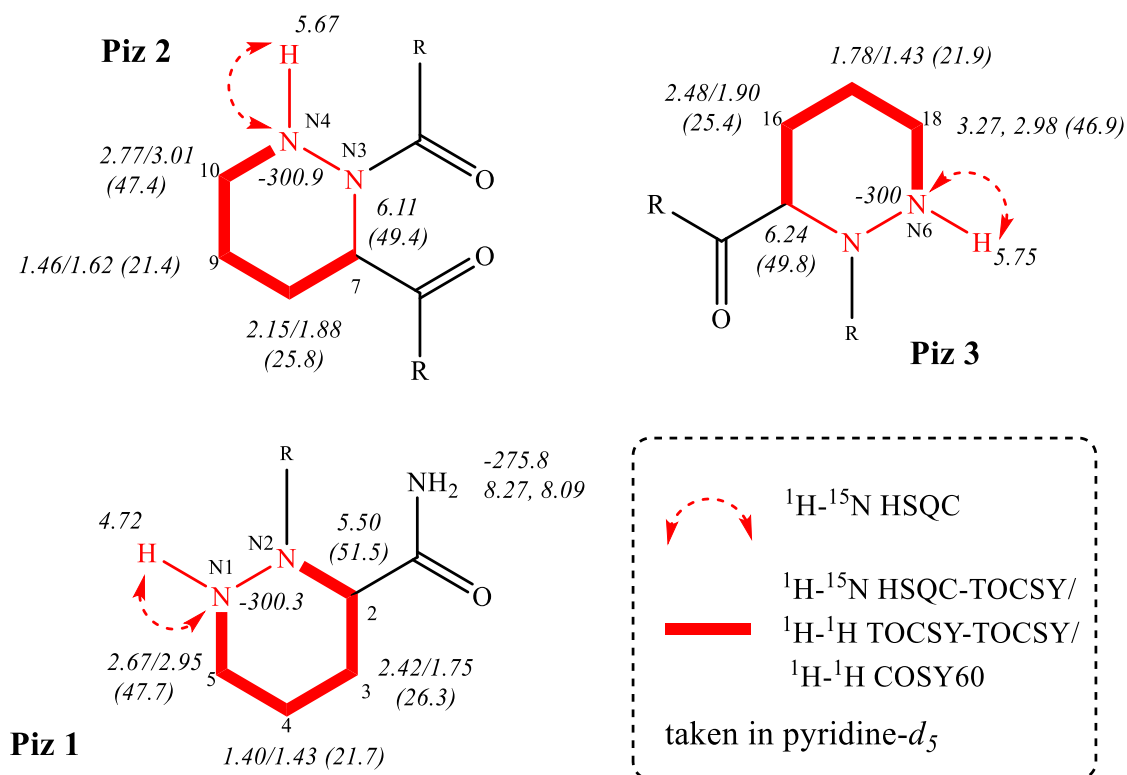
Figure 4-28. Chemical shift assignments for the three Piz amino acids of dentigerumycin F.

148
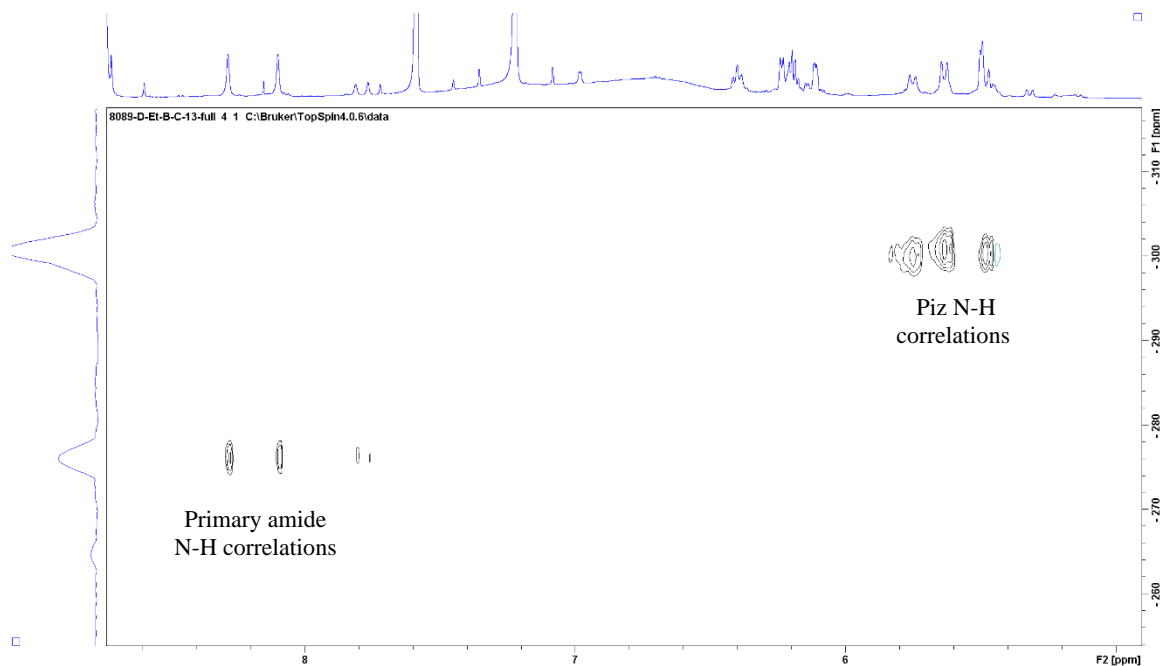
Figure 4-29. ^{1}H-^{15}N HSQC NMR spectrum of dentigerumycin F, recorded at 600 MHz in pyridine-d_5, showing the three Piz N-H correlations along with the primary amide N-H$_2$ correlations.
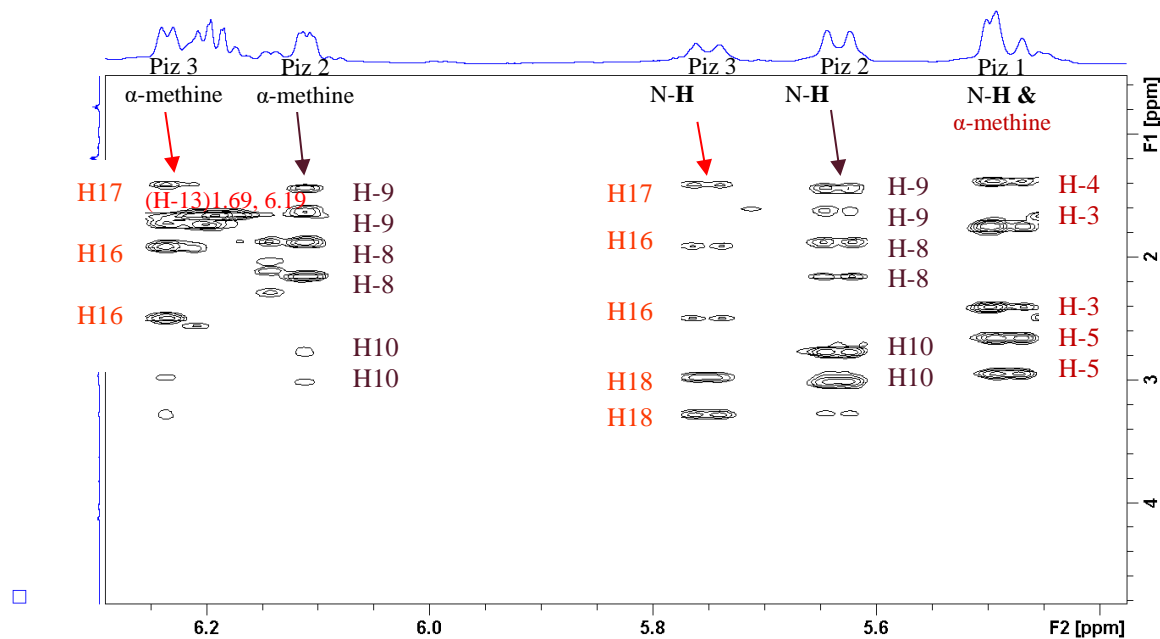
Figure 4-30. Expansion of ^{1}H-^{1}H TOCSY NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5 demonstrating the three Piz spin systems with scalar coupling from the Piz nitrogen-attached proton as well as the Piz α-methine protons.

149
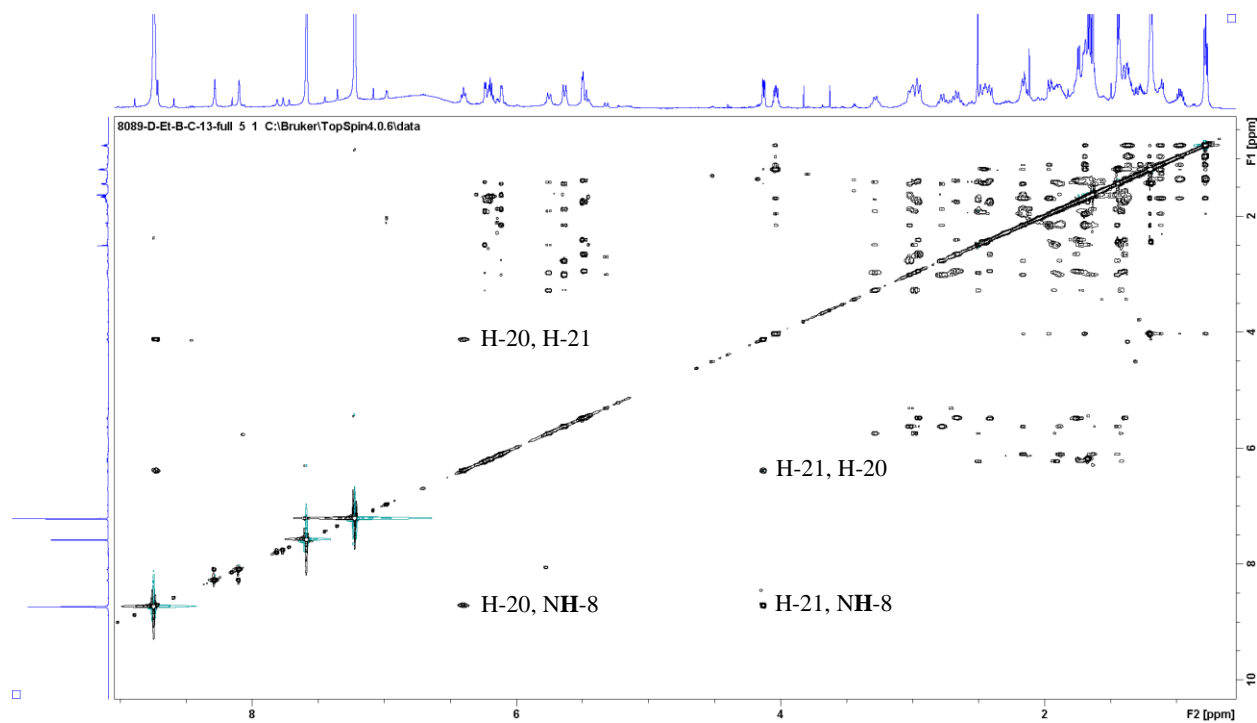
Figure 4-31. ^{1}H-^{1}H TOCSY NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5 also demonstrating the clear correlations between H-20, H-21 and NH-8.
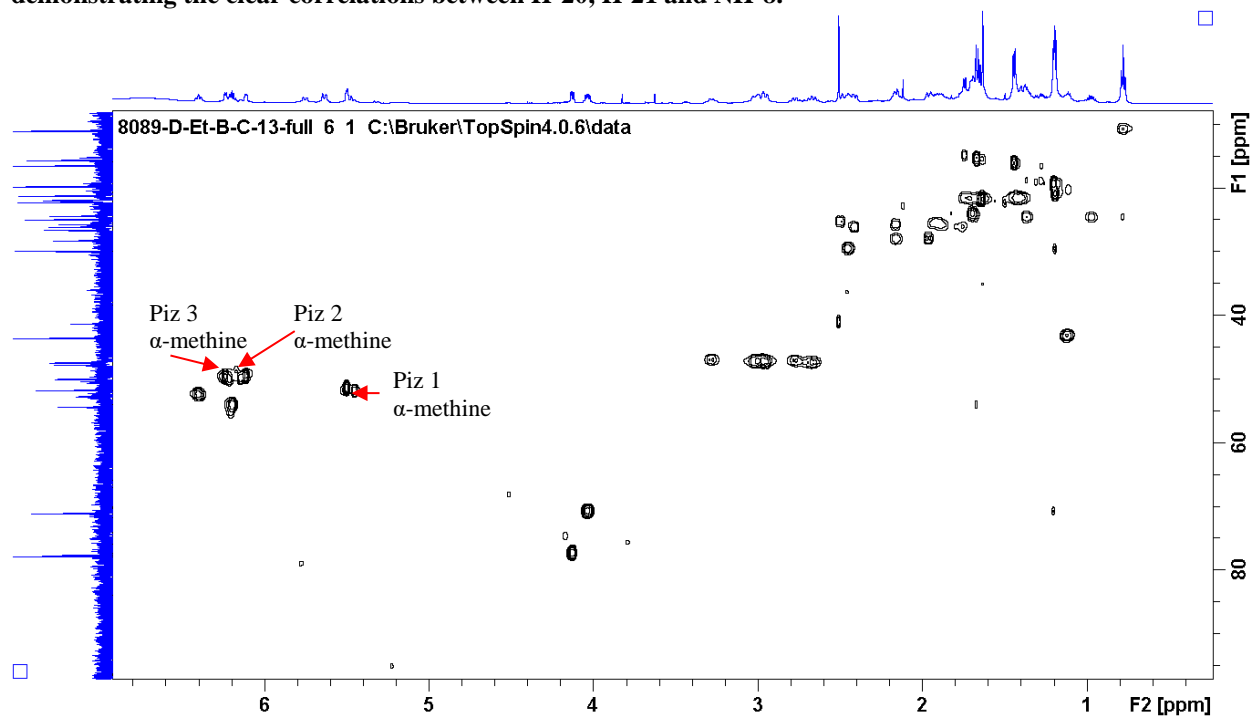
Figure 4-32. ^{1}H-^{13}C HSQC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5.
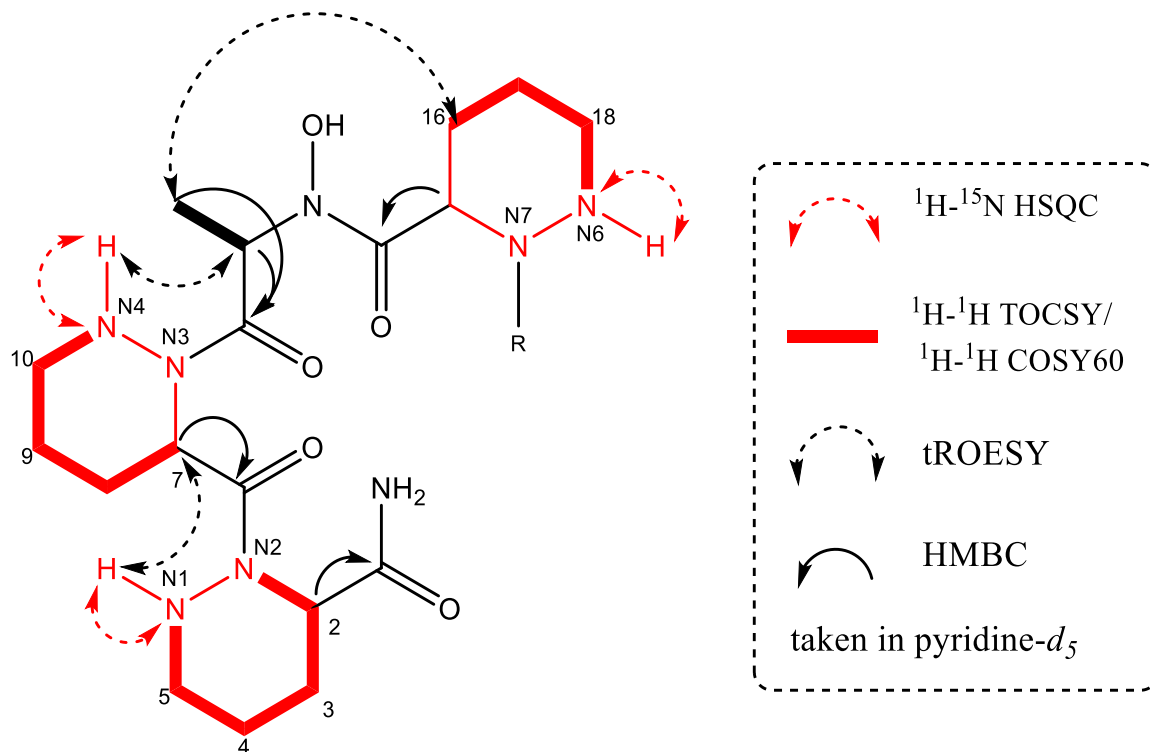
Figure 4-33. Connecting the core amino acids for dentigerumycin F.

The presence of the N-hydroxylated alanine was confirmed both via COSY60 and ^{1}H-^{1}H TOCSY correlations between the methyl doublet resonating at δ_H 1.67 (H-13) and the deshielded α-methine proton resonating at δ_H 6.19 (H-12) (**Figure 4-34**), the presence of ^{1}H-^{15}N LRHSQC correlations between the methyl doublet and the shielded nitrogen (δ_H 1.67, δ_N -197.6) suggestive of a hydroxamate moiety (**Figure 4-36**), and a ^{1}H-^{13}C HMBC correlation between the methyl doublet (δ_H 1.67) and the deshielded α-carbon (δ_C 54.2, C-12) (**Figure 4-35**). The hydroxamate proton was not visible due to the basic nature of the pyridine NMR solvent. tROESY correlations between adjacent α-methine's and the nitrogen-attached Piz protons established the placement of Piz 1, Piz 2 and *N*-OH-Ala (**Figure 4-37**). Since the nitrogen-attached hydroxyl was a very broad resonance in pyridine-d_5, it could not be used to establish the connection between *N*-OH-Ala and

Piz 3. However, tROESY correlations were visible between the methyl resonances at H-13 and one of the proton resonances at H-16 (**Figure 4-38**). 3D models demonstrated that the distance between H-13 and H-16 was 2.4A, well within the through-space range for strong NOEs.[221] Thus, the amino acid sequence of **4.2** was confirmed as Piz, Piz, *N*-OH-Ala, Piz (**Figure 4-33**).
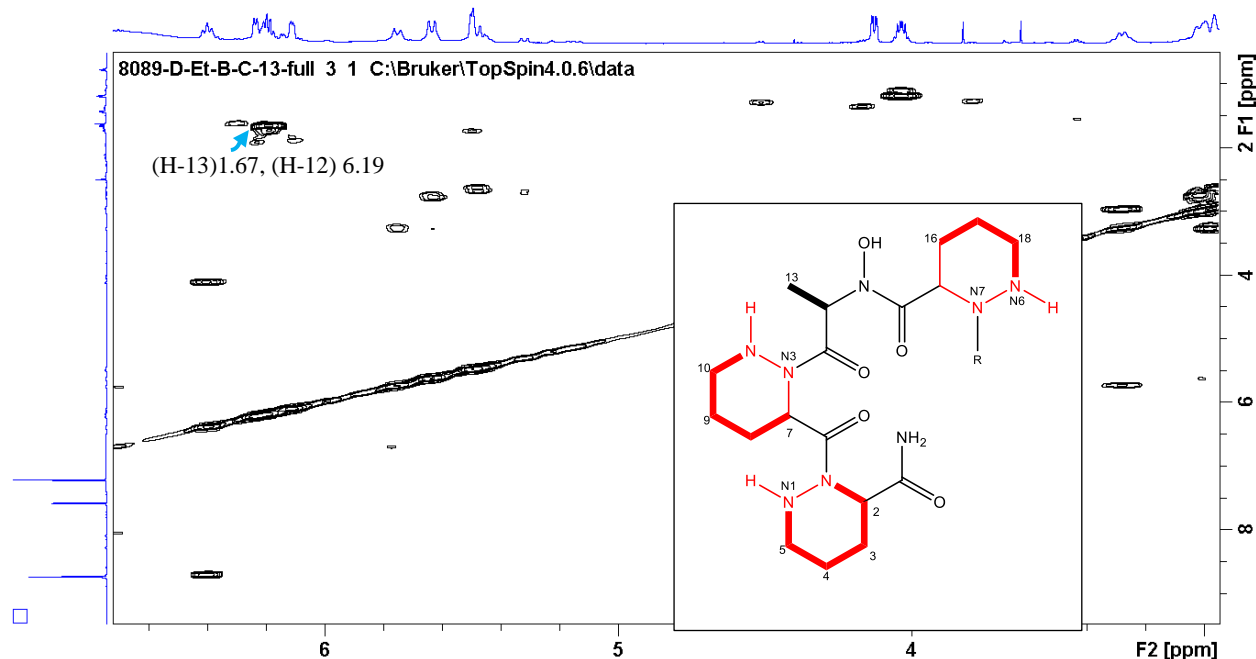
Figure 4-34. Expansion of ¹H-¹H COSY60 NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-*d*₅ which shows the correlation between neighboring H-13 and H-12 protons.

Figure 4-35. Expansion of ^{1}H-^{13}C HMBC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-*d5* which shows the correlation between the H-13 resonance and C-12 and C-11.

Figure 4-36. ^{1}H-^{15}N LRHMQC demonstrating correlations between *N*-OH-Ala's methyl protons and the nitrogen-attached hydroxyl group for dentigerumycin F, at 600 MHz in pyridine-*d5*.

153
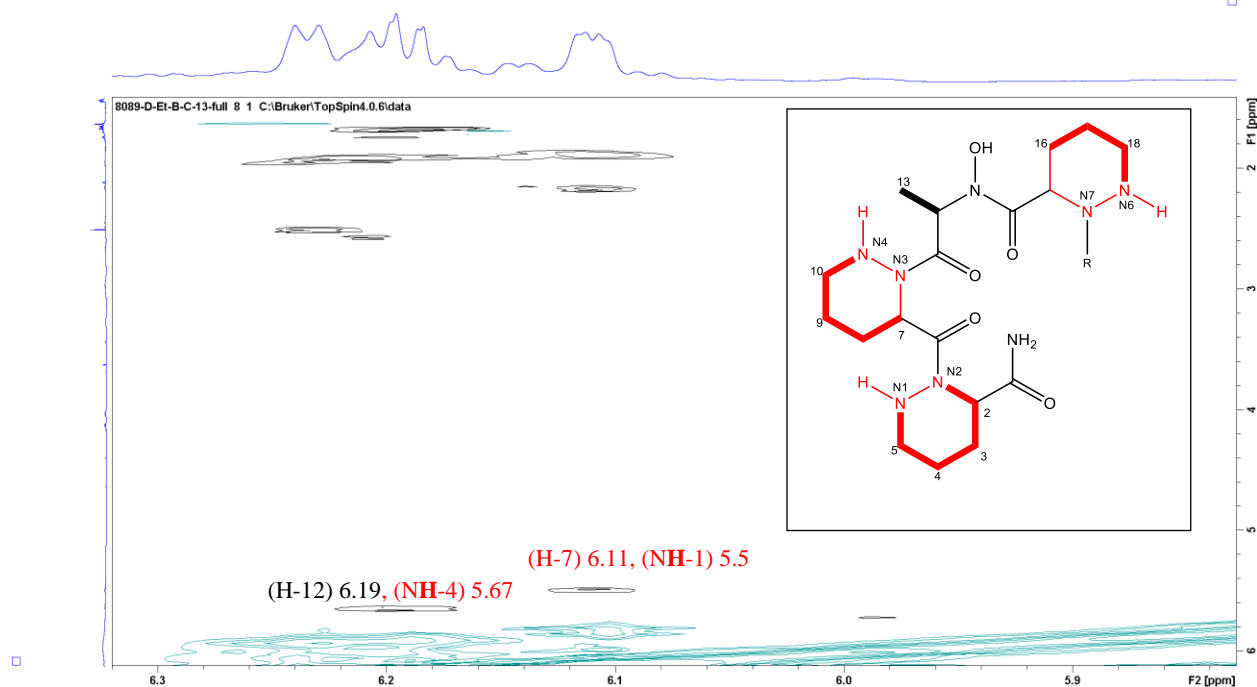
Figure 4-37. Expansion of tROESY NMR spectrum for dentigerumycin F, recorded at 600 MHz in pyridine-*d₅*, demonstrating the through-space correlations connecting Piz 1, Piz 2 and the *N*-OH-Ala amino acids.
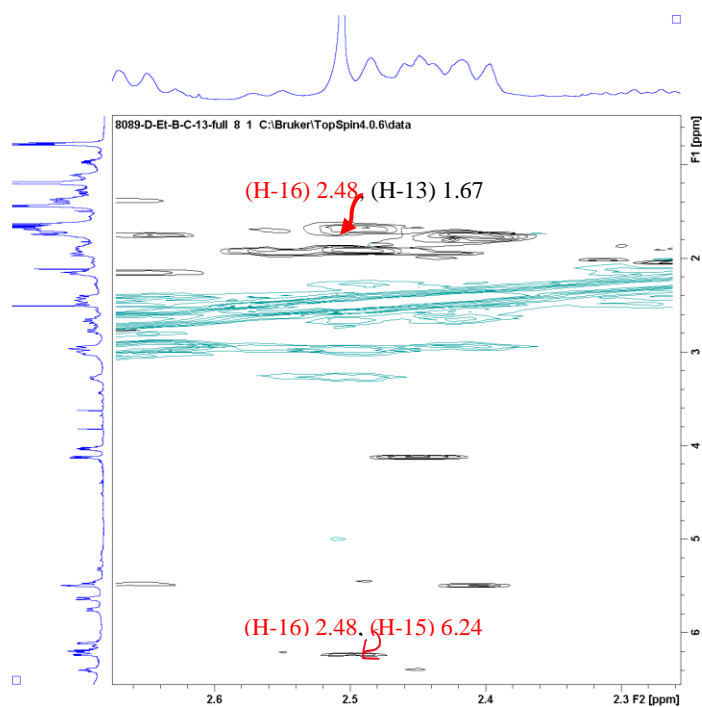
Figure 4-38. Expansion of the tROESY NMR spectrum for dentigerumycin F, recorded at 600 MHz in pyridine-*d₅*, demonstrating the through-space correlations connecting Piz 3 and the *N*-OH-Ala amino acids.

154
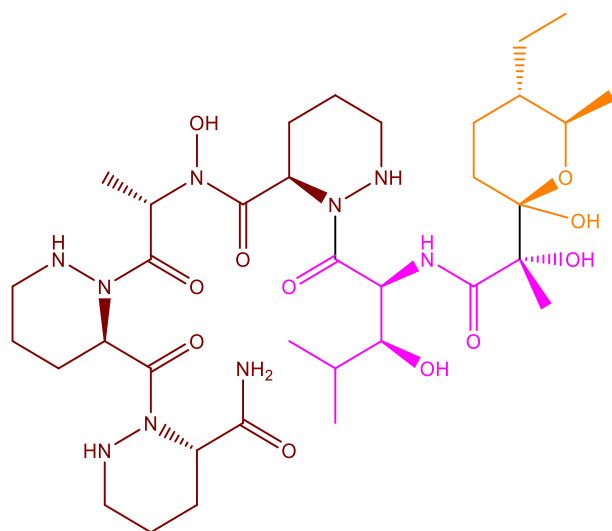
Fragment A : $C_{18}H_{31}N_8O_5$ Calculated mass: 439.2417
Fragment B : $C_9H_{16}NO_4$ Calculated mass: 202.1079
Fragment C : $C_8H_{15}O_2$ Calculated mass: 143.1072

Figure 4-39. Building up fragments of dentigerumycin F to help confirm the structure of the PKS portion.

To differentiate the structures of constitutional isomers dentigerumycin F (**4.2**) and

dentigerumycin D (**4.8**) fragments of the molecules were assessed for mass and formula

differences. The first point of difference was seen in the core amino acid mass (**Figure 4-39**);

Fragment A has a molecular formula of $C_{18}H_{31}N_8O_5$ while the comparable fragment in

dentigerumycin D[237] has a molecular formula of $C_{17}H_{29}N_8O_5$ (molecular formula for this

fragment of dentigerumycin D was determined using ChemDraw on the published structure). For

4.2, this left, as in **4.1**, seventeen carbons, thirty protons, one nitrogen, and six oxygen atoms that

were yet to be assigned within the structure. Recorded in pyridine-d_5, the hemiketal/dehydro

equilibrium (discussed in **Subsection 4-3-3**) for dentigerumycin F (**4.2**) was not as pronounced;

however, it was still a complicating factor in the structure elucidation. Through NMR analysis, it

became evident that the remainder of the molecule was identical to **4.1**. Notably, **4.2** also

contained a methyl singlet at C-32 (δ_C 21.7, δ_H 1.64) which showed clear HMBC correlations into a quaternary carbon at δ_C 78.2 (C-26), δ_C 100.2 (C-27) along with a carbonyl carbon at δ_C 178.5 (C-25) (**Figure 4-42).** The hydroxyl substituent of the C-26 quaternary carbon was not visible in the data due to the basic nature of the deuterated pyridine utilized as NMR solvent. However, the deshielded carbon resonances at C-26 and the singlet methyl at C-32 also provided evidence that there were no coupled protons attached to C-26. Further analysis of the 2-D NMR data revealed the characteristic β-OH-Leucine in the TOCSY spectrum with a spin system inclusive of a methine doublet of doublets at H-21 (δ_H 4.12), demonstrating TOCSY correlations to a methine at H-22 (δ_H 2.45), methyl doublets at H-23 (δ_H 1.44, δ_C 16.0) and H-24 (δ_H 1.19, δ_C 21.1) (**Figure 4-45),** a methine proton at H-20 (δ_H 6.40) and an amide proton at δ_H 8.71 (NH-8) (**Figure 4-31**). COSY data demonstrated cross peaks from the H-22 methine to the two methyl doublets at H-23 and H-24 (**Figure 4-44**). Strong HMBC correlations from the N-H amide proton at δ_H 8.71 to the C-25 carbonyl at δ_C 178.5; from H-20 to the carbonyls at C-19 (δ_C 174.2) and C-25; and from H-21 to C-19, built the connection between the NRPS and PKS portions of the molecule (**Figure 4-40**). Fragment B seen in **Figure 4-39,** was then constructed with a partial molecular formula of $C_9H_{16}NO_4$.
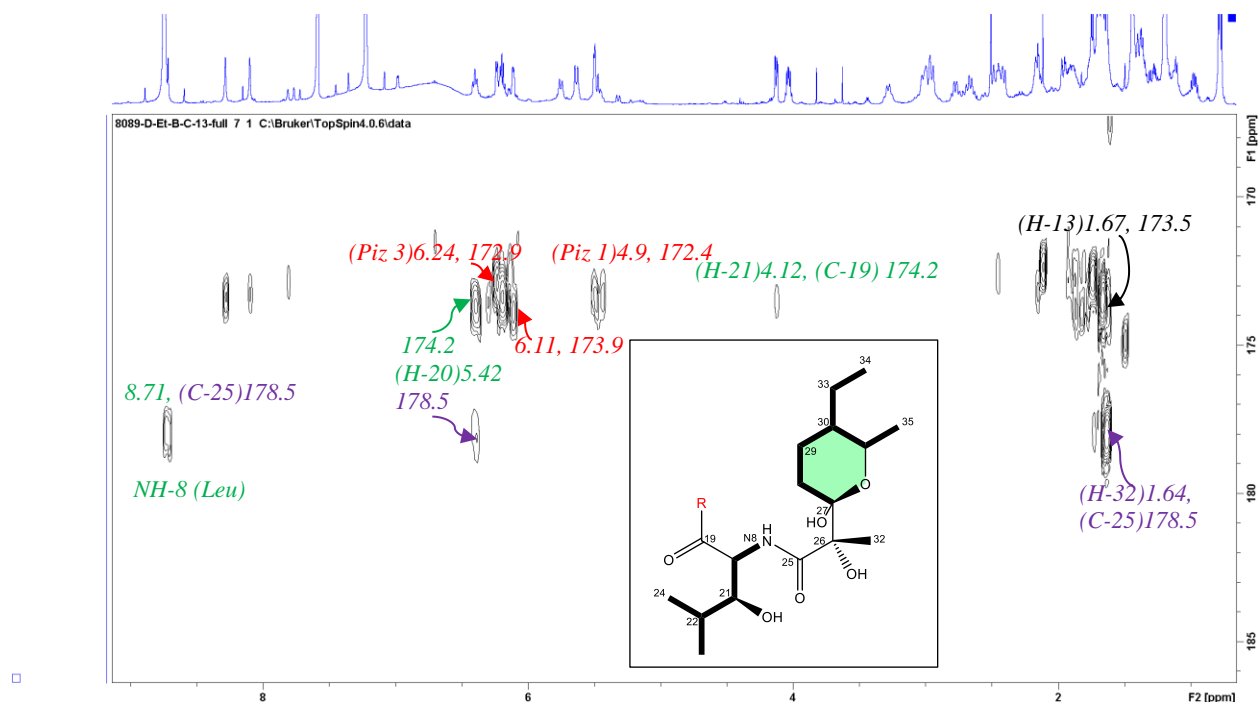
Figure 4-40. Expansion of the ^{1}H-^{13}C HMBC spectrum for 4.2 demonstrating correlations connecting the NRPS and PKS portions of the molecule through H-20, H-21 and NH-8 into C-19 and C-25. Also seen are the correlations which linked Piz protons to adjacent carbonyls. Recorded at 600 MHz in pyridine-d_5.

With this information, it was seen that there were eight carbons, fifteen protons and two oxygens that were yet to be assigned within the structure. With this information in hand, it was possible to elucidate the remainder of the molecule. Through analysis of the 2-D data, it became evident that the remaining $C_8H_{15}O_2$ formed a disubstituted oxane ring that was identical to the construction of the PKS sidechain seen in the free acid. The oxane ring was again revealed through gCOSY60, ^{1}H-^{1}H TOCSY and HMBC correlations. First, the COSY spectrum allowed clarification of the suspected proton methyl groups (**Figure 4-44**). The H-34 methyl shift (δ_H 0.77, δ_C 10.7), had a shielded chemical shift often associated with aliphatic methyl groups and was well separated from the β-OH-Leucine two methyls in pyridine-d_5 such that its triplet coupling pattern was very clear. This triplet splitting pattern was confirmed via the 2D *J*-resolved NMR experiment (**Figure 4-48**). In pyridine-d_5, the resonance for one of the leucine's

157

methyl groups was overlapping with the H-35 methyl peak so the COSY60 correlations (**Figure 4-44**) were utilized to confirm assignment. Then, the TOCSY spectrum allowed the spin system construction (**Figure 4-45**), building from the H-31 methine proton (δ_H 4.04) to the methyl doublet at H-35 (δ_H 1.9) on one side and from the H-31 methine proton to a shielded methine proton at H-30 (δ_H 1.11, δ_C 43.1) along to the methylene and methyl protons of H-33a/b (δ_H 1.39, 0.98) and H-34 (δ_H 0.77). This spin system also showed extension through to the H-29 and H-28 methylene protons, which was then confirmed with COSY60 correlations of the H-29 methylene (δ_H 1.70) to the distinctive (**Figure 4-46**) H-30 methine.

PKS sidechain

Figure 4-41. Correlations confirming the β-OH-Leucine amino acid characteristic of these compounds along with the unique substitution pattern in the PKS portion of dentigerumycin F. As well, the correlations that link the amino acid constructed portion of the molecule with the polyketide constructed portion of the molecule are seen.

Again, HMBC correlations (**Figure 4-41 and 4-47**) between the C-30 resonating at δ_C 43.1 and the methyl at H-35 (δ_H 1.19) along with the deshielded C-31 resonating at δ_C 70.8 and a methylene proton from H-33 (δ_H 0.98) confirmed the adjacent nature of the tetrahydropyran

substitutions. This unique ring pattern distinguished **(4.2)** as being a novel linear piperazic acid-containing free amide. As in dentigerumycin G (**4.1**), the absolute configuration of the molecule was assigned using the crystal structure of incarnatapeptin A (**4.3**) *Vida infra* as they are produced from the same biosynthetic pathway along with the similarity in NMR resonances for the amino acids and PKS-NRPS link.
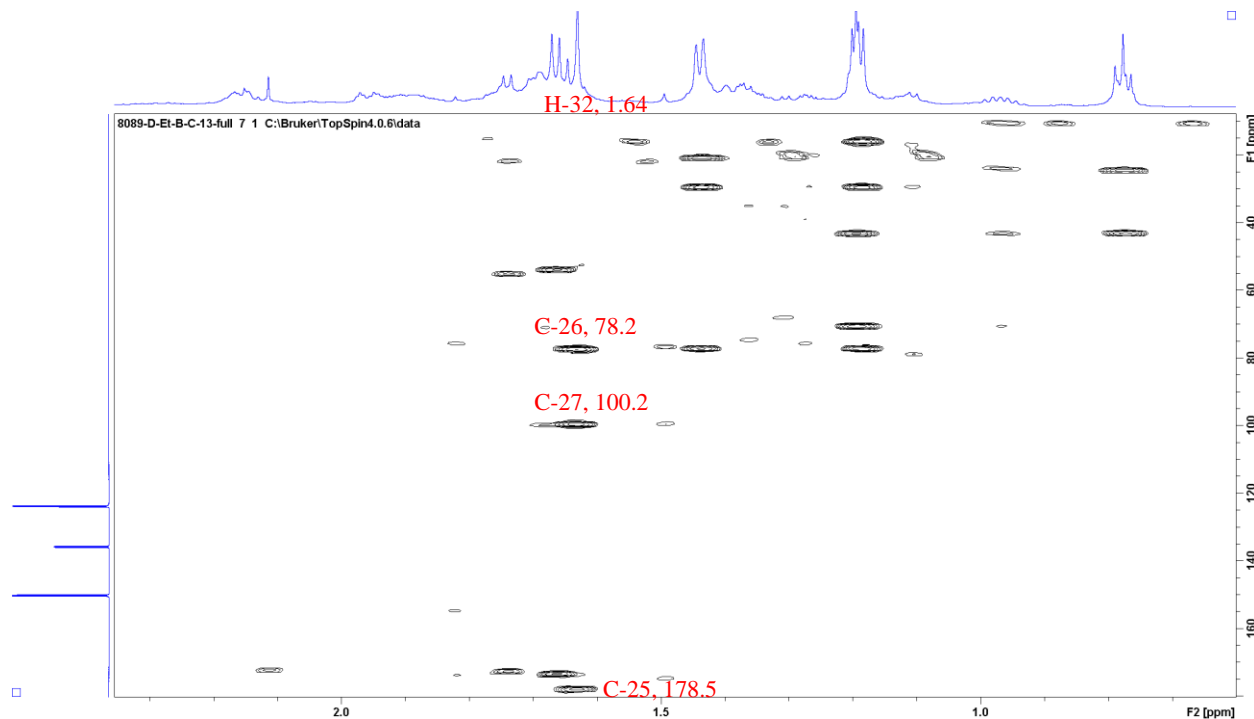
Figure 4-42. ^{1}H-^{13}C HMBC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5 demonstrating correlations from the methyl singlet in the PKS unit to the adjacent quaternary carbons.
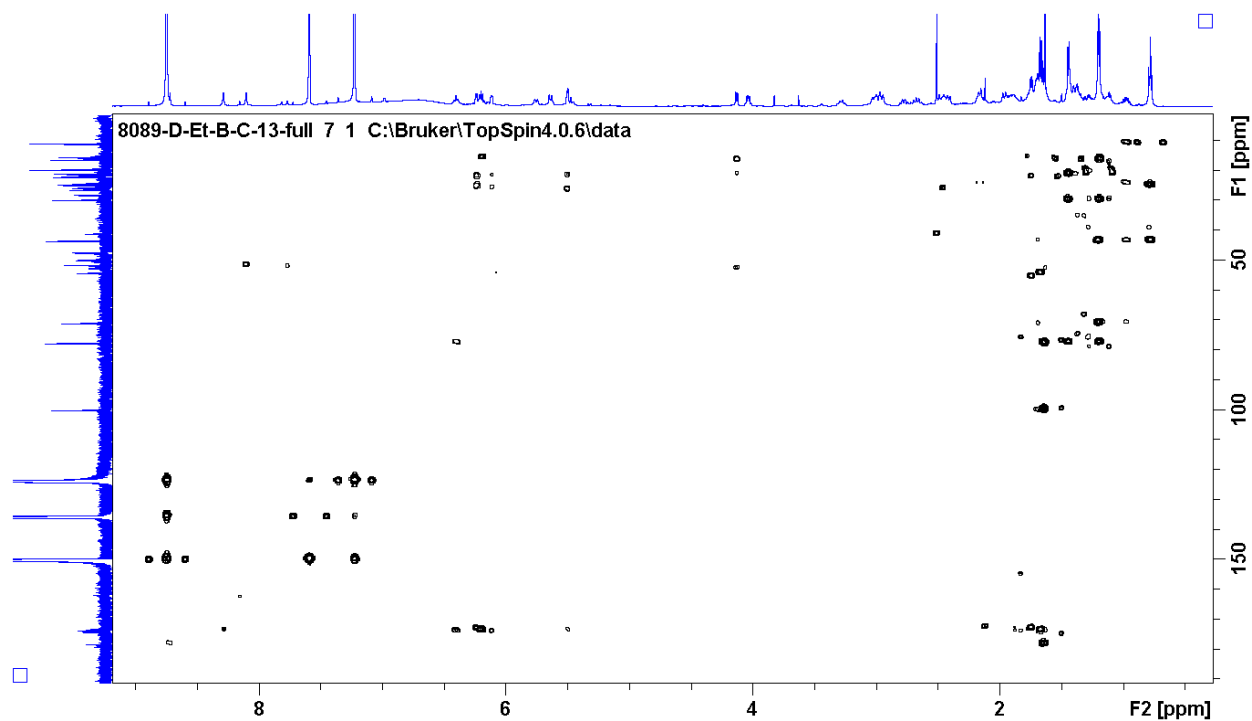
Figure 4-43. Full ^{1}H-^{13}C HMBC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5.
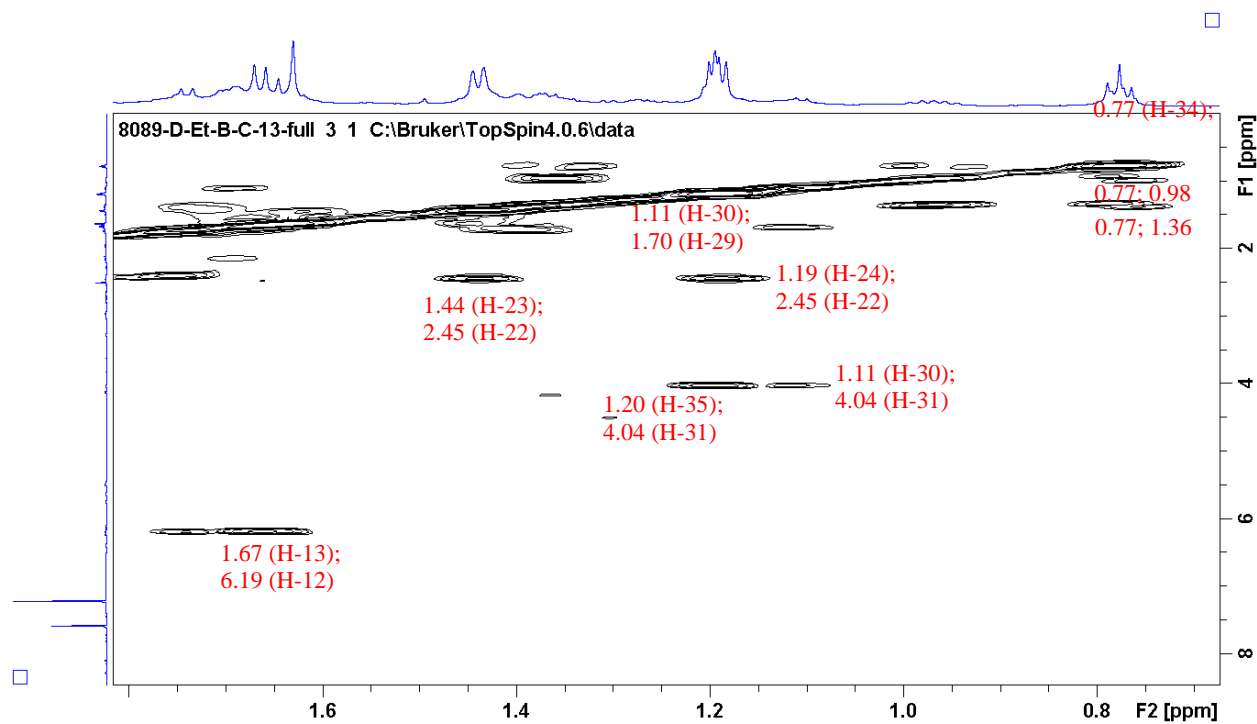
Figure 4-44. Expansion of the ^{1}H-^{1}H COSY60 NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5 wherein the correlations connecting the neighboring protons of the oxane ring are seen.
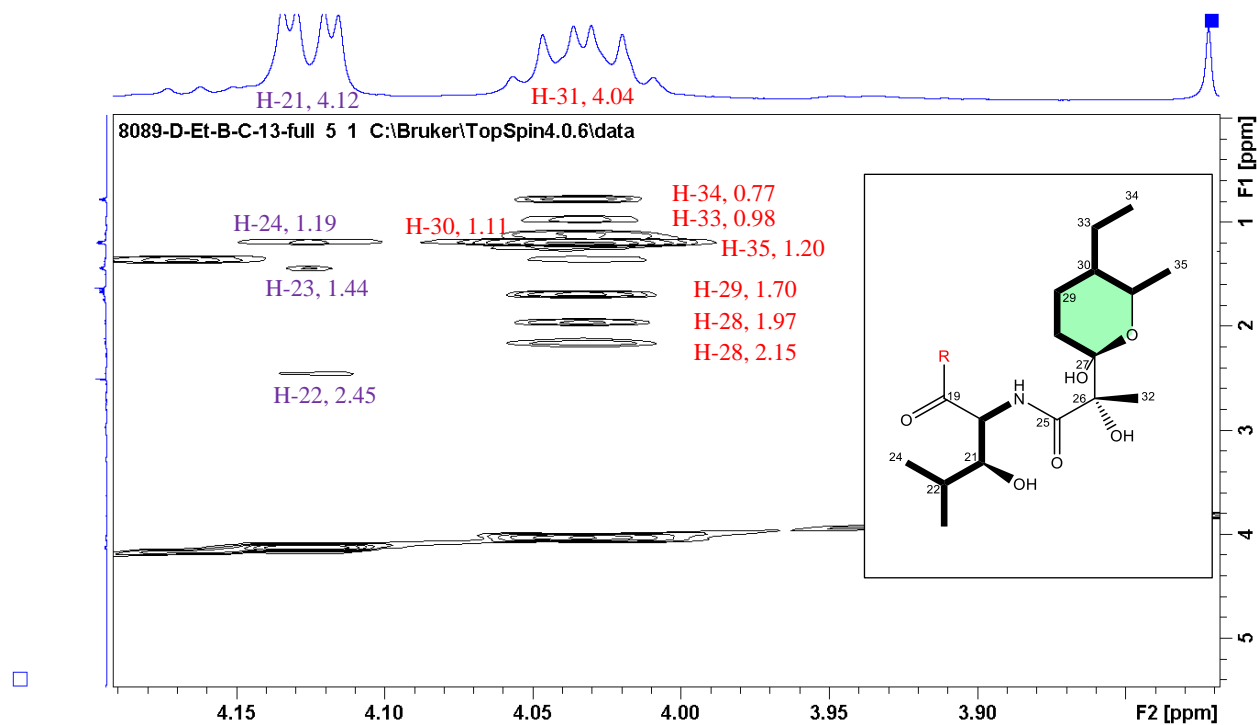
Figure 4-45. ^{1}H-^{1}H TOCSY NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5.
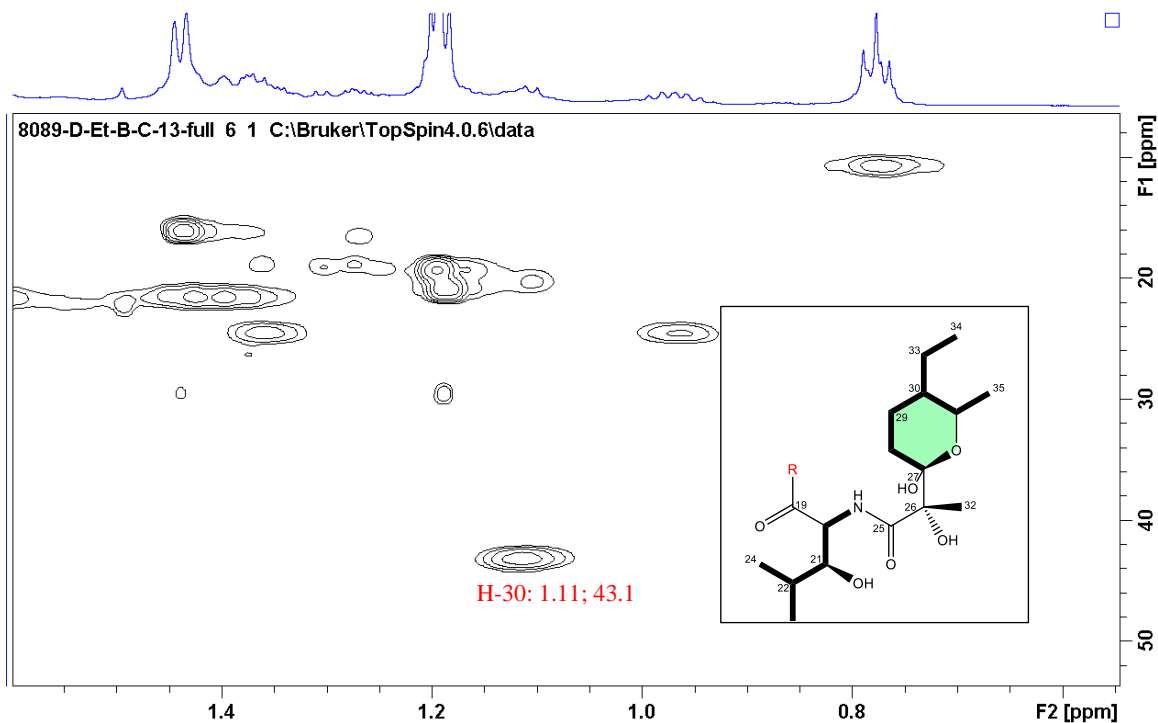
Figure 4-46. Expansion of the ^{1}H-^{13}C HSQC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-d_5 demonstrating the characteristic H-30 to C-30 ^1J correlation.

161
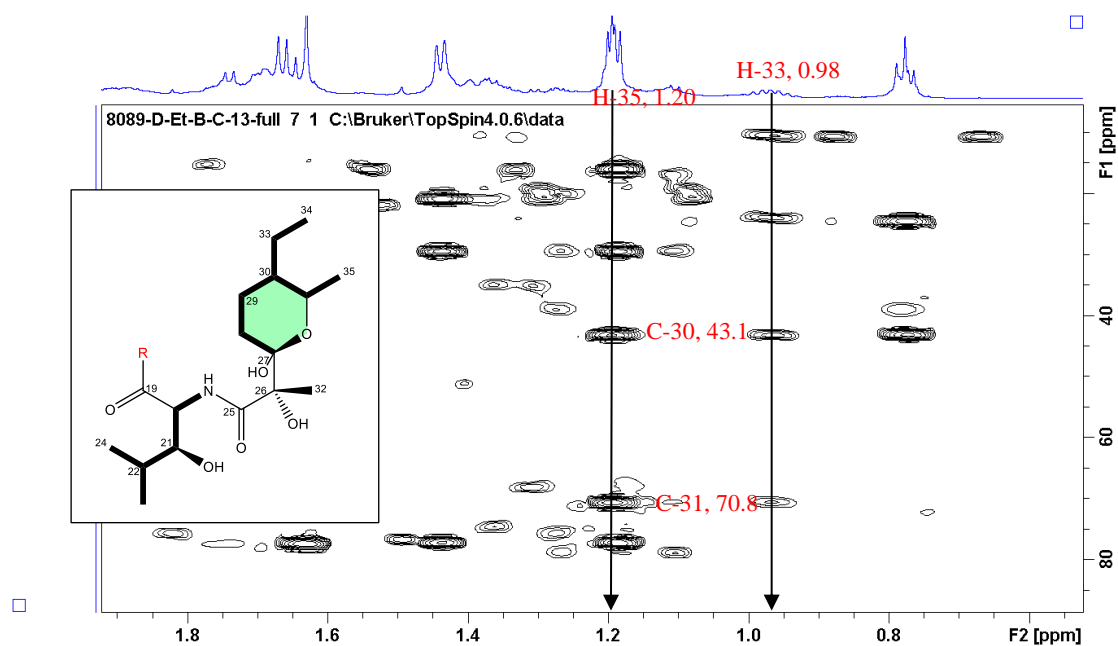
Figure 4-47. Expansion of ¹H-¹³C HMBC NMR spectrum of dentigerumycin F recorded at 600 MHz in pyridine-*d₅* demonstrating correlations from the methyl doublet at H-35 into the oxane ring and from the methylene at H-33 into the same carbons of the oxane ring.
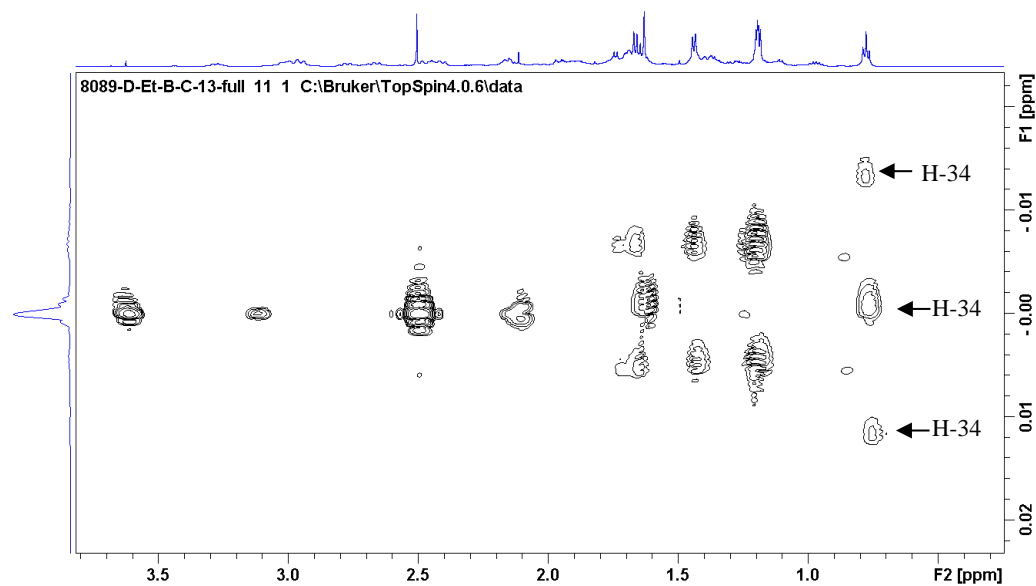
Figure 4-48. 2D *J*-resolved spectrum for dentigerumycin F. This helped confirm the identity of the substituents on the oxane ring via the triplet splitting pattern at H-34. Recorded at 600 MHz in pyridine-*d₅*.

			8089-13: Dentigerumycin F (4.2)					
			Pyridine-d_5		**DMSO-d_6**			
			Hemiketal Form		Hemiketal Form		Olefinic Form	
Residue	Position		δ_C/δ_N	δ_H (J = Hz)	δ_C/δ_N	δ_H (J = Hz)	δ_C/δ_N	δ_H (J = Hz)
L-Piz Amide	1	C	173.4		172.8		172.8	
	2	CH	51.5	5.50 m	50.3	4.82 dd (1.5, 5.8)	50.3	4.82 dd(1.49, 5.78)
	3 a/b	CH$_2$	26.3	1.75, 2.42 m	24.7	1.76, 2.06	24.7	1.76, 2.06
	4 a/b	CH$_2$	21.7	1.43, 1.40 m	21.3	1.45, 1.50	21.3	1.45, 1.50
	5 a/b	CH$_2$	47.7	2.67, 2.95 m	46.7	2.58, 3.00	46.7	2.58, 3.00
	NH-1	NH	-300.3	5.5	-301.0	4.99 m	-301.0	4.99 m
	CONH$_2$		-275.8	8.27, 8.09	-274.3	7.56/7.20	-274.3	7.56/7.20
D-Piz 2	6	C	173.9		172.0		172.0	
	7	CH	49.4	6.11 dd (2.9; 6.4)	47.3	5.53 m	47.3	5.53 m
	8 a/b	CH$_2$	25.8	1.88, 2.15 m	24.6	1.76,1.99 m	24.6	1.76,1.99 m
	9 a/b	CH$_2$	21.4	1.62, 1.46 m	20.7	1.41,1.50 m	20.7	1.41,1.50 m
	10 a/b	CH$_2$	47.4	3.01, 2.77 m	46.2	2.58, 2.96 m	46.2	2.58, 2.96 m
	NH-4	NH	-300.9	5.67	-300.1	5.11 m	-300.1	5.11 m
N-OH-L-Ala	C11	C	173.5		171.9			
	12	CH	54.2	6.19 m	53.3	5.53 m	53	5.55 m
	13	CH$_3$	16.1	1.67 d (6.8)	14.6	1.33 d (6.9)	14.8	1.35 d (6.8)
	N5		-197.6		-204.0			
	N5-OH			11.57 v. broad		9.54 br.s		
D-Piz 3	14	C	172.9		171.6		171.6	
	15	CH	49.8	6.24	47.8	5.47 m	47.8	5.47 m

	16 a/b	CH$_2$	25.4	2.48, 1.90 m	24.9	2.1, 1.77	24.9	2.1, 1.77
	17 a/b	CH$_2$	21.9	1.78, 1.43 m	21.6	1.49, 1.34	21.6	1.49, 1.34
	18 a/b	CH$_2$	46.9	3.27, 2.98 m	45.3	2.86 m	45.3	2.8 m
	NH-6	NH	-300.0	5.75	-300.0	5.1 m	-300.0	5.08 m
β-S-OH-L-Leu	19	C	174.2		170.9		170.9	
	20	CH	52.3	6.40 m	50.1	5.49 m	49.8	5.42 m
	21	CH	77.4	4.12 dd (2.7, 8.4)	75.8	3.39 m	75.6	3.41 m
	22	CH	29.6	2.45 m	27.8	1.85 m	28.8	1.69 m
	23	CH$_3$	16.0	1.44 d (6.87)	15.1	0.82 d (6.87)	14.7	0.82 d (6.87)
	24	CH$_3$	21.1	1.19 d (6.87)	20.2	0.83 d (6.87)	20.2	0.82 d (6.87)
	NH-8	NH	-264.9	8.71 d (9.38)	-264.9	7.70 d (9.38)	-267.0	7.23 d (10.02)
	21-OH			Not visible				
PKS	25	C	178.5		176.6		172.3	
	26	C	78.2		76.9		75.2	
	27	C	100.2		98.9		154.1	
	28 a/b	CH$_2$	28.0	1.95 m	28.0	2.1, 1.9 m	93.1	4.81
	29 a/b	CH$_2$	26.0	1.70 m	22.8	1.63, 1.53 m	24.1	1.68, 1.46 m
	30	CH	43.1	1.11 m	41.7	0.97 m	38.6	1.26 m
	31	CH	70.8	4.04 m	69.0	3.59 m	74.1	3.35 m
	32	CH$_3$	21.7	1.64 s	20.7	1.21 s	23.7	1.27 s
	33 a/b	CH$_2$	24.6	1.36, 0.98 q (6.9)	23.7	1.43, 0.99 m	24.1	1.06, 0.98 m
	34	CH$_3$	10.7	0.77, t (6.9)	10.7	0.82 t (6.9)	11.3	0.87 t
	35	CH$_3$	19.4	1.19, d (6.5)	18.6	0.99 d (6.5)	19.1	1.14 d (6.5)
	26-OH			Not visible		5.45		5.45
	27-OH			Not visible		5.88 s		

Table 4-2. NMR shifts of dentigerumycin F in both pyridine-d_5 and DMSO-d_6. Both hemiketal and dehydrated forms are seen in DMSO-d_6.

4.3.3 Investigations into the Equilibria Observed in Dentigerumycins F And G

Dentigerumycin F (**4.2**) and G (**4.1**) were initially isolated using HPLC conditions that included 0.05% TFA to remove peak tailing. The samples were also left at room temperature for up to thirty-six hours while removing the transfer solvent and solvent traces. NMR analysis of the natural products isolated under these conditions revealed multiple resonances for some atoms of the molecules (**Figure 4-49 a, Figure 4-51 and 4-52**). There was initial uncertainty if the signals were from separate atoms or different molecules. Obtaining a ^{1}H spectrum quickly after removing from storage at 4 °C demonstrated that specific chemical shifts were changing in intensity in cold temperatures (**Figure 4-49 b**). Next, a ^{1}H NMR spectrum from the
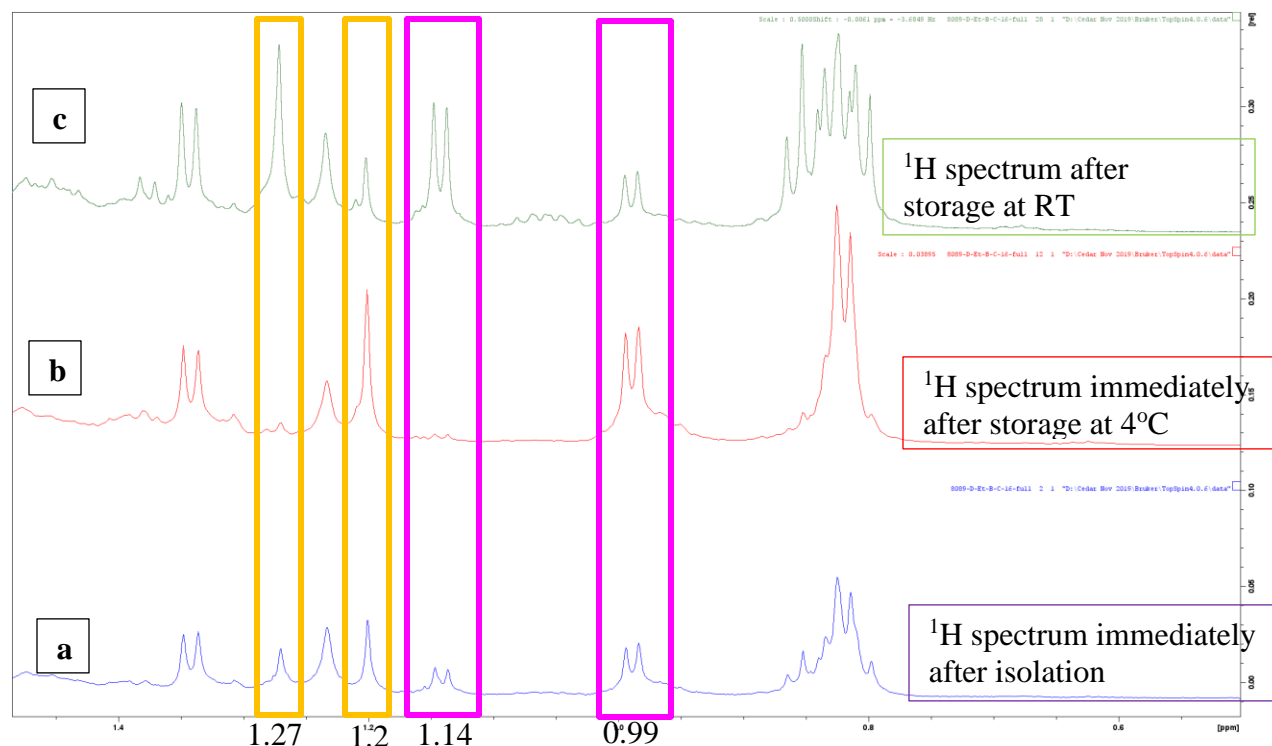
Figure 4-49. Comparison of proton spectra of 4.1 collected at various times and temperatures a) ^{1}H NMR spectrum obtained after isolation and sample preparation. b) ^{1}H NMR spectrum obtained immediately after storage in the fridge. c) ^{1}H NMR spectrum obtained after the sample was let sit at RT. The pink box highlights the H-35 methyl doublet, while the orange boxes highlight the methyl singlet at C-32. Recorded at 600 MHz in DMSO-*d6*.

sample sitting at room temperature revealed a return to the mixture of signals (as revealed by key proton peaks) (**Figure 4-49 c**). Significant changes to some δ_H and δ_C in the polyketide portion and the hydroxy leucine of the molecule were shown to be reversible depending on the temperature of the sample and the solvent in which they were stored (**Tables 4.1 and 4.2**; **Figures 4-50 to 4-52**). Cold temperatures favored the hemiketal form or natural form of **4.1**: resonances were seen for a quaternary carbon at δ_C 98.8 (C-27); a methyl singlet δ_C 21.5, δ_H 1.20 (C-32); a methyl doublet δ_H 0.99 (C-35); a methylene δ_C 26.54, δ_H 1.62 and 1.54 (C-28); and a
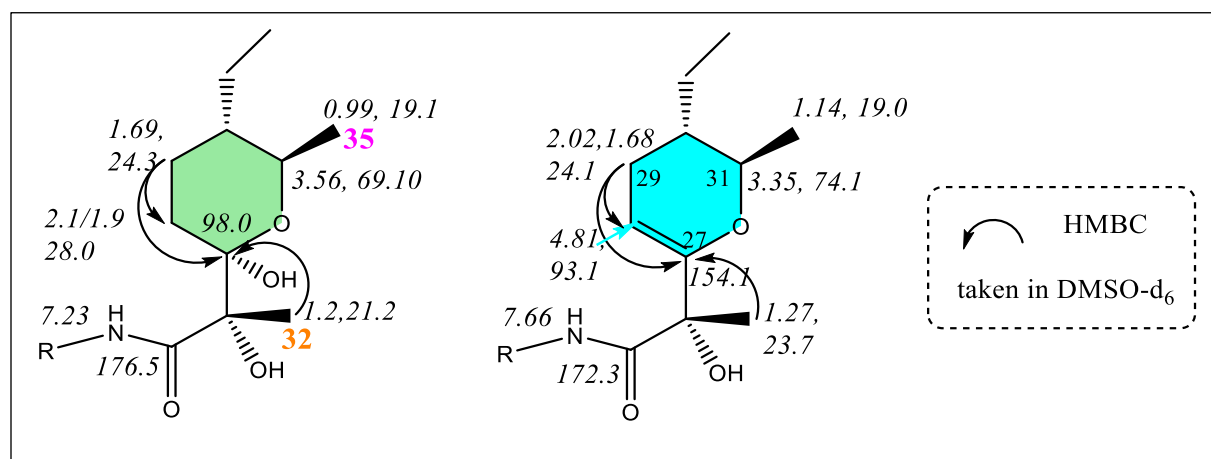
Figure 4-50. Oxane systems of dentigerumycin G (4.1) in DMSO-d_6, showing the resonances seen for each form. H-35 highlighted in pink, and H-32 highlighted in orange to match Figure 4-49.

methine at δ_C 69.1, δ_H 3.56 (C-31). When samples of **4.1** were left at room temperature (~26 °C) for the time needed to collect a data set (~20 hrs) or longer (3 or more days), those same resonances changed. In the dehydrated form, quaternary carbon C-27 was seen to resonate at δ_C 152.9; the C-32 methyl singlet changed to δ_H 1.27, δ_C 23.5; the methyl doublet at C-35 was deshielded (δ_H 1.27); C-28 now had an olefinic proton (δ_H 4.84) attached to a very deshielded carbon (δ_C 94.6), and the C-31 methine was found at δ_C 74.1, δ_H 3.35. The molecule was thus

seen to equilibrate between a hemiketal at C-27 and an olefin at C-27/C-28 depending on the

sample's temperature.
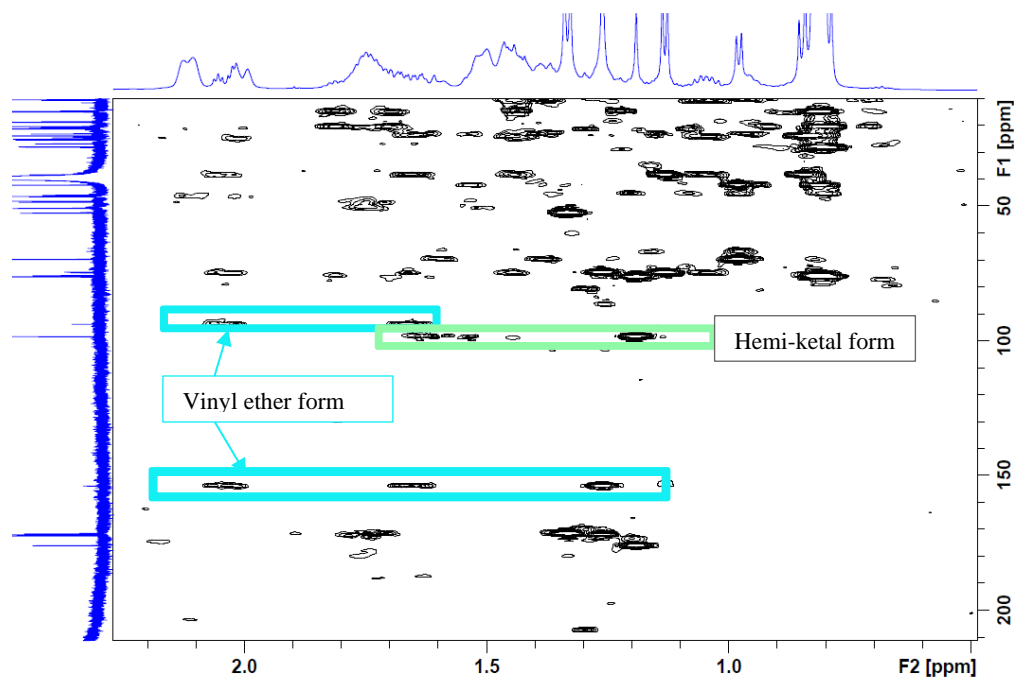
Figure 4-51. ^{1}H-^{13}C HMBC of room temperature dentigerumycin G (4.1) (isolated with TFA as HPLC pH buffer) showing HMBC correlations of both equilibrating forms. Recorded in DMSO-d_6 at 600 MHz.
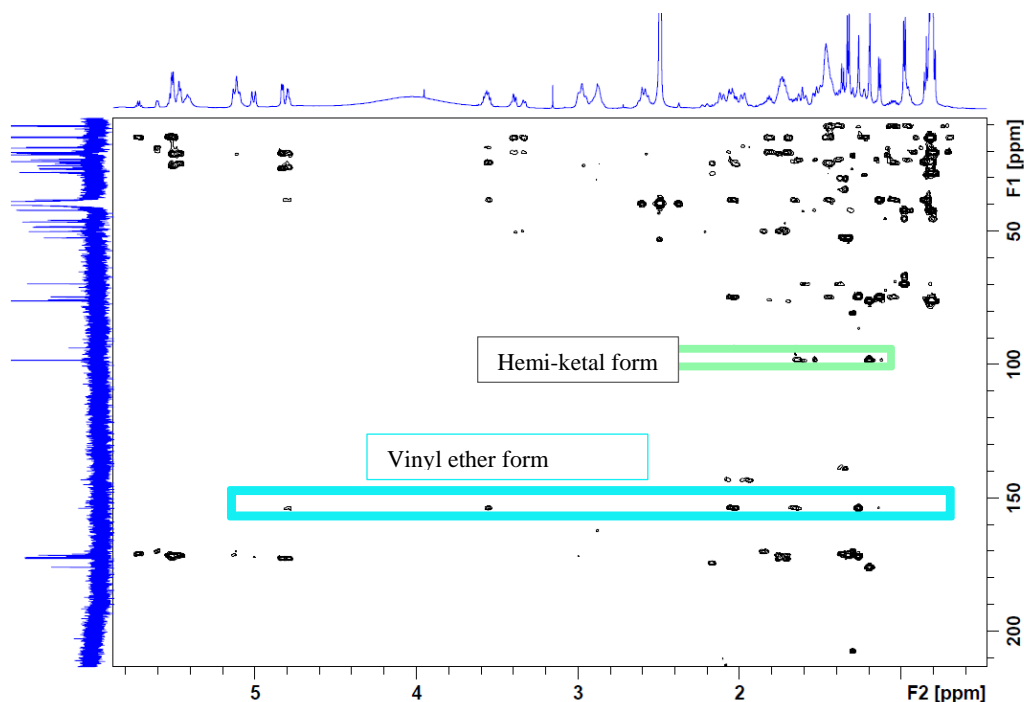
Figure 4-52. Full ^{1}H-^{13}C HMBC of room temperature dentigerumycin F (4.2) showing the HMBC correlations for both equilibrating forms. Recorded in DMSO-d_6 at 600 MHz.

The DMSO-d_6 (freezing point: 18.5 °C) solvent prevented obtaining NMR samples at cooler temperatures. Due to this, 2D spectra were challenging to obtain without resonances of both forms present. In order to attempt to obtain spectra at colder temperatures, the DMSO-d_6 solvent was removed via lyophilizer, after which CD$_3$OD was added to the NMR tube containing the sample. Proton spectra of this sample showed that all residual TFA from initial isolation was also removed through this process. NMR data of this sample recorded at room temperature in CD$_3$OD showed a majority of the dehydrated form of the molecule, with the C-31 methine correlation δ_C 74.1, δ_H 3.35 displaying an intense peak as well as C-28 as deshielded olefinic proton (**Figure 4-53; Figure 4-56 to Figure 4-57**). With the addition of 50 µL of D$_2$O to CD$_3$OD and data collection at 10 °C, the equilibrium favours the hemiketal form in an approximately 2:1 fashion. The C-31 methine correlation is observed at δ_C 69.1, δ_H 3.56 and the

168

correlation at C-28 as δ_H 4.81, δ_C 93.1 is no longer present in strong intensity (**Figure 4-54;**

Figure 4-56 to Figure 4-57).
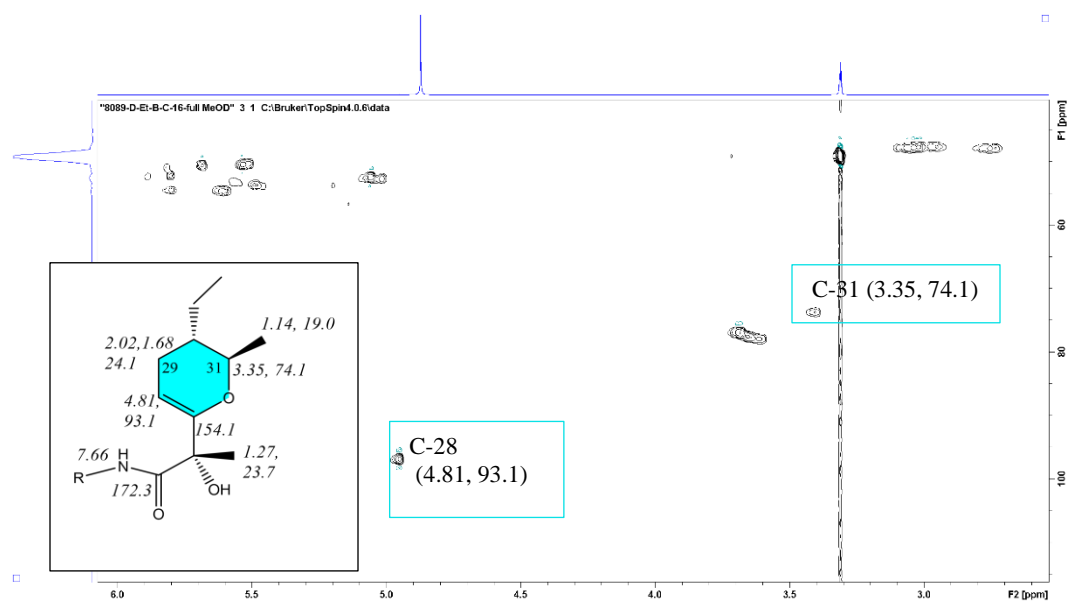
Figure 4-53. Expansion of ^{1}H-^{13}C HSQC of dentigerumycin G (4.1) in CD₃OD (at 600 MHz) and spectrum recorded at RT demonstrating a strong olefin signal at C-28.
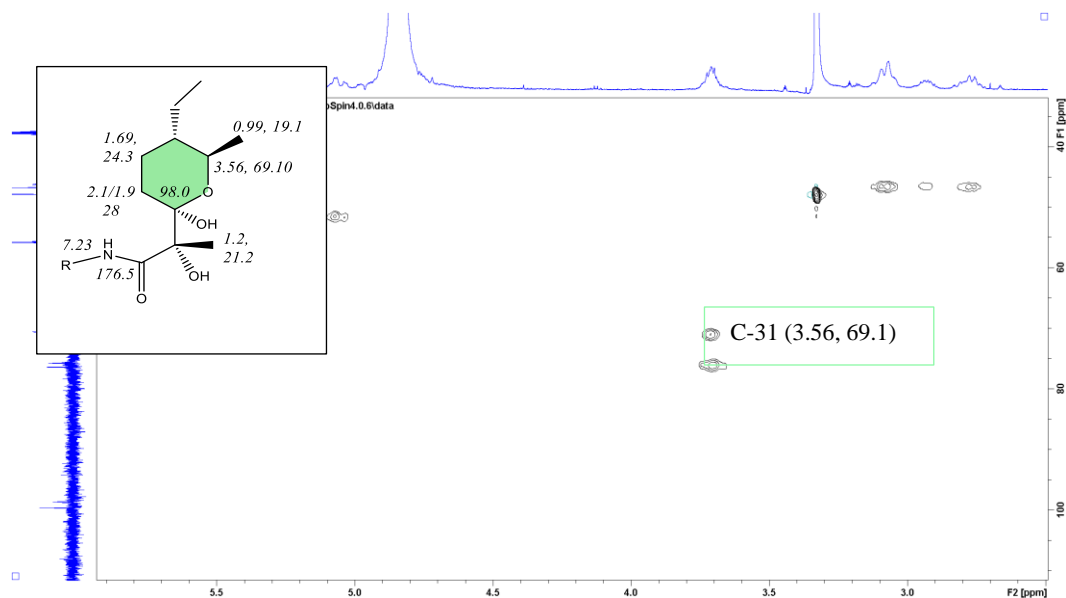
Figure 4-54. Expansion of ^{1}H-^{13}C HSQC of dentigerumycin G (4.1) in CD$_3$OD with 50 uL of D$_2$O added and spectrum recorded at 10 ºC where the olefinic correlation at C-28 is not seen. Recorded at 600 MHz.
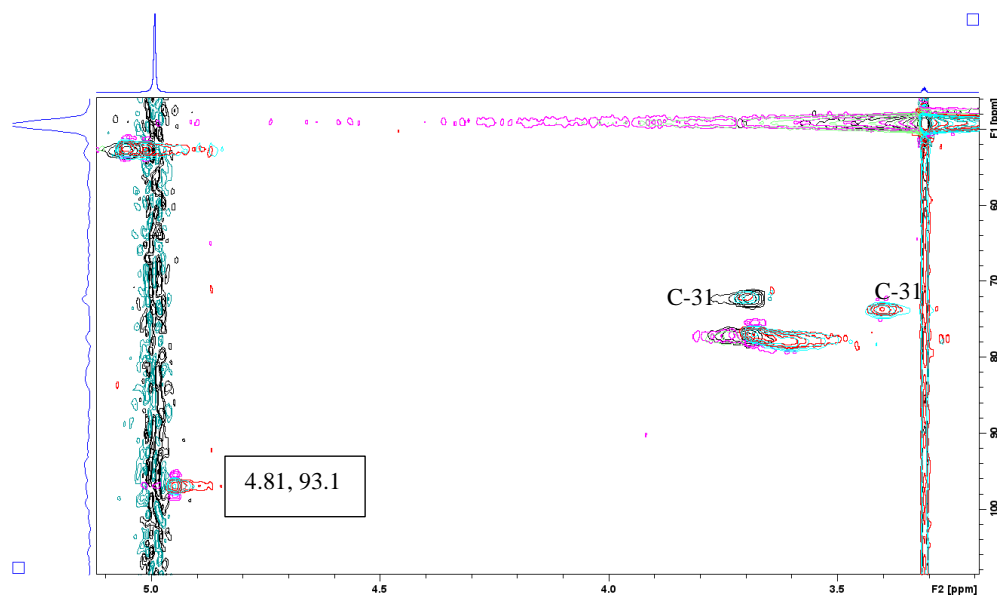
Figure 4-55. Expansion of overlapping ^{1}H-^{13}C HSQC spectra of 4.1 in CD$_3$OD. Red/pink correlations are found in CD$_3$OD let sit at RT. Teal correlations are found in CD$_3$OD after sitting in the fridge for a week. Black correlations are found in CD$_3$OD with 50 uL of D$_2$O added with spectra recorded at 10 ºC. Recorded at 600 MHz.
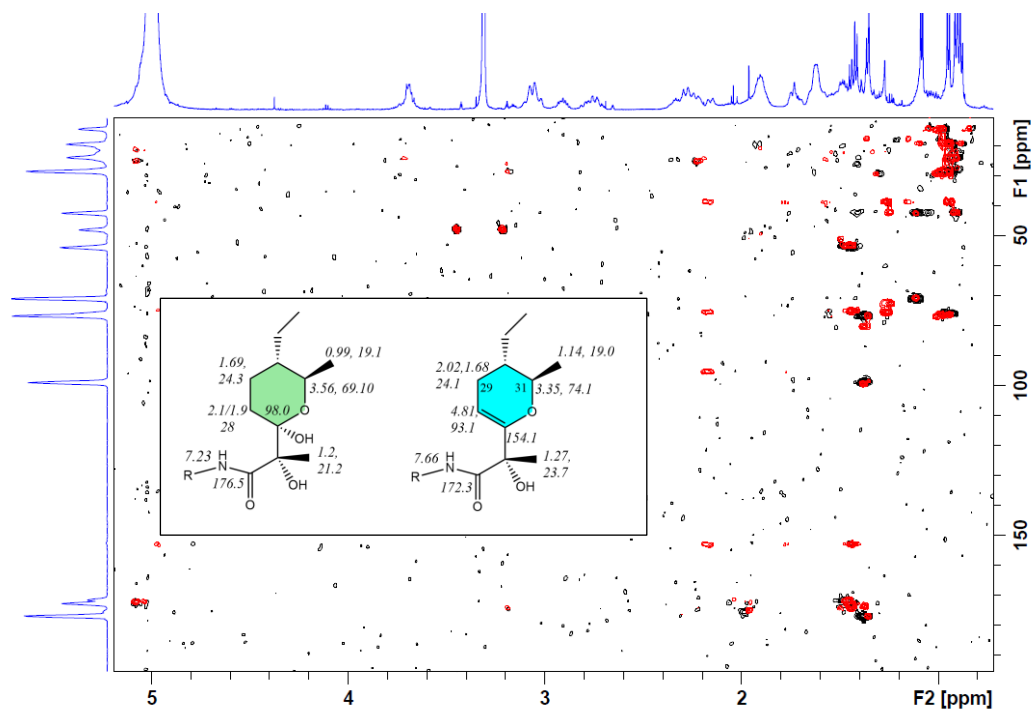
Figure 4-56. Overlapping ^{1}H-^{13}C HMBC NMR spectra of dentigerumycin G (4.1) in CD₃OD. Red correlations are found in CD₃OD let sit at RT. Black correlations are found in CD₃OD with 50 uL of D₂O added with spectra recorded at 10 °C, 600 MHz. Distinct differences are observable.

Hemi-Ketal Form of **4.1** [M+Na] m/z 807.5

Enol-like Form of **4.1** [M+H] m/z 767.4
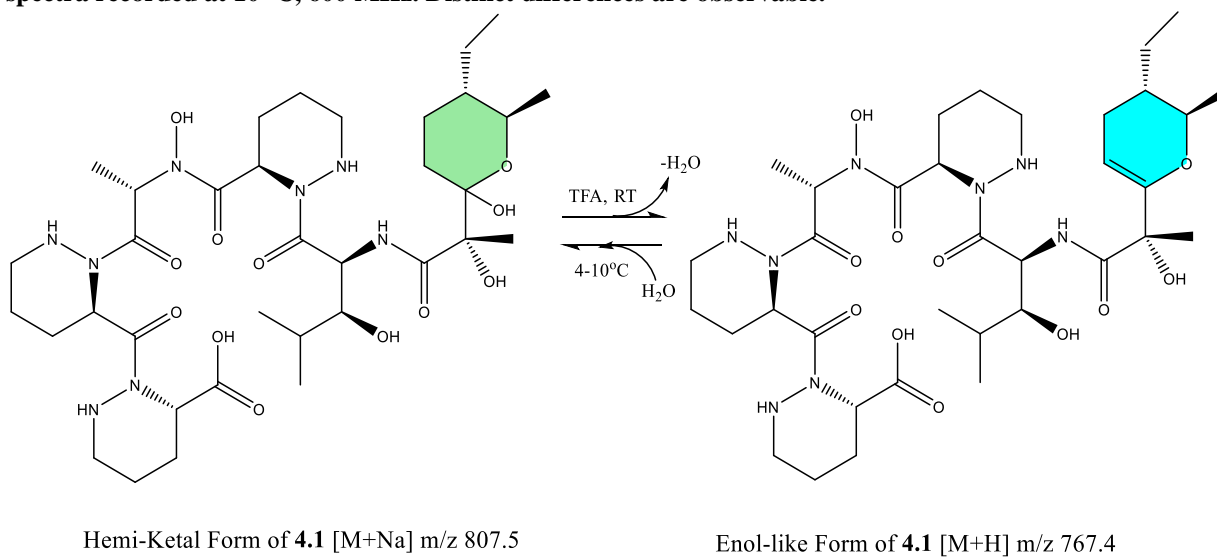
Figure 4-57. Scheme of equilibrium seen in both (4.1) and (4.2) when isolated using 0.05% TFA as a pH buffer during HPLC purification and stored in DMSO-*d₆*. *Not shown in a caption:* CD₃OD with all acid removed favors the dehydrated form, addition of small amounts of deuterated water without the addition of further acid favors the hemiketal form once again.

Overall, both the primary amide (**4.2**) and the free acid (**4.1**) were seen to engage in an equilibrium between hemiketal and vinyl ether forms when isolated with 0.05 TFA% in the HPLC solvent. In DMSO-d_6 at RT, there is an approximate 1:1 ratio between the two forms seen in data sets for **4.1** collected after prolonged RT exposure (**Figure 4-55**). In CD$_3$OD, without the presence of water, the equilibrium more strongly favours the dehydrated form. With the addition of deuterated water to the CD$_3$OD and chilling of the sample, the hemiketal form is dominant. More experiments are required to understand the mechanism driving this transformation.

4.2 and **4.1** were also seen to be undergoing an occasional equilibrium between a hemiketal and ketone form (weakly) as seen in **Figure 4-58.** The hemiketal/ketone equilibration, similar to that shown in verucopeptin,[239] was minor and visible in the NMR spectra of **4.1** only through the appearance of very weak HMBC correlations from the methyl singlet at C-32 (δ_H 1.30, δ_C 21.3) into a ketone carbonyl shift, δ_C 206.9. While verucopeptin has been previously shown to demonstrate an equilibrium between a ketone and hemiketal form[239,240] (a molecular property recently harnessed by researchers to synthesize verucopeptin and some novel derivatives[180]); the equilibrium between a hemiketal and an vinyl ether form at room-temperature has never been demonstrated for an oxane ring with piperazic-containing compounds or any natural products that I could uncover. As mentioned, isolation of (**4.1**) and (**4.2**) with formic acid (FA) as pH modifier proved to produce NMR evidence of only the stable hemiketal form (**Appendix B**).
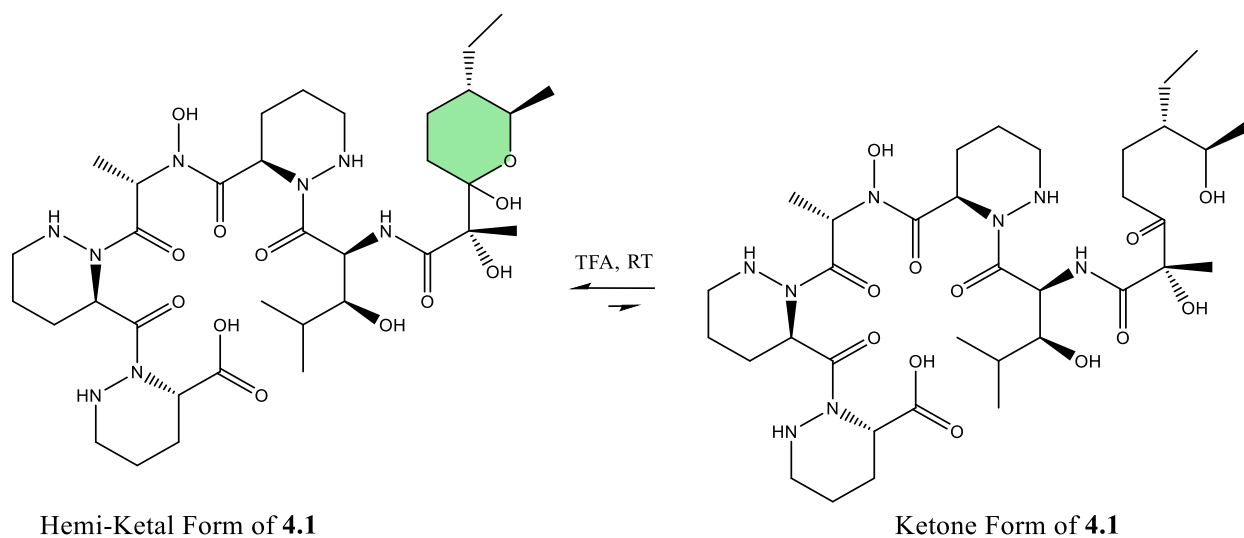
Figure 4-58. The tautomerization in 4.1 between a hemiketal and ketone form.

4.3.4   Structure Elucidation of Incarnatapeptin A (4.3)

After working on the structures of the more abundant **4.1** and **4.2**, the proton spectrum of incarnatapeptin A (**4.3**) was collected at multiple times throughout 2-D data set collection with multiple ^{1}H spectra showing no fluctuations in proton peaks. This observation suggested that incarnatapeptin A (**4.3**) had a unique polyketide region compared to **4.1** and **4.2,** which drove interest in solving the structure despite the minimal material available. In order to corroborate that this HPLC peak was indeed a Piz-containing compound, initial work done to solve the structure with available material was able to show exchangeable protons at δ_H 4.74 (δ_N -301.1), and 5.12 (δ_N -300.5) through a one-bond correlation between proton and nitrogen resonances in the ^{1}H-^{15}N HSQC spectrum **(Figure 4-61)**.

Piz 3

N-OH Ala

PKS

Piz 2

Incarnatapeptin A (**4.3**)

Piz 1
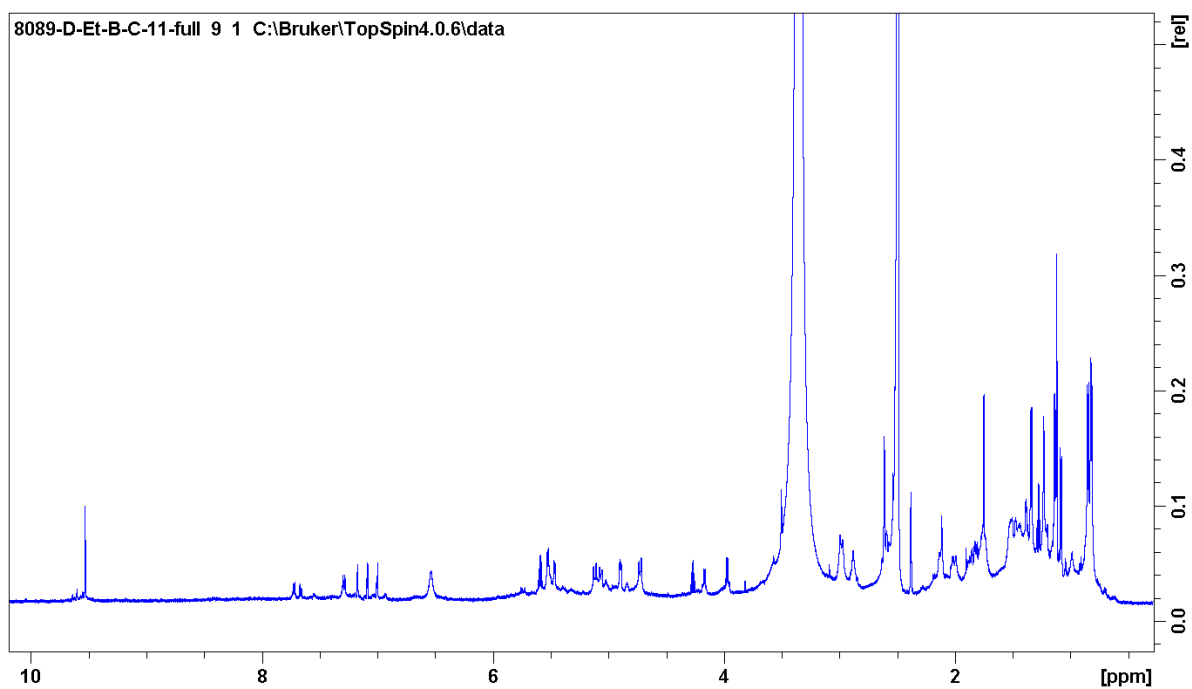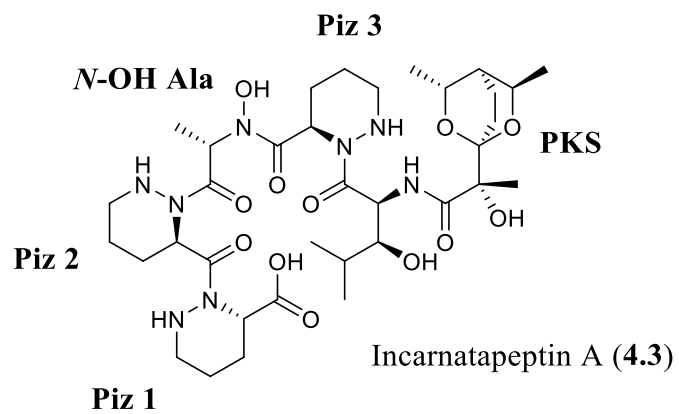
Figure 4-59. ^{1}H spectrum for author isolated incarnatapeptin A (**4.3**) in DMSO-d_6 at 600 MHz.

174
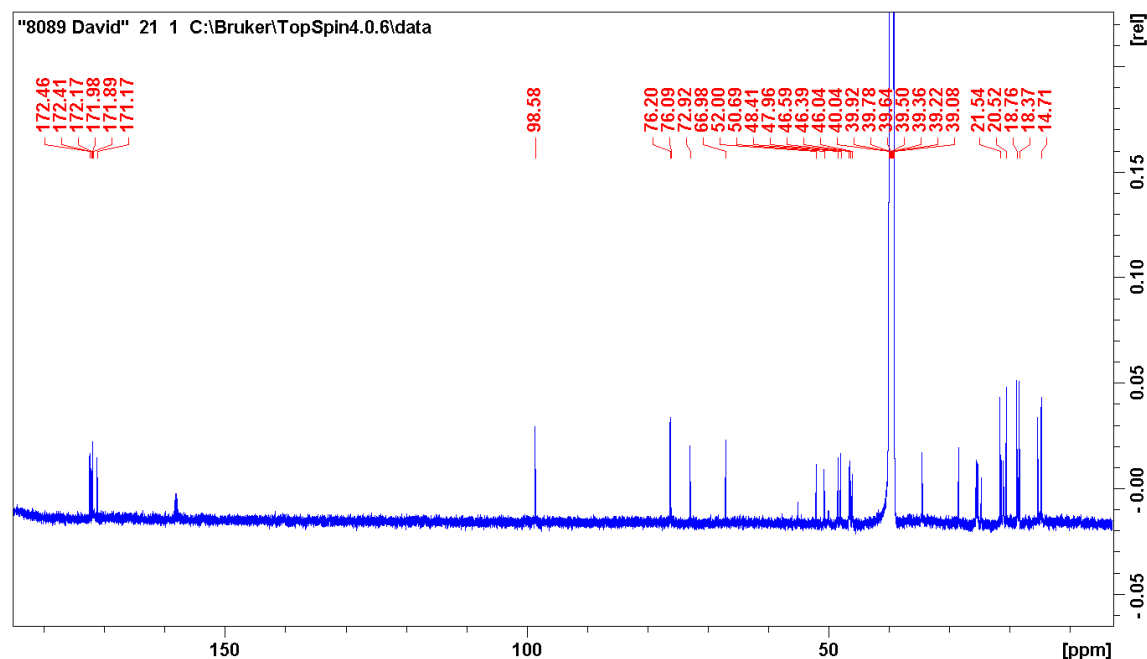
Figure 4-60. ^{13}C spectrum collected by author for incarnatapeptin A in DMSO-d_6 at 150 MHz.

Incarnatapeptin A (**4.3**) gave a [M - H]$^-$ ion at *m/z* 781.4097 and a [M + Na]$^+$ ion at *m/z* 805.4075 in the HRESIMS appropriate for a molecular formula of $C_{35}H_{58}N_8O_{12}$ that requires 11 sites of unsaturation. Examination of the NMR data confirmed that **4.3** is a piperazic acid-containing nonribosomal peptide-polyketide (NRP-PK) hybrid molecule with the ^{1}H, gTOCSY/gHSQC/gNHSQC/gHMBC/gNHSQC-TOCSY NMR resonances indicating the presence of multiple piperazic-acid spin systems, an N-hydroxylated alanine, a β-OH-leucine, and a region of clear PKS origin.
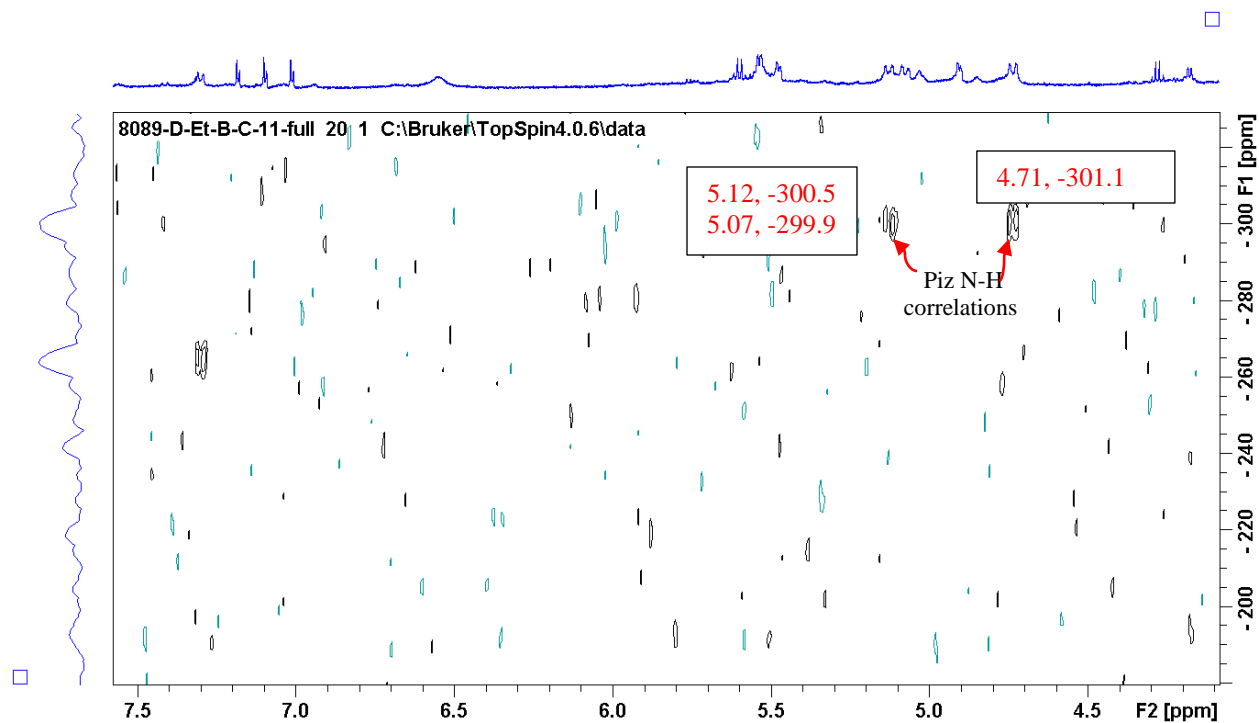
Figure 4-61. ^{1}H-^{15}N HSQC spectrum for author isolated incarnatapeptin A in DMSO-d_6, at 600 MHz showing the Piz N-H correlations.
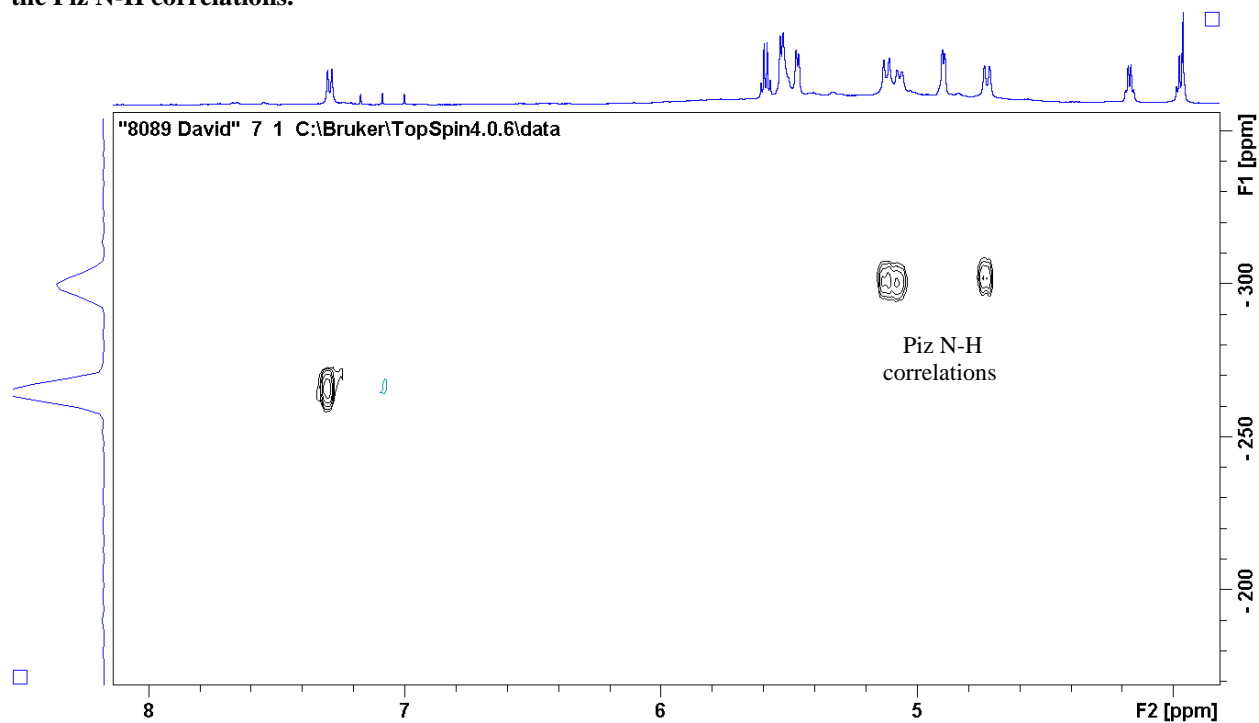
Figure 4-62. ^{1}H-^{15}N HSQC spectrum collected by author for incarnatapeptin A in DMSO-d_6 at 600 MHz.

176
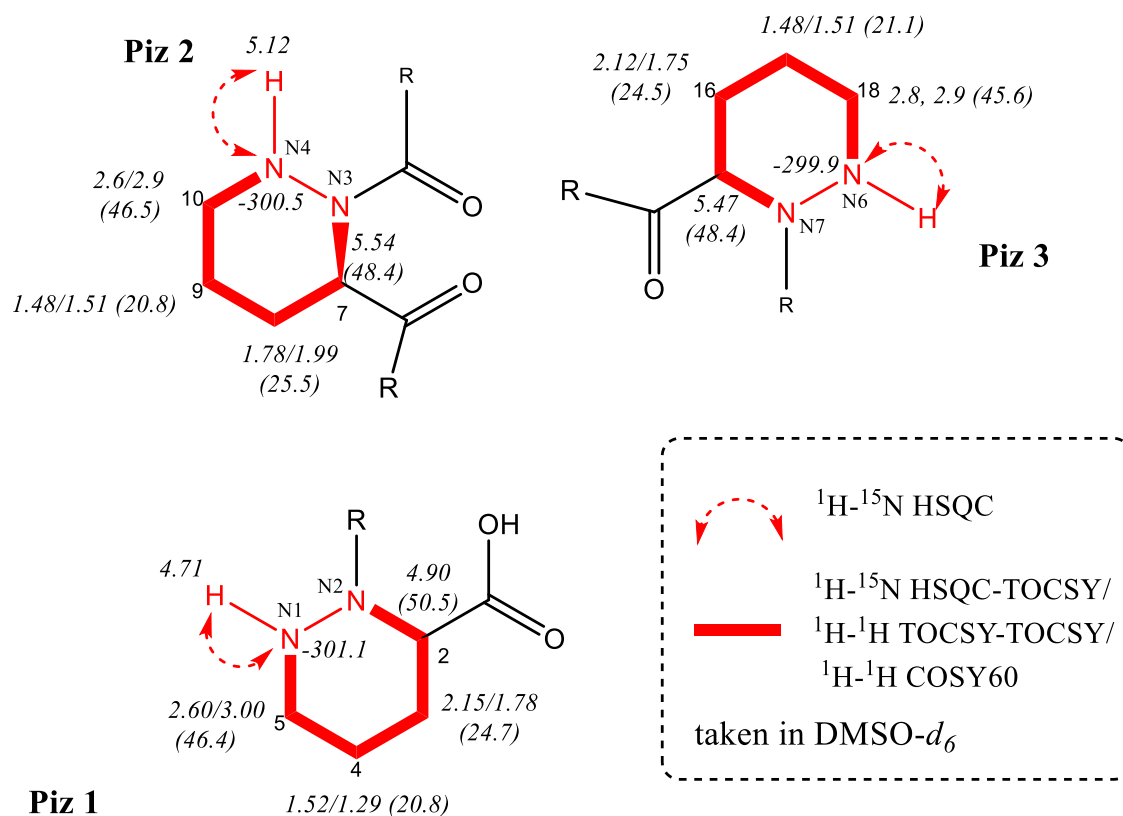
Figure 4-63. Chemical shift assignment for the three Piz amino acids of incarnatapeptin A.

In the ^{1}H-^{1}H TOCSY spectrum (**Figure 4.65**), exchangeable protons were shown to have typical Piz spin systems with δ_H 4.74 (NH-1) demonstrating correlations into H-5a/b, H-4a/b, H-3a/b, and H-2a/b; δ_H 5.07 (NH-4) showing correlations into H-10a/b, H-9a/b, H-8a/b, and H-7a/b; while δ_H 5.12 (NH-6) had correlations into the methylene protons of H-18, H-17, H-16, and H-15 (**Figure 4.63**). A carboxylic acid-like ^{1}H broad resonance at δ_H 13.26 suggested the presence of a free acid. With spectra recorded from more material isolated by David Williams (**Figure 4.62; Appendix C**), the third exchangeable proton resonance was confirmed at 5.07 (δ_N -299.4) in the ^{1}H-^{15}N HSQC, and piperazic acid spin systems were supported with a ^{1}H -^{15}N HSQC-TOCSY spectrum (**Figure 4.64**). With the initial materials, the HSQC spectrum (**Figure 4.66**) supported this compound as a Piz-containing natural product with typical deshielded

resonances C-5 (δ_C 45.6), C-10 (δ_C 46.50) and C-18(δ_C 46.35) for methylene carbons attached to the Piz N-H's.
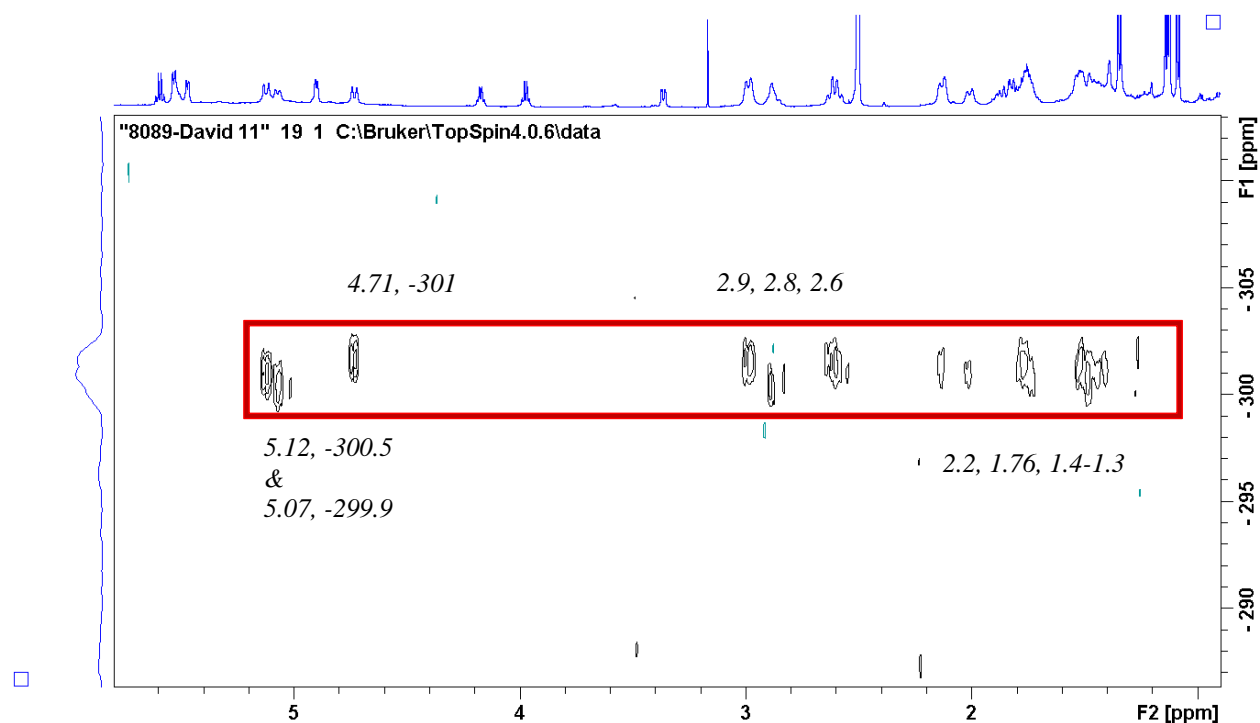
Figure 4-64. ^{1}H-^{15}N HSQC-TOCSY spectrum collected by author for incarnatapeptin A in DMSO-d_6 at 600 MHz. It illustrates the characteristic Piz spin systems.
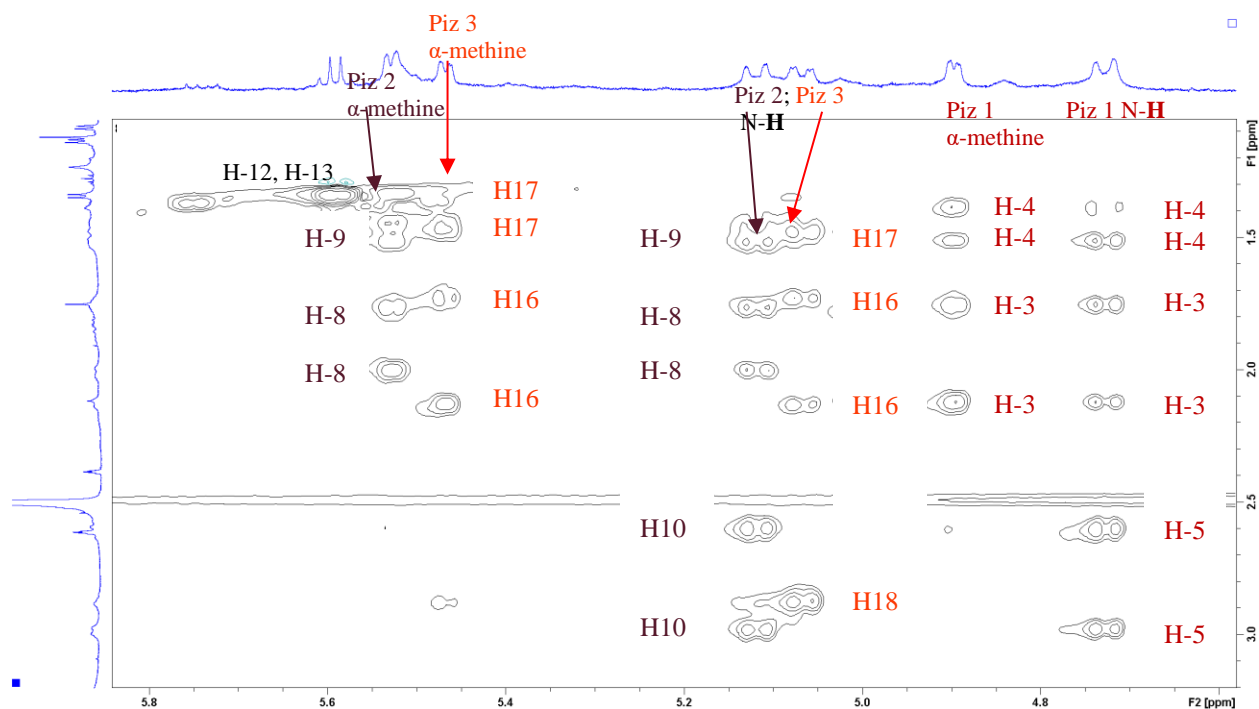
Figure 4-65. Expansion of ^{1}H-^{1}H TOCSY spectrum for author isolated incarnatapeptin A, recorded in DMSO-d_6 at 600 MHz, demonstrating the Piz spin systems for all three piperazic acids. Also seen are the correlations between the methyl (H-13) and the α-methine (H-12) of the *N*-OH-Ala amino acid.
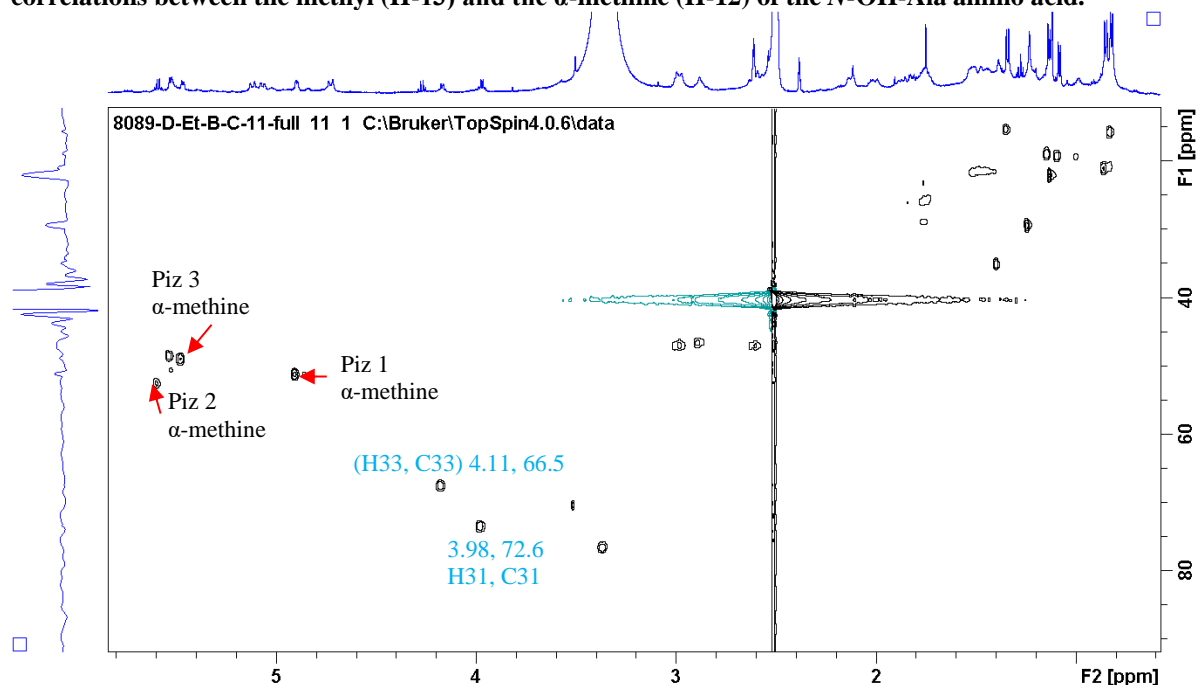
Figure 4-66. ^{1}H-^{13}C HSQC spectrum for author isolated incarnatapeptin A in DMSO-d_6 at 600 MHz, where the three Piz α-methine resonances with ^1J correlations into their respective carbon resonances are shown. Piz 3 and Piz 2 have unusually deshielded α-methines in line with other characterized Piz-containing natural products. In turquoise; the similar proton-carbon correlations of the bicyclic PKS ring.

179
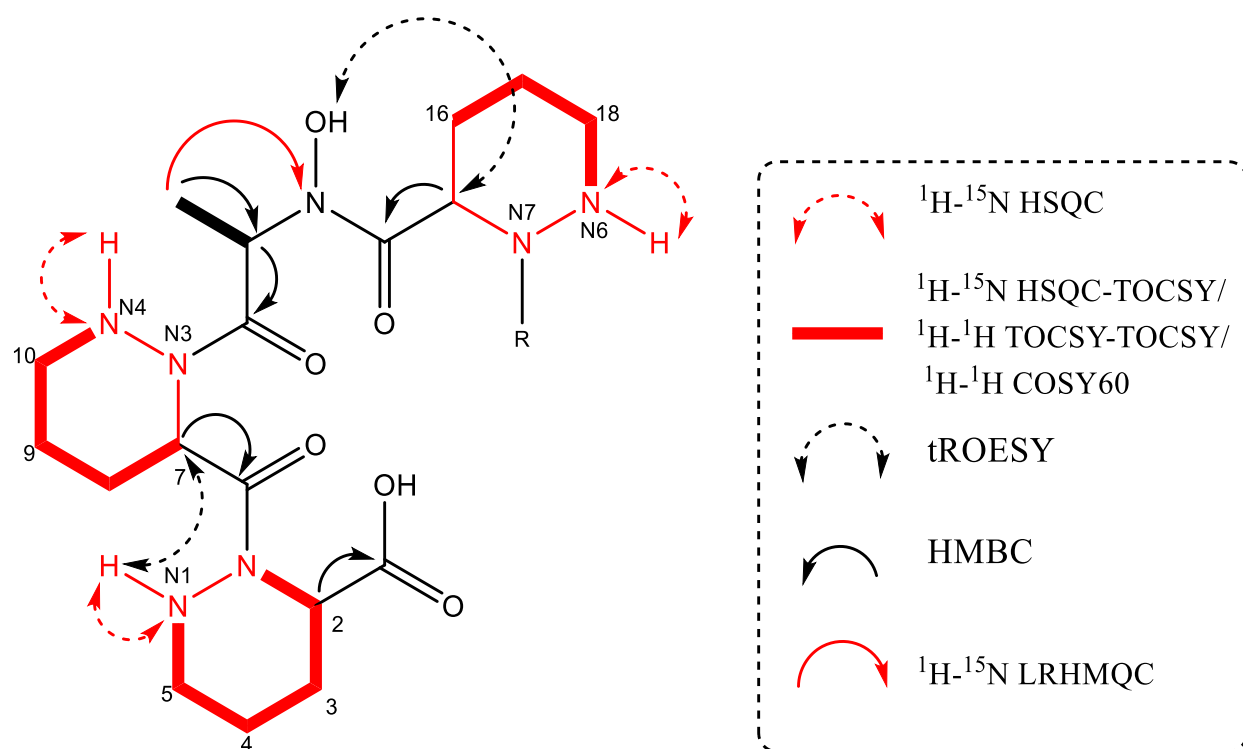
Figure 4-67. Connecting the core amino acids for incarnatapeptin A.

The presence of an N-hydroxylated alanine - identical to that seen in incarnatapeptin C and D -was confirmed via COSY60, HMBC, and ^{1}H-^{15}N LRHSQC correlations. In the COSY60 spectrum, a methyl doublet at δ_H 1.35 (H-13) correlating to a deshielded α-methine proton resonating at δ_H 5.59 (H-12) suggested an Ala amino acid (**Figure 4-69**). An ^{1}H-^{13}C HMBC correlation between the methyl doublet (δ_H 1.67, H-13) and the deshielded α-carbon (δ_C 54.2, C-12) confirmed this assignment (**Figure 4-74**). A hydroxymate-like proton resonance at δ_H 9.53 (**Figure 4-59**) suggested the presence of an *N*-OH-Ala, which was confirmed by ^{1}H-^{15}N LRHSQC correlations between the methyl doublet and the shielded nitrogen (δ_H 1.67, δ_N -200.6) (**Figure 4-71**). tROESY correlations between adjacent α-methine's and either the nitrogen-attached Piz proton or the nitrogen-attached hydroxyl of Ala were instrumental in building the correct placement of the amino acids (**Figure 4-67 and Figure 4-68**). At this juncture, the

limited material did not provide sufficient data to link *N*-OH-Ala and Piz 2. However, the

similarity of the amino acid data of **4.3** to **4.1** strongly suggested that the **4.3** had the same amino

acid sequence as **4.1,** for which evidence linking the *N*-OH-Ala and Piz 2 were present.

Subsequently, a tROESY correlation between the α-methine of *N*-OH-Ala and Piz 2 was

confirmed from data collected on **4.3** material that David Williams obtained (**Appendix C**).

Thus, the core amino acid sequence of incarnatapeptin A (**4.3**) was confirmed as a free acid Piz,

Piz, *N*-OH-Ala, Piz (**Figure 4.67**).
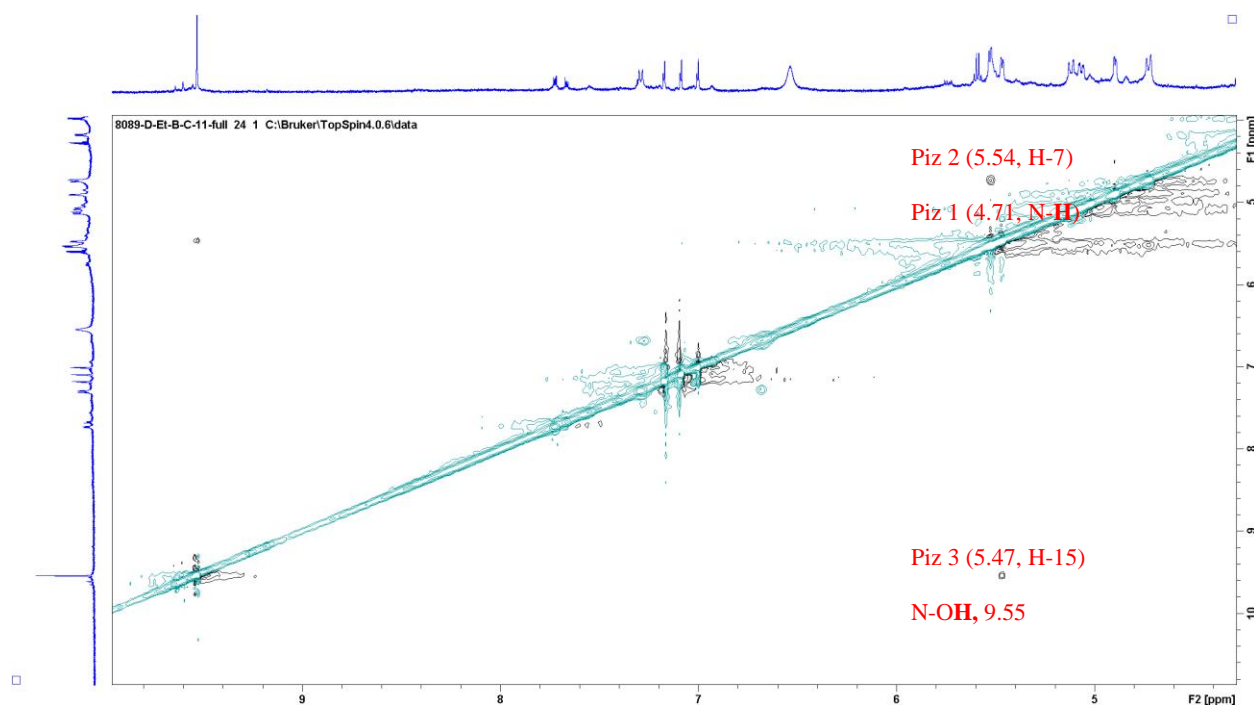
Figure 4-68. tROESY correlations between the protons of the nitrogen-attached hydroxyl/proton and α-methine proton leading to the assigned structure of 4.3. Recorded in DMSO-*d₆* at 600 MHz.
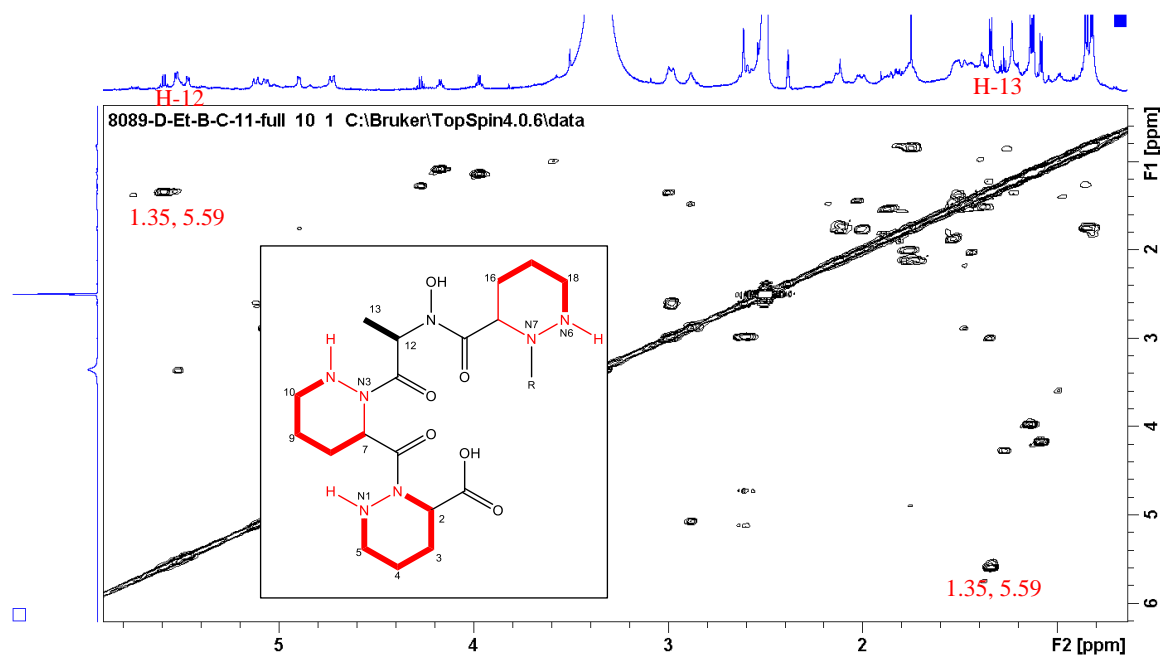
Figure 4-69. ^{1}H-^{1}H COSY spectrum for author isolated incarnatapeptin A in DMSO-d_6 at 600 MHz.
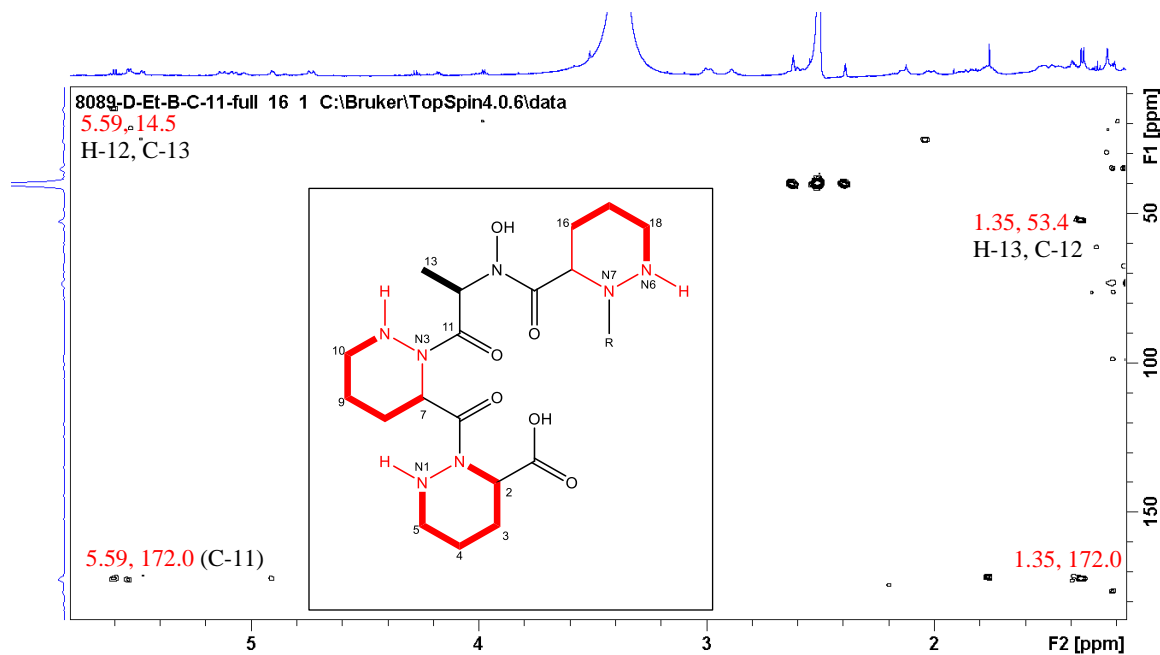
Figure 4-70. ^{1}H-^{13}C HMBC spectrum for author isolated incarnatapeptin A in DMSO-d_6 at 600 MHz.

182
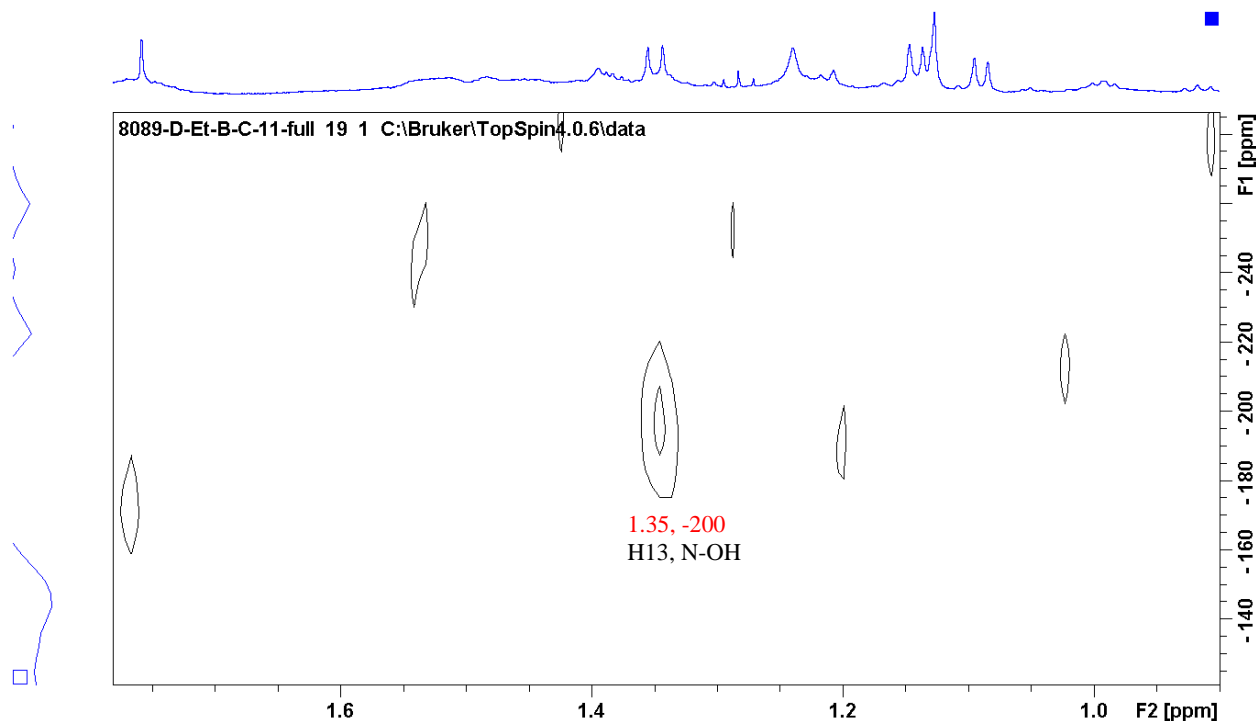
Figure 4-71. ^{1}H-^{15}N LRHMQC spectrum for author isolated incarnatapeptin A in DMSO-d_6 at 600 MHz demonstrating the correlation between the methyl group of the *N*-OH-Ala and the nitrogen resonance of the hydroxamate moiety.

Further analysis of the 2D data revealed the standard Piz-containing 19-membered depsipeptides β-OH-Leucine with methyl doublets at C-23 (δ_C 15.13, δ_H 0.82) and C-24 (δ_C 20.2, δ_H 0.86) showing HMBC correlations from δ_H 0.86 into C-21 (δ_C 75.82) (**Figure 4.74)**, and the g^{1}H-^{1}H COSY spectrum revealing three adjacent methines at H-20 (δ_H 5.50), H-21 (δ_H 3.36), and H-22 (δ_H 1.75) as well as a cross peak from H-20 to an amide proton (NH-8; δ_H 7.29) (**Figure 4.72 and 4.73)**. Analysis of the 2D data demonstrated that the PKS linkage between the β-OH Leu and the oxane ring was present with a methyl singlet at C-32 (δ_C 21.16, δ_H1.12) showing HMBC correlations into a carbinol carbon at C-26 (δ_C 76.40) (**Figure 4.74**). No other amino acid signatures were present; thus, **4.3** was seen to be a linear peptide like **4.1** and **4.2**.
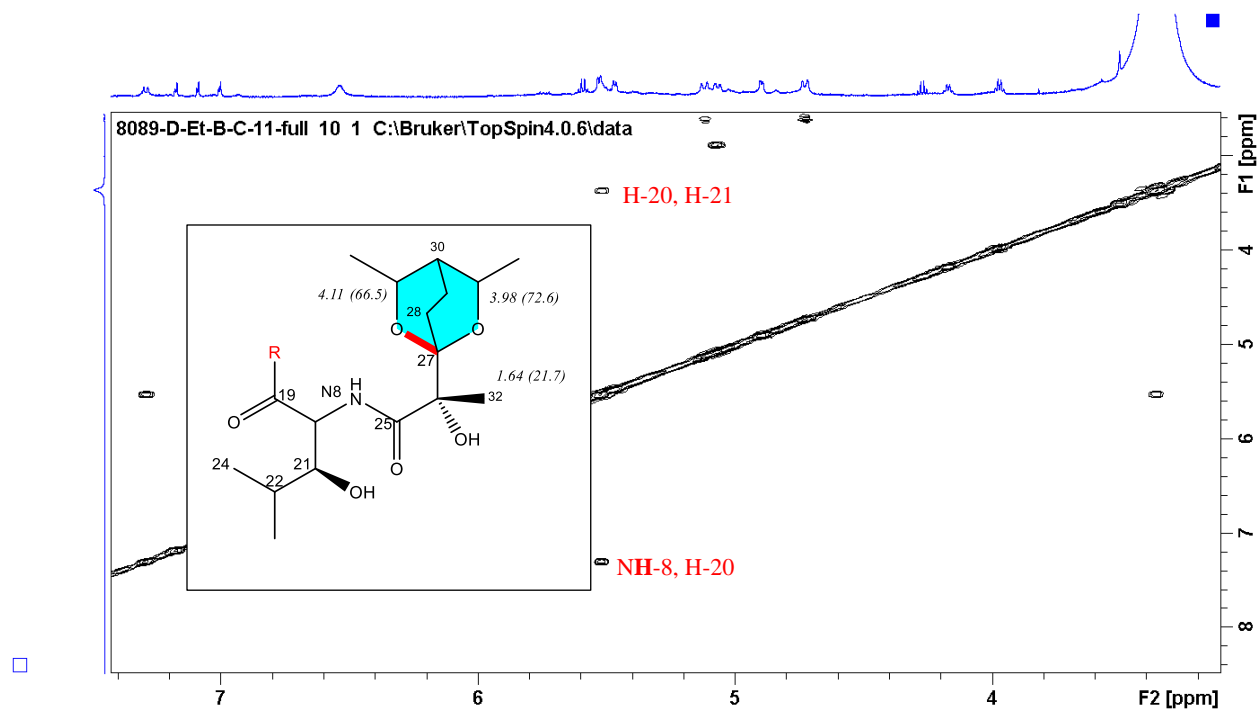
Figure 4-72. ^{1}H-^{1}H COSY spectrum for author isolated incarnatapeptin A in DMSO-d_6 at 600 MHz. Shown are the proton correlations in the conserved leucine residue.
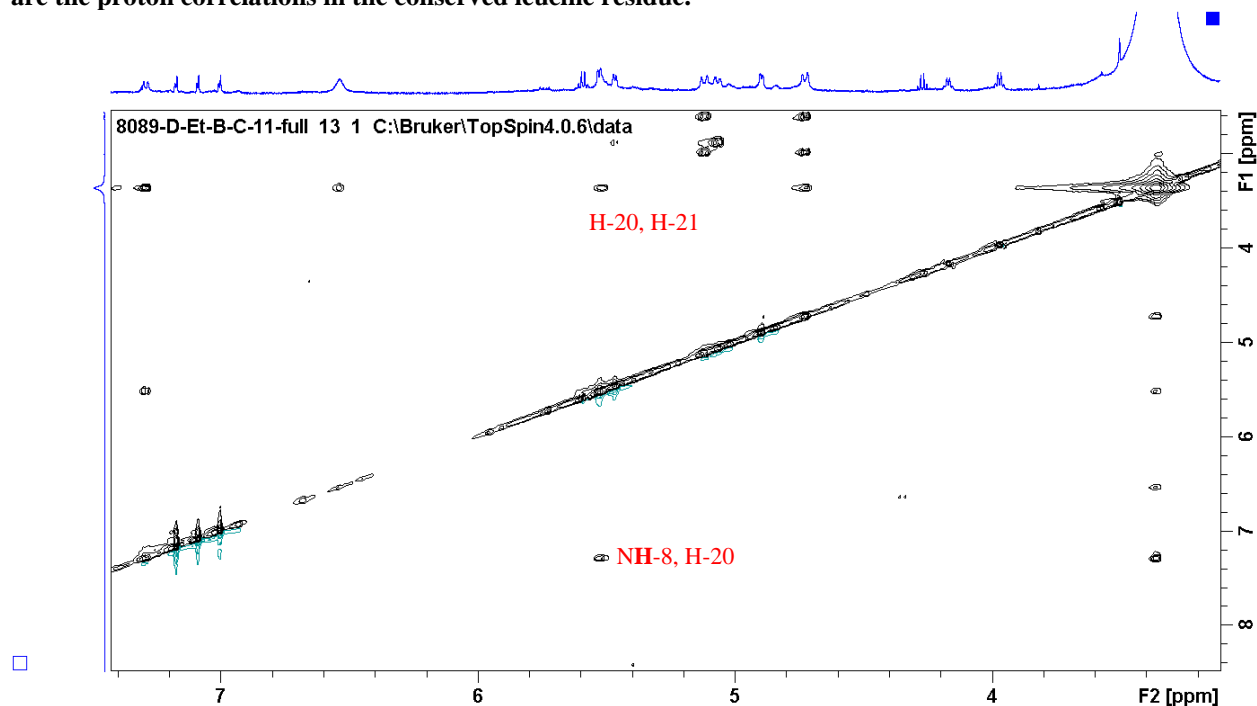
Figure 4-73. ^{1}H-^{1}H TOCSY spectrum for author isolated incarnatapeptin A, recorded in DMSO-d_6 at 600 MHz. The correlations supporting the construction of the leucine amino acid are seen.

184
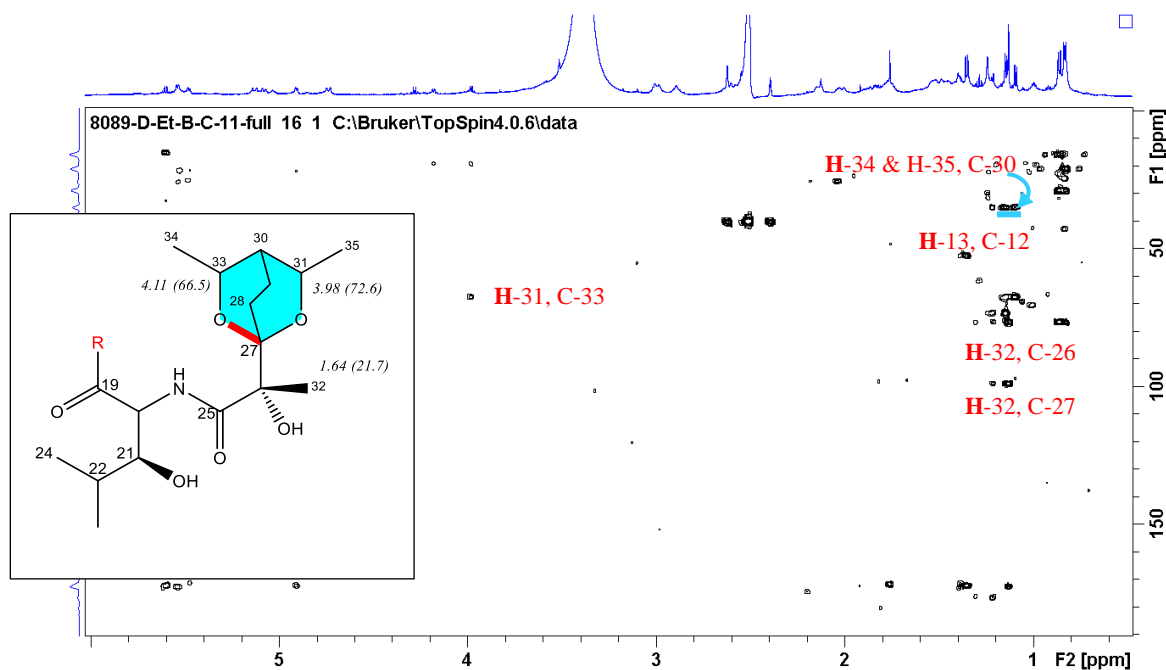
Figure 4-74. Full ^{1}H-^{13}C HMBC spectrum for author isolated incarnatapeptin A, recorded in DMSO-d_6 at 600 MHz. and showing key HMBC correlations.
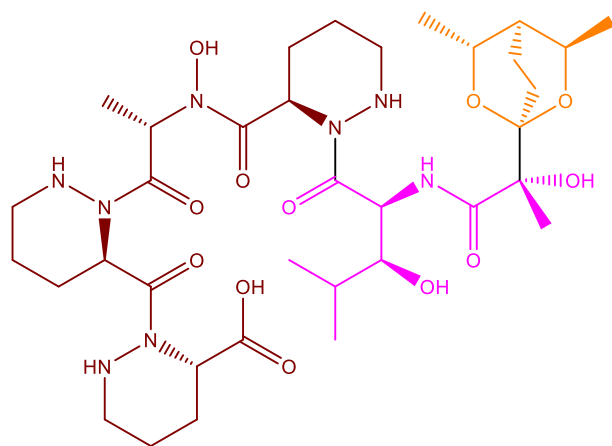
Fragment A : $C_{18}H_{30}N_7O_6$
Fragment B : $C_9H_{16}NO_4$
Fragment C : $C_8H_{13}O_2$

Figure 4-75. Building up fragments of incarnatapeptin A to help confirm the structure of the PKS portion.

Considering an *m/z*: [M-H] = 781.4097, mass analysis of the solved fragments seen in **Figure 4.75** left a small fragment C with eight carbons, thirteen protons, and two oxygen atoms to be accounted for. Initially, however, due to the low amount of compound, the low S/N ratio in the mass spectrum cast uncertainty on the accuracy of this *m/z*: [M - H] = 781.4097. Nevertheless, it was immediately evident that there was a substantial difference in the oxane ring's substitution pattern, with resonances demonstrating a somewhat symmetrical PKS ring system. Indeed, there was a third methyl doublet in the overall structure (*N*-OH-Ala methyl doublet, C-35 methyl doublet); this methyl doublet at δ_H 1.09 was showing COSY correlations into a deshielded methine proton at δ_H 4.11 (δ_C 66.53). This correlation was very similar to the H-35/H-31 relationship, as shown in the COSY spectrum in **Figure 4.76** and the deshielded carbon at δ_C 66.53 suggested connection to an electronegative oxygen.
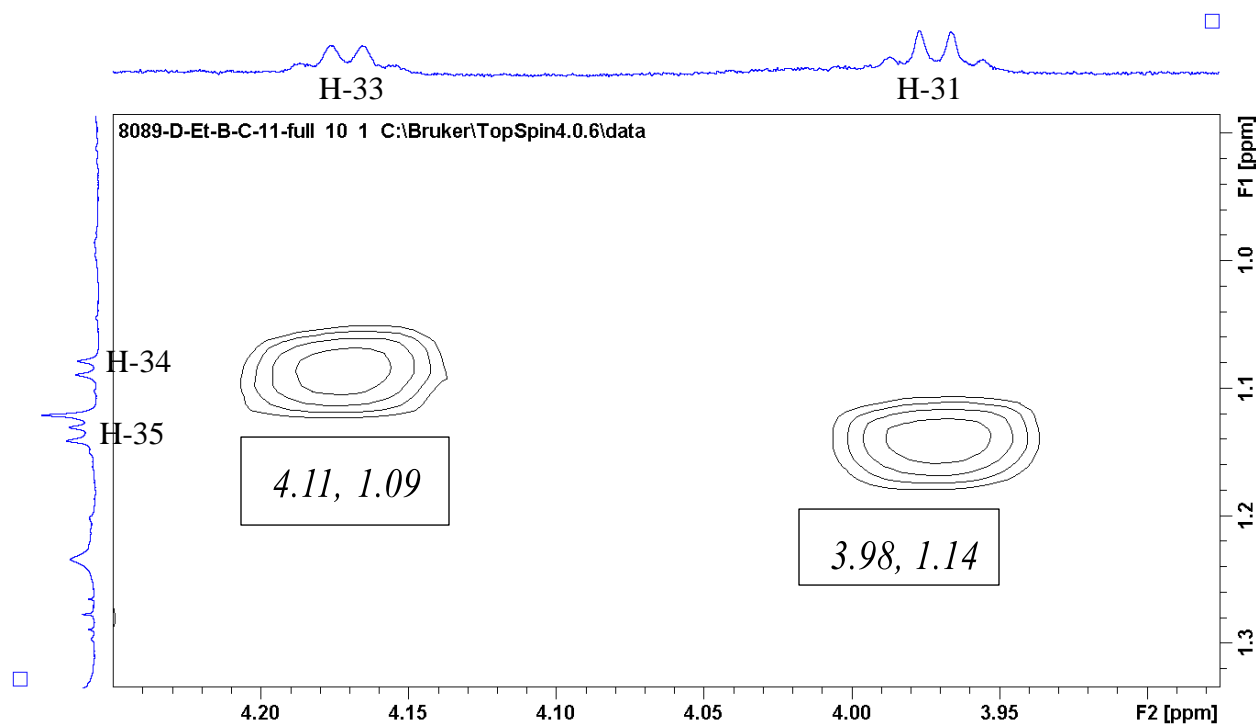
Figure 4-76. Expansion of ¹H-¹H COSY60 spectrum for incarnatapeptin A demonstrating the similar oxane correlations. Recorded in DMSO-*d₆* at 600 Mhz.

In summary, the data set for **4.3** demonstrated two carbinol-like carbons C-31 (δ_C 72.58) and C-33 (δ_C 66.53) each with a methine quartet H-31(δ_H 3.98) and H-33 (δ_H 4.11) (**Figure 4-66 in turquoise**). The methine quartets resonating at δ_H 3.98 (H-31) and δ_H 4.11 (H-33) each demonstrate ^{1}H-^{1}H COSY correlations into methyl doublets H-35 (δ_H 1.14) and H-34 (δ_H 1.09), respectively (**Figure 4-76**). Additionally, the H-34 and H-35 methyl doublets demonstrated HMBC correlations to the methine carbon at C-30 (δ_C 34.03) (**Figure 4-74, blue arrow**), setting up the tentative spin system seen in **Figure 4-77**. As will be discussed, the bond shown in red had no immediate supporting data. Additionally, two secondary carbons at δ_C 25.2 (C-28) and δ_C 18.4 (C-29) were observed with methylene protons (H-28a/b, 1.84/1.75; H-29a/b, 1.88/.54) demonstrating HMBC correlations into C-30 (δ_C 34.0) and C-31 (δ_C 72.6). In addition,
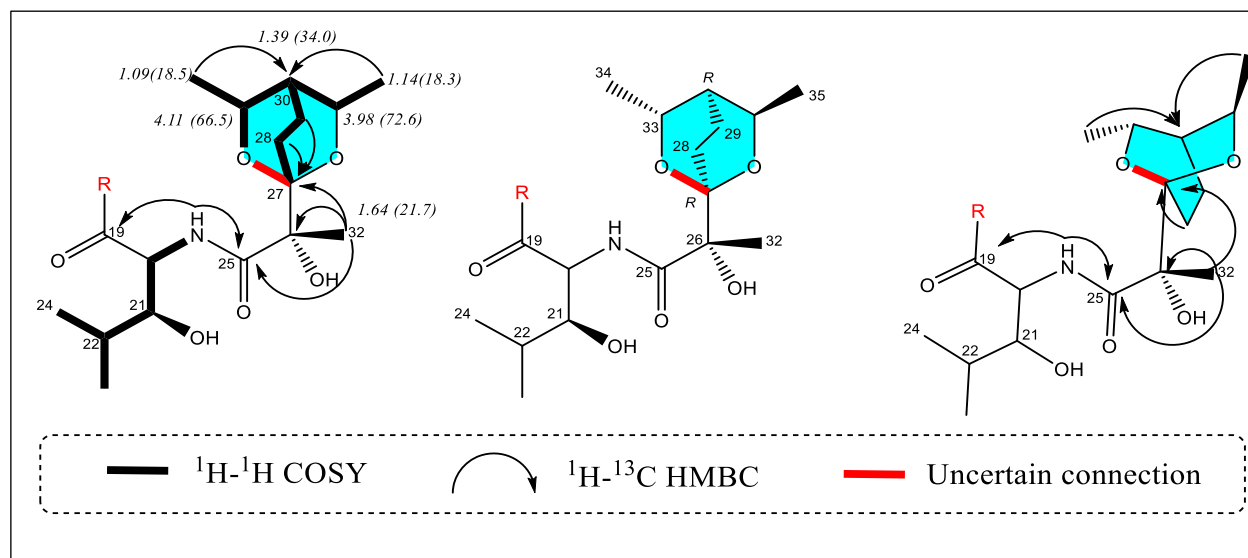
Figure 4-77. Left-hand panel: COSY and HMBC correlations confirming the hydroxy-leucine amino acid and the PKS portion of the molecule. Center panel: configuration in bicyclic ring. Right-hand panel: 3D representation of the bicyclic ring.

the methylene protons of H-28 (δ_H 1.84, 1.75) showed strong HMBC correlations into the quaternary carbon at δ_C 98.5 (C-27) (**Figure 4.77**). With the available NMR data, two possible oxane ring variants were proposed (**Figure 4-78**). Strong evidence supported an oxygenated

187

ethyl side chain at C-33 and C-34. Consideration of the calculated mass of 782.42 Daltons led to the further conclusion that this hydroxylated ethyl side chain would, in fact be involved in forming a bicyclic system. However, with the available NMR data, it was difficult to determine which of two possible variants, 1) 2,6-dioxabicyclo[2.2.2]octane moiety or 2) a single ring with a hydroxylated ethyl side chain was the correct structure.
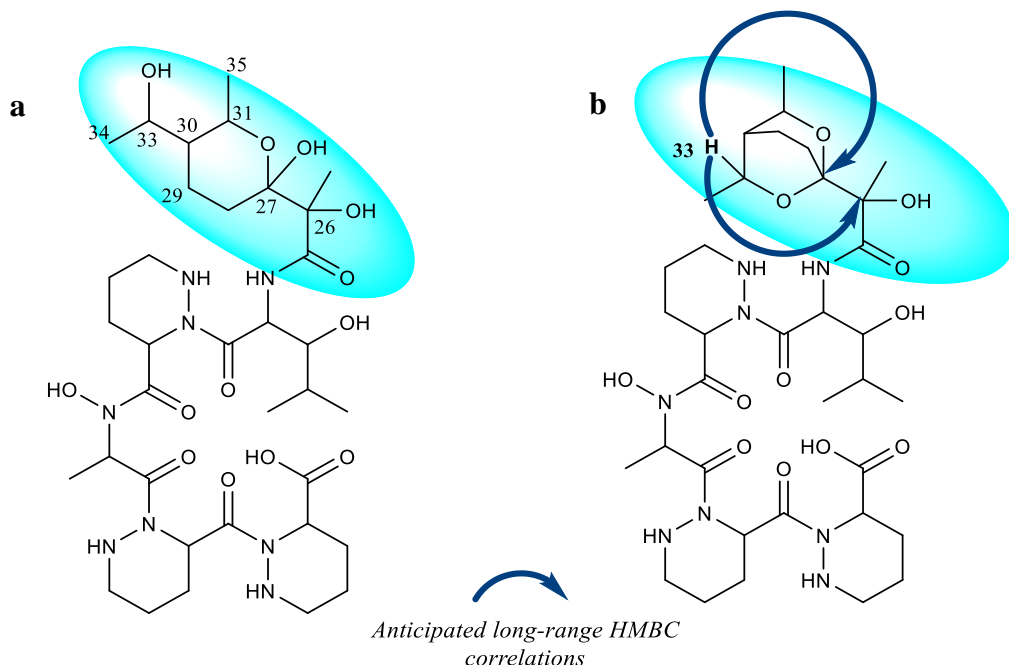
Figure 4-78. Initial NMR data could not distinguish between a) A possible monocyclic variant for 4.3's PKS portion given initial data or b) A possible bicyclic variant for 4.3's PKS portion given the initial data. LR-HSQMBC NMR experiment was conducted in order to see ^4J and ^5J coupling.

An HMBC correlation between H-33 to C-27 or C-26 would have demonstrated connection between the C-33 oxygen and the C-27 quaternary center, thus confirming a bicyclic system. However, this correlation was not present. On the alternate structure, the absence of the correlation did not confirm the single ring. Therefore, in an attempt to see this correlation, a LR-HSQMBC NMR experiment,[216] which can detect atoms with four to five bonds separating them was conducted in the hopes of seeing a correlation either between H-33 to C-27 or between H-33

188

to C-26. No signal between the methine proton at H-33 and the quaternary carbon at C-27 or C-26 was seen.

Residual trifluoroacetic acid (TFA) in the NMR sample prevented consistent observation of the hydroxyl proton resonances. It was thus difficult to ascertain if an artifact of the isolation process was preventing detection of a hemiketal hydroxyl proton resonance or if it was truly absent from the structure. To overcome this second limitation, David Williams carried out a methyl esterification of the natural product to generate a methyl ester for **4.1 (Appendix D)** and purified it by HPLC without TFA. NMR data obtained for the methyl ester of incarnatapeptin A had visible hydroxyl protons (**Appendix D**) with: ^{1}H-^{1}HCOSY correlations between H-21 (δ_H 3.36) of the 3-β-hydroxy leucine and an exchangeable doublet without nitrogen attachment at δ_H 4.73; the appearance of an exchangeable singlet at δ_H 4.91 (O-10H-10) at as well as additions of methoxy resonances for the free ester (δ_H 3.67, δ_C 51.97 with HMBC correlations into δ_C 170.9) and the transformed N-OH (δ_H 3.74, δ_C 63.34 with g^{15}NLRHMQC correlations into N-5 (δ_N -173.0)). Hydroxyl protons for the typical hemiketal at C-27 were not visible, nor was there a hydroxyl proton with correlations into C-33. However, a definitive assignment was still not obtainable from the NMR data, and there were multiple stereocenters to assign. Fortunately, the methyl ester (**Appendix D**) appeared to be crystallizable, and the crystals subsequently obtained were suitable for single-crystal X-ray diffraction analysis (**Appendix E**). The structure determined by X-ray diffraction (**ORTEP Diagram in Figure 4-79**) was consistent with the NMR data for **4.3,** which confirmed that the PKS portion of the natural product contained an unusual 2,6-dioxabicyclo[2.2.2]octane moiety along with confirming the absolute configuration of **4.3**.
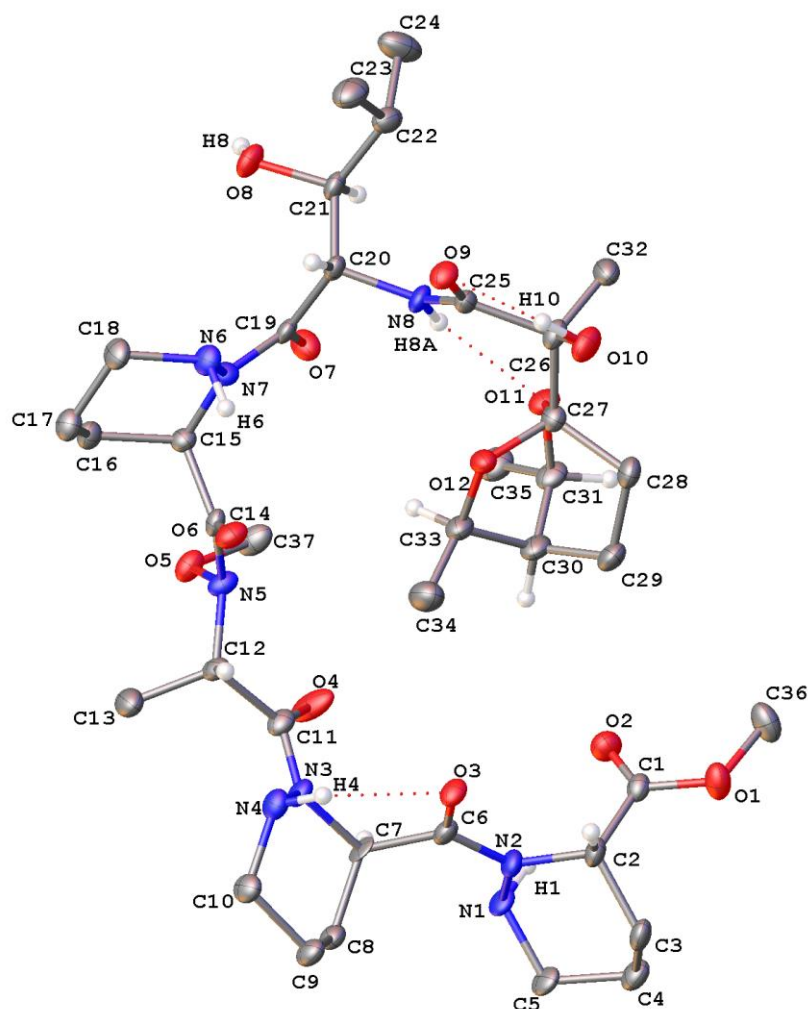
Figure 4-79. X-ray crystal figure produced for the methyl ester of incarnatapeptin A.

. To the best of our knowledge, this 2,6-dioxabicyclo[2.2.2]octane moiety has not been reported in a natural product or from synthesis. Interestingly, the X-ray structure confirmed that the chiral carbons of C-31 and C-33 in the bicyclic moiety were both *R*-configured (**Figure 4.77**), resulting from the relative trans configuration for the attached methyl branches of C-35 and C-34, respectively. With both C-31 and C-33 having an *R*-configuration and thus the same Cahn-Ingold-Prelog hierarchy, challenges arose in assigning the absolute configuration of the stereocenters at C-27 and C-30. Application of revised CIP rules, following the Cahn-Ingold-

Prelog rules for central chirality, along with analysis of the Xray data through ChemDraw™ helped to assign configuration at C-27 (*R*) and C-30 (*R*).[241,242]

Residue	Position		δ_C/δ_N	δ_H (*J* in Hz)
8089-11: Incarnatapeptin A (4.3)				
			DMSO-*d₆*	
L-Piz 1	1	C	172.0	
	2	CH	50.5	4.90, bdd (5.5, 1.8)
	3 a/b	CH₂	24.7	1.78/2.15 m
	4 a/b	CH₂	20.8	1.29/1.52 m
	5 a/b	CH₂	46.4	2.60/3.0 m
	NH-1	NH	-301.1	4.71, d (12.3)
	COO<u>H</u>	OH		13.26
D-Piz 2	6	C	172.5	
	7	CH	48.4	5.54 d (5.8)
	8 a/b	CH₂	25.5	1.99/1.78 m
	9 a/b	CH₂	20.8	1.48/1.51 m
	10 a/b	CH₂	46.5	2.6/2.9 m
	NH-4	NH	-300.5	5.12 d (12.6)
	N3	C		
N-OH-L-Ala	C11	CH	172.0	
	12	CH3	53.4	5.59 q (7.1)
	13	C	14.5	1.35 d (7.1)
	N5		-200.6	
	N5-O<u>H</u>			9.55 s
D-Piz 3	14	C	171.1	
	15	CH	48.4	5.47 dd (6.3, 1.4)
	16 a/b	CH₂	24.5	2.12/1.75 m
	17 a/b	CH₂	21.1	1.48/1.51 m
	18 a/b	CH₂	45.6	2.8 m
	NH-6	NH	-299.9	5.07 d(11.16)
β-OH-L-Leu	19	C	172.6	
	20	CH	49.6	5.50 m
	21	CH	75.8	3.36 dd (8.6, 2.1)
	22	CH	28.1	1.75 m

	23	CH$_3$	15.1	0.82 d (6.8)
	24	CH$_3$	20.2	0.86 d (6.8)
	NH-8	NH	-264.9	7.29 d (9.7)
	C21-OH			Not visible
PKS	25	C	172.0	
	26	C	76.4	
	27	C	98.5	
	28 a/b	CH$_2$	25.2	1.84/1.75 m
	29 a/b	CH$_2$	18.4	1.88/1.54 m
	30	CH	34.0	1.39 m
	31	CH	72.6	3.98 q (6.5)
	32	CH$_3$	21.2	1.12 s
	33	CH	66.5	4.11 q (6.4)
	34	CH$_3$	18.5	1.09 d (6.4)
	35	CH$_3$	18.3	1.14 d (6.5)
	26-OH			Not visible

Table 4-3. NMR assignments for incarnatapeptin A (4.3) in DMSO-d_6, recorded at 600 MHz.

4.3.5 Incarnatapeptin B

The cyclic relative of incarnatapeptin A (**4.3**), the structure elucidation of incarnatapeptin B

(**4.4**) was completed by David Williams after he isolated the molecule from *Streptomyces*

Incarnatus NRRL 8089 extracts as a very minor metabolite[243] during his work to purify more of

4.3 and help establish the bicyclic nature of the PKS-derived moiety. The structure of **4.4** is

given here as it is important to the discussion of the putative biosynthetic pathway and

bioactivity. The data used to elucidate the molecule can be found in the 2020 Organic Letters

paper where it was published.[243]

Incarnatapeptin B **(4.4)**
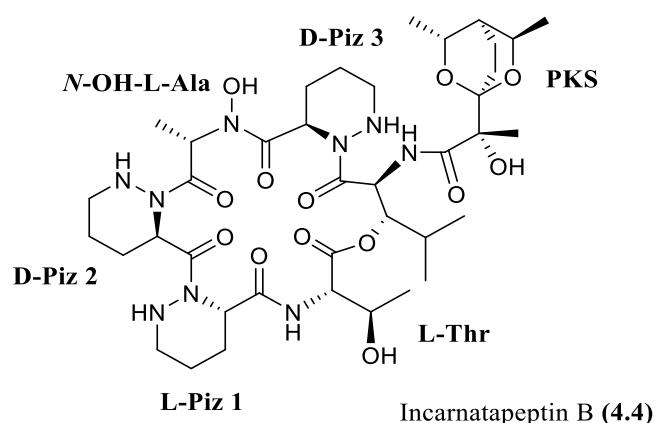
4.4   Biosynthetic Proposal for the Incarnatapeptins and Dentigerumycins F and G

As discussed in **Section 3.3.4**, analysis of the genome of *Streptomyces incarnatus* NRRL-8089 (using antiSMASH) allowed us to identify an ~90 kb gene cluster contained within the CP011498.1 contig as the putative incarnatapeptin biosynthetic gene cluster (**Scheme 3-1**, **Figure 4.80**). The gene cluster was identified as the sole gene cluster encoding a KtzT homolog within the full genome. The arrangement of genes in this gene cluster are highly related to the gene clusters for the piperazic acid-containing natural products polyoxypeptins, the aurantimycins (**Figure 2-5**) and verucopeptin.[173,180] Specifically, as seen in **Figure 3-22**, *Streptomyces incarnatus* NRRL-8089's CP011498.1 gene cluster contains the same number of PKS modules as both the polyoxypeptins and the aurantimycins, with similar to identical domains (**Table 3-9**). The CP011498.1 gene cluster also contains a similar number of NRPS modules with similar to identical domains (**Table 3-9**) to the gene clusters of the
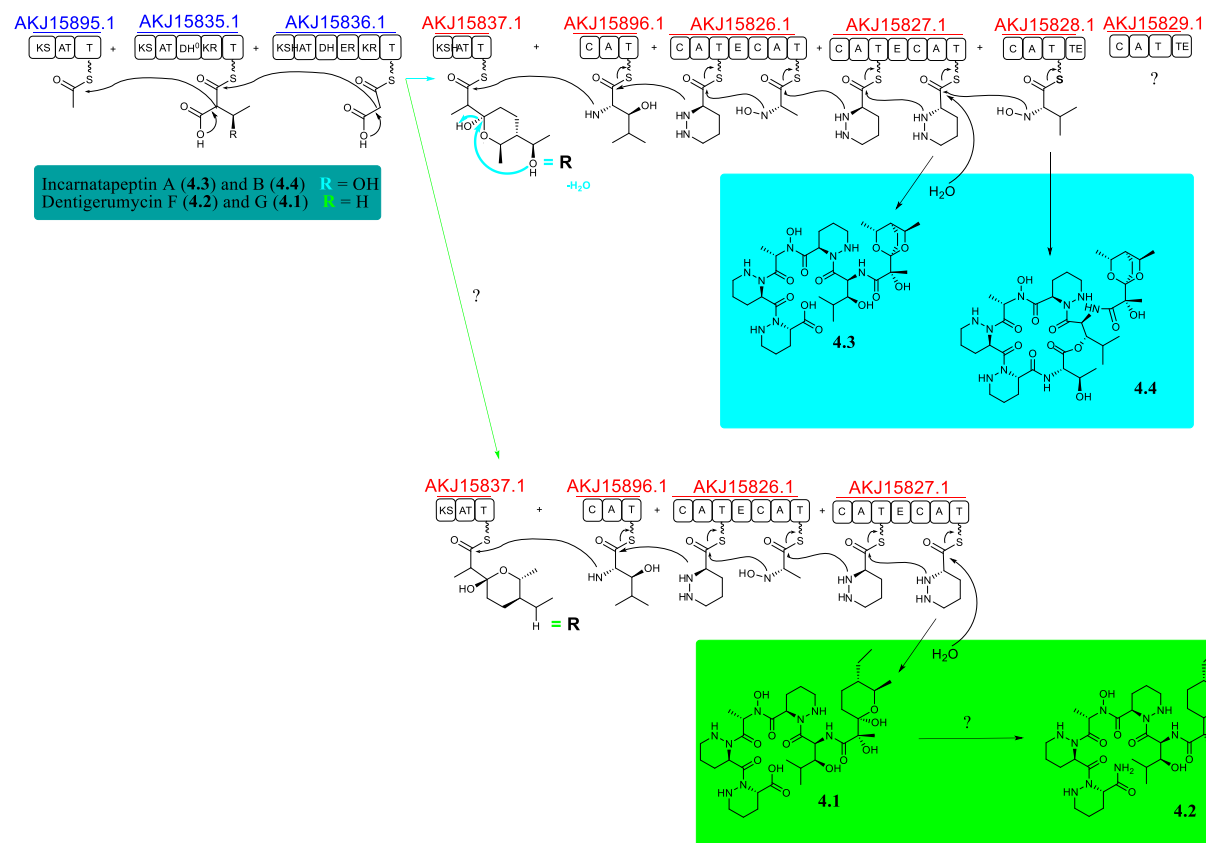
Figure 4-80. Biosynthetic proposal for the incarnatapeptins and dentigerumycins F and G.

characterized molecules. The similarity in the gene clusters is expected due to the overall

similarities in these molecules' structures, where a similar PKS-derived oxane moiety is linked to

amino acids through a highly conserved β-OH-Leu. One unique feature of **4.3** and **4.4** is the

polyketidic portion of these Piz-containing compounds. The transformation which leads to the

unusual bicyclic system of incarnatapeptins A (**4.3**) and B (**4.4**) does not have a literature

precedent. Additionally, the gene cluster is producing both the bicyclic moiety of **4.3** and **4.4**

along with the vicinal methyl/ethyl substituents of dentigerumycins F (**4.2**) and G (**4.1**). **4.3** and

4.4 have a modified C-33/C-34 compared to **4.1** and **4.2** (**Figure 4.81a**), this difference could

arise from either pre- or post-assembly modifications. As seen in **Figure 4.81b**, possible

biosynthetic origins of this bicyclic

194
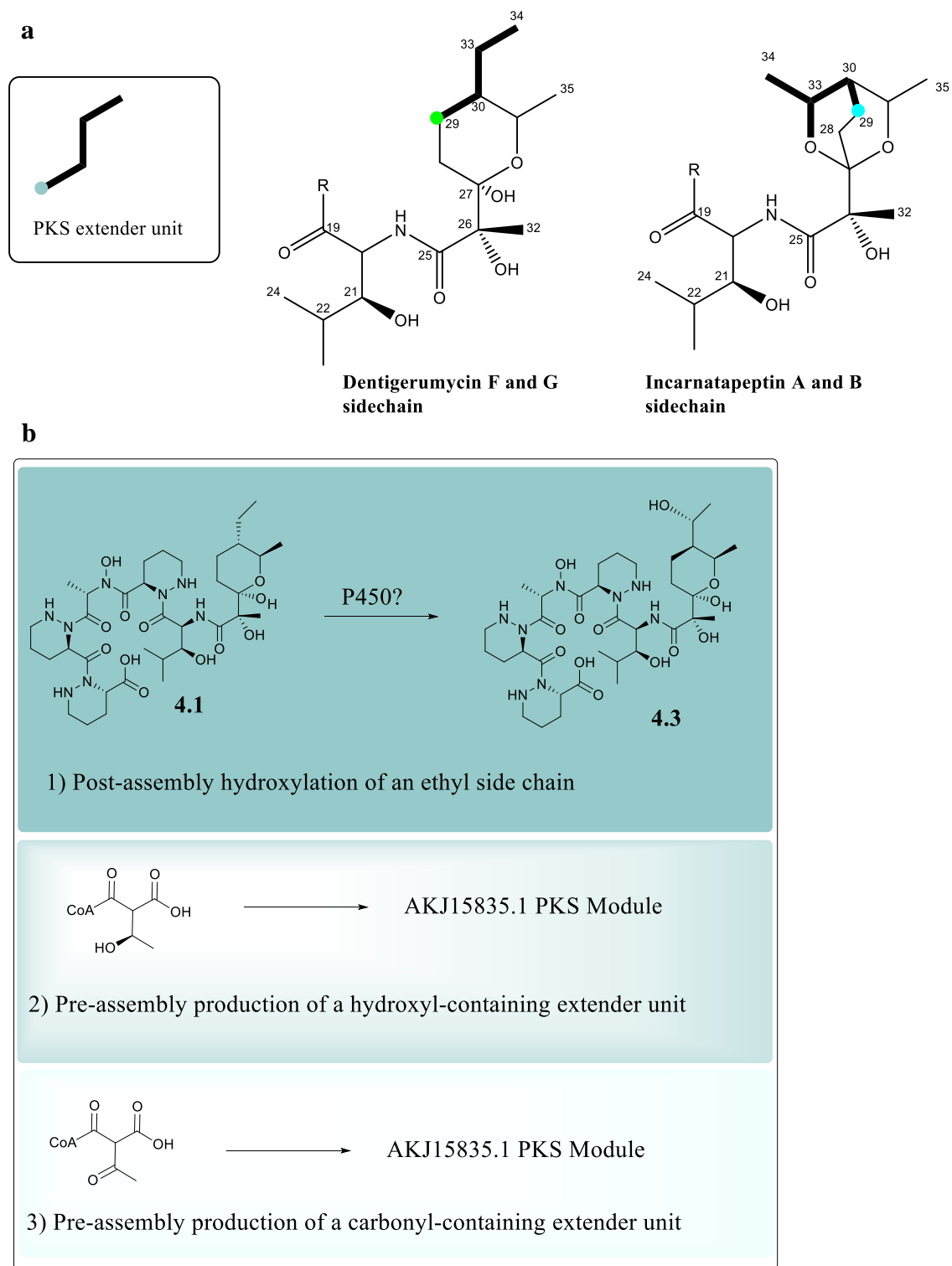
Figure 4-81. Visualizing putative biosynthesis pathways a) Visualizing potential PKS extender units within the structure of the incarnatapeptins and dentigerumycin F and G. b) Possible origins of the oxygenated side chain of the incarnatapeptins.

system could be either **1**) a post-assembly hydroxylation of an ethyl sidechain derived from an ethylmalonyl-CoA extender unit, **2**) pre-production of a hydroxyl-containing polyketide extender unit, or **3**) pre-production of a carbonyl-containing extender unit that was reduced in post-assembly. In all three routes, the bicyclic moiety of the incarnatapeptins would be formed through, likely, spontaneous dehydration. This dehydration could involve the formation of the second six-membered ring through nucleophilic attack of the oxygen of the side chain hydroxyl onto the electrophilic carbon of an oxonium ion intermediate (**Figure 4-80,** oxonium ion intermediate not shown due to space limitations). If the two different extender units are constructed in a pre-assembly manner, **4.3** and **4.4**'s C-29, C-30, C-33, and C-34 chain may come from a 3-hydroxy butyryl-CoA extender unit; whereas **4.1** and **4.2**'s C-29, C-30, C-33, and C-34 could likely come from an ethylmalonyl-CoA extender unit (**Figure 4.81a** and **Figure 4.82a**). It is difficult to say with certainty the exact transformations and genes responsible without further labelling or knock-out studies.

However, there is a suite of genes between AKJ15838.1 and AKJ15843.1 in the CP011498.1 contig from *Streptomyces incarnatus* NRRL 8089 (**Figure 4.82b**) that are associated with the construction of unusual PKS branched-chain extender units.[118,182,244] It is common for ethylmalonyl-CoA to be constructed from crotonyl-CoA through a crotonyl-CoA carboxylase/reductase (CCR) which is present at AKJ15838.1.[118] This alone would not explain the biosynthesis of both extender units. Given that the machinery is present to construct ethylmalonyl-CoA, the hydroxylation of C-33 may occur with the oxidation of the alkyl chain through a cytochrome P450 in post-assembly (**Figure 4.81b**). However, the only available cytochrome P450 is quite a distance upstream on the other side of the transpose used to hypothesize an approximate putative cluster boundary (**Table 3-10**). Given the rest of the genetic
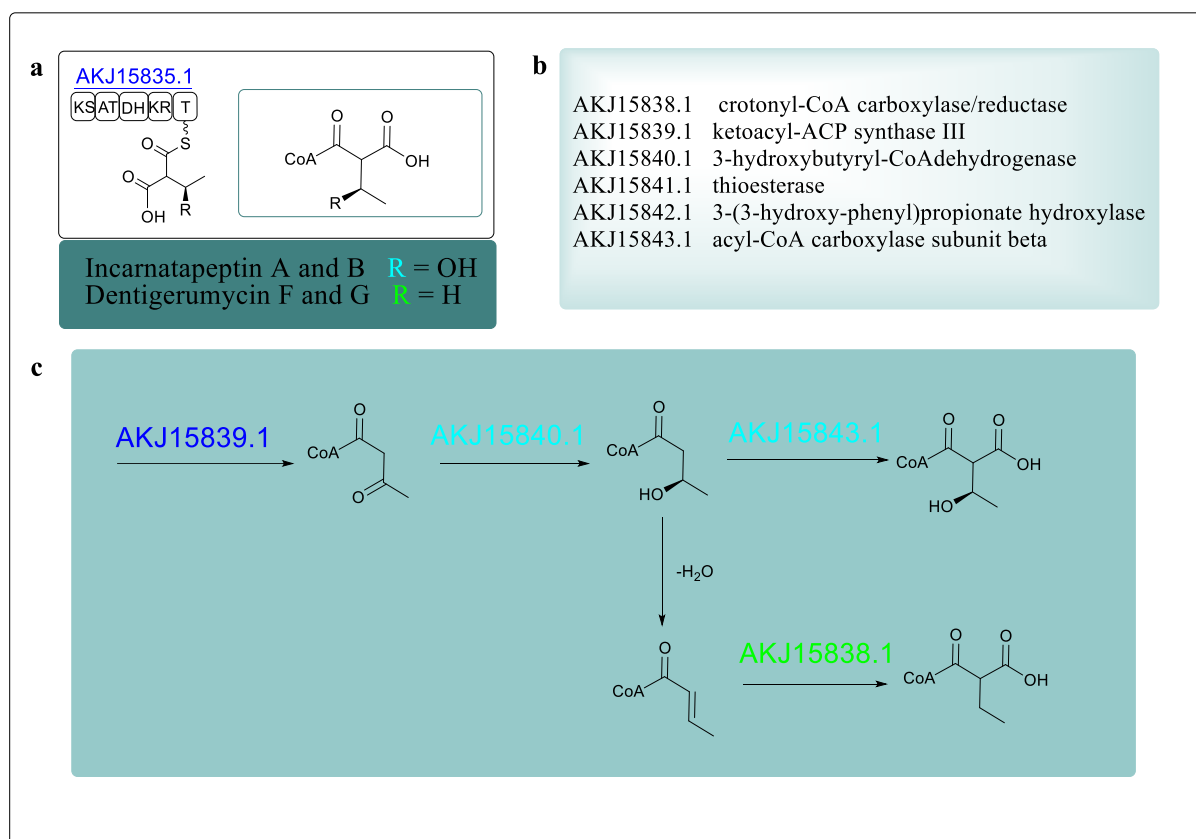
Figure 4-82. Proposed biosynthetic pathway for the PKS extender units leading to the incarnatapeptins and dentigerumycin F and G. a) Two different extender units maybe loaded for the incarnatapeptins and dentigerumycins F and G. b) The CCR suite found in the CP011498.1 contig of *Streptomyces incarnatus* NRRL 8089 c) Possible biosynthetic pathway leading to the two unique extender units.

machinery present between AKJ15838.1 and AKJ15843.1, I believe a plausible pathway is seen

in **Figure 4.82C** with an acetoacetyl-CoA constructed through AKJ15839.1, a ketoacyl-

ACP synthase III, whereby the carbonyl is transformed into an alcohol through the 3-

hydroxybutyryl-CoA dehydrogenase found at AKJ15840.1. Once the 3-hydroxybutyryl-CoA is

biosynthesized, dehydration would lead to the crotonyl precursor. Then, through the action of the

crotonyl-CoA carboxylase/reductase found at AKJ15838.1, the ethylmalonyl-CoA extender unit

would be produced (**Figure 4.82C**). These proposed transformations mirror a series of

characterized biosynthetic transformations in the divergolides[245], which adds plausibility to the

pre-assembly of two distinct extender units, as seen in **Figure 4.82**.

Also for consideration are the feeding studies for polyoxypeptin, which demonstrated that polyoxypeptin's unusual extender unit (at the same loading position as the incarnatapeptins' C33/34) was coming directly from an isoleucine (Ile) pathway.[246] If the incarnatapeptins follow this biosynthetic logic, in this case, one would imagine that threonine was the precursor to a hydroxylated extender unit that was loaded onto the AKJ15835.1 PKS module as in **Figure 4.81b-2**. However, there is some possibility that those feeding experiments simply reflect a crotonyl pathway derivative whose branching chain is via a fatty acid source with Ile a known precursor.[118] On the other hand, the published (but possibly truncated) polyoxypeptin gene cluster does not possess a CCR suite and, as such, could have a unique route to this PKS extender unit (as opposed to gene clusters that do show CCR genes). Nevertheless, given the presence of the CCR suite and other genes that reflect transformations seen in the divergolides, in the case of the incarnatapeptins, where I have access to the full putative gene cluster, I think the evidence points away from a threonine-based extender unit.

4.5   Bioactivity

4.5.1 Antimicrobial Assays

The incarnatapeptins A, dentigerumycins F and G were tested for antimicrobial activity in a minimum inhibitory concentration assay against *E.Coli* 8161 and MRSA. Incarnatapeptin A and dentigerumycin G demonstrated no antimicrobial activity. Incarnatapeptin B (**4.4**) was not tested due to limited material. Dentigerumycin F had conflicting results. Initially, due to availability, samples of dentigerumycins F and G that were initially collected when optimizing HPLC conditions were tested in the antimicrobial assays. When these samples were tested, dentigerumycin F demonstrated potent antimicrobial activity against *E.Coli* cultivar 8161 (MIC = 1 ug/mL or 1.27 nMol, **Table 4-4a**) and some activity against MRSA (**Table 4-4b**). However,

when dentigerumycins F collected in optimized HPLC conditions was retested, it showed no antimicrobial activity. This was a puzzling result. However, given the very small distance between the HPLC peaks (less than 1 minute) called 8089.Et.B.C.13 (dentigerumycin F) and 8089.Et.B.C.14 (even in the optimized HPLC conditions), I believe that it is highly likely the initial incarnatapeptin C sample tested contained enough of the compound named 8089.Et.B.C.14 to produce these results, and this is the likely antimicrobial compound. Unfortunately, the tiny amount of the pure fraction of 8089.Et.B.C.14 that was collected was placed into pyridine-d_5 for NMR data collection, where, as previously discussed, it underwent an unsolved degradation that left it unsuitable for bioassay testing. See **Appendix F** for excel sheets with complete assay data.

a	Dentigerumycin F concentration	*E. coli* 8161 % alive	b	Dentigerumycin F concentration	MRSA % alive
	Blank			Blank	
	20 mg/ml	0		64 mg/ml	0
	4 mg/ml	0		32 mg/ml	2
	2 mg/ml	0		16 mg/ml	3
	1 mg/ml	0		8 mg/ml	4
	500 ng/ml	2		4 mg/ml	4
	250 ng/ml	96		2 mg/ml	3
	50 ng/ml	99		1 mg/ml	75
	10 ng/ml	97		0.5 mg/ml	70
	2 ng/ml	95		0.25 mg/ml	80
	0.4 ng/ml	124		0.125 mg/ml	76
	DMSO + cells	100		DMSO + cells	100

Table 4-4. Antibacterial MIC results for dentigerumycin F that were not reproducible upon subsequent testing with a purer aliquot of the compound. a) *E. coli* **8161 results b) MRSA results**

4.5.2   Bioactivity of Incarnatapeptins in the Prostate Cancer Assay

Incarnatapeptins A (**4.3**, discussed in this thesis work) and B (**4.4,** isolated by David Williams) were tested for *in vitro* cytotoxicity to LNCaP and PC3 human prostate cancer cell lines. **4.1** and **4.2** were not tested in this assay due to the interest in the compounds with the

bicyclic ring system. Initially, both **4.3** and **4.4** were tested against both PC3 and LNCaP cells at a series of concentrations. PC3 cell growth was not impacted by incarnatapeptin A (**4.3**) or B (**4.4**) at 30 µM, whereas the growth of LNCaP cells appeared to be impacted by **4.4** at 15 and 30 µM as indicated by reduced cell density at 24 h (**Figure 4-83**). Due to the limited amount of **4.4** that was available, we were only able to quantify the LNCaP response at 30 uM (**Figure 4-83 & 4-84**). Using a marker that stains the DNA of nonviable cells revealed that incarnatapeptin B induced a 17-fold increase in non-viable LNCaP cells, whereas **4.3** had no apparent effect at 72 h (**Figure 4-83 & 4-84**). These limited data results suggest modest potency but selective sensitivity of LNCaP cells to incarnatapeptin B (**4.4**) compared to PC3 cells (**Figure 4-85**). Interestingly, the macrocyclic depsipeptide **4.4** was active at 30 uM, but the linear peptide analog **4.3** was inactive. If additional **4.4** becomes available, further studies are warranted to determine IC$_{50}$ values, underlying mechanisms, and additional SAR.

a   Incarnatapeptin A

LNCaP

PC3

| 0 μM | 7.5 μM | 15 μM | 30 μM |

b   Incarnatapeptin B

LNCaP

PC3

| 0 μM | 7.5 μM | 15 μM | 30 μM |
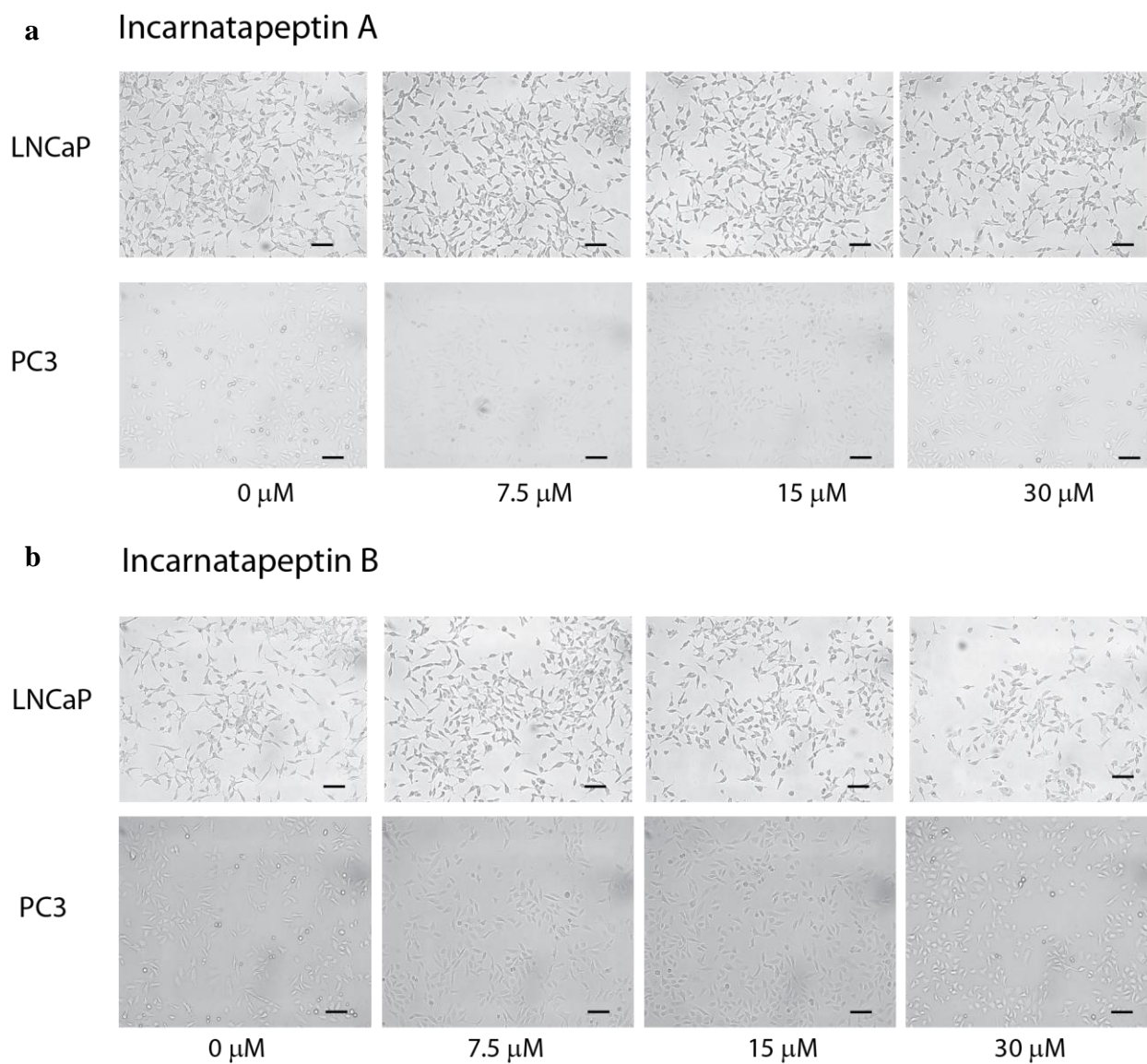
Figure 4-83. LNCaP and PC3 cells were exposed for 24 hours to increasing concentrations of incarnatapeptins A (4.3) and B (4.4). Micrograph scale bars: 100 µm.
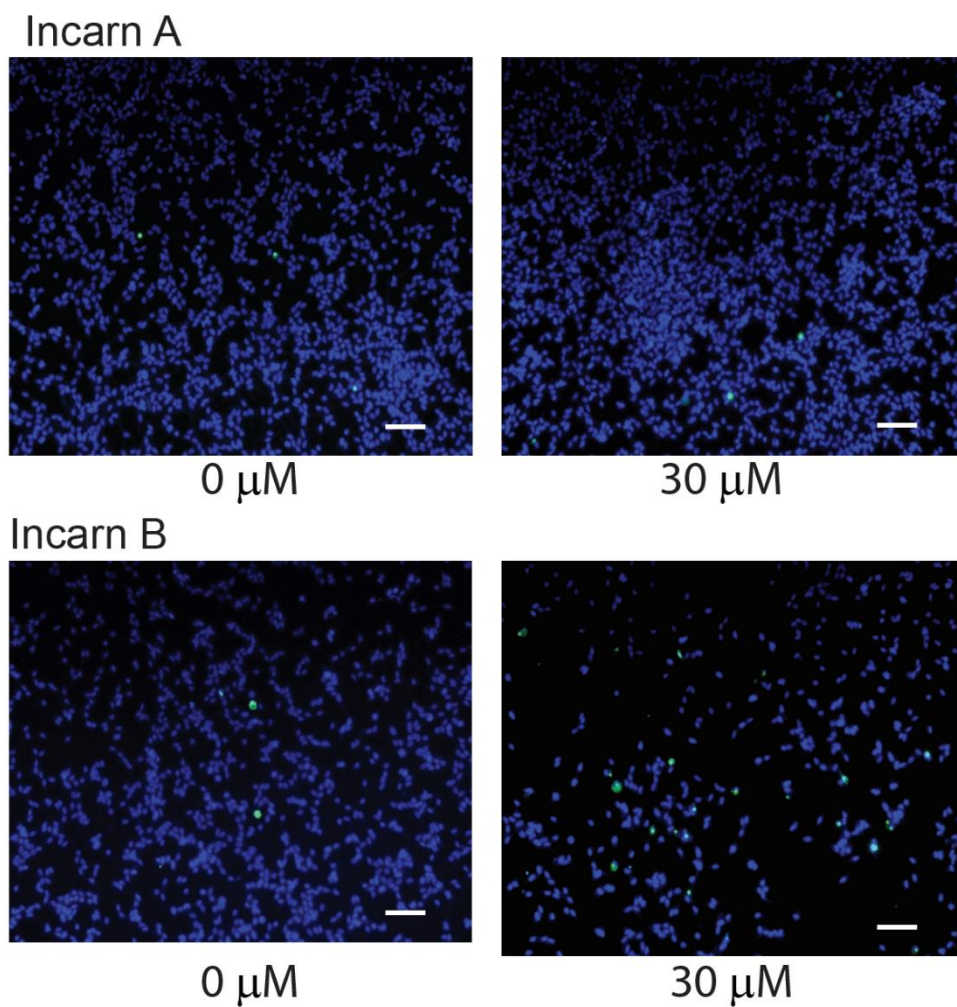
Figure 4-84. Incarnatapeptin A (4.3) induced cell death of LNCaP prostate cancer cellsand LNCaP. Incarnatapeptin B (4.4) induced cell death of LNCaP prostate cancer cells. Cell viability was measured using NucGreen Dead 488 Ready Probe and the number of positive green stained cells was counted.
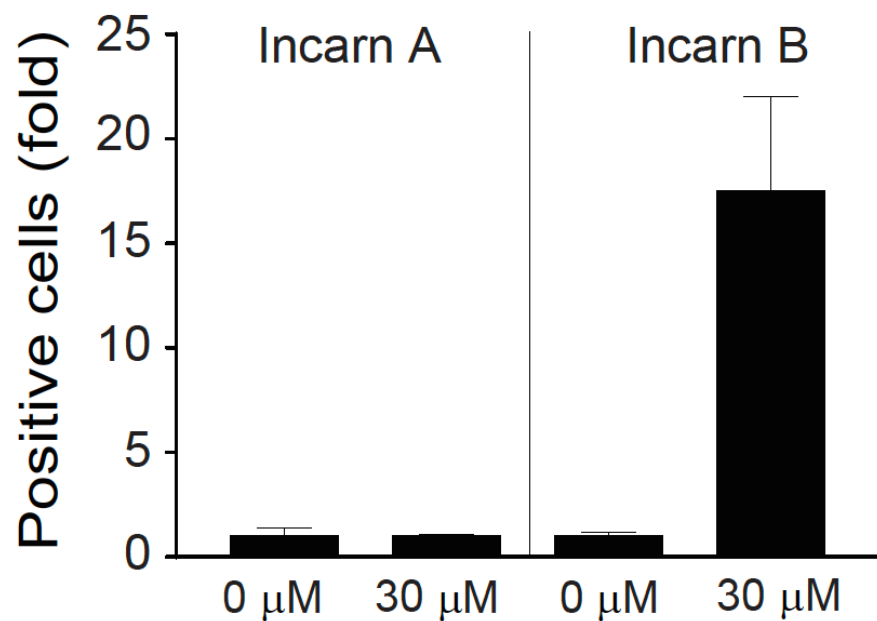
Figure 4-85. Viability of LNCaP cells treated with incarnatapeptins A (4.3) or B (4.4) was measured using NucGreen Dead 488 Ready Probe and the number of positive green stained cells was counted. Fold induction of positive (dead) cells after 72h of exposure is shown.

4.6   Summary and Conclusion
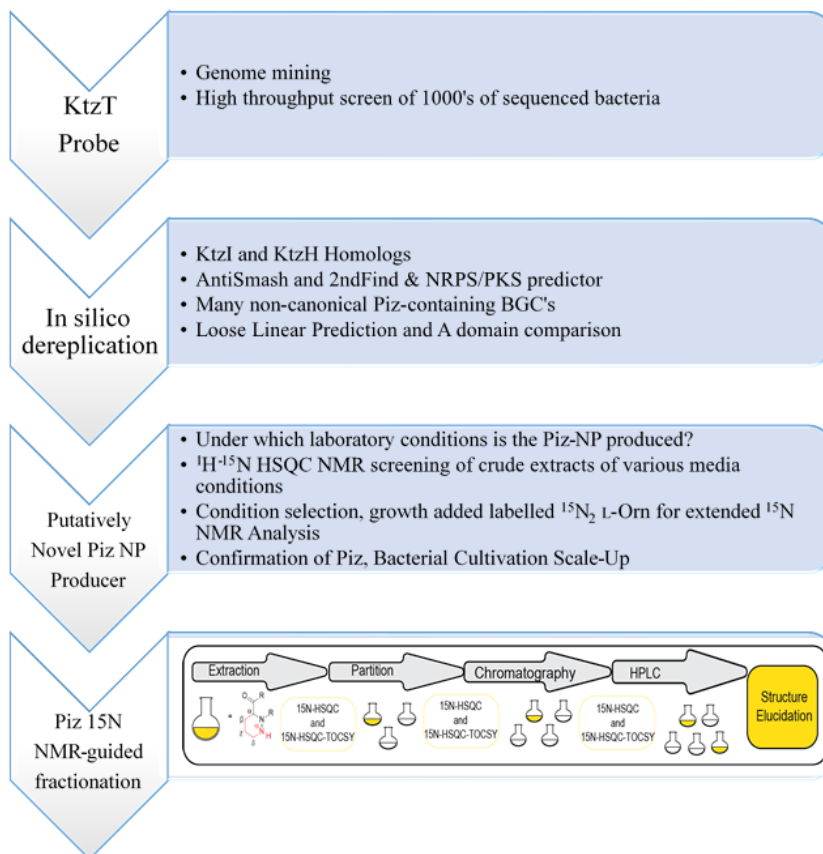
Figure 4-86. Combining genome mining and Piz ¹⁵N NMR-guided fractionation for the isolation of the incarnatapeptins and dentigerumycins F and G.
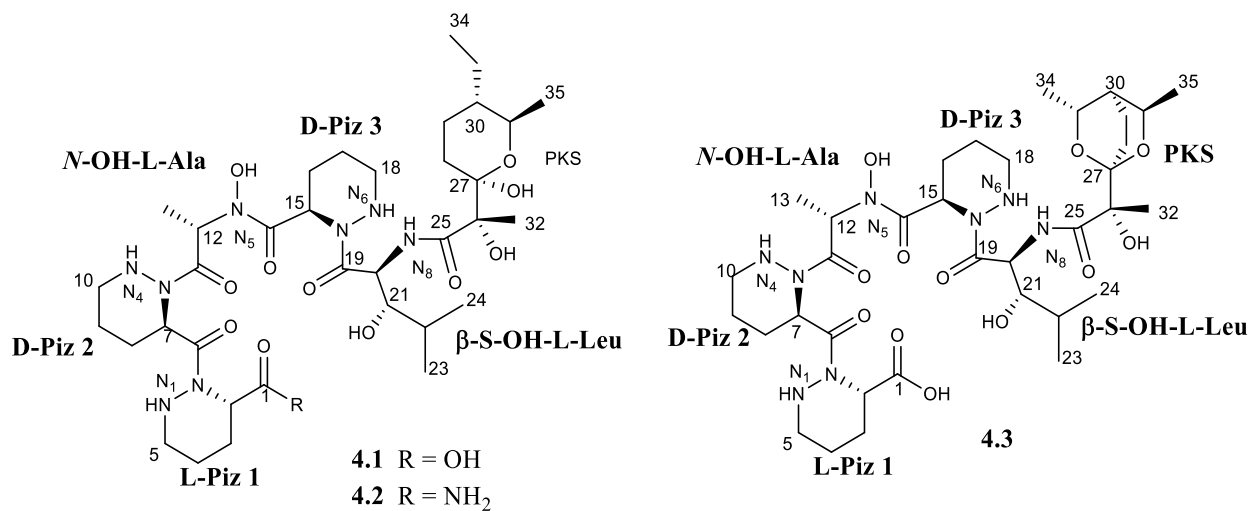
Figure 4-87. Structures of incarnatapeptin A (4.3), and dentigerumycins F (4.2) and G (4.1).

Throughout Chapters 2, 3 and 4, I have described the development of a new natural product discovery method combining genome mining and nitrogen NMR to target a specific functional-group of interest and the structure elucidation of the molecules discovered through this method as represented in **Figure 4-86**. First, genome mining was used to select strains that produced the desired moiety. Next, a specific nitrogen NMR signature was used to assess the laboratory growing conditions for Piz natural product by specific bacteria. Finally, the nitrogen NMR signature was used as a guide throughout the isolation process. This targeted genome mining/nitrogen NMR approach resulted in the discovery of four novel Piz-containing NRPS/PKS hybrid natural products, three of whose structures were elucidated as part of this thesis work (**Figure 4-87**). The incarnatapeptins A (**4.3**) and B (**4.4**) contain a never before seen bicyclic ring system, and incarnatapeptin B proved to be a selective cytotoxin.

Into the future, assay results point to the presence of an unsolved incarnatapeptin produced by *Streptomyces incarnatus* NRRL 8089 that has potent activity against *E. coli*. Combining the work that has already been completed with a bioassay-guided approach could uncover this powerful antibiotic. Dentigerumycin F (**4.2**) and G (**4.1**) were shown to contain an unusual equilibrium at the hemiketal functional group, knowledge of which could prove useful for studies of these normally stable 1-hydroxyoxane rings. Furthermore, there are very few families within these NRPS-PKS hybrids that show a divergence in the substitution pattern of the oxane ring when all the members of the family are derived from the same bacterial source (**Chapter 2**). There is only one other, the oleamycins, which look like they diverge in the likely loaded PKS extender unit. Characterizing the biosynthetic pathway for this crossroads could help both in engineering efforts and the ability to predict chemical phenotypes of biosynthetic gene clusters during genome mining efforts.

205

Broadly, the nitrogen NMR-guided approach could be expanded to target other nitrogen-containing natural products. Specifically, this method could be used to prospect for natural products containing nitrogen-nitrogen bond functional groups whose representation in the solved chemical space is inconsistent with possible compounds bioinformatically predicted from sequenced bacterial genomes. As this nitrogen NMR-guided approach is a targeted method, each moiety needs to be considered for a unique NMR signature. Early analysis determines that many unusual nitrogen-nitrogen bond functional groups contain such signatures (**Appendix G**) and that the solved chemical space could be rationally expanded using this targeted method. Finally, this method can also be used to target other Piz-containing molecules and could be enhanced by a tandem bioassay-guided approach.

4.7   Experimental Section

4.7.1 General Experimental Procedures

Optical rotations were measured on a Jasco P-1010 Polarimeter with sodium light (589 nm). The ^{1}H, ^{13}C and ^{15}N NMR spectra were recorded on a Bruker AV-600 spectrometer with a 5mm CPTCI cryoprobe. ^{1}H resonances are referenced to the residual DMSO-d_6 (δ 2.50), pyridine-d$_5$ (δ 7.22), ; CD$_3$OD (δ 3.31), and ^{13}C resonances are referenced to the DMSO-d_6 solvent peak (δ 39.51); pyridine-d$_5$ (δ 123.9); CD$_3$OD (δ 49.2). The ^{15}N NMR spectra are in reference to CH$_3$NO$_2$ (0 ppm). Low resolution ESIMS were recorded on Bruker HCT Ultra PTM Discovery System and high resolution ESIMS were recorded on a Waters / Micromass LCT time-of-flight (TOF) mass spectrometer equipped with an electrospray ion source. Labelled ^{15}N$_2$-L-Orn was obtained from Cambridge Isotope Laboratories, Inc and unlabelled L-Orn was obtained from either Sigma-Aldrich or Fisher Scientific. All other media components were obtained from Fisher Scientific. All solvents used for HPLC were Sigma-Aldrich HPLC grade

206

additionally filtered through a micron filter. Reversed-phase HPLC purifications were performed on a Waters 1525 Binary HPLC pump attached to a Waters 2998 Photodiode Array Detector and equipped with Empower Software.

Dentigerumycin F (4.2): Isolated as a clear oil; $[\alpha]^{25}_D$ -14.13' (0.47 g/100 mL, MeOH); ^{1}H (both pyridine-d_5 and DMSO-d_6, 600 MHz) and ^{13}C NMR (both pyridine-d_5 and DMSO-d$_{6,}$ 150 MHz); data can be found in **Table 4.2,** HRESIMS m/z: [M+Na]+ Calcd for $C_{35}H_{61}O_{11}N_9Na$, 806.4388; Found 806.4369.

Dentigerumycin G (4.1): Isolated as a clear oil; $[\alpha]^{25}_D$ -8.39' (0.73 g/100 mL MeOH); 1H (both DMSO-d_6 and CD$_3$OD, 600 MHz) and ^{13}C NMR (both DMSO-d_6 and CD$_3$OD, 150 MHz), data can be found in **Table 4.1**; HRESIMS m/z: [M+Na]+ Calcd for $C_{35}H_{60}O_{12}N_8Na$, 807.4228; Found 807.4237.

Incarnatapeptin A (4.3): Isolated as a clear oil; $[\alpha]^{25}_D$ +27.4° (c 2.8, MeOH); ^{1}H (DMSO-d_6, 600 MHz) and ^{13}C NMR (DMSO-d_6, 150 MHz), data can be found in **Table 4.3**; HRMS (TOFESI) m/z: [M-H]- Calcd for $C_{35}H_{57}N_8O_{12}$, 781.4096; Found 781.4097. HRMS (TOFESI) m/z: [M+Na]+ Calcd for $C_{35}H_{58}N_8O_{12}Na$, 805.4072; Found 805.4075.

4.7.2 Bioassay Experimental Methods

4.7.2.1   MIC Experimental Methods

E. coli 8161 and MRSA strains were inoculated into liquid growth media in falcon tubes. A dilution series of incarnatapeptin C and D dissolved in DMSO was placed into separate falcon tubes containing the human pathogen, along with a positive antibiotic control (dilution series of rifamycin) and a 100% control of DMSO. All falcon tubes were incubated at 37°C for 24 hours and then the growth of the pathogen was determined using optical density readings. The optical

density reading of the DMSO negative control was utilized as a baseline for which to measure the inhibition of growth of the tubes containing the positive control and the natural products. As such, the percentage of pathogens which were still alive was determined by OD readings over DMSO as 100% control.

4.7.2.2   Prostate Cancer Assay Experimental Methods

LNCaP and PC3 prostate cancer cell lines were plated in 24-well plates and incubated with 0, 7.5, 15.0 or 30.0 µM incarnatapeptin A (**4.3**) (Incarn A) or incarnatapeptin B (**4.14**) (Incarn B) for 24 h. Cell viability was determined by incubation for 30 min with the NucGreen Dead 488 Ready Probe (Thermo Fisher Scientific, Waltham, MA, USA). Between 500 to 2000 cells per field were counted using the AxioVision AC fluorescence inverted microscopy (Zeiss, Dubling CA, USA).

Chapter 5: Bacterial Library Construction for Bioassay-guided Discovery

5.1   Chapter Overview

This chapter will discuss the curation of a mini-library towards the discovery of bioactive natural products. The results of testing the mini-library across multiple biological activity assays will be shown. From those results, it will be seen that bioassay-guided fractionation was utilized as a tool to uncover molecules responsible for the bioactivity in multiple bacterial strains.

5.2   Strain Selection of Marine Bacteria

The goal of this research effort was to uncover novel bioactive natural products with a focus on bacterial metabolites. One of the challenges of natural product discovery, particularly in bacteria, is that, given no other indication of what an individual bacterial strain is producing, one must look through hundreds and even thousands of strains to find new or interesting molecules.[247] High-throughput sequence techniques are routinely employed to simultaneously look for a feature of interest from banks of hundreds or thousands of bacterial extracts and already isolated molecules.[247] I focused on producing a smaller library populated with bacterial strains predicted to have a higher than average tendency to produce novel molecules, based on the origin of the bacteria from diverse ecosystems,[11,248,249] or prediction that the strain should be "biosynthetically talented".[250] To support the search for novel bacterial natural products of putative biosynthetic interest, a mini library was constructed from bacterial strains curated from the larger Andersen marine bacteria library along with several terrestrial strains isolated by a member of Julian Davies's lab. The strains were generally selected to meet at least one of several criteria: (1) initial, promising anti-microbial screening results, (2) strains of genera such as *Salinispora* sp. which had previously been shown to produce diverse secondary metabolites,[248] (3) isolation from an invertebrate or other organism,[251] or (4) under-representation of that species

in natural product isolation attempts.[233] For a comprehensive list of the curated strains, see **Table 5.1.**

5.3   Bioassays for Guiding Fractionation

Each of the strains listed in **Table 5-1** was grown in small quantities, the crude extracts prepared into a 96 well plate (2 mg/well) and then sent for biological assays. Positive results were received in several bioassays. Bioassay-guided fractionation methodology was then applied to pursue isolation of pure bioactive compounds. The application of these methods led to the isolation of pure compounds in three of the seven assays: phosphatase inhibition, IL-6 anti-inflammatory and anti-microbial assays.

When screening crude extracts in biological assays, the identity and abundance of the active compounds are unknown. Initial screening is seeking benchmark activity values for each assay. As the separation process should increase the proportion of the active component in a fraction, assay results should show sustained or increasing activity values after each fractionation step. Bacterial strains can demonstrate variations in metabolite production, yet structure elucidation is dependent on sufficient quantities of a compound. These quantities are obtained by culturing multiple batches of a strain. Therefore, to be of continued interest, extracts from subsequent batches of a strain should be able to sustain levels of activity like that seen in the initial screening.

	NAME	Species	Isolated From:	Source*	Accession number		NAME	Species	Isolated from:	Source*
1	DCA3500	*Streptomyces sp.*	Cactus soil	T.S		38	RJA3119	*Salinispora sp.*	Dominican # 2	M.S.
2	DCA3201	*Streptomyces*	Cactus soil	T.S	EU594471	39	RJA3990	*Streptomyces cellulosae*	Georgia Straight #1	M.S.
3	DCA3503	*Streptomyces sp.*	Cactus soil	T.S		40	RJA4055	Streptomyces drozdowiczii.	Georgia Straight #3	M.S.
4	DCA3504	*Streptomyces sp.*	Whistler forest	T.S	GQ924538	41	RJA4089	unknown	*Hamigera*	M.I.
5	DCA3505	*Streptomyces sp.*	Mushroom soil	T.S	GQ924538	42	RJA4095	unknown	*Hamigera.* Ansell	M.I.
6	DCA3506	*Streptomyces sp.*	*Peltigera*	Lichen	GQ924538	43	RJA4096	unknown	*Hamigera.* Ansell	M.I.
7	DCA3507	*Streptomyces sp.*	*Lobaria*	Lichen	DQ221755	44	RJA4104	*Alpha proteobacteria B14*	*Hamigera.* Ansell	M.I.
8	DCA3508	*Streptomyces sp.*	*Flavocetraria*	Lichen	AB222068	45	RJA4108	unknown	*Hamigera.* Ansell	M.I.
9	DCA3509	*Streptomyces sp.*	*Flavocetraria*	Lichen	DQ231567	46	RJA4125	unknown	*Syringella*	M.I.
10	RJA471	*Kocuria Sp.*	Indian Arm	M.S.		47	RJA4127	unknown	*Syringella*	M.I.
11	RJA479	*Kocuria Sp.*	Indian Arm	M.S.		48	RJA4128	unknown	*Syringella*	M.I.
12	RJA494w	*Kocuria Sp.*	Indian Arm	M.S.		49	RJA4134	unknown	*Syringella*	M.I.
13	RJA494y	*Kocuria Sp.*	Indian Arm	M.S.		50	RJA4145	unknown	*Syringella*	M.I.
14	RJA496	*Kocuria Sp.*	Indian Arm	M.S.		51	RJA4150	*Algoriphagus winogradsky*	*Hamigera.* Ansell	M.I.
15	RJA515	*Kocuria Sp.*	Indian Arm	M.S.		52	RJA4193	unknown		M.I.
16	RJA526	*Kocuria Sp.*	Indian Arm	M.S.		53	RJA4198	*Arthrobacter sp.* KOPRa25422	*Boltenia villosa, B.C*	M.I.
17	RJA529	*Kocuria Sp.*	Indian Arm	M.S.		54	RJA4209	unknown		M.I.
18	RJA538	*Kocuria Sp.*	Indian Arm	M.S.		55	RJA4220	unknown	*Styela Clava, B.C.*	M.I.
19	RJA544	*Kocuria Sp.*	Indian Arm	M.S.		56	RJA4231	unknown	*Styela Clava, B.C.*	M.I.
20	RJA2910	*Streptomyces*	Bamfield # 5	M.S.		57	RJA4235	unknown	*Styela Clava, B.C.*	M.I.
21	RJA2998	*Salinispora sp.*	Dominicana #1	M.S.		58	RJA4241	unknown		M.I.
22	RJA3002	*Salinispora sp.*	Papau New	M.S.	JF922126	59	RJA4249	*Streptomyces sanglieri*	*Styela Clava, B.C.*	M.I.
23	RJA3003	*Salinispora sp.*	Papau New	M.S.		60	RJA4254	unknown	*Cliona Californiana,*	M.I.
24	RJA3004	*Salinispora sp.*	Papau New	M.S.		61	RJA4265	unknown	*Cliona Californiana,*	M.I.
25	RJA3005	*Salinispora sp.*	Papau New	M.S.		62	RJA4270	unknown	*Cliona Californiana,*	M.I.
26	RJA3006	*Salinispora sp.*	Papau New	M.S.	JF922127	63	RJA4274	unknown	*Ophlitaspongia, B.C*	M.I.
27	RJA3107	*Salinispora sp.*	Papau New	M.S.	JF837441	64	RJA4276	unknown	*Cliona Californiana,*	M.I.
28	RJA3109	*Salinispora sp.*	Papau New	M.S.	JF837442	65	RJA4296	unknown	*Phorbas* Sp. B.C.	M.I.
29	RJA3110	*Salinispora sp.*	Papau New	M.S.		66	RJA4300	unknown	*Phorbas* Sp. B.C.	M.I.
30	RJA3111	*Salinispora sp.*	Papau New	M.S.		67	RJA4303	*Streptomyces sp.* HB149 89%	*Phorbas* Sp. B.C.	M.I.
31	RJA3112	*Salinispora sp.*	Dominicana #2	M.S.		68	RJA4374	*Streptomyces sp*	RJA4020 & 4241	Fusant
32	RJA3113	*Salinispora sp.*	Dominicana #2	M.S.		69	RJA4375	*Streptomyces sp*	RJA4020 & 4241 #3	Fusant
33	RJA3114	*Salinispora sp.*	Papau New	M.S.		70	RJA4377	*Streptomyces sp*	RJA4020 & 4241 #5	Fusant
34	RJA3115	*Salinispora sp.*	Dominican #1	M.S.		71	RJA4379	*Streptomyces sp*	RJA4020 & 4173 #1	Fusant
35	RJA3116	*Salinispora sp.*	Papau New	M.S.		72	RJA4380	*Streptomyces sp*	RJA4020 & 4173 #3	Fusant
36	RJA3117	*Salinispora sp.*	Dominican #2	M.S.		73	RJA4382	*Streptomyces sp*	RJA4020 & 4173 #5	Fusant
37	RJA3118	*Salinispora sp.*	Papau New	M.S.		74	RJA4384	*Streptomyces sp*	RJA4020 & 4242 #11	Fusant

*T.S = Terrestrial Strain; M.S. = Marine Strain; M.I. = Marine Invertebrate

Table 5-1. List of strains selected for mini library. RJA number, species name if known, location or species origin, accession number if 16S RNA gene deposited.

5.3.1 Phosphatase Assay

Regulation of cellular function is critical to life. Reversible phosphorylation is an important mechanism of cellular regulation with dephosphorylation performed by phosphatases. There are several phosphatases of medical interest: PP1-c and calcineurin. Both of these proteins belong to the serine/threonine phosphatase class of enzymes, with the PP1 catalytic subunit dependent on binding with regulatory subunits for activity and specificity of action, while calcineurin has a calmodulin-binding catalytic subunit and a Ca^{2+}-binding regulatory subunit.[252] PP1-c inhibitors are often associated with cytotoxic activity. Calcineurin is involved in the immune system's responses to foreign bodies. Thus, calcineurin inhibitors, such as cyclosporin A, are universally important in modern transplant surgery, yet there are only a handful of known compounds, all with harmful side effects.[9]

5.3.2 Anti-Inflammatory Assays

Interleukin-6 (IL-6) is a keystone cytokine involved in regulating the immune response, capable of causing diseased states if it is either over-produced or under-produced.[253] On the one hand, the absence of IL-6 will allow pathogens and pathogen-related diseases to proliferate. On the other hand, elevated levels of IL-6 have been shown to be involved in both chronic inflammation diseased states such as psoriasis and arthritis and the development of certain cancers such as B cell malignancy and myeloma.[253,254] Thus, natural products that can act as lead compounds for inducing IL-6 pro-inflammatory or anti-inflammatory responses have many potential applications.

5.3.3 Anti-microbial Assays

The antibacterial technology first harnessed in penicillin has vastly decreased the health threat that once marked bacterial infections.[12] However, over-utilization of these natural products

has saturated the environment, leading to widespread antibiotic resistance to currently used medicines in human pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and E. coli.[255] This developing resistance drives the search for novel antibiotics and natural products with antibacterial activity that could be developed into new medicines. Additionally, there are 1.5 million yearly deaths from fungal infections worldwide, both from primary fungal infections and secondary fungal infections in immunosuppressed individuals.[256,257] This is in part due to the low number of antifungal drugs that can be clinically prescribed. Despite this, the search for novel antifungal compounds is underrepresented in natural product discovery programs, while new drugs are needed.[257] As such, *Candida albicans* is an excellent model fungal pathogen for use in bioassay-guided fractionation efforts.

Several qualitative assays may be used to assess the production of antibacterial compounds, which can also guide the isolation process. The standard disc-diffusion antimicrobial assay is quick and relatively easy.[258,259] The bioautography assay[260,261] combines the rapid results of thin-layer chromatography's separations with an assortment of biological tests.[262,263] It is particularly useful in targeting antibacterial/antifungal activity in TLC-Agar-Overlay bioautography assays.[71,260]

5.4   Results from Bioassay-guided fractionation

5.4.1 Phosphatase Assay Results

Extracts from five strains demonstrated PP-1c inhibition of interest in the pNPP assay. An interesting result in the pNPP assay will be any crude extract that shows at least 30% inhibition of activity while extracts demonstrating around 50% inhibition marking a strain of particular interest. RJA3005, a *Salinispora arenicola* strain, showed PP-1c and potent calcineurin inhibition in the pNPP assays and was scaled-up with initial results demonstrating a positive

increase in activity upon fractionation. This strain will be discussed further in **Chapter 6**. Prior

work in the Andersen lab demonstrated that *Streptomyces sp.* RJA4178 is a *Streptomyces* strain

whose extracts demonstrated PP-1c and calcineurin inhibition. Through bioassay-guided

fractionation efforts (**Scheme 5-1**), thienodolin (**5.1**) [264] was found as a very weak calcineurin

inhibitor, along with a family of known thiazole indole derivatives (**5.2, 5.3, 5.4**), one of which

was also a very weak calcineurin inhibitor. Interestingly, none of these compounds had been

reported in *Streptomyces* species previously, with **5.3** being an unpublished natural product

discussed in a patent for dye compounds,[265] and **5.2** and **5.4** being found in thermophilic

bacteria.[266–268]
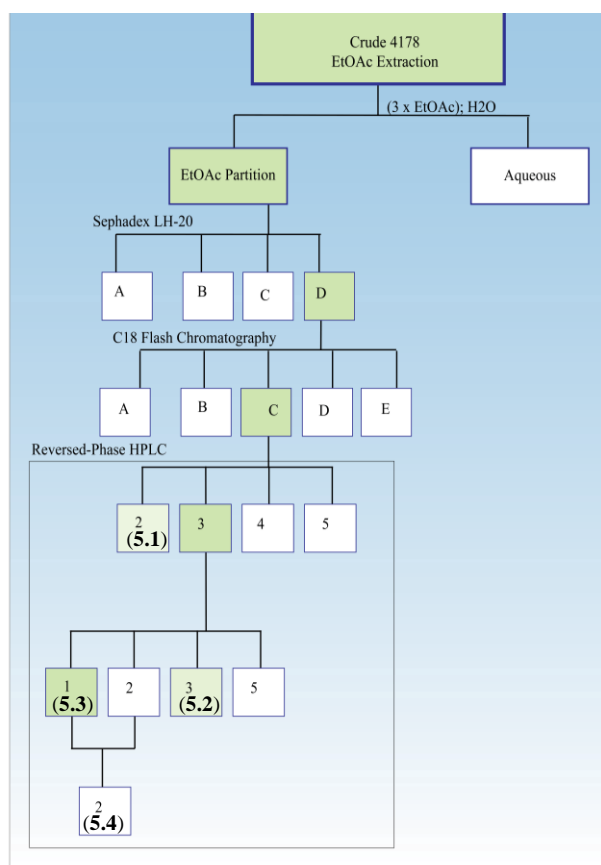
Scheme 5-1. Isolation scheme of *Streptomyces* sp. RJA4178. Darker green above is associated with a more pronounced activity.
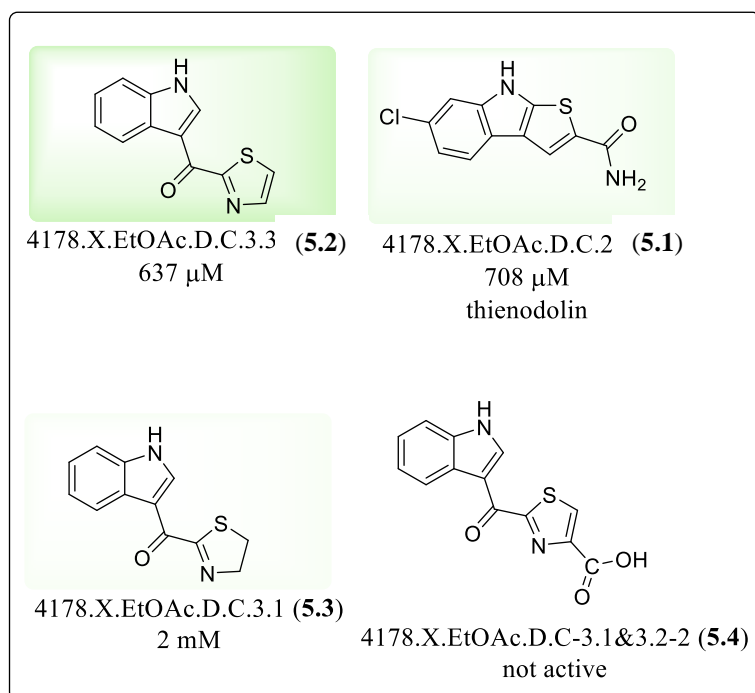
Figure 5-1. Compounds isolated from *Streptomyces* sp. RJA4178. Darker green above is associated with a more pronounced activity.

5.4.2 Anti-Inflammatory Assay Results

Positive anti-inflammatory results are considered of interest in the IL-6 whole-cell assay if the crude extract matches or exceeds the 88% inhibition displayed by the dexamethasone control. The crude extract of RJA4249 showed 97.9% IL-6 inhibition and was the focus of my efforts. RJA4249 was demonstrated to be a *Streptomyces sanglieri* species through 16s rRNA sequencing completed by the author. It was seen that the Streptomyces sp. RJA4249 was producing the known compound amicoumacin (**5.5**) which had only previously been reported from *Bacillus* species.[269] It was also producing various diketopiperazines (**5.6** - **5.8**) and several known cycloheximide analogues (**5.9** and **5.10**).[270] Through bioassay-guided fractionation efforts (**Scheme 5-2**), the fraction 4249.DCM.CDE.D.A.6 (**5.11**) was identified as the known natural product cycloheximide. Cycloheximide is a potent antifungal and antibiotic natural

215
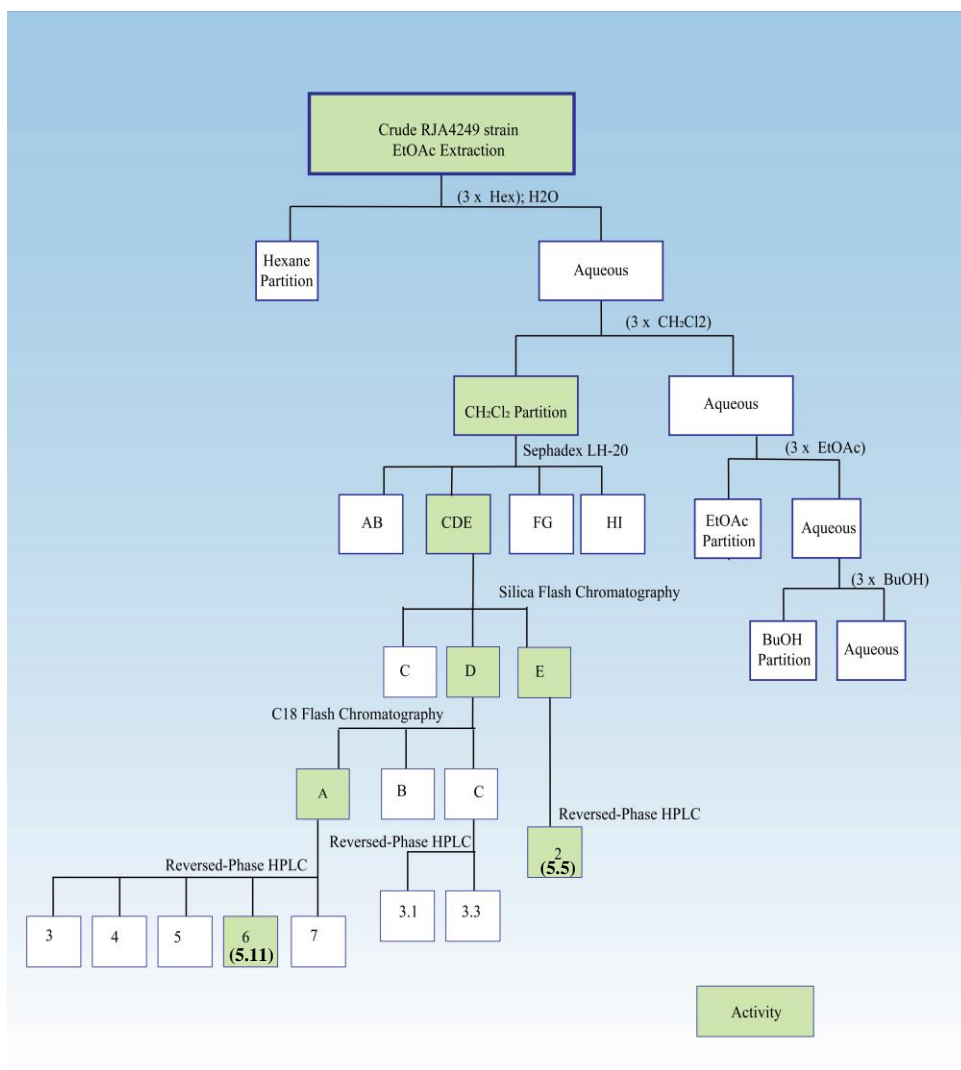
Scheme 5-2. Isolation scheme of the *Streptomyces sanglieri* RJA4249 leading to the known compounds; amicoumacin (5.5), cycloheximide (5.11), and cycloheximide analogs. Green above is associated with activity.

product that inhibits protein synthesis of proteins. While it was previously used as a medicine in cases of severe fungal-infections,[271] and for the potential treatment of Hodgkin's disease at low-doses, the observations of alteration in protein metabolism at the required doses along with nausea after prolonged use[272] resulted in discontinued use of cycloheximide as a common drug. Recent research has suggested that low doses of cycloheximide work well in conjunction with other medicine for anti-cancer therapy, but it has yet to be incorporated as a treatment.[273]

216
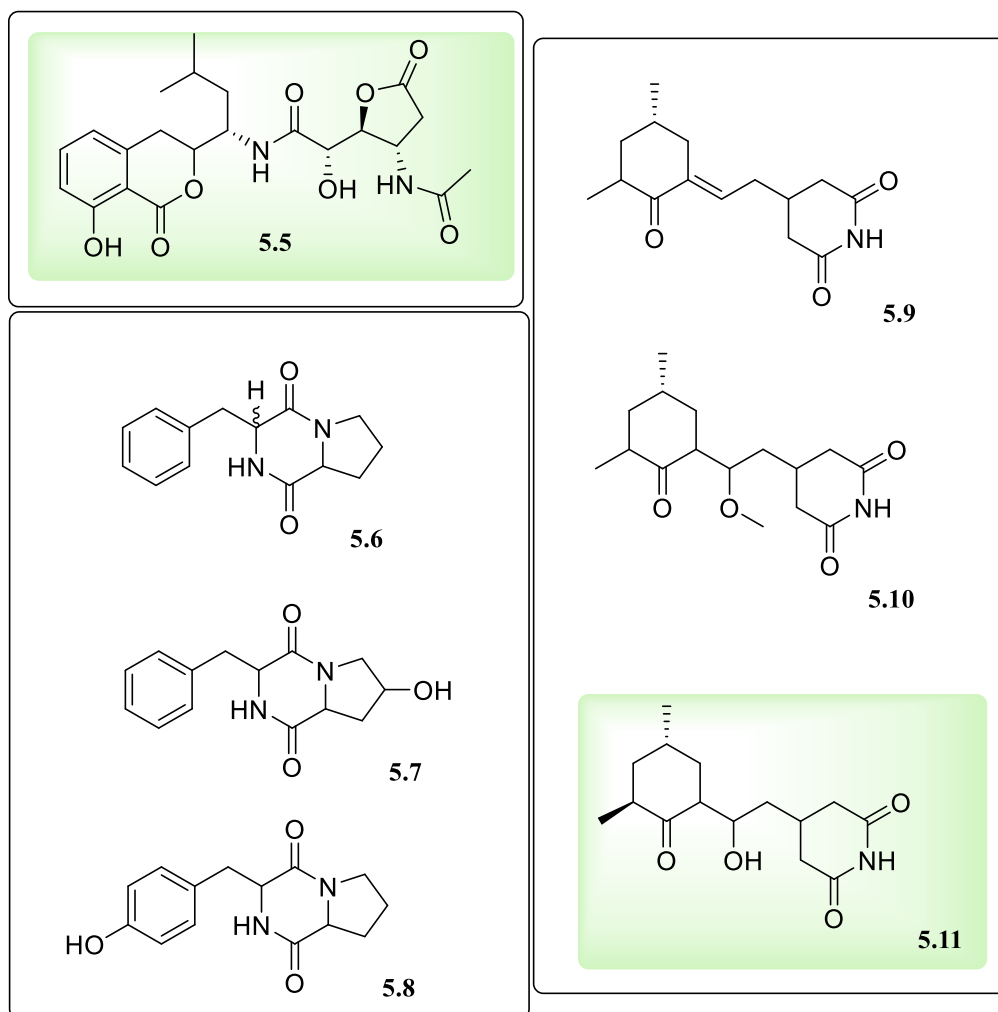
Figure 5-2. Compounds isolated from *Streptomyces sanglieri* RJA4249. Compounds highlighted in green were found through apparent activity. Green above is associated with activity.

Nevertheless, cycloheximide is routinely utilized in scientific labs and agricultural applications.[274] The fractions were retested along with a quantity of purchased cycloheximide; this demonstrated that there were equivalent anti-inflammatory responses, thus describing the origin of *Streptomyces* sp. RJA4249 potent anti-inflammatory results.

5.4.3 Anti-microbial Assay Results: Fusant Strains

The strains RJA4374-4384 are bacterial protoplast fusants, arising from a genetic technique that allows for homologous recombination between closely related strains.[275] These were prepared from various Andersen library *Streptomyces* strains. In all cases, a *Streptomyces caeruleus* RJA4020 (JF510466) isolated from Georgia Strait cold waters in Vancouver, British Columbia[249] and displaying no activity, was fused with a *Streptomyces* strain displaying potent anti-MRSA activity by Dr. D. Dalisay (**Figure 5.3**). These fusants demonstrated different morphology both from the parents and each other and, despite the strong antibacterial activity of some of the parents, the fusants lacked this trait (**Figure 5.3**). Given the potent antifungal activity of the extracts, a bioautography assay was chosen as the means to rapidly guide[262] the isolation of the antifungal compound.
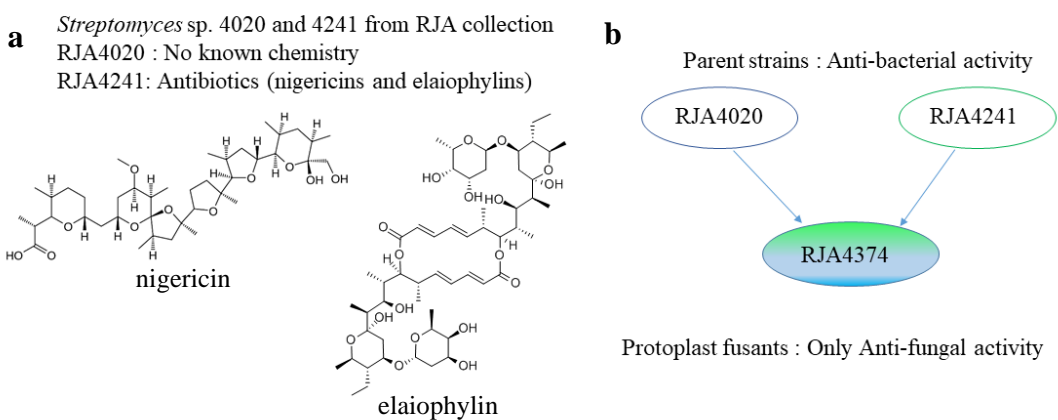
Figure 5-3. Fusant strains. a) Identity and chemistry being produced from parent strains. b) The change in observed activity led to interest in the fusant strains.

This bioautography assay-guided fractionation (**Scheme 5-3**) allowed the identification of the active components of fusant RJA4374 as members of the potent known antifungal antimycin family,[276] including antimycin A_{3ab} (4374.Hex.C.E.4.1 (**5.12**)) and antimycin A_{2ab} (4374.Hex.C.E.8.1 (**5.13**)). The isolation of these antimycins suggests that the fusant's *de novo*

antifungal activity most likely arises from the novel expression of a silent gene cluster, possibly

through insertion or deletion of a regulatory gene during the homologous recombination involved

in the fusant process.
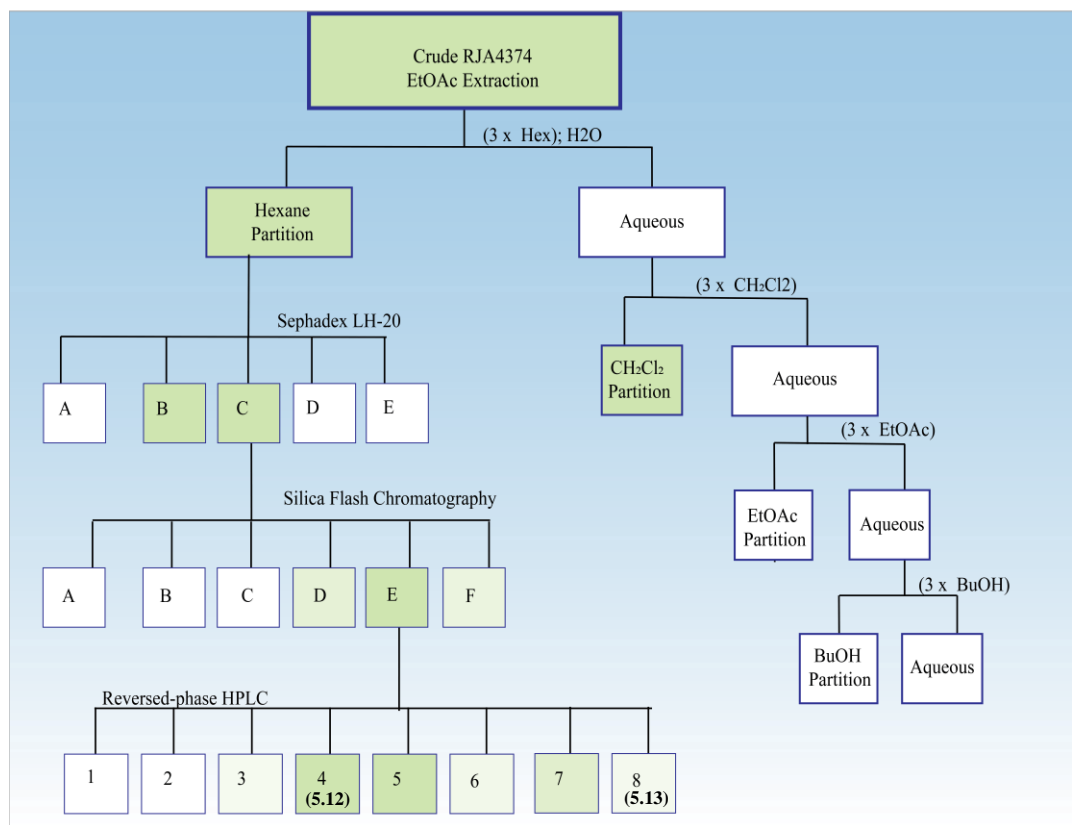
Scheme 5-3. Isolation scheme of the fusant RJA4374 leading to the known antifungal, antimycin. The darker green above is associated with a more pronounced activity.
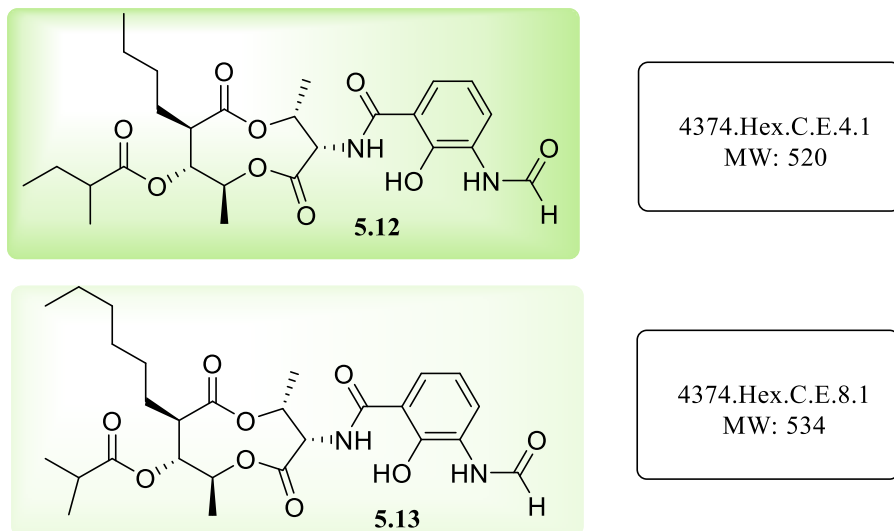
Figure 5-4. Antimycins responsible for the antifungal activity of the fusant RJA4374. There was seen to be some difference in activity. The darker green above is associated with a more pronounced activity.

5.5   Summary and Conclusion

A mini-library of marine and terrestrial bacterial were grown and assessed for potential bioactive drug leads through a small number of bioassays. Several of the bacterial cultural extracts demonstrated positive hits in the bioassays. Subsequently, the bioassays were used to guide the fractionation of the bacteria's complex metabolite pool. These bioassay-guided fractionations yielded known compounds. Several thiazole natural products not previously found from *Streptomyces* species were revealed. Amicoumacins were shown to be produced by *Streptomyces* species in addition to the original *Bacillus* producer. It was shown that the *de novo* antifungal activity of generated fusant strains of cold water bacteria had a novel expression of antimycins. Finally, through the construction of the mini-library, a *Salinispora arenicola* species from the RJA collection was seen to be producing some previously unknown natural products which will be discussed in **Chapter 6**.

Chapter 6: *Salinispora arenicola* RJA3005: Chemistry and Feeding Studies

6.1   Chapter Overview

Crude extracts from cultures of *Salinspora arenicola* RJA3005 were initially investigated due to promising bioassay results (**Chapter 5**), however, no metabolites could be isolated that had these activities following isolation. Nonetheless, NMR examination of purified fractions of the extract revealed natural products 6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-hydroxy-3-methyl-2H-pyran-2-one (**6.1**) and N-(3-hydroxy-5-(1-hydroxy-2-(4-hydroxy-3-methyl-2-oxo-2H-pyran-6-yl)propyl)phenyl)acetamide (**6.4**). **6.1** has been previously seen as the product of feeding an engineered *Amycolatopsis mediterranei* S699 strain bacterial mutant (-AHBA synthase) with synthetic aromatic precursors.[277,278] **6.4** has been previously seen as a shunt product of: an *Amycolatopsis mediterranei* S699 mutant engineered without multiple sections of the wild-type's post-PKS modification genes and an *A. mediterranei* S699 mutant with a *rifF* deletion.[277,279] As a product of these mutants, **6.4** was named SY4b but not completely characterized. As the natural product seen here, I will complete its characterization and refer to the natural product **6.4** by its chemical name 3-acetamide-5-hydroxyphenylpyrone.

Following structure elucidation of 6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-hydroxy-3-methyl-2H-pyran-2-one, derivatization was performed in an unsuccessful attempt to isolate related metabolites that were challenging to isolate. However, this derivatization allowed separation of the **6.1**'s tautomers, one of which was subsequently utilized to enable a Mosher's ester analysis. Consideration of the possible biosynthetic pathway leading to the isolated natural products led to the hypothesis that none of the known literature pathways to a 3,5-dihydroxy aromatic substitution pattern could be reasonably followed to make this molecule. To interrogate the biosynthetic assembly of 6.1's 3,5-dihydroxy aromatic substitution pattern, I carried out

feeding studies with $1,2$-$^{13}C_2$-acetate and U-$^{13}C_6$-glucose to track the ^{13}C coupling patterns. The analysis of these coupling patterns strongly suggested that the aromatic moiety of the natural product **6.1** is being produced through PEP and erythrose building blocks.

6.2   Salinispora: A Marine Obligate Genus and it's Natural Products

Salinispora is a marine obligate bacteria genus in the Micromonosporaceae family, Actinomycetales order, and the secondary metabolite prolific, Actinobacteria class and phylum. The first described members of this genus were reported in the literature some 15 years ago.[28] Indeed, the Salinispora and Micromonospora genera are very closely related, both exhibiting a leathery orange appearance until black spores develop. The significant phenotypic difference between the genera is Salinispora's higher salt requirement for growth.[280] Initially, two *Salinispora* species were reported, *Salinispora arenicola* and *Salinispora tropica*,[28] although a third, *Salinispora pacifica*, was quickly proposed[65] and later, confirmed.[29,281] Very recently, six new *Salinispora* species have been characterized, all demonstrating a closest relationship to *Salinispora pacifica*.[282] *Salinispora arenicola*, *Salinispora tropica*, and *Salinispora pacifica* species are distinguishable in part due to a different distribution of secondary metabolites and their corresponding gene clusters,[283,284] despite an overlap in ecological ranges.[65,281] All three are prolific producers of secondary metabolites, boasting genomes (~6 Mb) rich with natural product gene clusters, and the differential distribution of secondary metabolite genes means that they produce many different natural products.[67] *Salinipora pacifica* uniquely produces cyanosporaside A, salinipyrone A, lomaiviticin A, enterocin, and pacificanone A.[285] *Salinispora tropica* uniquely produces the potent and specific anti-cancer drug salinosporamide A (Marizomib),[31] sporolide A, salinilactam, sioxanthin, and antiprotealide.[285] *Salinispora arenicola*, perhaps the most abundant and widely dispersed of the species, produces arenicolide A, saliniquinone, cyclomarin A,

222

cyclomarazine, arenimycin, arenamide A, staurosporines, mevinolin and the related saliniketal A and rifamycin B.[67,285]

6.2.1   Biosynthesis of Rifamycin and Saliniketal

Rifamycin is a major metabolite of *S. arenicola* strains.[280] Additionally, *S. arenicola* strains produce multiple natural products, such as saliniketal A and salinisporamycin that share a biosynthetic pathway with rifamycin (**Figure 6.1**).[286,287] Key to all of these biosynthetic pathways is the starter unit 3-amino-5-hydroxybenzoic acid (AHBA), a glucose-derived metabolite loaded onto the PKS modules in a specialized loading module.[68,132,278,288] AHBA has a dedicated suite of genes adjacent to rifamycin PKS genes (**Figure 6.1, Table 6.1, Figure 6.2**) and is not made from shikimic acid.[68,132,289,290] Instead, AHBA biosynthesis, or the aminoshikimate pathway, closely mirrors the shikimic acid pathway. The shikimate pathway produces 3-deoxy-D-arabinoheptulosonate-7-phosphate (DAHP) through phosphoenolpyruvate (PEP) and erythrose building blocks. The aminoshikimate pathway utilizes AHBA synthase, a PLP-dependent aminotransferase, to ultimately produce a 3-amino-3-deoxy-D-arabinoheptulosonate-7-phosphate (aminoDAHP) through PEP and iminoerythrose building blocks. The AHBA synthase catalyzes the formation of the iminoerythrose and the amino dehydroshikimate equivalent to form AHBA (**Figure 6.2**) in the final step of the pathway.[132] In the shikimic acid pathway, the carbonyl functionality introduced in the first step is maintained until the final step, at which time it is reduced to an alcohol forming shikimate (**Figure 6.3**).[132]

Gene	Function
orf1	Hypothetical function
rifG	Aminodehydroquinate synthase
rifH	AminoDAHP synthase
rifI	Aminoquinate dehydrogenase
rifK	AHBA synthase
rifL	Oxidoreductase
rifM	Phosphatase
rifN	Kanosamine kinase
rifJ (far apart from the rest of the genes)	AminoDHQ dehydratase

Table 6-1. Gene cluster of required genes for production of AHBA.
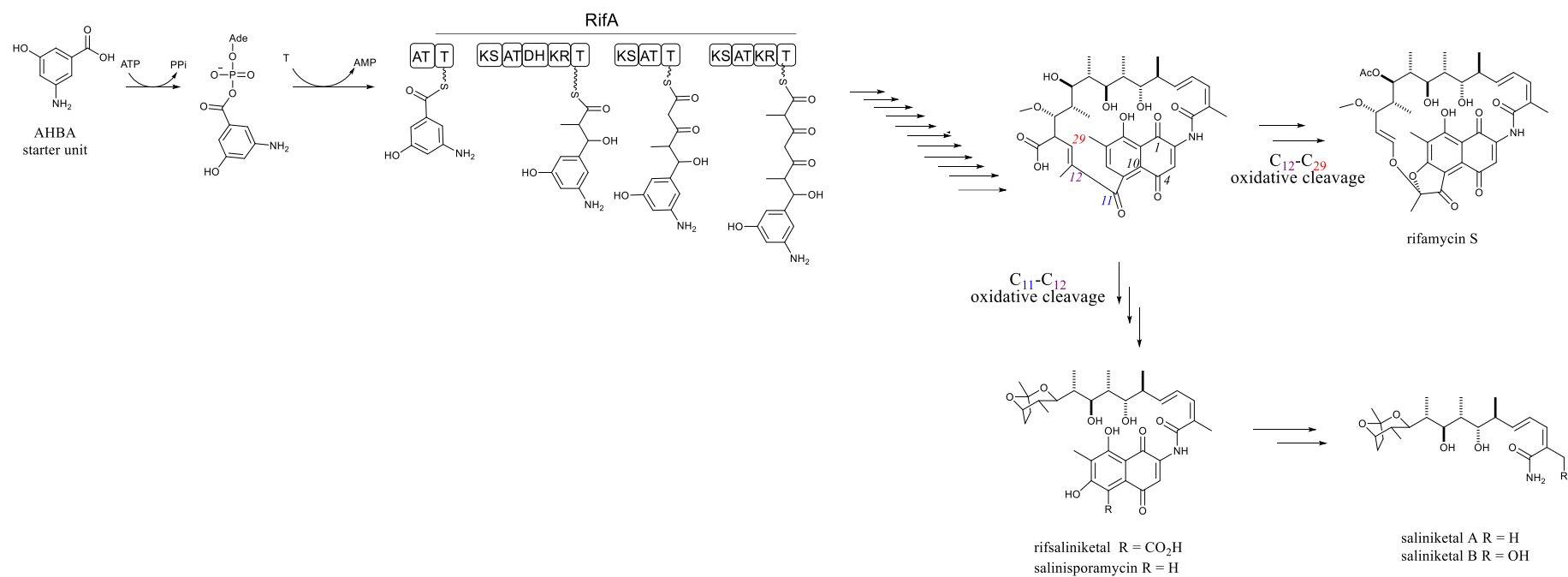
Figure 6-1. AHBA starter unit leading into the polyketide modules. Shared pathway leading to the rifamycins and saliniketals, depending on the carbon bond that is cleaved. Adapted from various sources.[68,132,278,288,289]
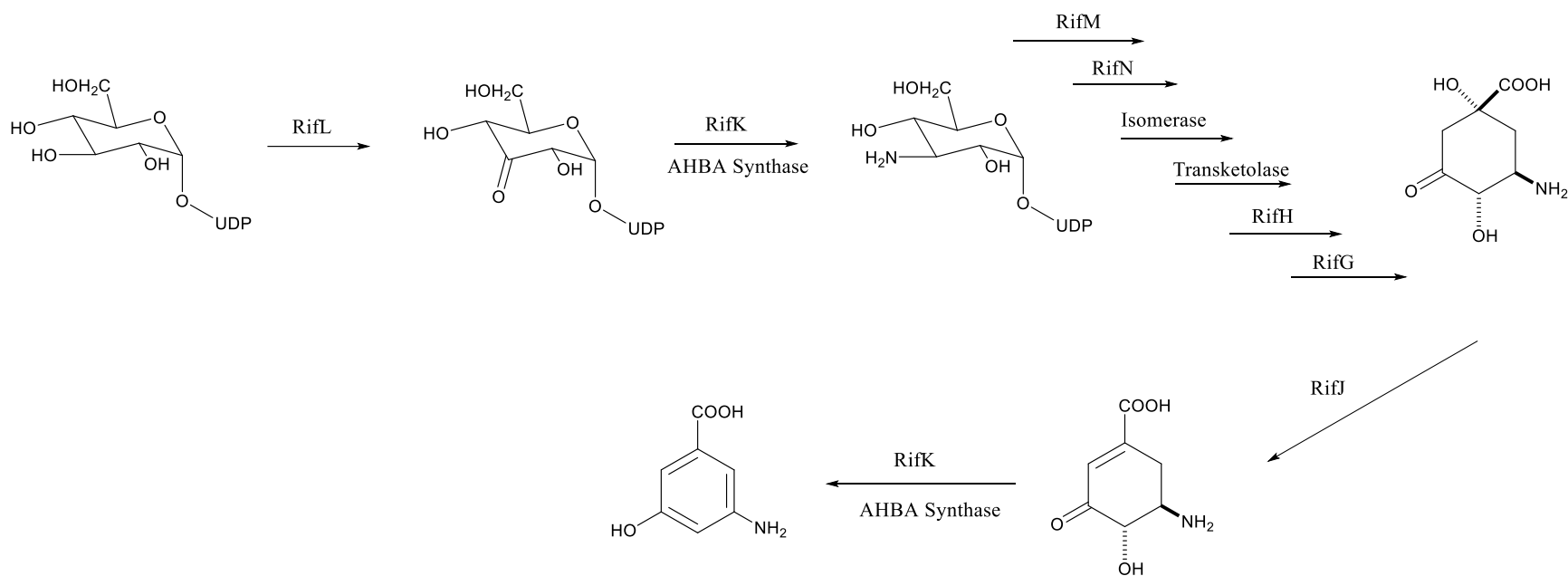
Figure 6-2. Key steps during the biosynthesis of AHBA from glucose.[132] Please see Table 6.1 for enzyme functions.

Dehydroshikimic acid
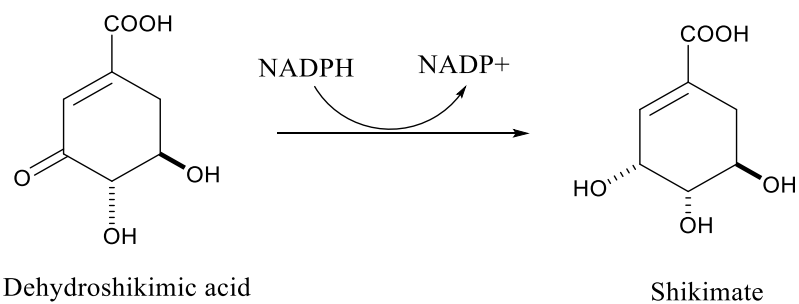
Shikimate

Figure 6-3. The final step in the shikimate pathway. This is a large difference between the shikimic acid pathway and the aminoshikimate (AHBA) pathway as the latter continues to aromatization.

6.3 Isolation and Structure Elucidation of Metabolites from *Salinispora arenicola* RJA3005

Salinispora arenicola RJA3005, a marine obligate strain isolated by Dr. Doralyn Dalisay from marine sediment collected in Papua New Guinea, was selected for further investigations following bioassay screening of the previously mentioned mini-library (**Table 5.1**). This strain showed promising activity in a phosphatase inhibition assay run (**Table 5-2**). Subsequent bioactivity-guided isolation efforts (**Scheme 6-1**) revealed final pure compounds that did not demonstrate the same bioactivity as the initial complex fractions. However, chemical prospecting of varying fractions revealed several related natural products.

S. arenicola RJA3005 was grown as lawns on MM1 solid agar prepared with seawater. After three weeks of growth, the solid agar was harvested and soaked in EtOAc. The EtOAc extracts were filtered, dried under vacuum, and subsequently partitioned between EtOAc and water. The resulting brown gum was subjected to size exclusion chromatography on a Sephadex LH-20™ column (4:1 MeOH:DCM) followed by step-gradient separation on a C18 flash column with the compounds of interest eluting with 4:1 H_2O:MeOH (Fraction C) and 3:2 H_2O:MeOH (Fraction E). Fraction C was a yellowish, oily solid that was subjected to semi-preparative reverse-phase HPLC (17:83 MeCN:H_2O with 0.05% Formic Acid (FA)) to reveal several diketopiperazines, tryptophan derivatives, shikimate pathway metabolites and 6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-hydroxy-3-methyl-2H-pyran-2-one (**6.1**) whose structure elucidation follows in **Section 6.3.1**. Fraction E was a light brown oily solid that was further separated via semi-preparative reverse-phase HPLC (17:83 MeCN:H_2O with 0.05% FA) to yield N-(3-hydroxy-5-(1-hydroxy-2-(4-hydroxy-3-methyl-2-oxo-2H-pyran-6-yl)propyl)phenyl)acetamide (**6.4**) whose structure elucidation is found in **Section 6.3.2**.
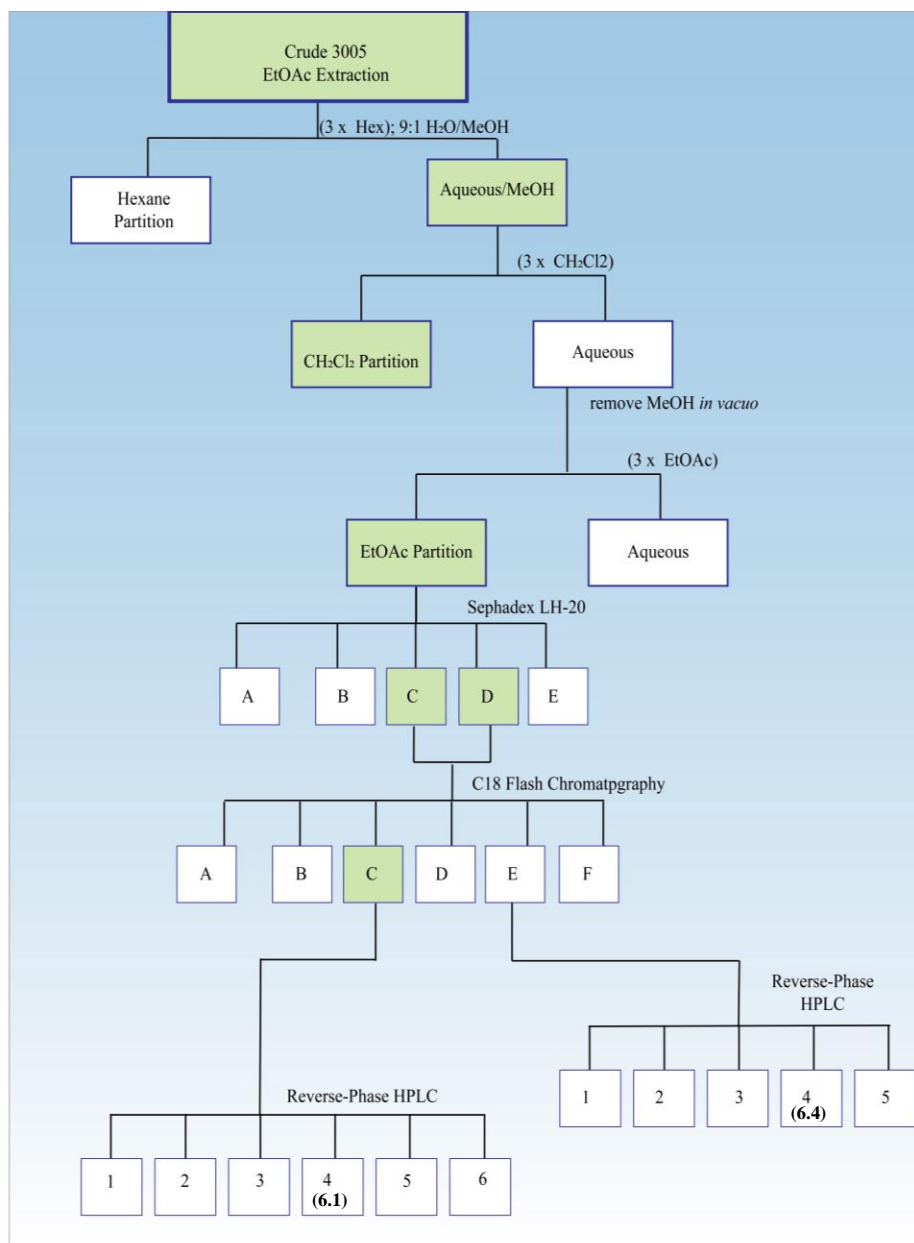
Scheme 6-1. Isolation scheme for *Salinispora arenicola* RJA3005 culture extract.

6.3.1 Structure Elucidation of 3,5-dihydroxyphenylpyrone (6.1)

3,5-dihydroxyphenylpyrone (**6.1**) gave a $[M + H]^+$ ion in the HRESIMS at *m/z* 293.1046 appropriate for a molecular formula of $C_{15}H_{16}O_6$ that requires eight sites of unsaturation. ^1H, ^{13}C, ^1H-^1H gCOSY60, ^1H-^{13}C gHSQC, ^1H-^{13}C gHMBC and ^1H-^1H tROESY NMR data recorded in DMSO-d*$_6$* (See **Table 6.2**) revealed **6.1** as a small metabolite that contains a disubstituted

benzene ring attached to a pyrone functionality. The interesting feature of **6.1** was the relatively

shielded aromatic protons at δ_H 6.16 (H-2 and H-6) and δ_H 6.09 (H-4) exhibiting a $^4J_{H-H} = 2.2$ Hz

which, in the context of aromatic protons, is usually indicative of *meta* coupling. Moreover, the
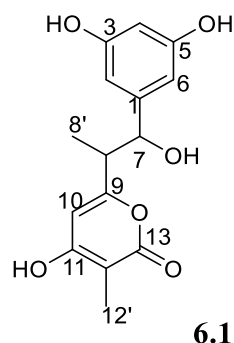
proton at δ_H 6.16 was integrated for two protons (Figure 6.6), indicating

some symmetry in the aromatic region. Confirming the symmetry was a

quaternary carbon at δ_c 157.9 (**Figure 6.6 and 6.7**) that was more abundant

than other quaternary carbons in the ^{13}C spectrum. Both proton chemical

shifts at δ_H 6.16 and δ_H 6.09 demonstrated HMBC correlations into the

carbon at δ_c 157.9 (C-3 and C-5) (**Figure 6-11**). As well, a hydroxyl proton at δ_H 9.15 was also

demonstrating HMBC correlations into δ_c 157.9. ^3J HMBC correlations were seen from the

protons at δ_H 6.16 (H-2, H-4) to the resonances δ_c 104.8 (C-2 and C-6) and 101.5 (C-4) in the

meta position to the proton thus helping to confirm the substitution pattern of the benzene ring.
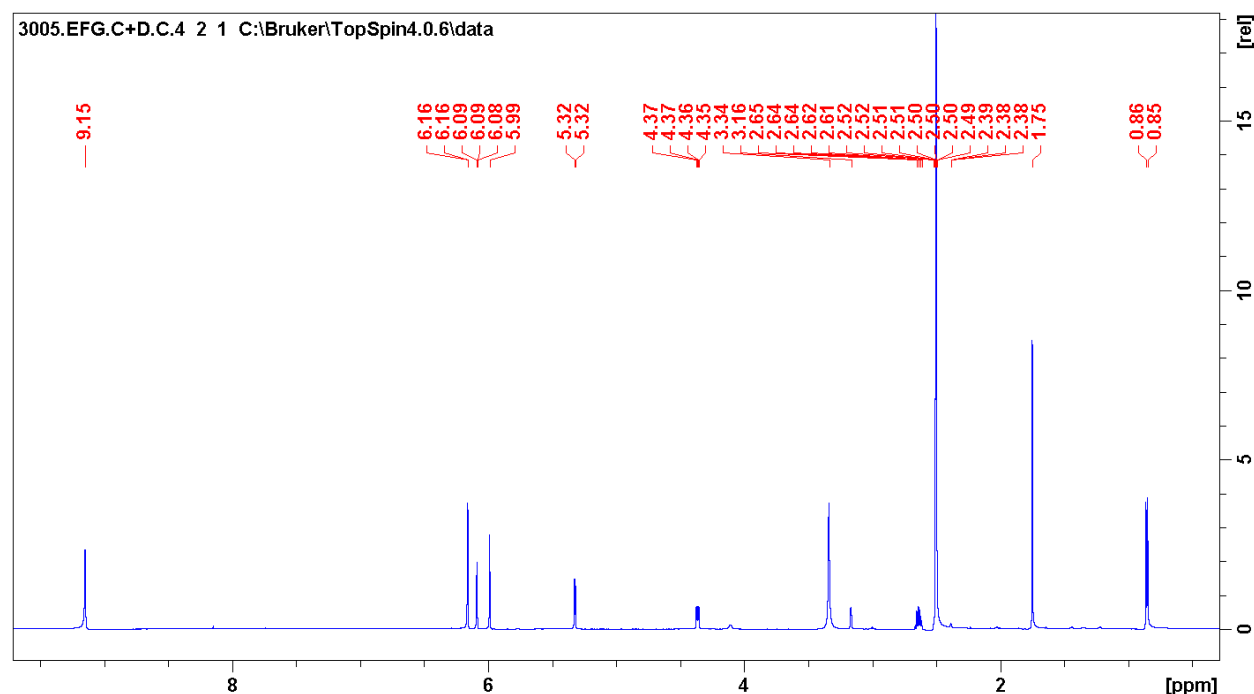
Figure 6-4. ^{1}H spectrum of 6.1, recorded in DMSO-d_6, at 600 MHz.

229
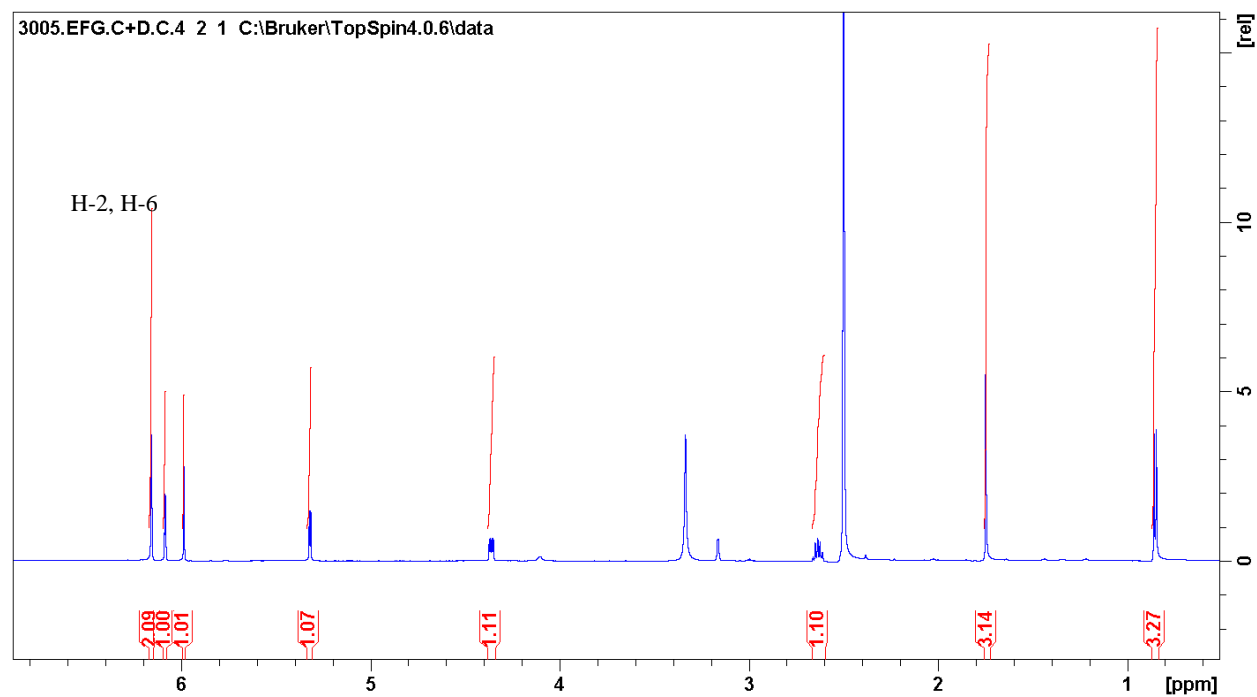
Figure 6-5. ^{1}H spectrum of 6.1 demonstrating integration, recorded in DMSO-d_6, at 600 MHz.
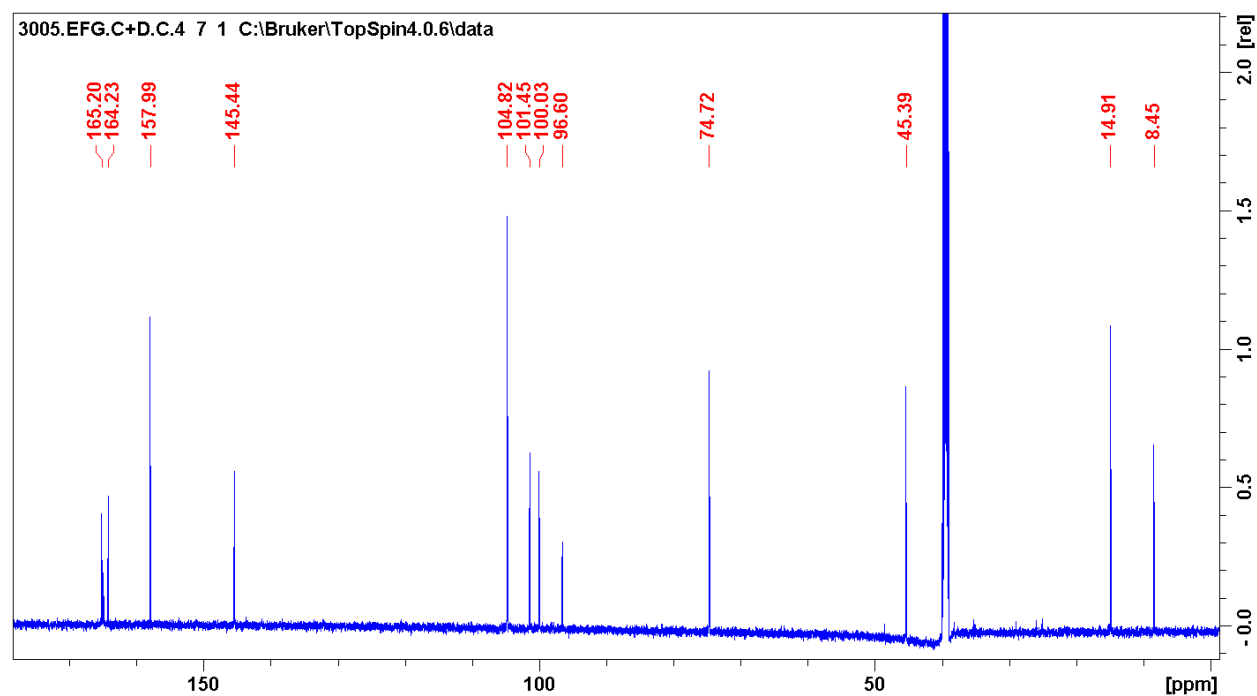
Figure 6-6. ^{13}C spectrum of 6.1, recorded in DMSO-d_6, at 150 MHz.
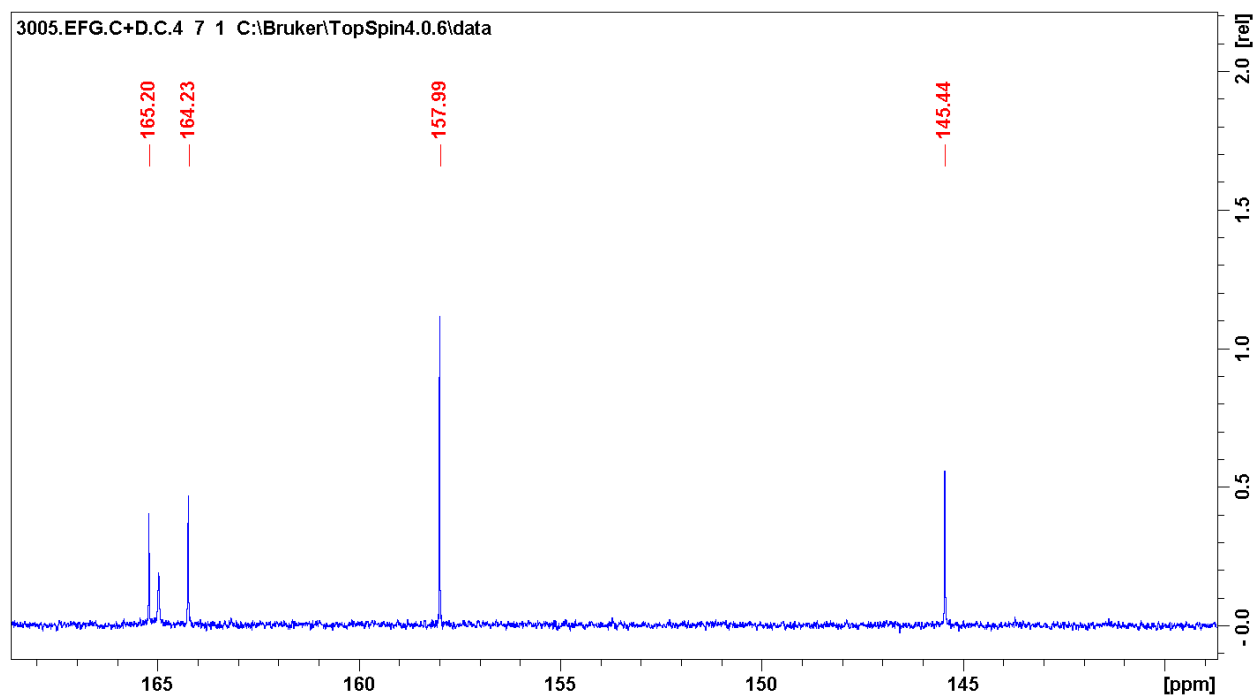
Figure 6-7. Downfield expansion of ^{13}C spectrum of 6.1, recorded in DMSO-d_6, at 150 MHz.
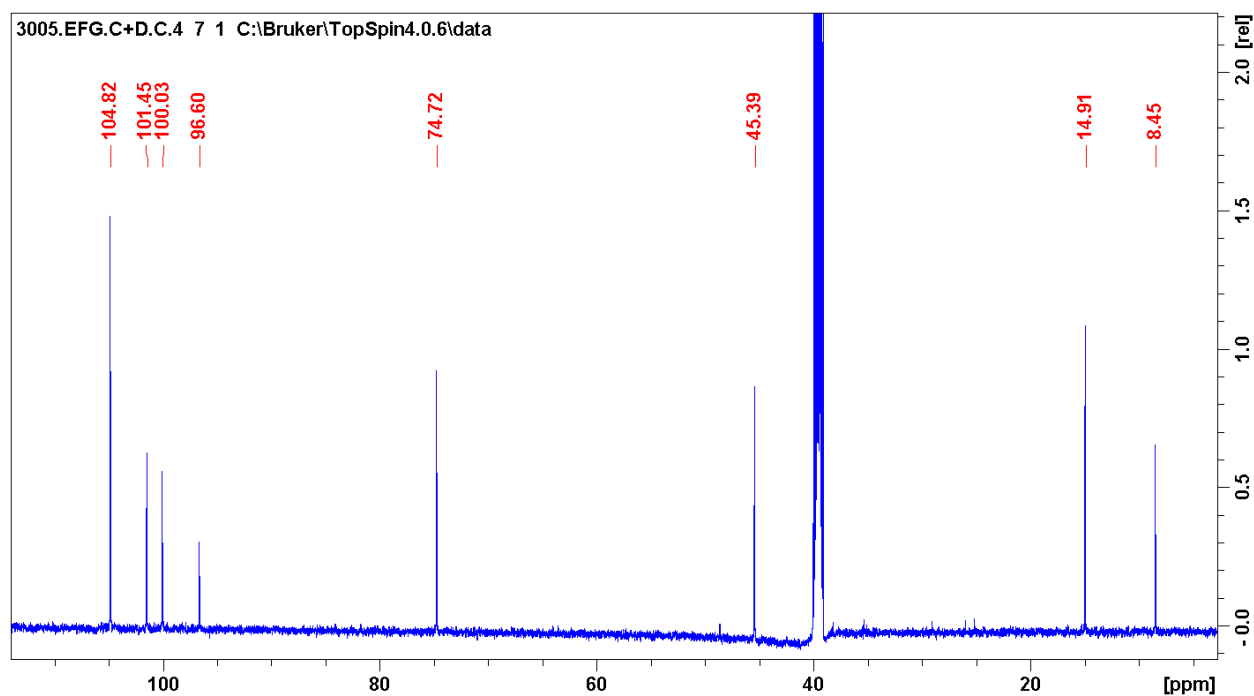
Figure 6-8. Expansion of ^{13}C spectrum of 6.1, recorded in DMSO-d_6, at 150 MHz.

A methine singlet at δ_H 6.16 (δ_c 100.0) anchored the elucidation of a pyrone ring (**Figure 6-11**) within the molecule with HMBC correlations into resonances δ_c 164.2 (C-9), δ_c 164.9 (C-

231

11) and δ_c 96.6 (C-12) (**Figures 6-11, 6-14, 6-16**). A methyl singlet at δ_H 1.75 (δ_c 8.5) demonstrated HMBC correlations into a carbonyl at C-13 (δ_c 165.2) and the resonance δ_c 100.0 (C-10, δ_H 6.16) (**Figure 6-18**). The methyl singlet at H-12' helped to link the benzene ring and the pyrone ring through unusual (and very weak) ^6J, ^7J, and ^8J HMBC correlations to C-8 (δ_c 45.4), C-8' (δ_c 14.8), and C-7 (δ_c 74.7) (**Figure 6-17**). More importantly, the methyl doublet at C-8' (δ_H 1.75, δ_c 14.8) showed HMBC correlations into the same resonance δ_c 164.2 (C-9) as the methine singlet at H-10 (**Figures 6-17, 6-18**). Additionally, the methine doublet of doublets at H-7 (δ_H 4.58, δ_c 74.7) demonstrated HMBC correlations into the quaternary carbons of the benzene (δ_c 145.5 (C-1)) and the pyrone ring (δ_c 164.2 (C-9)) along with the aromatic resonance δ_c 104.8 (C-2/6) (**Figure 6-18**). In the tROESY spectrum, H-7 showed strong through-space coupling to
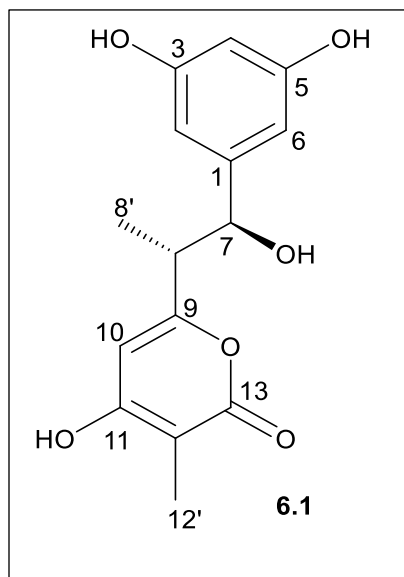
Figure 6-10. Numbering and structure of 6.1.
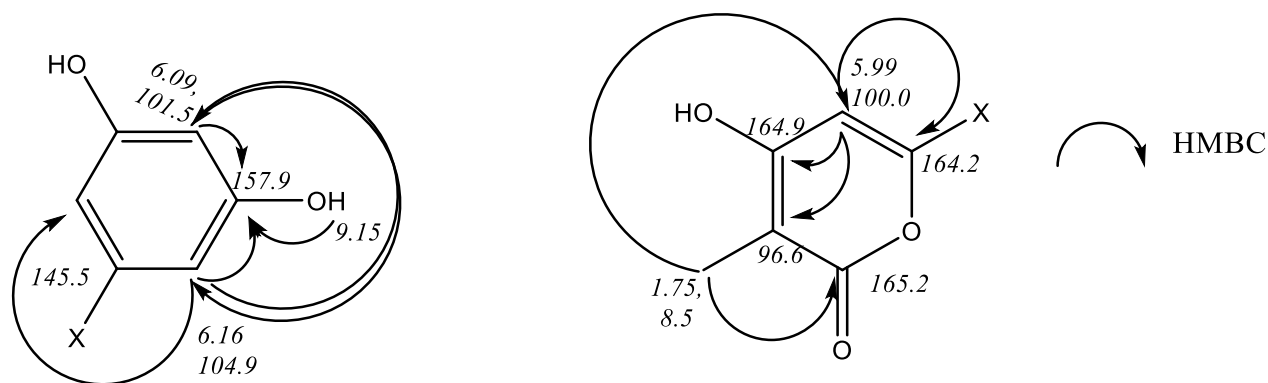
Figure 6-11. Selected HMBC correlations for the aromatic and pyrone regions of the 6.1 demonstrating key correlations.

the aromatic H-2 or H-4 proton (**Figure 6-20**) and the methine quartet at H-8 demonstrated

strong through-space correlations to both the methine singlet of the pyrone ring at H-9 and one of

the aromatic protons at H-2 or H-4. Taken all together, this data aided in the assignment of **6.1.**

6.1 is the first recorded isolation of this natural product from a wild-type bacterial culture, and

indeed as a biosynthesized molecule, as previous reports of **6.1** were from a mutated rifamycin

producer *Amycolatopsis mediterranei* S699 (-AHBA synthase) that was then fed 3,5-

dihydroxybenzoic acid.[277] It was noticed that although there was a sharp alcohol peak for OH-7

(δ_H 5.32), no such sharp peak was seen for OH-11, consistent with a possible keto-enol

isomerization. Work was ongoing to isolate several other clearly related compounds. However,

isolation was challenging, so derivatization of the fraction containing the compounds was

conducted to enable separation. Neither methylation nor acetylation allowed isolation of the

related compounds; however, methylation helped separate the isomers of **6.1**, which are seen as

two methyl ethers - **6.2** and **6.3** (**Figure 6-21**).
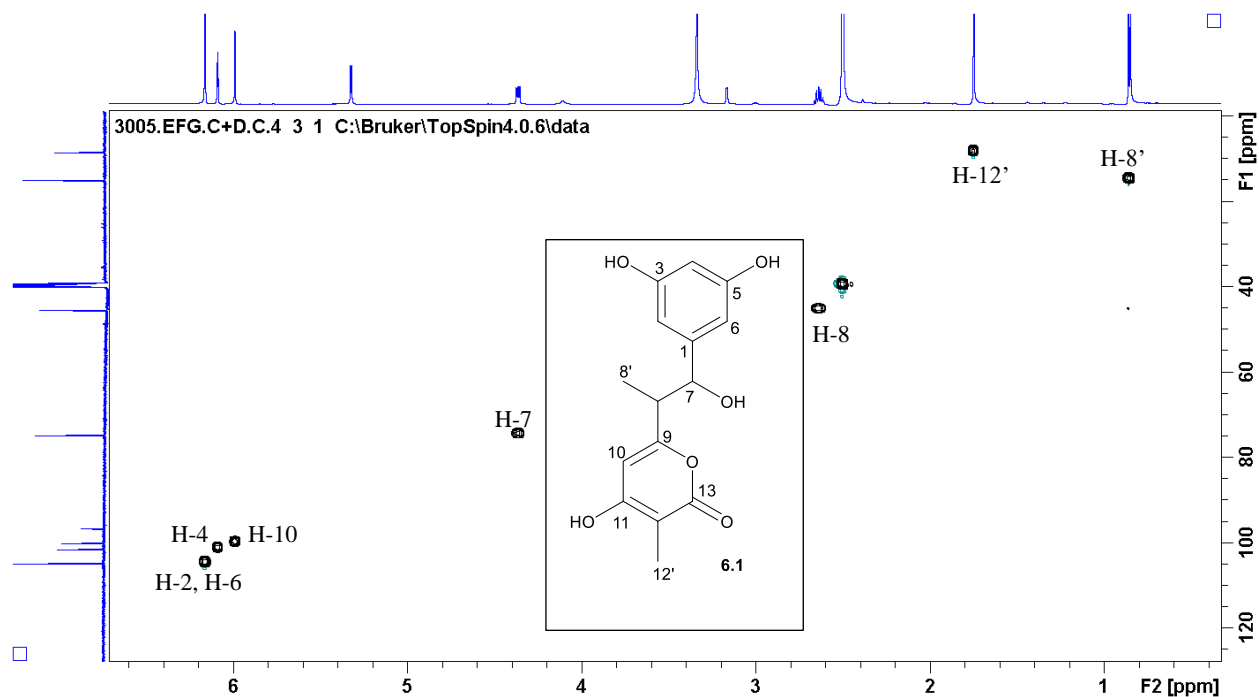
Figure 6-12. ^{1}H-^{13}C HSQC spectrum of 6.1, showing ^1J ^{1}H-^{13}C coupling, recorded in DMSO-d_6, at 600 MHz.
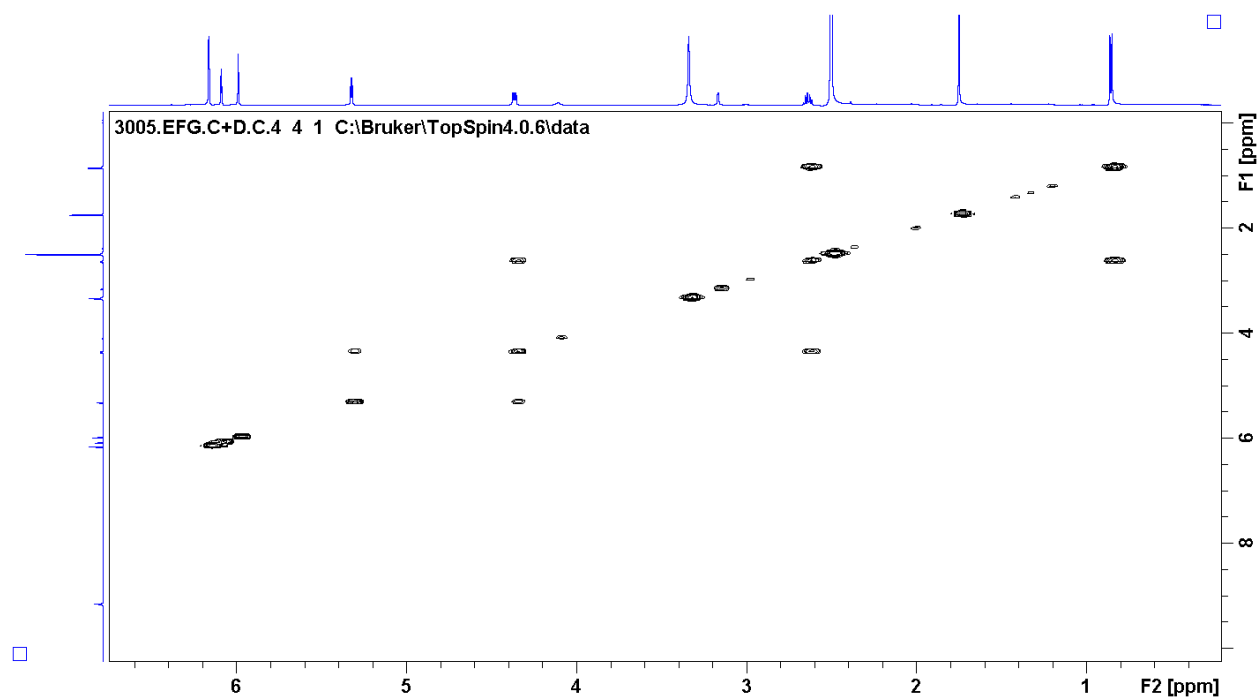
Figure 6-13. ^{1}H-^{1}H COSY60 spectrum of 6.1, recorded in DMSO-d_6, at 600 MHz. Few correlations were seen, which is in keeping with the many quaternary carbons of the molecule.
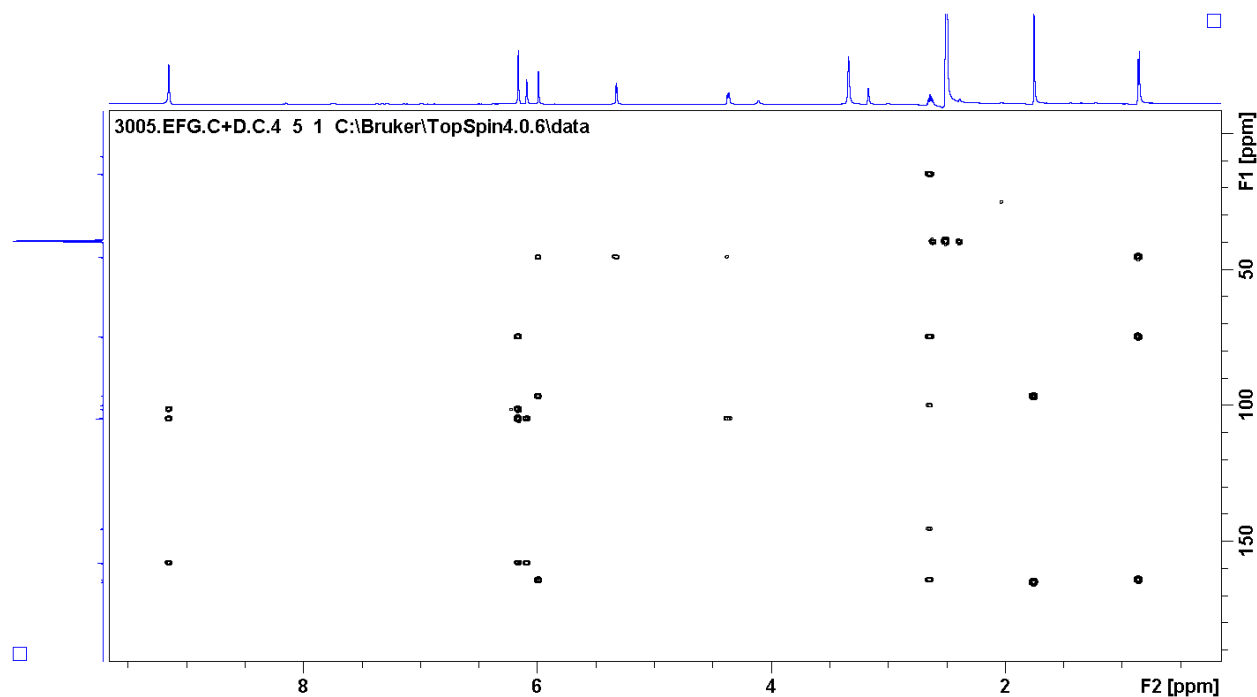
Figure 6-14. Full ^{1}H-^{13}C HMBC spectrum of 6.1, recorded in DMSO-d_6, at 600 MHz.
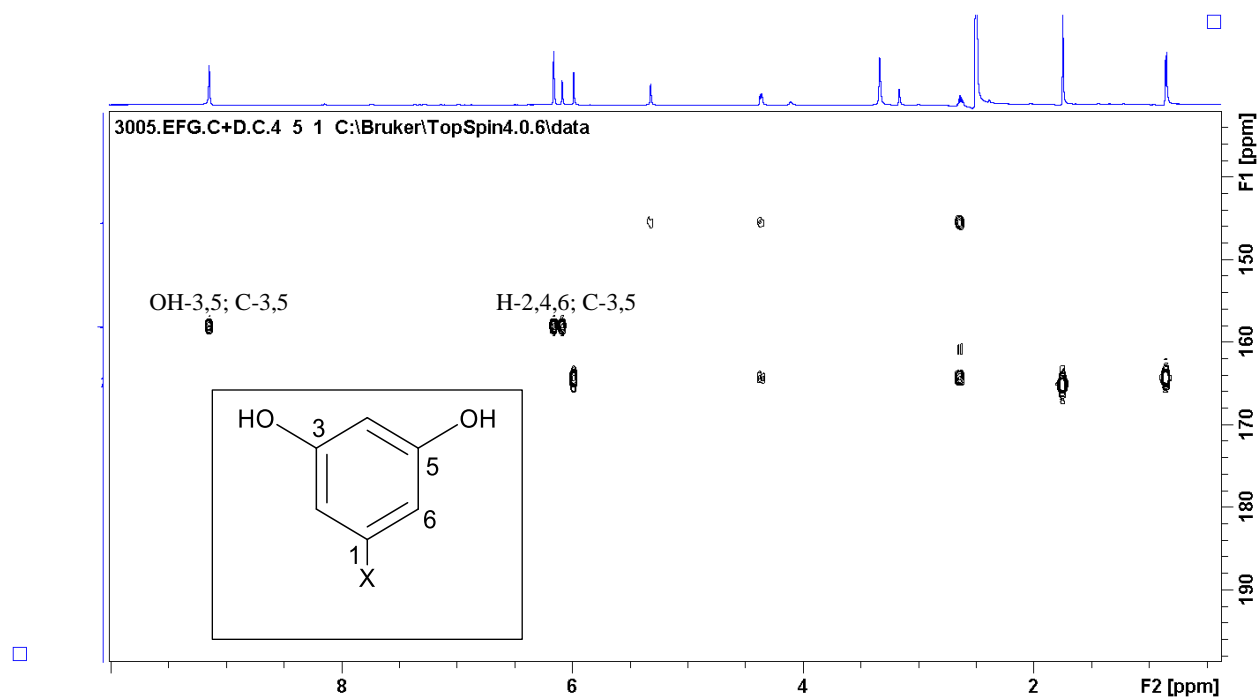
Figure 6-15. Expansion of ^{1}H-^{13}C HMBC resonances for 6.1 demonstrating the aromatic protons correlations into the quaternary carbons of C-3 and C-5. Recorded in DMSO-d_6, at 600 MHz.
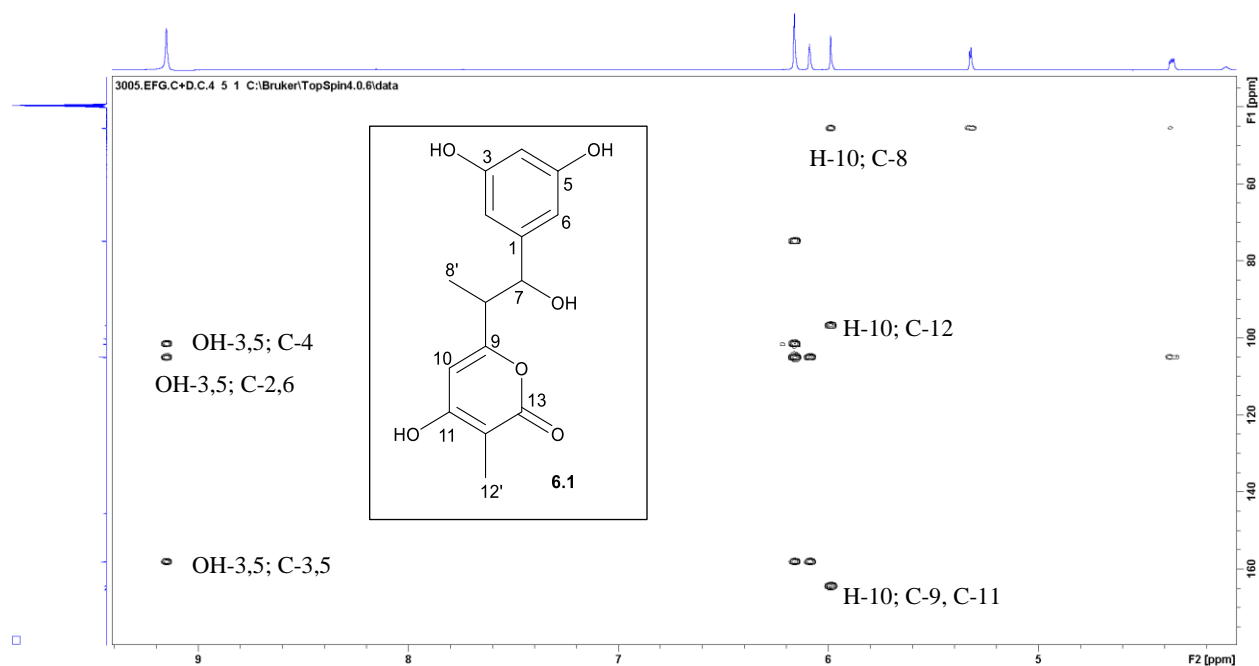
Figure 6-16. Expansion of ^1H-^{13}C HMBC spectrum of 6.1, showing select scalar couplings of protons into adjacent carbons. Recorded in DMSO-d_6, at 600 MHz.
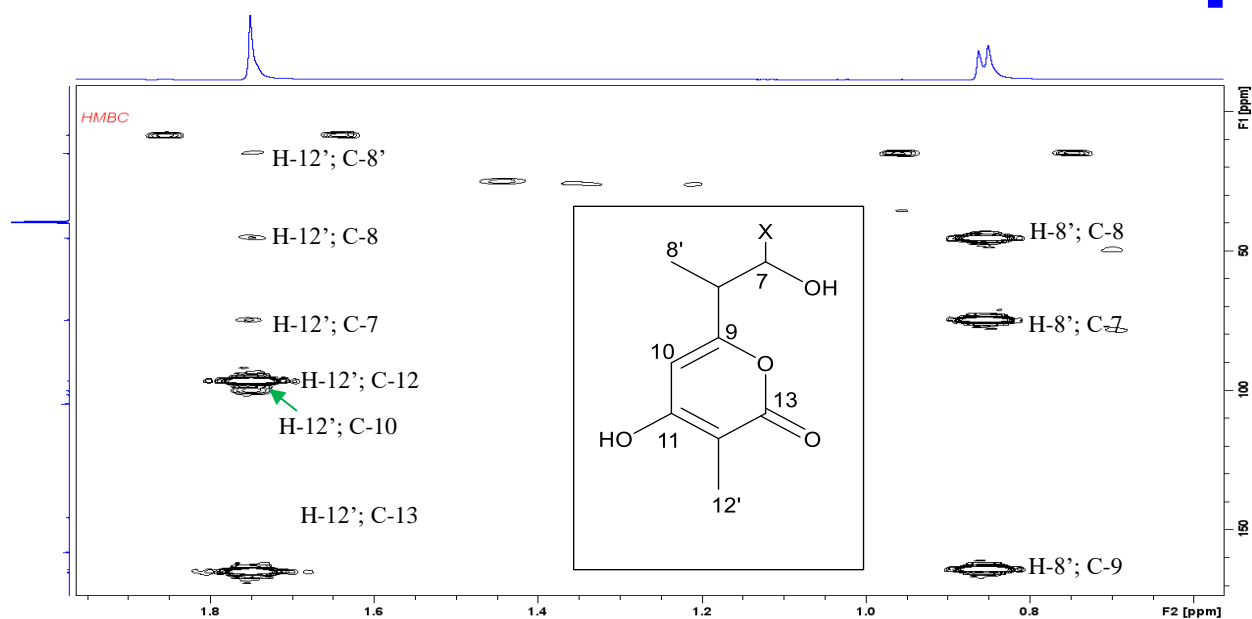
Figure 6-17. Expansion of ^1H-^{13}C HMBC spectrum of 6.1, showing select scalar couplings between the methyl singlet on the pyrone ring to the quaternary carbons as seen in Figure 6-18 below. Recorded in DMSO-d_6, at 600 MHz.

Figure 6-18. Key ¹H-¹³C HMBC and ¹H-¹H tROESY correlations connecting the two rings of 6.1.

Figure 6-19. Full ¹H-¹H tROESY spectrum of 6.1, recorded in DMSO-d_6, at 600 MHz.
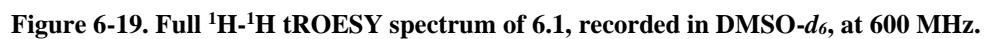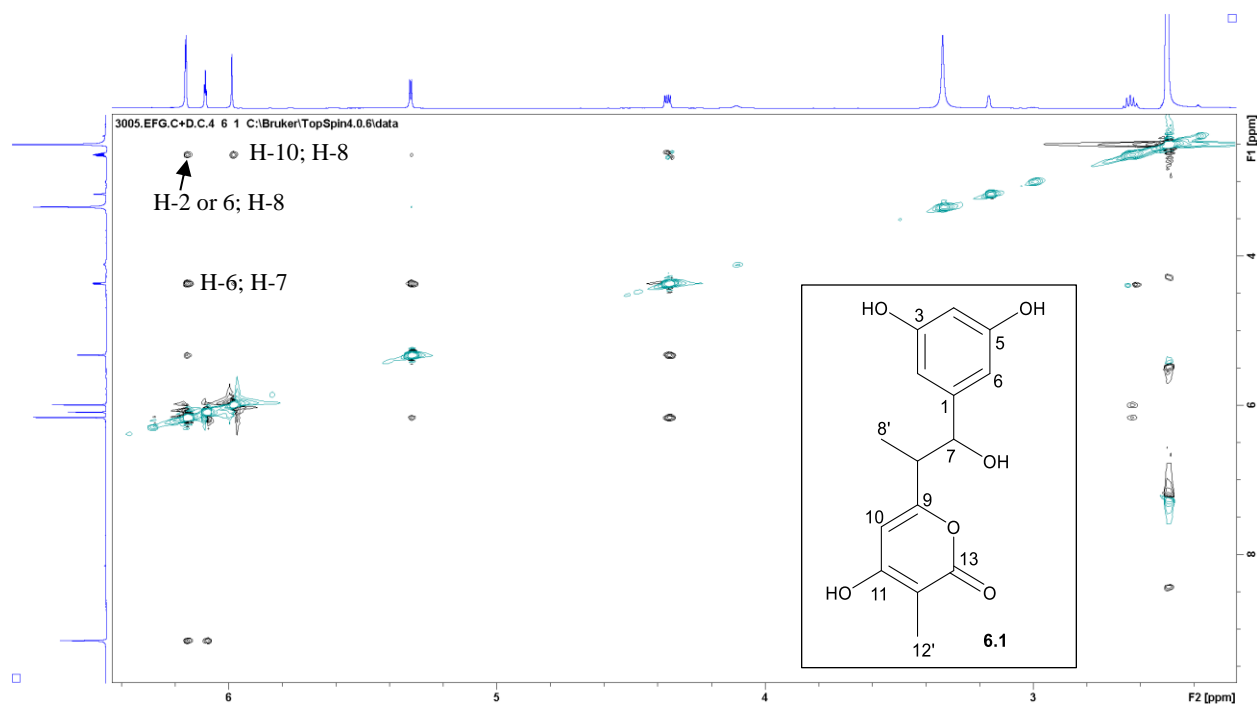
Figure 6-20. Expansion of the ^{1}H-^{1}H tROESY spectrum showing select through-space correlations that can be seen in Figure 6-18 and the diagram insert. Recorded in DMSO-d_6, at 600 MHz.
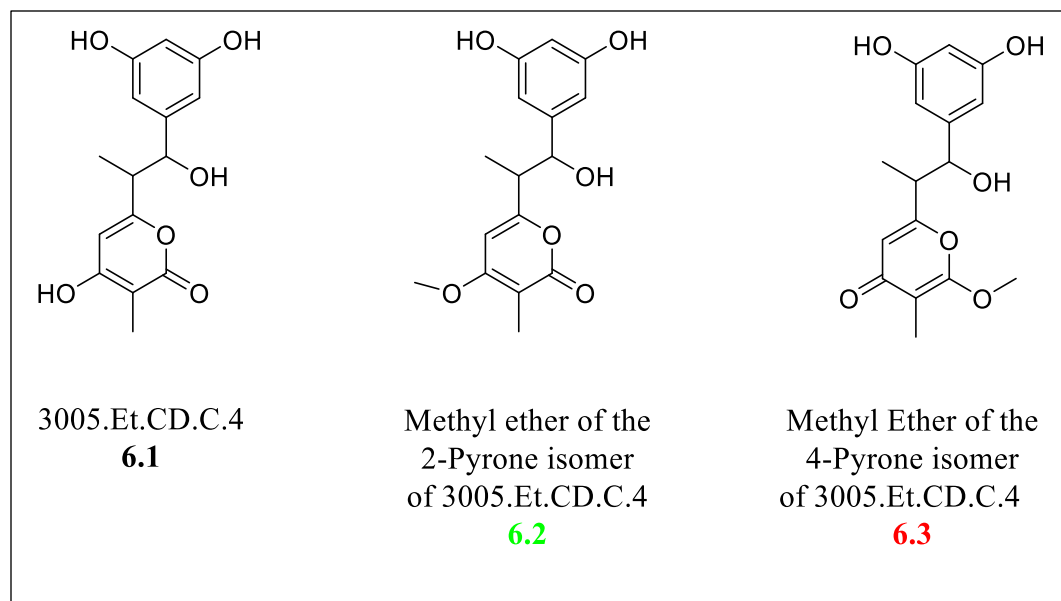
Figure 6-21. Structures of the natural product and methylation derivative which allowed separation of the pyrone tautomers.

6.3.1.1 Derivatization of 3,5-dihydroxyphenylpyrone (6.1)

In order to complete the methylation of the alcohol on the pyrone, a gentle method for methylation of natural products using trimethylsilyldiazomethane was employed (**Scheme 6-2**).[291] Upon completion of the methylation, the methyl ethers of the respective isomers were separated via reversed-phase HPLC (step-gradient: 0-60 min 17:83 MeCN:H_2O; 61mins to 120 mins 23:77 MeCN:H_2O). The earlier eluting isomer **6.3** was more abundant, while the later eluting isomer **6.2** was less abundant. Most chemical shifts were identical to the natural product **6.1,** with the major differences appearing in the pyrone ring.
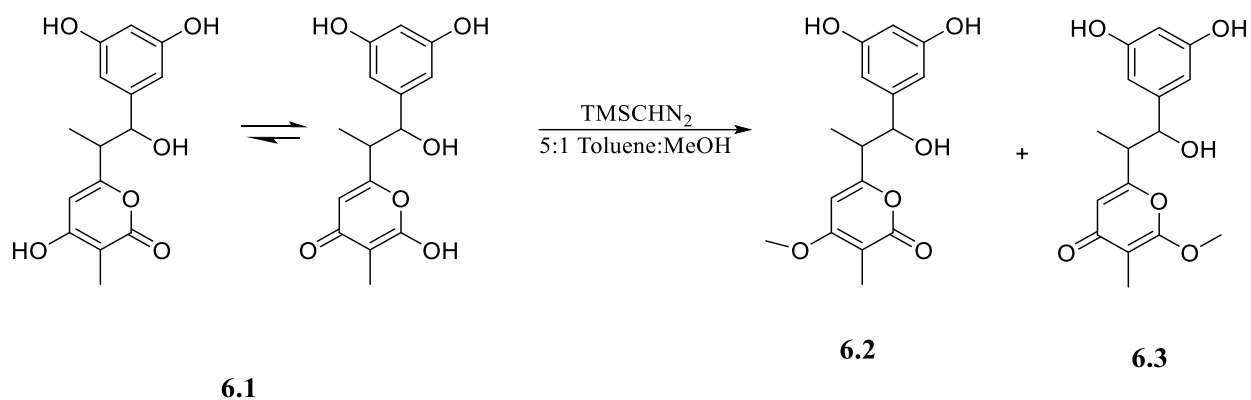
Scheme 6-2. Methylation of the tertiary alcohol at C-11 with trimethylsilyldiazomethane leading to separation of 6.1's tautomers.
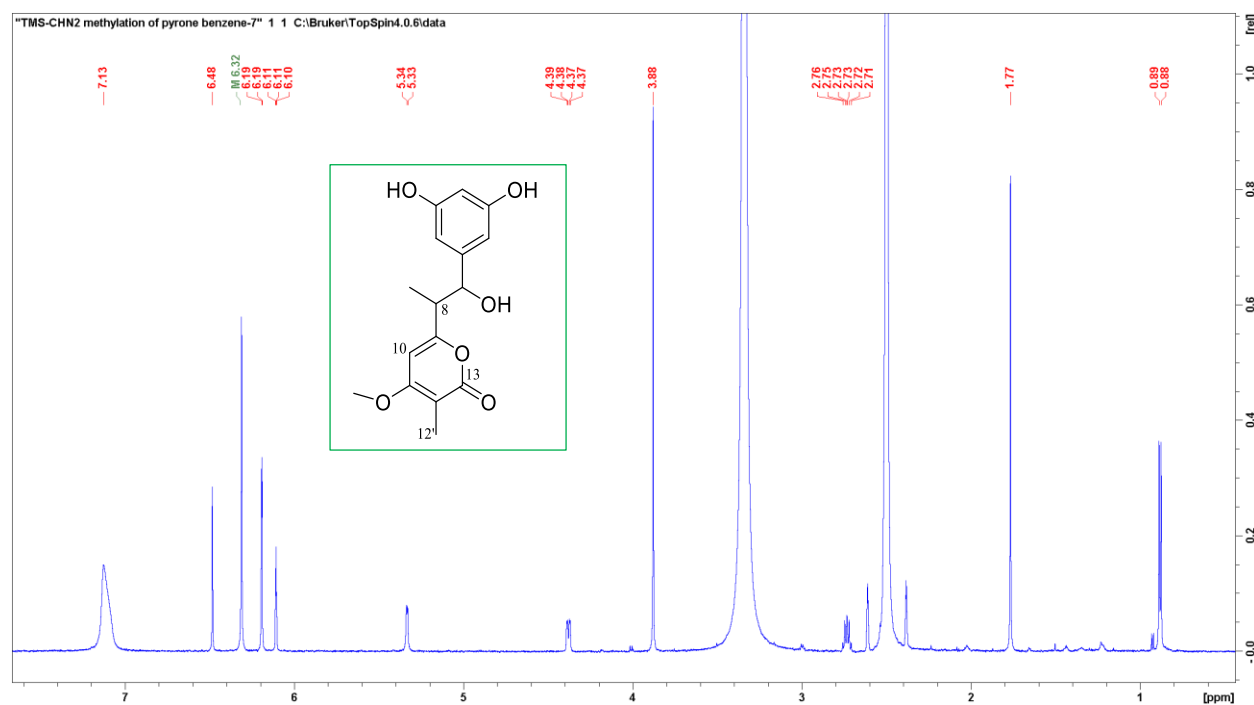
Figure 6-22. [1]H spectrum for methylation of the 2-pyrone isomer (6.2) of 3005.Et.D.C.4 (6.1). Recorded in DMSO-d_6, 600 MHz.
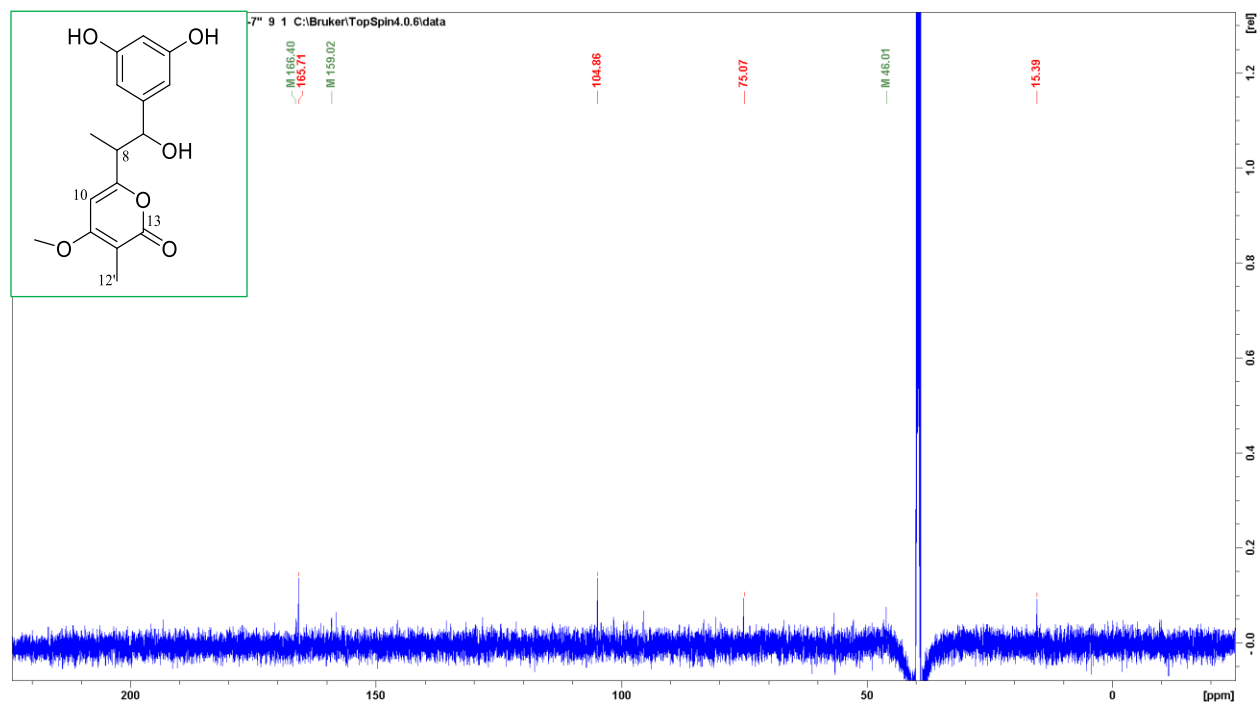
Figure 6-23. ^{13}C spectrum for methylation of the 2-pyrone isomer (6.2) of 3005.Et.D.C.4 (6.1). Recorded in DMSO-d_6, at 150 MHz.
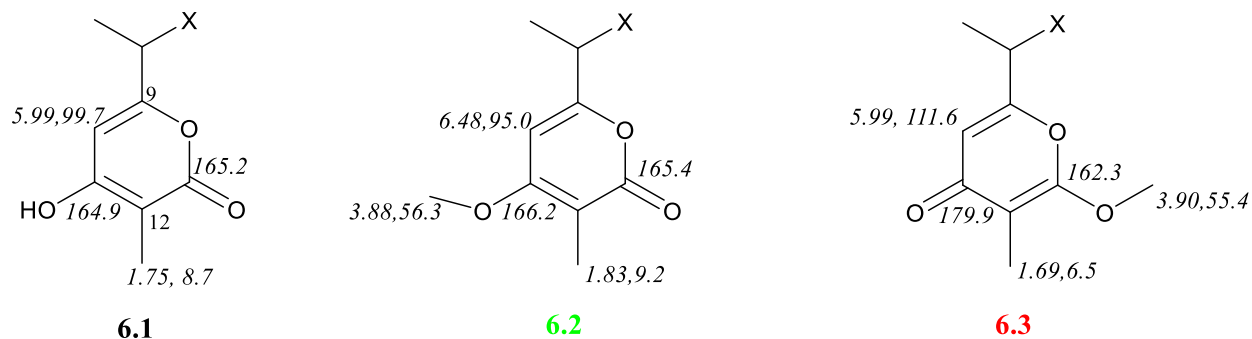
Figure 6-24. Chemical shifts for the natural product and its methylation derivatives demonstrating key changes.
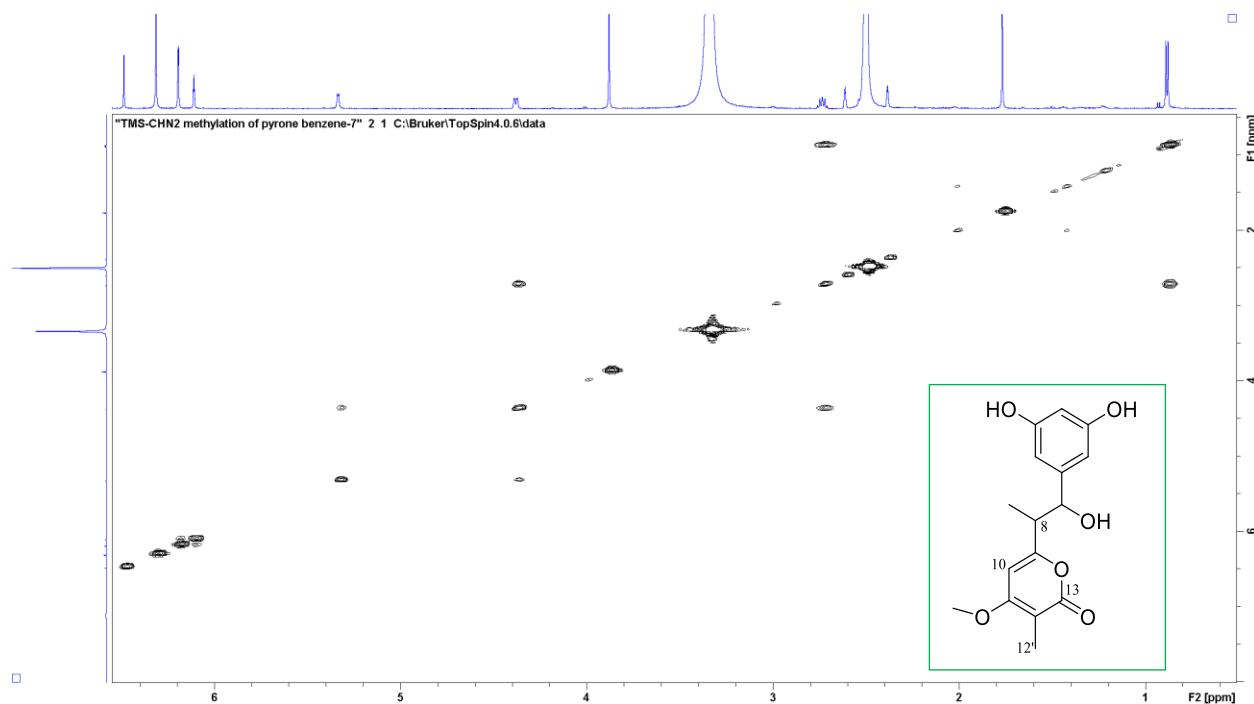
Figure 6-25. ^{1}H-^{1}HCOSY60 spectrum for methylation of the 2-pyrone isomer (6.2) of 3005.Et.D.C.4 (6.1). No changes were seen from the original natural product. Recorded in DMSO-d_6, at 600 MHz.

C-10: 95.0, 6:48
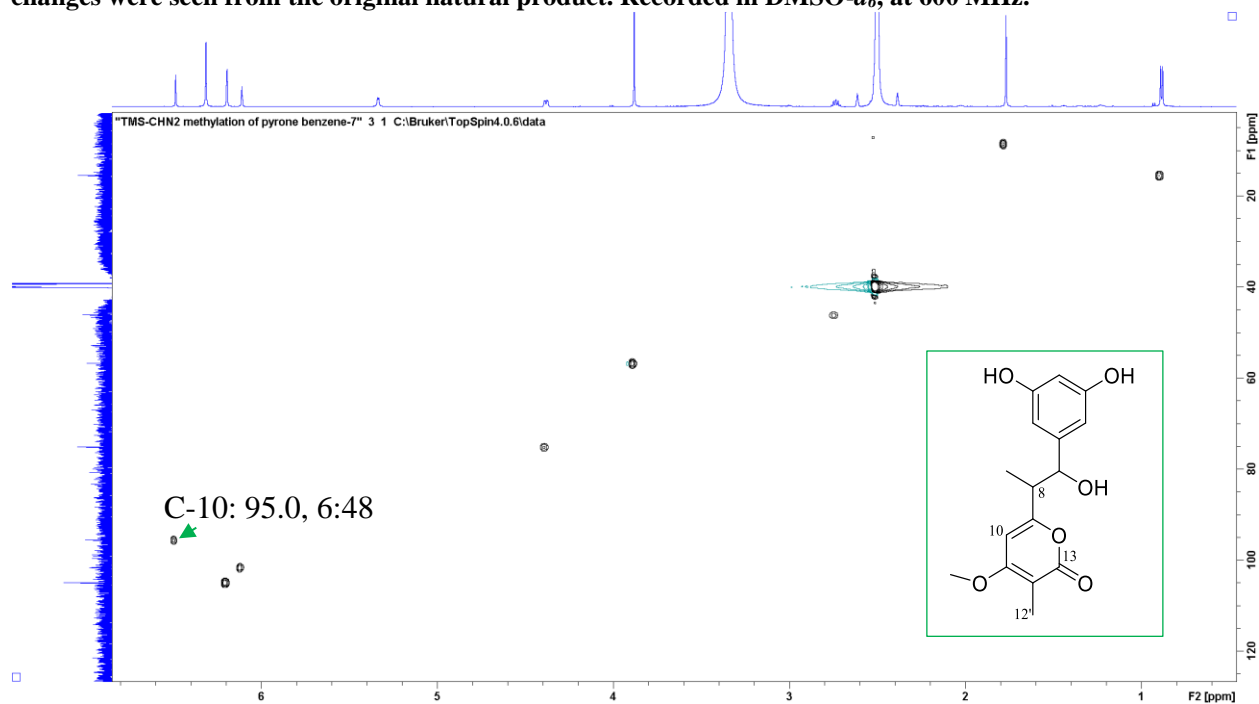
Figure 6-26. ^{1}H-^{13}C HSQC spectrum for methylation of the 2-pyrone isomer (6.2) of 3005.Et.D.C.4 (6.1), demonstrating a key change at the methine of C-10. Recorded in DMSO-d_6, at 600 MHz.
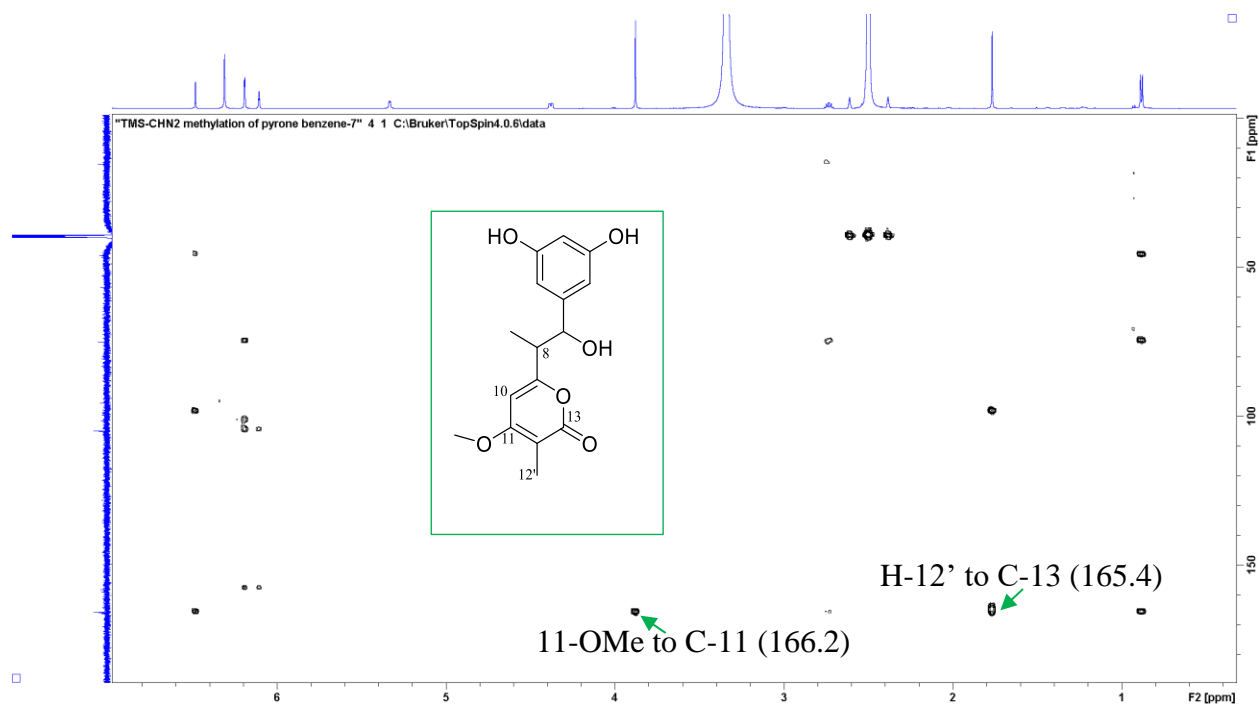
Figure 6-27. ^{1}H-^{13}C HMBC spectrum for methylation of the 2-pyrone isomer (**6.2**) of 3005.Et.D.C.4 (**6.1**) demonstrating changes in chemical shifts through the correlation from the methyl singlet at H-12' to the carbonyl at C-13. Recorded in DMSO-d_6, at 600 MHz.
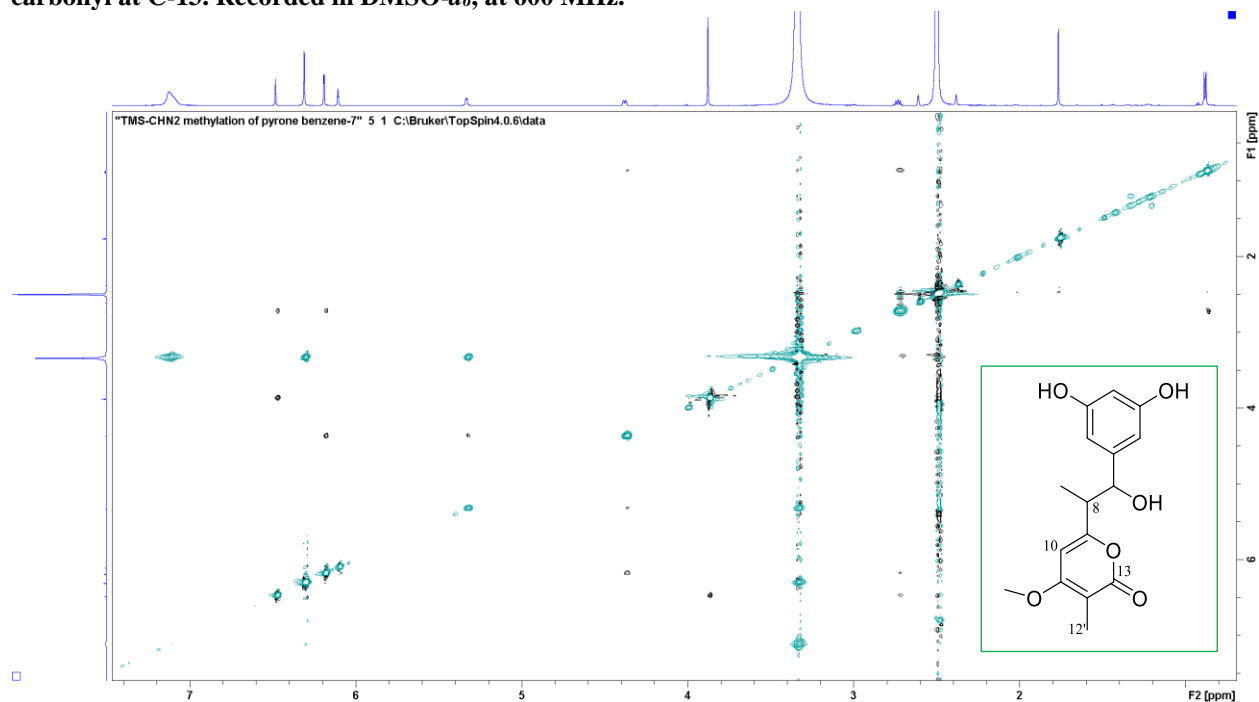
Figure 6-28. ^{1}H-^{1}H tROESY spectrum for methylation of the 2-pyrone isomer (**6.2**) of 3005.Et.D.C.4 (**6.1**). Recorded in DMSO-d_6, at 600 MHz.

Key differences between the 2-pyrone (**6.2**) and the 4-pyrone (**6.3**) isomers were seen at

C-11, C-13 and C-10 (**Figure 6-35**). In the 2-pyrone isomer (**6.2**) spectra, carbon shifts at δ_c

243

166.2 (C-11), δ_c 165.4 (C-13) and δ_c 95.0 (C-10) (**Figure 6-24 to 6-28**) displayed a closer match

to the dominant 2-pyrone form seen in the natural product [C-11 (δ_c 164.9), C-13 (δ_c 165.2) and

C-10 (δ_c 99.7)] (**Figure 6-24**). As such those spectra were assigned as the 2-pyrone form (**6.2**). In

the other set of spectra collected for the derivatized isomers, resonances at δ_c 179.9 (C-11), δ_c

162.3 (C-13), and δ_c111.6 (C-10) (**Figures 6-29 to 6-34**) were observed. The deshielded

resonance δ_c 179.9 (**Figure 6-33**) was assigned to the C-11 carbon in the 4-pyrone structure in

6.3 because it was a chemical shift more characteristic of a cyclic ketone adjacent to double

bonds.[292,293] Additionally, in **6.3** the C-10 resonance of δ_c 111.6 (**Figure 6-32**) matched an α

carbon of an α,β unsaturated ketone[293,294] whereas the resonance δ_c 95.0 of C-10 in **6.2** (**Figure**

6-26) is more characteristic of the β carbon of a conjugated lactone.[293,294]

Figure 6-29. Proton spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). Recorded in DMSO-*d₆*, at 600 MHz. There are large impurities gained while removing solvent traces on the lyophilizer. Given the small amount of material the impurities are relatively intense. The proton spectrum also shows chemical shift changes at H-10.

Figure 6-30. ^{13}C spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). Recorded in DMSO-d_6, at 150 MHz.
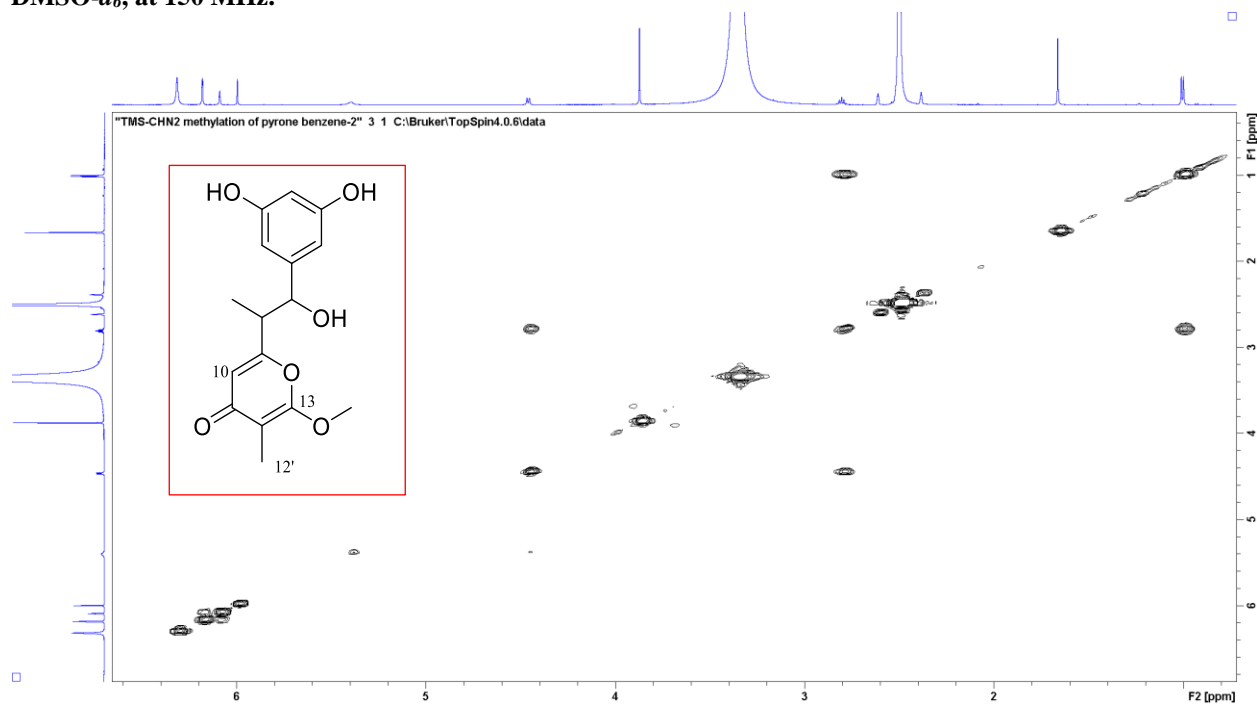
Figure 6-31. ^{1}H-^{1}H COSY60 spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). No changes are seen from the natural product. Recorded in DMSO-d_6, at 600 MHz.
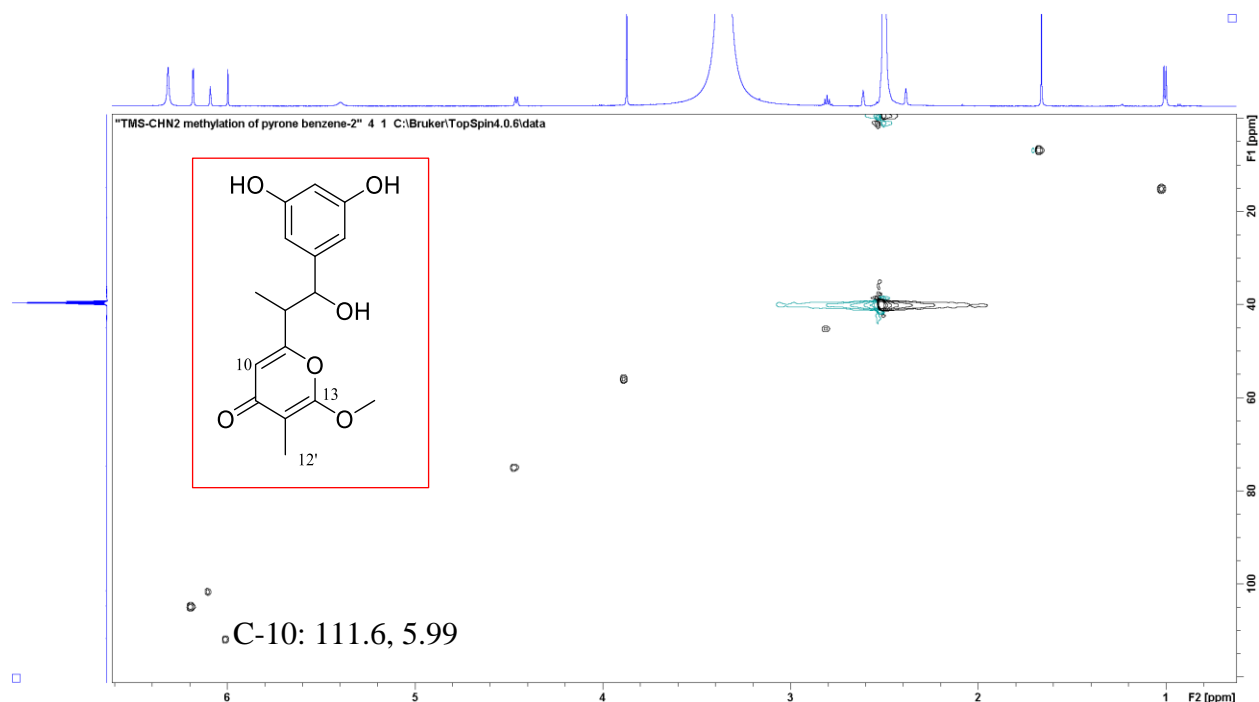
Figure 6-32. ^{1}H-^{13}C HSQC spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). Again C-10 shows chemical shift changes, not just from the natural product but also from the other methylated tautomer. C-10 is demonstrating a significant deshielding. Recorded in DMSO-d_6, at 600 MHz.
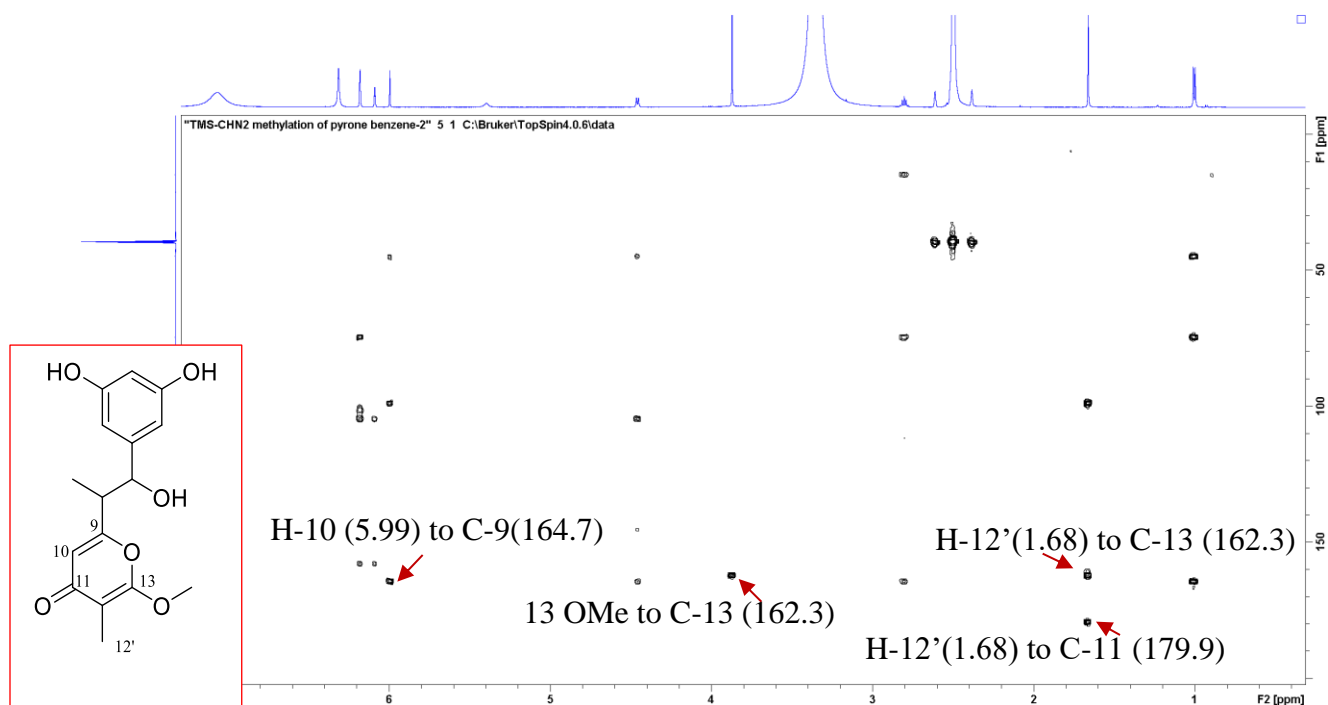
Figure 6-33. ^{1}H-^{13}C HMBC spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). The correlation from H-12' to C-13 allows the observation that C-13 carbon is significantly deshielded. Recorded in DMSO-d_6, at 600 MHz.
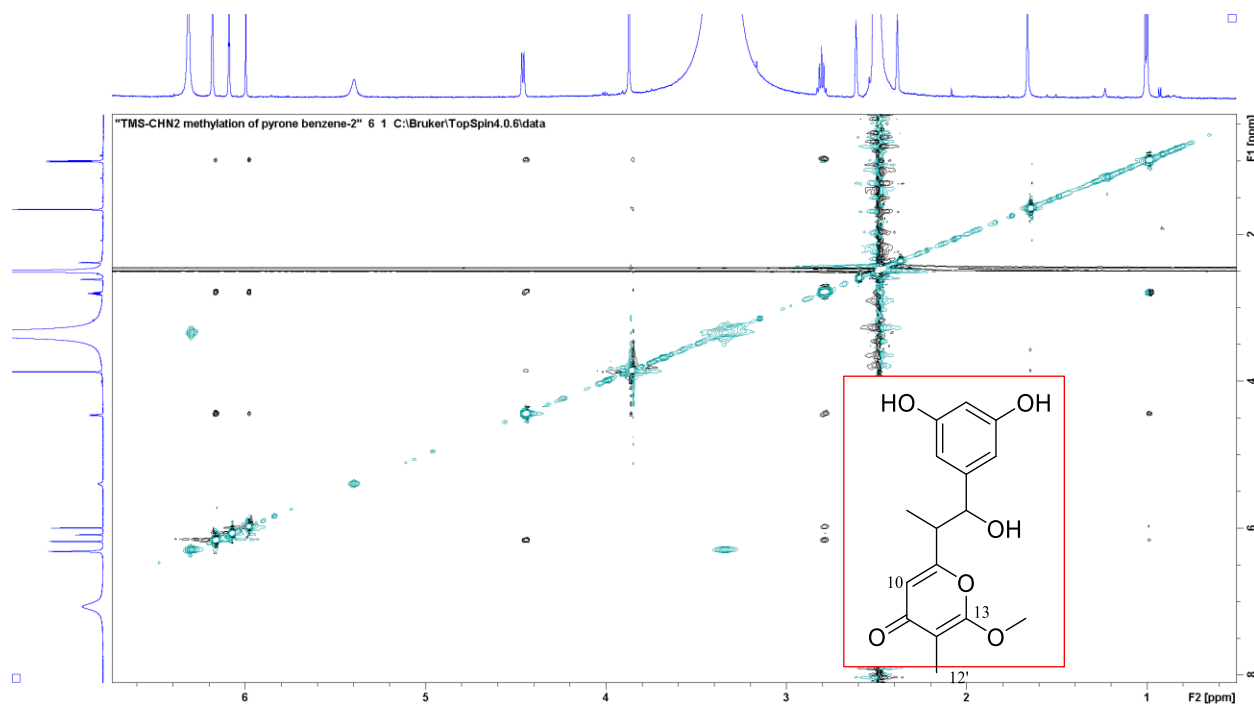
Figure 6-34. ^{1}H-^{1}H tROESY spectrum for methylation of the 4-pyrone form (6.3) of 3005.Et.D.C.4 (6.1). Recorded in DMSO-d_6, at 600 MHz.
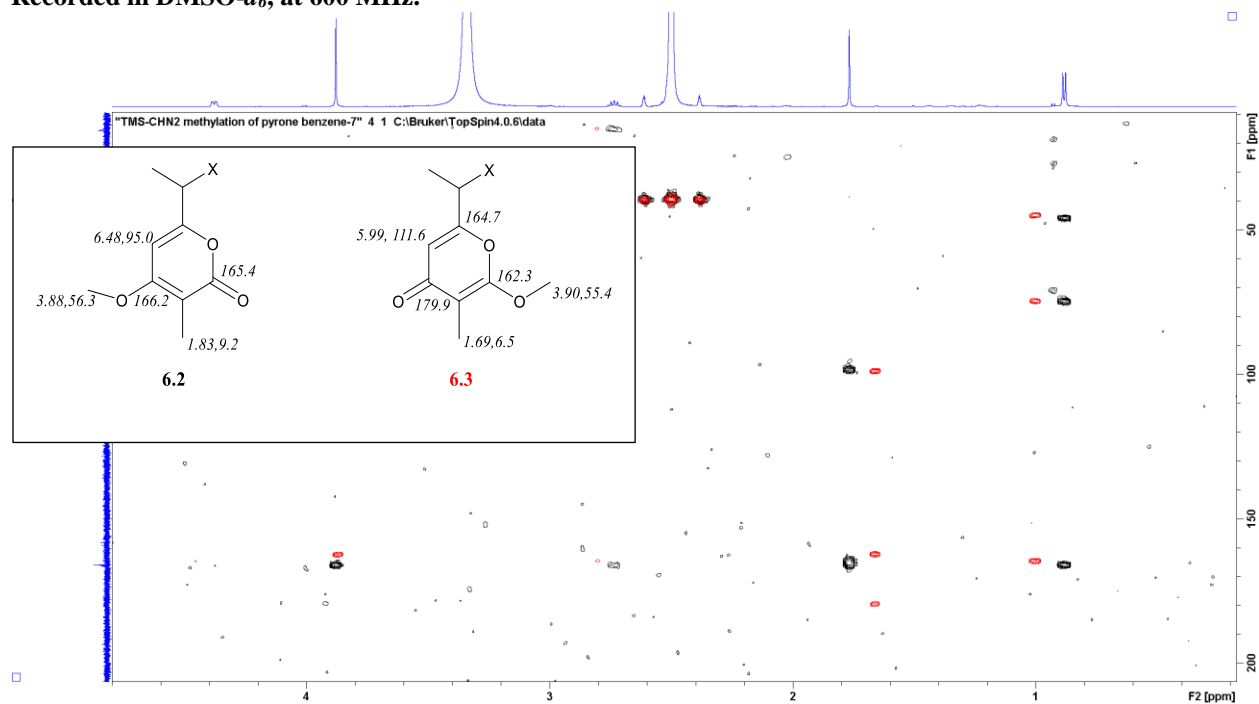
Figure 6-35. Overlay of the ^{1}H-^{13}C HMBC's of the 6.2 and 6.3. Black is the later eluting 2-pyrone (6.2), red is the earlier eluting 4-pyrone (6.3). The distinguishing δ_C 179.9 is visible in the red HMBC. Recorded in DMSO-d_6, at 600 MHz.

Position		DMSO-d_6					
		3,5-dihydroxy (**6.1**)		2-Pyrone Methyl Ether (**6.2**)		4-Pyrone Methyl Ether (**6.3**)	
		δ_C/δ_N	δ_H (*J* =Hz)	δ_C/δ_N	δ_H (*J* = Hz)	δ_C/δ_N	δ_H (*J* = Hz)
1	C	145.4		NV		145.2	
2	CH	104.8	6.16 d (2.2)	105.6	6.19 d (2.2)	105.9	6.17
3	C	157.9		158.6		158.6	
4	CH	101.5	6.09 t (2.2)	101.6	6.11 d (2.2)	101.6	6.09
5	C	157.9		158.6		158.6	
6	CH	104.8	6.16 d (2.2)	105.6	6.19 d (2.2)	105.9	6.17
7	CH	74.7	4.36 dd (4.0, 8.5)	73.6	4.40 dd (4.0, 8.5)	74.8	4.47 dd (4.0, 8.5)
7-OH	OH		5.32 d (4.0)		5.34, d (4.0)		5.40 d (4.0)
8	CH	45.4	2.64 q (~7.1)	45.8	2.78 q (~7.1)	46.6	2.81 q (~7.1)
8'	CH$_3$	14.9	0.85 d (7.1)	16.2	0.94 d (7.1)	14.9	1.02 d (7.1)
9	C	164.2		165.9		164.7	
10	CH	100.0	5.99 s	95.0	6.48 s	111.6	5.99 s
11	C	165.2		166.2		179.9	
11-OH/OMe	OH		NV	56.3	3.88 s		
12	C	96.6		98.3		100.4	
12'	CH$_3$	8.5	1.75 s	9.2	1.83 s	6.5	1.69 s
13	C	165.2		165.4		162.3	
13-OH/OMe	OMe					55.4	3.90 s
3 & 5 - OH	OH		9.15				

Table 6-2. NMR assignments for 6.1, 6.2 and 6.3. Recorded in DMSO-d_6 at 600 Mhz.

6.3.1.2   Assignment of Configuration

C-7 and C-8 are stereogenic carbons that required assignment because **6.1**, whose shifts and coupling constants were close to identical to those seen in the engineered compounds,[277] had not previously been reported as produced from a wild-type bacteria. Given the secondary alcohol

present at C-7, a Mosher's ester protocol was employed. The acid chloride reagents generally react with secondary alcohols to give the Mosher's esters.[295] Given the ability of the natural product to tautomerize and the desire to keep the subsequent NMR analysis as simple as possible, the methyl ether **6.3** was used to perform the Mosher's ester analysis. Since about 1 mg of **6.3** was available, an NMR tube-based Mosher's ester procedure was undertaken. This method completes the Mosher's ester reaction in the NMR tube without purification.[296] Two separate reaction tubes were set-up, one each for the *R* or *S* α-methoxy-α-trifluormethylphenylacetic acid (MPTA), with deuterated pyridine-d_5 employed as reaction and NMR solvent (**Figure 6-36**).[296] As **6.3** was residing in an NMR tube from previous data analysis, careful transfer was utilized to split **6.3** into two tubes. One tube was treated with the *S*-MTPA chloride, and the other tube was treated with the R-MTPA chloride (**Figure 6-36**). The tubes were shaken and then monitored by ^{1}H NMR until it appeared the reaction was complete at 4 hrs. Although the MTPA chlorides and acids were in large excess and swamped the proton spectrum, protons at H-2, H-4, H-6, H-7, H-8 and H-8' had correlations into distinct carbons that were not present in the excess reaction compounds (**Figures 6-37 to 6-42**). Aiding the assignment was the distinct coupling patterns at H-7 (doublet of doublets) and H-8 (quintet) (**Figures 6-38 and 6-41**). With this prior information and HSQC and gCOSY60 data, these protons were readily distinguishable, were assigned, and $\Delta\delta_{S-R}$ (**Table 6.3**) calculated. As seen in **Table 6.3**, protons at H-2, H-4, and H-6 and H-8 demonstrated positive changes in their chemical shifts when comparing the *S*-ester to the *R*-ester. This occurs when they are in the same plane as the phenyl moiety of Mosher's ester in the *R*-ester and experience anisotropic shielding (**Figure 6.43**). Conversely, the methyl proton at H-8' and the methine proton at H-10 experience negative changes when comparing the *S*-ester to the *R*-ester. This occurs when they are in the same plane as the phenyl moiety in the *S*-ester (**Figure**

6.43). These results directly support that C-7 is in the *S* configuration, which aligns with previous work done on configuration of the engineered compounds.[277] In the Mosher's ester analysis completed here, C-8' presented a strong negative difference between the *S*-ester and the *R*-ester (**Table 6-3**), suggesting an *R* configuration at C-8 (**Figure 6.43**).[295] Although these results can only provide possible support for the configuration at the C-8 stereocenter, in the 2015 paper examining the configuration of these compounds, they use computational evidence to assign the equivalent of C-8 in **6.3** as being in the *R* configuration. The chemical shifts at C-7 (H-7) and C-8 (H-8), as well as the coupling constants between H-7 and H-8, are nearly identical between the engineered compound and the natural product. The likelihood of the C-8 of **6.1** being in the *R* configuration is therefore strong and is represented as such here.
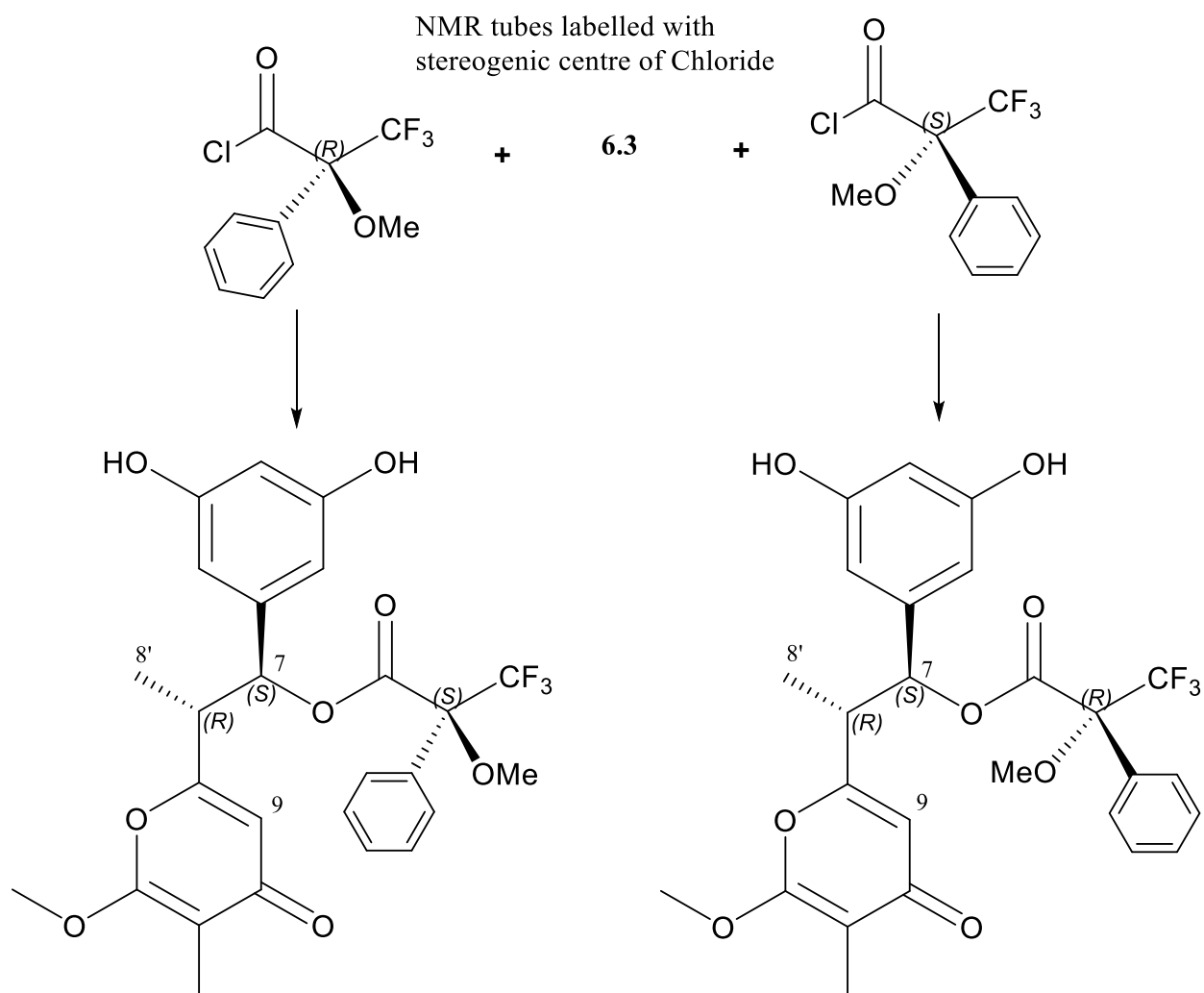
Figure 6-36. Mosher's ester reaction with 6.3, a derivative of 6.1. Two separate NMR tubes containing small amounts of 6.3 were each reacted with either the *S*-Mosher's chloride or the *R*-Mosher's chloride. Due to the priority rules, the chloride and the esters have the opposite configuration.
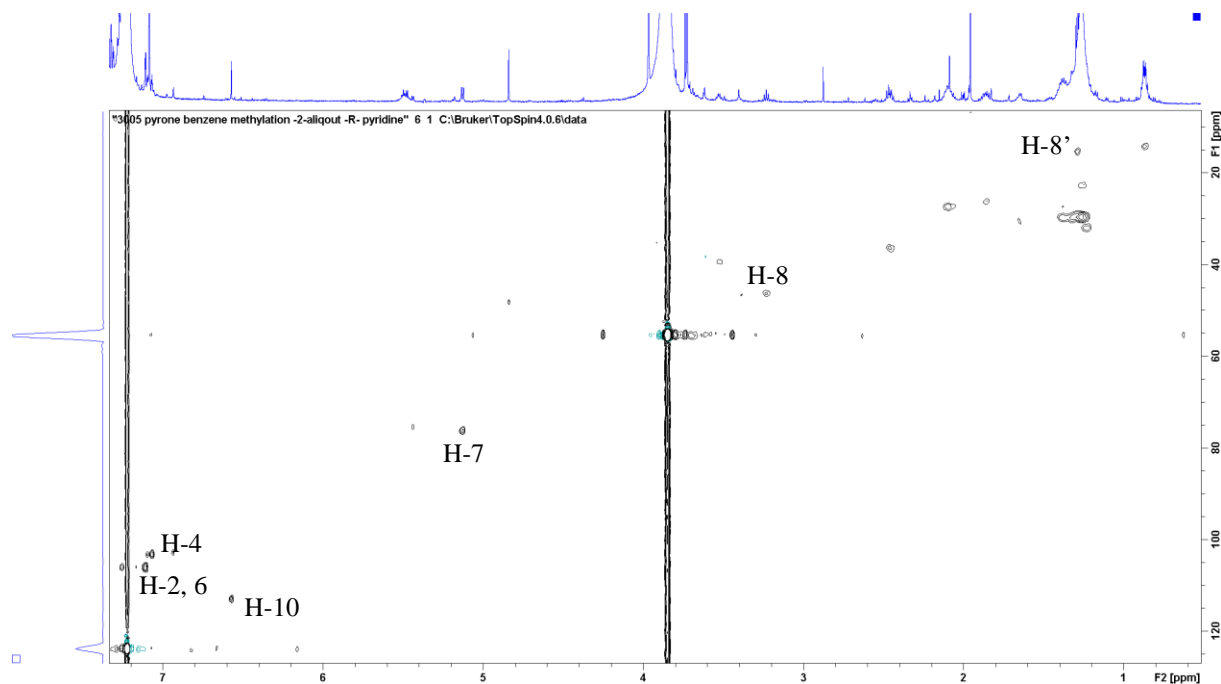
Figure 6-37. ^{1}H-^{13}C HSQC spectrum for *S*-Mosher's Ester(made from *R*-chloride) of 6.3. Recorded in pyridine-*d₅, at* 600 MHz.
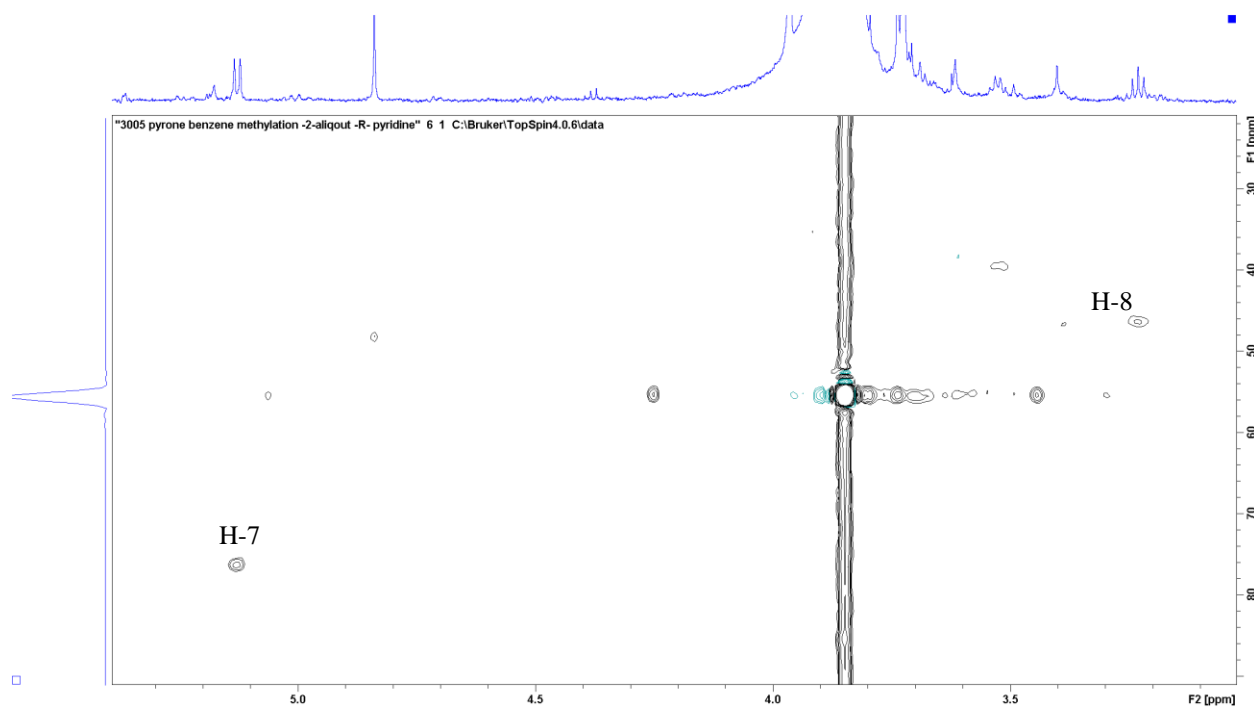
Figure 6-38. ^{1}H-^{13}C Expansion of HSQC spectrum for *S*-Mosher's Ester (made from *R*-chloride) of 6.3. Recorded in pyridine-*d₅,* at 600 MHz.
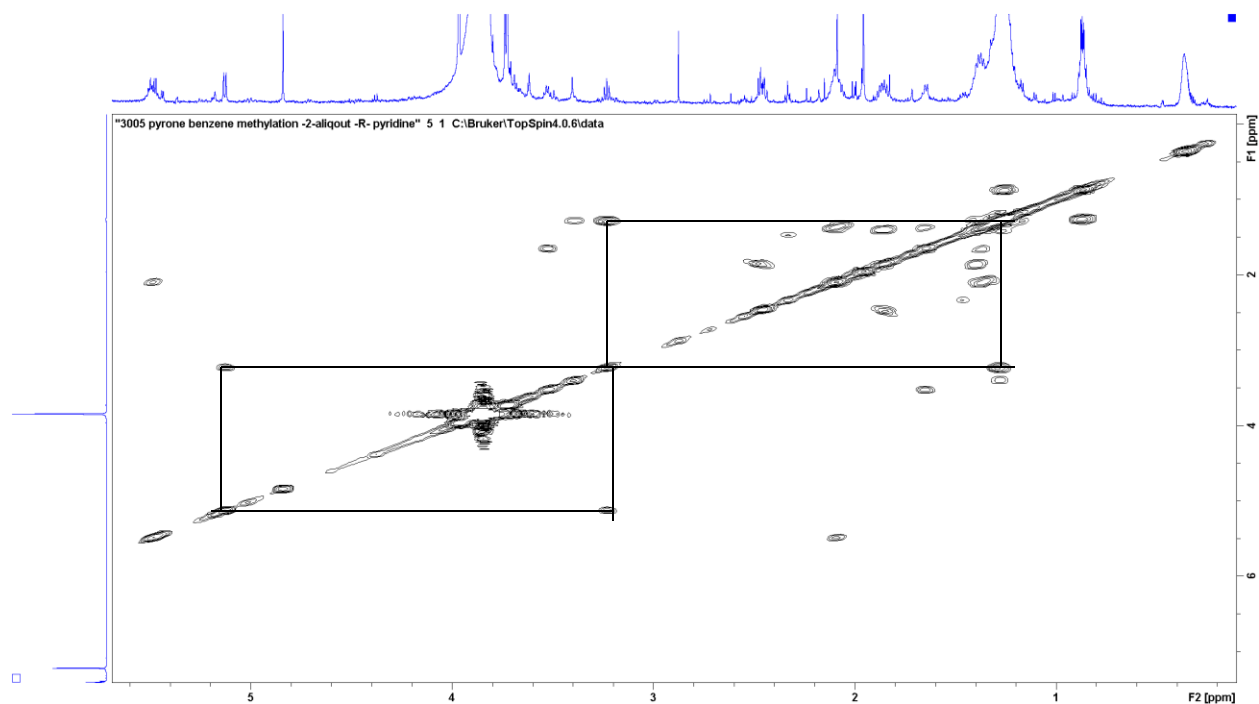
Figure 6-39. ^{1}H-^{1}H Expansion of COSY60 spectrum for *S*-Mosher's Ester (made from *R*-chloride) of 6.3. Connecting squares demonstrating key COSY correlations used to confirm chemical shift assignment. Recorded in pyridine-*d₅,* at 600 MHz.

Position		4-Pyrone Methyl Ether (6.3) (DMSO-d_6)		Measured chemical shift changes to 6.3 under Mosher's Ester (pyridine-d_5)		Calculated difference
				R-Chloride	S-Chloride	
Position		δ_C/δ_N	δ_H	S-MPTA	R-MPTA	$\Delta\delta S$-R
1	C	145.2				
2	CH	105.9	6.17	7.109	7.103	+0.006
3	C	158.6				
4	CH	101.6	6.09	7.068	7.066	+0.002
5	C	158.6				
6	CH	105.9	6.17	7.109	7.103	+0.006
7	CH	74.8	4.47	5.1276	5.1287	-0.001
7-OH	OH		5.40	Ester	Ester	
8	CH	46.6	2.81	3.233	3.231	+0.002
8'	CH$_3$	14.9	1.02	1.2893	1.3007	-0.011
9	C	164.1				
10	CH	111.6	5.99	6.5714	6.5786	-0.0072
11	C	164.7	No protons below this are shown as they were too complex to determine			

Table 6-3. Results from Mosher's ester NMR Analysis. ^{13}C shifts in DMSO-d_6 did not change in pyridine-d_5 as expected since carbon chemical shifts are not as influenced by solvent effects. ^{1}H shifts are reported as seen in pyridine-d_5.
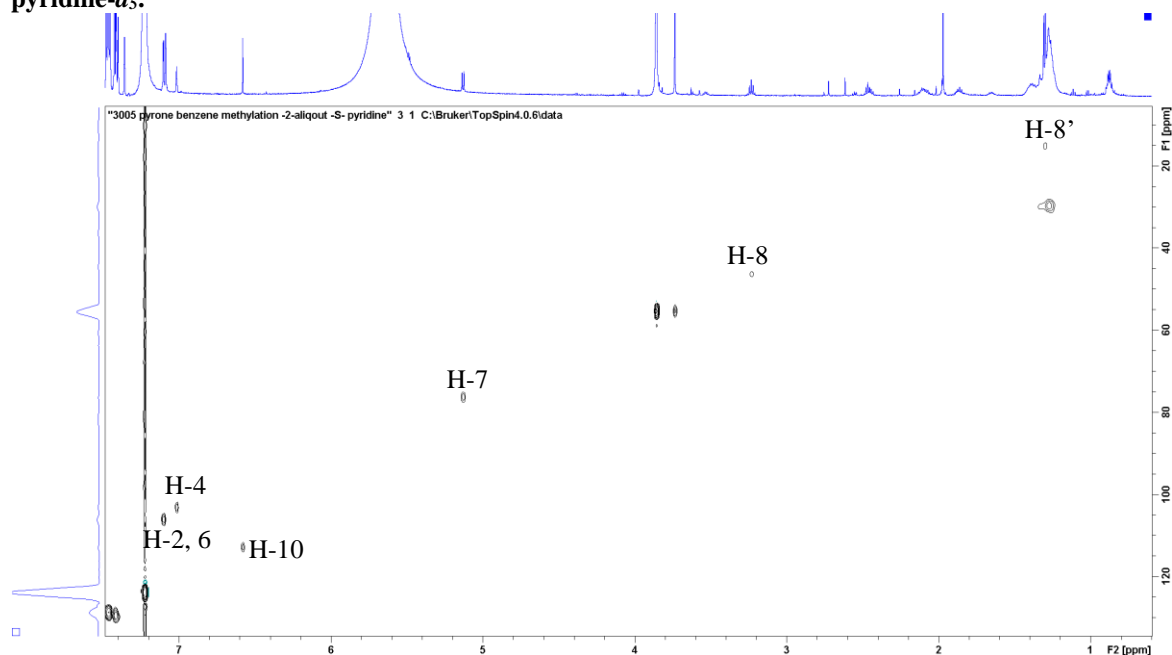
Figure 6-40. ^{1}H-^{13}C HSQC spectrum for R-Mosher's ester (made from S-chloride) of 6.3. Recorded in pyridine-d_5, at 600 MHz.
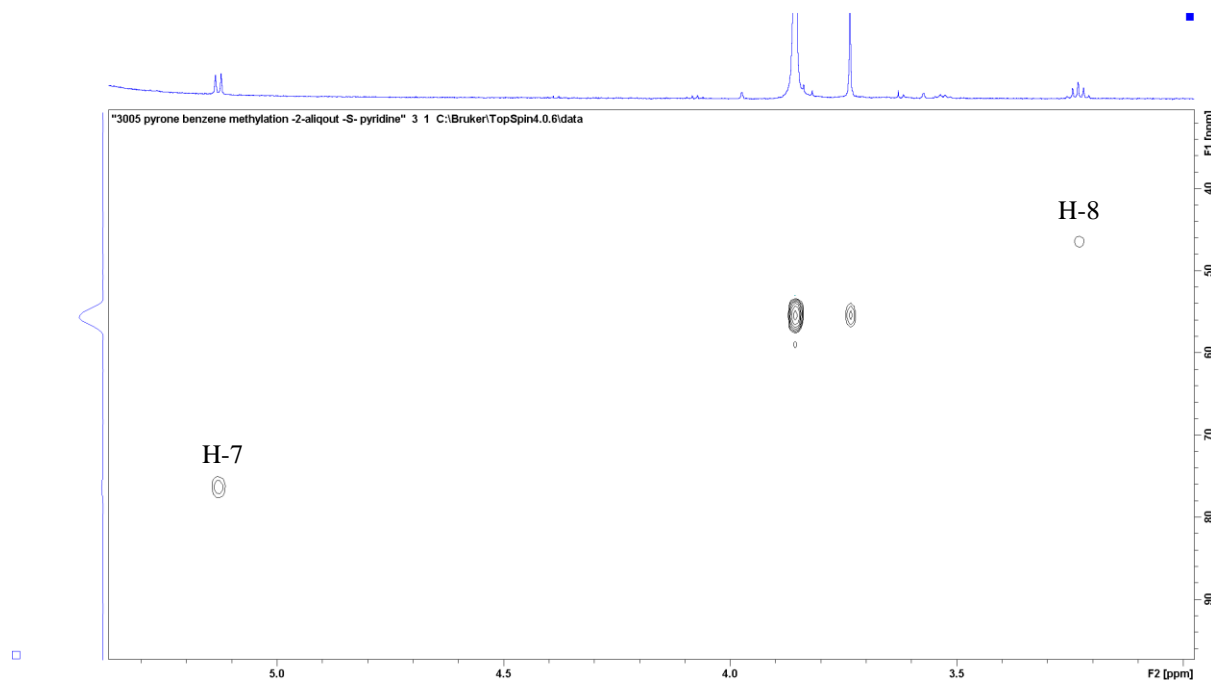
Figure 6-41. Expansion of ¹H-¹³C HSQC spectrum for *R*-Mosher's ester (made from *S*-chloride) of 6.3. Recorded in pyridine-*d₅,* at 600 MHz.
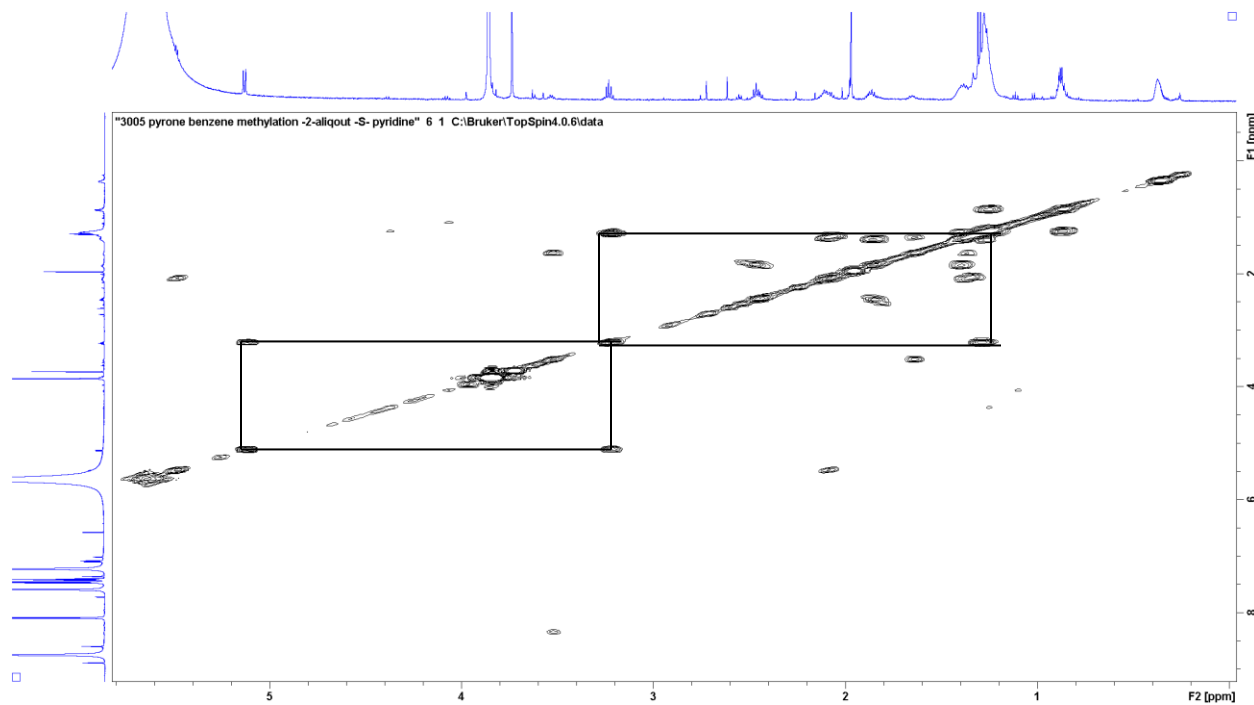
Figure 6-42. Expansion of ¹H-¹H COSY60 spectrum for *R*-Mosher's ester (made from *S*-chloride) of 6.3. Recorded in pyridine-*d₅,* at 600 MHz.

255
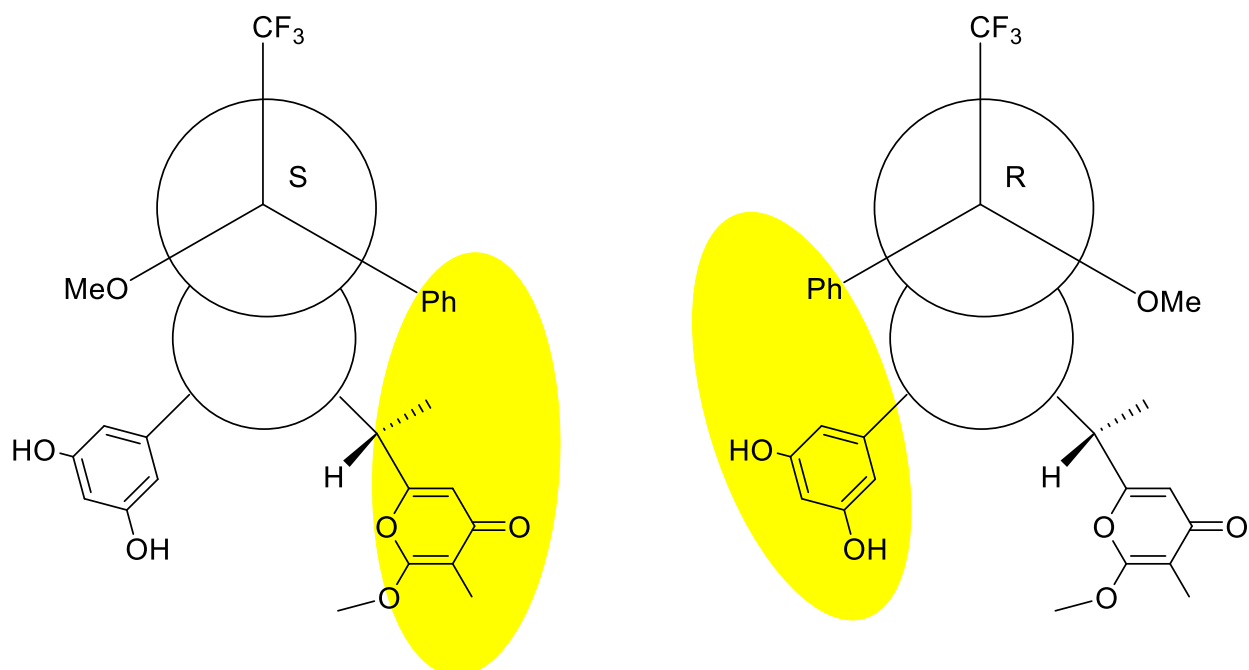
Figure 6-43. Newman projections of the Mosher's esters demonstrating in yellow the areas of the molecule 6.1 influenced by the anisotropic field of the phenyl moiety of the ester.

6.3.2 Structure elucidation of 3-acetamide-5-hydroxyphenylpyrone (6.4)

3-acetamide-5-hydroxyphenylpyrone (**6.4**) gave a $[M + H]^+$ ion in the HRESIMS at *m/z* 334.1297 appropriate for a molecular formula of $C_{17}H_{19}O_6N$ that requires 9 sites of unsaturation. Analysis of the ^1H, ^{13}C, ^1H-^1H gCOSY60, ^1H-^{13}C gHSQC, ^1H-^{13}C gHMBC and ^1H-^1H tROESY NMR data recorded in DMSO-d_6 (See **Table 6.3**) revealed **6.4** as a natural product that contains a disubstituted benzene ring attached to a pyrone functionality initially thought to be of complete PKS origin. The ^1H NMR spectrum of **6.4** was remarkably similar to **6.1** save for distinctive aromatic peaks (δ_H 6.89, 7.14 and 6.39) where the chemical shifts showed greater deshielding more typical of the standard aromatic region and with three peaks of equal intensity. Furthermore, the gHSQC (**Figure 6.47**) demonstrated that the three aromatic protons were coupled to varying carbons (δ_H 6.89, δ_C 108.5; δ_H 7.14, δ_C 106.1; and δ_H 6.39, δ_C 109.3). The HMBC spectrum demonstrated that the quaternary carbons at C-3 (δ_C 168.3) and C-5 (δ_C 157.0)

were also no longer identical. Furthermore, the ^{1}H-^{15}N HSQC spectrum (**Figure 6.52**) demonstrated an N-H proton at δ_H 9.80, which was demonstrating weak HMBC correlations into a carbonyl at δ_C 140.3 (**Figures 6.50 and 6.51**). A new methyl singlet at δ_H 2.04 (δ_C 24.2) was also showing HMBC correlations into the quaternary carbon at C-3. These pieces of data led to the assignment of an acetamide functional group as the second substituent of the benzene ring. The rest of the data showed either extreme similarity or identical chemical shifts to **6.1** and was assigned as such. The molecule **6.4** has previously been shown to be a product of a rifamycin producer *Amycolatopsis mediterranei* S699 that a) underwent extensive mutation to eliminate a 21kb DNA fragment[290] of the rifamycin gene cluster called the post-PKS modification genes,[278] or b) was a *rifF* (an acetyl-CoA acetyltransferase) deletion mutant.[279] However, to the best of my knowledge, **6.4** has not been previously isolated as a natural product from a wild-type bacteria.
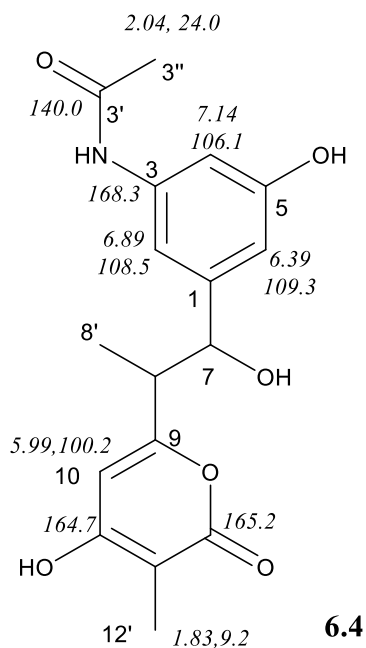
Figure 6-44. Numbering and structure for 3-acetamide-5-hydroxyphenylpyrone (6.4).
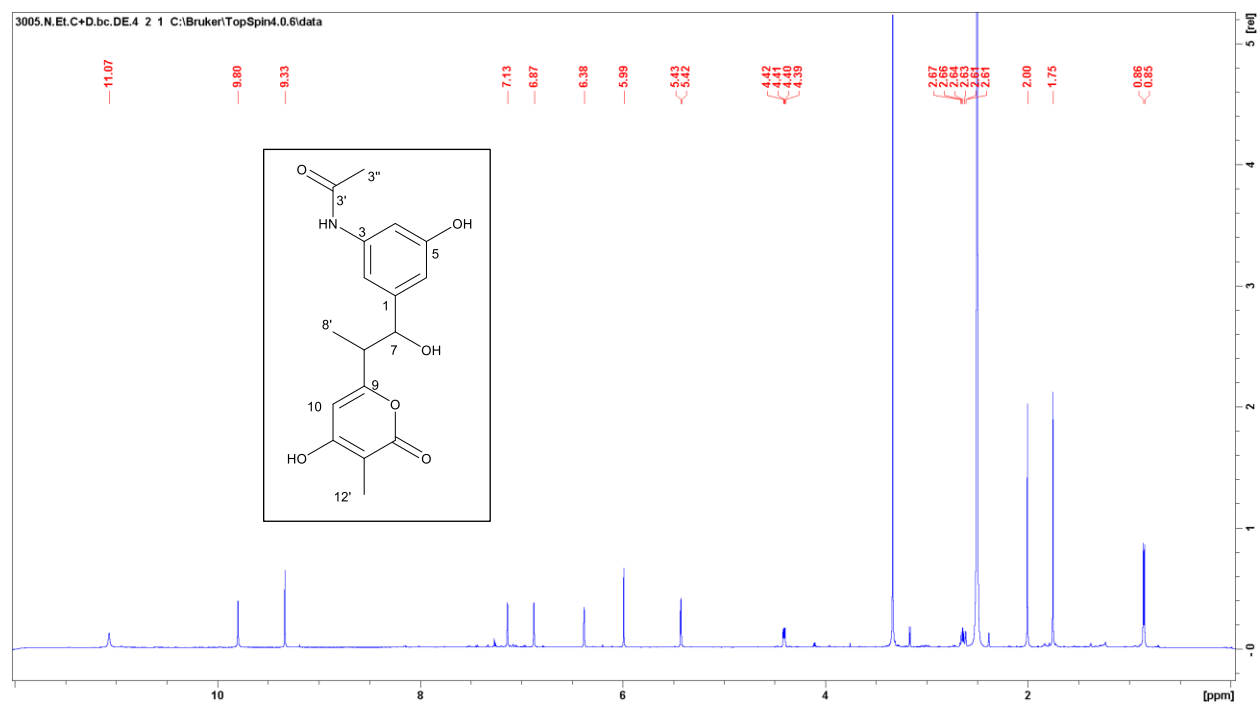
Figure 6-45. ¹H spectrum of 6.4, recorded in DMSO-d_6, at 600 MHz.
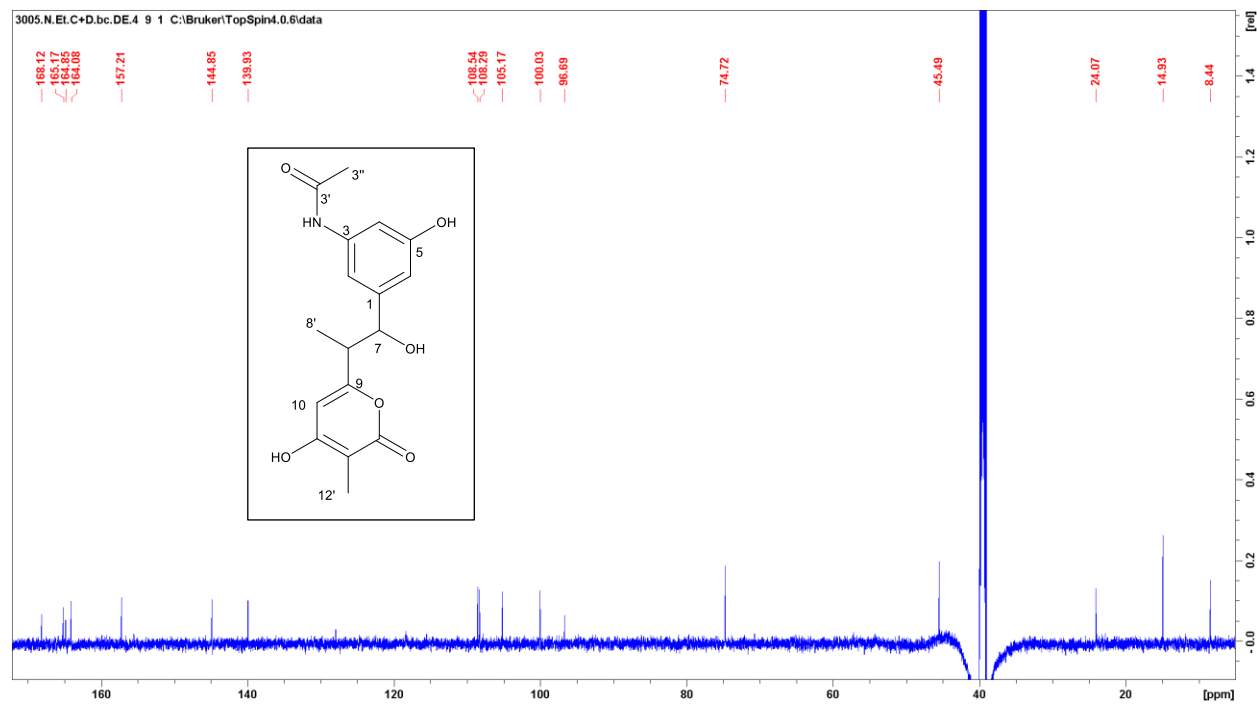
Figure 6-46. ¹³C spectrum of 6.4, recorded in DMSO-d_6, at 150 MHz.

258
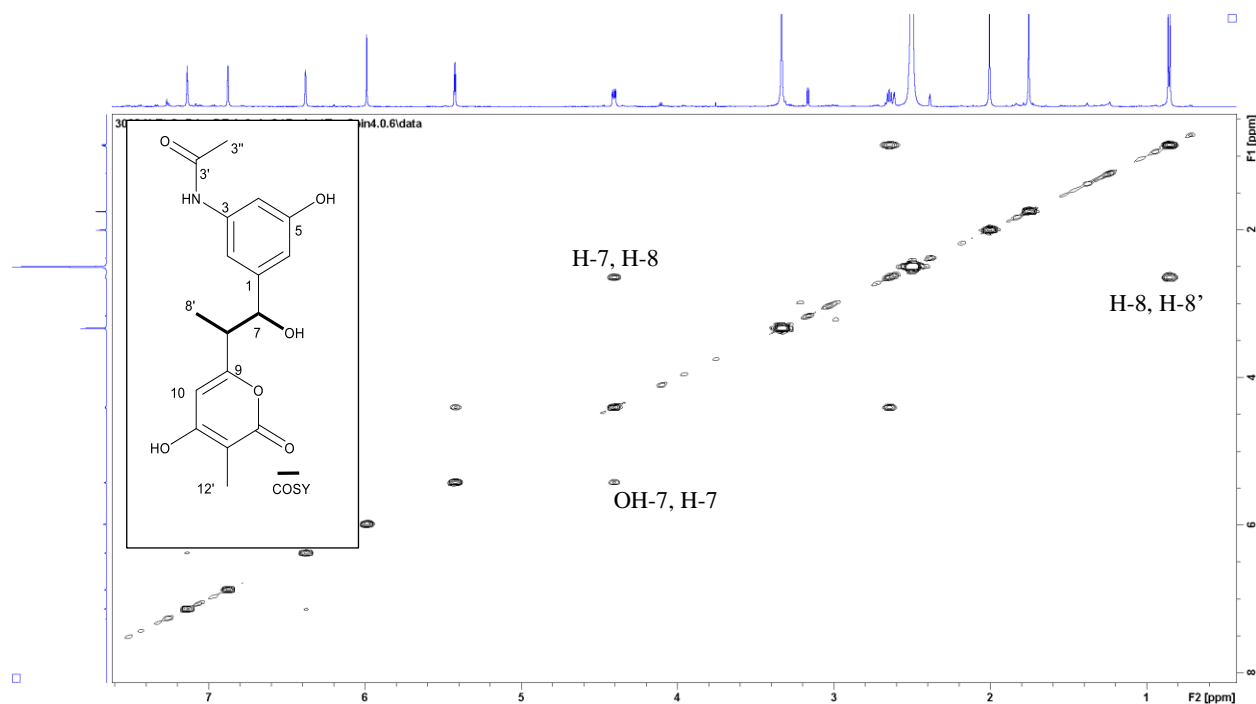
Figure 6-47. ^{1}H-^{1}H COSY60 spectrum of 6.4 demonstrating similar to identical correlations as 6.1 recorded in DMSO-d_6, at 600 MHz.
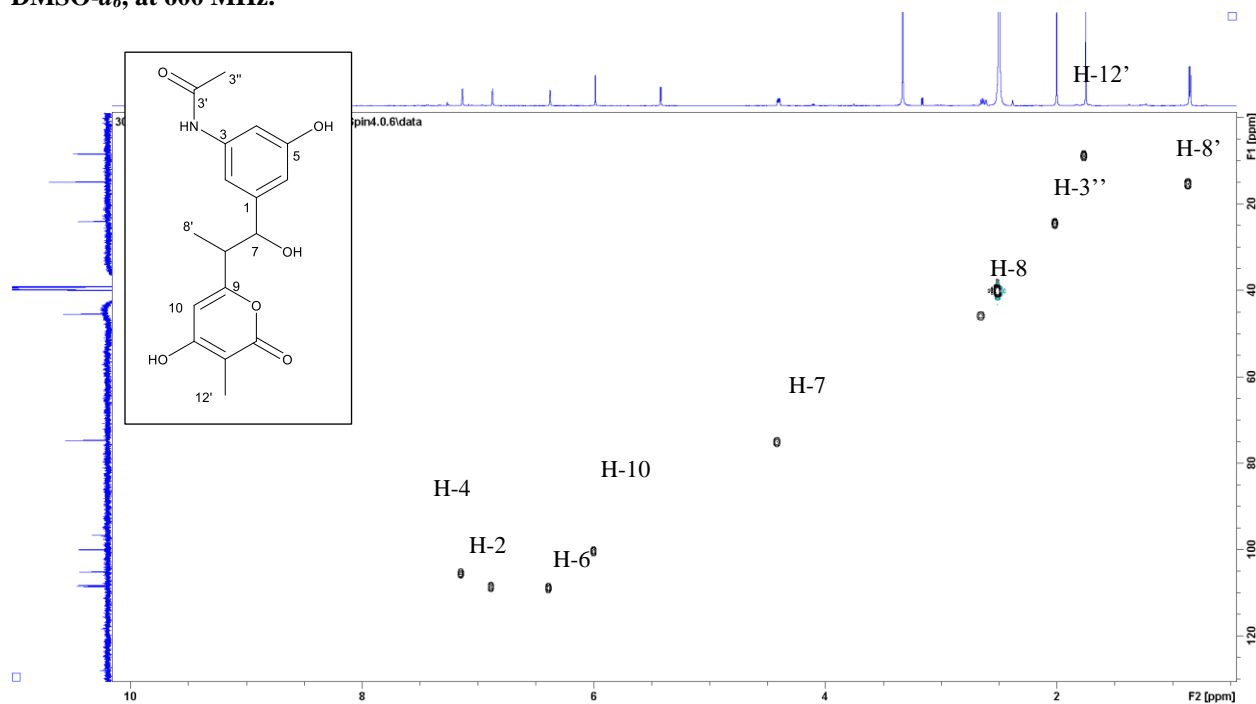
Figure 6-48. ^{1}H-^{13}C HSQC spectrum of 6.4 demonstrating similar carbon-proton correlations from H-7 to H-12' as 6.1 but unique carbon-proton correlations for aromatic H-2, H-4 and H-6, recorded in DMSO-d_6, at 600 MHz.

259

Figure 6-49. Full ^{1}H-^{13}C HMBC spectrum of 6.4, recorded in DMSO-d_6, at 600 MHz.

Figure 6-50. Key ^{1}H-^{13}C HMBC correlations for the ring systems of 6.4.

260
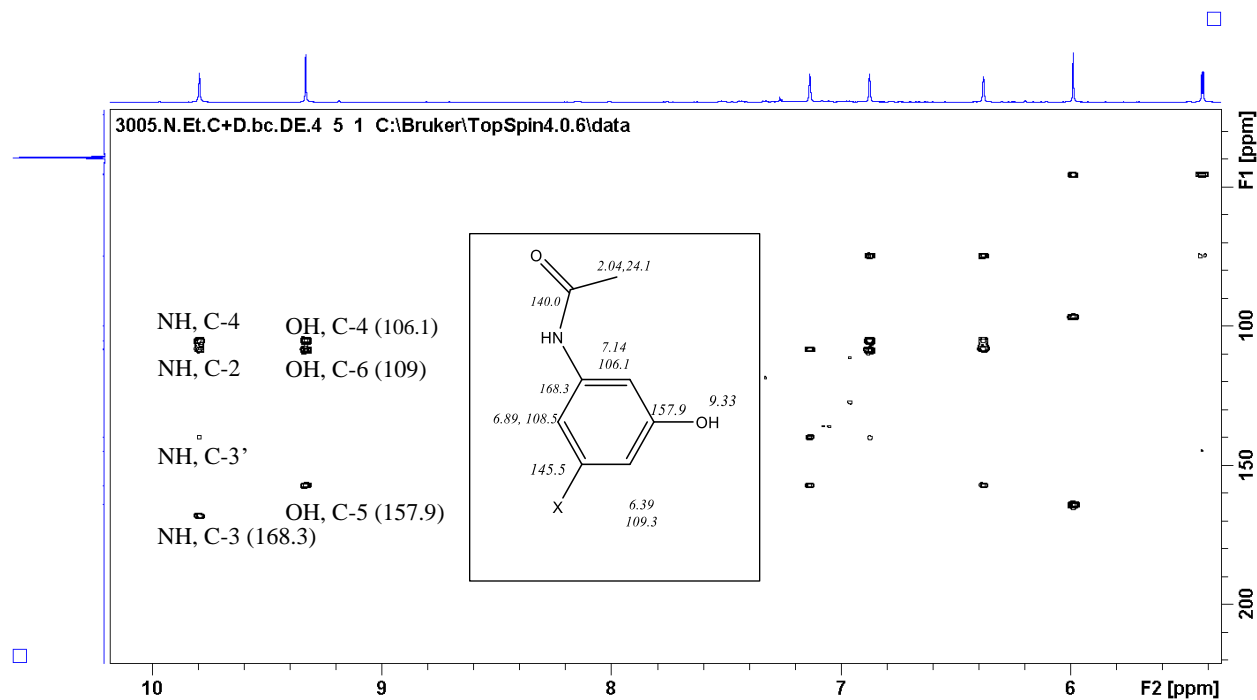
Figure 6-51. Expansion of the ^1H-^{13}C HMBC spectrum of 6.4 demonstrating key correlations leading to assignment of the asymmetric aromatic ring, recorded in DMSO-d_6, at 600 MHz.
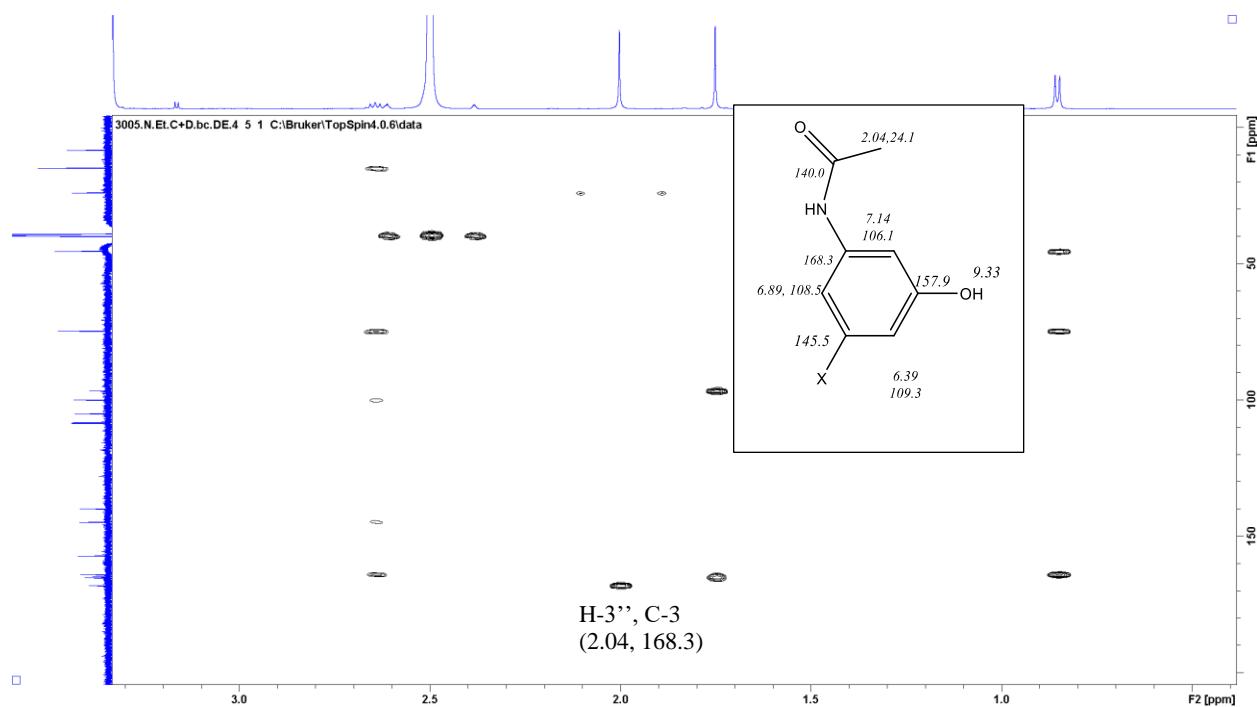
Figure 6-52. Expansion of the ^1H-^{13}C HMBC spectrum of 6.4 demonstrating the correlations leading to the assignment of an acetamide functionality, recorded in DMSO-d_6, at 600 MHz.
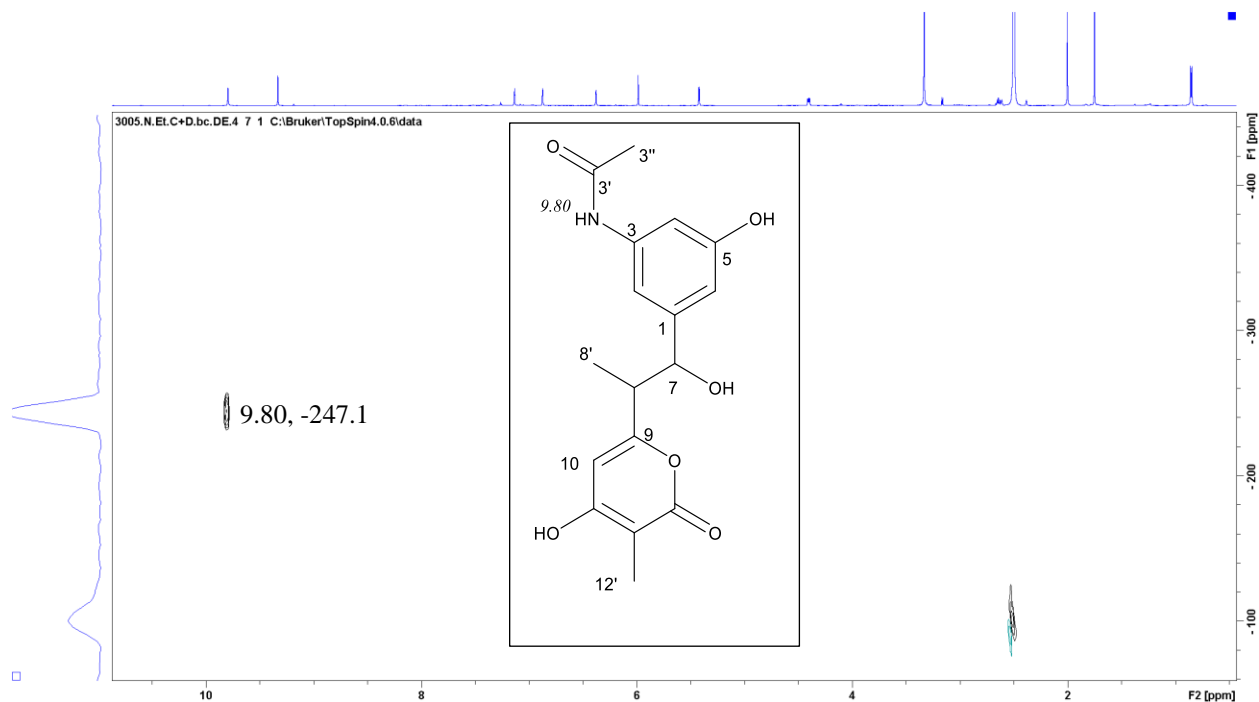
Figure 6-53. ^{1}H-^{15}N HSQC spectrum of 6.4 demonstrating the N-H correlation of the acetamide functionality, recorded in DMSO-d_6, at 600 MHz.
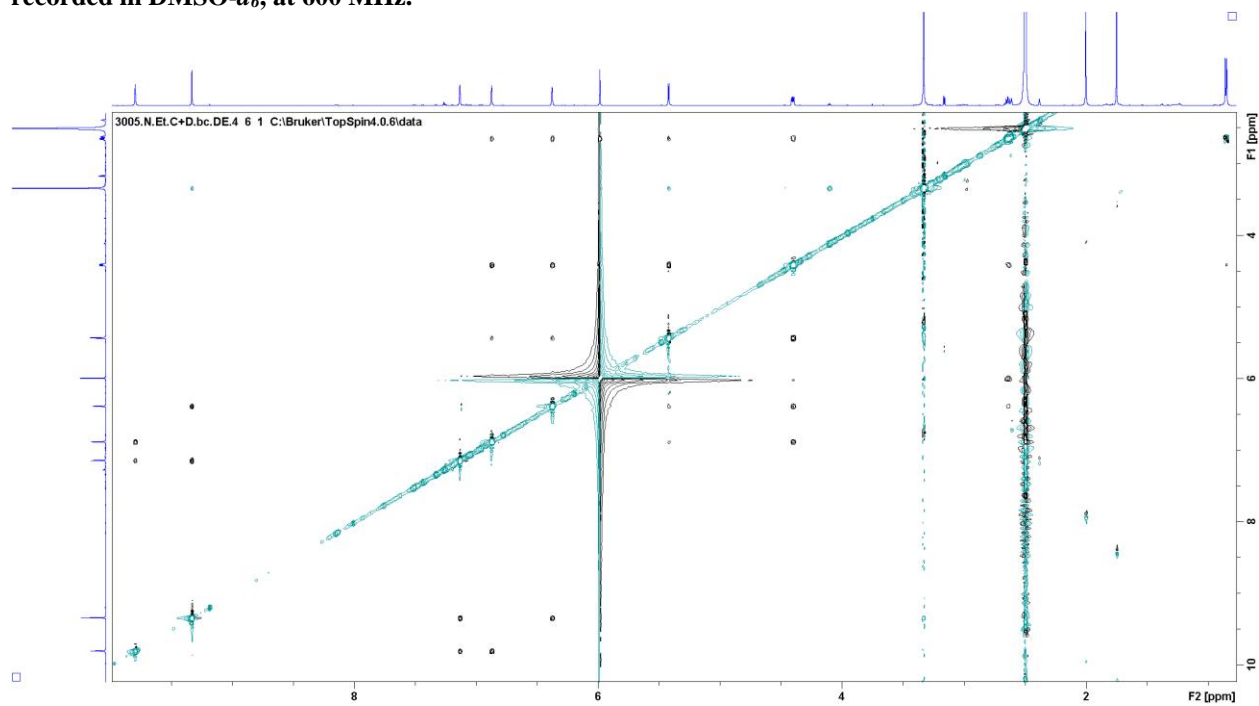
Figure 6-54. ^{1}H-^{1}H tROESY spectrum of 6.4, recorded in DMSO-d_6, at 600 MHz.

3-acetamide-5-hydroxyphenylpyrone (6.4)			
		DMSO-d_6	
Position		δ_C/δ_N	δ_H (J = Hz)
1	C	145.1	
2	CH	108.5	6.89 br.s
3'	HN**C**OCH$_3$	140.0	
3''	HNCO**C**H$_3$	24.2	2.01
3-NH	**HN**COCH$_3$	-247.1	9.80
3	C	168.3	
4	CH	106.1	7.14 br.t (1.68)
5	C	157.0	
5-OH	OH		9.33
6	CH	109.3	6.39 br. s
7	CH	74.8	4.42 dd (4.0, 8.5)
7-OH	OH		5.42 d (4.0)
8	CH	46.6	2.66 q (~7.1)
8'	CH$_3$	14.9	0.84 d (7.1)
9	C	164.1	
10	CH	100.7	5.99 s
11	C	164.7	
11-OH	OH		NV
12	C	96.8	
12'	CH$_3$	9.02	1.77 s
13	C	165.2	

Table 6-4. NMR assignments for 6.4. Recorded in DMSO-d_6 at 600 MHz.

6.4 Biosynthesis of *Salinispora arenicola* RJA3005's 6.1

Analysis of the possible biogenesis of **6.1** led to the conclusion that a 3,5-dihydroxybenzoic acid starter unit was likely because the pyrone moiety has largely been shown to form spontaneously from PKS extender units (**Figure 6-55**).[297] Analysis of the literature suggested that direct incorporation of a 3,5-dihydroxybenzoic acid as a starter
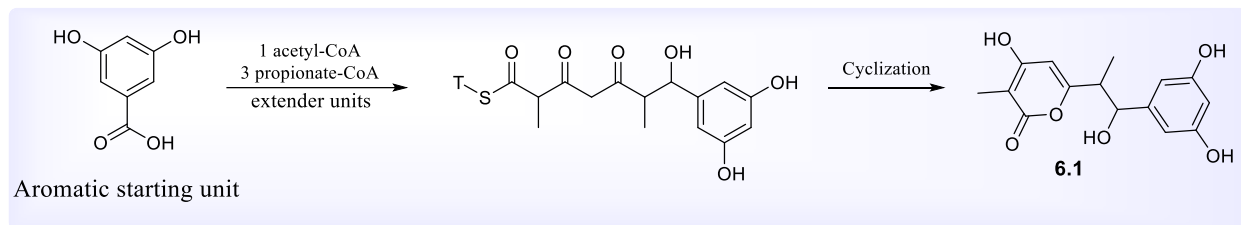
Figure 6-55. Likely pathway to the biosynthesis of 6.1. However, the biosynthetic pathway leading to the aromatic starting unit is uncertain.

unit for a polyketide synthase system had not previously been shown. Vancomycin and kendomycin show PKS III production of a 3,5-dihydroxyphenyl-2-oxoacetic acid precursor which then led to the production of extensively substituted aromatic precursors (**Figure 6-56**).[298,299] As a note, during biosynthetic characterizations of kendomycin, an initial proposal
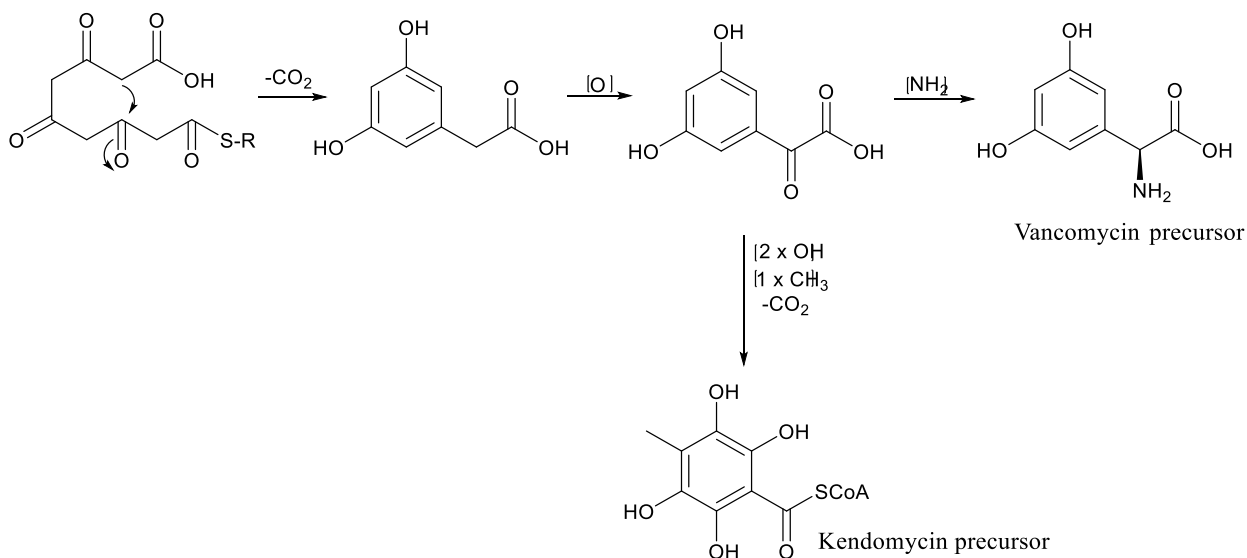
Figure 6-56. Vancomycin and kendomycin A use similar PKS III pathways to produce their aromatic starting units.

suggested that a 3,5-dihydroxybenzoic acid was uniquely prepared via a PKSIII module and then modified for incorporation into the natural product[300,301], but this proposal was later modified.
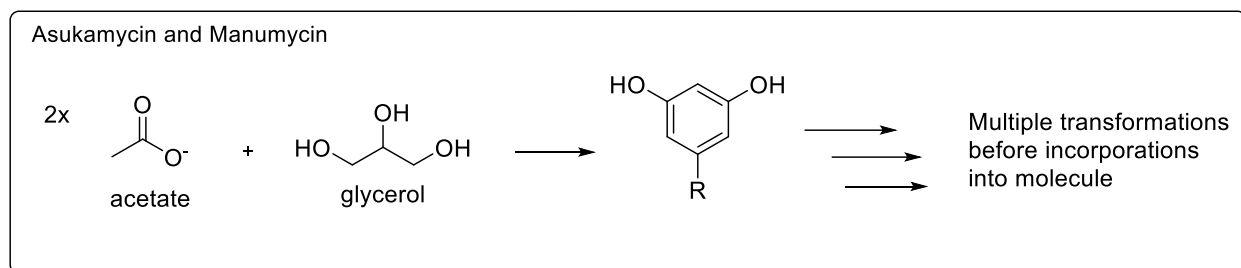
Figure 6-57. Asukamycin and manumycin incorporation of a 3,5-dihydroxy moiety during the production of their aromatic starting unit.

Asukamycin and manumycins also show incorporation of a heavily modified 3,5-dihydroxy moiety in biosynthesis of their starting unit but use a combination of acetate and glycerol building blocks (**Figure 6-57**).[302] Perhaps the most well-known example of a 3,5-dihydroxy aromatic moiety is resorcinol in plants; however, this occurs through cyclization of the terminal units (**Figure 6-58**).[106] As a note, this strategy for production of resorcinol's 3,5-dihydroxy
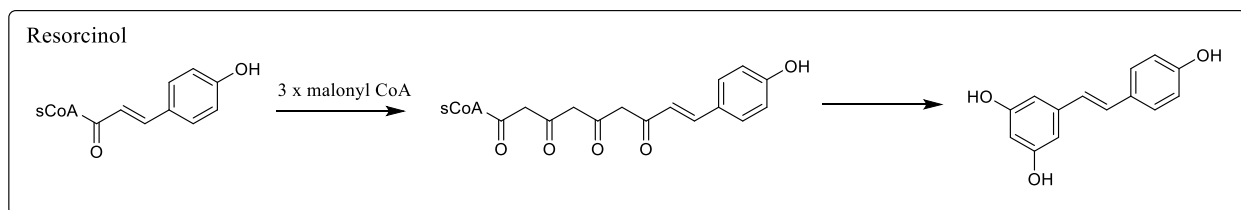
Figure 6-58. Biosynthetic strategy leading to resorcinol's 3,5-dihydroxyphenyl moiety.

substitution pattern is also found in microbes due to the cyclization of PKS units at the terminal end of fatty acid chains.[303,304] The unusual presentation of **6.1** was that both sides of the molecule had previous literature characterization from terminal cyclization of polyketide fragments. Consideration of these factors led to investigations into the biosynthesis of **6.1** using carbon-13 labelled precursors to investigate the three possible hypotheses (**Figure 6-59**) for the origin of an

aromatic starting unit. Lack of incorporation of acetate into the aromatic portion of **6.1** would rule out both the PKSIII or the manumycin pathways.
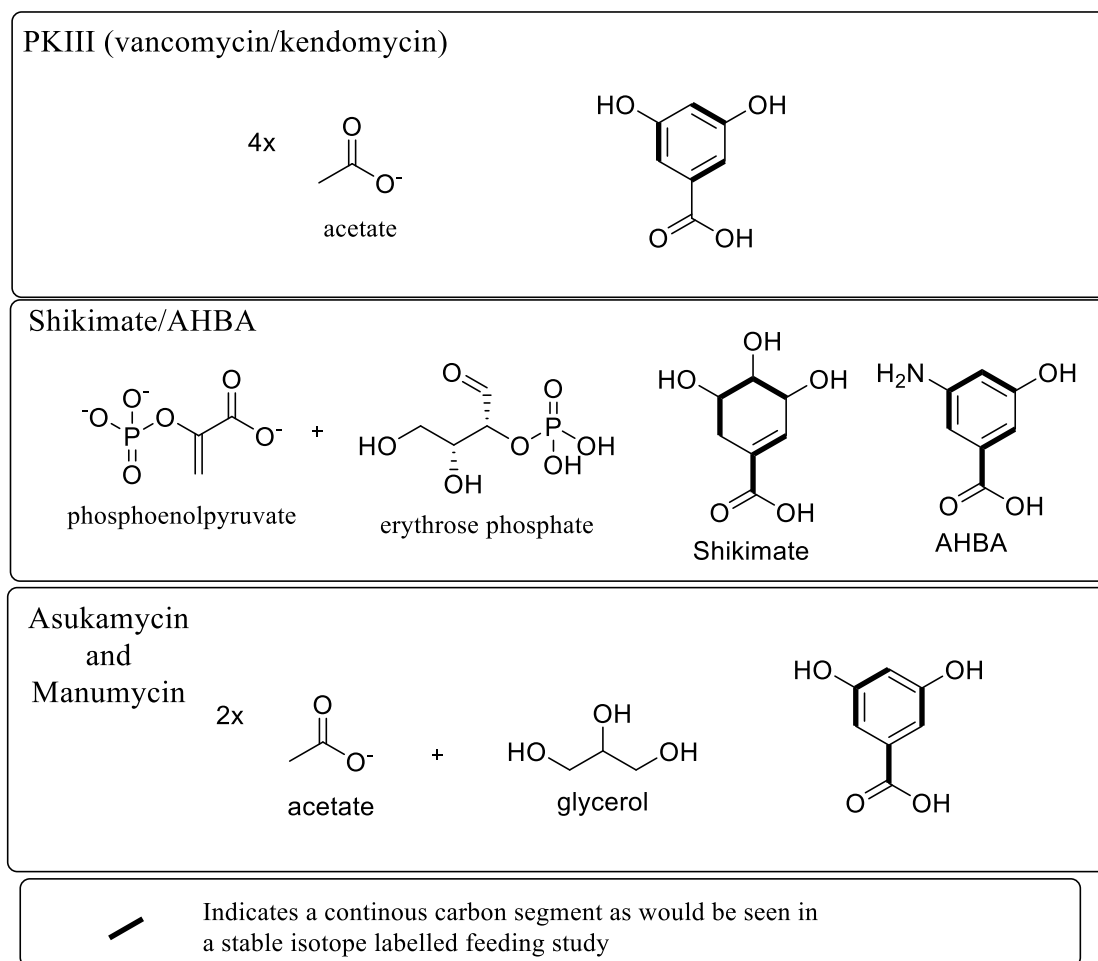
Figure 6-59. Three possible pathways to 6.1's proposed 3,5-dihydroxybenzoic acid starting unit with possible precursor carbons shown in the benzene ring.

6.4.1   Labelled Precursor Feeding Studies

6.4.1.1   1,2-^{13}C$_2$-acetate Feeding Experiments

If **6.1** was being produced entirely through a PKS pathway, then acetate would necessarily be incorporated throughout the molecule, including the benzene ring. As such 1,2-^{13}C$_2$-acetate was fed into cultures of *Salinispora arenicola* RJA3005. With doubly carbon-13 labelled acetate carbon-13 splitting patterns would be expected in the ^{13}C carbon NMR spectra. Analysis of those splitting patterns should reveal how the labelled building blocks were

266

incorporated.[133] As acetate is incorporated into the PKS extender units through the citric acid cycle this determines the expected patterns (**Figure 6-60**). *Salinispora arenicola* RJA3005 was
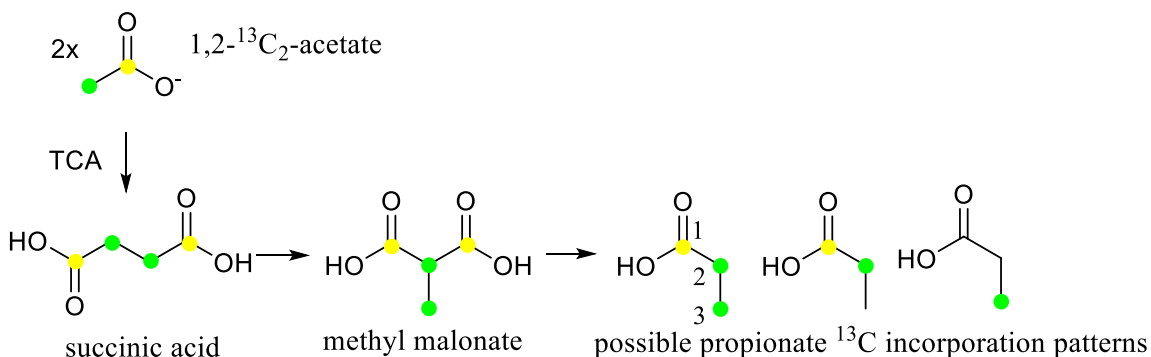
Figure 6-60. Doubly labelled acetate leads to varying ^{13}C incorporation patterns in PKS propionate extender units. Adjacent carbons from yellow and green highlights would lead to coupling patterns in the NMR spectra.

producing **6.1** in relatively small quantities, so 2L of poured agar medium was required to collect enough metabolites for appreciable carbon NMR data. 1,2-^{13}C$_2$-acetate was introduced into the solid agar of RJA3005. After growth, extraction and separation, several hundred micrograms of **6.1** was subjected to NMR analysis. Analysis of the carbon-13 coupling patterns observed in the ^{13}C NMR spectrum obtained from labelled **6.1** saw incorporation of the 1,2-^{13}C$_2$-acetate into the alkyl/pyrone portion of the molecule. Specifically, the strong singlet peak seen at C8' and C12' would come from incorporation of C-3 labelled propionyl-CoA units (**Figure 6-60**) overlapped with the resonance of the unlabelled molecules. The doublet coupling patterns seen at C8' (δ_C 14.9) (($J_{8'8} = 34$ Hz) and C12' (δ_C 8.4) (($J_{12'12} = 47.9$ Hz)) (**Figure 6-61**) demonstrated

incorporation of fully labelled propionyl-CoA units derived from succinyl-CoA in the citric acid

cycle (**Figure 6-60**), with splitting from a labelled neighbor at C-8 and C-12. Accordingly, it
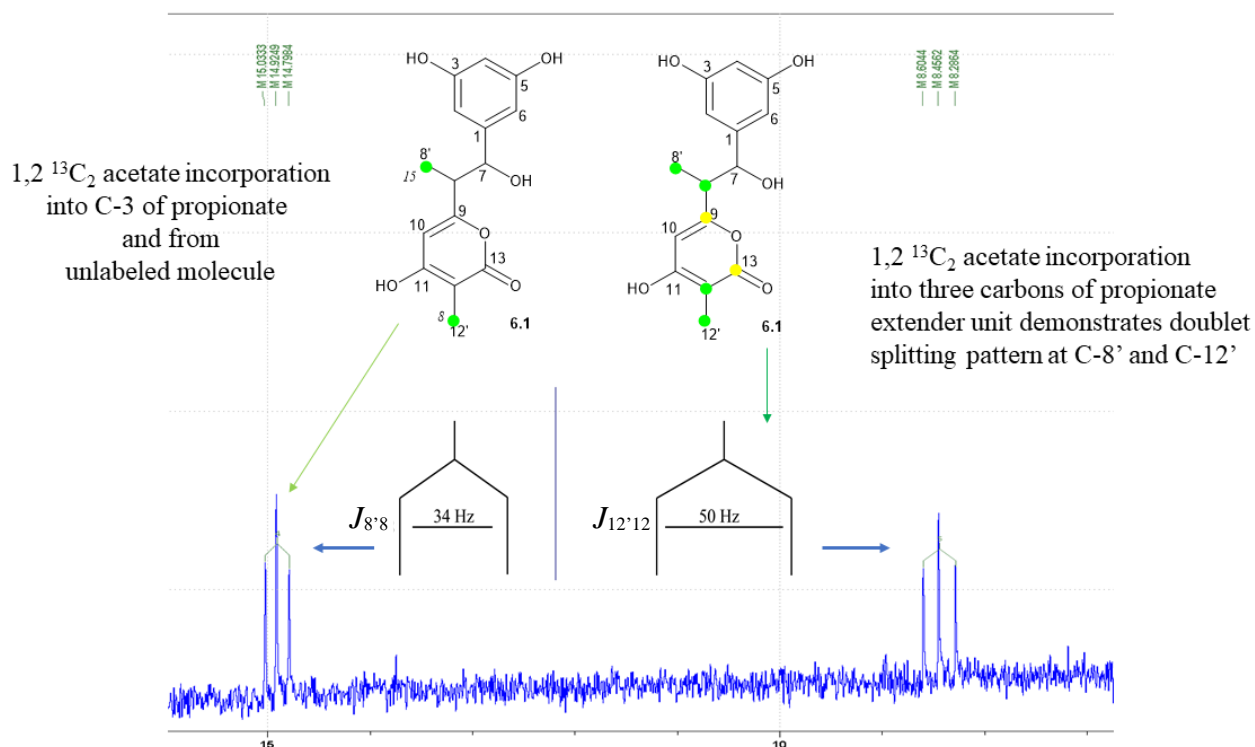
Figure 6-61. 1,2-^{13}C$_2$-acetate feeding demonstrates doublet splitting patterns for coupling between C-8'/C-8 and C-12'/C-12 overlapping the ^{13}C resonances of C-8' and C-12' from the unlabelled molecules of 6.1 as well as the singlets for C-8' and C-12'. ^{13}C spectrum recorded in DMSO-d_6, at 150 MHz.

would be expected that C-8 (δ_C 45.3) and C-12 (δ_C 99.6) would be split by either two labelled

neighbors if complete incorporation was seen in the propionyl-CoA units or by one neighbor if

incorporation as seen in C-1 and C-2 of the propionyl-CoA units (**Figure 6-60**). The resonance at

δ_C 99.6 (C-12) was not appreciably discernable. However, the resonance δ_C 45.3 (C-8)

demonstrated a complex splitting pattern consistent with being split by two neighbors in some

molecule and only one in others (**Figure 6-62**). In Figure 6-62, the coupling constant for J_{89} was

assumed to be the same as $J_{88'}$ since C-9 was not visible to measure, in reality there would
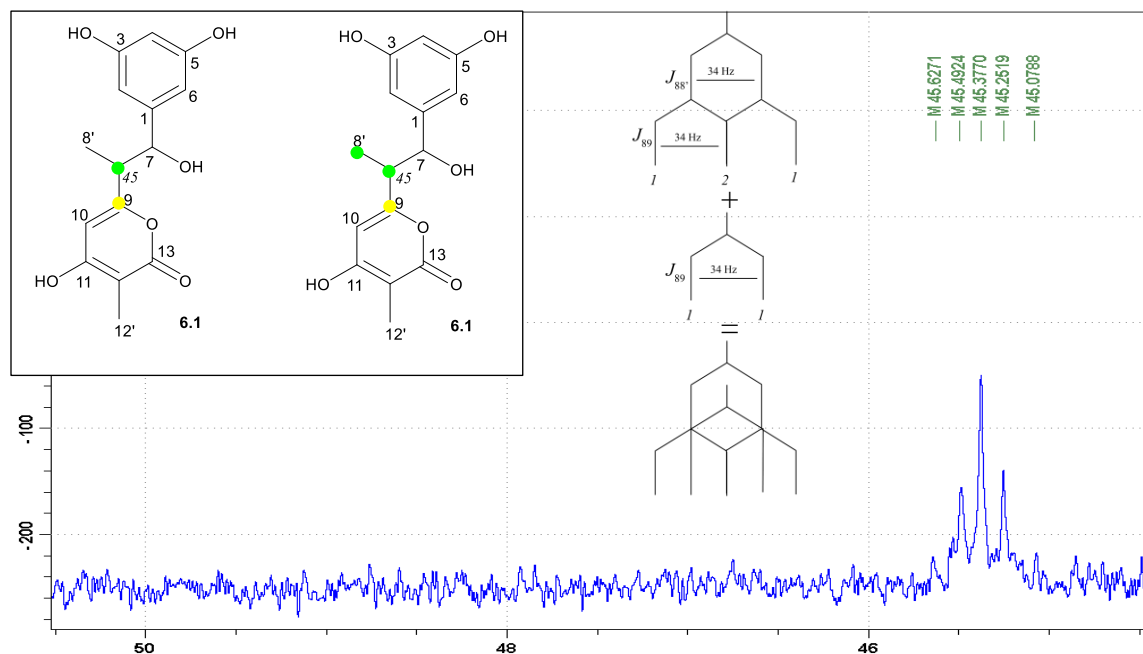
Figure 6-62. Acetate labelling incorporated into C-8 (δ_C 45). Showing a complex pattern likely arising from the overlap of greater labelling of adjacent C-9 and C-8 leading to a doublet and occasional labelling at C-9, C-8 and C-8' leading to a triplet. Both labelling patterns can arise from the production of propionate via acetate. The center peak arises from the unlabelled, uncoupled C-8 and from the center of the triplet. ^{13}C spectrum recorded in DMSO-d_6, at 150 MHz.

likely be slightly different coupling constants for J_{89} and $J_{88'}$ which adds to the complexity of the splitting pattern. There would also be expected signal at δ_C 45.3 from unlabelled molecules.

These results, showing incorporation of propionate units into alkyl portion of the molecule, mirrors the labeling patterns previously seen during rifamycin feeding studies.[279,286]

Additionally, the same carbon-13 spectrum of **6.1** demonstrating splitting patterns in the alkyl and pyrone portions of the molecule, failed to show incorporation of 1,2-^{13}C$_2$-acetate into the benzene portion of the molecule as carbon resonances for C-1 to C-7 demonstrated only carbon singlets (**Fig 6-63**). This result indicated the need for further feeding studies.
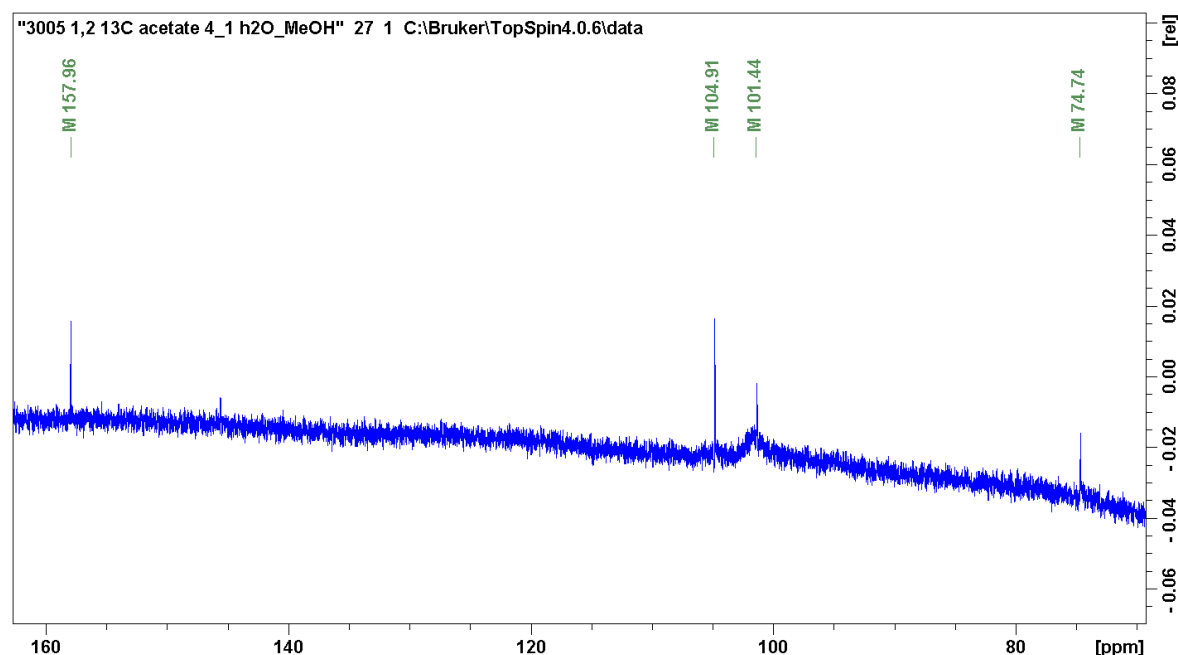
Figure 6-63. Carbon-13 spectrum of 6.1 with added 1,2-^{13}C$_2$-acetate demonstrating lack of coupling in C-2,4&6; C-3 & 5; and C-7. Recorded in DMSO-d_6, at 150 MHz.

6.4.1.2   U-^{13}C$_6$-Glucose Feeding Experiments

Although the 3,5-dihydroxy substitution pattern has not previously been shown to derive from a shikimate acid precursor, the presence of **6.4** in *Salinspora arenicola* RJA3005's culture extract pointed to the presence of mutations within the rifamycin pathway. The presence of this shunt product suggested the possibility that the aromatic starter unit in **6.1** could also be a rifamycin shunt product. The aromatic nature of the AHBA starting unit in rifamycin and the lack of incorporation of 1,2-^{13}C$_2$-acetate into the aromatic portion of **6.1** drove the investigation of this possibility. AHBA and shikimic acid follow the same biosynthetic pathway, as seen in **Figure 6.2,** save for the introduction of the nitrogen and a final aromatization via the AHBA synthase. Both AHBA and shikimic acid are derived from glucose through erythrose 4-phosphate (erythrose) and phosphoenolpyruvate (PEP) precursors. As such, U-^{13}C$_6$-glucose was added to the agar medium during growth of *Salinispora arenicola* RJA3005. As glucose is incorporated

270

into both acetate and shikimic acid metabolites, the expectation was that labelling would be found throughout the molecule (**Figure 6-64**). Analysis of the splitting patterns in the NMR ^{13}C spectrum of **6.1** after it was fed universally labelled glucose required consideration of genesis of the 3,5-dihydroxyphenyl moiety from either acetate precursors or PEP/erythrose (**Figure 6-64**). Incorporation of glucose-derived ^{13}C labelled acetate, either through a PKS III or through a shared acetate/glycerol origin (**Figure 6-59**), would see a preponderance of sharp doublets (**Figure 6-65**) for all the aromatic resonances and C-7 of **6.1**. Incorporation of
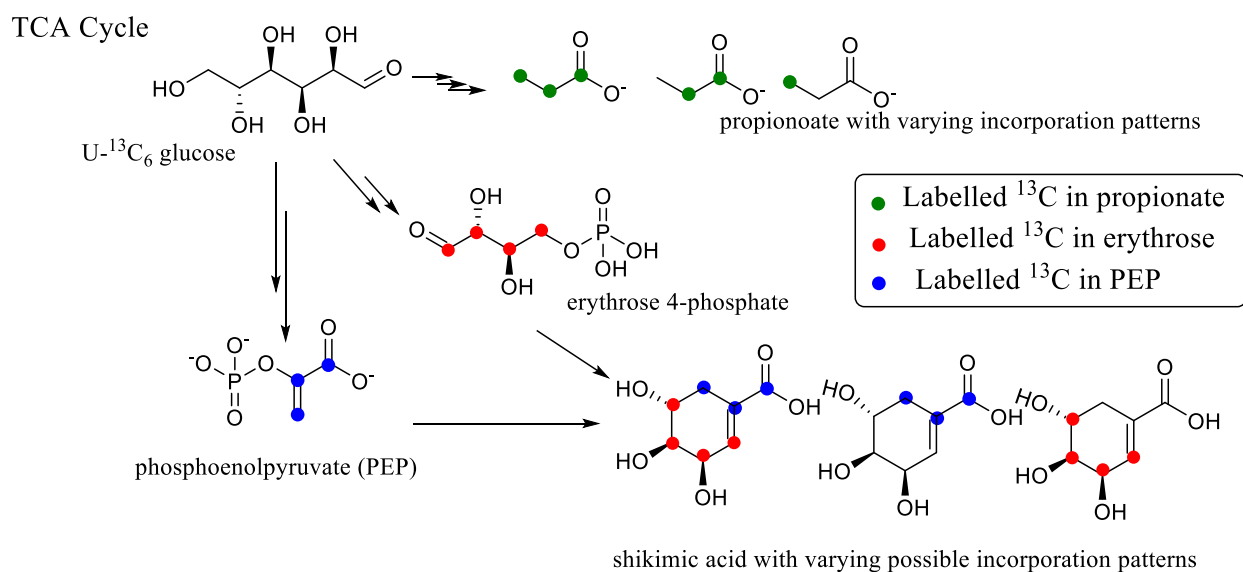
Figure 6-64. Glucose is a central metabolite that is incorporated into many molecules. However, when incorporated into acetate molecules through the citric acid cycle (TCA), it is incorporated in various ways, as seen in green here. Glucose is used to construct shikimic acid after being incorporated into two precursors, erythrose and PEP, which are incorporated into aromatic molecules stemming from shikimic acid as shown in red and blue above. The aminoshikimate pathway producing AHBA uses the same building blocks as the shikimic pathway leading to the same incorporation patterns as shikimic acid.

glucose-derived ^{13}C labelled PEP and erythrose, would see complex splitting throughout the molecule, mainly due to the different splitting patterns from the incorporation of the three-carbon PEP and the four-carbon erythrose in magnetically equivalent C-2 and C-6 in **6.1**'s benzene ring (**Figure 6-65**).
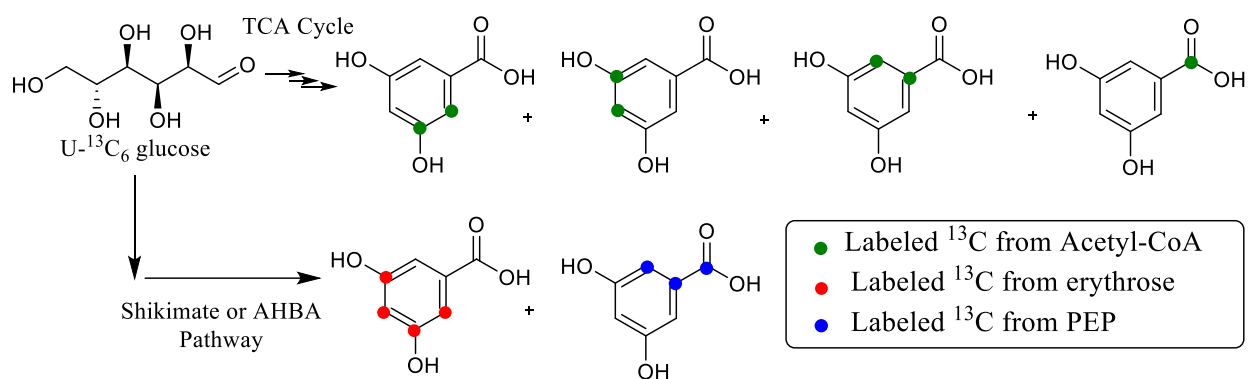
Figure 6-65. When examining the different pathways possible for biosynthetic construction of an aromatic ring, as seen in 6.1, both a PKS and a shikimate pathway were considered. All PKS units would need to incorporate as acetyl units which would when coupled together, lead to many doublets due to labelling shown in green. An aromatic ring built from universally labelled glucose through the shikimate or AHBA synthase pathways would have complex splitting patterns from multiple coupling partners, as seen in red and blue above.

Indeed, subsequent extraction, separation, and NMR analysis revealed multiple ^{13}C

carbon incorporation sites throughout the entire molecule. As glucose is incorporated into the

shikimic acid pathway and the citric acid cycle, careful interpretation of the coupling patterns

was required. Incorporation into C-8, C-8' and C-12' demonstrated a series of doublets matching

the incorporation of ^{13}C with 1,2-$^{13}C_2$-acetate addition in **Section 6.4.1.1** above. If the 3,5-

dihydroxyphenyl moiety were derived from acetate, then a predominance of doublets would be

expected for coupling partners C-2 (δ_C 105) and C-3 (δ_C 158); C-4 (δ_C 101) and C-5 (δ_C 158); C-

6 (δ_C 105) and C-1 (δ_C 145) with C-7 (δ_C 74) demonstrating a singlet peak (**Figure 6-65**).

Instead, coupling patterns for the aromatic resonances at C-2 & C-6, C-3 & C-5, C-1, C-4 as well

as at C-7, showed distinct complex splitting patterns in line with the incorporation of labelled

erythrose and PEP building blocks. The resonance at δ_C 158 demonstrated a specifically intricate

coupling pattern. Analysis of the coupling tree (**Figure 6-66**) demonstrated the overlap of a

triplet of doublets from C-3 (*ortho* coupling from C-2 and C-4 and *meta* coupling from C-5;
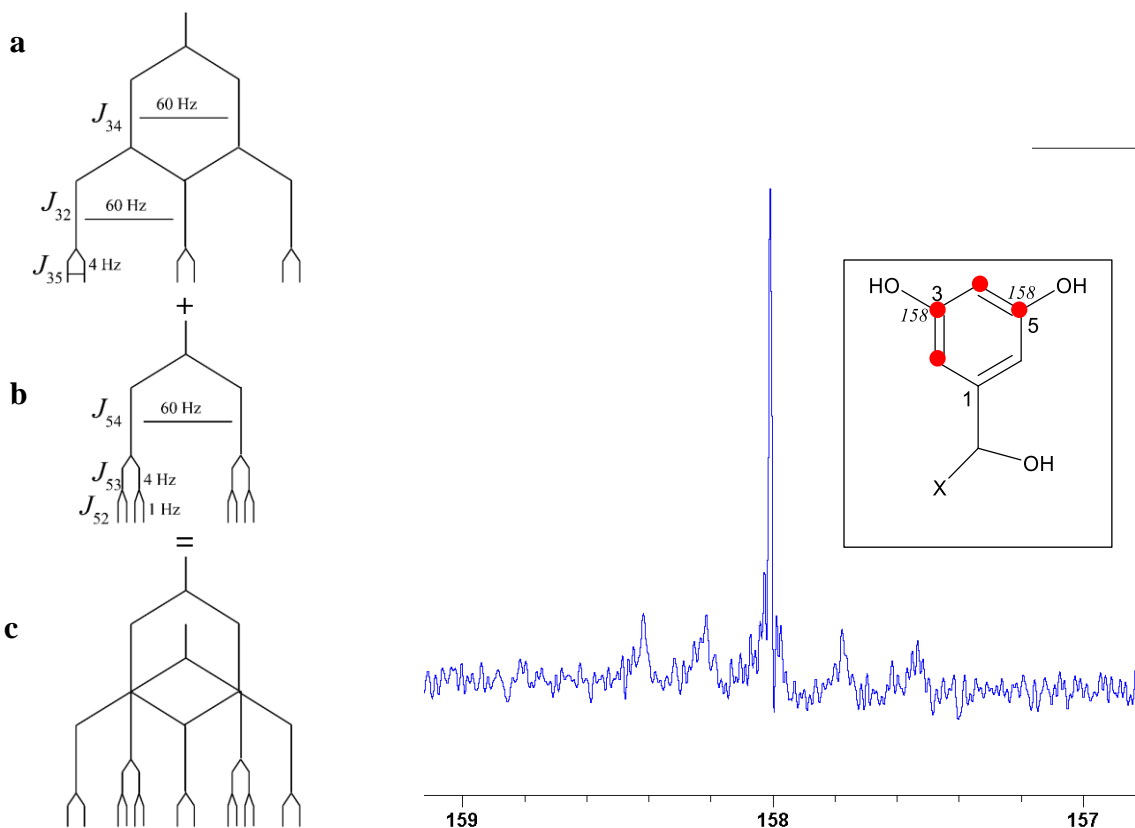
Figure 6-66. U-^{13}C$_6$-Glucose feeding experiments demonstrated complex coupling patterns around 158ppm in the ^{13}C spectrum. The most likely explanation for this complex coupling pattern is an intact fragment from an erythrose precursor (as seen in red dots) through the shikimate pathway that led to two different coupling patterns as seen in trees for a) C-3 and b) C-5. c) Overlapping coupling patterns of C-3 and C-5 lead to the observed ^{13}C spectrum. Recorded in DMSO-d_6, at 150 MHz.

J_{32}=~60Hz ; J_{34}= ~60 Hz; J_{35} = ~4 Hz) and a doublet of doublet of doublets from C-5 (*ortho* coupling from C-4, *meta* coupling from C-3 and *para* coupling from C-2; J_{54} = ~60 Hz; J_{53} = ~4 Hz; J_{52} = ~1 Hz). In **Figure 6-67,** the coupling pattern for C-4 (δ_C 101.4) with a triplet of doublets from ortho coupling with C-5 (J_{45} = ~60 Hz) and C-3 (J_{43} = ~60 Hz) and *meta* coupling with C-2 (J_{42} = ~4 Hz) is shown. However, there is a doublet seen overlapping with the C-4 signal, the origins of which is unclear. Interestingly this doublet appears to be very similar to the

273
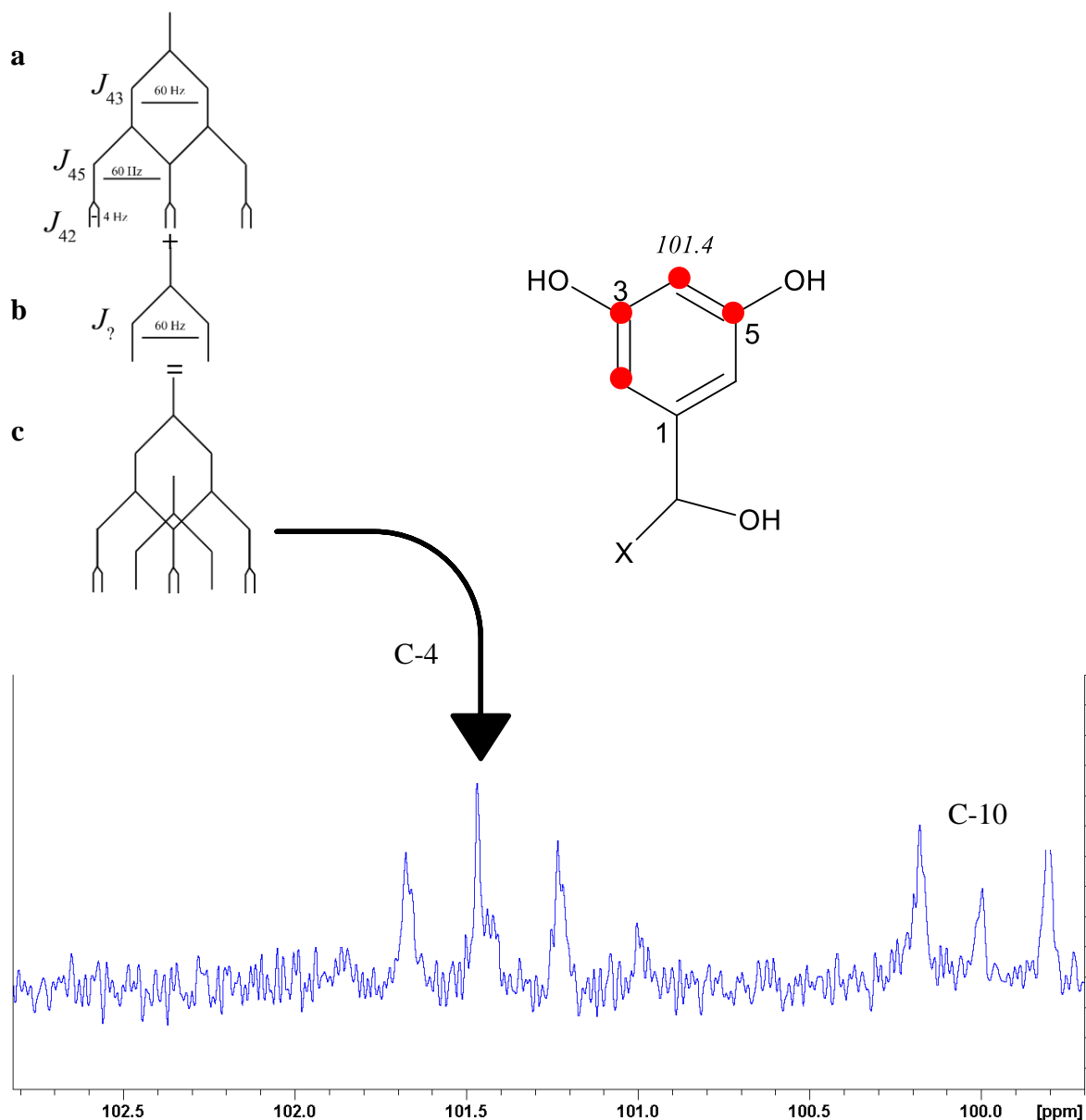
Figure 6-67. In the ¹³C spectrum, the resonance at δ_C 101.5 (C-4) in the U-¹³C₆-glucose feeding experiment demonstrated complex coupling patterns. a) Triplet of doublet splitting pattern as would arise from a labelled erythrose building block (as seen in red dots). b) The origin of the doublet with a *J* = 60 Hz is unclear. c) Overlapping coupling patterns of A and B lead to the observed ¹³C spectrum. Recorded in DMSO-*d₆,* at 150 MHz.

nearby doublet belonging to C-10 (δ_C 100). Next, in **Figure 6-68,** the coupling pattern for the

resonance at δ_C 104.8, assigned to both C-2 and C-6 of **6.1** is shown. Given the origin of C-2

from erythrose and C-6 from PEP, each would have one ortho coupling partner with

approximately the same coupling constant. Both should experience fine aromatic splitting,

274

which, when overlapped, leads to a very broad peak that can be seen to be due to multiple fine splitting in **Figure 6-68**. Analysis of the carbon resonance C-1 (δ_C 145) was difficult given the low signal of the quaternary carbon, yet a doublet of doublets (J_{16} = ~60 Hz and J_{17} = ~ 46 Hz) (**Figure 6-69**) could be discerned. This doublet of doublets would be expected from two coupling partners derived from a three-carbon PEP precursor. Finally, C-7 (δ_C 74.7), demonstrated a doublet of doublet arising from its neighbor at C-1 (J_{71} = ~46 Hz) and fine splitting from the labelled carbon at C-6 (J_{76} = ~3 Hz). Such coupling at C-7 was only possible if it was part of a three-carbon fragment from PEP.

The coupling pattern analysis led to the conclusion that the 3,5-dihydroxyphenyl moiety in **6.1** was likely produced through PEP and erythrose building blocks. At this stage, it is difficult to know whether this is through a modified aminoshikimate or shikimate pathway. However, the presence of **6.4** suggests there is undoubtedly a mutation somewhere in *Salinispora arenicola* RJA3005's rifamycin gene cluster. It has also been shown that the aminoshikamate pathway that produces aminoDAHP on its way to AHBA can produce the DAHP precursor of the shikimate pathway.[132]
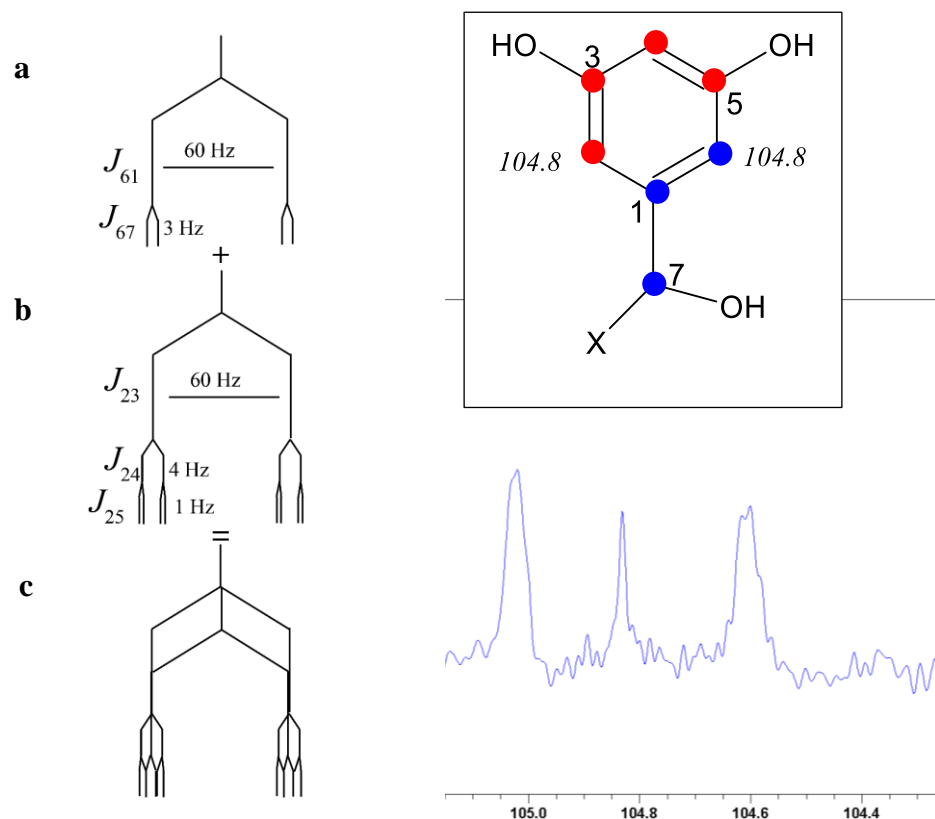
Figure 6-68. ^{13}C spectrum from the U-$^{13}C_6$-Glucose feeding experiments demonstrated a broad doublet containing multiple fine splitting around 104.8 ppm. Due to the symmetry of the molecule the 104.8 ppm chemical shift arises from both C-2 and C-6. The most likely explanation for this complex coupling pattern is a) An intact fragment from a PEP precursor (as seen in blue dots) leads to a doublet (60 Hz) of doublet (3 Hz) of doublets (1Hz) for C-6 and b) a erythrose precursor (as seen in the red dots) leading to a doublet (60 Hz) of doublets (3 Hz). c) Here the overlapping coupling patterns of C-2 and C-6 leading to the observed broad doublet with fine splitting at δ_C 104.8 in the ^{13}C spectrum are seen. Recorded in DMSO-d_6, at 150 MHz.
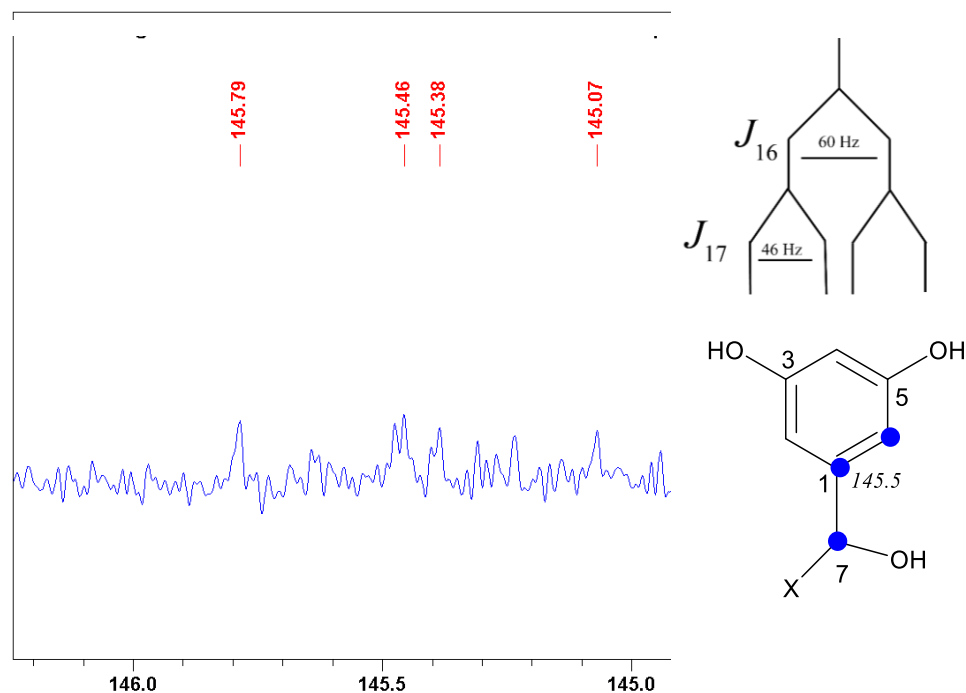
Figure 6-69. ^{13}C **spectrum from the U-**$^{13}C_6$**-Glucose feeding experiments demonstrated difficult to interpret weak signals at C-1 at 145 ppm. However, there is a strong suggestion of a doublet of doublet. The most likely explanation for this splitting pattern is an erythrose precursor (as seen in the blue dots) leading to a doublet (60 Hz) of doublets (46 Hz). Recorded in DMSO-**d_6**, at 150 MHz.**
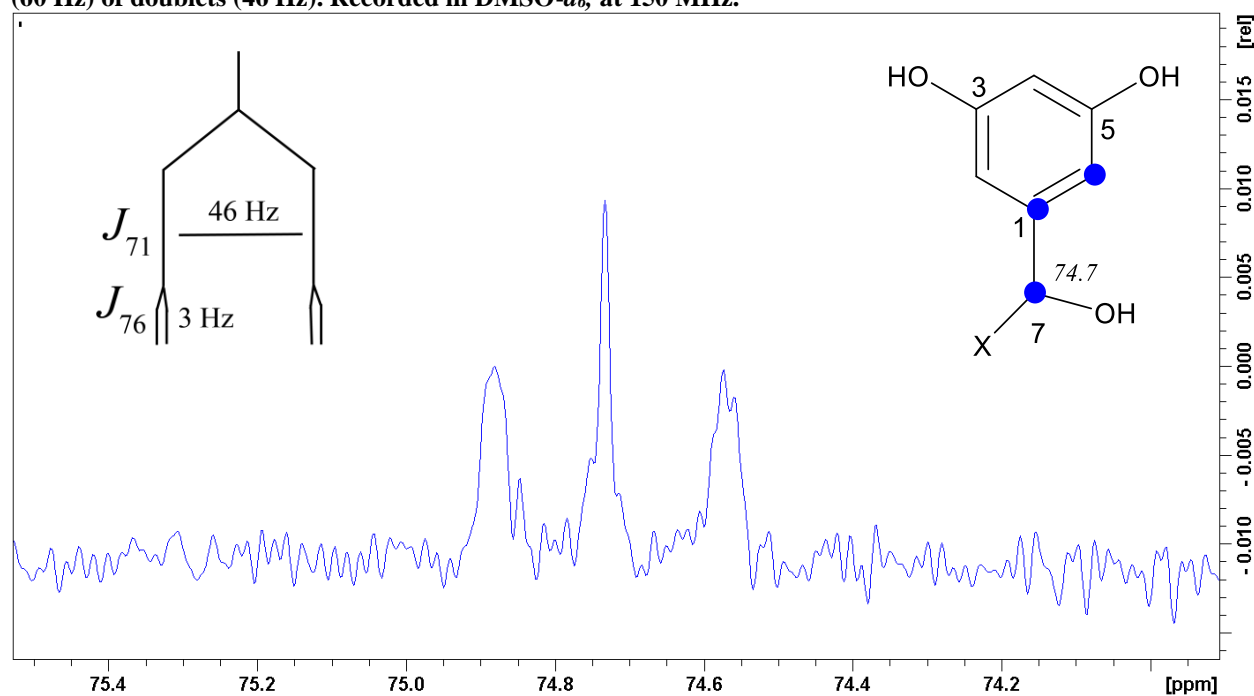
Figure 6-70. ^{13}C **spectrum from U-**$^{13}C_6$**-Glucose feeding experiments demonstrating a doublet with fine splitting for C-7 at 74.7 ppm. The most likely explanation for this splitting pattern is an erythrose precursor (as seen in the blue dots) leading to a doublet (46 Hz) with long range coupling lead to a second doublet (3 Hz). Recorded in DMSO-**d_6**, at 150 MHz.**

277

6.4.2 Bioinformatic Analysis of the *Salinspora arenicola* RJA3005's Draft Genome

It is difficult to predict the exact biosynthetic pathway from the above evidence. Analysis of the draft genome of *Salinspora arenicola* RJA3005 indicates the presence of both the rifamycin gene cluster and the AHBA synthase cluster with high similarity to other *Salinspora arenicolas* (**Figure 6-67**), although both are part of incomplete gene clusters in the draft genome for RJA3005. The closest AHBA synthase homolog resided on contig 17 of RJA3005's draft genome sequence. Contig 17 also bears other genes required for AHBA biosynthesis. It bears 100% similarity to other Salinispora arenicola with rifamycin homologous gene clusters. The only place of difference is a tiny gene (171 bp) with homology to hypothetical proteins in BLASTP searches.

Figure 6-71. Partial gene cluster comparison from the contig 17 of *Salinispora arenicola* RJA3005 as produced by antiSMASH. The AHBA synthase homolog is identical to other *S. arenicola* strains, save for one tiny gene whose closest homolog is a hypothetical protein. The yellow box is highlighting this solitary gene.

6.5 Bioactivity

There was enough of **6.1** to test the natural products in several bioassays. **6.1, 6.2** and **6.3** were subjected to the phosphatase assay with no activity. They were also subjected to anti-bacterial assays with no activity. They were tested against the human pathogen Candida albicans with no activity. There were some initial weak results from **6.1** as an antifungal against specific plant pathogens in a modified disc diffusion assay however they were considered too weak as a pure compound to consider pursuing. There were very small amounts of **6.4** isolated and as such it was not tested in any bioassay.

6.6 Summary and Conclusion

This project demonstrated the isolation of novel natural products through chemical prospecting from an initial survey of a biosynthetically talented *Salinispora arenicola* RJA3005 in a phosphatase assay. Although **6.4** and **6.1** have previously appeared in the literature, both compounds were only found from a mutated bacterial rifamycin-producer. **6.4** was shown to be a metabolic offshoot of a mutated rifamycin pathway. Additionally, **6.1** was only produced after a mutant was fed synthesized 3,5-dihydroxybenzoic acid, as such this is the first instance of it being shown as a true natural product. Intriguingly, feeding studies suggest that the resorcinol substitution pattern is formed via a modified shikimic acid or AHBA synthase pathway, which would be the first demonstration of such production. Recently, a *Salinispora arenicola* of Brazilian origin was shown to be producing a related molecule to both **6.1** and **6.4** except without any aromatic substitutions,[305] lending support to the feeding study results reported herein.

Initial bioinformatic analysis of a draft genome found only a small variant in the AHBA synthase gene cluster present in *Salinispora arenicola* RJA3005. Given the results from the

bioinformatic analysis, further knock-out studies of the AHBA synthase gene cluster of *S. arenicola* RJA3005 should be completed to confirm the results of the feeding studies. Furthermore, additional stable isotope feeding studies with labelled PEP, erythrose, DAHP and aminoDAHP could be undertaken to confirm the splitting patterns seen with the feeding of universally-labelled glucose. If indeed **6.1** is produced from a slightly modified AHBA synthase then the characterization of this modification along with the wild mutation leading to **6.4** could support engineering efforts of these types of molecules.

6.7   Experimental Section

6.7.1   General Experimental Procedures

All solvents used for partitioning and flash chromatography were Sigma-Aldrich HPLC grade, used without further treatment. For HPLC work, Sigma-Aldrich HPLC grade solvents were filtered through Milli-Q filter paper prior to use. Distilled water was used to prepare all growth media for terrestrial bacteria. Seawater collected by M. Leblanc was used to prepare growth media for all marine bacteria. Milli-Q H_2O was used for partitioning, reversed-phase Sep-Pak and HPLC. The ^1H, ^{13}C and ^{15}N NMR spectra were recorded on a Bruker AV-600 spectrometer with a 5 mm CPTCI cryoprobe. ^1H chemical shifts are referenced to the residual DMSO-d_6 peak (chemical shift δ 2.50) and ^{13}C chemical shifts are referenced to the DMSO-d_6 solvent peak (δ 39.51). The ^{15}N NMR spectra are in reference to CH_3NO_2 (0 ppm) with a variability of 1 ppm. Low resolution electrospray-ionization mass spectrometer (LRESIMS) were recorded on Bruker HCT Ultra PTM Discovery System and high resolution electrospray-ionization mass spectrometer (HRESIMS) were recorded on a Waters / Micromass LCT time-of-flight (TOF) mass spectrometer equipped with an electrospray ion source. Merck Type 5554

silica gel plates and Whatman MKC18F plates were used for analytical thin layer chromatography (TLC). Sephadex™ LH-20 purchased from Sigma was used for gel filtration chromatography. *para*-Anisaldehyde stain with subsequent heat application was used to visualize TLC spots. Sep-Pak SPE Silica and C18 cartridges (2 g, 5 g, and 10 g) were purchased from Waters. All media components were obtained from Fisher Scientific. Reversed-phase HPLC purifications were performed on a Waters 1525 Binary HPLC pump attached to a Waters 2998 Photodiode Array Detector and equipped with Empower Software.

6.7.2   Isolation of metabolites

Salinispora arenicola RJA3005 was isolated from marine sediment collected in Papua New Guinea by Dr. Doralyn D. Dalisay. Bacterial cultivation was carried out using nutrient rich media - Marine Medium 1 (MM1- soluble starch: 10.0 g; yeast extract: 4.0 g; peptone: 18.0 g; sea water (marine strains) or distilled fresh water (terrestrial strains): 1.0 L; KBr: 0.001 g and FeSO4.7H$_2$O: 0.0004 g. Solid media had agar added (2.0 g/L). *Salinispora arenicola* RJA3005 was grown for 14 days on solid agar until a thick mat of orange leathery textured bacterial colonies had formed. After 14 days, the agar and mycelia were cut into small squares and immersed in EtOAc for extraction. The EtOAc soaked agar was filtered through filter paper to separate the agar from the supernatant (3x, ~2 L solvent per 8 L growth medium). The EtOAc portions were then combined and the solvent removed by evaporation *in vacuo* to reveal a thick red/brown oily-solid, with large portions of the extract only soluble in MeOH until further separation. As such, the crude was dissolved in a 9:1 mixture of H$_2$O:MeOH (v/v). Subsequently, this crude extract was sequentially partitioned between first hexane/ H$_2$O and then CH$_2$Cl$_2$/H$_2$O after which time the aqueous partition had MeOH removed under reduced pressure to leave a

deep golden aqueous solution containing reddish brown oily-solids. At this time, EtOAc was added to completely dissolve the oily-solids and the EtOAc-soluble partition was removed and dried under reduced pressure to reveal a reddish golden viscous oil. The EtOAc-soluble partition was the active fraction in the phosphatase inhibition assay and was selected for further separation. This partition was then passed through a Sephadex® LH-20 size exclusion column (4:1 MeOH/ CH$_2$Cl$_2$) and fractioned based on TLC staining of smaller test tubes. Active fractions were then separated using step-gradients on flash chromatography (H$_2$O to MeOH), using C18 Sep-paks (2 g, 5 g, 10 g depending on initial amount of material grown). The fraction eluting in 4:1 H$_2$O:MeOH and 3:2 H$_2$O:MeOH (v/v) contained the NMR peaks of interest and were further separated using semi-preparative reversed-phase HPLC (InertSustain C18, 17:83 MeCN:H$_2$O with 0.05% FA).

3005.Et.CD.C.4: 6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-hydroxy-3-methyl-2H-pyran-2-one (6.1) : Isolated as a white powder; $[\alpha]^{25}_D$ -8.79° (c 3.5 g/100 mL, MeOH); ^{1}H (DMSO-d_6, 600 MHz) and ^{13}C NMR (DMSO-d_6, 150 MHz), see **Table 6-2**; HRMS (TOFESI) *m/z*: [M+H]+ Calcd for C$_{15}$H$_{17}$O$_6$, 293.1025; Found 293.1046.

6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-4-methoxy-3-methyl-2H-pyran-2-one (6.2): Isolated as an off-white powder; ^{1}H (DMSO-d$_6$, 600 MHz) and ^{13}C NMR (DMSO-d$_6$, 150 MHz)

6-(1-(3,5-dihydroxyphenyl)-1-hydroxypropan-2-yl)-2-methoxy-3-methyl-4H-pyran-4-one: (6.3): Isolated as an off-white powder; ^{1}H (DMSO-d$_6$, 600 MHz) and ^{13}C NMR (DMSO-d$_6$, 150 MHz);

3005.Et.CD.E.4 : N-(3-hydroxy-5-(1-hydroxy-2-(4-hydroxy-3-methyl-2-oxo-2H-pyran-6-yl)propyl)phenyl)acetamide (6.4): Isolated as a white powder; ^{1}H (both DMSO-d$_6$ 600 MHz) and ^{13}C NMR (both DMSO-d$_6$ 150 MHz), see **Table 6-3**; HRMS (TOFESI) *m/z*: [M+H]+ Calcd for $C_{17}H_{20}O_6N$, 334.1291; Found 334.1297.

6.7.3   Methylation

1.3 mg (0.0045 mmol) of **6.1** was dried in a vial and left overnight on the lyophilizer to remove any residual waters. A 5:1 (625 µL:125 µL) Toluene(anhydrous):MeOH v/v solution was added to the vial under inert conditions. Purchased trimethylsilyldiazomethane (*TMS*CHN$_2$, 3.35 µL, 0.6 mmol) was added dropwise to the stirring solution. The solution was left to stir overnight under inert conditions and monitored by TLC. Upon completion, the products were concentrated by removing the solvent and purified by reversed-phase HPLC (InertSustain C18; step gradient 0-60 min 17:83 MeCN:H$_2$O followed by 61 mins to 120 mins 23:77 MeCN:H$_2$O; methyl ester of the 4-pyrone isomer eluting between 55-57 min and the methyl ester of the enol isomer eluting between 75-80 mins)

6.7.4   Mosher's Ester

Compound **6.1** (1.3 mg) was divided into two clean NMR tubes which were dried completely on a lyophilizer under the vacuum of an oil pump. Deuterated pyridine (0.5 mL) was added to each NMR tube and either *R*-(-)-MTPA-Cl (3R, 10 µL) or *S*-(+)-MTPA-Cl (3S, 10 µL) were added into the NMR tube immediately under a N$_2$ gas stream and the top sealed with parafilm. Then the NMR tube was shaken carefully to mix the sample and MTPA chloride evenly. The reaction NMR tube was permitted to stand at room temperature and monitored every

1-2 h by ^{1}H NMR. The reaction was found to be complete after 4 h. Given the complex mixture and small amount of sample, COSY and HSQC experiments were recorded in order to assign the chemical shifts for both the (*R*)-MTPA ester derivative and (*S*)-MTPA derivative. NMR assignments (as was possible) are found in **Table 6.3**.

6.7.5   Feeding studies

1,2-^{13}C$_2$-acetate feeding: 1500 mg of 1,2-^{13}C$_2$-acetate was filter sterilized into 2 L of autoclaved MM1 agar medium prepared with sea water. This agar media, containing the labelled precursor, was poured into 6 stainless steel trays (37 mm X 20 mm) and *Salinispora arenicola* RJA3005, growing in liquid medium for 5-7 days, was spread over the surface of the cooled agar. The trays were covered with sterile aluminum foil and left to grow for 14 days. After 14 days and the formation of thick orange mat of bacterium, the agar and mycelia were cut into small squares and harvested into EtOAc. After a minimum of 24 hrs, the EtOAc extract was filtered through cotton wool and then remove under reduced pressure to reveal a deep yellow gold extract. After separation utilizing C18 Sep-Pak (10 g, step-gradient 4:1 H$_2$O:MeOH to 3:2 H$_2$O:MeOH) the fraction eluting with 3:2 H$_2$O:MeOH was placed onto the HPLC and separated by conditions previously determined for **6.1**.

U-^{13}C$_6$-glucose feeding: 850 mg of U-^{13}C$_6$-glucose was filter-sterilized into 2 L of autoclaved MM1 agar medium prepared with sea water. This agar media, containing the labelled precursor, was poured into 6 stainless steel trays (37 mm X 20 mm) and *Salinispora arenicola* RJA3005, growing in liquid medium for 5-7 days, was spread over the surface of the cooled agar. The trays were covered with sterile aluminum foil and left to grow for 14 days. After 14 days and the formation of thick orange mat of bacterium, the agar and mycelia were cut into small squares and

harvested into EtOAc. After a minimum of 24 hrs., the EtOAc extract was filtered through cotton wool and then removed under reduced pressure to reveal a deep yellow gold extract. After separation utilizing C18 Sep-Pak (10 g, step-gradient 4:1 H_2O:MeOH to 3:2 H2O:MeOH) the fraction eluting with 3:2 H_2O:MeOH was placed onto the HPLC and separated by conditions previously determined for **6.1**.

Chapter 7: Conclusion

The work presented in this thesis was conducted in an effort to discover novel natural products with potential therapeutic properties and to investigate the biosynthesis of natural products. Bacteria are a relevant source for exploration, given that they are the origin of many current pharmaceutical medicines. Both marine and terrestrial bacterial sources were explored utilizing a variety of discovery methods, including a ^{15}N NMR-guided approach developed through this thesis work and the robust bioassay-guided fractionation.

^{15}N NMR-guided discovery of Piz-containing natural products:

To summarize: In Chapters 2, 3 and 4, I demonstrated the development of a novel method for the targeted discovery of Piz-containing natural products. Chapter 2 describes the bioinformatic methodology utilized to harness the power of characterized genetic sequences to select for bacterial producers of desired natural products. Chapter 3 details the ^{15}N NMR experiments used to track Piz's diagnostic N-H correlation and N-H connected spin system from crude bacterial culture extract to the isolated natural product. ^{15}N$_2$-L-Orn, a stable isotope labelled precursor of Piz, was fed to the bacterial producer to amplify the N-H NMR signal, allowing its observation from crude extract to a single molecule. This amplified signal allowed observation of minute amounts of a Piz-containing natural product in the crude bacterial extract of a genome-mined bacterium. Thus, linking genome mining with NMR for targeted discovery allowed for the isolation of several novel very minor metabolites. Chapter 4 details the structure elucidation of those metabolites, including the structure elucidation of a novel moiety in chemical space. One of the members of this unique family was a bioactive natural product. Identification of novel drug leads is the first step in drug discovery that can lead to

pharmaceuticals development. Thus, Chapters 2, 3 and 4 demonstrate that natural product discovery methods rationally targeting moieties with rich associated bioactivity or structural uniqueness can lead us to our desired goals.

In future directions: This ^{15}N NMR-guided methodology is being applied to isolate Piz-containing natural products from diverse bacterial sources. Additionally, although this method was targeted to Piz compounds, nitrogen-nitrogen bond containing natural products and other nitrogen-containing moieties are underrepresented in elucidated natural products compared to their numbers in sequenced genomic databases. Utilizing genome mining to select species producing molecules of interest followed by careful analysis of unique ^{15}N NMR features would allow the elucidation of more significant numbers of these compounds. Coupled with feeding labelled precursors and additional nitrogen NMR experiments, various ^{15}N NMR features, including N-H correlations and longer-range nitrogen correlations, could be exploited for targeted isolation of N-N bond containing natural products.

Furthermore, Piz-containing natural products seem to be generally contained to the Actinobacteria bacterial phylum and have been most often found from *Streptomyces* species. There have been characterized gene clusters for many of the different families of natural products that contain Piz. Through bioinformatics analysis, it is also seen that there are many gene clusters for which the natural product phenotype is unknown. Within these characterized gene clusters and orphan gene clusters, there appears to be diversity of gene cluster organization found in members of different Actinobacteria genera and, to some degree, amongst the more closely related *Streptomyces* species. Utilizing this Piz-targeted ^{15}N NMR strategy to uncover Piz-containing natural products from various bacterial strains could reveal potential drug leads and new natural product families.

Finally, this method demonstrates how developing our understanding of biosynthetic pathways can couple with chemical knowledge for natural product discovery, expanding our natural product library in specific ways. Therefore, this method adds a tool to future discovery processes. As the chemical library grows and genotype linked to chemotype, understanding of the nuances in bacterial secondary metabolite gene cluster organization and genomic organization is enhanced. This information can, in turn, be utilized to uncover diversity present in whole families of natural products. Pursuing diverse members of interesting natural product classes may, in turn, lead to novel bioactivities of related compounds or similar bioactivities with novel mechanisms of action.

Bioassay-guided discovery of natural products:

To summarize: in Chapter 5, the isolation and structure elucidation of secondary metabolites from crude microbial extracts identified the biologically active natural products in their respective bioassays. In addition, the appearance of novel activity in manufactured bacterial fusants from several wild-type strains was pinpointed to the *de novo* expression of a known antifungal. This result could, in the future, help build an understanding of the mechanisms driving secondary metabolite operon expression due to genetic recombination events.

To summarize: in Chapter 6, the bioassay-guided exploration of a small library of crude extracts also led to the isolation and structure elucidation of natural products from wild-type members of the Salinispora genus. Several natural products were uncovered that had not previously been reported in the literature as naturally-derived molecules, although bioactivities for them are not clear. These natural products had previously been reported in the literature as the products of genetic intervention in the rifamycin pathway in the *Amycolatopsis mediterranei* S699 strain. These molecules not only serve as potential lead compounds but also expand the

known chemical library. Feeding studies were undertaken to understand the natural biosynthetic pathway of these natural products. The results of these feeding studies potentially point to modifications in the AHBA synthase cassette and the rifamycin pathway in the *Salinispora arenicola* RJA3005 strain.

Given the importance of rifamycin as a potent antibiotic, the AHBA synthase gene cluster's potential modification may be of particular interest for the future scope of this work. Knock-out studies would be necessary to unequivocally link **6.1** with the rifamycin gene cluster in *Salinispora arenicola* RJA3005. Previously, the incorporation of the 3,5-dihydroxybenzoic acid moiety into the rifamycin pathway was achieved by feeding synthetic precursors. In other pathways, the incorporation of this moiety has been the foundation for highly modified starting units produced through PKS III mechanisms. These results point to a novel microbial mechanism for biosynthesizing this moiety. It is also interesting to observe that the rifamycin gene cluster, a highly conserved gene cluster found in diverse bacterial species, could be subject to modifications that prevent rifamycin production as a main secondary metabolite. It could also be utilized in understanding how wild mutations can occur in this sizeable modular synthase.

A further avenue for future exploration is comparing both the rifamycin and AHBA synthase gene clusters and secondary metabolite production of the other Salinispora strains retrieved from the same or similar environments detailed in the curated mini-library. As our understanding of bacterial community metabolic pathways has been increasing over the last decade, this could build knowledge on how bacterial communities may be incorporating secondary metabolite gene cluster modifications. In the future, this could inform bacterial strain selection processes in the search for novel chemistry.

To conclude:

This thesis work demonstrates how genomic knowledge and chemical knowledge can be utilized in tandem for bacterial natural product discovery. Furthermore, it indicates that harnessing discovery methodologies tailored to our research goals can be a valuable strategy.

In conclusion, as a result of the thesis work presented above, a new method was developed for the strategic isolation of natural products using specific nitrogen NMR experiments to target genome-mined natural products. Additionally, novel natural products were isolated, and their structures were elucidated.

References

(1)   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.

(2)   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.

(3)   Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. Natural Products for Human Health: An Historical Overview of the Drug Discovery Approaches. *Nat. Prod. Res.* **2018**, *32* (16), 1926–1950. https://doi.org/10.1080/14786419.2017.1356838.

(4)   Saxena, S.; Raja, A. S. M. Natural Dyes: Sources, Chemistry, Application and Sustainability Issues. In *Roadmap to Sustainable Textiles and Clothing*; Muthu, S. S., Ed.; Textile Science and Clothing Technology; Springer Singapore: Singapore, 2014; pp 37–80. https://doi.org/10.1007/978-981-287-065-0_2.

(5)   Langenheim, J. *Plant Resins: Chemistry, Evolution, Ecology, Ethnobotany*; Timber Press: Portland, Oregon, 2003.

(6)   Burger, P.; Plainfossé, H.; Brochet, X.; Chemat, F.; Fernandez, X. Extraction of Natural Fragrance Ingredients: History Overview and Future Trends. *Chem. Biodivers.* **2019**, *16* (10). https://doi.org/10.1002/cbdv.201900424.

(7)   Desborough, M. J. R.; Keeling, D. M. The Aspirin Story - from Willow to Wonder Drug. *Br. J. Haematol.* **2017**, *177* (5), 674–683. https://doi.org/10.1111/bjh.14520.

(8)   Senator, H. SOME REMARKS ON SALICIN AND SALICYLIC ACID. *The Lancet* **1879**, *114* (2916), 79. https://doi.org/10.1016/S0140-6736(02)46379-9.

(9)   Moore, R.; Hernandez, D.; Valantine, H. Calcineurin Inhibitors and Post-Transplant Hyperlipidaemias: *Drug Saf.* **2001**, *24* (10), 755–766. https://doi.org/10.2165/00002018-200124100-00004.

(10)   E.A. Musson, R.; P.M. Smit, N. Regulatory Mechanisms of Calcineurin Phosphatase Activity. *Curr. Med. Chem.* **2011**, *18* (2), 301–315. https://doi.org/10.2174/092986711794088407.

(11)   Knight, V.; Sanglier, J.-J.; DiTullio, D.; Braccili, S.; Bonner, P.; Waters, J.; Hughes, D.; Zhang, L. Diversifying Microbial Natural Products for Drug Discovery. *Appl. Microbiol. Biotechnol.* **2003**, *62* (5–6), 446–458. https://doi.org/10.1007/s00253-003-1381-9.

(12)   Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic Discovery: History, Methods and Perspectives. *Int. J. Antimicrob. Agents* **2019**, *53* (4), 371–382. https://doi.org/10.1016/j.ijantimicag.2018.11.010.

(13)   Lewis, K. Recover the Lost Art of Drug Discovery. *Nature* **2012**, *485* (7399), 439–440. https://doi.org/10.1038/485439a.

(14)   Cragg, G. M.; Pezzuto, J. M. Natural Products as a Vital Source for the Discovery of Cancer Chemotherapeutic and Chemopreventive Agents. *Med. Princ. Pract.* **2016**, *25* (2), 41–59. https://doi.org/10.1159/000443404.

(15)   Chin, Y.-W.; Balunas, M. J.; Chai, H. B.; Kinghorn, A. D. Drug Discovery from Natural Sources. 15. https://doi.org/10.1007/BF02854894.

(16)   Cragg, G. M.; Newman, D. J. Medicinals for the Millennia. The Historical Record. *Ann. N. Y. Acad. Sci.* **2001**, *953a* (1 NEW VISTAS IN), 3–25. https://doi.org/10.1111/j.1749-6632.2001.tb11356.x.

(17)   Cohan, F. M. Bacterial Species and Speciation. *Syst. Biol.* **2001**, *50* (4), 513–524. https://doi.org/10.1080/10635150118398.

(18)   Andrews, J. H. Genetic Variation. In *Comparative Ecology of Microorganisms and Macroorganisms*; Andrews, J. H., Ed.; Springer: New York, NY, 2017; pp 25–68. https://doi.org/10.1007/978-1-4939-6897-8_2.

(19)   Albanese, D.; Donati, C. Strain Profiling and Epidemiology of Bacterial Species from Metagenomic Sequencing. *Nat. Commun.* **2017**, *8* (1), 2260. https://doi.org/10.1038/s41467-017-02209-5.

(20)   Lawson, P. A. The Phylum Actinobacteria. In *The Bifidobacteria and Related Organisms*; Elsevier, 2018; pp 1–8. https://doi.org/10.1016/B978-0-12-805060-6.00001-6.

(21)   Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete Salinispora Tropica. *Proc. Natl. Acad. Sci.* **2007**, *104* (25), 10376–10381. https://doi.org/10.1073/pnas.0700962104.

(22)   Robertsen, H. L.; Musiol-Kroll, E. M. Actinomycete-Derived Polyketides as a Source of Antibiotics and Lead Structures for the Development of New Antimicrobial Drugs. *Antibiotics* **2019**, *8* (4), 157. https://doi.org/10.3390/antibiotics8040157.

(23)   Berg, D.; Schedel, M.; Schmidt, R. R.; Ditgens, K.; Weyland, H. Naramycin B, an Antibiotic from Streptomyces Griseus Strain 587 with Herbicidal Properties-Fermentation, Isolation, and Identification. *Z. Für Naturforschung C* **1982**, *37* (11–12), 1100–1106. https://doi.org/10.1515/znc-1982-11-1210.

(24)   Hosaka, T.; Ohnishi-Kameyama, M.; Muramatsu, H.; Murakami, K.; Tsurumi, Y.; Kodani, S.; Yoshida, M.; Fujie, A.; Ochi, K. Antibacterial Discovery in Actinomycetes Strains with Mutations in RNA Polymerase or Ribosomal Protein S12. *Nat. Biotechnol.* **2009**, *27* (5), 462–464. https://doi.org/10.1038/nbt.1538.

(25)   Debono, M.; Barnhart, M.; Carrell, C. B.; Hoffmann, J. A.; Occolowitz, J. L.; Abbott, B. J.; Fukuda, D. S.; Hamill, R. L.; Biemann, K.; Herlihy, W. C. A21978C, a Complex of New Acidic Peptide Antibiotics. Isolation, Chemistry, and Mass Spectral Structure Elucidation. *J. Antibiot. (Tokyo)* **1987**, *40* (6), 761–777. https://doi.org/10.7164/antibiotics.40.761.

(26)   Bull, A. T.; Ward, A. C.; Goodfellow, M. Search and Discovery Strategies for Biotechnology: The Paradigm Shift. *Microbiol. Mol. Biol. Rev.* **2000**, *64* (3), 573–606. https://doi.org/10.1128/MMBR.64.3.573-606.2000.

(27)   Haefner, B. Drugs from the Deep: Marine Natural Products as Drug Candidates. *Drug Discov. Today* **2003**, *8* (12), 536–544. https://doi.org/10.1016/S1359-6446(03)02713-2.

(28)   Maldonado, L. A.; Fenical, W.; Jensen, P. R.; Kauffman, C. A.; Mincer, T. J.; Ward, A. C.; Bull, A. T.; Goodfellow, M. Salinispora Arenicola Gen. Nov., Sp. Nov. and Salinispora Tropica Sp. Nov., Obligate Marine Actinomycetes Belonging to the Family Micromonosporaceae. *Int. J. Syst. Evol. Microbiol.* **2005**, *55* (5), 1759–1766. https://doi.org/10.1099/ijs.0.63625-0.

(29)   Freel, K. C.; Edlund, A.; Jensen, P. R. Microdiversity and Evidence for High Dispersal Rates in the Marine Actinomycete "Salinispora Pacifica." *Environ. Microbiol.* **2012**, *14* (2), 480–493. https://doi.org/10.1111/j.1462-2920.2011.02641.x.

(30)   Stocker, R. Marine Microbes See a Sea of Gradients. *Science* **2012**, *338* (6107), 628–633. https://doi.org/10.1126/science.1208929.

(31)   Fenical, W. (Bill). Marine Microbial Natural Products: The Evolution of a New Field of Science. *J. Antibiot. (Tokyo)* **2020**, *73* (8), 481–487. https://doi.org/10.1038/s41429-020-0331-4.

(32)   Austin, B. Novel Pharmaceutical Compounds from Marine Bacteria. *J. Appl. Bacteriol.* **1989**, *67* (5), 461–470. https://doi.org/10.1111/j.1365-2672.1989.tb02517.x.

(33)   Mayer, R. J.; Davis, R. B.; Schiffer, C. A.; Berg, D. T.; Powell, B. L.; Schulman, P.; Omura, G. A.; Moore, J. O.; McIntyre, O. R.; Frei, E. Intensive Postremission Chemotherapy in Adults with Acute Myeloid Leukemia. *N. Engl. J. Med.* **1994**, *331* (14), 896–903. https://doi.org/10.1056/NEJM199410063311402.

(34)   Sagar, S.; Kaur, M.; Minneman, K. P. Antiviral Lead Compounds from Marine Sponges. *Mar. Drugs* **2010**, *8* (10), 2619–2638. https://doi.org/10.3390/md8102619.

(35)   Fenical, W.; Jensen, P. R.; Palladino, M. A.; Lam, K. S.; Lloyd, G. K.; Potts, B. C. Discovery and Development of the Anticancer Agent Salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* **2009**, *17* (6), 2175. https://doi.org/10.1016/j.bmc.2008.10.075.

(36)   Barbie, P.; Kazmaier, U. Total Synthesis of Cyclomarin A, a Marine Cycloheptapeptide with Anti-Tuberculosis and Anti-Malaria Activity. *Org. Lett.* **2016**, *18* (2), 204–207. https://doi.org/10.1021/acs.orglett.5b03292.

(37)   Renner, M. K.; Shen, Y.-C.; Cheng, X.-C.; Jensen, P. R.; Frankmoelle, W.; Kauffman, C. A.; Fenical, W.; Lobkovsky, E.; Clardy, J. Cyclomarins A−C, New Antiinflammatory Cyclic Peptides Produced by a Marine Bacterium (*Streptomyces* Sp.). *J. Am. Chem. Soc.* **1999**, *121* (49), 11273–11276. https://doi.org/10.1021/ja992482o.

(38)   *Handbook of Anticancer Drugs from Marine Origin*; Kim, S.-K., Ed.; Springer International Publishing: Cham, 2015. https://doi.org/10.1007/978-3-319-07145-9.

(39)   Jiménez, C. Marine Natural Products in Medicinal Chemistry. *ACS Med. Chem. Lett.* **2018**, *9* (10), 959–961. https://doi.org/10.1021/acsmedchemlett.8b00368.

(40)   Hanson, J. R. A Hundred Years in the Elucidation of the Structures of Natural Products. *Sci. Prog.* **2017**, *100* (1), 63–79. https://doi.org/10.3184/003685017X14876775256129.

(41)   Bevan, K.; Davies, J. S.; Hall, M. J.; Hassall, C. H.; Morton, R. B.; Phillips, D. A. S.; Ogihara, Y.; Thomas, W. A. The Monamycins, a New Family of Cyclodepsipeptide Antibiotics. *Experientia* **1970**, *26* (2), 122–123. https://doi.org/10.1007/BF01895528.

(42)   Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**, *2012*, 1–40. https://doi.org/10.1155/2012/282574.

(43)   Elyashberg, M. Identification and Structure Elucidation by NMR Spectroscopy. *TrAC Trends Anal. Chem.* **2015**, *69*, 88–97. https://doi.org/10.1016/j.trac.2015.02.014.

(44)   Breton, R. C.; Reynolds, W. F. Using NMR to Identify and Characterize Natural Products. *Nat. Prod. Rep.* **2013**, *30* (4), 501. https://doi.org/10.1039/c2np20104f.

(45)   Turner, D. L. Basic Two-Dimensional NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **1985**, *17*, 281–358. https://doi.org/10.1016/0079-6565(85)80011-X.

(46)   Kurita, K. L.; Linington, R. G. Connecting Phenotype and Chemotype: High-Content Discovery Strategies for Natural Products Research. *J. Nat. Prod.* **2015**, *78* (3), 587–596. https://doi.org/10.1021/acs.jnatprod.5b00017.

(47)   Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. Retrospective Analysis of Natural Products Provides Insights for Future Discovery Trends. *Proc. Natl. Acad. Sci.* **2017**, *114* (22), 5601–5606. https://doi.org/10.1073/pnas.1614680114.

(48)   Ventola, C. L. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *P T Peer-Rev. J. Formul. Manag.* **2015**, *40* (4), 277–283.

(49)   Anand, U.; Jacobo-Herrera, N.; Altemimi, A.; Lakhssassi, N. A Comprehensive Review on Medicinal Plants as Antimicrobial Therapeutics: Potential Avenues of Biocompatible Drug Discovery. *Metabolites* **2019**, *9* (11), 258. https://doi.org/10.3390/metabo9110258.

(50)   Ding, Y.; Sherman, D. H. The Role of Synthesis and Biosynthetic Logic. In *Comprehensive Natural Products II*; Elsevier, 2010; pp 559–579. https://doi.org/10.1016/B978-008045382-8.00052-6.

(51)   Grygorenko, O. O.; Volochnyuk, D. M.; Ryabukhin, S. V.; Judd, D. B. The Symbiotic Relationship Between Drug Discovery and Organic Chemistry. *Chem. – Eur. J.* **2020**, *26* (6), 1196–1237. https://doi.org/10.1002/chem.201903232.

(52)   Běhal, V. Enzymes of Secondary Metabolism in Microorganisms. *Trends Biochem. Sci.* **1986**, *11* (2), 88–91. https://doi.org/10.1016/0968-0004(86)90273-2.

(53)   Jackson, C. J.; Gillam, E. M. J.; Ollis, D. L. 9.20 Directed Evolution of Enzymes. 27.

(54)   Modolo, L. V.; Reichert, A. I.; Dixon, R. A. Introduction to the Different Classes of Biosynthetic Enzymes. In *Plant-derived Natural Products: Synthesis, Function, and Application*; Osbourn, A. E., Lanzotti, V., Eds.; Springer US: New York, NY, 2009; pp 143–163. https://doi.org/10.1007/978-0-387-85498-4_6.

(55)   Meinwald, J.; Eisner, T. Chemical Ecology in Retrospect and Prospect. *Proc. Natl. Acad. Sci.* **2008**, *105* (12), 4539–4540. https://doi.org/10.1073/pnas.0800649105.

(56)   Hartmann, T. The Lost Origin of Chemical Ecology in the Late 19th Century. *Proc. Natl. Acad. Sci.* **2008**, *105* (12), 4541–4546. https://doi.org/10.1073/pnas.0709231105.

(57)   Macintyre, L.; Zhang, T.; Viegelmann, C.; Martinez, I.; Cheng, C.; Dowdells, C.; Abdelmohsen, U.; Gernert, C.; Hentschel, U.; Edrada-Ebel, R. Metabolomic Tools for Secondary Metabolite Discovery from Marine Microbial Symbionts. *Mar. Drugs* **2014**, *12* (6), 3416–3448. https://doi.org/10.3390/md12063416.

(58)   Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2–3), 155–176. https://doi.org/10.1007/s10295-015-1723-5.

(59)   Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat. Rev. Drug Discov.* **2015**, *14* (2), 111–129. https://doi.org/10.1038/nrd4510.

(60)   Fewer, D. P.; Metsä-Ketelä, M. A Pharmaceutical Model for the Molecular Evolution of Microbial Natural Products. *FEBS J.* **2020**, *287* (7), 1429–1449. https://doi.org/10.1111/febs.15129.

(61)   Chevrette, M. G.; Currie, C. R. Emerging Evolutionary Paradigms in Antibiotic Discovery. *J. Ind. Microbiol. Biotechnol.* **2019**, *46* (3), 257–271. https://doi.org/10.1007/s10295-018-2085-6.

(62)   Soldatou, S.; Eldjarn, G. H.; Huerta-Uribe, A.; Rogers, S.; Duncan, K. R. Linking Biosynthetic and Chemical Space to Accelerate Microbial Secondary Metabolite Discovery. *FEMS Microbiol. Lett.* **2019**, *366* (13), fnz142. https://doi.org/10.1093/femsle/fnz142.

(63)   Scott, T. A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. *Nat. Rev. Chem.* **2019**, *3* (7), 404–425. https://doi.org/10.1038/s41570-019-0107-1.

(64)   Lind, A. L.; Wisecaver, J. H.; Smith, T. D.; Feng, X.; Calvo, A. M.; Rokas, A. Examining the Evolution of the Regulatory Circuit Controlling Secondary Metabolism

and Development in the Fungal Genus Aspergillus. *PLOS Genet.* **2015**, *11* (3), e1005096. https://doi.org/10.1371/journal.pgen.1005096.

(65)   Jensen, P. R.; Mafnas, C. Biogeography of the Marine Actinomycete Salinispora. *Environ. Microbiol.* **2006**, *8* (11), 1881–1888. https://doi.org/10.1111/j.1462-2920.2006.01093.x.

(66)   Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aguiñaga, N.; Chavarria, K. L.; Jensen, P. R. Diversity and Evolution of Secondary Metabolism in the Marine Actinomycete Genus Salinispora. *Proc. Natl. Acad. Sci.* **2014**, *111* (12), E1130–E1139. https://doi.org/10.1073/pnas.1324161111.

(67)   Letzel, A.-C.; Li, J.; Amos, G. C. A.; Millán-Aguiñaga, N.; Ginigini, J.; Abdelmohsen, U. R.; Gaudêncio, S. P.; Ziemert, N.; Moore, B. S.; Jensen, P. R. Genomic Insights into Specialized Metabolism in the Marine Actinomycete *Salinispora*: Genomic Insights into Specialized Metabolism. *Environ. Microbiol.* **2017**, *19* (9), 3660–3673. https://doi.org/10.1111/1462-2920.13867.

(68)   Floss, H. G.; Yu, T.-W. RifamycinMode of Action, Resistance, and Biosynthesis. *Chem. Rev.* **2005**, *105* (2), 621–632. https://doi.org/10.1021/cr030112j.

(69)   Weller, M. G. A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques. *Sensors* **2012**, *12* (7), 9181–9209. https://doi.org/10.3390/s120709181.

(70)   Renner, S.; van Otterlo, W. A. L.; Dominguez Seoane, M.; Möcklinghoff, S.; Hofmann, B.; Wetzel, S.; Schuffenhauer, A.; Ertl, P.; Oprea, T. I.; Steinhilber, D.; Brunsveld, L.; Rauh, D.; Waldmann, H. Bioactivity-Guided Mapping and Navigation of Chemical Space. *Nat. Chem. Biol.* **2009**, *5* (8), 585–592. https://doi.org/10.1038/nchembio.188.

(71)   Dudoit, A.; Mertz, C.; Chillet, M.; Cardinault, N.; Brat, P. Antifungal Activity of Brazilian Red Propolis Extract and Isolation of Bioactive Fractions by Thin-Layer Chromatography-Bioautography. *Food Chem.* **2020**, *327*, 127060. https://doi.org/10.1016/j.foodchem.2020.127060.

(72)   Broberg, A.; Menkis, A.; Vasiliauskas, R. Kutznerides 1−4, Depsipeptides from the Actinomycete Kutzneria Sp. 744 Inhabiting Mycorrhizal Roots of Picea Abies Seedlings. *J. Nat. Prod.* **2006**, *69* (1), 97–102. https://doi.org/10.1021/np050378g.

(73)   Schneider, O.; Simic, N.; Aachmann, F. L.; Rückert, C.; Kristiansen, K. A.; Kalinowski, J.; Jiang, Y.; Wang, L.; Jiang, C.-L.; Lale, R.; Zotchev, S. B. Genome Mining of Streptomyces Sp. YIM 130001 Isolated From Lichen Affords New Thiopeptide Antibiotic. *Front. Microbiol.* **2018**, *9*. https://doi.org/10.3389/fmicb.2018.03139.

(74)   Kersten, R. D.; Lane, A. L.; Nett, M.; Richter, T. K. S.; Duggan, B. M.; Dorrestein, P. C.; Moore, B. S. Bioactivity-Guided Genome Mining Reveals the Lomaiviticin Biosynthetic Gene Cluster in *Salinispora Tropica*. *ChemBioChem* **2013**, *14* (8), 955–962. https://doi.org/10.1002/cbic.201300147.

(75)   Wilson, B. A. P.; Thornburg, C. C.; Henrich, C. J.; Grkovic, T.; O'Keefe, B. R. Creating and Screening Natural Product Libraries. *Nat. Prod. Rep.* **2020**, *37* (7), 893–918. https://doi.org/10.1039/C9NP00068B.

(76)   Legerská, B.; Chmelová, D.; Ondrejovič, M.; Miertuš, S. The TLC-Bioautography as a Tool for Rapid Enzyme Inhibitors Detection - A Review. *Crit. Rev. Anal. Chem.* **2020**, 1–19. https://doi.org/10.1080/10408347.2020.1797467.

(77)   Sparks, T. C.; Wessels, F. J.; Lorsbach, B. A.; Nugent, B. M.; Watson, G. B. The New Age of Insecticide Discovery-the Crop Protection Industry and the Impact of Natural

Products. *Pestic. Biochem. Physiol.* **2019**, *161*, 12–22.
https://doi.org/10.1016/j.pestbp.2019.09.002.

(78)   Pharmazeutische Gesellschaft Der Deutschen Demokratischen Republik. Die Pharmazie
Jahrgang 30. 1975. **1975**. https://doi.org/10.24355/DBBS.084-201811091255-0.

(79)   Yang, J. Y.; Karr, J. R.; Watrous, J. D.; Dorrestein, P. C. Integrating '-Omics' and
Natural Product Discovery Platforms to Investigate Metabolic Exchange in Microbiomes.
Curr. Opin. Chem. Biol. **2011**, *15* (1), 79–87. https://doi.org/10.1016/j.cbpa.2010.10.025.

(80)   Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T.
S. Microbial Strain Prioritization Using Metabolomics Tools for the Discovery of Natural
Products. *Anal. Chem.* **2012**, *84* (10), 4277–4283. https://doi.org/10.1021/ac202623g.

(81)   Covington, B. C.; McLean, J. A.; Bachmann, B. O. Comparative Mass Spectrometry-
Based Metabolomics Strategies for the Investigation of Microbial Secondary
Metabolites. *Nat. Prod. Rep.* **2017**, *34* (1), 6–24. https://doi.org/10.1039/C6NP00048G.

(82)   Goering, A. W.; McClure, R. A.; Doroghazi, J. R.; Albright, J. C.; Haverland, N. A.;
Zhang, Y.; Ju, K.-S.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.
Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives
Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS
Cent. Sci.* **2016**, *2* (2), 99–108. https://doi.org/10.1021/acscentsci.5b00331.

(83)   Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.;
Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick,
W. H. Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A
Powerful Approach for Discovering Natural Products from Cyanobacteria. *J. Nat. Prod.*
2015, *78* (7), 1671–1682. https://doi.org/10.1021/acs.jnatprod.5b00301.

(84)   Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der
Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.;
Bandeira, N.; Dorrestein, P. C. Mass Spectral Molecular Networking of Living Microbial
Colonies. *Proc. Natl. Acad. Sci.* **2012**, *109* (26), E1743–E1752.
https://doi.org/10.1073/pnas.1203689109.

(85)   Liu, W.-T.; Ng, J.; Meluzzi, D.; Bandeira, N.; Gutierrez, M.; Simmons, T. L.; Schultz, A.
W.; Linington, R. G.; Moore, B. S.; Gerwick, W. H.; Pevzner, P. A.; Dorrestein, P. C.
Interpretation of Tandem Mass Spectra Obtained from Cyclic Nonribosomal Peptides.
Anal. Chem. **2009**, *81* (11), 4200–4209. https://doi.org/10.1021/ac900114t.

(86)   Townsend, C.; Furukawa, A.; Schwochert, J.; Pye, C. R.; Edmondson, Q.; Lokey, R. S.
CycLS: Accurate, Whole-Library Sequencing of Cyclic Peptides Using Tandem Mass
Spectrometry. *Bioorg. Med. Chem.* **2018**, *26* (6), 1232–1238.
https://doi.org/10.1016/j.bmc.2018.01.027.

(87)   Novák, J.; Lemr, K.; Schug, K. A.; Havlíček, V. CycloBranch: De Novo Sequencing of
Nonribosomal Peptides from Accurate Product Ion Mass Spectra. *J. Am. Soc. Mass
Spectrom.* **2015**, *26* (10), 1780–1786. https://doi.org/10.1007/s13361-015-1211-1.

(88)   Schroeder, F. C.; Gibson, D. M.; Churchill, A. C. L.; Sojikul, P.; Wursthorn, E. J.;
Krasnoff, S. B.; Clardy, J. Differential Analysis of 2D NMR Spectra: New Natural
Products from a Pilot-Scale Fungal Extract Library. *Angew. Chem. Int. Ed.* **2007**, *46* (6),
901–904. https://doi.org/10.1002/anie.200603821.

(89)   Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P.
W.; Schroeder, F. C. A Shortcut to Identifying Small Molecule Signals That Regulate

Behavior and Development in Caenorhabditis Elegans. *Proc. Natl. Acad. Sci.* **2009**, *106* (19), 7708–7713. https://doi.org/10.1073/pnas.0811918106.

(90)   Bingol, K.; Brüschweiler, R. Multidimensional Approaches to NMR-Based Metabolomics. *Anal. Chem.* **2014**, *86* (1), 47–57. https://doi.org/10.1021/ac403520j.

(91)   Miao, Z.; Jin, M.; Liu, X.; Guo, W.; Jin, X.; Liu, H.; Wang, Y. The Application of HPLC and Microprobe NMR Spectroscopy in the Identification of Metabolites in Complex Biological Matrices. *Anal. Bioanal. Chem.* **2015**, *407* (12), 3405–3416. https://doi.org/10.1007/s00216-015-8556-y.

(92)   Deyrup, S. T.; Eckman, L. E.; McCarthy, P. H.; Smedley, S. R.; Meinwald, J.; Schroeder, F. C. 2D NMR-Spectroscopic Screening Reveals Polyketides in Ladybugs. *Proc. Natl. Acad. Sci.* **2011**, *108* (24), 9753–9758. https://doi.org/10.1073/pnas.1107020108.

(93)   Nardella, F.; Margueritte, L.; Lamure, B.; Viéville, J. M. P.; Bourjot, M. Targeted Discovery of Bioactive Natural Products Using a Pharmacophoric Deconvolution Strategy: Proof of Principle with Eleganolone from Bifurcaria Bifurcata R. Ross. *Phytochem. Lett.* **2018**, *26*, 138–142. https://doi.org/10.1016/j.phytol.2018.05.034.

(94)   Gallé, J.-B.; Attioua, B.; Kaiser, M.; Rusig, A.-M.; Lobstein, A.; Vonthron-Sénécheau, C. Eleganolone, a Diterpene from the French Marine Alga Bifurcaria Bifurcata Inhibits Growth of the Human Pathogens Trypanosoma Brucei and Plasmodium Falciparum. *Mar. Drugs* **2013**, *11* (12), 599–610. https://doi.org/10.3390/md11030599.

(95)   Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. The Identification of Bacillaene, the Product of the PksX Megacomplex in Bacillus Subtilis. *Proc. Natl. Acad. Sci.* **2007**, *104* (5), 1506–1509. https://doi.org/10.1073/pnas.0610503104.

(96)   Halabalaki, M.; Vougogiannopoulou, K.; Mikros, E.; Skaltsounis, A. L. Recent Advances and New Strategies in the NMR-Based Identification of Natural Products. *Curr. Opin. Biotechnol.* **2014**, *25*, 1–7. https://doi.org/10.1016/j.copbio.2013.08.005.

(97)   Grkovic, T.; Pouwer, R. H.; Vial, M.-L.; Gambini, L.; Noël, A.; Hooper, J. N. A.; Wood, S. A.; Mellick, G. D.; Quinn, R. J. NMR Fingerprints of the Drug-like Natural-Product Space Identify Iotrochotazine A: A Chemical Probe to Study Parkinson's Disease. *Angew. Chem.* **2014**, *126* (24), 6184–6188. https://doi.org/10.1002/ange.201402239.

(98)   Yang, Z.; Wu, Y.; Zhou, H.; Cao, X.; Jiang, X.; Wang, K.; Wu, S. A Novel Strategy for Screening New Natural Products by a Combination of Reversed-Phase Liquid Chromatography Fractionation and ^{13}C NMR Pattern Recognition: The Discovery of New Anti-Cancer Flavone Dimers from Dysosma Versipellis (Hance). *RSC Adv.* **2015**, *5* (95), 77553–77564. https://doi.org/10.1039/C5RA13756J.

(99)   Wolff, T.; Santos, P. F. P.; Valente, L. M. M.; Magalhães, A.; Tinoco, L. W.; Pereira, R. C. A.; Guimarães, E. F. Piperamides from *Piper Ottonoides* by NMR and GC-MS Based Mixture Analysis. *J. Braz. Chem. Soc.* **2015**. https://doi.org/10.5935/0103-5053.20150226.

(100)   Shin, D.; Byun, W. S.; Moon, K.; Kwon, Y.; Bae, M.; Um, S.; Lee, S. K.; Oh, D.-C. Coculture of Marine Streptomyces Sp. With Bacillus Sp. Produces a New Piperazic Acid-Bearing Cyclic Peptide. *Front. Chem.* **2018**, *6*, 498. https://doi.org/10.3389/fchem.2018.00498.

(101)   Gross, H.; Stockwell, V. O.; Henkels, M. D.; Nowak-Thompson, B.; Loper, J. E.; Gerwick, W. H. The Genomisotopic Approach: A Systematic Method to Isolate Products

of Orphan Biosynthetic Gene Clusters. *Chem. Biol.* **2007**, *14* (1), 53–63. https://doi.org/10.1016/j.chembiol.2006.11.007.

(102) Cioni, J. P.; Doroghazi, J. R.; Ju, K.-S.; Yu, X.; Evans, B. S.; Lee, J.; Metcalf, W. W. Cyanohydrin Phosphonate Natural Product from *Streptomyces Regensis*. *J. Nat. Prod.* **2014**, *77* (2), 243–249. https://doi.org/10.1021/np400722m.

(103) Walsh, C. T.; Tang, Y. *Natural Product Biosynthesis*; The Royal Society of Chemistry: London, UK, 2017.

(104) Pott, D. M.; Osorio, S.; Vallarino, J. G. From Central to Specialized Metabolism: An Overview of Some Secondary Compounds Derived From the Primary Metabolism for Their Role in Conferring Nutritional and Organoleptic Characteristics to Fruit. *Front. Plant Sci.* **2019**, *10*. https://doi.org/10.3389/fpls.2019.00835.

(105) Ruchika; Naik, J.; Pandey, A. Chapter 15 - Synthetic Metabolism and Its Significance in Agriculture. In *Current Developments in Biotechnology and Bioengineering*; Singh, S. P., Pandey, A., Du, G., Kumar, S., Eds.; Elsevier, 2019; pp 365–391. https://doi.org/10.1016/B978-0-444-64085-7.00015-0.

(106) Dewick, P. M. *Medicinal Natural Products*; John Wiley & Sons, Ltd: Chichester, UK, 2001. https://doi.org/10.1002/0470846275.

(107) Chevrette, M. G.; Gutiérrez-García, K.; Selem-Mojica, N.; Aguilar-Martínez, C.; Yañez-Olvera, A.; Ramos-Aboites, H. E.; Hoskisson, P. A.; Barona-Gómez, F. Evolutionary Dynamics of Natural Product Biosynthesis in Bacteria. *Nat. Prod. Rep.* **2020**, *37* (4), 566–599. https://doi.org/10.1039/C9NP00048H.

(108) Banerji, A.; Jahne, M.; Herrmann, M.; Brinkman, N.; Keely, S. Bringing Community Ecology to Bear on the Issue of Antimicrobial Resistance. *Front. Microbiol.* **2019**, *10*. https://doi.org/10.3389/fmicb.2019.02626.

(109) Malik, V. S. Microbial Secondary Metabolism. *Trends Biochem. Sci.* **1980**, *5* (3), 68–72. https://doi.org/10.1016/0968-0004(80)90071-7.

(110) Vining, L. C. Secondary Metabolism, Inventive Evolution and Biochemical Diversity — a Review. *Gene* **1992**, *115* (1–2), 135–140. https://doi.org/10.1016/0378-1119(92)90551-Y.

(111) Osbourn, A. Secondary Metabolic Gene Clusters: Evolutionary Toolkits for Chemical Innovation. *Trends Genet.* **2010**, *26* (10), 449–457. https://doi.org/10.1016/j.tig.2010.07.001.

(112) Hodgson, D. A. Primary Metabolism and Its Control in Streptomycetes: A Most Unusual Group of Bacteria. In *Advances in Microbial Physiology*; Elsevier, 2000; Vol. 42, pp 47–238. https://doi.org/10.1016/S0065-2911(00)42003-5.

(113) Padilla, G.; Hindle, Z.; Callis, R.; Corner, A.; Ludovice, M.; Liras, P.; Baumberg, S. The Relationship Between Primary and Secondary Metabolism in Streptomycetes. In *Genetics and Product Formation in Streptomyces*; Baumberg, S., Krügel, H., Noack, D., Eds.; Federation of European Microbiological Societies Symposium Series; Springer US: Boston, MA, 1991; pp 35–45. https://doi.org/10.1007/978-1-4684-5922-7_6.

(114) Osbourn, A. Gene Clusters for Secondary Metabolic Pathways: An Emerging Theme in Plant Biology: Figure 1. *Plant Physiol.* **2010**, *154* (2), 531–535. https://doi.org/10.1104/pp.110.161315.

(115) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of

Nonmodular Enzymes. *Proc. Natl. Acad. Sci.* **2014**, *111* (25), 9259–9264. https://doi.org/10.1073/pnas.1401734111.

(116) Walsh, C. T. Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **2004**, *303* (5665), 1805–1810. https://doi.org/10.1126/science.1094318.

(117) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed.* **2009**, *48* (26), 4688–4716. https://doi.org/10.1002/anie.200806121.

(118) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **2009**, *26* (1), 90–114. https://doi.org/10.1039/B801658P.

(119) Katsuyama, Y.; Ohnishi, Y. Type III Polyketide Synthases in Microorganisms. In *Methods in Enzymology*; Elsevier, 2012; Vol. 515, pp 359–377. https://doi.org/10.1016/B978-0-12-394290-6.00017-3.

(120) Finking, R.; Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **2004**, *58* (1), 453–488. https://doi.org/10.1146/annurev.micro.58.030603.123615.

(121) Marahiel, M. A. A Structural Model for Multimodular NRPS Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 136–140. https://doi.org/10.1039/C5NP00082C.

(122) Du, L.; Sánchez, C.; Shen, B. Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. *Metab. Eng.* **2001**, *3* (1), 78–95. https://doi.org/10.1006/mben.2000.0171.

(123) Miyanaga, A.; Kudo, F.; Eguchi, T. Protein–Protein Interactions in Polyketide Synthase–Nonribosomal Peptide Synthetase Hybrid Assembly Lines. *Nat. Prod. Rep.* **2018**, *35* (11), 1185–1209. https://doi.org/10.1039/C8NP00022K.

(124) Fisch, K. M. Biosynthesis of Natural Products by Microbial Iterative Hybrid PKS–NRPS. *RSC Adv.* **2013**, *3* (40), 18228. https://doi.org/10.1039/c3ra42661k.

(125) Lin, S.; Huang, T.; Shen, B. Tailoring Enzymes Acting on Carrier Protein-Tethered Substrates in Natural Product Biosynthesis. In *Methods in Enzymology*; Elsevier, 2012; Vol. 516, pp 321–343. https://doi.org/10.1016/B978-0-12-394291-3.00008-3.

(126) Olano, C.; Méndez, C.; Salas, J. A. Post-PKS Tailoring Steps in Natural Product-Producing Actinomycetes from the Perspective of Combinatorial Biosynthesis. *Nat. Prod. Rep.* **2010**, *27* (4), 571. https://doi.org/10.1039/b911956f.

(127) Wilson, D. J.; Patton, S.; Florova, G.; Hale, V.; Reynolds, K. A. The Shikimic Acid Pathway and Polyketide Biosynthesis. *J. Ind. Microbiol. Biotechnol.* **1998**, *20* (5), 299–303. https://doi.org/10.1038/sj.jim.2900527.

(128) Dewick, P. M. The Biosynthesis of Shikimate Metabolites. *Nat. Prod. Rep.* **1995**, *12* (2), 101. https://doi.org/10.1039/np9951200101.

(129) Ganem, B. From Glucose to Aromatics: Recent Developments in Natural Products of the Shikimic Acid Pathway. *Tetrahedron* **1978**, *34* (23), 3353–3383. https://doi.org/10.1016/0040-4020(78)80222-1.

(130) Park, S. R.; Tripathi, A.; Wu, J.; Schultz, P. J.; Yim, I.; McQuade, T. J.; Yu, F.; Arevang, C.-J.; Mensah, A. Y.; Tamayo-Castillo, G.; Xi, C.; Sherman, D. H. Discovery of Cahuitamycins as Biofilm Inhibitors Derived from a Convergent Biosynthetic Pathway. *Nat. Commun.* **2016**, *7*, 10710. https://doi.org/10.1038/ncomms10710.

(131) Chen, S.; von Bamberg, D.; Hale, V.; Breuer, M.; Hardt, B.; Muller, R.; Floss, H. G.; Reynolds, K. A.; Leistner, E. Biosynthesis of Ansatrienin (Mycotrienin) and

Naphthomycin. Identification and Analysis of Two Separate Biosynthetic Gene Clusters in Streptomyces Collinus Tu 1892. *Eur. J. Biochem.* **1999**, *261* (1), 98–107. https://doi.org/10.1046/j.1432-1327.1999.00244.x.

(132) Floss, H. G.; Yu, T.-W.; Arakawa, K. The Biosynthesis of 3-Amino-5-Hydroxybenzoic Acid (AHBA), the Precursor of MC7N Units in Ansamycin and Mitomycin Antibiotics: A Review. *J. Antibiot. (Tokyo)* **2011**, *64* (1), 35–44. https://doi.org/10.1038/ja.2010.139.

(133) Werner, I.; Bacher, A.; Eisenreich, W. Retrobiosynthetic NMR Studies with ^{13}C-Labeled Glucose: FORMATION OF GALLIC ACID IN PLANTS AND FUNGI. *J. Biol. Chem.* **1997**, *272* (41), 25474–25482. https://doi.org/10.1074/jbc.272.41.25474.

(134) Eguchi, R.; Ono, N.; Hirai Morita, A.; Katsuragi, T.; Nakamura, S.; Huang, M.; Altaf-Ul-Amin, Md.; Kanaya, S. Classification of Alkaloids According to the Starting Substances of Their Biosynthetic Pathways Using Graph Convolutional Neural Networks. *BMC Bioinformatics* **2019**, *20* (1), 380. https://doi.org/10.1186/s12859-019-2963-6.

(135) O'Connor, S. E. Alkaloids. In *Comprehensive Natural Products II*; Elsevier, 2010; pp 977–1007. https://doi.org/10.1016/B978-008045382-8.00013-7.

(136) Herbert, R. B. The Biosynthesis of Plant Alkaloids and Nitrogenous Microbial Metabolites. *Nat. Prod. Rep.* **1995**, *12* (5), 445–464. https://doi.org/10.1039/NP9951200445.

(137) K, M. Solanum Alkaloids and Their Pharmaceutical Roles: A Review. *J. Anal. Pharm. Res.* **2015**, *3* (6). https://doi.org/10.15406/japlr.2016.03.00075.

(138) Prebble, D. W.; Holland, D. C.; Robertson, L. P.; Avery, V. M.; Carroll, A. R. Citronamine A, an Antiplasmodial Isoquinoline Alkaloid from the Australian Marine Sponge *Citronia Astra*. *Org. Lett.* **2020**, acs.orglett.0c03633. https://doi.org/10.1021/acs.orglett.0c03633.

(139) Mo, S.; Sydor, P. K.; Corre, C.; Alhamadsheh, M. M.; Stanley, A. E.; Haynes, S. W.; Song, L.; Reynolds, K. A.; Challis, G. L. Elucidation of the Streptomyces Coelicolor Pathway to 2-Undecylpyrrole, a Key Intermediate in Undecylprodiginine and Streptorubin B Biosynthesis. *Chem. Biol.* **2008**, *15* (2), 137–148. https://doi.org/10.1016/j.chembiol.2007.11.015.

(140) Jacob, F.; Monod, J. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J. Mol. Biol.* **1961**, *3* (3), 318–356. https://doi.org/10.1016/S0022-2836(61)80072-7.

(141) Molnár, I. Secretory Production of Homologous and Heterologous Proteins by Recombinant Streptomyces: What Has Been Accomplished? *Bioprocess Technol.* **1994**, *19*, 81–104.

(142) Price, M. N.; Arkin, A. P.; Alm, E. J. The Life-Cycle of Operons. *PLoS Genet.* **2006**, *2* (6), e96. https://doi.org/10.1371/journal.pgen.0020096.

(143) Rocha, E. P. C. The Organization of the Bacterial Genome. *Annu. Rev. Genet.* **2008**, *42* (1), 211–233. https://doi.org/10.1146/annurev.genet.42.110807.091653.

(144) Medvedev, Z. A. Molecular-Genetic Mechanisms of Morphogenesis and Intracellular Biochemical Differentiation of Monochromosomal Bacterial Cells. In *Molecular-Genetic Mechanisms of Development*; Medvedev, Z. A., Ed.; Springer US: Boston, MA, 1970; pp 79–125. https://doi.org/10.1007/978-1-4684-1764-7_3.

(145) Gerlt, J. A.; Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu. Rev. Biochem.* **2001**, *70* (1), 209–246. https://doi.org/10.1146/annurev.biochem.70.1.209.

(146) Land, M.; Hauser, L.; Jun, S.-R.; Nookaew, I.; Leuze, M. R.; Ahn, T.-H.; Karpinets, T.; Lund, O.; Kora, G.; Wassenaar, T.; Poudel, S.; Ussery, D. W. Insights from 20 Years of Bacterial Genome Sequencing. *Funct. Integr. Genomics* **2015**, *15* (2), 141–161. https://doi.org/10.1007/s10142-015-0433-4.

(147) Serin, E. A. R.; Nijveen, H.; Hilhorst, H. W. M.; Ligterink, W. Learning from Co-Expression Networks: Possibilities and Challenges. *Front. Plant Sci.* **2016**, *7*. https://doi.org/10.3389/fpls.2016.00444.

(148) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25* (1), 25–29. https://doi.org/10.1038/75556.

(149) Falkowski, P. G.; Fenchel, T.; Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **2008**, *320* (5879), 1034–1039. https://doi.org/10.1126/science.1153213.

(150) Koonin, E. V.; Aravind, L.; Kondrashov, A. S. The Impact of Comparative Genomics on Our Understanding of Evolution. *Cell* **2000**, *101* (6), 573–576. https://doi.org/10.1016/S0092-8674(00)80867-3.

(151) Hawley, A. K.; Torres-Beltrán, M.; Zaikova, E.; Walsh, D. A.; Mueller, A.; Scofield, M.; Kheirandish, S.; Payne, C.; Pakhomova, L.; Bhatia, M.; Shevchuk, O.; Gies, E. A.; Fairley, D.; Malfatti, S. A.; Norbeck, A. D.; Brewer, H. M.; Pasa-Tolic, L.; del Rio, T. G.; Suttle, C. A.; Tringe, S.; Hallam, S. J. A Compendium of Multi-Omic Sequence Information from the Saanich Inlet Water Column. *Sci. Data* **2017**, *4* (1), 170160. https://doi.org/10.1038/sdata.2017.160.

(152) Cuadrat, R. R. C.; Ionescu, D.; Dávila, A. M. R.; Grossart, H.-P. Recovering Genomics Clusters of Secondary Metabolites from Lakes Using Genome-Resolved Metagenomics. *Front. Microbiol.* **2018**, *9*, 251. https://doi.org/10.3389/fmicb.2018.00251.

(153) Winter, J. M.; Behnken, S.; Hertweck, C. Genomics-Inspired Discovery of Natural Products. *Curr. Opin. Chem. Biol.* **2011**, *15* (1), 22–31. https://doi.org/10.1016/j.cbpa.2010.10.020.

(154) Goodwin, S.; McPherson, J. D.; McCombie, W. R. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nat. Rev. Genet.* **2016**, *17* (6), 333–351. https://doi.org/10.1038/nrg.2016.49.

(155) Metzker, M. L. Sequencing Technologies — the next Generation. *Nat. Rev. Genet.* **2010**, *11* (1), 31–46. https://doi.org/10.1038/nrg2626.

(156) van Dijk, E. L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34* (9), 666–681. https://doi.org/10.1016/j.tig.2018.05.008.

(157) Haft, D. H.; DiCuccio, M.; Badretdin, A.; Brover, V.; Chetvernin, V.; O'Neill, K.; Li, W.; Chitsaz, F.; Derbyshire, M. K.; Gonzales, N. R.; Gwadz, M.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Yamashita, R. A.; Zheng, C.; Thibaud-Nissen, F.; Geer, L. Y.; Marchler-Bauer, A.; Pruitt, K. D. RefSeq: An Update on Prokaryotic Genome Annotation and Curation. *Nucleic Acids Res.* **2018**, *46* (D1), D851–D860. https://doi.org/10.1093/nar/gkx1068.

(158)   Pearson, W. R. An Introduction to Sequence Similarity ("homology") Searching. *Curr. Protoc. Bioinforma.* **2013**, *Chapter 3*, Unit3.1. https://doi.org/10.1002/0471250953.bi0301s42.

(159)   Medema, M. H.; Takano, E.; Breitling, R. Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30* (5), 1218–1223. https://doi.org/10.1093/molbev/mst025.

(160)   Bhagwat, M.; Aravind, L. *PSI-BLAST Tutorial*; Humana Press, 2007.

(161)   Ziemert, N.; Alanjary, M.; Weber, T. The Evolution of Genome Mining in Microbes – a Review. *Nat. Prod. Rep.* **2016**, *33* (8), 988–1005. https://doi.org/10.1039/C6NP00025H.

(162)   Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106* (8), 3468–3496. https://doi.org/10.1021/cr0503097.

(163)   Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H. AntiSMASH 4.0-Improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41. https://doi.org/10.1093/nar/gkx319.

(164)   Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39* (suppl_2), W362–W367. https://doi.org/10.1093/nar/gkr323.

(165)   Bachmann, B. O.; Ravel, J. Chapter 8 Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. In *Methods in Enzymology*; Elsevier, 2009; Vol. 458, pp 181–217. https://doi.org/10.1016/S0076-6879(09)04808-3.

(166)   Goering, A. W.; McClure, R. A.; Doroghazi, J. R.; Albright, J. C.; Haverland, N. A.; Zhang, Y.; Ju, K.-S.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L. Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS Cent. Sci.* **2016**, *2* (2), 99–108. https://doi.org/10.1021/acscentsci.5b00331.

(167)   Liu, W.; Sun, F.; Hu, Y. Genome Mining-Mediated Discovery of a New Avermipeptin Analogue in Streptomyces Actuosus ATCC 25421. *ChemistryOpen* **2018**, *7* (7), 558–561. https://doi.org/10.1002/open.201800130.

(168)   Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the Mosaic Structure of Trans-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. *Nat. Biotechnol.* **2008**, *26* (2), 225–233. https://doi.org/10.1038/nbt1379.

(169)   Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; Moreno, P.; Piel, J. Automated Structure Prediction of Trans-Acyltransferase Polyketide Synthase Products. *Nat. Chem. Biol.* **2019**, *15* (8), 813–821. https://doi.org/10.1038/s41589-019-0313-7.

(170)   Scott, T. A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. *Nat. Rev. Chem.* **2019**, *3* (7), 404–425. https://doi.org/10.1038/s41570-019-0107-1.

(171)   Le Goff, G.; Ouazzani, J. Natural Hydrazine-Containing Compounds: Biosynthesis, Isolation, Biological Activities and Synthesis. *Bioorg. Med. Chem.* **2014**, *22* (23), 6529–6544. https://doi.org/10.1016/j.bmc.2014.10.011.

(172) Ciufolini, M.; Xi, N. Synthesis, Chemistry and Conformational Properties of Piperazic Acids. *Chem. Soc. Rev.* **1998**, *27* (6), 437. https://doi.org/10.1039/a827437z.

(173) Morgan, K. D.; Andersen, R. J.; Ryan, K. S. Piperazic Acid-Containing Natural Products: Structures and Biosynthesis. *Nat. Prod. Rep.* **2019**, *36* (12), 1628–1653. https://doi.org/10.1039/C8NP00076J.

(174) Kunihiro, S.; Kaneda, M. Glomecidin, a Novel Antifungal Cyclic Tetrapeptide Produced by Streptomyces Lavendulae H698 SY2. *J. Antibiot. (Tokyo)* **2003**, *56* (1), 30–33. https://doi.org/10.7164/antibiotics.56.30.

(175) Hosaka, T.; Ohnishi-Kameyama, M.; Muramatsu, H.; Murakami, K.; Tsurumi, Y.; Kodani, S.; Yoshida, M.; Fujie, A.; Ochi, K. Antibacterial Discovery in Actinomycetes Strains with Mutations in RNA Polymerase or Ribosomal Protein S12. *Nat. Biotechnol.* **2009**, *27* (5), 462–464. https://doi.org/10.1038/nbt.1538.

(176) Yoshida, M.; Sekioka, N.; Izumikawa, M.; Kozone, I.; Takagi, M.; Shin-ya, K.; Doi, T. Total Synthesis and Structure Elucidation of JBIR-39: A Linear Hexapeptide Possessing Piperazic Acid and γ-Hydroxypiperazic Acid Residues. *Chem. - Eur. J.* **2015**, *21* (7), 3031–3041. https://doi.org/10.1002/chem.201406020.

(177) Nakagawa, M.; Hayakawa, Y.; Furihata, K.; Seto, H. Structural Studies on New Depsipeptide Antibiotics, Variapeptin and Citropeptin. *J. Antibiot. (Tokyo)* **1990**, *43* (5), 477–484.

(178) Sakai, Y.; Yoshida, T.; Tsujita, T.; Ochiai, K.; Agatsuma, T.; Saitoh, Y.; Tanaka, F.; Akiyama, T.; AKiNAGA, S.; Mizukami, T. GE3, a Novel Hexadepsipeptide Antitumor Antibiotic, Produced by Streptomyces Sp.I. Taxonomy, Production, Isolation, Physico-Chemical Properties, and Biological Activities. *J. Antibiot. (Tokyo)* **1997**, *50* (8), 659–664. https://doi.org/10.7164/antibiotics.50.659.

(179) Maskey, R. P.; Fotso, S.; Sevvana, M.; Usón, I.; Grün-Wollny, I.; Laatsch, H. Kettapeptin: Isolation, Structure Elucidation and Activity of a New Hexadepsipeptide Antibiotic from a Terrestrial Streptomyces Sp. *J. Antibiot. (Tokyo)* **2006**, *59* (5), 309–314. https://doi.org/10.1038/ja.2006.44.

(180) Zhang, L.; Wang, Y.; Huang, W.; Wei, Y.; Jiang, Z.; Kong, L.; Wu, A.; Hu, Z.; Huang, H.; Xu, Q.; Li, L.; Deng, X. Biosynthesis and Chemical Diversification of Verucopeptin Leads to Structural and Functional Versatility. *Org. Lett.* **2020**, *22* (11), 4366–4371. https://doi.org/10.1021/acs.orglett.0c01387.

(181) Umezawa, K.; Ikeda, Y.; Kawase, O.; Naganawa, H.; Kondo, S. Biosynthesis of Polyoxypeptin A: Novel Amino Acid 3-Hydroxy-3-Methylproline Derived from Isoleucine. *J. Chem. Soc. Perkin 1* **2001**, No. 13, 1550–1553. https://doi.org/10.1039/b101942m.

(182) Wilson, M. C.; Moore, B. S. Beyond Ethylmalonyl-CoA: The Functional Role of Crotonyl-CoAcarboxylase/Reductase Homologs in Expanding Polyketide Diversity. *Nat Prod Rep* **2012**, *29* (1), 72–86. https://doi.org/10.1039/C1NP00082A.

(183) Hu, Y.; Qi, Y.; Stumpf, S. D.; D'Alessandro, J. M.; Blodgett, J. A. V. Bioinformatic and Functional Evaluation of Actinobacterial Piperazate Metabolism. *ACS Chem. Biol.* **2019**, *14* (4), 696–703. https://doi.org/10.1021/acschembio.8b01086.

(184) Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. Cloning and Characterization of the Biosynthetic Gene Cluster for Kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (42), 16498–16503.

(185) Heemstra, J. R.; Walsh, C. T. Tandem Action of the O2- and FADH2-Dependent Halogenases KtzQ and KtzR Produce 6,7-Dichlorotryptophan for Kutzneride Assembly. *J. Am. Chem. Soc.* **2008**, *130* (43), 14024–14025. https://doi.org/10.1021/ja806467a.

(186) Jiang, W.; Heemstra, J. R.; Forseth, R. R.; Neumann, C. S.; Manaviazar, S.; Schroeder, F. C.; Hale, K. J.; Walsh, C. T. Biosynthetic Chlorination of the Piperazate Residue in Kutzneride Biosynthesis by KthP. *Biochemistry* **2011**, *50* (27), 6063–6072. https://doi.org/10.1021/bi200656k.

(187) Neumann, C. S.; Walsh, C. T. Biosynthesis of (−)-(1 *S, 2 R*)-Allocoronamic Acyl Thioester by an Fe ^{II} -Dependent Halogenase and a Cyclopropane-Forming Flavoprotein. *J. Am. Chem. Soc.* **2008**, *130* (43), 14022–14023. https://doi.org/10.1021/ja8064667.

(188) Strieker, M.; Nolan, E. M.; Walsh, C. T.; Marahiel, M. A. Stereospecific Synthesis of Threo- and Erythro-Beta-Hydroxyglutamic Acid during Kutzneride Biosynthesis. *J. Am. Chem. Soc.* **2009**, *131* (37), 13523–13530. https://doi.org/10.1021/ja9054417.

(189) Zolova, O. E.; Garneau-Tsodikova, S. KtzJ-Dependent Serine Activation and O-Methylation by KtzH for Kutznerides Biosynthesis. *J. Antibiot. (Tokyo)* **2014**, *67* (1), 59–64. https://doi.org/10.1038/ja.2013.98.

(190) Guo, Z.; Li, P.; Chen, G.; Li, C.; Cao, Z.; Zhang, Y.; Ren, J.; Xiang, H.; Lin, S.; Ju, J.; Chen, Y. Design and Biosynthesis of Dimeric Alboflavusins with Biaryl Linkages via Regiospecific C-C Bond Coupling. *J. Am. Chem. Soc.* **2018**. https://doi.org/10.1021/jacs.8b10136.

(191) Du, Y.-L.; Dalisay, D. S.; Andersen, R. J.; Ryan, K. S. N-Carbamoylation of 2,4-Diaminobutyrate Reroutes the Outcome in Padanamide Biosynthesis. *Chem. Biol.* **2013**, *20* (8), 1002–1011. https://doi.org/10.1016/j.chembiol.2013.06.013.

(192) Du, Y.; Wang, Y.; Huang, T.; Tao, M.; Deng, Z.; Lin, S. Identification and Characterization of the Biosynthetic Gene Cluster of Polyoxypeptin A, a Potent Apoptosis Inducer. *BMC Microbiol.* **2014**, *14*, 30. https://doi.org/10.1186/1471-2180-14-30.

(193) Zhao, H.; Wang, L.; Wan, D.; Qi, J.; Gong, R.; Deng, Z.; Chen, W. Characterization of the Aurantimycin Biosynthetic Gene Cluster and Enhancing Its Production by Manipulating Two Pathway-Specific Activators in Streptomyces Aurantiacus JA 4570. *Microb. Cell Factories* **2016**, *15* (1). https://doi.org/10.1186/s12934-016-0559-7.

(194) Oh, D.-C.; Poulsen, M.; Currie, C. R.; Clardy, J. Dentigerumycin: A Bacterial Mediator of an Ant-Fungus Symbiosis. *Nat. Chem. Biol.* **2009**, *5* (6), 391–393. https://doi.org/10.1038/nchembio.159.

(195) Shin, D.; Byun, W. S.; Moon, K.; Kwon, Y.; Bae, M.; Um, S.; Lee, S. K.; Oh, D.-C. Coculture of Marine Streptomyces Sp. With Bacillus Sp. Produces a New Piperazic Acid-Bearing Cyclic Peptide. *Front. Chem.* **2018**, *6*, 498. https://doi.org/10.3389/fchem.2018.00498.

(196) Liu, J.; Wang, B.; Li, H.; Xie, Y.; Li, Q.; Qin, X.; Zhang, X.; Ju, J. Biosynthesis of the Anti-Infective Marformycins Featuring Pre-NRPS Assembly Line *N* -Formylation and *O* -Methylation and Post-Assembly Line *C* -Hydroxylation Chemistries. *Org. Lett.* **2015**, *17* (6), 1509–1512. https://doi.org/10.1021/acs.orglett.5b00389.

(197) Li, Q.; Qin, X.; Liu, J.; Gui, C.; Wang, B.; Li, J.; Ju, J. Deciphering the Biosynthetic Origin of L - *Allo* -Isoleucine. *J. Am. Chem. Soc.* **2016**, *138* (1), 408–415. https://doi.org/10.1021/jacs.5b11380.

(198) Parry, R. J.; Li, W. The Biosynthesis of Valanimycin. Further Evidence for the Intermediacy of a Hydroxylamine in N-N Bond Formation. *J. Chem. Soc. Chem. Commun.* **1994**, No. 8, 995. https://doi.org/10.1039/c39940000995.

(199) Neumann, C. S.; Jiang, W.; Heemstra, J. R.; Gontang, E. A.; Kolter, R.; Walsh, C. T. Biosynthesis of Piperazic Acid via N5-Hydroxy-Ornithine in Kutzneria Spp. 744. *Chembiochem Eur. J. Chem. Biol.* **2012**, *13* (7), 972–976. https://doi.org/10.1002/cbic.201200054.

(200) Du, Y.-L.; He, H.-Y.; Higgins, M. A.; Ryan, K. S. A Heme-Dependent Enzyme Forms the Nitrogen–Nitrogen Bond in Piperazate. *Nat. Chem. Biol.* **2017**, *13* (8), 836–838. https://doi.org/10.1038/nchembio.2411.

(201) NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2015**, *44* (D1), D7–D19. https://doi.org/10.1093/nar/gkv1290.

(202) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. AntiSMASH 3.0—a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2015**, *43* (W1), W237–W243. https://doi.org/10.1093/nar/gkv437.

(203) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. AntiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res.* **2019**, *47* (W1), W81–W87. https://doi.org/10.1093/nar/gkz310.

(204) Chevrette, M. G.; Aicheler, F.; Kohlbacher, O.; Currie, C. R.; Medema, M. H. SANDPUMA: Ensemble Predictions of Nonribosomal Peptide Chemistry Reveal Biosynthetic Diversity across Actinobacteria. *Bioinforma. Oxf. Engl.* **2017**, *33* (20), 3202–3210. https://doi.org/10.1093/bioinformatics/btx400.

(205) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H. AntiSMASH 4.0—Improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41. https://doi.org/10.1093/nar/gkx319.

(206) Williams, David. E.; Dalisay, D. S.; Patrick, B. O.; Matainaho, T.; Andrusiak, K.; Deshpande, R.; Myers, C. L.; Piotrowski, J. S.; Boone, C.; Yoshida, M.; Andersen, R. J. Padanamides A and B, Highly Modified Linear Tetrapeptides Produced in Culture by a *Streptomyces* Sp. Isolated from a Marine Sediment. *Org. Lett.* **2011**, *13* (15), 3936–3939. https://doi.org/10.1021/ol2014494.

(207) Parella, T.; Sánchez-Ferrando, F. Improved Multiplicity-Edited ADEQUATE Experiments. *J. Magn. Reson.* **2004**, *166* (1), 123–128. https://doi.org/10.1016/j.jmr.2003.10.010.

(208) Martin, G. E. Chapter 5 - Using 1,1- and 1,n-ADEQUATE 2D NMR Data in Structure Elucidation Protocols. In *Annual Reports on NMR Spectroscopy*; Webb, G. A., Ed.; Academic Press, 2011; Vol. 74, pp 215–291. https://doi.org/10.1016/B978-0-08-097072-1.00005-4.

(209) Becker, E. D. Chapter 13 - Elucidation of Molecular Structure and Macromolecular Conformation. In *High Resolution NMR (Third Edition)*; Becker, E. D., Ed.; Academic

Press: San Diego, 2000; pp 347–367. https://doi.org/10.1016/B978-012084662-7/50057-0.

(210) Becker, E. D. Chapter 3 - Instrumentation and Techniques. In *High Resolution NMR (Third Edition)*; Becker, E. D., Ed.; Academic Press: San Diego, 2000; pp 49–82. https://doi.org/10.1016/B978-012084662-7/50047-8.

(211) Witanowski, M.; Webb, G. A. Nitrogen NMR Spectroscopy. In *Annual Reports on NMR Spectroscopy*; Elsevier, 1972; Vol. 5, pp 395–464. https://doi.org/10.1016/S0066-4103(08)60439-1.

(212) Martin, G. E.; Williams, A. J. Utilizing Long-Range 1 H–15 N 2-D NMR Spectroscopy for Chemical Structure Elucidation and Confirmation. In *eMagRes*; American Cancer Society, 2010. https://doi.org/10.1002/9780470034590.emrstm1083.

(213) Marek, R.; Lycka, A. 15N NMR Spectroscopy in Structural Analysis. *Curr. Org. Chem.* **2002**, *6* (1), 35–66. https://doi.org/10.2174/1385272023374643.

(214) Morris, G. A.; Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **1979**, *101* (3), 760–762. https://doi.org/10.1021/ja00497a058.

(215) Freeman, R. *Spin Choreography: Basic Steps in High Resolution NMR*; Spektrum ; University Science Books: Oxford : Sausalito, Calif, 1997.

(216) Williamson, R. T.; Buevich, A. V.; Martin, G. E.; Parella, T. LR-HSQMBC: A Sensitive NMR Technique To Probe Very Long-Range Heteronuclear Coupling Pathways. *J. Org. Chem.* **2014**, *79* (9), 3887–3894. https://doi.org/10.1021/jo500333u.

(217) Mulder, F. A. A.; Filatov, M. NMR Chemical Shift Data and Ab Initio Shielding Calculations: Emerging Tools for Protein Structure Determination. *Chem Soc Rev* **2010**, *39* (2), 578–590. https://doi.org/10.1039/B811366C.

(218) Deev, S. L.; Paramonov, A. S.; Shestakova, T. S.; Khalymbadzha, I. A.; Chupakhin, O. N.; Subbotina, J. O.; Eltsov, O. S.; Slepukhin, P. A.; Rusinov, V. L.; Arseniev, A. S.; Shenkarev, Z. O. [15] N-Labelling and Structure Determination of Adamantylated Azolo-Azines in Solution. *Beilstein J. Org. Chem.* **2017**, *13*, 2535–2548. https://doi.org/10.3762/bjoc.13.250.

(219) Ohki, S.; Kainosho, M. Stable Isotope Labeling Methods for Protein NMR Spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53* (4), 208–226. https://doi.org/10.1016/j.pnmrs.2008.01.003.

(220) Chokkathukalam, A.; Kim, D.-H.; Barrett, M. P.; Breitling, R.; Creek, D. J. Stable Isotope-Labeling Studies in Metabolomics: New Insights into Structure and Dynamics of Metabolic Networks. *Bioanalysis* **2014**, *6* (4), 511–524. https://doi.org/10.4155/bio.13.348.

(221) Bax, A. Two-Dimensional NMR and Protein Structure. *Annu. Rev. Biochem.* **1989**, *58* (1), 223–256. https://doi.org/10.1146/annurev.bi.58.070189.001255.

(222) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. 1H, 13C and 15N Chemical Shift Referencing in Biomolecular NMR. *J. Biomol. NMR* **1995**, *6* (2), 135–140. https://doi.org/10.1007/BF00211777.

(223) Martin, G. E.; Solntseva, M.; Williams, A. J. Applications of 15N NMR Spectroscopy in Alkaloid Chemistry. In *Modern Alkaloids*; John Wiley & Sons, Ltd, 2007; pp 409–471. https://doi.org/10.1002/9783527621071.ch14.

(224) Williamson, R. T.; Márquez, B. L.; Gerwick, W. H. Use of 1H-15N PEP-HSQC-TOCSY at Natural Abundance to Facilitate the Structure Elucidation of Naturally Occurring Peptides. *Tetrahedron* **1999**, *55* (10), 2881–2888. https://doi.org/10.1016/S0040-4020(99)00081-2.

(225) A. Ciufolini, M.; Xi, N. Synthesis, Chemistry and Conformational Properties of Piperazic Acids. *Chem. Soc. Rev.* **1998**, *27* (6), 437. https://doi.org/10.1039/a827437z.

(226) Weller, M. G. A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques. *Sensors* **2012**, *12* (7), 9181–9209. https://doi.org/10.3390/s120709181.

(227) Morgan, K. D.; Williams, D. E.; Patrick, B. O.; Remigy, M.; Banuelos, C. A.; Sadar, M. D.; Ryan, K. S.; Andersen, R. J. Incarnatapeptins A and B, Nonribosomal Peptides Discovered Using Genome Mining and 1H/15N HSQC-TOCSY. *Org. Lett.* **2020**, *22* (11), 4053–4057. https://doi.org/10.1021/acs.orglett.0c00818.

(228) Ta, L.; Gosa, L.; Nathanson, D. A. Biosafety and Biohazards: Understanding Biosafety Levels and Meeting Safety Requirements of a Biobank. In *Biobanking*; Yong, W. H., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2019; Vol. 1897, pp 213–225. https://doi.org/10.1007/978-1-4939-8935-5_19.

(229) Sharrar, A. M.; Crits-Christoph, A.; Méheust, R.; Diamond, S.; Starr, E. P.; Banfield, J. F. Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. *mBio* **2020**, *11* (3), e00416-20, /mbio/11/3/mBio.00416-20.atom. https://doi.org/10.1128/mBio.00416-20.

(230) Oshima, K.; Hattori, M.; Shimizu, H.; Fukuda, K.; Nemoto, M.; Inagaki, K.; Tamura, T. Draft Genome Sequence of *Streptomyces Incarnatus* NRRL8089, Which Produces the Nucleoside Antibiotic Sinefungin. *Genome Announc.* **2015**, *3* (4). https://doi.org/10.1128/genomeA.00715-15.

(231) Florent, J.; Lumel, J.; Mancy, D. Aplha,Delta-DIAMINOACID AND ITS LACTAM. 4,189,349, February 19, 1980.

(232) Greule, A.; Marolt, M.; Deubel, D.; Peintner, I.; Zhang, S.; Jessen-Trefzer, C.; De Ford, C.; Burschel, S.; Li, S.-M.; Friedrich, T.; Merfort, I.; Lüdeke, S.; Bisel, P.; Müller, M.; Paululat, T.; Bechthold, A. Wide Distribution of Foxicin Biosynthetic Gene Clusters in Streptomyces Strains – An Unusual Secondary Metabolite with Various Properties. *Front. Microbiol.* **2017**, *8*. https://doi.org/10.3389/fmicb.2017.00221.

(233) Bader, C. D.; Panter, F.; Müller, R. In Depth Natural Product Discovery - Myxobacterial Strains That Provided Multiple Secondary Metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480. https://doi.org/10.1016/j.biotechadv.2019.107480.

(234) Fritz, S.; Rajaonison, A.; Chabrol, O.; Raoult, D.; Rolain, J.-M.; Merhej, V. Full-Length Title: NRPPUR Database Search and in Vitro Analysis Identify an NRPS-PKS Biosynthetic Gene Cluster with a Potential Antibiotic Effect. *BMC Bioinformatics* **2018**, *19* (1), 463. https://doi.org/10.1186/s12859-018-2479-5.

(235) Parry, R. J.; Arzu, I. Y.; Ju, S.; Baker, B. J. Biosynthesis of Sinefungin: On the Mode of Incorporation of L-Ornithine. *J. Am. Chem. Soc.* **1989**, *111* (24), 8981–8982. https://doi.org/10.1021/ja00206a066.

(236) Niitsuma, M.; Hashida, J.; Iwatsuki, M.; Mori, M.; Ishiyama, A.; Namatame, M.; Nishihara-Tsukashima, A.; Matsumoto, A.; Takahashi, Y.; Yamada, H.; Otoguro, K.; Shiomi, K.; Ōmura, S. Sinefungin VA and Dehydrosinefungin V, New Antitrypanosomal

Antibiotics Produced by Streptomyces Sp. K05-0178. *J. Antibiot. (Tokyo)* **2010**, *63* (11), 673–679. https://doi.org/10.1038/ja.2010.102.

(237) Wyche, T. P.; Ruzzini, A. C.; Beemelmanns, C.; Kim, K. H.; Klassen, J. L.; Cao, S.; Poulsen, M.; Bugni, T. S.; Currie, C. R.; Clardy, J. Linear Peptides Are the Major Products of a Biosynthetic Pathway That Encodes for Cyclic Depsipeptides. *Org. Lett.* **2017**, *19* (7), 1772–1775. https://doi.org/10.1021/acs.orglett.7b00545.

(238) Webb, G. A.; Belton, P. S.; McCarthy, M. J. *Annual Reports on NMR Spectroscopy*; Academic Press: London, 1995.

(239) Sugawara, K.; Toda, S.; Moriyama, T.; Konishi, M.; Oki, T. Verucopeptin, a New Antitumor Antibiotic Active against B16 Melanoma. II. Structure Determination. *J. Antibiot. (Tokyo)* **1993**, *46* (6), 928–935. https://doi.org/10.7164/antibiotics.46.928.

(240) Yoshimura, A.; Nishimura, S.; Otsuka, S.; Hattori, A.; Kakeya, H. Structure Elucidation of Verucopeptin, a HIF-1 Inhibitory Polyketide–Hexapeptide Hybrid Metabolite from an Actinomycete. *Org. Lett.* **2015**, *17* (21), 5364–5367. https://doi.org/10.1021/acs.orglett.5b02718.

(241) Cahn, R. S.; Ingold, C.; Prelog, V. Specification of Molecular Chirality. *Angew. Chem. Int. Ed. Engl.* **1966**, *5* (4), 385–415. https://doi.org/10.1002/anie.196603851.

(242) Prelog, V.; Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem. Int. Ed. Engl.* **1982**, *21* (8), 567–583. https://doi.org/10.1002/anie.198205671.

(243) Morgan, K. D.; Williams, D. E.; Patrick, B. O.; Remigy, M.; Banuelos, C. A.; Sadar, M. D.; Ryan, K. S.; Andersen, R. J. Incarnatapeptins A and B, Nonribosomal Peptides Discovered Using Genome Mining and ^{1}H/^{15}N HSQC-TOCSY. *Org. Lett.* **2020**, acs.orglett.0c00818. https://doi.org/10.1021/acs.orglett.0c00818.

(244) Mo, S.; Kim, D. H.; Lee, J. H.; Park, J. W.; Basnet, D. B.; Ban, Y. H.; Yoo, Y. J.; Chen, S.; Park, S. R.; Choi, E. A.; Kim, E.; Jin, Y.-Y.; Lee, S.-K.; Park, J. Y.; Liu, Y.; Lee, M. O.; Lee, K. S.; Kim, S. J.; Kim, D.; Park, B. C.; Lee, S.; Kwon, H. J.; Suh, J.-W.; Moore, B. S.; Lim, S.-K.; Yoon, Y. J. Biosynthesis of the Allylmalonyl-CoA Extender Unit for the FK506 Polyketide Synthase Proceeds through a Dedicated Polyketide Synthase and Facilitates the Mutasynthesis of Analogues. *J. Am. Chem. Soc.* **2011**, *133* (4), 976–985. https://doi.org/10.1021/ja108399b.

(245) Ding, L.; Maier, A.; Fiebig, H.-H.; Görls, H.; Lin, W.-H.; Peschel, G.; Hertweck, C. Divergolides A–D from a Mangrove Endophyte Reveal an Unparalleled Plasticity in Ansa-Macrolide Biosynthesis. *Angew. Chem. Int. Ed.* **2011**, *50* (7), 1630–1634. https://doi.org/10.1002/anie.201006165.

(246) Umezawa, K.; Ikeda, Y.; Kawase, O.; Naganawa, H.; Kondo, S. Biosynthesis of Polyoxypeptin A: Novel Amino Acid 3-Hydroxy-3-Methylproline Derived from Isoleucine. *J. Chem. Soc. Perkin 1* **2001**, No. 13, 1550–1553. https://doi.org/10.1039/b101942m.

(247) Liu, X.; Bolla, K.; Ashforth, E. J.; Zhuo, Y.; Gao, H.; Huang, P.; Stanley, S. A.; Hung, D. T.; Zhang, L. Systematics-Guided Bioprospecting for Bioactive Microbial Natural Products. *Antonie Van Leeuwenhoek* **2012**, *101* (1), 55–66. https://doi.org/10.1007/s10482-011-9671-1.

(248) 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.

(249) Dalisay, D. S.; Williams, D. E.; Wang, X. L.; Centko, R.; Chen, J.; Andersen, R. J. Marine Sediment-Derived Streptomyces Bacteria from British Columbia, Canada Are a Promising Microbiota Resource for the Discovery of Antimicrobial Natural Products. *PLoS ONE* **2013**, *8* (10), e77078. https://doi.org/10.1371/journal.pone.0077078.

(250) Baltz, R. H. Gifted Microbes for Genome Mining and Natural Product Discovery. *J. Ind. Microbiol. Biotechnol.* **2017**, *44* (4–5), 573–588. https://doi.org/10.1007/s10295-016-1815-x.

(251) Harvey, A. Strategies for Discovering Drugs from Previously Unexplored Natural Products. *Drug Discov. Today* **2000**, *5* (7), 294–300. https://doi.org/10.1016/S1359-6446(00)01511-7.

(252) Barford, D.; Das, A. K.; Egloff, M.-P. THE STRUCTURE AND MECHANISM OF PROTEIN PHOSPHATASES: Insights into Catalysis and Regulation. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27* (1), 133–164. https://doi.org/10.1146/annurev.biophys.27.1.133.

(253) Hunter, C. A.; Jones, S. A. IL-6 as a Keystone Cytokine in Health and Disease. *Nat. Immunol.* **2015**, *16* (5), 448–457. https://doi.org/10.1038/ni.3153.

(254) Neurath, M. F.; Finotto, S. IL-6 Signaling in Autoimmunity, Chronic Inflammation and Inflammation-Associated Cancer. *Cytokine Growth Factor Rev.* **2011**, *22* (2), 83–89. https://doi.org/10.1016/j.cytogfr.2011.02.003.

(255) Davies, J.; Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **2010**, *74* (3), 417–433. https://doi.org/10.1128/MMBR.00016-10.

(256) Krysan, D. J. The Unmet Clinical Need of Novel Antifungal Drugs. *Virulence* **2017**, *8* (2), 135–137. https://doi.org/10.1080/21505594.2016.1276692.

(257) Berman, J.; Krysan, D. J. Drug Resistance and Tolerance in Fungi. *Nat. Rev. Microbiol.* **2020**, *18* (6), 319–331. https://doi.org/10.1038/s41579-019-0322-2.

(258) Heifetz, C. L.; Chodubski, J. A.; DeCarlo, M. O.; Fisher, M. W. Disc-Agar Diffusion Microbiological Assay Procedure for Determining Serum and Urine Levels of Sulfacytine and Other Sulfonamides. *Appl. Microbiol.* **1971**, *21* (5), 893–898. https://doi.org/10.1128/AEM.21.5.893-898.1971.

(259) Florey, H. *Antibiotics* **1949**, *1*.

(260) Meyers, E.; Erickson, R. C. Bioautography of Antibiotics on Thin Layer Chromatograms. *J. Chromatogr. A* **1967**, *26*, 531–532. https://doi.org/10.1016/S0021-9673(01)98921-6.

(261) Usdin, E.; Shockman, G. D.; Toennies, G. Tetrazolium Bioautography1. *Appl. Microbiol.* **1954**, *2* (1), 29–33. https://doi.org/10.1128/AEM.2.1.29-33.1954.

(262) Legerská, B.; Chmelová, D.; Ondrejovič, M.; Miertuš, S. The TLC-Bioautography as a Tool for Rapid Enzyme Inhibitors Detection - A Review. *Crit. Rev. Anal. Chem.* **2020**, 1–19. https://doi.org/10.1080/10408347.2020.1797467.

(263) Dewanjee, S.; Gangopadhyay, M.; Bhattacharya, N.; Khanra, R.; Dua, T. K. Bioautography and Its Scope in the Field of Natural Product Chemistry. *J. Pharm. Anal.* **2015**, *5* (2), 75–84. https://doi.org/10.1016/j.jpha.2014.06.002.

(264) Kanbe, K.; Okamura, M.; Hattori, S.; Naganawa, H.; Hamada, M.; Okami, Y.; Takeuchi, T. Thienodolin, a New Plant Growth-Regulating Substance Produced by a Streptomycete Strain: I. Taxonomy and Fermentation of the Producing Strain, and the Isolation and Characterization of Thienodolin. *Biosci. Biotechnol. Biochem.* **1993**, *57* (4), 632–635. https://doi.org/10.1271/bbb.57.632.

(265)   The New Ultraviolet Ray Absorbing Material and Its Manufacturing Method. JP 11269175, October 5, 1999.

(266)   Park, J.-S.; Yabe, S.; Shin-ya, K.; Nishiyama, M.; Kuzuyama, T. New 2-(1′H-Indole-3′-Carbonyl)-Thiazoles Derived from the Thermophilic Bacterium Thermosporothrix Hazakensis SK20-1T. *J. Antibiot. (Tokyo)* **2015**, *68* (1), 60–62. https://doi.org/10.1038/ja.2014.93.

(267)   Igarashi, Y.; Yamamoto, K.; Ueno, C.; Yamada, N.; Saito, K.; Takahashi, K.; Enomoto, M.; Kuwahara, S.; Oikawa, T.; Tashiro, E.; Imoto, M.; Xiaohanyao, Y.; Zhou, T.; Harunari, E.; Oku, N. Ktedonoketone and 2'-Oxosattabacin, Benzenoid Metabolites from a Thermophilic Bacterium Thermosporothrix Hazakensis in the Phylum Chloroflexi. *J. Antibiot. (Tokyo)* **2019**, *72* (9), 653–660. https://doi.org/10.1038/s41429-019-0195-7.

(268)   Jansen, R.; Mohr, K. I.; Bernecker, S.; Stadler, M.; Müller, R. Indothiazinone, an Indolyl Thiazolyl Ketone from a Novel Myxobacterium Belonging to the Sorangiineae. *J. Nat. Prod.* **2014**, *77* (4), 1054–1060. https://doi.org/10.1021/np500144t.

(269)   Itoh, J.; Omoto, S.; Nishizawa, N.; Kodama, Y.; Inouye, S. Chemical Structures of Amicoumacins Produced by *Bacillus Pumilus*. *Agric. Biol. Chem.* **1982**, *46* (11), 2659–2665. https://doi.org/10.1080/00021369.1982.10865491.

(270)   Bing, H.; Cui, C.; Bing, C.; Xin-sheng, Y. (-)-Anhydrocycloheximide, a New Cell Cycle Inhibitor, and l-Cycloheximide Produced by Streptomyces Flavoretus. *Chin J Med Chem* **2004**, *14* (14), 370–372.

(271)   Boshes, L. D. Fungus Infections of the Central Nervous System: Experience in Treatment of Cryptococcosis with Cycloheximide (Actidione). *AMA Arch. Neurol. Psychiatry* **1956**, *75* (2), 175. https://doi.org/10.1001/archneurpsyc.1956.02330200069008.

(272)   Young, C. W.; Dowling, M. D. Antipyretic Effect of Cycloheximide, an Inhibitor of Protein Synthesis, in Patients with Hodgkin's Disease or Other Malignant Neoplasms. **1975**, *35*, 8.

(273)   Pajak, B.; Orzechowska, S.; Gajkowska, B.; Orzechowski, A. Bisindolylmaleimides in Anti-Cancer Therapy - More than PKC Inhibitors. *Adv. Med. Sci.* **2008**, *53* (1). https://doi.org/10.2478/v10039-008-0028-6.

(274)   Georg, L. K. USE OF A CYCLOHEXIMIDE MEDIUM FOR ISOLATION OF DERMATOPHYTES FROM CLINICAL MATERIALS. *Arch. Dermatol.* **1953**, *67* (4), 355. https://doi.org/10.1001/archderm.1953.01540040013003.

(275)   Baltz, R. H. Genetic Methods and Strategies for Secondary Metabolite Yield Improvement in Actinomycetes. *Antonie Van Leeuwenhoek* **2001**, *79* (3/4), 251–259. https://doi.org/10.1023/A:1012020918624.

(276)   Joynt, R.; Seipke, R. F. A Phylogenetic and Evolutionary Analysis of Antimycin Biosynthesis. *Microbiology* **2018**, *164* (1), 28–39. https://doi.org/10.1099/mic.0.000572.

(277)   Bułyszko, I.; Dräger, G.; Klenge, A.; Kirschning, A. Evaluation of the Synthetic Potential of an AHBA Knockout Mutant of the Rifamycin Producer *Amycolatopsis Mediterranei*. *Chem. - Eur. J.* **2015**, *21* (52), 19231–19242. https://doi.org/10.1002/chem.201503548.

(278)   August, P. R.; Tang, L.; Yoon, Y. J.; Ning, S.; Müller, R.; Yu, T.-W.; Taylor, M.; Hoffmann, D.; Kim, C.-G.; Zhang, X.; Hutchinson, C. R.; Floss, H. G. Biosynthesis of the Ansamycin Antibiotic Rifamycin: Deductions from the Molecular Analysis of the Rif

Biosynthetic Gene Cluster of Amycolatopsis Mediterranei S699. *Chem. Biol.* **1998**, *5* (2), 69–79. https://doi.org/10.1016/S1074-5521(98)90141-7.

(279) Yu, T.-W.; Shen, Y.; Doi-Katayama, Y.; Tang, L.; Park, C.; Moore, B. S.; Richard Hutchinson, C.; Floss, H. G. Direct Evidence That the Rifamycin Polyketide Synthase Assembles Polyketide Chains Processively. *Proc. Natl. Acad. Sci.* **1999**, *96* (16), 9051–9056. https://doi.org/10.1073/pnas.96.16.9051.

(280) Jensen, P. R.; Gontang, E.; Mafnas, C.; Mincer, T. J.; Fenical, W. Culturable Marine Actinomycete Diversity from Tropical Pacific Ocean Sediments. *Environ. Microbiol.* **2005**, *7* (7), 1039–1048. https://doi.org/10.1111/j.1462-2920.2005.00785.x.

(281) Ahmed, L.; Jensen, P. R.; Freel, K. C.; Brown, R.; Jones, A. L.; Kim, B.-Y.; Goodfellow, M. Salinispora Pacifica Sp. Nov., an Actinomycete from Marine Sediments. *Antonie Van Leeuwenhoek* **2013**, *103* (5), 1069–1078. https://doi.org/10.1007/s10482-013-9886-4.

(282) Román-Ponce, B.; Millán-Aguiñaga, N.; Guillen-Matus, D.; Chase, A. B.; Ginigini, J. G. M.; Soapi, K.; Feussner, K. D.; Jensen, P. R.; Trujillo, M. E. Six Novel Species of the Obligate Marine Actinobacterium Salinispora, Salinispora Cortesiana Sp. Nov., Salinispora Fenicalii Sp. Nov., Salinispora Goodfellowii Sp. Nov., Salinispora Mooreana Sp. Nov., Salinispora Oceanensis Sp. Nov. and Salinispora Vitiensis Sp. Nov., and Emended Description of the Genus Salinispora. *Int. J. Syst. Evol. Microbiol.* **2020**, *70* (8), 4668–4682. https://doi.org/10.1099/ijsem.0.004330.

(283) Nemergut, D. R.; Costello, E. K.; Hamady, M.; Lozupone, C.; Jiang, L.; Schmidt, S. K.; Fierer, N.; Townsend, A. R.; Cleveland, C. C.; Stanish, L.; Knight, R. GLOBAL PATTERNS IN THE BIOGEOGRAPHY OF BACTERIAL TAXA. *Environ. Microbiol.* **2011**, *13* (1), 135–144. https://doi.org/10.1111/j.1462-2920.2010.02315.x.

(284) Wawrik, B.; Kutliev, D.; Abdivasievna, U. A.; Kukor, J. J.; Zylstra, G. J.; Kerkhof, L. Biogeography of Actinomycete Communities and Type II Polyketide Synthase Genes in Soils Collected in New Jersey and Central Asia. *Appl. Environ. Microbiol.* **2007**, *73* (9), 2982–2989. https://doi.org/10.1128/AEM.02611-06.

(285) Jensen, P. R.; Moore, B. S.; Fenical, W. The Marine Actinomycete Genus Salinispora: A Model Organism for Secondary Metabolite Discovery. *Nat. Prod. Rep.* **2015**, *32* (5), 738–751. https://doi.org/10.1039/C4NP00167B.

(286) Wilson, M. C.; Gulder, T. A. M.; Mahmud, T.; Moore, B. S. Shared Biosynthesis of the Saliniketals and Rifamycins in *Salinispora Arenicola* Is Controlled by the *Sare1259 -* Encoded Cytochrome P450. *J. Am. Chem. Soc.* **2010**, *132* (36), 12757–12765. https://doi.org/10.1021/ja105891a.

(287) Feng, Y.; Liu, J.; Carrasco, Y. P.; MacMillan, J. B.; De Brabander, J. K. Rifamycin Biosynthetic Congeners: Isolation and Total Synthesis of Rifsaliniketal and Total Synthesis of Salinisporamycin and Saliniketals A and B. *J. Am. Chem. Soc.* **2016**, *138* (22), 7130–7142. https://doi.org/10.1021/jacs.6b03248.

(288) Kang, Q.; Shen, Y.; Bai, L. Biosynthesis of 3,5-AHBA-Derived Natural Products. *Nat Prod Rep* **2012**, *29* (2), 243–263. https://doi.org/10.1039/C2NP00019A.

(289) Kim, C.-G.; Yu, T.-W.; Fryhle, C. B.; Handa, S.; Floss, H. G. 3-Amino-5-Hydroxybenzoic Acid Synthase, the Terminal Enzyme in the Formation of the Precursor of MC $_7$ N Units in Rifamycin and Related Antibiotics. *J. Biol. Chem.* **1998**, *273* (11), 6030–6040. https://doi.org/10.1074/jbc.273.11.6030.

(290) Xu, J.; Wan, E.; Kim, C.-J.; Floss, H. G.; Mahmud, T. Identification of Tailoring Genes Involved in the Modification of the Polyketide Backbone of Rifamycin B by

Amycolatopsis Mediterranei S699. *Microbiology* **2005**, *151* (8), 2515–2528. https://doi.org/10.1099/mic.0.28138-0.

(291)   Presser, A.; Hefner, A. Trimethylsilyldiazomethane , A Mild and Efficient Reagent for the Methylation of Carboxylic Acids and Alcohols in Natural Products. *Monatshefte Fer ChemieChemical Mon.* **2004**, *135* (8). https://doi.org/10.1007/s00706-004-0188-4.

(292)   Pretsch, E.; Bühlmann, P.; Badertscher, M. *Structure Determination of Organic Compounds: Tables of Spectral Data*, 4th, rev.enl. ed ed.; Springer: Berlin, 2009.

(293)   Fujimoto, H.; Nozawa, M.; Okuyama, E.; Ishibashi, M. Six New Constituents from an Ascomycete, Chaetomium Quadrangulatum, Found in a Screening Study Focused on Monoamine Oxidase Inhibitory Activity. **2003**, *51* (3), 5.

(294)   Marr, D. H.; Stothers, J. B. [13]C N.M.R. STUDIES: PART VI. CARBON-13 SPECTRA OF α,β-UNSATURATED CARBONYL COMPOUNDS. *Can. J. Chem.* **1965**, *43* (3), 596–607. https://doi.org/10.1139/v65-077.

(295)   Hoye, T. R.; Jeffrey, C. S.; Shao, F. Mosher Ester Analysis for the Determination of Absolute Configuration of Stereogenic (Chiral) Carbinol Carbons. *Nat. Protoc.* **2007**, *2* (10), 2451–2458. https://doi.org/10.1038/nprot.2007.354.

(296)   Su, B.-N.; Park, E. J.; Mbwambo, Z. H.; Santarsiero, B. D.; Mesecar, A. D.; Fong, H. H. S.; Pezzuto, J. M.; Kinghorn, A. D. New Chemical Constituents of *Euphorbia q Uinquecostata* and Absolute Configuration Assignment by a Convenient Mosher Ester Procedure Carried Out in NMR Tubes. *J. Nat. Prod.* **2002**, *65* (9), 1278–1282. https://doi.org/10.1021/np0202475.

(297)   Awakawa, T.; Crüsemann, M.; Munguia, J.; Ziemert, N.; Nizet, V.; Fenical, W.; Moore, B. S. Salinipyrone and Pacificanone Are Biosynthetic By-Products of the Rosamicin Polyketide Synthase. *ChemBioChem* **2015**, *16* (10), 1443–1447. https://doi.org/10.1002/cbic.201500177.

(298)   Bode, H. B.; Zeeck, A. Biosynthesis of Kendomycin: Origin of the Oxygen Atoms and Further Investigations. *J. Chem. Soc. Perkin 1* **2000**, No. 16, 2665–2670. https://doi.org/10.1039/b003362f.

(299)   Moore, B. S.; Hertweck, C.; Hopke, J. N.; Izumikawa, M.; Kalaitzis, J. A.; Nilsen, G.; O'Hare, T.; Piel, J.; Shipley, P. R.; Xiang, L.; Austin, M. B.; Noel, J. P. Plant-like Biosynthetic Pathways in Bacteria: From Benzoic Acid to Chalcone [1]. *J. Nat. Prod.* **2002**, *65* (12), 1956–1962. https://doi.org/10.1021/np020230m.

(300)   Bode, H. B.; Zeeck, A. Structure and Biosynthesis of Kendomycin, a Carbocyclic Ansa-Compound from Streptomyces. *J. Chem. Soc. Perkin 1* **2000**, No. 3, 323–328. https://doi.org/10.1039/a908387a.

(301)   Biosynthesis and Attachment of Novel Bacterial Polyketide Synthase Starter Units. *Nat. Prod. Rep.* **2002**, *19* (1), 70–99. https://doi.org/10.1039/b003939j.

(302)   Thiericke, R.; Zeeck, A.; Robinson, J. A.; Beale, J. M.; Floss, H. G. The Biosynthesis of Manumycin; Origin of the Oxygen and Nitrogen Atoms. *J. Chem. Soc. Chem. Commun.* **1989**, No. 7, 402. https://doi.org/10.1039/c39890000402.

(303)   Awakawa, T.; Fujita, N.; Hayakawa, M.; Ohnishi, Y.; Horinouchi, S. Characterization of the Biosynthesis Gene Cluster for Alkyl-O-Dihydrogeranyl-Methoxyhydroquinones in Actinoplanes Missouriensis. *ChemBioChem* **2011**, *12* (3), 439–448. https://doi.org/10.1002/cbic.201000628.

(304) Funabashi, M.; Funa, N.; Horinouchi, S. Phenolic Lipids Synthesized by Type III Polyketide Synthase Confer Penicillin Resistance on *Streptomyces Griseus*. *J. Biol. Chem.* **2008**, *283* (20), 13983–13991. https://doi.org/10.1074/jbc.M710461200.

(305) da Silva, A. B.; Pinto, F. C. L.; Silveira, E. R.; Costa-Lotufo, L. V.; Costa, W. S.; Ayala, A. P.; Canuto, K. M.; Barros, A. B.; Araújo, A. J.; Marinho Filho, J. D. B.; Pessoa, O. D. L. 4-Hydroxy-Pyran-2-One and 3-Hydroxy-N-Methyl-2-Oxindole Derivatives of Salinispora Arenicola from Brazilian Marine Sediments. *Fitoterapia* **2019**, *138*, 104357. https://doi.org/10.1016/j.fitote.2019.104357.

(306) Blair, L. M.; Sperry, J. Natural Products Containing a Nitrogen–Nitrogen Bond. *J. Nat. Prod.* **2013**, *76* (4), 794–812. https://doi.org/10.1021/np400124n.

(307) Le Goff, G.; Ouazzani, J. Natural Hydrazine-Containing Compounds: Biosynthesis, Isolation, Biological Activities and Synthesis. *Bioorg. Med. Chem.* **2014**, *22* (23), 6529–6544. https://doi.org/10.1016/j.bmc.2014.10.011.

(308) He, H.-Y.; Henderson, A. C.; Du, Y.-L.; Ryan, K. S. Two-Enzyme Pathway Links l-Arginine to Nitric Oxide in N-Nitroso Biosynthesis. *J. Am. Chem. Soc.* **2019**, *141* (9), 4026–4033. https://doi.org/10.1021/jacs.8b13049.

(309) Pereira, M. M.; Alcântara, A. F. de C.; Piló-Veloso, D.; Raslan, D. S. NMR Structural Analysis of Braznitidumine: A New Indole Alkaloid with 1,2,9-Triazabicyclo[7.2.1] System, Isolated from Aspidosperma Nitidum (Apocynaceae). *J. Braz. Chem. Soc.* **2006**, *17* (7), 1274–1280. https://doi.org/10.1590/S0103-50532006000700012.

(310) Helaly, S. E.; Pesic, A.; Fiedler, H.-P.; Süssmuth, R. D. Elaiomycins B and C: Alkylhydrazide Antibiotics from *Streptomyces* Sp. BK 190. *Org. Lett.* **2011**, *13* (5), 1052–1055. https://doi.org/10.1021/ol1031014.

(311) Matsuda, K.; Hasebe, F.; Shiwa, Y.; Kanesaki, Y.; Tomita, T.; Yoshikawa, H.; Shin-ya, K.; Kuzuyama, T.; Nishiyama, M. Genome Mining of Amino Group Carrier Protein-Mediated Machinery: Discovery and Biosynthetic Characterization of a Natural Product with Unique Hydrazone Unit. *ACS Chem. Biol.* **2017**, *12* (1), 124–131. https://doi.org/10.1021/acschembio.6b00818.

(312) Le Goff, G.; Martin, M.-T.; Servy, C.; Cortial, S.; Lopes, P.; Bialecki, A.; Smadja, J.; Ouazzani, J. Isolation and Characterization of α,β-Unsaturated γ-Lactono-Hydrazides from *Streptomyces* Sp. *J. Nat. Prod.* **2012**, *75* (5), 915–919. https://doi.org/10.1021/np300026p.

Appendices

Appendix A   NMR of Unsolved Fractions of *Streptomyces incarnatus* NRRL 8089
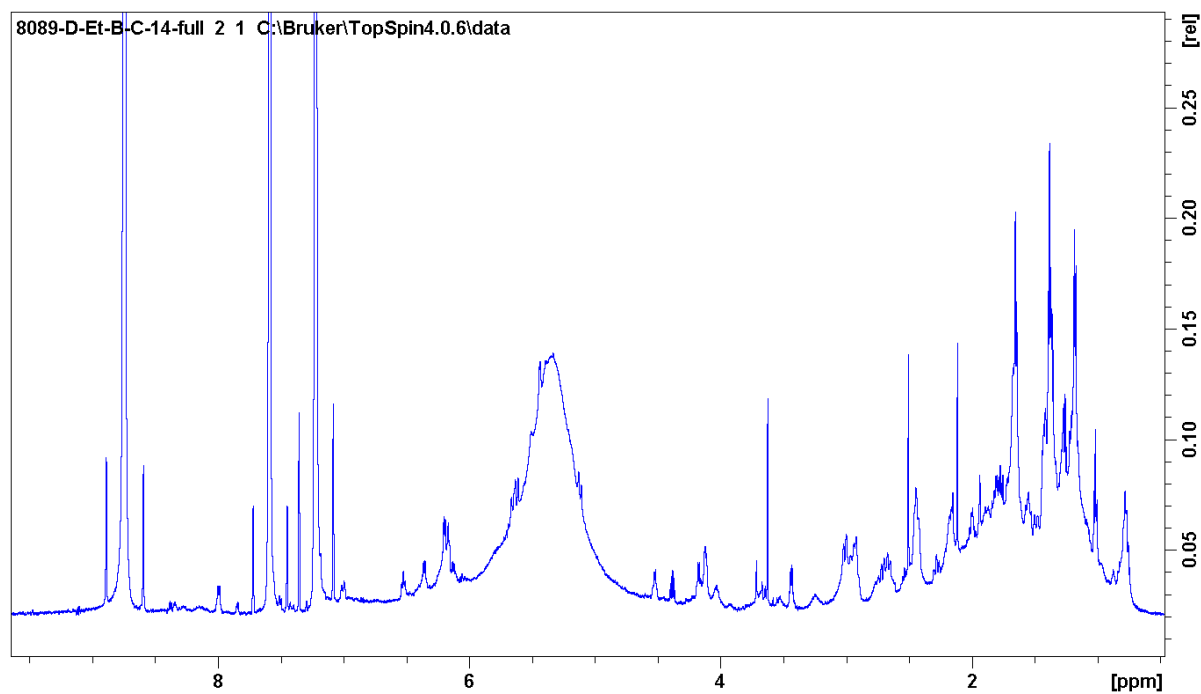
Figure A- 1. ^{1}H NMR spectrum of 8089.Et.B.C-14, recorded at 600 MHz in pyridine-d_5.
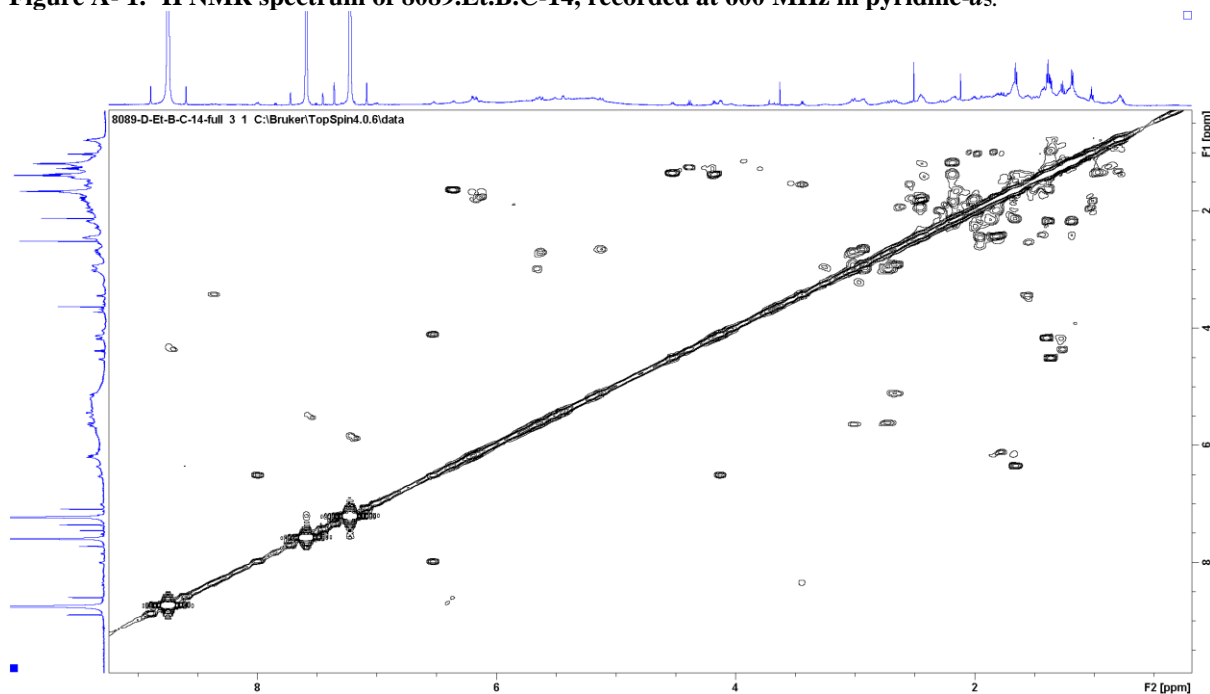
Figure A- 2. COSY60 NMR spectrum of 8089.Et.B.C-14, recorded at 600 MHz in pyridine-d_5.
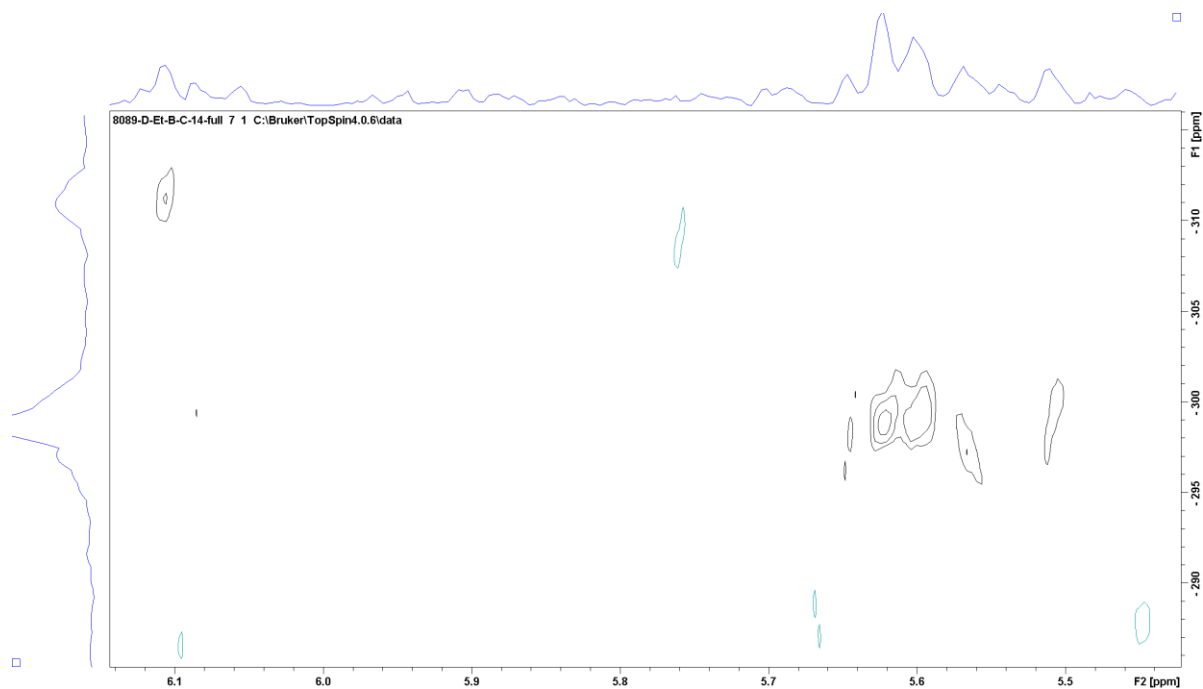
Figure A- 3. ^{1}H-^{15}N HSQC NMR spectrum of 8089.Et.B.C-14, demonstrating Piz N-H correlations recorded at 600 MHz in pyridine-d_5.
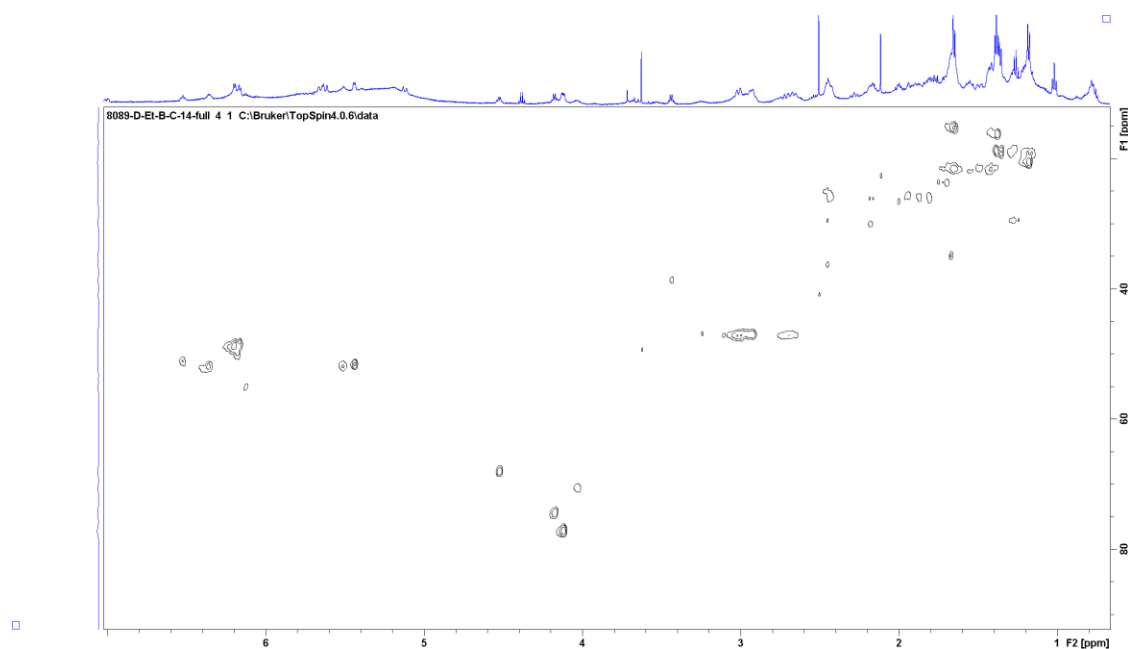
Figure A- 4. ^{1}H-^{13}C HSQC NMR spectrum of 8089.Et.B.C-14, demonstrating unique proton-carbon correlations at 600 MHz in pyridine-d_5.
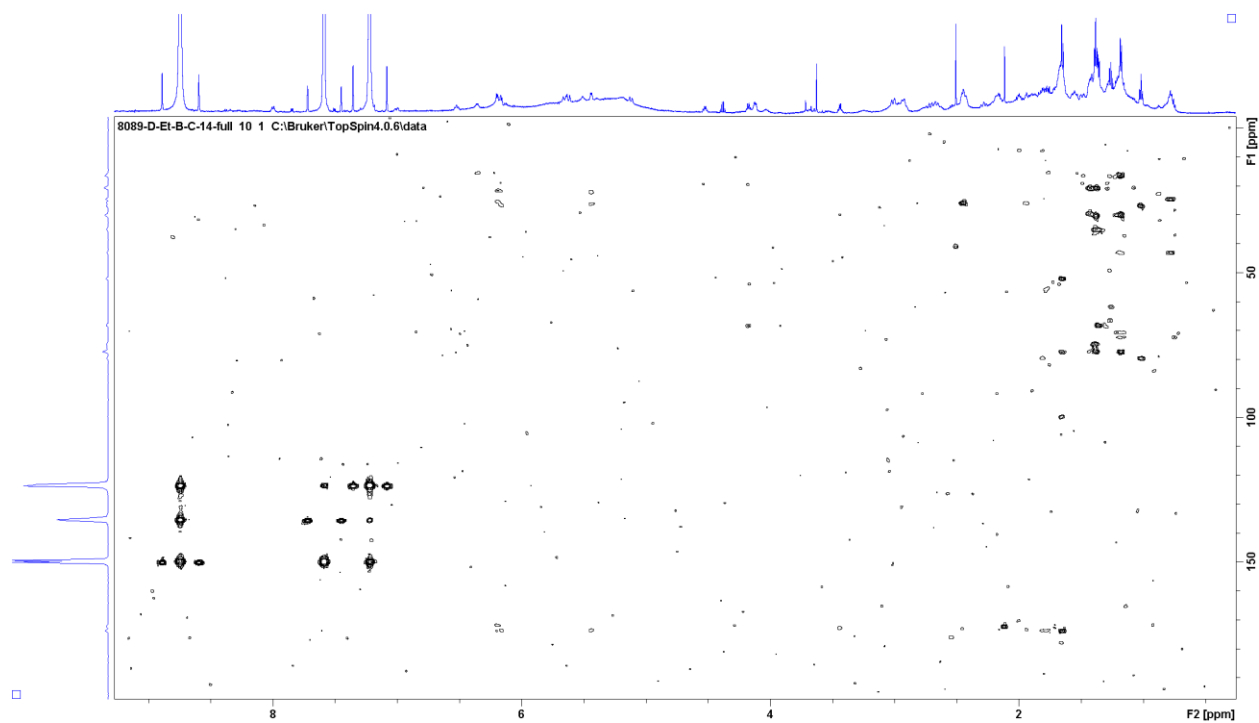
Figure A- 5. ¹H-¹³C HMBC NMR spectrum of 8089.Et.B.C-14, demonstrating unique proton-carbon correlations at 600 MHz in pyridine-*d*₅.
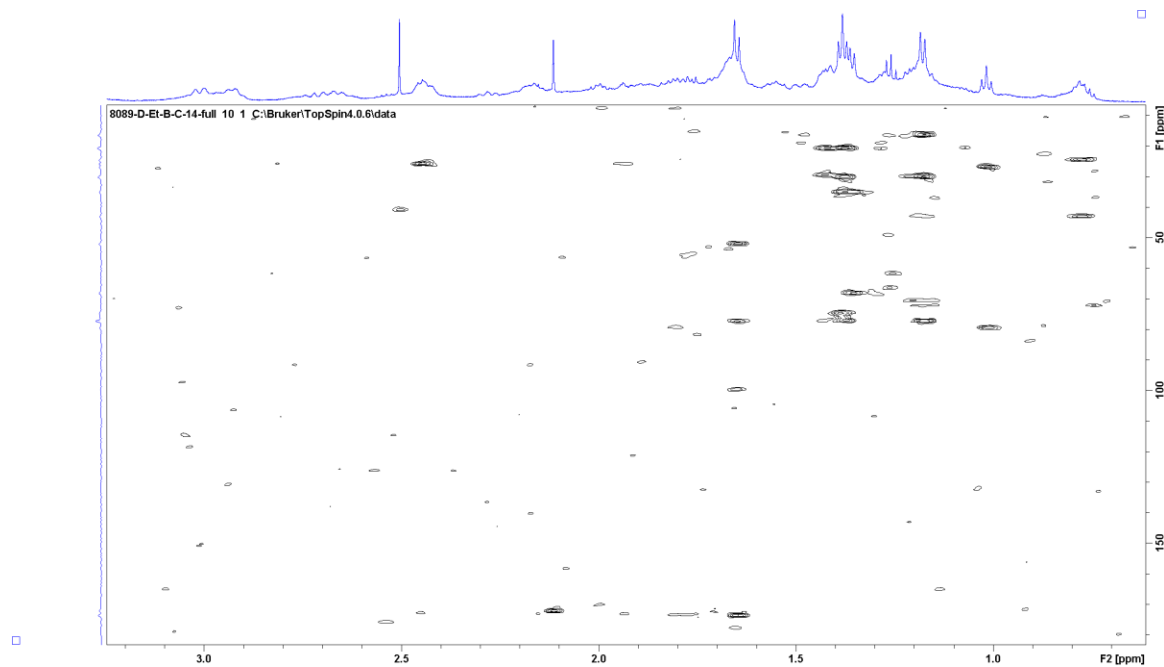
Figure A- 6. Expansion of the ¹H-¹³C HMBC NMR spectrum of 8089.Et.B.C-14, demonstrating unique proton-carbon correlations at 600 MHz in pyridine-*d*₅.
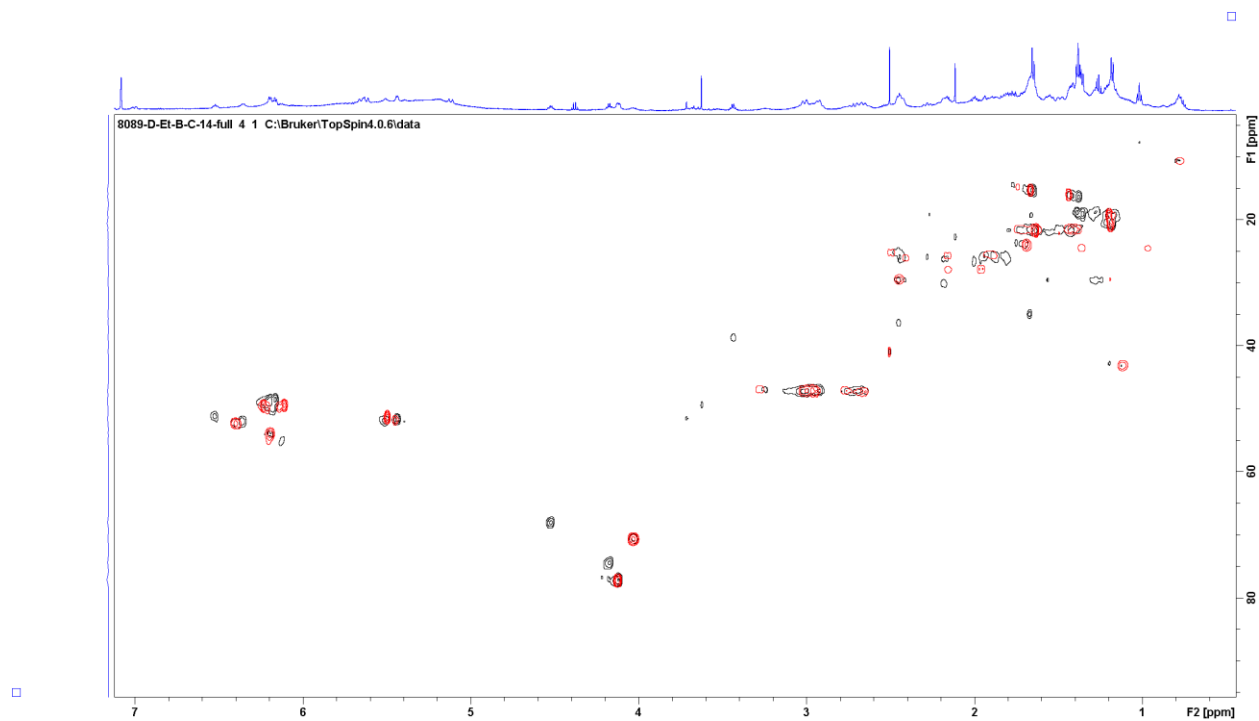
Figure A- 7. ^{1}H-^{13}C HSQC NMR spectrum of 8089.Et.B.C-14, demonstrating unique proton-carbon correlations at 600 MHz in pyridine-d_5 in black. Red correlations are from 8089.Et.B.C-13.
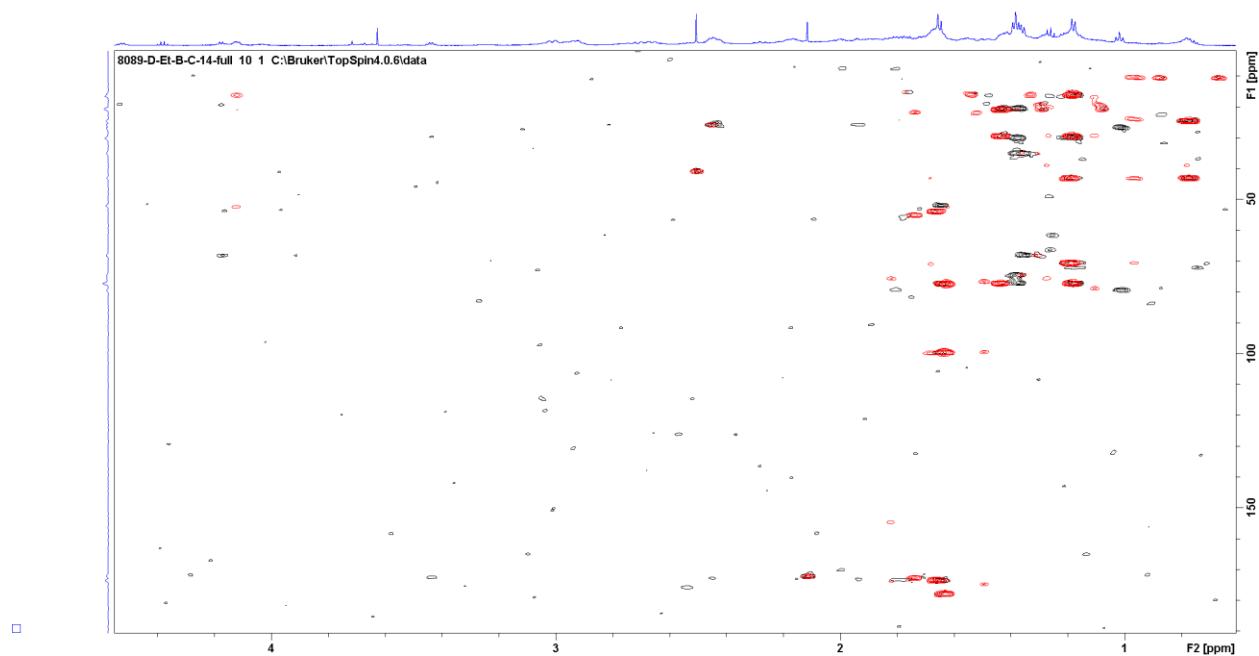
Figure A- 8. ^{1}H-^{13}C HMBC NMR spectrum of 8089.Et.B.C-14, demonstrating unique proton-carbon correlations at 600 MHz in pyridine-d_5 in black. Red correlations are from 8089.Et.B.C-13.

317

Appendix B NMR Spectra of Dentigerumycin F and G Isolated with Formic Acid taken by Dave Williams, Unpublished

Amide-FA 1 1 C:\Bruker\TopSpin4.0.6\data
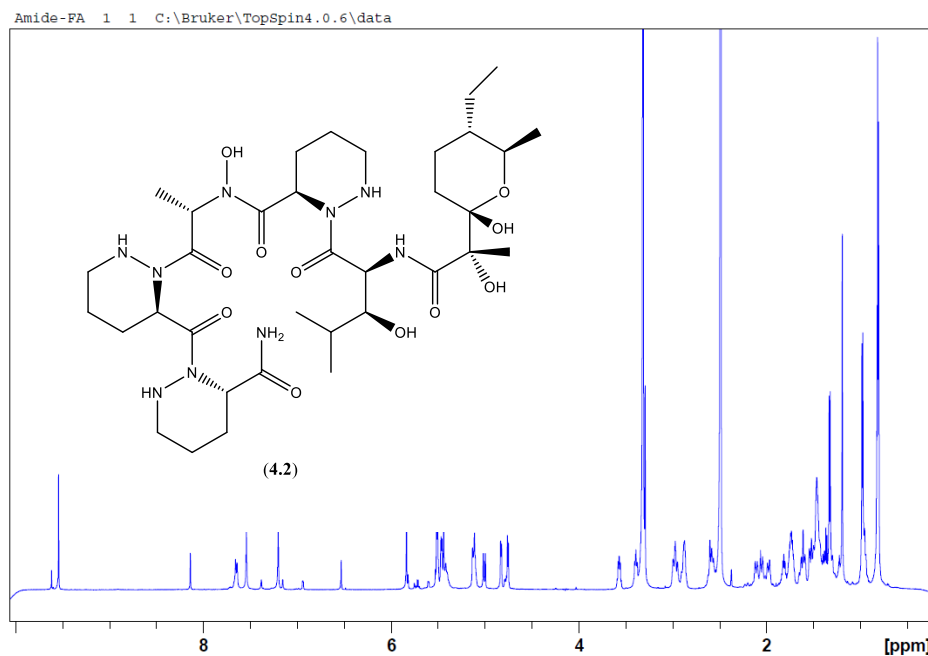
Figure A- 9. ^{1}H NMR spectrum of dentigerumycin F (4.2) recorded at 600 MHz in DMSO-d$_6$.

Amide-FA 11 1 C:\Bruker\TopSpin4.0.6\data
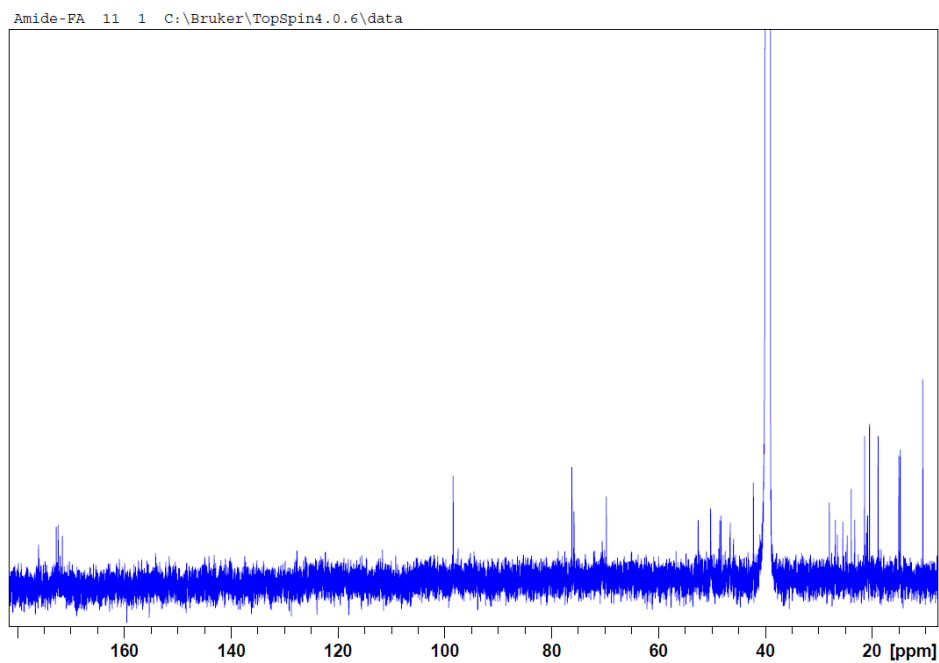
Figure A- 10. ^{13}C NMR spectrum of dentigerumycin F (4.2) recorded at 150 MHz in DMSO-*d*$_6$.

318
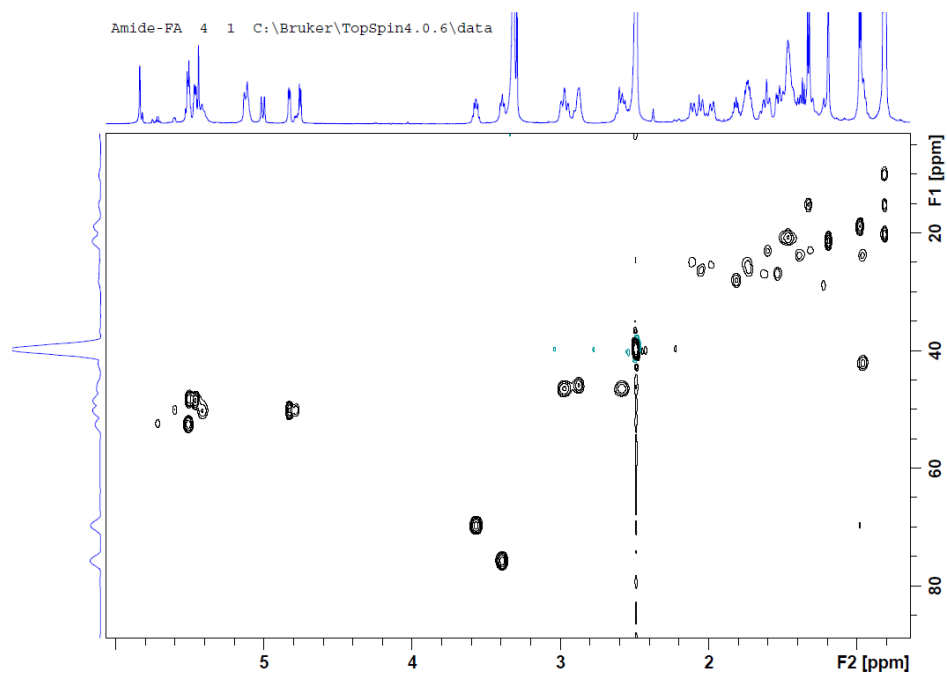
Figure A- 11. ^{1}H-^{13}C HSQC NMR spectrum of dentigerumycin F (4.2) recorded at 600 MHz in DMSO-d_6.
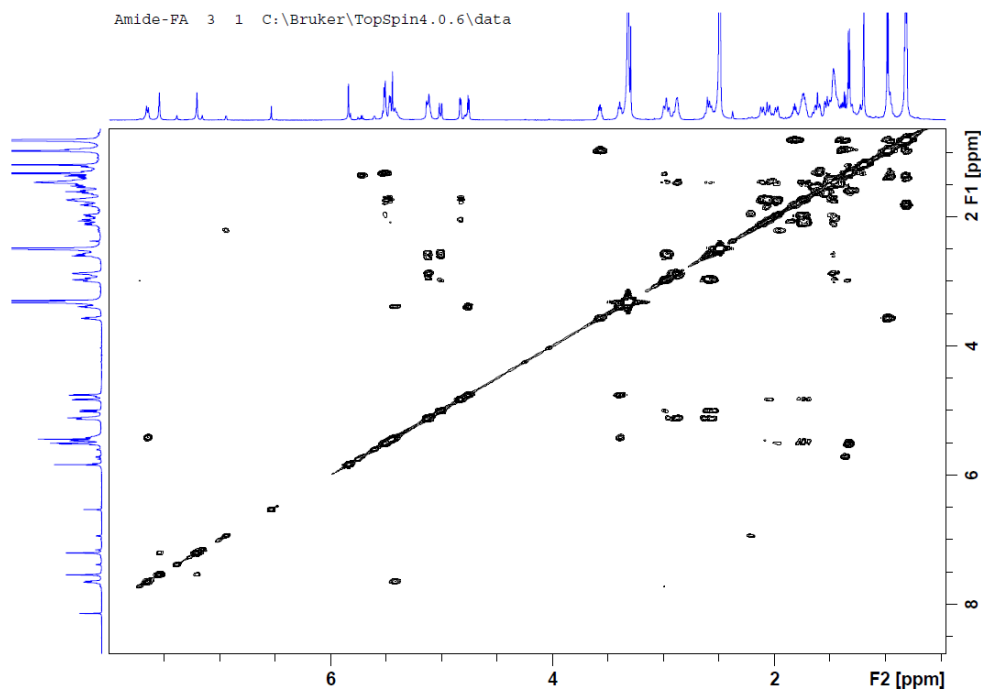
Figure A- 12. ^{1}H-^{1}H gCOSY NMR Spectrum of Dentigerumycin F (4.2) recorded at 600 MHz in DMSO-d_6.
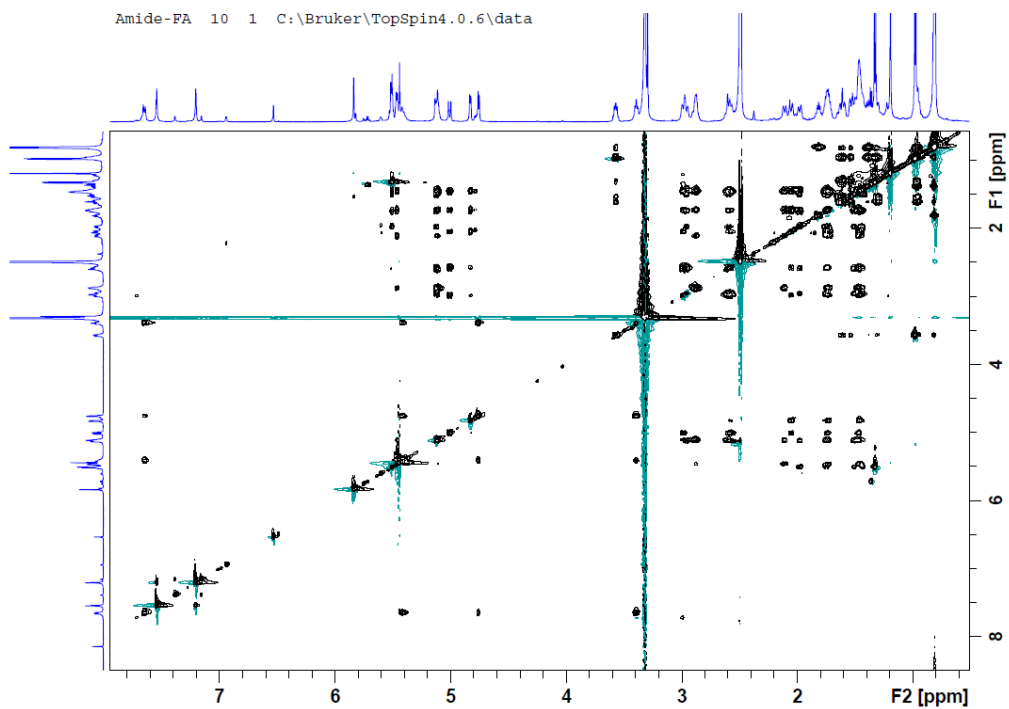
Figure A- 13. ^{1}H-^{1}H gTOCSY NMR Spectrum of Dentigerumycin F (4.2) recorded at 600 MHz in DMSO-d_6.
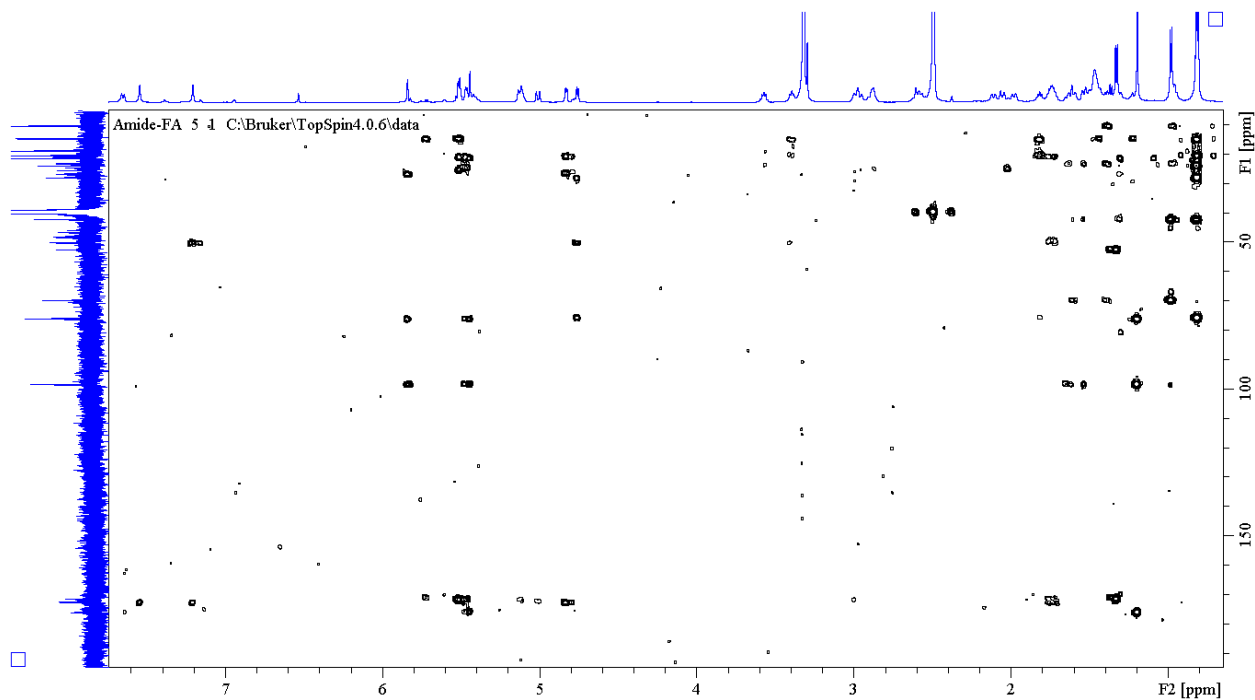
Figure A- 14. ^{1}H-^{13}C gHMBC NMR Spectrum of Dentigerumycin F (4.2) recorded at 600 MHz in DMSO-d_6.

320
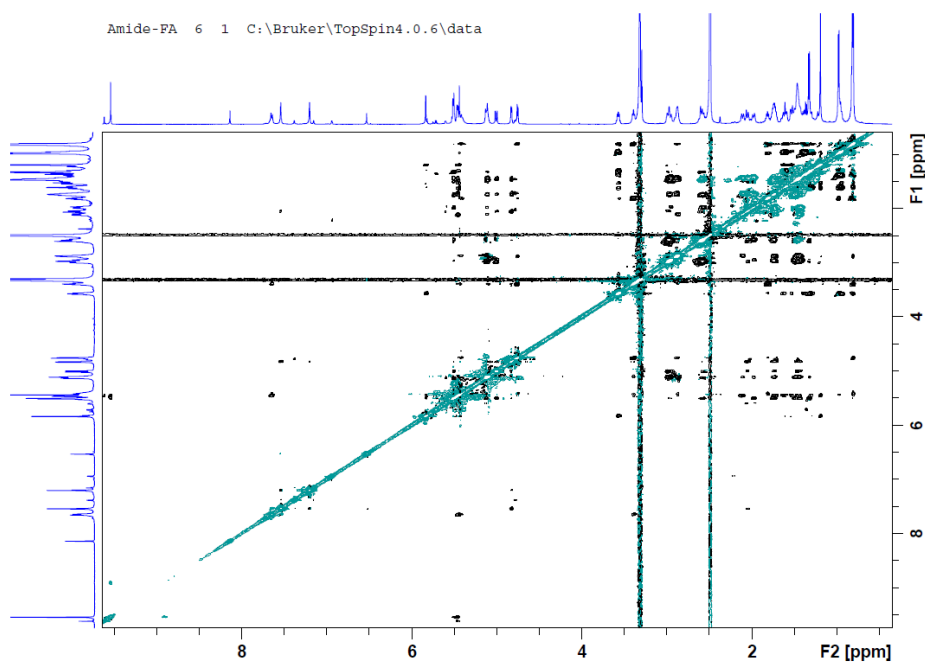
Figure A- 15. tROESY NMR Spectrum of Dentigerumycin F (4.2) recorded at 600 MHz in DMSO-*d₆*.
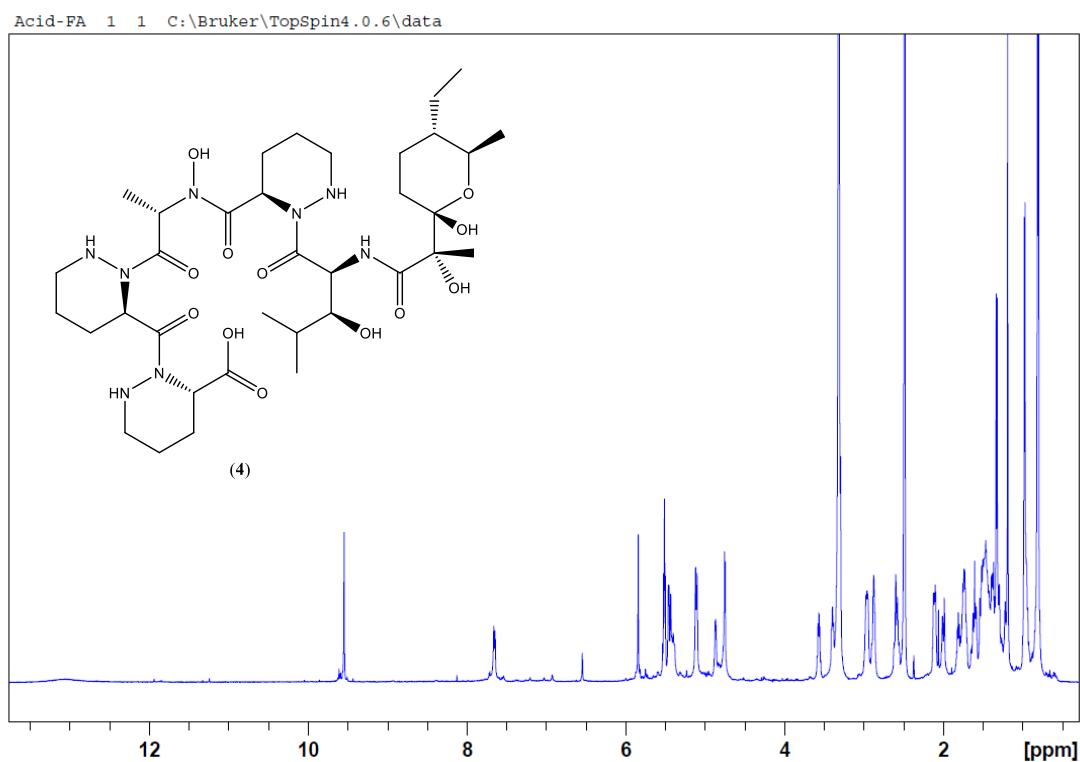
Figure A- 16. ¹H NMR Spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.
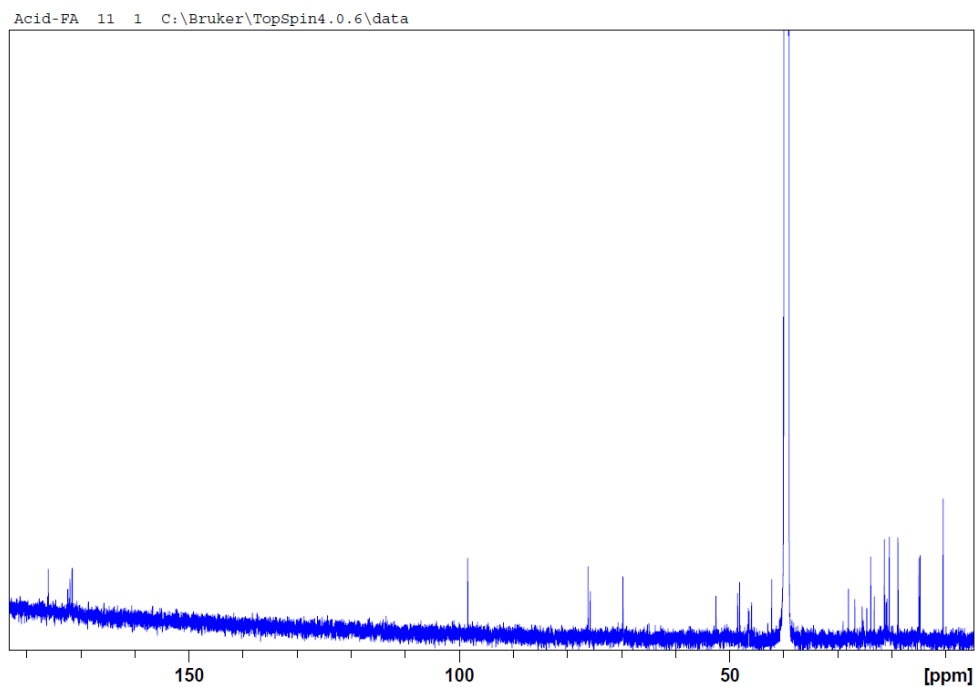
Figure A- 17. ¹³C NMR spectrum of Dentigerumycin G (4.1) recorded at 150 MHz in DMSO-*d₆*.
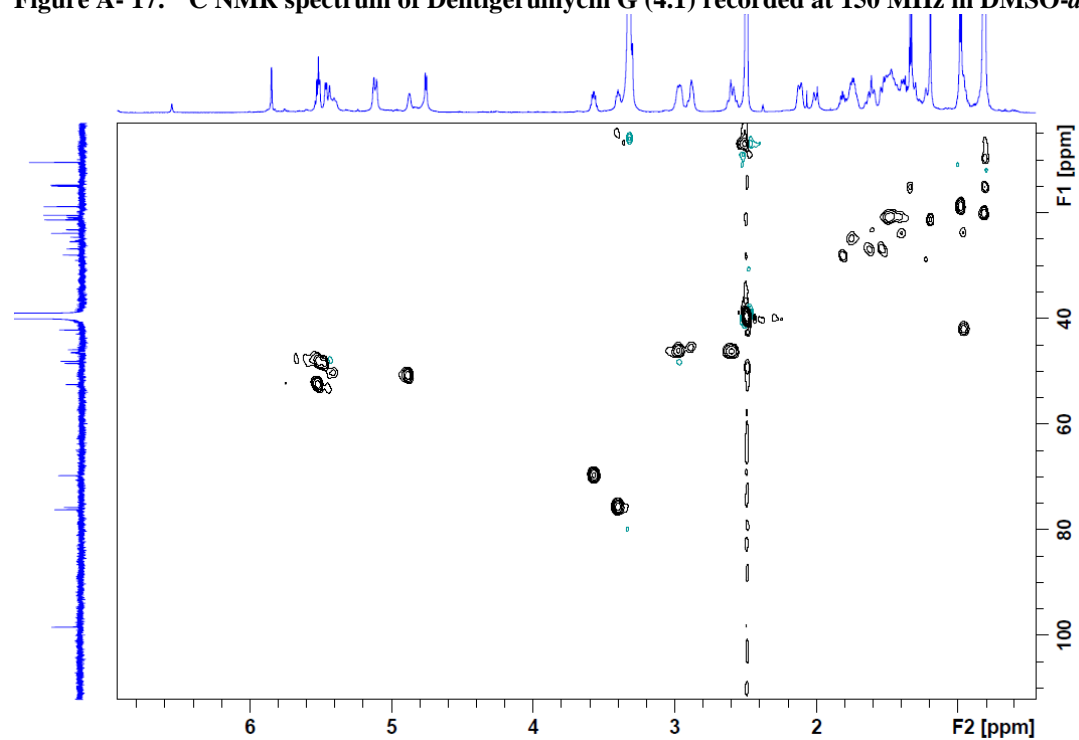
Figure A- 18. ¹H-¹³C HSQC NMR spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.
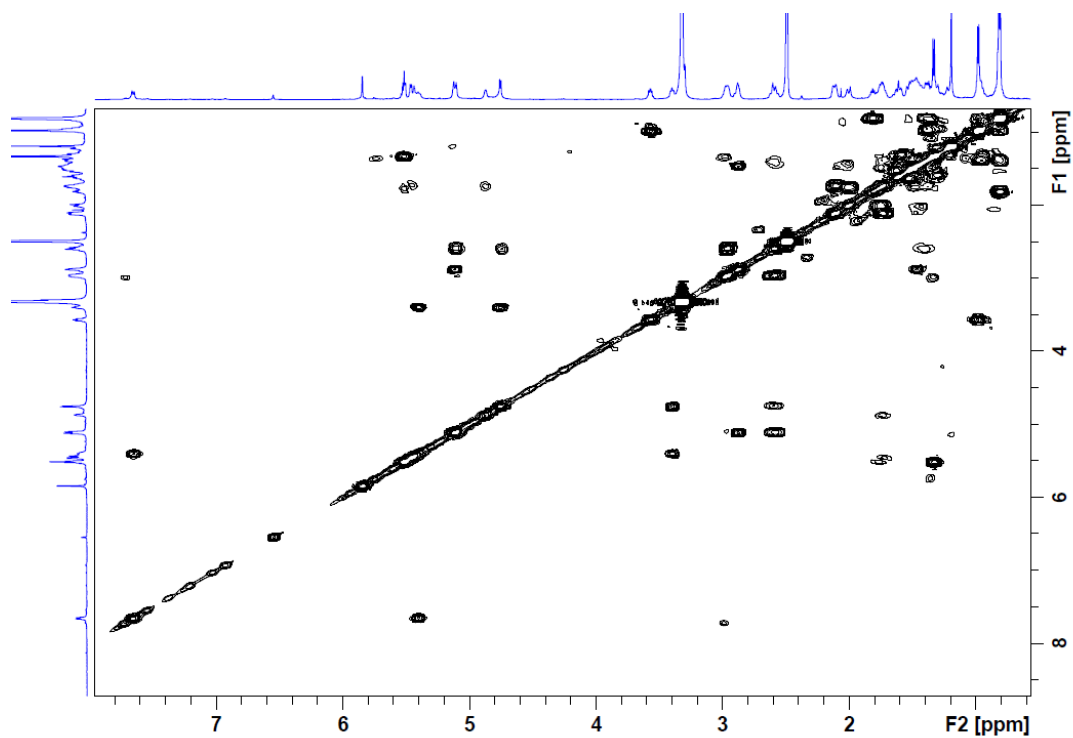
Figure A- 19. ¹H-¹H gCOSY NMR spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.
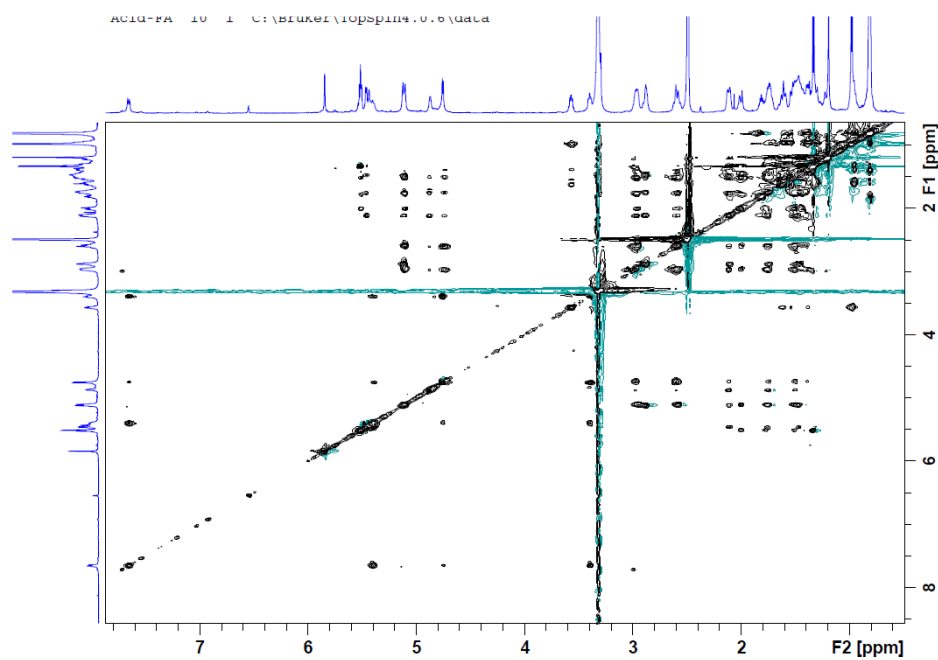
Figure A- 20. ¹H-¹H gTOCSY NMR spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-d₆.
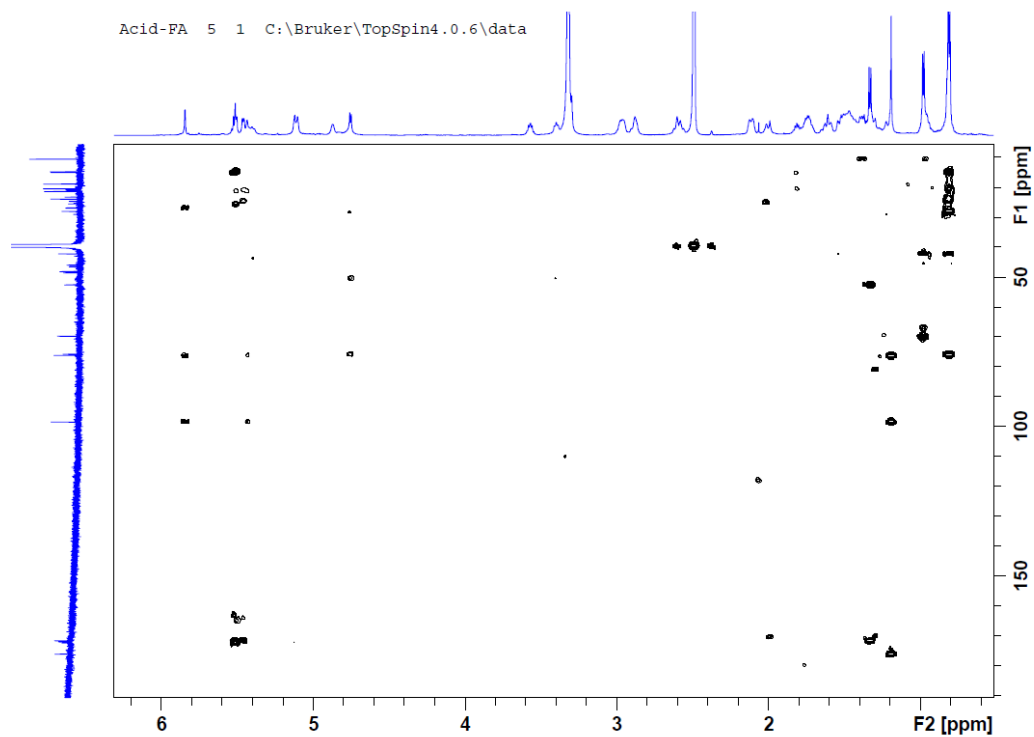
Figure A- 21. ¹H-¹³C gHMBC NMR spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.
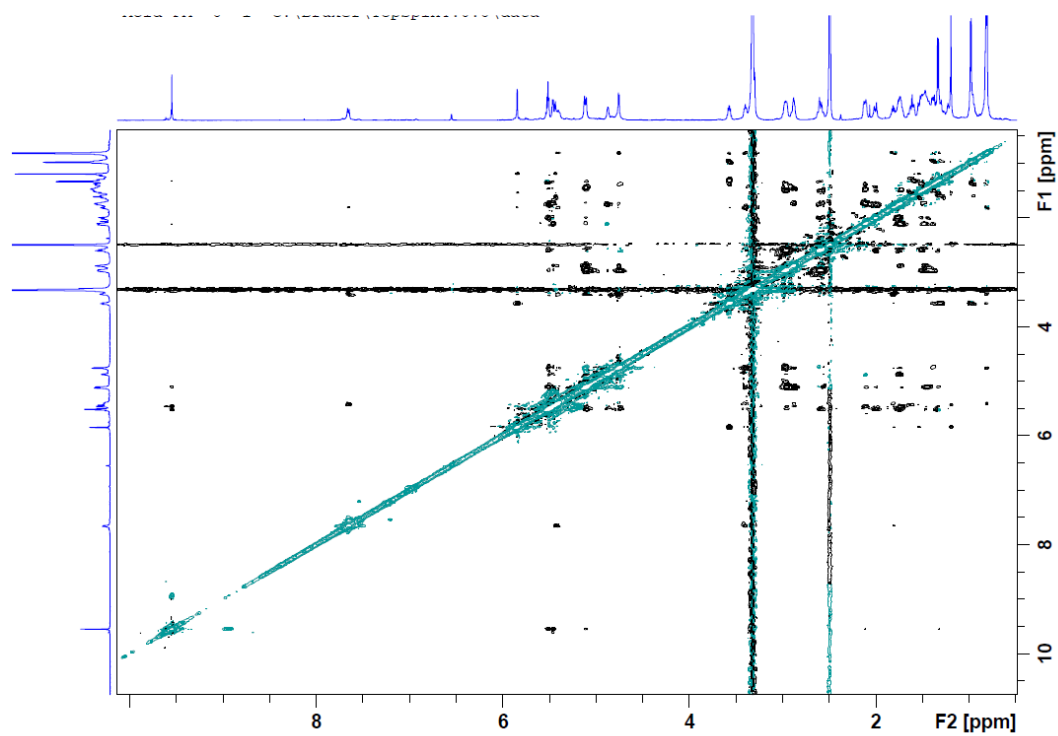
Figure A- 22. tROESY NMR spectrum of Dentigerumycin G (4.1) recorded at 600 MHz in DMSO-*d₆*.

Appendix C NMR spectra of Incarnatapeptin A as taken by Dave Williams, Published
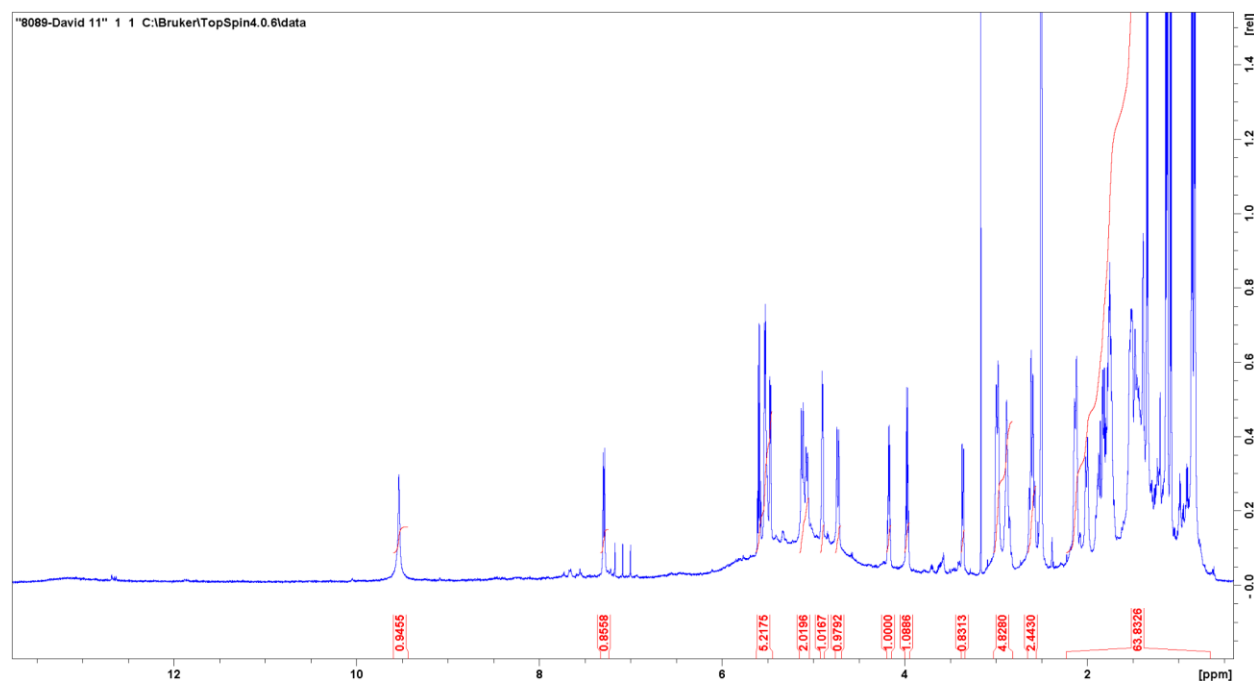
Figure A- 23. ^{1}H NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-d_6.
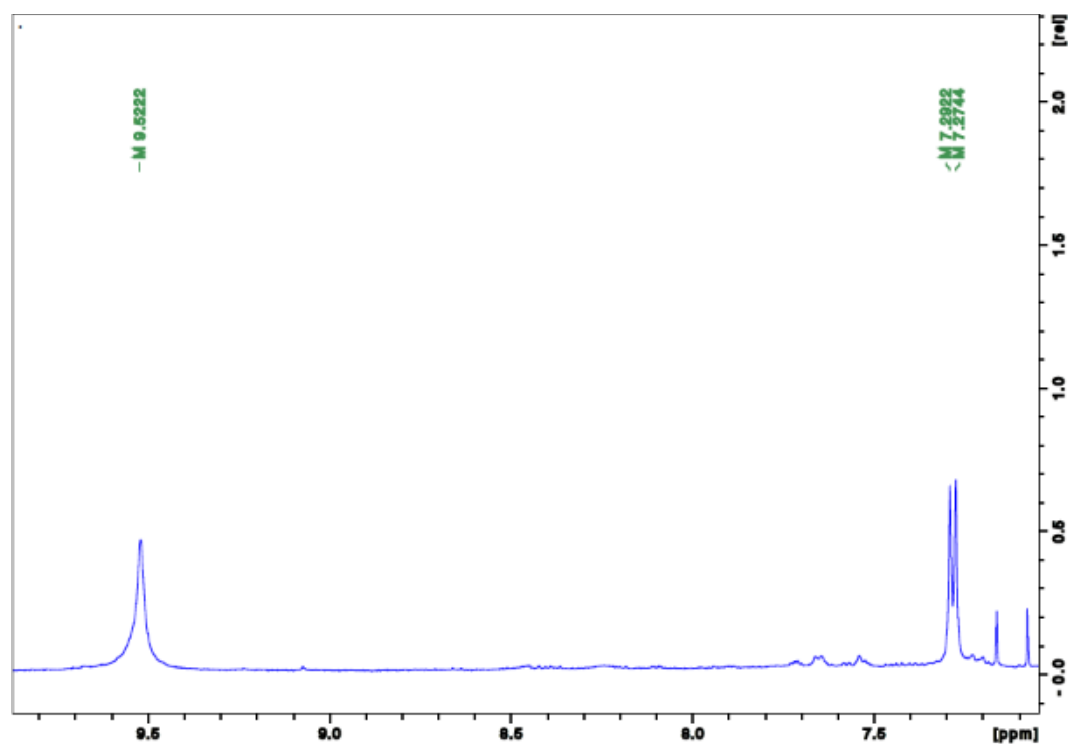
Figure A- 24. Expansion of ^{1}H NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-d_6.
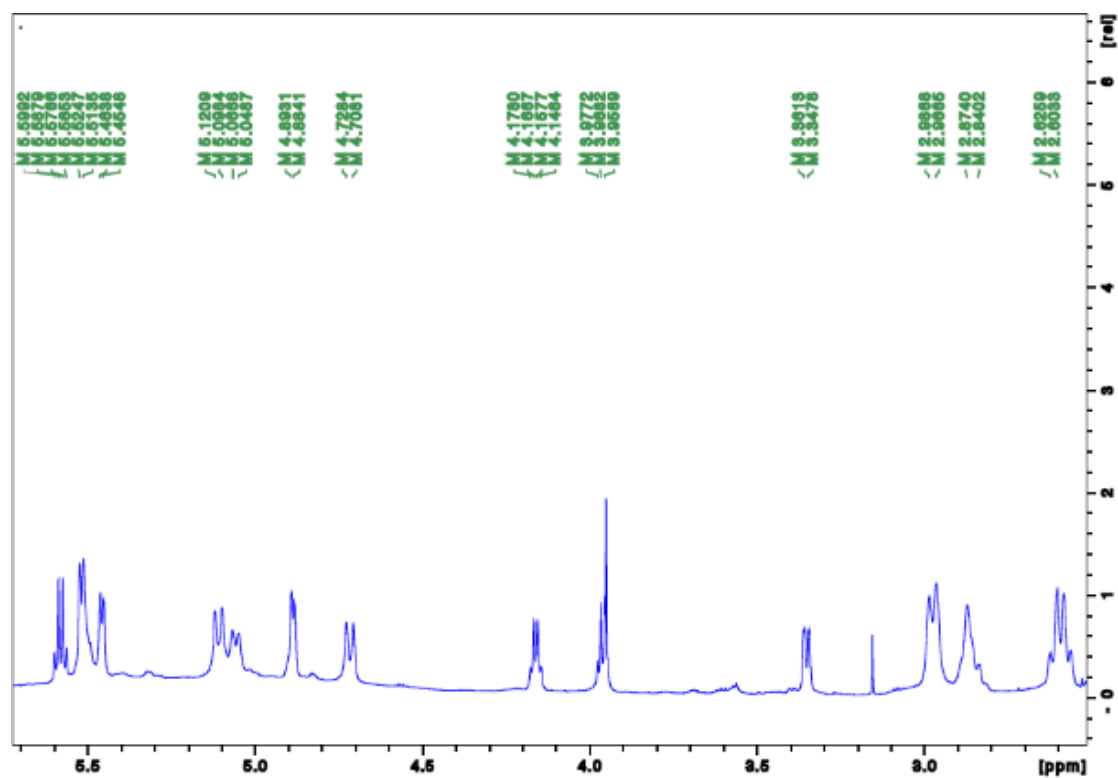
Figure A- 25. Expansion of ^{1}H NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
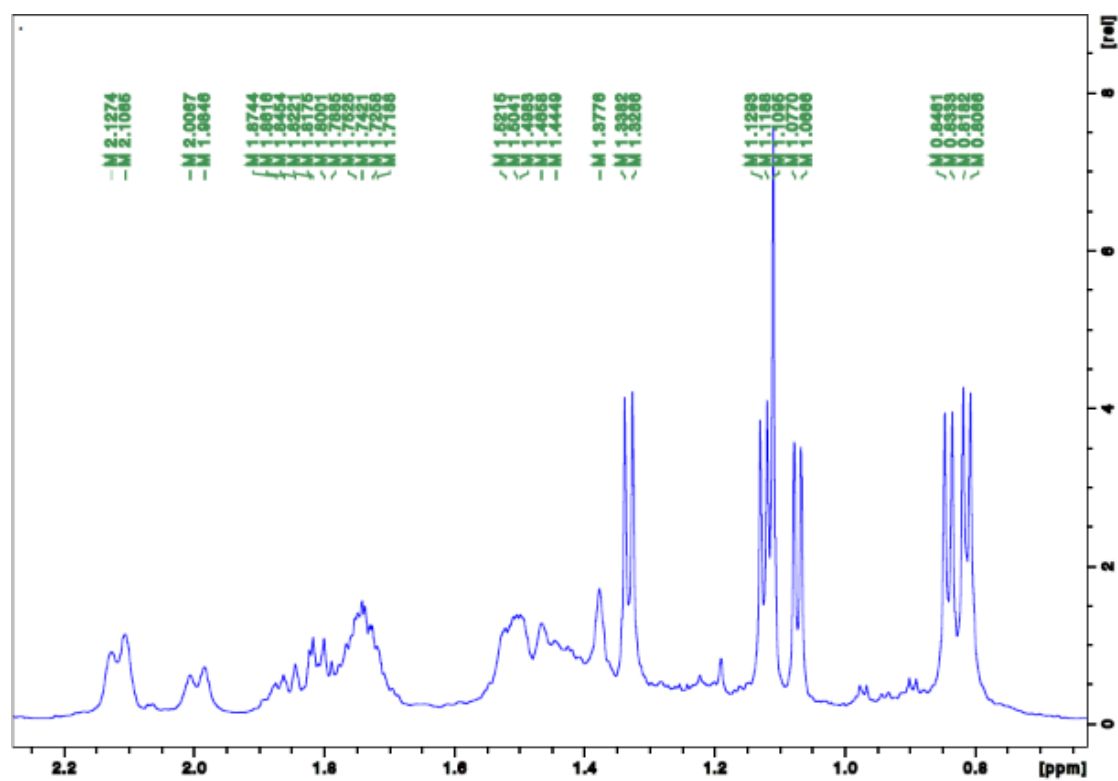
Figure A- 26. Expansion of ^{1}H NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
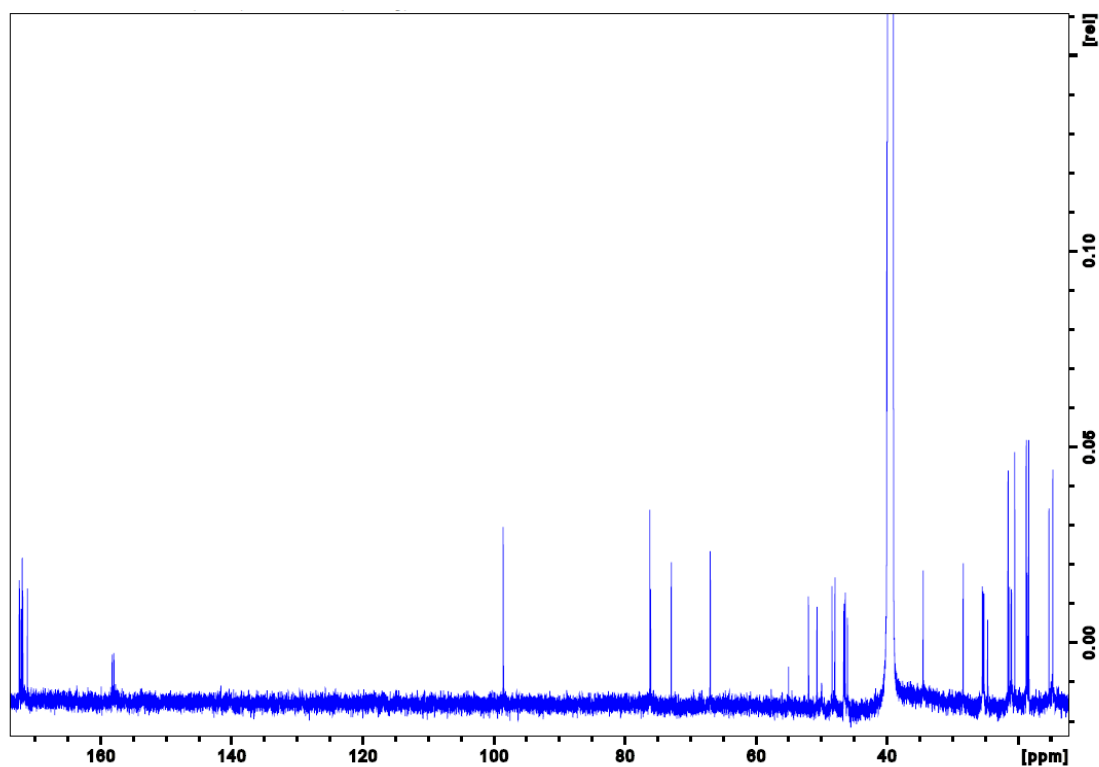
Figure A- 27. ^{13}C NMR spectrum of Incarnatapeptin A (4.3) recorded at 150 MHz in DMSO-d_6.
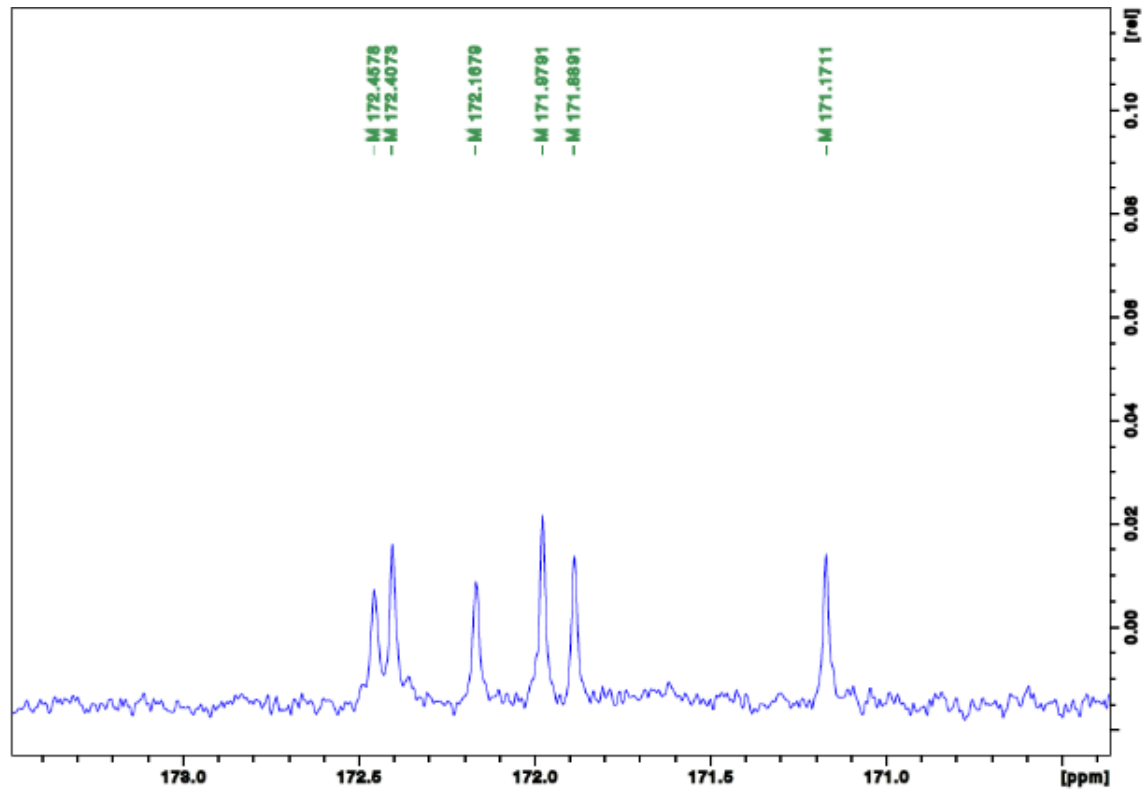
Figure A- 28. Expansion of ^{13}C NMR spectrum of Incarnatapeptin A (4.3) recorded at 150 MHz in DMSO-d_6.
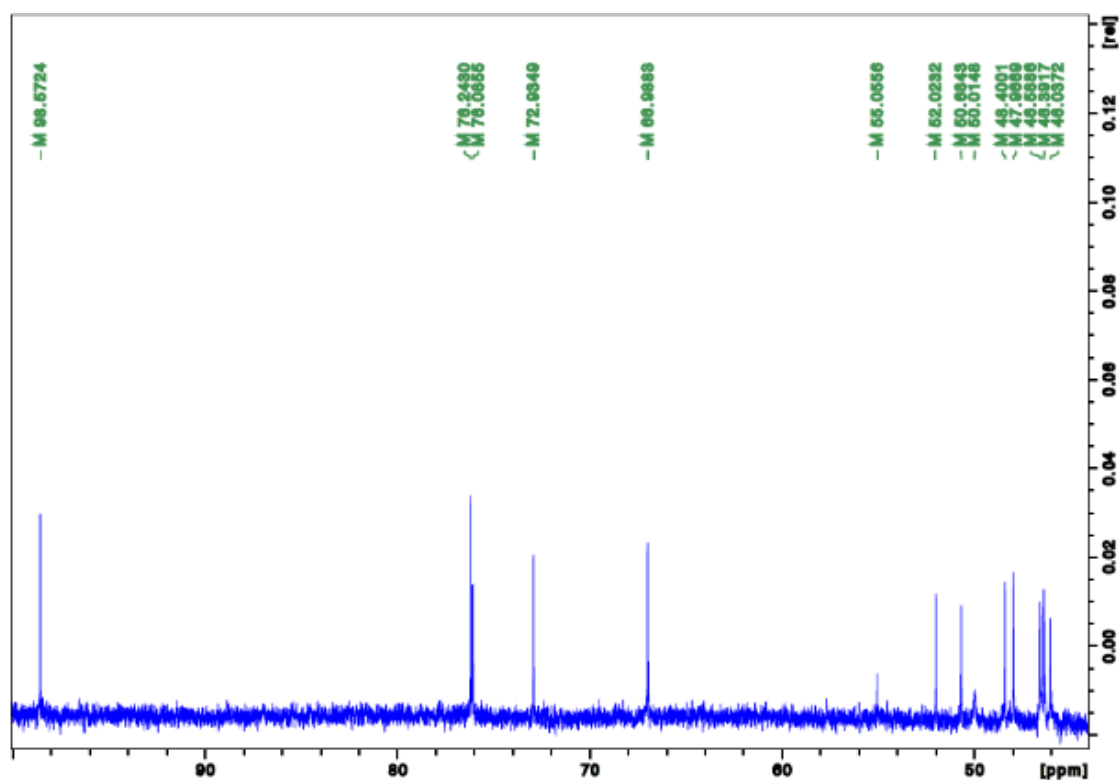
Figure A- 29. Expansion of ^{13}C NMR spectrum of Incarnatapeptin A (4.3) recorded at 150 MHz in DMSO-d_6.
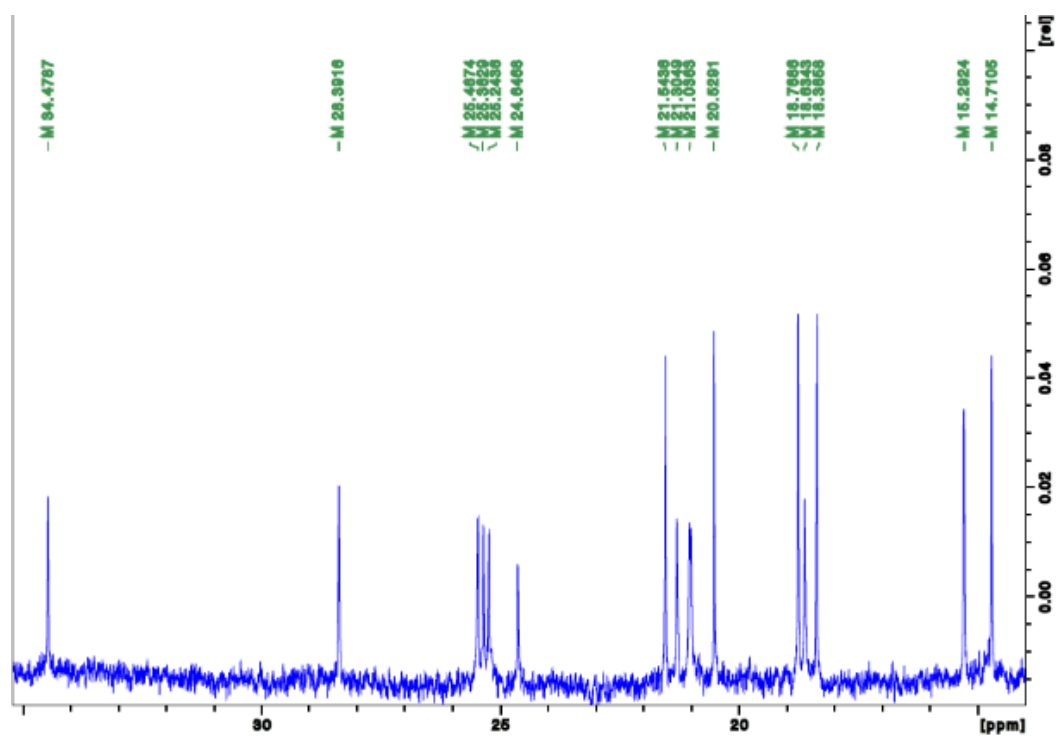
Figure A- 30. Expansion of ^{13}C NMR spectrum of Incarnatapeptin A (4.3) recorded at 150 MHz in DMSO-d_6.

328
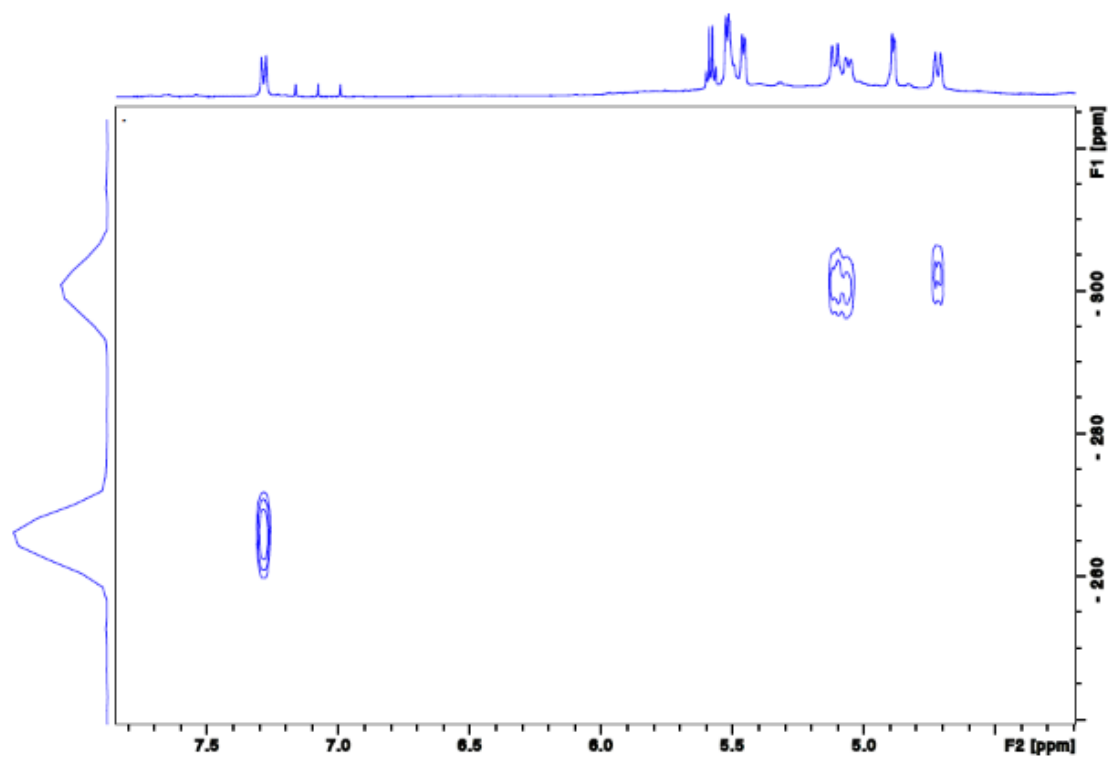
Figure A- 31. ¹H-¹⁵N HSQC NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
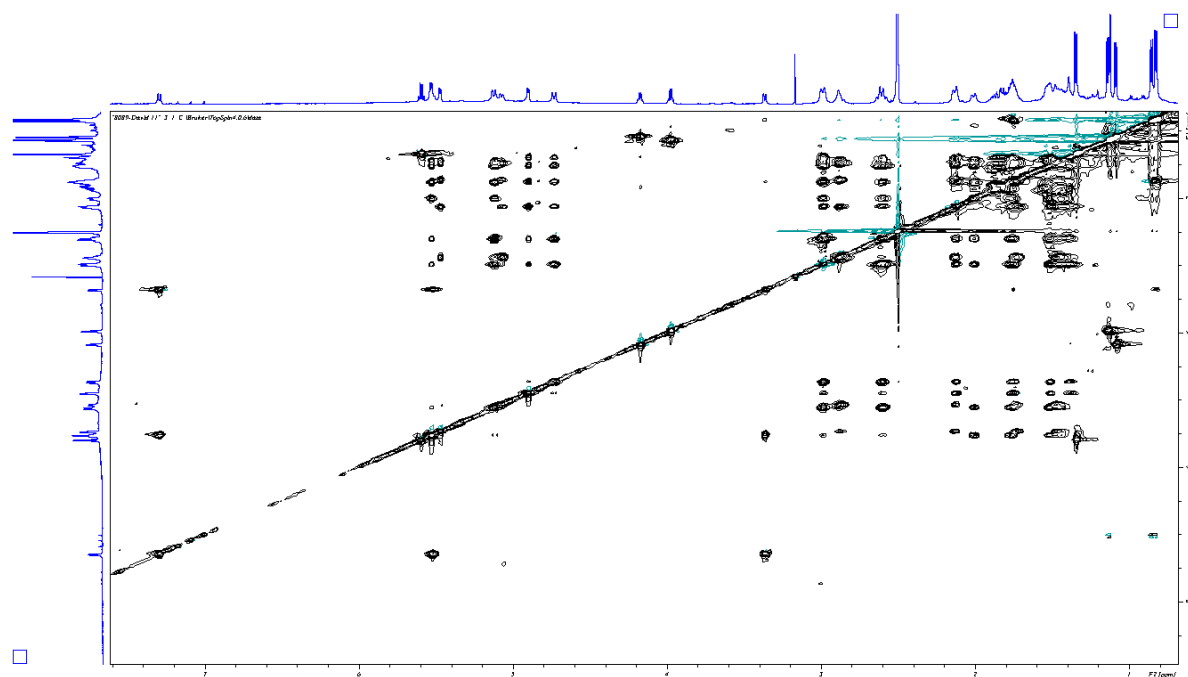
Figure A- 32. ¹H-¹H TOCSY NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
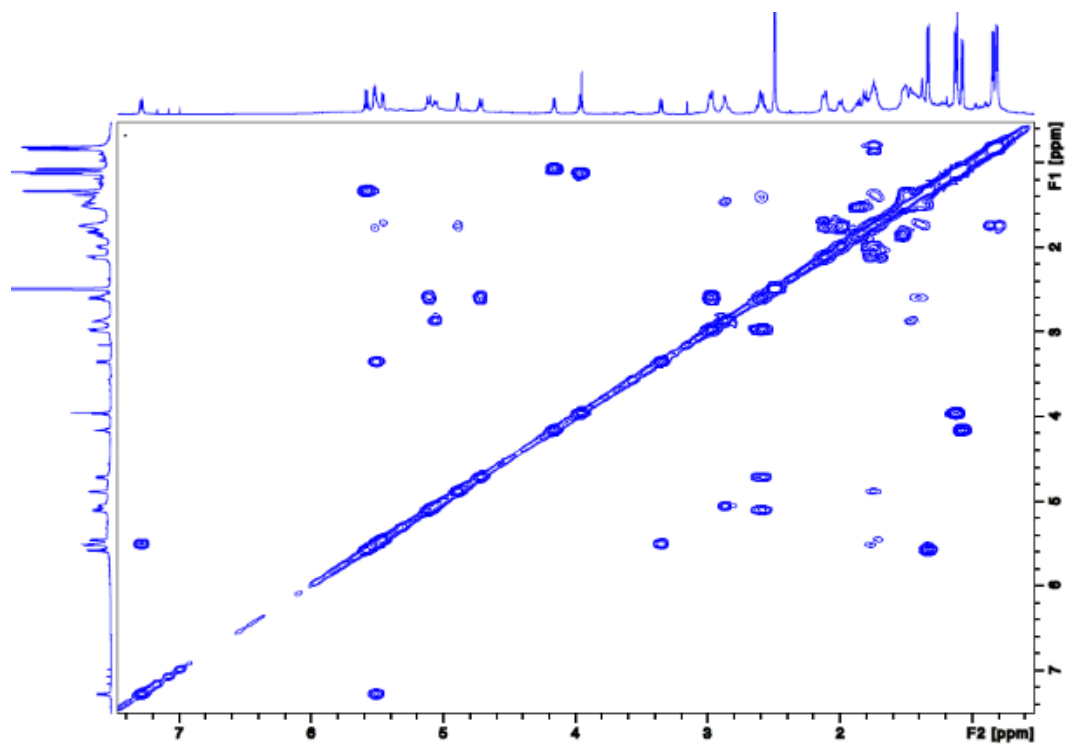
Figure A- 33. g¹H-¹HCOSY60 NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
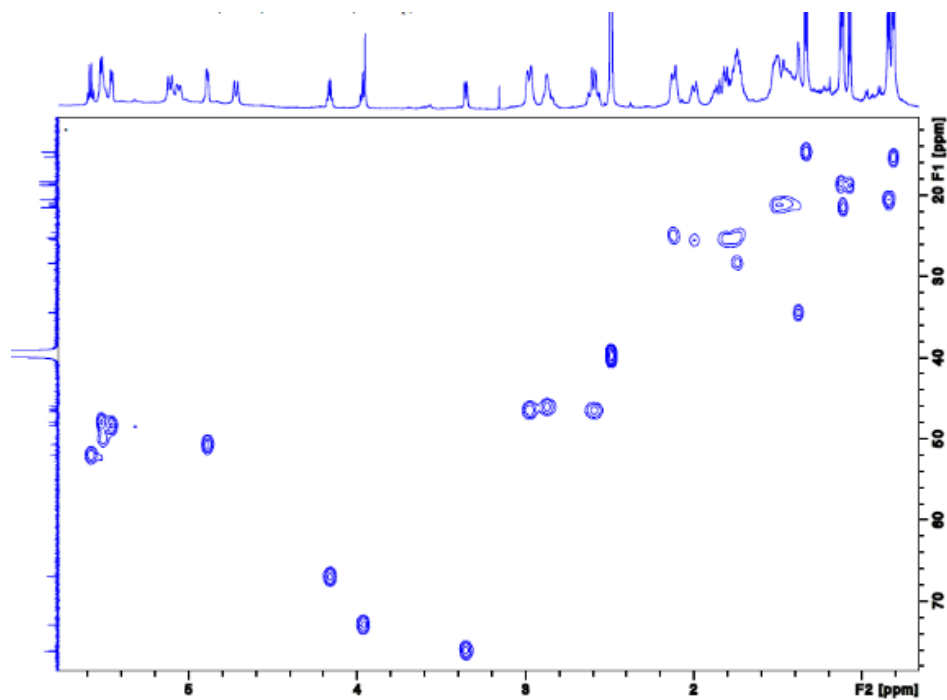
Figure A- 34. ¹H-¹³C HSQC NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
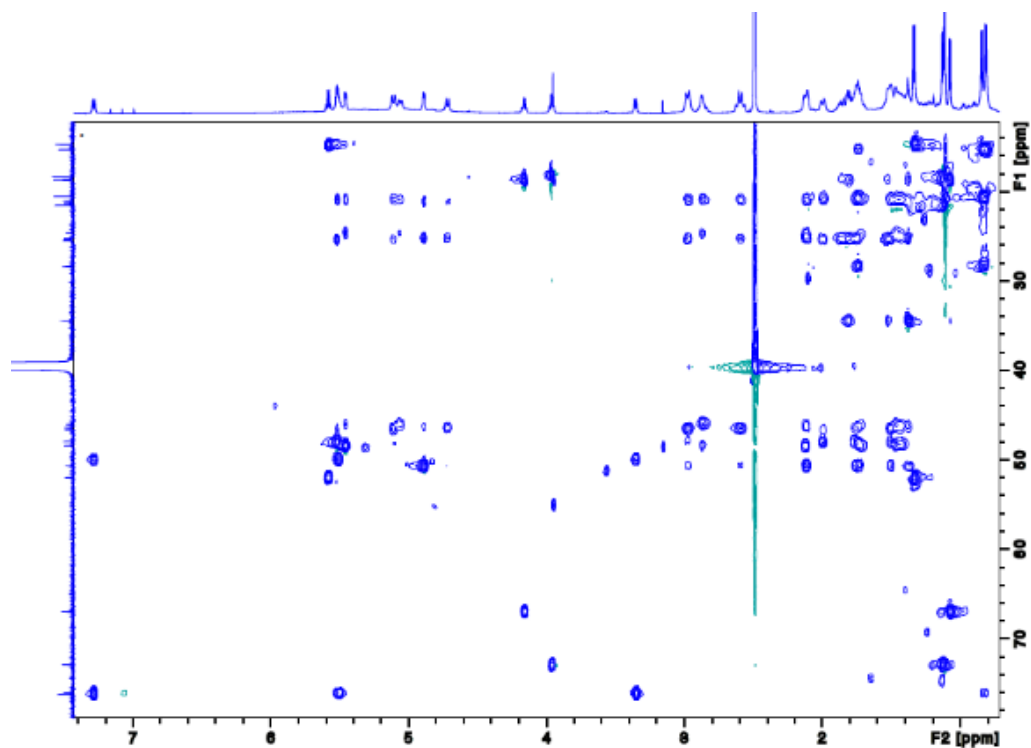
Figure A- 35. ¹H-¹³C HSQC-TOCSY NMR spectrum of incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
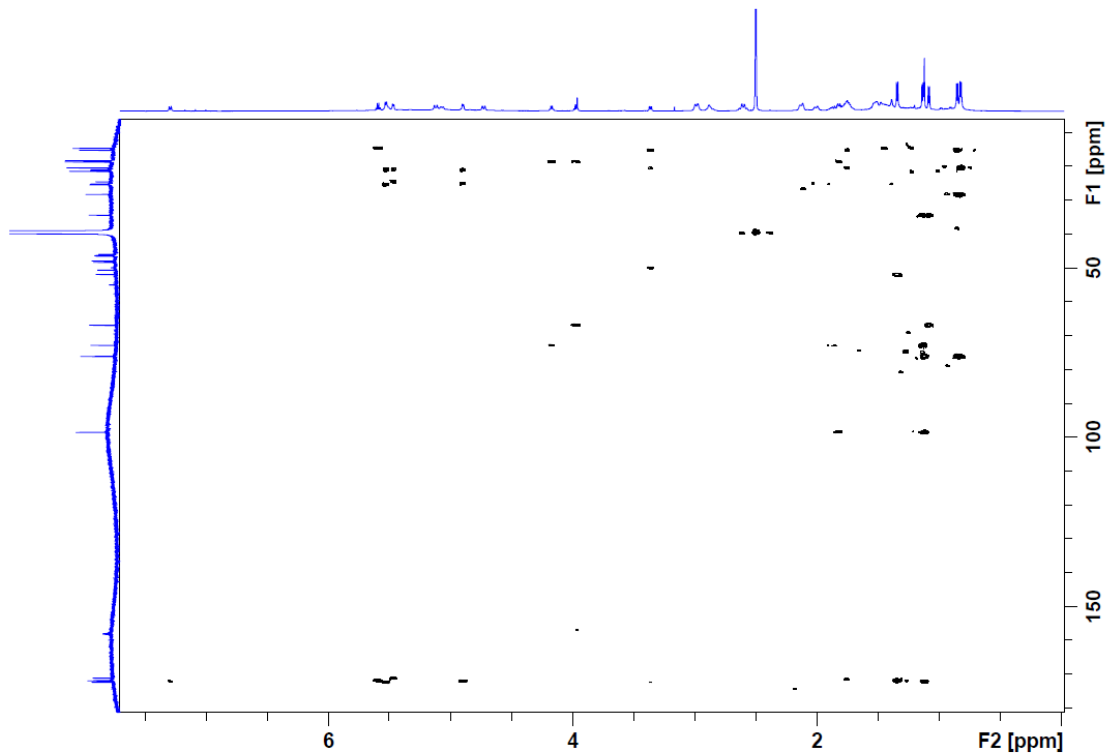
Figure A- 36. g¹H-¹³C HMBC NMR spectrum of incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.
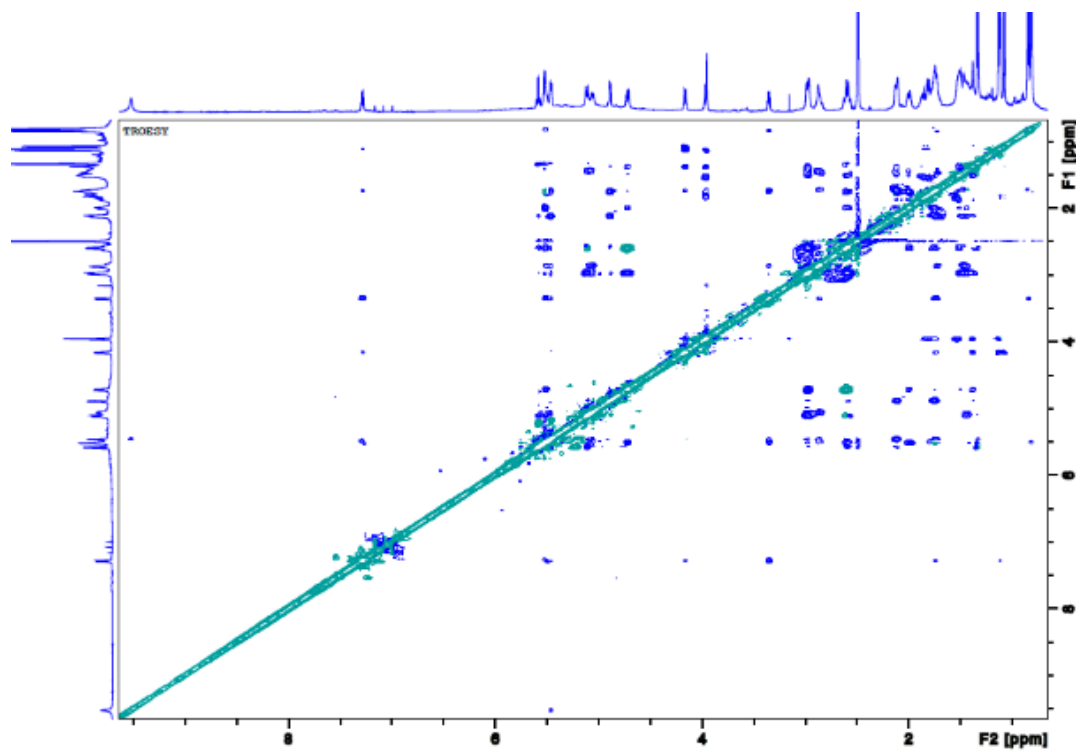
**Figure A- 37. tROESY NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.**
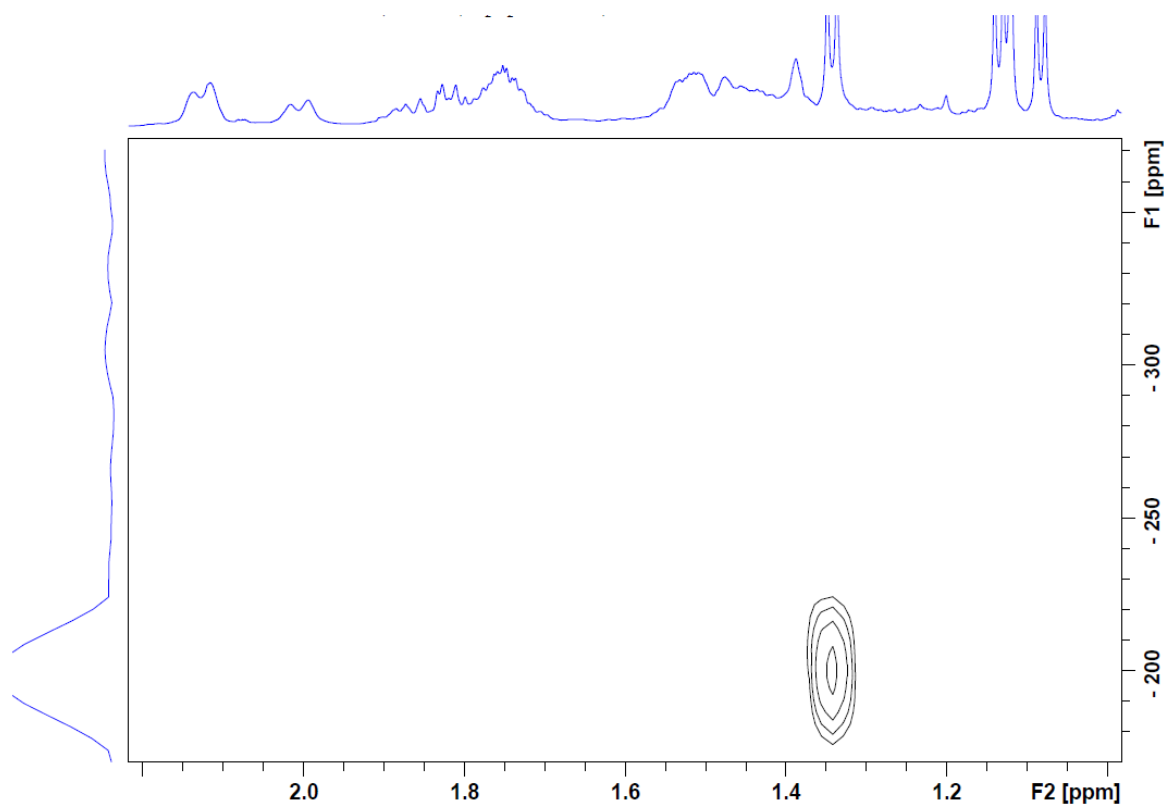
Figure A- 38. LR¹⁵NHMQC NMR spectrum of Incarnatapeptin A (4.3) recorded at 600 MHz in DMSO-*d₆*.

Appendix D Experimental and NMR Spectra of the Methyl Ester* of Incarnatapeptin A

taken by Dave Williams, Published.

*The methyl ester is not in the thesis work but the crystal structure is essential for stereochemical assignment.

EXPERIMENTAL:

Methyl ester of Incarnatapeptin A (A.1): 2.6 mg of incarnatapeptin A (**4.3**) was treated with 0.3 ml of 0.6 M trimethylsilyldiazomethane in hexanes, 0.3 mL of anhydrous MeOH and 0.5 mL of C_6H_6. The reaction mixture was sealed and left stirring for 16 h at rt. After evaporation of the reagents, the sample was purified via C_{18} reversed-phase HPLC using an InertSustain, 5 µm, 25 x 1 cm column with 11:9 MeCN/H_2O as eluent to yield 1.1 mg of product with a positive ion HRMS (TOFESI) *m/z*: [M+Na]+ Calcd for $C_{37}H_{62}N_8O_{12}Na$, 833.4385; Found 833.4390. ^1H (DMSO-d_6, 600 MHz) and ^{13}C NMR (DMSO-d_6, 150 MHz).

A.1
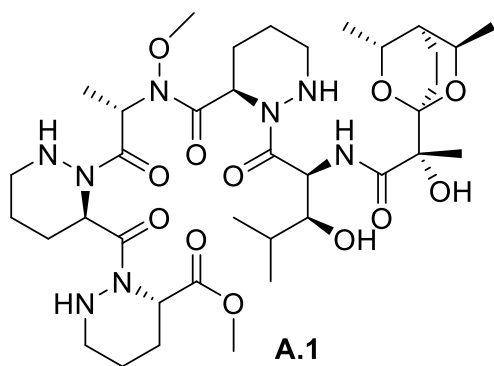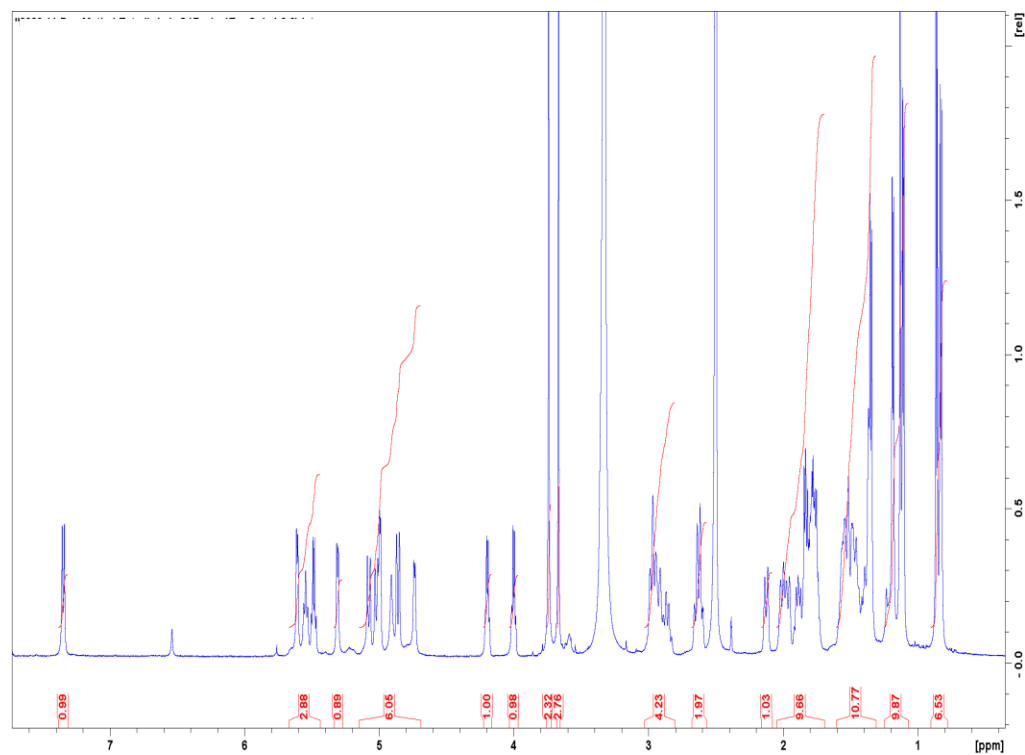
Figure A- 39. ^{1}H NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 600 MHz in DMSO-d_6.
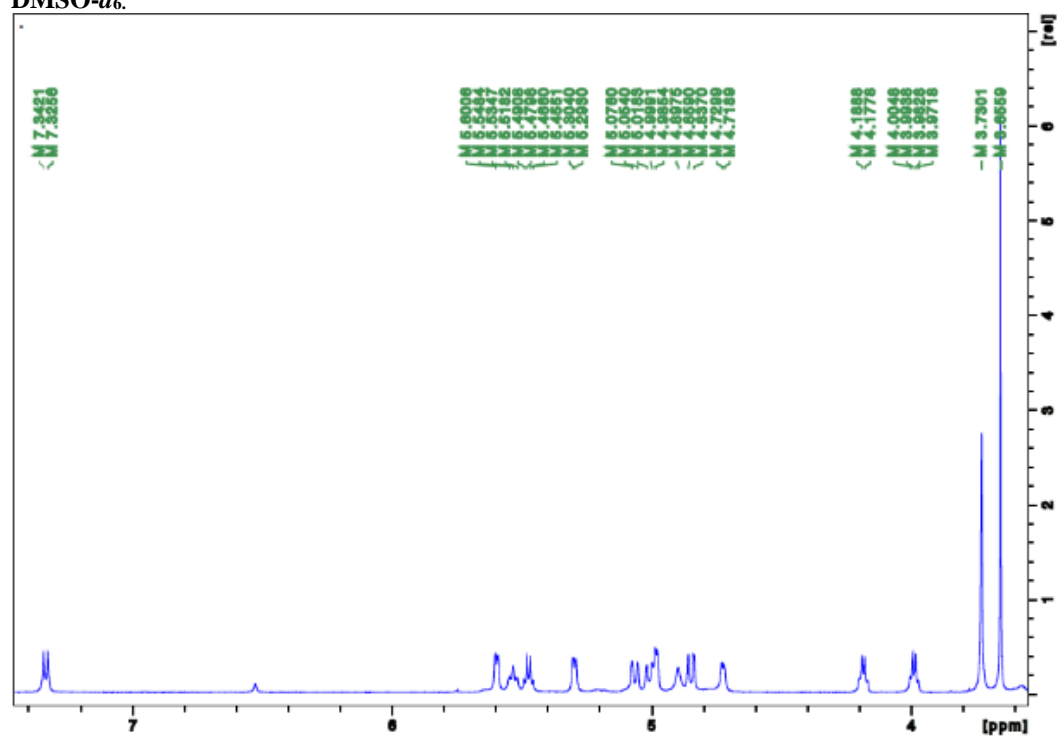
Figure A- 40. Expanded ^{1}H NMR spectrum of the methyl ester of incarnatapeptin A (A.1) with peak picking recorded at 600 MHz in DMSO-d_6.
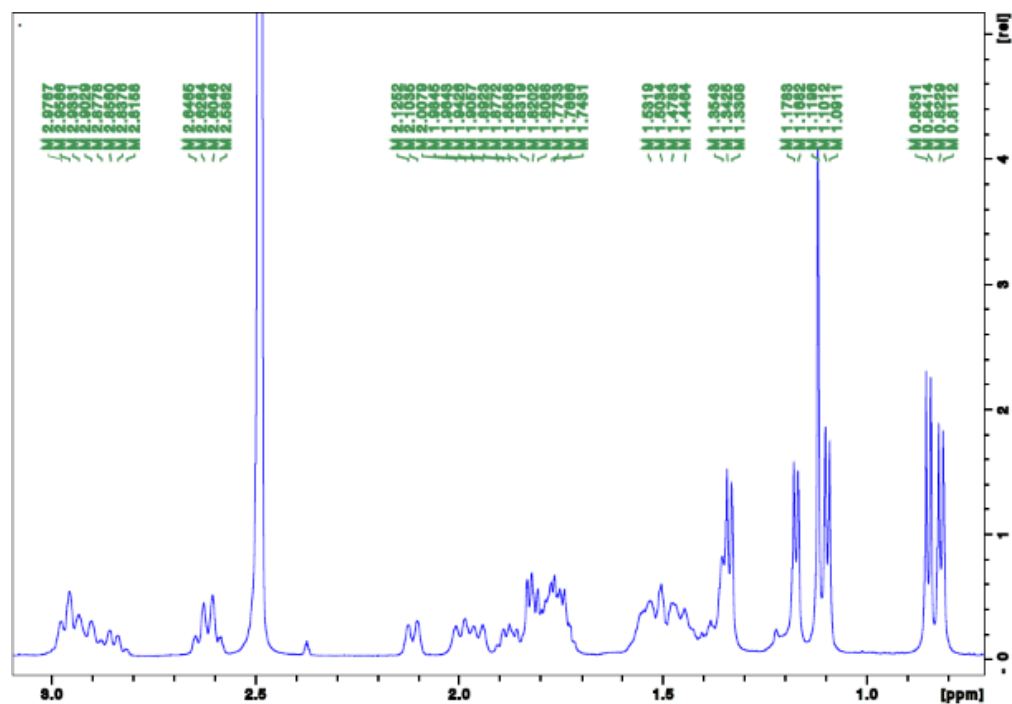
Figure A- 41. Expanded ¹H NMR spectrum of the methyl ester of incarnatapeptin A (A.1) with peak picking recorded at 600 MHz in DMSO-*d₆*.
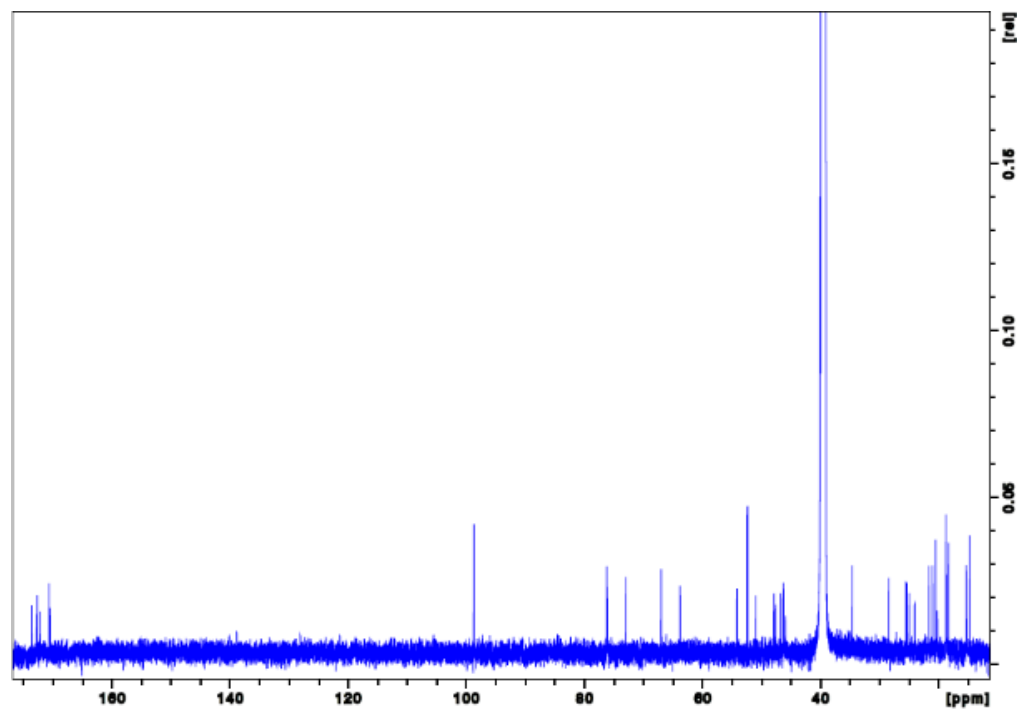
Figure A- 42. ¹³C NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 150 MHz in DMSO-*d₆*.
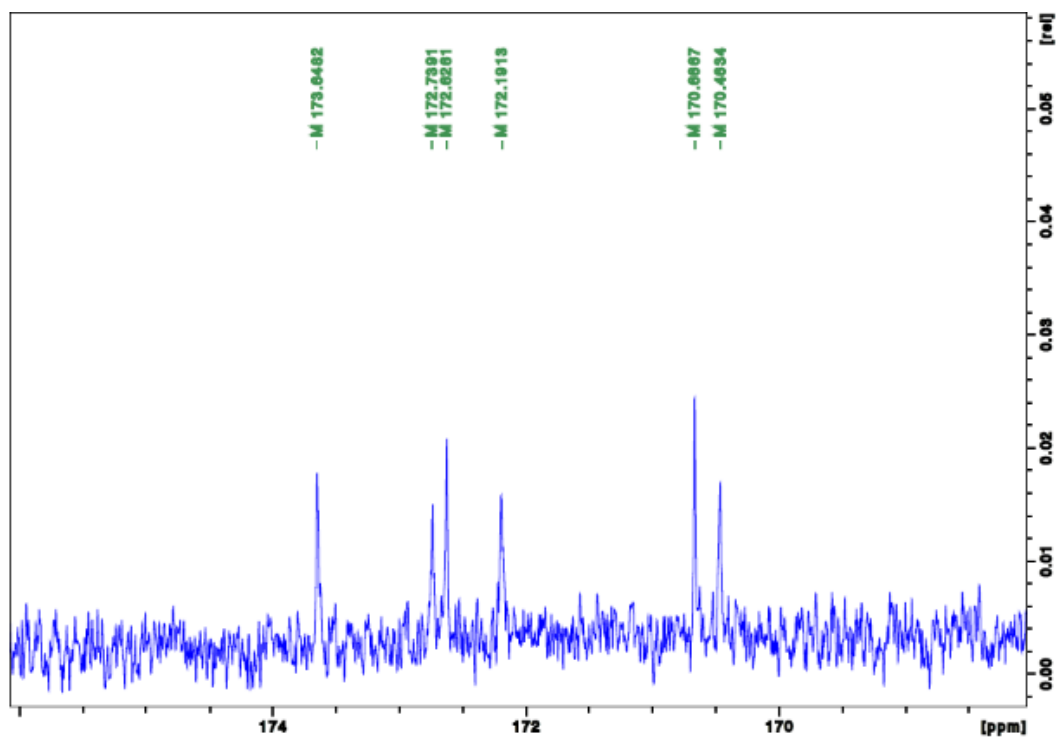
**Figure A- 43.** Expanded ¹³C NMR spectrum of the methyl ester of incarnatapeptin A (A.1) with peak picking recorded at 150 MHz in DMSO-*d₆*.
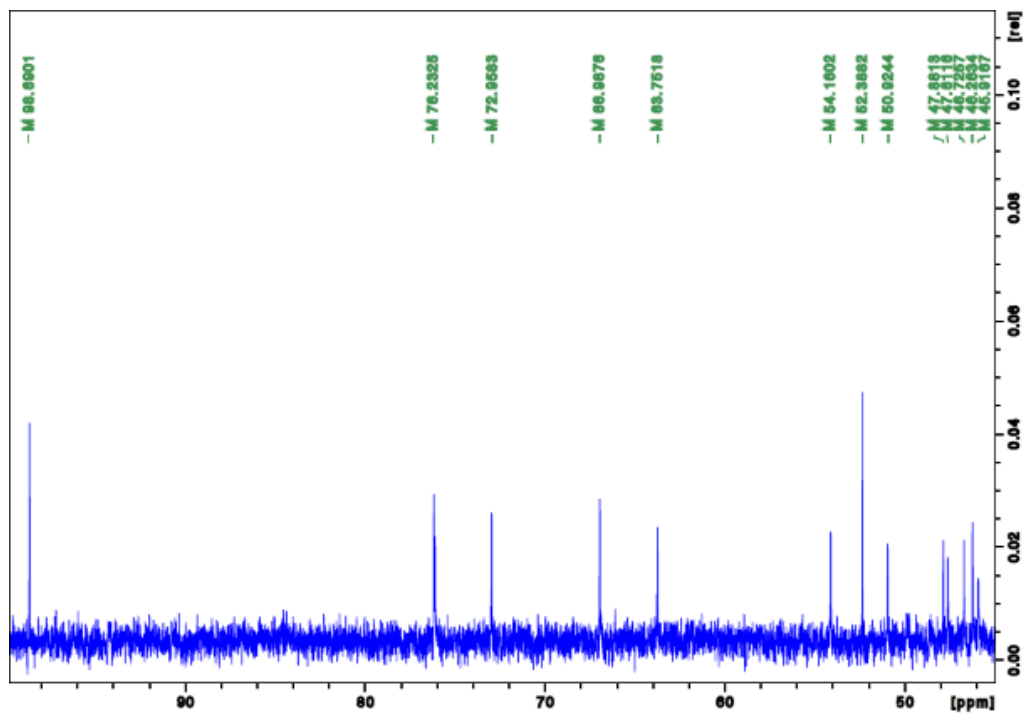
Figure A- 44. Expanded ¹³C NMR spectrum of the methyl ester of incarnatapeptin A (A.1) with peak picking recorded at 150 MHz in DMSO-*d₆*.

336
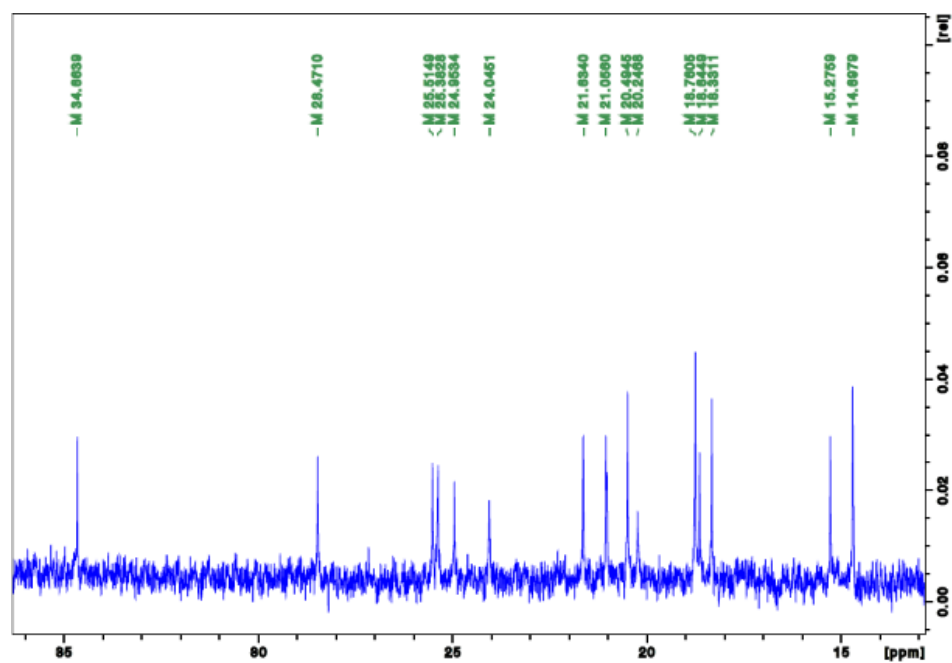
Figure A- 45. Expanded ^{13}C NMR spectrum of the methyl ester of incarnatapeptin A (A.1) with peak picking recorded at 150 MHz in DMSO-d_6.
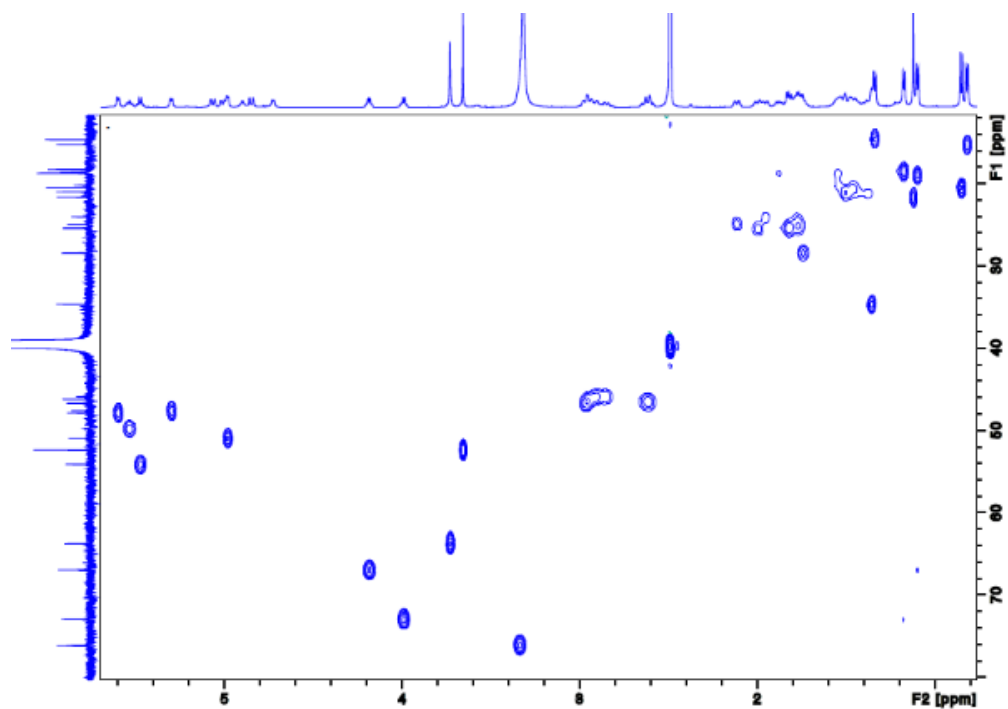
Figure A- 46. ^{1}H-^{13}C HSQC NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 600 MHz in DMSO-d_6.

337
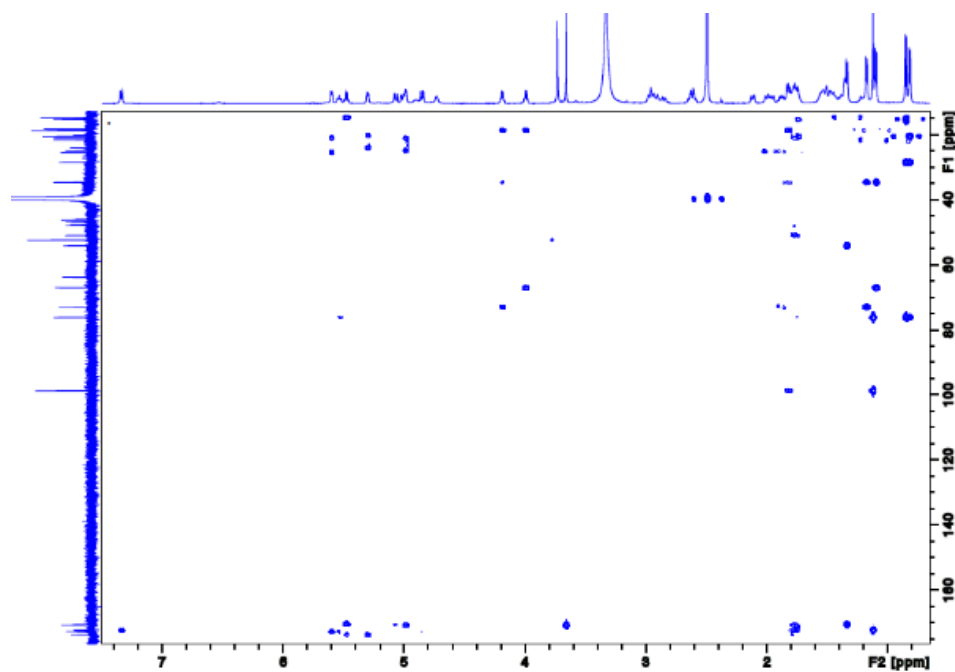
Figure A- 47. ¹H-¹³C gHMBC NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 600 MHz in DMSO-*d₆*.
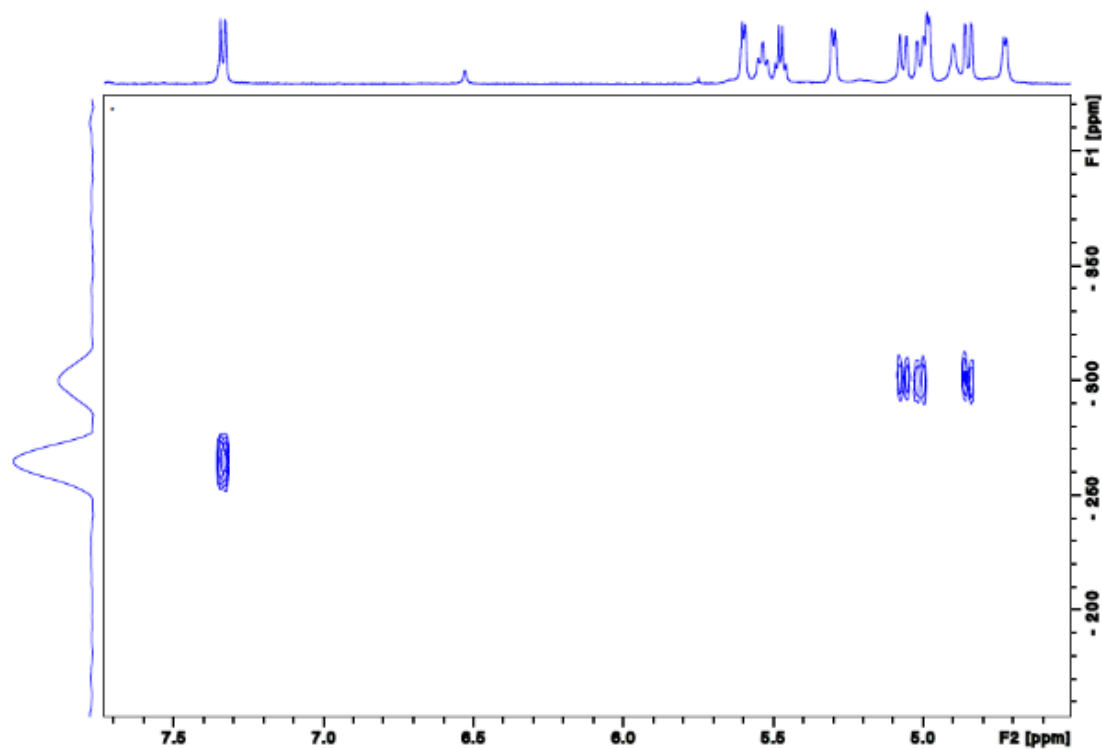
Figure A- 48. ¹H-¹⁵N gHSQC NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 600 MHz in DMSO-*d₆*.

338
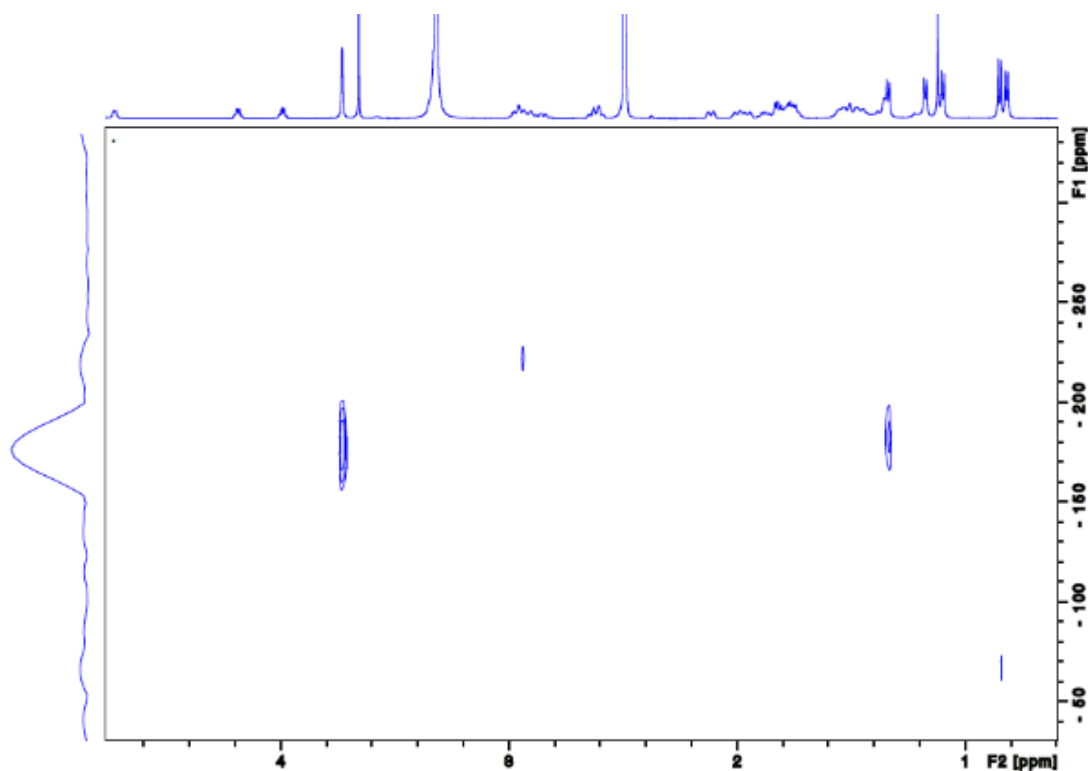
Figure A- 49. ^{1}H-^{15}N LRHMQC NMR spectrum of the methyl ester of incarnatapeptin A (A.1) recorded at 600 MHz in DMSO-d_6.
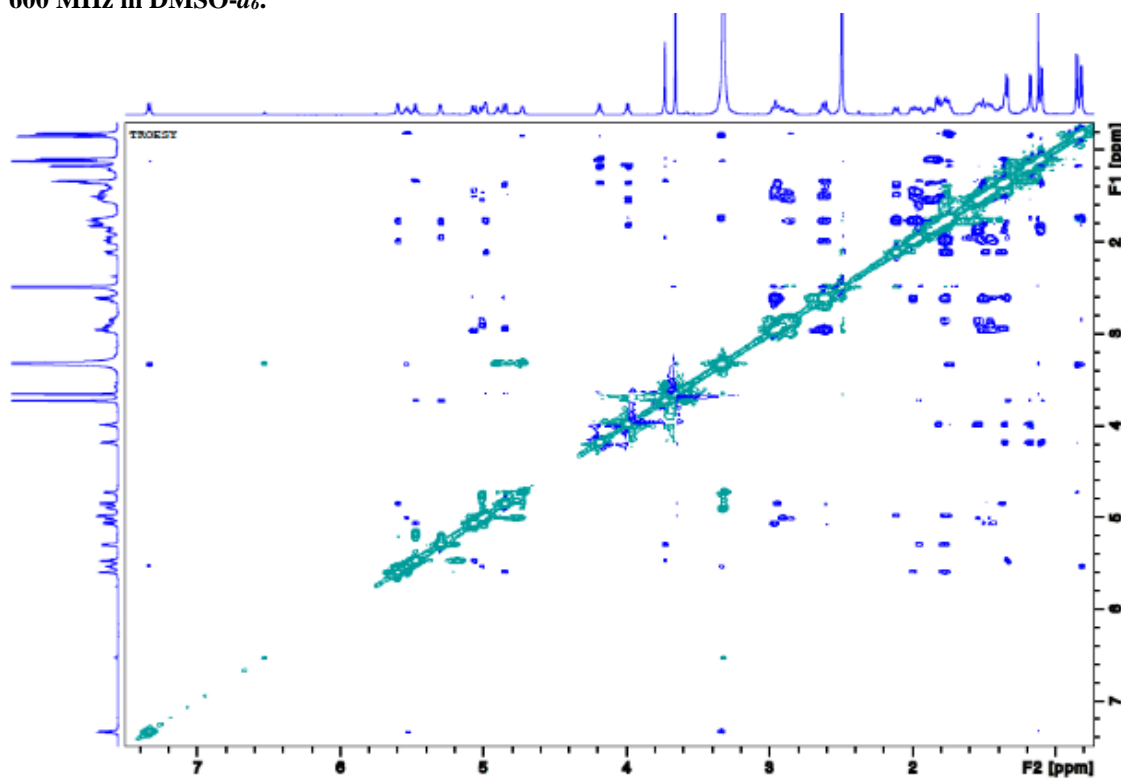
Figure A- 50. tROESY NMR spectrum of the methyl ester of incarnatapeptin (A.1) recorded at 600 MHz in DMSO-d_6.

Appendix E Crystal Data and Experimental Data provided by Dr Brian O. Patrick
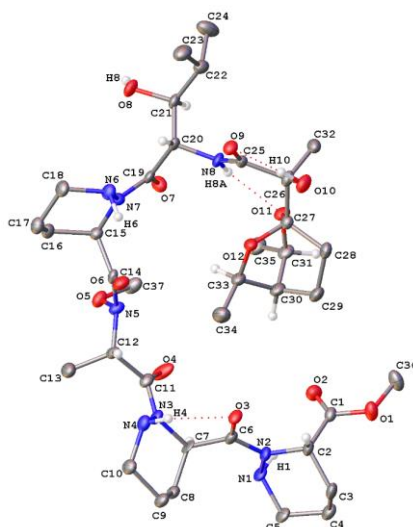
Crystal Data and Experimental

Experimental. Single colourless blade-shaped crystals of ra178 were recrystallised from methanol by slow evaporation. A suitable crystal 0.22×0.14×0.05 mm3 was selected and mounted on a mylar loop in oil on a Bruker APEX II area detector diffractometer. The crystal was kept at a steady T = 90(2) K during data collection. The structure was solved with the XT (Sheldrick, 2015) structure solution program using the Intrinsic Phasing solution method and by using Olex2 (Dolomanov et al., 2009) as the graphical interface. The model was refined with version 2018/3 of XL (Sheldrick, 2015) using Least Squares minimisation.

Crystal Data. $C_{37}H_{62}N_8O_{12}$, Mr = 810.94, triclinic, P1 (No. 1), a = 9.4700(2) Å, b = 11.1262(3) Å, c = 12.0492(3) Å, α = 65.774(2)°, β = 67.771(2)°, γ = 76.655(2)°, V = 1067.41(5) Å3, T = 90(2) K, Z = 1, Z' = 1, μ(CuKα) = 0.787, 34680 reflections measured, 5446 unique (Rint = 0.0868) which were used in all calculations. The final wR2 was 0.0929 (all data) and R1 was 0.0407 (I > 2(I)).

Compound	Methyl ester of Incarnatapeptin A
Formula	C37H62N8O12
Dcalc./ g cm-3	1.262
$\square$/mm-1	0.787
Formula Weight	810.94
Colour	colourless
Shape	blade
Size/mm3	0.22×0.14×0.05
T/K	90(2)
Crystal System	triclinic
Flack Parameter	0.04(14)
Hooft Parameter	0.12(12)
Space Group	P1
a/Å	9.4700(2)
b/Å	11.1262(3)
c/Å	12.0492(3)
$\square$/°	65.774(2)
$\square$/°	67.771(2)
$\square$/°	76.655(2)
V/Å3	1067.41(5)
Z	1
Z'	1
Wavelength/Å	1.54178
Radiation type	CuK$\square$
$\square$min/°	4.246
$\square$max/°	56.286
Measured Refl.	34680
Independent Refl.	5446
Reflections with I > 2(I)	4663
Rint	0.0868
Parameters	543
Restraints	3
Largest Peak	0.167
Deepest Hole	-0.238
GooF	1.053
wR2 (all data)	0.0929
wR2	0.0860
R1 (all data)	0.0547
R1	0.0407

Structure Quality Indicators

Reflections: | d min (Cu) 0.93 | I/σ 18.1 | Rint 8.68% | complete 97%

Refinement: | Shift 0.000 | Max Peak 0.2 | Min Peak -0.2 | GooF 1.053 | Flack .04(14)

A colourless blade-shaped crystal with dimensions 0.22×0.14×0.05 mm3 was mounted on a mylar loop in oil. Data were collected using an Bruker APEX II area detector diffractometer equipped with an Oxford Cryosystems low-temperature device operating at T = 90(2) K. Data were measured using □ and □ scans of 1.0 ° per frame for between 15 and 60 s using CuK□ radiation (microfocus sealed X-ray tube, 45 kV, 0.60 mA). The total number of runs and images was based on the strategy calculation from the program APEX3. The maximum resolution that was achieved was □ = 56.286° (0.93 Å).

The diffraction pattern was indexed and the unit cell was refined using SAINT (Bruker, V8.38A, after 2013) on 9880 reflections, 28% of the observed reflections. Data reduction, scaling and absorption corrections were performed using SAINT (Bruker, V8.38A, after 2013). The final completeness is 98.80 % out to 56.286° in □.

A multi-scan absorption correction was performed using SADABS-2016/2 (Bruker,2016/2) was used for absorption correction. wR2(int) was 0.0935 before and 0.0779 after correction. The ratio of minimum to maximum transmission is 0.8453. The □/2 correction factor is not present. The absorption coefficient □ of this material is 0.787 mm-1 at this wavelength (□ = 1.542Å) and the minimum and maximum transmissions are 0.815 and 0.964.

The structure was solved and the space group P1 (# 1) determined by the XT (Sheldrick, 2015) structure solution program using Intrinsic Phasing and refined by Least Squares using version 2018/3 of XL (Sheldrick, 2015). All non-hydrogen atoms were refined anisotropically. All C—H

hydrogen atom positions were calculated geometrically and refined using the riding model, while all N—H and O—H hydrogen atoms were refined freely.

The Flack parameter was refined to 0.04(14). Determination of absolute structure using Bayesian statistics on Bijvoet differences using the Olex2 results in 0.12(12). Note: The Flack parameter is used to determine chirality of the crystal studied, the value should be near 0, a value of 1 means that the stereochemistry is wrong and the model should be inverted. A value of 0.5 means that the crystal consists of a racemic mixture of the two enantiomers.

Table 7-1: Fractional Atomic Coordinates (×104) and Equivalent Isotropic Displacement Parameters (Å2×103) for ra178. Ueq is defined as 1/3 of the trace of the orthogonalised Uij.

Atom	x	y	z	Ueq
O1	11222(4)	2612(4)	3001(3)	29.8(10)
O2	9981(4)	4361(4)	3566(3)	28.9(9)
O3	7760(3)	5561(3)	1164(3)	21.5(8)
O4	6121(4)	7999(5)	3000(4)	38.5(11)
O5	2849(4)	8746(4)	4412(3)	24.1(9)
O6	1558(4)	6658(4)	3558(3)	25.7(9)
O7	933(4)	5572(3)	7584(3)	25.6(9)
O8	-2602(4)	4721(4)	9412(4)	26.3(9)
O9	-124(4)	1634(3)	7280(3)	23.3(9)
O10	2384(4)	153(4)	6854(4)	30.4(10)
O11	4038(4)	2413(3)	7347(3)	23.9(9)
O12	3412(4)	2860(3)	5522(3)	23.8(9)
N1	10948(4)	6767(4)	1102(5)	21.4(11)
N2	10038(5)	5893(4)	1125(4)	20.4(10)
N3	6469(4)	8033(4)	1037(4)	19.9(10)
N4	5810(5)	7891(5)	216(4)	22.5(11)
N5	2953(4)	7908(4)	3759(4)	19.2(10)
N6	-1231(5)	5717(4)	5669(4)	21.4(10)
N7	-294(4)	6112(4)	6134(4)	18.0(10)
N8	802(5)	2977(4)	7800(4)	19.3(11)
C1	10560(6)	3853(5)	2768(5)	20.4(13)
C2	10616(6)	4508(5)	1377(5)	21.9(13)
C3	12225(6)	4411(5)	413(5)	26.0(14)
C4	13252(6)	5275(6)	430(6)	28.2(14)
C5	12506(5)	6685(6)	199(5)	27.3(14)
C6	8620(6)	6328(5)	1037(5)	20.8(13)
C7	8152(5)	7801(5)	749(5)	19.7(13)
C8	8909(6)	8635(6)	-655(5)	22.6(13)

Atom	x	y	z	Ueq
C9	8286(6)	8447(6)	-1566(5)	26.5(14)
C10	6539(6)	8707(6)	-1121(5)	25.9(13)
C11	5566(6)	8097(5)	2195(5)	22.3(13)
C12	3844(5)	8368(5)	2387(5)	21.0(13)
C13	3437(6)	9838(6)	1723(6)	32.0(15)
C14	1678(5)	7262(5)	4172(5)	18.0(12)
C15	404(5)	7359(5)	5374(5)	19.3(12)
C16	-806(6)	8495(5)	4990(5)	22.2(13)
C17	-1767(6)	8121(6)	4428(5)	24.8(13)
C18	-2403(6)	6809(5)	5335(5)	26.5(13)
C19	62(5)	5275(5)	7207(5)	18.9(12)
C20	-557(5)	3925(5)	7910(5)	17.9(12)
C21	-1370(6)	3700(5)	9335(5)	21.6(13)
C22	-1943(6)	2318(5)	10133(5)	25.9(13)
C23	-3243(6)	2109(6)	9800(6)	36.4(16)
C24	-2412(8)	2059(7)	11571(6)	43.9(17)
C25	915(6)	1948(5)	7471(5)	18.5(12)
C26	2420(6)	1026(5)	7424(5)	21.9(13)
C27	3814(6)	1844(5)	6569(5)	21.3(13)
C28	5287(6)	1077(5)	6014(5)	26.6(14)
C29	6474(6)	2087(6)	5071(6)	30.8(15)
C30	5854(6)	3401(6)	5295(5)	28.4(14)
C31	5434(6)	3098(6)	6724(5)	26.2(14)
C32	2458(6)	249(5)	8785(5)	28.6(14)
C33	4396(6)	3948(6)	4913(5)	27.0(14)
C34	4610(7)	4485(7)	3504(6)	40.4(17)
C35	5125(6)	4242(6)	7167(6)	35.2(16)
C36	11276(7)	1905(6)	4301(6)	35.7(16)
C37	3601(6)	8080(6)	5397(5)	29.3(14)

Table 7-2: Anisotropic Displacement Parameters (×104) ra178. The anisotropic displacement factor exponent takes the form: -2□2[h2a*2 × U11+... +2hka* × b* × U1

Atom	U11	U22	U33	U23	U13	U12
O1	36(2)	26(2)	30(2)	-11.3(18)	-18.8(19)	8.1(19)
O2	29(2)	29(2)	26(2)	-11(2)	-6.4(18)	0.3(18)
O3	12.5(19)	22(2)	32(2)	-11.4(17)	-6.4(16)	-2.4(16)
O4	14(2)	77(3)	31(2)	-27(2)	-8.2(18)	-1.1(19)
O5	18.2(19)	29(2)	32(2)	-18.4(19)	-8.9(17)	1.8(16)
O6	16.3(19)	32(2)	28(2)	-13(2)	-2.4(17)	-4.0(16)
O7	24(2)	25(2)	29(2)	-4.1(17)	-14.2(17)	-6.1(16)
O8	19.1(19)	29(2)	35(2)	-21.0(18)	-7.6(17)	6.0(16)
O9	16.2(19)	24(2)	30(2)	-10.4(18)	-6.4(16)	-3.5(16)
O10	23(2)	28(2)	43(3)	-21(2)	-8.3(19)	3.7(17)
O11	14.1(19)	29(2)	27(2)	-10.3(18)	-3.5(16)	-5.9(16)

Atom	U11	U22	U33	U23	U13	U12
O12	15.0(19)	25(2)	24(2)	-2.8(17)	-3.8(16)	-2.1(16)
N1	9(2)	26(3)	30(3)	-10(2)	-7(2)	-0.6(19)
N2	13(2)	20(3)	27(3)	-8(2)	-9(2)	2.8(19)
N3	11(2)	27(3)	26(3)	-13(2)	-7(2)	1.5(19)
N4	17(2)	27(3)	29(3)	-12(2)	-12(2)	1(2)
N5	13(2)	27(3)	22(3)	-13(2)	-3.8(19)	-4.9(19)
N6	18(2)	27(3)	26(3)	-11(2)	-13(2)	0(2)
N7	16(2)	20(2)	19(3)	-6(2)	-5.5(19)	-5(2)
N8	12(2)	20(3)	27(3)	-12(2)	-6(2)	3.4(19)
C1	15(3)	18(3)	28(3)	-8(3)	-8(2)	0(2)
C2	16(3)	21(3)	27(3)	-10(3)	-9(2)	6(2)
C3	21(3)	25(3)	29(3)	-12(3)	-10(3)	10(2)
C4	17(3)	34(4)	32(3)	-14(3)	-8(2)	4(3)
C5	10(3)	37(4)	31(3)	-9(3)	-8(2)	2(2)
C6	14(3)	26(3)	20(3)	-10(3)	-4(2)	4(3)
C7	3(3)	30(3)	27(3)	-12(3)	-5(2)	2(2)
C8	13(3)	24(3)	27(3)	-7(3)	-4(2)	-1(2)
C9	14(3)	35(4)	26(3)	-9(3)	-6(2)	2(2)
C10	16(3)	32(3)	27(3)	-6(3)	-10(2)	0(2)
C11	13(3)	25(3)	29(3)	-10(3)	-6(3)	-3(2)
C12	11(3)	28(3)	20(3)	-7(3)	-4(2)	-1(2)
C13	18(3)	32(4)	33(4)	-3(3)	-7(3)	0(3)
C14	15(3)	18(3)	22(3)	-9(3)	-9(2)	5(2)
C15	14(3)	21(3)	20(3)	-6(2)	-5(2)	-2(2)
C16	20(3)	21(3)	25(3)	-9(3)	-6(2)	2(2)
C17	20(3)	27(3)	31(3)	-11(3)	-13(3)	2(2)
C18	17(3)	29(3)	33(3)	-7(3)	-13(3)	1(2)
C19	11(3)	21(3)	22(3)	-10(3)	-4(2)	5(2)
C20	14(3)	19(3)	19(3)	-9(2)	-5(2)	4(2)
C21	15(3)	21(3)	28(3)	-12(3)	-6(2)	5(2)
C22	23(3)	25(3)	26(3)	-10(3)	-1(3)	-3(2)
C23	24(3)	36(4)	44(4)	-14(3)	-2(3)	-8(3)
C24	59(4)	37(4)	25(3)	-7(3)	-1(3)	-14(3)
C25	17(3)	17(3)	17(3)	-6(2)	-3(2)	1(2)
C26	16(3)	24(3)	29(3)	-13(3)	-9(2)	3(2)
C27	17(3)	22(3)	23(3)	-7(3)	-7(2)	0(2)
C28	19(3)	25(3)	32(3)	-9(3)	-10(3)	6(2)
C29	18(3)	34(4)	31(3)	-11(3)	-2(3)	5(3)
C30	14(3)	30(3)	33(4)	-8(3)	0(2)	-5(2)
C31	14(3)	31(3)	34(4)	-12(3)	-5(3)	-5(2)
C32	18(3)	25(3)	32(4)	-1(3)	-6(3)	-1(2)
C33	16(3)	25(3)	30(3)	-6(3)	-1(2)	-3(2)
C34	31(3)	47(4)	29(4)	-4(3)	-7(3)	-1(3)
C35	23(3)	40(4)	44(4)	-16(3)	-10(3)	-6(3)
C36	48(4)	27(4)	33(4)	-4(3)	-25(3)	2(3)

Atom	U11	U22	U33	U23	U13	U12
C37	19(3)	47(4)	28(3)	-18(3)	-10(3)	-1(3)

Table7-3: Bond Lengths in Å for ra178.

Atom	Atom	Length/Å
O1	C1	1.339(6)
O1	C36	1.452(6)
O2	C1	1.204(6)
O3	C6	1.239(6)
O4	C11	1.225(6)
O5	N5	1.414(5)
O5	C37	1.463(6)
O6	C14	1.232(6)
O7	C19	1.236(6)
O8	C21	1.433(6)
O9	C25	1.241(6)
O10	C26	1.414(7)
O11	C27	1.422(6)
O11	C31	1.463(6)
O12	C27	1.417(6)
O12	C33	1.465(6)
N1	N2	1.427(6)
N1	C5	1.474(6)
N2	C2	1.456(7)
N2	C6	1.347(6)
N3	N4	1.426(6)
N3	C7	1.480(6)
N3	C11	1.354(6)
N4	C10	1.464(7)
N5	C12	1.469(7)
N5	C14	1.369(7)
N6	N7	1.434(6)
N6	C18	1.477(7)
N7	C15	1.457(7)
N7	C19	1.360(6)
N8	C20	1.465(6)
N8	C25	1.328(7)
C1	C2	1.513(7)
C2	C3	1.540(7)
C3	C4	1.527(8)
C4	C5	1.522(8)
C6	C7	1.522(8)
C7	C8	1.528(7)
C8	C9	1.527(7)
C9	C10	1.531(7)
C11	C12	1.532(7)
C12	C13	1.527(8)

Atom	Atom	Length/Å
C14	C15	1.520(7)
C15	C16	1.542(7)
C16	C17	1.527(7)
C17	C18	1.513(8)
C19	C20	1.518(7)
C20	C21	1.530(7)
C21	C22	1.537(8)
C22	C23	1.520(8)
C22	C24	1.533(8)
C25	C26	1.548(7)
C26	C27	1.545(7)
C26	C32	1.517(8)
C27	C28	1.528(7)
C28	C29	1.545(8)
C29	C30	1.536(8)
C30	C31	1.517(8)
C30	C33	1.534(7)
C31	C35	1.503(8)
C33	C34	1.497(8)

Table 7-4: Bond Angles in ° for ra178.

Atom	Atom	Atom	Angle/°
C1	O1	C36	115.1(4)
N5	O5	C37	110.6(4)
C27	O11	C31	112.5(4)
C27	O12	C33	111.7(4)
N2	N1	C5	110.1(4)
N1	N2	C2	118.4(4)
C6	N2	N1	119.8(4)
C6	N2	C2	121.5(4)
N4	N3	C7	118.8(4)
C11	N3	N4	120.0(4)
C11	N3	C7	119.7(4)
N3	N4	C10	110.2(4)
O5	N5	C12	115.4(4)
C14	N5	O5	115.2(4)
C14	N5	C12	119.4(4)
N7	N6	C18	109.4(4)
N6	N7	C15	119.2(4)
C19	N7	N6	120.3(4)
C19	N7	C15	120.2(4)
C25	N8	C20	123.3(4)
O1	C1	C2	110.6(5)
O2	C1	O1	124.3(5)

Atom	Atom	Atom	Angle/°
O2	C1	C2	125.1(5)
N2	C2	C1	109.1(4)
N2	C2	C3	109.5(4)
C1	C2	C3	113.8(4)
C4	C3	C2	109.7(4)
C5	C4	C3	110.3(4)
N1	C5	C4	113.3(5)
O3	C6	N2	121.0(5)
O3	C6	C7	122.2(4)
N2	C6	C7	116.7(4)
N3	C7	C6	108.8(4)
N3	C7	C8	108.6(4)
C6	C7	C8	112.4(4)
C9	C8	C7	112.3(5)
C8	C9	C10	108.9(4)
N4	C10	C9	113.3(4)
O4	C11	N3	121.0(4)
O4	C11	C12	123.3(5)
N3	C11	C12	115.6(5)
N5	C12	C11	110.9(4)
N5	C12	C13	111.3(4)
C13	C12	C11	110.2(4)
O6	C14	N5	120.5(5)
O6	C14	C15	121.5(5)
N5	C14	C15	118.0(5)
N7	C15	C14	110.7(4)
N7	C15	C16	110.6(4)
C14	C15	C16	109.5(4)
C17	C16	C15	111.2(4)
C18	C17	C16	109.7(5)
N6	C18	C17	114.3(4)
O7	C19	N7	121.1(5)
O7	C19	C20	119.6(4)
N7	C19	C20	119.2(4)
N8	C20	C19	104.9(4)
N8	C20	C21	108.1(4)
C19	C20	C21	110.4(4)
O8	C21	C20	107.1(4)
O8	C21	C22	111.5(4)
C20	C21	C22	113.7(4)
C23	C22	C21	112.7(5)
C23	C22	C24	111.0(5)
C24	C22	C21	110.4(5)
O9	C25	N8	124.8(5)
O9	C25	C26	118.3(5)

Atom	Atom	Atom	Angle/°
N8	C25	C26	116.7(5)
O10	C26	C25	106.3(4)
O10	C26	C27	108.7(4)
O10	C26	C32	109.9(4)
C27	C26	C25	110.1(4)
C32	C26	C25	109.0(4)
C32	C26	C27	112.6(4)
O11	C27	C26	105.7(4)
O11	C27	C28	110.8(4)
O12	C27	O11	109.7(4)
O12	C27	C26	107.5(4)
O12	C27	C28	107.5(4)
C28	C27	C26	115.6(4)
C27	C28	C29	107.6(4)
C30	C29	C28	107.5(4)
C31	C30	C29	107.2(5)
C31	C30	C33	108.4(4)
C33	C30	C29	109.3(4)
O11	C31	C30	106.5(4)
O11	C31	C35	107.0(4)
C35	C31	C30	117.9(5)
O12	C33	C30	107.0(4)
O12	C33	C34	107.7(5)
C34	C33	C30	116.4(5)

Table 7-5: Torsion Angles in ° for ra178.

Atom	Atom	Atom	Atom	Angle/°
O1	C1	C2	N2	-174.0(4)
O1	C1	C2	C3	-51.3(6)
O2	C1	C2	N2	7.1(7)
O2	C1	C2	C3	129.7(5)
O3	C6	C7	N3	16.5(7)
O3	C6	C7	C8	-103.8(5)
O4	C11	C12	N5	-25.6(7)
O4	C11	C12	C13	98.1(6)
O5	N5	C12	C11	78.2(5)
O5	N5	C12	C13	-44.9(5)
O5	N5	C14	O6	169.7(4)
O5	N5	C14	C15	-8.1(6)
O6	C14	C15	N7	39.0(6)
O6	C14	C15	C16	-83.1(6)
O7	C19	C20	N8	-60.7(6)

Atom	Atom	Atom	Atom	Angle/°
O7	C19	C20	C21	55.6(6)
O8	C21	C22	C23	53.2(6)
O8	C21	C22	C24	-71.4(6)
O9	C25	C26	O10	13.9(6)
O9	C25	C26	C27	131.5(5)
O9	C25	C26	C32	-104.5(5)
O10	C26	C27	O11	-164.2(4)
O10	C26	C27	O12	78.8(5)
O10	C26	C27	C28	-41.3(6)
O11	C27	C28	C29	-66.4(5)
O12	C27	C28	C29	53.4(5)
N1	N2	C2	C1	70.5(5)
N1	N2	C2	C3	-54.7(6)
N1	N2	C6	O3	-173.8(5)
N1	N2	C6	C7	7.4(7)
N2	N1	C5	C4	-51.7(6)
N2	C2	C3	C4	52.7(5)
N2	C6	C7	N3	-164.6(4)
N2	C6	C7	C8	75.0(6)
N3	N4	C10	C9	53.7(6)
N3	C7	C8	C9	-50.6(6)
N3	C11	C12	N5	157.5(4)
N3	C11	C12	C13	-78.8(6)
N4	N3	C7	C6	-70.9(6)
N4	N3	C7	C8	51.8(6)
N4	N3	C11	O4	167.9(5)
N4	N3	C11	C12	-15.1(7)
N5	C14	C15	N7	-143.2(4)
N5	C14	C15	C16	94.6(5)
N6	N7	C15	C14	-71.0(5)
N6	N7	C15	C16	50.5(6)
N6	N7	C19	O7	176.2(4)
N6	N7	C19	C20	-0.5(6)
N7	N6	C18	C17	53.1(6)
N7	C15	C16	C17	-48.8(6)
N7	C19	C20	N8	116.1(5)
N7	C19	C20	C21	-127.6(5)
N8	C20	C21	O8	174.2(4)
N8	C20	C21	C22	-62.1(5)
N8	C25	C26	O10	-170.5(4)
N8	C25	C26	C27	-52.9(6)
N8	C25	C26	C32	71.2(6)
C1	C2	C3	C4	-69.7(6)
C2	N2	C6	O3	-0.2(7)
C2	N2	C6	C7	-179.0(4)

Atom	Atom	Atom	Atom	Angle/°
C2	C3	C4	C5	-54.1(6)
C3	C4	C5	N1	54.7(6)
C5	N1	N2	C2	53.6(6)
C5	N1	N2	C6	-132.6(5)
C6	N2	C2	C1	-103.2(5)
C6	N2	C2	C3	131.6(5)
C6	C7	C8	C9	69.9(5)
C7	N3	N4	C10	-53.5(6)
C7	N3	C11	O4	1.8(8)
C7	N3	C11	C12	178.8(4)
C7	C8	C9	C10	53.8(6)
C8	C9	C10	N4	-55.4(6)
C11	N3	N4	C10	140.3(5)
C11	N3	C7	C6	95.4(5)
C11	N3	C7	C8	-142.0(5)
C12	N5	C14	O6	25.7(7)
C12	N5	C14	C15	-152.0(4)
C14	N5	C12	C11	-137.9(5)
C14	N5	C12	C13	99.0(5)
C14	C15	C16	C17	73.4(6)
C15	N7	C19	O7	2.7(7)
C15	N7	C19	C20	-174.1(4)
C15	C16	C17	C18	52.0(6)
C16	C17	C18	N6	-55.6(6)
C18	N6	N7	C15	-51.7(5)
C18	N6	N7	C19	134.7(4)
C19	N7	C15	C14	102.6(5)
C19	N7	C15	C16	-135.9(5)
C19	C20	C21	O8	60.0(5)
C19	C20	C21	C22	-176.3(4)
C20	N8	C25	O9	-3.5(8)
C20	N8	C25	C26	-178.9(4)
C20	C21	C22	C23	-68.0(5)
C20	C21	C22	C24	167.3(4)
C25	N8	C20	C19	-133.7(5)
C25	N8	C20	C21	108.5(5)
C25	C26	C27	O11	79.8(5)
C25	C26	C27	O12	-37.3(6)
C25	C26	C27	C28	-157.4(4)
C26	C27	C28	C29	173.4(5)
C27	O11	C31	C30	20.4(5)
C27	O11	C31	C35	147.4(4)
C27	O12	C33	C30	20.0(6)
C27	O12	C33	C34	145.8(4)
C27	C28	C29	C30	13.1(6)

Atom	Atom	Atom	Atom	Angle/°
C28	C29	C30	C31	51.6(5)
C28	C29	C30	C33	-65.7(6)
C29	C30	C31	O11	-73.0(5)
C29	C30	C31	C35	166.8(5)
C29	C30	C33	O12	48.6(6)
C29	C30	C33	C34	-71.8(6)
C31	O11	C27	O12	-71.1(5)
C31	O11	C27	C26	173.3(4)
C31	O11	C27	C28	47.4(5)
C31	C30	C33	O12	-68.0(5)
C31	C30	C33	C34	171.6(5)
C32	C26	C27	O11	-42.2(5)
C32	C26	C27	O12	-159.2(4)
C32	C26	C27	C28	80.7(6)
C33	O12	C27	O11	45.6(5)
C33	O12	C27	C26	160.0(4)
C33	O12	C27	C28	-74.9(5)
C33	C30	C31	O11	44.9(6)
C33	C30	C31	C35	-75.3(6)
C36	O1	C1	O2	-3.1(7)
C36	O1	C1	C2	178.0(4)
C37	O5	N5	C12	-115.8(4)
C37	O5	N5	C14	98.8(5)

Table 7-6: Hydrogen Fractional Atomic Coordinates (×104) and Equivalent Isotropic Displacement Parameters (Å2×103) for ra178. Ueq is defined as 1/3 of the trace of the orthogonalised Uij.

Atom	x	y	z	Ueq
H8	-2581.84	5065.79	9909.07	39
H10	1390(110)	400(90)	6780(90)	110(30)
H1	10990(60)	6450(50)	1950(50)	22(15)
H4	6040(70)	7030(70)	320(60)	37(19)
H6	-610(80)	5550(70)	4850(70)	70(20)
H8A	1670(80)	3210(60)	7920(60)	51(19)
H2	9921.86	4066.68	1239.54	26
H3A	12672.26	3479.62	644.63	31
H3B	12154.22	4707.86	-460.44	31
H4A	13429.95	4906.52	1271.35	34
H4B	14255.37	5275.82	-245.13	34
H5A	13154.52	7210.27	275.09	33
H5B	12459.11	7083.47	-688.44	33
H7	8445.52	8086.78	1312.01	24
H8B	8734.53	9580.86	-753.54	27
H8C	10027.29	8389.38	-891.83	27

Atom	x	y	z	Ueq
H9A	8580.59	7533.54	-1565.38	32
H9B	8723.18	9068.63	-2451.14	32
H10A	6263.58	9652.05	-1221.74	31
H10B	6130.7	8534.88	-1681.39	31
H12	3587.63	7862.11	1977.5	25
H13A	2332.03	10000.54	1863.73	48
H13B	3989.65	10100.77	799.69	48
H13C	3728.68	10356.09	2080.18	48
H15	850.6	7560	5902.5	23
H16A	-1485.27	8708.61	5754.89	27
H16B	-287.14	9294.52	4347.67	27
H17A	-2618.25	8814.07	4295.45	30
H17B	-1124.78	8054.79	3584.93	30
H18A	-3135.96	6920.17	6136.28	32
H18B	-2978.55	6558.47	4938.55	32
H20	-1254.65	3823.87	7510.16	21
H21	-630.61	3811.97	9690.05	26
H22	-1068.1	1655.25	9935	31
H23A	-4110.19	2761.74	9963.2	55
H23B	-3563.11	1215.45	10332.08	55
H23C	-2892.29	2215.24	8891.55	55
H24A	-1533.27	2132.7	11774.2	66
H24B	-2751.98	1167.7	12067.98	66
H24C	-3249.98	2712.19	11790.95	66
H28A	5089.93	599.53	5558.98	32
H28B	5674.67	421.78	6711.91	32
H29A	7471.48	1768.93	5228.89	37
H29B	6622.39	2209.54	4174.09	37
H30	6643.26	4053.57	4792.7	34
H31	6270.46	2478	7030.68	31
H32A	1576.26	-278.7	9276.46	43
H32B	2417.02	863.37	9193.44	43
H32C	3406.74	-339.09	8762.49	43
H33	3857.76	4660.89	5267.88	32
H34A	5118.52	3791.32	3145.07	61
H34B	5241.22	5233.37	3074.13	61
H34C	3608.41	4783.36	3371.43	61
H35A	4377.29	4902.21	6805.89	53
H35B	6079.78	4644.44	6877.85	53
H35C	4714.51	3925.77	8107.05	53
H36A	11687.6	986.34	4406.35	54
H36B	10240.48	1920.75	4912.43	54
H36C	11936.89	2331.63	4463.59	54
H37A	3077.83	7296.14	6038.46	44
H37B	3555.2	8686.98	5815.47	44

Atom	x	y	z	Ueq
H37C	4673.48	7812.08	4997.98	44

Table 7-7: Hydrogen Bond information for ra178.

D	H	A	d(D-H)/Å	d(H-A)/Å	d(D-A)/Å	D-H-A/deg
O8	H8	O31	0.84	1.96	2.786(5)	168.4
O10	H10	O9	0.95(10)	1.85(9)	2.566(5)	130(8)
N4	H4	O3	0.90(7)	2.24(6)	2.897(6)	129(5)
N8	H8A	O11	0.99(7)	2.17(6)	2.865(5)	126(5)
C37	H37C	O4	0.98	2.22	2.991(7)	134.3

1-1+x,+y,1+z

Citations

O.V. Dolomanov and L.J. Bourhis and R.J. Gildea and J.A.K. Howard and H. Puschmann, Olex2: A complete structure solution, refinement and analysis program, J. Appl. Cryst., (2009), 42, 339-341.
Sheldrick, G.M., A short history of ShelX, Acta Cryst., (2008), A64, 339-341.
Software for the Integration of CCD Detector System Bruker Analytical X-ray Systems, Bruker axs, Madison, WI (after 2013

Appendix F Experimental Data for MIC of Incarnatapeptins

MIC with MRSA

Incarnatapeptin C   (8089-C-Et-B-C-13)

OD @ 595nm				Mean	Net OD	% alive
Blank	0.0469	0.0495	0.0428	0.0464	0.0000	
DMSO + cells	0.6096	0.6656	0.6117	0.6290	0.5826	100
64μg/ml	0.0464	0.0463	0.0523	0.0483	0.0019	0
32μg/ml	0.0896	0.0452	0.0442	0.0597	0.0133	2
16μg/ml	0.047	0.0542	0.0921	0.0644	0.0180	3
8μg/ml	0.0548	0.0867	0.0598	0.0671	0.0207	4
4μg/ml	0.0405	0.0812	0.0859	0.0692	0.0228	4
2μg/ml	0.0534	0.0878	0.0446	0.0619	0.0155	3
1μg/ml	0.5133	0.5167	0.4183	0.4828	0.4364	75
0.5μg/ml	0.5194	0.4133	0.4225	0.4517	0.4053	70
0.25μg/ml	0.5079	0.4921	0.5353	0.5118	0.4654	80
0.125μg/ml	0.4772	0.4721	0.5192	0.4895	0.4431	76
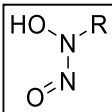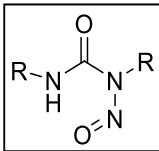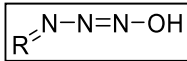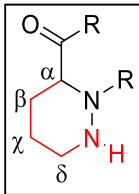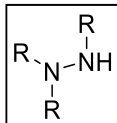
MIC with E.coli 8161
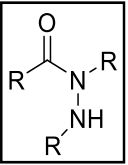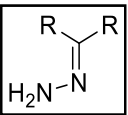
Rifampicin

OD @ 595nm				Mean	Net OD	% alive
Blank	0.0472	0.0478	0.0474	0.0475	0.0000	
DMSO + cells	0.133	0.1394	0.1345	0.1356	0.0881	100
64 µg/ml	0.0611	0.0614	0.0619	0.0615	0.0140	0
12.8 µg/ml	0.0533	0.0535	0.0519	0.0529	0.0054	0
2.56 µg/ml	0.0475	0.0472	0.0487	0.0478	0.0003	0
512 ng/ml	0.0477	0.0478	0.0475	0.0477	0.0002	0
102.4 ng/ml	0.0469	0.047	0.0476	0.0472	-0.0003	0
20.48 ng/ml	0.0475	0.047	0.0471	0.0472	-0.0003	0
4.096 ng/ml	0.0998	0.1117	0.1035	0.1050	0.0575	65
0.8192 ng/ml	0.1392	0.1379	0.1326	0.1366	0.0891	101
0.1638 ng/ml	0.1372	0.1388	0.1334	0.1365	0.0890	101
0.0328 ng/ml	0.1412	0.1572	0.1727	0.1570	0.1095	124

Incarnatapeptin D   (8089-16-acid)

OD @ 595nm				Mean	Net OD	% alive
Blank	0.0475	0.0476	0.0474	0.0475	0.0000	
DMSO + cells	0.1352	0.1444	0.1413	0.1403	0.0928	100
64 µg/ml	0.05	0.0492	0.0524	0.0505	0.0030	3
12.8 µg/ml	0.1396	0.1474	0.1351	0.1407	0.0932	100
2.56 µg/ml	0.1376	0.1385	0.1393	0.1385	0.0910	98
512 ng/ml	0.1418	0.1456	0.1413	0.1429	0.0954	103
102.4 ng/ml	0.141	0.1436	0.1381	0.1409	0.0934	101
20.48 ng/ml	0.1388	0.1424	0.1279	0.1364	0.0889	96
4.096 ng/ml	0.1411	0.1434	0.1316	0.1387	0.0912	98
0.8192 ng/ml	0.1377	0.1414	0.1336	0.1376	0.0901	97
0.1638 ng/ml	0.1359	0.1316	0.1375	0.1350	0.0875	94
0.0328 ng/ml	0.1407	0.1621	0.1782	0.1603	0.1128	122

Incarnatapeptin C   (8089-C-Et-B-C-13)

OD @ 595nm				Mean	Net OD	% alive
Blank	0.047	0.0639	0.0471	0.0527	0.0000	
DMSO + cells	0.1296	0.1259	0.1359	0.1305	0.0778	100
20 µg/ml	0.0475	0.0487	0.0475	0.0479	-0.0048	0
4 µg/ml	0.0482	0.0473	0.0469	0.0475	-0.0052	0
2 µg/ml	0.0471	0.048	0.0467	0.0473	-0.0054	0
1 µg/ml	0.047	0.047	0.0473	0.0471	-0.0056	0
500 ng/ml	0.051	0.0605	0.0514	0.0543	0.0016	2
250 ng/ml	0.1284	0.1278	0.1267	0.1276	0.0749	96
50 ng/ml	0.1315	0.1285	0.1283	0.1294	0.0767	99
10 ng/ml	0.1264	0.1279	0.1299	0.1281	0.0754	97

Appendix G Other N-N Bond Natural Product Functional Groups

Examples of Nitrogen-Nitrogen Functional Groups found in Natural Products. Part 1 of 2						Citations
Reported ^{15}N NMR chemical shifts are described as well as the availability of genomic information						+++ [306,307]
						++ [308,309]

N-N moiety	Structure of N-N moiety	Nitrogen NMR		Genome Mining	Natural products containing this moiety^{+++}
		^{15}N NMR in various solvents referenced to NH$_{3(l)}$ (CH$_3$NO$_2$)	Experiment	Known Clusters	
N-nitroso					
Nitrosohydroxylamines		Not reported	N/A	Yes	Fragin
					Alanosine
					Dopastin
Nitrosoureas		**Streptozocin^{++}** (in DMSO-d$_6$)	^{15}NHMBC	Yes	Streptozocin
		δ NHCO**N**CH$_3$NO 162.9(-217.6)			
		δ NHCONCH3**NO** 267.1(-113.4)			
		δ NHCONC**H3**NO 3.10			
N-hydroxytriazene		Not reported		Yes	Triascins
Cyclic Hydrazine		**Piperazic Acid**			Piperazic Acid
		See Chapter 3 for Piz nitrogen NMR		Yes	
Linear Hydrazines		**Braznitidumine^{++}** (in DMSO-d6)	^{15}NHSQC		Munroniamide
		δ **N**NH not reported			
		δ N**N**H 55 (-325.5)		No	Ostrerine A
		δ NN**H** 2.00			Braznitidumine

Examples of Nitrogen-Nitrogen Functional Groups found in Natural Products. Part 2 of 2

Reported ^{15}N NMR chemical shifts are described as well as the availability of genomic information

N-N moiety	Structure of N-N moiety	Nitrogen NMR		Genome Mining	Natural products containing this moiety[+++]	Citations
		^{15}N NMR in various solvents referenced to $NH_{3(l)}$ (CH_3NO_2)	Experiment	Known Clusters		++[310–312]
N-nitroso						
Hydrazides		**Elaiomycin B**[++] (in DMSO-d6)			Fosfazinomycins	
		δ CO**N**HCO 144.1 (-236.4)			Hydrazidomycins	
		δ CON**N**HCO 140.9 (-239.6)	^{15}NHSQC		Agaritine	
		δ CONN**H**CO 10.65	^{15}NHMBC	Yes	Anthglutin	
		Geralcin A[++] (in DMF-d7)			Xanthodermine	
		δ CO**N**NHCO 133.5 (-247)			Pygmeine	
		δ CON**N**HCO 137.0 (-243.5)			Spinamycin	
		δ CONN**H**CO 10.44			Elaiomycin	
		Geralcin B[++] in (CD_2Cl_2)			Geralcin	
		δ CO**N**NHCO 147.2 (-233.3)				
		δ CON**N**HCO 142.3 (-238.2)				
		δ CONN**H**CO 8.01				
Hydrazones		**peptide s56-p1**[++] (in H2O+D2O)			Katorazone	
		R=**N**NH$_2$ 361.4 (-19)	^{15}NHMBC	Yes	Yoropyrazone	
		R=N**N**H$_2$ 121 (-259.5)				
		R=NN**H**$_2$ not reported				