Title

The Discovery of Novel Bioactive Small Molecule Natural Products Through Metabolomic Method Development and Genome Mining

Author

Zill, Nicholas

Publication Date

2022

Peer reviewed|Thesis/dissertation

The Discovery of Novel Bioactive Small Molecule Natural Products Through Metabolomic Method Development and Genome Mining

By

Nicholas A. Zill

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Chemical Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Wenjun Zhang, Chair
Professor Markita Landry, Additional Member
Professor Matthew Traxler, ASM

Spring 2022

Abstract
The Discovery of Novel Bioactive Small Molecule Natural Products Through Metabolomic Method Development and Genome Mining
By
Nicholas A. Zill
Doctor of Philosophy in Chemical Engineering
University of California, Berkeley
Professor Wenjun Zhang, Chair

Natural products have been utilized by humans since the beginning of recorded history. Some of the first documented uses of these compounds comes in the form of plant-based extracts and mixtures of various poultices for a wide range of medicinal uses. Natural product small molecule discovery serves as an essential pipeline for novel therapeutics, improves our understanding of human disease, and supplies innovative chemistries for the synthesis and derivatization of compounds. This work will be split amongst the development and application of natural product discovery methodologies (Chapters 2 and 3) and the combination of existing methodologies to discover molecules pertinent to disease (Chapters 4 and 5). Each of these chapters will deal with the discovery, identification, purification, and characterization of small molecules with a key focus on metabolomics as a focal tool.

Dedication

Dedicated to my supportive and loving parents, Mark and Christine, my brother, Mike, and my dear friends, Weronika Jakubowska, Surya Parker, and Tommy Hoffmann. As well as to the beautiful cohorts of companions that I've had the privilege of calling my friend: Lana, Shelby, Dan, Pat, Ethan, AJ, Jordan, Colin, Sammy, and Ryan. Don't let your memes be dreams. Thank you.

Acknowledgements

This work involves the contributions of numerous members of the lab of Wenjun Zhang at UC Berkeley. I'm endlessly grateful for the time, hard work, and opportunities that have come from working with this truly elite group of scientists. I am particularly thankful to Wenjun Zhang for her support and grateful for her steady mentorship and brilliance.

This work has been financially supported by the Chan Zuckerberg Biohub and the Hong Kong Institute of Science and Technology Collaborative Research Fund.

Streptococcal genetic manipulations detailed in Chapters 4 and 5 were performed by Colin Barber. *Lactobacillus* strains detailed in Chapter 5 were provided by the Coates Lab. NMR characterizations discussed in Chapters 2 and 4 and detailed in the appendixes were performed by Dr. Yongle Du.

Table of Contents

Chapter 1: Introduction

1.1 History of Natural Products

Natural products have been utilized by humans since the beginning of recorded history. Some of the first documented uses of these compounds comes in the form of plant-based extracts and mixtures of various poultices for a wide range of medicinal uses. Natural product small molecule discovery serves as an essential pipeline for novel therapeutics, improves our understanding of human disease, and supplies innovative chemistries for the synthesis and derivatization of compounds[1].

1.2 Function of Secondary Metabolites in Microorganisms

Secondary metabolites serve a wide variety of functions in microorganisms. These metabolites are defined as small molecules that are not required for the replication or growth of a microorganism. Detailed below is a non-exhaustive list of the many functions that these molecules can exhibit.

1.2.1 Antimicrobial Activity

Antimicrobial activity is likely the most famous of secondary metabolite functions. Since the discovery of antibiotics like penicillin from the *Penicillium* mold, the human world has been dramatically changed by access to these molecules[2]. Among microbes, these molecules serve as active weaponry between organisms to fight for space and nutrients. The repurposing of these molecules to fight human infections has allowed for an era of unprecedented disease treatment.

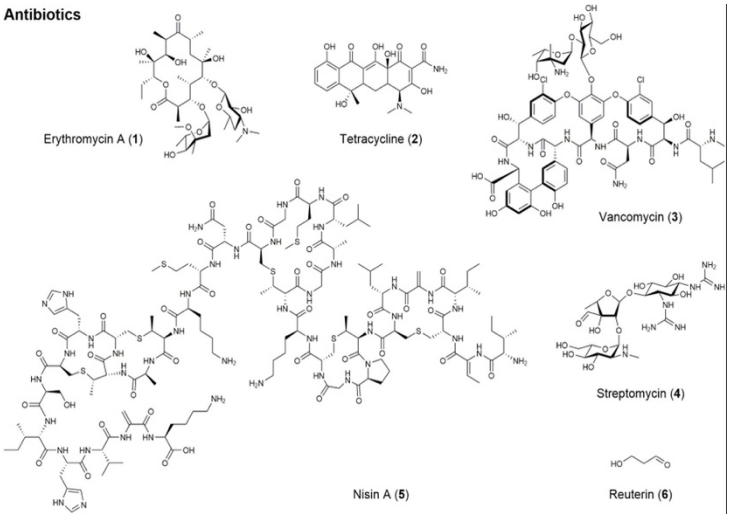


Figure 1:A list of natural product antibiotics adapted from Pham et al. 2019

The production of these molecules has travelled across many microbial classes, with *Actinomyces*, specifically *Streptomyces*, being some of the most proliferative producers. This privileged genus represents a wide range of bacteria that populates environments including land terrain, the ocean, plants, and other living eukaryotes. These strains have been responsible for more than 70% of the antibiotics used in human medicine to date[1,3]. Chapter 2 discusses the use of these strains as sources of novel bioactive small molecules.

While antibiotics have proved undeniably essential towards the advancement of human society, the rate of antibiotic discovery has slowed dramatically in recent years[4]. Since microorganisms are constantly evolving due to selective pressures, this has and will continue to lead to antibiotic resistance species that can be enormously dangerous towards human health. This can be observed in pathogens like *Pseudomonas aeruginosa* which has

developed resistances to "last resort" antibiotics like the carbapenems and cephalosporins[5]. While many factors play into this roadblock for managing effective infection treatment, two factors will be thoroughly discussed throughout this work: the limitations of current methodology for antibiotic discovery, and the need for small molecule mechanistic comprehension other than those directed solely at cytotoxicity.

1.2.2 Motility/Adhesion

Another useful feature of secondary metabolites can come in the form of motility and adhesion. The ability to control movement in a local environment is a highly desirable trait. Secondary metabolites can induce adhesion as well as biofilm formation in cells. *Streptococcus mutans* are a genus of gram positive coccus bacteria with the ability to form biofilms on human teeth. These acid producing bacteria can cause enamel decay and the formation of carries with at high cell densities. Small molecules, like mutanofactin, have previously been implicated in modulating the cell surface hydrophobicity of these strains, playing a significant role in biofilm formation[6]. Other pathogens have been shown to modify biofilm formation using small molecule mechanisms as well[7,8].

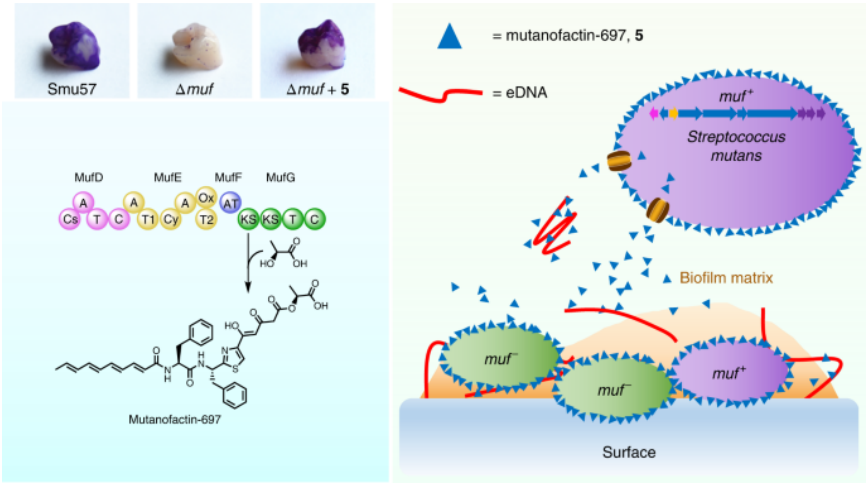


Figure 2: Biological role of mutanofactin in S. mutans; visual adapted from Li et al. 2021

1.2.3 Nutrient Acquisition

Small molecules can also play a role in the acquisition of nutrients. Most famously these molecules are frequently used in the acquisition of minerals. Metallophores, and especially siderophores, are a widespread class of secondary metabolites that scavenge key metals for cellular function[9]. Due to the defense strategy of withholding of iron present in some mammals including humans, many pathogens have developed these specialized small molecules to combat for access to these metals[10].

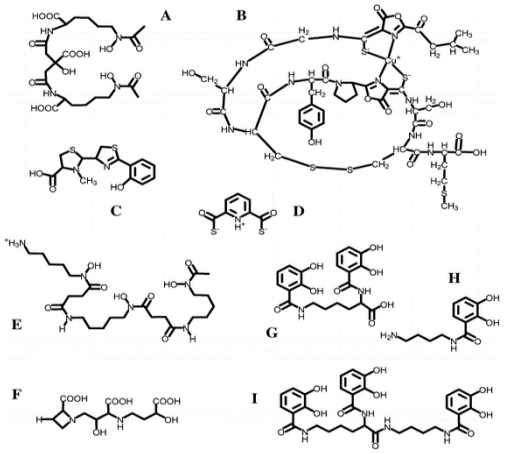


Figure 3: Various metallophores adapted from Kraemer et al. 2015

1.2.4 Environmental Resistance

In addition to nutrient acquisition, small molecules can also assist in defensive functions against threats. Small molecules have been detected in *Streptococcus mutans* that modify local pH environments favorably for the organism[11]. Additionally, oxidative stress has been shown to significantly influence secondary metabolite production in microbial species[12]. Small molecules can also play signaling roles in defense against antimicrobials. One strategy involves the management of radical oxidative species, either to increase ROS to degrade antibiotics, or decrease them to prevent damage induced by the same molecules[13,14]. The more recent focus on modulating oxidative stress and "antioxidants" in the human world is a strategy long employed by the microbial community.

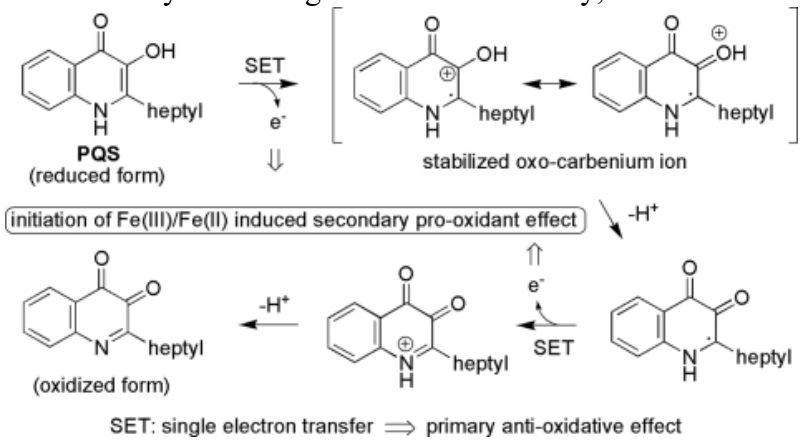


Figure 4: Oxidative stress modifying activity of PQS adapted from Haussler et al. 2008

1.2.5 Extracellular Signaling

A broad class of secondary metabolites exists that plays a role in extracellular signaling. Many molecules act to allow the detection of organisms from the same strain. This can allow microbes to expedite communicate in response to threats and indicate spatial restraints. Bacteria like *Erwinia carotovora* utilize quorum sensing molecules to trigger production of antibiotics. During late log to early stationary phase, this molecule triggers the production of β-lactam antibiotics[15]. Additionally a number of microbes utilize small molecules to signal and influence community compositions, such is the case with many fungal secondary metabolites[16,17].

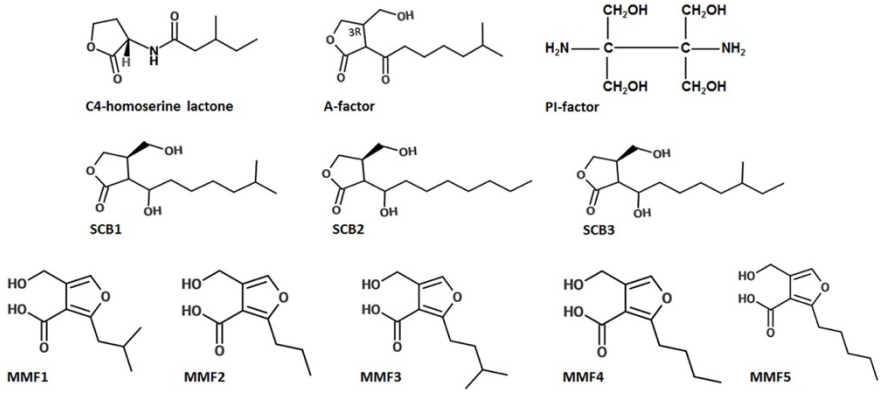


Figure 5: Representative quorum sensing molecules from actinobacteria taken from Li et al. 2016

1.3 Natural Products as a Source of Modern Medicine

Even with the use of highly constrained definitions, hundreds of thousands of small molecules have been discovered and their role in society has only grown as the human population has[18]. During 2019 alone, amoxicillin, a natural product penicillin type antibiotic that treats a number of microbial infections, was prescribed over 31 million times just in the US[19]. Amongst the many therapeutic activities of natural products, a core theme of targeted cytotoxicity emerges for much of early secondary metabolite discovery. Focus on antibiotics, antifungals, antiparasitics, antitumor, and antiviral small molecules has led to a flourishing class of therapies targeted at killing undesirable infections or cells.

Outside of infection, natural products have played a vital role in disease treatment for a cornucopia of human disease. A few examples include: physostigmine, a reversible cholinesterase inhibitor used to treat glaucoma; rapamycin, an immunosuppressant that inhibits the proliferation of T cells; and Tetrahydrocannabinol, a cannabinoid used to treat pain and appetite suppression in cancer patients[20–22]. Even non-prescribed molecules can have truly dramatic health impacts. The daily consumption of omega-3 fatty acids, found in foods like fish and flaxseed, has been shown to decrease the risk of heart disease in at risk patient groups by over 25%[23]. This dramatic effect encourages the continued discovery of this high impact small molecules.

As the scope of diseases has expanded for natural products, the need for more high throughput screening methods has become vital. Traditional cytotoxicity assays represent both a bottleneck to discovery and an incomplete characterization of what defines useful bioactivity. New approaches might include examining secondary metabolites generated by pathogens that diverge from cytotoxic behavior, such as the many bioactivities mentioned in Section 1.2. Furthermore, exploring the methodologies used to detect bioactive small molecules can provide rich insight and accelerate the discovery of high impact small molecules.

1.4 Existing Methodologies in Natural Product Discovery

Comparative metabolomics and bioactivity guided fractionation are two key methods currently used for identifying novel secondary metabolites. Figure 6 overviews the workflow of these two techniques and highlights many of their downfalls.

Most early work in discovery within the natural product space emerged from bioactivity guided fraction. Bioactivity based fractionation uses a specified phenotype, commonly some form of cytotoxicity, and seeks complex bacterial or plant extracts to test against[24]. Once a sample meets a set criterion, the extract is fractionated. typically using chromatographic techniques (though other techniques like distillation are not uncommon). The fractions are then screened iteratively and further fractionated until a major metabolite responsible for the activity is identified, either through mass spectrometry, NMR, or crystallization.

The bioactivity guided fractionation process has several issues. The lengthy duration of the iterative screening process serves as a time intensive roadblock to compound discovery and makes the discovery of compounds that degrade readily extremely challenging. In addition, a very common feature of microbes is the production of “families” of compounds from a given gene cluster, as well as upstream shunt products[25]. Because these families have varying structures, chromatographic separations can become highly convoluted by multiple bioactive fractions. The chromatographic

methods and processing methods in themselves can degrade the molecules as well. One particularly pressing issue in bioactivity based fractionation is the problem of rediscovery. Many plants and microbes utilize the same bioweapons and accordingly, this form of screening can expend enormous resources to only rediscover a known compound. While various metabolomic and database methods are attempting to resolve this problem, the process is still in its infancy[26]. Bioactivity based fraction additionally provides an extremely narrow look into the bioactivity of the molecule. This results in a very low-throughput, low return result from screenings of extracts. This methodology misses the potential of extracts deemed inactive for the given assay to be viable in other medical contexts. Lastly, compounds that may have biologically relevant titer at the microbial scale, may not produce sufficient metabolite to appear on the bioassays performed. This insensitivity of this approach misses what could be very promising molecules of insufficient titer.

Comparative metabolomics is a relatively newer technology, with the first appearance of the term appearing only as early as 1997[27]. This technique typically relies on differential expression of a trait between an organism that otherwise shares identical features. This is most commonly achieved through genetic manipulation of the producer, though there are examples of usage in phenotypic settings such as the comparison of planktonic and biofilm metabolite production. This differential expression ideally results

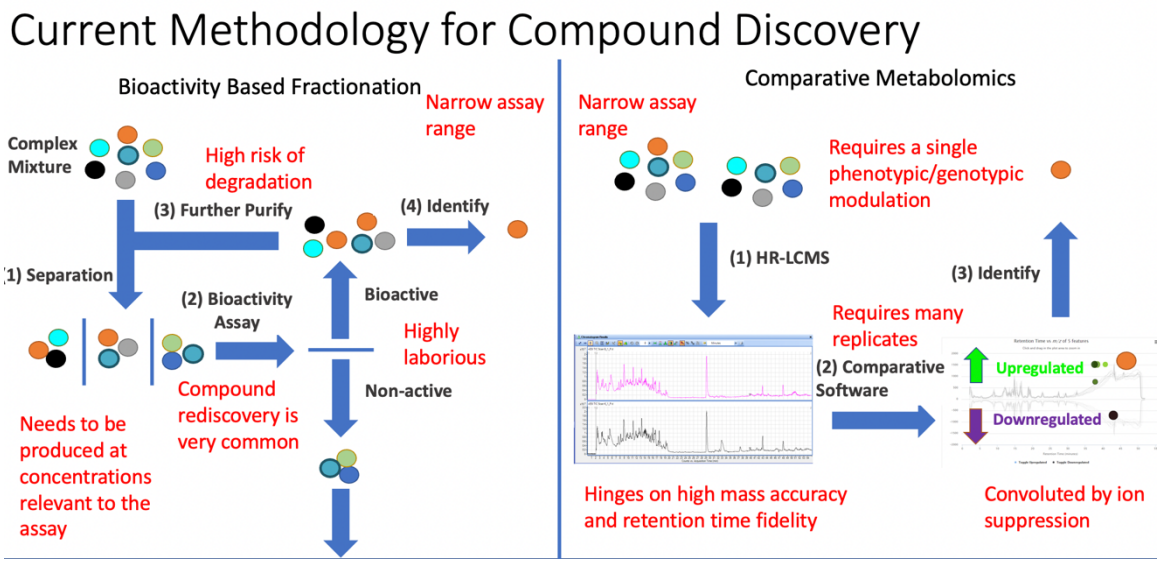


Figure 6: Methodology and drawbacks of bioactivity based fractionation and comparative metabolomics

in complete silencing or novel expression of a target compound, leading to the identification and structural elucidation of the compound. The respective samples are extracted and run through HR-LCMS, where their varying metabolite expression can be quantified by various software packages such as MS-DIAL.

This type of screening has prominent issues as well. The incomplete silencing or expression of the trait, in addition to ion suppression, can lead to inaccurate and subjective quantifications of relative metabolite abundance. Additionally, the procedure requires high mass accuracy LCMS as well as many replicates to provide statistical significance, an expensive and time-intensive prospect. Of particular concern, manipulating genetics

directly in strains is still reserved to a very small subset of existing microbes, as it is in itself time intensive and challenging.

The challenges that face these methodologies leads to a need for another form of natural product elucidation. While facing many issues, comparative metabolomics can still serve as a very useful tool for screening, especially when genetics are available. Chapters 3 and 4 include the application of this methodology to *Streptococcus mutans*.

1.5 Labeling as a Tool in Natural Product Discovery

Novel efforts to label small molecules for the purpose of discovery has spanned a variety of approaches. Imaging techniques including Raman and fluorescence probes allow for the visualization of these molecules in biologically relevant environments. Mass probes on the other hand, are of particular use due to their minimal deviation from molecular and biological characteristics, as well as their ability to be retained following biosynthetic derivatization.

Targeted isotope tagging in particular is a widely wielded technology with applications in the elucidation of biosynthetic pathways[28], mechanistic comprehension[29], and even in the context of molecular discovery[7]. A particularly common methodology is the feeding of labeled primary constituents like amino acids or sugars to track metabolic flux through a cell. Untargeted isotope tagging is a much less explored space. Tagging broadly across classes of molecules is rarely performed due a variety of difficulties. One example of using this technique includes the derivatization of hydroxyl groups on molecules with a Stiglich Esterification in order to fix on mass tags[30]. Another example includes the use of untargeted bromination to label proteins[31]. These methods are limited by the functional groups that can be derivatized, the reactive chemistry that they incorporate, and the capacity to detect the label through mass spectrometry. Missing in this toolbox of biochemistry, is the ability to target the broad spectrum of secondary metabolites produced by the cell in a scaffold-independent manner. Such a methodology would be greatly helpful for the tagging and screening of small molecules for various bioactive functions.

Chapters 2 and 3 sets out to establish and apply a novel labeling methodology for natural product discovery. Chapters 4, and 5 provide a different twist into the drug discovery pipeline. These chapters illustrate the deductive power of pairing various labeling strategies with comparative metabolomics. They also focus on discovery of small molecules outside of the paradigm of cytotoxicity.

Chapter 2: Deuterium Adduct Bioactivity Screening (DABS)

2.1 Introduction

It was set out to develop a technique for the rapid, universal, and economically feasible labeling of all natural product small molecules. It was particularly desirable to develop a methodology for detecting these molecules through mass spectrometry without any prior knowledge of their properties. This system could then serve as a broad band method for surveying and detecting metabolites of interest in natural product extracts. The candidate labeling system deemed most suitable to coincide with these requirements was the deuteration of small molecules with deuterium oxide (D_2O). All future references to H_2O will be referring the common protiated version of water, in opposition to fully deuterated D_2O.

In all known fatty acid biosynthetic pathways, a reduction of NADPH to NADP+ during elongation results in the delivery of a hydrogen originally sourced from water[32]. Along with a variety of other hydride transfers delivered to carbons through primary metabolism, this exchange results in the dispersion of hydrogens specifically sourced from water through nearly all pathways involved in the manufacturing of small molecules. Accordingly, the addition of D_2O to a growing cell culture results in widespread scattering of the label. Bacteria and fungi were the first targets for this labeling due to their quick doubling time and their diverse proliferative production of secondary metabolites. Isotope labels can cause cytotoxic effects due to the isotope effect and its impact on molecular hydrophobicity. Previous work has also been done to illustrate these organisms' tolerance to extremely high concentrations of D_2O, many tolerating a total exchange from H_2O[32]. This cytotoxic resilience, paired with the ensured

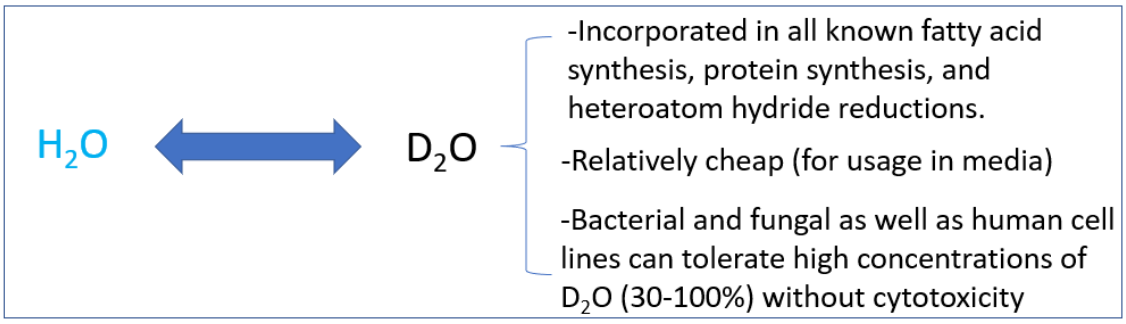


Figure 7: D2O labeling benefits for whole metabolome labeling

label incorporation, provides a versatile mass labeling strategy for a diverse source of natural products. Another goal was to utilize a labeling strategy that would not only supply a mass tag detectable through comparisons between a control and labeled group, but one that could be detected solely from the unique nature of its isotopic distribution. This system would provide a sampling simplicity and minimize the extensive equipment time required by traditional comparative metabolomic discovery methods. An inspiring strategy used by the Bertozzi group to detect brominated proteins looked at the average expected isotopic distribution for bromine within a typical protein[31]. If a comparable strategy could be applied, this would allow a deuterated group of compounds to be applied in a variety of assays and have both their components and derivatizes easily detectable in liquid chromatography mass spectrometry, without reference to an

undeuterated control. The derivatives were of particular interest because it could allow the tags to being altered in various microbial assembly lines and still retain their unique signature. This would allow one to observe small molecule incorporation of deuterated precursors, as well as observe systems with pro-drug, synergistic, or defensive interactions.

When using deuterated water in combination with other proton sources such as unadulterated water and other media components, a unique variety of isotopologues can be generated from a small molecule. Show in Figure 8 an example molecule is illustrated with the molecular formula $C_{30}H_8N_2O_4$. This formula is arbitrarily chosen due to the low hydrogen number facilitating explanation. In a situation where incorporation accounts to 25% deuterium, a curve of isotopologues should exist maximizing at 25% of the eight hydrogens (6H-2D). A predictable percentage of molecules would statistically incorporate no deuteriums (8H-0D) as well as every combination possible. Since these molecules all elute on the HPLC column at approximately the same time, the resulting distribution creates an elevated mid-region of isotope frequency. At high hydrogen values or deuterium incorporation maximizing around 50%, the curve begins to resemble a gaussian distribution. This is due to the principle that samplings of binomial distributions (in this case H or D) begin to resemble gaussian distributions as the samplings go to infinity[33]. This shape is shown to deviate such a significant way from molecules in a given mass range, that deuterated molecules can be detected purely through the unique nature of their isotopologues.

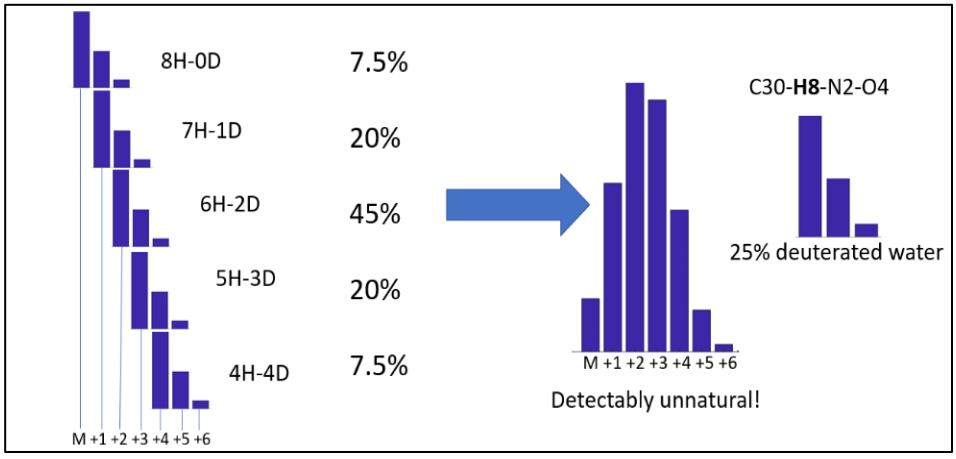


Figure 8: Deuterated distribution of small molecule expected with 25% deuterated water incorporation

2.2 Computational Methodology and Assay Techniques

2.2.1 Deuterium Tagging and Identification

A database including all potential combinations of formulas between 100-1500 Da was generated using the atoms carbon, nitrogen, oxygen, and hydrogen. These atoms are the most common occurring in biomolecules and allow a representative sampling of most natural products. A Matlab script queried "chemcalc.org", a formula prediction server, recursively until complete coverage of the 100-1500 Da mass range was covered. This list contained formulas that are not biologically or physically relevant (e.g., CH_{90}), due to

electron configurations or the sp orbitals available to any given atom. In order to address this, the formula list was curated by applying several of the Seven Golden Rules (SGR), a set of predicted formula distributions based off chemical frequency in a given mass range.[34] The SGR method addresses unlikely molecular formulas using both physical constraints (e.g. balanced charge states) and also uses a database of existing natural products to provide typical configurations of small molecules versus extremely rare formulas. These rules were used to eliminate candidate formulas that were more than three standard deviations away from typical formulas, allowing for a much more representative list of potential molecular formulas possible.

From the curated formula list, the expected isotopic distribution of a given formula was predicted using natural isotopologue frequencies of each atom[35] with a Matlab script. These frequencies are constant across all natural products and all naturally occurring molecules. A second script was utilized that predicted the isotopic distribution of a compound that is labeled with an extra neutron from deuterium approximately 6% of the time. This number was selected as beneath the lower limit of deuterium incorporation

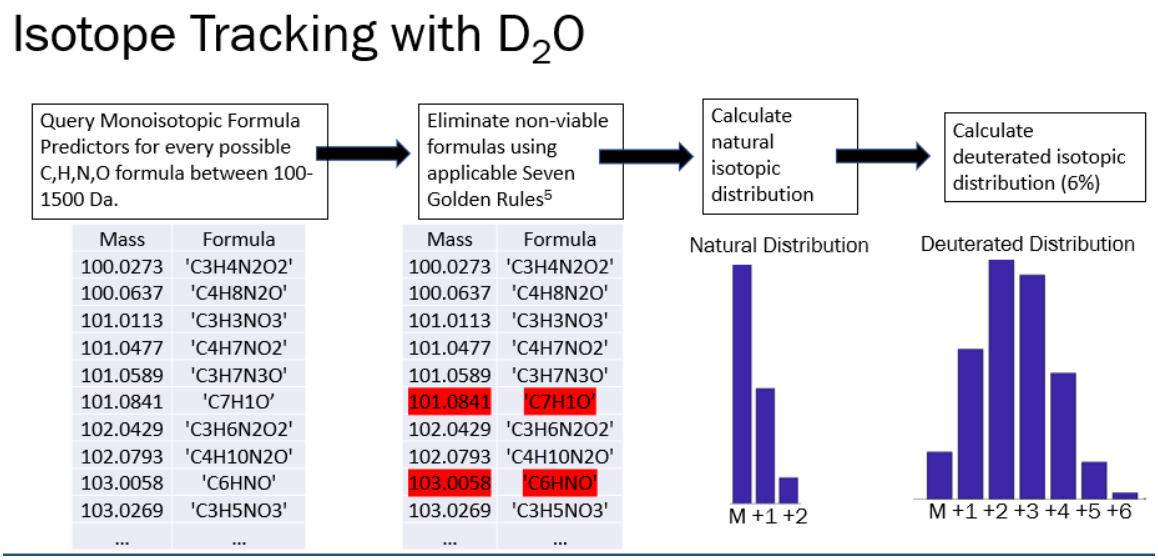


Figure 9: Methodology for molecular formula combinations, elimination of non-viable formulas, and predicted isotope distributions amongst natural and deuterated abundances

observed in known biomolecules at the concentrations used experimentally (25% D_2O). Although water contributes to a number of hydrogens within the metabolome, hydrogens can also come from other sources such as sugars or peptides found in the media components. Figure 9 provides an example of this sampling technique.

One method that sufficiently differentiated the deuterated and naturally occurring isotopologues was the analysis of the (M+1) and (M+2) peaks relative to the (M) peak. These refer to the base peak with no extra neutrons (M), one extra neutron (M+1), and two extra neutrons (M+2) respectively. Using the isotopic distributions, the data was selected for the local maxima of (M+1) and (M+2) peaks in any given mass range. For small molecules, the ratios of these higher peaks over the base peak are typically quite small. In contrast, for deuterated small molecules, the ratios are typically quite large. These respective ratios are plotted in Figure 1 with blue as the natural distribution and red as the

deuterated distribution. A curve representing the maximum isotopic distributions observed in naturally occurring samples is shown as the green line. Statistics regarding the frequency of deuterated samples exceeding this curvature were predicted to examine deuterium's detection capabilities. Combining both data sets, molecules that are deuterated supersede the maxima line 93.5% of the time. This means that with 93.5% of molecules deuterated

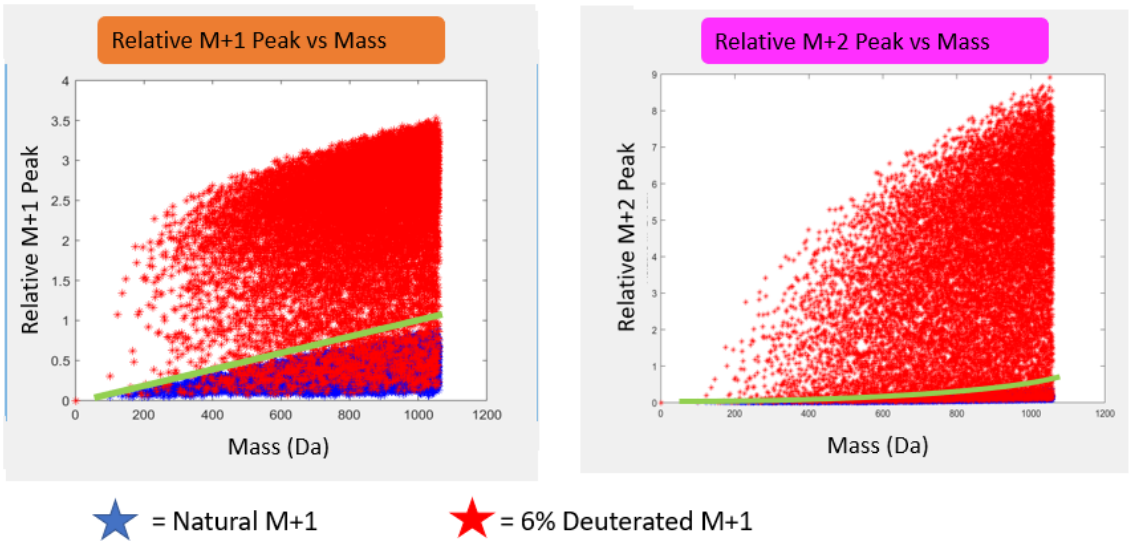


Figure 10: (M+1)/M and (M+2)/M relative isotopic abundances depcicted for all CHNO chemical formulas between 100-1100. The blue star illustrates natural distributions, while the red depicts deuterated distributions.

to 6%, using this criterion allows a high level assurance that the molecule has been artificially tagged and virtually eliminates false negatives. Compounds failing to pass this requirement are typically highly unsaturated molecules with low hydrogen and high oxygen counts. Curiously, a significant deviation in carbon formula abundance was also illustrated, with low carbon counts prohibiting detection. This is expected to be due to the low carbon count constraining the max number of hydrogens based off SGR predictions.

Mass spectrometry data was collected using Mass Hunter Data Acquisition. These files were exported into the preprocessing software MS-Dial. MS-Dial performs retention time corrections and collapses centroided data into single peaks per ion (peak picking). A Matlab script "D_2O Sorter" was written to predict whether a given small molecule had been labeled. The Matlab code for this program is included in Appendix C. This script's input included the peak picked data from MS-Dial. It sorts for peaks expected to be isotopes of each other. Significant retention time drift was allowed between samples due to hydrophobic secondary isotope effects induced by deuterium. Each addition of a deuterium atom slightly increased the molecules polarity. These compiled peak lists were then screened with their isotope distributions compared to the natural isotope maxima predictions of the given mass region. Depending on whether they met the (M+2) and (M+1) relative height requirements, the peaks were sorted into respective "Labeled" and "Unlabeled" matrixes. These matrixes provide mass, retention time, and abundance data. These results can then be manually observed and referenced back to the original Mass Hunter data to determine which masses correspond to the labeled compounds of interest.

2.2.2 Cell Binding Assay

In order to overcome the limitations of traditional bioactive screening assays, a cell binding assay was used. This assay has been used in human cancer cell lines to screen to identify compound localization in cells and assess bioactive featires[30]. To perform the assay, one mixes a recipient cell line with the compounds of choice. The recipient cell line is washed after incubation to remove weak binders. Drug bioactivity typically emerges from binding to a specific moiety, whether that's a receptor or cytosol localized protein. This binding encourages localization of the molecule inside or on the surface of the cells. Accordingly, cell localization can provide a valuable method for observing bioactivity in a very broad context. Using this method as a drug screening tool emerges as a relevant decision if the extracts of choice can be appropriately tagged, allowing for tracking of the molecules after incubation and any following derivatization.

2.2.3 Deuterium Adduct Bioactivity Screening Assay

The combination of deuterium labeling and the cell binding assay is dubbed Deuterium Adduct Bioactivity Screening (DABS). This methodology takes advantage of the holistic tracking provided by deuterium, and the broad spectrum of bioactivities identified by cell binding, to produce an assay that provides the means to identify novel natural products with a wide range of bioactive potential. The procedure for this assay is detailed in Figure 11. In short, a producer strain is grown in deuterated water and extracted to generate a complex mixture of labeled natural products. This extract is incubated with a recipient strain that serves to test the binding of these small molecules to the cell line. Following incubation, the recipient cell line is washed to remove non-binders, and the strain is extracted and screened on a mass spectrometer. The file is run through the deuterium program as detailed in Appendix C and the list of labeled compounds in generated. This workflow identifies molecules provided by the producer strain that showed some form of cell binding that could translate to bioactivity. This method additionally allows the observation of derivatized binders as the mass tag is retained following derivatization.

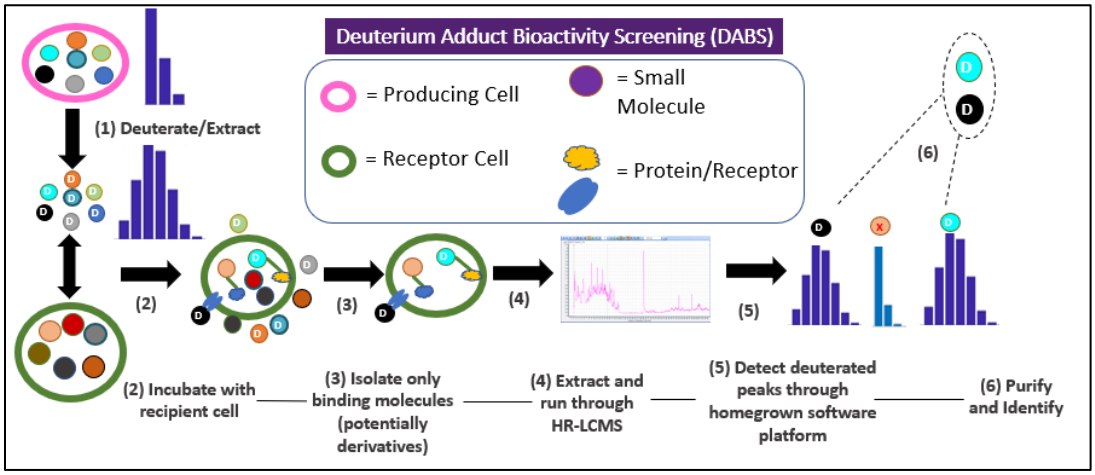


Figure 11: Deuterium Adduct Bioactivity Screening workflow

When comparing DABS with other technologies for identifying bioactive small molecules, several advantages become obvious. DABS can reliably identify strong cell binders independent of native microbe compound abundance. Advantageously, DABS

does not need comparative software or replicates to reliably identify a candidate. The process is conducive to a high throughput workflow, decreasing researcher manpower and the potential for compound degradation. The platform can easily identify families of binding compounds and doesn't rely on chromatographic separation for efficacy. One key advantage of DABS over comparative metabolomics is the lack of reliance on mass accuracy, retention time fidelity, or ion suppression in mass spectrometry. The platform only needs isotopic fidelity, which is a feature that can enable complex extract metabolite identification with even very low-resolution mass spectrometers. The platform requires very little compound to screen, with only the threshold of detection by mass spectrometry serving as the lower limit. This is a radical transformation from bioactivity screening methods practiced currently. By circumventing this limit, one can expand screening sensitivities to molecules at nanomolar concentrations and lower in bacterial extracts. Very importantly, DABS allows for the assessment of bioactive natural products emerging from strains of microbes that have no genetic tools established. By circumventing this challenging barrier, the platform greatly expands the scope of microbes that can be surveyed by researchers. By expanding the strains accessible, the molecules identifiable, and the equipment constraints usable, this platform can be a valuable research tool and lead to remarkable discoveries. Table 1 details the various platform features.

Table 1: Qualitative examination of existing methodology for the discovery of natural product secondary metabolites

Features	Bioactivity Based Fractionation	Comparative Metabolomics	Deuterium Adduct Bioactivity Screening
Time to Compound Identification	Weeks-Months	Days	Days
Necessary Compound Abundance	Very high	Low	Very Low
Labor Intensity	High	Low	Low
Screening Rate of Molecules	Very slow	Very slow	Very high throughput
Bioactivity Comprehensivness	Very Narrow	Very Narrow	Very Broad
Instrumentation Requirements	Inexpensive	Very expensive	Inexpensive

2.3 Validation

2.3.1 Validation of Deuterated Detection

Strains known to produce a number of known natural products were selected. This list comprised of a variety of *Streptomyces* with a prominent and well characterized metabolite present. Although *Streptomyces* produce dozens of uncharacterized metabolites, these known small molecules provide a proof of concept for labeling.

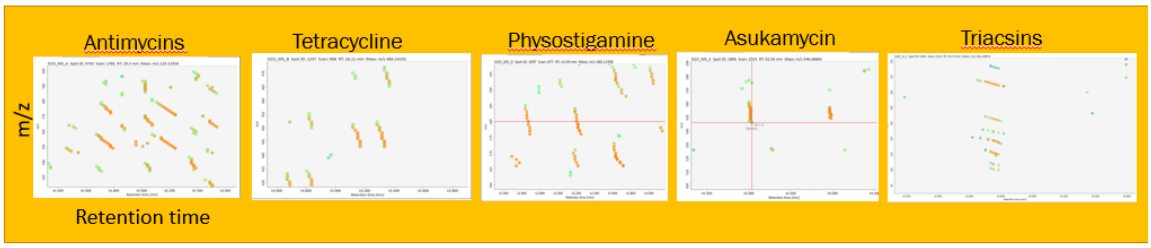
The various *Streptomyces* strains were cultured using Mannitol Soya (MS) media on agar for 10 days at 30 C. The strains were extracted 1:1 with organic solvent using 95:5 ethyl acetate:methanol and incubated for 30 minutes during extraction. The extracts were dried and resuspended at 10x in DMSO.

The human cancer cell lines were grown to 90% confluency at 37 C, 5% CO_2 in DMEM + 5% FPS in 12 well plates. The cells were incubated with the *Streptomyces* extracts for 4 hours in the same conditions. The cells were gently washed three times with prewarmed PBS. Following the wash, the cells were suspended in 1mL of PBS and scraped to suspend the cells. These were transferred to a 15 mL falcon tube and extracted 1:1 with organic solvent using 95:5 ethyl acetate:methanol. The resulting extract was dried down and resuspended at 10x in methanol. 100uL human cell culture equivalent was injected onto the LCMS. LCMS procedures are detailed in Appendix A.1.

Table 3: Growth conditions for various strains used in the DABS platform

Strain	Media	Growth Conditions
Actinomyces – Validation	MS-agar	10 days at 30 C
Streptomyces albus	MS-agar	10 days at 30 C
Streptomyces aureofaciens	MS-agar	10 days at 30 C
Streptomyces griseofuscus	MS-agar	10 days at 30 C
Streptomyces nodosus asukaensis	MS-agar	10 days at 30 C
Streptomyces tsukabensis	MS-agar	10 days at 30 C
Actinomyces – zillamycin purification		
Streptomyces griseofuscus	MS-agar	10 days at 30 C
Human Cell Lines		
HeLa	DMEM + 5% FBS	37 C to 90% Confluency
U87	DMEM + 5% FBS	37 C to 90% Confluency

Each of the strains was grown in deuterated water, extracted, and run on LCMS. Following the workflow detailed in Section 2.2, each of the extracts were screened and had a labeled list generated.



Strains	Known Product	Product Detected by Software?	Features of Incorporation
S. albus J1074	Antimycins	Detected	• Drift is compound dependent, though somewhat consistent amongst the compound • 8-10% incorporation is consistent • Many other products labeled
S. aureofaciens	Tetracycline	Detected	
S. griseofuscus	Physostigamine	Detected	
S. nodosus asukaensis	Asukamycin	Detected	
S. tsukabensis	Triacsins	Detected	

Figure 12: Known natural products depicted on an m/z vs retention time axis with additional details about the deuterium software kit

On all of these lists, the test metabolite was readily detected by the software and their deuterated distributions as depicted by MS-Dial are shown in Figure 1. Each strain had well over 50 other small molecules predicted to be deuterated. This is encouraging because it confirms a wealth of metabolites existing in these strains that have yet to have any characterization.

2.3.2 Validation of Cell Binding

The strains selected for qualification of the DABS platform included many of the previously examined *Streptomyces*. *S. albus* has been a target organism by the Zhang lab in the past for its production of antimycins. Specifically, cell binding assays using Raman labeled antimycins in HeLa cells had shown the methodology to be effective.[30] This labeling work was repeated with deuterated antimycins in the place of the Raman tags. Control species were used to ensure appropriate binding is observed. Figure 8 shows the breakdown of these molecules tested. Tetracycline has known bioactivity in many cell types but fails to penetrate mammalian cells. Physostigmine has both transport capabilities and bioactivity in HeLa cells but binds through covalent mechanisms. Since this strategy should only detect extractable non-covalent binders, this compound served as a control to ensure wash steps appropriately removed small molecules that could diffuse into the cells. Asukamycin was also examined, a molecule that has known activity in HeLa cells. This molecules mode of action is unknown, but the compound shares many scaffold similarities to manumycin type antibiotics, all covalent binders. Accordingly, its lack of detection fit with expectations.

Table 2:Validation of DABS through known Streptomyces product screening

Strain	Known Product	Mammalian Cell Uptake	Binding Mechanism	Expected Cell Binding Detection	Post Cell Binding Detection
S. albus	Antimycins	Yes	Non-covalent	**Detected**	**Detected**
S. aureofaciens	Tetracycline	No	Non-covalent	**Not Detected**	**Not Detected**
S. griseofuscus	Physostigmine	Yes	Covalent	**Not Detected**	**Not Detected**
S. nodosus asukeansis	Asukamycin	Yes	Likely covalent	**Not Detected**	**Not Detected**

2.4 Discovery of Zillamycin through DABS

DABS was performed using a panel of *Streptomyces* and a panel of human cancer cell lines. Excitingly, a novel metabolite was detected from the deuterated *Streptomyces griseofuscus* extract. This metabolite, with an m/z of 468.43, appears to bind both HeLa and U87 cancer cells (cervical and brain cancer cell lines respectively). Due to the chronic issue of dereplication in natural product discovery, this metabolites MS2 spectra was paired against a database of known metabolites using the Global Natural Products Social Molecular Networking (GNPS) server[36]. This service compares the various MS2 spectra present in the sample against each other and against a larger database of known molecules. If a sufficient number of fragments or neutral losses are shared between molecules, the molecules are paired together and assigned a similarity score. This provides a means of not only screening out molecules that have been previously discovered, but also screening for molecules that are very similar to metabolites previously discovered. This metabolite shows no similarity with known metabolites from *S. griseofuscus* or from the broader metabolite databases. All of these features incentivized the purification and structural elucidation of the molecule.

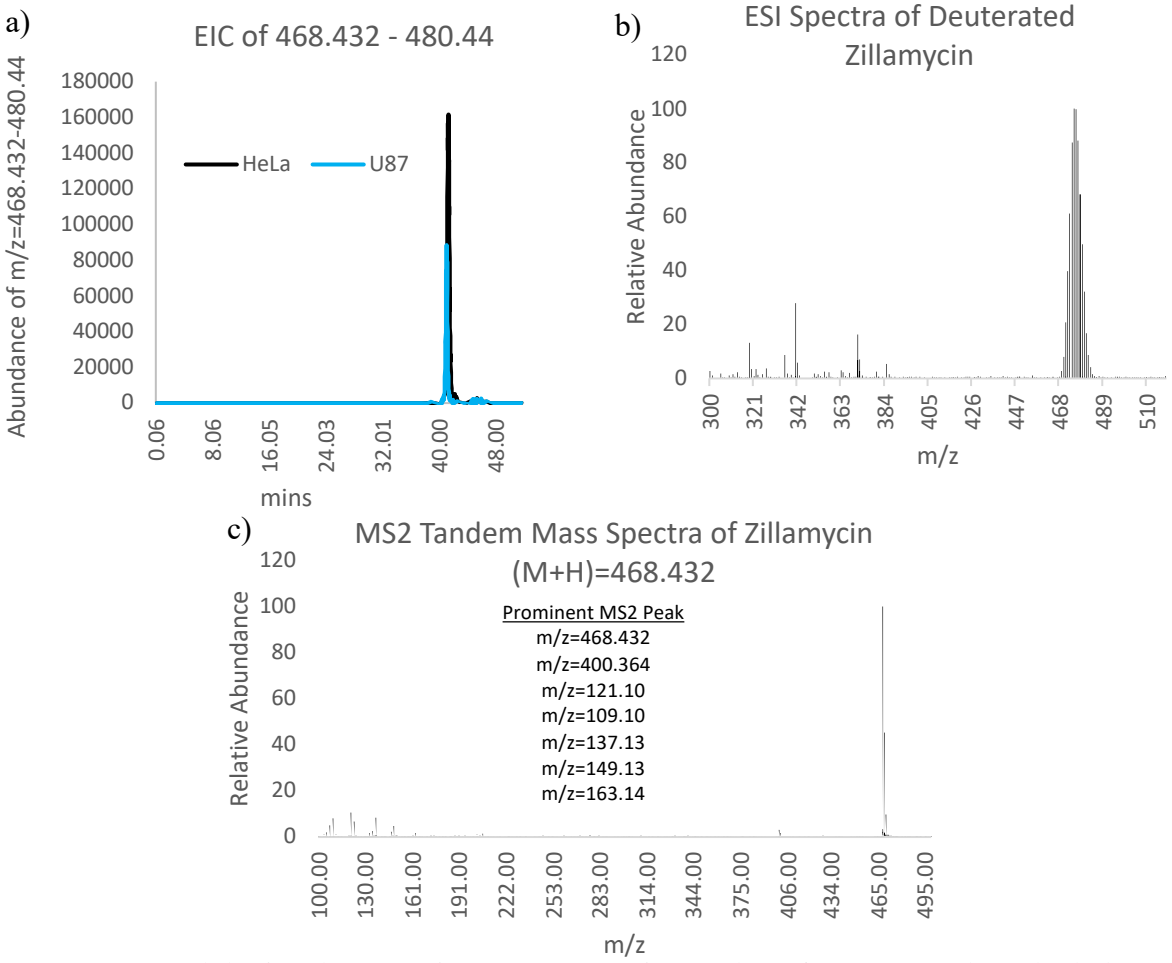


Figure 13: A metabolite from the extract of Streptomyces griseofuscus with an m/z=468.432 was detected as both deuterated and cell binding, satisfying the DABS criteria. a) Extracted Ion Chromatogram (EIC) of m/z=468.432 and it numerous isotopologues following DABS of S. griseofuscus extracts against HeLa cells (black) and U87 cells (blue). b) The electrospray ionization (ESI) spectra of deuterated zillamycin, shown as the grouped set of peaks in the 468-480 m/z region. c) The tandem mass spectra (MS2) of m/z=468.432. Prominent peaks are listed in order of abundance.

2.5 Purification of Zillamycin

2.5.1 Growth Conditions Optimizations

Streptomyces are relatively slow growing microorganisms, taking typically 5-10 days to reach stationary phase on plates[37]. Given the slow growing conditions of these organisms, a number of production schemes were tested to find optimal growth conditions for zillamycin. This included a variety of medias such as MS and ISP2, a variety of aeration conditions including liquid culture shaken growth and growth on agar plates, as well as sub-optimizations of candidate medias.

Streptomyces griseofuscus grown for 12 days on a lawn on MS-agar plates was identified to have the best production characteristics of 468. Of particular note, a variety of soy flours were tested including "defatted soy flour", "Bob's Red Mill Soy Flour", and "Kinako Soy Flour", all producing variable amounts of the 468 molecule. "Bob's Red Mill Soy Flour" was shown to produce the greatest amount of 468. This variation should be noted for future intended purification of the molecule.

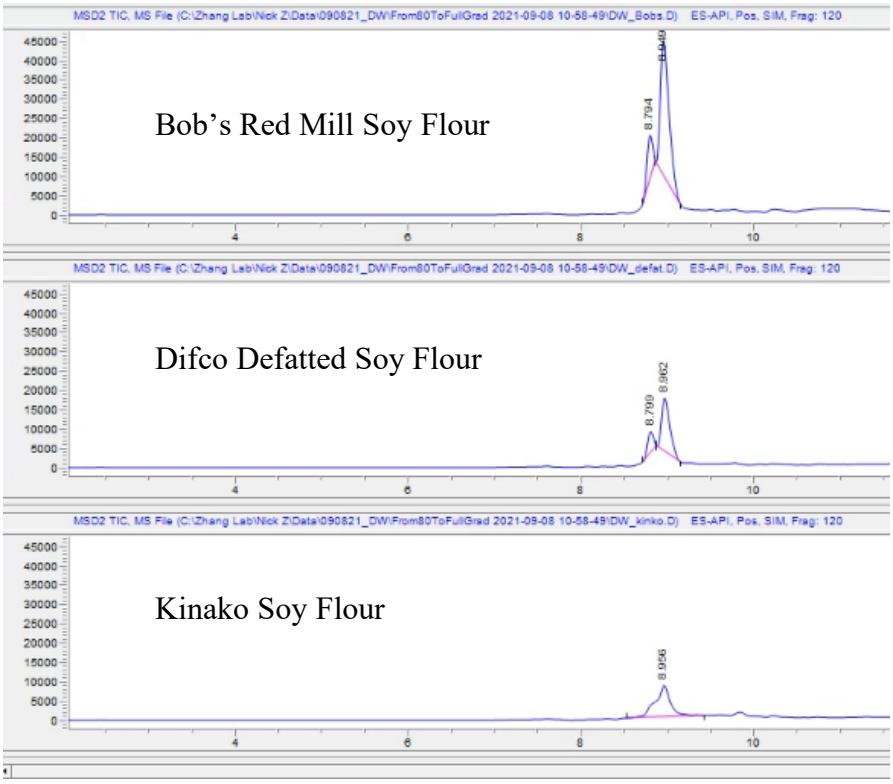


Figure 13: Abundance of zillamycin under various MS Soy Flour formulations. Bob's Red Mill Soy Flour showed best production

2.5.2 Large Scale Production and Purification

A number of extractions were tested with the metabolite, henceforth dubbed zillamycin. A variety of common organic solvents were used to test the best extraction.

80:20 ethyl acetate:methanol was shown to have the best extraction efficiency at small scale. This was used for the remainder of the extraction.

4Ls of *S. griseofuscus* was grown on MS-agar plates as lawns for 12 days. Upon the last day, the plates were extracted with 80:20 ethyl acetate:methanol. This organic phase was centrifuged to collapse the interphase, and the organic phase was separated from the remaining components. This extract was dried down using a rotovap and resuspended in 5 mLs methanol. The extract was quite oily and experienced phase separation at lower volumes of methanol.

This extracted sample was run through an LH-20 size exclusion column, with fractions being tested by LCMS for presence of 468. Appendix F.1C shows the elution of the compound as detected through fractions tested on a mass spectrometer. The fraction containing 468 was dried and resuspended in 1mL of methanol. These contents were separated using an SP-C18 packed column. The column was washed with acetonitrile + 1% TFA and methanol, before eluting the 468 compound with methanol + 1% TFA.

This fraction was concentrated to 1 mL of methanol and injected onto an HPLC in 100uL injections. Specifics of the HPLC method are detailed in appendix F.1B. The fraction containing eluted at 20 mins with a UV maximum of 208nm.

2.5.3 Structural Elucidation of Zillamycin

The resulting HPLC fraction was concentrated and resuspended in D_3 Chloroform. NMR was performed and the structure was characterized, discussed in Appendix F.1D. The compound has a six chain isoprenyl tail with a guanidine moiety fixed through a C-N bond. This compound's closest relatives are other guanidine alkaloids like metformin and galegine[38,39]. This is encouraging given the history of guanidine alkaloids serving as medications, most commonly for their effect on pancreatic cells[40].

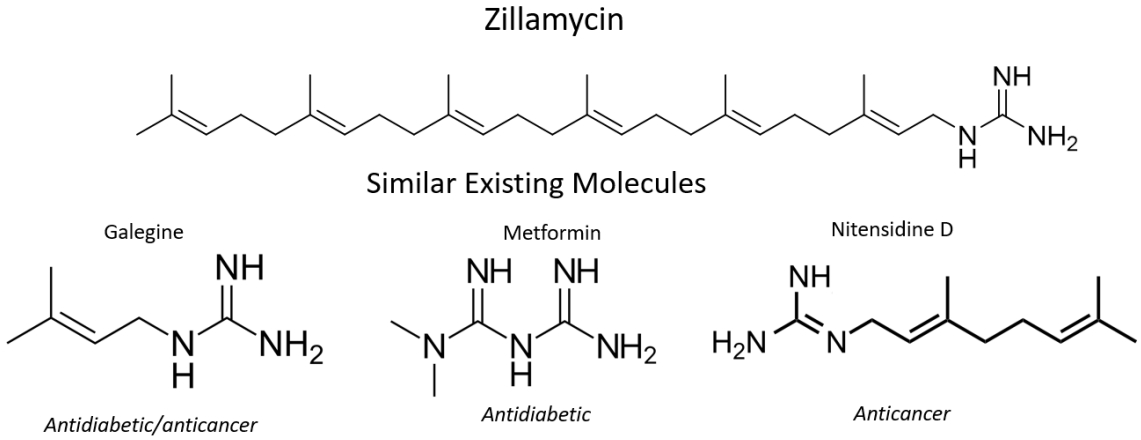


Figure 14: Zillamycin structure and guanidine alkaloids with similar structures

2.5.5 Some Biosynthetic Considerations

While the exact biosynthesis of zillamycin is unknown, the sequenced genome can provide some insight. Isoprenyl biosynthesis is ubiquitous to all life[41]; in *Streptomyces griseofuscus*, a variety of genes can be detected that could produce the isoprenyl chain. For attachment of the guanidine moiety, only one major enzymatic class is common. These amidinotransferases are less common than isoprenyl moieties and are only present in two areas on the genome. Unfortunately, no isoprenyl biosynthetic gene co-localizes with the amidinotransferase. This could indicate that the genes responsible for synthesis of this molecule are not part of a cluster, or that the responsible genes are biosynthetically not yet well understood. The genome of *S. griseofuscus* for this project was sequenced in house. Amidinotransferase similarities were detected using AmtA, an amidinotransferase involved in the biosynthesis of phaselotoxins produced by the *Pseudomonas* family[42]. Two different amidinotransferases were detected in the *S. griseofuscus* genome; the Blast result is displayed in **Error! Reference source not found.**. Interestingly, all other public *S. griseofuscus* genomes don't contain any amidinotransferase genes, indicating a unique feature among this strain.

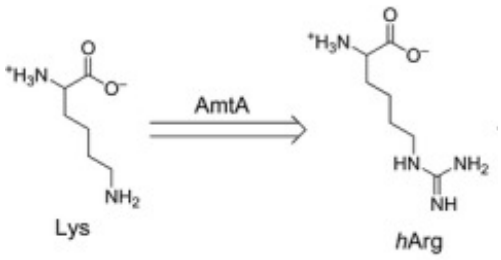


Figure 15: Amidinotransferase mechanisms adapted from Li et al. 2016

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Gri_R1_(paired)_trimmed_(paired)_contig_27		125	125	92%	1e-33	35.16%	43302	Query_47069
☑	Gri_R1_(paired)_trimmed_(paired)_contig_51		71.2	71.2	98%	8e-15	24.05%	186779	Query_47093

Figure 16: Blast result searches using AmtA, an amidinotransferase produced by the Pseudomonas family. Two independent amidinotransferases were detected on the genome

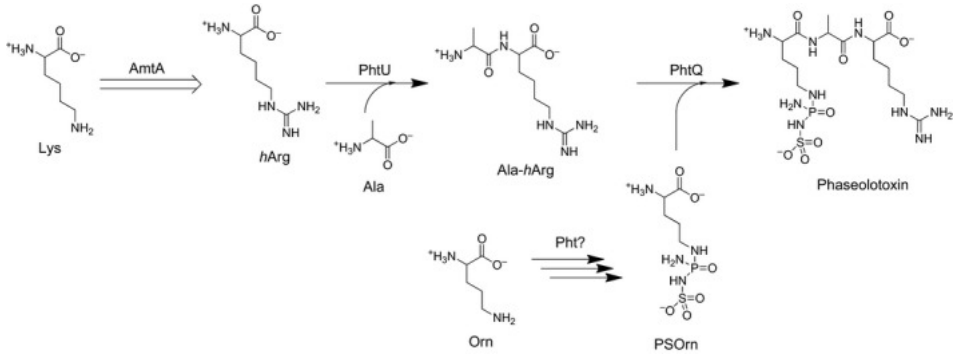


Figure 17: Mechanism of AmtA, the amidinotransferase used to blast for comparable genes in Streptomyces griseofuscus

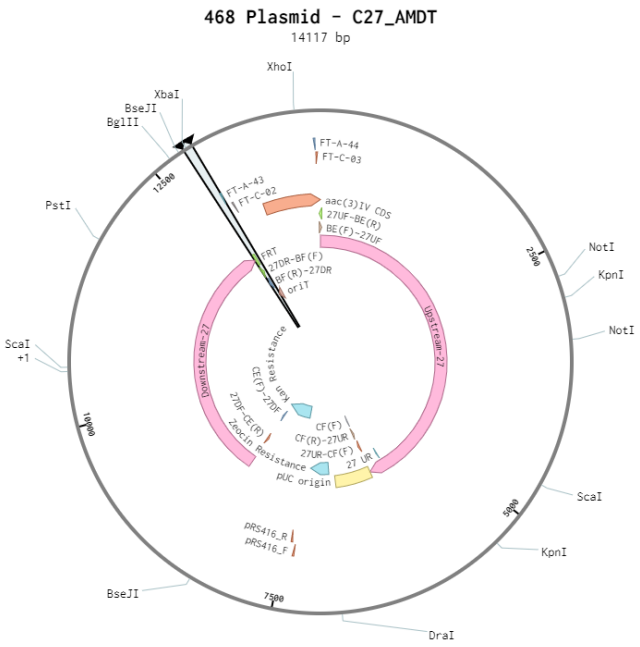


Figure 18: Representative plasmid constructed for the knockout of the two amidinotransferases located on the griseofuscus genome.

Attempts were made to construct gene knockouts of the *S. griseofuscus* amidinotransferases, given their limited presence on the genome. *Streptomyces* strains typically require large homologous arms to produce double crossover constructs through conjugation. The constructs designed for this work are shown in the Figure below. Homologous arms of at least 4000 base pairs were selected around the respective amidinotransferase sites. These were intended to combine in a Gibson assembly with a backbone containing an *E. coli* origin of replication with an apramycin resistance marker, a conjugation origin of transfer, and a Streptomyces translatable kanamycin resistance marker. Unfortunately, the initial PCRs performed to generate the fragments were not successful. A variety of primer and polymerases were utilized with limited success. Due to the extremely high GC content of *Streptomyces* strains, PCR reactions with appropriate melting temperatures can be quite challenging[37]. This may have contributed to the inability to produce the final constructs. Future work should be considered better understand the biosynthetic nature of zillamycin.

2.6 Biological Activity of Zillamycin

2.6.1 Microbial Binding

This purified compound's ability to affix to microorganisms was assessed as well. A panel of organisms including *Candida albicans*, *Pseudomonas aeruginosa*, and *Streptococcus mutans* were assessed for zillamycin binding. Each culture was grown overnight in BHI. The following day, 1 mL samples were aliquoted from the overnight cultures and were inoculated with 5 mg/L zillamycin. These samples were cultured for 4 hours at 37 C. Following culturing, each strain was spun down and washed with fresh PBS for 3x. Following the third wash, these samples were extracted in methanol and injected

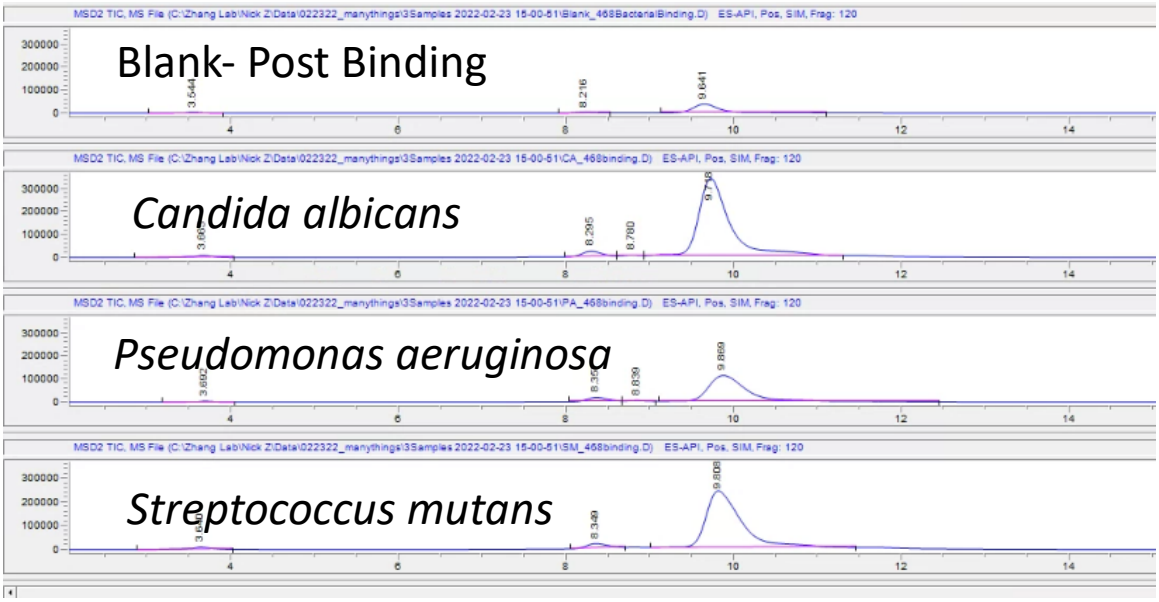


Figure 19: Microbial binding assay performed on Candida albicans, Pseudomonas aeruginosa, and Streptococcus mutans with Zillamycin.Time is on the x-axis, and abundance on the y-axis. Significant binding was observed in all three strains.

onto the mass spectrometer.

Evidence shows that the zillamycin compound binds significantly to each of the strains tested. The compound appears to have binding activity that is spans both the microbial and human cell architecture. Microbes were used for the remaining studies due to their rapid growth and easy handling conditions.

2.6.2 Antimicrobial Activity

The antimicrobial activity of zillamycin was assessed. This antimicrobial assay used *Pseudomonas aeruginosa, Streptococcus mutans,* and *Candida albicans*. zillamycin was added up to 50uM. These samples were performed in duplicate allowed to grow for two days in BHI media at 37C.

Significant inhibitory was observed in the *Streptococcal* strain, a gram positive bacterium. This was reproduced and is show in a growth curve in duplicate below.

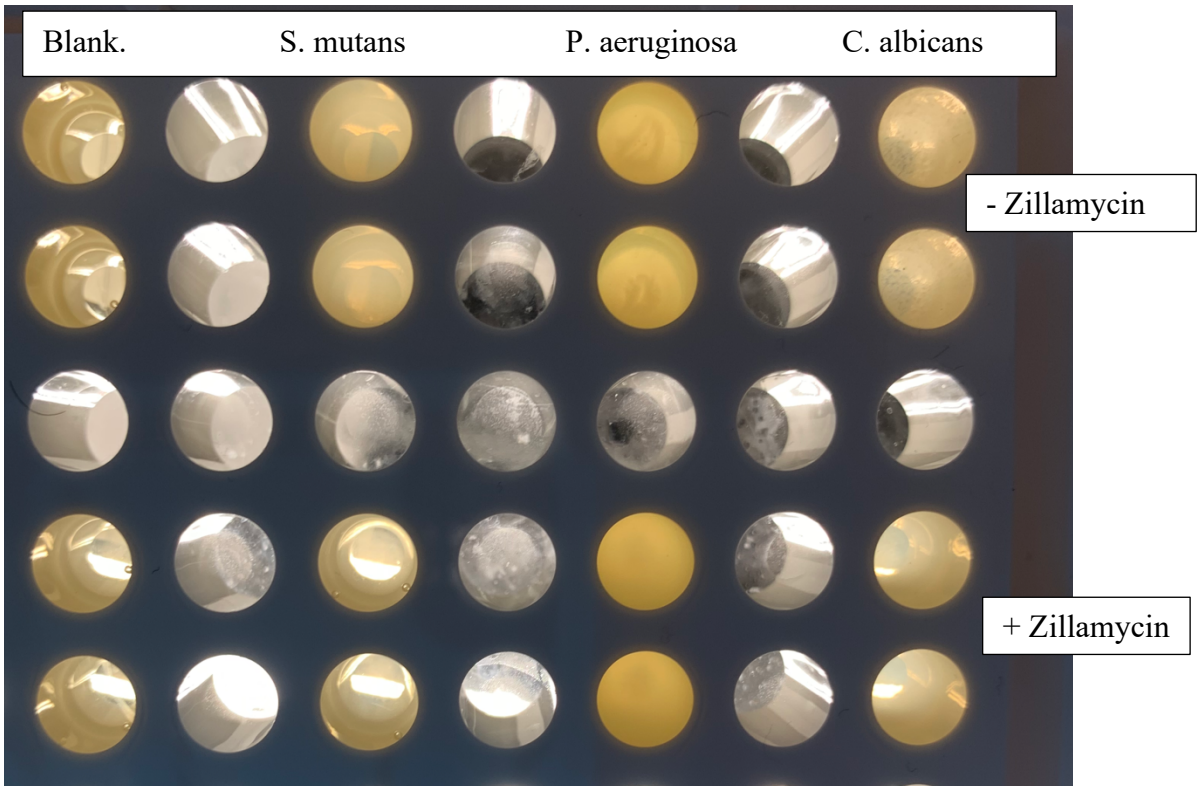


Figure 20: Microbial inhibitory activity of 50 uM zillamycin to Streptococcus mutans, Candida albicans, and Pseudomonas aeruginosa over two days. Significant inhibition was observed in the Streptococcal strain

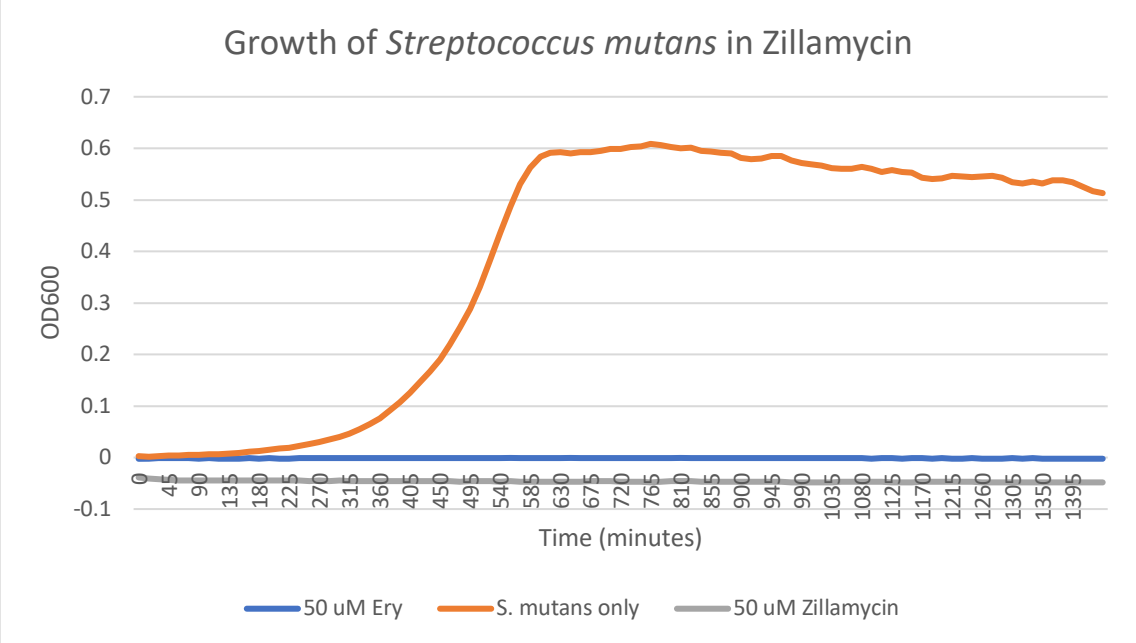


Figure 21: Growth of Streptococcus mutans with a DMSO blank (red line), 50 uM erythromycin (blue line), or 50 uM zillamycin (green line) in duplicate. Significant inhibition is observed of the strain when cultured with zillamycin.

2.6.3 Compound Localization

This led to a consideration that the compounds binding may be localized in the cellular membrane, a feature universal to each of these organisms. The localization could be tested using a subcellular fractionation assay. The resulting fractions include a sample of the unlysed cellular components, supernatant, cell wall, cell membrane, and the cytosol.

Within the cellular components, 468 seems to significantly accumulate in the cellular membrane as expected. Very little ends up in the cell wall, which makes sense given that the compound binds to human cells, which have no cell wall. A small amount ends up in the cytosol so there is some penetration of the compound into the cell. This makes sense entropically given that the compound is able to permeate the membrane as well.

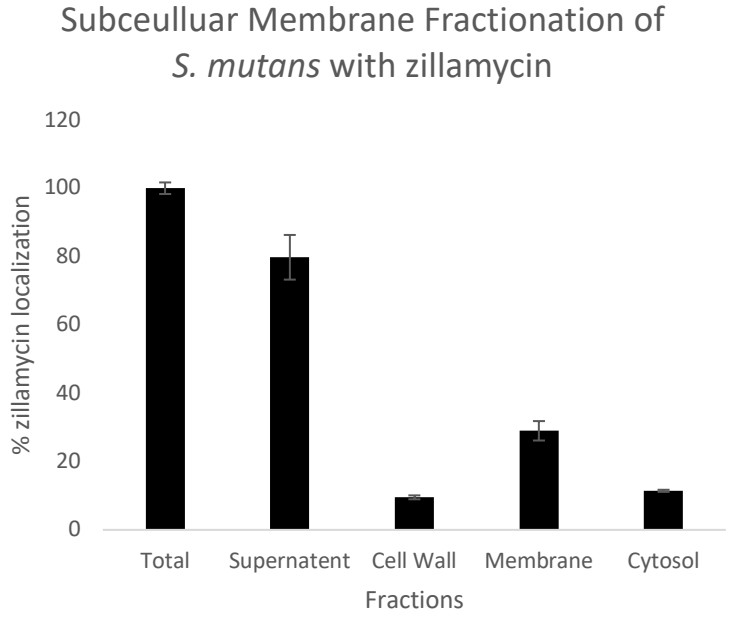


Figure 22: Subcellular membrane fractionation of Streptococcus mutans incubated for 4 hours with Zillamycin. Results indicate that the compound significantly localizes with the cellular membrane.

Chapter 3: Deuterium Labeling for the Identification and Screening of Novel Diterpenoids

3.1 Introduction

Terpenoids are a class of natural product defined as having varying incorporation of C_5 isoprene units[41]. These can be classified into a number of categories such as mono-, di-, and tri- terpenoids[43]. These molecules have diverse applications including antimicrobials, anti-inflammatory, immunosuppressants, antivirals, anticancer drugs like Taxol, or more recently antiparasitics like the malarial drug artemisinin[44,45].

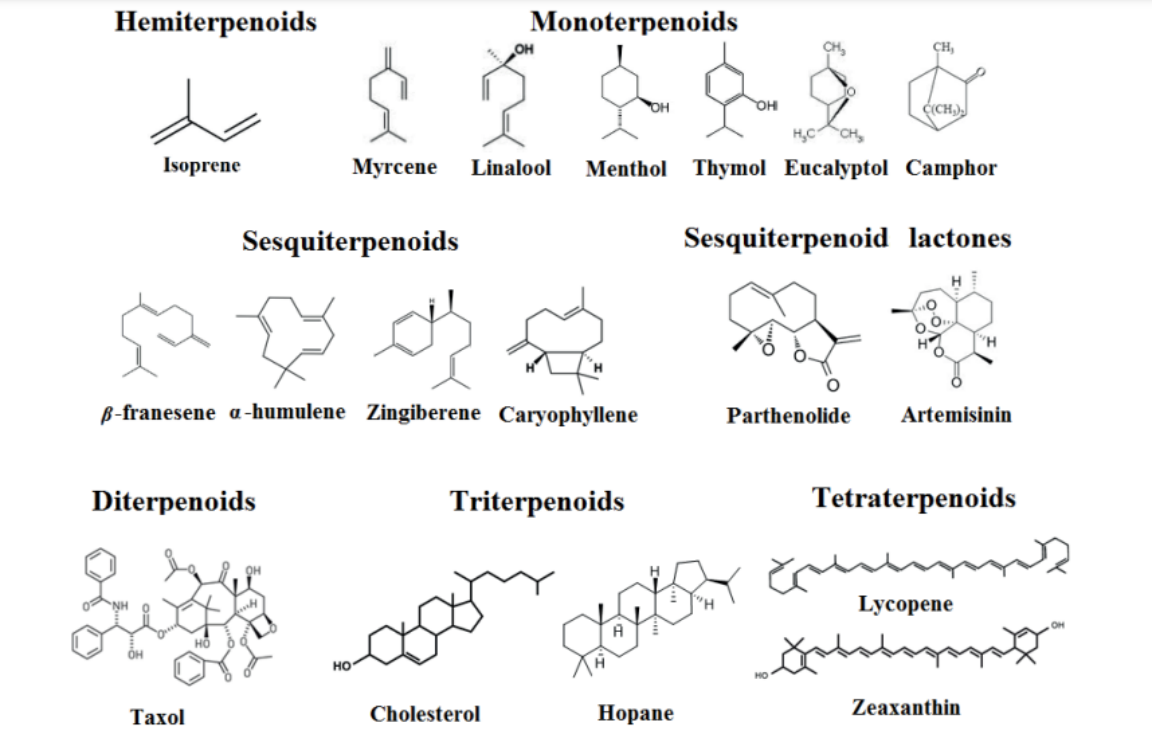


Figure 23: A collection of terpenoid small molecules with various applications in human health adapted from Pervine et al. 2021

Diterpenes are a class of terpenoids characterized by their biosynthesis with four isoprene units[46]. These molecules have served as a source for new medications for decades. In addition to their medicinal role, these molecules also play roles in fields such as food flavorings, fragrances, cosmetics, and vitamins[47]. More than 18,000 diterpenoids have been discovered from both plant and microbial species. These molecules represent a powerful tool in human medicine and their continued discovery is imperative.

This chapter features the methodological utility of deuterium labeling discussed in Chapter 2, paired with novel diterpenoid discovery.

3.2 Targeted Application to Brascilicardin A Analogs

Immunosuppressant drugs serve as medicines for a wide variety of purposes but are most commonly prescribed for patients with organ transplants[48]. These drugs can serve as a lifeline to prevent rejection of the transplanted organ by the host. These drugs also treat

a wide variety of auto-immune conditions. Accordingly, novel immunosuppressant drugs are greatly desired in order to expand the efficacy, safety, and number of conditions treated by these drugs.

Brasilicardin A is a known diterpene immunosuppressant drug produced by the strain *Nocardia brasiliensis*, a highly pathogenic gram-positive bacterium[49]. This compound is a tricyclic metabolite that contains two sugar moieties, rhamnose and N-acetylglucosamine, as well as an amino acid moiety. Brasilicardin A is an extremely promising candidate for use as an immunosuppressant: it appears to have a unique mode of action and shows a higher potency and reduced toxicity profile compared to current standards for immunosuppressants. While this compound represents an excellent clinical candidate for immunosuppressant use, three major hurtles limit research performed with the molecule. *Nocardia brasiliensis* is a BSL2 organisms that is a human pathogen; accordingly, large scale bioreactor fermentations at a clinical scale are not practical or safe. Secondly the strain's gene cluster is incomplete as the late stage glycosylation event comes from an unknown enzyme not thought to be localized with the cluster. This means that attempts at heterologous expression can only produce incomplete scaffolds for the compound. Lastly, the highly complex final scaffold sets a high barrier to any attempts at fully synthetic routes achieving commercial viability. One potential avenue to alleviate each of these barriers would be to find a safer strain of bacteria that produces a similar small molecule.

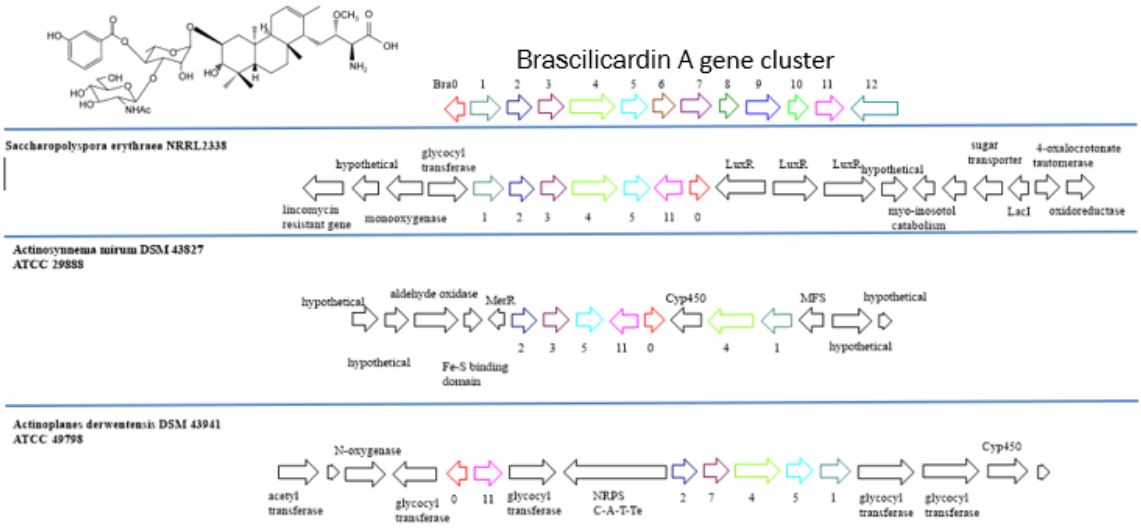


Figure 24: Brasilicardin A gene cluster blasted against a variety of bacterial genomes. Show beneath are three strains with a high degree of homology to the cluster.

The known biosynthetic gene cluster for Brasilicardin A was BLAST against a variety of known bacterial genomes, shown in Figure 2. Multiple strains with comparable biosynthetic capability were identified, including *Saccharopolyspora erythraea* NRRL2338, *Actinosynnema mirum* DSM 43827 ATCC 29888, and *Actinoplanes derwentensis* DSM 43941 ATCC 49798. Of the sequenced strain databases, these respective strains are expected to produce a small molecule diterpenoid that shares scaffold

characteristics that most closely resembles with Brasilicardin A. Given the disjointed genes for the manufacture of Brasilicardin A, one faces the prospect that the homologous clusters may have similarly separated genes in the other strains. Even with a complete gene set, both transfer of the relatively large clusters and then finding conditions for expression present a large roadblock towards characterization of the metabolite. Identification of the end product produced by these gene clusters is a challenging prospect. Focus was originally placed on *S. erythraea* as the strain is well characterized due to its study as the original erythromycin producer[50].

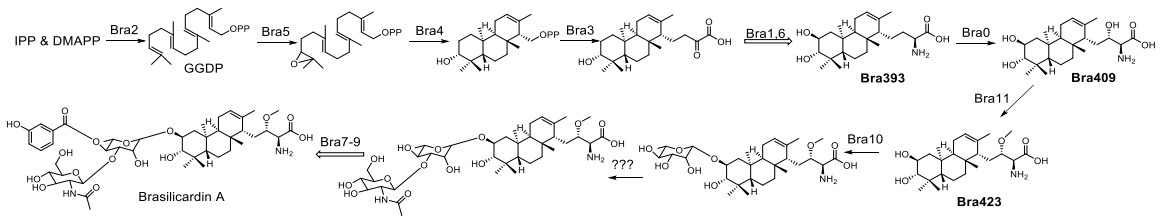


Figure 25: Brasilicardin A proposed biosynthetic pathway

The Brasilicardin A genes Bra0-Bra6 and Bra11 have homologous representatives covered well in each of the three strains, with the best coverage by *S. erythraea*. It was proposed that using a labeling strategy may allow for identification of the downstream products produced by the strains. A number of metabolites were predicted to be purifiable including

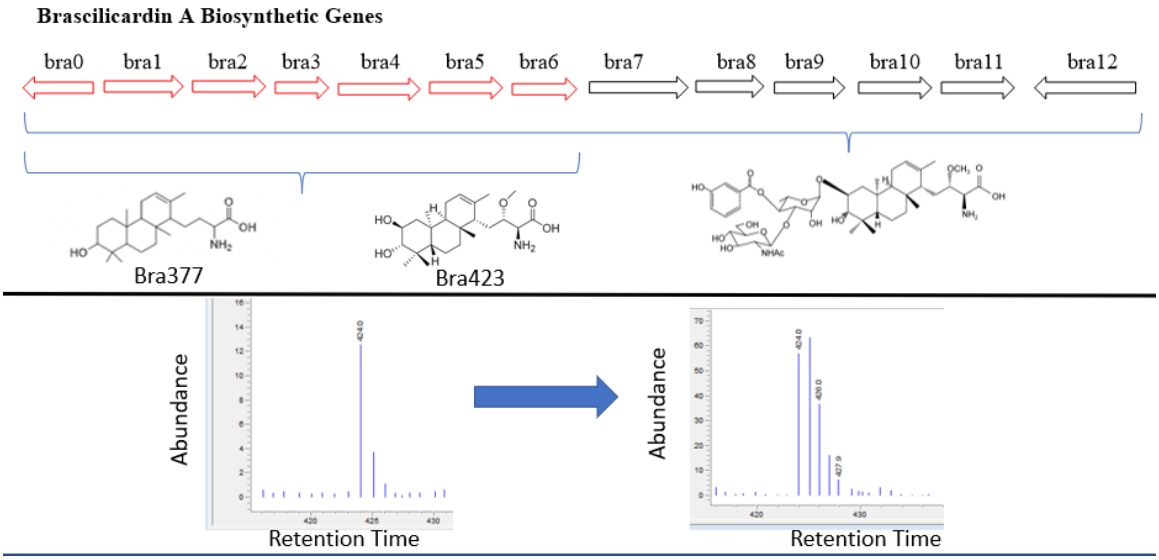


Figure 26: a) Brasilicardin A biosynthetic genes and the various purified intermediates. b) deuterated pattern of one of the purified scaffold intermediates

Bra393 and Bra423. To keep the scaffold as generic as possible to increase likely incorporation by the homologous gene clusters, Bra377 was selected to be partially purified. This molecule corresponds to the Bra393 precursor with the absence of the Bra6 gene, meaning the molecule only incorporates a single hydroxyl group. In order to track the incorporation of this metabolite into the *S. erythraea* manufacturing line, a labeling methodology was required to tag this metabolite. This paired quite well with the universal labeling technique discussed in Chapter 2, with the use of deuterated water. With the

addition of a deuterated tag to the precursor scaffold, the downstream products isotope ratios should change to a distinct signature recognized by the D_2O Sorter program.

A construct containing these genes as well as an MEP and GGPPS pathway was generated. The latter two sets of genes provide upstream material universal to all diterpene molecules, supplementing production of the scaffold[51]. The cassette was expressed heterologously in *Streptomyces albus* R1 using a Golden Gate Assembly and production of a compound with an m/z=378. A similar construct containing Bra0, Bra6, and Bra11 produced a compound with an m/z=424.3. These compounds match the proposed metabolites predicted by the proposed biosynthetic pathway. The *S. albus* R1 strain was cultured in 25% D_2O and the isotope signature changed predictably to the unique deuterated pattern. These scaffolds and examples of deuterium incorporation are illustrated in Figure 26.

The Bra393 precursor was purified and validated through LCMS, as shown in **Error! Reference source not found.**. This purified deuterated scaffold was fed to the *S. e rythraea* strain. *S. erythraea* was cultured in A media overnight and 1 mM of purified Bra393 was added to the culture. This mix was incubated for 24 hrs and extracted using HP20 resin, eluted with methanol, dried, and resuspended in methanol to run on an Agilent 6545 QToF. The files were exported to .abf and run through MS-Dial, a peak picking and comparative metabolomics software. These peak picked files were additionally exported to the D_2O Sorter program to detect deuteration.

By sampling at various time points, consumption and derivatization of the scaffold could be observed in the NRRL2338 strain. D_2O Sorter detected the precursors unique spectra before and after incubation. Excitingly, an additional metabolite was detected after incubation sharing the unique deuterated distribution. This molecule, (M+Na+)=599.40 and referred to as Diter599 is shown in Figure 27. This incorporation strongly implies that this metabolite is a downstream metabolite on the *S. erythraea* analogous Brasilicardin A gene cluster.

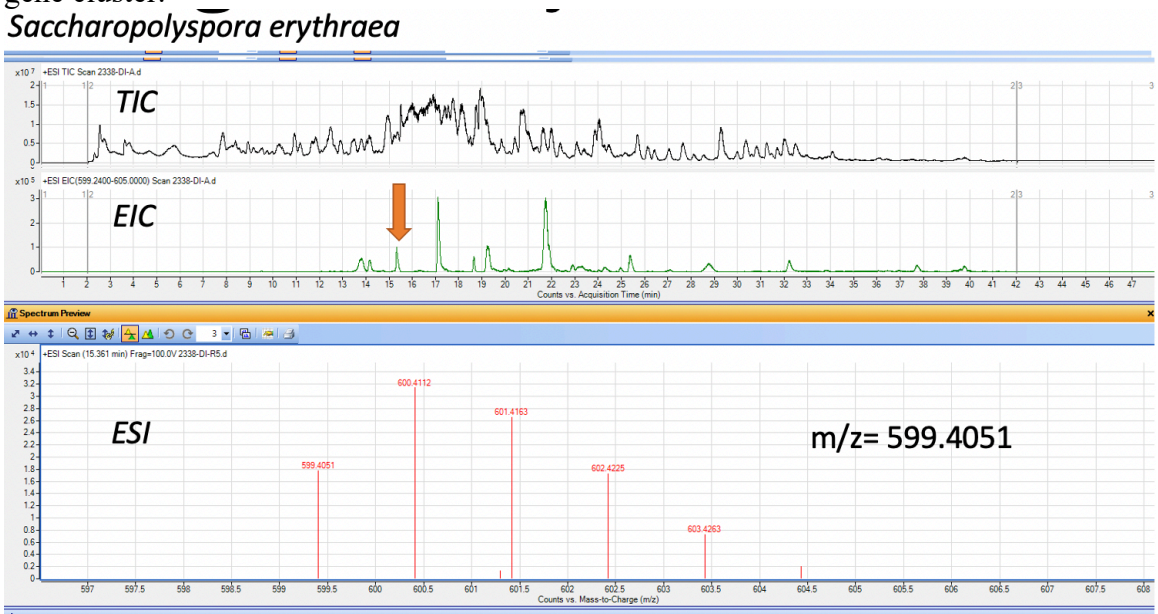


Figure 27: The TIC, EIC, and ESI of Saccharapolyspora erythraea incubated with the Bra393 deuterated scaffold. The EIC and ESI illustrated the specific peak, m/z=599.40, that became deuterated upon incubation.

This compound's peaks significant mass derivatization from the original precursor, differing by approximately 175 Da. This shows that the inputted precursor molecule is dramatically altered from an upstream diterpenoid scaffold, to a complex downstream product. The rapid application and detection from the deuterated platform illustrate the potential to discover novel clinically relevant small molecules using this technique. Deuterated small molecule labeling serves as a potent tool for therapeutic discovery.

3.3 Efforts in Untargeted Applications to Marine Libraries

3.3.1 Introduction

The success utilizing the targeted deuterium scaffold inspired a broader approach. The particular usefulness of deuterium labeling upstream precursors when genetic tools are not available promoted the idea of using the precursor to screen libraries for diterpenoids. Diterpenoids have been discovered from both plants and microorganisms in both terrestrial and marine settings. Marine-derived microbial terpenoids are of particular interest and have been detected from marine flora, fauna, and sediments. Relatively few terpenoids have been isolated from these organisms, but genomics studies indicate that these microbes are expected to have diverse potential in producing these molecules. Of those discovered, many have displayed antimicrobial, anticancer, anti-inflammatory, and antiviral activity[43].

Given the massive untouched biosynthetic potential of these organisms, strategies to increase discovery of these molecules seems prudent. Many organisms represented in libraries of marine microbes are unsequenced and may be challenging to genetically modify. This presents a significant barrier to high throughput analysis of terpenoid discovery in these organisms. It was proposed that using a highly modifiable scaffold for diterpenoids as that established in Chapter 3.2 paired with a deuterium labeling technique, may allow for rapid screening of these strain libraries to isolate novel bioactive metabolites.

3.3.2 Scaffold Purification

The Brasilicardin upstream scaffolds, Bra423, was purified from *Streptomyces albus* R1 modified to contain the necessary gene cluster. This strain was inoculated from a single colony into 5 mLs TSB. After 2 days of growth, the starter culture was inoculated into 10 L of 8% deuterated A-media, for 5 days. This culture was spun down using a centrifuge at 4000 g for 30 minutes. The supernatant was poured off and vacuumed through filter paper. For extraction, XAD4 polymeric resin was added at a 5g/100mL basis. This culture was incubated overnight shaken at 4C. The following day, the XAD4 was collected using filter paper and was extracted in 1 L of methanol for 1 hour. The resulting extract was rotovaped to dry and resuspended in 10 mLs of methanol. This extract was chromatographically separated through an LH20 size exclusion column. The fractions containing Bra423 were dried and resuspended in 1 mL of methanol. This extract was injected onto an RP-HPLC column and the fractions collected. The diterpenoid scaffold is UV silent so fractions were collected and run on the mass spec to detect those containing Bra423. Purity was checked using HR-LCMS. This is illustrated in **Error! Reference s ource not found.**.

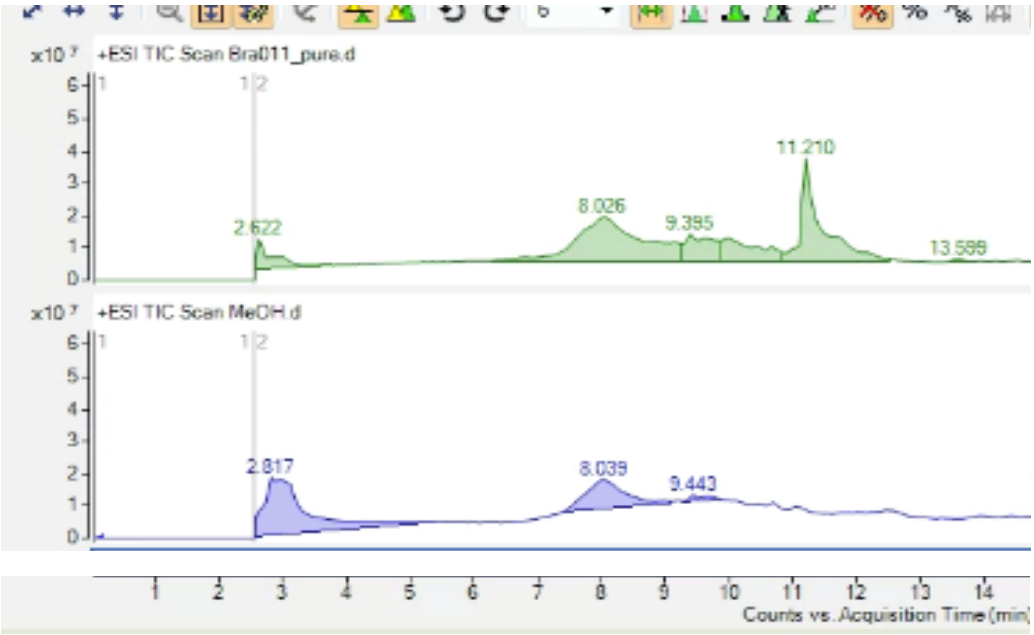


Figure 28: Mass spectrometry graph illustrating the purity of the 424 scaffold. The peak at 11.2 mins is the scaffold.

3.3.3 Marine Library Screening with Deuterated Diterpenoid Scaffold

The marine library bacteria were precultured in 1 mL of marine broth for 4 days at 220 rpm and 30 C. The bacteria were transferred to fresh broth after that time and 2 uL of Bra423 (10 mg/mL) was added to the mixture. The broth was extracted with 4 mL of 95:5 ethyl acetate:methanol at the end of the run and concentrated into 80uL methanol. 5 uL of this sample was injected onto LC-MS. As a control, 2uL of Bra423 was dissolved in 80uL of methanol and run on the LCMS to ensure appropriate sensitivity within the instrument.

Bra423 was purified using 10 Ls of media to generate 1 mg of purified precursor. This purified precursor sent to collaborators at the Hong Kong Institute of Science and Technology who have access to an extensive marine library. The bacteria were cultured as detailed in section 3.4 and complemented with the scaffold. After a sufficient incubation time, the samples were extracted and run on the mass spectrometer. This data was fed to the D_2O Sorter program in order to analyze and detect deuterated compounds.

Bra423 was successfully detected as a deuterated precursor when present in the strains. No additional metabolites exceeding the size of m/z=424.3 were detected during the initial screening, though many smaller metabolites suspected to be Bra423 degradation products were present. Interestingly, Bra423 showed widely variable consumption by strains as shown in Figure 29, leading to interest in what became of the deuterated metabolite. Larger molecules like Diter599 discovered in section 3.2, as well as Brasilicardin A itself, extracts quite poorly in the traditional organic solvent mixture 95:5 ethyl acetate:methanol. It's possible that significant derivative products were not detected in the initial screen because of poor extraction efficiency. Regardless, this trial paves the way for the methodology of using substantially derivatized upstream precursors as a way of detecting late assembly line products.

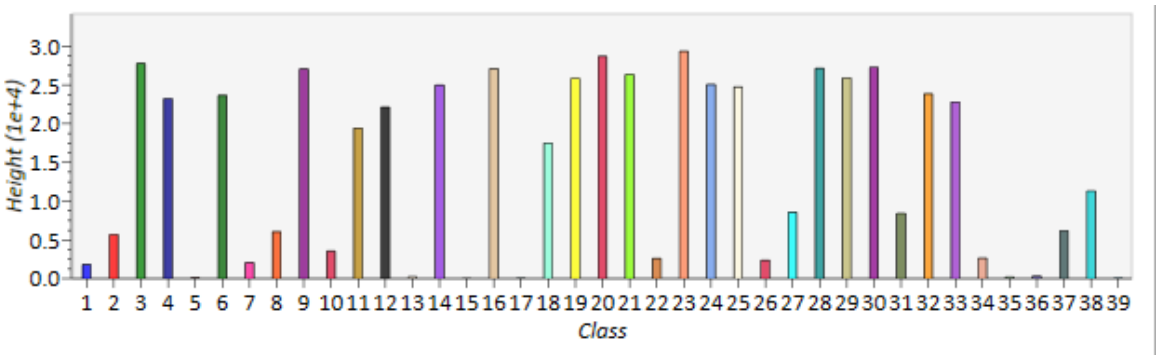


Figure 29: A variety of marine strains were screened with the addition of the deuterated scaffold. Each number represents a different organism and how much of the deuterated scaffold remains after incubation

Chapter 4: Identification and Biological Activity of mutanoclumpin, a *Streptococcus mutans* Secondary Metabolite

4.1 Introduction

4.1.1 The Oral Microbiota

The microbiotas of the human body have become a rapidly growing field of study due to the significant health outcomes associated with these communities. While the fields origins lay with Antony va Leeuwenhoek's observations of "little living animacules prettily moving" in 1674, the oral microbiome has come into importance as a key factor in disease and health[52]. The oral microbiome can play a role in dental caries (cavities), rhematic fever and heart disease, sepsis, and a wide variety of infections[53].

4.1.2 Streptococcus mutans

One known agent of disease of the oral microbiome is comprised of strains of the class *Streptococcus mutans*. These gram positive cocci are facultative anaerobes and tend to colonize on the gums and teeth. When sufficient density is achieved, these lactic acid producing bacteria serve as core etiological agents in the damage to tooth enamel and formation of cavities. Over 2.3 billion people suffer from caries in permanent teeth and the disease costs over $245 billion globally each year in direct treatment costs and productivity losses[54,55].

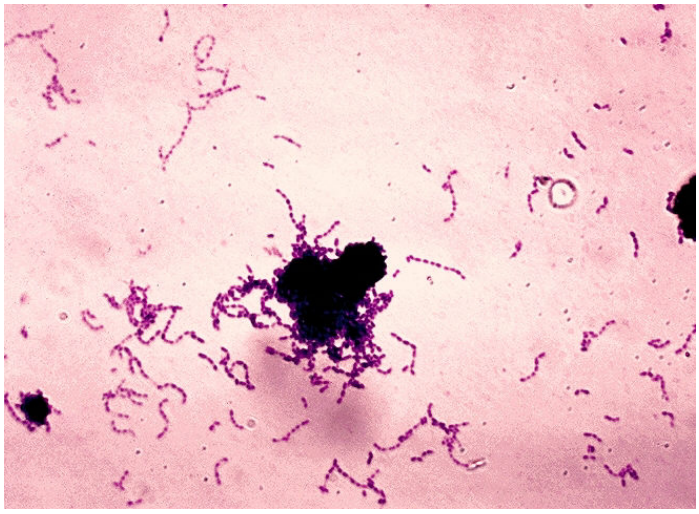
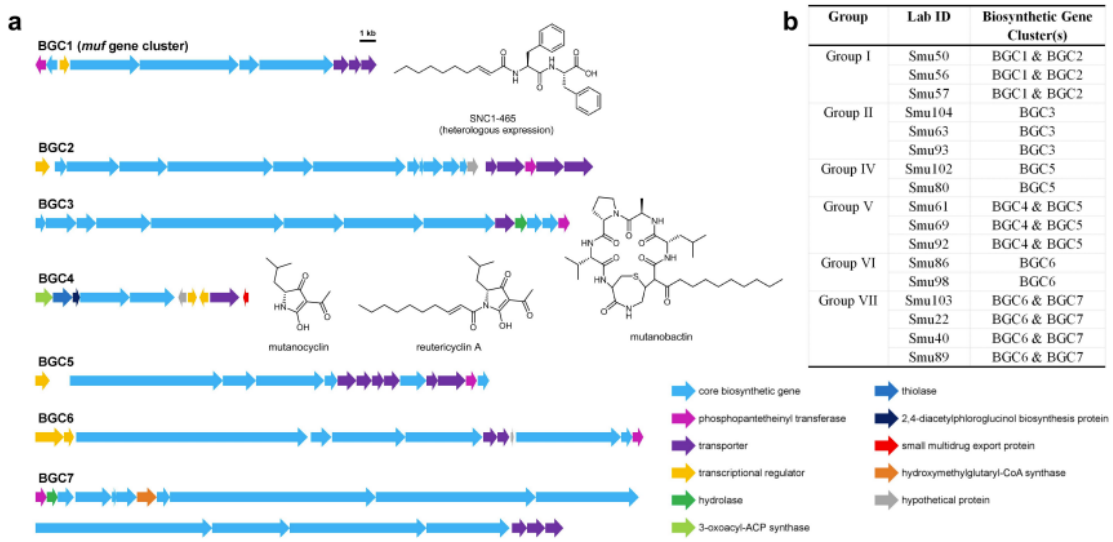


Figure 30: Streptococcus mutans microscopy image from the Centers for Disease Control and Prevention's Public Health Image Library #1043

These bacteria represent a significant threat to human health and utilize a variety of mechanisms to colonize the mouth. One such factor that can contribute to pathogenicity and survival of the strain is the use of secondary metabolite small molecules in order to promote a variety of survival mechanism. These can range from increased oxygen tolerance, acidurity, antimicrobials, quorum sensing molecules, and modifications and improvements to biofilm formation. A list was compiled of reoccurring NRPS/PKS gene clusters that occur in many *Streptococcus mutans* species[6], shown below in Figure 31. These gene clusters have revealed molecules with a variety of functions including antibiotics and compounds that modify cell surface hydrophobicity and biofilm attachment.



b

Group	Lab ID	Biosynthetic Gene Cluster(s)
Group I	Smu50	BGC1 & BGC2
	Smu56	BGC1 & BGC2
	Smu57	BGC1 & BGC2
Group II	Smu104	BGC3
	Smu63	BGC3
	Smu93	BGC3
Group IV	Smu102	BGC5
	Smu80	BGC5
Group V	Smu61	BGC4 & BGC5
	Smu69	BGC4 & BGC5
	Smu92	BGC4 & BGC5
Group VI	Smu86	BGC6
	Smu98	BGC6
Group VII	Smu103	BGC6 & BGC7
	Smu22	BGC6 & BGC7
	Smu40	BGC6 & BGC7
	Smu89	BGC6 & BGC7

Figure 31: a) a variety of biosynthetic gene clusters detected in Streptococcus mutans b) and some strains associated with the clusters. This is adapted from Li et al. 2021.

4.1.3 Biofilm Formation

Biofilms are a particularly pernicious challenge when combating human disease. Besides *S. mutans*, many pathogenic strains are known to use biofilms such as *Pseudomonas aeruginosa* and *Escherichia coli*. Biofilms produced by microorganisms generally confer a variety of favorable conditions to the bacteria. Adhesion to a surface allows the microorganism to establish and sustain a community in a desirable location. The thickness of the biofilm allows for antimicrobial resistance and bacterial persistence due to the inability of many antibiotics and environmental agents to fully penetrate these areas. Lastly though not exhaustively, biofilms allow for high concentrations of microbes to grow near each other which can improve efficiency of the microorganisms in a variety of functions like nutrient acquisition[56]. These biofilms are still poorly understood and are quite diverse, but some have identified small molecules responsible for causing or modifying the biofilm[57]. This chapter identifies a novel small molecule with bioactivity for modifying and promoting biofilm formation in many streptococcal strains through autoaggregation.

4.2 Compound Identification

4.2.1 Gene Knockouts

Strains of Streptococcus mutans S1B were genetically modified to product a BGC5 gene knockout using competence inducing peptides. These peptides have been used historically to induce competence in a variety of *S. mutans* strains[58]. Using this strategy, a core biosynthetic gene in the BGC5 gene cluster was replaced with a spectinomycin resistance marker. With the absence of this gene, production of the downstream BGC5 product is expected to be entirely abolished. This work was performed by Colin Barber of the Zhang Lab.

4.2.2 Comparative Metabolomics

A variety of culturing conditions were tested for the wild type (WT) and BGC5 knockout (KO. Four different BGC5 containing strains along with their KOs were examined. Additionally various culture conditions were tested such as the effects of oxygenation using shaking and standing cultures at 37 C. Two different medias were used: Streptococcal Chemically Defined Media (CDM) and Brain Heart Infusion (BHI). The recipe for CDM can be found in Appendix E. These strains were extracted after 16 hours using 95:5 ethyl acetate:methanol, a common general organic extraction solvent. Comparative metabolomics was performed an Agilent 6545 QToF and RP C18 column. Details regarding the instrumentation can be found in Appendix A. WT and KO data files were uploaded to MS-Dial and compounds of varying concentration identified. Details regarding comparative metabolomic workflows can be found in Appendix B.

A variety of masses were detected varying between the WT and KO strains. An extended discussion and list can be found in Section 4.6 and Table 6. The highest abundance was a molecule with an m/z=584.33, referred to as "584". This molecule was eliminated from the KO strains and showed up in high abundance among all of the WT strains.

Although comparative metabolomics can provide a base list of compounds varying between the strains, it does not provide a definitive answer that the most abundant peak is indeed the expected metabolite. This is partly due to the fact that changes in local gene expression in a strain can cause global shifts in metabolic expression. Even if the gene of interest metabolite has been eliminated, the large abundance metabolite variations observed in comparative metabolomics could be due to side effects in expression. To combat this and in theme with the additional chapters, additional strategies can be employed to add confidence to the pursued metabolite.

The first requisite would be the presence of the metabolite in all BGC5 containing strains. Fortunately, a variety of these strains proved to be genetically modifiable. Below, Figure 32 shows the extracted ion chromatographs (EICs) of a number of BGC5 containing strains and their respective KOs. The deletion of this metabolite from the KOs and the

strong abundance of the metabolite in each strains provides a higher level of confidence that 584 indeed the core product.

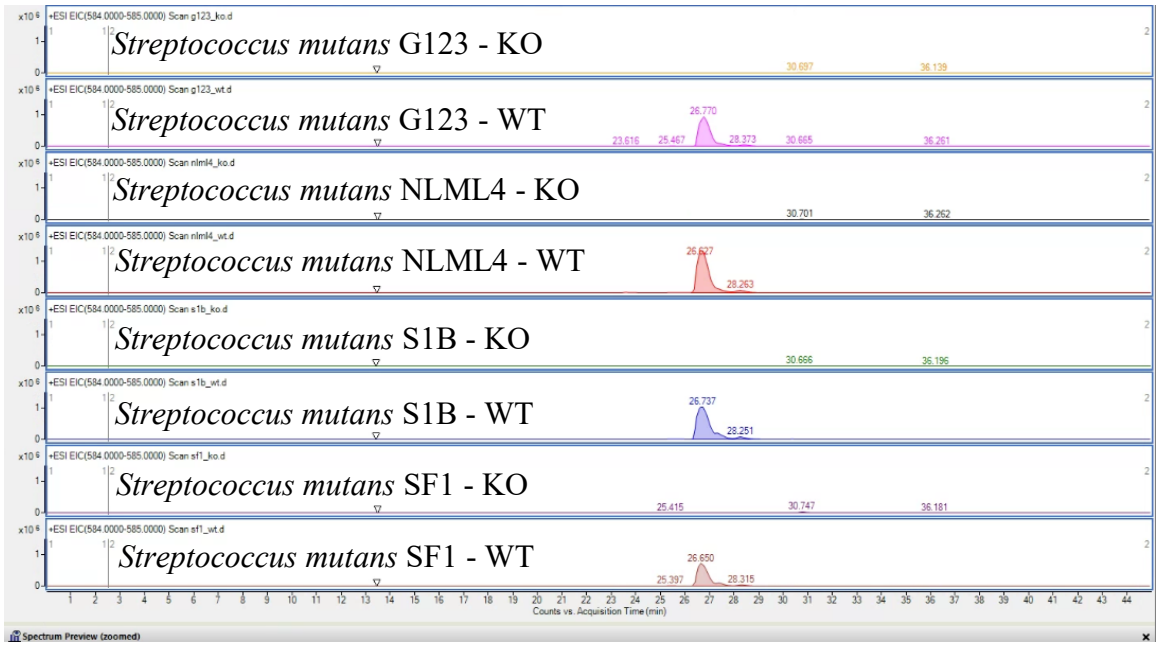


Figure 32: Extracted Ion Chromatograms of m/z=584.30 of BGC5 containing strains and their respective knockouts

Additionally, the BGC5 gene cluster contains an A domain prediction that is expected to load serine as a substrate. This prediction comes from antismash prediction with the gene cluster. If serine is loaded onto this module, it's expected that the final metabolite will contain a derivatized form of serine in its scaffold. To test if 584 included this serine moiety, unmodified serine and serine with N^{15} were fed into two cultures of *Streptococcus mutans* S1B. These samples cultured and extracted as discussed above and the sample run on LCMS. 584, along with a variety of metabolites showed significant incorporation of their isotopic distributions, as illustrated by an elevated ($M+H^{+}+2$) peak. This is illustrated in Figure 33.

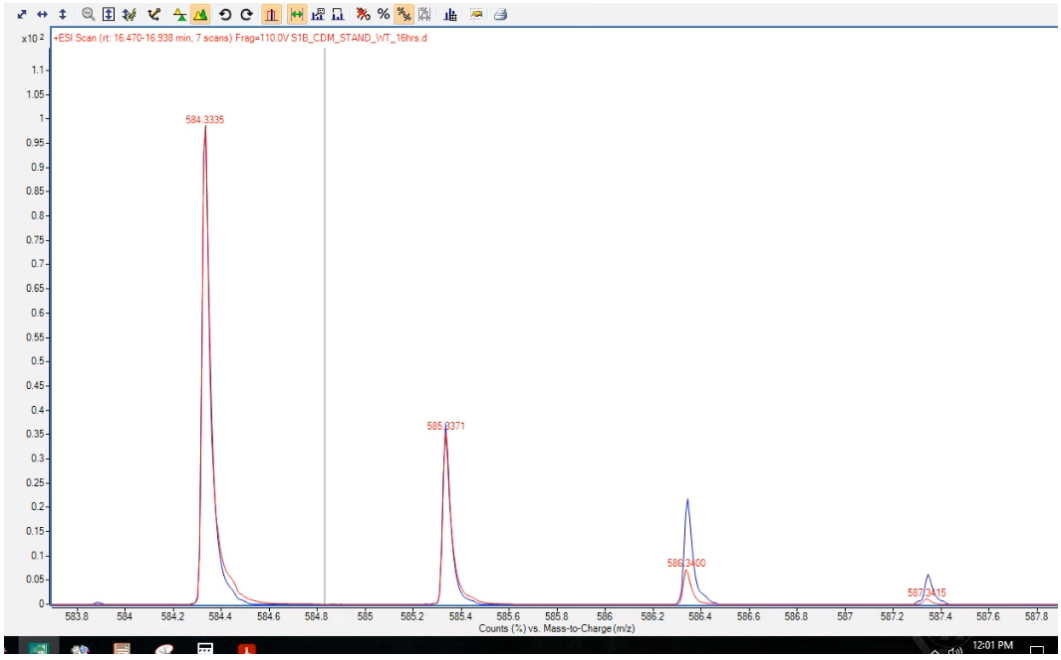


Figure 33: Streptococcus mutans S1B fed unlabeled serin (red) and labeled serine (blue). A significant increase in the (M+2) peak is observed in the labeled compound

4.3 Compound Purification and Structural Elucidation

Reaching purifiable titers of unidentified small molecules sufficient to structurally characterize through NMR can be a challenging prospect. On the production side, biologically relevant titers of the molecule can run in ranges beneath nanograms per microliter. Ensuring that the culture is extracted with the maximum accumulated product is pertinent. On the extraction side, identifying optimal extraction procedures that maximum compound partitioning while minimizing contaminants is desirable. Lastly, establishing downstream purification procedures that separate out what is initially mixture of hundreds of compounds to single molecule is quite a challenge.

4.3.1 Growth Conditions Optimizations

Streptococcus mutans S1B's was monitored using an OD spectrophotometer, while compound production was monitored using an LCMS Single Quad. Plotted below is the growth curve of *Streptococcus mutans* S1B overlayed with the abundance of mutanoclumpin as detected by mass spec. Optimal compound production was observed at late exponential phase, corresponding to approximately 18 hrs of growth. This is shown in Figure 34.

Additionally, various media formulations were tested to observe any changes in 584 abundance. Of particular note, compound abundance increases were observed with the addition of 1 mM serine to the CDM media mixture.

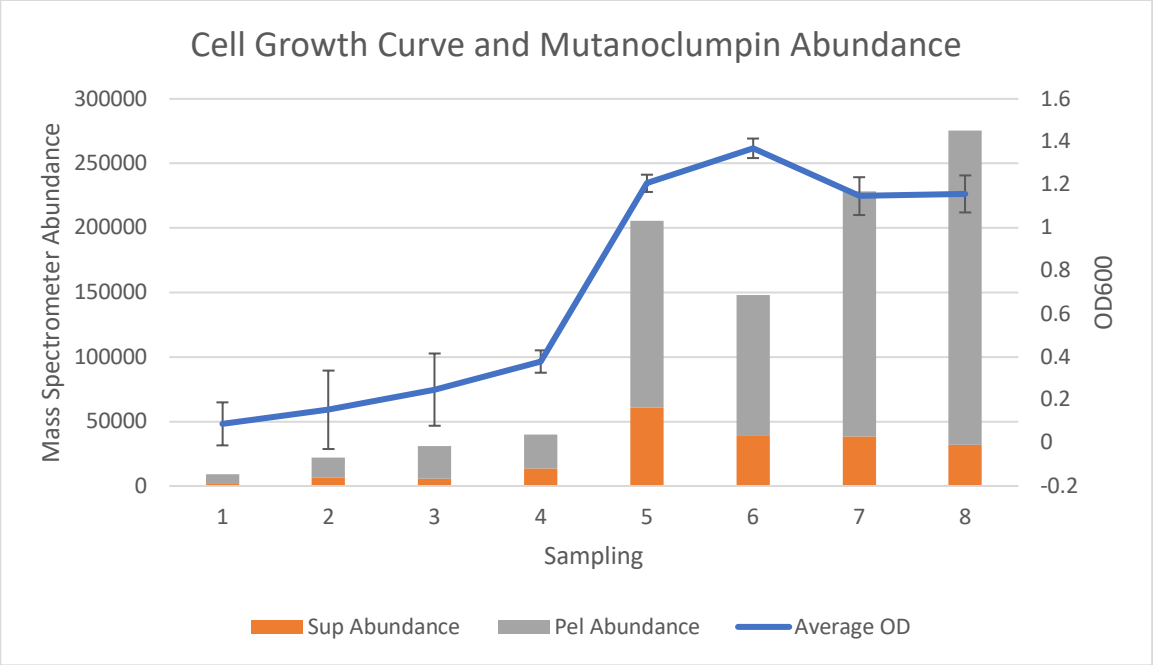


Figure 34: Streptococcus mutans S1B growth curve plotted with compound concentrations of mutanoclumpin. The orange bars show concentration of the molecule in the supernatent, while the gray bar illustrates the pellet.

4.3.2 Large Scale Production and Purification

Various extraction conditions were tested to optimize 584 extraction efficiency. A variety of common organic solvents and a few solid state polymeric resins were tested. While the initial comparative metabolomics work was performed using 95:5 ethyl acetate:methanol, comparable compound abundance was observed using a methanol pellet extraction technique, with significantly fewer contaminating metabolites.

S1B WT was started in 5 mLs of CDM media from a glycerol stock. This starter culture was added to a 50 mL of CDM and allowed to grow for a second night. This culture was used to inoculate a 10 L of CDM media to an OD=0.01 and allowed to grow for 16 hours. At 16 hours the culture was spun down using a floor centrifuge and the pellet was collected. The pellet was resuspended in 1 L of methanol and sonicated. This methanol was spun down in the bucket centrifuge to remove the pellet and rotovaped to dry down.

Upon drying down and concentrating, an HPLC injection solvent of acetonitrile was selected. While methanol is frequently used during these steps, acetonitrile was effective at crashing out viscous contaminants from the extract, but still effectively resuspended 584. This feature allowed for a skip from liquid-liquid extraction straight to HPLC. Frequently, a packed column chromatography step like size exclusion is required to concentrate the extract down effectively, as was discussed in Chapter 3.

This concentrated extract was resuspended in 1 mL of acetonitrile and all insoluble components spun down and removed. This compound was injected in 200uL aliquots onto an Agilent HPLC, discussed in Appendix A. The purified fraction containing the two predominant isomers was collected and dried down for NMR.

4.3.4 Structural Elucidation

This compound was resuspended in D3 Chloroform and NMR was performed to identify the structure of the molecule. Upon identification of the structure, the compound will henceforth be called "mutanoclumpin". There are two isomers of mutanoclumpin with the deviation appearing at the tertiary carbon on the serine-derived moiety. This compound has two cyclizations, a macrocyclization attached to a C10 fatty acyl chain, and an inner cyclization involving the serine moiety. There is additionally incorporation of a phenylalanine, and a glutamate moiety modified with an O-methyl group.

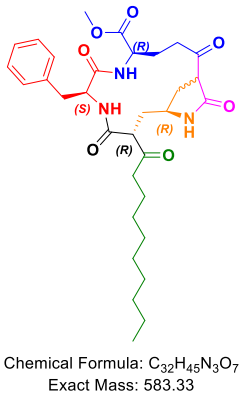


Figure 35: Mutanoclumpin Structure

4.4 Bioactivity Elucidation

4.4.1 Antimicrobial Activity

One of the most common and well-studied modes of small molecule bioactivity is that of antimicrobial defense. The discovery of antibiotics in the 1920s has since provided a cornucopia of medicinal treatments for human beings and is frequently dubbed "The Antibiotic Revolution"[59].

Five strains were tested: *Bacillus subtills* (a gram positive, grown in LB), *Pseudomonas aeruginosa* (a gram negative, grown in LB), *Mycobacterium smegmatis* (a mycobacterium, grown in 7H9), *Saccharomyces cerevisiae* (a yeast, grown in YPD), and *Streptococcus salivarius* (a fraternal strain and co-oral inhibitor, grown in BHI). These strains cover a broad swath of microbial diversity. Concentrations up to 1 mM mutanoclumpin were tested using a 96 well plate format. Cells were diluted from an OD=1.0 at 600nm by 20,000x following an overnight culture. Antimicrobial activity would be visible through decreases or absence of visible opacity in the test cultures. Cultures were examined after 48 hrs of growth. The plate is displayed in Figure 36.

Figure 36: Antimicrobial bioactivity of mutanoclumpin paired against Bacillus subtilis, Pseudomonas Aeruginosa, Saccharomyces cerevisiae, Mycobacterium smegmatis, and Streptococcus salivarius

The strains tested showed no visible growth deviations at concentrations up to 1 mM. This excitingly implied that the mode of activity provided by this molecule may come from less well studied bioactive mechanisms.

4.4.2 Oxidative Stress

Facultative anaerobes have variable preferences for dissolved oxygen concentrations. While oxygen can serve as a high energy electron recipient in the respiratory cascade, at high concentrations oxidative stress can result in damage to microbial DNA. Bacteria have a variety of mechanisms for dealing with oxidative stress: one such example is the use of secondary metabolites, either for signaling purposes or to absorb oxidative damage. To inspect responses to oxidative stress, H_2O_2 can be used as a proxy for dissolved oxygen radicals. This data is displayed in Figure 37.

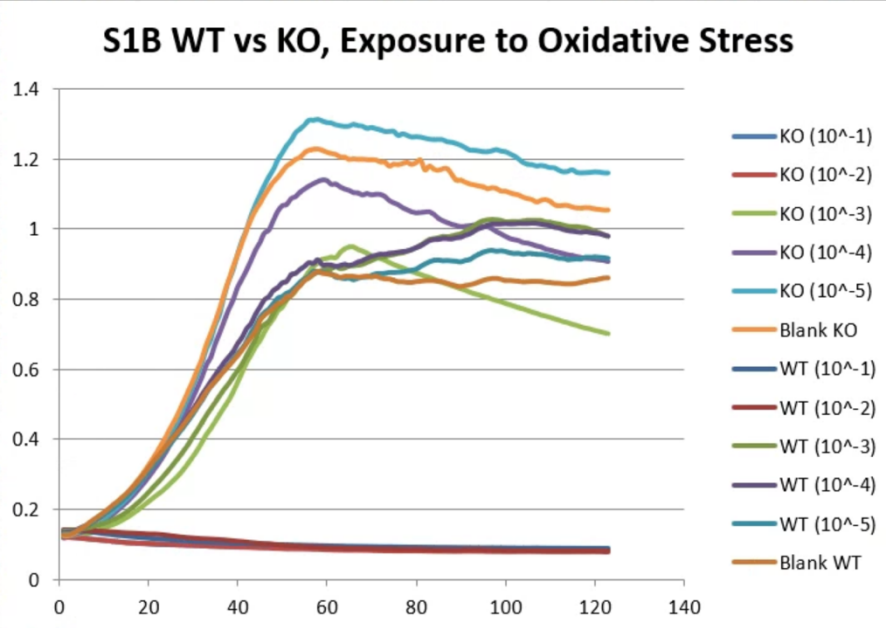


Figure 37: Growth curves using an OD spectrophotometer of Streptococcus mutans S1B wild Type and knockout under various conditions of oxidative stress. (10^-1) corresponds to a hydrogen peroxide concentration of 0.1% with each following sample being an order of magnitude lower in H_2O_2 concentration

WT and KO strains were inoculated at OD=0.01 in CDM media onto a 96 well plate read by a spectrophotometer. Concentrations of H_2O_2 ranging from 0%-0.1% were added to the respective cultures.

No significant deviation in growth conditions was observed.

4.4.3 Acidurity

S. mutans are lactic acid producing bacteria. This trait contributes to the synthesis of dental caries and also provides a hostile environment for competing microbes[11]. In order to survive these conditions, these microbes have a variety of mechanisms such as the use of proton pumps to modify local cellular pH. Regulation of these mechanisms can be controlled by signaling molecules or small molecules can directly impact solution pH. Examining the ability for strains to survive briefly under highly acidic conditions can provide insight into the tools used by the strain.

An acid challenge assay was utilized to examine this effect. In brief, *Streptococcus mutans* S1B WT and KO were exposed to CDM adjusted to a low pH. The cells are incubated and sampled over an hour in order to detect survival of the cells in the acidic solution[60]. The results are displayed in Figure 38.

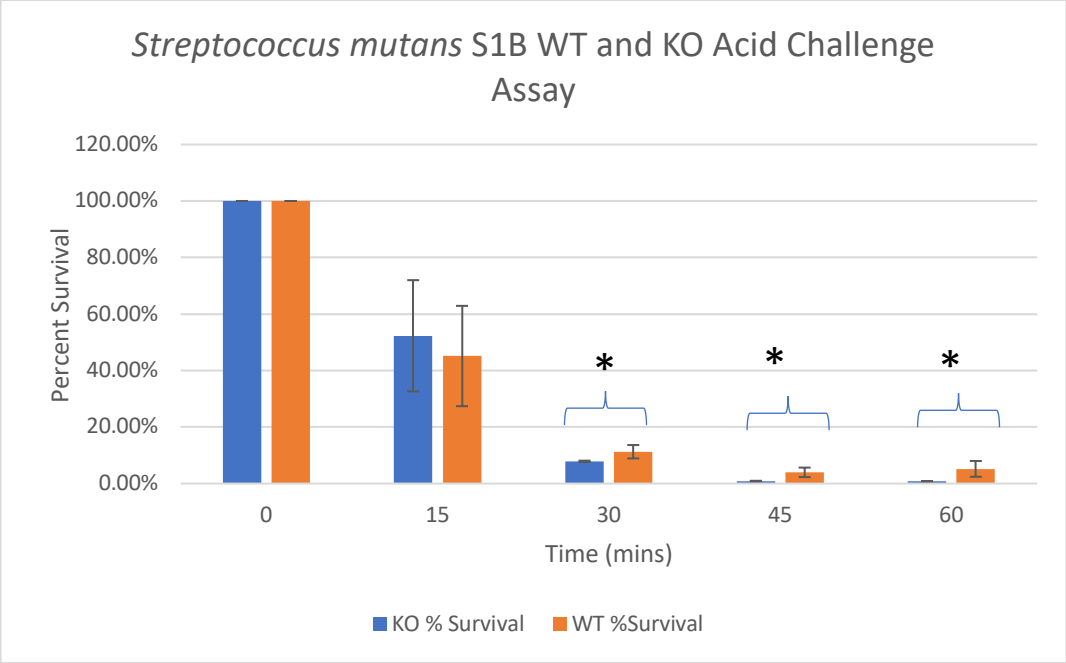


Figure 38: Acid challenge assay using Streptococcus mutans S1B WT and KO. (*) indicates a statistical confidence of α <0.05.

Interestingly, statistically significant deviations in acid survivorship were observed at the 30-60 minute range during the assay. If the presence of the BGC5 gene cluster is responsible for the effect, this could be due to a variety of factors. The metabolite produced by the strain could be directly modifying the local pH or playing signaling role in activating pH adjusting mechanisms. This could also be due to a secondary effect from the metabolite, such as biofilm formation inhibiting acid penetration.

4.4.4 Metal Chelation

Another mode of small molecule assistance comes in the form of metal chelation. Many transition metals are required for enzymes to function properly. Cells can generate small molecules that can chelate these metals and be uptaken into the cell.

An indirect examination of this chelation came through the supplementation of common metals to the WT and KO strains. Metals were supplemented at 100 uM concentrations to examine any changes in growth. Deviations in growth rates between the WT and KO strains might indicate improved acquisition of the metals by the WT strain. This is illustrated in Figure 39.

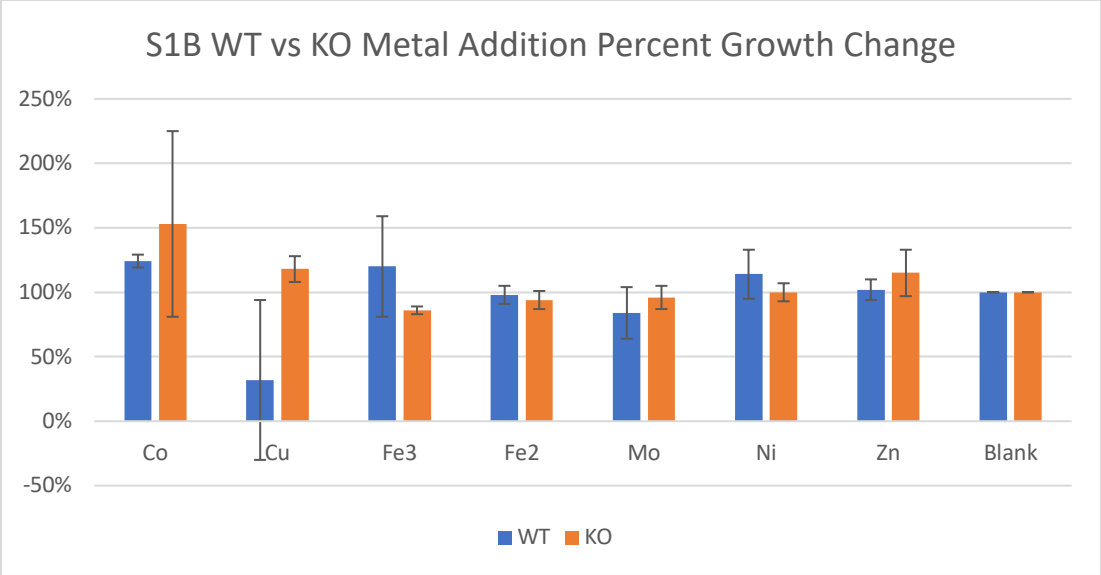


Figure 39: Comparisons of relative end point growth of Streptococcus mutans S1B Wild Type vs. Knockout when supplemented with various metals

While a high degree of variability was observed in some of the more cytotoxic metals like copper and cobalt, no significant change favoring the wild type strain's production of mutanoclumpin was observed. This deterred further investigation of metal chelating activity.

4.4.5 Antimicrobial Resistance

Small molecules can also participate directly and indirectly with antimicrobial resistance. Similar to oxidative stress, molecules can serve as part of a signaling pathway, or participate directly in reactive chemistry to generate microbial antibiotic resistance.

A variety of antibiotics for which *S.* mutans is susceptible were tested. These include erythromycin, tetracycline, and ampicillin. Each of these compounds was dosed at their working concentrations of 50 ug/mL, 100 ug/mL, and 50 ug/mL respectively as a control. Then samples were given 4x dilutions serially of each of the antibiotics until growth was observed. Each of the antibiotics has a different mode of action: erythromycin binds to the 23S ribosomal RNA molecule in the 50S subunit of the bacterial ribosome, tetracycline inhibits protein synthesis by preventing the attachment of aminoacycl-tRNA to the ribosome, and ampicillin interferes with cell wall synthesis. The data for this assay is shown in Figure 40 with the relative end point OD600 of each sample.

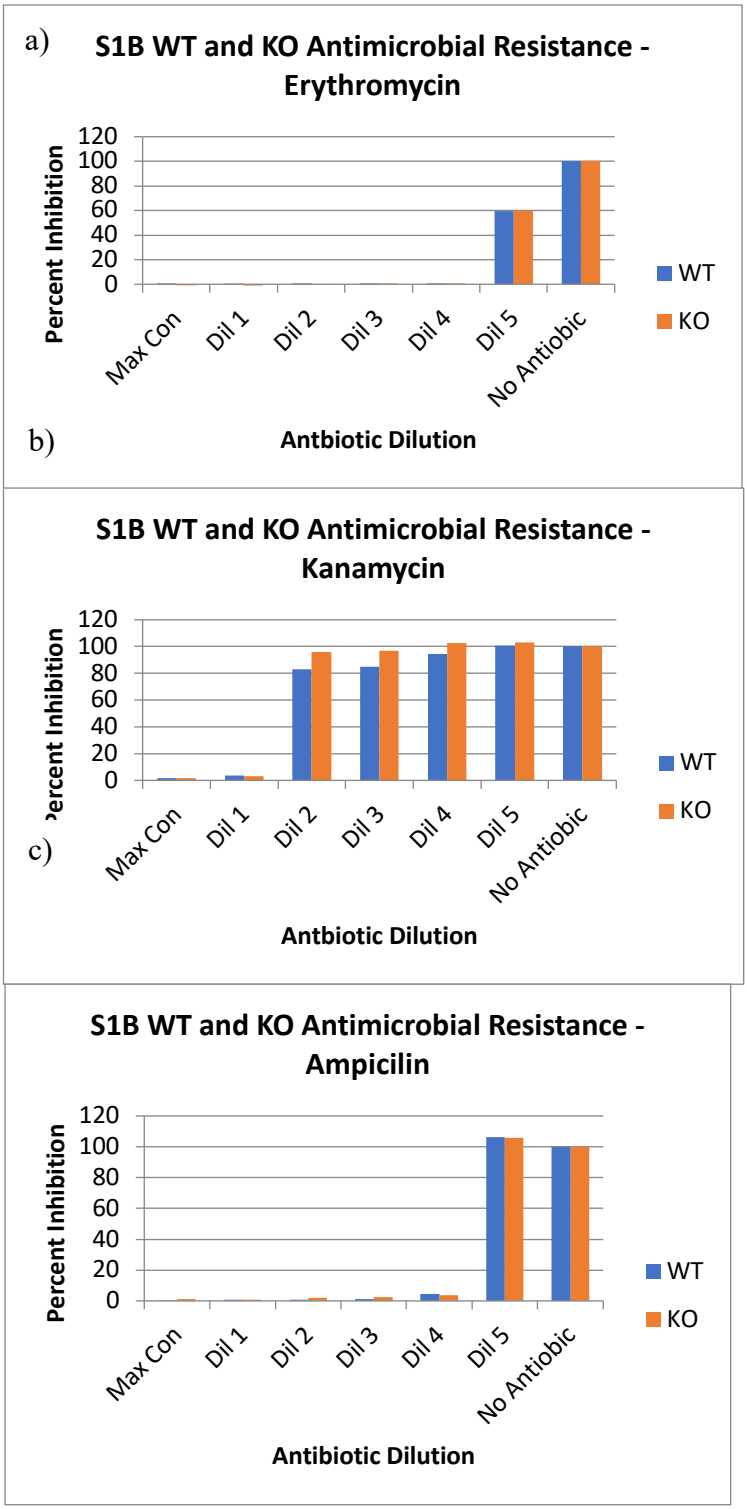


Figure 40: Sublethal Antimicrobial resistance assays using a) erythromycin, b) kanamycin, and c) ampicillin.

No significant inhibition variations were observed for the three different antibiotics. This indicated that the BGC5 metabolite does not appear to play a role in antimicrobial resistance.

4.4.6 Invasion Assay

Since these strains colonize the human mouth and teeth, advantages that confer adhesion to human cells may prove to be quite viable strategies for survival. Utilizing an invasion assay with HeLa and Caco2, cervical cancer and colon cancer cells respectively, WT vs KO strains were compared for adhesion.

In brief, bacterial cells were grown to exponential phase and incubated with human cancer cells for two hours. After two hours, cancer cells were washed in triplicate with PBS. The cells were resuspended in surfactant to inactivate the human cells. This was plated onto BHI plates and cell counts were taken.

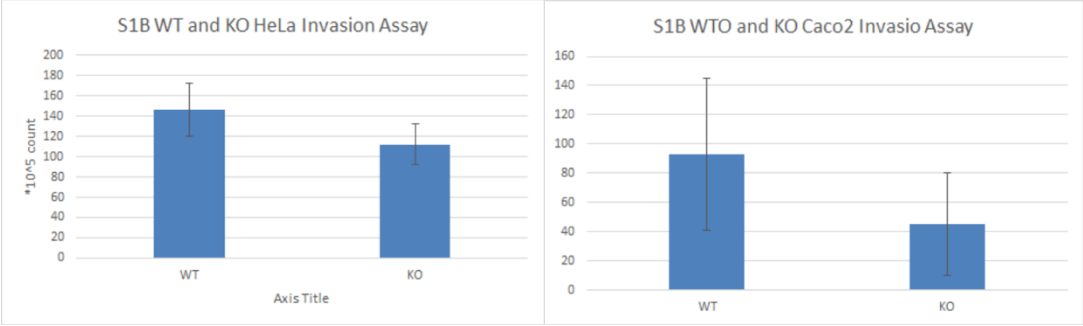


Figure 41: Human cancer cell invasion assay of the wild type and mutant Streptococcus mutans S1B strains using a) HeLa and b) Caco2 cell lines. Assay was performed in triplicate.

No statistically significant differences were observed between the two strains with each providing similar bacterial cell counts.

4.4.7 Carbohydrate Usage

Given the variety of carbon substrates that exist within the human mouth, testing growth on a variety of sugars can provide insight into if mutanoclumpin modifies sugar metabolism.

CDM media was modified with a number of different sugars. The WT and KO strains were grown in each of these conditions; end point OD and production of 584 were observed. Most interestingly, the strains were observed and photographed to show any growth deviations. Figure 42 illustrates pictures of each strain.

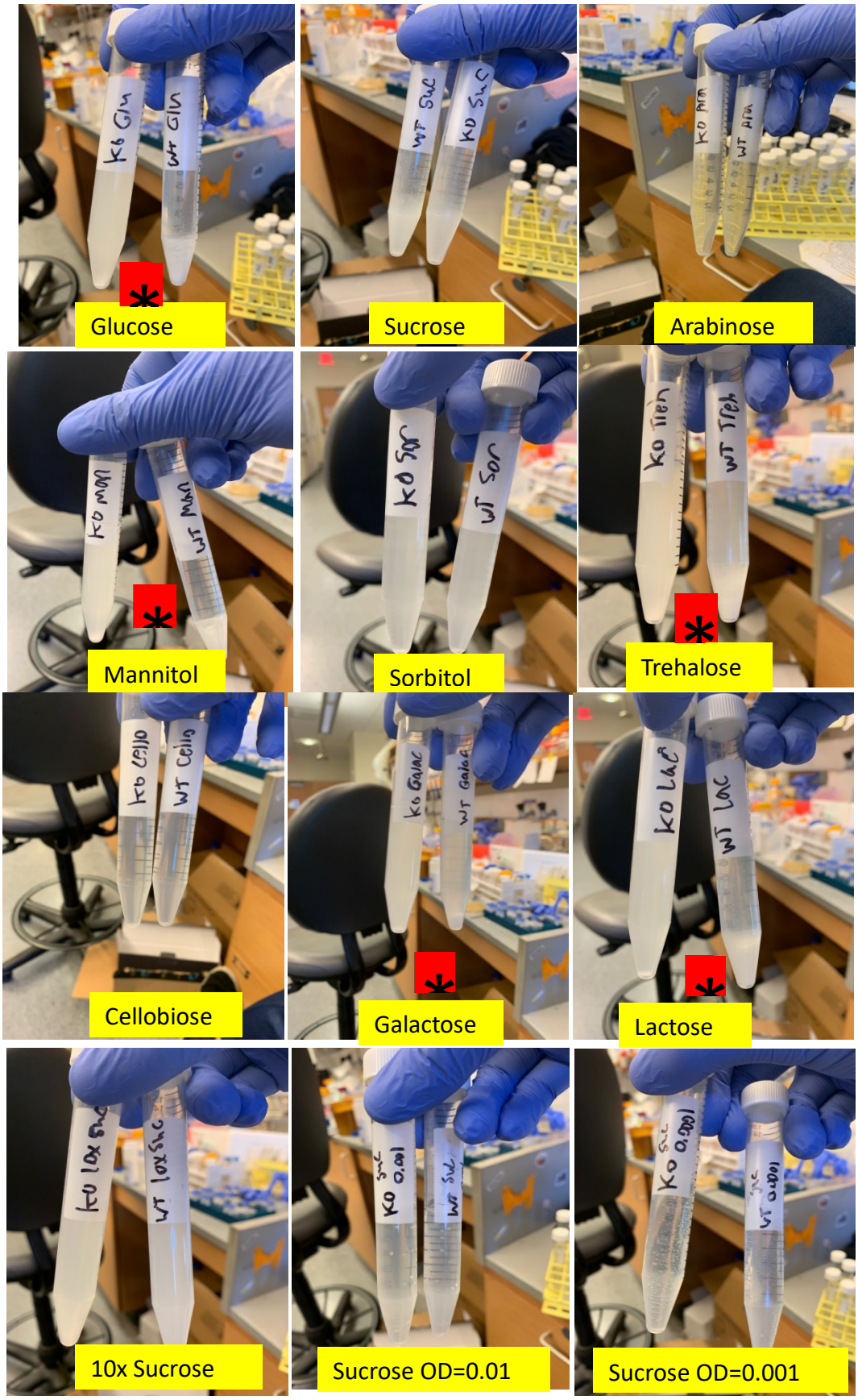


Figure 42: Streptococcus mutans S1B Wild Type and Knockout incubated with a variety of sugars in CDM media. The knockout is show on the left and the wild type on the right. The red box and white star indicate a visually significant difference between growth of the two strains

Table 3: Various end point ODs of Streptococcus mutans S1B wild type and mutant grown in a variety of different sugars.

Sugar	OD WT	OD KO

Glucose	2.69	3.18
Sucrose	1.01	1.11
Arabinose	0.24	0.24
Cellulobiose	0.27	0.31
Galactose	3.07	2.84
Lactose	2.91	1.83
Mannitol	2.51	2.70
Sorbitol	1.34	1.10
Trehalose	2.91	2.93
10 x sucrose	2.47	1.85

Some significant variations were observed from the sugar incorporation assay. There were wide differences between end point ODs between a number of the carbohydrates. Very significantly, the cultures were observed to flocculate variably depending on the sugar used and irrespective of the end point OD. This provides interesting insight into mutanoclumpin's role in clumping of the cell cultures. The BGC5 gene cluster is clearly playing a significant role with the autoaggregation of the strain.

4.4.8 Biofilm Formation and Autoaggregation

Given the results of the carbohydrate usage assay, there was a strong indication that the BGC5 metabolite may play a role in biofilm synthesis.

Submerged biofilm formation assays (SBFA) were performed to test biofilm capacity of the strain. In short, the strain inoculated at OD=0.01 into a 2 mL 12 well plate samples of CDM + 5% BHI. The samples were allowed to grow over 16 hours. After 16 hours, the plates were washed and stained using the standard crystal violet staining protocol[61].

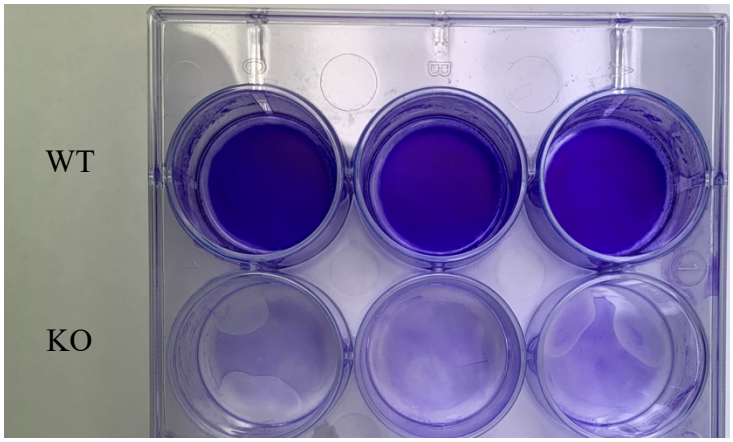


Figure 43: Crystal Violet Biofilm Staining Assay on Streptococcus mutans S1B wild type and mutant grown in CDM media in triplicate.

The knockout strain illustrated abrogation of biofilm retention to the surface. While a slight biofilm appears on the plate surface of the knockout, it is easily sloughed during the washing steps. The WT on the other hand forms a durable biofilm that resists washing. This indicates that BGC5 plays a vital role in biofilm formation and maintenance.

Chemical complementation was performed to see if recovery of the BGC5 phenotype was possible. Mutanoclumpin was added to vials at a variety of time points and

autoaggregation data was taken. Excitingly, addition of the mutanoclumpin small molecule to the knockout shows strong recovery of the autoaggregatory phenotype. When added to plates, visible autoaggregation is observed in the knockout. Currently complementation with mutanoclumpin hasn't conclusively recreated biofilm formation: this is expected to be due to the complexities of both perturbing the biofilm during molecule addition and delicate timing factors not yet well understood. Fortunately, the aggregation phenotype associated with the strains is quite reproducible as shown in Figure 46.

A variety of sugars were additionally tested to determine the presence or the effect with different carbons sources.

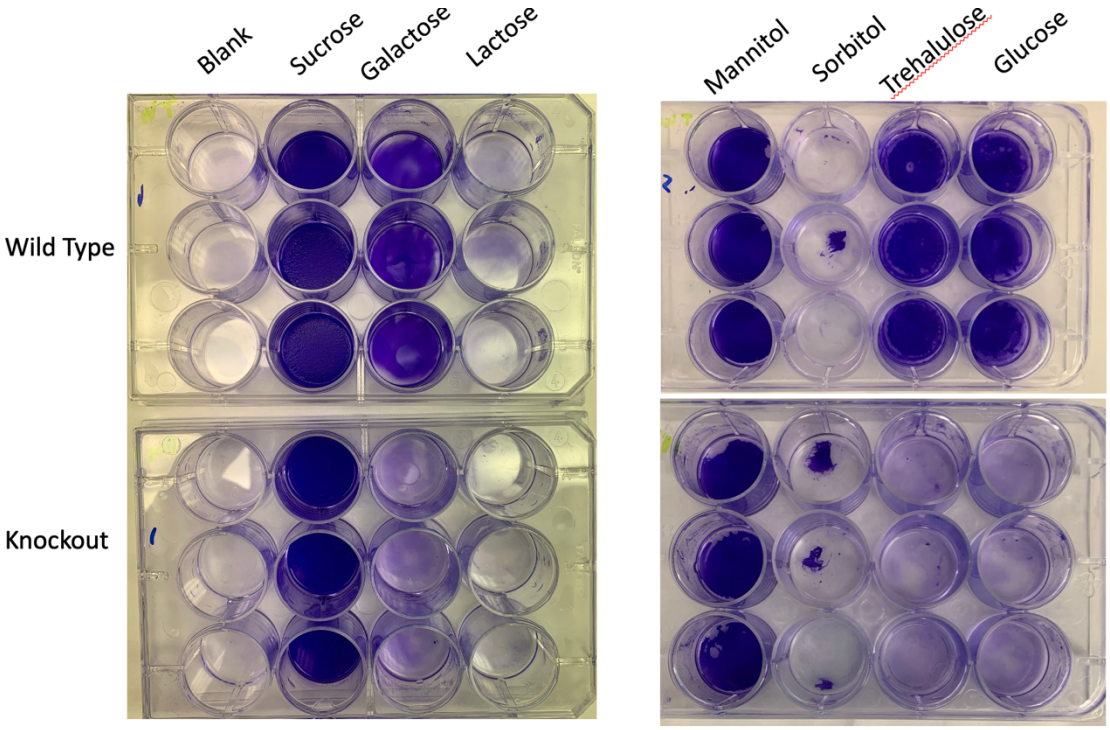


Figure 44: Crystal Violet Stainiing Assay of Streptococcus mutans S1B Wild Type and Knockout cultured in CDM media with a variety of carbon sources

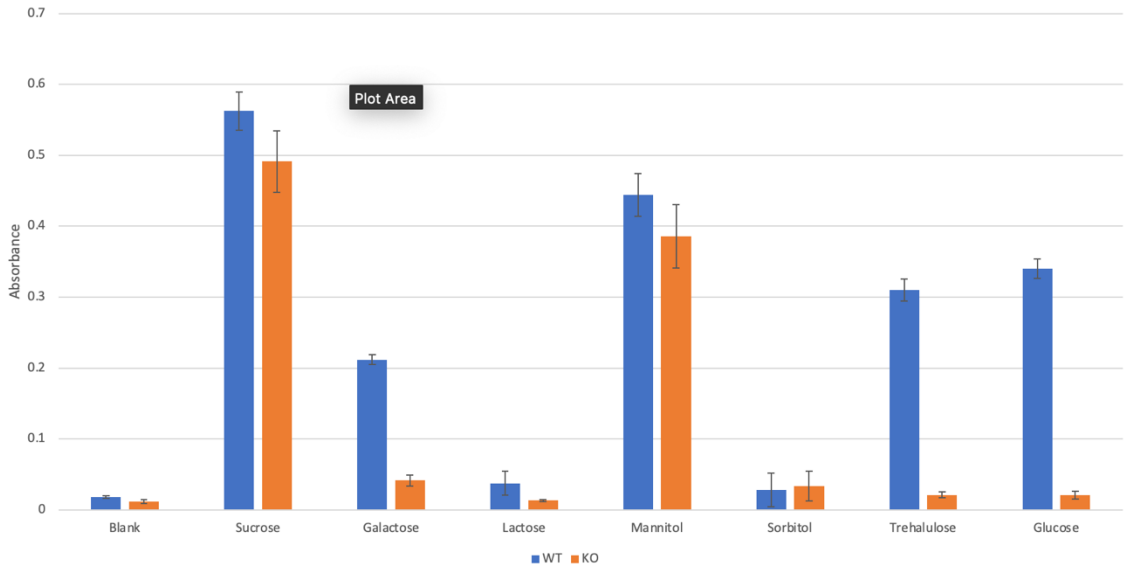


Figure 45: Crystal Violet Staining assay results quantified using an OD Spectrophotometer of Streptococcus mutans S1B Wild Type and Knockout cultured in a variety of carbon sources

Results showed meaningful variations in biofilm formation particularly pronounced in medias containing glucose, trehalose, and galactose. Medias with sucrose, typically associated with biofilm formation, did not show statistically significant variations in biofilm formation. This could imply that in the presence of sucrose, the effect of mutanoclumpin is negligible or not necessary to product adhesion.

Because of the high fidelity of aggregation, future work involving complementation was screened using an aggregation assay. This assay observes the OD of the settled culture supernatant, relative to the culture when disturbed. Figure 46 shows the results of this assay with the wild type strain and the knockout that disrupts production of mutanoclumpin. These culturing were performed in CDM with glucose due to the strong phenotype observed in the variable sugar assay. It also shows a knockout that's production is expected to prevent O-methylation of mutanoclumpin, evidenced by mass spectrometry data.

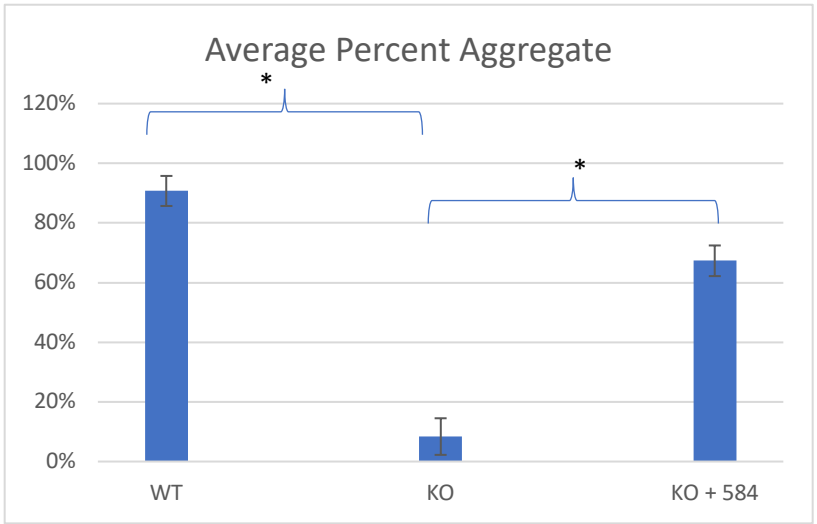


Figure 46: Autoaggregation assay of Streptococcus S1B Wild Type, PKS Knockout, and a Gene Knockout suspected to prevent O-methylation of mutanoclumpin

The last complementation performed observed the usage of mutanoclumpin for non-BGC5 containing strains of bacteria: both Streptococcal and otherwise. Figure 47 and 49 shows the addition of mutanoclumpin to *Streptococcus mutans* U2A and *Bacillus subtilus* and their respective autoaggregation. No statistically significant autoaggregation was observed, implying that mutanoclumpin is specific to the BGC5 containing strains.

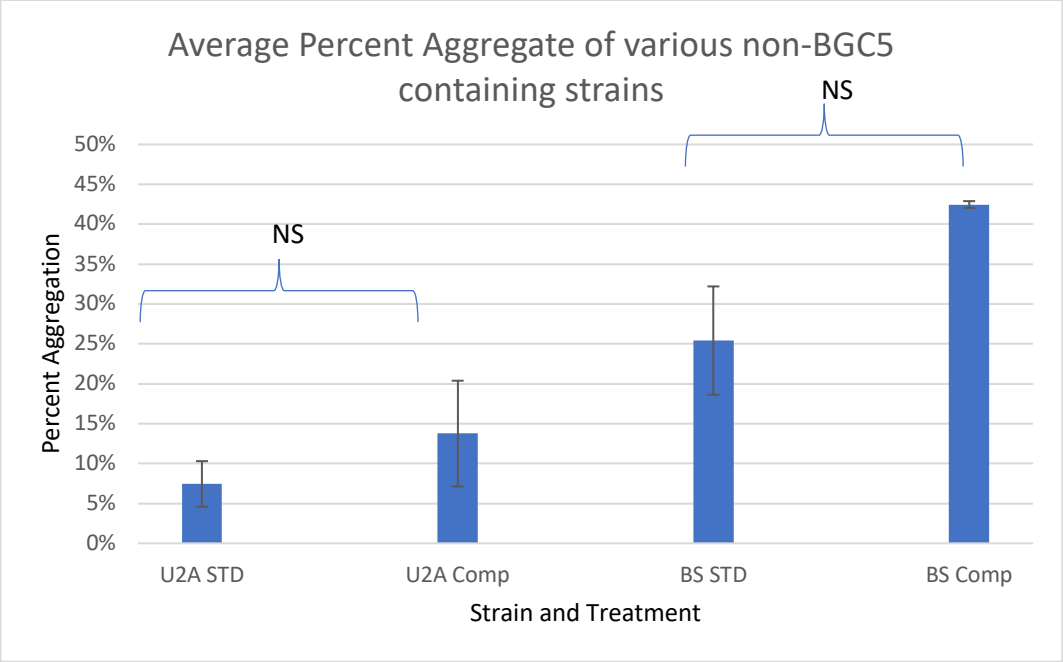


Figure 47: Effects of mutanoclumpin on a variety of non-BGC5 containing strains including Streptococcus mutans U2A and Bacillus subtilus. Neither strain showed significant increases in autoaggregation.

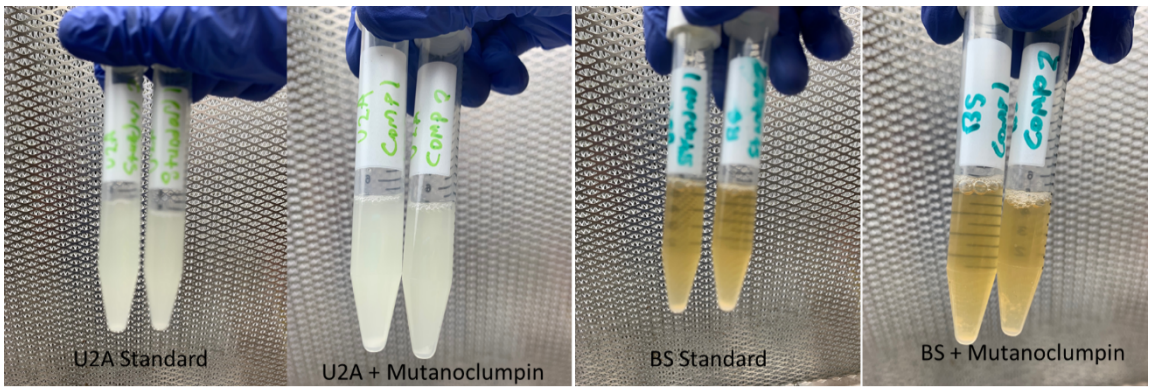


Figure 48: Pictures of various non-BGC5 containing strains including Streptococcus mutans U2A and Bacillus subtilus cultured with mutanoclumpin. Neither showed significant increases in autoaggregation

4.4.9 Compound Binding Affinity

Given the interesting bioactivity associated with the small molecule, especially in the compound's association with the cell pellet, preliminary characterizations of the compounds kinetic and thermodynamic equilibrium with its binding site were tested.

Firstly, a wash assay was performed to observe if the rate at which the compound could be removed from the cell pellet. This is shown in Figure 50. Interestingly, minimal removal of the compound was observed over a series of six washes at a 1:1 wash:cell culture equivalent volume. This implies very tight binding of the small molecule to its target site.

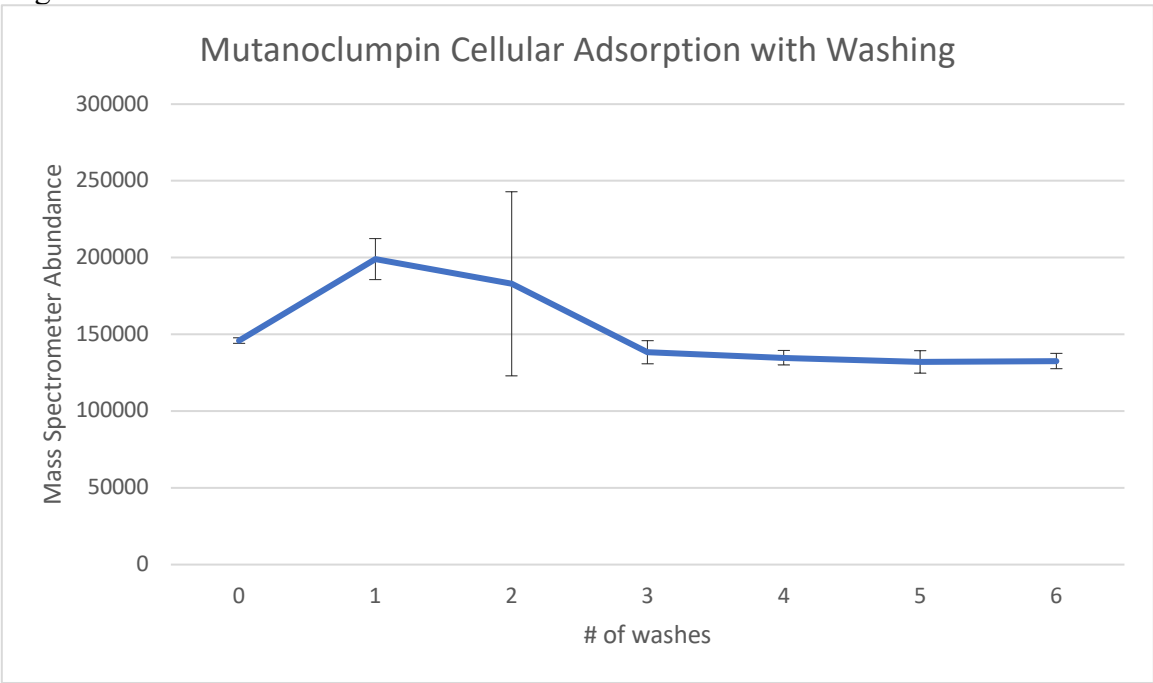


Figure 49: Binding affinity of mutanoclumpin to the cell pellet as observed through repeated washes of the cell pellet with PBS.

A follow-up to this assay was performed to test the kinetics of this binding. mutanoclumpin was added to the knockout strain and samples were tested for mutanoclumpin adhesion to the cell over a period ranging from one minute to two hours. Interestingly, the compound finds its binding site within one minute and see almost full adsorption by that time. This is illustrated in Figure 51.

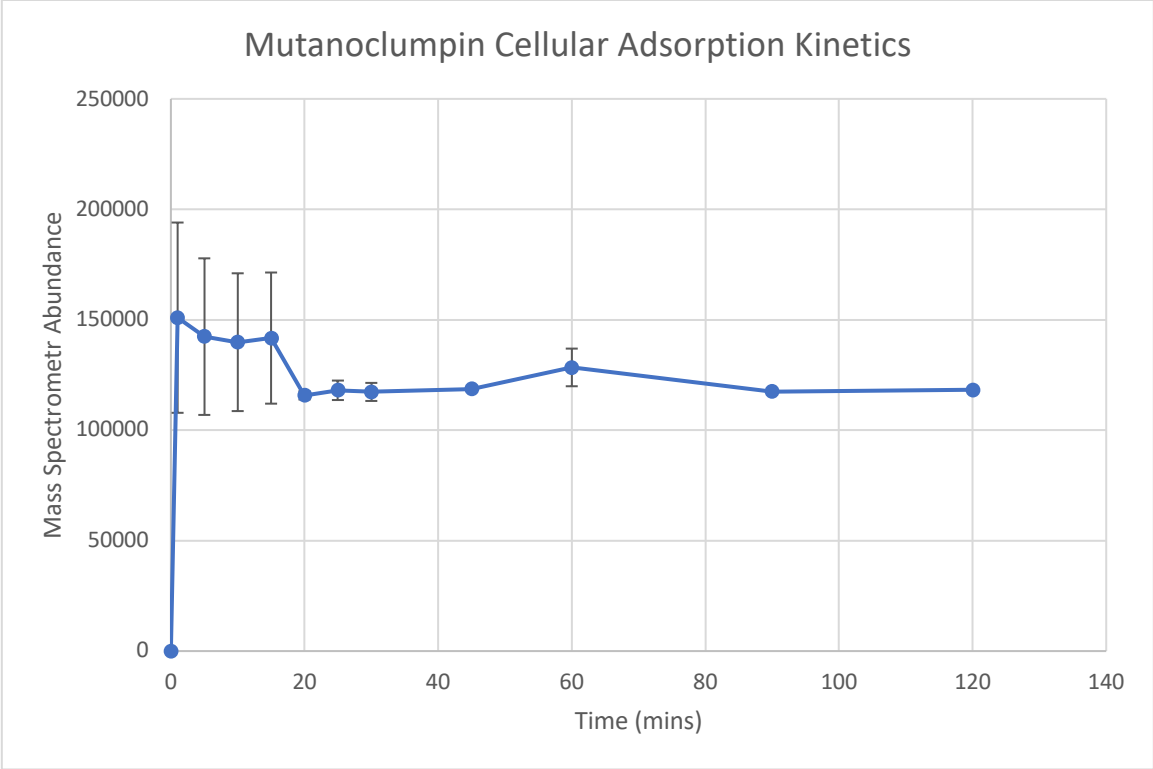


Figure 50; Association of mutanoclumpin with Streptococcus S1B cell pellet over the period of two hours. Full adsorption of the compound is observed within a minute of addition.

4.5 Purification of Intermediate Metabolites and Structural Elucidation by Mass Spectrometry

4.5.1 BGC5 Biosynthesis

Establishing the biosynthetic pathway of the molecule can help discover novel enzyme functionality and provide insight into the mechanisms by which the molecule is synthesized. BGC5 has a variety of biosynthetic genes responsible for the synthesis of the molecule. A proposed scheme is detailed in Figure 52. It is proposed that BGC5 biosynthesis starts off with the loading of a C_{10} fatty acid followed by an extension with malonyl CoA. Next a phenylalanine is loaded, followed by a glutamate. Next the chain is extended again with malonyl CoA, and then a serine moiety is loaded. Next a small ring cyclization occurs with the serine hydroxyl. It is currently unclear if the macrocyclization happens spontaneously or which enzymes might be responsible. The methyltransferase appears to act non-specifically on a number of the upstream products.

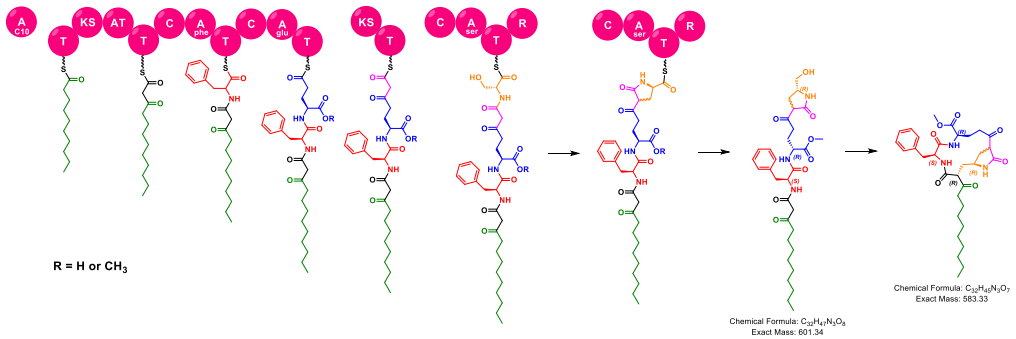


Figure 51: Proposed biosynthetic pathway of mutanoclumpin

4.5.2 Intermediate Purifications

To confirm or refute the proposed biosynthesis, a variety of purifications were performed to identify other key intermediate metabolites. Table 4 lists a variety of metabolites detected in the WT strains. The compounds (m/z=602.33) and (m/z=586.34) were chosen to be purified from the strains. These will be referred to as Mutclu602 and Mutclu586 respectively. Both of these compounds used similar extract and purification procedures as listed in Section 4.3. Their NMR spectra are included in Appendix G.

Both structures appeared to be variations of mutanoclumpin without the macrocyclization. Mutclu602 is expected to precede the macrocyclization event, though it is uncertain if this mass is an off-chain product or a precursor to mutanoclumpin.

Compound Mutclu586 showed similar structures to Mutclu602 but contained what appeared to be an alanine in replacement of the serine moiety. To confirm this, labeled C13 alanine was fed into cultures of S1B WT. The mass spectrometry data, shown in Figure 56, identifies that the serine moiety is indeed a misincorporated alanine, preventing macrocyclization to the final molecule.

4.5.3 Structural Elucidation through MSMS

While purification of all BGC5 related metabolites was not feasible, structures may still be proposed using tandem mass spectrometry. A list of notable BGC5 metabolites was generated. Utilizing the MS2 of Mutclu602, mutanoclumpin, and Mutclu586, considerations could be made as to the likely structure of the various noted masses.

For mutanoclumpin, MS2 peaks were assigned utilizing CFM-ID to assist with fragmentation predictions. Of particular importance, the (m/z=419.25) peak was identified as a y and b amide bond break around the phenylalanine and a loss of water from the fatty acid chain. This water loss is also observed in the (m/z=566.32) fragment. The (m/z=419.25) peak is useful because the mass loss (-165.1), can only occur following a macrocyclization. Another noted peak included the a and y amide bond breaks around the phenylalanine resulting an (m/z=120.08) fragment. This is a common phenylalanine peptide fragment[62]. Lastly an (m/z=265.12) is expected to be an a and b amide bond break around the phenylalanine as well as a loss of water.

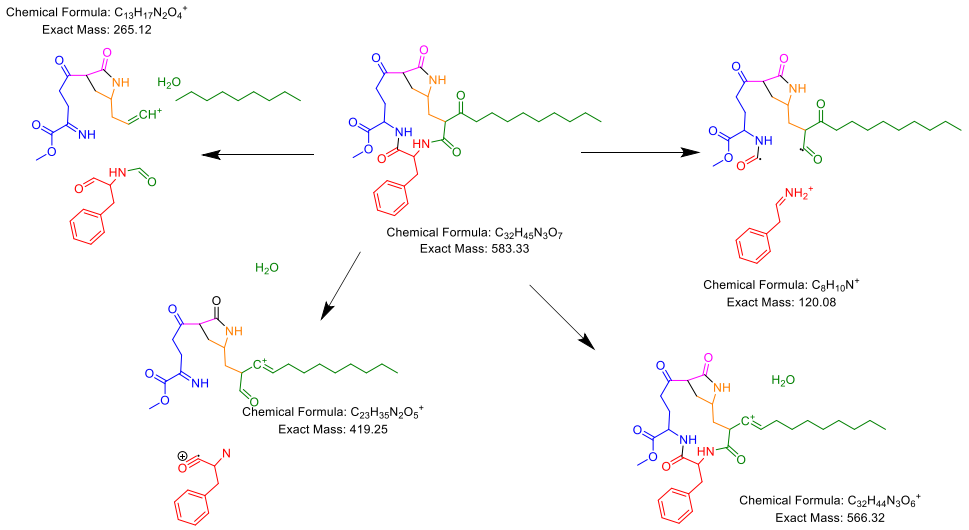


Figure 52: Proposed mutanoclumpin MS2 spectra fragments

For Mutclu602, the bond fragmentation follows a very similar pattern. The key (m/z=120.08) metabolite found in the macrocyclized molecule reasonably appears in linear fragment. This key fragment can serve to provide some confidence in late assembly line masses determined to be related to BGC5. There is a fragment (m/z=181.16) which occurs with a phenylalanine b amide fragmentation and a water loss. Lastly there is a (m/z=241.12) mass which occurs after a phenylalanine y fragmentation and a water loss. The (m/z=241.12) mass, is of particular relevance because it is only expected to occur in linear molecules.

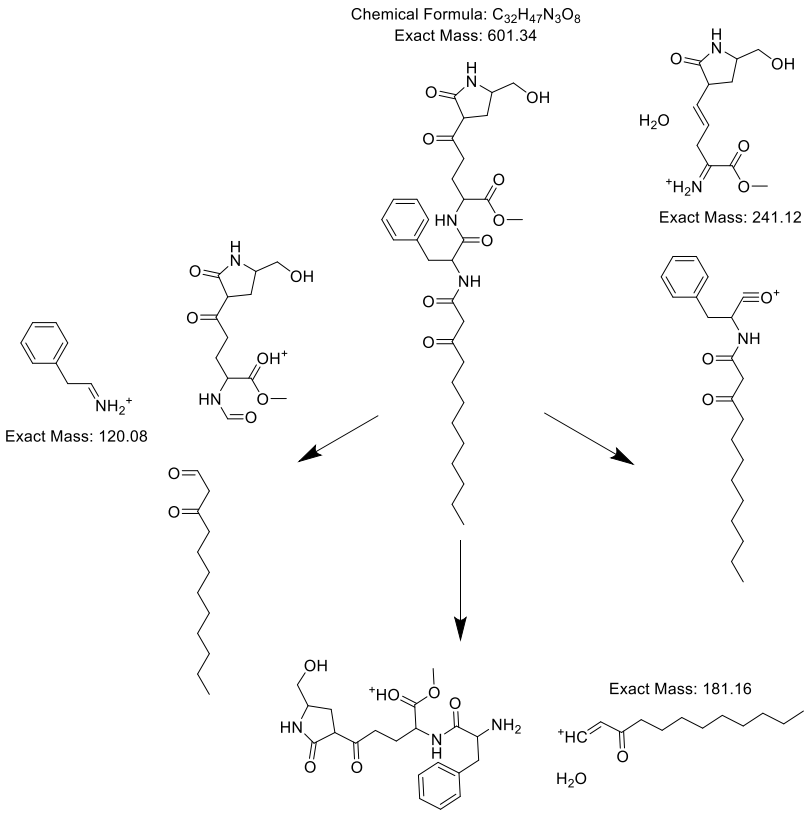


Figure 53: Proposed Mutclu602 MS2 spectra fragments

For Mutclu586, the key (m/z=120.08) metabolite remains. With NMR structural characterization, one can see what likely appears to be the incorporation of an alanine moiety in place of the serine. This was confirmed using isotope labeled alanine, illustrated in Figure 56. There is an (m/z=165.16) prominent fragment that is expected to be a phenylalanine b amide fragmentation with two water losses. Otherwise, the fragmentation resembles the Mutclu602 linear characteristic fragment, containing an (m/z=241.12) style fragment with a C-OH group removed (m/z=225.12).

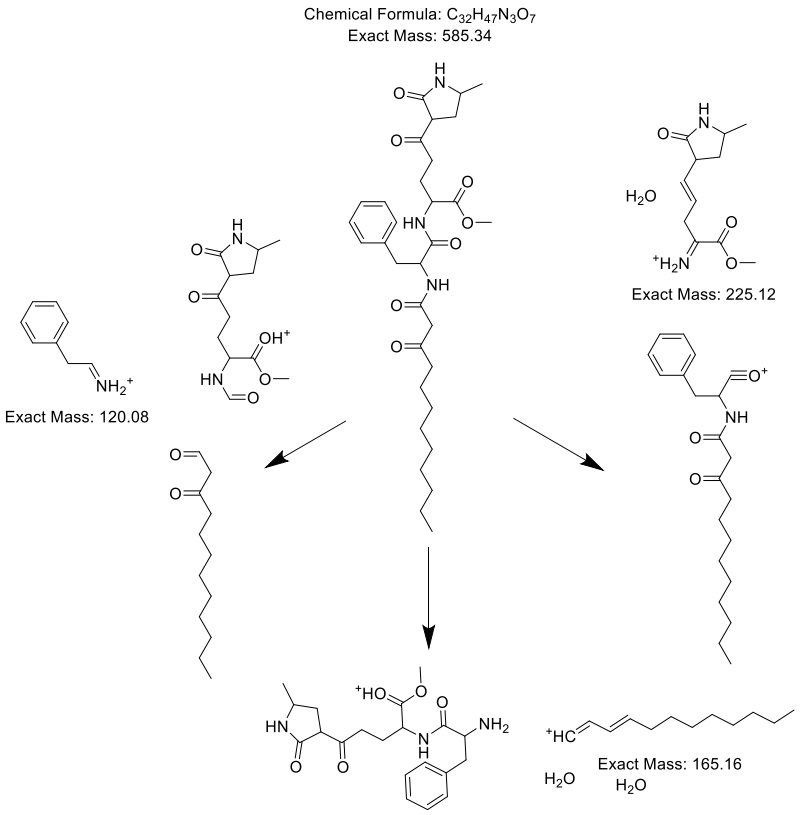


Figure 54: Proposed Mutclu602 MS2 spectra fragments

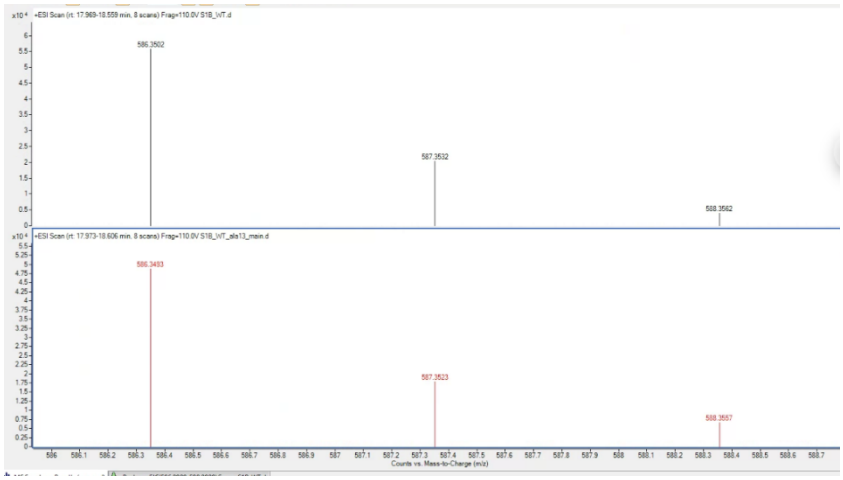


Figure 55: Streptococcus mutans fed unlabeled (black) and labeled (red) alanine. A significant increase in the (M+2) peak can be observed in the Mutclu586 molecule

With the NMR structures of these three molecules and their assigned peaks, a variety of intermediate structures can be tentatively assigned. Table details the various prediction for a variety of BGC5 related metabolites predicted to appear. Four characteristic variations were observed. The first is the ability for the compound to macrocyclize. All compounds with this showed a characteristic (-165.1) loss during fragmentation. Next the cyclization of the serine moiety or its substitution with alanine was

observed. Alanine fragments always showed a characteristic (m/z=225.12) fragment as previously discussed. If the alanine moiety hasn't be cyclized, the linear fragments show an additional (m/z=243.12) matching the added hydroxyl group. The serine moieties show matching fragments of (m/z=241.12) and (m/z=257.12) respectively. All of these observations were also confirmed using the serine and alanine labeling experiments to show appropriate incorporation of the proposed amino acid. Next desmethylated molecules showed a (-14) mass loss on all fragments containing the glutamate moiety. Lastly, the fatty acid chain length was detected by process of elimination of known fragments.

The wide variety of molecules detected suggests a few observations. The BGC5 starting domain seems flexible to the fatty acid loaded onto the molecule, especially for

Table 4: Proposed structural features of various mutanoclumpin associated metabolites based off MS2 spectra. Bold molecules have NMR structural confirmation

m/z	Key MS2 Fragmentation Peaks	Macrocycle	Serine/Alanine Cyclization/Substitution	Fatty Acid Chain	Methylation
560.29	120.08, 181.08, 227.10, 392.18	Linear	Cyclized - Serine	C8	Demethylated
574.31	120.08, 181.08, 241.12, 406.19	Linear	Cyclized - Serine	C8	Methylated
556.30	120.07, 182.08, 265.11, 391.22, 538.29	Cyclized	Cyclized - Serine	C8	Methylated
558.31	120.08, 165.10, 225.12	Linear	Cyclized - Alanine	C8	Methylated
602.34	**120.08, 181.09, 241.12**	**Linear**	**Cyclized - Serine**	**C10**	**Methylated**
620.34	120.08, 168.06, 181.09, 241.12	Linear	Linear - Serine	C10	Methylated
630.33	120.08, 181.08, 225.12, 241.12	Linear	Linear - Alanine	C12	Methylated
584.33	**120.08, 182.08, 265.11, 419.25,566.32**	**Cyclized**	**Cyclized - Serine**	**C10**	**Methylated**
586.34	**120.08, 165.10, 182.11, 225.12**	**Linear**	**Cyclized - Alanine**	**C10**	**Methylated**
604.32	120.08, 154.04, 225.12, 243.09	Linear	Linear - **Alanine**	C10	Methylated
588.32	120.08, 181.09, 227.10	Linear	Linear - Serine	C10	Demethylated
570.31	120.08, 168.06, 405.23, 552.30	Linear	Cyclized - Serine	C10	Demethylated

those of shorter chain length. The A domain used to attach serine to mutanoclumpin seems highly susceptible to incorporation of alanine. This prevents macrocyclization of the molecule due to the missing hydroxyl group. This also certainly confirms that the serine hydroxyl is vital to macrocyclization occurring. Lastly missing methylation of glutamate carboxylic acid doesn't appear to interfere with molecule construction.

Chapter 5: Identification and Preliminary Purification of a *Streptococcus mutans* Secondary Metabolite Discovered through Genome Mining

5.1 Introduction

As discussed in Chapter 4, *Streptococcus mutans* have a pantheon of biosynthetic capabilities, many of which are uncharacterized. Mutanoclumpin is one example of a unique small molecule with the ability to accentuate biofilm formation and autoaggregation in these species, a characteristic vital to survival in the oral microbiome. This small molecule was synthesized by the BGC5 gene cluster. This chapter will explore efforts into the discovery and characterization of the BGC6 gene cluster.

The BGC6 gene cluster, show below is a complex NRPS/PKS system that shows remarkable biosynthetic potential, and is widely present in *S. mutans* strains. Portions of this manufacturing line also appears to be present in the lactic acid bacterium *Lactococcus lactis* KF146. The larger gene cluster illustrated is the type II variation of the gene cluster, which is present in a variety of strains including *S. mutans* U2A.

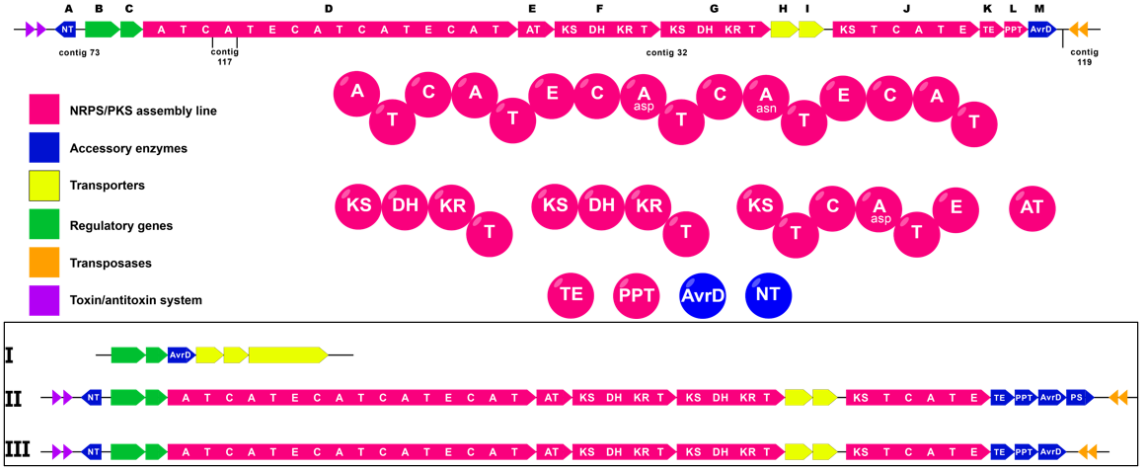


Figure 56: Biosynthetic gene cluster of BGC6

This BGC was illustrated to confer resistance to oxidative stress when grown on plant leaf tissue lysate. Given the extensive variation between the three common noted formations of the gene cluster (I, II, and II), these molecules may not be expected to perform the same cellular function between strains. Regardless, due to the unique enzymes and frequency amongst *Streptococcal* strains, the cluster is expected to be performing a vital function within these pathogens. A variety of labeling and genetic tools were utilized to approach the identification and characterization of this molecule.

5.2 Strain Targets

5.2.1 Lactococcus lactis KF147

Genetic tools were established by the Coates group to generate a knockout of the gene cluster in the KF147 strain. This KO illustrated a significant downregulation of transcription of the BGC6 gene cluster as illustrated by the Coates group through qPCR.

The *L. lactis* WT and KO strains were cultured in a variety of conditions. The first approach was to try a variety of media conditions. These conditions included BHI, ISP2,

LB, R2A, R5, and YPD media. The strains were grown in liquid cultures as well as on agar plates for 24 hours. Additionally, each condition was tested with the addition of 0.1% H_2O_2. As discussed in Chapter 4 Section 4.2, the addition of H_2O_2 can emulate oxidative stress caused by environmental oxygen exposure. Each of these conditions were extracted using 95:5 ethyl acetate:methanol. The dried down concentrates were injected in methanol with 100 uL cell culture equivalent onto HR-LCMS. These files were exported to MS-Dial, where they were compared.

Due to the incomplete silencing of the genes, comparative metabolomics in this case had to compare relative abundances rather than expect complete elimination of this compound. Additionally, compound abundance was expected to be upregulated during conditions of oxidative stress. So, the H_2O_2 containing WT sample should show an upregulation of this compound.

Unfortunately, none of the metabolites present met these criteria; nor was there a prominent metabolite that met solely the downregulation in KO criteria. This could be due to incomplete silencing of the BGC6 gene cluster by the knockout or a lack of expression coming from the wild type strain.

5.2.2 Streptococcus mutans U2A

Another approach for detecting a BGC6 metabolite would be performing genetic manipulation of some of the *Streptococcal* strains containing BGC6. While most BGC6 strains were resistant to transformation by competence inducing peptides (for reasons unknown), the attempts resulted in one strain, *S. mutans* U2A, that's genome could be edited. This allowed for the generation of a core biosynthetic gene KO, with the replacement of an Erythromycin resistance marker.

Similar to the BGC5 work, initial efforts were made to screen the BGC6 WT and KO strains, while performing comparative metabolomics. The initial medias tested included BHI and CDM media. These strains were also grown in liquid cultures in standing and shaking conditions at 37C for 16hrs.

5.3 Compound Identification

5.3.1 Gene Knockouts

As mentioned in Section 5.2.2, knockouts were generated of a core biosynthetic gene cluster in BGC6. This KO should abrogate all production of the downstream target.

5.3.2 Comparative Metabolomics

Unlike the BGC5 metabolite, upon performing extractions and metabolomics, no major peaks were initially detected that showed reliable production in the WT. A convoluting factor was also the growth rates of the strains. While the deletion of the BGC5 gene cluster doesn't noticeable affect growth rates with the strain, deletion of the BGC6 cluster causes a significant growth defect. *Streptococcal* strains are known to transiently produce a number of different signaling peptides. With variations in growth rate and expression in the cell, this causes comparative metabolomics to become extremely challenging to produce replicatable and reliable data. Without the presence of a persistent and high abundance metabolite, discerning an appropriate target becomes unviable.

In addition to the variable growth rates, the wild type *S. mutans* U2A strain showed significant variation in metabolome profiles between runs. This required many replicates to provide meaningful differentiation between the WT and KO strains.

A variety of approaches were employed to attempt to detect this small molecule. The first step was variation in the medias of choice. Two distinct medias, BHI and CDM, were selected for comparative metabolomics.

Next, a variety of culture conditions were trialed to attempt to detect the small molecule. These variations are made with the goal of recreating conditions that induce production by the BGC6 gene cluster. These variations included growth conditions in shaking cultures, standing cultures, growth on agar plates, and the addition of sand.

Table 5: Various conditions in BHI for BGC6 metabolite expression with (X) meaning no detection and (O) meaning detection

BHI	Shaking	Standing	Sand
Ethyl Acetate	X	X	X
95:5 EtAc: MeOH	X	X	X
3:1 Chlor: MeOH	X	X	X
2:1 Chlor:MeOH	X	X	X
Butanol	X	X	X
Pellet Methanol	X	X	X
XAD resin	X	X	X
HP20 resin	X	X	X

Lastly, a variety of extraction conditions were trialed to best extract the small molecule. Given the unknown structure of the compound, a variety of solvents were trialed. These include ethyl acetate, 95:5 ethyl acetate:methanol, 3:1 chloroform:methanol, 2:1 chloroform:methanol, pellet methanol extractions, butanol, XAD solid state resin, and HP20 solid state resin. Ideally this combination of solvents covers a wide swath of hydrophobicities with the goal of identifying a condition where the small molecule is detected. All of these are detailed in Table 5 and Table 6. Appendix contains further details regarding the detection of these molecules.

Table 6: Various conditions in CDM for BGC6 metabolite expression with (X) meaning no detection and (O) meaning detection

CDM	Shaking	Standing	Sand
Ethyl Acetate	X	X	X
95:5 EtAc: MeOH	X	X	X
3:1 Chlor: MeOH	X	X	X
2:1 Chlor:MeOH	X	X	X
Butanol	X	O	O
Pellet Methanol	X	O	O
XAD resin	X	O	O
HP20 resin	X	O	O

The data was screened using MS-Dial, a comparative metabolomics software. Most of the conditions showed no significant accumulation of a wild type compound that was absent consistently in the knockout. Fortunately, a few conditions showed significant accumulation of a family of small molecules present in a consistent retention time range on the instrument. Since molecules of a family are expected to have similar scaffold, many also have similar hydrophobicities. The predominant metabolite, m/z=473, looked like promising candidate for representing the major BGC6 metabolite. This compound was dereplicated using the molecular networking software GNPS. This metabolite was shown to network with a variety of the detected metabolites and was not shown to be a known molecule. This is in line with the BGC6 metabolites highly unique gene cluster.

Table 7: Peak picked list of metabolites appearing in BGC6 containing strain and absent in the KO; appears in descending order of abundance.

Detected Metabolites	Retention Time
473.26	**9.71**
545.28	9.88
589.29	9.88
597.29	10.46
615.32	10.92
641.31	10.2

Similar to BGC5, the gene cluster contains A domain proteins expected to incorporate asparagine and aspartic acid as amino acids. To confirm the identity of the small molecule, labeling experiments using N^{15} labeled amino acids were utilized. Shown in the figure below, incorporation of both molecules was observed, providing very strong evidence of the identity of the compound. This promoted an attempt to develop a method for purification.

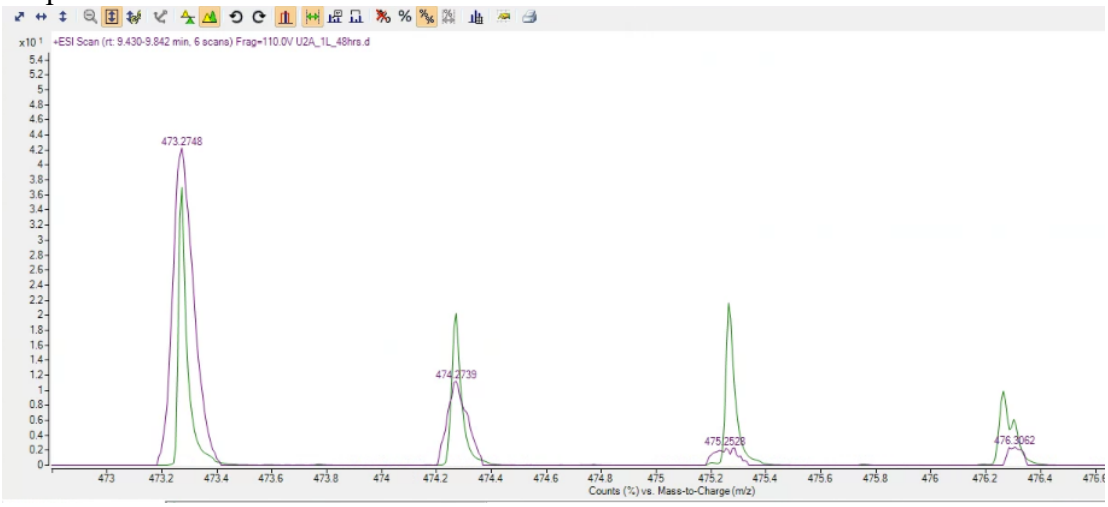


Figure 57: Streptococcus mutans U2A cultured with unlabeled (purple) and labeled (green) asparagine. A significant increase can be observed in both the (M+1) and (M+2) peaks of the m/z=473.3 molecule

Of similar significance, many different BGC6 growing strains were tested for production of 473. Fortunately, most strains in the collection containing BGC6 showed production of the 473 molecule. While U2A appears to be the highest producer, *Streptococcus mutans* NLML1 showed appreciable concentrations of the ion as well.

5.4 Compound Purification

Like BGC5, a variety of efforts were implemented to test the best production conditions and extraction optimizations for this small molecule.

5.4.1 Extraction Optimization

Various extraction conditions were tested to optimize 473 extraction efficiency. A variety of common organic solvents and a few solid-state polymeric resins were tested, shown in Figure 58. The solvent shown to best extract 473 was butanol + 1% acetic acid. Unfortunately, later purification efforts showed compound degradation under acidic conditions were observed. This led to large scale purifications using butanol without acid instead.

5.4.2 Growth Conditions Optimizations

Optimal compound production was observed between 12-16 hours. The window is

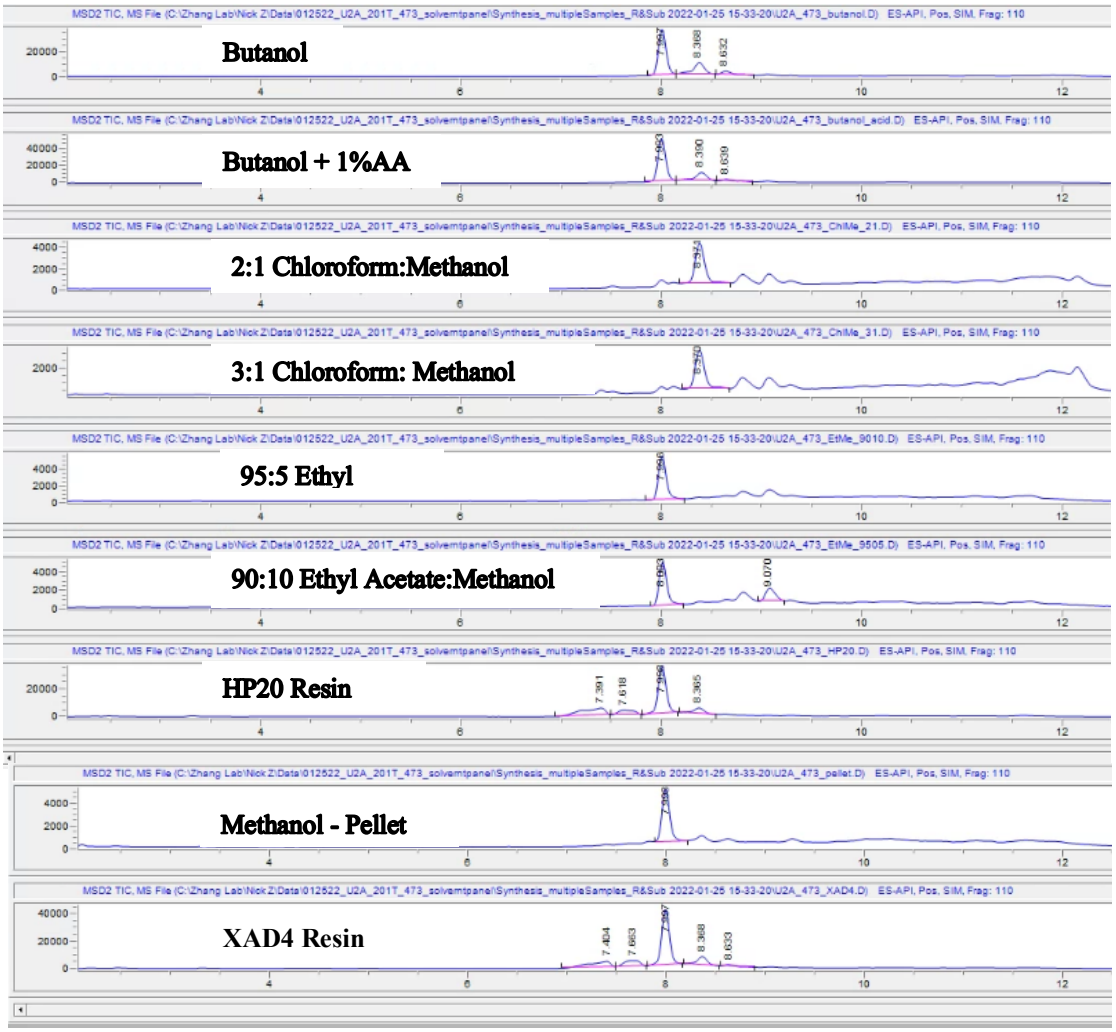


Figure 58: Solvent panel for extraction of the 473 metabolite from Streptococcus mutans U2A

larger than that observed in BGC5 due to the high sample to sample variability observed during culturing of this strain. This variability also produced a highly inconsistent amount of 473 between runs. This led to optimizations to attempt to increase the production of 473 within the strain.

Given that CDM is a chemically defined media (detailed in Appendix), components can be added and removed easily to produce optimal media conditions. It was proposed that the variation in compound production may be emerging from slight variations in the composition of the media. To test this phenomenon, a variety of CDM media components were supplemented to observe for any increases in production. Additionally, both amino acids shown to be incorporated, asparagine and aspartic acid, were supplemented as well.

A variety of supplements were shown to slightly elevate production. This led to a variation of the CDM media dubbed "CDM + BGC6 supplements".

5.4.3 Purification Conditions Optimizations

Streptococcus mutans U2A was started in 5 mLs of CDM media from a glycerol stock. This culture was used to inoculate a 4 L of CDM + BGC6 supplements media to an OD=0.01 and allowed to grow for 12-16 hours. Within that window, regular checks of compound production on a Single Quad mass spectrometer allowed for observation of the peak compound window. At this time, the culture was extracted using 1:1 butanol. This butanol was evaporated using a rotovap and the resulting dried compounds was resuspended in 5 mLs of methanol.

This concentrate was added to a packed LH20 size exclusion column suspended in methanol. Fractions were collected and tested on the mass spectrometer until the 473 compound eluted. This compound was dried down using a rotovap and the dried compounds were resuspended in 1 mL of methanol.

Given some observation of compound degradation observed using butanol + 1% acetic acid, a solvent system of water:acetonitrile without any acid was selected for HPLC. After optimizing a method, 200uL aliquots were injected onto an Agilent HPLC using an isocratic 85:15 water:acetonitrile method. Due to the low concentrations of the compound, it is still uncertain if the molecule contains a significant UV peak. Fractions were collected and run on the mass spectrometer to detect the elution of 473.

Concentrations of this compound were estimated on the mass spectrometer. It was determined that purified compound was at far too low abundance to perform NMR. This led to continued efforts to upregulate production of the compound.

5.5 Continued Efforts

To help confer production stability to a strain, native promoters that may be regulated by complex unknown mechanisms are sometimes replaced with constitutive promoters. Initial work into this space has shown that the replacement of the BGC6 native promoter with another *Streptococcal* constitutive promoter may help to stabilize production. Efforts will be continued to provide stable expression and eventually, structural and bioactive characterization of this important small molecule.

Chapter 6: Conclusions

Natural products have and continue to serve as an irreplaceable source of human medicine and provides insight into microbiological questions and understanding of disease. At the forefront of this field should be the development of methodology to further explore and characterize challenging microbial systems. Chapters 2 and 3 detailed some attempts to generate new strategies for upstream high throughput discovery of bioactive natural products. Each of these illustrated both the methodological approach, as well as their application in the discovery of novel compounds. Chapters 4 and 5 illustrated the usage of existing methodologies with both resound success and limited success respectively. Given that both take place in *Streptococcus mutans* species, it's a highly relevant example of how current methods in discovery can rely significantly on good fortune. Increasing the toolbox available to probes these systems is vital towards the continued discovery of therapies. Since 1970, no new class of antibiotics has been discovered. Hundreds of molecules are discovered each year, yet these molecules haven't revealed any novel class of antimicrobial for over 50 years now. However, each year that continues has seen a rise in antimicrobial resistance leading to disease and death. There are millions of gene clusters that produce novel natural products that have yet to be studied. This real issue comes not from the existence of these molecules, but from the means by which researchers seek them. Bioactivity based fraction is a common strategy used today that has existed for almost a century. The need for new methods to explore these spaces is pressing and necessary. These molecules have the ability to shape the human world and specifically, disease, in truly fantastical ways. It is pertinent that we refine what we explore, as well as how we explore it.

References

(1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661. https://doi.org/10.1021/acs.jnatprod.5b01055.

(2) Alexander Fleming Discovery and Development of Penicillin - Landmark - American Chemical Society https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/fleming penicillin.html (accessed Mar 10, 2022).

(3) Pham, J. V.; Yilma, M. A.; Feliz, A.; Majid, M. T.; Maffetone, N.; Walker, J. R.; Kim, E.; Cho, H. J.; Reynolds, J. M.; Song, M. C.; et al. A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Front. Microbiol.* **2019**, *10* (JUN), 1404. https://doi.org/10.3389/FMICB.2019.01404/BIBTEX.

(4) Conly, J. M.; Johnston, B. L. Where Are All the New Antibiotics? The New Antibiotic Paradox. *Canadian Journal of Infectious Diseases and Medical Microbiology*. 2005. https://doi.org/10.1155/2005/892058.

(5) Pang, Z.; Raudonis, R.; Glick, B. R.; Lin, T. J.; Cheng, Z. Antibiotic Resistance in Pseudomonas Aeruginosa: Mechanisms and Alternative Therapeutic Strategies. *Biotechnol. Adv.* **2019**, *37* (1), 177–192. https://doi.org/10.1016/J.BIOTECHADV.2018.11.013.

(6) Li, Z. R.; Sun, J.; Du, Y.; Pan, A.; Zeng, L.; Maboudian, R.; Burne, R. A.; Qian, P. Y.; Zhang, W. Mutanofactin Promotes Adhesion and Biofilm Formation of Cariogenic Streptococcus Mutans. *Nat. Chem. Biol.* **2021**, *17* (5), 576–584. https://doi.org/10.1038/S41589-021-00745-2.

(7) Richards, J. P.; Cai, W.; Zill, N. A.; Zhang, W.; Ojha, A. K. Adaptation of Mycobacterium Tuberculosis to Biofilm Growth Is Genetically Linked to Drug Tolerance. *bioRxiv* **2019**, 663369. https://doi.org/10.1101/663369.

(8) Worthington, R. J.; Richards, J. J.; Melander, C. Organic & Biomolecular Chemistry Small Molecule Control of Bacterial Biofilms. **2012**. https://doi.org/10.1039/c2ob25835h.

(9) Kraemer, S. M.; Duckworth, O. W.; Harrington, J. M.; Schenkeveld, W. D. C.; Kraemer, S. M.; Schenkeveld, W. D. C.; Duckworth, O. W.; Harrington, J. M. Metallophores and Trace Metal Biogeochemistry. *Aquat. Geochemistry* **2015**, *21*, 159–195. https://doi.org/10.1007/s10498-014-9246-7.

(10) Collins, H. L. Withholding Iron as a Cellular Defence Mechanism-Friend or Foe? https://doi.org/10.1002/eji.200838505.

(11) Matsui, R.; Cvitkovitch, D. Acid Tolerance Mechanisms Utilized by Streptococcus Mutans. *Future Microbiology*. 2010. https://doi.org/10.2217/fmb.09.129.

(12) Fountain, J. C.; Bajaj, P.; Pandey, M.; Nayak, S. N.; Yang, L.; Kumar, V.; Jayale, A. S.; Chitikineni, A.; Zhuang, W.; Scully, B. T.; et al. Oxidative Stress and Carbon Metabolism Influence Aspergillus Flavus Transcriptome Composition and Secondary Metabolite Production OPEN. *Nat. Publ. Gr.* **2016**. https://doi.org/10.1038/srep38747.

(13) Häussler, S.; Becker, T. The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in Pseudomonas Aeruginosa Populations. *PLoS Pathog.* **2008**, *4* (9). https://doi.org/10.1371/JOURNAL.PPAT.1000166.

(14) Perry, E. K.; Meirelles, L. A.; Newman, D. K. From the Soil to the Clinic: The Impact of Microbial Secondary Metabolites on Antibiotic Tolerance and Resistance. https://doi.org/10.1038/s41579-021-00620-w.
(15) Barnard, A. M. L.; Bowden, S. D.; Burr, T.; Coulthurst, S. J.; Monson, R. E.; Salmond, G. P. C. Quorum Sensing, Virulence and Secondary Metabolite Production in Plant Soft-Rotting Bacteria. https://doi.org/10.1098/rstb.2007.2042.
(16) Li, W.-J.; Albarracín, V. H.; Habib, N.; Jangid, K.; Polkade, A. V; Mantri, S. S.; Patwekar, U. J. Quorum Sensing: An Under-Explored Phenomenon in the Phylum Actinobacteria. **2016**. https://doi.org/10.3389/fmicb.2016.00131.
(17) Rangel, L. I.; Hamilton, O.; De Jonge, R.; Bolton, M. D. Fungal Social Influencers: Secondary Metabolites as a Platform for Shaping the Plant-Associated Community. **2021**. https://doi.org/10.1111/tpj.15490.
(18) Banerjee, P.; Erehman, J.; Bj¨, B.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II-a Database of Natural Products. *Nucleic Acids Res.* **2014**, *43*, 935–939. https://doi.org/10.1093/nar/gku886.
(19) Mikulic, M. Number of amoxicillin prescriptions in the U.S. from 2004 to 2019.
(20) Triggle, D. J.; Mitchell, J. M.; Filler, R. The Pharmacology of Physostigmine. **1998**.
(21) Li, J.; Kim, S. G.; Blenis, J. Rapamycin: One Drug, Many Effects. *Cell Metab.* **2014**, *19* (3), 373–379. https://doi.org/10.1016/J.CMET.2014.01.001.
(22) Ng T, G. V. Tetrahydrocannabinol (THC) - StatPearls - NCBI Bookshelf https://www.ncbi.nlm.nih.gov/books/NBK563174/ (accessed Mar 10, 2022).
(23) Omega-3 fatty acids and cardiovascular disease - PubMed https://pubmed.ncbi.nlm.nih.gov/25720716/ (accessed Mar 10, 2022).
(24) Weller, M. G. A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques. *Sensors (Basel).* **2012**, *12* (7), 9181. https://doi.org/10.3390/S120709181.
(25) Abegaz, B. M.; Kinfe, H. H. Secondary Metabolites, Their Structural Diversity, Bioactivity, and Ecological Functions: An Overview. *Phys. Sci. Rev.* **2019**, *4* (6). https://doi.org/10.1515/PSR-2018-0100/XML.
(26) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; et al. Reproducible Molecular Networking of Untargeted Mass Spectrometry Data Using GNPS. *Nat. Protoc.* https://doi.org/10.1038/s41596-020-0317-5.
(27) Bachmann, B. O.; Covington, B. C.; Mclean Ab, J. A. Natural Product Reports Comparative Mass Spectrometry-Based Metabolomics Strategies for the Investigation of Microbial Secondary Metabolites Comparative Mass Spectrometry-Based Metabolomics Strategies for the Investigation of Microbial Secondary Metabolites. **2017**, *34*, 1–124. https://doi.org/10.1039/c6np00048g.
(28) Harris, N. C.; Sato, M.; Herman, N. A.; Twigg, F.; Cai, W.; Liu, J.; Zhu, X.; Downey, J.; Khalaf, R.; Martin, J.; et al. Biosynthesis of Isonitrile Lipopeptides by Conserved Nonribosomal Peptide Synthetase Gene Clusters in Actinobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (27), 7025–7030. https://doi.org/10.1073/pnas.1705016114.
(29) Kushner, D. J.; Baker, A.; Dunstall, T. G. Pharmacological Uses and Perspectives of Heavy Water and Deuterated Compounds. *Can. J. Physiol. Pharmacol.* **1999**,

77 (2), 79–88. https://doi.org/10.1139/y99-005.
(30) Seidel, J.; Miao, Y.; Porterfield, W.; Cai, W.; Zhu, X.; Kim, S.; Bhattarai-kline, S.; Min, W.; Zhang, W. Live-Cell Imaging Analysis of Antimycin-Type Depsipeptides via Bioorthogonal Stimulated Raman Scattering Microscopy. **2019**, 8–11.
(31) Palaniappan, K. K.; Pitcher, A. A.; Smart, B. P.; Spiciarich, D. R.; Iavarone, A. T.; Bertozzi, C. R. Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics. *ACS Chem. Biol.* **2011**, *6* (8), 829–836. https://doi.org/10.1021/cb100338x.
(32) Berry, D.; Mader, E.; Lee, T. K.; Woebken, D.; Wang, Y.; Zhu, D.; Palatinszky, M.; Schintlmeister, A.; Schmid, M. C.; Hanson, B. T.; et al. Tracking Heavy Water (D2O) Incorporation for Identifying and Sorting Active Microbial Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (2), E194–E203. https://doi.org/10.1073/pnas.1420406112.
(33) Relationship between Binomial and Normal Distributions | Real Statistics Using Excel http://www.real-statistics.com/binomial-and-related-distributions/relationship-binomial-and-normal-distributions/ (accessed Sep 3, 2019).
(34) Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. *BMC Bioinformatics* **2007**, *8*, 1–20. https://doi.org/10.1186/1471-2105-8-105.
(35) Mass Spectrometry: Isotope Effects - Chemistry LibreTexts https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Instrumental_Analysis/Mass_Spectrometry/Mass_Spectrometry%3A_Isotope_Effects (accessed Sep 1, 2019).
(36) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.
(37) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. Practical *Streptomyces* Genetics. *John Innes Found. Norwich, Engl.* **2000**.
(38) Lastair, A.; Ood, J. J. W.; Ailey, J. B.; Ath, M. R. C. P.; Obert, R.; Urner, C. T. *Metformin*; 1996.
(39) Mooney, M. H.; Fogarty, S.; Stevenson, C.; Gallagher, A. M.; Palit, P.; Hawley, S. A.; Hardie, D. G.; Coxon, G. D.; Waigh, R. D.; Tate, R. J.; et al. Mechanisms Underlying the Metabolic Actions of Galegine That Contribute to Weight Loss in Mice. *Br. J. Pharmacol.* **2008**, *153*, 1669–1677. https://doi.org/10.1038/bjp.2008.37.
(40) Wang, H. X.; Ng, T. B. Natural Products with Hypoglycemic, Hypotensive, Hypocholesterolemic, Antiatherosclerotic and Antithrombotic Activities. *Life Sci.* **1999**, *65* (25), 2663–2677. https://doi.org/10.1016/S0024-3205(99)00253-2.
(41) Boronat, A.; Rodríguez-Concepción, M.; Boronat, A.; Rodríguez, Á. M. Terpenoid Biosynthesis in Prokaryotes. *Adv. Biochem. Eng. Biotechnol.* **2014**, *148*, 3–18. https://doi.org/10.1007/10_2014_285.
(42) Li, M.; Chen, L.; Deng, Z.; Zhao, C. Characterization of AmtA, an

Amidinotransferase Involved in the Biosynthesis of Phaseolotoxins. *FEBS Open Bio* **2016**, *6* (6), 603. https://doi.org/10.1002/2211-5463.12071.
(43) González, Y.; Torres-Mendoza, D.; Jones, G. E.; Fernandez, P. L. Marine Diterpenoids as Potential Anti-Inflammatory Agents. *Mediators Inflamm.* **2015**, *2015*. https://doi.org/10.1155/2015/263543.
(44) Costa Nicolaou, K.; Dai, W.-M.; Kiplin Guy, R. T. *Chemistry and Biology of Taxol*.
(45) White, N. J. Qinghaosu (Artemisinin): The Price of Success. *Science (80-.).* **2008**, *320* (5874), 330–334. https://doi.org/10.1126/SCIENCE.1155165/ASSET/7E1AC132-8134-454E-B17B-0C5101BDE146/ASSETS/GRAPHIC/320_330_F6.JPEG.
(46) James Hanson, D. R. Diterpenoids. **2006**. https://doi.org/10.1039/b516326a.
(47) Perveen, S. Introductory Chapter: Terpenes and Terpenoids; 2021. https://doi.org/10.5772/intechopen.98261.
(48) Wallemacq, P. E. Therapeutic Monitoring of Immunosuppressant Drugs. Where Are We? *Clin. Chem. Lab. Med.* **2004**, *42* (11), 1204–1211. https://doi.org/10.1515/CCLM.2004.242/MACHINEREADABLECITATION/RIS.
(49) Usui, T.; Nagumo, Y.; Watanabe, A.; Kubota, T.; Komatsu, K.; Kobayashi, J.; Osada, H. Brasilicardin A, a Natural Immunosuppressant, Targets Amino Acid Transport System L. *Chem. Biol.* **2006**, *13* (11), 1153–1160. https://doi.org/10.1016/j.chembiol.2006.09.006.
(50) York, N.; Flynn, E. H.; Sigal, M. V; Wiley, P. F.; Gerzon, K. Erythromycin : Properties and Degradation Studies Erythromycin. I. Properties and Degradation Studies1. **1954**.
(51) Zulak, K. G.; Bohlmann, J. Terpenoid Biosynthesis and Specialized Vascular Cells of Conifer Defense. *Journal of Integrative Plant Biology*. 2010. https://doi.org/10.1111/j.1744-7909.2010.00910.x.
(52) Sanketh, D.; Amrutha, N. Oral Microbial Flora in Health. *World J. Dent.* **2013**. https://doi.org/10.5005/jp-journals-10015-1242.
(53) Deo, P. N.; Deshmukh, R. Oral Microbiome: Unveiling the Fundamentals. *Journal of Oral and Maxillofacial Pathology*. 2019. https://doi.org/10.4103/jomfp.JOMFP_304_18.
(54) Vujicic, M.; Listl, S. An Economic Perspective of the Global Burden of Dental Caries. *ACFF*.
(55) Vos, T.; Allen, C.; Arora, M.; Barber, R. M.; Brown, A.; Carter, A.; Casey, D. C.; Charlson, F. J.; Chen, A. Z.; Coggeshall, M.; et al. Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 310 Diseases and Injuries, 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet* **2016**. https://doi.org/10.1016/S0140-6736(16)31678-6.
(56) Eick, S. Biofilms. *Monographs in Oral Science*. 2020. https://doi.org/10.1159/000510184.
(57) Worthington, R. J.; Richards, J. J.; Melander, C. Small Molecule Control of Bacterial Biofilms. *Organic and Biomolecular Chemistry*. 2012. https://doi.org/10.1039/c2ob25835h.
(58) Kaspar, J.; Underhill, S. A. M.; Shields, R. C.; Reyes, A.; Rosenzweig, S.; Hagen,

S. J.; Burne, R. A. Intercellular Communication via the ComX-Inducing Peptide (XIP) of Streptococcus Mutans. *J. Bacteriol.* **2017**. https://doi.org/10.1128/JB.00404-17.
(59) Dosani, S. Penicillin Man: Alexander Fleming and the Antibiotic Revolution. *BMJ* **2005**. https://doi.org/10.1136/bmj.330.7481.50-a.
(60) Wen, Z. T.; Burne, R. A. LuxS-Mediated Signaling in Streptococcus Mutans Is Involved in Regulation of Acid and Oxidative Stress Tolerance and Biofilm Formation. *J. Bacteriol.* **2004**. https://doi.org/10.1128/JB.186.9.2682-2691.2004.
(61) O'Toole, G. A. Microtiter Dish Biofilm Formation Assay. *J. Vis. Exp.* **2010**, No. 47. https://doi.org/10.3791/2437.
(62) Scheubert, K.; Hufsky, F.; Böcker, S. Computational Mass Spectrometry for Small Molecules. *J. Cheminform.* **2013**, *5* (3). https://doi.org/10.1186/1758-2946-5-12.
(63) Reifycs Abf Converter https://www.reifycs.com/AbfConverter/ (accessed Mar 28, 2022).
(64) CompMS | MS-DIAL http://prime.psc.riken.jp/compms/msdial/main.html (accessed Mar 28, 2022).
(65) Van De Rijn, I.; Kessler, R. E. Growth Characteristics of Group A Streptococci in a New Chemically Defined Medium. *Infect. Immun.* **1980**. https://doi.org/10.1128/iai.27.2.444-448.1980.

Appendix

A. Equipment, Chromatographic Resins, and Solid State Resins

A.1 Mass spectrometry

Extracted metabolites were run on an Agilent QToF 6545 mass spectrometer with an Agilent C18 analytical column and an Agilent 1240 DAD detector upstream. The instrument is run at 100-1700 Da mass range and the fragmenter is set to 110 V. These files are recorded by Agilent Mass Hunter Workstation Data Acquisition Version B.09.00.

A.2 HPLC

Concentrated extracts were purified using an Agilent 1200 HPLC with a Sephadex C18 column and an Agilent 1200 DAD detector upstream.

A.3 NMR

A 900 MHz NMR was utilized to solve structures. Approximately 1 mg of compounds was suspended in $CDCl_3$.

A.4 Various Resins

For size exclusion chromatography, a packed column of sephadex LH20 resin with an eluent of methanol.

For packed reverse phase chromatography, C18 SP columns were utilized.

For solid state extractions, XAD4 resin was utilized.

B. Mass Spectrometry Data Analysis Workflow

Samples are extracted and concentrated so that 100uL cell culture equivalent is injected in 10uL of methanol onto the mass spectrometer. The data is saved in the Agilent mass spectrometry data format, .d.

These .d files are converted into .abf files using Reifycs Analysis Base File Converter[63]. These files are uploaded to MSDial[64] using the base parameters with the following adjustments: MS1 tolerance=0.01 Da, Retention time tolerance = 0.15 min. The MSDial performs peak picking, and the alignment result is exported as a .txt file which is converted into an .xls file in excel.

These files are then uploaded to Matlab in the "D_2O Sorter" program. The "D_2O Sorter" program utilizes three columns: m/z, retention time, and abundance. These peaks are examined in order to pair relevant isotopologues. Once paired, the relative abundances are screened versus the maxima expected for non-deuterated small molecules within the given mass range. The chosen peaks are placed into a file with details of their selection. Results can be exported within the program using the write to excel feature, or results can be read within the program.

C. D_2O Software Detection Code

Below is the "D_2O Sorter" code used to identify deuterated small molecules

file=xlsread('Insert Excel File Name')

```
%file=readtable([namesave{po} '.txt'],'Delimiter', ' ');


%Parameters
error1=0.1;%
RTerror=.08; %0.08mins, allowed retention time error/dalton
MinAbund=0; %Minimum allowed abundnace
ratio1toM=0.09;
ratio2to1=0.01;

if(individual==true)
%file=file(3:end,[2,3,7]);
%file(:,[1 2])=file(:,[2 1]);
mzdata=[];
mzdata(:,1:3)=file(:,1:3);
mzdata=sortrows(mzdata,1);%Sorts the m/z's into rention time order and matches them
with their according intensity value
elseif(individual==false)
file=table2array((file(4:end,[2,3,27])))
file(:,[1 2 3])=file(:,[2 1 3])
mzdata=[];
mzdata(:,1:3)=cell2mat(file(:,1:3));
mzdata=sortrows(mzdata,1)
end

%For Specific Searching
if(true)
mzlimit=[400,500];
rtlimit=[33,37];
mzdata(mzdata(:,1)>mzlimit(2),:)=[];
mzdata(mzdata(:,1)<mzlimit(1),:)=[];
mzdata(mzdata(:,2)>rtlimit(2),:)=[];
mzdata(mzdata(:,2)<rtlimit(1),:)=[];
end

%MZ Isotope Sort

mzlist=zeros(60,length(mzdata));
abundancelist=zeros(60,length(mzdata));
rtlist=zeros(60,length(mzdata));

 for sw=1:length(mzdata);%
     mztemp=zeros(45,1);
     abundtemp=zeros(45,1);
     rttemp=zeros(45,1);
```

```
mztemp(1,1)=mzdata(sw,1);
abundtemp(1,1)=mzdata(sw,3);
rttemp(1,1)=mzdata(sw,2);

error=error1; %%gives a ppm error range


  rtdiff=abs(mzdata(sw,2)-mzdata(:,2));
  integerdiff=round(mzdata(:,1)-mzdata(sw,1));
  rtmatchesself =rtdiff<RTerror*integerdiff; %& rtdiff>-.005; %makes the triangle that gives a larger range with biger integer differences
  mzmatches = abs(mzdata(:,1)-mzdata(sw,1)- abs(1.00627*round(mzdata(:,1)-mzdata(sw,1))))<error & mzdata(:,1)-mzdata(sw,1)<10;

 %Here we want to ensure that the M has a viable M+1 and M+2 before we let if fiend on our higher order ones.
  abund1match=((mzdata(:,3).*(integerdiff==1)))>0;
  correctabund1match=abund1match&mzmatches&rtmatchesself;
  abund2match=((mzdata(:,3).*(integerdiff==2)))>0;
  correctabund2match=abund2match&mzmatches&rtmatchesself;

abundboth=any(correctabund1match)&any(correctabund2match);

  if(abundboth)
  up1=find(correctabund1match&rtmatchesself,1);
  mida1=rtmatchesself(sw+1:up1-1);
  up2=find(correctabund2match&rtmatchesself,1);
  mida2=rtmatchesself(up1+1:up2-1,1);
  midders=any(mida1)&any(mida2);
  else
  midders=0;
  end

  % if()
 %  mids1=mzdata(mzdata(:,1)<up1(1)&mzdata(:,1)>mzdata(sw,1),:)-mzdata(sw,1); %These lines produce comparable values between 0-1 for us to compare fragments
  % mids2=mzdata(mzdata(:,1)<up2(1)&mzdata(:,1)>up1(1),:)-up1(1);
  % mids3=mzdata(mzdata(:,1)<up3(1)&mzdata(:,1)>up2(1),:)-up2(1);
  % end
  % end


  all=rtmatchesself&mzmatches;

if(any(mzdata(all,2)~=mzdata(sw,2)))
```

```
    tooperfect=false;
  else
    tooperfect=true;
  end


    if(any(all)&abundboth&~midders&~tooperfect) %accounts for xcms fuck-ups and averages in rt/mz based on abundance and adds abundances

        position=round(mzdata(all,1)-mzdata(sw,1));
        mztemp(position+1,1)=mzdata(all,1);
        rttemp(position+1,1)=mzdata(all,2);
        abundtemp(position+1,1)=mzdata(all,3);
        mzdata(all,:)=0;
    end

    mzlist(1:length(mztemp),sw)=mztemp;
    abundancelist(1:length(abundtemp),sw)=abundtemp;
    rtlist(1:length(rttemp),sw)=rttemp;

end


  %This portion removes all of the zeros from the code

%gets rid of rows with all zeros
abundancelist(~any(mzlist,2),:)=[];
rtlist(~any(mzlist,2),:)=[];
mzlist(~any(mzlist,2),:)=[];

abundancelist(:,~any(mzlist,1))=[];  %gets rid of columns with any zeros
rtlist(:,~any(mzlist,1))=[];
mzlist(:,~any(mzlist,1))=[];


 % with all the values that made it past retention time
%correction stuff
AAA=[];
if(size(mzlist,1)>1)
AAA(:,1)=mzlist(1,:)';
AAA(:,2)=sum(mzlist~=0,1)';
AAA(:,3)=rtlist(1,:)'; %averages their two retention times for display
AAA(:,4)=abundancelist(1,:)'; %averages abundances for display
AAA(:,5)=(abundancelist(2,:))';
AAA(:,6)=(abundancelist(3,:))';
if(size(abundancelist,1)>3)
```

```
AAA(:,7)=(abundancelist(4,:))';
end
%AAA(:,8)=(abundancelist(5,:))';
AAA(:,8)=AAA(:,4)./AAA(:,4);
AAA(:,9)=AAA(:,5)./AAA(:,4);
AAA(:,10)=AAA(:,6)./AAA(:,4);
AAA(:,11)=AAA(:,7)./AAA(:,4);

AAA(:,15)=sum(abundancelist(1:end,:))';
AAA(:,16)=0.00078254.*AAA(:,1)+1*0.0476934;
AAA(:,17)=0.000000287705706*(AAA(:,1).^2)+0.000039169586817*AAA(:,1)+1*0.001
723397268419;
AAA(:,18)=0.00115025.*AAA(:,1)+1*2.3113929;
AAA(:,19)=-
0.000001637010966*(AAA(:,1).^2)+0.008100842649*AAA(:,1)+1*2.087517543739501;

AAA(AAA(:,2)<2,:)=[]; %removes all samples outside of abundance range

else
AAA=[]
end


% Applies Formula Removals (using my max molecular formula predictor)
  min=(AAA(:,5)./AAA(:,4))<AAA(:,16)|(AAA(:,6)./AAA(:,4))<AAA(:,17);

%min=(AAA(:,5)./AAA(:,4))<realistic*peaktempsmin(:,2)|(AAA(:,6)./AAA(:,4))<realistic
*peaktempsmin(:,3)|(AAA(:,7)./AAA(:,4))<realistic*peaktempsmin(:,4);
  max=(AAA(:,5)./AAA(:,4))>AAA(:,18)|(AAA(:,6)./AAA(:,4))>AAA(:,19);

%max=AAA(:,6)>1.1*peaktempsmax(:,2)|AAA(:,7)>1.1*peaktempsmax(:,3)|AAA(:,8)>1
.1*peaktempsmax(:,4)|AAA(:,9)>1.1*peaktempsmax(:,5)



eval(['AAA_' namesave{po} '=AAA']);
eval(['AA_fake_' namesave{po} '=AA_fake']);
eval(['AA_real_' namesave{po} '=AA_real']);
eval([namesave{po} '_results={AA_fake_' namesave{po} ', AA_real_' namesave{po} ',
AAA_' namesave{po} ', peaktempsmin}' ])

eval(['save ' namesave{po} '_results ' namesave{po} '_results'])

namesave{po}

end
```

```
%% Table
for poq=1
%Makes an actual table
r={'M','# ++', '# 1-3', 'RT','Abundance M','Abund M+1/M','Abund M+2/M+1','RT slope','R^2'};
f=figure('position',[600 800 300 500],'units','normalized','outerposition',[0 0 1 1]);
uitable(f,'data',AAA,'columnname',r,'position',[0 0 700 600],'ColumnWidth', 'auto')

toc;
end
```

D. *Streptococcal* Gene Cluster Identifications

167 S mutans publicly available genomes were imported into antismash. These results were input into Bigscape in order to illustrate the phylogenetic distribution of BGCs present in the strains. These gene clusters were analyzed and paired down to seven key reoccurring gene clusters.

E. Various Media Formulations

MS Media
20 g mannitol
20 g soy flour (Bob's red mill preferred)
20 g agar
1L Tap Water

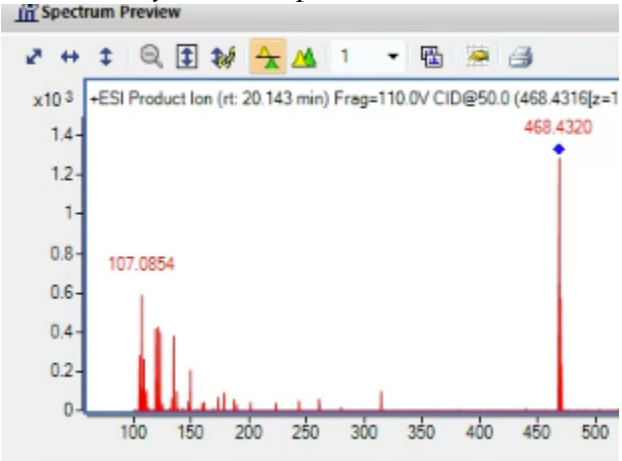
A Media
20 g soluble starch
10 g glucose
5 g peptone
5 g yeast extract
4 g NaCl
0.5 g K_2HPO_4
0.5g $MgSO_4*7H_2O$
2g $CaCO_3$
1L DI water

Streptococcal Chemical Defined Media (CDM)
Details of this media can be found by Rijn et al. 1980[65]

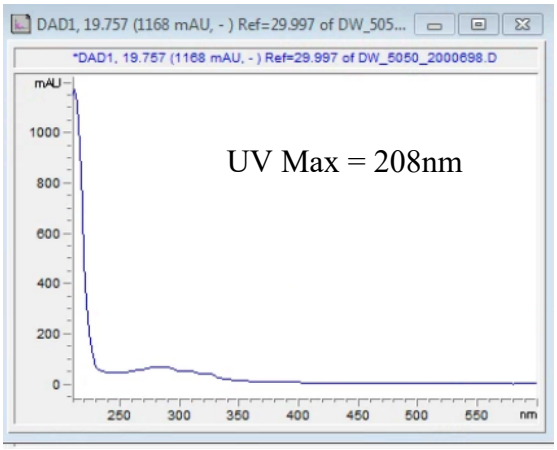
- Media was only kept for up to 3 days to noticeable declines in growth rates with old media

F. Molecule NMR, MS2, UV, and LH20

F.1A Zillamycin MS2 Spectra

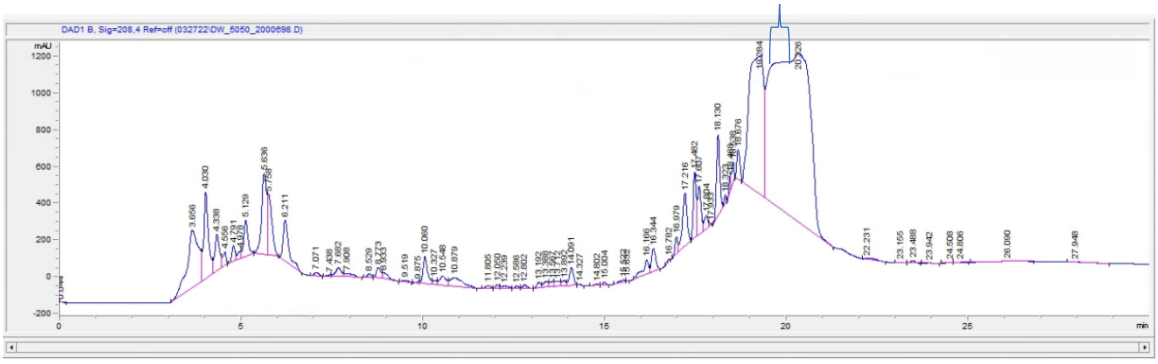


F.1B Zillamycin UV and HPLC Purification

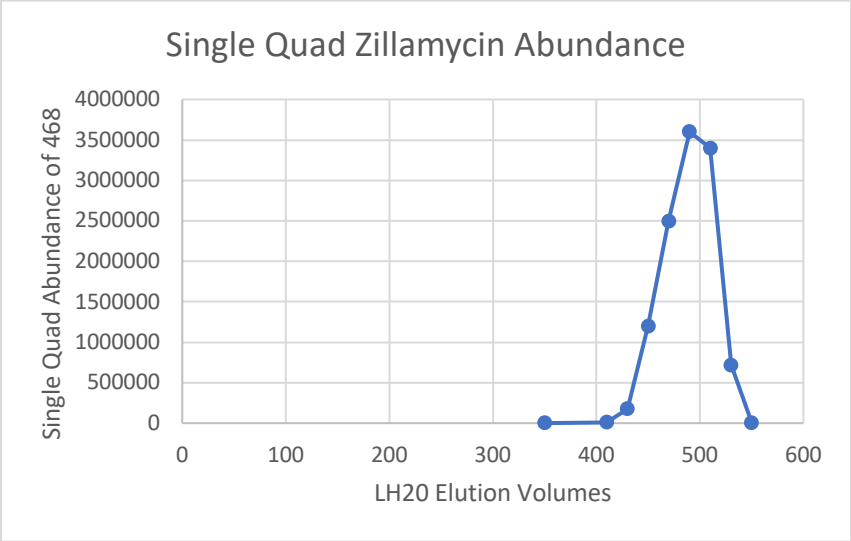


H2O:ACN + 0.1% TFA
50:50->0:100 0-15 mins
0:100 Hold 15-30 mins

m/z=468.40



F.1C Zillamycin LH20 Purification



F.1D Zillamycin NMR

Structural Elucidation of Compound 468. Compound 468 had the molecular formula $C_{31}H_{53}N_3$ based on a proton adduct ion at *m/z* 468.4313 $[M+H]^+$ in its positive ion HRESIMS spectrum. Analysis of its ^1H NMR, and ^{13}C NMR spectroscopic data of compound 468 (Table 1) indicated that compound 468 included one guanidine moiety and six isopentene parts. Furthermore, the structure of compound 468 was confirmed as a triterpene guanidine through investigating its COSY and HMBC spectra.

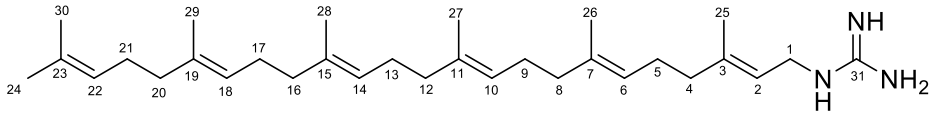
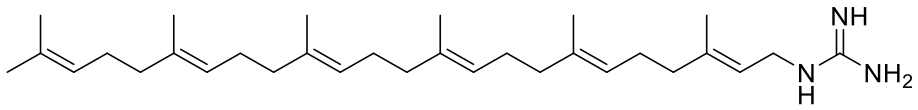


Table F.1D-S1 The NMR Data of Compound 468 **1** in DMSO-d_6					
Position	δ_H (J in Hz)	δ_C (C type)	Position	δ_H (J in Hz)	δ_C (C type)
1	3.71 dd 6.7, 5.2	39.8, CH_2	17	2.02 m	26.0, CH_2
2	5.18 t 6.7	118.6, CH	18	5.09 m	124.0, CH
3		139.7, C	19		134.3, C
4	1.99 m	38.8, CH_2	20	1.93 m	38.9, CH_2
5	2.05 m	25.9, CH_2	21	2.02 m	26.2, CH_2
6	5.09 m	123.6, CH	22	5.07 m	124.1, CH
7		134.7, C	23		130.6, C
8	1.93 m	39.2, CH_2	24	1.63 s	25.5, CH_3
9	2.02 m	26.0, CH_2	25	1.64 s	16.1, CH_3
10	5.07 m	123.9, CH	26	1.56 s	15.8, CH_3
11		134.2, C	27	1.55 s	15.8, CH_3
12	1.93 m	39.1, CH_2	28	1.55 s	15.8, CH_3
13	2.02 m	26.0, CH_2	29	1.55 s	15.8, CH_3
14	5.09 m	123.9, CH	30	1.55 s	17.5, CH_3
15		134.4, C	31		156.6, C
16	1.93 m	39.2, CH_2	-NH-	7.50, t 5.2	



Compound 468 **1** (^{1}H NMR, DMSO-d_6 at 900 MHz)

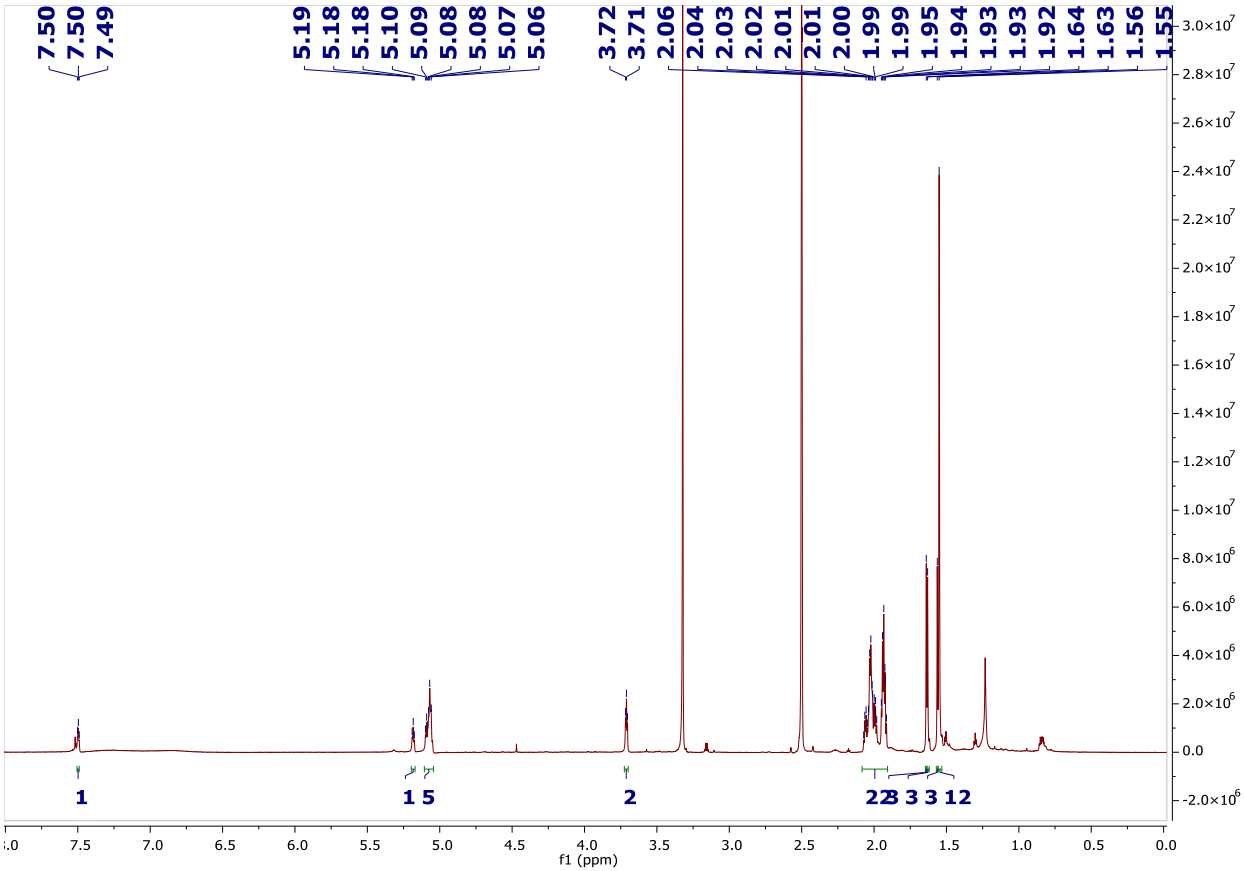
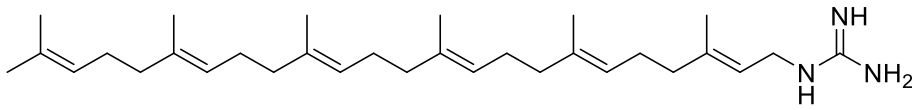


Figure F.1D-S1. ^{1}H NMR spectrum of Compound 468 **1**.



Compound 468 **1** (^{13}C NMR, DMSO-d_6 at 900 MHz)

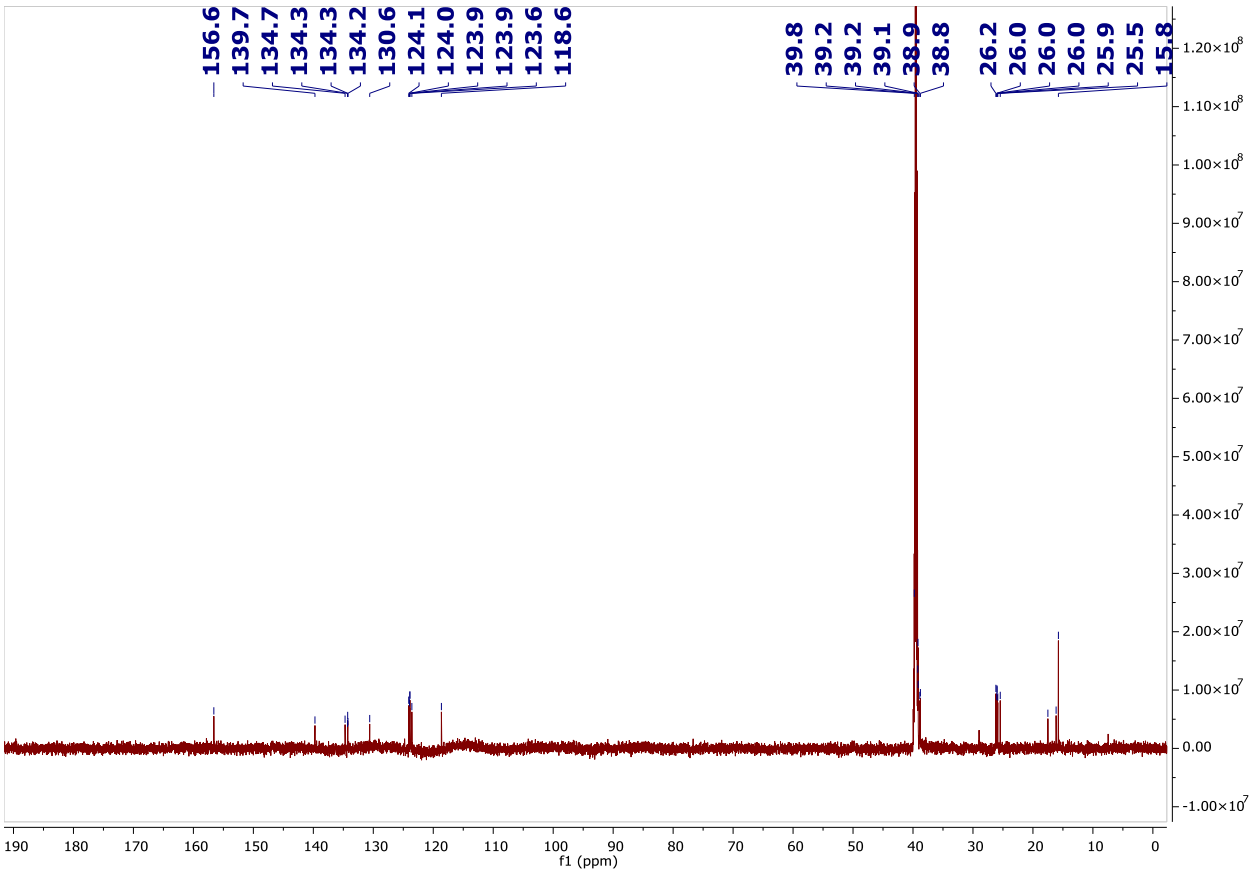
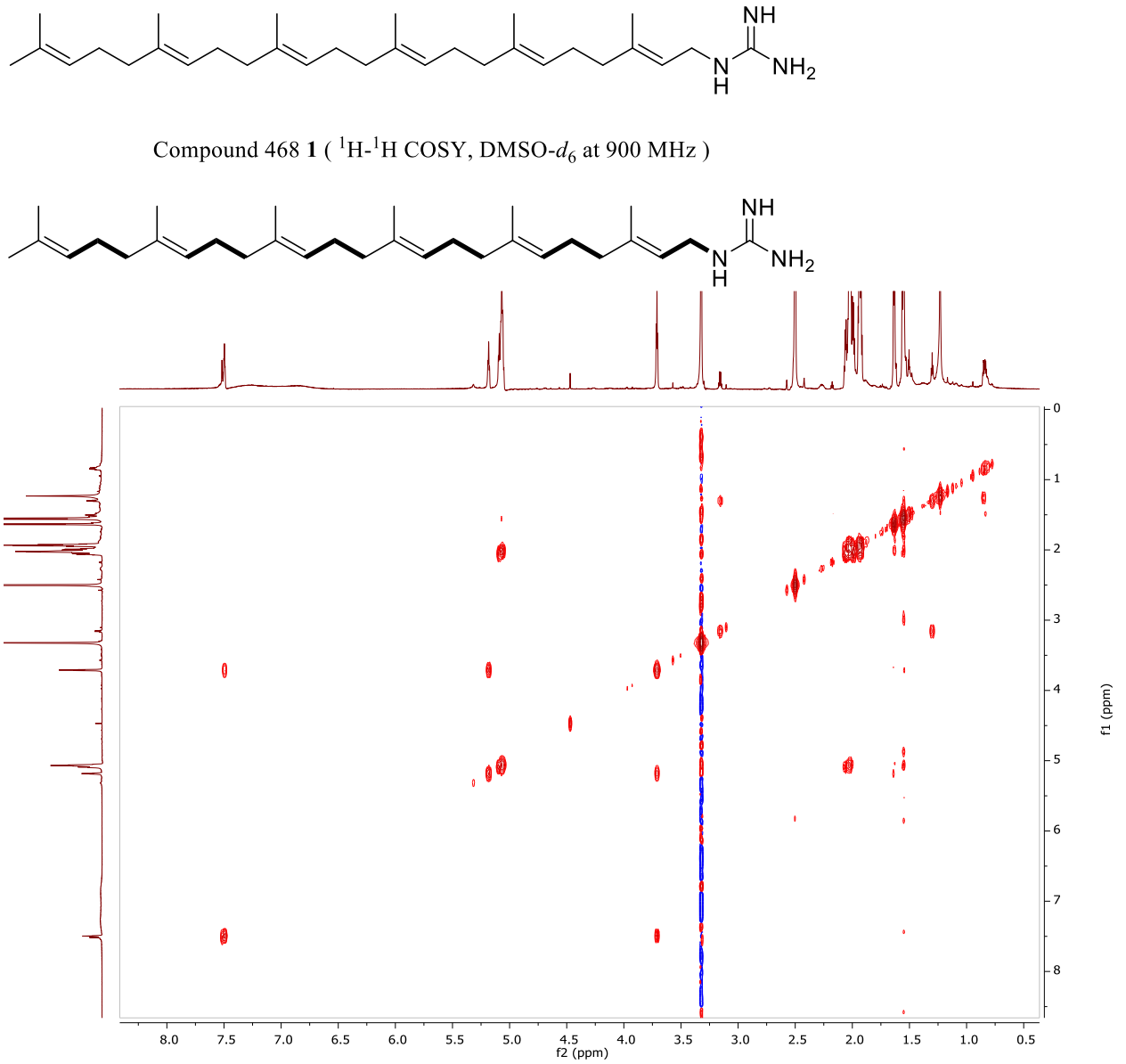


Figure F.1D-S2. ^{13}C NMR spectrum of Compound 468 **1**.



Compound 468 **1** (^{1}H-^{1}H COSY, DMSO-d_6 at 900 MHz)

Figure F.1D-S3. ^{1}H-^{1}H COSY spectrum of Compound 468 **1**.

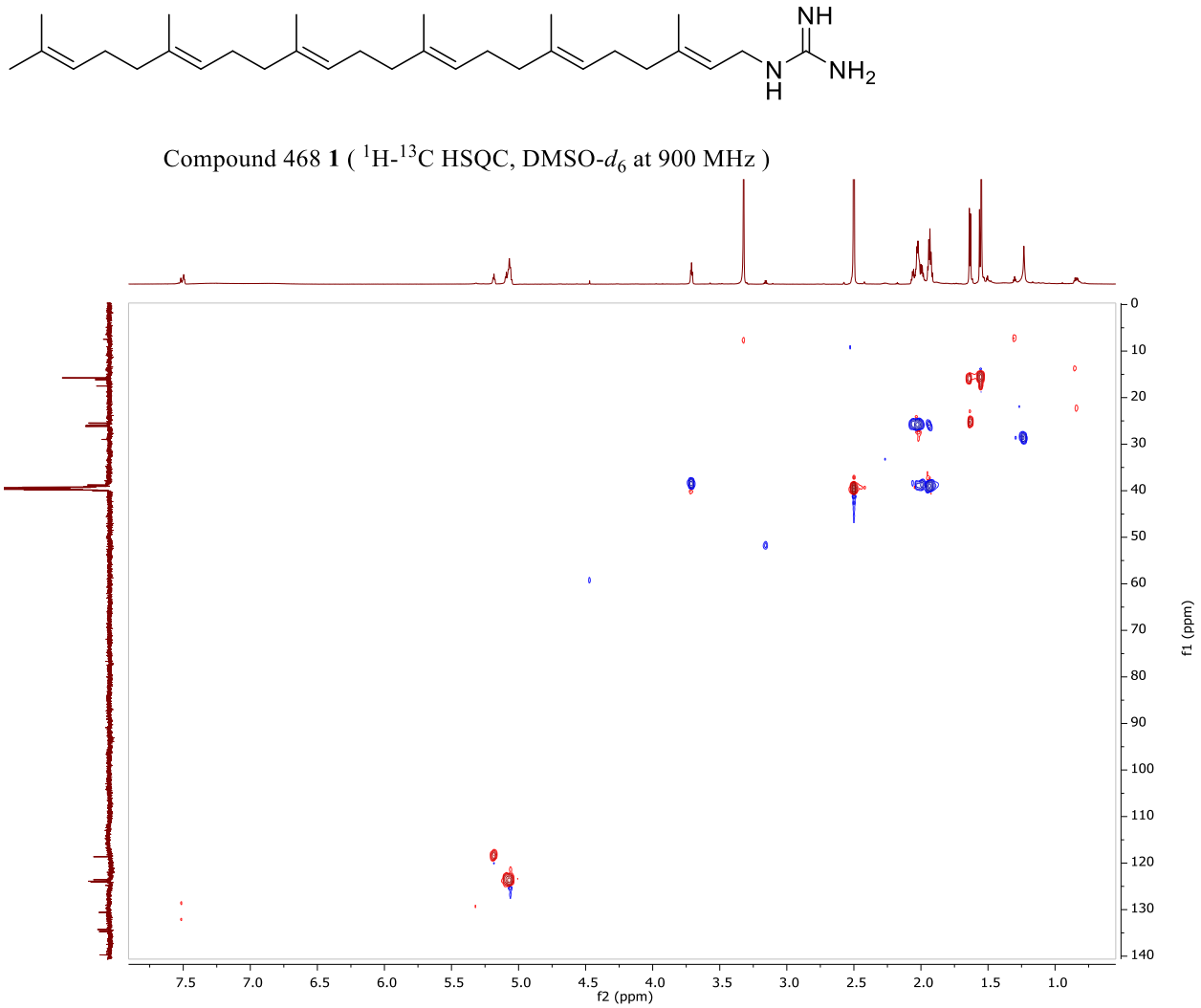
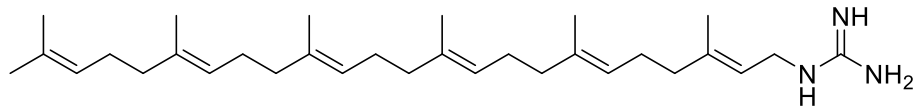


Figure F.1D-S4. ^{1}H-^{13}C HSQC spectrum of Compound 468 **1**.



Compound 468 **1** (^{1}H-^{1}H COSY, DMSO-d_6 at 900 MHz)

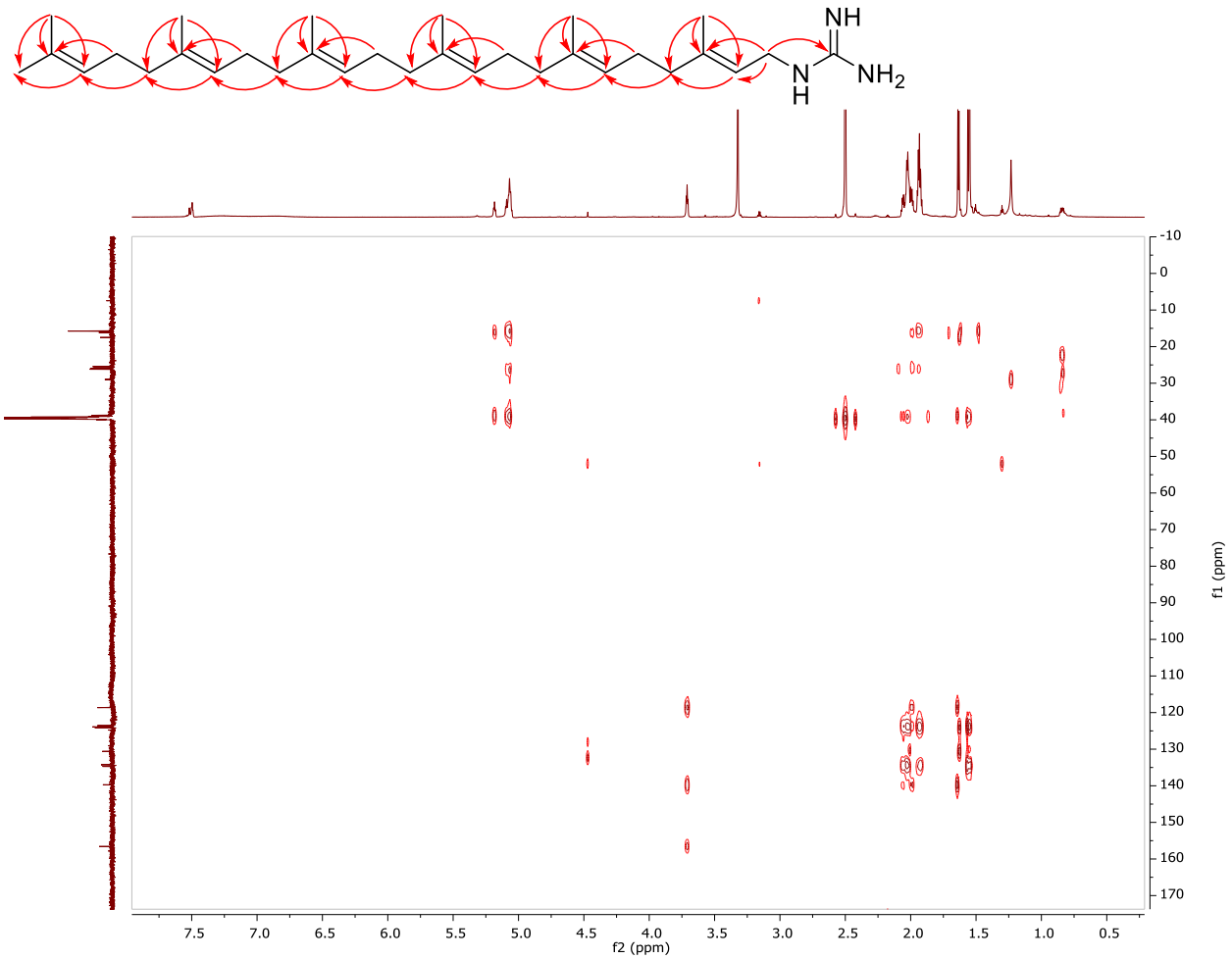
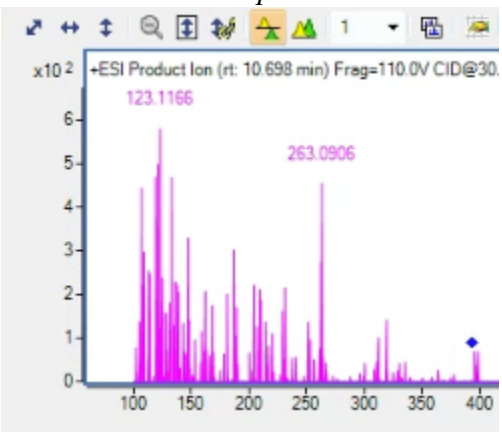


Figure F.1D-S5. ^{1}H-^{13}C HMBC spectrum of Compound 468 **1**.

F.2A Bra393 MS2 Spectra

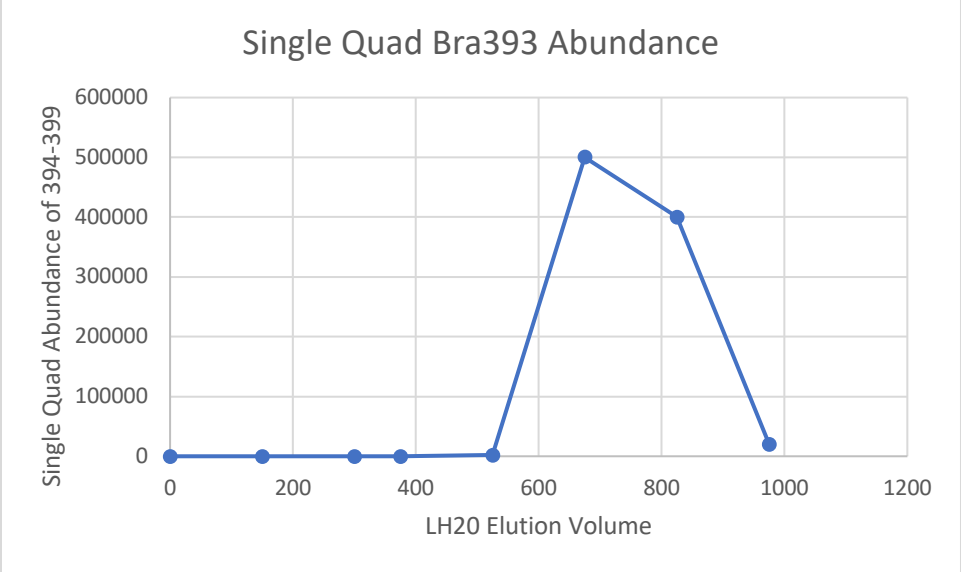


F.2B Bra393 UV and HPLC Purification

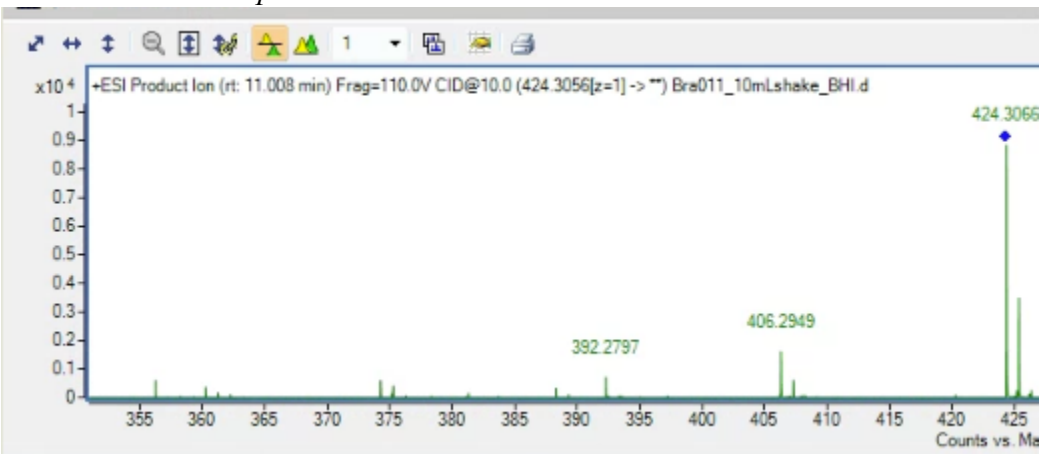
Bra393 is UV silent

Bra393 was only partially purified for isotopic feeding using LH20

F.2C Bra393 LH20 Purification



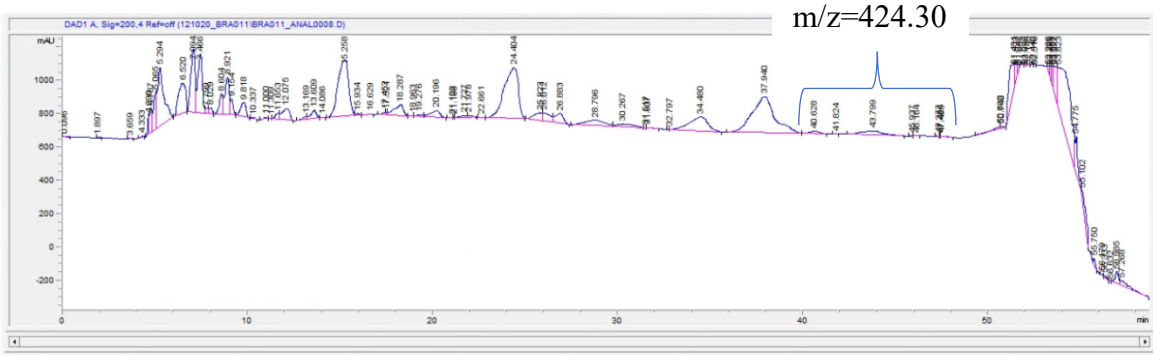
F.3A Bra423 MS2 Spectra



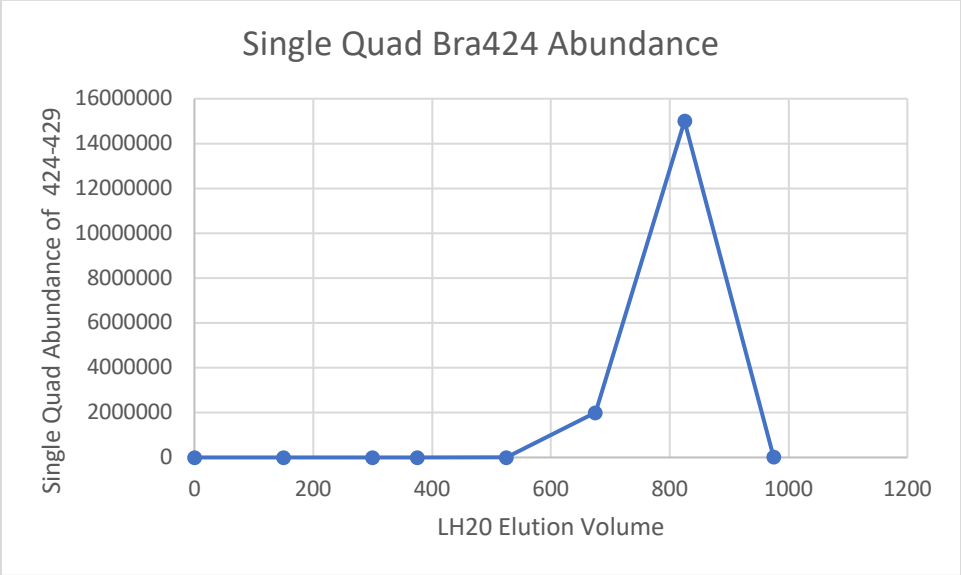
F.3B Bra423 UV and HPLC Purification

Bra423 is UV silent

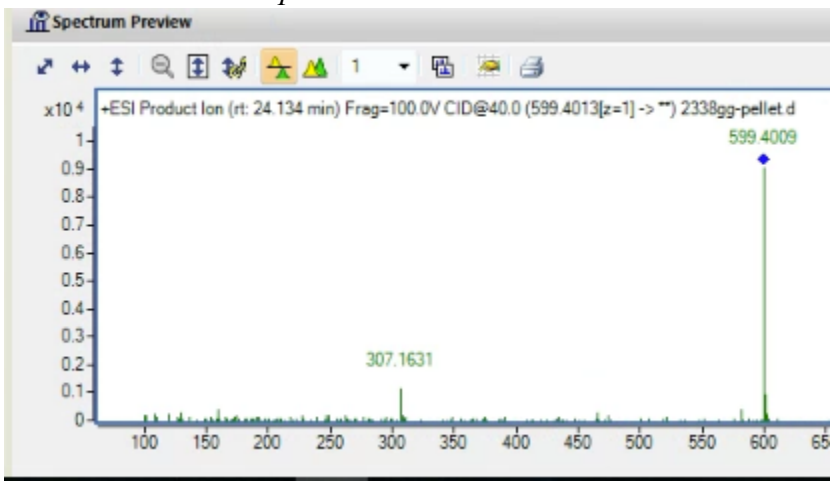
H2O:ACN
80:20 Hold 0-47 min
0:100 Hold 47-60mins
m/z=424.30



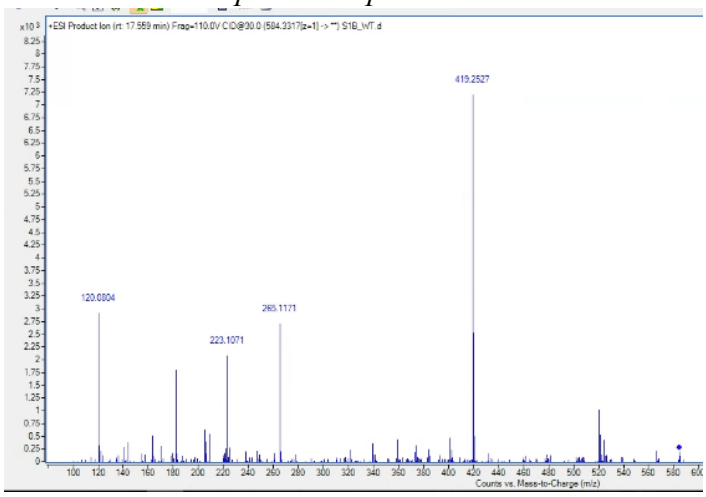
F.3C Bra423 LH20 Purification



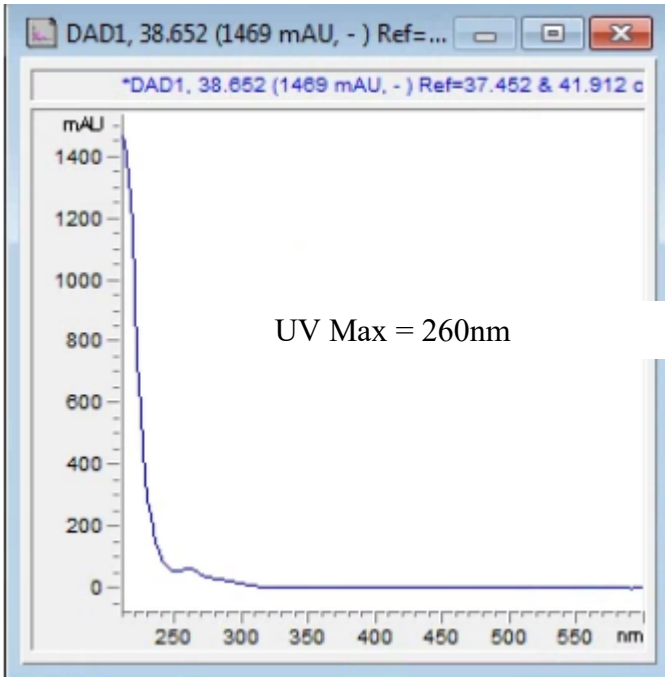
F.4A Diter599 MS2 Spectra



F.5A Mutanoclumpin MS2 Spectra



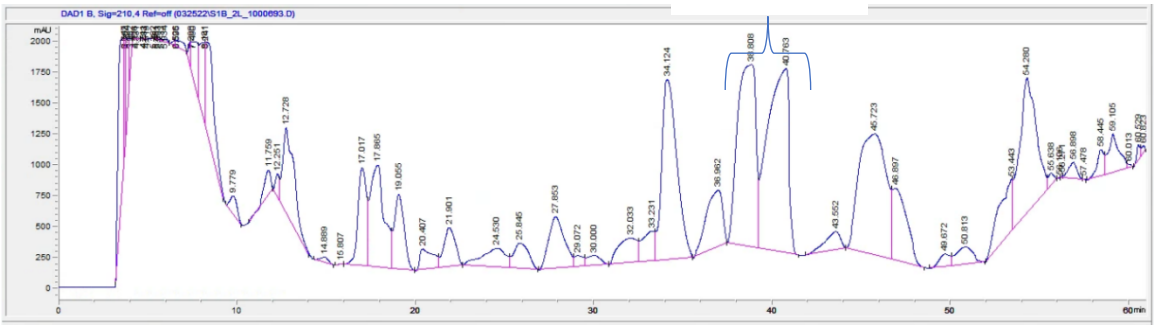
F.5B Mutanoclumpin UV and HPLC Purification



UV Max = 260nm

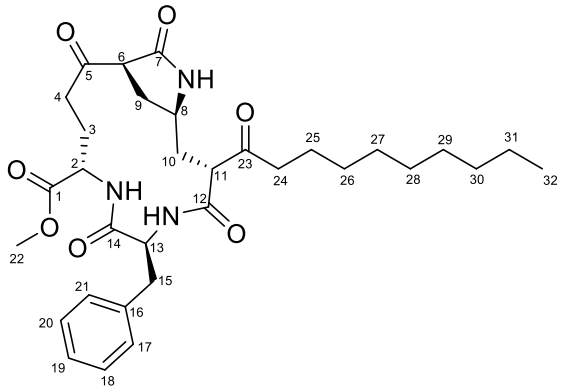
H2O:ACN
50:50->40:60 0-40 min
40:60 Hold 40-50mins
0:100 50-63 mins

m/z=584.33



F.5C Mutanoclumpin NMR

Structural Elucidation of Compound 583. Compound 583 was isolated as an amorphous yellow solid. It had the molecular formula $C_{32}H_{45}N_3O_7$ based on a proton adduct ion at *m/z* 584.3347 $[M+H]^+$ and a sodium adduct ion at *m/z* 606.3152 $[M+Na]^+$ in its positive ion HRESIMS spectrum. Its ^{1}H NMR, COSY, and HSQC spectra displayed signals for one phenylalanine residue (δ_H 9.29, -CO-NH-12; δ_H 4.74, H-13; δ_H 2.97, H-15a; δ_H 2.81, H-15b; δ_H 7.17 – 7.24, 5H, H-17/18/19/20/21), one glutamic residue (δ_H 8.97, -CO-NH-2; δ_H 4.24, H-2; δ_H 2.17, H-3a; δ_H 1.66, H-3b; δ_H 2.54, H-4a; δ_H 2.37, H-4b), one methoxy group (δ_H 3.59, H_3-22), one long chain moiety (δ_H 2.21, H-24a; δ_H 1.94, H-24b; δ_H 1.23 – 1.29, H_2-25/26/27/28/29/30/31; δ_H 0.86, H_3-32), one amide group (δ_H 8.27, -CO-NH-7), two methylene groups (δ_H 1.98, H-9a; δ_H 1.84, H-9b; δ_H 2.24, H-10a; δ_H 1.28, H-10b), and three methine groups (δ_H 3.07, H-6; δ_H 3.48, H-8; δ_H 3.59, H-11). Furthermore, the ^{13}C NMR spectrum showed signals for two ketones (δ_C 205.1, C-5; δ_C 204.6, C-23) in compound 583. The presence of 2-pyrrolidone unit was indicated by their COSY correlations between H-6 and H-9a/b, and H-9a/b and H-8, as well as ^2J-HMBC correlations between -CO-NH-7 and C-7 (δ_C 172.4)/C-8 (δ_C 51.8) and ^3J-HMBC correlations between -CO-NH-7 and C-6 (δ_C 57.5)/C-9 (δ_C 31.0). Meanwhile, the macrocyclic skeleton was verified by their COSY correlations between H-8 and H-10a/b, and H-10a/b and H-11, besides ^2J-HMBC correlations between H-4/6 and C-5, H-11 and C-12 (δ_C 167.5), and H-13 and C-14 (δ_C 171.4) and ^3J-HMBC correlations between H-6 and C-4 (δ_C 34.0), H-13 and C-12, and H-15 and C-14. The side chain 1-decanone was connected to C-11 through ^2J-HMBC correlations between H-11/24 and C-23. Additionally, the methoxy group was linked to the glutamic residue, based on ^3J-HMBC correlation between H_3-22 and C-1 (δ_C 171.9). Finally, the absolute configurations of C-2 and C-13 are both assigned as *S* due to these two residues are from the natural amino acids, and the remaining chiral centers are assigned as 6*S*, 8*R*, 11*R* in compound 583 **1a** and 6*R*, 8*R*, 11*R* in compound 583 **1b** via comparison between the experimental and computational NMR data.



Compound 583 **1a**

Table F.5C-S1. The NMR Data of Compound 583 **1a** in DMSO-d_6					
Position	δ_H (*J* in Hz)	δ_C (C type)	Position	δ_H (*J* in Hz)	δ_C (C type)
1		171.9, C	16		137.3, C
2	4.24 td 8.1, 4.1	49.8, CH	17/21	7.22 m	129.2, CH

3	2.17 m	22.9, CH_2	18/20	7.24 m	128.1, CH
	1.66 m		19	7.17 t 7.3	126.3, CH
4	2.54 m	34.0, CH_2	22	3.59 s	52.1, CH_3
	2.37 dt 18.6, 4.3		23		204.6, C
5		205.1, C	24	2.21 m	39.9, CH_2
6	3.07 d 10.6	57.5, CH		1.94 m	
7		172.4, C	25	1.29 m	22.9, CH_2
8	3.48 m	51.8, CH	26	1.23 m	28.7, CH_2
9	1.98 dd 13.1, 6.4	31.0, CH_2	27	1.23 m	28.8, CH_2
	1.84 m		28	1.23 m	28.9, CH_2
10	2.24 m	32.1, CH_2	29	1.23 m	28.9, CH_2
	1.28 m		30	1.23 m	31.4, CH_2
11	3.59 m	54.7, CH	31	1.26 m	22.2, CH_2
12		167.5, C	32	0.86 t 6.8	14.0, CH_3
13	4.74 td 9.2, 6.4	53.2, CH	2-NH-	8.97 br	
14		171.4, C	7-NH-	8.27 s	
15	2.97 dd 13.9, 6.4	37.6, CH_2	12-NH-	9.29 br	
	2.81 dd 13.9, 9.2				

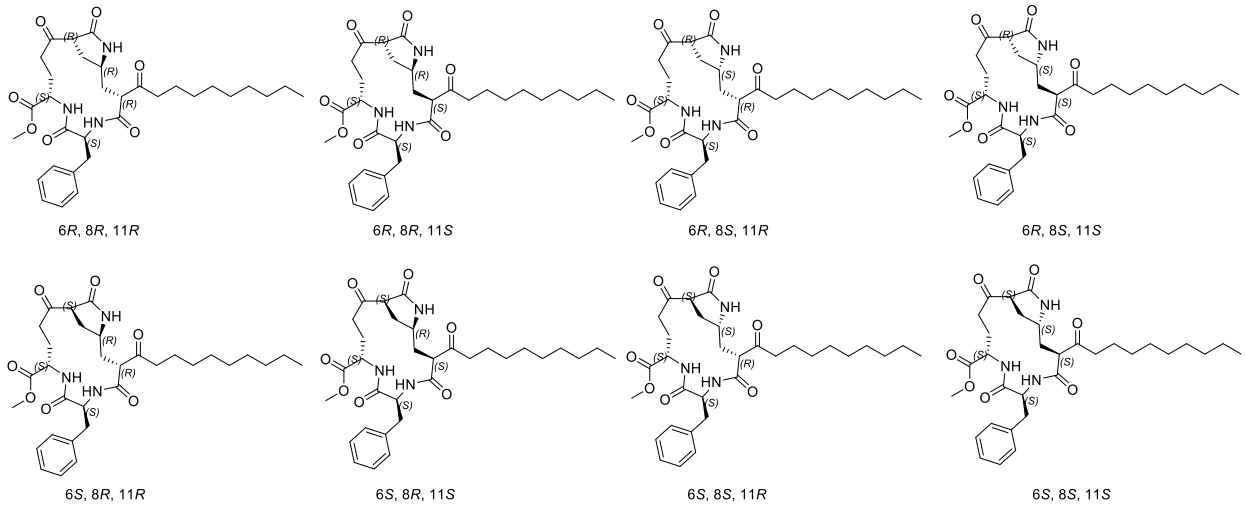
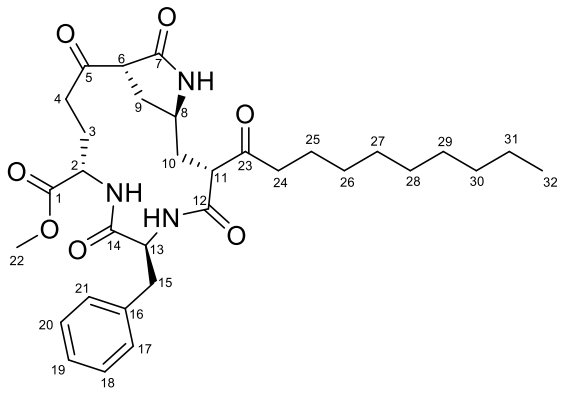


Table F.5C-S2. The Predicted ^{1}H NMR Data of Compound 583 **1a** in DMSO-d_6									
Position	**1a**	6*R*, 8*R*, 11*R*	6*R*, 8*R*, 11*S*	6*R*, 8*S*, 11*R*	6*R*, 8*S*, 11*S*	6*S*, 8*R*, 11*R*	6*S*, 8*R*, 11*S*	6*S*, 8*S*, 11*R*	6*S*, 8*S*, 11*S*
2	4.24	4.32	4.41	4.36	5.18	4.00	4.04	4.54	4.44
3a	2.17	2.42	2.27	2.46	2.71	2.33	2.32	2.22	2.54
3b	1.66	1.59	1.56	2.01	2.17	1.83	1.72	1.73	2.25
4a	2.54	2.67	2.91	3.12	2.91	3.08	3.57	2.41	3.01
4b	2.37	2.62	2.35	3.07	2.19	2.89	2.84	2.33	2.40
6	3.07	3.01	3.05	3.05	3.52	3.17	3.26	3.60	2.78
8	3.48	3.75	4.19	3.47	4.40	3.94	4.07	3.55	4.06
9a	1.98	2.05	2.39	2.94	2.33	2.11	2.36	2.36	2.01
9b	1.84	2.03	1.97	2.16	2.06	1.50	2.32	1.94	1.97
10a	2.24	1.85	2.09	2.69	2.23	2.57	2.49	2.11	2.27
10b	1.28	1.80	1.88	0.74	2.11	1.82	1.73	2.10	1.88
11	3.59	3.77	3.52	3.26	3.04	3.39	3.25	3.05	3.52
13	4.74	4.62	4.49	4.53	3.86	4.71	3.74	4.53	4.64
15a	2.97	3.18	3.39	3.33	3.56	3.13	3.36	3.37	3.30
15b	2.81	2.89	2.95	2.88	3.30	2.46	3.40	2.96	2.86
17	7.22	7.58	7.53	7.66	7.60	7.57	7.55	7.64	7.47
18	7.24	7.66	7.61	7.71	7.67	7.60	7.65	7.67	7.59
19	7.17	7.61	7.56	7.64	7.58	7.57	7.57	7.60	7.56
20	7.24	7.67	7.64	7.69	7.65	7.61	7.62	7.65	7.62
21	7.22	7.62	7.62	7.64	7.58	7.48	7.50	7.58	7.51
24a	2.21	2.69	2.73	2.53	2.95	2.54	2.77	2.26	2.89
24b	1.94	2.45	2.53	2.21	2.77	2.42	2.75	2.07	2.75
MAE		0.26	0.29	0.35	0.49	0.30	0.42	0.27	0.30
RMSE		0.31	0.36	0.43	0.57	0.34	0.50	0.35	0.39

Table F.5C-S3. The Predicted ^{13}C NMR Data of Compound 583 **1a** in DMSO-d_6

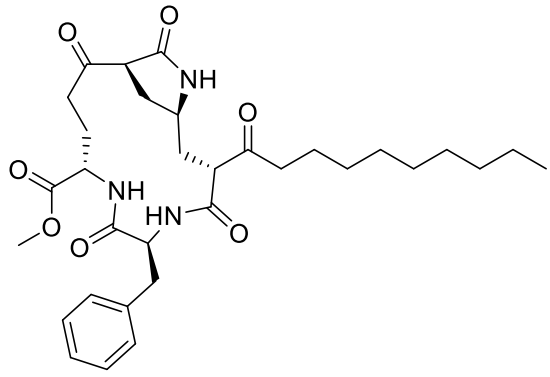
Position	**1a**	6*R*, 8*R*, 11*R*	6*R*, 8*R*, 11*S*	6*R*, 8*S*, 11*R*	6*R*, 8*S*, 11*S*	6*S*, 8*R*, 11*R*	6*S*, 8*R*, 11*S*	6*S*, 8*S*, 11*R*	6*S*, 8*S*, 11*S*
1	171.4	169.4	169.7	168.8	169.1	169.9	171.3	169.2	170.1
2	49.8	52.1	53.2	54.8	53.8	55.3	53.2	53.1	52.3
3	22.9	29.3	32.4	24.6	27.4	24.9	23.2	29.6	24.1
4	34.0	35.6	36.6	36.0	43.0	36.9	38.3	39.1	33.8
5	205.1	207.3	206.0	211.6	217.8	205.8	209.6	207.3	209.4
6	57.5	59.9	59.2	55.1	53.2	56.4	56.6	46.4	57.7
7	172.4	165.4	166.0	165.9	166.1	168.3	164.6	165.8	168.4
8	51.8	51.3	52.6	51.8	50.7	52.7	50.6	52.6	51.6
9	31.0	34.6	31.1	24.6	30.0	27.5	27.6	31.4	35.6
10	32.1	32.6	38.4	35.1	34.6	31.2	34.5	33.0	35.8
11	54.7	56.3	58.2	57.0	56.7	56.2	57.8	59.3	57.3
12	167.5	165.8	164.9	164.4	167.1	163.8	166.8	167.3	163.9
13	53.2	56.1	57.8	56.8	61.2	60.9	60.6	57.9	56.6
14	171.4	166.9	167.7	166.3	166.4	168.4	166.7	167.2	166.4
15	37.6	36.9	34.9	38.1	33.3	36.4	32.8	35.5	36.3
16	137.3	136.0	137.0	136.3	138.5	135.9	138.4	136.8	136.9
17	129.2	125.7	126.1	125.9	125.4	125.3	125.5	125.3	125.8
18	128.1	125.2	125.0	125.2	125.3	125.3	125.2	125.4	125.0
19	126.3	123.2	123.0	123.3	122.8	123.5	122.8	123.2	123.0
20	128.1	125.2	125.2	125.4	125.2	125.4	125.1	125.3	125.0
21	129.2	125.9	124.9	125.5	125.8	125.2	125.8	125.8	125.5
23	204.6	211.9	211.7	205.0	207.7	205.4	207.1	204.3	219.9
24	39.9	42.1	44.2	39.8	40.8	38.6	40.4	38.2	43.5
MAE		2.88	3.43	2.95	3.86	2.61	3.03	3.19	3.22
RMSE		3.42	4.08	3.51	4.79	3.11	3.63	4.06	4.37



Compound 583 **1b**

Table F.5C-S4. The NMR Data of Compound 583 **1b** in DMSO-d_6					
Position	δ_H (J in Hz)	δ_C (C type)	Position	δ_H (J in Hz)	δ_C (C type)
1		171.8, C	16		137.1, C
2	4.31 ddd 12.3, 8.4, 3.5	49.4, CH	17/21	7.21 d 7.0	128.9, CH
3	2.17 m	23.2, CH_2	18/20	7.24 m	127.9, CH
	1.61 tt 14.1, 3.5		19	7.18 t 7.0	126.1, CH
4	2.53 m	33.7, CH_2	22	3.60 s	51.7, CH_3
	2.37 dt 18.5, 3.5		23		204.8, C
5		n.d.	24	2.21 m	40.0, CH_2
6	3.07 dd 9.7, 2.2	57.1, CH		1.97 dt 17.6, 7.3	
7		n.d.	25	1.31 m	22.8, CH_2
8	3.49 m	51.5, CH	26	1.23 m	28.6, CH_2
9	1.88 m	30.7, CH_2	27	1.23 m	28.6, CH_2
10	2.24 m	30.1, CH_2	28	1.23 m	28.6, CH_2
	1.29 m		29	1.23 m	28.6, CH_2
11	3.59 m	54.4, CH	30	1.24 m	31.1, CH_2
12		n.d.	31	1.27 m	22.2, CH_2
13	4.78 m	52.5, CH	32	0.85 t 7.0	13.8, CH_3
14		171.2, C	2-NH-	8.79 d 8.4	
15	2.98 dd 13.7, 6.8	37.4, CH_2	7-NH-	8.28 s	
	2.79 dd 13.7, 8.6		12-NH-	9.00 d 9.7	

Table F.5C-S5. The Predicted ^{1}H NMR Data of Compound 583 **1b** in DMSO-d_6									
Position	**1b**	6*R*, 8*R*, 11*R*	6*R*, 8*R*, 11*S*	6*R*, 8*S*, 11*R*	6*R*, 8*S*, 11*S*	6*S*, 8*R*, 11*R*	6*S*, 8*R*, 11*S*	6*S*, 8*S*, 11*R*	6*S*, 8*S*, 11*S*
2	4.31	4.32	4.41	4.36	5.18	4.00	4.04	4.54	4.44
3a	2.17	2.42	2.27	2.46	2.71	2.33	2.32	2.22	2.54
3b	1.61	1.59	1.56	2.01	2.17	1.83	1.72	1.73	2.25
4a	2.53	2.67	2.91	3.12	2.91	3.08	3.57	2.41	3.01
4b	2.37	2.62	2.35	3.07	2.19	2.89	2.84	2.33	2.40
6	3.07	3.01	3.05	3.05	3.52	3.17	3.26	3.60	2.78
8	3.49	3.75	4.19	3.47	4.40	3.94	4.07	3.55	4.06
9a	1.88	2.05	2.39	2.94	2.33	2.11	2.36	2.36	2.01
9b	1.88	2.03	1.97	2.16	2.06	1.50	2.32	1.94	1.97
10a	2.24	1.85	2.09	2.69	2.23	2.57	2.49	2.11	2.27
10b	1.29	1.80	1.88	0.74	2.11	1.82	1.73	2.10	1.88
11	3.59	3.77	3.52	3.26	3.04	3.39	3.25	3.05	3.52
13	4.78	4.62	4.49	4.53	3.86	4.71	3.74	4.53	4.64
15a	2.98	3.18	3.39	3.33	3.56	3.13	3.36	3.37	3.30
15b	2.79	2.89	2.95	2.88	3.30	2.46	3.40	2.96	2.86
17	7.21	7.58	7.53	7.66	7.60	7.57	7.55	7.64	7.47
18	7.24	7.66	7.61	7.71	7.67	7.60	7.65	7.67	7.59
19	7.18	7.61	7.56	7.64	7.58	7.57	7.57	7.60	7.56
20	7.24	7.67	7.64	7.69	7.65	7.61	7.62	7.65	7.62
21	7.21	7.62	7.62	7.64	7.58	7.48	7.50	7.58	7.51
24a	2.21	2.69	2.73	2.53	2.95	2.54	2.77	2.26	2.89
24b	1.97	2.45	2.53	2.21	2.77	2.42	2.75	2.07	2.75
MAE		0.25	0.29	0.36	0.50	0.31	0.43	0.27	0.31
RMSE		0.31	0.36	0.44	0.57	0.35	0.51	0.35	0.39



Compound 583 **1a** (^{1}H NMR, DMSO-d_6 at 900 MHz)

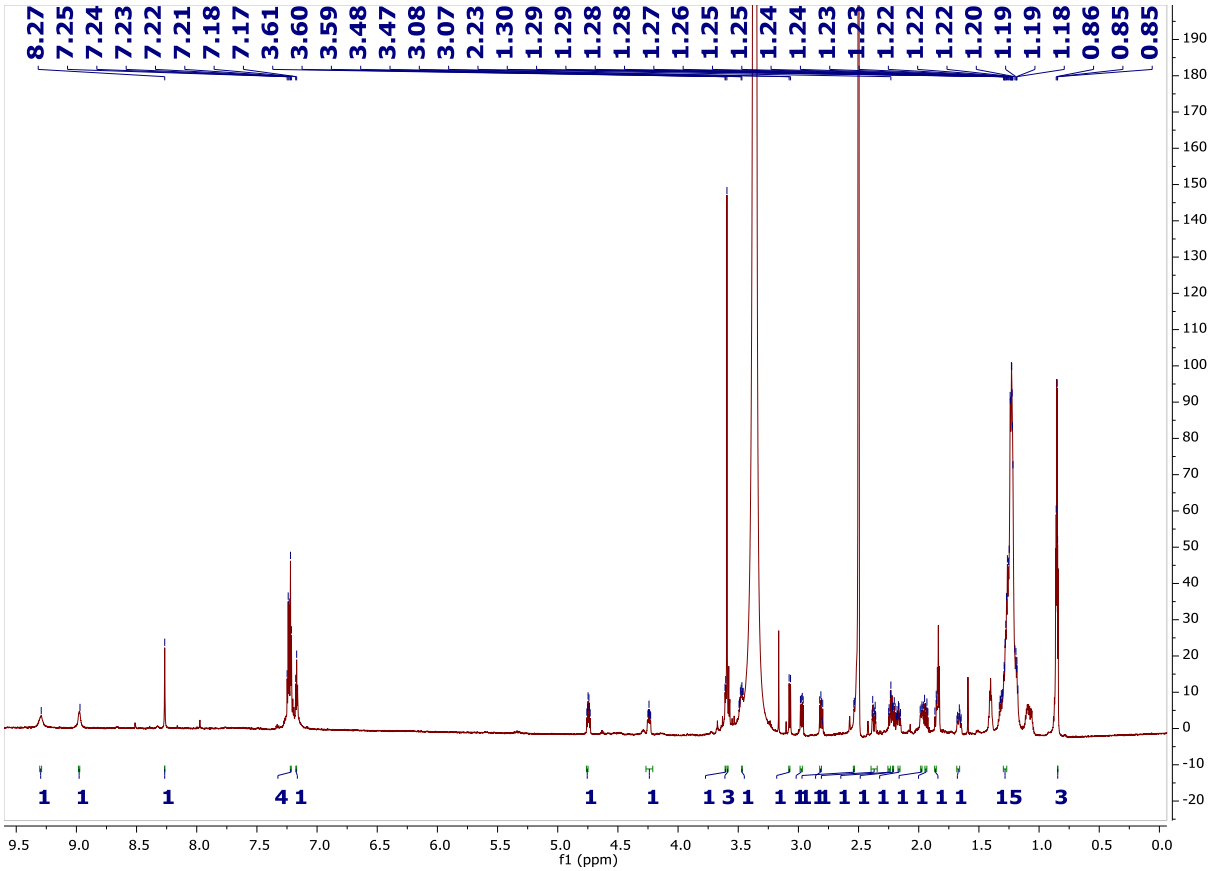
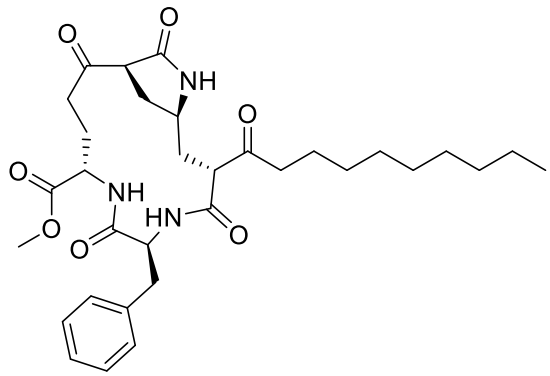


Figure F.5C-S1. ^{1}H NMR spectrum of Compound 583 **1a**.



Compound 583 **1a** (^{13}C NMR, DMSO-d_6 at 900 MHz)

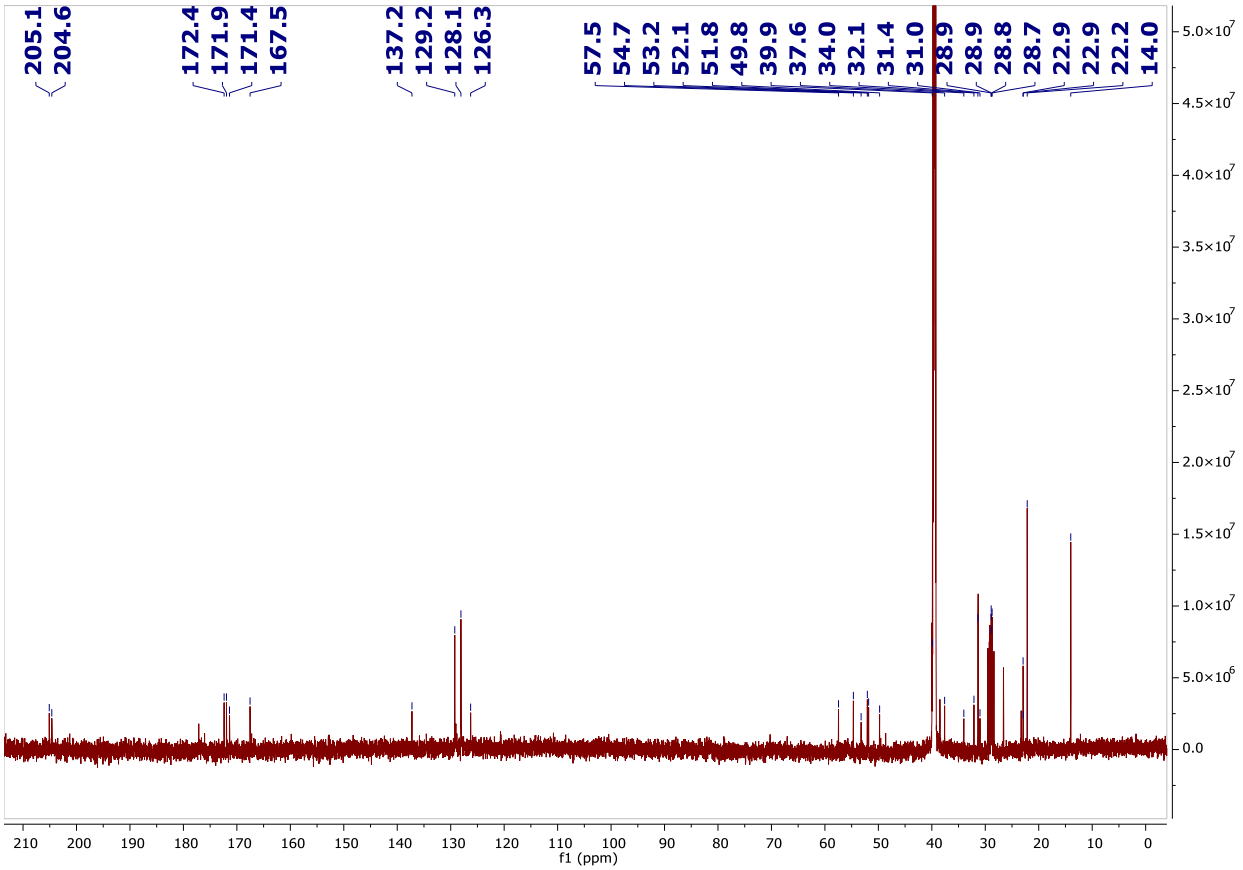
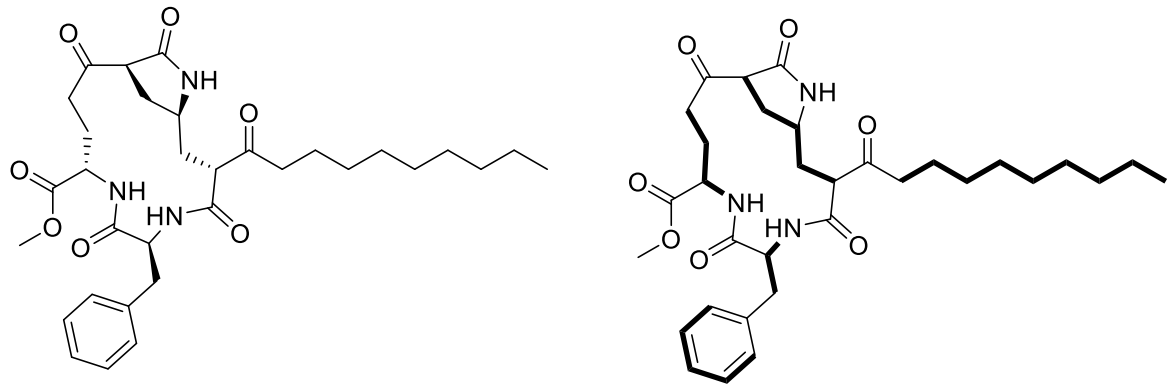


Figure F.5C-S2. ^{13}C NMR spectrum of Compound 583 **1a**.



Compound 583 **1a** (^{1}H-^{1}H COSY, DMSO-d_6 at 900 MHz)

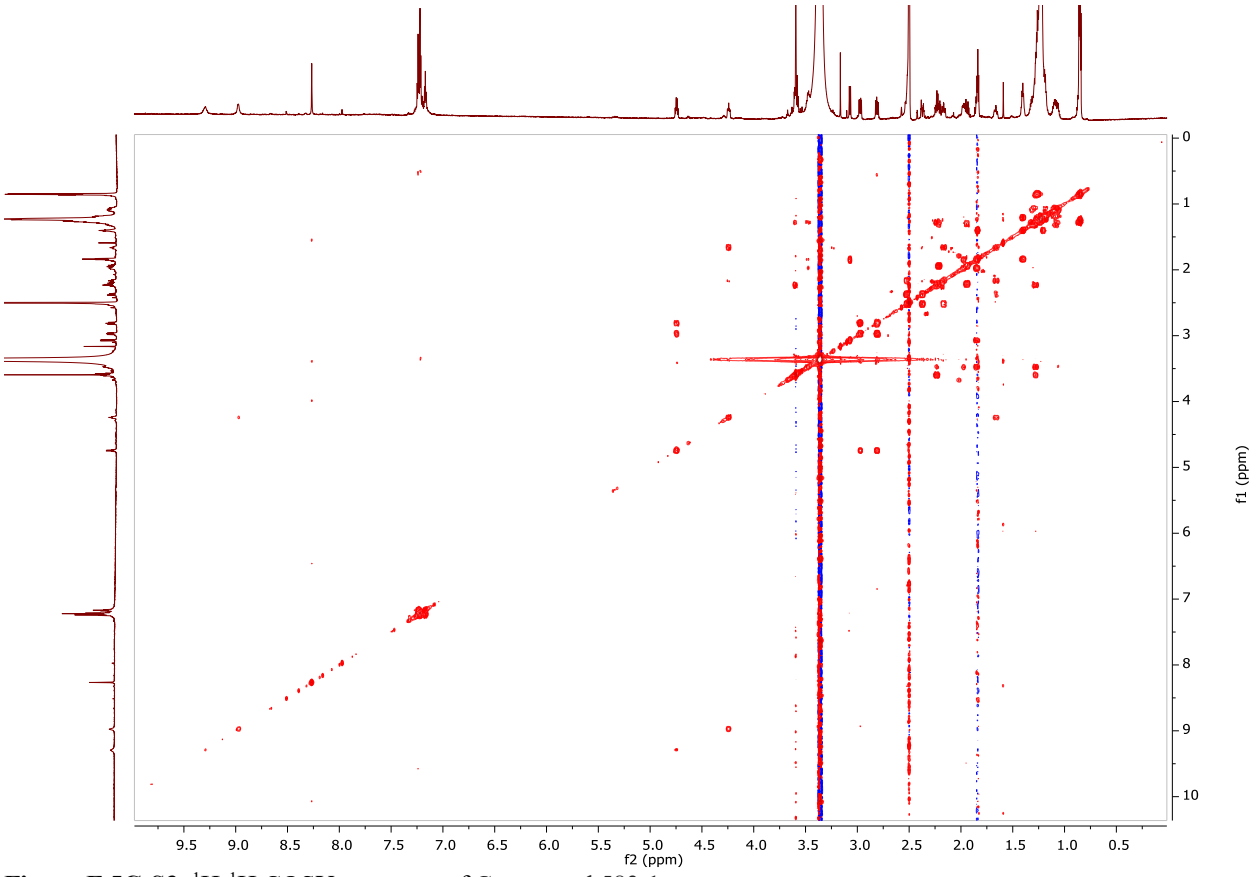
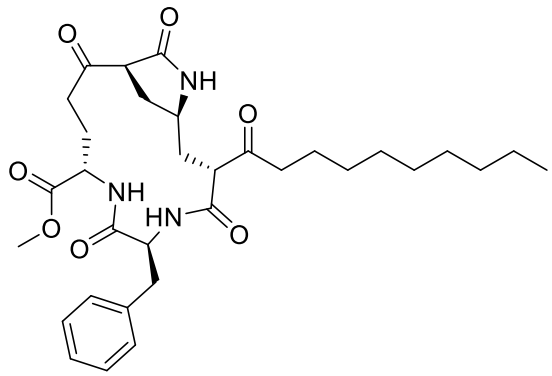


Figure F.5C-S3. ^{1}H-^{1}H COSY spectrum of Compound 583 **1a**.



Compound 583 **1a** (^{1}H-^{13}C HSQC, DMSO-d_6 at 900 MHz)

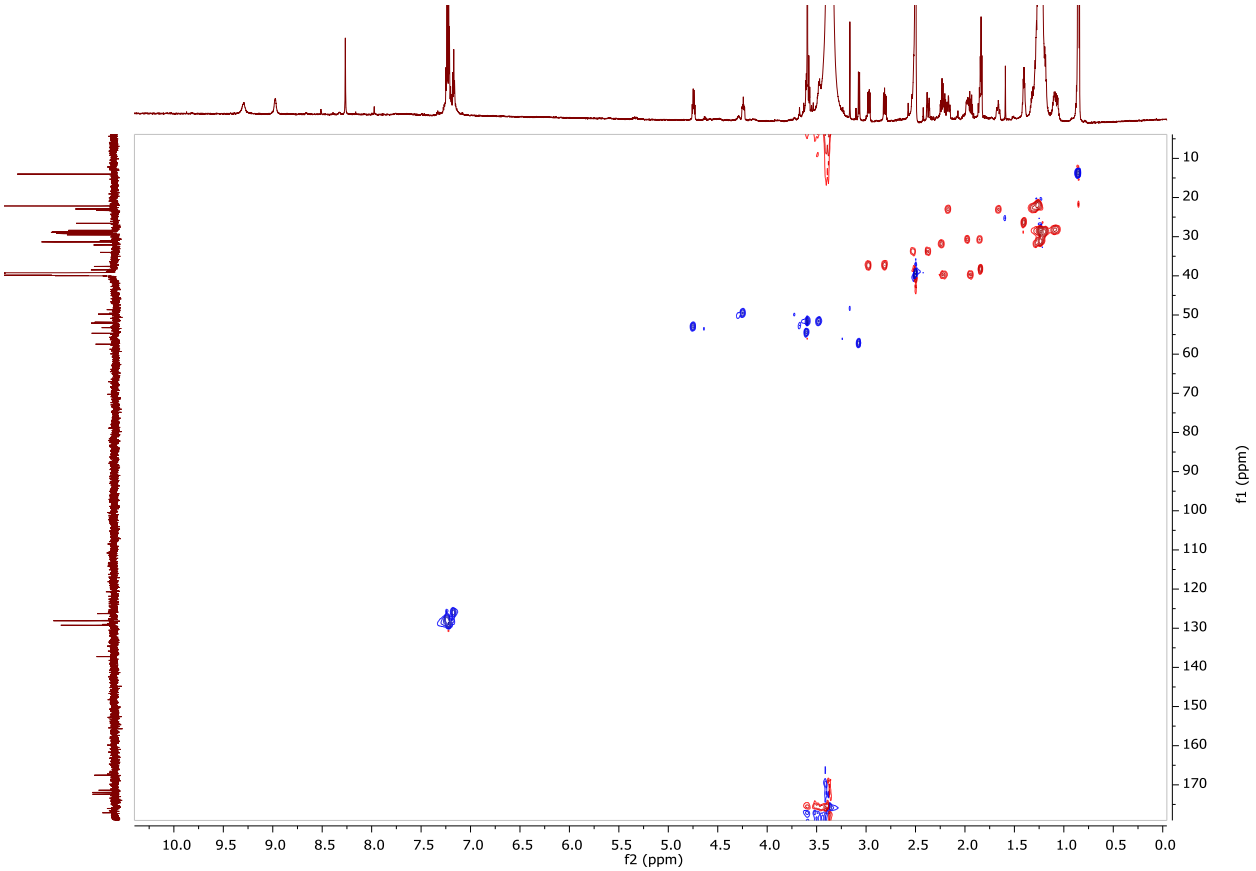


Figure F.5C-S4. ^{1}H-^{13}C HSQC spectrum of Compound 583 **1a**.

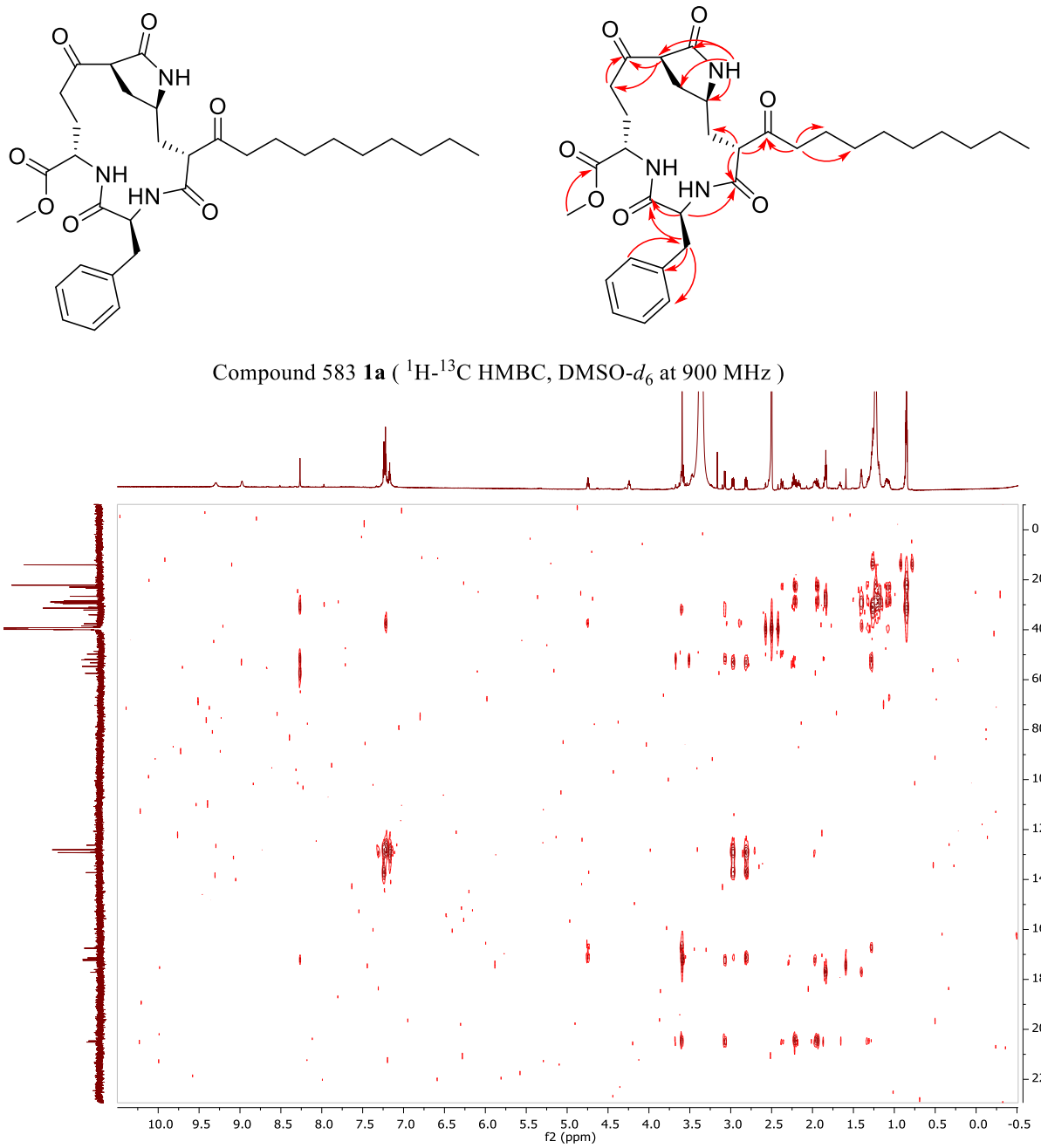
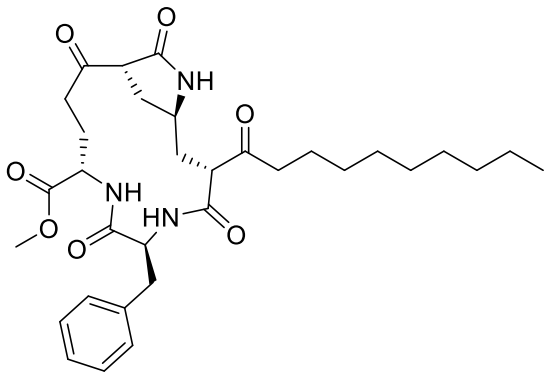


Figure S5. ^{1}H-^{13}C HMBC spectrum of Compound 583 **1a**.



Compound 583 **1b** (^{1}H NMR, DMSO-d_6 at 900 MHz)

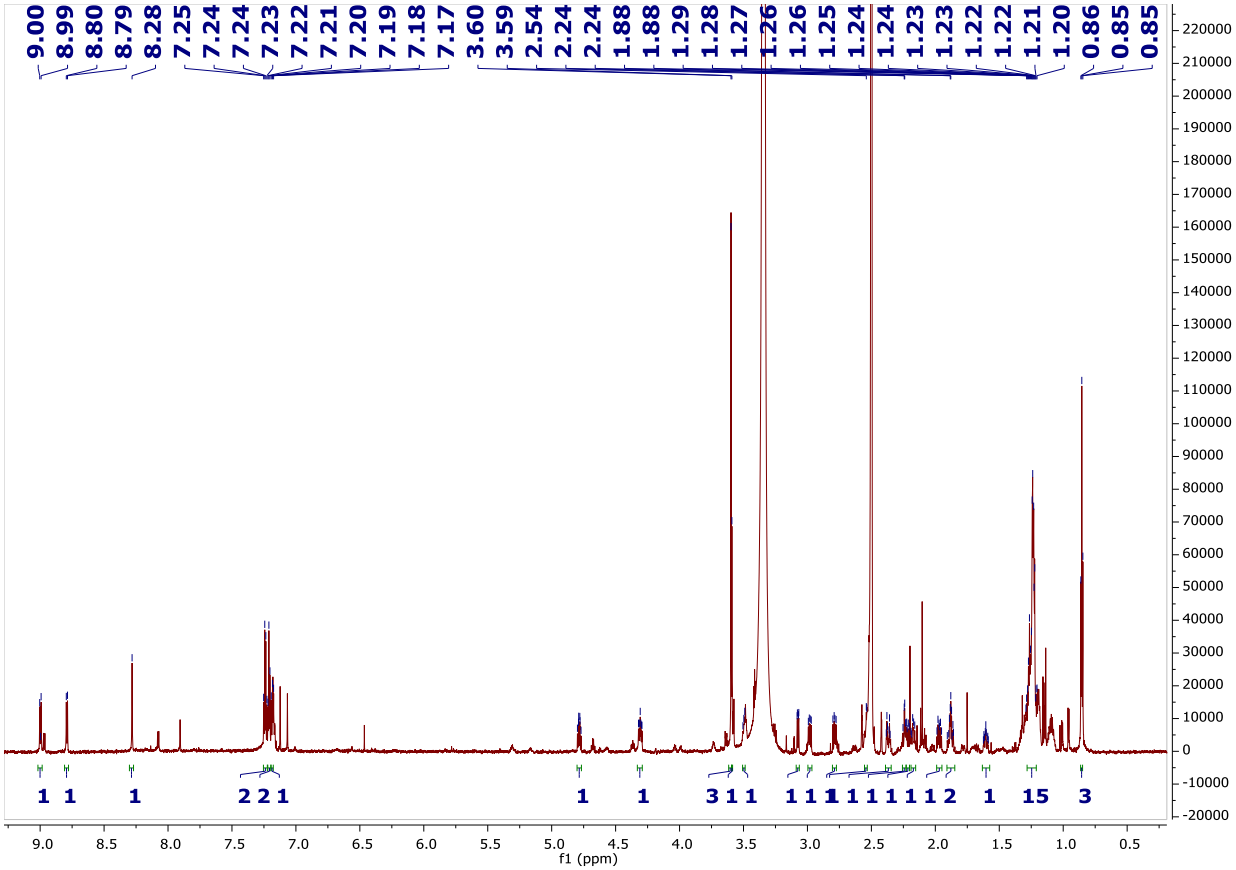
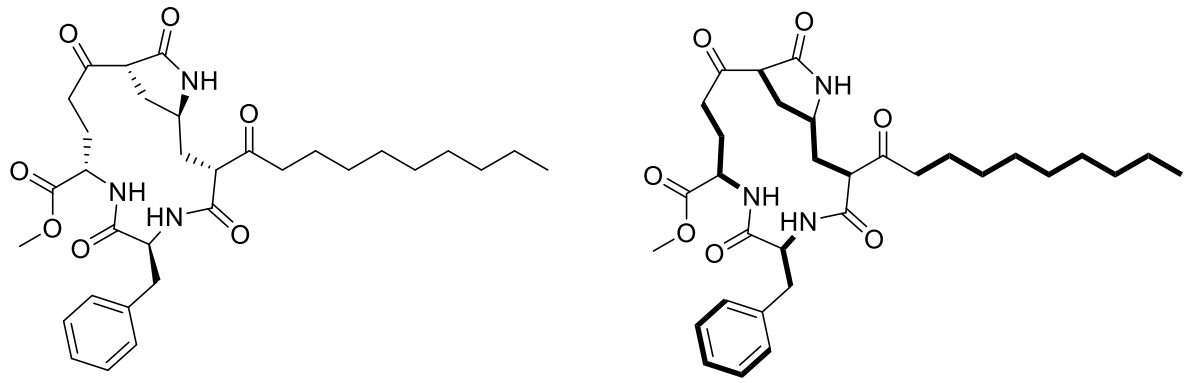


Figure F.5C-S6. ^{1}H NMR spectrum of Compound 583 **1b**.



Compound 583 **1b** (^{1}H-^{1}H COSY, DMSO-d_6 at 900 MHz)

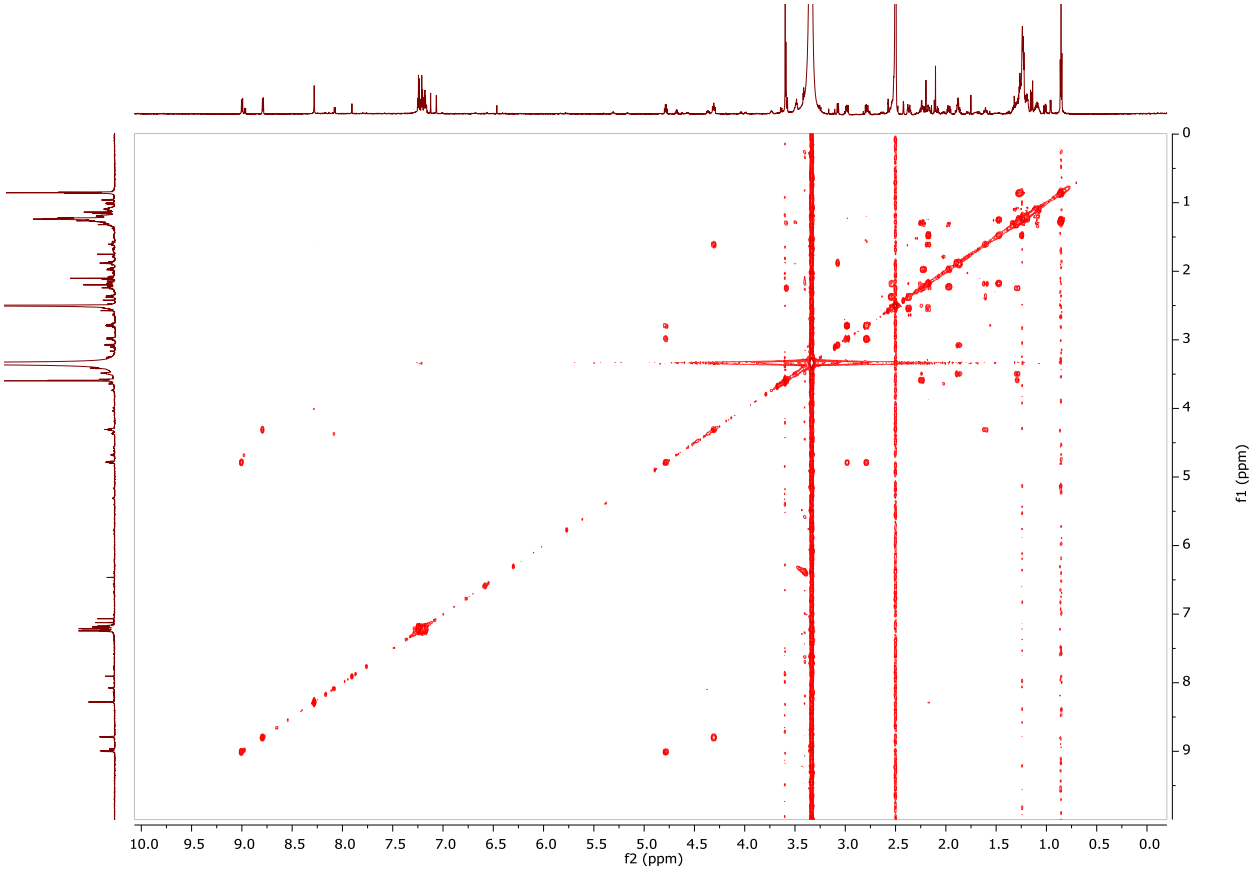
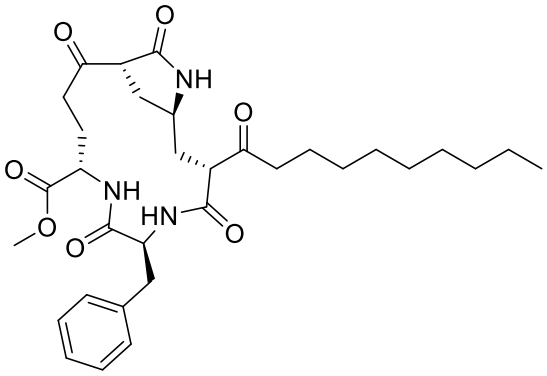


Figure F.5C-S7. ^{1}H-^{1}H COSY spectrum of Compound 583 **1b**.



Compound 583 **1b** (^{1}H-^{13}C HSQC, DMSO-d_6 at 900 MHz)

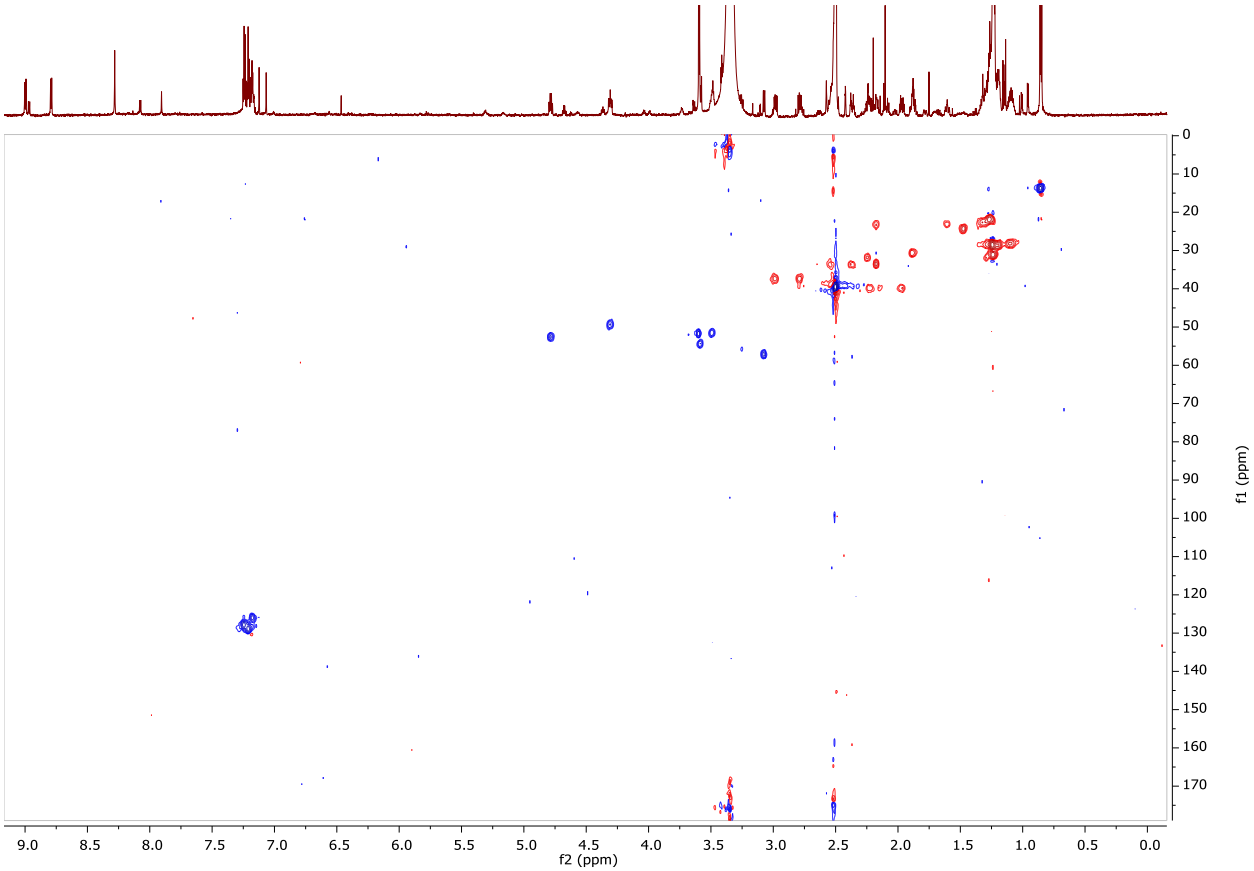


Figure F.5C-S8. ^{1}H-^{13}C HSQC spectrum of Compound 583 **1b**.

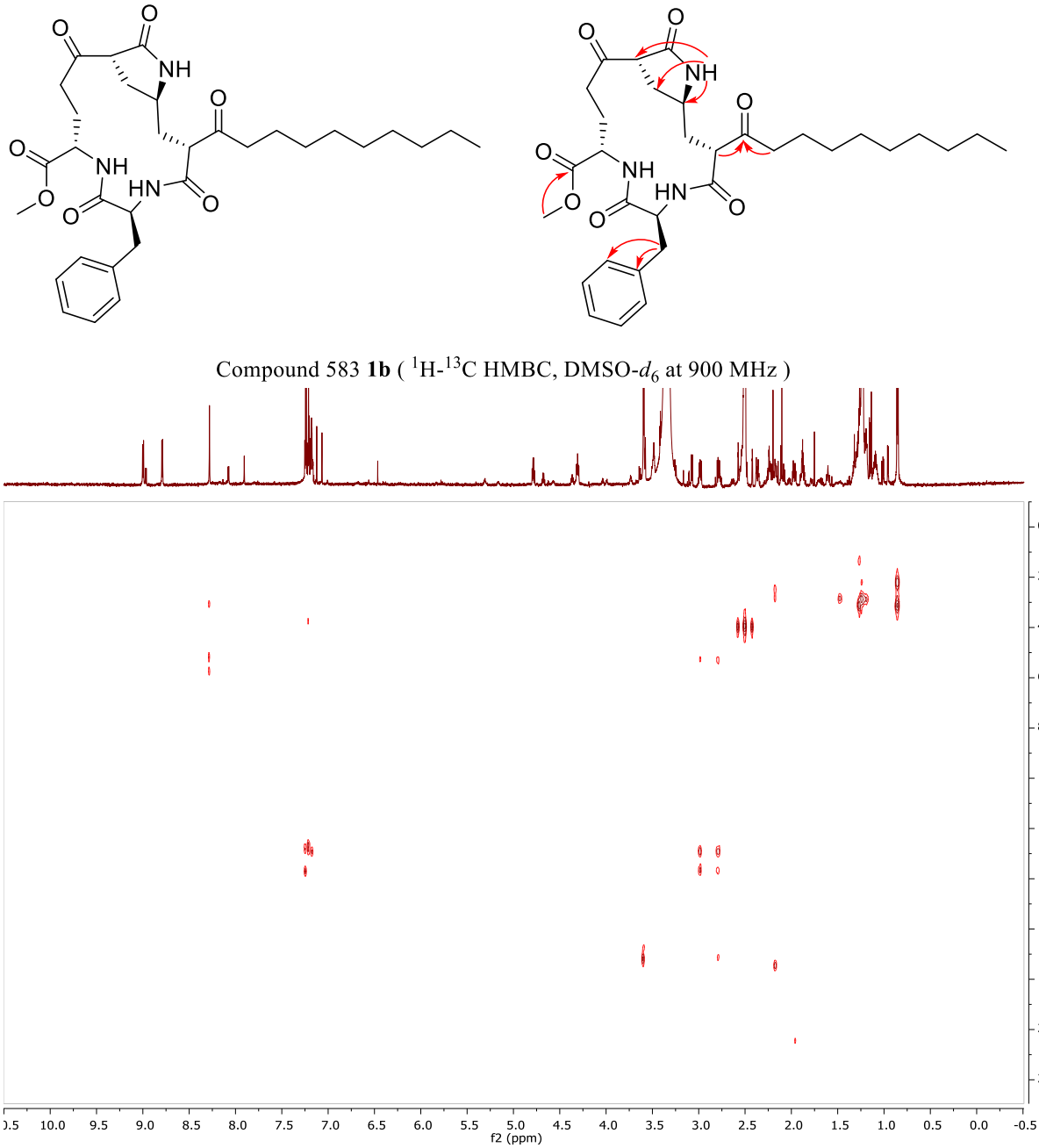
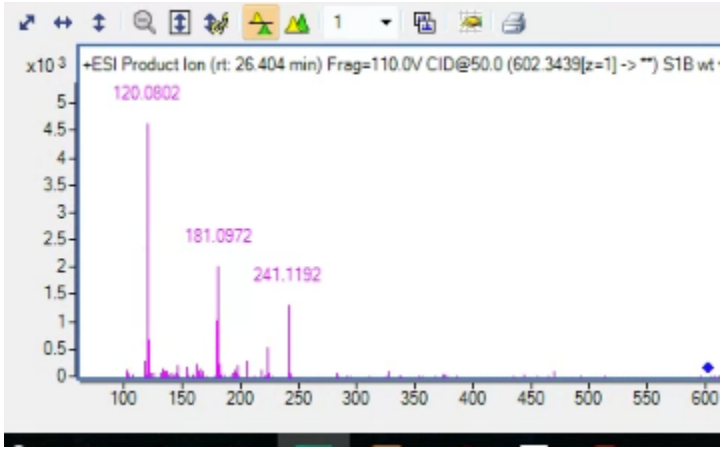
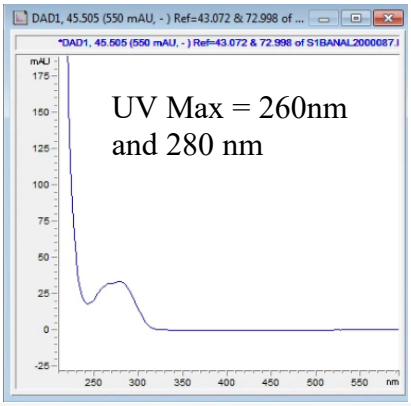


Figure F.5C-S9. ^{1}H-^{13}C HMBC spectrum of Compound 583 **1b**.

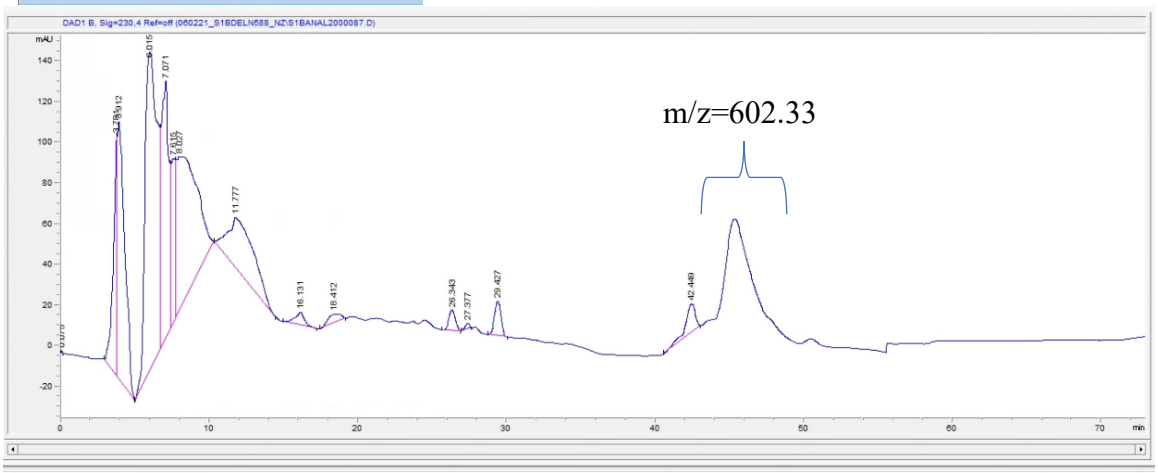
F.6A Mutclu602 MS2 Spectra



F.6B Mutclu602 UV and HPLC Purification

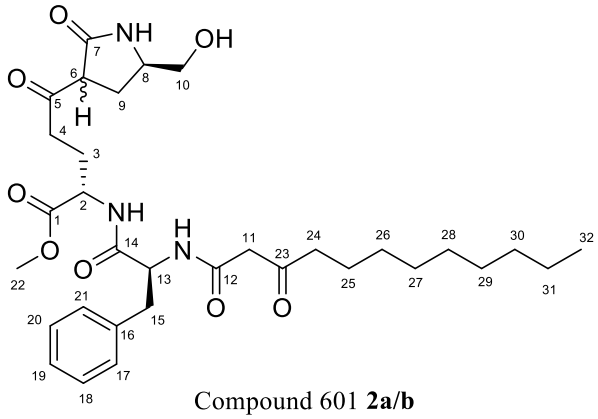


H2O:ACN
52:48->40:60 0-50 mins
0:100 50-63 mins



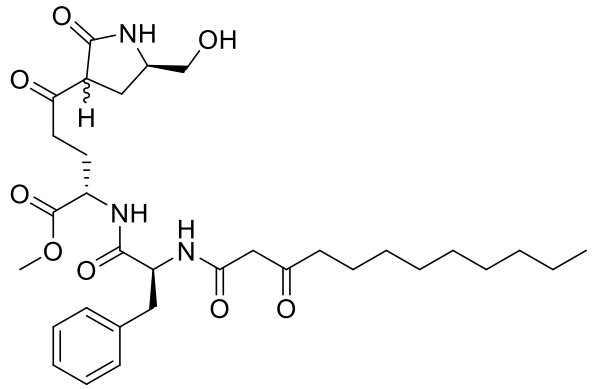
F.6C Mutclu602 NMR

Structural Elucidation of compound 601. Compound 601 was isolated as an amorphous yellow solid with two isomers. Its positive ion HRESIMS revealed a peak for a molecular ion at *m/z* 602.3436 $[M + H]^+$, corresponding to the molecular formula $C_{32}H_{47}N_3O_8$. The differences of 18 Da between the molecular weights of compound 583 and compound 601 implied the structure of compound 601 might be an additional product of water and compound 583. Furthermore, a side-by-side comparison of its NMR spectroscopic data with those of compound 583 showed them to be similar except for the C-10 and C-11 positions. The largest differences were observed in the HSQC spectrum for one methine group at C-11 position (δ_H 3.59; δ_C 54.7) and one methylene group at C-10 position (δ_H 2.24, 1.28; δ_C 32.1) in compound 583 compared to the methylene (δ_H 3.27, 3.22; δ_C 50.4) and the downfield methylene (δ_H 3.32, 3.22; δ_C 64.3) at the same positions in compound 601, indicating that the C-10 C-11 bond was cleavage in compound 601. With all the locations accounted for, the hydroxyl group was connected to C-10, and this assignment was verified by ^3J HMBC from H-9 (δ_H 2.40, 1.84) to C-10 (δ_C 64.3), and -CO-NH-7 (δ_H 7.95) to C-10. Finally, their absolute configurations of compound 601 **2a/b** were assigned as the same as compound 583 **1a/b**.



Compound 601 **2a/b**

Table F.6C-S1. The NMR Data of Compound 601 **2a/b** in DMSO-d_6				
	2a		**2b**	
Position	δ_H (J in Hz)	δ_C (C type)	δ_H (J in Hz)	δ_C (C type)
1		172.1, C		172.1, C
2	4.22 m	51.3, CH	4.26 m	51.2, CH
3	1.90 m; 1.82 m	24.6, CH_2	2.00 m; 1.76 m	25.0, CH_2
4	2.93 m	38.4, CH_2	2.93 m	38.3, CH_2
	2.58 ddd 14.5, 9.2, 5.5		2.64 dt 18.9, 7.5	
5		205.8, C		205.5, C
6	3.62 m	53.9, CH	3.61 m	54.2, CH
7		172.2, C		171.9, C
8	3.49 m	53.5, CH	3.48 m	53.5, CH
9	2.40 m; 1.84 m	25.3, CH_2	2.05 dd 10.3, 5.9	24.6, CH_2
10	3.32 m	64.3, CH_2	3.29 m	64.7, CH_2
11	3.27 d 15.0	50.4, CH_2	3.27 d 15.0	50.4, CH_2
	3.22 d 15.0		3.21 d 15.0	
12		165.9, C		165.9, C
13	4.58 dt 9.0, 4.4	53.6, CH	4.58 dt 9.0, 4.4	53.6, CH
14		171.2, C		171.3, C
15	3.03 dd 13.8, 3.1	37.6, CH_2	3.03 dd 13.7, 4.4	37.6, CH_2
	3.73 dd 13.8, 3.6		2.74 dd 13.7, 3.8	
16		137.7, C		137.7, C
17/21	7.21 m	129.2, CH	7.21 m	129.2, CH
18/20	7.25 m	128.0, CH	7.25 m	128.0, CH
19	7.18 m	126.3, CH	7.18 m	126.3, CH
22	3.61 s	51.9, CH_3	3.61 s	51.9, CH_3
23		204.9, C		204.9, C
24	2.26 t 7.4	41.5, CH_2	2.27 t 7.4	41.5, CH_2
25	1.33 m	22.8, CH_2	1.33 m	22.8, CH_2
26	1.12 m	28.4, CH_2	1.12 m	28.4, CH_2
27	1.23 m	28.7, CH_2	1.23 m	28.7, CH_2
28	1.23 m	28.9, CH_2	1.23 m	28.9, CH_2
29	1.23 m	28.9, CH_2	1.23 m	28.9, CH_2
30	1.23 m	31.3, CH_2	1.23 m	31.3, CH_2
31	1.27 m	22.1, CH_2	1.27 m	22.1, CH_2
32	0.86 t 7.0	14.0, CH_3	0.86 t 7.0	14.0, CH_3
2-NH-	8.48 d 7.5		8.49 d 7.5	
7-NH-	7.95 s		7.95 s	
12-NH-	8.36 d 9.0		8.36 d 9.0	



Compound 601 **2a/b** (^{1}H NMR, DMSO-d_6 at 900 MHz)

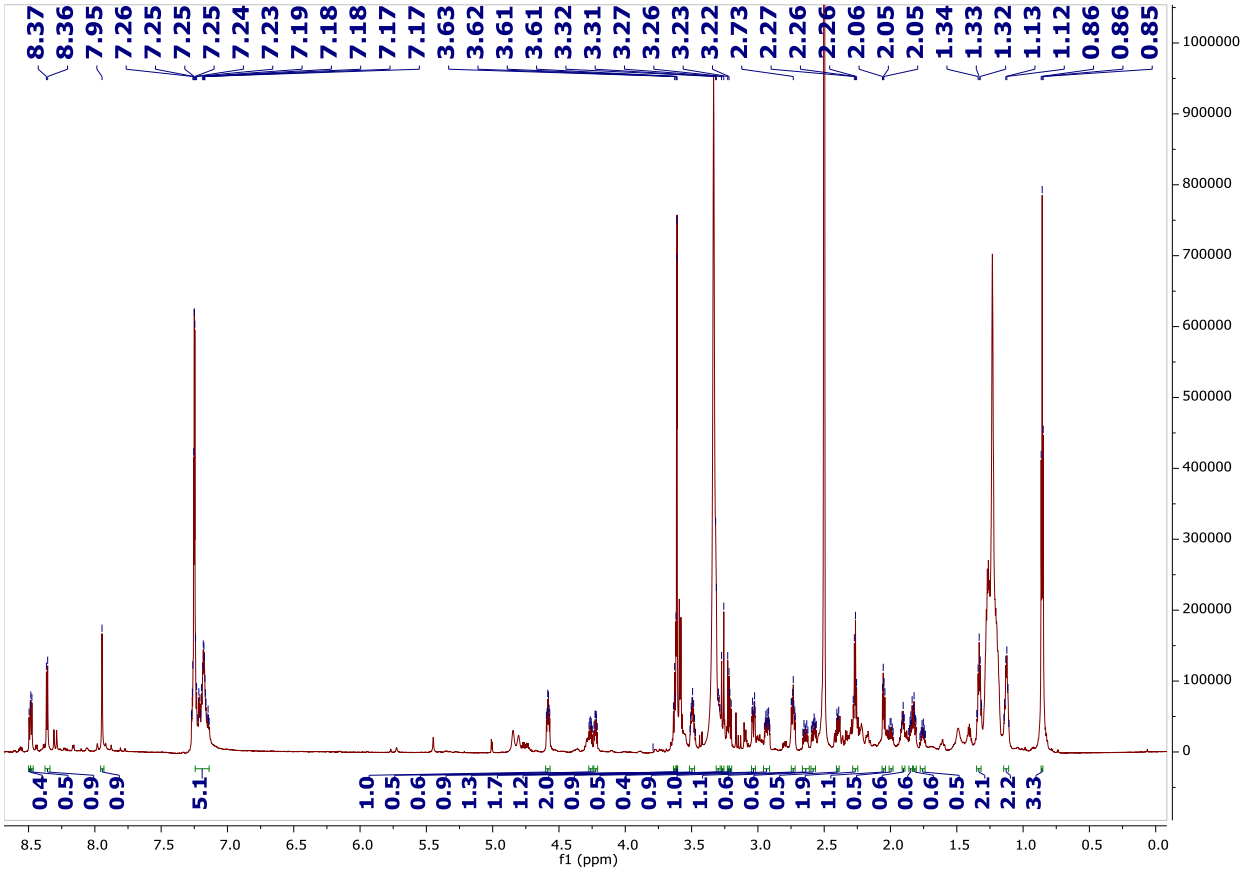
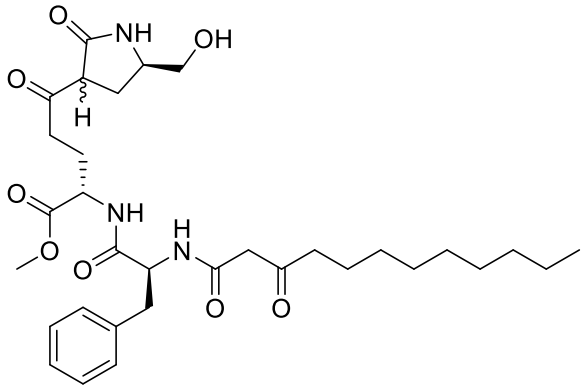


Figure F.6C-S1. ^{1}H NMR spectrum of Compound 601 **2a/b**.



Compound 601 **2a/b** (^{13}C NMR, DMSO-d_6 at 900 MHz)

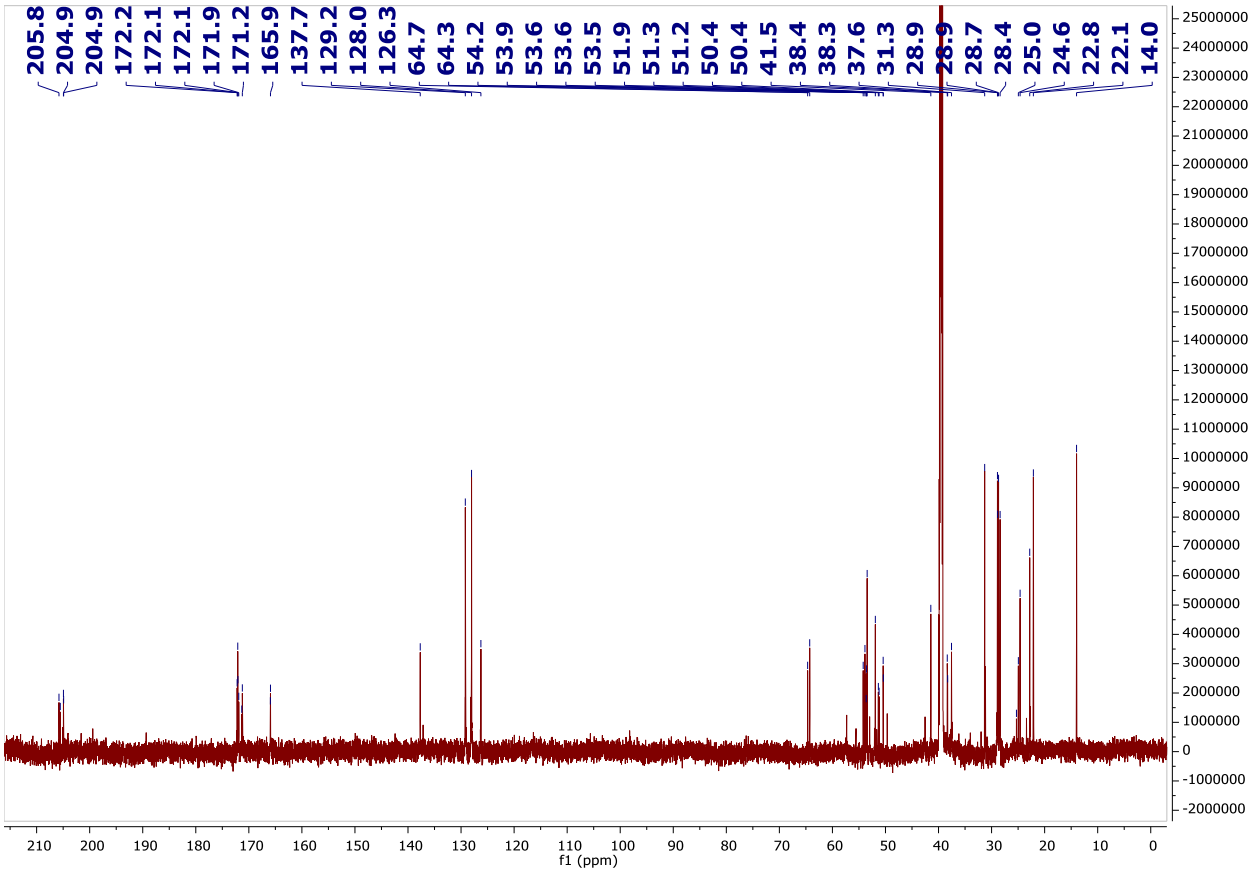
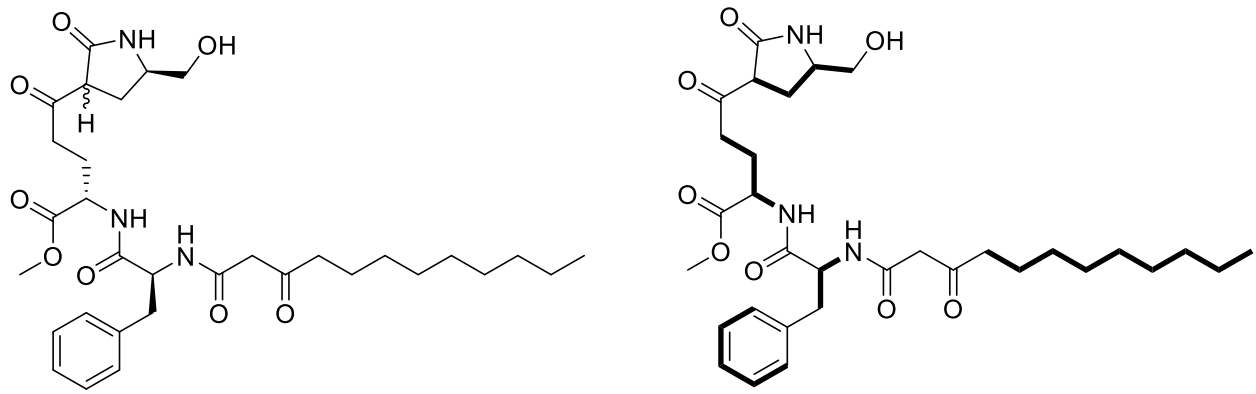


Figure F.6C-S2. ^{13}C NMR spectrum of Compound 601 **2a/b**.



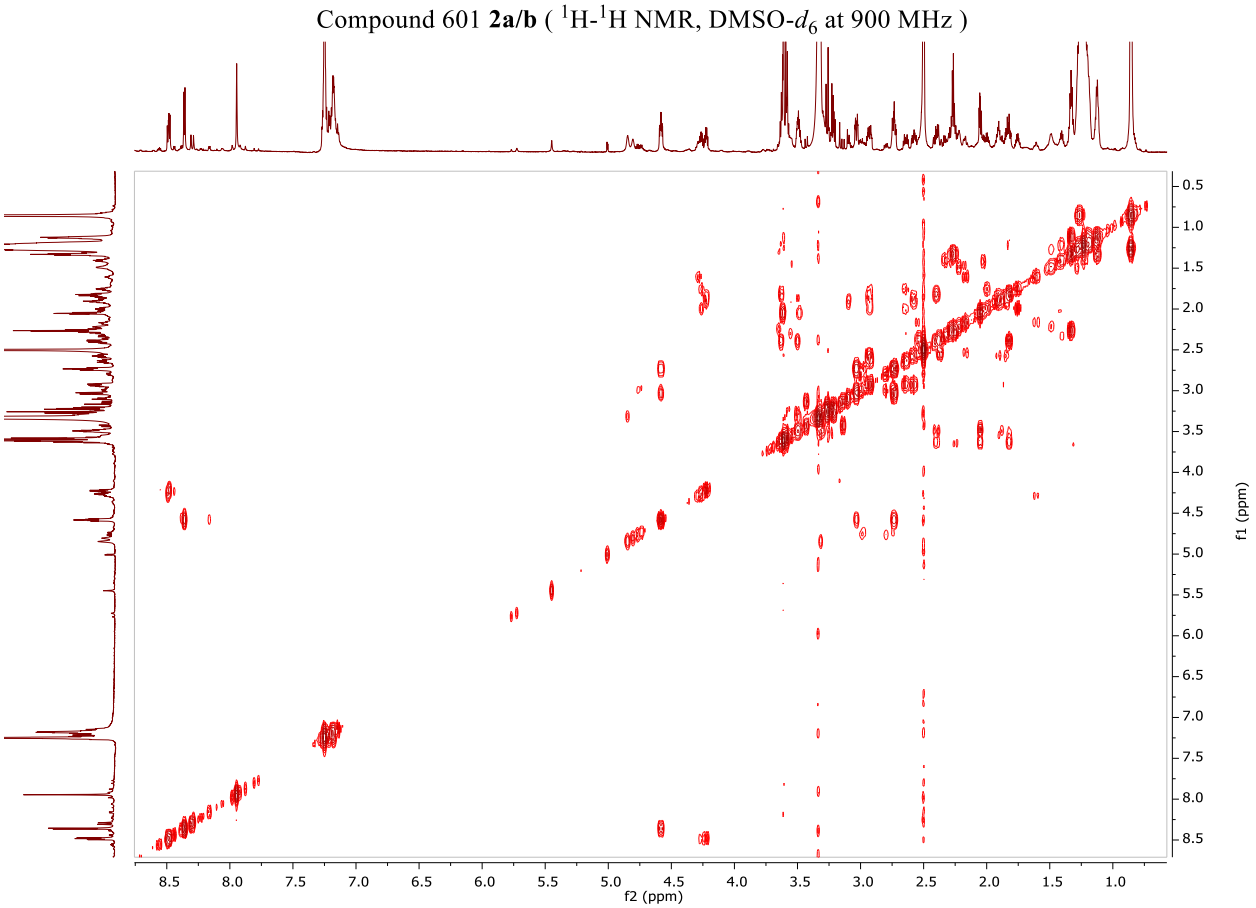
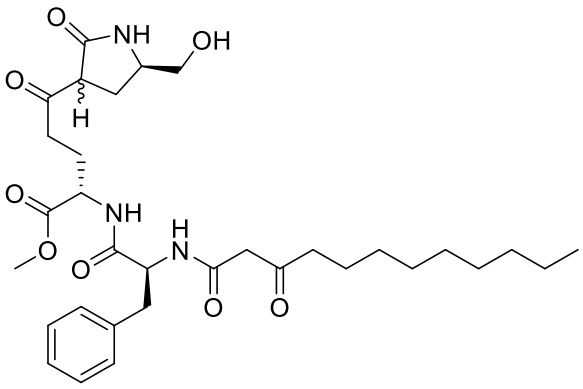


Figure F.6C-S3. ^{1}H-^{1}H COSY spectrum of Compound 601 **2a/b**.



Compound 601 **2a/b** (^{1}H-^{13}C HSQC, DMSO-d_6 at 900 MHz)

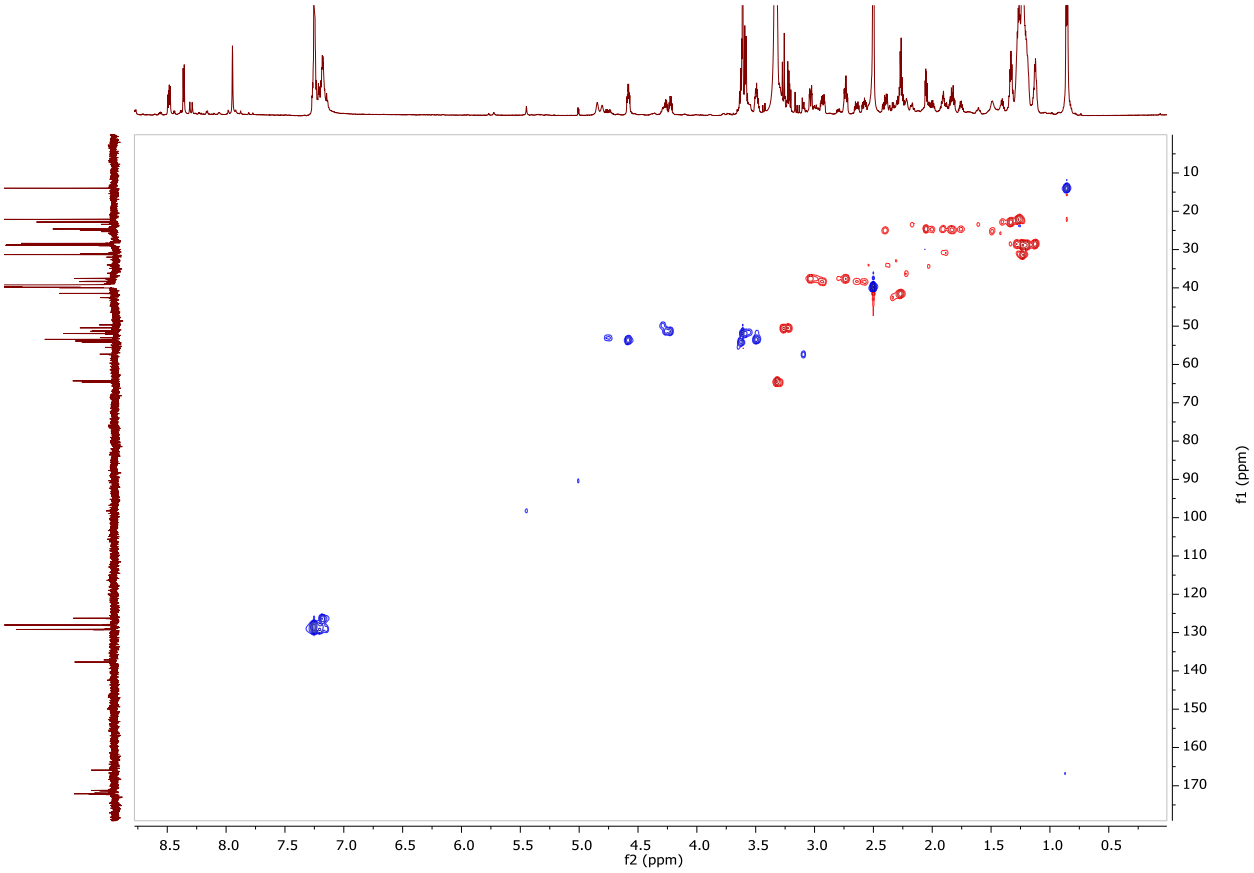


Figure F.6C-S4. ^{1}H-^{13}C HSQC spectrum of Compound 601 **2a/b**.

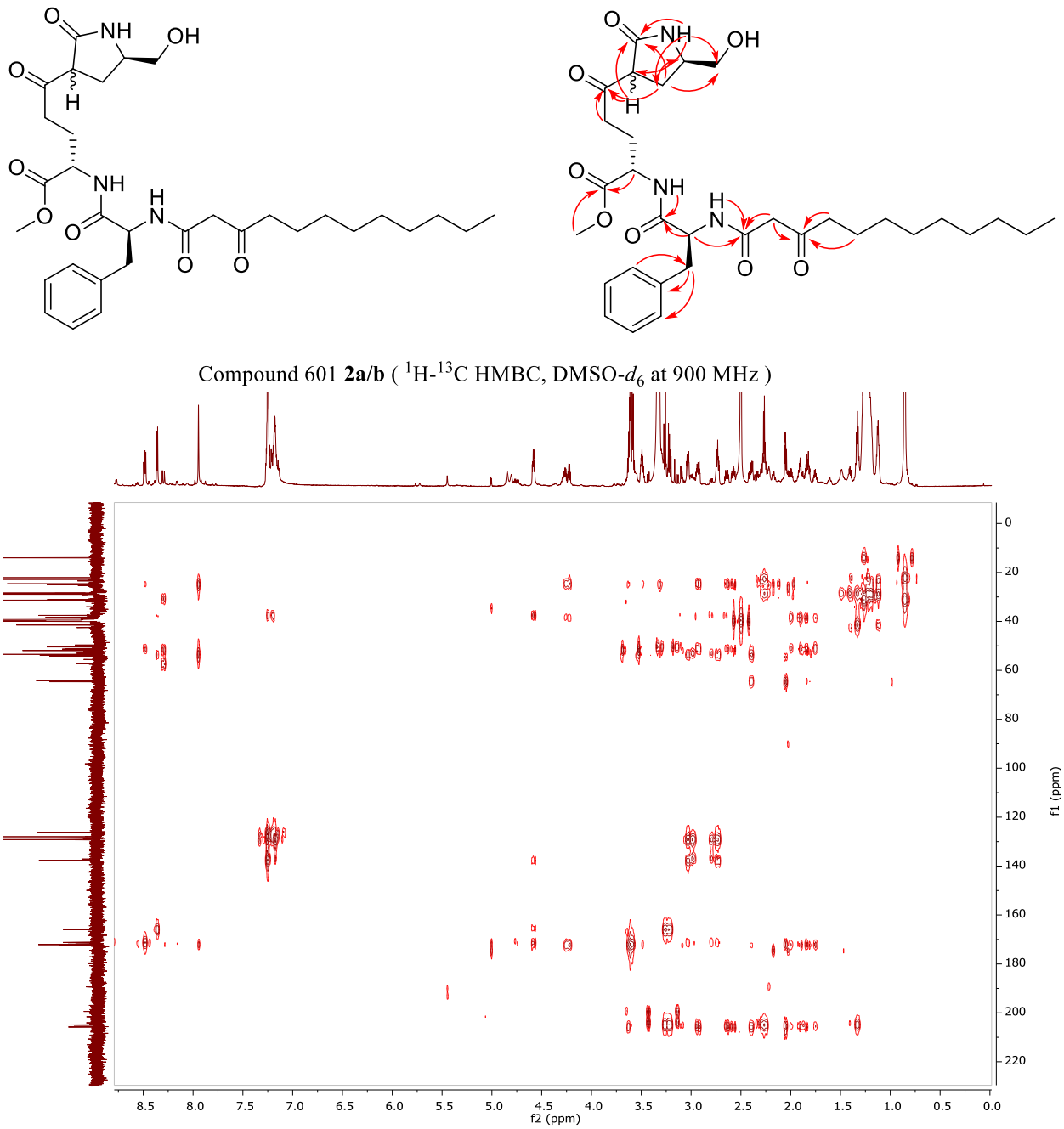
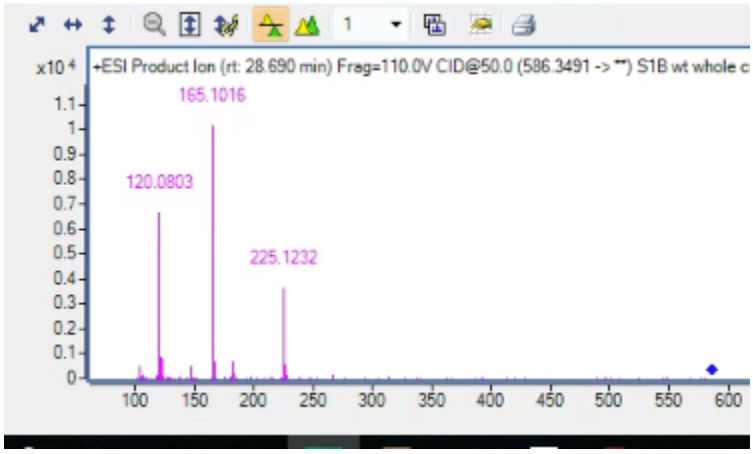
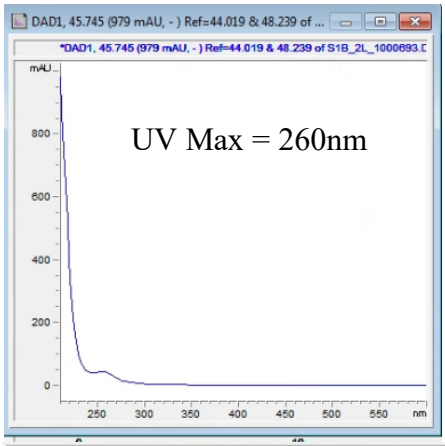


Figure F.6C-S5. ^{1}H-^{13}C HMBC spectrum of Compound 601 **2a/b**.

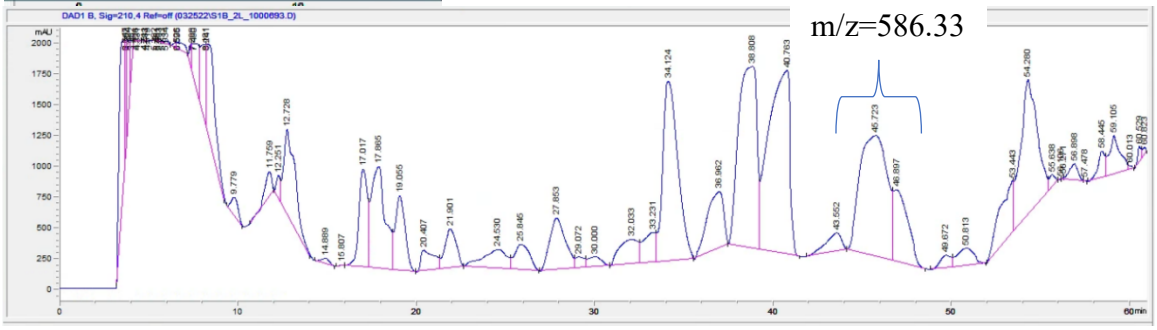
F.7A Mutclu586 MS2 Spectra



F.7B Mutclu586 UV and HPLC Purification



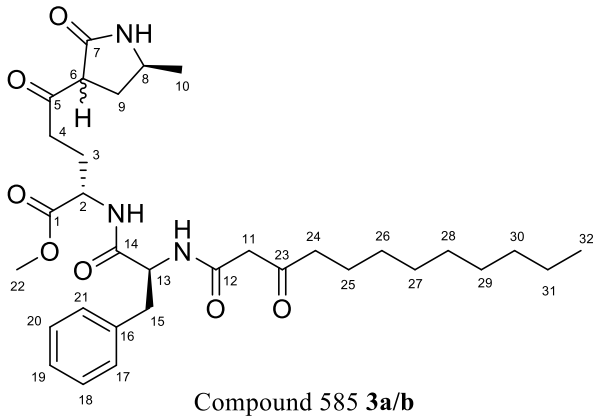
H2O:ACN
50:50->40:60 0-40 min
40:60 Hold 40-50mins
0:100 50-63 mins



F.7C Mutclu586 NMR

Structural Elucidation of compound 585. Compound 585 was also isolated as an amorphous yellow solid with two isomers and had the molecular formula $C_{32}H_{47}N_3O_7$ based on its positive mode HRESIMS, which showed an intense peak at m/z 586.3487 $[M + H]^+$. The differences of 16 Da between the molecular weights of compound 585 and compound 601 implied the structure of compound 601 might be a deoxygenated product

of compound 601. Moreover, through a side-by-side comparison of its NMR spectroscopic data with those of compound 601, the HSQC spectrum showed that one carbon signal at δ_C 64.3 (CH_2) in compound 601 was shifted to δ_C 22.0 (CH_3) in compound 585, which indicated the deoxygenation occurred at C-10. Consequently, their absolute configurations of compound 585 **3a/b** were assigned as the same as compound 601 **2a/b**.

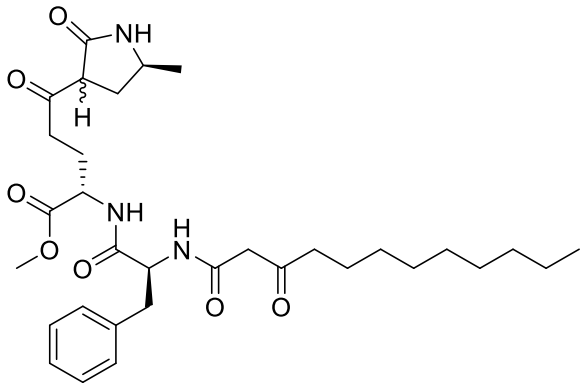


Compound 585 **3a/b**

Table F.7C-S1. The NMR Data of Compound 585 **3a/b** in DMSO-d_6

	3a		**3b**	
Position	δ_H (J in Hz)	δ_C (C type)	δ_H (J in Hz)	δ_C (C type)
1		171.8, C		172.1, C
2	4.22 m	51.3, CH	4.26 m	51.2, CH
3	1.89 m; 1.83 m	24.6, CH_2	2.00 m; 1.76 m	24.6, CH_2
4	2.91 m	38.2, CH_2	2.91 m	38.1, CH_2
	2.60 m		2.64 dt 18.9, 6.8	
5		205.6, C		205.7, C
6	3.65 m	54.3, CH	3.64 m	54.9, CH
7		171.6, C		171.9, C
8	2.47 m	30.5, CH_2	2.16 dt 12.8, 8.8	30.4, CH_2
	1.61 ddd 12.8, 9.2, 5.1		1.85 m	
9	3.61 m	47.6, CH	3.56 m	47.1, CH
10	1.08 d 5.7	22.0, CH_3	1.09 d 5.7	22.0, CH_3
11	3.26 d 15.0; 3.22 d 15.0	50.4, CH_2	3.26 d 14.6; 3.21 d 14.6	50.4, CH_2
12		165.9, C		165.9, C
13	4.57 ddd 8.4, 3.5, 3.3	53.7, CH	4.57 ddd 8.4, 3.5, 3.3	53.7, CH
14		171.3, C		171.3, C
15	3.03 dd 14.1, 3.5	37.5, CH_2	3.03 dd 14.1, 3.5	37.5, CH_2
	2.74 dd 14.0, 3.3		2.73 dd 14.0, 3.3	
16		137.7, C		137.7, C
17/21	7.24 m	129.2, CH	7.24 m	129.2, CH
18/20	7.25 m	128.0, CH	7.25 m	128.0, CH
19	7.18 m	126.3, CH	7.18 m	126.3, CH
22	3.61 s	51.9, CH_3	3.61 s	51.9, CH_3
23		204.9, C		204.9, C
24	2.27 m	41.5, CH_2	2.27 m	41.5, CH_2
25	1.33 m	22.4, CH_2	1.33 m	22.4, CH_2
26	1.13 m	28.4, CH_2	1.13 m	28.4, CH_2
27	1.24 m	28.7, CH_2	1.24 m	28.7, CH_2

28	1.24 m	28.9, CH_2	1.24 m	28.9, CH_2
29	1.24 m	28.9, CH_2	1.24 m	28.9, CH_2
30	1.23 m	31.3, CH_2	1.23 m	31.3, CH_2
31	1.27 m	22.1, CH_2	1.27 m	22.1, CH_2
32	0.86 t 7.1	14.0, CH_3	0.86 t 7.1	14.0, CH_3
2-NH-	8.51 d 7.0		8.51 d 7.1	
7-NH-	7.99 s		7.98 s	
12-NH-	8.41 d 8.4		8.40 d 8.4	



Compound 585 **3a/b** (^{1}H NMR, DMSO-d_6 at 900 MHz)

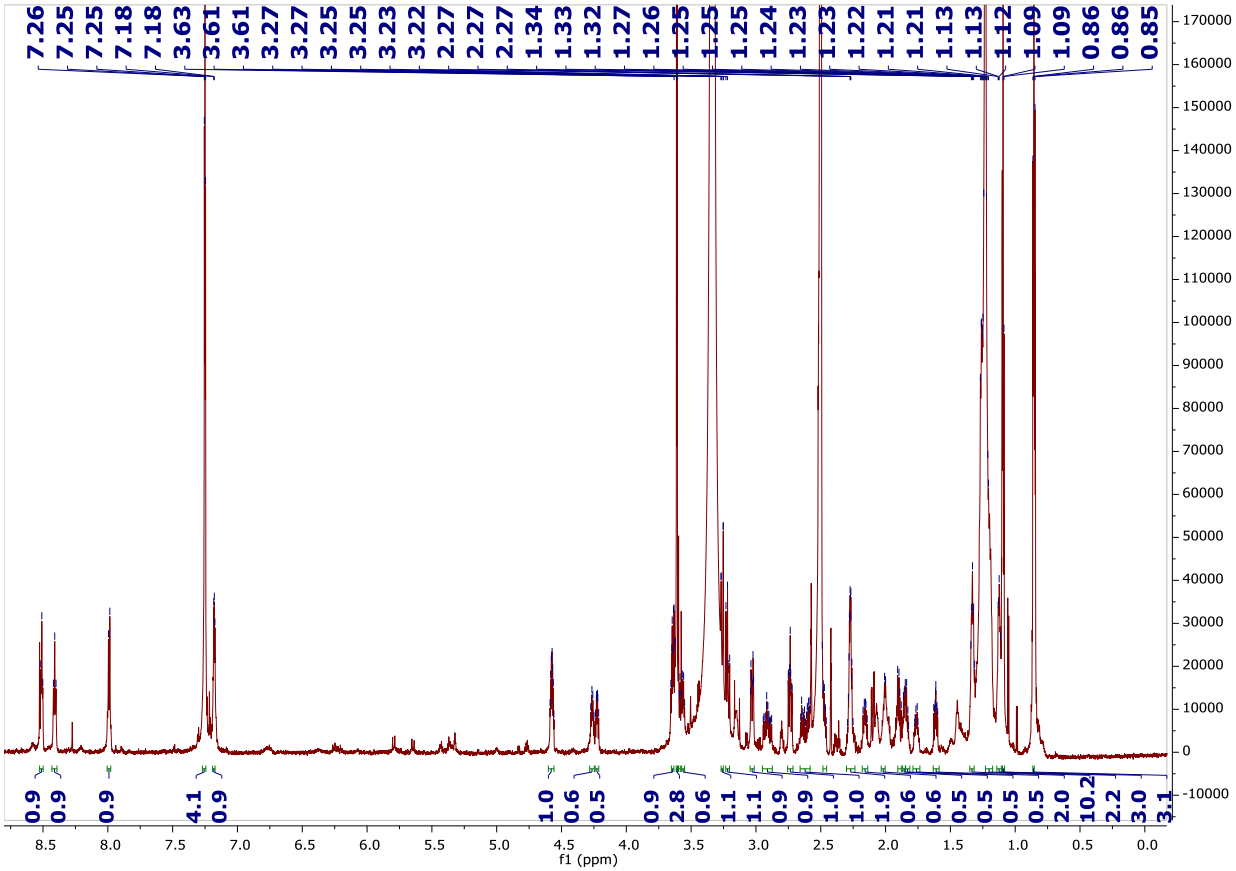
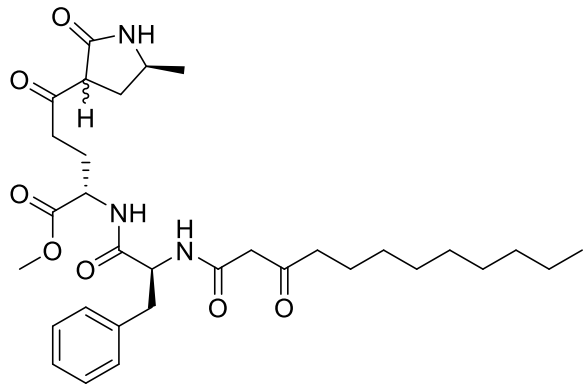


Figure F.7C-S1. ^{1}H NMR spectrum of Compound 585 **3a/b**.



Compound 585 **3a/b** (^{13}C NMR, DMSO-d_6 at 900 MHz)

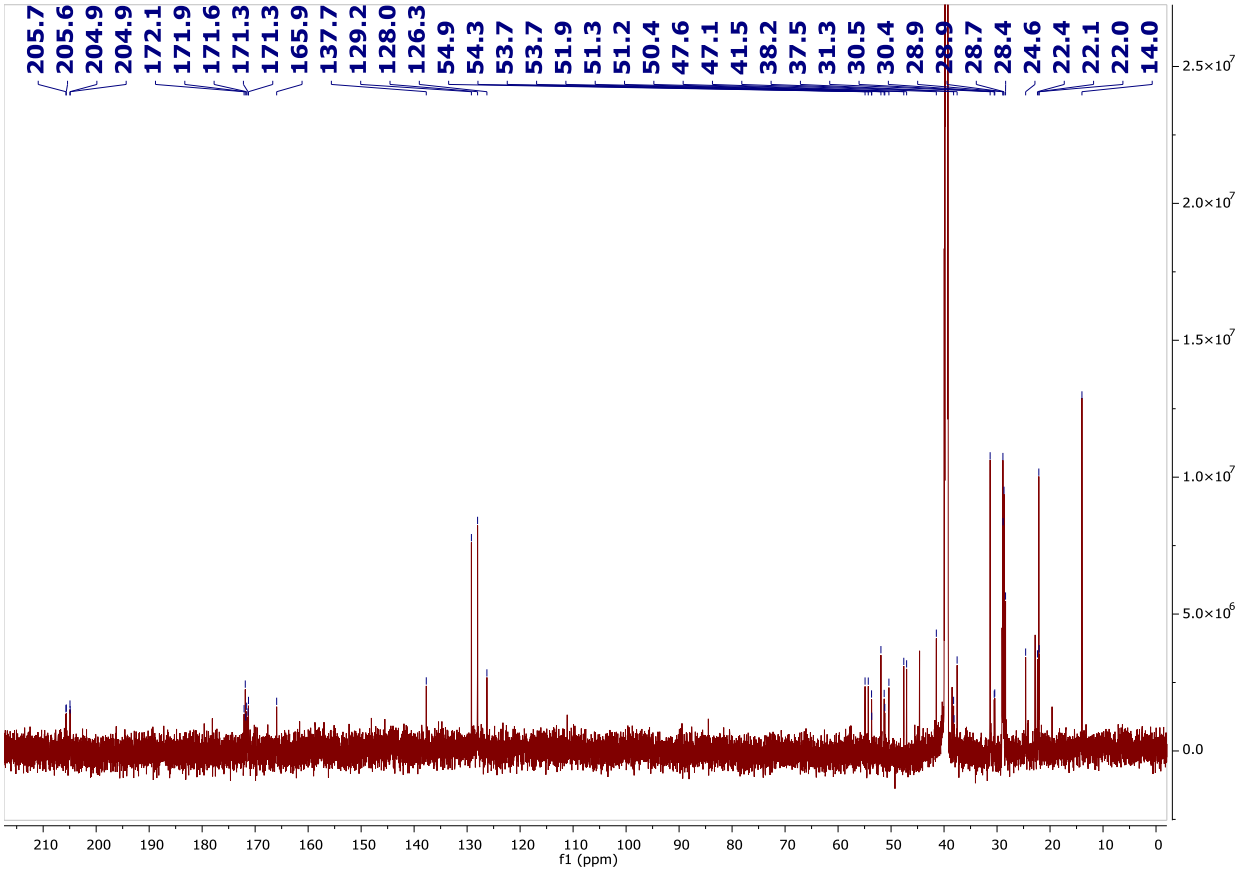
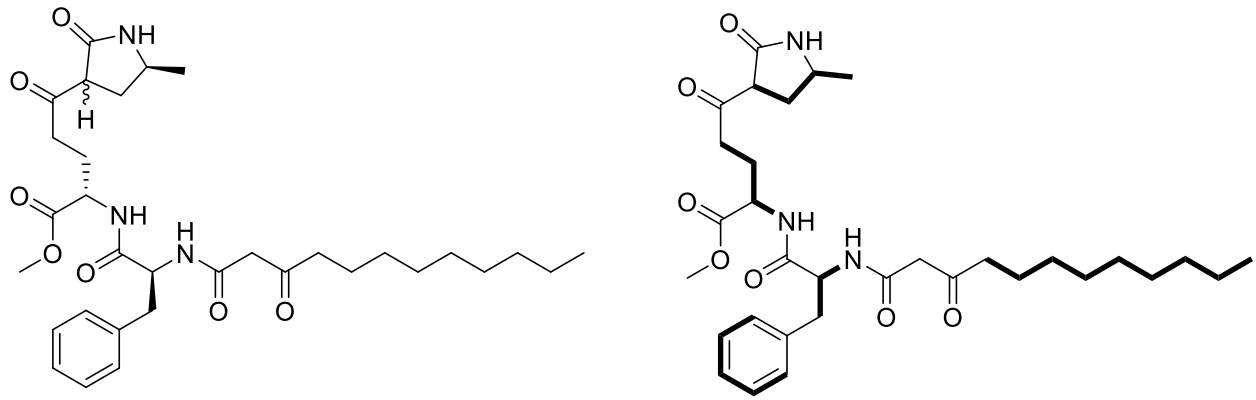


Figure F.7C-S2. ^{13}C NMR spectrum of Compound 585 **3a/b**.



Compound 585 **3a/b** (^{1}H-^{1}H COSY, DMSO-d_6 at 900 MHz)

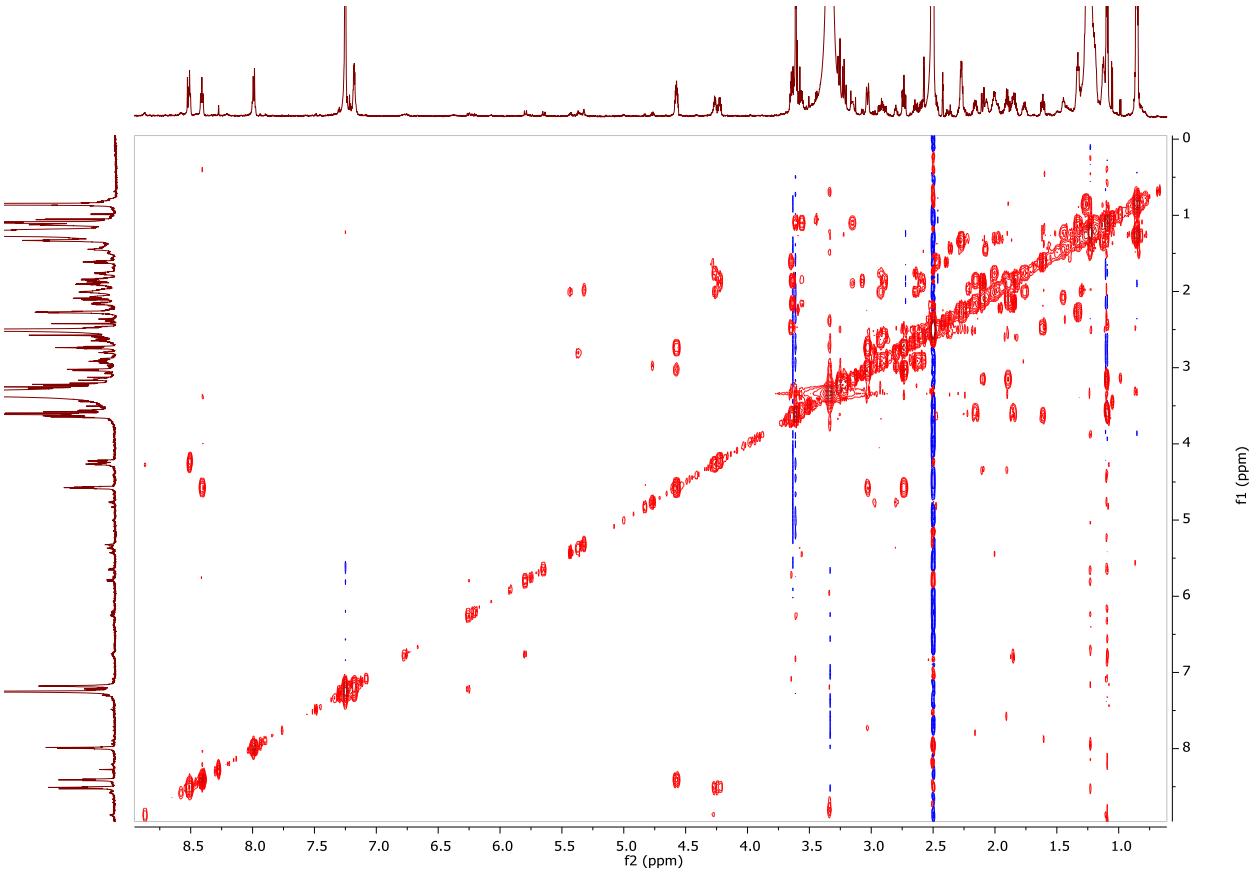
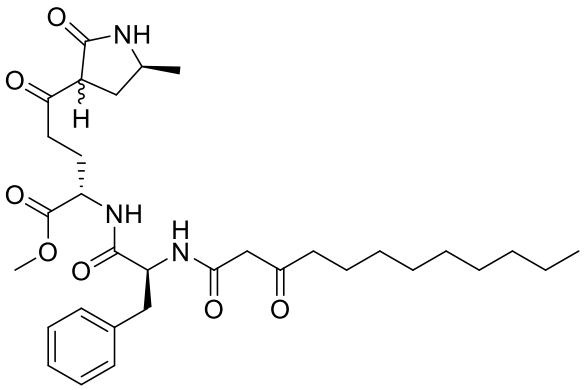


Figure F.7C-S3. ^{1}H-^{1}H COSY spectrum of Compound 585 **3a/b**.



Compound 585 **3a/b** (^{1}H-^{13}C HSQC, DMSO-d_6 at 900 MHz)

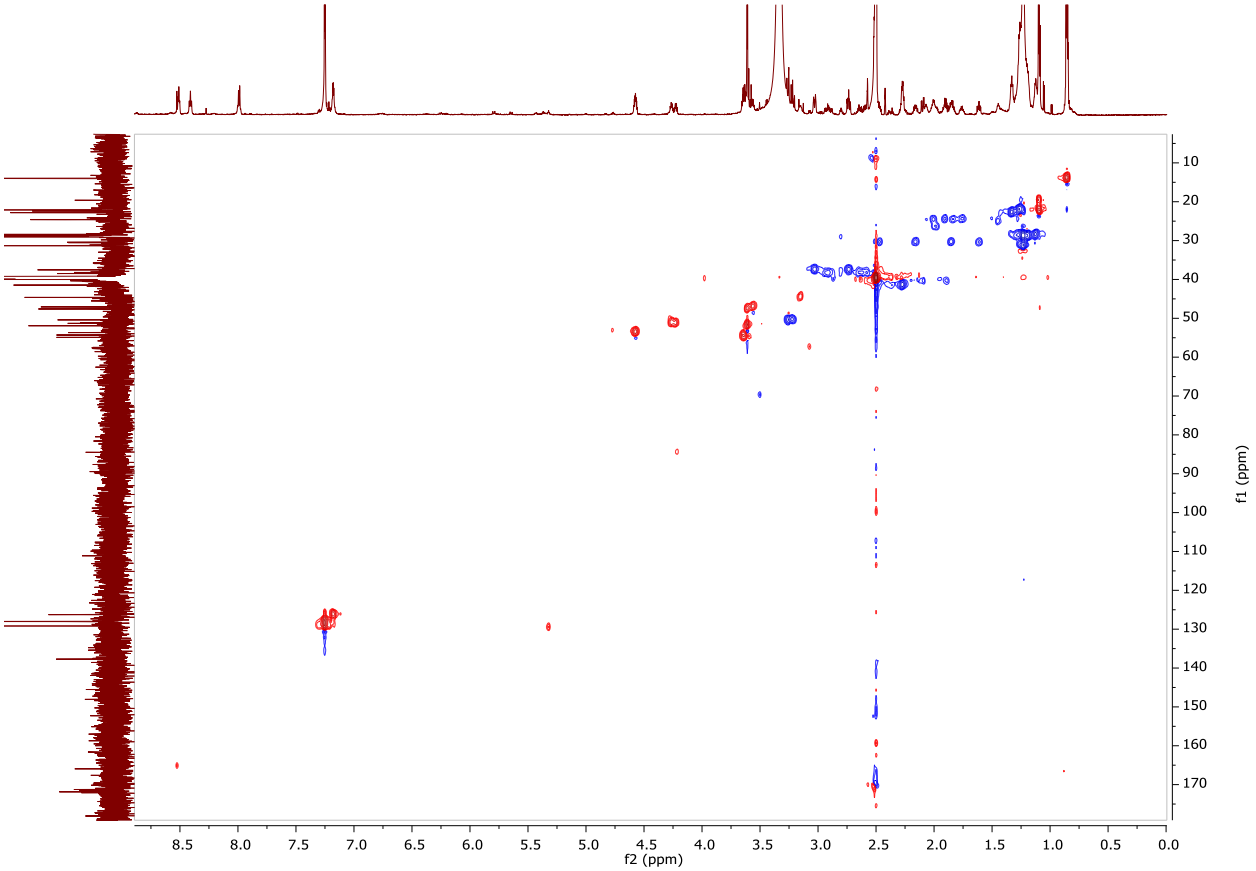


Figure F.7C-S4. ^{1}H-^{13}C HSQC spectrum of Compound 585 **3a/b**.

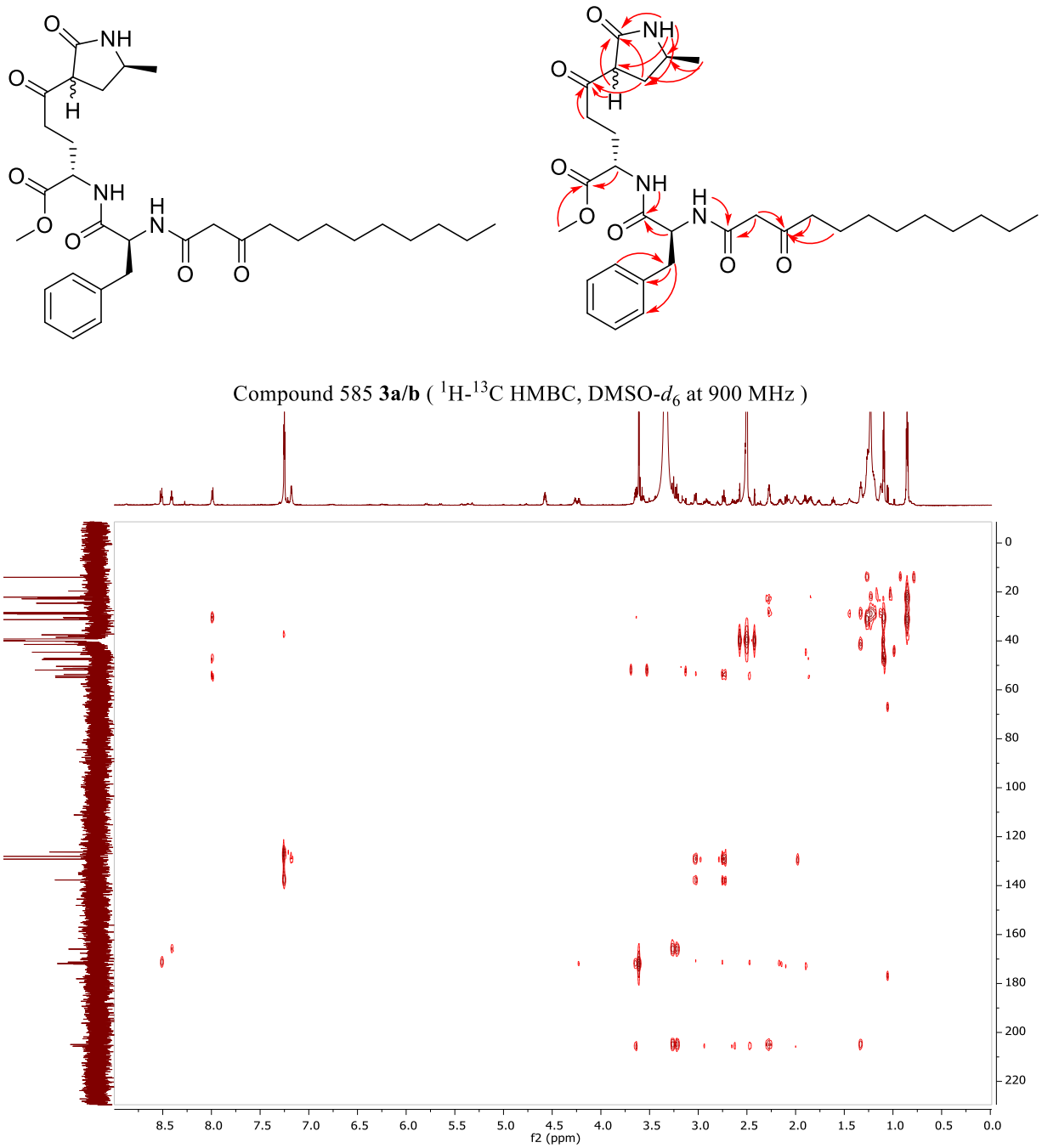
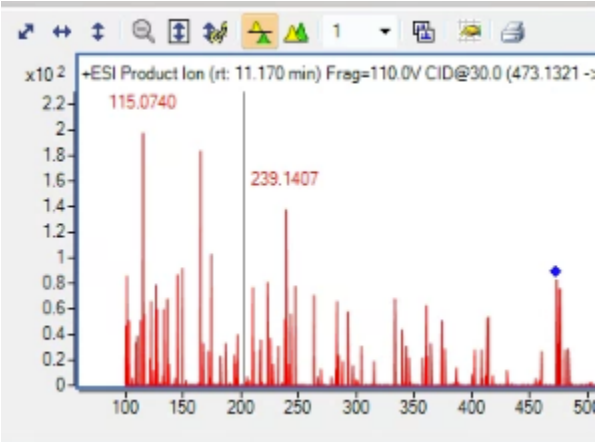


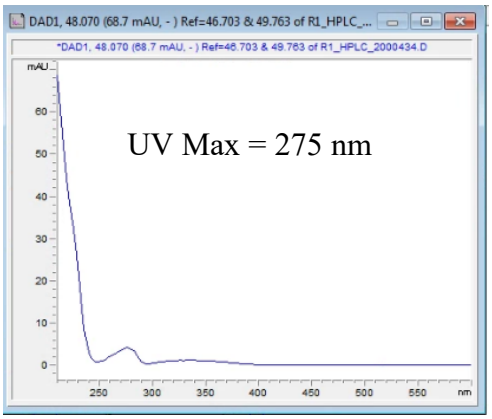
Figure F.7C-S5. ^{1}H-^{13}C HMBC spectrum of Compound 585 **3a/b**.

F.8A m/z=473.27 MS2 Spectra

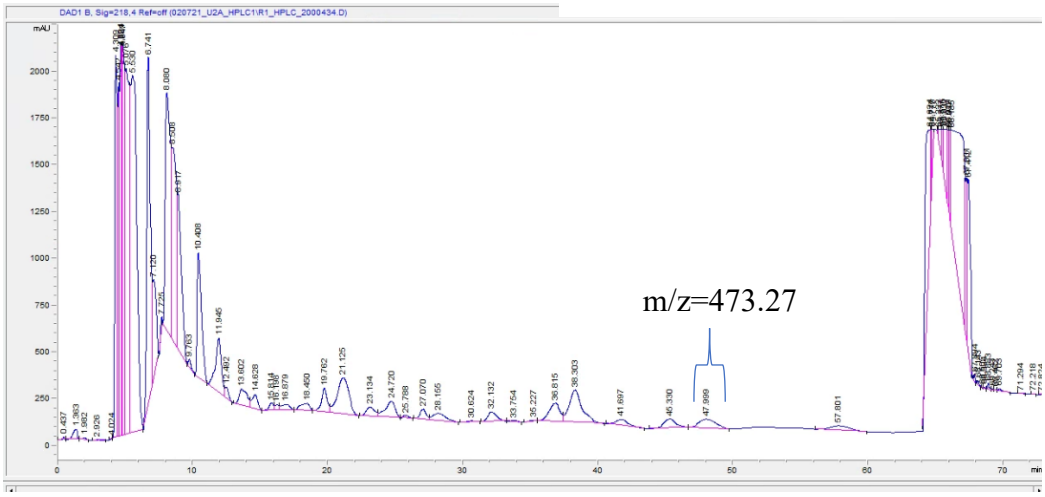


F.8B m/z=473.27 UV and HPLC Purification

The UV spectra of 473.27 is tentative given the current low abundance of compound



H2O:ACN + 1% TFA
85:15 Hold 0-60 mins
0:100 60-73 mins



F.8C m/z=473.27 LH20 Purification