Title

Tagging Small Molecule Natural Products with Functional Chemical Probes To Enable New Techniques in the Discovery and Development of Pharmaceuticals

Permalink

https://escholarship.org/uc/item/8c13q79d

Author

Seidel, Jeremy

Publication Date

2020

Peer reviewed|Thesis/dissertation

Tagging Small Molecule Natural Products with Functional Chemical Probes
To Enable New Techniques in the Discovery and Development of Pharmaceuticals

By

Jeremy A. Seidel

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
Professor Aaron Streets

Summer 2020

Abstract

Humans have exploited small molecule natural products have exploited throughout history. Recently, advances in chemistry and biology have led to an explosion in drug development, revolutionizing the treatment of infectious diseases, cancer, and other ailments. However, current techniques have already been exploited nearly to their limits. The work described in this dissertation aims to expand the pharmacological potential of small molecule natural products by developing a range of techniques that exploit functional chemical tags that are chemically bonded to the small molecules. These tags enable natural products to be studied in new ways using techniques such as Raman microscopy, computational analysis of LCMS data, and optical lock-in detection of FRET. By fundamentally altering small molecule natural products, we have raised the possibility to discover and study drug-like molecules that would be inaccessible using traditional techniques.

This is thesis is dedicated to my wife, Victoria.

Her love and support made this work possible.

Table of Contents

iii

List of Figures

List of Schemes

List of Tables

Acknowledgements

Chapter 1: A Brief History of Natural Products Pharmacology
Drugs are the product of the tools we use to study them

Small molecule natural products have been used for their medicinal properties since prehistoric times. Over the past 10,000 years, humans across the world have discovered and utilized plants with potent biological activity and learned to utilize them as entheogens, analgesics, sedatives, and even cures for deadly disease. In the past 200 years, this progress has greatly accelerated dramatically as a result of systematic screening, and advanced chemical knowledge allowing for active compound isolation. More recently, a shift to high-tech screening methods and the advent of revolutionary new drug modalities holds the promise of uncovering even more potent natural product drugs to cure a wider range of cancers, neurological and infectious diseases.

Human use of natural products evolved along with the development of human civilization. Prehistoric people used plants with an easily determined cause and effect, especially psychoactive compounds. Writing allowed for more rigorous documentation allowed for the discovery of antipyretic compounds, use of sedatives for surgery, and even lead to the discovery of a cure for malaria and an early form of chemotherapy. More recently, modern chemistry techniques have allowed for the isolation and characterization of bioactive compounds from natural sources. When combined with systematic screening methods, this led to the discovery of many antibiotics and anticancer compounds. In fact, natural products still represent over 60% of antibiotics and chemotherapy drugs.[1] In the past 20 years, targeted drug discovery and chemical biology have allowed researchers to develop small molecule therapies, including natural products, for an even wider range of human diseases. Natural product pharmacology is tightly intertwined with technological progress in general.

In order to navigate the future of natural product pharmacology, it is essential to understand its history. From prehistoric humans consuming psilocybin mushrooms to the advent of modern pharmacology, there is a common lesson to be learned: Drugs are the product of the tools we use to study them. Therefore, expanding the drug development toolset is essential to the advancement of medicine. The work presented in this thesis focused on forging a new class of drug discovery tools: functional tags that are directly attached to small molecule natural products. These tags include:

(1) vibrational tags to aid in spectroscopic imaging (Section 1.4.2)
(2) isotopic labels for streamlined LCMS analysis (Section 1.5.2)
(3) photo-switching fluorophore tags leverage a state-of-the-art FRET technique to illuminate protein-ligand binding behavior in live cells (Section 1.5.3)

1.1 Pre-historic use of Natural Products

Humans have been using primitive forms of herbal medicine since pre-history and it is assumed that these prehistoric humans were using plants, such as aloe vera or rosemary, for their medicinal properties. One particularly compelling piece of evidence was the discovery of a frozen mummified corpse in the Italian Alps, who lived 5,300 years ago. The man, who suffered from a parasitic infection of the intestines, carried birch fungus in a leather pouch.[2] Because birch fungus

is known to have anti-parasitic and laxative effects, researchers believe this man may have been consciously treating his infection.[2] Given this evidence, it seems reasonable to conclude that the laxative, anti-infective, and analgesic effects of these natural products would have been readily apparent and incredibly useful to early people. Knowledge of these plants was likely considered immensely valuable and passed down through generations as a verbal tradition.

The earliest recorded, deliberate use of a biologically active substance dates to 9,000-year-old African cave painting that depicts the ritual use of psilocybin mushrooms.[3] Psilocybin mushrooms are the most abundant source of the naturally occurring psychoactive compound psilocybin. Psilocybin is a pro-drug that is converted to psilocin inside the human body, which then crosses the blood-brain barrier and agonizes serotonin 5-HT2A receptors in the brain. It is easy to imagine how this potent psychoactive substance could have been discovered without rigorous science. Early humans would have likely eaten a wide variety of mushrooms, and the psychoactive effect from eating psilocybin mushrooms would have been obvious. Furthermore, these mushrooms have virtually non-existent toxicity.[4] Therefore, prohibition of their use in early society would have been unlikely. An easily identified natural producer, with potent psychoactive effects, and an excellent safety profile was essentially "low hanging fruit" for early humans to discover.

Based on accounts of psilocybin intoxication causing visions and spiritual experiences, it is believed that ritual use of psychedelic mushrooms may have formed the basis for early religious beliefs and ceremonies. Indeed, ritual use of plants containing small molecules that agonize 5-HT2A serotonin receptors have been established independently at least three other times throughout history. Peyote, a cactus containing the 5-HT2A agonist mescaline, has been used by Native Americans dating to at least 5,700 years ago.[5] Ayahuasca, a tea brewed from two plants containing the 5-HT2A agonist DMT and a monoamine oxidase inhibitor (MAOI) has been used in religious ceremonies in the Amazon Rainforest for at least 1,000 years.[6] Several species of morning glory, a rich source of ergine (a precursor to lysergic acid diethylamide (LSD)), have been similarly used as entheogens throughout Latin America for several hundred years.[7] Such widespread use of 5-HT2A receptor agonists illustrates a key point. Early people used these plants because their effects were readily apparent and presented obvious utility without rigorous documentation or sophisticated scientific methods.

Prehistoric use of psychoactive natural products also includes the use of other drug modalities. South American people began chewing coca leaves in a communal setting for their stimulating effect about 8,000 years ago.[8] It is believed that the stimulating effects of cocaine, a promiscuous monoamine reuptake inhibitor and central nervous stimulant helped early people work and hunt longer. Caffeine, the most commonly consumed drug today, is used in a similar manner by modern humans. However, the discovery of caffeine came much later, around 3,000 years ago in China with the use of tea and around 1,000 years ago in Latin America with the use of coffee.[9]

Contemporaneously with the use of coca, but across the globe in Asia, early people began cultivating hemp for fiber. However, the earliest documented medicinal use of cannabis only dates back 4,700 years, so it is not known if early people knew of its pharmacological potential.[10] This delay in the exploitation of cannabis's psychoactive properties may be attributable to the fact that THCA must be decarboxylated by heating to yield THC, which is usually achieved by cooking or

smoking the plant. This illustrates that the preparation of natural products is important to their pharmacological activity and may have presented another barrier to their discovery by early people.

1.2 Writing Drives Progress in Natural Product Pharmacology

The advent of writing in ancient Sumer 5,500 years ago would go on to revolutionize all aspects of human society, and natural product pharmacology was no exception. Early writings depict the use of willow bark, a source of the compound salicin, which is a relative of and precursor to salicylic acid and acetyl salicylic acid, commonly known as aspirin.[11] Salicin, like aspirin, is an antipyretic drug and pain reliever. A fever reducing drug may seem commonplace today, but it would have represented a revolutionary life-saving treatment at the time. In order to recognize the potent pharmacological potential of willow bark, early doctors would have needed to correctly diagnose a fever before administering treatment. Only through extensive experience, formalized training, and record-keeping were ancient Sumerians able to harness the incredible potential of this plant.

Ancient Greeks began using opium over 3,000 years ago.[12] At first glance, this may seem to a simple continuation of the exploitation of psychoactive compounds since prehistoric times. Indeed, opium use also dates to prehistoric times when it was used as a pain reliever and abused as an addictive drug. However, the Greeks demonstrated unprecedented prowess in utilizing opium's sedative and analgesic properties for the advancement of human healthcare. Opium was dosed precisely in patients prior to surgical operations, including an ancient form of eye surgery.[13] A patient with cataracts would be administered opium, after which the surgeon would use a sterile needle to push the cataract out of the way of the pupil. Later the Greeks even developed tools that would allow them to remove the cataract from the eye after it was displaced, ensuring that it would not obscure the patient's vision again. All these advances would have been impossible without the ever-present use of medicinal natural products leading up to this time, and the countless scholars who wrote about them.

While there are many ancient herbal traditions from civilizations across the globe, none are more extensive, more impressive, or more controversial than Traditional Chinese Medicine (TCM). TCM is infamous for promoting unfounded remedies, like rhino horn or pangolin scales, that drive poaching of endangered species to the brink of extinction. However, it is also the source of several astonishing success stories. As early as 2,200 years ago, Chinese herbalists began using the leaves of Artemisia to brew a treatment for malaria.[14] Artemisinin, the anti-malarial compound from Artemisia, was isolated and characterized in 1972 as the result of a research program that sought to rediscover antimalarial treatments by reviewing TCM texts. Artemisinin, a molecule derived from an ancient herbal remedy, is now considered the standard treatment for malaria. For ancient people to discover this treatment, they had to be able to properly diagnose malaria, and rigorously measure the effect of the treatment across many patients. In short, the discovery of Artemisinin was founded on rigorous medical record keeping. TCM also described the bark of the *Camptotheca acuminata* tree as a treatment cancer 1,600 years ago.[15] Camptothecin, the active component from that herbal remedy, is now a commonly used chemotherapy drug. The rigorous record keeping of early Chinese doctors and scholars enabled them to make uncanny discoveries

in the field of natural product pharmacology. However, such blunt methods did not enable them to discover treatments – or exclude ineffective treatments – nearly as well as modern scientific methods.

1.3 Modern Science Revolutionizes Natural Product Pharmacology

Beginning in the early 1800s, chemistry techniques were applied to natural medicines, allowing for the isolation and characterization of individual bioactive compounds from the natural producer. Between 1803-1817 Friedrich Wilhelm Adam Serturner performed a series rigorous separations and recrystallizations on crude opium in conjunction with animal and human testing to isolate morphine.[16] He later used the crystalline compound to treat his own tooth pain and performed a series of dose response studies to determine a safe dose of morphine, about 15 mg. Later, Serturner established a pharmaceutical lab that isolated many other alkaloids, such as quinine and strychnine. By isolating the individual chemical components from medicinally active and toxic plants, Serturner had established the field of pharmacology as a rigorous scientific discipline.

The isolation of individual bioactive compounds from natural producers continues to influence modern medicine. Cannabis has been used for thousands of years for its psychoactive and therapeutic effects, including use by the ancient Assyrians to treat epilepsy.[17] However, little was known about its chemical composition until the mid-1900s when scientists first isolated cannabinol (CBN), tetrahydrocannabinol (THC), and cannabidiol (CBD) in the 1940s.[18] Isolating and characterizing these compounds allowed researchers to individually study their effects. Raphael Mechoulam's research team tested CBD to treat patients with seizures in the early 1980s with good results. After over 20 years of regulatory interference, scientists gradually proved CBD's effectiveness in treating epilepsy, for which it won FDA approval in 2018. Although the identification of pure CBD as a treatment for epilepsy seems like a triumph of modern science, there is evidence that early people across the Middle East and India knew of marijuana's anticonvulsant properties as early as 3,000 years ago.[19] The truly modern innovation is that epilepsy patients can use CBD to alleviate convulsions without the psychoactive, inebriating effects of THC.

Throughout the late 19th and early 20th centuries, synthetic compounds were favored as a source of new drugs. However, Alexander Fleming's discovery of the antibiotic activities of Penicillium mold in 1928 would go on ignite an explosion of natural product drug discovery. The story of Fleming's now-famous discovery describes the observation that a staphylococcus culture plate contaminated with penicillium mold exhibited a ring of inhibition around the mold in which the bacteria could not grow. Fleming went on to replicate this experiment and isolate a crude form of the active component, which he called penicillin.

After Fleming's initial discovery, research on Penicillin fell dormant until 1939, when Howard Florey and Ernst Chain organized an effort to manufacture a pure, stable form of penicillin. Penicillium mold was grown in liquid culture, rather than on agar plates, and a sophisticated system of organic extraction and aqueous back-extraction was used to recover the penicillin.[20] The team then tested penicillin on rats and found it could be used to treat bacterial infections. At this point, pharmaceutical companies stepped in to help with scale up and manufacturing, and penicillin

became a life-saving drug during World War II. Eventually, Dorothy Hodgkin characterized penicillin's structure by X-ray crystallography in 1945. While Fleming is given sole credit for Penicillin's discovery, it's exploitation as an antibiotic is a true wonder of interdisciplinary science, spanning biology, chemistry, chemical engineering, and physics.

Eager to replicate Fleming's discovery of a life-saving drug, scientists began systematically screening for antibacterial compounds in the 1950s. Scientists systematized the screening of natural product extracts against bacteria, pioneering the field of phenotypic assays for drug screening. By seeking to replicate Fleming's discovery on a massive scale, scientists successfully discovered dozens of new antibiotics, ushering in the so-called "golden age of antibiotics." During this time, dozens of new antibiotics were isolated, characterized, and approved for treatment of many infectious diseases.[21]

The earliest anticancer drug discovery efforts followed a similar workflow to antibacterial drug discovery. The traditional natural product drug discovery process starts with cell culture of microbes (typically actinomycetes or fungi) to produce a mixture of potentially bioactive compounds. This step may also include plants or other producing organisms. Following cell culture, an organic extraction is used to recover compounds from the production media or plant material. This organic extract is then dried and the residue is re-dissolved in DMSO. This DMSO solution is then added to indicator cell cultures to determine if any compounds in the mixture inhibit cell growth. Following inhibition assays the entire cell extract is fractionated by HPLC, and each fraction is retested by an inhibition assay to isolate the active compound from an extract. Once an active fraction is identified, it usually needs to be re-fractionated, resulting in an iterative process that can take months.

Despite early success in anticancer drug discovery, these chemotherapies are plagued by a common problem: off-target toxicity towards healthy cells. While this problem could be mitigated by finding compounds that are selectively toxic towards certain cell lines, off-target toxicity remains a problem to this day. This off-target toxicity causes the side effects of chemotherapy: fatigue, hair loss, nausea, and immunosuppression. Using cytotoxic compounds, the fight against cancer is reduced to a war of attrition. Who will die first, the cancer or the patient?

1.4 Mode of Action Studies

Because phenotypic assays account for interaction with the entire proteome, small molecules discovered by this method may have known medicinal or toxic properties long before their molecular target or mode of action is known. Identifying the direct molecular target for a bioactive small molecule presents several advantages in the drug development process. Drug candidates with known targets can be optimized to bind more strongly to that target and avoid binding to off-target proteins that may contribute to undesirable side effects. Moreover, identifying a drug's mode-of-action often elucidates previously unknown therapeutic targets. At several points throughout the history of biology, identifying the mode-of-action of a bioactive natural product has contributed to rapid advances in our understanding of the biochemical and genetic drivers of life. Indeed, scientists have often named proteins after the natural products which target them:

- Cannabinoid receptors are named for the cannabinoids from cannabis
- Opioid receptors are named for the opiates that bind to them
- mTOR stands for "molecular target of Rapamycin"
- Many others are named for endogenous small molecules (i.e. serotonin receptors)

The relationship between small molecule natural products and fundamental biology is a fascinating case study in the unpredictable twists and turns of scientific progress. Rapamycin, a small molecule natural product isolated from *streptomyces hygroscopicus*, stands out for its impact on the fields of immunosuppression, oncology, and aging science.

Rapamycin was first isolated in 1972 as the result of a multi-year effort to identify bioactive molecules from soil bacteria samples originating from the island of Rapa Nui. Soon after its isolation it was found to have cytostatic anticancer activity, meaning that it halted the cell cycle rather than directly killing cells. It wasn't until 1994 that the molecular target of Rapamycin (mTOR) was discovered by cloning a gene that mediated Rapamycin toxicity in yeast.[22] Further study of mTOR revealed it to the central hub of a complex signaling pathway that integrated various signals from across the cell. Rapamycin's impressive variety of bioactive effects, including anticancer, immune suppression, and anti-aging properties, is owed to its interaction with such an integral signaling protein. Eventually, Rapamycin was developed into an immunosuppressant drug to prevent transplant organ rejection and was approved by the FDA in 1999. A Rapamycin derivative was later approved to treat kidney cancer in 2009. Currently, Rapamycin is undergoing clinical trials to exploit its anti-aging properties.[23] While it is easy to focus on Rapamycin's incredible bioactivity, its impact on elucidating the mTOR signaling pathway stands as an outstanding contribution to our understanding of the basic processes of life, growth, and aging.

The relationship between Rapamycin and mTOR is just one of many examples in which a bioactive natural product uncovered previously unimagined advances in biology. Given its immense value to drug development and science in general, protein target identification has been the subject of extensive methods development efforts. Among the most common and preferred methods are affinity purification and genetic interaction methods.[24] Affinity purification starts by immobilizing the small molecule of interest on a gel or resin. Then a mixture of proteins, usually from cell lysate, is eluted over the immobilized small molecule. The proteins that are retained or "pulled down" are determined to be putative protein targets. This method is quite laborious and biases the target identification in favor of soluble proteins. Genomic techniques focus on identifying a genetic link between the small molecule's bioactivity and a particular protein. While this method is less biased, it is quite labor intensive. Moreover, proteins identified by this method are not necessarily the binding partner of the small molecule drug but may be biologically associated with the protein target. Given the challenges of traditional protein target identification methods, chemical biologists have been searching for a faster, less biased, and more reliable workflow.

1.4.1 Mass Spectrometry-Based Chemical Proteomics

Liquid chromatography mass spectrometry (LCMS) has emerged as one of the most powerful techniques in chemical biology. While its applications to metabolomics and proteomics are well-known, its application to small molecule protein target identification stands out as a masterful application of the technique. Chemical proteomics aims to identify which sites across the proteome a given bioactive small molecule binds to. The basic approach starts by incubating live cells with the bioactive small molecule, while leaving a control sample without the compound of interest. Then isotopically labeled reactive probes that covalently link to specific protein sites, such as reactive cysteines, are incubated with each cell line. The active sites that are obscured by the small molecule of interest are not available for derivatization with the probe, leading to a decreased amount of labeling in the test sample. Then, the cell samples are lysed, their proteins digested, and the samples combined for one LCMS run. The protein targets are then identified by comparing the isotopic labeling ratios of each protein. Those that were more weakly labeled with the isotopic probe used in the test sample are considered putative hits for further testing. This technique has been used to successfully identify many protein targets for small molecule natural products. For example, Withaferin A was determined to target PP2R1A in breast cancer cells, leading to growth inhibition.[25]

While competitive activity-based chemical proteomics has been enormously successful, it is subject to several limitations. The competitive nature of the assay means that the test is subject to variation within the control and test samples, including variations caused by the bioactive compound. It is also limited to studying a compounds' interaction with reactive side chains on proteins, such as cysteine or lysine. Recent work has used light-activated reactive groups, such as the diazirine ring, to track small molecules' non-covalent interactions across the proteome.[26] However, this technique is both synthetically challenging, and requires laborious data analysis to exclude non-specific interactions. While chemical proteomics represents an impressive leap forward for small molecule target identification, there remains ample room for innovative new approaches to this problem.

1.4.2 Tagging: Imaging Natural Products

An emerging technique to obtain mechanistic information about small molecule drugs, including natural products, is to image their distribution within live cells. Typically, the small molecule natural product is tagged with a fluorophore, as in a recent study of apoptolidins.[27] apoptolidins have drawn much attention for being highly cell-selective cytotoxic agents. The authors used selective esterification to derivatize apoptolidins A and H with a fluorescent functionality while retaining the compounds' bioactivity. Fluorescent imaging revealed the subcellular distribution and provided new insights into the compound's mode of action. While this was an impressive use of semi-synthetic natural product probes, this technique presents a few drawbacks. The large fluorescent moiety is prone to reduce bioactivity and alter the subcellular distribution of small molecules, especially those that are smaller than the relatively large apoptolidins. Therefore, the use of fluorescent probes to study smaller small molecules is undesirable.

Recently, the development of bioorthogonal Raman microscopy has allowed smaller molecules to be tagged with smaller vibrational tags, enabling the imaging of smaller small molecules.Raman microscopy is a cutting-edge imaging technique that takes advantage of the chemical specificity of Raman spectroscopy to visualize the chemical composition of cells. This technique has so far been very useful for visualizing protein distribution (via amide bonds and methyl groups) and lipid structures (via alkyl CH2 groups). Recently, many groups have begun using bioorthogonal Raman tags to selectively visualize chemical species in live cell. Most commonly, the Raman tag is a simple terminal alkyne. Alkynes are useful because they present a strong peak in the cell-silent region of the Raman spectrum, limiting interference from endogenous species. Unfortunately, terminal alkynes are limited to visualizing high-concentration species due to low signal intensity. In order to detect species at lower concentrations, Yamakoshi et. al. measured the relative Raman intensity of many conjugated alkyne tags and found them to be much more sensitive.[28] Adding a conjugated aromatic ring further enhanced the signal. This phenyl diyne probe is smaller than a fluorophore, but nearly 100 times as sensitive as a terminal alkyne, making it extremely useful for visualizing small molecules. Three years later, Lee et. al. used this tag to track cholesterol storage in C. Elegans.[29] In this dissertation, the structure-activity-distribution relationship of antimycin-type depsipeptides is elucidated via bioorthogonal Raman microscopy.

1.5 Rational Drug Design

To avoid the shortcomings of drugs discovered by cytotoxicity assays, researchers began turning to targeted drug discovery, searching for compounds that inhibited pharmacologically relevant proteins. The most famous example of this approach was the development of Imatinib, a compound that inhibits the protein BCR-ABL. Notably, researchers set their sights on BCR-ABL because a hereditary mutation to the gene was known to be the singular cause of Chronic Myelogenous Leukemia. The resulting drug, marketed under the name Gleevec, is remarkable for its selectivity in killing cancer cells. This selectivity allowed for effective cancer treatment with relatively minor side effects compared to traditional cytotoxic chemotherapies. Indeed, Gleevec is unique among chemotherapies because patients can take the drug throughout the course of their lives ensuring continuous protection against relapse.

Today, there are many rational drug development programs targeting diverse protein classes including tyrosine kinases, farnesyl-protein transferases, and many more. Most of these research programs rely on high throughput protein binding assays. Although the methods vary slightly, the central characteristic of these protein binding assays is that a compound library is screened for compounds that bind strongly to a purified protein. Despite its success in generating new lead compounds, this approach fails to address three key problems in pharmacology:

- Proteins often exist as part of complex structures inside the cell, so binding to a pure protein *ex vivo* may not be conserved *in vivo*.
- There are countless proteins in the cell and many opportunities for off-target binding.

- Effective drugs must be able to cross the cell membrane and resist quick metabolic decomposition within the cell.

The next generation of anticancer drug screening methods will need to address these challenges by examining the interaction of cells with specific targets inside living cells. Cell-based assays like chemical proteomics are now employed for drug discovery purposes as well as target identification. However, they are limited by synthetic challenges, and an emphasis on covalent ligands. There remains a pressing need for screening and target ID methods capable of identifying non-covalent ligands for valuable therapeutic targets. Several new techniques, including some developed in this dissertation, aim to address this challenge directly.

1.5.1 Small Molecule Microarrays

Increasing throughput has been a major goal of drug discovery assays for many decades. Small molecule microarrays attempt to rise to this challenge by chemically bonding diverse small molecule libraries to a solid surface. In this manner arrays of tens of thousands of small molecules are packaged into a highly organized, compact format. Cell lysates can then be washed across the microarray. This allows for many proteins, including disordered proteins and protein complexes that are short-lived outside of the cell, to be screened against a massive small molecule library in one step. Proteins captured by the microarray are either fluorescent fusion proteins or labeled via immunofluorescence reagents so that the ligand-protein pair of interest can be identified via fluorescence microscopy.[30] Small molecule microarrays have proven useful for identifying ligands against MYC or other transcription factors that are typically considered undruggable proteins.[31] Small molecule microarrays screening against cell lysate are considered ideal for discovering ligands against transcription factors, because these intrinsically disordered proteins exist as more stable complexes inside the cell. The success of small molecule microarrays stands as a testament to the value of identifying ligands that bind to proteins in a more natural environment.

1.5.2 Tagging: Cell Binding Assays

A great deal of the success of traditional drug screens is based on their rigorous criteria for selection. A compound must cause a measurable phenotypic change to be discovered during a phenotypic assay. Similarly, it must bind to a specific protein target to be discovered during a ligand binding assay. However, these strict selection rules have also greatly limited the range of drug compounds that can be discovered. While small molecule microarrays made great strides in addressing these challenges, they are limited by their solid-state format; a physical surface precludes the molecules' diffusion into living cells, the most biologically relevant environment short of living tissue.

For this reason, we have developed the cell binding assay as an intentionally broader, less strenuous initial screen for bioactive compounds. Before the cell binding assay, a natural product extract is tagged with bioorthogonal stable isotope tags which allows compounds from the extract to be easily separated from endogenous compounds via LC/MS analysis. During the cell binding assay, compounds are filtered out based on their affinity to a living cell. The logic is quite simple:

in order to affect any form of bioactivity, a small molecule drug must interact with the cell by binding to proteins within the cell, so non-binding compounds can be ignored. The cell binding assay doesn't discriminate based on a specific protein target or rely easily measurable phenotypic outcome. Instead, all possible protein targets and all possible phenotypic outcomes are considered at once.

1.5.3 Tagging: FRET-based Ligand Binding Assays

Here, we propose a new approach to protein binding assays that addresses these problems: perform the protein binding assay in live cells. This technique once again hinges on the ability to chemically tag small molecule natural product – this time with a fluorescent label. It is possible to detect the binding of a fluorescently labeled small molecule to a specific GFP-labeled protein amidst the intricate cellular machinery using Forster Resonance Energy Transfer (FRET). The basic principle of this screening platform is that small molecule-protein interactions can be measured via FRET when both the drug and the protein are labeled with compatible fluorophores or fluorescent proteins.

FRET relies on the transfer of energy between two fluorophores in close proximity (usually within 10 nm). FRET's nickname, "the molecular ruler," refers to this strict dependence on the distance between the two fluorophores.[32] In practical terms, FRET can be observed as the emission of light from a low-energy acceptor fluorophore while exciting only the high-energy donor fluorophore. Similarly, a corresponding drop in the donor fluorophore emission intensity can be observed.

1.6 Chimeric Small Molecules

Traditional ligand binding assays, small molecule microarrays, the cell binding assay, and FRET-based ligand binding assays all share a major limitation: they screen for ligand binding, not a change in protein function. The entire surface of the protein, not just the active site, is considered in these assays. As such, many compounds discovered by these methods will ultimately be inactive ligands which bind to a protein without changing its function. However, a new drug modality, the chimeric small molecule, holds the promise of imparting bioactivity to inactive binders.

Chimeric small molecules are comprised of two separate small molecules conjoined by a flexible linker, such as PEG. One side of the chimeric small molecule binds to the protein which is targeted for functional alteration while the other side recruits a second protein that effects that protein function. Essentially, chimeric small molecules serve as a catalyst to bring two proteins together and encourage their interaction. In the original example of this paradigm, the protein target and a ubiquitin ligase are recruited by separate sides of the chimeric small molecule, encouraging ubiquitination of the target protein.[33] Ubiquitination marks the protein target for degradation, effectively causing a decrease in protein function. Since the advent of targeted protein degradation, many new drug modalities have been proposed for chimeric small molecules.[34] However, the basic premise remains the same: inactive small molecule ligands are converted into active drug-like molecules by recruiting a second protein.

From a strategic perspective, chimeric small molecules have opened the entire protein surface for ligand screening. This development will contribute to growing interest in ligand binding assays and other non-phenotypic drug screening methods. Indeed, chimeric small molecules represent an especially attractive end-use for FRET-based ligand screening for a variety of reasons. FRET-based screening will allow for natural product ligand discovery against high-value protein targets, including membrane proteins and intrinsically disordered proteins which are notoriously difficult to study outside of living cells. These ligands can then be incorporated into chimeric small molecules that will create an active drug, regardless of where the ligand engages the protein. Just as chimeric small molecules have expanded the targetable landscape of proteins beyond the active site, they will spur a revolutionary change in the methods and means of drug discovery assays.

1.7 References

1. G. M. Cragg & J. M. Pezzuto. *Med Princ and Pract,* 2016, **25(2)**, 41-59.
2. F. Grunewald et al. *Intern J Med Mushrooms,* 2018, **20(12)**: 1135-1147.
3. G. Samorinini. *Integration*, 1992, **2(3)**, 69-78.
4. Zhuk O, Jasicka-Misiak I, Poliwoda A, et al. *Toxins (Basel)*, 2015, **7(4)**, 1018-1029.
5. Bruhn, Jan & Smet, Peter & El-Seedi, Hesham & Beck, Olof. *Lancet*, 2002, 359. 1866.
6. Miller, Melanie J.; Albarracin-Jordan, Juan; Moore, Christine; Capriles, José M. *Proceedings of the National Academy of Sciences*, 2019, **116 (23)**, 11207–11212.
7. Carod-Artal, FJ. *Neurologia,* 2015, **30(1)**, 42–9.
8. Dillehay; et al. *Antiquity,* 2010, **84(326)**, 939–953.
9. Houyuan Lu et al. *Nature,* 2016, **6**, 8955.
10. Ren M, Tang Z, Wu X, Spengler R, Jiang H, Yang Y, Boivin N. *Science Advances*, 2019, **5(6)**, 1391.
11. J. Mahdi. *Journal of Saudi Chemical Society*, 2010, **14(3)**, 317-322.
12. P. G. Kritikos & S. P. Papadaki. *Journal of the Archaeological Society of Athens,* 1967.
13. Lois N. Magner. *A History of Medicine*. 1992. CRC Press. p. 91.
14. J. Wang, C. Xu, Y. K. Wong, et al. *Engineering,* 2019, **5(1)**, 32-39.
15. E. Martino, S. D. Volpe, E. Terribile, et al. *Bioorg Med Chem Lett*, 2017, **27(4)**, 701-707.
16. C. Crishnamurti, and SSC Chakra Rao. *Indian J Anaesth*, 2016, **60(11)**, 861–862.
17. E. Russo. *Chemistry & Biodiversity*, 2007, **4(8)**, 1614-1648.
18. V. Di Marzo. *Trends in Pharmacological Sciences,* 2006, **27(3)**, 134-140.
19. E. Perucca. *J Epilepsy Res*, 2017, **7(2)**, 61-76.
20. R. Gaynes. *Emerg Infect Dis*, 2017, **23(5)**, 849-853.
21. K. Gould. *Journal of Antimicrobial Chemotherapy*, 2016, **71(3)**, 572–575.
22. E.J. Brown et al. *Nature*, 1994, **369**, 756–8.
23. M. B. Blagosklonny. *Aging*, 2019, **11(19)**, 8048-8067.
24. M. Schenone et al. *Nat Chem Biol*, 2017, **9(4)**, 232-240.
25. E. A. Grossman. *Cell Chem Biol*, 2017, **24(11)**, 1368-1376.
26. J. Gao et al. *J. Am. Chem. Soc.*, 2018, **140(12)**, 4259-4268.
27. S. M. DeGuire et al. *Angew Chem Int Ed Engl*, 2015, **54(3)**, 961-964.
28. H. Yamakoshi et al. *JACS*, 2012, **134(51)**, 20681-20689.

29.   H. J. Lee et al. *Scientific Reports*, 2015, **5**, 7930.
30.   M. Uttamachandani et al. *Current Opinion in Chemical Biology,* 2005, **9(1)**, 4-13.
31.   N. B. Struntz et al. *Cell Chemical Biology,* 2019, **26(5),** 711-723.
32.   L. Stryer & R. P. Haugland. *Proc Natl Acad Sci U S A,* 1967, **58**, 719–726.
33.   K. M. Sakamoto et al. *Proc Natl Acad Sci U S A*, 2001, **98(15)**, 8554-9.
34.   C. Borsari et al. *J. Med. Chem.,* 2020, **63(5)**, 1908-1928.

Chapter 2: Vibrational tags to aid in spectroscopic imaging
Structure–activity–distribution relationship study of anti-cancer antimycin-type depsipeptides

Nature's small molecules have played an enormous role in the history of medicinal and pharmaceutical chemistry. For example, it has been estimated that over 70% of anti-cancer small molecule treatments are natural products, their derivatives or mimics.[1] Although the pharmaceutical value of natural products has been widely recognized, it is still difficult to transform medicinally active natural products into drugs. One of the major challenges is to understand the complex interplay between natural products and the network of cellular machinery beyond the specific protein targets.[2] This has spurred the development of advanced imaging techniques to obtain views of natural products in cells, but often in a static and destructive manner using bulky fluorescent probes.[3,4] An improved imaging technique, which provides dynamic views of natural product uptake and distribution in live cells, will have a profound impact on natural product-based drug discovery and development.[5] Tagging natural products with a bio-orthogonal alkyne functionality coupled with the Stimulated Raman Scattering (SRS) microscopy offers such promise.[6,7]

Raman imaging has evolved greatly over the past decade, with improved sensitivity, resolution, and scanning speeds offered by the latest SRS technology.[8-14] More importantly, compared with other vibrational imaging platforms such as Coherent Anti-Stokes Raman Scattering microscopy, SRS is free of spectral distortion and non-resonant background thus enables quantitative determination of Raman reporters. The vibrational Raman reporter can be as simple as an alkyne,[6,15] delivering chemical specificity and biocompatibility for natural product visualization and quantification in complex living systems with minimal activity perturbation of compounds. Compared to fluorescent imaging, SRS imaging offers additional advantages of minimal phototoxicity and photobleaching, allowing prolonged dynamic imaging of tagged natural products within live cells. SRS microscopy has recently been used to image various alkyne-tagged small-molecule derivatives that have high local intracellular concentrations.[16-21] Despite the great potential, few natural products have been imaged using SRS microscopy to probe their intracellular behavior.

Here, we've applied SRS imaging to study antimycin-type depsipeptides, a class of complex natural products that have attracted recent attention due to their anti-cancer potential.[22] This family of natural products share a common structural skeleton consisting of a macrocyclic ring with an amide linkage to a 3-formamidosalicylate unit, and primarily differ in the size of their macrolactone ring (Figure 2-1). The well-recognized members of this family are the 9-membered antimycins, for which multiple modes of action have been proposed, including inhibition of mitochondrial electron transport chain,[23] anti-apoptotic proteins Bcl2/Bcl-xL,[24] K-Ras plasma membrane localization,[25] and ATP citrate lyase activity.[26] The levels of contributions of these different mechanisms are unclear. Much less is known about the 15-membered neoantimycins, despite the fact that they have also shown promising anti-cancer activities toward various cancer cell lines.[27] The inhibitory activity of K-Ras plasma membrane localization was shown to be shared between antimycins and neoantimycins,[25] but neoantimycins lacked the Bcl-xL inhibitory activity

and were demonstrated to inhibit the expression of GRP-78,[28,29] a molecular chaperone in the endoplasmic reticulum (ER) that promotes protein folding and provides resistance to both chemotherapy and hypoglycemic stress.[30] To gain additional insights into anti-cancer activities of antimycin-type depsipeptides, we performed an SADR study of both antimycin and neoantimycin against live cancer cells.
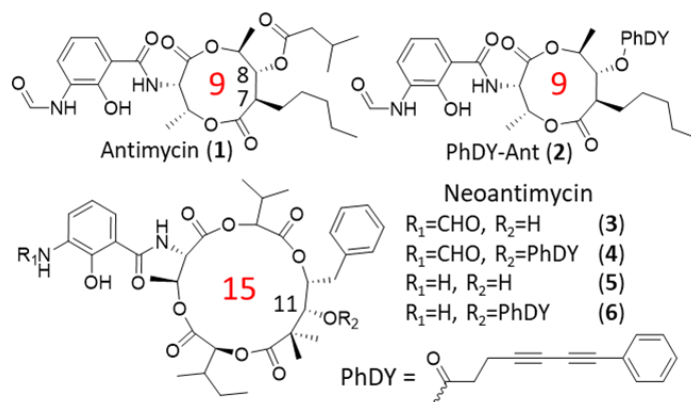
Figure 2-1. Structures of selected antimycin-type depsipeptides and their alkyne-tagged derivatives. The ring size is indicated in red numbers.

2.1 Probe Design and Validation

To increase the sensitivity of SRS imaging toward bioactive small molecules which are typically used in the low to mid micromolar range,[31] we chose to employ a conjugated diyne with a terminal phenyl ring as a Raman tag. This tag is known to possess an increased Raman scattering cross section due to conjugation within the poly-yne chain and the presence of an aryl end-capping group also improves the stability of poly-ynes.[16] This tag has recently been used in separate studies to image a phenyl-diyne anisomycin derivative in mammalian cells and to track the distribution of a cholesterol derivative in *Caenorhabditis elegans* where a detection limit of ~30 μM was attained.[16,32] Since bioactive antimycins naturally have high structural variations at the C-7 alkyl and C-8 acyloxy moieties (Figure 2-1) and the previous introduction of an alkyne side chain at C-8 did not significantly change cytotoxic activity nor binding of compounds to cancer cells,[4] we reasoned that the Raman tag could be readily introduced at this position with minimal functional perturbation. To prepare phenyl-diyne antimycin (PhDY-Ant, **2**), C-8 deacylated antimycin was first purified from the culture of *Streptomyces albus ΔantB* in which the last step of C-8 acyloxy formation is abolished in antimycin biosynthesis due to the deletion of the dedicated C-8 acyltransferase AntB.[33] A phenyl-diyne carboxylic acid was chemically synthesized and then coupled to the purified deacylated antimycin via Steglich esterification to yield **2** (Scheme 2-1, Figure S1). As expected, the MTT proliferation assays with both HeLa (human cervical cancer) and MCF-7 (human breast cancer) cell lines confirmed that PhDY-Ant retained a comparable activity to the natural antimycin (Table 2-1, Figure S2).
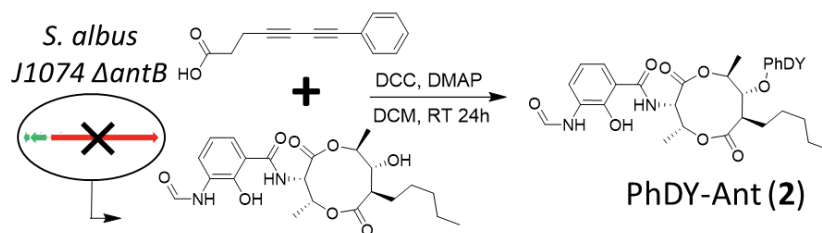
Scheme 2-1. Synthesis of PhDY-Ant (2).

Table 2-1. GR50 values (µM) for selected antimycins and neoantimycins against two cancer cell lines in vitro.

Compound	HeLa	MCF-7
Antimycin (**1**)	20.2 ± 4.3^a	2.1 ± 0.5
PhDY-Ant (**2**)	31.8 ± 10.4	28.8 ± 9.7
Neoantimycin (**3**)	38.7 ± 8.2	24.4 ± 5.5
PhDY-NeoA (**4**)	$> 100^b$	40.9 ± 22.8
Deformylated NeoA (**5**)	> 1000	> 1000

[a] All data were collected in at least triplicate and GR50 values were averaged between at least three biological replicates with error given in standard error of the mean (SEM).

[b] GR50 could not be determined due to solubility limitations.

PhDY-Ant (**2**) was then incubated with HeLa cells and SRS images were acquired by tuning the frequency difference between the pump and Stokes lasers to be resonant with intracellular components such as proteins (CH$_3$, 2940 cm^{-1}) and lipids (CH$_2$, 2845 cm^{-1}), and in the bio-orthogonal region of the Raman spectrum (alkyne, 2251 cm^{-1}; off-resonance, 2000 cm^{-1}). **2** was used at solution concentrations ranging from 1-100 µM to probe the detection limit. Intracellular signal could be distinguished at concentrations as low as 10 µM, and contrast was dramatically improved by increasing the solution concentration of **2** to 50 µM (Figure 2-2). The absolute intracellular concentration of **2** was determined to be ~ 1.74 mM from the dosing concentration of 50 µM, showing a 35-fold enrichment of this compound in cells. To confirm that the observed signal was driven by the activity of antimycin, a control experiment was performed by incubating 50 µM PhDY tag with cells. No signal of compound was detected inside cells (Figure 2-2), suggesting that the observed Raman signal was not an off-target affect caused by the PhDY tag alone. We next probed the antimycin uptake rate and mechanism. Time-resolved imaging of **2** uptake into live HeLa cells showed that compound uptake was nearly immediate, reaching 75% of the maximum within six minutes (Figure 2-3). **2** appeared to rapidly distribute throughout the cytoplasm of the cells and persist through prolonged incubation. In addition, a low-temperature (4°C) uptake study was performed to investigate possible mechanisms of compound uptake. **2** was

15

absorbed at comparable levels at both 4°C and 37°C (Figure 2-4), suggesting that PhDY-Ant crossed the cell membrane through passive diffusion.
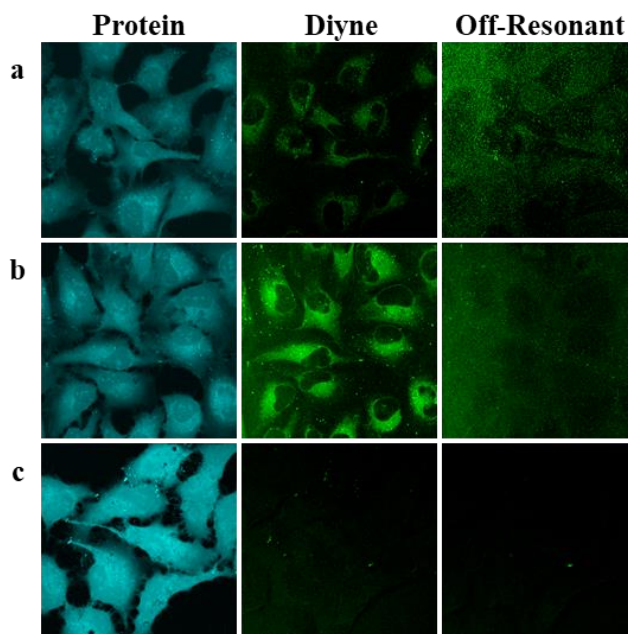
Figure 2-2. Intracellular PhDY-Ant (**2**) and phenyl-diyne tag imaging at various concentrations in HeLa cells. Cells were imaged 2 h post PhDY-Ant addition at 10 µM (a) and 50 µM (b). Phenyl-diyne tag (50 µM) was imaged as a negative control (c). SRS images were acquired by tuning the frequency difference between the pump and Stokes lasers to be resonant with protein (CH$_3$, 2953 cm^{-1}), diyne (2251 cm^{-1}), off-resonance (2000 cm^{-1}).
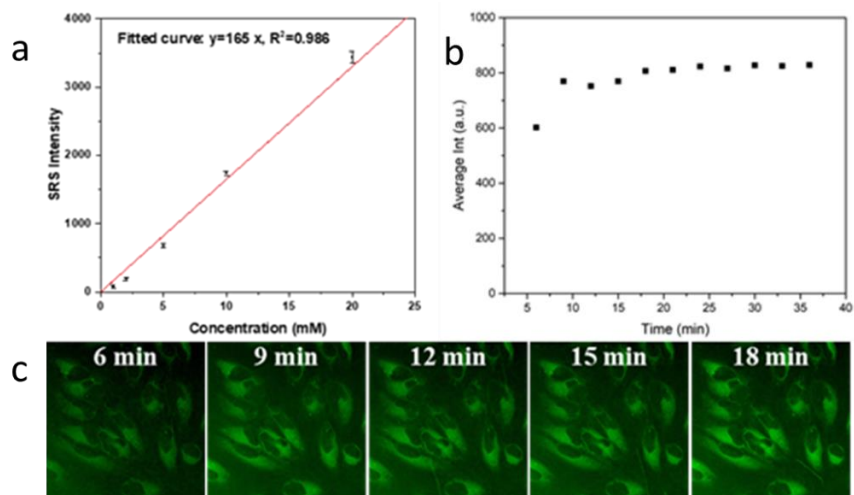
Figure 2-3. Time course of PhDY-Ant (**2**) uptake in HeLa cells. a) Calibration curve showing relationship of PhDY-Ant concentration with signal intensity. b) Average signal intensity over time. c)SRS images used to analyze the time course experiment. HeLa cells incubated with 50 µM PhDY-Ant (**2**).
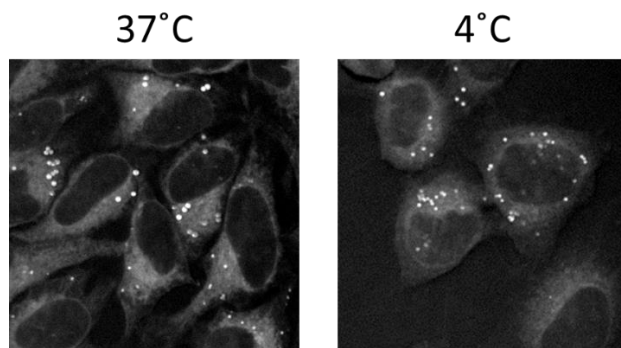
Figure 2-4. Temperature dependent uptake of PhDY-Ant (**2**). SRS images showing uptake of 50 µM **2** during 2h incubation at 4˚C and 37 ˚C. **2** was still absorbed into HeLa cells at 4˚C, indicating that the uptake mechanism is energy-independent.

The non-destructive nature of SRS imaging allows follow-up studies on compound distribution inside the cell, as well as correlating compound uptake with any possible phenotypic changes to cellular composition using dual-color and multi-modal approaches. For example, using a multi-modal approach to probe the subcellular localization of **2**, HeLa cells were treated with ER-Tracker Green and Mito-Tracker Deep Red, cell-permeable fluorescent stains selective for the endoplasmic reticulum and mitochondria, respectively. Inspection of the merged images demonstrated that **2** correlated well with ER-Tracker (Figure 2-5). This colocalization agrees with one of its known direct protein targets, Bcl$_2$, an anti-apoptotic protein localized primarily in the ER.[34] Notably, a prolonged incubation led to a decrease of correlation between ER-Tracker and **2** (Figure 2-6), possibly due to compound dissociation from targets and/or translocation of protein targets. In spite of the known activity of antimycin in inhibiting the mitochondrial electron transport chain by binding to the quinone reduction site Q_i of the cytochrome bc_1 complex,[23] no significant colocalization of **2** with Mito-Tracker was found (Figure 2-5). In addition, no obvious phenotypic changes were observed in lipids or proteins during the eight-hour incubation, although the merged images for **2** and the lipid channel showed a strong correlation, especially in lipid droplets (Figure 2-2). This correlation is likely caused by the lipophilic nature of **2** rather than any specific binding. Further image analysis using profile plots demonstrated that the localization of **2** was best explained by a combination of ER-Tracker and lipids (Figure 2-7). This result showed that the intracellular behavior of antimycin was not totally dictated by specific protein binding nor non-specific absorption. Instead, there was a complex interplay between antimycin, its protein targets, and the lipid-rich regions of the cell.
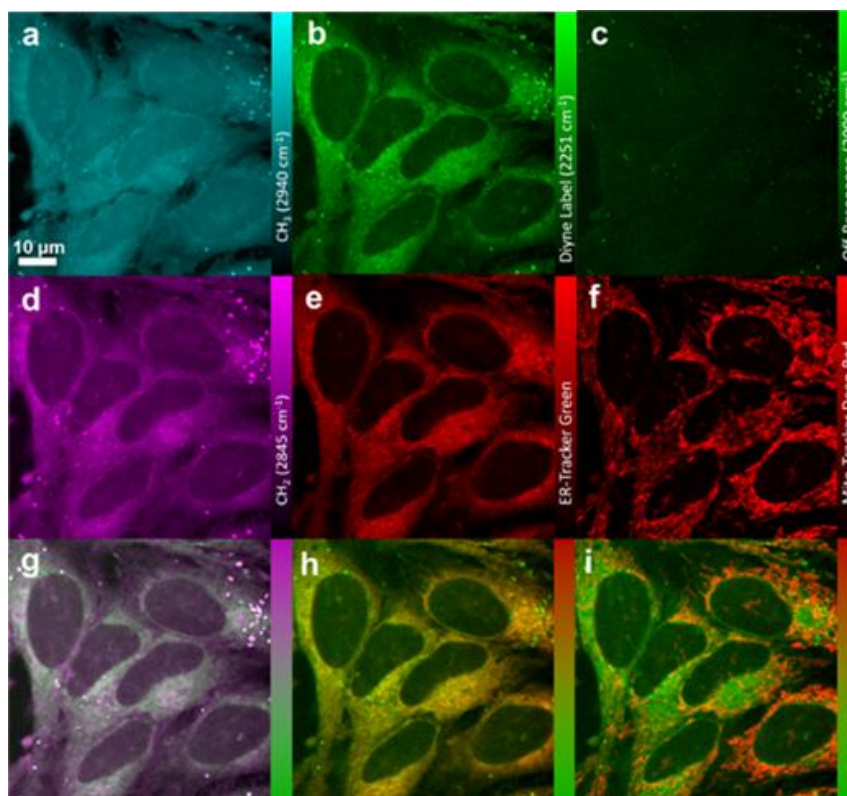
Figure 2-5. SRS and fluorescence imaging of PhDY-Ant (**2**) in HeLa cells. **a**) CH$_3$ channel at 2940 cm^{-1} representing proteins. **b**) Diyne label at 2251 cm^{-1}. **c**) Off-resonance channel at 2000 cm^{-1}. **d**) CH$_2$ channel at 2845 cm^{-1}, representing lipids. **e**) Confocal fluorescence imaging of ER-Tracker excited at 488 nm. **f**) Confocal fluorescence imaging of Mito-Tracker excited at 635 nm. **g**) Overlay image of d) lipids and b) diyne label. **h**) Overlay image of e) ER-Tracker and b) diyne label. **i**) Overlay image of f) Mito-Tracker and b) diyne label.

Figure 2-6. Time-dependent colocalization of PhDY-Ant (**2**) in HeLa cells. Pearson's correlation coefficients were calculated using the coloc2 plugin for imageJ in triplicate.

Figure 2-7. Line plot analysis of PhDY-Ant (**2**) in HeLa cells. Each line plot shows poor correlation between Mitotracker and the diyne channel. Each line plot shows good correlation between the lipid channel and the diyne channel. The ER-tracker and diyne channels are very well-correlated, including some features that are excluded on the lipid channel.

We next analyzed the distribution of PhDY-Ant (**2**) in MCF-7 and compared it to HeLa cells to probe if the distribution is cell-line specific. Similar to HeLa cells, 2 colocalized with ER-Tracker, but not Mito-Tracker (Figure 2-8). These data suggest that localization of antimycin in the ER is conserved across different cancer cell lines. In addition, MCF-7 cells showed a much smaller number of lipid droplets than HeLa cells, but instead contained highly lipid-rich regions at the intercellular boundaries that did not attract **2** (Figure 2-8). This result further suggests that the characteristics and location of lipids are also important for enrichment of antimycin. Indeed, a profile analysis showed that certain areas of localization of **2** in MCF-7 cells were best correlated with ER-Tracker while others were best correlated with the lipid channel (Figure 2-9).

19
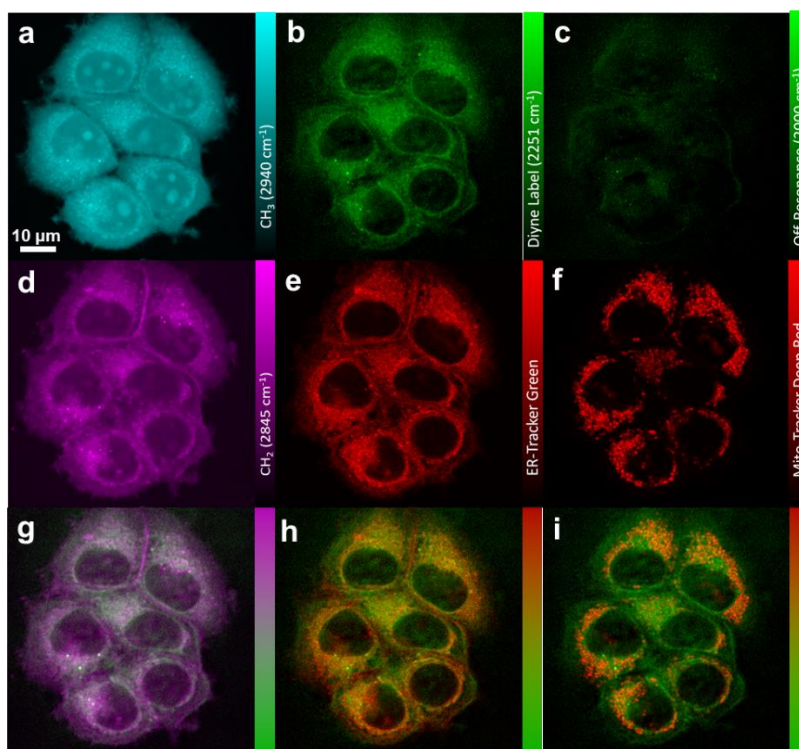
Figure 2-8. SRS and fluorescence imaging of PhDY-Ant (**2**) in MCF-7 cells. a) CH3 channel at 2940 cm-1 representing proteins. b) Diyne label at 2251 cm-1. c) Off-resonance channel at 2000 cm-1. d) CH2 channel at 2845 cm-1, representing lipids. e) Confocal fluorescence imaging of ER-Tracker excited at 488 nm. f) Confocal fluorescence imaging of Mito-Tracker excited at 635 nm. g) Overlay image of d) lipids and b) diyne label. h) Overlay image of e) ER-Tracker and b) diyne label. i) Overlay image of f) Mito-Tracker and b) diyne label.

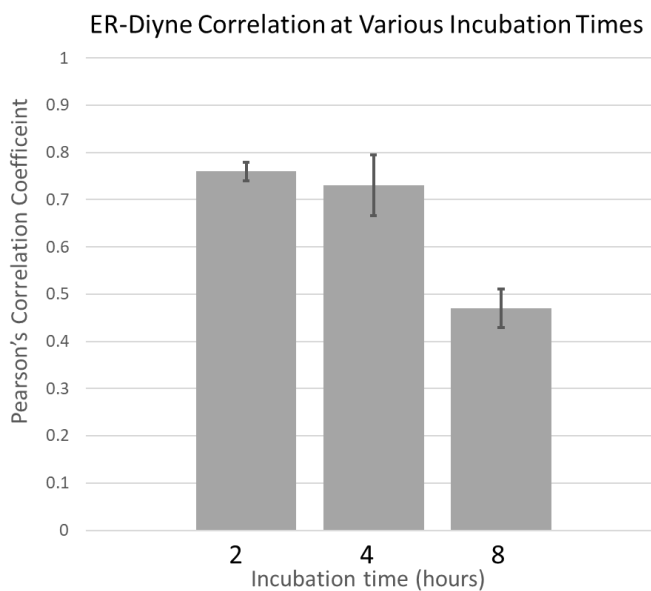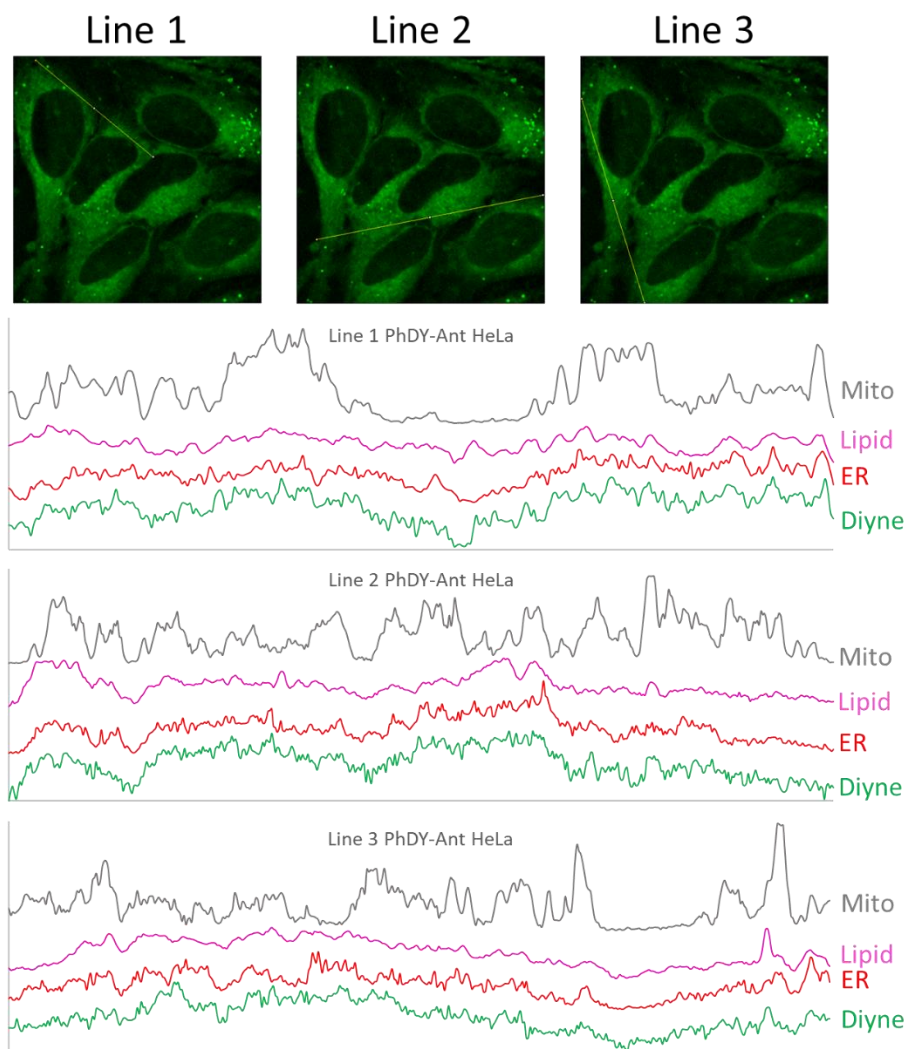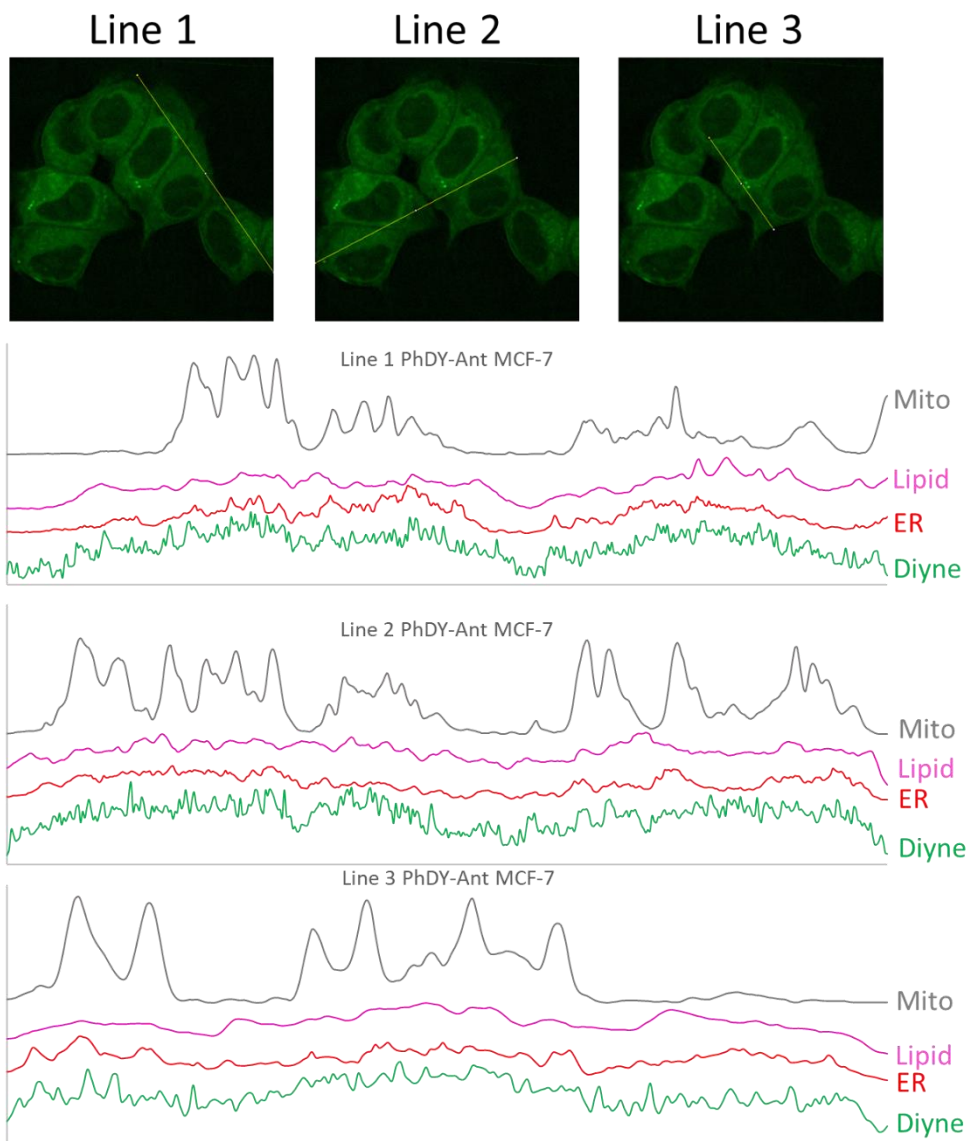**Figure 2-9.** Line plot analysis of PhDY-Ant (**2**) in MCF-7 cells. Each line plot shows poor correlation between Mitotracker and the diyne channel. The features of the diyne channel can be explained by a combination of the lipid and ER-tracker cannels.

Compared to antimycin, the molecular mechanism for the ring-expanded neoantimycin to inhibit cancer cell growth is much less known, and no direct protein target of neoantimycin has been identified. In addition, limited structure-activity relationship studies have been performed on the molecular scaffold of neoantimycin. It is yet to be determined if a similar tagging strategy, the esterification of the macrolactone C-11 hydroxyl moiety that is naturally present in neoantimycin (Figure 2-1), can be adopted to produce a neoantimycin derivative that is suitable for imaging analysis while retaining its anti-cancer activity. Neoantimycin was purified from the culture of *S. orinoci* and subjected to esterification by a phenyl-diyne carboxylic acid to generate phenyl-diyne

neoantimycin (PhDY-NeoA, **4**) (Scheme 2-2, Figure S3). The MTT proliferation assays with both HeLa and MCF-7 cells indicated that **4** had a slightly decreased but significant bioactivity, although its GR50 value against HeLa cells could not be determined due to solubility limitation (Table 2-1, Figure S4). This response of cell lines to the C-11 modification is consistent with a recent report in which oxidation of the same hydroxyl to ketone of neoantimycin led to slightly increased IC_{50} values against multiple cancer cell lines.[27] The *N*-formyl group has been conserved in the antimycin-type depsipeptides and linked to respiration inhibition for antimycin.[22] To probe the role of the *N*-formyl group in anti-cancer activities of neoantimycin, we produced deformylated neoantimycin (**5**) and its tagged version (**6**) through acid degradation of **3** and **4**, respectively (Scheme 2-3, Figure 2-10, Figure S5). The growth of HeLa and MCF-7 cells was not inhibited upon treatment of up to 1 mM of **5** (Table 2-1, Figure S4), demonstrating the critical role of this moiety for anti-cancer activity of the 15-membered neoantimycin. This is in contrast to the 18-membered antimycin-type depsipeptides of which deformylation did not significantly decrease the inhibitory activity toward various cancer cell lines,[35,36] suggesting different modes of action for 15- and 18-membered compounds. Nonetheless, the generation of both active and inactive tagged neoantimycins provided an opportunity to investigate possible differential uptake of these compounds in live cells.
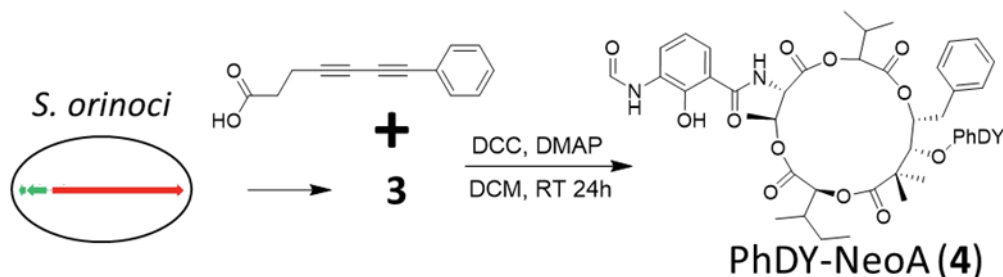
Scheme 2-2. Synthesis of PhDY-NeoA (**4**).
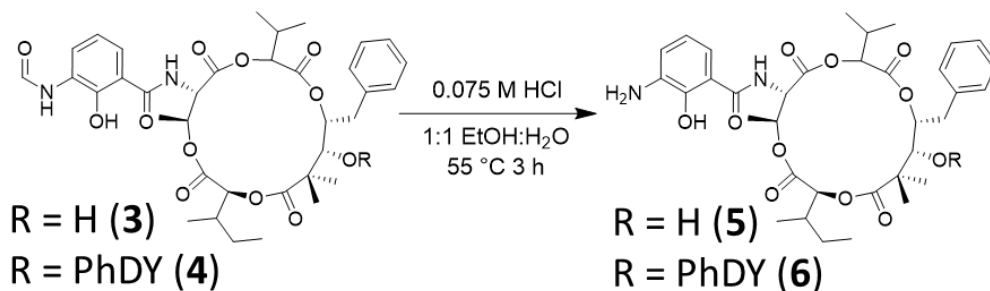
Scheme 2-3. Synthesis of deformylated neaontimcyin (**5**) and deformylated PhDY-NeoA (**6**).

22
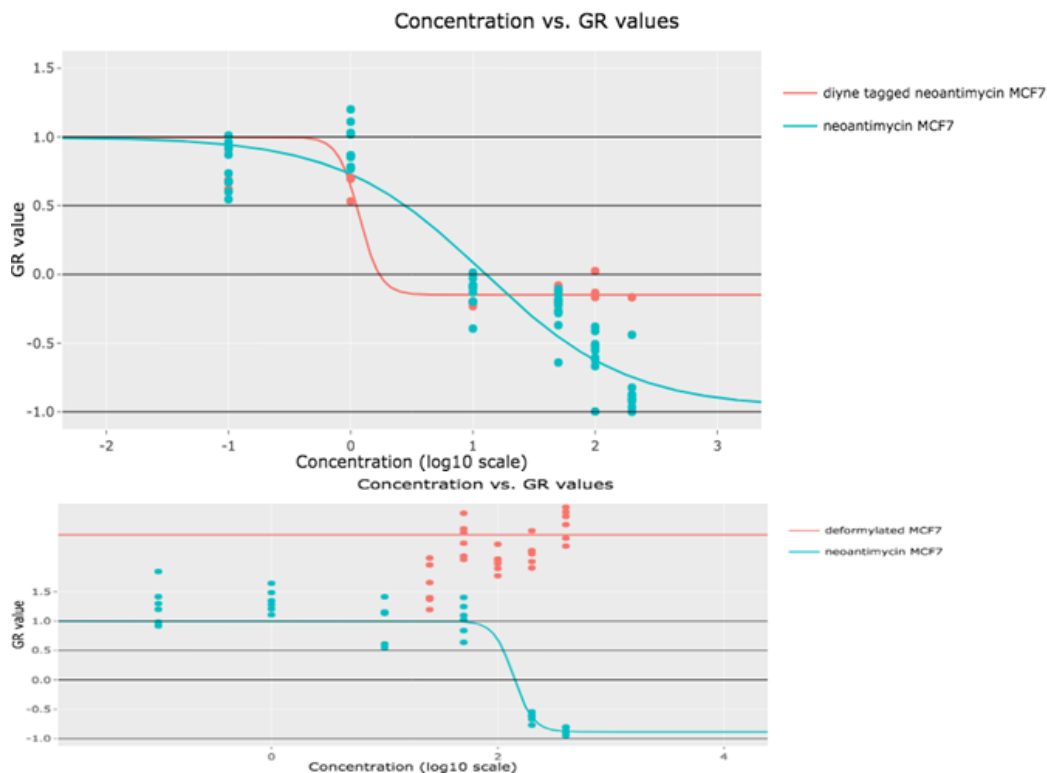
Figure 2-10. Dose response curve for various neoantimycins in MCF-7 cells. Concentration of compounds is in µM (log scale). Data were collected in triplet to sextuplet and growth response curves were generated as described in methods. The curves shown are representative of three individual biological replicates. A) Neoantimycin (**3**) and PhDY-NeoA (**4**) (200 µM to 0.10 µM), dose response curves in MCF-7 cells show the Raman-active tag at C-11 is minimally perturbing. The plateau at higher concentrations of PhDY-NeoA is most likely due to solubility limits of the compound. B) Neoantimycin (**3**) (1.0 mM to 0.10 µM), and deformylated neoantimycin (**5**) (400 µM to 25 µM) dose response curves in MCF-7 cells demonstrate the importance of the *N*-formyl group.

PhDY-NeoA (**4**) and deformylated PhDY-NeoA (**6**) were then subjected to SRS imaging analysis with both HeLa and MCF-7 cells. Enrichment of both compounds in lipid droplets was observed for both cell lines, which is likely due to the lipophilic nature of compounds rather than any specific binding to targets (Figure 2-11 and Figure 2-12). In addition, **4** was detected by SRS throughout the cytoplasm of the MCF-7 cells with an estimated intracellular concentration of ~ 0.82 mM from a dosing concentration of 100 µM. **4** was also detected throughout the cytoplasm of the HeLa cells, although with a decreased signal intensity, demonstrating a positive correlation of compound intracellular enrichment to its cytotoxic activity. This is particularly true for intracellular enrichment of **6**, for which very weak SRS signals were detected (except within lipid droplets) in either cell line (Figure 2-11). Further analysis using dual-color and multi-modal approaches showed that the intracellular distribution of PhDY-NeoA (**4**) differed from that of PhDY-Ant (**2**). In particular, **4** showed no correlation with Mito-Tracker and a very weak correlation with ER-Tracker in MCF-7 cells (Figure 2-12). Line plot analysis showed that **4** was much better correlated with lipids than with ER-Tracker (Figure 2-13), suggesting that neoantimycin does not target ER. This observation is consistent with previous reports that the known antimycin target, Bcl$_2$/Bcl-x$_L$, is not a neoantimycin protein target.[37]
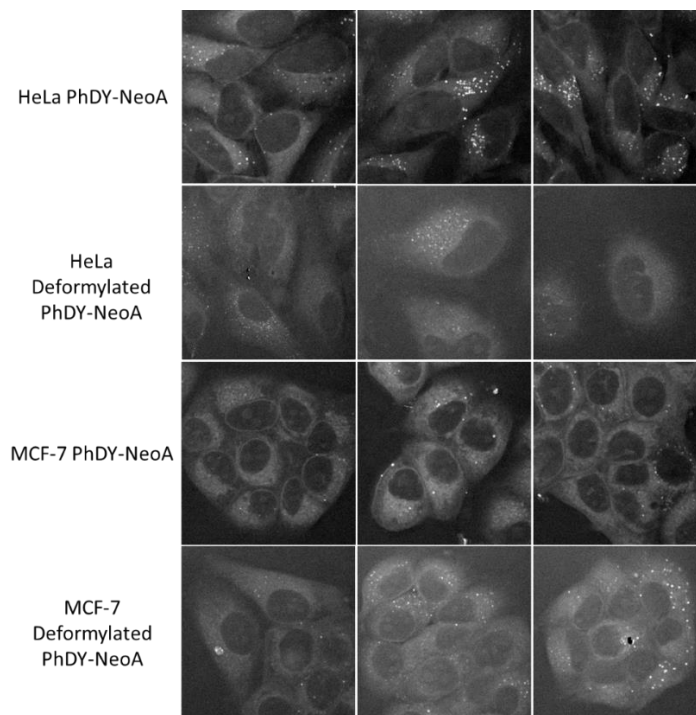
Figure 2-11. Raman imaging of the diyne channel at 2251 cm^{-1} shows the differential uptake of PhDY-NeoA (**4**) and deformylated PhDY-NeoA (**6**) in HeLa and MCF-7 cells.
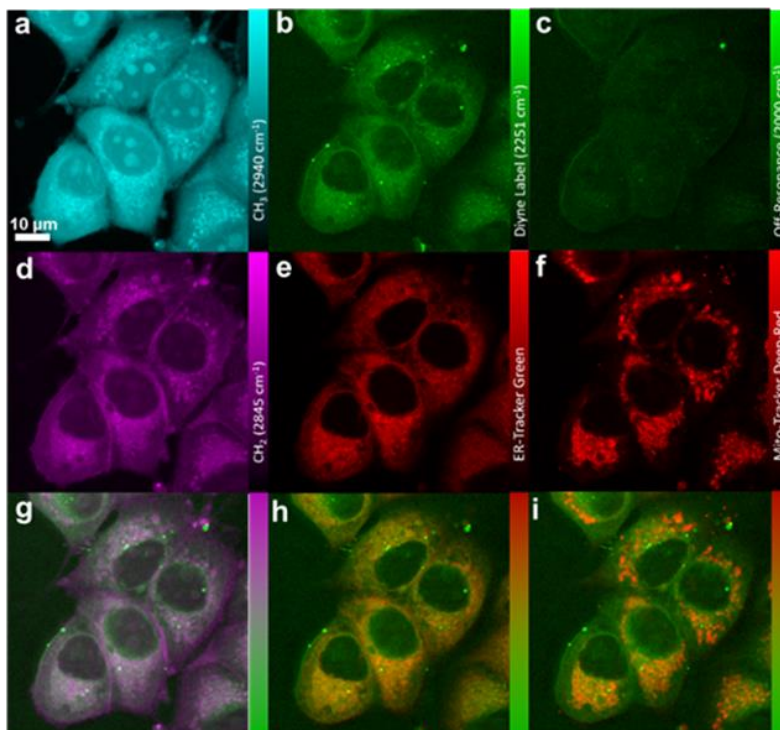
Figure 2-12. SRS and fluorescence imaging of PhDY-NeoA (**4**) in MCF-7 cells. **a**) CH$_3$ channel at 2940 cm^{-1} representing proteins. **b**) Diyne label at 2251 cm^{-1}. **c**) Off-resonance channel at 2000 cm^{-1}. **d**) CH$_2$ channel at 2845 cm^{-1}, representing lipids. **e**) Confocal fluorescence imaging of ER-Tracker excited at 488 nm. **f**) Confocal fluorescence imaging of Mito-Tracker excited at 635 nm. **g**) Overlay image of d) lipids and b) diyne label. **h**) Overlay image of e) ER-Tracker and b) diyne label. **i**) Overlay image of f) Mito-Tracker and b) diyne label.
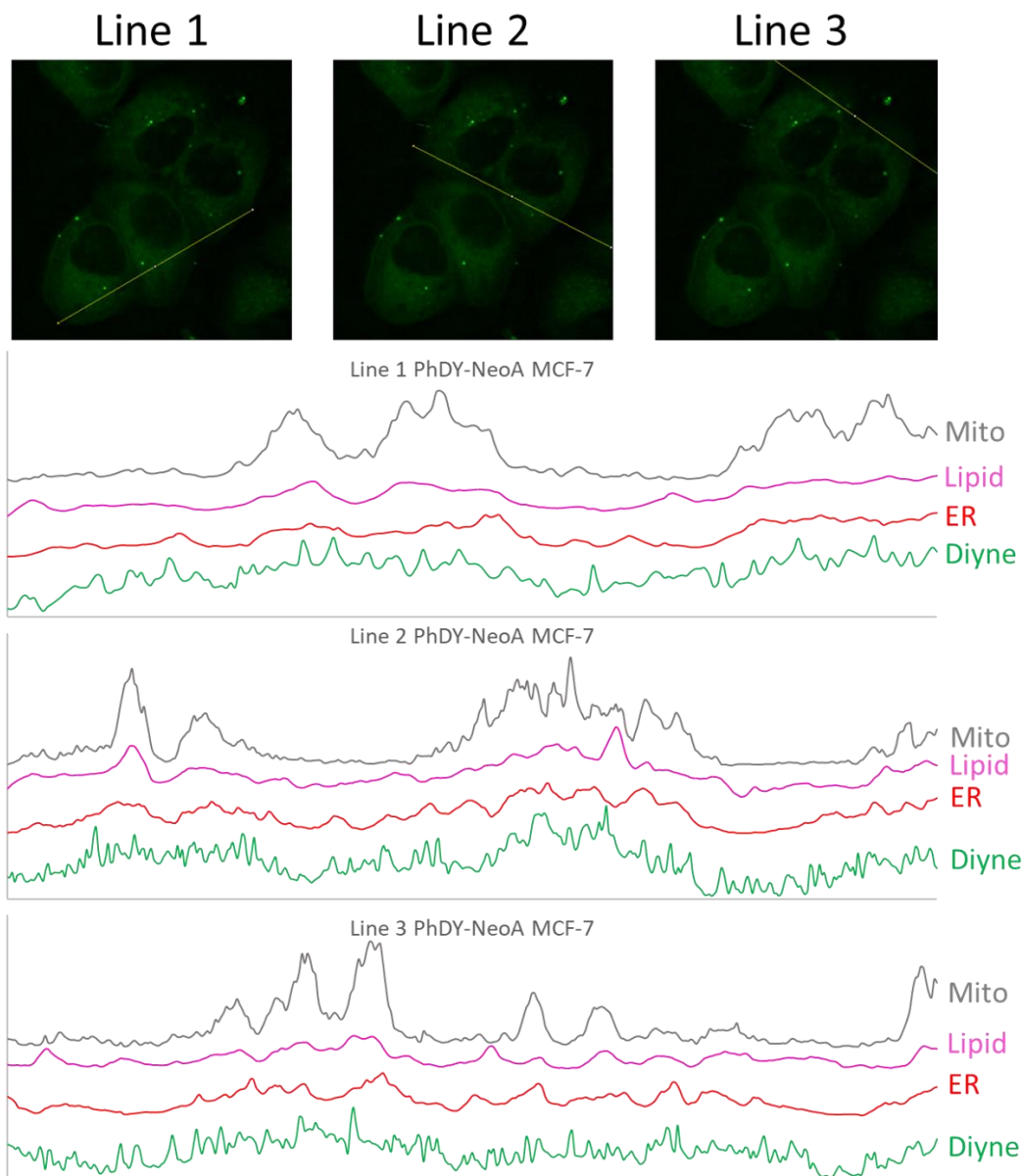
Figure 2-13. Line plot analysis of PhDY-NeoA (**4**) in MCF-7 cells. Each line plot shows poor correlation between Mitotracker and the diyne channel. Each line plot shows good correlation between the lipid channel and the diyne channel. The ER-tracker and diyne channel are less well-correlated than the diyne-lipid pair.

In summary, both the 9- and 15- membered antimycin-type depsipeptides have been subjected to the SADR study in live cancer cells. This work provides the first global and dynamic view of the interplay between these anti-cancer complex natural products and the complicated network of cellular machinery. We confirmed the high tolerance of the C-8 modification of the 9-membered antimycin for its anti-cancer activity and showed the passive while facile uptake of antimycin by live cancer cells. Interestingly, the primary localization of the 9-membered antimycin was demonstrated to be in the endoplasmic reticulum despite the previous known protein targets of

antimycin in various cellular organelles. We also showed that the anti-cancer activity of the 15-membered neoantimycin was dependent on the *N*-formyl moiety and less sensitive toward the C-11 modification. Importantly, a different intracellular localization of the 15-membered neoantimycin compared to the 9-membered antimycin was revealed. Our results further demonstrated that the intracellular enrichment and distribution of these compounds were driven by their potency and specific protein targets, as well as the lipophilic properties of compounds. This new integrative workflow of SADR study on bioactive natural products is expected to extend beyond the traditional SAR study, complement existing biochemical and proteomic techniques in the mode-of-action study of natural products, and facilitate efforts in reducing off-target effects and improving efficacy of candidate compounds in the early stages of drug discovery.

2.2 Experimental Methods

2.2.1 Spontaneous Raman Spectroscopy

All spontaneous Raman spectra were acquired with a commercial Raman microscope (Xplora, Horiba Jobin Yvon). All diyne labelled compounds were dissolved in dimethyl sulfoxide (DMSO) with a concentration of 50 mM for Raman spectrum acquisition. A 532 nm diode laser beam was guided through a 50×, 0.75 N.A. air objective (NPLAN EPI, Leica) to excite samples. All spectra were obtained with 5 s acquisition time and 10 times averaging. All Raman spectrum data were processed with LabSpec 6 software.

2.2.2 General cell culture

MCF7 and HeLa cells were obtained from the UCB Cell Culture Facility cultured in DMEM (Gibco) supplemented with 10% (vol/vol) fetal bovine serum, GlutaMAX, penicillin, and streptomycin. Cells were maintained at 37 °C with 5% CO_2 in a water saturated incubator.

2.2.3 Sample preparation for SRS and fluorescence imaging of live cells

HeLa and MCF-7 cells were seeded on round coverslips with 800 μL complete growth media in four-well plates 2 days before the experiment. On the day of the experiment, the culture media was replaced with 500 μL fresh culture media supplemented with fluorescent organelle markers (MitoTracker™ Deep Red FM, Invitrogen; ER-Tracker™ Green (BODIPY™ FL Glibenclamide), Invitrogen) and diyne-labelled compounds at specific concentrations. The cells were incubated for designated time in the incubator until washed with phosphate buffered saline (PBS, Gibco) three times. The coverslips with cells were placed onto imaging spacers (SecureSeal™ imaging spacers, Sigma-Aldrich) and attached to glass slides. The space between coverslips and glass slides was filled with PBS solution to preserve the live cells for a short period.

2.2.4 Confocal Fluorescence Microscopy

All fluorescence imaging was performed with a commercial Olympus FV1200 confocal microscope with standard laser excitation and bandpass filter set for ER-Tracker and Mito-Tracker. The objectives were either a $25\times$ 1.05 N.A. water objective (XLPlan N, Olympus) or a $60\times$ 1.2 N.A. water objective (UPlanAPO/IR, Olympus). The images were acquired with Olympus FV10 software and processed with ImageJ.

2.2.5 Stimulated Raman Scattering (SRS) Microscopy

SRS imaging was performed with the same microscope setup as the fluorescence imaging but with customized laser input. A commercial laser source (picoEmerald, Applied Physics & Electronics, Inc.) was utilized to produce both the Pump and Stokes beams for SRS. The wavelength of Stokes beam was 1064 nm. The beam was modulated at 8 MHz by an electro-optic modulator (EOM). The Pump beam had tunable wavelengths within a range of 720 nm to 990 nm. Both beams had 6 ps pulse width and 80 MHz repetition rate. The two beams were spatially and temporally overlapped and tightly focused onto the sample, after which Pump loss and Stokes Gain were generated. Both transmitted beams were collected by a 1.4 N.A. oil condenser. The Stokes beam was filtered off with a high O.D. bandpass filter (890/220 CARS, Chroma Technology), while the Pump beam was detected by a silicon photodiode (FDS1010, Thorlabs) with a DC voltage of 64 V. The output current was terminated by a 50 Ω terminator and demodulated by a lock-in amplifier (SR844, Stanford Research Systems) at 8 MHz frequency. The Pump loss signal at each pixel was sent to the FV10 analog channel to generate the images. To obtain sufficient resolution and sensitivity, all live cell images were acquired with 100 us time constant and 512×512 pixel number. The CH_3 channel at 2940 cm^{-1} was obtained with 40 mW Pump power and 100 mW Stokes power on sample. The CH_2 channel at 2845 cm^{-1}, diyne channel at 2251 cm^{-1} and off-resonance channel at 2000 cm^{-1} were all obtained with 100 mW Pump power and 100 mW Stokes power on sample.

2.2.6 Image Analysis

All images were generated by Olympus FV 10 software. Images were taken under strictly the same conditions (pump and Stokes power, digital gain and offset) when comparison is performed.

The Raman beams and the fluorescent beams are not in-line in the microscope set-up. As a result, the two modalities have a slight offset in the images that they acquire. Before any image analysis or background reduction was performed, this misalignment had to be corrected. MATLAB code was used to align the images to achieve the highest possible colocalization coefficient between the Raman and fluorescent channels. The non-overlapping regions of the different image channels were then cropped off immediately. Subsequent analysis of overlaid images showed that the many channels were very well-aligned.

All image analysis was performed in FIJI. Because the images had varying levels of background, some background reduction had to be performed to facilitate the analysis. Every image channel

(except for -CH3) of every image set was subjected to an identical rolling ball background reduction with a radius of 200 pixels. This relatively large radius ensured against dimming or outright removing features with a homogenous intensity. The relative brightness of images were occasionally adjusted via the multiply function in imageJ prior to merging the images. This was purely to help the reader observe colocalization between two channels at different intensities. The image intensities were not modified for any non-merged images.

2.2.7 Cell viability assays

HeLa or MCF7 were grown in clear, flat bottom 96 well plates (Olympus plastics). MTT cell growth assay reagents (EMD Millipore) were used as purchased. In brief, to cells in 100 µL DMEM media, 10 µL of MTT solution was added. The assay was allowed to develop at 37 °C for 4 h. After MTT formazan formation 100 µL 0.04 M HCl in isopropanol was added to each well and mixed until all solid MTT formazan was dissolved. The 96-well plates were read for absorbance at 570 nm using a Tecan M1000 plate reader.

For cytotoxicity studies, compounds, dissolved in DMSO, were added to wells (final DMSO concentration of 0.5-1%, referenced to control wells containing DMSO alone) at 20% cell confluency. Cells were incubated with compounds for 48 h at 37 °C prior to MTT addition.

Growth response (GR) values were obtained by getting a baseline cell count using the MTT assay performed concurrently with cytotoxic compound addition in a separate plate as described by Hafner *et al.*[1] GR50 values and plots were obtained using the GR metrics software package accessed through github (https://github.com/sorgerlab/gr50_tools).[38] All data were collected in at least triplicate and GR50 values were averaged between at least three biological replicates with error given in standard error of the mean (SEM).

2.2.8 General synthetic methods

All reagents were purchased from commercial suppliers and used without further purification. Reaction progress was monitored by thin-layer chromatography on silica gel 60 plates (aluminum back, EMD Millipore) and visualized by UV light or by LC/MS (Agilent 6120 single quadrupole). Samples were analyzed by reverse phase HPLC on a C18 column (Eclipse Plus 3.5 µM 100 x 4.6 mm) with the following elution protocol: A 40-80% MeCN in H_2O (0.1% FA) linear gradient over 3 min, followed by a 80-100% MeCN in H_2O (0.1% FA) over 22 min with a flow rate of 0.5 mL/min. Compounds were purified by flash column chromatography using Fisher Scientific 230-400 mesh, 60 Å, silica gel. HPLC purifications were performed on an Agilent 1260 HPLC with a semipreparative scale were performed using an Altima C18 column 5 µM (150 x 10 mm). NMR spectra were acquired with a Bruker Biospin spectrometer with a cryoprobe. All spectra were acquired at 298 K. [1]H spectra were acquired at 900 MHz, ^{13}C spectra were acquired at 226 MHz. Coupling constants (*J*) are provided in Hz and chemical shifts reported in ppm relative to residual non-deuterated NMR solvent. High resolution mass spectra were collected using an Agilent

Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument. UV data was obtained on an Agilent 1260 series DAD.

2.2.9 Phenyl-diyne acid synthesis

The phenyl-diyne acid tag was synthesized according to the method described by Balaraman *et. al.*[39] First, $Cu(Ac)_2$ monohydrate was added to 50 mL of methylene chloride in a round bottom flask until saturation was achieved. Then, 5 equivalents of piperidine was added, followed by 5 equivalents of phenylacetylene and 1 equivalent of 4-pentynoic acid. The mixture was left exposed to air for the duration of the reaction. After the reaction mixture had stirred at room temperature for 8 hours, the excess $Cu(Ac)_2$ was removed by centrifugation. The clarified reaction mixture was then concentrated to 10 mL in vacuo. The $Cu(Ac)_2$ that had precipitated was once again removed by centrifugation. The concentrated, clarified reaction mixture was then fractionated via silica gel chromatography. The fractions containing the acid product were identified via TLC analysis and combined. After evaporating the combined fractions in vacuo, the purity of the product was confirmed via LC-MS and UV-vis analysis.
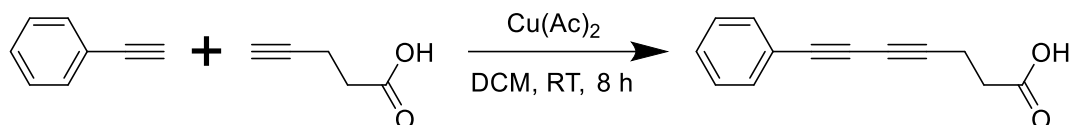
2.2.10 Deacylated antimycin production

Deacylated antimycin was produced using an engineered *Streptomyces albus* J1074 strain with the *antB* gene deleted in-frame through double crossover. A seed culture was grown in two separate aliquots of 5 mL tryptic soy broth at 30˚C for 48 hours. Next, 1 mL of the seed culture was added to 10 separate 1 L aliquots of mannitol soy broth in 4 L plastic flasks. After incubating for 6 days at 30˚C and 200 RPM, the solids were removed from the fermentation media via centrifugation. The clarified media was then extracted with 50 g/L of XAD-16 resin. The resin was then washed with 10 L of distilled water and subsequently extracted with 3 L of methanol. The methanol extract was then dried in vacuo and purified by silica gel column chromatography using a gradient of 0% to 100% ethyl acetate in methylene chloride in 25% increments. The fractions containing deacylated antimycin were combined, evaporated, dissolved in methanol and further purified by HPLC on a semi-preparative C18 column (10 mm i.d., 250 mm length, Vydac) with a linear gradient of 25 to 95% acetonitrile (vol/vol) over 20 min, and 95% acetonitrile (vol/vol) for a further 10 min in H2O with 0.1% (vol/vol) formic acid (FA) at a flow rate of 2 mL/minFractions were collected manually and concentrated under vacuum. Fractions containing *S. albus* Δ*antB* antimycins were pooled and further separated by HPLC using an Agilent Eclipse Plus C18 column (4.6 x 100 mm) with a linear gradient of 25 to 95% acetonitrile (vol/vol) over 20 min, and held at 95% acetonitrile (vol/vol) for 10 min in H2O with 0.1% (vol/vol) formic acid (FA) at a flow rate of 1 mL/min. Total yield was roughly 5 mg/L.

2.2.11 Neoantimycin production

Neoantimycin was produced using a wild-type strain of *Streptomyces orinoci*. *S. orinoci* was grown three separate 5 mL cultures of TSB for 72 hours. After that 50 µL of seed culture was added to 40 separate 150 mL plastic shake flasks each containing 50 mL of mannitol soy broth. After 6 days of fermentation, the broth was clarified by centrifugation. The 2 L of broth was extracted two times consecutively with 2 L of ethyl acetate. The ethyl acetate extract was dried with sodium sulfate and evaporated in vacuo. After redissolving the residue in 20 mL of ethyl acetate, it was fractionated by silica gel chromatography using a gradient of 0% to 100% ethyl acetate in methylene chloride in 25% increments. The fractions with neoantimycin were identified via LC-MS and combined. The fractions containing neoantimycin were combined, evaporated, dissolved in methanol and further purified by HPLC on a semi-preparative C18 column (10 mm i.d., 250 mm length, Vydac) with a linear gradient of 25 to 95% acetonitrile (vol/vol) over 20 min, and 95% acetonitrile (vol/vol) for a further 10 min in H2O with 0.1% (vol/vol) formic acid (FA) at a flow rate of 2 mL/min. Fractions containing *S. orinoci* neoantimycins were pooled and further separated by HPLC using an Agilent Eclipse Plus C18 column (4.6 x 100 mm) with a linear gradient of 25 to 95% acetonitrile (vol/vol) over 20 min, and held at 95% acetonitrile (vol/vol) for 10 min in H2O with 0.1% (vol/vol) formic acid (FA) at a flow rate of 1 mL/min. Total yield was roughly 10 mg/L.

2.2.12 Esterification

Deacylated antimycin (4.00 mg and 0.00886 mmol) and neoantimycin (6.20 mg and 0.00886 mmol) were converted to PhDY-Ant and PhDY-NeoA respectively via Steglich esterification with phenyl-diyne acid. Phenyl-diyne acid (1.93 mg and 0.00975 mmol), DCC (2.01 mg and 0.00975 mmol), and DMAP (1.19 mg and 0.00975 mmol) were added to methylene chloride (5 mL) stirred in a round bottom flask and chilled in an ice bath. The round bottom was then evacuated and flushed with and kept under inert atmosphere. After stirring for one hour, the alcohol (either deacylated antimycin or neoantimycin) was added to the flask. The reaction was allowed warm to room temperature and stirred for 24 hours. The reaction mixture was then extracted three times with water. The fractions were combined and back-extracted with methylene chloride. The two methylene chloride fractions were then combined and evaporated in vacuo. After dissolving the product in methanol, it was purified via HPLC using an Agilent Eclipse Plus C18 column (4.6 x 100 mm) with a linear gradient of 55 to 95% acetonitrile (vol/vol) over 20 min, and 95% acetonitrile (vol/vol) for a further 10 min in H2O with 0.1% (vol/vol) formic acid (FA) at a flow rate of 2 mL/min. Total yield was roughly 70% for PhDY-Ant and 35% for PhDY-NeoA.

2.2.13 Deformylated Neoantimycin Production

Neoantimycin (**2**), 0.050 g, was dissolved in 0.25 mL EtOH followed by 0.25 mL 0.03 M HCl and the reaction was stirred at 55 °C for 3 h. The reaction was quenched with 0.80 mL 100 mM ammonium bicarbonate (pH 7) and the EtOH was evaporated *in vacuo*. The solution was then purified by HPLC (semi-preparative, reversed phase using the following elution protocol: 30%

MeCN for 2 min, followed by a linear gradient of 30-90% MeCN in H_2O over 24 min, with a flow rate of 1 mL/min) to yield a white solid (0.011 g, 24%).

2.2.14 Deformylated PhDY-Neoantimycin Production

PhDY-NeoA (**7**), 0.0085 g, was dissolved in 0.30 mL EtOH followed by 0.30 mL 0.03 M HCl and the reaction was stirred at 55 °C for 8 h. The reaction was quenched with 0.80 mL 100 mM ammonium bicarbonate (pH 7) and the EtOH was evaporated *in vacuo*. The solution was then purified by HPLC (semi-preparative, reversed phase, using the following elution protocol: 40% MeCN for 2 min, followed by a linear gradient of 40-90% MeCN in H_2O over 24 min, with a flow rate of 1 mL/min) to yield a light brown solid (0.0029 g, 35%). See NMR table below for spectral information. MS (ESI) calcd for $C_{48}H_{55}N_2O_{12}$ [M+H]$^+$ *m/z* 851.3750, found 851.3754.

2.3 Summary

Small-molecule natural products have been an essential source of pharmaceuticals to treat human diseases, but very little is known about their behavior inside dynamic, live human cells. Here, we demonstrate the first structure-activity-distribution relationship (SADR) study of complex natural products, the anti-cancer antimycin-type depsipeptides, using the emerging bioorthogonal Stimulated Raman Scattering (SRS) Microscopy. Our results show that the intracellular enrichment and distribution of these compounds are driven by their potency and specific protein targets, as well as the lipophilic nature of compounds.

2.4 References

1.   D. J. Newman and G. M. Cragg, J Nat Prod, 2016, 79, 629-661.
2.   J. A. Prescher and C. R. Bertozzi, Nat Chem Biol, 2005, 1, 13.
3.   S. M. DeGuire, D. C. Earl, Y. Du, B. A. Crews, A. T. Jacobs, A. Ustione, C. Daniel, K. M. Chong, L. J. Marnett, D. W. Piston, B. O. Bachmann and G. A. Sulikowski, Angew Chem Int Ed Engl, 2015, 54, 961-964.
4.   Y. Yan, J. Chen, L. Zhang, Q. Zheng, Y. Han, H. Zhang, D. Zhang, T. Awakawa, I. Abe and W. Liu, Angew Chem Int Ed Engl, 2013, 52, 12308-12312.
5.   J. R. W. Conway, N. O. Carragher and P. Timpson, Nat Rev Cancer, 2014, 14, 314.
6.   L. Wei, F. Hu, Y. Shen, Z. Chen, Y. Yu, C. C. Lin, M. C. Wang and W. Min, Nat Methods, 2014, 11, 410-412.
7.   S. Hong, T. Chen, Y. Zhu, A. Li, Y. Huang and X. Chen, Angew Chem Int Ed Engl , 2014, 53, 5827-5831.
8.   L. Wei, F. Hu, Z. Chen, Y. Shen, L. Zhang and W. Min, Acc Chem Res, 2016, 49, 1494-1502.
9.   W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton and A. N. Hulme, Chem Soc Rev, 2016, 45, 2075-2089.

10. C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang and X. S. Xie, Science, 2008, 322, 1857-1861.
11. R. C. Prince, R. R. Frontiera and E. O. Potma, Chem Rev, 2017, 117, 5070-5094.
12. J.-X. Cheng and X. S. Xie, Science, 2015, 350, aaa8870.
13. C. H. Camp Jr and M. T. Cicerone, Nat Photonics, 2015, 9, 295.
14. C. Krafft, M. Schmitt, I. W. Schie, D. Cialla-May, C. Matthäus, T. Bocklitz and J. Popp, Angew Chem Int Ed Engl, 2017, 56, 4392-4430.
15. H. Yamakoshi, K. Dodo, A. Palonpon, J. Ando, K. Fujita, S. Kawata and M. Sodeoka, J Am Chem Soc, 2012, 134, 20681-20689.
16. W. J. Tipping, M. Lee, A. Serrels, V. G. Brunton and A. N. Hulme, Chem Sci, 2017, 8, 5606-5615.
17. M. M. Gaschler, F. Hu, H. Feng, A. Linkermann, W. Min and B. R. Stockwell, ACS Chem Biol, 2018, 13, 1013-1020.
18. H. Yamakoshi, K. Dodo, M. Okada, J. Ando, A. Palonpon, K. Fujita, S. Kawata and M. Sodeoka, J Am Chem Soc, 2011, 133, 6102-6105.
19. S. F. El-Mashtoly, D. Petersen, H. K. Yosef, A. Mosig, A. Reinacher-Schick, C. Kötting and K. Gerwert, Analyst, 2014, 139, 1155-1161.
20. Z. Zhao, Y. Shen, F. Hu and W. Min, Analyst, 2017, 142, 4018-4029.
21. F. Hu, Z. Chen, L. Zhang, Y. Shen, L. Wei and W. Min, Angew Chem Int Ed Engl , 2015, 54, 9821-9825.
22. J. Liu, X. Zhu, S. J. Kim and W. Zhang, Nat Prod Rep, 2016, 33, 1146-1165.
23. L. S. Huang, D. Cobessi, E. Y. Tung and E. A. Berry, J Mol Biol, 2005, 351, 573-597.
24. S. P. Tzung, K. M. Kim, G. Basanez, C. D. Giedt, J. Simon, J. Zimmerberg, K. Y. Zhang and D. M. Hockenbery, Nat Cell Biol, 2001, 3, 183-191.
25. A. A. Salim, K. J. Cho, L. Tan, M. Quezada, E. Lacey, J. F. Hancock and R. J. Capon, Org Lett, 2014, 16, 5036-5039.
26. C. J. Barrow, J. J. Oleynek, V. Marinelli, H. H. Sun, P. Kaplita, D. M. Sedlock, A. M. Gillum, C. C. Chadwick and R. Cooper, J Antibiot (Tokyo), 1997, 50, 729-733.
27. Y. Zhou, X. Lin, S. R. Williams, L. Liu, Y. Shen, S.-P. Wang, F. Sun, S. Xu, H. Deng, P. F. Leadlay and H.-W. Lin, ACS Chem Biol, 2018, 13, 2153-2160.
28. M. Izumikawa, J. Y. Ueda, S. Chijiwa, M. Takagi and K. Shin-ya, J Antibiot (Tokyo), 2007, 60, 640-644.
29. Y. Umeda, S. Chijiwa, K. Furihata, S. Sakuda, H. Nagasawa, H. Watanabe and K. Shin-ya, J Antibiot (Tokyo), 2005, 58, 206-209.
30. A. S. Lee, Nat Rev Cancer, 2014, 14, 263-276.
31. T. Zhu, S. Cao, P.-C. Su, R. Patel, D. Shah, H. B. Chokshi, R. Szukala, M. E. Johnson and K. E. Hevener, J Med Chem, 2013, 56, 6560-6572.
32. H. J. Lee, W. Zhang, D. Zhang, Y. Yang, B. Liu, E. L. Barker, K. K. Buhman, L. V. Slipchenko, M. Dai and J.-X. Cheng, Sci Rep-Uk, 2015, 5, 7930.
33. M. Sandy, X. Zhu, Z. Rui and W. Zhang, Org Lett, 2013, 15, 3396-3399.
34. A. Schinzel, T. Kaufmann and C. Borner, Biochim Biophys Acta - Mol Cell Res, 2004, 1644, 95-105.

35. G. R. Pettit, R. Tan, R. K. Pettit, T. H. Smith, S. Feng, D. L. Doubek, L. Richert, J. Hamblin, C. Weber and J. C. Chapuis, J Nat Prod, 2007, 70, 1069-1072.
36. G. R. Pettit, T. H. Smith, S. Feng, J. C. Knight, R. Tan, R. K. Pettit and P. A. Hinrichs, J Nat Prod, 2007, 70, 1073-1083.
37. S. A. Vanner, X. Li, R. Zvanych, J. Torchia, J. Sang, D. W. Andrews and N. A. Magarvey, Mol Biosyst, 2013, 9, 2712-2719.
38. M. Hafner, M. Niepel, M. Chung, P. K. Sorger, Nature Methods 2016, 13, 521-527.
39. K. Balaraman, V. Kesavan, Nature Methods 2010, 20, 3461-3466.
40. Y. Takeda, T. Masuda, T. Matsumoto, Y. Takechi, T. Shingu, H. G. Floss, Journal of Natural Products 1998, 61, 978-981.

Chapter 3: Shotgun chemical derivatization of crude natural product extracts
Using isotopic labeling and mass spectrometry to enable a novel live cell binding assay

Traditional bioactivity assays used to discover antibiotics and anticancer therapies fall into one of two distinct categories: phenotypic assays, or protein-ligand binding assays. For the past century, nearly all small molecule medicines have been discovered by either a phenotypic assay or a binding assay. Presently, pharmaceutical progress is constrained by the limits of these assays. Phenotypic assays are limited to the discovery of bioactive compounds that cause an established, measurable phenotypic change, most commonly cell death or cell growth inhibition. While these cytotoxic compounds can sometimes be used as chemotherapy, they also cause dangerous side effects as a result of toxicity to healthy tissue. On the other hand, ligand binding assays focus on finding small molecules that inhibit or otherwise interact with an individual target protein. Unfortunately, the results of these assays which rely on examining the ligand interacting with pure proteins outside of the cellular environment, have proven unreliable for predicting bioactivity either in tissue culture or in vivo. To escape from the limitations of these two drug discovery paradigms, a screening method that captures the full complexity of living cells while remaining open to discovering new modes of bioactivity is necessary.

3.1 Current Paradigms in Drug Discovery

Phenotypic assays are both the simplest, and most productive biological assays to date. The most common phenotypic assays are simply a measure of compound toxicity towards cells in culture. Despite the great variation in methods of measuring cell death and growth inhibition, all cytotoxicity assays answer the same question: is this compound toxic to this cell type? It is not surprising, then, that the original chemotherapies that were discovered by this method, including taxols and camptothecins, all caused general toxicity in patients.[1] The well-known side-effects of chemotherapy, hair loss, immune suppression, fatigue, and death, are all a result of how these compounds were discovered. Screening for toxic compounds will reveal toxic compounds.

Ligand binding assays have allowed researchers to develop drugs with better therapeutic profiles by targeting proteins in disease pathway. However, they are limited to screening against one water soluble protein target at a time. Moreover, because a pure protein in solution does not accurately model the cellular environment, ligand binding detected in vitro is often not conserved in vivo.[2]

Despite the limited scope of phenotypic assays, they present a key advantage: they measure the effectiveness of small molecules in the crowded, chaotic, and interconnected environment inside a living cell. Knowing this, researchers have attempted to broaden the scope of phenotypic assays beyond toxicity with high content screening (HCS). During HCS, compounds are screened for the effect they have on cell size and morphology.[3] Practically speaking, cells must be stained and imaged to detect phenotypic changes that are more subtle than cell death or growth inhibition. Although much work has been done to increase the efficiency and accuracy of HCS, it is still plagued by slow throughput and difficulty elucidating previously unknown modes of action. This is because image analysis algorithms are based on phenotypic changes from known compounds

with known modes of action.[4] This strict limitation to inductive reasoning prevents HCS from meaningfully expanding the pharmaceutical arsenal.

3.2 A New Approach: Cell Binding Assays

To broaden the scope of pharmacological activity compatible with high-throughput screening, we have developed the live cell binding assay as a broader initial screen for bioactive compounds (Figure 3-1). During a cell binding assay, compounds in a crude natural product extract are filtered out based on their binding affinity to a living cell. The reasoning guiding this approach is that in order to affect any form of bioactivity, a small molecule drug must interact with the cell by binding to a molecular target within the cell, so non-binding compounds can be assumed to have no bioactivity. The cell binding assay doesn't discriminate based on a specific protein target or rely easily measurable phenotypic outcome. Instead, all possible protein targets and all possible phenotypic outcomes that may be caused by any small molecule in the extract are considered at once.

Cell binding assays start with a source of diverse natural products, such as dried plant material and extract the compounds with an organic solvent. We then use an ester-coupling reaction to attach chemical labels to hydroxyl-bearing compounds throughout the extract. The chemical label is a one-to-one ratio of $^{13}C_3$-propanoate and $^{12}C_3$-propanoate, which imparts a bioorthogonal isotopic distribution onto the tagged compounds. Following chemical derivatization, the tagged cell extract is contacted with HeLa cells. After several hours of incubation, several wash steps are performed to remove any unbound, or weakly bound compounds from the natural product extract. Next, the cancer cells are lysed, their proteins are denatured, and the bound compounds are extracted in one step by contacting the HeLa cell suspension with chloroform. The mixture of bound compounds is then analyzed by LCMS. Computational data analysis allows for rapid identification of any strongly bound compounds. As a whole, this process acts to exclude all weakly bound compounds from more labor-intensive compound isolation and bioactivity assays.

The first step in the process was methanol extraction of the natural producer, either cultured bacteria or dried plant material. The methanol extracts are then concentrated under flowing nitrogen, and the solid mass of each extract was determined. The residues were re-dissolved in anhydrous THF to a constant concentration of 10 extract residue mg per mL. A Steglich esterification was then performed on each extract. The resulting isotopically labeled extracts were then employed in the cell binding assay and subsequent computational data analysis.

The Steglich esterification was selected to perform the chemical labeling for a few reasons. Of principal importance, it targets hydroxyl and phenol functional groups, two of the most prevalent functionalities across all natural products. In addition, the reaction can be performed at room temperature without the use of strongly acidic or basic solutions, allowing most small molecule natural products to remain intact throughout the course of the reaction. Lastly, carbodiimide coupling reagents can be used that precipitate out of solution once they are converted alkyl urea products, allowing for a very simple workup. The DMAP can be used in a small enough catalytic amount to avoid the need for separation. In short, the Steglich esterification can derivatize the

majority of small molecule natural products without leading to degradation, and the product mixture can be easily worked up to remove excess reagents that may perturb the cell binding assay.
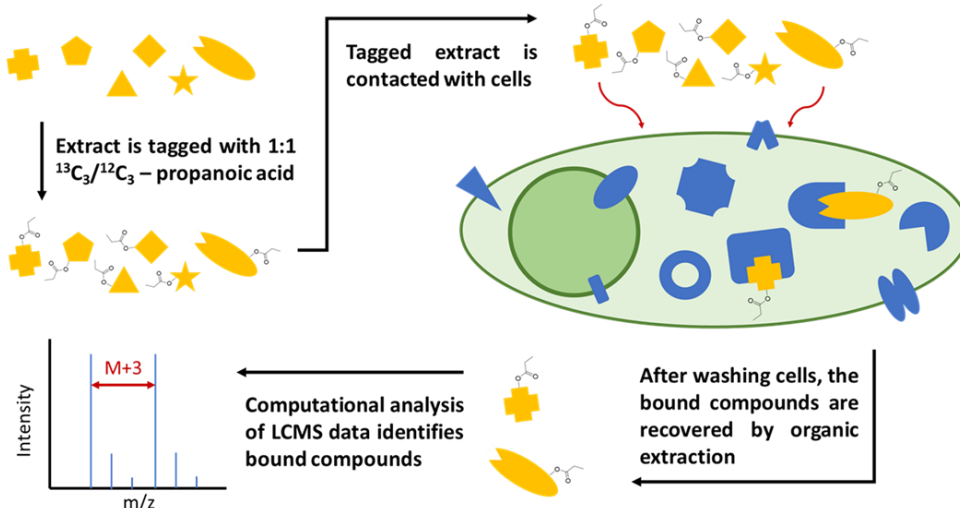

Figure 3-1. Workflow diagram for the live cell binding assay. First, a crude natural product extract is tagged with isotopically labeled propanoic acid. Then, the tagged extract is incubated with cells. Unbound and loosely bound compounds are washed off of the cells, which are then lysed and extracted with chloroform. The cell lysate/extract is then analyzed via LCMS, after which isotopically labeled compounds can be identified with computational data analysis.

3.3 Antimycin Proof-of-Concept: Extract Labeling and Binding Assay

Nature is an abundant source of unique, bioactive small molecules. Actinomycetes and plants are perhaps the two richest sources of medicinal compounds, so we relied on them as a source of crude extracts. However, it was important to prove that this process worked in a well-known system. As a proof-of concept, we performed shotgun derivatization on a *s. albus* J1074 ΔantB extract using a variety of acid labels and performed a cell binding assay.

The ΔantB mutant of *s. albus* J1074 was used to produce a deacylated antimycins.[5] The deacylated antimycins were extracted in the same manner but left as part of the crude extract with no purification steps. A Steglich esterification was then performed on the entire extract using one of two different acid reagents. In the first case, PhDY-acid was used in a 4:1 extract-to-acid mass ratio, to tag the deacylated Antimycins along with many other compounds in the crude extract. The PhDY acid was then used as a unique UV tag to determine the number of compounds labeled in the entire extract. The isoabsorbance plot from HPLC-UV analysis in Figure 3-2 below shows that over 40 PhDY derivatives were generated. Three of these products were PhDY-antimycins with varying alkyl sidechains. Following extract derivatization, PhDY-tagged extract was incubated with HeLa cells and imaged via SRS Microscopy. In this case, the antimycin localization was convoluted by the binding of many other compounds. This experiment showed that it is essential to use a labeling method that would allow us to deconvolute individual compounds in a cell binding assay. HPLC-UV and microscopic imaging of labeled extracts simply would not provide useful information.

In the second Steglich Esterification of S. Albus J1074 ΔantB extract, a 1:1 ratio of butyric acid and d7-butyric acid were used as a bioorthogonal isotopic label. This isotopically labeled mixture of acid introduced a distinctively unnatural, or bioorthogonal, pattern to the mass spectral patterns of all successfully esterified compounds in the extract. Following the isotopic label derivatization, the extract was analyzed by LCMS. Manual inspection of this data revealed the presence of several isotopically labeled Antimycins, as well as dozens of other isotopically labeled compounds; however, a thorough analysis of the data would require computational methods to selectively identify the isotopically labeled species.
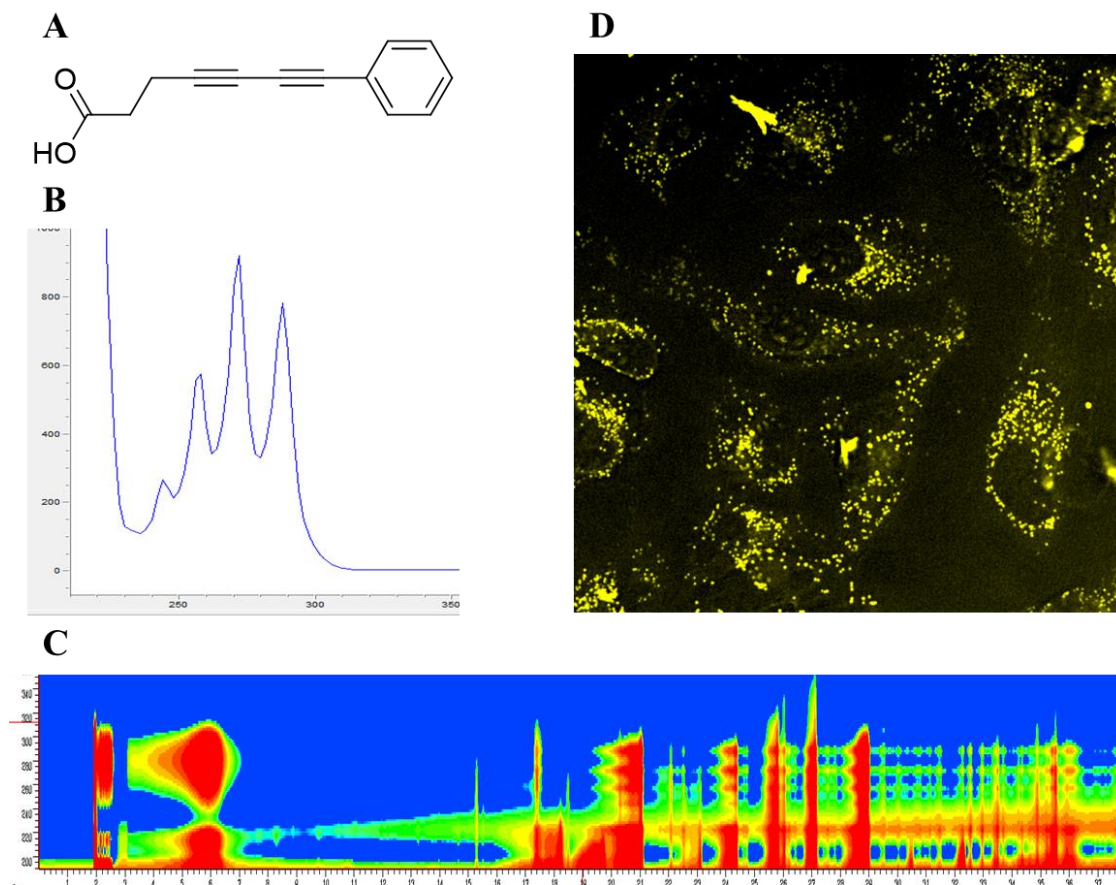
Figure 3-2. Labeling of ΔantB extract with PhDY acid. A) The structure of the PhDY acid tag used to derivatize the ΔantB extract. The UV signature of PhDY acid showing peaks at 245nm, 260nm, 275nm, and 290nm. C) The isoabsorbance plot of the tagged extract run on an HPLC showing the generation of many tagged species with the same UV absorbance signature. D) Diyne channel Raman image of cells incubated with 60 mg/mL tagged extract for two hours. The image shows tagged compounds accumulate in lipid droplets, giving little information on distribution.

To facilitate computational data analysis, the LCMS data set needed to be simplified into a table consisting of peaks with three variables: retention time, ion mass-to-charge ratio, and intensity. The LCMS data in RAW format was then converted to MZML using the opensource software msconvertgui. The MZML file was then processed using another opensource software called MSDIAL, which generated the aforementioned peak table.[6] The peak table was then subjected to a simple computational data analysis in Python with the following steps:

1. For any compound i, is there a compound with $mass_j = mass_i + 7.0439$ Daltons?
    a. 7D-7H = 7.0439 Daltons
    b. 2 ppm error allowed
2. Are the retention times RT_i & RT_j similar?
    a. 0.0 min $< RT_i - RT_j < 0.2$ min
    b. The asymmetry is because C-D bonds are more polar than C-H bonds, so deuterated compounds elute sooner during HPLC chromatography.
3. Are the peak intensities for I_i & I_j similar?
    a. $2/3 < I_i / I_j < 3/2$
    b. Similar intensities reflect the 1:1 reagent ratio.

This data analysis revealed that the derivatized ΔantB extract contained a remarkable 330 unique isotopically labeled compounds. This is particularly interesting, because bioinformatic analysis of *S. Albus* J1074 reveals only 26 secondary metabolite biosynthetic gene clusters (BGCs).[7] This can be understood by accounting for two phenomena: promiscuous biosynthetic enzymes, and premature release of shunt products. The deacylated antimycin itself serves as an example of how promiscuous enzymes lead to the production of many related compounds; three unique isotopically labeled Antimycins were identified that varied only in their alkyl chain length. Many BGCs also produce shunt products at various points in the biosynthesis of the finished secondary metabolite. In combination with substrate promiscuity, the formation of shunt products can lead to the production of many more compounds than the total number of BGCs, with even slight variations in compound structure resulting in diverse bioactivity.

Following the successful derivatization and computational data analysis of the ΔantB extract, the first cell binding assay was performed using HeLa cells. HeLa cells were selected as a model mammalian target cell because they are easy to culture and are very well characterized, with extensive cytotoxicity data available for many bioactive natural products. HeLa cells were cultured in a 10 cm petri dish using 10 mL of DMEM with glutaMAX and 10% FBS. Then, 10 mg of butyrate-labeled ΔantB extract in 100 μL DMSO was added directly to the media. The cells were incubated with the extract-rich media for 2 hours. After incubation, the media was removed, and the cells were washed with 10 mL PBS three consecutive times, for 5 minutes each. Then, 2 mL PBS was added to the dish, and the cells were gently scraped into the liquid. Finally, the 2 mL mixture of cells was contacted with 8 mL of chloroform and vortexed for 1 minute. The chloroform was used to simultaneously lyse the cells, denature proteins, and extract the hydrophobic natural products that had bound to the cell. After centrifugation, the organic phase was recovered, evaporated, and re-dissolved in 200 μL methanol.

The cell binding assay sample was then analyzed via LCMS and subjected to the same computational data analysis as the labeled extract. This analysis revealed that 30 isotopically labeled compounds bound to the HeLa cells and were recovered. These compounds included all three variants of Antimycin identified before the binding assay, despite the widely varying abundance of these three compounds. This experiment confirmed that isotopically labeled bioactive compounds could be bound to cells, and later recovered and identified with LCMS. Furthermore, more than 90% of the labeled secondary metabolites from the extract did not bind to cells and could be easily excluded from further analysis. This shows that the cell binding assays is both capable of recovering bioactive compounds and capable of excluding a significant portion of inactive compounds.

3.4 Screening Dietary Supplements for Bioactive Compounds

To screen for new bioactive compounds, we decided to rely on an easily accessible, consistent source of material rich in diverse natural products, plant-based dietary supplements. Dietary supplements were selected based on two criteria: they were made of raw, dried plant material from a single species, and the plant species used was known to contain a variety of structurally diverse small molecule secondary metabolites. Four common dietary supplements fit these criteria: Holy Basil, Ashwagandha, Kava Kava, and Valerian. All three extracts were subjected to the same extraction, chemical derivatization, cell binding assay, and data analysis.

First, the 2.5 grams of raw plant material was recovered from the dietary supplement capsules. This dried plant material was then contacted with 35 mL of pure methanol and shaken for 24 hours. After centrifugation, the methanol extract was recovered, and evaporated under flowing nitrogen. The residue mass was obtained after evaporation, and it was re-dissolved in anhydrous THF to a constant concentration of 10 mg residue per mL.

Then, five mL of each THF solution (50 mg residue) was transferred to separate reaction vessels. In each vessel, the same quantity of 4 reagents was added. Five additional mL of pure THF was added to each vessel. 35 µmoles each of $^{13}C_3$ and $^{12}C_3$-propanoic acid were added to the vessel for a total acid amount of 70 µmoles. One equivalent of dimethylaminopyrridine (DMAP) was added to each vessel along with 1.1 equivalents of diisopropylcarbodiimide (DIC). After shaking at 22ºC for 24 hours, the reactions were terminated by evaporation followed by re-dissolving in methanol. The methanol solvent was then evaporated, so that the tagged extract could be dissolved in DMSO to a concentration of 100 mg/mL.

After the extracts were tagged with $^{13}C_3/^{12}C_3$-propanoic acid, they were used in a cell binding assay. During the cell binding assay, each extract was added to cell culture medium, and HeLa cells were incubated with extract rich medium. During this time, high-affinity compounds concentrated within the cells. Several wash steps were performed to eliminate non-binding and weakly bound compounds from the cells. Cells were recovered by scraping into PBS, and the suspension was extracted using chloroform. This extract was concentrated and re-dissolved in methanol for LCMS analysis.

After the binding assay, both the labeled extracts and the cell lysate samples were subjected to HPLC analysis, and the isotopically labeled compounds were identified using the same computational method as described in the ΔantB proof-of-concept, with appropriately modified parameters. Most importantly, the target mass difference was changed to 3.0188 Daltons, and the RT tolerance was changed to a symmetrical -0.1 to 0.1 minutes because there is no significant difference in retention time for ^{13}C-labeled compounds.

LCMS analysis resulted in incredibly dense data sets for each labeled extract and its corresponding binding assay sample. To facilitate rapid data analysis, the raw LCMS data was subjected to pre-processing in MSDIAL, which yielded a peak table. Each peak was described by its retention time, m/z, and intensity. These three factors were sufficient to enable simple Python code (see Appendix) to rapidly identify the isotopically labeled compounds in each sample. This data analysis, summarized in Table 3-1, revealed several known bioactive compounds, and several ions that did not correspond to any known compound from its respective plant source. Each labeled extract contained between 300 and 800 isotopically labeled compounds, from which about 2.5% to 8% compounds present in the extract were successfully concentrated during the cell binding assay. There were many tagged ions present after the binding assay which were absent in the tagged extract. These ions are likely the product of cellular degradation of a tagged species, and therefore difficult to pursue. This analysis revealed several known bioactive compounds from the dietary supplement extracts as well as a multitude of unidentified ions.

Table 3-1. Summary of tagged ions identified from the dietary supplement labeling and cell binding assay experiments.

Herb Name	tagged ions	bound ions	hit ions	interesting hits
Holy Basil	773	126	56	8
Ashwagandha	436	44	34	5
Valerian	366	42	14	4
Kava	352	23	9	4

Ocimum tenuiflorum, commonly known as Holy Basil, is a known producer of many small molecules, including several bioactive compounds. Using the binding assay, tulsinol C was successfully identified (Figure 3-3, Figure S9). Tulsinol C, a member of the neolignoid family of small molecule natural products, has been reported to have strong antiparasitic activity against *leishmania normalis*.[8] However, it is not known to have any anticancer activity or other bioactive properties within mammalian cells. Absent from the cell binding assay is a family of bioactive isomers including betulinic acid and oleanolic acid. Betulinic acid is known for possessing anti-HIV activity while oleanolic acid is known as potent anti-inflammatory compound.[9] Bioactivity for betulinic derivatives generated by acylation at the hydroxyl group is well-known and includes the acylated betulinic acid derivative Beverimat, an experimental anti-HIV drug.[9] Based on this knowledge, we can conclude that acylation of betulinic acid did not cause a loss of bioactivity. Instead, it is likely that these compounds did not appear in the binding assay simply because they

do not form positive ions during LCMS analysis. Future experiments may benefit from LCMS analysis in both positive and negative mode to avoid missing potential bioactive compounds. Tulsinol C, as well as seven unidentified species, were selected for isolation and characterization (Table 3-2).
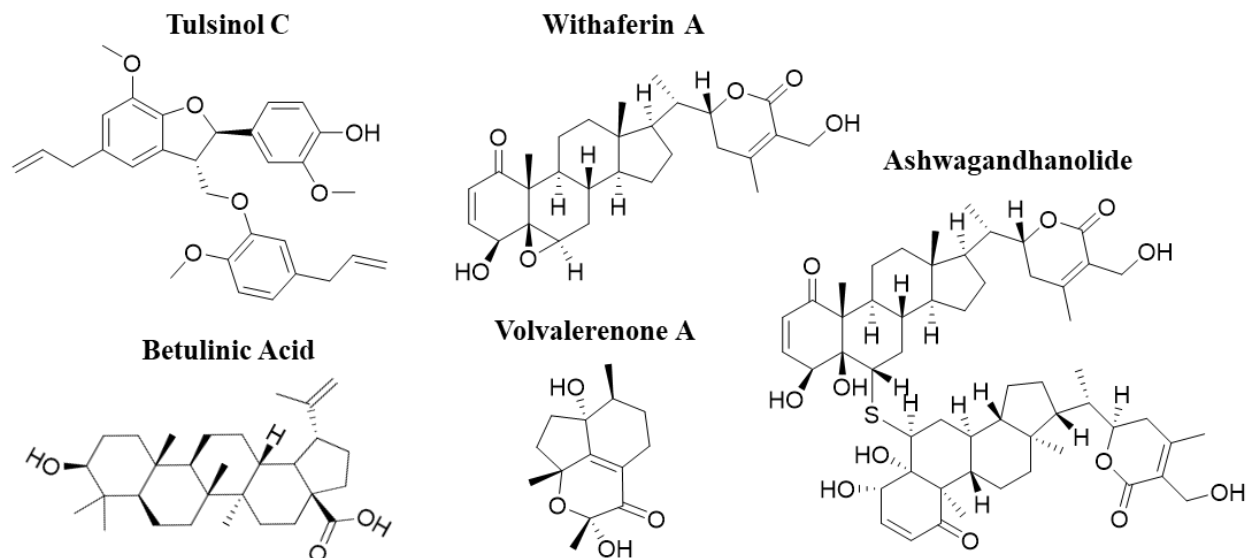

Figure 3-3. Selected small molecules isolated from various herbal supplements.

Withania somnifera, commonly known as Ashwagandha, is used in Ayurvedic medicine as an anti-anxiety and fitness supplement, though there is no scientific evidence to support these uses. Despite this, Ashwagandha is a source of several known bioactive compounds, including Ashwagandhanolide, and withaferin A (Figure 3-3, Figure S15, Figure S16). Ashwagandhanolide and withanolide sulfoxide, bioactive compounds isolated from *W. somnifera* with reported anticancer activity, were both successfully identified after the binding assay.[10,11] On the other hand, withaferin A is a known bioactive compound with reported anti-cancer, immunosuppressive, and anti-inflammatory activities, but it was not recovered from our cell binding assay. This may be because withaferin A binds covalently to its protein targets, an idea supported by its highly reactive epoxide group.[12] To address this gap in the analysis, the cell binding assay may be coupled with a MS-based chemoproteomic analysis like the one described by Grossman et. al.[12] However, that analysis is outside of the scope of this work, which focuses on non-covalent lignads. Ashwagandhanolide, as well as five unidentified species were selected for isolation and characterization (Table 3-2).

Table 3-2. Ions identified from various dietary supplements that have been selected for isolation and characterization based on the cell binding assay.

Source	m/z	Reten. time	Intensity	adduct	Parent ion m/z	Molecule exact mass	Supplementary Figure
H. Basil	588.3742	28.9	1.03E+05		532.348		Figure S6
H. Basil	777.3252	30.2	9.77E+04	Na+	721.299	698.3092	Figure S7
H. Basil	747.3146	30.3	2.19E+05		691.2884		Figure S8
H. Basil	545.2531	31.4	5.21E+05	H+	489.2269	488.2191	Figure S9
H. Basil	742.3588	31.8	1.09E+05	NH4+	686.3326	668.2982	Figure S10
H. Basil	761.3293	31.7	3.59E+05	Na+	705.3031	682.3133	Figure S11
H. Basil	655.5043	32.6	7.03E+04		599.4781		Figure S12
H. Basil	729.3036	33.8	1.12E+05		673.2774		Figure S13
Ashwa	719.3771	22.9	3.49E+05		663.3509		Figure S14
Ashwa	1047.5519	23.6	1.17E+05	H+	991.5257	990.5179	Figure S15
Ashwa	1031.5559	24.5	2.62E+05	H+	975.5297	974.5219	Figure S16
Ashwa	1135.6299	24.9	6.08E+04		1079.6037		Figure S17
Ashwa	567.235	29.8	1.73E+05	Na+	511.2088	488.219	Figure S18
Valerian	269.1018	30.1	2.01E+04		213.0756		Figure S19
Valerian	558.3575	30.8	9.68E+04		502.3313		Figure S20
Valerian	586.3598	34	5.23E+04		530.3336		Figure S21
Valerian	591.3426	31.5	5.05E+04		535.3164		Figure S22
Kava	342.3003	33.4	4.80E+04		286.2741		Figure S23
Kava	586.3598	34.1	1.54E+04		530.3336		Figure S24
Kava	313.1798	35.4	1.06E+05		257.1536		Figure S25
Kava	588.3752	35.6	1.32E+04		532.349		Figure S26

Valeriana officinalis, commonly known as valerian, is a popular dietary supplement derived from the root of the plant. While it has been advertised as an anti-anxiety supplement, with activity attributed to valerenic acid, there is little evidence to support this claim. However, valerian root is known as a source of many unique terpinoids, including vlolvalerenone A (Figure 3-3).[13] With such a wide diversity of unique small molecule scaffolds, valerian was an attractive candidate for screening via the cell binding assay. Four unidentified ions have been selected for isolation and characterization (Table 3-2).

Piper methysticum, commonly known as Kava Kava, is a cash crop native to the pacific islands. There it is used for its sedative and anxiolytic properties, and is commonly used as a recreational drug, known for its euphoric and inebriating effects at higher doses.[14] These psychoactive properties are the result of a family of small molecules known collectively as kavalactones. However, none of these known psychoactive compounds possess a hydroxyl group or other

nucleophilic group available to be tagged with the isotopically labeled propionic acid. Nevertheless, the binding assay revealed four unidentified ions, all with masses much larger than the typical kavalactone, which have been selected for isolation and characterization (Table 3-2).

3.5 Experimental Section

3.5.1 Extraction

For each dietary supplement, exactly 2.5 grams of raw plant material was recovered from the dietary supplement capsules – roughly five capsules. This dried plant material was then contacted with 25 mL of pure methanol and shaken continuously for 24 hours. After centrifugation, the methanol extract was recovered, and evaporated under flowing nitrogen. The residue mass was obtained after evaporation, and it was re-dissolved in anhydrous THF to a constant concentration of 10 mg extract residue per milliliter.

3.5.2 Extract labeling

Chemical Derivatization was performed on a mass basis proportional to the amount of extract residue for each extract. Five mL of each THF solution (50 mg residue) was transferred to separate reaction vessels. In each vessel, the same quantity of 4 reagents was added. Five additional mL of pure THF was added to each vessel. 35 µmoles each of $^{13}C_3$ and $^{12}C_3$-propanoic acid were added to the vessel for a total acid amount of 70 µmoles. One equivalent of dimethylaminopyrridine (DMAP) was added to each vessel along with 1.1 equivalents of diisopropylcarbodiimide (DIC). After shaking at 22ºC for 24 hours, the reactions were terminated by evaporation followed by re-dissolving in methanol. The methanol solvent was then evaporated, so that the tagged extract could be dissolved in DMSO to a concentration of 100 mg/mL.

3.5.3 Cell binding assay

After the extracts were tagged with $^{13}C_3/^{12}C_3$-propanoic acid, they were used in a cell binding assay. HeLa cells were cultured in 10 cm dishes using 10 mL per well of media (DMEM with glutaMAX and 10% FBS). Then, 10 mg of labeled extract in 100 µL DMSO was added directly to fresh media in falcon tubes. After centrifugation, the four extract-rich mediums were substituted for growth medium in three dishes. The cells were incubated with the extract-rich media for an hour. Then the cells were washed with 5 mL PBS and incubated with 10 mL PBS for 30 minutes. After incubation, the media was removed, and the cells were washed with 5 mL PBS three consecutive times, for 5 minutes each. Then, 3 mL PBS was added to the dish, and the cells were gently scraped into the liquid. Finally, the separate suspensions of extract-rich cells were contacted with 12 mL of chloroform and vortexed for one minute each. The chloroform was used to simultaneously lyse the cells, denature proteins, and extract the hydrophobic natural products that had bound to the cell. After centrifugation, the organic phase was recovered, evaporated, and re-dissolved in 200 µL methanol.

3.5.4 <u>LCMS analysis</u>

Samples were analyzed on a Thermofisher Orbitrap mass spectrometer after separation on a C18 column with a 30-minute elution gradient from 0% - 100% acetonitrile in an aqueous mixture with 0.1% formic acid. MS1 data was collected with a m/z range from 500 – 1700, with an error less than two parts per million for known metabolites.

3.5.5 <u>Computational data analysis</u>

Thermofisher .RAW files were compressed into .MZML files using MSconvertGUI after which they were analyzed using the preset parameters in MSDIAL. The peak table was generated using MSDIAL and saved as .csv file in excel. Subsequent analysis in custom MATLAB software selectively identified isotopically labeled compounds based on the peak table.

3.6 Summary

Small-molecule natural products have a rich history as a source of medicinal compounds, serving as antibacterial, antifungal, and anticancer antibiotics; however, most screening methods rely on either purified proteins, or the systematic measurement of easily observable phenotypic changes, such as cell death. Here, we demonstrate the shotgun chemical derivatization of natural products in crude extracts of plant-based dietary supplements. The chemical label, a one-to-one ratio of $^{12}C_3$- and $^{13}C_3$-propanoic acid, is attached to alcohol functional groups, which are present on most bioactive small molecules, via Steglich esterification. This chemical derivatization imparts a bioorthogonal mass spectrum onto any tagged molecule, which is easily identified via computational analysis of LCMS data. We then demonstrate the use of these isotopically labeled extracts and the associated computational data analysis for a live cell binding assay. This novel bioactivity assay selectively identifies compounds with a high affinity for living mammalian cells in culture while filtering out compounds that are only loosely bound to the target cells. This assay allows the screening of small molecules from a crude extract that bind to any molecular target in the target cells, regardless of the phenotypic impact, or effective dose of the bioactive molecule. Here, we use these techniques to successfully rediscover several known bioactive small molecules and find many unidentified compounds from well-studied sources that are promising candidates for further characterization.

3.7 References

1.   Ravi Chari. *Acc. Chem. Res,* 2008, **41(1)**, 98–107.
2.   J. Renaud and M. Delsuc. *Current Opinion in Pharmacology,* 2009, **9(5)**, 622-628.
3.   M. Bickle. *Anal Bioanal Chem,* 2010, **398,** 219–226.
4.   F. Gasparri. *Expert opinions in Drug Discovery,* 2009, **4(6)**, 643-657.
5.   M. Sandy et. al. *Org. Lett,* 2013, **15(13)**, 3396–3399.
6.   H. Tsugawa et. al. *Nature Methods,* 2015, **12**, 523–526.

7.   Yousra Ahmed et. al. *Scientific Reports,* 2017, **7**.
8.   A. Suzuki et. al. *Chemical and Pharmaceutical Bulletin,* 2009, **57(3)**, 245-251.
9.   P. Yogeeswari et. al. *Current Medicinal Chemistry,* 2005, **12(6)**, 657-666.
10.   G. Subbaraju et. al. *J Nat Prod,* 2006, **69(12)**, 1790-1792.
11.   V. Mulagabal et. al. *Phytotherapy Research,* 2009, **23(7)**, 987-992.
12.   E. A. Grossman. *Cell Chem Biol*, 2017, **24(11)**, 1368-1376.
13.   P. Wang et. al. *Tetrahedron Letters,* 2010, **51(41)**, 5451-5453.
14.   J. Sarris et. al. *Australian and New Zealand Journal of Psychology,* 2011, **45(1)**, 27-35.

Chapter 4: FRET-based Protein Binding Assays in the Cellular Environment
Optical lock-in detection of FRET will enable new methods in drug development

This chapter serves as a roadmap to combine natural product tagging with FRET for applications in chemoproteomics and drug discovery. While this project is early in development, its long-term potential deserves discussion.

We propose the development of a FRET detection technique with unprecedented throughput and sensitivity by utilizing a photoswitching fluorophore to isolate FRET behavior from background fluorescence. This technique, known as optical lock-in detection of FRET (OLID-FRET), has previously been used in imaging application[1], but we will focus on optimizing it for a 96-well plate assay. The proposed optimization of this technique will enable high throughput protein binding assays in live cells, an entirely new methodology for drug discovery and development.

4.1 Introduction

Studying the complex machinery of life, with incredible chemical diversity and structural organization at a submicroscopic scale, has proven to be one of the greatest challenges in science. While the average size of a eukaryotic protein is five nanometers, typical confocal microscopy is limited to a resolution fifty times that large. More advanced techniques can achieve an astounding twenty nanometer resolution, but it is limited to low throughput applications.[2] As such, researchers are currently limited to either choose high throughput experiments to study a diverse array of proteins, or low throughput experiments to study the organization of individual proteins. FRET provides chemically selective proximity detection at a resolution of 5-10 nm, so it has the potential to accommodate these two conflicting goals. Unfortunately, traditional FRET detection is extremely sensitive to background fluorescence, making it unreliable in a high-throughput assay. OLID-FRET effectively eliminates this background signal, allowing for rapid, sensitive, and robust proximity measurements even inside living cells.

Traditional FRET relies on measuring fluorescence emission from a FRET acceptor after exciting the FRET donor. In contrast, OLID-FRET relies on the use of a photoswitching fluorophore as the FRET acceptor to repeatedly quench and unquench a FRET donor (Figure 4-1). Though many FRET pairs are compatible with this technique, we will focus specifically on the nitroBIPS-GFP system developed by our collaborator, Gerard Marriott's lab at UC Berkeley.[1]
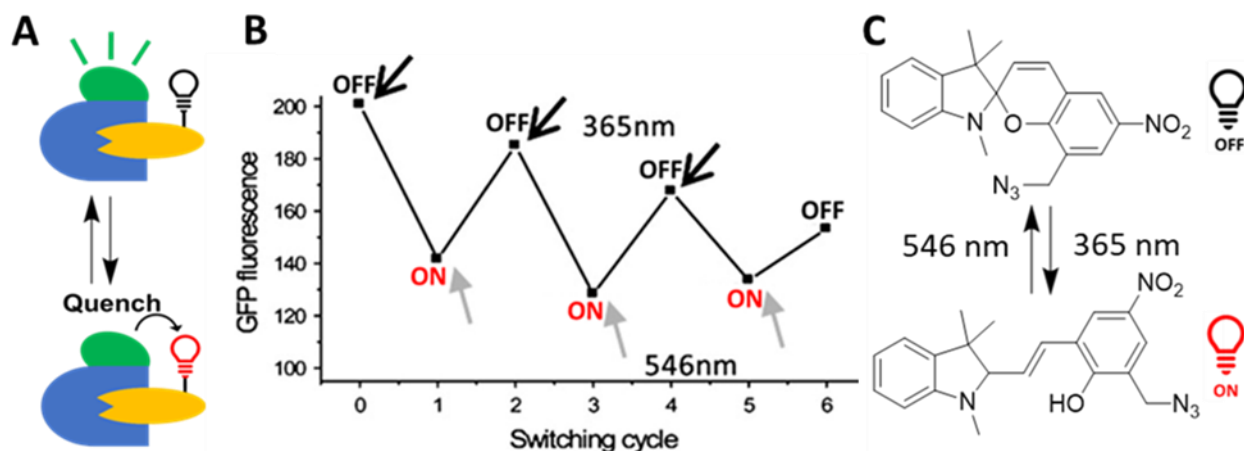
Figure 4-1. Schematic of the detection of OLID-FRET. A) A nitroBIPS-labeled drug is used to quench and unquench a GFP label on its protein target. B) The sawtooth pattern in GFP signal is generated over many switching cycles.[1] C) nitroBIPS molecular structure can be reversibly switched on and off with 365nm and 546nm light respectively.

4.2 Assay Optimization

The main technical challenge of this process is to transition OLID-FRET from under the microscope to a 96-well plate reader. To make this transition, we require a test case that allows for a strong positive control and an exhaustive range of negative controls. The simplest case is a nitroBIPS-labeled small molecule drug and an engineered cell line with a GFP fused to the drug's protein target. For this purpose, we have purchased a library of 4000 engineered yeast cell lines with GFP fusion proteins covering 75% of the proteome.[4] We have also successfully synthesized a nitroBIPS-labeled antifungal drug, echinocandin B (Scheme 4-1). Its structure was confirmed by NMR (Figure S27). Its bioactivity was confirmed with a typical inhibition assay with a minimum inhibitory concentration of 32 μM in *S. cerevisiae* S288C. Echinocandin B binds selectively to 1,3-ß-glucan synthase, so we can expect nitroBIPS-Echinocandin to produce an OLID-FRET signal when incubated with yeast expressing GFP-labeled 1,3-ß-glucan synthase but produce no signal with any other GFP-labeled protein.

With a solid test system available, we can start working optimizing our OLID-FRET detection platform. Broadly, the challenges in this project fall into two categories: experimental design and data processing. In addition to basic parameters like compound concentration, experimental design optimizations will focus on ensuring that the nitroBIPS probe is efficiently switched between states without bleaching GFP. In each well, GFP fluorescence will be measured over the course of several switching cycles, generating a separate function for each well on a plate. Each individual function must then be scored for its similarity to the sawtooth pattern produced by continuously measuring GFP fluorescence while switching nitroBIPS (Figure 4-1B). In previous work, optical lock-in detection could measure FRET efficiency as low as 0.1%, a feat that is impossible with ordinary FRET detection.[1]
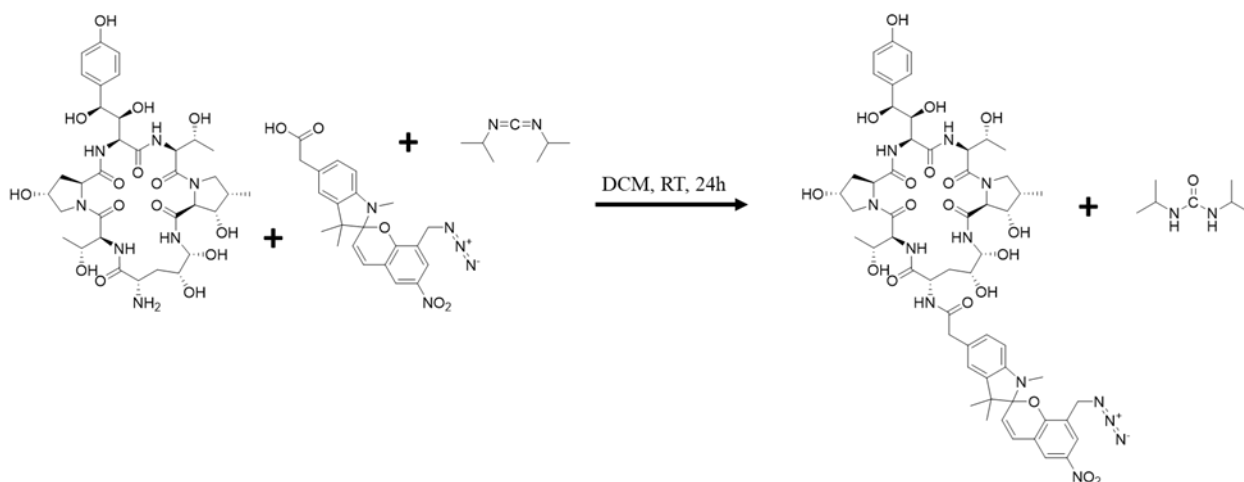
Scheme 4-1. NitroBIPS echinocandin is synthesized by DIC amide coupling of azido-nitroBIPS acid with the echinocandin B nucleus.

4.3 Antifungal Target Identification

During the next stage of this project, OLID-FRET will be employed for high throughput target identification (Figure 4-2). We will generate nitroBIPS derivatives for dozens of antifungal compounds with unknown modes of action. After determining which of these derivatives retain their bioactivity, the active derivatives will be screened against the entire GFP fusion protein yeast library. This will serve as a striking demonstration of OLID-FRET as a rapid and robust technique for mode of action studies.
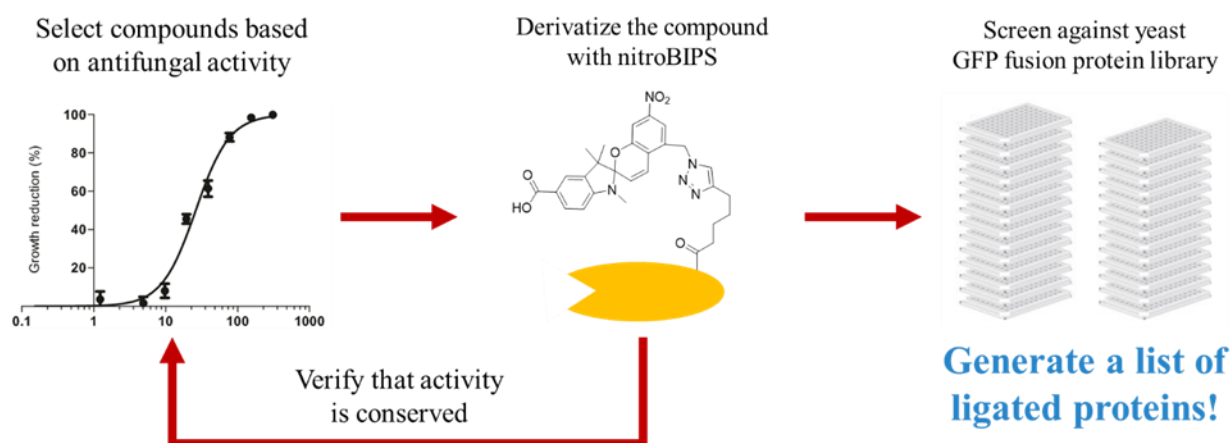
Figure 4-2. An overview of the workflow for identifying protein targets for antifungal natural products.

This work is intended to expand the range of antifungal drug classes, an important and unmet challenge in medicine. Presently, major antifungal drugs are limited to five classes: polyenes,

azoles, allylamines, flucytosine, and echinocandins. These drug classes are defined by their modes of action. Amazingly, three of these drug classes kill fungi by interfering with the synthesis or function of ergosterol, an important component of the fungal cell membrane. Polyenes bind directly to ergosterol, causing pore formation in fungal cell walls and leading to loss of the cytoplasm. Azoles and allylamines target different proteins in the biosynthetic pathway for ergosterol. Flucytosine inhibits DNA biosynthesis while Echinocandins interfere with cell wall biosynthesis by inhibiting 1,3-ß-glucan synthase. Due to this incredibly limited range of antifungal drug modalities, antifungal drug resistance is a growing threat.[3]

4.4 Drug Discovery

The main obstacle to using FRET to probe drug-protein interactions is that both the target protein and drug molecule must be fluorescently labeled. There is an abundance of validated GFP fusion protein genes available to purchase, and the genetic techniques to incorporate them into a cell line are well-established. In fact, choosing which of the plethora of potential targets to screen first is the real challenge. Here, we consider well-studied oncogenes with no associated drug that also have validated, GFP-labeled genes readily available. These include MYC, KRAS, and PTPRA among many others.[5, 6, 7]
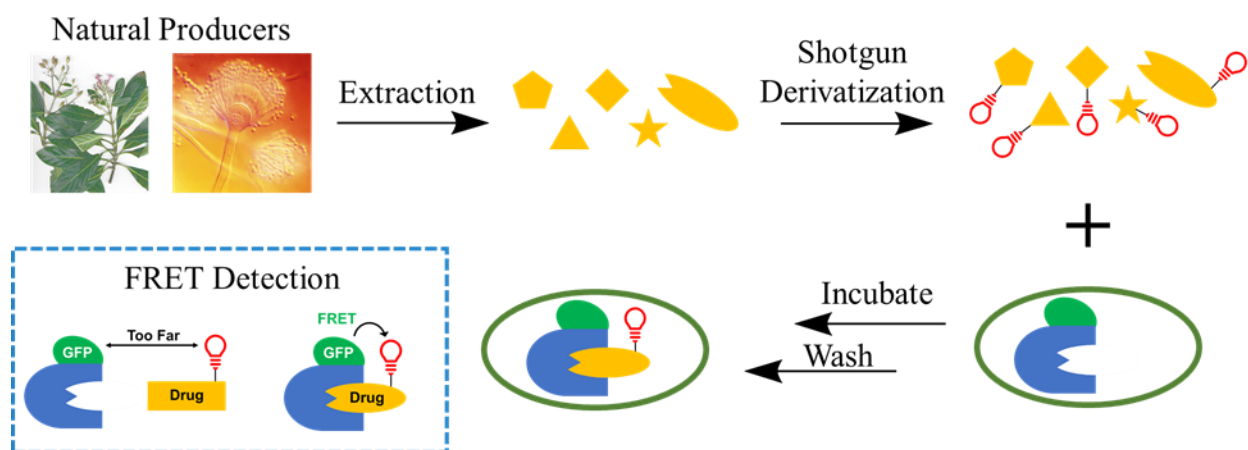
Figure 4-3. An overview of the workflow for performing drug screening via shotgun derivatization and live cell ligand binding assays.

The remaining challenge lies in conjugating the nitroBIPS tag to small molecules with sufficiently high throughput to facilitate drug discovery efforts. Fortunately, the shotgun chemical derivatization described in Chapter 3 is well-suited for this workflow (Figure 4-3). Ideally, this chemical derivatization could be employed to tag hundreds of crude extracts in multi-well plates However, the traditional Steglich esterification is incompatible with DMSO due to a reaction between carbodiimides and DMSO (Scheme 4-2A).[8] This can be avoided by using a symmetric anhydride as the tag, avoiding the need for a coupling reagent (Scheme 4-2B). While this reaction

is theoretically compatible with DMSO, the hydrolysis of the acid anhydride means that a very hygroscopic reagent like DMSO is not ideal (Scheme 4-2C).
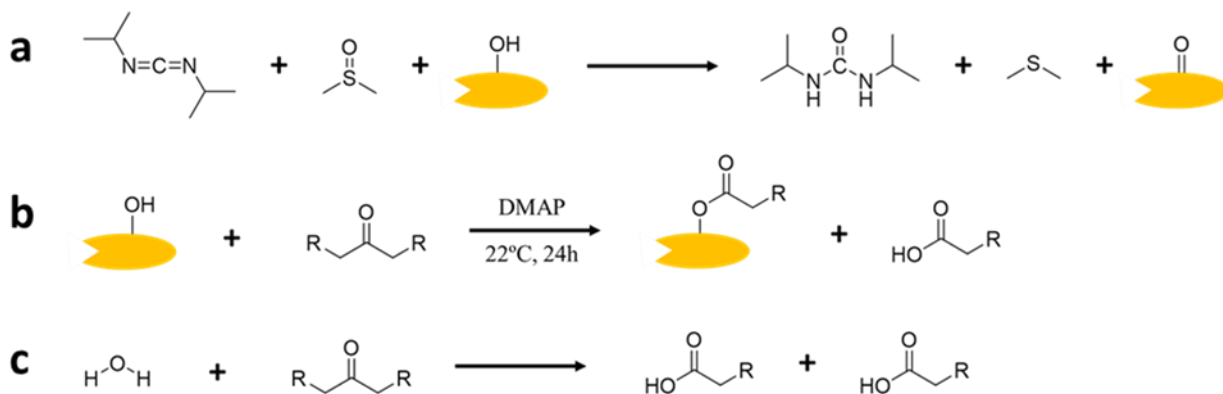
Scheme 4-2. A) Moffat oxidation occurs when diisopropylurea and DMSO are combined with an alcohol. B) Symmetric anhydride esterification can be catalyzed by DMAP without use of a coupling reagent. C) Symmetric anhydrides are easily hydrolyzed by water in the solvent.

Because extract libraries are typically shipped in DMSO, and multi-well plates that are not airtight, achieving a water-free environment is unrealistic. However, a compromise can be reached by diluting the DMSO-solvated extract into acetonitrile to perform the reaction. A preliminary study focused on derivatizing erythromycin with 6-hyptynoic anhydride in various ratios of acetonitrile to DMSO showed that good conversions could be achieved at higher acetonitrile amounts (Figure 4-4). While this study is limited, it serves as evidence that high-throughput derivatization of crude extract libraries is possible with this modified Steglich Esterification procedure.

The methods described so far are useful in determining if any compounds in an extract bind to a specific protein. However, the arduous task of isolating the compound of interest remains. The cell binding assays described in chapter three can be used to expedite the process compound isolation and repeat of the OLID-FRET assay on pure compounds. Although this will still be quite laborious, the overall procedure still presents several advantages over traditional drug discovery techniques. The compounds discovered through this approach will have proven themselves to bind to a specific, pharmaceutically valuable protein target inside of the chaotic environment of a living cell. This represents a meaningful change in assay results interpretation compared either phenotypic assays or traditional ligand binding assay.
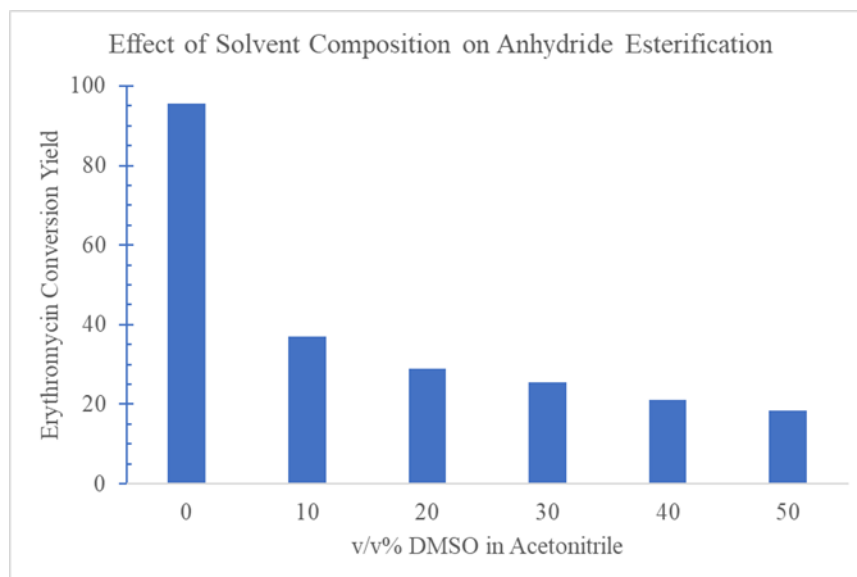
Figure 4-4. Symmetric anhydride esterification achieves higher yields in lower ratios of DMSO to Acetonitrile.

It will be necessary to re-optimize the OLID-FRET assay conditions in mammalian cells to facilitate screening for anticancer compounds. To support this proof-of-concept work, we have synthesized a nitroBIPS functionalized derivative of Kabiramide C (Scheme 4-3). The synthesis was confirmed by LCMS analysis. This synthesis is based on the work of Pretchprayoon et. al. that used similar fluorescent Kabiramide derivatives to image actin depolymerization.[9] Kabiramide C is a desirable probe because it is a potent binder to the barbed end of G-actin in cells.[10] To complete this proof of concept, nitroBIPS-kabiramide C will be incubated with mammalian cells expressing GFP-actin, and the OLID-FRET signal will be measured. Once completed, this will prove that binding of a nitroBIPS-tagged small molecule to a GFP fusion protein, can be detected via OLID-FRET in live mammalian cells. As such, the future completion experiment represents an important step in developing the proposed drug discovery platform.
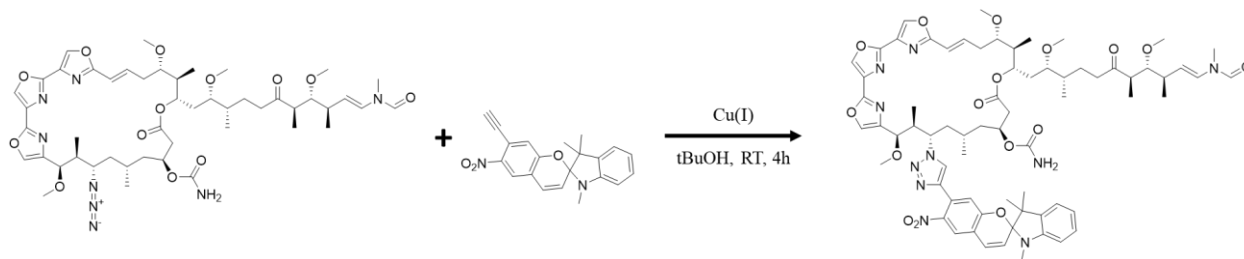
Scheme 4-3. nitroBIPS kabiramide C was formed by Cu(I) catalyzed click chemistry between azido-kabiramide C and alkyne-nitroBIPS.

4.5 Experimental Section

4.5.1 Echinocandin Derivatization

Pure echinocandin nucleus was purchased from BOC sciences while the nitroBIPS probe was provided by the Marriott lab. Echinocandin nucleus (10.0 mg and 0.0125 mmol) was converted to nitroBIPS-echinocandin via DIC amide coupling esterification with acid-functionalized azido-nitroBIPS. NitroBIPS acid (5.46 mg and 0.0125 mmol), and DIC (1.74 mg and 0.0138 mmol) were added to methylene chloride (10 mL) stirred in a round bottom flask and chilled in an ice bath. The round bottom was then evacuated and flushed with and kept under inert atmosphere. After stirring for one hour, the echinocandin nucleus was added to the flask. The reaction was allowed warm to room temperature and stirred for 24 hours. The reaction mixture was then extracted three times with water. The fractions were combined and back-extracted with methylene chloride. The two methylene chloride fractions were then combined and evaporated in vacuo. After dissolving the product in methanol, it was purified via HPLC.

4.5.2 Antifungal Activity Assays

Several colonies from 24-hour-oldculture of S. cerevisiae S288C on PDA plate were picked and suspended in 5 mL of sterile 0.85% saline. The resulting suspension was vortexed for 15 seconds and the cell density was adjusted with a spectrophotometer by adding sterile saline to increase the transmittance to that produced by a 0.5 McFarland standard at 530 nm. A working suspension was made by a 1:50 dilution followed by a 1:20 dilution with PDB medium. 100 mL of the working suspension was inoculated into each well of a 96-well plate. Antifungal agents were dissolved at concentrations 100 times higher than the highest desired test concentration in DMSO. Then a two times dilution series was prepared with DMSO. Working solutions were made by a 1:100 dilution with PDB medium. 100 mL of the working solutions were inoculated to each 100 mL cell suspension. The plates were cultured at 37ºC for 48 hours. MICs were defined as the lowest concentration that prevents visible growth of the microorganism.

4.5.3 Symmetric Anhydride Synthesis

Synthesis of 6-heptynoic anhydride was carried out via coupling of 6-heptynoic acid. 6-heptynoic acid (100 mg and 0.793 mmol) and DIC (150 mg and 1.19 mmol) were added to methylene chloride (10 mL) and allowed to react at room temperature for 24 hours. The symmetric anhydride was recovered via silica gel chromatography using a gradient from 100% methylene chloride to 100% ethyl acetate. Fractions were tested via thin layer chromatography in conjunction with iodine to visualize the alkyne bonds. Fractions were combined and evaporated under vacuum, yielding an oily product.

4.5.4 Erythromycin Esterification

In each of six separate reactions, erythromycin (1.00 mg and 0.00136 mmol) was esterified with 6-heptynoic anhydride (0.32 mg and 0.00136 mmol). Each reaction was catalyzed with DMAP

(0.16 mg and 0.00136 mmol). The reactions were run in 6 different ratios of DMSO in acetonitrile on a volume basis (0%, 10%, 20%, 30%, 40% and 50%). After allowing the reaction to run at room temperature for 24 hours, each sample was analyzed via LCMS. Reaction yields were determined by calculating the area of peaks obtained from the extracted ion chromatograms for erythromycin and the ester product. Yields ranged from 95.5% in pure acetonitrile to as low as 18.5% in equal parts acetonitrile and DMSO.

4.5.5 Kabiramide C Click Chemistry

Azido-kabiramide C was derivatized with alkyne-nitroBIPS. Azido-kabiramide C and alkyne-nitroBIPS were provided by the Marriott Lab. Azido-kabiramide C (10 mg and 0.0103 mmol), alkyne-nitroBIPS (7.16 mg and 0.0207 mmol), copper sulfate pentahydrate (0.26 mg and 0.00103 mmol), and sodium ascorbate (4.10 mg and 0.0207 mmol) were added to 10 mL of tert-butanol and stirred in a round bottom flask. After reacting for 4 hours, the mixture was concentrated under nitrogen flow, and redissolved in 1 mL MeOH. NitroBIPS-kabiramide C was not purified, but the finished structure was confirmed by LCMS analysis (Figure S28).

4.6 Conclusion

OLID-FRET represents such a substantial leap forward for molecular-scale proximity detection that it has the potential for application across the life sciences. FRET is routinely used to detect protein-protein interactions, protein conformational changes, and is even used in some illicit drug tests. Optical lock-in detection has the potential to improve the sensitivity and reliability of FRET in all of these applications across academia and industry. However, in order to reach such wide-spread adoption, OLID-FRET must be extensively optimized, and the necessary tools and software must be made available to the scientific community.

4.7 References

1.   S. Mao et al. *Biophys. J*, 2008, **94(11)**, 4515-4524.
2.   C. Galbraith and J. Galbraith. *J. Cell Sci,* 2011, **124**, 1607-1611.
3.   L. Cowen et. al. *CSH Perspectives in Medicine*, 2015, **5(7)**.
4.   W. Huh et. al. *Nature*, 2003, **425**, 686-691.
5.   Soucek, L. & Evan, G. I. *Current Opinion in Genetics & Development*, 2010, **20**, 91–95.
6.   Mccormick, F. *Clinical Cancer Research, 2015,* **21,** 1797–1801.
7.   Du, Y. & Grandis, J. R. *Chinese Journal of Cancer*, 2015, **34**, 61–69.
8.   K. E. Pfitzner and J. G. Moffatt. *J Am Chem Soc*, 1965, **87(24)**, 5661–5670.
9.   C. Petchprayoon et. al. *Bioconj Chem,* 2005, **12(6)**, 1382-1389.
10.   J. Tanaka et. al. *Proc Nat Acad Sci,* 2003, **100(24)**, 13851-13856.

Appendix

1. Python Code

```python
import numpy as np
import pandas as pd

# import dataset
data = np.genfromtxt('HolyBa_HeLa.csv',delimiter=',')

# define columns of table as separate variables
time = data[0:,0]
mass = data[0:,1]
intensity = data[0:,2]

# Define and edit constants
mC13 = 13.003355-12  # mass C13 - mass C12 (amu)

ppm = 2
error = ppm/1000000

mintime = -6/60  # minutes
maxtime = 6/60  # minutes

ratio1 = 10/10*0.67  # Peak 2 intensity/Peak 1 intensity lower bound
ratio2 = 10/10*1.5  # Peak 2 intensity/Peak 1 intensity upper bound

# create a hypothetical list of M+3 values from the data
n = len(time)  # The length of my dataset
m3guess = list()

i = 0
while i < n:
    m3 = mass[i] + 3*mC13
    m3guess.append(m3)
    i = i + 1

# compare m3guess to mass

i = 500
j = 0

dataMAT = []
```

```
while i < n - 500:
    j = i - 500
    while j < i + 500:
        if mass[i] > m3guess[j] * (1 - error) and mass[i] <
m3guess[j] * (1 + error) and time[i] > time[j] + mintime and time[i]
< time[j] + maxtime and intensity[i] > intensity[j]*0.8*ratio1 and
intensity[i] < intensity[j]*0.8*ratio2:
            dataMAT.append([mass[j], time[j], intensity[j], mass[i],
time[i], intensity[i] ])
            print(dataMAT[-1])
            break
        else:
            j = j + 1
    i = i + 1

data = pd.DataFrame(dataMAT,columns = ['mass 1','time 1','intensity
1','mass 2','time 2','intensity 2'])
```

2. NMR Tables

PhDY-Ant (5)

^{1}H NMR (900 MHz, DMSO-d_6) ∂ 12.50 (s, 1H), 9.59 (s, 1H), 9.26 (d, J = 8.3 Hz, 1H), 7.96 (d, J = 7.9 Hz, 1H), 7.88 (d, J = 6.6 Hz, 1H), 7.53 (d, J = 7.2 Hz, 2H), 7.44 (t, J = 7.5 Hz, 1H), 7.40 (t, J = 7.7 Hz, 2H), 6.93 (t, J = 7.9 Hz, 1H), 5.68 (d, J = 8.4 Hz, 1H), 5.52 (p, J = 6.8 Hz, 1H), 5.25 (app t, J = 7.9 Hz, 1H), 4.67 (dq, J = 9.5, 6.3 Hz, 1H), 3.31-3.30 (overlap m, J = 8.6 Hz, 1H), 2.72 (d, J = 3.8 Hz, 4H), 2.17 (dt, J = 10.5, 3.0 Hz, 1H), 2.03 – 1.95 (m, 1H), 1.83 (tdd, J = 11.2, 6.2, 2.9 Hz, 1H), 1.51-1.48 (m, 1H), 1.32 (d, J = 6.3 Hz, 3H), 1.27 (d, J = 6.8 Hz, 3H), 1.09 – 1.02 (m, 2H), 0.85 (d, J = 6.6 Hz, 3H), 0.84 (d, J = 6.7 Hz, 3H). ^{13}C NMR (226 MHz, DMSO-d_6) ∂ 174.9, 170.0, 169.8, 168.8, 151.1, 132.4, 129.6, 128.8, 127.2, 127.0, 124.3, 120.7, 118.4, 115.4, 84.9, 76.2, 76.0, 75.0, 74.2, 70.8, 64.9, 53.8, 52.1, 36.1, 34.1, 27.5, 26.5, 22.6, 22.1, 18.4, 15.20, 15.06, 13.97. MS (ESI) calcd for $C_{35}H_{39}N_2O_9$ [M+H]$^+$ m/z 631.2650 found 631.2649.

PhDY-NeoA (6)

^{1}H NMR (900 MHz, DMSO-d_6) ∂ 12.62 (s, 1H), 9.64 (s, 1H), 9.31 (br s, 1H), 8.13 (s, 1H), 7.98 (d, J = 7.9 Hz, 1H), 7.74 (d, J = 7.7 Hz, 1H), 7.52 (d, J = 7.7 Hz, 2H), 7.44 (t, J = 7.3 Hz, 1H), 7.39 (t, J = 7.6 Hz, 2H), 7.32 – 7.27 (m, 4H), 7.22 (t, J = 6.9 Hz, 1H), 6.92 (br s, 1H), 5.39 – 5.34 (m, 2H), 5.18 (d, J = 6.1 Hz, 1H), 5.13 (d, J = 3.4 Hz, 1H), 5.08 (dt, J = 10.0, 3.6 Hz, 1H), 4.95 (dd, J = 8.4, 5.8 Hz, 1H), 3.82 (app t, J = 5.4 Hz, 1H), 2.91 (dd, J = 14.6, 9.7 Hz, 1H), 2.81 (dd, J = 14.5, 3.6 Hz, 1H), 2.73-2.71 (m, 4H), 2.26-2.22 (m, 1H), 1.69-1.66 (m, 1H), 1.36-1.32 (m, 1H), 1.32 (d, J = 6.3 Hz, 3H), 1.28 (s, 3H), 1.11 – 1.07 (m, 1H), 1.06 (s, 3H), 1.00 (d, J = 6.8 Hz, 3H), 0.88 (d, J = 6.9 Hz, 3H), 0.77 (d, J = 6.9 Hz, 3H), 0.75 (t, J = 7.4 Hz, 2H). ^{13}C NMR (226 MHz, DMSO) ∂ 186.5, 184.2, 179.4, 179.3, 173.0, 169.7, 168.4, 163.5, 159.0, 137.5, 135.2, 132.8, 132.6, 130.0, 129.6, 129.3, 128.82, 128.78, 127.0, 121.1, 118.5, 115.6, 95.2, 85.3, 80.0, 79.1, 77.1, 75.4, 74.7, 74.6, 69.5, 62.6, 56.1, 44.6, 38.6, 34.9, 34.5, 30.0, 24.1, 22.5, 19.1, 18.4, 17.3, 16.6, 15.6, 15.5, 11.8. MS (ESI) calcd for $C_{49}H_{55}N_2O_{13}$ [M+H]$^+$ m/z 879.3699, found 879.3706.

Deformylated NeoA (7)

^{1}H NMR (900 MHz, MeOD) ∂ 7.3 (app d, J = 4.4, 4H), 7.2 (m, 2H), 6.93 (dd, J = 7.7, 1.5, 1H), 6.72 (t , J = 7.9, 1H), 5.64 (m, 1H), 5.42 (d , J = 3.7, 1H), 5.19 (d , J = 4.2, 1H), 5.1 (d , J = 3.3, 1H), 5 (dt , J = 8.4, 3.9, 1H), 4.04 (d , J = 5.1, 1H), 3.05 (dd , J = 14.7, 9.3, 1H), 2.93 (dd , J = 14.6, 4.1, 1H), 2.34 (m, 1H), 1.8 (m, 1H), 1.5 (m, 1H), 1.4 (d , J = 6.4, 3H), 1.35 (s, 3H), 1.23 (m, 1H), 1.18 (s, 3H), 1.11 (d , J = 6.9, 3H), 0.98 (d , J = 6.8, 3H), 0.92 (d , J = 6.83, 3H), 0.88 (t , J = 7.5, 3H). ^{13}C NMR (226 MHz, MeOD) ∂ 181.4, 174.0, 172.8, 170.6, 169.7, 138.2, 138.1, 130.7, 130.31, 130.30, 129.6, 127.8, 125.2, 124.8, 120.1, 119.7, 81.8, 80.4, 78.6, 76.5, 71.7, 56.8, 46.1, 40.1, 36.3, 31.4, 30.8, 25.2, 23.0, 19.5, 18.9, 17.3, 17.0, 15.6, 12.0. MS (ESI) calcd for $C_{35}H_{47}N_2O_{11}$ [M+H]$^+$ m/z 671.3174, found 671.3236.

Deformylated PhDY-NeoA (8)

Deformylated PhDY-NeoA (8) NMR table includes 2D NMR data to ensure PhDY tag had not migrated during acid degradation. Atom numbers kept the same as Takeda *et al.*[3] to avoid confusion.
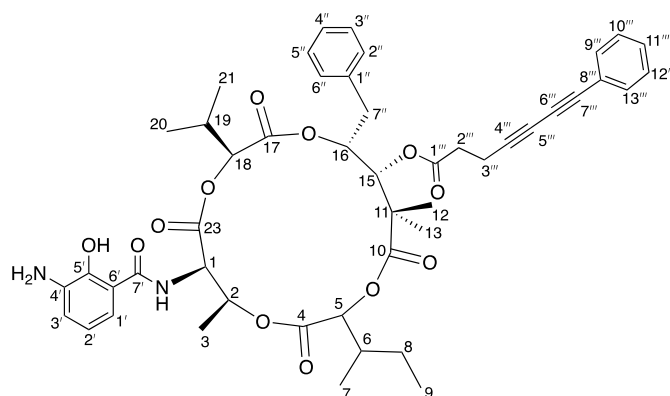
position	δ_H [ppm]	integration	mult. (J in Hz)	^{1}H-^{13}C HSQC	^{1}H-^{13}C HMBC	δ_C [ppm]
1	5.14	1	br s	1		57.0
2	5.63	1	m	2		72.0
3	1.43	3	d (6.32)	3	2, 1, 19, 13	17.3
4						169.9
5	5.11	1	d (3.3)	5	4, 19, 3	78.5
6	1.5	1	m	8	3	23.4
7	0.96	3	d (6.86)	7	5, 8	15.9
8	1.26	1	m	8		25.1
	0.94	1	m	8	7	
9	0.91	3	t (7.46)	9	8	12.0
10						181.5
11						46.0
12	1.18	3		12	10, 18, 11, 13	18.9
13	1.35	3		13	10, 18, 11, 12	23.0
15	4.06	1	d (4.84)	15	1''', 2''', 8, 7	76.4
16	4.98	1	dt (9.40, 3.91)	16		82.0
17						181.4
18	5.42	1	d (3.68)	18	10, 4, 13	80.3
19	2.34	1	m	19	12	31.3
20	1.01	3	d (6.89)	20	5, 19, 21	17.0
21	1.13	3	d (6.92)	21	5, 19, 20	19.6
23						169.7
1'	7.67	1	br s	1'		124.5
2'	8.09	1	br s		1'	129.7
3'	6.62	1	br s			96.1
4'						133.1
1"						138.3
6'						118.0
7'						162.8
5'						135.0
2", 6", 3", 5"	7.29	4	app d (4.4)	2", 6", 3", 5"	4", 1", 7"	130.3
4"	7.22	1	p (4.35)	4"		127.7
7"	3.04	1	dd (14.72, 9.41	7"	2", 6", 3", 5", 16	36.3
	2.93	1	dd (14.68, 4.01	7"	2", 6", 3", 5", 16	
1'''						171.2
2'''	2.74	2	m	2''', 3'''	1''', 4''', 5''', 6''', 7'''	36.6
3'''	2.79	2	m	2''', 3'''	1''', 4''', 3'''	16.6
4'''						83.7
5'''						74.9
6'''						66.4
7'''						75.8
8'''						123.2
9''', 13'''	7.46	2	m	9''', 13'''	9''', 13''', 8''', 7'''	133.4
10''', 12'''	7.35	2	m	10''', 12'''	8''', 10''', 12'''	129.5
11'''	7.38	1	t t (7.4, 1.3)	11'''		130.1
(NH)	1.89	1	br s			

58

3. NMR, MS, and UV-vis characterization of PhDY-Ant (5)

A
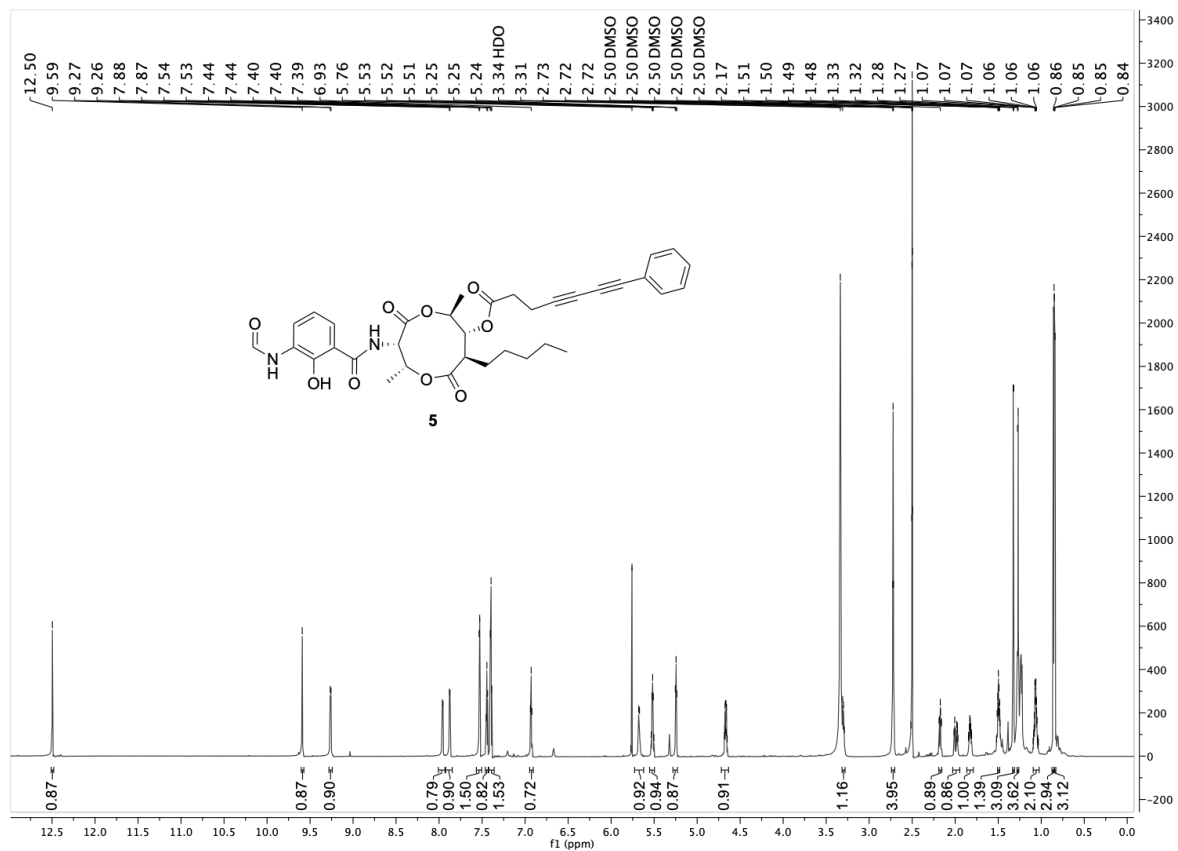
Figure S1: NMR, MS, and UV-vis characterization of PhDY-Ant (5). A) ^{1}H NMR spectrum (DMSO-d_6, 900 MHz) of **2**.

B
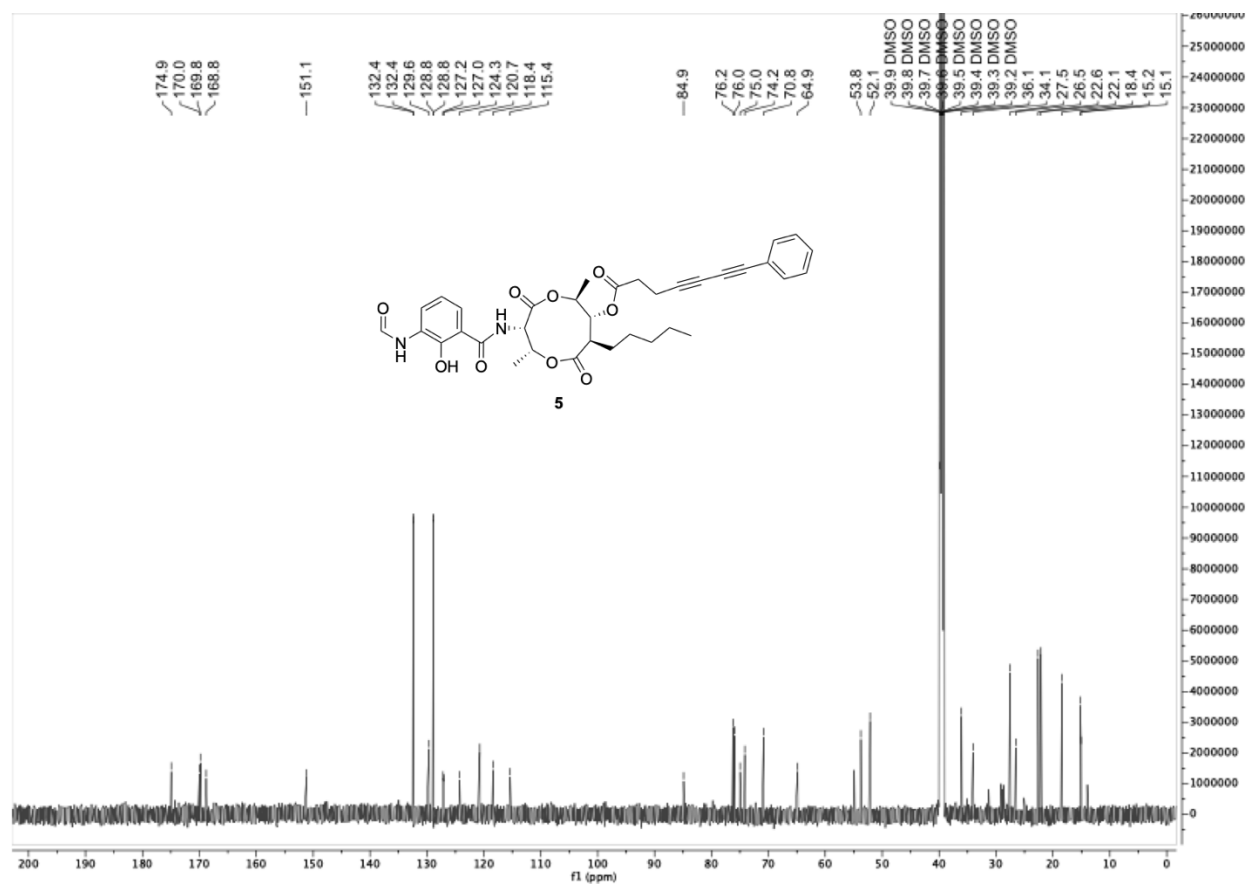
Figure S1: B) ^{13}C NMR spectrum (DMSO-d_6, 226 MHz) of **2**.

C
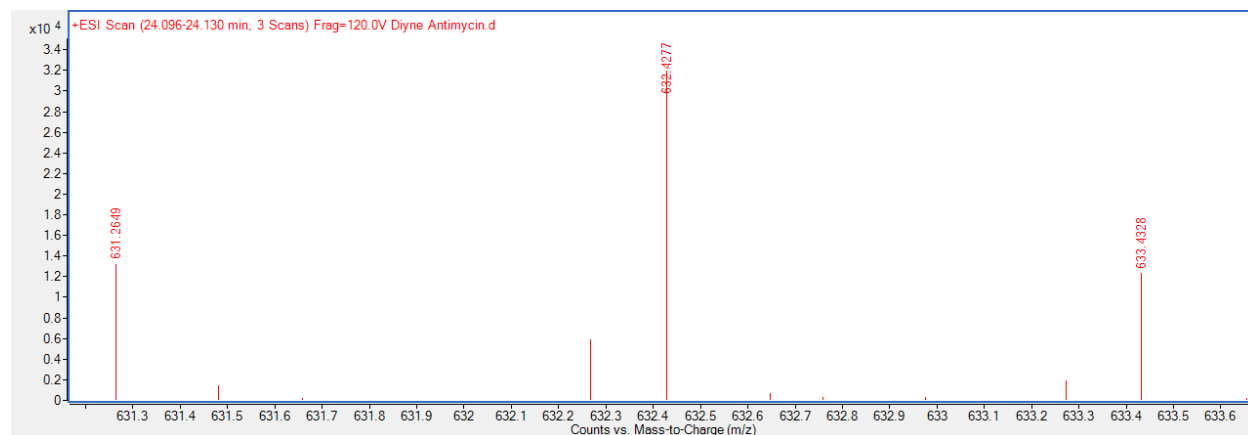
Figure S1: C) Mass spectrum of **2**. Calcd for $C_{49}H_{55}N_2O_{13}$ [M+H]$^+$ m/z 631.3650, found 631.3649.

60

D
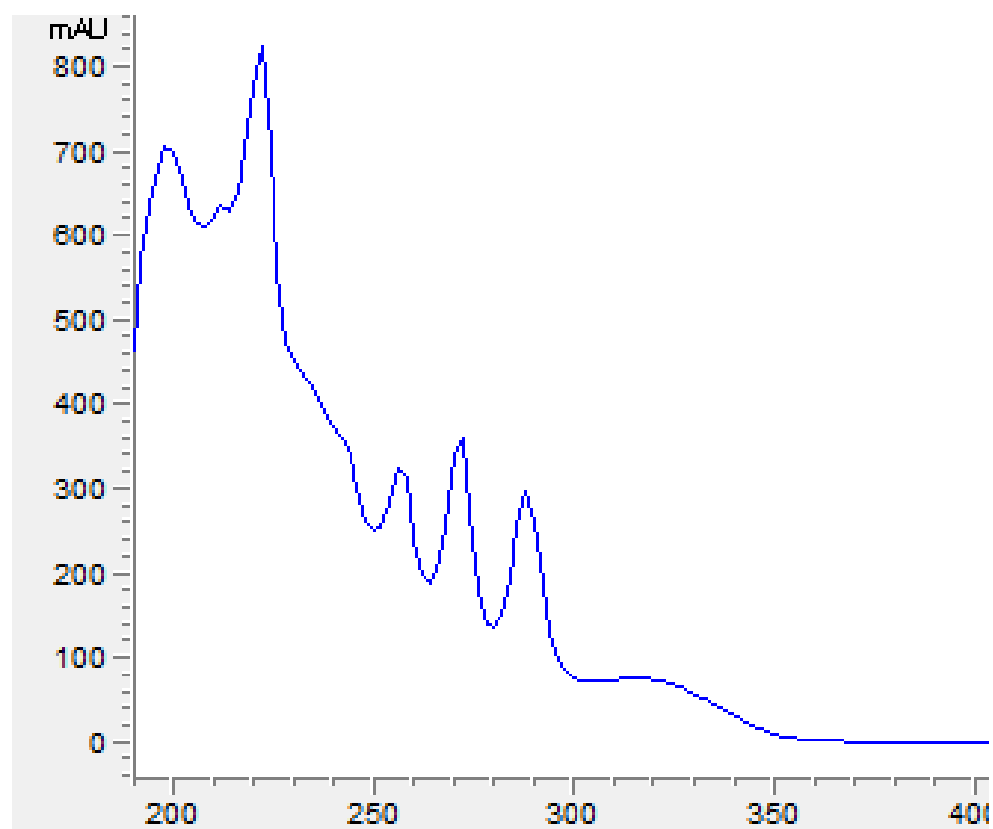
Figure S1: D) UV-vis spectrum of **2**.

4. PhDY-Ant (2) and antimycin (1) dose response curves
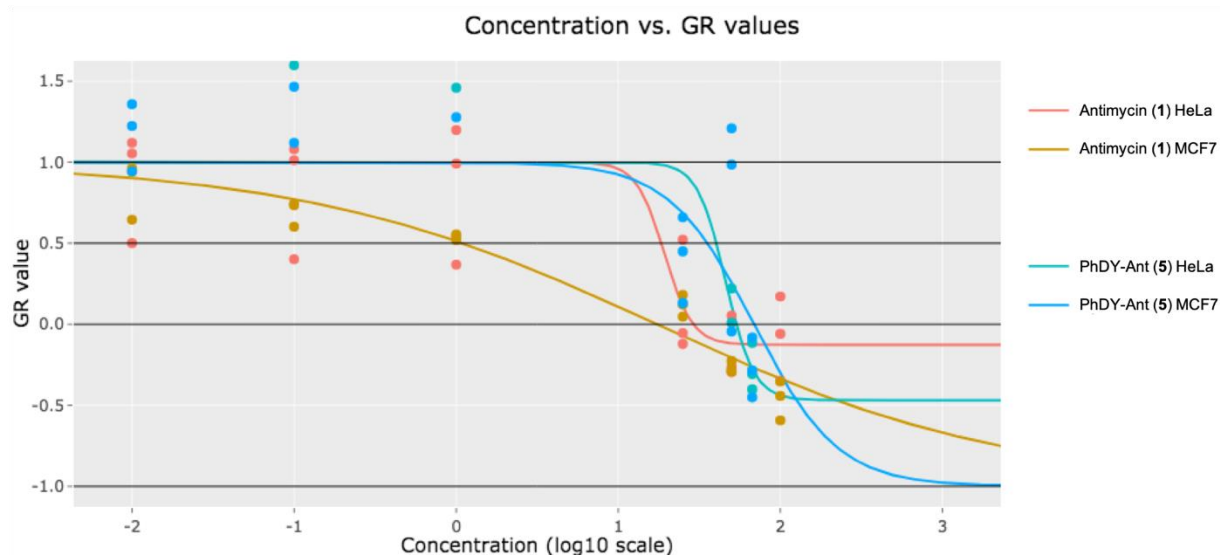
Figure S2. PhDY-Ant (**2**) and antimycin (**1**) dose response curves. Antimycin (100 µM to 0.010 µM, red and gold), and PhDY-Ant (**2**) (67 µM to 0.010 µM, teal and blue lines) dose response curves in HeLa (red and teal lines) and MCF-7 (gold and blue lines) cells show the PhDY Raman tag minimally perturbs activity. Concentration of compounds is in µM (log scale). Data were collected in triplet and growth response curves were generated as described above. The curve shown is representative of at least three biological replicates.

5. NMR, MS, and UV-vis characterization of PhDY-NeoA (4)

A
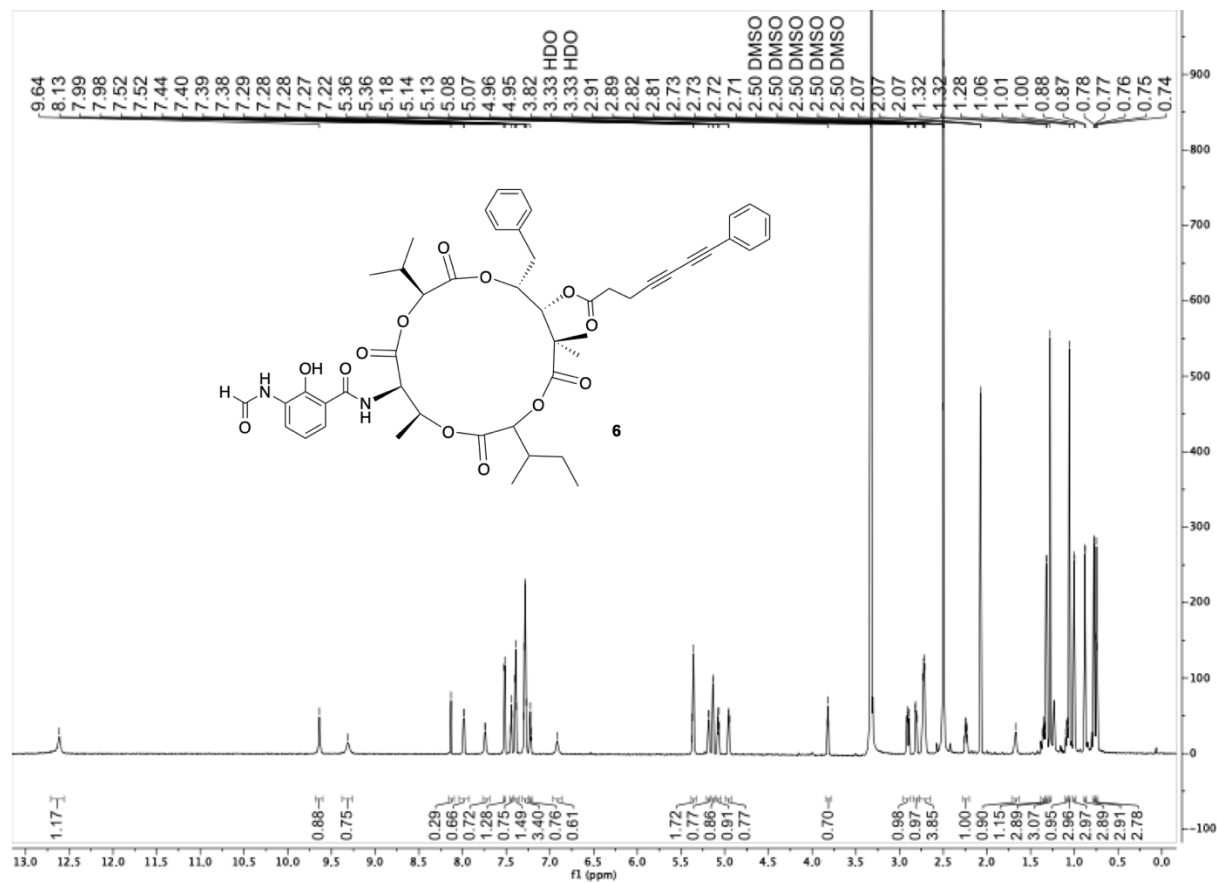
Figure S3. NMR, MS, and UV-vis characterization of PhDY-NeoA (**4**). A) ^{1}H NMR spectrum (DMSO-d_6, 900 MHz) of **4**.

B
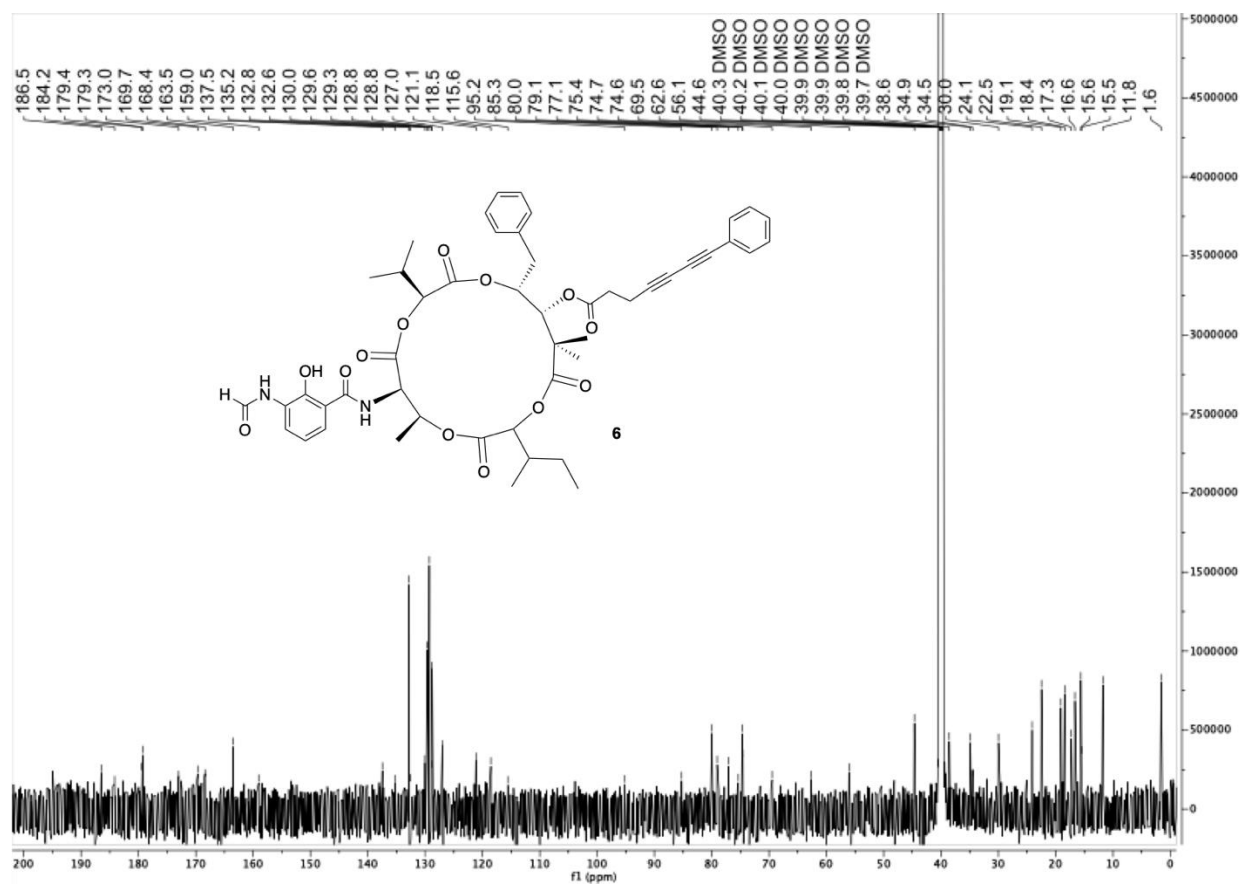
Figure S3. B) ^{13}C NMR spectrum (DMSO-d_6, 900 MHz) of **4**.

C
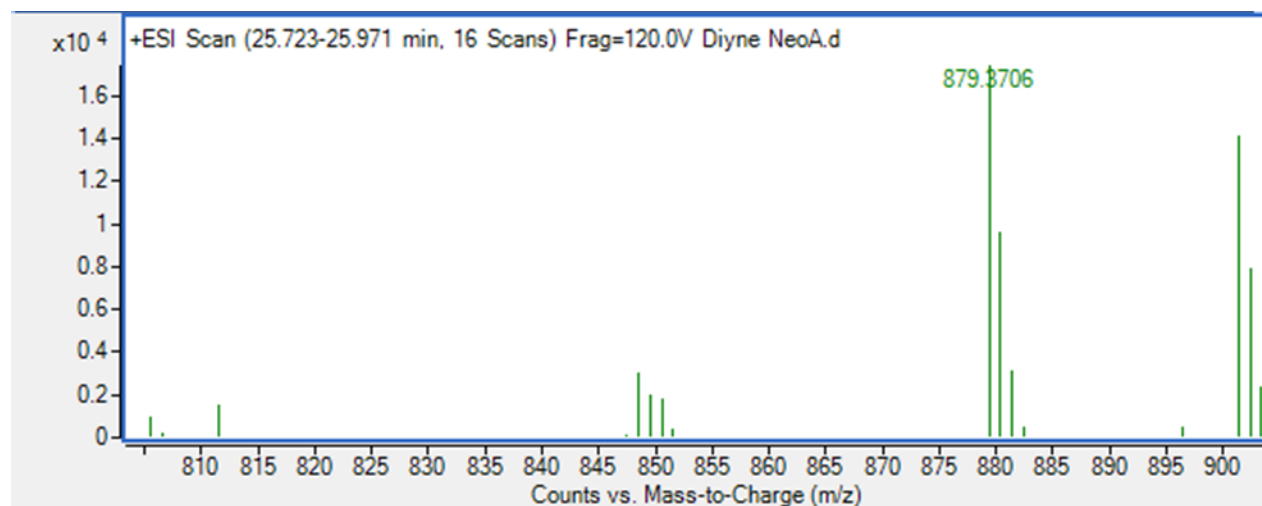
Figure S3. C) Mass spectrum of **4**. Calcd for $C_{49}H_{55}N_2O_{13}$ [M+H]$^+$ *m/z* 879.3699, found 879.3706.

D
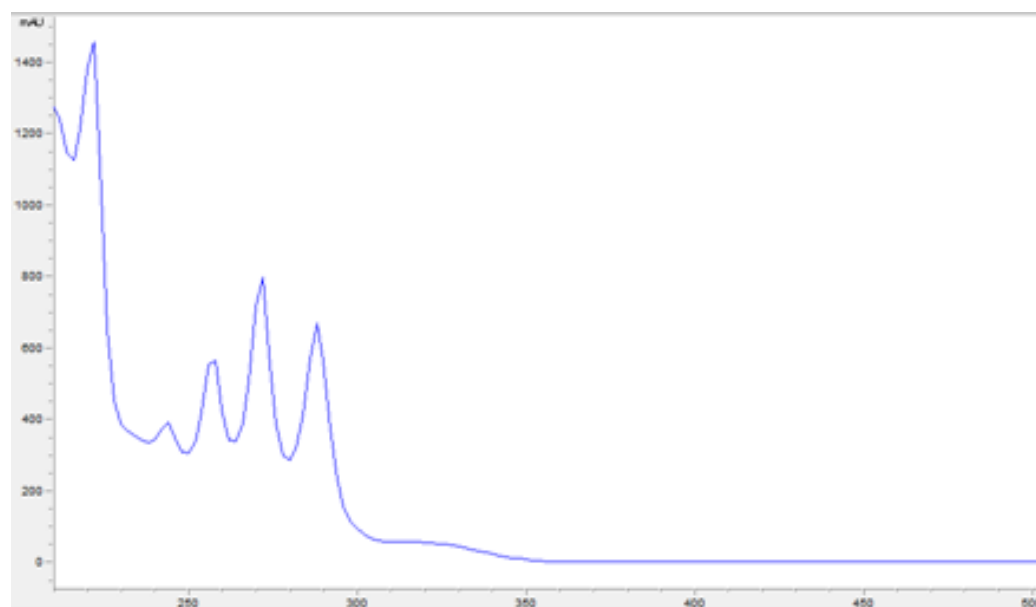
Figure S3. D) UV-vis spectrum of **4**.

6. NMR, MS, and UV-vis characterization of deformylated neoantimycin (5)

A
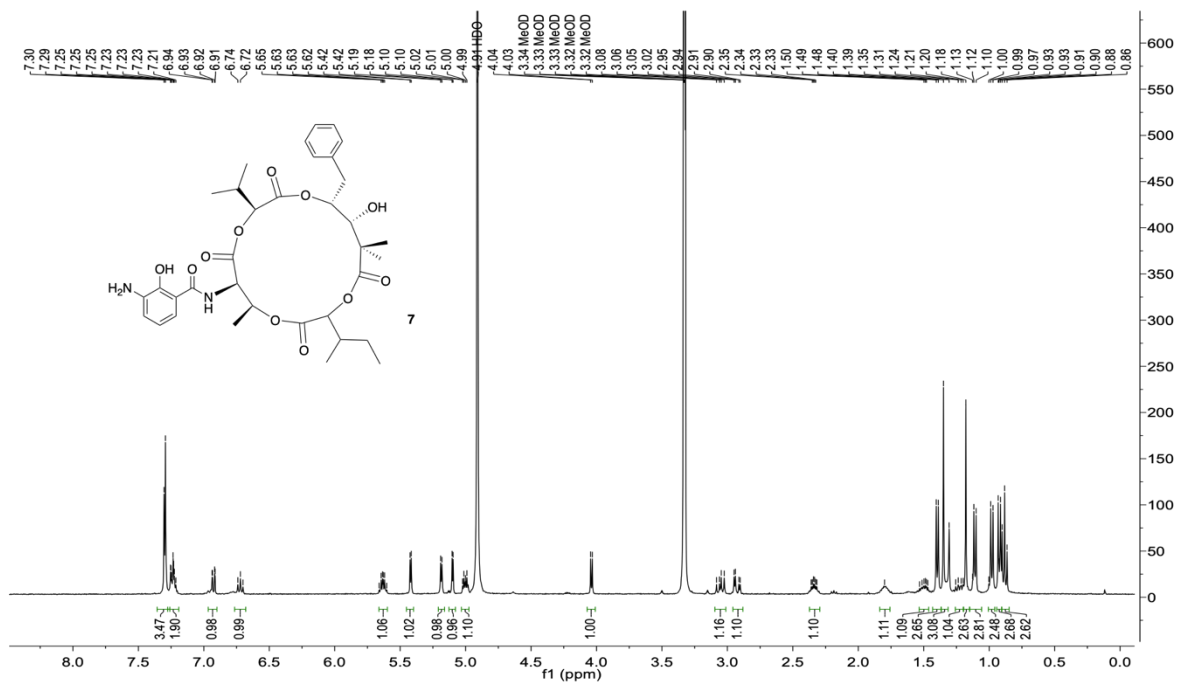
Figure S4. NMR, MS, and UV-vis characterization of deformylated neoantimycin (**5**). A) ^{1}H NMR spectrum (DMSO-d_6, 900 MHz) of **5**.

B
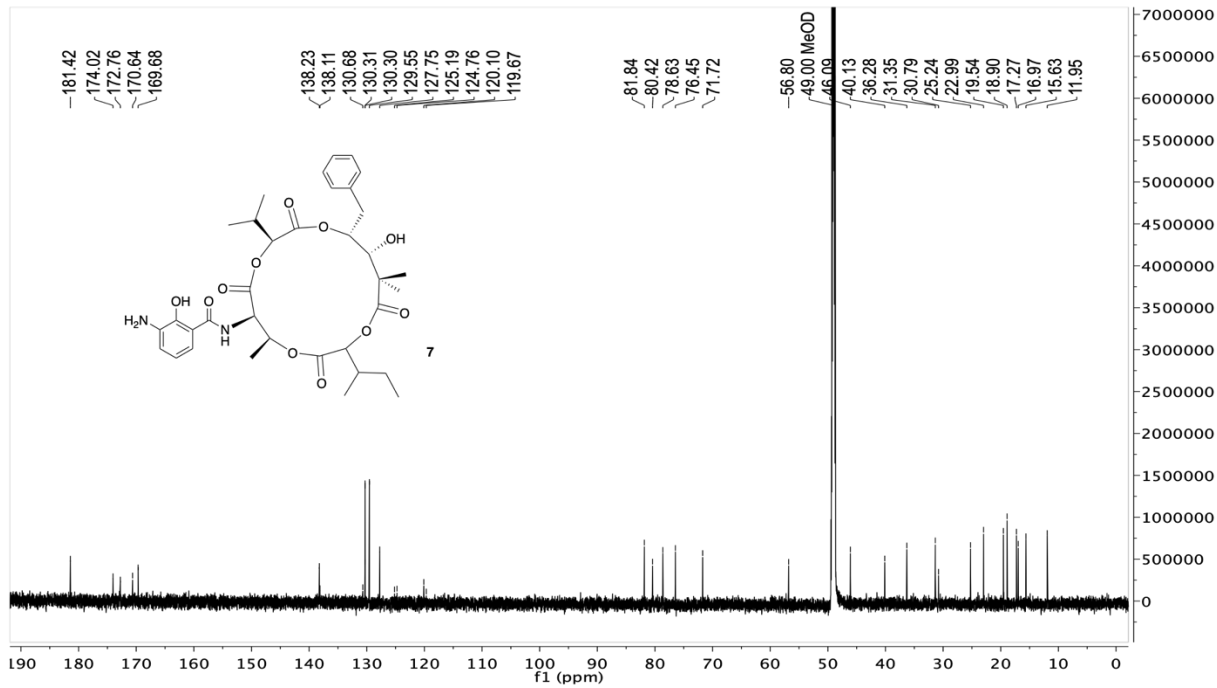
Figure S4. B) ^{13}C NMR spectrum (DMSO-d_6, 226 MHz) of **5**.

C

+ESI Scan (12.704 min) Frag=100.0V wp-01-71-f3.d

x10 7

667.9159
670.3094
*671.3236
672.3234
673.3244
674.3267
675.3421
676.3694
680.0100

Counts vs. Mass-to-Charge (m/z)
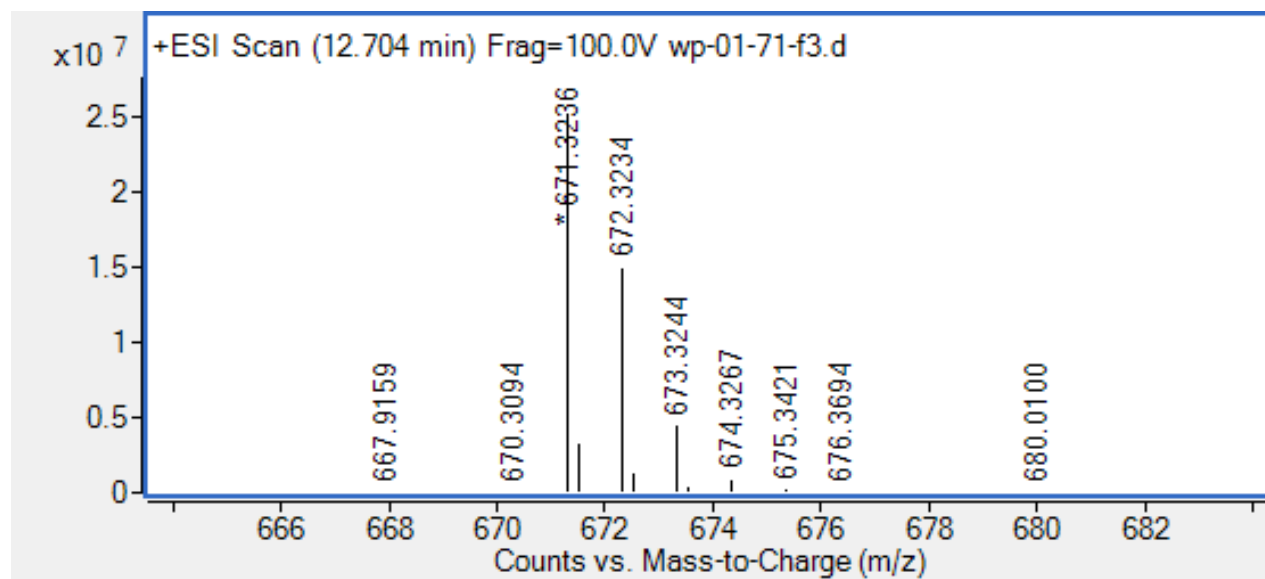
Figure S4. C) Mass spectrum of compound **5**.

D
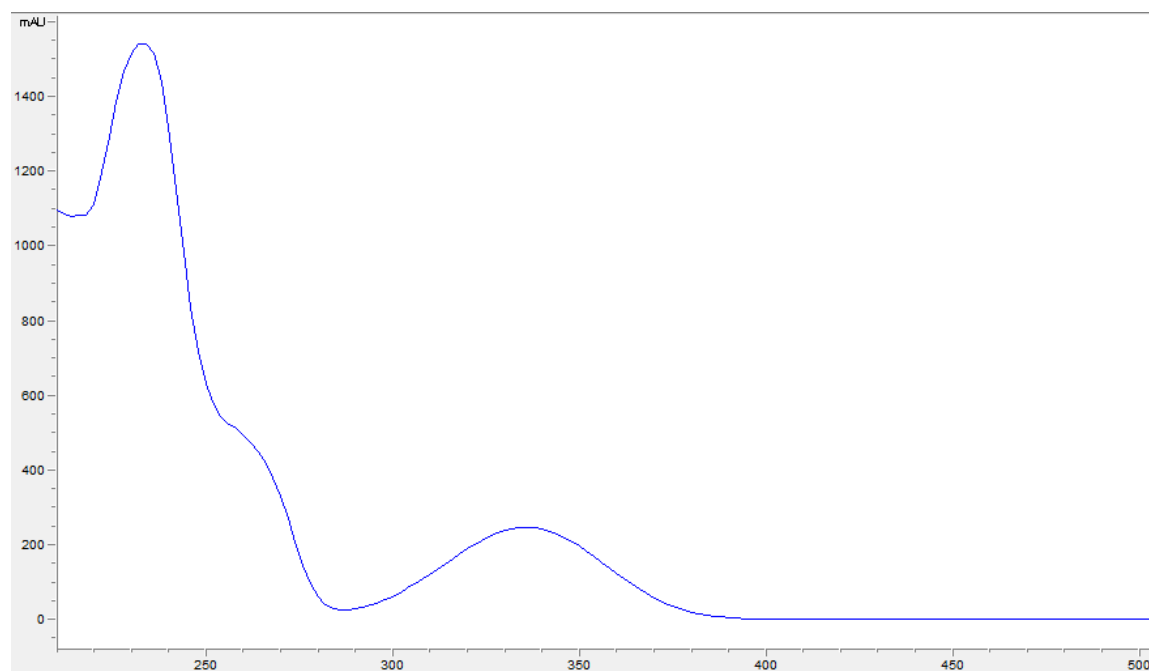
Figure S4. C) UV-vis spectrum of **5**.

7. NMR, MS and UV-vis characterization of deformylated PhDY-NeoA (6)

A
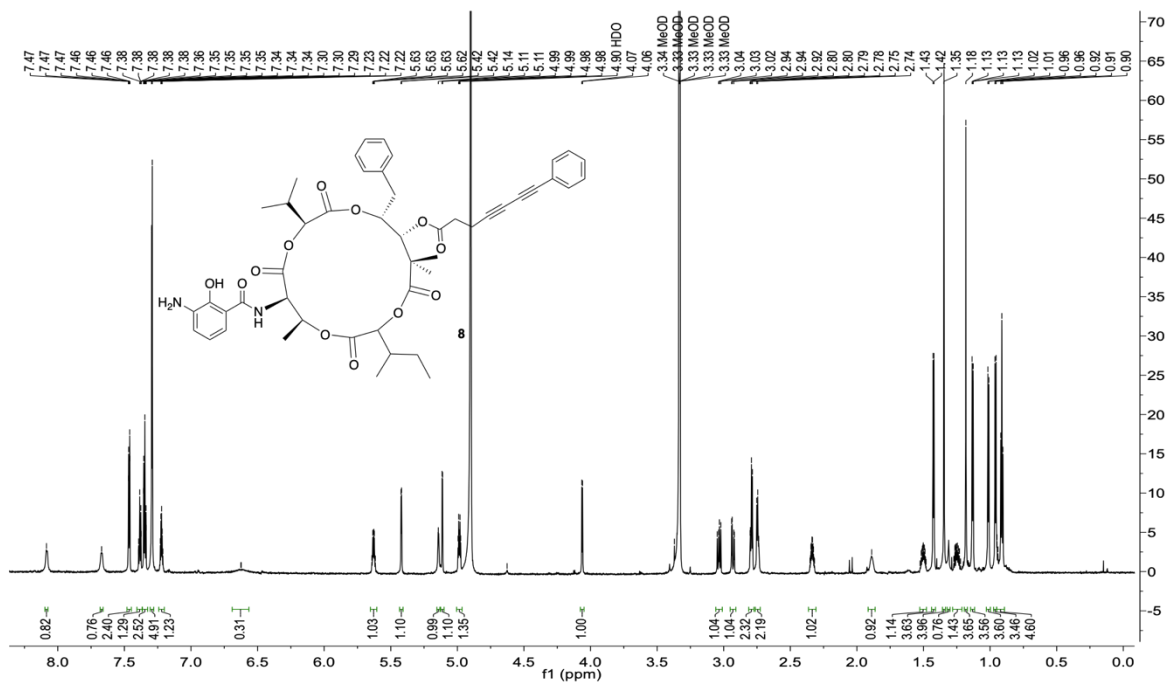
Figure S5. NMR, MS and UV-vis characterization of deformylated PhDY-NeoA (**6**). A) ^{1}H NMR (DMSO-d_6, 900 MHz) of **6**.

B
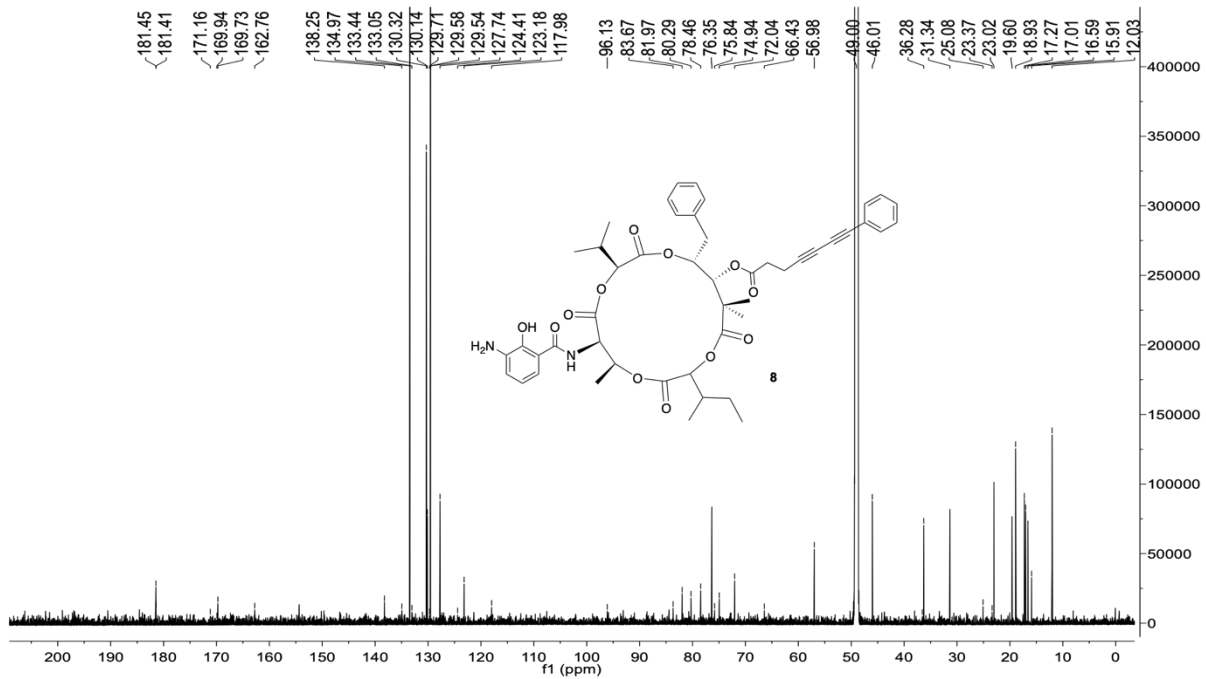
Figure S5. B) ^{13}C NMR (DMSO-d_6, 226 MHz) of **6**.

C
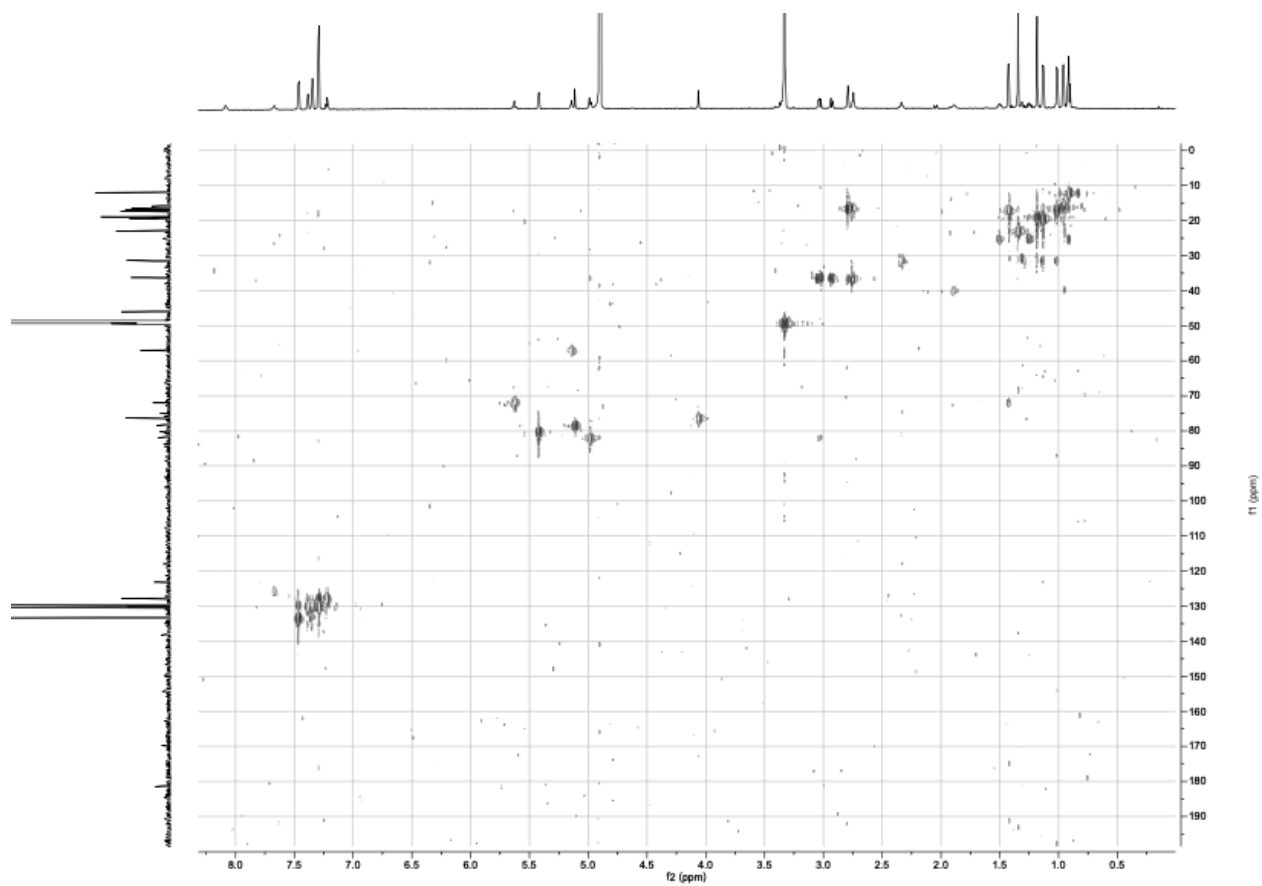
Figure S5. C) HSQC spectrum (DMSO-d_6, 900 MHz) of **6.**

D
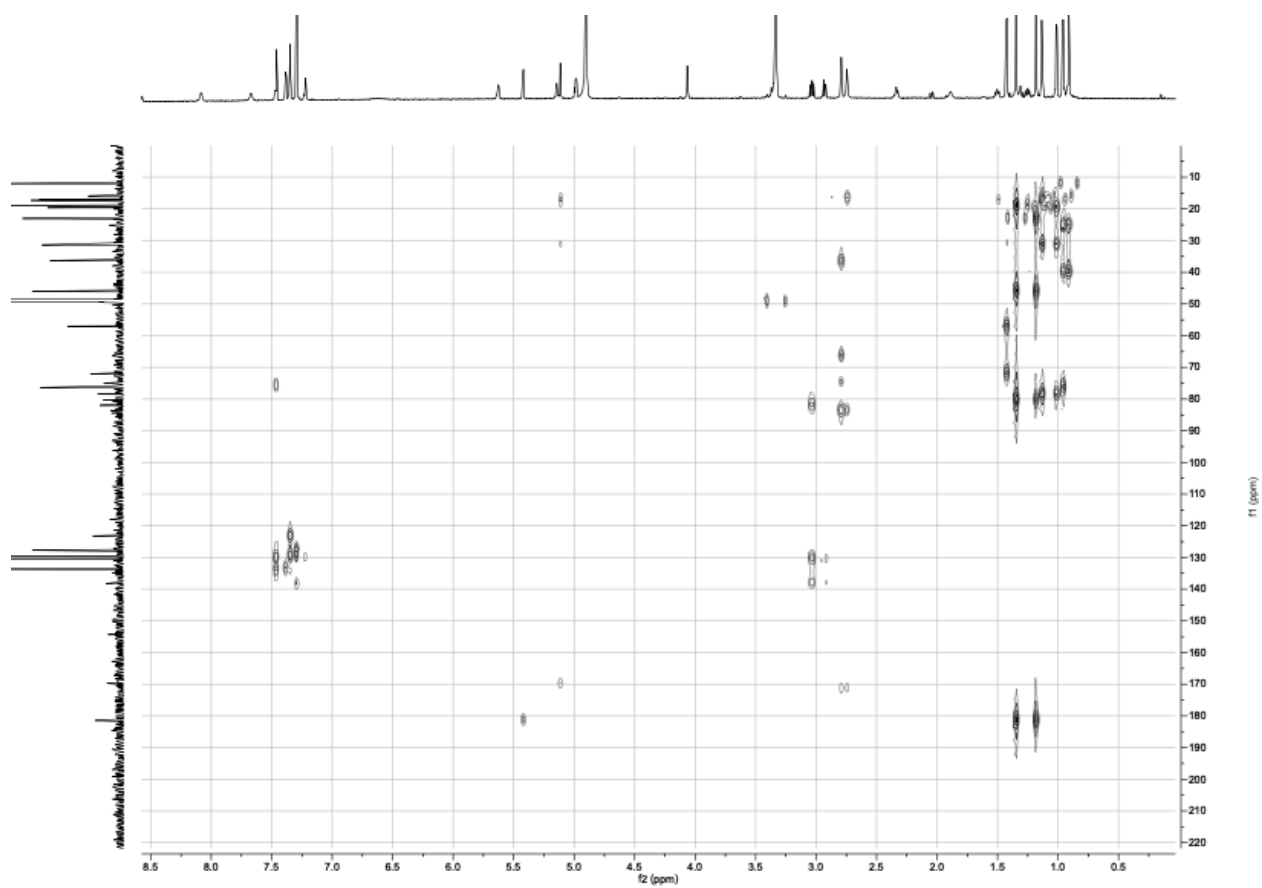
Figure S5. D) HMBC spectrum (DMSO-d_6, 900 MHz) of **6.**

E
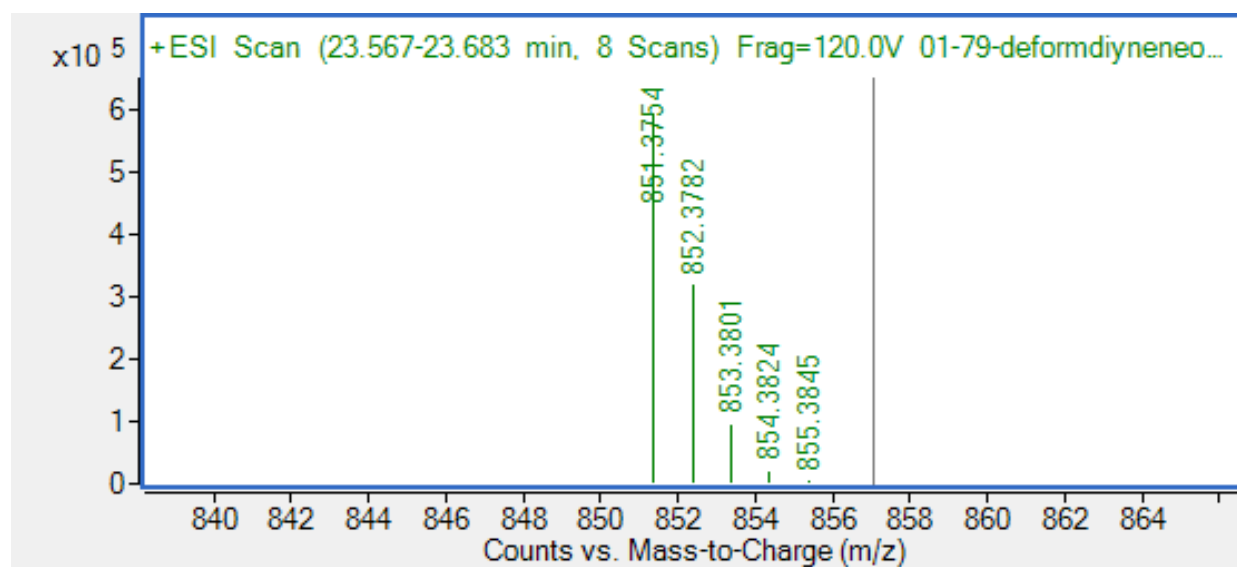
Figure S5. E) Mass spectrum of **6**. MS (ESI) calcd for $C_{48}H_{55}N_2O_{12}$ [M+H]$^+$ *m/z* 851.3750, found 851.3754.

F
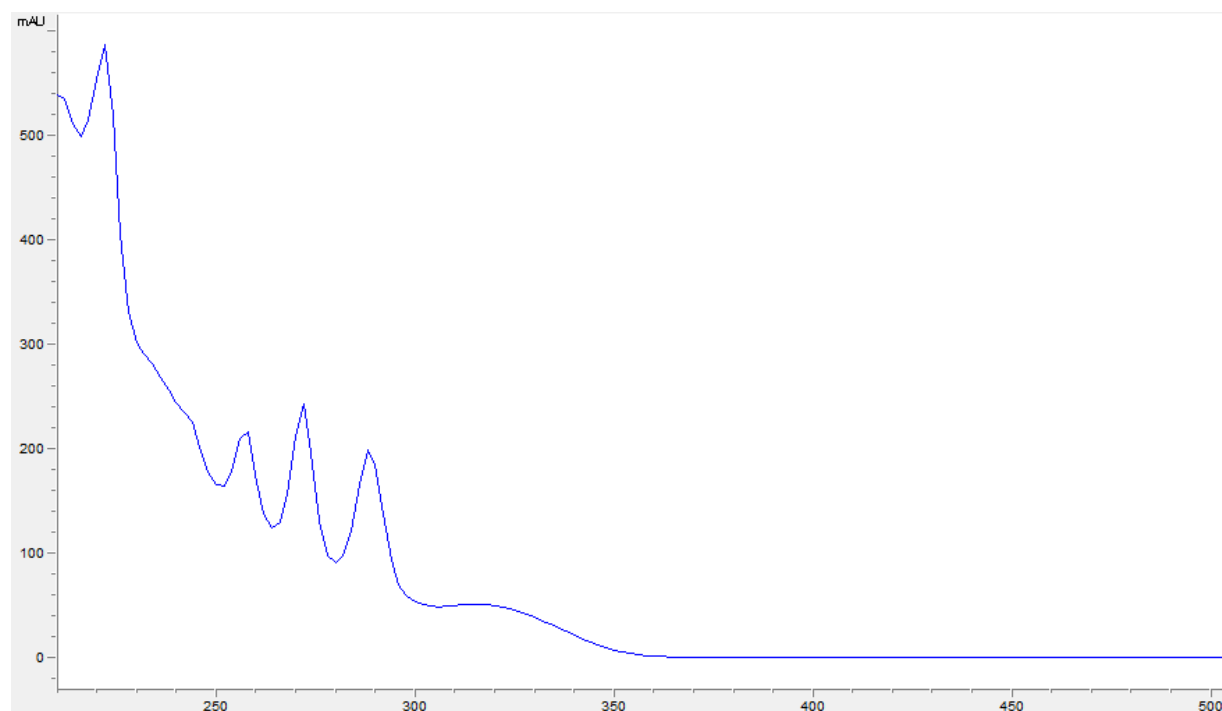
Figure S5. F) UV-vis spectrum of compound (**6**).

8. Mass Spectrometry Data for Crude Natural Products Extracts

RT: 0.0 - 60.0

NL:
1.29E5
m/z=
588.37-
588.37 MS
3_HolyBa_
Steg
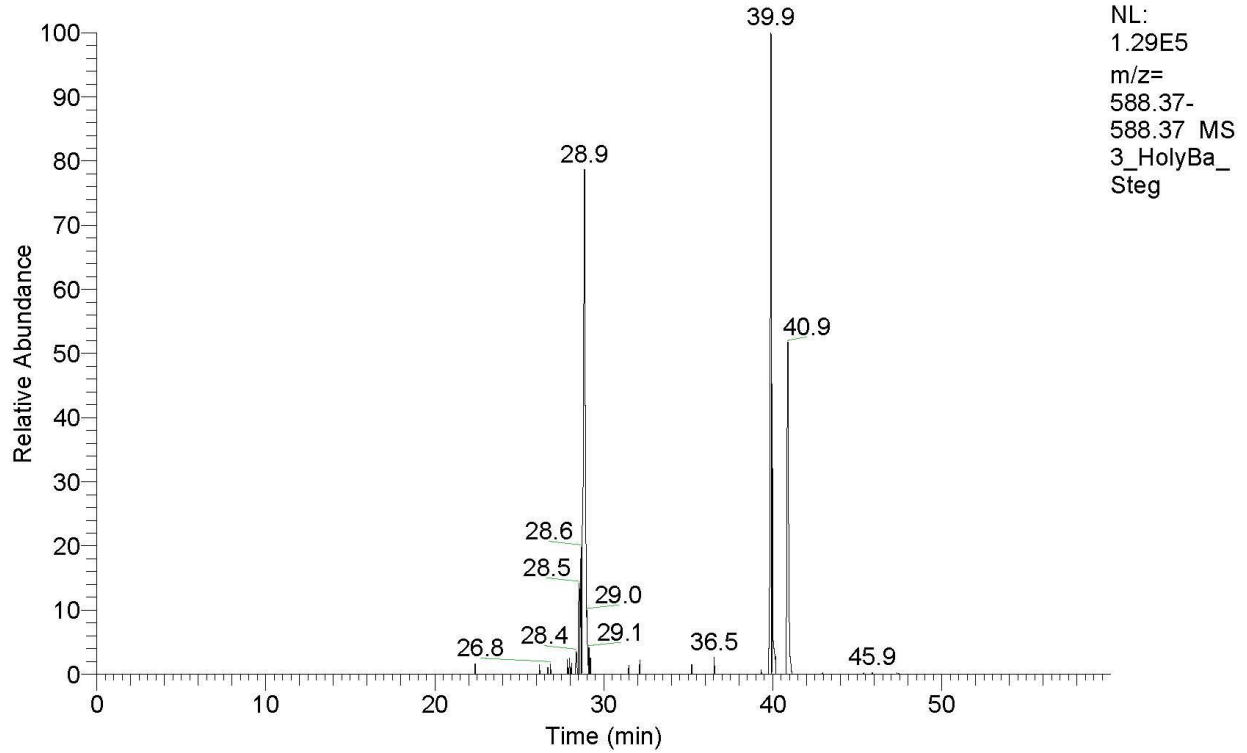
3_HolyBa_Steg #900-903   RT: 28.80-28.89   AV: 4   NL: 7.38E4
T: FTMS + p ESI Full ms [100.00-2000.00]
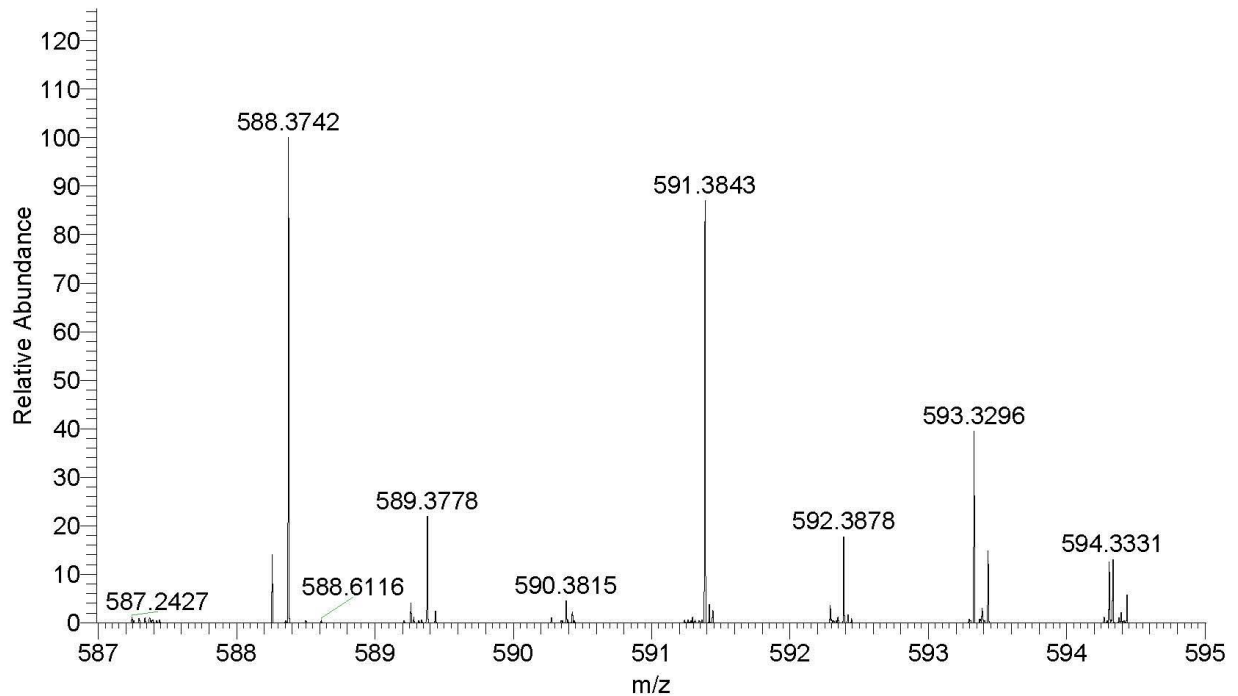
Figure S6: Holy Basil m/z 588

74
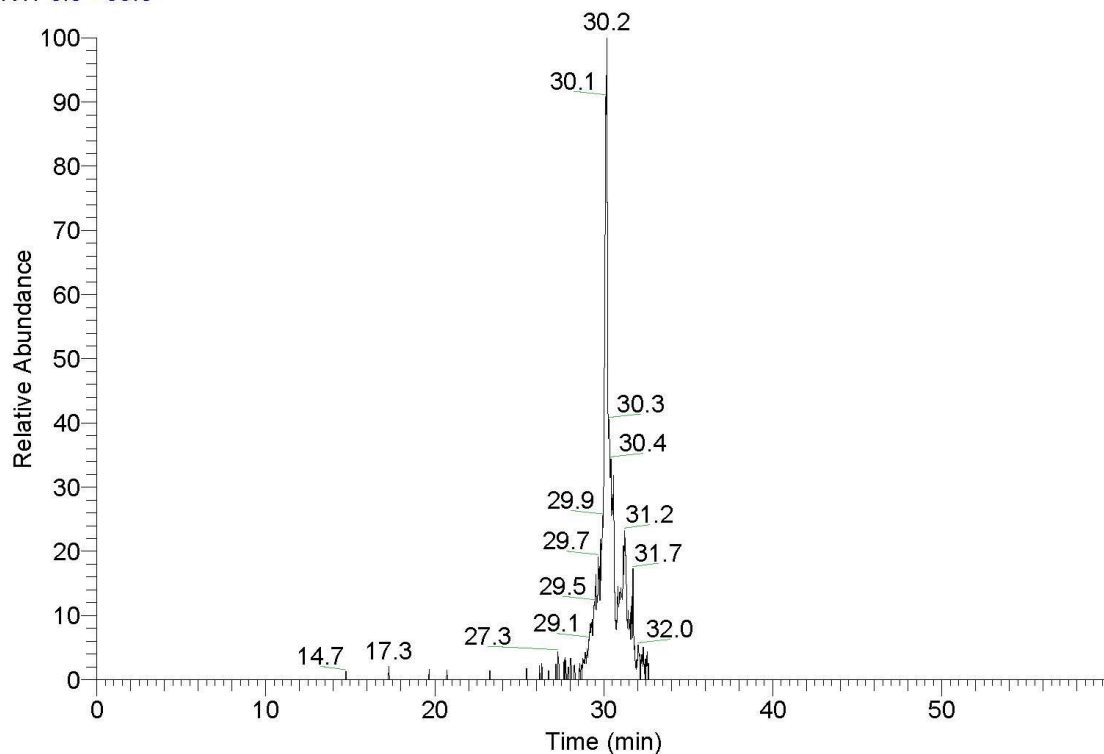
RT: 0.0 - 60.0

NL:
9.77E4
m/z=
777.32-
777.33 MS
3_HolyBa_
Steg

3_HolyBa_Steg #940-945   RT: 30.04-30.20   AV: 6   NL: 7.20E4
T: FTMS + p ESI Full ms [100.00-2000.00]
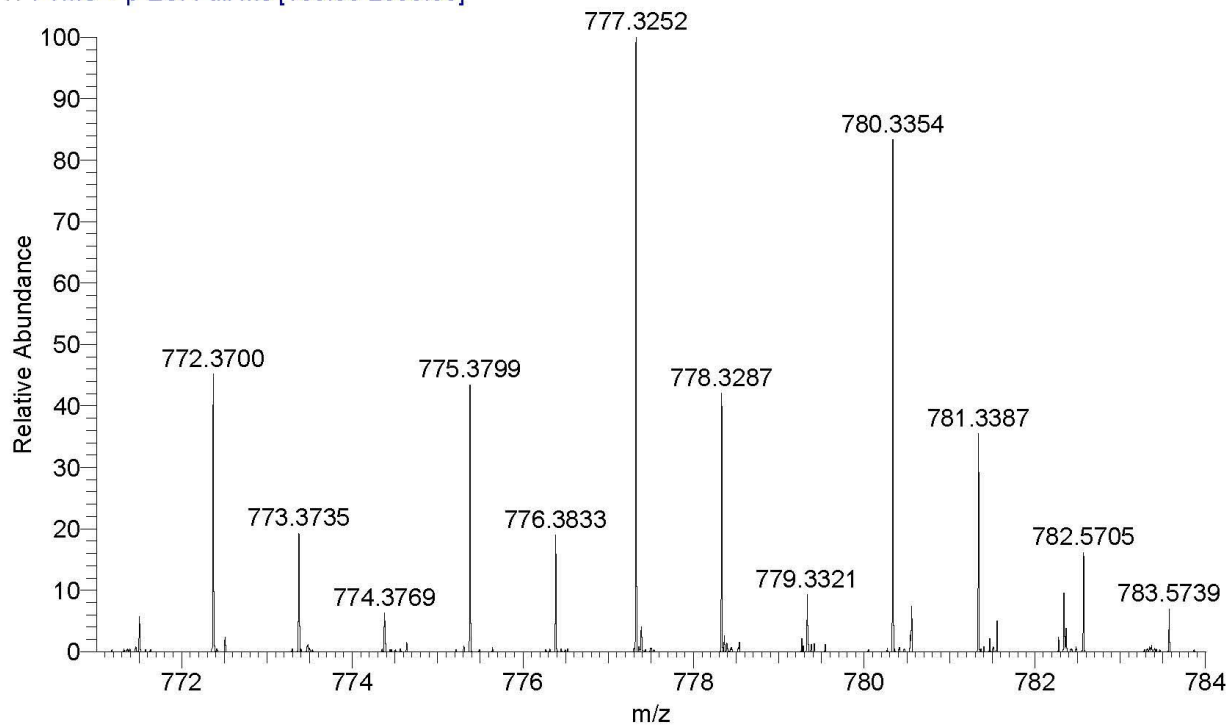
Figure S7: Holy Basil m/z 777

RT: 0.0 - 60.0

NL:
2.19E5
m/z=
747.31-
747.32 MS
3_HolyBa_
Steg
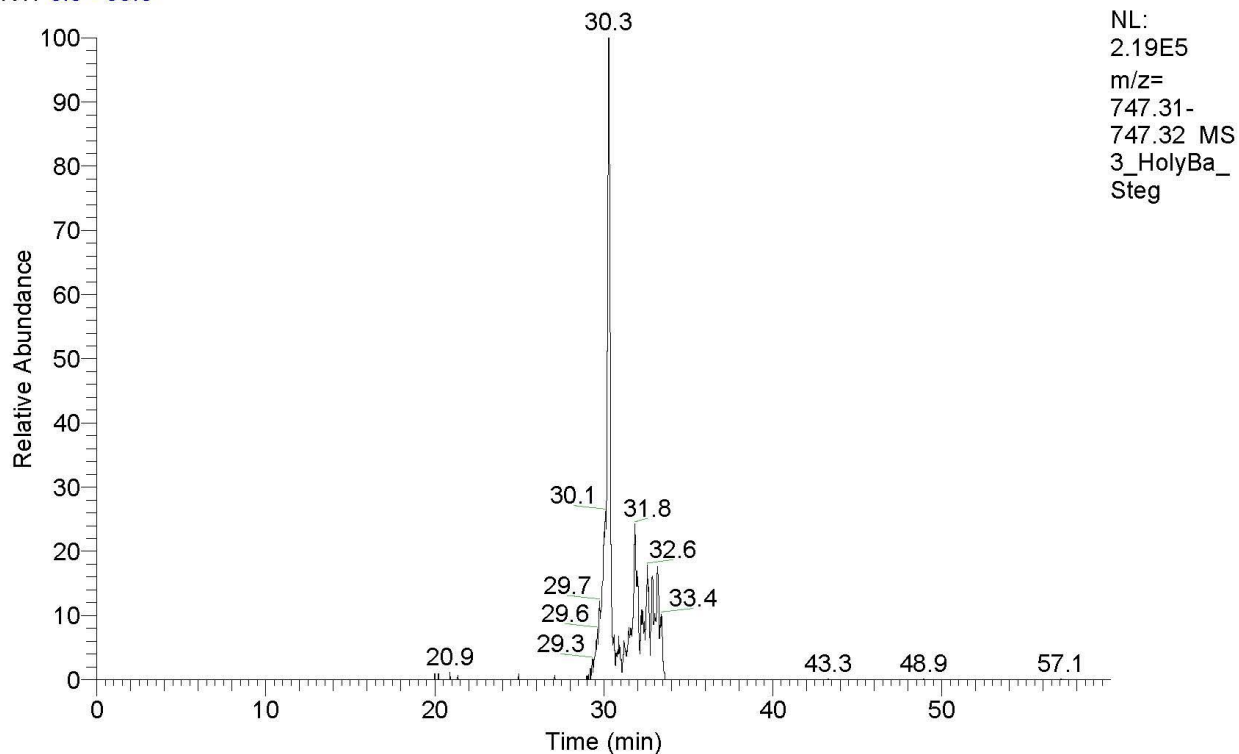
3_HolyBa_Steg #947-949   RT: 30.26-30.32   AV: 3   NL: 1.91E5
T: FTMS + p ESI Full ms [100.00-2000.00]
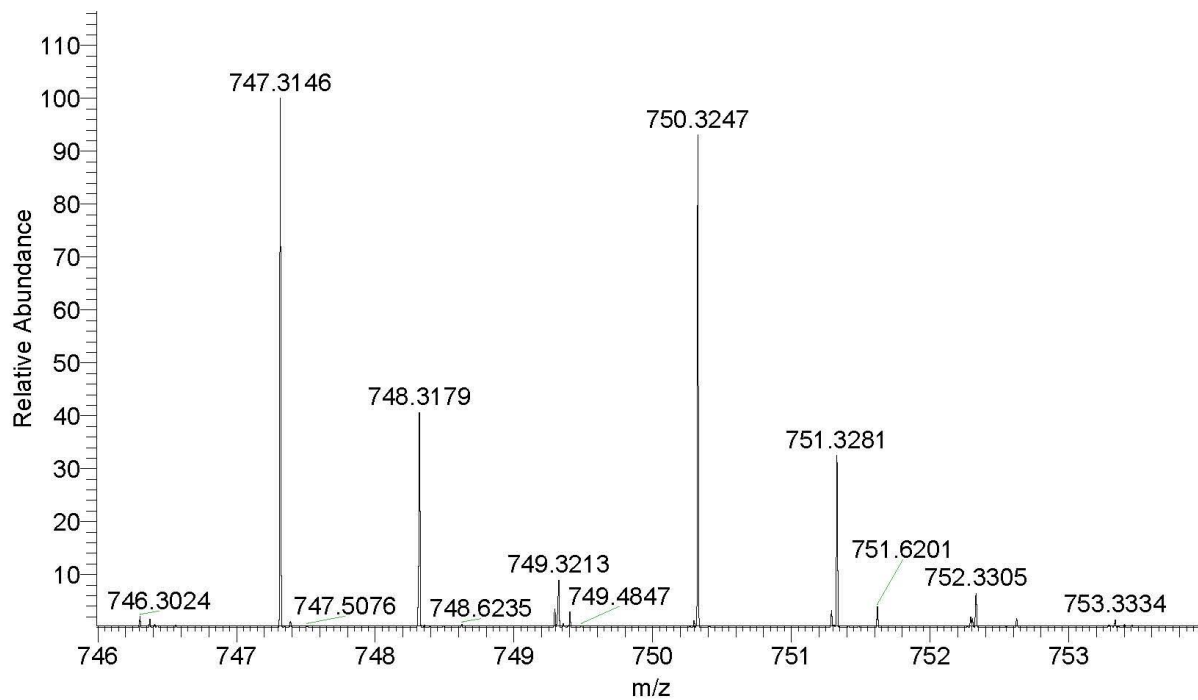
Figure S8: Holy Basil m/z 747

RT: 0.0 - 60.0

NL:
5.66E5
m/z=
545.25-
545.25 MS
3_HolyBa_
Steg
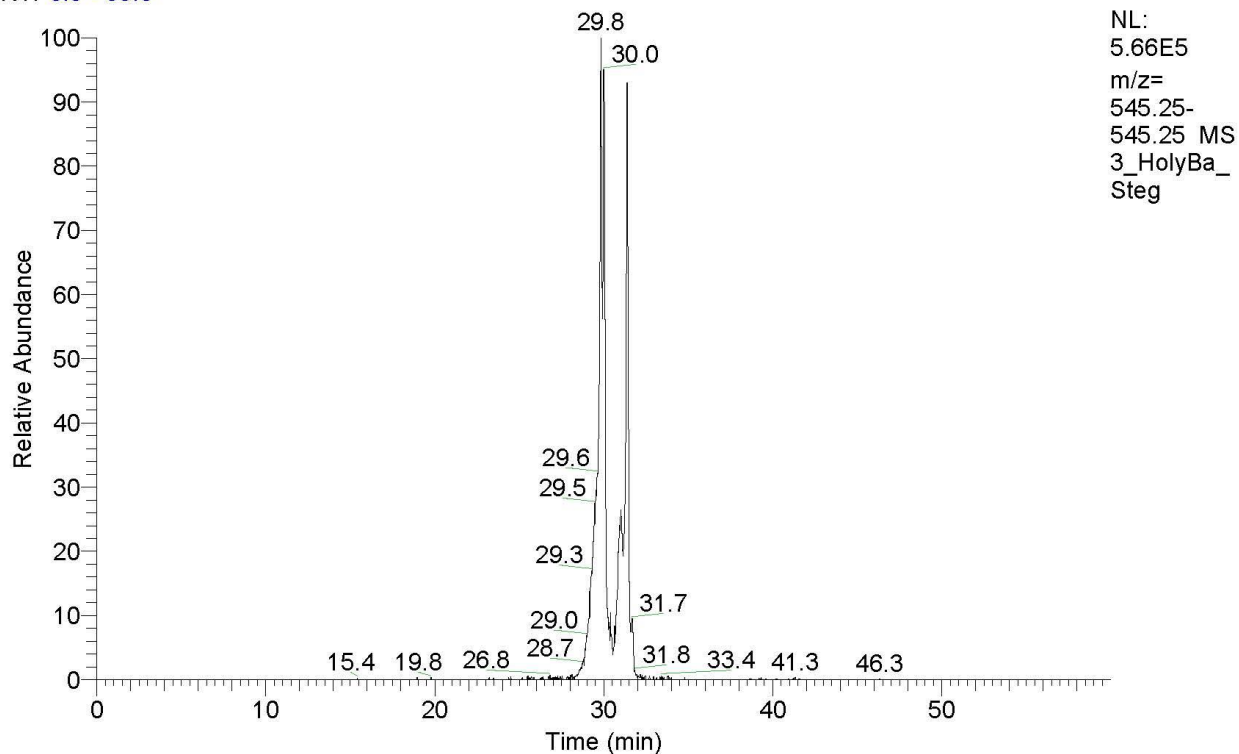
3_HolyBa_Steg #978-985   RT: 31.22-31.44   AV: 8   NL: 2.98E5
T: FTMS + p ESI Full ms [100.00-2000.00]
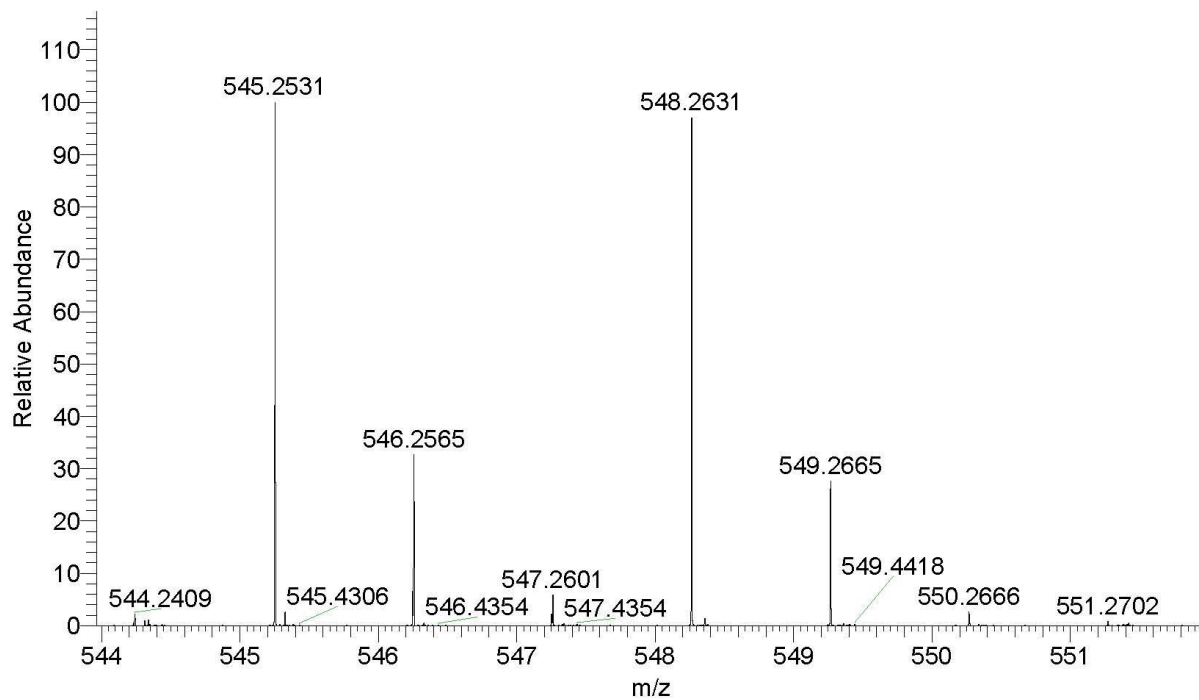
Figure S9: Holy Basil m/z 545

RT: 0.0 - 60.0

NL:
1.09E5
m/z=
742.36-
742.36 MS
3_HolyBa_
Steg
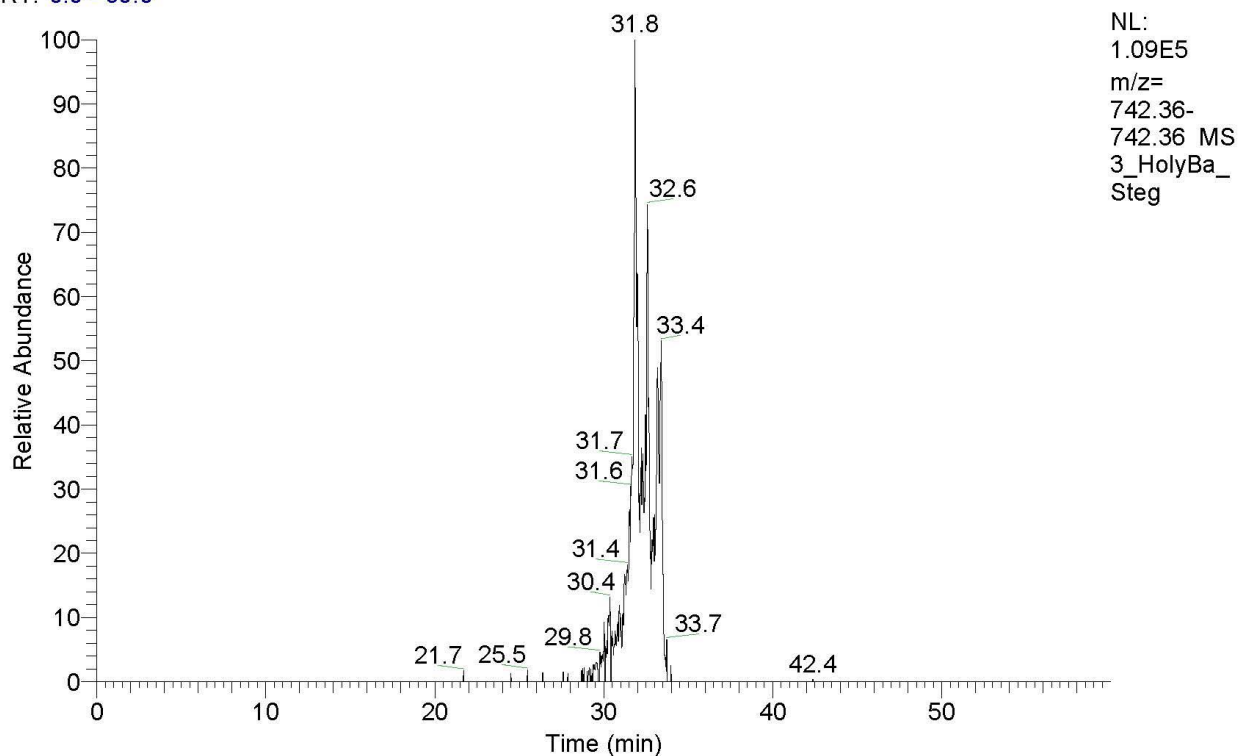
3_HolyBa_Steg #997-999   RT: 31.81-31.87   AV: 3   NL: 8.58E4
T: FTMS + p ESI Full ms [100.00-2000.00]
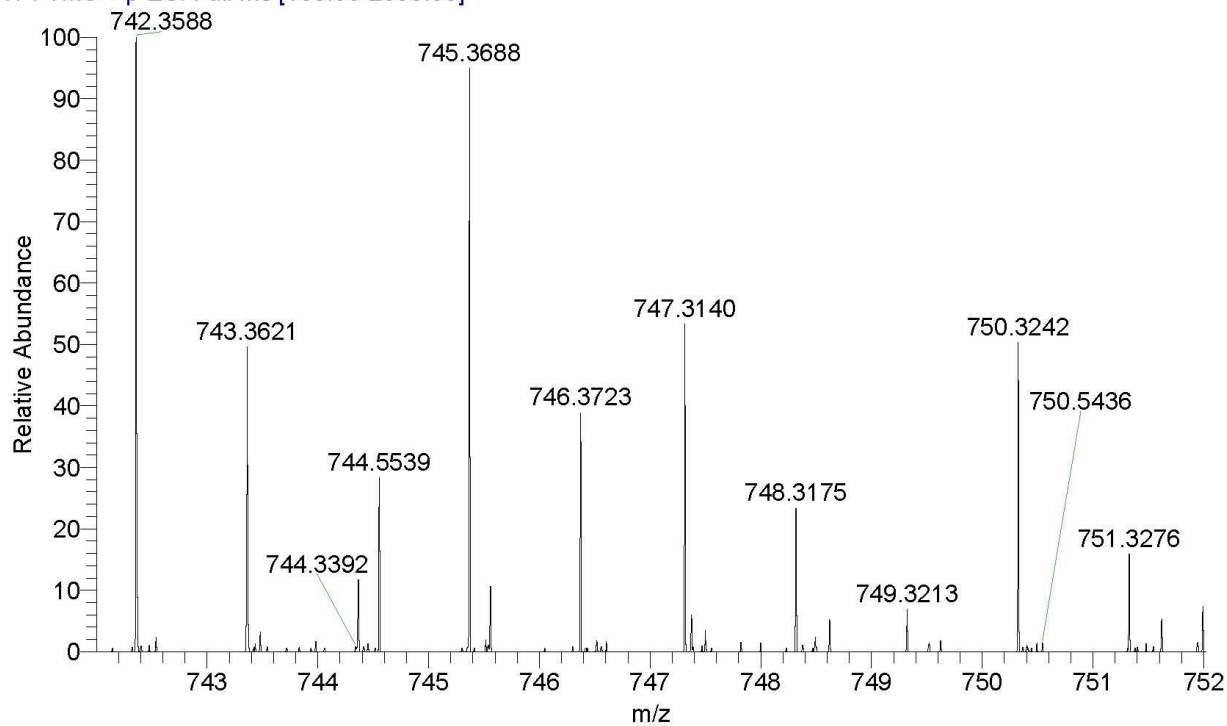
Figure S10: Holy Basil m/z 742
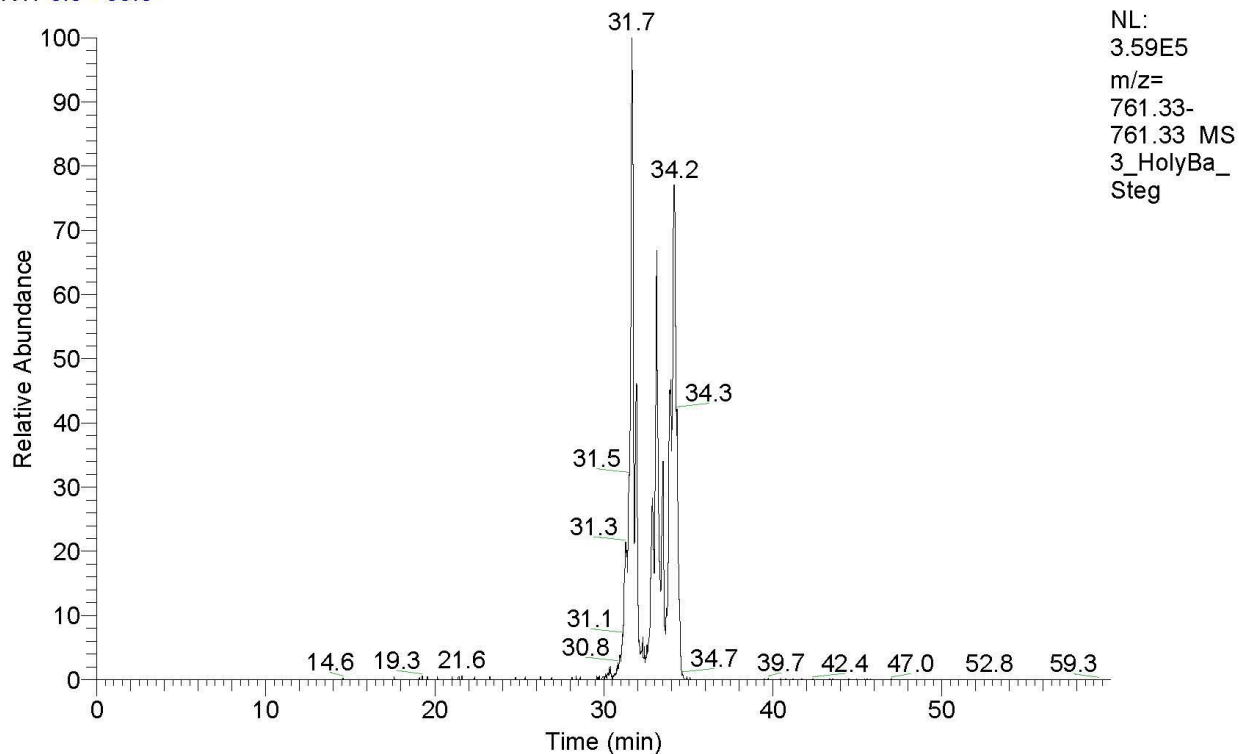
RT: 0.0 - 60.0
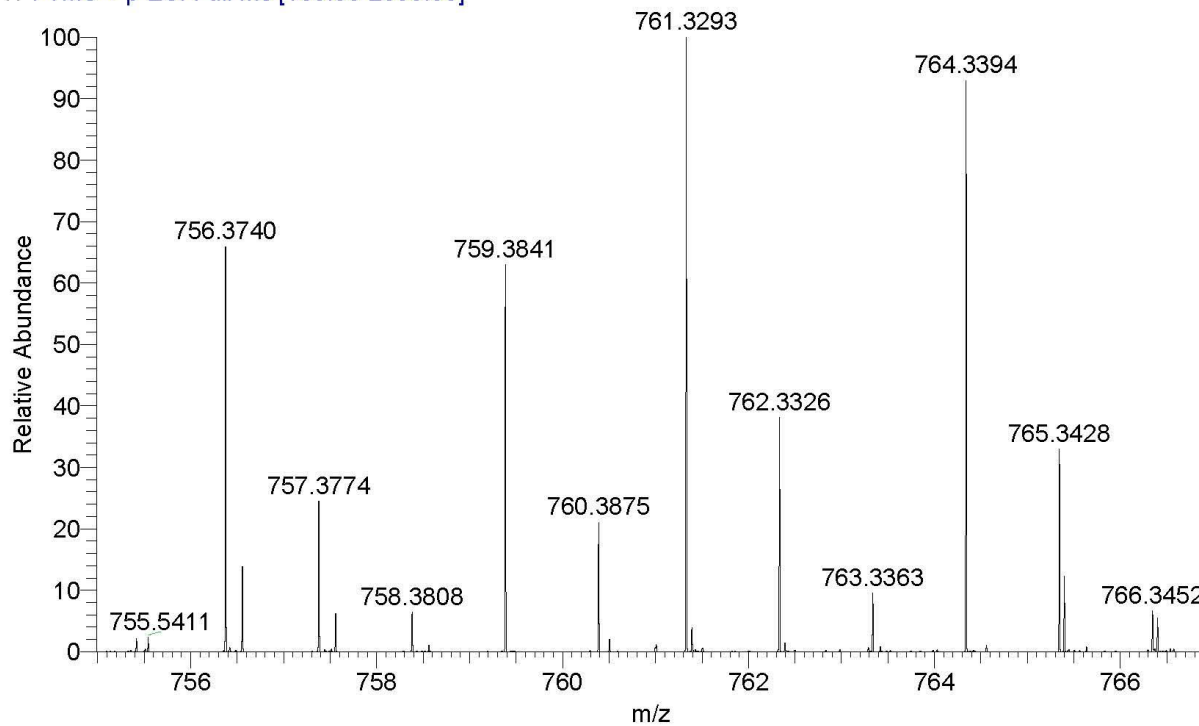
Figure S11: Holy Basil m/z 761

RT: 0.0 - 60.0

NL:
7.03E4
m/z=
655.50-
655.51 MS
3_HolyBa_
Steg
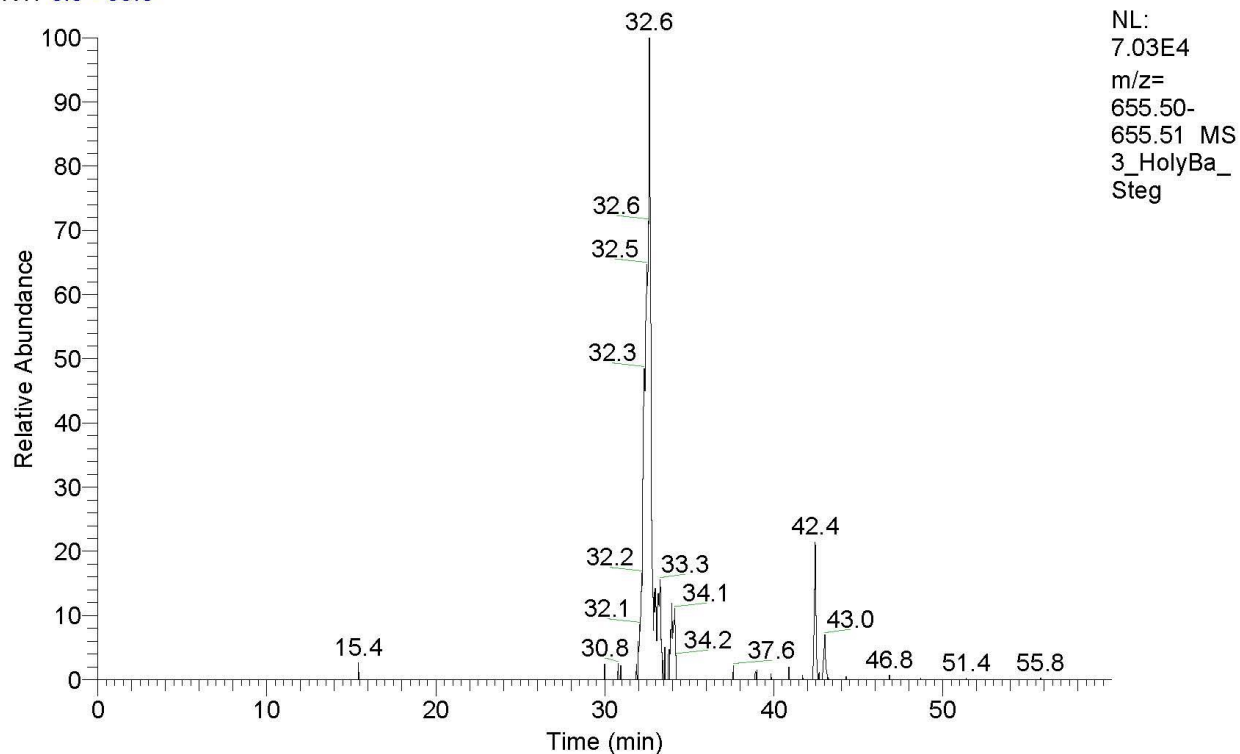
3_HolyBa_Steg #1023-1025   RT: 32.61-32.67   AV: 3   NL: 5.58E4
T: FTMS + p ESI Full ms [100.00-2000.00]
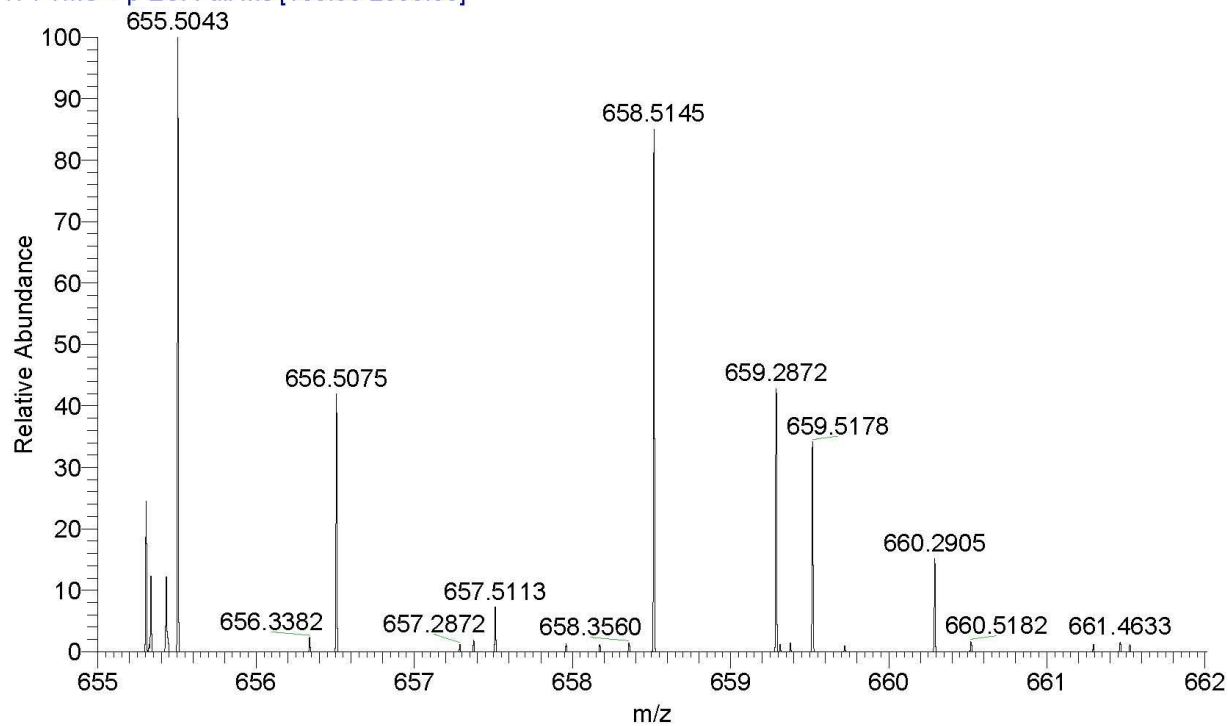
Figure S12: Holy Basil m/z 655
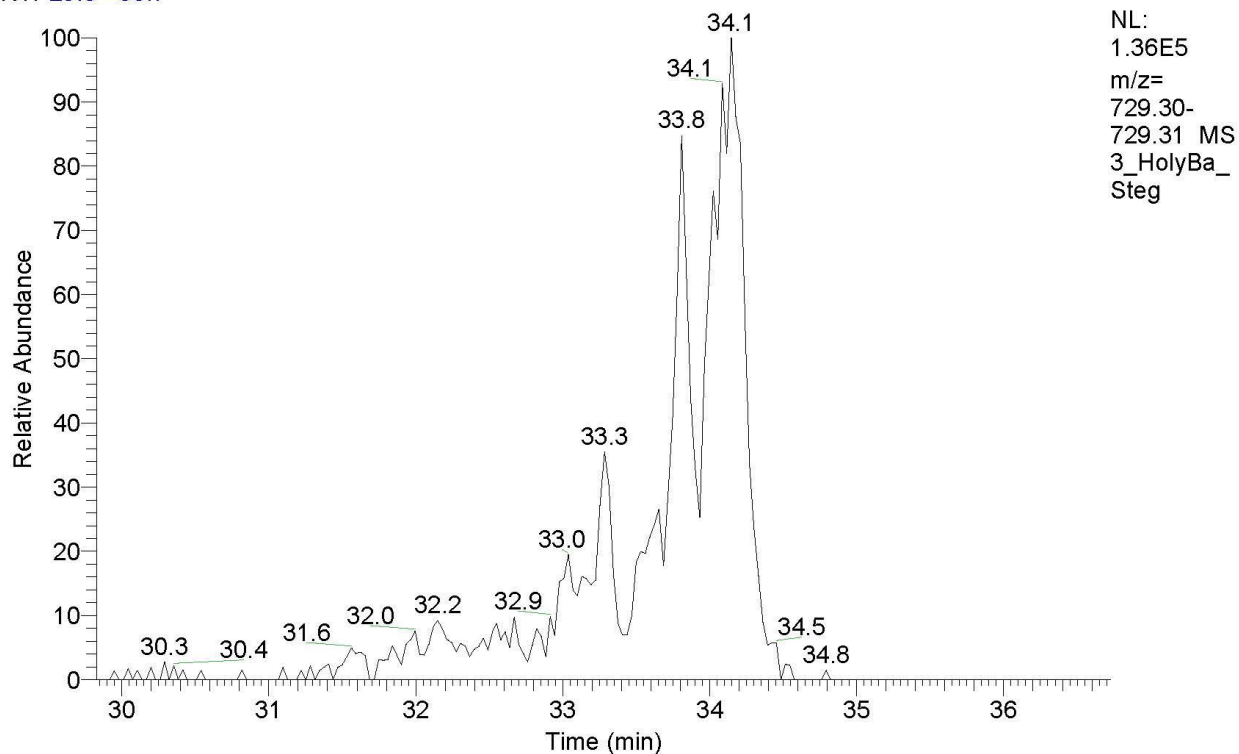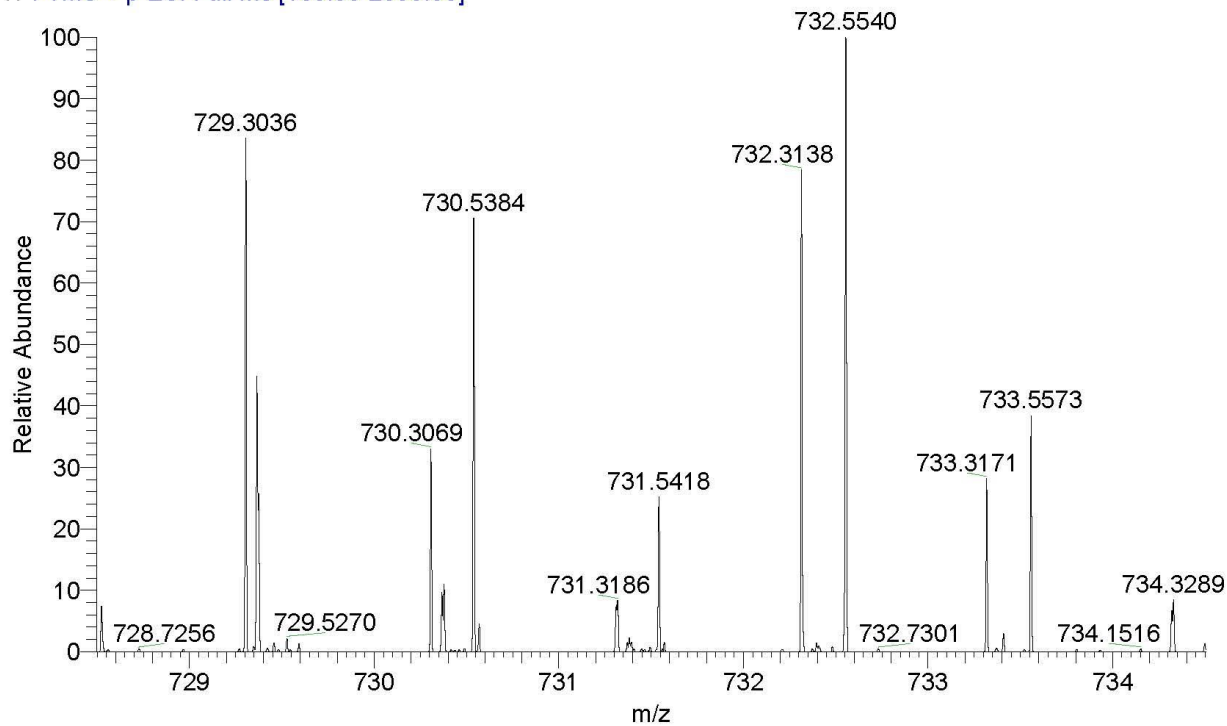
Figure S13: Holy Basil m/z 729

RT: 0.0 - 60.0

NL:
3.49E5
m/z=
719.37-
719.38 MS
3_Ashwa_S
teg
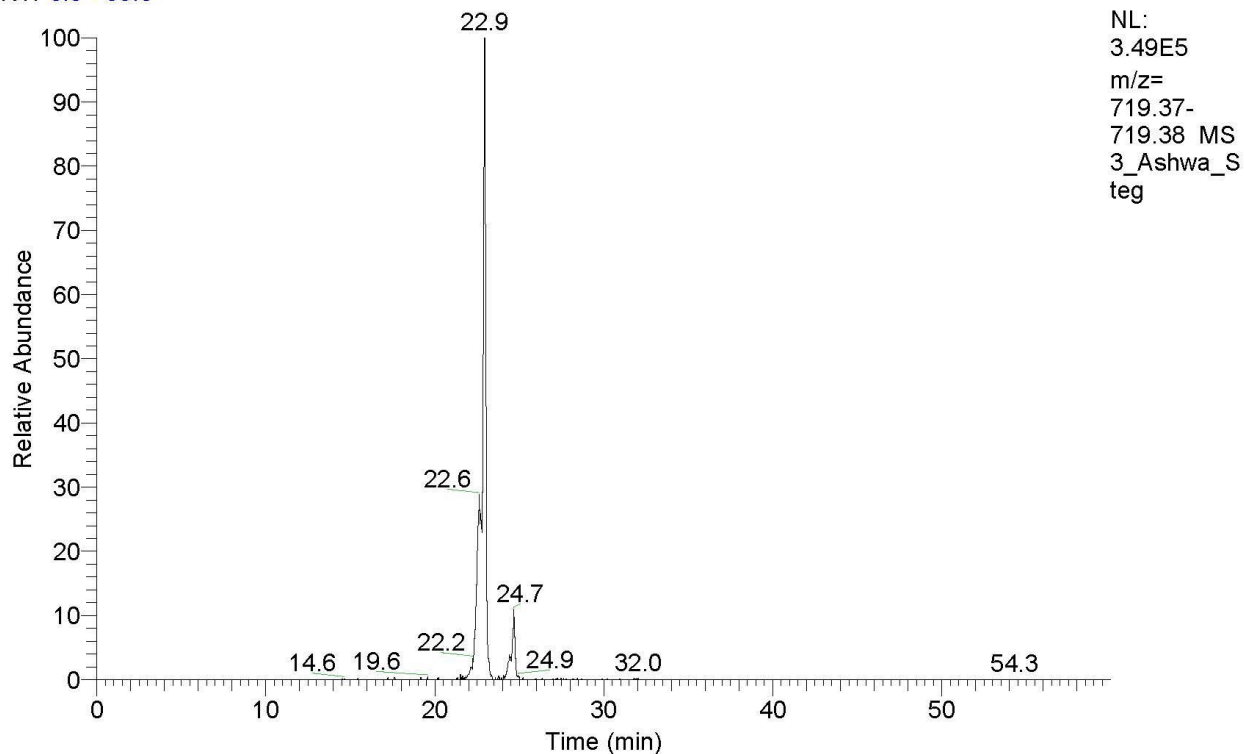
3_Ashwa_Steg #666   RT: 22.98   AV: 1   NL: 2.80E5
T: FTMS + p ESI Full ms [100.00-2000.00]
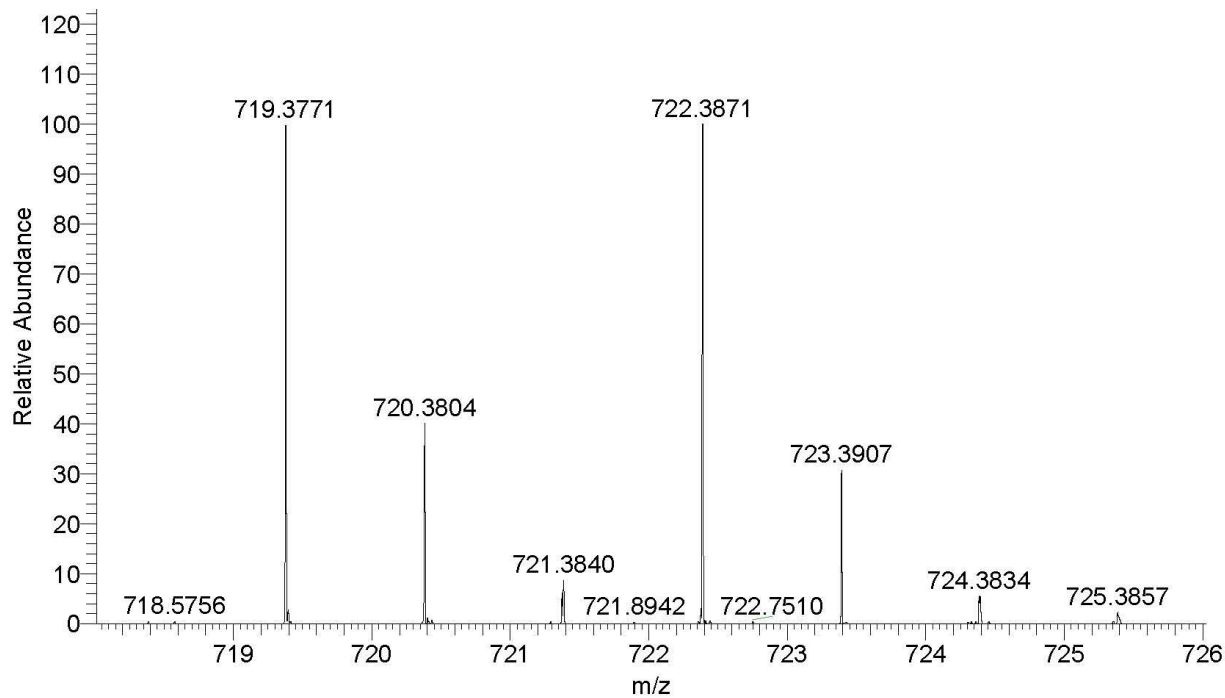
Figure S14: Ashwagandha m/z 719

RT: 0.0 - 60.0

NL:
1.17E5
m/z=
1047.55-
1047.55
MS
3_Ashwa_S
teg
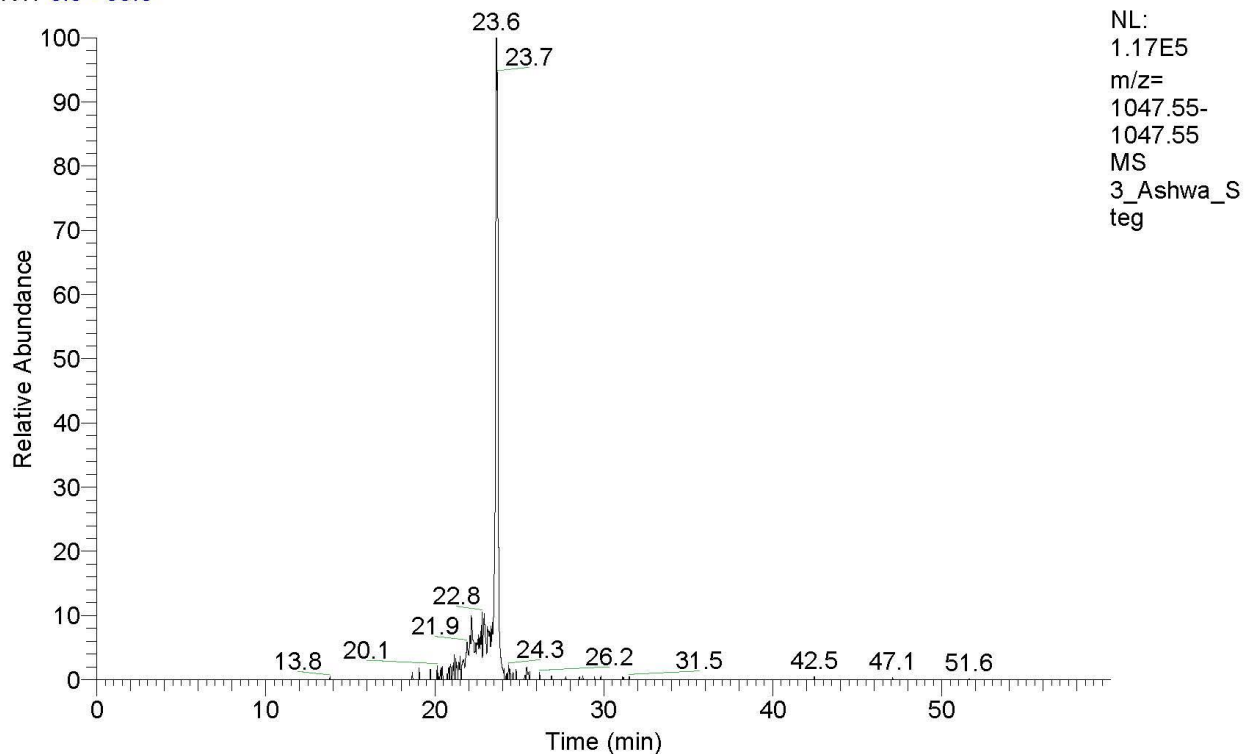
3_Ashwa_Steg #685-690   RT: 23.58-23.74   AV: 6   NL: 8.88E4
T: FTMS + p ESI Full ms [100.00-2000.00]
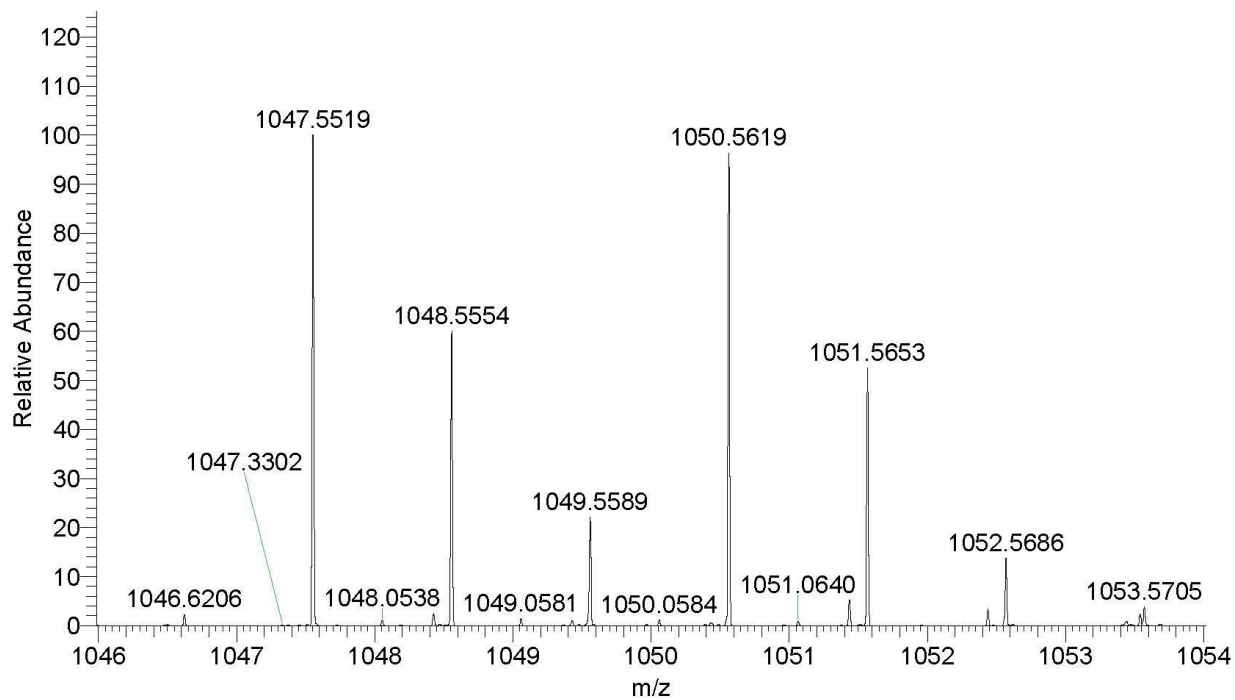
Figure S15: Ashwagandha m/z 1047

RT: 0.0 - 60.0

NL:
2.62E5
m/z=
1031.55-
1031.56
MS
3_Ashwa_S
teg
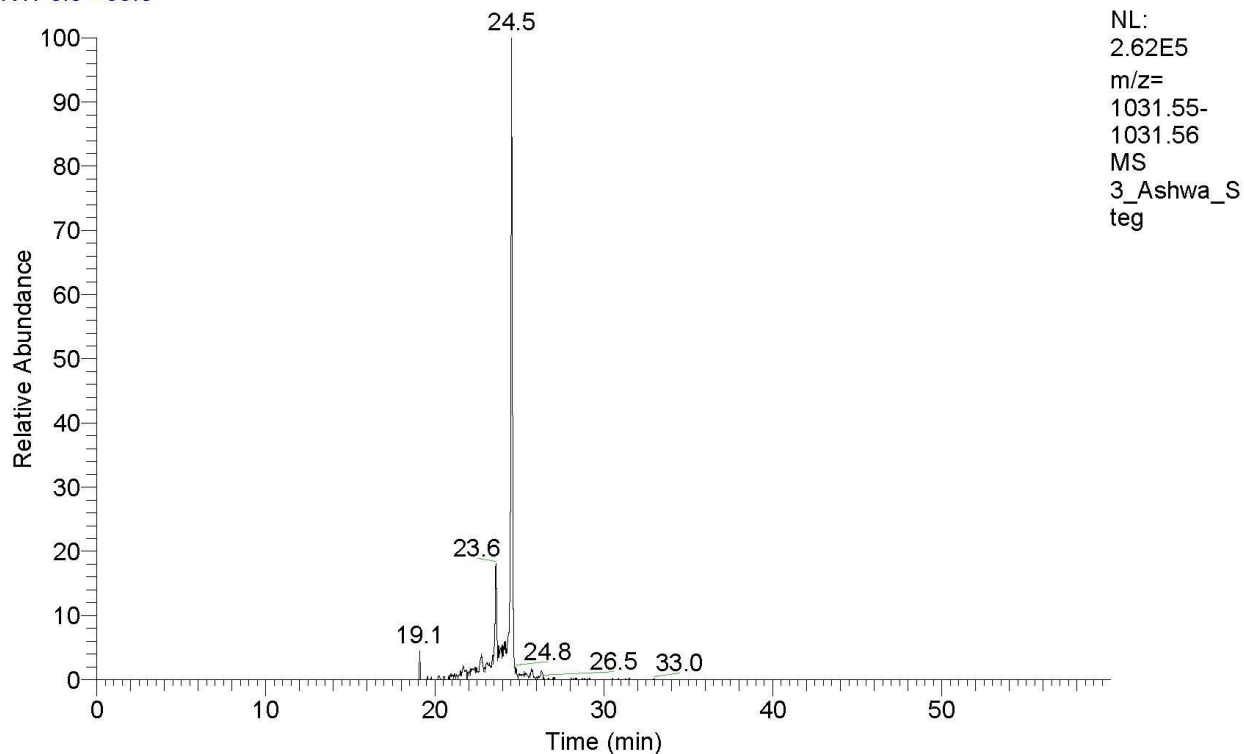
3_Ashwa_Steg #713-716　RT: 24.48-24.57　AV: 4　NL: 1.70E5
T: FTMS + p ESI Full ms [100.00-2000.00]
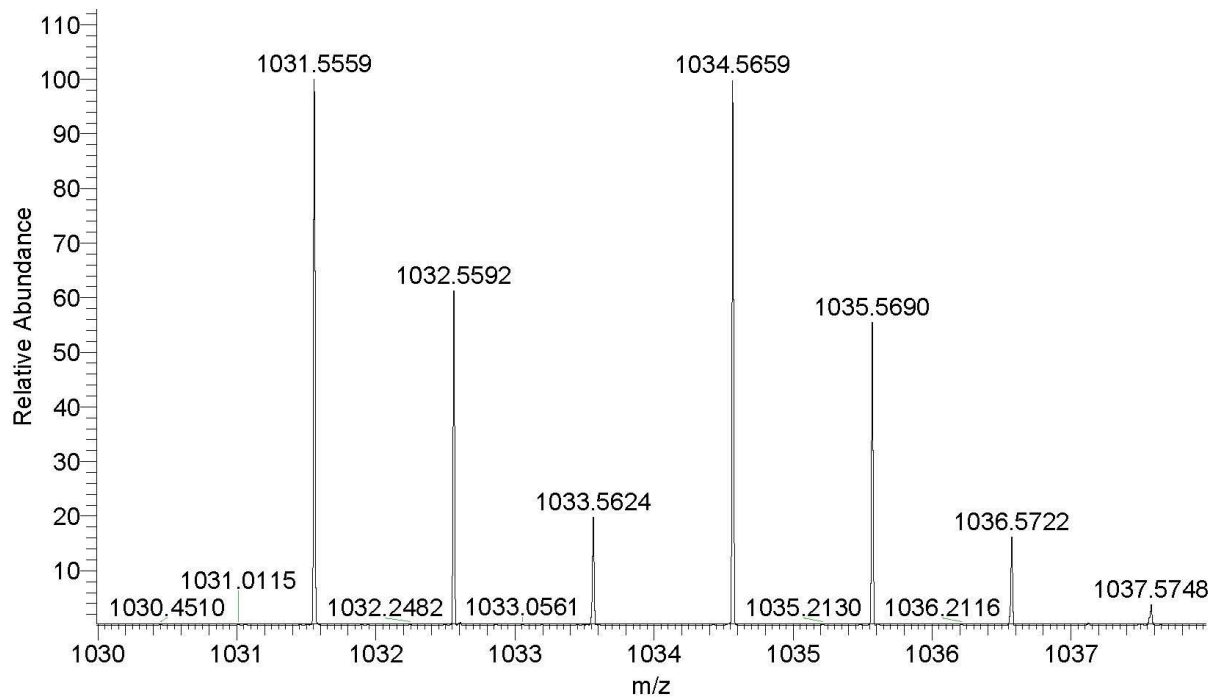
Figure S16: Ashwagandha m/z 1031
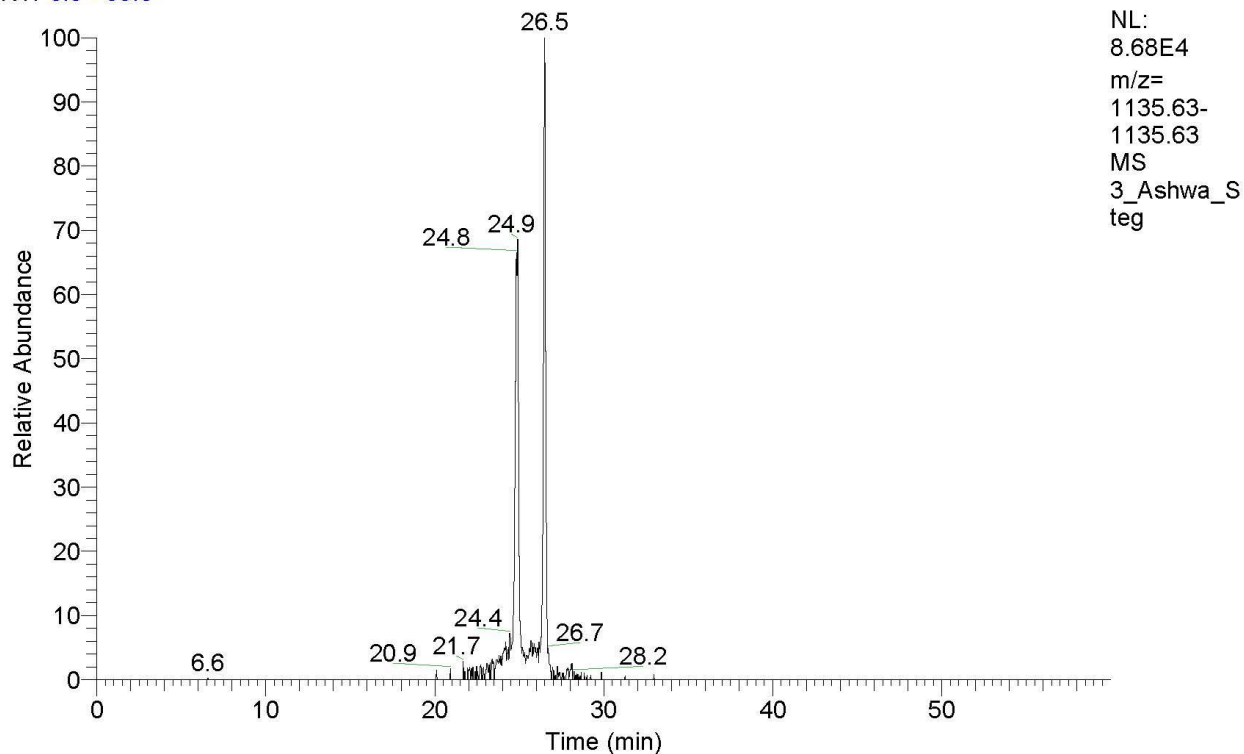
RT: 0.0 - 60.0
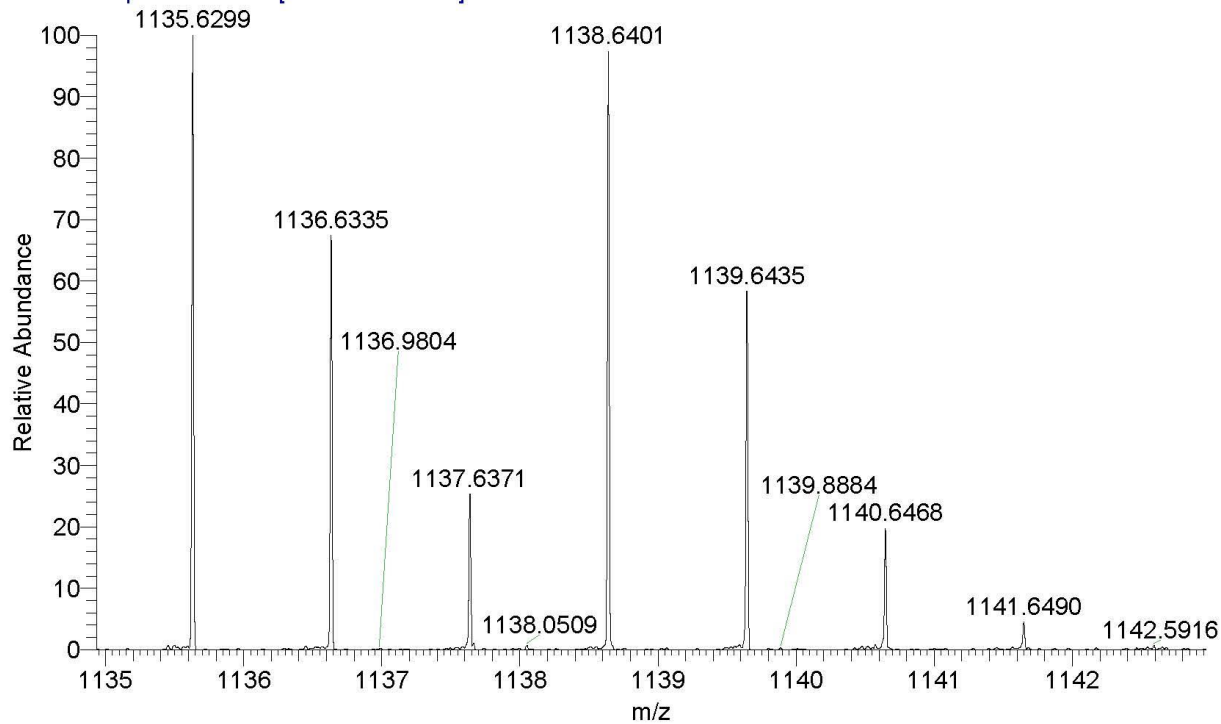
Figure S17: Ashwagandha m/z 1135

RT: 0.0 - 60.0

NL:
1.88E5
m/z=
567.23-
567.24 MS
3_HolyBa_
Steg
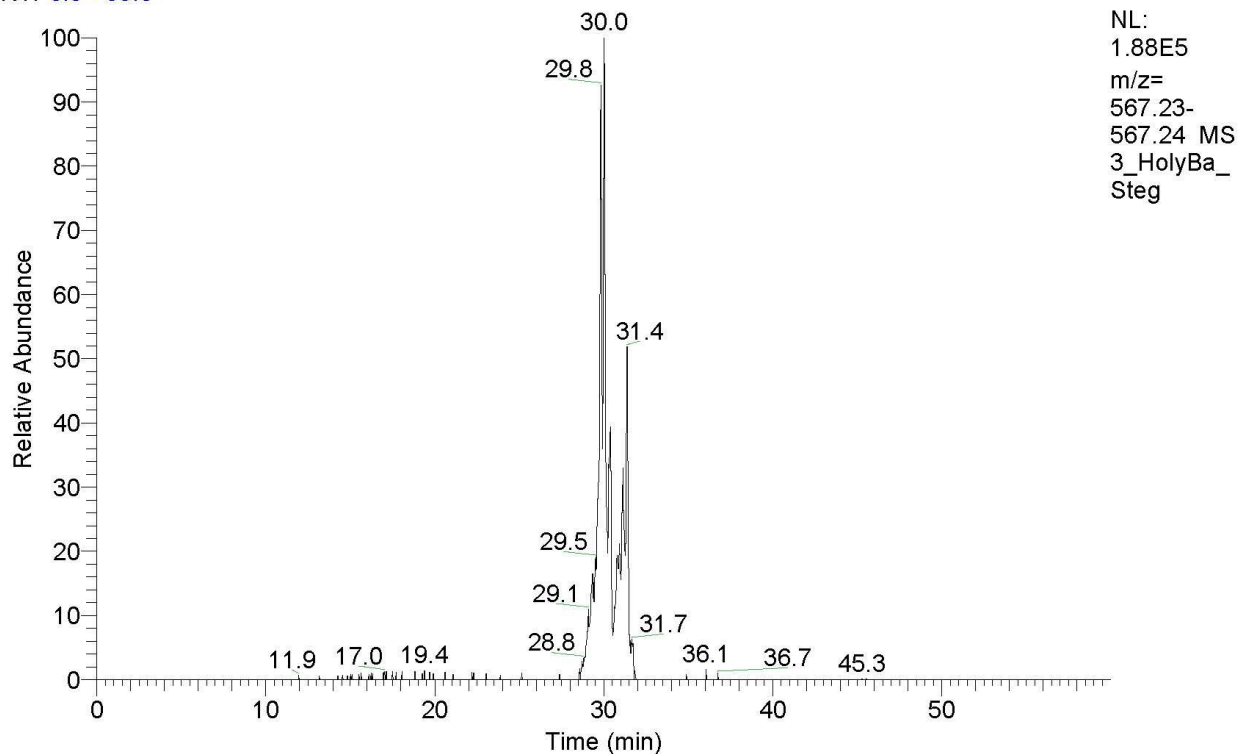
3_HolyBa_Steg #931-935   RT: 29.76-29.89   AV: 5   NL: 1.32E5
T: FTMS + p ESI Full ms [100.00-2000.00]
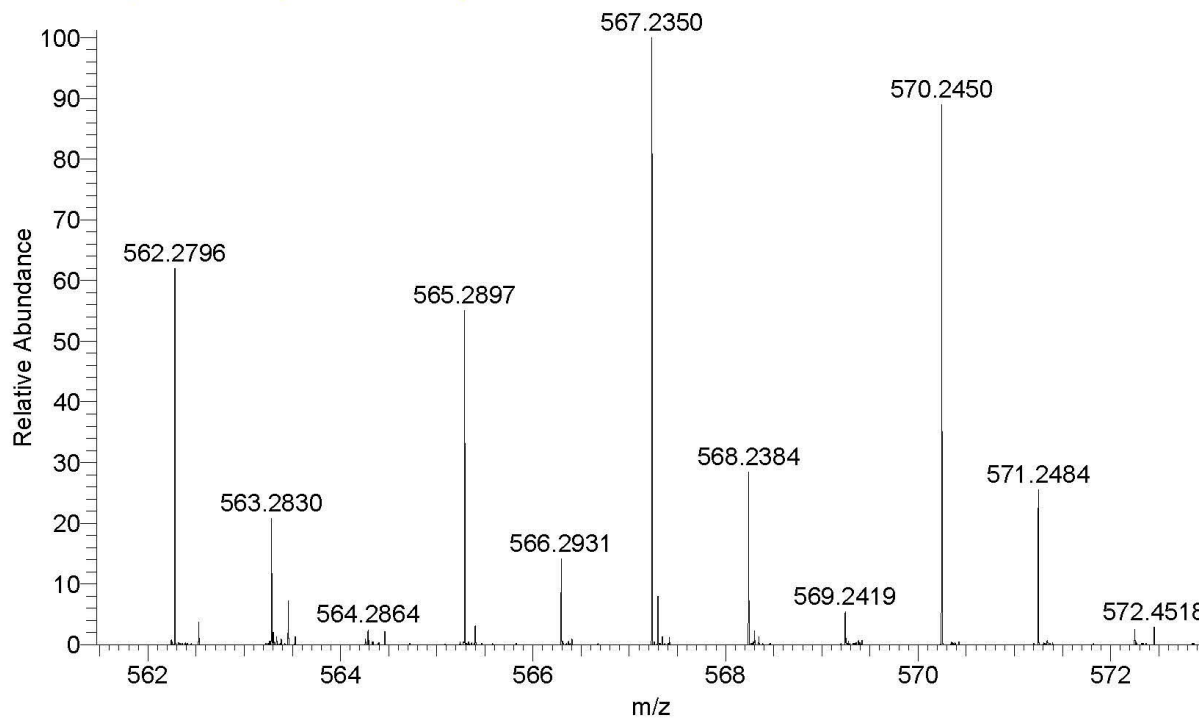
Figure S18: Ashwagandha m/z 567
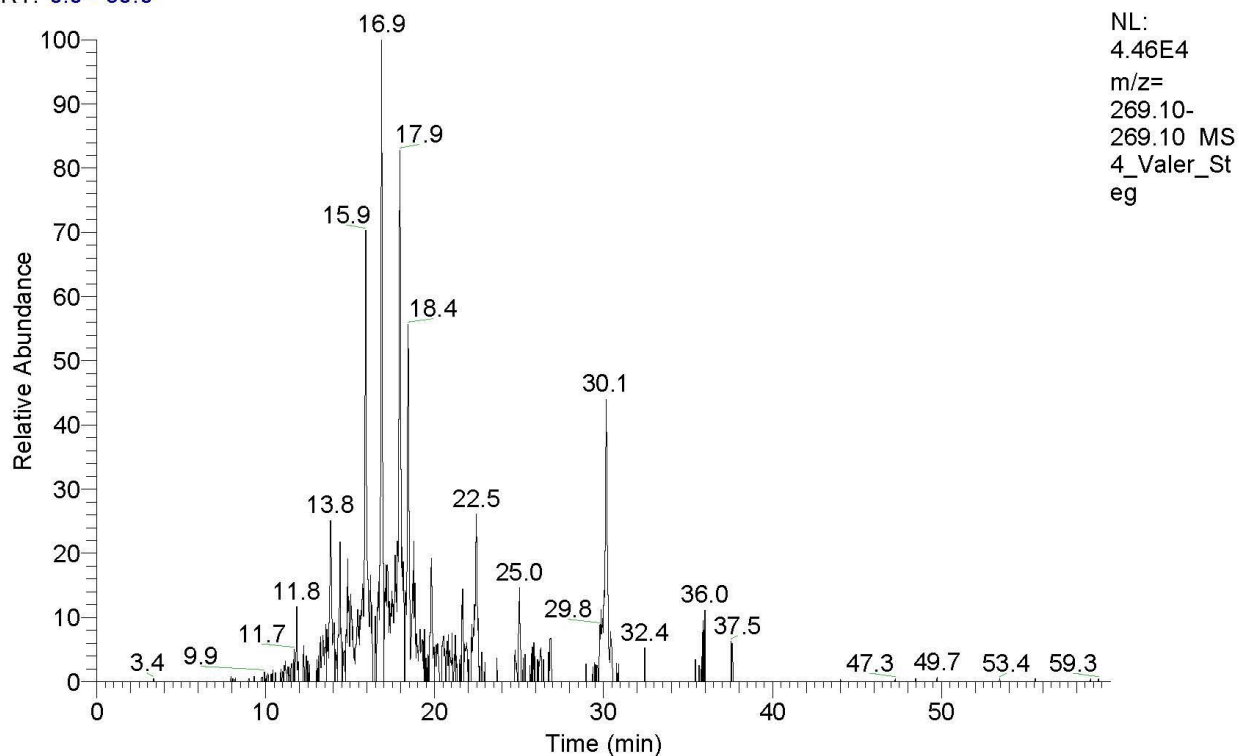
Figure S19: Valerian m/z 269

RT: 0.0 - 60.0
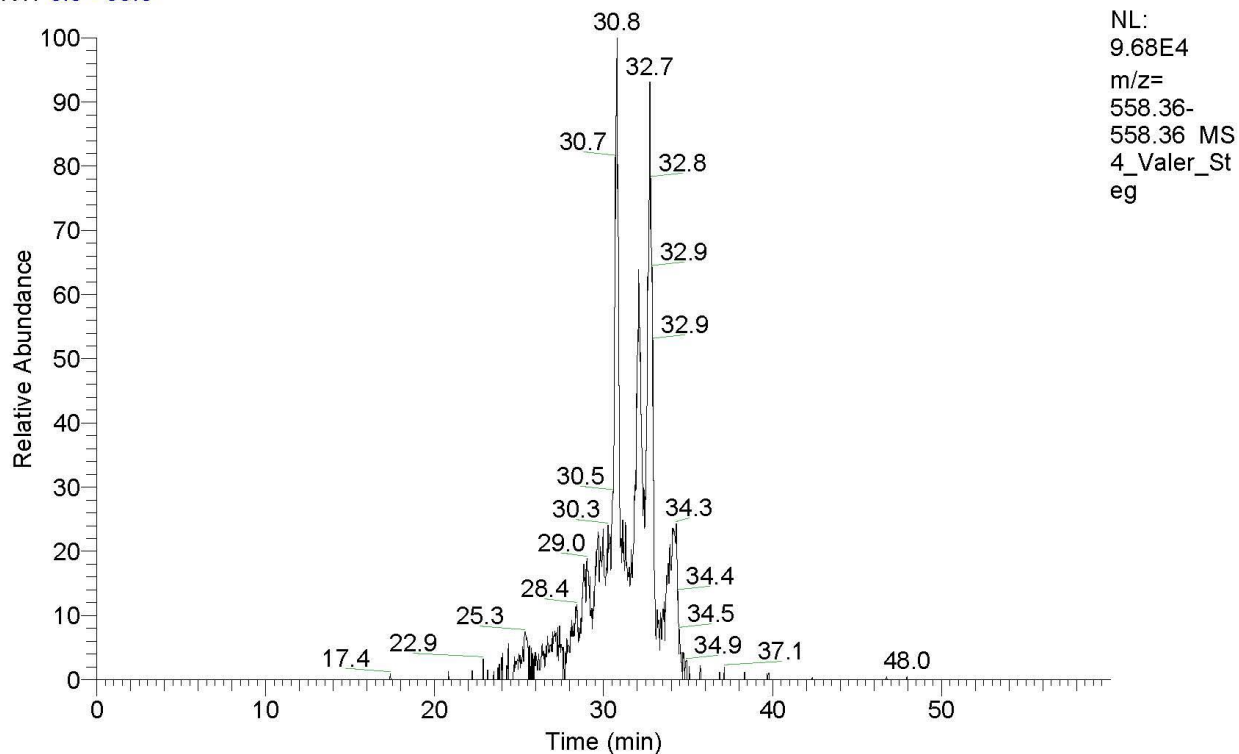
4_Valer_Steg #957-967   RT: 30.61-30.91   AV: 11   NL: 9.76E4
T: FTMS + p ESI Full ms [100.00-2000.00]
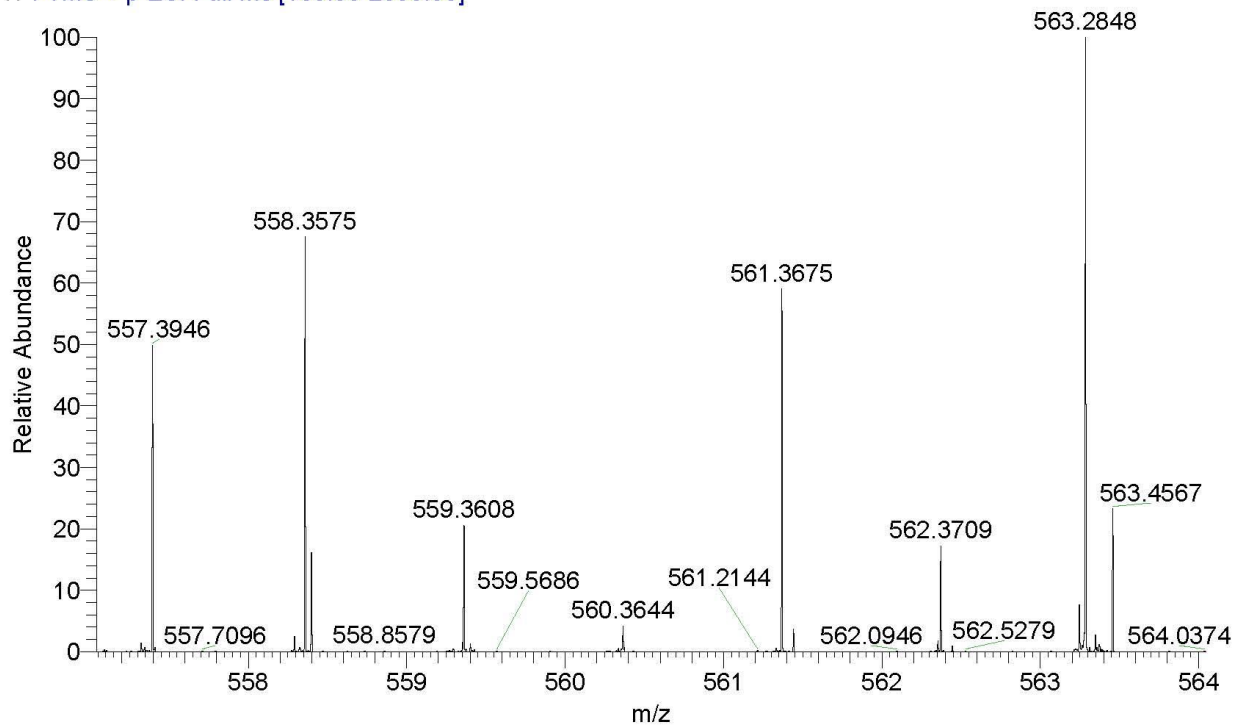
Figure S20: Valerian m/z 558
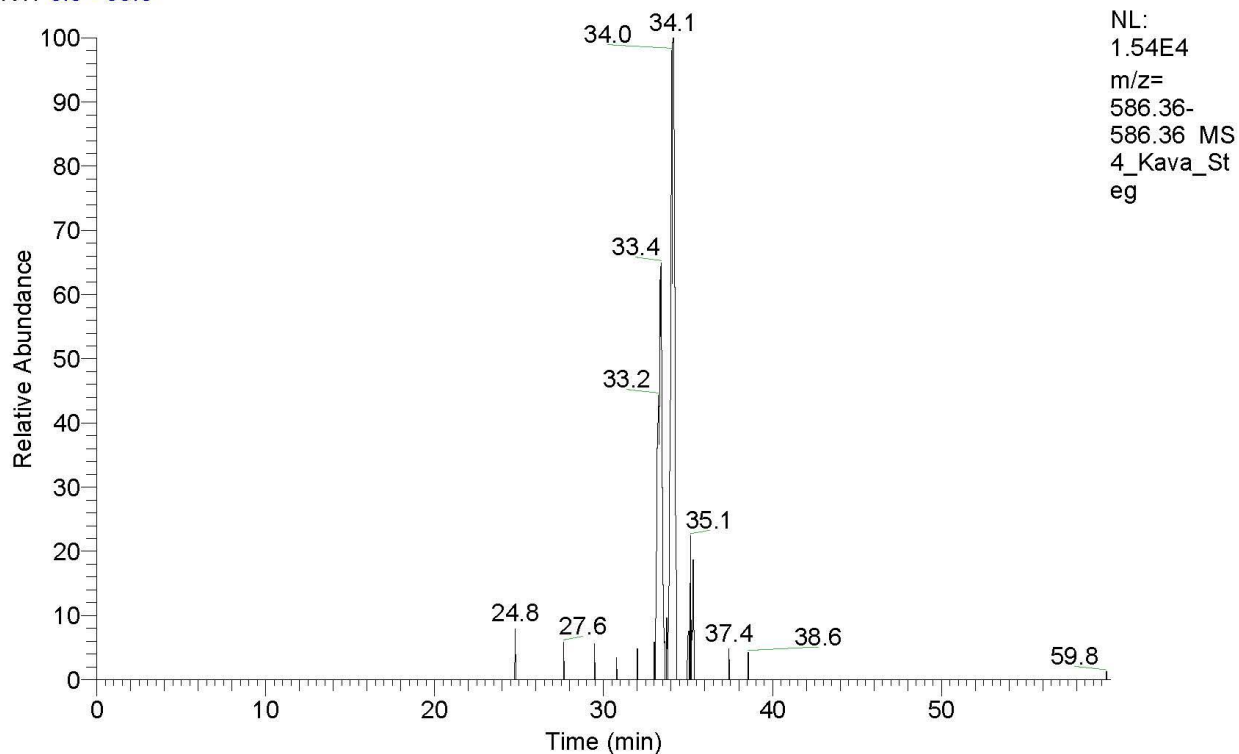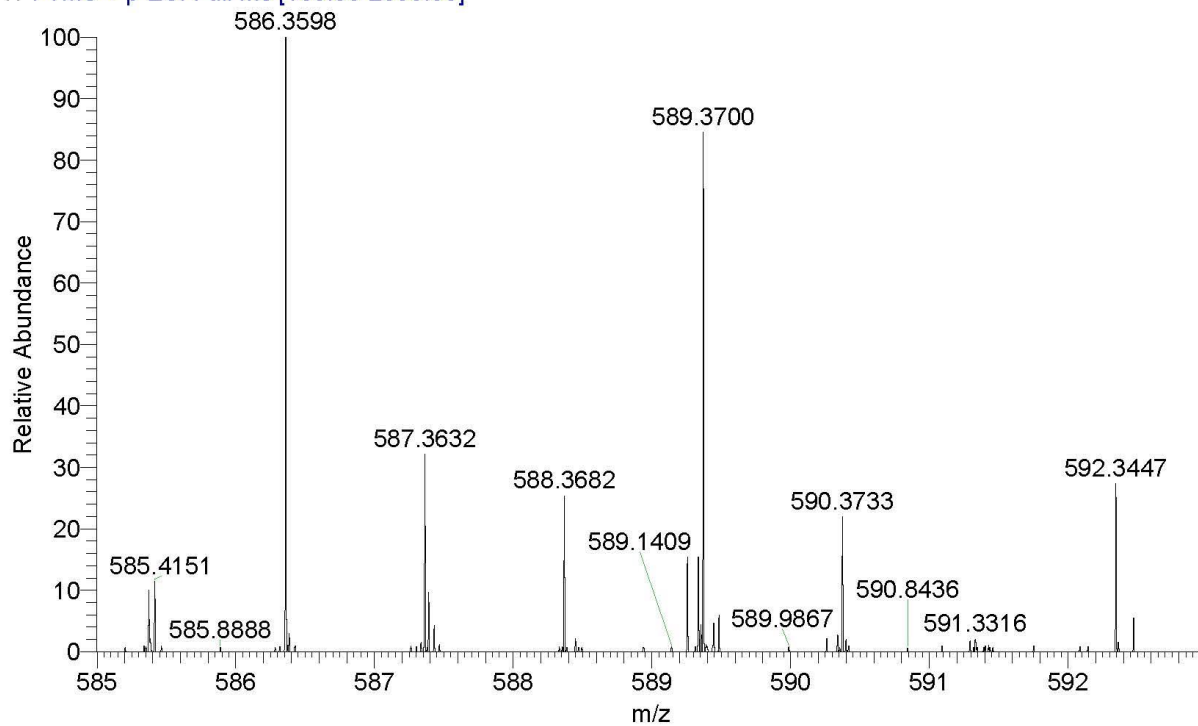
Figure S21: Valerian m/z 586

RT: 0.0 - 60.0
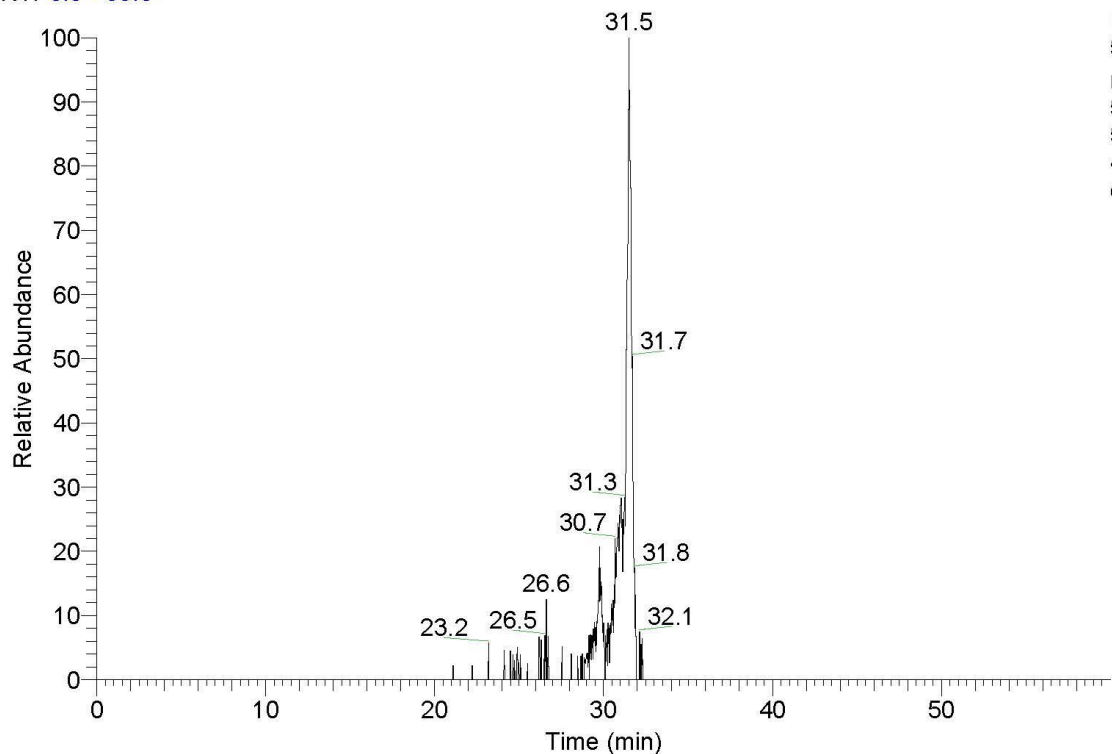
NL:
5.05E4
m/z=
591.34-
591.34 MS
4_Valer_St
eg

4_Valer_Steg #991-994   RT: 31.66-31.75   AV: 4   NL: 1.98E4
T: FTMS + p ESI Full ms [100.00-2000.00]
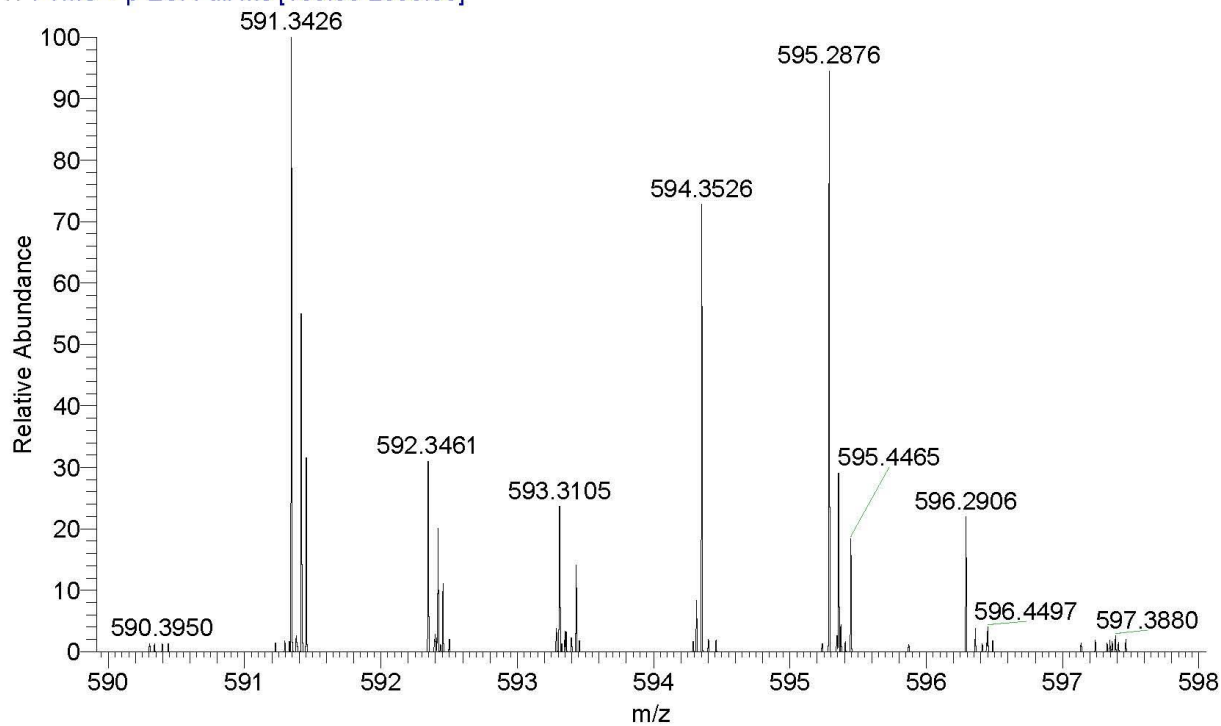
Figure S22: Valerian m/z 591

RT: 0.0 - 60.0

NL:
4.80E4
m/z=
342.30-
342.30 MS
4_Kava_St
eg
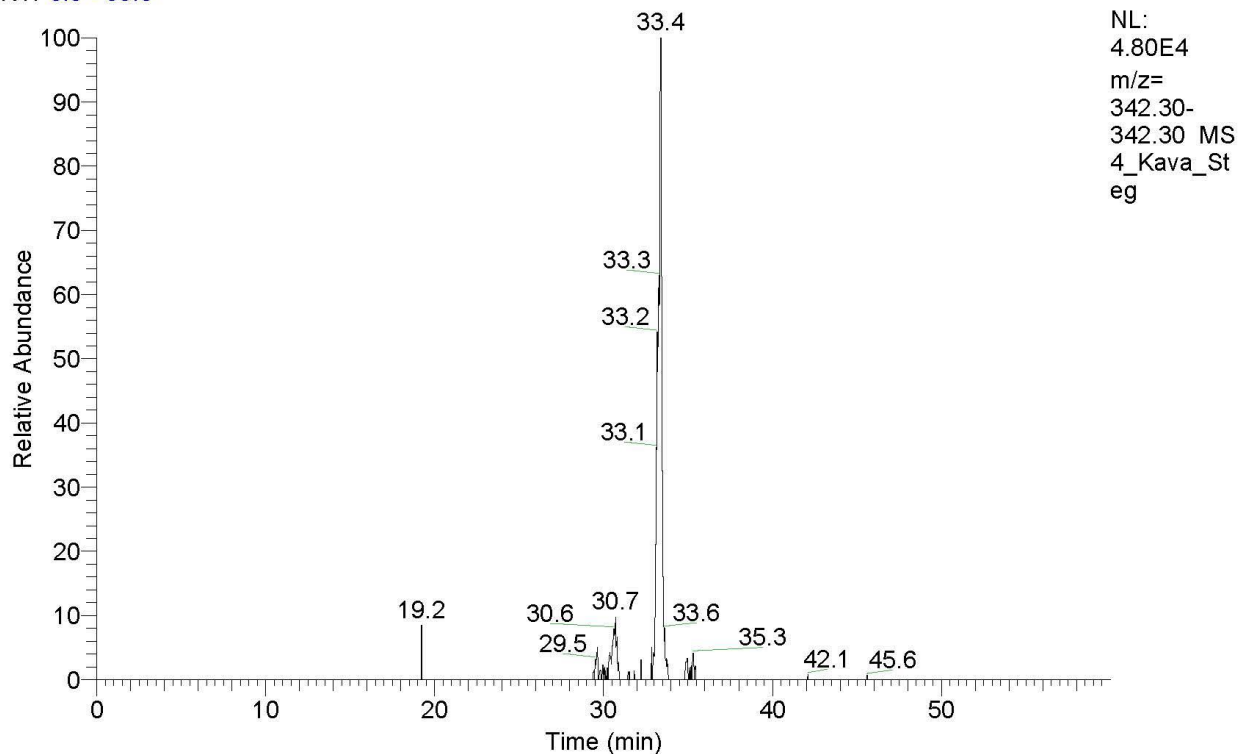
4_Kava_Steg #1018-1021   RT: 33.33-33.42   AV: 4   NL: 4.09E4
T: FTMS + p ESI Full ms [100.00-2000.00]
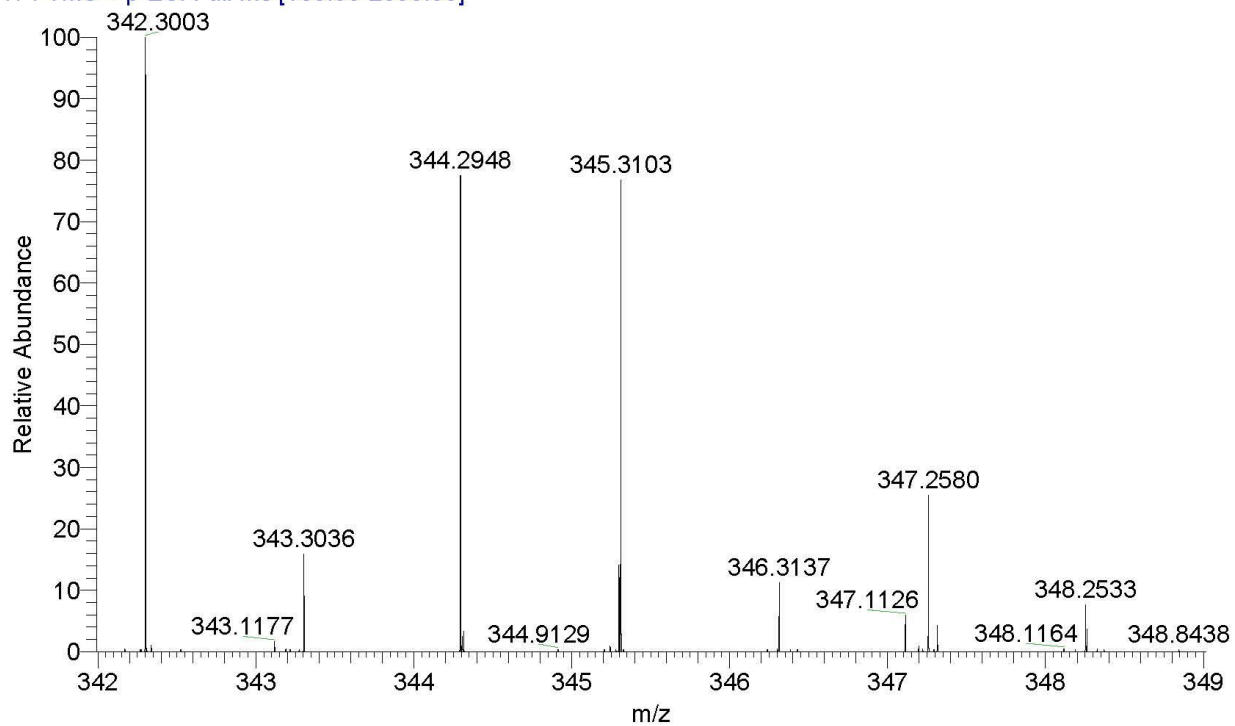
Figure S23: Kava m/z 342
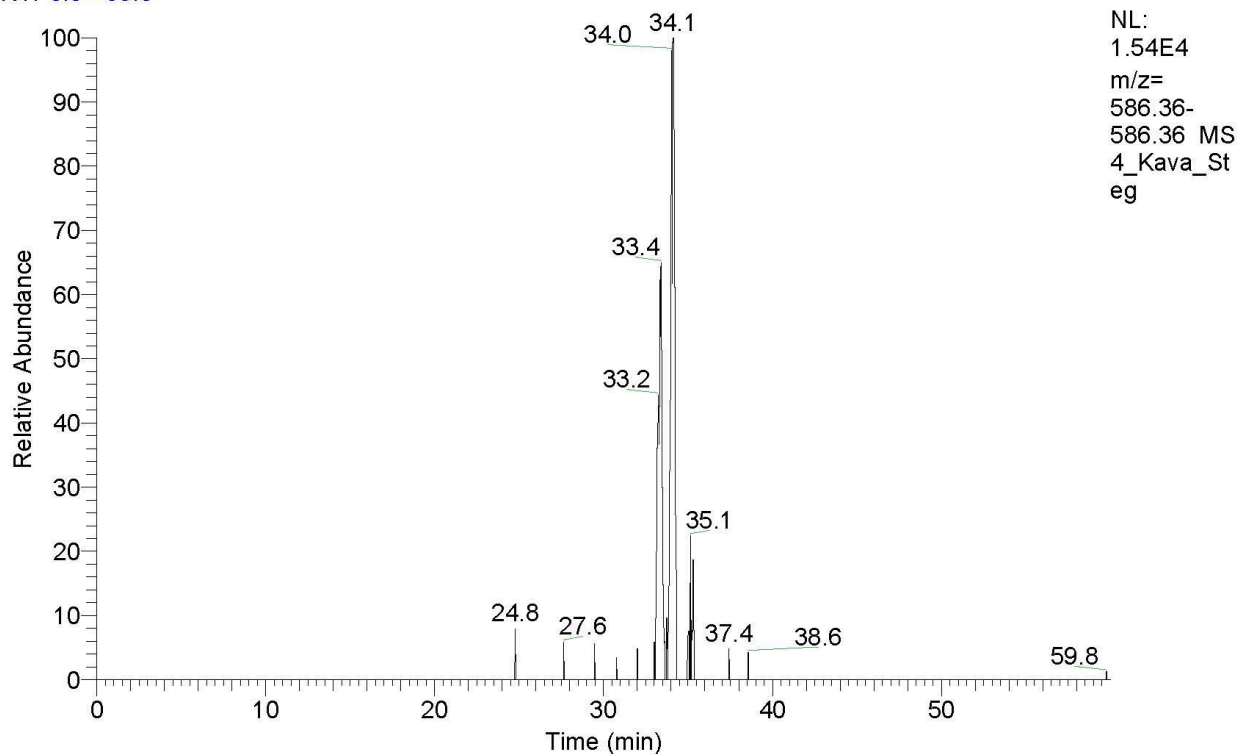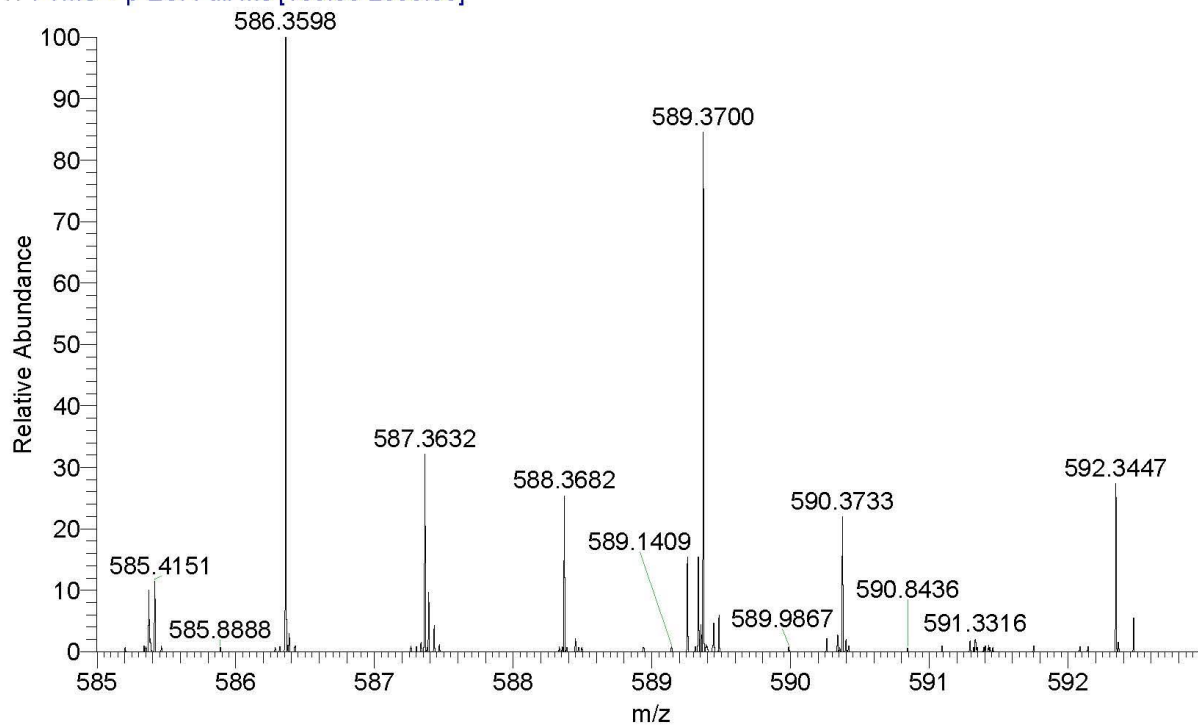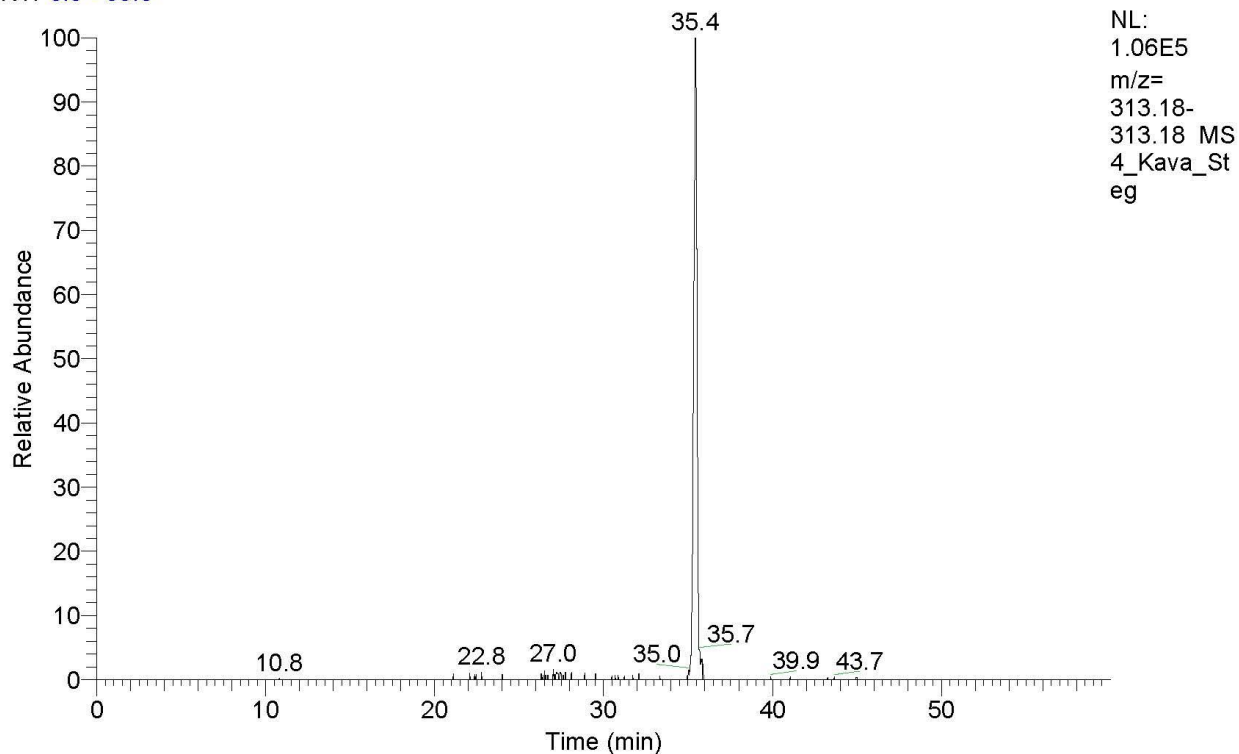
Figure S24: Kava m/z 586

Figure S25: Kava m/z 313

RT: 0.0 - 60.0

NL:
1.32E4
m/z=
588.37-
588.38 MS
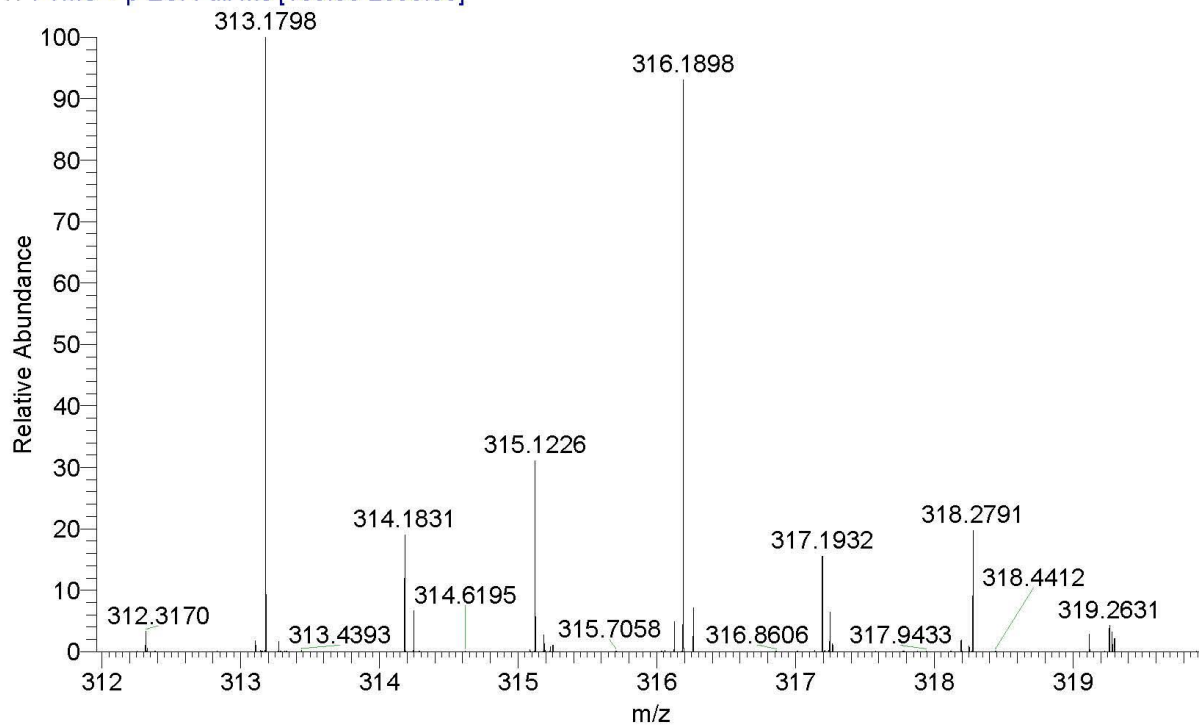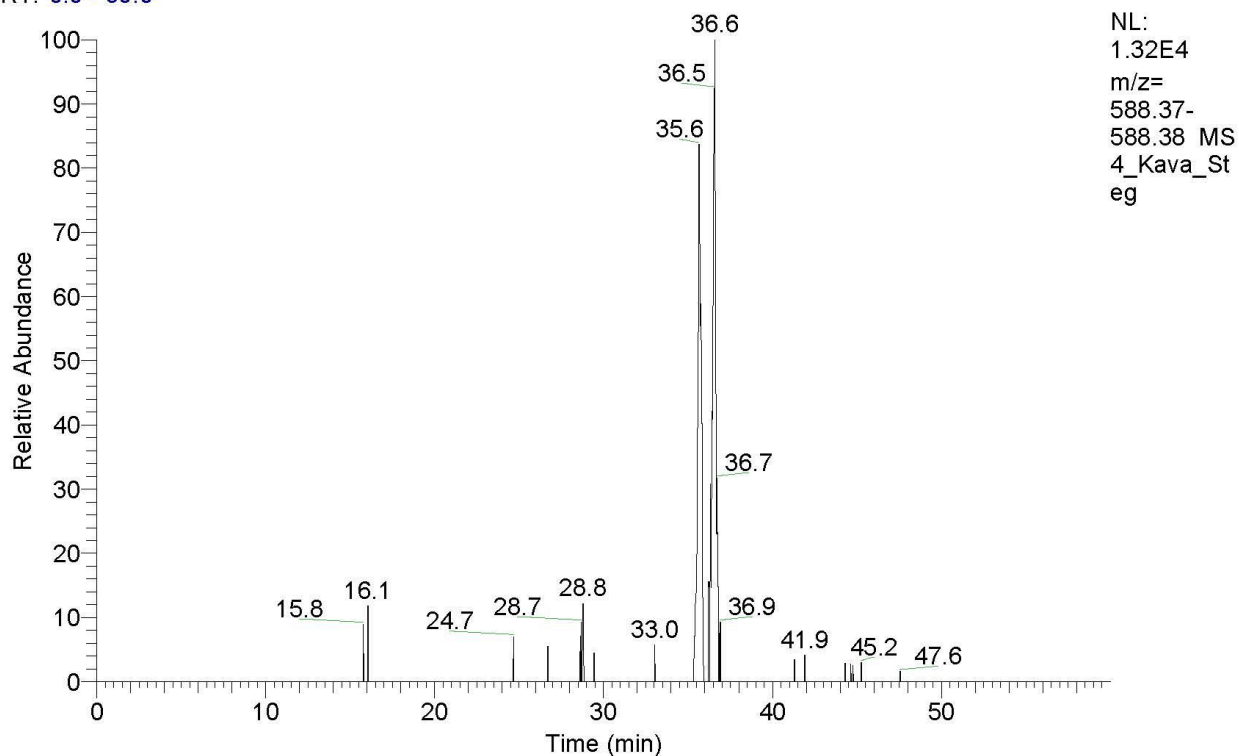4_Kava_St
eg

4_Kava_Steg #1088-1094 RT: 35.55-35.73 AV: 7 NL: 7.58E3
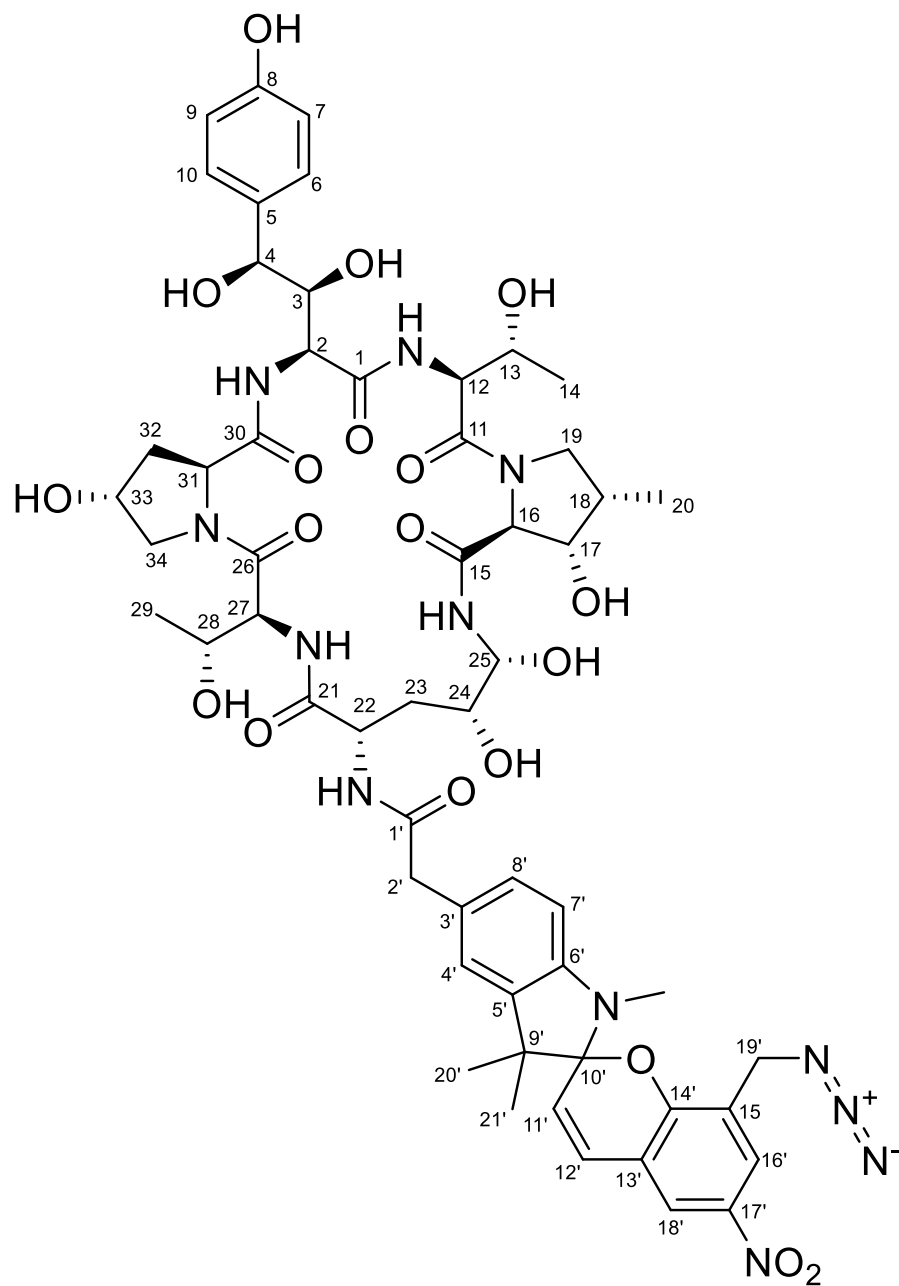T: FTMS + p ESI Full ms [100.00-2000.00]
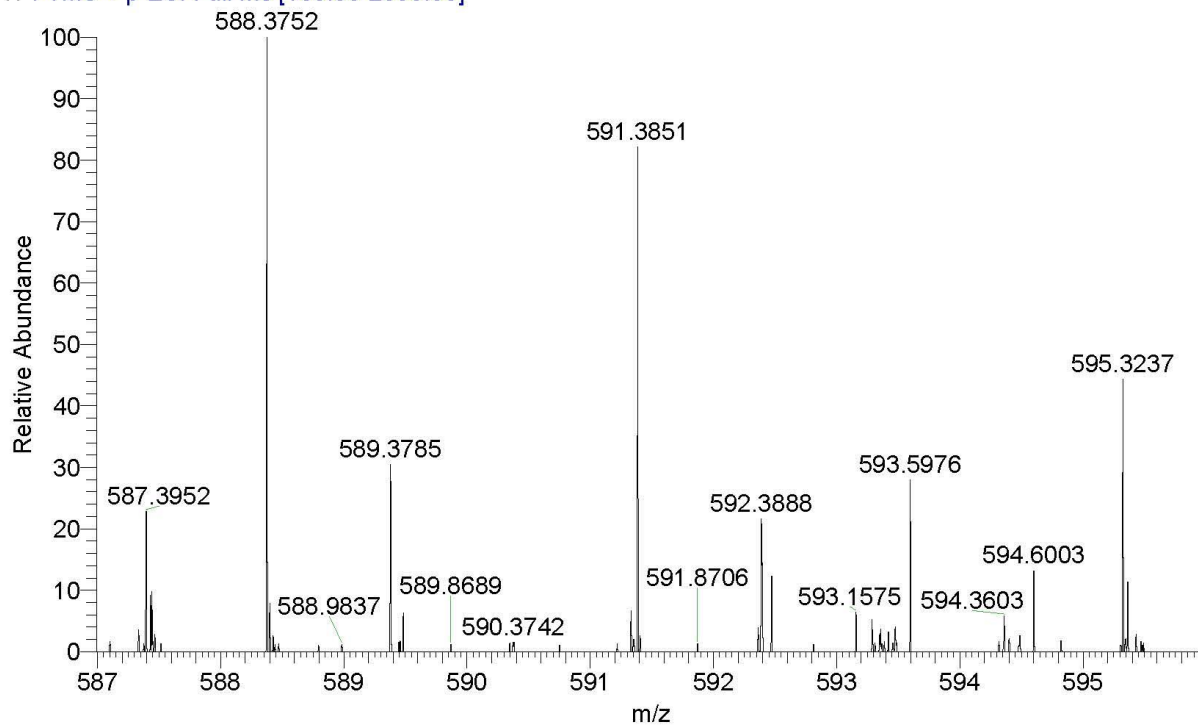
Figure S26: Kava m/z 588

94

A

Figure S27: A) Structure of nitroBIPS echinocandin.

B

Position	Proton	Carbon	Position	Proton	Carbon
1		169.0, C	19	3.18 m; 3.84 m	51.3, CH2
2	NT	NT	20	0.94 d 6.8	11.1, CH3
3	3.91 m	75.7, CH	21		NT
4	4.18 m	73.2, CH	22	4.21 m	49.7, CH
5		132.4, C	23	1.67 m; 1.88 m	34.7, CH2
6./ 10	7.00 d 8.3	128.0, CH	24	3.75 m	69.3, CH
7./ 10	6.67 d 8.3	114.7, CH	25	4.98 m	72.7, CH
8		156.8, C	26		170.5, C
11		168.1, C	27	4.74 m	56.7, CH
12	4.63 dd 9.6, 5.0	55.8, CH	28	4.33 m	65.6, CH
13	3.95 m	67.8, CH	29	1.04 m	19.4, CH3
14	1.04 m	19.4, CH3	30		170.8, C
15		169.3, C	31	4.31 m	60.9, CH
16	4.19 m	68.4, CH	32	1.86 m; 2.20 m	37.4, CH2
17	4.02 m	73.2, CH	33	4.40 m	69.1, CH
18	2.34 m	37.1, CH	34	3.81 m; 3.67 m	55.7, CH2
1'		170.4, C	12'	7.23 d 10.5	127.7, CH
2'	3.40 m	41.4, CH2	13'		119.2, C
3'		127.6, C	14'		157.0, C
4'	7.02 m	121.8, CH	15'		122.8, C
5'		135.6, C	16'	8.16 m	125.2, CH
6'		145.6, C	17'		140.1, C
7'	6.52 d 8.0	107.1, CH	18'	8.24 m	122.4, CH
8'	7.00 m	128.1, CH	19'	4.20 m; 4.28 m	47.4, CH2
9'		127.6, C	20'	1.12 s	25.9, CH3
10'		107.0, C	21'	1.22 s	19.7, CH3
11'	6.01 d 10.5	121.2, CH			

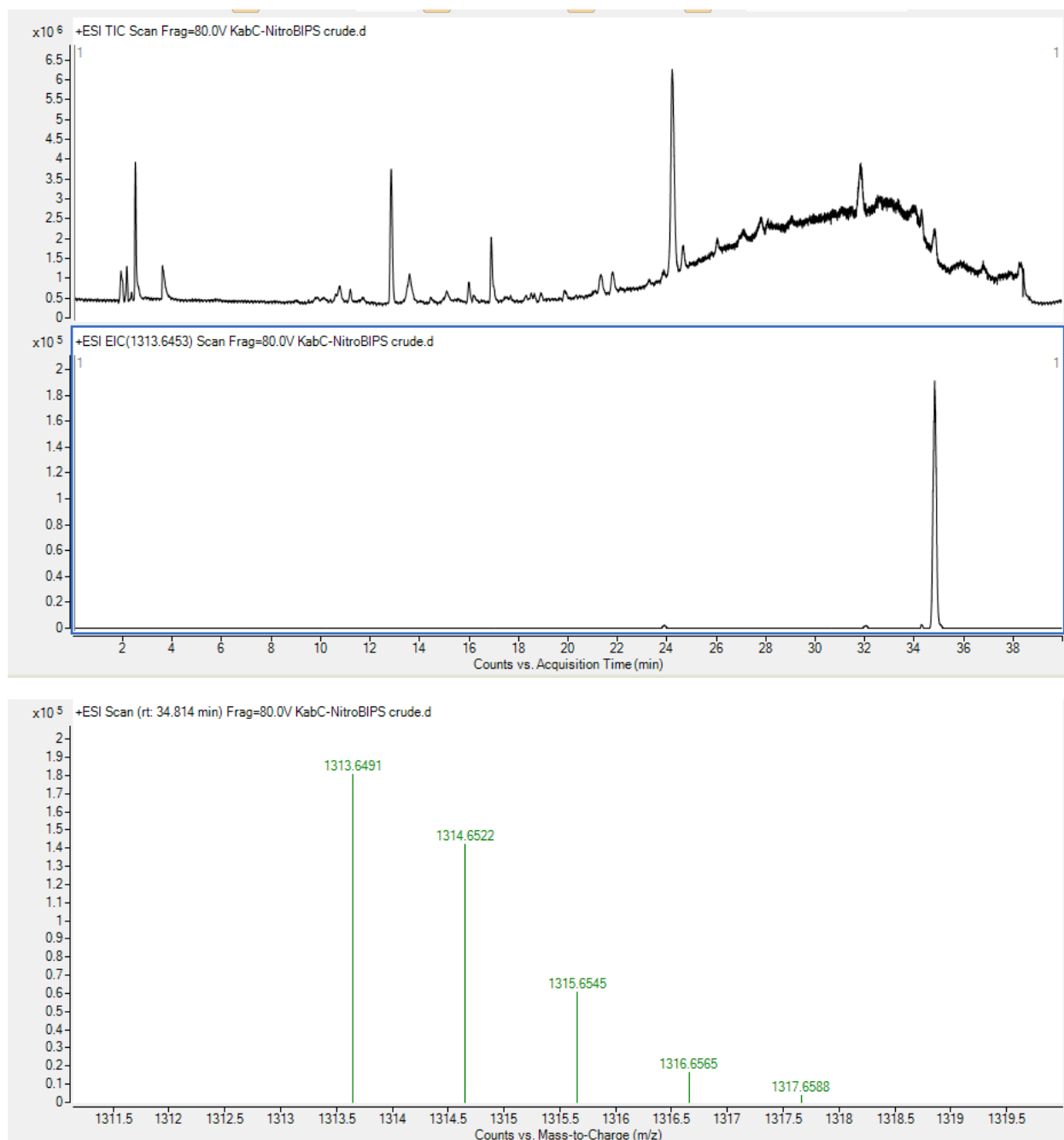**Figure S27: B)** NMR table for nitroBIPS-echinocandin.

Figure S28: Mass spectra for crude nitroBIPS-Kabiramide C. The total ion chromatogram is shown on top. The Extracted ion chromatogram is shown in the middle. The Mass spectrum including a measured monoisotopic mass of 1313.6491 is on the bottom. This is a 2.89 ppm error relative to the calculated exact mass of 1313.6453 for nitroBIPS-Kabiramide C. This is well within the error range of the equipment and indicates a successful synthesis.