UNDERWOOD, ETHAN B., M.S. Discovery of Conserved Peptide Sequences to Gain Insight into Cyclopeptide Alkaloid Formation in *Ceanothus americanus*. (2022)
Directed by Dr. Jonathan R. Chekan. 44 pp.

Cyclopeptide alkaloids are a large and abundant class of plant cyclopeptides with over 200 known analogs and a range of bioactivities. While these molecules have been known and studied for decades, their biosynthetic pathways remain a mystery. Utilizing a transcriptome-mining approach, we have linked cyclopeptide alkaloids from *Ceanothus americanus* to dedicated ribosomally synthesized and post-translationally modified peptide (RiPP) precursor peptides. This approach has become a way to uncover precursor peptides to search for putative cores which could lead to new cyclopeptide alkaloid compound discoveries. Our approach yielded two new putative cores which were confirmed by metabolomic, fragmentation, and transcriptome data. It has also confirmed the core sequences for the more than ten known compounds produced by *C. americanus*. Some of the compounds produced by *C. americanus* are known to have anti-plasmodial, anti-inflammatory, anti-bacterial, anti-fungal, and some sedative activity.[1] *C. americanus*, also known as New Jersey Tea, has been used for many generations in folk medicine to treat blood coagulation, pressure, spleen pain, and even cancer.[2] These results have allowed us to explore the plant for new putative cores and has set the groundwork for determination of the genes responsible for the biosynthetic pathways and bioactivity studies of the known and new compounds.

DISCOVERY OF CONSERVED PEPTIDE SEQUENCES TO GAIN INSIGHT INTO CYCLOPEPTIDE ALKALOID FORMATION IN *CEANOTHUS AMERICANUS*

by

Ethan B. Underwood

A Thesis
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro

2022

Approved by

Dr. Jonathan R. Chekan
Committee Chair

APPROVAL PAGE

This thesis written by Ethan B. Underwood has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Jonathan Chekan

Committee Members

Dr. Nicholas Oberlies

Dr. Norman Chiu

October 17, 2022

Date of Acceptance by Committee

October 17, 2022

Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

I.A. Background

I.A.1. Overview of Alkaloids

Alkaloids are an important class of diverse compounds which include a basic nitrogen in a heterocyclic ring and are typically derived from amino acids. The term alkaloid stems from the Arabic term ‘al qualja’ which roughly translates to ashes of plants.[3] To date, more than 12,000 alkaloids across 150 different families have been identified in plants[4] and around twenty percent of flowering plants contain alkaloids. These alkaloids serve as important secondary metabolites in the plant whose roles are proposed to be involved in metabolism, catabolism, end products or waste of these pathways, nitrogen storage, defensive agents against predators, and growth regulators. Alkaloids have been found in many plants used by native people groups to treat a variety of ailments for generations.[4] This background of use in folk medicine reinforces the idea that alkaloids are a prime class to pursue for potential drug targets.

The term peptide alkaloid was coined by the Goutarel group in the early 1960s and cyclopeptide alkaloids are a subdivision of this subclass.[5] Cyclopeptide alkaloids contain 13,14, or 15-membered macrocycle with an ether bond and strylamine moiety. They are found commonly in the plant kingdom across various families including but not limited to Rhamnaceae, Sterculiaceae, Pandaceae, Rubiaceae, Uritcaceae, Hymenocardiaceae, and Celastraceae.[6] Cyclopeptide alkaloids have been unknowingly used in folk medicine across the globe for generations. Uses in teas, salves, and ointments to remedy different ailments have been shown. These compounds are known to have many biological and physiological activities

including but not limited to anti-plasmodial[7], anti-inflammatory[8], anti-bacterial[9], and anti-fungal[10] biological activities. Other activities include sedative, and hypnotic activities.[11]

Cyclopeptide alkaloids can be found in many tissues of the plant however, the abundancies of these compounds are dependent on various variables like growth stage, region, and soil nutrient levels. Isolated mixtures are often overly complex and contain many constituents which can make analysis difficult. Extractions seeking alkaloid products often yield very little product amounts ranging from 0.01 to 1% w/w. This variable yield can make high throughput screenings very difficult to perform due to the lack of product.[12] One of the characteristic aspects of cyclopeptide alkaloids is that their fragmentation patterns are well studied and predictable. This predictability allows for guided analysis of mass spectrometric data.

While many plants are known to produce cyclopeptide alkaloids, *Ceanothus americanus* is particularly noteworthy because it produces at least ten cyclopeptide alkaloids (**Figure 1**). The known cyclopeptide alkaloid compounds known to be produced by *C. americanus* are well studied and published in a review by Tan and Zhou from 2006.[13]

I.A.2. Cyclopeptide Alkaloid Structure in *Ceanothus americanus*

The known compounds found in *C. americanus* are Type 1A1 cyclopeptide alkaloids. Type 1 cyclopeptide alkaloids are classified as basic compounds containing *p* or *m-ansa* structure with a 13, 14, or 15-membered ring where the 1, 3 or 1, 4 positions of the benzene ring are used to form an ether bridge.[13] This type also contains one styrylamine moiety, two or three ring-bound α-amino acid residues, with the possibility for one or two side-chain *N*-methyl or *N*,*N*-dimethyl α-amino acid residues.[13] Stereochemical assignments are inconsistently reported in the literature, because of this only frangulanine is shown with stereochemistry. Frangulanine has

been crystalized[14] and synthezied[15], giving strong evidence for its stereochemical assignments. The remaining cyclopeptide alkaloids do not have strong enough evidence to support stereochemical assignments therefore, they have been left unassigned and are shown below in **Figure 1**. Three of the cyclopeptides have some ambiguity about the identity of one amino acid therefore, all options are shown.

Figure 1. Cyclopeptide Alkaloids Known to be Produced by *C. americanus*

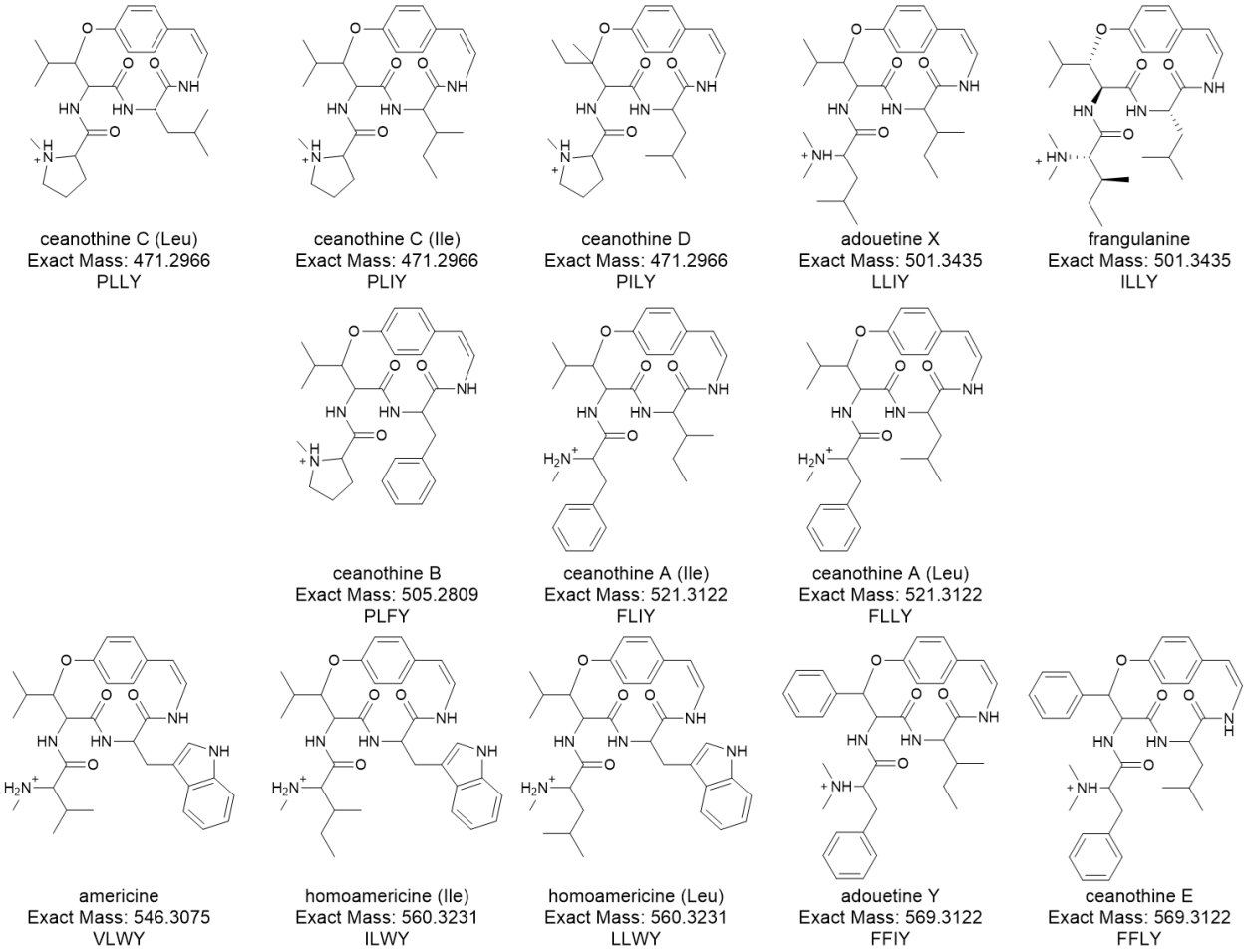


Fig. 1 Depicted above are the 9 known cyclopeptide alkaloids produced by C. americanus shown as [M+H+].

While little is known about how these cyclopeptide alkaloids are formed, we hypothesize that these specific cyclopeptide alkaloids are ribosomally synthesized and post-translationally

modified peptide (RiPP) natural products because of their peptidic features and the nature of the modifications observed across all of the compounds.

I.A.3 Overview on Ribosomally Synthesized and Post-translationally Modified Peptide (RiPP) Natural Products

Ribosomally synthesized and post-translationally modified peptide (RiPP) natural products are secondary metabolites that are biosynthesized by the ribosome and subsequently modified by tailoring enzymes. RiPP biosynthesis can often by categorized by different biosynthetic routes, with three shown in **Figure 2** below.

Figure 2. RiPP Biosynthetic Route Overview

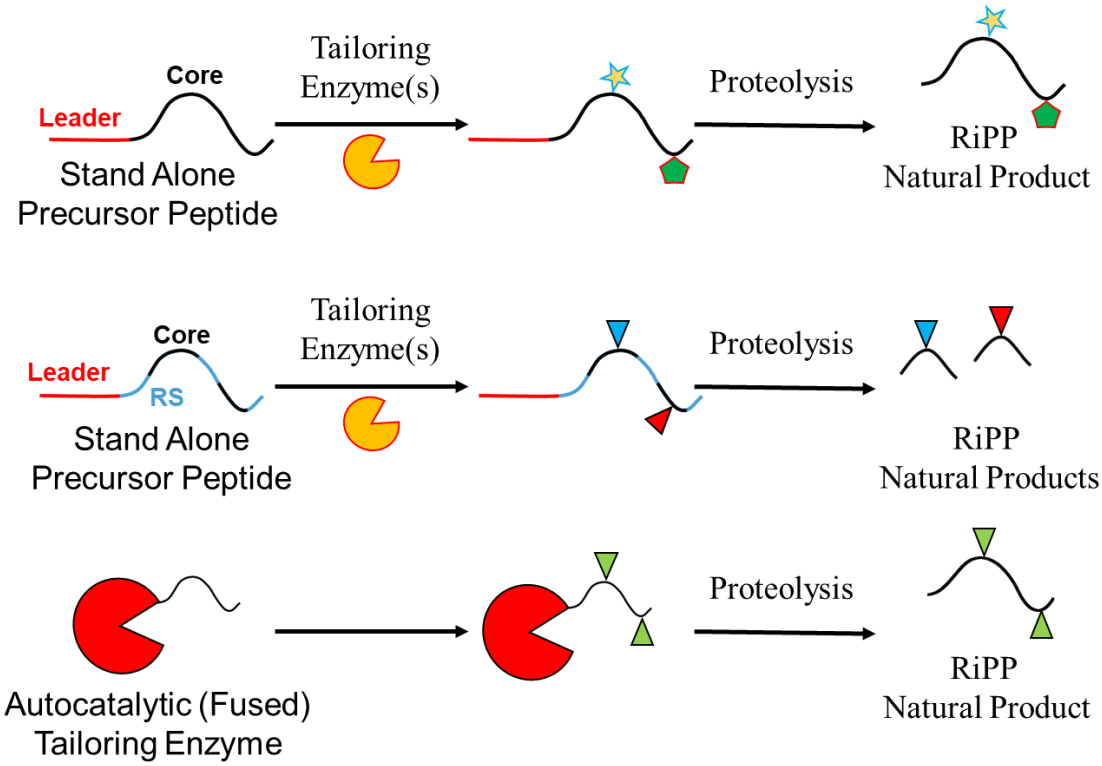


Fig. 2 RiPP natural products can follow a variety of biosynthetic motifs.

Two of the three RiPP biosynthetic pathways contain precursor peptides composed of leader peptides, cores, and sometimes recognition sequences. The leader and recognition sequence are responsible for guiding the post-translational modifications to the core sequence. The core

sequence ultimately composes the final secondary metabolite product. The third and final motif shows the core fused to the enzyme. An autocatalytic reaction where the core acts as the catalyst for the enzyme which modifies the core and detaches through proteolysis to give the RiPP natural product.[16] RiPP precursor peptide gene sequences can also follow a variety of motifs shown in **Figure 2**. The genes can code for a single core, or they can code for multiple cores which can be the same throughout the gene or can vary from core to core.

Ribosomally synthesized and post-translationally modified peptide natural products as a class are very diverse and consists of over 40 families.[17] Some examples include cyanobactins, thiopeptides, lasso peptides, lantibiotics, and sactipeptides. Nisin is a well-known lantibiotic which has multiple post-translational modifications. These modifications yield five different cyclizations via thioether bridges and multiple dehydrations. The precursor peptide consists of a 23 amino acid leader peptide and a 34 amino acid core. The biosynthesis moves through four modifications and this process yields the 34 amino acid core as the mature RiPP natural product Nisin.[18] Nisin is used as a natural preservative in food products to inhibit pathogenic food borne bacteria.

Another family of RiPP natural products, Cyanobactins are produced by a diversity of cyanobacteria and are small, cyclized peptides. These cyanobacteria are found in a variety of environments including but not limited to, terrestrial, marine, and freshwater.[19] Most cyanobactins consist of five to ten cyclized amino acids and are bioactive. These RiPP natural products contain modifications like oxazoles, oxazolines, thiazoles, thiazolines and disulfide bridges. One example of a cyanobactin is axinellin C, axinellin C was isolated from a Fijian marine sponge, *Stylotella aurantium*.[20] The structures of nisin and axinellin C are shown in **Figure 3.**

Structures of Nisin and Axinellin C. Nisin and axinellin C are two examples of RiPP natural products.

Figure 3. Structures of Nisin and Axinellin C

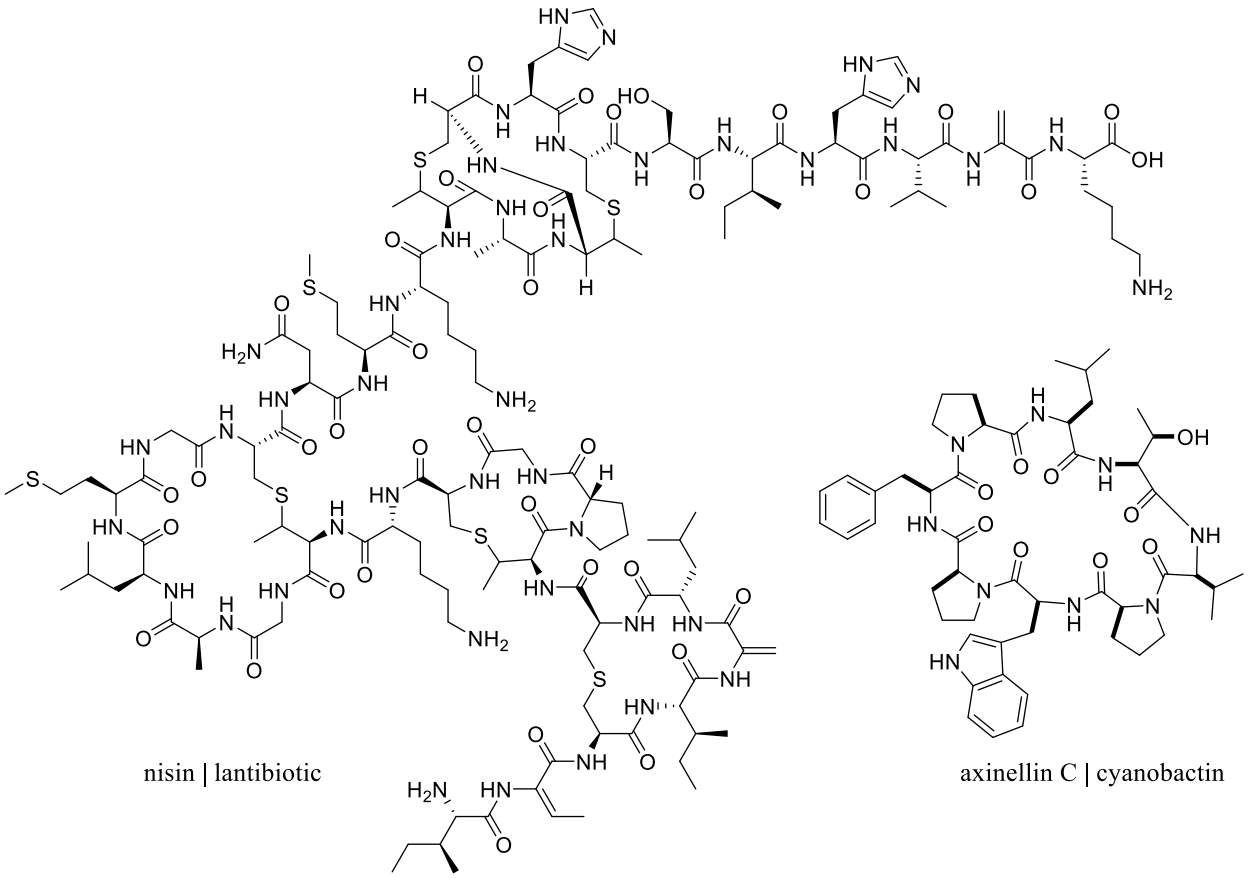


Fig. 3 Structures of Nisin and Axinellin C. Nisin and axinellin C are two examples of RiPP natural products.

While bacteria are most often recognized as RiPP producers, this family of natural products is also found in plants and fungi. In particular, burpitides are a new emerging class of RiPP natural products seen in plants like *C. americanus*. These burpitides are classified by the covalent linkage between side chain members that result in a side-chain macrocyclization of the core amino acid sequence, some examples are shown in **Figure 4**. In the past, sequenced burpitide RiPP pathways have been shown to contain a BURP-domain that was fused to the

precursor peptide, these fused systems have been shown to be unique to plants.[21]

Figure 4. Examples of Burpitide Natural Products.

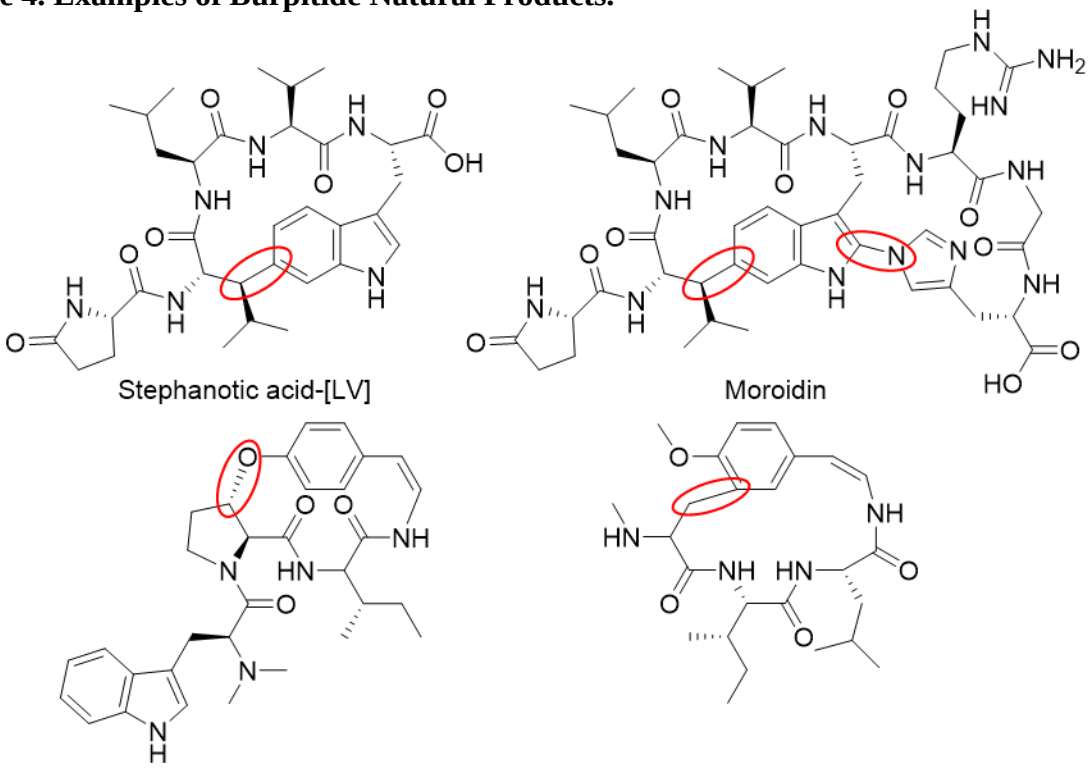


Fig. 4 The covalent side-chain linkage is notated by a red circle.

Work published by the Kersten group showed that BURP-domain cyclases are copper-dependent and autocatalytic enzyme systems.[22] These BURP-domains are responsible for the post-translational modifications that result in the mature RiPP natural product. Originally these BURP-domains had been thought to only exist as fused to the gene that included the precursor peptide. However, there is some evidence that split BURP systems could exist where the BURP-domain is not included in the gene with the precursor peptide and are called split BURP-domain systems.[23]

I.B. Basis for Study

I.B.1. Fighting the Low Yields of Traditional Extraction Methods and Sample Acquisition

While the plant kingdom has been a treasure trove for potential drugs the extraction and purification process is often tedious and has incredibly variable yields for the target compound or compounds. Low yields require large amounts of plant samples to overcome this challenge so that a group can move onto bioactivity, method of action, and drug development studies. Even total synthesis pathways, which typically require many steps are often neither cost nor time effective.[24, 25] Finding a way to elucidate the biosynthetic routes and genes that these compounds are formed by could allow for plasmid-based protein production on a more effective scale than extraction and total synthesis pathways alone.

I.B.2. Sample Acquisition

Discovery of biosynthetic pathways to produce RiPP natural products will allow for more diverse studies across the kingdoms. For example, while *C. americanus* is widely available and a good target for bioactive RiPP natural products it is not an endangered species. Other bioactive RiPP natural products could be present in endangered or rare species, which would make sourcing a challenge or even unethical. If we can determine a way to elucidate the biosynthesis pathways using a relatively small amount of material, we can further study the RiPP natural products found in endangered species while preventing further harm to these species.

I.C. Significance

I.C.1. A New Route to Ribosomally Synthesized and Post-translationally Modified Peptide Natural Product Discovery

The Chekan group has interests in RiPP natural products. However, they have always started a project with publicly available bioinformatic data and end the project with sequence and compound presence and functionality studies. This project was the reversal of that route and started with past research from the 1950s on cyclopeptide alkaloids from *C. americanus*. This route uses compound presence and confirmation and ends with sequencing, transcription, and translation of genomic data to study and search for linkage of cyclopeptide alkaloids to precursor peptides. This route allows for precursor peptide guided discovery and targeted identification of undiscovered analogs. The project used compounds that have been known for decades but the biosynthesis pathways were never uncovered. This project aimed to uncover these pathways and along the way found a method to predict new molecules.

I.C.2. Formulation of a new Route to Compound Synthesis

Discovery of the biosynthesis pathway and the genes required to generate these compounds would allow for large scale production of pure compounds utilizing plasmid-based protein production in *E. coli* or in plant-based systems. While the full biosynthetic pathway has not been elucidated for these compounds, the progress made thus far has kickstarted the process for other plants that share these compounds. If these compounds are found in other species and one or more of those species has a full genome published publicly it can aid in determination of the genes required for a compound's production. We hypothesize that all the cyclopeptide

alkaloids, regardless of source, would have many of the same enzymes for the post-translational modifications across the board as those found in *C. americanus*.

CHAPTER II: EXPERIMENTAL

II.A. Approach

II.A.1. Certification of Previous Research on Cyclopeptide Alkaloids Found in *Ceanothus americanus*

Previous published research methods[1] served as a proof of concept that these cyclopeptide alkaloids could be extracted and studied. We build upon these approaches to both extract and partially purify the cyclopeptide alkaloids to separate out many of the other metabolites we were not interested in. Using a method optimized by an undergraduate student in the lab, Imani Khedi, we compared nursery raised plants with commercially available root power. After extraction we opted for ultra-high-performance liquid chromatography tandem mass spectrometer (UHPLC-MS/MS) to analyze the extracts and fragmentation data to confirm the presence of known cyclopeptides in our samples. This process allowed us to solidify the hypothesized core sequences that would be present in our *C. americanus* sequencing data.

II.A.2. Validation of Compound Presence Leads to RNA Extraction and Transcriptomics

After confirming the compound presence between nursery grown plants and commercial root powder, we moved to RNA extraction. We opted to find a method that worked for leaf, stem, and root tissues from *C. americanus* so that we could correlate the presence of precursor peptide transcripts with cyclopeptide alkaloids in specific *C. americanus* tissues. After determining the success of this method, we sent the RNA to the High Throughput Sequencing Facility (HTSF) at the University of North Carolina at Chapel Hill for quality checks and sequencing. The raw sequencing reads were processed with a Trinity[26] and Transdecoder[27] pipeline to assemble and translate the sequencing data. This pipeline allowed us to search the

transcriptome for the putative cores of the known compounds to identify putative precursor peptides.

II.A.3. CD-Hit Guided Discovery of the Recognition Sequences and Stand-alone Precursor Discovery

Based on the cyclopeptide alkaloid metabolomic data, we predicted the core sequences that could lead to the observed natural products. We searched the transcripts for these core sequences to generate 2189 possible precursor peptide sequences. We utilized these with the software CD-Hit[28] to cluster transcripts by similarity and reduce the number of false positives and identify transcripts with high levels of conservation that only vary in the predicted core sequences. This clustering would allow us to confirm the conserved sequences while limiting the number of false positives within the transcripts in order to determine what constitutes a precursor peptide. Using these newly uncovered precursor peptides, we searched the assembled transcriptome for additional precursor peptide sequences in order to find new, unpublished cores that could match to unknown masses found in the metabolomic and fragmentation datasets.

II.B. Methodology

II.B.1. Known Cyclopeptide Alkaloid Confirmation and Comparison

The comparison of nursery raised plants and commercially available root powder and confirmation of the known cyclopeptide alkaloids produced by *C. americanus* was completed in a three-step process. This process consisted of extraction, fractionation, and LC-MS analysis. Commercial root powder extractions were completed by Imani Khedi, who also developed the extraction protocol. This extraction and fractionation method was completed in eight steps followed with UHPLC-MS/MS analysis.

I first focused on taking this method and applying it to live plants to confirm the presence of the cyclopeptide alkaloids. The extraction process was completed by preparing one to two grams of plant tissue by a two-step process. This preparation process consisted of lyophilization of the tissue followed by grinding into a fine powder. After the tissues were prepared, they were first extracted in petroleum ether overnight, this first extraction was followed by vacuum filtration and the filtrate was transferred for a follow up overnight extraction in methanol. The methanol extraction was also vacuum filtered, and the liquid phase was kept and dried down to residue by rotational evaporation. The residue was resuspended in 25 mL of 0.1% formic acid and water. After resuspension, the extract was fractionated using a Hypersep C18 200 mg cartridge (Thermo Scientific) following the fraction compositions listed in **Table 1**.

Table 1. Fraction Composition for Molecule Confirmation and Comparison

Sample	Composition
Blank	100 % Acetonitrile
Flow-through	Collected flow-through from loading the cartridge
15 % Fraction	15 % Acetonitrile in Water with Formic Acid (0.1%)
30 % Fraction	30 % Acetonitrile in Water with Formic Acid (0.1%)
45 % Fraction	45 % Acetonitrile in Water with Formic Acid (0.1%)
60 % Fraction	60 % Acetonitrile in Water with Formic Acid (0.1%)
100 % Fraction	100 % Acetonitrile in Water with Formic Acid (0.1 %)

The fractions were collected in scintillation vials and dried down in the Savant SpeedVac SPD120. The remaining sample residues were resuspended in 50 µL of 50% aqueous acetonitrile (ACN) with 0.1% formic acid. 15 µL of the resuspended samples were transferred into inserts held in autosampler vials which were inserted into a solvent tray with the blank and analyzed on Q-Exactive Plus Orbitrap Mass Spectrometer (Thermo Scientific). The Q-Exactive was ran in positive ion mode with a mass range from 300-1000 mass to charge (m/z) at collision energies 20, 25, and 30 eV. The liquid chromatography set-up was a BEH 1.7 µm C18 reverse phase 130

Å 2.1x50 mm column with a ramping gradient: 0 min, 5% B; 1 min, 5% B; 11 min, 100% B; 12 min, 100% B; 12.1 min 5% B; 13 min, 5% B where solvent A was water with formic acid (0.1 %) and solvent B was acetonitrile with formic acid (0.1 %). UHPLC-MS/MS data was analyzed using Qual Browser Thermo Xcalibur software package (v.3.0.63 Thermo Scientific).

II.B.2. RNA Extraction, Sequencing, and Transcriptomics

Finding an RNA extraction method that worked for the leaf, stem, and root of the plant proved to be difficult. This difficulty led to three different methods being tried until the problem was solved. The first method was a plant RNA extraction kit developed by Qigen, the second was a phenol, chloroform, and isoamyl alcohol extraction, and the third was a cetrimonium bromide-based extraction. This final method worked for all of the target tissues which were leaf, stem, and root. The stock solutions required for extraction buffer are described in **Table 2** below.

Table 2. RNA Extraction Buffer Contents

Component	Buffer Solution Makeup (40 mL)
100 mM Tris-HCl at pH 8.0	4 mL
2 M Sodium Chloride (NaCl)	16 mL of 5 M NaCl
20 mM Ethylenediamine Tetraacetic Acid (EDTA)	1.6 mL 0.5 M EDTA
2 % Cetrimonium Bromide (CTAB)	0.8 g CTAB
50 mM Dithiothreitol (DTT)	0.306 g DTT

This method published by Sim, Ho, and Phang[1] consisted of 20 steps and started with tissue preparation. Plant tissue was collected and weighed between 0.1 and 0.3 grams after collection, the tissue was flash frozen in liquid nitrogen and ground to a coarse powder in-order-to prevent coagulation in the extraction buffer. After grinding, the coarse powder was collected into a microcentrifuge tube that was chilled in liquid nitrogen to prevent spattering of the tissues. The extraction buffer was added to the microcentrifuge tube at a ratio of 1:10 (w/vol) and was gently vortexed for maximum contact of the extraction buffer with the coarse tissues. Sodium dodecyl sulfate (SDS) was added to the tube to reach a content of two percent for the entire solution.

After the addition of SDS the tube was vortexed again and then placed into a 4 ℃ centrifuge which was ran for ten minutes at 15,900 times gravity. After centrifugation the aqueous phase was transferred into a new microcentrifuge tube. To this tube, 0.3 (vol/vol) of ethanol was added and mixed through inversion. After the addition of absolute ethanol, the chloroform-isoamyl alcohol (24:1) solution was added at a ratio of 1:1 (vol/vol) and was centrifuged again for ten minutes under the same settings. After centrifugation the supernatant was transferred to a new microcentrifuge tube to which 8 M lithium chloride was added to reach a final concentration of 2 M. This solution was mixed by inversion eight times and was placed into the -80 ℃ freezer for at least two hours to induce precipitation of the RNA pellet. After the incubation period, the tube was centrifuged for 30 minutes, and the liquid phase was removed to retain the pellet. The remaining pellet was dissolved in diethyl pyrocarbonate (DEPC) treated water and was put through a repeat ethanol wash, chloroform-isoamyl alcohol purification, centrifugation, precipitation by lithium chloride, and a minimum of a two-hour incubation period in the -80 ℃ freezer again. After this second purification the tube was also centrifuged for 30 minutes, the liquid phase was removed, and ice-cold 70% ethanol was added to the remaining pellet. This pellet-ethanol mixture was then placed into the speed vac to remove the ethanol and purify the pellet. After complete removal of the ethanol, the pellet was dissolved in 25 μL of DEPC treated water. The dissolved pellets were subjected to analysis by spectra drop and a Biotum RNA quantification kit and kept in the -80 ℃ freezer for storage.

The most promising RNA extraction samples, which had the highest concentrations of RNA and consistent absorbances for purity, were submitted to the High Throughput Sequencing Facility (HTSF) lab at UNC Chapel Hill for quality checks and sequencing. The RNA sequencing was completed using a KAPA RNA HyperPrep kit with RiboErase. After sequencing

by the HTSF lab the returned data was run through a Trinity and TransDecoder pipeline. Trinity software was used with default settings to assemble the data to create a de novo assembled transcriptome. This transcriptome was then processed with TransDecoder to give translated transcripts. The transcripts were analyzed to determine core amino acid sequences for the known cyclopeptide alkaloids produced by *C. americanus*.

II.B.3. Bioinformatic Analysis and CD-Hit Guided Discovery

The last piece in the pipeline was CD-Hit which allowed for clustering of the transcripts by a preset similarity percentage. This clustering was completed using the University of California San Diego web server[29] at a similarity minimum of 40%. After clustering, the grouped transcripts were analyzed to look for clusters made up of different putative cores within transcripts that had high levels of amin acid sequence similarity. This allowed for identification of precursor peptides and enabled us to search for potential new core sequences to match unknown hits in the metabolomic data.

II.B.4. Extraction and Purification Method Determination

After confirmation of the core sequences and presence of the compounds of interest was completed, we moved into extraction, purification, and isolation method determination. The difficulty with isolating the molecules of interest is that their masses are very similar. However, before we tackle this issue, we aimed to generate clean crude extracts. This method tested different liquid-liquid extraction liquids against one another to determine which gave the cleanest mass spectra data. We developed a methanol extraction of root powder which is extracted overnight twice, and vacuum filtered in between extractions. The liquid and solid phases are kept and the solid is reextracted. After the second extraction, the liquid phases are combined and dried to a residue on the rotational evaporator. The residue is resuspended in basic water at pH 10 and

is then put through multiple rounds of liquid-liquid extraction. The liquid-liquid extraction pipeline tested dichloromethane, ethyl acetate, and *n*-butanol and tested them individually. After eight rounds of liquid-liquid extractions using the three liquids we tested, each layer was tested on mass spec to determine the location and contamination levels of the samples.

CHAPTER III: RESULTS AND DISCUSSION

III.A. Molecule Confirmation and Comparison

III.A.1. Metabolomic Analysis of Live *Ceanothus americanus* Plants

To aid in the metabolic analysis of nursery raised plants, molecular networks were generated using the global natural products social molecular networking (GNPS) site. This resource compares mass spectra by one of two paths. The first path is comparison to deposited spectra held in the library and the second is comparing the input spectra to one another. The path chosen for the data was the input versus input comparison. Initially we had thought the cyclopeptide alkaloids were found across each of the leaf, stem, and root tissues. However, it was discovered that the root samples caused contamination and carry over into the leaf and stem samples. The comparison shown in **Figure 5** showed that the known cyclopeptide alkaloids are found in the roots of the plant alone and are related by their fragmentation patterns observed in the metabolomic data.

Figure 5. GNPS Molecular Network

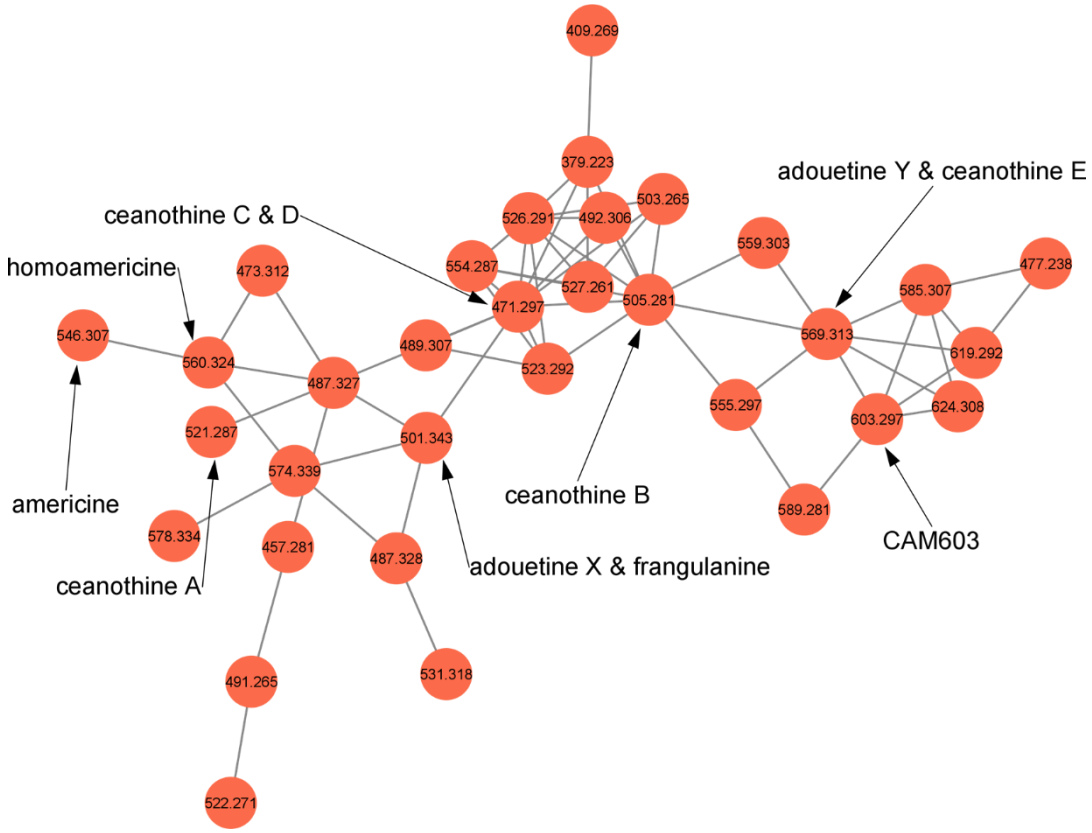


Fig. 5 GNPS Molecular Network. The data shown indicates that the known compounds are found within the root tissues of the plant only.

III.A.2. Molecule Presence Confirmation by Mass Spectroscopy

Molecule presence was determined by searching for the molecular weight of the known compound and analyzing the fragmentation data in the metabolomic datasets. Cyclopeptide alkaloids fragment in predictable ways which were used to confirm the presence of specific molecules. **Figures 6 through 12** detail the fragmentation data that confirmed the compound presence and suggested hypothesized amino acid sequences while leaving some ambiguity for isomers

Figure 6. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

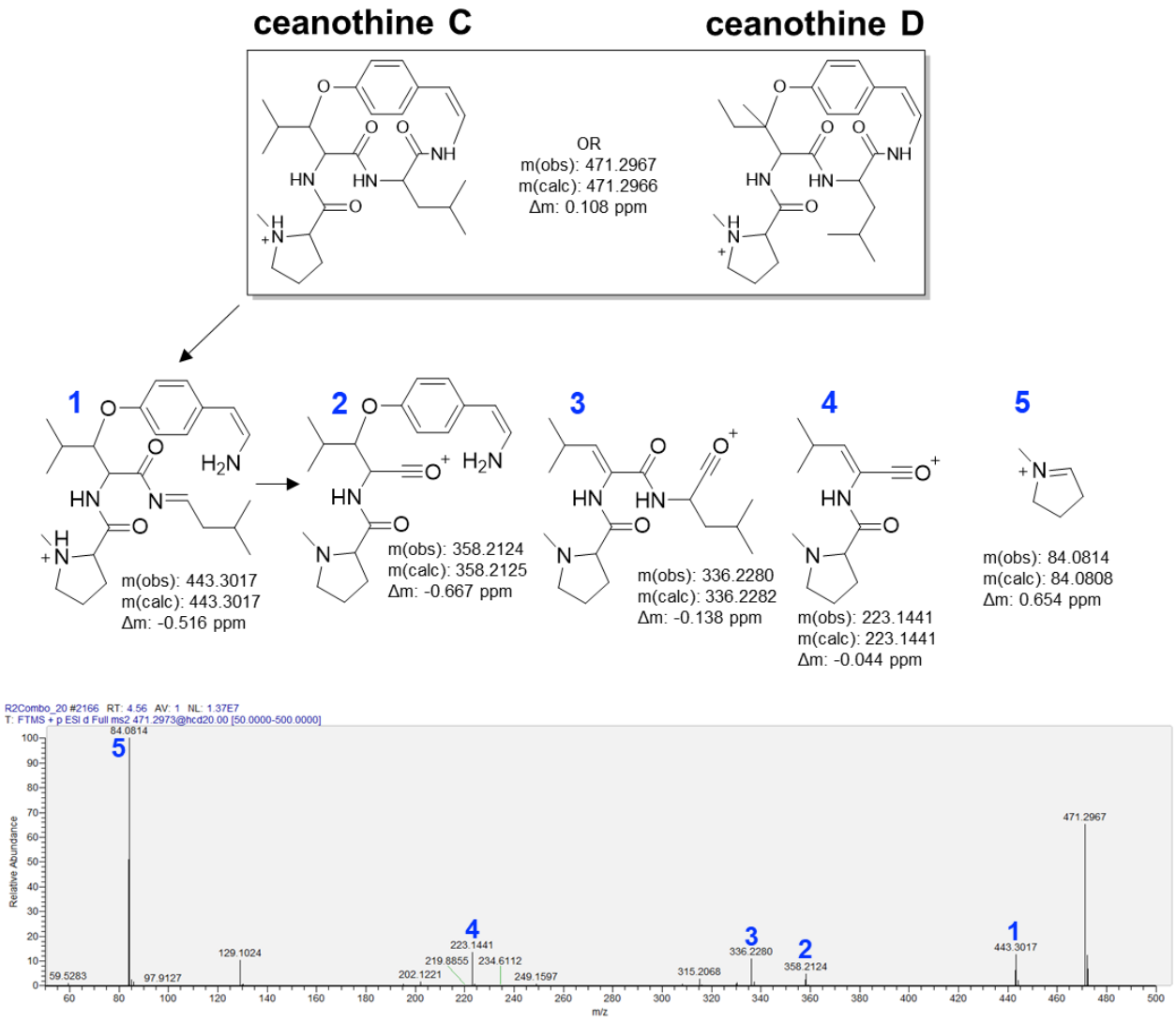


Fig. 6 MS/MS spectrum of a feature from *C. americanus* root extract. This feature is consistent with ceanothine C or ceanothine D.

Figure 7. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

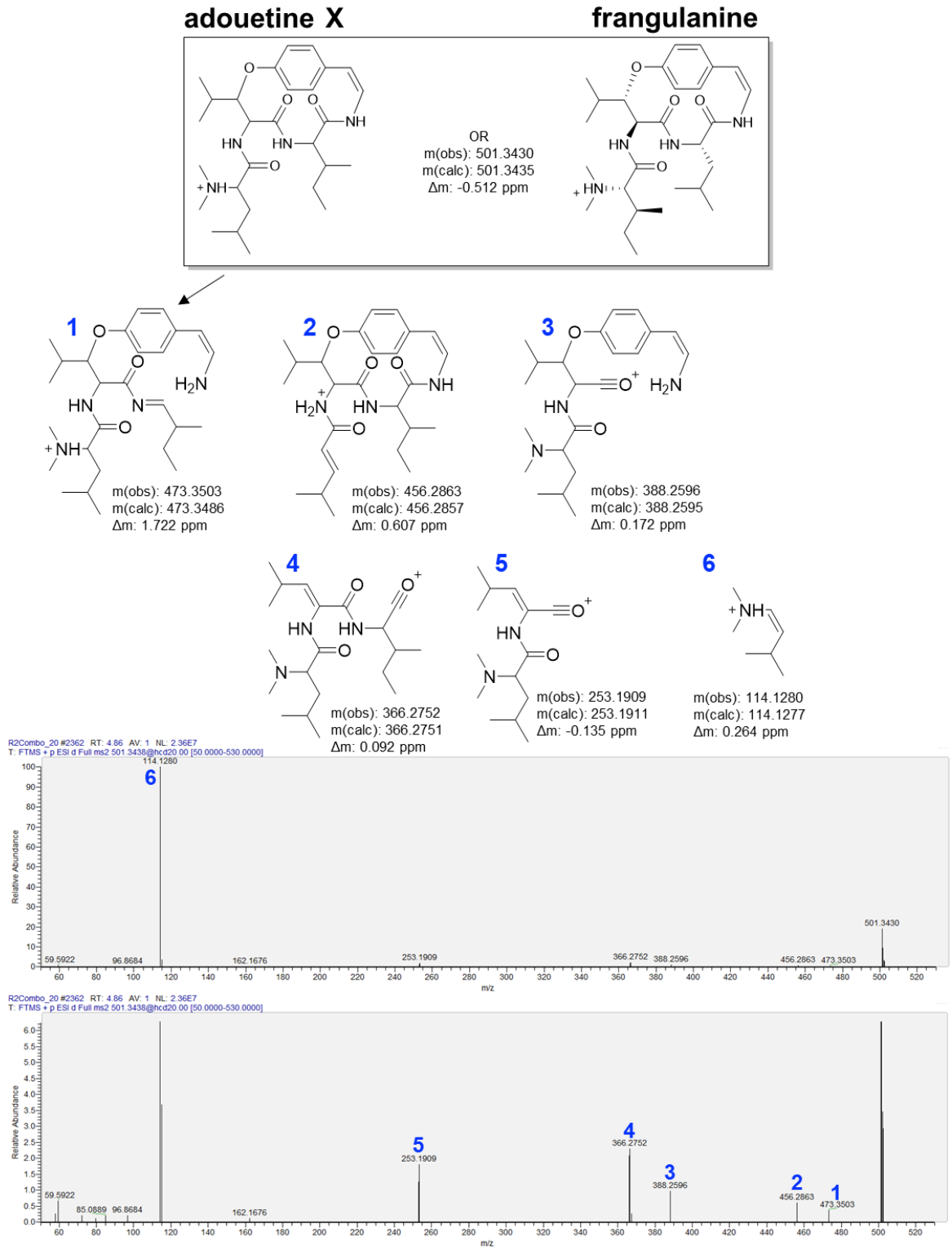


Fig. 7 MS/MS spectrum of a feature from *C. americanus* root extract. This feature is consistent with adouetine X or frangulanine.

Figure 8. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

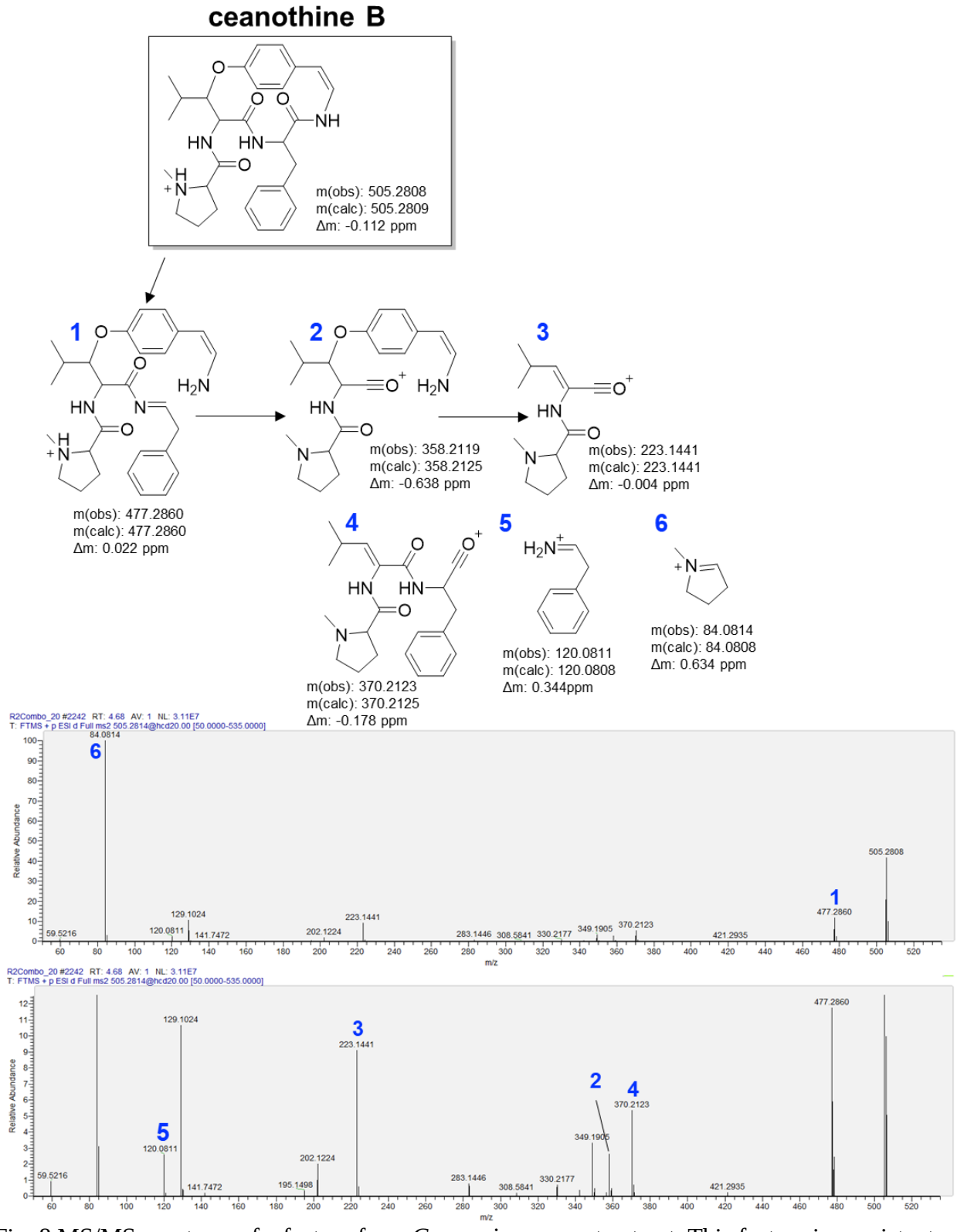


Fig. 8 MS/MS spectrum of a feature from *C. americanus* root extract. This feature is consistent with ceanothine B.

Figure 9. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

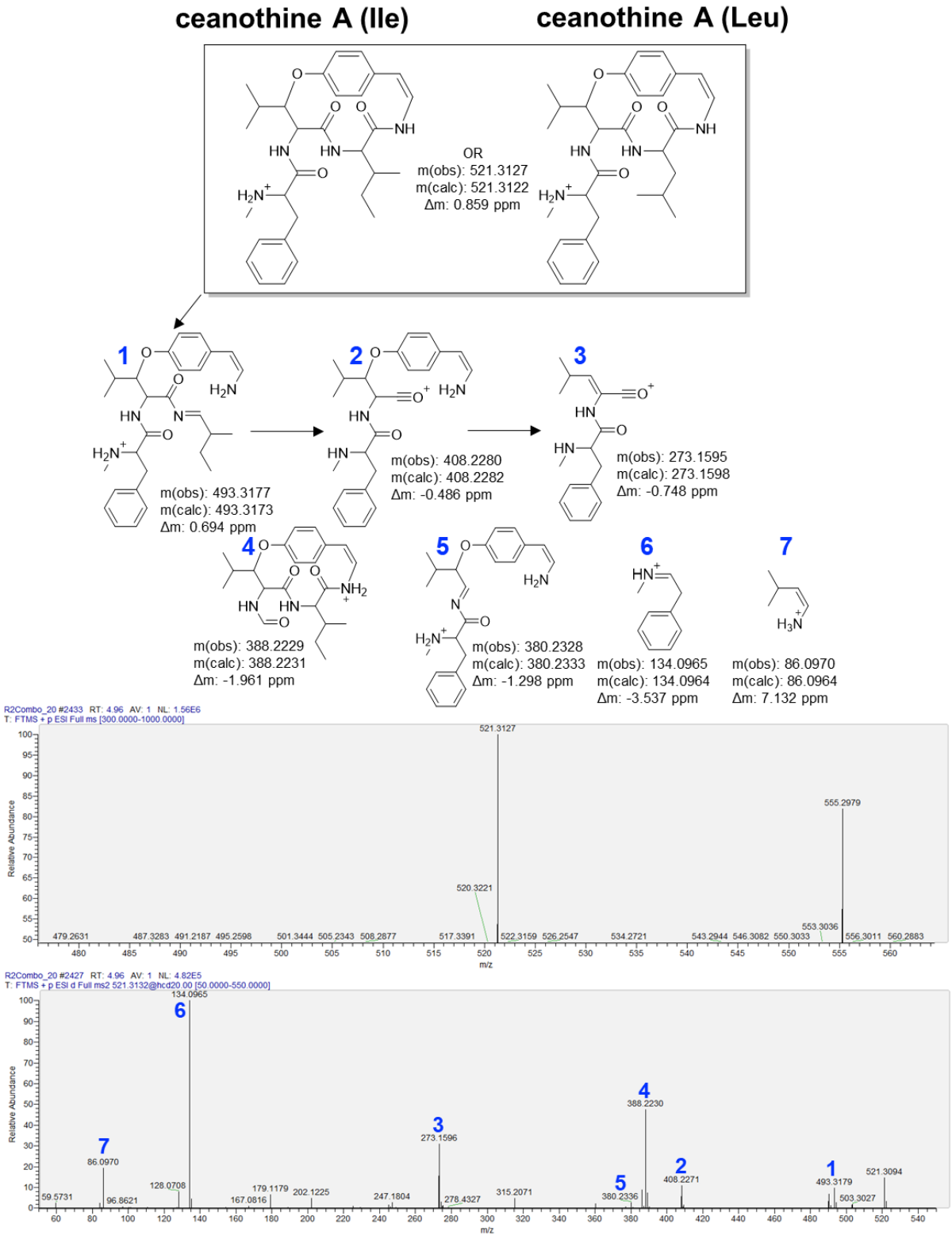


Fig. 9 MS/MS Spectrum of a feature from *C. americanus* root extract. This feature is consistent with ceanothine A.

Figure 10. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

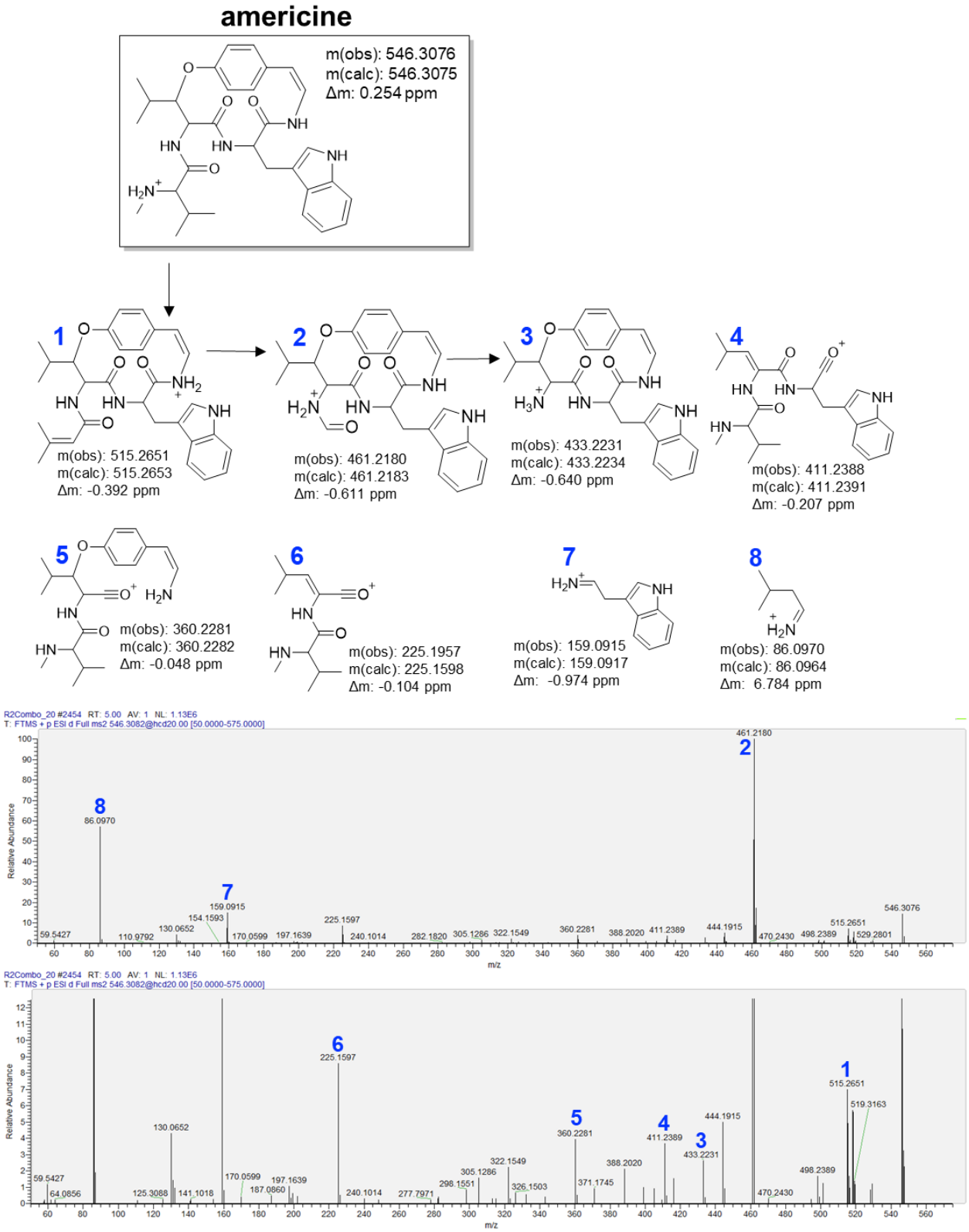


Fig. 10 MS/MS Spectrum of a feature from *C. americanus* root extract. This feature is consistent with americine.

Figure 11. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

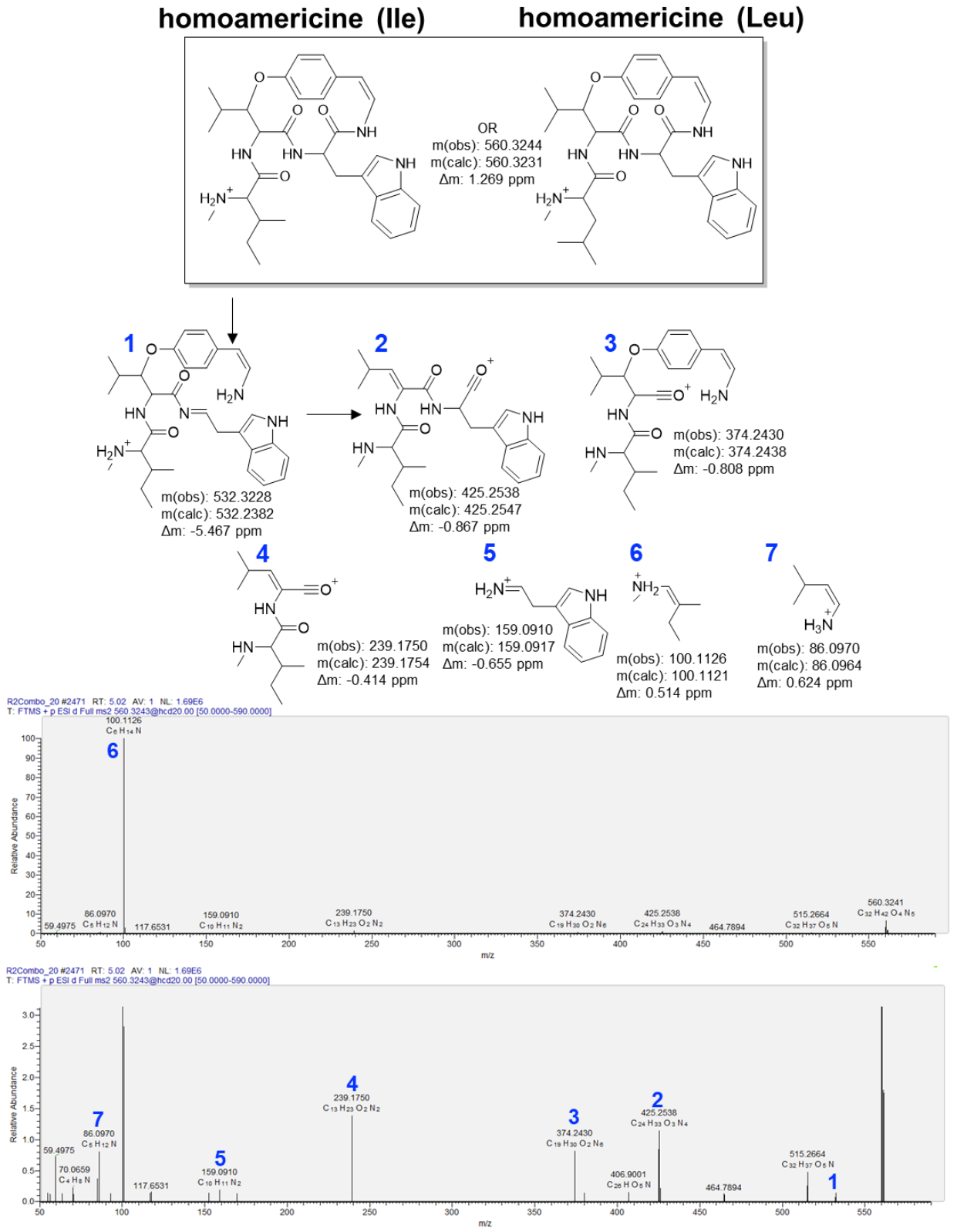


Fig. 11 MS/MS Spectrum of a feature from *C. americanus* root extract. This feature is consistent with homoamericine.

Figure 12. MS/MS Spectrum of a Feature from *C. americanus* Root Extract

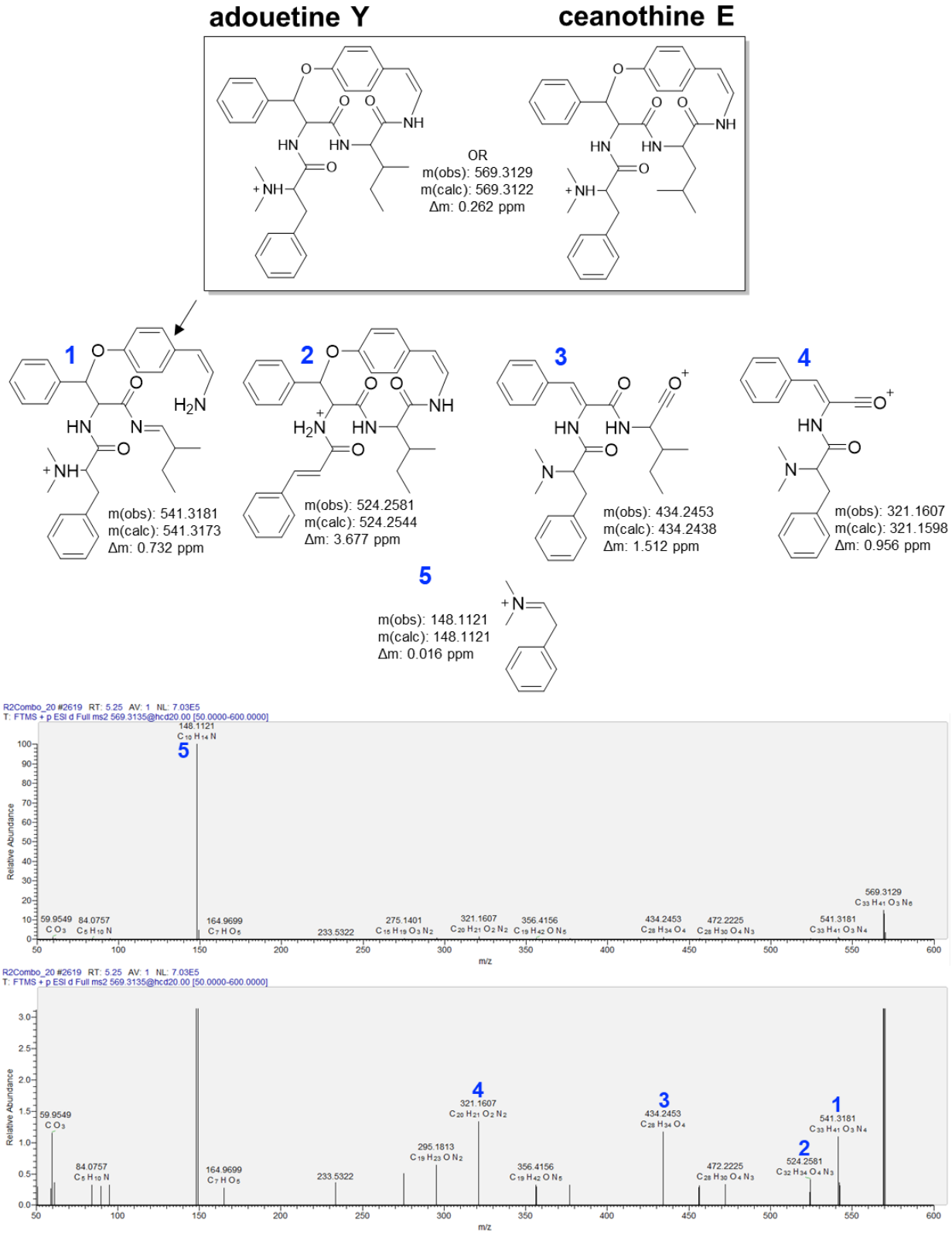


Fig. 12 MS/MS Spectrum of a feature from *C. americanus* root extract. This feature is consistent with adouetine Y and ceanothine E.

III.B. RNA Extraction and Analysis Results

Most of the RNA extraction samples yielded between one and 2.5 μg of RNA by the Biotum RNA quantification dye-based analysis, the results are shown in **Table 3**. Some of the samples stand out with negative absorbance values which indicated no RNA presence. The spectra drop analysis of the RNA helped to indicate initial purities. Examples of good and poor purity samples are shown in **Figures 13** and **14**. The ratios indicative of good quality RNA are a 1:1 ratio between 230 nm and 280 nm and a 2:1 ratio between 260 nm and 280 nm absorbances. Of the eight samples that were sent to the HTSF lab, five passed quality control tests and were sequenced. Sequencing yields for the three 5-inch plant samples were as follows: 21.1 million 2x100 paired end (PE) reads for leaf, 25.0 million 2x100PE reads for stem, and 22.5 million 2x100PE reads for root. The Trinity assembly made up of data from the three samples from the 5-inch plant yielded 67,599 total transcripts with 41,548 Trinity genes and a 42.30 % GC content. The leaf, stem, and root samples were also assembled separately. The root assembly alone generated 2189 transcripts that contained one of the putative cores. Sequencing and Trinity assembly data were deposited to the NCBI but will not be made available until publication under the accessions in appendix A.

Table 3. RNA Extraction Data

Sample	Type	Dye Abs	Total ug	Sample Name	UNC CQ Results	Plant Sample
Oct 11-12	Leaf	545056	2.320557	4L 10/11	Sequenced	5-inch height plants - From Prarie Nursery, Westfield, WI 53964
Oct 22-25	Stem	325854	1.290876	S4 10/25	Sequenced	
Oct 22-25	Root	502870	2.122393	R4 10/25	Sequenced	
Nov 2-6	Leaf	330787	1.314048	La 11/5	-	16-inch height plants - From Taylor's Nursery, Raleigh, NC 27610
Nov 2-6	Stem	498695	2.102781	S4 11/5	-	
Nov 2-6	Root	21542	-0.1386	-	-	
Nov 8-10	Leaf	646141	2.795395	L 11/8	Sequenced	
Nov 8-10	Stem	-34957	-0.404	-	-	
Nov 8-10	Root	57043	0.028161	-	-	
L4 8/23	Leaf	157450	0.470384	-	-	
S4 8/23	Stem	280178	1.059836	S4 8/23	-	
NL4 8/23	Leaf	232551	0.831088	NL4 8/23	Sequenced	

Figure 13. Poor RNA Purity Peak Example

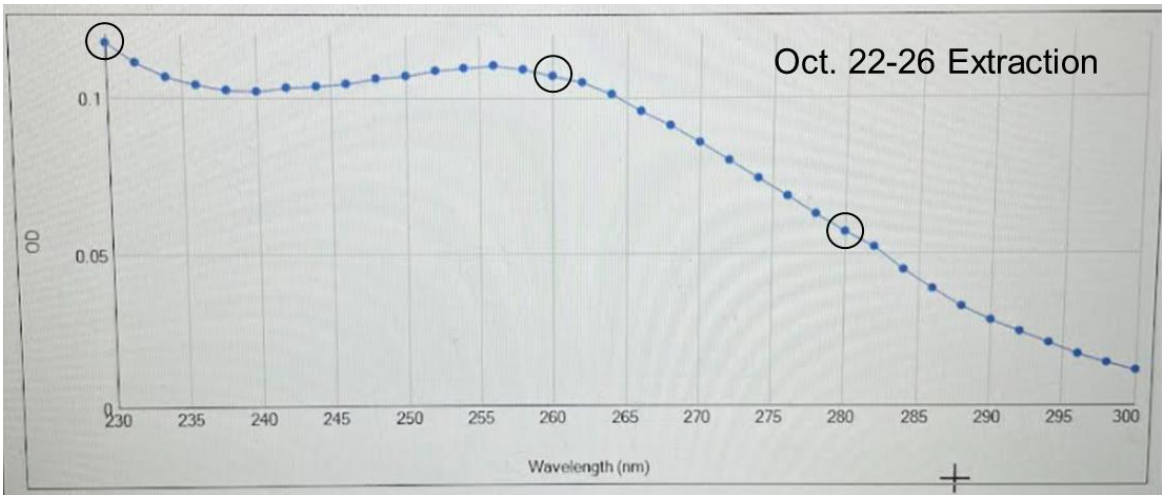


Fig. 13 This sample from the October 22-26 extraction is indicative of an impure RNA sample, where the absorbance at 230 nm is higher than both the 260 and 280 nm absorbances.

Figure 14. Good RNA Purity Peak Example

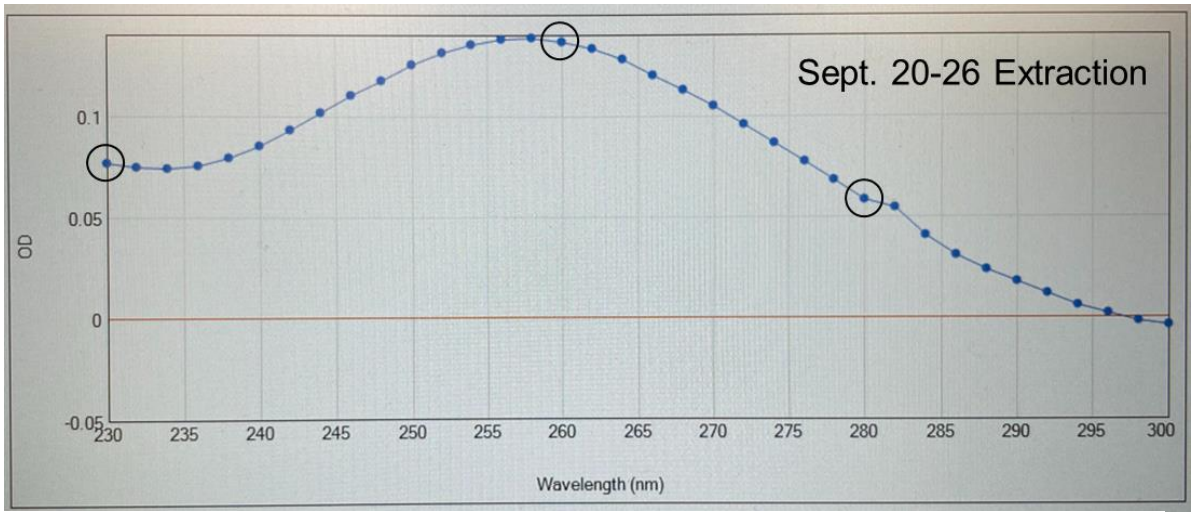


Fig. 14 This sample from the September 20-26 extraction is indicative of a RNA sample with good purity, where the absorbances at 230 and 280 nm are almost equal, and the 260 nm absorbance is higher than the other peaks.

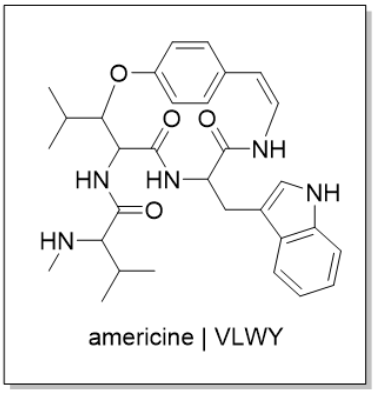
III.C. Bioinformatic Analysis Results

III.C.1. Stand-alone Precursor Discovery

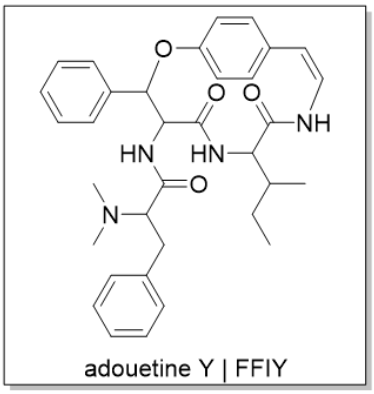
Precursor peptide discovery was completed by searching the transcripts for the hypothesized cores of the known compounds. Using the CD-Hit clustering approach described above, a total of 2189 precursor peptide transcripts that contained putative cores were identified. The 2189 transcripts were clustered into 1321 clusters. The cores in these transcripts directly correlated with the structures of the known cyclopeptide alkaloids we observed on our metabolomic analysis. The cluster which was selected for analysis was chosen because it contained a variety of core sequences with sequence similarity percentages above 60 but below 100 percent. These cores were cataloged from the transcripts shown in **Figure 15** below and correspond to known cyclopeptide alkaloids. The confirmation of hypothesized cores allowed for determination of the recognition sequence. Analysis of the core and surrounding sequences with multiple sequence alignment (MAFFT) is shown in **Figure 16** below, which highlights a highly conserved sequence found around the cores which was revealed to be SVDPSGN.

Figure 15. Stand-alone Precursor Confirmation.

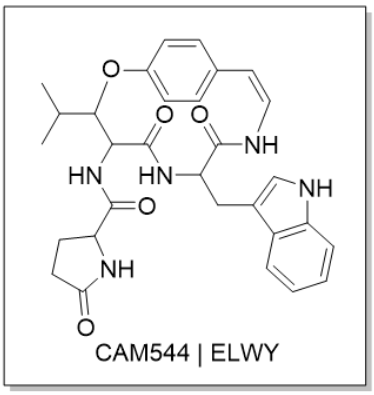
>Ceanothus_americanus_27101
MKTFFALFAFSSLLLLSSTITARKEPEYLKTVIENQPILEVLQGV
LDLITKSGNGGRVAKDISVDPSGNILWYHGNKNAEAKSKDDGEHK
DDFSVDPSGNELWYHDNKNAEAKSKADGEHKDDFSVDPSGNELWY
HDNKNAEAKSINPSGN**VLWY**HGNKNAEAKSINPSGN**VLWY**HGNKN
AEAKSKADGEHKDDFSVDPSGN**VLWY**HGKKIVELDSETIAEQVAK
AFSDAPSGTFILYHGNKNGESNSKASGEGAAKDSFVDPSVFDK



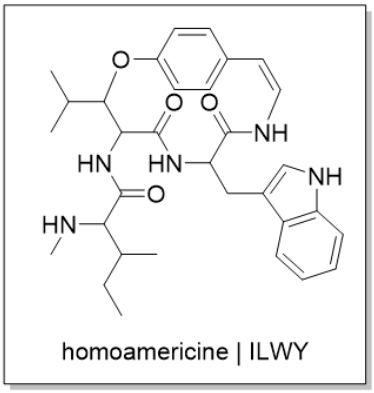
>Ceanothus_americanus_27097
PKKHFEFFEVDIKMKTFFALFAFSSLLLLSSTITARKEPEYLKTV
IENQPILEVLQGVLDLITKSGNGGRVAKDISVDPSGNILVYHGNK
NGEHVPKDASVDPSGNILLYHGNQNGKQVAKDFSVDPSGN
FFIYHNNKNAAAKSKADG



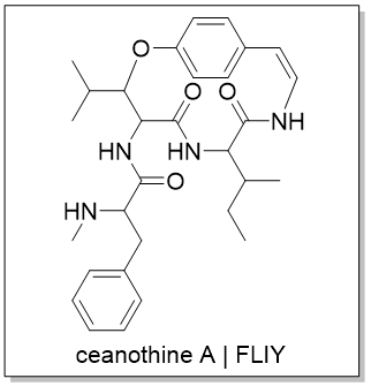
>Ceanothus_americanus_27105
SVDPSGN**ELWY**HDNKNAEAKSKADGEHKDDFSVDPSGN**ELWY**HDNK
NAEAKFLWYHGNKNAEAKSKADGEHKDDFSVDPSGNVLWYHGKKIV
ELDSETIAEQVAKAFSDAPSGTFILYHGNKNGESNSKASGEGAAKD
SFVDPSVFDK



>Ceanothus_americanus_27100
ILWYHGNKNAEAKSKDDGEHKDDFSTDPSGNFFLYHGNKIVELNSE
TNIAEQVAKAFSVAPSGTFIINLGNKMVSKM

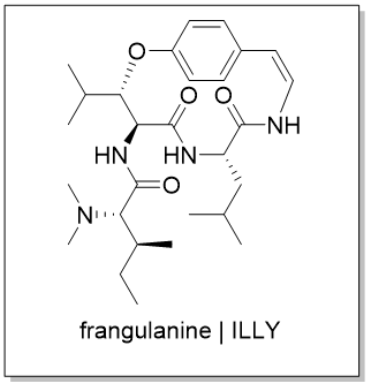


>Ceanothus_americanus_27108
MKTFFVLLAFSSLLLSRTITARKEPAEYAKTIMENQPIPVLEVIQGVL
DLIAKSSRNGGRLAKDLSVDPLSNWLIYRGNQNGENVAKDLKDLSVDP
SSN**FLIY**RGNHNGENAAKDLSVDPLSNWLIYRGNQNGENVAKDLKDLS
VDPSSN**FLIY**RGNHNEENVAKDLSVDPSSN**FLIY**RGNHNEENVAKDLS
VDPSSN**FLIY**RGNHNEENVAKDLSVDPSSN**FLIY**HGNKNADPNSKANG
EGVAKGISPDHHDQ



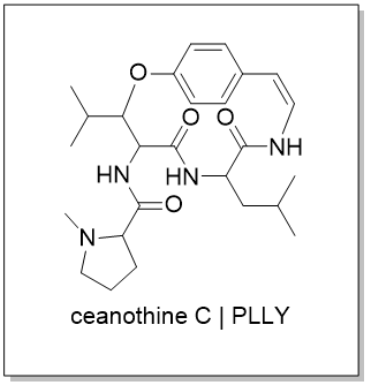
ceanothine A | FLIY

>Ceanothus_americanus_27111
GNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFFYHNNKNAVAKSK
ADEGELKDFSVDPSGNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNF
FFYHNNKNAVAKSKADEGELKDFSVDPSGN**ILLY**HGNKNAEVKSNTDG
KHKEDFSVDPSGNPLFYHGNKNDDFSVDPSGNPLFYHGNKNGEH



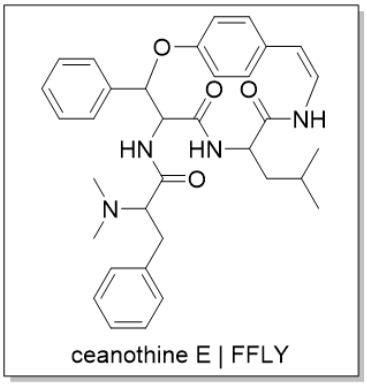
frangulanine | ILLY

>Ceanothus_americanus_27116
GNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFFYHNNKNAVAKSKA
DEGELKDFSVDPSGNILLYHGNKNAEVKSNTDGKHKEDFSVDPSGN**PLL**
YHDNKNDDFSVDPSGN**PLLY**HDNKNGEHKDGFSVDPSGNPLFYHGNKND
DFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGN**PLLY**HDNKNDDFSVDPS
GNPLFYHGNKNGEHKDGFSVDPSGNPLFYHGNKNGEHKDGFSVDPSG



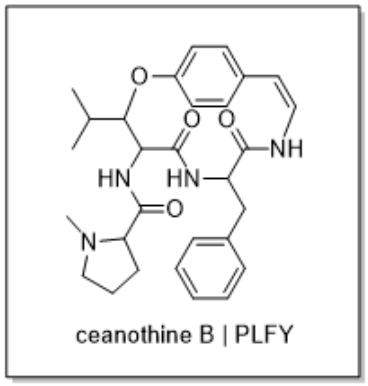
ceanothine C | PLLY

>Ceanothus_americanus_27100
ILWYHGNKNAEAKSKDDGEHKDDFSTDPSGN**FFLY**HGNKIVELNSETNI
AEQVAKAFSVAPSGTFIINLGNKMVSKM



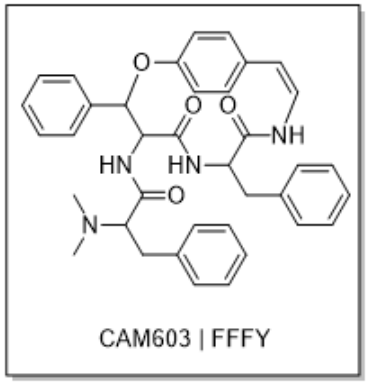
ceanothine E | FFLY

>Ceanothus_americanus_27112
GNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFFYHNNKNAVAKSKA
DEGELKDFSVDPSGNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFF
YHNNKNAVAKSKADEGELKDFSVDPSGNILLYHGNKNAEVKSNTDGKHK
EDFSVDPSGNPLLYHDNKNDDFSVDPSGNPLLYHDNKNGEHKDGFSVDP
SGNPLFYHGNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLLY
HDNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLFYHGNKNGE
HKDGFSVDPSG



ceanothine B | PLFY

>Ceanothus_americanus_27112
GNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFFYHNNKNAVAKSKA
DEGELKDFSVDPSGNFFFYHNNKNAVAKSKADEGELKDFSVDPSGNFFF
YHNNKNAVAKSKADEGELKDFSVDPSGNILLYHGNKNAEVKSNTDGKHK
EDFSVDPSGNPLLYHDNKNDDFSVDPSGNPLLYHDNKNGEHKDGFSVDP
SGNPLFYHGNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLLY
HDNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLFYHGNKNGE
HKDGFSVDPSG



CAM603 | FFFY

Fig. 15 Precursors were confirmed by searching for the putative cores contained in the transcripts and by metabolomic fragmentation data.

Figure 16. Weblogo Generated with *C. americanus* Precursor Peptides

Ceanothus americanus
(cyclopeptide alkaloids)

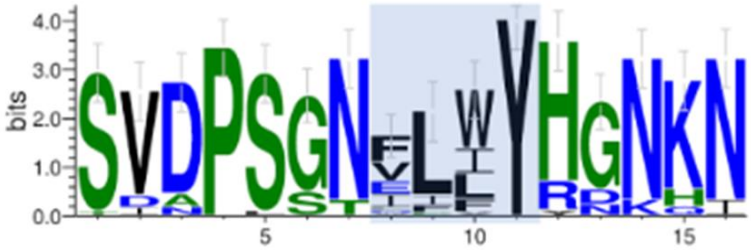


Fig. 16 The highlighted region is the predicted core motif and reveals a conserved Tyrosine, recognition sequences, and variable cores.

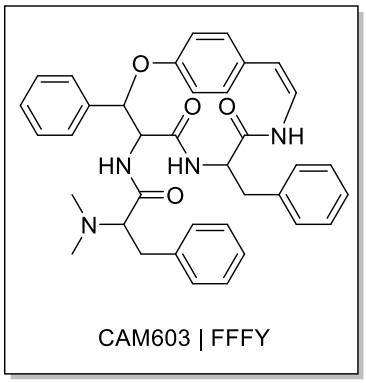
III.C.2. Precursor Peptide Guided Discovery

Using the newly uncovered precursor peptide sequences, the transcriptomic data was searched discover cores that were not known. This process yielded two unknown cores, F-F-F-Y and E-L-W-Y shown in **Figures 17** and **18** respectively. Analysis of our metabolomic data revealed two features consistent in both m/z and fragmentation to these new cyclopeptides predicted by transcriptomics. Work to extract and purify these compounds is underway. The method utilizes methanol-based extraction that is repeated three times overnight followed by removal of the excess methanol. The left-over residue is resuspended in basic water (pH 10) and put through eight rounds of liquid-liquid extraction using dichloromethane. After the liquid-liquid extractions, the sample is dried down and stored for purification by HPLC.

Figure 17. CAM603 Presence

A

>Ceanothus_americanus_27112
GN**FFFY**HNNKNAVAKSKADEGELKDFSVDPSGN**FFFY**HNNKNAVAKSKA
DEGELKDFSVDPSGN**FFFY**HNNKNAVAKSKADEGELKDFSVDPSGN**FFF**
YHNNKNAVAKSKADEGELKDFSVDPSGNILLYHGNKNAEVKSNTDGKHK
EDFSVDPSGNPLLYHDNKNDDFSVDPSGNPLLYHDNKNGEHKDGFSVDP
SGNPLFYHGNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLLY
HDNKNDDFSVDPSGNPLFYHGNKNGEHKDGFSVDPSGNPLFYHGNKNGE
HKDGFSVDPSG



B

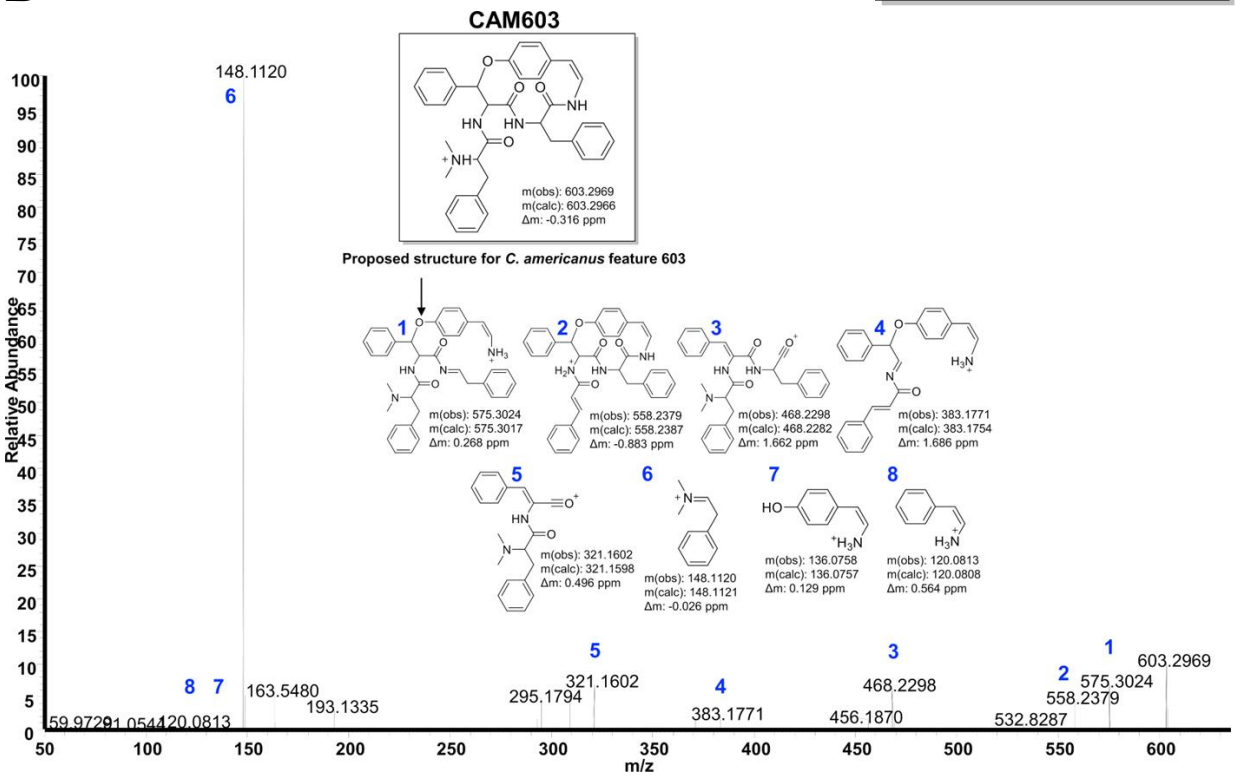
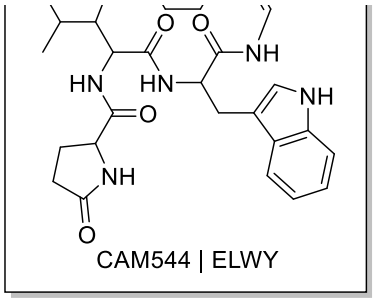


Fig. 17 (A) Transcript Data Showing Presence of an Unknown Core. (B) MS/MS spectrum of a feature from *C. americanus* root extract consistent with a cyclopeptide derived from the putative F-F-F-Y core.

Figure 18. CAM544 Presence

>Ceanothus_americanus_27105
SVDPSGN**ELWY**HDNKNAEAKSKADGEHKDDFSVDPSGN**ELWY**HDNK
NAEAKFLWYHGNKNAEAKSKADGEHKDDFSVDPSGNVLWYHGKKIV
ELDSETIAEQVAKAFSDAPSGTFILYHGNKNGESNSKASGEGAAKD
SFVDPSVFDK



B

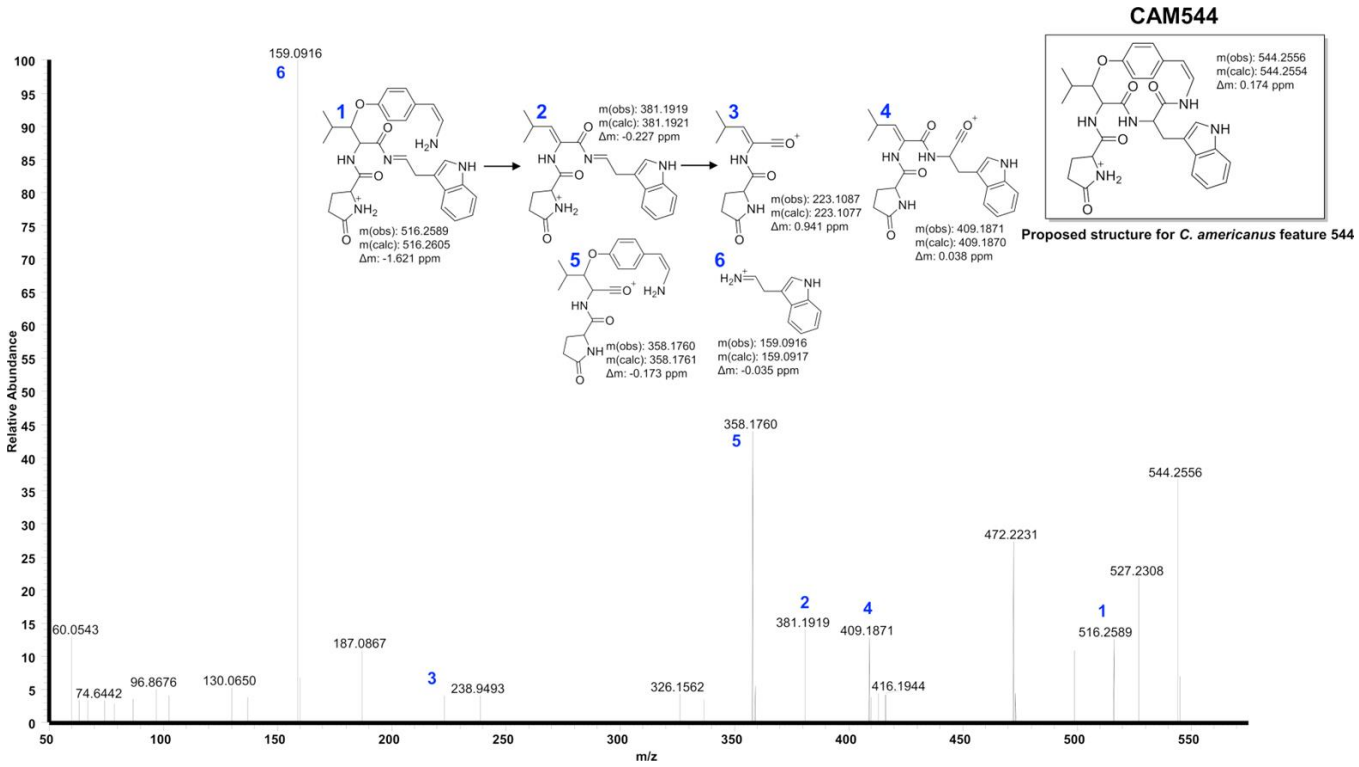


Fig. 18 (A) Transcript Data Showing Presence of an Unknown Core. (B) MS/MS spectrum of a feature from *C. americanus* root extract consistent with a cyclopeptide derived from the putative E-L-W-Y core.

III.D. Extraction and Purification Method Determination Results

Mass spectrum samples of each liquid extraction yielded the results shown in Figures **19 through 21**. Out of these experiments, the one that yielded the best signal to noise ratios while retaining high absorbance and intensity in the range our molecules are seen (4-6 minutes) was dichloromethane. Dichloromethane reduced the number of peaks outside of the four-to-six-minute runtime range while retaining signal intensity in that same range. From here will come resuspension of the dichloromethane-extracted residues and isolation attempts on high throughput liquid chromatography.

Figure 19. Dichloromethane Liquid-Liquid Extraction

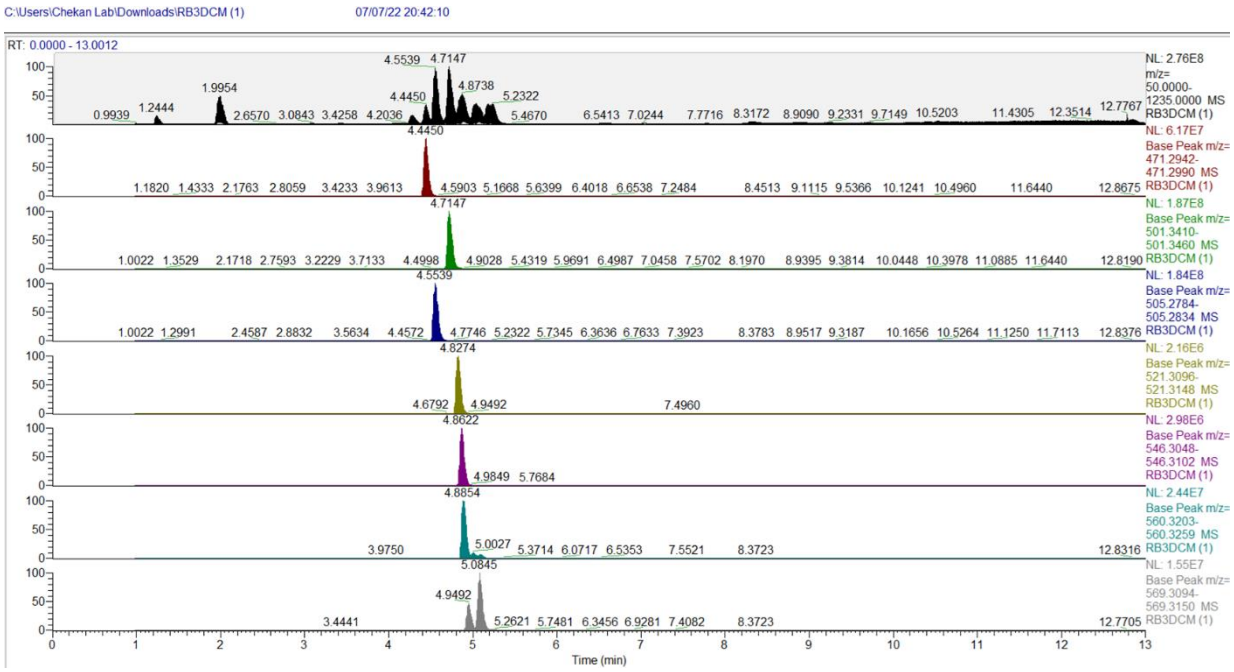


Fig. 19 Dichloromethane Liquid-Liquid Extraction. Dichloromethane reduced the peaks outside of the target runtime while retaining the masses of interest shown as base peaks.

Figure 20. Ethyl Acetate Liquid-Liquid Extraction

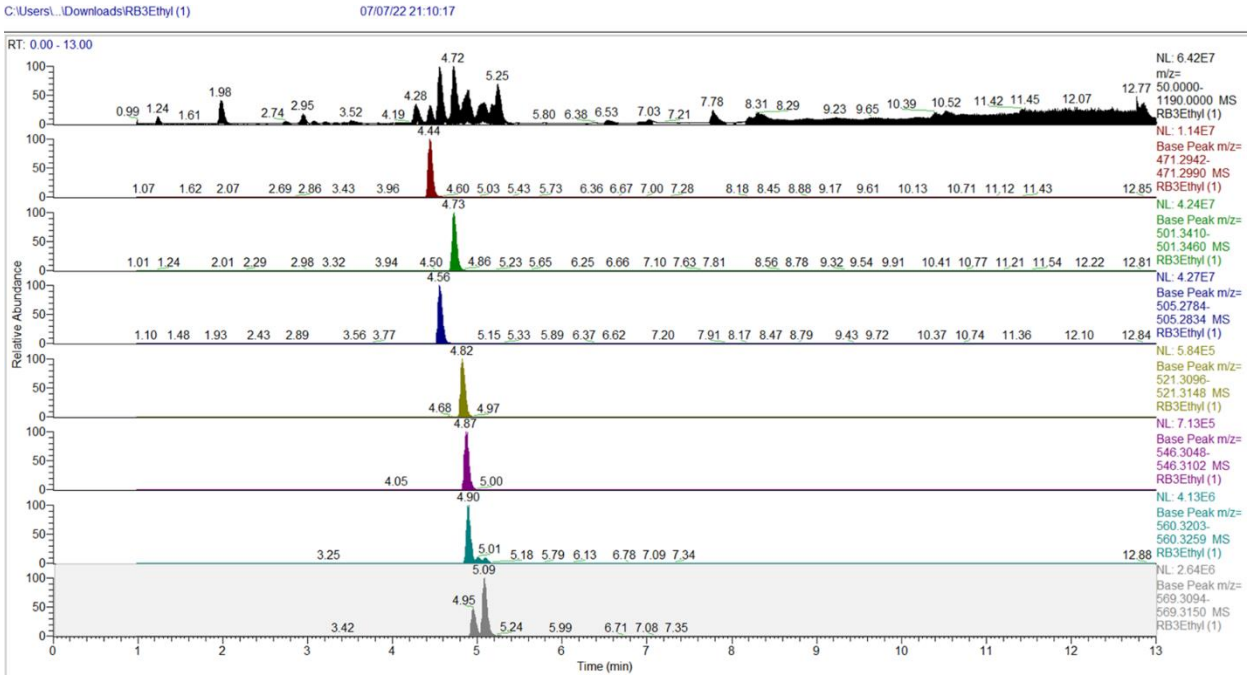


Fig 20 Ethyl acetate reduced the peaks outside of the target runtime but not to the level that dichloromethane had.

Figure 21. *n*-Butanol Liquid-Liquid Extraction

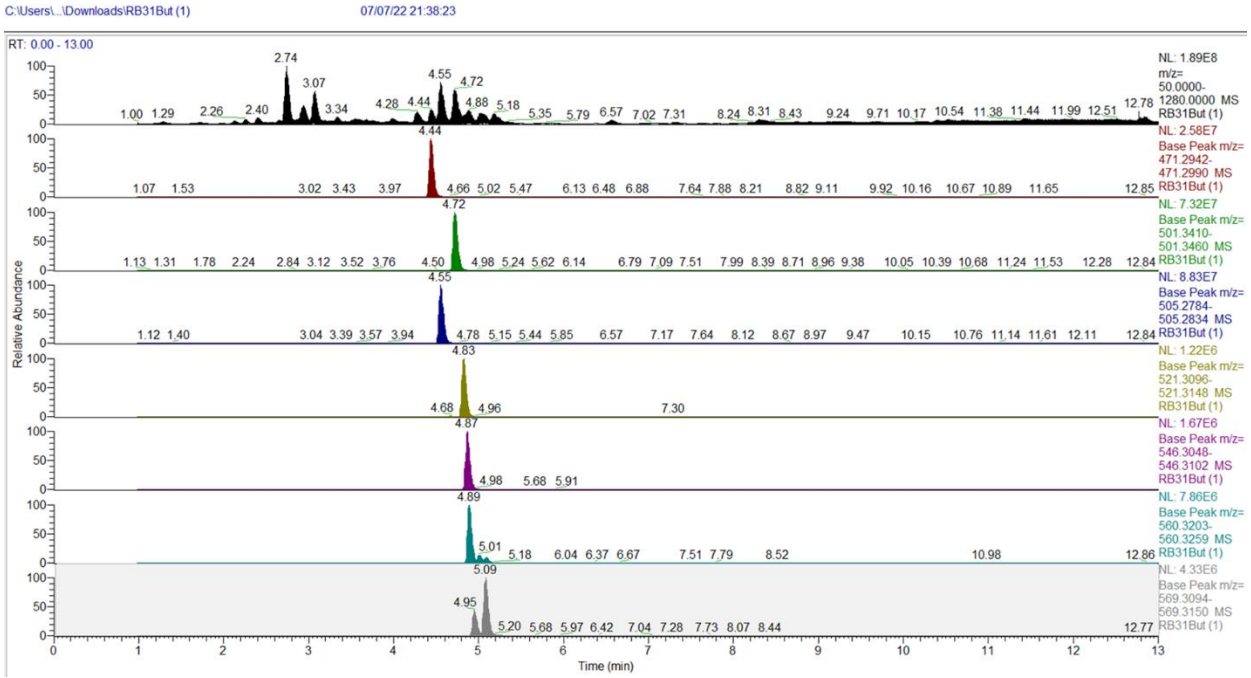


Fig. 21 The *n*-Butanol extraction reduced the peaks outside of the target runtime but is also reduced the relative intensity of the masses of interest.

III.E. Unsuccessful and Inconclusive Results

Of the RNA extraction methods that we used, only one of them successfully extracted quality RNA from the leaf, stem, and root samples. While the other methods worked for some tissues, they did not work for all of them. The Qigen RNA extraction kit did not yield any quality RNA over 16 different attempts. The phenol, chloroform, isoamyl alcohol-based extraction worked for the leaf tissues very well but did not successfully extract quality RNA from the stem or root across eight different attempts. However, this method allowed us to test different variables like incubation time in the -80 ℃ freezer, precipitation by lithium chloride at various concentrations and times, and ethanol wash step parameters. Any incubation in lithium chloride at or greater than two molar for more than 72 hours resulted in little to no RNA presence. This allowed us to find the optimal time for precipitation without denaturation to be between 24 and 36 hours. The ethanol wash step was critical in eliminating contaminants like left over salts, however, any percentage over 75% ethanol wash was likely to degrade the RNA. The ethanol would also denature the RNA if it was not removed in a timely manner. The RNA would be washed and would have to be immediately placed into the SPD 120 speed vac to remove the ethanol wash, or the RNA would degrade.

CHAPTER IV: CONCLUSIONS

IV.A. Cyclopeptide Alkaloids Found in *C. americanus* are RiPP Natural Products

The ability to hypothesize and eventually confirm cores by transcriptomics and fragmentation data opens the door for cyclopeptide alkaloid recognition as RiPP natural products because these natural products follow RiPP biosynthesis routes.

IV.B. A New Way to Confirm and Predict RiPP Natural Products

Utilizing RiPP biosynthesis motifs we can confirm and predict known and unknown RiPP natural products and reinforce that plants are a treasure trove for potential natural products. By determining the precursor peptides, we can link these precursors with the cores and determine the absolute sequences of the known cores. This approach allows for predictive analysis of potential putative cores like the two examples shown earlier.

IV.C. Future Endeavors

Future research endeavors will consist of biosynthetic pathway and gene determination in-order-to open the doors for bacteria or plant-based production of these compounds for research into bioactivity and enzyme engineering studies. There will also be more work on extraction and purification of these compounds while the biosynthetic pathways are determined. This work will open the door for studies searching for RiPP natural products across all plants.

REFERENCES

(1) Sim, M. C., Ho, C. L., Phang, S. M. A simple and effective method for RNA isolation and cDNA library construction from the brown seaweed Sargassum polycystum (Fucales, Phaeophyceae). *Journal of Applied Phycology*, **2013**, *25*(5), 1277–1285.

(2) Gruenwald J, Brendler T, Jaenicke C. PDR for Herbal Medicines. 1st ed. Montvale, NJ: Medical Economics Company, Inc., **1998**.

(3) Croteau, R., Kutchan, T. M., & Lewis, N. G. Secondary metabolites. *Comprehensive Biotechnology*, **2000**, 131–143.

(4) Ziegler, J., & Facchini, P. J. Alkaloid biosynthesis: Metabolism and trafficking. *Annual Review of Plant Biology*, **2008**, 59, 735–769.

(5) Pais, M.; Mainil, J.; Goutarel, R. The Adouetins X, Y and Z, Alkaloids of Waltheria Americanan L. (sterculia). *Annales Pharmaceutiques Francaises*, **1963**, 21, 139–146.

(6) *The Alkaloids Chemistry and Pharmacology*, Brossi, A. vol. 21, Academic Press, **1983**.

(7) Temal-Laib, T., Chastanet, J., & Zhu, J. A convergent approach to cyclopeptide alkaloids: Total synthesis of sanjoinine G1. *Journal of the American Chemical Society*, **2002**,*124*(4), 583–590.

(8) Goyal, M., Ghosh, M., Nagori, B. P., & Sasmal, D. Analgesic and antiinflammatory studies of cyclopeptide alkaloid fraction of leaves of Ziziyphus nummularia. *Saudi Journal of Biological Sciences*, **2013**, *20* (4), 365–371.

(9) Morel, A. F., Araujo, C. A., Da Silva, U. F., Hoelzel, S. C. S. M., Záchia, R., & Bastos, N. R. Antibacterial cyclopeptide alkaloids from the bark of Condalia buxifolia. *Phytochemistry*, **2002**, *61*(5), 561–566.

(10) Gournelis, D. C., Laskaris, G. G., & Verpoorte, R. Cyclopeptide alkaloids. *Natural Product Reports*, **1997**, *14*(1), 75–82.

(11) Chigumba, D. N., Mydy, L. S., de Waal, F., Li, W., Shafiq, K., Wotring, J. W., Mohamed, O. G., Mladenovic, T., Tripathi, A., Sexton, J. Z., Kautsar, S., Medema, M. H., & Kersten, R. D. Discovery and biosynthesis of cyclic plant peptides via autocatalytic cyclases. *Nature Chemical Biology* **2021**, *18*(1), 18–28.

(12) Servís, R. E., Kosak, A. I., Tschesche, R., Frohberg, E., & Fehlhaber, H. Peptide Alkaloids from Ceanothus americanus L. (Rhamnaceae). **1969**, *273*(23).

(13) Tan, N. H., & Zhou, J. Plant cyclopeptides. *Chemical Reviews*, **2006**, *106*(3), 840–895.

(14) U. Schmidt, M. Zäh, A. Lieberknecht, *J. Chem. Soc. Chem. Commun.* **1991**, 1002–1004.

(15) F. Teufel, J. J. Almagro Armenteros, A. R. Johansen, M. H. Gíslason, S. I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, *Nat. Biotechnol.* **2022**, DOI 10.1038/s41587-021-01156-3.

(16) Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3954855/ (accessed Sep 19, 2021).

(17) Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; Heel, A. J. van; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Burgos, M. J. G.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; Wezel, G. P. van; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; Donk, W. A. van der. New developments in Ripp discovery, enzymology and engineering. https://pubs.rsc.org/en/content/articlelanding/2021/np/d0np00027b (accessed Oct 8, 2021).

(18) Khusainov, R.; van Heel, A. J.; Lubelski, J.; Moll, G. N.; Kuipers, O. P. Identification of essential amino acid residues in the Nisin dehydratase nisb. https://www.frontiersin.org/articles/10.3389/fmicb.2015.00102/full (accessed Aug. 5, 2021).

(19) Sivonen, K., Leikoski, N., Fewer, D. P., & Jokela, J. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Applied microbiology and biotechnology*, **2010** *86*(5), 1213–1225. https://doi.org/10.1007/s00253-010-2482-x

(20) Tabudravu, J. N.; Morris, L. A.; Bosch, J. J. K.-van den; Jaspars, M. Axinellin C, a proline-rich cyclic octapeptide isolated from the Fijian marine sponge stylotella aurantium. https://www.sciencedirect.com/science/article/abs/pii/S0040402002008980 (accessed Aug 8, 2021).

(21) Sun, H., Wei, H., Wang, H., Hao, P., Gu, L., Liu, G., Ma, L., Su, Z., Yu, S. Genome-wide identification and expression analysis of the BURP domain-containing genes in gossypium hirsutum. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6615115/ (accessed Sep 21, 2021).

(22) Chigumba, D.N., Mydy, L.S., de Waal, F. *et al.* Discovery and biosynthesis of cyclic plant peptides via autocatalytic cyclases. *Nat. Chem. Biol.* **2021**, *18*, 18–28 DOI 10.1038/s41589-021-00892-6.

(23) Z. Zhang, Q. Shi, B. Wang, A. Ma, Y. Wang, Q. Xue, B. Shen, H. Hamaila, T. Tang, X. Qi, A. R. Fernie, J. Luo, X. Li, *Plant J.* **2022**, *109*, 1116–1133.

(24) Temal-Laib, T., Chastanet, J., & Zhu, J. A convergent approach to cyclopeptide alkaloids: Total synthesis of sanjoinine G1. *Journal of the American Chemical Society*, **2002**, *124*(4), 583–590.
(25) Lee, J.; Joullié, M. M. Total synthesis of the reported structure of ceanothine D via a novel macrocyclization strategy. *Journal of Chemical Science*, **2018**, *9*, 2432-2436. DOI 10.1039/C8SC00234G (accessed Oct 12, 2021).
(26) Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **2011**, *29*(7), 644–652.
(27) Software, B. I. A. Bioinformatics applications note. *Bioinformatics*, **2005**, *21*(8), 1745–1746.
(28) Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **2012**, *28*(23), 3150-3152.
(29) Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, *Bioinformatics*. **2010**, *26*, 680–682.

APPENDIX A: PUBLIC SUBMISSION DATA

C. amercanus Sequencing (NCBI)

BioProject: PRJNA840870

BioSamples: SAMN28561782 (Leaf), SAMN28561783 (Stem), SAMN28561784 (Root)

SRA: SRR19334864 (Leaf), SRR19334863 (Stem), SRR19334862 (Root)

Herbarium Sample (University of North Carolina at Chapel Hill, NC)

Ceanothus americanus

Underwood #01

NCU Accession 677816