EXPLORING CYSTEINE-RICH BOTANICAL BIOACTIVE PEPTIDES VIA MASS SPECTROMETRY

Nicole C. Parsley

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the School of Arts and Sciences.

Chapel Hill
2020

Approved by:

Leslie M. Hicks

Matthew R. Redinbo

Gary L. Glish

Marcey L. Waters

Bo Li

ABSTRACT

Nicole C. Parsley: Exploring Cysteine-rich Botanical Bioactive Peptides via Mass Spectrometry
(Under the direction of Leslie M. Hicks)

A rise in clinical multidrug-resistant microbes threatens global health, agriculture, and economy. Botanical natural product peptides are an underexplored source of antimicrobial therapeutics with novel and unique chemistries to infuse an increasingly dry drug pipeline. However, complex natural product matrices, inherent sequence variability, low abundance, and dynamic / circadian expression challenge the identification of novel antimicrobial peptides (AMPs): analytical methods are needed to explore the expansive chemical space offered by natural product AMPs.

Mass spectrometry (MS) drives primary natural product discovery, where modern instruments coupled with ultra-high performance liquid chromatography (UHPLC) can identify and resolve low abundance peptidyl species in complex mixtures with unprecedented limits of detection, resolving power, and mass accuracy. As such, the Hicks lab has developed PepSAVI-MS (**S**tatistically-guided bio**A**ctive **Pep**tides prioritized **Vi**a **M**ass **S**pectrometry), an adaptable method leveraging mass spectrometry, sensitive bioactivity assays, and statistical analysis to rapidly identify AMPs in natural product extracts.

Despite significant biochemical diversity, conserved features among AMPs can guide PepSAVI-MS identification of novel bioactive sequences. Antimicrobial peptides are often cysteine-rich; extensive disulfide formation among cysteine residues greatly enhances AMP stability and thus highlights the therapeutic attractiveness of this molecular class. Chemical derivatization strategies can be exploited to detect the presence of cysteine-rich peptides (CRPs) in complex matrices by mass shift analysis, where extracts before and after the reduction of

disulfides and alkylation of cysteine residues produces a characteristic mass shift observable by LC-MS. Mass shift analysis can be leveraged to identify and prioritize putatively bioactive mass species in a sea of botanical macromolecules and rich secondary metabolites typically obscuring analyses (Chapters 2-5).

Cyclotides are a class of head-to-tail cyclized, plant-derived CRPs with diverse and potent intrinsic bioactivities. Examination of *Viola* spp. via PepSAVI-MS and mass shift analysis has revealed new anticancer and antifungal bioactivities of a known cyclotide, cycloviolacin O8 (Chapter 3), and six novel antibacterial cyclotides, cycloviolacins I1-6 (Chapter 4). Elements of the discovery process can be modified to improve AMP identification, including alternative botanical tissue sources (Chapter 7), target pathogens (Chapter 3), mass spectral fragmentation techniques (Chapter 3), and bioactivity assay formats (Chapter 6). Furthermore, cyclic peptides recalcitrant to MS/MS fragmentation often challenge MS-based sequence characterization; as such, common cyclotide MS/MS "fingerprint" ions are determined and rapidly provide molecular information for novel cyclotide mass species prior to full sequence elucidation (Chapter 5). Herein, robust analytical methods support the exploration of complex botanical extracts and expand the biochemical repertoire of natural product AMPs.

To the hoofed and wing-ed beasts that clawed their way out of my nightmares to haunt my home.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMPs	Antimicrobial peptides
AMR	Antimicrobial resistance
ATCC	American type culture collection
cyO2	Cycloviolacin O2
cyO8	Cycloviolacin O8
cyI1-6	Cycloviolacins I1-6
Da	Daltons
ESKAPE	*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacter* species
HPLC	High performance liquid chromatography
IC	Inhibitory concentration
IDA	Information-dependent acquisition
LC-MS	Liquid chromatography-mass spectrometry
m/z	Mass to charge ratio
MDR	Multi-drug resistance
MHB	Mueller Hinton broth
MIC	Minimum inhibitory concentration

MOAs	Mechanisms of action
MPA	Mobile phase A
MPB	Mobile phase B
MS	Mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MW	Molecular weight
N2(l)	Liquid nitrogen
O.D.	Optical density
PepSAVI-MS	Statistically guided bioactive peptides prioritized via mass spectrometry
PMMA	Polymethylmethacrylate
PP	Polypropylene
PS	Polystyrene
PTMs	Post-translational modifications
PVPP	Polyvinylpolypyrrolidone
RFU	Relative fluorescence units
RiPPs	Ribosomally-synthesized, post-translationally modified peptide natural products
RPLC	Reversed phase-liquid chromatography

SCX	Strong cation exchange
SSPS	Solid-phase peptide synthesis
TIC	Total ion chromatogram

CHAPTER 1: Introduction

1.1 Modern Antimicrobial Resistance

Rapidly emerging antimicrobial resistance (AMR) poses a significant threat to public health and is responsible for substantial global economic burden. Common bacterial and fungal infections are becoming untreatable as new mechanisms for AMR develop, accelerated by the misuse and overuse of antimicrobial drugs.[1, 2] The U.S. Centers for Disease Control and Prevention (CDC) estimate that > 2.8 million antibiotic-resistant infections occur in the United States every year, in which 35,000 cases are fatal.[3] Decreased productivity in U.S. health care systems due to AMR costs a staggering $35 billion in economic loss every year.[1, 4] Unless measures are taken to reduce and prevent further AMR, it is predicted that 10 million deaths globally will result from AMR annually by the year 2050, costing a global sum of $100 trillion.[5-7]

We are already living in a "post-antibiotic" era.[3] Emerging multi- and pan-drug resistant microbes evade conventional therapeutics and challenge the effective treatment of nosocomial infections. In particular, the *ESKAPE* pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) are responsible for an increase in clinical multidrug resistant infections associated with a higher risk of mortality and escalating health care costs.[8, 9] Named for their ability to "escape" the effects of antimicrobial drugs,[10] the *ESKAPE* pathogens, consisting of both Gram positive and Gram negative bacteria, employ diverse resistance strategies [11] and are the focus of extensive AMR research efforts.

Development of resistance is an ancient and natural evolutionary process[12]; however, AMR is accelerated by the global misuse of antimicrobials at the clinic/hospital level, the high

accessibility of unregulated ("over-the-counter") antimicrobials, and non-surveillance of AMR development in developing nations.[13] An estimated 30-50% of prescribed antibiotic treatment regimens, including drug choice and treatment duration, are incorrect or unnecessary.[1, 14] Once developed, microbial resistance can spread rapidly throughout medical facilities via horizontal gene transfer,[15] exacerbated by aerosolized bacteria [16] and bacterial vectors, such as cockroaches, spreading common nosocomial pathogens (i.e. *S. aureus*, *P. aeruginosa*, *Escherichia coli*, *K. pneumoniae*, and carbapenem-resistant Enterobacteriaceae).[17] Furthermore, agricultural processes contribute significantly to global AMR, where the excessive use of antimicrobial treatments to support and sustain crop and livestock vitality has increased the prevalence of multidrug resistant pathogens in the food chain. Alarmingly, 80% of antibiotics sold in the United States are used to promote growth and prevent microbial infection in healthy livestock.[4, 18] Animal-based foods are the source of most foodborne microbial diseases, and are thus major outlets for spreading multidrug resistant genes.[19]

Antimicrobial stewardship strategies have been proposed to mitigate AMR through increased infection cure rates and decreased AMR development, adverse effects, treatment failures, and hospital burden.[20] Analysis of local antibiotic use/resistance data in hospitals and antibiotic cycling coupled with public education can drive responsible antibiotic usage; however, care must be taken to not compromise patient care.[21] While antimicrobial stewardship offers tremendous promise towards the regulation of AMR, developing and implementing effective stewardship programs requires extensive collaboration among medical professionals[22-25] and local/state public health officials.[26, 27] Although AMR will occur naturally, antimicrobial stewardship practices may slow or prevent widespread AMR; as such, until antibiotic stewardship programs are developed and standardized, the prevalence of clinical and agricultural AMR and its

implications for morbidity, mortality, and economy, necessitates novel antimicrobial leads to combat multidrug resistant pathogens.

1.2 Shortcomings of the Drug Discovery Pipeline

Antimicrobial development and commercialization was one of the most important and defining features of the twentieth century. In the early 1900's "pre-antibiotic" era, infectious disease spread by highly virulent microorganisms caused significant morbidity and mortality[28]: pneumonia deaths soared to 40%,[29, 30] 70% of amputations during World War I resulted from untreatable wound infections,[29, 31] and the average human life expectancy, in a time when infectious disease was the leading cause of death, was a meager 47 years.[32] The advent of the first antibiotics, i.e. the organoarsenic Salvarsan (1910), sulfonamide Protonsil (1935), and β-lactam penicillin (1945), immediately and dramatically improved quality of life and laid the foundation for continuing antimicrobial drug discovery efforts.[33] Subsequent years (1950-1970) witnessed the "Golden Age" of antibiotic discovery in which pharmaceutical companies, motivated by facile microbial-sourced antibiotic identification and high profits, produced over half of the antimicrobial drugs used today.[34]

Despite the clinical success of antimicrobial therapies, the drug pipeline has run dry initiating a modern "pre-antibiotic" era. Practical and financial factors have contributed to the modern dearth of viable antimicrobial therapeutic options; resistance to an antimicrobial evolves typically < 10 years[35] (sometimes even as few as two)[36] after its release to the market, whereas new drug candidates require at least 10-12 years to transcend the drug pipeline to clinical relevance.[37] Further disincentivizing industry involvement, the identification of novel chemistries after the initial flourish of antibiotic discovery presented significant analytical and biological challenges – the latter particularly in overcoming penetration barriers and active

transport mechanisms (e.g. efflux pumps) in difficult-to-treat Gram-negative pathogens.[38] As such, most "new" antibiotics are in fact chemically-modified forms of existing drugs, with few new classes of antibiotics discovered in the last 50 years (e.g. teixobactin[39]). Success rates of clinical drug development reveal that only 20% of infectious disease products entering into phase 1 human clinical trials will ever be approved for human use.[40] Additionally, drugs that treat chronic conditions generate significantly more revenue than those that treat acute ailments (e.g. infectious disease), and are thus more attractive targets for pharmaceutical interests. As infectious disease is responsible for one-third of present day global mortality, innovative solutions are required to infuse the drug pipeline with novel antimicrobial compounds.

1.3 Bioactive Peptides for Antimicrobial Therapy

Bioactive peptides represent a vast, underexplored source of novel chemistries with tremendous therapeutic potential. Peptide-based compounds exhibit substantial sequence/structural diversity, post-translational modifications (PTMs), and unique mechanisms of action and may serve to complement traditional small molecule therapeutics as antimicrobial agents.[41] The molecular size and chemical diversity of peptides can be leveraged to achieve specific advantages over other drug classes; peptidyl compounds often exhibit high affinity/specificity for a molecular target, reducing toxicity and off-target effects,[41, 42] and can participate in the disruption of protein-protein interactions that elude small molecules.[43, 44] Furthermore, peptides may offer tissue-penetration, used alone as innately bioactive molecules or in mixtures to enable the delivery/targeting of other therapeutic payloads.[45, 46] Additionally, established biosynthetic (e.g. bacterial/yeast expression systems) and synthetic (e.g. solid-phase peptide synthesis) methods are available to produce diverse peptide-based compounds on a

commercial scale, where reduced production costs elevate peptides as attractive drug candidates.[41, 47, 48]

Inherent instability *in vivo* and low oral bioavailability has historically thwarted advances in therapeutic peptide development. Extreme pH environments and metabolism via digestive enzymes in the gastrointestinal (GI) tract (e.g. pepsin)/secreted by luminal bacteria hydrolyze amide bonds and rapidly destabilize/degrade proteinaceous compounds.[49, 50] Intestinal absorption is challenged by the molecular size and chemical properties of peptide drugs.[44] Accordingly, most clinically available protein-based drugs are delivered via intramuscular, subcutaneous, or intravenous injection; however, oral delivery of medicines is preferable regarding patient compliance, ease of administration, and production cost.[44, 50]

Advances in peptide synthesis techniques may alleviate peptide instability *in vivo* and improve peptide pharmacologic performance/facilitate oral delivery. Stapling via a hydrocarbon "brace" can lock peptides into a bioactive structure, e.g. α-helical proteins, thus improving stability and resistance to proteolysis while enhancing target specificity and cell penetrating properties.[51, 52] Cyclization through side chains (e.g. disulfides), backbone interactions (e.g. "head-to-tail cyclization"), side chain to backbone linkages, or via artificial linkers may enhance peptide stability by limiting proteolysis and conformational entropy while preserving biological function.[53, 54] Additionally, peptides can be grafted into highly stable, cyclic molecular scaffolds, where native amino acid segments are interchanged with peptides exhibiting desired bioactivities.[55]

1.4 Natural Product Peptides as Lead Compounds

Antimicrobial peptides (AMPs) are ancient defense molecules; AMPs have offered a competitive advantage for countless biological species throughout history, evolving in potency and complexity alongside organism development. Expressed by all forms of life, AMPs demonstrate broad functions;[56, 57] > 2600 naturally occurring AMPs have been identified among AMP databases.[58, 59] In higher organisms, ribosomally-synthesized AMPs are fundamental to innate immune response, defending against microbial and viral pathogens, whereas prokaryotes may offensively secrete AMPs to secure resources and territory.[56] Antimicrobial peptides often exhibit broad spectrum activity and may function only in synergism as AMP combinations or with small molecule antibiotics.[57, 60] AMPs often achieve bioactivity via the insertion of an induced or permanent amphipathic motif into a target membrane – effecting physical disruption and subsequent permeabilization.[61, 62] Target specificity is largely dictated by membrane composition [63]; the membranolytic mechanism of action (MOA) employed by AMPs has historically prevented the development of widespread AMP resistance in nature, requiring remodeling of basic membrane structure.[56] Furthermore, intracellular MOAs have been identified, where the translocation of AMPs across pathogen membranes enables the disruption of key cellular process, e.g. DNA/RNA and protein synthesis.[56, 61, 64] With the recent surge of clinical multidrug resistant pathogens, AMPs have become increasingly attractive therapeutic candidates with broad spectrum activity and novel MOAs, as standalone drugs or in combination therapies.

Accordingly, numerous commercially-available drugs are derived from natural sources or are a semi-synthetic derivative of natural products.[65] Bioactive peptides have debuted in the clinic as promising therapeutic agents, where peptide antibiotics such as vancomycin are listed on the

World Health Organization's List of Essential Medicines.[66] Approved in 1971, triple antibiotic ointment (brand name Neosporin) is a Neomycin/polymyxin B/bacitracin broad-spectrum antibacterial topical treatment, where polymixin B and bacitiran are peptide antibiotics. Colistin is a 60 year old antibiotic commonly used as the 'last-line' treatment for numerous multi-drug resistant Gram-negative infections.[67] Despite the therapeutic value peptide-based drugs present, however, only seven AMPs are currently approved by the Food and Drug Administration (FDA),[68, 69] with nine and 27 AMPs in preclinical and clinical trials, respectively.[70] Interestingly, all FDA-approved AMPs are either bacterially-derived or semi-synthetic glycopeptides, though AMPs in clinical trials have broad origins – from mammals to insects, fungi, and marine sponges and snails.[69, 71] Though the limitations of peptide-based drugs often discourage natural product AMP drug development, this chemical class should be pursued for the broad activity spectrum, reduced occurrence of resistance, and synergistic potential it presents.

1.5 Botanical Antimicrobial Peptides

Plants are sessile organisms whose survival necessitates the expression of abundant secondary metabolites, enabling environmental adaptations and protection against pathogens.[72, 73] As such, botanical peptides are a rich source of novel chemistries with potent and diverse biological activities, though plant AMPs are greatly underrepresented in the clinic. Antimicrobial peptides generated through immune response pathways can be constitutively expressed or induced upon pathogen stress/environmental signals.[74] Ranging in size from two (e.g. cyclic dipeptides in green tea[75]) to > 50 residues (e.g. defensins[76, 77]), plant AMPs exhibit significant sequence and secondary/tertiary structural diversity. However, certain characteristics are commonly seen among plant AMPs, such as positive charge (facilitating interactions with negatively-charged phospholipids on target membranes), significant

hydrophobicity/amphipathicity (e.g. amphipathic α-helix), and enhanced stability through extensive disulfide bonding.[78, 79] Numerous plant AMP families have been characterized (e.g. defensins, knottins, thionins, and lipid transfer proteins[76]), and are grouped by similarities in sequence, number and spacing of cysteine residues, and/or tertiary structure.[78, 80]

1.6 Sources of Botanical Antimicrobial Peptides

Traditional medicines are a rich source of bioactive compounds, where botanical remedies have been a mainstay of African[81, 82] and Eastern[83] ethnomedicines for centuries; it is estimated that 80% of the World's emerging populations rely on traditional medicines[81] Historically, essential oils, extracts, and poultices were used for the treatment of microbial and viral infections (cough, cold, fever), inflammatory disease (skin disease, migraine) and oxidative stress (cancer, hypertension).[84, 85] In some instances, modern medicine has concurred with traditional usage, where analyses of the chemical constituents present in traditional medicinal plants have revealed abundant bioactive secondary metabolites with potent bioactivities.[86-88] The expansive biochemical repertoire presented by traditional botanical medicines encourages further exploration and evaluation of these species as putative clinical therapeutics with potent and diverse biological activities.

1.7 Methods for Natural Product Discovery

Bioassay-guided fractionation is a long established technique [89, 90] where a complex bioactive sample is subject to iterations of activity testing, fractionation, and subsequent reanalysis of activity until the bioactive constituent(s) is(are) identified.[91] Sample separations are primarily performed via liquid chromatography (LC), though other separation methods may be necessary (gel/thin layer/gas chromatography),[92] and analysis of fraction constituents is achieved with mass spectrometry (MS) or spectroscopic methods (e.g. nuclear magnetic resonance).[91] For decades,

bioassay-guided fractionation has been leveraged to identify the bioactive constituents in botanical natural product extracts with activity against microbial pathogens, including Gram-positive/negative bacteria and fungi,[93-97] allows for a direct measure of bioactivity, and can be tailored to any bioactivity assay. However, this strategy often favors the re-identification of highly abundant/active and previously-characterized compounds and requires significant sample quantities, necessitating the development of novel and sensitive analytical methods for the discovery of less abundant bioactive constituents.

Genome mining is a bioinformatics strategy wherein the genomes of sequenced organisms are 'mined' for biosynthetic gene clusters to predict novel chemical species. Targeted biosynthetic gene clusters are identified via homology to known, classified gene clusters producing enzymes that participate in the non-ribosomal assembly of bioactive secondary metabolites.[98] Sophisticated computational algorithms are used to predict biosynthetic gene clusters, e.g. BiosyntheticSPAdes[99] and antiSMASH[100], and the structures of putative bioactive compounds resulting from tailoring enzymes, e.g. PRISM.[101] Biosynthetic gene clusters can be cloned and heterologously expressed in biological systems to produce novel, putatively bioactive products. Recent advances in high-throughput genetic sequencing have generated a wealth of publicly-available genomic information. The Platform of the Joint Genome Institute (JGI IMG-ABC) boasts the largest collection of automatically mined gene clusters (> 960,000 putative gene clusters); however, most of these clusters and their products are not yet characterized.[102] Genome mining can be used as a powerful complement to proteomics experiments, however this technique requires sequence information, significant bioinformatics analyses/processing, and does not provide a direct measurement of bioactivity.

Novel methodologies are needed to identify low abundance bioactive species expressed in minimal quantities of natural product extracts derived from species lacking genetic information. With the implementation of high resolving power mass spectrometry, sensitive bioactivity assays, and sophisticated statistical analyses, identification of AMPs from complex natural products is viable via the PepSAVI-MS (**S**tatistically-guided bio**A**ctive **PEP**tides prioritized **VI**a **M**ass **S**pectrometry) pipeline developed in the Hicks laboratory.[103] Following this pipeline, peptide libraries are screened for growth inhibition in adaptable, high-throughput bioactivity assays and the relative abundances of peptidyl species in each fraction are determined via mass spectrometry. Statistical analysis via an elastic net regression model ranks the most probable peptidyl species responsible for the observed bioactivity. PepSAVI-MS does not require multiple iterations of fractionation, nor knowledge of existing gene clusters. PepSAVI-MS uniquely allows for simultaneous contributions from multiple molecular species contributing to synergistic interactions. The components of peptide libraries can be tailored readily to contain constitutive or inducible AMPs and AMPs derived from alternative source materials, e.g. bacterial and fungal secretomes[104, 105] or various plant tissues (Chapter 7). The flexibility of the PepSAVI-MS pipeline allows for the incorporation any bioactivity metric, including alternative assay formats, such as traditional disc-diffusion assays[104, 105] or microfluidic platforms (Chapter 6). Strategies to prioritize target species or focus the analysis on a specific protein class of interest (e.g. cyclotides), such as derivatization and mass shift analysis, will further streamline the process of rapid bioactive peptide discovery (Chapters 2-5). Mass spectrometry-based methods have enabled the identification and/or sequence characterization of novel AMPs in traditional botanical species[106, 107] (Chapters 3-5). Isolation of the putative bioactive peptides

detected in complex plant fractions will allow for subsequent characterization and facilitates future mechanistic studies (e.g. MOA).

REFERENCES

1. Ventola, C.L., *The antibiotic resistance crisis: part 1: causes and threats.* P T, 2015. **40**(4): p. 277-83.
2. *2018 Antimicrobial Resistance Fact Sheet*. March 8, 2020.
3. CDC, *Antibiotic Resistance Threats in the United States, 2019*. 2019, U.S. Department of Health and Human Services, CDC: Atlanta, GA.
4. Aslam, B., et al., *Antibiotic resistance: a rundown of a global crisis.* Infect Drug Resist, 2018. **11**: p. 1645-1658.
5. Jiang, T. and X.S. Chen, *Outcome Impacts Due to Pathogen-Specific Antimicrobial Resistance: A Narrative Review of Published Literature.* Int J Environ Res Public Health, 2020. **17**(4).
6. Mendelson, M. and M.P. Matsoso, *The World Health Organization Global Action Plan for antimicrobial resistance.* S Afr Med J, 2015. **105**(5): p. 325.
7. Tacconelli, E., et al., *Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis.* Lancet Infect Dis, 2018. **18**(3): p. 318-327.
8. Mulani, M.S., et al., *Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review.* Front Microbiol, 2019. **10**: p. 539.
9. Founou, R.C., L.L. Founou, and S.Y. Essack, *Clinical and economic impact of antibiotic resistance in developing countries: A systematic review and meta-analysis.* PLoS One, 2017. **12**(12): p. e0189621.
10. Ma, Y.X., et al., *Considerations and Caveats in Combating ESKAPE Pathogens against Nosocomial Infections.* Adv Sci (Weinh), 2020. **7**(1): p. 1901872.
11. Santajit, S. and N. Indrawattana, *Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens.* Biomed Res Int, 2016. **2016**: p. 2475067.
12. D'Costa, V.M., et al., *Antibiotic resistance is ancient.* Nature, 2011. **477**(7365): p. 457-61.
13. Chokshi, A., et al., *Global Contributors to Antibiotic Resistance.* J Glob Infect Dis, 2019. **11**(1): p. 36-42.
14. Luyt, C.E., et al., *Antibiotic stewardship in the intensive care unit.* Crit Care, 2014. **18**(5): p. 480.
15. Lerminiaux, N.A. and A.D.S. Cameron, *Horizontal transfer of antibiotic resistance genes in clinical environments.* Can J Microbiol, 2019. **65**(1): p. 34-44.
16. Wu, B., et al., *Detection of microbial aerosols in hospital wards and molecular identification and dissemination of drug resistance of Escherichia coli.* Environ Int, 2020. **137**: p. 105479.

17. Donkor, E.S., *Nosocomial Pathogens: An In-Depth Analysis of the Vectorial Potential of Cockroaches.* Trop Med Infect Dis, 2019. **4**(1).

18. Bartlett, J.G., D.N. Gilbert, and B. Spellberg, *Seven ways to preserve the miracle of antibiotics.* Clin Infect Dis, 2013. **56**(10): p. 1445-50.

19. Perez-Rodriguez, F. and B. Mercanoglu Taban, *A State-of-Art Review on Multi-Drug Resistant Pathogens in Foods of Animal Origin: Risk Factors and Mitigation Strategies.* Front Microbiol, 2019. **10**: p. 2091.

20. CDC, *Core Elements of Hospital Antibiotic Stewardship Programs.*, C. US Department of Health and Human Services, Editor. 2019: Atlanta, GA.

21. Wendt, S., et al., *[Antibiotic stewardship (ABS). Part 1: Basics]*. Internist (Berl), 2020.

22. Adre, C., R.L.P. Jump, and S.S. Spires, *Recommendations for Improving Antimicrobial Stewardship in Long-Term Care Settings Through Collaboration.* Infect Dis Clin North Am, 2020. **34**(1): p. 129-143.

23. Olans, R.D., N.B. Hausman, and R.N. Olans, *Nurses and Antimicrobial Stewardship: Past, Present, and Future.* Infect Dis Clin North Am, 2020. **34**(1): p. 67-82.

24. Sukumar, S., et al., *Think before you prescribe: how dentistry contributes to antibiotic resistance.* Aust Dent J, 2020. **65**(1): p. 21-29.

25. Robinson, E.D., et al., *Collaborative Antimicrobial Stewardship for Surgeons.* Infect Dis Clin North Am, 2020. **34**(1): p. 97-108.

26. Evans, C.D. and J.W.S. Lewis, *Collaborative Antimicrobial Stewardship in the Health Department.* Infect Dis Clin North Am, 2020. **34**(1): p. 145-160.

27. Fletcher, S., *Understanding the contribution of environmental factors in the spread of antimicrobial resistance.* Environ Health Prev Med, 2015. **20**(4): p. 243-52.

28. Jayachandran, S., *Pre-antibiotics Era to Post-antibiotic Era.* J Indian Acad Oral Med Radiol, 2018. **30**: p. 100-101.

29. Friedman, N.D., E. Temkin, and Y. Carmeli, *The negative impact of antibiotic resistance.* Clin Microbiol Infect, 2016. **22**(5): p. 416-22.

30. Bartlett, J.G. and L.M. Mundy, *Community-acquired pneumonia.* N Engl J Med, 1995. **333**(24): p. 1618-24.

31. Hirsch, E.F., *"The Treatment of Infected Wounds," Alexis Carrel's contribution to the care of wounded soldiers during World War I.* J Trauma, 2008. **64**(3 Suppl): p. S209-10.

32. Yoshikawa, T.T., *Antimicrobial resistance and aging: beginning of the end of the antibiotic era?* J Am Geriatr Soc, 2002. **50**(7 Suppl): p. S226-9.

33. Aminov, R.I., *A brief history of the antibiotic era: lessons learned and challenges for the future.* Front Microbiol, 2010. **1**: p. 134.

34. Davies, J., *Where have All the Antibiotics Gone?* Can J Infect Dis Med Microbiol, 2006. **17**(5): p. 287-90.

35. De Waele, J.J., J. Boelens, and I. Leroux-Roels, *Multidrug-resistant bacteria in ICU: fact or myth.* Curr Opin Anaesthesiol, 2020. **33**(2): p. 156-161.

36. Baker, S.J., et al., *Technologies to address antimicrobial resistance.* Proc Natl Acad Sci U S A, 2018. **115**(51): p. 12887-12895.

37. Mohs, R.C. and N.H. Greig, *Drug discovery and development: Role of basic biological research.* Alzheimers Dement (N Y), 2017. **3**(4): p. 651-657.

38. Theuretzbacher, U., et al., *The global preclinical antibacterial pipeline.* Nat Rev Microbiol, 2019.

39. Ling, L.L., et al., *A new antibiotic kills pathogens without detectable resistance.* Nature, 2015. **517**(7535): p. 455-9.

40. Hay, M., et al., *Clinical development success rates for investigational drugs.* Nat Biotechnol, 2014. **32**(1): p. 40-51.

41. He, R., et al., *Peptide Conjugates with Small Molecules Designed to Enhance Efficacy and Safety.* Molecules, 2019. **24**(10).

42. Otvos, L., Jr. and J.D. Wade, *Current challenges in peptide-based drug discovery.* Front Chem, 2014. **2**: p. 62.

43. Lee, A.C., et al., *A Comprehensive Review on Current Advances in Peptide Drug Development and Design.* Int J Mol Sci, 2019. **20**(10).

44. Lau, J.L. and M.K. Dunn, *Therapeutic peptides: Historical perspectives, current development trends, and future directions.* Bioorg Med Chem, 2018. **26**(10): p. 2700-2707.

45. Ruoslahti, E., *Peptides as targeting elements and tissue penetration devices for nanoparticles.* Adv Mater, 2012. **24**(28): p. 3747-56.

46. Kurzrock, R., et al., *Safety, pharmacokinetics, and activity of GRN1005, a novel conjugate of angiopep-2, a peptide facilitating brain penetration, and paclitaxel, in patients with advanced solid tumors.* Mol Cancer Ther, 2012. **11**(2): p. 308-16.

47. Baeshen, N.A., et al., *Cell factories for insulin production.* Microb Cell Fact, 2014. **13**: p. 141.

48. Chandrudu, S., P. Simerska, and I. Toth, *Chemical methods for peptide and protein production.* Molecules, 2013. **18**(4): p. 4373-88.

49. Hamman, J.H., G.M. Enslin, and A.F. Kotze, *Oral delivery of peptide drugs: barriers and developments.* BioDrugs, 2005. **19**(3): p. 165-77.

50. Renukuntla, J., et al., *Approaches for enhancing oral bioavailability of peptides and proteins.* Int J Pharm, 2013. **447**(1-2): p. 75-93.

51. Ali, A.M., et al., *Stapled Peptides Inhibitors: A New Window for Target Drug Discovery.* Comput Struct Biotechnol J, 2019. **17**: p. 263-281.

52. Verdine, G.L. and G.J. Hilinski, *Stapled peptides for intracellular drug targets.* Methods Enzymol, 2012. **503**: p. 3-33.

53. Clark, R.J., et al., *Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13767-72.

54. Iwai, H. and A. Pluckthun, *Circular beta-lactamase: stability enhancement by cyclizing the backbone.* FEBS Lett, 1999. **459**(2): p. 166-72.

55. Craik, D.J. and J. Du, *Cyclotides as drug design scaffolds.* Curr Opin Chem Biol, 2017. **38**: p. 8-16.

56. Mahlapuu, M., et al., *Antimicrobial Peptides: An Emerging Category of Therapeutic Agents.* Front Cell Infect Microbiol, 2016. **6**: p. 194.

57. Hancock, R.E. and G. Diamond, *The role of cationic antimicrobial peptides in innate host defences.* Trends Microbiol, 2000. **8**(9): p. 402-10.

58. Mookherjee, N., et al., *Antimicrobial host defence peptides: functions and clinical potential.* Nat Rev Drug Discov, 2020.

59. Wang, G., X. Li, and Z. Wang, *APD3: the antimicrobial peptide database as a tool for research and education.* Nucleic Acids Res, 2016. **44**(D1): p. D1087-93.

60. Kampshoff, F., M.D.P. Willcox, and D. Dutta, *A Pilot Study of the Synergy between Two Antimicrobial Peptides and Two Common Antibiotics.* Antibiotics (Basel), 2019. **8**(2).

61. Brogden, K.A., *Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?* Nat Rev Microbiol, 2005. **3**(3): p. 238-50.

62. Tossi, A., L. Sandri, and A. Giangaspero, *Amphipathic, alpha-helical antimicrobial peptides.* Biopolymers, 2000. **55**(1): p. 4-30.

63. Yin, J., et al., *Mechanisms of bactericidal action and resistance of polymyxins for Gram-positive bacteria.* Appl Microbiol Biotechnol, 2020.

64. Yeaman, M.R. and N.Y. Yount, *Mechanisms of antimicrobial peptide action and resistance.* Pharmacol Rev, 2003. **55**(1): p. 27-55.

65. Munuswamy, H., et al., *A review on antimicrobial efficacy of some traditional medicinal plants in Tamilnadu.* Journal of Acute Disease, 2013. **2**(2): p. 99-105.

66. *World Health Organization model list of essential medicines: 21st list 2019,* W.H. Organization, Editor. 2019, World Health Organization.

67. Nation, R.L. and J. Li, *Colistin in the 21st century.* Curr Opin Infect Dis, 2009. **22**(6): p. 535-43.

68. Lei, J., et al., *The antimicrobial peptides and their potential clinical applications.* Am J Transl Res, 2019. **11**(7): p. 3919-3931.

69. Divyashree, M., et al., *Clinical Applications of Antimicrobial Peptides (AMPs): Where do we Stand Now?* Protein Pept Lett, 2020. **27**(2): p. 120-134.

70. Koo, H.B. and J. Seo, *Antimicrobial peptides under clinical investigation.* Peptide Science, 2019. **111**(5).

71. Kang, X., et al., *DRAMP 2.0, an updated data repository of antimicrobial peptides.* Sci Data, 2019. **6**(1): p. 148.

72. Wink, M., *Plant Secondary Metabolites Modulate Insect Behavior-Steps Toward Addiction?* Front Physiol, 2018. **9**: p. 364.

73. Wink, M., *Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective.* Phytochemistry, 2003. **64**(1): p. 3-19.

74. Zaynab, M., et al., *Role of secondary metabolites in plant defense against pathogens.* Microb Pathog, 2018. **124**: p. 198-202.

75. Yamamoto, K., et al., *Development of LC-MS/MS analysis of cyclic dipeptides and its application to tea extract.* Biosci Biotechnol Biochem, 2016. **80**(1): p. 172-7.

76. Salas, C.E., et al., *Biologically active and antimicrobial peptides from plants.* Biomed Res Int, 2015. **2015**: p. 102129.

77. Stotz, H.U., J.G. Thomson, and Y. Wang, *Plant defensins: defense, development and application.* Plant Signal Behav, 2009. **4**(11): p. 1010-2.

78. Nawrot, R., et al., *Plant antimicrobial peptides.* Folia Microbiol (Praha), 2014. **59**(3): p. 181-96.

79. Zhang, L.J. and R.L. Gallo, *Antimicrobial peptides.* Curr Biol, 2016. **26**(1): p. R14-9.

80. Tam, J.P., et al., *Antimicrobial Peptides from Plants.* Pharmaceuticals (Basel), 2015. **8**(4): p. 711-57.

81. Mahomoodally, M.F., *Traditional medicines in Africa: an appraisal of ten potent african medicinal plants.* Evid Based Complement Alternat Med, 2013. **2013**: p. 617459.

82. Abdullahi, A.A., *Trends and challenges of traditional medicine in Africa.* Afr J Tradit Complement Altern Med, 2011. **8**(5 Suppl): p. 115-23.

83. Park, H.L., et al., *Traditional medicine in china, Korea, and Japan: a brief introduction and comparison.* Evid Based Complement Alternat Med, 2012. **2012**: p. 429103.

84. Mbuni, Y.M., et al., *Medicinal Plants and Their Traditional Uses in Local Communities around Cherangani Hills, Western Kenya.* Plants (Basel), 2020. **9**(3).

85. Chang, N., et al., *Indigenous Uses and Pharmacological Activity of Traditional Medicinal Plants in Mount Taibai, China.* Evid Based Complement Alternat Med, 2017. **2017**: p. 8329817.

86. Gran, L., *On the effect of a polypeptide isolated from "Kalata-Kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus.* Acta Pharmacol Toxicol (Copenh), 1973. **33**(5): p. 400-8.

87. Dong, J., *The Relationship between Traditional Chinese Medicine and Modern Medicine.* Evid Based Complement Alternat Med, 2013. **2013**: p. 153148.

88. Taylor, J.L.S., et al., *Towards the scientific validation of traditional medicinal plants.* Plant Growth Regulation, 2001. **34**: p. 23-37.

89. Hollenbeak, K.H. and F.J. Schmitz, *Aplysinopsin: antineoplastic tryptophan derivative from the marine sponge Verongia spengelii.* Lloydia, 1977. **40**(5): p. 479-81.

90. Wani, M.C., et al., *Plant antitumor agents. XVI. 6alpha-Senecioyloxy-chaparrinone, a new antileukemic quassinoid from Simaba multiflora.* Lloydia, 1978. **41**(6): p. 578-83.

91. Roberts, G.K., et al., *Finding the bad actor: Challenges in identifying toxic constituents in botanical dietary supplements.* Food Chem Toxicol, 2019. **124**: p. 431-438.

92. Weller, M.G., *A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques.* Sensors (Basel), 2012. **12**(7): p. 9181-209.

93. Yang, X., et al., *Isolation of an antimicrobial compound from Impatiens balsamina L. using bioassay-guided fractionation.* Phytother Res, 2001. **15**(8): p. 676-80.

94. Jamil, M., et al., *Isolation of antibacterial compounds from Quercus dilatata L. through bioassay guided fractionation.* Ann Clin Microbiol Antimicrob, 2012. **11**: p. 11.

95. Malheiros, A., et al., *Antifungal activity of drimane sesquiterpenes from Drimys brasiliensis using bioassay-guided fractionation.* J Pharm Pharm Sci, 2005. **8**(2): p. 335-9.

96. Si, W., et al., *Bioassay-guided purification and identification of antimicrobial components in Chinese green tea extract.* J Chromatogr A, 2006. **1125**(2): p. 204-10.

97. Neghabi-Hajiagha, M., et al., *A bioassay-guided fractionation scheme for characterization of new antibacterial compounds from Prosopis cineraria aerial parts.* Iran J Microbiol, 2016. **8**(1): p. 1-7.

98. Ward, A.C. and N.E. Allenby, *Genome mining for the search and discovery of bioactive compounds: the Streptomyces paradigm.* FEMS Microbiol Lett, 2018. **365**(24).

99. Meleshko, D., et al., *BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs.* Genome Res, 2019. **29**(8): p. 1352-1362.

100. Blin, K., et al., *antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.* Nucleic Acids Res, 2019. **47**(W1): p. W81-W87.

101. Skinnider, M.A., et al., *PRISM 3: expanded prediction of natural product chemical structures from microbial genomes.* Nucleic Acids Res, 2017. **45**(W1): p. W49-W54.

102. Ziemert, N., M. Alanjary, and T. Weber, *The evolution of genome mining in microbes - a review.* Nat Prod Rep, 2016. **33**(8): p. 988-1005.

103. Kirkpatrick, C.L., et al., *The "PepSAVI-MS" Pipeline for Natural Product Bioactive Peptide Discovery.* Anal Chem, 2017. **89**(2): p. 1194-1201.

104. Kirkpatrick, C.L., et al., *Fungal Secretome Analysis via PepSAVI-MS: Identification of the Bioactive Peptide KP4 from Ustilago maydis.* J Am Soc Mass Spectrom, 2018.

105. Kirkpatrick, C.L., et al., *Exploring bioactive peptides from bacterial secretomes using PepSAVI-MS: identification and characterization of Bac-21 from Enterococcus faecalis pPD1.* Microb Biotechnol, 2018. **11**(5): p. 943-951.

106. Moyer, T.B., et al., *PepSAVI-MS Reveals a Proline-rich Antimicrobial Peptide in Amaranthus tricolor.* J Nat Prod, 2019. **82**(10): p. 2744-2753.

107. Parsley, N.C., et al., *PepSAVI-MS reveals anticancer and antifungal cycloviolacins in Viola odorata.* Phytochemistry, 2018. **152**: p. 61-70.

CHAPTER 2: Cysteine-rich Antimicrobial Peptides: Identification via Mass Spectrometry

2.1 Introduction

Antimicrobial peptides (AMPs) are promising lead compounds (Chapter 1); however, the varied biophysical and biochemical characteristics of AMPs (secondary/tertiary structure, post-translational modification, molecular weight, and bioactivities) present additional hurdles towards rapid and accurate bioactive peptide identification. Furthermore, the expression and abundance of AMPs are dynamic and influenced by environment, pathogen stress, season, and circadian rhythms.[1-4] As such, the bioactive chemical repertoires of promising botanical species remain largely unexplored; new approaches are needed to overcome complexity, diversity, and periodicity to efficiently mine the vast chemical space accessed by AMPs.

Early AMP identification and characterization was achieved with traditional biochemical methods (e.g. peptide isolation and sequencing via Edman degradation).[5] However, significant advances in instrumentation have elevated mass spectrometry (MS) as a prime analytical technique with unrivaled experimental flexibility, specificity, sensitivity, and limits of detection. As such, MS-based peptidomics drive modern AMP discovery and characterization, with the ability to identify and resolve low abundance peptide species in complex natural product extracts.[5] Orthogonal tandem MS methods,[6] such as collision-induced dissociation (CID) and electron transfer dissociation (ETD), and chemical derivatization strategies facilitate characterization of peptides in complex extracts without prior isolation. However, extreme chemical diversity often complicates comprehensive bioactive peptide identification. Additional strategies are needed to expedite the accurate identification and characterization of natural product peptides.

2.2 Cysteine-rich Antimicrobial Peptides

Despite significant sequence and structural diversity among AMP families, cysteine-rich peptides (CRPs) are highly represented among botanical AMP constituents[7-10] (Table 2.1). Cysteine-rich peptides orchestrate various elements of physiology, reproduction, growth/development, and defense against pathogens.[11-15] Disulfide bonding among cysteine residues scaffolds CRPs, affording stability via reduced conformational entropy and enhanced enzyme and chemical resistance; as such, bioactive CRPs are attractive drug candidates with the potential to survive harsh physiological conditions. Cysteine/disulfide number and cysteine spacing along the peptide backbone contribute to the classification of botanical AMP families,[8] e.g. knottins exhibit three conserved disulfides,[16-18] whereas defensins may contain three to five disulfide bonds.[19] Additionally, CRPs can be found in ultra-stable, cyclic conformations; cyclic CRPs often require *in vitro* enzymatic digestion (e.g. endoproteinase Glu-C) to cleave the cyclic backbone and facilitate MS^2 sequence characterization (as is demonstrated in Chapters 3-6). The diversity of CRPs can vary substantially among sources, while the comprehensive characterization of the suite of CRPs in a single botanical species can elude even the most sensitive analytical methods.

Conserved features among AMP families, e.g. high cysteine content, can guide MS-based methods towards the identification of putative bioactive sequences. As such, chemical derivatization strategies can be exploited to detect CRPs in complex natural product matrices by mass shift analysis. With this approach, extracts are subject to chemical reduction and subsequent cysteine derivatization via alkylation; mass species possessing cysteine residues produce a mass shift readily observable with mass spectrometry. As each cysteine residue is expected to be singly derivatized (e.g. a carbamidomethyl group is added to each cysteine when

using iodoacetamide) (Figure 2.1, Table 2.1), this method identifies both the presence and quantity of cysteine residues in a given mass species. Mass shift analysis is often the first step towards the identification of CRPs in botanical extracts.[20, 21]

2.3 Methods

2.3.1 Mass Shift Analysis

Reduction of disulfide-bound cysteine residues and subsequent cysteine derivatization with iodoacetamide prevents disulfides from reforming and results in a predictable mass shift proportional to the number of cysteine residues in a peptide (Figure 2.1). Samples subject to mass shift analysis must be split (approximately in half) prior to reduction/alkylation: half are subject to reduction/alkylation, and the other half are set aside to assess intact masses. Both samples will be analyzed via LC-MS to identify mass shifts before and after reduction/alkylation consistent with the presence of multiple cysteine residues/disulfides.

Protocol:

1. Ion-exchange or reversed-phase liquid chromatography (LC) fractions are often used for mass shift analysis. The concentration of these samples can be crudely estimated from the 280 nm LC trace; 280 nm peaks in cyclotide-containing fractions often reach ~3000-4000 mV.
2. After elution from the LC system, samples are dried in a vacuum concentrator and resuspended in 200 µL LC-MS grade water to achieve a "1X" concentration.
    a. If samples are eluted from a reversed-phase system, solvent (organic) evaporation in the CentriVap is sufficient.

b. If samples are eluted from an ion-exchange system (aqueous with a volatile salt, e.g. ammonium formate), desalting must be performed by solid-phase extraction (detailed in step 6) and subsequently dried in the CentriVap. Alternatively, fractions can be dried in the CentriVap, reconstituted in 1.5 mL MilliQ water, dried again in the CentriVap, and this process repeated at least three times to remove volatile salts. Often, the latter process fails to completely desalt samples, and is highly dependent on CentriVap performance.

3. Dilute approximately 10 µL of 1X fraction in 40 µL 100 mM ammonium bicarbonate, pH 7.8 (total volume 50 µL). Depending on the anticipated number of experiments to be performed with this material, the amount of 1X fraction can be scaled up. For example, if reducing, alkylating, and enzymatically-digesting for both mass shift analysis and MS/MS sequence coverage, 30 µL or more 1X sample is preferable; scale up the amount of ammonium bicarbonate buffer accordingly (e.g. 30 µL sample in 120 µL ammonium bicarbonate buffer, for a total volume of 150 µL).
4. Reduce samples with 10 mM dithiothreitol (DTT), working concentration, at 45 °C, shaking at 850 rpm for 30 min. Fresh DTT must be prepared daily in 100 mM ammonium bicarbonate, pH 7.8.
5. Alkylate the reduced samples with 100 mM iodoacetamide (IAM), working concentration, at room temperature (~ 25 °C), shaking at 850 rpm for 15 min in the dark.
    a. Iodoacetamide is not stable in solution; as such, care must be taken to cover IAM stock and IAM-containing sample with tin foil to reduce light exposure.
    b. Fresh IAM must be prepared daily in 100 mM ammonium bicarbonate, pH 7.8.

c. Additionally, IAM reactions must be quenched, e.g. with excess DTT, shortly after the 15 minute reaction time, as the frequency of cysteine over-alkylation and off-target alkylation of non-cysteine residues via N- and O-linkages, increases with time.[22, 23] In the Hicks laboratory, we commonly see peptide N-terminus and tryptophan alkylation if alkylation is allowed to proceed for longer than ~15 min.

6. Remove interfering contaminants / salts via solid phase extraction (SPE) with Pierce C_{18} spin columns (Thermo Scientific) per manufacturer protocols, with one exception as described below. Samples must be acidified to pH 3 with 5% formic acid or trifluoroacetic acid prior to SPE.
   a. Pierce C_{18} spin columns: After wetting resin with 50/50 water/methanol, and prior to equilibrating resin in 95/5/0.1% water/acetonitrile/trifluoroacetic acid, perform a "pre-elution" step by adding 200 µL of 70/30/0.1% acetonitrile/water/trifluoroacetic acid.
   b. High organic pre-elution is not included in the manufacturer protocol; however, it is an essential step to mitigate polymer contamination derived from the spin column device prior to loading samples. Without this pre-elution step, polymers, particularly polyethylene glycol (PEG), will leach into the final sample elution and negatively affect subsequent LC-MS analysis.
7. Dry SPE elutions in a vacuum concentrator (CentriVap) and resuspend in 0.1% formic acid in LC-MS grade water for LC-MS/MS analysis. Resuspend as 1X concentration (the amount of sample used originally for reduction and alkylation, e.g. 10 µL).
   a. Often, cyclotide-containing samples at 1X concentration must be diluted 100X to achieve ~1 µg per 5 µL injection.

8. Optional step: Even in a reduced and alkylated form, cyclic CRPs are recalcitrant towards MS/MS fragmentation, challenging sequencing efforts. As such, enzymatic digestion is leveraged to pre-linearize cyclic species *in vitro*. Perform Glu-C digestion to cleave cyclic species at the C-terminus of glutamic acid residues by incubating reduced and alkylated samples with 1:200 enzyme:substrate (endoproteinase Glu-C:sample) at 37 °C, shaking at 850 rpm, for 3 h. Perform SPE as described previously (Step 6).
    a. Endoproteinase Glu-C cleaves at the C-terminus of aspartic acid residues ~ 100 X slower than that of glutamic acid residues. As such, longer incubation times (e.g. 12+ h) may increase aspartic acid cleavage.

2.3.2 Mass Shift Data Acquisition

Protocol:

1. Inject ~1 μg of acidified intact, reduced/alkylated, or reduced/alkylated/Glu-C digested sample onto the nano-LC-ESI-MS/MS.
    a. All mass shift data described herein were acquired on a NanoAcquity (Waters, Milford, MA) coupled to a TripleTOF5600 MS (AB Sciex, Framingham, MA) equipped with a Symmetry C_{18} trap column (100 Å, 5 μm, 180 μm × 20 mm, Waters) and HSS T3C_{18} analytical column (100 Å, 1.8 μm, 75 μm × 250 mm, Waters). A flow rate of 0.3 μL/min and a 30 minute linear ramp of 5%−50% mobile phase B (mobile phase A, 1% formic acid in water; mobile phase B, 1% formic acid in acetonitrile) were used for UPLC peptide separation. The TripleTOF5600 MS was operated in positive-ion, high-sensitivity mode.

2. MS experiments consisted of a 2 s cycle, with an MS survey scan using a mass range of 350−1600 Da in 250 ms, and IDA where the first 20 MS^2 spectra were collected over a mass range of 100-1800 Da in ~87 ms each.

2.3.3 Mass Shift Data Analysis

Protocol:

1. Generate de-isotoped peak lists for the intact and reduced/alkylated samples using Progenesis QI for Proteomics software (Nonlinear Dynamics, v.2.0) by importing intact and reduced/alkylated .wiff files into separate Progenesis files (do not align). Use a retention time filter of 14−45 min and maximum charge of +10.
2. Export "Peptide ion data" .csv files (one for each of the intact and reduced/alkylated samples) from Progenesis.
3. Paste the intact and reduced/alkylated mass (Da) columns from the two .csv files generated in Progenesis (step 2) into a single .csv file, with the intact and reduced/alkylated masses in columns 1 and 2, respectively.
4. Use the Python Mass Shift Analysis script (https://github.com/hickslab/MassShiftAnalysis) to identify mass shifts corresponding to a defined number of disulfide bonds/cysteine residues.
    a. The mass shift searched by the Python script can be modified by opening the script (in Notepad++ or IDLE) and modifying the number of disulfides/cysteine residues searched and mass of alkylating agent (Figure 2.2, Table 2.2).
5. Numerous mass species (often > 100) will be detected in the mass shift analysis of a given botanical species. The accuracy of peak assignments and mass shifts must be confirmed via manual interrogation of intact and reduced/alkylated mass spectral data.

2.4 Results

Mass shift analysis was used to probe the extent of putative CRPs containing three, four, or five disulfide bonds in botanical extracts (Table 2.3). Viola spp. consistently reported abundant disulfide-rich masses, where *Viola odorata*, *Viola inconspicua*, and *Viola communis* exhibited numerous mass shifts corresponding to three disulfides – consistent with cyclotides. Analyses of all plant materials demonstrated mass shifts consistent with at least three disulfide bonds, however, manual investigation of the mass spectra of numerous botanical species (e.g. *Psychotria zombamontana*, *Larrea tridentata*, *Datura stramonium*, and *Hyoscyamus niger*) did not support masses exhibiting the anticipated mass shifts, likely a result of poor peak picking or spectral noise. *Trichosanthes kirilowrii* is of particular interest for follow-up studies, exhibiting numerous mass shifts consistent with three disulfides – all demonstrating similar high abundance MS/MS "fingerprint" ions, indicating that these masses contain similar sequence motifs. Additionally, *Cucurbita foetidissima* and *Cucurbita pepo* warrant further investigation; MS/MS spectra of masses exhibiting three disulfide mass shifts should be inspected for sequence information.

2.5 Conclusion

Conserved AMP features can inform mass spectral methods for the identification and characterization of novel sequences in highly complex natural product extracts. Cysteine-rich peptide identification can be achieved through an established cysteine derivatization strategy and mass shift analysis. Mass species exhibiting mass shifts consistent with three or more disulfides must be further characterized. Tandem mass spectrometry can be leveraged to elucidate the putative CRP sequence, and bioactivity assays can be performed on peptide library fractions containing the mass species (Chapters 3-4).

2.6 Tables

Table 2.1. Disulfide characteristics of common AMP structural classes. Hevein-like and defensin classes exhibit significant cysteine/disulfide diversity, as such, only one possible cysteine pattern is represented.

Structural class	Size (kDa)	No. Disulfides	Disulfide Pattern
Knottin	3	3	C1–C4, C2–C5, C3–C6
Hevein-like peptides	4	3 to 5	C1–C4, C2–C5, C3–C6
Thionin	5	4	C1–C8, C2–C5, C3–C6, C4–C7
Defensin (γ-thionin)	5	3 to 5	C1–C8, C2–C5, C3–C6, C4–C7
Lipid transfer protein	9	4	C1–C6, C2–C3, C4–C7, C5–C8

Table 2.2. Common alkylating agents used for cysteine derivatization.

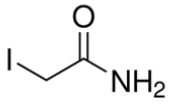
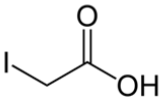
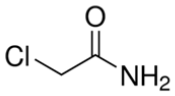
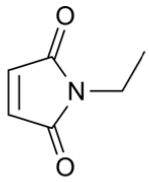
Alkylating Agent	$Cys^{(red)}$ Mass Shift (Da)	$Cys^{(ox)}$ Mass Shift (Da)	Structure
Iodoacetamide (IAM)	58.02	57.02	
Iodoacetic acid (IAA)	59.02	58.02	
Chloroacetamide (CAA)	58.02	57.02	
N-ethylmaleimide (NEM)	126.13	125.13	

Table 2.3. Botanical materials analyzed via mass shift analyses for three, four, and five disulfides (masses > 1500 Da). *Thoroughly investigated via manual search of MS data. † Manually searched, however, should be verified with more thorough investigation. NA = Data not available.

Source Material					Mass Shift: Disulfides		
Latin Name	Common Name	Family	Tissue	Dry/Fresh	Three	Four	Five
Cucurbita foetidissima	Buffalo gourd	Cucurbitaceae	Seed	-	37	16	5
Cucurbita maxima "Warty Goblin"	Warty Goblin	Cucurbitaceae	Seed	-	0†	NA	NA
Cucurbita pepo"Jack-O-Lantern"	Jack-O-Lantern	Cucurbitaceae	Seed	-	0†	20	19
Cucurbita pepo	Field pumpkin	Cucurbitaceae	Flower	-	51	28	24
Datura stramonium	Jimsonweed	Solanceae	Aerial	Fresh	0*	42	21
Helianthus annuus "Mammoth"	Sunflower	Asteraceae	Seed	-	84	64	49
Helianthus annuus "Skyscraper"	Sunflower	Asteraceae	Seed	-	31	26	21
Hyoscyamus niger	Henbane	Solanceae	Aerial	Fresh	0*	207	184
Lagenaria siceraria	Bottle gourd	Cucurbitaceae	Seed	-	14*	26	17
Larrea tridentata	Creosote bush	Zygophyllales	Seed	-	0*	2	7
Pisum sativum "Alaska"	Alaska pea	Fabaceae	Seed	-	0*	55	58
Pisum sativum var. saccharatum	Sugar snap pea	Fabaceae	Seed	-	0†	NA	NA
Psychotria zombamontana	Red bird-berry	Rubiaceae	Aerial	Dry	0*	4	7
Trichosanthes kirilowii	Snake gourd	Cucurbitaceae	Seed	-	5*	46	39
Viola inconspicua	-	Violaceae	Aerial	Fresh	44*	107	78
Viola inconspicua	-	Violaceae	Root	Fresh	44†	NA	NA
Viola odorata	Sweet violet	Violaceae	Aerial	Fresh	25*	NA	NA
Viola communis, March	-	Violaceae	Aerial	Fresh	54	100	82

2.7 Figures

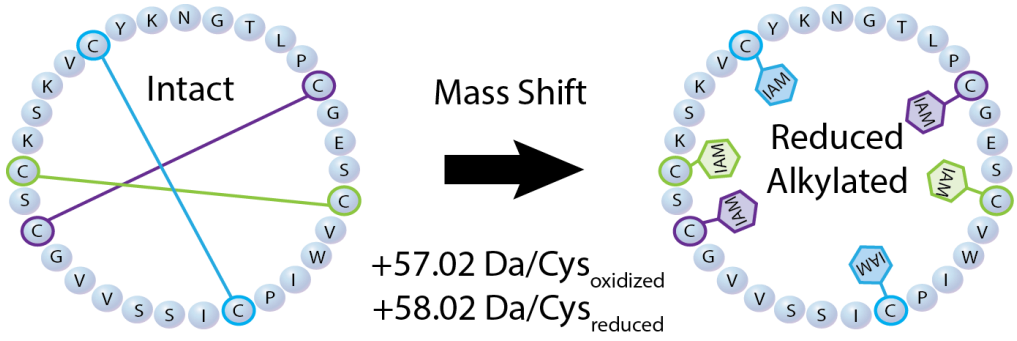
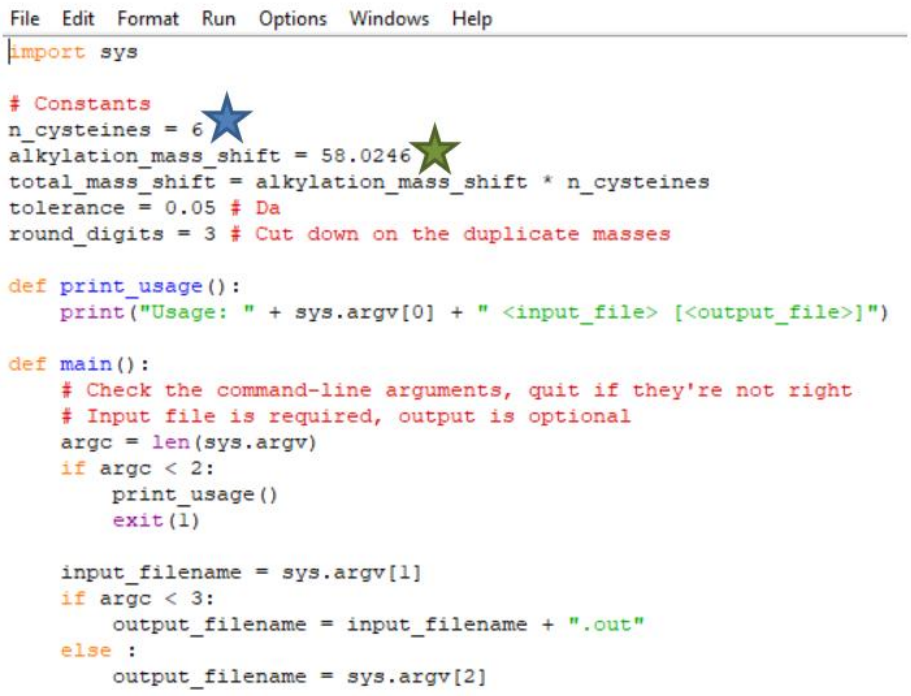


Figure 2.1. Reduction and alkylation of cysteine residues facilitates mass shift analysis, where each cysteine participating in a disulfide bond (oxidized) and free cysteine residue (reduced) will produce a mass shift of 57.02 and 58.02 Da with alkylation via iodoacetamide, respectively.



```
File  Edit  Format  Run  Options  Windows  Help
import sys

# Constants
n_cysteines = 6
alkylation_mass_shift = 58.0246
total_mass_shift = alkylation_mass_shift * n_cysteines
tolerance = 0.05 # Da
round_digits = 3 # Cut down on the duplicate masses

def print_usage():
    print("Usage: " + sys.argv[0] + " <input_file> [<output_file>]")

def main():
    # Check the command-line arguments, quit if they're not right
    # Input file is required, output is optional
    argc = len(sys.argv)
    if argc < 2:
        print_usage()
        exit(1)

    input_filename = sys.argv[1]
    if argc < 3:
        output_filename = input_filename + ".out"
    else :
        output_filename = sys.argv[2]
```

Figure 2.2. Modifications to the mass shift analysis Python script can be made by opening MassShiftAnalysis.py in any text editor (e.g. Notepad++, IDLE); cysteine number (blue star) and alkylating agent mass (green star) are easily edited to accommodate new experimental parameters.

REFERENCES

1. Gruber, C.W., et al., *Distribution and evolution of circular miniproteins in flowering plants.* Plant Cell, 2008. **20**(9): p. 2471-83.
2. Trabi, M., et al., *Variations in cyclotide expression in viola species.* J Nat Prod, 2004. **67**(5): p. 806-10.
3. Adams, S. and I.A. Carre, *Downstream of the plant circadian clock: output pathways for the control of physiology and development.* Essays Biochem, 2011. **49**(1): p. 53-69.
4. Roden, L.C. and R.A. Ingle, *Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant-pathogen interactions.* Plant Cell, 2009. **21**(9): p. 2546-52.
5. Martinez, O.F., et al., *Deciphering bioactive peptides and their action mechanisms through proteomics.* Expert Rev Proteomics, 2016. **13**(11): p. 1007-1016.
6. Zubarev, R., *Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry.* Expert Rev Proteomics, 2006. **3**(2): p. 251-61.
7. Kader, J.C., *Lipid-Transfer Proteins in Plants.* Annu Rev Plant Physiol Plant Mol Biol, 1996. **47**: p. 627-654.
8. Tam, J.P., et al., *Antimicrobial Peptides from Plants.* Pharmaceuticals (Basel), 2015. **8**(4): p. 711-57.
9. Slavokhotova, A.A., et al., *Hevein-Like Antimicrobial Peptides of Plants.* Biochemistry (Mosc), 2017. **82**(13): p. 1659-1674.
10. Stec, B., *Plant thionins--the structural perspective.* Cell Mol Life Sci, 2006. **63**(12): p. 1370-85.
11. Kumari, G., et al., *Cysteine-Rich Peptide Family with Unusual Disulfide Connectivity from Jasminum sambac.* J Nat Prod, 2015. **78**(11): p. 2791-9.
12. Marshall, E., L.M. Costa, and J. Gutierrez-Marcos, *Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development.* J Exp Bot, 2011. **62**(5): p. 1677-86.
13. Liu, X., et al., *Expansion and evolutionary patterns of cysteine-rich peptides in plants.* BMC Genomics, 2017. **18**(1): p. 610.
14. Loo, S., et al., *Bleogens: Cactus-Derived Anti-Candida Cysteine-Rich Peptides with Three Different Precursor Arrangements.* Front Plant Sci, 2017. **8**: p. 2162.
15. Shen, Y., et al., *Potentides: New Cysteine-Rich Peptides with Unusual Disulfide Connectivity from Potentilla anserina.* Chembiochem, 2019. **20**(15): p. 1995-2004.
16. Rees, D.C. and W.N. Lipscomb, *Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 A resolution.* J Mol Biol, 1982. **160**(3): p. 475-98.

17. Postic, G., et al., *KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling.* Nucleic Acids Res, 2018. **46**(D1): p. D454-D458.

18. Gelly, J.C., et al., *The KNOTTIN website and database: a new information system dedicated to the knottin scaffold.* Nucleic Acids Res, 2004. **32**(Database issue): p. D156-9.

19. Stotz, H.U., J.G. Thomson, and Y. Wang, *Plant defensins: defense, development and application.* Plant Signal Behav, 2009. **4**(11): p. 1010-2.

20. Narayani, M., A. Chadha, and S. Srivastava, *Cyclotides from the Indian Medicinal Plant Viola odorata (Banafsha): Identification and Characterization.* J Nat Prod, 2017. **80**(7): p. 1972-1980.

21. Burman, R., et al., *Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae.* Front Plant Sci, 2015. **6**: p. 855.

22. Muller, T. and D. Winter, *Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents.* Mol Cell Proteomics, 2017. **16**(7): p. 1173-1187.

23. Boja, E.S. and H.M. Fales, *Overalkylation of a Protein Digest with Iodoacetamide.* Analytical Chemistry, 2001. **73**(15): p. 3576-3582.

CHAPTER 3: PepSAVI-MS reveals anticancer and antifungal cycloviolacins in *Viola odorata*

3.1 Introduction

Antimicrobial peptides (AMPs) have guided the coevolution of countless biological species over hundreds of millions of years, the arsenal of naturally-occurring molecular weapons growing in potency and complexity as organisms compete for resources and real-estate. Conserved across all domains of life, AMPs are ancient defense molecules with broad functions.[1-4] In multicellular organisms, highly diverse ribosomally-synthesized AMPs are fundamental to the innate immune response, driving off invading bacterial, fungal, and viral pathogens, whereas prokaryotes may wield narrow-spectrum bactericidal AMPs to procure their place in ecological niches.[5-8] Acting by physical disruption of target membranes via an induced or permanent amphipathic motif,[9, 10] the mechanism of action (MOA) employed by most known AMPs has historically prevented the development of widespread AMP resistance in nature, requiring major remodeling of basic membrane structure.[11] Antimicrobial peptides often exhibit broad spectrum activity across diverse phyla, however, AMPs can be highly specialized, targeting specific and unique membrane constituents.[12, 13] Unlike neutrally-charged mammalian cells, both cancer and microbial cells rich with anionic molecules can be selectively targeted by AMPs with cell death caused by either membranolytic or intracellular mechanisms.[7, 13-16] With

the recent surge of clinical multidrug resistant pathogens and cancers, AMPs have become increasingly attractive therapeutic candidates boasting novel and diverse MOAs.[7, 17, 18]

Bioactive peptides exist in all organisms, however the abundance and diversity of these molecules is vastly underexplored as their identification in a sea of macromolecules and small molecule metabolites presents significant analytical challenges. With the implementation of high resolving power mass spectrometry, sensitive bioactivity assays, tailored and streamlined automated data processing and statistical analyses, identification of AMPs from complex natural extracts is viable via the PepSAVI-MS pipeline.[19, 20] Following this pipeline, peptide libraries are screened for growth inhibition in adaptable, high-throughput bioactivity assays and the relative abundances of peptidyl species in each fraction are determined via mass spectrometry. Statistical analysis via an elastic net regression model ranks the most probable peptidyl species responsible for the observed bioactivity. PepSAVI-MS does not require multiple iterations of fractionation as in bioassay-guided fractionation,[21] nor knowledge of existing gene clusters as in genome mining.[22] PepSAVI-MS uniquely allows for simultaneous contributions from multiple molecular species contributing to synergistic interactions. The components of peptide libraries can be tailored readily to contain constitutive or inducible AMPs by modifying experimental conditions. While a necessary and unique benefit of PepSAVI-MS, the ability to account for synergistic effects as well as contributions from multiple peptidyl species greatly increases the number of putative bioactive targets. As such, strategies to prioritize target species or focus the analysis on a specific protein class of interest (e.g. cyclotides as presented herein) will further streamline the process of rapid bioactive peptide discovery.

Cyclotides are a class of cyclic, disulfide-rich, plant-derived peptides found primarily in the Violaceae and Rubiaceae families that boast a diverse range of innate bioactivities ranging from antibacterial, anticancer, anti-HIV, mulloscicidal and insecticidal [23-29] indicating a highly stable, evolvable scaffold. Approximately 400 unique cyclotide sequences have been documented; however, tens of thousands up to 150,000 are estimated to exist in nature.[30-33] The extraordinary stability conferred by the highly complex cyclic cysteine knot (CCK) motif found in all known cyclotides makes the conserved three-dimensional structure of this peptide class an ideal scaffold for protein engineering.[34-37] Naturally-occurring and modified cyclotides have already found application in both medicine and commercial agriculture. Grafting therapeutic peptides into cyclotide scaffolds alleviates the issue of peptide instability *in vivo*, and has the potential to deliver a variety of highly specific and efficacious peptide-based therapeutics via oral administration and to intracellular targets.[37, 38] Meanwhile, Sero-X, a potent bioinsecticide derived from the cyclotide-containing extract of *Clitorea ternatea*, was approved for use on cotton plants in Australia in early 2017. Highly effective at controlling insect pests, Sero-X does not harm beneficial insects such as bees and ladybugs, and exemplifies the next level of safe and effective biocontrol that can be achieved with peptide-based treatments.[39]

While mass spectrometry has been used to successfully identify many cyclotides, the standard approach includes pre- and post-reduction and alkylation with iodoacetamide to reveal any peaks with a mass shift of 348.16 Da[32] and linearization using endoproteinase Glu-C via a single conserved glutamic acid residue,[40] or other enzymes,[41] enabling efficient fragmentation to obtain primary protein sequence, possibly in tandem with transcriptome-mining,[42] While successful, sample loss from the numerous sample preparation steps hinders the sequence elucidation of less abundant peptidyl species. Likely, the challenges associated with purifying

structurally similar cyclotides and cyclotide sequencing are why hundreds of putative cyclotide masses have been identified via reduction/alkylation methods, but only a fraction of these masses have been sequenced.[30, 31] A top-down approach in which cyclotides are analyzed undigested would ameliorate sample loss issues,[43] but has not been demonstrated for cyclotides. While the commonly used collision induced dissociation (CID) is a practical method for short, linear peptides or enzymatically-digested proteins, CID often provides insufficient fragmentation of intact cyclic peptides. Ultraviolet photodissociation (UVPD) is an emerging technique that yields highly complex but rich fragmentation data, generating several possible types of fragmentation events (e.g. *a*, *b*, *c*, *x*, *y*, and *z*) (Supplemental Figure 1).[44] UVPD has been used to fragment circular (stapled) peptides,[45] but no cyclotides have been sequenced with UVPD previously.

The botanical species *Viola odorata* L. (Violaceae) abundantly produces >30 known, unique cyclotide sequences with potent and diverse bioactivities,[46] and may harbor up to 166 cyclotide species as indicated by mass shift analysis.[31] A *V. odorata* peptide library was screened against breast, prostate, and ovarian cancer cell lines for validation of the PepSAVI-MS pipeline,[19] Constituents of this complex botanical extract demonstrated substantial cancer cell cytotoxicity in all assays, particularly across cyclotide-containing library fractions. While known and novel cycloviolacin O2 (cyO2) activity was verified, our results indicated additional cyclotide species responsible for the activity observed in the cancer cell panel. Further analysis presented herein revealed another cyclotide, cycloviolacin O8 (cyO8), as a putative anticancer peptide (Figure 1). Isolated cyO8 demonstrated micromolar bioactivities against MDA-MB-231 breast, PC-3 prostate, and OVCAR-3 ovarian cancer cell lines. Additionally, the *V. odorata* library was shown to exhibit robust activity against the filamentous fungus *Fusarium graminearum*, responsible for the devastating disease Fusarium Head Blight (FHB); however, the bioactive

constituents were not identified.[19] Herein, PepSAVI-MS revealed the *F. graminearum* antifungal activity of cyO8, which was confirmed with isolated peptide. While cy08 has known anthelmintic activity against nematode larvae,[47] this work expands the reported bioactivities to anticancer and antifungal for this cyclotide. Additionally, a simple reduction/alkylation strategy was implemented to further refine molecular targets revealed with PepSAVI-MS, identifying several putative anticancer and antifungal cyclotides that require further molecular characterization. As these masses are low in abundance, traditional MS-based sequencing methods are not sufficient and alternative sequence elucidation techniques are required. Using cyO8 as a representative species, we show the complexity of cyclic-peptide 193 nm UVPD fragmentation and attempt to discern common cyclotide fragmentation patterns to inform application to future *de novo* sequencing algorithms.

3.2 Results

3.2.1 Mining *V. odorata*

Validation of the PepSAVI-MS pipeline produced numerous, previously uncharacterized *V. odorata* molecular species as highly ranked anticancer and antifungal targets.[19] Further exploration indicated that the original filtering conditions imposed were highly stringent, i.e. filtering out all *m/z* outside of the bioactivity region excluded peptides that might be detected via mass spectrometry but not yet high enough abundance to be revealed in the bioactivity assessment. As such, we refined the filtering conditions to include one strong cation exchange (SCX) fraction on either side of the bioactivity region and remodeled the cancer panel bioactivity to produce refined lists of ranked putative anticancer compounds with bioactivities against MDA-MB-231 breast, PC-3 prostate, or OVCAR-3 ovarian cancer cell lines (Supplemental Table 1). The *V. odorata* cyclotide cyO8 (MW 3225.42 Da) was ranked among the top 20

contributors to ovarian and prostate cancer cell cytotoxicity. Additionally, a clear region of bioactivity was seen in the cyclotide-containing fractions 25-31 against *F. graminearum* (Figure 2). Statistical modeling of the bioactivity region revealed cyO8, in both the +3 and +4 charge states, as a highly ranked antifungal candidate (Supplemental Table 1). Additionally, mass 3152.41 Da was ranked in the top 20 contributors to *F. graminearum* antifungal activity, and may represent the known cyclotides cycloviolacin O3 (MW 3152.38 Da) or cycloviolacin O7 (MW 3152.41 Da) from *V. odorata* or Vitri A (MW 3152.38) from *Viola tricolor*, as some cyclotide sequences are not unique to a single botanical species.

3.2.2 Validation of novel bioactivities revealed by PepSAVI-MS

Ranked peptides identified via PepSAVI-MS must be isolated for sequence and structure characterization and verification of putative bioactivities. To confirm the anticipated anticancer and antifungal activities of cyO8, *V. odorata* SCX fractions were further processed with reversed-phase LC to achieve cyO8 isolation from the complex mixture of plant peptides. Purified cyO8 was shown to be active against the three cancer cell lines tested. Inhibitory concentration (IC_{50}) values were determined for each cell line: MDA-MB-231 breast (1.15 µM), PC-3 prostate (1.05 µM), and OVCAR-3 ovarian (0.80 µM) (Figure 3). CyO8 cytotoxicity against non-cancerous human dermal fibroblast cells was determined to be ~3X less (3.13 µM) than cancer cell lines (Figure 3). Isolated cyO8 was shown to be active against *F. graminearum* with an IC_{50} of 25-30 µM (a two-fold serial dilution was performed and visually assessed for the minimum inhibitory concentration; after 72 h the well containing 24 µM cyO8 showed only the beginnings of fungal growth).

3.2.3 Revealing novel, putatively bioactive *V. odorata* cyclotides

Taking advantage of the highly conserved three disulfide bonds in cyclotides, reduction and alkylation of *V. odorata* combined SCX fractions 12-45 produced numerous mass shifts of 348.16 ± 0.05 Da, consistent with the alkylation of three disulfide bonds.[32] Interrogation of this data resulted in 25 candidate masses likely belonging to the cyclotide family (Table 1). Eight putative cyclotide species correlated to masses identified in PepSAVI-MS bioactivity screens against cancer cell lines and/or *F. graminearum*: 3154.38 Da, 3156.36 Da, 3165.49 Da, 3166.48 Da, 3170.38 Da, 3171.38 Da, 3184.39 Da, and 3257.37 Da (Table 1).

3.2.4 Characterization of 3257 Da putative cyclotide

Reduction, alkylation, and linearization with Glu-C of the most abundant 3257.37 Da containing *V. odorata* SCX fraction resulted in incomplete yet informative CID MS^2 spectra (Figure 4). The 32 Da mass difference between the putative cyclotide and the known, highly abundant cyclotide, cyO8, could be localized to a tryptophan residue, showing a mass shift of the peak *m/z* 533 to *m/z* 565, and likely represents a double oxidation of the side chain, a known post-translational modification in cyclotides.[33] Of the 24 remaining candidate cyclotide masses, some are likely known cyclotides with singly (oxindolyalanine, *oia*) or doubly-oxidized tryptophan (N-formylkynurenine, *nfk*) residues based on mass differences of multiples of 16 Da: 3154.38 Da ($cyO2_{oia}$), 3184.39 Da ($cyO3_{nfk}$), and 3170.38 Da ($cyO2_{nfk}$). However, several putative bioactive cyclotides, 3156.36 Da, 3165.49 Da, 3166.48 Da, and 3171.38 Da remain uncharacterized.

3.2.5 UVPD fragmentation of cyO8

Fragmentation of cyclic molecules must occur twice to observe fragment ions, as the first cleavage results only in ring opening and no initial mass change. Incomplete fragmentation of

intact cyclotides with traditional tandem MS approaches (e.g. CID, ECD/ETD) and sample loss from numerous reduction, alkylation, and digestion steps has traditionally hindered the sequence determination of low abundance cyclotide species.[30, 31] Generating rich fragmentation data of all ion types,[48] high energy UVPD has the potential to circumvent these sequencing challenges, particularly if fragmentation patterns can be established to aid in *de novo* sequence assignment. Absorption of high energy (193 nm) photons produced during UVPD allows access to excited electronic states that result in extensive fragmentation and production of numerous types of fragment ions. CyO8 fragmentation obtained by 193 nm UVPD using complex *V. odorata* library fractions yielded the spectrum presented in Figure 5A. Predominant fragmentation events during UVPD occurred at the C-terminus of the two proline residues in the cyO8 sequence; however, fragmentation was seen throughout most sequence iterations (Figure 5B). Peak labels in Figure 5A indicate the residue in position 1 of the identified fragment sequence and the ion type (for example, *y4*(8) is the sequence that results in an initial fragmentation event between residues 7 and 8 and a second cleavage corresponding to a *y4* fragment). The majority of fragment ions were not observed from any one particular sequence iteration, and only about 50% of the cyO8 residues could be confidently assigned. Although the cyO8 precursor mass was abundant, MS^2 fragments generated by UVPD were low intensity because the total abundance was distributed among numerous fragment ions. Spectral overlap precluded assignments in the *m/z* 800-900 and 1000-1200 regions, and isobars or fragment ions within a 30 ppm mass tolerance led to sequence ambiguity, in some cases up to five different sequence possibilities for one observed *m/z* (Supplemental Tables 2-6).

3.3 Discussion

Previously, we have shown that cyclotide-containing *V. odorata* fractions exhibit anti-cancer bioactivities against MDA-MB-231 breast, PC-3 prostate, and OVCAR-3 ovarian cancer cell lines.[19] In addition to the known activity of cyO2, refinement of data filtering criteria in PepSAVI-MS revealed additional known and putative cyclotides with possible antiproliferative activity against cancer cell lines. Specifically, cyO8 which is a known nematocidal cyclotide, was ranked highly in our modeling. Subsequent isolation and targeted bioassays revealed anticancer activities of cyO8 at low micromolar IC_{50} values. Additionally, cyO8 demonstrated approximately three times less cytotoxicity against human dermal fibroblast cells than to the human cancer cell lines. Commonly used pharmaceuticals targeting breast, ovarian, and prostate cancers exhibit IC_{50} values from low nanomolar to low micromolar concentrations, pushing cyO8 to the higher end of the treatment range.

While innate antibacterial, anticancer, and antiviral cyclotide bioactivities have been thoroughly explored over the past 35 years, the antifungal properties of cyclotides have seldom been reported, and typically pertain only to common lab strains of Candida sp.[49, 50] We are excited to report cyclotide antifungal activity against the agriculturally-relevant, filamentous fungus *F. graminearum*, with a MIC value of 25-30 μM. Interestingly, several of the *V. odorata* fractions promoted fungal growth in the *F. graminearum* bioactivity assays, highlighting a potential issue that may be encountered when screening additional fungal species with the PepSAVI-MS pipeline and the need to filter datasets wider than the observed bioactivity region. Fortunately, cyO8 was a potent fungal inhibitor, as weaker activities may not be observed if growth promotion masks bioactivity. The appearance of the semi-abundant cyO3/cyO7 in the *F. graminearum* top 20 ranked compounds is consistent with recently published work in which

cyO3 was identified as a low micromolar inhibitor of the opportunistic human pathogenic yeast, *Candida albicans*.[49] Though exceptionally stable and resistant to harsh conditions, cyclotides make particularly promising topical biofungicides as the eventual, natural degradation of these molecules controls environmental concentrations and prevents unwanted accumulation in soils.[13]

Several putative cyclotide masses identified from the pre- and post- reduction/alkylation analysis were found among the ranked lists for the human cancer cell lines and *F. graminearum*. Of these masses, we anticipate several may be known cyclotides with oxyindolyalanine or N-formylkynurenine tryptophan modifications. As most of these species are present in very low abundances, we have only been able to confirm the identity of 3257.37 Da as a doubly-oxidized tryptophan cyO8 ($cyO8_{nfk}$) via CID tandem mass spectrometry. Tryptophan oxidation has been documented previously in cyclotides as the solvent-exposed tryptophan residue is more susceptible to modification, and is thought to play a role in cyclotide degradation pathways *in vivo*.[33, 51-53] The 24 remaining putative cyclotide species have yet to be characterized. Several of these potential anticancer and antifungal cyclotide species have recently been reported by other groups in large scale reduction/alkylation experiments to identify novel cyclotides; 3154 Da and 3171 Da were reported by Göransson *et al.*,[30] whereas similar masses, 3166.23 Da, 3170.36 Da, and 3171.37 Da, were reported by Srivastava *et al*.[31] Four masses never reported in *V. odorata* appearing on the human cancer cell line and *F. graminearum* ranked lists, 3156.36 Da, 3165.49 Da, 3166.48 Da, and 3171.38 Da, are prioritized for sequence elucidation and activity confirmation. Though unknown in *V. odorata*, all of the proposed cyclotide masses identified herein match known cyclotide sequences in other botanical species and may be identical in sequence. As some cyclotides are shared among numerous plant species, there are also known isobaric cyclotides with unique sequences even within the same plant; further sequence

characterization is necessary. By combining a simple reduction/alkylation strategy with the PepSAVI-MS pipeline, we have been able to prioritize ranked *m/z* species more-likely to be contributing to observed bioactivities.

Traditional mass spectrometry-based cyclotide sequencing approaches utilize a reduction, alkylation, and enzymatic digestion strategy to linearize target peptides as intact cyclotides produce minimal fragmentation upon traditional collision induced dissociation. While linearized cyclotides are easier to fragment, resulting MS/MS spectra are often incomplete and do not allow for confident sequence identification of novel targets, with as little as 20% sequence coverage (Figure 4). Furthermore, increased sample handling steps significantly increase sample loss and result in minimal signal for peptide fragmentation and identification. Herein we demonstrate a top-down UVPD approach, using cyO8 as a model, to reduce sample loss and facilitate the sequence elucidation of less abundant cyclotide species. To estimate the potential complexity of the MS/MS spectra generated for cyclic peptides, the numbers of fragment ions were manually calculated for cyO8 considering a maximal charge of +4. Assuming initial and secondary cut sites at all 31 cyO8 backbone positions and always a *b*/*y* initial cleavage, UVPD fragmentation would result in ~22,000 *a*, *b*, *c*, *x*, *y*, and *z* theoretical ions compared to ~11,000 *a*, *b*, and *y* ions with CID (considering 31 residues, each with a 30 fragment ion series, four charge states, and six or three ion types for UVPD or CID, respectively). In comparison, theoretical fragmentation of the 31 residue linear cyO8 peptide results in ~720 *a*, *b*, *c*, *x*, *y*, and *z* fragments for UVPD and ~360 *a*, *b*, and *y* fragments with CID, including fragment ion charge states ranging from +1 to +4 and no internal fragment ions. Even with reduced and alkylated disulfide bridges, both required fragmentation events can occur at any position in the ring and leads to a combinatorial explosion of theoretical fragment ions. As such, the 193 nm UVPD fragmentation of reduced and alkylated

cyO8 in *V. odorata* library fractions resulted in MS^2 spectra that were overly data-rich and spectral overlap made peak assignments challenging. Though predominant fragmentation sites were observed at the two highly constrained proline sites, cleavage occurred throughout most cyO8 variations and few other discernible patterns emerged during analysis to inform *de novo* assignments of uncharacterized cyclotides. Most fragment types, *a*, *b*, *c*, *x*, *y*, and *z* ions, were seen in roughly the same abundances, except possibly a small preferential for *b*/*y* ion types. Additionally, ambiguity associated with isobaric fragment ions or fragments similar in *m/z* introduced uncertainty into the assignment of cyO8 peaks. Ambiguity among similar but non-identical fragment ions may be remedied by producing even higher resolution mass spectra, which may be feasible for targeted experiments where an increase in acquisition time would be insignificant for only one or a few *m/z* species. UVPD can conceivably be used to verify minor amino acid modifications in known sequences to rapidly aid in the characterization of engineered products. Additionally, UVPD strategies employing lower energy wavelengths (351 nm) in conjunction with chromogenic derivitizing tags may allow for *de novo* sequencing of unknown cyclotides.[54] For future *de novo* assignments of cyclic molecules fragmented with UVPD, a sophisticated algorithm is necessary in which more common fragmentation patterns are established and used to make decisive fragment assignments based on probability. Multistage mass spectrometry employing MS^n coupled with machine learning could lend a tremendous amount of information to aid in *de novo* sequencing, particularly using UVPD fragmentation; however, this technique requires large datasets for algorithm training, that are as of yet unavailable, and specialized instrumentation.[43]

3.4 Conclusions

Peptide-based treatments offer new and effective alternatives in drug discovery and sustainable agricultural practices, and cyclotides are promising candidates with enhanced stability and a host of diverse bioactivities. PepSAVI-MS revealed novel anticancer and antifungal activities of the known cyclotide, cyO8, and herein those activities have been validated and characterized. The work described here illuminates the possibility of re-mining known natural product sources via PepSAVI-MS with the likelihood that valuable bioactive constituents were previously overlooked and may be revealed. We also demonstrate the utility of a reduction/alkylation strategy to identify putative cyclotides with results prioritized via PepSAVI-MS to expedite the prioritization and validation of bioactive molecules. UVPD of a cyclotide was demonstrated for the first time using cyO8 as a model sequence and assessed for potential in *de novo* applications. Several putative cyclotide species with suspected bioactive properties have been identified via PepSAVI-MS; once characterized these novel cyclotides can add to the substantial sequence and functional diversity already seen in this ever-growing peptide family. This work adds to the growing body of knowledge concerning these fascinating knotted molecules whose seemingly boundless bioactivities will undoubtedly be harnessed to combat emerging microbial resistance in agriculture and medicine.

3.5 Materials and Methods

3.5.1 Preparation of Plant Material and Library Generation

Viola odorata L. (Violaceae) peptide library was prepared as described previously.[19] Briefly, *Viola odorata* seeds (Strictly Medicinal Seeds, OR) were grown to mature rosettes in a laboratory greenhouse at 17.5-20.3 °C on a 14/10 h light/dark cycle. Aerial tissue was harvested, manually ground, and proteins extracted in a 10% acetic acid buffer. Centrifugation was used to

pellet insoluble material and the remaining extract was filtered (0.45 μm stericup filtration; Millipore), concentrated to remove high-molecular-weight molecules (MWCO 30 kDa; Millipore), and dialyzed into 5 mM ammonium formate (Fluka) pH 2.7 to remove small-molecules (0.1−1 kDa cutoff; SpectrumLabs). The *V. odorata* peptide library was prepared by fractionating crude extract with a 47 min strong cation exchange (SCX) method (PolySulfethyl A column (100 mm × 4.6 mm, 3 μm particles, PolyLC) using a salt gradient (linear ramp from 5 mM ammonium formate, 20% acetonitrile, pH 2.7 to 500 mM ammonium formate, 20% acetonitrile, pH 3.0). Fractions were collected in one-min intervals, desalted, and stored at 4 °C.

3.5.2 Bioactivity Assays

V. odorata peptide library was tested against *F. graminearum* in a 96-well plate format bioactivity assay, as described previously.[19] Each plate well contained 30 μL minimal media, 10 μL *F. graminearum* spore culture, and 10 μL *V. odorata* library fraction, antibiotic (Hygromycin B, Corning), or water, and was prepared in triplicate. Minimal media is a modified Czapek-Dox formulation, a standard growth media for laboratory fungus propagation [55, 56]. Plates were incubated at 25 °C, shaking at 300 rpm. After 48 h, BacTiterGlo (Promega) was added to each well to quantify the amount of ATP present, indicative of fungal cell growth, and luminescence was measured with a 500 ms integration time.

3.5.3 Statistical Modeling

LC-MS/MS data of the *V. odorata* SCX library fractions (fractions 15-40 for cancer cell lines, 16-33 for *Fusarium*) was processed using the PepSAVI-MS software and filtered for compounds of interest by applying retention time, mass, and charge-state limits. MS features of the same charge state were binned together within a 0.05 Da window and features with (1) intensities <100 in the activity region, (2) charge states >+10, and (3) intensities >1% of

maximum peak intensity in non-activity regions were excluded. Data reduction and modeling were tailored to cancer cell line/Fusarium activity profile. Data filtering of SCX library fractions was as follows: MDA-MB-231 breast (15-24); PC-3 prostate (17-24); OVCAR-3 ovarian (17-23); *Fusarium* (23-31). Activity regions were defined for each cancer cell line: MDA-MB-231 breast (16-23); PC-3 prostate (18-23); OVCAR-3 ovarian (18-22), *Fusarium* (25-30). Resulting *m/z* meeting these criteria were modeled with an elastic net regression with a quadratic penalty parameter specification of 0.001.

3.5.4 Reversed-Phased Isolation of cyO8

CyO8-containing SCX fractions were subject to reversed-phase isolation on a Waters ACQUITY UPLC H-Class Bio System equipped with an Agilent Zorbax extended C18 column (20k PSI, 300 Å, 5 µm, 4.6 mm x 250 mm). Separation of cyO8 was achieved with a flow rate of 2 mL/min and a linear ramp of 5-25% mobile phase B from 0-5 min and 25-50% mobile phase B from 5-20 min (mobile phase A, water with 0.1% formic acid; mobile phase B, acetonitrile with 0.1% formic acid). Reversed-phase fractions were manually collected, dried down in a vacuum centrifuge, resuspended in water, and fractions containing cyO8 were combined for use in Fusarium conidia germination and cancer/fibroblast cell viability assays. CyO8 purity was assessed via LC-MS, and considering peaks >1% of the abundance of cyO8, purity was deemed to be 95% for the cancer cell line assays and >91% for the *Fusarium graminearum* assays (separate preparations) with no single impurity exceeding 2% (except doubly oxidized cyO8 at 3.8% in the *F. graminearum* MIC assay).

3.5.5 Cancer Cell Line/Fibroblast Viability Assay

MDA-MB-231 breast cancer cells, PC-3 prostate cancer cells, OVCAR-3 ovarian cancer cells cells (all cancer cell lines from ATCC, Manassas, VA), or human dermal fibroblasts (HDF;

Lonza Inc., Walkersville, MD, USA) were seeded in triplicate wells of a flat-bottom 96-well plate at a concentration of 5×10^3 cells/well. Cancer cells were in DMEM supplemented with 10% heat-inactivated fetal bovine serum, 100 U/mL penicillin, 100 µg/mL streptomycin, 2mM of L-glutamine, and 5 mM of HEPES buffer (all from Invitrogen) while HDFs were in Fibroblast Cell Basal Medium (Lonza) supplemented with 2% fetal bovine serum plus 0.1% insulin, 0.1% recombinant human basic fibroblast growth factor, and 0.1 % gentamicin sulfate/amphotericin, all from a proprietary Fibroblast Cell Medium BulletKit (Lonza). Cells were cultured for 12 h at 37^{o}C in a humidified atmosphere containing 10% CO_2 (cancer cells) or 5% CO_2 (HDF) in order to form adherent monolayers. The medium was then removed, replaced with fresh complete DMEM with or without cyO8, and the cells were cultured for an additional 24 h. Two hours before the end of culture, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) solution (Sigma Aldrich, Oakville, ON, Canada) was added to each well to a final concentration of 0.5 mg/mL to measure mitochondrial succinate dehydrogenase activity. After 2 h, culture supernatants were discarded and formazan crystals were solubilized in 0.1 mL of dimethyl sulfoxide. Absorbance was measured at 570 nm using an ASYS Expert 96 microplate reader (Montreal Biotech Inc., Kirkland, QC, Canada). Average absorbance in each group was normalized relative to the medium control and the percent reduction in viability, as reflected by the change in mitochondrial succinate dehydrogenase activity, was determined.

3.5.6 Minimum Inhibitory Concentration Assay

MDA-MB-231 breast cancer cells, PC-3 prostate cancer cells, OVCAR-3 ovarian cancer cells, or HDFs were cultured as described above in the absence or presence of different concentrations of cyO8 and inhibitory concentration (IC_{50}) values of cyO8 were calculated from the resulting dose response curves. To determine the cyO8 MIC against *Fusarium graminearum*,

a two-fold serial dilution was performed in a 96-well plate format. CyO8 concentration was determined using a NanoDrop and an extinction coefficient of 7365 $M^{-1}cm^{-1.}$ In each well, 30 µL minimal media, 10 µL Fusarium spore culture, and 10 µL purified cyO8 was added (for a final concentration of 12, 24, or 48 µM), and the plate was incubated at 22 °C for 72 h, shaking at 250 rpm, and assessed visually for MIC.

3.5.7 Reduction, Alkylation, and Glu-C digest of *V. odorata* library

V. odorata peptide library fractions were reduced with 10 mM dithiothreitol (Sigma-Aldrich) at 45 °C, 850 rpm, 30 min, and alkylated with 100 mM iodoacetamide (Sigma –Aldrich) at 25 ^{O}C, 850 rpm, 15 min. Pierce C18 spin columns (Thermo Scientific) were used to remove interfering contaminants from the *V. odorata* samples, which were subsequently dried down in a vacuum concentrator and resuspended in acidified water (0.1% formic acid) for LC-MS/MS analysis. Digested *V. odorata* samples were reduced and alkylated as described above, then incubated 1:50 enzyme: substrate with endoproteinase Glu-C enzyme (Sigma) at 37 °C for 3 h.

3.5.8 LC-MS/MS Analysis of *V. odorata* library

Five microliters of each intact, reduced/alkylated, and reduced/alkylated/glu-C digested *V. odorata* sample were injected onto a nano-LC-ESI-MS/MS platform: nanoAcquity (Waters, Milford, MA) coupled to a TripleTOF5600 (AB Sciex, Framingham, MA). Samples were initially injected onto a trap column (NanoAcquity UPLC 2G-W/M Trap 5 µm Symmetry C18, 180 µm × 20 mm: Waters) before entering the analytical C18 column (10k PSI, 100 Å, 1.8 µm, 75 µm × 250 mm: Waters). LC separation of peptides was achieved with a flow rate of 0.3 µL/min and a linear ramp of 5%−50% B (mobile phase A, 1% formic acid in water; mobile phase B, 1% formic acid in acetonitrile) over 30 min. The MS was operated in positive-ion, high-sensitivity mode with the MS survey spectrum using a mass range of 350−1600 Da in 250 ms

and information-dependent acquisition (IDA) of CID MS/MS data. The first 20 features above 150 counts threshold with a charge state of +2 to +5 were fragmented using rolling collision energy (±5%). Deisotoped peak lists for the intact and reduced/alkylated samples were generated using Progenesis QI for Proteomics software (Nonlinear Dynamics, v.2.0) with a retention time filter of 14−45 min. "Peptide ion data" was exported from Progenesis and analyzed for the characteristic reduction/alkylation mass shift corresponding to three disulfide bonds (348.16 Da) using Python.

3.5.9 UVPD Analysis of cyO8

Reduced and alkylated cy08 was targeted for analysis following loading of a complex mixture of cyclotides on an in-house packed C-18 (3.5 μm particle size, X-Bridge) trap column (3.5 cm New Objective Integrafrit capillary, 100 μm i.d.) and separation using an in-house packed C-18 (3.5 μm, X-Bridge) analytical column (15 cm New Objective Picofrit capillary, 75 μm i.d.) on a Ultimate 3000 UHPLC system (ThermoFisher Scientific). Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid at a net flow rate of 300 nL/min. The applied gradient is as follows: 0-8 min at 2% B; 8-10 min increasing linearly from 2 to 20% B; 10-32 min increasing linearly from 20 to 40% B; rapid increase to 80% B at 32 min and holding until 34 min; 34 min decreasing to 2% B to return the mobile phase to its initial condition. 1 μL of sample at a concentration of approximately 1 mM total protein content was injected on column per experiment. The chromatography system was coupled to Thermo Scientific Orbitrap Fusion Lumos mass spectrometer (San Jose, CA, USA) outfitted with a 193 nm excimer laser (Coherent Excistar 500 Hz, Santa Clara, CA, USA) to allow photodissociation, as implemented and described previously.[57] UVPD was performed using 2 pulses at 2 mJ per pulse in the low pressure linear

ion trap. The Orbitrap mass spectrometer was operated in the positive ion mode with a spray voltage of 1800 V. The 3+ and 4+ charge states of cy08 were targeted during the experiments with a 100 ppm mass tolerance during the MS1 spectrum (60000 FT resolving power, 5 x 10^5 AGC target). The MS/MS spectra were collected using the following parameters: quadrupole isolation with an isolation width of 4 m/z; 15000 FT resolving power; 2 x 10^5 AGC target; 2 μscans averaged per spectrum. Theoretical fragment libraries were generated with mMass, an open source mass spectrometry tool to aid in the characterization of cyclic and linear peptides.[58-60] Theoretical fragments were manually assigned to cyO8 MS^2 peaks within 30 ppm error.

3.6 Tables

Table 3.1. Putative cyclotide species identified with a reduction/alkylation mass shift experiment and assays in which mass was ranked as a potential anticancer/antifungal contributor. Masses identified in recent cyclotide screens are indicated. All masses can be matched to known cyclotide masses in other botanical species; however, this is not necessarily an indication of identity.

Mass (Da)	Assay	Detected in screen
3597.55		Yes
3568.68		Yes
3514.60		
3499.54		
3497.54		Yes
3483.49		
3440.51		
3364.41		
3350.47		
3344.68		Yes
3332.44		
3330.48		
3297.37		Yes
3269.37		Yes
3257.37	Prostate cancer/ *Fusarium*	
3243.45		Yes
3208.41		
3186.42		Yes
3184.39	Breast/Prostate/ Ovarian/*Fusarium*	
3171.38	Ovarian	Yes
3170.38	Breast	
3166.48	Breast/Prostate/ *Fusarium*	
3165.49	Ovarian/Prostate	
3156.36	*Fusarium*	Yes
3154.38	*Fusarium*	

3.7 Figures

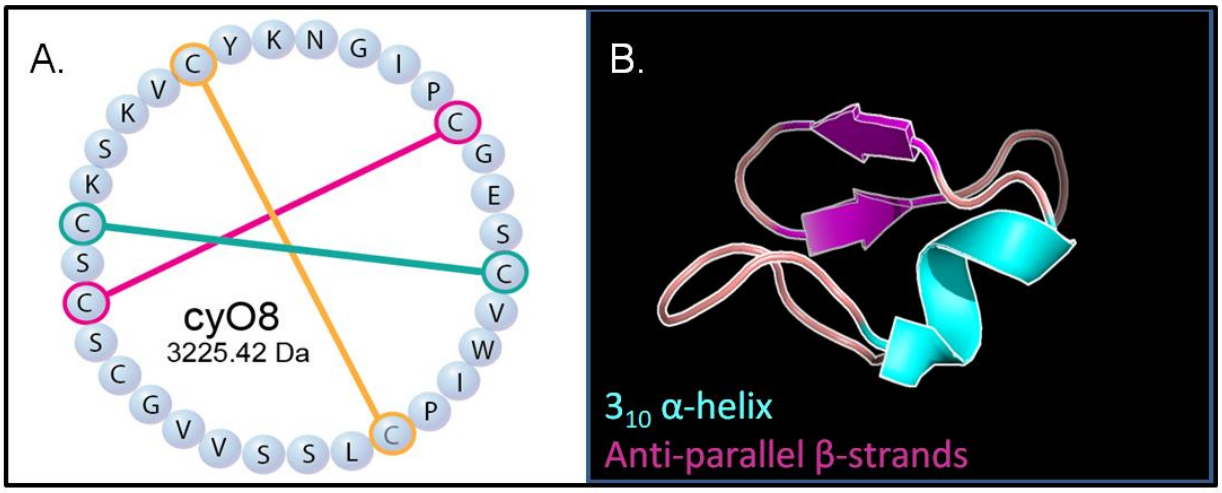


Figure 3.1. Cycloviolacin O8 (cyO8), a known anthelmintic cyclotide from *Viola odorata*. A. CyO8 sequence, MW 3225.42 Da, showing disulfides. B. Predicted three-dimensional cyO8 structure, not showing disulfides. Magenta arrows and cyan domain represent anti-paralled β-sheets and a 3_{10} α-helix, respectively (cybase.org (Mulvenna et al., 2006)); figure generated in Pymol).

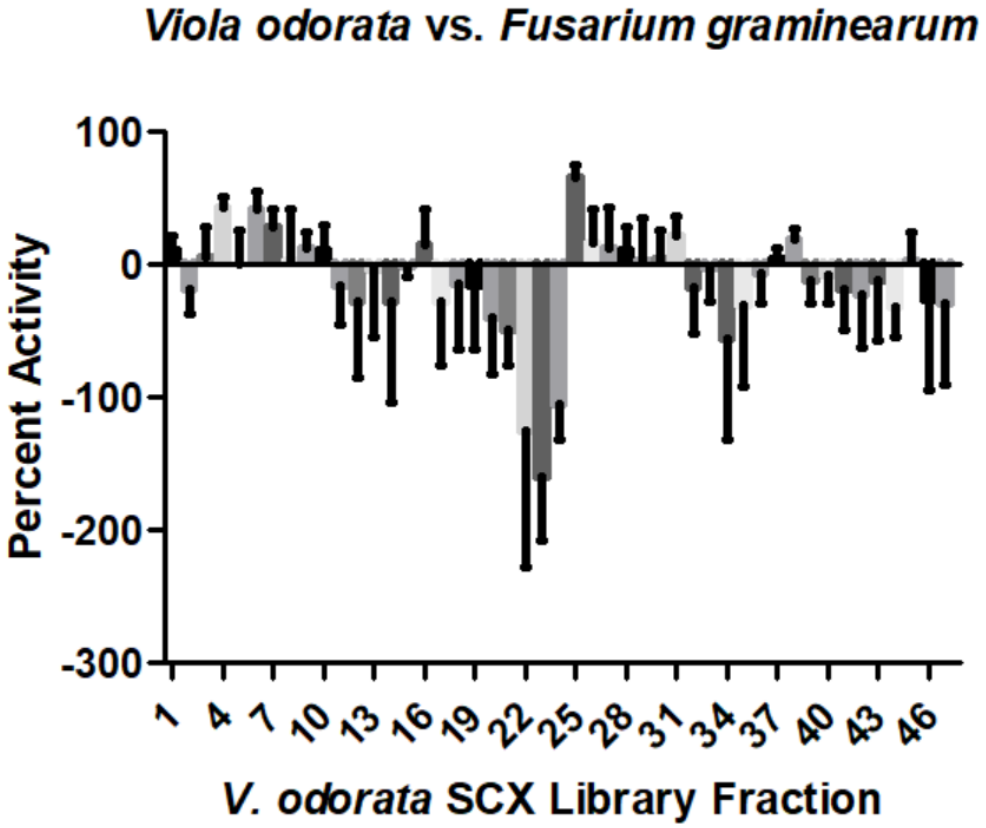


Figure 3.2. *Viola odorata* SCX fraction library versus *Fusarium graminearum* bioactivity assay. Data are represented as mean ± SD of three replicates. Antifungal bioactivity is seen in *V. odorata* fractions 25-31.

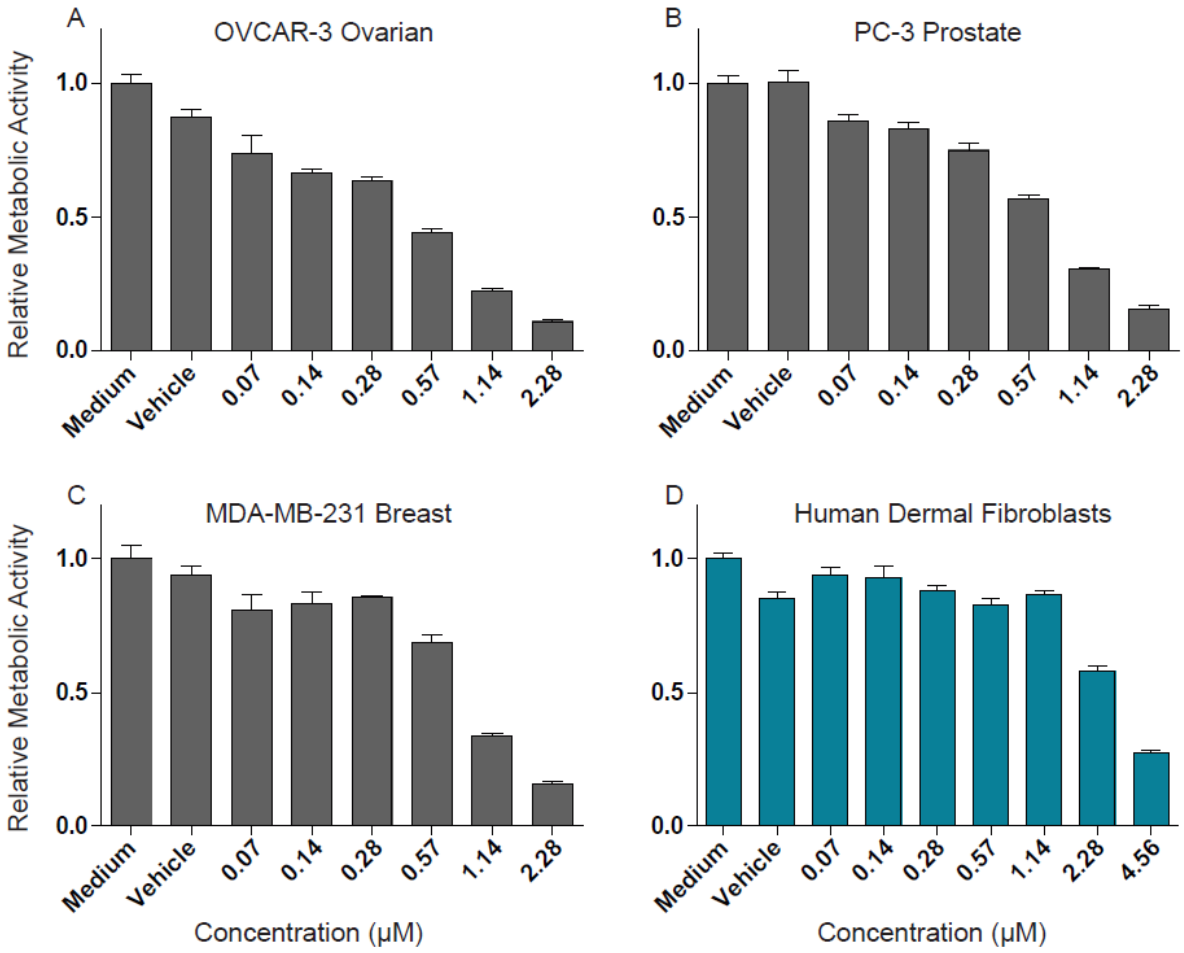


Figure 3.3. Inhibitory concentration (IC_{50}) determination of cyO8 against human cancer cell lines. Purified cyO8 tested against human cancer cell lines, OVCAR-3 ovarian (A), PC-3 prostate (B), and MDA-MB-231 breast (C), and against non-cancerous human dermal fibroblasts (D). Vehicle indicates cells are cultured with water in place of cyO8. Data are represented as mean ± SD of three replicates.

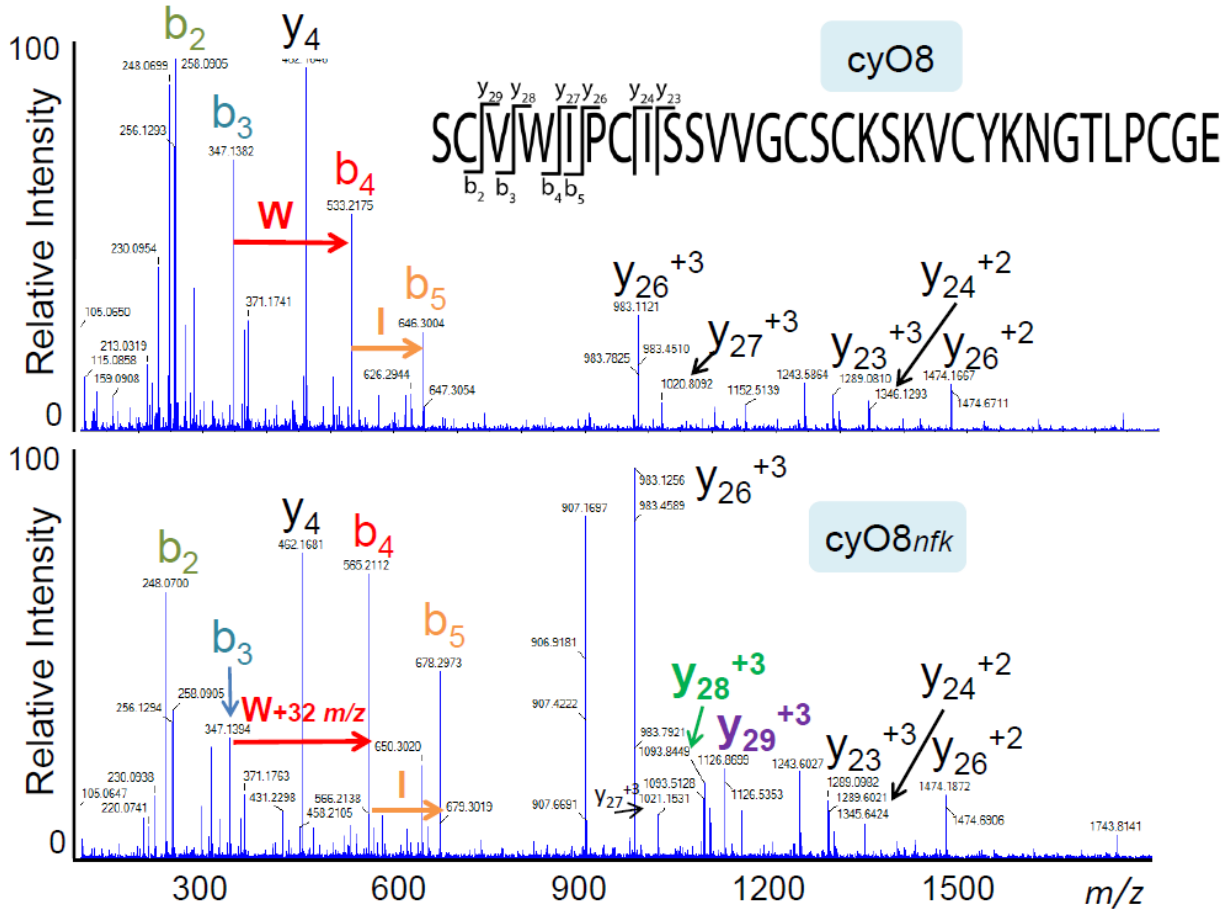


Figure 3.4. Collision induced dissociation (CID) fragmentation of cyO8. CID MS^2 spectra of reduced, alkylated, gluC digested cyO8 (+4) (top) and reduced, alkylated, gluC digested cyclotide mass 3257.37 Da (+4) (bottom) showing 32 Da double-oxidation mass shift localized to the tryptophan residue of $cyO8_{nfk}$ (cyO8 N-formylkynurenine). Fragment ions are colored the same in both spectra to aid in quick identification.

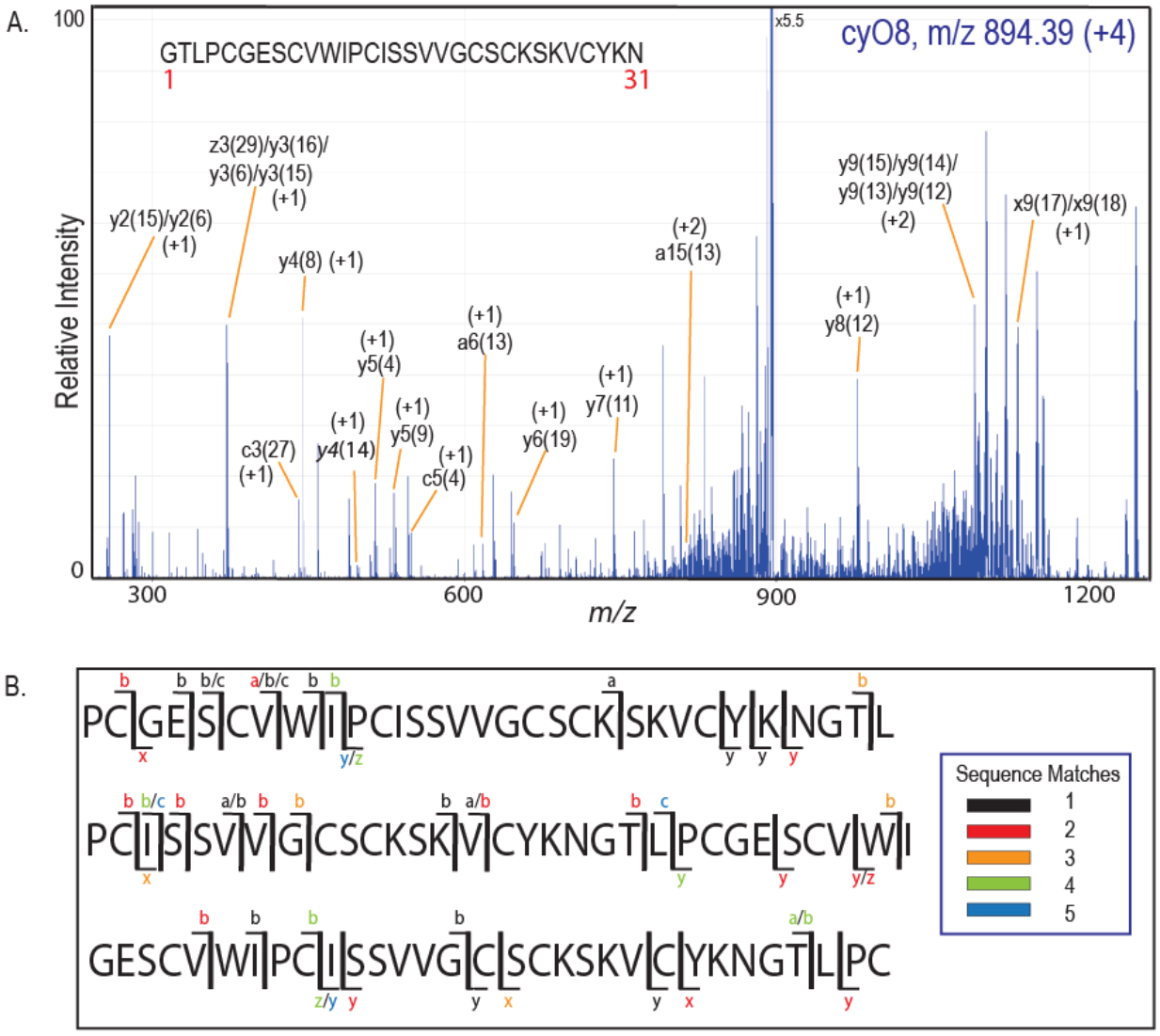


Figure 3.5. Ultraviolet Photodissociation (UVPD) fragmentation of cyO8. A. MS/MS of (+4, *m/z* 894.39) cyO8 (2 pulses, 2 mJ/pulse) shows abundant, low intensity fragment ions. Fragments are labeled by the fragment ion type and the numeric index (in parentheses) from the sequence map shown indicating the initial cleavage site (e.g. *a15*(13) indicates an initial cleavage between residues 12 and 13 and a second fragmentation event resulting in an *a15* fragment ion). B. Representative cyO8 sequence iterations showing UVPD MS^2 coverage. Ion types are color-coded based on the number of ambiguous fragment ion matches.

REFERENCES

1. Ajesh, K. and K. Sreejith, *Peptide antibiotics: an alternative and effective antimicrobial strategy to circumvent fungal infections.* Peptides, 2009. **30**(5): p. 999-1006.

2. Tassanakajon, A., K. Somboonwiwat, and P. Amparyup, *Sequence diversity and evolution of antimicrobial peptides in invertebrates.* Dev Comp Immunol, 2015. **48**(2): p. 324-41.

3. Peschel, A. and H.G. Sahl, *The co-evolution of host cationic antimicrobial peptides and microbial resistance.* Nat Rev Microbiol, 2006. **4**(7): p. 529-36.

4. Pasupuleti, M., A. Schmidtchen, and M. Malmsten, *Antimicrobial peptides: key components of the innate immune system.* Crit Rev Biotechnol, 2012. **32**(2): p. 143-71.

5. Kommineni, S., et al., *Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract.* Nature, 2015. **526**(7575): p. 719-22.

6. Ganz, T., *The role of antimicrobial peptides in innate immunity.* Integr Comp Biol, 2003. **43**(2): p. 300-4.

7. Mahlapuu, M., et al., *Antimicrobial Peptides: An Emerging Category of Therapeutic Agents.* Front Cell Infect Microbiol, 2016. **6**: p. 194.

8. Loo, S., et al., *Bleogens: Cactus-Derived Anti-Candida Cysteine-Rich Peptides with Three Different Precursor Arrangements.* Front Plant Sci, 2017. **8**: p. 2162.

9. Zhang, L.J. and R.L. Gallo, *Antimicrobial peptides.* Curr Biol, 2016. **26**(1): p. R14-9.

10. Sato, H. and J.B. Feix, *Peptide-membrane interactions and mechanisms of membrane destruction by amphipathic alpha-helical antimicrobial peptides.* Biochim Biophys Acta, 2006. **1758**(9): p. 1245-56.

11. Peters, B.M., M.E. Shirtliff, and M.A. Jabra-Rizk, *Antimicrobial peptides: primeval molecules or future drugs?* PLoS Pathog, 2010. **6**(10): p. e1001067.

12. Herbel, V. and M. Wink, *Mode of action and membrane specificity of the antimicrobial peptide snakin-2.* PeerJ, 2016. **4**: p. e1987.

13. Lee, D.W. and B.S. Kim, *Antimicrobial cyclic peptides for plant disease control.* Plant Pathol J, 2015. **31**(1): p. 1-11.

14. Guzman-Rodriguez, J.J., et al., *Plant antimicrobial peptides as potential anticancer agents.* Biomed Res Int, 2015. **2015**: p. 735087.

15. Rotem, S. and A. Mor, *Antimicrobial peptide mimics for improved therapeutic properties.* Biochim Biophys Acta, 2009. **1788**(8): p. 1582-92.

16. Brogden, K.A., *Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?* Nat Rev Microbiol, 2005. **3**(3): p. 238-50.

17. Bahar, A.A. and D. Ren, *Antimicrobial peptides.* Pharmaceuticals (Basel), 2013. **6**(12): p. 1543-75.

18. Craik, D.J., et al., *Ribosomally-synthesised cyclic peptides from plants as drug leads and pharmaceutical scaffolds.* Bioorg Med Chem, 2017.

19. Kirkpatrick, C.L., et al., *The "PepSAVI-MS" Pipeline for Natural Product Bioactive Peptide Discovery.* Anal Chem, 2017. **89**(2): p. 1194-1201.

20. Kirkpatrick, C.L., et al., *Fungal Secretome Analysis via PepSAVI-MS: Identification of the Bioactive Peptide KP4 from Ustilago maydis.* J Am Soc Mass Spectrom, 2018.

21. Britton, E.R., et al., *Biochemometrics to Identify Synergists and Additives from Botanical Medicines: A Case Study with Hydrastis canadensis (Goldenseal).* J Nat Prod, 2018. **81**(3): p. 484-493.

22. Scheffler, R.J., et al., *Antimicrobials, drug discovery, and genome mining.* Appl Microbiol Biotechnol, 2013. **97**(3): p. 969-78.

23. Gerlach, S.L., et al., *Anticancer and chemosensitizing abilities of cycloviolacin 02 from Viola odorata and psyle cyclotides from Psychotria leptothyrsa.* Biopolymers, 2010. **94**(5): p. 617-25.

24. Burman, R., et al., *Evaluation of toxicity and antitumor activity of cycloviolacin O2 in mice.* Biopolymers, 2010. **94**(5): p. 626-34.

25. Pranting, M., et al., *The cyclotide cycloviolacin O2 from Viola odorata has potent bactericidal activity against Gram-negative bacteria.* J Antimicrob Chemother, 2010. **65**(9): p. 1964-71.

26. Wang, C.K., et al., *Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis.* J Nat Prod, 2008. **71**(1): p. 47-52.

27. Plan, M.R., et al., *Backbone cyclised peptides from plants show molluscicidal activity against the rice pest Pomacea canaliculata (golden apple snail).* J Agric Food Chem, 2008. **56**(13): p. 5237-41.

28. Fensterseifer, I.C., et al., *Effects of cyclotides against cutaneous infections caused by Staphylococcus aureus.* Peptides, 2015. **63**: p. 38-42.

29. Craik, D.J., et al., *Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif.* J Mol Biol, 1999. **294**(5): p. 1327-36.

30. Burman, R., et al., *Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae.* Front Plant Sci, 2015. **6**: p. 855.

31. Narayani, M., A. Chadha, and S. Srivastava, *Cyclotides from the Indian Medicinal Plant Viola odorata (Banafsha): Identification and Characterization.* J Nat Prod, 2017. **80**(7): p. 1972-1980.

32. Gruber, C.W., et al., *Distribution and evolution of circular miniproteins in flowering plants.* Plant Cell, 2008. **20**(9): p. 2471-83.

33. Hellinger, R., et al., *Peptidomics of Circular Cysteine-Rich Plant Peptides: Analysis of the Diversity of Cyclotides from Viola tricolor by Transcriptome and Proteome Mining.* J Proteome Res, 2015. **14**(11): p. 4851-62.

34. Camarero, J.A., *Cyclotides, a versatile ultrastable micro-protein scaffold for biotechnological applications.* Bioorg Med Chem Lett, 2017. **27**(23): p. 5089-5099.

35. Thell, K., et al., *Oral activity of a nature-derived cyclic peptide for the treatment of multiple sclerosis.* Proc Natl Acad Sci U S A, 2016. **113**(15): p. 3960-5.

36. White, A.M. and D.J. Craik, *Discovery and optimization of peptide macrocycles.* Expert Opin Drug Discov, 2016. **11**(12): p. 1151-1163.

37. Wong, C.T., et al., *Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment.* Angew Chem Int Ed Engl, 2012. **51**(23): p. 5620-4.

38. Henriques, S.T., et al., *The Prototypic Cyclotide Kalata B1 Has a Unique Mechanism of Entering Cells.* Chem Biol, 2015. **22**(8): p. 1087-97.

39. APVMA, *PUBLIC RELEASE SUMMARY on the evaluation of the new active Clitoria ternatea in the product Sero-X Insecticide*. 2016, Australian Pesticides and Veterinary Medicines Authority: Kingston, Australia. p. 28.

40. Poth, A.G., et al., *Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins.* ACS Chem Biol, 2011. **6**(4): p. 345-55.

41. Chan, L.Y., et al., *A new family of cystine knot peptides from the seeds of Momordica cochinchinensis.* Peptides, 2013. **39**: p. 29-35.

42. Koehbach, J., et al., *Cyclotide discovery in Gentianales revisited--identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants.* Biopolymers, 2013. **100**(5): p. 438-52.

43. Mohimani, H., et al., *Sequencing cyclic peptides by multistage mass spectrometry.* Proteomics, 2011. **11**(18): p. 3642-50.

44. Shaw, J.B., et al., *Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation.* J Am Chem Soc, 2013. **135**(34): p. 12646-51.

45. Crittenden, C.M., et al., *Exploitation of the Ornithine Effect Enhances Characterization of Stapled and Cyclic Peptides.* J Am Soc Mass Spectrom, 2016. **27**(5): p. 856-63.

46. Ireland, D.C., M.L. Colgrave, and D.J. Craik, *A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability.* Biochem J, 2006. **400**(1): p. 1-12.

47. Colgrave, M.L., et al., *The anthelmintic activity of the cyclotides: natural variants with enhanced activity.* Chembiochem, 2008. **9**(12): p. 1939-45.

48. Brodbelt, J.S., *Photodissociation mass spectrometry: new tools for characterization of biological molecules.* Chem Soc Rev, 2014. **43**(8): p. 2757-83.

49. Stromstedt, A.A., et al., *Bactericidal activity of cyclotides where phosphatidylethanolamine-lipid selectivity determines antimicrobial spectra.* Biochim Biophys Acta, 2017. **1859**(10): p. 1986-2000.

50. Tam, J.P., et al., *An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 8913-8.

51. Plan, M.R., et al., *The cyclotide fingerprint in oldenlandia affinis: elucidation of chemically modified, linear and novel macrocyclic peptides.* Chembiochem, 2007. **8**(9): p. 1001-11.

52. Burman, R., et al., *Cytotoxic potency of small macrocyclic knot proteins: structure-activity and mechanistic studies of native and chemically modified cyclotides.* Org Biomol Chem, 2011. **9**(11): p. 4306-14.

53. Ravipati, A.S., et al., *Understanding the Diversity and Distribution of Cyclotides from Plants of Varied Genetic Origin.* J Nat Prod, 2017. **80**(5): p. 1522-1530.

54. Robotham, S.A., et al., *UVnovo: A de Novo Sequencing Algorithm Using Single Series of Fragment Ions via Chromophore Tagging and 351 nm Ultraviolet Photodissociation Mass Spectrometry.* Anal Chem, 2016. **88**(7): p. 3990-7.

55. Leslie, J.F. and B.A. Summerell, *The Fusarium Laboratory Manual, first ed*. first ed ed. 2006, Iowa: Blackwell Publishing. 388.

56. Correll, J.C., C.J.R. Klittich, and J.F. Leslie, *Nitrate non-utilizing mutants of Fusarium oxysporum and their use in vegetative compatibility tests.* Phytopathology, 1987. **77**: p. 1640-1646.

57. Klein, D.R., D.D. Holden, and J.S. Brodbelt, *Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry.* Anal Chem, 2016. **88**(1): p. 1044-51.

58. Niedermeyer, T.H. and M. Strohalm, *mMass as a software tool for the annotation of cyclic peptide tandem mass spectra.* PLoS One, 2012. **7**(9): p. e44913.

59. Strohalm, M., et al., *mMass data miner: an open source alternative for mass spectrometric data analysis.* Rapid Commun Mass Spectrom, 2008. **22**(6): p. 905-8.

60. Strohalm, M., et al., *mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data.* Anal Chem, 2010. **82**(11): p. 4648-51.

61. Mulvenna, J.P., C. Wang, and D.J. Craik, *CyBase: a database of cyclic protein sequence and structure.* Nucleic Acids Res, 2006. **34**(Database issue): p. D192-4.

CHAPTER 4: Viola "inconspicua" No More: An Analysis of Antibacterial Cyclotides

4.1.1 Introduction

As the world rapidly evolves into a post-antibiotic era, it is necessary to examine the successes and limitations of standard approaches for new lead discovery, including re-evaluation of previous and current traditional medicines. Botanical remedies derived from the >500-membered genus *Viola* (Violaceae) have been a mainstay of Chinese,[1] Middle Eastern,[2] Indian,[3] and African[3,4] ethnomedicine systems for generations. *Viola* essential oils, extracts, and poultices were used classically in the treatment of bacterial and viral infections (cough, cold, fever), in addition to inflammation (skin disease, migraine) and oxidative stress (cancer, hypertension).[3] True to age-old beliefs, modern examination of the chemical constituents present in *Viola spp.* has revealed a wealth of bioactive secondary metabolites with broad-spectrum antibacterial activity,[5,6] inhibition of inflammation pathways,[7] antioxidant activity towards the treatment of hypertension[8] and oxidative stress,[9] as well as selective cytotoxic effects on drug-resistant[10] and triple-negative[11] cancers. The impressive biochemical repertoire presented by this genus highlights its clinical relevance and suggests that further exploration of its bioactive potential may unearth additional potent and diverse biological activities.

Numerous structurally and functionally distinct molecules orchestrate the biological activities characteristic of *Viola spp.*; however, a group of small ribosomally synthesized cyclic plant

peptides, the cyclotides, are a primary driver of this bioactive powerhouse genus. Cyclotides are defined by their unique three-dimensional structure - the highly stable cyclic cysteine knot motif (CCK) generated via three conserved disulfides and a cyclic backbone forged through a standard peptide bond.[12] Six "loops" or sequence regions exist between the six cysteine residues and are numbered consecutively from the first cysteine residue after the N-terminal cyclic ligation site to the C-terminus. Loops 1 and 4 are largely conserved among known cyclotide variants, while loop 6 has been subject to substantial evolutionary diversity, contributing to >400 distinct cyclotide species (Cybase.org), with innumerous variations thought to await discovery.[13-15] Cyclotide expression, both in quantity and identity, is highly dynamic among different botanical species and even within the same species arising from seasonal variation, environmental conditions, geographical location, and external cues, dramatically increasing the sequence space sampled.[16,17] Novel cyclotide motifs are often emerging,[18] suggesting that current attempts to elucidate the extent of sequence diversity in this peptide class have barely scratched the surface of variations that may exist in Nature.

As diverse as their sequences, cyclotides exhibit a range of robust biological activities. Discovery of the first cyclotide, kalata B1, followed the observation that African women consumed *Oldenlandia affinis* tea to speed childbirth and the uterotonic activities of this concoction later attributed to its cyclotide content.[19] Subsequent characterization of numerous additional cyclotide species over the last three decades has illuminated a world of intrinsically bioactive compounds. Believed to participate in host defense,[20,21] the antimicrobial[22] and insecticidal activities demonstrated by cyclotide family have made these fascinating peptides obvious candidates for potential clinical and agricultural use. As membranolytic pore-formers, cyclotides achieve cytotoxicity via electrostatic-driven interactions and an amphipathic tertiary

structure to facilitate target membrane disruption. Affecting cells with net negative membrane potentials, the largely nondiscriminatory cyclotide mechanism of action has led to additional clinically important anti-HIV[23,24] and antineoplastic bioactivities.[11,25-27]

Viola inconspicua Blume is a weedy perennial prone to rapid dispersion through abundant cleistogamous flowers distributed among the grasslands, fields, and forests of Asia. This commonly hybridized *Viola* species is used in Chinese herbal remedies and was recently implicated in antifungal activity against the dermatophyte *Trichophyton rubrum.*[28] As few as three cyclotide sequences (kalata S, vinc A, vinc B) have been identified in *V. Inconspicua,*[29,30] thus warranting further examination of this underexplored botanical species to reveal previously uncharacterized cyclotide sequences.

Herein, analysis of *V. inconspicua* revealed 41 uncharacterized cyclotide candidates and the known cyclotides viba 11, cyO8 and cyO9. To assess bioactivity of this cyclotide-rich species, a *V. inconspicua* peptide library was assayed against *E. coli* ATCC 25922 and the highly virulent and rifampin/streptomycin resistant pathogen *Klebsiella pneumoniae* VK148[31] Cyclotide-containing library fractions demonstrated robust antibacterial bioactivity and the most abundant putative cyclotides masses in these fractions were prioritized for sequence characterization. Six novel cyclotide sequences were revealed (cycloviolacins I1-I6, cyI1-cyI6), four of which contain a previously unknown “TLNGNPGA” motif in the hypervariable loop 6 region.

Cyclotide sequences containing this novel motif were characterized for biological activity and secondary structure. Antibacterial activities of cyI3-cyI6 library fractions and isolated cyclotide species cyI3 and cyI4 were investigated. Determination of minimum inhibitory concentrations of cyclotide mixtures may begin to elucidate the synergistic effects among the numerous cyclotide species expressed in planta, likely enhancing the bioactive potencies of

individual cyclotide members. Molecular modeling of the cyclotide cycloviolacin I3 (cyI3, 3378 Da) with Robetta[32] suggests these novel-motif containing species maintain secondary and tertiary structure to known NMR-based cyclotide structures.

4.2 Results and Discussion

4.2.1 Discovery of Novel *V. inconspicua* Cyclotides

Identification of putative cyclotide masses in complex botanical extracts has been achieved routinely via mass spectrometry, where the reduction and alkylation of three conserved disulfides characteristic of cyclotides produces a telltale mass shift of +348.16 Da.[33,34] This analysis of the *V. inconspicua* peptidome resulted in 41 putative cyclotide masses and three previously identified cyclotides, Viba 11 (*Viola tricolor, Viola baoshanensis, Viola philippica*), cyO8 (*Viola odorata, Viola adunca),* and cyO9 (*Viola odorata, Viola biflora)* (Table S1, Supporting Information). This novel complement of cyclotide-like masses suggests a much larger population of cyclotidyl species present in *V. inconspicua* than reported in the literature. Although the 41 masses identified via reduction alkylation mass shift analysis have not yet been mapped to specific cyclotides in *V. inconspicua,* 10 of these putative cyclotide masses are isobaric with cyclotides known in other botanical species (Cybase.org). Without full sequence characterization, however, it is not possible to confirm the identity of these masses.

4.2.2 Exploration of *V. inconspicua* Antibacterial Activities

Sequence diversity within cyclotides is a likely contributor to the broad range of activities exhibited by this bioactive family of molecules. *V. inconspicua* reversed-phase peptide libraries produced regions of robust bioactivity (50-100 %) against the Gram-negative bacteria *E. coli* ATCC 25922 and *K. pneumoniae* VK148 (Figure 1). LC-MS/MS analysis of this peptide library revealed bioactive fractions to be highly populated with both novel and known cyclotides (Figure

2), specifically library fractions 24, 25, 27, and 28. Six dominant and uncharacterized masses in these bioactive regions [3236.4 Da (cycloviolacin I1, cyI1), 3259.4 Da (cyI2), 3378.5 Da (cyI3), 3391.4 Da (cyI4), 3406.5 Da (cyI5), and 3433.5 Da (cyI6)] were thus prioritized for molecular characterization and validation of bioactivity.

4.2.3 Sequence Characterization of Six Novel Cyclotides

Multiple material harvests were required to amass sufficient material for characterization, and notably, abundances of putative and confirmed cyclotide sequences identified in this study were variable, as has been described in a previous study.[35] The cause of these fluctuations, whether seasonal, temperature-related, or pathogen-induced, was indiscernible. This highlights the dynamic nature of cyclotide expression and suggests that cyclotides described in current literature are only a snapshot of possible sequences that exist in Nature.

The six targeted masses (cyI1-cyI6) were sequenced via LC-MS/MS analysis of reduced, alkylated, and Glu-C enzymatic digests via de novo sequence assignment. Assigned cyO8/cyO9 MS/MS spectra aided the de novo sequencing of cyclotides cyI1-2 (Figures S1 and S2, Supporting Information), whereas MS/MS fragment ions of the highly abundant cyclotide cyI3 (Figure 3) were used to guide the sequence elucidation of cyclotides 4-6 (Figures S3-S5, Supporting Information). Low abundance MS/MS fragmentation within loop five required additional sequence verification with chymotrypsin digestion and LC-MS/MS analysis. Additionally, chymotrypsin selectivity for leucine residues and homology with known cyclotide sequences were used to discern isoleucine/leucine identities in several positions.

Four newly found sequences (cyI3-cyI6) demonstrated a novel "TLNGNPGAC" motif, deemed motif "Type I," in the hypervariable loop 6 and were on the high end of the mass range typically exhibited by cyclotides (~2800-3700 Da, with most sequences clustering under 3400

Da [33]) (Table 1). The remaining two sequences (cyI1 and cyI2) were more homologous to known cyclotides cyO8/cyO9 based on cluster analysis, motif "Type II" (Figure 4). Several parallels exist among the novel sequences identified in this work and the cytotoxic vibi cyclotides from the alpine violet, *Viola biflora.*[36] CyI1 (3236.4 Da) was found to contain a tyrosine residue in loop two often occupied by a tryptophan residue (as in known cyclotides vibi H), a conserved mutation unlikely to affect activity. Cyclotides cyI1 and cyI2 contain a 'GTXPCGE' motif, but contain a phenylalanine residue in the third position similar to only two other known cyclotides, vibi G and vibi J. Often occupied primarily by hydrophobic residues, loop three of cyclotide cyI6 contains a basic residue (lysine), common to only vibi J and ~10 other known cyclotides[36] (cybase.org). All cyclotides cyI1-cyI6 contain an asparagine residue in loop 6, a critical feature involved in the intracellular cyclization process and highly conserved throughout cyclic cyclotide variants.[37-39] Interestingly, Type II cyclotides lack the loop 6 basic residue found in most cyclotide species; chemically masking this lysine or arginine residue decreases the activity of cyclotides cyO2 and varv A[40] and may contribute to the weak activity of cyI3-cyI6.

Examination of the tandem mass spectrometry data revealed that the two distinct sequence types, i.e. Type I and II motifs, detected in *V. inconspicua* are discernable via characteristic MS/MS fingerprint ions (Figure 5). The previously uncharacterized high molecular weight motifs represented by Type II cyclotides (cyI3-cyI6) were easily identified by the distinct C-terminus containing y-ions resulting from commonly used collision-induced dissociation (CID) fragmentation, *m/z* 590.2, 365.1, and 205.1, whereas Type I sequences are indicated by the highly abundant N-terminus containing b-ion series *m/z* 248.1, 347.1, 533.2, 646.3. Sample losses from reduction, alkylation, and enzymatic digestion often complicate sequence elucidation of novel cyclotide species. Using these highly abundant *m/z* fingerprints, MS/MS data can

provide us with some information about mass identity and sequence type, though overlap in the MS1 may lead to ambiguity from "mixed" populations of Type I and II cyclotide fingerprints (Table S1, Supporting Information). Additionally, reversed-phase C_{18} retention times of cyO8/cyO9-like sequences cyI1 and cyI2 (Type I) were distinct from those containing the novel loop 6 motif, cyI3-cyI6 (Type II) (Figure 2). Interestingly, both types are similar in hydrophobic content and isoelectric point, however, the distribution of these residues may modify the tertiary structure. For example, a tyrosine residue found in position one of loop six in Type I sequences is in position one of loop three of type II cyclotides (Figure 4).

4.2.4 Bioactivity Assessment

Due to the high sequence homology of cyclotides cyI3-cyI6, chromatographic isolation of quantities sufficient for bioactivity characterization of individual cyclotides was challenging. However, the early reversed-phase retention time of cyI4 allowed separation from other novel Type II cyclotides cyI3, 5, and 6 in enough quantity for bioactivity assays against both *E. coli* and *K. pneumoniae*, and the high abundance of cyI3 allowed for isolation of enough material for bioactivity testing solely against *E. coli*. Minimum inhibitory concentration of isolated cyI4 was not achieved at a concentration > 60 μM against both *E. coli* and *K.* pneumoniae and cyI3 exhibited ~50% activity against *E. coli* at 60 μM. In contrast, fractions predominantly containing cyclotides cyI3-cyI6 demonstrated an MIC of 18 μM and 35 μM against *E. coli* ATCC and *K. pneumoniae*, respectively. As the library fractions generated in this work are representative of the relative concentrations found in physiological conditions of the botanical specimens used, fractions containing mixed cyclotide species allow for synergistic interactions potentially occurring in Nature. Though it is posited that the ability of cyclotides to self-associate in solution may affect membranolytic activity,[41] synergistic effects of different cyclotide sequences on

bioactivity is largely unexplored.[42] Regardless, cyclotide sequence diversity seen in planta is likely a multifaceted defense strategy evolved to combat diverse and seasonal stress conditions,[33] supported by the observation that small sequence changes influence and tune cyclotide hemolytic ability, and thus target specificity.[35] Interestingly, the observed bioactivity of Type II cyclotides is significantly less potent than the bactericidal cyclotide cyO2 (MIC 5 and <12.5 μM against *E. coli* and *K. pneumoniae*, respectively).

All new cyclotides were found to belong to the bracelet subfamily, defined by a high isoelectric point and the lack of a *cis*-proline in loop 5. Unlike non-ribosomally synthesized bacterial-derived cyclic peptides, which tend to be smaller in size, cyclotides can form various elements of secondary structure. Molecular modeling of cyclotide cyI3 with Robetta generated five homology-based models and predicted an overall three-dimensional and secondary structure (a short α-helix in loop 3 and anti-parallel β-sheets with loops 4 and 5) similar to that of cycloviolcin O2 (cyO2), the prototypical bracelet cyclotide (Figure 6).

4.2.5 Concluding Remarks

New compounds with potent bioactivity and novel mechanisms of action are needed to supplement the increasingly ineffective antimicrobial therapeutics used in the modern clinic, particularly those with efficacy against Gram-negative bacterial species. Herein, we have established the antibacterial activity of a *V. inconspicua* peptide library against the Gram-negative pathogens *E. coli* ATCC 25922 and *K. pneumoniae* VK148. Forty-one putative cyclotide masses were identified in *Viola inconspicua*. Six novel sequences (cyI1-cyI6) highly abundant in bioactive library fractions were determined, expanding the already substantial sequence diversity of this highly stable and potent peptide class. Notably, each motif type was readily identifiable via characteristic MS/MS fingerprint ions that may be used in future analysis

of cyclotide contents of botanical extracts. MIC values determined for library fractions containing novel cyclotides cyI3-cyI6 were in the low micromolar range, whereas MIC values of isolated cyclotides cyI3 and cyI4 were considerably higher. Due to the high sequence homology of cyI3-cyI6, the discrepancy in MIC values between isolated and complex cyclotide samples could indicate synergistic interactions among different sequences; however, it is possible that minor sequence variations in cyI5 and cyI6 could account for this increase in potency. Molecular modeling of a cyclotide with the novel loop six "Type I" motif provided evidence for standard cyclotide bracelet sub-family secondary structure elements. This work adds to the sequence diversity of the highly bioactive and stable cyclotide class.

4.3 Experimental Section

4.3.1 Plant Material

Viola inconspicua seeds (Mountain Gardens, Burnsville, NC) were grown in a laboratory greenhouse at 17.5-20.3 °C on a 14/10 h light/dark cycle. Mature aerial tissue was harvested with immediate flash freezing and storage at -80 °C until extraction. *V. inconspicua* greenhouse specimens have been deposited at the UNC Herbarium under accession numbers NCU00303107 and NCU00303108. Specimens are publicly available for viewing on the SERNEC (Southeast Regional Network of Expertise and Collections) website (http://sernecportal.org/portal/index.php).

4.3.2 Peptide Library Preparation

V. inconspicua strong cation exchange (SCX) peptide library was prepared as described previously.[41] Briefly, peptidyl constituents were extracted from 100 g fresh weight material in 300 mL 10% acetic acid buffer and selected for by passing extracts through a high molecular weight cutoff filter (MWCO 30 kDa; Millipore) to remove large proteins and by dialyzing into 5

mM ammonium formate pH 2.7 to remove small-molecules (0.1−1 kDa cutoff; SpectrumLabs). A *V. inconspicua* SCX peptide library was generated with an SCX method (PolySulfethyl A column, 100 mm × 4.6 mm, 3 μm particles, PolyLC) using a linear salt gradient (mobile phase A: 5 mM ammonium formate, 20% acetonitrile, pH 2.7; mobile phase B: 500 mM ammonium formate, 20% acetonitrile, pH 3.0). SCX library fractions were desalted via vacuum centrifugation. A reversed-phase peptide library was subsequently prepared by injecting combined SCX library fractions 15-40 (to avoid collecting early eluting small molecules) onto a Phenomenex Jupiter C_{18} column (300 Å, 5 μm, 4.6 mm x 150 mm) with a flow rate of 0.5 mL/min and a 25 minute linear ramp (mobile phase A, 95/5 water/acetonitrile with 0.1% TFA; mobile phase B, acetonitrile with 0.1% TFA). Reversed-phase library fractions were dried down via vacuum centrifugation and resuspended in water.

4.3.3 Mass Shift Analysis

V. inconspicua SCX peptide library fractions 15-40 were combined, reduced with 10 mM dithiothreitol (45 °C, 850 rpm, 30 min) and alkylated with 100 mM iodoacetamide (25 ˚C, 850 rpm, 15 min). Salts were removed by solid phase extraction with Pierce C_{18} spin columns (Thermo Scientific) prior to LC-MS/MS analysis.

4.3.4 Antibacterial Bioactivity Assays

The *V. inconspicua* reversed-phase peptide library was assayed against *E. coli* ATCC 25922 and clinical isolate *Klebsiella pneumoniae* VK148 (rifampin and streptomycin resistant derivative of ATCC 43816) in a 96-well plate format, as described previously[41]. Bioassay plate wells each contained 30 μL Mueller-Hinton medium, 10 μL bacterial culture, and 10 μL *V. inconspicua* library fraction. Positive and negative controls consisted of antibiotic (*E. coli:* ampicillin, 100 μg/mL; *K. pneumoniae* and *S. aureus:* erythromycin 100 μg/mL) and sterile

water, respectively. Plates were incubated at 37 °C, 250 rpm for 4 h. Resazurin was added to each well to a final concentration of 1 mM and incubated for approximately 1 h. Cell viability was quantified using an excitation wavelength of 544 nm and emission wavelength of 590 nm. Assays were performed in triplicate.

4.3.5 Proteolysis for Sequence Characterization

To avoid overly complex LC-MS/MS spectra, reduction and alkylation was performed as described above on individual SCX fractions. Reduced and alkylated material was incubated 1:50 enzyme:substrate with endoproteinase Glu-C enzyme (Sigma) or Chymotrypsin (Sigma) at 37 °C for 3 h.

4.3.6 LC-MS/MS Analysis

Approximately 1 μg of acidified reduced/alkylated, reduced/alkylated/glu-C digested, or reduced/alkylated/chymotrypsin digested *V. inconspicua* sample was injected onto a nano-LC-ESI-MS/MS platform consisting of a NanoAcquity (Waters, Milford, MA) coupled to a TripleTOF5600 MS (AB Sciex, Framingham, MA). Front-end UPLC separation of peptides was achieved on an HSS T3C_{18} column (100 Å, 1.8 μm, 75 μm × 250 mm, Waters), after passing a Symmetry C_{18} trap column (100 Å, 5 μm, 180 μm × 20 mm, Waters), with a flow rate of 0.3 μL/min and a 30 minute linear ramp of 5%−50% B (mobile phase A, 1% formic acid in water; mobile phase B, 1% formic acid in acetonitrile). The TripleTOF5600 MS was operated in positive-ion, high-sensitivity mode with the MS survey spectrum using a mass range of 350−1600 Da in 250 ms. Targeted CID MS/MS data was acquired for the six novel cyclotides sequences using reduced/alkylated and digested samples and a collision energy (CE) of 40. Deisotoped peak lists for the intact and reduced/alkylated samples were generated using Progenesis QI for Proteomics software (Nonlinear Dynamics, v.2.0) with a retention time filter

of 14−45 min and a maximum charge of +10. “Peptide ion data” was exported from Progenesis and analyzed for the characteristic reduction/alkylation mass shift corresponding to three disulfide bonds (348.16 Da) using Python.

4.3.7 Isolation of CyI3 and CyI4

CyI3 and cyI4 isolation was achieved by collecting cyclotide-containing SCX fractions, desalting with SPE, and injecting onto an Agilent Zorbax extended C_{18} column (300 Å, 5 µm, 4.6 mm x 250 mm) with a flow rate of 2 mL/min and a linear ramp of 5-25% mobile phase B from 0-5 min and 25-50% mobile phase B from 5-20 min (mobile phase A, water with 0.1% formic acid; mobile phase B, acetonitrile with 0.1% formic acid). CyI3/cyI4 reversed-phase peaks from multiple runs were manually collected, dried down in a vacuum concentrator, resuspended in water, and combined for subsequent MIC testing. CyI3/cyI4 purity was analyzed by LC-MS analysis on a ThermoScientific Q Exactive HF-X via direct infusion.

4.3.8 Determination of Minimum Inhibitory Concentration

The cyI3-cyI6 concentration in *V. inconspicua* reversed-phase library fractions and isolated cyI3/cyI4 samples was determined using a NanoDrop (Thermo) and an extinction coefficient of 8855 $M^{-1}cm^{-1}$. MIC values were assessed with a two-fold serial dilution performed in a 96-well plate format. In each well, 30 µL minimal medium, 10 µL bacterial culture, and 10 µL library fraction was added, and the plate was incubated at 37 °C for 4 h, shaking at 250 rpm, and both OD_{600} and resazurin were used to establish MIC values. The assay was performed in triplicate for all concentrations (except the highest concentration of cyI3, in which availability of material limited replicates to two).

4.3.9 Molecular Modeling of Cyclotide CyI3

The three-dimensional structure of cyI3 was modeled with the Robetta server[32], an automated structure prediction server based on the Rosetta software package (Rosetta Commons). Circulin B (PDB 2ERI) was chosen as the homology model template. The five resulting structure predictions were aligned in Pymol (Schrödinger) with the prototypical bracelet cyclotide, cyO2.

4.4 ACKNOWLEDGEMENTS

The authors would like to thank Carol Ann McCormick (NCU herbarium, Chapel Hill, NC) and Harvey Ballard (Ohio University) for curation and identification of *Viola* specimens. NSF MRI (CHE-1726291) supported the purchase of the Q-Exactive HF-X mass spectrometer, and we thank Dr. Brandie Ehrmann for training on this instrument. This work was funded by NIH (1R01GM125814-01) to L.M.H. and N.C.P. acknowledges NIH Biophysics training grant support (T32 GM008570).

4.5 Tables

Table 4.1. Six Novel Cyclotide Sequences (cyI1-cyI6) Identified in *Viola inconspicua*.

Identity	Mass (Da)	Sequence
cyI1	3236.4	SCVYIPCISSVVGCSC-KSKVCYKNGT-FPCGE
cyI2	3259.4	SCVWIPCISSVVGCSC-KSKVCYKNGT-FPCGE
cyI3	3378.5	TCIWGKCYSASIGCSCSKYKVCTLNGNPGACGE
cyI4	3391.4	TCVWGKCYSASIGCSCNKYKVCTLNGNPGACGE
cyI5	3406.5	TCIWGKCYSASIGCSCSRYKVCTLNGNPGACGE
cyI6	3433.5	TCIWGKCYSAKIGCSCSKYKICTLNGNPGACGE

4.6 Figures

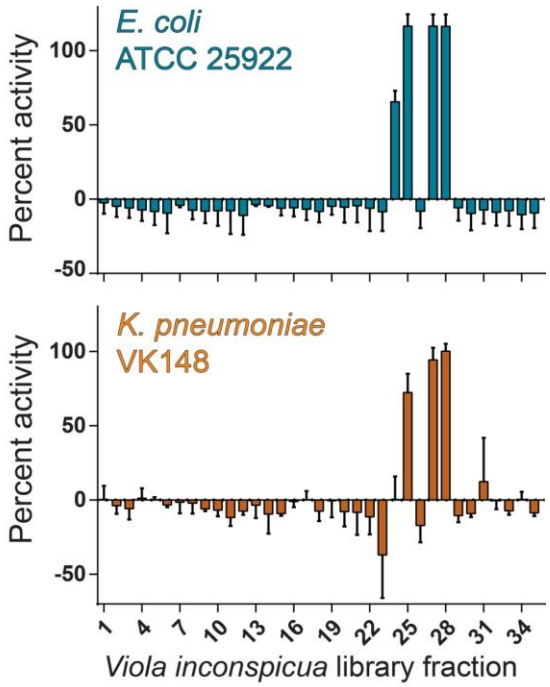


Figure 4.1. *V. inconspicua* peptide library exhibits robust bioactivity against *E. coli* ATCC 25922 (top) and *K. pneumoniae* VK148 (bottom).

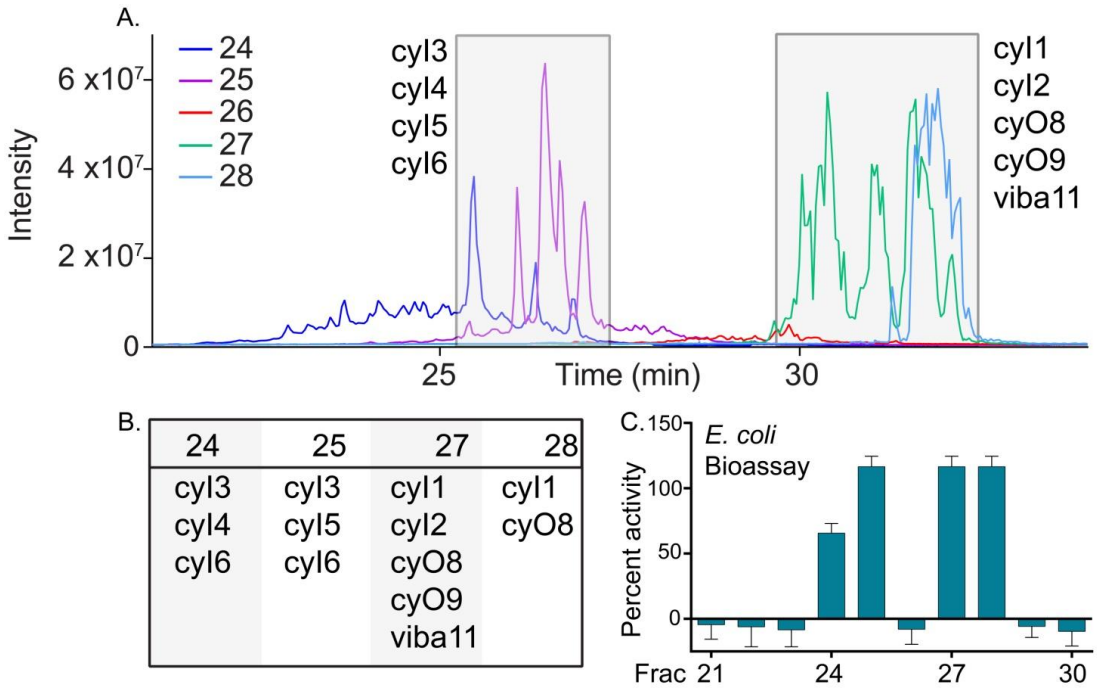


Figure 4.2. Bioactive *V. inconspicua* peptide library fractions contain abundant novel cyclotides (cyI1-cyI6). (A) Overlaid total ion chromatograms (TICs) of *V. inconspicua* reversed-phase library fractions 24-28. (B) Abundant cyclotide species present in each of the bioactive fractions. (C) *V. inconspicua* vs. *E. coli* ATCC 25922 assay bioactivity region matches elution profile of novel cyclotide species cyI1-cyI6.

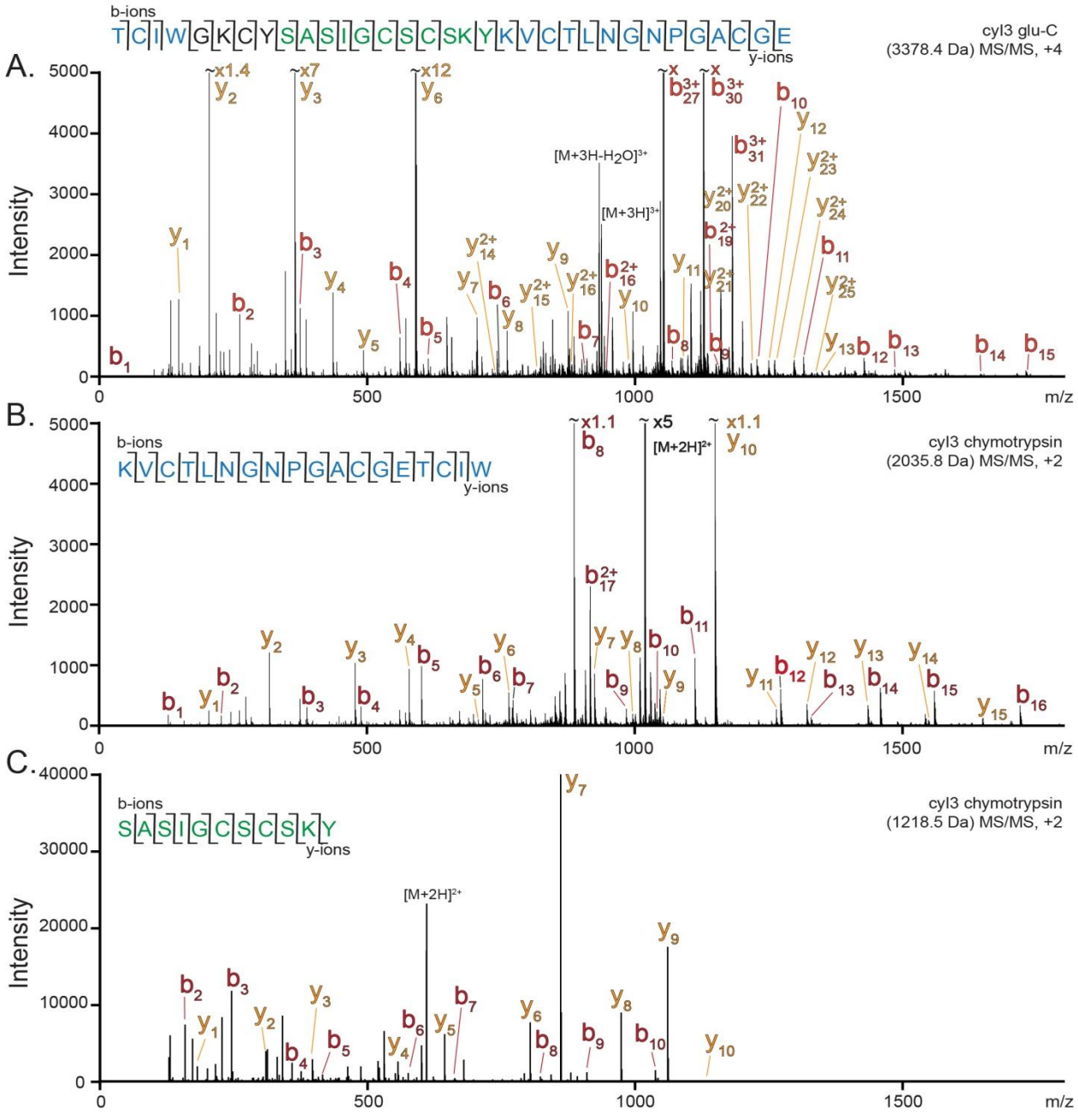


Figure 4.3. *V. inconspicua* cyclotide cyI3 (3378.4 Da) CID MS/MS spectra of (A) gluC linearized cyI3 (+4), (B) cyI3 chymotrypsin fragment 2035.8 Da (+2), and (C) cyI3 chymotrypsin fragment 1218.5 Da (+2). Spectra are annotated with *b* and *y* ion series, and coverage is shown on sequence maps above each spectrum.

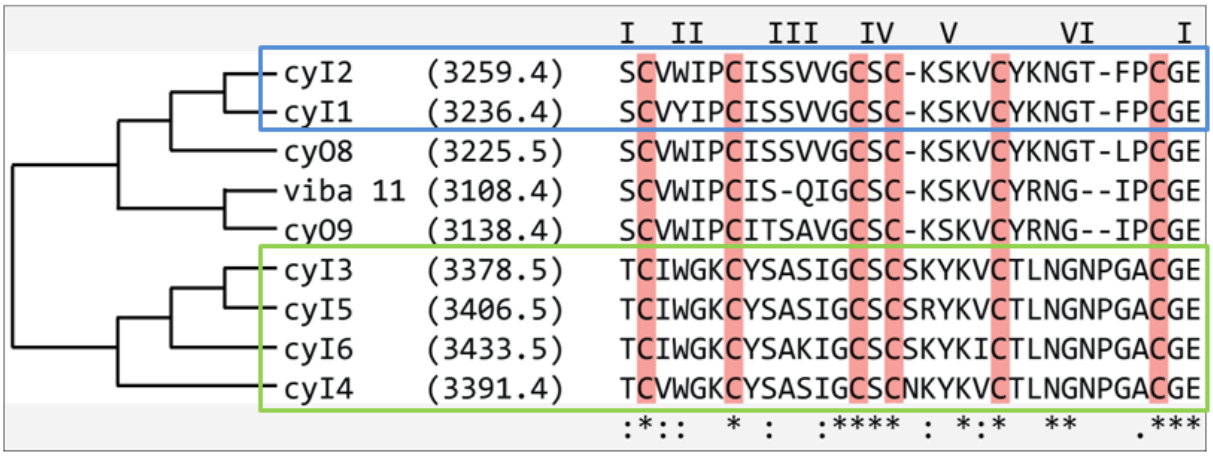


Figure 4.4. Alignment of novel cyclotide sequences cyI1-cyI6 with previously characterized cyclotides viba 11, cyO8, and cyO9. Loops are indicated by roman numerals above sequences. Homology is indicated by tree on left. Asterisks on bottom of sequences indicate high conservation, where colons indicate conservative mutations.

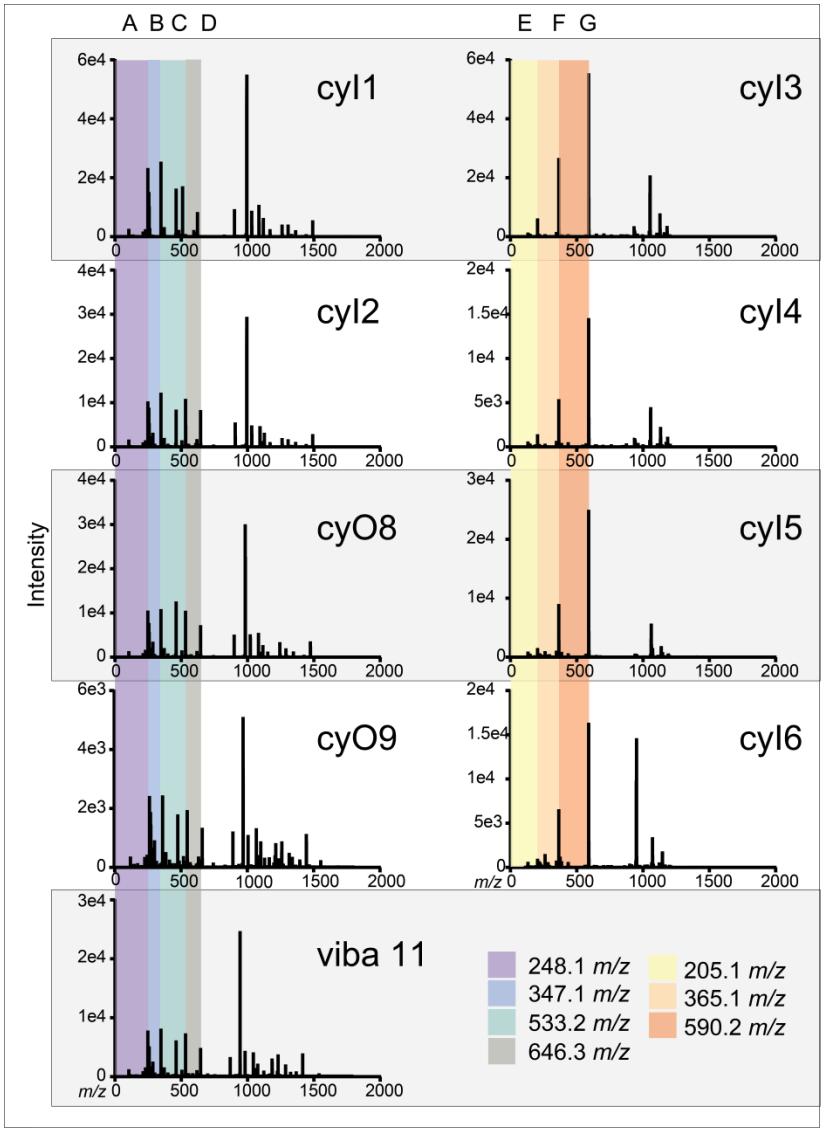


Figure 4.5. CID MS/MS spectra of GluC-digested *V. inconspicua* cyclotides contain several highly abundant fragment ions characteristic of motif type, either Type I (left column) or the novel Type II motif reported herein (right column). Fragments are color-coded throughout spectra.

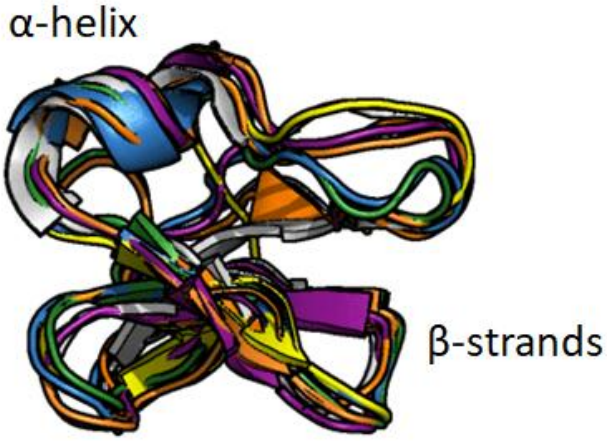


Figure 4.6. Overlay of Robetta predicted novel cyclotide cyI3 structures with prototypical bracelet cyclotide cyO2 (white stucture, PDB 2KNM). Generated in PyMol (Shrödinger). CyI3 predicted structures conform to known bracelet subfamily secondary and tertiary structure.

REFERENCES

1. Muluye, R.A., Y. Bian, and P.N. Alemu, *Anti-inflammatory and Antimicrobial Effects of Heat-Clearing Chinese Herbs: A Current Review*. J Tradit Complement Med, 2014. **4**(2): p. 93-8.

2. Feyzabadi, Z., et al., *A Critical Review on Phytochemistry, Pharmacology of Viola odorata L. and Related Multipotential Products in Traditional Persian Medicine.* Phytother Res, 2017. **31**(11): p. 1669-1675.

3. Chandra, D., et al., *Phytochemical and Ethnomedicinal Uses of Family Violaceae.* Current Research in Chemistry, 2015. **7**(2): p. 44-52.

4. Gran, L., *Oxytocic principles of Oldenlandia affinis.* Lloydia, 1973. **36**(2): p. 174-8.

5. Witkowska-Banaszczak, E., et al., *Antimicrobial activity of Viola tricolor herb.* Fitoterapia, 2005. **76**(5): p. 458-61.

6. Gautam, S.S., Navneet, and S. Kumar, *The antibacterial and phytochemical aspects of Viola odorata Linn. extracts against respiratory tract pathogens.* Proceedings of the National Academy of Sciences, India Section B, 2012. **82**(4): p. 567-572.

7. Jeong, Y.H., et al., *Anti-inflammatory effects of Viola yedoensis and the application of cell extraction methods for investigating bioactive constituents in macrophages.* BMC Complement Altern Med, 2016. **16**: p. 180.

8. Siddiqi, H.S., et al., *Studies on the antihypertensive and antidyslipidemic activities of Viola odorata leaves extract.* Lipids Health Dis, 2012. **11**: p. 6.

9. Habibi, E., et al., *Modulatory effects of Viola odorata flower and leaf extracts upon oxidative stress related damage in an experimental model of ethanol-induced hepatotoxicity.* Appl Physiol Nutr Metab, 2018.

10. Deng, W., et al., *Anticancer activity of Oldenlandia diffusa and Viola Philippica Car.* Journal of Cancer Research Updates, 2013. **2**: p. 87-94.

11. Alipanah, H., M.R. Bigdeli, and M.A. Esmaeili, *Inhibitory Effect of Viola odorata Extract on Tumor Growth and Metastasis in 4T1 Breast Cancer Model.* Iran J Pharm Res, 2018. **17**(1): p. 276-291.

12. Saether, O., et al., *Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1.* Biochemistry, 1995. **34**(13): p. 4147-58.

13. Zhang, J., et al., *Transcriptomic screening for cyclotides and other cysteine-rich proteins in the metallophyte Viola baoshanensis.* J Plant Physiol, 2015. **178**: p. 17-26.

14. Weidmann, J. and D.J. Craik, *Discovery, structure, function, and applications of cyclotides: circular proteins from plants.* J Exp Bot, 2016. **67**(16): p. 4801-12.

15. Hellinger, R., et al., *Peptidomics of Circular Cysteine-Rich Plant Peptides: Analysis of the Diversity of Cyclotides from Viola tricolor by Transcriptome and Proteome Mining.* J Proteome Res, 2015. **14**(11): p. 4851-62.

16. Trabi, M., et al., *Variations in cyclotide expression in viola species.* J Nat Prod, 2004. **67**(5): p. 806-10.

17. Narayani, M., A. Chadha, and S. Srivastava, *Cyclotides from the Indian Medicinal Plant Viola odorata (Banafsha): Identification and Characterization.* J Nat Prod, 2017. **80**(7): p. 1972-1980.

18. Niyomploy, P., et al., *Discovery and Characterization of Cyclotides from Rinorea Species.* J Nat Prod, 2018.

19. Gran, L., *On the effect of a polypeptide isolated from "Kalata-Kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus.* Acta Pharmacol Toxicol (Copenh), 1973. **33**(5): p. 400-8.

20. Slazak, B., et al., *How Does the Sweet Violet (Viola odorata L.) Fight Pathogens and Pests - Cyclotides as a Comprehensive Plant Host Defense System.* Front Plant Sci, 2018. **9**: p. 1296.

21. Slazak, B., et al., *Immunolocalization of cyclotides in plant cells, tissues and organ supports their role in host defense.* Planta, 2016. **244**(5): p. 1029-1040.

22. Pranting, M., et al., *The cyclotide cycloviolacin O2 from Viola odorata has potent bactericidal activity against Gram-negative bacteria.* J Antimicrob Chemother, 2010. **65**(9): p. 1964-71.

23. Wang, C.K., et al., *Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis.* J Nat Prod, 2008. **71**(1): p. 47-52.

24. Ireland, D.C., et al., *Cyclotides as natural anti-HIV agents.* Biopolymers, 2008. **90**(1): p. 51-60.

25. Gerlach, S.L., et al., *Anticancer and chemosensitizing abilities of cycloviolacin 02 from Viola odorata and psyle cyclotides from Psychotria leptothyrsa.* Biopolymers, 2010. **94**(5): p. 617-25.

26. Lindholm, P., et al., *Cyclotides: a novel type of cytotoxic agents.* Mol Cancer Ther, 2002. **1**(6): p. 365-9.

27. Burman, R., et al., *Evaluation of toxicity and antitumor activity of cycloviolacin O2 in mice.* Biopolymers, 2010. **94**(5): p. 626-34.

28. Sun, Z., et al., *Identification of medicinal species and antifungal property of a Dong ethnic drug.* Int J Clin Exp Med, 2014. **7**(12): p. 5004-9.

29. Ravipati, A.S., et al., *Understanding the Diversity and Distribution of Cyclotides from Plants of Varied Genetic Origin.* J Nat Prod, 2017. **80**(5): p. 1522-1530.

30. Ravipati, A.S., *Discovery and mode of action of cyclotides*, in *Institute for Molecular Bioscience*. 2016, University of Queensland: Australia.

31. Pechous, R.D., et al., *In vivo transcriptional profiling of Yersinia pestis reveals a novel bacterial mediator of pulmonary inflammation.* MBio, 2015. **6**(1): p. e02302-14.

32. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W526-31.

33. Burman, R., et al., *Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae.* Front Plant Sci, 2015. **6**: p. 855.

34. Gruber, C.W., et al., *Distribution and evolution of circular miniproteins in flowering plants.* Plant Cell, 2008. **20**(9): p. 2471-83.

35. Ireland, D.C., M.L. Colgrave, and D.J. Craik, *A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability.* Biochem J, 2006. **400**(1): p. 1-12.

36. Herrmann, A., et al., *The alpine violet, Viola biflora, is a rich source of cyclotides with potent cytotoxicity.* Phytochemistry, 2008. **69**(4): p. 939-52.

37. Craik, D.J. and U. Malik, *Cyclotide biosynthesis.* Curr Opin Chem Biol, 2013. **17**(4): p. 546-54.

38. Nguyen, G.K., et al., *Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis.* Nat Chem Biol, 2014. **10**(9): p. 732-8.

39. Saska, I., et al., *An asparaginyl endopeptidase mediates in vivo protein backbone cyclization.* J Biol Chem, 2007. **282**(40): p. 29721-8.

40. Troeira Henriques, S. and D.J. Craik, *Cyclotide Structure and Function: The Role of Membrane Binding and Permeation.* Biochemistry, 2017. **56**(5): p. 669-682.

41. Kirkpatrick, C.L., et al., *The "PepSAVI-MS" Pipeline for Natural Product Bioactive Peptide Discovery.* Anal Chem, 2017. **89**(2): p. 1194-1201.

CHAPTER 5: Exploring the diversity of cysteine-rich natural product peptides via MS/MS fingerprint ions

*****Manuscript submitted for publication. Authors: Nicole C. Parsley, Owen L. Williams, and Leslie M. Hicks.

5.1 Introduction

A global rise in antimicrobial resistance necessitates the identification and characterization of novel antimicrobial compounds to supplement and increasingly dry drug pipeline. Natural product peptides are a source of antimicrobial therapeutics with unique chemistries; however, highly complex natural product matrices, inherent sequence variability, and low abundance challenge the identification of new antimicrobial peptide (AMP) sequences. Antimicrobial peptide secondary/tertiary structure, charge, post-translational modification, and molecular weight are non-uniform, ranging from 2-4 residue, negatively-charged non-ribosomally synthesized, linear bacterial lipopeptides to >40 residue, basic, and highly disulfide-bound plant defensins. Additionally, the expression and abundance of AMPs are dynamic, influenced by environmental factors (e.g. season, stress) and circadian rhythms. Thus, despite the >3000 AMPs reported among databases (dbaas.org, AMPdb.org), the chemical suites of many sources remain obscured by extract complexity, diversity, and periodicity, necessitating new approaches to explore the true AMP sequence diversity in nature.

Mass spectrometry (MS) drives primary discovery, where modern instruments coupled with ultra-high performance liquid chromatography (UHPLC) can identify and resolve low abundance peptidyl species in complex mixtures with unprecedented limits of detection, resolving power, and mass accuracy. Advances in multistage MS (MS^n) allow for the determination of extract constitution at the MS^1 level as well as the identity of constituents via tandem MS.

Fragmentation methods (e.g. CID, ETD, EThcD) and orthogonal chemical derivatization strategies enable the characterization of diverse and difficult-to-sequence mass species without prior isolation. Even so, the hundreds to thousands of compounds present in extracts complicate comprehensive peptide identification/sequence elucidation, and additional strategies are needed to explore the depths of natural product peptide sequence diversity. Conserved features among AMP families can inform MS-based methods towards the exploration of novel sequences in natural product extracts.

Cysteine-rich peptides (CRPs) represent the majority of characterized botanical natural product peptides[1] and are produced extensively in plants to mediate various elements of physiology, reproduction, growth/development, and defense against pathogens.[2-6] Disulfide bonding among cysteine residues scaffolds these highly stable and enzymatically-resistant molecules; CRPs are typically classified by cysteine number and pattern along the backbone.[1] Chemical derivatization strategies can be exploited to detect the presence of CRPs in complex matrices by mass shift analysis, where extracts before and after the reduction and alkylation of Cys residues produces an easily observable characteristic mass shift. Mass shift analysis is often the first step towards the identification of CRPs in botanical extracts,[7,8] however this alone can lead to under- or over-estimation of the number of unique CRPs. Additionally low quality MS/MS spectra typical of untargeted, data dependent experiments may preclude full sequence characterization.

Mass spectral fingerprinting is a method of profiling MS ions as chemical signatures that are characteristic of classes of molecules, organisms, or other target group, and can be derived from *in silico* database-based prediction or from experimental data. Currently, MS fingerprinting is primarily performed at the MS^1 level, without or complementary to MS/MS data. Database-

based protein identification with peptide mass fingerprinting often employs *in-vitro* enzyme digestion prior to MS, where a unique combination of peptides is used as a chemical "fingerprint" of a given protein when compared to those predicted from the database. With this approach, increasing sample complexity generates increasing peptide interference and may challenge accurate protein identification/reduce confidence in protein assignments.[9] Spectral patterns acquired via MALDI-TOF and processed with advanced statistical algorithms have proven invaluable as a tool to discriminate bacterial[10-17] and fungal[18-20] species, including drug resistant strains, though the specific molecular origin of diagnostic ions is often unknown in such chemometric approaches. MS fingerprinting has been exploited extensively in the field of natural products to discriminate chemical composition as it relates to the geographical origin[21,22] and purity/adulteration[23-27] of a variety of natural substances. Additionally, it has been applied to the identification of CRPs at the MS^1 level to reveal misidentification/mislabeling of herbal medicines.[28] MS/MS data is routinely collected during LC-MS experiments and can provide information about the type/characteristics of components in a complex matrix without complete sequence elucidation. Common and abundant CRP MS/MS fingerprint ions can be leveraged to prioritize the identification of novel peptidyl species and reduce the re-characterization of known CRPs.

Knottins are low molecular weight (~3 kDa), functionally diverse CRPs characterized by three conserved disulfide bonds and a canonical 'knotted' cysteine connectivity of C1-C4, C2-C5, and C3-C6.[29] Peptides conforming to the knottin structural class are further subdivided into families based on sequence homology: cyclotides are head-to-tail cyclized knottins found largely in the Violaceae, Rubiaceae, and Cucurbitaceae botanical families. Cyclotides possess diverse, innate bioactivities with >400 sequences known, and thousands more are thought to await

discovery. Three cyclotide subfamilies exist (bracelets, Möbius, and trypsin inhibitors) where regions between the six Cys residues (loops 1-6) facilitate sequence comparison among cyclotide species (Figure 1). Largely homologous, the bracelet and Möbius cyclotides are differentiated by the absence or presence of a cis-Trp/Pro or cis-Tyr/Pro bond in loop 5, respectively, in addition to some regions of sequence homology specific to each subfamily. Despite sharing a cyclic knot structure, trypsin inhibitor sequences are more dissimilar to the aforementioned subfamilies and offer a comparatively limited range of bioactivities.[30]

The number and identity of CRP species, such as cyclotides, can dramatically vary among natural product extracts, where the comprehensive characterization of the suite of CRP masses in a single botanical species can challenge even the most sensitive and accurate analytical methods. Cysteine derivatization and mass shift analysis are commonly used to identify CRP constituents, and cyclic CRPs require *in vitro* enzymatic digestion (e.g. endoproteinase Glu-C) to cleave the cyclic backbone and facilitate MS^2 sequence characterization. Regardless, complete sequence characterization is often challenged by low abundance cyclotide species in complex matrices, numerous sample preparation steps, and incomplete MS/MS fragmentation.

Regions of sequence homology among CRPs can be exploited to guide the identification of novel mass species, where shared sequence motifs are reflected in MS/MS spectra as "fingerprint" ions. Herein, theoretical MS/MS fingerprint ions most representative of the bracelet and Möbius cyclotide subfamilies are identified. Proof-of-principle is achieved by coupling a mass shift analysis, where CRPs are detected by a +348.16 Da mass shift consistent with the reduction and alkylation of three disulfide bonds, with a "fingerprint" analysis of the reduced, alkylated, and enzymatically-digested botanical extracts from the cyclotide-producing Viola genus. Common *m/z* fingerprints observed with these data are discussed and sequence

information derived from fingerprint analysis is cross referenced with CyBase to identify sequences of interest. False discovery rate is determined via *in silico* fragmentation of the Swissprot database and enumeration of the instances in which these diagnostic fingerprint ions would be produced by a non-cyclotide mass. Aerial and root tissue harvested from the same greenhouse grown Viola plants and wild-grown Viola aerial material harvested in two different seasons was used to demonstrate the utility of MS/MS fingerprint ions and assess changing peptide populations, focusing on the bracelet and Möbius cyclotides abundant in Viola spp. Without full sequence characterization or targeted runs, and only low quality MS/MS spectra, we demonstrate that a rapid assessment of MS/MS fingerprint ions can reveal novel cyclotide masses, modified cyclotides, and masses that are likely known cyclotide sequences. We extend our analysis to CRPs in the seeds of the well characterized Cucurbitaceae species, *Lagenaria siceraria* (bottle gourd). Highly abundant MS/MS fingerprint ions are identified of the known linear trypsin inhibitors LSTI-II and III.[31] Mass shift analysis is then coupled with the identification of fingerprint ions in unknown mass species to identify and subsequently sequence three new *L. siceraria* CRPs, LSCRP-I, II, and III.

Thus, we herein demonstrate that low *m/z* fingerprint ions can indicate whether a mass is (a) likely a cyclotide, (b) belongs to a cyclotide subfamily, (c) a potentially new cyclotide motif type that warrants full sequence characterization, (d) an oxidized or over-alkylated cyclotide, aiding in dereplication, (e) can inform what types of cyclotides are changing in abundance among sample types (e.g. tissue samples, seasonal variation), and (f) can differentiate multiple unique peptide species close or identical in mass that may have been overlooked otherwise based on mass alone.

5.2 Experimental

5.2.1 Fragment Database Generation

Fragmentation databases were generated by reading .fasta sequences of cyclotide (CyBase.au.org, 433 sequences, accessed January 5, 2020) and Swissprot (558,712 sequences, accessed January 5, 2020) into an in-house developed C++ program (accessible at github.com/hickslab, Repositories: FragmentSearch) that simulates Glu-C digestion, *b/y* ion fragmentation, and compares the resulting set of *in silico* predicted fragments with a user-specified experimentally-derived list of fingerprint fragment masses within a mass tolerance, defaulting to 0.02 Daltons. For false discovery rate (FDR) determination, all cyclotide species (208 sequences) were removed from the Swissprot database prior to *in silico* manipulation. Monoisotopic fragment masses were calculated, where Cys has a fixed modification of +57.0214 Da. Digestion with Glu-C was simulated by splitting each sequence at the C-terminus of each glutamic acid residue and the resulting theoretical *b/y* ions of the linear sequence predicted.

5.2.2 Plant Material and Extraction

Viola communis aerial material was harvested from North Carolina residential areas (35°55'25.6"N, 78°56'14.5"W) in the months of March and December 2019. *Viola inconspicua* was grown in a greenhouse (14/10 h light/dark cycle, 17.5-20.3 °C) and aerial and root material was harvested from the same plant. *Lagenaria siceraria* seeds were purchased from Strictly Medicinal (Eugene, OR). Viola aerial and root material was ground under liquid nitrogen to a fine powder, and *L. siceraria* seed material was ground with a seed mill. Peptides were extracted with 10% acetic acid buffer (1:3 w/v aerial and root tissues, 1:10 w/v seed material). Peptidyl constituents were selected for by filtering extracts through a 30 kDa MWCO filter (Millipore) and dialyzing with a 0.1−1 kDa cutoff membrane (Spectrum Labs) into 5 mM ammonium

formate pH 2.7. Remaining small molecule components were removed by strong cation exchange (PolySulfoethyl A column, 100 mm × 4.6 mm, 3 μm particles, PolyLC) using a linear salt gradient (mobile phase A: 5 mM ammonium formate, 20% acetonitrile, pH 2.7; mobile phase B: 500 mM ammonium formate, 20% acetonitrile, pH 3.0). Peptide-containing fractions were pooled, salts were removed via solid-phase extraction (SPE) cartridges (Waters Sep-Pak, 500 mg), and elutions were dried in a vacuum centrifuge (Labconco).

5.2.3 Mass Shift Analysis and Proteolysis

Material was reduced (10 mM dithiothreitol, 45 °C, 850 rpm, 30 min) and alkylated (100 mM iodoacetamide, 25 ˚C, 850 rpm, 15 min). Reduced and alkylated material was incubated 1:200 enzyme:substrate with endoproteinase Glu-C enzyme (Sigma) in 100 mM ammonium bicarbonate, pH 7.8, at 37 °C for 3 h. Pierce SPE C_{18} spin columns were used to remove salts (Thermo Scientific) prior to LC-MS analysis.

5.2.4 LC-MS/MS Analysis.

One μg of acidified reduced/alkylated, reduced/alkylated/Glu-C digested sample was injected onto a nano-LC-ESI-MS/MS platform comprised of a NanoAcquity (Waters, Milford, MA) coupled to a TripleTOF5600 MS (AB Sciex, Framingham, MA). A Symmetry C_{18} trap column (100 Å, 5 μm, 180 μm × 20 mm, Waters) and HSS T3C_{18} analytical column (100 Å, 1.8 μm, 75 μm × 250 mm, Waters) were used for nUPLC separation of peptides, with a flow rate of 0.3 μL/min and a 30 minute linear ramp of 5%−50% B (mobile phase A, 1% formic acid in water; mobile phase B, 1% formic acid in acetonitrile). The TripleTOF5600 MS was operated in positive-ion, high-sensitivity mode with the MS survey spectrum using a mass range of 350−1600 Da in 250 ms and IDA where the first 20 MS^2 spectra were collected with a mass range of 100-1800 Da each in 87 ms. Targeted CID MS/MS data was acquired for the three LSTI

masses, 2605.93 Da (869.65 *m/z*, +3), 2705.02 Da (902.68 *m/z*, +3), 2762.01 Da (921.68 *m/z*, +3), using the reduced/alkylated samples and a collision energy (CE) of 40 and a CE spread of 5. Mass shift analysis was performed by identifying 348.16 Da mass shifts in Progenesis QI for Proteomics (Nonlinear Dynamics, v.2.0) using "Peptide ion data" deisotoped peak lists with a retention time filter of 14−45 min and a maximum charge of +10.

5.3 Results

5.3.1 Cyclotide Fingerprint Ions

Cyclotide sequences (CyBase.au.org) were Glu-C cleaved and fragmented *in silico* to produce theoretical *b/y* MS/MS ions. Fingerprint ions of cyclotides and cyclotide subfamilies were assessed for abundance in the theoretical database, initially focusing on the more common bracelet and Möbius species (Table 1). Möbius cyclotides are defined primarily by an aromatic-Pro motif in loop 5; total Möbius cyclotides, cis-Trp/Pro and cis-Tyr/Pro Möbius sequences were separately examined. Necessitated by design, all cyclotides contain a 147.08 Da fragment ion corresponding to a C-terminal Glu residue. Notably, ~87% of all known cyclotide species contain the C-terminal residues 'CGE,' providing a 148.06, 205.08, and 365.11 *m/z* fingerprint handle for masses likely belonging to the cyclotide family. The Swissprot database [32] was then used to estimate the FDR for these cyclotide fingerprint ions based on the frequencies of identifying sequences/fragment ions characteristic of cyclotides in non-cyclotide masses. The likelihood of finding the cyclotide-indicative fragment ions 148.06, 205.08, 365.11 *m/z* in non-cyclotide masses was calculated to be 97%, 0.46%, and 0.02%, respectively, where the likelihood of identifying the 148.06 and 205.08 *m/z* fragment ions in the same spectrum would be 0.45% and of identifying all three fragment ions would be only 0.005% (Figure 1). This indicates that the 'CGE' motif is unlikely to be found in non-cyclotide mass species; therefore

these fragment ions in combination can be used as a basis to assign novel putative cyclotides from LC-MS/MS data.

More specific molecular information can be attained from other highly abundant MS/MS fingerprint ions. In addition to the presence/absence of the cis-Trp/Pro or cis-Tyr/Pro motif in loop 5, bracelet and Möbius cyclotides offer regions of sequence homology within subfamilies. As such, the *b*-ions 248.07, 347.14, and *y*-ion 462.17 *m/z* are seen primarily in bracelets, where *b*-ions 102.05, 262.09, 409.15 *m/*z and *y*-ions 464.18 and 561.23 *m/*z are seen largely in Möbius cyclotides (Figure 1, Table 1, summarized in Table 2). Differences between the cis-Trp/Pro and cis-Tyr/Pro Möbius sequences are seen primarily in the cis-Trp/Pro *y*-ion series: 464.18, 674.32, 561.23, and 731.34 *m/z* (Table 1). The percent of bracelet and Möbius cyclotides in CyBase that contain a given MS/MS fingerprint ion can be calculated (Table 2). Additionally, the percent in which a given MS/MS fingerprint ion belongs to a bracelet or Möbius sequence can be calculated (Table 2), and indicates the specificity of a given sequence to a subfamily (e.g. if a sequence motif is identified, does it suggest a bracelet or Möbius identity, aiding in sequence characterization). Bracelet sequences outnumber Möbius sequences ~2:1, consistent with calculated Bracelet:Möbius ID values of ~67:33. Though this does not preclude the identification of these fragment ions in the alternative subfamily, these ions can be used to predict cyclotide subfamily type in addition to contributing additional evidence that a mass belongs to the cyclotide family. A conserved aromatic residue in bracelet subfamily loop 2 (N-terminal sequence 'SCVX(Aromatic)') can provide additional sequence information, where instances of Phe, Tyr, and Trp incorporation into this *b4* position are roughly equal (Table 2) and can aid in the discrimination of sequence identities. In addition to these theoretical cyclotide, bracelet, and Möbius MS/MS fingerprint ions, experimentally observed fragment ions (Table 2) can be used to

quickly differentiate cyclotides close in mass, identify common modifications (e.g. oxidation, alkylation), and identify likely new/known sequences in tandem with database searching towards a more accurate representation of sequence diversity in a plant extract, and are discussed in detail through examples below.

5.3.2 Seasonality

Antimicrobial peptide expression can be seasonal, where plants respond to different environmental conditions and biotic/abiotic pressures via proteomic changes. *Viola communis* material harvested from residential areas in the months of March and December 2019 (spring and winter) was reduced, alkylated, and Glu-C digested for mass shift and fingerprint ion analysis. Total ions chromatograms (TICs) of the intact *V. communis* seasonal samples suggest that the extract components are of roughly the same composition, however, differ in abundance (Figure 2A). It should be noted that numerous factors can affect peptide expression and abundance *in planta*, and though material harvested in the spring and winter months is used herein as a illustrative example of seasonal variation, the determination of true seasonal differences would require multiple material harvests to assess the consistency and reproducibility of peptide expression/abundance measurements in each sample.

Approximately 200 *V. communis* masses produced a characteristic +348.16 Da mass shift upon reduction and alkylation and were within the cyclotide mass range of 2500-3700 Da. Sequence elucidation of all 200 mass species would take considerable time, additional targeted experiments to improve MS/MS fragmentation, and would result in the identification of known cyclotides from other botanical species as well as duplicate sequences present in the same plant species with modifications (e.g. oxidation) that obscure sequence identity. Thus, the use of cyclotide-indicative fingerprint ions identified herein can guide sequencing prioritization.

Novel species. Mass 3508.57 Da (878.15 *m/z*, +4, and the neutral, intact mass (NI), calculated by subtracting 348.16 Da for cysteine modification and 18 Da for Glu-C digestion, = 3142.40 Da) produced very few MS/MS fragment ions that could be used towards full sequence elucidation, however, cyclotide *y*-ion fragments 148.06, 205.08, and 365.11 *m/z* corresponding the C-terminal sequence 'CGE' are observed (Figure 2B). Additionally, the common bracelet specific *b*-ions 248.06, 347.13, and *y*-ion 462.16 *m/z* are present, corresponding to the N-terminal and C-terminal sequences 'SCV' and 'PCGE,' respectively (Table 1). Additionally, a 510.20 *m/z* fragment is present and corresponds to a Phe in the aromatic *b4* position, indicating an N-terminal sequence 'SCVF.' The neutral, intact mass 3142.40 Da is within 1 Da of four entries in CyBase: mden F, viphi F, mra5, and kalata B18, identified in the botanical species *Melicytus dentatus, Viola philippica, Melicytus ramiflorus,* and *Oldenlandia affinis,* respectively. Comparing these known sequences with the MS/MS fingerprint ions, mden F and viphi F contain N-terminal 'TCF' and 'SCVF' sequences, respectively, while mra5 and kalata B18 share the N-terminal 'SCVY' motif with 3142.42 Da but both contain a C-terminal 'PCAE' motif. Thus, based on the known sequences, all of these peptides can be ruled out strictly from the four N- and four C-terminal residues/predicted fingerprint ions, indicating that 3142.42 Da may be a novel mass and should be pursued further for sequence characterization.

Dereplication. Natural product discovery is plagued by the re-identification of known compounds, and dereplication strategies are key to the discovery of novel mass species. Like the putatively novel mass above, mass 3518.61 Da (880.66 *m/z*, +4, and NI = 3152.45 Da), did not produce enough fragmentation for full sequence determination, however, the characteristic cyclotide fingerprint ions 205.08 and 365.11 *m/z* were observed, in addition to the bracelet *b*-ions 248.07, 347.13, and the y-ion 462.16 *m/z,* indicating that this mass is a cyclotide in the bracelet

subfamily (Figure 2C). The MS/MS spectrum also contains the *b4* aromatic peak 533.31 *m/z*, indicating that this mass contains an N-terminal sequence 'SCVW' (Table 2). The neutral, intact mass 3152.45 Da searched on CyBase is within 22 ppm of the known cytotoxic bracelet cyclotide, vitri A (3152.38 Da), identified in the botanical species *Viola tricolor*, *Viola biflora*, and *Psychotria leptothyrsa*. The observed MS/MS fingerprint ions are consistent with this cyclotide, with a 'SCVW' N-terminal motif and a 'PCGE' C-terminal motif. Though there could still be internal same mass amino acid substitutions making 3152.45 Da a unique sequence, based on the presence of vitri A among numerous botanical species, mass 3152.45 Da should be de-prioritized for sequence characterization when targeting novel mass species. It is notable that the abundance of this mass appears to increase from the December to March harvests, mirroring an increase in pathogen pressure with changing seasons; the use of internal standards to quantify the differential abundance of the cytotoxic vitri A[33,34] species may inform subsequent studies assessing seasonal peptide expression profiles.

Modifications. Present in both March and December, mass 3550.61 Da (888.66 *m/z*, +4, and NI = 3184.45 Da), has the characteristic cyclotide fingerprint *y*-ions 148.06, 205.08, and 365.11 *m/z* and the bracelet *b*-ions 248.07, 347.13 and *y*-ion 462.16 *m/z*; however, we see a shift of the aromatic *b*-ion peak 533.20 *m/z* to 565.20 *m/z* (Figure 2D), indicating N- and C-terminal sequences 'SCVW(+32)' and 'PGCE,' respectively. This mass shift is commonly observed experimentally in which bracelet type cyclotides are modified with a doubly oxidized loop two Trp residue (+32 Da) [35], and may result from *in vivo* or experimental conditions (e.g. acidic environments). The true intact, neutral mass of this cyclotide is thus calculated by subtracting the 32 Da contributed by di-oxidation, and is consistent with a doubly oxidized vitri A. Without this observation, 3184.45 Da may be misidentified as a new cyclotide species or reassigned

incorrectly: the neutral doubly-oxidized mass is identical to two CyBase cyclotides, Mobo B and Kalata B16. Mobo B can be eliminated from consideration immediately, as the N-terminal sequence motif is 'TCAK' and would not produce the 248.07 and 347.14 *m/*z ions indicative of the common bracelet N-terminal motif 'SCV.' Kalata B16 has the N-terminal 248.07 and 347.13 *m/*z ions, however, these are followed by a Tyr residue, representing the motif 'SCVY,' and is inconsistent with the 565 *m/z* ion, as well as the -32 Da, *b*–ion 533.20 *m/z* peak derived from the vitri A 'SCVW' motif. Additionally, kalata B16 has a C-terminal motif 'PCAE,' and would not contain the *y*–ions 205.08, 365.11, and 462.16 *m/z*. Here an assessment of the MS^2 spectrum quickly identifies a common modification and demonstrates the use of fingerprint ions in dereplication strategies essential to MS-based AMP analysis of natural product extracts.

Expanding chemical diversity. Cyclotides with different sequences can have the same mass (e.g. cycloviolacins O2 and O9), and without MS^2 fingerprint ions the chemical diversity of cyclotide species in a given sample may be underestimated. In the *V. communis* December reduced, alkylated, and Glu-C digested spectra, common cyclotide fingerprint ions in the MS^2 spectra of the masses 3565.65 (892.42 *m/z*, +4, and NI = 3199.49 Da), 3565.61 (892.41 *m/z*, +4, and NI = 3199.45 Da), and 3564.57 Da (892.15 *m/z*, +4, and NI = 3198.41 Da) (Figures 2E, 2F, and 2G, respectively) are present. Upon closer examination, several characteristic fragment ions indicate that these masses are comprised of at least three different peptidyl species. All share the same base characteristic cyclotide *y*-ion 365.11 *m/z* and bracelet specific *b*-ions 248.07 and 347.14 *m/z* and *y*–ion 464.17 *m/z*, however, the *b*-ion loop two aromatic residue is clearly shifted to 494.21, 510.20, and 565.20 *m/z* corresponding to a Phe, Tyr, and doubly-oxidized Trp in 3199.49, 3199.45, and 3198.41 Da, respectively. Standard cyclotide-identifying fingerprint ions couple with the shifting of the tell-tale aromatic peak has allowed quick and facile discrimination of

three unique sequences close in mass, which may have been overlooked as a single mass otherwise. Masses 3199.49 and 3199.45 Da are within 1 Da of three masses on CyBase, and mass 3166.41 Da (3198.41 Da with 32 Da subtracted to account for a di-oxidation) is within 1 Da of four masses on CyBase. Using the information gathered here, these masses can be subsequently analyzed for possible novelty/redundancy with known sequences as described above.

Linear species. Masses exhibiting a 348.16 Da mass shift but failing to appear at the predicted reduced, alkylated, and Glu-C digested *m/z* are either not visible after the multiple sample preparation steps from low abundance, have >1 or 0 Glu residues, the off-target Glu-C cleavage at Asp residues was significant (Glu-C cleaves ~100x slower at Asp residues), or the original mass was linear. The neutral intact mass 3004.26 Da is present in both the March and December harvests, however, despite the initial high abundance in both samples, this mass is not present at the anticipated reduced/alkylated/Glu-C digested *m/z*. The reduced and alkylated *V. communis* samples, however, contain the expected +348.16 Da mass in the MS^1 (3352.41 Da) and unlike cyclic cyclotide variants, exhibited fragmentation in the MS^2. Upon initial observation, no cyclotide fingerprint ions were present and indicates this is either a non-cyclotide mass – and would not have any cyclotide fingerprint ions – or a linear cyclotide variant, and thus would not be cleaved at the loop one Glu residue (*in vivo* cyclotide ligation occurs in loop six, in which would be the natural terminus of a linear variant). The intact neutral mass 3004.26 Da is within 10 ppm of the cyclotide violacin A, a linear cyclotide with a Möbius cis-Trp/Pro motif in loop 5, found in the botanical species *Viola odorata* and *Psychotria leptothyrsa*. *In silico* prediction of violacin A aligns with several prominent fragment ions in the reduced and alkylated MS^2 spectrum (data not shown). Returning to the reduced, alkylated, and Glu-C digested spectrum,

the predicted digested violacin A mass is present in the MS^1 spectrum, where now the linear mass is cleaved into two discrete peptides via an internal Glu residue. The MS^2 from the larger sequence (2648.15 Da) resulting from the Glu-C digest of the linear violacin A cyclotide contains the expected Möbius fragment *b*-ions 262.09 and 409.15 *m/z* (Figure 2H). Interestingly, violacin A has been shown to be less biologically active than other cyclotide species [36] and in contrast to the putative vitri A sequence, the relative abundance appears to significantly decrease from the March harvest to the December harvest.

5.3.3 Tissue-specificity

Novel species. Greenhouse grown *Viola inconspicua* aerial and root material, derived from the same plants and harvested/extracted simultaneously, was assessed for fingerprint ions towards the identification of tissue-specific cyclotide species. Differential cyclotide expression among various plant tissues has been described, and likely relates to pathogen-specific interactions. Differences in the TICs of the *V. inconspicua* aerial and root material suggest that a significant portion of the extract constituents are unique to tissue type (Figure 3A). Again, distinct MS^2 patterns can be used to quickly discern two unique sequences close in mass, 3784.45 (947.12 *m/z*, +4, and NI = 3418.29 Da) and 3784.49 Da (947.13 *m/z*, +4, and NI = 3418.33 Da), and present in both *V. inconspicua* aerial and root MS1 spectra. Both MS^2 spectra support bracelet cyclotide identities, however the aromatic *b4* indicator peak is shifted to 533.21 (Figure 3B) and 494.20 *m/z* (Figure 3C), representing 'SCVW' and 'SCVF' motifs in 3784.45 and 3784.49 Da, respectively. There are no known cyclotides on CyBase.org within 1 Da of the neutral intact mass 3418.29 or 3418.33 Da, suggesting that both of these sequences are novel, unique, and should be prioritized for sequence characterization.

Fingerprint combinations. Though single fingerprint ions can have significant diagnostic importance (e.g. the *b4* bracelet aromatic peak), it is essential to regard combinations of fingerprint ions during analysis. A prominent 262.09 *m/z b*-ion present in both 3705.43 Da (1236.15 *m/z*, +3, NI = 3339.27 Da) and 3745.36 Da (1249.46 *m/z*, +3, NI = 3379.20 Da) spectra (Figure 3D and E, respectively) could represent the N-terminal sequence 'TC,' a motif seen primarily in Möbius cyclotides (Table 2) and taken alone could result in the inaccurate assessment of cyclotide type. The 3705.43 Da (Figure 3D) spectrum contains the subsequent *b*-ion 361.15 *m/z* present in some cis-Trp/Pro Möbius species (Table 1) and an abundant 524.20 *m/z* peak, indicating a potential 'TCVY' motif. Only 'TCVF' and 'TCVW' motifs are seen in CyBase cyclotides, exclusively in bracelet species. Though 'TCV' occurs more often in Mobius, ~9% of bracelets contain a 'TCV' N-terminal motif (data not shown), and all instances where 'TCV' is followed by an aromatic residue, these sequences belong to bracelet, compared with a common 'TCVG' motif in Möbius species – and indicates that this sequence could be a novel bracelet cyclotide. In contrast, the 3745.36 Da MS^2 spectrum also contains an abundant 262.09 *m/z* ion, but in a different combination of prominent peaks, including the cyclotide *b*-ions 205.08 and 365.11 *m/z* and a high abundance 590.20 *m/z* ion. Common *in vitro* alkylation with iodoacetamide (+57.02 Da) can result in the over-alkylation of cyclotide residues. The solvent-exposed bracelet tryptophan is particularly susceptible to over-alkylation, shifting this diagnostic fragment ion to 590.22 *m/z* (Figure 3F). The true reduced and alkylated mass is then calculated as 57.02 Da less, and there should not be a correlated mass in the intact sample MS1 spectra (as this modification occurs upon reduction and alkylation). However, an over-alkylated Trp residue in the *b4* position in the 3745.36 Da spectrum is incompatible with the presence of the 262.09 *m/z* peak, where these ions indicate the unique N-terminal sequences 'TC' and 'SCVW(Alk).'

Recently, five novel bracelet cyclotides were identified in *V. inconspicua* containing an unusual bracelet 'TCV/I' N-terminal motif and a C-terminal 'PGACGE' motif, deemed "Type II" motifs [37,38]. Here, the 'proline effect,' or the tendency for increased fragmentation at the Pro N-terminus [39], generates a high abundance peak at 590.22 *m/z*, and is only coincidentally the approximate *m/z* of an over-alkylated Trp residue in the common 'SCVW' bracelet motif. These three examples described here illustrate the necessity of considering all possible fingerprint ion information in spectra for more accurate mass assignments. Interestingly, using this information to quickly assess all collected MS^2 in both *V. inconspicua* aerial and root material, it appears that Type II cyclotides are expressed solely in the aerial tissue, indicating a biological significance of this motif type, and opening the door to subsequent biological activity assessments.

5.3.4 Trypsin inhibitor fingerprint analysis

Mass spectral fingerprinting is not limited to bracelet and Möbius subfamilies, and can be translated to the facile identification and prioritization of trypsin inhibitor CRPs. Three highly homologous *Lagenaria siceraria* seed trypsin inhibitors have been characterized previously, LLTI-I, II, and III, and conform to the three disulfide connectivity of the cysteine-knot motif, albeit with a linear backbone [31]. Mass shift analysis of *L. siceraria* material revealed ~20 masses demonstrating a +348.16 Da mass shift upon derivatization. Of these, masses 3666.53 Da (917.64 *m/z*, +4, NI = 3318.37 Da) and 3795.61 Da (949.91 *m/z*, +4, NI = 3447.43 Da) were consistent with the *L. siceraria* trypsin inhibitors LLTI-II and LLTI-III, respectively (Table 3). MS^2 spectra of 3666.53 and 3795.61 Da reveal distinct and shared, *y*-ion fingerprint ions (*y2*: 236.07, y5: 543.22, *y8*: 945.39, *y9*: 1044.46, *y10*: 1204.49, *y11*: 1319.51, *y12*: 1390.56, and *y13*: 1503.64 *m/z*) consistent with known LSTI sequences (Figure 4A, B, Table 3). Exploration of the reduced and alkylated *L. siceraria* MS data resulted in the identification of an additional,

unknown mass 3852.61 Da (964.16 *m/z*, +4, NI = 3504.45 Da), differing from the mass of LLTI-III by 57.02 Da and demonstrating the characteristic LLTI-II and III fingerprint ions. A third trypsin inhibitor, LLTI-I, has been characterized in *L. siceraria* and differs in sequence from LSTI-III only by a modified N-terminal Glu (to a pyroglutamic acid residue, -18 Da) and the addition of a C-terminal Gly residue (where the mass of Gly is 57.02 Da). Relying on this mass difference, it may be assumed that 3504.45 Da is a LLTI-I sequence variant with an unmodified N-terminal Glu residue, differing only by a single Gly, and may be prioritized for sequence characterization to investigate this possibility. However, closer examination of the MS^2 spectrum (Figure 4C) reveals the same fingerprint *y*–ion series as in LLTI-II and III, and is incompatible with the LLTI-I sequence derived from an extra C-terminal Gly (which would offset all *y*-ions by the mass of Gly). With this information, it is likely that 3852.61 Da is an over-alkylated LLTI-III sequence. Though the site of alkylation cannot be immediately localized with the observed MS/MS fingerprints, as with the bracelet aromatic peak, this demonstrates the use of fingerprint ions in other CRPs to prioritize or de-prioritize mass species of interest.

New sequence types of novel cysteine-rich peptides may be detected through mass shift analysis complemented with MS^2 fingerprint ion identification, where masses that belong to the same class of CRPs can be grouped based on similar highly abundant MS^2 ions prior to sequence characterization. The three *L. siceraria* masses 2605.93 Da (869.65 *m/z*, +3, NI = 2257.76 Da), 2705.02 Da (902.68 *m/z*, +3, NI = 2356.86 Da), 2762.01 Da (921.68 *m/z*, +3, NI = 2413.85 Da) were detected with mass shift analysis to likely contain three disulfides and are not consistent with any characterized *L. siceraria* CRPs. MS^2 spectra of the reduced and alkylated mass species contained similar, highly abundant fragment ions (Figure 4D-F), indicating that these masses belong to the same sequence family of CRPs. Complete sequence elucidation via tandem MS

revealed three novel, linear sequences, confirmed to contain six Cys residues, differing each by only one amino acid residue, and are named LSCRP-I, II, and III (Table 3). The initially observed abundant fragment ions 300.17, 460.20, 717.28, 903.36, 1066.41, 1226.45, 1341.48, and 1501.52 *m/z* present among the LSCRP spectra correlate to a conserved *b*-ion series and can be used as a handle to identify additional LSCRP species, either in *L. siceraria* or perhaps other Cucurbitcaeae species. The biological function and activities, if any, of these peptides is unknown; however, the cysteine-rich nature of these small, homologous peptides dictates that follow up studies concerning biological activity should be pursued. Here, a standard mass shift analysis experiment coupled with on the fly MS^2 assessment has aided the facile identification of a small set of novel CRPs in *L. siceraria.*

5.4 Conclusion

Stable and commonly bioactive CRPs are typically discovered via mass shift analysis. However, the accurate assessment of unique CRP species in a given botanical species is often challenged by same mass species, post-translational modifications or those derived from sample handling, and incomplete MS^2 fragmentation. Mass spectral fingerprint ions can be leveraged to gain additional information about a mass species prior to full sequence characterization and with only poor quality MS^2 spectra. Orthogonal methods, e.g. transcriptomics, can be utilized to aid the accurate sequence characterization of targeted masses demonstrating MS/MS fingerprint ions. Herein we identify sets of mass spectral fingerprint ions characteristic of the CRP cyclotide family, which may indicate a mass belongs to a specific cyclotide subfamily, and “tell-tale” ions that are of importance when discriminating putative cyclotide species, including common oxidation and over-alkylation ions observed experimentally. Cyclotide-containing *Viola* material is used as proof-of-principle, where seasonal and tissue-specific cyclotide variation is explored.

Fingerprint ions derived from a third type of CRP, the trypsin inhibitors, are assessed for a single plant species, *L. siceraria.* Combining mass shift analysis with the identification of prominent MS^2 fingerprint ions is then used to identify three novel CRPs. We demonstrate that abundant mass spectral fingerprint ions can be used to quickly discern masses of interest in complex matrices and masses that are already characterized, aiding prioritization of the most promising novel mass species in a natural product sample for characterization.

5.5 Acknowledgments

The authors would like to thank Harvey Ballard, Ohio University, and Carol Ann McCormick, University of North Carolina Chapel Hill Herbarium (NCU), for curation of *Viola* specimens.

5.6 Funding

This work was funded by NIH (GM125814-01) to L.M.H. and the UNC Graduate School Dissertation Completion Fellowship to N.C.P.

5.7 Supporting Information

No supporting information.

5.8 Tables

Table 5.1. Top 15 most abundant theoretical *b/y* MS/MS fingerprint ions in, from left to right, all cyclotides, the bracelet subfamily, the Möbius subfamily, the cis-Trp/Pro Möbius, and the cis-Tyr-Pro Möbius. Masses are shown in the +1 charge state (+1 *m/z*).

All Cyclotides (423)				Bracelets (294)				All Möbius (129)				WP Möbius (93)				YP Möbius (36)			
m/z	No.	Ion Type	%	*m/z*	No.	Ion Type	%	*m/z*	No.	Ion Type	%	*m/z*	No.	Ion Type	%	*m/z*	No.	Ion Type	%
148.06	423	*y*	100	148.06	294	*y*	100	148.06	129	*y*	100	148.06	93	*y*	100	148.06	36	*y*	100
205.08	373	*y*	88	205.08	246	*y*	84	205.08	127	*y*	98	205.08	91	*y*	98	205.08	36	*y*	100
365.11	368	*y*	87	365.11	243	*y*	83	365.11	125	*y*	97	365.11	90	*y*	97	365.11	35	*y*	97
88.04	242	*b*	57	88.04	226	*b*	77	102.05	110	*b*	85	102.05	82	*b*	88	102.05	28	*b*	78
248.07	242	*b*	57	248.07	226	*b*	77	262.09	110	*b*	85	262.09	82	*b*	88	262.09	28	*b*	78
347.14	207	*b*	49	347.14	205	*b*	70	409.15	54	*b*	42	464.18	44	*y*	47	409.15	23	*b*	64
575.25	194	*y*	46	462.17	180	*y*	61	561.23	48	*y*	37	674.32	43	*y*	46	575.25	14	*y*	39
462.17	182	*y*	43	575.25	153	*y*	52	464.18	48	*y*	37	561.23	42	*y*	45	478.20	14	*y*	39
102.05	163	*b*	39	494.21	78	*b*	27	674.32	47	*y*	36	731.34	41	*y*	44	523.20	11	*b*	31
262.09	163	*b*	39	607.29	78	*b*	27	731.34	45	*y*	35	1001.48	31	*y*	33	537.25	11	*b*	31
533.22	87	*b*	21	632.27	77	*y*	26	575.25	41	*y*	32	845.38	31	*y*	33	688.33	9	*y*	25
632.27	80	*y*	19	704.34	76	*b*	26	478.20	41	*y*	32	409.15	31	*b*	33	466.18	7	*b*	19
494.21	78	*b*	18	864.37	76	*b*	26	523.20	36	*b*	28	575.25	27	*y*	29	510.20	7	*b*	19
607.29	78	*b*	18	533.22	71	*b*	24	845.38	35	*y*	27	478.20	27	*y*	29	567.22	7	*b*	19
704.34	76	*b*	18	646.30	70	*b*	24	1001.48	34	*y*	26	361.15	26	*b*	28	88.04	7	*b*	19

Table 5.2. Diagnostic *b/y* MS/MS fingerprint ions indicative of (A) all cyclotides, (B) bracelets, (C) Möbius, and (D) notable and often highly abundant bracelet ions that provide valuable information towards the identification of a mass species. **m/z* species generated from common, often unintentional, modifications (di-oxidation, alkylation) of the aromatic bracelet peak; values for percent bracelet, percent Möbius, bracelet identification, and Möbius identification represent a false positive value, and were calculated from the number of times this *m/z* would be produced by a cyclotide sequence in the CyBase database. †All %Bracelet and Bracelet ID values derived from this sequence; however, %Möbius and Möbius ID values were calculated from a same mass (510.20 *m/z*) but unique sequence: the bracelet and mobius values here are not derived from the same amino acid sequence, but both product the 510.20 *m/z* ion, necessitating that the 510.20 *m/z* value be assigned to a bracelet 'SCVY' motif only when observed in combination with 248.07 and 347.14 *m/z b*-ions.

All cyclotides							Bracelets						
m/z	Type	Sequence	%Bracelet	%Möbius	Bracelet ID	Möbius ID	*m/z*	Type	Sequence	%Bracelet	%Möbius	Bracelet ID	Möbius ID
148.06	*y*	E	100	100	70	30	248.07	*b*	SC	77	12	93	7
205.08	*y*	EG	84	98	66	34	347.14	*b*	SCV	70	2	99	1
365.11	*y*	EGC	83	97	66	34	462.17	*y*	EGCP	61	2	99	1
Möbius							Informative Bracelet Peaks						
m/z	Type	Sequence	%Bracelet	%Möbius	Bracelet ID	Möbius ID	*m/z*	Type	Sequence	%Bracelet	%Möbius	Bracelet ID	Möbius ID
102.05	*b*	T	18	85	33	67	494.20	*b*	SCVF	27	0	100	0
262.09	*b*	TC	18	85	33	67	510.20	*b*	SCVY†	20	9	85	15
409.15	*b*	TCF	4	42	18	82	533.21	*b*	SCVW	24	0	100	0
464.18	*y*	EGCV	1	37	6	94	565.21	*b*	SCVW(diOX)*	1	0	100	0
561.23	*y*	EGCVP	5	37	25	75	590.23	*b*	SCVW(Alk)*	1	0	100	0

Table 5.3. Sequences and molecular weight (Da) of *Lagenaria siceraria* known trypsin inhibitors LLTI-I, II, and III (top) and previously uncharacterized *Lagenaria siceraria* cysteine-rich peptides LSCRP-I, II, and III, bottom. Bolded, blue Glu residue (E) at the LLTI-I N-terminus is modified to a pyroglutamic acid.

LLTI-I	**E**RRCPRIYMECKHDSDCLADCVCLEHGICGG	3504.44 Da
LLTI-II	-RRCPRIYMECKHDSDCLADCVCLEHGICG-	3318.37 Da
LLTI-III	ERRCPRIYMECKHDSDCLADCVCLEHGICG-	3447.41 Da
LSCRP-I	WICPCWYCDCNPCTCPKTE-	2605.96 Da
LSCRP-II	WICPCWYCDCNPCTCPKTEV	2706.14 Da
LSCRP-III	WICPCWYCDCNPCTCPKTER	2763.06 Da

5.9 Figures

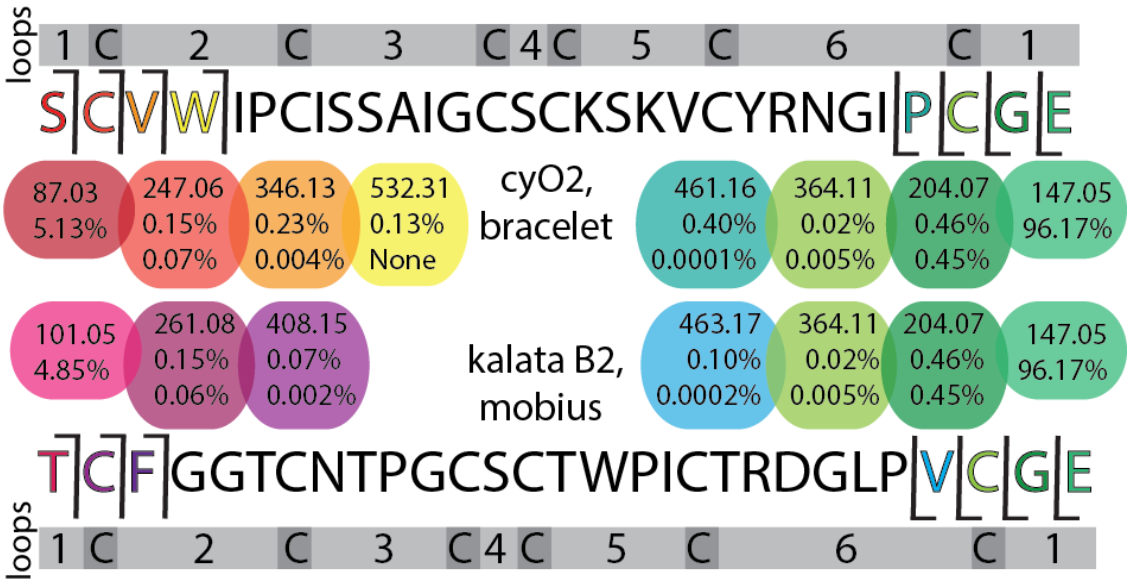


Figure 5.1. Fingerprint ions derived from the Glu-C digestion and *b/y* fragmentation of the prototypical cyclotides, cyO2, (top, bracelet) and kalata B2 (bottom, Möbius). Fragment ion color-coding aligns to the sequence colors; below the fragment ion neutral mass (Da) is the chance of finding that ion in the MS^2 of a non-cyclotide, followed by the chance of finding that ion in combination with those preceding it towards the nearest termini. Loop regions between cysteine residues are numbered above and below sequences in grey and allow for facile sequence comparison among cyclotide sequences.

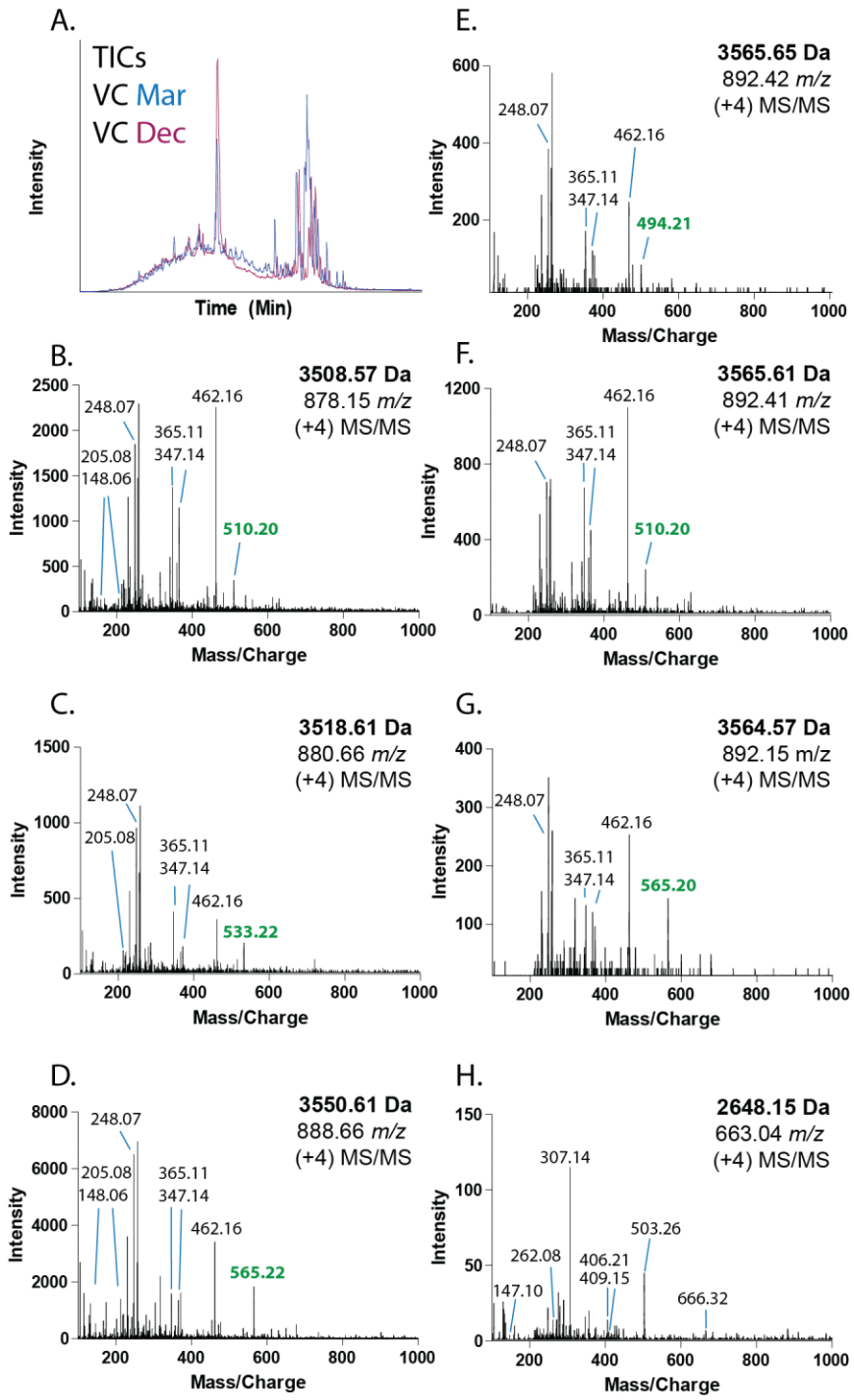


Figure 5.2. (A) Total ion chromatograms (TICs) of *V. communis* harvested in March (blue) and December (purple) tissues. (B-H) MS^2 spectra of *V. communis* reduced, alkylated, and Glu-C digested seasonal material. Common fingerprint ions are labeled; the bracelet aromatic peak often differentiates masses and is bolded in green in B-G.

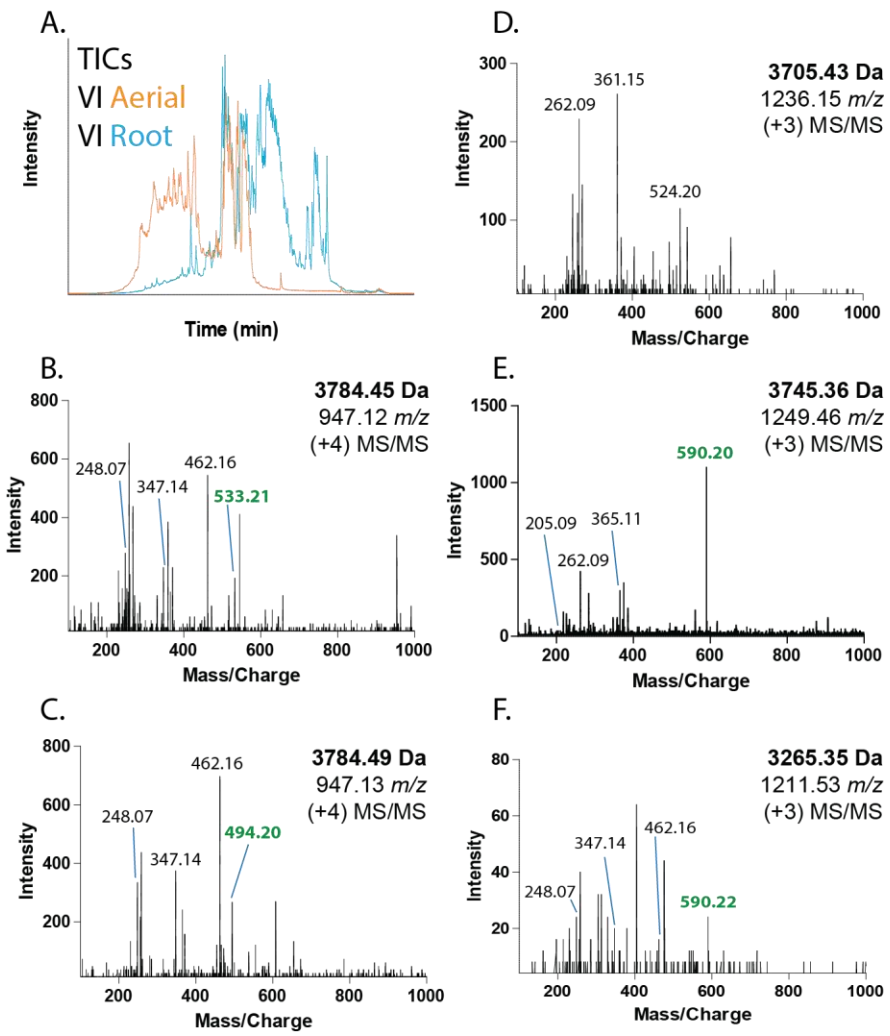


Figure 5.3. (A) Total ion chromatograms (TICs) of *V. inconspicua* aerial (orange) and root (blue) tissues. (B-F) MS^2 spectra of *V. inconspicua* reduced, alkylated, and Glu-C digested tissue-specific material. Common fingerprint ions are labeled; the bracelet aromatic peak often differentiates masses and is bolded in green in B, C, E, and F.

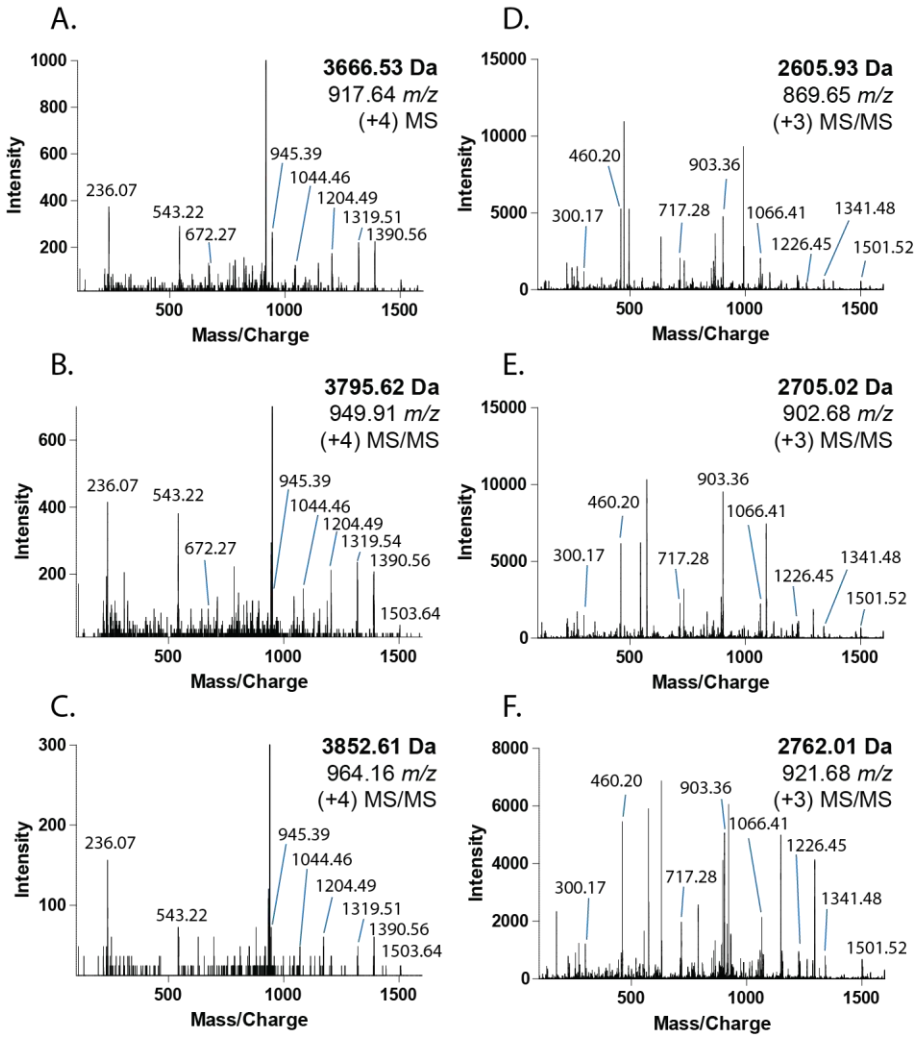


Figure 5.4. MS^2 spectra of *L. siceraria* known trypsin inhibitors (A-C) and novel cysteine-rich sequences (D-F); fingerprint ions common to each set of CRP are labeled.

REFERENCES

1. Tam, J.P., et al., *Antimicrobial Peptides from Plants.* Pharmaceuticals (Basel), 2015. **8**(4): p. 711-57.

2. Kumari, G., et al., *Cysteine-Rich Peptide Family with Unusual Disulfide Connectivity from Jasminum sambac.* J Nat Prod, 2015. **78**(11): p. 2791-9.

3. Marshall, E., L.M. Costa, and J. Gutierrez-Marcos, *Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development.* J Exp Bot, 2011. **62**(5): p. 1677-86.

4. Liu, X., et al., *Expansion and evolutionary patterns of cysteine-rich peptides in plants.* BMC Genomics, 2017. **18**(1): p. 610.

5. Loo, S., et al., *Bleogens: Cactus-Derived Anti-Candida Cysteine-Rich Peptides with Three Different Precursor Arrangements.* Front Plant Sci, 2017. **8**: p. 2162.

6. Shen, Y., et al., *Potentides: New Cysteine-Rich Peptides with Unusual Disulfide Connectivity from Potentilla anserina.* Chembiochem, 2019. **20**(15): p. 1995-2004.

7. Narayani, M., A. Chadha, and S. Srivastava, *Cyclotides from the Indian Medicinal Plant Viola odorata (Banafsha): Identification and Characterization.* J Nat Prod, 2017. **80**(7): p. 1972-1980.

8. Burman, R., et al., *Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae.* Front Plant Sci, 2015. **6**: p. 855.

9. Eriksson, J. and D. Fenyo, *Modeling mass spectrometry-based protein analysis.* Methods Mol Biol, 2011. **694**: p. 109-17.

10. Clark, C.M., et al., *Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function.* Proc Natl Acad Sci U S A, 2018. **115**(19): p. 4981-4986.

11. Mahe, P., et al., *Automatic identification of mixed bacterial species fingerprints in a MALDI-TOF mass-spectrum.* Bioinformatics, 2014. **30**(9): p. 1280-6.

12. Christner, M., et al., *Rapid identification of bacteria from positive blood culture bottles by use of matrix-assisted laser desorption-ionization time of flight mass spectrometry fingerprinting.* J Clin Microbiol, 2010. **48**(5): p. 1584-91.

13. Liu, H., et al., *Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry.* Appl Environ Microbiol, 2007. **73**(6): p. 1899-907.

14. Edwards-Jones, V., et al., *Rapid discrimination between methicillin-sensitive and methicillin-resistant Staphylococcus aureus by intact cell mass spectrometry.* J Med Microbiol, 2000. **49**(3): p. 295-300.

15. Holland, R.D., et al., *Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 1996. **10**(10): p. 1227-32.

16. Krishnamurthy, T., P.L. Ross, and U. Rajamani, *Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 1996. **10**(8): p. 883-8.

17. Vargha, M., et al., *Optimization of MALDI-TOF MS for strain level differentiation of Arthrobacter isolates.* J Microbiol Methods, 2006. **66**(3): p. 399-409.

18. Gruenwald, M., et al., *MALDI-TOF mass spectrometry fingerprinting: A diagnostic tool to differentiate dematiaceous fungi Stachybotrys chartarum and Stachybotrys chlorohalonata.* J Microbiol Methods, 2015. **115**: p. 83-8.

19. Chalupova, J., et al., *Identification of fungal microorganisms by MALDI-TOF mass spectrometry.* Biotechnol Adv, 2014. **32**(1): p. 230-41.

20. Kondori, N., et al., *Analyses of black fungi by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS): species-level identification of clinical isolates of Exophiala dermatitidis.* FEMS Microbiol Lett, 2015. **362**(1): p. 1-6.

21. Kharbach, M., et al., *Selected-ion flow-tube mass-spectrometry (SIFT-MS) fingerprinting versus chemical profiling for geographic traceability of Moroccan Argan oils.* Food Chem, 2018. **263**: p. 8-17.

22. Sawaya, A.C., et al., *Electrospray ionization mass spectrometry fingerprinting of propolis.* Analyst, 2004. **129**(8): p. 739-44.

23. Morgado Schmidt, E., et al., *Characterization of Royal Jelly by Electrospray Ionization Mass Spectrometry Fingerprinting.* Mass Spectrometry & Purification Techniques, 2015. **01**(01).

24. Acevska, J., et al., *Fingerprinting of morphine using chromatographic purity profiling and multivariate data analysis.* Journal of Pharmaceutical and Biomedical Analysis, 2015. **109**: p. 18-27.

25. Chen, P., et al., *Discrimination among Panax species using spectral fingerprinting.* J AOAC Int, 2011. **94**(5): p. 1411-21.

26. Catharino, R.R., et al., *Characterization of vegetable oils by electrospray ionization mass spectrometry fingerprinting: classification, quality, adulteration, and aging.* Anal Chem, 2005. **77**(22): p. 7429-33.

27. Wu, Z., R.P. Rodgers, and A.G. Marshall, *Characterization of vegetable oils: detailed compositional fingerprints derived from electrospray ionization fourier transform ion cyclotron resonance mass spectrometry.* J Agric Food Chem, 2004. **52**(17): p. 5322-8.

28. Huang, J., et al., *Cysteine-Rich Peptide Fingerprinting as a General Method for Herbal Analysis to Differentiate Radix Astragali and Radix Hedysarum.* Front Plant Sci, 2019. **10**: p. 973.

29. Islam, S.M.A., C.M. Kearney, and E. Baker, *Classes, Databases, and Prediction Methods of Pharmaceutically and Commercially Important Cystine-Stabilized Peptides.* Toxins (Basel), 2018. **10**(6).

30. Weidmann, J. and D.J. Craik, *Discovery, structure, function, and applications of cyclotides: circular proteins from plants.* J Exp Bot, 2016. **67**(16): p. 4801-12.

31. Hamato, N., et al., *Purification and characterization of serine proteinase inhibitors form gourd (Lagenaria leucantha Rusby var. Gourda Makino) seeds.* Biosci Biotechnol Biochem, 1992. **56**(2): p. 275-9.

32. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.

33. Tang, J., et al., *Isolation and characterization of cytotoxic cyclotides from Viola tricolor.* Peptides, 2010. **31**(8): p. 1434-40.

34. Svangard, E., et al., *Cytotoxic cyclotides from Viola tricolor.* J Nat Prod, 2004. **67**(2): p. 144-7.

35. Parsley, N.C., et al., *PepSAVI-MS reveals anticancer and antifungal cycloviolacins in Viola odorata.* Phytochemistry, 2018. **152**: p. 61-70.

36. Ireland, D.C., et al., *Discovery and characterization of a linear cyclotide from Viola odorata: implications for the processing of circular proteins.* J Mol Biol, 2006. **357**(5): p. 1522-35.

37. Parsley, N.C., et al., *Viola "inconspicua" No More: An Analysis of Antibacterial Cyclotides.* J Nat Prod, 2019. **82**(9): p. 2537-2543.

38. Foreman, D.J., et al., *Gas-Phase Sequencing of Cyclotides: Introduction of Selective Ring Opening at Dehydroalanine via Ion/Ion Reaction.* Anal Chem, 2019. **91**(24): p. 15608-15616.

39. Huo, D., T. Qin, and L. Zu, *Energetic switch of the proline effect in collision-induced dissociation of singly and doubly protonated peptide Ala-Ala-Arg-Pro-Ala-Ala.* J Mass Spectrom, 2019. **54**(1): p. 55-65.

CHAPTER 6: Implementation of microfluidics for antimicrobial susceptibility assays: issues and optimization requirements

*Manuscript to be submitted for publication. Authors: Nicole C. Parsley, Amanda L. Smythers, and Leslie M. Hicks

6.1 Introduction

Major advances in 20th century antimicrobial therapies significantly decreased mortality associated with microbial infections; however, a recent global surge of antimicrobial resistance (AMR) has made the treatment of common bacterial and fungal infections increasingly challenging. Once viable antimicrobial therapies are now obsolete as new mechanisms for AMR develop, accelerated by the "misuse and overuse" of antimicrobial drugs.[1] Additionally, most recent commercial drug discovery efforts have targeted chronic ailments, ushering in a decline in antimicrobial approvals over the past three decades.[2] As such, the future of global health relies on an infusion of new antimicrobial therapeutics in the drug discovery pipeline.

Natural product bioactive peptides are a vast and largely untapped source of antimicrobial chemistries expressed by all life, offering innate antimicrobial, antiviral, and anticancer properties.[3] Natural product drug discovery has often favored the identification and isolation of small molecule constituents; however, antimicrobial peptides (AMPs) offer a complementary activity profile with the potential for increased selectivity and efficacy against pathogens.[4, 5] Peptide-based therapeutics debuted in the clinic in WWII with the soil bacilli-derived tyrothricin (a mixture of gramicidins and tyrocidines), a potent topical antimicrobial agent that predated the commercialization of penicillin.[6] Despite the characterization of >3000 AMPs since then,[7] only seven peptide antimicrobial drugs are currently approved by the FDA.[8] Antimicrobial peptide discovery, characterization, and ultimately the path to clinical relevance is hindered by extreme

diversity in sequence, size (2 to >50 residues), charge, post-translational modification (e.g. cyclic, disulfide-bound), and natural product complexity.

To address this gap, PepSAVI-MS was developed to identify low abundance bioactive peptides in complex natural product extracts.[9-13] PepSAVI-MS is a top-down peptidomics approach that leverages modern mass spectrometry and relies on traditional bioassays for bioactivity characterization (e.g. 96-well plate or disk-diffusion). The sample quantities needed (~ micrograms) are often a limiting factor towards the rapid identification of novel AMPs. Plate-based assays offer increased sensitivity and higher throughput compared with traditional agar disk-diffusion assays, and are largely accessible in inexpensive, commercially-available 6, 12, 24, 96, and 384-well formats. Although higher density wells allow for greater throughput and require less sample volume per well, wells with smaller volumes (e.g. 384-well) can inhibit pathogen growth due to challenges with oxygenation[14] and mixing.[15, 16] Plate-based assays require optimization regarding pathogen identity, cell seeding density,[17] aeration,[14] media choice,[18, 19] and assay length,[18] and are unable to replicate physiological conditions.

Optical density (O.D)[20] and fluorescence-based measurements[21, 22] are among the most commonly used methods to determine cell viability in plate-based assays. Optical density reads are often hindered by a low dynamic range and are highly susceptible to inhomogeneities of the solution tested, which can be problematic with bacterial cells prone to clumping and biofilm formation.[20] Frequently used as a measurement directly proportional to cell density, O.D. must be calibrated based on specific species and measurement conditions.[20] Metabolically-driven fluorescent dyes, e.g. resazurin/resorufin, have a large dynamic range, however; these are subject to extensive non-specific interactions with non-bacterial samples components (particularly thiol and carboxylic acid moieties) – a major cause of false positives/negatives.[23] Resazurin must be

intracellularly aerobically reduced by metabolically active cells to fluorescent resorufin, requiring >30 min incubation time prior to analysis,[24, 25] and the ability of bacterial species to participate in this reduction can be strain dependent due to differences in membrane permeability. While resazurin is effectively used for endpoint analysis, the further reduction of resorufin into a colorless non-fluorescent product makes it challenging to monitor cell growth over time.[26] Despite common use in testing for antibiotic susceptibility, both O.D. and fluorescence-based bioactivity assays often require significant optimization prior to reliable use.[20, 25, 27] Importantly, plate-based assays offer only a macroscopic view of indirectly measured bioactivity and provide no insight into cellular effects of treatment. Alternative assay formats with minimized optimization requirements and sample consumption are needed to enable rapid and accurate bioactivity assessment.

Emerging and highly versatile microfluidic technologies offer an alternative format for bioactivity testing; operating at micrometer scales and down to picoliter volumes, microfluidic platforms offer significant experimental flexibility via customizable fabrications, where channels and chambers may be precisely designed to answer specific biological questions. These technologies enable bioactivity assessment through microscopic detection of bacterial growth/inhibition, facilitating the direct evaluation of AMP mechanism of action, and boast minute sample/reagent requirements, decreased assay times, and highly controlled environmental conditions.[28] Microfluidics have been used to assay “unculturable” bacteria while screening for bioactive compounds with higher sensitivity and in a 4-24x decreased time frame needed for traditional approaches.[29] Furthermore, microfluidic platforms may better imitate physiological microsystems (e.g. blood vessels and anoxygenic environments), supporting clinically-relevant quantitative evaluation of AMP efficacy that supersedes that of bulk assays.[30]

Microfluidic platforms are adaptable, allowing for cell viability analysis via fluorescent/cell staining dyes and through direct cell counting with the acquisition of brightfield images.[31] Metabolically/biomolecule-specific fluorescent and cell staining dyes (e.g. Hoechst 33342) can be incorporated into microfluidic platforms to provide insight into mechanism of action and the differential effects of antibiotic treatment on inhomogeneous cell populations. Although the use of fluorescence-based assays must often contend with cellular phototoxicity during fluorophore excitation, the ability to gain complementary data such as membrane potential, membrane integrity, and DNA degradation may be invaluable in advancing the drug discovery pipeline. Alternatively, brightfield imaging of microfluidic channels/chambers offers a simple, direct, and dye-free measurement of cell viability[32] through bacterial cell counts and may provide insight into the antibacterial mechanism of action[33]; thus, brightfield imaging is a preferred metric for cell viability during microfluidic platform optimization. Regardless, brightfield imaging still presents analytical challenges, where the imaging of only a cross-section of a three-dimensional channel and the inhomogeneity of cells throughout microfluidic channels/chambers can mislead accurate interpretation of bacterial growth.[34]

Herein, a microfluidics-based bioactivity assay is compared with that of a traditional bulk 96-well plate in order to assess its potential to elevate PepSAVI-MS AMP identification and biological characterization. A natural product peptide library with previously reported antibacterial bioactivity[10] was generated from the cyclotide-expressing botanical species *Viola inconspicua* and screened against *Escherichia coli* ATCC 25922 (Figure 1). Bioactivity is determined through a polymethylmethacrylate (PMMA) microfluidic platform coupled with confocal brightfield microscopy and compared to that of polypropylene (PP) and polystyrene (PS) plate-based assays using both optical density and resazurin-based fluorescence

measurements. Despite similarities in the activity profiles of PP and PS plate-based assays, *V. inconspicua* library bioactivity was highly variable among plate-based and microfluidic platforms. These results show that significant optimization must be pursued prior to the incorporation of microfluidic technologies into established antimicrobial susceptibility pipelines.

6.2 Methods

6.2.1 Plant material

Viola inconspicua seeds (Mountain Gardens, Burnsville, NC) were grown in a laboratory greenhouse (14/10 light/dark, 17.5-20.3 °C) to mature rosettes. *V. inconspicua* greenhouse specimens have been submitted to the UNC Herbarium under accession numbers NCU00303107 and NCU00303108 and can be viewed on the SERNEC (Southeast Regional Network of Expertise and Collections) website (http://sernecportal.org/portal/index.php).

6.2.2 Peptide library

Plant material was extracted and peptide library prepared as previously described [9, 10] (Figure 1). Briefly, *V. inconspicua* leaf tissue was harvested, ground under liquid nitrogen, and extracted into 10% acetic acid (1:3 w/v, 4 h, 4 ˚C, with stirring). Extract was sterile-filtered (0.22 μm, Corning), filtered to remove high molecular weight components (30 kDa MWCO, Millipore), and dialyzed into 5 mM ammonium formate, 20% acetonitrile, pH 2.7 to remove molecules < 1 kDa (0.1-1 kDa MWCO, SpectrumLabs). Remaining small molecule components were removed by injecting the extract onto strong cation exchange (PolySulfoethyl A column, 100 mm × 4.6 mm, 3 μm particles, PolyLC) using a linear salt gradient (5 mM ammonium formate, 20% acetonitrile, pH 2.7 to 500 mM ammonium formate, 20% acetonitrile, pH 3.0) and a flow rate of 0.5 mL/min. All late-eluting peptide-containing fractions were collected and combined, desalted with a C_{18} solid-phase extraction (100 mg, Waters Sep-Pak), dried down in a vacuum

concentrator, and resuspended in water. A *V. inconspicua* reversed-phase peptide library was prepared by injecting this sample onto a reversed-phase column (Phenomenex Jupiter C_{18} column, 300 Å, 5 μm, 4.6 mm x 150 mm) where mobile phase A (MPA) consisted of 95/5 water/acetonitrile with 0.1% TFA; mobile phase B (MPB) consisted of acetonitrile with 0.1% TFA, using the following gradient: 0 % MPB from 0-10 m, 0-20 % MPB from 10–13 m, 20-40 % MPB from 13-33 m, 40-100 % from 33-43 m, and hold at 100 % B from 43-47 m, a flow rate of 0.5 mL/min, and collecting 1 min fractions. *V. inconspicua* peptide reversed-phase library fractions were dried down via vacuum centrifugation and resuspended in acidified LC-MS grade water.

6.2.3 LC-MS/MS Analysis

V. inconspicua library fractions 34-39 (~ 1 μg) were injected onto a nano-LC-ESI-MS/MS consisting of a NanoAcquity LC (Waters, Milford, MA) coupled to a TripleTOF5600 MS (AB Sciex, Framingham, MA). Front-end UPLC separation of peptides was accomplished via a Symmetry C18 trap column (100 Å, 5 μm, 180 μm × 20 mm, Waters) and an HSS T3C18 column (100 Å, 1.8 μm, 75 μm × 250 mm, Waters), with a flow rate of 0.3 μL/min and a 30 min linear ramp of 5%−50% B (mobile phase A, 1% formic acid in water; mobile phase B, 1% formic acid in acetonitrile). The TripleTOF5600 MS was operated in positive-ion, high-sensitivity mode, and an IDA method with a cycle time of 2 s, the MS survey spectrum using a mass range of 350−1600 Da in 250 ms, and the first 20 *m/z* from low to high *m/z* selected for MS/MS, acquiring each MS/MS over ~87 ms. Deisotoped peak lists for each library fraction were generated using Progenesis QI for Proteomics software (Nonlinear Dynamics, v.2.0) with a retention time filter of 14−45 min and a maximum charge of +10. “Peptide ion data” was

exported from Progenesis and used to inform the relative abundances of *m/z* across *V. inconspicua* library fractions for PepSAVI-MS statistical analysis.

6.2.4 Bacterial cultures

Escherichia coli ATCC 25922 frozen glycerol stocks were streaked onto Mueller-Hinton broth (MHB, BD difco) agar plates and grown overnight at 37 °C. A colony from each plate was used to inoculate 5 mL of MHB in a culture tube, grown 8-16 h at 37 °C, 250 rpm, and subsequently used to inoculate plate-based assay and microfluidic samples as described below.

6.2.5 Plate-based assay

PP and PS 96-well plates were used to assess *V. inconspicua* library bioactivity in triplicate. Bacteria were adapted to liquid growth as described above, diluted back to OD_{600} = 0.25 in MHB, and grown for 1 h prior to the experiment to regain active growth. Each sample assayed consisted of 10 µL bacterial culture (diluted for a final concentration of OD_{600} = 0.1 in well), 20 µL 1X MHB, 10 µL 2X MHB, and 10 µL *V. inconspicua* library fraction. *V. inconspicua* library fractions 34-39 were assayed and positive and negative controls were prepared by replacing library fraction with ampicillin and water, respectively. Plate assays were incubated at 37 °C for 4 h, shaking at 250 rpm. At t=4 h, OD_{600} reads were acquired on a SpectraMax M5 (Molecular Devices). For fluorescence analysis, 1 µL resazurin sodium salt (Sigma) was added to each well for a final concentration of 1 mM and the plates were incubated at 37 °C, 250 rpm. After 1 h or 1.5 h (t = 5 h or 5.5 h total assay time) for PS and PP plates, respectively, fluorescence of each well was measured (SpectraMax M5), with excitation at 544 nm and emission at 590 nm.

6.2.6 Microfluidics

An 8-channel PMMA microfluidic chip (Microfluidic-Chipshop, Germany) with channels 100 µm x 100 µm x 18 mm and PP Luer fluidic interface was used for the microfluidic assay in

singlet. Each channel was flushed with 100 µL 2.5 M NaOH, 100 µL sterile water, and equilibrated with 100 µL MHB prior to inoculation. After initial adaptation to liquid growth as described above, bacterial cultures were diluted to OD_{600} = 0.5 in MHB. Approximately 40 µL of each sample was loaded into a channel via pipette tip (until excess flowed through the outlet), consisting of the same proportions of bacterial culture (diluted to a final in-chip concentration of OD_{600} = 0.1), 1X MHB, 2X MHB, and *V. inconspicua* library fraction. *V. inconspicua* library fractions 34-39 were assayed and positive and negative controls were prepared by replacing library fraction with ampicillin and water, respectively. The microfluidic chip was incubated at room temperature for 1 h prior to bacterial growth assessment via confocal brightfield imaging.

6.2.7 Microscopy

Brightfield contrast and fluorescence images were acquired on a Zeiss LSM710 confocal laser-scanning microscope, with PlanApo 40X and 10X 0.45NA objective lenses, respectively. Brightfield and fluorescence images were collected using 561 nm and 405 nm lasers, respectively, on Zeiss ZEN software (Car Zeiss, INC. NY, USA). Bacterial cell viability was assessed via manual counting of bacterial cells in brightfield images.

6.2.8 Cell counts

Differences in cell counts (Δcell) were calculated by subtracting initial cell counts from final cell counts for each replicate, averaged, and standard deviation calculated. Bioactivity calculations were performed using Δcell counts for each channel, each replicate, in the following equation:

$$\%\ Activity = \left[1 - \left(\frac{V.inconspica\ fraction\ \Delta cell - positive\ control\ \Delta cell}{negative\ control\ \Delta cell - positive\ control\ \Delta cell}\right)\right] * 100$$

6.3 Results and Discussion

6.3.1 Cyclotide library

Cyclotide-containing peptide libraries consistently demonstrate potent antibacterial activity in 96-well plate assay format,[10] and were thus utilized for comparison analyses of plate-based and microfluidics activity assays. *V. inconspicua* library fractions 34-39 were analyzed for peptidyl constituents via LC-MS and chosen for the high intensity UV traces and late retention time characteristic of cyclotide species. Mass spectral (MS1) analysis revealed that all library fractions contain masses corresponding to cyclotides or putative cyclotide species previously identified in *V. inconspicua* (Figure 2).[10] Highly abundant masses present in the *V. inconspicua* library fractions are consistent with the masses of the fully characterized cyclotides cyO8 (fractions 38-39), cyO9 (fractions 35-38), viba 11 (fractions 35-36), cyI1 (fractions 35-36), cyI2 (fractions 38-39), and cyO8/cyO9 oxidized variants (fractions 34/38). Additional high abundance masses 3271 Da (fraction 34), 3275 Da (fraction 34-36), and 3229 Da (fraction 37) were identified as putative cyclotide species based on a mass shift analysis,[10] however, these sequences have not yet been characterized. The presence of different cyclotide species throughout the fraction subset, with at least one cyclotide mass in every fraction, and the propensity of cyclotides to exhibit antimicrobial activities, supports the use of this fraction library for assay comparisons.

6.3.2 Bioactivity of 96-well plate assays

V. inconspicua library fractions 34-39 were assessed for bioactivity in conventional PP and PS 96-well plates (Figure 3). Optical density and fluorescence measurements conferred significant bioactivity in the *V. inconspicua* library fractions 38 and 39. However, activity was dependent on the measurement of cell viability (absorbance versus fluorescence) and the plate material (PP

versus PS). Among the active fractions in the PP and PS plate assays, optical density reads averaged 60 ± 5 and 37 ± 16 % less bioactivity than fluorescent reads for fractions 38 and 39, respectively; accordingly, fluorescent resazurin/resorufin measurements can achieve > 6-fold higher signal to background[25] in comparison to O.D. measurements.[35] Polypropylene and PS optical density-based activity measurements of fraction 38 were within error of each other, at 46 ± 6 and 51 ± 8 % activity, respectively. Likewise, fluorescence measurements of fraction 38 for both plate materials were within error with 81 ± 5 and 80 ± 2 % for PP and PS. However, significant differences were seen in the bioactivities of PP O.D. and fluorescence measurements of fraction 39, PS O.D. and fluorescence measurements of fraction 39, and the non/less active fractions 34-37 in the PP plate. Significant growth promotion was seen only in the PP O.D. measurements of fractions 34-37. Generally, both PP and PS O.D. reads exhibited higher standard deviations in activity measurements versus fluorescence-based measurements, reflecting the inhomogeneity of bacterial cultures.

6.3.3 Bioactivity in microfluidic assays

V. inconspicua library fraction activity was assessed with a simple 8 straight-channel microfluidic chip (Figure 1) to explore the merits of a microfluidic bioassay format. Whereas traditional "bulk" assays (e.g. 96-well plate) often require a minimum of 50 µL of sample mixture in each well[36-38] (10 µL extract, 40 µL of media and cells for each replicate),[9-11, 39] microfluidic devices utilize sample sizes in the nanoliter/picoliter range. However, connectors and interfaces between microfluidic channels/chambers and injection sites can generate significant dead volume, requiring excess sample to ensure complete filling of microfluidic compartments. Although <1 µL was theoretically needed to fill each microfluidic test channel, 25 µL was used during injection to completely fill the channel and Leur tip-fitted wells, only

decreasing the required sample quantity by a factor of two when compared with a 96-well plate format. Additionally, microfluidic chips are made from plastic materials, e.g. PMMA or polydimethylsiloxane (PDMS), that are prone to degradation upon reuse and deformations during manufacturing. These imperfections were visible under a confocal microscope and contributed to varying refractive indices, resulting in poor image quality and reduced standardization.

In contrast to the plate-based assays, no bioactivity was observed in the microfluidic chip – with growth promotion dominating in most channels (Figures 4A-4C). Extreme growth variation among experimental replicates is seen by large standard deviations in bacterial growth across all channels; only positive control antibiotic-containing channels showed negligible deviation in cell count differences (final-initial cell counts) among experimental replicates, where averaging even negative control (water) fractions resulted in an average cell difference (Δcell counts) of 68 ± 23 cells (Figure 4B). Notably, despite care to homogenize master sample mixes prior to loading, samples within the channel were highly heterogeneous. Large variations in cell counts among focal planes and chamber latitudes and visible cell clumping at random intervals were likely the source of inaccurate or erratic cell counts throughout this experiment. While using Z-stacks could enable counting across focal planes, the imperfections in fabricated microfluidic chips prevented standardized Z-stacks across different channels. Poor reproducibility across replicates was likely a result of channel imperfections, varying refractive indexes, and the three-dimensional nature of the channel.

6.3.4 Comparison of plate-based and microfluidic assays

Bioactivity differences in fractions 38/39 between plate-based and microfluidic assays were striking. The microfluidic chip used was comprised of PMMA with PP Leur connectors on the inlets/outlets of the device, while the 96-well assay plates were comprised of PP and PS.

Polymethylmethacrylate is a biocompatible hydrophilic plastic, while PP/PS are inexpensive hydrophobic plastics used to produce the majority of assay plates. Thus, hydrophobic/amphipathic biomolecules (e.g. AMPs) are subject to extensive PP/PS surface binding via hydrophobic interactions.[40, 41] Though PP materials bind polar molecules less than PS, the opacity of PP hinders the measurement of bacterial growth via optical density reads.[41] The surface properties of PMMA, including charge densities and reactive functional groups, are significantly impacted by the initial fabrication method and temperature, with potential impact on cell growth, homogeneity, and bioavailability of bioactive compounds.[42, 43] In any case, subsequent activity assays may benefit from passivation of assay plates/chips with inert biomolecules (e.g. polyethylene glycol, bovine serum albumin) prior to bacterial inoculation and bioactivity assessment in order to lessen the effect of surface interactions on experimental results.

Hydrophobicity. The *V. inconspicua* peptide library used in all analyses was generated via reversed-phase LC, where later eluting fractions (e.g. 38/39) contain hydrophobic cyclotide species. Major concerns regarding peptide non-specific binding to plastic surfaces[44, 45] may explain a decrease in peptide concentration of hydrophobic peptides in hydrophobic activity assay plates. However, our experimental results counter this reasoning, as increasingly hydrophobic cyclotides are active in assays with hydrophobic material (PP/PS). Surface binding interactions are complicated by the propensity for cyclotides to oligomerize; for example, the cyclotide kalata B2 self-associates into tetramers and octamers in buffer where the 3-D hydrophobic face characteristic of cyclotides is "quenched,"[46] and this behavior can affect surface binding interactions.[47] It can therefore be surmised that peptide binding is not only determined by hydrophobic interactions; rather, a complex combination of hydrophobic

interactions, peptide conformations and orientations, as well as media, can affect the susceptibility of peptide adsorption to various surfaces.[43, 48]

Oxygenation. Oxygenation plays an important role in bacterial growth, response to environmental cues, and antibiotic susceptibility.[49] Significantly different oxygen levels are experienced in microfluidic systems and plate-based assays. PMMA has low gas permeability,[50] with longer assays (>1 hour) requiring an infusion of new media containing dissolved oxygen to support bacterial growth. Contrastingly, plate-based assays incorporate considerably more oxygen when shaking during incubation. The clinical relevance of bioactivity assay formats must be assessed: while assays are typically performed in ambient oxygen levels, decreased oxygen may be more representative of certain clinical environments (e.g. burns, abscesses, oral cavity).[49] Although measured activity differences between microfluidic and plate-based assays are striking and could originate in part from the increased oxygenation with 96-well plates, producing the same activity profiles in different bioassay formats may be irrelevant as no assay format imitates true biological conditions. While there is no substitute for dynamic biological systems, *in vivo* systems are also flawed: animal models often do not correlate directly to human biology and cannot resolve the same mechanistic detail as *in vitro* assays.[51] As such while challenges exist between high throughput bioactivity assays, they are still the first line of identification and characterization of novel antimicrobial compounds.

6.4 Conclusion

Metrics of compound bioactivity include a host of assay platforms and reagents, where specialty assay designs and smart combinations of dyes and fluorophores allow us to peer beyond basic bioactivity assessment to the mechanisms of action and specific cellular effects of antimicrobial treatments. While microfluidic technologies are an emerging alternative to

conventional bulk activity assays with the potential for high-throughput antimicrobial compound discovery and characterization,[28] the integration of microfluidic devices into current drug discovery platforms requires careful parameter consideration and significant optimization. Polypropylene and PS plate-based assays, largely confer similar bioactivity profiles for matched samples. However, dramatic differences in bioactivity for plate-based vs. microfluidic assays are observed; where the latter demonstrated no significant activity or growth promotion in all channels. Though the basis of these differences is unclear, major influences could be derived from (a) the plastic materials comprising each assay device and/or (b) the growth/assay conditions. Importantly, the biological relevance, the benefit of a specific assay type, and the merit of using multiple assay platforms for orthogonal compound identification is not clear, and may be complicated by AMP/pathogen identity. Microfluidic devices may offer an alternative activity assay form with the potential for significant experimental flexibility in design and output; however, significant optimization of a microfluidic platform is required before standardization and common laboratory use.

6.5 Acknowledgments

The authors would like to thank Tony Perdue (Imaging Core Facility, The University of North Carolina at Chapel Hill) for microscopy training and experimental advice. This work was funded by NIH (GM125814) to L.M.H. and the UNC-CH Graduate School Dissertation Completion Fellowship to N.C.P.

6.6 Figures

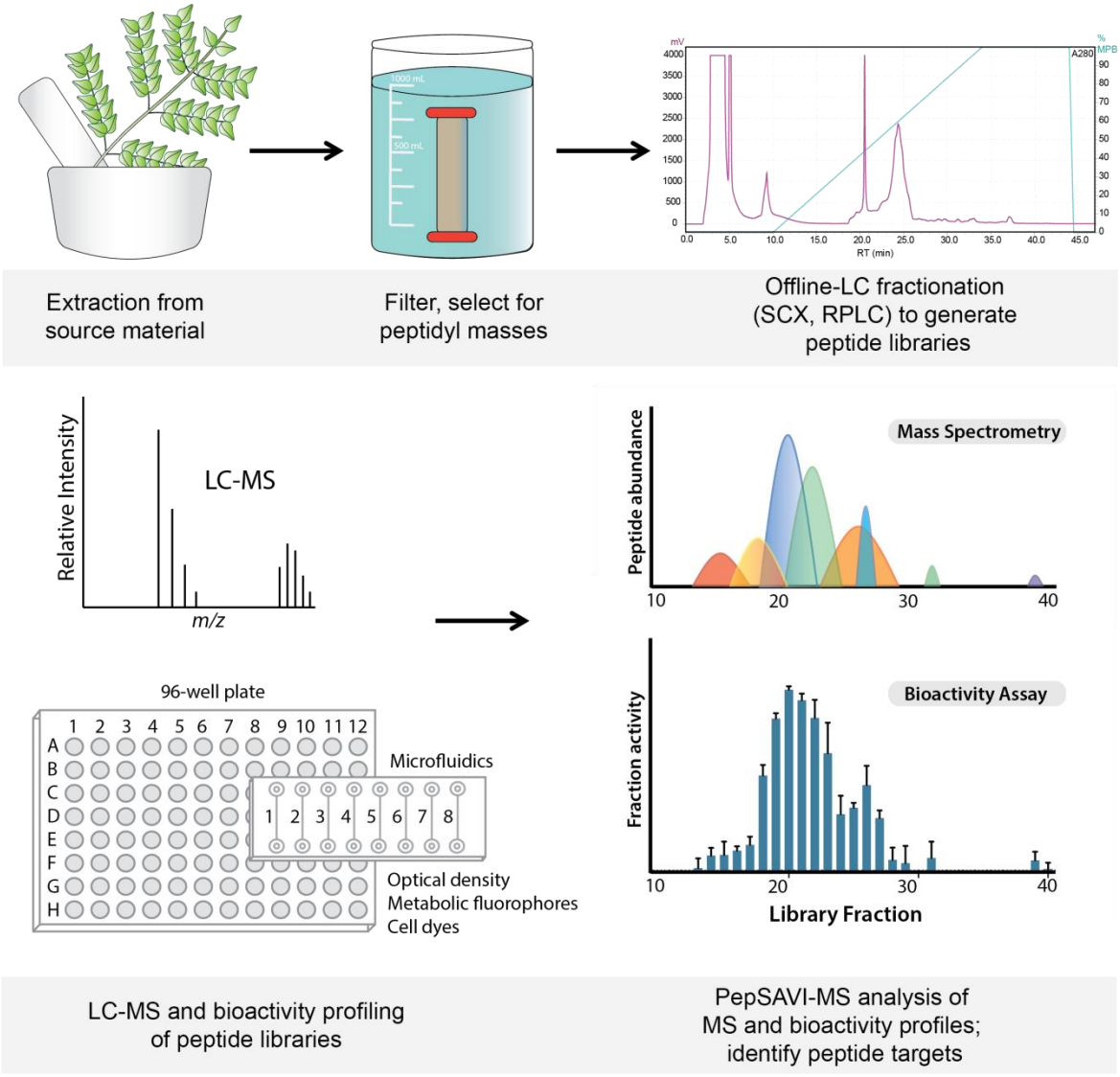
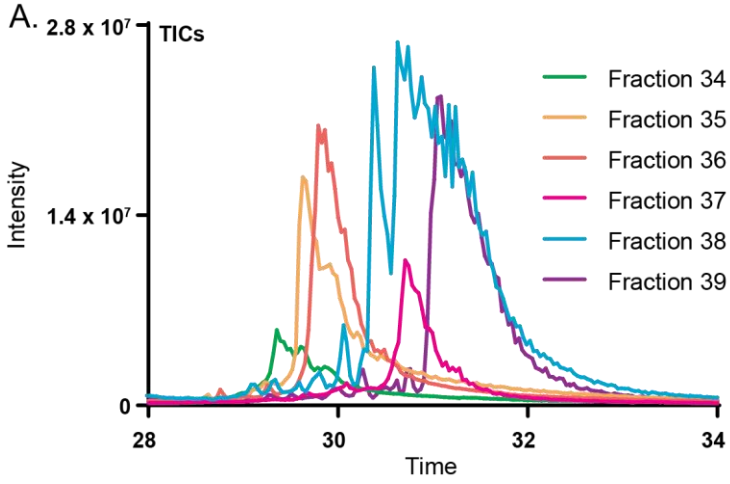


Figure 6.1. Workflow from starting material extraction through filtering and fractionation steps, generating a peptide library (top). Libraries are screened via LC-MS and for bioactivity (bottom), and subsequent PepSAVI-MS statistical analysis guides the identification of putative bioactive peptide species.

A.



B.

F34	F35	F36
CyO8(ox) CyO9(ox) 3271 Da 3275 Da	Viba 11 CyO9 Cyl1 3275 Da	Viba 11 CyO9 Cyl1 3275 Da
F37	**F38**	**F39**
3229 Da CyO9	CyO8 CyO9 CyO8(diox) Cyl2	CyO8 Cyl2

Figure 6.2. Overlaid total ions chromatograms (TICs) of *V. inconspicua* peptide library fractions 34-39 (A) and cyclotide constituent analysis of each fraction (B). Oxidized and doubly oxidized species are denoted by (ox) and (diox), respectively.

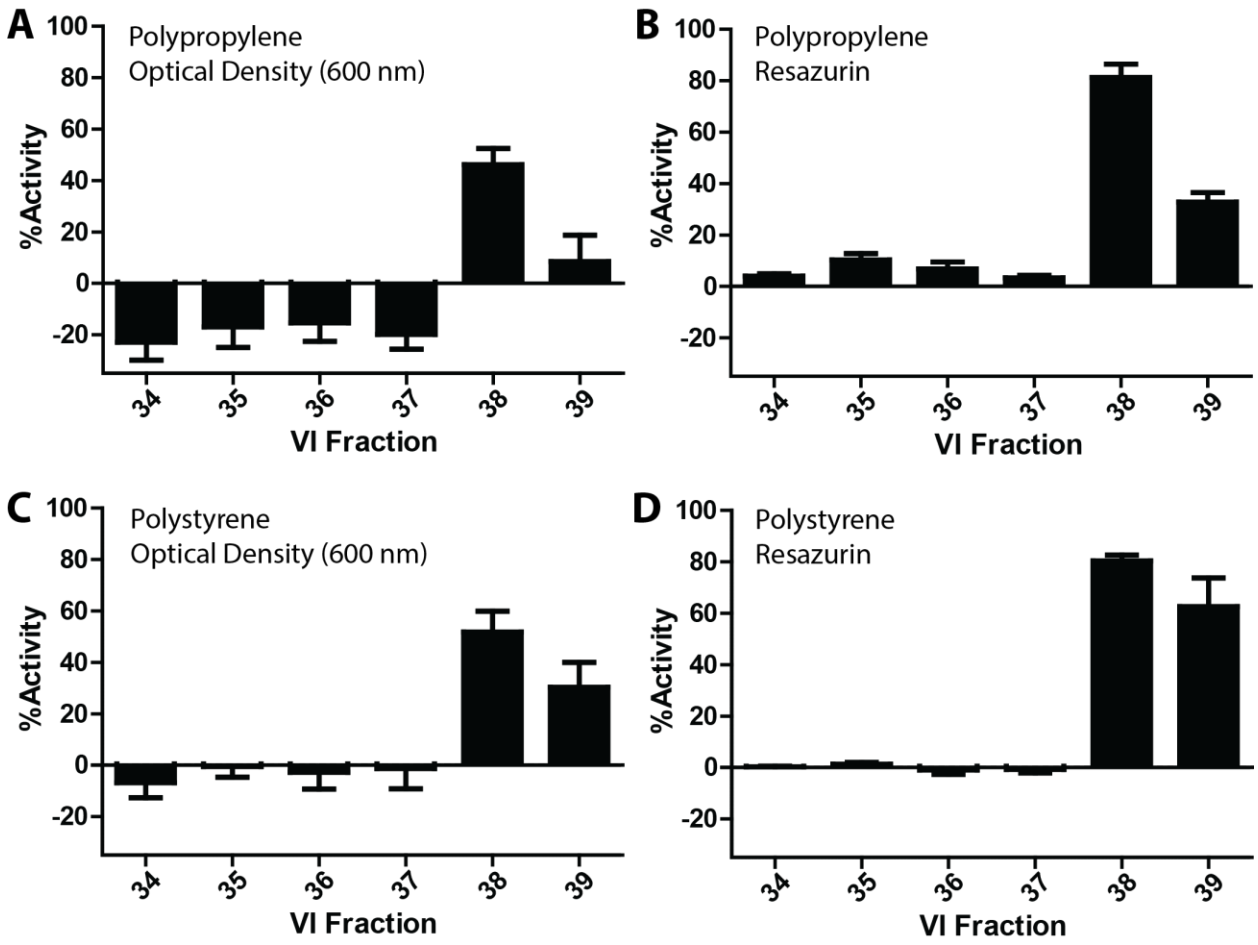


Figure 6.3. Bioactivity profiles of a *V. inconspicua* library against *E. coli* ATCC 25922 in a 96-well plate format and cell viability assessment with optical density in polypropylene (A), fluorescence in polypropylene (B), optical density in polystyrene (C), and fluorescence in polystyrene (D). Assays performed in triplicate, error bars represent standard deviation.

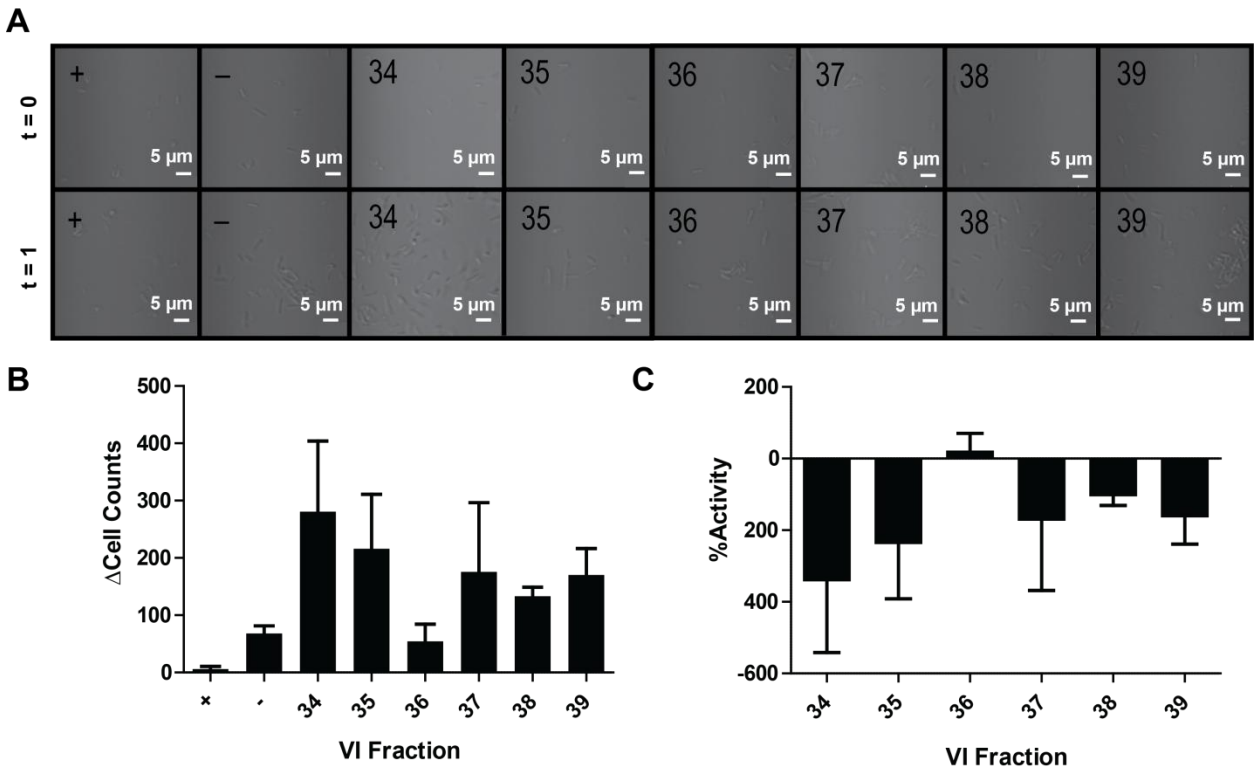


Figure 6.4. Activity of peptide fractions as measured through brightfield microscopy. Microfluidic channels are imaged at the initial assay time point and after 1 h of growth (A). Differences in initial to final cells counts are calculated from brightfield images in positive control (+), negative control (-), and *V. inconspicua* library fraction replicates (B). The percent bioactivities of *V. inconspicua* fractions are calculated from cell count differences (C).

REFERENCES

1. *2018 Antimicrobial Resistance Fact Sheet*. March 8, 2020.

2. Cole, S.T., *Who will develop new antibacterial agents?* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1645): p. 20130430.

3. Chen, C.H. and T.K. Lu, *Development and Challenges of Antimicrobial Peptides for Therapeutic Applications.* Antibiotics, 2020. **9**(1): p. 24.

4. Fosgerau, K. and T. Hoffmann, *Peptide therapeutics: current status and future directions.* Drug Discovery Today, 2015. **20**(1): p. 122-128.

5. Lázár, V., et al., *Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides.* Nature Microbiology, 2018. **3**(6): p. 718-731.

6. Wenzel, M., et al., *The Multifaceted Antibacterial Mechanisms of the Pioneering Peptide Antibiotics Tyrocidine and Gramicidin S.* mBio, 2018. **9**(5).

7. Wang, G., X. Li, and Z. Wang, *APD3: the antimicrobial peptide database as a tool for research and education.* Nucleic Acids Res, 2016. **44**(D1): p. D1087-93.

8. Divyashree, M., et al., *Clinical Applications of Antimicrobial Peptides (AMPs): Where do we Stand Now?* Protein Pept Lett, 2020. **27**(2): p. 120-134.

9. Kirkpatrick, C.L., et al., *The "PepSAVI-MS" Pipeline for Natural Product Bioactive Peptide Discovery.* Anal Chem, 2017. **89**(2): p. 1194-1201.

10. Parsley, N.C., et al., *Viola "inconspicua" No More: An Analysis of Antibacterial Cyclotides.* J Nat Prod, 2019. **82**(9): p. 2537-2543.

11. Moyer, T.B., et al., *PepSAVI-MS Reveals a Proline-rich Antimicrobial Peptide in Amaranthus tricolor.* J Nat Prod, 2019. **82**(10): p. 2744-2753.

12. Kirkpatrick, C.L., et al., *Fungal Secretome Analysis via PepSAVI-MS: Identification of the Bioactive Peptide KP4 from Ustilago maydis.* J Am Soc Mass Spectrom, 2018.

13. Kirkpatrick, C.L., et al., *Exploring bioactive peptides from bacterial secretomes using PepSAVI-MS: identification and characterization of Bac-21 from Enterococcus faecalis pPD1.* Microb Biotechnol, 2018. **11**(5): p. 943-951.

14. Duetz, W.A. and B. Witholt, *Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions.* Biochemical Engineering Journal, 2004. **17**(3): p. 181-185.

15. Walling, L., et al., *Mixing in 384-well plates: issues, measurements, and solutions.* Assay Drug Dev Technol, 2007. **5**(2): p. 265-75.

16. Regnault, C., D.S. Dheeman, and A. Hochstetter, *Microfluidic Devices for Drug Assays.* High Throughput, 2018. **7**(2).

17. Che, P., L. Wang, and Q. Li, *The development, optimization and validation of an assay for high throughput antiviral drug screening against Dengue virus.* Int J Clin Exp Med, 2009. **2**(4): p. 363-73.

18. Agarwal, R.K., et al., *Optimization of microtitre plate assay for the testing of biofilm formation ability in different Salmonella serotypes.* International Food Research Journal, 2011. **18**(4): p. 1493-1498.

19. Gomez-Lopez, A., et al., *Analysis of the influence of Tween concentration, inoculum size, assay medium, and reading time on susceptibility testing of Aspergillus spp.* J Clin Microbiol, 2005. **43**(3): p. 1251-5.

20. Stevenson, K., et al., *General calibration of microbial growth in microplate readers.* Sci Rep, 2016. **6**: p. 38828.

21. Riss, T.L., et al., *Cell Viability Assays*, in *Assay Guidance Manual*, G.S. Sittampalam, et al., Editors. 2004: Bethesda (MD).

22. Osaka, I. and P.S. Hefty, *Simple resazurin-based microplate assay for measuring Chlamydia infections.* Antimicrob Agents Chemother, 2013. **57**(6): p. 2838-40.

23. Neufeld, B.H., et al., *Small Molecule Interferences in Resazurin and MTT-Based Metabolic Assays in the Absence of Cells.* Anal Chem, 2018. **90**(11): p. 6867-6876.

24. Chen, J.L., T.W.J. Steele, and D.C. Stuckey, *Metabolic reduction of resazurin; location within the cell for cytotoxicity assays.* Biotechnol Bioeng, 2018. **115**(2): p. 351-358.

25. Kim, H.J. and S. Jang, *Optimization of a resazurin-based microplate assay for large-scale compound screenings against Klebsiella pneumoniae.* 3 Biotech, 2018. **8**(1): p. 3.

26. O'Brien, J., et al., *Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity.* Eur J Biochem, 2000. **267**(17): p. 5421-6.

27. Uzarski, J.S., et al., *Essential design considerations for the resazurin reduction assay to noninvasively quantify cell expansion within perfused extracellular matrix scaffolds.* Biomaterials, 2017. **129**: p. 163-175.

28. Liu, Z., N. Banaei, and K. Ren, *Microfluidics for Combating Antimicrobial Resistance.* Trends Biotechnol, 2017. **35**(12): p. 1129-1139.

29. Behera, B., et al., *Emerging technologies for antibiotic susceptibility testing.* Biosensors and Bioelectronics, 2019. **142**: p. 111552.

30. Kim, K.P., et al., *In situ monitoring of antibiotic susceptibility of bacterial biofilms in a microfluidic device.* Lab Chip, 2010. **10**(23): p. 3296-9.

31. Duncombe, T.A., A.M. Tentori, and A.E. Herr, *Microfluidics: reframing biological enquiry.* Nat Rev Mol Cell Biol, 2015. **16**(9): p. 554-67.

32. Sun, X., et al., *A microfluidic platform for profiling biomechanical properties of bacteria.* Lab Chip, 2014. **14**(14): p. 2491-8.

33. Uddin, R., et al., *New Evidence for the Mechanism of Action of a Type-2 Diabetes Drug Using a Magnetic Bead-Based Automated Biosensing Platform.* ACS Sens, 2017. **2**(9): p. 1329-1336.

34. Golchin, S.A., et al., *A microfluidic system for long-term time-lapse microscopy studies of mycobacteria.* Tuberculosis (Edinb), 2012. **92**(6): p. 489-96.

35. Vukomanovic, M. and E. Torrents, *High time resolution and high signal-to-noise monitoring of the bacterial growth kinetics in the presence of plasmonic nanoparticles.* J Nanobiotechnology, 2019. **17**(1): p. 21.

36. Sarker, S.D., L. Nahar, and Y. Kumarasamy, *Microtitre plate-based antibacterial assay incorporating resazurin as an indicator of cell growth, and its application in the in vitro antibacterial screening of phytochemicals.* Methods, 2007. **42**(4): p. 321-4.

37. Teh, C.H., et al., *Determination of antibacterial activity and minimum inhibitory concentration of larval extract of fly via resazurin-based turbidometric assay.* BMC Microbiol, 2017. **17**(1): p. 36.

38. Fowler, P.W., et al., *Automated detection of bacterial growth on 96-well plates for high-throughput drug susceptibility testing of Mycobacterium tuberculosis.* Microbiology, 2018. **164**(12): p. 1522-1530.

39. Parsley, N.C., et al., *PepSAVI-MS reveals anticancer and antifungal cycloviolacins in Viola odorata.* Phytochemistry, 2018. **152**: p. 61-70.

40. Jones, E., S. Michael, and G.S. Sittampalam, *Basics of Assay Equipment and Instrumentation for High Throughput Screening*, in *Assay Guidance Manual*, G.S. Sittampalam, et al., Editors. 2004: Bethesda (MD).

41. Stromstedt, A.A., J. Felth, and L. Bohlin, *Bioassays in natural product research - strategies and methods in the search for anti-inflammatory and antimicrobial activity.* Phytochem Anal, 2014. **25**(1): p. 13-28.

42. Becker, H. and L.E. Locascio, *Polymer microfluidic devices.* Talanta, 2002. **56**(2): p. 267-87.

43. Horinek, D., et al., *Peptide adsorption on a hydrophobic surface results from an interplay of solvation, surface, and intrapeptide forces.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 2842-7.

44. Goebel-Stengel, M., et al., *The importance of using the optimal plasticware and glassware in studies involving peptides.* Anal Biochem, 2011. **414**(1): p. 38-46.

45. Chico, D.E., R.L. Given, and B.T. Miller, *Binding of cationic cell-permeable peptides to plastic and glass.* Peptides, 2003. **24**(1): p. 3-9.

46. Rosengren, K.J., et al., *The self-association of the cyclotide kalata B2 in solution is guided by hydrophobic interactions.* Biopolymers, 2013. **100**(5): p. 453-60.

47. Ballet, T., et al., *Protein conformational changes induced by adsorption onto material surfaces: an important issue for biomedical applications of material science.* Bulletin of the Polish Academy of Sciences: Technical Sciences, 2010. **58**(2).

48. Roach, P., D. Farrar, and C.C. Perry, *Interpretation of protein adsorption: surface-induced conformational changes.* J Am Chem Soc, 2005. **127**(22): p. 8168-73.

49. Gupta, S., N. Laskar, and D.E. Kadouri, *Evaluating the Effect of Oxygen Concentrations on Antibiotic Sensitivity, Growth, and Biofilm Formation of Human Pathogens.* Microbiol Insights, 2016. **9**: p. 37-46.

50. Ochs, C.J., et al., *Oxygen levels in thermoplastic microfluidic devices during cell culture.* Lab Chip, 2014. **14**(3): p. 459-62.

51. Petronilho, S., et al., *In vitro and in vivo studies of natural products: A challenge for their valuation. The case study of chamomile (Matricaria recutita L.).* Industrial Crops and Products, 2012. **40**: p. 1-12.

CHAPTER 7: Future Work and Conclusions

Strategies to increase AMP identification, prioritize putative bioactive species, and characterize targets are available to further streamline the process of rapid AMP discovery. Exploration of alternative source materials, extraction methods, peptide library preparation, and bioactivity assays may increase AMP hit rates. Orthogonal methods, e.g. mass shift analysis and transcriptomics, can be leveraged to aid in rapid AMP identification. Biological characterization of novel AMPs can be achieved via minimum bactericidal concentration and mechanism of action assays.

7.1 Increasing Hit Rates in Bioactivity Screening

Over 50 plant species from diverse botanical families, chosen largely for traditional therapeutic use, have been extracted and screened for biological activity over the course of my dissertation research (Table 7.1). Few extracts have demonstrated potent and/or reproducible bioactivity against selected microbial pathogens; this low 'hit rate' necessitates ongoing screening efforts in the Hicks laboratory. A majority of the material analyzed for activity was derived from plant aerial tissues and seeds, as these are often the most abundant by mass and/or facile to harvest, and extracted via an acidic aqueous method as described[1] (Chapters 3-7). However, deviations to include alternative source materials and extraction types may serve to access a greater chemical space towards the identification of novel bioactive peptides.

7.1.1 Plant growth conditions

Plant growth conditions may be altered to effect induction of novel AMPs in botanical species. Abiotic (salinity, pH) and biotic (bacterial, fungal pathogens) factors influence plant vitality, where plant stress responses occur through overlapping signaling pathways.[2] Plant

survival relies on a chemically diverse complement of constitutively expressed and stress-induced AMPs.[3] Snakin-2 is an antibacterial and antifungal AMP constitutively expressed in potato tubers; however, snakin-2 is upregulated by wounding, abscisic acid treatment, and fungal infection.[3] In a laboratory greenhouse, biotic and abiotic stresses can be introduced to upregulate a novel and otherwise undetectable suite of botanical AMPs and captured during subsequent harvest. However, specialized facilities necessary to perform pathogen-induced experiments are needed to prevent pathogen spread beyond experimental bounds. Alternatively, abiotic stress can be applied to induce a pathway response similar to that of biotic stress, although experimental conditions are often difficult to control (e.g. maintaining desired salinity in soil may necessitate a hydroponics set-up). Finally, a method to increase AMP diversity in source material is the wild harvest of botanical materials. Plants gathered from nature have adapted to and survived pathogens and environmental stresses; however, material availability is often limited, species identifications can be challenging, and exact growth conditions are unknown – significantly impacting experimental reproducibility. Despite these limitations, abiotic stress (salt) and wounding experiments and wild harvest of source materials have been pursued in the Hicks lab to enhance AMP expression/identification.

7.1.2 Analysis of all plant tissues

Analysis of all plant tissues, including seeds, flowers, bark, fruits, and roots, has resulted in the identification of numerous biologically active peptides.[4-14] Ideally, all available tissues for a botanical species would be screened for bioactivity and profiled via LC-MS/MS, as AMPs may be differentially expressed among different tissue types[8, 15, 16] and may be related to the pathogen stress associated with that tissue. For example, proteinase inhibitors are structurally-diverse and highly bioactive CRPs, and include Kunitz and Bowman–Birk type inhibitors[17-20] and small

molecular weight knottin-type trypsin inhibitors expressed abundantly in the seeds of Cucurbitaceae (gourd) and Fabaceae (pea) species.[21] Trypsin/chymotrypsin inhibition prevents the proteolytic digestion and subsequent absorption of proteins (e.g. when consumed by a predator),[22] where decreased nutritional availability, "antinutrition," serves to inhibit predator (e.g. seed-feeding insects) growth and may thwart future consumption.[23] Cyclotide expression has also been found to be tissue specific; an analysis of violet *Viola hederacea* revealed a set of late reversed-phase eluting (hydrophobic) cyclotide species exclusively produced in tissues directly in contact with soil (roots, runners, and bulbs).[24]

Exploration of cyclotide tissue-specificity is being pursued in the Hicks laboratory via mass shift analysis. LC-MS analysis of *V. inconspicua* root and aerial tissues demonstrated highly variable total ion chromatograms (TICs), and thus likely differ significantly in peptidyl constituents (Figure 7.1). Mass shift analysis revealed a suite of putative cyclotide species differentially expressed in root and aerial tissues; though some cyclotide species were expressed in both materials, most were seen in only one (Table 7.2). Early reversed-phase LC retention times suggest putative aerial cyclotide masses are collectively more hydrophilic than root cyclotides. Interestingly, type II cyclotides (Chapter 4) are expressed only in the aerial tissue, indicating a possible biological significance of this motif. Sequence elucidation of tissue-specific mass species and assessment of bioactivity may reveal relationships among sequence types and anticipated pathogen/environment pressure; as such, molecular and biological characterization efforts are currently underway.

7.1.3 Alternative extractions

Alternative extractions may capture additional AMP species with bioactivity potential. Although aqueous extractions are often employed towards the identification of novel botanical peptides,[25, 26] organic extractions, particularly dichloromethane, acetonitrile, ethanol, and methanol, are commonplace in the extraction of bioactive AMPs.[27-31] Botanical species extracted in water and devoid of bioactivity (Table 7.1) may be re-extracted in organic solvents or 50/50 water/organic solvent. An initial comparison should be pursued using cyclotide-dense Viola material, which is often abundant in mass and contains large quantities of diverse cyclotide species. Organic solvents have been used to extract cyclotide constituents[31]; performing simultaneous extractions of the same Viola material with both aqueous and organic solvents may reveal differentially/similarly extracted cyclotide constituents when subsequently analyzed via LC-MS/MS, and may inform the value of future multi-solvent extractions.

7.1.4 Expansion of bioactivity screens

While > 50 plants have been screened for bioactivity, only limited pathogen species have been assayed against. Botanical peptide libraries are initially assayed against the control strain *E. coli* ATCC 25922, a Gram-negative bacteria standard for antimicrobial susceptibility testing; if promising activity is identified, peptide libraries are subsequently assayed against multi-drug resistant human pathogens (*ESKAPE*). Activity assays of few plant fungal pathogens (*Ustilago maydis*[32] and *Fusarium graminearum* – Chapter 3) have been pursued. Extension of bioactivity assessment to additional plant specific pathogens, which may be the direct target of plant AMPs, e.g. *Xylella fastidiosa*,[33-35] *Pseudomonas syringae*,[35] and Fusarium spp.,[36, 37] may reveal potent bioactivities in libraries bereft of bioactivity against human pathogens.

7.1.5 Synergy

Synergy is an important mechanism by which AMPs may achieve bioactivity[38-40]; while fractionation of peptide libraries is essential for PepSAVI-MS statistical analyses, this process may separate peptide interactions necessary for antimicrobial activities. Thus, peptide libraries are now assessed for bioactivity in the form of library fractions and as "superfractions," the latter of which is a single sample consisting of all combined library fractions (Figure 7.2). Active superfractions and inactive peptide libraries indicate possible synergy among extract constituents, and suggest an alternative fractionation approach is needed, either via orthogonal chromatography or through fraction combinations. For example, library fractions can be systematically combined until bioactivity is reestablished. Alternatively, large groups of peptide fractions can be assayed for activity, divided into smaller fraction groups, and reassayed until activity is lost (Figure 7.3). In either case, mass spectrometric analysis of fractions before/after activity is lost/gained may reveal the compounds necessary for activity. Interestingly, some superfractions have lacked bioactivity where peptide libraries generated from the same extract material exhibit a bioactive region, and may result from antagonistic interactions of peptide species normally spatially separated/compartmentalized in the plant cell.

7.1.6 Assay Optimization

Amphipathic AMPs interact with a variety of material surfaces,[41] and as such, materials used throughout extraction and sample preparation steps may affect downstream AMP availability. Plastics constituting dialysis membranes, conical vials and other plastic consumables for sample collection/concentration, and bioactivity assay plates may interact with and bind select extract constituents. The predictability of which extract constituent chemistries may be affected by a specific material is limited; multiple plastic types (or glass, when appropriate) may improve

bioactive peptide identification. For example, dramatic differences in activity profiles and even the identity of bioactive species may be observed in polystyrene, polypropylene, and polymethylmethacrylate bioactivity assay formats (Chapter 6). Ultimately, the most physiologically relevant material is unclear, and thus protocols with reduced sample preparation steps are preferred to lessen AMP losses through surface interactions.

7.2 Improving Molecular Identification Prioritization/Characterization

7.2.1 Orthogonal library preparation

Peptide libraries generated and analyzed via PepSAVI-MS[1] often contain >10,000 detected mass spectral features. Sample preparation steps are taken to reduce sample complexity prior to statistical analyses; however, abundant sample constituents hinder the rapid and accurate identification of novel bioactive mass species. Though requiring at least 2X the amount of starting material, orthogonal chromatographies may be employed to reduce the number of putative bioactive species in a dataset. Fractionation of the same natural product extract with ion-exchange, reversed-phase, or other chromatographies can produce peptide libraries with differentially separated extract constituents. Subsequent bioactivity and mass spectral analyses would ideally reveal regions of active library fractions containing unique suites of mass species; overlap among the mass identities in independent chromatographic bioactive regions reduces the number of putative bioactive compounds and enables the elimination of masses not present in all active regions.

7.2.2 Conserved sequence/structural features

Conserved sequence/structural features (e.g. cysteine-rich) can be leveraged to aid the rapid and accurate identification of novel AMPs (Chapters 2-6); additional characteristics typical of

bioactive peptides can be targeted with chemical derivatization strategies coupled with mass spectrometry.

Mass species cyclized via one or both termini may be detected through labeling approaches, where the lack of an N- and/or C-terminus prevents the addition of derivatization agents targeting primary amines or carboxyl moieties, respectively. In contrast to cysteine derivatization, this method would result in the lack of an expected mass shift upon derivatization; as such, this method is less robust than strategies where mass shifts are expected to be observed. For example, derivatization with *O*-phthalaldehyde/N-acetylcysteine (OPA/NAC) labels primary amines[42] and carboxyl groups can be converted to tertiary/quaternary amines through a variety of reagents,[43] where the absence of a mass shift would indicate a blocked N-/C-terminus, respectively, and potential cyclization. However, three dimensional accessibility and other modifications (e.g. N-terminal pyroglutamic acid) can inhibit termini labeling and generate false positives. Additionally, the derivatization of amino acid side chains with primary amines (the ε-amine of Lys) or carboxyl groups (Asp, Glu) can convolute general analyses; though it may provide more specific information. Möbius cyclotides, for example, contain a cyclic backbone, thus no N-terminus, and lack basic residues (e.g. Lys), where cyclotides in the bracelet cyclotide subfamily, though cyclic, contain abundant Lys and Arg residues. As such, cursory attempts at identifying novel Möbius mass species may be achieved through OPA/NAC derivatization of cyclotide-containing extracts and the subsequent identification of non-shifting masses.

7.2.3 Transcriptomics

Transcriptomics can be coupled with mass spectrometry/bioactivity-guided AMP identification. Over 1000 transcriptomes from the Viridiplantae (green plants) are publicly available (One Thousand Plant Transcriptomes Initiative, onekp.com).[44] Transcriptomes may

otherwise be generated from a small amount of plant material (50-100 mg), however, require significant initial costs (~$2,000) and *de novo* transcriptome assembly may be difficult to implement in species with no genomic information. Transcriptomes can be processed to predict a proteomic profile for a botanical species, where predicted masses can be compared with mass spectrometry-derived masses in bioactive library regions. Transcriptome-predicted peptides matched via mass can be easily fragmented *in silico* (e.g. http://prospector.ucsf.edu), and predicted fragments matched to experimental MS/MS data. Additionally, same mass species (Ile/Leu) can be quickly discerned using transcriptomics data, without the need for additional multistage (MS^3) experiments.

7.3 Biological Characterization

Peptides confirmed to possess antimicrobial activities must be biochemically characterized. Bioactive peptides are often characterized by a minimum inhibitory concentration (MIC) against a pathogen; however, MIC cannot distinguish between bacteriostatic or bactericidal activity. Minimum bactericidal concentration (MBC) activity assays can be performed to determine the number of viable cells after treatment with an AMP. Minimum bactericidal concentration values determined at different time points and with different AMP concentrations can be used to generate time-kill curves towards an estimate of a bacterial killing rate and killing concentration dependence.[45]

Understanding mechanism of action (MOA) may inform therapeutic treatment regimens and possible avenues of microbial resistance, in addition to providing sequence/structure-activity relationships that may guide future AMP design towards the modulation of specific biochemical targets. Established AMP MOAs are largely membrane-acting, thus methods to investigate MOAs often assess AMP effects on membranes. Cytoplasmic membrane disruption assays are

performed with a variety of cell- and nucleic acid-staining dyes. SYTOX Green, for example, is a cell-impermeable nucleic acid stain: fluorescence is only observed when cell membranes are compromised, allowing cellular entry and subsequent nucleic acid interaction by the SYTOX Green dye.[45, 46] Complementary cellular leakage assays identify "leaked" cellular components, such as ATP quantified by bioluminescent ATP-binding dyes,[45] or other intracellular molecules (DNA/RNA,[47] physiologically-relevant ions[48]). Alternatively, methods to characterize intracellular MOAs may assess changes in cellular gene expression via transcriptomic/proteomic analyses.[49]

In vitro antimicrobial activity assays (e.g. 96-well plate format, in rich bacteriological media) likely do not recapitulate the biochemical conditions *in vivo*: characterization of AMP stability and activity in appropriate environments are necessary to prioritize the most promising lead compounds. As such, assays to characterize serum, pH, thermal, salt, and enzymatic stability can inform potential routes of drug delivery (e.g. intravenous versus oral delivery methods experience difference biophysical and biochemical obstacles).[50]

7.4 Antimicrobial Peptide Production

Activity confirmation and bioactivity/MOA characterization of novel AMPs require significant quantities of purified AMP. Innate low abundance and variable AMP expression often preclude large-scale AMP isolation from a source organism. Solid-phase peptide synthesis (SPPS) may access large amounts of material, however, post-translational modifications may be difficult to reproduce. Additionally, higher molecular weight AMPs (> 35 residues) may present increased SPPS production costs and the numerous chemical protection-deprotection steps associated with larger AMP species increase the generation of side-products.[51, 52] Alternatively, heterologous bacterial,[52-55] fungal,[55-59] plant,[55, 60-62] and algal[63-65] expression systems have been

utilized for transgenic peptide production and may offer lower manufacturing costs, easy affinity purification, and the ability to co-express enzymes necessary to tailor a final AMP product. Fusion tags, such as glutathione S-transferase (GST) or maltose-binding protein (MBP), can be incorporated into the translated AMP product to reduce host toxicity, evade cellular proteolytic degradation, and aid in subsequent purification.[52, 66]

7.5 Conclusions

As multidrug-resistant microbes are becoming alarmingly prevalent in the modern world, new methods are needed to generate novel antimicrobial therapeutics. Botanical natural product peptides are a vast and underexplored drug class with novel chemistries and potent bioactivity potential. Extract complexity, inherent variability in AMP sequence, post-translational modifications, and secondary/tertiary structure, and low abundance/inconsistent AMP expression challenge the identification and isolation of new AMP species. Mass spectrometry-based methods (e.g. MS^n) coupled with ultra-high performance liquid chromatography (UHPLC) enable the exploration of complex extracts and the identification/characterization of low abundance peptidyl species. Despite significant biochemical diversity, the identification of conserved AMP features coupled with PepSAVI-MS can be leveraged towards the MS-based identification of novel bioactive sequences (Chapters 2-5). Tandem mass spectrometry enables the sequence characterization of novel AMP species without peptide isolation (Chapter 4) as well and can rapidly provide molecular information prior to full sequence characterization (Chapter 5). Elements of the discovery process can be modified to enhance AMP identification and characterization, including alternative botanical tissue sources (e.g. aerial, roots [Chapter 7]), pathogens (e.g. bacteria, fungi [Chapter 3]), mass spectral fragmentation techniques (e.g. gas-phase linearization[67] and ultraviolet photodissociation [Chapter 3]), bioactivity assay formats and

cell viability readouts (e.g. 96-well plate versus microfluidic format, absorbance versus fluorescent detection [Chapter 6]). This work expands the therapeutic potential of natural product bioactive peptides and highlights the application of robust analytical tools to explore the natural complexity presented by plants.

7.6 Tables

Table 7.1. Botanical species extracted and assayed for activity. RP = Reversed-phase LC. SCX = Strong cation exchange. *SM = small molecules likely responsible for activity. *NR = Not reproducible.

Latin name	Common Name	Abbreviation	Family	Tissue type	Location	Known bioactive compounds	Activity
Amaranthus hypochondriacus	Prince's feather	AHs	Amaranthaceae	Seed	Strictly Medicinal	Flavonoids, Phenolic acids, Anthocyanins, Tannis, Phytosterols	SF/RP lib: *E. coli,* none
Anchusa azurea	Bugloss	AAz	Boraginaceae	Aerial	UNC Greenhouse	Flavonoids, Phenolic acids	SCX lib: *K. pneumoniae* activity *SM
Anchusa azurea	Bugloss	AAzs	Boraginaceae	Seed	Strictly Medicinal	Flavonoids, Phenolic	SF: *E. coli,* none
Anethum graveolens	Dill	AGs	Apiaceae	Seed		Tannins, Terpenoids, Alkaloids, Flavonoids	SF/RP lib: *E. coli* *NR
Astragalus membranaceous	Huangqi	AM	Fabaceae	Aerial	UNC Greenhouse	Astragalus	SCX lib: *E. coli,* active *NR
Atropa belladonna	Belladonna	AB	Solanaceae	Aerial	UNC Greenhouse	Tropane alkaloids	SCX lib: *E. coli* , active *SM
Azadirachta indica	Neem	AI	Meliaceae	Aerial, dried	Ademola Oyagemi	Triterpenoid	SF: *E. coli,* none
Brassica napus	Canola	BN	Brassicaceae	Aerial	UNC Greenhouse	Flavonoids, Carotenoids	SF: *E. coli,* none
Canabis sativa	Hemp	CS	Cannabaceae	Aerial	UNC Greenhouse	Cannabinoids	SCX lib: *E. coli,* active *NR
Capsicum annuum	Pepper	CA	Solanaceae	Aerial	UNC Greenhouse	Carotenoids, Tocopherols, Capsaicinoids	SF: *E. coli,* >80% activity *SM
Cucurbita foetidissima	Buffalo gourd	CFs	Cucurbitaceae	Seed	Strictly Medicinal	Cucurbitacins, Foetidissimosides	SF: *E. coli,* none
Cucurbita pepo	Zucchini	CPf	Cucurbitaceae	Flower	UNC Greenhouse	Carotenoids, Phenolic acids, Flavonols	SF: *E. coli,* 100% active *SM
Datura stramonium	Angel's trumpet	DS	Solanaceae	Aerial	UNC Greenhouse	Alkaloids, Tannins	SCX lib: E. coli, active *SM
Daucus carota	Wild carrot	DC	Apiaceae	Inflorescence	Durham	Phenolics, Catotenoids	SF: *E. coli,* none
Dictamnus albus var. purpureus	Burning bush	DA	Rutaceae	Seed	Strictly Medicinal	-	SF: *E. coli,* none
Digitalis lanata	Foxglove	DL	Plantaginaceae	Aerial	UNC Greenhouse	Digitalins	SCX lib: *A. baumanii, K. pneumoniae* activity *SM
Digitaria sp.	Crabgrass	DSp	Poaceae	Inflorescence	Durham	-	SF: *E. coli,* none
Echinacea purpurea	Purple coneflower	EPu	Asteraceae	Aerial	UNC Greenhouse	Caffeic acid derivatives, Alkylamides	SF: *E. coli,* none
Echinacea purpurea	Purple coneflower	EPus	Asteraceae	Seed	Strictly Medicinal	Caffeic acid derivatives, Alkylamides	SF: *E. coli* , none; RP lib: *E. coli* , active *SM
Glechoma hederacea	Alehoof	GH	Lamiaceae	Aerial	UNC Greenhouse	Phenolics, Flavonoids	SF: *E. coli,* none
Hyoscyamus niger	Henbane	HN	Solanaceae	Aerial	UNC Greenhouse	Alkaloids, Lignans, Flavanoids	SF: *E. coli,* >40% activity, *S. aureus,* 50% activity *NR
Juniper sp.	-	Jb	Cupressaceae	Berry	Chapel Hill	Flavonoids	SF/RP lib: *E. coli,* none
Lagenaria siceraria	Bushel basket gourd	LSs	Cucurbitaceae	Seed	Stictly Medicinal	Alkaloids, Phenols, Tannins, Flavonoids	SF: *E. coli,* none
Larrea tridentata	Creosote bush	LTs	Zygophyllaceae	Seed	Strictly Medicinal	Phytoestrogens, Phenols, Flavonoids	SF: *E. coli,* none
Momordica charantia	Bitter melon	MCh	Cucurbitaceae	Aerial	UNC Greenhouse	Tryspin inihibtors	SF: *E. coli,* none
Nepenthes ventricoa	Pitcher plant	NV	Nepentaceae	Pitcher	UNC Greenhouse	-	SF: *E. coli,* none
Nepetia cataria	Catnip	NC	Lamiaceae	Aerial	UNC Greenhouse	Nepetalactone	SF: *E. coli,* none
Ocimum sanctum	Tusli/Holy basil	OS	Lamiaceae	Aerial	UNC Greenhouse	-	SF: *E. coli,* none
Packera aurea	Golden ragwort	PAu	Asteraceae	Aerial	UNC arboretum	Alkaloids	SF: *E. coli,* none
Perilla frutescens	Perilla	PF	Lamiaceae	Aerial	Mountain Gardens	-	SF: *E. coli,* none
Phytolacca americana	Pokeberry	PA	Phytolaccaceae	Aerial	UNC Greenhouse	Alkaloids	SF: *E. coli,* >20% activity
Planktothrix agardhii	Cyanobacteria	PA	Microcoleaceae	Cell	Wendy Strangman	Microcystins	NA
Planktothrix rubescens	Cyanobacteria	PR	Microcoleaceae	Cell	Wendy Strangman	Microcystins	NA
Plantago major	Broadleaf plantain	PM	Plantaginaceae	Inflorescence	Durham	Aucubin	SF: *E. coli,* none
Portulaca oleracea	Purslane	PO	Portulacaeae	Aerial	UNC Greenhouse	Portulene	SF: *E. coli,* none
Psychotria zombamontana	Red bird berry	PZ	Rubiaceae	Aerial, dried	Abimbola	-	SF/SCX lib: *E. coli,* none
*Pyschotria capensis *Organic*	Bird berry	PC	Rubiaceae	Aerial, dried	Abimbola	-	NA
Pyschotria capensis	Bird berry	PC	Rubiaceae	Aerial, dried	Abimbola	-	SF/SCX lib: *E. coli,* none
Silybum marianum	Milk thistle	SM	Asteraceae	Aerial	UNC Greenhouse	Silymarin	SCX lib: *E. coli, K. pneumoniae* activity *NR
Silybum marianum	Milk thistle	SMs	Asteraceae	Seed	Strictly Medicinal	Silymarin	SF/RP lib: *E. coli,* none
Trichosanthes kirilowii	Guālóu	TKs	Cucurbitaceae	Seed	Strictly Medicinal	Trichosanthin	SF: *E. coli* , none
Verbascus thapsus	Mullein	VT	Scrophulariaceae	Aerial	UNC Greenhouse	Rotenone	SF: *E. coli,* none
Vicia sativa	Vetch	VS	Fabaceae	Aerial	American Meadows Seeds	-	SF: *E. coli,* none
Viola communis	Violet	VC	Violaceae	Aerial	Durham	Cyclotides	SF: *E. coli,* 95% activity
Viola inconspicua	Violet	VI	Violaceae	Aerial	UNC Greenhouse	Cyclotides	RP lib: *E. coli, K. pneumoniae* activity
Viola inconspicua	Violet	VI	Violaceae	Roots	UNC Greenhouse	Cyclotides	RP lib: *Verticillium* activity
Viola odorata	Sweet violet	VO	Violaceae	Aerial	UNC Greenhouse	Cyclotides	SCX lib: cancer cell activity
Vitis rotundifolia	Muscadine	VR	Vitaceae	Aerial	UNC Greenhouse	Tannins, Anthocyanins Flavonoids	SF: *E. coli,* none
Withania somnifera	Ashwagandha	WSs	Solanaceae	Seed	Strictly Medicinal	Withanolides	SF: *E. coli,* none
Withania somnifera	Ashwagandha	WS	Solanaceae	Aerial	UNC Greenhouse	Withanolides	SCX library: *F. graminearum* activity
Zingiber officinale	Ginger	ZO	Zingiberaceae	Rhizome	Durham (Commercial)	Gingerol	SF: *E. coli,* none; SCX library: *F. graminearum* activity

Table 7.2. Mass shift analysis of *V. inconspicua* aerial and root material revealed several putative, tissue-specific cyclotide species (masses > 2500 Da).

Aerial		Root	
Mass (Da)	RA Mass (Da)	Mass (Da)	RA Mass (Da)
2577.35	2925.47	3025.22	3373.34
3069.44	3417.61	3030.24	3378.39
3140.46	3488.64	3055.28	3403.38
3151.70	3499.86	3057.22	3405.32
3225.43	3573.62	3071.28	3419.40
3226.48	3574.67	3085.30	3433.42
3245.45	3593.64	3091.24	3439.35
3248.93	3597.11	3095.33	3443.43
3293.50	3641.69	3117.27	3465.40
3294.51	3642.69	3118.25	3466.36
		3128.28	3476.39
		3129.28	3477.41
		3141.25	3489.41
		3146.23	3494.34
		3147.23	3495.42
		3160.24	3508.42
		3162.18	3510.35
		3170.31	3518.46
		3211.37	3559.47
		3226.38	3574.50
		3240.38	3588.50
		3254.27	3602.44
		3451.37	3799.53

7.7 Figures

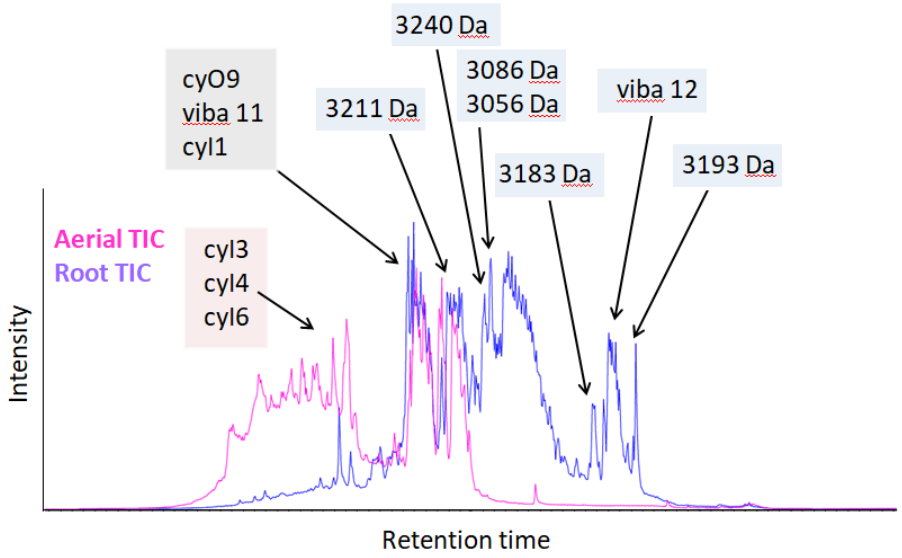


Figure 7.1. Overlaid *V. inconspicua* aerial (pink) and root (purple) total ion chromatograms (TICs). Early reversed-phase LC retention times suggest putative aerial cyclotide masses (pink box) are collectively more hydrophilic than root cyclotide masses (blue box); some masses were shared between tissues (grey box).

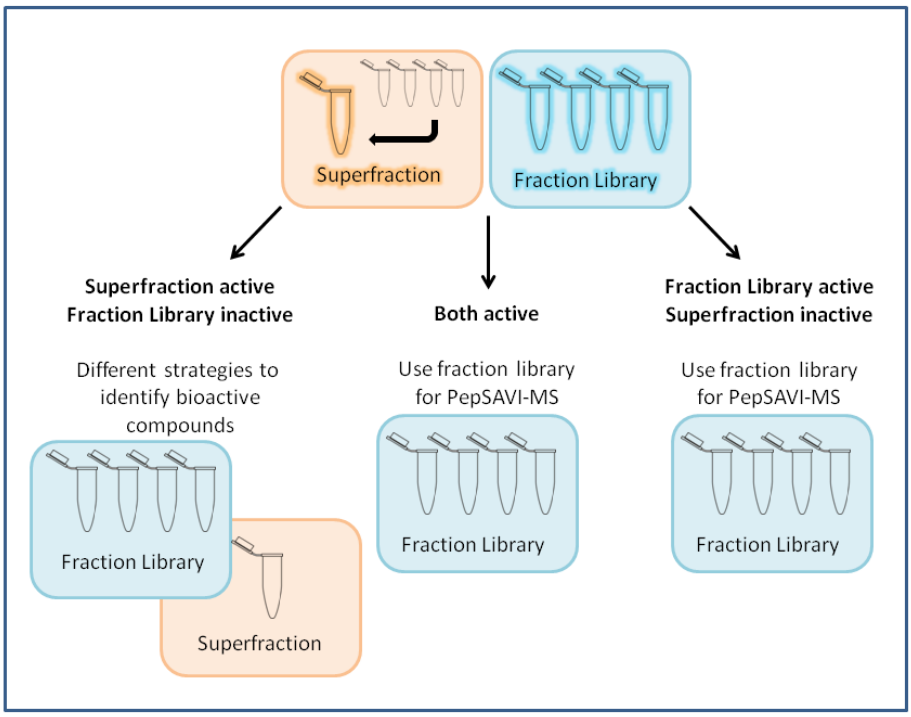


Figure 7.2. Botanical fraction libraries and "superfractions" are assayed against pathogens; strategies to identify bioactive species vary based on activity outcome (fraction library, superfraction, or both active).

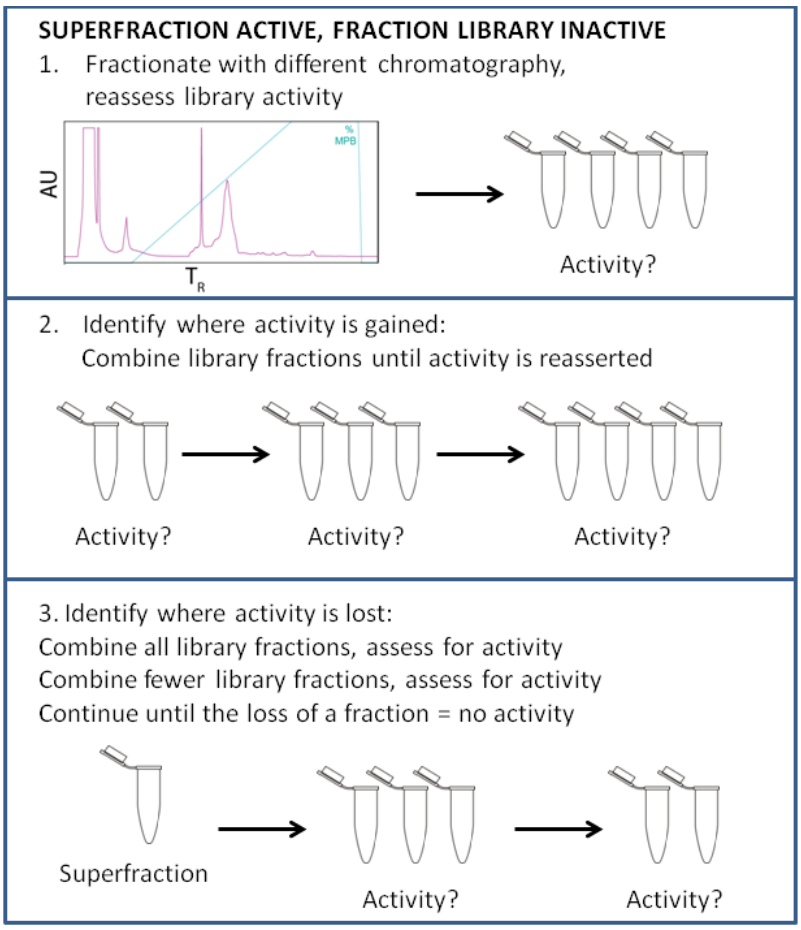


Figure 7.3. Superfraction activity and peptide library inactivity may indicate synergism among extract constituents: alternative fractionation options can aid in the identification of bioactive compounds.

REFERENCES

1. Kirkpatrick, C.L., et al., *The "PepSAVI-MS" Pipeline for Natural Product Bioactive Peptide Discovery.* Anal Chem, 2017. **89**(2): p. 1194-1201.

2. Chojak-Kozniewska, J., E. Kuzniak, and J. Zimny, *The Effects of Combined Abiotic and Pathogen Stress in Plants: Insights From Salinity and Pseudomonas syringae pv lachrymans Interaction in Cucumber.* Front Plant Sci, 2018. **9**: p. 1691.

3. Berrocal-Lobo, M., et al., *Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection.* Plant Physiol, 2002. **128**(3): p. 951-61.

4. Van den Bergh, K.P., et al., *Ee-CBP, a hevein-type antimicrobial peptide from bark of the spindle tree (Euonymus europaeus L.).* Meded Rijksuniv Gent Fak Landbouwkd Toegep Biol Wet, 2002. **67**(2): p. 327-31.

5. Dos Santos, L.A., et al., *Purification and characterization of peptides from Capsicum annuum fruits which are alpha-amylase inhibitors and exhibit high antimicrobial activity against fungi of agronomic importance.* Protein Expr Purif, 2017. **132**: p. 97-107.

6. Meneguetti, B.T., et al., *Antimicrobial Peptides from Fruits and Their Potential Use as Biotechnological Tools-A Review and Outlook.* Front Microbiol, 2016. **7**: p. 2136.

7. Alvarez, C.A., et al., *Identification of Peptides in Flowers of Sambucus nigra with Antimicrobial Activity against Aquaculture Pathogens.* Molecules, 2018. **23**(5).

8. Tavares, L.S., et al., *Biotechnological potential of antimicrobial peptides from flowers.* Peptides, 2008. **29**(10): p. 1842-51.

9. Astafieva, A.A., et al., *Discovery of novel antimicrobial peptides with unusual cysteine motifs in dandelion Taraxacum officinale Wigg. flowers.* Peptides, 2012. **36**(2): p. 266-71.

10. Ogbole, O.O., et al., *Evaluation of peptide-rich root extracts of Calliandria portoriscensis (Jacq.) Benth (Mimosaceae) for in vitro antimicrobial activity and brine shrimp lethality.* BMC Complement Med Ther, 2020. **20**(1): p. 30.

11. Gao, X., et al., *Identification and antimicrobial activity evaluation of three peptides from laba garlic and the related mechanism.* Food Funct, 2019. **10**(8): p. 4486-4496.

12. Van den Bergh, K.P., et al., *Five disulfide bridges stabilize a hevein-type antimicrobial peptide from the bark of spindle tree (Euonymus europaeus L.).* FEBS Lett, 2002. **530**(1-3): p. 181-5.

13. Vieira, M.E., et al., *Isolation, characterization and mechanism of action of an antimicrobial peptide from Lecythis pisonis seeds with inhibitory activity against Candida albicans.* Acta Biochim Biophys Sin (Shanghai), 2015. **47**(9): p. 716-29.

14. Al Akeel, R., et al., *Purification and MIC analysis of antimicrobial proteins from Cucumis sativus L. seeds.* BMC Complement Altern Med, 2018. **18**(1): p. 121.

15. Tam, J.P., et al., *Antimicrobial Peptides from Plants.* Pharmaceuticals (Basel), 2015. **8**(4): p. 711-57.

16. Campos, M.L., et al., *The role of antimicrobial peptides in plant immunity.* J Exp Bot, 2018. **69**(21): p. 4997-5011.

17. Xu, Y., et al., *Preparation and Irreversible Inhibition Mechanism Insight into a Recombinant Kunitz Trypsin Inhibitor from Glycine max L. Seeds.* Appl Biochem Biotechnol, 2020.

18. Shamsi, T.N., et al., *Trypsin Inhibitors from Cajanus cajan and Phaseolus limensis Possess Antioxidant, Anti-Inflammatory, and Antibacterial Activity.* J Diet Suppl, 2018. **15**(6): p. 939-950.

19. Ruan, J.J., et al., *Coix lacryma-jobi chymotrypsin inhibitor displays antifungal activity.* Pestic Biochem Physiol, 2019. **160**: p. 49-57.

20. Major, I.T. and C.P. Constabel, *Functional analysis of the Kunitz trypsin inhibitor family in poplar reveals biochemical diversity and multiplicity in defense against herbivores.* Plant Physiol, 2008. **146**(3): p. 888-903.

21. Kennedy, A.R., *The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent.* Am J Clin Nutr, 1998. **68**(6 Suppl): p. 1406S-1412S.

22. Aviles-Gaxiola, S., C. Chuck-Hernandez, and S.O. Serna Saldivar, *Inactivation Methods of Trypsin Inhibitor in Legumes: A Review.* J Food Sci, 2018. **83**(1): p. 17-29.

23. Sowbaghya, A.Y., et al., *Proteins and trypsin inhibitors in seeds of various plants.* Indian Journal of Entomology, 2019. **81**(1): p. 177.

24. Trabi, M. and D.J. Craik, *Tissue-specific expression of head-to-tail cyclized miniproteins in Violaceae and structure determination of the root cyclotide Viola hederacea root cyclotide1.* Plant Cell, 2004. **16**(8): p. 2204-16.

25. Zhang, Y. and K. Lewis, *Fabatins: new antimicrobial plant peptides.* FEMS Microbiol Lett, 1997. **149**(1): p. 59-64.

26. Aliahmadi, A., et al., *A simple method for primary screening of antibacterial peptides in plant seeds.* Iran J Microbiol, 2011. **3**(2): p. 104-8.

27. Wang, C.K., et al., *Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis.* J Nat Prod, 2008. **71**(1): p. 47-52.

28. Nostro, A., et al., *Extraction methods and bioautography for evaluation of medicinal plant antimicrobial activity.* Lett Appl Microbiol, 2000. **30**(5): p. 379-84.

29. Mostafa, A.A., et al., *Antimicrobial activity of some plant extracts against bacterial strains causing food poisoning diseases.* Saudi J Biol Sci, 2018. **25**(2): p. 361-366.

30. Hammami, R., et al., *Detection and extraction of anti-Listerial compounds from Calligonum comosum, a medicinal plant from arid regions of Tunisia.* Afr J Tradit Complement Altern Med, 2011. **8**(3): p. 322-7.

31. Ireland, D.C., M.L. Colgrave, and D.J. Craik, *A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability.* Biochem J, 2006. **400**(1): p. 1-12.

32. Kirkpatrick, C.L., et al., *Fungal Secretome Analysis via PepSAVI-MS: Identification of the Bioactive Peptide KP4 from Ustilago maydis.* J Am Soc Mass Spectrom, 2018.

33. Vanhove, M., et al., *Population structure and adaptation of a bacterial pathogen in California grapevines.* Environ Microbiol, 2020.

34. Purcell, A.H. and S.R. Saunders, *Fate of Pierce's Disease Strains of Xylella fastidiosa in Common Riparian Plants in California.* Plant Dis, 1999. **83**(9): p. 825-830.

35. Mansfield, J., et al., *Top 10 plant pathogenic bacteria in molecular plant pathology.* Mol Plant Pathol, 2012. **13**(6): p. 614-29.

36. Gordon, T.R., *Fusarium oxysporum and the Fusarium Wilt Syndrome.* Annu Rev Phytopathol, 2017. **55**: p. 23-39.

37. Munkvold, G.P., *Fusarium Species and Their Associated Mycotoxins.* Methods Mol Biol, 2017. **1542**: p. 51-106.

38. Hanson, M.A., et al., *Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach.* Elife, 2019. **8**.

39. Yan, H. and R.E. Hancock, *Synergistic interactions between mammalian antimicrobial defense peptides.* Antimicrob Agents Chemother, 2001. **45**(5): p. 1558-60.

40. Matsuzaki, K., et al., *Mechanism of synergism between antimicrobial peptides magainin 2 and PGLa.* Biochemistry, 1998. **37**(43): p. 15144-53.

41. Goebel-Stengel, M., et al., *The importance of using the optimal plasticware and glassware in studies involving peptides.* Anal Biochem, 2011. **414**(1): p. 38-46.

42. Bruckner, H., R. Wittner, and H. Godel, *Automated enantioseparation of amino acids by derivatization with o-phthaldialdehyde and n-acylated cysteines.* J Chromatogr, 1989. **476**: p. 73-82.

43. Frey, B.L., et al., *Chemical derivatization of peptide carboxyl groups for highly efficient electron transfer dissociation.* J Am Soc Mass Spectrom, 2013. **24**(11): p. 1710-21.

44. One Thousand Plant Transcriptomes, I., *One thousand plant transcriptomes and the phylogenomics of green plants.* Nature, 2019. **574**(7780): p. 679-685.

45. Raheem, N. and S.K. Straus, *Mechanisms of Action for Antimicrobial Peptides With Antibacterial and Antibiofilm Functions.* Front Microbiol, 2019. **10**: p. 2866.

46. Lebaron, P., P. Catala, and N. Parthuisot, *Effectiveness of SYTOX Green stain for bacterial viability assessment.* Appl Environ Microbiol, 1998. **64**(7): p. 2697-700.

47. Yasir, M., D. Dutta, and M.D.P. Willcox, *Mode of action of the antimicrobial peptide Mel4 is independent of Staphylococcus aureus cell membrane permeability.* PLoS One, 2019. **14**(7): p. e0215703.

48. Hou, X., et al., *Mechanism of antimicrobial peptide NP-6 from Sichuan pepper seeds against E. coli and effects of different environmental factors on its activity.* Appl Microbiol Biotechnol, 2019. **103**(16): p. 6593-6604.

49. Tomasinsig, L., et al., *Genome-wide transcriptional profiling of the Escherichia coli response to a proline-rich antimicrobial peptide.* Antimicrob Agents Chemother, 2004. **48**(9): p. 3260-7.

50. Dong, N., et al., *Short Symmetric-End Antimicrobial Peptides Centered on beta-Turn Amino Acids Unit Improve Selectivity and Stability.* Front Microbiol, 2018. **9**: p. 2832.

51. Wibowo, D. and C.X. Zhao, *Recent achievements and perspectives for large-scale recombinant production of antimicrobial peptides.* Appl Microbiol Biotechnol, 2019. **103**(2): p. 659-671.

52. Deng, T., et al., *The heterologous expression strategies of antimicrobial peptides in microbial systems.* Protein Expr Purif, 2017. **140**: p. 52-59.

53. Azari, M., S. Asad, and M.R. Mehrnia, *Heterologous production of porcine derived antimicrobial peptide PR-39 in Escherichia coli using SUMO and intein fusion systems.* Protein Expr Purif, 2020. **169**: p. 105568.

54. Fan, F., Y. Wu, and J. Liu, *Expression and purification of two different antimicrobial peptides, PR-39 and Protegrin-1 in Escherichia coli.* Protein Expr Purif, 2010. **73**(2): p. 147-51.

55. Parachin, N.S., et al., *Expression systems for heterologous production of antimicrobial peptides.* Peptides, 2012. **38**(2): p. 446-56.

56. Kuddus, M.R., et al., *Expression, purification and characterization of the recombinant cysteine-rich antimicrobial peptide snakin-1 in Pichia pastoris.* Protein Expr Purif, 2016. **122**: p. 15-22.

57. Mao, R., et al., *Optimization of expression conditions for a novel NZ2114-derived antimicrobial peptide-MP1102 under the control of the GAP promoter in Pichia pastoris X-33.* BMC Microbiol, 2015. **15**: p. 57.

58. Sang, M., et al., *Expression and characterization of the antimicrobial peptide ABP-dHC-cecropin A in the methylotrophic yeast Pichia pastoris.* Protein Expr Purif, 2017. **140**: p. 44-51.

59. Cao, J., et al., *Yeast-Based Synthetic Biology Platform for Antimicrobial Peptide Production.* ACS Synth Biol, 2018. **7**(3): p. 896-902.

60. Holaskova, E., et al., *Antimicrobial peptide production and plant-based expression systems for medical and agricultural biotechnology.* Biotechnol Adv, 2015. **33**(6 Pt 2): p. 1005-23.

61. Cary, J.W., et al., *Transgenic expression of a gene encoding a synthetic antimicrobial peptide results in inhibition of fungal growth in vitro and in planta.* Plant Sci, 2000. **154**(2): p. 171-181.

62. Poon, S., et al., *Co-expression of a cyclizing asparaginyl endopeptidase enables efficient production of cyclic peptides in planta.* J Exp Bot, 2018. **69**(3): p. 633-641.

63. Boto, A., J.M. Perez de la Lastra, and C.C. Gonzalez, *The Road from Host-Defense Peptides to a New Generation of Antimicrobial Drugs.* Molecules, 2018. **23**(2).

64. Potvin, G. and Z. Zhang, *Strategies for high-level recombinant protein expression in transgenic microalgae: a review.* Biotechnol Adv, 2010. **28**(6): p. 910-8.

65. Rasala, B.A., et al., *Robust expression and secretion of Xylanase1 in Chlamydomonas reinhardtii by fusion to a selection gene and processing with the FMDV 2A peptide.* PLoS One, 2012. **7**(8): p. e43349.

66. Kim, D.S., et al., *A new prokaryotic expression vector for the expression of antimicrobial peptide abaecin using SUMO fusion tag.* BMC Biotechnol, 2019. **19**(1): p. 13.

67. Foreman, D.J., et al., *Gas-Phase Sequencing of Cyclotides: Introduction of Selective Ring Opening at Dehydroalanine via Ion/Ion Reaction.* Anal Chem, 2019. **91**(24): p. 15608-15616.