Discovery and characterization of novel peptide biologics in herbal medicine

Nguyen, Kien Truc Giang

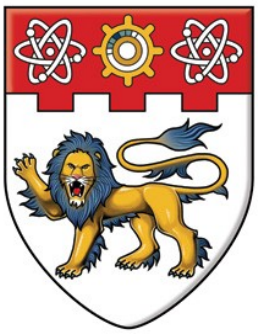
2011

Nguyen, K. T. G. (2011). Discovery and characterization of novel peptide biologics in herbal medicine. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/46476

https://doi.org/10.32657/10356/46476

DISCOVERY AND CHARACTERIZATION OF NOVEL PEPTIDE BIOLOGICS IN HERBAL MEDICINE



NANYANG TECHNOLOGICAL UNIVERSITY

NGUYEN KIEN TRUC GIANG

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2011

Acknowledgements

I would like to express sincere gratitude to my thesis advisor, Professor James P. Tam, for all of his guidance and supports. He has inspired me the interest in doing research. He has taught me how to plan for experiments, analyze data and think critically. I have learnt a lot from him not only research experiences but also life skills, management and leadership styles. I feel very fortunate to have the opportunity to work with him in this project.

I would like to thank Dr Wang Wei for his guidance and instructions of general molecular biology techniques and plant extractions.

I would like to thank Professor Sze Siu Kwan for teaching me mass spectrometry analysis.

I would like to thank my thesis advisory committee members, Professor Sze Siu Kwan and Professor Liu Chuan Fa, for all of their helpful suggestions, discussions and support through all these years.

I also would like to thank to the following friends and colleagues for their generous contributions to my project: Mrs. Young Soon, Ms Nguyen Quoc Thuc Phuong, Ms Nguyen Thi Kim Ngan, Ms Luu Thanh Thuy, Ms. Neo Tuan Ling, Ms Loo Shining, Ms Chiu Sheau Ming, Mrs Bian Yulan, Dr. Ardina Gruber, Mrs. Cai Lin, Mr. Zhang Sen, Dr. Sui Xiaogang, Dr. Wang Sujuan, Dr. Eom Khee Dong, Mrs. Loo Hann-Ne, Mr Lim Weihan, Mr Ye Weijian, Mr Antony Hardjojo, Mr. Garrick and Mr. Sean.

Finally, I would like to thank my family for their constantly emotional support during my project.

Table of Contents

List of Tables

List of Figures

Abbreviations

3D three dimensional, AP aminopeptidase, AEP asparaginyl endopeptidase, BLAST basic local alignment search tool, *C. chartacea Chassalia chartacea*, CCK cyclic cystine knot, CID collision induced dissociation, Co Company, CRPs cysteine-rich peptides, CT C-terminal, Da Dalton, DCM dichloromethane, DTT dithiothreitol, ER endoplasmic reticulum, EtOH ethanol, *H. biflora Hedyotis Biflora, H. corymbosa Hedyotis corymbosa*, H. diffusa *Hedyotis diffusa,* HPLC high performance liquid chromatography, IAA iodoacetamide, LC liquid chromatography, MALDI matrix assisted laser desorption ionization, *M. citrifolia Morinda Citrifolia,* MeOH methanol, MS mass spectrometry, MS/MS tandem mass spectrometry, NT N-terminal, NTPP N-terminal propeptide, NTR N-terminal repeat, *O. affinis Oldenlandia affinis*, RACE rapid amplification of cDNA end, *P. ginseng Panax ginseng, P. notoginseng Panax notoginseng, P. quinquefolius Panax quinquefolius,* PLCP Papain-like cysteine protease, RP reverse phase, SCX strong cation exchange, TFA trifluoroacetic acid, TOF time of flight, UV ultra violet, VTS vacuolar targeting signal.

Abstract

Peptide and protein biologics are therapeutics and prophylactics known for their desirability of high specificity and low toxicity. They display a wide range of biological activities such as antimicrobial, immunomodulation, neurotransmission and hormonal functions. Most biologics are derived from animal sources with few representatives from plants. In herbal medicine, these biologics have not received much attention with the common perception that they are unstable and unavailable as a source of active principles in decoctions. This bias is explained by the intrinsic instability of peptides and proteins during the concoctions preparation, poor absorption through the gastrointestinal tract and their susceptibility to enzymatic and acidic hydrolysis under harsh environment inside our stomach.

My hypothesis is that plant biologics may constitute a group of unexplored active principles in traditional herbal medicine. Research in this thesis focuses on plant cysteine-rich peptides (CRPs) such as cyclotides, uncyclotides, knottins, heveins and plant defensins. These peptides have recently gained great interest among the scientific community due to their remarkable stability and diverse biological functions. The major aims of this project are: (1) discovery of novel biologics from medicinal plants using tandem mass spectrometry, (2) characterization of their thermal and enzymatic stabilities, tissue specificity, genetic structures and disulfide connectivities, (3) pharmacological profilings of their biological activities for therapeutic applications. The proposed study may

provide lead compounds for drug development and aid in quantitative profiling of active ingredients in commercial viable medicinal herbs.

In this thesis, I have successfully identified more than 120 novel peptides comprising of cyclotides, uncyclotides, defensins and various CRPs from several medicinally important plants. By far only four uncyclotides have been reported, and thus the discovery of over 20 uncyclotides in this work is a surprising and provides insights about the biosynthetic processing of cyclic peptides. I also provide the first description of defensin-like peptides in cyclotide-producing plants. In addition, two novel classes of cysteine-rich peptides, ginsentide and morintide, have been identified from ginseng and *Morinda citrifolia,* respectively. Furthermore, I also expand the distribution of cyclotides to a new plant family, the Fabaceae.

Biophysical characterization of novel CRPs showed that most of them display high stability against heat and enzymatic digestions together with their amphipathic structures making them the likely active principles in herbal medicine. Activity screenings showed that many CRPs display potent antimicrobial, cytotoxic and hemolytic properties. Biologics with interesting pharmacological profiles could be used as lead compounds for drug development.

Chapter 1

Introduction

1. Biologics in herbal medicine

Herbal medicine is one of the oldest forms of medicine, existing for more than thousand years and almost as ancient as human civilization. It has a pivotal role in medicinal history and its effectiveness for treatment of human diseases has been well recognized [1]. In fact, about 25% of the pharmaceutical products prescribed worldwide come from plants [2]. However, most of the herbal derived compounds are small chemical entities with MW<500 Da, including a wide range of compounds such as volatile oils, tannins, alkaloids, polyphenols and flavonoids. These small compounds have been favored by big pharmaceutical companies due to their ease of production, low cost and oral bioavailability. Despite these advantages, there is a big stagnant in the number of new small molecules approved in the recent years [3]. This is because of the low specificity of small molecules owing to their small footprints, which often causes unwanted and unpredictable side effects and long-term toxicity.

Biologics are peptides and proteins with therapeutic potentials. They have MW considerably >500 Da and large footprint, which confers them high specificity and low toxicity. Currently, biologics represent the largest group of therapeutics, adjuvants and synthetic vaccines under different phases of clinical trials [4]. Despite a long history of prevalent usage in plant products and herbal

concoctions, few plant biologics are known. Past investigations of the active principles in herbal medicine have entirely focused on small molecules and missed the full landscape of bioactive biologics. Indeed, a data base of >20,000 plant-derived molecules with known or undetermined medicinal values are small molecules or non-peptide compounds. A contributing factor to this neglect is the intrinsic instability of peptides and proteins against heat during decoction preparations or their susceptibility to enzymatic and acidic hydrolysis during ingestion. In addition, few biologics are orally bioavailable through gut absorption or uptake, irrespective of their source of origin [5]. However, recent literature precedents suggest otherwise.

Cumulative evidence shows that several classes of cysteine-rich peptides (CRPs) in plants such as defensins, thionins, A1bs (also known as leginsulins), knottins and cyclotides are highly stable [6-9] (Fig. 1). They are very common in plants and genomic analysis of two model plants *Arabidopsis thaliana* and *Oryza sativa* (rice) reveal a surprise abundance of genes encoding for CRPs which account for up to 2-3 % of the gene repertoire of each model species [10]. Although the primary sequences, biochemical properties and functions may differ greatly, CRPs possess multiple disulfide bridges that cross brace their structures conferring them thermal, chemical and enzymatic stability [7, 9]. In addition, most possess an amphipathic structure which may promote their absorption through gut. These characteristics make CRPs the likely active principles in herbal medicine.

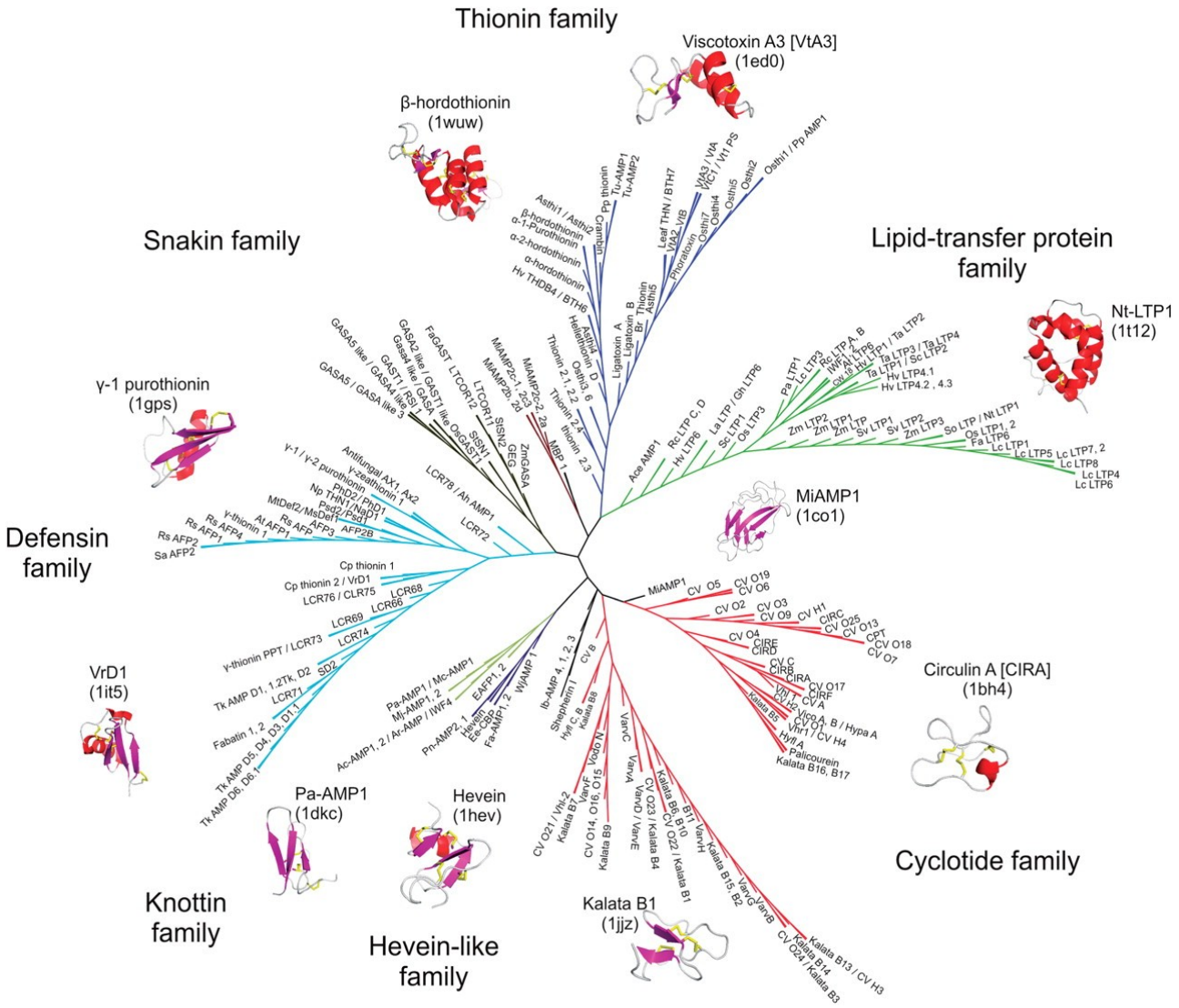


Figure 1. Phylogenetic tree of plant biologics in the PhytAMP database. Primary sequences of 271 plant CRPs were used for the sequence alignment and tree construction. *Figure taken from Hammami et al. [11].*

2. Cyclotides

Cyclotides are macrocyclic peptides belonged to a new class of CRPs that are recently gaining interest due to their exceptional stability and diverse biological functions. They typically contain 28-37 amino acid residues. They are gene-encoded as linear precursors which are subsequently processed to produce mature cyclic peptides [12, 13]. Despite their small size, they can be considered as mini proteins or biologics because of their well defined 3D structures containing an end-to-end circular peptide backbone cross braced by a cystine knot with three disulfide bonds (SS) in an arrangement of Cys I-IV, II-V, III-VI. The absence of both termini and the presence of a knotted SS arrangement in a close-end peptide structure confer cyclotides extraordinary stability against proteolysis as well as thermal and chemical degradations [9, 14, 15]. These defining characteristics also distinguish cyclotides from other plant-derived cystine-rich peptides such as the plant defensins or thionins which also contain a knotted SS-arrangement but having an open-end structure [6].

2. 1 Discovery of cyclotides

The story of cyclotide first came about when the Norwegian doctor Lorents Gran discovered the use of a traditional herbal decoction by a native African tribe to accelerate childbirth in 1970s [16]. The boiled decoction prepared from *Oldenlandia affinis* plant is ingested or applied topically to stimulate uterus contraction. The uterotonic agent was identified to be a small peptide which was named kalata B1 after the native tongue for the *O. affinis* “kalata kalata” [16].

Three more cyclotides were identified through drug screening programs in 1980s and 1990s [17, 18]. The structure of kalata B1, however, was only reported in 1995 [19], and the confirmation of all four cyclotides by chemical synthesis soon followed in 1998 [20, 21] . At present, about 150 cyclotides have been identified in three plant families of the Rubiaceae, Violaceae and Cucurbitaceae. They display a diverse range of biological activities, including anti-viral, anti-bacterial, and hormonal functions [22, 23]. Because the amino acid sequences of nearly 100 cyclotides have been reported, it has been suggested that they may surpass the well-known plant defensins in number and diversity [24, 25].

2. 2 Structural characteristic of cyclotides

Cyclotides comprise of 28-37 amino acids in an end-to-end cyclized manner with no free N- or C-termini [26]. They are further defined by a unique cyclic-cystine-knot motif (CCK), where the first two disulfide bonds, Cys I-IV and Cys II-V, and their interconnecting backbone segments are penetrated by a third disulphide bond Cys III-VI [19] (Fig. 2). Although the cystine-knot structure has been found in a diverse class of proteins such as toxins, growth factors and inhibitory proteins in both plants and animals [27], cyclotides are the only class of protein with a close-end backbone [28]. It is due to this CCK motif that cyclotides are able to withstand thermal, chemical and enzymatic degradations [9]. The CCK arrangement also forces some hydrophobic residues to be exposed outward together with the hydrophilic and charged residues rendering cyclotides an amphipathic structure [26].

Cyclotides consist of six individual loops with each loop defined by a section of residues between two successive cysteine residues. Dependent on the presence or absence of cis-Pro peptide bond in loop 5, cyclotides can be divided into two subfamilies, Möbius and bracelet, respectively. The cis-Pro peptide bond causes a twist in the circular backbone of Möbius cyclotides, a unique structural feature that is absent in bracelet cyclotides. Representatives of the Möbius and bracelet cyclotides can be seen from kalata B1 and cycloviolacin O1, respectively (Fig. 3). These two peptides although display similar overall fold, their surface hydrophobic patches are located at entirely different positions. The hydrophobic patch of kalata B1 constitutes of residues in loop 2 and loop 5, while it is predominantly involved in residues of loop 2 and 3 in cycloviolacin O1.

2. 3 Sequence conservation of cyclotides

Cyclotides contain six absolutely conserved cysteine residues which divide cyclotides into six individual loops. All loops differ in size and diversity making them tolerant to substitutions but still maintain certain conserved residues. The sequence alignment of selected cyclotides is shown in Fig. 4 to demonstrate their conservation patterns. Loop 1 and 4 form the backbone segments of the CCK scaffold and hence are the most conserved loops [26]. Loop 1 has the consensus sequences of (G/A) E (T/S), and loop 4 comprises only a single residue Thr, Ser or much less common Tyr and Lys residue. The first residue of loop 1 is always a

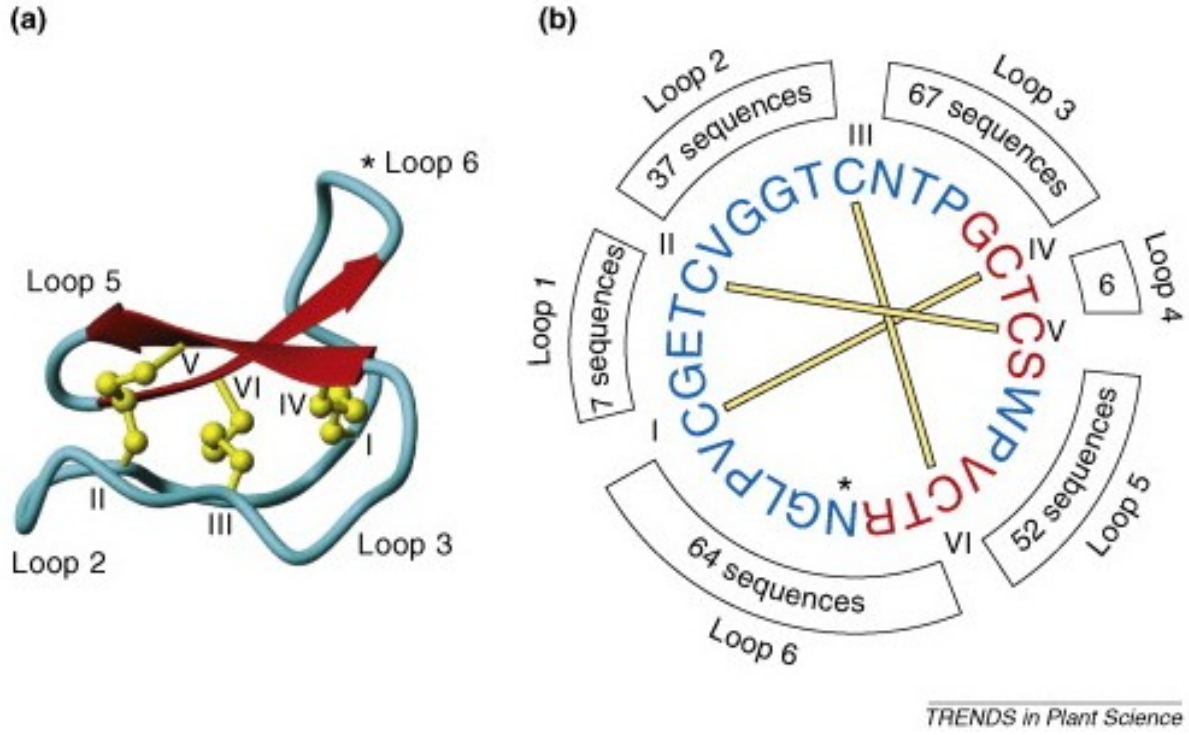


Figure 2. Schematic structure of the prototypic cyclotide kalata B1. (a) 3D structure of kalata B1 shows the cystine-knot motif with the cysteinyl residues labeled in Roman numerals (in yellow) and the distorted β-sheet (in red). (b) The primary sequence of kalata B1 is shown using one-letter amino acid codes. Diversity of cyclotides is indicated by the numbers of known loop sequences. The ligation site is located at the Asn of loop 6 and indicated with an asterisk (*). *Figure taken from Craik, D. J. [22].*

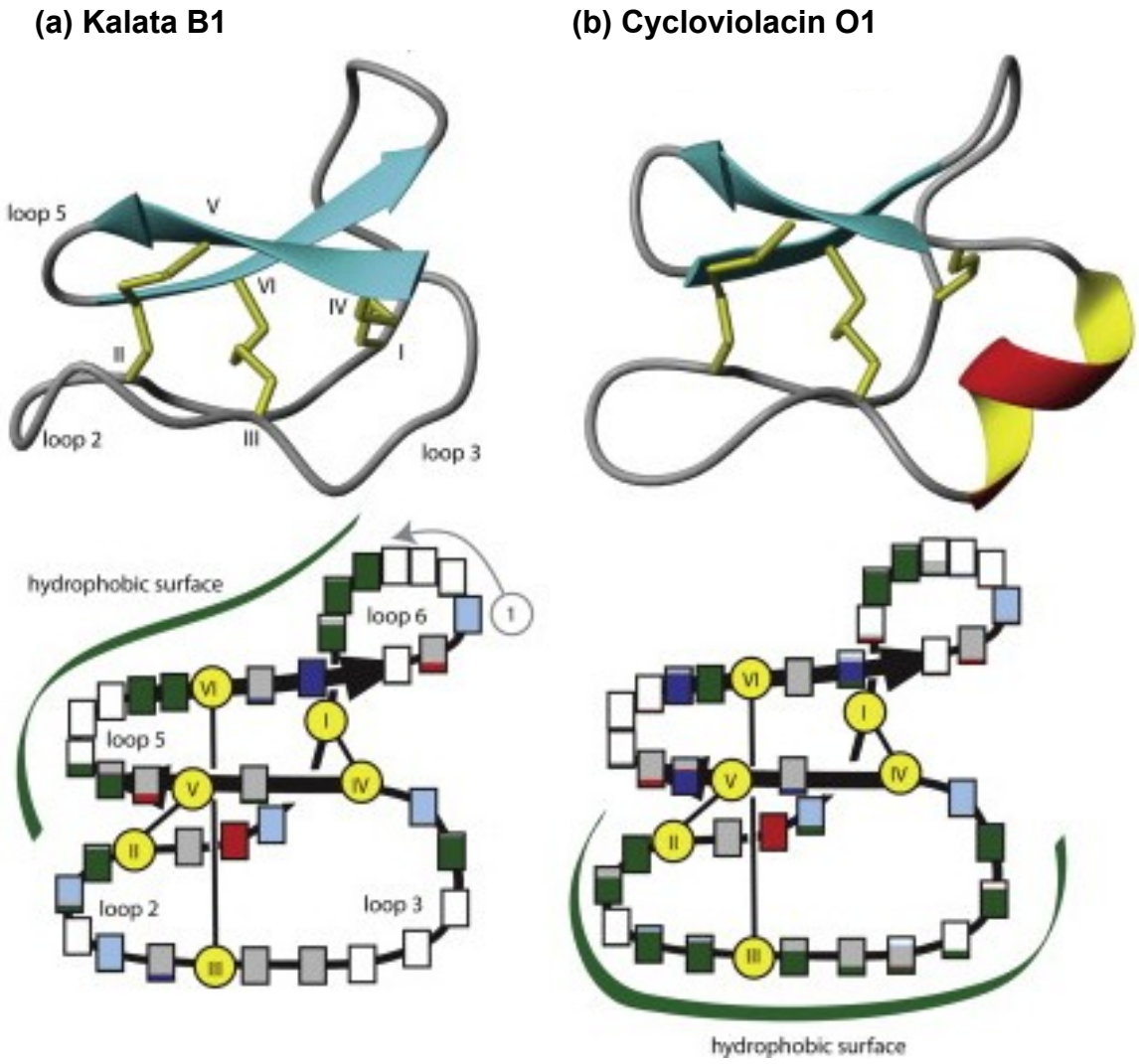


Figure 3. Representative structures of Möbius and bracelet cyclotides. (a) Kalata B1 has a Möbius structure characterized by a cis-Pro residue in loop 5 that leads to a conceptual twist. (b) Cycloviolacin O1 of the bracelet family. About two-third of cyclotides belong to this family. *Figure taken from Wang et al. [29].*

small side chain amino acid such as Gly and Ala. The second residue is Glu which is absolutely conserved in all known cyclotides due to its function to stabilize the overall framework by forming multiple hydrogen bonds with other residues in loop 3 [26].The third residue is either Ser or Thr of which bracelet cyclotides have a higher propensity for Ser residue while Möbius cyclotides prefer Thr residue at this position.

The amino acid compositions of loop 2 and 3 are quite different between two subfamilies. Loop 2 of Möbius cyclotides usually contains a Gly-turn whereas a Pro-turn is often found in the bracelet cyclotides [26]. Loop 3 also displays significant difference between two subfamilies. Möbius cyclotides usually contain four residues as compared to six or seven residues in bracelet cyclotides. While there is a relatively high conservation in loop 3 of Möbius cyclotides, this loop is quite diverse and often features a short helical segment in bracelet cyclotides.

Finally, the defining difference of both families lies in loop 5, where clusters of cationic residues are found in the bracelet cyclotides and hydrophobic residues including the characteristic cis-Pro residue are found in the Möbius cyclotide. In loop 6, both subfamilies display great diversity yet conserving a Gly-Asn/Asp sequence that plays an important role in the biosynthesis of cyclotides [26].

Möbius

cycloviolacin H3	GLPV-CGETCFGGT-C---NTPGCICD--PWPVCTRN
cycloviolacin O14	GSIPACGESCFKGK-C---YTPGCSCS--KYPLCAKN
cycloviolacin O15	GLVP-CGETCFTGK-C---YTPGCSC---SYPICKKN
cycloviolacin O16	GLP--CGETCFTGK-C---YTPGCSC---SYPICKKIN
cycloviolacin O22	GLPI-CGETCVGGT-C---NTPGCTC---SWPVCTRN
cycloviolacin O23	GLPT-CGETCFGGT-C---NTPGCTCD-SSWPICTHN
kalata B1	GLPV-CGETCVGGT-C---NTPGCTC---SWPVCTRN
kalata B2	GLPV-CGETCFGGT-C---NTPGCSC---TWPICTRD
kalata B3	GLPV-CGETCFGGT-C---NTPGCTC---DPWICTRD
kalata B4	GLPV-CGETCVGGT-C---NTPGCTC---SWPVCTRD
kalata B12	GS-L-CGDTCFVLG-C---NDSSCSC---NYPICVKD
varv A	GLPV-CGETCVGGT-C---NTPGCSC---SWPVCTRN
varv B	GLPV-CGETCFGGT-C---NTPGCSCD--PWPMCSRN
varv C	GVPI-CGETCVGGT-C---NTPGCSC---SWPVCTRN
vhl-2	GLPV-CGETCFTGT-C---YTNGCTCD--PWPVCTRN
vibi A	GLPV-CGETCFGGT-C---NTPGCSC---SYPICTRN
vibi B	GLPV-CGETCFGGT-C---NTPGCTC---SYPICTRN
violacin A	SAIS-CGETCFKFK-C---YTPRCSC---SYPVC-K-
violapeptide 1	GVPV-CGETCVGGT-C---NTPGCSC---SRPVCTXN
vodo M	GAPI-CGESCFTGK-C---YTVQCSC---SWPVCTRN

Bracelet

circulin A	GIP--CGESCVWIP-C-ISAALGCSC---KNKVCYRN
circulin C	GIP--CGESCVFIP-C-ITSVAGCSC---KSKVCYRN
circulin F	AIP--CGESCVWIP-C-ISAAIGCSC---KNKVCYR-
cycloviolacin H1	GIP--CGESCVYIP-C-LTSAIGCSC---KSKVCYRN
cycloviolacin H2	SAIA-CGESCVYIP-C--FIPGGCSC---RNRVCYLN
cycloviolacin O2	GIP--CGESCVWIP-C-ISSAIGCSC---KSKVCYRN
cycloviolacin O3	GIP--CGESCVWIP-C-LTSAIGCSC---KSKVCYRN
cycloviolacin O4	GIP--CGESCVWIP-C-ISSAIGCSC---KNKVCYRN
cycloviolin B	GTA--CGESCYVLP-C---FTVGCTC---TSSQCFKN
cycloviolin C	GIP--CGESCVFIP-C-LTTVAGCSC---KNKVCYRN
cycloviolin D	GFP--CGESCVFIP-C-ISAAIGCSC---KNKVCYRN
hyfl I	GIP--CGESCVFIP-C-ISGVIGCSC---KSKVCYTN
hyfl J	GIA--CGESCAYFG-C--WIPGGCSC---RNKVCYFN
hyfl K	GTP--CGESCVYIP-C-FTAVVGCTC---KDKVCYLN
hyfl L	GTP--CAESCVYLP-C-FTGVIGCTC---KDKVCYLN
hypa A	GIP--CAESCVYIP-CTITALLGCSC---KNKVCY-N
kalata B16	GIP--CAESCVYIP-CTITALLGCKC---QDKVCY-D
kalata B17	GIP--CAESCVYIP-CTITALLGCKC---KDQVCY-N
kalata B5	GTP--CGESCVYIP-C-ISGVIGCSC---TDKVCYLN
palicourein	GDPTFCGETCRVIPVCTYSAALGCTCDDRSDGLCKRN

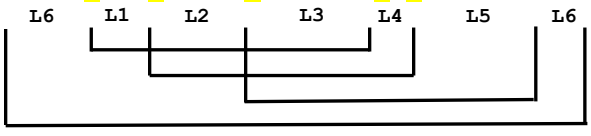


Figure 4. Sequence comparison of selected cyclotides. Cyclotides can be divided into two subfamilies Möbius and bracelet based on the presence or absence of cis-Pro bond in loop 5, respectively. *Sequences taken from Cybase [30].*

2. 4 Occurrence of cyclotides

Cyclotides are currently found in three plant families: the Rubiaceae, Violaceae and Cucurbitaceae. Of these families, the Violaceae is the richest source of cyclotides with an individual species may produce up to 100 different cyclotides [25]. The Violaceae is a small family of flowering plants consisting of about 900 species in 23 genera [31]. They can be found throughout the world but are most abundant in temperate regions. It has been demonstrated that cyclotides have widespread distribution and likely to be present in almost, if not all, Violaceae species [32], irrespective of plant morphology (herb/shrub/tree) or life cycle (annual/ perennial). In addition, each individual species possesses a unique arsenal of cyclotides that can potentially use as species-specific markers for botanical authentication.

The ubiquitous nature of cyclotides in the Violaceae family is contradictive to their dispersed distribution in the Rubiaceae family. It has been predicted that cyclotides present in approximately 5 - 10 % of the Rubiaceae species [33]. With more than 13,000 species in 637 genera, the Rubiaceae family is one of the largest of the angiosperms, ranking number four after the Asteraceae, Orchidaceae and Fabaceae [34]. It consists of mostly tropical trees, shrubs and herbs. The differences in geographical locations and growing temperatures between the Rubiaceae and Violaceae species predispose them to different ranges of pests and pathogens, which may lead to a different need in cyclotide and other secondary metabolite defenses.

In the Cucurbitaceae, only two cyclotides, MCoTI-I and MCoTI-II, have been reported from the seeds of *Momordica cochinchinensis* [35]. Their primary

sequences are entirely different from those of the Rubiaceae and Violaceae families. In fact, these two peptides belong to the trypsin squash inhibitor family but happen to possess a cyclic backbone, and thus were grouped to the cyclotide family. Trypsin squash inhibitors are CRPs of similar size and cystine-knot structure as cyclotides but are opened at both ends. They bind trypsin with high affinity and act as potent inhibitors [36].

2. 5 Biological activities of cyclotides and mechanism of action

Cyclotides are produced in plants for defense against pests and pathogens based on their potent insecticidal and antimicrobial activities [21, 37]. They also display a wide range of other biological activities including anti-HIV [38], cytotoxicity [39], hemolysis [40], uterotensin [16] and neurotensin [18] (Table 1). The diverse functions of cyclotides have been postulated to associate with their membrane-active properties [41]. Electrophysiological experiments show that kalata B1 exerts its action by forming multimeric transmembrane pores [42] with channel-like activity (Fig. 5). This causes leakage of the cellular contents and cell death.

Interestingly, Möbius and bracelet cyclotides have been shown to have different membrane binding modes despite sharing a conserved CCK motif [29]. This difference is attributed to the location of the hydrophobic patches on the peptide surfaces which mediate the membrane interactions. The hydrophobic patch of Möbius cyclotides mainly involves in residue in loop 2 and 5 as compared to loop 2 and loop 3 in bracelet cyclotides. This may explain the differences in biological activities of these two peptide subfamilies.

Table 1. Bioactivities of selected cyclotides and their sources

Cyclotide	Plant Source	Activity	References
Kalata B1	*Oldenlandia affinis*	Uterotensin	[16]
		Antimicrobial	[21]
		Insecticidal	[37]
Violapeptide I	*Viola tricolor*	Hemolysis	[43]
Cyclopsychotride A	*Psychotria longipes*	Neurotensin antagonism	[18]
		Antimicrobial	[21]
Circulins A and B	*Chassalia parvifolia*	Anti-HIV	[17]
		Antimicrobial	[21]
Cycloviolin A – D	*Leonia cymosa*	Anti-HIV	[44]
Palicourein	*Palicourea condensata*	Anti-HIV	[45]
Cycloviolacin O2	*Viola odorata*	Cytotoxicity	[46]
		Antifouling	[47]

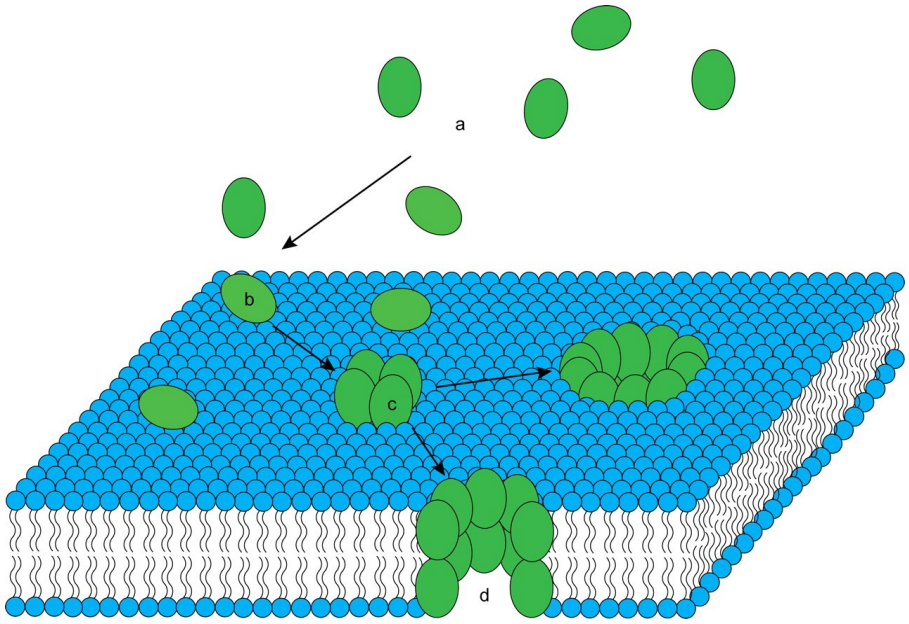


Figure 5. Proposed model for the pore formation of kalata B1 in the lipid membrane. (a) Free kalata B1 exists as monomer form in the solution. (b) Membrane bound kalata B1. (c) Kalata B1 self-associates on the membrane to form tetramer. (d) Final pore formation consisting of multimeric kalata B1 molecules. *Figure taken from Huang et al. [42].*

2. 6 Gene architecture of cyclotides in the Violaceae and Rubiaceae family

Cyclotides are ribosomally synthesized as linear precursor proteins which are post translationally modified to join the N- and C- termini to produce mature cyclic peptides. Their encoding genes are first cloned from the African plant *O. affinis* of the Rubiaceae family [37]. DNA plot analysis of cyclotide genes in this plant revealed up to 12 hybridized bands, indicating that cyclotides derived from repertoire of a multigene family [37]. Cyclotide-encoding cDNA sequences have also been characterized in several species of the Violaceae [12, 48, 49]. The predicted precursors from these clones share low sequences homology with those characterized from *O. affinis,* presumably due to distant evolutionary relationship between the Rubiaceae and Violaceae family. The gene architecture, however, is remarkably similar in these two families [12] but is significantly different from the Fabaceae as described later in chapter 7.

Each cyclotide gene encodes for five major domains [12] consisting of the ER signal sequence, N-terminal propeptide (NTPP), one to three mature cyclotide domains with each preceded by a short N-terminal repeat region (NTR), and a C-terminal propeptide (CTPP) (Fig. 6). Apart from the mature peptides, the endogenous functions of each domain are not entirely understood. The presence of ER signal in the first part of the protein presumably directs the precursor to the ER lumen, which provides a suitable environment for the oxidative folding to occur [50]. The compartmentalization of the ER enables a distinct set of folding catalysts such as protein disulfide isomerase (PDI) to facilitate the formation and isomerisation of disulphide bonds [50]. Recently, a novel PDI was discovered from the African plant *O. affinis* and has been demonstrated to assist the correct folding of cyclotides [51].

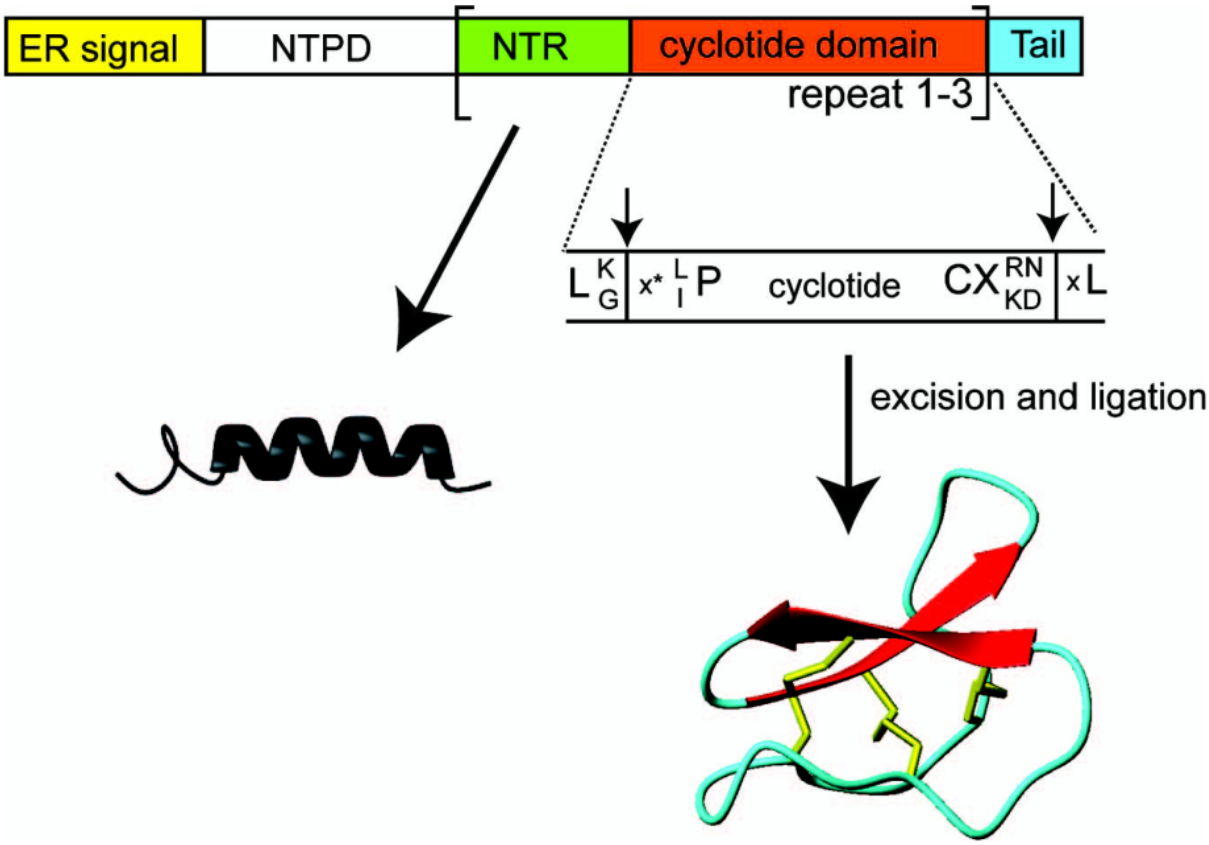


Figure 6. Schematic representation of a cyclotide precursor protein. Each precursor protein consists of an ER signal sequence, N-terminal propeptide (NTPP), N-terminal repeat (NTR) followed by one to three mature cyclotide domains separated by NTR regions of about 25 to 30 amino acids, and finally a C-terminal tail of variable length. The conserved residues around the processing sites are shown and the cleavage sites are arrowed. X represents any amino acid. *Figure taken from Gunasekera et al. [52].*

The ER signal is followed by the NTPP domain that is not tightly conserved in either composition or number of amino acids. It has been proposed that the length of the NTPP is not important factor that whether the mature peptide will be open- or close-end [24]. In addition, the fact that each gene may encode for up to three cyclotide domains but only one NTPP, suggests that it probably has no direct role in the folding or detoxification of the mature peptides. The endogenous function of this domain thus remains to be characterized. It is possible that it may harbor protein sorting signal required for intracellular trafficking of cyclotides to a specific organelle.

Located downstream to this domain is the NTR region, which is highly conserved within species but not between species [12]. Structural characterization of this region from the *O. affinis* (Rubiaceae) and *V. odorata* (Violaceae) showed that they have similar 3D fold despite of having no apparent sequence homology [12]. In both species, the NTR regions adopt a structurally conserved α-helical motif. The conserved structural element across two different plant families suggests they may have functional importance in oxidative folding, processing or detoxification of the mature macrocyclic peptides [12].

Adjacent to the NTR region is the cyclotide domain, which is approximately 30 residues in length and displays a diverse range of sequence diversity. Sequence alignment of cyclotide precursors from *O. affinis*, *Viola odorata* and *V. tricolor* led to identification of several highly conserved amino acids flanking at both site of the mature peptide. A Leu residue located at one residue away from both N- and C-terminal processing points appears to have a

functional importance as it is strongly conserved in both Rubiaceae and Violaceae family [53]. It is likely that this residue is required for recognition by processing enzymes that release the mature cyclotides from their embedded precursors.

The final part in the cyclotide genes comprises of a short hydrophobic tail (also called CTPP) of around 3 to 7 residues. It contains a conserved tripeptide motif, which appears to be essential for recognition by the asparaginyl proteinase to catalyze the backbone cyclization [54, 55]. The removal of the entire CTPP from *oak1* gene abolished the production of circular kalata B1 and led to formation of only linear forms [55]. In addition, it has also been proposed that the vacuolar targeting signal (VTS) is located within the CTPP domain [52]. Many vacuolar proteins are known to possess C-terminal VTS that have no common sequence identity except the short stretches of hydrophobic amino acids [56]. However, there is no information about the final destination of cyclotides as well as no evidence that this sequence is able to direct cyclotides to the plant central vacuole.

2. 7 Biosynthesis processing

The processing of cyclotide precursors to produce mature cyclic peptides is a multiple steps mechanism (Fig. 7). Study by Gillon *et al.* [55] using kalata B1 precursor as a model protein showed that the earliest event begins with the cleavage of ER signal sequence, which primary function is probably to facilitate the translocation of the newly synthesized polypeptides chain into the ER lumen. This cleavage occurs co- or post-translationally and generates a stable product detected *in vivo* as an 11 kDa intermediate containing NTPP, NTR, cyclotide

domain and CTPP [55]. Inside the ER lumen, the folding and disulfide bond formation of the nascent protein are assisted by the high oxidative redox potential and ER residential proteins such as PDI and chaperon proteins. Only polypeptide chains that are correctly folded allow leaving ER. Misfolded proteins are retained and subsequently undergo ER-associated degradation [57]. The strict control in oxidative folding process provides explanation why only one isomer with a knotted connectivity of Cys I-IV, II-V, III-VI being formed out of the 15 possible disulfide isomers.

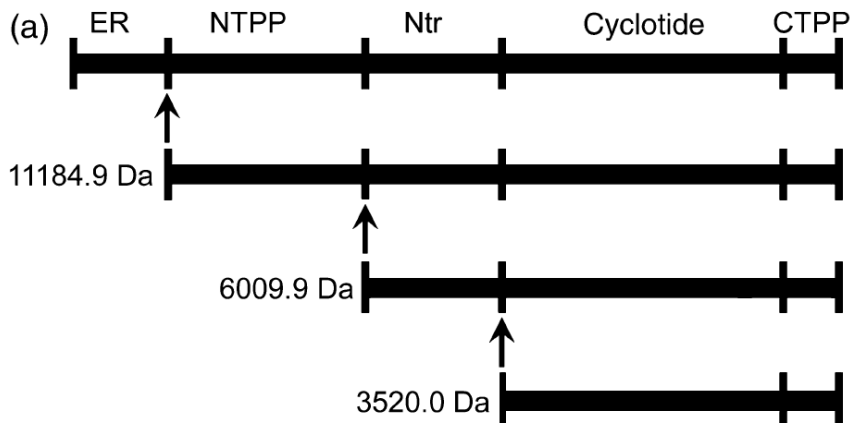


Figure 7. Schematic representation of kalata B1 precursor and potential linear intermediates during the biosynthesis process. Arrows indicate the processing sites. *Figure taken from Amanda et al. [55].*

Once formation of the disulfide bonds is complete, the linear intermediate is probably transported from the ER to the Golgi apparatus for further processing. A second cleavage occurs between NTPP and NTR to generate a 6 kDa linear intermediate. It is not known when this process happens and mediates by which proteases. Golgi apparatus and its associated vesicles in plants are known to

contain a number of processing enzymes [58]. It is possible that this second proteolytic event happens during the Golgi trafficking process.

The fate of the 6 kDa intermediate afterward is uncertain. It is not known whether it will enter the default secretory pathway to the intercellular fluid or targeted to the plant central vacuole. Although many plant defense peptides are found to localize in the intercellular fluid which constitutive a physical barrier to infection [59], there is no information about the cellular distribution of cyclotides. Both plant vacuole and intercellular fluid compartments are peptidase-rich environments, which allow further processing of the 6 kDa intermediate.

The final excision and cyclization of the cyclotide domain require at least three enzymatic-assisted events: two proteolytic reactions at the N- and C-termini and one transpeptidation reaction to close the peptide backbone. The N-terminal cleavage locates after the conserved Leu-X dipeptide motif and less frequently after Asn residue prior to the first residue of the cyclotide domain. This process presumably happens before the C-terminal cleavage to generate a 3.5 kDa linear precursor. Enzyme catalyzing this process, however, is still unidentified.

Cleavage at the C-terminus occurs after the highly conserved Asn and occasionally Asp residues. There is strong evidence that asparaginyl endopeptidase (AEP) is responsible for both processing at the C-terminus and transpeptidation reaction [54, 55]. AEP is an endopeptidase that specifically recognizes Asn and cleaves after its C-terminal side to generate an acyl-enzyme intermediate. It also cleaves after the Asp residue but at a slower rate. The CTPP

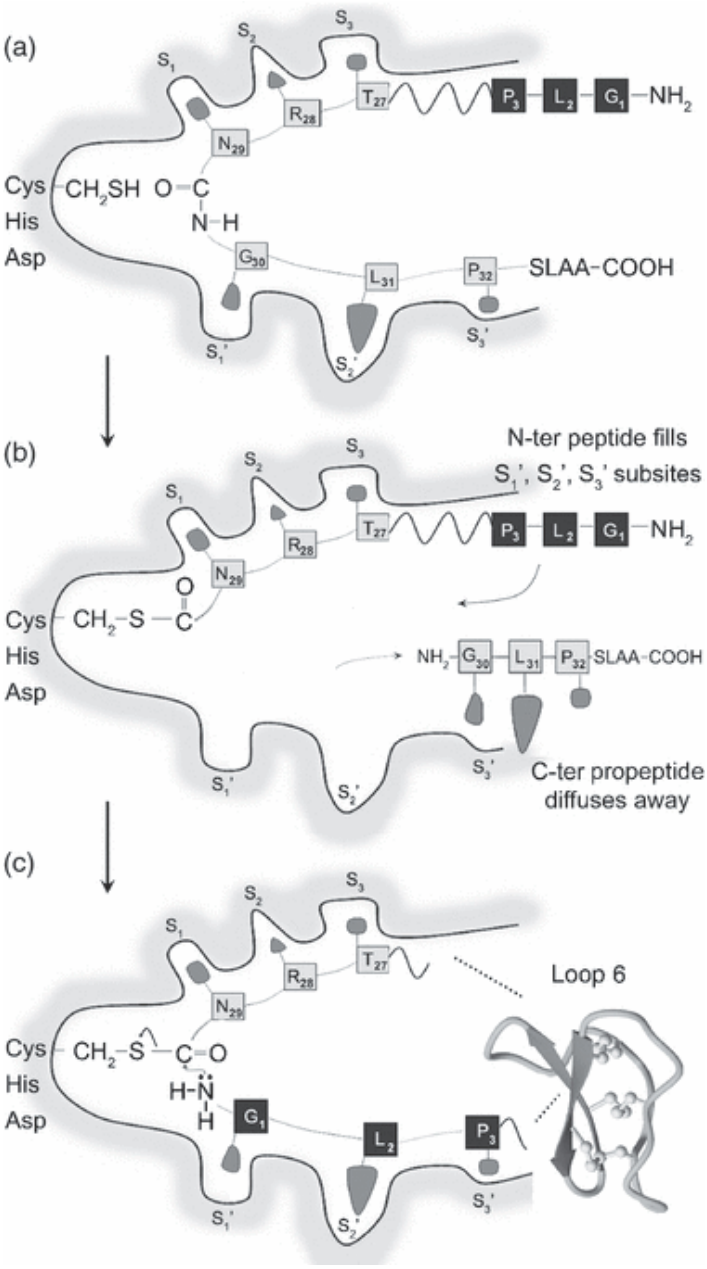


Figure 8. Proposed model for backbone cyclization of kalata B1 mediated by an asparaginyl endopeptidase (AEP). (a) Binding of the conserved Asn and residues at the C-terminal processing site to the active site of AEP. (b) AEP cleaves after the C-terminal site of Asn residue forming an acyl-enzyme intermediate and releasing the C-terminal propeptide. (c) Residues at the N-terminal site occupy the newly vacated S_1', S_2' and S_3' binding sites. Deacylation and backbone cyclization mediated by a nucleophilic attack of the amino group of the N-terminal Gly residue. *Figure taken from Saska et al. [54].*

subsequently diffuses away after the cleavage allowing the N-terminal residues to occupy the S1' subsite and initiate a nucleophilic attack to join the N- and C-termini together (Fig. 8) [54]. It should be noted that under this model the N-terminal cleavage must happen before the C-terminal cleavage to liberate the free amino group on the N-terminal residue allowing for the aminolysis reaction.

3. Uncyclotides

In 2006, Ireland *et al.* [24] discovered the first linear cyclotide violacin A. Two additional linear versions of kalata B9 and kalata B10 cyclotides were identified in *O. affinis* [60]. However, Plan *et al.* [60] suggest that these two linear cyclotides are likely degraded products resulted from cleavage of Asn-Gly ligation bond via the formation of anhydride succinimide intermediate. In 2010, another linear cyclotide, psyle C, was identified from the Micronesian plant *Psychotria leptothyrsa* [61]. All these linear cyclotides are of Möbius subfamily and the linear analogs of bracelet cyclotides are yet to be discovered.

Table 2. Sequences of uncyclotides hitherto discovered

Uncyclotide	Primary sequence	MW	References
Violacin A	SAISCGETCFKFKCYTPRCSCS-YPVCK	2840	*[24]*
Psyle C	KLCGETCFKFKCYTPGCSCS-YPFCK	3004	*[61]*
Kalata B9L	GSVFNCGETCVLGTCYTPGCTCNTYRVCTKD	3288	*[60]*
Kalata B10L	G-LPTCGETCFGGTCNTPGCSCSSWPICTRD	3046	[60]

Homology search using BLAST reveals several putative cyclotide-like sequences at mRNA level in common crop plants such as rice (*Oryza sativa*), maize (*Zea mays*), and wheat (*Triticum aestivum*) [32]. These cyclotide-like peptides, however, are yet to be confirmed at protein level and thus remain uncertain whether they are cyclic or linear. Since the term “linear cyclotide” is contradictory, the term “uncyclotides” will be used in this thesis to describe the linear variations of cyclotides, which share the same cystine arrangement, and high sequence homology, but are unzipped at both ends as linear, open-end peptides.

3. 1 Stability of uncyclotides

It has been demonstrated that Violacin A maintains a well defined 3D structure and is stable against endopeptidases such trypsin and thermolysin [24]. Linear analogs of kalata B1 also display similar thermal, chemical and enzymatic resistances. This suggests that it is the cystine knot rather than the cyclic back bone accounts for the stability of cyclotides. Violacin A, however, is susceptible to digestions by aminopeptidases. This highlights the additional advantage of cyclic backbone for resistance against exopeptidases.

3. 2 Biosynthesis of uncyclotides

The biosynthesis of uncyclotides is still not well understood. Of the four uncyclotides isolated, only violacin A has gene characterized, which provides most of our understanding about the bioprocessing of uncyclotides. The cDNA

encoding for violacin A reveals a non-sense mutation that introduces a premature stop codon located at one residue prior to the C-terminal Asn/Asp. This mutation prevents the translation of the C-terminal Asn/Asp, a key residue essential for the backbone ligation. Similar mutation may have happened in psyle C as this ligation residue is also absent in its primary sequence. The linear nature of violacin A and psyle C is thus genetic pre-determined, whereas linear analogs of kalata B9 and B10 are likely resulted from chemical degradation.

3. 3 Biological activities of uncyclotides

Violacin A displays a marked reduction in hemolytic activity as compared to kalata B1 [24]. It has been demonstrated that linearization of cyclotides causes loss of activity indicated by a complete lack of hemolytic properties of various acyclic permutants of kalata B1 [40]. Disruption of the cyclic backbone may cause loss of some structural features essential for bioactivities. Interestingly, psyle C maintains the cytotoxicity comparable to several cyclotides such as varv A and varv E [61]. This naturally occurring uncyclotide may have evolved new structural features to compensate for the absence of the cyclic backbone.

4. Application of cyclotides as peptide-drug scaffold

The cyclic cystine knot in cyclotides provides a very stable and compact core structure which serves as a promising template for grafting of bioactive epitopes. A short peptide segment with a known binding target can be incorporated into the cyclotide scaffold. This allows generation of therapeutic

miniproteins with high binding affinity and selectivity against a specific target for medical applications. This approach has been successfully applied for development of novel VEGF-A antagonists by introducing a VEGF-A antagonist peptide onto the cyclotide framework [62]. The application of cyclotides as peptide-drug scaffold is very promising and provides a cheaper alternative to monoclonal antibodies and large, complex therapeutic proteins.

5. Aims of this study

The beneficial effects of botanical extracts for treatment of human diseases have been long recognized. Most are based on prescriptions of traditional folklore medicine that contain many undetermined components. Some of these reports are more specific and are based on small molecules <500 Da such as alkaloids and triterpenoids, saponins and polyphenols. These studies, however, lack a complete landscape of small chemicals and peptide biologics in the botanical extracts responsible for their therapeutic applications.

This thesis focuses on the identification and characterization of druggable biologics in medicinal plants with a particular focus on cyclotides, uncyclotides and several individual groups of small cysteine-rich-peptides. My hypothesis is that these biologics may constitute a novel group of unexplored active principles in herbal medicine. New sequencing technique will be investigated to facilitate the rapid discovery of their primary structures and disulfide connectivities in plants. This aspect will be explored in chapter 3 together with the discovery of novel cyclotides and uncyclotides in multiple *Hedyotis* species. This chapter also provides the first description of the bracelet and hybrid uncyclotides. Furthermore, the biodegradable pathway of cyclotides and uncyclotides will also be studied. In chapter 4, I will report an in-depth characterization of novel cyclotides and uncyclotides from a local medicinal plant *Chassalia chartacea*. Biological activities, structure-activity relationships and genetic characterization of the novel peptides will also be described in detail. Chapter 5 reports the discovery of an unusual class of highly anionic CRPs possessing an inhibitor cystine-knot motif.

The most anionic peptide has -7 net charge and this is the first time such anionic CRPs have ever been isolated. Chapter 6 studies the CRPs content in ginseng, the most famous Chinese herb. A novel class of CRP with unique disulfide pattern resembling a pretzel knot was characterized. Their heat stability and biological activities will also be described. Chapter 7 reports the discovery of cyclotides in a new plant family, the Fabaceae. Genetic characterization of novel cyclotides reveals unexpectedly that they are originated from a chimeric precursor consisting of cyclotide and leginsulin domains. Such chimeric gene arrangement provides a likely evidence of horizontal gene transfer or convergent evolution of CRPs in plants. Chapter 8 reports a new post-translational pathway describing how plant generates the diversity of CRPs beyond the DNA coding capacity. Finally, additional 19 CRP sequences characterized in various medicinal plants but not studied in detail are summarized in Table A1 (Appendix A). Overall, my study reports the discovery of over 120 novel CRPs and provides new insight about their diversity and evolution in plants.

Chapter 2

Materials and Methods

1. Materials

1. 1 Chemicals and Reagents

All the chemicals and reagents used in this study were of analytical or molecular biology grade and purchased from the following companies:

Trifluoroacetic acid (TFA)	Merck
Acetonitrile (ACN)	Fisher
Dichloromethane (DCM)	Merck
Methanol (MeOH)	Merck
Ethanol (EtOH)	Merck
Isopropanol	Fisher
Butanol	Fisher
Acetic acid	Merck
Dithiothreitol (DTT)	Sigma-Aldrich
Iodoacetamide (IAA)	Sigma-Aldrich
Tris[2-carboxyethyl]phosphine (TCEP)	Sigma-Aldrich
Alpha-Cyano-4-hydroxycinnamic Acid (CHCA)	Sigma-Aldrich
Ammonium bicarbonate (NH_4HCO_3)	Sigma-Aldrich
Sodium citrate	Sigma-Aldrich
Sodium chloride	Sigma-Aldrich

Potassium chloride	Sigma-Aldrich
Calcium chloride	Sigma-Aldrich
Hydrochloric acid (HCl)	Sigma-Aldrich
Sulfuric acid (H_2SO_4)	Sigma-Aldrich
IPTG	Fermentas
dNTP nucleotide mix	Fermentas
Triton-X	Bio-Rad
X-gal	Fermentas
Agarose	Bio-Rad
C18 media	Grace Davison Discovery Sciences

1. 2 Enzymes

Trypsin	Roche
Chymotrypsin	Roche
Endoproteinase Glu-C (EndoGlu-C)	Roche
Pepsin	Sigma-Aldrich
Pfu DNA polymerase	Fermentas
Taq DNA polymerase	NEB
T4 ligase	Fermentas

1. 3 Cell culture media

LB agar	BD, Difco
LB broth	BD, Difco
Trypticase soy broth (TSB)	Sigma-Aldrich

RPMI	PAA Laboratories
DMEM	PAA Laboratories

1. 4 Kits

Plasmid extraction	QIAGEN
QIAquick PCR Purification	QIAGEN
DNA extraction	Invitrogen
RNA extraction	Invitrogen
3'-RACE	Invitrogen
5'-RACE	Invitrogen
ACCQ-tag Ultra	Waters
Elisa IL-1, Il-8, IL-6 and TNF-α	BioLegend

1.5 Cell lines

HeLa	Gift from Prof Tam
A549	Gift from Prof Lars
THP-1	Gift from Dr Tan Suet Mien

1.6 Bacterial strains

Escherichia coli	ATCC 25922
Pseudomonas aeruginosa	ATCC 10145
E. faecalis	ATCC 13883
Staphylococcus aureus	ATCC 12600
Staphylococcus epidermis	ATCC 14990
Streptococcus salivarius	ATCC 13419
Streptococcus mutant	ATCC 25175

Candida albicans	ATCC 11006
Candida tropicalis	ATCC 750
Candida kefyr	ATCC2512

1.7 Others

Primers	1st Base
pGEM T-easy vector	Promega

1.8 Plant materials

Plant materials were collected at the herb garden of Nanyang Technological University (Singapore) and the countryside of Vietnam. Voucher samples were authenticated by Herbarium Department (Singapore Botanical Garden). Plants examined in this thesis included: *Hedyotis biflora, Hedyotis corymbosa, Hedyotis diffusa, Hedyotis herbacea, Hedyotis congesta, Hedyotis auricularia, Oldenlandia affinis, Chassalia chartacea, Morinda citrifolia, Panax ginseng, Panax notoginseng, Panax quinquefolius, Clitoria ternatea, Clitoria fairchildiana, Clitoria laurifolia, Lasianthus tomentosus, Achyranthis bidentatae, Alternanthera sessilis, Allamanda carthatica, Alstonia scholaris, Talinum Triangulare, Viola yedoensis, Alternanthera paronychioides, Borreria alata and Hedyotis sp.*

2. Methods

2.1 Proteomics

2.1.1 MALDI-TOF MS and MS/MS

Mass spectrometry was performed on the ABI 4800 MALDI-TOF/TOF system (Applied Biosystems, Framingham, MA, USA). The instrument was equipped with a solid-state laser (diode pumped Nd:YAG laser) pulsing at a

repetition rate of 200 Hz. An 8 mg/mL solution of CHCA in 60% acetonitrile, 0.05% TFA was used as a MALDI matrix. Samples were mixed thoroughly with the matrix solution at the ratio of 1:1 (v/v) and 0.5 µL of the mixture was spotted onto a target plate. The instrument was calibrated externally using a mixture of peptide standards obtained from Sigma- Aldrich (MSCal1).

Both MS and MS/MS spectra were acquired using dual-stage reflectron mirror. The laser intensity was set between 3000-4000. Accelerating voltages applied for MS and MS/MS measurements were 20 and 8 kV, respectively. The mass spectra were accumulated up to 1000 and 5000 shots in MS and MS/MS mode, respectively. In MS/MS mode, collision energy of 1 kV was applied and nitrogen was used as a collision gas in collision-induced dissociation experiments.

2.1.2 High-performance liquid chromatography (HPLC)

HPLC was run on a Shimadzu system equipped with UV detector at 220, 254 and 280 nm. For RP-HPLC, Grace Vydac 250 x 22 mm C18 columns (5 µm particle size and 300 Å pore size) were used for preparative purifications at a flow rate of 6 mL/min; 250 x 10 mm and 250 x 4.6 mm C18 columns were used for semi-preparative and analytical purifications at a flow rate of 3 mL/min and 1 mL/min, respectively. For ion exchange, PolyLC 250 x 9.4 mm and 250 x 4.6 mm polysulfoethyl A columm were used for semi-preparative and analytical purifications at a flow rate of 3 mL/min and 1 mL/min, respectively.

2.1.3 MS profiling of CRP constituents in the plant extracts

Around 100 mg material of each plant species was macerated in liquid nitrogen and extracted with 500 μL of 50% ethanol. The plant debris was removed by filtration. The aqueous extracts were subjected to MALDI-TOF MS to scan for mass signals between the 2-8 kDa range. If the signals were weak, the CRP constituents could be concentrated by using ZipTips purification after diluting the extracts four-fold. Plant extracts showing positive signal in the desired mass range were subjected further for S-reduction and S-alkylation to verify the disulfide content.

2.1.4 Isolation and purification of CRPs

Plant materials of each species were ground using kitchen blenders and extracted with 60% ethanol. After removal of plant debris, the aqueous extracts were partitioned with half volume of DCM for defatting. The ethanol/water fractions were dried in vacuum and redissolved in 10% ethanol. The concentrated extracts were loaded onto flash column packed with C18 media. The column was washed with 20% ACN and eluted with 80% ACN to obtain the CRP-enriched fractions. Eluents were concentrated by rotary evaporator and redissolved in 10% ethanol. Isolation of individual peptides was then achieved through several stages of reversed-phase high-performance liquid chromatography (RP-HPLC). Preparative RP-HPLC was performed at linear gradient from 10% to 70% ACN (250 x 22 mm Grace Vydac column, flow rate 6 mL/min). Semi-preparative RP-HPLC was carried out on same linear gradient of ACN (250 x 10 mm Grace

Vydac column, flow rate 3 mL/min). Strong cation exchange SCX was performed at 60-min linear gradient from 0 mM to 500 mM KCl solution (PolyLC 250 x 9.4 mm column, flow rate 3mL/min). Final purification achieved by analytical RP-HPLC (250 x 4.6 mm column, flow rate 1 mL/min). A general scheme for purification of CRPs in plant was summarized in Figure 9.

2.1.5 S-Reduction and S-Alkylation

Around 20 μg of each peptide was dissolved in 100 μL of 100 mM NH_4HCO_3 buffer, pH 7.8 containing 10 mM DTT. The peptides were incubated for 2 hr at 37°C. Two-fold excess of IAA over the total thiol was added and incubated for 1 hr at 37°C. S-alkylated peptides were purified by RP-HPLC.

2.1.6 Enzymatic digestion and sequence determination

Lyophilized S-alkylated peptides were dissolved in 5 μL of NH_4HCO_3 buffer (50 mM, pH 7.8) and incubated with endoproteinase Glu-C (EndoGlu-C), trypsin or chymotrypsin at a final peptide-to-enzyme ratio of 50:1. The digestions were allowed to proceed for 1 hr. Peptide fragments resulting from the digestions were first examined by MALDI-MS followed by MALDI-MS/MS analysis using a 4800 Proteomics Analyzer MALDI–TOF/TOF mass spectrometer. Primary peptide sequences were obtained by interpreting the b- and y-ions formed during the MS/MS fragmentation. Assignments of isobaric residues Ile/Leu and Lys/Gln were based on enzymatic digestion patterns, quantitative amino acid analysis, cDNA sequences, and homology to known CRPs.

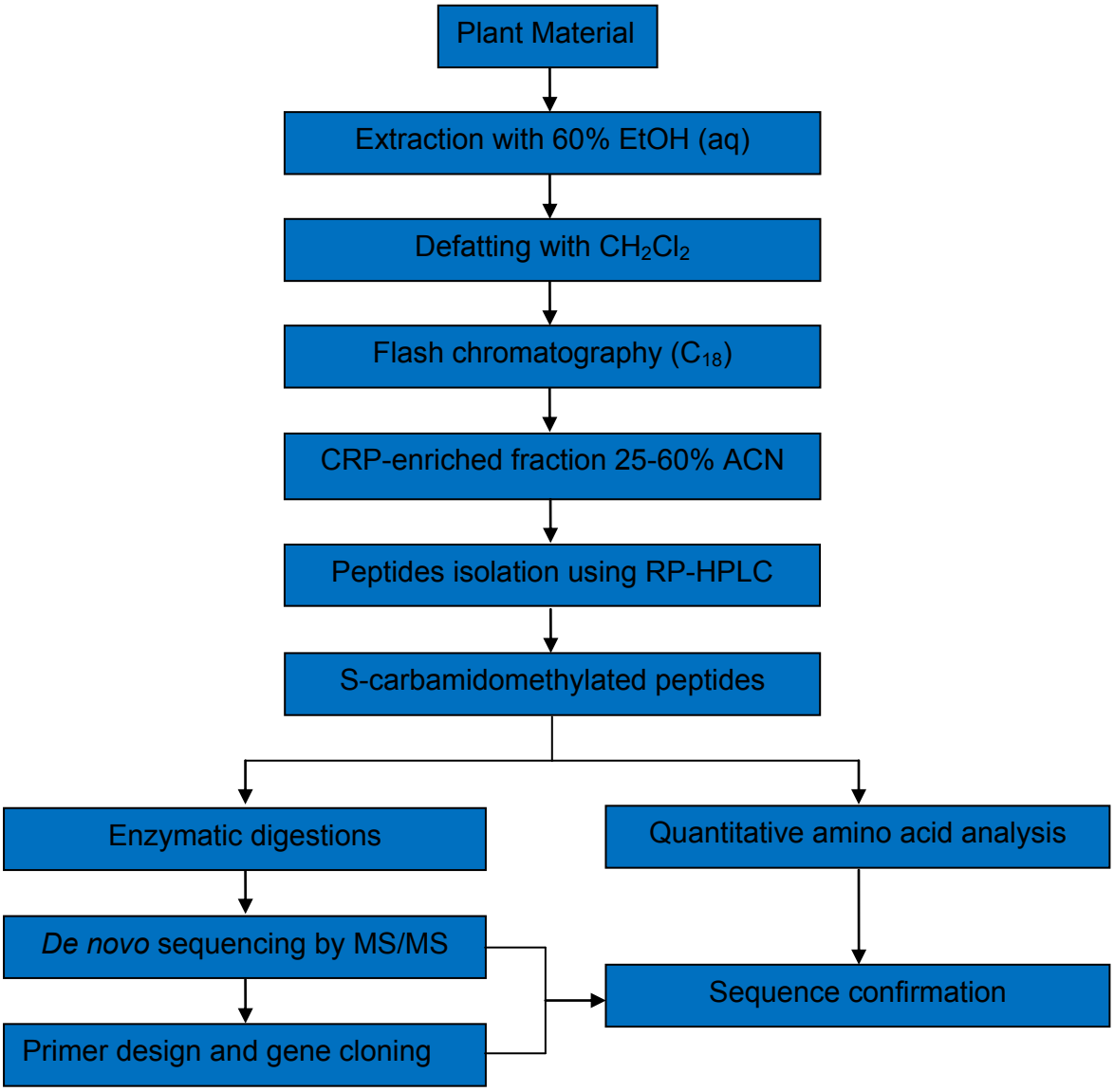


Figure 9. A general scheme for isolation and characterization of plant CRPs.

2.1.7 Disulfide mapping

Around 0.2 mg of each peptide sample was partially reduced in 500 µL of 100 mM citrate buffer pH 3.0, 20 mM TCEP at 37°C for 35 min. Subsequently, NEM (N-ethylmaleimide) powder was added directly to a final concentration of 50 mM and incubated at 37°C for another 15 min. The reaction was quenched by immediate injection of samples on a Vydac C18 column (250 x 4.6 mm) at a flow rate of 1 mL/min. Intermediate species were separated with a linear gradient of 0.3% min^{-1} of 10-60% buffer B and analyzed with MALDI-TOF MS to verify the number of NEM-alkylated cysteines. NEM-alkylated intermediate species were then fully reduced with 20 mM DTT, incubated at 37°C for 60 min. The reduced peptides were S-alkylated with 40 mM IAA, incubated at 37°C for 30 min before stopping the reaction by injection into HPLC. S-alkylated peptides were lyophilized and re-dissolved in 20 µL of 10% ACN, 20mM NH_4HCO_3, pH 7.8 for enzymatic digestions. Fragments obtained were sequenced by MS/MS.

2.1.8 Quantitative amino acid analysis

For hydrolysis, about 10 µg of each peptide sample was added into an insert tube and placed in a reaction vessel from Eldex (Napa, CA, USA). After closing the reaction vials, the solutions were dried under vacuum using an Eldex hydrolysis/derivatization station. To the bottom of each vial, 200 µL of 6 M HCl solution containing 1.0% phenol was added. The vials were sealed under vacuum after three alternating vacuum-nitrogen flushing steps and placed in an oven at 110 ºC for 24 h. After acid hydrolysis, the samples were dried under vacuum.

For derivatization, 20 µL of borate buffer was added to each hydrolyzed sample. Next, 5 µL of reagent solution from the AccQ-Tag Ultra derivatization kit (Waters) was added. The reaction mixture was mixed immediately and heated to 55 ºC for 10 min. After cooling down, an aliquot of the reaction mixture was used for UPLC injection.

UPLC was performed on an Acquity system (Waters), equipped with a 2996 photodiode array detection (DAD) system. The column used was AccQ-Tag Ultra C_{18} column (100 x 2.1, 1.7 µm particle size) from Waters. The flow rate was 0.7 mL/min and the column temperature was kept at 55 ºC. The injection volume was 2 µL and the detection wavelength was set at 260 nm. The solvent system consisted of two eluents: (A) AccQ-Tag Ultra eluent A concentrate (5%, v/v) and water (95%, v/v); (B) AccQ-Tag Ultra eluent B. The following gradient elution was used: 0–0.54 min, 99.9% A–0.1% B; 5.74 min, 90.9% A–9.1% B; 7.74 min, 78.8% A–21.2% B; 8.04 min, 40.4% A–59.6% B; 8.70–10 min, 99.9% A–0.1% B. Empower 2 (Waters) software was used for system control and data acquisition.

2.1.9 Spectrophotometric determination of protein concentration

Absorbances at 280 nm wavelength of peptide solutions prepared in MilliQ water were measured by using NanoPhotometer (Implemen). Peptide concentrations were determined based on the Beer-Lambert Law:

$$A = \varepsilon \cdot l \cdot c$$

where A is absorbance, ε is the molar absorption coefficient ($M^{-1}cm^{-1}$) and *l* is the cell path length (cm). The molar absorption coefficient of a protein at 280 nm wave length was calculated using the following equation [63]:

$$\varepsilon_{280} = (5500 \cdot n\text{Trp}) + (1490 \cdot n\text{Tyr}) + (125 \cdot n\text{SS})$$

where *n*Trp is the number of Trp residues, *n*Tyr is the number of Tyr residues, and *n*SS is the number of disulfide bonds in the protein; and the numbers indicate the the molar absorbances for tryptophan (Trp), tyrosine (Tyr), and cystine, respectively.

2.1.10 Homology modeling

The work is done in the courtesy of Dr. Xiaogang Sui. To build a computer model for a novel CRP, blast search was first performed to identify homologs of known structures. The 3-D structures were built by MODELLER 9v7 (available at http://salilab.org/modeller/) [64] using homolog structures as templates. The generated computer models were evaluated by DOPE, MolPDF and GA341. The structure with lowest energy was used. The structures were analyzed and represented by Pymol (available at http://www.pymol.org/).

2.2 Genomics

2.2.1 RNA extraction

Fresh plant materials were used immediately after collection for extraction of total RNA using the PureLink™ Micro-to-Midi Total RNA Purification System (Invitrogen). Quality and concentration of RNA was assessed by measuring the

A_{260}/A_{280} ratio using Nanophotometer (IMPLEN). RNA was also analyzed by running on agarose gel electrophoresis. RNA was stored at -80°C and used for production of cDNA.

2.2.2 cDNA gene cloning using 3′-RACE and 5′-RACE PCR

cDNA libraries were prepared from total RNA using Smarter™ RACE cDNA Amplification Kit (Takara). First, 3′-RACE PCR was conducted using degenerate primers designed based on MS/MS determined sequences. The PCR products were cloned into pGEM-T Easy vector (Promega) and transformed in to *E. coli* bacterial cells. Plasmids were extracted using Qiagen kit and DNA sequencing was done by 1st Base. Afterward, a series of specific primers against 3′ untranslated region of the newly obtained genetic sequences were used for 5′-RACE PCR. The full transcripts were then assembled from the two partial genes.

2.2.3 DNA extraction and amplification of CRP genes from DNA template

DNA was extracted from fresh plant materials using Pure-Link plant total DNA purification kit (Invitrogen). PCR reactions on DNA templates were then conducted with specific primers designed against 5′ and 3′ untranslated regions of each CRP gene. PCR products were purified, cloned into pGEM-T Easy vector (Promega) and sent for sequencing.

2.2.4 Phylogenetic tree construction

For *H. biflora*, tree was constructed using seven cyclotides precursors including hedyotide B1 and B2 (*H. biflora*), kalata B1 (GenBank: AAL05477.1),

kalata B2 (P58454.1), kalata B3/6 (AAL05478.1) kalata B7 (AAL05478.1), hcf-1 (CB083237.1). The precursor protein sequences were aligned using ClustalW. Phylogenetic tree was visualized by Tree View software.

For *Clitoria ternatea*, phylogenetic tree was constructed using translated precursors of A1bs/leginsulins, cyclotides, and cliotides. A1b/leginsulin sequences used included four from *Pisum sativum* (CAB82859, CAE00465, CAE00467, CAE00468), two from *Medicago truncatula* (AC233561, CAE00461, CAE00462, ACJ85909), one from *Glycine max* (BAA04219), and one from *Glycine soja* (CAA09880). Cyclotide sequences included two from *Viola baoshanensis* (DQ851860, ABI48957), two from *Viola odorata* (P58440, AY630565), one from *Gloeospermum blakeanum* (GQ438777), and one from *Hedyotis centranthoides* (FJ211188). Cliotide precursors of cliotide T1, T2, T3, T7, and T11 were used. The precursor protein sequences were aligned using COBALT multiple alignment [65]. Phylogenetic tree was generated using fast minimum evolution method and visualized by Tree View software [66].

2.3 Bioassays

2.3.1 Antimicrobial assay

A sensitive two-stage radial diffusion assay [67] was conducted to test the antimicrobial activities of isolated CRPs against various microbial strains. An aliquot containing 4 x 10^6 colony forming units was mixed with 10 mL of molten underlay gel (composed of 10mM phosphate buffer saline (PBS), 1% trypticase soy broth (TSB) and 1% agarose) and poured into a round Petri dish (Greiner Bio One) to form a uniform layer. After the gel solidified, six wells were punched in an

evenly spaced array using a 2.5-mm diameter gel puncher. An aliquot of 3.5 µL of a serial two-fold dilution of each CRP at six different concentrations was added to each well. Each peptide concentration was tested in duplicate. D4R (a synthetic antimicrobial peptide) was used as positive control [68]. The dishes were incubated at 37^{o}C for 3 hours to allow the diffusion of the peptides into the underlay gel. Each underlay gel was then covered with 10 mL of nutrient-rich overlay gel (containing 6% TSB (w/v) and 1% agarose). The dishes were incubated at 37^{o}C overnight and the diameters of the clear zones surrounding the wells were measured (not including the diameter of the well). Zone diameters were then converted to units (0.1 mm = 1 unit) and plotted against log of concentrations. The Minimal Inhibitory Concentrations (MICs) were determined from the x intercept of the dose-response curve.

2.3.2 Cytotoxicity assay

Trypsinized HeLa cells or A549 cells were counted using C-chip DHC-N01 (iNCYTO), and diluted in RPMI media to a density of 20,000 cells/mL. 100 µL of this cell suspension (2,000 cells) was loaded into each well of a 96-well microtiter plate (Nunc) and incubated for 24 hr. Each tested CRP was serially diluted in PBS, and 5 µL was dispensed into each well. After incubation for 72 hours, 20 µL of 5 mg/mL Thiazolyl Blue Tetrazolium Bromide (MTT) solution was added into each well. The plate was incubated for 3.5 hr, 37°C. The media was removed and 150 µL of dimethyl sulfoxide was added for cell lysis. The absorbance at 590 nm was measured with a reference filter at 620 nm using an automatic plate reader

(Safire^{2TM} Tecan). Survival Index (SI) was calculated as percentage of growth inhibition in comparison with the control wells (added PBS only). Cyclotides' cytotoxicity was indicated by IC_{50} values (concentration that gives an SI of 50%).

2.3.3 Hemolytic assay

Blood type A was taken from a healthy volunteer. Red blood cells (RBCs) were collected by centrifugation at 1,000 rpm, washed three times with PBS and re-suspended in PBS to give a final 1% suspension. 100 µL of this suspension was added to each well of a 96-well microtiter plate (Nunc). Each tested CRP was dissolved in water and serially diluted in PBS, and then 5 µL was dispensed into each well containing the RBCs suspension. Each peptide concentration was tested in triplicate. The plate was incubated at 37 °C for 1 hr and centrifuged at 1,000 rpm for 6 min. Aliquots of 60 µL were transferred to a new 96 well U-bottomed microtiter plate (Nunc).The absorbance was measured at 415 nm using an automatic plate reader Safire^{2TM} Tecan. The level of hemolysis was calculated as the percentage of maximum lysis (1% Triton X-100 as positive control) after adjusting for minimum lysis (PBS as negative control). The cyclotide concentration causing 50% hemolysis (HD_{50}) was calculated.

2.3.4 Immunomodulation assay

THP-1 or RAW cells were suspended in RPMI medium at the density of 1 x 10^6 cells/mL and placed in six-well culture plates. The cells were treated with various peptide concentrations for 6 hours with or without 2-hour pretreatment with lipopolysaccharides (LPS). The supernatants were collected by

centrifugation at 1,000 rpm and stored at -20^{o}C until assayed. The concentrations of tumor necrosis factor-alpha (TNF-α), interleukin-6 (IL-6), and interleukin-8 (IL-8) were determined by enzyme-linked immunosorbent assay (ELISA) using ELISA MAX™ Deluxe Sets (BioLegend, USA).

2.4 Stability assays

2.4.1 Heat stability

Peptide solutions dissolved in MilliQ water (0.2 mg/mL) was incubated at 100^{o}C for one to two hr to examine their thermal stability. At specific times, aliquots of 25 μL peptide solutions were removed and monitored by analytical RP-HPLC. The percentage of peptides remaining after heating was calculated based on their relative area on the RP-HPLC profiles as compared as to those without heating.

2.4.2 Enzymatic stability

Peptides were prepared in appropriate buffers at the final concentration of 0.2 mg/mL. Trypsin, endoproteinase Glu-C (EndoGlu-C) and pepsin were prepared as 0.5 mg/mL stock solutions. These enzymes were diluted in appropriate buffers and added to the peptide solutions to give a final enzyme to peptide ratio of 1:50 (w/w). Digestion by trypsin and EndoGluC was carried out in 100 mM NH_4HCO_3 buffer, pH 7.8. Digestion by pepsin was carried out in 100 mM sodium citrate buffer, pH 2.5. All the digestions were conducted at 37 ^{o}C. At specific time intervals, aliquots of 25 μL peptide solutions were removed and

immediately monitored by analytical RP-HPLC. The relative areas of peptides after digestions based on the RP-HPLC profiles were compared to those without enzymatic treatments to examine their stability. A linear synthetic peptide with primary sequence as TAYGRKFFSL was used as control.

2.4.3 Acid stability

Peptides were prepared in 0.2 M HCl solution at the final concentration of 0.2 mg/mL. They were incubated at 37^{o}C. At specific time intervals, the reactions were stopped and immediately monitored by RP-HPLC to examine for the acid stability.

Chapter 3

Characterization of Novel Cyclotides and Uncyclotides from *Hedyotis/Oldenlandia Species*

1. Introduction

Hedyotis and *Oldenlandia* are closely related genera of the Spermacoceae tribe (Rubiaceae family) distributed mainly in tropical or subtropical regions. They comprise of about 400 species of annual and perennial herbs with a few being shrubs or trees [69]. The scientific names of these two genera *Oldenlandia* and *Hedyotis* have been used interchangeably and their generic delimitation has caused confusion ever since they were first proposed by Linnaeus in 1753 [70]. Some taxonomists favor the recognition of *Hedyotis* and *Oldenlandia* as distinct genera [69], whereas others merge *Oldenlandia* into *Hedyotis* as a single genus [69]. For simplicity, I will follow the latter and accept *Hedyotis* as the name of the combined genus.

In traditional medicine, over 20 *Hedyotis* species have been used for treatment of diseases and in healing practices [71]. The most popular among these are *H. diffusa* and *H. corymbosa* which are active principles in several Chinese remedies such as *bai hua she she cao*, *peh hue juwa chi cao* and *feibao* syrup. They are taken for treatment of cancers, infections and other diseases. Phytochemical investigation of *Hedyotis* species was first published in 1933 upon examining the active components of the medicinal plant *H. auricularia* [71]. Since then, over 50 novel compounds have been isolated from various member of the

Hedyotis genus. These compounds have highly divergent structures comprising of alkaloids, anthraquinones, flavonoids, iridoids, triterpenoids, sterols, lignans and a number of other compounds [71].

The prototypic cyclotide, kalata B1, is the first CRP reported from *Hedyotis* genus by Lorent Gran in the 1970s [16] when investigating the active principles of *Oldenlandia affinis* (syn *Hedyotis affinis*) responsible for accelerating contractions and child birth. Subsequent studies of this medicinal plant have led to the discovery of more than 20 novel cyclotides with diverse primary structures [72]. However, similar study to examine the cyclotide constituents in other medicinal *Hedyotis* species was not performed. With several local *Hedyotis* species having long history of medical usages, they constitute a valuable treasure trove of druggable cyclotides awaiting discovery.

In this chapter, I report my work on the isolation and characterization of novel cyclotides and CRPs from important *Hedyotis* species commonly used in traditional medicine. They include *H. affinis*, *H. biflora, H. diffusa*, *H. corymbosa*, *H. herbacea*, *H. pinifolia*, *H. congesta* and *H. auricularia*. Screening of these species resulted in isolation of novel cyclotides from *H. affinis*, *H. biflora* and *H. diffusa,* and unexpectedly many of them have open-end structures. The expression level of cyclotides varies greatly among different species with more than 12,000-fold difference between *O. affinis* and *H. diffusa.* Novel cyclotides also display tissue-specificity and geographical differences. These findings provide a better understanding about cyclotide constituents in medicinally important *Hedyotis* species and their potential roles as active principles in traditional medicines.

2. Results

2.1 Cyclotide screening of medicinally important *Hedyotis* species

Seven medicinally important *Hedyotis* species was screened for the presence of cyclotides. These included *H. biflora, H. corymbosa*, *H. diffusa, H. herbacea*, *H. pinifolia*, *H. congesta* and *H. auricularia*. The occurrence of cyclotides was determined based on three main criteria: hydrophobicity (desorbed from C18 at 25-55% ACN), mass range (2500 to 4000 Da) and disulfide content (three cystine bonds) [33]. Around 500 mg plant material of each species was homogenized in 2 mL ethanol/water (1:1, v/v). The extracts were diluted five-fold and subjected to C18 solid phase extraction (SPE). Small molecules and polar substances that potentially interfere with the mass spectrometry signals were removed by washing the SPE column with 20% ACN. Cyclotides typically did not elute under this condition due to their hydrophobicity, and usually desorbed from SPE column between 25 to 55% ACN. To ensure complete desorption, the putative cyclotide-containing fraction was eluted with 80% ACN. This fraction was profiled and fingerprinted by MALDI-TOF MS.

Plant extracts of *H. biflora* and *H. diffusa* showed positive signals in the desired mass range indicative of cyclotides (Fig. 10). The remaining *Hedyotis* species gave negative result under our screening procedure. It is possible that cyclotides in these species may occur in low-abundant level that escaped our detection. The disulfide content of these putative cyclotide compounds was then examined by comparing the mass difference before and after S-alkylation with IAA. Each S-alkylated half-cystine residue caused a mass increase of 58 Da.

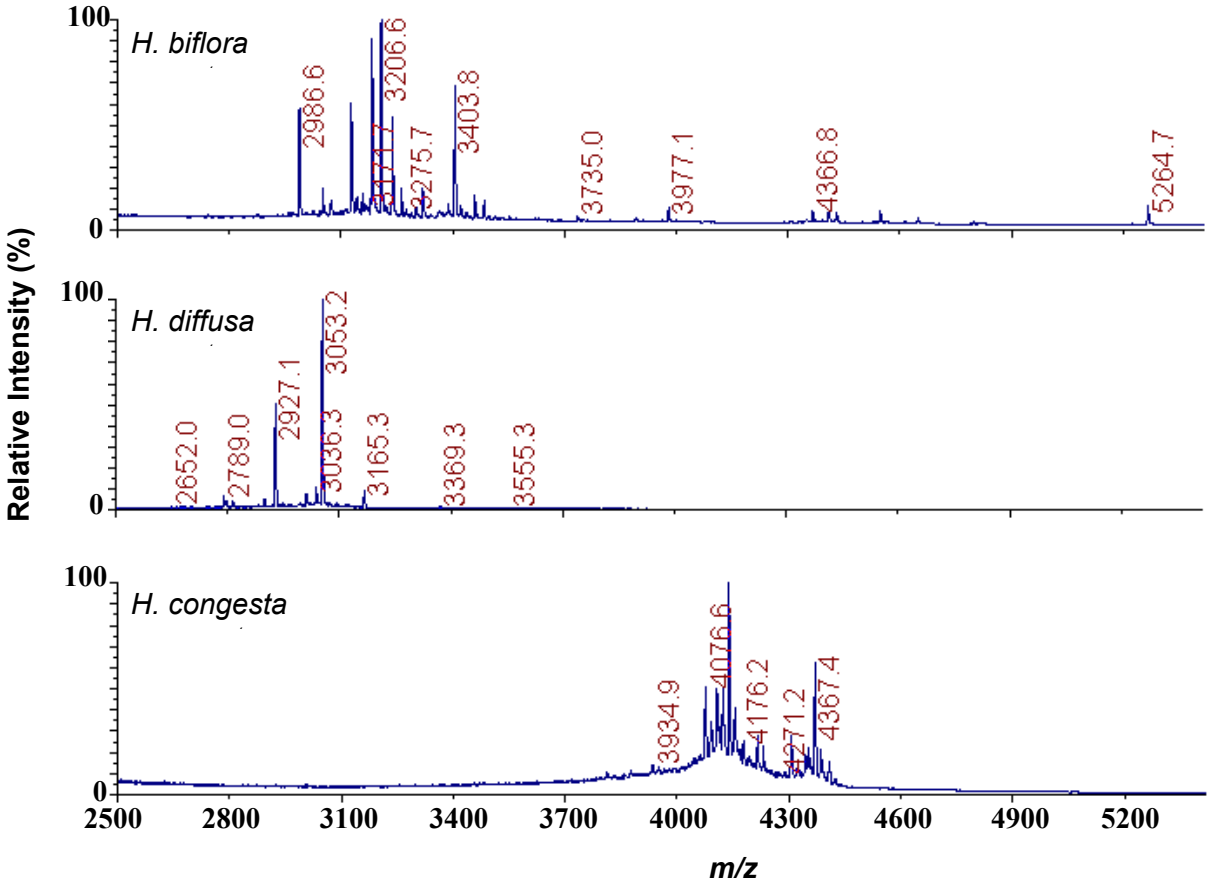


Figure 10. MS profiling of CRPs in *H. biflora, H. diffusa* and *H. congesta. H. biflora* and *H. diffusa* show cluster of peptide masses around 3000 Da region characteristic of cyclotides. *H. congesta*, on other hand, shows cluster of peptides around 4 kDa region indicative of defensin-like compounds

Most of these compounds displayed a mass shift of 348 Da indicating the presence of three cystine bonds and suggestive the occurrence of cyclotides. It is interesting to note that the extracts of *H. biflora* and *H. congesta* showed several peptide masses at 4000 to 5000 Da range. These peptides contain four disulfide bonds indicated by a mass shift of 464 Da. This demonstrated the presence of other CRPs beside cyclotides in these species.

2.2 Plant extraction and isolation of individual cyclotides

To provide sufficient samples for sequence determination, larger scale extraction was carried out as described in Materials and Methods. Briefly, plant samples (~1 kg) were grounded and extracted three times with 60% EtOH and partition with DCM for defatting. The ethanol/water fraction was dried in vacuum and dissolved in 200 mL of 10% ethanol. This concentrated extract was subjected to flash column packed with 100 g of C18 media. The column was washed with 20% ACN and eluted with 80% ACN to obtain the cyclotide-enriched fraction. Isolation of individual cyclotides was then achieved by repetitive RP-HPLC.

2.3 Sequence determination by MALDI-CID-MS/MS

Primary structure determination of cyclotides replies on ESI-CID-MS/MS with a few examples by MALDI-CID-MS/MS. Although the later has several advantages over the former [73], it has yet to receive a widespread acceptance. In this work, I showed that MALDI-CID-MS/MS allowed rapid and complete sequence determination of cyclotides. In either approach, sequencing of cyclotides requires breakage of their disulfides and linearization of the cyclic backbone. Since complete sequence coverage is rarely obtained for peptide

larger than 2 kDa, cyclotides after S-reduction and S-alkylation are often cleaved into smaller fragment suitable for MS/MS [74].

To examine whether peptides larger than 2 kDa can be sequenced directly with MALDI-CID-MS/MS, a putative cyclotide with mass of 3156 Da isolated from *H. biflora* (later designated as hedyotide B7) was S-reduced with DTT and S-alkylated with IAA. The S-alkylation was confirmed by a mass shift of 348 Da and gave rise to a major peak at *m/z* 3504 Da. The S-alkylated peptide was subjected to tryptic digestion yielding a single fragment with *m/z* at 3522 Da. The mass increase of 18 Da after tryptic digestion suggested the hydrolysis of a cyclic amide backbone.

The 3522-Da-peptide was then subjected directly to tandem mass spectrometry without any additional enzymatic digestion to generate smaller fragments. I found that this peptide was effectively fragmented by MALDI-CID-MS/MS (Fig. 11A). *De novo* interpretation of this spectrum gave complete primary sequence of hedyotide B7. Full sequence coverage of a 30-residues-cyclotide was thus obtained in a single MS/MS spectrum. Blast search revealed a high sequence homology (70% identity) between hedyotide B7 and kalata B5 from *O. affinis* confirming that it was a new member of the cyclotide family.

For comparison with tryptic digestion which gave a positive charged Arg/Lys at the C-terminus, I also fragmented the S-alkylated hedyotide B7 with Glu-C to determine the effect of C-terminal residue's charge state on the MS/MS fragmentation efficiency. Glu-C digestion yielded a single fragment with the same *m/z* at 3522 Da as the tryptic digested peptide but retaining Glu residue with a negative charge at the C-terminus.

As shown in Fig. 11B, the fragmentation of Glu-C digested peptide was less efficient and resulted in incomplete sequence coverage. Fragment ions from

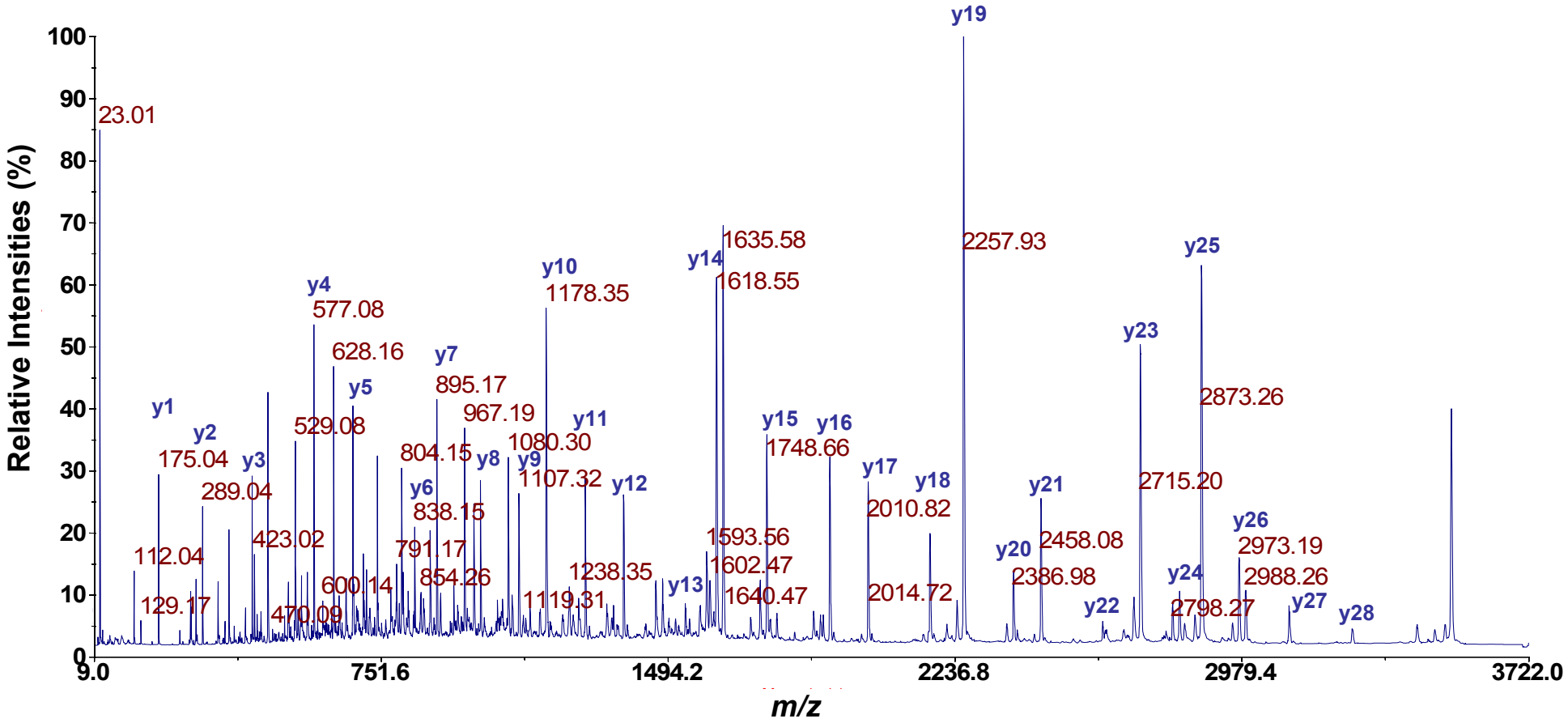
y1 to y11 and the corresponding b-series ions were missing due to the suppression effect of C-terminal acidic residue and the absent of basic residue for charge retention. More intense fragment ions from y12 to y29 were observed due to the present of internal Arg residue resulted from Glu-C digestion. *De novo* interpretation of the MS/MS spectrum provide sequence coverage of 19 residues from y12 to y29, whereas the sequence information of 11 residues correspond to y1 to y11 was missing. Tryptic digestion thus has a clear advantage as compare to Glu-C digestion for sequencing of high mass peptides.

2.4 Sequence verification by amino acid analysis and enzymatic digestion

An inherent limitation of *de novo* sequencing approach using MS/MS is its inability to distinguish isobaric residues, i.e. Ile/Leu and Gln/Lys. To confirm the novel sequences, I used a combination of enzymatic digestions and quantitative amino acid analysis. The first approach was facilitated by proteases that cleave the peptide bond after a specific residue. Trypsin cleaves sequences after Lys or Arg but not Gln residue and thus was used for Lys/Gln assignment. Similarly, chymotrypsin was used for distinguishing Ile and Leu. Chymotrypsin preferably cuts after Phe, Trp, Tyr, Met and Leu but not Ile. The Ile/Leu assignment, however, is often complicated as they are usually followed by a Pro residue in the cyclotide sequences. The X-Pro peptide bonds are known to resist against degradation by most proteolytic enzymes. Under these circumstances, additional information from quantitative amino acid analysis to provide the total number of Ile/Leu residues in combination with sequence homology to known cyclotides are usually sufficient for the Ile/Leu assignment.

A

b-ions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	V	C	Y	L	N	G	T	P	C	A	E	S	C	V	Y	L	P	C	V	T	A	V	I	G	C	T	C	Q	N	R
y-ions	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1



B

b-ions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	S	C	V	Y	L	P	C	V	T	A	V	I	G	C	T	C	Q	N	R	V	C	Y	L	N	G	T	P	C	A	E
y-ions	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

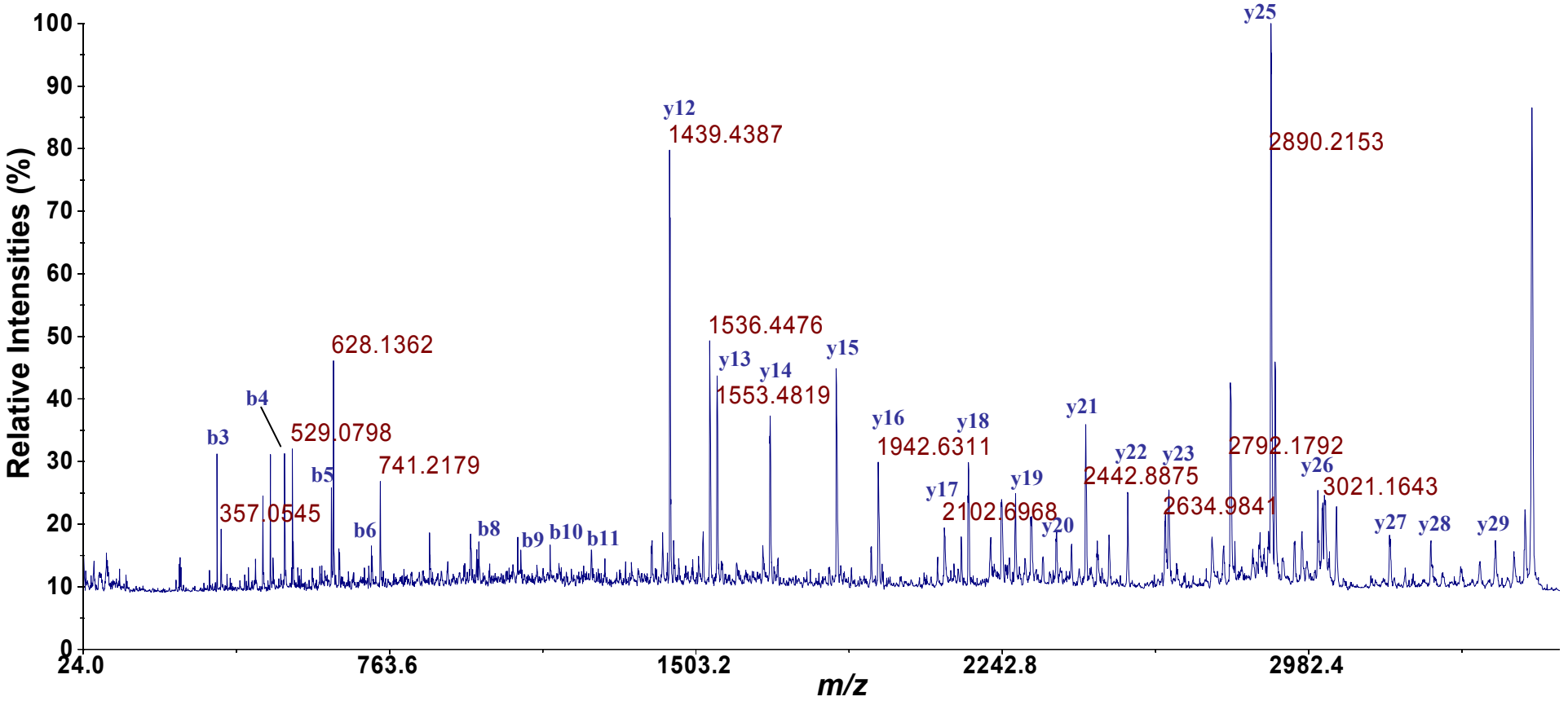


Figure 11. MS/MS spectrum of linearized S-alkylated hedyotide B7 upon digestion with (A) trypsin and (B) endoproteinase Glu-C. Complete sequence coverage of 30 residue cyclotides was obtained in single spectrum.

2.5 Novel cyclotides in *Hedyotis biflora*

Exhaustive fingerprinting of cyclotides in *H. biflora* led to the discovery of thirty six novel peptides comprising of 21 cyclotides and 15 uncyclotides (Table 3). The total yield of cyclotides and uncyclotides in *H. biflora* was estimated to be around 0.5-0.8 mg/kg. It should be noted that plant samples were collected in two different regions, Viet Nam and Singapore, to increase the number of detected cyclotides and to study the geographical variation of cyclotides expression. All 21 novel cyclotides are of bracelet subfamily, and the 15 novel uncyclotides can be subdivided into bracelet-like and hybrid-like subfamilies dependent on their sequence homology. The discovery of a large number of uncyclotides in a single species was unexpected due to the novelty of uncyclotides, with only two naturally occurring examples being reported. This suggests that their occurrence may be more common than previously thought. In addition, my work also provides the first description of bracelet and hybrid uncyclotides as all known examples of uncyclotides are of Möbius subfamily.

Table 3. Novel cyclotides and uncyclotides in Hedyotis biflora.

Peptide	Sequence	Residues	MW[a]	Sample[b]	Plant Parts
Loop	6 1 2 3 4 5 6				
B1	G--TRCGETCFVLPC-WSAKFGCYCQKGF--CYRN	30	3403	V/S	Aerial
B2*	IQCGESCVWIPC-ISSAWGCSCKNKI--CSS	28	2986	V/S	Aerial
B3	G--ISCAETCVLIPC-ISSVIGCTCQNKR--CYKN	30	3181	V/S	Root
B4	IPCGESCAFIPC-LTSLLGCTCQNKV--CYRDE	30	3260	V/S	Root
B5	GGAVPCGETCVYIPC-ITAAIGCSCQNNV--CYHN	32	3236	V/S	Root
B6	G--TPCAETCIFIPCTVTALIGCSCQNQI--CYKN	31	3270	V	Root
B7	G--TPCAESCVYIPC-VTAVLGCTCQNRV--CYLN	30	3156	V/S	Root
B8	G--TPCAESCVYIPCTVTALIGCSCQNQI--CYKN	31	3258	V	Root
B9	G-AVPCGETCVYIPC-ITAAIGCSCQNNV--CYHN	32	3179	V/S	Root
B10*	REECGETCYILPC-VT--PDCICSGGQ--CYKIE	29	3206	S	Root
B11*	EECGETCYILPC-VT--PDCICSGGQ--CYKIE	28	3050	S	Root
B12	GG-VPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	34	3440	V	Root
B13	G--IPCGESCAFIPC-LTSLLGCTCQNKV--CYRD	30	3170	V/S	Root
B14	G--VPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	34	3383	V	Root
B15*	G--IPCGESCAFIPC-LTSLLGCTCQNKV--CYRDE	30	3317	V/S	Root
B16*	AVPCGETCVYIPC-ITAAIGCSCQNNV--CYHN	32	3140	V/S	Root
B17	G--NPCGESCVYIPC-ITTVVGCSCQNSV--CYHN	30	3126	V/S	Root
B18[c]	GGAVPCGETCVYIPC-ITAAIGCSCQNKV--CYRD	32	3270	S	Root
B19[c]	G--IPCGESCVIIPC-ISTVIGCTCENKV--CLRN	30	3100	S	Root
B20[c]	GGAVPCGESCVYIPC-ITTVVGCSCQNSV--CYHN	32	3239	S	Root
B21[c]	G--TPCGESCVFIPC-VTTVIGCTCQNRV--CYLN	30	3156	S	Root
B22	G--IPCGESCIFIPC-FTAALGCSCQNKV--CYRN	30	3173	V	Root
B23	G--ISCGESCVFIPC-ISTVLGCSCQNKV--CYKN	30	3131	V	Root
B24	G-TIPCGESCIFIPC-ISTVIGCSCKNNV--CYRN	31	3270	V	Root
B25	G--TSCGEGCVFIPC-ITAVIGCSCQNKV--CYLN	30	3058	V	Root
B26*	QCGESCVWIPC-ISSAWGCSCKNKI--CSS	27	2873	V/S	Aerial
B27*	Q⁺CGESCVWIPC-ISSAWGCSCKNKI--CSS	27	2856	V/S	Aerial
B28[c]	FIPC-VTTVIGCTCSQNKL-CYLN	n.a	n.a	S	Aerial
B29*	VPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	34	3344	V	Root
B30*	NPCGESCVYIPC-ITTVVGCSCQNSV--CYHN	30	3087	V/S	Root
B31*	Q⁺ECGETCYILPC-VT--PDCICSGGQ--CYKIE	28	3032	S	Root
B32*	ECGETCYILPC-VT--PDCICSGGQ--CYKIE	27	2921	S	Root
B33*	G--TRCGETCFVLPC-WSAKFGCYCQKGF--CYRN	30	3421	V/S	Aerial
B34*	G--TRCGETCFVLPC-WSAKFGCYCQKGF--CYRNE	30	3550	V/S	Aerial
B35*	TRCGETCFVLPC-WSAKFGCYCQKGF--CYRN	30	3364	V/S	Aerial
B36*	TRCGETCFVLPC-WSAKFGCYCQKGF--CYRNE	30	3493	V/S	Aerial

[a] Molecular weights (Da). [b] Specimens collected from Vietnam (V) or Singapore (S). [c] Putative cyclotides predicted from cDNA and DNA sequences. Basic residues (R, K, H) are highlighted in cyan, acidic residues (D, E) are in red. Asterisk indicates uncyclotides. Q^+ is pyroglutamic acid.

2.6 Novel uncyclotides from *H. diffusa*

MS profile of *H. diffusa* extracts demonstrated the expression of several unique masses corresponding to cyclotides. Around 1 kg plant samples were collected and extracted as described in Materials and Methods. The total amount of cyclotides produced in *H. diffusa* was about ten-fold lower than *H. biflora* with the approximate yield around 0.1 mg/Kg. This low expression level complicated the isolation and purification of cyclotides from this species.

Eleven novel peptides was purified and sequenced by MS/MS (Table 4). They were designated as hedyotide D1 to D11. Surprisingly, despite sharing high sequence homology with known cyclotides, none of them possess a cyclic backbone and are thus classified as uncyclotides. These novel uncyclotides can be divided into two groups: Möbius-like (hedyotide D1-D5) and bracelet-like uncyclotides (hedyotide D6-11) based on sequence homology. The former are characteristic by a pyroglutamyl residue at the N-terminus instead of the highly conserved Gly as conventional cyclotides. They have Möbius-like character retaining the highly conserved cis-Pro residue on loop 5 (hedyotide D1 and D2) which often causes a conceptual twist on the circular backbone. Interestingly, this conserved proline turn is replaced by a Gly turn in hedyotide D3, D4 and D5, which also allows a sharp turn on the peptide backbone. The later comprise of hedyotide D6 to D11 which share a high sequence homology among each others. Hedyotide D6 and D7 appear to be truncated versions of hedyotide D8. These three peptides probably originate from the same precursor. Similarly, hedyotide D9, 10 and D11 may also derive from a single precursor protein. They have bracelet-like character with the exception of loop 5 which contains an unusual sequence DTTTYD. This loop is formed entirely by hydroxyl-bearing residues and does not share homology to any known cyclotides.

Table 4. Summary of uncyclotides in *Hedyotis diffusa*

Uncyclotides	Amino acid sequence [a]	MW (Da)[b]
Möbius		
hD1	Q*LPICGETCVLGTCYTPG--CRCQYPI--CVR	3053
hD2	Q*LPICGETCVLGRCYTPN--CRCQYPI--CVR	3165
hD3	Q*AF-CGETCILGKCYTPG--CSCHTGI--CIK	2813
hD4	Q*AF-CAETCILGKCYTPG--CSCHTGI--CIK	2827
hD5	Q*AF-CGETCILGTCYTPG--CRCTAGI--CIK	2789
Bracelet		
hD6	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKS	3369
hD7	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKSV	3468
hD8	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKSVS	3555
hD9	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKI	3581
hD10	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKIV	3494
hD11	VG-CYERCVWGPCISKIVGCSCDTTTYDCVKIVS	3395

(a) Isobaric residues Ile/Leu were assigned based on chymotryptic digestion and sequence homology. (b) Molecular weights are reported as monoisotopic mass.

2.7 A novel regional specific cyclotide from *Oldenlandia affinis*

Examination the MS profile of *O. affinis* extract identified a novel cyclotide with *m/z* at 3014 Da. This cyclotide ranked third in abundance based on RP-HPLC and MS profiles (Fig. 12). Interestingly, it appeared to be regional specific as its molecular weight was not reported in the African specimen by Craik *et al.* despite their extensive efforts to fingerprint the cyclotide constituents in *O. affinis* [60]. This novel cyclotide was found to expressed in both aerial and root tissues. Its primary sequence was determined by MALDI-TOF MS/MS and confirmed by quantitative amino acid analysis. This novel cyclotides was named hedyotide A1, with hedyotide A stemmed from the species name *Hedyotis affinis* (syn *O. affinis*). Hedyotide A1 belongs to Möbius subfamily with the conserved cis-Pro residue in loop 5.

2.8 Novel defensin-like-peptides from *Hedyotis biflora*

Re-analyzing the MS profiles of *H. biflora* extracts at a broader mass range (2-10 kDa) revealed a wide variety of other CRPs resembling plant defensins. To my knowledge no defensin-like peptide has been reported in cyclotide-producing plants. These peptides are more hydrophilic than cyclotides and eluted between 25 to 35 % ACN. Most of them contained four disulfide bonds as indicated by S-alkylation with iodoacetamide resulted in a mass increase of 464 Da. This implied that a single plant may produce several different types of CRPs to diversify their defense mechanism or serving for other endogenous functions.

Two novel peptides with *m/z* at 4544 and 4645 Da, designated as HBDL-1 and HBDL-2 (*Hedyotide biflora* defensin like), were isolated and sequenced by MS/MS. Their primary structures were confirmed by Edman degradation (Table 5). HBDL-1 has identical sequence to HBDL-2 except one extra residue at the N-terminus. They are probably derived from the same precursor protein.

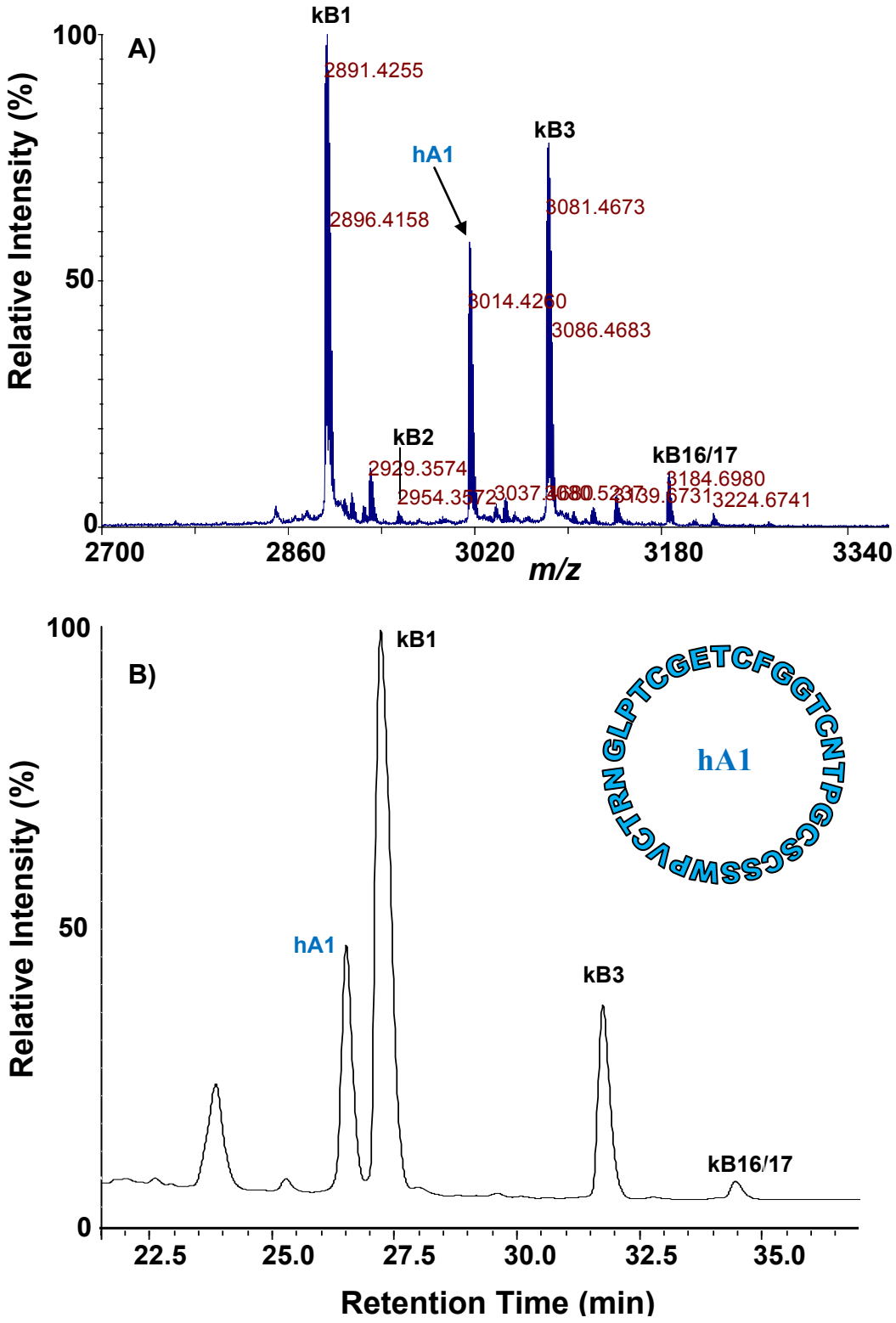


Figure 12. Profiling of crude extract from *Oldenlandia affinis* root tissues. (A) The MS profile shows kalata B1, B3 and hedyotide A1 as three most abundant peptides. (B) RP-HPLC trace shows retention times and relative intensities of different cyclotides. Arrows indicate the novel cyclotide, hedyotide A1. Its primary sequence was shown in circular form.

In addition, the Met-oxidized derivative of HBDL-2 (HBDL-2_{ox}) was also identified indicated by a mass increase of 16 Da. HBDL-1, HBDL-2 and HBDL-2_{ox} all contain four disulfide bonds with a unique cysteine pattern entirely different from plant defensins or any known CRPs. They may thus represent a novel group of CRP with unknown functions. It is interesting to note that HBDL-2 starts with an acidic residue Asp on the N-terminus and ending with a basic residue His on the C-terminus. The positive and negative charge on each end may help to bring its N- and C-termini close together by electrostatic interaction, possibly forming a pseudocyclic structure.

Table 5. Defensin-like peptides from *Hedyotis bilfora*

Peptides	Sequence	MW
HBDL1	**TDKCIPYGGNCAGMTTSCCASCYCDIRFIDGKPTPVCDGYQGCH**	**4645**
HBDL2	**DKCIPYGGNCAGMTTSCCASCYCDIRFIDGKPTPVCDGYQGCH**	**4544**
HBDL2$_{ox}$	**DKCIPYGGNCAGMTTSCCASCYCDIRFIDGKPTPVCDGYQGCH**	**4560**
PSBTN$_f$	**DVAEPKPGTTEAKKKYAPVCVTMPTAKICHK**	**3340**

Molecular weights are reported as monoisotopic mass. Basic residues (R,K,H) are colored in cyan, acidic residues (D,E) are in red. Oxidized Met residue is highlighted in gray and Cys residues are in yellow

Another hydrophilic peptide with *m/z* at 3340 Da peptide was also successful sequenced by MS/MS. This peptide is highly basic (+4 net charge) and contains a single disulfide bond. Blast search revealed a high sequence homology to

chloroplast photosystem protein PSBTN, which was highly conserved in several plant species. This suggested that this peptide may be a fragmented derivative of PSBTN protein. It was thus named $PSBTN_f$ (PSBTN fragment).

There are still many defensin-like peptides in *H. biflora* yet to be sequenced. Since most of them are quite large, sequencing using MS/MS is more challenging in comparison with cyclotides. An alternative method for determining novel sequences is by genetic approach. Edman degradation can be used to obtain sequence information for the first 10 or 15 residues, which can be used for designing degenerative primers to clone the whole encoding genes. A combination with genetic approach offers several advantages because much less amount of plant samples is needed and there is a possibility of cloning other closely related low abundant peptides that share the conserve precursor sequences.

2.9 Cloning of hedyotide encoding cDNAs from *H. biflora*

This part of work was performed in collaboration with Mr Zhang Sen and Dr Wang Wei. To clone the hedyotide encoding cDNAs from *H. biflora*, degenerate primers targeting loop 1 sequences CGETC and CGESC were used for 3′-RACE amplifications. This gave the DNA products of around 450 bp in the agarose gel, which were extracted and cloned into the TOPO vector. After transformation, thirty bacterial colonies were picked up randomLy for analysis. This resulted in the identification of two unique partial clones, which encoded for hedyotide B1 and B2, respectively. Reverse specific primers based on sequence of each unique clone were then used for 5′-RACE amplification. After PCR reactions, two bands of approximately 300 bp were obtained. Sequencing of these two bands gave the remaining sequences of hedyotide B1 and B2 precursors. Their full-length clones were designated as *HBC1* and *HBC2* (*Hedyotis biflora* clone). They have similar

overall arrangement with the *oak1* and *hcf-1* clones from *O. affinis and H. centranthoides,* which each contains an ER signal sequence, N-terminal propeptide, cyclotide domain and a hydrophobic tail (Fig. 13).

To facilitate rapid discovery of new cyclotide sequences without need for 5′-RACE amplification, sequence alignments were carried out to identify the conserved elements in the upstream region. Two conserved sequences were found with one located at the ER region (AACVGVL) and the other at the NTR region (AAEIQFKKE). These two sequences were targeted as forward primers for cDNA gene cloning. Another 11 unique cyclotide genes were amplified named HBC3-HBC13. Four of them encoded for novel cyclotides that yet to be characterized at protein level. In addition, full length clone of *HBC3, HBC4, HBC6, HBC8* and *HBC12* were completed by 5′-RACE. The predicted amino acid sequences from all the clones were summarized in Fig. 14.

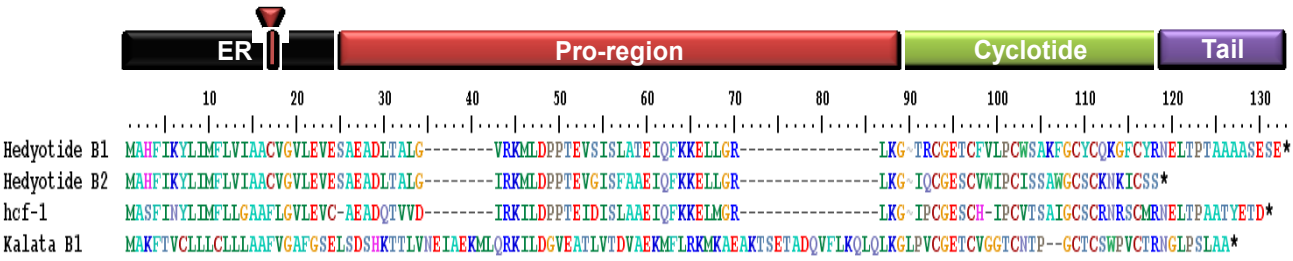


Figure 13. Genetic organization of hedyotide precursors. Their primary sequences were deduced from cDNA clones and aligned with hcf-1 and kalata B1 precursor (OaK1) using Clustal W. ER signal is predicted by SignalP-HMM software. They share overall arrangement containing an ER signal sequence, a pro-region, mature cyclotide and a short tail domain. The inverted triangle indicates the intron location. Asterisk indicates stop codon.

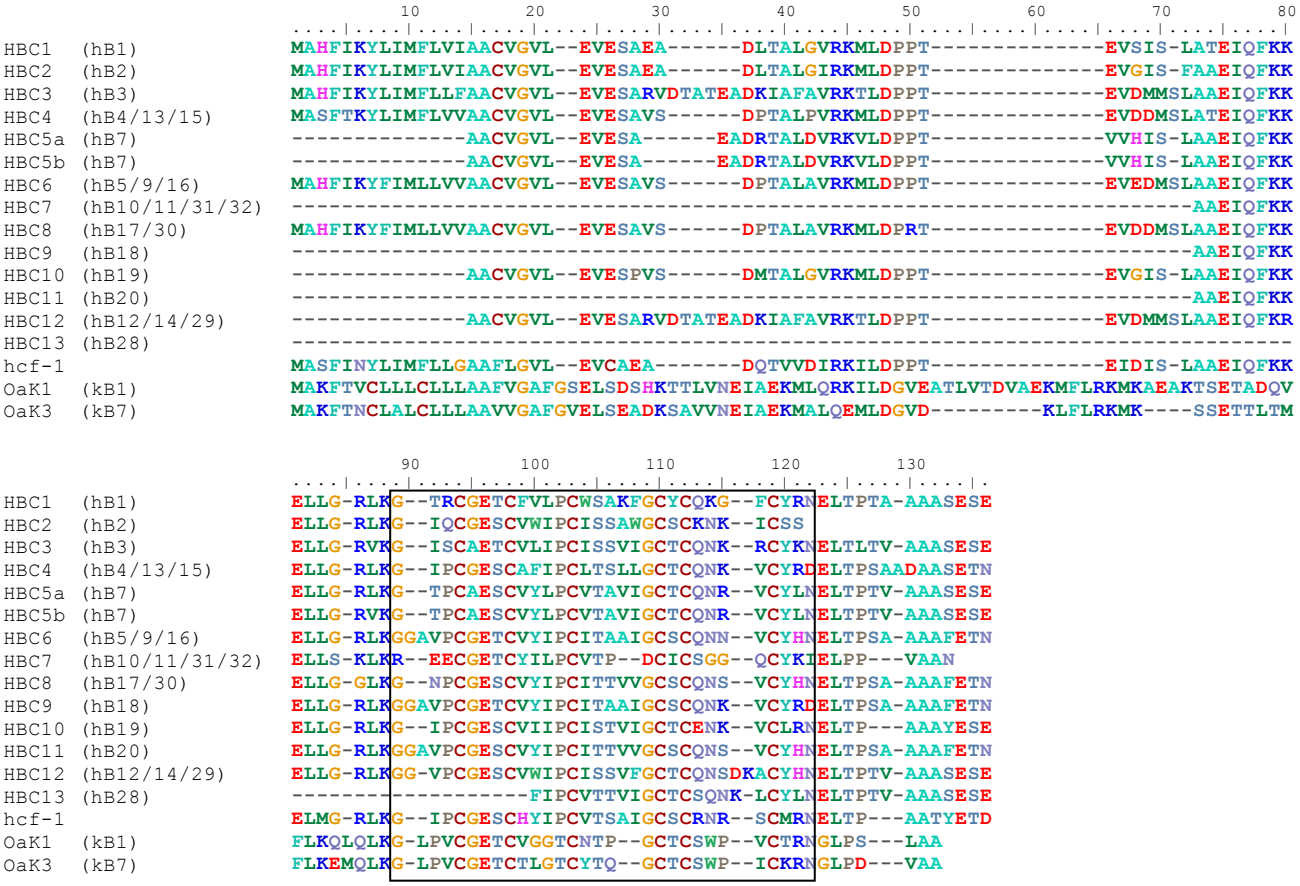


Figure 14. Translated precursor sequences of hedyotide genes from *H. biflora.* Thirteen clones were obtained using degenerate primers based on two conserved sequences, AACVGV and AAEIQFKK, for 3′-RACE PCR followed by and 5′-RACE PCR. Encoding genes of hcf-1, kalata B1 and B7 were aligned together for comparison. Cyclotide domains are boxed.

2.10 Isolation of DNA clones of cyclotide precursors from *H. biflora*

This part of work was performed in collaboration with Dr Wang Wei and Mr Zhang Sen. To compare the hedyotide genetic structures at DNA and mRNA level, encoding genes of hedyotide B1 and hedyotide B2 were cloned from genomic DNA using the cDNA-derived sequences as primers. In both cases, the DNA sequences revealed a single intron (104 nucleotides) locating within the ER signal region (Fig. 12). The introns are highly conserved with 99% sequence identity between two sequences as compared to 94% sequence identity of the exon regions. This is surprising as introns are usually evolved more rapidly and thus more poorly conserved than exons [75]. The high degree of conservation may reflect the possibility of functional constraints or presence of *cis*-regulatory elements within the intron regions.

Besides the DNA clones of hedyotide B1 and B2, another DNA clone was also obtained encoding for a novel predicted cyclotide sequence (hedyotide B21). However, a frame-shift mutation is found in the NTPP region, which results in several premature stop codons. Sequence alignment with *HBC8* clone (encoding for hedyotide B17) revealed an insertion mutation of a cytosine base in the exon region that caused the translational frame-shift (Fig. 15). Consequently, no peptide mass corresponding to this putative cyclotide was found at the protein level. It is thus highly likely that this novel DNA clone may represent a pseudo-cyclotide gene that has been turned off during the evolutionary process.

2.11 Cloning of defensin-like genes from *H. biflora*

Encoding genes of HBDL-1 and HBDL-2 were cloned using degenerate primers targeted GNCAGM sequences in 3′-RACE PCR. Full length clones were subsequently obtained by 5′-RACE PCR. This resulted in identification of a single

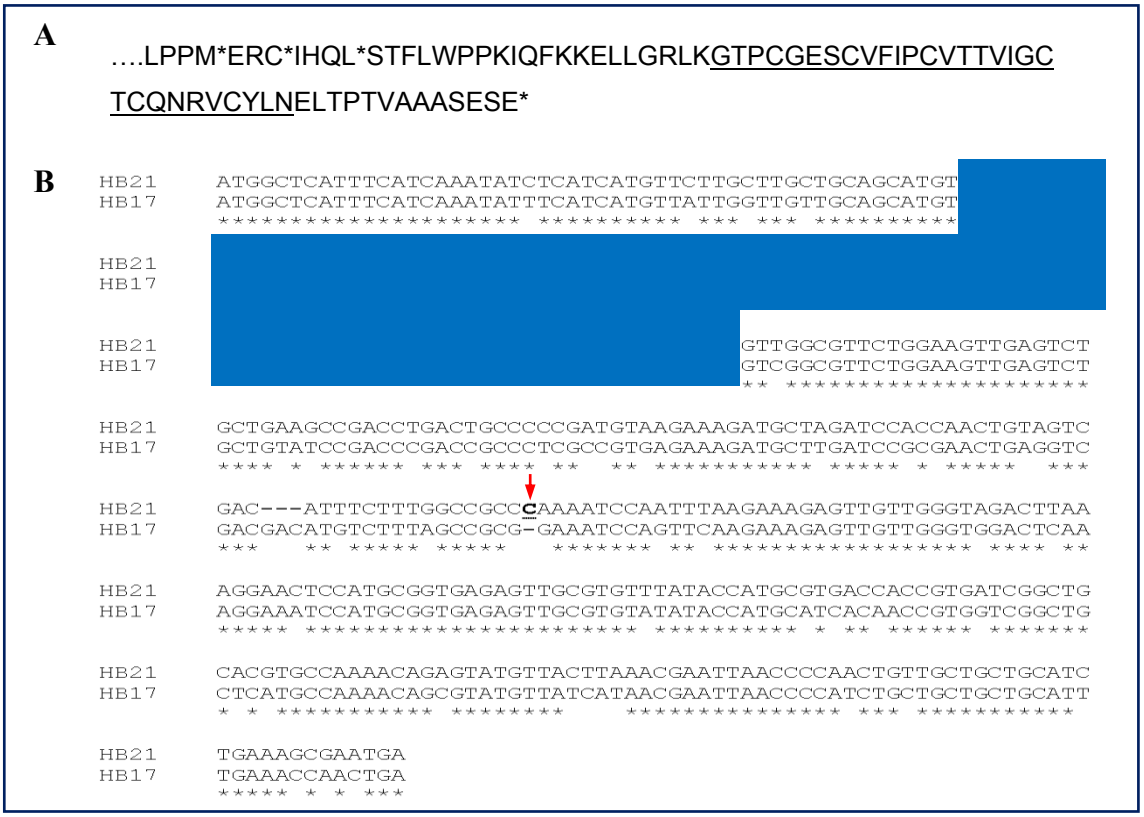


Figure 15. DNA clone of hedyotide B21. A) Predicted translated precursor of hedyotide B21 shows several stop codons (asterisk). The mature cyclotide domain is underlined. B) Sequence alignment of hedyotide B17 cDNA and B21 DNA clone reveal a translational frame-shift induced by a cytosine base insertion in the exon region (indicated by the red arrow). The intron sequence is shaded in blue.

cDNA clone containing mature peptide domain for both HBDL-1 and HBDL-2 (Fig. 16). This supports my prior hypothesis that they are derived from a single precursor. Translated amino acid sequence of HBDL-1 and HBDL-2 precursor shows that it consists of three major domains commonly observed in CRPs: ER signal sequence, N-terminal propeptide and mature peptide domain followed directly by a stop codon.

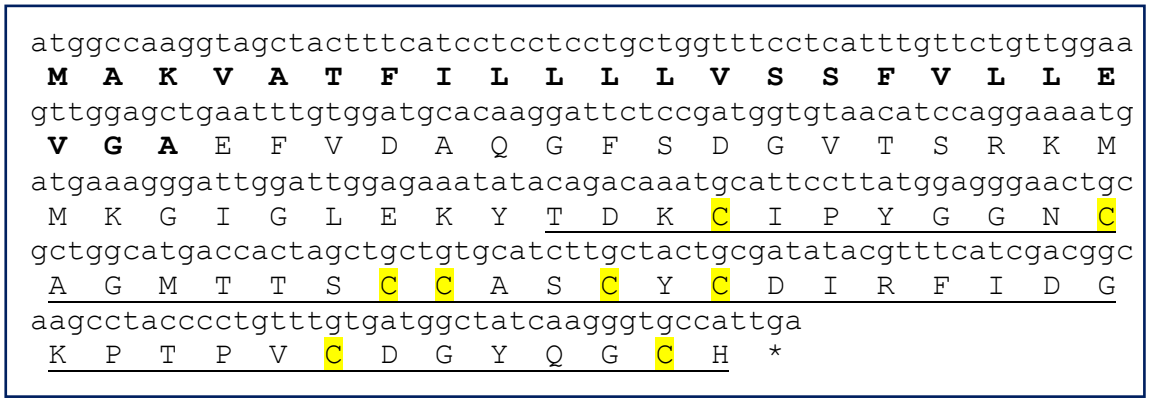


Figure 16. Amino acid sequence of HBDL-1 and HBDL-2 precursor protein deduced from cDNA clone. The sequence contains three major domains: ER signal sequence (in bold), N-terminal propeptide and sequence for mature peptide (underlined).

2.12 Cloning of hedyotide genes from *Hedyotis diffusa*

This part of work was performed in collaboration with Ms Nguyen Quoc Thuc Phuong. Encoding cDNAs of hedyotide D9, D10 and D11 were cloned using degenerate primers targeted loop 1 sequence CYERC in 3′-RACE PCR. Full length clone, designated as *HDC1* (*Hedyotis diffusa* clone), was subsequently obtained by 5′-RACE PCR (Fig. 17). Attempt for cloning of hedyotide D1 to D5 was unsuccessful due to non-specific amplification when using loop 1 sequence CGETC as primers. Translated amino acid sequence of *HDC-1* cloned showed that its genetic arrangement is similar to other rubiaceous cyclotides, which includes ER, NTPP, uncyclotide domain and CTPP region.

Figure 17. cDNA sequence of hedyotide D9-11 and their respective translated precursor protein. The predicted signal peptide is bold, and the mature peptide is underlined. Asterisk indicates the stop codon. Cysteine residues are highlighted in yellow.

2.13 Relative expression of cyclotides in different *Hedyotis* species

To provide a relative expression level of cyclotides in different *Hedyotis* species, a standard cyclotide with *m/z* at 3127 Da isolated from *Clitoria ternatea* (see chapter 7) was introduced into the extracts of *H. biflora, H. diffusa* and *O. affinis.* For each *Hedyotis* species, 50 mg plant materials were homogenized with 250 μl of 50% EtOH. Different amount of standard cyclotide was then added to the plant extracts in such a way that its relative area in MS was at an approximate level of endogenous cyclotides' expression.

At 0.0625 μM concentration, the standard peptide is approximately one-half and ten-fold the endogenous level of cyclotides in *H. biflora* and *H. diffusa*, respectively (Fig. 18). For *O. affinis*, it requires up to 80 μM of the standard peptide to have a comparable relative area to the endogenous cyclotides. The cyclotide expression levels are thus remarkably different among these *Hedyotis* species.

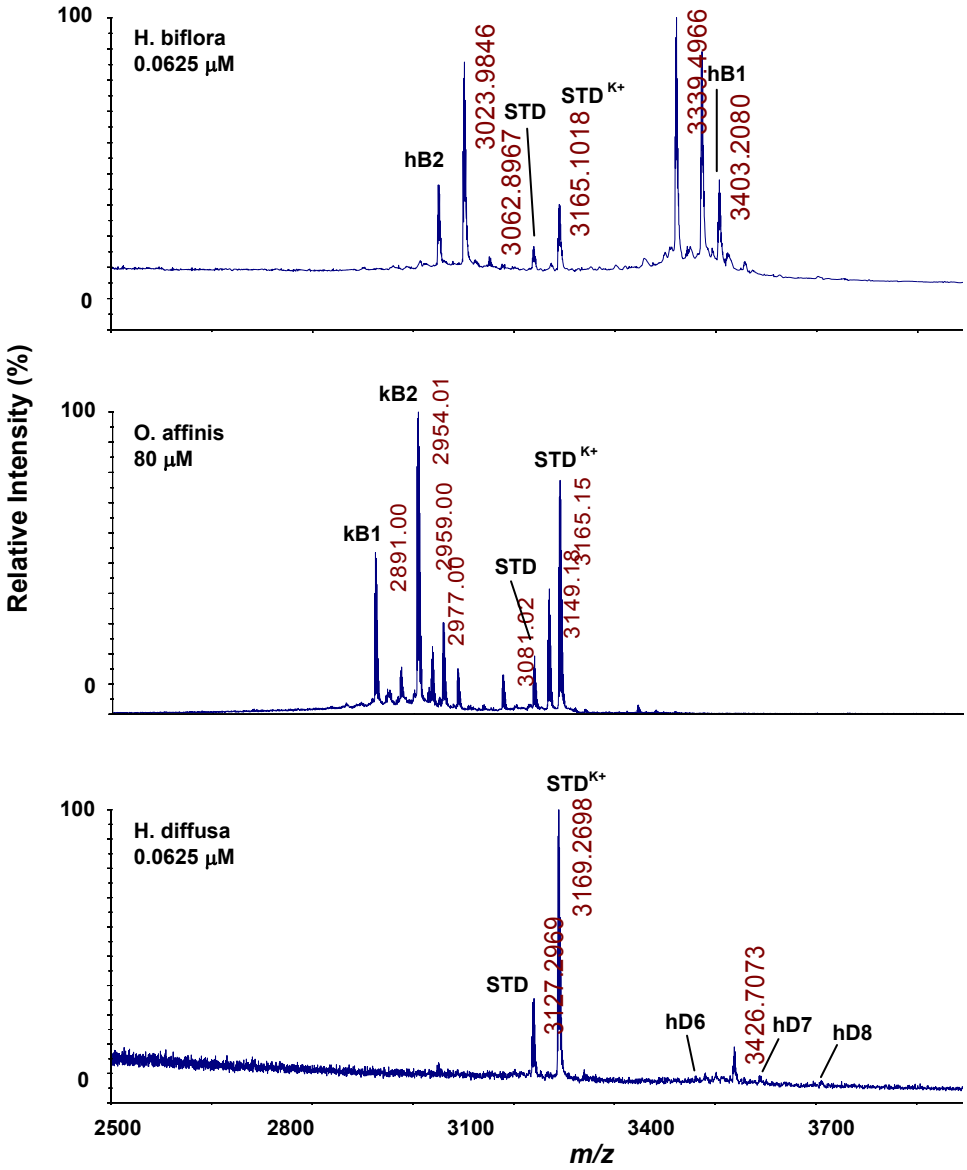


Figure 18. Relative expression of cyclotides and uncyclotides in different *Hedyotis* species. STD and STD^{K+} are the standard peptide and its potassium binding form. STD are added to the plant extract in such a way that its relative area is comparable the endogenous cyclotide expression.

O. affinis has highest expression level followed by *H. biflora* and *H. diffusa.* The relative amount of cyclotides in *O. affinis* is over 12,000-fold (80/(0.0625/10)) higher than *H. diffusa*. This is consistent with the reported yields of cyclotides obtained from plant extraction of *H. diffusa* (0.1 mg/Kg) and *O. affinis* (1000 mg/Kg).

Phylogenetic analysis

A phylogenetic tree was constructed using seven translated cyclotide precursors from three *Hedyotis* species: hedyotide B1 and B2 (*H. biflora*), kalata B1, B2, B3/6 and B7 (*O. affinis*), and hcf-1 (*H. centranthoides*). As shown in Fig. 19, hedyotide precursors have a closer evolutionary relationship to hcf-1 than to kalata precursors. Sequence alignment using bl2seq showed that hedyotide B1 and hcf-1 precursor share 61% sequence identity, whereas only 27% identity between hedyotide B1 and kalata B1. This suggested that *H. biflora* is more closely related to *H. centranthoides* than to *O. affinis.* Our finding is consistent with previous study by Karehed *et al.* using chloroplast and nuclear DNA marker [76] to construct the phylogeny of Rubiaceae species.

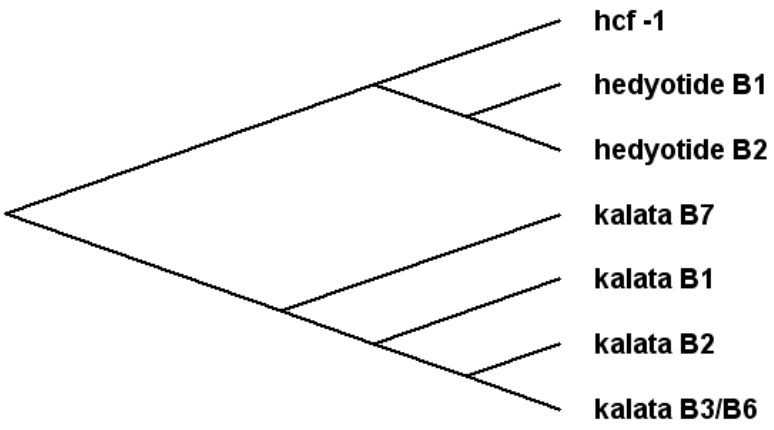


Figure 19. Phylogenetic analysis of cyclotide precursors in *Hedyotis* genus. Tree was constructed using seven cyclotides precursors: hedyotide B1 and B2 (*H. biflora*), kalata B1, B2, B3/6 and B7 (*O. affinis*), hcf-1 (*H. centranthoides*).

2.14 Tissue- and region-specific expressions of cyclotides in *H. biflora*

To determine if novel cyclotides in *H. biflora* were expressed universally or in a tissue-specific manner, different plant parts including roots, stems and leaves were extracted separately. MS profiling of these extracts revealed a substantial difference in the expression patterns between the aboveground and underground parts (Fig. 2). In the aerial tissues, leaves and stems had similar profiles with hedyotide B1 and B2 as major cyclotide constituents. They are relatively specific to the aerial tissues with minor expression in the roots. Majority of other cyclotides were found predominantly in the underground parts. It is common that many defense molecules are found in the roots where plants have to face a diverse number of microbes, pests and pathogens.

Next, we compared the mass spectra of plant samples collected in Vietnam and Singapore to study the regional influence on the cyclotides' composition. We found that the cyclotide profiles of the aerial tissues did not change significantly, whereas the root tissues displayed a striking geographic variation (Fig. 20). Several cyclotides expressed specifically in one sample but were absent in the other. The relative expression levels were also different in both samples. It is noteworthy that the most abundant cyclotides in the roots appear to be regional-specific. Hedyotide B6, for example, was the most abundant cyclotide in plants collected in Vietnam but was nearly missing in the Singapore specimen. On the other hand, hedyotide B10 Da was highly expressed only in the Singapore sample and was not detected in the Vietnam sample under our experimental conditions.

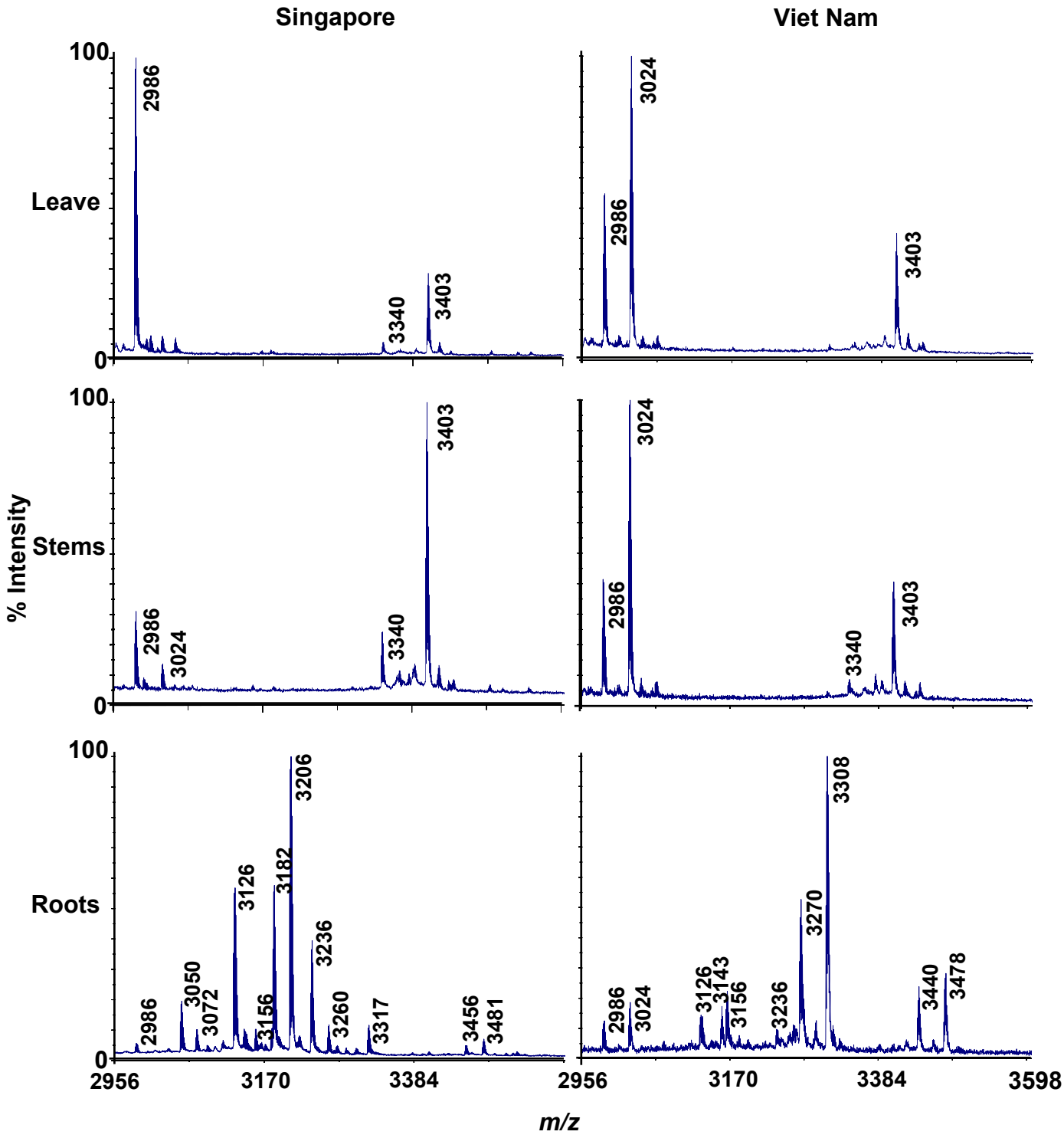


Figure 20. Tissue- and region-specific distribution of cyclotides in *Hedyotis biflora.* Plant specimens were collected in two different countries Singapore and Vietnam. They were divided into three different plant parts: leaves, stems and roots and profiled separately by mass spectrometry.

2.15 Top-down mapping of hedyotide B2 disulfide connectivity

As hedyotide B2 is the first known example of bracelet uncyclotide, it is of interest to elucidate its disulfide connectivity. To facility a rapid disulfide mapping, a top-down approach using tandem mass spectrometry was developed. The overall strategy involved in S-tagging of hedyotide B2 with two different alkylation reagents using NEM and IAA followed by direct MALDI-CID-MS/MS analysis. Although peptide sequencing by MS/MS normally limits for peptides smaller than 2.5 kDa, we were able to obtain nearly complete sequence information across the entire backbone of S-alkylated hedyotide B2 up to 3.6 kDa by optimizing the laser power and number of shots per spectrum. Our approach requires no prior proteolytic digestion step as conventional bottom-up strategy, thus allowing rapid characterization of disulfide pattern.

Native hedyotide B2 was first partially S-reduced with TCEP at pH 3.5. This generated a series of isoforms with one or two disulfide bonds being reduced. The released thiols were immediately alkylated with excess NEM. The whole process was performed under acidic condition at pH 3.5 to avoid the scrambling of disulfide linkage [77, 78]. Partially S-reduced and S-alkylated peptides were then purified by RP-HPLC.

Seven chromatography-separated peaks were collected (Fig. 21) and analyzed by MALDI-TOF MS. Since each NEM-modified residue caused a mass increase of 126 Da, the mass gained after S-alkylation was used to deduce the number of reduced disulfide bonds. Peak 1 contained the native peptide with intact disulfide bridges. Peak 2 and 3 had *m/z* at 3238 Da, corresponding to two NEM-labeled cysteine residues and two intact disulfide bonds (2SS species). Peak 4, 5 and 6 had *m/z* at 3490 Da, corresponding to four NEM-labeled cysteine residues

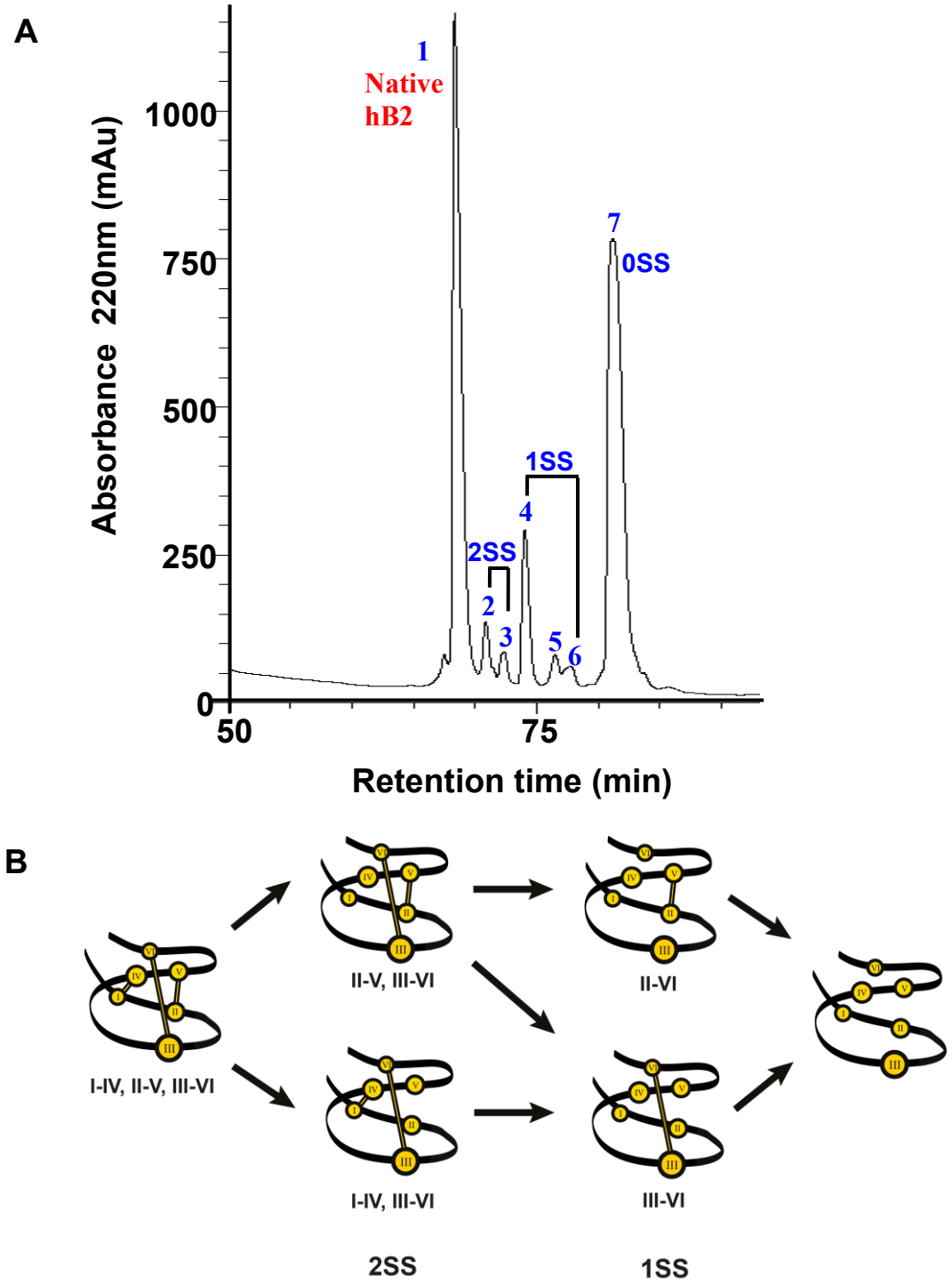
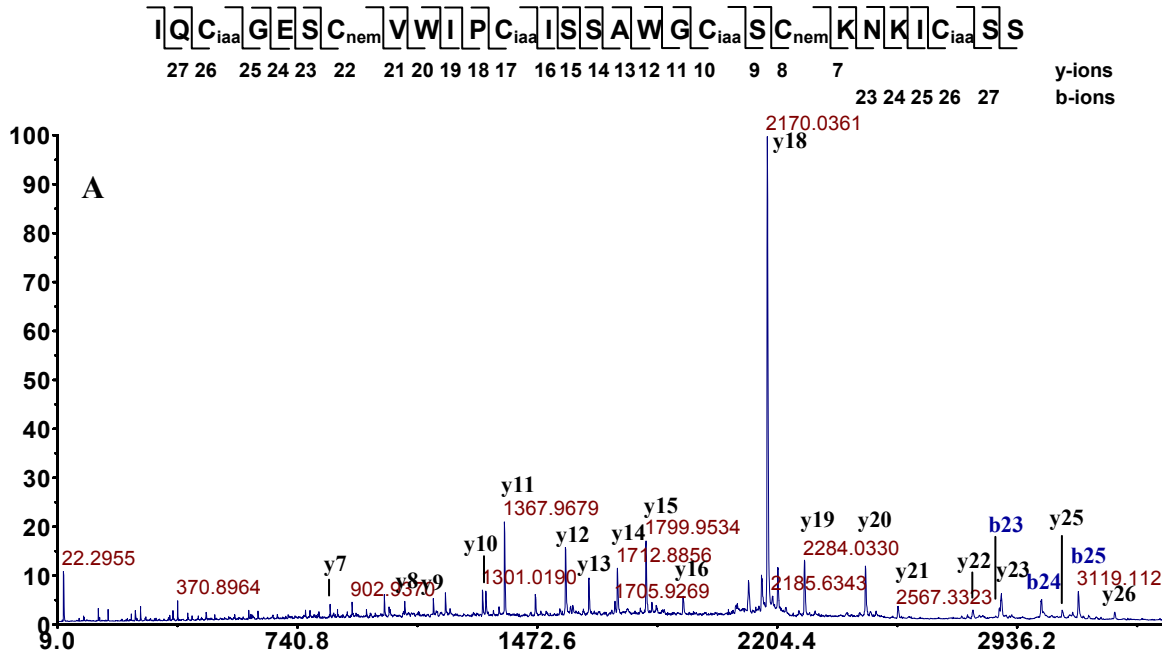
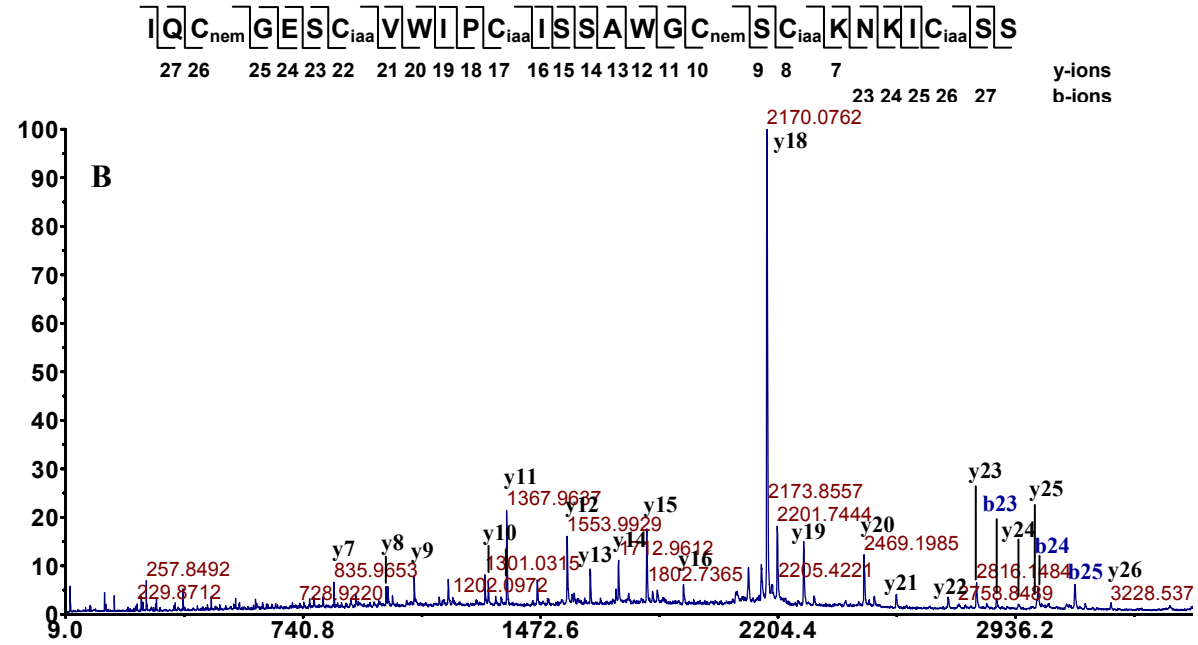


Figure 21. Disulfide mapping of hedyotide B2. (A) HPLC separation of partial reduction and S-NEM alkylated hedyotide B2. Peak 1 is native peptides; peak 2 and 3 are intermediates with two intact disulfide bonds (2SS); peak 4, 5 and 6 are intermediates with one intact disulfide bond (1SS); peak 7 is fully S-NEM alkylated hedyotide B2. (B) Schematic presentation of hedyotide B2 unfolding pathway. Under our experimental condition, the I-IV bond are likely to be reduced first followed by the II-V and lastly by III-VI bond.


IQCiaaGESCnemVWIPCiaaISSAWGCiaaSCnemKNKICiaaSS
27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7
y-ions
23 24 25 26 27
b-ions
A
2170.0361
y18
y11
1367.9679
y15
1799.9534
y19
y20
2284.0330
b23
y25
y10
y12
y14
1712.8856
b25
22.2955
y7
y8
y9
y13
y16
y21
y22
y23
b24
3119.112
370.8964
902.9370
1301.0190
1705.9269
2185.6343
2567.3323
y26
100
90
80
70
60
50
40
30
20
10
0
9.0
740.8
1472.6
2204.4
2936.2


IQCnemGESCiaaVWIPCiaaISSAWGCnemSCiaaKNKICiaaSS
27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7
y-ions
23 24 25 26 27
b-ions
B
2170.0762
y18
y11
1367.9637
y12
1553.9929
y15
2173.8557
y23
b23
y25
2201.7444
y19
y20
2469.1985
y24
b24
y10
y14
1712.9612
y7
y8
y9
y13
y16
257.8492
835.9653
1301.0315
1802.7365
2205.4221
y21
y22
2816.1484
b25
y26
229.8712
728.9220
1202.0972
2758.8489
3228.537
100
90
80
70
60
50
40
30
20
10
0
9.0
740.8
1472.6
2204.4
2936.2

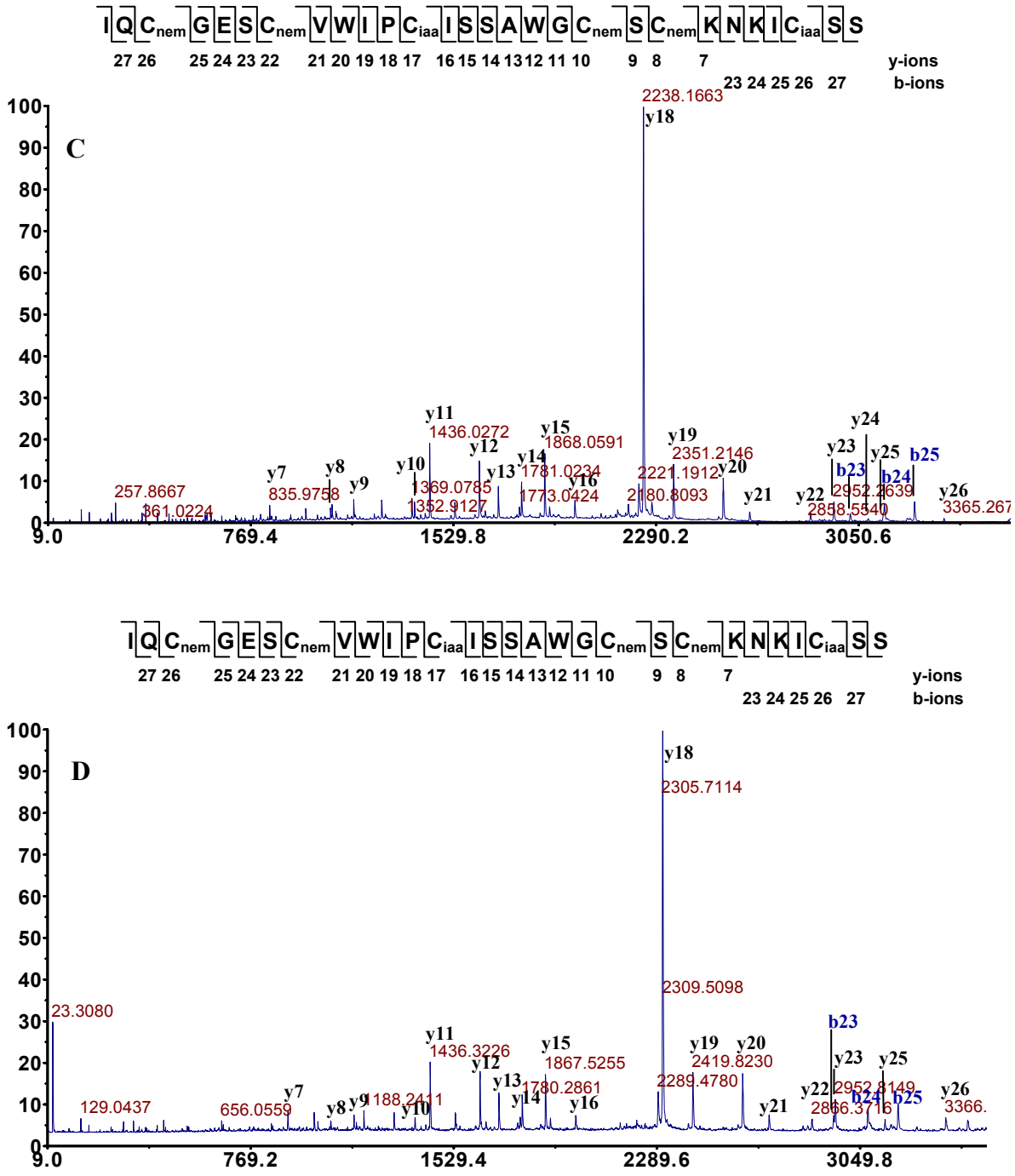


Figure 22. MALDI-TOF MS/MS spectra of S-alkylated hedyotide B2 double-tagged with IAA and NEM. (A), (B), (C) and (D) are S-alkylated derivatives of intermediate species in peak 2, 3, 4 and 6, respectively. Peak 5 contains isoforms of intermediate species in peak 4 and 6 with identical MS/MS spectra and is not shown here.

and one remaining disulfide bond (1SS species). Peak 7 had all three disulfide bond reduced and labeled by NEM (0SS species). It is worth noting that the peptide became more hydrophobic upon S-NEM labeling, probably attributed to the hydrophobicity of the NEM-group. To obtain the disulfide connectivity, 2SS and 1SS species in peak 2, 3 and peak 4, 5, 6, respectively, were fully S-reduced and S-tagged with a second alkylation reagent using IAA. S-alkylated peptides were then analyzed directly by MALDI-CID-MS/MS without prior proteolytic digestion.

Figure 22 shows the MS/MS spectra of S-alkylated hedyotide B2 containing both NEM and IAA labeling. Information from the differential S-tagging was used for deducing the connectivity. S-alkylated species of peak 2 showed S-NEM labeling on Cys II and V, and S-IAA labeling on the remaining cysteine residues, establishing the connectivity as Cys II-V. Similarly, analyzing the MS/MS spectrum of S-alkylated species of peak 3 gave the connectivity of Cys I-IV. Disulfide pattern of 1SS species in peak 4, 5 and 6 were characterized by a mix labeling of four S-NEM and two S-IAA groups. Accordingly, the connectivity of peak 4 and 6 was established as Cys III-VI and Cys II-V, respectively. Peak 5 contained a mixture of two 1SS intermediates with disulfide connection of Cys III-VI and Cys II-V. The coelution of two different disulfide-linkage species in peak 5 was probably due to the formation of stereoisomers resulting from the introduction of a new chiral center on the cysteine side chain upon S-NEM labeling [77]. Our results provided the first chemical evidence that hedyotide B2, an uncyclotide, has the same cystine knot arrangement of Cys I-IV, II-V, III-VI as cyclic counterparts.

2.16 Biodegradation of hedyotide B1 and B2

To provide an understanding about the biodegradation course of cyclotides, aerial materials of *H. biflora* were kept at room temperature until slightly rotten

(around two weeks) before extraction. MS analysis of storage materials and fresh plant samples revealed several degradation products of hedyotide B1 and B2 (Fig. 23). A number of new peptides masses corresponding to novel cyclotides were also observed. They were probably induced during the storing process or peptide fragments from large proteins.

As shown in Fig. 23, both hedyotide B1 and B2 were prone to Trp-oxidation. The indole ring of Trp side chain is known to have three major oxidative states namely Trp_{oia}, Trp_{nfk}, Trp_{kyn}, which cause a mass shift of +4, +16 and +32, respectively [60]. For hedyotide B1, all three oxidative derivatives were observed. A significant percentage of hedyotide B1 (>50%) was being oxidized justified from the MS ratio of the native and the oxidized products. Hedyotide B1 contains a single Trp residue located on loop 3, which is exposing on the peptide surface making it relatively susceptible to oxidation. For hedyotide B2, it was also undergone similar oxidative process but was complicated by having two Trp residues in the primary sequence. It is thus possible to form up to 15 different oxidation products. They can be easily recognized under RP-HPLC by the characteristic loss of 280 nm absorbance due to the modification of the indole ring. Thirteen of them were observed at molecular weight level. Their masses and possible oxidation state were summarized in Table 6.

Beside Trp-oxidation, hedyotide B2 was also susceptible to exopeptidases degradation owing to its open-end structure. The Ile residue at the N-terminus was trimmed giving rise to hedyotide B26 with a *m/z* at 2873 Da. Subsequently, the Gln residue of hedyotide B26 at the N-terminus spontaneously cyclized to form a pyroglutamyl derivative generating hedyotide B27 with a *m/z* at 2856 Da. The loss of N-terminal residue was probably due to the open-end structure of hedyotide B2

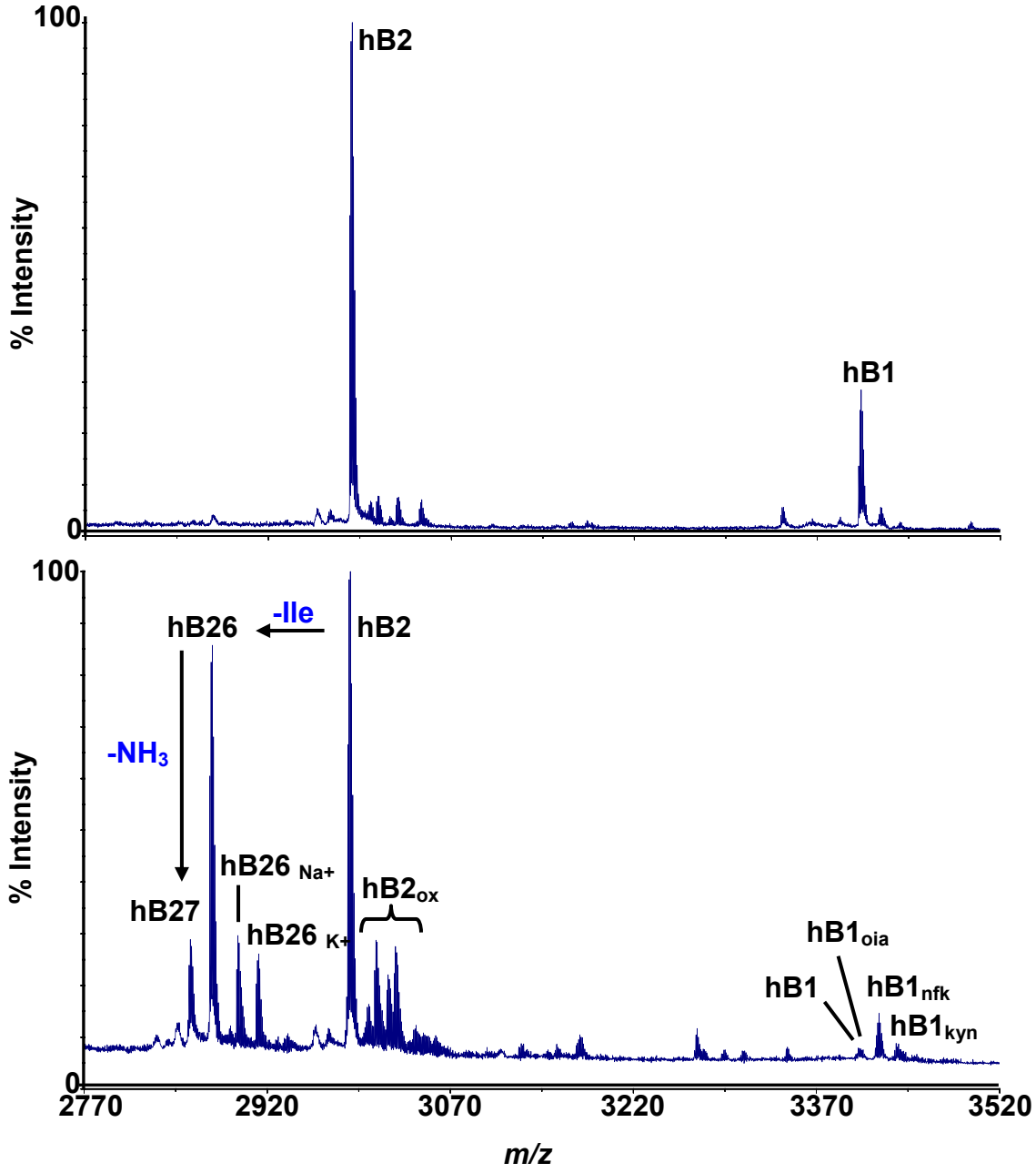


Figure 23. Biodegradation of hedyotide B1 and B2. (A) MS profile of freshly collected plants. (B) MS profile of stored plant samples shows several degradation products of hedyotide B1 and B2.

Table 6. List of hedyotide B2 modification products. Y/N = Yes/No

Peptide	Amino Acid Sequence	MW	Observed
B2	IQCGESCVW..IPCISSAW..GCSCKNKICSS	3002	Y
B2a	IQCGESCVW$_{oia}$IPCISSAW..GCSCKNKICSS	3002	Y
B2b	IQCGESCVW..IPCISSAW$_{oia}$GCSCKNKICSS	3002	Y
B2c	IQCGESCVW$_{nfk}$IPCISSAW..GCSCKNKICSS	3018	Y
B2d	IQCGESCVW..IPCISSAW$_{nfk}$GCSCKNKICSS	3018	Y
B2e	IQCGESCVW$_{kyn}$IPCISSAW..GCSCKNKICSS	2990	Y
B2f	IQCGESCVW..IPCISSAW$_{kyn}$GCSCKNKICSS	2990	Y
B2g	IQCGESCVW$_{oia}$IPCISSAW$_{oia}$GCSCKNKICSS	3018	Y
B2h	IQCGESCVW$_{nfk}$IPCISSAW$_{nfk}$GCSCKNKICSS	3050	N
B2i	IQCGESCVW$_{kyn}$IPCISSAW$_{kyn}$GCSCKNKICSS	2994	N
B2j	IQCGESCVW$_{oia}$IPCISSAW$_{nfk}$GCSCKNKICSS	3034	Y
B2k	IQCGESCVW$_{nfk}$IPCISSAW$_{oia}$GCSCKNKICSS	3034	Y
B2l	IQCGESCVW$_{oia}$IPCISSAW$_{kyn}$GCSCKNKICSS	3006	Y
B2m	IQCGESCVW$_{kyn}$IPCISSAW$_{oia}$GCSCKNKICSS	3006	Y
B2n	IQCGESCVW$_{nfk}$IPCISSAW$_{kyn}$GCSCKNKICSS	3022	Y
B2o	IQCGESCVW$_{kyn}$IPCISSAW$_{nfk}$GCSCKNKICSS	3022	Y
B26	QCGESCVW..IPCISSAW..GCSCKNKICSS	2873	Y
B26a	QCGESCVW$_{oia}$IPCISSAW..GCSCKNKICSS	2889	N
B26b	QCGESCVW..IPCISSAW$_{oia}$GCSCKNKICSS	2889	N
B26c	QCGESCVW$_{nfk}$IPCISSAW..GCSCKNKICSS	2905	N
B26d	QCGESCVW..IPCISSAW$_{nfk}$GCSCKNKICSS	2905	N
B26e	QCGESCVW$_{kyn}$IPCISSAW..GCSCKNKICSS	2877	N
B26f	QCGESCVW..IPCISSAW$_{kyn}$GCSCKNKICSS	2877	N
B26g	QCGESCVW$_{oia}$IPCISSAW$_{oia}$GCSCKNKICSS	2905	N
B26h	QCGESCVW$_{nfk}$IPCISSAW$_{nfk}$GCSCKNKICSS	2937	N
B26i	QCGESCVW$_{kyn}$IPCISSAW$_{kyn}$GCSCKNKICSS	2881	N
B26j	QCGESCVW$_{oia}$IPCISSAW$_{nfk}$GCSCKNKICSS	2921	N
B26k	QCGESCVW$_{nfk}$IPCISSAW$_{oia}$GCSCKNKICSS	2921	N
B26l	QCGESCVW$_{oia}$IPCISSAW$_{kyn}$GCSCKNKICSS	2893	N

B26m	QCGESCVW$_{kyn}$IPCISSAW$_{oia}$GCSCKNKICSS	2893	N
B26n	QCGESCVW$_{nfk}$IPCISSAW$_{kyn}$GCSCKNKICSS	2909	N
B26o	QCGESCVW$_{kyn}$IPCISSAW$_{nfk}$GCSCKNKICSS	2909	N
B27	Q*CGESCVW..IPCISSAW..GCSCKNKICSS	2856	Y
B27a	Q*CGESCVW$_{oia}$IPCISSAW..GCSCKNKICSS	2872	N
B27b	Q*CGESCVW..IPCISSAW$_{oia}$GCSCKNKICSS	2872	N
B27c	Q*CGESCVW$_{nfk}$IPCISSAW..GCSCKNKICSS	2888	N
B27d	Q*CGESCVW..IPCISSAW$_{nfk}$GCSCKNKICSS	2888	N
B27e	Q*CGESCVW$_{kyn}$IPCISSAW..GCSCKNKICSS	2860	N
B27f	Q*CGESCVW..IPCISSAW$_{kyn}$GCSCKNKICSS	2860	N
B27g	Q*CGESCVW$_{oia}$IPCISSAW$_{oia}$GCSCKNKICSS	2888	N
B27h	Q*CGESCVW$_{nfk}$IPCISSAW$_{nfk}$GCSCKNKICSS	2920	N
B27i	Q*CGESCVW$_{kyn}$IPCISSAW$_{kyn}$GCSCKNKICSS	2864	N
B27j	Q*CGESCVW$_{oia}$IPCISSAW$_{nfk}$GCSCKNKICSS	2904	N
B27k	Q*CGESCVW$_{nfk}$IPCISSAW$_{oia}$GCSCKNKICSS	2904	N
B27l	Q*CGESCVW$_{oia}$IPCISSAW$_{kyn}$GCSCKNKICSS	2876	N
B27m	Q*CGESCVW$_{kyn}$IPCISSAW$_{oia}$GCSCKNKICSS	2876	N
B27n	Q*CGESCVW$_{nfk}$IPCISSAW$_{kyn}$GCSCKNKICSS	2892	N
B27o	Q*CGESCVW$_{kyn}$IPCISSAW$_{nfk}$GCSCKNKICSS	2892	N

W_{oia}, W_{nfk}, W_{kyn} indicate three major oxidative states of the indole ring of the Trp side chain [60]. Q* indicate pyroglutamic acid.

making it susceptible to exopeptidases released during plant-degradation process. It should be note that hedyotide B26 and B26 are also prone to oxidation, and each may generate a set of 15 different oxidation products. Hedyotide B2 thus can give rise up to 47 different modification variants, including the formation of hedyotide B26, B27 and their derivatives. My results provide the first understanding about the biodegradation of cyclotides. This also highlights the additional advantage of cyclic backbone for being resistant against exopeptidase degradations.

2.17 Antimicrobial activity

Due to low expression level of cyclotides in *H. biflora* and *H. diffusa*, only hedyotide B1 and B2 were obtained in sufficient amount for bioassays. Antibacterial activity of hedyotide B1 and B2 were tested against four bacterial strains: *E. coli, S. aureus, S. epidermis* and *S. salivarius.* Hedyotide B1 was active against *E. coli* and *S. salivarius* with MIC value of 3.4 and 5.9 μM, respectively (Table 7). Hedyotide B2 was inactive against all tested bacteria strains up to 80 μM concentration. Neither peptide had inhibitory activity against *S. aureus* and *S. epidermis*.

Table 7. Antibacterial activities of hedyotide B1 and B2

Organism	MIC[a] (μM)		
	hB1	hB2	D4R[b]
E. coli	3.4	>80	1.29
S. salivarius	5.9	>80	0.96
S. epidermidis	>80	>80	0.71
S. aureus	>80	>80	1.9

[a] Minimal Inhibitory Concentration. [b] Synthetic antibacterial peptide as positive control

3. Discussion

3.1 Distribution and relative expression of cyclotides in *Hedyotis* genus

My findings of cyclotides in *H. biflora and H. diffusa* provide the first report of annual *Hedyotis* herbs classified as cyclotide-bearing plants. The morphologies of other *Hedyotis* species vary from perennial herbs (O*. affinis*), shrubs (*H. centranthoides*) and perennial woody trees (*H. terminalis*). Interestingly, the amount of cyclotides' production differs vastly among *Hedyotis* species. The relative expression of cyclotides in *O. affinis*, for example, is approximately 12,000-fold higher than in *H. diffusa*. It has been previously believed that most cyclotide-containing plants express cyclotides at very high levels (i.e., kalata B1 at 1 g/Kg dry weight) [33]. It is the first time that only a trace amount of cyclotides is reported to be produced in a given plant. However, it is uncertain why there is such a large difference in the relative expression of cyclotides among related species of the same genus. Besides the variation in concentration, the cyclotide landscape also varies greatly among plants. *O. affinis* produces both Möbius and bracelet cyclotides where as *H. biflora* appears to produce only bracelet and hybrid cyclotides/uncyclotides, and *H. diffusa* produces only uncyclotides.

Intriguingly, despite the discovery of cyclotides in multiple *Hedyotis* species, not all plants in this genus express cyclotides. *Hedyotis corymbosa,* for example, is a widely distributed species of *Hedyotis* genus. It is closely related to *O. affinis* and a detail comparison has been studied extensively by Datta and Sen, and Grant et al. [79]. Regardless of its closed relationship with *H. biflora* and *O. affinis, H. corymbosa* has been shown to be a non-cyclotide containing species [33]. This was confirmed in my work with no trace amount of cyclotides being detected in *H. corymbosa* extract under my experimental conditions. This raises the question why cyclotides have

been lost in several *Hedyotis* plants, or more broadly in the Rubiaceae lineages during their course of evolution. Cyclotides are known to possess a wide-range of defense-related functions such as antimicrobial [21], anti-HIV [38] and insecticidal [37]. It is intriguing to speculate how plants ensure their fitness in the absence of cyclotide-defense armory. It is possible that plants may have developed alternative protection strategies in compensation for the loss of cyclotide expression.

3.2 Rapid approach of primary structure determination by MALDI-CID-MS/MS

Primary structures of hedyotides were determined by MALDI-CID-MS/MS, which offers several advantages over the traditional ESI-CID-MS/MS, i.e. salt tolerance, rapid analysis and high throughput [80]. Its current application for *de novo* sequencing, however, is mostly limited for small peptides as CID fragmentation rarely yields complete sequence coverage for peptides bigger than 2 kDa [74]. Here, I have shown that large peptides up to 3.5 kDa can be fragmented efficiently using MALDI-CID-MS/MS. Sequence information across the entire backbone of cyclotides up to 30 residues were obtained in a single MS/MS spectrum. My method provides a rapid strategy for primary structure determination of cyclotides.

The key to my strategy is the usage of trypsin for digestion to generate precursor peptides with a positive charge at the C-terminus. This effectively facilitates the ionization of the fragment ions, especially the y-series ions. Using hedyotide B7 as an example, I demonstrated that alteration of the charge state at the C-terminus by using Glu-C instead of trypsin for digestion had a dramatic effect on the CID fragmentation efficiency. Glu-C treatment yielded peptides with C-terminal Glu residue which suppressed the protonation. Consequently, fragment ions devoid

of internal basic residues were not detected. Tryptic digestion thus generates more suitable peptides for *de novo* sequencing than Glu-C digestion.

3.3 Sequence analysis of novel cyclotides from *H. biflora*

All novel hedyotides isolated from *H. biflora* belong to bracelet subfamily except hedyotide B10, B11, B31 and B32 which contain hybrid elements from both subfamilies. Interestingly, none of them are of Möbius subfamily despite our exhaustive effort in characterization of over thirty new peptides. Möbius cyclotides though generally are less common than bracelet variants; most plants do produce both types. It is thus unusual for *H. biflora* to produce only one type of cyclotides and devoid of the others.

In attempt to provide a plausible explanation of this bias, statistical examination of the overall net charge of 125 cyclotides from cybase (81 bracelet and 44 Möbius) revealed that novel hedyotides have similar distribution pattern to Möbius rather than bracelet cyclotides which are usually more cationic. The majority of hedyotides fall into the neutral and negative net charge region (Fig. 24). The resembling of hedyotides to Möbius pattern may reflect a functional redundancy in which hedyotides may have evolved in such a way to complement for the loss of Möbius cyclotides. Functional redundancy is a common phenomenon in multigene protein families.

Another interesting finding about the cyclotide constituents of *H. biflora* is the presence of a large number of uncyclotides. Thus far, only four uncyclotides have been isolated with all belonging to Möbius subfamily [24, 60, 61]. Surprisingly, 15 linear variants have been isolated from *H. biflora* alone. They can be categorized into

bracelet and hybrid uncyclotides, and are also the first examples of uncyclotides from these two subfamilies hitherto reported.

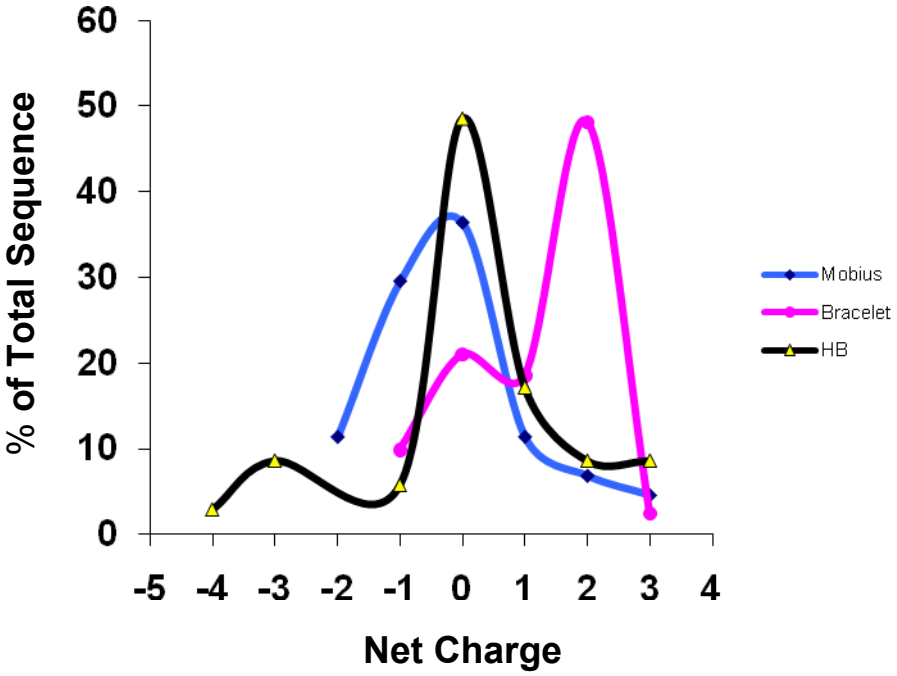


Figure 24. Net charge distribution of cyclotides. Graph was plotted using 35 hedyotide sequences from *H. biflora,* and a database of 81 bracelet and 44 Möbius cyclotides from Cybase (Appendix B).

The novel hedyotides also displayed high sequence diversity with new sequence elements. Residues in loop 1, for example, are typically highly conserved with the consensus sequence of G/A-E-S/T. The GEG sequence in hedyotide B25 features an exception to this rule with Gly occupying the third position. The discovery of hedyotide B25 at the protein level also supports the existence of mra30a, a mRNA predicted cyclotide with GEG triad in loop 1, from the Violaceae species *Melicytus ramiflorus* [53].

Most hedyotide loop 2 residues are hydrophobic and feature a conserved motif of H-H-I/L-P, (H=hydrophobic). The hydrophobic character of loop 2 is probably important for membrane interaction as suggested in previous studies [29, 81]. The terminal proline is required for turn geometry which explains for its conservation. Most bracelet cyclotides usually contain one aromatic residue in first or second position of loop 2. Hedyotide B3 represents one of the few cyclotides that have no aromatic residue in this loop.

Loop 3 is the most variable of all hedyotide loops and contains six to seven residues. In many bracelet cyclotides, it forms a short alpha helix, which explains why bulky side-chain aromatic amino acid such as Trp, Phe and Tyr are rarely observed in this loop. Hedyotide B1 stands out as the only reported cyclotides having two aromatic residues in loop 3. In addition, it is also the most basic cyclotides (+3) and the only known example having six aromatic residues as opposed to usual two to three residues in most cyclotides. Hedyotide B2 is another example containing Trp residue in loop 3.

Loop 4 contains a single residue and has high propensity for hydroxyl-bearing amino acids such as Thr and Ser. Interestingly, this position is occupying by the aromatic residue Tyr in hedyotide B1 and the hydrophobic residue Ile in hedyotide B10.

Loop 5, perhaps, is the most distinctive trait of hedyotides with all the loop sequences being novel and retaining a general hydrophilic character. Unlike the majority of bracelet cyclotides which carry two basic residues in loop 5 (more than 65% of the total sequences), most hedyotides contain either one or no basic residue in this loop making them overall less cationic. In hedyotide B3, loop 5 ends with a basic residue Arg (QNKR) instead of the highly conserved Val residue. Little is

known about the structural and functional role of Val at this position and why it is so conserved. In hedyotide B12, loop 5 have an extended sequence with six residues (QNSDKA) relative to four found in most other cyclotides.

Loop 6 is the longest of all hedyotide loops. It usually contains six residues but can accommodate longer sequence. Hedyotide B24 and B9, for example, have seven and eight residues in this loop, respectively. The longest known loop 6 can harbor up to 10 residues such as cycloviolacin Y1 [82] and tricyclon A [83] suggesting the flexibility of this loop for epitope grafting in protein engineering using cyclotides as a scaffold. Several residues in loop 6 are highly conserved, especially Asn and Gly which are required for the backbone cyclization. Other conserved residues include Tyr at the first position, Arg/Lys/His at the second position and Pro at the last position.

3.4 Analysis of novel uncyclotides from *H. diffusa*

Hedyotis diffusa is the most well-known herb of the *Hedyotis* genus famous for its anti-cancer property [84]. Examination of its cyclotide content resulted in characterization of 11 novel peptides which all possess an open-end structure despite of their high sequence homology with known cyclotides from the Rubiaceae and Violaceae. This is surprising as the backbone cyclization machinery in cyclotide-producing plants is known to be highly efficient with only two naturally occurring uncyclotides hitherto reported [24, 61]. A closer examination of their primary structures shows that two key residues, N-terminal Gly and C-terminal Asn, which normally require for backbone cyclization are absent in all cases. This may account for their inability to form a cyclic structure.

The novel uncyclotides can be subdivided into two groups: Möbius -like and bracelet-like cyclotides. They display significant sequence variation from known cyclotides. In the first group, it appears that the absence of the cyclic backbone is compensatory by a pyroglutamyl residue at the N-terminus, which is known to protect the peptide from exopeptidase degradations. They also display substantial variation in loop 4, i.e. Arg instead of the conserved Ser/Thr, and loop 5 with all loop sequences are novel. Interestingly, the characteristic cis-proline turn in loop 5 of Möbius cyclotides is replaced with a glycine turn in hedyotide D3, D4 and D5.

Uncyclotides in the second group display extensive sequence substitutions on loop 1 and loop 5. Loop 1 is the most conserved of all cyclotide loops tolerant to only a small range of amino acids variation. It usually has the consensus sequence of G/A-E-T/S. Hedyotide D6-11 share the same loop 1 primary sequence, which is entirely different from conventional cyclotides, except the absolutely conserved Glu residue. Small side-chain amino acid Gly/Ala at the first position of loop 1 is substituted by a bulky Tyr residue. The neutral hydroxyl bearing side-chain Ser/Thr at the third position is replaced with the basic amino acid Arg. Loop 5 of these peptides has also being modified with particular enrichment of hydroxyl containing residues DTTTYD. These modifications are probably to maintain the structure stability in compensation for loop 1 substitutions.

Thus far, no “close-end” cyclotides have been detected in *H. diffusa.* It is possible that they may present at low concentration making them difficult for isolation and sequence determination. The cyclic backbone thus may be not an essential requirement in this medicinal plant. This may be partly due to the energetic requirement for plants to produce cyclic peptides, which requires an additional step to ligate the N- and C-termini together. Therefore, it is probably more economic for

plants to produce linear peptides. Cyclic peptides, however, do have advantages in term of stability and resistance against exopeptidases.

It is also worth noting that *H*. *diffusa* is an annual herb with relatively small leaves as compared to *O*. *affinis* which is a perennial with larger leaves. It is thus possible that *H. diffusa* with a shorter life span and smaller leaves, it is less susceptible to insect attack than *O. affinis* and hence may not require "extremely-stable" peptides for its defense mechanism. The evolution pressure to conserve the cyclic backbone is thus less stringent than *O. affinis*, which provides a possible explanation for the wide spread of uncyclotides in *H. diffusa*.

3.5 Top-down mapping of hedyotide B2 disulfide linkage

The discovery of hedyotide B2 as the first linear variant of bracelet subfamily has stimulated my interest in characterization of its disulfide pattern. To my knowledge, cystine connectivity of linear cyclotides has not been to be established, although they have been proposed to possess the same knotted arrangement [24]. Here, I mapped the connectivity of hedyotide B2 using a top-down approach instead of conventional bottom-up strategy where proteins are subjected to proteolysis prior to MS analysis. Although bottom-up approach provides a straightforward strategy for characterizing the cystine linkages, it requires extra steps of proteolytic digestion and purification, which is not desirable for rapid disulfide mapping. In addition, it is common that many of the digested fragments are not observed in MS due to inappropriate charge, mass or insolubility upon digestion [85]. Valuable information may be lost making it difficult for complete characterization of a protein. Therefore, by eliminating the proteolytic digestion step and subjecting the whole peptides

directly to MS/MS analysis, I overcome these limitations allowing rapid characterization of disulfide connectivity.

In my approach, hedyotide B2 was sequentially tagged with NEM and IAA followed by direct by MS/MS analysis. In the first S-alkylation with NEM, five different intermediate species eluted in peak 2-6 were collected. In previous work by Goransson *et al.* [77] for disulfide characterization of kalata B1, only two intermediate species were isolated, resulting in identification of Cys II-V and III-VI linkages. The third connectivity of Cys I-IV was only obtained by deduction. MS characterization of five intermediate species in my study provided chemical evidences for all three cystine linkages establishing the knotted arrangement of hedyotide B2.

In addition to the characterization of disulfide connectivity, the analysis of intermediate products also provides additional information about the unfolding mechanism of uncyclotides. Intermediate species at various reduction stages and possible unfolding pathway are presented in Fig. 20B. Cys III-VI appears to be the most stable disulfide bond as no 2SS species with breakage of Cys III-VI was isolated under my experimental conditions. In addition, 1SS species with intact Cys III-VI is the most abundant among the five intermediate species. This finding is consistent with our understanding of cystine knot arrangement as Cys III-VI is most shielded and buried inside the cystine core. Our results are also consistent with the previous study [77] on kalata B1 unfolding pathway, where the penetrating disulfide bond (Cys III-VI) appears to be most stable and the last to be broken.

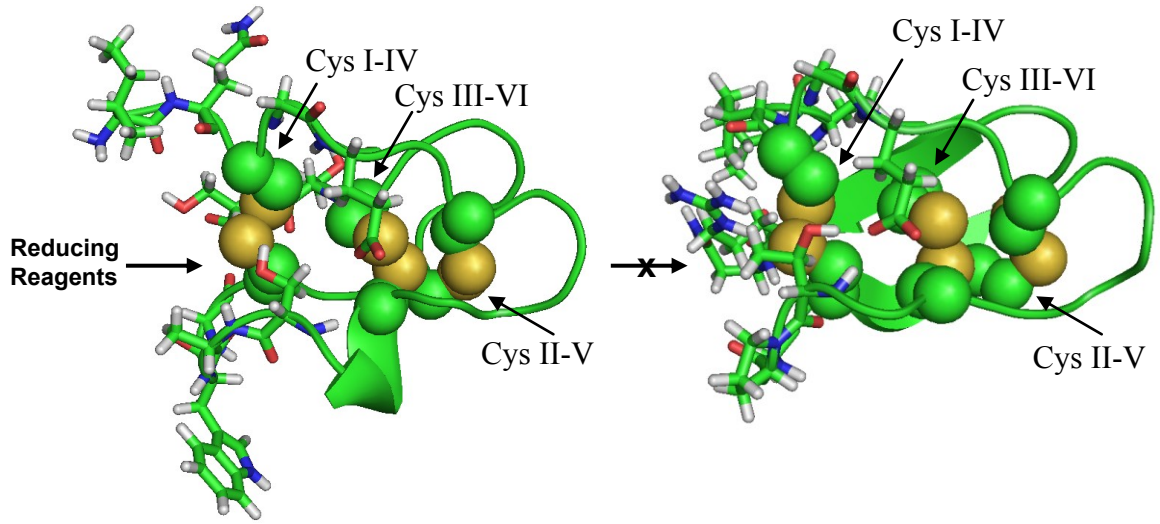


Figure 25. Computer models of hedyotide B2 and kalata B1 (PDB, 1NB1). Disulfide bonds are in sphere and the surrounding amino acids are in stick configuration. In hedyotide B2, its open-end structure increases the accessibility of Cys I-IV to reducing reagents, whereas in kalata B1 this bond is sterically hindered by the cyclic backbone. Cys I-IV, II-V and III-VI disulfide bonds are indicated by the arrows.

Reduction order of the other two disulfide bonds is more difficult to predict since both 2SS species with breakage of Cys I-IV and II-V were isolated. However, no 1SS species with intact Cys I-IV was identified, suggesting that this disulfide bond is first to be broken. This is contrasted with previous study by Gorranson *et al.* [77], in which Cys II-V was proposed to be reduced first. The difference in the reduction order is likely attributed to the cyclic and linear nature of kalata B1 and hedyotide B2, respectively. A computer model of hedyotide B2 based on circulin A was used for comparison with kalata B1 structure (Fig. 25). In kalata B1, Cys II-V is most surfaced exposed and thus most readily reduced [77]. Cys I-IV is shielded by the cyclic backbone limiting its access to reducing reagents. In hedyotide B2, its open-ends structure increases the accessibility of Cys I-IV making it more prone to reduction

than Cys II-V. The reduction order of Cys I-IV and II-V thus may be reversed in linear as compared to cyclic forms.

3.6 Biosynthesis pathway of hedyotide B2

Comparing the amino acid sequence with the known database, the linear structure of hedyotide B2 is likely genetic predetermined and unlikely resulted from ring-opening or chemical degradation. Its mature sequence lacks both highly conserved residues N-terminal Gly and C-terminal Asn which are required for backbone cyclization. cDNA encoding clone of hedyotide B2 (*HBC2*) revealed a nonsense mutation that introduces a premature stop codon precisely at the conserved Asn position. This residue has been proposed to act as recognition site for a putative asparaginyl endopeptidase (AEP), an enzyme believed to catalyze the peptide cyclization [54]. The open-end nature of hedyotide B2 is probably owing to the absence of this key residue. Similar mutation has been reported in violacin A, the first identified linear cyclotide from *V. odorata* [24]. The stop codon in this peptide, however, occurs at one residue prior to the conserved Asn. Uncyclotides generally could be formed if nonsense mutations occur at any position in loop 6 that prevent the translation of the conserved Asn residue.

The *HBC2* clone also reveals that the conserved Gly residue at the N-terminus is indeed present in the precursor protein but is absent in the mature peptide. It is uncertain whether hedyotide B2 is synthesized through a direct cleavage of the precursor protein resulting in the mature product without the Gly residue, or through a short-live intermediate with the N-terminal Gly and subsequently undergoing proteolysis. The later scenario seems more likely to happen as sequence comparison with hedyotide B1 and rubiaceous cyclotides

suggested that the cleavage sites are highly conserved and located between Lys and the conserved Gly. Trimming of Gly residue is probably caused by aminopeptidases which are quite ubiquitous in plants. This process appears to be highly efficient as mature peptide with the conserved Gly residue was not detected in the plant extracts under our isolation procedure.

3.7 Biodegradation of cyclotides in *Hedyotis biflora*

Characterization of stored plant samples provides an early understanding about the degradation of cyclotides in the natural environment. Several biodegraded products of hedyotide B1 and B2 were formed during the prolong storage of plant materials. They can be classified into two categories: Trp-oxidized products and truncated products. In the first category, hedyotide B1 and B2 formed three and thirteen putative Trp-oxidized products, respectively. The modifications of Trp have been proposed to be facilitated by exposure to sun light [60]. In my study, the plant materials were kept at room temperature away from sun light suggesting that other factors beside UV released during the plant decomposition process may accelerate the Trp oxidation.

The second category affects only hedyotide B2 owing to its open-end structure whereas hedyotide B1 with a cyclic backbone is not affected. N-terminal trimming of hedyotide B2 led to the formation of hedyotide B26 and B27. No detectable amount of these two peptides was found in either fresh or prolonged storage of pure hedyotide B2 at room temperature. This suggests the involvement of external exopeptidases in the trimming process. The decomposition of plant materials are known to occur through several stages. It begins with leaching of water and releases of most water soluble compounds [86]. The trimming of hedyotide B2 is

probably facilitated by aminopeptidase liberated during this leaching process. Another early event is the breakdown of the plant materials which provide nutrients and surface area for bacteria and fungi colonization [86]. These microorganism produce a great diversity of enzymes possibly contributed to the degradation of hedyotide B2. Knowledge of these degradation pathways provides a better understanding about the half-life and potential applications of uncyclotides.

3.8 Antimicrobial activity

Hedyotide B1 and B2 show a clear difference in their antimicrobial actions. Hedyotide B1 was active against both Gram-positive (*S. salivarius*) and Gram-negative bacteria (*E. coli*) with moderate MIC value suggesting its defense-related functions. Hedyotide B2, however, was inactive against all tested strains. Previous studies have shown that linearization of cyclotide results in loss of activity [40, 87]. It is possible that backbone cyclization may provide an additional epitope on loop 6 important for biological activities. Though devoid of antimicrobial activity, the high expression level of hedyotide B2 in aerial tissues implying that it may serve crucial physiological functions in plants.

4. Conclusions

In summary, this study described an in-depth characterization of novel cyclotides, uncyclotides and CRPs from medicinally important *Hedyotis* species. It has greatly enhanced our knowledge on several aspects of cyclotides and uncyclotides i.e. genetic structures, disulfide pattern and tissue-specificity. In addition, we also provide basic understanding about the biodegradation of cyclotides and uncyclotides in plants. Knowledge of these degradation pathways provides a

better understanding about the half-life and potential applications of cyclotides in agriculture, therapeutics intervention and drug development.

Chapter 4

Novel Cyclotides from *Chassalia chartacea* and Effects of Methionine oxidation on Cyclotide Bioactivities

1. Introduction

Within the Rubiaceae, the *Chassalia* is amongst the earliest genera discovered to produce cyclotides [17]. Subsequent studies on cyclotides, however, seldom characterized other species from this genus. Until now, cyclotides have been isolated from only two *Chassalia* species: *C. parviflolia* and *C. discolor.* The discovery of cyclotides in *C. parviflolia* originally based on anti-HIV bioassay-guided isolation [17], and led to the isolation of six novel cyclotides, circulin A-F. They exhibited anti-HIV activity that ranged from 50 to 275 nM depending on the viral strains and cell lines used in the assay [17, 88]. In a follow-up study, circulin A and B were successfully synthesized using SPPS and demonstrated to possess potent antimicrobial activity against a broad range of Gram-negative and Gram-positive bacteria [21]. The discovery of cyclotides in the second *Chassalia* species, *C. discolor,* was through a random screening program that aimed to understand the distribution and diversity of cyclotides in the Rubiaceae family [33]. One novel cyclotide was isolated from this species. Its biological activity, however, was not investigated further.

This chapter describes my work on the isolation and characterization of novel cyclotides and uncyclotides from *Chassalia chartacea* (synonym *C. curviflora*), a local species commonly found in Singapore forest. The plant is a medium-sized tree around 2-3 m height producing either white or red inflorescences (Fig. 26). It is used in Malay traditional medicine for treatment of malaria, coughs, childbirth, cuts, wounds and ulcers [89]. Active principles responsible for these activities are unknown. My study has led to the discovery of 16 novel cyclotides and uncyclotides from *C. chartacea.* Their genetic structures, antimicrobial, anti-tumor and hemolytic activities were also investigated, which provided basic understanding about their biosynthetic pathway and structure-activity relationship.

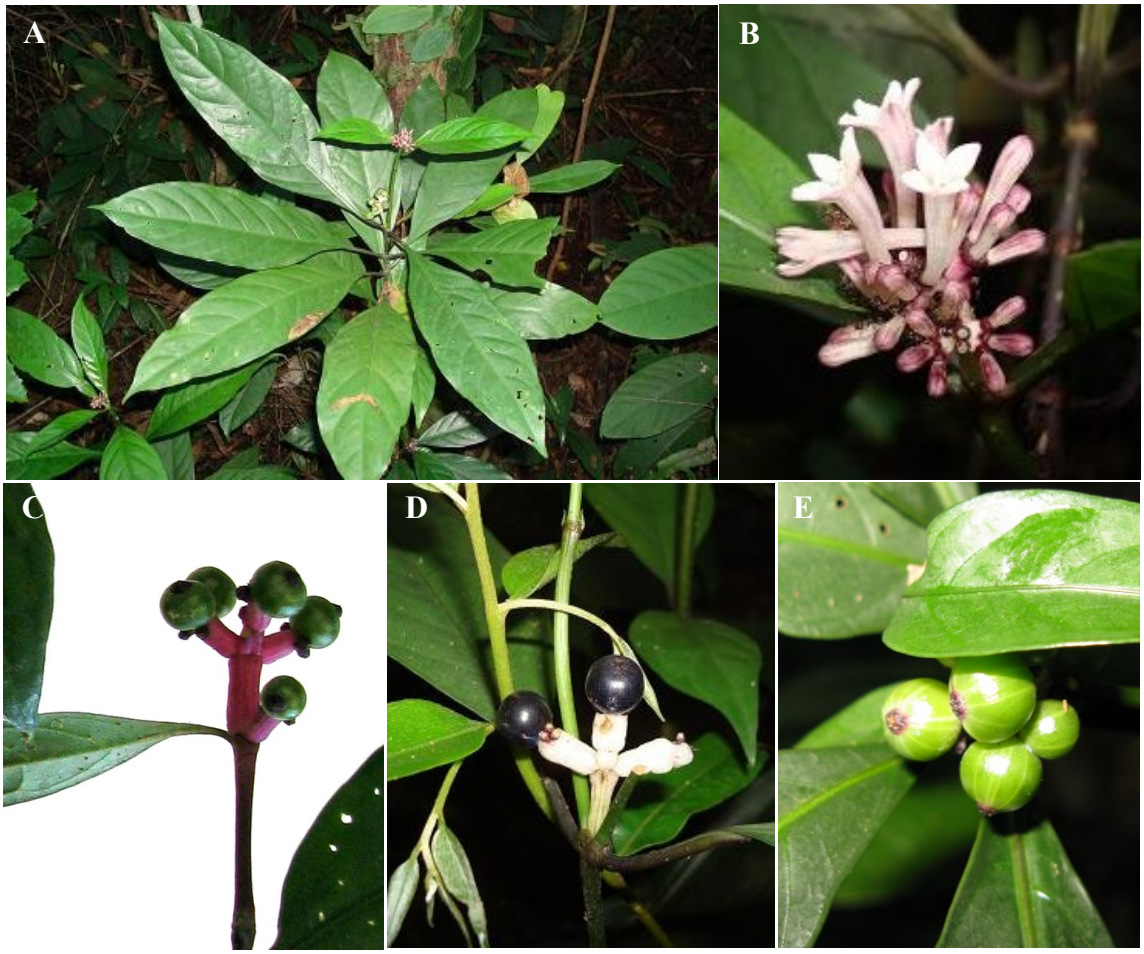


Figure 26. Chassalia chartacea. (A) *Chassalia chartacea* is a medium-sized tree around 2-3 meters tall. (B) Its inflorescence has pinkish white color. (C, D) There are two variations of *Chassalia chartacea* with red or white pedicel. (E) Young fruits have green color while the mature fruits are black. Figure taken from http://www.natureloveyou.sg.

2. Results

2.1 Isolation and discovery of novel cyclotides from *Chassalia chartacea*

Around 40 g of *C. chartacea* plant was collected and extracted with 10% EtOH. After several round of repetitive RP-HPLC, 15 peptides were isolated and sequenced by MS/MS, which comprised of circulin A, nine novel cyclotides, three novel uncyclotides and two Met-oxidation derivatives (Table 8). These novel CRPs were termed “Chassatides C” followed by numbers according to their order of discovery. Their primary sequences were confirmed by amino acid analysis, enzymatic digestion patterns and homology to known cyclotides.

Majority of novel chassatides belong to bracelet subfamily except two Möbius sequences (chassatide C1 and C12) and two hybrid sequences (chassatide C10 and C13). Although Möbius and hybrid cyclotides are not more difficult to be isolated or sequenced, they are less common in most examined species so far. Chassatide C10 and C13 are considered as hybrid in the sense that they share a shortened loop 3 similar to Möbius cyclotides, but lacking the *cis*-Pro residue in loop 5 like other bracelet cyclotides. Chassatide C7, C8 and C11 have open-end structures and thus classified as uncyclotides. They all lack the C-terminal Asn/Asp residue that is crucial for cyclization. These three peptides also do not possess the *cis*-Pro residue in loop 5, and therefore are categorized into bracelet subfamily.

Despite these variations, a number of cyclotide sequences appeared to be highly homologous between themselves or with known cyclotides. Chassatide C1 and C12, for example, differed by only one residue in loop 6 making them difficult to be separated in RP-HPLC. As the result, chassatide C1 and C12 were tested as a mixture in the subsequent bioassays. Other similar sets of chassatides such

chassatide C3 and C4, chassatide C5 and C6. These chassatides are also different from one another by a single residue.

Besides the characterization of novel cyclotides and uncyclotides, I also reported here the isolation of two Met-oxidized derivatives of chassatide C2 and C11, named as chassatide C2A and C11A, respectively. The oxidation of the Met residues in loop 3 resulted in a mass increase of 16 Da corresponding to the addition of an oxygen molecule. This oxidation event may have disrupted the hydrophobic patch formed by hydrophobic residues [90], leading to a decrease in hydrophobicity. As the result, the oxidized peptides were eluted earlier elution in RP-HPLC (Fig. 27).

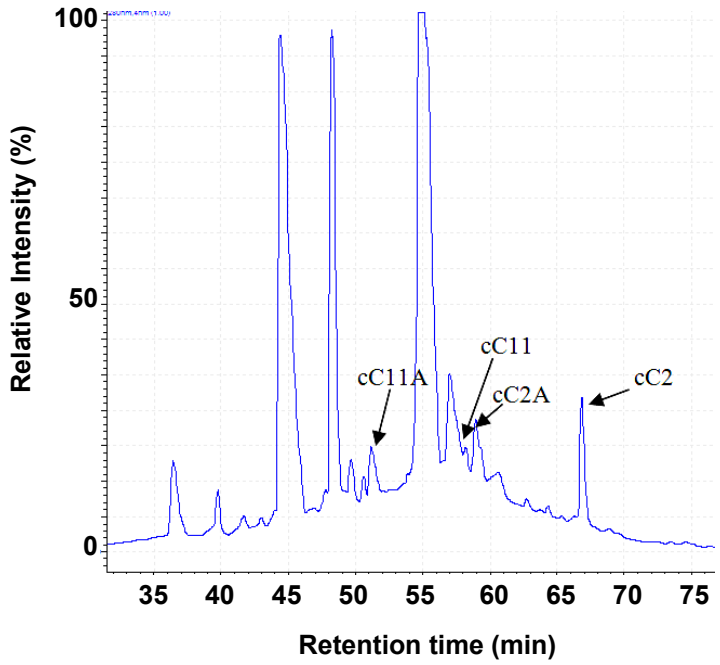


Figure 27. HPLC profile of Chassalia chartacea fruit. Relative hydrophobicity of Met-oxidized chassatide C2A and C11A was shown in comparison to their respective native peptides.

Table 8. Novel cyclotides and uncyclotides in *Chassalia chartacea*

Chassatide	Amino acid sequence	MW (Da)[a]	Structure
cC1	G-DACGETCFTGIC-FT--AGCSCNPWPTCTRN	2991	Möbius
cC2	G-IPCAESCVWIPCTITALMGCSCKNNV-CYNN	3283	Bracelet
cC2A	G-IPCAESCVWIPCTITALMGCSCKNNV-CYNN	3299	Bracelet
cC3	G-IPCGESCVWIPC-ISSALGCSCKNKV-CYRN	3166	Bracelet
cC4[b]	G-IPCGESCVWIPC-ISAALGCSCKNKV-CYRN	3150	Bracelet
cC5	GVIPCGESCVFIPC-ISSVVGCSCKNKV-CYRN	3240	Bracelet
cC6	GVIPCGESCVFIPC-ISSVIGCSCKNKV-CYRN	3254	Bracelet
cC7[c]	IPCGESCVWIPC-ITAIAGCSCKNKV-CYT	2956	Bracelet
cC8[c]	AIPCGESCVWIPC-ISTVIGCSCSNKV-CYR	3085	Bracelet
cC9	G-IPCGEACDFIPC-VTTVIGCSCKDKV-CYNN	3128	Bracelet
cC10	G-EYCGESCYLIPC-FT--PGCYCVSRQ-CVNKN	3212	Hybrid
cC11[c]	IPCGESCVWIPC-ISGMFGCSCKDKV-CYS	3009	Bracelet
cC11A[c]	IPCGESCVWIPC-ISGMFGCSCKDKV-CYS	3025	Bracelet
cC12	G-ASCGETCFTGIC-FT--AGCSCNPWPTCTRN	2963	Möbius
cC13	EYCGESCYLIPC-FT--PGCYCVSRQ-CVNKN	3155	Hybrid
cC14	G-FPCAESCVYIPCTVTALLGCSCRNRV-CYRN	3374	Bracelet
cC15	G-IPCAESCVYIPCTITALFGCSCKDKV-CYNN	3291	Bracelet
cC16[d]	G-IPCAESCVYIPCTITALLGCSCKDKV-CYKN	3271	Bracelet
cC17[d]	G-VPCAESCVYIPCTITALFGCSCKDKV-CYNN	3277	Bracelet
cC18[c,d]	IPCGESCVYIPC-ISAVLGCSCQNKV-CYR	3002	Bracelet
cC19[d]	G-IPCGESCVFIPC-ISALLGCSCSNKV-CYNN	3070	Bracelet

(a) Molecular weights are in monoisotopic masses. (b) Chassatide C4 has identical primary sequence to circulin A. (c) Chassatide C7, C8, C11 and C18 are uncyclotides. (d) Predicted sequences from cDNA. Cys residues are highlighted in yellow. Oxidized Met residues are in gray.

Cyclotide-expression profiles in *Chassalia chartacea*

Five different plant parts including leaves, stems, roots, pedicels, fruits and roots were extracted separately to study the tissue-specific distribution of chassatides in *C. chartacea*. As shown in Fig. 28, leaves and roots have strikingly different expression patterns. Leaves express predominantly two cyclotides, chassatide C1 and C12 which both are Möbius. On the contrary, roots produce mainly bracelet cyclotides and uncyclotides, with virtually undetectable presence of Möbius subfamily. Interestingly, fruits, pedicels and stems share similar MS profiles, and their chassatide contents appear to be a combined expression of both roots and leaves. This suggested these tissues are heavily protected indicative of their importance for plant survival.

2.2 Cloning of chassatide encoding genes from *Chassalia chartacea*

This work was performed in collaboration with Ms Nguyen Quoc Thuc Phuong. Encoding cDNA of chassatides was cloned by 3′-RACE PCR using degenerate primers targeted loop 1 sequences. Full length genes were subsequently obtained by 5′-RACE PCR. This resulted in identification of six full length clones encoding for chassatide C2, C8, C12, C14, C15 and C18 (novel), and five partial clones encoding for chassatide C7, C11, and three novel cyclotides, chassatide C16, C17 and C19. A premature stop codon precisely at the C-terminal Asn position is found in the precursors of chassatide C7, C8 and C11, which provides a likely explanation for their open-end structures. This is the first time cyclotide genes have been characterized from another genus, the *Chassalia*, besides the *Hedyotis* in the Rubiaceae family.

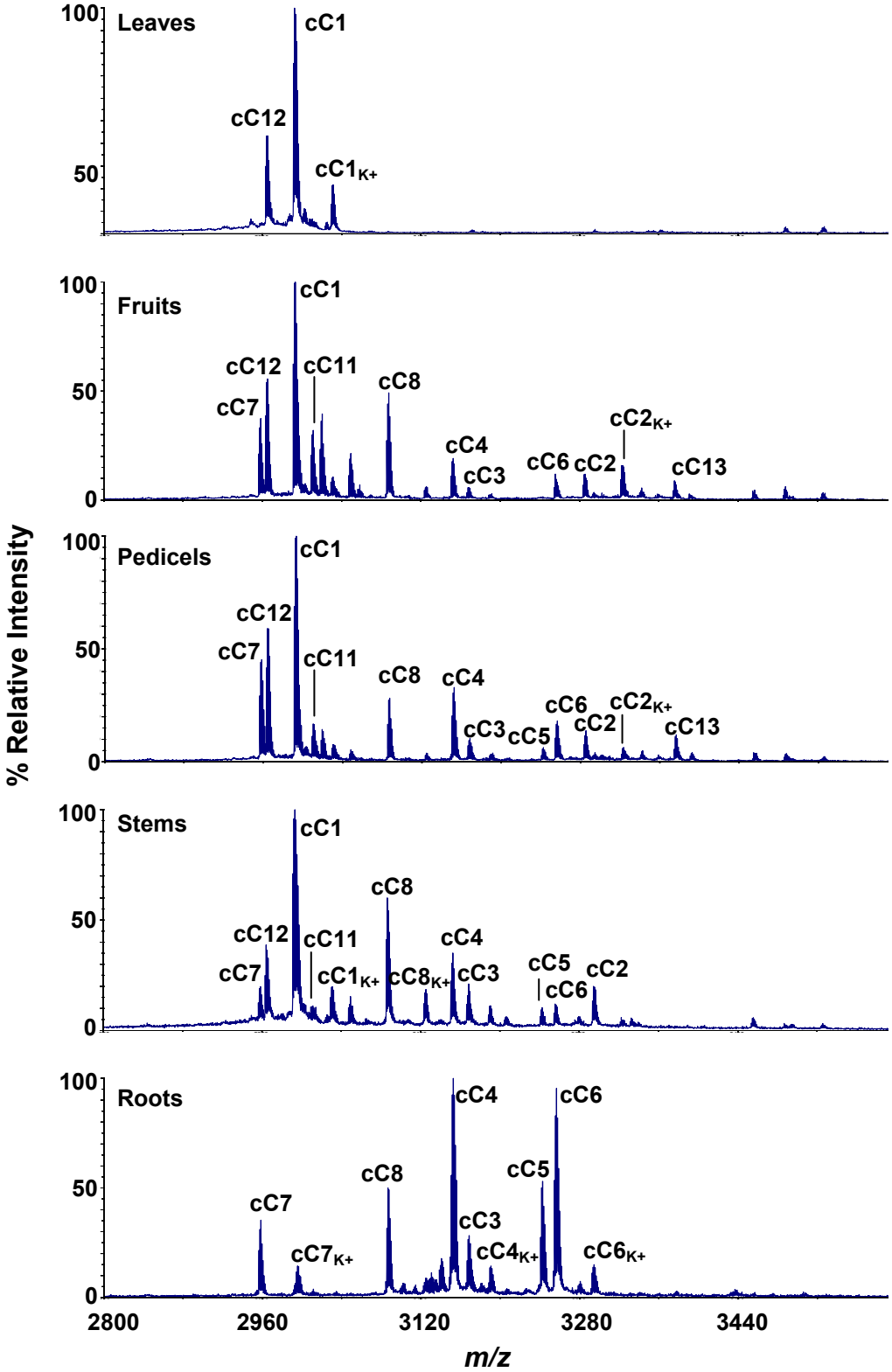


Figure 28. Tissue-specific distribution of chassatides in *Chassalia chartacea.* Plant specimens were divided into five different parts: leaves, stems, roots, pedicels and fruits. They were extracted and profiled separately by mass spectrometry.

Translated precursors of chassatide genes are shown in Fig. 29. Their primary sequences were aligned with precursor proteins of hedyotide B1 and B2 (*H. biflora*), kalata B1 and B7 (*O. affinis*), hedyotide D6 (*H. diffusa*) and hcf-1 (*H. centranthoides*), which showed a pronounced difference among these genes. Chassatide precursors are significantly shorter than the others with only 75-78 residues as compared to kalata B1 (124 residues) or hedyotide B1 (107 residues). They have the shortest precursors of all known cyclotides hitherto isolated. Their NTPP, NTR and CTPP domains are all reduced in size. The NTR domain in particular is over simplified and nearly absent in several chassatides such as chassatide C2, C12 and C14. It is interesting to note that their CTPPs comprises of only two residues “EL” as compared to at least four residues in most cyclotides.

2.3 Antimicrobial activity of novel chassatides

Antimicrobial activity of chassatides C1/C12 (tested as a mixture due to co-elution), C2, C4, C7, C8, C10, C11 and C11A were tested against three bacteria strains: *E. coli*, *S. aureus* and *S. epidermis*. D4R, an antimicrobial peptide synthesized in our laboratory, was used as a positive control. Chassatide C1/C12 and C2 were not effective against all tested strains. The remaining chassatides displayed moderate activity against *E. coli* with MIC value ranged between 3.72 μM to 10.57 μM (Fig. 30A, Table 9). They appear to be specific for Gram-negative bacteria and did not show any inhibitory activity on two tested Gram-positive strains *S. aureus* and *S. epidermis*.

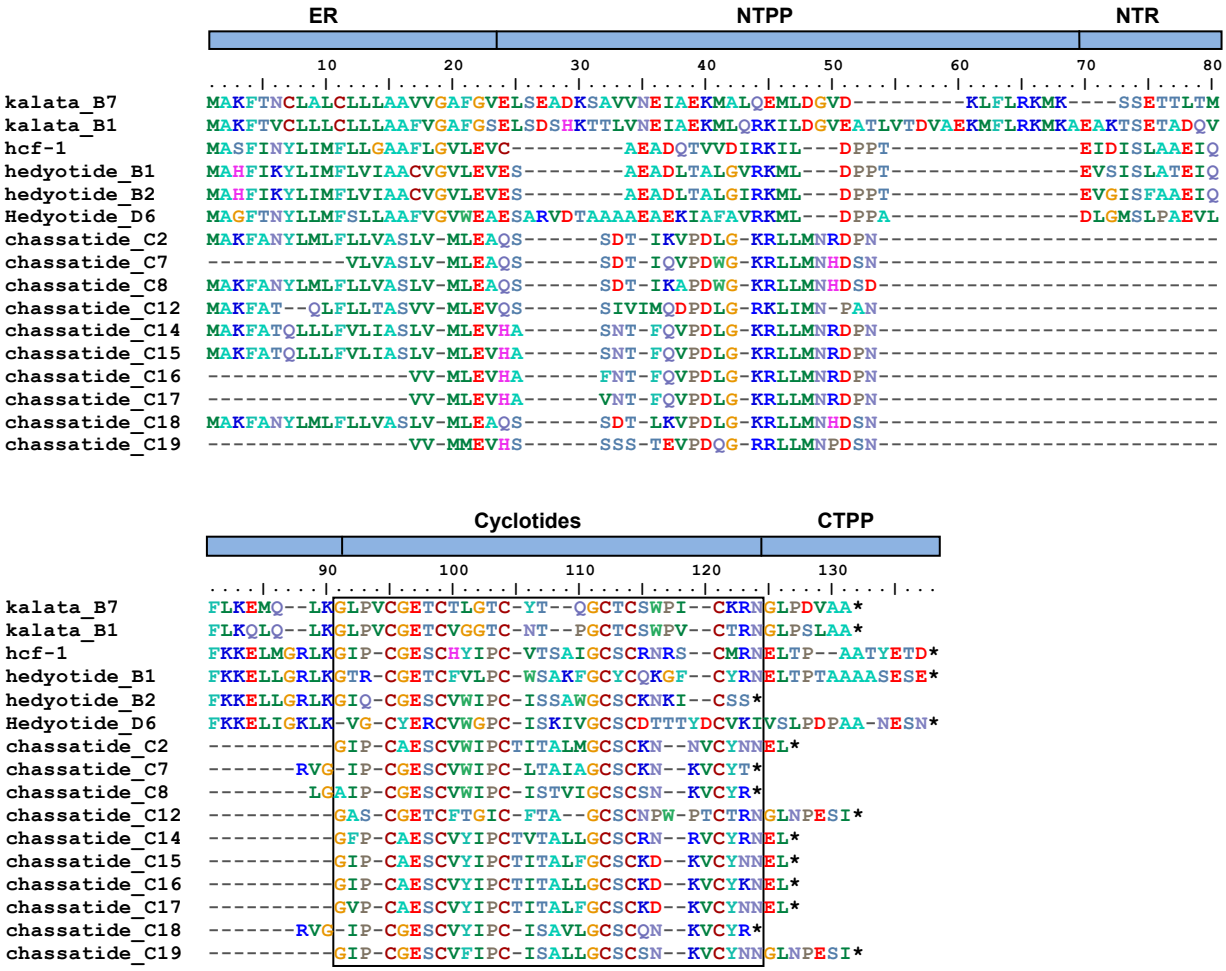


Figure 29. Multiple sequence alignment of chassatide precursors with kalata B1, kalata B7, hcf-1, hedyotide B1, B2 and D6 precursors. The mature cyclotide domains are boxed. Asterisk indicates stop codon.

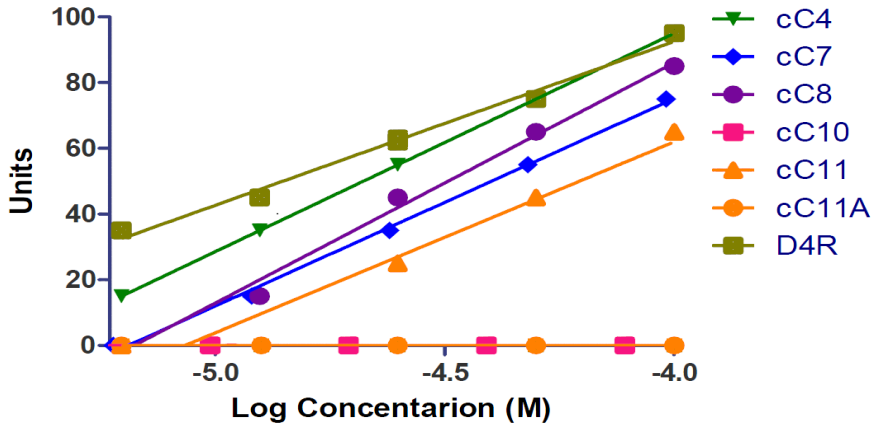
2.4 Cytotoxicity of novel chassatides

Cytotoxicity was assessed using the MTT assay, which measures living cell mitochondrial activity as an indicator of cell survival. Five concentrations (0.625 μM, 1.25 μM, 2.5 μM, 5 μM and 10 μM) within known cyclotide cytotoxic values were examined. Their IC_{50} interpolated from the dose-response profiles were summarized in Table 9. All the tested cyclotides showed very sharp dose response curves (Fig. 30B), in which 0-100% inhibition was generated within a 2-3-fold peptides dilution.

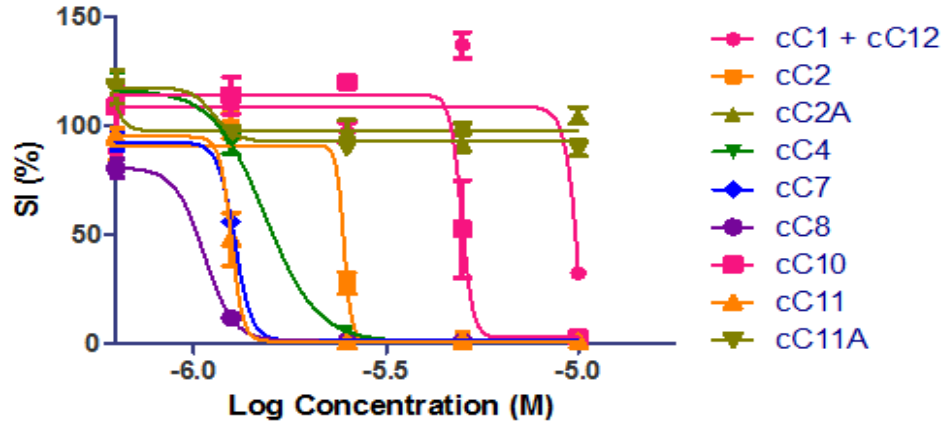
2.5 Hemolytic activity

Hemolytic effect of chassatides was conducted using human type A erythrocytes. Melittin, a known peptide from bee venom with strong hemolytic activity, was used as positive control. As shown in Fig. 30C, variations in primary sequences resulted in marked difference in hemolytic activity. Their HD_{50} ranged from 11.6 to 51.9 μM (Table 9), and were significantly less potent than melittin (3.21 μM). Among ten chassatides tested, three uncyclotides, chassatide C7, C8 and C11 were the most active peptides with HD_{50} of 11.6 μM, 25.5 μM and 13.3 μM, respectively. This is surprising since open-end variants of kalata B1 were totally inactive [87, 91]. These uncyclotides in this instance were not only retained the hemolytic activity but also the most potent one among tested chassatides.

A) Antimicrobial



B) Cytotoxicity



C) Hemolysis

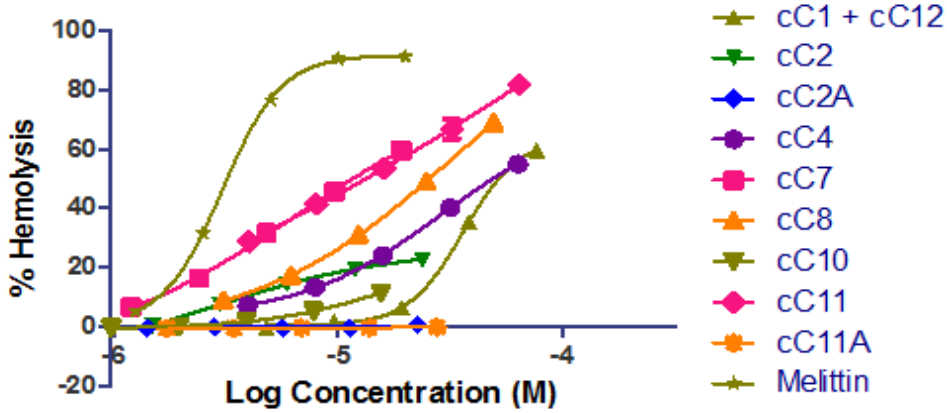


Figure 30. Biological activity studies of chassatides. (A) Antimicrobial activity of chassatides against *E. coli*. (B) Cytotoxicity of chassatides on Hela cell. (C) Hemolytic activity of chassatides on RBC. DiaGrams were drawn using GraphPad Prism. Each point represents the mean ± SEM for three replicates.

Table 9. Antimicrobial, cytotoxic and hemolytic effects of chassatides from *Chassalia chartacea*. Abbreviations are as follows: Not determined (ND), melittin (MT).

Chassatide	MIC (μM)			Cytotoxicity	Hemolysis
	Gram Negative	Gram Positive		IC_{50} (μM)	HD_{50} (μM)
	E. coli	*S. aureus*	*S. epidermis*		
cC1/12	>100	>100	>100	9.8	51.9
cC2	>100	>100	>100	2.4	>23.8
cC2A	ND	ND	ND	>10	>22.6
cC4	3.7	>100	>100	1.6	48.8
cC7	6.4	>100	>100	1.2	11.6
cC8	6.6	>100	>100	1.0	25.5
cC10	>100	>100	>100	5.0	>15.6
cC11	8.5	>100	>100	1.2	13.3
cC11A	>100	>100	>100	>10	>27.6
hA1	10.5	>100	>100	4.9	11.3
kB1	ND	ND	ND	4.8	18.6
MT	ND	ND	ND	ND	3.2
D4R	1.3	ND	ND	ND	ND

3. Discussion

3.1 Distribution of cyclotides in *Chassalia* genus

Chassalia chartacea is the third species of its genus found to produce cyclotides. Little is known about the distribution of cyclotides in this genus. Before this work, only three *Chassalia* species was screened including two cyclotide-positive plants , *C. parviflolia* and *C. discolor,* and one non-cyclotide containing plant, *C. Comm. ex Poir. sp* [17, 33]. Although the statistical sample size of the specimens is insufficient with only four *Chassalia* species tested, it appears that the occurrence of cyclotides in this genus is also sporadic similar to the *Hedyotis*. This suggests a complicated evolution and distribution of cyclotides in different genus of the Rubiaceae family.

Thus far, only seven cyclotides have been purified from *Chassalia* species of which six came from *C. parviflolia* and one from *C. discolor.* In-depth analysis of cyclotide content from *C. chartacea* in my work have led to the discovery of 16 new cyclotides and uncyclotides, which triples the number of cyclotides isolated in this genus. However, more works are clearly needed to unlock the cyclotide gold mine in the *Chassalia* which comprises of more than 70 species [33].

3.2 Genetic structure and biosynthetic processing of novel Chassatides

With no prior cyclotide genes being isolated from other genus besides the *Hedyotis* of the Rubiaceae family, it is of interest to characterize the cDNAs encoding for novel chassatides. Ten unique cDNAs were successfully amplified of which four encoded for novel cyclotides. Sequence comparison with known cyclotide genes from *O. affinis, H. biflora* and *H. centranthoides* revealed several interesting features about chassatide genes.

First, they have the shortest precursors of all known cyclotide clones. Chassatide C12, C14 and C15 are 37% to 39% shorter than kalata B1 precursor. The reduced size may provide an economic advantage for the plants as less energy will be required for cyclotides' production. This is a plausible explanation as *C. chartacea* is an understory tree where sun light is a limiting factor. When less light is available, less energy is available for the plants. For this reason, most understory plants are evolved to be efficient energy-users [92]. The reduced size of chassatide genes may be part of the shade-tolerant mechanism developed for plant adaptation to low-light condition.

Second, the NTR domain is absent in several if not all chassatides which is also the main reason accounted for the reduced size of chassatide precursors. This suggests that this domain is not essential for cyclotide biosynthesis [93]. *In vitro* experiment has also shown that it has no role in cyclotide folding and thus explaining its abridgement in chassatide precursors.

Third, the CTPPs of chassatide C2 and C14 to C17 consist of only two residues "EL", the shortest CTPPs documented for cyclotide genes. It has been previously proposed that the C-terminal tripeptide motif is essential for production of cyclic peptides [55]. This finding suggests that only two residues on the CTPPs are sufficient for the biosynthesis of cyclotides.

Four, both chassatide C14 and C16 precursors contain a tetrapeptide motif RNEL and KNEL at their C-termini, respectively. These peptide motifs resemble the classical ER retention signal, C-terminal KDEL sequence. In fact, mutagenesis studies have demonstrated that variants of the KDEL such as RDEL and KNEL can direct intracellular retention [94]. It is thus of interest to study if chassatide C14 and C16 precursor will be retained on ER or be processed for extracellular transport.

Fifth, residues at the N-terminal processing site of the cyclotide domain of chassatide genes are different from those of the *Hedyotis* origin. The N-terminal cleavage of chassatides occurs after the X-Asn dipeptide motif instead of the highly conserved Leu-Lys sequence in the *Hedyotis*. Although the X-Asn dipeptide motif has been found in several cyclotide genes of the Violaceae family, this is the first time it is reported for the Rubiaceae. The processing at the Asn residue also suggests that AEP may also be responsible for the N-terminal processing.

Next, I will discuss about the biosynthesis processing of three uncyclotides, chassatide C7, C8 and C11. Their synthesis processes are probably similar to hedyotide B2, in which genetic determination is the key factor explaining their open-end structures. These uncyclotides share several key features. First, a premature stop codon is found at the C-terminal Asn position which inhibits its translation. As mentioned in chapter 3, the C-terminal Asn residue has a crucial function in backbone cyclization and its deficiency will lead to uncyclotide formation. Second, N-terminal Gly residue is present in the precursor proteins of chassatide C7 and C11 but was trimmed off in the mature peptides. Like hedyotide B2, no mature products with N-terminal Gly were isolated. Interestingly, when N-terminal Gly is substituted by Ala in the case of chassatide C8, it was not cleaved off and remained in the mature sequence. This suggested the susceptibility of N-terminal Gly to exopeptidase degradations. However, more examples will be needed to establish a solid rule about uncyclotide formation.

3.3 Biological activities study of novel chassatides

Cyclotides have been proposed to exert their biological activities by permeabilization of the lipid membrane of the target cells, ultimately led to pore

formation and cell death [42]. In this work, the membrano-active property of novel chassatides was assessed by using antimicrobial, hemolysis and cytotoxicity assays, which evaluates their effects on four different membrane types of bacteria (Gram +ve and -ve), erythrocytes and HeLa cells.

Structure, net charge and hydrophobicity (Table 10) show varying degrees of correlation to the activity of chassatides. The end result of membrane permeabilization usually determines by the combined effects of electrostatic and hydrophobic interactions between peptides and target membranes. Before going into details about the structure-activity relationships of novel chassatides, I would like to discuss about the differences in the structures of different membrane types used in the study. First, the envelope of Gram-negative bacteria consists of outer and inner membranes, interspaced by a peptidoglycan layer [95]. Highly negative charged lipopolysaccharides (LPS) located on the outer membrane contributing to the overall negative charge of Gram-negative bacteria envelope [95]. Second, Gram-positive bacteria have no outer membrane instead compensating by a very thick peptidoglycan layer. Embedded in the cell wall are teichoic and teichuroic acids which give the overall negative charge of Gram-positive surface [96]. Third, the eukaryotic membranes of erythrocytes and HeLa cells are electrically neutral at physiological pH and stabilized by a high content of cholesterol [97]. Taken together, bacterial membranes are generally more negatively charged than mammalian membranes. Therefore, I hypothesize that electrostatic interaction would play a dominant role on peptide interaction with bacterial cells, whereas hydrophobic interaction would be the decisive factor on the peptide-induced damage of the mammalian membranes. The final outcome, however, will depend on both type of interactions and probably additional structural factors such as geometry, shape and

Table 10. Structure, net charge and hydrophobic ratio of *Chassatide*

Cyclotide	Structure	Net Charge	Hydrophobic Ratio (%)
cC1	Möbius	-1	41
cC2	Bracelet	0	51
cC2A	Bracelet	0	ND
cC3	Bracelet	+1	46
cC4	Bracelet	+1	50
cC5	Bracelet	+2	48
cC6	Bracelet	+2	48
cC7	Linear	+1	53
cC8	Linear	+1	51
cC9	Bracelet	-1	40
cC10	Hybrid	0	37
cC11	Linear	0	50
cC11A	Linear	0	ND
cC12	Möbius	0	41

Abbreviations: Chassatide C (cC). Not determined (ND)

topology of the peptides and the target membranes.

To demonstrate the above hypothesis, I would first examine the results from antimicrobial assay. Net charge or electrostatic interaction appears to be the dominant factor determined the antimicrobial activity. Except for chassatide C11, all the peptides with neutral or negative net charges are inactive against bacteria. Chassatides with +1 net charge display bactericidal activity on *E. coli*. Interestingly, none of the peptides are active against two tested Gram-positive strains. It is possible that the thick peptidoglycan layer of Gram-positive bacteria would have protected them from cellular damage by preventing the interaction between chassatides and the bacterial membranes.

For cytotoxicity, there is strong influence of hydrophobicity on chassatides activity. Those with hydrophobicity ratio more than 50% (chassatide C2, C4, C7, C8 and C11) are the most effective peptides. Consistently, chassatide C10 (37%) and chassatide C1/12 (41%) with hydrophobicity ratios less than 50% have weaker cytotoxicity. Net charge in this instance has less importance role. Chassatide C2, for example, displayed strong cytotoxicity with IC50 values of 2.44 μM despite its neutral charge.

Similar trend appears to be true for hemolytic activity. Those with strong cytotoxicity also display potent hemolytic effects. However, with the same hydrophobic ratio and net charge, their HD_{50} values do vary suggesting that additional factors may contribute for the activity. Chassatide C4 and C8, for examples, have a two-fold difference in their HD_{50} although they have same +1 net charge and comparable hydrophobicity.

3.4 Functional correlations of the cyclic backbone and biological activities

The discovery of three naturally occurring uncyclotides from *C. chartacea*, perhaps, is one of the most important finding of this work. Characterization of their biological activities will provide new understanding about functional correlations between linear and cyclic forms of cyclotides. Previous studies on synthetic linear variants of kalata B1 and the uncyclotide violacin A have reported loss or reduced hemolytic activity presumably due to their linear structures [24, 40]. Surprisingly, three uncyclotides isolated in *C. chartacea* not only maintained the membranolytic activities, but also were the most active of all tested chassatides. Chassatide C7, C8 and C11 exhibited stronger cytotoxic (IC_{50} 1.2 μM, 1.0 μM and 1.2 μM) and hemolytic (HD_{50} 11.6 μM, 25.5 μM, 13.3 μM) effects than several cyclotides such as chassatide C4 (IC_{50} 1.61 μM, HD_{50} 48.8 μM) and kalata B1 (IC_{50} 4.8 μM, HD_{50} 18.6 μM) under the same experimental conditions. This suggested that the cyclic backbone may be not an essential requirement for biological activities.

Indeed, cytotoxic activity has been reported for the uncyclotide psyle C in a recent study [61]. My works on chassatide C7, C8 and C11 have confirmed that uncyclotides possess a comparable membranolytic activity as their cyclic counterparts. The circular backbone though may provide an additional advantage in term of stability, it is not an indispensible functional requirement. In the case of violacin A, it seems that its hydrophilic nature is the main reason behind the low hemolytic value. For kalata B1, opening of its cyclic backbone may cause loss of a functional epitope important for activities. While it is still uncertain about the physiological roles of the cyclic backbone, it is clear that plants do display a wide variety of both linear and cyclic variants, and they are likely to have functional significance in plants.

3.5 Effects of methionine-oxidation on cyclotides bioactivities

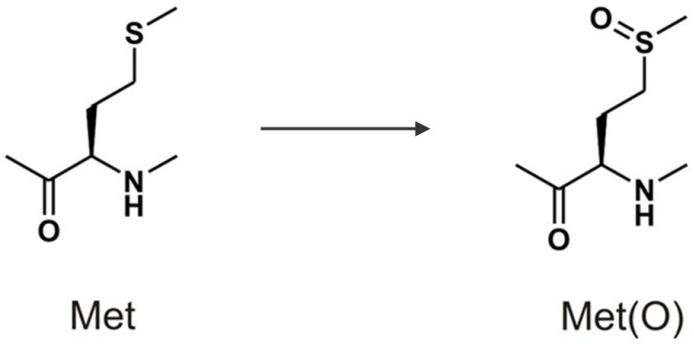


Figure 31. Oxidation of methionine to methionine sulfoxide.

The effect of Met-oxidation on cyclotides is another interesting topic observed in this study. The sulfur atom of methionine can be oxidized to sulfoxide by the addition of an oxygen atom (Fig. 31). This phenomenon has been observed in a number of native proteins and is especially relevant in aging tissues [98]. Oxidation of Met causes a significant change in its biophysical properties and often leads to alteration of protein functions. In the normal state, Met has long and non-polar side chain, which becomes polar upon oxidation. The hydrophobicity index of methionine sulfoxide (MetO) has been estimated to be similar to that of Asn [99]. Therefore, the oxidation of Met can be considered as a substitution of a hydrophobic for a hydrophilic amino acid, and is expected to have pronounced structural and functional consequences.

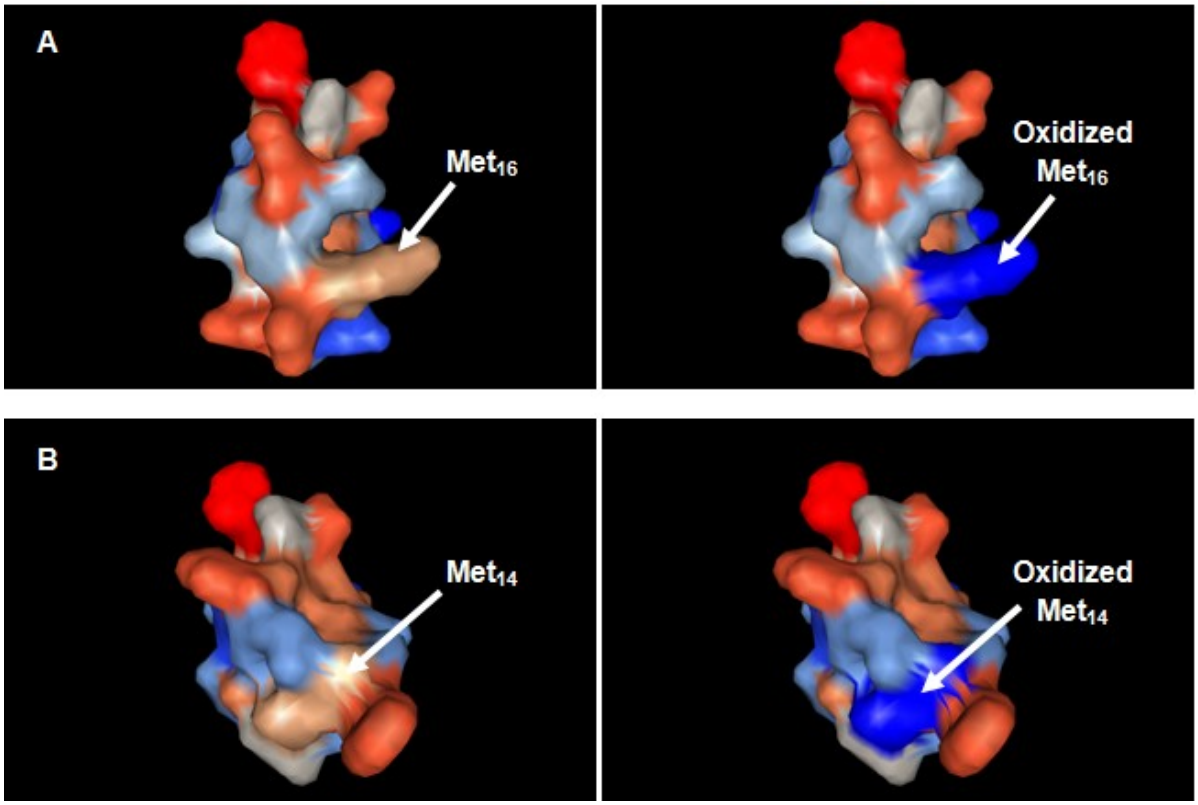


Figure 32. Effect of methionine oxidation on the hydrophobic patch. (A) Chassatide C2 and oxidized Chassatide C2A. (B) Chassatide C11 and oxidized Chassatide C11A. Red color indicates hydrophobic regions, while blue color indicates hydrophilic regions.

Computer models of chassatide C2A and C11A show a disruption of the hydrophobic patches in loop 3 due to Met-oxidation (Fig. 32). It has been proposed that bracelet cyclotides reply on patches of hydrophobic residues stretching over loop 2 and loop 3 for membrane interactions [29]. These hydrophobic patches are interrupted in oxidized chassatides C2A and C11A explaining why they displayed a complete loss of cytotoxic, hemolytic and antimicrobial properties. It should be noted that chassatides C2A and C11A possess cyclic and linear structures, respectively. The hydrophobic patches are thus important for both cyclotides and uncyclotides for exerting their biological activities. A similar phenomenon has been reported for Möbius cyclotides of which hydrophobic patches involve mainly on loop 2 and loop 5.

In this study, the oxidation occurred on Trp residue in loop 5 which also led to abrogation of hemolytic activity [60].

4. Conclusions

The work presented here contributes to our understanding about the diversity and distribution of cyclotides in the *Chassalia* genus. From *C. chartacea*, I have successfully isolated 16 novel cyclotides, two Met-oxidized derivatives, circulin A and ten unique cDNA clones. These novel chassatides display a tissue-specific distribution suggesting that they may have differential physiological functions in different plant parts. In addition, I also studied the medicinal values of novel chassatides by examining their antimicrobial, cytotoxic and hemolytic activities, which provides a more comprehensive knowledge about the membrano-active property of cyclotides.

Chapter 5

Morintides: an Unusual Class of Highly Anionic Cysteine-Rich Peptides from *Morinda Citrifolia*

1. Introduction

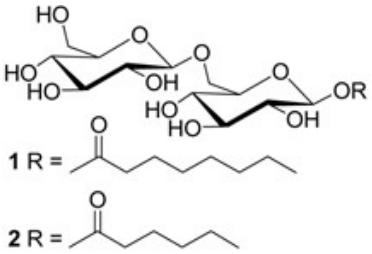
Morinda Citrifolia L. (Noni) is an evergreen tree (Fig. 33) distributed across the tropics and belongs to the tribe of Morindeae of the Rubiaceae family. It has been used in folk remedies for over 2000 years in Polynesia since the time they migrate from South-east Asia [100]. Almost every part of Noni plant is used by the Polynesians for treatment of a variety of illnesses such as arthritis, diabetes, high blood pressure, muscle aches, pains, menstrual difficulties, headaches, heart disease, AIDS, cancers, gastric ulcers, sprains, mental depression, senility, poor digestion, atherosclerosis, blood vessel problems, and drug addiction (28). In TCM, it is known as *Hai Ba Ji* and used to treat impotency, edema and back pains. It also has long traditional usages as folk medicine in India and the Pacific Islands [101]. Children in the island nation of Kiribati, for example, consumed the Noni leaves for treatment of night blindness. The plant is also eaten as raw or steamed vegetable in Indonesia, Thailand and several Asian countries [102].

Noni products mostly sold as fruit juice have been marketed in USA, Canada, Japan, Australia, Mexico, Norway and Hong Kong since 1996 as dietary supplements [101]. They were approved for selling in Europe as Novel Food in 2003. The popularity of Noni products grew rapidly attributed to claims for their “super

health“ benefits and a cure for all diseases. Today, there are over a hundred companies marketing Noni juice worldwide with estimated sale figures reach US$ 2 billion annually.

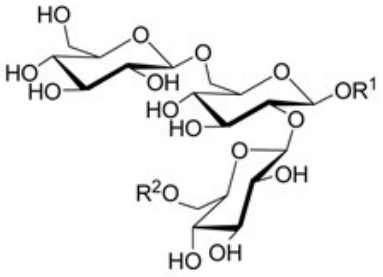
Figure 33. Morinda citrifolia (Noni) is a small-to-medium size tree belonged to Morindeae tribe of the Rubiaceae family. It produces small white flowers and edible fruits all year round. The plant is well-known for its rich medicinal values and has been used for over 2000 years in folk remedies by the native Polynesians. (Figure taken from Flora de Filipinas, Francisco Manuel Blanco)

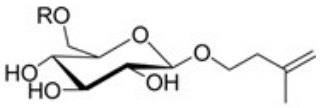
With multiple health benefits and a large market for juice products, Noni has attracted several medical researches with over 215 papers published since 2000 and 40 papers in year 2009 alone (search "Noni" or "Morinda citrifolia"; ISI Web of Knowledge, November 2010). Pharmacological studies on Noni juice have led to identification of numerous secondary metabolites containing a wide spectrum of anthraquinones, glycosides, flavonoids and iridoids (Fig. 34).



1 R =

2 R =





8 R = H

9 R = Glc

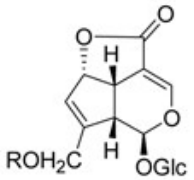
3 R^1 = R^2 = H

4 R^1 = R^2 = H

5 R^1 = R^2 = H

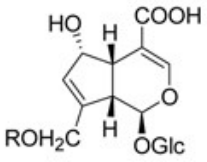
6 R^1 = R^2 =

7 R^1 and R^2 =



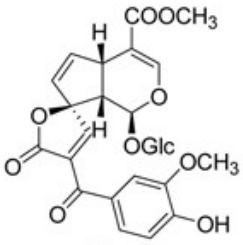
10 R = $COCH_3$

13 R = H

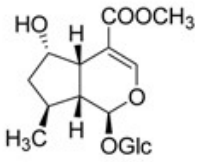


11 R = $COCH_3$

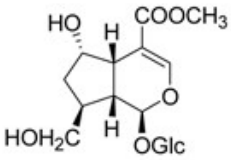
12 R = H



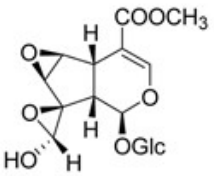
14



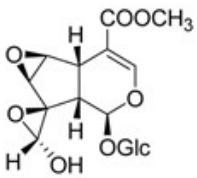
15



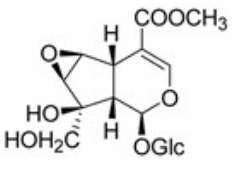
16



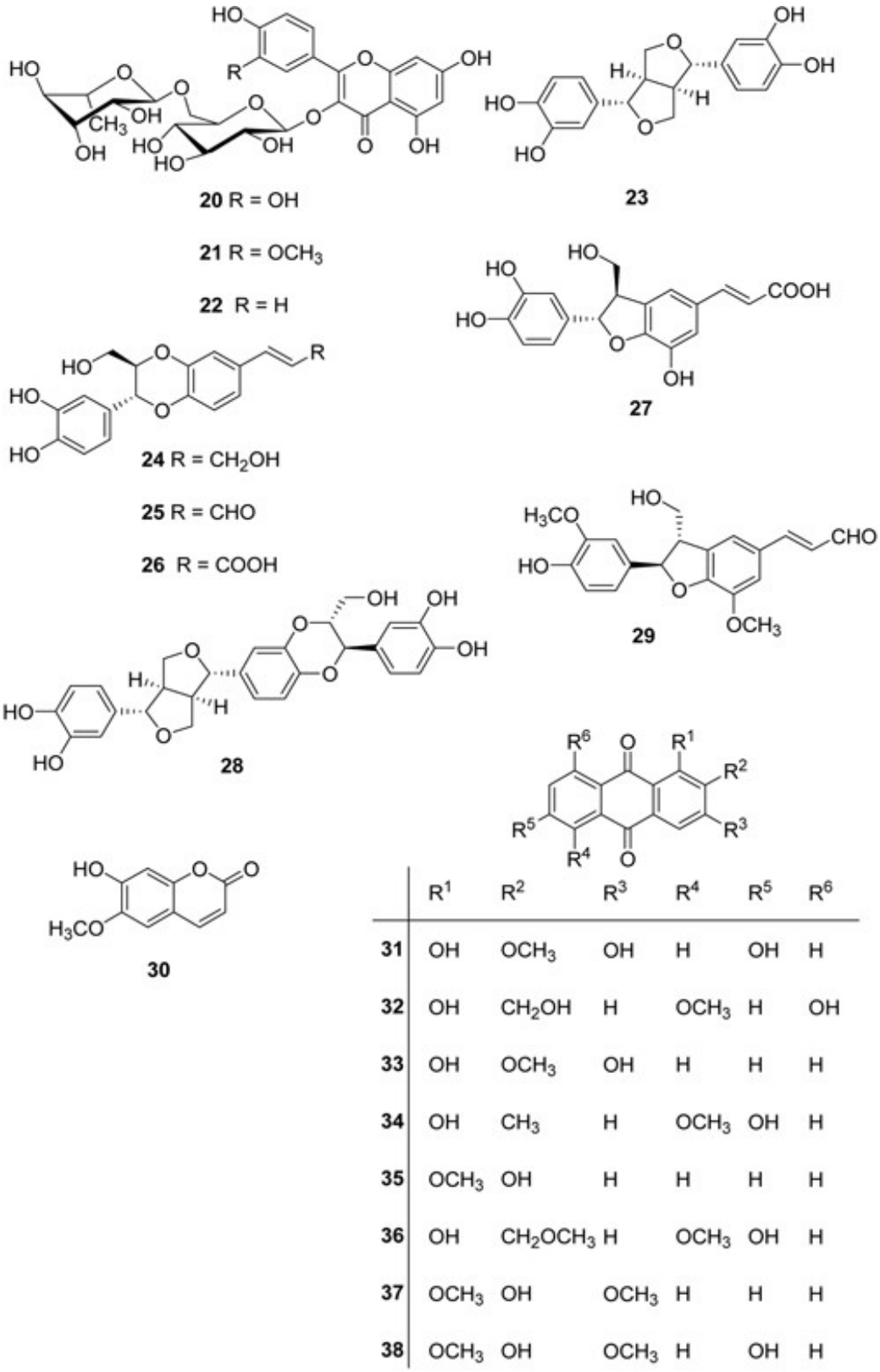
17



18



19



	R1	R2	R3	R4	R5	R6
31	OH	OCH_3	OH	H	OH	H
32	OH	CH_2OH	H	OCH_3	H	OH
33	OH	OCH_3	OH	H	H	H
34	OH	CH_3	H	OCH_3	OH	H
35	OCH_3	OH	H	H	H	H
36	OH	CH_2OCH_3	H	OCH_3	OH	H
37	OCH_3	OH	OCH_3	H	H	H
38	OCH_3	OH	OCH_3	H	OH	H

Figure 34. Secondary metabolites identified in Noni fruits contained a wide spectrum of fatty acids, glycosides, anthraquinones, flavonoids and iridoids. *Figure taken from Potterat et al.[103].*

Despite these extensive researches, active components characterized from the fruits are mostly restricted to small molecules. No bioactive peptides or CRPs have been isolated in this plant. Since Noni belongs to the Rubiaceae family, preliminary attempts to screen for the occurrence of cyclotides have been carried out but failed to detect their presence [33]. As such, I aim to screen, isolate and characterize if other CRPs besides cyclotides may be present in this medicinal plant. My work has led to the discovery of a suite of novel anionic cysteine-rich peptides containing three disulfide bonds and unusually low isoelectric points. Their antimicrobial, cytotoxicity, hemolytic and immunomodulatory activities were also investigated.

2. Results

2.1 Screening of novel CRPs in *Morinda citrifolia*

Roots, barks, leaves, flowers and fruits of *M. citrifolia* were extracted with 50% ethanol and profiled by MS. Peaks within the region of 2500 Da to 5000 Da corresponding to putative CRPs were documented. Their monoisotopic masses were presented in Table 11, and their respective MS spectra were shown in Fig. 35. Most of these compounds displayed a mass shift of 348 Da upon S-alkylation with IAA suggested the presence of three disulfide linkages indicative of CRPs.

A marked difference in CRPs expression was noted between different plant parts indicative of their tissue-specificity. Leaves and fruits produce a rich source of CRPs with at least 30 unique masses, of which several are tissue specific. Flowers, barks and roots produce less diverse number of CRPs. Barks and roots have similar CRP profiles whereas most CRPs expressed in flowers are unique suggestive of their specialized functions.

2.2 Isolation, purification and sequencing of morintides

Eleven novel CRPs were purified from *M. citrifolia* using the same procedure as that employed for cyclotides as described in chapter 3 and 4. They were designated as morintide C1 to C11. These novel morintides share several features with cyclotides such as having three disulfide bonds, mass ranges of 3000-4000 Da and eluting between 20-60% ACN in RP-HPLC (Fig. 36). They are, however, neither cyclotides nor uncyclotides with completely different primary sequences as determined by tandem mass spectrometry.

Table 11. List of CRPs expressed in different tissues of M. citrifolia. Molecular weights were reported as monoisotopic mass.

Tissue Parts	Mass Peaks
Fruits	3166, 3207, 3240, 3254, 3322, 3358, 3364, 3429, 3446, 3472, 3486, 3508, 3514, 3551, 3601,3626, 3697, 3728, 3735, 3750, 3751, 3766, 3777, 3779, 3788, 3792, 3807, 3817, 3828, 3846,3851, 3871, 3884, 3890, 3900, 3918, 3936, 3947, 3976
Leaves	3166, 3192, 3207,3240, 3322, 3349, 3364, 3470, 3513, 3537, 3551, 3566, 3582, 3626, 3748, 3828, 3846, 3854, 3870, 3888, 3900, 3918, 3964, 3981, 3999, 4007, 4024, 4041
Flowers	3166, 3322, 3360, 3467, 3495, 3505, 3620 3697, 3735,3766, 3773, 3806, 3846, 3886, 3918, 393,
Barks	3207, 3268, 3324, 3349, 3364, 3403, 3513, 3531, 3551, 3771, 3779,3788, 3807, 3811, 3828, 3846, 3903, 3918
Roots	2895, 2917, 3513, 3551, 3568, 3582, 3828, 3846, 3873, 3884, 3900, 3918, 3981, 3999, 4024, 4038, 4041

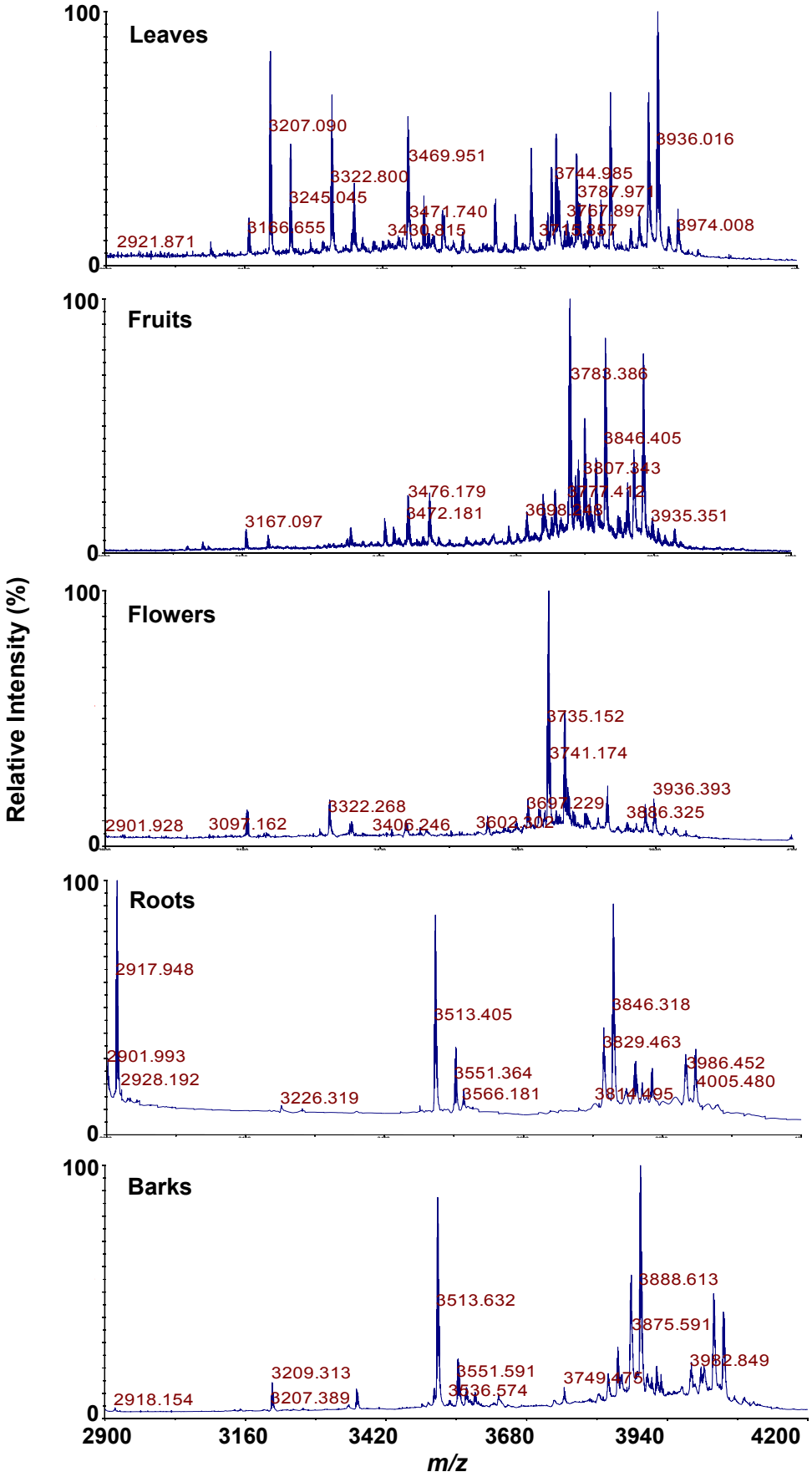
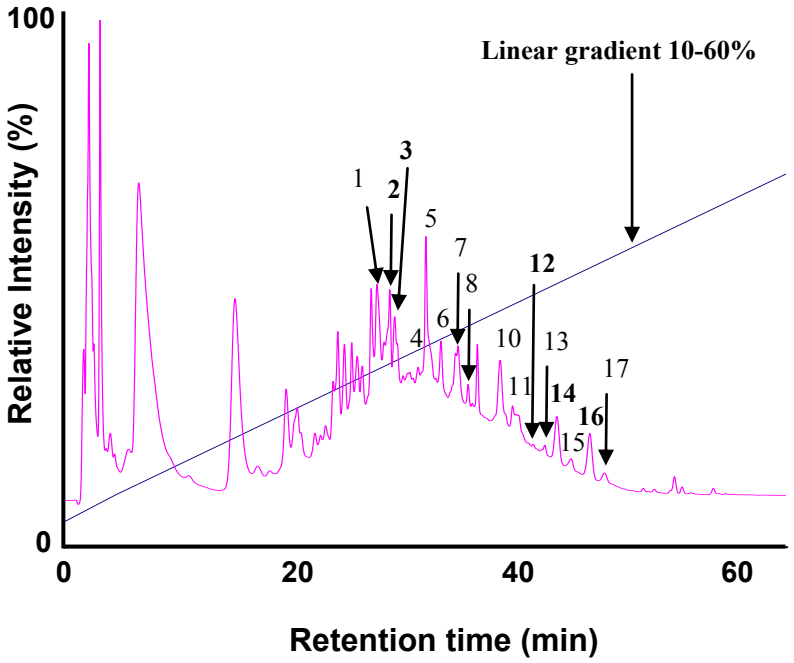


Figure 35. Tissue-specific profiling of M. citrifolia by MS. Five different plant parts were collected including leaves, fruits, flowers, barks and roots. Each tissue sample were extracted separately and profiled by MS.



Peak	Mass (Da)
1	3322, 3828
2	**3846**
3	**3918**
4	**3364**, 3828, **3918**
5	3728, 3766, 3871, 3890
6	3750
7	3551, 3728
8	**3513**
9	3792
10	3240, 3254
11	3166, 3182, 3780, 3802, 3818, 3851
12	**3779**
13	3133
14	**3751**
15	3789, 3822
16	**3777**
17	3792, 3817, 3846

Figure 36. RP-HPLC profile of M. citrifolia fruit demonstrated the presence of a wide array of CRPs . Fruits were extracted with 10% ethanol and the supernatant was fractionated by RP-HPLC using a linear gradient from 10-60% ACN for 60 min. CRP masses are presented in the table. Due to coelution, some peaks contain multiple masses. Bolded peaks and masses represent CRPs purified and identified as morintides in this report.

MS/MS spectrum of morintide C1 and C2 are shown in Fig. 37 and Fig. 38 as examples. To determine their primary structures, native peptides were first S-reduced with DTT, followed by enzymatic digestion. EndoGlu-C digest of the S-reduced morintide C1 yielded two major fragments of masses 2165 Da and 1610 Da (Fig. 36A). Sequencing of the 2165 Da fragment gave the N-terminal half of morintide C1 (Fig. 36B) as DDECPATLCSSDDDCGGCKCE. Analysis of the 1610 Da fragment gave the C- terminal half (Fig. 36C) as TFGLPLVLGSCYILD. Collectively, its primary sequence was determined as DDECPATLCSSDDDCG-

GCKCETFGLPLVLGSCYILD. Likewise, EndoGlu-C digestion of morintide C3 resulted in two fragments with masses of 2174 Da and 1629 Da (Fig. 37), which together gave the sequence of morintide C3 as DNVCPDTLCSSDADCGPCK-CETFGLPLSMGSCYILQ. The primary structures of remaining morintides were determined in a similar manner (Table 12). No cyclotide was found in the fruits or leaves, consistent with earlier reports of their absence [33].

To confirm the MS/MS-derived sequences, Edman degradations were carried out for morintide C2 and C5. This helped to verify the primary structure of morintide C2, but not morintide C5 due to the N-terminal block consistent with the presence of pyroglutamyl residue at its N-terminus. For other morintides, their sequences were verified by a combination of amino acid analysis, enzymatic digestion using trypsin and chymotrypsin and homology assignment based on morintide C2 sequence.

Novel morintides have open-end structures and are unusually rich of acidic residues Asp and Glu giving them a net negative charge ranging from -1 to -7. They do not show homology to any known CRPs and thus represent a novel family of anionic cysteine-rich peptides. To my knowledge, there are no known CRPs with net charge below -5 have been isolated. Morintide C1 and C2 are thus the most anionic CRPs with net charges at -7. They both contain only one basic residue and up to eight acidic residues while having only 36 amino acids in size. This is the first report of CRPs with such anionicity being characterized.

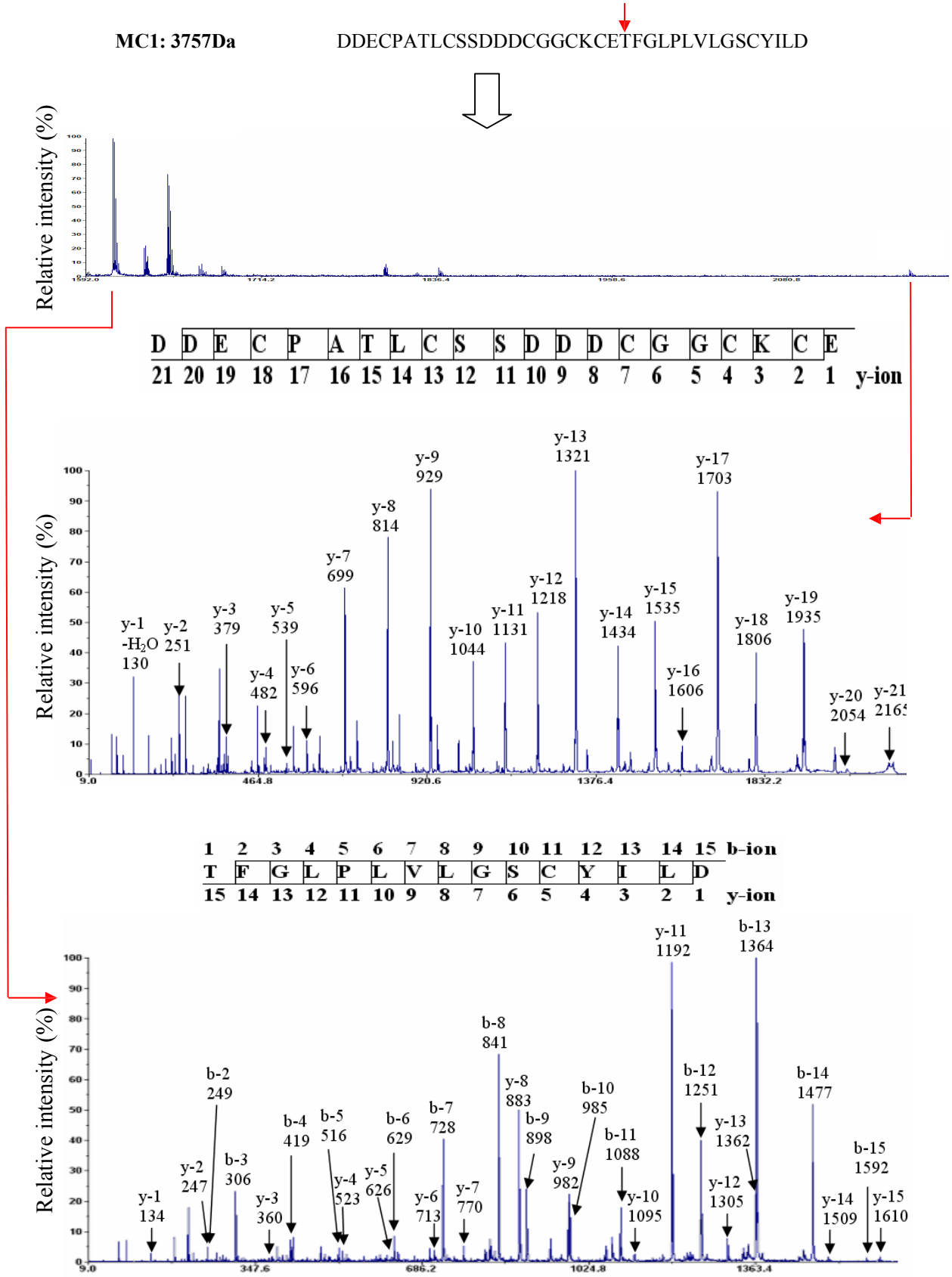


Figure 37. MS/MS sequencing of morintide C1. **A.** Enzymatic cleavage of morintide C1 by EndoGlu-C. The resultant ions were fragmented by MS/MS. **B.** MS/MS spectra of the 2165-Da-fragment. The y-ion sequence is shown. **C.** MS/MS spectra of the 1610-Da-fragment. Both b- and y-ion series are shown. Combining both sets of sequences gave the overall sequence of MC1 as DDECPATLCSSDDDCGGCKCETFGLPLVLGSCYILD.

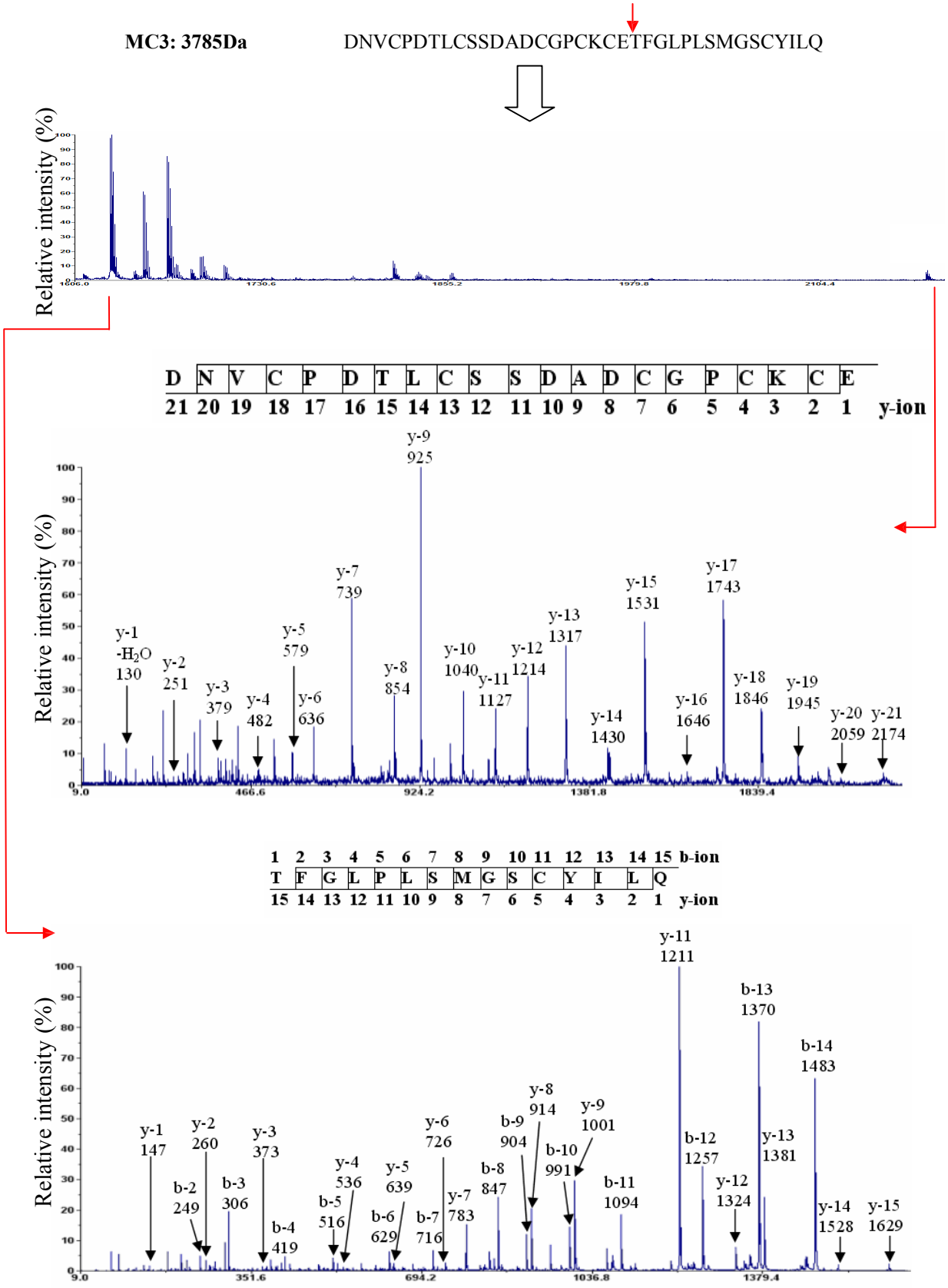


Figure 38. MS/MS sequencing of morintide C3. A. Enzymatic cleavage of morintide C3 by EndoGlu-C. The resultant ions were fragmented by MS/MS. **B.** MS/MS spectra of the 2174-Da-fragment. The y-ion sequence is shown. **C.** MS/MS spectra of the 1629-Da-fragment. Both b- and y-ion series are shown. Combining both sets of sequences gave the overall sequence of MC3 as DNVCPDTLCSSDADCGPCKCETFGLPLSMGSCYILQ.

Table 12. Summary of morintides from Morinda citrifolia. Their respective molecular weights, net charge, yields and disulfides are shown. Morintide C1-5 were purified from fruits, and morintide C6-11 were isolated from leaves.

Morintide	Primary sequence	MW[1]	Net Charge	Yield (mg/kg)
mC1	DDECPATLCSSDDDCGG--C-KCETFGLPLVLGSCYILD	3751	-7	5
mC2	DDECPAILCTSDDDCGG--C-KCETFGLPLVLGSCYILD	3777	-7	10
mC3	DNVCPDTLCSSDADCGP--C-KCETFGLPLSMGSCYILQ	3779	-5	5
mC4	DVLCPNIQCTADYQCGPY-C-DCEPPDGLRR---CYWVG	3846	-3	5
mC5	Q^{+}DLCPDIQCTEDYECGPY-C-DCEPPDGLRR---CYWVG	3918	-6	10
mC6	ECVNVPCDNDIFCQKLGCIQCVVPIIGGKV--CV	3403	-1	-
mC7	QECVNVPCDNDIFCQKLGCIQCVVPIIGGKV--CV	3531	-1	-
mC8	Q^{+}ECVNVPCDNDIFCQKLGCIQCVVPIIGGKV--CV	3513	-1	1
mC9[2]	QECVNVPCDSDIFCQKLGCIQCVVPIIGGKV--CV	3504	-1	-
mC10	RLCPRPFCTDDSLCGP--C-KCVIYSGLDYGG-CE	3364	-1	-
mC11	LCPRPFCTDDSLCGP--C-KCVIYSGLDYGG-CE	3214	-2	-

[1] Monoisotopic mass (Da). [2] Sequence predicted from cDNA clone. Cysteine residues are highlighted in yellow. Negatively charged residues are in red and positively charged residues are in blue. Q^{+} indicates pyroglutamic acid

2.3 Genetic characterization of novel morintides

This work was performed in collaboration with Ms Nguyen Quoc Thuc Phuong and Mr Ye Weijian. Partial nucleotide sequence of morintide C6 was obtained by using degenerate primers targeted PCDNDLF sequence for 3′-RACE PCR. Primers based on 3′ UTR of morintide C6 cDNA were then used for 5′-RACE PCR. Interestingly, this led to the discovery of a new morintide named as morintide C7, which suggests that the 3′ UTR may be conserved for both morintides. This is not surprising as primary sequences of morintide C6 and C7 are highly similar and different from one another by a single residue. They may thus have highly homologous precursor proteins explaining their cross PCR amplification.

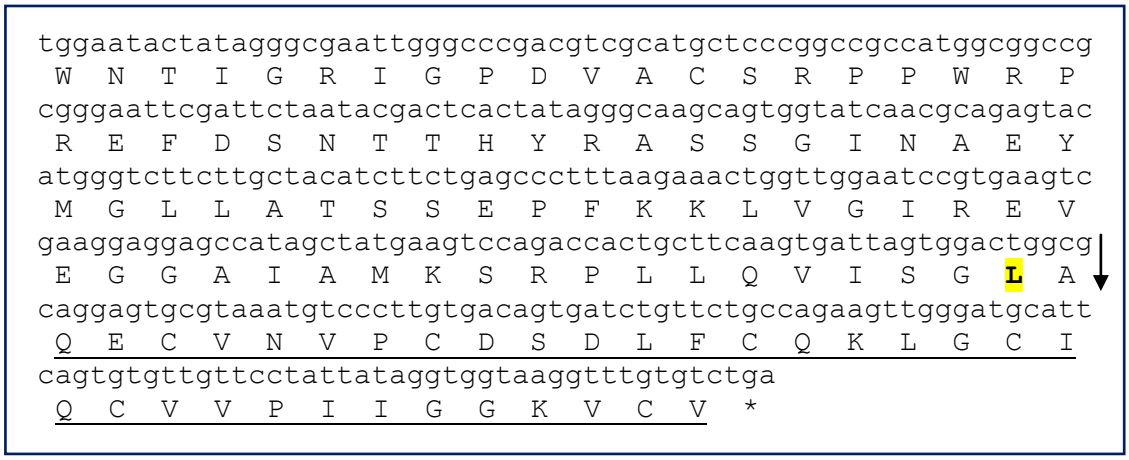


Figure 39. Translated precursor sequence of morintide C7. Predicted mature peptide domain is underlined. Arrow indicates the N-terminal processing site. The conserved Leu residue located at two residues away from the cleavage site is highlighted in yellow.

Translated amino acid sequence of morintide C7 was shown in Fig. 39. Based on its sequence similarity with morintide C6, the predicted mature sequence was Q^{+}ECVNVPCDSDLFCQKLGCIQCVVPIIGGKVCV. The precursor protein contained no C-terminal prodomain with a stop codon followed directly the mature peptide sequence. The cleavage site at the N-terminus of morintide C7 is similar to

that of cyclotides, which features a highly conserved Leu residue located at two amino acids away from the processing site.

2.4 Disulfide mapping of morintide C2

With no known homology, there is an essential need to elucidate the disulfide connectivity of novel morintides, which is important for understanding of their 3D structures. Morintide C2 was selected as a representative for disulfide mapping due to its abundance. The peptide was double S-tagged with IAA and NEM using step-wise partial reduction alkylation strategy followed by top-down MS/MS to elucidate the disulfide structures. The approach used here was similar to that used for mapping disulfides of hedyotide B2.

My first attempt for partial reduction of morintide C2 with TCEP at pH 3, 37^0C for 30 min, followed by 30 min S-alkylation with NEM did not result in significant intermediates. Under the same condition, approximately 40 % of hedyotide B2 has been reduced (Fig. 20, chapter 3). It appears that morintide C2 is more resistant to TCEP-mediated reduction than hedyotide B2 due to its high anionicity. Morintide C2 with eight acidic residues is likely to electrically repel TCEP which structure contains three carboxylic acid groups. The reduction temperature thus was raised to 45 degree and incubated for 60 min instead of 30 min. This yielded a mixture of intermediates containing 1SS species had MW of 4281 Da (one SS bond and four S-NEM groups) and 2SS species had MW of 4029 Da (two SS bond and two S-NEM groups). Despite numerous attempts to separate these intermediates, they tended to co-elute and partially eluted together with the native peptide and fully S-NEM labeled peptide (RN) indicating a similarity in hydrophobicity. These intermediates were only separated upon complete S-reduction with DTT and S-alkylation with IAA (Fig. 40B). This resulted in three chromatography-separated peaks containing double S-tagged

peptides. Two peaks had *m/z* at 4397 Da corresponding to two S-IAA and four S-NEM labeled groups; one peak had *m/z* at 4261 Da corresponding to four S-IAA and two S-NEM groups.

MS/MS fragmentations of these double S-tagged peptides enabled the identification of the S-alkylators on each cysteine residue, thereby allowing for determination of the cystine linkages (Fig 39C). Their MS/MS spectra were shown in Fig. 41. The sites of the S-NEM alkylation of the 4261-Da-peptide directly established the Cys I-IV connectivity. Similarly, the locations of the S-IAA alkylation of two 4397-Da-species established the connection of the remaining two disulfide bonds as Cys III-VI and Cys II-V. The combination of these data led to the complete disulfide connectivities of morintide C2 as Cys I-IV, Cys II-V and Cys III-VI, which formed a cystine-knot structure (Fig. 39D).

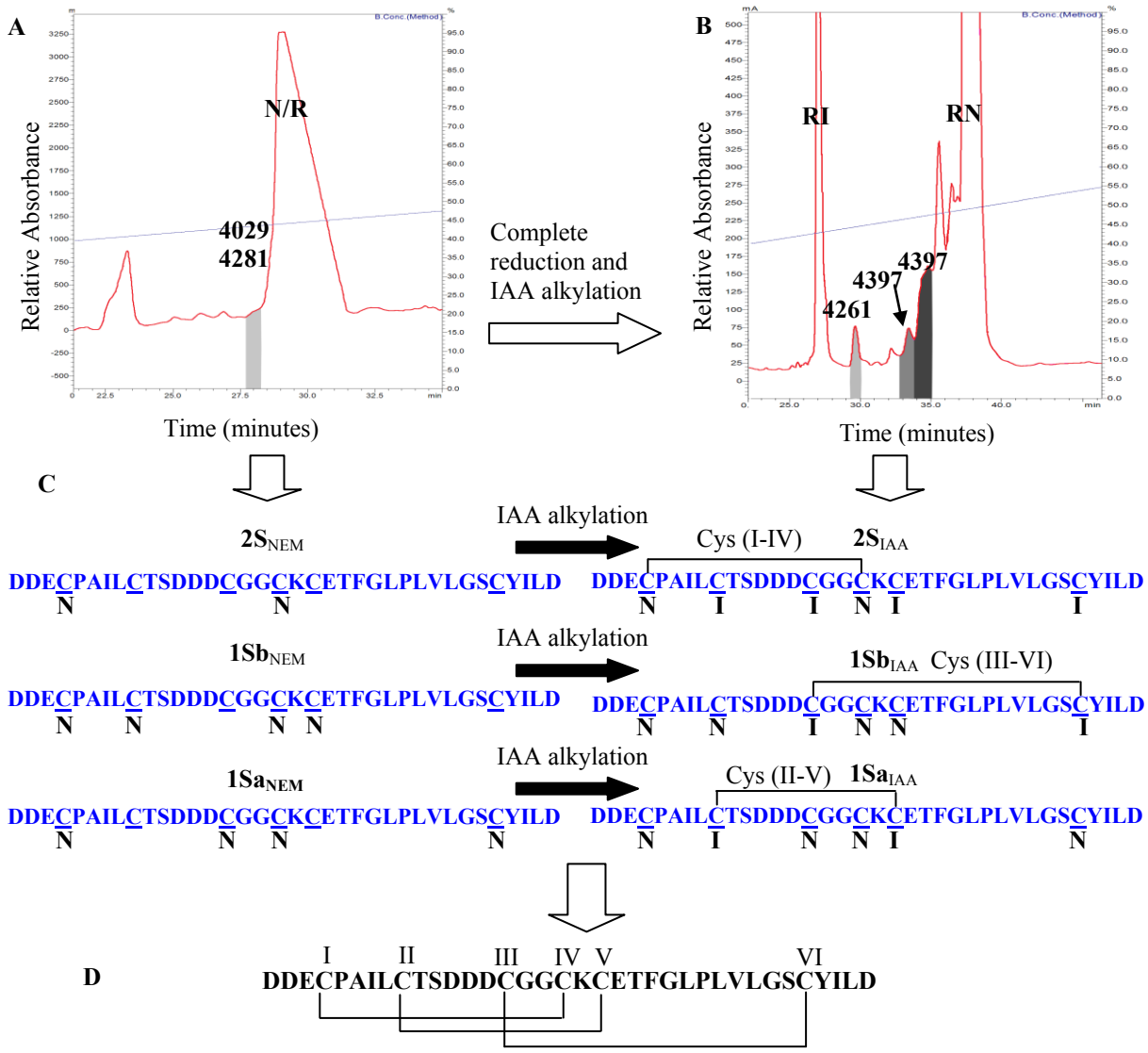


Figure 40. Disulfide mapping of morintide C2. (A) HPLC profile after partial reduction and alkylation of morintide C2 with TCEP and NEM, respectively. The partially reduced intermediates, 1SS and 2SS species, were co-eluted near the native (N) and fully reduced (R) peptides (highlighted in gray). (B) HPLC profile of the complete S-reduction and S-alkylation of the intermediates with DTT and IAA, respectively. **C)** Primary amino acid sequences of the intermediates tagged with NEM (N) and/or IAA (I) after determination of the sites of the different alkylators by MS/MS. **D)** Combination of the above data led to the deduction of the complete set of disulfides as Cys I-IV, Cys II-V and Cys III-VI.

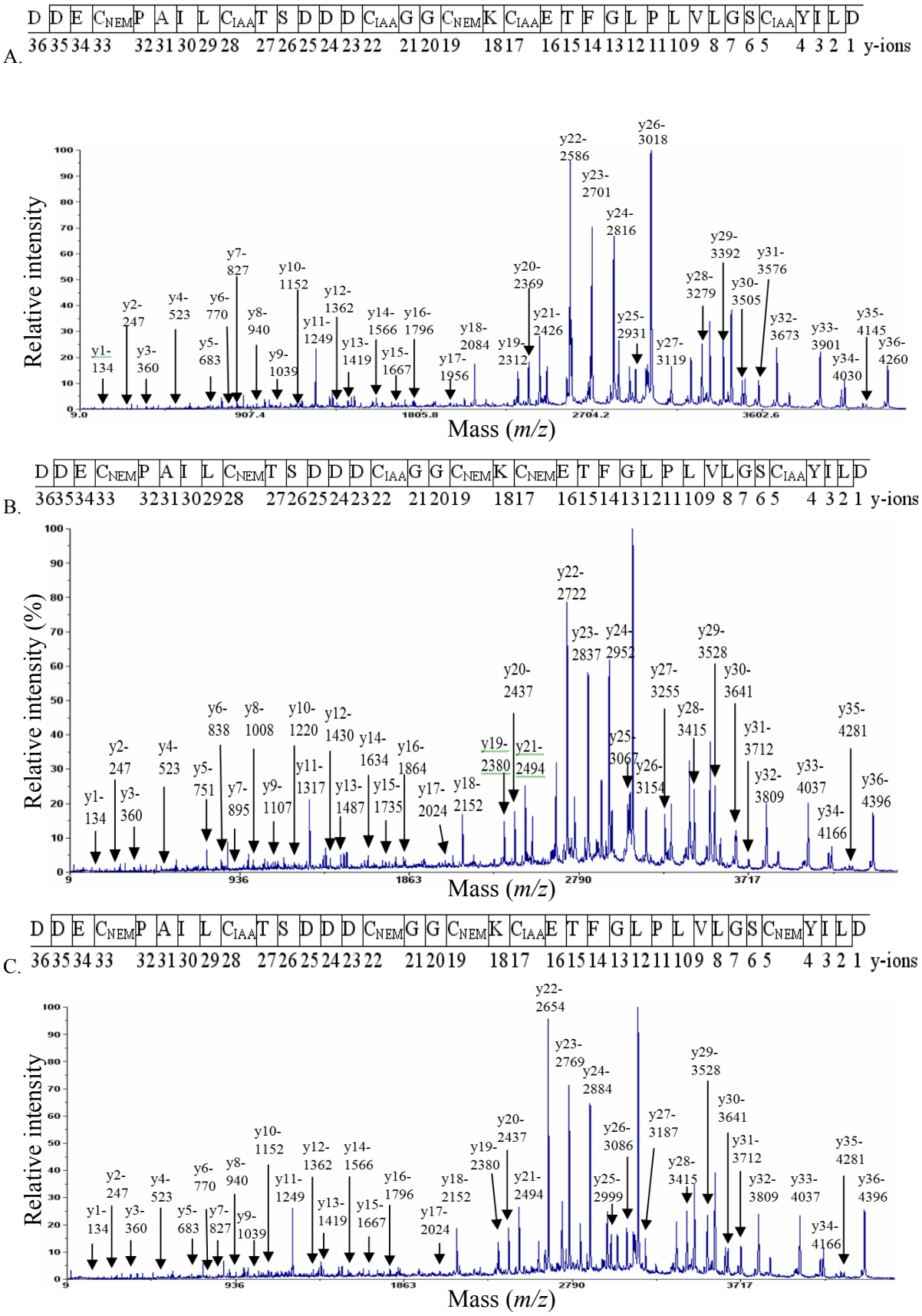


Figure 41. MS/MS spectra of double S-tagged morintide C2. C_{NEM} : S-NEM cysteine; C_{IAA}: S-IAA cysteine. A) MS/MS spectra of 4261-Da peptide showed a C_{NEM}-C_{IAA}-C_{IAA}-C_{NEM}-C_{IAA}-C_{IAA} alkylation pattern, establishing the Cys I-IV connectivity; B) MS/MS data of 4397-Da peptide showed a C_{NEM}-C_{NEM}-C_{IAA}-C_{NEM}-C_{NEM}-C_{IAA} alkylation pattern establishing the Cys III-VI connectivity; C) MS/MS data of 4397-Da peptide showed a C_{NEM}-C_{IAA}-C_{NEM}-C_{NEM}-C_{IAA}-C_{NEM} alkylation pattern establishing the Cys II-V connectivity.

2.5 Homology modeling

Computer models of morintides C1-5 were constructed based on NMR structures of kalata B1 [104]. The backbone, cystine-knot motif, charge and hydrophobic distribution of each morintide are shown in Figure 42. All peptides display a distorted, compact globular structure, where the interior is occupied by the cystine-knot. They have amphipathic structures with clustering of hydrophobic and polar residues.

2.6 Thermal and enzymatic stability tests

Given our aim of discovering druggable biologics, it necessitated that such peptides are chemically, thermally and enzymatically stable. As such, stability assays [105] were performed to quantitate the chemical, thermal and enzymatic resistance of morintides by comparing the relative RP-HPLC areas before and after the treatment.

Morintides were tested for their heat stability by incubation at 100^{o}C for a period of 10, 15, 30, 45 and 60 min. Their respective RP-HPLC profiles after 10 min and 30 min of heat treatments were shown in Fig. 43A and 43B as examples. Retention times of morintide C1 to C3 were unchanged, with less than 15% decline in peak intensity over 1 hr, indicating their heat stability. In contrast, morintide C4 and C5 were denatured as indicated by a retention-time shift and reduction in their native peak intensity. The percentage degradation of morintides C1, C2, C3, C4 and C5 after 60 min heating were as follows 15%, 6%, 3%, 70% and 88%, respectively (Fig. 43C).

Enzymatic stabilities of morintides were tested on four different proteases - trypsin, chymotrypsin, EndoGlu-C and pepsin. A linear synthetic peptide was used

as positive control. The HPLC traces of morintides after enzymatic digestions are shown in Fig. 44. Morintide C1, C2 and C3 were resistant to all enzymes tested experiencing less than 5% decline in concentration. Morintide C4 and C5, however, were unstable to trypsin, chymotrypsin and EndoGlu-C with over 80% peptide degraded after 1 hr. Surprisingly, MC4 and MC5 were resistant to pepsin. It is possible that the low pH condition using for pepsin digestion may confer some protection for the peptides.

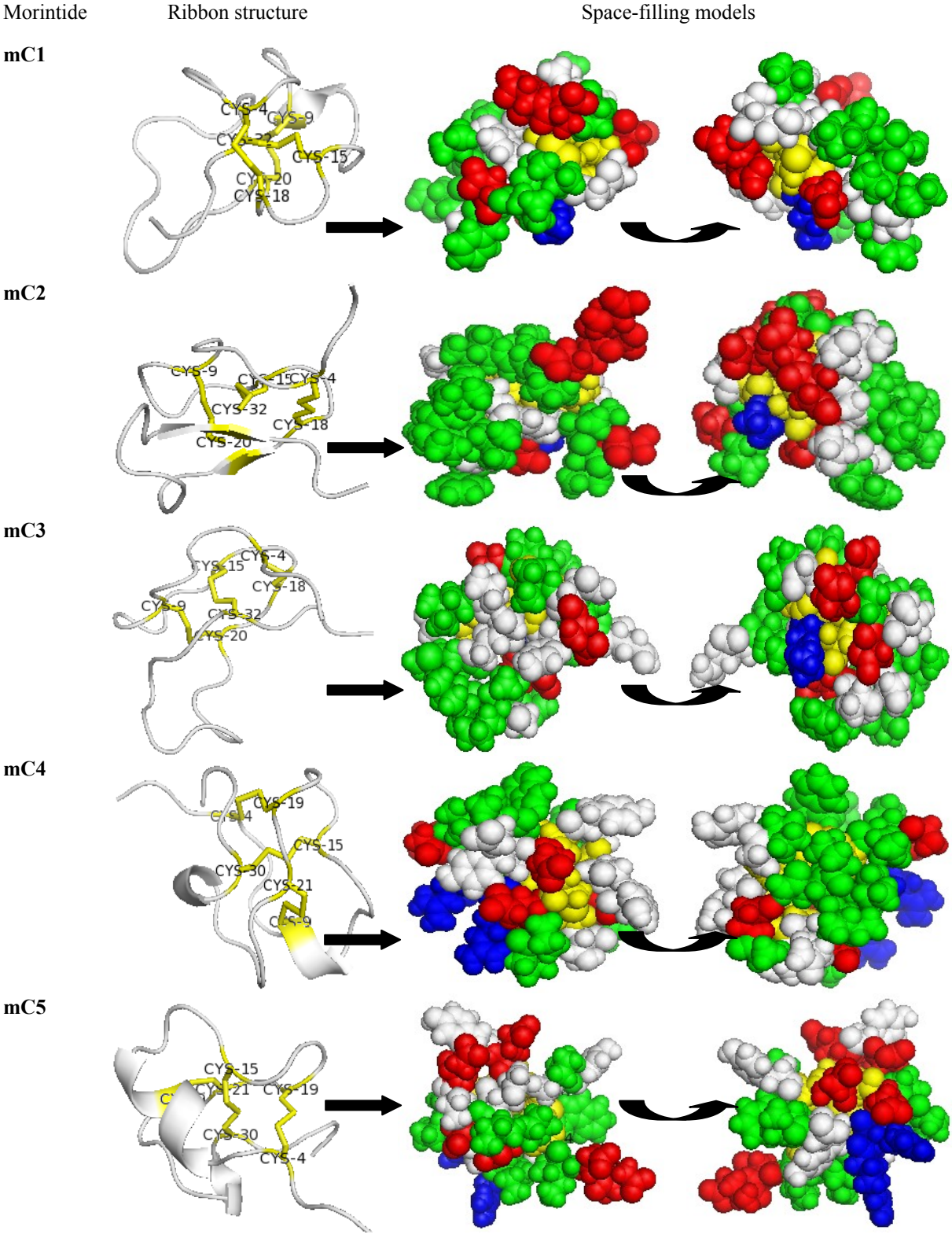


Figure 42. Computer models of mC1 to mC5. The lowest energy structures of morintides were chosen for these computer models. (A-C) represented mC1, D to F represented MC2, G to I represented MC3, J to L represented MC4, and M to O represented MC5. Figures **A, D, G , J** and **M** were presented in ribbon style, demonstrating the inhibitor cystine-knot (ICK) motif. The ring formed by Cys I-IV and

Cys II-V was penetrated by Cys III-VI. Cysteine residues were labeled and colored yellow. Disulphide bonds were presented in stick. Figures **B, E, H, K**, and **N** represented space-filling models, demonstrating the surface nature of the respective morintides in exactly the same orientation as their ribbon structures. These were rotated 180^0 to give Figures **C, F, I, L** and **O**. The individual residues was colored based on their properties, with green, white, yellow, blue, and red representing hydrophobic, hydrophilic, cysteines, basic and acidic residues, respectively. (Modeled using Pymol)

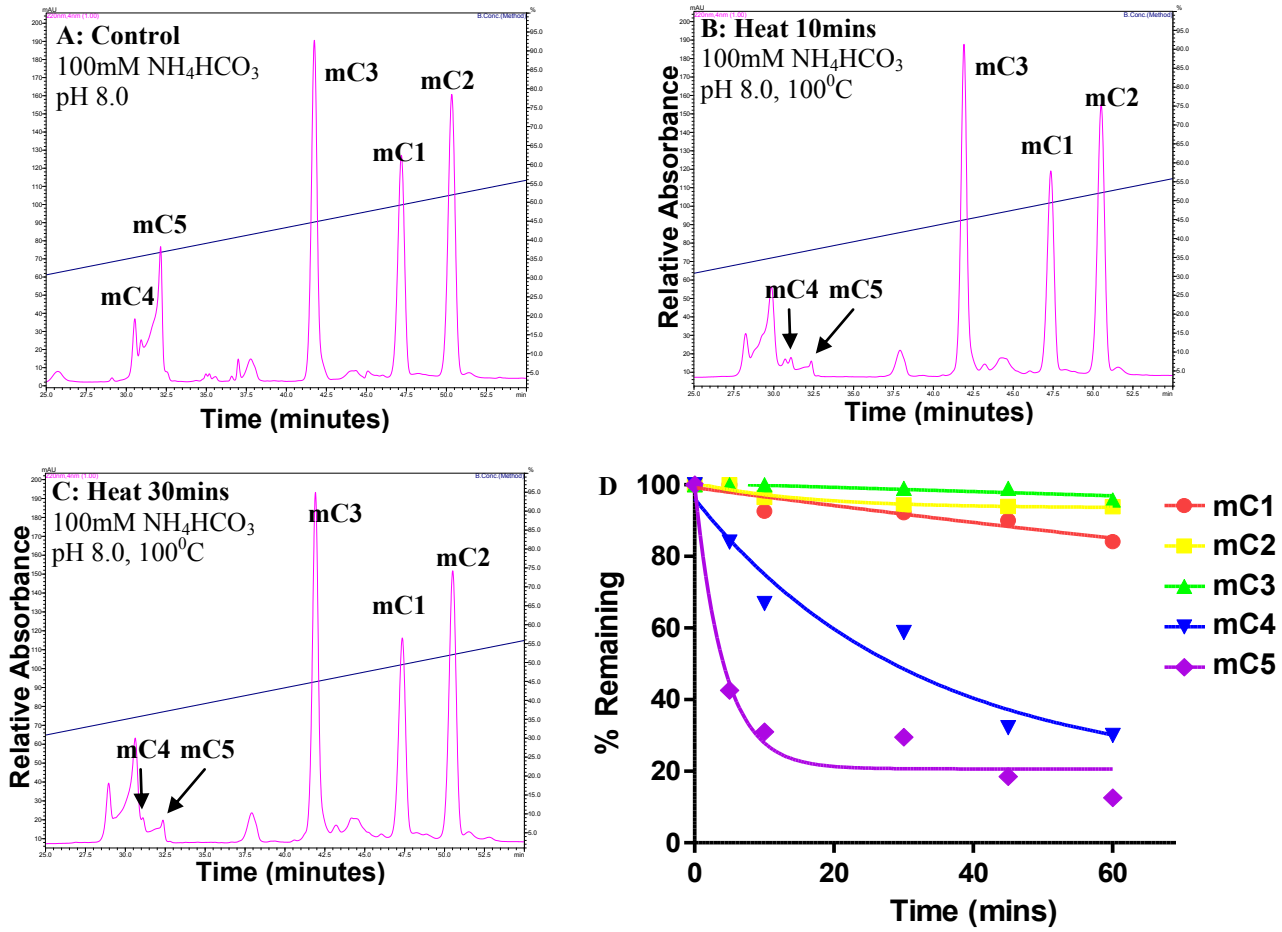


Figure 43. Effects of heat on stability of morintides**. A.** RP-HPLC profile of morintides in 100mM NH_4HCO_3, pH 8.0, 100^0C without heating (control). Peaks corresponding to each morintide were labeled. **B, C.** RP-HPLC profile of morintides in 100mM NH_4HCO_3, pH 8.0, 100^0C after 10 and 30 min. D**.** Percentage of morintides remaining when heating at 100^0C at different time intervals. Control was kept at 37^0C. The maximal degradation for each morintide was as follows: 15% for mC1, 6% for mC2 and 3% for mC3, 70% for mC4 and 88% for mC5.

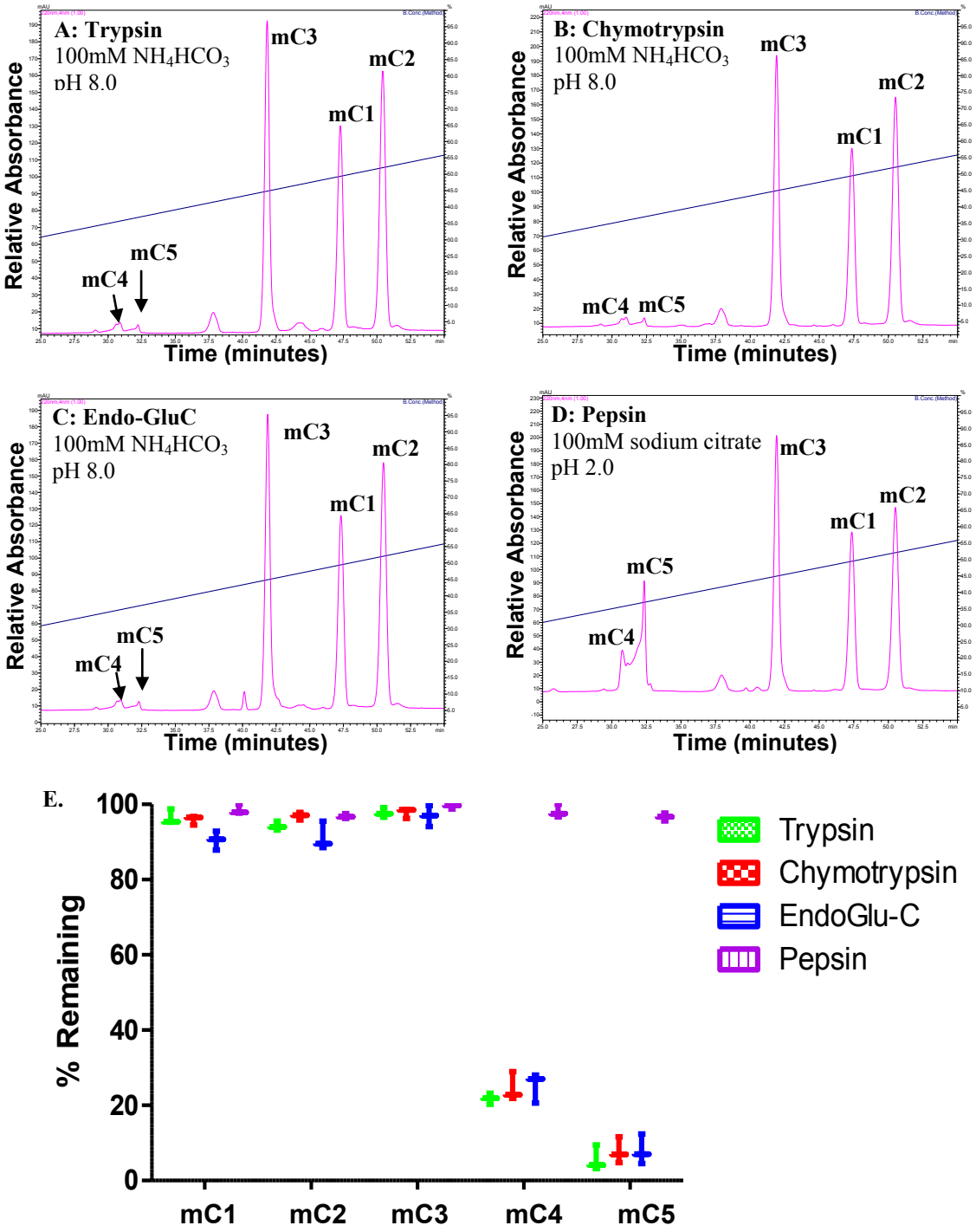


Figure 44. Effects of different enzymatic conditions on the stability of morintides. **A-D.** RP-HPLC profile of morintides exposed to trypsin (A), chymotrypsin (B), endoGlu-C (C) in 100mM NH_4HCO_3, pH 8.0, or pepsin (D) in 100mM sodium citrate, pH 2.0, at 37^0C for 1h. Peaks corresponding to each morintide were labeled. **E.** Box and whiskers plot showing the percentage of morintides remaining under each enzymatic stress condition after 1h.

2.7 Cytotoxicity

To assess for cytotoxicity, effects of morintide C1-5 on A549 and HeLa cells were evaluated using MTT assay. For each peptide, six different concentrations were examined (2.5, 5, 10, 20, 40, 80 μM). All morintides tested were not cytotoxic to both cell lines up to 80 μM concentration. Morintides also found not to induce cell proliferation at all tested concentrations.

2.8 Hemolytic assay

Hemolytic effects of morintide C1-5 were assessed on Human type A erythrocyte. For each peptide, five different concentrations were examined (3.125, 6.25, 12.5, 25, 50 μM). No significant hemolysis was observed for all tested morintides up to 50 μM concentration.

2.9 Antimicrobial assay

Radial diffusion assay was performed to assess the antimicrobial properties of morintide C1-5 on nine different microorganisms. These included five Gram-positive bacteria (*S. epidermis, S. aureus, S. mutants, S. salivarius* and *M. luteus*), one Gram-negative (*E. coli*), and three fungal strains (*C. albicans, C. tropicalis* and *C. kefyr*). None of the morintides were active against these strains.

The assay was then repeated with Zn^{2+} supplementation, which had been postulated to act as a cationic salt bridge between anionic microbial surfaces and anionic antimicrobial peptides [106]. Despite these changes, no antimicrobial activities were observed for morintides up to 100 μM. This could be due to the low concentration (100μM) of morintides tested. Brogden et al reported that the minimum

bactericidal concentrations of anionic pulmonary surfactant was 100-600μM with Zn^{2+} supplementation [107].

2.10 Immunomodulation assay

Morinda citrifolia has been traditionally used for treatment of inflammation and pain. Such observed effects are most likely arisen from immunomodulatory effects. Immunomodulation assays were thus performed on THP-1 cells pretreated with LPS to determine if morintides (mC1-5) possessed any immunomodulatory properties. No significant changes in cytokine profiles of IL-6, IL-8, IL-10 and TNF-α were observed for morintides up to concentration of 10 μM.

3. Discussion

Morinda citrifolia is a well-known medicinal plant with broad therapeutic effects. The increasing popularity of fruit juice products worldwide has attracted many scientific researches to investigate its phytochemicals and constituents. This led to discovery of a large number of secondary metabolites which all are small molecule entities with MW less than 500 Da. The work described here is a pilot study for CRP constituents in *Morinda citrifolia* which are likely active phytocomponents attributed for its multiple health benefit. A novel family of highly anionic cystine-knot peptides named morintide was discovered. Over 40 unique masses in the range of 3 to 4 kDa were documented in different tissue parts, of which 11 primary sequences were obtained. They are the most anionic knottin peptides hitherto isolated with net charge as low as -7. This is also the first time another class of knottin, other than cyclotides, being isolated from a member of the Rubiaceae family.

Morintides isolated in this study were achieved by using similar approach as that used for cyclotide purification. They share three key characteristics with cyclotides including late retention time (eluted between 20 to 60% ACN), mass range (3 to 4 kDa) and a knotted structure. They are, however, distinguished from cyclotides or uncyclotides by having entirely different primary sequences, open-end structures and highly anionic nature. Blast search revealed no sequence homology of morintides to any known CRPs indicating that morintides represent a new class of CRPs. It is interesting to note that *M. citrifolia* do not produce cyclotides as other members of the Rubiaceae family [33]. It instead produces morintides as probable CRPs replacement of cyclotides deemed to benefit the plant during its course of evolution.

Morintides, with a consensus sequence of $CX_4CX_5CX_{2-4}CX_{1-2}CX_{8-11}C$, belong to the inhibitor cystine-knot (ICK) family with a general consensus sequence of $CX_{3-7}CX_{3-6}CX_{0-5}CX_{1-4}CX_{4-13}C$ [108]. Their knotted structures represented by morintide C2 were confirmed by double-tagged S-alkylation combined with top-down MS/MS approach. Peptides adopted this ICK scaffold are usually stable and have a range of interesting biological activities. They were typically small CRPs with 26-48 residues and almost all have a net positive charge [109]. In contrast, morintides are highly anionic, possessing charges from -1 to -7. This represents as a departure from past reports that plant CRPs were neutral to positively charged. Morintides extend the diversity of CRPs to a new window pushing the anionicity limit to -7 in net charge.

Stability testings showed that morintide C1-3 were notably stable to heat and enzymatic digestion which are characteristic properties of highly bridged peptides [110]. However, not all morintides were equally stable. Morintide C4 and C5 were susceptible to heat and endopeptidases. It is likely due to the destabilizing effects of the Glu-Pro or Asp-Gly bond present in their primary sequences. Through the anchimeric assistance of the carboxylic side chain, Asp and Glu are prone to hydrolysis at the carboxylic side, with the formation of a five or six-member ring transition state, respectively. Coupled with the bond strain presented in the tertiary amide bond of Pro or the little steric hindrance of the Gly side chain, the Glu-Pro and Asn-Gly bonds are highly susceptible to hydrolysis. Such differences in heat and enzymatic stability observed in morintides, makes them ideal candidates for structure-stability study of cystine-knot miniproteins.

Biological activities testing aimed to characterize the membranolytic effects of morintides revealed that they are not cytotoxic, hemolytic or bactericidal. This is not surprising given that most membrano-active peptides are generally cationic, which

facilitates their binding to the anionic microbial surfaces leading to destruction or solubilization of the cell membranes [111]. Similarly, the membranes of cancer cells typically carry a net negative charge due to increased expression of anionic lipid such as phosphatidylserine [111]. The highly anionic nature of morintides probably impairs their binding to the negatively charged surface of bacterial and cancer cells explaining their lack of activities. In addition, morintides are also not hemolytic despite containing a significant percentage of hydrophobic residues comparable to cyclotides. The reason for this low hemolytic activity is not entirely understood. It is possible that electrostatic attraction though less important is contributed for peptide interactions with the negatively charged lipids located in the inner membrane of the cells. The presence of multiple acidic residues in the morintide molecules would not only nullify such interaction but also cause an electrostatic repulsion between morintides and the erythrocyte membrane.

The unusual primary structures and relative non-toxicity of morintides on both prokaryotes and eukaryotes cells tested in this work raise the question about their physiological functions in plants. It is probable that morintides may participate in insect defense, the most significant herbivores that constitute about 75% of all animal species, by inhibiting certain enzymes, proteins or pathways that cause death or interfere with their growth and development. Alternatively, they may participate in signaling pathways or play unidentified functions in plants. It is interesting to note that morintide C1 and C2 possess clusters of acidic residues. Computer models show that they are divided into two pockets resembling two pairs of “acidic claws” which probably involve in divalent metal ion chelation (Fig. 45A and B). Asp and Glu residues were known to bind divalent ions such as Zn^{2+}, Mn^{2+} and Mg^{2+}. For example, bacterial transposases and retroviral integrases possess a conserved DDE

triad (Fig. 44C) as part of a catalytic pocket in which the carboxylic groups co-ordinate a Mg^{2+}ion [112]. Metal binding properties may play many important roles in plant metabolism.

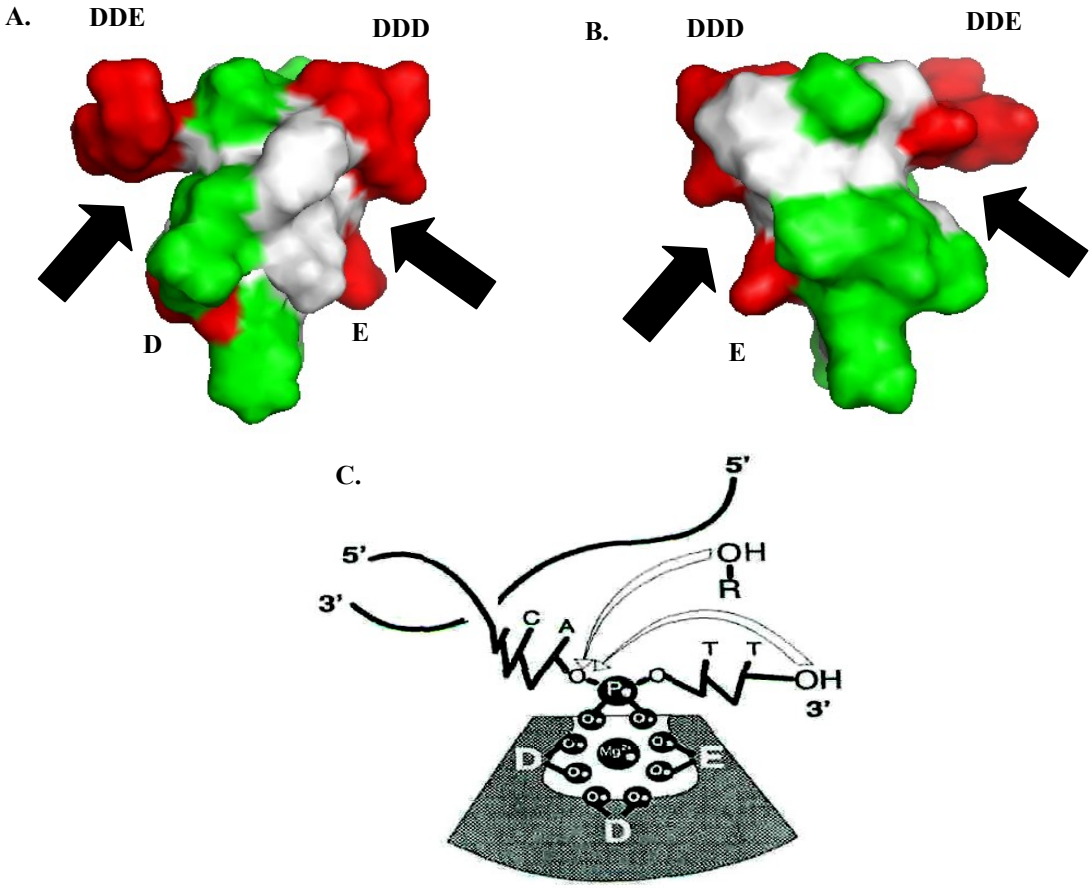


Figure 45. Comparison of DDE motif and computer model of morintide C1. **A.** Space filling model of morintide C1 demonstrating a pair of 'acidic claws'. Black arrows showed putative pocket for chelation of divalent ions. Acidic residues were labeled and shown in red. Hydrophobic residues were shown in green. **B.** Figure A flipped 180^0C. **C.** Example of DDE motif showing the acidic groups coordinating Mg^{2+}ion together with a backbone phosphate group of DNA, rendering the associated phosphodiester bond susceptible to nucleophilic attack (open arrows). *Figure Taken from Polard et al. [113].*

Although the biological activities of morintides remain to be elucidated, their high stability to heat, chemicals, and enzymes coupled with the lack of hemolytic and low cytotoxic effects allow them to be exploited as a drug scaffold. Bioactive-peptide epitopes grafted into morintide loops between successive Cys residues would significantly improve their proteolytic resistance and thermal stability. Knowledge of each loop involvement in folding, stability and biological activities would be an interest area for future research to determine the tolerance of this scaffold for sequence modification. Among morintide loops, peptide sequences between Cys V and Cys VI of morintide C1-3 stretching over 11 residues are sufficiently long and thus preferred sites for protein engineering to bind diverse targets.

4. Conclusions

Morintides belong to a novel class of anionic CRPs possessing an inhibitor cystine-knot. They do not exhibit the properties typically associated with CRPs, such as cytotoxic, hemolytic and antimicrobial activities. This is not surprising given that most CRPs isolated to date are predominantly cationic whereas morintides are highly anionic. Although the biological properties of morintides remain to be elucidated, on the basis of similarities with metal-ion binding motifs, it is possible that the conserved D and E residues in morintides may participate in the binding of divalent metal ions, such as Mn^{2+} and Mg^{2+}. Some morintides are remarkably stable to heat denaturation and enzyme proteolysis coupled with their lack of cytotoxic and hemolytic properties, they present as interesting protein scaffold for drug design.

Chapter 6

Molecular Basis of Heat-Stable Ginsentides in Three Ginseng species

1. Introduction

Ginseng, directly translated as “man essence”, is one of the most valued medicinal plants in TCM. It is often known as “King of all Herbs” and was once prized more than gold to the ancient Chinese. Ginseng (*Panax ginseng*) was used by the royals and the Chinese Emperors since long before the Han Dynasty (203 BC- 220 AD) with the belief that it is an elixir of life capable of promoting health and longevity. The efficacy of Ginseng was known to the West by the 18th century, and since then numerous medical researches have been carried out to provide scientific evidences of this highly acclaimed herbs [114]. In the recent times, the demand for ginseng has increased significantly as a healthy food supplement.

Ginseng is the collective name for 11 species of slow-growing perennial plants belonging to the *Panax* genus of the Araliaceae family. They grow best in the cool weather climate and are often found in the mountain area. Most studies of ginseng were performed on three common species: *Panax ginseng* (Asian ginseng), *Panax quinquefolius* (American ginseng) and *Panax notoginseng* (notoginseng). Phytochemistry characterization of these three species have led to the discovery of over 107 chemical entities which are categorized into 18 groups [115]. Among them, ginsenosides [116], the bitter components found in the roots, are believed to be the main bioactive ingredients for ginseng pharmacological actions, and large amount of researches have been focused on them. Studies have shown that ginsenosides have various effects on CNS both inhibitory and stimulatory properties. They have also

demonstrated anticarcinogenic, immunomodulatory effects and a number of other biological activities [114].

The content of ginsenosides varies in different ginseng species and different ginseng forms depending on their processing. *Panax ginseng* alone is available in several forms of fresh, white and red ginseng. Fresh ginseng is ginseng in its original raw form. White ginseng, more commonly known as dried ginseng, is formed by air drying raw ginseng under the sun. Red ginseng is formed through the process of steaming raw ginseng, stripped of their skins at 100°C until it is reddish-brown. It has been demonstrated that red ginseng has a higher ginsenoside content than white ginseng due to the conversion of malonyl-ginsenosides to their corresponding ginsenosides by hydrolysis [117]. As the result, different ginseng forms have different pharmacological activities and thus different usages when prescribed to the patients.

Changing the focus from ginsenosides and small molecules to biologically active peptides and proteins, this study reports for the findings of novel CRPs named ginsentides from three ginseng species: *Panax ginseng*, *Panax notoginseng* and *Panax quinquefolius*. Instead of possessing a cystine knot motif [118] commonly seen in many plant CRPs, ginsentides have a novel pattern which resembles a pretzel-knot structure, greatly confining their conformational space. Biophysical and functional characterizations of novel ginsentides were subsequently investigated providing early understanding of their molecular structures and activities.

2. Results

2.1 CRPs' profiling of *P. ginseng, P. notoginseng* and *P. quinquefolius*

Primary screening was done for three ginseng species: *Panax ginseng, Panax notoginseng* and *Panax quinquefolius*. Mass spectrum of these plant extracts revealed that they contain a group of peptides with strong *m/z* intensity within the range of 2900-3500 Da similar to that of cyclotides and knottins. They also have similar hydrophobicity which eluted at 40-50% ACN from RP-HPLC. Despite these similarities, S-reduction and S-alkylation studies revealed that they have four cystine linkages, which is unusual for peptides of their mass range. As far as I am aware of, these peptides have not been studied before. They may thus represent a novel group of disulfide-rich peptides in ginseng.

As shown in Fig. 46, each ginseng species have a unique CRPs' profile. MS spectrum of *Panax ginseng* displayed three abundant peaks at 3054, 3084 and 3122 Da. Panax notoginseng and Panax quinquefolius each displayed a single abundant mass at 3054 and 3071 Da, respectively. Since these three ginseng species are commonly misidentified, their unique MS patterns can be used as a rapid and reliable method for identification and authentication.

2.2 Purification of novel ginsentides from *Panax ginseng* and *Panax notoginseng*

Plant samples of each ginseng species were pulverized and extracted with water. Novel CRPs designated as ginsentides were purified by repetitive RP-HPLC. They were eluted between 40-50% of ACN on RP-HPLC. Four novel ginsentides were isolated from *Panax ginseng* with *m/z* at 3054, 3084, 3099 and 3141 Da named ginsentide G1-4, and one from *Panax notoginseng* with *m/z* at 3054 Da named ginsentide N1. On average, 1 kg of dried ginseng roots yielded 10-15 mg of ginsentides and 1 kg of dried ginseng flowers yielded 100-200 mg of ginsentides.

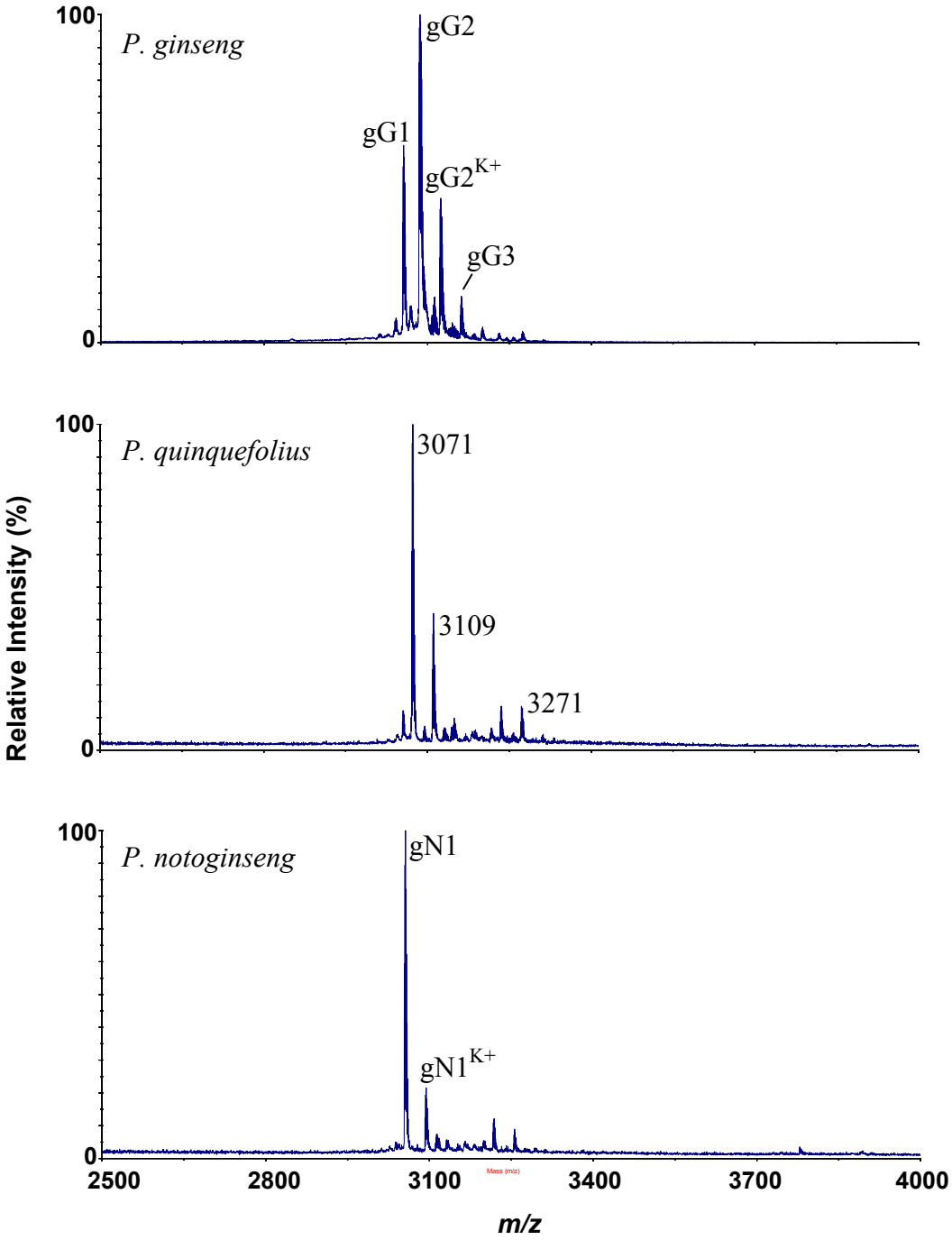


Figure 46. MS profiles of the three ginseng species. Root samples of *P. ginseng, P. quinquefolius, P. notoginseng* were extracted with water and profiled by MS after Ziptip. Peaks with K+ label are potassium-binding adducts.

2.3 Sequencing of ginsentides

To determine the primary sequences of novel ginsentides, they were S-reduced and S-alkylated with DTT and IAA, respectively. S-alkylation was confirmed by a mass increase of 464 Da indicative of four disulfide linkages. S-alkylated peptides were subjected to digestions by trypsin or chymotrypsin followed by MS/MS analysis. Their primary sequences were summarized in Table 13. All ginsentides have eight cysteine residues and are unusually rich of Gly amino acid. Ginsentide G1 for examples contains 9 Gly out of 31 residues in length. They also possess unique cysteine spacing entirely different from any known CRPs. It should be noted that ginsentide G1 and N1 have identical primary structures suggesting their functional importance to conserve the peptide sequence in both *P. ginseng* and *P. notoginseng*.

Blast search to identify homology in the database revealed the encoding genes of ginsentide G1-4 from *P. ginseng*. Matching gene for ginsentide N1, however, was not found in the database. Ginsentide encoding genes are known as ginseng-specific abundant proteins (GSAPs) identified from a random gene screening program [119]. Their mRNA transcripts are highly expressed in the rhizomes which ranked third among 17,605 ESTs in the ginseng cDNA library [119]. Although isolation of these genes was performed, there was no information about their biosynthesis processing as well as functional and structural characterizations of the mature peptides. Homology search also revealed a striking similarity in the primary sequence of ginsentides and ω-conotoxin from *Conus californius.* This suggests that ginsentides may have ion-channel related activities.

Table 13. Novel ginsentides from *Panax ginseng* and *Panax notoginseng*

Ginsentide	Primary Sequence	Mass (Da)	Species
gG1	C KSGGAW C GFDPHG CC GN C G C LVGF C YGTG C	3054	*P. ginseng*
gG2	C KSSGAW C GFDPHG CC GN C G C LVGF C YGTG C	3084	*P. ginseng*
gG3	GC KSSGAW C GFDPHG CC GN C G C LVGF C YGTG C	3141	*P. ginseng*
gG4	C KSAGTW C GFDPHG CC GS C G C LVGF C YGVS C	3099	*P. ginseng*
gN1	C KSGGAW C GFDPHG CC GN C G C LVGF C YGTG C	3070	*P. notoginseng*
CC1[b]	RYC SDSGGW C GLDPEL CC NSS C FVL CG	2868	C. *canifornicus*
CC2[b]	REC SESGEW C GLDPAL CC GSS C FFT CN	2898	C. *canifornicus*

[a] Number of amino acids, [b] ω-conotoxin from *Conus californicus* denoted as cC1 and cC2 are shown for comparison. Cysteine residues are highlighted in yellow. Positively charged residues are written in red. Negatively charged residues are written in blue. Polar residues are written in green while non-polar residues not colored. Residues indicated with * are those which we were unable to confirm as its signal was too weak to detect by MALDI-TOF MS/MS.

2.4 Tissue specificity

To test for tissue specificity of ginsentides, small-scale screenings of *P. ginseng* flowers, leaves and roots were performed. Around 200 mg of pulverized specimens of each tissue part was dissolved in 1 mL of water. Removal of small molecules from the extracts was done using zip-tip. The resulting samples were then profiled using MALDI-TOF MS. From Figure 47, it can be observed that ginsentides were specifically expressed in the roots and flowers of ginseng but were absent in the leaf tissues. This distribution supports the traditional usages of ginseng in which roots and flowers are more commonly used than leaves. Specific expression of ginsentides in the flowers and roots may also be useful in fighting against counterfeit products that fake herbs or ginseng leaves instead of roots are used for manufacturing.

2.5 Connectivity mapping of ginsentide G1

Strategy to determine disulfide connectivity was to partially S-reduced followed by S-alkylated of ginsentides using TCEP and NEM, respectively, to generate intermediates with one to three remaining disulfide bonds. Intermediates containing one remaining disulfide bond are known as 1SS species, with two remaining disulfide bonds, 2SS species and three remaining disulfide bonds, 3SS species. These intermediate species were subsequently fully S-reduced and S-tagged with a second alkylation reagent using IAA. After tryptic digestion and analyzing with MS/MS, connectivity of ginsentide G1 derived based on S-alkylated pattern of IAA and NEM (Fig. 48) was as follows: Cys I-IV, Cys II-VI, Cys III-VII and Cys V-VIII (Fig. 49). Its disulfide linkages are novel with no homology to any known CRPs.

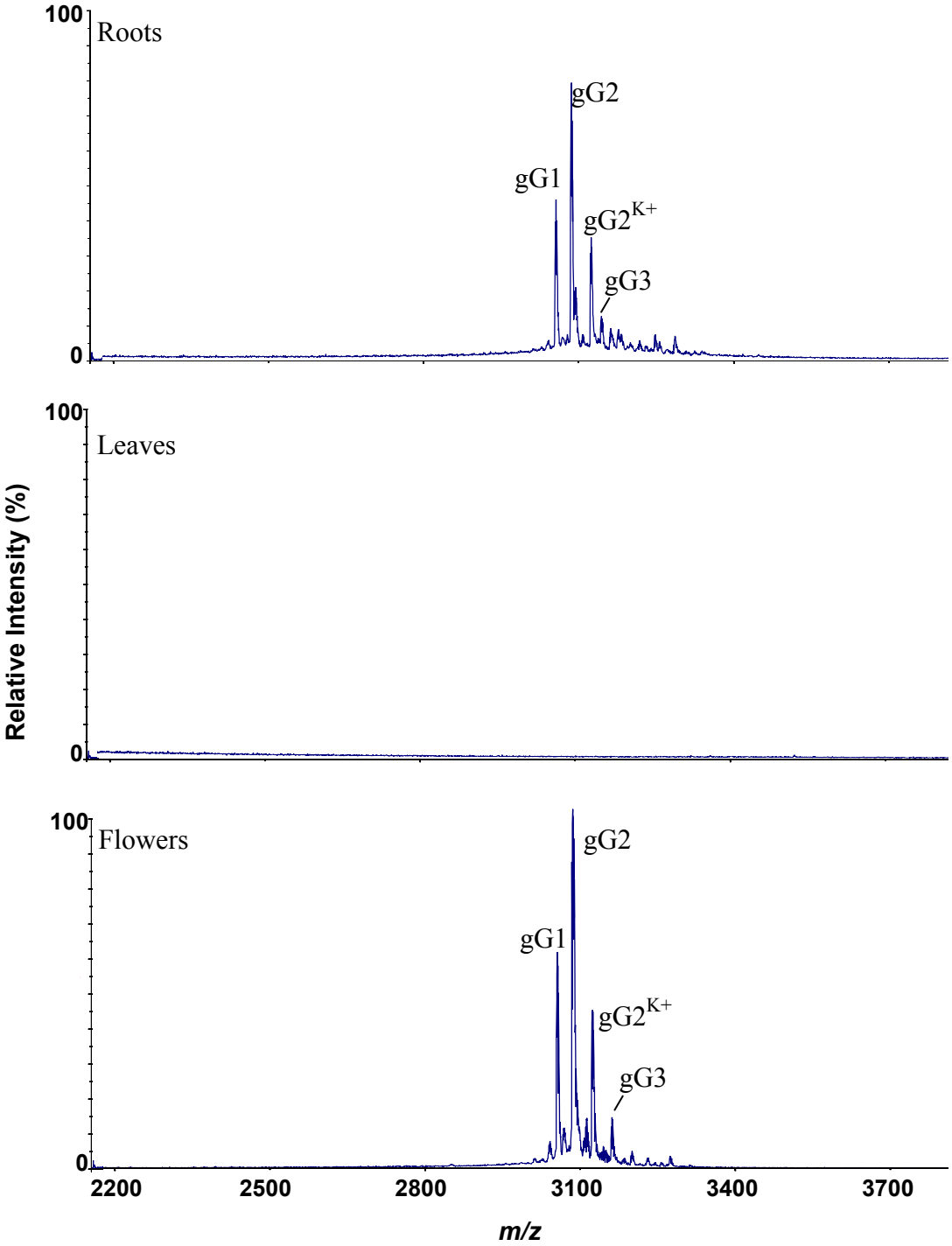
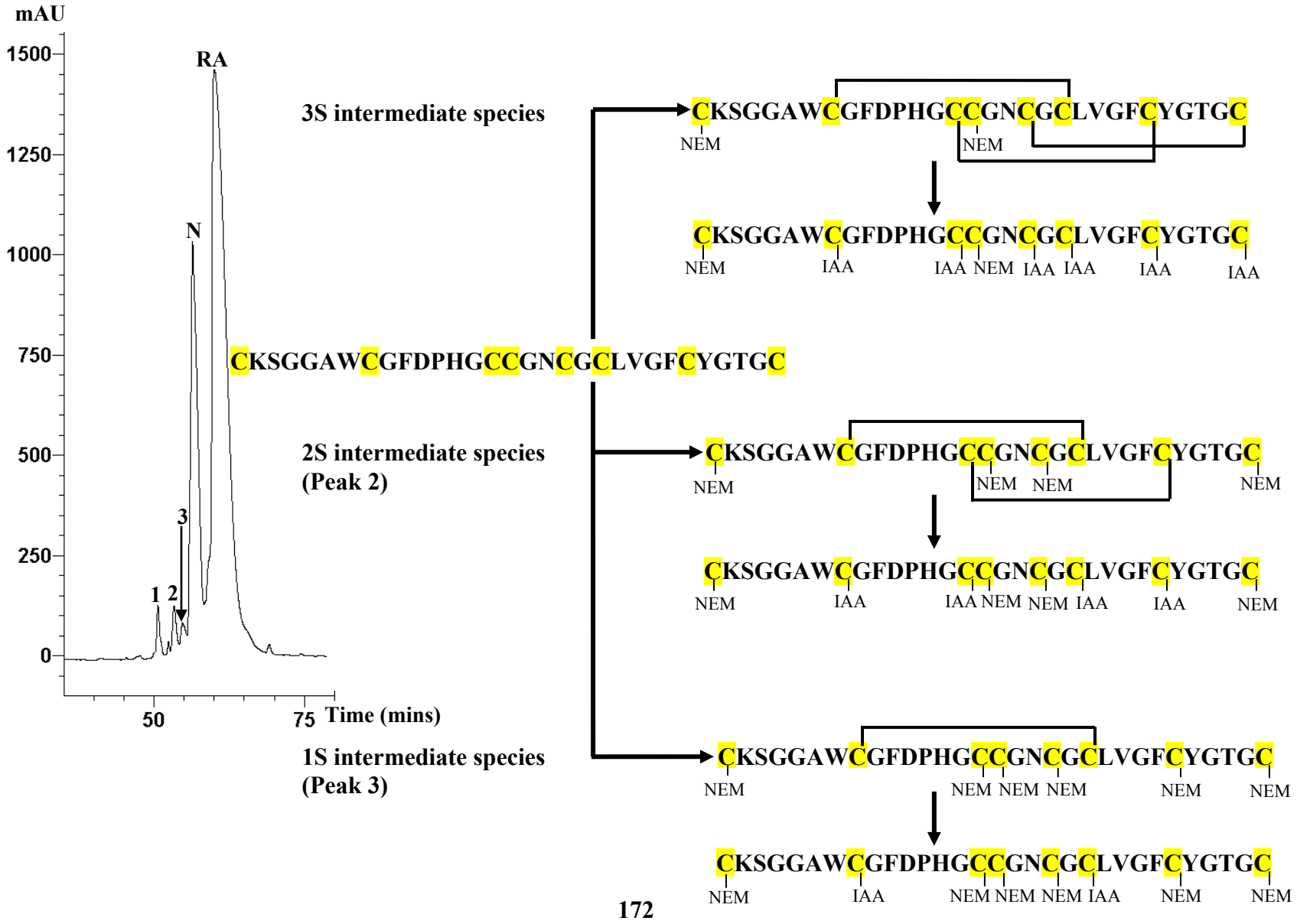
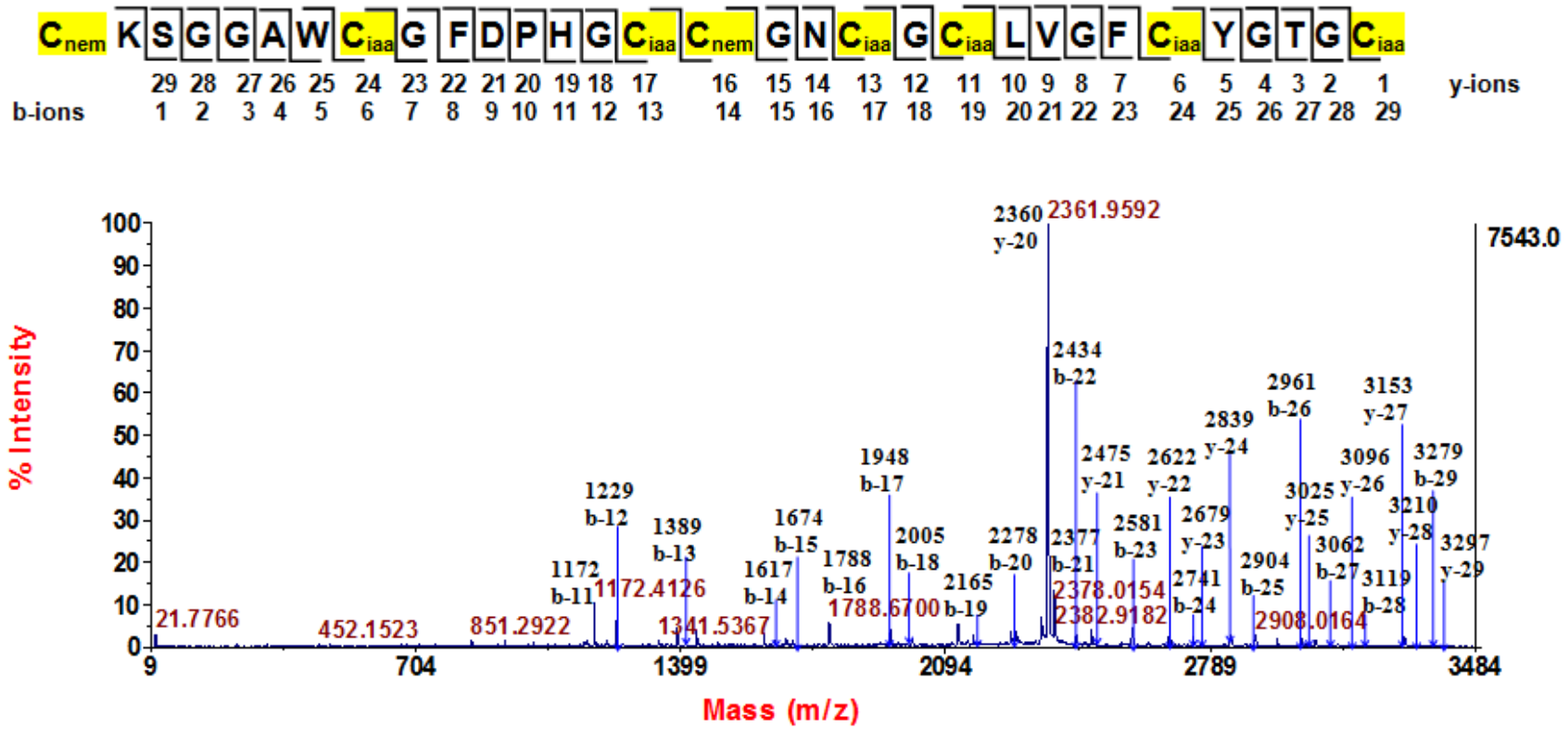


Figure 47. Tissue-specificity of ginsentides. MS profiles of the plant extract of ginseng roots, leaves and flowers show that ginsentides specifically expressed in the roots and flowers and are absent in the leaves.

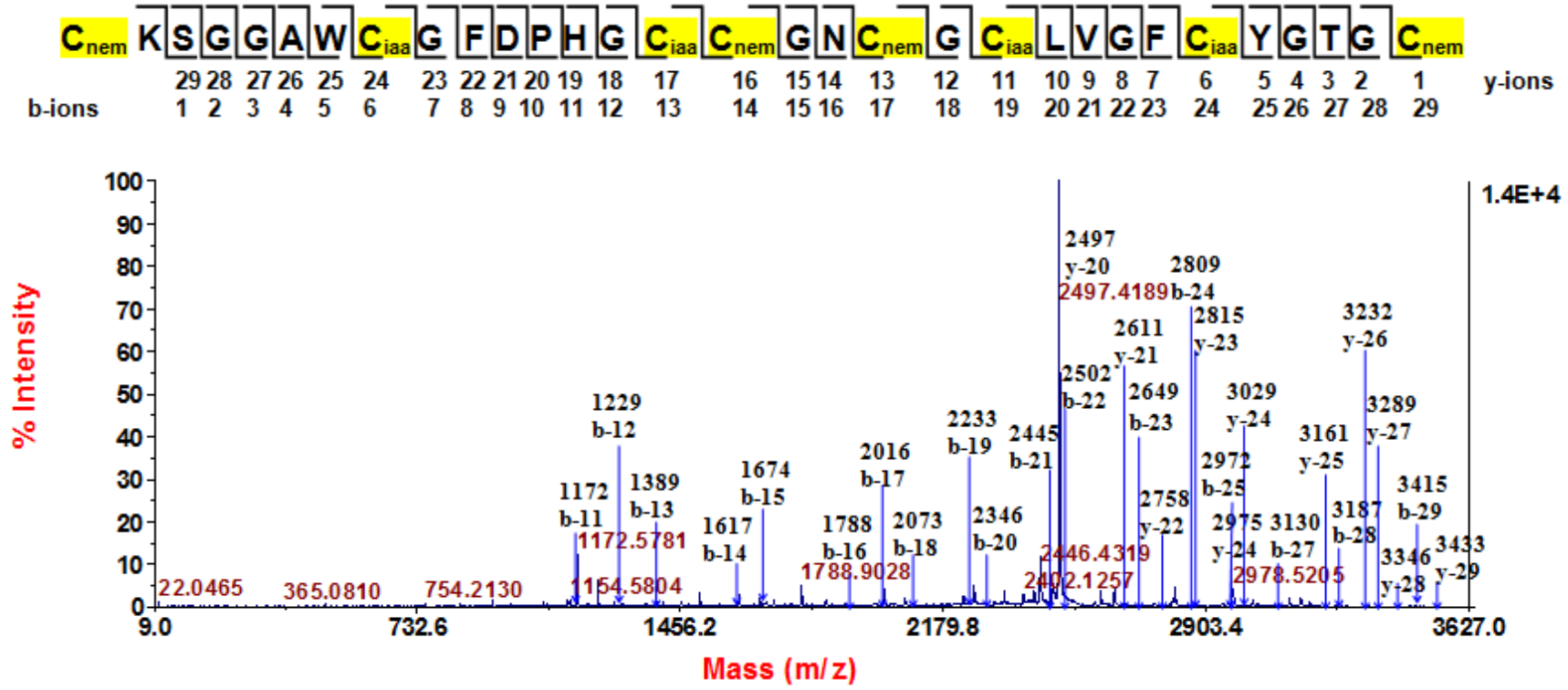
A)


mAU
1500
1250
1000
750
500
250
0
RA
N
3
1
2
50
75
Time (mins)
CKSGGAWCGFDPHGCCGNCGCLVGFCYGTGC
3S intermediate species
2S intermediate species
(Peak 2)
1S intermediate species
(Peak 3)
NEM
IAA

B) 3S intermediate species



C) 2S intermediate species



D) 1S intermediate species

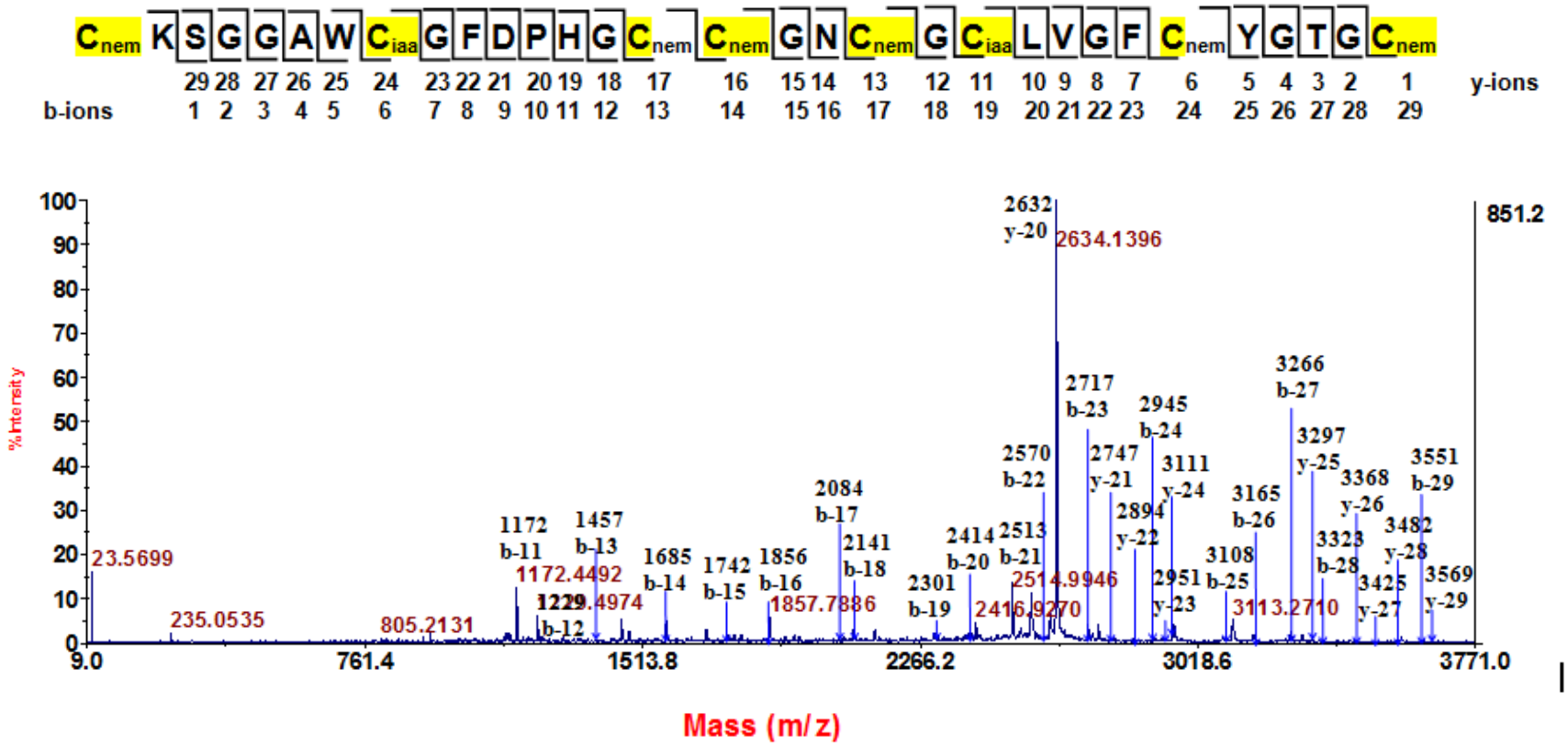


Figure 48. Disulfide mapping of ginsentide G1**. A)** HPLC profile of the partially S-reduced and S-alkylated ginsentide G1. Peak 1, 2, 3, N and RA contained the 3SS, 2SS, 1SS, native peptide and fully S-NEM alkylated peptides, respectively. Schematic presentation of disulfide mapping of ginsentide G1 is also shown. **B-D)** MS/MS profile of trypsin digested intermediate species after S-alkylation with IAA. The two NEM tagged-cysteines in the 3SS intermediate species revealed Cys I-IV bonding. The four NEM tagged-cysteines combined with the result from 2SS intermediate species reveal a Cys V-VIII bonding. With the last combined information from the 1SS intermediate species, we were able to derive the novel disulfide connectivity of GS1 to be Cys I-IV, Cys II-VI, Cys III-VII and Cys V-VIII.

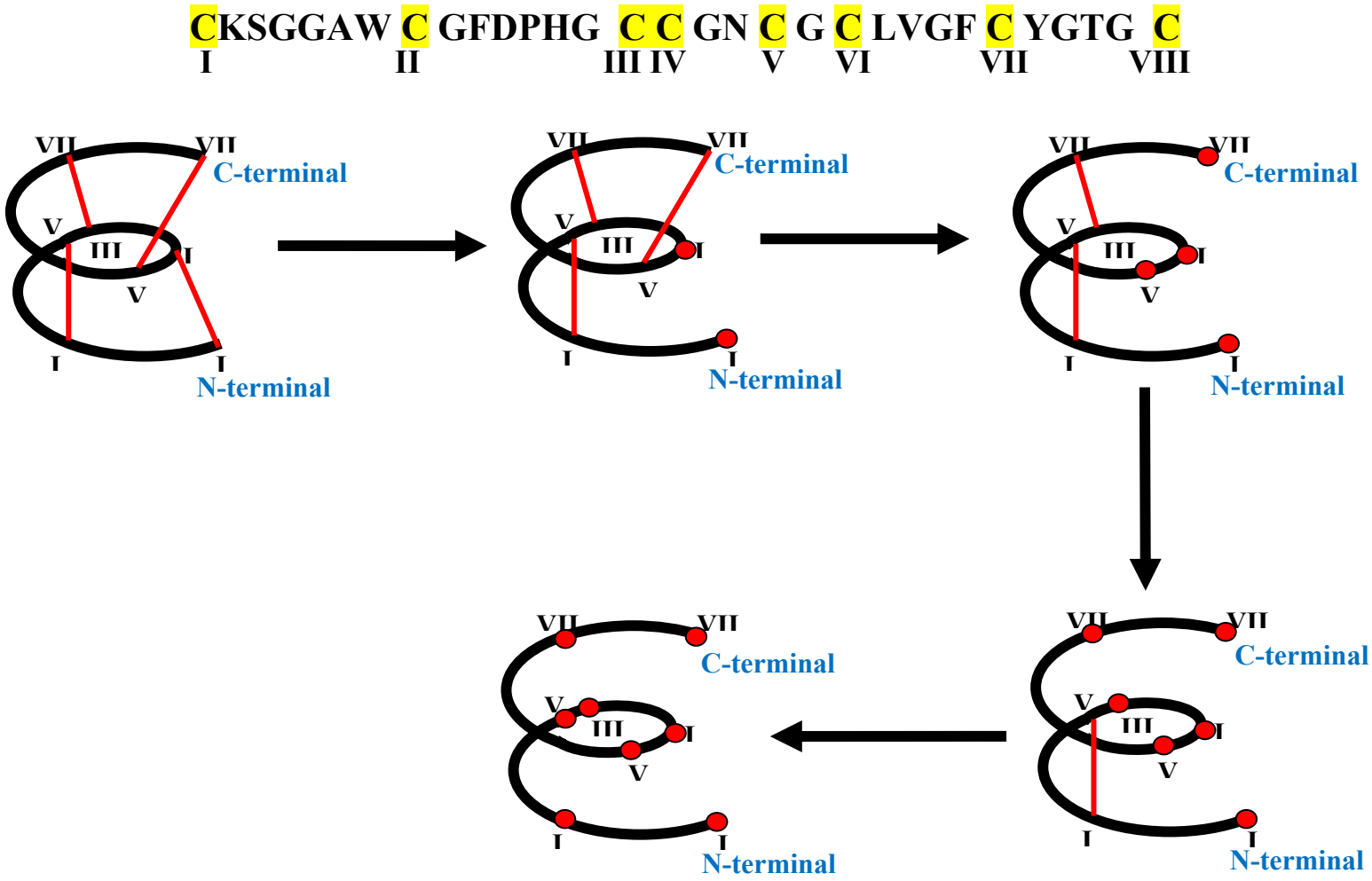


Figure 49. The unfolding pathway of ginsentide G1. Under the experimental conditions used, the I-IV bond (3S species) was the first to be reduced, followed by the V-VIII bond (2S species) and lastly the III-VII bond (1S species).

2.6 Thermal stability assay

Stability assays were performed mainly on ginsentide G1 and G2. They were dissolved in water and heated at 100^{o}C for 15, 30 and 60 min. The percentages of ginsentides remaining were quantified based on their relative areas on the RP-HPLC chromatoGrams before and after the heat treatment. The results show that ginsentides are relatively stable to heat with more than 90% ginsentides survived after 30 min of heat treatment (Fig. 50, 51). They are, however, slowly degraded upon prolong heating. After 60 min of incubation, approximately 73% and 87% of ginsentide G1 and G2, respectively, remained in the solution. The degradation products though not characterized in this study were more hydrophilic than ginsentides evident by few new peaks eluted earlier in the HPLC trace, and these increased in intensity with increased incubation time.

2.7 Enzyme stability

Enzymatic stability of ginsentide G1 was conducted using three different enzymes commonly found in the digestive system including trypsin, chymotrypsin and pepsin. It was incubated with these proteases at 37^{0}C for 1 hr. As shown in Fig. 52, ginsentide G1 were stable to all three enzymes with only a slight drop in peak intensity. The control peptide, having primary sequence as TAYGRKFFSL, was mostly digested under the same condition (Fig. 53).

2.8 Acid stability

Ginsentide G1 was dissolved in 0.2 M HCl and incubated at 37^{o}C to examine its stability to low acidic condition similar to those inside the stomach. No hydrolysis or denaturation was observed after two hours incubation indicated by identical HPLC profiles as compared to control sample without acid treatment (Fig. 54).

2.9 Bioassays

Ginsentide G1 and G2 were tested for a range of biological activities similar to those described for morintides including antimicrobial, cytotoxicity, hemolysis and immunomodulation. The results showed that ginsentides were neither bactericidal nor bacteriostatic for all nine tested fungi and bacteria strains up to 100 μM concentration. They did not lyse RBC up to 50 μM concentration. They were also not cytotoxic against HeLa and A549 cell lines at 30 μM concentration. Immunomodulatory assays reveal that ginsentides did not up regulate or down regulate the expression of TNF-a, IL-1, IL-6, IL-8 and IL-10 produced by THP-1 cells at five tested concentration (0.5, 1, 2, 4, 8 μM).

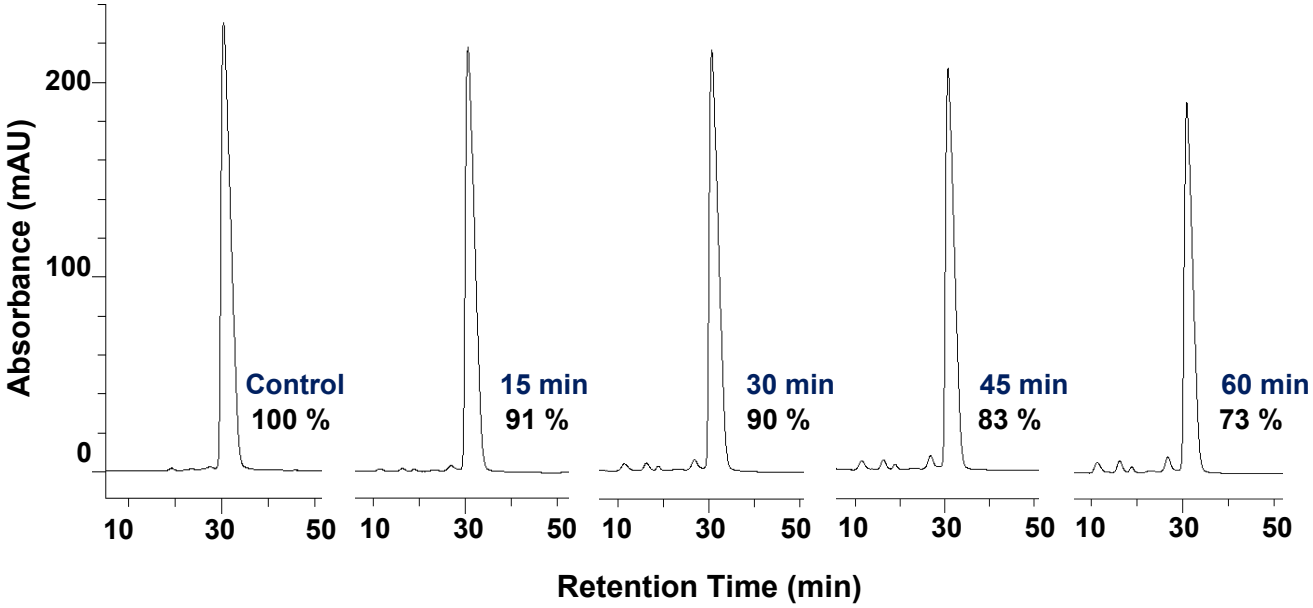


Figure 50. Heat stability of ginsentide G1 monitored by RP-HPLC. Ginsentide G1 was heated for 15, 30, 45 and 60 min at 100°C. The relative area of the remaining ginsentide after heating was compared with the non-heated heated sample (control).

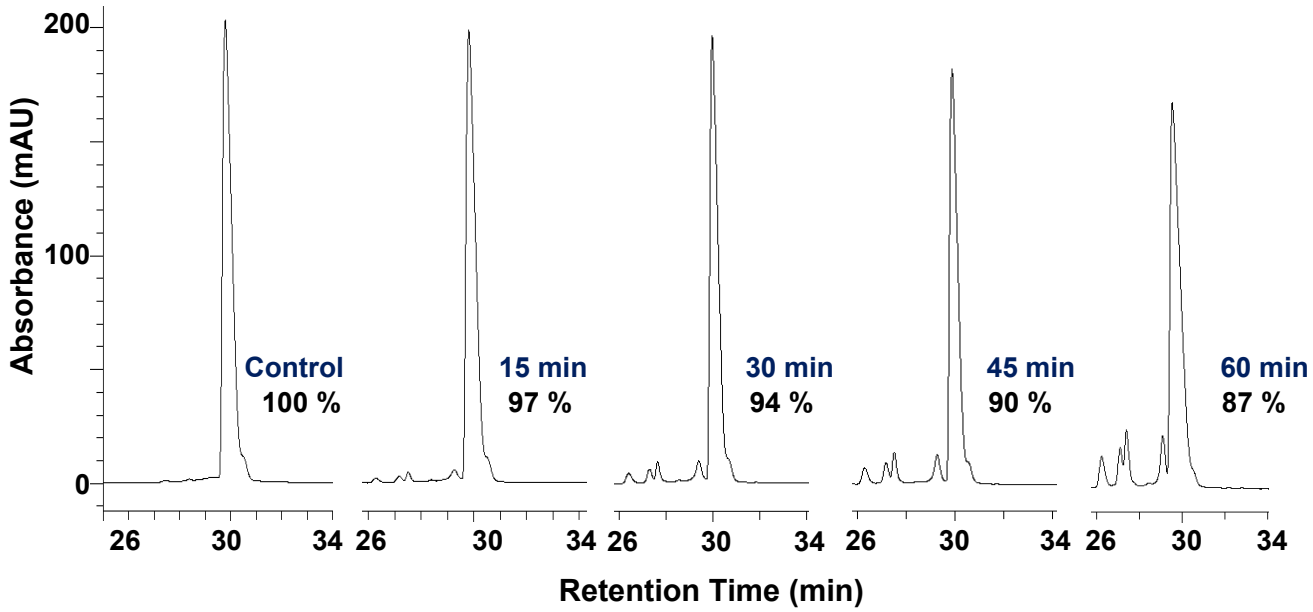


Figure 51. Heat stability of ginsentide G2 monitored by RP-HPLC. Ginsentide G2 was heated for 15, 30, 45 and 60 min at 100°C. The relative area of the remaining ginsentide after heating was compared with the non-heated heated sample (control).

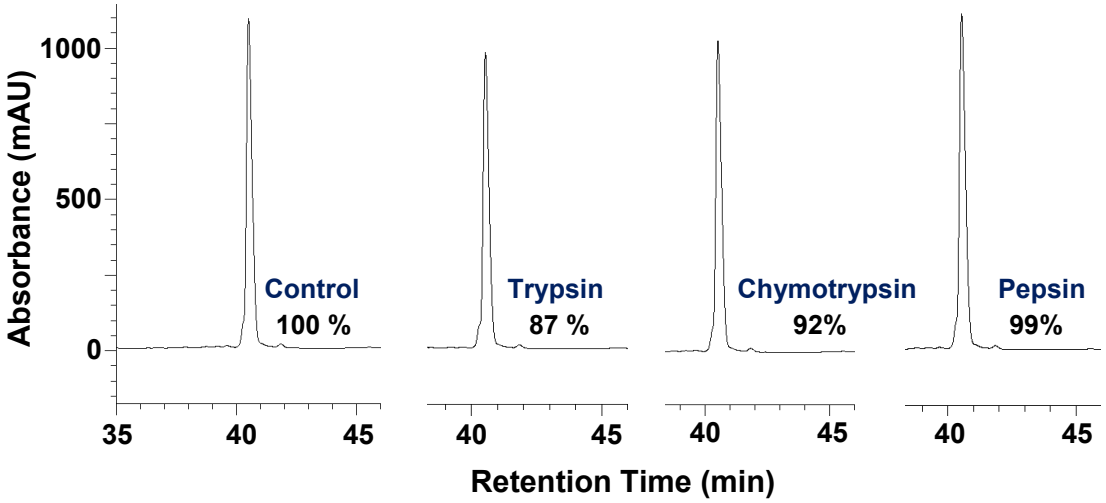


Figure 52. Enzymatic stability of ginsentide G1. HPLC profiles of ginsentide G1 incubated at 37^{o}C for 1hr with or without the presence of trypsin, chymotrypsin and pepsin. Over 80% of the peptide remain in the solution suggest their stability against enzymatic digestion.

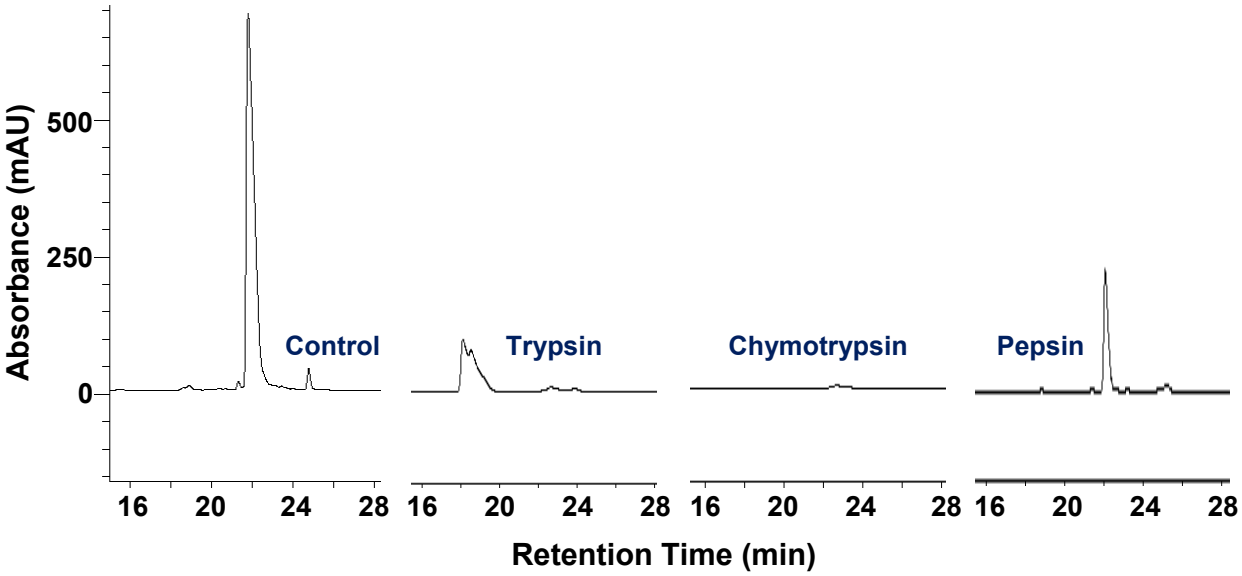


Figure 53. HPLC profiles of the control synthetic peptide in the presence of proteolytic enzymes. Incubation at 37^{o}C for 1 hr show rapid degradation suggests its susceptibility to enzymatic degradation.

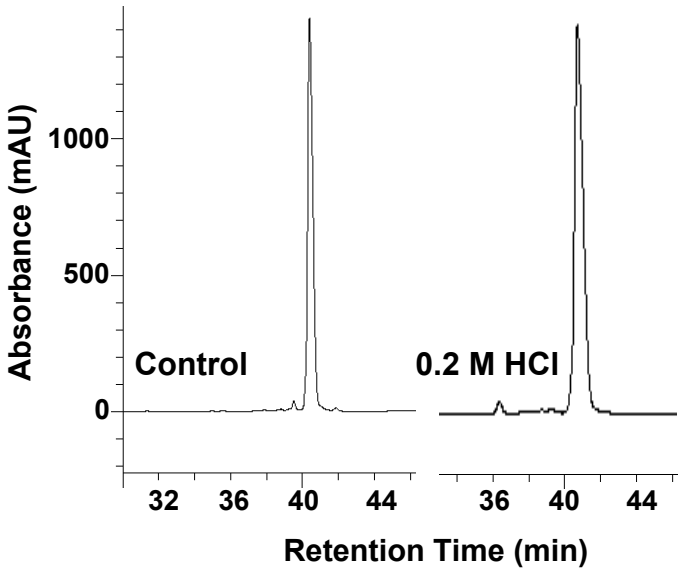


Figure 54. Acid Stability Assay of ginsentide G1**.** HPLC profile of GS1 with or without (control) the presence of 0.2 M HCl incubated at 37^{o}C for two hours**.**

3. Discussion

Ginseng, the “King of Herbs“, is probably the most well studied medicinal plant of all time. Phytocomponents discovered in ginseng thus far mostly consists of small molecular compounds such as ginsenosides, polysaccharides, polyacetylenic alcohols and fatty acids [120]. Little is known about its CRP constituents. In this study, I have isolated and characterized five novel ginsentides from *Panax ginseng* and *Panax notoginseng*. This is the first time that CRPs in ginseng have been reported. They displayed extraordinary stability against acid, heat and digestive enzymes suggesting that they may play an important role in the medicinal properties of ginseng.

3.1 Ginsentides as species-specific markers for rapid authentication

With over 5,000 research papers on ginseng, its phytochemistry has been extensively investigated since early 19th century. Its CRP constituents, surprisingly, have never been elucidated for all these years despite of their high abundance in ginseng plants. MS profiling of three popular ginseng species including *P. ginseng, P. notoginseng* and *P. quinquefolius* revealed unique ginsentide patterns for each species, which could serve as species-specific markers for rapid identification and authentication of different ginseng species. Authentication is vital to ensure accuracy and efficacy of herbs used during prescription. It is also important to ensure the healthcare of the consumers and to fight against counterfeit products.

Currently, some of the common methods used to distinguish different ginseng species are TLC, DNA fingerprinting and HPLC [121, 122]. These methods are laborious, time-consuming and have low sensitivity. The use of MS profiling, a highly sensitive analytical technique, focused on ginsentides would provide a fast, high

throughput and high sensitivity method for identification and authentication of ginseng species.

3.2 Expression and tissue specificity of ginsentides

Ginsentide mRNAs are the third most abundant transcripts in Korean ginseng rhizome [119] as well as in American ginseng root [123]. They rank number one as the most highly expressed proteins in the American ginseng flowers [124]. Their expressions are apparently lower in leaf tissues than in roots and flowers. Tissue-specificity study in this work is also consistent with these reports. Flowers gave the highest yield of ginsentides followed by roots. Ginsentides, however, were not detected in dried leaves under our experimental conditions. It is possible that leaves may produce a trace amount of ginsentides that escape our detection.

3.3 Connectivity mapping

With novel cysteine spacing pattern, it is of interest to characterize their disulfide linkages. Ginsentides have a unique cysteine-rich motif of CCXXCXC pattern and contain four disulfide bonds instead of usual three as most peptides of their size. Comparing with three-disulfide-bond CRPs, mapping of those with four disulfide structures are often more complicated with 105 possible connectivity patterns for each peptide. To address this problem, I utilized the step-wise S-reduction and S-alkylation method proposed by Gray *et al.* [125]. Combined with MALDI-TOF MS/MS, I have unequivocally determined the connectivity of disulfide bonds and demonstrated that they have a pattern of Cys I-IV, II-VI, III-VII and V-VIII. This is a novel scaffold that has not been reported for any known plant CRPs. Ginsentides possess a pseudo-cyclic structure with both N- and C-terminal cysteines cross-bracing by four disulfide bonds, forming a pretzel-like structure.

Besides the determination of cystine connectivity, the results obtained also reveal the susceptibility and order of reduction of individual disulfide bonds during reductive unfolding of ginsentide G1 as shown in Fig. 48. More solvent exposed Cys I-IV and Cys V-VIII bonds were first reduced. The remaining two bonds of Cys III-VII and I-VI broke down later as they are more shielded and thus less susceptible to attacks by the reducing agents. The reduction orders are as followed Cys I-IV > Cys V-VIII > Cys III-VII > Cys II-VI. Knowledge of their unfolding pathway will provide better understanding about their biosynthesis pathway and will greatly facilitate their chemical synthesis using SPPS by connecting the difficult-to-form bonds first enabling efficient synthesis.

3.4 Stability of ginsentides

Highly cross-linked structures of ginsentides conferred by four cystine linkages render them compactness and stability against heat, acid and digestive enzymes. Ginsentides, however, do slowly degrade overtime upon prolong heating. This explains their absence in red ginseng since long steaming of raw *Panax ginseng* at 100°C is required for red ginseng production. This harsh condition probably caused major degradation of ginsentides, which may explain the differences in pharmacological effects of different ginseng forms.

Stability of ginsentides also depends on their primary structures. Ginsentide G2, for examples, is slightly more stable to heat than ginsentide G1 as degradation proceeded slower for it. This may be attributed to a single residue difference between their primary sequences. The Gly_4 residue of ginsentide G1 is replaced by Ser in ginsentide G2. Since the adjacent residue at position 5 is also a glycine residue, ginsentide G1 with two consecutive Gly residues (GG bond) provide a higher conformational flexibility than ginsentide G2 (SG bond) at the same position.

The less rigid structure of ginsentide G1 allows increased solvent accessibility, making it more susceptible to hydrolysis and degradation.

3.5 Biosynthesis processing of ginsentides

Ginsentide encoding genes were discovered in a random genomic study of ginseng cDNA libraries. They were first described as transcripts that ranked third in expression and did not show homology to any known sequences deposited in Genbank [119]. Since they are ginseng-specific and highly abundant in ginseng, they were name ginseng-specific abundant protein (GSAP). GSAP cDNAs encode for precursor proteins consisting of 117-121 amino acids. However, information about their biosynthesis processing and the formation of mature products, the ginsentides, was not known until this study.

Ginsentide precursor proteins comprise of three major domains commonly seen in many plant CRPs (Fig. 55). They include a putative ER signal sequence, a propeptide region and a single ginsentide domain followed directly by a stop codon. The processing from precursor proteins to mature ginsentides probably required at least two proteolytic events. The first event is catalyzed by signal peptidase which cleaves off the ER signal peptide allowing the precursor proteins to depart from ER and enter the secretory pathway. This proteolytic processing probably happens after the folding of the ginsentide domain inside the ER. The second event is catalyzed by unknown protease that cleaves at the N-terminal side of the ginsentide domain. Although the exact identity of this protease is not known, there is a high possibility that it belongs to PLCPs (similar to N-terminal processing enzyme of cyclotides) with the characteristic conserved Leu residue two residues away from the cleavage side. This, however, does not explain the formation of ginsentide G4 which has an

additional Gly residue at the N-terminus as compared to other ginsentides. An alternative route of processing would provide a possible explanation for its formation.

Figure 55. Translated sequences of ginsentide precursors deduced from cDNA clones obtained from Genbank [119]. GSAP1 encodes for ginsentide G2 and G3, GSAP2 encodes for ginsentide G4 and GSAP3 encodes for ginsentide G1. The putative ER signal is predicted by SignalP-HMM software and shown in blue. The putative NTPP domain is shown in purple and the ginsentide domain in black. Cysteine residues are highlighted in yellow. Asterisks indicate stop codon.

3.6 Bioactivity of ginsentides

Transgenic expression of the whole ginsentide genes in Arabidopsis showed enhanced root growth [119]. They are up-regulated under water deficit condition [119]. Endogenous functions of ginsentides are thus related to water-stress response. To determine if ginsentides possess interesting biological activities that can be used for therapeutic applications, several bioassays were performed to screen for their pharmacological activities. They were found to have no observable antibacterial, cytotoxic, immunomodulatory or hemolytic functions. Ginsentides thus do not possess membranolytic activities like cyclotides despite their similarities in size, hydrophobicity and electrostatic properties. The large difference in primary sequences may explain for such difference in bioactivities. Another aspect worth

further investigation is to examine ion-channel related functions of ginsentides especially on Ca^{2+} channel considering their high sequence similarity with ω-conotoxin from *Conus californius.*

4. Conclusions

Ginsentides are no doubt among the top-expression proteins in ginseng, ranked third in rhizomes and number one in flowers. It is surprised that such abundant proteins have been ignored for all these year despites numerous research papers on ginseng. The unusually high abundance of ginsentides suggests their bioactive importance. Disulfide mapping of ginsentides showed that they possess a novel pretzel-knot structure which confers them the compactness and high resistance to heat, acid and digestive enzymes. Although biological activities testing of ginsentides were not determined in this study, I believed that ginsentides may constitute one of the active ingredients responsible for the therapeutic properties of ginseng.

Chapter 7

Discovery of Heat-stable Cyclotides and Their Chimeric Precursors in the Fabaceae Plant *Clitoria ternatea*

1. Introduction

Clitoria ternatea (CT) is a perennial climber well-known for its butterfly-shaped vivid-blue flowers. It belongs to the Fabaceae family and displays a broad-spectrum of medicinal usages in almost every plant part [126]. It is often cited in the Ayuverdic system of Indian medicines as an effective antidepressant, antimicrobial, antipyretic and nerve tonic for enhancing memory [126]. It is also used as an alternative medicine in America and other tropical Asian countries [126]. In Cuba, decoctions of roots and flowers are reported to have emmenagogue properties which promote menstruation and uterine contraction. The same but stronger decoction is used as a vaginal douche. The seeds are also laxative, expulse intestinal parasites, antihelminthic, diuretic, antipyretic and antidotal [126]. Studies on animals have shown that the aqueous extracts of the flowers and leaves have antihyperglycemic effect in rats [127]. Decoctions of roots and leaves elicit a wide spectrum of activities on central nervous system and have shown to enhance acetylcholine content in rat hippocampus [128].

Preliminary phytochemical screening of CT extracts in recent studies showed that the biological active fractions were rich in peptides and proteins while showing negative tests for alkaloids, saponins, flavonoids, coumarins and lignans [128, 129]. In these studies, the plant extracts were prepared by boiling the pulverized plant powder with hot water. Although the exact chemical components are unidentified, it

is plausible to speculate that the active principles are heat-stable proteins, which are indicative of CRPs in the CT extracts.

As a part of discovery program of plant biologics in traditional medicines, we profiled the heat-stable CRPs of *Clitoria ternatea* with the expectation to identify novel A1b-like peptides. A1bs (also known as leginsulins) are open-end CRPs characteristic of the Fabaceae family and have been shown to present abundantly in several legume species [130, 131]. They consist of about 35 to 40 amino acid residues and three disulfide bridges [130]. A1bs are highly stable peptides and able to survive heat, acids and digestive enzymes in the porcine stomach and intestine [132, 133]. They have been shown to possess several biological activities such as insecticidal and hormonal functions in plants [134, 135]. They also affect mammalian physiological functions such as regulation of glucose metabolism in mice by binding to the voltage-dependent anion channel 1 (VDAC-1) [136].

Screening of CT extract led to identification of 35 novel CRPs, cliotide T1-14 (cT1-14), in which the number indicates the order of discovery. These CRPs, surprisingly, are not A1bs but turn out to be cyclotides characterized by their cyclic-cystine-knot motif. This is the first time cyclotides are reported from the Fabaceae family besides three known cyclotide-producing families: Rubiaceae, Violaceae and Cucurbitaceae.

Fabaceae is the second most economically important family after the Poaceae (grass family) [137]. It contains more than 19,000 species and is the third largest family of flowering plants [138, 139]. In addition to its agricultural importance as food crops, the Fabaceae is a valuable source of medicinal compounds. It is thus of great interest to discover the presence of cyclotides in the Fabaceae family potentially unlocking a novel source of druggable biologics.

In this work, I report the isolation, structural, functional and genetic characterizations of the novel cliotides from every plant part including leaves, stems, roots, nodules, flowers and seeds. My results show unequivocally that cliotides, at the protein level, belong to the cyclotide and not A1b family. They are found to preserve the cystine-knot motif, active against *E. coli* and cytotoxic to HeLa cell. These findings suggest that cliotides may be the active ingredients responsible for the indications of CT as anti-infectives and anti-tumors in traditional medicines [140]. Genetic characterization of novel cliotides, however, reveals unexpectedly that cliotides are derived from an entirely new arrangement differing from the known cyclotides of Rubiaceae and Violaceae families. Cliotides, at the gene level, have chimeric structures containing genetic elements from both A1b of the Fabaceae and cyclotides of the Rubiaceae and Violaceae families. This unusual arrangement suggests a novel mode of biosynthetic processing of cyclotides in the Fabaceae family and provides hints about their evolution in plants.

2. Results

2.1 Screening of heat-stable biologics in CT flowers

Heat-stable CRPs were extracted from CT flowers by incubation with boiling water at 100^{o}C for 30 min. The aqueous soluble fraction was separated from the denatured protein precipitates and other plant debris by centrifugation. A control experiment without the heat treatment was performed in parallel. The aqueous extracts were subjected to MALDI-TOF MS to detect the presence of putative biologics within the 1 to 10 kDa mass range. The mass spectrum revealed a group of peptides with strong *m/z* intensity at around 3 kDa (Fig. 56A). Comparison of the RP-HPLC profiles before and after the heat treatment revealed that >80% of these peptides survived the hot water incubation (Fig 54B, 54C). Their disulfide contents were subsequently analyzed by S-reduction with DTT and S-alkylation with IAA followed by MS analysis. The number of disulfide bonds was deduced by comparing the mass difference before and after reductive S-alkylation. Each S-alkylated half-cystine residue causes a mass shift of 58 Da. Most of these compounds displayed a mass shift of 348 Da suggesting the presence of three disulfide bridges, a common structural feature found in many plant biologics.

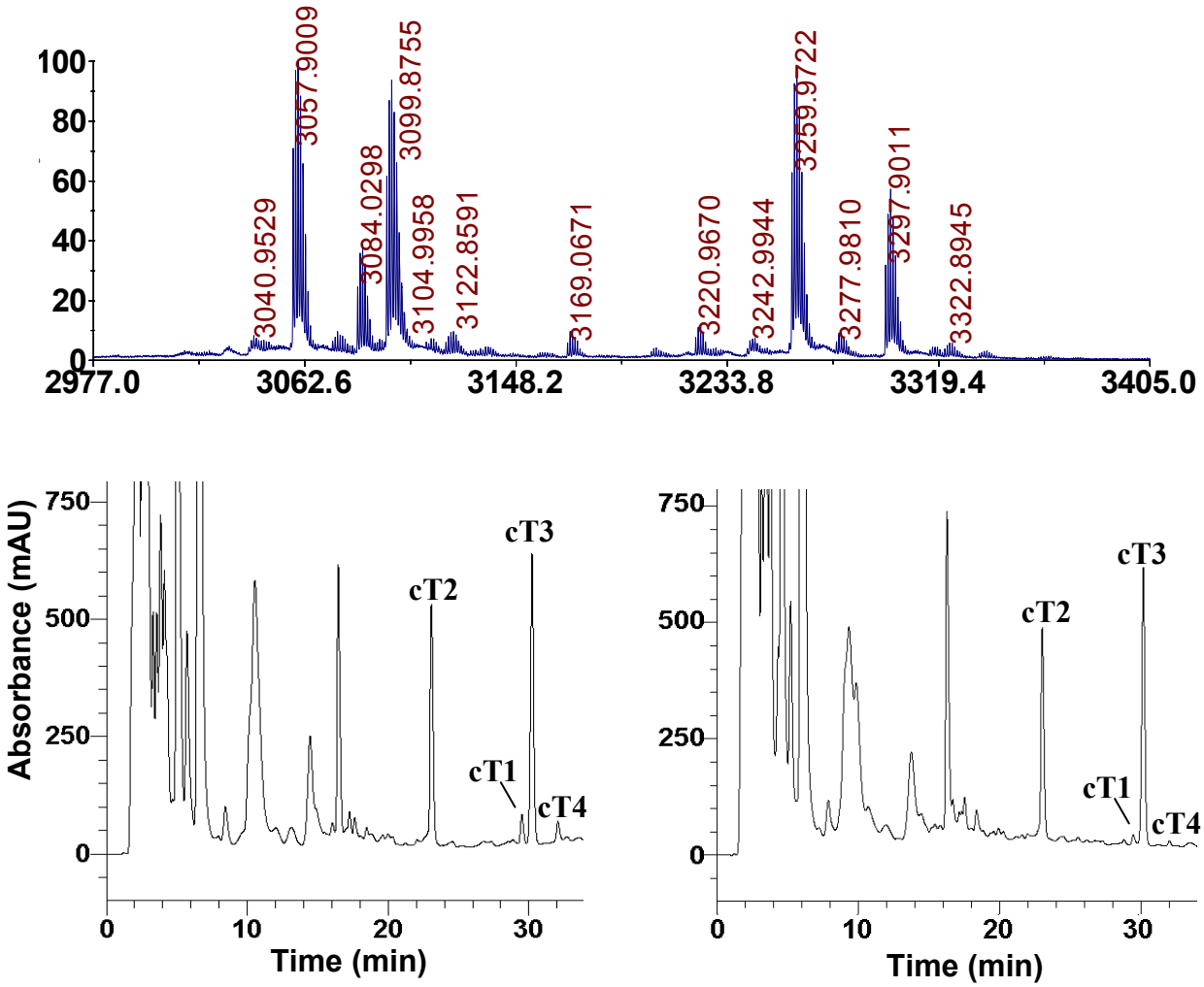


Figure 56. Thermal stability of biologics in the C. ternatea flowers. (A) MS spectrum of the heat stable aqueous extract of the *C. ternatea* flower. (B, C) HPLC profile of the flower extract before (B) and after (C) heat treatment at 100^{o}C for 30 min. The main peaks were labeled according to their cliotide contents.

2.2 Peptide isolation and sequencing: cyclotides as heat-stable biologics in CT

To isolate sufficient CRP samples for *de novo* sequencing and biological activities studies, 1 kg of fresh plant materials (whole plant) were collected and extracted with hot water (100^{o} C). Using RP-HPLC, we isolated 33 novel CRPs. Initial attempts for direct sequencing of the intact CRPs using MS/MS were unsuccessful due to insufficient fragmentation. In addition, proteolytic treatments with enzymes such as trypsin, chymotrypsin, or Glu-C failed to generate any detectable fragments. The resistance of the native CRPs against enzymes and CID fragmentation was in part attributed to the presence of cross-linking disulfide bridges that provide structural stability, but also suggestive of a highly compact framework such as a macrocyclic without free N- or C-terminal end.

To elucidate their primary sequences, each HPLC-purified peptide was SS-reduced, S-carbamidomethylated and digested with trypsin, chymotrypsin or Endo Glu-C. Generated fragments were sequenced by tandem mass spectrometry. As an example, MS/MS sequencing of cliotide T1 is shown in Fig. 57. Cliotide T1 and its S-alkylated form had *m/z* at 3084 and 3432 Da, respectively. Enzymatic digestion of S-alkylated cliotide T1 by Endo Glu-C gave a single fragment with a mass increase of 18 Da suggesting the hydrolysis of a cyclic amide backbone. Its circular structure was confirmed by tryptic and chymotryptic digestions which yielded end-to-end overlapping fragments. Sequence analysis of cliotide T1 revealed 84% identity to circulin A, a prototypic cyclotides, further suggesting that it is a novel member of the cyclotide family. Using this approach, we have successfully sequenced 33 novel cyclotides and uncyclotides, cliotide T1 to cliotide T35 (Table 14).

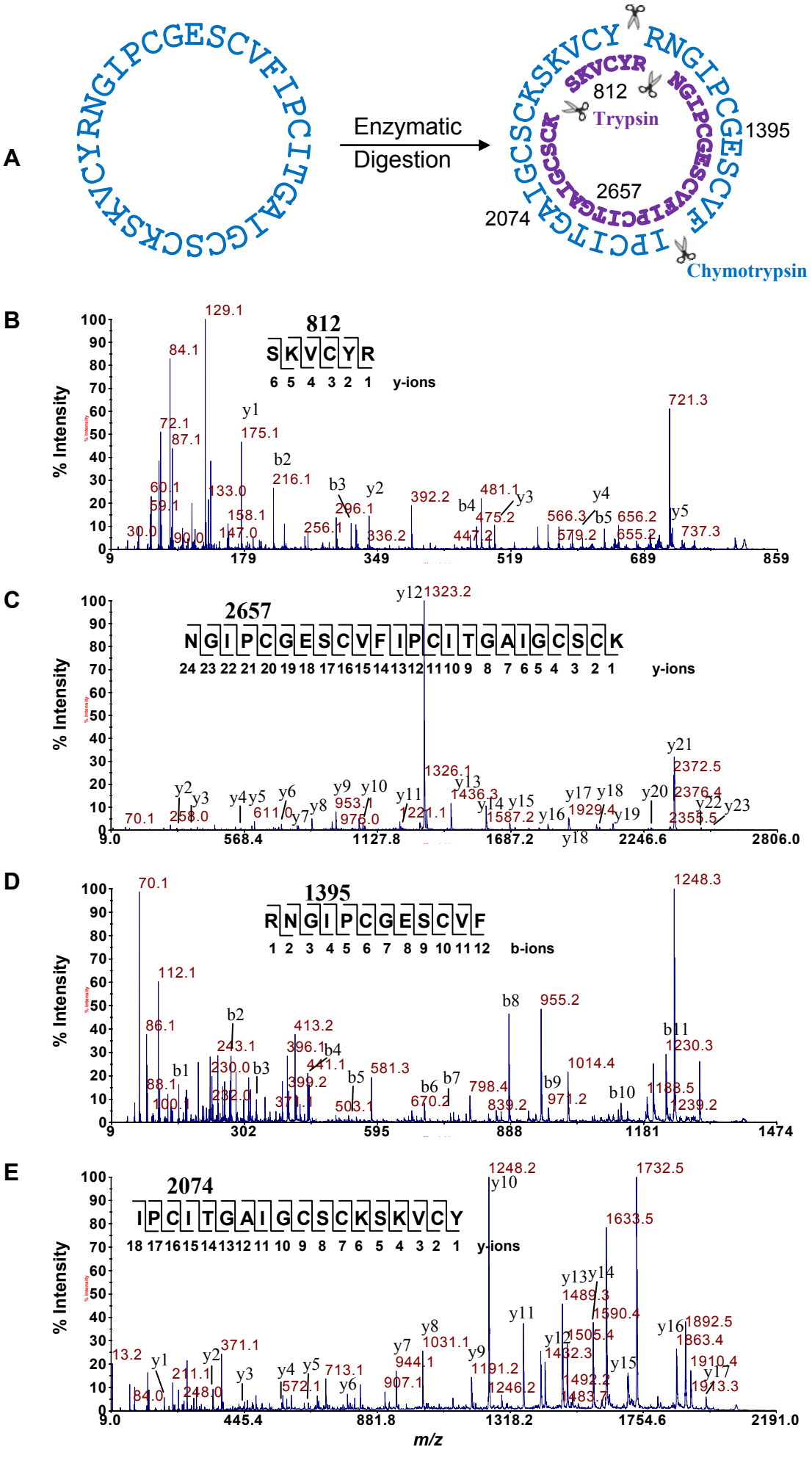


Figure 57. MS/MS sequencing of cliotide T1. (A) Tryptic digestion of S-alkylated cliotide T1 yielded 812- and 2657-Da fragments; and chymotryptic digestion yielded 1395- and 2074-Da fragments. (B-E) MS/MS spectra of digested fragments. Analysis of these spectra gave overlapping sequences indicative of a cyclic backbone of cliotide T1.

Table 14. Novel cyclotides in Clitoria ternatea

Cliotides	Amino Acid Sequence	MW	Method	Yield
cT1	GIPCGESCVFIPC-ITGAIGCSCKSKVCYRN	3084	P,cDNA	3
cT2	GEFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3260	P,cDNA	5
cT3	GLPTCGETCTLGTC-YV--PDCSCSWPICMKN	3058	P,cDNA	70
cT4	GIPCGESCVFIPC-ITAAIGCSCKSKVCYRN	3098	P,cDNA	3
cT5	GIPCGESCVFIPC-ISTVIGCSCKNKVCYRN	3169	P,cDNA	1
cT6	SIPCGESCVYIPC-ITTIVGCSCKNSVCYSN	3119	P	
cT7	GIPCGESCVFIPCTVTALLGCSCKDKVCYKN	3227	P,cDNA	50
cT8	GIPCGESCVFIPC-ISSVVGCSCKSKVCYNN	3072	P,cDNA	50
cT9	GIPCGESCVFIPC-ITTVVGCSCKNKVCYNN	3127	P,cDNA	50
cT10	GVPCAESCVWIPCTVTALLGCSCKDKVCYLN	3251	P,cDNA	50
cT11	GVIPCGESCVFIPC-ISTVIGCSCKNKVCYRN	3268	P,cDNA	
cT12	GVPCAESCVYIPCTVTALLGCSCKDKVCYKN	3243	P	50
cT13	GIPCAESCVWIPCTVTALLGCSCKDKVCYLN	3265	P	15
cT14	GVPCAESCVFIPCTITALLGCSCKDKVCYKN	3241	P	5
cT16	RIPCGESCVWIPCTITALVGCACHEKVCYKS	3363	cDNA	
cT17	KIPCGESCVWIPC-LTGYFGCYCQSKVCYHN	3377	cDNA	
cT18	GTVPCGESCVFIPC-ITGIAGCSCKNKVCYLN	3155	P,cDNA	
cT19	GSVIKCGESCLLGKC-YT--PGCTCSRPICKKN	3124	P,cDNA	
cT20	SVIKCGESCLLGKC-YT--PGCTCSRPICKKN	3067	P,cDNA	
cT21	GSVIKCGESCLLGKC-YT--PGCTCSRPICKKD	3125	P,cDNA	
cT22	GSAIRCGESCLLGKC-YT--PGCTCDRPICKKN	3152	P,cDNA	4
cT23	SAIRCGESCLLGKC-YT--PGCTCDRPICKKN	3095	P,cDNA	
cT24	AIRCGESCLLGKC-YT--PGCTCDRPICKKN	3026	P,cDNA	
cT25	IRCGESCLLGKC-YT--PGCTCDRPICKKN	2955	P,cDNA	
cT26	GSAIRCGESCLLGKC-YT--PGCTCDRPICKKD	3153	P,cDNA	1.5
cT27	DIQCAETCVHSPC-I---GPCYCKHGIICYRN	3102	P	0.5
cT28	GSVIGCGETCLRGRC-YT--PGCTCDHGICKKN	3107	P	0.4
cT29	VIGCGETCLRGRC-YT--PGCTCDHGICKKN	2981	P	
cT30	GLPICGETCFTGTC-YT--PGCTCSYPVCKKN	3022	P	
cT31	DTTPCGESCVWIPC-VSSIVGCSCQNKVCYQN	3299	P	
cT32	DTIPCGESCVWIPC-ISSIIGCSCKDKVCYHN	3349	P	

cT33	GEFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3278	P,cDNA
cT34	EFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3221	P,cDNA
cT35	FLKCGESCVQGEC-YT--PGCSCDWPICKKN	3092	P,cDNA
cT36	LKCGESCVQGEC-YT--PGCSCDWPICKKN	2945	P,cDNA

[a]Average isotopic molecular weight. Cys residues that form the cystine knot are highlighted in yellow, Pro residues that form the Möbius structure in gray, negatively charged residues in red and positively charged residues in blue.

2.3 cDNA cloning revealed chimeric structures of cliotide encoding genes

To determine their encoding genes and to provide insight about their biosynthesis, we designed degenerate primers for 3′-RACE PCR based on MS/MS determined sequences. A total of 11 partial clones were obtained. Afterward, a series of specific primers based on newly obtained genetic sequences were used for 5′-RACE PCR from which nine complete cliotide genes were obtained. These clones were named *ctc1, ctc2, ctc3, ctc4, ctc5, ctc7, ctc8, ctc9, ctc10, ctc11, and ctc15* according to their cliotide inserts.

The translated sequences of cliotide clones are shown in Fig. 58. They feature an entirely different arrangement from previously published cyclotide precursors from Rubiaceae and Violaceae families, such as *oak1* and *voc1* [12], which contain an endoplasmic reticulum (ER) signal sequence, an N-terminal pro-domain (NTPP), an N-terminal repeat region (NTR), a cyclotide domain, and a short C-terminal tail. Sequence analysis revealed that cliotide precursors contained no NTPP and NTR regions with the ER signal sequence followed directly by the cyclotide domain. In addition, a novel cysteine-rich domain was found at the C-terminal end of the precursor proteins separated from the cyclotide domain by a short putative linker region. Blast search showed that this novel domain was highly homologous to A1a domain (Albumin-1 chain a) of A1 genes found in many legume species [141, 142]. Surprisingly, the gene architecture of cliotides displayed striking similarity to those of A1 genes (Figure 4) which contain a signal peptide followed by A1b domain, linker peptide, and A1a domain [143]. The A1b domain in pea seeds encodes for a 37-residue insecticidal peptide containing three disulfide bonds [142]. The cliotide genes shared exactly the same arrangement with A1 genes except that the A1b domain was replaced by the cyclotide domain. The homology of cliotide

genes with both A1 genes (Fabaceae) and cyclotide genes (Rubiaceae, Violaceae) makes them naturally occurring chimeric genes.

2.4 Cliotide genes contain a single intron at the ER region

Known families of CRP genes are generally characterized by the presence of a single intron within the ER domain [10, 144-147]. To compare the genetic structure of cliotides at both DNA and mRNA level, five cliotide precursor genes including cliotide T2, T3, T7, T8 and T10 were cloned from leaf DNA using cDNA-derived sequence as primers. The DNA clones revealed a single intron located in the signal peptide region of the cliotide genes, which is similar to leginsulin genes from Fabaceae and cyclotide genes from Rubiaceae family (Fig. 59). Thus, the single-intron architecture of cliotide DNA genes is consistent with genes of CRP families which include not only cyclotides and leginsulin, but also plant defensins and thionins [144-146]. It is interesting to note that cliotide T7 gene contains no intron similar to those characterized from Violaceae family.

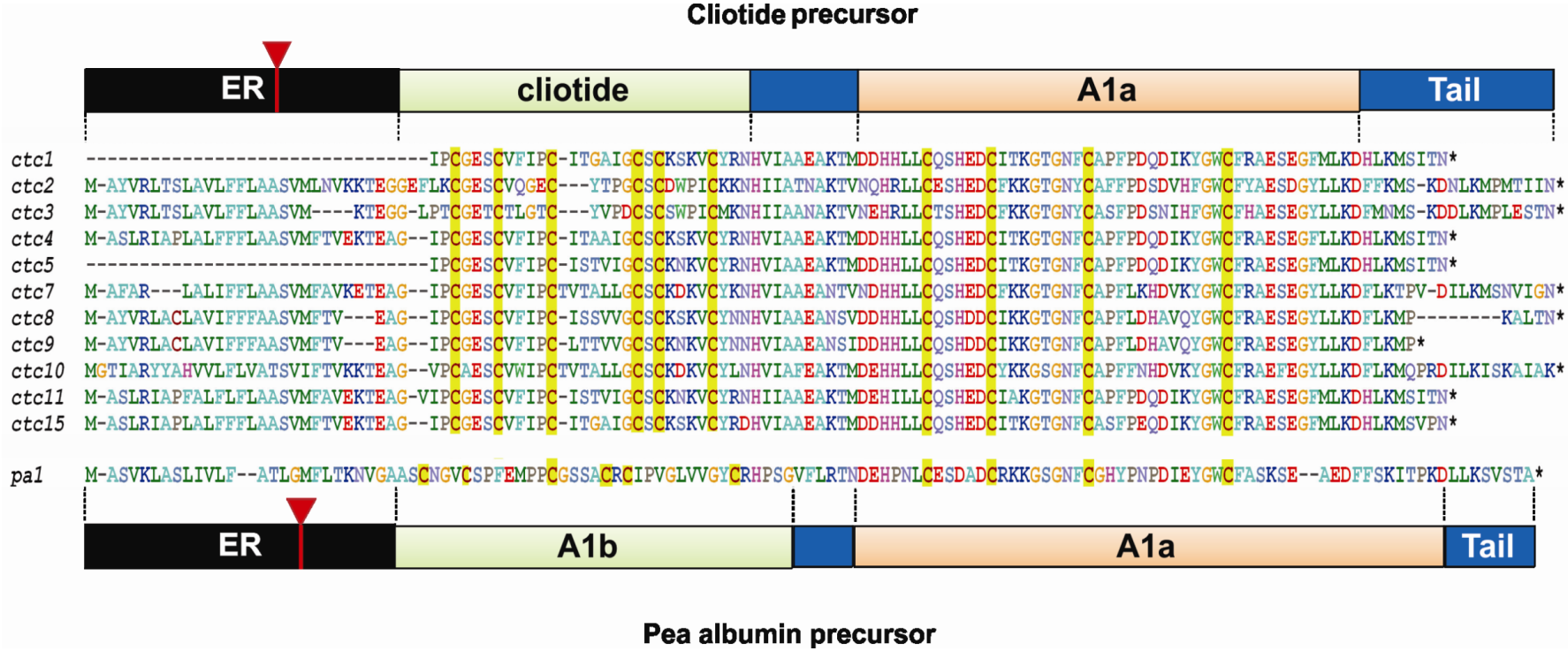


Figure 58. Deduced amino acid sequences of cliotide precursors. They display a novel gene arrangement as compared to previously known precursors from Rubiaceae and Violaceae family. The precursors contain an ER signal region directly followed by the mature cyclotide domain, a short linker region, A1a domain and a short C-terminal tail. This gene organization resembles that of pea albumin 1 (PA1) precursor except that the A1b domain is replaced by cyclotide domain. The *pa1* gene can be found in GenBank with the assession number CAB82859. The inverted triangles indicate the introns location.

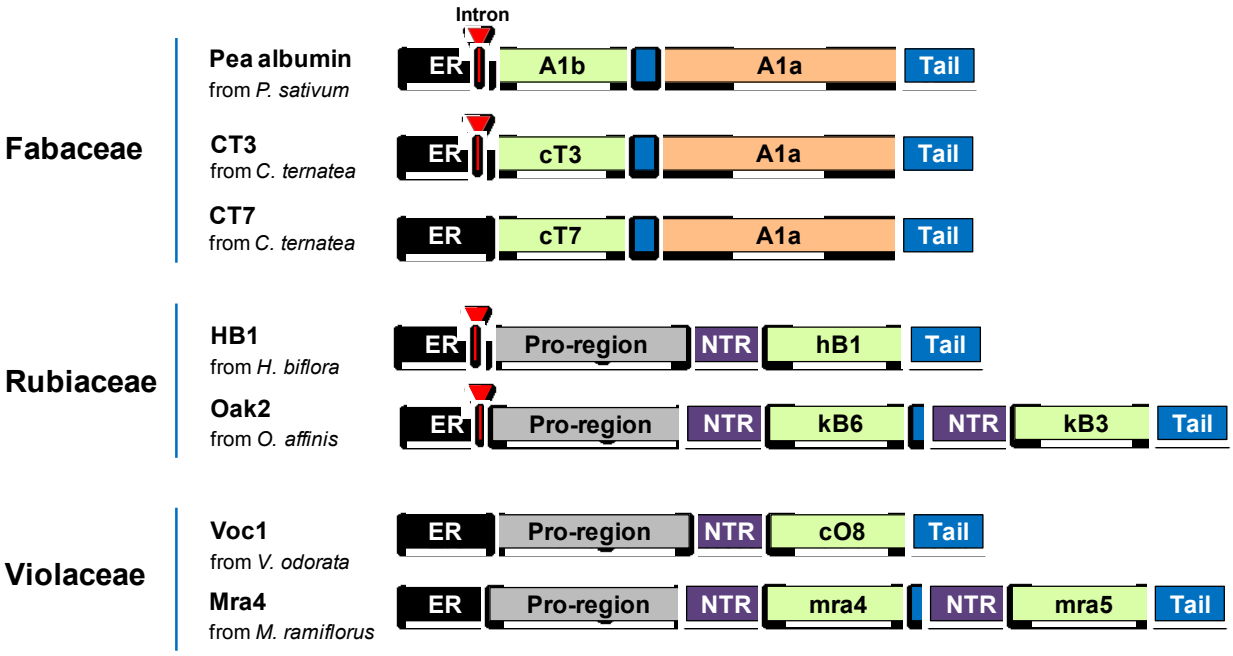


Figure 59. Schematic comparison of the genetic arrangements of cyclotide (Rubiaceae and Violaceae), cliotide and A1 (Fabaceae) [53, 148]. Signal peptides are highlighted in black, N-terminal prodomain in gray N-terminal repeat in purple, A1b and cyclotide domains in green, A1a in orange. The inverted triangles indicate the introns located in the signal peptide regions.

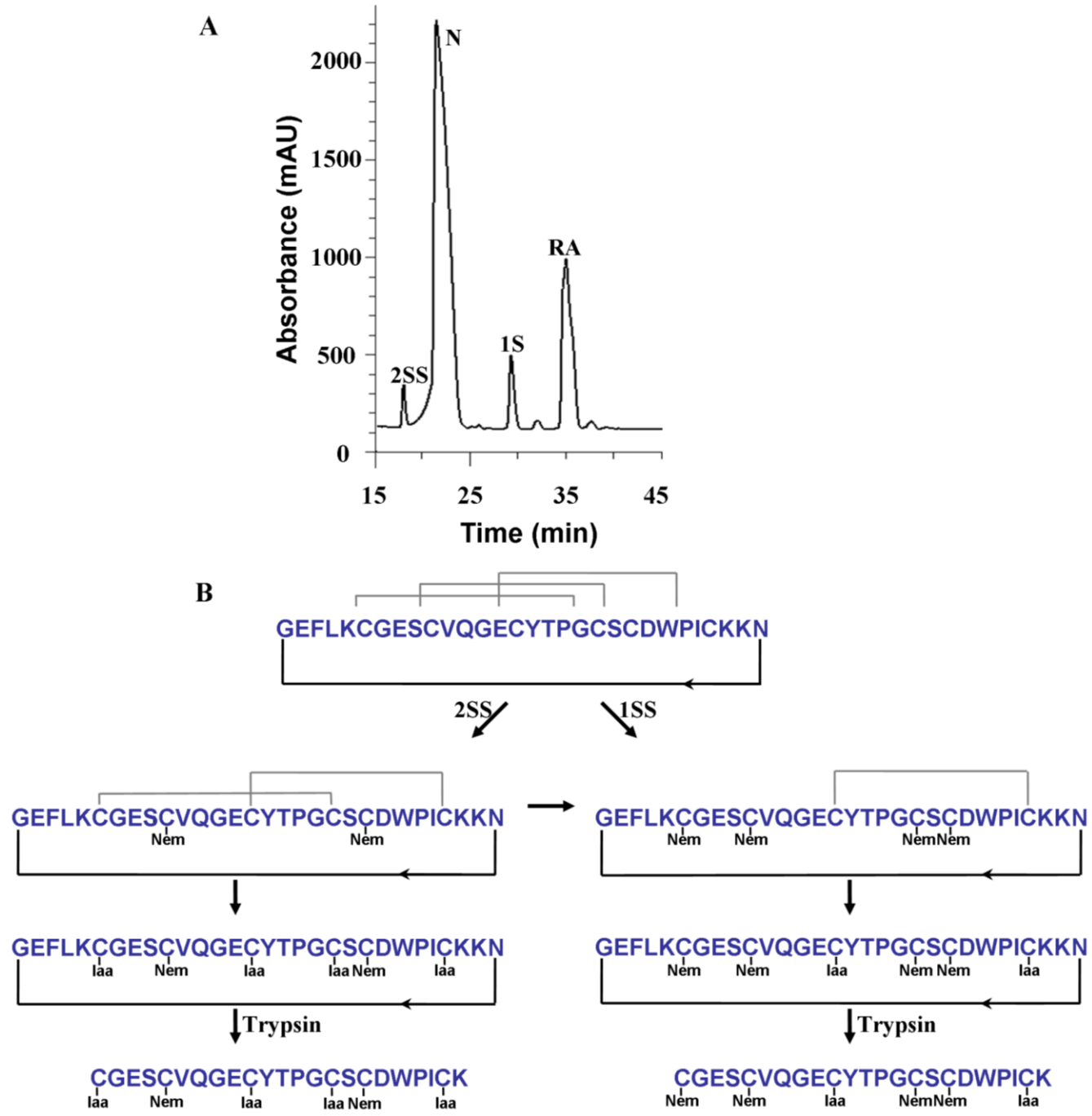
2.5 Disulfide mapping of cliotide T2: New precursor, same Connectivity

The disulfide connectivity of novel cliotides was sought to determine if they possessed a different arrangement than the cystine-knot motif found in cyclotides. Cliotide T2, present in high abundance in the floral tissue, was selected as a representative. The connectivity of cliotide T2 was determined by a differential S-tagging strategy through a sequential S-reduction and S-alkylation [78]. The SS bonds of native cliotide T2 were first partially ruptured with TCEP to generate a series of isoforms with one or two disulfide bonds being reduced. The released thiols were immediately S-alkylated with an excessive amount of NEM. The whole process was performed under acidic condition at pH 3.5 to avoid the scrambling of disulfide linkage. The partially S-reduced and S-alkylated peptides were purified by RP-HPLC.

Four chromatography-separated peaks were collected Fig. 60 and analyzed by MALDI-TOF MS to determine the number of NEM-tagged cysteines. Each NEM-tagged cysteine caused a mass increase of 126 Da. The mass gain after S-alkylation was then used to deduce the number of reduced disulfide bonds. Peak 1 had *m/z* at 3512 Da, which was 252 Da larger than the original peptide (3260 Da). Based on the mass gain, we could assign the intermediates eluted in peak 1 the 2SS intermediate having two NEM-labeled cysteine residues and two intact disulfide bonds, peak 2 the native peptide with intact cystine core, peak 3 the 1SS species with four NEM-tagged cysteine residues and one remaining disulfide bond (*m/z* at 3764 Da), and Peak 4 with all six cysteines tagged by NEM. It is noteworthy that the 2SS species was more hydrophilic than the native peptide whereas the 1SS species and fully NEM-tagged peptides became more hydrophobic upon alkylation.

To obtain the connectivity, the remaining disulfide bonds of 1SS and 2SS species were again S-reduced with DTT and then S-tagged with IAA, a second alkylation reagent. S-alkylated peptides were digested with trypsin and the resulting fragments were analyzed by MS/MS. Tryptic digestion generated three proteolytic products with two fragments containing no cysteine residue (K and NGEFLK) and thus were universal for both IAA derivatives of 1SS and 2SS species (Fig. 59B). The third fragment was unique for each isoforms and contained all of the six cysteine residues. MS/MS analysis of this fragment provided unambiguous assignment of disulfide bridges based on the differential S-tagging of NEM and IAA on the cysteine residues.

The connectivity of 2SS species was established as Cys II-V with two NEM-tagged groups on Cys II and Cys V and four IAA-tagged groups on the remaining cysteine residues (Fig. 59C). Similarly, the connectivity of 1SS species was identified as Cys III-VI with four NEM-tagged groups on Cys I, II, IV, and V and two IAA-tagged groups on Cys III and Cys VI (Fig. 59D). The third disulfide bond of Cys I-IV was obtained by deduction. Taken together, these results provided evidence for the knotted cystine arrangement of cliotide T2, a disulfide connectivity common to both cyclotides and A1bs.


A
N
RA
1S
2SS
2000
1500
1000
500
0
Absorbance (mAU)
15
25
35
45
Time (min)
B
GEFLKCGESCVQGECYTPGCSCDWPICKKN
2SS
1SS
Nem
Iaa
Trypsin
CGESCVQGECYTPGCSCDWPICK

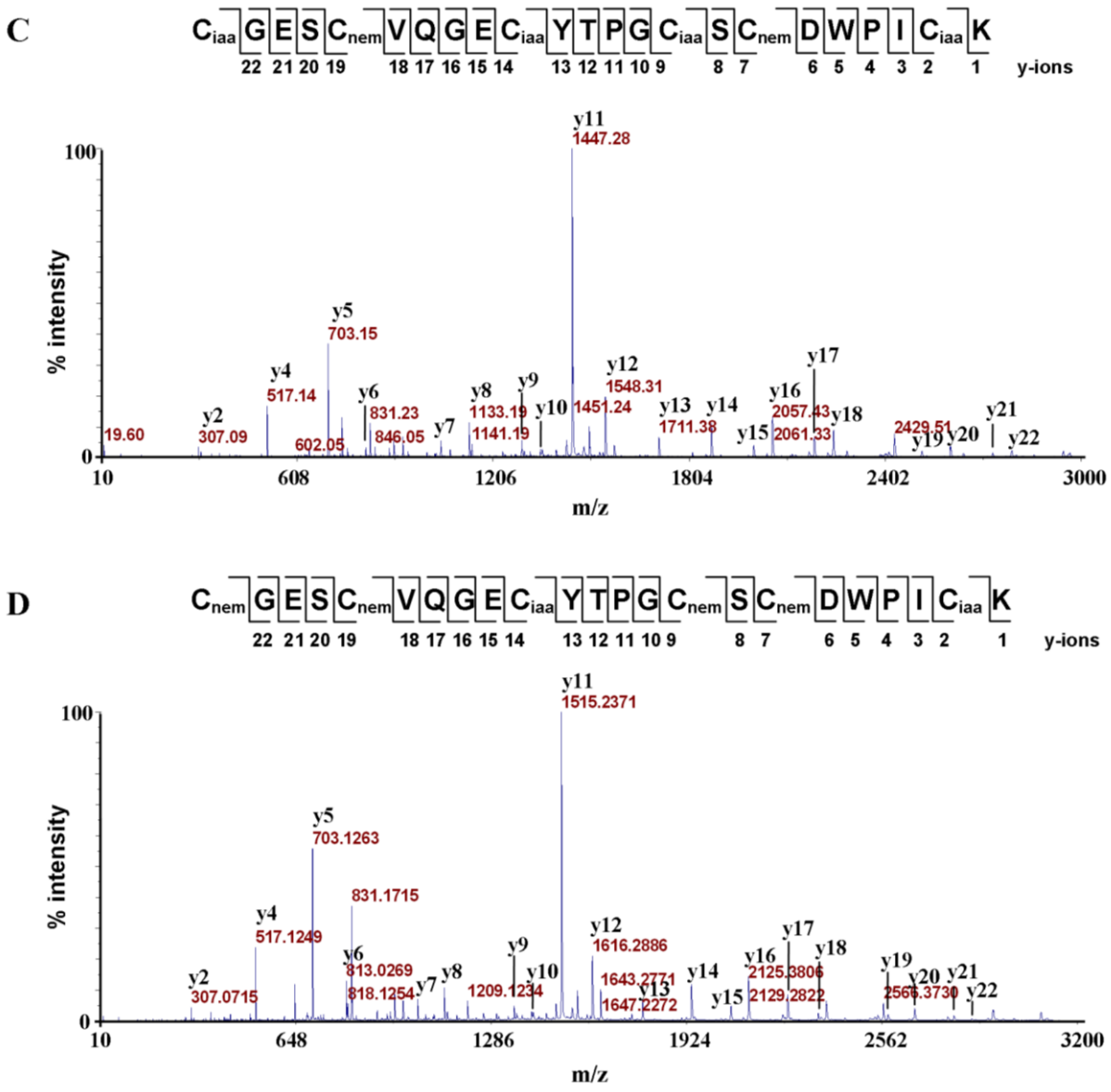


Figure 60. Disulfide mapping of cliotide T2. (A) RP-HPLC profile of partially S-reduced and S-NEM labeled cliotide T2. 1SS and 2SS are intermediate species with one and two intact disulfide bonds respectively. N is native peptide; RA is fully S-NEM labeled peptide. (B) Schematic presentation of disulphide mapping of cliotide T2. 1SS and 2SS species were isolated, reduced with DTT and subsequently S-alkylated with IAA. S-alkylated peptides were treated with trypsin and fragments are sequenced by MS/MS. (C, D) MALDI-TOF MS/MS spectra of trypsin-digested peptide fragments.

2.6 Tissue-specific distribution

To determine if the novel cliotides are expressed universally or in a tissue-specific manner, MALDI-TOF MS was utilized to profile their expression in different plant parts. A total of eight different tissues were collected including leaves, stems, roots, flowers, pods, seeds, nodules and shoots. They were extracted separately using 50% ethanol and the extracts were analyzed by mass spectrometry. The mass spectra revealed that each tissue has a unique profile distinct by the number and amount of cliotide being expressed (Fig. 61). Each tissue produced at least 10 different cliotides but only a few specific cliotides dominated the expression profiles. For instance, cliotide T2 and cliotide T3 were highly expressed in the flowers and pods, accounting for >70% of total cyclotides in these tissues. Similarly, cliotide T8, cliotide T9 and cliotide T10 were the major constituents in seeds. However, not all cliotides are tissue-specific. Certain cliotides such as cliotide T1 and cliotide T4 were found to express in all examined tissues, suggesting their house-keeping role in plant defense.

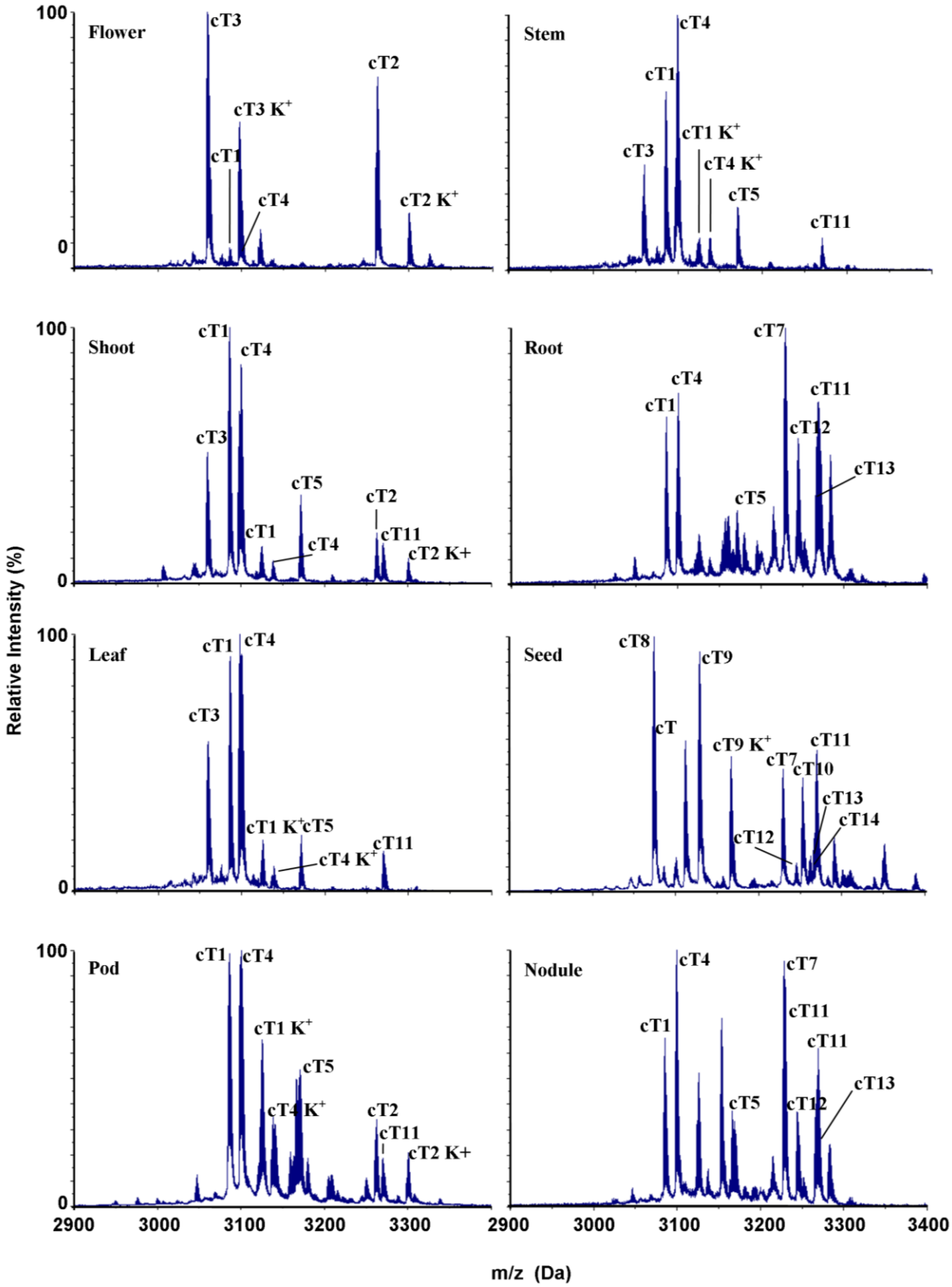


Figure 61. Tissue-specific distribution of cliotides in Clitoria ternatea. Eight different plant parts were collected including leaf, flower, pod, shoot, seed, stem, root and nodule. The extracts were profiled by mass spectrometry. The peaks are labeled according to the masses of corresponding cliotides. Peaks with K+ labeled indicate the potassium adduct (+38 Da).

2.7 Antimicrobial activity

To assess the antimicrobial activity, four cliotides (cliotide T1-4) were tested against a panel of five different bacterial strains using radial diffusion assay. Their selection was primarily based on abundance and structural characteristics with two Möbius and two bracelet cyclotides. Table 14 summarizes their antimicrobial activity with D4R, a synthetic bactericidal peptide dendrimer, as positive control [68]. The bracelet subgroup including cliotide T1 and cliotide T4 showed strong antimicrobial activity against three tested Gram-negative bacteria, with the MIC values at low μM range. Among these strains, *E. coli* is most susceptible to cliotides' bactericidal activity. Cliotide T1 and T4, however, were relatively ineffective against both tested Gram-positive bacteria *S. aureus* and *E. faecalis*. The Möbius subgroup comprising of cliotide T2 and cliotide T3 were inactive against all five tested bacterial strains up to 100 μM concentration.

2.8 Hemolytic activity

Cyclotides are membranolytic and cause lysis of red blood cells, and hemolysis has been a common assay for cyclotides. Hemolytic effect of cliotide T1-4 was conducted using human type A erythrocytes. Melittin, a known peptide from bee venom with strong hemolytic activity, was used as positive control. Cliotide T2 was relatively non-hemolytic and only lysed about 18% of the RBC at 40 μM. The HD_{50} of the other three cliotides ranged from 7.19 to 13.1 μM (Table 15), which were significantly less potent than melittin (3.21 μM).

2.9 Cytotoxicity against HeLa cell line

The cytotoxic property of cliotide T1-4 was assessed by MTT assay, which monitored mitochondrial activity in living cells, hence indicating cell viability. HeLa

cell line was selected as a model for the assay. Five concentrations were tested for each cliotides and IC_{50} values were determined from the survival curves. All of the four tested cliotides displayed sharp dose-response curves and IC_{50} values ranging from 0.6 to 8.01 μM (Table 15). Cliotide T1 and T4 were the most potent with IC_{50} value at 0.66 and 0.6 μM, respectively. Cliotide T3 has moderate cytotoxic effect with IC_{50} at 2.0 μM and cliotide T2 was the least cytotoxic with IC_{50} at 8.01 μM.

Table 15. Antibacterial, hemolytic and cytotoxic activities of selected cliotides

Peptide	MIC[a] (μM)			HD_{50}[b]	IC_{50}[c]
	E. coli	*K. pneumonia*	*P. aeruginosa*		
cT1	1.1	2.7	4.7	7.1	0.6
cT2	>100	>100	>100	>100	8.0
cT3	>100	>100	>100	13.1	2.0
cT4	1.0	5.5	7.5	8.4	0.6
D4R[d]	1.3	1.9	1.5	-	-
Melittin[e]	-	-	-	3.2	-

[a] Minimal Inhibitory Concentration. . [b] HD50 refers to the hemolytic dose cause 50% cell lysis. [c] IC50 refers to the inhibitory concentration to cause 50% death of HeLa cells. [d] Synthetic antibacterial peptide is used as positive control for antibacterial assay. [e] Melittin is used as positive control for hemolytic assay.

3. Discussion

This work describes the discovery of cyclotides as heat-stable biologics in *Clitoria ternatea.* 36 novel cyclotides, cliotides T1-36, have been characterized from the plant extracts. Their tissue-specific distributions profiled by mass spectrometry show that cliotides are present in every plant parts, including flowers, leaves, roots, nodules and seeds. The presence of cliotides in nodules is the first known example of cyclotides found in this tissue. It is also noteworthy because of their high abundance and possible roles in controlling the symbiosis of the host plant and nitrogen-fixing bacteria.

The identification of cliotides as cyclotides and not A1bs are unequivocal. A1bs previously discovered in the legumes are CRPs of similar sizes as cyclotides, but contain an open-end structure. Cyclotides have close-end structures with a head-to-tail cyclized peptide backbone. Using a combination of enzymatic fragmentations and MS sequencing, all 14 cliotides are shown to be cyclic proteins. Disulfide connective mapping of cliotide T2 confirms that it retains a cystine knot motif similar to cyclotides. The cyclic-cystine-knot arrangement has been shown to be the main contributing factor for the structural stability and heat resistance of cyclotides [9]. Although there are few examples of linear cyclotides, sequence and cysteine-spacing comparisons of these two classes of CRPs also firmly place cliotides in the cyclotide family. Cliotides contain the highly conserved Glu at loop 1 and the Asn-Gly/Ser dipeptide motif at loop 6 of the cyclotide family. The Asn-Gly/Ser sequence is the ligation site in the bioprocessing of a linear cliotide precursor to the cyclic form, and which will be discussed later.

The disulfide mapping of cliotide T2 provides useful information about its unfolding mechanism based on intermediate species isolated. The abundance of 2SS species formed by breaking of the Cys II-V bond suggests that this cystine linkage is most susceptible to reducing reagents. 1SS species is subsequently generated by reducing of Cys I-IV bond with the remaining of Cys III-VI bond as the most stable disulfide linkage. This is consistent with our understanding of cyclotides' cystine knot structure that the penetrating disulfide bond of cystine-knot motif (Cys III-VI) is the most shielded and buried deep inside the cystine-core explaining its resistance against reducing reagents [77].

Functionally, cyclotides are characterized by their membrane-active properties for a wide variety of biological actions [42, 149]. Among these, antimicrobial, hemolysis and cytotoxicity have been studied to dissect the membrane binding ability of cyclotides on three different membrane types of bacteria, erythrocytes and cancer cells [149]. These bioactivities were thus selected for functional evaluations of novel cliotides. Our results show that cliotides retain the characteristic functions of cyclotides with similar potency. Cliotide T1 and T4 are bactericidal against *E. coli* with MIC value at approximately 1 μM as compared to 1.55 μM of cyclopsychotride [21]. They are toxic to HeLa cell at IC_{50} of 0.6-8.01 μM range comparable to cytotoxic effects of several cyclotides such as varv A (IC_{50} 2.7-6.35 μM), varv F (IC_{50} 2.6-7.4 μM), and cycloviolacin O2 (IC_{50} 0.1-0.3 μM) [150]. They also have similar hemolytic activities (HD_{50} 7.9-13.4 μM) as compared to kalata B1 (11.7 μM) [82].

The discovery of cyclotides in CT expands their occurrence to a new plant family, the Fabaceae, also known as the legume family or pea family (Fig. 62). This finding is significant since the Fabaceae is the third largest family of flowering

plants having immense importance in medicine, agriculture and ecology. With more than 19,000 species distributed in almost all terrestrial habitats together with their unique ability to fix atmospheric nitrogen [137], the Fabaceae no doubt plays an essential role in maintaining the ecosystem and environment. They constitute part of nearly all human diet as a source of high protein food and vegetable oils. It is thus intriguing to speculate that the heat- and enzyme-stable cyclotides may have long been entering human food chain without our recognition. A possible and likely route is through animal feed since CT is a highly palatable forage legume generally preferred by livestock over other legumes and has been widely used as pastures for grazing of livestock in several countries such as India and Australia [151].

The presence of cyclotides in the Fabaceae also sheds new light about their distribution in plants. Thus far, the discovery of cyclotides has been confined to three plant families: the Rubiaceae, the Violaceae and the Cucurbitaceae with no report on their occurrence in other families [33, 35]. Our findings of cyclotides in the Fabaceae family suggest that their occurrence may be more diverse than originally thought. Recently, putative cyclotide-like genes have been reported in several monocot plants suggestive of the likely ancestral genes of circular proteins [32]. Screening by Gruber and colleagues in 2008 also provided evidences for the possible existence of cyclotide-like CRPs in Apocynaceae family but no primary sequences have been reported to date [33]. With increasing interest and research on cyclotides, we believe that more cyclotide-producing families will be discovered in future.

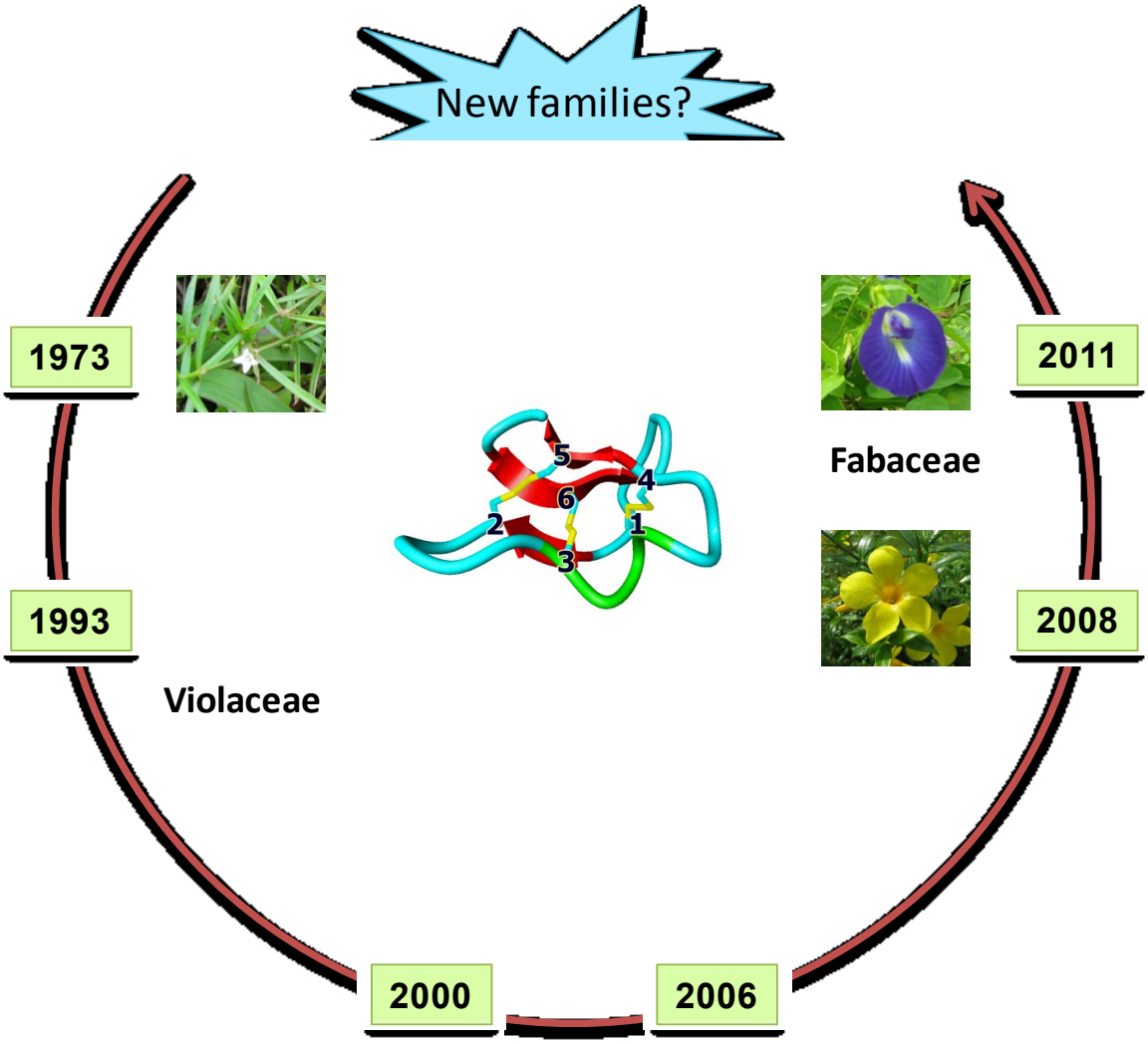


Figure 62. Distribution of cyclotides in different plant families and timeline of discovery.

Extraction of cyclotides in our study is achieved by using hot water demonstrating their heat stability. Although numerous extraction techniques, often involving organic solvents, have been used by the research community to isolate active principles, decoction preparation by soaking the herbs in boiling water is the most relevant method in traditional medicine. Heating of the plant extract at 100°C for 30 min showed nearly identical cliotide profiles before and after the treatments. Our results also demonstrate the thermal stability of cyclotides not only as pure compounds dissolved in water [9] but also in the

complex mixture of the plant extract which contains many herbal derived, small-molecule chemicals. Thus, our findings provide experimental support that cyclotides can survive heat treatment during the decoction preparation process and may be one of the active principles attributed for the medicinal values of CT.

Similar to many CRPs, our results show that the cliotide genes contain a single intron embedded in the ER signal peptide domain [144, 146]. Intriguingly, genetic characterization of cliotides also reveals unexpectedly that cliotides are derived from an entirely new arrangement differing from the known cyclotides of Rubiaceae and Violaceae families. Cliotide genes contain a hybrid coding region consisting of peptide domains from two unrelated genes. They present a novel case of genetic swapping between cyclotide and leginsulin genes resulting in new hybrid entities with half native, half foreign. In most legume species, the leginsulin genes encode for three major domains: ER, A1b and A1a [148]. Cliotide genes share essentially the same arrangement except the exchange of A1b for the cyclotide domain. This unusual arrangement raises another question if A1b-like peptides still express in CT, or if they have been totally replaced by cyclotides. Our attempts to identify A1b-like peptides in the CT extract have thus far failed to identify similar peak in the known mass range of A1bs. It is possible that they may be modified and present in low abundance that escapes our detection.

Cliotide precursors, unlike their cyclotide counterparts, are devoid of the NTPP and NTR regions typically separating the ER signal from the mature cyclotide domain. This suggests that cliotides are N-terminally processed by signal peptidases (SPase I) (Fig. 63). The absence of the NTPP and NTR regions also implies these protein domains may not be essential for the

biosynthesis of cyclotides. They instead may have other endogenous functions in plant. At the C-terminus, the cyclotide domains are probably cleaved by AEP (asparaginyl endopeptidase) as in Rubiaceae and Violaceae species indicating by the presence of the highly conserved Asn at the processing site. AEP is a cysteine proteinase and forms with the C-terminal Asn a reactive thioester bond that leads a head-to-tail ligation to afford the cyclized structure of a cyclotide. Interestingly, the conserved Leu residue locating at two residues downstream of the C-terminal Asn processing site is replaced by Val or Ile in cliotide genes. This reflects a difference in specificity at P2' position of AEP in CT.

The genetic displacement of A1b by cyclotides raises another question about the functional overlapping between these two peptide families. They are similar in size, with both possessing a cystine-knot scaffold and insecticidal activity [152]. A1b is known to exert physiological functions in plant such as growth, differentiation and cell proliferation by binding to a 43-kDa receptor locating at the plant cell wall [153, 154]. In contrast, little is known about the physiological functions of cyclotides, their intracellular location and their binding receptors in plants. The slow progress on cyclotides immunohistochemistry research is partly due to their low immunogenicity and difficulty to make antibody against cyclotides. Nevertheless, the substitution of A1bs for cyclotides in CT suggests that they may share certain similar functions and provide hints about the physiological functions of cyclotides.

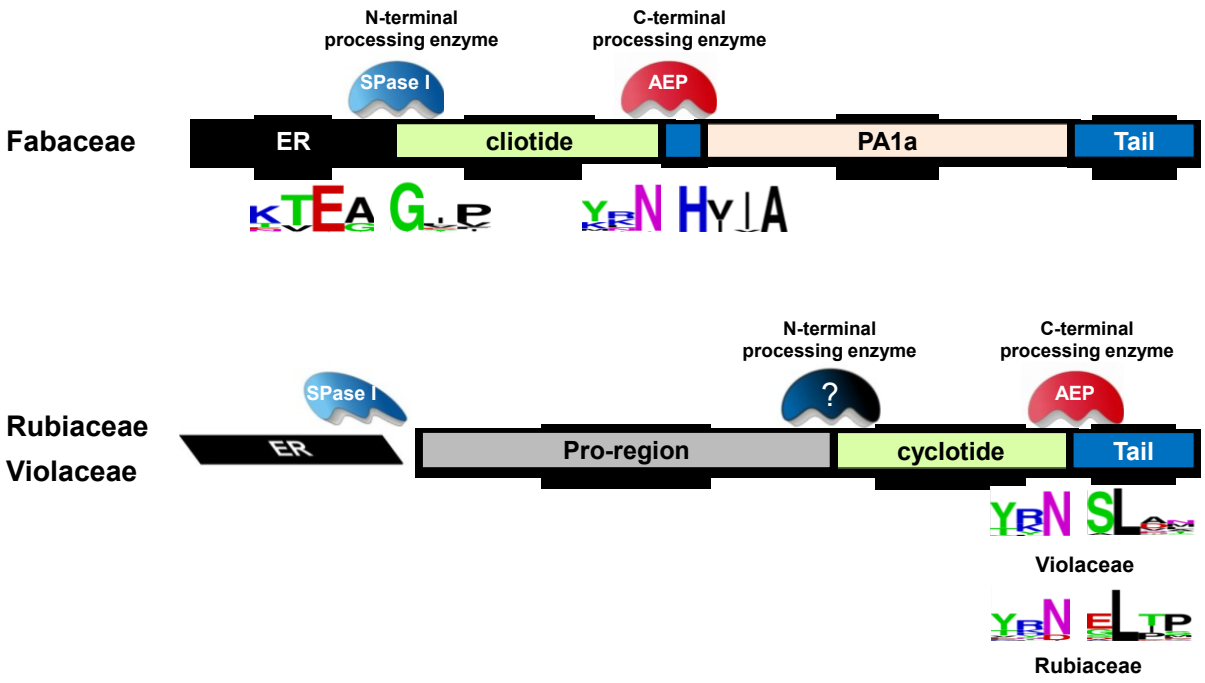


Figure 63. Proteolytic enzymes involved in cyclotide bioprocessing**.** In *C. ternatea* (Fabaceae), the signal peptidase I (SPaseI) responsible for ER excision is also the N-terminal processing enzyme of cliotides. Meanwhile, in Rubiaceae and Violaceae families, another unknown enzyme(s) is required to process N-terminus cleavage. The C-termini of three family are processed by the same Asparaginyl Endopeptidase (AEP) enzyme because of the highly conserved N/D residue at the processing site as shown in the sequence logos. The sequence logos were built using WebLogo server [155, 156].

How does the chimeric structure of cliotide genes occur? The cyclotide-encoding domains in CT may arise by vertical heredity, horizontal gene transfer or convergent evolution. In vertical heredity, cyclotide genes would have to exist in the ancestral legume lineages and are inherited in CT by means of sexual reproduction. In horizontal gene transfer (HGT) event, CT acquires the cyclotide domains horizontally from other species. Such scenario is supported by two lines of evidences. First, there is an unexpected cluster of CT with several Rubiaceae and Violaceae species using phylogenetic tree constructed with cyclotide,

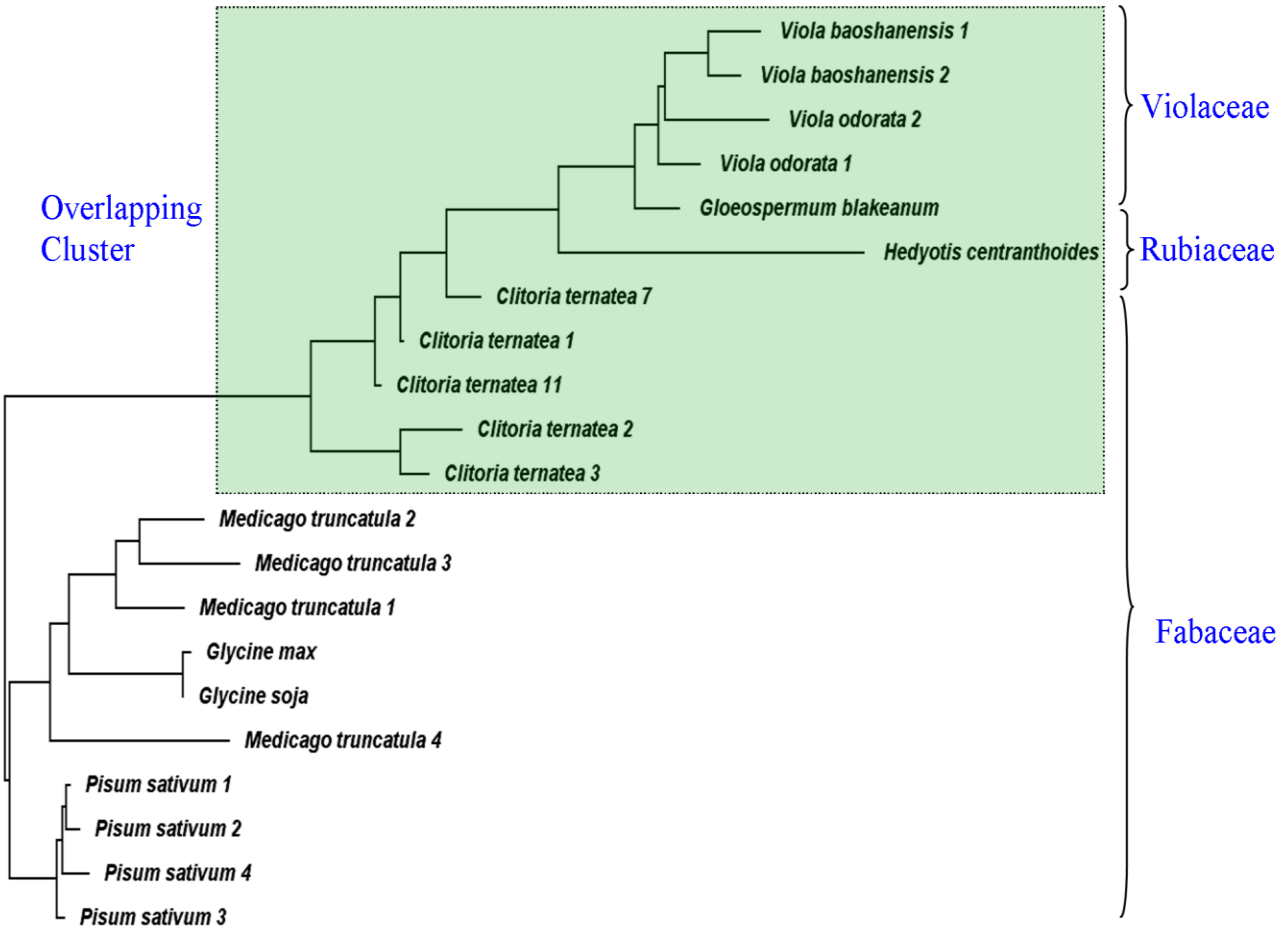


Figure 64. Phylogenetic tree showing the relationship among cyclotide, cliotide and A1 genes from different Rubiaceae, Violaceae and Fabaceae species. The precursor protein sequences were aligned using COBALT multiple alignment. The phylogenetic tree shows cluster of *C. ternatea* with several species of Rubiaceae and Violaceae families.

leginsulin and cliotide genes (Fig. 64). It should be noted that three different gene types are used for constructing the phylogenetic tree because of the chimeric nature of the cliotide genes. Second, DNA database search of 22 Fabaceae species, including two plants with full genome sequences *Glycine max* and *Medicago truncalata*, identified orthologs of leginsulin genes but none of them contains the cyclotide domain. It is thus possible that CT may have received the cyclotide genes from donor species of Rubiaceae or Violaceae families.

The third possibility is convergent evolution from A1b to cyclotides. Both are similar in size, share a cystine knotted structure and overlap in certain functions. For this to happen, it would require massive mutations and substitutions for complete conversion from A1b to cyclotides. An interesting sample of convergent evolution has recently been reported for the platypus [157]. The toxin genes identified from this animal show striking homology to those found in other venomous species such as snakes, lizards, starfish and sea anemones. Genetic analysis reveals that these venoms have evolved independently from different origins but end up in developing similar families of molecules [157, 158].

Genetic characterization of cyclotides in the Fabaceae may open up new avenues of research allowing genetic manipulation and production of cyclotide-based therapeutics in agriculturally important Fabaceae plants. There has been an increasing interest in using cyclotides as molecular scaffolds for grafting of bioactive epitopes. The current bottlenecks for chemical synthesis of these engineering peptides are high production cost and uncertainty in obtaining the correct folding. The use of plants as bioreactors can provide a cost-efficient solution and rapid production of cyclotides for medical and industrial applications.

4. Conclusions

Our data provide the first description of cyclotides in the Fabaceae. The novel cliotides possess sequences and structural elements of cyclotides. They maintain membrane-active properties of cyclotides despite heat treatment, and thus may constitute the putative active principles in *C. ternatea*. More significantly, we show, also for the first time, that their precursors are chimeras, half from cyclotide and the other half from A1a and A1b (leginsulin), with cyclotide domain displacing the A1b domain in the precursor. Their chimeric structures provide a likely evidence of horizontal gene transfer or convergent evolution of plant nuclear genomes. Collectively, my findings shed new understanding about the biosynthetic mechanism of cyclotides in the Fabaceae and their evolution in plants.

Chapter 8

Fuzzy Splicing as Novel Mechanism to Generate Chemical Diversity in Plant Defense

1. Introduction

Plants are photoautotrophs capable of capturing sun energy to convert carbon dioxide into organic compounds. They are primary food source for myriads of bacteria, fungi, invertebrates, vertebrates and even other plants. Since they cannot move to escape these multitudes of enemies, plants have developed sophisticated and remarkable diversity of secondary metabolites to ward off attackers allowing them to survive and thrive on the earth. High chemical diversity is important for plants survival as they are constantly attacked by multiple pathogens and different molecular types may cover gaps in each-others activity spectrum or act synergistically [159].

Much understanding of the molecular mechanism to generate chemical diversity in plant defense comes from the biosynthesis of low-mass compounds such as alkaloids, terpenoids, polyketides, and nonribosomal peptides. The majority of these secondary metabolites are non-gene encoded and produced from general building blocks via complex biosynthesis machinery involving multiple processing enzymes. Often, these secondary metabolic pathways favor molecular diversity where one pathway produces multiple products, as many as 100 in some pathways [160]. The biosynthesis promiscuity thus usually leads to a spectrum of diverse minor products.

For gene-encoded CRPs such as cyclotides and defensins, the molecular strategies for generating of their diversity are mostly limited to one mature domain for one mature peptide [37]. It might be more appropriate to describe it as 'one-to-one' model. Under this model, plants increase the CRPs diversity by having many genes or multi-domain precursors through duplication and divergence. CRP genes such as cyclotides and defensins usually encode for one to three mature domains and each domain leads to formation of a single mature peptides [37].

In this chapter, I report a new phenomenon, 'one-to-many', describing how plants expand the proteome beyond the DNA coding capacity by employing one domain to generate many products different in forms and sequences. The proteome diversification is achieved by a novel post translational modification pathway that occurs through a series of promiscuous proteolytic processings and ligation reactions, a process termed 'fuzzy splicing'. I discovered that a single cyclotide encoding domain may yield up to five mature products comprising of a mixture of both cyclic and linear forms with variation in primary sequences. The 'fuzzy splicing' process is probably controlled by the N- and C-terminal processing enzymes of the mature cyclotide domain. This process enables an economic usage of genetic materials and allows the expansion of cyclotides and uncyclotides inventory in plants.

2. Results and Discussion

2.1 Identification of cyclotide and uncyclotide clusters with common core sequences

Sequence analysis revealed several clusters of cyclotides and uncyclotides sharing common core sequences but differing at the N/D-G ligation junction (for cyclotides), or at the N- and C-terminal ends (for uncyclotides) (Table 16). Each cluster contains a nested set of 2-5 unique peptides distinguished from each other by either forms (cyclic or linear) or primary sequences (extensions or truncations of one or few residues from the common core sequences). Five such nested peptide sets are found in *H. biflora*, four in *C. ternatea,* two in *H. diffusa* and one in *C. chartacea*. Sequence comparisons with their respective precursors suggest that nested peptides in each cluster are derived from a same precursor protein.

The first peptide listed in each set is the main product and denoted as 100% (Table 16). Relative molar percentage of other peptides compared to the first one was calculated based on their relative area on the MS spectra. Occurrence of these peptide clusters was reproducible and biologically relevant as extractions of fresh plant samples were performed immediately after collection with the addition of a cocktail of protease inhibitors to block proteolytic degradations.

Table 16. Fuzzy Splicing in Plants

Peptide	Sequence	MW	Form	Family	RI(%)
Clitoria ternatea					
cT2	GEFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3260	C	M	100
cT33	GEFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3278	L	M	<5
cT34	EFLKCGESCVQGEC-YT--PGCSCDWPICKKN	3221	L	M	<5
cT35	FLKCGESCVQGEC-YT--PGCSCDWPICKKN	3092	L	M	<5
cT36	LKCGESCVQGEC-YT--PGCSCDWPICKKN	2945	L	M	<5
cT19	GSVIKCGESCLLGKC-YT--PGCTCSRPICKKN	3124	C	M	100
cT20	SVIKCGESCLLGKC-YT--PGCTCSRPICKKN	3067	C	M	<5
cT21	GSVIKCGESCLLGKC-YT--PGCTCSRPICKKD	3125	C	M	20
cT22	GSAIRCGESCLLGKC-YT--PGCTCDRPICKKN	3152	C	M	100
cT23	SAIRCGESCLLGKC-YT--PGCTCDRPICKKN	3095	C	M	<5
cT26	GSAIRCGESCLLGKC-YT--PGCTCDRPICKKD	3153	C	M	20
cT24	AIRCGESCLLGKC-YT--PGCTCDRPICKKN	3026	L	M	7
cT25	IRCGESCLLGKC-YT--PGCTCDRPICKKN	2955	L	M	<5
cT28	GSVIGCGETCLRGRC-YT--PGCTCDHGICKKN	3107	C	M	100
cT29	VIGCGETCLRGRC-YT--PGCTCDHGICKKN	2981	L	M	-
Hedyotis biflora					
B1	GTRCGETCFVLPC-WSAKFGCYCQKGFCYRN	3403	C	B	100
B33	GTRCGETCFVLPC-WSAKFGCYCQKGFCYRN	3421	L	B	<5
B34	GTRCGETCFVLPC-WSAKFGCYCQKGFCYRNE	3550	L	B	<5
B35	TRCGETCFVLPC-WSAKFGCYCQKGFCYRN	3364	L	B	<5
B36	TRCGETCFVLPC-WSAKFGCYCQKGFCYRNE	3493	L	B	<5
B4*	IPCGESCAFIPC-LTSLLGCTCQNKVCYRDE	3260	L	B	100
B13	GIPCGESCAFIPC-LTSLLGCTCQNKVCYRD	3170	C	B	49
B15*	GIPCGESCAFIPC-LTSLLGCTCQNKVCYRDE	3317	L	B	77

B5	GGAVPCGETCVYIPC-ITAAIGCSCQNNVCYHN	3236	C	B	100
B9	GAVPCGETCVYIPC-ITAAIGCSCQNNVCYHN	3179	C	B	8
B16	AVPCGETCVYIPC-ITAAIGCSCQNNVCYHN	3140	L	B	17
B12	GGVPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	3440	C	B	100
B14	GVPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	3383	C	B	8
B29	VPCGESCVWIPC-ISSVFGCTCQNSDKACYHN	3344	L	B	3
B10*	REECGETCYILPC-VT--PDCICSGGQCYKIE	3206	L	H	100
B11*	EECGETCYILPC-VT--PDCICSGGQCYKIE	3050	L	H	13
B31*	pEECGETCYILPC-VT--PDCICSGGQCYKIE	3032	L	H	<5
B32*	ECGETCYILPC-VT--PDCICSGGQCYKIE	2921	L	H	<5
B17	GNPCGESCVYIPC-ITTVVGCSCQNSVCYHN	3126	C	B	100
B30	NPCGESCVYIPC-ITTVVGCSCQNSVCYHN	3087	L	B	<5
Chassalia chartacea					
cC10	GEYCGESCYLIPC-FT--PGCYCVSRQCVNKN	3212	C	H	-
cC13	EYCGESCYLIPC-FT--PGCYCVSRQCVNKN	3155	C	H	-
Hedyotis diffusa					
hD8	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKSVS	3555	L	B	100
hD7	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKSV	3468	L	B	76
hD6	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKS	3369	L	B	54
hD9	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKI	3581	L	B	-
hD10	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKIV	3494	L	B	-
hD11	VG-CYERCVWGPC-ISKIVGCSCDTTTYDCVKIVS	3395	L	B	-

[a]Average isotopic molecular weight. Cys residues that form the cystine knot are highlighted in yellow, Pro residues that form the Möbius structure in gray, negatively charged residues in red and positively charged residues in blue.

2.2 Fuzzy splicing as a mechanistic explanation for nested formation of cyclotides and uncyclotides

Since the discovery of the first cyclotide-encoding gene, it has been evident that biosynthesis of cyclotides requires two precise cleavages at the N- and C-termini followed by a ligation reaction [54]. A single cyclic product is formed as the result. This model apparently fulfilled the formation of all cyclotides hitherto discovered. However, it fails to explain the formation of multiple cyclic and linear forms from a single cyclotide domain observed in my work. The biosynthetic processing of cyclotides may be thus more complicated than previously thought.

To provide a mechanistic explanation for this new phenomenon, I propose here a novel post translational pathway termed “fuzzy splicing” to describe a “fuzzy excision” of a cyclotide domain followed by a “fuzzy ligation” of the newly formed N- and C-termini. The fuzzy processings of the cyclotide precursors can occur at three different stages: N-terminal cleavage, C-cleavage and backbone cyclization. This novel pathway allows generation of many products, not just one from a single cyclotide domain and thus greatly enhances the molecular diversity of cyclotides in plant defense.

2.3 Fuzzy splicing at the C-terminal processing

Nearly all cyclotides in the database contain a highly conserved Asn/Asp residue in loop 6 which acts as the recognition site for AEP. Based on sequence comparison of the cyclotide precursors, this processing enzyme has specific requirements for Asn/Asp residue at the S_1 position and Leu (Rubiaceae/Violaceae) or Ile/Val (Fabaceae) residue at the S_2’ position.

Residues at the S1' position tend to be variable between species but remain relatively conserved within species. S1' position usually adopts small side chain residues (Ser, Gly, Ala) in the Violaceae, Gly or Glu in the Rubiaceae and His in the Fabaceae. Residues at the S3' position are variable and appear to be non-essential for enzyme recognition as suggested from the biosynthesis of chassatides discussed in chapter 4.

The conventional C-terminal processing of cyclotides cleaves the precursor proteins precisely at the Asn/Asp residue followed by a backbone cyclization. Interestingly, I found that the precursor proteins are also processed at the S1' position if this site is occupied by Glu residue i.e. hedyotide B1 and B10 precursors. The structure similarity between Glu and Asn/Asp residues probably allows its promiscuous recognition by AEP. Recent evidence suggests that enzyme promiscuity may be more widespread than previously recognized and this process is related to the evolvability of proteins and biosynthesis of secondary metabolites [161]. Alternatively, cleavage at the C-terminal side of Glu may be catalyzed by a glutamyl-specific endopeptidase (GEP). Hydrolysis of glutamyl bond forms an intermediate that is unable to cyclized and led to uncyclotide formation. This allows generation of forms and sequences diversity, both cyclic and linear structures from a single cyclotide domain by processing at the C-terminal Asn/Asp or Glu, respectively.

The ratio of cyclotide and uncyclotide provides an additional information about the kinetic of enzyme catalysis. There is a significant difference of cyclotide and uncyclotide compositions among peptide clusters dependent on whether Asn or Asp located at the C-terminal processing site. If it is Asn residue, AEP and/or GEP will preferably cleave after its carboxylic side generating

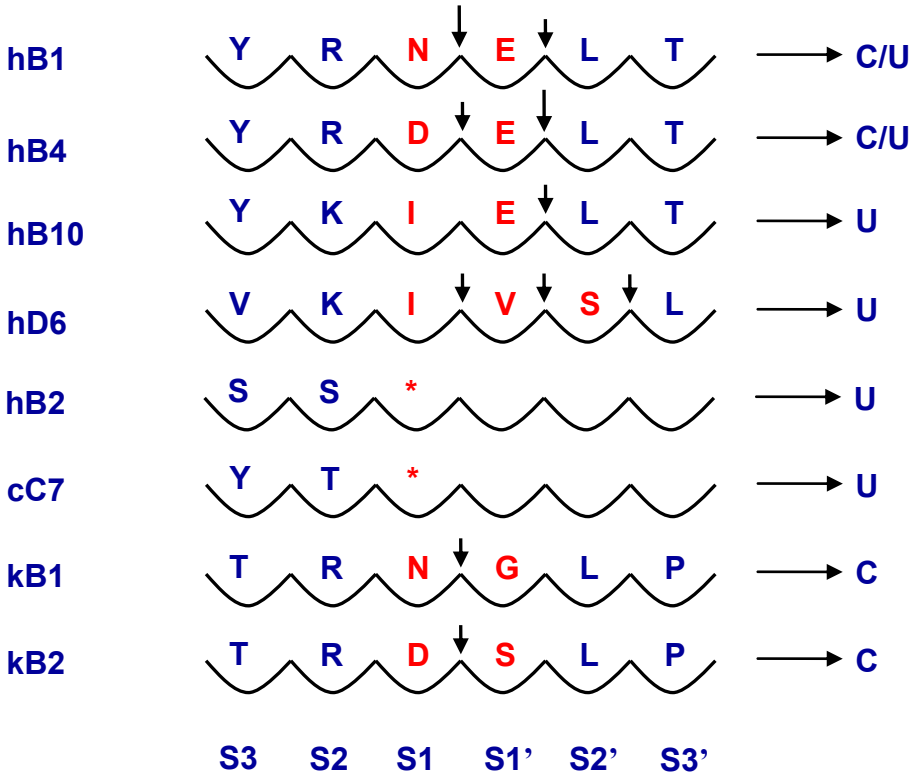


Figure 65. Summary of C-terminal cleavage site in hedyotide B1, B10, B4/13/15, kalata B2 and kalata B6. Vertical arrows indicate the scissile peptide bond. Horizontal arrows indicate the formation of cyclotides (C) or uncyclotide (U). S3 to S3' indicate the enzyme active site.

cyclotides as the main products, and uncyclotides forming by cleaving after the Glu residue as the minor products. The processing of hedyotide B1 precursor, for example, generates one dominant cyclic form (cutting after Asn), and four minor linear forms (cutting after Glu). Interestingly, if Asp residue locates at the C-terminal processing site, AEP and/or GEP will cleave both Asp and Glu at similar rates with a slight preference for Glu indicated by a higher percentage of uncyclotide (hedyotide B4 and B15) than cyclotide (hedyotide B13) formation. AEP has been shown to hydrolyze the asparaginyl bond 100-fold faster than the aspartyl bond. And in this instance, hydrolysis of Glu appears to be slightly faster than Asp but significantly slower than asparaginyl bond hydrolysis. The order of hydrolysis rate are thus as follow N>>E>D.

2.4 What happen if C-terminal Asn/Asp is absent due to mutations?

Two scenarios, nonsense or missense mutations, cause the loss of C-terminal Asn/Asp residue effectively preventing the backbone cyclization. As a result, only uncyclotides are formed. The first scenario has been explained in detail for the formation of uncyclotides such as hedyotide B2 and chassatide C7 in the earlier chapters. Briefly, a premature stop codon caused by a nonsense mutation inhibits the translation of Asn/Asp residue. Since it is a stop codon, the C-terminal processing and ligation step are bypassed and only uncyclotides are formed.

In the second scenario, missense mutations cause substitution of the C-terminal Asn/Asp for other amino acids. By far, AEP is the only known enzyme that participates in the post translational modification of cyclotide precursors at the C-terminus. This raises an interesting question how the precursor protein is

processed in the absence of this highly conserved Asn residue. My results showed that other alternative pathways might have taken place in the absence of the conserved Asn residue.

The C-terminal processing in this instance is determined by the adjacent residue at the S1' position. If the Glu residue occupies this position such as hedyotide B10 and B11, the precursor protein will be processed after its C-terminal side by AEP and/or GEP. Interestingly, if S1' position is not Glu but other amino acids, a series of linear species with the successive addition of residues constituting the CTPP are observed. This is exemplified by the formation of hedyotide D6-8 and hedyotide D9-11. Their precursor proteins are unlikely to be processed by AEP or GEP but likely by a carboxypeptidase that progressively trims the C-terminal end.

2.5 Fuzzy splicing at the N-terminus

The N-terminal processing of cyclotides is still not well understood. In the Rubiaceae and Violaceae, the enzyme catalyzing this process appears to have a specific requirement for Leu/Val at the S2 position while residues at the S1 position are not important. In the Fabaceae, this process is likely to be carried out by signal peptidases. A single cyclotide is believed to be formed as the result. Careful examination of the low-abundant products revealed that several different cyclotides and uncyclotides generated from a same precursor protein are likely attributed by an aberrant processing at the N-terminus. It is noteworthy that the formation of uncyclotides in this case is not genetic predetermined owing to the introduction of a stop codon but instead by the bioprocessing.

Isolated peptides in each cluster show variation in length and are truncated from the N-terminal side of the ligation junction suggesting that an editing event must happen after the N-terminal cleavage. The editing enzymes probably trim the N-terminal extensions of the linear precursors produced after the N-terminal cleavage, forming a series of linear intermediates which subsequently undergo C-terminal processing and backbone ligation. The resulting products can be either cyclotides or uncyclotides dependent on which amino acids located at the N-terminus. Residues that allow backbone cyclization are called permissive residues whereas those that do not are called non-permissive residues (given that Asp/Asn located at the C-terminus).

The identity of a permissive or non-permissive residue was established mainly based on conservation pattern. Most cyclotides contain a highly conserved N-terminal Gly (>90% of cyclotide sequences) and occasionally Ser, Asp and His residues. These are permissive residues since they allow cyclotide formation. For non-permissive residues, their identities were mainly based on sequences discovered in my work since only two known naturally occurring uncyclotides has been published and their formations are both due to the introduction of a stop codon. Sequence analysis of uncyclotides shows that Ala, Asn, Thr, Val, Phe, Leu and Ile are non-permissive residues. Intriguingly, Glu is a permissive residue in the Rubiaceae but non-permissive in the Fabaceae.

What are the likely factors that determine if an amino acid is permissive or non-permissive? Based on current understanding of cyclotide biosynthesis, the N-terminal residue must be able to accommodate the S1' subsite of the AEP enzyme to enable the backbone cyclization. Small side-chain amino acids such as Gly and Ser are preferable at this position based on their high conservation at

the ligation junction. Interestingly, it appears that there is a slight difference in specificity at the S1' subsite of AEPs among different plant families deduced from the residues at the C-terminal processing site. In the Rubiaceae, this site often occupied by Glu, Gly or Ser, whereas in the Violaceae it is occupied by Ser, Gly or Ala and in the Fabaceae by His. This explains why Glu is permissive residue in the Rubiaceae but non-permissive in the Fabaceae.

2.6 Fuzzy ligation

Backbone ligation is the final step in the biosynthesis of cyclotides. This reaction presumably happens right after the proteolytic cleavage of the C-terminal side of the highly conserved Asn/Asp residue which led to the formation of an acyl-enzyme intermediate. This intermediate is attacked by the nucleophilic amino group of the N-terminal residue to form new peptide bond to complete the cyclization. This ligation reaction is thought to be highly efficient and produced only cyclotide as a single product. Evidence from my work suggests that this reaction may be more promiscuous than previously thought.

A number of uncyclotides such as hedyotide B33 and cliotide T33 containing both ligation permissible residues, an N-terminal Gly and a C-terminal Asn, were indentified suggesting that the cyclization process is leaky and a small percentage of the acyl-enzyme intermediate is hydrolyzed by water. The deacylation reaction thus can be considered as a competing process between a hydrolysis reaction by water to form uncyclotide and an aminolysis reaction by the amino group of the N-terminal residue to form cyclotide (Fig. 24). This allows generation of two distinct peptides with identical primary sequences but different in forms, cyclic and linear.

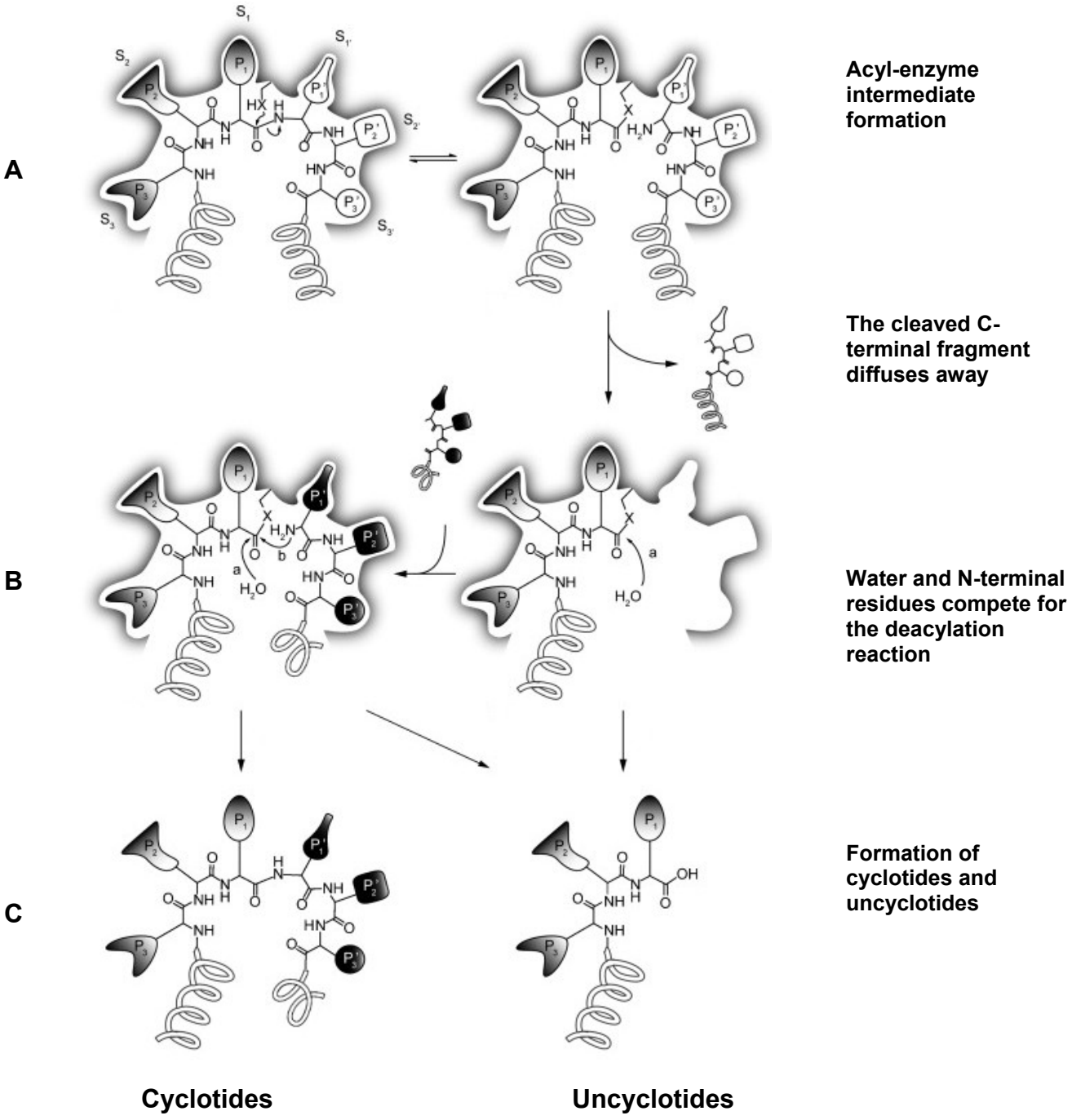


Figure 66. Proposed mechanism for cyclotides and uncyclotides formation. Figure is adapted from Celia et al (30). (A) Formation of acyl-enzyme intermediate between AEP and C-terminal processing sites (i.e YRDELTP for HB4). AEP can recognize either Glu or Asn/Asp residue at P_1 position. (B) The cleaved C-terminal fragment diffuses away. Water and amino group of N-terminal residues compete for the deacylation reaction. (C) If P1 = Asn/Asp cyclotides formation are favored, whereas if P1 = Glu uncyclotides formation are favored.

2.7 Fuzzy splicing is a bug or a feature?

Generation of multiple products from a single precursor may imply that the biosynthesis of cyclotides is intrinsically sloppy and unable to produce a single product. However, the high fidelity of most primary metabolite synthesis pathways such as cholesterol and amino acids suggests that plants do have the capacity for high accuracy synthesis. Numerous results now indicate that most secondary metabolite synthetic pathway are promiscuous and usually make many products. The biosynthesis of gibberellin for example produces at least 136 distinct chemical entities and only a few have known biological activities [160]. Fuzzy splicing is a novel mechanism demonstrated for the first time that how plants generate chemical diversity for gene-encoded CRPs such as cyclotides in addition to the well-understanding biosynthetic pathway of many small-molecule secondary metabolites.

High chemical diversity is often advantageous for plant defense but it usually incur costs. Evolution would thus favor organisms that produce high diversity with minimizing cost. Fuzzy splicing allows rapid expansion of diversity without additional cost by hijacking the available bioprocessing pathway to form a cluster of new products from a single precursor. Furthermore, all the minor products could possess potent bioactivities or cover gap in each other activity spectrum or act synergistically.

Cyclotide biosynthesis liked many small-molecule secondary metabolic pathways often produces only one or a few dominant products and most other compounds in trace amount. The concentration of the major and minor products could be differed by a few orders of magnitude. Interestingly, mutations at the

ligation junction could alter the enzyme-substrate affinities and lead to a new quantitative profile. The major products could become the minor products and vice versa. This is exemplified by the substitution of the C-terminal Asn for Asp or other residues. When Asn locates at the C-terminal position, cyclotides is the major component and uncyclotide is the minor component. When Asp locates at the C-terminal position (i.e. hedyotide B4/13/15), the profile is shifted with uncyclotide becomes the major and cyclotide become the minor component. When Asn is mutated to other residues (i.e. hedyotide B10/11), the biosynthetic pathway produces solely uncyclotide without the cyclic counterpart. Such shifts in cyclotide and uncyclotide profiles could result in marked change in biological activity. Fuzzy splicing thus allows rapid evolution of diversity because any mutations at the ligation junction potentially result in a new quantitative profile.

In summary, the biosynthetic processing of cyclotides is fuzzy and capable of generating many products, not just one as previously thought. This greatly increases the molecular diversity of chemical warfare in plant defense. Being unable to run away from invaders and having to face against a wide range of pests and pathogens, a high chemical diversity is essential for plant survival and evolution.

Chapter 9

Summary, Conclusions and Future Outlook

With a focus interest in biologics, this study examines the diversity of CRPs in herbal medicine. My study reports the discovery of over 120 novel CRPs and provides new insights about their diversity and evolution in plants. For this purpose, chapter 3 describes the development of new sequencing technique to facilitate the rapid discovery of primary structures and disulfide connectivities of CRPs in plants. By employing tryptic digestion, the amino acid sequences of a full length cyclotide can be obtained in a single MS/MS spectrum. A novel approach was also developed for rapid disulfide characterization using top-down MS/MS. This technique surpasses the conventional bottom-up strategy by omitting the proteolytic digestion step and thus greatly speeds up the disulfide characterization process.

Chapter 3 also describes the discovery of cyclotides and uncyclotides in various *Hedyotis* species. They display hypervariable primary structures with no individual residues being absolutely conserved except the cystine scaffold. They also display tissue-specificity with entirely different profiles expressed in different plant tissues. Interestingly, geographic investigation of cyclotide compositions in *H. biflora* demonstrated a striking variation of plant samples collected in Vietnam and Singapore. This supports the traditional concept of *"dao di yao cai"* of Chinese medicine (authentic region herbal materials), which believes that only herbs grow in a specific region displayed the desired medicinal values. Different soil conditions and environmental factors may trigger the expression of different set of CRPs led to variations in the medicinal profiles.

I have further shown that the landscape of cyclotides and uncyclotides varies greatly among the *Hedyotis* species. *O. affinis* has the highest expression level of all known cyclotide-producing plants and is approximately 12,000 fold higher than the expression level in *H. diffusa.* In addition, it produces mostly cyclotides while only uncyclotides are produced in *H. diffusa.* Another major finding in this chapter is the discovery of a large number of uncyclotides with 11 sequences in *H. diffusa* and 15 sequences in *H. biflora*. This finding contributes significantly to the uncyclotide database with only two naturally occurring uncyclotides reported violacin A and psyle C. Finally, chapter 3 highlights the difference in the biodegradation pathway of cyclotides and uncyclotides. Both are susceptible to air oxidation of Trp residue. However, only uncyclotides are prone to exopeptidase degradation whereas cyclotides being cyclic are protected against proteolysis.

Chapter 4 reports the discovery of 21 chassatides from *Chassalia chartacea*, which comprises of 14 novel cyclotides and 4 novel uncyclotides, two Met-oxidized derivative and circulin A. This is the most in-depth analysis of cyclotides content in a *Chassalia* species. Bioassays of novel chassatides show unexpectedly that uncyclotides are the most active of all tested chassatides with stronger cytotoxic and hemolytic effects than their cyclic counterparts. This is contrary to the previous belief that the cyclic backbone is essential for activities and ring-opening would cause loss of cyclotide functions. My finding suggests that the cyclic backbone may provide additional stability advantages but not an indispensible functional requirement. Another finding in this work is the effect of Met-oxidation on cyclotides and uncyclotides bioactivities. Oxidation of Met disrupts the hydrophobic patch on loop 3, which causes total loss of activity

suggesting the importance of this loop for membrane interaction. Finally, genetic characterization of novel chassatides unveils several interesting characters of their precursor proteins. They have the shortest precursors of all known cyclotides. The reduction in length is mostly attributed to the absence of the NTR domain. This enables the economic biosynthesis of cyclotides which is beneficial for understory plants like *Chassalia chartacea.*

Chapter 5 provides the first description of morintides from the *Morinda citrifolia*. Morintides are assigned to a new class of CRPs due to their unusually high anionicity (net charge -1 to -7) in contrast to the basic nature of most plant CRPs. Disulfide mapping showed that morintides adopt a knotted configuration and thus belong to the knottin family. Morintide C1-3 are stable to heat and enzymes whereas morintide C4 and C5 are not, probably attributed to the presence of EP and NG bond in their primary sequences. Interestingly, *M. citrifolia* although belongs to the Rubiaceae family, it produces morintides instead of cyclotides. These two classes of CRPs do share several structural similarities such as similar size, knottin scaffold and hydrophobicity. Their biological activities, however, are entirely different. Unlike cyclotides, morintides are devoid of membranolytic properties displaying no antimicrobial, cytotoxicity or hemolysis. The presence of multiple Asp and Glu in their sequences suggests that morintides may bind to divalent ion such as Zn^{2+}, Cu^{2+} and Mn^{2+} which play important role in plant metabolism. These findings suggest that multiple classes of CRPs may have evolved independently within the Rubiaceae family.

Chapter 6 investigates the CRP contents in ginseng, the most well-known medicinal plant in Chinese medicine. A novel class of CRP, termed ginsentide, with unique disulfide pattern resembling a pretzel knot was characterized. Four

novel ginsentides were isolated from Korean ginseng and one from notoginseng. They are specifically expressed in flowers and roots but not in leaves. Ginsentides display a highly compact structure with four disulfide bridges as compared to the usual three found in similar size CRPs. They display high stability to heat and enzymatic digestion and thus likely contribute to the active principles in ginseng. Cell-based testings of ginsentide functions show that they display no membranolytic activities. Their primary function in plants has been shown to be related to water-stress responses. Considering the high abundance of ginsentides and the long history of medicinal usage of ginseng, future investigation of ginsentides for therapeutic applications should be of great interest.

Chapter 7 reports the discovery of cyclotides as heat-stable CRPs in *Clitoria ternatea* and expands the distribution of cyclotides to a new plant family, the Fabaceae. A panel of 36 novel cyclotides and uncyclotides (cliotide T1-T34) was isolated. Cliotides were shown unambiguously to be cyclotides and not A1bs, as determined by their sequence homology, disulfide connectivity, and membrane-active properties indicated by their antimicrobial activity against *E. coli* and cytotoxicity to HeLa cells. I also show that cliotides are prevalent in *C. ternatea* and are found in every plant tissues examined, including flowers, seeds, stems, nodules and roots. Genetic characterization of novel cyclotides reveals unexpectedly that they are originated from a chimeric precursor consisting of cyclotide and leginsulin domains. Their chimeric structures are likely originated from either horizontal gene transfer or convergent evolution in plant nuclear genomes which are exceedingly rare events. Such atypical genetic arrangement

also implies a different mechanism of biosynthetic processing of cyclotides in the Fabaceae and provides new understanding about their evolution in plants.

Chapter 8 reported a novel post translation pathway termed “fuzzy splicing” describing how plants generate multiple products different in forms and sequences from a single cyclotide encoding domain. This is achieved through a series of promiscuous proteolytic processings and ligation reactions of the cyclotide precursors. The fuzzy splicing process is probably controlled by the N- and C-terminal processing enzymes of the mature cyclotide domain. This process enables the economic usage of genetic materials and allows the expansion of cyclotides and uncyclotides inventory in plants. Fuzzy splicing therefore helps to increase the molecular diversity and provide a better adaptation essential for plant survival.

In conclusion, this thesis provides a pilot study exploring the biologics in herbal medicine with the discovery of over 120 novel CRPs. They represent a rich pharmacological goldmine that has yet to receive much attention in the past drug-screening programs. These novel CRPs display highly diverse primary sequences and biological functions. They also exhibit high stability against heat and enzymatic degradations. These advantages warrant future study to investigate their pharmacological profiles for therapeutic applications.

Publications

1. Giang K. T. Nguyen, Sen Zhang, Ngan T. K. Nguyen, Phuong Q. T. Nguyen, Ming S. Chiu, Antony Hardjojo and James P. Tam. Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Cyclotide and A1a Domains in the Fabaceae family. (Submitted to *J. Biol. Chem.*)

2. Giang K. T. Nguyen, Weihan Lim, Phuong Q. T. Nguyen and James P. Tam. Novel cyclotides from *Chassalia chartacea*: genetic characterizations and effects of Methionine oxidation on cyclotide bioactivities. (In preparation)

3. Giang K. T. Nguyen, Weijian Ye, Phuong Q. T. Nguyen and James P. Tam. An unusual highly anionicity of cystine-knot miniproteins from *Morinda citrifolia*. (In preparation)

4. Giang K. T. Nguyen, Shining Loo and James P. Tam. Discovery and Characterization of ginsentide, a novel type of disulfide-rich miniproteins from *Panax ginseng, P. notoginseng* and *P. quinquefolius*. (In preparation)

6. Giang K. T. Nguyen, Sen Zhang, Wei Wang, Clarence T. T. Wong, Ngan T. K. Nguyen and James P. Tam. Discovery of novel cyclotides and uncyclotides in *Hedyotis biflora*: Top-down Disulfide Mapping of Hedyotide B2 (In preparation).

7. Giang K. T. Nguyen, Phuong Q. T. Nguyen and James P. Tam. Novel uncyclotide in *Hedyotis diffusa* (In preparation).

Poster Presentation

5th international peptide symposium in conjuction with 47th Japanese Peptide Symposium, December 4-9, 2010. Title: “Rapid Approach for de novo Sequencing of Cyclotides”

Invited Talk

4th Korea-Singapore International Conference on Bioscience & Biotechnology (KSICBB2009). Seoul, Korea (December 2009). Title: "Discovery of Novel Cyclotides and Uncyclotides in *Hedyotis Biflora"*

Appendix A

Table A1. List of additional CRPs discovered in multiple medicinal plants

Peptide	Primary Sequence	MW
Alstonia scholaris **(Apocynaceae)**		
Alstotide S1	CRPYGYRCDGVINQCCDPYHCTPPLIGICL	3366
Alstotide S2	CRPYGYRCDGVINQCCDPYRCTPPLIGICL	3385
Alstotide S3	CVPRFGRCDGIINQCCDPYLCTPPLVGICT	3250
Allamanda carthatica **(Apocynaceae)**		
Allatide C1	CIAHYGKCDGIINQCCDPWLCTPPIIGICL	3256
Allatide C2	CRPVGTRCDGVINQCCDPYWCTPPIYGWCK	3426
Allatide C3	CRPVGTRCDGVINQCCDPYWCTPPIYGWCK	3490
Allatide C4	CIAHYGKCDGIINQCCDPWLCTPPIIGFCL	3290
Allatide C5	CVSHYGKCDGIINQCCDPWLCTPPIIGFCL	3292
Alternanthera sessilis **(Amaranthaceae)**		
Altertide S1	APGECKHGRCPPGICCSQYGYCGTGPAYCG	3029
Altertide S2	APGECKHGRCPPGICCSQYGYCGTGPAYC	2972
Achyranthis bidentatae **(Amaranthaceae)**		
Achytide B1	NCESGTSCIPGAQHNCCSGVCVPIVTIFYGVCY	3418
Lasianthus tomentosus **(Rubiaceae)**		
Lasiantide T1	pQACLGHVCETDQYCQNLGCYRCVTSGSAKICI	3443

Talinum Triangulare **(*Portulacaceae*)**

Talitide T1	SCVEADGSCGPLFPCCSGITCSFDIYQFRFKCW	3663
Talitide T2	DSCVEANGSCGPLFPCCAGLSCTFALSEFRFKCW	3642
Talitide T3	ADDNCVNEVGGICNISERINCCPGIYCDSFGIGHNGVCVIPVS	4450

Viola yedoensis **(Violaceae)**

Viotide Y1	GLPVCGETCFGGTCNTPGCICEWPVCTRN	2993

Alternanthera paronychioides **(Amaranthaceae)**

Altertide P1	CGRPGVTCGFSNPSTVCCPPCVCDFTFDADVCFGSC	3686

Borreria alata **(Rubiaceae)**

Boritide A1	CFSDYDCSALNCVCKFFEWHYGGCV	-

Hedyotis sp **(Rubiaceae)**

	VGCYERCVWAPCISKIVGCSCDTTTYDCVKS	3383

Appendix B

Alignment of Cyclotide Sequences

Cyclotide	Sequence
Möbius	
cycloviolacin B6	GLPV-CGETCVGGTCNTPGCGCS--WPVCTRN
cycloviolacin B7	GLPV-CGETCVGGTCNTPGCACS--WPVCTRN
cycloviolacin B17	GLPI-CGETCTLGTCYTVGCTCS--WPICTRN
cycloviolacin H3	GLPV-CGETCFGGTCNTPGCICDP-WPVCTRN
cycloviolacin O14	GSIPACGESCFKGKCYTPGCSCSK-YPLCAKN
cycloviolacin O15	GLVP-CGETCFTGKCYTPGCSCS--YPICKKN
cycloviolacin O16	GL-P-CGETCFTGKCYTPGCSCS--YPICKKIN
cycloviolacin O21	GLPV-CGETCVTGSCYTPGCTCS--WPVCTRN
cycloviolacin O22	GLPI-CGETCVGGTCNTPGCTCS--WPVCTRN
cycloviolacin O23	GLPT-CGETCFGGTCNTPGCTCDSSWPICTHN
cycloviolacin O24	GLPT-CGETCFGGTCNTPGCTCDP-WPVCTHN
hymo B	CGETCVTGTCYTPGCACD--WPVCKRD
kalata B1	GLPV-CGETCVGGTCNTPGCTCS--WPVCTRN
kalata B2	GLPV-CGETCFGGTCNTPGCSCT--WPICTRD
kalata B3	GLPV-CGETCFGGTCNTPGCTCDP-W-ICTRD
kalata B4	GLPV-CGETCVGGTCNTPGCTCS--WPVCTRD
kalata B6	GLPV-CGETCFGGTCNTPGCSCSS-WPICTRN
kalata B7	GLPV-CGETCTLGTCYTQGCTCS--WPICKRN
kalata B10	GLPT-CGETCFGGTCNTPGCSCSS-WPICTRD
kalata B11	GLPV-CGETCFGGTCNTPGCSCT--DPICTRD
kalata B12	GS-L-CGDTCFVLGCNDSSCSCN--YPICVKD
kalata B13	GLPV-CGETCFGGTCNTPGCACDP-WPVCTRD
kalata B14	GLPV-CGESCFGGTCNTPGCACDP-WPVCTRD
kalata B15	GLPV-CGESCFGGSCYTPGCSCT--WPICTRD
mram 11	GHPT-CGETCLLGTCYTPGCTCK--RPVCYKN
mram 12	GSAILCGESCTLGECYTPGCTCS--WPICTKN
mram 13	GHPI-CGETCVGNKCYTPGCTCT--WPVCYRN
varv A	GLPV-CGETCVGGTCNTPGCSCS--WPVCTRN
varv B	GLPV-CGETCFGGTCNTPGCSCDP-WPMCSRN
varv C	GVPI-CGETCVGGTCNTPGCSCS--WPVCTRN
varv D	GLPI-CGETCVGGSCNTPGCSCS--WPVCTRN
varv E	GLPI-CGETCVGGTCNTPGCSCS--WPVCTRN
varv F	GVPI-CGETCTLGTCYTAGCSCS--WPVCTRN
varv G	GVPV-CGETCFGGTCNTPGCSCDP-WPVCSRN
varv H	GLPV-CGETCFGGTCNTPGCSCET-WPVCSRN
vhl-2	GLPV-CGETCFTGTCYTNGCTCDP-WPVCTRN
vibi A	GLPV-CGETCFGGTCNTPGCSCS--YPICTRN
vibi B	GLPV-CGETCFGGTCNTPGCTCS--YPICTRN
vibi C	GLPV-CGETCAFGSCYTPGCSCS--WPVCTRN
vibi D	GLPV-CGETCFGGRCNTPGCTCS--YPICTRN
violacin A (linear)	SAIS-CGETCFKFKCYTPRCSCS--YPVC-K-
violapeptide 1	GVPV-CGETCVGGTCNTPGCSCS--RPVCTXN
vodo M	GAPI-CGESCFTGKCYTVQCSCS--WPVCTRN
vodo N	GLPV-CGETCTLGKCYTAGCSCS--WPVCYRN

Bracelet

```
cyclopsychotride A      SIP--CGESCVFIP-CTVTALLGCSC--KSKVCYKN
circulin A              GIP--CGESCVWIP-C-ISAALGCSC--KNKVCYRN
circulin B              GVIP-CGESCVFIP-C-ISTLLGCSC--KNKVCYRN
circulin C              GIP--CGESCVFIP-C-ITSVAGCSC--KSKVCYRN
circulin D              KIP--CGESCVWIP-C-VTSIFNCKC--ENKVCYHD
circulin E              KIP--CGESCVWIP-C-LTSVFNCKC--ENKVCYHD
circulin F              AIP--CGESCVWIP-C-ISAAIGCSC--KNKVCYR
CD-1                    GADGFCGESCVYIP-C-ISYLVGCSCDTIEKVCKRN
cycloviolacin B1        GIP--CGESCVYLP-C-FTAPLGCSC--SSKVCYRN
cycloviolacin B2        GIP--CGESCVWIP-C-LTATIGCSC--KSKVCYRN
cycloviolacin B3        GIP--CAESCVYLP-C-VTIVIGCSC--KDKVCY-N
cycloviolacin B4        GIP--CAESCVWIP-CTVTALLGCSC--KDKVCY-N
cycloviolacin B8        GIP--CGEGCVYLP-C-FTAPLGCSC--SSKVCYRN
cycloviolacin B9        GIP--CGESCVWIP-C-LTAAIGCSC--SSKVCYRN
cycloviolacin B10       GVP--CGESCVWIP-C-LTSAIGCSC--KSSVCYRN
cycloviolacin B11       GIP--CGESCVLIP-C-ISSVIGCSC--KSKVCYRN
cycloviolacin B12       GVIP-CGESCVFIP-C-ISSVIGCSC--KSKVCYRN
cycloviolacin B13       GAG--CIETCYTFP-C-ISEMINCSC--KNSRCQKN
cycloviolacin B14       GIP--CGESCVWIP-C-ISSAIGCSC--KNKVCYRK
cycloviolacin B15       GIP--CGESCVWIP-C-ISGAIGCSC--KSKVCYRN
cycloviolacin B16         TIPCAESCVWIP-CTVTALLGCSC--KDKVCY-N
cycloviolacin H1        GIP--CGESCVYIP-C-LTSAIGCSC--KSKVCYRN
cycloviolacin H2        SAIA-CGESCVYIP-C--FIPGGCSC--RNRVCYLN
cycloviolacin H4        GIP--CAESCVWIP-CTVTALLGCSC--SNNVCY-N
cycloviolacin O1        GIP--CAESCVYIP-CTVTALLGCSC--SNRVCY-N
cycloviolacin O2        GIP--CGESCVWIP-C-ISSAIGCSC--KSKVCYRN
cycloviolacin O3        GIP--CGESCVWIP-C-LTSAIGCSC--KSKVCYRN
cycloviolacin O4        GIP--CGESCVWIP-C-ISSAIGCSC--KNKVCYRN
cycloviolacin O5        GTP--CGESCVWIP-C-ISSAVGCSC--KNKVCYKN
cycloviolacin O7        SIP--CGESCVWIP-CTITALAGCKC--KSKVCY-N
cycloviolacin O8        GTLP-CGESCVWIP-C-ISSVVGCSC--KSKVCYKN
cycloviolacin O9        GIP--CGESCVWIP-C-LTSAVGCSC--KSKVCYRN
cycloviolacin O10       GIP--CGESCVYIP-C-LTSAVGCSC--KSKVCYRN
cycloviolacin O11       GTLP-CGESCVWIP-C-ISAVVGCSC--KSKVCYKN
cycloviolacin O13       GIP--CGESCVWIP-C-ISAAIGCSC--KSKVCYRN
cycloviolacin O17       GIP--CGESCVWIP-C-ISAAIGCSC--KNKVCYRN
cycloviolacin O18       GIP--CGESCVYIP-CTVTALAGCKC--KSKVCY-N
cycloviolacin O19       GTLP-CGESCVWIP-C-ISSVVGCSC--KSKVCYKD
cycloviolacin O20       GIP--CGESCVWIP-C-LTSAIGCSC--KSKVCYRD
cycloviolacin O25       DIF--CGETCAFIP-CITHVPGTCSC--KSKVCYFN
cycloviolacin Y4        GVP--CGESCVFIP-C-ITGVIGCSC--SSNVCYLN
cycloviolacin Y5        GIP--CAESCVWIP-CTVTALVGCSC--SDKVCY-N
cycloviolin A           GVIP-CGESCVFIP-C-ISAAIGCSC--KNKVCYRN
cycloviolin B           GTA--CGESCYVLP-C---FTVGCTC--TSSQCFKN
cycloviolin C           GIP--CGESCVFIP-C-LTTVAGCSC--KNKVCYRN
cycloviolin D           GFP--CGESCVFIP-C-ISAAIGCSC--KNKVCYRN
hcf                     GIP--CGESCHYIP-C-VTSAIGCSC--RNRSCMRN
htf                     GIP--CGDSCHYIP-C-VTSTIGCSC--TNGSCMRN
hyfl A                  SIS--CGESCVYIP-CTVTALVGCTC--KDKVCYLN
hyfl D                  GSVP-CGESCVYIP-C-FTGIAGCSC--KSKVCYYN
hyfl I                  GIP--CGESCVFIP-C-ISGVIGCSC--KSKVCYTN
hyfl J                  GIA--CGESCAYFG-C--WIPGGCSC--RNKVCYFN
hyfl K                  GTP--CGESCVYIP-C-FTAVVGCTC--KDKVCYLN
hyfl L                  GTP--CAESCVYLP-C-FTGVIGCTC--KDKVCYLN
```

```
hypa A                  GIP--CAESCVYIP-CTITALLGCSC--KNKVCY-N
kalata B5               GTP--CGESCVYIP-C-ISGVIGCSC--TDKVCYLN
kalata B16              GIP--CAESCVYIP-CTITALLGCKC--QDKVCY-D
kalata B17              GIP--CAESCVYIP-CTITALLGCKC--KDQVCY-N
kalata B18              GVP--CAESCVYIP-C-ISTVLGCSC--SNQVCYRN
mram 1                  GSIP-CGESCVYIP-C-ISSLLGCSC--KSKVCYKN
mram 2                  GIP--CAESCVYIP-C-LTSAIGCSC--KSKVCYRN
mram 4                  GSIP-CGESCVFIP-C-ISSVVGCSC--KNKVCYKN
mram 5                  GTIP-CGESCVFIP-C-LTSAIGCSC--KSKVCYKN
mram 6                  GSIP-CGESCVYIP-C-ISSLLGCSC--ESKVCYKN
mram 7                  GSIP-CGESCVFIP-C-ISSIVGCSC--KSKVCYKN
mram 8                  GIP--CGESCVFIP-C-LTSAIGCSC--KSKVCYRN
mram 9                  GVP--CGESCVWIP-C-LTSIVGCSC--KNNVCTLN
mram 10                 GVIP-CGESCVFIP-C-ISSVLGCSC--KNKVCYRN
mram 14                 GSIP-CGEGCVFIP-C-ISSIVGCSC--KSKVCYKN
palicourein             GDPTFCGETCRVIPVCTYSAALGCTCDDRSDGLCKRN
vhl                      SIS-CGESCAMISFC-FTEVIGCSC--KNKVCYLN
vhr1                    GIP--CAESCVWIP-CTVTALLGCSC--SNKVCY-N
vibi E                  GIP--CAESCVWIP-CTVTALIGCGC--SNKVCY-N
vibi F                  GTIP-CGESCVFIP-C-LTSALGCSC--KSKVCYKN
vibi G                  GTFP-CGESCVFIP-C-LTSAIGCSC--KSKVCYKN
vibi H                  GLLP-CAESCVYIP-C-LTTVIGCSC--KSKVCYKN
vibi I                  GIP--CGESCVWIP-C-LTSTVGCSC--KSKVCYRN
vibi J                  GTF--CGESCVWIP-C-ISKVIGCAC--KSKVCYKN
vibi K                  GIP--CGESCVWIP-C-LTSAVGCPC--KSKVCYRN
vico A                  GSIP-CAESCVYIP-C-FTGIAGCSC--KNKVCYYN
vico B                  GSIP-CAESCVYIP-C-ITGIAGCSC--KNKVCYYN
vitri A                 GIP--CGESCVWIP-C-ITSAIGCSC--KSKVCYRN
```

Hybrid

```
cycloviolacin B5        GRL---CGERCVIERTRAWCRTVGCICSLHTLECVRN
PS                      GFIP--CGETCIWDKTC-HAAGCSC-SVANIC-VRN
cycloviolacin Y1        GGTIFDCGETCFLGTC--YTPGCSC-GNYGFCYGTN
cycloviolacin Y2        GGTIFDCGESCFLGTC--YTAGCSC-GNWGLCYGTN
cycloviolacin Y3        GGTIFDCGETCFLGTC--YTAGCSC-GNWGLCYGTN
kalata B8               GSVL-NCGETCLLGTC--YTTGCTC-NKYRVC-TKD
tricyclon A             GGTIFDCGESCFLGTC--YTKGCSC-GEWKLCYGTN
tricyclon B             GGTIFDCGESCFLGTC--YTKGCSC-GEWKLCYGEN
kalata B9               G-SVFNCGETCVLGTC--YTPGCTC-NTYRVCTKD
hyfl B                  G-SPIQCAETCFIGKCYTEELGCTC--TAFLCMKN
hyfl C                  G-SPRQCAETCFIGKCYTEELGCTC--TAFLCMKN
hyfl F                  SIS---CGETCTTFNC--WIPNCKCNHHDKVCYWN
hyfl E                  GEIP--CGESCVYPLC--FLPNCYC--RNHVCYLN
mram 3                  GI-P--CGESCVYLPCFTTIIGCKC--QGKVCYH
```

Cyclic trypsin inhibtors

```
MCoTI-2                 GGVCPKILKKCRRDSDCPGACICRGNGYCGSGSD
MCoTI-1                 GGVCPKILQRCRRDSDCPGACICRGNGYCGSGSD
```

The sequences are aligned within their respective subfamily. Sequences were obtained from Cybase [30].

References

1. Kamboj VP: **Herbal medicine**. *Current Science* 2000, **78**(1):35-39.
2. Rates SMK: **Plants as source of drugs**. *Toxicon* 2001, **39**(5):603-613.
3. Wong G: **Biotech scientists bank on big pharma's biologics push (vol 27, pg 293, 2009)**. *Nat Biotechnol* 2009, **27**(5):485-485.
4. Obradovic M, Mrhar A, Kos M: **Market Uptake of Biologic and Small-Molecule-Targeted Oncology Drugs in Europe**. *Clin Ther* 2009, **31**(12):2940-2952.
5. Baumann A: **Early development of therapeutic biologics - Pharmacokinetics**. *Curr Drug Metab* 2006, **7**(1):15-21.
6. Thomma BP, Cammue BP, Thevissen K: **Plant defensins**. *Planta* 2002, **216**(2):193-202.
7. Lay FT, Anderson MA: **Defensins--components of the innate immune system in plants**. *Curr Protein Pept Sci* 2005, **6**(1):85-101.
8. Hu Z, Dun X, Zhang M, Zhu H, Xie L, Wu Z, Chen Z, Xu T: **PA1b, a plant peptide, induces intracellular [Ca2+] increase via Ca2+ influx through the L-type Ca2+ channel and triggers secretion in pancreatic beta cells**. *Sci China C Life Sci* 2007, **50**(3):285-291.
9. Colgrave ML, Craik DJ: **Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot**. *Biochemistry (Mosc)* 2004, **43**(20):5965-5975.
10. Silverstein KAT, Moskal WA, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA: **Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants**. *Plant J* 2007, **51**(2):262-280.
11. Hammami R, Ben Hamida J, Vergoten G, Fliss I: **PhytAMP: a database dedicated to antimicrobial plant peptides**. *Nucleic Acids Res* 2009, **37**(Database issue):D963-968.
12. Dutton JL, Renda RF, Waine C, Clark RJ, Daly NL, Jennings CV, Anderson MA, Craik DJ: **Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins**. *J Biol Chem* 2004, **279**(45):46858-46867.

13. Jennings CV, Rosengren KJ, Daly NL, Plan M, Stevens J, Scanlon MJ, Waine C, Norman DG, Anderson MA, Craik DJ: **Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Mobius strips exist in nature?** *Biochemistry* 2005, **44**(3):851-860.
14. Craik DJ, Cemazar M, Wang CK, Daly NL: **The cyclotide family of circular miniproteins: nature's combinatorial peptide template**. *Biopolymers* 2006, **84**(3):250-266.
15. Craik DJ, Daly NL, Bond T, Waine C: **Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif**. *J Mol Biol* 1999, **294**(5):1327-1336.
16. Gran L: **Isolation of oxytocic peptides from *Oldenlandia affinis* by solvent extraction of tetraphenylborate complexes and chromatography on sephadex LH-20**. *Lloydia* 1973, **36:207-208.**
17. Gustafson KR SR, Henderson LE, Parsons IC, Kashman Y, Cardellina JH, McMahon JB, Buckheit RB, Pannell LK, Boyd MR: **Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree Chassalia parvifolia**. *J Am Chem Soc* 1994, **116:9338-8**.
18. Witherup KM, Bogusky MJ, Anderson PS, Ramjit H, Ransom RW, Wood T, Sardana M: **Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from Psychotria longipes**. *J Nat Prod* 1994, **57**(12):1619-1625.
19. Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG: **Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1**. *Biochemistry (Mosc)* 1995, **34**(13):4147-4158.
20. Tam JP, Lu YA: **A biomimetic strategy in the synthesis and fragmentation of cyclic protein**. *Protein Sci* 1998, **7**(7):1583-1592.
21. Tam JP, Lu YA, Yang JL, Chiu KW: **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):8913-8918.
22. Craik DJ: **Circling the enemy: cyclic proteins in plant defence**. *Trends Plant Sci* 2009, **14**(6):328-335.
23. Ireland DC, Colgrave ML, Daly NL, Craik DJ: **The discovery and development of a natural combinatorial peptide template: the cyclotides**. *Adv Exp Med Biol* 2009, **611**:477-478.

24. Ireland DC, Colgrave ML, Nguyencong P, Daly NL, Craik DJ: **Discovery and characterization of a linear cyclotide from *Viola odorata*: implications for the processing of circular proteins**. *J Mol Biol* 2006, **357**(5):1522-1535.
25. Ireland DC, Colgrave ML, Craik DJ: **A novel suite of cyclotides from *Viola odorata*: sequence variation and the implications for structure, function and stability**. *Biochem J* 2006, **400**(1):1-12.
26. Rosengren KJ, Daly NL, Plan MR, Waine C, Craik DJ: **Twists, knots, and rings in proteins. Structural definition of the cyclotide framework**. *J Biol Chem* 2003, **278**(10):8606-8616.
27. Pallaghy PK, Nielsen KJ, Craik DJ, Norton RS: **A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides**. *Protein Sci* 1994, **3**(10):1833-1839.
28. Craik DJ: **Plant cyclotides: circular, knotted peptide toxins**. *Toxicon* 2001, **39**(12):1809-1813.
29. Wang CK, Colgrave ML, Ireland DC, Kaas Q, Craik DJ: **Despite a conserved cystine knot motif, different cyclotides have different membrane binding modes**. *Biophys J* 2009, **97**(5):1471-1481.
30. Wang CK, Kaas Q, Chiche L, Craik DJ: **CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering**. *Nucleic Acids Res* 2008, **36**(Database issue):D206-210.
31. Tokuoka T: **Molecular phylogenetic analysis of Violaceae (Malpighiales) based on plastid and nuclear DNA sequences**. *J Plant Res* 2008, **121**(3):253-260.
32. Mulvenna JP, Mylne JS, Bharathi R, Burton RA, Shirley NJ, Fincher GB, Anderson MA, Craik DJ: **Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins?** *Plant Cell* 2006, **18**(9):2134-2144.
33. Gruber CW, Elliott AG, Ireland DC, Delprete PG, Dessein S, Goransson U, Trabi M, Wang CK, Kinghorn AB, Robbrecht E *et al*: **Distribution and evolution of circular miniproteins in flowering plants**. *Plant Cell* 2008, **20**(9):2471-2483.
34. Razafimandimbison SG, McDowell TD, Halford DA, Bremer B: **Molecular phylogenetics and generic assessment in the tribe Morindeae (Rubiaceae-**

Rubioideae): How to circumscribe *Morinda L.* to be monophyletic? *Molecular Phylogenetics and Evolution* 2009, **52**(3):879-886.

35. Hernandez JF, Gagnon J, Chiche L, Nguyen TM, Andrieu JP, Heitz A, Trinh Hong T, Pham TT, Le Nguyen D: **Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure**. *Biochemistry (Mosc)* 2000, **39**(19):5722-5730.
36. Laure HJ, Faca VM, Izumi C, Padovan JC, Greene LJ: **Low molecular weight squash trypsin inhibitors from *Sechium edule* seeds**. *Phytochemistry* 2006, **67**(4):362-370.
37. Jennings C, West J, Waine C, Craik D, Anderson M: **Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis***. *Proc Natl Acad Sci U S A* 2001, **98**(19):10614-10619.
38. Gustafson KR, McKee TC, Bokesch HR: **Anti-HIV cyclotides**. *Curr Protein Pept Sci* 2004, **5**(5):331-340.
39. Svangard E, Goransson U, Hocaoglu Z, Gullbo J, Larsson R, Claeson P, Bohlin L: **Cytotoxic cyclotides from *Viola tricolor***. *J Nat Prod* 2004, **67**(2):144-147.
40. Barry DG, Daly NL, Clark RJ, Sando L, Craik DJ: **Linearization of a naturally occurring circular protein maintains structure but eliminates hemolytic activity**. *Biochemistry (Mosc)* 2003, **42**(22):6688-6695.
41. Svangard E, Burman R, Gunasekera S, Lovborg H, Gullbo J, Goransson U: **Mechanism of action of cytotoxic cyclotides: cycloviolacin O2 disrupts lipid membranes**. *J Nat Prod* 2007, **70**(4):643-647.
42. Huang YH, Colgrave ML, Daly NL, Keleshian A, Martinac B, Craik DJ: **The biological activity of the prototypic cyclotide kalata b1 is modulated by the formation of multimeric pores**. *J Biol Chem* 2009, **284**(31):20699-20707.
43. Schöpke T, Hasan, Agha MI, Kraft, R, Otto, A, Hiller, K: **Haemolytisch aktive Komponenten aus *Viola tricolor L.* und *Viola arvensis* murray**. *Scientia Pharmaceutica* 1993, **61**:145 - 153.
44. Hallock YF, Sowder RC, 2nd, Pannell LK, Hughes CB, Johnson DG, Gulakowski R, Cardellina JH, 2nd, Boyd MR: **Cycloviolins A-D, anti-HIV macrocyclic peptides from *Leonia cymosa***. *J Org Chem* 2000, **65**(1):124-128.

45. Bokesch HR, Pannell LK, Cochran PK, Sowder RC, McKee TC, Boyd MR: **A novel anti-HIV macrocyclic peptide from *Palicourea condensata***. *J Nat Prod* 2001, **64**(2):249-250.
46. Lindholm P, Goransson U, Johansson S, Claeson P, Gullbo J, Larsson R, Bohlin L, Backlund A: **Cyclotides: A novel type of cytotoxic agents**. *Mol Cancer Ther* 2002, **1**(6):365-369.
47. Goransson U, Sjogren M, Svangard E, Claeson P, Bohlin L: **Reversible antifouling effect of the cyclotide cycloviolacin O2 against barnacles**. *J Nat Prod* 2004, **67**(8):1287-1290.
48. Simonsen SM, Sando L, Ireland DC, Colgrave ML, Bharathi R, Goransson U, Craik DJ: **A continent of plant defense peptide diversity: cyclotides in Australian *Hybanthus* (Violaceae)**. *Plant Cell* 2005, **17**(11):3176-3189.
49. Zhang J, Liao B, Craik DJ, Li JT, Hu M, Shu WS: **Identification of two suites of cyclotide precursor genes from metallophyte *Viola baoshanensis*: cDNA sequence variation, alternative RNA splicing and potential cyclotide diversity**. *Gene* 2009, **431**(1-2):23-32.
50. Margittai E, Banhegyi G: **Oxidative folding in the endoplasmic reticulum: Towards a multiple oxidant hypothesis?** *FEBS Lett* 2010, **584**(14):2995-2998.
51. Gruber CW, Cemazar M, Clark RJ, Horibe T, Renda RF, Anderson MA, Craik DJ: **A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins**. *J Biol Chem* 2007, **282**(28):20435-20446.
52. Gunasekera S, Daly NL, Anderson MA, Craik DJ: **Chemical synthesis and biosynthesis of the cyclotide family of circular proteins**. *IUBMB Life* 2006, **58**(9):515-524.
53. Trabi M, Mylne JS, Sando L, Craik DJ: **Circular proteins from Melicytus (Violaceae) refine the conserved protein and gene architecture of cyclotides**. *Org Biomol Chem* 2009, **7**(11):2378-2388.
54. Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ: **An asparaginyl endopeptidase mediates in vivo protein backbone cyclization**. *J Biol Chem* 2007, **282**(40):29721-29728.
55. Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA: **Biosynthesis of circular proteins in plants**. *Plant J* 2008, **53**(3):505-515.

56. Bednarek SY, Raikhel NV: **Intracellular Trafficking of Secretory Proteins**. *Plant Mol Biol* 1992, **20**(1):133-150.

57. Hebert DN, Molinari M: **In and out of the ER: Protein folding, quality control, degradation, and related human diseases**. *Physiol Rev* 2007, **87**(4):1377-1408.

58. Otegui MS, Herder R, Schulze J, Jung R, Staehelin LA: **The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies**. *Plant Cell* 2006, **18**(10):2567-2581.

59. Regalado AP, Ricardo CPP: **Study of the intercellular fluid of healthy Lupinus albus organs - Presence of a chitinase and a thaumatin-like protein**. *Plant Physiol* 1996, **110**(1):227-232.

60. Plan MR, Goransson U, Clark RJ, Daly NL, Colgrave ML, Craik DJ: **The cyclotide fingerprint in *Oldenlandia affinis*: elucidation of chemically modified, linear and novel macrocyclic peptides**. *Chembiochem* 2007, **8**(9):1001-1011.

61. Gerlach SL, Burman R, Bohlin L, Mondal D, Goransson U: **Isolation, characterization, and bioactivity of cyclotides from the Micronesian plant *Psychotria leptothyrsa***. *J Nat Prod*, **73**(7):1207-1213.

62. Gunasekera S, Foley FM, Clark RJ, Sando L, Fabri LJ, Craik DJ, Daly NL: **Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides**. *J Med Chem* 2008, **51**(24):7697-7704.

63. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T: **How to Measure and Predict the Molar Absorption-Coefficient of a Protein**. *Protein Sci* 1995, **4**(11):2411-2423.

64. Fiser A, Sali A: **Modeller: Generation and Refinement of Homology-Based Protein Structure Models**. In: *Methods in Enzymology*. Edited by Charles W. Carter, Jr., Robert MS, vol. Volume 374: Academic Press; 2003: 461-491.

65. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences**. *Bioinformatics* 2007, **23**(9):1073-1079.

66. Page RD: **TreeView: an application to display phylogenetic trees on personal computers**. *Comput Appl Biosci* 1996, **12**(4):357-358.

67. Lehrer RI, Rosenman M, Harwig SS, Jackson R, Eisenhauer P: **Ultrasensitive assays for endogenous antimicrobial polypeptides**. *J Immunol Methods* 1991, **137**(2):167-173.

68. Tam JP, Lu YA, Yang JL: **Antimicrobial dendrimeric peptides**. *Eur J Biochem* 2002, **269**(3):923-932.

69. S. NEUPANE SDaTJM: **The *Hedyotis–Oldenlandia–Kohautia* Complex (Rubiaceae) in Nepal: A Study of Fruit, Seed and Pollen Characters and Their Taxonomic Signigicance**. *Edinburgh Journal of Botany* 2009, **66**:371-390

70. Wang R-J: **Two New Species of Hedyotis (Rubiaceae: Hedyotideae) from Hainan, China**. *NOVON* 2008, **18: 264–268**.

71. Ahmad NHLaR: **Phytochemical studies and pharmacological activities of plants in genus *Hedyotis*/*Oldenlandia*** *Studies in Natural Products Chemistry* 2006, **33**(13):1057-1090.

72. Mylne JS, Wang CK, van der Weerden NL, Craik DJ: **Cyclotides are a component of the innate defence of *Oldenlandia affinis***. *Biopolymers*.

73. Yang Y, Zhang S, Howe K, Wilson DB, Moser F, Irwin D, Thannhauser TW: **A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics**. *J Biomol Tech* 2007, **18**(4):226-237.

74. Goransson U, Broussalis AM, Claeson P: **Expression of Viola cyclotides by liquid chromatography-mass spectrometry and tandem mass spectrometry sequencing of intercysteine loops after introduction of charges and cleavage sites by aminoethylation**. *Anal Biochem* 2003, **318**(1):107-117.

75. Lin H, Zhu W, Silva JC, Gu X, Buell CR: **Intron gain and loss in segmentally duplicated genes in rice**. *Genome Biol* 2006, **7**(5):R41.

76. Karehed J, Groeninckx I, Dessein S, Motley TJ, Bremer B: **The phylogenetic utility of chloroplast and nuclear DNA markers and the phylogeny of the Rubiaceae tribe Spermacoceae**. *Mol Phylogenet Evol* 2008, **49**(3):843-866.

77. Goransson U, Craik DJ: **Disulfide mapping of the cyclotide kalata B1. Chemical proof of the cystic cystine knot motif**. *J Biol Chem* 2003, **278**(48):48188-48196.

78. Gray WR: **Disulfide structures of highly bridged peptides: a new strategy for analysis**. *Protein Sci* 1993, **2**(10):1732-1748.

79. Gran L, Sandberg F, Sletten K: ***Oldenlandia affinis* (R&S) DC. A plant containing uteroactive peptides used in African traditional medicine**. *J Ethnopharmacol* 2000, **70**(3):197-203.

80. Canas B, Lopez-Ferrer D, Ramos-Fernandez A, Camafeita E, Calvo E: **Mass spectrometry technologies for proteomics**. *Brief Funct Genomic Proteomic* 2006, **4**(4):295-320.

81. Kamimori H, Hall K, Craik DJ, Aguilar MI: **Studies on the membrane interactions of the cyclotides kalata B1 and kalata B6 on model membrane systems by surface plasmon resonance**. *Anal Biochem* 2005, **337**(1):149-153.

82. Wang CK, Colgrave ML, Gustafson KR, Ireland DC, Goransson U, Craik DJ: **Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis**. *J Nat Prod* 2008, **71**(1):47-52.

83. Mulvenna JP, Sando L, Craik DJ: **Processing of a 22 kDa precursor protein to produce the circular protein tricyclon A**. *Structure* 2005, **13**(5):691-701.

84. Gupta S, Zhang D, Yi J, Shao J: **Anticancer activities of *Oldenlandia diffusa***. *J Herb Pharmacother* 2004, **4**(1):21-33.

85. Reid GE, McLuckey SA: **'Top down' protein characterization via tandem mass spectrometry**. *J Mass Spectrom* 2002, **37**(7):663-675.

86. Gupta MK, Shrivastava P, Singhal PK: **Decomposition of young water hyacinth leaves in lake water**. *Hydrobiologia* 1996, **335**(1):33-41.

87. Daly NL, Craik DJ: **Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and beta-sheet formation in the macrocyclic polypeptide kalata B1**. *J Biol Chem* 2000, **275**(25):19068-19075.

88. Gustafson KR, Walton LK, Sowder RC, Jr., Johnson DG, Pannell LK, Cardellina JH, Jr., Boyd MR: **New circulin macrocyclic polypeptides from *Chassalia parvifolia***. *J Nat Prod* 2000, **63**(2):176-178.

89. Ali M: **Medicinal Plant Diversity and Vegetation Analysis of Logged over Hill Forest of Tekai Tembeling Forest Reserve, Jerantut, Pahang**. *Journal of Agricultural Science* 2010, **2**(3).

90. Shenkarev ZO, Nadezhdin KD, Sobol VA, Sobol AG, Skjeldal L, Arseniev AS: **Conformation and mode of membrane interaction in cyclotides. Spatial structure of kalata B1 bound to a dodecylphosphocholine micelle**. *FEBS J* 2006, **273**(12):2658-2672.

91. Simonsen SM, Daly NL, Craik DJ: **Capped acyclic permutants of the circular protein kalata B1**. *FEBS Lett* 2004, **577**(3):399-402.

92. Skillman JB, Winter K: **High photosynthetic capacity in a shade-tolerant Crassulacean acid metabolism plant - Implications for sunfleck use, nonphotochemical energy dissipation, and susceptibility to photoinhibition**. *Plant Physiol* 1997, **113**(2):441-450.

93. Gunasekera S. DNL, Clark R. J., Simonsen S. M., Foly F. M., Saska I., Lin F., Wade J. D., Anderson M. A., Craik D. J.: **The conserved N-terminal repeat (NTR) region of plant cyclotides does not assist folding of the precursor in vitro**. *American Peptide Symposium* 2007.

94. Andres DA, Rhodes JD, Meisel RL, Dixon JE: **Characterization of the Carboxyl-Terminal Sequences Responsible for Protein Retention in the Endoplasmic-Reticulum**. *J Biol Chem* 1991, **266**(22):14277-14282.

95. Bos MP, Robert V, Tommassen J: **Biogenesis of the gram-negative bacterial outer membrane**. *Annu Rev Microbiol* 2007, **61**:191-214.

96. Schaffer C, Messner P: **The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together**. *Microbiology-Sgm* 2005, **151**:643-651.

97. Glukhov E, Stark M, Burrows LL, Deber CM: **Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes**. *J Biol Chem* 2005, **280**(40):33960-33967.

98. Stadtman ER, Van Remmen H, Richardson A, Wehr NB, Levine RL: **Methionine oxidation and aging**. *Biochim Biophys Acta* 2005, **1703**(2):135-140.

99. Black SD, Mould DR: **Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Posttranslational or Cotranslational Modifications**. *Anal Biochem* 1991, **193**(1):72-82.

100. Wang MY, West BJ, Jensen CJ, Nowicki D, Su C, Palu AK, Anderson G: ***Morinda citrifolia* (Noni): a literature review and recent advances in Noni research**. *Acta Pharmacol Sin* 2002, **23**(12):1127-1141.

101. Potterat O, Hamburger M: ***Morinda citrifolia* (Noni) fruit - Phytochemistry, pharmacology, safety**. *Planta Med* 2007, **73**(3):191-199.

102. West BJ, Tani H, Palu AK, Tolson CB, Jensen CJ: **Safety tests and antinutrient analyses of noni (*Morinda citrifolia* L.) leaf**. *J Sci Food Agric* 2007, **87**(14):2583-2588.

103. Potterat O, Dalsgaard P, Dieterle F, Paululat T, Kuhn T, Hamburger M: **Phytochemical investigation of noni fruit (*Morinda citrifolia*) and noni-derived commercial products**. *Planta Med* 2006, **72**(11):964-965.

104. Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG: **Elucidation of the Primary and Three-Dimensional Structure of the Uterotonic Polypeptide Kalata B1**. *Biochemistry* 1995, **34**(13):4147-4158.

105. Bakshi M, Singh S: **Development of validated stability-indicating assay methods--critical review**. *Journal of Pharmaceutical and Biomedical Analysis* 2002, **28**(6):1011-1040.

106. Brogden KA, Mark A, Paul BM, Brian FT: **Antimicrobial peptides in animals and their role in host defences**. *International journal of antimicrobial agents* 2003, **22**(5):465-478.

107. Brogden KA, DeLucca AJ, Bland J, Elliott S: **Isolation of an ovine pulmonary surfactant-associated anionic peptide bactericidal for *Pasteurella haemolytica***. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(1):412-416.

108. Norton RS, Pallaghy PK: **The cystine knot structure of ion channel toxins and related polypeptides**. *Toxicon* 1998, **36**(11):1573-1583.

109. Craik DJ, Daly NL, Waine C: **The cystine knot motif in toxins and implications for drug design**. *Toxicon* 2001, **39**(1):43-60.

110. Colgrave ML, Craik DJ: **Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot†**. *Biochemistry* 2004, **43**(20):5965-5975.

111. Schweizer F: **Cationic amphiphilic peptides with cancer-selective toxicity**. *Eur J Pharmacol* 2009, **625**(1-3):190-194.

112. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM: **Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases**. *Mol Cell Biol* 1992, **12**(5):2331-2338.

113. Polard P, Chandler M: **Bacterial transposases and retroviral integrases**. *Molecular Microbiology* 1995, **15**(1):13-23.
114. Attele AS, Wu JA, Yuan CS: **Ginseng pharmacology: multiple constituents and multiple actions**. *Biochem Pharmacol* 1999, **58**(11):1685-1693.
115. Jia L, Zhao Y, Liang XJ: **Current evaluation of the millennium phytomedicine- ginseng (II): Collected chemical entities, modern pharmacology, and clinical applications emanated from traditional Chinese medicine**. *Curr Med Chem* 2009, **16**(22):2924-2942.
116. Sanada S, Kondo N, Shoji J, Tanaka O, Shibata S: **Studies on Saponins of Ginseng .1. Structures of Ginsenoside-Ro,Ginsenoside-Rb, Ginsenoside Rc and Ginsenoside Rd**. *Chem Pharm Bull (Tokyo)* 1974, **22**(2):421-428.
117. Dharmananda S: **The Nature of Ginseng: From Traditional Use to Modern Research**. *Internet Journal Of The Institute For Traditional Medicine And Preventive Health Care* Sep 2002.
118. Gran L: **Oxytocic Principles of *Oldenlandia-Affinis***. *Lloydia-the Journal of Natural Products* 1973, **36**(2):174-178.
119. Ha YI, Lim JM, Ko SM, Liu JR, Choi DW: **A ginseng-specific abundant protein (GSAP) located on the cell wall is involved in abiotic stress tolerance**. *Gene* 2007, **386**(1-2):115-122.
120. Lee FC: **Facts about ginseng, the Elixir of Life**: Hollyn International Corp., Elizabeth, NJ; 1992.
121. Chan TWD, But PPH, Cheng SW, Kwok IMY, Lau FW, Xu HX: **Differentiation and authentication of *Panax ginseng, Panax quinquefolius*, and ginseng products by using HPLC/MS**. *Anal Chem* 2000, **72**(6):1281-1287.
122. Yap KYL, Chan SY, Lim CS: **The reliability of traditional authentication - A case of ginseng misfit**. *Food Chemistry* 2008, **107**(1):570-575.
123. Sun C, Li Y, Wu Q, Luo HM, Sun YZ, Song JY, Lui EMK, Chen SL: **De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis**. *BMC Genomics* 2010, **11**:-.
124. Wu Q, Song JY, Sun YQ, Suo FM, Li CJ, Luo HM, Liu Y, Li Y, Zhang XW, Yao H *et al*: **Transcript profiles of Panax quinquefolius from flower, leaf and root**

bring new insights into genes related to ginsenosides biosynthesis and transcriptional regulation. *Physiologia Plantarum* 2010, **138**(2):134-149.

125. Gray WR: **Disulfide structures of highly bridged peptides: A new strategy for analysis**. *Protein Sci* 1993, **2**(10):1732-1748.

126. Mukherjee PK, Kumar V, Kumar NS, Heinrich M: **The Ayurvedic medicine *Clitoria ternatea*--from traditional use to scientific assessment**. *J Ethnopharmacol* 2008, **120**(3):291-301.

127. P. Daisy KSaMR: **Antihyperglycemic and antihyperlipidemic effects of Clitoria ternatea Linn. in alloxan-induced diabetic rats**. *African Journal of Microbiology Research* 2009, **3** (5):5.

128. Rai KS, Murthy KD, Karanth KS, Nalini K, Rao MS, Srinivasan KK: **Clitoria ternatea root extract enhances acetylcholine content in rat hippocampus**. *Fitoterapia* 2002, **73**(7-8):685-689.

129. Rai KS, Murthy KD, Karanth KS, Rao MS: **Clitoria ternatea (Linn) root extract treatment during growth spurt period enhances learning and memory in rats**. *Indian J Physiol Pharmacol* 2001, **45**(3):305-313.

130. Louis S, Delobel B, Gressent F, Rahioui I, Quillien L, Vallier A, Rahbé Y: **Molecular and biological screening for insect-toxic seed albumins from four legume species**. *Plant Science* 2004, **167**(4):705-714.

131. Da Silva P, Strzepa A, Jouvensal L, Rahioui I, Gressent F, Delmas AF: **A folded and functional synthetic PA1b: an interlocked entomotoxic miniprotein**. *Biopolymers* 2009, **92**(5):436-444.

132. Dun XP, Wang JH, Chen L, Lu J, Li FF, Zhao YY, Cederlund E, Bryzgalova G, Efendic S, Jornvall H *et al*: **Activity of the plant peptide aglycin in mammalian systems**. *FEBS J* 2007, **274**(3):751-759.

133. Le Gall M, Quillien L, Gueguen J, Rogniaux H, Seve B: **Identification of dietary and endogenous ileal protein losses in pigs by immunoblotting and mass spectrometry**. *J Nutr* 2005, **135**(5):1215-1222.

134. Hanada K, Nishiuchi Y, Hirano H: **Amino acid residues on the surface of soybean 4-kDa peptide involved in the interaction with its binding protein**. *Eur J Biochem* 2003, **270**(12):2583-2592.

135. Rahioui I, Laugier C, Balmand S, Da Silva P, Rahbe Y, Gressent F: **Toxicity, binding and internalization of the pea-A1b entomotoxin in Sf9 cells**. *Biochimie* 2007, **89**(12):1539-1543.

136. Dun XP, Li FF, Wang JH, Chen ZW: **The effect of pea albumin 1F on glucose metabolism in mice**. *Peptides* 2008, **29**(6):891-897.

137. Zhu H, Choi HK, Cook DR, Shoemaker RC: **Bridging model and crop legumes through comparative genomics**. *Plant Physiol* 2005, **137**(4):1189-1196.

138. Van de Peer Y, Maere S, Meyer A: **The evolutionary significance of ancient genome duplications**. *Nat Rev Genet* 2009, **10**(10):725-732.

139. Wojciechowski MF, Lavin M, Sanderson MJ: **A phylogeny of legumes (Leguminosae) based on analyses of the plastid matK gene resolves many well-supported subclades within the family**. *Am J Bot* 2004, **91**(11):1846-1862.

140. Devi BP, Boominathan R, Mandal SC: **Anti-inflammatory, analgesic and antipyretic properties of Clitoria ternatea root**. *Fitoterapia* 2003, **74**(4):345-349.

141. Watanabe Y, Barbashov SF, Komatsu S, Hemmings AM, Miyagi M, Tsunasawa S, Hirano H: **A Peptide That Stimulates Phosphorylation of the Plant Insulin-Binding Protein - Isolation, Primary Structure and Cdna Cloning**. *Eur J Biochem* 1994, **224**(1):167-172.

142. Louis S, Delobel B, Gressent F, Duport G, Diol O, Rahioui I, Charles H, Rahbe Y: **Broad screening of the legume family for variability in seed insecticidal activities and for the occurrence of the A1b-like knottin peptide entomotoxins**. *Phytochemistry* 2007, **68**(4):521-535.

143. Tan JZ, Lou CF, Zhang GY, Sun BY, Ping YJ: **[cDNA cloning and sequence analysis of leginsulin gene in broad bean (Vicia faba)]**. *Yi Chuan* 2003, **25**(2):168-172.

144. Meyer B, Houlne G, Pozueta-Romero J, Schantz ML, Schantz R: **Fruit-specific expression of a defensin-type gene family in bell pepper. Upregulation during ripening and upon wounding**. *Plant Physiol* 1996, **112**(2):615-622.

145. Chen JJ, Chen GH, Hsu HC, Li SS, Chen CS: **Cloning and functional expression of a mungbean defensin VrD1 in *Pichia pastoris***. *J Agric Food Chem* 2004, **52**(8):2256-2261.

146. Carvalho Ade O, Gomes VM: **Plant defensins--prospects for the biological functions and biotechnological properties**. *Peptides* 2009, **30**(5):1007-1020.

147. Graham MA, Silverstein KAT, VandenBosch KA: **Defensin-like genes: Genomic perspectives on a diverse superfamily in plants**. *Crop Science* 2008, **48**:S3-S11.

148. Higgins TJV, Chandler PM, Randall PJ, Spencer D, Beach LR, Blagrove RJ, Kortt AA, Inglis AS: **Gene Structure, Protein-Structure, and Regulation of the Synthesis of a Sulfur-Rich Protein in Pea-Seeds**. *J Biol Chem* 1986, **261**(24):1124-1130.

149. Daly NL, Rosengren KJ, Craik DJ: **Discovery, structure and biological activities of cyclotides**. *Adv Drug Deliv Rev* 2009.

150. Lindholm P, Goransson U, Johansson S, Claeson P, Gullbo J, Larsson R, Bohlin L, Backlund A: **Cyclotides: a novel type of cytotoxic agents**. *Mol Cancer Ther* 2002, **1**(6):365-369.

151. Hall TJ: **Adaptation and Agronomy of Clitoria-Ternatea L in Northern Australia**. *Tropical Grasslands* 1985, **19**(4):156-163.

152. Gruber CW, Cemazar M, Anderson MA, Craik DJ: **Insecticidal plant cyclotides and related cystine knot toxins**. *Toxicon* 2007, **49**(4):561-575.

153. Hirano H: **Structures and functions of leginsulin and leginsulin-binding protein**. *Protein Structure - Function Relationship* 1996:91-96

298.

154. Hirano H, Komatsu S, Watanabe Y, Barbashov SF, Nishizawa NK, Kajiwara H, Karibe H: **Structure and Function of Plant Leginsulin and Leginsulin-Binding Protein**. *J Cell Biochem* 1995:480-480.

155. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences**. *Nucleic Acids Res* 1990, **18**(20):6097-6100.

156. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator**. *Genome Res* 2004, **14**(6):1188-1190.

157. Whittington CM, Papenfuss AT, Bansal P, Torres AM, Wong ES, Deakin JE, Graves T, Alsop A, Schatzkamer K, Kremitzki C *et al*: **Defensins and the convergent evolution of platypus and reptile venom genes**. *Genome Res* 2008, **18**(6):986-994.

158. Whittington CM, Papenfuss AT, Locke DP, Mardis ER, Wilson RK, Abubucker S, Mitreva M, Wong ESW, Hsu AL, Kuchel PW *et al*: **Novel venom gene discovery in the platypus**. *Genome Biol* 2010, **11**(9):-.

159. Odintsova TI, Egorov TA, Musolyamov A, Odintsova MS, Pukhalsky VA, Grishin EV: **Seed defensins from *T. kiharae* and related species: genome localization of defensin-encoding genes**. *Biochimie* 2007, **89**(5):605-612.

160. Fischbach MA, Clardy J: **One pathway, many products**. *Nat Chem Biol* 2007, **3**(7):353-355.

161. Carbonell P, Faulon JL: **Molecular signatures-based prediction of enzyme promiscuity**. *Bioinformatics*, **26**(16):2012-2019.