NANYANG TECHNOLOGICAL UNIVERSITY

DISCOVERY AND CHARACTERIZATION OF NOVEL CYSTEINE-RICH PEPTIDES IN MEDICINAL PLANTS

TAN WEI LIANG

SCHOOL OF BIOLOGICAL SCIENCES

2017

Discovery and Characterization of Novel Cysteine-rich peptides in Medicinal Plants

TAN WEI LIANG

A thesis submitted to the Nanyang Technological University
In partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2017

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Prof James P Tam for the opportunity to pursue a PhD in his laboratory under his guidance. I am thankful for the opportunities he had provided me over the past four years allowing me to fully experience all aspects of research life. In addition, he imparted his extensive experience through innumerable lessons allowing me to develop myself beyond my expectations. He is indeed a source of inspiration and motivation throughout my PhD journey. I would also like to thank my co-supervisor, Assoc. Prof. Kathy Luo Qian and my thesis advisory committee, Assoc. Prof Koh Cheng Gee and Assoc. Prof Newman Sze, for their invaluable guidance and advice throughout my PhD. I am also grateful to Prof. Rolf Hilgenfeld for the opportunity to work in his laboratory and learn about X-ray crystallography in the University of Lübeck.

Next I would like to thank my laboratory members as this thesis would not have been possible without the help and support from them. In particular, I would like to thank Dr. Wong Ka Ho for the support and guidance he has given me. I am grateful for having him as a great friend and working partner. I would also like to thank Dr. Nguyen Kien Truc Giang for teaching me when I just began my PhD journey, allowing me to quickly get up to speed in my research. I also appreciate the advice and help of Assoc. Prof Liu Chuan Fa, Assoc. Prof Liu Ding Xiang and Dr Huang Mei. I am thankful to have known Kini Shruthi Gopalkrishna, Geeta Kumari, Stephanie Victoria Tay, Huang Jiayi and Yunjiao in the laboratory, the camaraderie that we shared allowed us to achieve many milestones together. I am also thankful to have made many friends in the lab including Clarene Chan, Dr Hemu Xinya, Lim Wei Qin, Dr Aida S. Maqueda, Dr. Xiao Tian Shu, Dr. Nguyen Quoc Thuc Phuong, Dr Qiu Yibo, Dr Fang Minyi, and many others. I am immensely grateful to have the many URECA and FYP students who worked with me and learn along together.

Support is very important in the journey towards achieving a PhD. For that, I am extremely blessed to have a great family who are very supportive of me. They are

always there for me to lend a hand or ear. My parents provided a safe house so that I can freely pursue my dreams. It is because of their unconditional support and love that I am able to achieve what I have achieve today.

My wife Caris has been my greatest supporter throughout this journey. Patiently waiting for me while I complete my experiments without complains, even making the effort to travel long distance to meet. She tries her best to cheer me up when stress level is high, in addition to being understanding. I am glad to be able to meet this amazing lady during my PhD journey.

Table of Contents

List of Figures

List of Tables

Abbreviations

4-VP	4-Vinylpyridine
A. alternata	*Alternaria alternata*
A. brassicicola	*Alternaria brassicola*
A. cathartica	*Allamanda cathartica*
A. hypocondriacus	*Amaranthus hypochondriacus*
ACN	Acetonitrile
BLAST	Basic Local Alignment Search Tool
C. albicans	*Candida albicans*
C. lunata	*Curvularia lunata*
C. tropicalis	*Candida tropicalis*
CHCA	α-Cyano-4-hydroxycinnamic acid
CKAI	Cystine-knot α-amylase inhibitor
COSY	Correlation spectroscopy
CPA	Carboxypeptidase A
CPI	Carboxypeptidase inhibitor
CRP	Cysteine-rich peptide
DTT	Dithiothreitol
E. europaeus	*Euonymus europaeus*
E. longifolia	*Eurycoma longifolia*
E. ulmoides	*Eucommia ulmoides*
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FA	Formic acid
F. oxysporum	*Fusarium oxysporum*
HCl	Hydrochloric acid

HIV	Human Immunodeficiency virus
HPLC	High performance liquid chromatography
IAM	Iodoacetamide
K_i	Inhibition constant
L. barbarum	*Lycium barbarum*
MALDI	Matrix Assisted Laser Desorption/Ionization
MCP	Metallocarboxypeptidase
MIC	Minimal Inhibitory Concentration
MS	Mass spectrometry
MS/MS	Tandem Mass Spectrometry
NEM	N-ethylmaleimide
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
PBS	Phosphate buffered saline
PCI	Potato carboxypeptidase inhibitor
R. solani	*Rhizoctonia solani*
RNA	Ribonucleic acid
RPHPLC	Reversed phase liquid chromatography
S-SAD	Sulfur single-wavelength anomalous dispersion
SCX	Strong Cation exchange
tBLASTn	translated nucleotide BLAST
TCEP	Tris(2-carboxyethyl)phosphine
TCM	Traditional Chinese Medicine
TFA	Trifluoroacetic acid
TMA	Tenebrio molitor α-amylase

TOF	Time of Flight
UDA	Urtica dioica agglutinin
UPLC	Ultra-performance liquid chromatography
WCI	Wolfberry carboxypeptidase inhibitor
W. religiosa	*Wrightia religiosa*

Abstract

Small molecules and proteins represent two major families of pharmaceuticals used clinically. In between these two families, in terms of molecular size, are the disulfide-constrained peptides, a class of compounds that have their drug-like advantages of both small molecules and proteins. Disulfide-constrained peptides share the advantages of proteins for high on-target specificity and low off-target adverse side effects. They also have the robustness of small molecules to tolerate thermal, chemical and enzymatic degradation. Currently, naturally-occurring constrained peptides in plants are an underexplored chemical space in drug discovery.

The objective of my thesis is the discovery and characterization of novel cysteine-rich peptides in medicinal plants. They include *Lycium babarum*, the plant which produces wolfberries, the popular functional food and herb, from which a novel carboxypeptidase inhibitor and a new class of cysteine-rich peptides, lybatides containing a disulfide-stapled helix, were isolated. The wolfberry carboxypeptidase inhibitor inhibits the activity of carboxypeptidase A comparable to the potato carboxypeptidase inhibitor and may account for the anti-thrombotic effect usually associated to wolfberry. In contrast, lybatides isolated from the root bark of the same tree display a structure of naturally-occurring stapled peptides. This shows that one plant is able to produce different cysteine-rich peptides. In *Eurycoma longifolia*, commonly known as Tongkat Ali a popular aphrodisiac in Malaysia, a novel 10C-hevein-like peptide, elongtide 1, was isolated and characterized. The structure and disulfide connectivity of eL1 was determined, confirming the previously predicted disulfide connectivity of the 10C-hevein-like peptide subclass. Together, my thesis expands the

existing knowledge of cysteine-rich peptides and enriches the number of existing disulfide-constrained peptide scaffolds for drug design and peptidyl therapeutic development.

Chapter 1 Introduction

1.1 Peptides as an underexplored chemical space in drug discovery

Commonly used drugs are broadly divided into two major categories, the traditional small molecular drugs that are usually <500 Da and the large protein biologics that occupy the chemical space >10,000 Da. During the 20th Century, natural products or natural product derived molecules represent a key part of drug development. [1]. These leads were either based on bioactivity screening of small molecules or rational design that was ligand-, mechanism- or receptor-based. This enabled the development of many major pharmaceutical successful drugs in the treatment of a wide range of diseases [2].

Small molecular drugs are widely favored by pharmaceutical companies due to several characteristics that they possess. They are usually orally bioavailable, metabolically stable and cheap to produce. However, despite a recent increase in the discovery of additional drug targets, the corresponding increase in the approval of small molecule drugs has not been observed [3].

In the later part of the 20th century, new classes of therapeutics that are on the opposite end of the size spectrum of small molecules began to emerge. This trend was possible due to the development of improved molecular biology tools and advances in recombinant protein expression. These factors coupled with the recognition that proteins have higher potency and specificity for their targets, lead to fewer off-target side-effects, a disadvantage which is one of the biggest drawbacks of using small molecular drugs. These newly developed protein drugs, which include growth factors

and in particular antibodies, are referred to as biologics. However, due to their large size, these biologics are not suitable for oral administration and requires other drug delivery methods like injection and intranasal delivery [4].

With the advent of the genome sequencing efforts at the beginning of the 21st century [5], many predicted that there would be a sudden increase in drug development from the discovery of a wide multitude of novel drug targets. While it is true that technologies in genomics have led to a huge amount of gene expression data, this did not translate into a large amount of validated drug targets [3]. However, it is believed that protein-protein interactions are likely to be involved with many of the new targets obtained from the sequencing efforts, featuring interaction sites with shallow grooves spanning across large surface areas [3]. This class of target is something that small molecular drugs traditionally struggle with and inaccessible for proteins which are membrane impermeable. However, they would be for peptides.

More than 7000 naturally occurring peptides, often involved in crucial physiological roles, have been identified including hormones, neurotransmitters and anti-infectives [6-8]. Peptides are selective and efficacious signaling molecules in general targeting specific cell surface receptors such as G protein-coupled receptors or ion channels. This specificity translates into relatively safe, tolerant and efficient profiles in humans as compared to small molecules [9]. In addition, production of peptide therapeutics is generally cheaper than protein-based therapeutics [4].

The global peptide drug market is predicted to rise from US$14.1 billion in 2011 up to an estimated US$25.4 billion in 2018 [10]. Currently, more than 60 US Food and

Drug Administration (FDA)-approved peptide medicines are available on the market and this number is expected to increase further [11].

Peptides are very suitable for site-specific modification that would allow fine tuning on the potency and specificity of the peptide drug. With rapid development of transcriptomics and proteomics technologies, the use of protein-protein interaction as targets in drug design, resulting in increased efforts on the development of proteins and peptide drugs as drugs. However, low oral bioavailability and metabolic stability, coupled with higher cost in mass production is a major concern in the widespread implementation of peptide and proteins as drugs.

1.2 Cysteine-rich peptide in plants

Plants possess a multitude of defense mechanisms that allow them to survive against harsh environmental factors and pathogenic attacks [12]. The thick cell wall constitutes part of the first line of defense against pathogenic invasions. Pathogens that are able to get past the cell wall are confronted with the plant's cellular innate immune system, comprising of many different defense-related chemical compounds including tannins, polyphenolic compounds, enzyme inhibitors [13] and pathogenesis-related proteins [14].

Plants have been used as a remedy for various illnesses and diseases since prehistoric times. Despite the rapid development of modern medicine in many developing countries, 70-80% of the world's population, according to the World Health Organization (WHO) rely on herbal medicine as the primary form of healthcare [15]. In fact, approximately 80% of the herbal drugs that are in used in recent times have a usage identical to their ethnomedicinal use [15]. However, their mechanisms of actions remain a mystery and continue to be a challenge in natural product development as substantial efforts are required to purify the complex mixture of natural products to single compounds for assays.

Natural products have been a lead-inspiring source in drug discovery. Considerable efforts to isolate and characterize the multitudes of bioactive principles in medicinal plants that are responsible for their therapeutic effects have been made. Most of the studies focus on secondary metabolites produced by the plants [16]. These include Taxol, a chemotherapy agent used in the treatment of several cancer isolated

from *Taxus brevifolia*, quinine, used as an Anti-malarial drug that is isolated from *Cichona officinalis*, morphine, an alkaloid widely used as an analgesic isolated from *Papaver somniferum* and Penicilin, antibiotic drug isolated form the *Penicillium ascomycetous* fungi. Although biologics are playing an increasing role in drug development, accounting for about 30% of approved FDA drugs in 2000 -2010, small molecules remain a key part in drug development pipeline [17]. This is especially in natural products where much of the focus remains in the chemical space of secondary metabolites. In general, proteinaceous compounds are usually dismissed as potential drugs because they are thought to be easily denatured upon heating, digested by proteolytic enzymes during digestion and have low oral bioavailability [18].

Plant cysteine-rich peptides (CRPs) are ubiquitous and found to be involved in many different roles in plant physiology, including cell signaling, reproduction and defense [19]. The multiple cysteine residues present in these CRPs enables the formation of multiple disulfide bonds in different configurations. This result in highly compact and constrained structures allowing them to resist thermal, chemical and proteolytic degradation [20]. In this work, we are interested in plant CRPs that range from 2 kDa to 6 kDa.

Cysteine motifs form a large basis in the classification of plant CRP families where they display a characteristic cysteine pattern and spacing. They are highly evolvable with their scaffold tolerating hypervariable sequences, giving rise to promiscuity in function within each peptide family. The number of CRPs isolated from plants has exceeded thousands and is set to increase further in the future. Analysis of sequences and genomic datamining with the cysteine motifs of these CRPs revealed

that CRPs are highly under-predicted in plants [21]. It had been hypothesized that CRPs account for up to 3% of the expressed proteins in model plants such as rice and *Arabidopsis*.

Plant CRPs are classified into different families based on their sequences, cysteine motifs and disulfide connectivity, which determines their tertiary structure. Table 1.1 shows the major families of CRPs found in plants, including thionins, defensins, hevein-like peptides and knottin-type peptides.

Table 1.1. List of major plant CRP families.

Peptide family	S-S No.	Representative member			Structural motif
		Name	AA No.	Disulfide motif	
Thionin	3-4	Crambin	46	2-C-0-C-11-C-8-C-5-C-7-C-6	
		β-Purothionin	45	2-C-0-C-7-C-3-C-8-C-3-C-1-C-7-C-6	Gamma (Γ) fold β1-α1-α2-β2-coil motif
Defensin	4-5	NaD1	47	2-C-10-C-5-C-3-C-9-C-6-C-1-C-3-C	CSαβ motif β1-coil-α-β2-β3
		PhD1	47	2-C-3-C-6-C-5-C-2-C-0-C-9-C-6-C-1-C-3-C	
Hevein	3-5	Ac-AMP1	29	3-C-4-C-4-C-0-C-5-C-6-C-1	
		Hevein	43	2-C-8-C-4-C-0-C-5-C-6-C-5-C-3-C-2	Gly & Cys rich Central β strands & (short helical) side coils
		EAFP1	41	2-C-3-C-3-C-4-C-0-C-5-C-6-C-5-C-1-C-1-C-2	
Knottin	3	PAFP-S (linear)	38	2-C-6-C-8-C-0-C-3-C-10-C-3	Cystine knot Short β strands & coils
		Kalata B1 (cyclic)	29	4-C-3-C-4-C-4-C-1-C-4-C-3-	Cyclic Cystine knot Short β strands & coils
α-Hairpin	2	Ec-AMP1	37	6-C-3-C-13-C-3-C-8	α-hairpin α1-turn-α2
Jasmintide	3	jS1	27	2-C-2-C-5-C-6-C-3-CC-3	

1.3 Classification of CRPs based on their cysteine framework

CRPs contain a high amount of cysteine residues in their sequences that are involved in the formation of disulfide bonds. Disulfide bonds play an important role in the stabilization of the peptide structure which confers it its characteristic stability. They also have been known to assist in the correct folding of proteins by forming disulfide bridges [22]. The cysteine framework of a cysteine-rich peptide is comprised of two parts, the cysteine spacing and arrangement of cysteine residues in the amino acid sequences, and disulfide connectivity, that is how the cystine residues connect to each other. The combination of the two factors results in an enormous amount of variability and structural diversity.

With plant CRPs in the 2 – 6 kDa range usually containing six to ten cysteines residues, the possibility for a multitude of different disulfide connectivity would be high, about 176 combinations. However, most of the disulfide connectivity observed can be generally classified into two main groups, the type I cystine-knot disulfide connectivity and the type II symmetric disulfide connectivity (Figure 1.1).

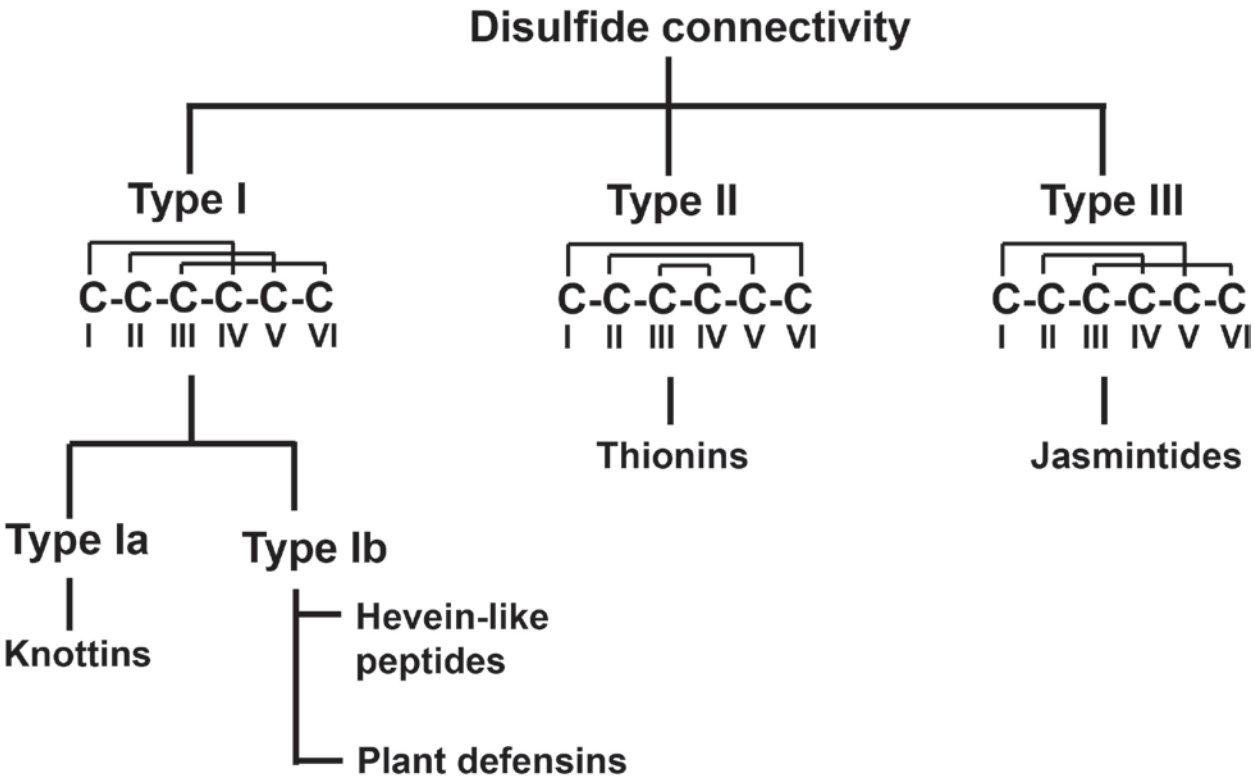


Figure 1.1. Plant CRPs classification based on disulfide connectivity. Type I CRPs contains a cystine knot disulfide connectivity core. Type II CRPs contains a symmetric disulfide connectivity.

The type I cystine-knot disulfide connectivity are displayed by CysI-CysIV, CysII-CysV and Cys III-CysVI. The cystine-knot are one of most commonly observed disulfide connectivity amongst plant CRPs and are usually found in knottins and 6C-hevein-like peptides [23]. These CRPs only contain six cysteine residues which form the cystine knot and are termed type Ia. The cystine knot structural motif is formed when a disulfide bond pierced through a ring that is formed by the backbone interconnected to the other two disulfide bonds (Figure 1.2). The cystine-knot disulfide connectivity can be expanded through the inclusion of additional disulfide bonds such as observed in 8C-hevein-like peptides and defensins [24], these are termed type Ib. The additional disulfide bonds do not usually disturb the knot topology.

In contrast, the type II symmetric disulfide connectivity is displayed by CysI-CysVI, CysII-CysV and CysIII-CysIV. The symmetric is usually observed in plant CRPs such as thionins and α-hairpinins [25]. The disulfide bonds usually do not crisscross, thus there is no disulfide knot formation. Additional disulfide, as seen in 8C thionins, usually continues the symmetric pattern of the disulfide bonds.

Recently, our laboratory identified another disulfide connectivity pattern that was not previously reported for plant CRPs in jasmintides, isolated from *Jasminum sambac* [26]. Jasmintides display a CysI-CysV, CysII-CysIV and CysIII-VI disulfide connectivity different from the type I and type II disulfide connectivity and is thus termed type III.

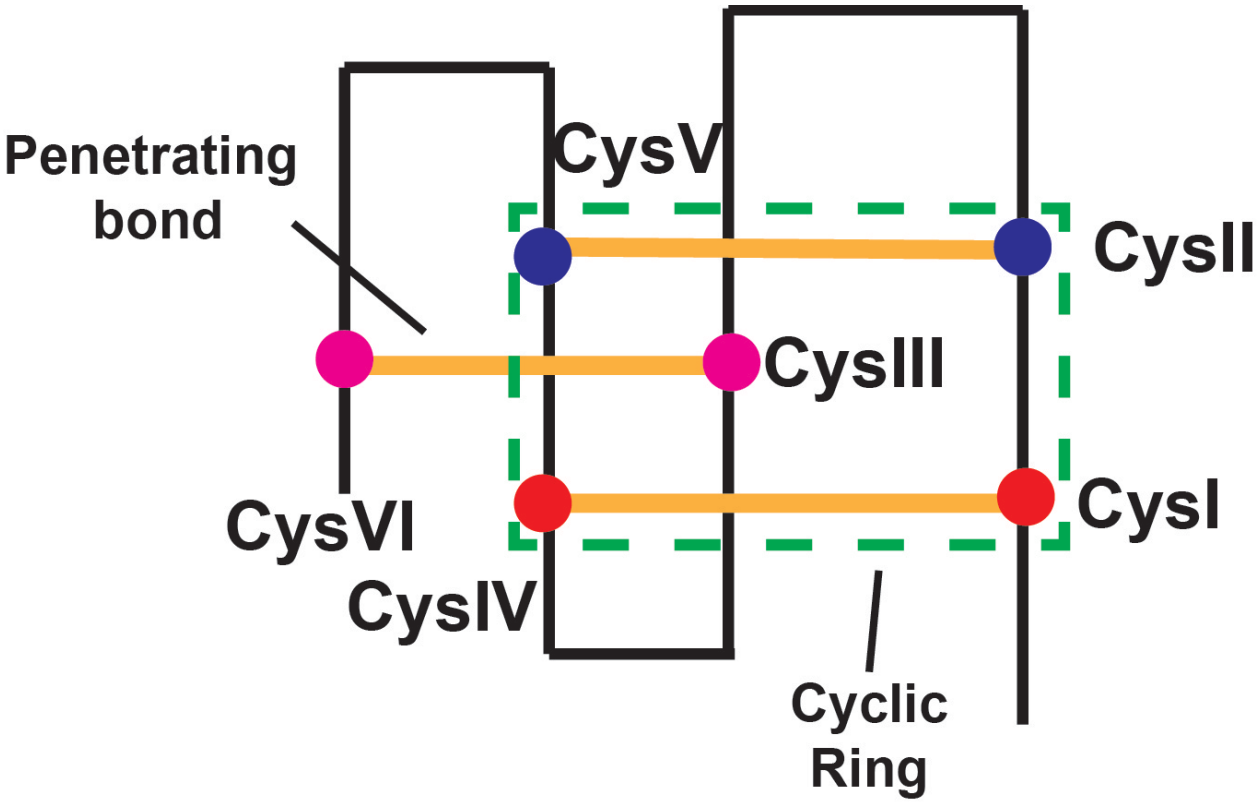


Figure 1.2. Schematic of the cystine-knot connectivity. Cysteine pairs are represented with small circle with different colours and numbered I to VI from N- to C-terminus. The ring, formed by the backbone and two disulfide bonds, is highlighted in green, with the third disulfide bond (CysIII-VI) penetrating it.

The cysteine spacing in CRPs determines the number and length of inter-cysteinyl loops that the peptide possess. The CC motif, with two adjacent cysteine residues and no amino acid in between, is often a diagnostic element in CRP families (Table 1.2). For example, thionins contain a CC motif at their N-terminus, displaying a cysteine spacing of CC-C-C-C-C-C-C, whereas trans-defensins contain a CC motif at the C-terminus as C-C-C-C-C-C-CC [27]. The presence of the CC motif in the middle can be found in CRPs such as hevein-like peptides, cystine-knot α-amylase inhibitors and jasmintides[26,28,29].

As observed in defensins, the CXC motif, present in cis-defensins, and the CC motif, present in trans-defensins, results in difference in overall structure [30]. The disulfides of the cystine residues in the CXC motif points in the same direction, whereas the disulfides of the cystine residues in the CC motif points in the opposite direction. This would allow cysteine-rich peptides that share the same disulfide connectivity to have different overall structures as in the case of defensins.

Together, the disulfide connectivity and cysteine spacing of cysteine-rich peptides combine to give rise to a multitude of different structures. Conoserver, a database for conotoxins, show that there are a total of 27 different frameworks in conotoxins, a much higher number than the frameworks of plant cysteine-rich peptides that were reported (Table 1.3).

Table 1.2. Sequence alignment of plant CRPs from different CRP families.

Family	Peptide	No. of SS	Amino Acid Sequence	Ref.
Defensins	NaD1	4	-RECKTESNTFPGICITKPPCRKAC---ISEKFTDGHC---SKILRRCLCTKPC----	[31]
α-hairpinin	Ec-Amp	2	--GSGR--------GSCRSQCMRR---HEDEPWRVQEC-----VS---QCRRRRGGGD	[32]
Thionins	β-Purothionin	4	-KSCCK--------STLGRNCYNLCRARGAQKLCANVCR--CKLTSGLSCPKDFPK--	[33]
Jasmintides	jS1	3	-QLCLQ--------CRSNSDCN----------IIWRIC-----RDG--CCNVI-----	[34]
Knottins	cT1	3	GIPCGE--------SCVFIPCI----------TGAIGC-----SCKSKVCYRN-----	[35]
Hevein-like peptides	gB5	4	-DPTCS--------VLGDFKCN-----PGRCCSKFNYCGSTAAYCGPGNCIAQCP---	[36]

Table 1.3. Conotoxin cystine frameworks and plant CRP equivalent.

Framework	Cystine Pattern	# Cys	Connectivity	Plant CRPs with similar framework	Ref
I	CC-C-C	4	I-III, II-IV		Gray,W.R. et al. (1981)
II	CCC-C-C-C	6			Ramilo,C. et al. (1992)
III	CC-C-C-CC	6			Sato,S. et al. (1983)
IV	CC-C-C-C-C	6	I-V, II-III, IV-VI	6C-Thionins	Fainzilber,M. et al. (1995)
V	CC-CC	4	I-III, II-IV		Walker,C.S. et al. (1999)
VI/VII	C-C-CC-C-C	6	I-IV, II-V, III-VI	6C-HLPs, CKAIs	Olivera,B.M. et al. (1984)
VIII	C-C-C-C-C-C-C-C-C-C	10			England,L.J. et al. (1998)
IX	C-C-C-C-C-C	6	I-IV, II-V, III-VI	Cyclotides, PCI	Lirazan,M.B. et al. (2000)
X	CC-C.[PO]C	4	I-IV, II-III		Balaji,R.A. et al. (2000)
XI	C-C-CC-CC-C-C	8	I-IV, II-VI, III-VII, V-VIII		Jimenez,E.C. et al. (2003)
XII	C-C-C-C-CC-C-C	8			Brown,M.A. et al. (2005)
XIII	C-C-C-CC-C-C-C	8			Aguilar,M.B. et al. (2005)
XIV	C-C-C-C	4	I-III, II-IV	α-Hairpinins	Moller,C. et al. (2005)
XV	C-C-CC-C-C-C-C	8		8C-HLPs	Peng,C. et al. (2008)
XVI	C-C-CC	4			Pi,C. et al. (2006)
XVII	C-C-CC-C-CC-C	8			Yuan,D.D. et al. (2008)
XVIII	C-C-CC-CC	6			Chen,J.S. et al. (1999)
XIX	C-C-C-CCC-C-C-C-C	10			Chen,P. et al. (2008)
XX	C-CC-C-CC-C-C-C-C	10			Loughnan,M.L. et al. (2009)
XXI	CC-C-C-C-CC-C-C-C	10			Möller,C. and Marí,F. (2011)
XXII	C-C-C-C-C-C-C-C	8		8C-Defensins	Elliger,C.A. et al. (2011)
XXIII	C-C-C-CC-C	6			Ye,M. et al. (2012)
XXIV	C-CC-C	4			Luo,S. et al. (2013)
XXV	C-C-C-C-CC	6			Aguilar,M.B. et al. (2013)
XXVI	C-C-C-C-CC-CC	8			Bernáldez,J. et al. (2013)
XXVII	C-CC-C-C-C	6			Kancherla,A.K. et al. (2015)
NR	CC-C-C-C-C-C-C	8	I-VIII, II-VII, III-VI, IV-V	8C-Thionins	Mak AS and Jones BL (1976)
NR	C-C-C-C-CC-C-C-C-C	10	I-X, II-V, III-VII, IV-VIII, VI-IX	10C-Defensins	Lay FT et al. (2003)
NR	C-C-C-CC-C-C-C-C-C	10	I-V, II-IX, III-VI, IV-VII, VIII-IX	EAFP, WAMP	Andreev Y A et al. (2012)
NR	C-C-CC-C-CC-C-C-C	10	I-IV, II-V, III-VI, VII-X, VIII-IX	Ee-CBP	Van den Bergh KP et al. (2004)

*NR denotes framework not reported in conotoxins

1.4 Major Cysteine-rich Peptides Families

1.4.1 Defensins

Plant defensins are cationic CRPs consisting of 45-54 amino acid residues that belong to the innate immune system of plants primarily targeted against fungal pathogens [37,38]. They are highly sequence diverse although they share a highly conserved molecular structure. They were originally classified as part of the thionins CRP family due to limited sequence identity, however, were later found, based on their structural features that they are unrelated [39]. Plants defensins can be separated into two main groups, 8C plant defensins containing eight cysteines and 10C plant defensins containing ten cysteines (Figure 1.3).

Plant defensins are apparently ubiquitously expressed in almost all plants in the plant kingdom [37,40]. They were identified in different plant tissues with a majority in seeds and roots [41] but also tubers [28], leaves, pods and flowers [42]. Two different precursor organization have been identified in plant defensins, plant C8 class I defensins which only contain the signal peptide and mature peptide domain and the plant C8 class II defensins which have the common three domain precursor organization with the mature peptide domain flanked by the signal peptide and the C-terminal peptide domain [30].

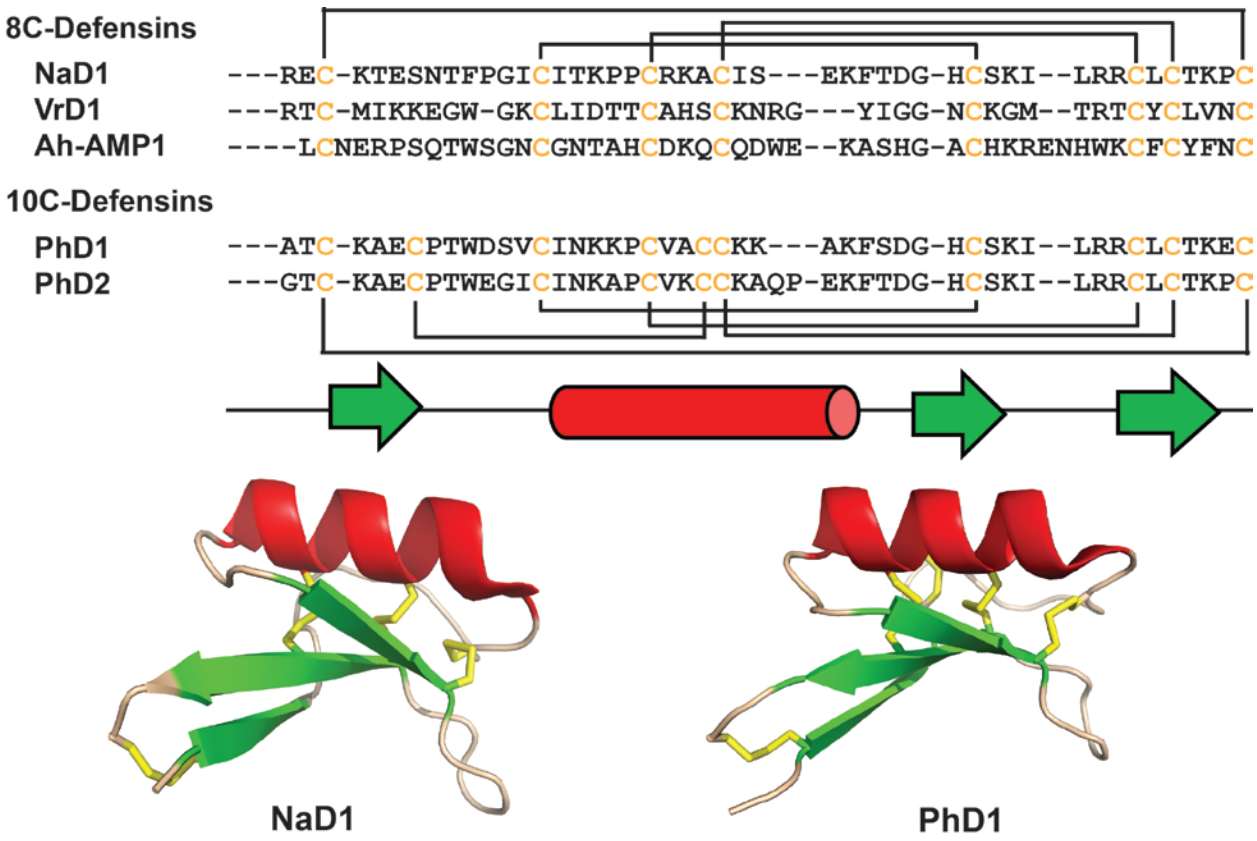


Figure 1.3. Sequences and structures of representative members of plant defensins family. The α-helix present is represented by red and β-strands present is represented by green. (PDBID: NaD1 4AB0, PhD1 1N4N)

Plant defensins typically contains four to five disulfide bonds in the structure and are classified under the cis-defensins superfamily based on the orientation of the disulfides from the C-terminal β-strand [27]. Plant defensins display a cystine-stabilized αβ motif, characterized by four disulfide bonds stabilizing an α-helix and a triple stranded β-sheet [31]. The characteristics of the Cys-stabilized αβ (CSαβ) motif were first characterized in charybdotoxin [43,44]. The secondary structure of this motif is displayed in a β1-coil-α-β2-β3 pattern where the α-helix is parallel to three antiparallel β-strands. The disulfide bonds connecting the middle parallel β-strand to the α-helix stabilizes it. Peptides with the CSαβ motif differs from other cystine-stabilized helical peptides in that the α-helix is stabilized by the C-terminal segment of the peptide with the stabilization scaffold made up of a β-sheet instead of other secondary structural elements. CSαβ peptides are also characterized by the presence of a cystine knot in the structure with an additional exterior disulfide bond that connects the N- and C-terminal of the peptide. PhDs, plant defensins isolated from *Petunia hybrid*, containing ten cysteine residues maintain the conserved disulfide structure with an addition of the extra disulfide bond connecting the N-terminal β-strand to the α-helix [42]. The high number of disulfide bonds present in plant defensins result in them being stable against harsh conditions like high temperature, low pH and proteolytic degradation [37,45].

Plant defensins have been shown to exhibit various biological activities including antifungal, antibacterial and enzyme inhibition activities [46,47]. Plant defensins also play a role in response against pathogenic attack and plant growth and development, in addition to being antimicrobial [19]. It have been demonstrated that NaD1, a plant defensin isolated from *Nicotiana alata*, inhibits the growth of *Botrytis cinerea* and

Fusarium oxysporus [31,48]. ω-Hordothionin, isolated from barley endosperm, shows inhibition in the translational activity in eukaryotic and prokaryotic cell-free systems [49]. Plant defensins S1α1, S1α2 and S1α3, isolated from *Sorghum bicolor* seeds, were able to inhibit the activity of α-amylases [50]. Plant defensins isolated from *Vigna uguiculata* seeds and from the plant *Cassia fistula* were also shown to be able to inhibit trypsin activity [51,52]. Expression of AlfAFP, isolated from *Medicago sativa* seeds, in transgenic potato plants increased the resistance of the plant against the fungal pathogen *Verticillium dahliae* [53].

The antimicrobial activity displayed by plant defensins have been attributed to their positive charge and amphipathic nature. This allows plant defensins to bind to the negatively charged membranes of microbes by interacting with specific binding sites [54,55]. It was shown that this interaction leads to the influx and efflux of cations such as Ca^{2+} and K^{+} [56]. It was also demonstrated with NaD1 that cell membrane permeabilization was not caused via nonspecific insertion of the peptide into membranes but likely through a process that requires a specific receptor on the cell wall, resulting in it entering the cell and inducing reactive oxygen species oxidative stress [57]. These characteristics of defensins enhanced binding of defensins to bacterial membrane allowing them to better, resulting in increased anti-bacterial effects.

1.4.2 Knottins

Plant knottins are a superfamily of plant CRPs that contains approximately 30-40 amino acid residues. They are one of the smallest in size amongst plant antimicrobial CRPs, however is one of the most diverse in terms of biological functions displayed. They usually contain six cysteine residues in their amino acid sequences forming highly conserved disulfide bonds in the CysI-CysIV, CysII-CysV and CysIII-CysVI connectivity (Figure 1.4). This leads to a formation of a cystine knot motif in the three dimensional structure of the peptide. Plant knottins include inhibitors of trypsin, carboxypeptidase and α-amylase and cyclotides, which are cyclic forms of knottins. Although a modified version of the knottin cystine arrangement is also observed in plant defensins and hevein-like peptides, they differ from knottins in terms of their cysteine spacing and disulfide core structure.

The knottin-type peptides were collectively named as knottins, cystine knot inhibitor peptides, as when they were discovered, they were protease inhibitors that share the common cystine-knot motif in their structure. The potato carboxypeptidase inhibitor, from the carboxypeptidase inhibitor subfamily and discovered in 1974, is the first prototypic knottin scaffold discovered [58]. The use of knottins allows the differentiation of cystine-knot inhibitor peptides from the growth factor cystine-knot peptides that are expressed in animals [59]. The penetrating bond of the cystine knot motif is CysIII-CysVI in knottins while in growth factor cystine-knot peptides its CysI-CysIV [60]. Plant knottins have been hypothesized to be one of the largest group of antimicrobial peptides plant peptides even above defensins in terms of the variation of molecular forms and sequence diversity [23].

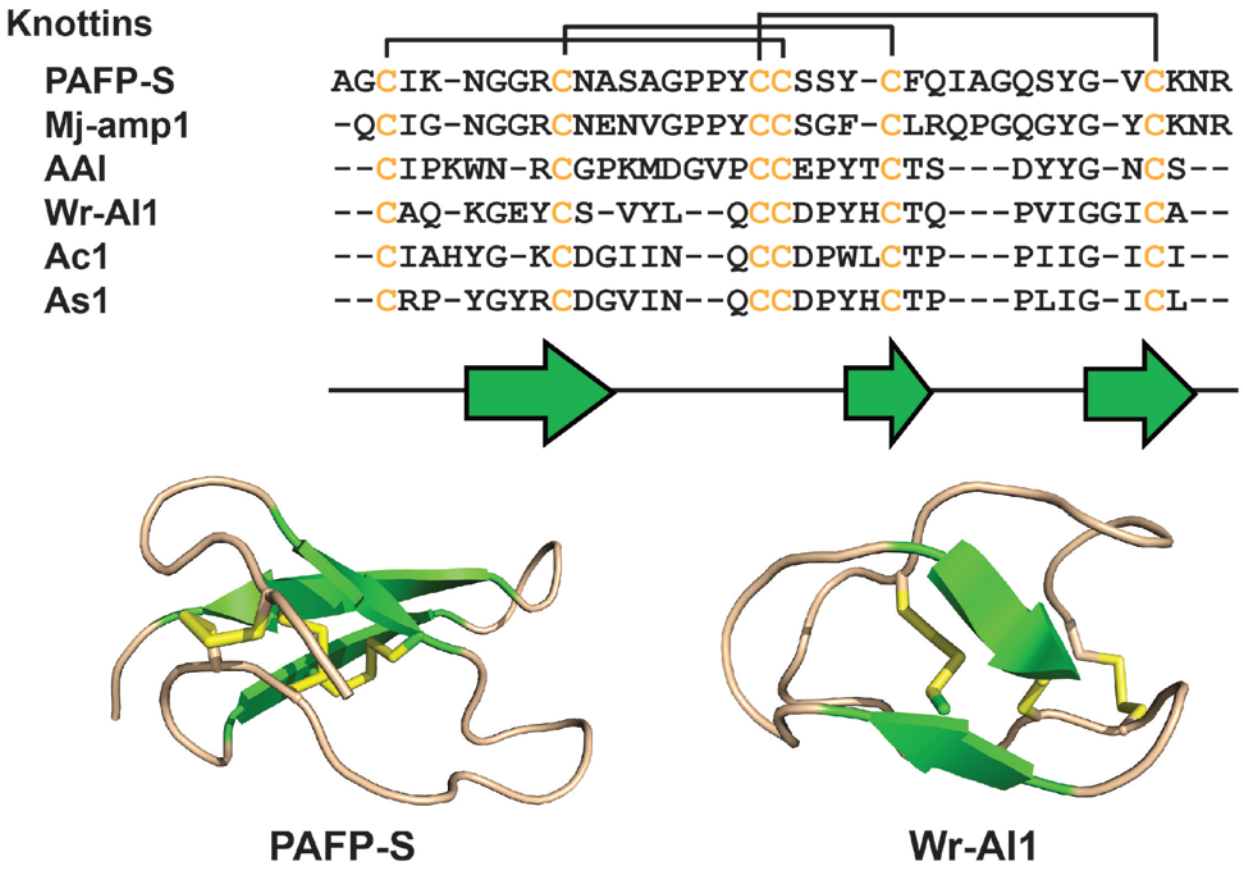


Figure 1.4. Sequences and structures of representative members of plant knottin family. The β-strands present is represented by green. (PDBID: PAFP-S 1DKC, Wr-AI1 2MI9)

Knottins are not only found in plants but in many other organisms as well, including fungi, insects and spiders. The presence of identical or related cystine-knot peptide scaffold found across different kingdom of organisms is an example of parallel evolution of protein structures. Cyclotides and their linear forms have been observed to be only expressed in plants and predicted to be distributed amongst different plant families, including dicot plants of the Rubiaceae, Violaceae, Curcurbitaceae, Fabacaeae and Solanaceae family and monocot Poaceae family [35,61-66]

Cystine-knot α-amylase inhibitors (CKAIs) are inhibitors of the α-amylase enzymes. They are expressed in plants and first isolated from *Amaranthus hypocondriacus* [67]. Out of the seven known families of proteinaceous α-amylase inhibitors, CKAIs are the smallest in terms of molecular size [68]. They differ from other knottins in that they are proline-rich with at least one of the proline residues existing in a cis-conformation [69]. CKAIs are commonly found to be expressed in plants of the Amarathaceae family, including *Wrightia religiosa*, *Allamanda cathartica* and *Alstonia scholaris* [20,28,70].

The common knottin structural motif was first defined in 1994 as a cystine-knot and a triple-stranded β-sheet with a long loop between the first and second β-strand [71]. The first two disulfide bonds (CysI-CysIV and CysII-CysV) and their connecting backbone forms an embedded ring pierced by the third disulfide bond (CysIII-CysVI). The two disulfide bonds (CysII-CysV and CysIII-CysVI) in the knottin scaffold are found to be highly conserved and demonstrated to be sufficient in maintaining the Cys-stabilized β-sheet motif in squash trypsin inhibitors and potato carboxypeptidase inhibitors [72,73]. Knottins belonging to the cyclotide and trypsin inhibitor sub families

are found to exist in two different molecular forms, cyclic and linear, based on the presence or absence of backbone cyclization. Although cyclic knottins of the squash trypsin subfamily are often classified as cyclotides, they share little sequence identity apart from their Cys residues. Both linear (acyclotides) and cyclic (cyclotides) forms of the cyclotide subfamily have been isolated from plants.

Although a common knottin motif was shared between members of the knottin-type peptide family, they display high sequence diversity within the highly conserved CRP scaffold. The ability of the knottin scaffold to tolerate a wide variety of sequences confers it high potential to be used as a scaffold in drug design and peptide engineering. The engineering of knottins through the substitution and/or addition of several consecutive amino acids in the loops without compromising the structural integrity of the molecule have been demonstrated to provide new bioactivity or increased stability [74].

Plant knottin-type peptides, especially cyclotides, have been shown to display high stability against thermal, chemical and enzymatic degradation [20,75,76]. Several studies on cyclotides show that the disulfide bonds present in the knottin scaffold contributes to its stability, while exopeptidase resistance is accounted by the cyclized backbone [66,77]. An alternate mechanism in which knottins resist degradation by exopeptidase without a cyclic backbone structure is through the formation of a pseudocyclic structure as seen in CKAIs [20]. In wrightides, CKAIs isolated from *W. religiosa*, formation of disulfide bonds at the ultimate or penultimate amino acid residues at the N- and C-terminus protects against exopeptidase degradation. The structure of wrightides shows that the termini were able to loop back to the peptide core via the disulfide bonds especially at the N-terminal, resulting in a pseudocyclic structure.

Similar structural features are observed in CKAIs isolated from *Allamanda. cathartica* and *Amaranthus. hypocondriacus*.

Owing to the ability of the knottin scaffold to tolerate highly diverse sequences, knottin-type peptides exhibit a multitude of bioactive activities, from hormone-like functions to enzyme-inhibitory, cytotoxic antimicrobial and anti-HIV activities [71]. “Peptide promiscuity” has been observed in knottin-type peptides where several cystine-knot peptides sharing identical scaffold structures have been demonstrated to be involved in multiple or different biological functions.

The different subfamilies of knottins were demonstrated to possess a broad variety of bioactivities. Several knottin-type peptides exhibit anti-microbial activities and were thus identified as plant antimicrobial peptides (AMPs), including PAFP-S, Mj-AMPs and Psacotheasin. PAFP-S, isolated from *Phytolacca americana* is a knottin that exhibit antifungal activity [78]. Mj-AMP1 and Mj-AMP2 from the seeds of *Mirabilis jalapa* have been demonstrated to possess broad spectrum antimicrobial activity. They inhibit the growth of 12 fungal strains and two Gram-positive bacteria, but were found to not inhibit Gram-negative bacteria and cultured human cells [79]. Psacotheasin have also been shown to be antimicrobial with a minimal inhibitory concentration of 12.5-25 μM.

Several knottin-type peptides were also shown to be inhibitors of enzymes such as α-amylase, carboxypeptidase A or trypsin. Cystine-knot α-amylase inhibitors (CKAI) isolated from plants in the Amaranthaceae and Apocynaceae families are amongst the smallest peptide inhibitors of α-amylase activity. Wr-AI1 and Wr-AI2 isolated from *Wrightia religiosa* was shown to be an inhibitor of α-amylase from *Tenebrio molitor* α-

amylase (TMA) but not the activity of fungal or mammalian α-amylases [20]. Similarly, Allotide Aac4 was observed to display α-amylase inhibitory activity against TMA [70]. However, due to the variation of their N-terminal sequences and the presences of high amount of cis-prolines, allotides do not interact with TMA in the same way as other CKAIs like amaranth α-amylase inhibitor and wrightides. The specificity of CKAIs to insect α-amylases allows them to confer pest resistance in plants and not affect the α-amylase in mammalian digestive system. This suggests that CKAIs may be an attractive target for further development to increase insect resistance in transgenic plants.

Other knottin-type enzyme inhibitors include squash trypsin inhibitors and the knottin-type carboxypeptidase inhibitor. The first reported squash trypsin inhibitor was isolated from cucurbita maxima of the Cucurbitaceae family [80]. These CRPs typically have peptide backbones that are linear except MCoTI-I and MCoTI-II isolated from *Momordica cochinchinensis* which are cyclical in nature [81]. Knottin-type carboxypeptidase inhibitors include potato carboxypeptidase inhibitor [58] and tomato carboxypeptidase inhibitor [82]. Instead of β-strands that are typically observed in knottin-type peptides, potato carboxypeptidase contains long loops. The C-terminal tail inserts into and binds to the active site of carboxypeptidase A in a stopper-like manner reinforced by the secondary binding sites [83,84].

Cyclotides are typically cyclic knottin-type peptides that are 29- 37 amino acids and in length with end to end peptide bond formation resulting in a cyclical backbone (Figure 1.5). However, linear variants of cyclotides, acyclotides, had been identified from *Panicum laxum,* and *Viola odorata* as well [65,66]. Cyclotides are deduced to play a role in the defense of the plant from their insecticidal and antimicrobial activities [85-

87]. Kalata b1, the first cyclotide to be isolated was reported in 1973 from *Oldenlandia affinis* it possess uterotonic activity and was believed to be the main component in the medicinal tea used to accelerate childbirth [88,89]. Cyclotides have been demonstrated to display a wide range of pharmacological functions including antimicrobial, anti-HIV, anti-tumor and immune suppression activities [90,91]. Cyclotides have thus far been reported to be isolated mainly in plants of the Rubiaceae, Violaceae and Solanaceae families. Their highly conserved structures contain highly variable amino acid sequences and are classified into two main structural typed: möbius and bracelet. Möbius types contains a twist in the cyclic backbone and a cis-Pro at loop 5 while the twist and cis-Pro is not observed in the bracelet type [61,92]. However, there are no significant differences in the general scaffold structure between the two types. Sónia et al., showed that kalata B1 can enter cells though both endocytosis and direct membrane translocation. Kalata B1 initiates both pathways by targeting phosphatidylethanolamine phospholipids at the cell surface resulting in the induction membrane curvature [93].

Knottin-type peptides that are membranolytic usually possess an amphipathic nature similar to other AMPs, allowing them to interact with membranes to exhibit their antimicrobial effects. Hydrophobic patches surrounded by several hydrophilic residues were observed on the surface of the PAFP-s and kalata B1 peptide structure. In contrast, thionins and plant defensins are strongly cationic, however, electrostatic interactions between cyclotides and membranes would not be as strong as they are usually neutral or weakly cationic at physiological pH [94]. Previous studies investigated the interactions of kalata B1 with membranes using the detergent

dodecylphosphocholine *in vitro*. It was demonstrated that binding largely involved strong hydrophobic interactions between the loops of kalata B1 and hydrophobic tail of the detergent and weaker interactions between the positively charged amino acid residues of the peptide and the polar head of the detergent [95]. However, since the loops of cyclotides contain hyper variable sequences, leading to different distribution of hydrophobic residues in different cyclotides, similar studies conducted on other cyclotides suggest that different cyclotides may bind to membranes through other mechanisms [96]. One such mechanism exhibited by kalata B1, is the interaction with phospholipids containing phosphatidylethanolamine (PE) headgroups that is further modulated by nonspecific peptide-lipid hydrophobic interactions that are favored in raft-like membranes [97].

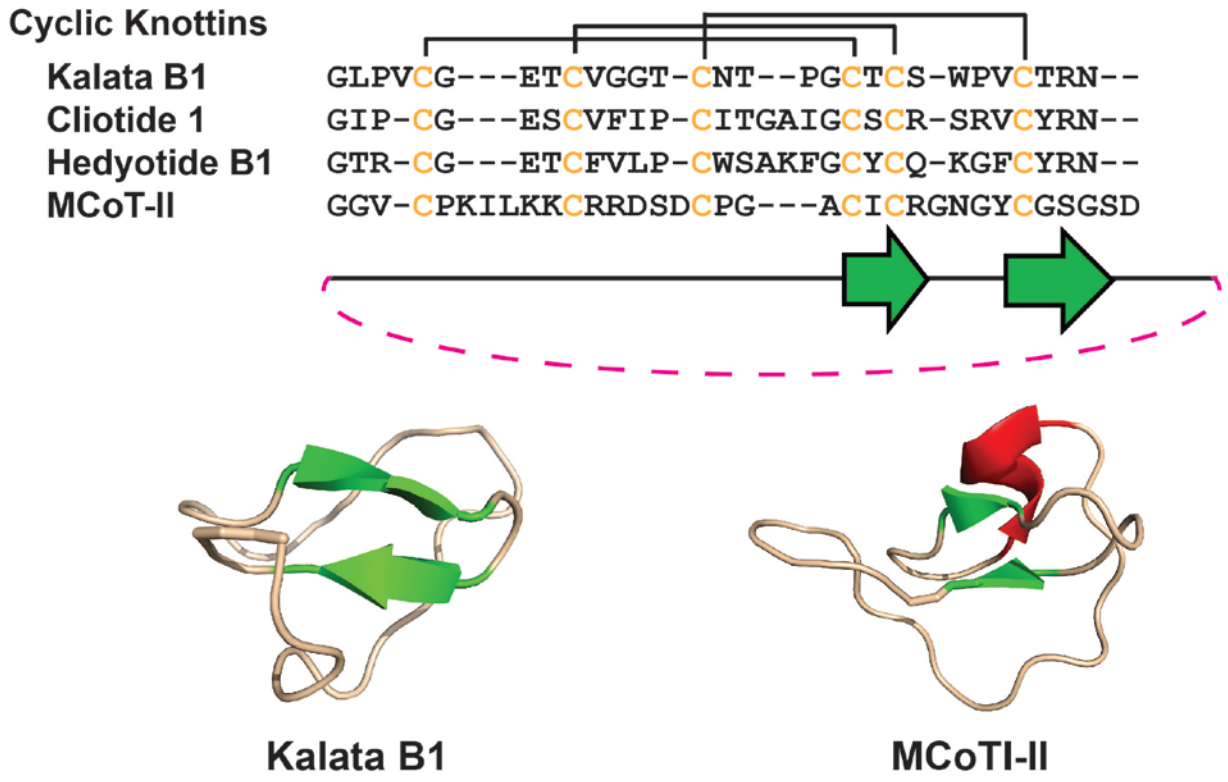


Figure 1.5. Sequences and structures of representative members of cyclic plant knottin family. The α-helix present is represented by red and β-strands present is represented by green. (PDBID: Hevein 1HEV, WAMP 2LB7, EAFP 1P9Z)

1.4.3 Hevein-like peptides

Hevein-like peptides are basic peptides of 29-45 amino acids in length with three to five disulfide bonds. They are usually rich in Gly and contain a chitin binding domain that is highly conserved and can be found in the hevein domain of lectins as well [98,99]. The chitin binding domain usually consists of a Ser and several aromatic residues.

The prototypic hevein-like peptide, hevein was first isolated from the latex extract of the rubber tree *Hevea brasiliensis* in 1960 [100]. It was identified as the protein component of rubber latex with highest abundance and was demonstrated to display strong antifungal activity *in vitro*[101]. The sequence of hevein was found to be similar to the Cys/Gly rich domains of chitinases [102] and chitin-binding lectins [103]. The precursor of hevein is 204 amino acid residues long arranged in a three domain structure, including a signal peptide, followed by a mature domain encoding for hevein and ending with a 144 amino acid residue long C-terminal tail [104]. Additionally, the C-terminal tail was found to contain a Barwin domain. The sequence of hevein was found to be homologous to *Urtica dioica* agglutinin (UDA), a lectin that binds to carbohydrates like chitin containing repeating units of N-acetylglucosamine. UDA was demonstrated to inhibit chitin-containing fungi *in vitro* [105]. Similarly, inhibition of fungal growth was observed when chitin-containing fungal strains were treated with hevein [101]. Different from other antimicrobial peptides like thionins [106], defensins [40] and cyclotides [35], the antifungal activity of hevein is likely due to the presence of the chitin-binding domain.

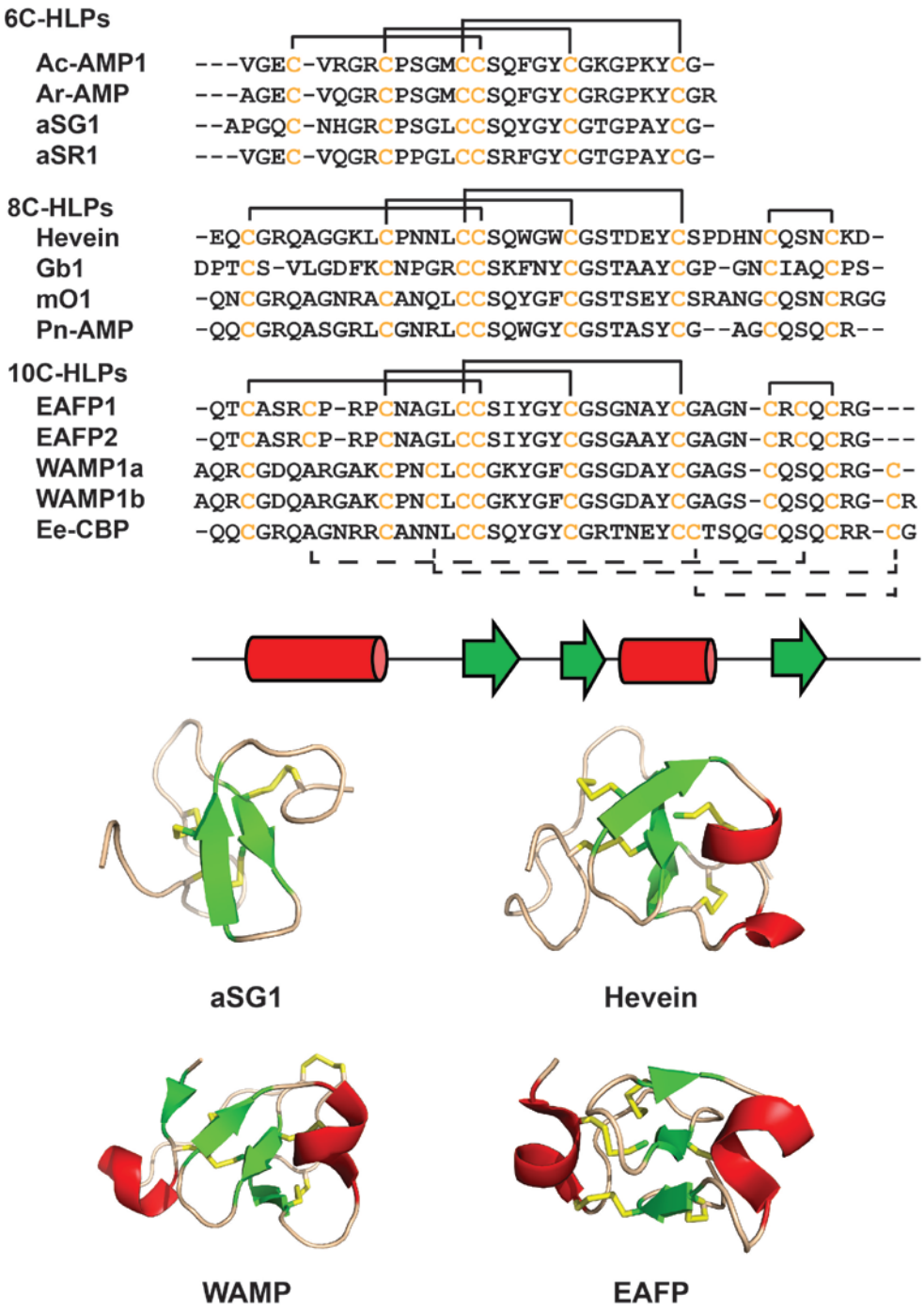


Figure 1.6. Sequences and structures of representative members of hevein-like peptide family. The α-helix present is represented by red and β-strands present is represented by green. (PDBID: Hevein 1HEV, WAMP 2LB7, EAFP 1P9Z)

Hevein-like peptides are divided into three sub-classes based on the number of cysteine residues present, namely 6C-, 8C- and 10C-hevein-like peptides [23] (Figure 1.6). 6C-hevein like peptides are usually 29-30 amino acid residues long with six cysteines. Currently, 13 6C-hevein-like peptides, from plants in the Amaranthaceae family, have been isolated and characterized [29,107-109]. 6C-hevein peptides lack the fourth disulfide bond at the C-terminal region of the mature peptide compared to 8C- and 10C-hevein-like peptides and are thus considered as “truncated” variants of hevein-like peptides.

The majority of hevein-like peptides belong to the 8C-hevein-like peptide sub-class, including hevein, Pn-AMPs from *Pharbtis nil* [110], Fa-AMPs from *Fagopyrum esculentum* [111] and gingkgotides from *Ginkgo biloba* [36]. These hevein-like peptides are usually 37-44 amino acids in length. 10C-hevein-like peptides retain the core disulfide motif as observed in 8C-hevein-like peptides with an extra disulfide bond. Unlike in 6C- and 8C-hevein-like peptides where the disulfide structures are fully conserved, the additional disulfide bond found in 10C-hevein-like peptides are promiscuous, resulting in 10C-hevein-like peptides having the most diverse structure amongst the three subclasses of hevein-like peptides. The promiscuity of the fifth disulfide bond in 10C-hevein-like peptides may result in diverse functionality among the 10C-hevein-like peptide subclass. This phenomenon is similarly observed in other CRPs including defensins and conotoxins where differences in the primary sequences result in different activity in defensins or specificity for different ion channels in conotoxins [112,113].

Table 1.4. Intercysteinyl loops of 6C-, 8C- and 10C-hevein-like peptides

	Amino Acid Sequence
Inter-Cys loop	1 / 2 / 3 / 4 / 5 / 6 / 7
Peptide	`                *  *  *       *`
Ac-AMP1	-VGECVRG----RCPSGMCCSQFGYCGKGPKYCG------------
Ar-AMP	-AGECVQG----RCPSGMCCSQFGYCGRGPKYCG-R----------
aSG1	APGQCNHG----RCPSGLCCSQYGYCGTGPAYCG------------
aSR1	-VGECVQG----RCPPGLCCSRFGYCGTGPAYCG------------
Hevein	--EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKD--
Gb1	-DPTCSVLGDFK-CNPGRCCSKFNYCGSTAAYCG-PGNCIAQCPS--
mO1	--QNCGRQAGNRACANQLCCSQYGFCGSTSEYCSRANGCQSNCR-GG
Pn-AMP	--QQCGRQASGRLCGNRLCCSQWGYCGSTASYCG--AGCQSQCR---
EAFP1	--QTCASRCP-RPCNAGLCCSIYGYCGSGNAYCG-AGNCRCQCRG--
EAFP2	--QTCASRCP-RPCNAGLCCSIYGYCGSGAAYCG-AGNCRCQCRG--
WAMP1a	-AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCG-AGSCQSQCRGC-
WAMP1b	-AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCG-AGSCQSQCRGCR
Ee-CBP	--QQCGRQAGNRRCANNLCCSQYGYCGRTNEYCCTSQGCQSQCRRCG

* Residues in the chitin binding domain motif

Based on the cysteine spacing, 6C-, 8C- and 10C-hevein like peptides can be divided into four, six and seven intercysteinyl loops respectively (Table 1.4). The highly conserved chitin binding domain is located at loop three extending to the end of loop four for the Tyr. The chitin binding domain consist of a Ser, Gly in WAMPs, one Gly and three aromatic residues displayed as S-X-(F/W/Y)-X-(F/W/Y)-C-G-X_4-Y arranged around a cystine core of three to five disulfide bonds [99,114]. In wheat germ agglutinin, urtica dioica agglutinin and several chitinases, multiple repeats of chitin binding domains are observed where they are fused to an unrelated catalytic domain [98]. It was hypothesized that the chitinases elicits its antifungal effect after binding to the chitin-rich fungal walls through the chitin binding domain. However, hevein-like peptides that only contain one unit of chitin binding domain and not fused to another catalytic domain possess stronger antifungal activity than their protein counterpart. The chitin binding domain is also of great medical importance as it is thought to play an important part in the latex fruit syndrome where individuals are observed to have increased sensitivity to fresh fruits like banana if they were allergic to natural rubber latex. The cross-reactivity of allergens have been hypothesized to be caused by IgEs recognizing the chitin-binding domain of class I chitinases in fruits as a result of their similarity in sequence to hevein found in latex [115].

The structure of hevein was elucidated by both X-ray crystallography [116] and NMR [117].It was shown that a cystine core was formed by the first three disulfide bonds while the fourth disulfide is located at the C-terminus. Since then, hevein had been used to study peptide-carbohydrate interactions. In a study by Asensio et. al., the NMR structure of the hevein and N, N'-diacetylchitobiose complex revealed that amino

acid residues Ser19, Trp21, Trp23 and Tyr30 in hevein, later termed as the chitin-binding motif, were involved in the binding interaction [118]. The aromatic residues stabilize the protein carbohydrate complex through π-stacking and if they are Tyr or Typ, they also contribute through hydrogen bonding [25]. Similar to hevein, hevein-like peptides possess a secondary structure arrangement of coil-β1-β2-coil-β3 with varying number of short turns in the two long coiled regions [23]. This secondary structural arrangement differs from the β1-coil-α-β2-β3 arrangement observed in plant defensins and the β1-α1-α2-β2 arrangement observed in thionins [23]. The C-terminal segment of hevein-like peptides usually contains a short helical segment [117], while two anti-parallel β-strands forms a β-sheet with the core stabilized by multiple disulfide bonds.

The general structure of 6C-hevein-like peptides include a two-stranded anti-parallel β-sheet and a short helical region that is stabilized by three disulfide bonds. In terms of its primary amino acid sequences, 6C-hevein-like peptides like Ar-AMP are similar to 8C-hevein-like peptides except the lack of the C-terminal segment consisting of the seventh and eight cysteine residues [109].The preparation of the synthetic truncated hevein, HEV32, supports the hypothesis that the N-terminal domain is able to form a structurally and functionally autonomous entity, similar to other 6C-hevein-like peptides [119]. Nuria Aboitiz, et al. further show that the absence of the C-terminus of hevein reduces the binding affinity of chitooligosaccharide ligands by 20%. This supports what Muraki et al. had suggested that the C-terminal aid in the stability of the binding conformation [120].

Compared to 6C-hevein-like peptides, 8C-hevein-like peptides and 10C-hevein-like peptides, contains an extension of the C-terminal sequence which contains a fourth

disulfide bond. This additional C-terminal sequence constitutes the third β-strand in the antiparallel β-sheet found in 8C- and 10C-hevein-like peptides, and is orientated parallel to the fourth disulfide bond. Cy-AMP1 and Cy-AMP2, isolated from *Cyas revolute*, have a different cysteine motif as the other 8C-hevein-like peptides, even though 8 cysteines are present in the molecule. Instead of having an extra fourth disulfide bond at the C-terminus, two cysteines were found at the 10^{th} and 11^{th} position of the amino acid sequences instead. This may indicate that Cy-AMP1 and Cy-AMP2 belong to a new class of hevein-like peptide more similar to the 6C-hevein-like peptides instead [121].

The main difference between 10C-hevein-like peptides and the other sub-classes of hevein-like peptides is the presence of the fifth disulfide bond [23]. Interestingly, while the other disulfide bonds found in hevein-like peptides are highly conserved, the location of the fifth disulfide bonds found in 10C-hevein-like peptides are highly promiscuous. The 3D structure of 10C-hevein-like peptides changes depending on where the fifth disulfide bond is located. This results in them being the most structurally diverse sub-class of the hevein-like peptides (Figure 1.7). For example, the fifth disulfide bond in EAFP connects Cys7 at the N-terminus to the C-terminus Cys37, resulting in a rigid conformation with a cationic surface [122]. Whereas in WAMP-1a, the fifth disulfide bond connects the C-terminal to loop 1, resulting in a structure that is more similar to chitinases [123].

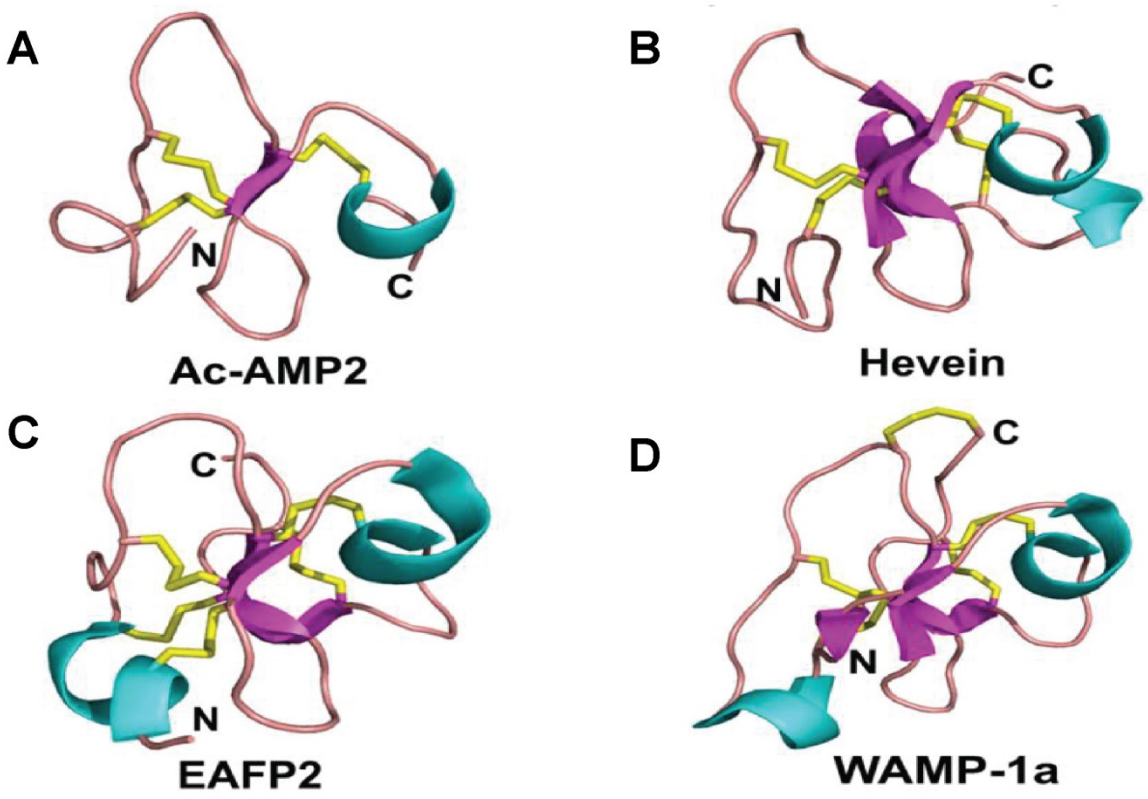


Figure 1.7. NMR structures and scheme of binding interaction. Tertiary structures of (A) Ac-AMP2, (B) hevein, (C) EAFP-2 and (D) WAMP-1a. The secondary structure is depicted by different colors, cyan- α-helix, magenta- β-sheet, yellow- disulfide bonds. (PDBID: Ac-AMP2 1MMC, Hevein 1HEV, EAFP 1P9Z, WAMP 2LB7)

Van Parijs et al. first studied the antifungal activity of hevein as its sequence is homologous with UDA, a protein that is able to inhibit the growth of chitin-containing fungi. Although there is a distinct lack of a catalytic or chitinase domain, inhibition of fungal growth was observed in eight chitin-containing fungi with IC_{50} ranging from 22-312 μM. It was demonstrated that heat was not able to deactivate the anti-fungal activity. It was observed that treatment with hevein on fungal strains results in morphological changes including the formation of thick hyphae [101]. Similar to hevein, hevein-like peptides were demonstrated to be able to inhibit the growth of chitin-containing fungi and are thought to be involved in the defense of the plant against a broad range of fungal pathogenic attacks [98]. 6C-hevein-like peptides were found to possess stronger anti-fungal activity than hevein. However, in addition to inhibiting fungal growth, some 8C and 10C-hevein-like peptides, such as Pn-AMP, Ee-CBP and WAMPs isolated from *Pharbitis nil*, *Euonymus europaeus* and *Triticum kiharae* respectively, were demonstrated to inhibit Gram-positive bacteria growth as well [110,124,125].

Although the anti-fungal activity of hevein-like peptides has been well established, the exact mechanism is still unclear. Currently, there are three hypothesis proposed on the mechanism of action of hevein-like peptides. Van Parijs et al first hypothesized that the small size and chitin binding property of hevein-like peptides could enable the penetration of the fungal cell wall and inhibit nascent chitin chains, thus interfering hyphal growth [101,126]. The chitin-binding property could also disrupt the balance between chitin-synthesis and the hydrolysis of pre-formed chitin chains that is essential for hyphae growth. However, this would not be sufficient to explain the inhibitory activity against non-chitin containing fungi of other hevein-like peptides. An alternative

mechanism of action is proposed by Koo et al., based on confocal studies. It was shown that Pn-AMP was able to penetrate the fungal cell walls at the tips of hyphae and septa after 15 minutes incubation, resulting in the breakage of cytoskeleton followed by the release of cytoplasmic material [110]. The ability to penetrate the fungal cell wall could be attributed to their highly basic pI (12.02), which would be able to explain the broad spectrum anti-fungal activity of highly basic hevein-like peptides. Finally, in a recent study by S Lavokhotova et al., it was found that WAMPs are able to inhibit fungal metalloproteases due to sequence homology with class I/IV chitinases. The fungal metalloprotease FV-cmp cleaves the chitinase at the Gly-Cys site of the chitin-binding domain. Interestingly, WAMPs was found to not only be stable to the degradation of the enzyme, but they could also inhibit the proteolytic activity against chitinases. The inhibition is likely attributed to the Ser residue located in between the Gly-Cys cleavage site in WAMPs which would still encourage binding but prevents the catalytic action of the enzyme, thus resulting in inhibition of the enzyme's activity. This would allow the plant chitinases to then function properly and inhibit fungal growth by degrading the fungal cell wall [127].

The anti-bacterial activity against Gram-positive bacterial strains of hevein-like peptides is likely attributed to the chitin-binding domain of the hevein-like peptides binding to the peptidoglycan layers accounting for approximately 90% of the dry weight of Gram-positive bacteria. Composition of peptidoglycan is similar to chitin, comprising repeated units of N-acetylglucosamine and N-acetylmuraminic acid linked by β-(1-4)-glycosidic bonds, which may aid in the binding of anti-microbial peptides on the bacterial cell surface which results in membrane permeabilization [128]. Accumulation of the

peptide on the surface would cause a loss in membrane integrity and cell death. Together, this demonstrates that hevein-like peptides display diverse modes of actions against microbes from their sequences, net charges and the cell wall composition of the pathogens.

1.4.4 Thionins

Thionins are CRPs of about 45-48 amino acid residues containing six to eight cysteine residues and are generally cationic [39]. The first CRP classified as a thionin, α-purothionin was isolated from wheat endosperm and display antimicrobial activity [77]. Thionins are classified based on similarity to the α-purothionin folds and consists of two main groups, 6C thionins containing six cysteines and 8C thionins containing eight cysteines (Figure 1.8).

Thionins are found to be distributed in various monocots and dicots and are expressed in various plant tissues including roots, leaves and seeds [39]. The expression of thionins in plants is found to be inducible via infection by microbial infection [129]. Thionins are ribosomally synthesized and are expressed in the propeptide form, which are then activated upon pathogenic infection [130,131]. The precursor organization of thionins is the three domain precursor similar to many other CRPs, a signal peptide, followed by a mature domain and a C-terminal domain that is acidic. It was hypothesized that the acidic C-terminal domain keeps the cationic thionin in the inactivate state, with its cleavage resulting in the active form [132].

Thionins are characterized by two antiparallel α-helices and a double stranded antiparallel β-sheet stabilized by three to four disulfide bonds [39]. The disulfide connectivity of CysI-CysVIII, CysII-CysVII, CysIII-CysVI and CysIV-CysV is conserved throughout 8C thionins, with the disulfide bond between CysII-CysVII absent in 6C thionins [23]. Thionins also share a secondary structural motif of β1-α1-α2-β2. The disulfide bond CysI-CysVIII connects β1 to the C-terminal coil, CysII-CysVII connects

the β-strands, while CysIII-CysVI and CysIV-CysV stabilizes the two α-helices. Together, this results in thionins having a gamma fold with an overall structure resembling the capital “Γ” greek letter, comprising of a long arm and a short arm with a groove in between. The structure of 8C thionins shares the same Γ shape with some difference in the C-terminal region when compared to 6C thionins [133,134]. The two α-helices make up the long arm, while the short arm is composed of the β-sheet. It was proposed that the groove in the middle of the two arms plays an important role in the interaction between thionins and membrane lipids [135]. The outer surface of the long arm contains most of the hydrophobic amino acid side chains, while hydrophilic side chains are clustered within the groove or outer face of the short arm.

Thionins were initially known as plant toxins as they display cytotoxicity on bacteria [77], fungi [136], plant and animal cells [137,138], thus suggesting that thionins is involve in the defense system of plants. The cytotoxic effect have been postulated to be a result of the cationic thionin interacting with the target cell membrane, however, the actual mechanism have not been confirmed. A model is proposed by Stec et al., to explain the cytotoxic activity of thionin. They propose that the thionins interact with the anionic phospholipids on the cell membrane leads to cell permeabilization and cell lysis [135].

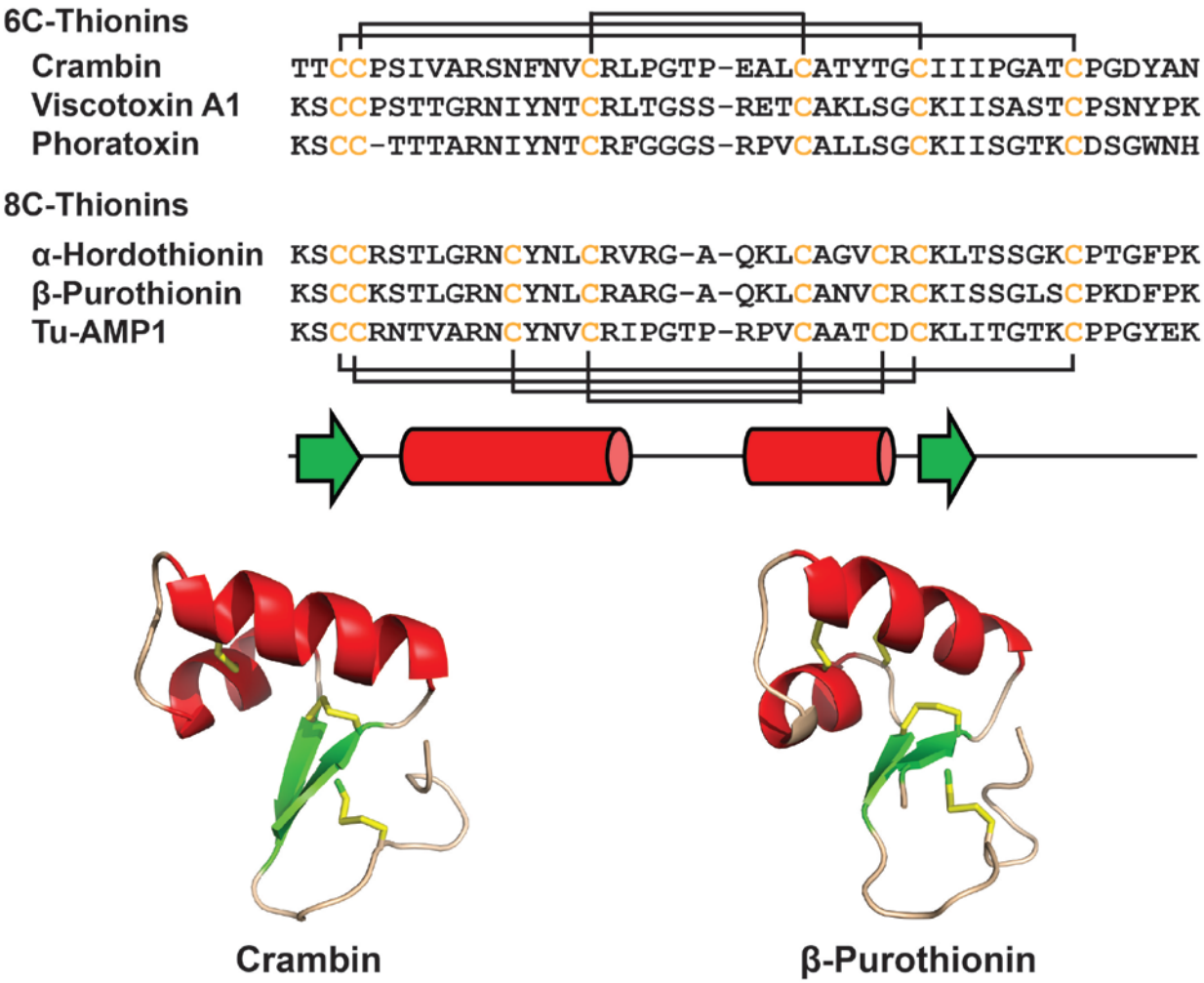


Figure 1.8. Sequences and structures of representative members of plant thionin family. The α-helix present is represented by red and β-strands present is represented by green. (PDBID: Crambin 3NIR, β-purothionin 1BHP)

1.4.5 α-Hairpinin

The CRPs of the α-hairpinin family are plant defense peptides that are rich in Lys and Arg. They share a helix-loop-helix secondary structural motif. Both α-helices in the structure are orientated antiparallel to each other with two disulfide bonds stabilizing the tertiary structure (Figure 1.9). They are structurally distinguished from other plant CRPs containing multiple β-strands such as defensins, knottin-type peptides and thionins.

Several α-hairpinins have been discovered in a variety of agricultural crops. Additionally, some members have been shown to be processed from multimodular precursor proteins [139,140]. MBP-1 and MiAMP2, isolated from maize kernel, *Zea mays*, and *Macadamia integrifolia* respectively, exhibit antifungal activity against several plant pathogenic fungi *in vitro* [141,142]. Ec-AMP1, isolated from the seeds of *Echinochloa crus-gali* was demonstrated to be able to inhibit the growth of several phytopathogenic fungi. It was demonstrated through confocal microscopy that Ec-AMP1 binds to the surface of the fungal conidia, internalize and accumulate in the cytoplasm without affecting the integrity of the membrane [32]. Other biological activities exhibit by members of the α-hairpinin family includes the trypsin inhibition activity of BWI-2c, isolated from seeds of buckwheat *Fagopyrum esculentum* and anti-HIV activity of Luffin P1, isolated from seeds of *Luffa cylindrical*.

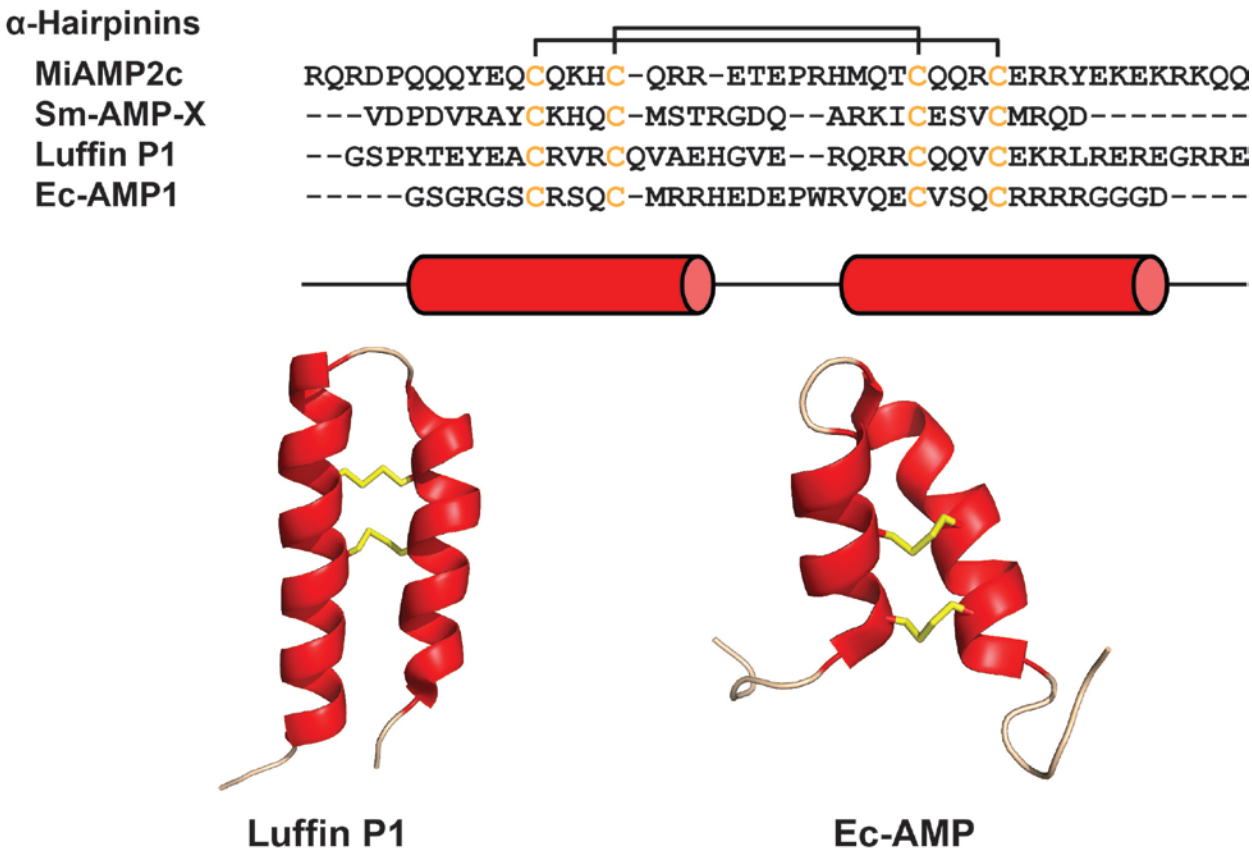


Figure 1.9. Sequences and structures of representative members of α-hairpinin family. The α-helix present is represented by red. (PDBID: Luffin P1 2I37, Ec-AMP 2I2R)

1.5 Biosynthesis of CRPs

CRPs are mini proteins that are gene-encoded and synthesized ribosomally. The precursor of CRPs usually contains multiple domains, including at least one signal peptide domain and one mature peptide domain. The presence of the signal peptide indicates that they may be secretory proteins that are directed to the endoplasmic reticulum for translation. In that case, the signal peptide directs the precursor protein after synthesis to its destination where it is being excised by type I signal peptidase (SPase I) during the translocation [143]. The 16-30 amino acid long signal peptide contains an N-terminal region that is hydrophilic followed by a hydrophobic region and a C-terminal region where the SPase I cleavage site is located [144]. However, it should be noted that the amount of direct evidence for the biosynthesis of plant CRPs is limited.

The maturation of CRPs involves several steps. Following the removal of the signal peptide from the precursor protein after targeting to the ER lumen, folding and disulfide bond formation occurs in the highly oxidative environment. This process is assisted with protein disulfide isomerase and chaperone to ensure that the prepropeptides folds correctly [145]. If the peptides are misfolded, they will remain inside the ER lumen waiting for ER-associated degradation [146,147]. The correctly folded prepropeptides will then leave the endoplasmic reticulum and be transported to the pre-vacuolar compartment and vacuole where further proteolytic processing will take place. The propeptides are then cleaved by proteases in order to release the mature cysteine-rich peptides. Much remains to be known regarding the enzymes involved in the biosynthesis of CRPs. However, recently our lab identified one of the enzymes,

butelase, involved in the processing of cyclotides expressed by the plant *Clitoria ternatea* [148].

CRPs display a wide variety of precursor gene organization (Figure 1.10). The four-domain precursor organization, comprising of an ER signal peptide, prodomain, a mature domain and a short C-tail, is commonly observed in plant knottins such as cyclotides and carboxypeptidase inhibitors [65,86]. The three-domain arrangement: a mature domain flanked by a signal peptide and a C-terminal tail of about 12-15 amino acids, is commonly observed in plant CRPs like thionins, 8C- and 10C-hevein-like peptides and knottins such as CKAIs. However, in defensins, there are two types of precursor structure organization. The plant C8 class II defensins have a similar three domain organization as thionins, whereas the plant class I defensin precursors is a two domain organization that do not contain the C-tail domain, similar to 6C-hevein like peptides. The highly diverse precursor organization suggests that plants may utilize different mechanisms in gene expression modulation in order to better adapt and survive.

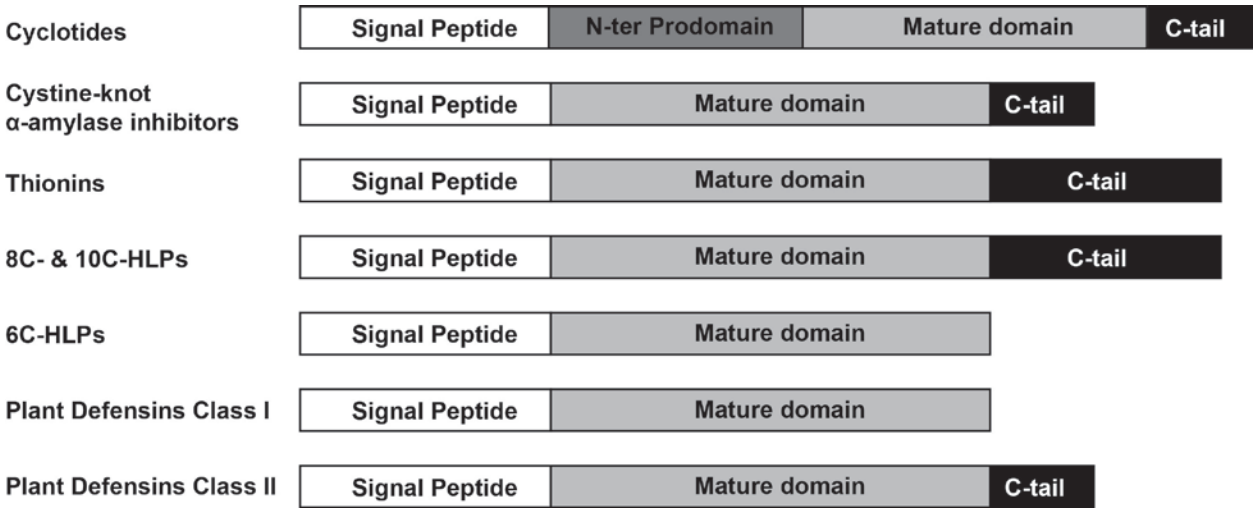


Figure 1.10. **Precursor organization of the major plant CRP families.** Cyclotides exhibit a four domain precursor gene organization. Cystine-knot α-amylase inhibitors, thionins, 8C- and 10C-HLPs exhibit similar three domain precursors where the mature domain is flanked by the signal peptide and C-tail. 6C-HLP display a two domain precursor and are thought to be truncated versions of 8C-HLPs. Plant defensins display two major types of precursor gene organization, the two domain precursor of the Plant Defensins Class I similar to 6C-HLPs and the more common three domain precursors of the Plant defensins Class II.

1.6 Cysteine-rich peptides as drug leads and scaffolds

Peptides have been garnering increasing interests as drug leads and for pharmaceutical developments as they possess the higher specificity of protein therapeutics. However, linear peptides are easily degraded by proteolytic enzymes in the gastrointestinal tract and serum, followed by removal from the blood stream via the kidneys. As such, injection remains the most common delivery method for peptide drugs, thus limiting their use as pharmaceuticals. Their poor stability, which results in short half-lives, coupled with poor oral bioavailability make them an unattractive target for drug development. As a result, the search for a method to develop peptidyl drugs that possess the stability of small molecules while maintaining the specificity of protein therapeutics that could occupy the chemical space between these two groups is ongoing.

Several methods have been developed in order to increase the stability of peptide drugs including cyclization [149], peptide stapling [150], introduction of unnatural amino acid residues [151] and the grafting of linear bioactive peptides into natural hyper-stable scaffolds [152,153]. Grafting of linear bioactive peptides to a scaffold was first done by Vita et al., in 1955 [154]. They engineered a metal-binding site onto charybdotoxin, a cysteine-rich scorpion toxin. The engineering capacity of a cystine-knot scaffold was also demonstrated by Li et al., where they had grafted a p53 inhibitor into a scorpion toxin named BmBKTx1 resulting in an anti-tumor peptide. They also modified several cationic residues to allow cell penetration of the resulting peptide molecule [155]. They also show that the cystine-knot motif was not affected by the introduction of an external peptide, the disulfide bonds were able to form correctly. In addition, the grafted peptide

was able to penetrate the cell membrane and inhibit the p53 process to elicit its anti-tumor activity.

Our laboratory had also grafted a linear bradykinin B1 receptor antagonist DALK and DAK into the scaffold of the cyclotide kB1. The bradykinin B1 receptor is found to be involved in the stimulation of chronic inflammation and inhibition of the receptor could potentially treat inflammatory pain. However, clinical use of its linear antagonists is limited as they are easily degraded by peptidases. The grafting of the linear bradykinin receptor antagonists into the stable scaffold not only result in increased stability against degradation but also increased bioavailability. They also displayed significant pain inhibition as demonstrated in a writhing assay done on mice when the drug is administered orally when compared to its linear analogs [152]. In addition, the use of CRPs as scaffolds for grafting can also be seen in examples like the grafting of the same bradykinin antagonist into a SFTI scaffold and the grafting of a thrombopoietin-mimetic peptides into the EETI-II scaffold [153,156].

Altogether, these various studies show that the CRP scaffolds are highly amicable to modification as their loops are able to tolerate a wide variety of amino acid residues within their intercysteinyl loops. They are demonstrated to confer increased stability without affecting the biological activity of the linear peptide epitopes grafted as well. With the increasing shift in paradigm of therapeutics to peptidyl drugs in clinics, the discovery and application of additional hyper-stable scaffolds gains increasing importance.

1.7 Lycium barbarum

Wolfberry, also known as Goji, is a traditional food and herb in East Asia. It has been gaining popularity in North America and Europe and is found in numerous commercial products. *L. barbarum*, the wolfberry plant, plays an important role in TCM where not only its fruits, the wolfberry, but its root bark is used as a medical herb as well [157].

L. barbarum, from the Solanacaea family, grows to one to three metres tall and is a deciduous shrub with lanceolate to ovate leaves (Figure 1.11). The dark red berries are oblong measuring up to 2 cm in length with a bitter to sweet taste. *L. barbarum* is thought to originate from the Mediterranean Basin, while it is distributed widely in the warmer regions of the world such as the Mediterranean and Southwest and Central Asia. It is also cultivated in North America and Australia. However, majority of the wolfberries that are commercially produced are from the Ningxia Hui region in north-central China.

Although TCM uses mainly the fruit (枸杞子) and root bark (地骨皮) of *L. barbarum*, the leaves and seeds have been mentioned in some medical records, like Yao Xing Lun. Wolfberries are used in TCM for the treatment of a variety of diseases, including blurry vision, infertility, abdominal pain, dry cough and headache. The root barks are utilized in the treatment of night sweating, chronic low grade fever, cough, hemoptysis and hematuria.

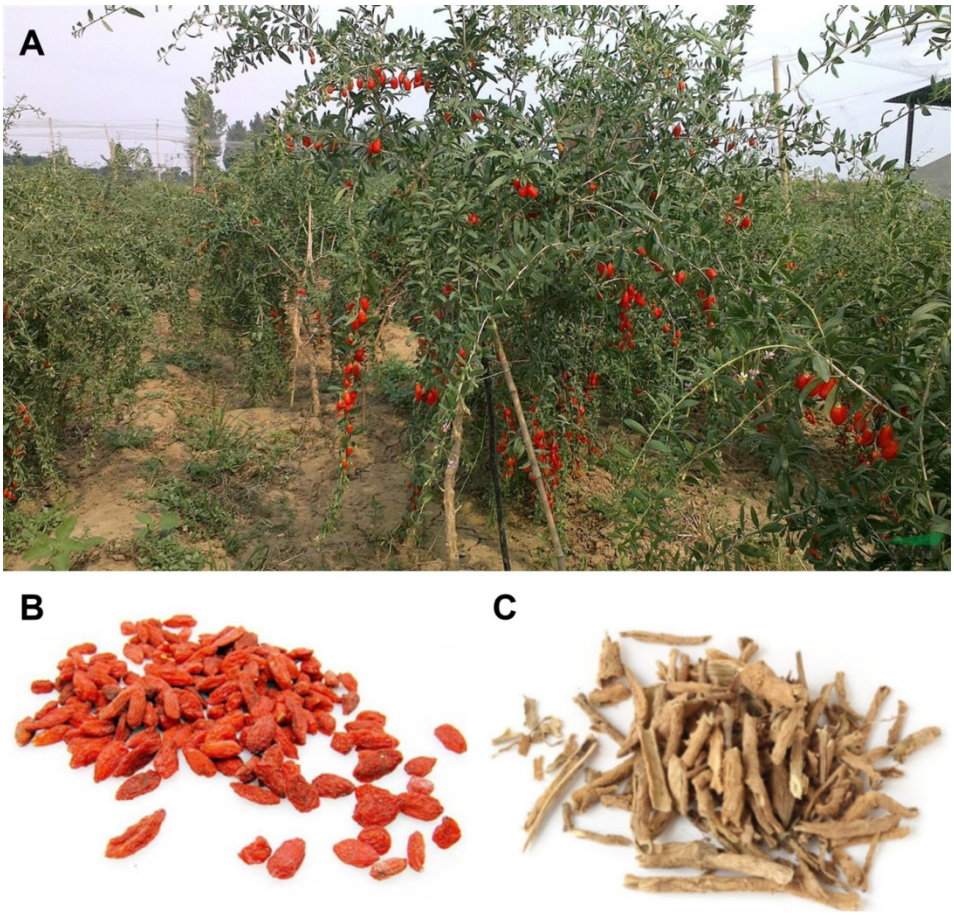


Figure 1.11. *Lycium barbarum* plant. The plant is thought to be originated from the Mediterranean Basin. Currently, it is being cultivated in the Ningxia Hui region in north-central China for commercial purposes. (A) *L. barbarum* plant. (B) Fruit of *L. barbarum*, wolfberry in dried herb form. (C) Root bark of *L. barbarum* in dried herb form, also known as cortex lycii radicis.

Polysaccharides make up one of the most abundant group of compounds extracted from wolfberries with 23% yield after optimization of the extraction procedures [158]. The polysaccharide fraction contains a mixture of highly branched polysaccharides and proteoglycans. Another major group of metabolites found in wolfberry fruit extracts are carotenoids, which increases during the ripening process [159]. In addition, the fruit also contains vitamins [160,161], flavanoids [162], as well as essential oils and fatty acids [163].

The extracts of the root of *L. barbarum* consist of a variety of secondary metabolites including alkaloids [164], polyphenolic compounds [165] and terpenoids [166]. A group of cyclic peptides, lyciumins A-D were also isolated from the extract of *L. barbarum* root bark [167].

Antioxidant activities have been demonstrated in various investigations on *L. barbarum* fruits [168]. These properties have been attributed mainly to polysaccharides [169,170] and flavonoid [162] present in the fruit. The mechanism of actions have been though to involve the chelation of metal ions and free radical scavenging [170]. Other pharmacological activity displayed by wolfberries include the lowering of plasma cholesterol [171], improving insulin sensitivity [172], immunomodulation [173-175] and anti-cancer properties [176-178].

In contrast, different pharmacological activities are displayed by the extracts of root bark. They include alleviating high blood pressure, reducing blood glucose and anti-pyretic activity in animal models [164]. The alkaloid kukomine A has been demonstrated to contribute to the hypotensive effect [164]; however, the peptides isolated from the

root bark, lyciumins A and B, may be involved as well as they had been shown to inhibit angiotensin converting enzyme and renin [179]. Aqueous extracts of the root bark have also been demonstrated to protect against the toxic effects induced by CCl_4 in vivo and to have anti-inflammatory activity [180].

1.8 Eurycoma longifolia

Eurycoma longifolia, also known as tongkat ali, is a popular traditional medical herb native to South East Asia. It is well-known in its use as an aphrodisiac and for its Anti-malarial effects in Asia. In Malaysia, it is highly regarded as an adaptogen, gaining it the title of “Malaysian ginseng” [181]. *E. longifolia* is a tall, lean, shrubby tree from the Simaroubaceae family, usually growing in sandy soil (Figure 1.12). It has compound leaves that are pinnate in shape and are green in colour. The flowers are small, red, unisexual and are arranged densely together [182].

Traditionally, *E. longifolia* is a very popular folk medicine in South East Asia due to its claimed aphrodisiac properties and in the treatment of malarial fever [183]. It is traditionally administered orally as a water decoction of the roots. In recent times, it is also available in more convenient forms such as raw crude powders of the root and capsules or as additives in many commercial products such as coffee and teas. It is estimated that 21,000 kg of *E. longifolia* are being harvested a year with demand for it reaching >54,000 kg [184].

E. longifolia roots contains a wide variety of chemical compounds which would account for the multitude of pharmacological effects associated with its traditional use. They include quassinoids, alkaloids, triterpenes and bioactive steroids [185]. Quassinoids, a group of nortriterpenoids, represents a major portion of the phytochemical components extracted from the root [186]. These include various types of eurycommanone, eurycomanols and eurycomalactones [187,188].

Figure 1.12. *Eurycoma longifolia* plant cultivated mainly in the South East Asian region. It is also known as “Malaysian Ginseng” as it is used in the treatment of a multitude of illnesses in traditional folk medicine. (A) *E. longifolia* plant. (B) Root of *E. longifolia* plant. (C) Dried herb form of the root of *E. longifolia*.

Eurycomanone, a major quassinoid from *E. longifolia* extract had been found to significantly increase the production of testosterone in a dose dependent manner of rat testicular Leydig cell-rich interstitial cells [189]. Several other studies involving *E. longifolia* extracts demonstrated that it is able to improve male fertility in vivo and in clinical trials [190-192] as well. 7-MCPA, a β-carboline alkaloid, isolated from the hairy-root cultures was shown to be able to activate Nrf2 via a ROS-dependent p38 MAPK pathway inhibiting inflammation resulting in potential for use in the treatment of inflammatory diseases [193].

Aims

Two categories of successful drugs are the small molecules which are less than 500 Da, and proteins or biologics which are larger than 10,000 Da. In between, peptides from medicinal plants occupy a chemical space that is relatively less developed in drug discovery research. Peptides, especially CRPs, could fill the gap of novel pharmaceuticals and may represent a class of compounds that possess the best of both worlds. CRPs retain the specificity and the potency of the larger biologics to inhibit protein-protein interactions and the robustness of the small molecular drugs which are characterized by high metabolic stability and cheap production cost. In this thesis, I aim to develop the underexplored chemical space for drug development through identification and characterization of CRPs to elucidate their potential to be developed for use as therapeutics agents or as scaffold in peptidyl drug engineering.

The specific aims of my thesis are to isolate and characterize novel CRPs from:

1. the highly popular wolfberry, *Lycium barbarum*
2. the root bark of *Lycium barbarum*, a commonly used herb in Traditional Chinese Medicine
3. Eurycoma longifolia, Tongkat Ali, root, a popular herb widely known as "Malaysian Ginseng"

In my thesis, I focus on a special class of peptides, cysteine-rich peptides, which are highly stable and contain 15 to 60 amino acids. CRPs are expressed in plants,

specifically *L. barbarum* and *Eurycoma longifolia*. Using a combination of transcriptomic and peptidomic approach, I was able to identify and characterize several cysteine-rich peptides in these medicinal plants. A carboxypeptidase inhibitor was identified in the fruit of *Lycium barbarum*, a highly popular TCM herb and functional food. In the root bark of the same plant, two lybatides were isolated. Lybatides display a novel cysteine spacing and disulfide connectivity that was not reported in plant CRPs. The structure of lybatide 2 was analyzed by X-ray crystallography and it was revealed that the overall structure resembles a disulfide stapled α-helical peptides. Finally, a 10C-hevein-like peptide, elongtide, was isolated and characterized in the root of *E. longifolia*. This is the first reported hevein-like peptide found to be expressed in the root of a plant. The structure and disulfide connectivity of elongtide was also determined by NMR, thus confirming the previously predicted disulfide connectivity of this sub-class of 10C-hevein-like peptides.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Chemical reagents

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

Acetic acid	Merck
Acetonitrile (ACN)	Fisher
Agarose	Bio-Rad
Ammonium bicarbonate (NH_4HCO_3)	Sigma-Aldrich
C18 media	Grace Davison Discovery Sciences
Dichloromethane (DCM)	Merck
Dithiothreitol (DTT)	Sigma-Aldrich
dNTP nucleotide mix	Fermentas
Ethanol (EtOH)	Merck
Formic acid (FA)	Sigma-Aldrich
Iodoacetamide (IAM)	Sigma-Aldrich
Isopropanol	Fisher
Methanol	Merck
Sodium chloride	Sigma-Aldrich

2.1.2 Plant materials

E. longifolia were obtained from Singapore botanic gardens and Thong Hup Gardens Pte, Singapore. *L. barbarium* and *lycii radicis was* purchased from a local herb distributor, Hung Soon Pte Ltd, Singapore and authenticated by Mr. Li Cai Ming. A specimen was deposited at the Nanyang Technological University herbarium under voucher number TL-LB-20140302 and TL-LB-20150520, respectively.

2.1.3 Fungal strains

Five phyto-pathogenic fungal strains were acquired from China Center of Industrial Culture Collection, including *Curvularia lunata* (CICC 40301), *Fusarium oxysporum* (CICC 2532), *Rhizoctonia solani* (CICC 40259), *Alternaria alternata* (CICC 2465) and *Alternaria brassiciola* (CICC 2646).

2.1.4 RNA extraction

Plant material was homogenized using liquid nitrogen. TRIzol® Reagent (Life Technologies) was added to the homogenized sample and incubated at room temperature for 5 min to allow complete dissociation of the nucleoprotein complex. Chloroform (0.2 mL) was added to the sample and incubated for 3 minutes followed by centrifugation at 12,000 rpm for 15 min. Equal volume of isopropanol and 25% salt solution (1.2 M sodium chloride + 0.8 M sodium acetate) were added to the RNA containing aqueous phase and incubated at room temperature for 10 min to allow precipitation of RNA. The samples were then centrifuged at 12000 rpm for 10 min and the RNA pellet was washed with 75% ethanol (1 mL). The pellet was dried and resuspended in diethylpyrocarbonate (DEPC) water (30 µL). Concentration and purity of

extracted RNA was measured using NanoPhotometer™ (Implen, Northstar Scientific, Germany).

2.1.5 Sequence analysis

Sequencing results were analyzed using the BioEdit software. ExPasy translate tool (http://www.expasy.ch/tools/dna.html) was used to predict the amino acid sequences of the clones. Prediction of ER signal sequences of the precursor sequences was performed using SignalP 4.0 prediction server (http://www.cbs.dtu.dk/services/SignalP/) [194,195]. Logo sequences were built with WebLogo application, version 3 (http://weblogo.berkeley.edu/) [196].

2.2 Proteomics

2.2.1 HPLC and UPLC analysis

Shimadzu systems were used for high performance liquid chromatography (HPLC). Reverse-phase-HPLC (RP-HPLC) was performed using Phenomenex C18 columns (particle size, 5 μm; pore size, 300 Å; Hesperia, CA, USA) with dimensions of 250 x 22 mm, 250 x 10 mm, and 250 x 4.6 mm for preparative, semi-preparative and analytical HPLC at a flow rate of 5 mL/min, 3mL/min and 1 mL/min respectively. A PolyLC polysulfoethyl A column was used for strong cation exchange (SCX)-HPLC with dimensions of 250 x 9.4 mm and 250 x 4.6 mm run at flow rates of 3 and 1 mL/min flow rates, respectively.

2.2.2 MALDI-TOF MS and MS/MS

Mass spectrometry analysis was performed using ABI 4800 MALDI-TOF/TOF system (Applied Biosystems, Framingham, MA, USA). A saturated solution of CHCA in 60% ACN, 0.05% TFA was used as the matrix for matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) and MS/MS. Samples were mixed in a 1:1 ratio (v/v) with the matrix and spotted on the target plate.

The reflectron mode with laser intensity set between 3000-4000 was used to scan both MS and MS/MS spectra. Average spectra for MS and MS/MS were obtained from 1000 and 5000 shots with an accelerating voltage of 20 kV and 8 kV respectively.

2.2.3 Protein extraction and purification

Dried herbal samples (1-3 kg) of *Lycium barbarium* and Cortex Lycii radicis, obtained from Hung Soon Pte Ltd, and Eurycoma longifolia (obtained from Thong Hup Gardens Pte Ltd, were homogenized in water (1:1 ratio). Debris was filtered using a muslin cloth and the filtrate was centrifuged at 8000 rpm for 10 min at 4°C. After centrifugation, the supernatant was concentrated, filtered and loaded on a C18-flash column. Increasing concentration of ethanol (20%-70%) was used for elution from the column. Presence of desired peptides in the eluents was confirmed by MALDI-TOF MS and these fractions were pooled and purified using several dimensions of RP-HPLC and SCX-HPLC. A linear gradient from buffer A (5% ACN, 20 mM KH_2PO_4; pH 3) to buffer B (5% ACN, 0.5 M KCl, 20 mM KH_2PO_4; pH 3) was used for SCX-HPLC. Fractions from SCX-HPLC that contained desired peptides were pooled and purified by RP-HPLC using buffer A (0.1% TFA in water) and buffer B (0.1% TFA in 100% ACN).

2.2.4 De novo sequencing with MALDI-TOF MS/MS

Approximately 50 µg of purified peptide was dissolved in 100 mM ammonium bicarbonate (NH_4HCO_3) buffer and was incubated with 50 mM dithiothreitol (DTT) at 37°C for 45 min to reduce disulfide bonds. The reduced disulfides were alkylated with iodoacetamide (IAM) (300 mM) or 4-vinylpyridine (4-VP) (300 mM) for 2 h at 37°C. Digestion of alkylated samples was done at room temperature for 5 min with trypsin, chymotrypsin, and endo-GluC digestion. *De novo* sequencing of peptides was performed based on both b- and y-ion series in the tandem mass spectrometry (MS/MS) profiles.

2.2.5 LC-ESI-MS/MS analysis

The reduced-alkylated peptide sample was desalted using Millipore ZipTips and lyophilized. The peptide was re-dissolved in 0.1% formic acid (FA) before MS analysis. A Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific, Bremen, Germany) coupled with an Orbitrap Elite mass spectrometer (Thermo Scientific Inc., Bremen, Germany) was used to perform LC/MS-MS analysis. Elution was performed over a 60 min gradient from eluent A (0.1% FA) to eluent B (90% ACN/0.1% FA). The LTQ Tune Plus software (Thermo Fisher Scientific, Bremen, Germany) was used to set the mass spectrometer to positive mode for data acquisition using. A Michrom's Thermo CaptiveSpray nanoelectrospray ion source (Bruker-Michrom, Auburn, USA) was used to generate the spray. A Full FT-MS (350-2000 m/z, resolution 60.000, with 1 µscan per spectrum) was alternated with Full FT-MS and a FT-MS/MS scan applying 27%, 30% and 32% normalized collision energy in high-energy collisional dissociation (110-2000 m/z, resolution 30.000, with 2 µscan averaged per MS/MS spectrum) for data

acquisition where three intense ions with a charge greater than +2 and a mass difference of 3 Da were isolated and fragmented. Source voltage of 1.5 kV and capillary temperature of 250°C were used. Automatic gain control was set to 1×106 for full scan-MS and MS/MS. PEAKS studio (version 7.5, Bioinformatics Solutions, Waterloo, Canada) was used to process data from LC-MS/MS analysis with parent error tolerance and a fragment error tolerance of 10 ppm and 0.05 Da respectively.

2.2.6 Disulfide Mapping of Peptide

The peptide lyba1 was partially reduced in 100 µL of 50 mM Tris(2-carboxyethyl)phosphine (TCEP) and 10% ACN at 65°C for 2 min. Subsequently, N-ethylmaleimide (NEM) was added to a final concentration of 250 mM and incubated at 37°C for 1 h. The intermediates were then fractionated with analytical RP-HPLC with a Vydac C18 column (particle size: 5 µm, 250 x 4.6 mm; GRACE, USA) at a flow rate of 0.3 mL/min with a gradient of 23-37% over 80 min with Buffer A (0.1% TFA) and buffer B (100% ACN, 0.1% TFA). Intermediate species were collected and analyzed using MALDI-TOF MS. Each intermediate species was then fully reduced with 50 µL of 20 mM DTT, 25 mM ammonium bicarbonate in 20% (v/v) ACN at 37℃ for 1 h and alkylated with 40 mM IAM at 37℃ for 2 h. Alkylated samples were then examined by MALDI-TOF/TOF and sequenced by tandem mass spectrometry.

2.2.7 Spectrophotometric determination of protein concentration

The Beer-Lambert law was used to calculate concentrations of purified peptides according to the equation

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

Where,

A : the absorbance at 280 nm of peptide solution in miliQ water measured on the Nanophotometer (Implemen, Germany)

ε : molar absorption coefficient ($M^{-1}cm^{-1}$)

l : cell path length (cm)

The theoretical ε value of a protein at 280 nm was calculated as follows [197]:

ε280 = (5500. *n*Trp) + (1490 . *n*Tyr) + (125 . *n*SS)

where *n*Trp : the number of Trp residues

*n*Tyr : the number of Tyr residues

*n*SS : the number of disulfide bonds

2.3 Structural analysis

2.3.1 NMR spectroscopy

To perform nuclear magnetic resonance (NMR) experiments, samples containing approximately 0.5-0.6 mM peptide were prepared in a 20 mM sodium phosphate buffer (pH 7.0) containing 50 mM NaCl and 0.01% NaN_3 in D_2O or 10% D_2O. Nuclear Overhauser effect spectroscopy (NOESY) experiments were acquired with 200 and 300 ms mixing times [198,199]. Total correlation spectroscopy (COSY) [200] data were recorded with a mixing time of 69 or 78 ms using MLEV17 spin lock pulses [201]. Vicinal coupling constants were measured using the double-quantum-filtered (DQF)-COSY [202] and one-dimensional (1D)-^{1}H-NMR experiments. All 2D-

NMR data were recorded in the phase sensitive model using the time-proportional phase increment method [203], with 2048 data points in the t2 domain and 512 points in the t1 domain. Slowly-exchanging amide protons were identified by immediate acquisition of a series of 1D-experiments after dissolving the lyophilized peptide in a D_2O solution. The water signal was suppressed using water-gated pulse sequences [204] or excitation sculpting [205] combined with pulsed-field gradients. All NMR data were processed using Bruker TOPSPIN 2.1 or NMRPipe [206] programs on a Linux workstation and analyzed using Sparky 3.12 software. DQF-COSY spectra were processed on 8192 x 1024 data matrices to obtain a maximum digital resolution for coupling constant measurements. Sodium 3-(trimethylsilyl)-1-propanesulfonate (DSS-d6) was used as internal reference.

2.3.2 Structure calculations

To determine the structure of the peptide eL1, 2D ^{1}H, ^{1}H-TOCSY and NOESY spectra were recorded using a Bruker 800 MHz NMR spectrometer (Bruker, IL, USA) equipped with a cryogenic probe in 298K. The peptide concentration was 1.2 mM. The sample contained 95% H_2O and 5% D_2O (pH3.5). The mixing times of TOCSY and NOESY experiments were 80 ms and 200 ms respectively. The carrier frequency of ^{1}H was at 4.735 ppm. The spectrum width was 12 ppm. The data was processed using the software NMRpipe. The analysis and assignment were done using the software Sparky 3.115. Structure calculation was conducted using the software CNSsolve 1.3. The distance restraints were obtained based on the intensities of NOE cross peaks, which were divided to 3 classes: strong, $0 < d \leq 1.8$ Å; medium, $1.8 < d \leq 3.4$ Å; and weak, $3.4 < d \leq 5$ Å. Hydrogen bond restraints were obtained from H/D exchange NMR experiment.

The lyophilized peptide was dissolved in 100% D_2O, and 1D NMR spectra were recorded after varied time. The last spectrum was recorded after 24 h. The distance between HN and O was defined as 2.2-0.6 Å, and the distance between N and O was defined as 3.3-0.8 Å. The structure was verified using the online server to generate the ramachandran plot (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and displayed using Chimera version 1.6.2.

The NMR sample of WCI was prepared by dissolving WCI in 95% H2O/5% D_2O or 99.9% D_2O directly (~ 1 mM protein and pH 6). All NMR experiments were carried out on a Bruker 600-MHz NMR spectrometer equipped with a cryogenic probe. Two-dimensional homonuclear ^{1}H-^{1}H TOCSY and NOESY spectra were acquired using the mixing times of 80 and 200 ms, respectively. The 2D data were acquired at 298 K. The NMR spectra were processed with NMRPIPE [207]. The amides involved in hydrogen bonding were identified by hydrogen-deuterium exchange (1D ^{1}H) experiments [89].

Sequence-specific assignments were achieved with 2D TOCSY and NOESY, and NMRSPY (http://yangdw.science.nus.edu.sg/Software&Scripts/NMRspy/index.htm) was used to assign NOEs from 2D NOESY results. Distance restraints were derived from the peak intensities of the assigned NOEs. Three-dimensional structures of WCI were reconstructed using CYANA 2.0 [208]. Structures were displayed and analyzed with PYMOL and PROCHECK-NMR respectively.

2.3.3 X-Ray Crystallographic Experiments

Purified lyba2 was dissolved in deionized water at 10 mg/mL. The sample was then used to perform crystallization experiments at 18°C with a Phoenix crystallization robot (Art Robbins, USA) employing the sitting-drop vapor-diffusion method, with mixing 0.3 µL of peptide and 0.3 µL of reservoir to equilibrate against 75 µL reservoir solution. Two commercially available screens were used: Index™ (Hampton Research, USA), and Structure Screen1/2 MD1-01/02 (Molecular Dimensions, United Kingdom). Tiny crystals of lyba2 were observed under condition No. 30 of MD1-01 (2.0 M ammonium sulfate, 0.1 M sodium HEPES pH 7.5, 2% PEG400). Optimized crystals were subsequently obtained within three months from 1.6 M ammonium sulfate, 0.1 M sodium citrate tribasic dihydrate pH 4.5, 2% PEG400) with mixing 2 µL of peptide and 2 µL of reservoir to equilibrate against 500 µL reservoir.

Optimized crystals were cryo-protected by 20% glycerol and flash-cooled in liquid nitrogen. A native dataset to 1.48 Å resolution was collected at a wavelength of 0.9184 Å, and a sulfur single-wavelength anomalous dispersion (S-SAD) data set to 1.95 Å was collected at a wavelength of 1.8000 Å, both at beamline P11 of the PETRA III storage ring (DESY, Hamburg). Data collection followed the recommendations for S-SAD phasing published by Weiss et al. [209]. Both diffraction data sets were processed with XDS [209,210]. The space group was $P2_1$, with unit-cell parameters a = 19.32 Å, b = 54.65 Å, c = 42.40 Å, β=100.60°. The structure of lyba2 was determined by the SAD method. 20 of 24 sulfur atoms were found using SHELXD [211]. The correct hand for the substructure was determined and a preliminary poly(Ala) chain was modeled with SHELXE [211]. This initial model was then rebuilt and refined using Coot [212] and

REFMAC 5 [213]. The refined structure was visualized using Pymol (Schrödinger; http://www.pymol.org/) and UCSF Chimera [214].

2.4 Stability Assays

2.4.1 Heat stability assay

Purified peptides (10 µg) was added to 100 µL of H_2O and incubated at 100°C for 1 h and then analyzed by UPLC. Short linear peptides RLYRRGRLYRRNHV (RV-14) synthesized in our laboratory served as controls and DALK, a thermostable peptide was used as an internal standard. Peaks from UPLC, at wavelength 220 nm and 254 nm, were analyzed by MALDI-TOF MS.

2.4.2 Proteolytic enzyme stability assay

Purified peptides (10 µg) were incubated with trypsin, pepsin and chymotrypsin (100 µL) at 37°C for 6 h at a final peptide to enzyme ratio of 20: 1(mol/mol). At each time point, 50 µL of the sample was injected in UPLC to assess degradation. RV-14 was used as controls.

2.4.3 Serum stability assay

Stability of isolated and purified peptides in human serum was assessed according to the protocol by Jenssen *et al.* [215]. Briefly, 25% human serum (Sigma Aldrich, Singapore) was prepared in PBS and incubated at 37°C for 15 min. Purified peptides (100 µg) were added to the temperature-equilibrated serum (300 µL) and incubated at 37°C. At specified time points, 50 µL aliquots of the sample was added to

100 µL of 95% ethanol and incubated at 4°C for 15 min to precipitate serum proteins. The sample was then centrifuged at 18,000 rpm for 2 min to pellet down the precipitated proteins. The supernatant (50 µL) was injected in UPLC and the chromatographs were analyzed to determine presence and extent of degradation. Linear peptides RV-14 was used as controls.

2.5 Bioassays

2.5.1 Antimicrobial Assays

The antifungal activity of eL1 was tested against *Alternaria alternata*, *Alterneria brassiciola*, *Curvularia lunata*, *Fusarium oxysporum* and *Rhizoctonia solani* using the disc diffusion assay as described by Ye *et al.* [216]. *A. niger* was grown on malt extract agar plates while the remaining fungal strains were grown on potato dextrose agar plates, at 25°C. When sufficient mycelial growth at 75% of the plate was observed, a hole was punched in the fungal culture and was transferred to a new agar plate and incubated for 48 h-72 h at 25°C till a radial mycelial colony was formed. Paper discs (6 mm) inoculated with 20 µl of peptides dissolved in MilliQ water (17.5 µg- 70 µg) were placed at the growing ends of the mycelia at an equal distance of 1 cm and incubated for 24 h at 25°C. Deionized water was used as a negative control. Formation of crescent-shaped inhibition zones indicated susceptibility of fungi to test peptides.

2.5.2 Microbroth dilution assay

The half maximal inhibitory concentration (IC_{50}) of peptides was assessed using the microbroth dilution assay [217]. Fungal spores were seeded in half-strength potato dextrose broth or half-strength malt extract at a final density of 2.5 X 10^3 cells/mL. To 80

µL of the spore suspension, 20 µL aliquots of different concentration of peptides were added and incubated at 25°C for 24 h. After incubation, the cells were fixed with 100 % methanol for 30 min followed by staining for 30 min with 1% (w/v) methylene blue in 0.01 M borate buffer. Water was used to wash off excess dye and elution was performed using 1:1 (v/v) ethanol/ 0.1N HCl. Absorbance was read at 650 nm using Infinite@ 200 PRO Tecan microplate reader (Tecan Group Ltd, Germany). Control wells were treated with half-strength media. Percentage inhibition was calculated as 100 times the ratio of absorbance of treated samples to control samples. A dose-response curve using the "Log-inhibitor vs response (variable slope)-four parameters" function was computed using GraphPad Prism 6 for Windows GraphPad Software, San Diego California USA, www.graphpad.com.

2.5.3 Radial Diffusion assay

Candida albicans and *Candida tropicalis* fungal strains were cultured in tryptic soy broth (TSB) at 37^{o}C. The anti-fungal activity of eL1 and WCI was examined with the radial diffusion assay as described by Lehrer *et al.* [218]. Briefly, the yeast phase of *C. albicans* were washed with sterile cold 10 mM sodium phosphate buffer (10 mL), pH 7.4 and resuspended in sterile cold sodium phosphate buffer (5 mL). The optical density of the subculture was measured to determine the concentration of yeast cells. 4 X 10^{6} CFU was added to 10 mL of molten underlay gel containing 10 mM of NaCl and poured into a 60 mm X 15 mm petridish. 3.2 mm diameter wells were punched into the underlay gel in a 4 X 3 array and 5 µL of different concentration of eL1 was added to each well. After the addition of peptide, the plates were inverted and incubated for 3 hours at 37^{o}C. 10 mL of molten overlay gel was then poured above the underlay gel and the plates

were incubated overnight. The diameters of the clear zones around each well were measured and a minimum inhibitory concentration (MIC) was calculated.

2.5.4 Cytotoxicity assay

PrestoBlue™ Cell Viability Reagent (Invitrogen) was used to test the cytotoxicity of purified peptides. Vero (African green monkey kidney cells), Huvec-CS (Human umbilical vein endothelial cells) and A549 (adenocarcinomic human alveolar basal epithelial cells) were seeded at a density of 5 X 10^4 cells/mL in a 96-well plate and incubated with 100 µM of purified peptides for 24 h at 37°C. After incubation, PrestoBlue™ reagent (10 µl) was added to the wells followed by incubation for 2 h at 37°C. Fluorescence was measured as prescribed by the manufacturer. 1% triton-X100 was served as the positive control.

2.6 EST-Based data mining

2.6.1 Translated nucleotide based search for putative cysteine-rich peptides

Database searches were conducted using methods modified from literature [25,219]. TBLASTn was employed to search for ESTs encoding putative CRP precursors in two databases, the National Center for Biotechnology Information (NCBI) [220] and the 1000 Plants Project (OneKP), via queries using eL1, WCI, lyba1 and lyba2 peptide sequences [221]. The maximum target sequence and expected threshold were set at 1000 and 10, respectively. The accession numbers of all the putative cysteine-rich peptide sequences are provided in the appendix E. The sequences were

manually selected based on the following criteria: (1) The open reading frame must contain a stop codon following the C-terminal tail; (2) The translated amino acid sequence must not contain any untranslated amino acid residues, such as X; and (3) The mature peptide must contain six, eight or ten cysteine residues. The sequences were then submitted to SignalP 4.0 [222] for identification of signal peptides and were aligned using ClustalW [223]. Replicate sequences from same plants or different species of the same genus sharing an identical full-length precursor were deleted from the dataset.

2.6.2 Data analysis

The stability, cytotoxicity and antimicrobial data were analyzed by Student's t-test. The results were expressed as the mean ± standard error of the mean (SEM), where p-values less than 0.05 were considered statistically significant. Aligned sequences were visualized using Weblogo [196]. The phylogenetic tree was constructed and annotated using the online tool called, Interactive tree of life (iTOL) [224].

Chapter 3 Identification and characterization of a carboxypeptidase inhibitor from the medicinal herb *Lycium barbarum* (wolfberry)

3.1 Introduction

Plants have developed many survival strategies that involve a variety of molecules to defend against pathogenic infections, herbivore attacks, and environmental stress [225,226]. However, most earlier research efforts largely focused on small molecules and secondary metabolites, and ignored larger molecules such as peptides and proteins. There is recent evidence showing that plant CRPs, in particular those associated with protease inhibition, play an important role in plant defense functions [20,227]. These CRPs are widely expressed in plants and also participate in signal regulation and growth [228]. Protease inhibitors in plants typically exhibit insecticidal functions, and can inhibit important digestive enzymes of insects to impair their nutrient uptake [229,230]. Other CRPs such as cyclotides, defensins, thionins, snakins, and hevein-like peptides are found to be commonly involved in defending plants against microbes and biotic stresses [231,232]. CRPs also display a wide range of pharmacological properties, and have immunomodulatory, anti-cancer, and uterotonic effects [88,91,233]. However, the therapeutic importance of plant CRPs isolated from medicinal herbs has not been explored.

Metallocarboxypeptidases (MCPs) are enzymes that hydrolyze C-terminal residues of peptides and proteins. MCPs plays a role in a wide variety of physiological processes including digestion [234], blood coagulation/fibrinolysis [235], neuropeptide

processing [236], inflammation [237], and cancer progression [238]. The activities of MCPs can be modulated by exogenous carboxypeptidase inhibitors (CPIs) [239] that are found in many organisms such as plants [240], leeches [241], parasites [242], and marine annelids [243]. CPIs vary in size from 4 to 18 kDa, with the smallest inhibitors isolated from plants of the Solanaceae family.

The first and also the smallest proteinaceous CPI was isolated from potato 40 years ago [240]. Since then, eight other exogenous proteinaceous CPIs, of which only one is from plants, have been isolated and characterized (Table 3.1) [82,240]. The prototypic potato and tomato CPIs are small in size (39 residues) with three disulfide bridges stabilizing the structure forming a cystine knot, where the third disulfide bond pierce through a ring formed by the two other disulfide bonds and the intervening backbone segments [20,69]. This knotted arrangement generally confers high resistance to thermal, acidic and endoproteolytic degradation [65,75].

Plant CPIs inhibit MCP activity through a mechanism that mimics substrate binding, wherein the C-terminal tail interacts with the groove of the carboxypeptidase active site [244]. The ability of these CPI peptides to inhibit the activity of carboxypeptidase B (thrombin activatable fibrinolysis inhibitor) [245] could be used as fibrinolytic agents to treat thrombotic diseases [246-248]. Moreover, plant CPIs can affect inflammation [237], tumor progression [238], and angiotensin processing [249], making them possible therapeutic agents to treat a variety of diseases involved these pathways.

Table 3.1. Summary of natural occurring carboxypeptidases inhibitors discovered and characterized. As below, no other carboxypeptidases inhibitors for plants were reported after the two carboxypeptidases inhibitors discovered in potato and tomato.

Natural occurring carboxypeptidase inhibitors	No. of amino acids	Organism	References
Potato carboxypeptidase inhibitor	**39**	**Potato**	**Ryan CA. *et al*., 1974**
Tomato carboxypeptidase inhibitor	**37**	**Tomato**	**Hass& Ryan *et al*., 1980**
Leech carboxypeptidase inhibitor	**66**	**Leech**	**Reverter D. *et al*., 1998**
Tick carboxypeptidase inhibitor	**75**	**Ticks**	**Arolas JL. *et al*., 2005**
***H. Longicornis* carboxypeptidase inhibitor**	**77**	**Ticks**	**Gong H. *et al*., 2007**
***A. Suum* carboxypeptidase inhibitor**	**67**	**Parasite**	**Sanglas L. *et al*., 2009**
***N. Versicolor* carboxypeptidase inhibitor**	**53**	**Marine Mollusk**	**Covaleda G. *et al*., 2012**
***S. Magnifica* carboxypeptidase inhibitor**	**165**	**Marine Annelid**	**Alonso-Del-Rivero M. *et al*., 2012**
Latexin	**25 kDa**	**Human**	**Pallares I. *et al*., 2005**

The fruit of *Lycium barbarum*, also known as wolfberry or Goji berry, is a well-known medicinal ingredient that has been used for more than 2300 years in TCM [250]. Wolfberry has since gained popularity in Europe and America as a heath supplement and anti-aging remedy [157]. The fruits and root bark of wolfberry are commonly used in TCM, although leaves and seeds also reportedly have anti-inflammatory activity [251]. According to TCM theory, wolfberries are used as a mild tonic for the liver, kidneys, and lungs and can be used to treat blurry vision, infertility, abdominal pain, headache, fatigue, and dry cough [251-253]. Polysaccharides, a complex mixture of highly branched polysaccharides and proteoglycans, represent one of the most abundant components of wolfberry, with a yield of 23% (w/w) based on dried fruit extraction [158]. Another major group of metabolites in wolfberries are carotenoids, the levels of which increase during fruit ripening [159]. The fruits also contain several vitamins, including riboflavin, thiamin, and ascorbic acid (~42 mg/100 g dried mass) [160,161]. Non-proteinogenic amino acids identified in these fruits include taurine, γ-aminobutyric acid, and betaine [254,255]. Flavanoids [162] as well as essential oils and fatty acids [163] have also been found in wolfberry extracts. The antioxidant properties of wolfberry extracts have mainly been attributed to these polysaccharide [169,170] and flavonoid [162] components, although betaine could also contribute to the antioxidant activity of this fruit [256]. Furthermore, polysaccharides of wolfberry extracts have been shown to lower plasma cholesterol [171], improve insulin sensitivity [172], stimulate the immune system [173-175], and display anti-cancer properties [176-178].

This chapter describes the identification and characterization of a novel plant CPI, the wolfberry carboxypeptidase inhibitor (WCI), from an aqueous extract of the *Lycium*

barbarum fruit. At 37 amino acid residues, WCI is currently the smallest member of the CPI family, but the amino acid sequence of WCI is highly homologous to the CPI from potato (PCI). WCI contains 6 cysteine residues, which are arranged as a robust cystine knotted conformation with three disulfide bonds. The presence of this knotted arrangement may enable WCI to resist thermal, acidic, and endopeptidase degradation. Interestingly, WCI is also resistant to exopeptidases, including both amino and carboxypeptidases. We describe the NMR structure of WCI and also performed a tBLASTn search of the National Center for Biotechnology Information (NCBI) and 1kp-project (www.onekp.com) database. This search identified 26 additional homolog sequences from 19 plant species, which greatly expands the number of plant CPIs from the current 3 to 29. The discovery of the wide occurrence of CPIs among plant species can further our understanding of their sequence diversity, and enables their development as therapeutics for various diseases.

3.2 Results and Discussion

3.2.1 Screening for natural peptidyl products in *L. barbarum*

To screen for the presence of CRPs, *L. barbarum* fruits were ground in liquid nitrogen, extracted with distilled water and semi-purified by C18 SPE. MS analysis of these extracts showed that *L. barbarum* expressed a peptide with m/z 3991 (Figure. 3.1A). To confirm that this peptide is a CRP, it was reduced by dithiothreitol and alkylated by IAM. Since each disulfide bond present in a peptide causes a mass shift of 116 Da (2 x 58 Da), the mass increase of 348 Da indicated that this peptide included three disulfide bonds and six cysteines (Figure. 3.1B).

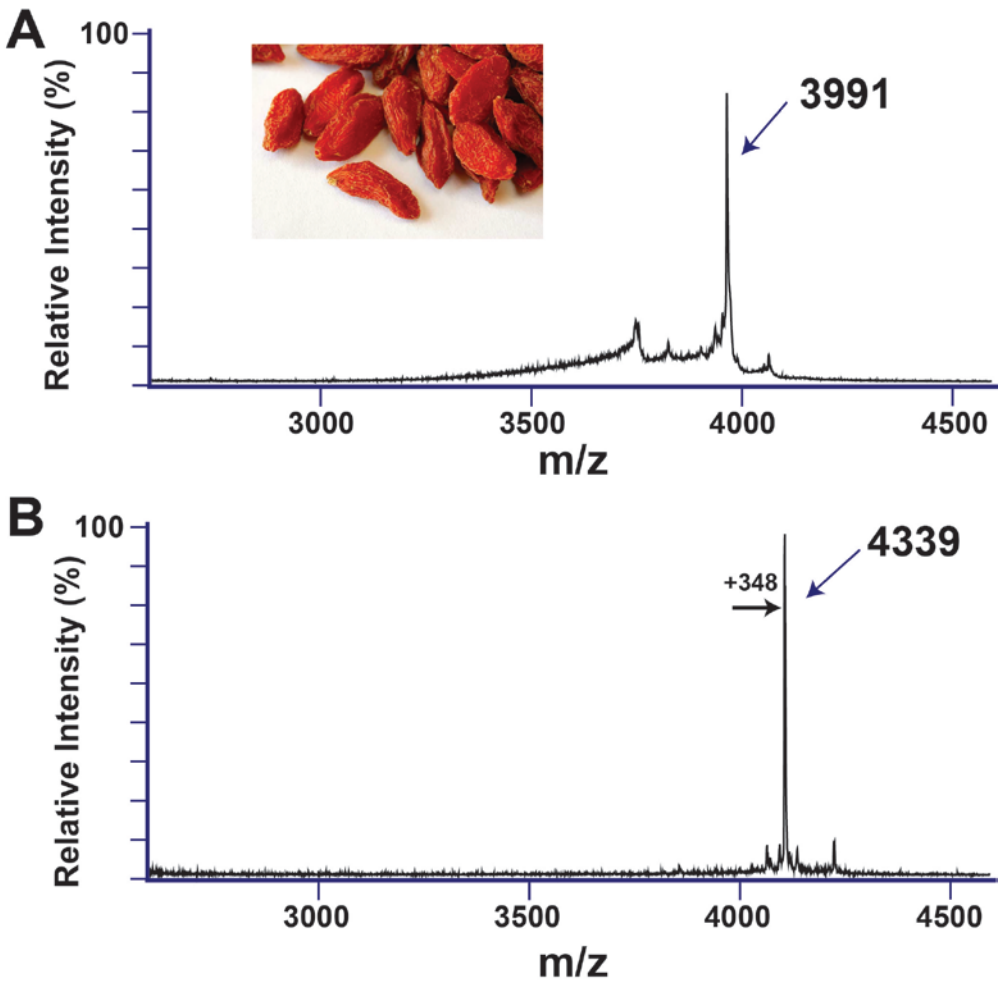


Figure 3.1 MS profiles of wolfberry aqueous extract. A) MS profile of a truncated WCI isolated from wolfberry. The truncated WCI is missing a glycine residue at the C-terminal compared to the full WCI amino acid sequence obtained from the transcriptome. B) Reduction and alkylation with dithiothreitol and Iodoacetamide respectively. The shift of +348 Da on the mass spectra shows that the peptide isolated from the water extract of wolfberry contains 3 disulfide bonds.

3.2.2 Isolation and sequence determination of a novel carboxypeptidase inhibitor from *L. barbarum*

To determine the sequence of the putative peptide and to obtain enough samples for amino acid sequencing, a larger scale extraction of 500 g *L. barbarum* fruit was conducted. Following multiple rounds of RP-HPLC, a peptide with m/z 3991 was isolated. The isolated peptide was *S*-reduced with DTT following digestion with trypsin or chymotrypsin without *S*-alkylation.

Tryptic digestion of the reduced peptide generated two fragments with m/z values of 3265 and 751, while chymotrypsin digestion generated two fragments with m/z values of 2907 and 1108. Tandem mass spectrometry was then used to sequence these enzyme-generated fragments, which revealed the peptide amino acid sequence as pQDDPTCGKPCNTMDDCSNGWFCQACWNSRKTCGPFV where a pyroglutamyl residue is located at the N-terminus (Figure 3.2). This novel peptide, named wolfberry carboxypeptidase inhibitor (WCI), shared 70% sequence homology with the potato carboxypeptidase inhibitor (PCI) and is one of the smallest CPIs isolated to date. The sequence was confirmed with a tBLASTn search of the EST database in the 1kp-project (www.onekp.com). Although an additional glycine was found at the C-terminus in the precursor sequence, this residue was absent from the aqueous extract of wolfberry fruit.

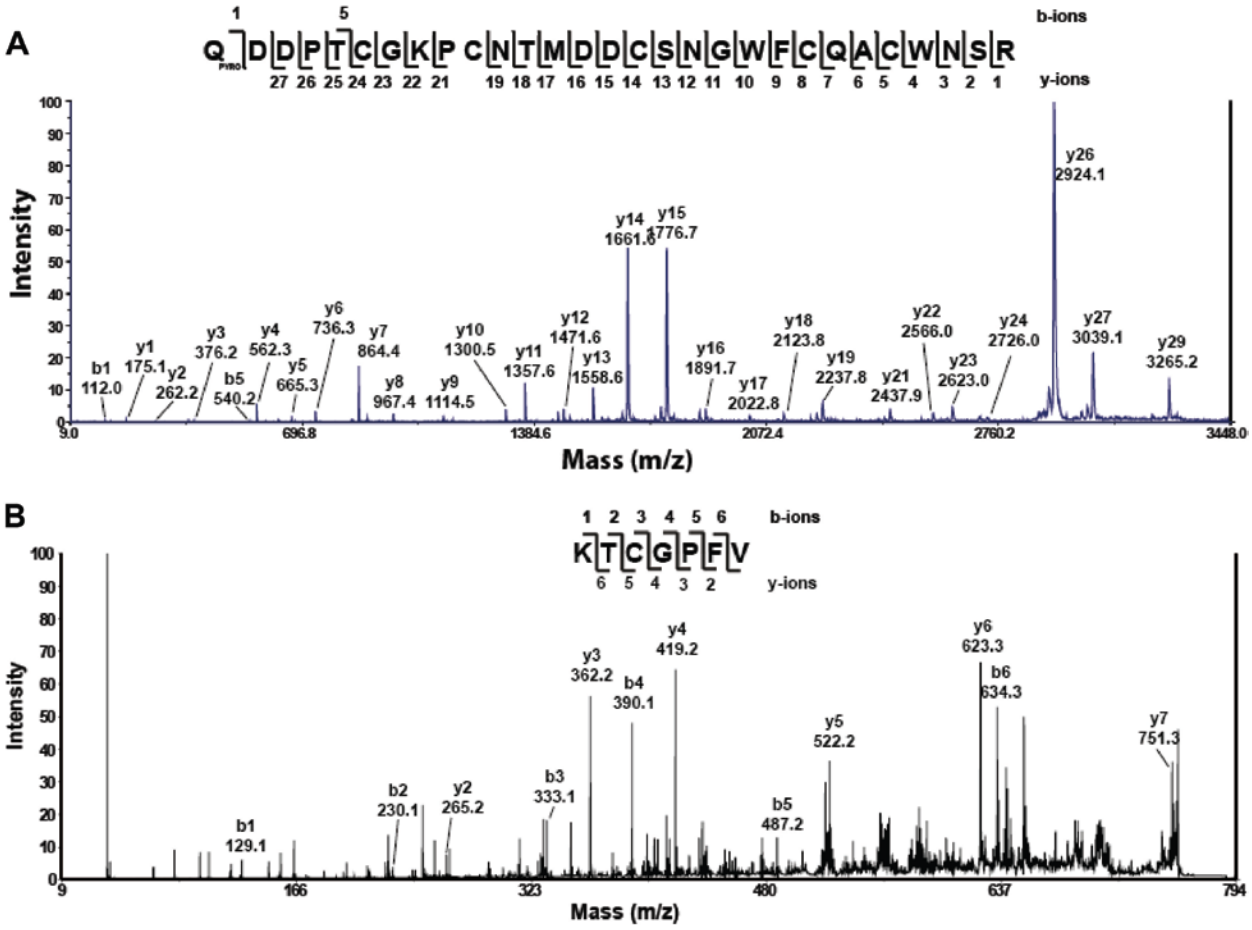


Figure 3.2. MS/MS sequencing of WCI. Trypsin digestion of WCI produces two fragments. **A.** MS/MS spectrum of the 3265 Da fragment. **B.** MS/MS spectrum of the 751 Da fragment. Combining both sets of sequences gave the overall sequence of WCI as pQDDPTCGKPCNTMDDCSNGWFCQACWNSRKTCGPFV.

3.2.3 Inhibitory activity of Wolfberry Carboxypeptidase Inhibitor

The equilibrium dissociation constant for the WCI and carboxypeptidase A complex was determined with a Burke-Lineweaver plot by varying the inhibitor concentration while maintaining a fixed enzyme concentration against increasing concentrations of substrate. WCI is a competitive inhibitor of MCP with a Ki value of 6.62 nM, which is comparable to that of PCI (Ki = 5 nM) (Figure 3.3). The WCI and PCI sequences showed conservation of 27 amino acid residues. Notably, the C-terminal tail amino acid sequence is highly conserved with the CGPFVG sequence in PCI. This conservation is consistent with the finding that the C-terminal tail of PCI is the primary binding site that governs the interaction between the peptide and the carboxypeptidase A active site [84]. Moreover, the equivalent of Val38 in the PCI C-terminus, which the X-ray structure showed to be a key residue for the inhibitory activity of PCI towards carboxypeptidase [257], is also present in WCI. Meanwhile, Trp22, Phe23, and Trp28 in the highly conserved sequence of WFCQACWN in WCI likely play an important functional role in the secondary binding site of this peptide, as suggested by a previous study that examined the effect of chemical modification of equivalent residues in PCI [258]. One of the main sequence differences between WCI and PCI is the replacement of the C-terminal Tyr with Phe. However, as suggested by results from a study by Cristina et al. showed that substitution of the C-terminal Tyr of PCI with Phe did not significantly affect the Ki [84].

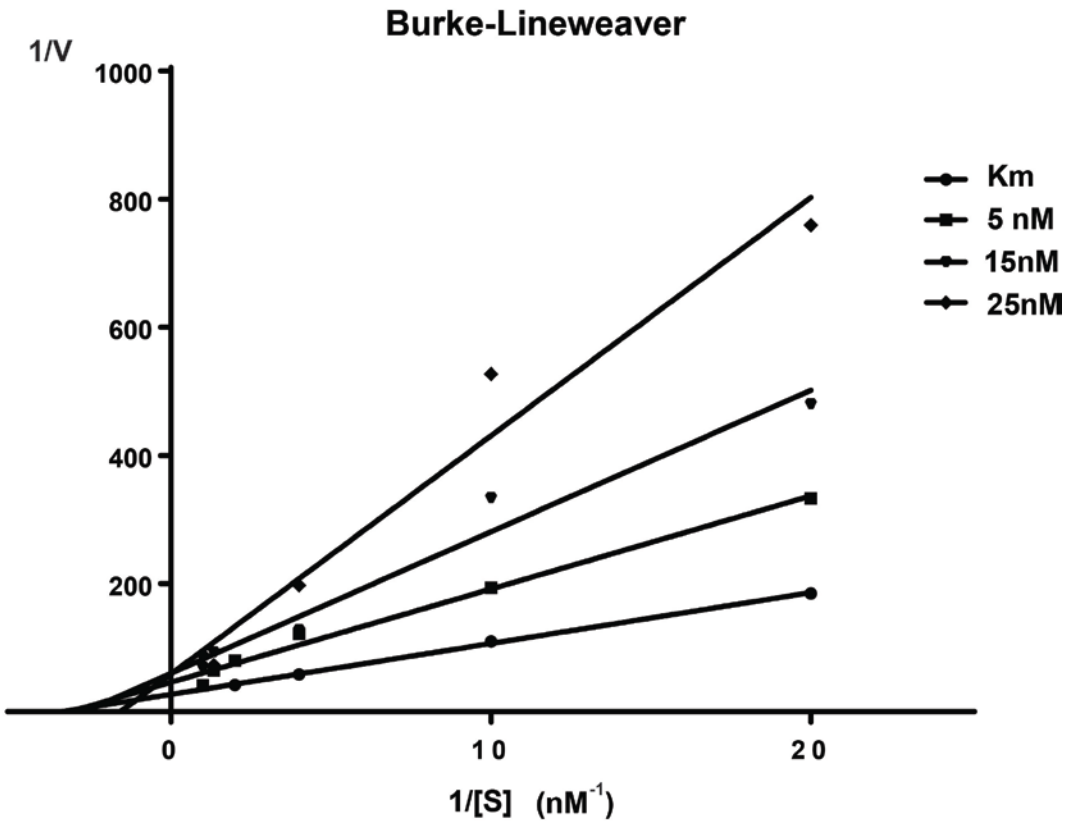


Figure 3.3. Lineweaver-Burke plot of the inhibition kinetics of WCI and bCPA on N-(4-methoxyphenylazoformyl)-phenylalanine. The plot shows that WCI is a competitive inhibitor of bCPA similar to other carboxypeptidases inhibitors isolated from other organisms. The K_i was determined to be 6.62 nM.

3.2.4 NMR analysis of WCI

To determine the 3D structure of WCI, a 2D NMR analysis was performed. From the distance, dihedral angle, and hydrogen bond restraints derived from the 1H-NMR experiments (Table 3.2), the WCI solution structures show that this peptide adopts a cystine knot scaffold that is similar to PCI in that it has the same disulfide connectivity: CysI-IV, CysII-V and CysIII-VI, where CysIII-VI is the disulfide bond the pierces the ring formed by the backbone and the remaining disulfide bonds. The overall structure also shows high homology with PCI, which is made up of a globular core and a C-terminal tail. The similarity of the structure and sequence of WCI and PCI suggests that the C-terminal tail of these peptides is not highly mobile, but rather maintains a preferred orientation to facilitate carboxypeptidase binding, as can be seen in a superposition of WCI on PCI in contact with carboxypeptidase A (Figure 3.4).

Table 3.2. NMR Experimental and structural statistics of WCI.

NOE constraints	400
Intraresidue ($\|i - j\| = 0$)	159
Sequential ($\|i - j\| = 1$)	117
Medium-range ($1 < \|i - j\| < 5$)	36
Long-range ($\|i - j\| \geq 5$)	88
Dihedral angle restraints	3
Hydrogen bonds	15
PROCHECK-NMR Ramachandran plot (%)	
Most favored region	46.4
Additionally allowed region	50.0
Generously allowed region	3.6
Disallowed region	0
Average maximum violations per structure	
Distance (Å)	0.037 $\pm$ 0.0002
Van der waals (Å)	3.5 $\pm$ 0.1
CYANA target function value (Å2)	1.52 $\pm$ 0.0095
Average rmsd to mean structure (Å)	
All backbone atoms (1–30)	0.26 $\pm$ 0.10
All heavy atoms (1–30)	0.64 $\pm$ 0.08

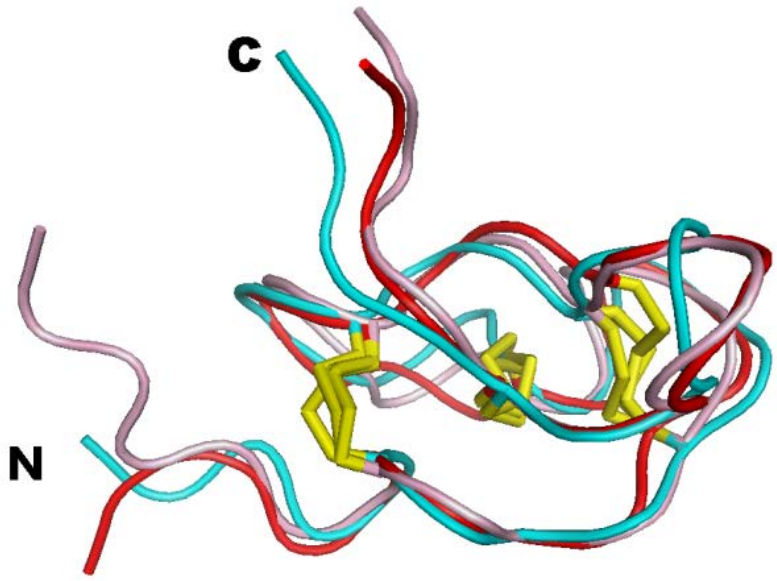


Figure 3.4. NMR 3D structure of WCI. The 3D structure of WCI was solved by NMR and found to be homologous to the potato carboxypeptidase inhibitor. Superposition of WCI and PCI is shown here. Both structures show a compact core with three disulfide bonds arranged into a cystine knot motif. The ring formed by Cys I-IV and Cys II-V was penetrated by Cys III-VI. Both structures contain a C-terminal tail which is the primary active site in the inhibition of carboxypeptidase A.

3.2.5 WCI stability

To examine the metabolic stability of WCI, the peptide was subjected to heat and acid treatment, as well as tryptic digestion treatment. WCI is stable to acid and heat at 100 °C for 1 h, when more than 95% of WCI remained intact (Figure 3.5). The WCI peptide is able to resist tryptic digestion up to 3 h, where >95% of the peptide remains intact.

Similar to other CRPs [20], the compact structure conferred by the cross-linking of the 3 disulfide bonds increases the stability of WCI against thermal, acid, and enzymatic degradation. Furthermore, the presence of an N-terminal pyroglutamate has been shown to protect peptides against degradation by aminopeptidases [259], and this residue together with the carboxypeptidase inhibiting C-terminal tail likely contributes to the ability of WCI to resist exopeptidase degradation. Although WCI is a linear peptide, its stability against endopeptidases and exopeptidases is comparable to pseudocyclic peptides such as wrightides and allotides, where the disulfide bonds at the ultimate or penultimate residues protects both the N- and C-terminus [20,260], and cyclic peptides from the cyclotide family [227].

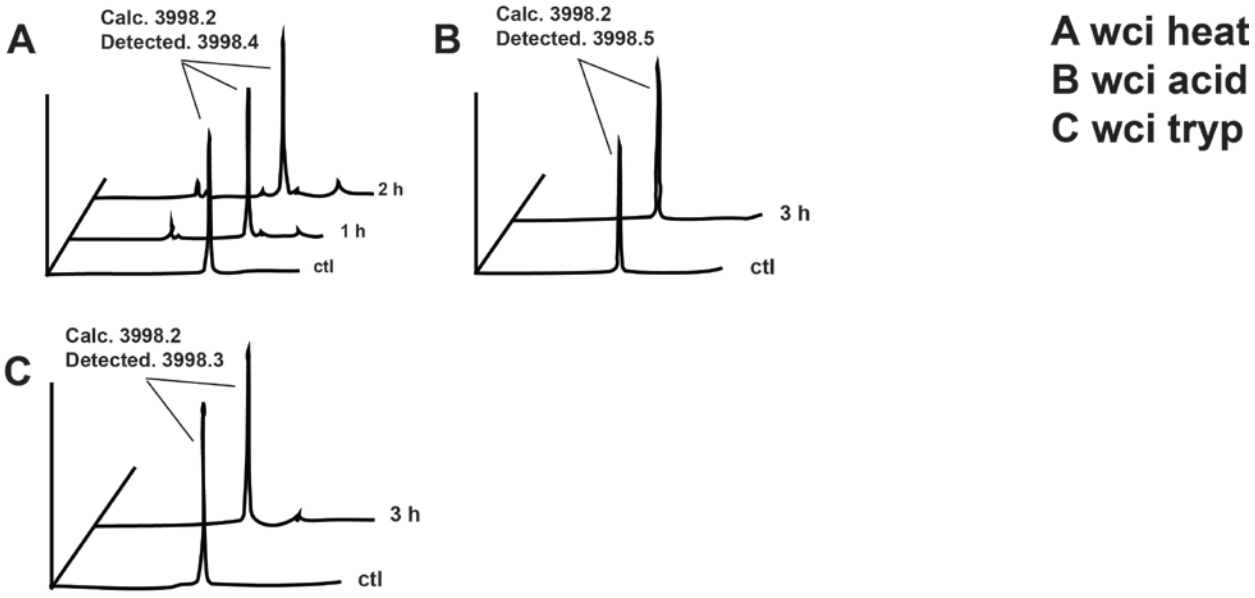


Figure 3.5. Thermal, acidic and enzymatic stability of WCI. (A) RP-HPLC profile of WCI heated at 100°C for 1 h and 2 h. Control was kept at 37°C. Comparison of the area under the peak of the chromatogram revealed that 95% of WCI remained after heating. (B)**.** RP-HPLC profile of WCI in 0.2 M HCl. Control was kept in ddH_2O. Comparison of the area under the peak of the chromatogram revealed that 98% of WCI remained after 3 h incubation. (C) RP-HPLC of WCI incubated with 1 ng/µl trypsin in 100 mM ammonium bicarbonate buffer (pH 7.8). Comparison of the area under the peak of the chromatogram revealed that 99% of WCI remained after 3 h incubation.

3.2.6 WCI antimicrobial activity

The antimicrobial activity of WCI was tested on two different bacterial and two different fungal strains, including *E. coli*, *S. aureus*, *C. albicans*, and *C. tropicalis*. However, no observable antimicrobial activity was detected against these four microbes at peptide concentrations up to 100 μM.

3.2.7 Identification of additional plant CPIs.

Using the sequence of WCI as a query sequence to search the EST databases of NCBI GenBank and the 1kp-project (www.onekp.com), we identified an additional 26 homologous peptide sequences across 19 different plant species from two families, Solanaceae and Asteraceae. Table 3.3 lists the 29 precursor peptide sequences from various plant sources aligned using a similarity alignment approach (ClustalW2).

Table 3.3. Summary of precursor peptides of putative carboxypeptidase inhibitors from expressed sequence tag database in NCBI. Carboxypeptidase inhibitors in *S. tuberosum* and *S. lycopersicum* were reported by Hass G.M. *et al.*

These peptides share many common molecular characteristics. First, the precursor protein and mature peptides of the plant CPIs have similar lengths (ranging between 34 and 39 amino acids for the mature peptides, Table 3.4). Second, they share a conserved cysteine arrangement and a three disulfide bond motif. Third, these peptides contain a short five amino acid C-terminal tail in the mature peptide that can interact with the active site of the carboxypeptidase. This short C-terminal tail plays a crucial role in the inhibitory action of these peptides, presumably because this region can insert into the groove of the enzyme and remain bound to the protein after the C-terminal residue of the peptide is cleaved by the target enzyme [84]. Based on these molecular characteristics, the peptides that show sequence similarity to WCI may also exhibit carboxypeptidase inhibition activity.

The cysteine residues in PCI, tomato carboxypeptidase inhibitor, WCI, and the putative CPIs identified from the tBLASTn search are highly conserved. They contain a cysteine motif, the cystine knot, which belongs to a class of peptides known as knottins. This cystine knot, or T-knot, scaffold is shared with other peptides such as cyclotides, plant trypsin inhibitors, and ω-conotoxin [261].

Table 3.4. Summary of peptides that are homologous to WCI identified from expressed sequence tag database in NCBI and 1kp project (www.onekp.com). The sequence alignment was prepared using bioedit with Clustal W2.

Peptides	Amino Acid Sequences	Plant Species
WCI*	QDD--PTCGKPCNTMDDCSNGWFCQACWNSRKTCGPFV-	*Lycium barbarum*
WCI	QDD--PTCGKPCNTMDDCSNGWFCQACWNSRKTCGPFVG	*Lycium barbarum*
PCI	QQHADPICNKPCKTHDDCSGAWFCQACWNSARTCGPYVG	*Solanum tuberosum*
PCI2	QEDQDPICGKPCNTHDDCSDAWLCQACWNFRKTCGPFVG	*Solanum tuberosum*
ToCI	QQ-YDPVCHKPCSTQDDCSGGTFCQACWRFAGTCGPYVG	*Solanum lycopersicum*
ToCI2	QEDQDPVCGKPCTTHDDCSEAWLCQACWNFRQTCGPFVG	*Solanum lycopersicum*
DMCI1	QEDQDPTCGKHCNTMDDCADGWFCQACYNFRKTCGPFVS	*Datura metel*
DMCI2	QEDQDPTCGKPCNTHDDCSGGWLCQACWNFRKTCEPFVG	*Datura metel*
LCCI1	QND--PTCGKSCNTMADCSEGWFCQACWNTPKTCGPFVR	*Lycopersicon cheesmanii*
SDCI1	QEDQDPTCGKPCTTMDDCSEGWFCQACWNFRQTCGPFVG	*Solanum dulcamara*
SDCI2	QYNRDPTCDKSCNTHADCSGGWFCQACWNSKKTCGPYVG	*Solanum dulcamara*
SLCI1	QQ-YDPVCHKPCSTQDDCSGGTFCQACWRFAGTCGPYVG	*Solanum lasiophyllum*
SLCI2	QV--DPVCYKPCKTHDDCSGALYCEACRRAAGSCGPWRS	*Solanum lasiophyllum*
BSCI	QEDQDPTCGKPCNTMVDCADGWFCQACYNSRKTCGPYVH	*Brugmansia sanguinea*
SPCI1	QEDQDPTCGKPCNTMDDCSEGWFCQACWNSRQTCGPFVG	*Solanum ptychanthum*
SPCI2	QDD--PTCGKPCNTFDDCSGGWYCQACWTVPKTCGPLVR	*Solanum ptychanthum*
SSCI	QND--PTCGKPCNTMADCSEGWFCGACWNTPRTCRPFVR	*Solanum sisymbriifolium*
SXCI1	QAD--ATCGKPCNTMADCSEGWFCQACWNTPKTCGPFVR	*Solanum xanthocarpum*
STCI1	QDD--PTCGKSCNTMADCSEGWFCQACWNTPKTCGPFVR	*Solanum torvum*
STCI2	QDD--PTCGKSCNTMADCFEGWFCQACWNTPKTCGPFVR	*Solanum torvum*
SMCI1	QND--PTCGKSCNTMADCSEGWFCQACWNTPKTCGPFVR	*Solanum melongena*
SMCI2	HND--PVCGKSCNTMADCSEGWFCQACWNTPKTCGPFVR	*Solanum melongena*
SCCI	QRDPDPICNKPCKTHDDCSGAWFCQACWNSARTCGPYVG	*Solanum chacoense*
SHCI	QQ-YDPVCNKPCSTQDDCSGGTFCQACWRFAGTCGPYVG	*Solanum habrochaites*
NCI	QEQYDATCGKPCNTRDDCSKGWLCSECYNFRKTCGPLIG	*Nicotiana tabacum*
NCI2	QEQYDATCGKPCNTRDDCSSGWLCSECYNFSKTCGPLVG	*Nicotiana tabacum*
NCI3	QEQYDATCGKPCNTRDDCSKGWLCSECYNFRKTCGPLIG	*Nicotiana tabacum*
NSCI1	QEQYDATCGKPCNTRDDCSKGWLCSECYNFRKTCGPLIG	*Nicotiana sylvestris*
WSCI	QY--DPTCETPCNTHSDCVRGWLCQACWNSKKTCNPYVT	*Withania somnifera*
AACI	QY--DPTCGTPCNTHSDCVRGWLCQACWNSKKTCNPYVT	*Artemisia annua*

In addition to the cysteine residues, the C-terminal tail is also highly conserved among the identified CPIs. Of the five residues that comprise the tail, the highly conserved Val is a key residue that plays an important role in binding to carboxypeptidase A [257]. Val38 establishes two hydrogen bonds and a coordinate bond with the Zn^{2+} atom of the carboxypeptidase, thus stabilizing the entire enzyme-inhibitor complex after cleavage at PCI Gly39 [244]. Of interest are the CPIs that were identified in plants from the Nicotiana genus. In the NCI, NCI3, and NSCI peptides, Val38 is replaced with Ile, which has a similar side chain structure. In fact, a mutagenesis study by Miguel et al. showed that the replacement of Val with Ile does not affect the Ki of PCI since the β and γ-methyl groups, which are in the appropriate position to be buried in the active site of carboxypeptidase A, are present in both residues [257].

Gly39 and Pro36, which do not interact with carboxypeptidase directly but are important to maintain the rigidity and orientation of the C-terminal tail, are also highly conserved among carboxypeptidase inhibitors. The absolute conservation of Pro36 in the 29 sequences highlights the evolutionary importance of this residue to CPI function. Moreover, replacement of Pro with Gly at this position resulted in a Ki for CPA that is 100-fold higher than the wild type peptide.

The third residue of the C-terminal tail displayed strong preference for aromatic residues. The amino acid at this position does not play a part in inhibitor binding, but rather performs a key role in interacting with the enzyme after the C-terminal residue is cleaved. Indeed, Marino-Buslje et al. showed that the third amino acid of PCI can be replaced with Phe or Tyr without affecting the Ki, as is reflected in the conservation of

Phe or Tyr, with 51.7% and 31.0% respectively, of the identified peptides having one of these residues in that position [84].

The C-terminal residue usually acts as a substrate, allowing the enzyme to cleave off when the inhibitor interacts with the enzyme. Most CPIs have Gly as the C-terminal residue, although about a third of the putative CPI sequences have Arg instead. Since the last residue acts as a substrate, having Arg at this position may allow the inhibitor to preferentially interact with carboxypeptidase B, which has a substrate preference for basic residues at the C-terminus. However, it was shown with Phe in PCI that for preferred substrates of carboxypeptidase A, having a more suitable substrate may not result in better binding as there may be interference in the docking of the C-tail into the active site [84]. As seen in PCI, LCI, and TCI, peptides that lack the C-terminal residue have virtually identical inhibitory activity to the full-length peptide, and this finding may explain the higher variability of the C-terminal amino acid residues among the putative CPIs [84,241,262].

In addition to the high conservation of the C-terminal tail, Trp22 and Trp28 are also conserved among 86% and 76%, respectively, of the putative CPI sequences. These residues appear at the secondary binding site of the peptides and are involved in the interaction between the inhibitor/enzyme complexes. The side chain of PCI Trp22 interacts with Phe277 and Arg276 of carboxypeptidase A (CPA), while Trp28 interacts with CPA residues Tyr198, Ser199, and Leu202. Among the putative CPIs, 20.6% of the sequences replace Trp with Tyr at this residue; the conservation of the aromatic amino acid may thus allow the inhibitor to maintain interactions with carboxypeptidases. The amino acid at position 23 is also highly conserved as hydrophobic amino acids Phe,

Tyr, and Leu. The high conservation of a hydrophobic amino acid in this region may result in large hydrophobic patches on the peptide surface, stabilized by disulfide bonds that is a common feature of cystine knot peptides [158]. This hydrophobic patch, which is the secondary binding site of CPIs, together with the C-terminal tail, completes the carboxypeptidase inhibitor domain of the plant CPI peptide family (Table 3.3).

Another amino acid residue that is highly conserved among the identified inhibitor sequences is Gln or Glu at the N-terminus. The N-terminal Glu/Gln can spontaneously convert [74] into pyroglutamic acid [263,264], which confers resistance against degradation by aminopeptidases [265].

Although multiple amino acid residues are involved in the function and structure of the CPI peptides are well conserved, some regions carry a variety of amino acids (Figure 3.6). This sequence diversity may indicate that, like social insects, the immune system of plants may respond differently to selection pressure caused by environmental microbes and pathogens [266]. This line of thinking is supported by insect carboxypeptidases that are resistant to specific plant CPIs [267]. This high sequence variability could allow CPI peptides to be used as templates for further bioengineering modification to confer additional functions.

The analyses of sequences that were identified in the transcriptome databases allowed us to predict that these peptides may also have carboxypeptidase inhibition activities, thus suggesting their role as putative CPIs. This study greatly expanded the number of CPIs identified in plants from the current 3 to 29. Moreover, the widespread presence of CPIs in plants of the Solanaceae family indicates that these peptides may

play an important role in plant self-defense mechanisms, and could protect plants from insect attack by inhibiting the digestion enzymes found in the gut of insects, that in turn restricts nutrient uptake [268].

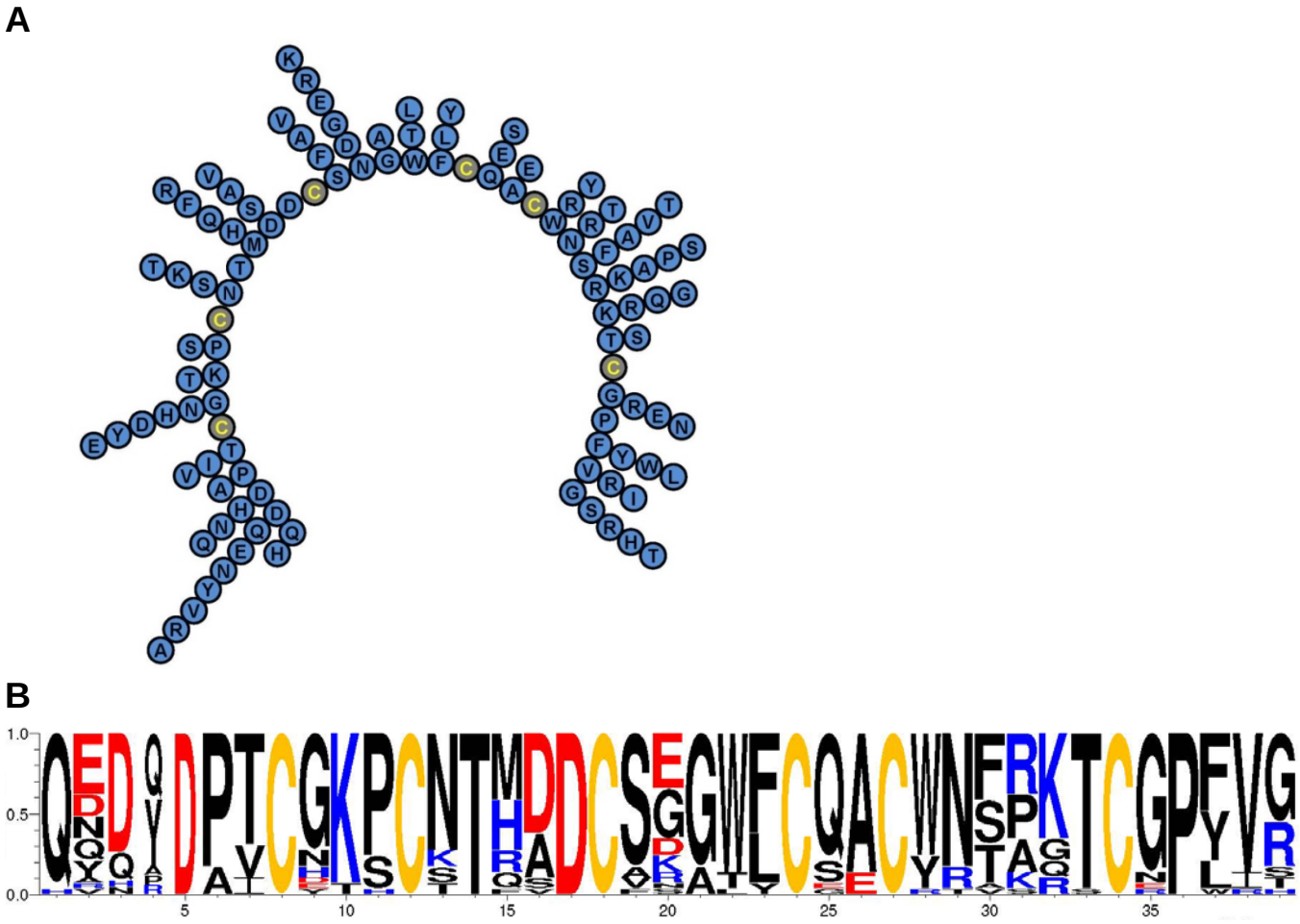


Figure 3.6. Sequence diversity of putative plant carboxypeptidase inhibitors. (A)Diversity wheel of plant carboxypeptidases inhibitors. The sequence diversity of the plant carboxypeptidases inhibitors is indicated for peptides found in this study. Variations of the amino acids for each position to the carboxypeptidases, generated by alignment of the Wolfberry carboxypeptidases inhibitor backbone, are indicated in a radial formation on the outside of the backbone circle. The cysteine residues that are absolutely conserved are highlighted in grey. (B) The Sequences of 34 identified plant carboxypeptidase inhibitors have been compared using a sequence logo to highlight their amino acid variation. Conserved cysteine residues are in yellow.

3.2.8 Plant CPIs follow the CRP biosynthesis pathway

Analysis of the putative plant CPI precursor sequences revealed that they consist of an ER signal domain, an N-terminal prodomain, a CPI domain, and a short C-terminal prodomain (Figure 3.7A). The presence of a signal peptide in the precursor sequence shows that these peptides likely follow the conventional pathway for secretory proteins that is similar for most CRPs [81], including cyclotides from Rubiaceae and Violaceae [86,227,269], squash trypsin inhibitors from *Momordica cochinchinensis* [270], and towel gourd trypsin inhibitors [271].

Figure 3.7 shows the precursor sequences of three plant CPIs and three animal CPIs. As compared to animal CPIs, plants CPIs contain an additional C-tail prodomain. This pro-region is hypothesized to aid peptide folding, and may be involved in vacuolar sorting and modulating inhibitor activity prior to removal by specific proteases in the vacuolar compartment [272]. The cleavage site between the N-terminal pro domain and the mature peptide domain is also highly conserved, with 93.1% of plant CPIs cleaved after K-L-F/L (Figure 3.7B). This finding suggests that the same processing enzyme may promote the release of the mature CPI peptide from the N-terminal pro domain, and also supports that these peptides belong to the same family.

The precursor sequences of plant and animal CPIs (Figure 3.7A) show that they may have slight differences in biosynthesis. The animal CPI precursors contain only two domains, the signal peptide and the mature domain, and lack the N-terminal proregion that is present in the plant inhibitors. This difference may allow animal cells to produce the mature CPI peptide without additional processing steps to increase the production

rate of these inhibitors. These differences in processing may have arisen from the different roles these inhibitors play in the cells of the organism.

It should be noted that diverse structures are used to organize cystine knot peptide precursors in the plant kingdom. These structures include multiple repeats of cystine knot peptides, as was reported for the Violaceae family cyclotides [37,148], chimeric precursors of both cystine knot and other types of proteins in *Clitoria ternatea* [35], and the above four-domain precursor structure of plant CPIs. Due to the significant diversity in precursor structures, understanding the differences in genetic sequences between and within each peptide family will allow us to better understand their biosynthetic mechanisms and how they might be applied for crop protection and pharmaceuticals.

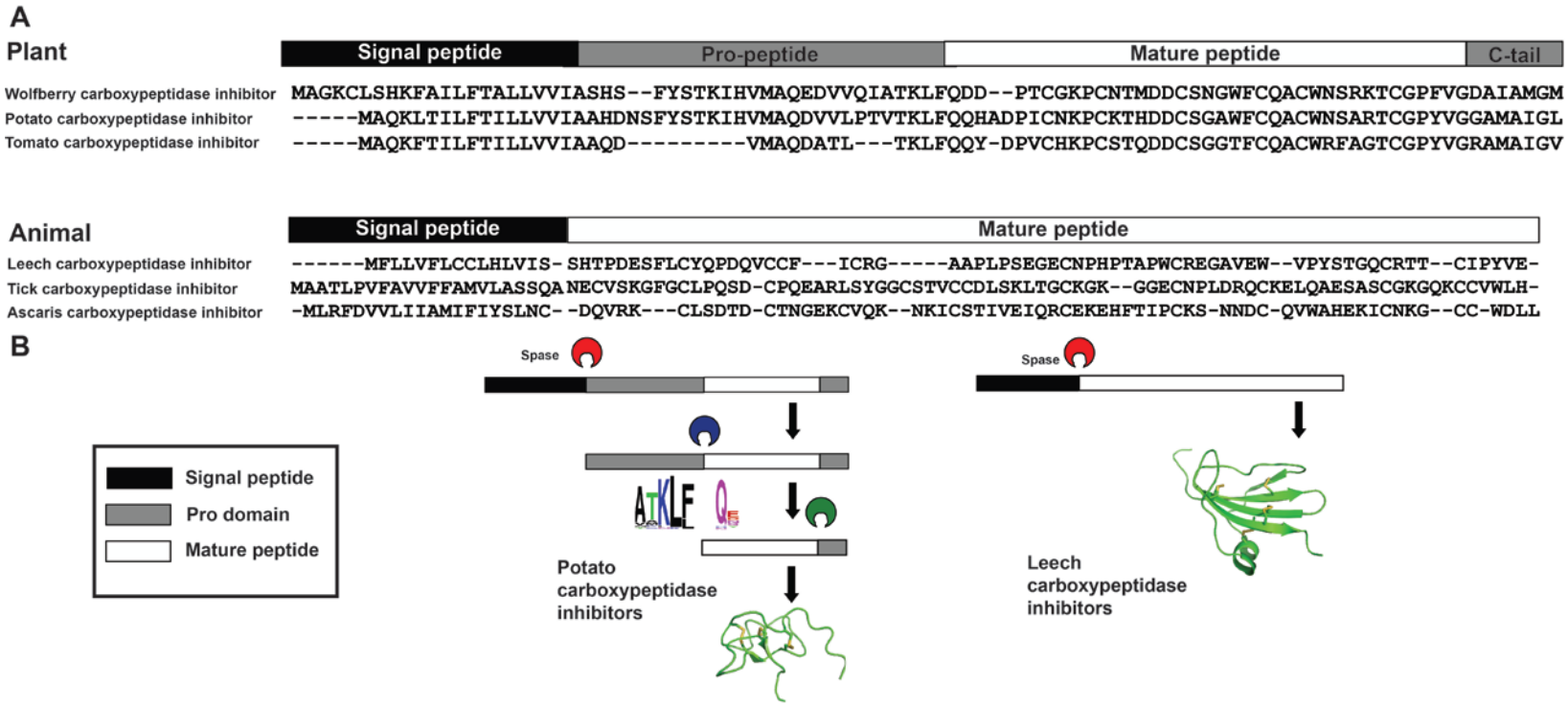


Figure 3.7. Precursor peptide sequence alignment and biosynthesis pathway of plant and animal carboxypeptidases inhibitors, plant trypsin inhibitor and plant alpha amylase inhibitors. (A) Alignment of peptide precursors of plant carboxypeptidases inhibitors and other plant cystine knot peptides. Processing site of the C-terminus of the pro-domains is not similar when comparing plant carboxypeptidases (KLF) and other plant cystine knot peptides (VX). (B) Comparison of the biosynthesis pathways between plant carboxypeptidase inhibitors, other plant cystine knot peptides and animal carboxypeptidases inhibitors. Signal peptide is removed from the full precursors by enzyme Spase. Plant peptides require additional cleavage by processing enzyme at the C-terminal of the N-terminal prodomain to release the mature peptide, while plant carboxypeptidase inhibitors require one last processing at the n-terminal of the mature domain to remove the c-terminal domain.

3.2.9 Evolution of CPIs

The different carboxypeptidases that have been identified and characterized are an example of convergent evolution that occurs between phyla, while divergence occurs within the phylum. An example of convergent evolution is seen in an inhibitor isolated from leeches, wherein the 66 residue peptide has neither sequence nor folding similarity with plant CPIs, but nonetheless can inhibit carboxypeptidase activity at nanomolar concentrations. The cysteine-rich structure of this leech peptide produces a 3D structure that is cross-connected by disulfide bridges. However, the C-terminal tail is homologous with plant CPIs, which suggests that it has a similar mechanism of inhibition [241]. Other proteins that display similar inhibitory activity as plant CPIs are found in marine mollusks [56] and parasites [273]. An inhibitor found in ticks inhibits carboxypeptidase by anchoring to the surface of the target enzyme in a double-headed manner that is not observed with other CPIs [27]. CPIs in animals display higher inhibition of carboxypeptidases than plant CPIs, despite having similar inhibitor modes, and this difference may be due to the larger size of these peptides (two or three times larger), which allow stronger interactions between inhibitors and enzymes that in turn produce a lower Ki. Although different CPIs possess strikingly different peptide sequences, they still exhibit similar functions in their inhibition of carboxypeptidase enzymes.

With increased interest in the discovery of additional carboxypeptidases and the roles these enzymes play in physiology, as well as the promise of natural peptides as therapeutics, analysis of our findings from an evolutionary and drug discovery point of view would be valuable as it may provide additional guidance for the development of novel tools or drug leads.

An unbiased tBLASTn search allowed the prediction of the extent of expression and availability of a particular protein family in different plant species and also homolog diversity. By examining the molecular diversity of identified peptides in the CPI family, we may be able to correlate the conservation of residues to peptide function. However, the accuracy and significance of these data is limited by the information available in databases, and thus cross validation and confirmation of search findings is needed at the peptidomic level.

3.2.10 Therapeutic potential of CPIs and biologic engineering

Results from the stability assays performed in this study demonstrated the stability of WCI against thermal, acidic, enzymatic, and serum treatment. WCI displays the robustness of small molecules, such that it could likely resist degradation in the digestive tract and in human serum, which would enhance its oral bioavailability. Carboxypeptidases have been found to play many physiological roles, and thus CPIs could exert pharmacological effects in the regulation of thrombotic process [274], inflammation [275], and cancer progression [276], while also having possible uses as pesticides [277]. PCI has already been proposed as a leading compound for development of an anti-thrombotic drug [278] and for development of insecticidal parasite strategies [279].

The use of CPIs as anti-thrombotic drugs is especially compelling. CPIs isolated from plants [274], leeches [30], ticks [27], and ascaris parasites [238] are known to inhibit the thrombin activatable fibrinolysis inhibitor (TAFI), a regulator of fibrinolysis [96]. Inhibition of activated TAFI promotes endogenous thrombolysis in jugular clots in a rabbit model [280], while the anti-thrombotic activity of PCI showed activity in a murine model of ferric chloride-induced vena cava thrombosis [281]. Furthermore, reductions in tissue factor-induced renal microthrombosis in rats were also demonstrated following the administration of CPIs [274]. In particular, the CPI from leeches was successfully marketed as the anti-thrombotic drug Refludan©, although this drug was withdrawn in 2012. The above studies show that TAFI inhibition is widely exhibited among naturally occurring proteinaceous CPIs, which suggests that, by virtue of its sequence and

structural conservation, WCI may also inhibit TAFI, although experimental data are needed to confirm this possibility.

The additional CPIs identified in this work may provide further knowledge that can be applied to the development of inhibitors with higher specificity to target different carboxypeptidases, or to counteract the co-evolutionary capacity of other organisms to adapt to natural PCIs through the expression of proteases that are insensitive to these natural inhibitors [267,282].

The discovery of CRPs is of great significance for drug discovery efforts. The high tolerance of the conserved inter-cysteine loops allows most CRPs to adopt a similar scaffold. In particular, plant CPIs adopt cystine knot scaffolds, which are small and extraordinarily resistant to thermal and endopeptidase degradation [261]. Thus, cystine knots such as those contained in plant CPIs are an appealing feature that could be exploited in potential therapeutics [283,284]. The small size of these cystine knot peptides makes them suitable for chemical synthesis and their high tolerance to sequence variation and spacing of the cysteine residues could allow these plant CPIs to be utilized as scaffolds in protein engineering that could afford novel bioactivities in a manner that is similar to the grafting of DALK or DAK, a bradykinin antagonist, onto the cyclotide kalata B1 scaffold [152]. Given that the mechanisms of carboxypeptidases are so well characterized, it may be possible to use CPIs as a scaffold onto which a functional peptide could be grafted, while leaving selected amino acid residues that are known to play important roles in the inhibitory action of the peptide intact. Through this grafting method bifunctional peptide therapeutics could be created.

3.3 Conclusion

Since the discovery of the potato carboxypeptidase inhibitor in 1975 [19], only 2 CPIs have been identified in plants. The discovery of WCI in this study and the subsequent tBLASTn search greatly expanded the number of plant CPIs in the knottin family from 2 to 29, one of which was expressed in Asteraceae family plant that was not previously known to express CPIs. WCI and the other CPIs identified in this paper displayed high conservation of the cysteine motif and scaffold while containing variable amino acid sequences in loops between each cysteine residue. This sequence analysis will allow us to further our understanding of these peptides and provide insight into the distribution, evolution, and biosynthesis of cysteine-rich proteins. Meanwhile, study of the different bioactive components in wolfberry extracts will provide a better understanding of the various pharmacological applications of this herb, and promote its use as component of complementary alternative treatments.

The carboxypeptidase inhibitory activity of WCI may indicate that it may be responsible for the indication of the wolfberry herb in the treatment of blood circulatory illness in TCM as it may exhibit similar effects as the anti-thrombotic activity of potato carboxypeptidase inhibitor.

Overall, this study expanded our knowledge of *L. barbarum* peptides and the evolution, distribution, and molecular diversity of plant CPIs, thus contributing to characterization of cystine knot peptides as well as to drug discovery and engineering of bioactive peptides.

Chapter 4 Lybatides from *Lycium barbarum* Contain An Unusual Cystine-stapled Helical Peptide Scaffold

4.1 Introduction

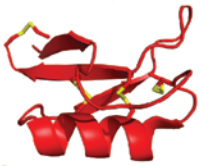
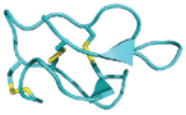
Lycium barbarum is a deciduous shrub native to southeastern China from the Solanaceae family [157]. The fruit of the plant is also known as the highly popular herb wolfberries or goji berries. However, the root bark of *L. barbarum* (地骨皮 or DiGuPi in Chinese) is also in Traditional Chinese Medicine used for the treatment of chronic low-grade fever, cough, hemoptysis and hematuria, diabetes mellitus and hypertension. A wide variety of secondary metabolites has been isolated from the cortex of *L. barbarum* root, including alkaloids, flavonoids and flavone glycosides. In addition, a group of small cyclic peptides, known as licyumins A - D, with molecular weights <1 kDa, are identified in the root bark extract and have been demonstrated to inhibit renin and angiotensin-converting enzymes [285]. However, no bioactive peptides >2kDa have been reported.

Bioactive compounds from medicinal plants have long inspired chemical structures in drug discovery. There are fewer studies on natural products that focus on peptides as compared to small molecules and secondary metabolites in plants [85,286]. Even in studies seeking to identify peptides from plants as potential bioactive compounds, efforts have been largely placed on small cyclic peptides. This bias is the result of a general perception that peptides >2 kDa are not stable and are readily denatured the herbal decoction process or during digestion in the gastrointestinal tract. This would be true for larger peptides and proteins with molecular weight >8 kDa.

However, CRPs, a class of peptides with a molecular range of 2 - 6 kDa and 3 - 5 disulfide bonds, are stable against heat, acid and proteolytic degradation [20]. As a group, CRPs in this defined chemical space are a great source of leads and inspiration in the development of useful drugs in medicinal plants [23,287].

Plant CRPs that falls within the chemical space of 2 - 6 kDa can be arbitrarily classified according to their structure into two major groups (Table 4.1). They are the cystine-stabilized α-helical (CSα) peptides and cystine-stabilized β-peptides. The CSα peptides contain a dominant cystine-stapled helix with at least three turns in their structure and are usually found in plant CRPs with >40 residues such as plant defensins and plant thionins. In contrast, Cystine-stabilized β-peptides generally contain short β-strands and no well-formed α-helix and are found in plant CRPs with <40 residues, such as knottins and hevein-like peptides [23]. The helical conformation of CSα peptides is usually stabilized by another well-defined secondary structure such as an α-helix or one or more β-strands. This phenomenon can be observed in the cystine-stabilized α/β motif (CSαβ) of plant defensins [27,39], and the cystine-stabilized α-α helical motif (CSαα) of thionins [288].

Table 4.1. Major families of cysteine-rich peptides in plants.

Structural Motif	Peptide Structure	Peptide	Family
Gamma (Γ) fold β1-α1-α2-β2-loop motif		β-Purothionin	Thionins
CSαβ motif β1-loop-α-β2-β3		NaD1	Defensins
Gly & Cys rich Central β strands & (short helical) side loops		gB5	Hevein-like peptides
Cystine knot Short β stands and loops		Wr-A1	Knottins

In this chapter, I report the identification and characterization of two novel cystine-stapled helical CRPs, namely lybatide 1 and 2 (lyba1 and lyba2), from the root bark of *L. barbarum*. Structural determination by X-ray crystallography of lybatide2 reveals a helix in its structure that is not stabilized by another well-defined secondary structure, different from the structural features of CSαβ and CSαα peptides. Analysis of the cysteine arrangements, disulfide connectivity and overall three-dimensional structure of lybatides to other cysteine-rich peptides revealed that lybatides represent a new family of CRP with an unusual cystine-stabilized helical structure that is not observed in other plant CRPs.

4.2 Results

4.2.1 Isolation of lyba1 and lyba2 from *L. barbarum*

Mass spectrometry analysis of the aqueous extract of *L. barbarum* root bark revealed the presence of two peptides with m/z 3509 and 3616 Da (Figure 4.1A). Presence of eight cysteine residues in these peptides was confirmed with a mass shift 848 Da after dithiothreitol reduction and *S*-alkylation with 4-vinylpyridine (Figure 1B, 1C). The extraction and purification process was then scaled up to isolate the two peptides, lyba1 and lyba2. Yield of lyba 1 and lyba 2 were 20 mg and 30 mg, respectively, from 3 kg of dried plant material. In order to determine their amino-acid sequences, lyba1 and lyba2 were reduced fully with dithiothreitol and *S*-alkylated with 4-vinylpyridine followed by digestion with trypsin or chymotrypsin. Two fragments of m/z values of 2523 and 1853 Da was generated after Trypsin digestion of the *S*-alkylated lyba1, while four fragments with m/z values of 1577, 2799, 2490 and 1886 Da were generated after chymotrypsin digestion. Tandem mass spectrometry sequencing of the fragments revealed the amino acid sequence of the 32-residue Lyba1 as DSCSEYCSNNSCPYCDGQKLYTLCCINTCCPS (Figure 4.2). Similar process was repeated to obtain the sequence of the 33-residue Lyba2 as DSCSEYCSNRCPSCDGQTQTQYTLCCINICCPS (Figure 4.3). Homologous expressed sequence tag sequences from the OneKP transcriptome database (www.onekp.com) was utilized for the assignment of isobaric Ile and Leu. It was observed, after aligning the lybatide sequences, that they contain a highly conserved unique cysteine motif of eight cysteine residues arranged as C–C–C–C–CC–CC. Of particular note, two consecutive CC motifs were observed at the C-terminus of the lybatides. The lybatides

isolated from *L. barbarum* are anionic with a theoretical isoelectric point of 4.03 and are serine-rich, where five out of 32 or 33 amino acid residues are serine residues, respectively.

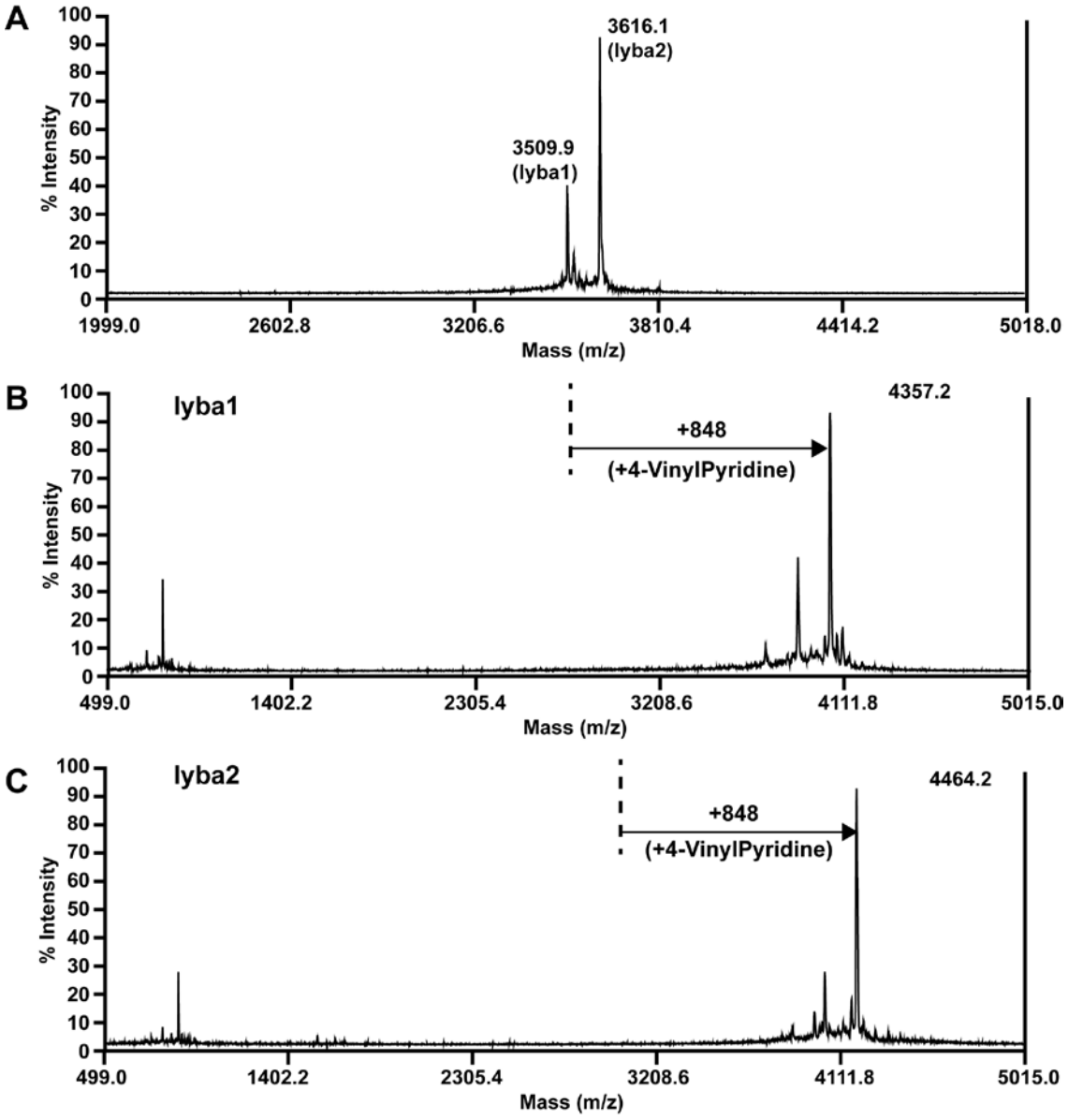


Figure 4.1. MALDI mass spectrum of lybatides, lyba1 and lyba2. A) Mass spectrum displaying two main peaks at m/z 3509, lyba1, and m/z 3616, lyba2. B) 4-vinylpyridine alkylation of lyba1 after reduction by dithiothreitol, showing a m/z shift of 848. C) 4-vinylpyridine alkylation of lyba2 after reduction by dithiothreitol, showing a m/z shift of 848.

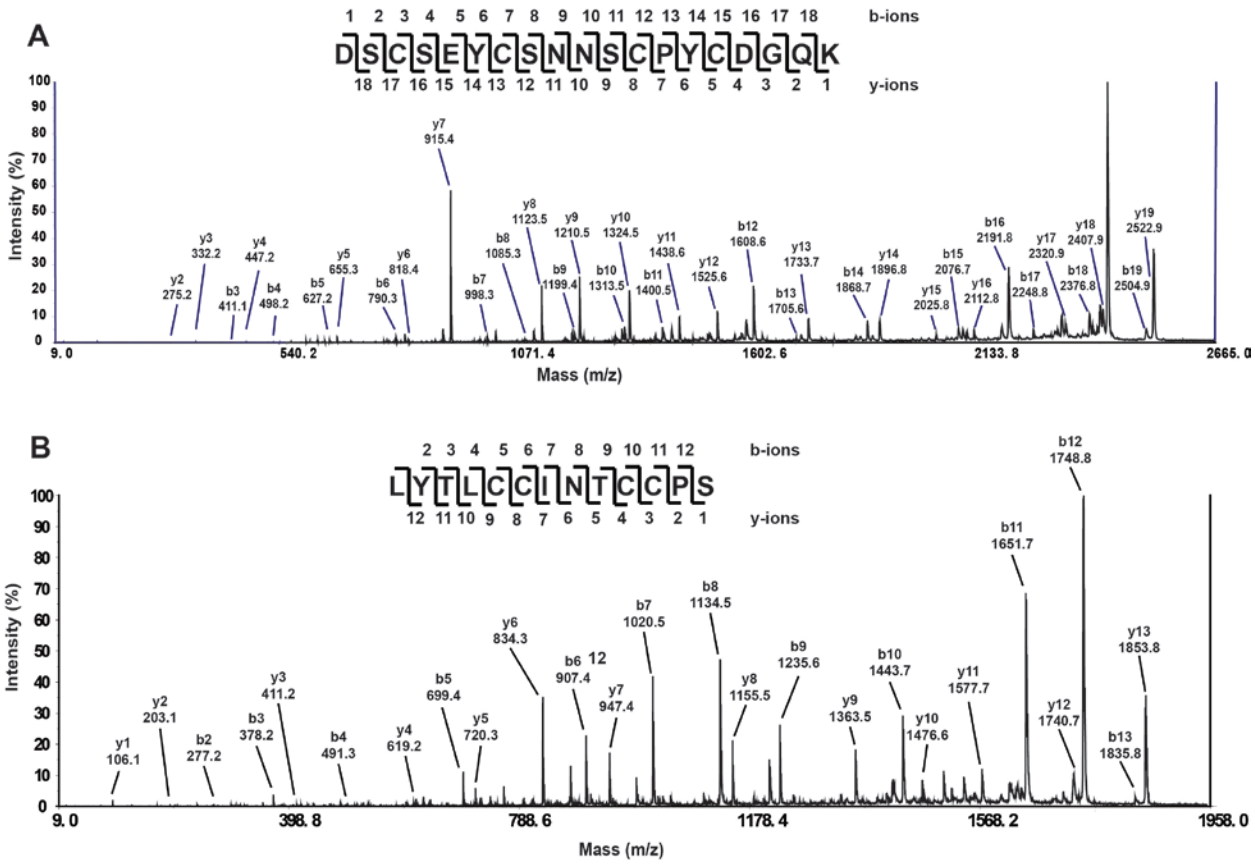


Figure 4.2. *De novo* sequencing of lyba1. Tryptic digestion of *S*-alkylated lyba1 produced two fragments, one with m/z of 1853 (A) and another with m/z of 2523 (B). Tandem mass spectrometry was used to determine the primary peptide sequence as DSCSEYCSNNSCPYCDGQKLYTLCCINTCCPS.

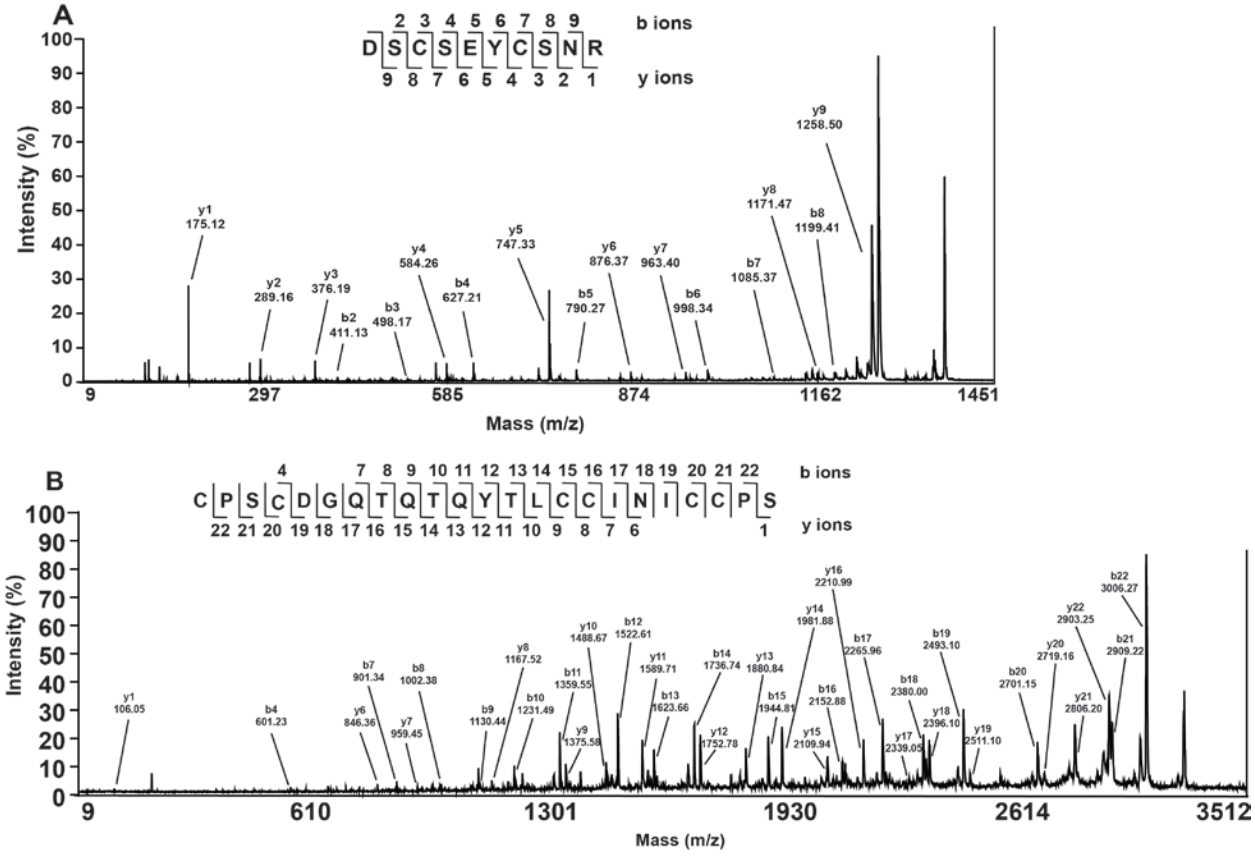


Figure 4.3. *De novo* sequencing of lyba2. Tryptic digestion of *S*-alkylated lyba2 produced two fragments, one with m/z of 1372 (A) and another with m/z of 3109 (B). Tandem mass spectrometry was used to determine the primary peptide sequence as DSCSEYCSNRCPSCDGQTQTQYTLCCINICCPS.

4.2.2 Determination of the disulfide connectivity of lyba1

A stepwise partial alkylation strategy, where lyba1 is S-tagged alkylated with iodoacetamide and N-ethylmaleimide (NEM), followed by MS/MS analysis, was utilized to elucidate the disulfide connectivity of lybatides [20]. In summary, a mixture of intermediates species containing 1-SS, 2-SS and 3-SS of lyba1 was obtained through partial reduction by TCEP and *S*-alkylation by NEM. RP-HPLC was then used to purify the partially S-alkylated intermediates, followed by fully reduction with DTT and S-alkylation with IAM to obtain three major groups of mixed alkylated products. These products analyzed with MALDI-TOF MS/MS.

MS/MS fragmentation patterns of the mixed S-alkylated peptides allow us to determine the cysteine linkages of the peptide by observing differentially S-alkylated cysteine residues by NEM and IAM (Figure 4.4). Depending on the ratio of NEM-modified and IAM-modified cysteine residues, five differentially S-alkylated products, two with a mass of 4109 Da, one with a mass of 4245 Da and two with a mass of 4381 Da, were produced. MS/MS sequencing of the two NEM-modified cysteines species sharing a mass of 4109 Da established the disulfide connectivity of CysIII–CysVII and CysII–CysVIII. Likewise, the locations of the IAM-modified cysteines of the two peptides sharing a mass of 4381 Da established the remaining disulfide bonds as CysI–CysVI and CysIV–CysV. Combining the two sets of data led to the complete disulfide connectivity of lyba1 as CysI–CysVI, CysII–CysVIII, CysIII–CysVII and CysIV–CysV.

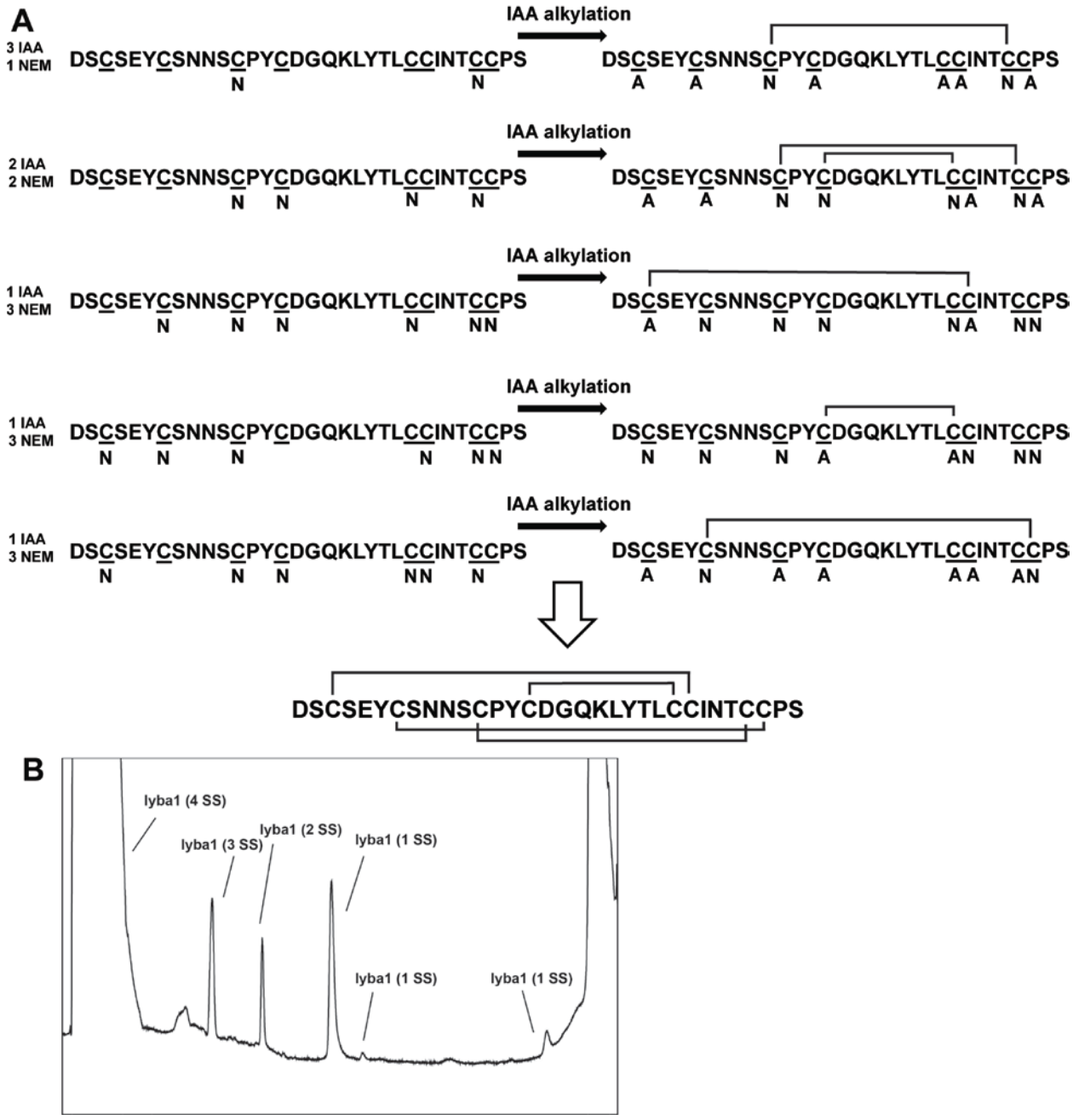


Figure 4.4. Disulfide mapping of lyba1. The disulfide structure was elucidated by S-tagged alkylation with iodoacetamide (IAA) and N-ethylmaleimide (NEM) of the peptide using a stepwise partial reduction alkylation strategy followed by MS/MS analysis. **A.** Five partially reduced intermediates were alkylated with NEM followed by full reduction by dithiothreitol and S-alkylation with IAA. MS/MS-analysis of the differently S-alkylated peptides allows us to deduce the disulfide connectivity of lyba1. **B.** HPLC chromatogram of the intermediate species. **N** indicates cysteine residues alkylated with NEM and **A** indicates cysteine residues alkylated with IAA.

4.2.3 X-ray crystal structure of lyba2

The more easily crystallized lyba2 peptide was used to determine the 3D structure with X-ray crystallography. The three-dimensional structure at 1.48-Å was determined by the sulfur single-wavelength anomalous dispersion (S-SAD) method [209] (Figure 4.5). Table 4.2 presents the final refinement statistics. There are three copies of the peptide molecule (named A, B and C) per asymmetric unit (ASU) of the crystal. The root-mean-square deviations (for Cα atoms) between A and B, A and C, and B and C are 0.44, 0.46 and 0.36 Å, respectively. The last residue (Ser 33) of molecule A is undefined due to incomplete electron density. The Rcryst and Rfree of the final model are 0.110 and 0.137, respectively.

Analysis of the diffraction of the crystal structure revealed the presence of a turn (Ser2–Tyr6), a one-turn 3_{10} helix (Pro12–Cys14), and an α helix (Gln19–Leu24 in molecule A, Gln19—Cys25 in molecules B and C) that progress into a distorted π helix (Cys25–Cys30 in molecule A, Cys26–Cys30 in molecules B and C; Figure 4.5A) in lyba2. A π helix is a widened α-helix with 4.1 residues per turn (instead of 3.6) and hydrogen bonds between atoms N (i+5) and O (i). The four disulfide bonds present are formed between Cys3–Cys26, Cys7–Cys31, Cys11–Cys30 and Cys14–Cys25 according to the X-ray structure (Figure 4.5A), which agrees with our disulfide mapping results. Cys3, Cys7 and Cys11, located in the N terminal loop, are stabilized by disulfide bonds with Cys26, Cys31 and Cys30 in the π helix. Cys14 is located at the C-terminus of the one-turn 3_{10} helix and connects with Cys25 of the π helix (molecule A) and α helix (molecules B and C). These four disulfide bonds result in lyba2 forming a compact and constrained structure.

Table 4.2. Data collection and refinement statistics for lyba2

	Diffraction data	
	Native	Sulfur-SAD
Data collection statistics		
Space group	$P2_1$	$P2_1$
Unit-cell dimensions (Å, °)		
	a = 19.32, *b* = 54.65	*a* = 19.37, *b* = 54.76,
	c = 42.40	*c* =42.49
	β = 100.60	β = 100.59
Wavelength (Å)	0.9184	1.8000
V_m (Å^3/Da)	2.04	2.04
Solvent content (%)	39.8	39.8
Resolution range (Å)	41.68-1.48 (1.56-1.48)	41.77-1.95 (2.06-1.9
Number of unique reflections	14465 (2057)	6382 (879)
R_{merge}[1]	0.033 (0.039)	0.037 (0.203)
R_{pim}[2]	0.014 (0.017)	0.023 (0.173)
Completeness (%)	99.6 (98.4)	98.8 (96.5)
Mean I/σ(I)	44.3 (31.8)	9.7 (7.0)
$CC_{1/2}$[3]	0.999 (0.998)	0.997 (0.822)
Multiplicity	6.5 (6.3)	5.8 (3.9)
Wilson B-factor (Å^2)	4.2	14.2
Refinement statistics		
R_{cryst} (%)[4]	10.76	
R_{free} (%)[4]	13.70	
Protein atoms	731	
Water molecules	155	
Ligands	5	
Clashscore[5]	4	
r.m.s. deviation in bond lengths (Å)	0.02	
r.m.s. deviation in bond angles (°)	1.95	
Average *B*-factor for all atoms (Å^2)	10.6	
Ramachandran plot		
Residues in favored regions (%)	91.2	
Residues in additionally allowed regions (%)	8.8	
Residues in outlier regions (%)	0	

[1] $R_{merge} = \sum_{hkl}\sum_i \left| I_i(hkl) - \langle I(hkl)\rangle \right| / \sum_{hkl}\sum_i I_i(hkl)$.

[2] $R_{pim} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \times \sum_j \left| I_i(hkl) - \langle I(hkl)\rangle \right| / \sum_{hkl}\sum_j I_i(hkl)$

[3] $CC_{1/2}$ is the correlation coefficient determined by two random half data sets.

[4] $R_{cryst} = \sum_{hkl} \left| F_o(hkl) - F_c(hkl) \right| / \sum_{hkl} F_o(hkl)$. R_{free} was calculated for a test set of reflections (4.9%) omitted from the refinement.

[5] Clashscore is defined as the number of clashes calculated for the model per 1000 atoms (including hydrogens) of the model. Hydrogens were added by *MolProbity.*

The lyba2 crystal lattice is formed from the interactions between molecules A, B, and C, as well as their symmetry-related copies. Interestingly, we discovered that a short polyethylene glycol molecule (PEG, HO-(CH2-CH2-O)3-CH2-CH2-OH, here named “P40”) is able to stabilize the crystal lattice by filling a cleft between molecule C and its symmetry-related copy C* (symmetry operator: X-1, Y, Z; Figure 4.5B). The PEG molecule is well defined by electron density (Figure 3C). Its torsion angles about the C-O bonds are in the antiperiplanar range and those about the C-C bonds are in the synclinal range, as is common for polyethers [289]. The majority of its interactions with peptide copies C and C* consist of water-mediated hydrogen bonds (Figure 4.5C).

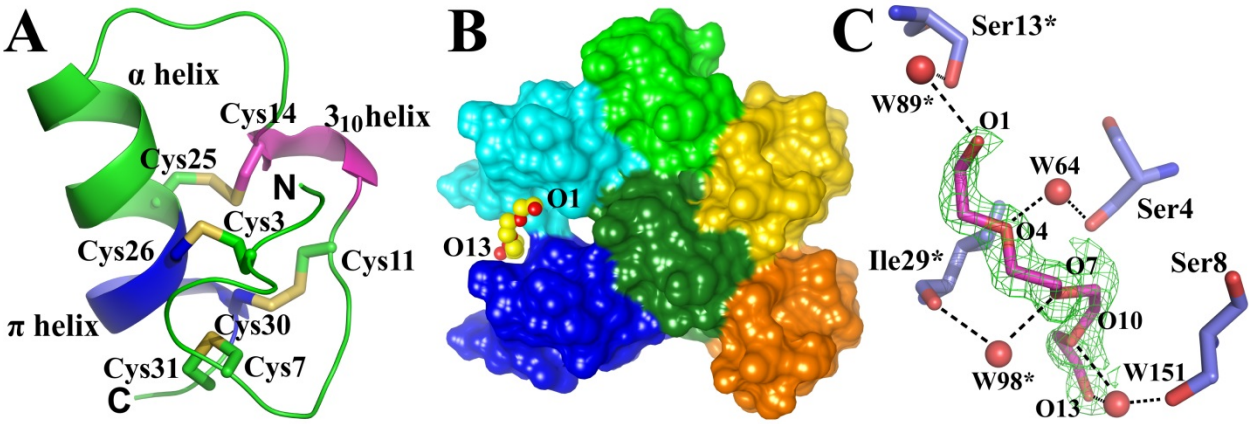


Figure 4.5. The three-dimensional structure of lyba2. (A) Cartoon view of the overall structure. Four disulfide bonds are formed by Cys3-Cys26, Cys7-Cys31, Cys11-Cys30 as well as Cys14-Cys25. The 3_{10} helix and π helix are colored purple and blue, respectively. (B) Packing of the 3 copies of lyba2 and their symmetry-mates in the crystal. Molecules A, B and C are colored green, pink and blue, respectively. The polyethylenglycol (PEG) molecule is shown in sphere view. The terminal oxygen atoms of the PEG are labeled. (C) Details of the interactions between PEG and lyba2. An Fo-Fc electron density of the PEG molecule (green; contoured at 2.5 σ level) is displayed. Hydrogen bonds between the PEG and lyba2 are indicated as dashed lines. Symmetry-related water molecules and amino-acid residues are labelled by an asterisk (*).

4.2.4 Metabolic stability of lybatides

CRPs are known for their compact structure and high resistance to heat and enzymatic degradation as a result of their well-defined disulfide core. The stability of CRPs allows them to be identified as putative bioactive peptides as they are able to survive the decoction processes in traditional herbal preparations and still possess bioavailability and pharmacological effects administered orally. Figure 4 shows that lyba1 is able to resist degradation and denaturation in harsh testing conditions. A total of 96% of lyba1 remained after heating in boiling water at 100°C for 2 h. Similarly, 99% of the peptide remained intact after incubation in 2 M HCl for 3 h, and 99% of the peptide was not degraded when incubated with trypsin for 3 h. (Figure 4.6)

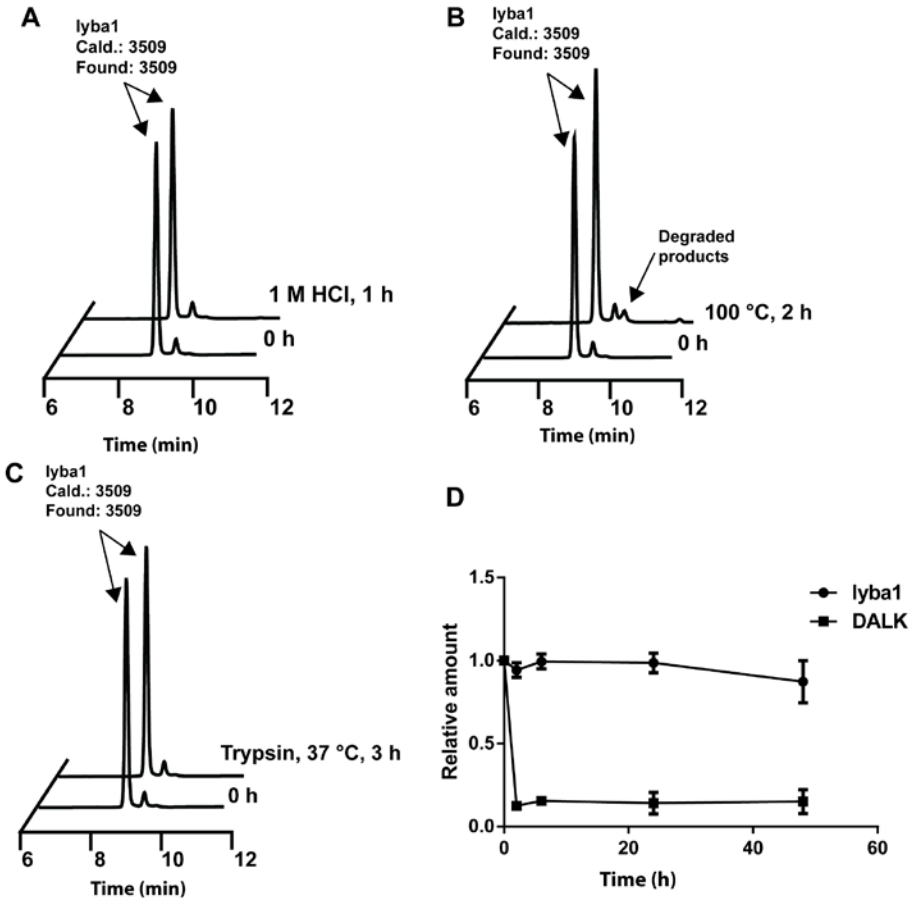


Figure 4.6. Stability assays of lybatide, lyba1. (A) Thermal stability of lyba1 incubated at 100°C for 2 h. (B) Acidic condition stability of lyba1 incubated in 1M HCl (pH 2.0) for 1 h. (C) Trypsin enzymatic stability of lyba1 incubated for 3 h in buffer as suggest by manufactures at 37◦C. (D) Stability of lyba1 incubated in human serum for 60 min at 37°C, DALK was used as a control. The molecular weight of the peak was determined by MALDI-MS.

4.2.5 Anti-bacterial activity and cytotoxicity of lyba1

Antimicrobial or cytotoxic activities have been found to be possessed by many CRPs including defensins, thionins and hevein-like peptides [23]. In order to determine if lybatides possess similar antimicrobial activity, a radial diffusion assay on four different microorganisms was performed with lyba1. No antimicrobial activity of lyba1 was observed against the specific strains tested up to 1 mM. In addition, MTT assays were conducted to evaluate the cytotoxic effect of lyba1 on three different cell lines which lyba1 did not display any cytotoxic effect when tested up to 100 μM (Figure 4.7).

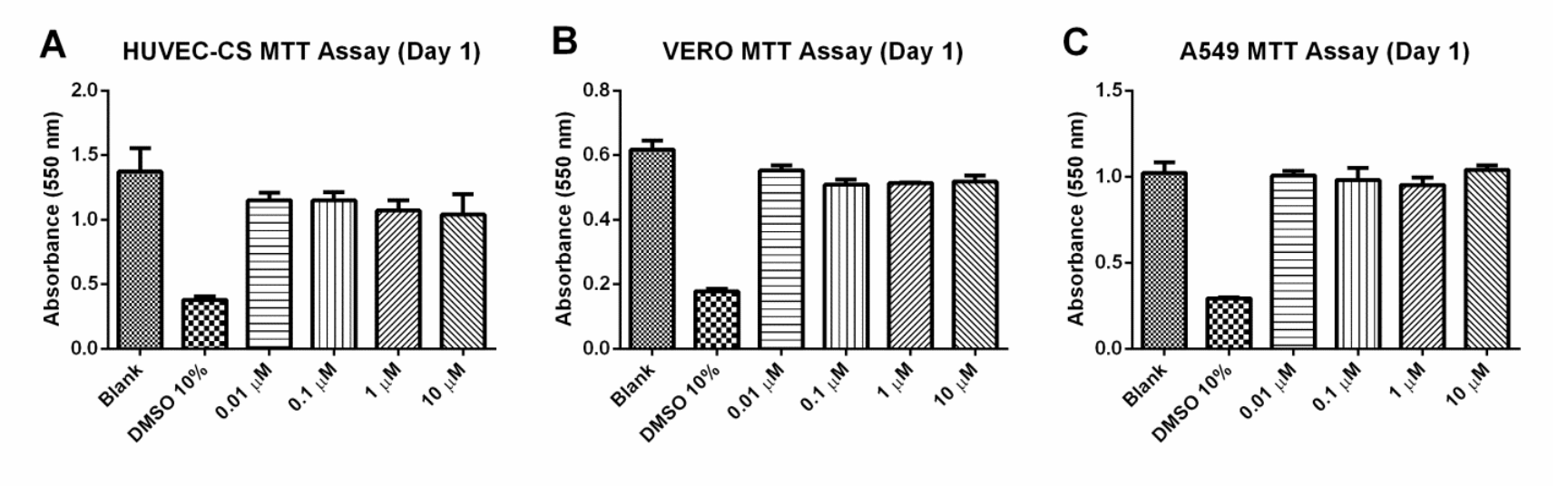


Figure 4.7. Cytotoxicity of lybatides. MTT assay of lybatides on (A) human embilical vein cord cells (HUVEC-CS), adenocarcinomic human alveolar basal epithelial cells (A549) and monkey kidney epithelial cells (vero). 10% DMSO were used as control. No significant difference was detect after one day of incubation with lyba1 from testing concentrations of up to 10 μM as compared to blank.

4.3 Discussion

In this study, we report the isolation and characterization of two novel helical CRPs, lyba1 and lyba2, from the aqueous extract of *L. barbarum* root bark. Lybatides contain eight cysteines residues, resulting in a highly constrained and compact structure. They do not show any sequence similarity to other known plant CRPs. Moreover, they exhibit a novel 3D structure with a unique cysteine framework (Table 4.3) that contains a dominant α-helix that is cystine-stapled and not stabilized by another well-defined secondary structure. When compared to defensins and thionins, the smaller mass of lybatides together with the presence of a cystine-stapled α-helix structural motif suggests that lybatides represent a new CRP family. This would allow lybatides to be used as a new scaffold in peptide drug engineering [153,290].

Table 4.3. Sequence alignment of plant CRPs from different CRP families.

Peptide Family	Peptide	No. of S-S	Amino Acid Sequence
Lybatides	lyba1	4	-DSCSE--------YCSNNSCP-YC--DGQ--KLYTLC---C-IN--TCCPS------
	lyba2	4	-DSCSE--------YCSN-RCP-SC--DGQTQTQYTLC---C-IN--ICCPS------
Defensins	NaD1	4	-RECKTESNTFPGICITKPPCRKAC---ISEKFTDGHC---SKILRRCLCTKPC----
α-hairpinin	Ec-Amp	2	--GSGR--------GSCRSQCMRR---HEDEPWRVQEC-----VS---QCRRRRGGGD
Thionins	β-Purothionin	4	-KSCCK--------STLGRNCYNLCRARGAQKLCANVCR--CKLTSGLSCPKDFPK--
Jasmintides	jS1	3	-QLCLQ--------CRSNSDCN----------IIWRIC-----RDG--CCNVI-----
Knottins	cT1	3	-DPTCS--------VLGDFKCN-----PGRCCSKFNYCGSTAAYCGPGNCIAQCP---
Hevein-like peptides	gB5	4	GIPCGE--------SCVFIPCI----------TGAIGC-----SCKSKVCYRN-----

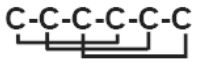
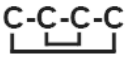
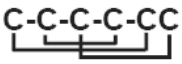
4.3.1 Lybatides represent a novel family of CRPs

MALDI MS/MS *de novo* sequencing of lybatides revealed the presence of eight cysteine residues (8C-CRPs) with a distinctive C-C-C-C-CC-CC arrangement. Interestingly, there are two consecutive CC motifs at the C-terminus of lybatide. The CC motif, with two adjacent cysteine residues without any amino acid in between, is often a diagnostic element in CRP families (Table 4.3). For example, a CC motif is found at the N-terminus of thionins, resulting in a cysteine spacing of CC-C-C-C-C-C-C, while in trans-defensins, one CC motif is found at the C-terminus as C-C-C-C-C-C-CC [27]. In addition, in CRPs such as cystine-knot α-amylase inhibitors, hevein-like peptides and jasmintides, there is a CC motif in the middle [26,28,29]. The presence of two consecutive CC motifs at the C-terminus of a CRP has not been reported thus far.

A novel disulfide connectivity of CysI–CysVI, CysII–CysVIII, CysIII–CysVII and CysIV–CysV was determined in lybatides from disulfide mapping. Currently, in plant CRPs, there are only three types of general disulfide connectivities reported. Of the three, CysI-V, CysII-IV and CysIII-VI, is only found in the most recently discovered jasmintides [25]. The remaining two are cystine knots found in defensins, knottins [58] or hevein-like peptides [29] and symmetrics found in thionins [133] and α-hairpinins [32]. The cystine knot motif is the more commonly expressed type of disulfide connectivity in plant CRPs between the two. The cystine connectivity in lybatides appears to be a hybrid between a cystine knot and a symmetric (Table 4.4). When the CysI–VI disulfide bond in lybatides is removed, the remaining disulfide bonds, CysII–VIII, CysIII–VII and CysIV–V, result in an arrangement of a symmetric [141,291].

In addition, lybatides are further distinguished from plant defensins and thionins as they are generally basic and hydrophobic, whereas the lybatides isolated are anionic and hydrophilic. This may be the reason why lybatides do not display the cytotoxic and antimicrobial effects commonly displayed by thionins and plant defensins. The presence of negatively charged phospholipids and negatively charged structures, such as teichoic acid in Gram positive and lipopolysaccharides in Gram-negative bacteria, at the membranes of bacteria results in a negatively charged environment. This would allow positively charged antimicrobial peptides like plant defensins and thionins to attach to the plasma membrane [47]. In addition, the hydrophobic nature of cysteine-rich peptides would allow interaction with the lipid bilayer and thus disrupt the integrity of the membrane resulting in cell death [292]. It is possible the anionic and hydrophilic nature of lybatides would prevent lybatide from interacting with bacterial membrane that are negatively charged.

Table 4.4. Major families of plant cystine-rich peptides

Peptide family	S-S No.	Representative member		
		Name	AA No.	Disulfide motif
Knottins	6	PCl	39	
Defensins	4	NaD1	47	
α-Hairpin	2	Ec-AMP1	37	
Thionins	4	β-Purothionin	45	
Jasmintide	3	jS1	27	
Lybatides	4	lyba2	33	

A non-plant CRP that shares the same cysteine framework as lybatides is the conotoxin RsXXIVA, a Cav2.2 channel inhibitor that is isolated from the venom duct of Conus regularis [293]. However, although they have the same cysteine framework, they do not share high sequence similarity (46.9%). CRP scaffolds have been known to be very tolerant to hypervariable amino acid sequences within the loops while maintaining similar overall structure, thus it is highly likely lybatides and RsXXIVA may share similar structural folds. Our X-ray crystallographic and disulfide connectivity studies may aid in the prediction of the structure of RsXXIVA since the structure of conotoxin RsXXIVA has not been reported yet. It would also be of interest to determine if lybatides would have calcium channel binding activity similar to RsXXIVA.

The solid-state tertiary structure of lyba2 revealed a dominant structural feature of an (α+π)-helix that is stabilized by three disulfide bonds. It is rare for a cystine-stabilized (α+π)-helix to be observed in a small disulfide-rich peptides, although an α-helix involved in disulfide bonds is commonly observed in CRP families such as thionins and plant defensins [294]. The CSαα and CSαβ motifs are two commonly known motifs of cystine-stabilized α-helix structures. The CSαα motif defined by an α-helix stabilized by another α-helix, with disulfide bonds cross-bracing between the two α-helices is found in thionins and α-hairpinins. In contrast, the CSαβ motif contains a β-sheet stabilizing an α-helix, with disulfide bonds connecting the α-helix to the central β-strand reinforcing the CSαβ structure is found in plant defensins [40].

Figure 4.8 presents five different types of α-helical peptides. Of the five, the CSαβ structural motif is the most well-known and commonly found in plant defensins (Figure 5B) composing of one α-helix and three antiparallel β-strands [30]. The CSαβ

motif was first characterized in charybdotoxin [43,44]. A β1-loop-α-β2-β3 pattern secondary structural motif is observed in the CSαβ motif, where the α-helix is parallel to three antiparallel β-strands [295]. Two disulfide bridges connected to the central β-strand of the antiparallel β-sheet stabilizes the α-helix [38]. The presence of a cystine knot, CysII–V, CysIII–VI and CysIV–VII, with an extra disulfide bond, CysI–VIII, in the structure (based on NaD1, a plant defensin isolated from Nicotiana alata) [31] is another feature of CSαβ peptides such as plant defensins. The CSαα structural motif is commonly found in thionins and α-hairpinins (Figure 5C and 5D). In thionins, the two α-helices and two anti-parallel β-strands are presented as a β1-α1-α2-β2-loop secondary structural motif, resulting in a CSαα structural motif [141]. The disulfide bonds cross brace between the two α-helices stabilizing it, with no disulfide bonds connecting the α-helix to either β-strand. Similarly, in α-hairpinins, the helix conformation is stabilized by another anti-parallel α-helix. The fifth type presented is the all-hydrocarbon-stapled helical peptide (Figure 4.8E), they are not found synthesized and will be further elaborated below.

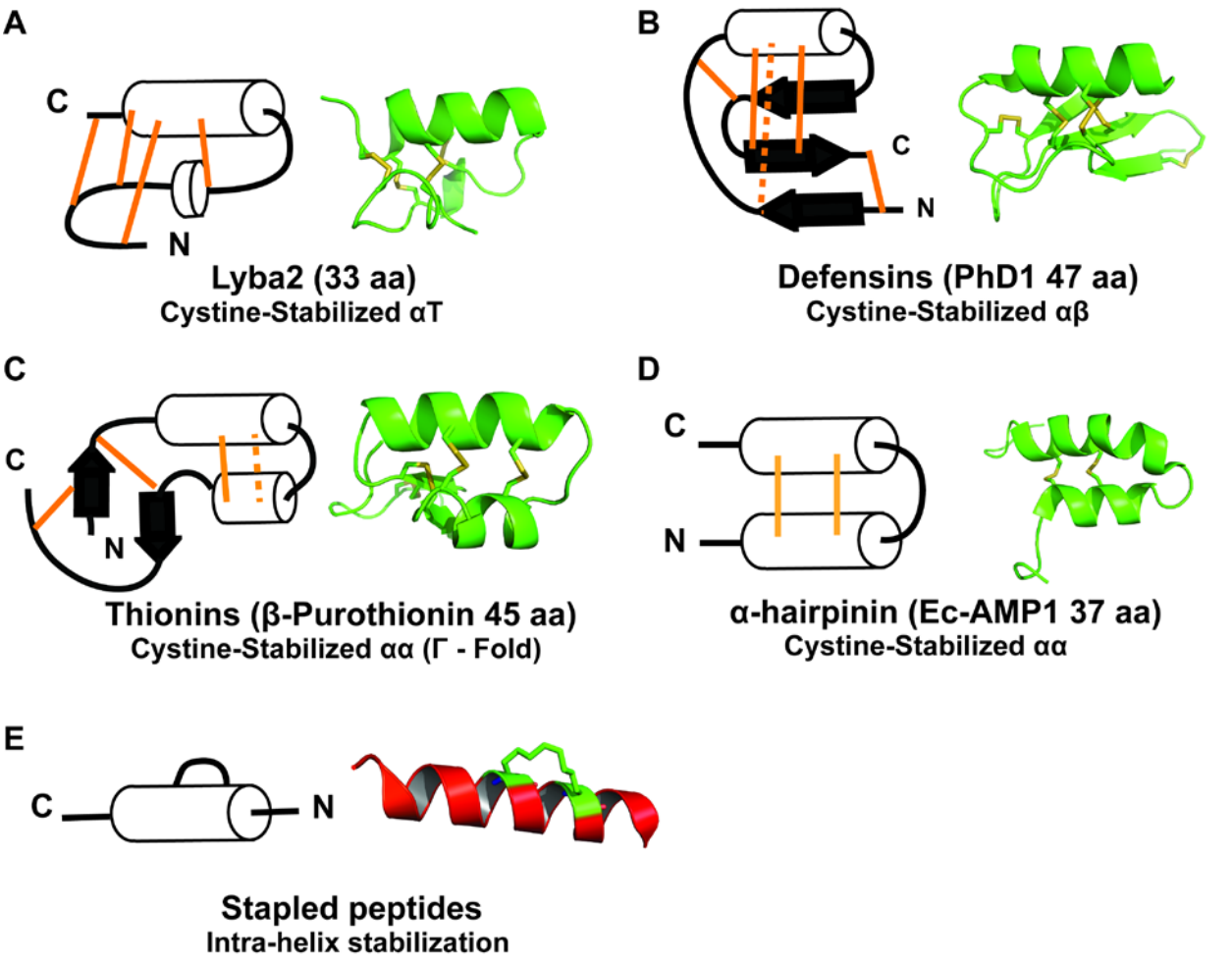


Figure 4.8. Cartoon and 3D structures of lybatide lyba2, defensin PhD1, thionin β-purothionin, α-hairpinin Ec-amp and stapled peptides. (A) Structure of lyba2. The N-terminal half of the peptide loops back along the (α+π)-helix stabilizing it through hydrogen and disulfide bonds. (B) The structure of plant defensin contains a β1-loop-α-β2-β3 pattern in the secondary structure, resulting in the CSαβ motif where the α-helix is stabilized by the anti-parallel β-sheet through disulfide bonds connecting the α-helix to the center strand of the β-sheet [42]. (C) The overall structure of thionins can be represented as by the Greek capital letter Γ. The long arm is made up of antiparallel α-helices and the short arm contains a region of extended conformation and a short anti-parallel β-sheet. Disulfide crosslinking occurs within the two arms separately [33]. (D) The two helices in α-hairpinins are stabilized by two parallel disulfide bonds [32]. (E) The stabilization of the helical structure of stapled peptides occurs within the helix and is stabilized by hydrocarbon chains [296]. (PDB ID: (B) 1N4N, (C) 1BHP, (D) 2L2R, (E) 4MZK)

Different from the α-helix – α-helix cystine stabilization found in thionins or the α-helix – β-sheets stabilization in plant defensins, the (α+π)-helix is stabilized by four disulfide bonds located in two turns (the π-turn 2 - 6 and the 310-helical turn 12 - 14) and one loop (7-11) in the N-terminal portion in lybatides. In addition, a hydrogen bond is formed between the main-chain amide of Cys3 (located in turn 2-6) and the side-chain of Tyr22 in the α-helix. A hydrogen bond is donated by the side-chain amide nitrogen of Asn9 (located in the 7-11 loop) to the main-chain C=O of Cys30 (in the π-helix). Lybatides exhibit a fold that is similar but in an opposite direction when compared to thionins, which results in a structure that resembles a capital “L” rather than the Γ fold found in thionins. When the structures of lybatides and certain members of the thionins and plants defensins are compared with TM-align [297] a TM score of 0.36 – 0.39 and 0.33 – 0.36, respectively (Table 4.5) (where a score of 0.5 indicates similar structural fold) was given.

One other interesting feature of the lybatide structure is that there is a distortion of the α-helix towards a π-helix in the C-terminal half of lybatides. This is likely to be the result of conformational stress that is induced by the disulfide bonds at the C-terminus (Cys30 and Cys31). This π-helix conformation is predicted to be unfavorable and unless it is associated with a functional advantage, would tend to be naturally selected against [298]. This suggests that the region where the π-helix is found may be the location of the active functional site of lybatides, as the π-helix has been shown to play a functional role within certain protein binding sites [299].

Table 4.5. TM-alignment score of cysteine-rich peptides with α-helix in the structure.

Peptide	Family	TM-alignment score
Crambin	Thionins	0.36316
B-purothionin	Thionins	0.39036
Viscotoxin	Thionins	0.37814
NaD1	Defensins	0.35518
PhD1	Defensins	0.33433
VrD1	Defensins	0.35667

4.3.2 Biosynthetic pathway of lybatides

Two homologous sequences with 78.8% (OSMU-2070462) and 69.7% (OSMU-2073250) similarity to lyba2 was obtained when performing a sequence query using t-BLASTn with lybatide sequences on the oneKP transcriptomic database (www.oneKP.com). From the homologous sequences obtained, we hypothesized that the precursor of lyba1 and lyba2 may present a similar a three-domain arrangement: a mature domain following the signal peptide and ending with a C-terminal tail of about 12-15 amino acids (Figure 4.9). This precursor structure arrangement is similar to the thionin precursors, where the signal peptide and an acidic C-terminal tail flanks the mature domain [23]. The three-domain precursor arrangement is also commonly observed in plant CRPs including cystine-knot α-amylase inhibitors, hevein-like peptides and knottins [26,36] as well. However, two types of precursor structure organization exist in defensins. A similar three domain organization can be observed in plant C8 class II defensins as lybatides and thionins, whereas the plant class I defensin precursors do not contain the C-tail domain [30].

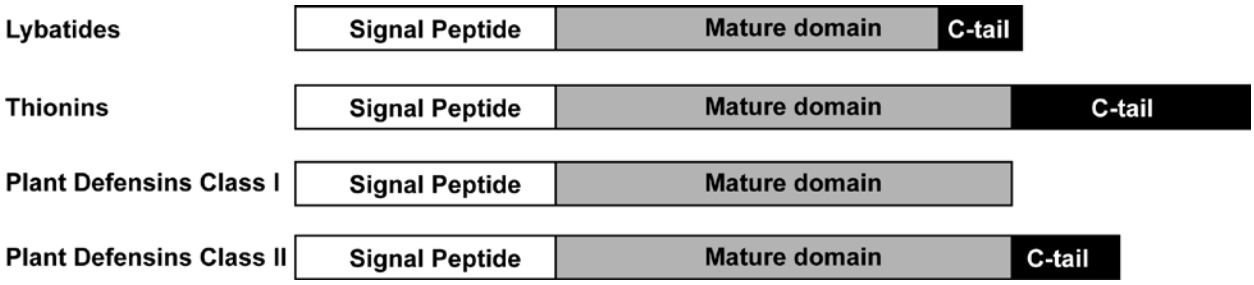


Figure 4.9. Precursor gene structure organization of lybatides. (A) Predicted precursor structure of lybatides based on homologous sequences. A signal peptide followed by the mature domain and a c-tail of about 12-15 amino acids. (B) Precursor structure of Thionins, where the mature domain is flanked by the signal peptide and an acidic c-tail. (C) Precursor structure of the plant class I defensin. They do not contain the c-tail domain (D) Precursor structure of the plant C8 class II defensins have a similar three domain organization as lybatides and thionins.

4.3.3 Lybatide as a helical peptide-displayed scaffold

Naturally occurring constrained peptides typically provide excellent scaffolds for peptide drug design and engineering as they generally evolved into shapes that are able to fit precisely into the binding pockets of their targets as well as being highly stable [287]. A feature that is of potential use for peptide engineering in lybatides is the three-turn (α+π)-helix. The α-helix is a structural motif that is commonly found in many bioactive peptides. Development of chemical methods for stabilization of bioactive conformations remains an active area of research [300]. Some of the approaches used to reinforce α-helical conformations include constraints by disulfide bonds [301], lactam bridges [302] and all-hydrocarbon crosslinks to create stapled peptides [18] (Figure 4.8E). In certain cases of stapled peptides, an increase in helical propensity often results in enhanced target-binding and metabolic stability. It has also been shown that membrane permeability increases which allows them to interrupt protein-protein interactions inside the cell. In this regard, lybatides may be considered to be naturally-occurring “stapled peptides”. However, there are some subtle differences. Intra-helix stapling of side chains of a helical strand is used to stabilize the helix in all-hydrocarbon stapled peptides, while cystine-stapled helical lybatides utilizes inter-strand disulfide bonds to stabilize the helical conformation.

The discovery of a new CRP scaffold from nature holds potential for use in protein and peptide drug engineering. Pease et al., [294] demonstrated that if the contact surface is exposed correctly and the secondary structure is correctly stabilized, disulfides used for stabilization will not necessarily affected the activity of a particular peptide fold. This may allow lybatides to serve as a scaffold, increasing the stability of

peptidyl drugs. Such concepts have been proven in the grafting of a bradykinin antagonist peptides onto the cyclotide kalata B1 scaffold [152]. In the case of lybatides, grafting of helical peptides may increase its helical conformation, resulting in enhanced activity and metabolic stability. In addition, the presence of the π-helix in the lybatide scaffold would allow for additional peptide drug engineering capacity, especially where a π-helix is needed rather than the conventional α-helix for binding to the target protein.

4.4 Conclusion

Lybatides exhibit a novel structural motif and cysteine framework that has not been reported in plant CRPs with MW <4 kDa. Different cysteine patterns may strongly influence the resulting three dimensional structures of cysteine-rich peptides. Within families of CRPs with similar disulfide arrangements, conservation of the three-dimensional structure, even when the overall sequence homology is low, is relatively strong. The discovery of lybatides expands the number of different families of plant CRPs that are known to occur in nature. These findings expand the molecular and structural diversity of plant CRPs with four to ten cysteine residues. The discovery of the lybatide scaffold increases the current repertoire of CRP-based templates that could be utilized as a cystine-stabilized helical template in further refinement in the design of peptide therapeutics and bioengineering [111].

Chapter 5 Elongtide: 10C-hevein-like peptides from the root of Tongkat Ali, *Eurycoma longifolia.*

5.1 Introduction

Eurycoma longifolia jack, commonly known as tongkat ali or Malaysian ginseng, is a dioecious evergreen tree from the Simaroubaceae family and native to the South East Asia tropical forests [185]. *E. longifolia* roots has been used as a local folk medicine for treating fever, malaria, ulcers and headache [303]. A broad range of small-molecule and secondary metabolites are found in *E. longifolia* roots, including quassinoids, eurycomanone and eurycomanol [183,304]. *E. longifolia* roots extract have been shown to exhibit anticancer [305], antimicrobial [306], cytotoxic [307], aphrodisiac [308] and antimalarial properties [309]. For example, three alkaloids, namely 9-methoxycanthin-6-one, canthin-6-one and 9-O-β-glucopyranoside, were demonstrated to be cytotoxic to human lung cancer cells (A-549) [309]. However, Tongkat Ali is sold mainly based on its putative aphrodisiac appeal.

Our interest in the discovery and design of orally active peptides has led us to undertake a program to study cysteine-rich peptides (CRPs) in medicinal plants [23,85]. CRPs, particularly those containing six, eight and ten cysteine (6C-, 8C- and 10C-CRPs) with molecular weights ranging from 2 to 6 kDa, are highly constrained by disulfide bonds, resulting in an extraordinarily compact structure [85]. These properties offer disulfide-constrained CRPs in the chemical space of 2-6 kDa high metabolic stability and potentials as leads for developing orally-active compounds [35,65,310].

Hevein and hevein-like peptides, containing 29 to 45 amino acids, are cysteine-rich peptides which fall into our chemical space of interest. Hevein, the major protein found in the latex of the rubber tree, *Hevea brasiliensis*, was isolated in 1960 [100]. Hevein and hevein-like peptides distinguish themselves from other CRPs with the presence of a chitin binding domain which allows them to bind to chitin, a polysaccharide found in the cell wall of fungi and in the exoskeleton of invertebrates [133,311]. Currently, more than 40 hevein-like peptides were reported from 14 different plant species in 10 different families, including several commercially important agricultural crops such as Poaceae (rice), Amaranthaceae (beet), Ginkgoaceae and Moringaceae [23,36,107,124,133]. They are isolated from various above-ground plant parts, including leaves, seeds and latex, but none from roots.

Hevein-like peptides are broadly classified based on the number of cysteine residues present, namely 6C-, 8C- and 10C-hevein-like peptides [23] (Table 5.1). The 8C-hevein-like peptides which fall between the 6C- and 10C- hevein-like peptides are regarded as the prototypic members because of their bipartite arrangement. Similar to the 6C- and 8C-hevein-like peptides, the N-terminal 30 residues of the 10C-hevein peptides contain a cystine core three-disulfide cystine knot and a chitin-binding domain at intercysteinyl loop 3 and 4. Similar to the 8C-hevein-like peptides, the C-terminus of the 10C-hevein peptides contain an extender fragment of 12-16 residues with one (8C-HLP) or two (10-C HLP) disulfide bonds [134]. In contrast, 6C-hevein-like peptides contain only the cystine core and no extender at the C-terminus, thus they are usually considered as a truncated version of 8C- and 10C-hevein-like peptides [134,312]. The presence of three to five disulfide bonds in hevein-like peptides results in a highly

compact disulfide-constrained structure which is known to be resistant to chemical, thermal and proteolytic denaturation [26,35,36,70]. These characteristics of hevein-like peptides and other plant CRPs suggests that they may be potential bioactive compounds in traditional medicines which are often prepared as decoctions and are administered orally [23].

Five 10C-hevein-like peptides were studied from three different plant families, including Ee-CBP from the spindle tree *Euonymus europaeus* of the Celastraceae [313], WAMP from *Triticum kiharae* of the Poaceae [124] and EAFP from *Eucommia ulmoides* of the Eucommiaceae [314]. The location of the fifth disulfide bond of the 10C-hevein-like peptide subfamily is promiscuous. It can be located either entirely in the C-terminus or the C-terminus and a loop in the cystine core. The fifth disulfide bond in Ee-CBP is formed between the C-terminus and loop 5 [313], while in EAFP, it is between loop 1 and 7 and in WAMP, it is between the C-terminus and loop 2 [314].

In the present study, we report the isolation and characterization of two 10C-hevein-like peptide, designated as elongtide eL1 and eL2, from the root of *E. longifolia*. Elongtide, eL1 consists of 10 cysteine residues, a chitin-binding domain and an N-terminal pyroglutamic acid. The presence of the multiple disulfide bonds confers eL1 resistance against thermal, acidic and proteolytic degradation. The discovery of eL1 furthers our understanding of their distribution and the occurrence amongst different plant species and provides insight into the molecular diversity of anti-pathogenic peptides in plants.

Table 5.1. Representative members of different sub-classes of hevein-like peptides

Peptides	Amino Acid Sequence
6C-HLPs	
Ac-AMP1	-VGECVRG----RCPSGMCCSQFGYCGKGPKYCG----------
Ar-AMP	-AGECVQG----RCPSGMCCSQFGYCGRGPKYCG-R----------
aSG1	APGQCNHG----RCPSGLCCSQYGYCGTGPAYCG----------
aSR1	-VGECVQG----RCPPGLCCSRFGYCGTGPAYCG----------
8C-HLPs	
Hevein	--EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKD--
Gb1	-DPTCSVLGDFK-CNPGRCCSKFNYCGSTAAYCG-PGNCIAQCPS--
mO1	--QNCGRQAGNRACANQLCCSQYGFCGSTSEYCSRANGCQSNCR-GG
Pn-AMP	--QQCGRQASGRLCGNRLCCSQWGYCGSTASYCG--AGCQSQCR---
10C-HLPs	
EAFP1	--QTCASRCP-RPCNAGLCCSIYGYCGSGNAYCG-AGNCRCQCRG--
EAFP2	--QTCASRCP-RPCNAGLCCSIYGYCGSGAAYCG-AGNCRCQCRG--
WAMP1a	-AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCG-AGSCQSQCRGC-
WAMP1b	-AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCG-AGSCQSQCRGCR

5.2 Results

5.2.1 Screening for Peptidyl Natural Products in *E. longifolia*

To determine the presence of CRPs in different parts of the E. longifolia plant, leaves, branches and roots were extracted with distilled water, fractionated with C18 solid phase extraction (SPE) and analyzed with MALDI-TOF MS. The mass spectrum profile of the root aqueous extract showed that a peptide with m/z of 4313, designated elongtide 1 (eL1) was expressed in the root and not the other plant parts of *E. longifolia* (Figure 5.1). Using hot-water extraction, no substantial signal of CRP in the mass region of 3 – 5 kDa was detected in the leaves, twigs, fruits and seeds of *E. longifolia*. Elongtide, eL1 was then reduced with dithiothreitol and *S*-alkylated with iodoacetamide. A mass shift of 1160 Da was observed, confirming the presence of five disulfide bonds.

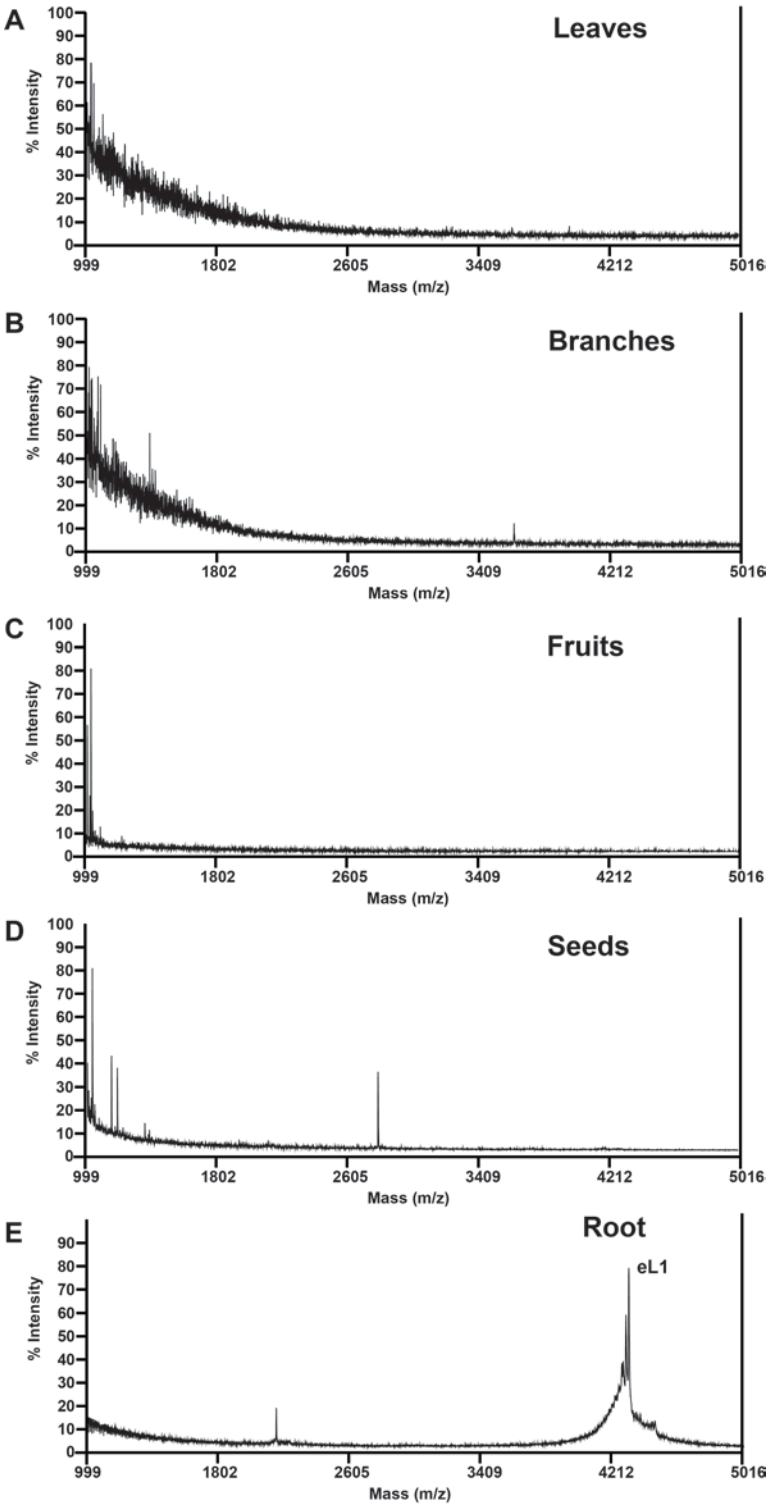


Figure 5.1 Mass Spectrometry profile of E. longifolia aqueous extract from different plant parts. The MALDI mass spectra show a peak at m/z 4313 in the aqueous extract of the root. No peaks at the 3-4 kDa range were detected at the (A) leaves, (B) branches, (C) fruits and (D) seeds.

5.2.2 Isolation and Sequence Determination of Elongtides from *E. longifolia*

A scale-up extraction of 500 g of *E. longifolia* root was performed to obtain a larger amount of peptide for sequencing. Two peptides with m/z 4313 and 4330, eL1 and eL2 respectively, were isolated and purified from the aqueous extract with several rounds of RP-HPLC and SCX-HPLC. Elongtide eL1 was then S-reduced and S-alkylated with dithiothreitol and iodoacetamide respectively, followed by digestion with trypsin or chymotrypsin. The sequences of the peptide fragments were determined with nanospray MS/MS. Analysis of the tandem MS spectra of the peptide fragments revealed the amino acid sequence as pyroQRCSAKKHCPGGECCSQFGYCGLSRAYCCKGCVADCGQIDC (Figure 5.2). Elongtide eL2 was also isolated with the same sequence as eL1 with glutamine instead of pyroglutamic acid at the N-terminus. Transcriptome data was used to determine the isobaric amino acid residues such as Ile/Leu. The sequence of eL1 revealed that it displays a chitin binding domain of Ser16, Phe18, Tyr20 and Tyr27.

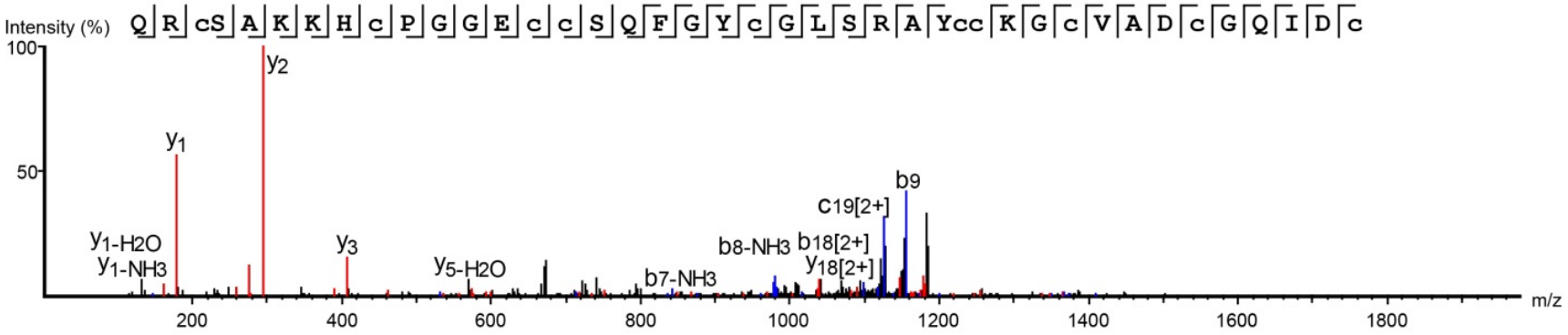


Figure 5.2. Mass spectra of elongtide, eL1, from LC-ESI-LTQ-Orbitrap MS/MS in positive ion mode. The purified eL1 was S-reduced by 20 mM dithiothreitol, S-alkylated with 200 mM iodoacetic acid and subsequently de-salted by C18 Ziptip. Assignments of isobaric amino acids such as Leu/Ile were confirmed by the transcriptome.

5.2.3 NMR Structure

The backbone assignment was done using 2D NOESY and TOCSY spectra. The solution structure of eL1 (Figure 5.3A) was determined with the restraints on distance, dihedral angle and hydrogen bonds derived from TOCSY and NOESY (Table 5.2). For each residue except for Pro, amide proton HNi had NOE correlation peaks with the side chain protons from the previous residue i-1. This allowed us to assign the peaks in NOESY and TOCSY spectra unambiguously. The structure of eL1 contains two short β strands (Cys14-Ser16 and Tyr20-Gly22) and one short helix (Arg25-Cys28). There is one proline (Pro10) in eL1, which adopts a trans conformation as confirmed by the presence of the NOE cross peak between Hδ of Pro10 and Hα of Cys9. Elongtide, eL1 contains five disulfide bonds. For each disulfide bond, the Hβs of the two cysteines had correlation peaks in NOESY spectrum (Figure 5.4). The disulfide bond CysI–CysIV and CysII–CysV ties the N-terminus to the strands β1 and β2. The disulfide bond CysIII–CysVI restricts the helix near the strand β1. The disulfide bond CysVIII and CysIX makes the C terminal tail fold like a clamp. The disulfide bond CysVII–CysX connects the C terminus to the helix (Figure 5.3B). These five disulfide bonds maintain the structure of eL1, and are responsible for the high stability of eL1. There are six positive charged residues (Arg2, Lys6, Lys7, His8, Arg25 and Lys30) and three negative charged residues (Glu13, Asp35 and Asp40) in eL1. The three negative residues are located opposite to the chitin-binding site. The positive charged residues are mainly distributed in the N- and C-terminii. The aromatic rings of Phe18, Tyr20 and Tyr27 are close to each other (Figure 5.3C). Hydrophobic interaction between these side chains likely contribute to the stability of chitin-binding site.

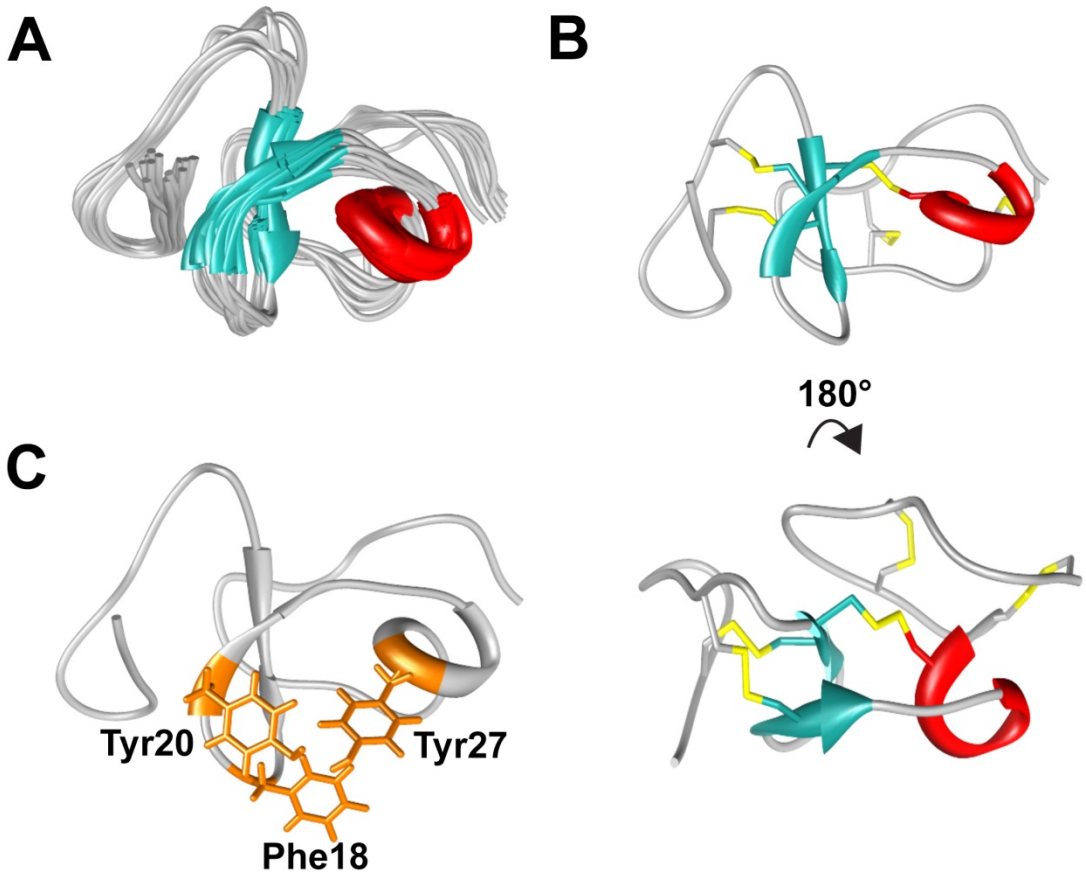


Figure 5.3. NMR solution structure of eL1. (A) Superposition of the eL1 backbone traces from the final 20 ensembles solution structures and restrained energy minimized structure. (B) The ribbon representation of the gB5 structure. The disulfide bonds are formed between CysI-CysIV, CysII-CysV, CysIII-CysVI, CysVII-CysX and CysVIII-CysVIX. (C) The residues Phe18, Tyr20 and Tyr27 have been shown to play an important role in binding toward chitin.

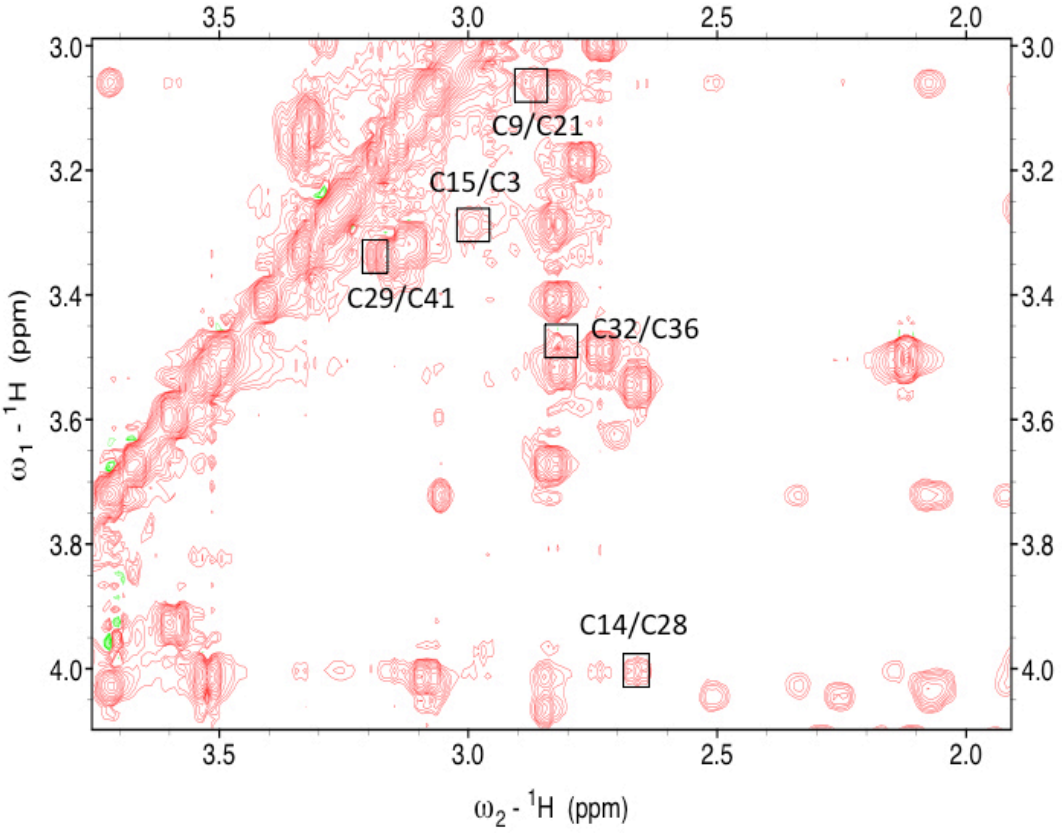


Figure 5.4. 2D NOESY spectrum eL1. The NOE cross peaks between Hβs of Cys3/15, Cys9/21, Cys14/28, Cys29/41 and Cys32/36 displayed by Sparky 3.115.

Table 5.2. Parameters of NMR structure calculation of eL1

Experimental Restraints and Structural Statistics of 20 Lowest-Energy Structures of eL1 among the 100 Structures Generated by CNSsolve 1.3		
NMR Distance Restraints		
	Intra-Residue NOE	*132*
	Sequential NOE	*140*
	Medium-Range NOE	*180*
	Long-Range NOE	*102*
	Hydrogen Bonds	*2*
Dihedral Angle		*5*
Structural Statistics (41 residues, D^{1}-P^{41})		
Violations per Structure		
	NOE Violation (Å)	*0.026±0.002*
	Maximum NOE Violation(Å)	*0.029*
Ramachandran Plot Region (41 residues)		
Residues in Most Favored Regions	*21*	*65.6%*
Residues in Additional Allowed Regions	*10*	*31.2%*
Residues in Generously Allowed Regions	*1*	*3.1%*
Residues in Disallowed Regions	*0*	*0%*
Number of End-Residues (excl. Gly and Pro)	*2*	
Number of Glycine Residues	*6*	
Number of Proline Residues	*1*	
Mean RMSD from the Average Coordinates (41 residues, Q^{1}-C^{41})		
	Backbone Atoms(Å)	*0.93±0.21*
	Heavy Atoms(Å)	*1.64±0.22*

5.2.4 Chitin binding activity of eL1

A chitin binding assay was performed to determine the chitin binding activity of eL1. In order to determine the amount of unbound peptides in the supernatant after incubation with chitin beads at 4 °C, the absorbance of the supernatant was measured at 280 nm. UPLC analysis showed that eL1 binds to chitin within 1 h (Figure 5.5A). Subsequently, the bound peptides were eluted by 0.5 M acetic acid. Figure 5.5B and 5.5C show that eL1 binds strongly to chitin and can only be eluted at 1 M NaCl or heating at 100 °C with 0.5 M acetic acid.

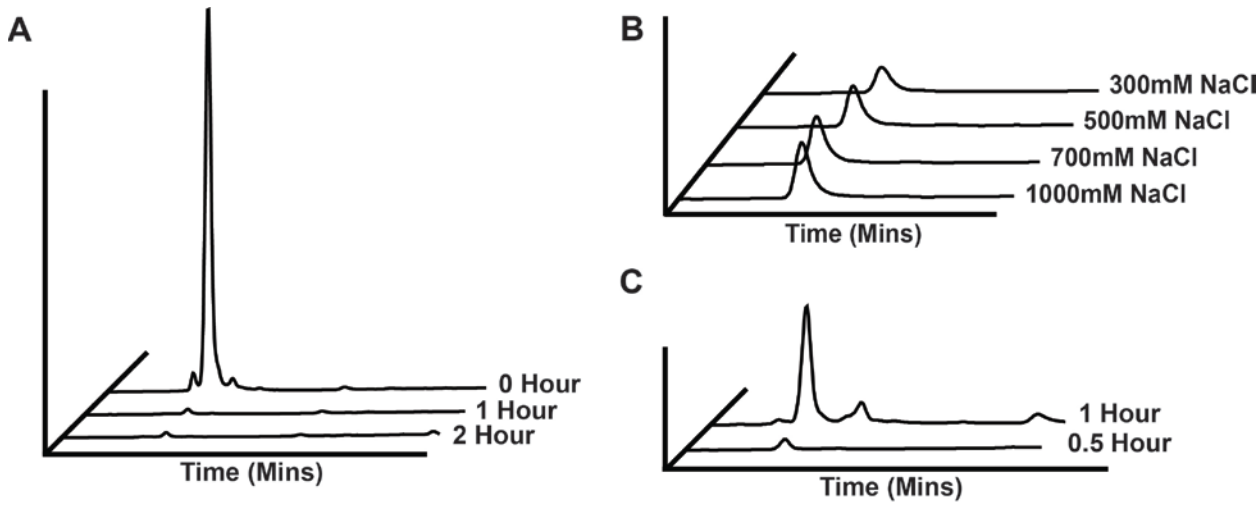


Figure 5.5. Chitin binding studies of eL1. (A) UPLC chromatogram showing the amount of eL1 detected in the supernatant of tubes 1 (0 Hour), 2 (1 Hour), and 3 (2 Hour). At 1 hour, up to 98.9% of eL1 was bounded to the chitin beads. At 2 hour, up to 99.1% of eL1 was bounded to chitin. (B) Elution of eL1 with sodium chloride buffer. The amount of eL1 eluted increased with increasing salt concentration, however, only 25% of eL1 is eluted at 1 M sodium chloride concentration. (C) Elution of eL1 with 0.5 M acetic acid and heated at 100°C, 1.09% of eL1 was eluted after 30 min and an additional 16.2% of eL1 was eluted after 1 h.

5.2.5 Metabolic Stability of eL1

Several stability assays were performed to demonstrate the stability of eL1 against thermal, acidic and enzymatic degradation. UPLC monitoring showed that eL1 is stable against heat degradation with more than 79% of eL1 remaining intact after incubation at 100 °C for 1 h (Figure 5.6A). Elongtide eL1 is also stable in 1 M of HCl, pH 2.0, and 80% remaining intact after incubating for 5 h (Figure S2B). Similarly, eL1 is resistant to tryptic digestion with >90% remaining intact after incubation with trypsin for 5 h (Figure S2C). The linear peptide RV-14 was used as a control. Stability assays against exopeptidases using aminopeptidase and carboxypeptidase A showed that eL1 displayed high stability against these enzymes, with >90% remaining intact after 18 h of treatment (Figure 5.6D and 5.6E).

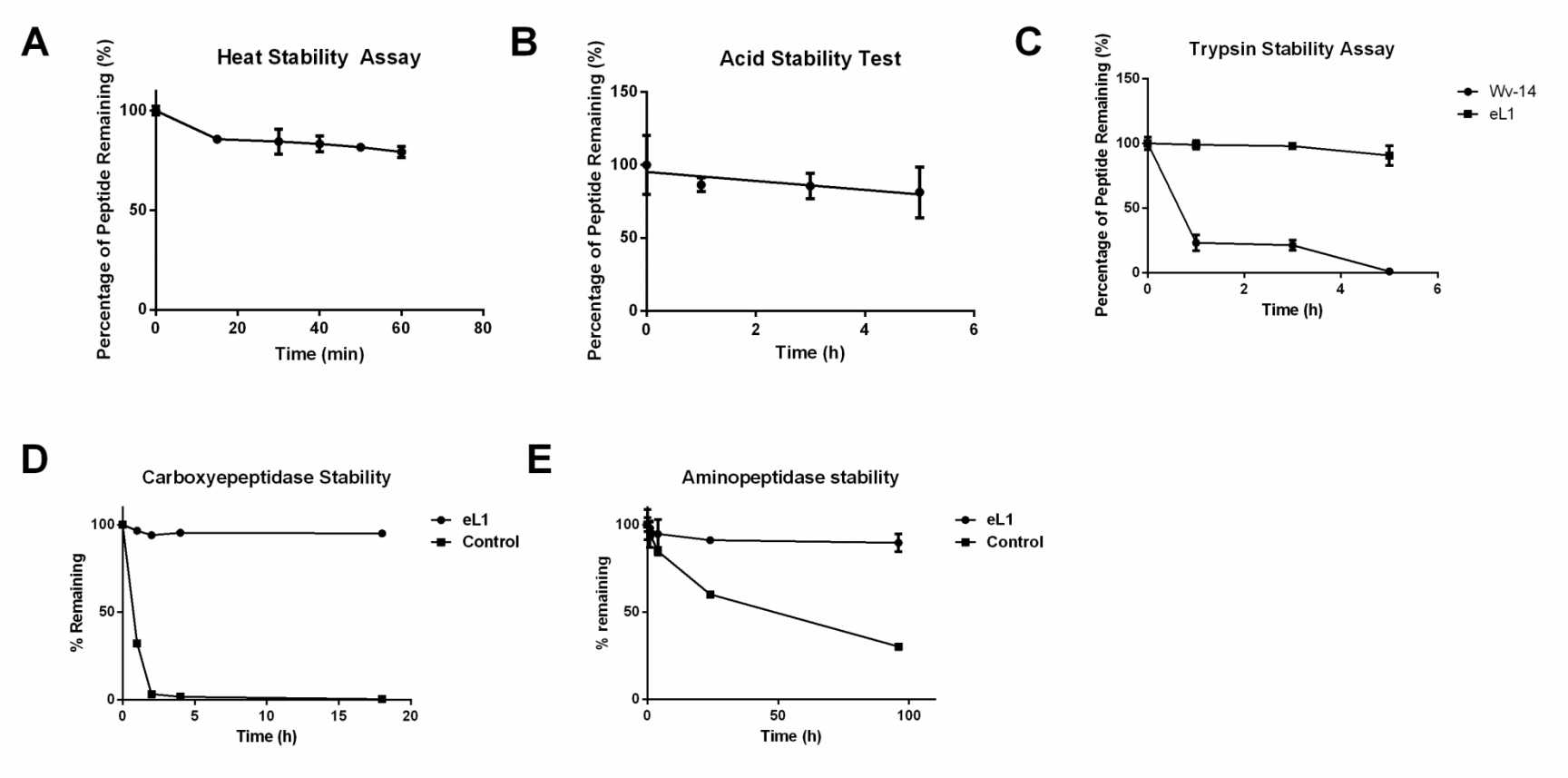


Figure 5.6 Stability assays of elongtides, eL1. (A) Acidic condition stability of eL1 incubated in 1M HCl (pH 2.0) for 5 h. (B) Thermal stability of eL1 incubated at 100◦C for 1 h. Exo- and endo-peptidase enzymatic stability of gB5 against (C) trypsin, (D) carboxypeptidase and (E) aminopeptidase, respectively for 5, 18 and 96 h in buffer as suggest by manufactures at 37◦C. RV-14 was used as control for enzymatic stability assays.

5.2.6 Anti-fungal Activity of eL1

In order to screen for potential anti-fungal activity of eL1, disk diffusion assay was performed. Table 5.3 summarizes the effect of eL1 against the mycelium growth of the five fungal strains. After 24 h incubation, a zone of inhibition was observed in the rim of paper disk, demonstrating that eL1 exerted anti-fungal activity against all five fungal strains. Micro-broth dilution assay was performed to further quantify the anti-fungal activity of eL1. The half maximal inhibitory concentration (IC50) of eL1 against *C. lunata*, *F. oxysporum*, *A. brassicicola*, *A. alternate* and *R. solani* were 87.1, 9.72, 26.78, 31.85 and 5.12 µg/mL after 24 h incubation at 25 °C.

Figure 5.7 illustrates the changes in the morphology of *A. brassicicola* prior and after eL1 treatment at 25 °C for 24 h. Treatment of eL1 resulted in shorter and highly branched hyphae when compared to control experiment (Figure 5.7A). Swollen hyphal tips and germination of fungal spores were also observed (Figure 5.7B). These morphological changes were observed to occur in a concentration-dependent manner.

To determine if the anti-fungal activity of eL1 is selective only to plant pathogenic fungal strains, radial diffusion assay was done on eL1 against *C. albicans* and *C. tropicalis*, two common human pathogenic strains. After overnight incubation, the clear zone of inhibition surrounding each well was measured and the MIC of eL1 against *C. albicans* was determined. Figure 6 shows the graph of the log concentration of eL1 and the diameter of the zone of inhibition observed. The MIC of eL1 against *C. albicans* and *C. tropicalis* was determined as 26.1 µM and 13.9 µM, respectively.

Table 5.3. Antifungal activity of eL1 on 5 common strains of plant pathogenic fungi

Fungal Strains	Inhibition (Disk diffusion Assay)	IC50 (Microbroth dilution assay) (µg/ml)
Fusarium oxysporum	**+**	**9.72**
Curvularia lunata	**+**	**87.1**
Rhizoctonia solani	**+**	**5.12**
Alternaria brassicola	**+**	**26.78**
Alternaria alternate	**+**	**31.85**

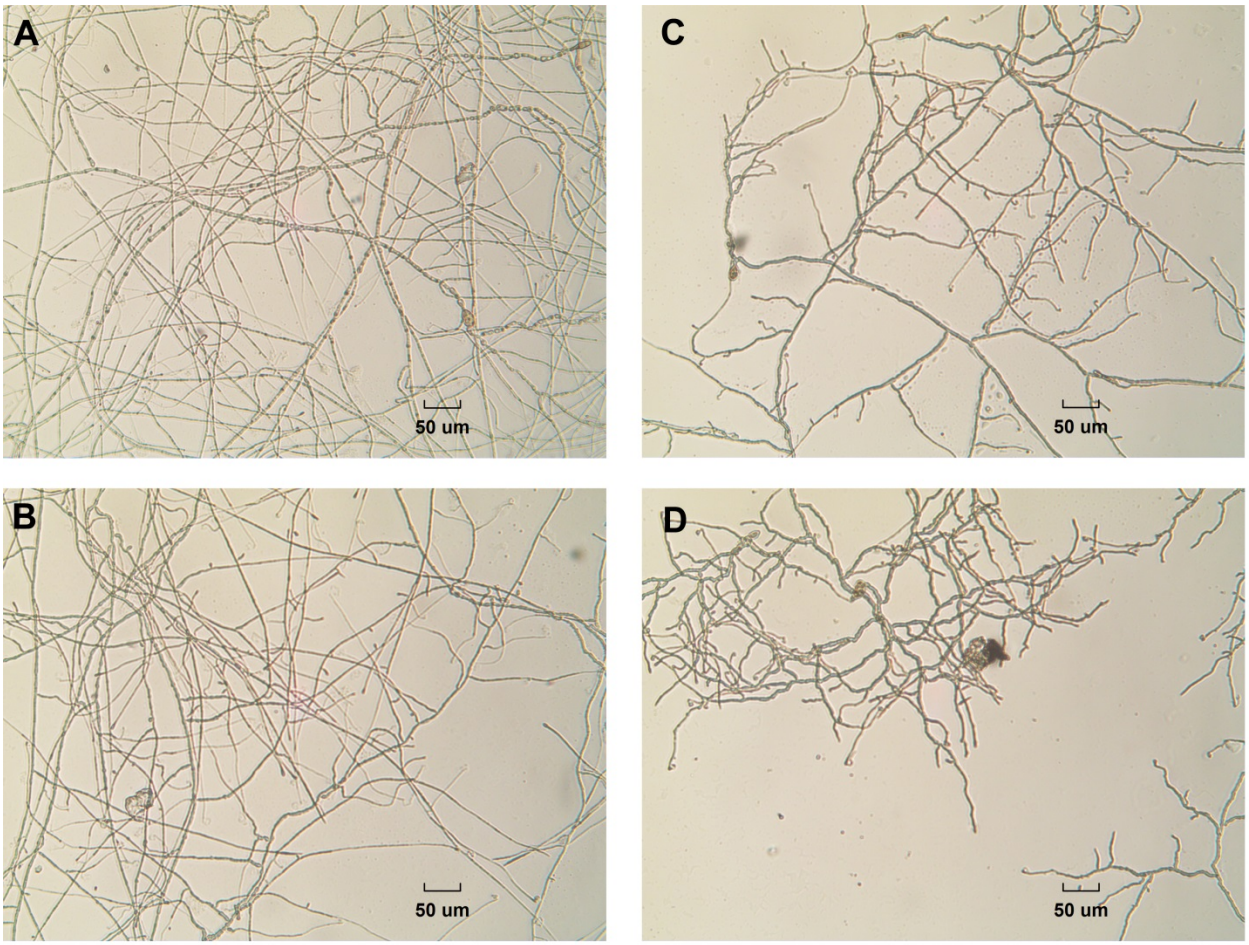


Figure 5.7. Microscopic views of the mycelium growth of *A. brassicicola* (100X). The well without the addition of eL1 (A) was used as a control and were co-incubated with (B) 25, (C) 100 and (D) 400 µg/mL eL1 in half strength potato dextrose broth at 25◦C for 24 h.

5.3 Discussion

Various studies have shown that hevein-like peptides act as antimicrobial agents, implying a role in the defense system of the plants [315]. The expression of hevein-like peptides in various parts of the plants and their cysteine-rich nature, similar to other cysteine-rich antimicrobial peptides, including thionins and cyclotides may indicate that they are involved in host defense. However, it is not certain if this is their primary function in plants. In this study, two root specific 10C-hevein-like peptides, eL1 and eL2, were isolated and characterized from the root of *E. longifolia* of the Simaroubaceae family. Unlike other hevein-like peptides, eL1 and eL2 are expressed in the roots of the plant and are not detected in the aerial parts. In contrast, cyclotide expression is observed to occur in almost all plant parts in Violaceae plants [315]. Since all hevein-like peptides isolated to date were isolated from aerial parts of the plant, the discovery of elongtides in the underground part of *E. longifolia* demonstrates tissue specific expression of hevein-like peptides. Additionally, this demonstrates that the underground parts of plants may be an underexplored region in the discovery of hevein-like peptides.

5.3.1 Distribution of 10C-HLPs

This is the first time hevein-like peptides were isolated from a plant in the Simaroubaceae family. Instead of being restricted in a selected number of plant families, distribution of hevein-like peptides may be more ubiquitously expressed in plants than previously thought. This may be due to much research focusing on hevein-like peptide before, however, as more research is done on hevein-like peptides in plants, more hevein-like peptides are discovered to be expressed in a much broader families of plants. The discovery of the presence of hevein-like peptides in both gymnosperms and

angiosperms seem to suggest that hevein-like peptides may be much more evolutionarily ancient. Similar to plant defensins, where it is apparently ubiquitously expressed throughout the plant kingdom [37], hevein-like peptides may be regarded as a family of plant defense peptides complementary to plant defensins in plants.

5.3.2 Sequence Comparison with other 10C-HLPs

Similar to other 10C-hevein-like peptides, eL1 contains 10 cysteine residues and a highly conserved chitin binding domain. Currently, only five other 10C-hevein-like peptides have been characterized, including Ee-CBP from *E. europaeus*, WAMP1 and WAMP2 from *T. kiharae* and EAFP1 and EAFP2 from *E. ulmoides* [124,313,314]. The cysteine framework of eL1 is similar to Ee-CBP, with a cystine connectivity of CysI–IV, CysII–V, CysIII–VI, CysVII–X and CysVIII–IX. However, apart from the cysteine spacing and the chitin binding domain, the remaining amino acid residues are highly variable with only 16.1 % similarity between eL1 and Ee-CBP. This demonstrates the hypervariability of the hevein-like peptide scaffold, being able to tolerate a variety of amino acid residues within the intercysteinyl loops.

5.3.3 Structural Comparison with other 10C-HLPs

This is the first time the molecular structure of a 10C-hevein-like peptide displaying a cysteine framework similar to Ee-CBP is elucidated. The disulfide bonds stabilize the global fold of the peptide in addition to the secondary structures and hydrophobic interactions, similar to other hevein-like peptides. Like other hevein-like peptides, eL1 displays a very compact structure, in which the disulfide bonds occupy the core of the structure. The cystine core topology in hevein-like peptides is

characterized by highly compact disulfide crossbracings. This motif may confer an evolutionary advantage to peptides that contains it as the cystine core has been suggested to act as a folding nucleus [123].

The overall fold of eL1 contains at least two anti-parallel β strands and a short α-helix, a common characteristic reported for several other hevein-like peptides and chitin-binding proteins including hevein [117], EAFP2 [316] and the chitin binding domain of the plant class 1 chitinase from rice [317]. Similar to EAFP2, the location of the hydrophobic and charged residues in eL1 result in a hydrophilic and hydrophobic patch on the opposite sides of the peptide conferring it an amphiphilic nature.

The fifth disulfide bond in eL1 (CysVII–X), similar to Ee-CBP remains at the C-terminal half of the peptide (Cys29-Cys41), connecting the C-terminal tail to loop five, unlike in EAFP (CysII–CysIX, connecting loop one and six) and WAMP-1a (CysIII–X, connecting loop 2 and the C-terminal tail) (Table 5.4). The fifth disulfide bond being wholly located only after the cystine core in eL1 would grant higher flexibility to the C-terminal tail of the peptide as compared to EAFP and WAMP-1a as it does not tether the C-tail to the cystine core of the peptide. Comparison between the 10C-hevein-like peptide structures suggests that the position of the fifth disulfide bond changes the overall topology of the molecular structure. Figure 5.8 shows how the position of the fifth disulfide bond changes the overall shape and charge distribution of the peptide. The C-terminal portion of the peptide wraps around different part of the cystine core depending on where the fifth disulfide bond tethers it to. Interestingly, none of them is positioned near the hydrophobic chitin binding domain of the peptide, maintaining the bipartite nature of the whole molecule. The distribution of cationic and hydrophobic regions on

the surface of the peptide is similarly affected as the C-terminal portion of the peptide is guided to different area of the structure. The change in the overall topology of the peptide, together with the distribution of surface charges, may affect the specificity of the peptide to different target proteins, thus resulting in the development of new functions within the same peptide family with a conserved core structure.

Table 5.4. Comparison of the primary peptide sequences of 10C-hevein-like peptides.

Peptide	Amino Acid Sequence	Mass (Da)	Charge	Approach
	N-Terminal Head | Loop 1 | Loop 2 | Loop 3 | Loop 4 | Loop 5 | Loop 6 | C-terminal Tail			
eL1	-QRCSAKKH---CPGGECCSQFGYCGLSRAYCCKG--CVADCGQIDC	4313	+1	T, P
Ee-CBP	-QQCGRQAGNRRCANNLCCSQYGYCGRTNEYCCTSQGCQSQCRRCG-	4992.5	+5	T, P
WAMP1a	AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCGAGS-CQSQCRGC--	4431.7	+3	G, P
WAMP1b	AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCGAGS-CQSQCRGCR-	4587.5	+4	G, P
EAFP1	-QTCASRCP-RPCNAGLCCSIYGYCGSGNAYCGAGN-CRCQCRG---	4201.4	+4	P
EAFP2	-QTCASRCP-RPCNAGLCCSIYGYCGSGAAYCGAGN-CRCQCRG---	4158.9	+4	P

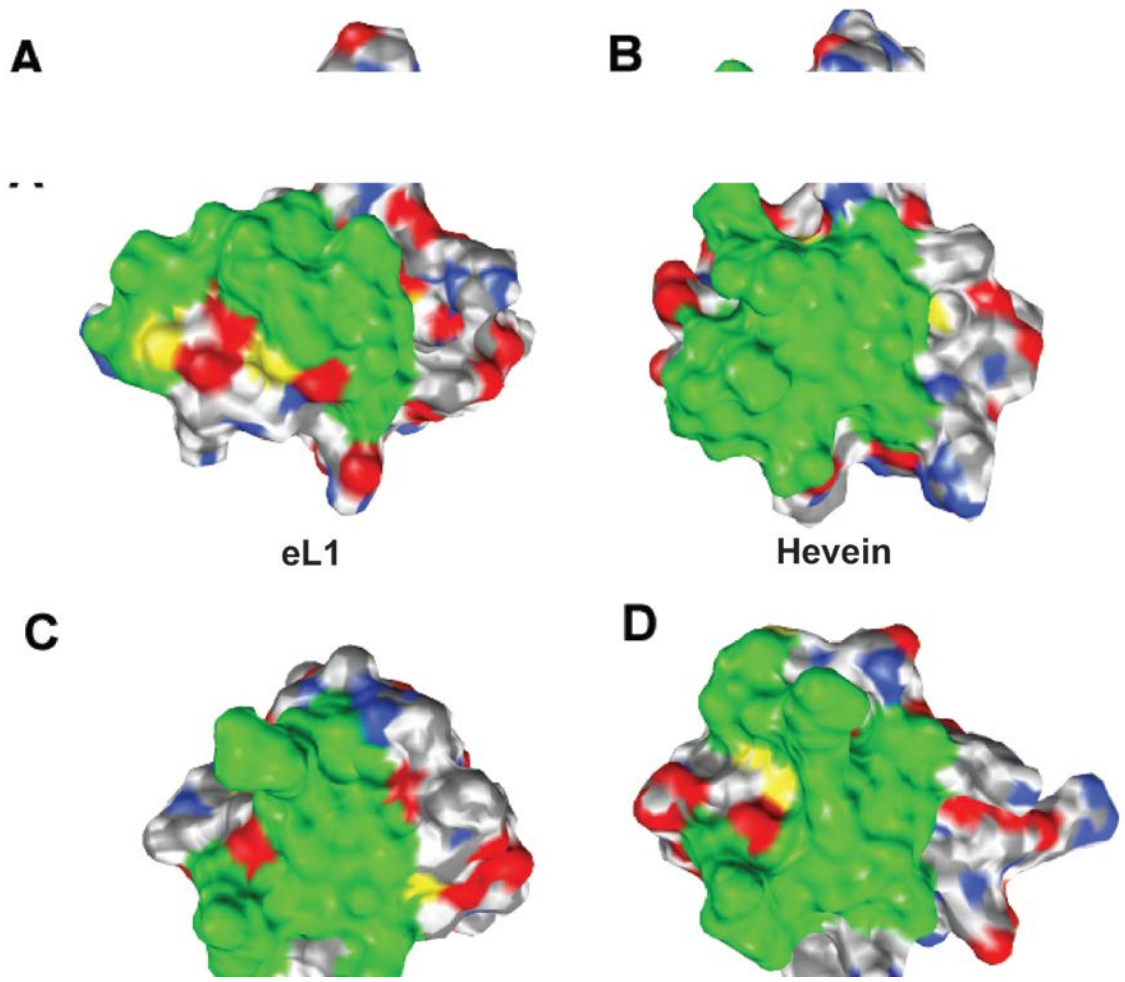


Figure 5.8. Surface topology of the structures of Hevein and 10C-HLPs with reported structure. (A) eL1, (B) Hevein, (C) EAFP1 and (D) WAMP1. The area highlighted green is the region of the backbone after the knottin core. It can be observed that the difference in the position or absence of the fifth disulfide bond affects the overall shape of the peptide, which may affect its binding to different target protein.

5.3.4 Stability of lybatides

Similar to other CRPs, the presence of high amount of disulfide bonds in eL1 increases the stability of the peptide, which enables it to withstand harsh conditions and resist thermal, acidic and enzymatic degradation [20]. This was shown in the stability assays performed where > 90% of the peptide remained after treatment. The presence of a pyroglutamic acid at the N-terminus and the last residue of the peptide being a cysteine confers it high resistance against exopeptidase digestion. This pseudocyclic feature is also observed in other CRPs like wrightides, cystine-knot α-amylase inhibitors isolated from *Wrightia religiosia* [20]. The high stability of eL1 suggests that it can potentially be used in the development of orally active peptidyl therapeutic as it would be able to survive the harsh environment in the gastrointestinal tract.

5.3.5 Anti-fungal activity of eL1

A chitin binding assay was performed to confirm the chitin binding activity of eL1. The results show that eL1 is able to bind to the chitin beads strongly, similar to other hevein-like peptides [29]. One of the major mechanisms thought to contribute to the anti-fungal effect of hevein-like peptides is the ability to bind to chitin [107], with the other being the small size and compact structural fold. This allows them to migrate through the pores of fungal cell wall to the fungal plasma membrane and work its action [313]. The binding of the hevein-like peptide to the chitin polymer present in the cell wall of the fungi prevents nascent chitin chains and β-glucan microfibrils from cross-linking, leading to a disruption of the morphology of the cell wall and inhibiting hyphal growth. The IC50 of the inhibitory effects of eL1 on fungal growth is comparable to other 10C-HLPs [107,124,313,314]. However, the addition of Ca^{2+} to the medium results in a

decrease in anti-fungal activity of eL1. This antagonistic effect of Ca^{2+} is similarly seen in other hevein-like peptides including Ee-CBP [313].

Most hevein-like peptides affect the morphology of germinating spores and/or the growing hypha of fungi in addition to inhibiting their growth [313]. We demonstrated that eL1 exhibit similar effects on the morphology of fungi when the fungal strain was grown in the presence of the peptide. Microscopic observations showed that fungal strains treated with eL1 displayed hyphal branching and stunted growth when compared to controls. These observations are in agreement with the results that were reported for hevein [101] and Ee-CBP [111]. Interestingly, growth inhibition was observed in *C. albicans* and *C. tropicalis*, both human pathogenic fungal strains, with treatment of eL1.

The sequence alignment of eL1 and other HLPs revealed a high conservation of the amino acids in the chitin-binding domain. This chitin binding domain consisting of Ser16, Phe18, Tyr20 and Tyr27 in eL1 is highly conserved and found in many other chitin-binding peptides and proteins, such as plant endochitinase [318], N-acetylglucosamine-specific lectin [108] and zymocin, a tRNA endonuclease [319]. Studies on hevein have shown that the key aromatic residues located within the chitin binding domain, Trp21 and Trp23 in hevein, stabilize the peptide carbohydrate complexes through the stacking of CH-π interactions. Tyr30 interacts through van der Waals interactions with the methyl group of the acetamide moiety of the sugar [320]. In addition, the hydroxyl groups of Tyr30 and Ser19 interacts with the hydroxyl group at third carbon atom of the sugar residue and with the carbonyl group of the acetamide moiety through hydrogen bonds [321,322]. The capacity for chitin binding and the amphiphilic nature of the peptide, where the cationic and hydrophobic residues are

clustered into specific regions of the peptide, was hypothesized to be important for antimicrobial activity. This has been observed in other antimicrobial peptides where the cationic sector is thought to facilitate binding to the negatively charged phospholipid of microbial cell membrane thereby disturbing normal membrane functions [292].

5.3.6 Biosynthesis of eL1 and hevein-like peptides

Hevein-like peptides are ribosomally synthesized and gene-encoded as linear precursors. These precursors are then processed to produce the mature hevein-like peptides. They typically exist in two different forms, the long and short forms. The long form consists of a mature peptide domain flanked by a signal peptide and a long C-terminal tail which includes a hinge domain and a protein cargo. The protein cargo domain usually encodes for a bioactive protein as seen in EAFP, from *E. ulmoides* Oliv, Ee-CBP from the spindle tree *E. europaeus*, and hevein. In contrast, the short form differed from the long form in a short C-tail without a protein cargo. However, analysis of the short C-tail of WAMPs revealed that it may be remnants of a chitinase gene due to a frameshift deletion as a pblast of the nucleotide sequences revealed partial homology to chitinase [323].

Similar to Ee-CBP from the spindle tree *E. europaeus*, eL1 displays a long-form precursor organization (Figure 5.9). Both of their cargo protein domains encode for a catalytic chitinase protein. A study comparing the cDNA of Ee-CBP and Ee-chitinase revealed that the main difference between the hevein-like peptide precursor sequence and the genuine chitinase is at the hinge/linker region. The hinge region of Ee-chitinase is rich in Gly residues whereas in Ee-CBP, no Gly-rich region is observed. Similarly, the

corresponding region of eL1 has no such Gly-rich motif, which may explain why eL1 was not processed with its cargo protein. However, the exact mechanism in which the hevein-like peptide precursor is processed to its mature chitin-binding domain only form remains to be elucidated.

The presence of the signal peptide domain indicates that these peptides follow the conventional pathway for secretory proteins that are commonly seen in CRPs, including towel gourd trypsin inhibitors [271], cyclotides [35,65,86] and other hevein-like peptides [125,323]. This precursor arrangement is similar to thionins but different from other CRPs such as cyclotides and defensins which contain an additional prodomain between the signal peptide and mature peptide domain [20,65].

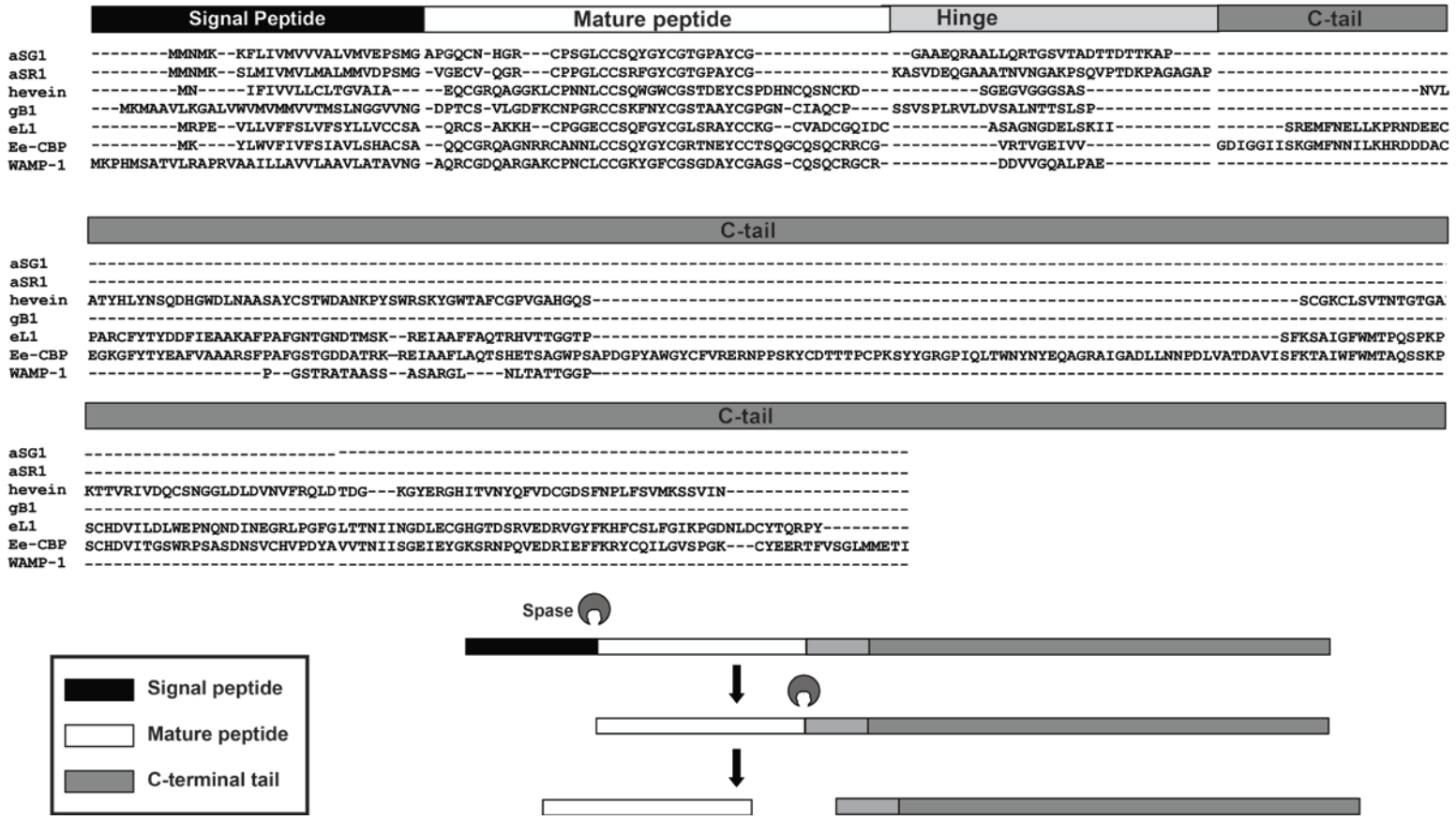


Figure 5.9. Gene alignment and biosynthesis pathway of hevein-like peptides. (A) Precursor sequences alignment of 6C-, aSG1 and aSR1, Hevein, 8C-, gB1, and 10C-, eL1, Ee-CBP and WAMP-1, hevein-like peptides. The precursors are divided into three major domains, including the signal peptide, mature hevein-like peptide domain and C-terminal tail, which contain a hinge domain and a cargo protein domain. (B) Biosynthesis pathways of hevein and hevein-like peptides. Signal peptide was removed from the full precursor sequences by SPase. The C-terminal tail is then cleaved by endopeptidase to release the mature peptides. In hevein and 10C-hevein-like peptides (except WAMPs), the C-terminal tail was coded for a bio-functional protein such as barwin-like protein or chitinase.

The precursor sequences of elongtides and other chitin-binding peptides and proteins in plants were aligned and analyzed by neighbor-joining clustering algorithm to evaluate the relationships between and displayed as a tree in Figure 5.10. From the tree, we can observe that eL1 is highly related to Ee-CBP as they share the same CRP scaffold and their precursor sequences both contain a protein cargo domain. This feature of a cargo protein is usually absent in 6C-HLPs and other 8C-HLPs which feature a short C-terminal tail instead. Hevein is the only reported 8C-hevein-like peptide to contain a cargo protein in the precursor sequence. The long C-tail of hevein encodes a Barwin-like protein that binds to tetrameric N-acetylglucosamine [104]. In contrast, a type I chitinase domain was identified in the C-tail of the 10C-hevein-like peptide, Ee-CBP [125,313]. A conserved domain BLAST on the NCBI database revealed the presence of the Glycosyl hydrolase 17 (NCBI Ascension no. pfam00332) domain in the C-tail of the precursor of eL1. The presence of cargo proteins in hevein and 10C-HLPs and the absence of it in 6C and 8C-HLPs may indicate that 6C- and 8C-HLPs may be truncated versions of 10C-HLPs, which they themselves may be artefacts of class I chitinases. This is seen in WAMP-1 and WAMP-2, where in silico translation of the cDNA in reading frames +3 and +2, respectively, revealed sequences that are highly homologous to the N-terminal portions of the catalytic domains of plant class I chitinases [323].

Similar phenomenon can also be observed in the 8C-hevein-like peptides, mO1 and mO2 isolated from *Moringa oleifera* [24]. The presence or absence of different cargo protein domain in HLP C-terminal precursor may provide clues about the evolution and relationship of hevein-like peptides and other chitin binding proteins,

which contains hevein-like domain. However, larger datasets of the different precursor sequences of different HLPs will be required in order to develop a more conclusive hypothesis.

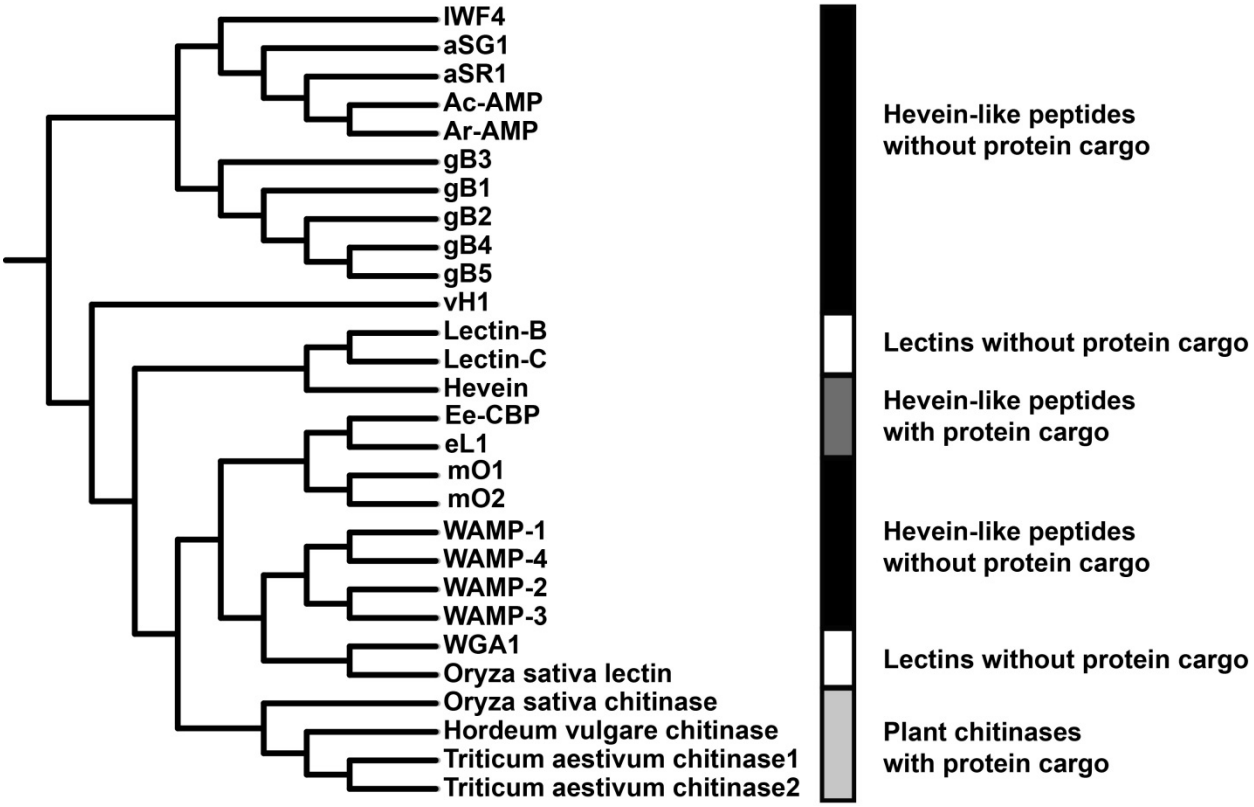


Figure 5.10. Tree of selected peptides and proteins containing chitin binding domains. The precursor sequences were aligned by MAFFT and the phylogenetic tree was generated by iTOL. eL1 was found to be closely related to Ee-CBP.

5.4 Conclusion

Here, we report the discovery and characterization of two root-specific 10C-hevein-like peptide isolated from *E. longifolia*. This is the first time hevein-like peptides are isolated from the roots of a plant. This revealed that hevein-like peptides may be more ubiquitously expressed than previously thought. We have also confirmed the predicted disulfide connectivity and elucidated a structure of the subclass of the 10C-hevein-like peptides which eL1 belongs to through NMR. This would allow us to observe how the promiscuity of the fifth disulfide bond in 10C-hevein-like peptides affects the overall structure of the peptide molecule. The discovery of elongtides and the elucidation of its structure provide insights into the distribution and molecular diversity of hevein-like peptides in planta.

Summary, Conclusion and Future Outlook

Naturally-occurring cysteine-rich peptides play important roles in the plant host-defense and physiology. They include cell signaling, defense and immunity [324]. This class of peptides possesses the advantages of small molecules and large molecules such as proteins. They are metabolically stable and could be potential therapeutics.

CRPs are classified into different families, each with a unique cysteine motif and structural framework [325]. In addition, CRPs are known to display many different biological functions. Prominent examples include the defensins which the Anderson laboratory had reviewed extensively on their structures and functions [326]. Defensins exhibit anti-microbial, enzyme inhibition and ion channel blocking activity. The molecular and structural diversity of cysteine-rich peptides inspired me to isolate and characterize novel cysteine-rich peptides other than defensins with a focus on medicinal herbs like *Lycium barbarum* and *Eurycoma longifolia*. My study examines this particularly neglected chemical space occupied by CRPs which could account for some of the bioactivities in medicinal plants. In addition, I explore their distribution, structures and molecular diversity across and within CRP families.

A summary of my findings are as follows:

Chapter 3 discusses the isolation and characterization of a novel carboxypeptidase inhibitor from the fruit of *Lycium barbarum*, the wolfberry carboxypeptidase inhibitor (WCI). Structural analysis of WCI by 2D-NMR revealed that the three disulfide bonds form a cystine knot which confers high stability against thermal,

acid and enzymatic degradation. Of the two major classes of disulfide connectivity, the cystine knot is by far the dominant disulfide connectivity known amongst the peptidyl natural products found in both plants and animals [327].

The K_i of WCI was determined by a burke-lineweaver plot and found to be 6.62 nM, which is comparable to the potato carboxypeptidase inhibitor (PCI). The amino acid sequence of WCI shows a high conservation with PCI, especially the primary and secondary binding site. This conservation can be further observed in the additional 26 amino acid sequences obtained in the expressed sequence tag transcriptomic database in NCBI.

Analysis of 29 homologous sequences of carboxypeptidase inhibitors obtained from the transcriptomic databases showed that their noncysteine residues are able to tolerate substitutions by a high variety of amino acid sequences. In contrast, the amino acid residues important for the inhibition function of the carboxypeptidase inhibitor are highly conserved. Studying the degree of conservation of these amino acid residues across a large number of homologous sequences may allow us to understand and to predict the importance of each individual amino acid residues contributing to the function of CRPs [328]. This is demonstrated by the high conservation of the C-terminal Val which is responsible for stabilizing the entire enzyme-inhibitor complex after cleavage of the C-terminal amino acid in carboxypeptidase [257].

Chapter 4 describes the isolation and characterization of two novel naturally occurring stapled peptide, called lybatides, from the root bark of *Lycium barbarum*. Lybatides are 32-32 amino acid residue long with eight cysteine residues. Similar to

other cysteine-rich peptides, lybatides was shown to exhibit high stability against heat, acid and enzymatic degradation.

Sequence determination by mass spectrometry revealed that lybatides display a cysteine pattern, C-C-C-C-CC-CC, a cysteine spacing that has not been reported in plant cysteine-rich peptides. The disulfide mapping showed that lybatides presents a new cystine framework in plant cysteine-rich peptides. X-ray analysis revealed that lybatide 2 displays an α-helix as a major structural feature stabilized by disulfide bonds [325]. This is different from the α-helix stabilization found in plant defensins, which contains an α-helix stabilized by β-strands, and in thionins, which contains an α-helix stabilized by another α-helix. Thus lybatides emerges as a new family of α-helical cystine-rich peptides.

Stapling of α-helical peptides would result in improvement of pharmacological performance by increasing affinity to their target, increased metabolic stability while conferring increased levels of cell penetration [329]. Promising examples are the hydrocarbon-stapled α-helical peptides, where all hydrocarbon chains were utilized to locked synthetic peptides into their bioactive α-helical fold [330]. The lybatide scaffold represents a naturally-occurring stapling of an α-helical peptide, which may exhibit similar properties

The characterization of the lybatide scaffold expands the number of CRP scaffold for peptidyl drugs development. The lybatide scaffold allows the grafting of bioactive peptides to increase metabolic stability and conformation preorganization which may enhance target binding in drug development [152].

Chapter 5 describes the isolation and characterization of a novel 10C-hevein-like peptide, named elongtide, eL1 from *Eurycoma longifolia* using proteomic and transcriptomic methods. Elongtide, eL1 is 41 amino acid residues in length. Sequence comparison with previously reported 10C-hevein-like peptides showed that elongtide belongs to the 10C-hevein-like peptide family similar to Ee-CBP [313]. Comparison between reported 10C-hevein-like peptides revealed that the location of the fifth disulfide bond is variable, resulting in a slightly different overall structure and charge distribution.

Chitin-binding assay demonstrated the strong chitin binding of eL1. Elution can only be achieved after heating at 100 °C in acidic conditions. This result suggests that eL1 may be able to inhibit the growth of phytopathogenic fungi. Indeed, elongtide eL1 was demonstrated to inhibit the growth of five phytopathogenic fungi. Morphological observation of *A. brassicicola* before and after eL1 treatment revealed that the fungi displayed shorter and highly branched hyphae. Swollen hyphal tips and germination of fungal spores were also observed. These morphological changes were observed to occur in a concentration-dependent manner. In addition, eL1 was also found to be able to inhibit human pathogenic fungi, *C. albicans* and *C. tropicalis*.

In conclusion, this thesis reports the isolation and characterization of novel CRPs from *L. barbarum* and *E. longifolia*. The WCI, lybatides and elongtides shows how nature introduces molecular diversity within and across different cysteine-rich peptides families.

The molecular diversity of the CRP scaffolds may represent a convergent evolution by nature to devise metabolically stable and compact peptides to solve a vast array of physiological and defense-related problems. Convergent evolution results when selection pressures combined with biophysical restrictions only favoring a few solutions within an adaptive landscape. This leads to natural selection narrowing the number of viable evolutionary lineages resulting in similar adaptations [331,332]. Convergent evolution has been observed in many biological phenomena, however, it is still uncommon at the peptide level [328]. The different families of CRPs with their unique cystine motif suggested peptides with different lineages selecting cystine-stabilized scaffolds as a solution towards stabilizing structural folds. In contrast, the hypervariable sequences and the variation of additional disulfide bonds within a CRP family may be a result of divergent evolution. Mutations and natural selection allows nature to diversify the amino acid sequences in a particular CRP scaffold resulting in a family of CRP performing a multitude of biological functions as can be seen in different knottin-type peptides inhibiting different enzymes when they share the same CRP scaffold [333]. Structural divergence can also be observed in elongtides of my study where the position of the additional disulfide bond results in slightly different structural folds.

The discovery of naturally occurring cysteine-rich peptides would increase our knowledge in the molecular diversity, distribution and occurrence. With increased understanding of the multitudes of CRP scaffolds in nature, it may allow us to design novel scaffolds, inspired by naturally occurring CRP scaffolds, refining them for developing peptide therapeutics [287]. Expansion of the repertoire of CRP scaffolds would allow us to develop metabolically stable and multifunctional peptidyl drugs by

grafting bioactive peptides into these hyperstable CRP scaffolds [152,153]. In addition, due to their stability and functional promiscuity, it is highly likely that they may be responsible for some of the pharmacological effects presented by medicinal plants. Taken together, this thesis expands the chemical space of naturally occurring bioactive compounds and opens up new avenues in the development and discovery of novel peptidyl therapeutics.

Publications and Presentations

1. **Tan, W.L.**; Wong, K.H.; Lei, J.; Sakai, N.; Tan, H.W.; Hilgenfeld, R.; Tam, J.P. Lybatides from lycium barbarum contain an unusual cystine-stapled helical peptide scaffold. *Sci Rep* **2017**, *7*, 5194.

2. Wong, K.H.; **Tan, W.L.**; Xiao, T.; Tam, J.P. Beta-ginkgotides: Hyperdisulfide-constrained peptides from ginkgo biloba. *Sci Rep* **2017**, *7*, 6140.

3. K. H. Wong, **W. L. Tan**, S. G. Kini, T. Xiao, A. Serra, S. K. Sze, and J. P. Tam, "Vaccatides: Antifungal Glutamine-Rich Hevein-Like Peptides from Vaccaria hispanica," Frontiers in plant science, vol. 8, p. 1100, 2017.

4. S. G. Kini, K. H. Wong, **W. L. Tan**, T. Xiao, and J. P. Tam, "Morintides: cargo-free chitin-binding peptides from Moringa oleifera," BMC plant biology, vol. 17, p. 68, Mar 31 2017.

5. Wong, K. H., **Tan, W. L.**, Serra, A., Xiao, T., Sze, S. K., Yang, D., & Tam, J. P. (2016). Ginkgotides: Proline-Rich Hevein-Like Peptides from Gymnosperm Ginkgo biloba. *Frontiers in plant science*, *7*.

6. Tam, J. P., Wang, S., Wong, K. H., & **Tan, W. L.** (2015). Antimicrobial peptides from plants. *Pharmaceuticals*, *8*(4), 711-757.

7. **W. L. Tan**, K. H. Wong, and J. P. Tam, "Lybatide: Naturally-Occurring Disulfide-Stapled Helical Peptides from Lycium barbarum," *The FASEB Journal,* vol. 31, pp. lb115-lb115, 2017.

8. **Tan, W. L.**, Wong, K. H., Kini, S. G., Maqueda, A. S., Xiao, T., Yang, D., Sze, S. K., Tam, J. P. Elongtide: 10C-hevein-like peptides from Eurycomma longifolia root. (In Preparation)

9. **Tan, W. L.**, Wong K. H., Nguyen, G. K. T., Xiao, T., Tam J. P. Identification and characterization of a carboxypeptidase inhibitor from the medicinal herb wolfberry Lycium barbarum (In Preparation)

10. **Tan, W. L.**, Wong K. H., Nguyen, G. K. T., Xiao, T., Tam J. P. Discovery of a novel carboxypeptidase inhibitor from the medicinal herb wolfberry, Lycium barbarum, Peptide Therapeutics Symposium, San Diego, California, USA, Oct 2015

11. **Tan, W. L.**, Wong K. H., Nguyen, G. K. T., Tam J. P. Discovery of a stable carboxypeptidases inhibitor from the aqueous extract of wolfberry, Lycium barbarum. International Peptide Symposium, Singapore, Dec 2015

References

1. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* **2012**, *75*, 311-335.
2. Cumulative nce introduction index, 1983-2011 (by indication). In *Annual reports in medicinal chemistry*, Manoj, C.D., Ed. Academic Press: 2012; Vol. Volume 47, pp 629-652.
3. Craik, D.J.; Fairlie, D.P.; Liras, S.; Price, D. The future of peptide-based drugs. *Chem Biol Drug Des* **2013**, *81*, 136-147.
4. Fosgerau, K.; Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug discovery today* **2015**, *20*, 122-128.
5. Park, S.T.; Kim, J. Trends in next-generation sequencing and a new era for whole genome sequencing. *International Neurourology Journal* **2016**, *20*, S76-83.
6. Padhi, A.; Sengupta, M.; Sengupta, S.; Roehm, K.H.; Sonawane, A. Antimicrobial peptides and proteins in mycobacterial therapy: Current status and future prospects. *Tuberculosis (Edinb)* **2014**, *94*, 363-373.
7. Buchwald, H.; Dorman, R.B.; Rasmus, N.F.; Michalek, V.N.; Landvik, N.M.; Ikramuddin, S. Effects on glp-1, pyy, and leptin by direct stimulation of terminal ileum and cecum in humans: Implications for ileal transposition. *Surg Obes Relat Dis* **2014**, *10*, 780-786.
8. Giordano, C.; Marchio, M.; Timofeeva, E.; Biagini, G. Neuroactive peptides as putative mediators of antiepileptic ketogenic diets. *Front Neurol* **2014**, *5*, 63.
9. Katz, C.; Levy-Beladev, L.; Rotem-Bamberger, S.; Rito, T.; Rudiger, S.G.D.; Friedler, A. Studying protein-protein interactions using peptide arrays. *Chem Soc Rev* **2011**, *40*, 2131-2145.
10. Peptide therapeutics market: Global industry analysis, size, share, growth, trends and forecast 2012–2018. *Transparency Market Research* **2012**.
11. Kaspar, A.A.; Reichert, J.M. Future directions for peptide therapeutics development. *Drug discovery today* **2013**, *18*, 807-817.
12. Green, T.R.; Ryan, C.A. Wound-induced proteinase inhibitor in plant leaves: A possible defense mechanism against insects. *Science* **1972**, *175*, 776-777.
13. van Loon, L.C.; Rep, M.; Pieterse, C.M. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* **2006**, *44*, 135-162.
14. Sels, J.; Mathys, J.; De Coninck, B.M.; Cammue, B.P.; De Bolle, M.F. Plant pathogenesis-related (pr) proteins: A focus on pr peptides. *Plant Physiol Biochem* **2008**, *46*, 941-950.
15. Akerele, O. Nature's medicinal bounty: Don't throw it away. *World Health Forum* **1993**, *14*, 390-395.
16. Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G. Counting on natural products for drug design. *Nat Chem* **2016**, *8*, 531-541.
17. Kinch, M.S. An overview of fda-approved biologics medicines. *Drug discovery today* **2015**, *20*, 393-398.
18. Baumann, A. Early development of therapeutic biologics--pharmacokinetics. *Curr Drug Metab* **2006**, *7*, 15-21.
19. Marshall, E.; Costa, L.M.; Gutierrez-Marcos, J. Cysteine-rich peptides (crps) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot* **2011**, *62*, 1677-1686.
20. Nguyen, P.Q.; Wang, S.; Kumar, A.; Yap, L.J.; Luu, T.T.; Lescar, J.; Tam, J.P. Discovery and characterization of pseudocyclic cystine-knot alpha-amylase inhibitors with high resistance to heat and proteolytic degradation. *Febs J* **2014**, *281*, 4351-4366.

21. Silverstein, K.A.; Moskal, W.A., Jr.; Wu, H.C.; Underwood, B.A.; Graham, M.A.; Town, C.D.; VandenBosch, K.A. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J* **2007**, *51*, 262-280.
22. Betz, S.F. Disulfide bonds and the stability of globular proteins. *Protein Sci* **1993**, *2*, 1551-1558.
23. Tam, J.P.; Wang, S.; Wong, K.H.; Tan, W.L. Antimicrobial peptides from plants. *Pharmaceuticals (Basel)* **2015**, *8*, 711-757.
24. Kini, S.G.; Wong, K.H.; Tan, W.L.; Xiao, T.; Tam, J.P. Morintides: Cargo-free chitin-binding peptides from moringa oleifera. *BMC Plant Biol* **2017**, *17*, 68.
25. Porto, W.F.; Souza, V.A.; Nolasco, D.O.; Franco, O.L. In silico identification of novel hevein-like peptide precursors. *Peptides* **2012**, *38*, 127-136.
26. Kumari, G.; Serra, A.; Shin, J.; Nguyen, P.Q.; Sze, S.K.; Yoon, H.S.; Tam, J.P. Cysteine-rich peptide family with unusual disulfide connectivity from jasminum sambac. *J Nat Prod* **2015**, *78*, 2791-2799.
27. Shafee, T.M.; Lay, F.T.; Hulett, M.D.; Anderson, M.A. The defensins consist of two independent, convergent protein superfamilies. *Mol Biol Evol* **2016**, *33*, 2345-2356.
28. Nguyen, P.Q.; Ooi, J.S.; Nguyen, N.T.; Wang, S.; Huang, M.; Liu, D.X.; Tam, J.P. Antiviral cystine knot alpha-amylase inhibitors from alstonia scholaris. *J Biol Chem* **2015**, *290*, 31138-31150.
29. Kini, S.G.; Nguyen, P.Q.; Weissbach, S.; Mallagaray, A.; Shin, J.; Yoon, H.S.; Tam, J.P. Studies on the chitin binding property of novel cysteine-rich peptides from alternanthera sessilis. *Biochemistry* **2015**, *54*, 6639-6649.
30. Shafee, T.M.; Lay, F.T.; Phan, T.K.; Anderson, M.A.; Hulett, M.D. Convergent evolution of defensin sequence, structure and function. *Cell Mol Life Sci* **2016**.
31. Lay, F.T.; Schirra, H.J.; Scanlon, M.J.; Anderson, M.A.; Craik, D.J. The three-dimensional solution structure of nad1, a new floral defensin from nicotiana alata and its application to a homology model of the crop defense protein alfafp. *J Mol Biol* **2003**, *325*, 175-188.
32. Nolde, S.B.; Vassilevski, A.A.; Rogozhin, E.A.; Barinov, N.A.; Balashova, T.A.; Samsonova, O.V.; Baranov, Y.V.; Feofanov, A.V.; Egorov, T.A.; Arseniev, A.S*., et al.* Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide ecamp1 from seeds of barnyard grass (echinochloa crus-galli). *J Biol Chem* **2011**, *286*, 25145-25153.
33. Stec, B.; Rao, U.; Teeter, M.M. Refinement of purothionins reveals solute particles important for lattice formation and toxicity .2. Structure of beta-purothionin at 1.7 angstrom resolution. *Acta Crystallogr D* **1995**, *51*, 914-924.
34. Kumari, G.; Serra, A.; Shin, J.; Nguyen, P.Q.T.; Sze, S.K.; Yoon, H.S.; Tam, J.P. Cysteine-rich peptide family with unusual disulfide connectivity from jasminum sambac. *J Nat Prod* **2015**, *78*, 2791-2799.
35. Nguyen, G.K.; Zhang, S.; Nguyen, N.T.; Nguyen, P.Q.; Chiu, M.S.; Hardjojo, A.; Tam, J.P. Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the fabaceae family. *J Biol Chem* **2011**, *286*, 24275-24287.
36. Wong, K.H.; Tan, W.L.; Serra, A.; Xiao, T.; Sze, S.K.; Yang, D.; Tam, J.P. Ginkgotides: Proline-rich hevein-like peptides from gymnosperm ginkgo biloba. *Front Plant Sci* **2016**, *7*, 1639.
37. Thomma, B.P.; Cammue, B.P.; Thevissen, K. Plant defensins. *Planta* **2002**, *216*, 193-202.
38. Broekaert, W.F.; Terras, F.R.; Cammue, B.P.; Osborn, R.W. Plant defensins: Novel antimicrobial peptides as components of the host defense system. *Plant Physiol* **1995**, *108*, 1353-1358.
39. Stec, B. Plant thionins--the structural perspective. *Cell Mol Life Sci* **2006**, *63*, 1370-1385.
40. Van der Weerden, N.L.; Anderson, M.A. Plant defensins: Common fold, multiple functions. *Fungal Biology Reviews* **2013**, *26*, 121-131.

41. 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*, 705-714.
42. Janssen, B.J.; Schirra, H.J.; Lay, F.T.; Anderson, M.A.; Craik, D.J. Structure of petunia hybrida defensin 1, a novel plant defensin with five disulfide bonds. *Biochemistry* **2003**, *42*, 8214-8222.
43. Bontems, F.; Roumestand, C.; Boyot, P.; Gilquin, B.; Doljansky, Y.; Menez, A.; Toma, F. 3-dimensional structure of natural charybdotoxin in aqueous-solution by h-1-nmr - charybdotoxin possesses a structural motif found in other scorpion toxins. *Eur J Biochem* **1991**, *196*, 19-28.
44. Cornet, B.; Bonmatin, J.M.; Hetru, C.; Hoffmann, J.A.; Ptak, M.; Vovelle, F. Refined 3-dimensional solution structure of insect defensin-a. *Structure* **1995**, *3*, 435-448.
45. Raj, P.A.; Dentino, A.R. Current status of defensins and their role in innate and adaptive immunity. *FEMS Microbiol Lett* **2002**, *206*, 9-18.
46. Stotz, H.U.; Thomson, J.G.; Wang, Y. Plant defensins: Defense, development and application. *Plant Signal Behav* **2009**, *4*, 1010-1012.
47. Carvalho Ade, O.; Gomes, V.M. Plant defensins--prospects for the biological functions and biotechnological properties. *Peptides* **2009**, *30*, 1007-1020.
48. Lay, F.T.; Brugliera, F.; Anderson, M.A. Isolation and properties of floral defensins from ornamental tobacco and petunia. *Plant Physiol* **2003**, *131*, 1283-1293.
49. Mendez, E.; Rocher, A.; Calero, M.; Girbes, T.; Citores, L.; Soriano, F. Primary structure of omega-hordothionin, a member of a novel family of thionins from barley endosperm, and its inhibition of protein synthesis in eukaryotic and prokaryotic cell-free systems. *Eur J Biochem* **1996**, *239*, 67-73.
50. Bloch, C., Jr.; Richardson, M. A new family of small (5 kda) protein inhibitors of insect alpha-amylases from seeds or sorghum (sorghum bicolar (l) moench) have sequence homologies with wheat gamma-purothionins. *FEBS Lett* **1991**, *279*, 101-104.
51. Wijaya, R.; Neumann, G.M.; Condron, R.; Hughes, A.B.; Polya, G.M. Defense proteins from seed of cassia fistula include a lipid transfer protein homologue and a protease inhibitory plant defensin. *Plant Sci* **2000**, *159*, 243-255.
52. Melo, F.R.; Rigden, D.J.; Franco, O.L.; Mello, L.V.; Ary, M.B.; Grossi de Sa, M.F.; Bloch, C., Jr. Inhibition of trypsin by cowpea thionin: Characterization, molecular modeling, and docking. *Proteins* **2002**, *48*, 311-319.
53. Gao, A.G.; Hakimi, S.M.; Mittanck, C.A.; Wu, Y.; Woerner, B.M.; Stark, D.M.; Shah, D.M.; Liang, J.; Rommens, C.M. Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nat Biotechnol* **2000**, *18*, 1307-1310.
54. Thevissen, K.; Osborn, R.W.; Acland, D.P.; Broekaert, W.F. Specific binding sites for an antifungal plant defensin from dahlia (dahlia merckii) on fungal cells are required for antifungal activity. *Mol Plant Microbe Interact* **2000**, *13*, 54-61.
55. Thevissen, K.; Warnecke, D.C.; Francois, I.E.; Leipelt, M.; Heinz, E.; Ott, C.; Zahringer, U.; Thomma, B.P.; Ferket, K.K.; Cammue, B.P. Defensins from insects and plants interact with fungal glucosylceramides. *J Biol Chem* **2004**, *279*, 3900-3905.
56. Thevissen, K.; Terras, F.R.; Broekaert, W.F. Permeabilization of fungal membranes by plant defensins inhibits fungal growth. *Appl Environ Microbiol* **1999**, *65*, 5451-5458.
57. van der Weerden, N.L.; Hancock, R.E.; Anderson, M.A. Permeabilization of fungal hyphae by the plant defensin nad1 occurs through a cell wall-dependent process. *J Biol Chem* **2010**, *285*, 37513-37520.
58. Ryan, C.A.; Hass, G.M.; Kuhn, R.W. Purification and properties of a carboxypeptidase inhibitor from potatoes. *J Biol Chem* **1974**, *249*, 5495-5499.

59. McDonald, N.Q.; Hendrickson, W.A. A structural superfamily of growth factors containing a cystine knot motif. *Cell* **1993**, *73*, 421-424.
60. Isaacs, N.W. Cystine knots. *Curr Opin Struct Biol* **1995**, *5*, 391-395.
61. Craik, D.J.; Daly, N.L.; 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*, 1327-1336.
62. Gruber, C.W.; Elliott, A.G.; Ireland, D.C.; Delprete, P.G.; Dessein, S.; Goransson, U.; Trabi, M.; Wang, C.K.; Kinghorn, A.B.; Robbrecht, E*., et al.* Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* **2008**, *20*, 2471-2483.
63. Poth, A.G.; Mylne, J.S.; Grassl, J.; Lyons, R.E.; Millar, A.H.; Colgrave, M.L.; Craik, D.J. Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (solanaceae). *J Biol Chem* **2012**, *287*, 27033-27046.
64. Poth, A.G.; Colgrave, M.L.; Lyons, R.E.; Daly, N.L.; Craik, D.J. Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc Natl Acad Sci U S A* **2011**, *108*, 10127-10132.
65. Nguyen, G.K.; Lian, Y.; Pang, E.W.; Nguyen, P.Q.; Tran, T.D.; Tam, J.P. Discovery of linear cyclotides in monocot plant panicum laxum of poaceae family provides new insights into evolution and distribution of cyclotides in plants. *J Biol Chem* **2013**, *288*, 3370-3380.
66. Ireland, D.C.; Colgrave, M.L.; Nguyencong, P.; Daly, N.L.; Craik, D.J. Discovery and characterization of a linear cyclotide from viola odorata: Implications for the processing of circular proteins. *J Mol Biol* **2006**, *357*, 1522-1535.
67. Chagolla-Lopez, A.; Blanco-Labra, A.; Patthy, A.; Sanchez, R.; Pongor, S. A novel alpha-amylase inhibitor from amaranth (amaranthus hypocondriacus) seeds. *J Biol Chem* **1994**, *269*, 23675-23680.
68. Svensson, B.; Fukuda, K.; Nielsen, P.K.; Bonsager, B.C. Proteinaceous alpha-amylase inhibitors. *Biochim Biophys Acta* **2004**, *1696*, 145-156.
69. Martins, J.C.; Enassar, M.; Willem, R.; Wieruzeski, J.M.; Lippens, G.; Wodak, S.J. Solution structure of the main alpha-amylase inhibitor from amaranth seeds. *Eur J Biochem* **2001**, *268*, 2379-2389.
70. Nguyen, P.Q.; Luu, T.T.; Bai, Y.; Nguyen, G.K.; Pervushin, K.; Tam, J.P. Allotides: Proline-rich cystine knot alpha-amylase inhibitors from allamanda cathartica. *J Nat Prod* **2015**, *78*, 695-704.
71. Pallaghy, P.K.; Nielsen, K.J.; Craik, D.J.; Norton, R.S. A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides. *Protein Sci* **1994**, *3*, 1833-1839.
72. Le-Nguyen, D.; Heitz, A.; Chiche, L.; el Hajji, M.; Castro, B. Characterization and 2d nmr study of the stable [9-21, 15-27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor eeti ii. *Protein Sci* **1993**, *2*, 165-174.
73. Heitz, A.; Le-Nguyen, D.; Chiche, L. Min-21 and min-23, the smallest peptides that fold like a cystine-stabilized beta-sheet motif: Design, solution structure, and thermal stability. *Biochemistry* **1999**, *38*, 10615-10625.
74. Kolmar, H. Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. *Curr Opin Pharmacol* **2009**, *9*, 608-614.
75. Colgrave, M.L.; Craik, D.J. Thermal, chemical, and enzymatic stability of the cyclotide kalata b1: The importance of the cyclic cystine knot. *Biochemistry* **2004**, *43*, 5965-5975.
76. Heitz, A.; Avrutina, O.; Le-Nguyen, D.; Diederichsen, U.; Hernandez, J.F.; Gracy, J.; Kolmar, H.; Chiche, L. Knottin cyclization: Impact on structure and dynamics. *BMC Struct Biol* **2008**, *8*, 54.
77. Fernandez de Caleya, R.; Gonzalez-Pascual, B.; Garcia-Olmedo, F.; Carbonero, P. Susceptibility of phytopathogenic bacteria to wheat purothionins in vitro. *Appl Microbiol* **1972**, *23*, 998-1000.

78. Gao, G.H.; Liu, W.; Dai, J.X.; Wang, J.F.; Hu, Z.; Zhang, Y.; Wang, D.C. Solution structure of pafp-s: A new knottin-type antifungal peptide from the seeds of phytolacca americana. *Biochemistry* **2001**, *40*, 10973-10978.
79. Cammue, B.P.; De Bolle, M.F.; Terras, F.R.; Proost, P.; Van Damme, J.; Rees, S.B.; Vanderleyden, J.; Broekaert, W.F. Isolation and characterization of a novel class of plant antimicrobial peptides form mirabilis jalapa l. Seeds. *J Biol Chem* **1992**, *267*, 2228-2233.
80. Polanowski, A.; Wilusz, T.; Nienartowicz, B.; Cieslar, E.; Slominska, A.; Nowak, K. Isolation and partial amino acid sequence of the trypsin inhibitor from the seeds of cucurbita maxima. *Acta Biochim Pol* **1980**, *27*, 371-382.
81. Hernandez, J.F.; Gagnon, J.; Chiche, L.; Nguyen, T.M.; Andrieu, J.P.; Heitz, A.; Trinh Hong, T.; Pham, T.T.; Le Nguyen, D. Squash trypsin inhibitors from momordica cochinchinensis exhibit an atypical macrocyclic structure. *Biochemistry* **2000**, *39*, 5722-5730.
82. Hass, G.M.; Hermodson, M.A. Amino acid sequence of a carboxypeptidase inhibitor from tomato fruit. *Biochemistry* **1981**, *20*, 2256-2260.
83. Arolas, J.L.; Lorenzo, J.; Rovira, A.; Vendrell, J.; Aviles, F.X.; Ventura, S. Secondary binding site of the potato carboxypeptidase inhibitor. Contribution to its structure, folding, and biological properties. *Biochemistry* **2004**, *43*, 7973-7982.
84. Marino-Buslje, C.; Venhudova, G.; Molina, M.A.; Oliva, B.; Jorba, X.; Canals, F.; Aviles, F.X.; Querol, E. Contribution of c-tail residues of potato carboxypeptidase inhibitor to the binding to carboxypeptidase a a mutagenesis analysis. *Eur J Biochem* **2000**, *267*, 1502-1509.
85. Tam, J.P.; Lu, Y.A.; Yang, J.L.; Chiu, K.W. 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*, 8913-8918.
86. 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*, 10614-10619.
87. Pranting, M.; Loov, C.; Burman, R.; Goransson, U.; Andersson, D.I. The cyclotide cycloviolacin o2 from viola odorata has potent bactericidal activity against gram-negative bacteria. *J Antimicrob Chemoth* **2010**, *65*, 1964-1971.
88. Gran, L. On the effect of a polypeptide isolated from "kalata-kalata" (oldenlandia affinis dc) on the oestrogen dominated uterus. *Acta Pharmacol Toxicol (Copenh)* **1973**, *33*, 400-408.
89. Saether, O.; Craik, D.J.; Campbell, I.D.; Sletten, K.; Juul, J.; Norman, D.G. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata b1. *Biochemistry* **1995**, *34*, 4147-4158.
90. Craik, D.J. Host-defense activities of cyclotides. *Toxins (Basel)* **2012**, *4*, 139-156.
91. Grundemann, C.; Thell, K.; Lengen, K.; Garcia-Kaufer, M.; Huang, Y.H.; Huber, R.; Craik, D.J.; Schabbauer, G.; Gruber, C.W. Cyclotides suppress human t-lymphocyte proliferation by an interleukin 2-dependent mechanism. *PLoS One* **2013**, *8*, e68016.
92. Rosengren, K.J.; Daly, N.L.; Plan, M.R.; Waine, C.; Craik, D.J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J Biol Chem* **2003**, *278*, 8606-8616.
93. Henriques, S.T.; Huang, Y.H.; Chaousis, S.; Sani, M.A.; Poth, A.G.; Separovic, F.; Craik, D.J. The prototypic cyclotide kalata b1 has a unique mechanism of entering cells. *Chem Biol* **2015**, *22*, 1087-1097.
94. Henriques, S.T.; Craik, D.J. Cyclotides as templates in drug design. *Drug discovery today* **2010**, *15*, 57-64.
95. Shenkarev, Z.O.; Nadezhdin, K.D.; Sobol, V.A.; Sobol, A.G.; Skjeldal, L.; Arseniev, A.S. Conformation and mode of membrane interaction in cyclotides. Spatial structure of kalata b1 bound to a dodecylphosphocholine micelle. *Febs J* **2006**, *273*, 2658-2672.

96. Wang, C.K.; Colgrave, M.L.; Ireland, D.C.; Kaas, Q.; Craik, D.J. Despite a conserved cystine knot motif, different cyclotides have different membrane binding modes. *Biophys J* **2009**, *97*, 1471-1481.
97. Henriques, S.T.; Huang, Y.H.; Rosengren, K.J.; Franquelim, H.G.; Carvalho, F.A.; Johnson, A.; Sonza, S.; Tachedjian, G.; Castanho, M.A.R.B.; Daly, N.L.*, et al.* Decoding the membrane activity of the cyclotide kalata b1 the importance of phosphatidylethanolamine phospholipids and lipid organization on hemolytic and anti-hiv activities. *Journal of Biological Chemistry* **2011**, *286*, 24231-24241.
98. Beintema, J.J. Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett* **1994**, *350*, 159-163.
99. Jimenez-Barbero, J.; Javier Canada, F.; Asensio, J.L.; Aboitiz, N.; Vidal, P.; Canales, A.; Groves, P.; Gabius, H.J.; Siebert, H.C. Hevein domains: An attractive model to study carbohydrate-protein interactions at atomic resolution. *Adv Carbohydr Chem Biochem* **2006**, *60*, 303-354.
100. Archer, B.L. The proteins of hevea brasiliensis latex. 4. Isolation and characterization of crystalline hevein. *Biochem J* **1960**, *75*, 236-240.
101. Van Parijs, J.; Broekaert, W.F.; Goldstein, I.J.; Peumans, W.J. Hevein: An antifungal protein from rubber-tree (hevea brasiliensis) latex. *Planta* **1991**, *183*, 258-264.
102. Boller, T.; Metraux, J. Extracellular localization of chitinase in cucumber. *Physiological and molecular plant pathology* **1988**, *33*, 11-16.
103. Rice, R.H.; Etzler, M.E. Subunit structure of wheat germ agglutinin. *Biochem Biophys Res Commun* **1974**, *59*, 414-419.
104. Broekaert, I.; Lee, H.I.; Kush, A.; Chua, N.H.; Raikhel, N. Wound-induced accumulation of mrna containing a hevein sequence in laticifers of rubber tree (hevea brasiliensis). *Proc Natl Acad Sci U S A* **1990**, *87*, 7633-7637.
105. Shibuya, N.; Goldstein, I.J.; Shafer, J.A.; Peumans, W.J.; Broekaert, W.F. Carbohydrate binding properties of the stinging nettle (urtica dioica) rhizome lectin. *Arch Biochem Biophys* **1986**, *249*, 215-224.
106. Carrasco, L.; Vazquez, D.; Hernandez-Lucas, C.; Carbonero, P.; Garcia-Olmedo, F. Thionins: Plant peptides that modify membrane permeability in cultured mammalian cells. *Eur J Biochem* **1981**, *116*, 185-189.
107. Nielsen, K.K.; Nielsen, J.E.; Madrid, S.M.; Mikkelsen, J.D. Characterization of a new antifungal chitin-binding peptide from sugar beet leaves. *Plant Physiol* **1997**, *113*, 83-91.
108. Broekaert, W.F.; Marien, W.; Terras, F.R.; De Bolle, M.F.; Proost, P.; Van Damme, J.; Dillen, L.; Claeys, M.; Rees, S.B.; Vanderleyden, J.*, et al.* Antimicrobial peptides from amaranthus caudatus seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. *Biochemistry* **1992**, *31*, 4308-4314.
109. Lipkin, A.; Anisimova, V.; Nikonorova, A.; Babakov, A.; Krause, E.; Bienert, M.; Grishin, E.; Egorov, T. An antimicrobial peptide ar-amp from amaranth (amaranthus retroflexus l.) seeds. *Phytochemistry* **2005**, *66*, 2426-2431.
110. Koo, J.C.; Lee, S.Y.; Chun, H.J.; Cheong, Y.H.; Choi, J.S.; Kawabata, S.; Miyagi, M.; Tsunasawa, S.; Ha, K.S.; Bae, D.W.*, et al.* Two hevein homologs isolated from the seed of pharbitis nil l. Exhibit potent antifungal activity. *Biochim Biophys Acta* **1998**, *1382*, 80-90.
111. Fujimura, M.; Minami, Y.; Watanabe, K.; Tadera, K. Purification, characterization, and sequencing of a novel type of antimicrobial peptides, fa-amp1 and fa-amp2, from seeds of buckwheat (fagopyrum esculentum moench.). *Biosci Biotechnol Biochem* **2003**, *67*, 1636-1642.
112. Yang, Y.F.; Cheng, K.C.; Tsai, P.H.; Liu, C.C.; Lee, T.R.; Lyu, P.C. Alanine substitutions of noncysteine residues in the cysteine-stabilized alphabeta motif. *Protein Sci* **2009**, *18*, 1498-1506.

113. Balaji, R.A.; Ohtake, A.; Sato, K.; Gopalakrishnakone, P.; Kini, R.M.; Seow, K.T.; Bay, B.H. Lambda-conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. Isolation and characterization from the venom of conus marmoreus. *J Biol Chem* **2000**, *275*, 39516-39522.
114. Asensio, J.L.; Canada, F.J.; Siebert, H.C.; Laynez, J.; Poveda, A.; Nieto, P.M.; Soedjanaamadja, U.M.; Gabius, H.J.; Jimenez-Barbero, J. Structural basis for chitin recognition by defense proteins: Glcnac residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem Biol* **2000**, *7*, 529-543.
115. Wagner, S.; Breiteneder, H. The latex-fruit syndrome. *Biochem Soc Trans* **2002**, *30*, 935-940.
116. Rodriguez-Romero, A.; Ravichandran, K.G.; Soriano-Garcia, M. Crystal structure of hevein at 2.8 a resolution. *FEBS Lett* **1991**, *291*, 307-309.
117. Andersen, N.H.; Cao, B.; Rodriguez-Romero, A.; Arreguin, B. Hevein: Nmr assignment and assessment of solution-state folding for the agglutinin-toxin motif. *Biochemistry* **1993**, *32*, 1407-1422.
118. Asensio, J.L.; Canada, F.J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J. The interaction of hevein with n-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose. *Eur J Biochem* **1995**, *230*, 621-633.
119. Aboitiz, N.; Vila-Perello, M.; Groves, P.; Asensio, J.L.; Andreu, D.; Canada, F.J.; Jimenez-Barbero, J. Nmr and modeling studies of protein-carbohydrate interactions: Synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *Chembiochem* **2004**, *5*, 1245-1255.
120. Asensio, J.L.; Canada, F.J.; Bruix, M.; Gonzalez, C.; Khiar, N.; Rodriguez-Romero, A.; Jimenez-Barbero, J. Nmr investigations of protein-carbohydrate interactions: Refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside. *Glycobiology* **1998**, *8*, 569-577.
121. Yokoyama, S.; Kato, K.; Koba, A.; Minami, Y.; Watanabe, K.; Yagi, F. Purification, characterization, and sequencing of antimicrobial peptides, cy-amp1, cy-amp2, and cy-amp3, from the cycad (cycas revoluta) seeds. *Peptides* **2008**, *29*, 2110-2117.
122. Huang, R.H.; Xiang, Y.; Tu, G.Z.; Zhang, Y.; Wang, D.C. Solution structure of eucommia antifungal peptide: A novel structural model distinct with a five-disulfide motif. *Biochemistry* **2004**, *43*, 6005-6012.
123. Dubovskii, P.V.; Vassilevski, A.A.; Slavokhotova, A.A.; Odintsova, T.I.; Grishin, E.V.; Egorov, T.A.; Arseniev, A.S. Solution structure of a defense peptide from wheat with a 10-cysteine motif. *Biochem Biophys Res Commun* **2011**, *411*, 14-18.
124. Odintsova, T.I.; Vassilevski, A.A.; Slavokhotova, A.A.; Musolyamov, A.K.; Finkina, E.I.; Khadeeva, N.V.; Rogozhin, E.A.; Korostyleva, T.V.; Pukhalsky, V.A.; Grishin, E.V*., et al.* A novel antifungal hevein-type peptide from triticum kiharae seeds with a unique 10-cysteine motif. *Febs J* **2009**, *276*, 4266-4275.
125. Van den Bergh, K.P.; Proost, P.; Van Damme, J.; Coosemans, J.; Van Damme, E.J.; Peumans, W.J. Five disulfide bridges stabilize a hevein-type antimicrobial peptide from the bark of spindle tree (euonymus europaeus l.). *FEBS Lett* **2002**, *530*, 181-185.
126. Wessels, J.G.H. A steady-state model for apical wall growth in fungi. *Acta Botanica Neerlandica* **1988**, *37*, 3-16.
127. Slavokhotova, A.A.; Naumann, T.A.; Price, N.P.; Rogozhin, E.A.; Andreev, Y.A.; Vassilevski, A.A.; Odintsova, T.I. Novel mode of action of plant defense peptides - hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases. *Febs J* **2014**, *281*, 4754-4764.
128. Malanovic, N.; Lohner, K. Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. *Biochim Biophys Acta* **2016**, *1858*, 936-946.

129. Epple, P.; Apel, K.; Bohlmann, H. An arabidopsis thaliana thionin gene is inducible via a signal transduction pathway different from that for pathogenesis-related proteins. *Plant Physiol* **1995**, *109*, 813-820.
130. Romero, A.; Alamillo, J.M.; Garcia-Olmedo, F. Processing of thionin precursors in barley leaves by a vacuolar proteinase. *Eur J Biochem* **1997**, *243*, 202-208.
131. Stevens, C.; Titarenko, E.; Hargreaves, J.A.; Gurr, S.J. Defence-related gene activation during an incompatible interaction between stagonospora (septoria) nodorum and barley (hordeum vulgare l.) coleoptile cells. *Plant Mol Biol* **1996**, *31*, 741-749.
132. Ponz, F.; Paz-Ares, J.; Hernandez-Lucas, C.; Carbonero, P.; Garcia-Olmedo, F. Synthesis and processing of thionin precursors in developing endosperm from barley (hordeum vulgare l.). *Embo J* **1983**, *2*, 1035-1040.
133. Stec, B.; Rao, U.; Teeter, M.M. Refinement of purothionins reveals solute particles important for lattice formation and toxicity. Part 2: Structure of beta-purothionin at 1.7 a resolution. *Acta Crystallogr D Biol Crystallogr* **1995**, *51*, 914-924.
134. Milbradt, A.G.; Kerek, F.; Moroder, L.; Renner, C. Structural characterization of hellethionins from helleborus purpurascens. *Biochemistry* **2003**, *42*, 2404-2411.
135. Stec, B.; Markman, O.; Rao, U.; Heffron, G.; Henderson, S.; Vernon, L.P.; Brumfeld, V.; Teeter, M.M. Proposal for molecular mechanism of thionins deduced from physico-chemical studies of plant toxins. *J Pept Res* **2004**, *64*, 210-224.
136. Ebrahim-Nesbat, F.; Behnke, S.; Kleinhofs, A.; Apel, K. Cultivar-related differences in the distribution of cell-wall-bound thionins in compatible and incompatible interactions between barley and powdery mildew. *Planta* **1989**, *179*, 203-210.
137. Evans, J.; Wang, Y.D.; Shaw, K.P.; Vernon, L.P. Cellular responses to pyrularia thionin are mediated by ca2+ influx and phospholipase a2 activation and are inhibited by thionin tyrosine iodination. *Proc Natl Acad Sci U S A* **1989**, *86*, 5849-5853.
138. Coulon, A.; Mosbah, A.; Lopez, A.; Sautereau, A.M.; Schaller, G.; Urech, K.; Rouge, P.; Darbon, H. Comparative membrane interaction study of viscotoxins a3, a2 and b from mistletoe (viscum album) and connections with their structures. *Biochemical Journal* **2003**, *374*, 71-78.
139. Utkina, L.L.; Andreev, Y.A.; Rogozhin, E.A.; Korostyleva, T.V.; Slavokhotova, A.A.; Oparin, P.B.; Vassilevski, A.A.; Grishin, E.V.; Egorov, T.A.; Odintsova, T.I. Genes encoding 4-cys antimicrobial peptides in wheat triticum kiharae dorof. Et migush.: Multimodular structural organization, instraspecific variability, distribution and role in defence. *Febs J* **2013**, *280*, 3594-3608.
140. Slavokhotova, A.A.; Rogozhin, E.A.; Musolyamov, A.K.; Andreev, Y.A.; Oparin, P.B.; Berkut, A.A.; Vassilevski, A.A.; Egorov, T.A.; Grishin, E.V.; Odintsova, T.I. Novel antifungal alpha-hairpinin peptide from stellaria media seeds: Structure, biosynthesis, gene structure and evolution. *Plant Mol Biol* **2014**, *84*, 189-202.
141. Duvick, J.P.; Rood, T.; Rao, A.G.; Marshak, D.R. Purification and characterization of a novel antimicrobial peptide from maize (zea mays l.) kernels. *J Biol Chem* **1992**, *267*, 18814-18820.
142. Marcus, J.P.; Green, J.L.; Goulter, K.C.; Manners, J.M. A family of antimicrobial peptides is produced by processing of a 7s globulin protein in macadamia integrifolia kernels. *Plant J* **1999**, *19*, 699-710.
143. Tuteja, R. Type i signal peptidase: An overview. *Arch Biochem Biophys* **2005**, *441*, 107-111.
144. von Heijne, G. Signal sequences. The limits of variation. *J Mol Biol* **1985**, *184*, 99-105.
145. Gruber, C.W.; Cemazar, M.; Clark, R.J.; Horibe, T.; Renda, R.F.; Anderson, M.A.; Craik, D.J. A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins. *J Biol Chem* **2007**, *282*, 20435-20446.
146. Hebert, D.N.; Molinari, M. In and out of the er: Protein folding, quality control, degradation, and related human diseases. *Physiol Rev* **2007**, *87*, 1377-1408.

147. Stevens, F.J.; Argon, Y. Protein folding in the er. *Semin Cell Dev Biol* **1999**, *10*, 443-454.
148. Nguyen, G.K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J.P. Butelase 1 is an asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* **2014**, *10*, 732-738.
149. Nguyen, G.K.; Qiu, Y.; Cao, Y.; Hemu, X.; Liu, C.F.; Tam, J.P. Butelase-mediated cyclization and ligation of peptides and proteins. *Nat Protoc* **2016**, *11*, 1977-1988.
150. Schafmeister, C.E.; Po, J.; Verdine, G.L. An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *J Am Chem Soc* **2000**, *122*, 5891-5892.
151. Ward, P.; Ewan, G.B.; Jordan, C.C.; Ireland, S.J.; Hagan, R.M.; Brown, J.R. Potent and highly selective neurokinin antagonists. *J Med Chem* **1990**, *33*, 1848-1851.
152. Wong, C.T.; Rowlands, D.K.; Wong, C.H.; Lo, T.W.; Nguyen, G.K.; Li, H.Y.; Tam, J.P. Orally active peptidic bradykinin b1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew Chem Int Ed Engl* **2012**, *51*, 5620-5624.
153. Qiu, Y.; Taichi, M.; Wei, N.; Yang, H.; Luo, K.Q.; Tam, J.P. An orally active bradykinin b1 receptor antagonist engineered as a bifunctional chimera of sunflower trypsin inhibitor. *J Med Chem* **2017**.
154. Vita, C.; Roumestand, C.; Toma, F.; Menez, A. Scorpion toxins as natural scaffolds for protein engineering. *Proceedings of the National Academy of Sciences* **1995**, *92*, 6404-6408.
155. Li, C.; Liu, M.; Monbo, J.; Zou, G.; Li, C.; Yuan, W.; Zella, D.; Lu, W.-Y.; Lu, W. Turning a scorpion toxin into an antitumor miniprotein. *J Am Chem Soc* **2008**, *130*, 13546-13548.
156. Krause, S.; Schmoldt, H.U.; Wentzel, A.; Ballmaier, M.; Friedrich, K.; Kolmar, H. Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists. *Febs Journal* **2007**, *274*, 86-95.
157. Potterat, O. Goji (lycium barbarum and l-chinense): Phytochemistry, pharmacology and safety in the perspective of traditional uses and recent popularity. *Planta Med* **2010**, *76*, 7-19.
158. Yin, G.H.; Dang, Y.L. Optimization of extraction technology of the lycium barbarum polysaccharides by box-behnken statistical design. *Carbohyd Polym* **2008**, *74*, 603-610.
159. Piao, M.; Murata, Y.; Zhu, B.; Shimoishi, Y.; Tada, M. Changes in carotenoid content and its composition during maturation of fructus lycii fruits. *Jpn J Food Chem* **2005**, *12*, 35-39 (CAN 144: 169735).
160. Toyoda-Ono, Y.; Maeda, M.; Nakao, M.; Yoshimura, M.; Sugiura-Tomimori, N.; Fukami, H. 2-o-(β-d-glucopyranosyl)ascorbic acid, a novel ascorbic acid analogue isolated from lycium fruit. *J Agric Food Chem* **2004**, *52*, 2092-2096.
161. Qi, Z.; Li, S.; Wu, J.; Qu, R.; Yang, Y.; Zhang, L.; Yang, X. Chemical constituents of fructus lycii and folium lycii – nutrients in fructus lycii and folium lycii. *Zhongyao Tongbao (Beijing, China)* **1986**, *11*, 169-171.
162. Le, K.; Chiu, F.; Ng, K. Identification and quantification of antioxidants in fructus lycii. *Food Chem* **2007**, *105*, 353-363.
163. Altintas, A.; Kosar, M.; Kirimer, N.; Baser, K.H.; Demirci, B. Composition of the essential oils of lycium barbarum and lycium ruthenicum fruits. *Chem Nat Comp* **2006**, *41*, 24-25.
164. Funayama, S.; Yoshida, K.; Konno, C.; Hikino, H. Structure of kukoamine a, a hypotensive principle of lycium chinense root barks. *Tetrahedron Lett* **1980**, *21*, 1355-1356.
165. Chu, Q.; Fu, L.; Lin, M.; Ye, J. Study on bioactive ingredients in cortex lycii by capillary zone electrophoresis with amperometric detector. *Fenxi Huaxue* **2005**, *33*, 1611-1614.
166. Noguchi, M.; Mochida, K.; Shingu, T.; Fujitani, K.; Kozuka, M. Sugiol and 5α-stigmastane-3,6-dione from the chinese drug “ti-ku-p'i” (lycii radicis cortex). *J Nat Prod* **1985**, *48*, 342-343.
167. Yahara, S.; Shigeyama, C.; Ura, T.; Wakamatsu, K.; Yasuhara, T.; Nohara, T. Cyclic peptides, acyclic diterpene glycosides and other compounds from lycium chinense mill. *Chem Pharm Bull* **1993**, *41*, 703-709.

168. Chang, R.C.C.; So, K.F. Use of anti-aging herbal medicine, lycium barbarum, against aging-associated diseases. What do we know so far? *Cell Mol Neurobiol* **2008**, *28*, 643-652.
169. Li, X.M.; Li, X.L.; Zhou, A.G. Evaluation of antioxidant activity of the polysaccharides extracted from lycium barbarum fruits in vitro. *Eur Polymer J* **2007**, *43*, 488-497.
170. Li, X.L.; Zhou, A.G. Evaluation of the antioxidant effects of polysaccharides extracted from lycium barbarum. *Med Chem Res* **2007**, *15*, 471-482.
171. Ma, L.; Chen, Q.; Yang, W.; Xi, S.; Wan, X.; Tang, X.; Yu, Y.; Kang, J. Effect of lycium barbarum polysaccharide against atherosclerosis in rabbits. *Zhengzhou Daxue Xuebao, Yixueban* **2005**, *40*, 328-330 (CAN 144: 324440).
172. Zhao, R.; Li, Q.; Xiao, B. Effect of lycium barbarum polysaccharide on the improvement of insulin resistance in niddm rats. *Yakugaku Zasshi* **2005**, *125*, 981-988.
173. Chen, Z.; Lu, J.; Srinivasan, N.; Tan, B.K.H.; Chan, S.H. Polysaccharide-protein complex from lycium barbarum l. Is a novel stimulus of dendritic cell immunogenicity. *J Immunol* **2009**, *182*, 3503-3506.
174. Gan, L.; Zhang, S.H.; Liu, Q.; Xu, H.B. A polysaccharide-protein complex from lycium barbarum upregulates cytokine expression in human peripheral blood mononuclear cells. *Eur J Pharmacol* **2003**, *471*, 217-222.
175. Peng, X.; Huang, J.; Qi, C.; Zhang, Y.X.; Tian, G.Y. Studies on chemistry and immuno-modulating mechanism of a glycoconjugate from lycium barbarum l. *Chin J Chem* **2001**, *19*, 1190-1197.
176. Zhang, M.; Chen, H.; Huang, J.; Li, Z.; Zhu, C.; Zhang, S. Effect of lycium barbarum polysaccharide on human hepatoma qgy7703 cells: Inhibition of proliferation and induction of apoptosis. *Life Sci* **2005**, *76*, 2115-2124.
177. Liu, X.; Sun, J.; Li, H.; Zhang, L.; Qian, B. Extraction and isolation of active component in fruit of lycium barbarum for inhibiting pc3 cell proliferation in vitro. *Zhongguo Zhongyao Zazhi* **2000**, *25*, 481-483 (CAN 134: 263473).
178. Li, G.; Sepkovic, W.; Bradlow, H.L.; Telang, N.T.; Wong, G.Y.C. Lycium barbarum inhibits growth of estrogen receptor positive human breast cancer cells by favorably altering estradiol metabolism. *Nutr Cancer* **2009**, *61*, 408-414.
179. Yahara, S.; Shigeyama, C.; Nohara, T. Structures of anti-ace and -renin peptides from lycii radicis cortex. *Tetrahedron Lett* **1989**, *30*, 6041-6042.
180. Lin, C.C.; Chuang, S.C.; Lin, J.M.; Yang, J.J. Evaluation of the anti-inflammatory hepatoprotective and antioxidant activities of lycium chinense from taiwan. *Phytomedicine* **1997**, *4*, 213-220.
181. Group, H.M.R. Compendium of medicinal plants used in malaysia. *Kuala Lumpur: Institute for Medical Research, Malaysia* **2002**, *345*.
182. Keng, H. *Orders and families of malayan seed plants*. NUS Press: 1987.
183. Jiwajinda, S.; Santisopasri, V.; Murakami, A.; Hirai, N.; Ohigashi, H. Quassinoids from eurycoma longifolia as plant growth inhibitors. *Phytochemistry* **2001**, *58*, 959-962.
184. Rehman, S.U.; Choe, K.; Yoo, H.H. Review on a traditional herbal medicine, eurycoma longifolia jack (tongkat ali): Its traditional uses, chemistry, evidence-based pharmacology and toxicology. *Molecules* **2016**, *21*, 331.
185. Bhat, R.; Karim, A.A. Tongkat ali (eurycoma longifolia jack): A review on its ethnobotany and pharmacological importance. *Fitoterapia* **2010**, *81*, 669-679.
186. Fiaschetti, G.; A Grotzer, M.; Shalaby, T.; Castelletti, D.; Arcaro, A. Quassinoids: From traditional drugs to new cancer therapeutics. *Current medicinal chemistry* **2011**, *18*, 316-328.
187. Chan, K.; Lee, S.; Sam, T.; Han, B. A quassinoid glycoside from the roots of eurycoma longifolia. *Phytochemistry* **1989**, *28*, 2857-2859.
188. Darise, M.; Kohda, H.; Mizutani, K.; Tanaka, O. Eurycomanone and eurycomanol, quassinoids from the roots of eurycoma longifolia. *Phytochemistry* **1982**, *21*, 2091-2093.

189. Low, B.-S.; Choi, S.-B.; Wahab, H.A.; Das, P.K.; Chan, K.-L. Eurycomanone, the major quassinoid in eurycoma longifolia root extract increases spermatogenesis by inhibiting the activity of phosphodiesterase and aromatase in steroidogenesis. *J Ethnopharmacol* **2013**, *149*, 201-207.
190. Ang, H.; Ngai, T.; Tan, T. Effects of eurycoma longifolia jack on sexual qualities in middle aged male rats. *Phytomedicine* **2003**, *10*, 590-593.
191. Ismail, S.B.; Wan Mohammad, W.M.Z.; George, A.; Nik Hussain, N.H.; Musthapa Kamal, Z.M.; Liske, E. Randomized clinical trial on the use of physta freeze-dried water extract of eurycoma longifolia for the improvement of quality of life and sexual well-being in men. *Evidence-Based Complementary and Alternative Medicine* **2012**, *2012*.
192. Tambi, M.I.B.M.; Imran, M.K. Eurycoma longifolia jack in managing idiopathic male infertility. *Asian journal of andrology* **2010**, *12*, 376.
193. Hai Dang, N.; Choo, Y.Y.; Tien Dat, N.; Hoai Nam, N.; Van Minh, C.; Lee, J.H. 7‐methoxy‐(9h‐β‐carbolin‐1‐il)‐(e)‐1‐propenoic acid, a β‐carboline alkaloid from eurycoma longifolia, exhibits anti‐inflammatory effects by activating the nrf2/heme oxygenase‐1 pathway. *Journal of cellular biochemistry* **2016**, *117*, 659-670.
194. Nielsen, H.; Engelbrecht, J.; Brunak, S.; vonHeijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **1997**, *10*, 1-6.
195. Bendtsen, J.D.; Nielsen, H.; von Heijne, G.; Brunak, S. Improved prediction of signal peptides: Signalp 3.0. *J Mol Biol* **2004**, *340*, 783-795.
196. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome Res* **2004**, *14*, 1188-1190.
197. Pace, C.N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to measure and predict the molar absorption-coefficient of a protein. *Protein Science* **1995**, *4*, 2411-2423.
198. Jeener, J.; Meier, B.H.; Bachmann, P.; Ernst, R.R. Investigation of exchange processes by 2-dimensional nmr-spectroscopy. *J Chem Phys* **1979**, *71*, 4546-4553.
199. Kumar, A.; Ernst, R.R.; Wuthrich, K. A two-dimensional nuclear overhauser enhancement (2d noe) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* **1980**, *95*, 1-6.
200. Davis, D.G.; Bax, A. Assignment of complex h-1-nmr spectra via two-dimensional homonuclear hartmann-hahn spectroscopy. *J Am Chem Soc* **1985**, *107*, 2820-2821.
201. Bax, A. Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance (1969)* **1985**, *65*, 355-360.
202. Rance, M.; Sorensen, O.W.; Bodenhausen, G.; Wagner, G.; Ernst, R.R.; Wuthrich, K. Improved spectral resolution in cosy 1h nmr spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun* **1983**, *117*, 479-485.
203. Wuthrich, K.; Billeter, M.; Braun, W. Pseudo-structures for the 20 common amino-acids for use in studies of protein conformations by measurements of intramolecular proton proton distance constraints with nuclear magnetic-resonance. *J Mol Biol* **1983**, *169*, 949-961.
204. Liu, M.; Mao, X.-A.; Ye, C.; Huang, H.; Nicholson, J.; Lindon, J. Improved watergate pulse sequences for solvent suppression in nmr spectroscopy. *Journal of Magnetic Resonance* **1998**, *132*, 125-129.
205. Hwang, T.L.; Shaka, A.J. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A* **1995**, *112*, 275-279.
206. Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe: A multidimensional spectral processing system based on unix pipes. *Journal of Biomolecular Nmr* **1995**, *6*, 277-293.
207. Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe: A multidimensional spectral processing system based on unix pipes. *Journal of biomolecular NMR* **1995**, *6*, 277-293.

208. Guntert, P.; Mumenthaler, C.; Wuthrich, K. Torsion angle dynamics for nmr structure calculation with the new program dyana. *J Mol Biol* **1997**, *273*, 283-298.
209. Weiss, M.S.; Sicker, T.; Hilgenfeld, R. Soft x-rays, high redundancy, and proper scaling: A new procedure for automated protein structure determination via sas. *Structure* **2001**, *9*, 771-777.
210. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.
211. Sheldrick, G.M. Experimental phasing with shelxc/d/e: Combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 479-485.
212. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501.
213. Murshudov, G.N.; Skubak, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. Refmac5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 355-367.
214. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. Ucsf chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25*, 1605-1612.
215. Jenssen H, A.S. Serum stability of peptides. *Methods Mol. Biol.* **2008**, *494*, 177-186.
216. Ye, X.Y.; Ng, T.B. A new antifungal peptide from rice beans. *J Pept Res* **2002**, *60*, 81-87.
217. Wiegand, I.; Hilpert, K.; Hancock, R.E. Agar and broth dilution methods to determine the minimal inhibitory concentration (mic) of antimicrobial substances. *Nat Protoc* **2008**, *3*, 163-175.
218. Lehrer, R.I.; Rosenman, M.; Harwig, S.S.S.L.; Jackson, R.; Eisenhauer, P. Ultrasensitive assays for endogenous antimicrobial polypeptides. *J Immunol Methods* **1991**, *137*, 167-173.
219. Marsh, A.J.; O'Sullivan, O.; Ross, R.P.; Cotter, P.D.; Hill, C. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics* **2010**, *11*, 679.
220. Boguski, M.S.; Lowe, T.M.; Tolstoshev, C.M. Dbest--database for "expressed sequence tags". *Nat Genet* **1993**, *4*, 332-333.
221. Matasci, N.; Hung, L.H.; Yan, Z.; Carpenter, E.J.; Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Ayyampalayam, S.; Barker, M.*, et al.* Data access for the 1,000 plants (1kp) project. *Gigascience* **2014**, *3*, 17.
222. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. Signalp 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* **2011**, *8*, 785-786.
223. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis tool web services from the embl-ebi. *Nucleic Acids Res* **2013**, *41*, W597-600.
224. Letunic, I.; Bork, P. Interactive tree of life (itol) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **2016**.
225. Acharya, K.; Pal, A.K.; Gulati, A.; Kumar, S.; Singh, A.K.; Ahuja, P.S. Overexpression of camellia sinensis thaumatin-like protein, cstlp in potato confers enhanced resistance to macrophomina phaseolina and phytophthora infestans infection. *Mol Biotechnol* **2013**, *54*, 609-622.
226. Zhang, J.; Hu, M.; Li, J.T.; Guan, J.P.; Yang, B.; Shu, W.S.; Liao, B. A transcriptional profile of metallophyte viola baoshanensis involved in general and species-specific cadmium-defense mechanisms. *J Plant Physiol* **2009**, *166*, 862-870.
227. Nguyen, G.K.; Zhang, S.; Wang, W.; Wong, C.T.; Nguyen, N.T.; Tam, J.P. Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry. *J Biol Chem* **2011**, *286*, 44833-44844.
228. Cheek, S.; Krishna, S.S.; Grishin, N.V. Structural classification of small, disulfide-rich protein domains. *J Mol Biol* **2006**, *359*, 215-237.

229. Pereira, P.J.; Lozanov, V.; Patthy, A.; Huber, R.; Bode, W.; Pongor, S.; Strobl, S. Specific inhibition of insect alpha-amylases: Yellow meal worm alpha-amylase in complex with the amaranth alpha-amylase inhibitor at 2.0 a resolution. *Structure* **1999**, *7*, 1079-1088.
230. Taylor, W.G.; Fields, P.G.; Elder, J.L. Insecticidal components from field pea extracts: Isolation and separation of peptide mixtures related to pea albumin 1b. *J Agric Food Chem* **2004**, *52*, 7491-7498.
231. Abdallah, N.A.; Shah, D.; Abbas, D.; Madkour, M. Stable integration and expression of a plant defensin in tomato confers resistance to fusarium wilt. *GM Crops* **2010**, *1*, 344-350.
232. Ng, T.B.; Cheung, R.C.; Wong, J.H.; Ye, X. Lipid-transfer proteins. *Biopolymers* **2012**, *98*, 268-279.
233. Koehbach, J.; Gruber, C.W. From ethnopharmacology to drug design. *Commun Integr Biol* **2013**, *6*, e27583.
234. Vendrell, J.; Querol, E.; Avilés, F.X. Metallocarboxypeptidases and their protein inhibitors: Structure, function and biomedical properties. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2000**, *1477*, 284-298.
235. Bajzar, L.; Manuel, R.; Nesheim, M.E. Purification and characterization of tafi, a thrombin-activable fibrinolysis inhibitor. *J Biol Chem* **1995**, *270*, 14477-14484.
236. Wei, S.; Segura, S.; Vendrell, J.; Aviles, F.X.; Lanoue, E.; Day, R.; Feng, Y.; Fricker, L.D. Identification and characterization of three members of the human metallocarboxypeptidase gene family. *J Biol Chem* **2002**, *277*, 14954-14964.
237. Zisman, L.S.; Keller, R.S.; Weaver, B.; Lin, Q.; Speth, R.; Bristow, M.R.; Canver, C.C. Increased angiotensin-(1-7)-forming activity in failing human heart ventricles: Evidence for upregulation of the angiotensin-converting enzyme homologue ace2. *Circulation* **2003**, *108*, 1707-1712.
238. Kayashima, T.; Yamasaki, K.; Yamada, T.; Sakai, H.; Miwa, N.; Ohta, T.; Yoshiura, K.; Matsumoto, N.; Nakane, Y.; Kanetake, H.*, et al.* The novel imprinted carboxypeptidase a4 gene (cpa4) in the 7q32 imprinting domain. *Hum Genet* **2003**, *112*, 220-226.
239. Arolas, J.L.; Vendrell, J.; Aviles, F.X.; Fricker, L.D. Metallocarboxypeptidases: Emerging drug targets in biomedicine. *Curr Pharm Des* **2007**, *13*, 349-366.
240. Hass, G.M.; Nau, H.; Biemann, K.; Grahn, D.T.; Ericsson, L.H.; Neurath, H. The amino acid sequence of a carboxypeptidase inhibitor from potatoes. *Biochemistry* **1975**, *14*, 1334-1342.
241. Reverter, D.; Vendrell, J.; Canals, F.; Horstmann, J.; Aviles, F.X.; Fritz, H.; Sommerhoff, C.P. A carboxypeptidase inhibitor from the medical leech hirudo medicinalis. Isolation, sequence analysis, cdna cloning, recombinant expression, and characterization. *J Biol Chem* **1998**, *273*, 32927-32933.
242. Sanglas, L.; Aviles, F.X.; Huber, R.; Gomis-Ruth, F.X.; Arolas, J.L. Mammalian metallopeptidase inhibition at the defense barrier of ascaris parasite. *Proc Natl Acad Sci U S A* **2009**, *106*, 1743-1747.
243. Covaleda, G.; del Rivero, M.A.; Chavez, M.A.; Aviles, F.X.; Reverter, D. Crystal structure of novel metallocarboxypeptidase inhibitor from marine mollusk nerita versicolor in complex with human carboxypeptidase a4. *Journal of Biological Chemistry* **2012**, *287*, 9250-9258.
244. Rees, D.C.; Lipscomb, W.N. Refined crystal structure of the potato inhibitor complex of carboxypeptidase a at 2.5 a resolution. *J Mol Biol* **1982**, *160*, 475-498.
245. Bouma, B.N.; Meijers, J.C. Thrombin-activatable fibrinolysis inhibitor (tafi, plasma procarboxypeptidase b, procarboxypeptidase r, procarboxypeptidase u). *J Thromb Haemost* **2003**, *1*, 1566-1574.
246. Walker, J.B.; Hughes, B.; James, I.; Haddock, P.; Kluft, C.; Bajzar, L. Stabilization versus inhibition of tafia by competitive inhibitors in vitro. *J Biol Chem* **2003**, *278*, 8913-8921.
247. Schneider, M.; Nesheim, M. Reversible inhibitors of tafia can both promote and inhibit fibrinolysis. *Journal of Thrombosis and Haemostasis* **2003**, *1*, 147-154.

248. Nagashima, M.; Werner, M.; Wang, M.; Zhao, L.; Light, D.R.; Pagila, R.; Morser, J.; Verhallen, P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* **2000**, *98*, 333-342.
249. Pereira, H.J.V.; Souza, L.L.; Salgado, M.C.O.; Oliveira, E.B. Angiotensin processing is partially carried out by carboxypeptidases in the rat mesenteric arterial bed perfusate. *Regul Peptides* **2008**, *151*, 135-138.
250. Xin, Y.F.; Wan, L.L.; Peng, J.L.; Guo, C. Alleviation of the acute doxorubicin-induced cardiotoxicity by lycium barbarum polysaccharides through the suppression of oxidative stress. *Food Chem Toxicol* **2011**, *49*, 259-264.
251. GA, S.; FP, S. *Chinese materia medica.* . American Presbyterian Mission Press: Shanghai, 1911.
252. C, B.; S, C.; E, S. *Chinese herbal medicine*. 3rd Edition ed.; Eastland Press, Inc: Seattle, 2004.
253. JK, C.; TT, C. *Chinese medical herbology and pharmacology*. Art of Medicine Press, Inc: City of Industry, CA, 2004.
254. Chen, S.; Wang, Q.; Gong, S.; Wu, J.; Yu, X.; Lin, S. Analysis of amino acid in fructus lycii. *Zhongguo Yaoke Daxue Xuebao* **1991**, *22*, 53-55 (CAN 115: 15369).
255. Cao, Y.; Zhang, X.; Chu, Q.; Fang, Y.; Ye, J. Determination of taurine in lycium barbarum l. And other foods by capillary electrophoresis with electrochemical detection. *Electroanalysis* **2003**, *15*, 898-902.
256. Ren, B.; Ma, Y.; Sheng, Y.; Gao, B. Protective action of lycium barbarum l. And betaine on lipid peroxidation of rbc membrane induced by hydrogen peroxide. *Zhongguo Zhongyao Zazhi* **1995**, *20*, 303-304 (CAN 124: 21718).
257. Molina, M.A.; Marino, C.; Oliva, B.; Aviles, F.X.; Querol, E. C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase-a. *Journal of Biological Chemistry* **1994**, *269*, 21467-21472.
258. Hass, G.M.; Ako, H.; Grahn, D.T.; Neurath, H. Carboxypeptidase inhibitor from potatoes - effects of chemical modifications on inhibitory activity. *Biochemistry* **1976**, *15*, 93-100.
259. Armentro.Rw; Doolittl.Rf. Pyrrolidonecarboxylyl peptidase - stabilization and purification. *Arch Biochem Biophys* **1969**, *132*, 80-&.
260. Nguyen, G.K.; Cao, Y.; Wang, W.; Liu, C.F.; Tam, J.P. Site-specific n-terminal labeling of peptides and proteins using butelase 1 and thiodepsipeptide. *Angew Chem Int Ed Engl* **2015**, *54*, 15694-15698.
261. Lin, S.L.; Nussinov, R. A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. *Nat Struct Biol* **1995**, *2*, 835-837.
262. Arolas, J.L.; Lorenzo, J.; Rovira, A.; Castella, J.; Aviles, F.X.; Sommerhoff, C.P. A carboxypeptidase inhibitor from the tick rhipicephalus bursa: Isolation, cdna cloning, recombinant expression, and characterization. *J Biol Chem* **2005**, *280*, 3441-3448.
263. Chelius, D.; Jing, K.; Lueras, A.; Rehder, D.S.; Dillon, T.M.; Vizel, A.; Rajan, R.S.; Li, T.; Treuheit, M.J.; Bondarenko, P.V. Formation of pyroglutamic acid from n-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal Chem* **2006**, *78*, 2370-2376.
264. Abraham, G.N.; Podell, D.N. Pyroglutamic acid. In *The biological effects of glutamic acid and its derivatives*, Springer: 1981; pp 181-190.
265. Rink, R.; Arkema-Meter, A.; Baudoin, I.; Post, E.; Kuipers, A.; Nelemans, S.A.; Akanbi, M.H.; Moll, G.N. To protect peptide pharmaceuticals against peptidases. *J Pharmacol Toxicol Methods* **2010**, *61*, 210-218.
266. Viljakainen, L.; Pamilo, P. Selection on an antimicrobial peptide defensin in ants. *J Mol Evol* **2008**, *67*, 643-652.

267. Bayes, A.; Comellas-Bigler, M.; de la Vega, M.R.; Maskos, K.; Bode, W.; Aviles, F.X.; Jongsma, M.A.; Beekwilder, J.; Vendrell, J. Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proc Natl Acad Sci U S A* **2005**, *102*, 16602-16607.
268. Ryan, C.A. Protease inhibitors in plants - genes for improving defenses against insects and pathogens. *Annu Rev Phytopathol* **1990**, *28*, 425-449.
269. Ireland, D.C.; Colgrave, M.L.; Craik, D.J. A novel suite of cyclotides from viola odorata: Sequence variation and the implications for structure, function and stability. *Biochem J* **2006**, *400*, 1-12.
270. Mylne, J.S.; Chan, L.Y.; Chanson, A.H.; Daly, N.L.; Schaefer, H.; Bailey, T.L.; Nguyencong, P.; Cascales, L.; Craik, D.J. Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* **2012**, *24*, 2765-2778.
271. Ling, M.H.; Qi, H.Y.; Chi, C.W. Protein, cdna, and genomic DNA sequences of the towel gourd trypsin inhibitor. A squash family inhibitor. *J Biol Chem* **1993**, *268*, 810-814.
272. Villanueva, J.; Canals, F.; Prat, S.; Ludevid, D.; Querol, E.; Aviles, F.X. Characterization of the wound-induced metallocarboxypeptidase inhibitor from potato. Cdna sequence, induction of gene expression, subcellular immunolocalization and potential roles of the c-terminal propeptide. *FEBS Lett* **1998**, *440*, 175-182.
273. Homandberg, G.A.; Litwiller, R.D.; Peanasky, R.J. Carboxypeptidase inhibitors from ascaris suum: The primary structure. *Arch Biochem Biophys* **1989**, *270*, 153-161.
274. Muto, Y.; Suzuki, K.; Sato, E.; Ishii, H. Carboxypeptidase b inhibitors reduce tissue factor-induced renal microthrombi in rats. *European journal of pharmacology* **2003**, *461*, 181-189.
275. Morser, J. Thrombomodulin links coagulation to inflammation and immunity. *Curr Drug Targets* **2012**, *13*, 421-431.
276. Sitja-Arnau, M.; Molina, M.A.; Blanco-Aparicio, C.; Ferrer-Soler, L.; Lorenzo, J.; Aviles, F.X.; Querol, E.; de Llorens, R. Mechanism of action of potato carboxypeptidase inhibitor (pci) as an egf blocker. *Cancer Lett* **2005**, *226*, 169-184.
277. Bayes, A.; de la Vega, M.R.; Vendrell, J.; Aviles, F.X.; Jongsma, M.A.; Beekwilder, J. Response of the digestive system of helicoverpa zea to ingestion of potato carboxypeptidase inhibitor and characterization of an uninhibited carboxypeptidase b. *Insect Biochem Molec* **2006**, *36*, 654-664.
278. Francesc, A.; Francesc, C.; Hans, F.; Jeanny, H.; Enrique, Q.; David, R.; P, S.C.; Josep, V. Inhibidor de metalocarboxipeptidasas como agente fibrinolitico. Google Patents: 2000.
279. AVILÉS PUIGVERT, F.X.; LORENZO RIVERA, J.; RODRÍGUEZ-VERA, M.; QUEROL MURILLO, E.; BAUTISTA MARUGÁN, M.; DÍEZ MARTIN, A.; BAUTISTA SANTA CRUZ, J.M. Therapeutic agents for treatment of malaria. WO/2008/077977

20.12.2007, 2008.

280. Minnema, M.C.; Friederich, P.W.; Levi, M.; von dem Borne, P.A.K.; Mosnier, L.O.; Meijers, J.C.M.; Biemond, B.J.; Hack, C.E.; Bouma, B.N.; ten Cate, H. Enhancement of rabbit jugular vein thrombolysis by neutralization of factor xi - in vitro evidence for a role of factor xi as an anti-fibrinolytic factor (vol 101, pg 10, 1998). *J Clin Invest* **1998**, *101*, 917-917.
281. Hohenstein, B.; Braun, A.; Amann, K.U.; Johnson, R.J.; Hugo, C.P.M. A murine model of site-specific renal microvascular endothelial injury and thrombotic microangiopathy. *Nephrol Dial Transpl* **2008**, *23*, 1144-1156.
282. Ryan, C.A. Proteinase inhibitor gene families: Strategies for transformation to improve plant defenses against herbivores. *Bioessays* **1989**, *10*, 20-24.
283. Tam, J.P.; Lu, Y.A. Synthesis of large cyclic cystine-knot peptide by orthogonal coupling strategy using unprotected peptide precursor. *Tetrahedron Lett* **1997**, *38*, 5599-5602.
284. Taichi, M.; Hemu, X.; Qiu, Y.B.; Tam, J.P. A thioethylalkylamido (tea) thioester surrogate in the synthesis of a cyclic peptide via a tandem acyl shift. *Org Lett* **2013**, *15*, 2620-2623.

285. Morita, H.; Yoshida, N.; Takeya, K.; Itokawa, H.; Shirota, O. Configurational and conformational analyses of a cyclic octapeptide, lyciumin a, from lycium chinense mill. *Tetrahedron* **1996**, *52*, 2795-2802.
286. Koehn, F.E.; Carter, G.T. The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **2005**, *4*, 206-220.
287. Bhardwaj, G.; Mulligan, V.K.; Bahl, C.D.; Gilmore, J.M.; Harvey, P.J.; Cheneval, O.; Buchko, G.W.; Pulavarti, S.V.; Kaas, Q.; Eletsky, A.*, et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **2016**, *538*, 329-335.
288. Richard, J.A.; Kelly, I.; Marion, D.; Auger, M.; Pezolet, M. Structure of beta-purothionin in membranes: A two-dimensional infrared correlation spectroscopy study. *Biochemistry* **2005**, *44*, 52-61.
289. Hilgenfeld, R.; Saenger, W. Structural chemistry of natural and synthetic ionophores and their complexes with cations. *Top Curr Chem* **1982**, *101*, 1-82.
290. Bernal, F.; Tyler, A.F.; Korsmeyer, S.J.; Walensky, L.D.; Verdine, G.L. Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide. *Journal of the American Chemical Society* **2007**, *129*, 2456-2457.
291. Teeter, M.M.; Mazer, J.A.; L'Italien, J.J. Primary structure of the hydrophobic plant protein crambin. *Biochemistry* **1981**, *20*, 5437-5443.
292. Nguyen, K.N.; Nguyen, G.K.; Nguyen, P.Q.; Ang, K.H.; Dedon, P.C.; Tam, J.P. Immunostimulating and gram-negative-specific antibacterial cyclotides from the butterfly pea (clitoria ternatea). *Febs J* **2016**, *283*, 2067-2090.
293. Bernaldez, J.; Roman-Gonzalez, S.A.; Martinez, O.; Jimenez, S.; Vivas, O.; Arenas, I.; Corzo, G.; Arreguin, R.; Garcia, D.E.; Possani, L.D.*, et al.* A conus regularis conotoxin with a novel eight-cysteine framework inhibits ca(v)2.2 channels and displays an anti-nociceptive activity. *Mar Drugs* **2013**, *11*, 1188-1202.
294. Pease, J.H.B.; Storrs, R.W.; Wemmer, D.E. Folding and activity of hybrid sequence, disulfide-stabilized peptides. *Proc Natl Acad Sci U S A* **1990**, *87*, 5643-5647.
295. Bruix, M.; Jimenez, M.A.; Santoro, J.; Gonzalez, C.; Colilla, F.J.; Mendez, E.; Rico, M. Solution structure of gamma 1-h and gamma 1-p thionins from barley and wheat endosperm determined by 1h-nmr: A structural motif common to toxic arthropod proteins. *Biochemistry* **1993**, *32*, 715-724.
296. Douse, C.H.; Maas, S.J.; Thomas, J.C.; Garnett, J.A.; Sun, Y.; Cota, E.; Tate, E.W. Crystal structures of stapled and hydrogen bond surrogate peptides targeting a fully buried protein-helix interaction. *ACS Chem Biol* **2014**, *9*, 2204-2209.
297. Zhang, Y.; Skolnick, J. Tm-align: A protein structure alignment algorithm based on the tm-score. *Nucleic Acids Res* **2005**, *33*, 2302-2309.
298. Cooley, R.B.; Arp, D.J.; Karplus, P.A. Evolutionary origin of a secondary structure: Pi-helices as cryptic but widespread insertional variations of alpha-helices that enhance protein functionality. *J Mol Biol* **2010**, *404*, 232-246.
299. Weaver, T.M. The pi-helix translates structure into function. *Protein Sci* **2000**, *9*, 201-206.
300. Walensky, L.D.; Bird, G.H. Hydrocarbon-stapled peptides: Principles, practice, and progress. *J Med Chem* **2014**, *57*, 6275-6288.
301. Jackson, D.Y.; King, D.S.; Chmielewski, J.; Singh, S.; Schultz, P.G. General-approach to the synthesis of short alpha-helical peptides. *Journal of the American Chemical Society* **1991**, *113*, 9391-9392.
302. Chen, S.T.; Chen, H.J.; Yu, H.M.; Wang, K.T. Facile synthesis of a short peptide with a side-chain-constrained structure. *J Chem Res-S* **1993**, 228-229.

303. Evans Schultes, R. Medicinal plants of east and southeast asia: Attributed properties. *Econ Bot* **1980**, *34*, 361-361.
304. Ebrahimi, F.; Ibrahim, B.; Teh, C.H.; Murugaiyah, V.; Lam, C.K. 1hnmr-based discriminatory analysis of eurycoma longifolia from different locations and establishing a profile for primary metabolites identification and quassinoids quantification. *Planta Med* **2017**, *83*, 172-182.
305. Tee, T.T.; Cheah, Y.H.; Hawariah, L.P.A. F16, a fraction from eurycoma longifolia jack extract, induces apoptosis via a caspase-9-independent manner in mcf-7 cells. *Anticancer Res* **2007**, *27*, 3425-3430.
306. Farouk, A.E. Antibacterial activity of eurycoma longifolia jack - a malaysian medicinal plant. *Saudi Med J* **2007**, *28*, 1422-1424.
307. Kuo, P.C.; Damu, A.G.; Lee, K.H.; Wu, T.S. Cytotoxic and antimalarial constituents from the roots of eurycoma longifolia. *Bioorgan Med Chem* **2004**, *12*, 537-544.
308. Ang, H.H.; Lee, K.L.; Kiyoshi, M. Sexual arousal in sexually sluggish old male rats after oral administration of eurycoma longifolia jack. *J Basic Clin Physiol Pharmacol* **2004**, *15*, 303-309.
309. Kuo, P.C.; Shi, L.S.; Damu, A.G.; Su, C.R.; Huang, C.H.; Ke, C.H.; Wu, J.B.; Lin, A.J.; Bastow, K.F.; Lee, K.H.*, et al.* Cytotoxic and antimalarial beta-carboline alkaloids from the roots of eurycoma longifolia. *J Nat Prod* **2003**, *66*, 1324-1327.
310. Nguyen, G.K.; Lim, W.H.; Nguyen, P.Q.; Tam, J.P. Novel cyclotides and uncyclotides with highly shortened precursors from chassalia chartacea and effects of methionine oxidation on bioactivities. *J Biol Chem* **2012**, *287*, 17598-17607.
311. Rinaudo, M. Chitin and chitosan: Properties and applications. *Prog Polym Sci* **2006**, *31*, 603-632.
312. Broekaert, W.F.; Cammue, B.P.A.; De Bolle, M.F.C.; Thevissen, K.; De Samblanx, G.W.; Osborn, R.W.; Nielson, K. Antimicrobial peptides from plants. *Critical Reviews in Plant Sciences* **1997**, *16*, 297-323.
313. Van den Bergh, K.P.; Van Damme, E.J.; Peumans, W.J.; Coosemans, J. Ee-cbp, a hevein-type antimicrobial peptide from bark of the spindle tree (euonymus europaeus l.). *Meded Rijksuniv Gent Fak Landbouwkd Toegep Biol Wet* **2002**, *67*, 327-331.
314. Huang, R.H.; Xiang, Y.; Liu, X.Z.; Zhang, Y.; Hu, Z.; Wang, D.C. Two novel antifungal peptides distinct with a five-disulfide motif from the bark of eucommia ulmoides oliv. *FEBS Lett* **2002**, *521*, 87-90.
315. Trabi, M.; Craik, D.J. Tissue-specific expression of head-to-tail cyclized miniproteins in violaceae and structure determination of the root cyclotide viola hederacea root cyclotide1. *Plant Cell* **2004**, *16*, 2204-2216.
316. Xiang, Y.; Huang, R.H.; Liu, X.Z.; Zhang, Y.; Wang, D.C. Crystal structure of a novel antifungal protein distinct with five disulfide bridges from eucommia ulmoides oliver at an atomic resolution. *J Struct Biol* **2004**, *148*, 86-97.
317. Kezuka, Y.; Kojima, M.; Mizuno, R.; Suzuki, K.; Watanabe, T.; Nonaka, T. Structure of full-length class i chitinase from rice revealed by x-ray crystallography and small-angle x-ray scattering. *Proteins* **2010**, *78*, 2295-2305.
318. Collinge, D.B.; Kragh, K.M.; Mikkelsen, J.D.; Nielsen, K.K.; Rasmussen, U.; Vad, K. Plant chitinases. *Plant Journal* **1993**, *3*, 31-40.
319. Jablonowski, D.; Schaffrath, R. Zymocin, a composite chitinase and trnase killer toxin from yeast. *Biochem Soc T* **2007**, *35*, 1533-1537.
320. Chavez, M.I.; Vila-Perello, M.; Canada, F.J.; Andreu, D.; Jimenez-Barbero, J. Effect of a serine-to-aspartate replacement on the recognition of chitin oligosaccharides by truncated hevein. A 3d view by using nmr. *Carbohyd Res* **2010**, *345*, 1461-1468.
321. Jimenez-Barbero, J.; Asensio, J.L.; Canada, F.J.; Poveda, A. Free and protein-bound carbohydrate structures. *Curr Opin Struc Biol* **1999**, *9*, 549-555.

322. Chavez, M.I.; Andreu, C.; Vidal, P.; Aboitiz, N.; Freire, F.; Groves, P.; Asensio, J.L.; Asensio, G.; Muraki, M.; Canada, F.J*., et al.* On the importance of carbohydrate-aromatic interactions for the molecular recognition of oligosaccharides by proteins: Nmr studies of the structure and binding affinity of acamp2-like peptides with non-natural naphthyl and fluoroaromatic residues. *Chem-Eur J* **2005**, *11*, 7060-7074.
323. Andreev, Y.A.; Korostyleva, T.V.; Slavokhotova, A.A.; Rogozhin, E.A.; Utkina, L.L.; Vassilevski, A.A.; Grishin, E.V.; Egorov, T.A.; Odintsova, T.I. Genes encoding hevein-like defense peptides in wheat: Distribution, evolution, and role in stress response. *Biochimie* **2012**, *94*, 1009-1016.
324. Gongora-Benitez, M.; Tulla-Puche, J.; Albericio, F. Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem Rev* **2014**, *114*, 901-926.
325. Tan, W.L.; Wong, K.H.; Lei, J.; Sakai, N.; Tan, H.W.; Hilgenfeld, R.; Tam, J.P. Lybatides from lycium barbarum contain an unusual cystine-stapled helical peptide scaffold. *Sci Rep* **2017**, *7*, 5194.
326. Lay, F.T.; Anderson, M.A. Defensins--components of the innate immune system in plants. *Curr Protein Pept Sci* **2005**, *6*, 85-101.
327. Daly, N.L.; Craik, D.J. Bioactive cystine knot proteins. *Curr Opin Chem Biol* **2011**, *15*, 362-368.
328. Gough, J. Convergent evolution of domain architectures (is rare). *Bioinformatics* **2005**, *21*, 1464-1471.
329. Kritzer, J.A. Stapled peptides: Magic bullets in nature's arsenal. *Nat Chem Biol* **2010**, *6*, 566-567.
330. Verdine, G.L.; Hilinski, G.J. Stapled peptides for intracellular drug targets. *Methods Enzymol* **2012**, *503*, 3-33.
331. Doolittle, R.F. Convergent evolution: The need to be explicit. *Trends in Biochemical Sciences* **1994**, *19*, 15-18.
332. Stern, D.L. The genetic causes of convergent evolution. *Nat Rev Genet* **2013**, *14*, 751-764.
333. Moore, S.J.; Leung, C.L.; Cochran, J.R. Knottins: Disulfide-bonded therapeutic and diagnostic peptides. *Drug Discovery Today: Technologies* **2012**, *9*, e3-e11.