NANYANG TECHNOLOGICAL UNIVERSITY

HEVEIN-LIKE PEPTIDES IN PLANTS: DISCOVERY, CHARACTERIZATION AND MOLECULAR DIVERSITY

KINI SHRUTHI GOPALKRISHNA

SCHOOL OF BIOLOGICAL SCIENCES

2016

Hevein-like Peptides in Plants: Discovery, Characterization and Molecular Diversity

KINI SHRUTHI GOPALKRISHNA

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

2016

Dedicated to Amma and Annu

Acknowledgements

The 4-year-long journey that culminates in this thesis has been a truly riveting one! Its completion is thanks to all the people that motivated, supported and challenged me along the way.

First and foremost, I express my heartfelt gratitude to my supervisor, Prof. James P. Tam for giving me the opportunity to pursue my research in his laboratory. His constant guidance and faith in me kept me motivated even through the tough times. He taught me that “there is no substitute for hard work” and that nothing is impossible without dedication and perseverance. I would also like to thank my co-supervisor, Assoc. Prof. Kathy Luo Qian and my thesis advisory committee for their invaluable guidance and advice throughout my PhD. I am deeply grateful to Prof. Dr. Thomas Peters from the University of Lübeck for giving me the opportunity to work in his laboratory and learn about NMR-titrations. I would also like to thank Assoc. Prof. Peter Dröge for his positive critique and Prof. P.V Subba Rao for being a source of inspiration and motivation throughout this journey.

This thesis would not have been possible without the motivation and support from my lab members. In particular, I would like to thank Dr. Nguyen Quoc Thuc Phuong for patiently teaching me the basics of extraction, isolation and cloning. I thank Dr. Wong Ka Ho whole-heartedly for guiding me with anti-fungal assays, EST-based data mining and for being a great friend. I am blessed to have found wonderful friends in my lab-mates, Tan Wei Liang, Dr. Nguyen Kien Truc Giang, Clarene Chan, Yunjiao and Geeta Kumari. Our daily lunch-breaks, dinners and outings will be cherished forever. I also wish to thank Dr. Alvaro Mallagaray and Dr. Sophie Weissbach from the University of Lübeck for making me feel at home in a new environment and helping me with the NMR-titration experiments. Special thanks to Areetha D’Souza and Dr. Shin Joon for helping me with NMR

experiments at NTU. I am thankful to all the URECA and FYP students who worked with me and helped me learn along with them.

A strong support system is essential to survive the rigorous journey towards obtaining a PhD. I am lucky to have met some wonderful people that I can call friends-for-life during this journey. My time in Singapore has truly been a pleasant one thanks to Akshita and Lakshya. With friends like them constantly by my side I knew I could face my toughest days with ease. This journey would have been a dull one without Annanya, my partner-in-crime and closest friend. I deeply appreciate Garima and Dilraj for their guidance, both professional and personal. My PhD would not have been possible without the encouragement from all my friends back home in Bangalore. I specially thank Paru and Prithvi for sticking by me despite the distance and constantly motivating me.

My fiancé, Akhil, has been my pillar of support, my confidant and my biggest motivator throughout this journey. I thank him for his unwavering faith in me and for putting a smile on my face even on my worst days. I appreciate his patience in dealing with me during the most stressful phase of my PhD. Whether it was treating me to ice-cream in the middle of the night or staying awake with me during an overnight experiment, he has been there for me all the way and I will be forever grateful to him for his unconditional love for me.

There are no words to express how thankful I am to my parents for trusting me with the decision to pursue my studies abroad and for allowing me to follow my dreams. They have celebrated every tiny success and consoled me through the failures during this journey for which I am truly grateful. It is their prayers and blessings that have helped me come this far. I hope this thesis makes them proud and I have lived up to their expectations.

Table of Contents

List of Figures

CHAPTER 1

CHAPTER 3

CHAPTER 4

CHAPTER 5

List of Tables

CHAPTER 1

CHAPTER 3

CHAPTER 4

Abbreviations

A. alternata	*Alternaria alternata*
A. brassiciola	*Alternaria brassiciola*
A. niger	*Aspergillus niger*
A. sessilis	*Alternanthera sessilis*
ACN	Acetonitrile
AEP	asparaginyl endopeptidase
AMP	antimicrobial peptide
BLAST	basic local alignment search tool
C. lunata	*Curvularia lunata*
cDNA	complementary deoxyribonucleic acid
COSY	correlation spectroscopy
CRP	cysteine-rich peptide
C-tail	C-terminal tail
D2O	deuterium oxide
DCM	Dichloromethane
DTT	dithiothreitol
DTT	Dithiothreitol
EDTA	ethylenediaminetetracetic acid
ER	endoplasmic reticulum (signal peptide)
ER	endoplasmic reticulum
EST	expressed sequence tags
EtOH	Ethanol
F. oxysporum	*Fusarium oxysporum*
FA	Formic acid
GlcNAc	N-acetylglucosamine
HPLC	high-performance liquid chromatography
IAA	Iodoacetamide
IC_{50}	half maximal inhibitory concentration

K_D	dissociation constant
kDa	kilo Dalton
M. oleifera	*Moringa oleifera*
MALDI-TOF MS	matrix-assisted laser desorption/ionization time of flight mass spectrometry
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NH4HCO3	Ammonium bicarbonate
NMR	Nuclear Magnetic Resonance
NOESY	nuclear overhauser effect spectroscopy
ORF	open reading frame
PCR	polymerase chain reaction
R. solani	*Rhizoctonia solani*
RACE	rapid amplification of cDNA ends
RP	reversed-phase
SCX	strong cation exchange
tblastn	translated nucleotide BLAST
TFA	Trifluoroacetic acid
UPLC	Ultra-performance liquid chromatography
V. dahilae	*Verticulum dahilae*

Abstract

Hevein-like peptides belong to a family of chitin-binding, cystine-knotted peptides. Based on the number of cysteine residues in hevein-like peptides they can be divided into three sub-classes, namely, 6C-, 8C- and 10C- hevein-like peptides. The objectives of my thesis are to isolate and characterize novel hevein-like peptides and to study their distribution and evolution in plants. From two herbs, *Alternanthera sessilis* and *Moringa oleifera*, six novel 6C-hevein-like peptides and three novel 8C-hevein-like peptides were isolated and purified. In addition, we gained insights into their chitin-binding interaction, anti-fungal activity and biosynthesis pathway.

Using an *in silico* approach, a total of 385 novel hevein-like peptides were found in 124 plant families including bryophytes and gymnosperms. In conclusion, my thesis expands the existing knowledge of hevein-like peptides regarding their structure; function and biosynthesis which may help to develop robust anti-fungal agents and transgenic plants with increased specificity and resistance to fungal infections.

Chapter 1

Introduction

1. Cysteine-rich peptides in plant defense

Plants are highly susceptible to microbial invasions and pathogenic attacks. The first line of defense to combat these attacks constitutes the thick cell wall [1]. Pathogens crossing this barrier are tackled by the cellular innate immune system comprised of defense-related chemical compounds such as phytoalexins [2], pathogenesis-related proteins [3], reactive oxygen species and enzyme inhibitors [4]. Inherently expressed plant defense compounds are also known to be beneficial for humans. Since prehistoric times, plants have been used for treatment of a number of ailments and diseases [5]. Long-standing proofs of this principle are traditional folk medicine practices like traditional Chinese medicine (TCM) and Ayurveda. Today, drugs of herbal origin constitute 25% of prescribed drugs worldwide [6]. In fact, out of approximately 120 herbal drugs used today, 96 have a usage identical to their ethnomedicinal use [5]. However, their mechanism of action is still an unsolved mystery.

Efforts have been made to isolate and characterize active principles in medicinal plants which are responsible for their therapeutic effects. Most of the isolated compounds are plant secondary metabolites like tannins, flavonoids and alkaloids which have a molecular weight less than 500 Dalton (Da). Some such popular herbal-small-molecule drugs include aspirin, isolated from the bark of *Salix alba* (willow tree) and *Spiraea salicifolia* and used as an anti-pyretic and analgesic [7]; morphine, an alkaloid from *Papaver somniferum* (opium) widely

used as an analgesic [8] and quinine, isolated from *Cinchona officinalis* trees, used as an anti-malarial drug [9]. Pharmaceutical industries prefer small-molecule drugs as their production is cost-effective and they have high oral bioavailability. However, a major drawback of these drugs is that their small size results in smaller footprint and hence reduced specificity, increased off-target effects and toxicity [10]. Therefore, recent focus has shifted from small molecules to macromolecular therapeutics like peptide biologics [11]. Peptide biologics are defined as proteinaceous compounds with therapeutic potential and a molecular weight above 500 Da [10]. Their large molecular weight confers them a large footprint resulting in high specificity and reduced off-target effects. However, the use of plant biologics as drugs is still debated due to the common notion that proteins and peptides are easily denatured on heating, are readily digested by proteolytic enzymes in the digestive tract and are poorly absorbed through the gut resulting in low bioavailability [12].

Advances in understanding of cysteine-rich peptides (CRPs) in recent years have helped to disprove the above misconceptions. CRPs are present throughout the plant kingdom and can be defined as peptides with a molecular weight between 2 kDa to 6 kDa containing a high number of cysteine residues [13]. CRPs contain an even number of cysteine residues that form intramolecular disulfide bonds in different configurations (Table 1.1) [14]. These intramolecular disulfide bonds render CRPs highly stable against thermal and enzymatic degradation [15, 16].

Table 1.1. Classification of plant cysteine-rich peptides.

Peptide family	Peptide name	Plant species	No. of residues	Cysteine pattern
Plant defensin	Rs-AFP2	*Raphanus sativus*	45 - 57	[1-9] - C - [8-11] - C - [3-5] - C - [3] - C - [9-11] - C - [4-8] - C - [1] - C - [3] -
Thionin	α- purothionin	*Hordeum vulgare*	42 - 47	[2] - CC - [7] - C - [3] - C - [8-10] - C - [3] - C - [1] - C [7] - C - [3-6]
Knottin-type peptides	Mj-AMP1	*Mirabilis jalapa*	36 - 38	[0-2] - C - [6] - C - [8] - CC- [3] - C - [10] - C - [3]
Cyclotide	Kalata B1	*Odenlandia affinis*	28 - 37	[3-5] - C - [3] - C - [4-5] - C- [4-7] - C - [1] - C - [4-7] - C - [2-4]
Impatiens	Ib-AMP1	*Impatiens balsamina*	20	[5] - CC- [8] - C - [3] - C
Hevein-like peptides	hevein	*Hevea brasiliensis*	29 - 45	[2-6] - C - [1-9] - C - [4-5] - CC- [5] - C - [3-4] - C - [1-5] - C - [2-6] - C - [2-

The consensus sequence and cysteine connectivity of each family are summarized from the PhytAMP database [17]. Numbers within brackets indicate the number of non-cysteine amino acid residues in the backbone. The disulfide connectivity is shown by solid lines while the cyclic backbone of cyclotides is indicated by a dotted line.

CRPs are known to have diverse biological roles and are an integral part of the plant defense system. They are expressed in peripheral tissues like leaves and flowers to combat pathogenic attacks. Some CRPs play non-defensive roles like regulation of reproduction, growth and development while others are toxic to insect-larvae [18], inhibit α-amylase [19] or are toxic to mammalian cells [20]. Some examples of CRPs include plant defensins [21], thionins [22] knottin-type peptides [23], cyclotides [24], and impatiens [25]. As a consequence of their diverse pharmacological profiles along with their enhanced resistance to thermal and enzymatic degradation, we hypothesize that CRPs are the active principles responsible for the therapeutic properties of medicinal herbs.

A class of CRPs called hevein-like peptides, comprising 29-45 amino acid residues are rich in glycine and cysteine residues [26]. Hevein-like peptides are known for their ability to bind chitin, a polymer of repeating N-acetylglucosamine units (GlcNAc), linked by β-(1-4)-glycosidic bonds and the primary constituent of fungal cell walls. As a consequence of their chitin-binding property, hevein-like peptides are potent anti-fungal agents against a number of chitin-containing fungi. These hevein-like peptides are the subject of this thesis.

2. Hevein-like peptides

2.1. Discovery of hevein-like peptides

The rubber tree (*Hevea brasiliensis*) is the source of the world's natural rubber with a global production of 12 million tons in 2014 (Fig. 1.1). Rubber is obtained from the latex of this plant which constitutes rubber particles, proteins, lipids and non-rubber particles such as alkaloids, tannins and starch. In 1953, Cook and Sekhar showed that latex was fractionated into three layers on centrifugation. A creamy rubber top layer, a C-serum intermediate layer and a bottom layer comprised of lutoids [27, 28]. Lutoids are vacuolar organelles that make up 12% of the latex and comprise proteins that play an integral role in stress-response, chitin degradation and plant defense [29, 30].

In the early 1950s Archer *et al*. conducted a study to elucidate the protein components of the bottom layer of rubber latex and reported that it comprised of several electophoretically distinct protein components [20, 27, 28]. In 1960, Archer and his group isolated a 10 kDa anionic non-glycosylated protein from the 'bottom layer' of fractionated latex and reported that it was rich in cysteine, aspartic acid and serine [31]. They named this protein hevein which stands for a protein "ein" isolated from *Hevea brasiliensis* "hev". Later in 1975, Walujono *et al.* sequenced hevein and presented its primary sequence in the International Rubber Conference held in Malaysia [32].

Figure 1.1. Natural rubber tapping from *Hevea brasiliensis* latex
The rubber tree is the major source of the world's natural rubber with a global production of 12 million tons in 2014. Picture adapted from https://goo.gl/images/hmYqIN.

\

Walujono and group showed that the hevein “protein” isolated by Archer was a dimer of a glycine and cysteine-rich peptide which was 43 amino acid residues in length [32]. The sequence of hevein was homologous to the cysteine/glycine-rich domain of chitin-binding proteins like basic chitinases from cucumber [33], tobacco [34], chitin-binding lectins from wheat [35] and stinging nettle lectin [36]. In 1990, Broekaert *et al.* isolated the cDNA clone (HEV1) of hevein from the latex of *Hevea brasiliensis [37]*. The hevein precursor was encoded by an open reading frame of 1080 bp and was 204 amino acid residues long. The hevein precursor comprised a 17 amino-acid-long signal peptide followed by a 43 amino acid long mature domain encoding for hevein and a long C-terminal tail which was 144 amino acid residues in length. Hevein was found to be homologous to *Urtica dioica* agglutinin (UDA), a lectin that bound specifically to sugars like chitin that contained repeating N-acetylglucosamine units (GlcNAc) and displayed an inhibitory effect against a myriad of chitin-containing fungi *in vitro* [36]. Due to sequence homology, anti-fungal activity of hevein was tested against chitin-containing fungal strains *in vitro* and it was observed that treatment with hevein resulted in inhibition of fungal growth [38]. The anti-fungal activity of hevein was attributed to the presence of a chitin-binding domain distinguishing it from other CRPs like thionins [20], defensins [39] and cyclotides [40].

After its discovery in 1960, the structure of hevein was solved by X-ray crystallography at 2.8 Å and 1.5 Å in 1991 [41] and 2004 [42] including NMR studies in 1993 [43]. These structural analyses showed that the first three disulfide bonds in hevein formed a stable cystine-knot motif and the fourth bond

flanked the C-terminal. The small size and chitin-binding property of hevein made it an interesting candidate to study peptide-carbohydrate interactions. Asensio *et al.* first reported the NMR structure of the complex between hevein and N, N'-diacetylchitobiose $(GlcNAc)_2$ in 1995. They showed that hevein bound to $(GlcNAc)_2$ with millimolar affinity and amino acid residues Ser-19, Trp-21, Trp-23 and Tyr-30 were involved in the binding interaction. This binding motif was later termed the chitin-binding domain [44].

Since its discovery in 1960, hevein was the only member of this CRP family for 32 years (Fig. 1.2). Later, from 1992-1997, there was a rapid increase in the discovery of peptides like Ac-AMPs and IWF-4 which comprised six cysteine residues and were homologous to cysteine/glycine rich domains of chitin-binding proteins and hevein [45, 46]. These peptides are 29-30 residues in length, bound to chitin in a reversible manner and inhibited the growth of phytopathogenic fungi with a higher potency than hevein and other chitin-binding proteins. These peptides were considered "truncated" variants of hevein since they lacked the fourth disulfide bond at the C-terminal.

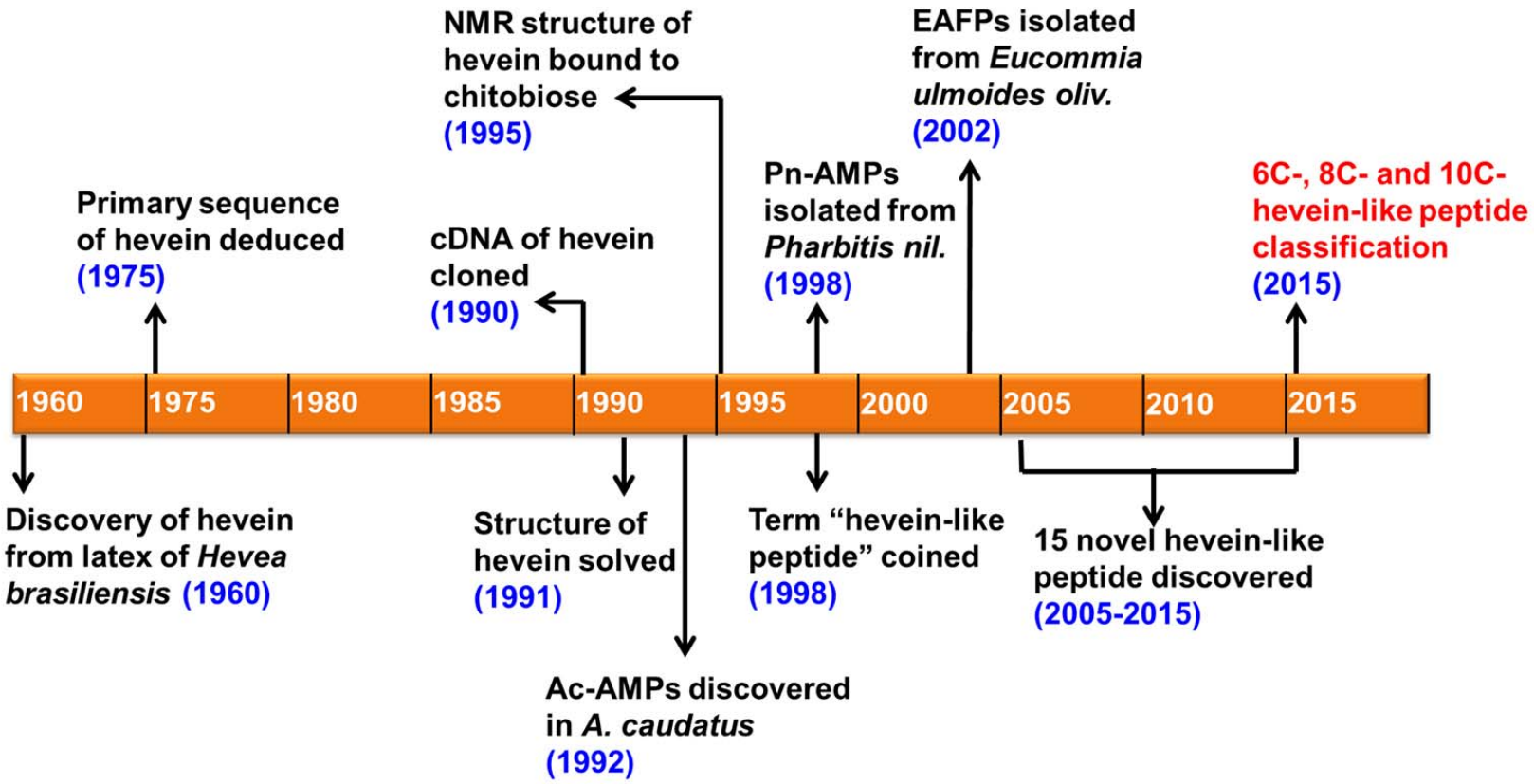


Figure 1.2. Timeline of major milestones in the hevein-like peptide family. Hevein, the prototypic member of the family, was first discovered in *Hevea brasiliensis* in 1960 and was sequenced in 1975. The cDNA of hevein was cloned in 1990 shortly after which the 3D-structure was solved by X-ray crystallography in 1991. Hevein remained the sole member of this family for 32 years after its discovery in 1960. In 1992, Ac-AMPs were discovered in *A. caudatus* and were considered "truncated" variants of hevein as they comprised six cysteine residues. The term "hevein-like peptide" was first coined in 1998 when Pn-AMPs were isolated from *Pharbitis nil.* In 2002, EAFPs, the first 10C-hevein-like peptides, were isolated from *Eucommia ulmoides*. In a span of 10 years from 2005-2015, a total of 15 new CRPs were discovered and the classification of this family into three sub-classes, 6C-, 8C- and 10C-hevein-like peptide was suggested.

The term “hevein-like peptide” was first coined in 1998 by Koo *et al.* who isolated two eight cysteine hevein homologs called Pn-AMPs from *Pharbitis nil* [47]. In 2002, natural variants of hevein that comprise ten cysteine residues were isolated from *Eucommia ulmoides* and named EAFP-1 and EAFP-2 *[48]*. NMR analysis of EAFP1 showed that it comprised similar cysteine connectivity as hevein with an additional disulfide bond connecting the C-terminal to the cystine core. Thus, the hevein-like peptide family is a highly diverse family in terms of its cysteine connectivity. Since 2003 the library of hevein-like peptides was expanded by the addition of 12 novel peptides. In a recent review published by our laboratory, we have classified hevein-like peptide into three sub-classes, namely 6C-, 8C- and 10C-hevein-like peptides based on their cysteine content [49]. Thus far there are a total of ten 6C-hevein-like peptides, five 8C-hevein-like peptides and five 10C-hevein-like peptides reported. Due to the diversity and widespread distribution of hevein-like peptides we hypothesize that they are potentially one of the chemical constituents that are constitutively expressed in plants.

2.2. Sequence diversity of hevein-like peptides

Hevein-like peptides contain 29-45 amino acid residues with strictly conserved cysteine and glycine residues. Thus far, four 6C-hevein-like peptides have been isolated and characterized, namely, two Ac-AMPs from *Amaranthus caudatus* [45], one IWF-4 from *Beta vulgaris* [46] and one Ar-AMP from *Amaranthus retroflexus* [50]. 6C-hevein-like peptides contain 29-30 amino acid residues and six cysteine residues.

The cysteine-spacing pattern of 6C-hevein-like peptides is X_3**C**-X_4**C**-X_4**CC**-X_5**C**-X_6**C**-$X_{1-2,}$ where X represents non-cysteine amino acid residues and the number represents the number of amino acid residues. The residues between two cysteines are termed a "loop" and named according to the preceding cysteine. The majority of hevein-like peptides including hevein contain eight cysteine residues. Currently, five 8C-hevein-like peptides reported are hevein, two Pn-AMPs from *Pharbitis nil* [47] and two Fa-AMPs from *Fagopyrum esculentum* (Buckwheat) [51]. These hevein-like peptides contain 37-44 amino acid residues with a cysteine-spacing pattern of iX_{2-6}**C**-X_{1-9}**C**-X_{4-5}**CC**-X_5**C**-X_6**C**-X_{1-5}**C**-X_{2-6}**C**-X_{2-10}.

In 10C-hevein-like peptides, modifications in the position of cysteine residues result in three distinct cysteine-spacing patterns which we have designated Type I, Type II and Type III. Type I, X_2**C**-X_3**C**-X_3**C**-X_4**CC**-X_5**C**-X_6**C**-X_4**C**-X_1**C**-X_1**C**-X_2 was observed in Ee-CBPs from *Euonymus europaeus* [52]. Type II, X_2**C**-X_8**C**-X_4**CC**-X_5**C**-X_6**CC**-X_4**C**-X_3**C**-X_2**C**-X_1 and Type III, X_{2-3}**C**-X_{4-8}**C**-X_2**C**-X_1**CC**-X_5**C**-X_6**C**-X_{3-4}**C**-X_3**C**-X_{2-3}**C**-X_{2-5} were previously reported in WAMPs isolated from *Triticum kiharae* [53] and EAFPs isolated from *Eucommia ulmoides [48]*, respectively. This promiscuity of the cysteine residues in 10C-hevein-like peptides may lead to functional diversity among the 10C-hevein-like peptide family. Similar observations have been made in defensins and conotoxins where minor modifications in the primary sequence result in alterations of defensin activity and specificity of contotoxins for distinct ion channels, respectively [54, 55].

Based on the cysteine spacing pattern of 6C-, 8C- and 10C-hevein-like peptides they can be divided into four, six and seven loops, respectively (Table 1.2). In

6C-hevein-like peptides each loop is four to five residues in length. Sequence comparison shows that Asn and Arg in loop 1, Pro in loop 2, Ser, Phe, in loops 3 and Tyr in loops 3 and 4 along with all Gly residues are absolutely conserved. The C-terminal comprises a conserved Gly and Arg residue while the N-terminal shows greater sequence diversity. In 8C-hevein-like peptides, loop 1 is the longest loop and is eight residues in length while the length of loop 5 is highly variable (3-5 residues). A conserved Glu residue is present in loop 1 which may play a role in stabilizing the framework of 8C-hevein-like peptides by forming hydrogen bonds with neighboring residues. A Leu residue in loop 2 and Ser, Glu, Trp and Gly residues in loop 3 are absolutely conserved. The sequence of loop 4 differs considerably from 6C-hevein-like peptides but comprises a conserved Tyr residue that is known to be crucial for chitin-binding activity. Loop 5 is the most variable while loop 6 shows higher sequence conservation and comprises a consensus sequence CQS(N/Q)C. The 10C-hevein-like peptides show the highest sequence variation and comprise a conserved Leu and Glu in loop 2 and loop 6, respectively (named according to Ee-CBP) (Table 1.2).

A striking feature that distinguishes the hevein-like peptide family from other CRPs is the presence of a hevein domain [56] or chitin-binding domain [57] between loops 3 and 4. Sequence alignment of the reported 6C-, 8C- and 10C-hevein-like peptides shows that the binding motif is conserved throughout the hevein-like peptide family (Table 1.2).

Table 1.2. Consensus sequences of reported hevein-like peptides.

Peptide	Peptide Sequence
Loop	1 2 3 4 5 6 7
	* * * *
Ac-AMP2	VGEC-VRG---RCPSGMCCSQFGYCGKGPKYCG----------
Ar-AMP	AGEC-VQG---RCPSGMCCSQFGYCGRGPKYCGR----------
IWF-4	SGECNMYG---RCPPGYCCSKFGYCGVGRAYCG----------
Hevein	-EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKD--
Fa-AMP	-AQCGAQGGGATCPGGLCCSQWGWCGSTPKYCG--AGCQSNCK---
Pn-AMP	-QQCGRQASGRLCGNRLCCSQWGYCGSTASYCG--AGCQSQCRS--
EAFP-1	-QTCASRCP-RPCNAGLCCSIYGYCGSGNAYCG-AGNCRCQCRG--
Ee-CBP	-QQCGRQAGNRRCANNLCCSQYGYCGRTNEYCCTSQGCQSQCRRCG
Wamp-1a	AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCG-AGSCQSQCRGC-

Sequences of reported hevein-like peptides are aligned and they cysteine residues are highlighted in yellow. Amino acid residuesbetween two cysteine residues are termed a loop and numbered according to the first cysteine residue in each loop. The conserved chitin-binding domain is indicated by asterisk.

The chitin-binding domain is a motif comprising a conserved serine residue, one glycine residue and three aromatic residues present in the sequence S-X-(F/W/Y)-X-(F/W/Y)-C-G-X_4-Y organized around a stable core of three to five disulfide bonds [26, 58]. Multiple repeats of the chitin-binding domain are found in proteins like wheat germ agglutinin (WGA) [59], *Urtica dioica* agglutinin (UDA) [36] and several plant chitinases where the chitin-binding domain is fused to an unrelated catalytic domain [60]. It was speculated that chitinases bound to chitin-rich fungal cell walls by virtue of the chitin-binding domain, eliciting anti-fungal effect. However, peptides like hevein [31], its natural variant, pseudohevein, with six amino acid alterations [61] as well as all 6C-, 8C- and 10C-hevein like peptides comprise a single chitin-binding domain which is not fused to a catalytic domain and are more potent anti-fungal agents than their protein counter-parts (Fig. 1.3). The chitin-binding domain is also of medical interest as it is the major allergen involved in the latex-fruit syndrome [62]. The latex-fruit syndrome is an increased sensitivity to fresh fruits like banana, avocado and water chestnuts in individuals who are allergic to natural rubber latex. It is hypothesized that cross-reactivity of allergens occurs as IgEs recognize the chitin-binding domain of Class I chitinases in fruits due to its similarity to hevein found in latex [63]. Thus, studies have been done to understand the structural features of the chitin-binding domain to gain insight into the distinct conformational epitope responsible for allergic reactions [64, 65].

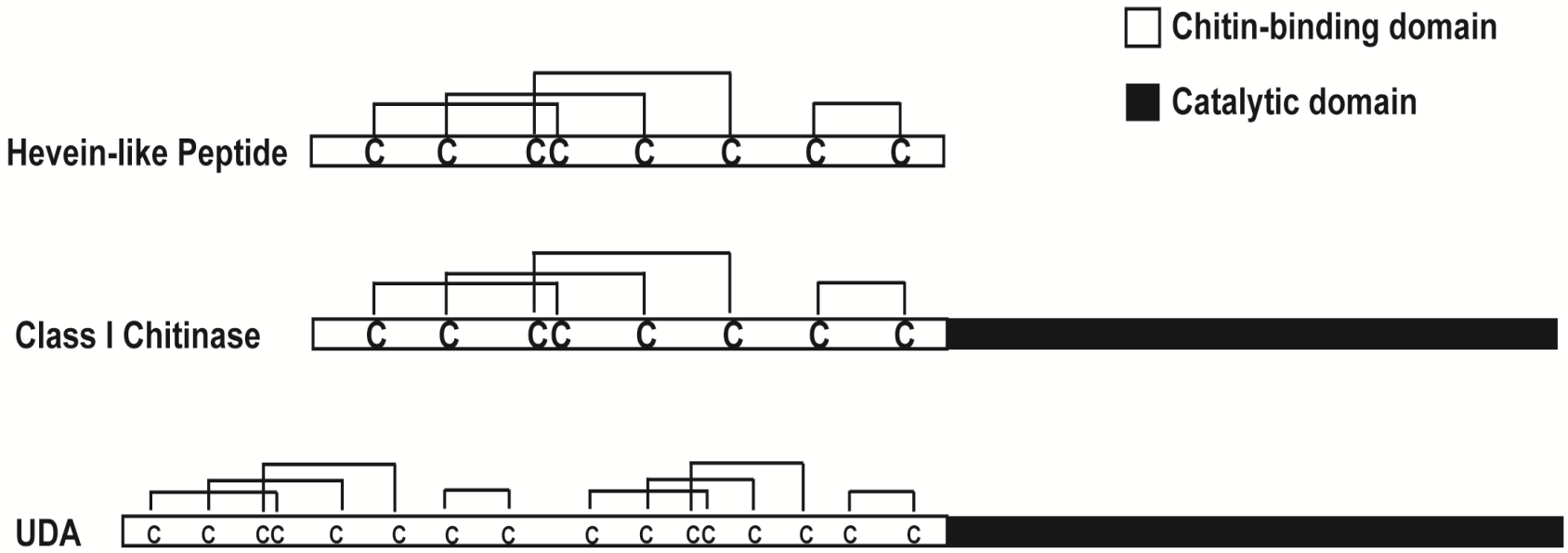


Figure 1.3. Schematic representation of primary structure of hevein-like peptides, Class I chitinases and Urtica dioica agglutinin (UDA). Hevein-like peptides comprise a single chitin-binding domain, class I chitinases comprise a single chitin-binding domain fused to a catalytic chitinase domain and UDA consists of multiple repeats of the chitin-binding domain linked to a catalytic domain.

2.3. Structure of hevein-like peptides

The secondary structure of hevein-like peptides is comprised of a coil-β1-β2-coil-β3 motif which varies in the number of short turns in the two long coiled regions [49]. This arrangement differs from other CRPs like defensins and thionins which comprise a CS$\alpha\beta$ and γ-fold structural motifs, respectively [49]. The C-terminal of hevein-like peptides generally comprises a short helical segment [66]. The central β-sheet of the hevein motif is formed by two anti-parallel β-strands and the core is stabilized by disulfide bonds [49]. The conserved cysteine residues form a highly stable knotted topology called the cystine knot motif [67]. The first two disulfides (Cys I-IV and Cys II-V) form a ring through which the third bond (Cys III-VI) penetrates (Fig. 1.4). Interestingly, this cystine-knot motif is widespread in a variety of phyla including animals, plants and fungi and is highly prevalent in venoms of spiders, scorpions and cone snails [68].

The 6C-hevein-like peptides are the smallest hevein-like peptides that can elicit anti-fungal effects on a myriad of chitin-containing fungi. The structure of these peptides contains two anti-parallel β-strands which form the main central β-sheet, an N-terminal coil region and a C-terminal helical coil. Two disulfide bonds link the N-terminal coil region to the two central β-strands while the third disulfide bond connects the C-terminal coil to the first β-strand (Fig. 1.5 A). This structural framework allows the residues involved in chitin-binding to be exposed on the surface [69].

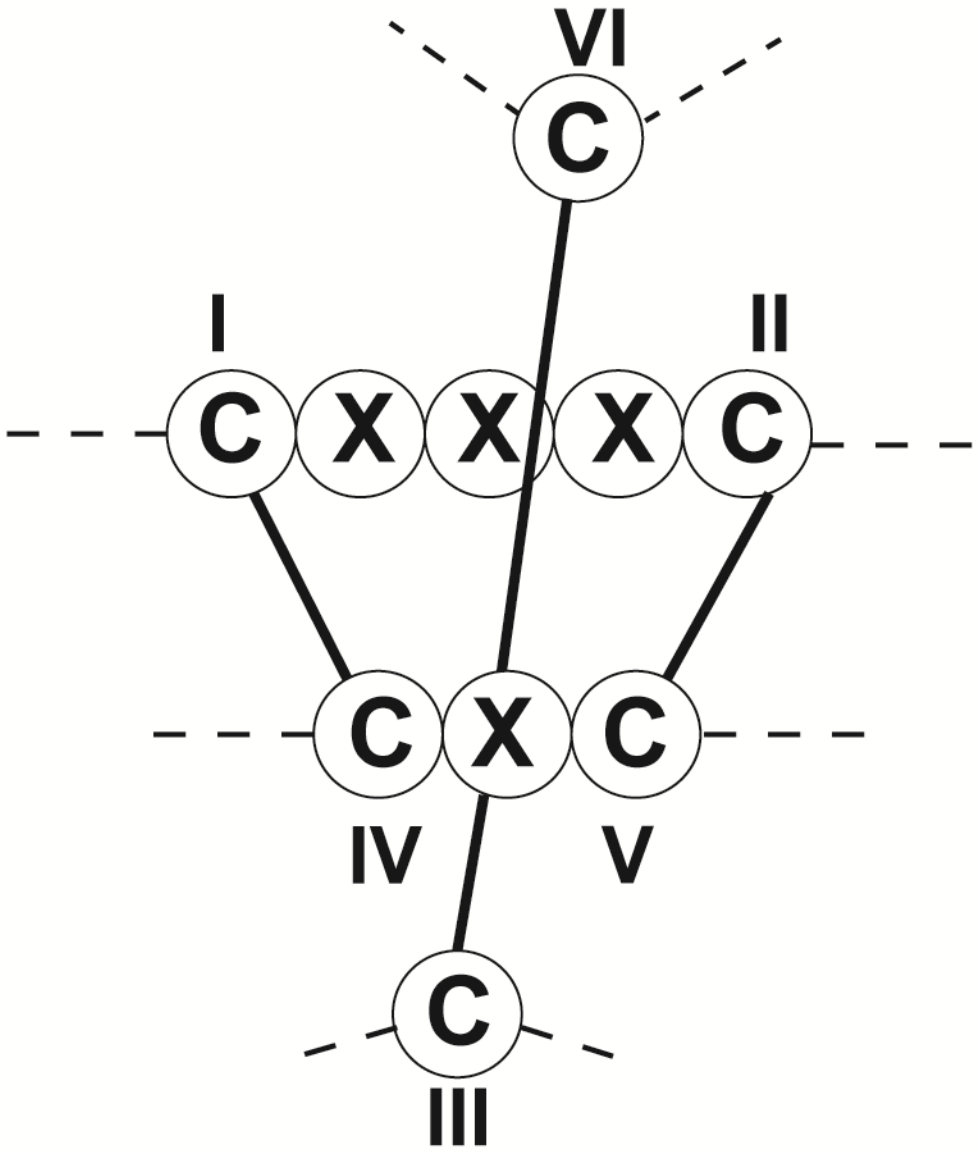


Figure 1.4. Cystine-knot motif. The motif has an embedded ring formed by two disulfide bonds between Cys I- Cys IV and Cys II- Cys V. A third disulfide bond between Cys III- Cys VI penetrates through this ring resulting in a knotted motif. The disulfide bonds are indicated by thick lines while dotted lines represent the peptide backbone. 'X' represents any non-cysteine amino acid

Comparison with 6C-hevein-like peptides shows that 8C-hevein-like peptides contain an additional loop at the C-terminal formed by the fourth disulfide bond. This loop constitutes the third β-strand in the antiparallel β-sheet formed in 8C-hevein-like peptides and is parallel to the fourth disulfide bond. Therefore the structure of 8C-hevein-like peptides is comprised of a central β-sheet formed by three antiparallel β-strands. (Fig. 1.5 B).

The 10C-hevein-like peptides contain five disulfide bonds but vary in the position of the fifth disulfide bond. Examples include EAFP1 and EAFP2 isolated from the bark of *Eucommia ulmoides Oliv.* [48], WAMP-1a and WAMP-1b isolated from *Triticum kiharae* seeds [53] and Ee-CBP from *Euonymus europaeus* [70]. A striking feature that distinguishes 10C-hevein-like peptides from 8C-hevein-like peptides is the presence of an additional disulfide bond within the molecule. The 10C-hevein-like peptides are the most structurally diverse sub-class of hevein-like peptides as their 3D-structure varies depending on the position of the fifth disulfide bond. NMR and X-ray crystallography of EAFP2 revealed that it contains a 3_{10} helix (Cys3-Arg6), an α-helix (Ala27-Cys31) and three anti-parallel β-sheets (Cys16-Ser18, Cys23-Ser25 and Cys35-Cys37). Residues 11-30 of EAFP1 constitute a hydrophobic patch in the structure and include the chitin-binding domain.

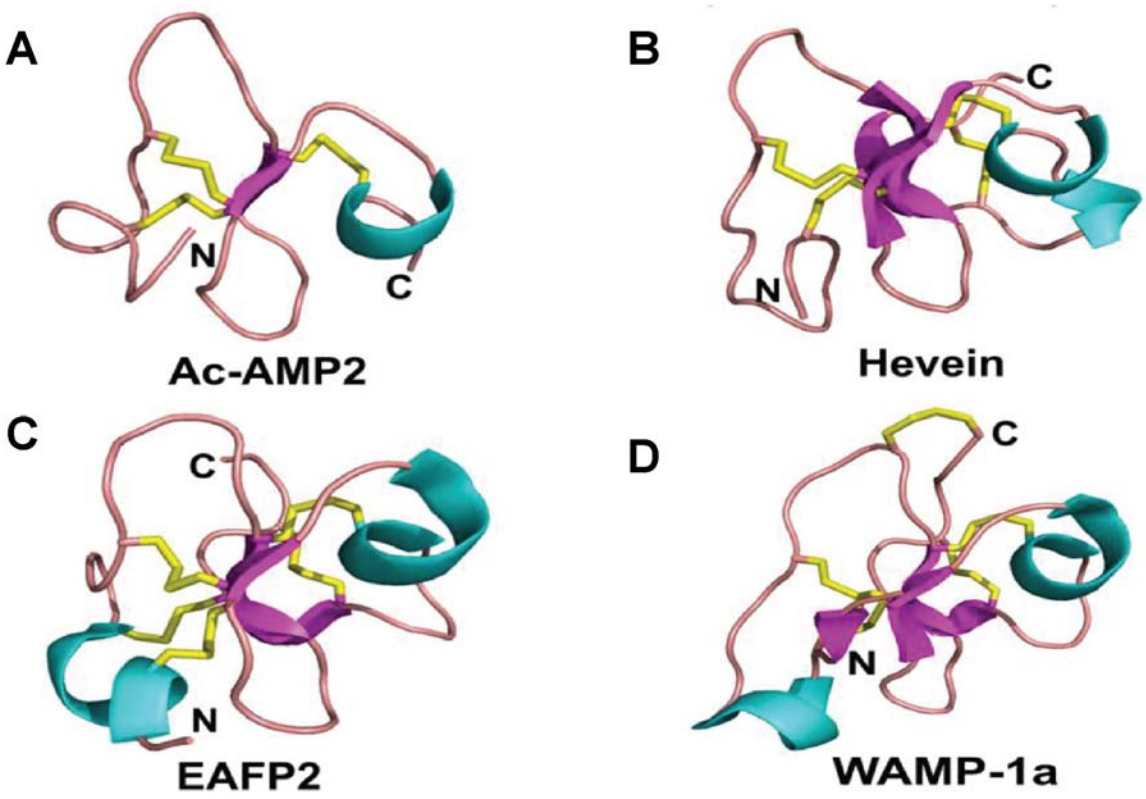


Figure 1.5. NMR structures and scheme of binding interaction. Tertiary structures of (A) Ac-AMP2 (63), (B) hevein (34), (C) EAFP-2 (66) and (D) WAMP-1a (68). The secondary structure is depicted by different colors, cyan- α-helix, magenta- β-sheet, yellow- disulfide bonds. Figure adapted from J.P Tam *et al*. (40).

Disulfide mapping indicated that the first four disulfide bonds of EAFPs are similar to hevein and the fifth disulfide bond connects N-terminal Cys 7 with Cys 37 at the C-terminal [71]. This connection of the N- and C-termini results in the rigidification of the conformation with the presence of a cationic surface. The amphiphilic nature of EAFP1 can be attributed to the presence of the cationic and hydrophobic patches in the structure. Structural analysis of EAFP2 revealed that the three Tyr residues of the chitin-binding domain are exposed to the solvent resulting in the formation of a hydrophobic pocket around the Ser residue. Hydrogen bonding between this Ser residue and the sugar further stabilize the orientation of the Ser for efficient binding to chitin oligosaccharides [72] (Fig. 1.5 C). The fifth disulfide bond of Wamp-1a connects the C-terminal to loop 1 of the structure resulting in a cysteine motif similar to chitinases [73] (Fig. 1.5 D). Similar to EAFP1 the fifth disulfide bond brings the positively charged residues of the N- and C-termini in close proximity resulting in a cationic surface [53].

2.4. The Hevein-Chito-Oligosaccharide complex

2.4.1. Active site involved in the binding interaction

The small size and easy availability of hevein and hevein-like peptides using isolation and purification [31], molecular biology [65] and chemical synthesis [74] makes them ideal candidates to study their chitin-binding interaction. Structural and thermodynamic data of the complex between chitin oligosaccharides $(GlcNAc)_{1-5}$ and hevein [75], pseudohevein [76], WGA [77], Ac-AMP2 [78], Hev32 (a truncated hevein mutant) [79] and Hev32S19D (a hevein-32 mutant

comprising a Ser19 to Asp19 mutation) [80] have been obtained using techniques like isothermal titration calorimetry (ITC) [66, 81], NMR spectroscopy [44, 66], X-ray crystallography [42] and ultracentrifugation [58]. These studies have unequivocally shown that aromatic residues occupying the chitin-binding domain stabilize the complex with chitin through CH-π stacking interactions and van der Waals contacts. A sequence logo was obtained by aligning the chitin-binding domains of 6C-, 8C- and 10C-hevein-like peptides to illustrate the sequence conservation and relative frequency of appearance of each amino acid at a specific position. Amino acid residues occupying the chitin-binding domain were numbered from I to V, based on their position in the sequence logo (Fig. 1.6 A). The serine residue at position I is conserved in all hevein-like peptides except WAMP-1b, where serine is replaced by glycine. In the second position, an aromatic amino acid is preferred with higher occurrence of Phe in 6C-hevein-like peptides, Trp in 8C-hevein-like peptides and Tyr in 10C-hevein-like peptides. Any aromatic amino acid is favored at position III while glycine and tyrosine at positions IV and V, respectively are absolutely conserved in all hevein-like peptides. X-ray and NMR studies on complexes of WGA, hevein and Ac-AMP2 with chitin oligosaccharides have clearly indicated the presence of two binding subsites, namely subsite +1 and subsite +2 at the structural level. The aromatic residues at position III and position II of the chitin-binding domain comprise subsite +1 and subsite +2, respectively (Fig. 1.6 A). Hydrogen bonding between Ser I, named according to the Weblogo of the chitin-binding domain, and an acetamide moiety of one GlcNAc residue stabilizes the complex.

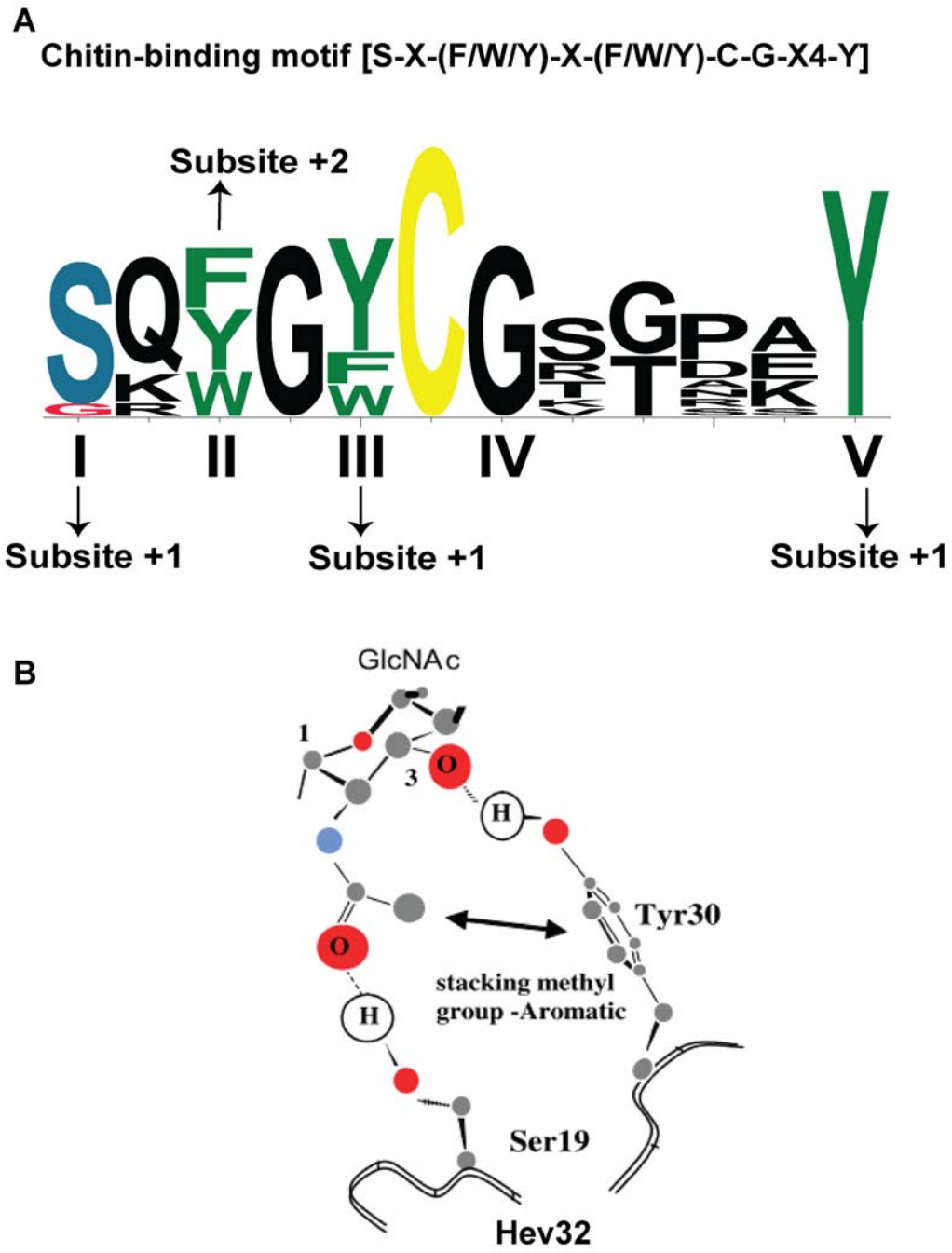


Figure 1.6. Sequence logo of the chitin-binding domain and scheme of binding interaction. (A) Amino acid residues in positions labelled I to V are involved in the binding interaction with chitin oligosaccharides. Residues at position I, III and V are part of subsite +1 and position II is a part of subsite +2. (B) A hydrogen bond between Ser-19 of HEV32 and the acetamide moiety and Tyr-30 with OH-3 of GlcNAc stabilize the binding interaction. A stacking interaction between the aromatic ring of Tyr-30 and the acetamide moiety further stabilizes the complex. Scheme adapted from M. Chavez *et al*. (63)

Another hydrogen bond between conserved Tyr V of the chitin-binding domain and OH-3 of the same GlcNAc unit was found to be important in the complex between WGA and chitin oligosaccharides [80, 82-84] (Fig. 1.6 B). Ser I and Tyr V of the chitin-binding domain are also part of subsite +1 and are important for binding. Thermodynamic data from ITC experiments have revealed that hevein binds chitin oligosaccharides with a K_a of 8,500 M^{-1} at 303K and the binding process is enthalpically driven (Table 1.3). The binding interaction was not calcium-dependent and the negative free energy (ΔG^0) and enthalpy (ΔH^0) values are indicative of van der Waals forces and hydrogen bond interactions [44].

2.4.2. Effect of mutations on carbohydrate binding

A number of studies have been done to understand the effect of mutation on the binding of hevein-like peptides to chitin oligosaccharides. To gain an insight into the role of the C-terminal of hevein in binding, Hev32, a hevein mutant comprising a C-terminal truncation of 10 residues was chemically synthesized by Aboitiz *et al*. Comparison of ΔG^0 values from calorimetry experiments of hevein and Hev32 with $(GlcNAc)_6$ showed that C-terminal truncation of hevein leads to a loss of 3.6 ± 0.7 kJ/mol in ΔG^0. This data suggests that there is no significant loss in binding affinity when the C-terminal of hevein is truncated. However, vant Hoff plots show that Hev32 has a larger ΔH^0 value compensated for by increased change in entropy (ΔS^0).

Table 1.3. Association constants (K_a) and thermodynamic parameters of the binding interaction of hevein-like peptide and chitin oligosaccharides

	K_a [M^{-1}]				Thermodynamic Parameters		
					ΔG	ΔH	ΔS
	T = 298 K	T = 303 K	T = 308 K	T = 313 K	[kJmol^{-1}]	[kJmol^{-1}]	[kJmol^{-1}]
Hevein*(ITC)		8,500			-22.6	-34.7	-41.4
Hevein*(NMR)	11,500	8,700	6,900	5,700	-23.0	-34.7	-44.7
Hevein**(NMR)	-	474,000	-	-	-32.6	-46.1	-45.0
Hev32*(NMR)	7,700	4,200	3,400	2,200	-21.8	-62.6	-136.0
Hev32S19D**(NMR)	2,800	2,400	2,200	1,400	-19.7	-33.7	-46.4
Hev32S19D*(NMR)	-	2,400					
Hev32S19D***(NMR)	150	110	90	80	-11.8	-26.2	-47.6
Ac-AMP2*(NMR)				1,000			

K_a and thermodynamic parameters between hevein, Hev32, Hev32S19D, Ac-AMP2 with $(GlcNAc)_3$ (*), $(GlcNAc)_5$ (**) and $GlcNH_2\beta(1\text{-}4)$ $(GlcNAc)_4$ (***) by NMR and ITC. K_a and thermodynamic parameters adapted from Chavez *et al.* and Verheyden *et al.* (*73, 75*).

This increased ΔS^0 value could be attributed to a reduced flexibility of the side chains of amino acid residues upon binding, due to the C-terminal truncation. Thus, Aboitiz *et al.* named the N-terminal 32 residues as the 'minimum hevein domain' required for effective binding to the chitin oligosaccharides [79]. Since Ser and aromatic residues occupying the chitin-binding domain are primarily involved in binding to chitin oligosaccharides, effect of mutations in each of these positions on chitin-binding was studied. To quantify the energy value of the hydrogen bond between Ser I and the acetamide moiety of the sugar, Chavez *et al.* mutated the Ser residue of Hev32 to Asp (Hev32S19D), as it is of the same chain length as Ser but lacks the side-chain hydrogen bond donor group (Fig. 1.7). Calorimetry experiments showed that there was a drastic decrease in the binding affinity of Hev32S19D with $(GlcNAc)_3$ as compared to Hev32. This decrease could be attributed to the elimination of the hydrogen bond between the peptide and the acetamide moiety of the sugar. Due to the loss of the hydrogen bond, there is not enough driving force for the rearrangement of the Trp II residue for efficient CH-π stacking interactions. Thus, it can be concluded that the intermolecular hydrogen bond between Ser I and the carbonyl of the acetamide group on the sugar is crucial for efficient binding interactions [80].

The relevance of Trp II in the chitin-binding domain of hevein, was analyzed by Asesnio *et al*. who performed ^{1}H-NMR titration and ITC experiments of pseudohevein, a natural mutant of hevein comprising a Tyr residue at position II of the chitin binding domain instead of Trp [76].

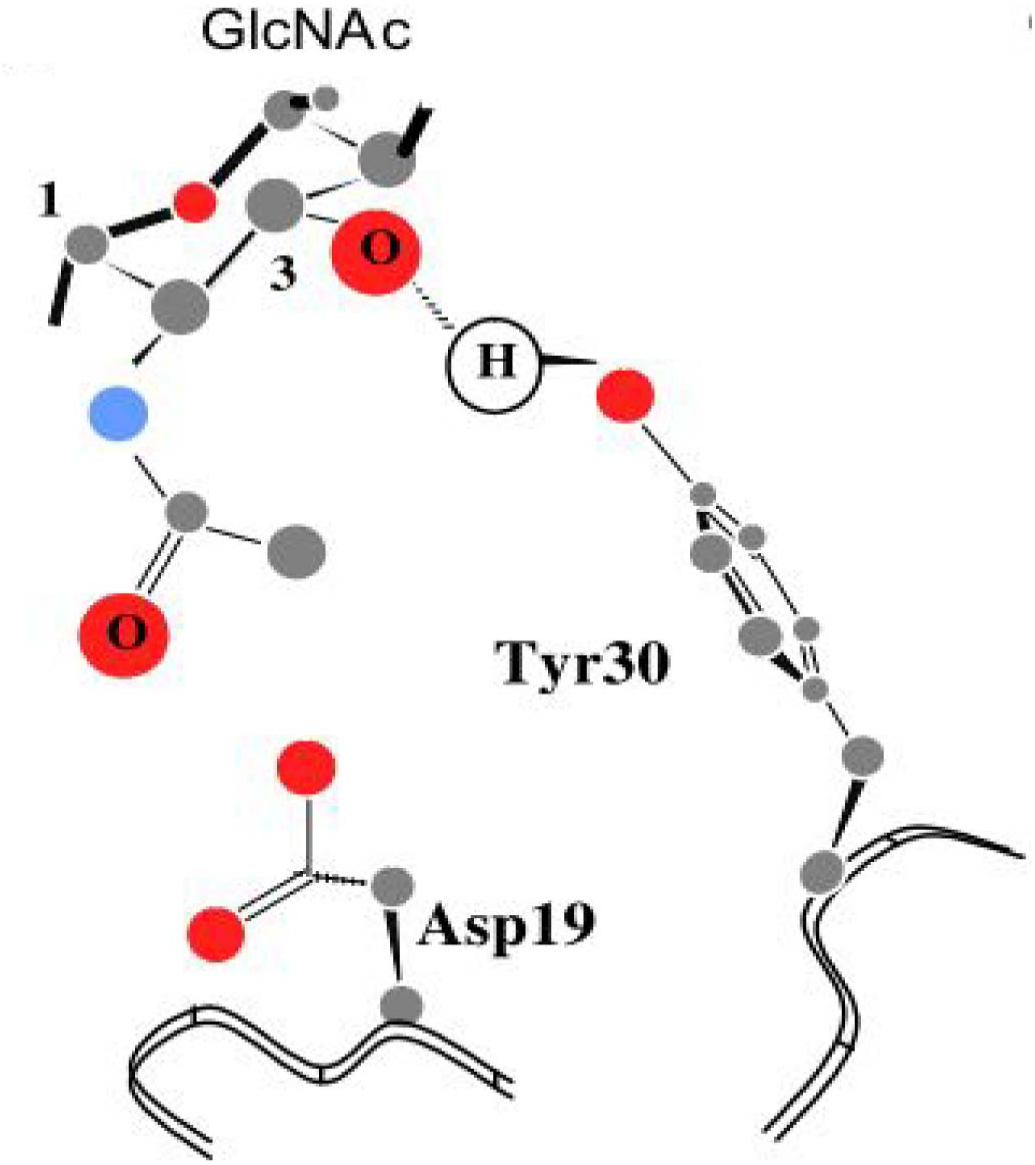


Figure 1.7. Effect of mutation on chitin-binding interaction. Mutation of the Ser-19 residue to Asp-19 prevents the formation of a hydrogen bond with the acetamide moiety, thus drastically affecting the binding affinity with GlcNAc. Scheme adapted from M. Chavez *et al*. (63)

Significant changes in chemical shift of a number of peaks in ^{1}H-NMR spectra indicated the formation of complexes between the sugar and pseudohevein. Comparison of ITC data for hevein and pseudohevein titrated with $(GlcNAc)_3$ revealed a difference of <1kJ/mol in ΔH^{0} values. This data suggests that there is no substantial change in binding affinity when Tyr replaces Trp II of the chitin-binding domain. In Ac-AMP2, position II is occupied by a phenylalanine residue.

Muraki *et al*. replaced Phe II with non-natural amino acid residues like 18-β-(cyclohexyl)alanine (Cha18), 18-β-(p-nitrophenyl) alanine (pNO2-Phe18) and 18-β-(pentafluorophenyl) alanine (F5-Phe18) which resulted in enhanced, unchanged and reduced binding affinity, respectively when compared with wild type Ac-AMP2 [74]. Replacement of each of the aromatic amino acid residues occupying the chitin-binding domain with an Ala residue resulted in reduction of affinity in the order Tyr III>Tyr IV>Phe II [74]. This data further validates that the stacking interaction between the aromatic side-chain and the apolar carbohydrate surface is essential for efficient binding to occur [85].

2.4.3. Multivalent nature of the binding interaction

From NMR and ITC experiments it was established that hevein-like peptides bind to chitin oligosaccharides with millimolar affinity. Initial modeling experiments on WGA showed that GlcNAc residues occupying the center of the polymer chain of chitin oligosaccharides would be sterically hindered and not be involved in binding [80, 82-84]. It was therefore suggested that hevein-like peptides bound chitin oligosaccharides only at the terminal non-reducing end of the sugar in a 1:1

stoichiometry. Computer modelling of the complex between hevein and $(GlcNAc)_4$ by Asensio *et al*. showed that the reducing end of the sugar is not bound by the peptide and is completely exposed to the solvent [58]. However, this hypothesis could not be used to explain the efficient recognition and binding of hevein-like peptides to the long chitin chain. Asensio *et al*. reported that the use of longer chitin chains $(GlcNAc)_n$, where n>2 are more appropriate models to gain insight on the location of the sugar in the binding site. They reported that from n=1 to n=3 a one-order-of-magnitude increase in binding affinity was observed [58]. This increase could be attributed to the fact that longer chitin chains provide more number of binding sites. Thus, this data suggests that more than one hevein molecule can bind to the sugar, thereby refuting the 1:1 stoichiometry hypothesis. This result was further validated by assessing the molecular weights of hevein-chitin oligosaccharide complexes with different pepetide:ligand ratios by analytical ultracentrifugation. For chitin oligosaccharides upto n=4, the chain length is too short to allow multiple binding sites, therefore 1:1 complexes were observed. However, when $(GlcNAc)_5$ and $(GlcNAc)_8$ were used, 2:1 and to an extent 3:1 peptide:sugar complexes were observed. To further validate these results, titration experiments of hevein with LacNAc were performed. Despite the presence of a terminal GalNAc residue instead of GlcNAc at the non-reducing end, changes in chemical shift were observed indicating that binding can occur at residues other than the non-reducing end of the sugar. Thus, it can be concluded that hevein binds chitin in a multivalent fashion, the binding process is enthalpically driven and is mediated by hydrogen bond interactions

and van der Waal's contacts between the sugar and the residues of the chitin-binding domain [77].

2.5. Biosynthesis of hevein-like peptides

In plants, hevein-like peptides are ribosomally synthesized peptides encoded by dedicated genes as linear precursors. These precursors are then processed to produce the mature hevein-like peptides. Insight into the distribution, evolution and precursor organization of hevein-like peptides has been obtained by genomic and transcriptome data analysis. Sequence alignment of precursors obtained by cDNA cloning and transcriptome data mining of 6C-, 8C- and 10C-hevein-like peptides reveal that the overall precursor organization is conserved (Fig. 1.8 A) [46, 47, 50, 70, 86-89]. The precursors are comprised of three domains, an N-terminal signal peptide domain, followed by a mature CRP domain and a C-terminal tail. This precursor organization suggests that hevein-like peptides are ribosomally synthesized and their biosynthesis involves the secretory pathway [90, 91]. In the secretory pathway, the N-terminal signal peptide directs the precursor peptide to the endoplasmic reticulum where the signal peptide is cleaved by SPaseI followed by cleavage of the C-terminal tail by an endopeptidase and subsequent release of the mature peptide (Fig.1.8 B). Hevein-like peptide precursors differ from CRPs like cyclotides that comprise an N-terminal pro-domain before the mature peptide domain [92-94].

An interesting feature of hevein-like peptide precursors is the presence of a C-terminal tail. The cDNA of hevein was cloned in 1990 and comprised a 144-

amino-acid-long C-terminal tail homologous to the Barley wound-inducing protein (Barwin) [88]. In 1993, the gene for Ac-AMP2 was cloned by De Bolle *et al.* who found that the C-terminal tail of the Ac-AMP2 precursor showed no sequence homology to any known proteins in the Swiss-Prot databank [86]. It was found that the C-terminal tail of Ac-AMP2 showed similar features as the carboxyl-terminal of Gramineae lectins and tobacco glucanase [86, 95]. These features include the presence of an N-glycosylation site [45], amino acid residues with short hydrophobic side chains and a valine residue eight amino acid residues ahead of the glycosylation site [96]. The carboxyl terminal of Gramineae lectins is known to be involved in the sorting of peptides to be transported to the vacuoles, thus it was speculated that the C-tail of Ac-AMP2 may play a similar role [86]. However this hypothesis was later disproved by De Bolle and group who studied the processing, sorting and intracellular localization of recombinant Ac-AMP2 in transgenic tobacco leaves [97]. Similarly, precursors of 6C- and 8C-hevein-like peptides, Ar-AMP, IWF-4 and Pn-AMPs, respectively, comprised C-terminal tails bearing no homology to any known proteins domains [46, 50, 98]. Cloning of the cDNA of Ee-CBPs, 10C-hevein-like peptides from *Euonymus europaeus* revealed that they are expressed as class-I-chitinase-like precursors and post-translational cleavage of the C-terminal chitinase domain yields the mature Ee-CBP peptide [52, 70]. However, in WAMPs, 10C-hevein-like peptides from *Triticum kiharae*, the C-terminal type-I chitinase domain observed in Ee-CBPs is absent and the C-terminal tail bears no homology to any known conserved domains.

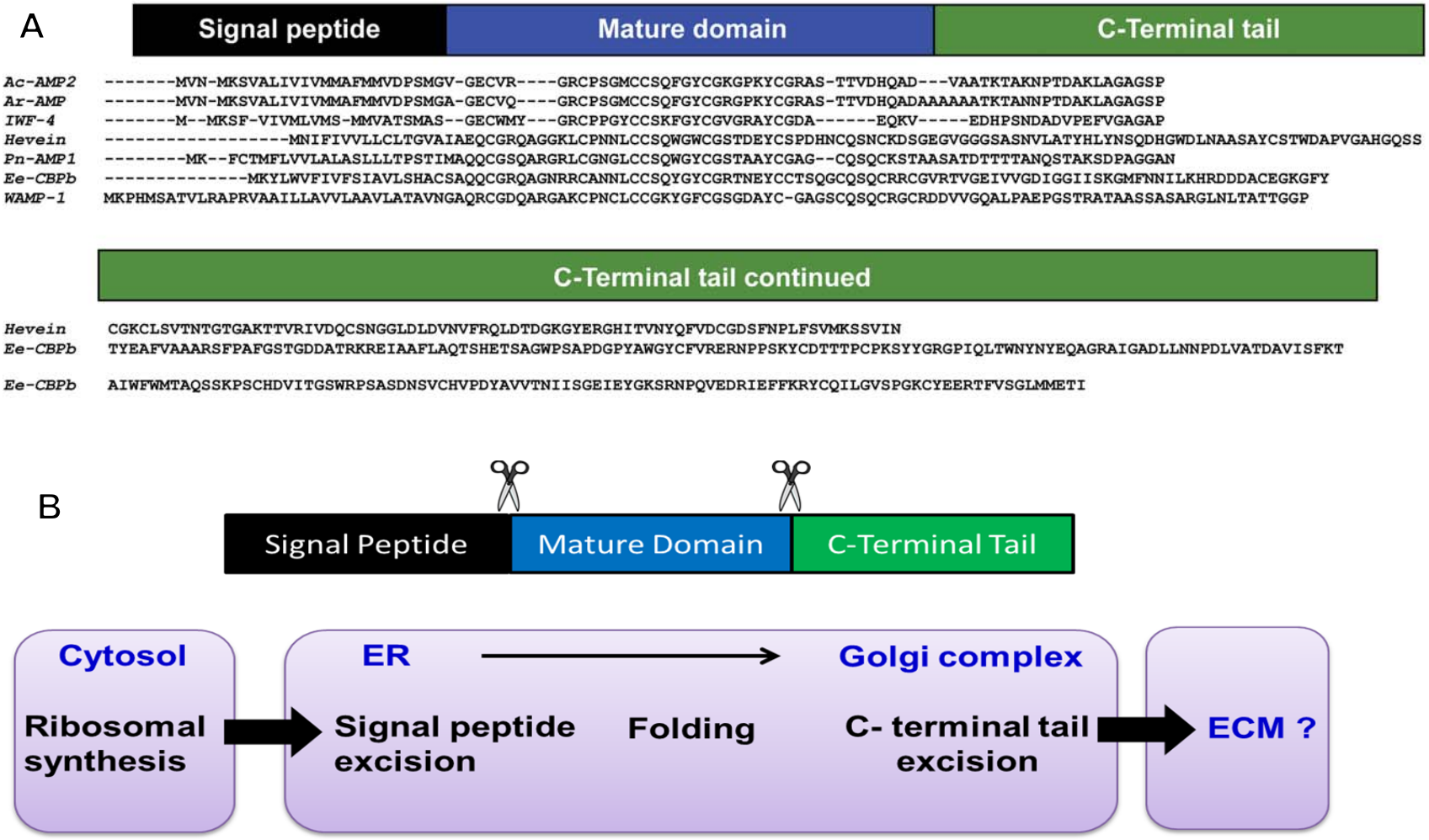


Figure 1.8. Biosynthesis of hevein-like peptides. (A) Sequence alignment of precursors of 6C-, 8C- and 10C-hevein-like peptides showed that the precursor organization is conserved and comprises a signal peptide, followed by a mature domain and a C-terminal tail. (B) Putative biosynthesis pathway of hevein-like peptides. ER- endoplasmic reticulum, ECM- extra-cellular matrix.

The promiscuity of the precursors of hevein-like peptides suggests that they are a highly diverse class of CRPs and further analysis of their precursor organization and biosynthesis may be useful in predicting their evolutionary relationship.

2.6. Anti-microbial activity of hevein-like peptides

The anti-fungal activity of hevein was studied by Van Parijs *et al*. due to its sequence homology with UDA, a carbohydrate binding protein that could inhibit the growth of chitin-containing fungi. Although hevein lacks the chitinase domain, it could inhibit the growth of eight chitin-containing fungi with IC_{50} in the range of 90-1250 µg/ml. This anti-fungal activity of hevein was retained upon heat treatment and resulted in morphological changes like the formation of thick hyphae with buds [38]. 6C-hevein-like peptides, Ac-AMP2, IWF-4 and Ar-AMP are relatively more potent inhibitors of fungal growth when compared to hevein (Table 1.4). Treatment with IWF-4 resulted in the formation of shorter, more branched hyphae as compared to the control cultures [46]. The anti-fungal property of Ac-AMP2 and IWF-4 was resistant to heat treatment but was antagonized in the presence of calcium ions [45, 46]. Ar-AMP (4-32 µM) had a protective effect on barley seedlings inoculated with *Helminthosporum sativum*, a fungus responsible for root rot [50].

Interestingly, 8C-hevein-like peptides like Pn-AMPs isolated from *Pharbitis nil* could inhibit the growth of chitin-containing fungi, *oomycetes* (chitin-free fungi) and gram-positive bacteria [47]. Similarly, Fa-AMPs isolated from *Fagopyrum*

esculentum could inhibit the growth of gram-positive and gram-negative bacteria apart from chitin-containing fungi [51].

The 10C-hevein-like peptides, EAFPs from *Eucommia ulmoides Oliv.,* could inhibit the growth of chitin-containing fungi with IC_{50} in the range of 18-155 µg/ml [48] while Ee-CBPs from *Euonymus europaeus* and WAMPs from *Triticum kiharae* were more potent than Ac-AMP2 in inhibiting phytopathogenic fungi and could also inhibit gram-positive bacteria [53, 99]. The anti-fungal activity of Ee-CBPs was similar to hevein and was inhibited in the presence of cations but was resistant to heat and resulted in morphological changes like bulbing, branching and stunted growth of hyphae accompanied by spore release.

Although the fungal growth inhibition property of hevein-like peptides is known, the mechanism of action is still not clear. Currently, three proposed mechanisms of action have been suggested. Initially, Van Parijs *et al*. speculated that since hevein is localized in the vacuole-derived lutoid bodies of the plant cells, it might be secreted with a cascade of lytic enzymes if a fungal pathogen breaks through the plasma membrane and enters the cell. As a consequence of its small size and ability to bind to chitin, hevein could then penetrate through fungal cell walls and inhibit growth by binding to nascent chitin chains and interfering with the "steady state" of hyphal growth [38, 100] . Subsequently, hevein's chitin-binding property could also create an imbalance between chitin-synthesis and hydrolysis of pre-formed chitin chains, which is essential for growth of hyphae. However, this mechanism is not sufficient to explain the inhibitory activity of Pn-AMPs and other hevein-like peptides against non-chitin-containing fungi.

Table 1.4. Summary of anti-fungal activity (IC_{50}) of hevein-like peptides.

Fungi	IC_{50} (µg/ml)							
Peptide	Ac-AMP2	IWF-4	Hevein	Pn-AMP1	Fa-AMP1	Ee-CBP	EAFP1	WAMP1a
Alternaria brassicola	4	-	-	-	-	3	-	-
Ascochyta lycopersici	-	-	-	-	-	-	155	-
Ascochyta pisi	8	-	-	-	-	-	-	-
Bipolaris sorokiniana	-	-	-	-	-	-	-	5
Bortrytis cinerea	8	-	500	16	-	1	-	20
Circospora beticola	-	2	-	-	-	-	-	-
Coleotrichum langenarium	-	-	-	10	-	-	-	-
Coleotrichum gossypii	-	-	-	-	-	-	35	-
Fusarium culmorum	2	-	600	-	-	3	-	-
Fusarium oxysporum	-	-	1250	10	19	15	46	5
Fusarium moniliforme	-	-	-	-	-	-	56	-
Fusarium solani	-	-	-	-	-	-	-	5
Geotrichum candidum	-	-	-	-	36	-	-	-
Phycomyces blakeseelanus	-	-	300	-	-	-	-	-
Phytopthora capsici	-	-	-	5	-	-	-	-
Phytopthora parasitica	-	-	-	3	-	-	-	-
Pyricularia oryzae	-	-	500	-	-	-	-	-
Saccharomyces cerevisiae	-	-	-	14	-	-	-	-
Septoria nodorum	-	-	500	-	-	-	-	-
Rhizoctonia solani	-	-	-	26	-	25	-	-
Trichoderma hamatum	3	-	90	-	-	100	-	-
Verticulum dahilae	8	-	-	-	-	-	-	-

Ac-AMP2 (36), IWF-4 (37), hevein (29), Pn-AMP1 (38), Fa-AMP1 (64), Ee-CBP (42), EAFP-1(39) and WAMP-1a (41). The IC50 is represented in µg/ml and '-' indicates that the peptide has not been tested against the fungal strains.

Koo *et al.,* proposed the second mechanism of action based on confocal studies which showed that Pn-AMPs penetrate through the fungal cell walls at growing tips of hyphae and septa within 15 minutes of incubation and result in breakage of the cytoskeleton and subsequent release of cytoplasmic material [47]. This ability to penetrate through the fungal cell wall could be a consequence of their highly basic pI (12.02) and is a suitable explanation for the broad-spectrum anti-fungal activity of highly basic hevein-like peptides. In a recent study by Slavokhotova *et al*., on WAMPs, a novel mode of action involving inhibition of fungal metalloproteases was reported. Due to sequence homology of WAMPs with class I/IV chitinases, their susceptibility to cleavage at the Gly-Cys site of the chitin-binding domain by the fungal metalloprotease FV-cmp was tested. Surprisingly, it was found that WAMPs were not only stable to degradation by the enzyme but could also inhibit its activity against chitinases. This inhibitory activity was attributed to the presence of a Ser residue between the Gly-Cys cleavage site in WAMPs that could prevent recognition by the enzyme. It was proposed that WAMPs bind to the fungal metalloprotease, thus preventing its binding to the plant chitinase which can then degrade the fungal cell wall and inhibit fungal growth [101].

Although the mechanism of action against chitin-free fungi is not known, the anti-bacterial activity of hevein-like peptides could be attributed to the composition of the cell wall of gram-positive bacteria. The peptidoglycan layer accounts for approximately 90% of the dry weight of gram-positive bacteria. The composition of peptidoglycan is similar to chitin and comprises repeating units of N-

acetylglucosamine and N-acetylmuraminic acid linked by β-(1-4)-glycosidic bonds interspersed with oligopeptide chains. It has been speculated that the peptidoglycan layer may assist in the docking of anti-microbial peptides on the bacterial cell surface subsequently causing membrane permeabalization [102]. The chitin-binding domain of hevein-like peptides can bind to N-acetylglucosamine of the peptidoglycan layer leading to accumulation of the peptide on the surface subsequently causing loss of membrane integrity and cell death. Together, this data shows that hevein-like peptides have diverse modes of action against microbes based on their sequence, net charge and cell-wall composition of the pathogens. More studies need to be done to better elucidate the mechanism involved in hevein-like peptide defense against fungi to develop potent anti-fungal agents.

3. *Alternanthera sessilis* and *Moringa oleifera*

A program to discover novel CRPs, with a focus on hevein-like peptides was initiated in our laboratory. During our screening process we came across two herbs, *Alternanthera sessilis* and *Moringa oleifera* which have a myriad of health benefits and exhibited an abundance of hevein-like peptides (Fig. 1.9).

A. sessilis is a perennial herb belonging to the Amaranthaceae family and is a rich source of vitamin A, iron, dietary fibre [103], β-carotene [104], palmitic, myristic and linoleic acid [105]. *A. sessilis* has also been used in folk medicine to treat a myriad of ailments. In Ayurveda, *A. sessilis* is used to treat itchy and inflamed skin [106] and in local Taiwanese medicine it is used to treat hepatitis, bronchitis and asthma. A boiled decoction of the leaves and shoots are used as an anti-hypertensive remedy [107]. Kumar *et al.* reported that the aqueous and ethanolic extracts of *A. sessilis* have anti-diabetic effects [108]. Subhashini T *et al.* clinically evaluated the anti-inflammatory effects of the chloroform and petroleum leaf extracts of *A. sessilis* and showed that the chloroform extract at 200 mg/kg was more effective than the petroleum extract [109]. This combination of dietary and medicinal benefits is the reason why *A. sessilis* is an important herb in Indian folk medicine, where ‘Food as Medicine’ is the major principle [110].

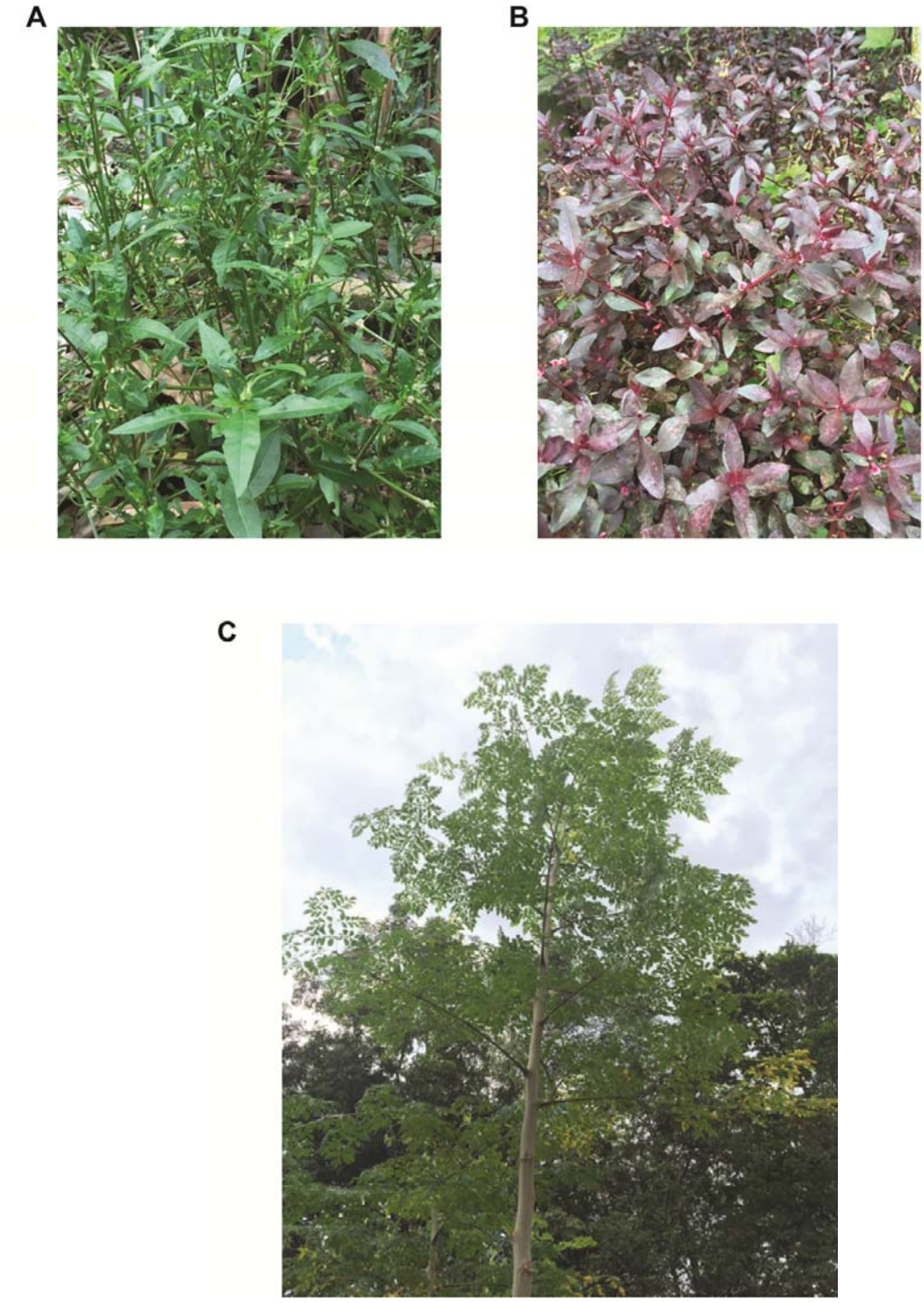


Figure 1.9. Aerial parts of plants selected for the study. (A) *Alternanthera sessilis* var. green (B) *Alternanthera sessilis* var. red (C) *Moringa oleifera.*

Moringa oleifera belongs to the flowering plant family Moringaceae and is commonly known as the drumstick tree or the horse-radish tree [111, 112]. It was once native to the foothills of the Himalayas but is now cultivated globally especially in tropical and sub-tropical areas [113, 114]. *M. oleifera* leaves are a rich source of vitamin C, calcium, potassium, β-carotene and natural antioxidants [111]. The stem-bark, flowers, seeds and roots are edible and nutritious, comprising compounds such as alkaloids, potassium, calcium, ascorbic acid and antioxidants [115, 116]. Since *M. oleifera* is rich in nutrients it is commonly referred to as the "Miracle Tree" to combat malnutrition around the world [117]. Aqueous and ethanol extracts of *M. oleifera* leaves are known to be anti-inflammatory, antimicrobial, anti-diabetic, anti-ulcer and have cholesterol lowering and blood pressure stabilizing effects [118].

The medicinal and nutritional benefits of *A. sessilis* and *M. oleifera* have been attributed to small-molecules thus far. Since the plant extracts were boiled before studying their medicinal properties, we speculate that the bioactive principles in these herbs could be heat-stable proteins or peptides like CRPs that are in high abundance.

4. Hypothesis and Objectives of this study

Hevein-like peptides are a highly diverse class of chitin-binding peptides that play an integral role in plant defense against microbes and insect pathogens. Therefore, insights into their sequence diversity, structure, function and biosynthesis will be beneficial for the development of peptide-based anti-fungal agents.

The aims of my thesis are:

1. To isolate and characterize novel 6C-hevein-like peptides from two varieties of *Alternanthera sessilis*
2. To isolate and characterize novel 8C-hevein-like peptides from *Moringa oleifera*
3. To study the distribution and diversity of hevein-like peptides in plants using EST-based bio-informatics approaches

This work describes the discovery of novel 6C- and 8C-hevein-like peptides from *A. sessilis* and *M. oleifera*, respectively. Genomic and transcriptomic analysis showed that the precursor organization is conserved among hevein-like peptides. It was shown that the binding affinity of 6C-hevein-like peptides to chitin oligosaccharides is comparable to hevein and aromatic residues in the chitin-binding domain are involved in binding. Both 6C- and 8C-hevein-like peptides could inhibit the growth of chitin-containing phytopathogenic fungi. Knowledge of the bioactivity and gene organization may be useful in the development of transgenic plants with enhanced resistance against fungi.

An insight into the distribution of hevein-like peptides in plants was obtained using EST-based data-mining in NCBI and OneKp databases. Interestingly, a total of 385 novel hevein-like peptides were found in 124 different plant families. These peptides were expressed in mosses, the first terrestrial plants on Earth and their expression has been retained in highly evolved angiosperms till date. This data shows that hevein-like peptides could be the largest family of plant CRPs and are widespread across the plant kingdom due to their integral role in defense against fungal pathogens.

Chapter 2

Materials and Methods

1. Materials

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

1.2. Kits

The kits for molecular cloning were purchased from Invitrogen (Life Technologies, USA), Qiagen (USA), Clontech (Takara Bio, Japan), and Promega (USA).

1.3. Plant materials

Alternanthera sessilis (green) and *Alternanthera sessilis* (red) were obtained from Singapore botanic gardens and grown in Herb garden of School of Biological Sciences, NTU, Singapore. *Moringa oleifera* was purchased from local markets in Singapore.

1.4. Fungal strains

Seven phyto-pathogenic fungal strains were acquired from China Center of Industrial Culture Collection, including *Curvularia lunata* (CICC 40301), *Fusarium oxysporum* (CICC 2532), *Aspergillus niger* (CICC 2089), *Verticillium dahliae* (CICC 2534), *Rhizoctonia solani* (CICC 40259), *Alternaria alternata* (CICC 2465) and *Alternaria brassiciola* (CICC 2646).

2. Genomics

2.1. 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 minutes to allow complete dissociation of the nucleoprotein complex. Chloroform (0.2 ml) was added sample was incubated for 3 minutes followed by centrifugation at 12,000 rpm for 15 minutes. 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).

2.2. DNA extraction

Samples were homogenized in liquid nitrogen. CTAB buffer (500 µl) was added to 50-60 mg of homogenized sample and incubated in a water bath at 55°C for 1 h with intermittent shaking. After incubation, chloroform (500 µl) was added to the samples followed by centrifugation at 16000 x *g* for 7 minutes. The aqueous layer was collected and 0.08 volumes of cold 7.5 M ammonium acetate and 0.54 volumes of cold isopropanol were added and incubated on ice for 30 minutes. The sample was centrifuged at 16000g for 3 minutes and the pellet was washed twice with 75% ethanol. After the pellet was dried it was resuspended in DEPC water (30 µl) and DNA concentration and integrity were determined using NanoPhotometer™ (Implen, Northstar Scientific, Germany) and 1.0% agarose gel electrophoresis.

2.3. Rapid amplification of cDNA ends (RACE) and PCR analysis

Total RNA extract was used as a template to generate the 3´- cDNA library using 3´RACE (rapid amplification of cDNA ends) system (Invitrogen) according to the manufacturer's protocol. Degenerate primers were designed using GeneRunner software and were used with universal adaptor primers (UAP/AUAP) to amplify the desired sequence using a T100 thermal cycler (BioRad). The PCR products were run on a 1% agarose gel and the target bands were excised and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and cloned with pGEM®-T Easy Vector System (Promega) using JM109 high-efficiency competent cells. Sequencing of insert containing plasmids was performed using 1st Base Company's service. Degenerate primers targeting regions PQCNHG (5´–CCTGGTCArTGyAAyCAyGG-3´) and QGYCGTG (5´-CAAGGTTATTGyGGnACnGG-3´) were used in the 3´-RACE polymerase chain reaction (PCR). Specific primers 5´-AAACACATGCCCAACAAAGAG-3´ and 5´-AATAATTATAACATACAGTGTTAT-3´ were designed against the 3´UTR region of the partial sequences obtained from 3´RACE and primed against the 5´cDNA obtained using SMARTer™ RACE cDNA Amplification Kit (Clontech). The complete protocol from PCR to sequencing was performed to obtain complete sequences of the transcripts of interest.

To determine presence and location of introns, specific primers were designed from obtained cDNA sequences and primed against a DNA template. Forward primers were designed from the 5´UTR or ER signal peptide and reverse primers from 3´UTR. The PCR set up was as follows: initial denaturation at 94°C for 5

min, main amplification for 35 cycles (denaturation at 94°C for 30 sec, annealing at Tm - 3°C for 30 sec, elongation at 72°C for 45 – 60 sec), final elongation at 72°C for 10 min, and cooling to 4°C. PCR products were tested on a 1.5% agarose electrophoresis, purified, cloned into pGEM-T Easy vector and sequenced as described.

2.4. 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 endoplasmic reticulum (ER) signal sequences of the precursor sequences was performed using SignalP 4.0 prediction server (http://www.cbs.dtu.dk/services/SignalP/) [119, 120]. Logo sequences were built with WebLogo application, version 3 (http://weblogo.berkeley.edu/) [121].

3. Proteomics

3.1. 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 (250 x 9.4 mm and 250 x 4.6 mm) 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.

3.2. MALDI-TOF MS and MS/MS

Mass spectrometry analysis was performed using the 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.

3.3. Protein extraction and purification

Fresh leaves (5-10 kg) of *Alternanthera sessilis* and *Moringa oleifera* were obtained from the Nanyang Technological University (Singapore) herb garden and 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).

3.4. 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) 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.

3.5 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.

3.6. 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 [122]:

$\varepsilon_{280} = (5500. n_{Trp}) + (1490 . n_{Tyr}) + (125 . n_{SS})$

where n_{Trp} : the number of Trp residues

4. Structural analysis

4.1. 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% NaN3 in D_2O or 10% D_2O. Nuclear Overhauser effect spectroscopy (NOESY) experiments were acquired with 200 and 300 ms mixing times [123, 124]. Total correlation spectroscopy (COSY) [125] data were recorded with a mixing time of 69 or 78 ms using MLEV17 spin lock pulses [126]. Vicinal coupling constants were measured using the double-quantum-filtered (DQF)-COSY [127] 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 [128], 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 [129] or excitation sculpting [130] combined with pulsed-field gradients. All NMR data were processed using Bruker TOPSPIN 2.1 or NMRPipe [131] 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.

4.2. Structure calculations

Solution structures of aSG1 were calculated by hybrid distance geometry and a simulated annealing protocol in torsion angle space with CNS 1.2 [132]. The three disulfide bonds were restrained in accordance with the disulfide bonding patterns based on the observed Hβ-Hβ NOEs from NOESY spectra by generation of covalent disulfide linkages during the initial molecular topology file generation stage using a CNS script. A total of 597 distance and 18 torsion angle constraints were used for structure calculations. NOE distance restraints were classified strong (1.8-3.0 Å), medium (1.8-3.5 Å), weak (1.8-5.0 Å), or very weak (1.8-6.0 Å) based on intensities derived from NOESY spectra recorded in 10% D_2O or 100% D_2O. To validate the disulfide bond connectivities, structure calculations & refinements were performed for 15 different disulfide bond patterns. Molecular topology files for 15 different disulfide bond connectivities were generated using molecular topology file generating script of CNS. 200 structures were calculated by distance geometry regularization and simulated annealing protocol out of which the lowest 18-20 structures were selected for each disulfide bond combination and average total energies were compared. Corrections for pseudo-atom representations were used for non-stereospecifically assigned methylene, methyl group, and aromatic ring protons

[128]. Backbone dihedral angle restraints were derived from 3JHNα coupling constants in DQF-COSY or 1D-^{1}H-NMR spectra in a H_2O solution [133, 134]. Backbone dihedral restraints were -55° ± 45° (3JHNα < 6 Hz) and -120° ± 50° (3JHNα > 8 Hz). Hydrogen bond donors were identified from proton-deuterium exchange experiments, and hydrogen bond acceptors were determined in the preliminary structure calculation stage. The 200 starting structures were generated and refined using a hybrid distance geometry-simulated annealing protocol [135-137] in the CNS 1.2 program. Finally, 18 final structures were selected by their total energy values for display and structural analysis. MOLMOL [138] and PyMOL [139] programs were used for structure visualization and PROCHECK-NMR [140] were used for structure validation.

4.3. NMR titration experiments

The binding affinity of aSG1 and aSR1 towards N, N´-diacetyl chitobiose $(GlcNAc)_2$ and N, N´, N´´-triacetyl chitotriose $(GlcNAc)_3$ was determined by monitoring chemical shifts in ^{1}H-NMR spectra from a series of titrations. Samples contained 0.6 mM peptide in 20 mM sodium phosphate buffer (pH 7.4) in D_2O, using 100 μM of DSS-d6 as internal reference. Increasing concentrations of chitobiose (0.036-50.8 mM), chitotriose (0.1-4.8 mM), and chitopentose (0.1-4.8 mM) were then titrated while keeping the peptide concentration constant. Differences in chemical shifts of characteristic protons were plotted against sugar concentrations and fitted using Origin 9.0 according to equation (1) [141]:

$$\Delta\delta obs = \Delta\delta max \frac{\langle [P]_t + [L]_t + K_D - \{([P]_t + [L]_t + K_D)^2 - 4[P]_t[L]_t\}^{1/2} \rangle}{2[P]_t} \quad (1)$$

where Δδobs is the change in the observed shift from the free state, Δδmax is the maximum shift change upon saturation, $(P)_t$ is the total concentration of the peptide, and $(L)_t$ is the total concentration of the sugar [141].

5. Stability Assays

5.1. Heat stability assay

Purified peptides were incubated at 100°C for 1 h and then analyzed by UPLC. Short linear peptides RLYRRGRLYRRNHV (RV-14) and WV-14 (WRLYRGRLYRRNHV) synthesized in our laboratory served as controls and DALK, a thermostable peptide was used as an internal standard. Peaks from UPLC were analyzed by MALDI-TOF MS.

5.2. Proteolytic enzyme stability assay

Purified peptides were incubated with trypsin, pepsin and chymotrypsin 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 and WV-14 were used as controls.

5.3. Serum stability assay

Stability of isolated and purified peptides in human serum was assessed according to the protocol by Jenssen *et al.* [142]. 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 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 WV-14 and RV-14 were used as controls.

6. Bioassays

6.1. Disc diffusion assay

Susceptibility of seven fungal strains, *Alternaria alternata*, *Alterneria brassiciola*, *Aspergillus niger*, *Curvularia lunata*, *Fusarium oxysporum*, *Rhizoctonia solani and Verticulum dahilae* to purified hevein-like peptides from *A.sessilis* and *M.oleifera* were tested using the disc diffusion assay as described by Ye *et al.* [143]. *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 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.

6.2. Microbroth dilution assay

The half maximal inhibitory concentration (IC_{50}) of peptides was assessed using the microbroth dilution assay [144]. 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.

6.3. Cytotoxicity assay

PrestoBlue™ Cell Viability Reagent (Invitrogen) was used to test the cytotoxicity of purified peptides. Vero (African green monkey kidney 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.

6.4. Hemolysis assay

Fresh type A+ blood was donated by a student volunteer. Erythrocytes from the blood (1 ml) were collected by centrifuging at 700 rpm for 15 min. The pellet was then washed with PBS (1ml) five times and centrifuged at 1000 rpm for 5 min. The centrifuged erythrocytes were diluted 100X in PBS to obtain the stock dispersion. Peptide samples were prepared in PBS and added to the stock dispersion in a 1:1 ratio in a 96-well plate and incubated for 4 h at 37°C with gentle shaking. After incubation, the plate was centrifuged at 1000 rpm for 5 min and the supernatants were transferred to a new plate. Absorbance was read at 415 nm in an Infinite@ 200 PRO Tecan microplate reader. Equal volume of 8% Triton-X100 and PBS were used as positive and negative controls, respectively.

7. EST-Based data mining

7.1. Translated nucleotide based search for putative hevein-like peptides

Database searches were conducted using methods modified from literature [145, 146]. tBLASTn was employed to search for ESTs encoding putative hevein-like peptide precursors in two databases, the National Center for Biotechnology Information (NCBI) [147] and the 1000 Plants Project (OneKP), via queries using all known hevein-like peptide sequences [148]. The retrieved sequences were then used as queries in subsequent tBLASTn searches until no novel sequences

were obtained. The maximum target sequence and expected threshold were set at 1000 and 10, respectively. The accession numbers of all the putative hevein-like peptide sequences are provided in the appendix E. Search results were exported to Microsoft Excel, and sequences with coverage lower than 80% were deleted. The remaining 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 [149] for identification of signal peptides and were aligned using ClustalW [150]. Replicate sequences from same plants or different species of the same genus sharing an identical full-length precursor were identified using ClustalW Phylogeny and deleted from the dataset. The elimination of Class I chitinases performed by running an InterPro scan using the BLAST2GO PRO software [151].

7.2. Data analysis

Statistical analyses were performed using GraphPad Prism software version 6.01 (GraphPad Software, CA, USA). The 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 [121]. The phylogenetic tree was constructed and annotated using the online tool called, Interactive tree of life (iTOL) [152].

Chapter 3

Isolation and Characterization of Novel 6C-Hevein-like Peptides from *Alternanthera sessilis*

1. Introduction

The Amaranthaceae family belongs to the order Caryophyllales of flowering plants and is the one of the most species-rich lineage with 180 species and 2500 genera [153]. This family comprises annual or perennial herbs and shrubs and very rarely vines and trees that grow in tropical climate. Plants of this family are rich in secondary metabolites like isofalvonoids, saponins, triterpenoids and betalain pigments [154]. Some Amaranthaceous plants like *Dysphania ambrosioides and Dysphania anthelmintica are* used as medicinal herbs. Species like *Amaranthus* and *Celosia* are commonly found in gardens as ornamental plants.

Here we studied two Amaranthaceae species, *Alternanthera sessilis* var. green and *Alternanthera sessilis* var. red which are perennial herbs commonly found in tropical and sub-topical areas and make up the staple diet in India and Africa [155]. The herbs comprise prostrate stems that grow upto 0.4 to 1.4 m high and often root at the nodes with broadly elliptical leaves that have petioles about 1-5 mm long petioles [156]. As its flowers are shaped like the eyes of a fish it is commonly called Matsyakashi (fish-eyed). Although the medicinal benefits of *A. sessilis* have been documented, most of the biological activities are attributed to the secondary metabolites in the plant extract. A larger chemical space

comprising peptide biologics has not been studied thus far. Previously, closely related plant species like *Amaranthus caudatus* and *Amaranthus retroflexus* have been reported to comprise 3-4 kDa 6C-hevein-like peptides. These peptides could bind to chitin and inhibit the growth of chitin-containing fungi [45, 50]. Thus, we examined if *A. sessilis* also expresses similar hevein-like peptides and studied their role in plant defense.

This chapter reports the isolation and characterization of six novel 2-4 kDa 6C-hevein-like peptides from red and green varieties of *Alternanthera sessilis*. These newly identified peptides were collectively named altides in accordance with the plant's genus name. In order to distinguish individual altides, they were named using the initial alphabets of the genus and species name followed by the variety and a number indicating the order of discovery. For example, aSG1 and aSR1 are altides from green and red varieties of *A. sessilis*, respectively. ^{1}H-NMR titration experiments with chitin oligosaccharides provide insight into their binding interaction with chitin. The anti-fungal, cytotoxic and haemolytic activities of altides have also been studied. Gene cloning of the peptide precursors has helped shed light on their biosynthesis. Taken together this work provides insight into the sequence, structure, bioactivity and biosynthesis of altides.

2. Results

2.1. Isolation and sequencing of 6C-hevein-like peptides from *A. sessilis*

Small-scale screening of red and green varieties of *A. sessilis* using MALDI-TOF MS revealed a cluster of peptides between 3 and 4 kDa indicating the presence of CRPs (Fig. 3.1). A large-scale extraction of 5-10 kg of *A. sessilis* leaves was performed to isolate the putative CRPs. The plant material was homogenized in equal volume (v/v) of water and centrifuged to remove plant debris. The filtered supernatant was loaded onto a C_{18} flash column, which was washed repeatedly with 5% ethanol to partially remove colored pigments and hydrophilic compounds. Elution of bound peptides was performed using 40%-70% ethanol, which were then purified by several rounds of SCX- and RP-HPLC.

Isolated and purified peptides were subjected to disulfide reduction (S-reduction) using dithiothreitol (DTT) to free the cysteine residues followed by alkylation of the free cysteines and subsequent digestion with trypsin, chymotrypsin and endo-GluC. The resulting peptide fragments were sequenced by tandem MS (Fig. 3.2). For example, S-reduced aSG1 had an m/z value of 3010, an increase of +6 Da, indicating the presence of six cysteine residues.

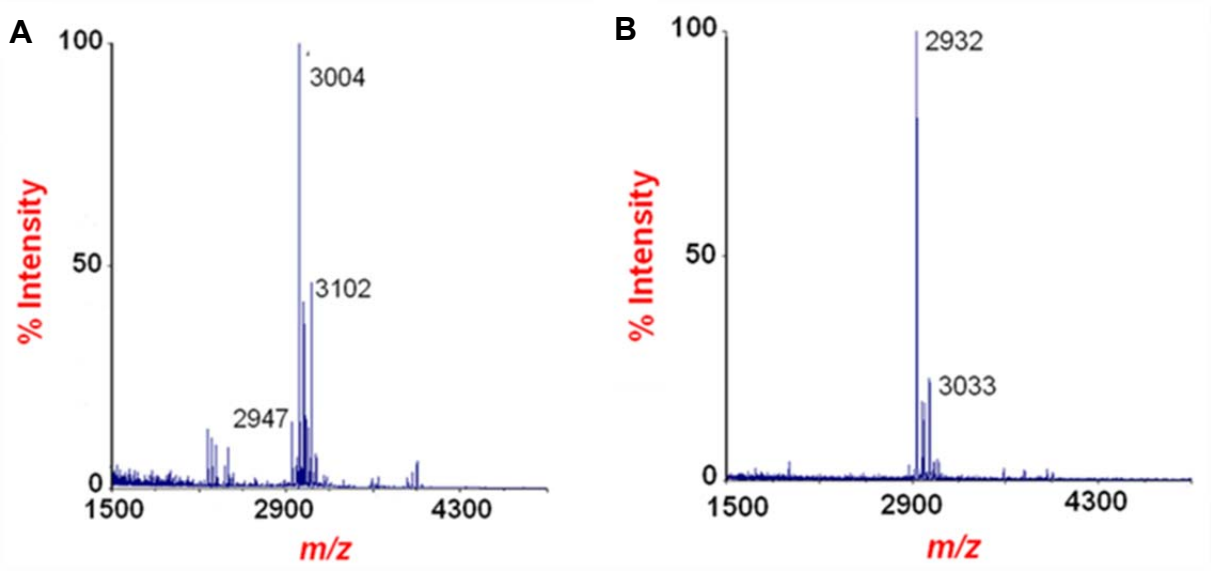


Figure 3.1. MALDI-TOF spectra of crude extracts. (A) *A.sessilis* var. green and (B) *A.sessilis* var. red. Clusters of peaks in the range of 2-3 kDa indicate the presence of putative CRPs.

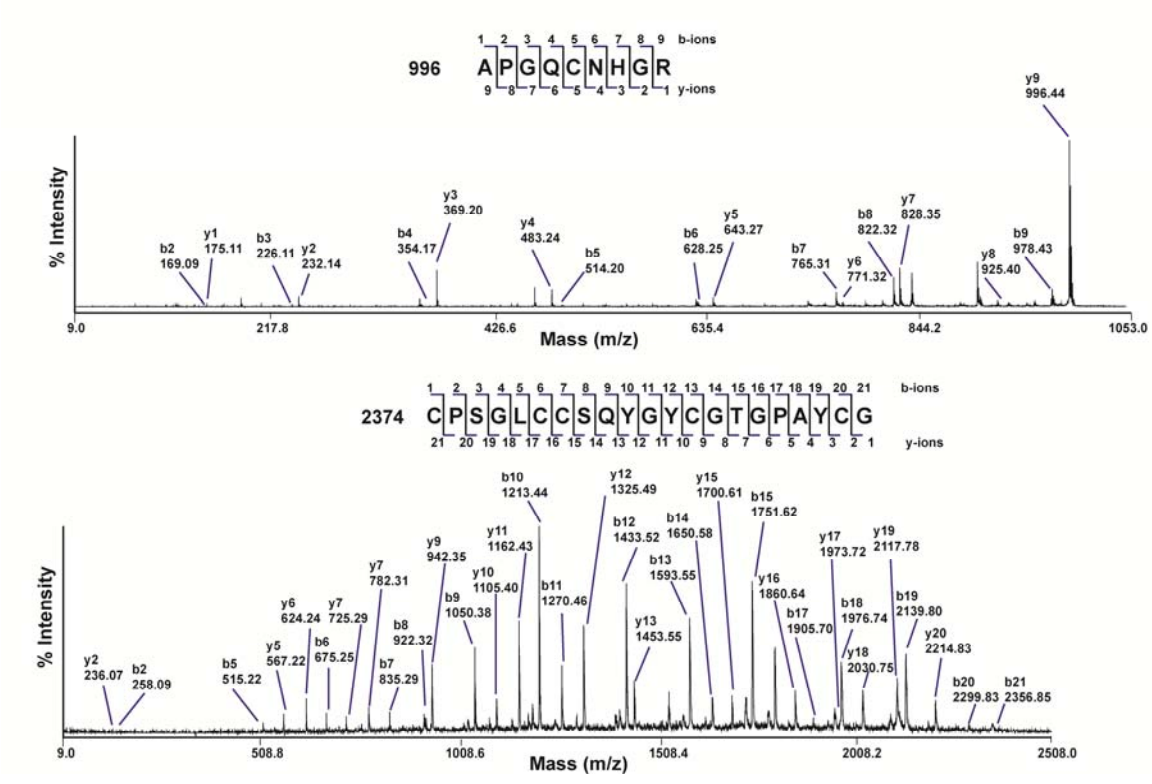


Figure 3.2. *De novo* sequencing of aSG1. Tandem MALDI-TOF/TOF MS/MS profiles of two tryptic fragments of aSG1 (m/z 996 and m/z 2374) provide the full aSG1 sequence.

One of the limitations of tandem MS sequencing is the difficulty in differentiating isobaric residues like Leu/Ile and Lys/Gln. Thus, assigned sequences were validated using gene cloning studies where genetic sequence of the precursors helps to unambiguously obtain the complete peptide sequence with the added advantage of providing information about the biosynthesis pathway of the peptide.

A total of six novel peptides, collectively named altides were isolated and sequenced. Table 3.1 summarizes the primary sequences of altides aSG1-aSG3 and aSR1-aSR3 from green and red varieties of *A. sessilis* respectively and previously reported 6C-hevein-like peptides. Altides are 29-31 amino acid residues in length and are rich in cysteine (6 residues), glycine (7 residues) and proline (2-3 residues) accounting for ~60% of the peptide sequence. Altide, aSR1 (29 amino acid residues) is the smallest hevein-like peptide isolated thus far. BLAST analysis revealed that altides belong to a family of hevein-like peptides with a single chitin-binding domain and shared ~80% homology with Ac-AMP2 isolated from *Amaranthus caudatus* [45]. The cysteine-spacing pattern X_{3-4}**C**-X_4**C**-X_4**CC**-X_5**C**-X_6**C**-X_{1-2} is highly conserved among altides. Apart from the conserved cysteines, all seven glycine residues, Arg-9, Pro-13, Ser-17, Tyr-21 and Tyr-28 (numbered according to aSG1) are absolutely conserved throughout the 6C-hevein-like peptide family. In altides isolated from the same variety, the sequence variability of the loops is minimal with difference in only 1 or 2 residues. Sequence comparison of altides from different varieties shows that loop 1 is the most variable while loop 2 is partially conserved and varies in only 1 or 2 residues.

Table 3.1. Consensus sequences of altides aligned with previously reported 6C-hevein-like peptides.

Peptide/ Loop	Peptide Sequence
	1 2 3 4
	I II III IV V VI
aSG1	APGQC-NHGRCPSGLCCSQYGYCGTGPAYCG-
aSG2	-AGEC-NHGRCPSGLCCSQYGYCGTGPRYCG-
aSG3	APGQC-NHGRCPSGLCCSQYGYCGTGPAYCGG
aSR1	-VGEC-VQGRCPPGLCCSRFGYCGTGPAYCG-
aSR2	APGEC-KHGRCPPGICCSQYGYCGTGPAYCG-
aSR3	APGEC-KHGRCPPGICCSQYGYCGTGPAYC--
Ac-AMP1	-VGEC-VRGRCPSGMCCSQFGYCGKGPKYCG-
Ac-AMP2	-VGEC-VRGRCPSGMCCSQFGYCGKGPKYCGR
Ar-AMP	-AGEC-VQGRCPSGMCCSQFGYCGRGPKYCGR
IWF-4	-SGECNMYGRCPPGYCCSKFGYCGVGRAYCG-

The conserved chitin-binding domain is indicated with an asterisk and the cystine-knot motif is depicted at the bottom of the table.

Loop 3 comprises three residues of the chitin-binding domain and is highly conserved in all altides except aSR1 where Gln-18 and Tyr-19 are replaced by Arg and Phe respectively. Interestingly, loop4 which immediately follows the chitin-binding domain is highly conserved in all altides except aSG2 where Ala27 is replaced by Arg. Overall, the sequence of aSR1 is the most variable as it lacks the N-terminal Ala residue and differs from other altides in 6 positions - Pro-2, Asn-6, His-7, Ser-12, Gln-18 and Tyr-19 are replaced by Val-2, Val-6, Gln-7, Pro-12, Arg-18 and Phe-19 respectively. This data shows that despite being from two distinct varieties, altides show some degree of sequence conservation indicating that they may have evolved from a common ancestor.

2.2. Gene cloning of altides

To obtain the full length cDNA sequence of altide precursors, 3'-RACE PCR was performed by designing a degenerate primer targeting the region QGYCGTG. After cloning products of desired size, 5'-RACE PCR was performed using primers designed to the 3'UTR to obtain the full precursor sequence. Two full-length gene sequences encoding for aSG1 and aSG2 were obtained and designated as *asg1* and *asg2* respectively (Fig. 3.3). To study the location of introns, genomic DNA was used as a template in the PCR reaction with specific primers targeting the untranslated region. Sequence comparison of DNA and RACE PCRs showed that no introns are located in the gene.

The deduced amino acid sequences of the full-length precursors of aSG1 and aSG2 are shown in Figure 3.4 A. The 81-amino acid long precursors contain three domains, a 25-amino acid long ER signal peptide followed by the 30-amino

acid long mature CRP domain and a short 25-amino acid long C-terminal tail. The signal peptide and C-tail sequences are highly conserved among aSG1 and aSG2 and differ by only one amino acid residue; Gly-55 (numbered according to aSG1) is replaced by Ser in aSG2. The precursor of aSR1 shares 70% sequence homology with altides from the green variety with variations at seven positions in the signal peptide, six positions in the mature domain and a highly variable C-tail. The presence of an ER signal sequence indicates that altides are ribosomally synthesized and bioprocessing follows the secretory pathway (Fig. 3.4 B).

asg1

```
5´ untranslated region
actacttaacagcggagttatgactttgtttctaaaata
agctctttgttgggcatgtgtttctttatgttgtatcttggctttcttatcgattggatg
gagaattaacgtagttgttatttcttctgttatccaatcatcatgttgtagtactcgtt

atgatgaacatgaagaagtttttgatcgttatggtagtggtagcc
 M  M  N  M  K  K  F  L  I▽V  M  V  V  V  A
ctagtgatggtggagccatcaatgggagcaccaggtcaatgcaaccatggacgttgccct
 L  V  M  V  E  P  S  M  G  A  P  G  Q  C  N  H  G  R  C  P
agtgggttgtgctgcagccagtatggttactgtggcactggccccgcttattgtggtggc
 S  G  L  C  C  S  Q  Y  G  Y  C  G  T  G  P  A  Y  C  G  G
gctgccgagcaacgcgctgctcttcttcagcgcaccggtagtgtcaccgcagacactact
 A  A  E  Q  R  A  A  L  L  Q  R  T  G  S  V  T  A  D  T  T
gacactactaaagctccatga
 D  T  T  K  A  P  *


3´ untranslated region
agtgttcaagctagaagataagaaaaaaaaaa
```

asg2

```
actacttaacagcggagttatgactttgtttctaaaataagctct


atgatgaacatgaagaagtttttgatccttatggta
 M  M  N  M  K  K  F  L  I  L  M  V▽
gtagtagccctagtgatggtggagccatcaatgggagcaggcgagtgcaaccatggacgg
 V  V  A  L  V  M  V  E  P  S  M  G  A  G  E  C  N  H  G  R
tgccccagtgggttgtgctgcagccagtatggttactgtggcactggcccccgttattgt
 C  P  S  G  L  C  C  S  Q  Y  G  Y  C  G  T  G  P  R  Y  C
ggtagcgctgccgagcaacgcgctgctcttcttcagcgcaccggtagtgtcaccgcagac
 G  S  A  A  E  Q  R  A  A  L  L  Q  R  T  G  S  V  T  A  D
actactgacactactaaagctccatga
 T  T  D  T  T  K  A  P  *


3´ untranslated region
tagagtgttcaagctagaagataagaaaaaaaaacaaaaaaaaa
```

Figure 3.3. Gene sequences of aSG1 and aSG2. The sequence of the signal peptide is in bold and an inverted triangle depicts the cleavage site of signal peptidase as predicted by Signal P 4.0 server. The mature domain is highlighted in blue and the conserved cysteine residues in yellow. From the sequence it is clear that no intron is present in the gene expressing aSG1. An asterisk is present below the stop codon.

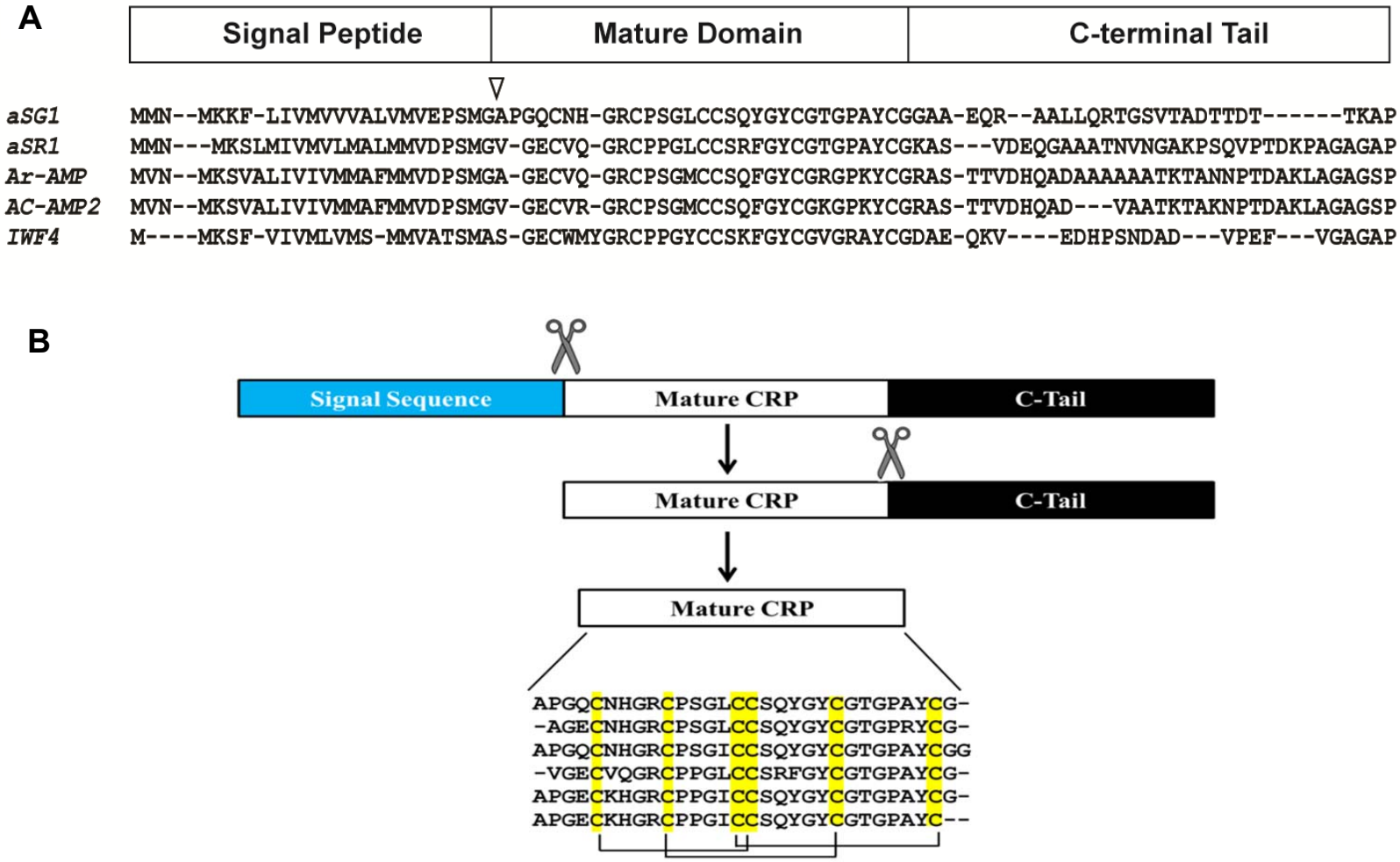


Figure 3.4. Biosynthesis of altides. (A) Aligned full length precursors of altides with Ar-AMP, Ac-AMP2 and IWF4. The precursor organization is conserved among 6C-hevein-like peptides and comprises a signal peptide, mature domain and C-terminal tail. The cleavage site of SPase is indicated with an inverted triangle. (B) Bioprocessing of altides follows the secretory pathway.

2.3. Structural analysis of aSG1

The spin-spin systems of aSG1 and approximately 98% of proton resonances were assigned unambiguously. The solution structure of aSG1 was calculated from a total of 597-NMR derived distance restraints and 18 dihedral angle restraints (Fig. 3.5 B). Root-mean square deviation values relative to the mean of 18 conformers for residues Gly3-Gly 30 were 0.28 Å for backbone atoms and 1.11 Å for heavy atoms respectively. Amide proton exchange experiments were performed to determine the sequential dαN(*i,i*+1) NOE connectivity, $^{3}J_{HN\alpha}$ values and hydrogen bond patterns. The secondary structure of aSG1 was found to comprise two extended β-sheets and tight turns. The C-terminal region (Pro-26-Gly-30) also comprised several strong d_{NN}(i,i+1) and medium or weak $d_{\alpha N}$(*i,i*+3), d_{NN}(*i,i*+2), and d_{NN}(*i,i*+3) NOEs indicating the presence of α-helical structure in this region. Thus, the structure of aSG1 comprise two short, anti-parallel β-sheets (β1, Cys-15-Cys-16, and β2, Gly-20−Gly-24) and one α-helical segment with several tight turns and loops with a well-defined structure based on medium and long-range NOEs. The disulfide connectivity of aSG1 was determined by calculating the structural energies for all 15 possible disulfide bond patterns. While most disulfide patterns showed high energy values from 200 to 760 kcal/mol, three patterns namely, Cys I−Cys IV/Cys II−Cys V/Cys III−Cys VI, Cys I−Cys V/Cys II−Cys IV/Cys III−Cys VI, and Cys I−Cys II/Cys III−Cys VI/Cys IV−Cys V showed low total energy values (30−66 kcal/ mol). However, the lowest energy value of 29.90 ± 0.87 kcal/mol was observed in pattern Cys I−Cys IV/Cys II−Cys V/Cys III−Cys VI indicating that this disulfide pattern is the most favorable

amongst all 15 patterns tested and is in good agreement with Hβ−Hβ connectivities observed from the NMR spectrum. PROCHECK analysis indicated that all residues were distributed in the allowed region of the Ramachandran map (Fig. 3.5 A). The overall topology of aSG1 was that of a compact globular molecule and seemed to be stabilized by several strong hydrogen bonds and three disulfide bridges (Fig. 3.5 C).

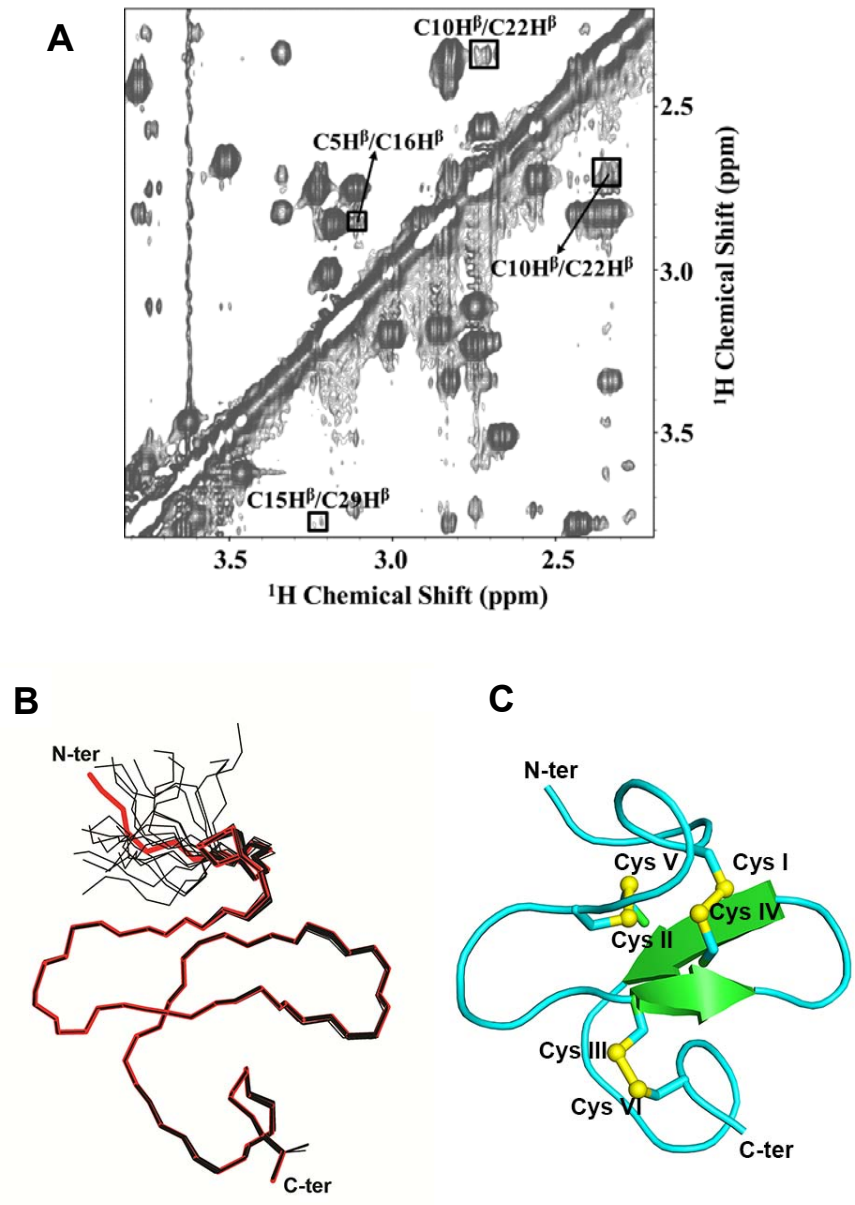


Figure 3.5. NMR structure of aSG1. (A) 2D-NOESY spectra showing the expanded aliphatic regions of aSG1 obtained with a mixing time of 300 ms in D2O at 25 °C. The correlations between cysteines forming disulfide bridges of aSG1 (Cys-5 Hβ−Cys-16 Hβ, Cys-10 Hβ−Cys-22 Hβ, and Cys-15 Hβ−Cys-29 Hβ) are represented by rectangular boxes. (B) Superposition of the backbone traces from the final 18 ensembles of solution structures and restrained energy minimized (REM) structure of aSG1 determined by NMR spectroscopy. The REM structure of aSG1 is colored red. (C) Ribbon representation of aSG1 REM structure. Three disulfide bridges (Cys I−Cys IV, Cys II−Cys V, and Cys III−Cys VI) are shown as balls and sticks.

2.4. Chitin-binding activity

The chitin-binding activity of altides was assessed by incubating purified altides (4 µg) with chitin beads for 4 h at 4°C in chitin binding buffer. UPLC chromatographs showed that altides, aSG1, aSR1 and aSR2 bound to chitin within 1 h (Fig. 3.6). To elute the peptides, two elution methods were tested. The first method involved the use of chitin elution buffers with increasing salt concentration (up to 1 M NaCl) while the second involved heating the peptide-bound chitin beads at 100°C for 30 min and 1 h in the presence of 0.5 M acetic acid [46]. It was observed that none of the altides could be eluted from the beads when upto 1 M NaCl was used while aSG1, aSR2 and aSR1 were eluted in 30 min and 1 h respectively after heating in acidic conditions (Fig. 3.6). These results indicate that altides bind strongly to chitin and require heating and acidic conditions for elution.

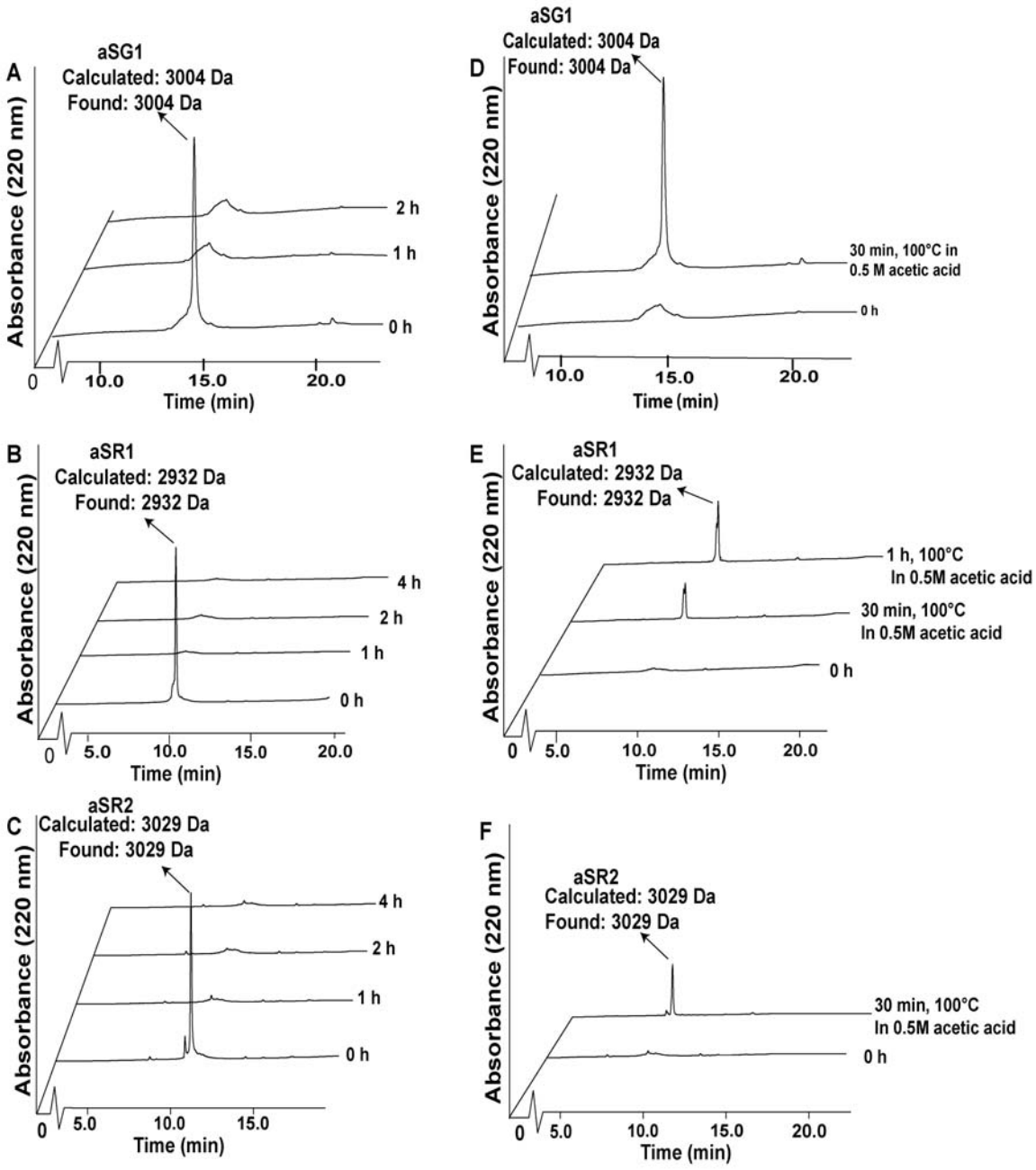


Figure 3.6. Chitin-binding activity of altides. Upon incubation with chitin beads for 1 h at 4°C, altides, (A) aSG1, (B) aSR1 and (C) aSR2 bound within 1 h. Elution of the bound peptides was performed in acidic environment and heating at 100°C. (D) aSG1 and (E) aSR1 eluted in 1 h while (F) aSR2 eluted in 30 min.

2.5. ^{1}H-NMR titration of altides with GlcNAc oligosaccharides

The binding motif and binding constant (K_D) of the interaction between altides and chitin oligosaccharides was deduced by performing ^{1}H-NMR titration experiments with chitobiose $(GlcNAc)_2$ and chitotriose $(GlcNAc)_3$. Keeping the peptide concentration constant (0.6 mM), increasing concentrations of chitobiose (0.036-50.8 mM) and chitotriose (0.1-4.8 mM) were titrated into the sample and ^{1}H-NMR spectra were recorded at each titration point.

^{1}H-NMR spectra of the titration of altides and chitin oligosaccharides showed significant changes in the chemical shifts of a number of peaks indicating the formation of a complex between altides and the oligosaccharides. When aSG1 and aSR1 were titrated with chitobiose and chitotriose, a difference of chemical shifts of the peaks between 7-8 ppm indicated the involvement of aromatic residues in binding (Fig. 3.7). From the assigned spectra of aSG1, it was deduced that the binding motif comprised Tyr-28, Tyr-19, Tyr-21, and Gln-18, which occupy the conserved chitin-binding domain. Similarly, titration of aSR1 with chitotriose showed shifts in peaks between 7-8 ppm indicating that aromatic residues in the chitin-binding region were involved in the binding interaction. The binding motif was deduced as Arg-17, Phe-19, Tyr-21, and Tyr-27 by overlaying the assigned spectra of aSG1 with aSR1 (Fig. 3.8).

The difference of chemical shift values were plotted against sugar concentration and fitted into equation 1 to calculate the K_D of the binding interaction. Interestingly, the K_D of the interaction between aSG1 and chitobiose was 9.6 ± 0.107 mM while K_D of aSR1 and chitotriose was 97 ± 17.8 µM (Fig. 3.8). This difference of two orders of magnitude can be due to the increase in number of binding sites when chitotriose is used.

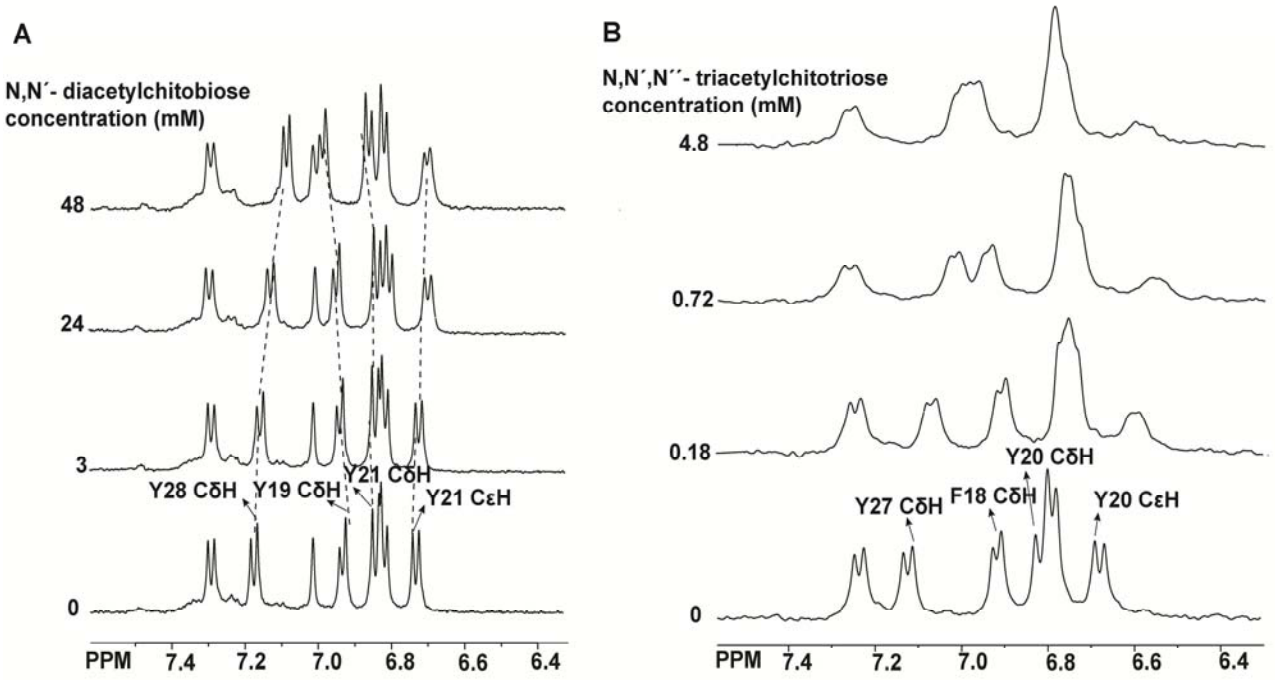


Figure 3.7. ^{1}H-NMR titration study of altides with chitin oligosaccharides. (A) 0.6 mM aSG1 was titrated with increasing concentration of chitobiose at pH 7.4. Variations in protons of tyrosine residues were followed. Sugar concentration is indicated on the left. (B) Increasing concentration of chitotriose was titrated into 0.6 mM aSR1. Dramatic changes in chemical shift and line broadening of peaks corresponding to aromatic residues were observed. Chitotriose concentration is indicated on the left.

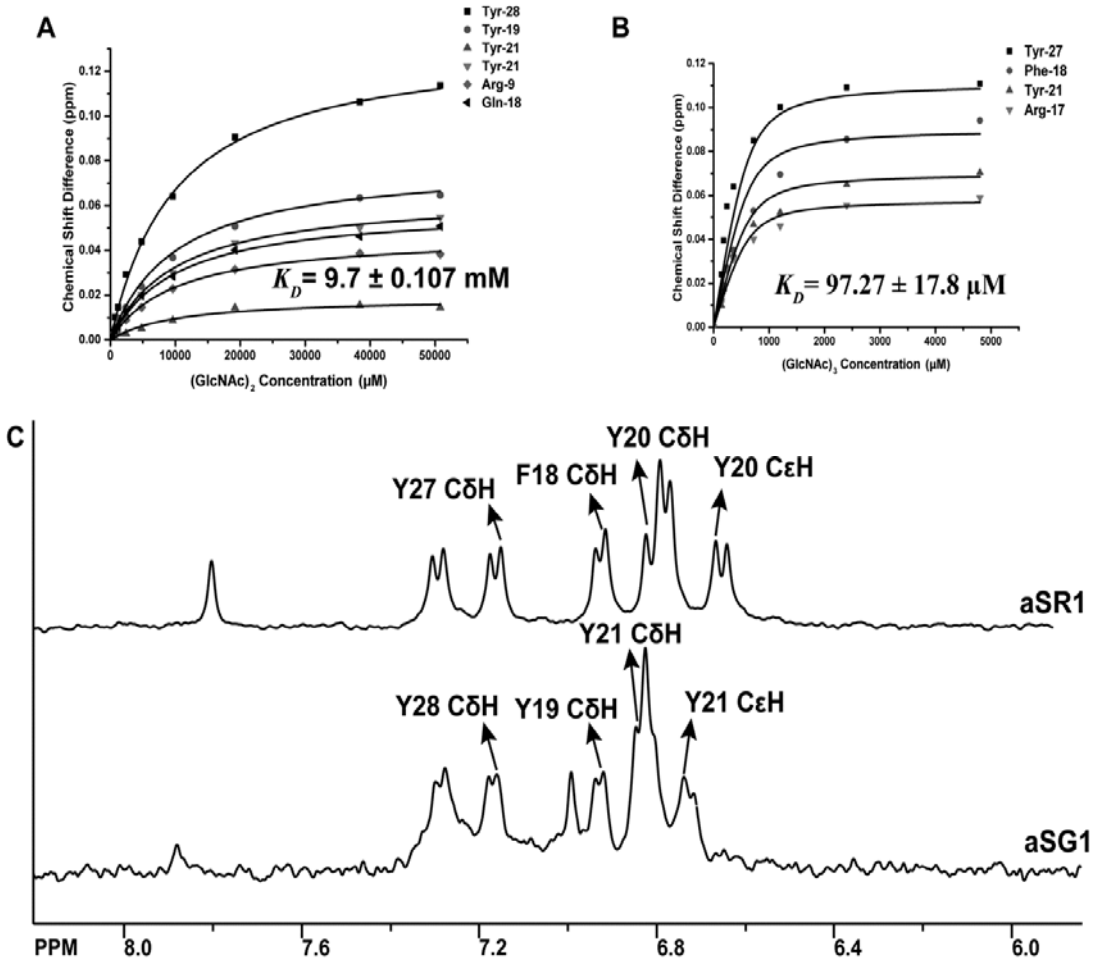


Figure 3.8. ^{1}H-NMR titration of altides with chitin oligosaccharides. (A) ^{1}H-NMR titration plot of aSG1 with chitobiose. (B) ^{1}H-NMR titration plot of aSR1 and chitotriose. (C) Overlay of 1H-NMR spectra of aSG1 and aSR1 indicating the residues involved in the binding interaction.

2.6. Anti-fungal activity of altides

The susceptibility of seven common phyto-pathogenic fungi (*Alternaria alternata, Alternaria brassiciola*, *Aspergillus niger*, *Curvularia lunata*, *Fusarium oxysporum, Rhizoctonia solani* and *Verticulum dahilae*) to altides was assessed using the disc diffusion assay. Discs of growing fungal mycelia were inoculated on Potato Dextrose Agar (PDA) or Malt Extract (ME) plates and incubated at 25°C for 48 - 72 h till a radial colony was formed. Paper discs (6 mm) were placed at the growing mycelial ends and impregnated with different concentrations of altides and incubated for 24 h at 25°C.

After incubation with 17.5 to 70 µg of aSG1, crescent-shaped inhibition zones were observed in *A. alternata*, *A. brassiciola*, *F. oxysporum* and *R. solani* indicating that aSG1 had an inhibitory effect on the growth of these fungi (Fig. 3.9 A). Altide aSR1 from the red variety of *A. sessilis* could inhibit the growth of *A. alternata*, *A. brassiciola, R. solani* and *C. lunata* (Fig. 3.9 B). Altides aSG1 and aSR1 were ineffective against *A. niger* at the tested concentration. Having established the susceptibility of fungal strains to altides, the half-maximal inhibitory concentration was calculated using the micro-broth dilution assay.

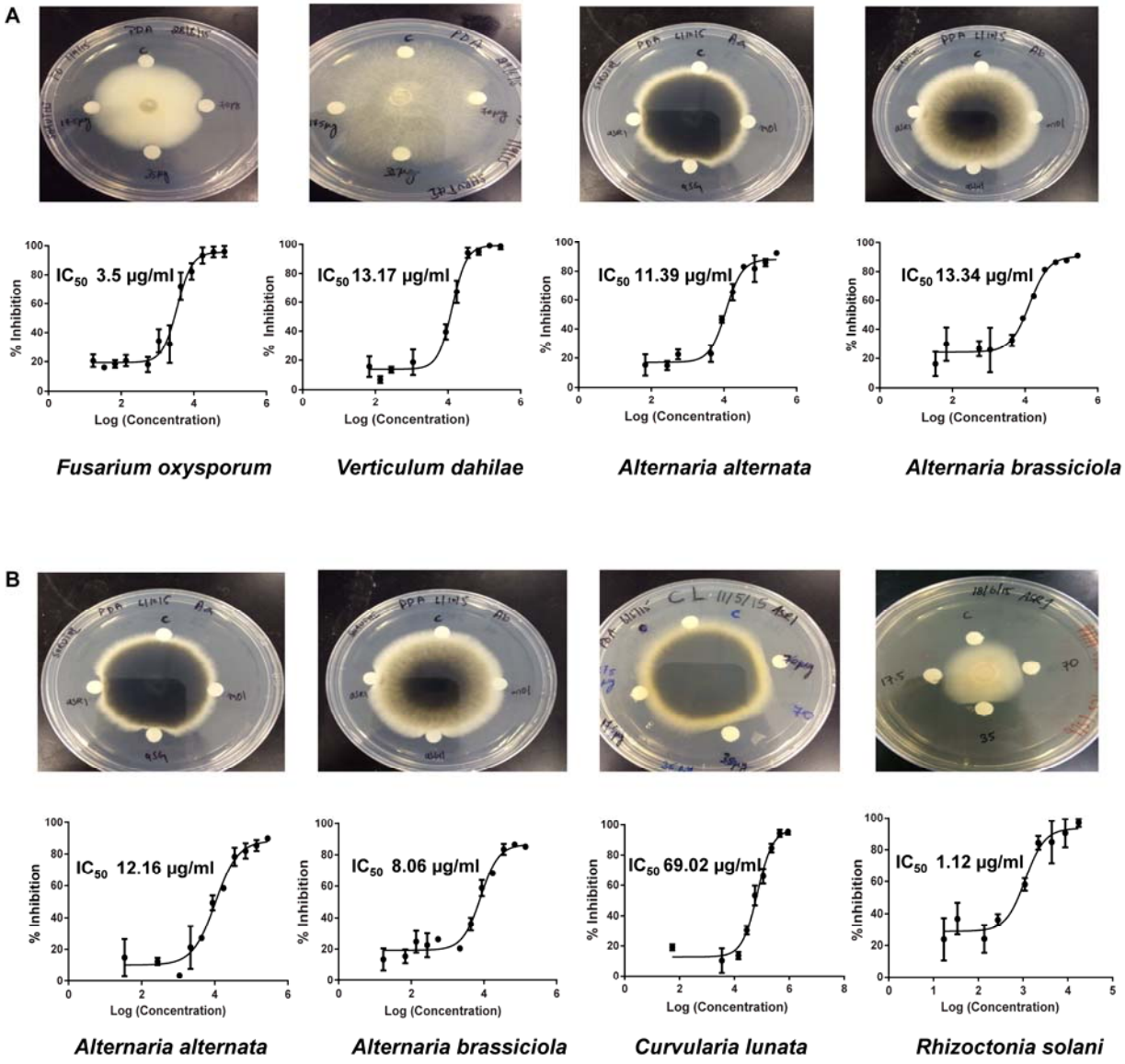


Figure 3.9. Anti-fungal activity of altides (A) aSG1 and (B) aSR1. Formation of arc-shaped inhibition zones in the radial disc diffusion assay indicated susceptibility of fungi against altides. The IC_{50} was calculated from the dose-response curve obtained from the microbroth dilution assay.

To obtain IC_{50} of altides against the susceptible fungi, 2.5 x 10^3 fungal spores were seeded in 96-well plates and incubated with different concentrations of altides for 24 h at 25°C. After incubation, the cells were fixed, stained with 0.1% methylene blue and absorbance was read at 650 nm. The IC_{50} was calculated from a dose-response curve generated by plotting percentage inhibition against peptide concentration (Fig. 3.9). Altide aSG1 demonstrated high potency against *F. oxysporum* with an IC_{50} 3.5 µg/ml while IC_{50} against *A. alternata, A. brassiciola* and *R. solani* ranged 11.39 µg/ml to 13.34 µg/ml. *R. solani* was most sensitive to anti-fungal activity of aSR1 demonstrated by a remarkably low IC_{50} of 1.12 µg/ml.

To gain insight into the putative mechanism of anti-fungal activity of altides, 1X10^3 fungal spores were seeded on 12-well slide chambers and incubated with different concentrations of altides for 24 h at 25°C. Bright field microscopy was performed to observe any morphological changes. From figures 3.10 and 3.11 it is observed that, the hyphae are stunted and more branched than the untreated control with a significant retardation of hyphal growth when *A. alternata, A. brassiciola* are treated with aSG1 and aSR1. Together this data indicates that altides are anti-fungal agents that inhibit the growth of chitin-containing phytopathogenic fungi by retarding hyphal elongation at growing ends of the mycelia.

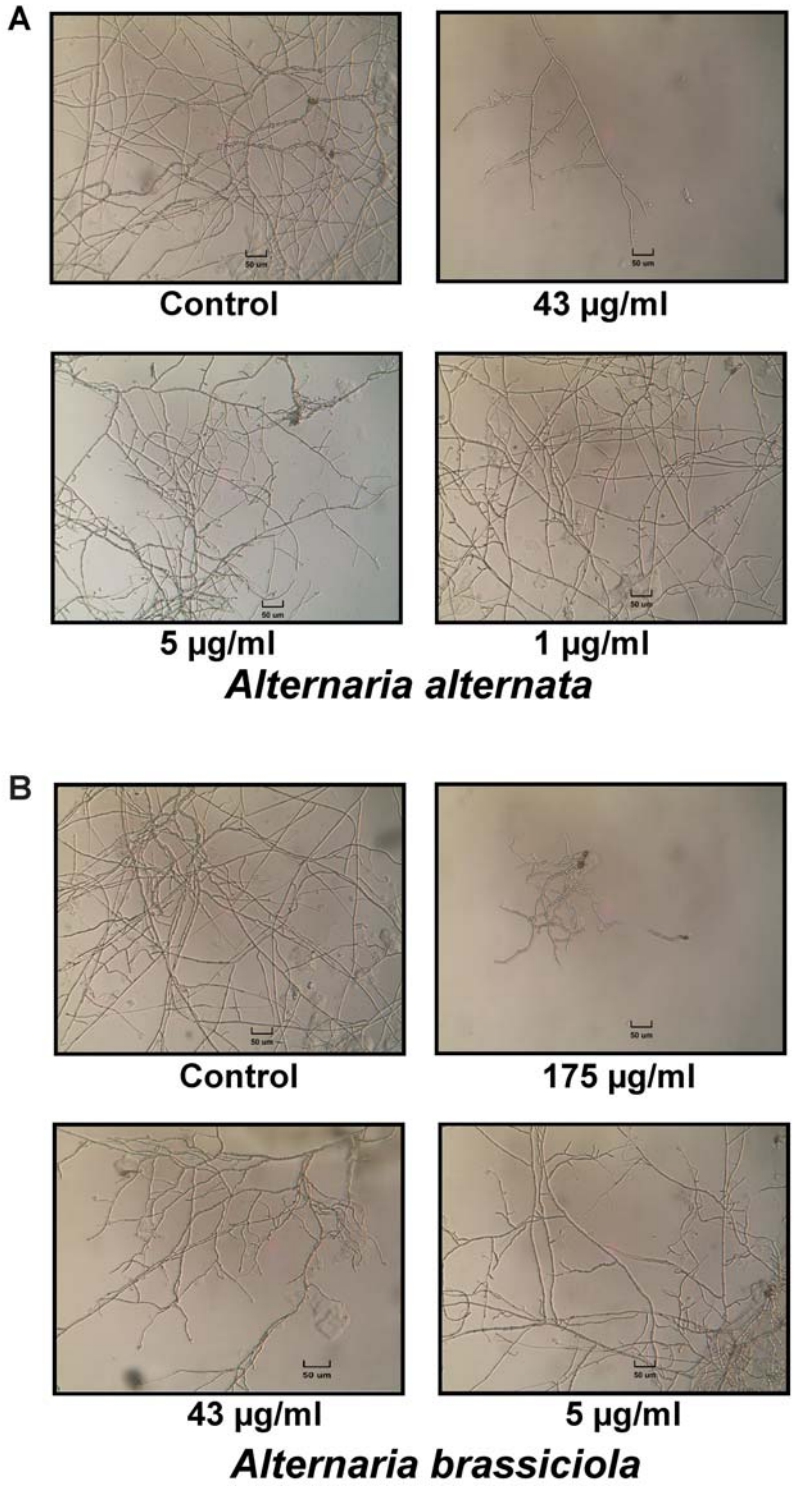


Figure 3.10. Hyphal growth inhibition of fungi treated with aSG1. (A) *Alternaria alternata* (B) *Alternaria brassiciola*. Formation of short stunted hyphae indicated that altides can inhibit hyphal growth at the growing tips of fungal mycelia.

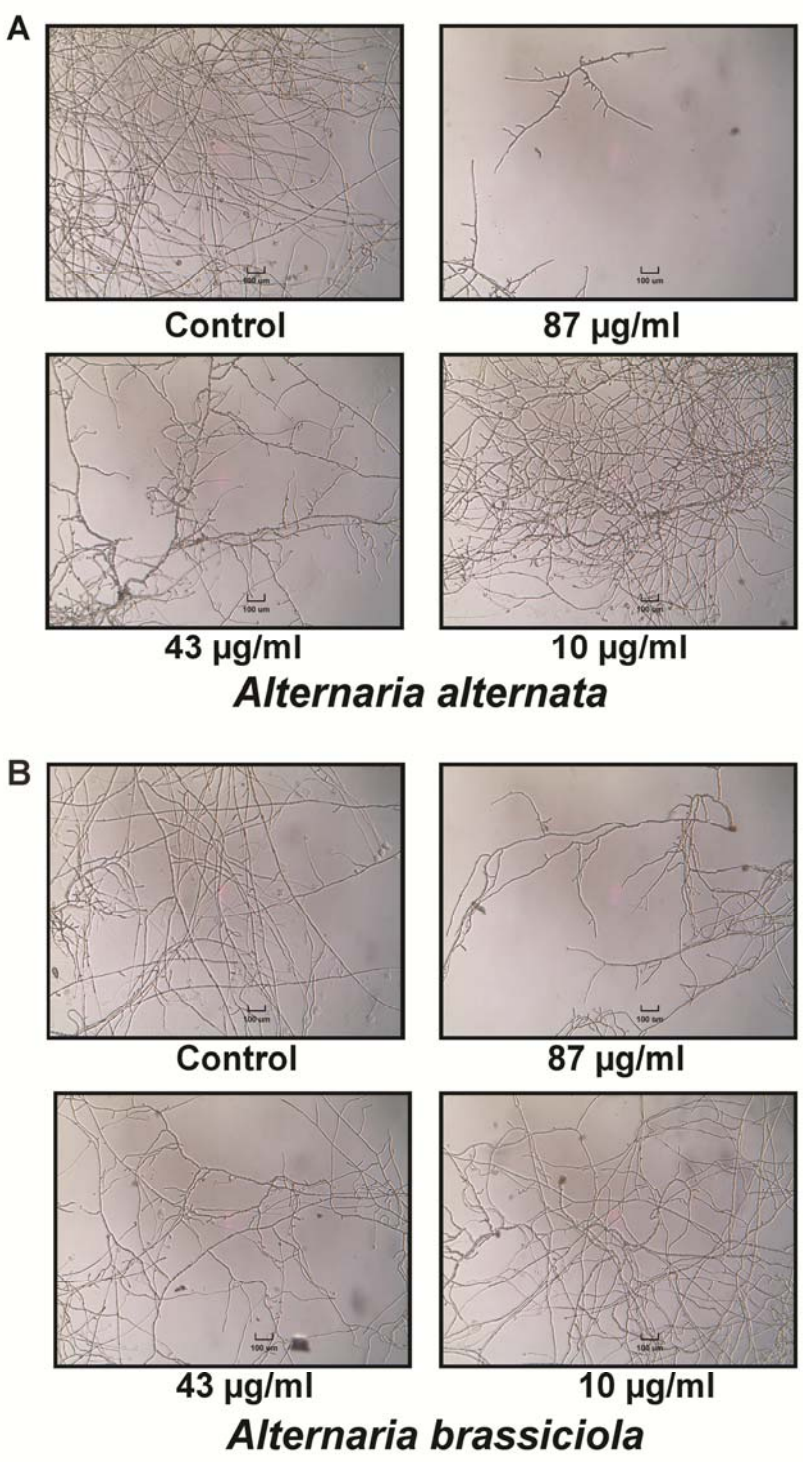


Figure 3.11. Hyphal growth inhibition of fungi treated with aSR1. Significant retardation of hyphal elongation was observed when (A) *Alternaria alternata* and (B) *Alternaria brassiciola* were treated with 87.5 µg/ml of aSR1.

2.7. Heat and proteolytic stability

Thermal and enzymatic degradation cause significant loss of peptides and proteins in herbal medicine. To assess the stability of altides to heat and proteolytic degradation, altides were heated at 100°C for 1 h or incubated with trypsin and pepsin for 6 h. Presence and extent of degradation was assessed from the UPLC chromatographs and MALDI-TOF MS. Greater than 90% of altides remained stable to thermal degradation after 1 h while >75% of altides were stable against proteolytic degradation by trypsin and pepsin after incubation for 6 h (Fig. 3.12). In the heat and enzyme stability tests, linear peptides DALK and RV-14 (RLYRRGRLYRRNHV), respectively, synthesized in-house served as controls. From the figure we can see that DALK degraded completely within 1 h of heat treatment and RV-14 within 6 h of incubation with trypsin and pepsin. Together, this data provides satisfactory evidence of the stability of altides against thermal and enzymatic degradation.

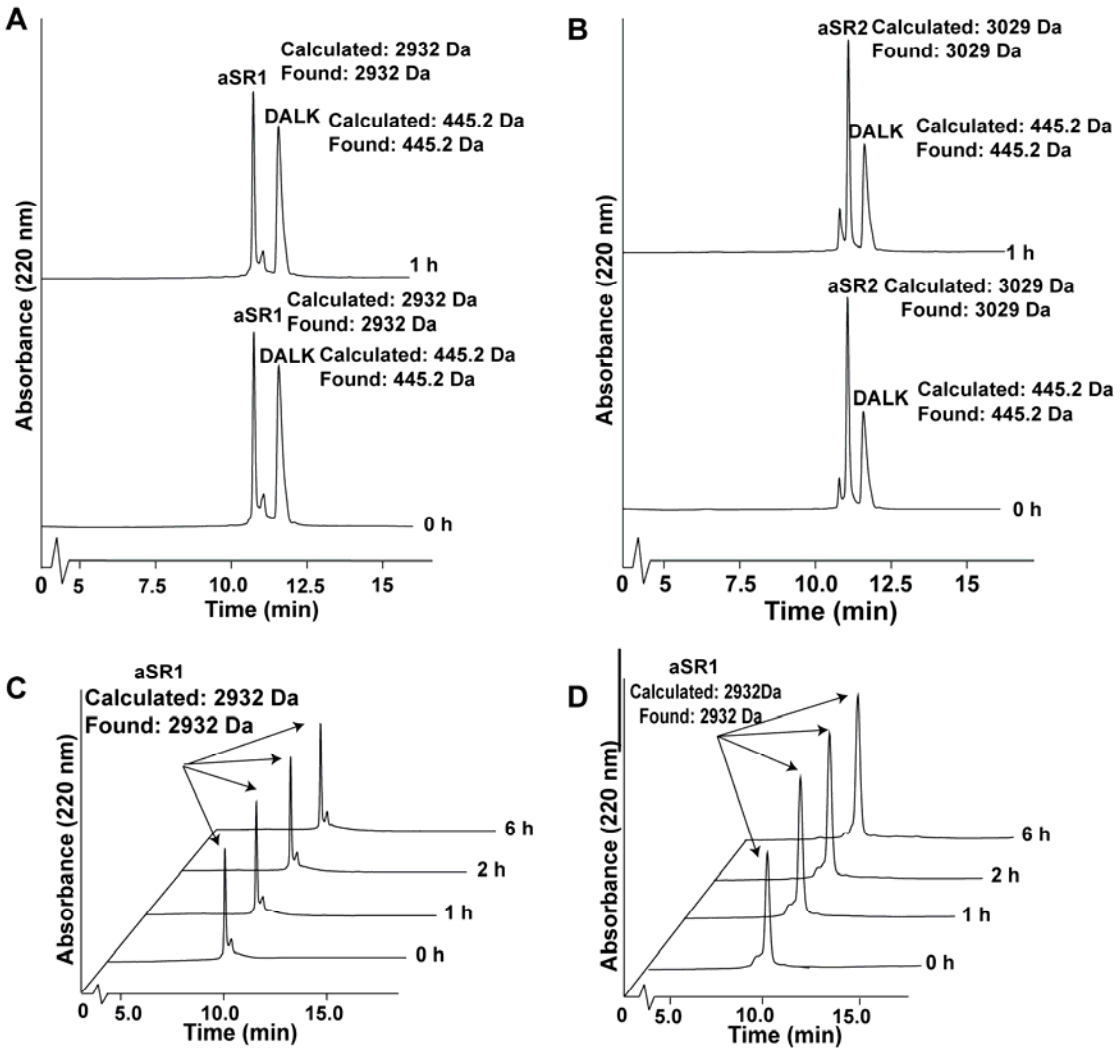


Figure 3.12. Stability assays of altides. Thermal stability of (A) aSR1 and (B) aSR2 at 100°C. DALK, a linear peptide synthesized in our laboratory was used as an internal standard. (C) Trypsin stability assay of aSR1. (D) Pepsin stability assay of aSR1. Altide, aSR1 remained stable to thermal and enzymatic degradation.

2.8. Bioactivity assays

Cytotoxic, hemolytic and anti-bacterial effects of altides, aSG1 and aSR1 were studied. In the cytotoxicity assay, Vero cells were treated with upto 100 μM of altides aSG1 and aSR1 and their viability was assessed using the PrestoBlue reagent. After treatment, it was found that altides did not demonstrate any cytotoxic effect on Vero cells at concentrations upto 100 μM. The hemolytic effect of altides was studied using human type O erythrocytes with 0.1% Triton X-100 as the positive control. Altides, aSG1 and aSR1 did not display any hemolytic effects upto 100 μM.

The anti-bacterial activity of altides, aSG1 and aSR1 was tested against gram-positive *Staphylococcus aureus* and gram negative *Escherichia coli* using the radial diffusion assay. Cyclotide, cT8, a potent anti-microbial CRP isolated from *Clitoria ternatea* in our laboratory was used as a positive control [40]. Altides, aSG1 and aSR1 showed negligible anti-bacterial effect at concentrations upto 100 μM.

3. Discussion

Hevein-like peptides are integral in plant defense against fungi and insect pathogens. They can be divided into three sub-classes, namely, 6C-, 8C- and 10C-hevein-like peptides based on their cysteine content [49]. Thus far, four 6C-hevein-like peptides have been reported, namely, two Ac-AMPs from *A. caudatus [45]*, Ar-AMP from *A. retroflexus* [50] and IWF-4 from *B. vulgaris [46].* In this study we have isolated and characterized six novel 6C-hevein-like peptides from two varieties of *A. sessilis*, expanding the existing library of 6C-hevein-like peptides. Altides are 29-30 amino acid residues in length and adopt the cystine-knot disulfide arrangement previously found in CRPs like cyclotides [40, 157] and cystine knot-type of α-amylase inhibitors (CKAIs) [158]. This knotted topology renders altides highly stable against thermal and enzymatic degradation. Altides bind to chitin oligosaccharides with dissociation constants in the micromolar range with aromatic residues occupying the chitin-binding domain mediating responsible the binding interaction. Altides could inhibit the growth of commonly occurring phyto-pathogenic fungi with IC_{50} in the range of 1.12 µg/ml to 69.02 µg/ml and were non-cytotoxic and non-hemolytic to Vero cells and type O erythrocytes respectively.

3.1. Sequence conservation of altides

Sequence comparison of altides within the green variety revealed >89% homologies with aSG1 and aSG3 differing by only one amino acid residue (Gly-31 of aSG3 is absent in aSG1). Altides of the red variety were 85% identical to each other where aSR2 and aSR3 differ by only one amino acid residue (Gly-30 of aSR2 is absent in aSR3). Sequence comparison of altides between red and green varieties showed 75%-90% homology.

Altides shared a sequence homology of 70%-90% with other 6C-, 8C- and 10C-hevein-like peptides from different plant families. Sequence alignment of altides with previously reported 6C-hevein-like peptides shows that the N-terminal is highly variable when compared to the C-terminal. The C-terminal Gly is conserved in all altides and IWF-4 but is replaced by Arg in Ac-AMP2 and Ar-AMP. A common observation in hevein-like peptides is the difference of one amino acid residue at the C-terminal. For example, aSG2 and aSG3 differ by only one glycine residue at the C-terminus. A similar observation was made among aSR2 and aSR3. This phenomenon could be due to different post-translational processing or proteolytic degradation by carboxypeptidase during the purification process [47]. The molecular diversity of altides is evident from our results as none of the loops are absolutely conserved and vary by 2-3 residues even in altides from the same plant species. Except for the six cysteine and seven glycine residues, variations are also observed in the chitin-binding domain where aSR1 comprises a Phe residue instead of Tyr. This variation may have arisen due to a point mutation in the gene encoding aSR1 and is unlikely to affect

the binding interaction with chitin as the aromatic ring of Phe can still facilitate CH-π stacking interactions [79].

Among all altides, aSR1 showed highest homology (79.6%) to Ac-AMP2 isolated from *A. caudatus*. This data suggests that there could be a close evolutionary relationship between altide aSR1 from *A. sessilis* var. red and *A. caudatus*. Although altides, Ac-AMPs, Ar-AMPs and IWF-4 are isolated from plants belonging to the Amranthaceae family, they show some sequence diversity but retain the same function. This suggests that 6C-hevein-like peptides evolved from a common ancestral gene that is probably retained throughout the Amaranthaceae family due to its integral role in plant defense. Screening more plants from Amaranthaceae and closely-related plant families may help gain further insight into the exact evolutionary relationship among 6C-hevein-like peptides.

3.2. Biosynthesis of altides

The precursor organization of altides comprises three domains - an ER signal domain, a mature CRP domain and a C-terminal tail. This precursor organization is similar to thionins [159] but differs from cyclotides [94, 157, 160], CKAIs [158] and ω-conotoxins [161] that comprise an N-terminal pro-domain before the mature domain and lack a C-terminal tail (Fig. 3.13). The presence of a signal peptide indicates that the bioprocessing of altides follows the secretory pathway as suggested for many CRP [91, 162, 163].

This feature distinguishes them from small peptides (5-12 amino acid residues in length) that are synthesized by multi-enzyme complexes [90]. An interesting feature of the precursor is the presence of a C-terminal tail of 25-33 amino acid residues that bears no sequence homology to any known conserved domains in the NCBI database. The role of the C-terminal tail is yet to be understood.

Sequence comparison of altide precursors with Ac-AMP2, IWF-4 and Ar-AMP showed that they were highly homologous with Ac-AMP2 (81%) and Ar-AMP (66%). The precursors of altides were five residues longer than IWF-4 and their signal peptides shared no sequence homology. Due to high sequence identity between the clones of altides, Ac-AMP2 and Ar-AMP it can be presumed that they evolved by gene duplication from the same ancestral gene by divergent evolution. In divergent evolution, a single ancestral gene undergoes variations to help the plant improve its survival and adapt to environmental changes [164].

Knowledge of the gene organization and precursor structure of altides can be used to develop transgenic crops with enhanced resistance to fungal pathogens. The transgenic approach could prevent the use of toxic chemical pesticides that are harmful for human consumption and adversely affect the environment. Thus far, 6C-hevein-like peptides have not been used to develop transgenic crops. However, closely related α-amylase inhibitors α-AI1 and α-AI2 have been cloned in plants which are deemed fit for human consumption [165]. Therefore, insight into the gene organization and bioprocessing of 6C-hevein-like peptide opens up new avenues for transgenic crop research and can be implemented for the development of pest-resistant crops.

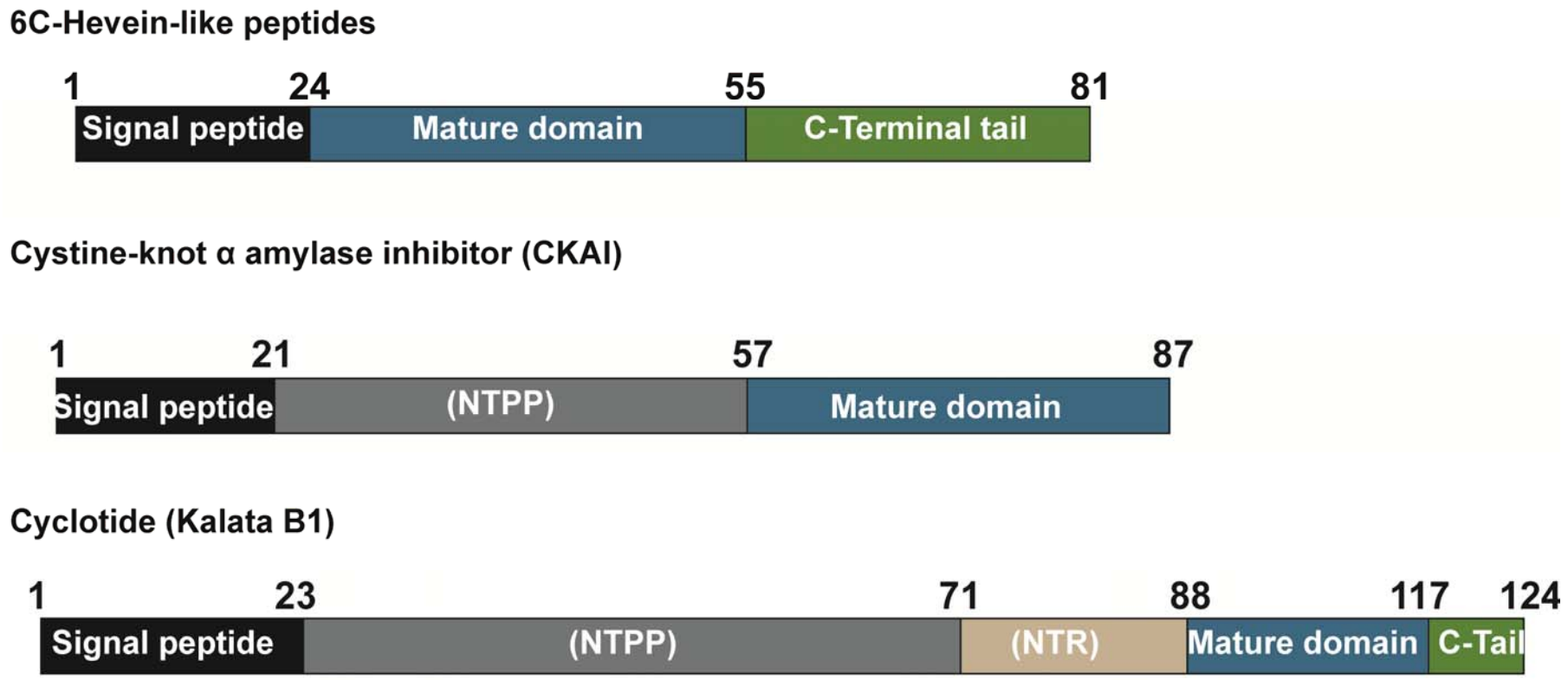


Figure 3.13. Comparison of the precursor organization of 6C-hevein-like peptides with other knottin-type CRPs. Precursors of hevein-like peptides differ from cyclotides and cystine-knot α-amylase inhibitors (CKAIs) as they do not comprise an N-terminal pro-domain (NTPP) or an N-terminal repeat (NTR) before the C-tail. CKAIs lack the C-tail while cyclotides comprise a short C-tail of 7 amino acid residues.

3.3. Binding interaction of altides with chitin oligosaccharides

Analysis of the binding interaction and the residues involved in binding of altides, aSG1 and aSR1 was performed by ^{1}H-NMR titrations with chitobiose and chitotriose respectively. From our experiments we found that aromatic residues, Tyr-19, Tyr-21, and Tyr-28 from aSG1 and Phe-18, Tyr-20, and Tyr-27 from aSR1 respectively are involved in the binding interaction (Fig 3.14). These aromatic amino acid residues occupy loops 3 and 4 of altides that comprise the conserved chitin-binding domain S-X-(F/W/Y)-X-(F/W/Y)-C-G-X_4-Y. Similar binding motifs were observed when Ac-AMP2, hevein, pseudohevein and WGA [76-78, 166] were titrated with chitin oligosaccharides. This data indicates that the chitin-binding domain is conserved throughout the hevein-like peptide family and the peptide-sugar complexes are stabilized by CH-π stacking interactions and van der Waal's contacts between the aromatic residues and the sugar moiety. The dissociation constant K_D of the binding interaction between aSG1 and chitobiose was two orders of magnitude higher than that of aSR1 and chitotriose. This significant difference can be due to the multivalent nature of the binding interaction [58], where increased number of binding sites in chitotriose lead to increased binding affinity. Similar observations were made when hevein was titrated with chitin oligosaccharides of increasing chain length $(GlcNAc)_n$ (n=1-3) where, a difference of one-order of magnitude was observed per GlcNAc residue.

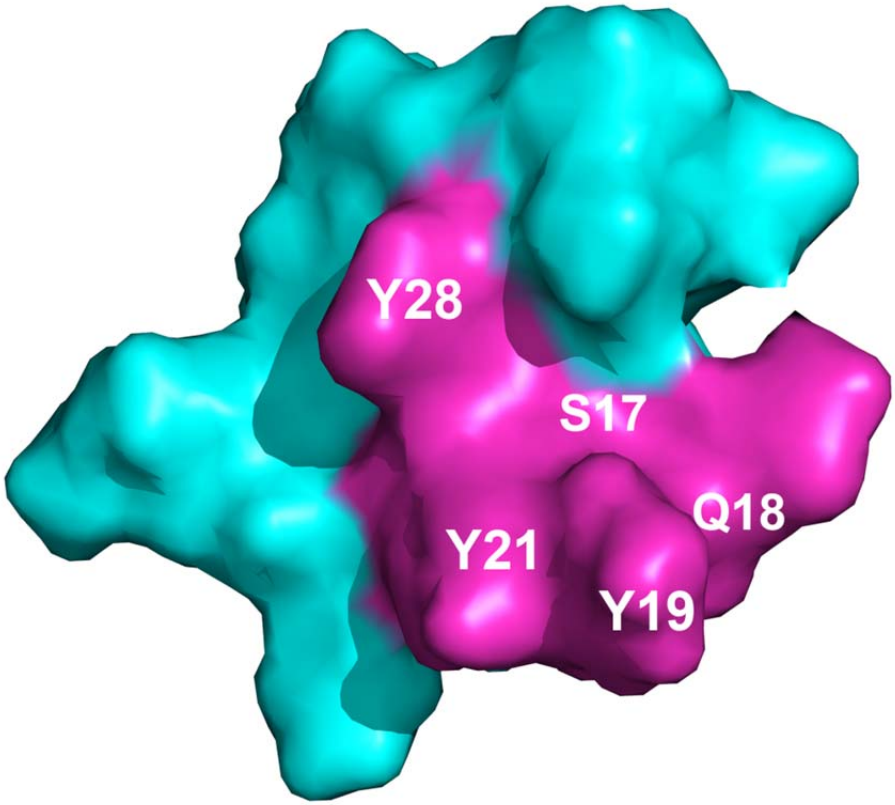


Figure 3.14. Chitin-binding motif of aSG1. ^{1}H-NMR titrations with chitobiose showed that Gln-18, Tyr-19, Tyr-21, and Tyr-28 were involved in the binding interaction.

The dissociation constant of altides is similar to hevein even though they comprise a C-terminal truncation indicating that the C-terminal does not play a major role in binding. This observation has previously been made by Aboitiz *et al.* who synthesized a truncated hevein analog (HEV32) and reported that there was no significant difference in the binding constant upon titration with chitotriose [79]. Together, this data provides insight into the nature of the binding interaction of altides with chitin oligosaccharides and shows that the chitin-binding domain is conserved throughout the hevein-like peptide family. Knowledge of the binding interaction can be used for efficient design of putative anti-fungal agents that have increased affinity for chitin. The dissociation constant K_D of the binding interaction between aSG1 and chitobiose was two orders of magnitude higher than that of aSR1 and chitotriose. This significant difference can be due to the multivalent nature of the binding interaction [58], where increased number of binding sites in chitotriose lead to increased binding affinity. Similar observations were made when hevein was titrated with chitin oligosaccharides of increasing chain length $(GlcNAc)_n$ (n=1-3) where, a difference of one order of magnitude was observed per GlcNAc residue.Together, this data provides insight into the nature of the binding interaction of altides with chitin oligosaccharides and shows that the chitin-binding domain is conserved throughout the hevein-like peptide family. Knowledge of the binding interaction can be used for efficient design of putative anti-fungal agents that have increased affinity for chitin.

3.4. Anti-fungal activity of altides

Six phytopathogenic fungal strains (*A. alternata*, *A. brassiciola*, *C. lunata*, *F. oxysporum*, *R. solani and V. dahilae)* were susceptible to treatment with altides aSR1 and aSG1 as indicated by the formation of arc-shaped inhibition zones at growing hyphal tips. The IC_{50} of altides against these fungi ranged from (1.12 – 69.02 µg/ml) depending on the peptide and strain tested and is comparable with other 6C-hevein-like peptides like, Ac-AMP2, IWF-4 and Ar-AMP [45, 46, 50]. The potency of all 6C-hevein-like peptides is higher than hevein whose IC_{50} is in the range of 90-1250 µg/ml.

Morphological changes like short and stunted hyphae with increased branching were observed when *A. alternata* and *A. brassiciola* were treated with aSG1 and aSR1. Similar morphology was observed when IWF-4 was incubated with *Cercospora beticola [46]*. These observations indicate that altides and IWF-4 inhibit fungal growth through the same mechanism. It was speculated that, as a consequence of their chitin-binding property, altides can disrupt the “steady-state” model of hyphal growth [100]. According to the “steady-state” model, during hyphal growth, the tip of the hypha is plastic, comprises nascent chitin chains and gradually gets rigidified at the base by cross-linking of newly-synthesized chitin chains with β-glucans. In this manner, a ‘steady-state’ of plasticity and rigidification is maintained for efficient hyphal growth. Due to the plasticity of the hyphal apex, the nascent chitin chains are highly susceptible to chitinases and anti-fungal agents. Thus, altides may inhibit the growth of fungi by binding to

these nascent chitin chains and preventing the rigidification process resulting in the formation of short and stunted hyphae in *A. alternata* and *A. brassiciola*.

Taken together, our data highlights the sequence diversity, biosynthesis pathway, chitin-binding interaction and anti-fungal property of a novel suite of 6C-hevein-like peptides from two varieties of *A. sessilis*. Our data has helped expand the existing library of 6C-hevein-like peptides from four to ten. The altide aSR1 is the smallest 6C-hevein-like peptide isolated thus far with a binding affinity to chitin oligosaccharides comparable to hevein. Although the anti-fungal activity of altides is established, further insights are needed into their putative mechanism of action. This data can be used to develop robust anti-fungal agents with increased potency against fungal pathogens.

Chapter 4

Morintides: Novel 8C-Hevein-like Peptides from *Moringa oleifera*

1. Introduction

Moringa oleifera is a drought-resistant tree that can grow to a height of 10 m and its trunk can reach a diameter of 45 cm [167]. Due to its high medicinal and nutritional value, *M. oleifera* is now incorporated in marketed health formulations which are remedies for a variety of health disorders (Fig. 4.1) [168]. *M. oleifera* seeds are the best known natural coagulants which are commonly used to purify turbid water. It is speculated that the coagulation properties are due to dimeric proteins in the seeds which are highly stable [169]. The seeds, stem, roots and leaves of *M. oleifera* are reported to exhibit anti-fungal effects [111, 118]. However, there is no report on the peptide and protein components that contribute to the putative therapeutic value of *M. oleifera* leaves. Since hevein-like peptides are primarily involved in plant defense against fungi, we screened the leaves of *M. oleifera* for putative hevein-like peptides.

Figure 4.1. Nutritional benefits of *Moringa oleifera*. Moringa leaves are sold as health formulations to combat malnutrition around the world. Figure adapted from http://www.tentree.com/blog/benefits-of-the-moringa-oleifera-tree/

In this study, we report three anti-fungal 8C-hevein-like peptides from *M. oleifera*, designated morintides and abbreviated as mO1, mO2 and mO3. Sequencing using LC-ESI-MS/MS showed that morintides mO1 and mO2 are 44 amino acid residues in length while mO3 is 35 amino acid residues long. Transcriptome data analysis revealed that full-length precursors of morintides contain three domains, a signal peptide, a mature CRP domain and a short C-terminal tail. Morintides are the first 8C-hevein-like peptides that have a significantly short C-terminal tail unlike other 8C- and 10C-hevein-like peptides that contain a long C-terminal tail encoding for another bioactive protein like Barwin or chitinase [88, 170, 171]. As a consequence of its chitin-binding property mO1 inhibited the growth of two phytopathogenic fungal strains with IC_{50} in the micromolar range by inhibiting the growth of fungal hyphae. Together our data provide an insight into the sequence, biosynthesis and anti-fungal activity of morintides.

2. Results

2.1. Screening and isolation of peptides from *M.oleifera*

Screening of *M. oleifera* leaves revealed a cluster of peptides around 4 kDa. A scale-up extraction of 3 kg of plant material was performed by blending in equal volume of water. Putative CRPs were isolated by multiple rounds of SCX- and RP-HPLC and isolated peptides were profiled using MALDI-TOF MS in the mass range of 2-5 kDa. The MS profile of *M. oleifera* indicated that two peptides with m/z of 4536.71 Da and 4463.76 Da were present in the leaf extract and were designated morintide mO1 and mO2 respectively. The purified putative CRPs were subjected to reduction with DTT for 2 h followed by alkylation with IAM for 1 h at 37°C. After reduction and alkylation of the disulfide bonds, a mass shift of +474 Da was observed indicating the presence of eight cysteine residues and confirming that the isolated peptides were CRPs (Fig. 4.2).

Sequencing of morintides was performed by nanospray MS/MS. The mass difference between b and y ions was used to deduce the primary sequence of mO1 and mO2. Since tandem MS cannot differentiate isobaric residues like Leu/Ile, they were assigned based on the transcriptome data. From proteomic and transcriptomic analysis, morintides were 44 amino acid residues in length.

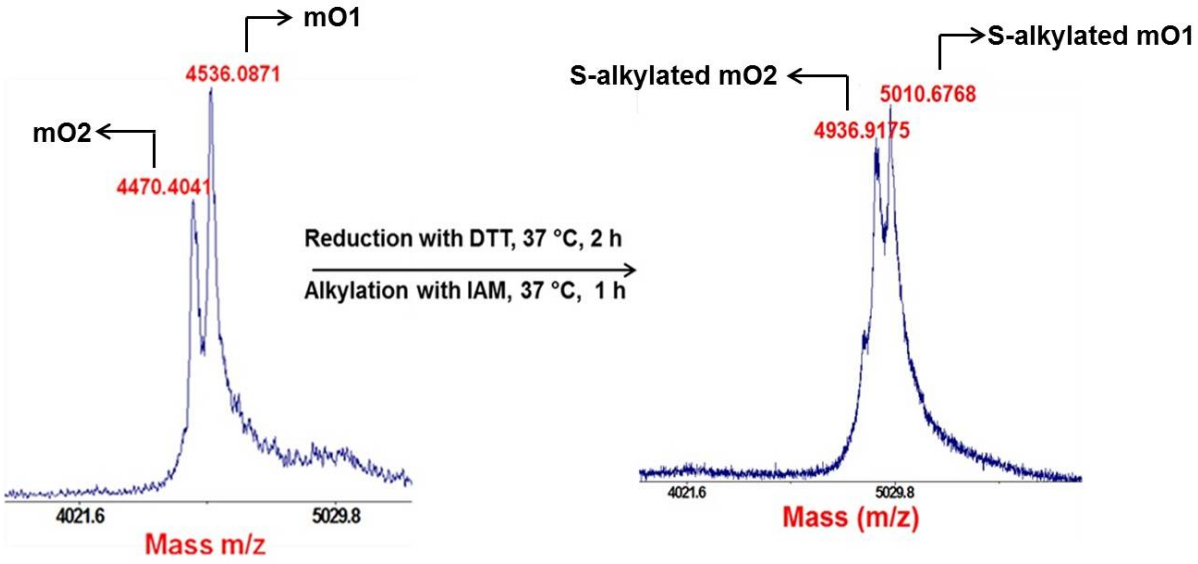
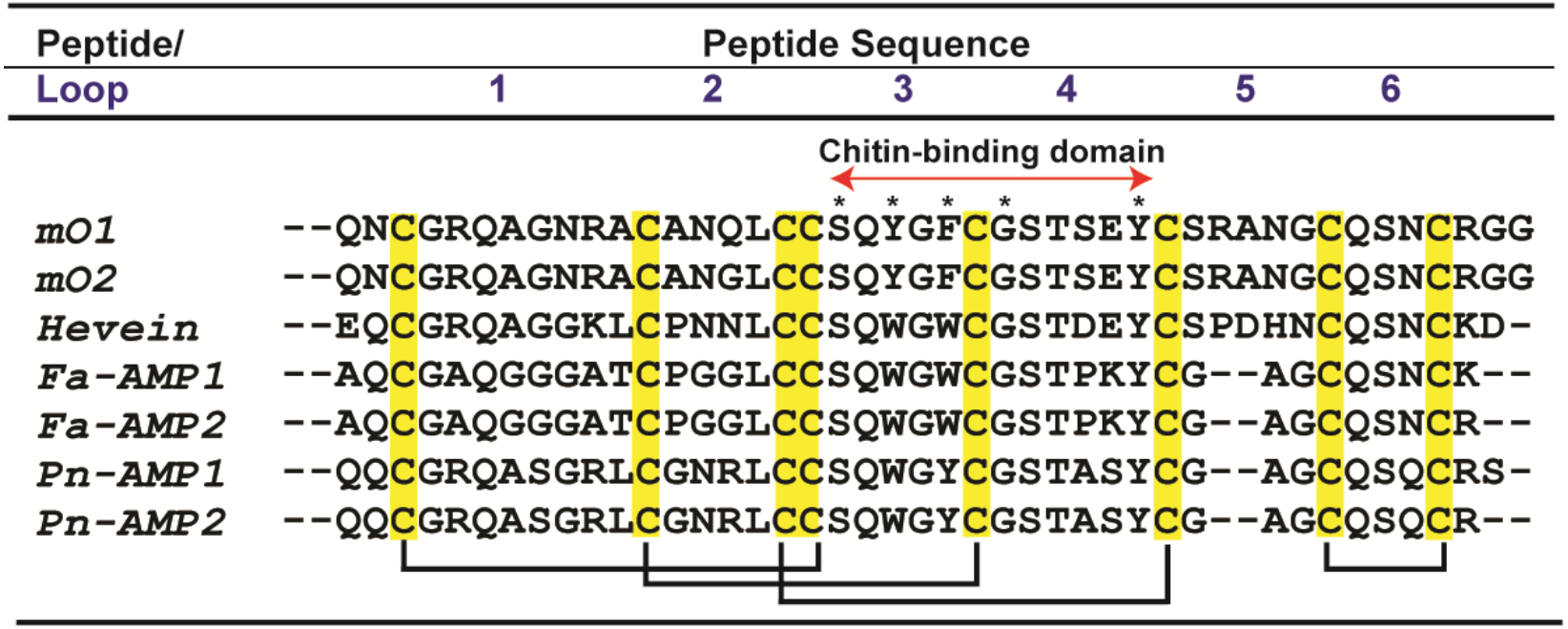


Figure 4.2. Reduction and alkylation of morintides. Mass spectra of reduced-alkylated morintides revealed that a peak with m/z 5010 was of highest abundance. The increase in mass by approximately 474 Da indicates that morintides comprise 8 cysteine residues.

The primary sequence comprised high content of cysteine (eight residues), glycine (seven residues), asparagine (five residues) and glutamine (five residues) making up nearly 50% of the peptide sequence. The glutamine residue at the N-terminal was spontaneously converted to pyroglutamic acid making the peptide resistant to degradation by exopeptidases. Sequence comparison of morintides revealed that their primary sequences are highly homologous and differ in one amino acid position: Gln-15 in mO1 was replaced by Gly-15 in mO2. Sequence alignment revealed a cysteine-spacing pattern of X_2**C**-X_8**C**-X_4**CC**-X_5**C**-X_6**C**–X_5**C**–X_3**C**–X_3 with an absolutely conserved chitin binding domain SQYGFCX_4Y in loops 3 and 4, where X represents any amino acid. BLAST analysis revealed that morintides belonged to the hevein-like peptide family, shared greater than 54% homology with 8C-hevein-like peptides and highest homology (64.3%) with hevein and Pn-AMPs (Table 4.1).

Table 4.1. Consensus sequences of morintides.



Sequence alignment with previously reported 8C-hevein-like peptides including hevein (29), Fa-AMPs (62) and Pn-AMP (38). The chitin-binding domain is located in loops 3 and 4 and is highlighted by a red arrow. Asterisks indicate the specific residues involved in binding to chitin and the disulfide pattern is indicated at the bottom of the table.

2.2. Modeled structure of morintide mO1

Structure modeling for mO1 was performed using the online sever SWISS-MODEL. The best three results were all based on the structures of hevein (PDB ID: 4wp4, 1q9b and 1hev) as templates. The model using the structure 1HEV had the best QMEAN score. The modeled structure has two β strands consisting of residues Cys17-Ser19 (β1) and Phe23-Gly25 (β2), forming antiparallel β-sheets (Figure 4.3 A) while residues Ser28-Cys31form a one-turn α-helix. From the model generated it was clear that mO1 is a typical 8C-hevein-like peptide with four disulfide bonds. Due to the high structural and sequence homology between hevein and mO1 it was presumed that the disulfide bond pattern would be the same in both peptides. The disulfide bonds Cys I-Cys IV and Cys II-Cys V link the N terminal loop and the strands β1 and β2 respectively. The disulfide bond Cys III-Cys VI makes the α-helix lie close to the strands β1 and β2. The disulfide bond Cys VII-Cys VIII makes the C-terminus bend, lying beside the strand β1. The two β strands consist all of hydrophobic residues, located in the center of the structure. The four positive charged residues (Arg-10, Arg-5, Arg-33 and Arg-42) are distributed in the N-terminus and C-terminus. There is only one negative charged residue (Glu-29) located in the C-terminal region. The structure of mO1 is very similar to hevein (Figure 4.3 B). For both mO1 and hevein, the two strands β1 and β2 consist mostly of hydrophobic residues. Hevein contains more charged residues than mO1, and has five negative charged residue (Glu-1, Asp-28, Glu-29, Asp-34 and Asp-43) and four positive charged residues (Arg-5, Lys-10, His-35 and Lys-42).

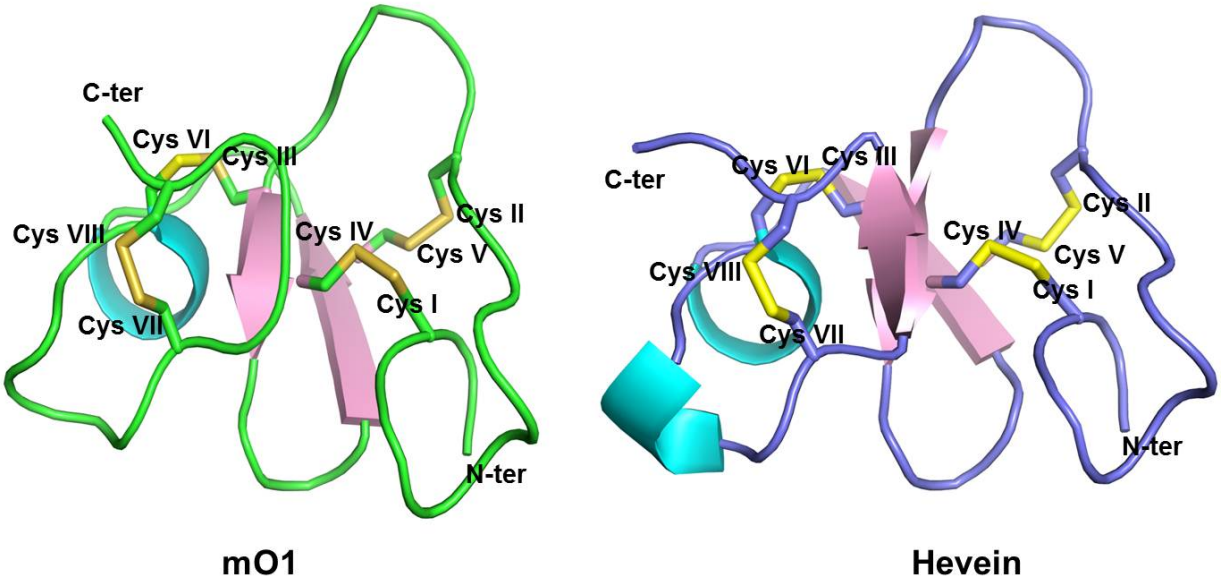


Figure 4.3. Modeled structure of morintide mO1. The structure of hevein modeled using hevein (1HEV) had the best QMEAN score. The modeled structure has two anti-parallel β-sheets and a one turn α-helix. The eight cysteine residues are presumed to have the same disulfide pairing as hevein, Cys I-Cys IV, Cys II-Cys V, Cys III-Cys VI, and Cys VII-Cys VIII.

2.3. Chitin binding activity

To assess the chitin binding activity of morintides, mO1 was incubated with chitin beads for 4 h at 4ºC. RP-UPLC analysis showed that mO1 bound to chitin beads within 1 h (Fig. 4.4 A). On using elution buffers with up to 1 M NaCl, mO1 could not be eluted from the beads indicating a strong binding interaction. The use of 0.5 M acetic acid with heating at 100°C effectively eluted mO1 in 1 h (Fig. 4.4 B). This data indicated that mO1 bound strongly to chitin beads as it could be eluted using a strong elution buffer and heating at high temperature.

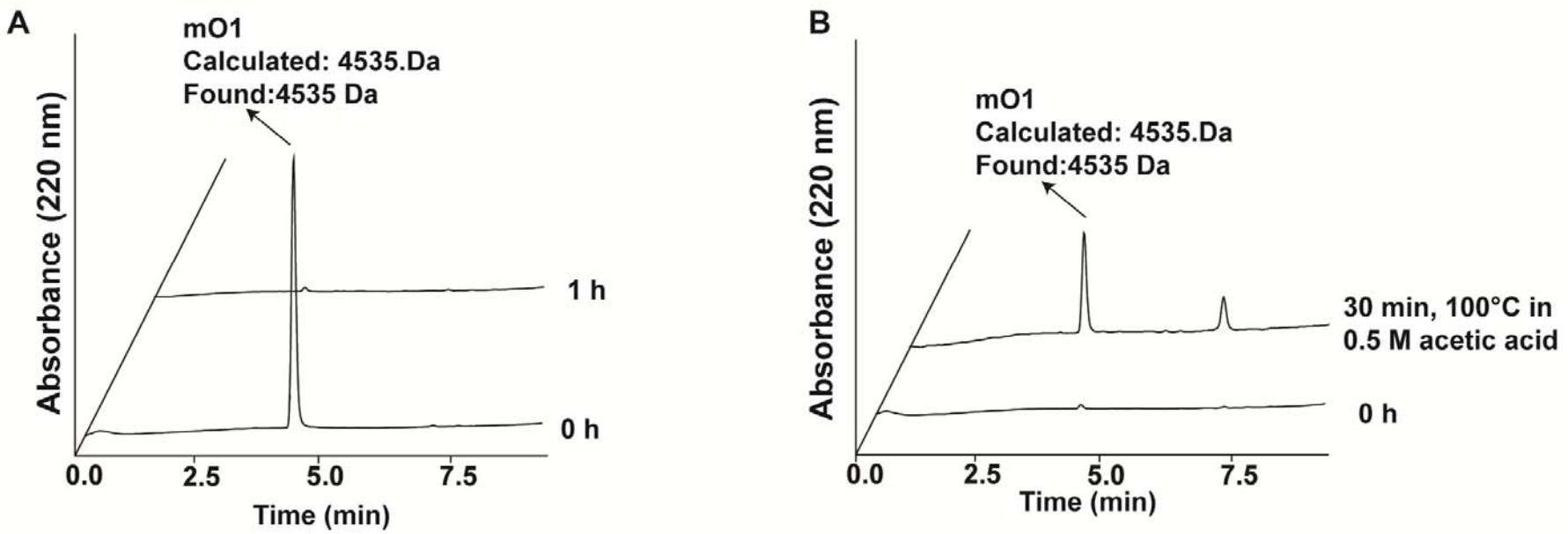


Figure 4.4. Chitin-binding activity of morintide mO1. (A) Incubation with chitin beads for 1 h at 4°C resulted in binding of mO1 within 1 h. (B) Elution of bound mO1 was performed at 100°C for 30 min in 0.5 M acetic acid.

2.4. Anti-fungal activity

The disc diffusion assay was performed to test the vulnerability of seven strains of phytopathogenic fungi to mO1. Fungi were inoculated and allowed to grow at 25°C for 24-72 h or until a radial colony was formed. Morintides of designated concentration were loaded on discs placed at the growing mycelial tips. Out of the seven strains *A. alternata* and *A. brassiciola* showed the formation of crescent-shaped inhibition zones indicating that they are susceptible to treatment with mO1 (Fig. 4.5 A).

To obtain the half-maximal inhibitory concentration (IC_{50}) of mO1, micro-broth dilution assay was performed against the susceptible fungi. After treatment with different morintide concentrations for 24 h, the cells were fixed with methanol and stained with 0.1% methylene blue. Absorbance was measured at 650 nm and a dose-response curve was generated from which the IC_{50} calculated was in the range of 25.5 µg/ml to 60.43 µg/ml after incubation for 24 h at 25°C depending on the fungal strain tested (Fig. 4.5 B).

To observe any morphological changes in fungi sensitive to mO1, the fungal spores were treated with mO1 and grown on slide chambers for 24 h at 25°C. Bright field microscopy showed that treatment of fungal spores with mO1 resulted in morphological changes like short and thick hyphae with swollen tips when compared to the untreated control. The mO1 treated fungal mycelia were closely bunched together and did not branch out as extensively the untreated fungi (Fig 4.6). Together these data suggest that mO1 inhibits fungal growth by retarding the growth of budding hyphae in a dose-dependent manner.

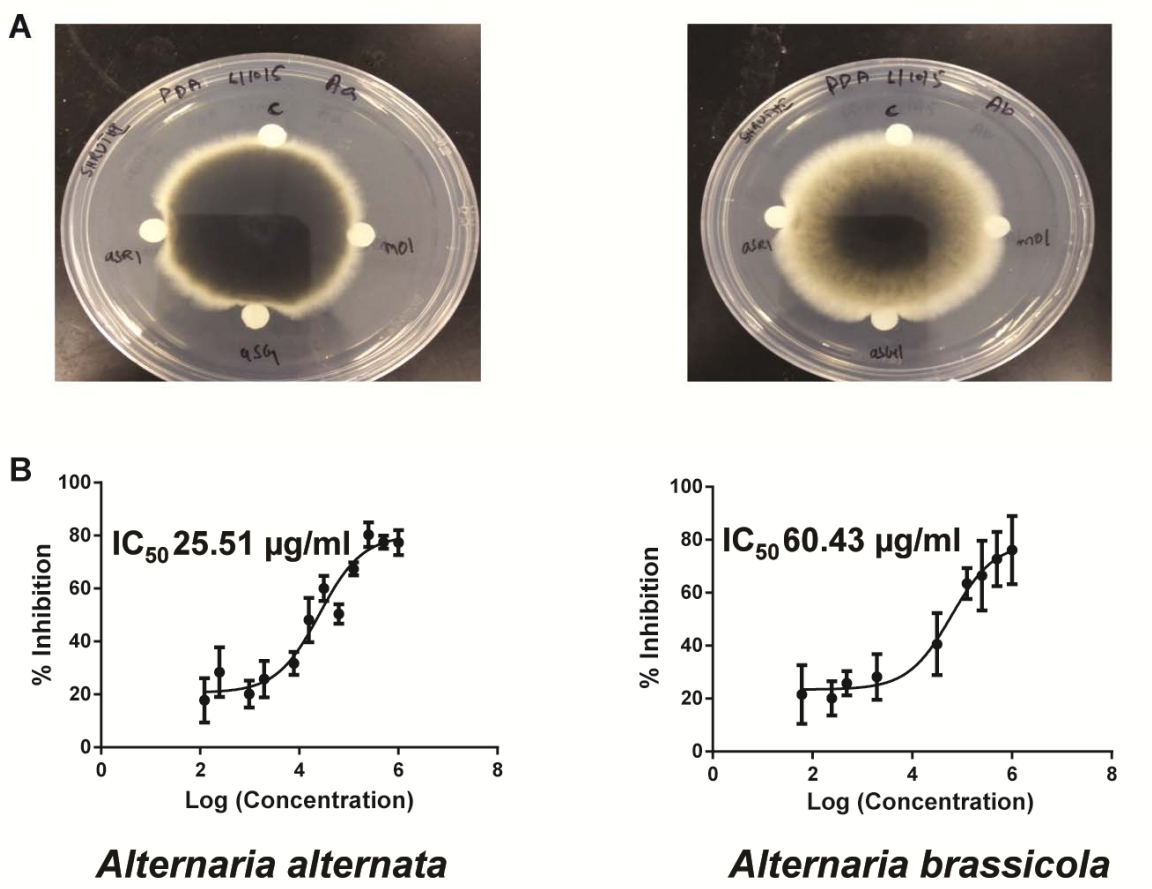


Figure 4.5. Anti-fungal activity of morintide mO1. (A) Formation of arc-shaped inhibition zones at the growing tips of fungal mycelia indicates the susceptibility of A. alternata and A. brassicola to mO1. (B) dose-response curves generated from the microbroth dilution assay were used to calculate the IC_{50} values.

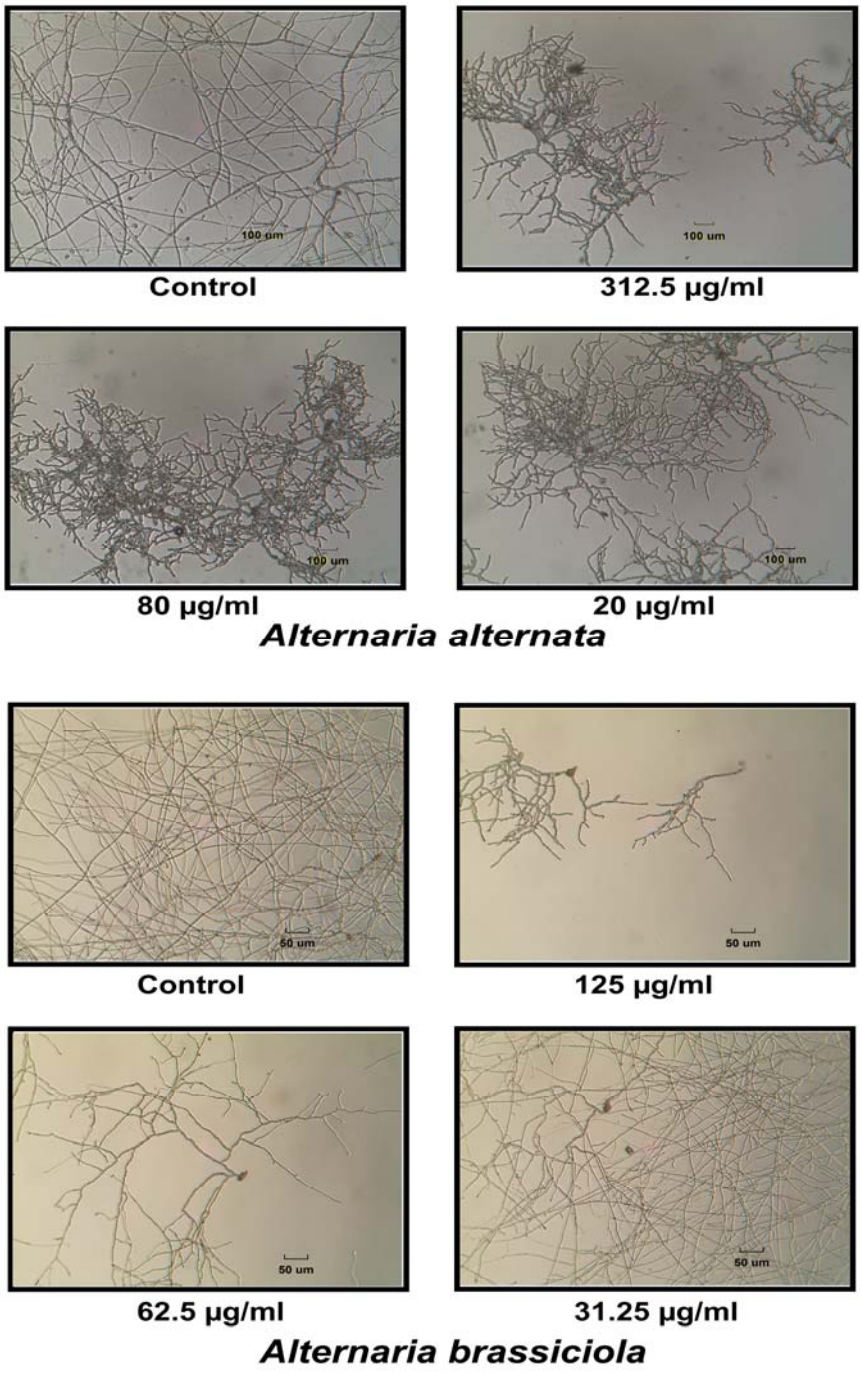


Figure 4.6. Hyphal growth inhibition assay of morintide mO1. Treatment of *A. alternata* and *A. brassiciola* spores with mO1 resulted in the formation of thicker and less branched hyphae than the control well indicating that mO1 could inhibit hyphal elongation consequently retarding the growth of fungi.

2.5. Heat and proteolytic stability

Traditional medicine requires heating therapeutic decoctions to 100°C for prolonged time periods and peptides are usually considered unstable to thermal degradation. In addition, bioavailability of therapeutic agents is significantly reduced due to susceptibility to proteolytic degradation. To study if mO1 could be the primary bioactive principle in decoctions of *M. oleifera*, stability of mO1 against thermal and enzymatic degradation was assessed. Morintide, mO1 was incubated at 100°C for 1 h or with pepsin and carboxypeptidase A for 6 h. By computing the area under the curve from RP-UPLC chromatograms it was observed that more than 95% of the peptide remained intact after heating for 1 h and on treatment with proteolytic enzymes (Fig. 4.7). More than 50% of linear control peptide RV-14 was degraded within 1 h of heating at 100°C while WV-14 was completely digested in 4 h by carboxypeptidase A and only 50% of WV-14 remained intact after 6 h of pepsin treatment. Thus, this data establishes the high stability of mO1 against thermal and enzymatic degradation and suggests that mO1 could be the active principle contributing to the medicinal value of *M. oleifera*.

Figure 4.7. Thermal and enzymatic stability assay. Morintide mO1 was stable after (A) heating at 100°C for 1 h while control peptide (B) RV-14 was degraded within 1 h. Enzymatic stability of mO1 against (A) carboxypeptidase and (B) pepsin was assayed upto 6 h. Control peptide WV-14 was degraded in the presence of (C) carboxypeptidase and (D) pepsin in 6 h.

2.6. Biosynthesis of morintides

A total of 78,561 transcripts with contig length between 201-22191 bp were obtained from transcriptome of RNA extracted from fresh *M. oleifera* leaves. These transcripts were analyzed using in-house software, Protein Analyzer 1.0 to look for conserved cysteine motifs. Three precursor sequences, designated *mo1, mo2* and *mo3* were obtained. Sequence comparison of precursors of *mo1* and *mo2* revealed that the signal peptide is absolutely conserved, the mature domain comprises a variation of one amino acid residue and the C-terminal tail varies at three positions, namely, Asp-67, Glu-71 and Gly-76 of *mo1* are replaced by Ala-67, Gly-71 and Ser-76 in *mo2*. The precursor sequence, *mo3* was detected only at the transcriptome level and contained a C-terminal truncation. The mature domain was different from mO1 by six amino acid residues: Gln-2, Asn-8, Ala-10, Ala-12, Ser-25, and Glu-28 in mO1 were replaced by Gln-2, Gly-8, Thr-10, Pro-12, Thr-25 and Ala-28 in mO3, respectively. These precursor sequences were aligned with previously reported 8C-hevein-like peptides including hevein from *H. brasiliensis [88]* and Pn-AMP1 from *Pharbitis nil* [47] (Fig. 4.8). From sequence alignment it was observed that morintides share similar precursor organization as other 8C-hevein-like peptides and comprise three domains, a 20-amino acid long ER signal peptide followed by a 44-amino acid long mature CRP domain and a short 15 amino acid-long C-terminal tail. The presence of an ER signal peptide indicates that morintides are ribosomally synthesized peptides [90] and their biosynthesis is through the secretory pathway [91]. A striking feature that distinguishes the precursors of mO1 from hevein is the significantly short C-tail

(15 residues). Hevein comprises a long C-terminal tail, 142 amino acid residues in length which encodes for a different bioactive protein. These bioactive proteins are usually lectins with a Barwin domain capable of binding to tetrameric N-acetylglucosamine [88]. This phenomenon leading to the truncation of the C-terminal tail may have occurred by divergent evolution since *M. oleifera* and *H. brasiliensis* belong to orders Brassicales and Malphigiales respectively which are two distinct clades within the plant kingdom.

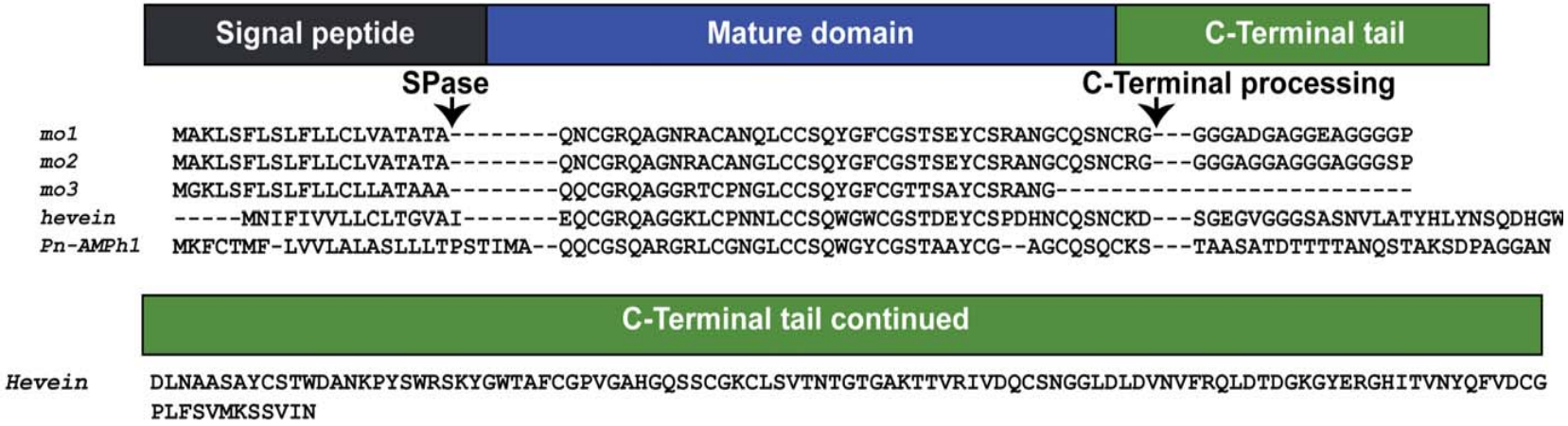


Figure 4.8. Sequence alignment of putative morintide precursors and 8C-hevein-like peptides. Precursor organization is conserved and comprises a signal peptide domain, a mature CRP domain and a C-terminal tail. The C-terminal tail of hevein is significantly longer than morintides, mO1 and mO2 as it encodes for a bioactive protein comprising the Barwin domain.

3. Discussion

In this study, three novel 8C-hevein-like peptides, collectively named morintides were isolated, sequenced and characterized by proteomic, transcriptomic and biochemical methods. Morintides, mO1 and mO2 are 44 amino acid residues in length, were detected at proteomic and transcriptomic level and differ by one amino acid residue: Gln-15 in mO1 is replaced by Gly-15 in mO2 while mO3 was 35 amino acid residues in length and detected only at the transcriptome level. Morintides are homologous to hevein and other 8C-hevein-like peptides like Pn-AMPs [47] and Fa-AMPs [51]. Structural analysis of mO1 revealed that three disulfide bonds were arranged to form the cystine knot motif and the fourth disulfide bond was located at the C-terminal. Morintide mO1 inhibited the growth of *A. alternata* and *A. brassiciola* in the disc diffusion assay with an IC_{50} of 25.5 µg/ml to 60.43 µg/ml, respectively. Transcriptomic analysis revealed that morintide precursors consist of three domains: a signal peptide domain, a mature CRP domain and a short C-terminal tail suggesting that their bioprocessed through the secretory pathway.

3.1. Sequence Comparison of Morintides

Sequence comparison of morintides with reported 8C-hevein-like peptides revealed that, in addition to the eight conserved cysteine residues, three glutamine (Gln-6, Gln-20, Gln-36), two serine (Ser-19, Ser-37) and Leu-16 were absolutely conserved in all 8C-hevein-like peptides.

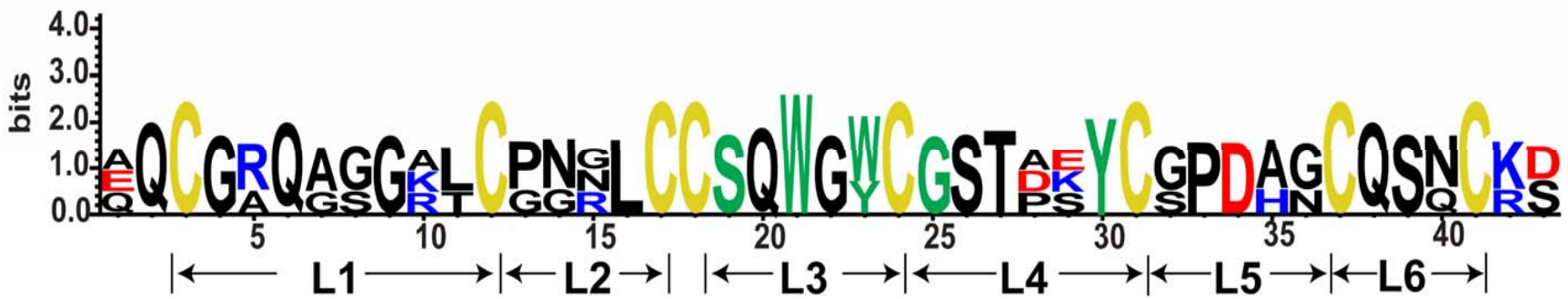


Figure 4.9. Sequence logo of aligned mature domains of 8C-hevein-like peptides. Based on the cysteine pattern the sequence can be divided into six loops. None of the loops are absolutely conserved indicating that there is high molecular diversity among 8C-hevein-like peptides. Loops 3 and 4 comprise the conserved chitin-binding domain which is highlighted in green.

Based on the cysteine-pattern the peptide sequences can be divided into six loops. The length of the loops is conserved with variation observed only in loop 5 of Fa-AMPs and Pn-AMPs that comprise three residues as opposed to five residues seen in hevein and mO1. The N-terminal of all 8C-hevein-like peptides except Fa-AMPs comprises a glutamate residue which is spontaneously converted to pyroglutamic acid conferring enhanced proteolytic stability against exopeptidases. The C-terminus usually comprises a positively charged residue followed by glycine, serine or aspartate residues in mO1, Pn-AMP1 and hevein respectively.

A sequence logo of the aligned morintides and 8C-hevein-like peptides was used to illustrate the similarity and frequency of occurrence of amino acid residues at each position (Fig. 4.9). The height of the stack indicates the overall sequence conservation while the frequency of an amino acid at a particular position is indicated by the height of the symbols. From the sequence logo it is clear that the extent of molecular diversity is high among 8C-hevein-like peptides as none of the loops show absolute sequence conservation. Loops 3 and 4 comprise the chitin-binding domain where a serine and glutamine are absolutely conserved. The aromatic amino acid residues in loop 3 are highly variable while a tyrosine residue in loop 4 is absolutely conserved. It is speculated that such variations in the primary sequences of closely-related peptides arise by divergent evolution and lead to functional diversity in the peptide family.

Similar modulation of function by natural diversity was observed in 10C-hevein-like peptides from *Triticum kiharae* where mutation of Ser to Gly in the chitin-binding domain reduced binding affinity to chitin but resulted in a gain of function. These natural variants could bind to fungal metalloproteases and inhibit the degradation of plant chitinases resulting in resistance to fungal infections in *Triticum kiharae* [101].

3.2. Biosynthesis of morintides

Transcriptome data analysis revealed that morintides are expressed as a three-domain precursor comprised of a signal peptide, a mature domain and a C-terminal tail and are processed via the secretory pathway [91]. In the secretory pathway, ribosomally synthesized peptides are directed to the endoplasmic reticulum where the signal peptide is cleaved by signal peptidase (SPase1) followed by cleavage of the mature domain from the C-terminal tail. Sequence alignment of morintide precursors with other 6C-, 8C- and 10C-hevein-like peptides showed that this precursor organization is conserved throughout the hevein-like peptide family with variations primarily in C-terminal domain length (Fig. 4.10). Sequence alignment revealed that the 15-amino acid long C-terminal domain of morintide is rich in glycine residues and differs from hevein and 10C-hevein-like peptides which have a significantly longer C-terminal tail (142-200 amino acid residues) coding for proteins with a Barwin domain or a glycine or serine-rich hinge region [170] followed by the catalytic domain of Class I chitinases [171]. The glycine-rich nature of the C-tail of morintides suggests that

it could be a remnant of the hinge-region in chitinases and morintides could be truncated chitinases.

The hevein-like peptide family is an interesting family of CRPs as it could be divided into three sub-classes making it an ideal family to study genetic diversity. Genetic divergence within the plant phyla result from mutations in the mature domain which also lead to functional diversification. For example, the signal peptide and C-terminal tail of morintide mO1 and mO2 precursors are highly homologous, but a point mutation in the mature domain results in replacement of Gln-15 in mO1 to Gly-15 in mO2. Similar genetic diversity has also been observed in other CRPs like cliotides cT4-cT12 of the cyclotide family isolated from *Clitoria ternatea* [93], allotides of cystine knot α-amylase family isolated from *Allamanda cathartica* [158] and α- and β-hordothionin of the thionin family isolated from *Hordeum vulgare* [172].

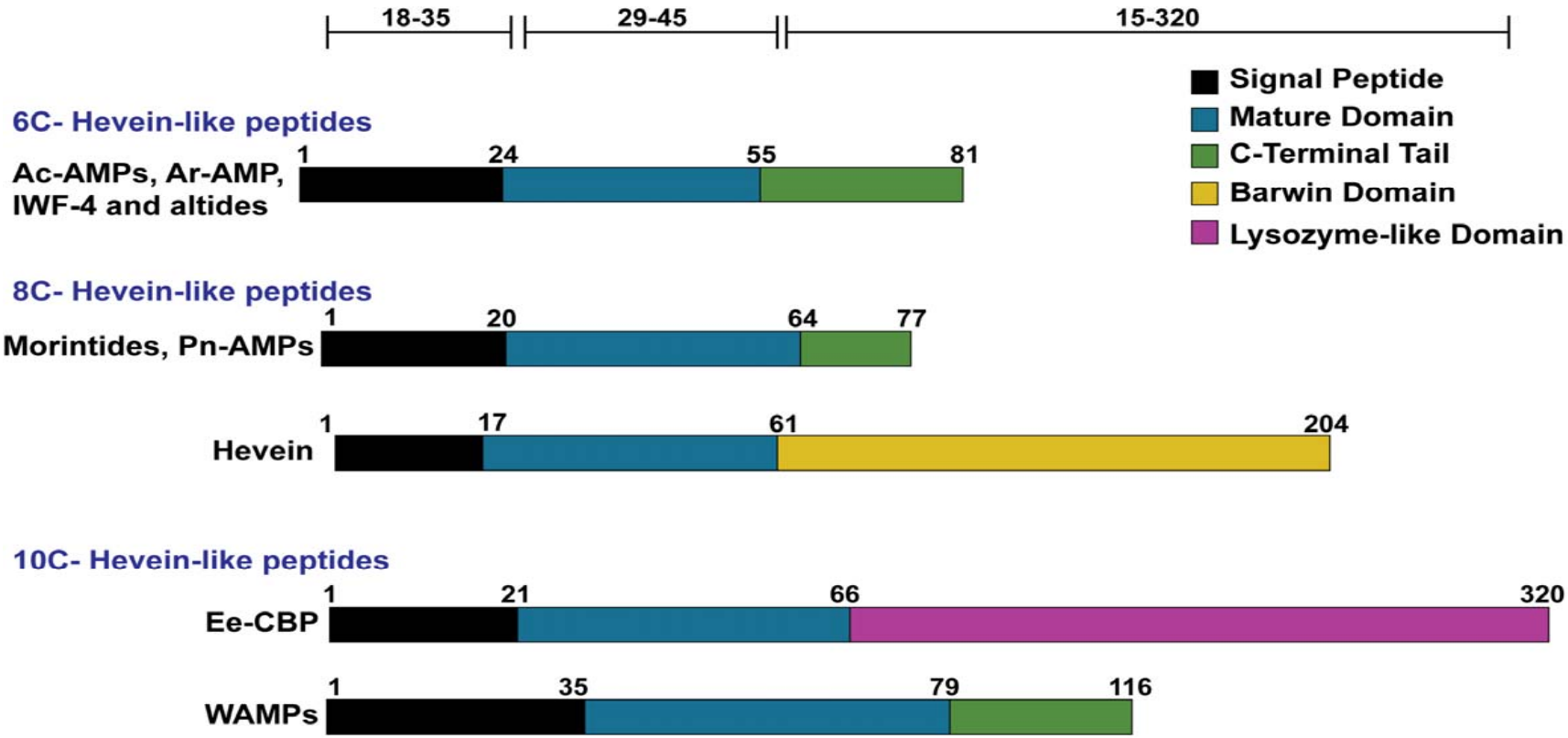


Figure 4.10. Representation of precursor arrangement in the hevein-like peptide family. The precursor organization is conserved among 6C-hevein-like peptides, 8C- and 10C-hevein-like peptides including hevein and Ee-CBP have long C-terminal tails coding for proteins with a Barwin domain or a chitinase domain. Morintides and WAMP-1a are exceptions of the 8C- and 10C-sub-classes and comprise short C-terminal tails with no homology to any known proteins in the NCBI conserved domains database.

A phylogenetic tree was constructed using the precursor sequences of morintides and nine other hevein-like peptides to study their evolutionary relationship (Fig. 4.11). Data analysis showed that morintides can be grouped as a separate cluster within the 8C-hevein-like peptide family, attributed by the significantly short C-terminal tail of morintides (15 amino acid residues) when compared to hevein (142 amino acid residues). The precursor structure of hevein is a molecular skeleton of hololectins like class-I/IV chitinases and comprises a chitin-binding domain followed by a long C-terminal chitinase domain [70]. A similar phenomenon was observed among 10C- hevein-like peptides where the C-terminal tail of WAMP-1 and WAMP-2 from *T. Kiharae* seeds lacks the chitinase domain and [89] is significantly shorter than Ee-CBP from *E. europaeus* [70]. Andreev *et al*. speculated that *wamp* genes were “remnants” of chitinase genes and were either formed by a frame shift deletion of a portion of the DNA sequence coding for the chitinase domain or by alternative splicing of chitinase pre-mRNA followed by reverse transcription and genomic integration. These “truncated” genes were then selected by evolution as they coded for anti-fungal peptides which conferred the plant with resistance to fungal infections [89]. The distinct nature of *wamp* genes also suggested that excision of the chitinase domain from class-I chitinases led to the evolution of class-II chitinases that lack this domain. It can be speculated that a similar phenomenon occurs in the evolution of 8C-hevein-like peptides which will be discussed in detail in chapter 5.

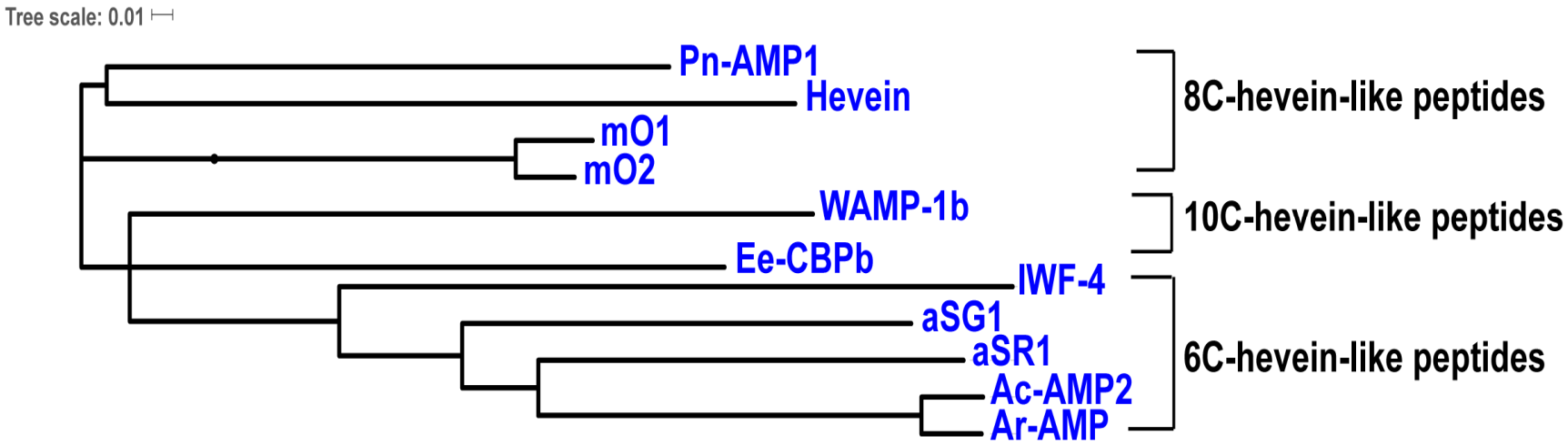


Figure 4.11. Phylogenetic tree of reported hevein-like peptides. 6C-, 8C- and 10C-hevein-like peptides are grouped in separate clusters. Morintides form a different cluster within the 8C-hevein-like peptide cluster due to their significantly short C-terminal tails.

3.3. Anti-fungal activity of morintides

Out of seven fungal strains tested, two strains, namely, *A. alternata, A. brassiciola* were susceptible to treatment with mO1 as indicated by arc-shaped inhibition zones around growing mycelia. The IC_{50} calculated from dose-response curves was in the range of 25.5 µg/ml to 60.43 µg/ml and is similar to other 8C-hevein-like peptides, Pn-AMPs and Fa-AMPs but more potent than hevein [47, 51]. Morphological changes like swollen hyphal tips, stunted growth and less-branched hyphae were observed on treatment of fungal spores with different concentrations of mO1. Similar changes in morphology have been observed with other 8C-hevein-like peptides including hevein [38] and Pn-AMPs [47]. However, the mechanism of action against fungi is still unclear. It was speculated that hevein inhibits growth of fungal pathogens by interrupting the steady-state model of hyphal elongation similar to 6C-hevein-like peptides. However, Pn-AMPs could also inhibit non-chitin containing fungi suggesting that there could be a different mechanism involved. Fluorescence studies on fungi treated with labelled Pn-AMPs indicated that the small size and highly basic PI of Pn-AMPs allows them to easily penetrate through the fungal cell wall. Once they reach the plasma membrane alterations in membrane polarity could cause leakage of cytoplasmic material subsequently inhibiting fungal growth [47, 52, 99]. The mechanism of action of morintides remains to be elucidated; however, the anti-fungal activity of morintides can be attributed to their chitin-binding property as they have neutral PI values.

3.4. Comparison of 6C- and 8C-hevein-like peptides

An intriguing feature of the hevein-like peptide family is the molecular diversity leading to the existence of three sub-classes with distinct cysteine content. Since we have studied two of these sub-classes we could gain further insight into their evolution and functional diversity. The 6C-hevein-like peptides are considered “truncated” versions of 8C-hevein-like peptides that comprise a C-terminal truncation of about 10 amino acid residues including two cysteines [45, 46, 50]. The 6C-hevein-like peptides are 29-30 amino acid residues in length and based on the cysteine-spacing pattern can be divided into four loops. The loop lengths are conserved with four to six amino acid residues per loop. In the 40-45 amino acid long 8C-hevein-like peptides, the sequence can be divided into six loops where loop 2 is significantly longer than loop 2 of 6C-hevein-like peptides and comprises eight amino acid residues. Loops 3 and 4 of 6C- and 8C-hevein-like peptides comprise the chitin-binding domain.

A sequence logo of the chitin-binding domains of 6C- and 8C-hevein-like peptides sheds light on the sequence conservation of this domain within each sub-class (Fig. 4.12). The overall sequence of the chitin-binding domain is conserved with variations only in the aromatic amino acid residues following the conserved Ser residue.

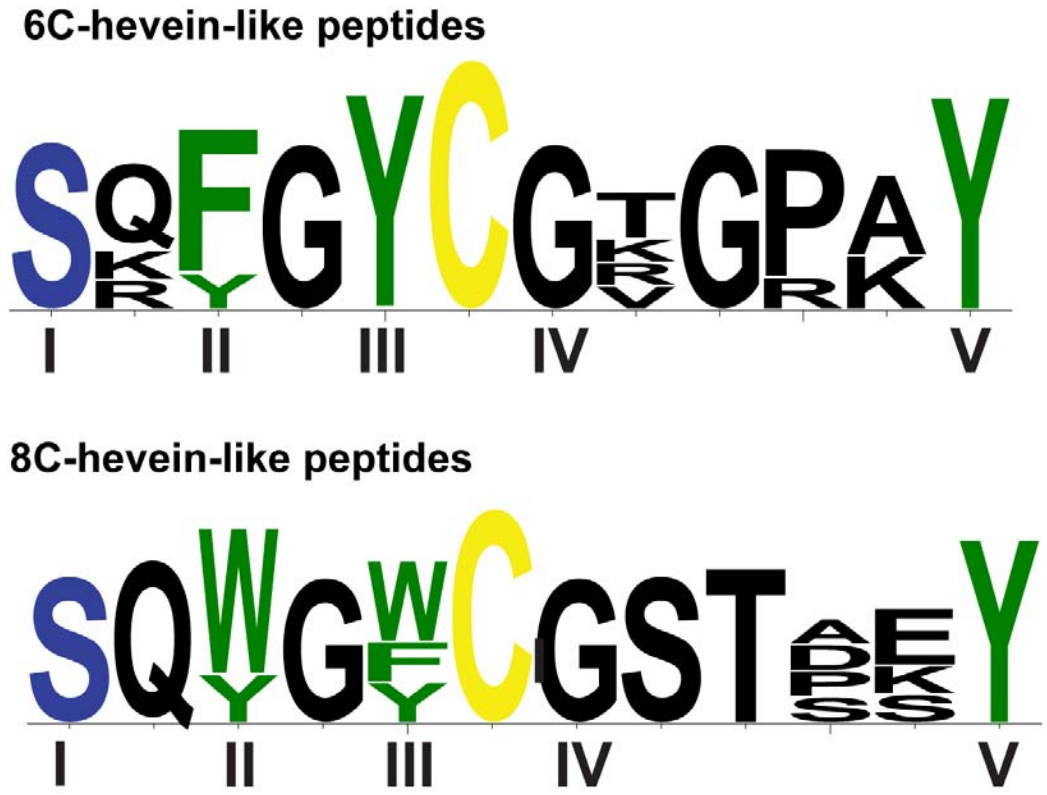


Figure 4.12. Comparison of chitin-binding domain of 6C- and 8C-hevein-like peptides. The Ser residue at position I, Gly at position IV and Tyr at position V are absolutely conserved. In 6C-hevein-like peptides Phe is preferred at position II while Tyr at position III is highly conserved. The aromatic amino acid residues at position II and III are variable in 8C-hevein-like peptides.

In 6C-hevein-like peptides a higher occurrence of Tyr and Phe is seen at position II of the chitin-binding domain while positions III and IV comprise an absolutely conserved Tyr residue. However, in 8C-hevein-like peptides, Trp and Tyr are preferred in positions II and III with an absolute conservation of Tyr residue at position IV. Despite these variations in amino acid sequences, altides and hevein bind with similar affinity to chitin oligosaccharides indicating that variations in the aromatic amino acid residues do not significantly affect binding affinity. The 8C-hevein-like peptides contain two additional loops due to the two additional cysteine residues. The C-terminal of altides and morintides comprise a Gly residue however, other 8C-hevein-like peptides comprise positively charged residues like Lys and Arg [32, 47, 51].

Sequence variations among peptides of the same family are known to increase functional diversity. Since one of the primary functions of hevein-like peptides in plants is defense against fungal infections, we studied the anti-fungal activity of altides and morintides against seven phytopathogenic fungi. Out of the seven strains tested, both altides and morintides could inhibit *A. alternata* and *A. brassiciola* but inhibitory activity against the remaining five strains was dependent on the peptides and strain being tested (Table. 4.2). For example, altide aSG1 could inhibit *V. dahilae* while mO1 showed no inhibitory activity against this strain. On comparing the potency of altides and morintides against *A. alternata* and *A. brassiciola* it was evident that altides were more potent inhibitors of these fungi. Thus, it can be speculated that variations in peptide sequence may have a role in broadening the spectrum of anti-fungal activity of hevein-like peptides.

Table 4.2. Comparison of anti-fungal activity of aSG1, aSR1 and mO1. *A. alternata* and *A. brassiciola* were sensitive to all three peptides whereas fungal growth inhibition of these peptides against the remaining strains was more strain-specific. "+" indicates inhibitory activity and "-" indicates no fungal growth inhibition.

Fungal strain	Anti-fungal activity		
	aSG1	aSR1	mO1
Alternaria alternata	+	+	+
Alternaria brassicola	+	+	+
Aspergillus niger	-	-	-
Curvularia lunata	-	+	-
Fusarium oxysporum	+	-	-
Rhizoctonia solani	-	+	-
Verticulum dahilae	+	-	-

Insights into the precursor organization of altides and morintides can help shed light on their evolutionary relationship. Comparison of the precursor sequences of altides and morintides shows that the precursor organization is conserved and comprises a signal peptide, a mature CRP domain and a C-terminal tail. The signal peptide of altides is four residues longer than morintides while the C-terminal tail is longer by 10 residues. It has previously been reported that the C-terminal tail may comprise sorting signals that can decide the fate of the peptide after secretion from the Golgi apparatus [173-176]. Thus we performed a BLAST analysis of the C-terminal tails to identify any sorting signals within the peptide sequence. However, no sequence homology with any conserved domains in the database was obtained thereby making it difficult to decipher their functional significance. While hevein is obtained from a chimeric precursor that encodes for another bioactive protein with a Barwin domain the short C-tail of morintides is similar to the 6C-hevein-like peptides [37]. Thus, it can be assumed that morintides and altides share a close evolutionary relationship although they are isolated from two different plant species. It is possible that morintides and altides arose by divergent evolution from hevein as their short length and integral role in defense against fungal infections were more energetically favorable for plants. This evolutionary relationship is discussed in more detail in Chapter 5.

CHAPTER 5

EST-Based *in silico* Identification of Hevein-like Peptides in Plants

1. Introduction

Bioactive peptides, commonly found in all organisms from bacteria to plants and animals, are highly diverse in their structure and display considerable functional diversity [177]. These bioactive peptides are interesting candidates to study the interplay between natural diversity and evolution in creating an arsenal of peptides with diverse biological functions [178]. The conventional approach of identifying putative bioactive peptides is bio-guided fractionation, where the crude plant extract is fractionated by HPLC and each fraction is screened for bio-activity [179]. Although this method has been used for decades in the discovery of novel peptide biologics, it is time-consuming and laborious. Thus, there is a need to develop a robust and high-throughput method by which novel peptides can be identified in plants at a rapid rate. Recently, a lot of progress has been made in the field of genomics, proteomics and transcriptomics providing a vast expanse of data and new avenues to discover bioactive peptides. So far, there are more than 73 million expressed sequences tags (ESTs) in the NCBI database (dbEST) which is expanding substantially each year [180]. A combination of bioinformatics and computational algorithms provides a platform to virtually screen a variety of plant genomes at an exponential rate as compared to the conventional approach [181]. This *in silico*-based approach has been

previously used for the identification of bacteriocins [182] and lantibiotics [183]. However, thus far, this method has not been widely used for screening CRPs.

The hevein-like peptide family is an interesting family of CRPs due to its natural sequence diversity allowing it to be divided into three sub-classes, namely, 6C-, 8C- and 10C-hevein-like peptides. The molecular diversity of hevein-like peptides and their role in plant defense against fungal pathogens makes them interesting candidates to study their evolution and distribution in plants. Thus far, ten 6C-hevein-like peptides, five 8C-hevein-like peptides and five 10C-hevein-like peptides have been isolated and characterized from Amaranthaceae (10 peptides), Euphorbiaceae (1 peptide), Polygonaceae (2 peptides), Convolvulaceae (2 peptides), Poaceae (2 peptides), Celastraceae (1 peptide) and Eucommiaceae (2 peptides) (Fig. 5.1). Although hevein-like peptides are widely distributed in plants, knowledge of their origin, evolution and distribution is still limited. Thus, the aim of this chapter is to gain insights into the expression, distribution and evolution of hevein-like peptides across different plant families using a high-throughput bioinformatics-guided approach.

Taking advantage of the vast expanse of the EST-based database a BLAST search was performed using the reported hevein-like peptides as queries to identify their potential homologs. The resulting hits were filtered based on characteristics of hevein-like peptides and a total of 385 novel hevein-like peptides were identified from 124 different plant families.

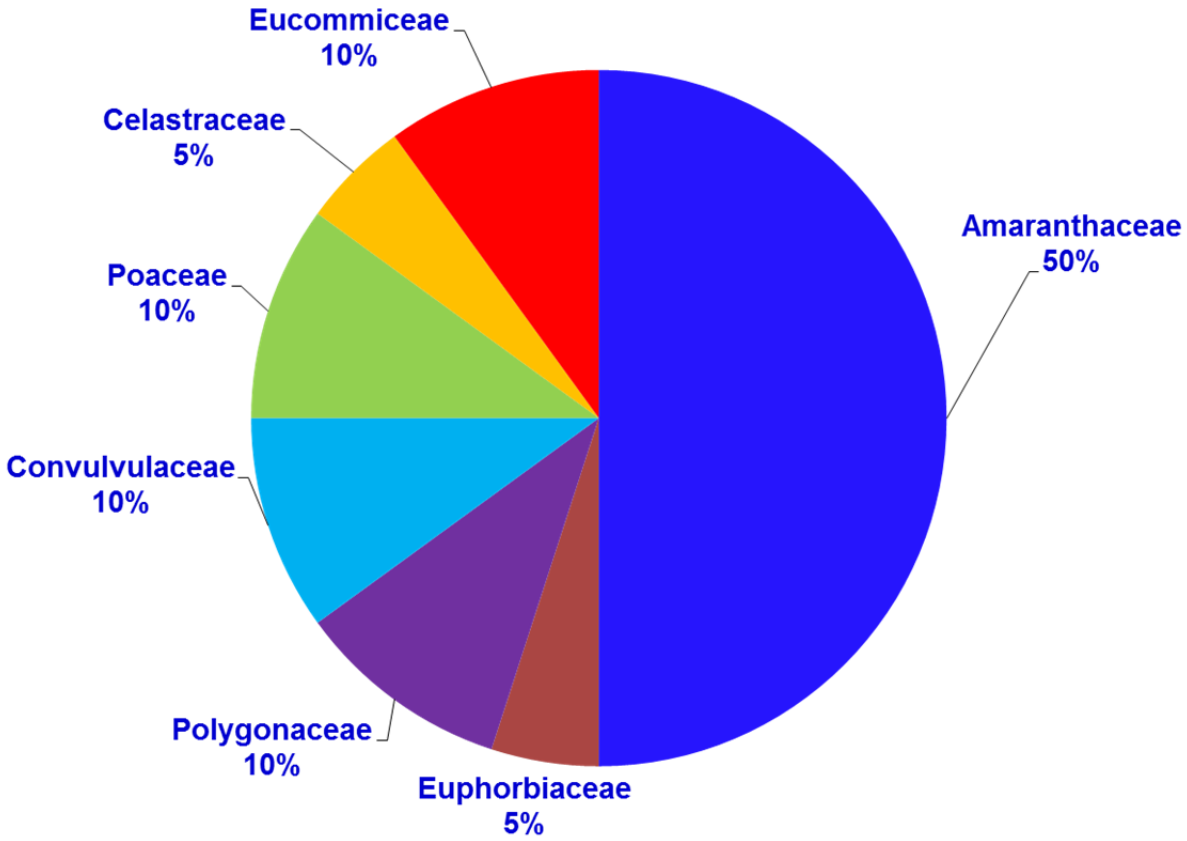


Figure 5.1. Distribution of reported hevein-like peptides in plants. A total of 20 hevein-like peptides have been reported to date and they are distributed in 7 different plant families. The Amaranthaceae family has the highest number of hevein-like peptides and accounts for 50% of the reported peptides.

A phylogenetic tree was constructed to understand the evolutionary pattern of hevein-like peptides and their biosynthesis pathway was explored. Sequence alignment revealed that the precursor organization is conserved throughout the hevein-like peptide family. The three-domain precursor is comprised of a signal peptide, a mature CRP domain and a C-terminal tail. In 8C- and 10C-hevein-like peptides, chimeric precursors were observed, where the C-terminal tails coded for other bioactive proteins like Barwin and rare lipoprotein-like double-psi beta barrel (RlpA-DPBB-1) domain. Together our data sheds light on the sequence diversity, distribution, evolution and biosynthesis of the hevein-like peptide family.

2. Results

2.1. EST-based discovery of hevein-like peptides in plants

Two major databases, namely NCBI (www.ncbi.nlm.nih.gov) and OneKp (www.**onekp**.com) were used to search for novel hevein-like peptide sequences in plants. The reported peptide sequences of 6C-, 8C- and 10C-hevein-like peptides were used as queries to search against the EST nucleotide database comprising nucleotide sequences translated in six open reading frames (ORFs). Initially, a total of 626 peptide sequences were obtained from both databases of which 24 were 6C-hevein-like peptides, 565 were 8C-hevein-like peptides and 37 were 10C-hevein-like peptides which were pooled into different datasets. There was a considerable overlap in the sequences obtained from both databases, thus, replicate sequences were filtered manually. After the first round of filtering a total of 161 sequences with identical precursor sequences from the same plant or from different species of the same genus were deleted. Of these deleted sequences, four were 6C-hevein-like peptides, 150 were 8C-hevein-like peptides and seven were 10C-hevein-like peptides. After filtering the data, a total of 465 putative hevein-like peptide sequences remained (Fig. 5.2 A). However, the likelihood of class I chitinase sequences appearing in the hevein-like peptide datasets is high since the chitin-binding domain of class I chitinases is essentially a hevein-like peptide linked to a catalytic domain by a hinge region [60, 96] (Fig. 5.2 B).

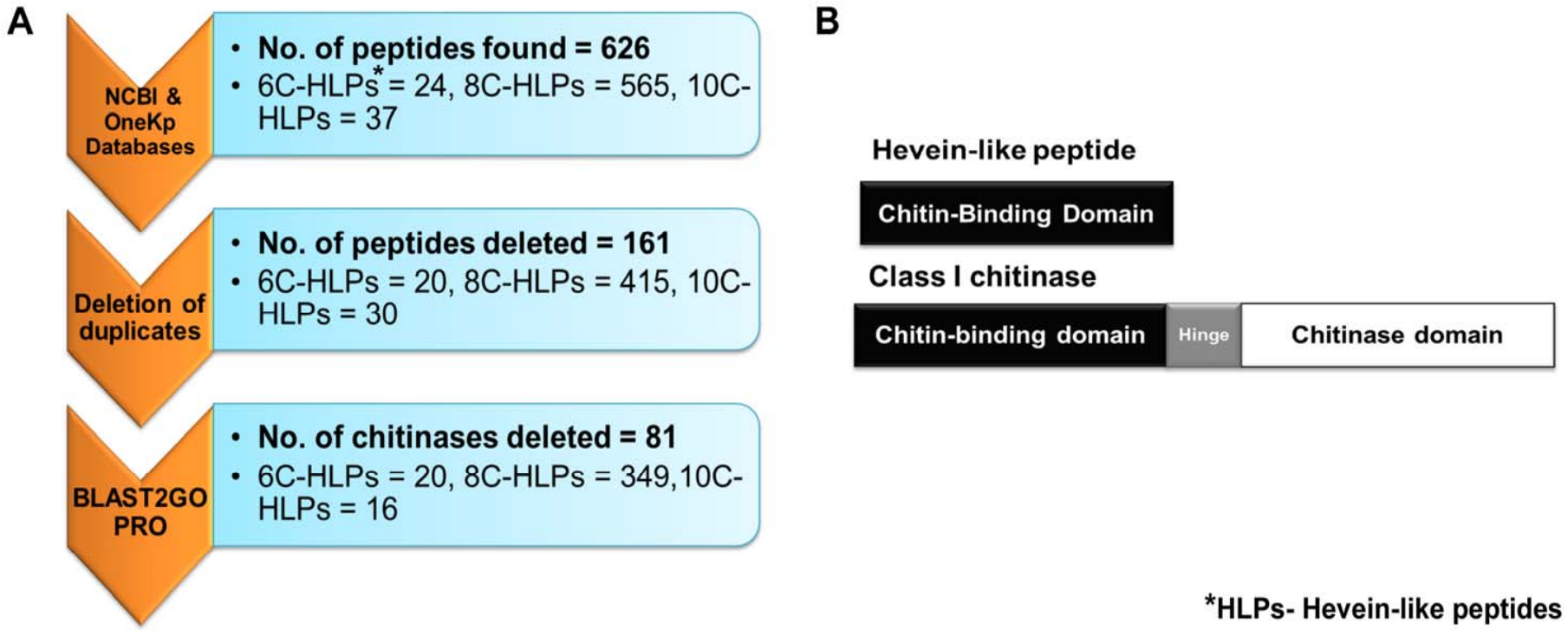


Figure 5.2. EST-based screening of hevein-like peptides. (A) Flow chart of data mining of putative hevein-like peptides from NCBI and OneKp databases. (B) Comparison of organization of hevein-like peptides and class I chitinases. The chitin-binding domain of class I chitinases is essentially a hevein-like peptide.

Therefore, the 6C-, 8C- and 10C-hevein-like peptide datasets were pooled and filtered a second time using the Blast2GO PRO software to eliminate chitinase sequences [151]. After scanning all the datasets, a total of 81 chitinase sequences were deleted. Thus, after two rounds of filtering, we obtained 385 putative hevein-like peptides, that is, 20, 6C-hevein-like peptides, 349, 8C-hevein-like peptides and 16, 10C-hevein-like peptides. This high incidence of hevein-like peptides in plants suggests that they are a widely distributed family of CRPs and makes them interesting candidates to study their origin, distribution and evolutionary relationship.

2.2. Distribution and Occurrence of hevein-like peptides in the plant kingdom

From EST-based data mining of hevein-like peptides, a total of 124 plant families belonging to 52 orders expressing hevein-like peptides were found from NCBI and OneKp databases (Fig. 5.3 A). The 20 novel 6C-hevein-like peptides were distributed in 18 plant species from 4 plant families only comprising dicot angiosperms. These families are derived from orders Asterales, Caryophyllales and Boraginales which are closely-related and grouped under the common clade Asterids. The Amaranthaceae family had the highest number of 6C-hevein-like peptides suggesting that they are an important source of these peptides. The novel 8C-hevein-like peptides were found in 271 plant species from 117 different plant families originating from non-vascular bryophytes (mosses) to more-evolved vascular plants including gymnosperms and angiosperms.

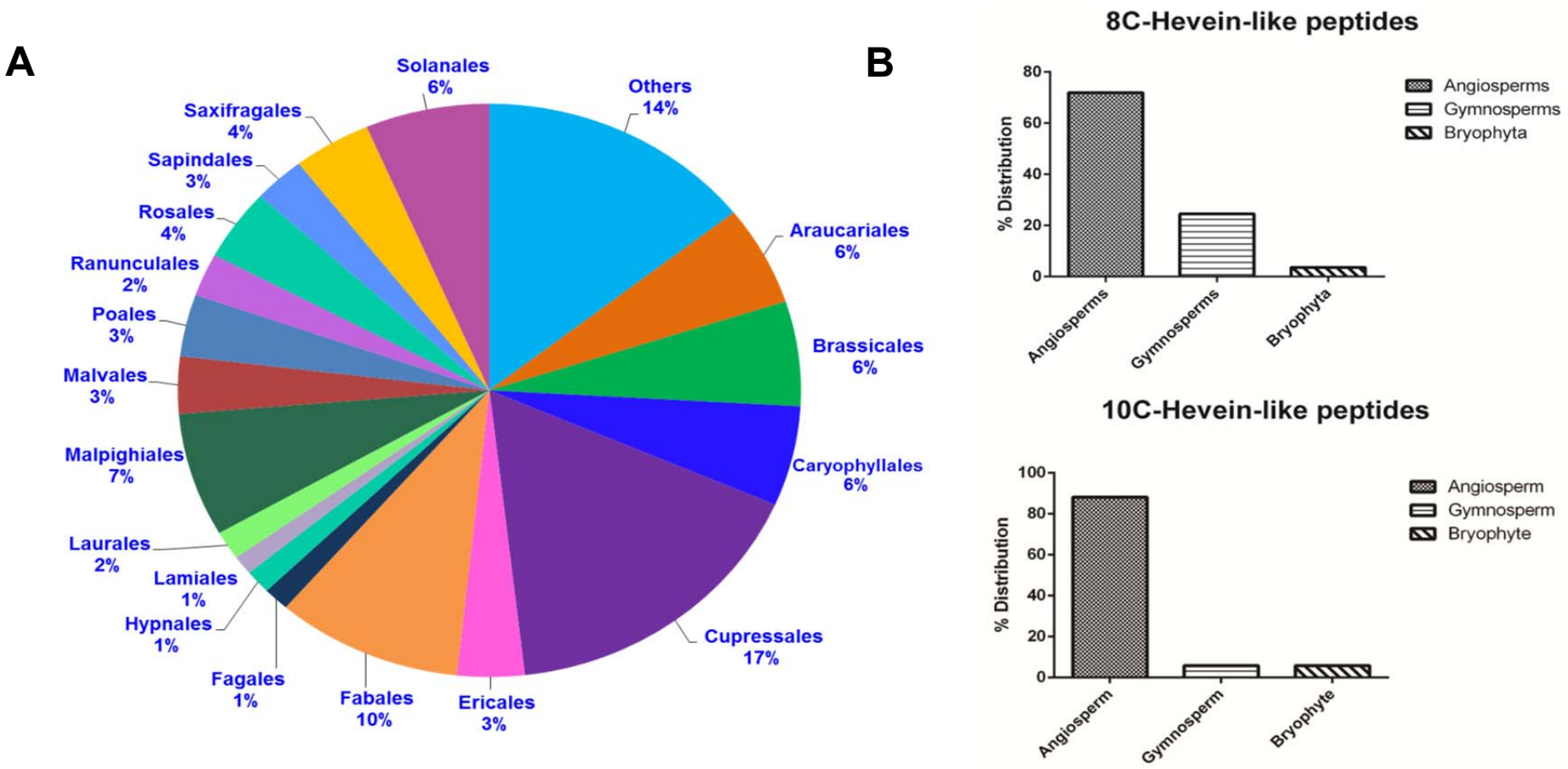


Figure 5.3. Distribution of putative hevein-like peptides in plants. (A) Novel putative hevein-like peptides were distributed in 124 plant families belonging to 52 orders. The order Cupressales comprising gymnosperms had the highest number of hevein-like peptides closely followed by Fabales from angiosperms. (B) 8C- and 10C-hevein-like peptides were distributed among angiosperms, gymnosperms and Bryophyta indicating that they are the most diverse and widespread sub-classes of hevein-like peptides.

A total of 3.6% 8C-hevein-like peptides were found in non-vascular plants, originating from the lineages of the first terrestrial plants on Earth, namely, Bryophyta and Marchantiophyta while 17% were distributed in the order Cupressales from gymnosperms and 10% in Fabales from angiosperms (Fig. 5.3 B). The 16 novel 10C-hevein-like peptides were distributed in 14 species from six different plant families. One sequence each was found in the Plagiochilaceae and Gnetaceae families of Marchantiophyta and gymnosperms, respectively. Majority of the 10C-hevein-like peptides were detected in monocot and dicot angiosperms with 88% in the Poaceae family and 6% in the Rutaceae family belonging to orders Poales and Sapindales, respectively (Fig. 5.3 B). Thus, it can be speculated that there is a close evolutionary relationship between 10C-hevein-like peptides from monocot and dicot angiosperms. Interestingly, 15 hevein-like peptides were found multiple times in more than one plant species. While some sequences were found in plants of the same genus and different species, others were found in plants belonging to the same family (Appendix A).

To further understand the distribution and evolution of hevein-like peptides we constructed cladograms of the putative hevein-like peptide producing plants. The expression of 6C-hevein-like peptides was restricted to eudicots belonging to Asterids and the closely-related, sister clade Caryophyllales [184] (Fig. 5.4 A).

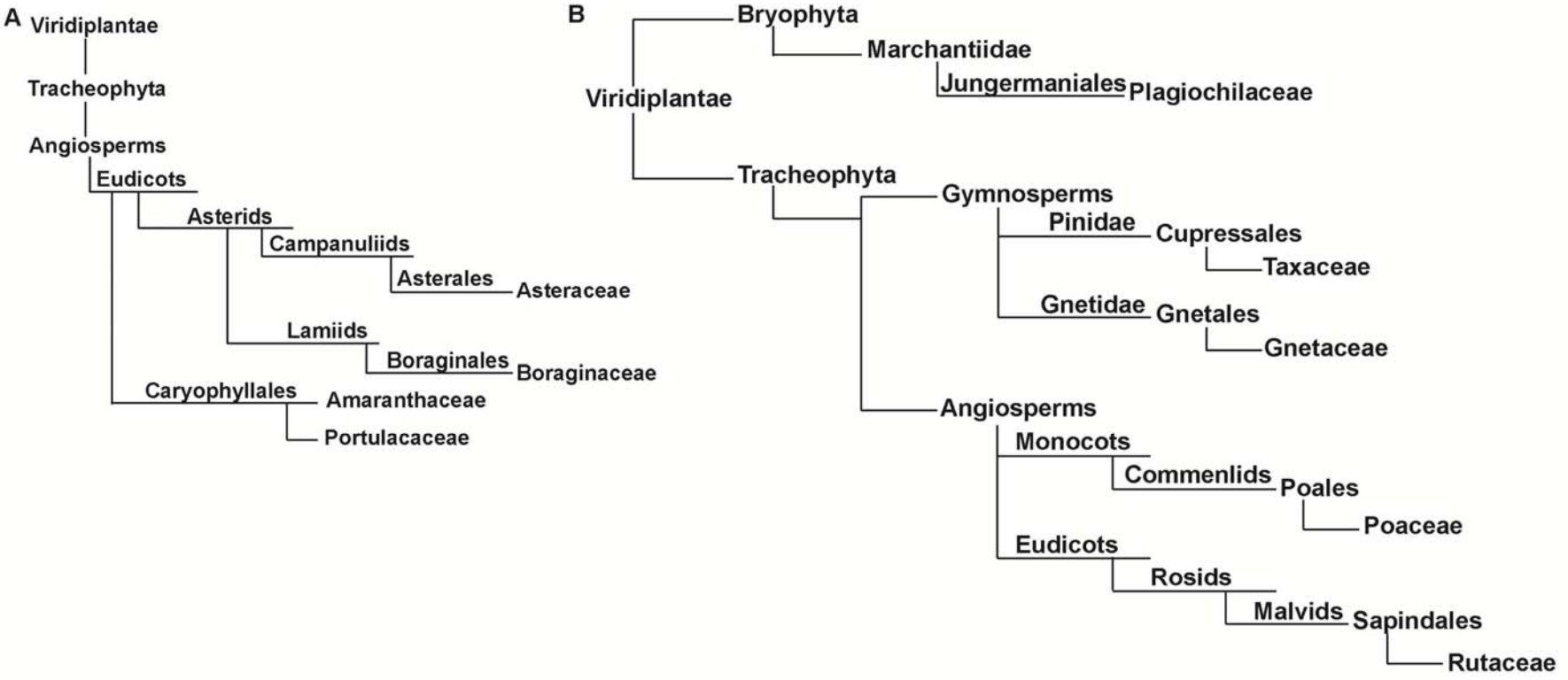


Figure 5.4. Summary cladogram of the distribution of hevein-like peptides. (A) 6C-hevein-like peptides were exclusively found in eudicots belonging to orders Asterids and Caryophyllales. (B) 10C-hevein-like peptides were distributed in Bryophyta and Tracheophyta including Gymnosperms and monocot and eudicot angiosperms.

Within Asterids, 6C-hevein-like peptides occur in Asterales and Boraginales suggesting that closely-related orders, Gentinanales and Solanales may also contain these peptides and examining plant families within these orders maybe a means of discovering new 6C-hevein-like peptides. In contrast to 6C-hevein-like peptides, 10C-hevein-like peptides occurred in Bryophyta and Tracheophyta (Fig. 5.4 B). However, the incidence of 10C-hevein-like peptides in bryophytes and gymnosperms was limited to Marchantiidae and Gnetidae, respectively.

Majority of 10C-hevein-like peptides were prevalent in monocots from the Poaceae family and eudicots from the Rutaceae family. The 8C-hevein-like peptides were widely dispersed in plants and occurred in Bryophyta and Tracheophyta including gymnosperms and angiosperms (Fig. 5.5). Among Bryophytes, 8C-hevein-like peptides were found in the basal grade Marchantiidae and the main clades like Jungermaniidae, Bryidae and Lycopodiidae suggesting that these peptides originated early in the evolution of Bryophytes. In gymnosperms, majority of 8C-hevein-like peptides were observed in the most abundant order Cupressales from the sub-class Pinidae and also in the less common Gnetales and Gingkoales indicating that these peptides are ubiquitous among gymnosperms. These peptides were also observed in ferns from the sub-class Polypodiidae.

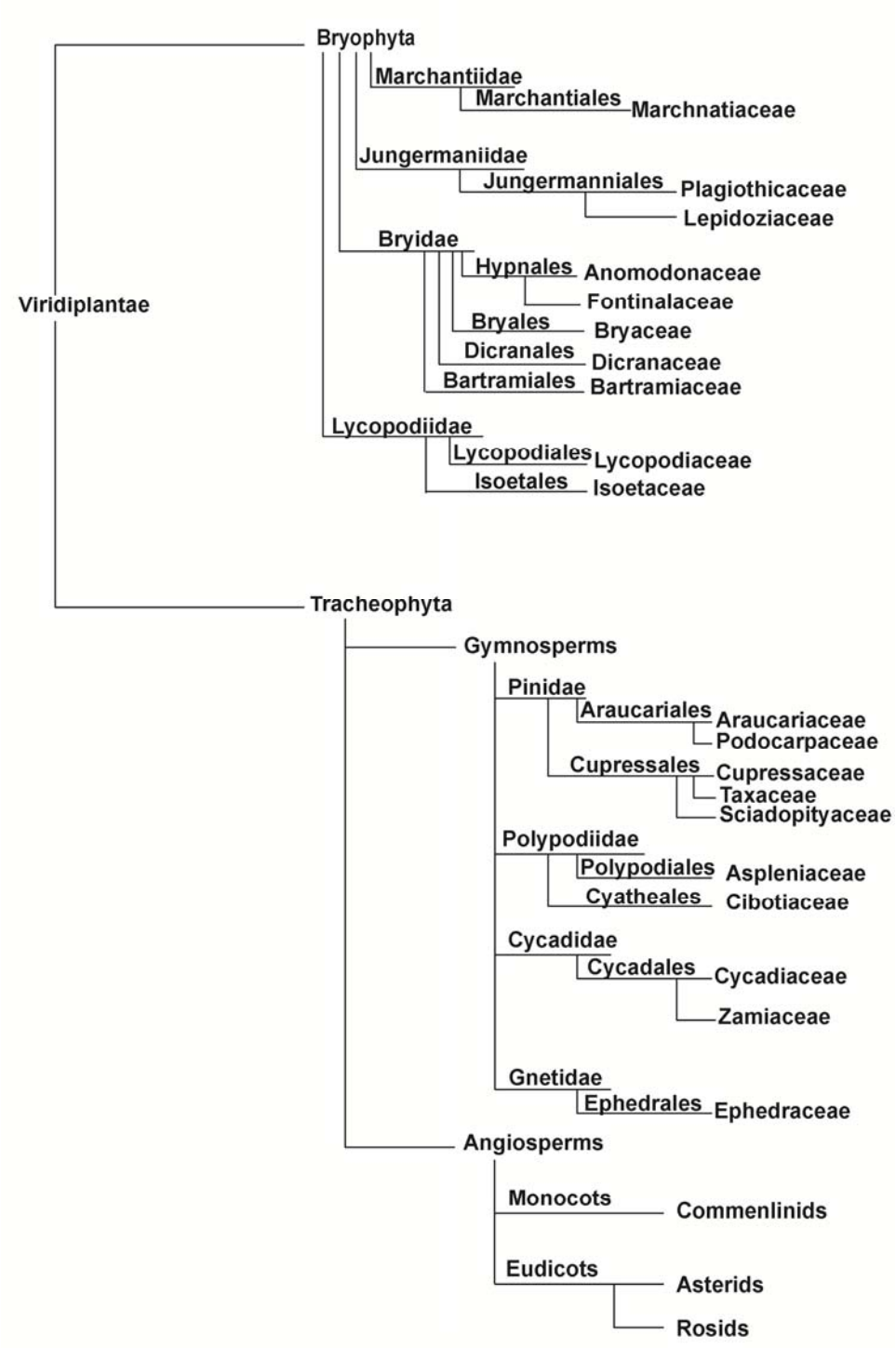


Figure 5.5. Summary cladogram of the distribution of 8C-hevein-like peptides in plants. The 8C-hevein-like peptides are the most diverse and widespread sub-class of the hevein-like peptide family and are distributed in Bryophyta, gymnosperms and angiosperms.

The 8C-hevein-like peptides occurred in orders Poales, Arecales and Asparagles belonging to the Commenlinids clade of monocot angiosperms. These peptides were also distributed in 85 plant families from higher order Rosids and Asterids comprising eudicot angiosperms. This widespread distribution of 8C-hevein-like peptides suggests that they are highly diverse and were probably the first member of the hevein-like peptide family to be expressed in plants. Taken together, the widespread expression of hevein-like peptides across the plant kingdom suggests that they are probably the largest family of CRPs and are indispensable to plants probably due to their defense roles against fungi and insect pests.

2.3. Sequence alignment and comparison of hevein-like peptides

To perform sequence comparison of the putative hevein-like peptides from different plant families, multiple sequence alignment using ClustalW was performed using previously reported hevein-like peptide sequences and the putative peptides obtained from databases. Similar to previously reported 6C- and 8C-hevein-like peptides, the putative peptides displayed conserved cysteine-spacing patterns. However, among the putative 10C-hevein-like peptides, a novel cysteine motif that differed from previously reported Type I, Type II and Type III was observed and designated Type IV (Fig. 5.6).

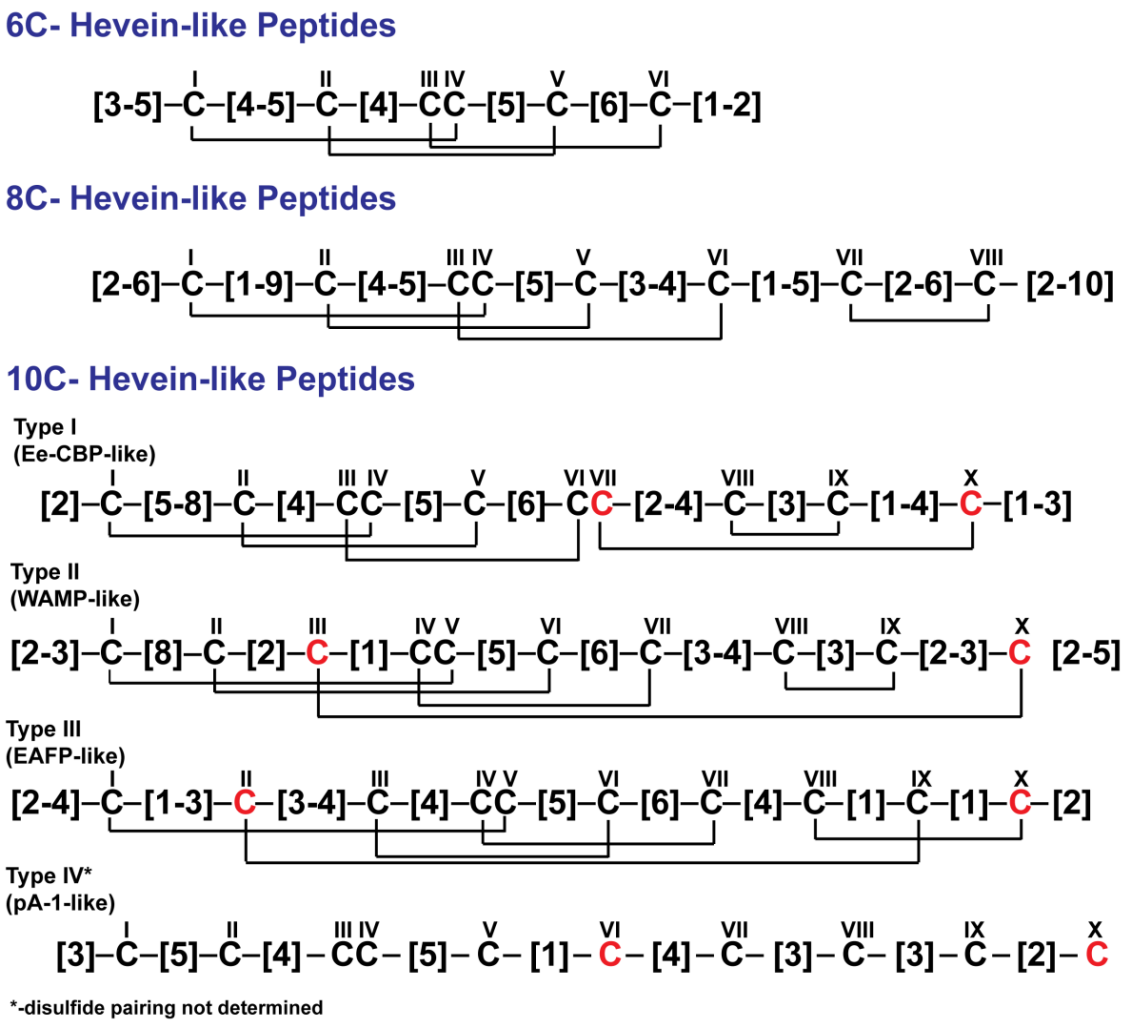


Figure 5.6. Cysteine-spacing and disulfide connectivity of hevein-like peptides. The putative 6C- and 8C-hevein-like peptides had similar cysteine-spacing patterns as their previously reported counter-parts. A novel cysteine-spacing motif was observed in a 10C-hevein-like peptide from *Plagiochila asplenioides* and was designated as Type IV. The disulfide connectivity of the new motif is yet to be determined.

2.3.4. 6C-Hevein-like peptides

From sequence alignment of five previously reported 6C-hevein-like peptide sequences and 20 putative peptides obtained from EST databases a sequence logo was obtained (Fig. 5.7). It was observed that the peptides are 29-30 amino acid residues in length and have a high content of glycine (7 residues), proline (2 to 3 residues) and aromatic amino acid residues (3 residues) which make up approximately 60% of the sequence (Appendix B). Apart from the cysteine residues, glycine at positions 2, 7, 12, 19, 22, 24 and 29, Ser-16, Pro-10, Tyr-20 and Tyr-27 are highly conserved (numbered according to Ac-AMP2). The 6C-hevein-like peptides can be divided into four loops which are about four to six residues in length. Loop 1 is the most variable while loop 2 comprises absolutely conserved residues Pro-10 and Gly-12 (Fig. 5.7). Loops 3 and 4 comprise the conserved chitin-binding domain SX(F/W)X(Y/W)CGX_4Y except one peptide from *Kochia scoparia* where the second aromatic residue in the chitin-binding domain is replaced by histidine. The disulfide bonds in previously reported 6C-hevein-like peptides form the stable cystine-knot motif where three disulfide bonds are formed between Cys I-Cys IV, Cys II-Cys V and Cys III-Cys VI. Due to the highly conserved cysteine pattern among all 6C-hevein-like peptides, we assume that the putative peptides also comprise the cystine-knot motif similar to altides and Ac-AMPs where the disulfide connectivity has been confirmed by NMR studies [45, 87].

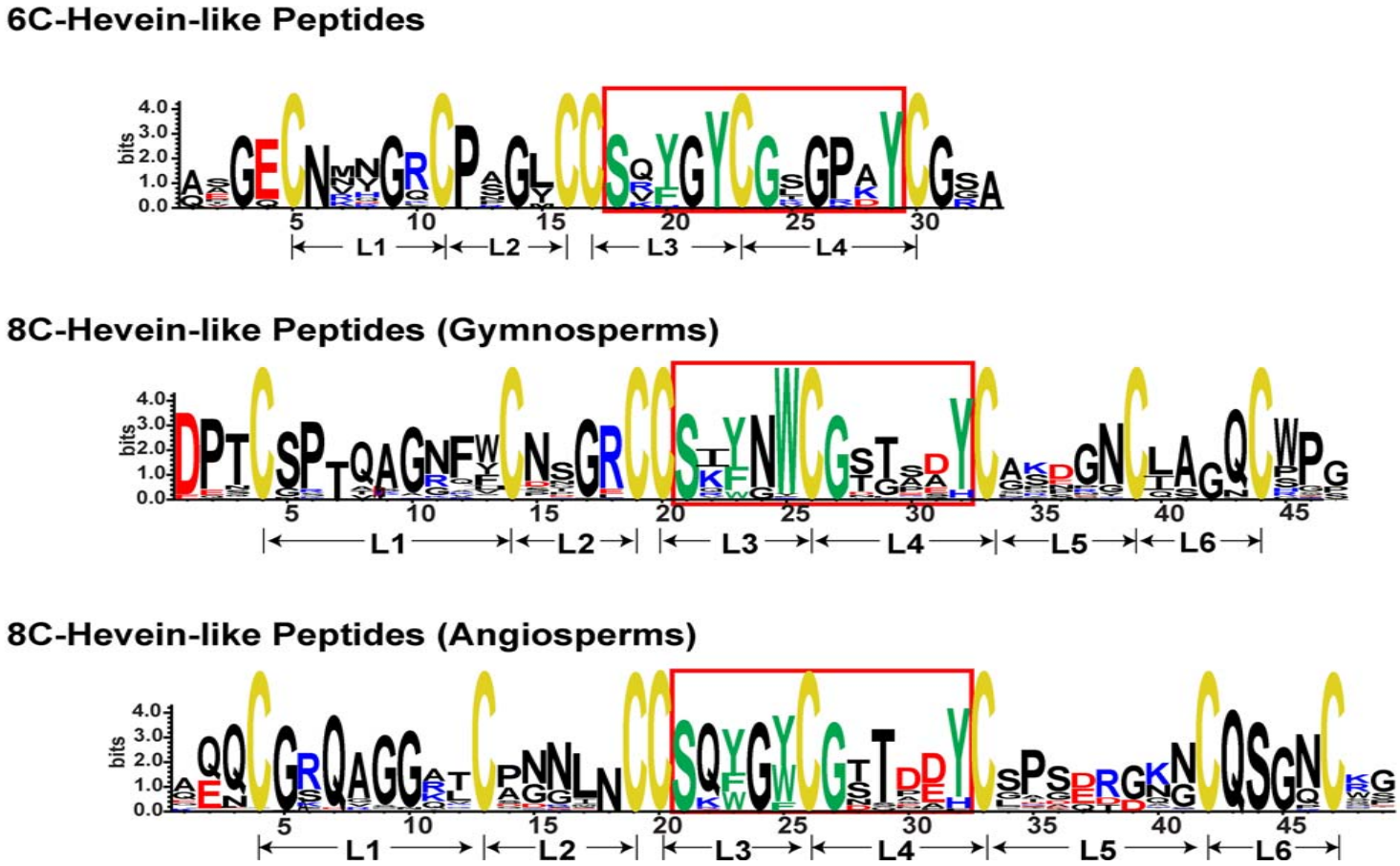


Figure 5.7. Sequence logo of putative 6C- and 8C-hevein-like peptides. The cysteine spacing is absolutely conserved among 6C- and 8C-hevein-like peptides. The 6C-hevein-like peptides can be divided into four loops and comprise the chitin-binding domain in loops 3 and 4 (highlighted in box). The N-terminal of 8C-hevein-like peptides from gymnosperms and angiosperms are starkly different but in both cases the composition of the chitin-binding domain is conserved.

2.3.5. 8C-Hevein-like peptides

Sequence alignment of 349 putative 8C-hevein-like peptides from EST databases with previously reported sequences revealed that the cysteine spacing is highly conserved and the peptide backbone between cysteine residues can be divided into six loops (Fig. 5.7). Interestingly, the N-terminal of sequences expressed in gymnosperms comprises the motif (D/EP) and differs from that of angiosperms and bryophytes which comprise the motif (A/E/QQN) (Appendix C). Although there is high sequence variability at the N- and C-termini, the length and sequences of loop 3 and loop 4, which comprises the chitin-binding domain, are conserved.

Along with the conserved cysteine residues, Ser-19 (numbered according to hevein) is absolutely conserved while Gly-25 is replaced by Val-25 or Asp25 in three sequences, namely, *Austrotaxus spicata*, *Dryas octopetela* and *Marchantia paleacea* (Appendix C). Each of these three sequences originates from a gymnosperm, an angiosperm and a liverwort respectively suggesting that there could be a close relationship between peptides from all three clades. Interestingly, Tyr-30 known to be integral for chitin-binding activity, is replaced by His-30 in all 8C-hevein-like peptides expressed in gymnosperms and in 19 sequences from eudicots belonging to Rosids, Asterids, Gunneridae, Ranunculaes and Asparagles. Structural analysis of hevein revealed that the first three disulfide bonds form the cystine knot motif while the fourth disulfide bond is formed between the last two cysteine residues [66]. Due to conserved cysteine-spacing

pattern between the putative peptides and hevein it can be assumed that the disulfide pairing is identical.

2.3.6. 10C-Hevein-like peptides

The three previously reported sequences were aligned with 16 putative 10C-hevein-like peptides (Appendix D) which revealed that there are four distinct cysteine spacing patterns designated Type I to Type IV.

Type I, similar to Ee-CBPs from *E. europaeus [52]*, is observed in three peptides from the order Sapindales comprising eudicot angiosperms. Sequence comparison with Ee-CBP showed that loop 1 of the putative hevein-like peptides is shorter than Ee-CBP and comprises four amino acid residues. Loop 2 shows high sequence conservation in peptides from the Rutaceae family but bears no homology with Ee-CBP. In loop 5 of the peptide from *Poncrius trifoliate*, Tyr-30 (numbered according to Ee-CBP) is replaced by His-30. Loop 6 and loop 8 are the most variable and loop 7 is partially conserved (Fig. 5.8). Based on the conserved cysteine spacing, it can be assumed that the disulfide connectivity is similar to Ee-CBP and follows the pattern Cys I-Cys IV, Cys II-Cys V, Cys III-Cys VI, Cys VII-Cys X and Cys VIII-Cys IX [99].

Nine monocot angiosperms from the order Poales displayed a cysteine-spacing similar to Type II previously observed in WAMPs from *T. kiharae [53]*. Sequence comparison with WAMP-1b from *T. kiharae* revealed that loop 1 was the most variable while loops 2 and 3 were highly conserved. Loop 2 comprised three residues while loop 3 comprised an absolutely conserved Leu residue.

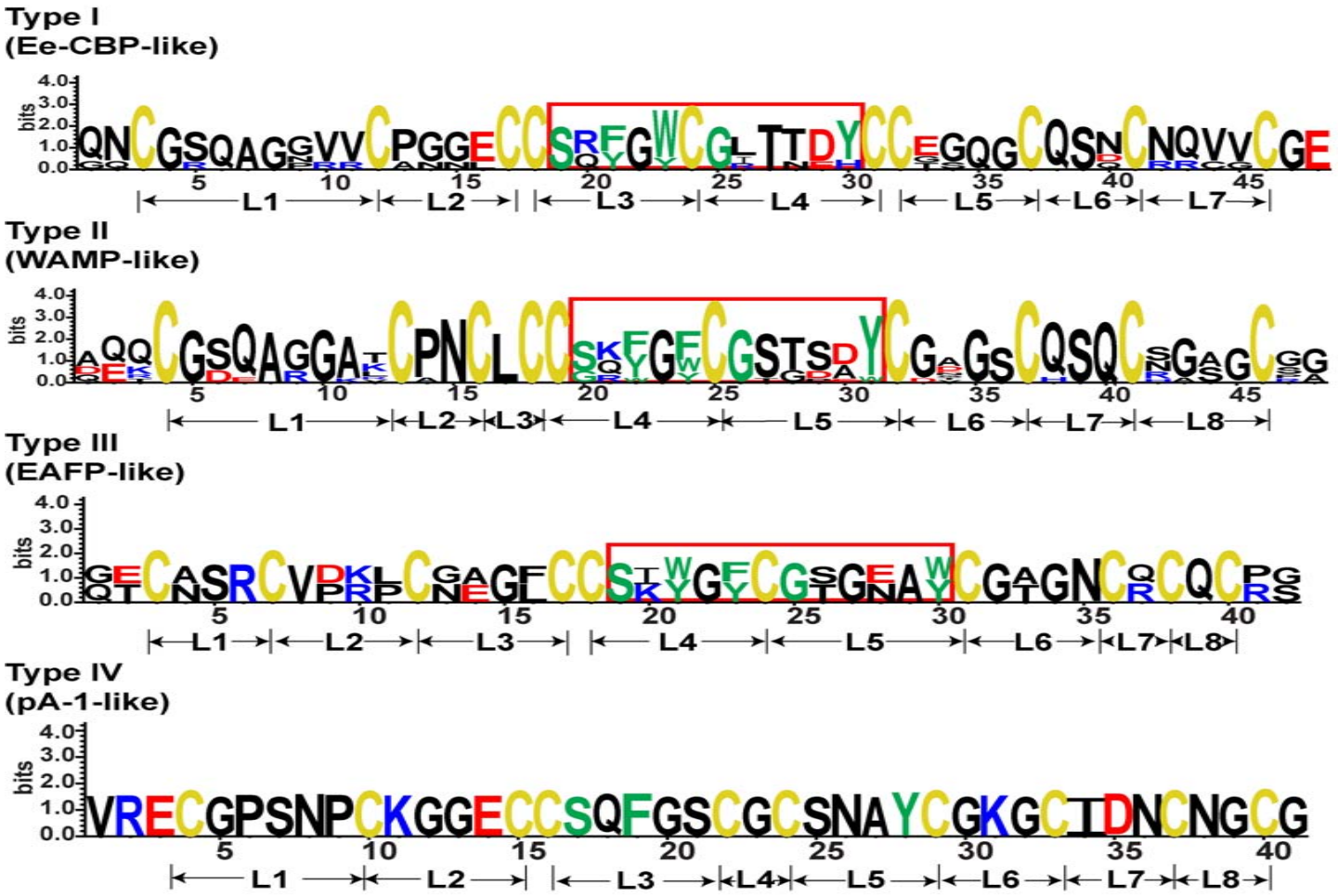


Figure 5.8. Sequence logo of putative 10C-hevein-like peptides. Four distinct cysteine-spacing patterns were observed among 10C-hevein-like peptides and designated Type I to Type IV. While Type I to Type III have been previously reported, a novel pattern, Type IV was observed in a Marchantiophyte called *Plagiochila asplenioides* with a modification in the chitin-binding domain. The chitin-binding domain is indicated in the box.

Loops 4 and 5 comprise the conserved chitin-binding domain. Interestingly, the conserved Ser-20 residue in the chitin-binding domain in loop 4 of peptides from *Triticum aestivum* and *Elymus spicatus* is replaced by Gly-20 similar to WAMPs reported from *Triticum kiharae* [53] (Appendix C). Loop 6 comprises the conserved motif GXGX, except in *Elymus spicatus*, where Gly-33 is replaced by Asp-33 (Fig. 5.8). Loop 7 comprises a conserved Glu-37 residue which is replaced by His-37 in one peptide from *Oryza sativa* Japonica. It can be assumed that the putative hevein-like peptides from Poales share the same disulfide connectivity as WAMPs from *Triticum kiharae*, Cys I-Cys V, Cys II-Cys VI, Cys III-Cys X, Cys IV-Cys VII, Cys VIII-Cys IX, since they belong to the same plant family [73].

The Type III cysteine-spacing pattern, similar to EAFPs from *E. ulmoides Oliv [48]*, was observed in one gymnosperm from the order Gnetidae. However, on sequence comparison with EAFP-1, no absolute sequence conservation was observed except for the cysteine residues (Fig. 5.8). This could be because the putative peptide is observed in a gymnosperm as opposed to EAFPs isolated from eudicot angiosperms. Despite the high sequence variation observed in the all loops of both peptides, loops 4 and 5 are more similar as they comprise the chitin-binding domain. Due to conserved cysteine-spacing it can be presumed that the disulfide pairing is similar to EAFPs.

A novel cysteine-spacing pattern, that we designated Type IV was observed in *Plagiochila asplenioides*, a Bryophyte from the order Marchantiidae. Intriguingly, the additional cysteine residue in this peptide replaces the Ser residue of the chitin-binding domain. This is the first instance of a 10C-hevein-like peptide from a Bryophyte with a substitution of the conserved Ser residue by a Cys residue in the chitin-binding domain. Further studies need to be done to isolate and elucidate the disulfide pattern, structure and chitin-binding affinity of the putative peptide.

From this data it is clear that 8C- and 10C-hevein-like peptides are the most diverse sub-classes of the hevein-like peptide family. This can be attributed to their widespread distribution across the plant kingdom from mosses and gymnosperms to highly evolved eudicots. In contrast, the 6C-hevein-like peptides seem to be more recent members of the hevein-like peptide family and comprise loops that are less variable than 8C- and 10C-hevein-like peptides. An interesting feature of the 10C-hevein-like peptides is the disulfide promiscuity resulting from four distinct cysteine-spacing patterns. Despite the variations in the length of the primary sequences of 6C-, 8C- and 10C- hevein-like peptides, the length of the loops comprising the chitin-binding domain are conserved across the hevein-like peptide family. Thus, this peptide family shows universality in diversity in their primary and secondary structure as shown by the great degree of diversity as well as conservation of certain key residues involved in binding chitin.

2.4. Precursor organization of hevein-like peptides

One of the advantages of using EST-based databases is that the full-length precursor sequence of the target peptides can be studied. Since knowledge of the sequence diversity was obtained from the primary and secondary structure of the putative peptides, we analyzed the precursor sequences to gain insights into their putative processing sites and pathway followed in their *in planta* synthesis. Gene cloning studies of 6C-, 8C- and 10C-hevein-like peptides have shown that the precursor organization is conserved throughout the family and comprises a signal peptide domain, a mature CRP domain and a C-terminal tail with variations only in domain length [86, 89, 98]. This precursor organization suggests that hevein-like peptides are ribosomally synthesized and are first translated as a three-domain precursor. The presence of a signal peptide suggests that the biosynthesis follows the secretory pathway. The biosynthesis of hevein-like peptides involves excision of the signal peptide from the mature domain followed by excision of the C-terminal tail and oxidative folding of the mature peptide to form the disulfide bonds.

Sequence alignment of the precursors of putative hevein-like peptides from the EST database revealed that they comprise a similar precursor structure. The residue at the N-terminal processing site is Ala or Gly in majority of the putative peptides sequences. However, 8C-hevein-like peptides from *Austrocedrus chilensis*, *Brassica napus*, *Callitris gracilis, Papuacedrus papuana* and *Populus trichocarpa* comprise a Ser residue while *Cycas micholitzii* comprises a Lys residue at the N-terminal cleavage site. The C-terminal of hevein-like peptide

precursors comprises multiple glycine repeats and resembles the hinge region of chitinases. Thus it is difficult to predict the exact processing site at the C-terminal tail.

An interesting feature of hevein-like peptide precursors is the diversity of their C-terminal tail. In 6C-hevein-like peptides the length of the precursor varies from 56 to 90 amino acid residues. BLAST analysis of the C-tail in the NCBI conserved domain database showed no homology to any known domains. The precursors of 8C- and 10C-hevein-like peptides are significantly longer than 6C-hevein-like peptides. This can be attributed to the chimeric nature of the precursors where the C-tails encode for other bioactive protein domains. In 8C-hevein-like peptides, the mature domain is followed by either a short (11-35 amino acid residues) or a long C-tail (35-250 amino acid residues) where the short C-tails shared no homology with any known protein sequences in the database. In 10C-hevein-like peptides, C-tails of peptides from Poaceae family comprise the lysozyme-like protein domain while those from Rutaceae family bear no sequence homology to any known proteins. The C-terminal tails bearing homology to the RlpA-DPBB-1 domain were only observed in five peptides expressed in Bryophyta. The Extensin domain was observed in one peptide from *Glycine max*. A phylogenetic tree was constructed using the precursor sequences of the putative hevein-like peptides to gain insight into their probable evolutionary pattern (Fig. 5.9).

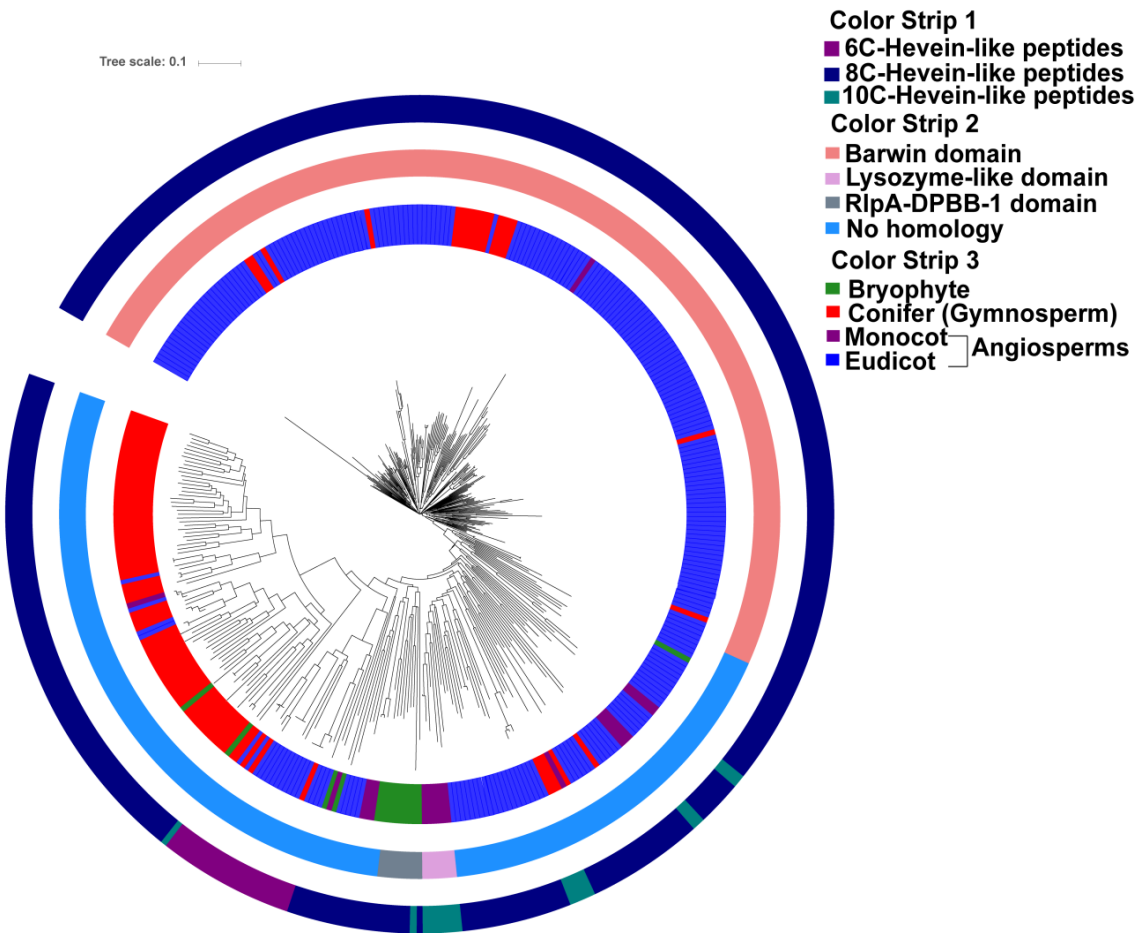


Figure 5.9. Phylogenetic Distribution of putative hevein-like peptides. (A) Phylogenetic tree of the putative hevein-like peptides. Color strips 1, 2 and 3 represent the number of cysteine residues, the C-terminal domains and the clade in which the peptides are expressed, respectively.

From the phylogenetic tree it is clear that the hevein-like peptides are grouped into distinct clusters based on their C-tail composition and the clade to which the plants belong. From the consolidated data of 6C-, 8C- and 10C-hevein-like peptide precursors, 48.8% of the C-tails comprised the Barwin domain, 5.5% comprised the lysozyme-like domain, 1.3% comprised the RlpA-DPBB1 domain and 0.8% comprised the Extensin domain (Fig. 5.10). Thus, although the precursor organization is conserved across the hevein-like peptide family, variations in processing sites and C-terminal tail composition indicate that there may be species or family-specific mechanisms or processing enzymes that need to be further elucidated.

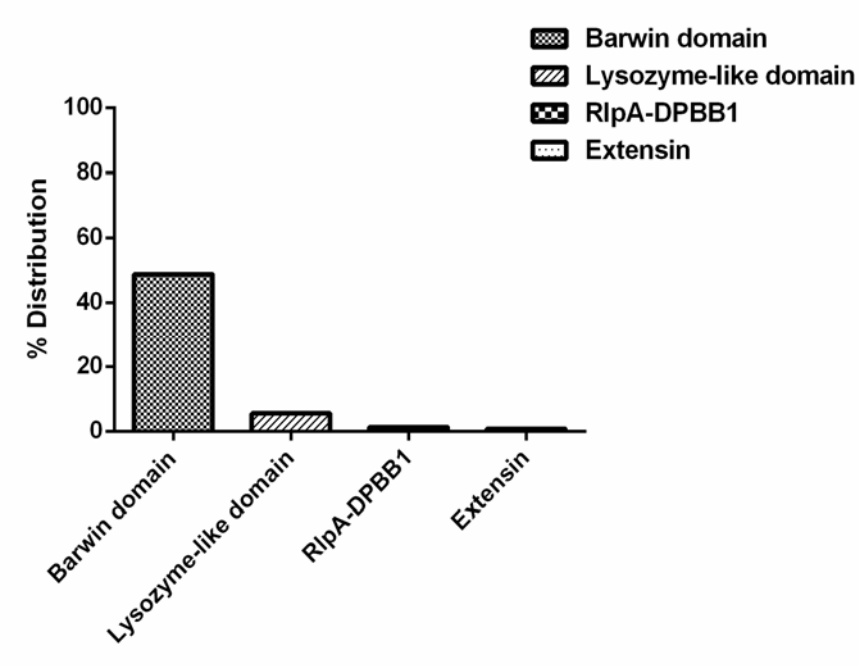


Figure 5.10. Distribution of C-tail domains in hevein-like peptides. Majority of the putative hevein-like peptide precursors comprise a C-terminal tail with a Barwin domain. The 10C-hevein-like peptide precursors from the Poaceae family comprise C-tails with a lysozyme-like domain while five peptides from the clade Bryophyta comprise the RlpA-DPBB-1 domain. The Extensin domain was observed in one peptide from *Glycine max*.

3. Discussion

3.1. Distribution and evolution of hevein-like peptides

Insights into the distribution and evolution of hevein-like peptides have been gained using EST-based nucleotide databases. A total of 385 putative hevein-like peptides were identified in 124 different plant families originating from non-flowering and flowering plants. Knowledge of the distribution of hevein-like peptides in the plant kingdom helps shed light on the origin and evolution of this peptide family. The occurrence of 8C-hevein-like peptides in Bryophytes suggests that they may have appeared about 450 million years ago when the first terrestrial plants (mosses and liverworts) diverged from their algal ancestors [185] (Fig. 5.11). Interestingly, 10C-hevein-like peptides were also found in the clade Marchantiidae of Bryophyta which suggests that 8C- and 10C-hevein-like peptides are the most ancient members of the hevein-like peptide family. The 6C-hevein-like peptides are relatively recent members of this peptide family as they are expressed exclusively in Asterids which diverged from core eudicot Rosids about 128 million years ago [186].

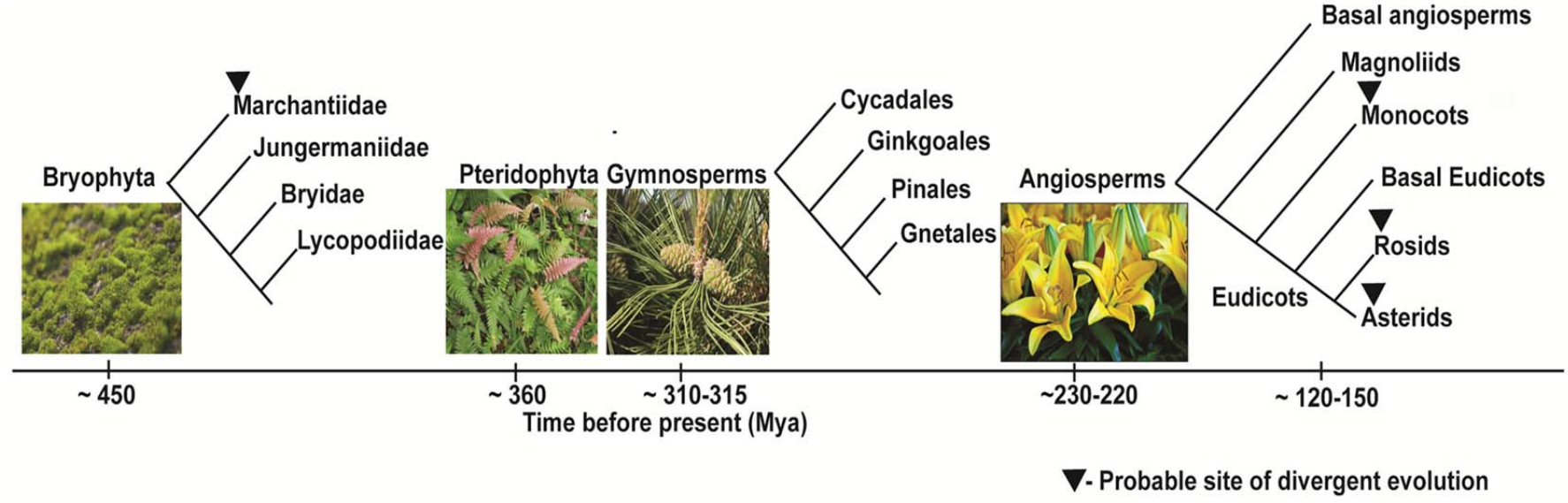


Figure 5.11. Summary cladogram showing the major evolutionary groups of hevein-like peptides. The timeline of evolution of hevein-like peptides is indicated as million years ago (Mya). The expression of 8C- and 10C-hevein-like peptides in Bryophyta suggests that they originated ~450 million years ago in the first terrestrial plants on Earth. The 6C-hevein-like peptides are more recent members of the hevein-like peptide family as they are only expressed in Asterids which diverged from class Rosids ~125 million years ago. It is speculated that the 6C- and 10C-hevein-like peptide sub-classes arose by divergent evolution from their 8C-hevein-like peptide ancestors in closely related plant species.

The distribution of 6C-hevein-like peptides is similar to cyclotides which have only been found in eudicot angiosperms [187]. However, the evolutionary pathway of 6C- and 10C-hevein-like peptides is still unclear.

The similarity in length of the 8C- and 10C-hevein-like peptides with variations only in the number and positions of cysteine residues in the mature domain allow us to speculate that they arose from point mutations in 8C-hevein-like peptides (Fig. 5.12). The four distinct cysteine-spacing patterns of 10C-hevein-like peptides are observed in orders Marchantiidae, Gnetidae, Poales and Sapindales. Since Marchantiidae is the most ancient among the three clades, it can be assumed that the origin of 10C-hevein-like peptides was in Bryophytes. Based on the summary cladogram in figure 5.10 it can be presumed that the distinct cysteine-spacing patterns of 10C-hevein-like peptides evolved from 8C-hevein-like peptides of Marchantiidae, Poales and Sapindales, respectively by divergent evolution. Thus, it is possible that after the divergence of 10C-hevein-like peptides from their 8C-ancestors, adaptive evolution and positive Darwinism played a role in retaining their expression due to the increased resistance to fungal pathogens in plants. The most recent members of the hevein-like peptide family are the 6C-hevein-like peptides which are exclusively expressed in eudicots. Majority of the 6C-hevein-like peptides are expressed in the Amaranthaceae family belonging to the order Caryophyllales.

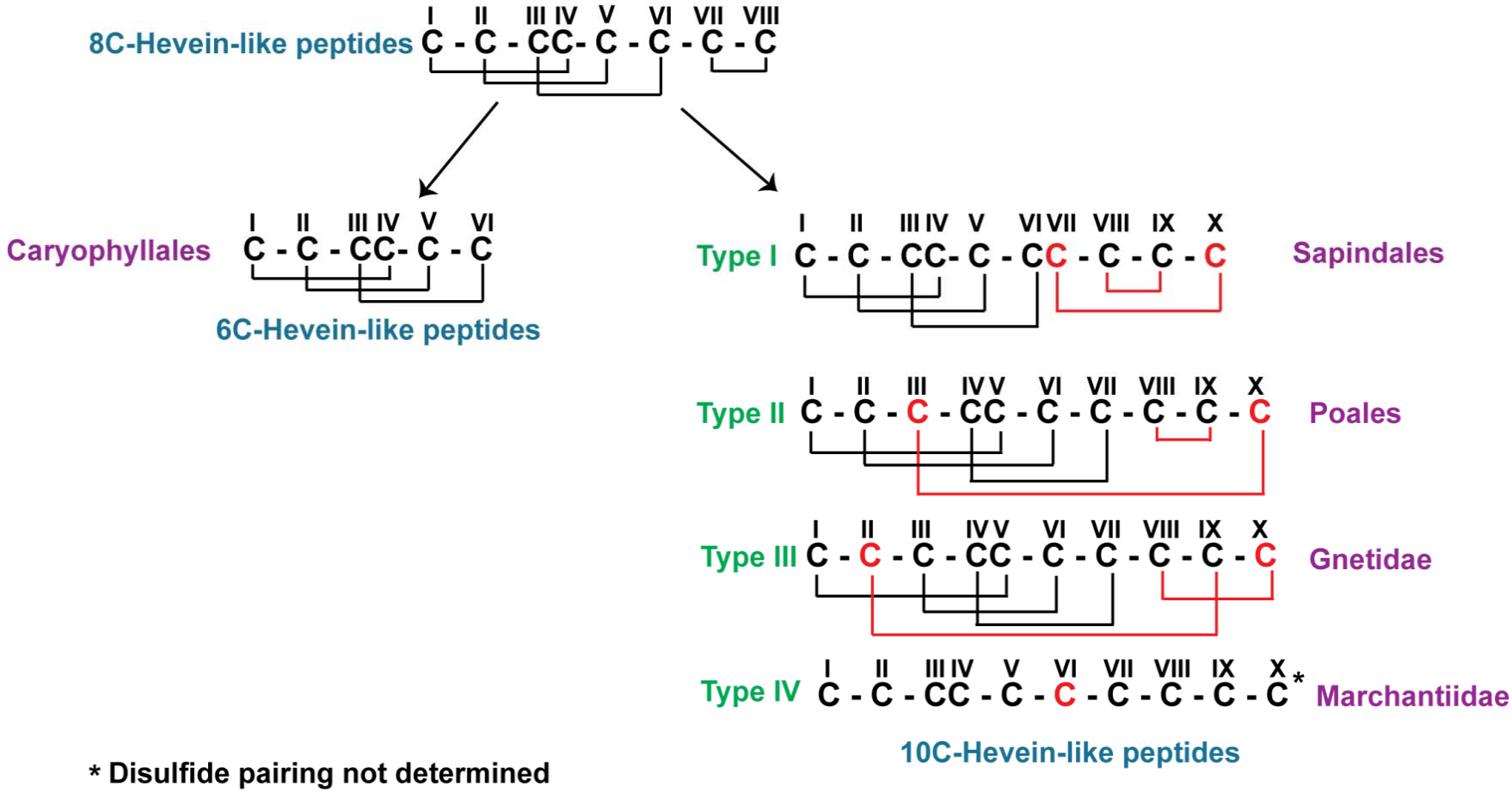


Figure 5.12. Putative evolutionary pathways of hevein-like peptides. The 10C-hevein-like peptide sub-class may have divergently evolved from 8C-hevein-like peptides in orders Sapindales, Poales, Gnetidae and Marchantiidae by point mutations resulting in two additional cysteine residues. The “truncation” of 8C-hevein-like peptides giving rise to 6C-hevein-like peptides may have occurred in order Caryophyllales.

Thus, it can be assumed that the truncation of the C-terminal comprising the last two cysteine residues of 8C-hevein-like peptides may have occurred in Caryophyllales and resulted in the divergent evolution of 6C-hevein-like peptides. However, the mechanism behind the truncation still remains to be elucidated. The 6C-hevein-like peptides could then be selected by positive Darwinism due to their small size and increased potency against fungal pathogens. A similar evolutionary pattern has been observed in chitinases and snake venom toxins that confer functional diversity and ability to target different prey [178, 188].

An alternate mechanism for the evolution of 10C-hevein-like peptides (WAMPs) from *Triticum kiharae* (wheat) has been suggested by Andreev *et al*. who proposed that WAMPs could be truncated Class I chitinases [89]. It is well known that Class I chitinases evolved by the incorporation of a chitin-binding domain in the gene coding for the catalytic domain by a transposition event [189]. On sequence comparison, they observed a striking similarity between the chitin-binding domains of Class I chitinases from cereals and WAMP-1b. Thus, they speculated that frame-shift deletions or alternative RNA splicing of chitinase genes may result in the expression of “truncated chitinases” comprising only the chitin-binding domain. This interesting observation suggests that, during the evolution of class I chitinases, a reverse phenomenon also took place where deletions in the chitinase gene resulted in the expression of a single chitin-binding domain lacking the catalytic domain. However, this theory is specific to WAMPs and cannot explain the distinct cysteine-spacing patterns of 10C-hevein-

like peptides from other plant families. Thus, more studies are required to gain a deeper insight into the origin of 10C-hevein-like peptides.

The evolution pattern of hevein-like peptides differs from cyclotides which arose from their linear counterparts by convergent evolution where, the mutations to insert essential Asp/Asn residues at the point of cyclization occurred independently in Asterids and Rosids about 125 million years ago [190]. Together, this data suggests that hevein-like peptides are a an ancient, highly diverse class of CRPs and that the sub-classes of hevein-like peptides arose due to divergent evolution between closely related plant species.

3.2. Universality in diversity of hevein-like peptides

The highly diverse nature and widespread distribution of hevein-like peptides in plants makes them interesting candidates to study molecular diversity and the effect of sequence variations on functional diversification. Our results highlight the high molecular diversity of hevein-like peptides as shown by the extensive variation in their loop lengths and residue composition. The 6C-hevein-like peptides show relatively high sequence conservation when compared to 8C- and 10C-hevein-like peptides. This could be attributed to their distribution in closely related plant families resulting in less sequence variation. In 8C-hevein-like peptides, apart from the eight cysteine residues, none of the other residues show absolute conservation. In 10C-hevein-like peptides only two glycine residues in loops 3 and 4 are absolutely conserved suggesting that they may play an important role in stabilizing the conformation of the peptide while binding to chitin.

It is commonly believed that the variable residues; not involved in biological functions are prone to mutations as opposed to more conserved residues and they are located in the interior pocket in the peptide [191]. Thus, the surface residues are more prone to insertions, deletions and substitutions. However, in the case of hevein-like peptides, mutations in important residues in the chitin-binding domain are observed. Interestingly, in the loops comprising the chitin-binding domain, the tyrosine residue known to be crucial for binding to chitin is replaced by histidine in 8C-hevein-like peptides from 2% of gymnosperms and 8% of angiosperms. A similar phenomenon is observed in a 10C-hevein-like peptide from *Poncrius trifoliate* belonging to the Rutaceae family and a 6C-hevein-like peptide from *Kochia scorpia*. Thus, it can be speculated that there may have been a gene transfer event from gymnosperms to angiosperms during the evolution of 8C- and 10C-hevein-like peptides. However, further analysis of the gene organization needs to be done to validate this hypothesis. We speculate that the chitin-binding property of these peptides may not be drastically affected since the imidazole ring of the histidine residue can retain the CH-π stacking interaction with the GlcNAc moiety. Intriguingly, the conserved serine residue of the chitin-binding domain is replaced by a glycine residue in 10C-hevein-like peptides from wheat species. A similar observation was made in WAMPs from *Triticum kiharae*, where this mutation led to decreased affinity for chitin oligosaccharides [53]. However, this loss of binding affinity is compensated by Ser-36 in loop 6 which not only confers resistance against fungal chitinases but also inhibits them [101]. The replacement by histidine and glycine residues in the

chitin-binding domain could also alter the net charge of the peptides, increase their cationicity and amphiphilicity conferring them anti-bacterial properties in addition to fungal defense [73]. Thus, it is possible that hevein-like peptides underwent multiple mutations during evolution to confer the plant an arsenal of defense peptides to target diverse microbial pathogens. Together, these observations suggest that the promiscuity of hevein-like peptides is integral for plant defense against multiple pathogens [192].

All hevein-like peptides are stabilized by a cystine-knot motif formed by the first three disulfide bonds [67]. This cystine-knotted scaffold has been observed in many CRPs and insect toxins [193, 194]. In hevein-like peptides this core cystine-knot motif is conserved and additional disulfide bonds are present in 8C- and 10C-hevein-like peptides. A similar phenomenon is observed in snake venom toxins, where, additional disulfide bonds are known to confer specificity to different receptors and also alter the receptor-binding efficiency [178]. However, thus far, the effect of additional disulfide bonds on the binding efficiency of hevein-like peptides has not been elucidated. Taken together, these observations show that despite the sequence diversity of hevein-like peptides, key features of the primary and secondary structure are conserved indicating the universality in diversity of this peptide family.

3.3. Precursor organization and biosynthesis of hevein-like peptides

Knowledge of the precursor organization of hevein-like peptides provides insights into their bio-processing pathway. The precursor organization comprising a signal peptide, a mature CRP domain and a C-terminal tail suggests that they are ribosomally synthesized peptides. Similar precursor structure is observed in thionins [159] and varies from precursors of cystine-knot α-amylase inhibitors (CKAIs) which comprise an N-terminal pro-domain before the mature domain [94]. It can be presumed that the folding of hevein-like peptides to form the cystine knot occurs in the endoplasmic reticulum by virtue of their signal peptide. Subsequently they may be transported to the Golgi apparatus or vacuoles where cleavage from the C-terminal tail occurs before secretion of the peptide [195, 196]. The signal-peptide guided biosynthesis of hevein-like peptides could facilitate natural evolution due to the relatively small number of processing enzymes and the high multiplicity of pathways towards similar types of precursors. Thus, the convergent evolution of processing pathways across different plant families confers an evolutionary advantage since access to chemical diversity is increased at a low genetic cost [197].

The natural diversity of hevein-like peptides is evident from the chimeric nature of the 8C- and 10C-hevein-like peptide precursors. Their C-terminal tails encode bioactive proteins with a Barwin domain or RlpA-DPBB-1) domain. The Barwin domain is a protein domain found in a lectin from Barley and is known to bind to tetrameric N-acetylglucosamine while the RlpA-DPBB-1 domain is a transglycosylase which targets glycan chains [88, 198]. A similar precursor

organization was observed in cyclotides from *Clitoria ternatea* where the C-tail comprised the legume albumin PA1a [93, 199]. However, the 6C-hevein-like peptides and some 8C- and 10C-hevein-like peptides comprise short C-terminal tails that have no homology to any known protein domains.

The diversity in precursor organization could lead to functional diversity of hevein-like peptides. The hevein-like peptides with chimeric precursors could function like “guided missiles”, where the chitin-binding domain of hevein-like peptides facilitates binding to the fungal cell wall and the Barwin and RIpA-DPBB-1 domains either increase binding efficiency or enzymatically digest the chitin chains, respectively. This mechanism involving enzymatic digestion of the chitin cell wall is similar to chitinases and cellulases [60, 200, 201]. Thus, it can be speculated that Class I chitinases evolved from 8C- or 10C-hevein-like peptide precursors comprising catalytic domains.

Evaluation of the important residues involved in the bio-processing of hevein-like peptide precursors was performed by aligning the putative peptide precursors. It was observed that excision between the signal peptide and the mature domain occurs at either a glycine or alanine residue. A variety of endoproteases in the vacuole or Golgi apparatus may be responsible for cleavage of the C-tail from the mature domain. One such family of endoproteases is the asparaginyl endoproteinases (AEPs) responsible for the cleavage and cyclization of cyclotides [202]. In hevein-like peptides, the C-terminal cleavage site is occupied by a glycine residue suggesting that a glycine endoproteinase might be responsible for cleavage of the mature domain from the C-tail.

3.4. Applications of hevein-like peptides

The discovery that hevein-like peptides are more numerous than originally anticipated will be of value in developing strategies to utilize this diverse class of CRPs for agrochemical and pharmaceutical applications.

3.4.7. Development of transgenic crops

Plant fungal pathogens adversely affect a myriad of agricultural crops leading to major losses in the agricultural industry and posing a threat to food safety [203]. Thus, effective and sustained control of agricultural crops against fungal pathogens is becoming increasingly important. Production of transgenic plants with increased resistance to fungal infections could alleviate the problem. However, expression of non-proteinaceous inhibitors in plants is cumbersome as a consequence of their complex bio-processing pathways [165]. Knowledge of the gene structure and *in vitro* anti-fungal activity of hevein-like peptides can be used for the development of transgenic plants with increased resistance to fungal pathogens. Previously, transgenic expression of 8C-hevein-like peptide, *pnAMP-h2* cDNA in tobacco plants has resulted in increased resistance to oomycete, *Phytophthora infestans* known to cause black shank disease [98]. Similarly, Gao *et al*. have successfully expressed a defensin gene in potato plants which demonstrated resistance against *V. dahilae* in greenhouse and field tests comparable to commonly used fumigants [204]. Thus, hevein-like peptides serve as less-toxic alternatives to commercially used fungicidal agents and their ease

of expression in plants makes them promising candidates for the development of transgenic crops.

3.4.8. Scaffold for development of peptidyl bioactives

Cysteine-rich peptides possessing a cystine-knotted structure serve as scaffolds to graft multiple linear bioactive peptides or proteins from plants, animals or microorganisms. Thus, cystine-knotted hevein-like peptides possess several features that make them interesting candidates for engineering stable bioactive peptides with increased bioavailability. First, hevein-like peptides are easily available by chemical synthesis or molecular biology methods. Second, cystine-knotted peptides are tolerant to sequence variations in multiple loops thus; the structure of hevein-like peptides makes them potential scaffolds for grafting linear bioactive peptides [205-207]. This technique has previously been used to graft bradykinin antagonists and melanocortin onto cyclotide kalata B1 [208, 209], human kallikrein-related peptidase 4 inhibitor onto sunflower trypsin inhibitor-1 [210] and angiogenic peptides onto *Momordica cochinca cochinchinensis* trypsin inhibitor-II [211]. Third, the cystine-knotted scaffold confers high resistance to degradation by proteolytic enzymes thereby increasing the bioavailability of grafted linear peptides [40, 158]. Due to the lectin-like property of hevein-like peptides, they can be used for the development of targeted therapeutics with specificity for glyco-peptides on the plasma membrane. These peptide conjugates would have the added advantage of increased bioavailability along

with high specificity resulting in lesser off-target side-effects which are common among small molecule drugs.

Thus, for the first time, the distribution and evolution of hevein-like peptides in plants have been studied using a bioinformatics-guided approach. The unprecedented distribution of hevein-like peptides in bryophytes, gymnosperms and angiosperms suggests that they are integral for plant defense against fungi. Insights into the sequence diversity of 6C-, 8C- and 10C- hevein-like peptides show that the key residues involved in binding to chitin are relatively more conserved as compared to the remaining residues in the sequence. On analyzing the precursor organization, it was found that 8C- and 10C-hevein-like peptides show great diversity in the C-terminal tail which encodes for other bioactive proteins like Barwin and RlpA-DPBB-1. Thus, combined knowledge of distribution, sequence diversity and biosynthesis allowed us to propose a probable evolutionary pathway of hevein-like peptides. We presume that 6C- and 10C-hevein-like peptides arose by divergent evolution from their 8C-hevein-like peptide ancestors in closely related plant families.

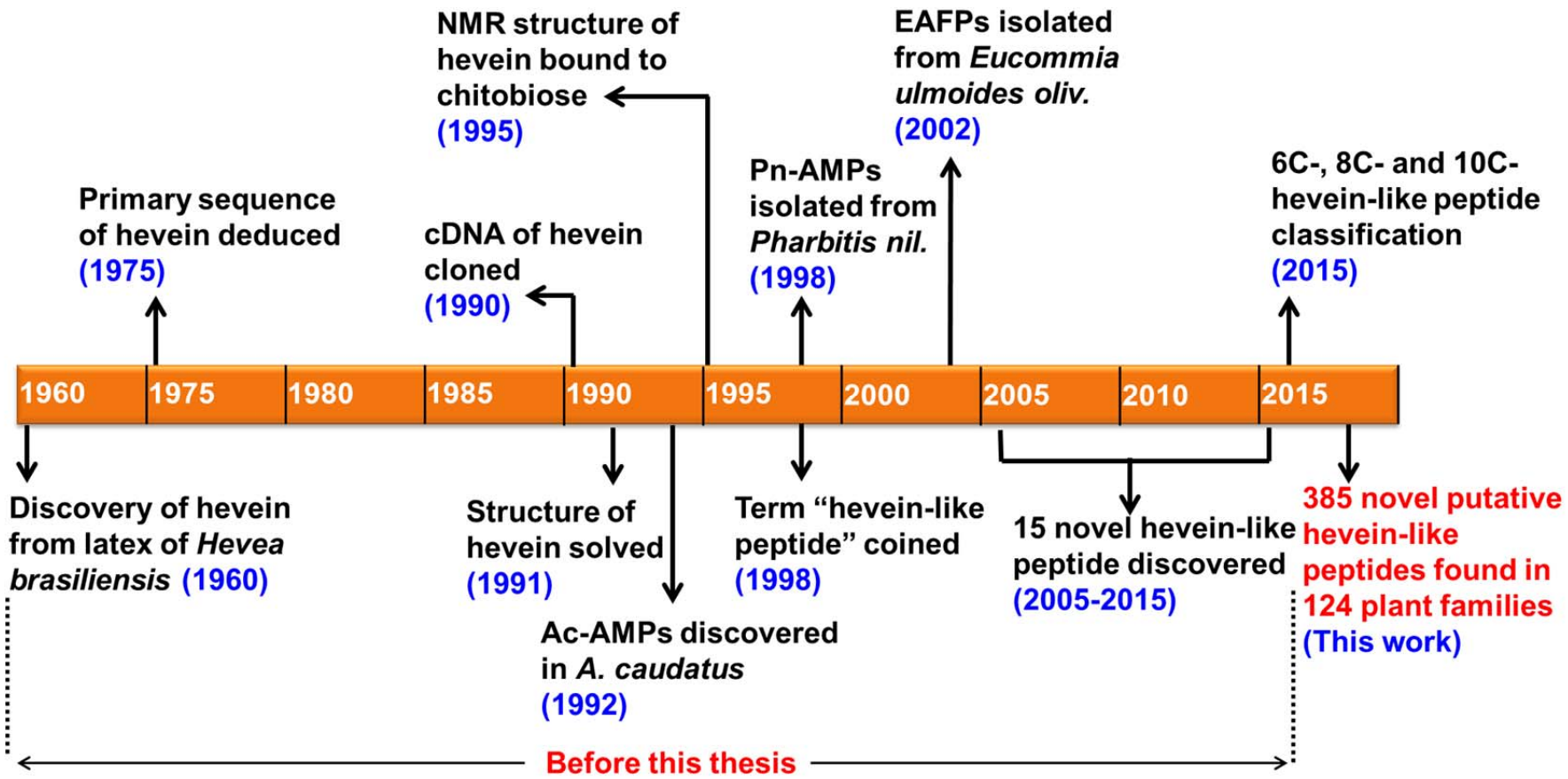


Figure 5.13. Updated timeline of hevein-like peptides. About 56 years after the discovery of hevein, our study on the distribution of hevein-like peptides in plants has expanded our knowledge on this highly diverse family of CRPs. We have discovered 385 novel hevein-like peptides across the plant kingdom and also suggested a possible evolutionary pathway for the hevein-like peptide family, the largest family of constitutively expressed CRPs.

Summary, Conclusion and Future Outlook

Hevein-like peptides are a family of cysteine-rich peptides that are rich in glycine and aromatic residues. They play an integral role in plant defense against fungal pathogens by binding to chitin, a polysaccharide of repeating N-acetylglucosamine units (GlcNAc) and the primary constituent of fungal cell walls. An interesting feature of the hevein-like peptide family is that they can be divided into three sub-classes, namely 6C-, 8C- and 10C-hevein-like peptides based on the number of cysteine residues they contain. The molecular diversity and defense role of hevein-like peptides in plants prompted us to isolate and characterize novel hevein-like peptides from medicinal herbs like *Alternanthera sessilis* and *Moringa oleifera*. We also used bioinformatics and proteomic methods to gain insights into the distribution, diversity and evolution of the hevein-like peptide family.

Chapter 3 discusses the isolation and characterization of six novel 6C-hevein-like peptides, collectively named altides, from red and green varieties of *A. sessilis*. This finding is significant as it expands the existing library of 6C-hevein-like peptides from four to ten. Our data signifies the high molecular diversity of altides as none of the loops are absolutely conserved even among altides from the same plant species. Structural analysis of aSG1 using 2D-NMR showed that the three disulfide bonds form a cystine knot conferring altides stability against thermal and enzymatic degradation.

From ^{1}H-NMR titration experiments we found that the binding affinity of altides, aSG1 and aSR1 to $(GlcNAc)_2$ and $(GlcNAc)_3$ was in the micromolar range and decreased by two-orders of magnitude with an increase in one GlcNAc unit. The binding motif comprised a conserved Ser, Gly and three aromatic residues occupying the chitin-binding domain. Since the binding interaction with chitin oligosaccharides was established, we studied the anti-fungal effects of altides against seven commonly occurring phytopathogenic fungi. Altides, aSG1 and aSR1 showed broad-spectrum anti-fungal effects against four of the seven tested strains with IC_{50} in the micromolar range. Microscopic analysis of fungi treated with altides showed that the hyphae were stunted and branching was reduced suggesting that altides inhibit fungi by retarding hyphal growth. Knowledge of the binding interaction with chitin oligosaccharides and anti-fungal property of altides can be used to develop robust anti-fungal agents with increased binding affinity to chitin.

Gene cloning studies of aSG1 revealed that altides are synthesized as a precursor protein and are processed to attain the mature peptide. The altide precursors comprise a signal peptide domain, a mature domain and a C-terminal tail suggesting that their biosynthesis is through a secretory pathway. Knowledge of the gene organization and precursor structure may be useful for the development of transgenic crops with increased resistance to fungal pathogens.

Chapter 4 describes the isolation and characterization of three novel 8C-hevein-like peptides, called morintides, from *M. oleifera* using proteomic and transcriptomic methods. Morintides mO1 and mO2 are 44 amino acid residues in

length and differed in only one amino acid residue while mO3 is 35 amino acid residues in length. Sequence comparison with previously reported 8C-hevein-like peptides showed the cysteine-spacing pattern is conserved and the sequences can be divided into six loops. The N-terminal of all 8C-hevein-like peptides comprises a Gln residue that is spontaneously converted to pyroglutamate, enhancing resistance against exopeptidases. Apart from the cysteine residues, three Gln, one Leu and two Ser residues are absolutely conserved suggesting that they may play a significant role in structure stabilization. Computer modeling of the structure of morintide mO1 revealed that the first three disulfide bonds form a cystine knot while the fourth disulfide bond is at the C-terminal.

Studies on the chitin-binding property of morintide, mO1, revealed that mO1 binds completely to chitin beads, and its elution required heating at 100°C under acidic conditions, suggesting a strong binding interaction between mO1 and chitin. As a consequence of its chitin-binding property, mO1 could inhibit the growth of two phytopathogenic fungi, *A. alternata* and *A. brassiciola*. Morphological changes like short and thick hyphae with reduced branching were observed on treatment of *A. alternata* and *A. brassiciola* with mO1. Transcriptomic analysis revealed that morintides have a similar precursor organization as other hevein-like peptides and comprise a signal peptide domain, mature domain and a C-terminal tail. However, the C-tail of morintides is only 15 amino acid residues in length in contrast to the 142 amino-acid-long C-tail of hevein which encodes a protein with a Barwin domain. This diversity of precursor

organization may hold clues to the evolutionary pattern of 8C-hevein-like peptides.

In chapter 5, the distribution, evolution and diversity of the hevein-like peptide family was studied using an EST-based in silico approach. A total of 385 novel hevein-like peptides were discovered in 124 plant families belonging to 52 orders. This widespread distribution of hevein-like peptides suggests that they are probably the largest family of constitutively expressed CRPs. Sequence alignment and comparison of 6C-, 8C- and 10C-hevein-like peptides shed light on the molecular diversity of this peptide family. While 6C- and 8C-hevein-like peptides had conserved cysteine-spacing patterns, 10C-hevein-like peptides had four distinct cysteine-spacing motifs which we designated Type I to Type IV resulting in distinct disulfide pairing. While cystine patterns Type I, Type II and Type III have been previously reported, Type IV is a novel motif found during our EST-based search in a Marchantiophyte, *Plagiochila asplenioides*. Further studies will be done to isolate this peptide and elucidate its disulfide pattern and structure. This promiscuity of the cysteine residues and disulfide bonds in 10C-hevein-like peptides further highlights the molecular diversity of this vast peptide family.

Knowledge of the precursor structure of hevein-like peptides was used to construct a phylogenetic tree and gain insight into their putative evolutionary pathway. A striking feature of hevein-like peptide precursors is the diversity of their C-terminal tail. While 6C-hevein-like peptides comprise short C-tails that bear no sequence homology to any known proteins, certain 8C- and 10C-hevein-

like peptides comprise chimeric precursors where the C-tails encode proteins with Barwin, RlpA-DPBB-1, lysozyme-like or Extensin domains. These chimeric precursors may function like “guided missiles” where the hevein-like peptide domains bind to GlcNAc units and the catalytic domains digest the chitin chains.

Phylogenetic analysis suggests that 8C-hevein-like peptides are the most ancient members of the hevein-like peptide family as they are expressed in Bryophytes, the first terrestrial plants on Earth. In addition, our analysis showed that 6C- and 10C-hevein-like peptides are evolved from their 8C-hevein-like peptide ancestors by divergent evolution. We propose that 10C-hevein-like peptides are originated from point mutations in 8C-hevein-like peptides resulting in the insertion of two additional cysteine residues, whereas, the 6C-hevein-like peptides are originated from the C-terminal truncations of 8C-hevein-like peptides. This hypothesis provides an evolutionary pathway of hevein-like peptides for the first time

In conclusion, this thesis presents the discovery, isolation and characterization of novel hevein-like peptides from *A. sessilis* and *M. oleifera* with insights into their binding interaction with chitin oligosaccharides and anti-fungal properties against phyto-pathogenic fungi. In addition, EST-based data mining has expanded the existing library of hevein-like peptides from 20 to 385. This unprecedented number of hevein-like peptides distributed across the plant kingdom further validates the integral role of this peptide family in plant defense. Our data also sheds light on the distribution and evolution of hevein-like peptides and further advances our understanding of this diverse family of CRPs. Taken together, this thesis highlights the molecular diversity of one of the largest families of

constitutively expressed CRPs and opens up new avenues for the development of non-toxic, peptide-based anti-fungal agents that can be of commercial and agricultural value.

Publications and Presentations

1. **Kini, S. G**., Nguyen, P. Q., Weissbach, S., Mallagaray, A., Shin, J., Yoon, H. S., and Tam, J. P. (2015) Studies on the Chitin Binding Property of Novel Cysteine-Rich Peptides from *Alternanthera sessilis*, *Biochemistry 54*, 6639-6649.

2. **Kini, S. G**., Wong, K. H., Tan, W. L., and Tam, J.P., $8C_S$-Hevein-like-Peptides: A Novel Sub-Class of Hevein-like Peptides Isolated from *Moringa oleifera* (in manuscript)

3. **Kini, S. G**., Wong, K. H., Tan, W. L., and Tam, J.P., EST-based in silico identification of novel hevein-like peptides in plants (in manuscript)

4. **Kini, S. G**., Nguyen, P. Q., Weissbach, S., Mallagaray, A., Shin, J., Yoon, H. S., and Tam, J. P Studies on the Chitin Binding Property of Novel Cysteine-Rich Peptides from *A. sessilis,* Peptide Therapeutics Symposium, San Diego, California, USA, Oct 2015

5. **Kini, S. G** and Tam, J. P., Hevein-like Peptides from *Alternanthera sessilis*: Bringing Down the "Walls" of Fungal Pathogens, International Peptide Symposium, Singapore, Dec 2015

References

1. Showalter AM: **Structure and Function of Plant-Cell Wall Proteins**. *The Plant Cell* 1993, **5**(1):9-23.
2. Selitrennikoff CP: **Antifungal Proteins**. *Applied and Environmental Microbiology* 2001, **67**(7):2883-2894.
3. Sels J, Mathys J, De Coninck BM, Cammue BP, De Bolle MF: **Plant Pathogenesis-Related (Pr) Proteins: A Focus on Pr Peptides**. *Plant Physiology and Biochemistry : PPB / Societe francaise de physiologie vegetale* 2008, **46**(11):941-950.
4. van Loon LC, Rep M, Pieterse CM: **Significance of Inducible Defense-Related Proteins in Infected Plants**. *Annual Review of Phytopathology* 2006, **44**:135-162.
5. Fabricant DS, Farnsworth NR: **The Value of Plants Used in Traditional Medicine for Drug Discovery**. *Environmental Health Perspectives* 2001, **109**(Suppl 1):69-75.
6. Rates SMK: **Plants as Source of Drugs**. *Toxicon : Official Journal of the International Society on Toxinology* 2001, **39**(5):603-613.
7. Wick JY: **Aspirin: A History, a Love Story**. *The Consultant Pharmacist : The Journal of the American Society of Consultant Pharmacists* 2012, **27**(5):322-329.
8. Norn S, Kruse PR, Kruse E: **History of Opium Poppy and Morphine**. *Dansk Medicinhistorisk Arbog* 2005, **33**:171-184.
9. Ball P: **History of Science: Quinine Steps Back in Time**. *Nature* 2008, **451**(7182):1065-1066.
10. Craik DJ, Fairlie DP, Liras S, Price D: **The Future of Peptide-Based Drugs**. *Chemical Biology & Drug Design* 2013, **81**(1):136-147.
11. Belting M, Wittrup A: **Developments in Macromolecular Drug Delivery**. *Methods in Molecular Biology* 2009, **480**:1-10.
12. Baumann A: **Early Development of Therapeutic Biologics - Pharmacokinetics**. *Current Drug Metabolism* 2006, **7**(1):15-21.

13. Hammami R, Ben Hamida J, Vergoten G, Fliss I: **Phytamp: A Database Dedicated to Antimicrobial Plant Peptides**. *Nucleic Acids Research* 2009, **37**:D963-D968.
14. Broekaert WF, Cammue BPA, DeBolle MFC, Thevissen K, DeSamblanx GW, Osborn RW: **Antimicrobial Peptides from Plants**. *Critical Reviews in Plant Sciences* 1997, **16**(3):297-323.
15. Colgrave ML, Craik DJ: **Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot**. *Biochemistry* 2004, **43**(20):5965-5975.
16. Thomma BPHJ, Cammue BPA, Thevissen K: **Plant Defensins**. *Planta* 2002, **216**(2):193-202.
17. Hammami R, Ben Hamida J, Vergoten G, Fliss I: **Phytamp: A Database Dedicated to Antimicrobial Plant Peptides**. *Nucleic Acids Research* 2009, **37**(Database issue):4.
18. Kramer KJ, Klassen LW, Jones BL, Speirs RD, Kammer AE: **Toxicity of Purothionin and Its Homologs to the Tobacco Hornworm, Manduca-Sexta (L) (Lepidoptera-Sphingidae)**. *Toxicology and Applied Pharmacology* 1979, **48**(1):179-183.
19. Bloch C, Richardson M: **A New Family of Small (5 Kda) Protein Inhibitors of Insect Alpha-Amylases from Seeds or Sorghum (Sorghum-Bicolor (L) Moench) Have Sequence Homologies with Wheat Gamma-Purothionins**. *FEBS Letters* 1991, **279**(1):101-104.
20. Carrasco L, Vazquez D, Hernandezlucas C, Carbonero P, Garciaolmedo F: **Thionins - Plant Peptides That Modify Membrane-Permeability in Cultured Mammalian-Cells**. *European Journal of Biochemistry* 1981, **116**(1):185-189.
21. Lay FT, Anderson MA: **Defensins--Components of the Innate Immune System in Plants**. *Current Protein & Peptide Science* 2005, **6**(1):85-101.
22. Fernandez de Caleya R, Gonzalez-Pascual B, Garcia-Olmedo F, Carbonero P: **Susceptibility of Phytopathogenic Bacteria to Wheat Purothionins in Vitro**. *Applied Microbiology* 1972, **23**(5):998-1000.

23. Cammue BPA, Debolle MFC, Terras FRG, Proost P, Vandamme J, Rees SB, Vanderleyden J, Broekaert WF: **Isolation and Characterization of a Novel Class of Plant Antimicrobial Peptides from Mirabilis-Jalapa L Seeds**. *Journal of Biological Chemistry* 1992, **267**(4):2228-2233.
24. Craik DJ, Daly NL, Bond T, Waine C: **Plant Cyclotides: A Unique Family of Cyclic and Knotted Proteins That Defines the Cyclic Cystine Knot Structural Motif**. *Journal of Molecular Biology* 1999, **294**(5):1327-1336.
25. Tailor RH, Acland DP, Attenborough S, Cammue BPA, Evans IJ, Osborn RW, Ray JA, Rees SB, Broekaert WF: **A Novel Family of Small Cysteine-Rich Antimicrobial Peptides from Seed of Impatiens Balsamina Is Derived from a Single Precursor Protein**. *Journal of Biological Chemistry* 1997, **272**(39):24480-24487.
26. Jiménez-Barbero J, Javier Cañada F, Asensio JL, 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**. *Advances in Carbohydrate Chemistry and Biochemistry* 2006, **60**:303-354.
27. Archer BL, Cockbain EG: **The Proteins of Hevea Brasiliensis Latex. 2. Isolation of the Alpha-Globulin of Fresh Latex Serum**. *The Biochemical Journal* 1955, **61**(3):508-512.
28. Archer BL, Sekhar BC: **The Proteins of Hevea Brasiliensis Latex. I. Protein Constituents of Fresh Latex Serum**. *The Biochemical Journal* 1955, **61**(3):503-508.
29. Dauzac J, Cretin H, Marin B, Lioret C: **A Plant Vacuolar System - the Lutoids from Hevea-Brasiliensis Latex**. *Physiologie Vegetale* 1982, **20**(2):311-331.
30. Wang XC, Shi MJ, Wang D, Chen YY, Cai FG, Zhang SX, Wang LM, Tong Z, Tian WM: **Comparative Proteomics of Primary and Secondary Lutoids Reveals That Chitinase and Glucanase Play a Crucial Combined Role in Rubber Particle Aggregation in Hevea Brasiliensis**. *Journal of Proteome Research* 2013, **12**(11):5146-5159.

31. Archer BL: **The Proteins of Hevea Brasiliensis Latex Isolation and Characterization of Crystalline Hevein**. *Journal of Biochemistry* 1960:236-240.
32. Walujono K, Scholma, R. A., Beintema, J. J., Mariono, A. &, Hahn AM: **Amino Acid Sequence of Hevein**. *Proceedings of the International Rubber Conference Kuala Lampur* 1975, **II**:518-531.
33. Boller T, Metraux JP: **Extracellular Localization of Chitinase in Cucumber**. *Physiological and Molecular Plant Pathology* 1988, **33**(1):11-16.
34. Shinshi H, Mohnen D, Meins F: **Regulation of a Plant Pathogenesis-Related Enzyme - Inhibition of Chitinase and Chitinase Messenger-Rna Accumulation in Cultured Tobacco Tissues by Auxin and Cytokinin**. *Proceedings of the National Academy of Sciences USA* 1987, **84**(1):89-93.
35. Rice RH, Etzler ME: **Subunit Structure of Wheat-Germ Agglutinin**. *Biochemical and Biophysical Research Communications* 1974, **59**(1):414-419.
36. Peumans WJ, De Ley, M., Broekaert, W.F. : **An Unusual lectin from Stinging Nettle (Urtica Dioica) Rhizomes**. *FEBS Letters* 1983, **177**:99-103.
37. Broekaert I, Lee HI, Kush A, Chua NH, Raikhel N: **Wound-Induced Accumulation of Mrna Containing a Hevein Sequence in Laticifers of Rubber Tree (Hevea Brasiliensis)**. *Proceedings of the Natural Academy of Sciences USA* 1990, **87**(19):7633-7637.
38. Jan Van Parijs WFB, Irwin J. Goldstein , and Willy J. Peumans **Hevein an Antifungal Protein from Rubber-Tree**. *Planta* 1990, **258-264**.
39. Van der Weerden NL, Anderson MA: **Plant Defensins: Common Fold, Multiple Functions**. *Fungal Biology Reviews* 2013, **26**(4):121-131.
40. Nguyen GK, Zhang S, Nguyen NT, Nguyen PQ, Chiu MS, Hardjojo A, Tam JP: **Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain**

a and Cyclotide Domains in the Fabaceae Family. *The Journal of Biological Chemistry* 2011, **286**(27):24275-24287.

41. Rodriguezromero A, Ravichandran KG, Sorianogarcia M: **Crystal-Structure of Hevein at 2.8 a Resolution**. *FEBS Letters* 1991, **291**(2):307-309.
42. Reyes-Lopez CA, Hernandez-Santoyo A, Pedraza-Escalona M, Mendoza G, Hernandez-Arana A, Rodriguez-Romero A: **Insights into a Conformational Epitope of Hev B 6.02 (Hevein)**. *Biochemical and Biophysical Research Communications* 2004, **314**(1):123-130.
43. Andersen NH, Cao B, Rodriguezromero A, Arreguin B: **Hevein - Nmr Assignment and Assessment of Solution-State Folding for the Agglutinin-Toxin Motif**. *Biochemistry* 1993, **32**(6):1407-1422.
44. Asensio JL, Canada FJ, Bruix M, Rodriguezromero A, Jimenezbarbero J: **The Interaction of Hevein with N-Acetylglucosamine-Containing Oligosaccharides - Solution Structure of Hevein Complexed to Chitobiose**. *European Journal of Biochemistry* 1995, **230**(2):621-633.
45. Willem F. Broekaert WM, Franky R. G. Terras, Miguel F. C. De Belle, Paul Proost,, Jozef Van Damme LD, Magda Claeys, Sarah B. Rees, Jozef Vanderleyden, and, Cammue BPA: **Antimicrobial Peptides from Amaranthus Caudatus Seeds with Sequence Homology to the Cysteine Glycine-Rich Domain of Chitin-Binding Proteins**. *Biochemistry* l992, **31**(17):4308-4314.
46. Klaus K. Nielsen JEN, Susan M. Madrid, and Jorn D. Mikkelsen: **Characterization of a New Antifungal Chitin-Binding Peptide from Sugar Beet Leaves**. *Plant Physiology* 1997:83-91.
47. Ja Choon Koo SYL, Hyun Jin Chun , Yong Hwa Cheong, Jae Su Choi,, Shun-ichiro Kawabata MM, Susumu Tsunasawa, Kwon Soo Ha,, Dong Won Bae C-dH, Bok Luel Lee, Moo Je Cho **Two Hevein Homologs Isolated from the Seed of Pharbitis Nil L. Exhibit Potent Anti-Fungal Activity**. *Biochimica et Biophysica Acta* 1998:80-90.

48. Ren-Huai Huanga YX, Xiao-Zhu Liub, Ying Zhanga, Zhong Hub, Da-Cheng Wanga: **Two Novel Antifungal Peptides Distinct with a Five-Disulfide Motif from the Bark of Eucommia Ulmoides Oliv**. *FEBS Letters* 2002, **521**:87-90.
49. James P. Tam SW, Ka H. Wong, Wei Liang Tan: **Antimicrobial Peptides from Plants**. *Pharmaceuticals* 2015, **8**.
50. 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**(20):2426-2431.
51. 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.)**. *Bioscience, Biotechnology, and Biochemistry* 2003, **67**(8):1636-1642.
52. Van den Bergh KP, Van Damme EJ, Peumans WJ, Coosemans J: **Ee-Cbp, a Hevein-Type Antimicrobial Peptide from Bark of the Spindle Tree (Euonymus Europaeus L.)**. *Mededelingen* 2002, **67**(2):327-331.
53. Odintsova TI, Vassilevski AA, Slavokhotova AA, Musolyamov AK, Finkina EI, Khadeeva NV, Rogozhin EA, Korostyleva TV, Pukhalsky VA, Grishin EV *et al*: **A Novel Antifungal Hevein-Type Peptide from Triticum Kiharae Seeds with a Unique 10-Cysteine Motif**. *The FEBS Journal* 2009, **276**(15):4266-4275.
54. Yang YF, Cheng KC, Tsai PH, Liu CC, Lee TR, Lyu PC: **Alanine Substitutions of Noncysteine Residues in the Cysteine-Stabilized Alphabeta Motif**. *Protein Science : A Publication of the Protein Society* 2009, **18**(7):1498-1506.
55. Balaji RA, Ohtake A, Sato K, Gopalakrishnakone P, Kini RM, Seow KT, Bay BH: **Lambda-Conotoxins, a New Family of Conotoxins with Unique Disulfide Pattern and Protein Folding. Isolation and**

Characterization from the Venom of Conus Marmoreus. *The Journal of Biological Chemistry* 2000, **275**(50):39516-39522.

56. Wright HT, Sandrasegaram G, Wright CS: **Evolution of a Family of N-Acetylglucosamine Binding Proteins Containing the Disulfide-Rich Domain of Wheat Germ Agglutinin**. *Journal of Molecular Evolution* 1991, **33**(3):283-294.
57. Drenth J, Low, B. W., Richardson, J. S., and Wright, C. S.: **The Toxin-Agglutinin Fold - a New Group of Small Protein Structures Organized around a 4-Disulfide Core**. *Journal of Biological Chemistry* 1980, **255**:2652-2655.
58. Juan L Asensio FJC, Hans-Christian Siebert, José Laynez,, Ana Poveda PMN, UM Soedjanaamadja, Hans-Joachim Gabius, Jiménez-Barbero aJ: **Structural Basis for Chitin Recognition by Defense Proteins: Glcnac Residues Are Bound in a Multivalent Fashion by Extended Binding Sites in Hevein Domains.** *Chemistry & Biology* 2000, **7**:529–543.
59. Wright HT, Brooks DM, Wright CS: **Evolution of the Multidomain Protein Wheat Germ Agglutinin**. *Journal of Molecular Evolution* 1984, **21**(2):133-138.
60. Beintema JJ: **Structural Features of Plant Chitinases and Chitin-Binding Proteins**. *FEBS Letters* 1994, **350**(2-3):159-163.
61. Soedjanaatmadja UM, Hofsteenge J, Jeronimus-Stratingh CM, Bruins AP, Beintema JJ: **Demonstration by Mass Spectrometry That Pseudo-Hevein and Hevein Have Ragged C-Terminal Sequences**. *Biochimica et Biophysica Acta* 1994, **1209**(1):144-148.
62. Diaz-Perales A, Collada C, Blanco C, Sanchez-Monge R, Carrillo T, Aragoncillo C, Salcedo G: **Class I Chitinases with Hevein-Like Domain, but Not Class Ii Enzymes, Are Relevant Chestnut and Avocado Allergens**. *The Journal of Allergy and Clinical Immunology* 1998, **102**(1):127-133.
63. Wagner S, Breiteneder H: **The Latex-Fruit Syndrome**. *Biochemical Society Transactions* 2002, **30**(Pt 6):935-940.

64. Karisola P, Alenius H, Mikkola J, Kalkkinen N, Helin J, Pentikainen OT, Repo S, Reunala T, Turjanmaa K, Johnson MS *et al*: **The Major Conformational Ige-Binding Epitopes of Hevein (Hev B6.02) Are Identified by a Novel Chimera-Based Allergen Epitope Mapping Strategy**. *The Journal of Biological Chemistry* 2002, **277**(25):22656-22661.

65. Karisola P, Mikkola J, Kalkkinen N, Airenne KJ, Laitinen OH, Repo S, Pentikainen OT, Reunala T, Turjanmaa K, Johnson MS *et al*: **Construction of Hevein (Hev B 6.02) with Reduced Allergenicity for Immunotherapy of Latex Allergy by Comutation of Six Amino Acid Residues on the Conformational Ige Epitopes**. *Journal of Immunology* 2004, **172**(4):2621-2628.

66. Andersen NH, Cao B, Rodriguez-Romero A, Arreguin B: **Hevein: Nmr Assignment and Assessment of Solution-State Folding for the Agglutinin-Toxin Motif**. *Biochemistry* 1993, **32**(6):1407-1422.

67. Paul K. Pallaghy KJN, David J. Craik,, Norton RS: **A Common Structural Motif Incorporating a Cystine Knot and a Triple-Stranded B-Sheet in Toxic and Inhibitory Polypeptides**. *Protein Science* 1994:1833-1839.

68. Zhu S, Darbon H, Dyason K, Verdonck F, Tytgat J: **Evolutionary Origin of Inhibitor Cystine Knot Peptides**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2003, **17**(12):1765-1767.

69. Martins JC, Maes D, Loris R, Pepermans HAM, Wyns L, Willem R, Verheyden P: **H-1 Nmr Study of the Solution Structure of Ac-Amp2, a Sugar Binding Antimicrobial Protein Isolated from Amaranthus Caudatus**. *Journal of Molecular Biology* 1996, **258**(2):322-333.

70. Van den Bergh KP, Rouge P, Proost P, Coosemans J, Krouglova T, Engelborghs Y, Peumans WJ, Van Damme EJ: **Synergistic Antifungal Activity of Two Chitin-Binding Proteins from Spindle Tree (Euonymus Europaeus L.)**. *Planta* 2004, **219**(2):221-232.

71. 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 Fivedisulfide Motif**. *Biochemistry* 2004, **43**:6005-6012.

72. Xiang Y, Huang RH, Liu XZ, Zhang Y, Wang DC: **Crystal Structure of a Novel Antifungal Protein Distinct with Five Disulfide Bridges from Eucommia Ulmoides Oliver at an Atomic Resolution**. *Journal of structural biology* 2004, **148**(1):86-97.

73. Dubovskii PV, Vassilevski AA, Slavokhotova AA, Odintsova TI, Grishin EV, Egorov TA, Arseniev AS: **Solution Structure of a Defense Peptide from Wheat with a 10-Cysteine Motif**. *Biochemical and Biophysical Research Communications* 2011, **411**(1):14-18.

74. Muraki M, Morii H, Harata K: **Chemically Prepared Hevein Domains: Effect of C-Terminal Truncation and the Mutagenesis of Aromatic Residues on the Affinity for Chitin**. *Protein Engineering* 2000, **13**(6):385-389.

75. Asensio JL, Canada FJ, 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**(6):569-577.

76. Asensio JL, Siebert HC, von Der Lieth CW, Laynez J, Bruix M, Soedjanaamadja UM, Beintema JJ, Canada FJ, Gabius HJ, Jimenez-Barbero J: **Nmr Investigations of Protein-Carbohydrate Interactions: Studies on the Relevance of Trp/Tyr Variations in Lectin Binding Sites as Deduced from Titration Microcalorimetry and Nmr Studies on Hevein Domains. Determination of the Nmr Structure of the Complex between Pseudohevein and N,N',N"-Triacetylchitotriose**. *Proteins* 2000, **40**(2):218-236.

77. Juan Felix Espinosa JLA, Jose Luis Garcia, Jose Laynez, Marta Bruix, Christine Wright,, Hans-Christian Siebert H-JG, Francisco Javier Canada and Jesus Jimenez-Barbero: **Nmr Investigations of Protein-**

Carbohydrate Interactions Binding Studies and Refined Three-Dimensional Solution Structure of the Complex between the B Domain of Wheat Germ Agglutinin and N,N',N''-Triacetylchitotriose. *European Journal of Biochemistry* 2000, **267**:3965-3978.

78. Verheyden PP, Jurgen Maes, Dominique Pepermans, Henri A M Wyns, Lode: **^{1}h Nmr Study of the Interaction of Triacetyl Chitotriose with Ac-Amp2, a Sugar Binding Antimicrobial Protein Isolated from *A.Caudatus***. *FEBS Letters* 1995, **370**:245-249.
79. Aboitiz N, Vila-Perello M, Groves P, Asensio JL, Andreu D, Canada FJ, 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 : A European Journal of Chemical Biology* 2004, **5**(9):1245-1255.
80. Chavez MI, Vila-Perello M, Canada FJ, 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**. *Carbohydrate Research* 2010, **345**(10):1461-1468.
81. Garcia-Hernandez E, Zubillaga RA, Rojo-Dominguez A, Rodriguez-Romero A, Hernandez-Arana A: **New Insights into the Molecular Basis of Lectin-Carbohydrate Interactions: A Calorimetric and Structural Study of the Association of Hevein to Oligomers of N-Acetylglucosamine**. *Proteins* 1997, **29**(4):467-477.
82. Wright CS: **2.2 a Resolution Structure Analysis of Two Refined N-Acetylneuraminyl-Lactose--Wheat Germ Agglutinin Isolectin Complexes**. *Journal of Molecular Biology* 1990, **215**(4):635-651.
83. Wright CS: **Structural Comparison of the Two Distinct Sugar Binding Sites in Wheat Germ Agglutinin Isolectin Ii**. *Journal of Molecular Biology* 1984, **178**(1):91-104.
84. Wright CS: **Crystal-Structure of a Wheat-Germ-Agglutinin Glycophorin-Sialoglycopeptide Receptor Complex - Structural Basis**

for Cooperative Lectin-Cell Binding. *Journal of Biological Chemistry* 1992, **267**(20):14345-14352.

85. Muraki M: **The Importance of Ch/Pi Interactions to the Function of Carbohydrate Binding Proteins**. *Protein and Peptide Letters* 2002, **9**(3):195-209.
86. De Bolle MFC, David KMM, Rees SB, Vanderleyden J, Cammue BPa, Broekaert WF: **Cloning and Characterization of a Cdna Encoding an Antimicrobial Chitin-Binding Protein from Amaranth, *Amaranthus Caudatus***. *Plant Molecular Biology* 1993, **22**(6):1187-1190.
87. Kini SG, Nguyen PQ, Weissbach S, Mallagaray A, Shin J, Yoon HS, Tam JP: **Studies on the Chitin Binding Property of Novel Cysteine-Rich Peptides from Alternanthera Sessilis**. *Biochemistry* 2015, **54**(43):6639-6649.
88. Broekaert I, Lee HI, Kush A, Chua NH, Raikhel N: **Wound-Induced Accumulation of Mrna Containing a Hevein Sequence in Laticifers of Rubber Tree (Hevea Brasiliensis)**. *Proceedings of the National Academy of Sciences USA* 1990, **87**(19):7633-7637.
89. Andreev YA, Korostyleva TV, Slavokhotova AA, Rogozhin EA, Utkina LL, Vassilevski AA, Grishin EV, Egorov TA, Odintsova TI: **Genes Encoding Hevein-Like Defense Peptides in Wheat: Distribution, Evolution, and Role in Stress Response**. *Biochimie* 2012, **94**(4):1009-1016.
90. Jennings C, West J, Waine C, Craik D, Anderson M: **Biosynthesis and Insecticidal Properties of Plant Cyclotides: The Cyclic Knotted Proteins from Oldenlandia Affinis**. *Proceedings of the National Academy of Sciences USA* 2001, **98**(19):10614-10619.
91. Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E: **A Novel Family in Medicago Truncatula Consisting of More Than 300 Nodule-Specific Genes Coding for Small, Secreted Polypeptides with Conserved Cysteine Motifs**. *Plant Physiology* 2003, **132**(1):161-173.

92. Giang KTN, Lian YL, Pang EWH, Phuong QTN, Tran TD, Tam JP: **Discovery of Linear Cyclotides in Monocot Plant Panicum Laxum of Poaceae Family Provides New Insights into Evolution and Distribution of Cyclotides in Plants**. *Journal of Biological Chemistry* 2013, **288**(5):3370-3380.
93. Giang KTN, Zhang S, Ngan TKN, Phuong QTN, Chiu MS, Hardjojo A, Tam JP: **Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family**. *Journal of Biological Chemistry* 2011, **286**(27):24275-24287.
94. Nguyen PQ, Wang S, Kumar A, Yap LJ, Luu TT, Lescar J, Tam JP: **Discovery and Characterization of Pseudocyclic Cystine-Knot Alpha-Amylase Inhibitors with High Resistance to Heat and Proteolytic Degradation**. *The FEBS Journal* 2014, **281**(19):4351-4366.
95. Shinshi H, Wenzler H, Neuhaus JM, Felix G, Hofsteenge J, Meins F: **Evidence for N-Terminal and C-Terminal Processing of a Plant Defense-Related Enzyme - Primary Structure of Tobacco Prepro-Beta-1,3-Glucanase**. *Proceedings of the National Academy of Sciences USA* 1988, **85**(15):5541-5545.
96. Broglie KE, Gaynor JJ, Broglie RM: **Ethylene-Regulated Gene Expression: Molecular Cloning of the Genes Encoding an Endochitinase from Phaseolus Vulgaris**. *Proceedings of the National Academy of Sciences USA* 1986, **83**(18):6820-6824.
97. De Bolle MF, Osborn RW, Goderis IJ, Noe L, Acland D, Hart CA, Torrekens S, Van Leuven F, Broekaert WF: **Antimicrobial Peptides from Mirabilis Jalapa and Amaranthus Caudatus: Expression, Processing, Localization and Biological Activity in Transgenic Tobacco**. *Plant Molecular Biology* 1996, **31**(5):993-1008.
98. Koo JC, Chun HJ, Park HC, Kim MC, Koo YD, Koo SC, Ok HM, Park SJ, Lee SH, Yun DJ *et al*: **Over-Expression of a Seed Specific Hevein-Like Antimicrobial Peptide from Pharbitis Nil Enhances Resistance to a**

Fungal Pathogen in Transgenic Tobacco Plants. *Plant Molecular Biology* 2002, **50**(3):441-452.

99. Van den Bergh KP, Proost P, Van Damme J, Coosemans J, Van Damme EJ, Peumans WJ: **Five Disulfide Bridges Stabilize a Hevein-Type Antimicrobial Peptide from the Bark of Spindle Tree (Euonymus Europaeus L.)**. *FEBS Letters* 2002, **530**(1-3):181-185.
100. Wessels JGH: **A Steady-State Model for Apical Wall Growth in Fungi**. *Acta Botanica Neerlandica* 1988, **37**(1):3-16.
101. Slavokhotova AA, Naumann TA, Price NP, Rogozhin EA, Andreev YA, Vassilevski AA, Odintsova TI: **Novel Mode of Action of Plant Defense Peptides - Hevein-Like Antimicrobial Peptides from Wheat Inhibit Fungal Metalloproteases**. *The FEBS Journal* 2014, **281**(20):4754-4764.
102. Malanovic N, Lohner K: **Gram-Positive Bacterial Cell Envelopes: The Impact on the Activity of Antimicrobial Peptides**. *Biochimica et Biophysica Acta* 2016, **1858**(5):936-946.
103. Anitha R, and Kanimozhi, S: **Pharmacognostic Evaluation of Alternanthera Sessilis (L.) R.Br.Ex.Dc**. *Pharmacognosy Journal* 2012, **4**:31-34.
104. Chandrika UG, Svanberg U, Jansz ER: **In Vitro Accessibility of Beta-Carotene from Cooked Sri Lankan Green Leafy Vegetables and Their Estimated Contribution to Vitamin a Requirement**. *Journal of the Science of Food and Agriculture* 2006, **86**(1):54-61.
105. Mehrotra NN OV: **Ayurvedic Rasayana Therapy and Rejuvenation (Kaya Kalp).** *Industrial Crops and Products* 2006, **19**:133-136.
106. Gayathri BM, Balasuriya K, Gunawardena GSPD, Rajapakse RPVJ, Dharmaratne HRW: **Toxicological Studies of the Water Extract of Green Leafy Vegetable Sessile Joy Weed (Alternanthera Sessilis)**. *Current Science* 2006, **91**(11):1517-1520.
107. Acharya EP, B.: **Ethno-Medicinal Plants Used by Bantar of Bhaudaha, Morang, Nepal**. *Our Nature* 2006, **4**:96-103.

108. M SK, G, Silpa Rani, Astalakshmi, N, Manasa, G, Vanaja, P, Sirisha, G, Slvvsnk, Swaroop Kumar: **Screening of Aqueous and Ethanolic Extracts of Aerial Parts of A.Sessilis for Antidiabetic Activity**. *Journal of Pharmacy Research* 2011, **4**(5):1528-1530.
109. Subhashini T, Krishnaveni B, C SR: **Anti- Inflammatory Activity of Leaf Extracts of Alternanthera Sessilis**. *Hygeia Journal for Drugs and Medicines* 2010, **2**(1):54-56.
110. Walter M, Merish, S, Tamizhamuthu, M: **Review of Alternanthera Sessilis with Reference to Traditional Siddha Medicine**. *International Journal of Pharmacognosy and Phytochemical Research* 2014, **6**(2):249-254.
111. Anwar F BM: **Analytical Characterization of *Moringa Oleifera* Seed Oil Grown in Temperate Regions of Pakistan**. *Phytotherapy Research* 2003, **20** 906–910.
112. Ramachandran C, Peter KV, Gopalakrishnan PK: **Drumstick (Moringa-Oleifera) - a Multipurpose Indian Vegetable**. *Economic Botany* 1980, **34**(3):276-283.
113. Mughal MH AG, Srivastava PS, Iqbal M. : **Improvement of Drumstick (Moringa Pterygosperma Gaertn.) – a Unique Source of Food and Medicine through Tissue Culture**. *Hamdard Medicine* 1999, **42**:37–42.
114. Somali MA BM, Al-Faimani SS: **Chemical Composition and Characteristics of *Moringa Peregrina* Seeds and Seed Oil**. *Journal of the American Oil Chemists' Society* 1984, **61**:85–86.
115. Faizi S SB, Saleem R, Aftab K, Shaheen F, Gilani AH: **Hypotensive Constituents from the Pods of *Moringa Oleifera***. *Planta Medica* 1998, **64**:225-228.
116. Ruckmani K KS, Anandan R, Jaykar B.: **Antioxidant Properties of Various Solvent Extracts of Total Phenolic Constituents from Three Different Agro-Climatic Origins of Drumstick Tree**. *Indian Journal of Pharmaceutical Science* 1998, **60**:33-35.

117. Palada MC: **Moringa (Moringa Oleifera Lam): A Versatile Tree Crop with Horticultural Potential in the Subtropical United States**. *Hortscience* 1996, **31**(5):794-797.
118. Anwar F, Latif S, Ashraf M, Gilani AH: **Moringa Oleifera: A Food Plant with Multiple Medicinal Uses**. *Phytotherapy Research* 2007, **21**(1):17-25.
119. Nielsen H, Engelbrecht J, Brunak S, vonHeijne G: **Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of Their Cleavage Sites**. *Protein Engineering* 1997, **10**(1):1-6.
120. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved Prediction of Signal Peptides: Signalp 3.0**. *Journal of Molecular Biology* 2004, **340**(4):783-795.
121. Crooks GE, Hon G, Chandonia JM, Brenner SE: **Weblogo: A Sequence Logo Generator**. *Genome Research* 2004, **14**(6):1188-1190.
122. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T: **How to Measure and Predict the Molar Absorption-Coefficient of a Protein**. *Protein Science* 1995, **4**(11):2411-2423.
123. Jeener J, Meier BH, Bachmann P, Ernst RR: **Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy**. *Journal of Chemical Physics* 1979, **71**(11):4546-4553.
124. Kumar A, Ernst RR, Wuthrich K: **A Two-Dimensional Nuclear Overhauser Enhancement (2d Noe) Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules**. *Biochemical and Biophysical Research Communications* 1980, **95**(1):1-6.
125. Davis DG, Bax A: **Assignment of Complex H-1-Nmr Spectra Via Two-Dimensional Homonuclear Hartmann-Hahn Spectroscopy**. *Journal of the American Chemical Society* 1985, **107**(9):2820-2821.
126. Bax A: **Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy**. *Journal of Magnetic Resonance (1969)* 1985, **65**(2):355-360.

127. Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wuthrich K: **Improved Spectral Resolution in Cosy 1h Nmr Spectra of Proteins Via Double Quantum Filtering**. *Biochemical and Biophysical Research Communications* 1983, **117**(2):479-485.
128. 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**. *Journal of Molecular Biology* 1983, **169**(4):949-961.
129. 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**(1):125-129.
130. Hwang TL, Shaka AJ: **Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients**. *Journal of Magnetic Resonance, Series A* 1995, **112**(2):275-279.
131. 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**(3):277-293.
132. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*: **Crystallography & Nmr System: A New Software Suite for Macromolecular Structure Determination**. *Acta Crystallographica Section D Biological Crystallography* 1998, **54**(Pt 5):905-921.
133. Pardi A, Billeter M, Wüthrich K: **Calibration of the Angular Dependence of the Amide Proton-Cα Proton Coupling Constants, 3jhnα, in a Globular Protein**. *Journal of Molecular Biology* 1984, **180**(3):741-751.
134. Driscoll PC, Gronenborn AM, Beress L, Clore GM: **Determination of the Three-Dimensional Solution Structure of the Antihypertensive and Antiviral Protein Bds-I from the Sea Anemone Anemonia Sulcata: A Study Using Nuclear Magnetic Resonance and Hybrid Distance**

Geometry-Dynamical Simulated Annealing. *Biochemistry* 1989, **28**(5):2188-2198.

135. Nilges M, Clore GM, Gronenborn AM: **Determination of Three-Dimensional Structures of Proteins from Interproton Distance Data by Hybrid Distance Geometry-Dynamical Simulated Annealing Calculations**. *FEBS Letters* 1988, **229**(2):317-324.
136. Nilges M, Gronenborn A, Brünger A, Clore M: **Determination of Three-Dimensional Structures of Proteins by Simulated Annealing with Interproton Distance Restraints. Application to Crambin, Potato Carboxypeptidase Inhibitor and Barley Serine Proteinase Inhibitor 2**. *Protein Engineering* 1988, **2**(1):27-38.
137. Nilges M: **A Calculation Strategy for the Structure Determination of Symmetric Demers by 1h Nmr**. *Proteins: Structure, Function, and Bioinformatics* 1993, **17**(3):297-309.
138. Koradi R, Billeter M, Wüthrich K: **Molmol: A Program for Display and Analysis of Macromolecular Structures**. *Journal of Molecular Graphics* 1996, **14**(1):51-55.
139. **The Pymol Molecular Graphics System, Version 1.7.4 Schrödinger, Llc**.
140. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **Procheck: A Program to Check the Stereochemical Quality of Protein Structures**. *Journal of Applied Crystallography* 1993, **26**(2):283-291.
141. Williamson MP: **Using Chemical Shift Perturbation to Characterise Ligand Binding**. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2013, **73**:1-16.
142. Jenssen H AS: **Serum Stability of Peptides**. *Methods Mol Biol* 2008, **494**:177-186.
143. Ye XY, Ng TB: **A New Antifungal Peptide from Rice Beans**. *The journal of peptide research : official journal of the American Peptide Society* 2002, **60**(2):81-87.

144. Wiegand I, Hilpert K, Hancock RE: **Agar and Broth Dilution Methods to Determine the Minimal Inhibitory Concentration (Mic) of Antimicrobial Substances**. *Nature Protocols* 2008, **3**(2):163-175.
145. Marsh AJ, O'Sullivan O, Ross RP, Cotter PD, 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.
146. Porto WF, Souza VA, Nolasco DO, Franco OL: **In Silico Identification of Novel Hevein-Like Peptide Precursors**. *Peptides* 2012, **38**(1):127-136.
147. Boguski MS, Lowe TM, Tolstoshev CM: **Dbest--Database for "Expressed Sequence Tags"**. *Nature Genetics* 1993, **4**(4):332-333.
148. Matasci N, Hung LH, Yan Z, Carpenter EJ, Wickett NJ, 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.
149. Petersen TN, Brunak S, von Heijne G, Nielsen H: **Signalp 4.0: Discriminating Signal Peptides from Transmembrane Regions**. *Nature Methods* 2011, **8**(10):785-786.
150. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R: **Analysis Tool Web Services from the Embl-Ebi**. *Nucleic Acids Research* 2013, **41**(Web Server issue):W597-600.
151. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2go: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research**. *Bioinformatics* 2005, **21**(18):3674-3676.
152. 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 Research* 2016.
153. Byng MJMCaJW: **The Number of Known Plants Species in the World and Its Annual Increase**. *Phytotaxa* 2016, **261**(3):201-217.
154. Niraimathi KL, Sudha V, Lavanya R, Brindha P: **Biosynthesis of Silver Nanoparticles Using Alternanthera Sessilis (Linn.) Extract and Their**

Antimicrobial, Antioxidant Activities. *Colloids and surfaces B, Biointerfaces* 2013, **102**:288-291.

155. Hossain AI, Faisal M, Rahman S, Jahan R, Rahmatullah M: **A Preliminary Evaluation of Antihyperglycemic and Analgesic Activity of Alternanthera Sessilis Aerial Parts**. *BMC Complement Alternative Medicine* 2014, **14**:169.
156. Gunasekera L: **Sessile Joyweed (Alternanthera Sessilis) a Popular Leafy Vegetable in South East Asia but Federal Noxious Weed in USA**. *Sixteenth Australian Weeds Conference* 2008:347-348.
157. Nguyen GKT, Lim WH, Nguyen PQT, Tam JP: **Novel Cyclotides and Uncyclotides with Highly Shortened Precursors from Chassalia Chartacea and Effects of Methionine Oxidation on Bioactivities**. *Journal of Biological Chemistry* 2012, **287**(21):17598-17607.
158. Nguyen PQ, Luu TT, Bai Y, Nguyen GK, Pervushin K, Tam JP: **Allotides: Proline-Rich Cystine Knot Alpha-Amylase Inhibitors from Allamanda Cathartica**. *Journal of Natural Products* 2015, **78**(4):695-704.
159. 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.)**. *The EMBO Journal* 1983, **2**(7):1035-1040.
160. Nguyen GKT, Zhang S, Wang W, Wong CTT, Nguyen NTK, Tam JP: **Discovery of a Linear Cyclotide from the Bracelet Subfamily and Its Disulfide Mapping by Top-Down Mass Spectrometry**. *The Journal of Biological Chemistry* 2011, **286**(52):44833-44844.
161. Ireland DC, Colgrave ML, Craik DJ: **A Novel Suite of Cyclotides from Viola Odorata: Sequence Variation and the Implications for Structure, Function and Stability**. *The Biochemical Journal* 2006, **400**(1):1-12.
162. Davids BJ, Reiner DS, Birkeland SR, Preheim SP, Cipriano MJ, McArthur AG, Gillin FD: **A New Family of Giardial Cysteine-Rich Non-Vsp Protein Genes and a Novel Cyst Protein**. *Plos One* 2006, **1**(1).

163. Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA: **Biosynthesis of Circular Proteins in Plants**. *Plant Journal* 2008, **53**(3):505-515.
164. Shubin N, Tabin C, Carroll S: **Deep Homology and the Origins of Evolutionary Novelty**. *Nature* 2009, **457**(7231):818-823.
165. Franco OL, Rigden DJ, Melo FR, Grossi-De-Sa MF: **Plant Alpha-Amylase Inhibitors and Their Interaction with Insect Alpha-Amylases**. *European Journal of Biochemistry* 2002, **269**(2):397-412.
166. Juan Luis Asensio FJC, Marta Bruix CG, Noureddine Khiar,, Jiménez-Barbero AR-RaJ: **Nmr Investigations of Protein–Carbohydrate Interactions Refined Three-Dimensional Structure of the Complex between Hevein and Methyl Chitobioside**. *Glycobiology* 1998, **8**(6):569–577.
167. JF M: **The Horseradish Tree, Moringa Pterigosperma (Moringaceae). A Boon to Arid Lands**. *Economic Botany* 1991, **45**: 318–333.
168. Mehta LK BR, Amin AH, Bafna PA, Gulati OD: **Effect of Fruits of Moringa Oleifera on the Lipid Profile of Normal and Hypercholesterolaemic Rabbits**. *Journal of Ethnopharmacology* 2003, **86**:191–195.
169. Ndabigengesere A NK, Talbot BG: **Active Agents and Mechanism of Coagulation of Turbid Waters Using *Moringa Oleifera***. *Water Research* 1995, **29**:703-710.
170. Araki T, Funatsu J, Kuramoto M, Konno H, Torikata T: **The Complete Amino Acid Sequence of Yam (Dioscorea Japonica) Chitinase. A Newly Identified Acidic Class I Chitinase**. *The Journal of Biological Chemistry* 1992, **267**(28):19944-19947.
171. Iseli B, Boller T, Neuhaus JM: **The N-Terminal Cysteine-Rich Domain of Tobacco Class I Chitinase Is Essential for Chitin Binding but Not for Catalytic or Antifungal Activity**. *Plant Physiology* 1993, **103**(1):221-226.
172. Hernandezlucas C, Royo J, Pazares J, Ponz F, Garciaolmedo F, Carbonero P: **Polyadenylation Site Heterogeneity in Messenger-Rna**

Encoding the Precursor of the Barley Toxin Beta-Hordothionin. *FEBS Letters* 1986, **200**(1):103-106.

173. Bednarek SY, Wilkins TA, Dombrowski JE, Raikhel NV: **A Carboxyl-Terminal Propeptide Is Necessary for Proper Sorting of Barley Lectin to Vacuoles of Tobacco**. *The Plant Cell* 1990, **2**(12):1145-1155.
174. Melchers LS, Sela-Buurlage MB, Vloemans SA, Woloshuk CP, Van Roekel JS, Pen J, van den Elzen PJ, Cornelissen BJ: **Extracellular Targeting of the Vacuolar Tobacco Proteins Ap24, Chitinase and Beta-1,3-Glucanase in Transgenic Plants**. *Plant Molecular Biology* 1993, **21**(4):583-593.
175. Neuhaus JM, Sticher L, Meins F, Boller T: **A Short C-Terminal Sequence Is Necessary and Sufficient for the Targeting of Chitinases to the Plant Vacuole**. *Proceedings of the National Academy of Sciences USA* 1991, **88**(22):10362-10366.
176. Bednarek SY, Raikhel NV: **The Barley Lectin Carboxyl-Terminal Propeptide Is a Vacuolar Protein Sorting Determinant in Plants**. *The Plant Cell* 1991, **3**(11):1195-1206.
177. Koehbach J, Attah AF, Berger A, Hellinger R, Kutchan TM, Carpenter EJ, Rolf M, Sonibare MA, Moody JO, Wong GKS *et al*: **Cyclotide Discovery in Gentianales Revisited-Identification and Characterization of Cyclic Cystine-Knot Peptides and Their Phylogenetic Distribution in Rubiaceae Plants**. *Biopolymers* 2013, **100**(5):438-452.
178. Kini RM, Doley R: **Structure, Function and Evolution of Three-Finger Toxins: Mini Proteins with Multiple Targets**. *Toxicon : Official Journal of the International Society on Toxinology* 2010, **56**(6):855-867.
179. Kitts DD, Weiler K: **Bioactive Proteins and Peptides from Food Sources. Applications of Bioprocesses Used in Isolation and Recovery**. *Current Pharmaceutical Design* 2003, **9**(16):1309-1323.
180. Boguski MS, Lowe TMJ, Tolstoshev CM: **Dbest - Database for Expressed Sequence Tags**. *Nature Genetics* 1993, **4**(4):332-333.

181. Kanehisa M, Bork P: **Bioinformatics in the Post-Sequence Era**. *Nature Genetics* 2003, **33**:305-310.
182. Walsh CJ, Guinane CM, Hill C, Ross RP, O'Toole PW, Cotter PD: **In Silico Identification of Bacteriocin Gene Clusters in the Gastrointestinal Tract, Based on the Human Microbiome Project's Reference Genome Database**. *BMC Microbiology* 2015, **15**.
183. Marsh AJ, O'Sullivan O, Ross RP, Cotter PD, Hill C: **In Silico Analysis Highlights the Frequency and Diversity of Type 1 Lantibiotic Gene Clusters in Genome Sequenced Bacteria**. *BMC Genomics* 2010, **11**.
184. Nickrent DL, Der JP, Anderson FE: **Discovery of the Photosynthetic Relatives of the "Maltese Mushroom" Cynomorium**. *BMC Evolutionary Biology* 2005, **5**.
185. Bennici A: **Origin and Early Evolution of Land Plants: Problems and Considerations**. *Communicative & Integrative Biology* 2008, **1**(2):212-218.
186. Wikstrom N, Kainulainen K, Razafimandimbison SG, Smedmark JE, Bremer B: **A Revised Time Tree of the Asterids: Establishing a Temporal Framework for Evolutionary Studies of the Coffee Family (Rubiaceae)**. *Plos One* 2015, **10**(5):e0126690.
187. Zhang J, Hua Z, Huang Z, Chen Q, Long Q, Craik DJ, Baker AJ, Shu W, Liao B: **Two Blast-Independent Tools, Cyperl and Cyexcel, for Harvesting Hundreds of Novel Cyclotides and Analogues from Plant Genomes and Protein Databases**. *Planta* 2015, **241**(4):929-940.
188. Bishop JG, Dean AM, Mitchell-Olds T: **Rapid Evolution in Plant Chitinases: Molecular Targets of Selection in Plant-Pathogen Coevolution**. *Proceedings of the National Academy of Sciences USA* 2000, **97**(10):5322-5327.
189. Shinshi H, Neuhaus JM, Ryals J, Meins F: **Structure of a Tobacco Endochitinase Gene - Evidence That Different Chitinase Genes Can Arise by Transposition of Sequences Encoding a Cysteine-Rich Domain**. *Plant Molecular Biology* 1990, **14**(3):357-368.

190. Gruber CW, Elliott AG, Ireland DC, Delprete PG, Dessein S, Goransson U, Trabi M, Wang CK, Kinghorn AB, Robbrecht E *et al*: **Distribution and Evolution of Circular Miniproteins in Flowering Plants**. *The Plant Cell* 2008, **20**(9):2471-2483.
191. Creighton TE, Darby NJ: **Functional Evolutionary Divergence of Proteolytic-Enzymes and Their Inhibitors**. *Trends in Biochemical Sciences* 1989, **14**(8):319-324.
192. Franco OL: **Peptide Promiscuity: An Evolutionary Concept for Plant Defense**. *FEBS Letters* 2011, **585**(7):995-1000.
193. Balaji RA, Sasaki T, Gopalakrishnakone P, Sato K, Kini RM, Bay BH: **Purification, Structure Determination and Synthesis of Covalitoxin-Ii, a Short Insect-Specific Neurotoxic Peptide from the Venom of the Coremiocnemis Validus (Singapore Tarantula)**. *FEBS Letters* 2000, **474**(2-3):208-212.
194. Gruber CW, Elliott AG, Ireland DC, Delprete PG, Dessein S, Goransson U, Trabi M, Wang CK, Kinghorn AB, Robbrecht E *et al*: **Distribution and Evolution of Circular Miniproteins in Flowering Plants**. *The Plant Cell* 2008, **20**(9):2471-2483.
195. Schaller A: **A Cut above the Rest: The Regulatory Function of Plant Proteases**. *Planta* 2004, **220**(2):183-197.
196. Muntz K, Shutov AD: **Legumains and Their Functions in Plants**. *Trends in Plant Science* 2002, **7**(8):340-344.
197. Arnison PG: **Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature**. *Natural Product Reports* 2013, **30**(1):108-160.
198. Jorgenson MA, Chen Y, Yahashiri A, Popham DL, Weiss DS: **The Bacterial Septal Ring Protein Rlpa Is a Lytic Transglycosylase That Contributes to Rod Shape and Daughter Cell Separation in Pseudomonas Aeruginosa**. *Molecular Microbiology* 2014, **93**(1):113-128.

199. Tao S, Lu Y, Zhang DY, Yang YF, Yang Y, Lu XX, Sai DJ: **Assessment of Oral Bioaccessibility of Organochlorine Pesticides in Soil Using an in Vitro Gastrointestinal Model**. *Environmental Science & Technology* 2009, **43**(12):4524-4529.
200. Araki T, Torikata T: **Structural Classification of Plant Chitinases: Two Subclasses in Class I and Class Ii Chitinases**. *Bioscience, Biotechnology, and Biochemistry* 1995, **59**(2):336-338.
201. Goldstein MA, Takagi M, Hashida S, Shoseyov O, Doi RH, Segel IH: **Characterization of the Cellulose-Binding Domain of the Clostridium Cellulovorans Cellulose-Binding Protein A**. *Journal of Bacteriology* 1993, **175**(18):5762-5768.
202. Tao FF, Yang YF, Wang H, Sun XJ, Luo J, Zhu X, Liu F, Wang Y, Su C, Wu HW *et al*: **Th1-Type Epitopes-Based Cocktail Pddv Attenuates Hepatic Fibrosis in C57bl/6 Mice with Chronic Schistosoma Japonicum Infection**. *Vaccine* 2009, **27**(31):4110-4117.
203. Price CL, Parker JE, Warrilow AGS, Kelly DE, Kelly SL: **Azole Fungicides – understanding Resistance Mechanisms in Agricultural Fungal Pathogens**. *Pest Management Science* 2015, **71**(8):1054-1058.
204. Gao AG, Hakimi SM, Mittanck CA, Wu Y, Woerner BM, Stark DM, Shah DM, Liang J, Rommens CM: **Fungal Pathogen Protection in Potato by Expression of a Plant Defensin Peptide**. *Nature Biotechnology* 2000, **18**(12):1307-1310.
205. Clark RJ, Daly NL, Craik DJ: **Structural Plasticity of the Cyclic-Cystine-Knot Framework: Implications for Biological Activity and Drug Design**. *The Biochemical Journal* 2006, **394**(Pt 1):85-93.
206. Norton RS, Pallaghy PK: **The Cystine Knot Structure of Ion Channel Toxins and Related Polypeptides**. *Toxicon : Official Journal of the International Society on Toxinology* 1998, **36**(11):1573-1583.
207. Krause S, Schmoldt HU, 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**(1):86-95.

208. Wong CT, Rowlands DK, Wong CH, Lo TW, Nguyen GK, Li HY, Tam JP: **Orally Active Peptidic Bradykinin B1 Receptor Antagonists Engineered from a Cyclotide Scaffold for Inflammatory Pain Treatment**. *Angewandte Chemie* 2012, **51**(23):5620-5624.
209. Eliasen R, Daly NL, Wulff BS, Andresen TL, Conde-Frieboes KW, Craik DJ: **Design, Synthesis, Structural and Functional Characterization of Novel Melanocortin Agonists Based on the Cyclotide Kalata B1**. *The Journal of Biological Chemistry* 2012, **287**(48):40493-40501.
210. Thongyoo P, Roque-Rosell N, Leatherbarrow RJ, Tate EW: **Chemical and Biomimetic Total Syntheses of Natural and Engineered Mcoti Cyclotides**. *Organic & Biomolecular Chemistry* 2008, **6**(8):1462-1470.
211. Chan LY, Gunasekera S, Henriques ST, Worth NF, Le SJ, Clark RJ, Campbell JH, Craik DJ, Daly NL: **Engineering Pro-Angiogenic Peptides Using Stable, Disulfide-Rich Cyclic Scaffolds**. *Blood* 2011, **118**(25):6709-6717.

Appendix

Appendix A. Replicate sequences of putative 6C- and 8C-hevein-like peptides in plants

6C-Hevein-like Peptides

Plant	Family	Peptide Sequence
Amaranthus retroflexus	Amaranthaceae	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
Chamaseyce mesebyranthemum	Euphorbiaceae	
Flaveria bidentis	Asteraceae	
Portulaca mauii	Portulacaceae	
Alternanthera sessilis	Amaranthaceae	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
Atriplex hortensis	Amaranthaceae	
Atriplex prostrata	Amaranthaceae	
Chamaseyce mesebyranthemum	Euphorbiaceae	
Heliotropium calcicola	Heliotropiaceae	

8C-Hevein-like Peptides

Plant	Family	Peptide Sequence
Lycopersicon cheesmanii	Solanaceae	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG
Solanum lycopersicum	Solanaceae	
Capsicum annuum	Solanaceae	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
Nicotiana tabacum	Solanaceae	
Actinidia chinensis	Actinidiaceae	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCGG
Actinidia deliciosa	Actinidiaceae	
Actinidia eriantha	Actinidiaceae	
Medicago truncatula	Fabaceae	
Arachis hypogea	Fabaceae	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG
Arachis ipensis	Fabaceae	
Glycine max	Fabaceae	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG
Widdringtonia cedarbergensis	Cupressaceae	
Populus nigra	Salicaceae	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
Populus tremula	Salicaceae	
Populus trichocarpa	Salicaceae	
Lathyrus sativus	Fabaceae	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCHG
Pisum sativum	Fabaceae	
Galega orientalis	Fabaceae	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCQG
Glycyrrhiza glabra	Fabaceae	
Bursera simaruba	Burseraceae	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDSG
Vitex agnus castus	Lamiaceae	
Eutrema salsugineum	Brassicaceae	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG
Thellungiella halophila	Brassicaceae	
Limnanthes douglassii	Limnanthaceae	QQCGSQAGGRTCPNNICCSQYGYCGTTEDYCSPSKNCQSNCWG
Trochodendron araliodes	Trochodendronaceae	
Lycium sp	Solanaceae	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG
Tellima breviflora	Saxifragaceae	
Juglans nigra	Juglandaceae	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS
Phellodendron amurense	Rutaceae	

Appendix B. Full-length precursors of putative 6C-hevein-like peptides

6C-Hevein-like Peptides

Plant name	Signal peptide	Mature domain	C-Terminal tail
Aerva lanata	MMSMKSCLMIMVIVALVMVEPSMG	TGECKHGKCPNGLCCSQYGYCGSGPAYCGGA	NEQQAALLQRTGVTNTGECKHGKCPSGLCCSQYG
Aerva persica	MMNMKKFLILMVVIALVMAEPSMG	AGECRNGRCPSGLCCSRYGYCGSGPAYCG	KSNMLHFFSVTAVSLRTLPELILEHNHIIVACLLTSKVNMYLST
Alternanthera caracasana	MMNMKKFLILMVVIALVMAEPSMG	AGECRNGRCPSGLCCSRYGYCGSGPAYCG	SAQEQHAALLQRNGGVTTDTTGAHP
Alternanthera sessilis	MMKSFVIMMVMMSLMVGTSMG	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG	NAEQQVAEQQVAHPASVNTTANLPQTRKVRAAPLGVNTP
Alternanthera sessilis	MVNMKSFVVIMVIMALVFVEPSMG	TGECVNGLCPRGLCCSQYGYCGKGKAYCGS	ADEQRAALLERTAAKHDAHVDGTTAP
Alternanthera sessilis	MMNMKKFLIVMVVVALVMVEPSMG	APGQCKHGRCPSGLCCSQYGYCGTGPAYCG	GAAEQRAALLQRTGSVTADTTDTTKAP
Amaranthus retroflexus	MNMKSLVVLMVILALMMVQPSMG	DGECVNGSCPNGLCCSQFGYCGSGPAYCG	GADEQRAALLQRIGVKTRKIPTTKITGAGAP
Amaranthus retroflexus	MANMNMKKFMMVMVVVALVMVGPSMG	EGECNNGQCPAGLCCSQYGYCGSGPDYCG	AAAQQRAALLKRTGATSATTTTQAP
Atriplex hortensis	MMKSFVIMMVMMSLMVGTSMG	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG	DAEQQGAQPASVNTTANLPQTRKVRAAPLGVETP
Atriplex prostrata	MMKSFVIMMVMMSLMVGTSMG	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG	NAEQQVAEQQVAHPASVNTTANLPQTRKVRAAPLGVNTP
Atriplex rosea	MMKMKSFVIIMVIMTLMVGTSMA	QSGQCTDGQCPNGLCCSQYGYCGSGPAYCG	GSVEQAQADPSTDNPQVVVGAKAP
Beta maritima	MKSFVIVMLVMSMMVATSMA	SGECNMYGRCPPGYCCSKFGYCGVGRAYCG	DAEQKVEDHPSNDADVPKFVGAGAP
Chamaseyce mesebyranthemum	MANMNMKKFMMVMVVVALVMVGPSMG	EGECNNGQCPAGLCCSQYGYCGSGPDYCG	AAAQQRAALLKRTGATSATTTTQAP
Chamaseyce mesebyranthemum	MMKSFVIMMVMMSLMVGTSMG	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG	NNAEQQVAEQQVAHPASVNTTANLPQTRKVRAAPLGVNTP
Flaveria bidentis	MANMNMKKFMMVMVVVALVMVGPSMG	EGECNNGQCPAGLCCSQYGYCGSGPDYCG	AAAQQRAALLKRTGATSATTTTQAP
Heliotropium calcicola	MMKSFVIMMVMMSLMVGTSMG	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG	NAEQQVAEQQVAHPASVNTTANLPQTRKVRAAPLGVNTP
Heliotropium tenellum	MMNMKKFLILMVVIALVMAEPSMG	AGECRNGRCPSGLCCSRYGYCGSGPAYCG	SAQEQXSVTAVSLRTLPELILEHNHIIVACLLTSKVNMYLST
Heliotropium texanum	MVVIALVMAEPSMG	AGECRNGRCPSGLCCSRYGYCGSGPAYCG	SAQEQHAALLQRTG
Kochia scopari	MMKSFVIMMVMMSLIVGTSMA	EGECKKGRCPRGLCCSKHGYCGRGPEYCG	RDAEQAVDTTDLLQARKVRAAPVGVKTP
Portulaca mauii	MANMNMKKFMMVMVVVALVMVGPSMG	EGECNNGQCPAGLCCSQYGYCGSGPDYCG	AAAQQRAALLKRTGATSATTTTQAP

Appendix C. Full-length precursors of putative 8C-hevein-like peptides

8C-Hevein-like Peptides

Plant name	Signal peptide	Mature domain	C-Terminal tail
Actinidia deliciosa	MKRAGMFGVWVLAILYLATDGAA	QFQCGRQAGGRKCPKKICCSKWGYCGTTPDFCLPSNGCQSNCKK	RLVDGGESASNVMATYHYYNPEQVGWDLRAVSSYCSTWDDNMPLAVRKKYGWA
A. deliciosa (cont.)			AFCGPAGPYGQASCGKCLRVTNTATGAETTVSIVDRCSNGGLDLDTGVFQGLD
A. deliciosa (cont.)			TDGQGNARGHLIVDYQFVNCG
Actinidia eriantha	MVGMFGVWAVAILYLAAGGAA	QQCGRQAGGRLCDNNLCCSQWGYCGNTDEYCLPSNNCQSNCKG	SGGDSASNVRATYHFYNPEQVGWDLRAVSAYCSTWDADKPLAWRKKYGWTAFC
Actinidia eriantha			GPVGPRGQASCGKCL
Actinidia setosa	MFGVWVLAILYLATDGAA	QYQCGRQAGGRKCPNKVCCSKWGYCGTTPEFCLPSKGCQSNCKK	RLVGGGESASNVMATYHYYNPEQVGWDLTAVSSYCSTWDANMPLALRKK
Actinidia chinensis	MKMRYCVLVSVLAILVIRGSA	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCGG	TPSPGPSGSDITNLIPRSTFDRMLLHRNDGACPGKNFYTYDGFIAAARSFG
Agathis robusta	MEYKNLVCVVLVLAVVIGESSVSVKG	DPTCSPAGQVYCNSGRCCSKFNWCGSTAAYCQRPNCIAQCWPG	VTALTLSESLLSLSNYTSEAATMXC
Ajuga reptans	MAFFLIQTKPLLTLTLTALFLAGGTQVSG	QNCGCASNECCSQWGYCGTTNDYCGRGCQSGPCFAP	PSTNGASVPDIVTDAFFNGIADQAA
Alangium chinense	MSKKQQSKMGRSINLCAALIACLTVAATA	QQCGRQAGGRLCPGNICCSQWGYCGTTDEYCLPSQNCQSNCKPSG	GGGESASNVRATYHFYNPEQVGWDLNAVSAYCSTWDANKPLAWRRRYGWTAFC
A.chinense (cont.)			GPAGPRGQAACGRCLRVTNTGTGTQTTVRIVDQCSNGGLDLDVGVFRQLDTDGQ
A.chinense (cont.)			GYARGHLIVNYQFVDCGDGLELNALLPILDFQ
Allium commutatum	MKTVLLVTTILALAFALFTNSYA	QQCGGQAGGALCSNGLCCSQYGYCGSTSDYCGPGCQSQCG	GGGAGGGAGGGGGGV
Alnus glutinosa	MERLSVCLVLFLCLLATATA	QQCGSQAGGKTCADNLCCSQYGYCGTTDDYCLPSNNCQSNCKSSG	DSSTGQSASNVRATYHFYNPEQNGWDLNAVSAYCSTWDANKPLAWRSKYGWTA
A.glutinosa (cont.)			FCGPVGPRGQDSCGKCLKVTNSATGAQTTVRIVDQCSNGGLDLDEGEFRKID
A.glutinosa (cont.)			TDGRGYAQGHLTVSYEFV
Amentotaxus argotaenia	MALYKLSVTVLVMLMFTAMLRVMA	DPTCSPAGNFFCNNGRCCSRFNWCGTGPSYCGRGNCIAQCP	PAVTALTSYSFSAKNLTSANP
Amentotaxus argotaenia	MKKSLLLIGFVLAVISTSSNVIVRG	DPTCSPAGRFFCNSGRCCSKFNWCGSTSDYCARPNCVAQCWPG	VTTLTYKSSLLTITNNTNP
Amentotaxus argotaenia	MVRTGTHHSTMAETRLKLAFYVLVLMAVKGLG	DPTCSPAGNFWCDTGRCCSIYNWCGSTAEYCAQGNCLAQCWSS	VAALNHYQYGLSLLNQQNLTTASLNNP
Amentotaxus argotaenia	MAKVKMLVFSILAIIVLDLGLGLPAFAA	QCGRQDRGRRCPRGLCCSQFGYCGSTPAYCCAGCQSQ	CKCPQGAAFATAEEQQSLLNIIP
Ammopiptanthus mongolicus	MKKVGVAGLVVFLCLIVTAIA	EQCGSQAGGKKCPNNLCCSQYGYCGNSEDYCSPSKNCQSNCWG	GGGGESTSNVRATYHYYQPEQHGWDLNAVSAYCSTWDAGKPYSWRSKYGWTAF
A. mongolicus (cont.)			CGPVGPHGQASCGKCLRVTNTGTGAQEIVRIVDQCSNGGLDLDVGVFNRLDTD
A. mongolicus (cont.)			GRGYQQGHLIVNYQFVDCGNELEELTNPSIPSILDALQ
Ammopiptanthus mongolicus	MKKVGVVGLVVFLCLIVTAIA	EQCGSQAGGKKCPNNLCCSQYGYCGDSEDYCSPSKNCQSNCWG	GGGGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDAGKPYSWRSKYGWTAFC
A. mongolicus (cont.)			GPVGPHGQASCGKCLRVTNTGTGAQEIVRIVDQCSNGGLDLDVGVFNRLDTDGR
A. mongolicus (cont.)			GYQQGHLIVNYQFADCGNELEKLTNPSLPSILDALQ
Anemone hupenhensis	MGNRNLSSVLLLCLIGVATA	ENCGRQVGGATCPGNHCCSQWGWCGTTDEHCLPSKNCQSNCRG	GGSSSGGGESATNVRATYHVYNPAGIGWDLYRASAYCSTWDGNKPLEWRKQYGW
A. hupenhensis (cont.)			TAFCGPVGPRGQASCGRCLRVTNTXTLQ
Anomodon attenuatus	MARHGEMLIVLVVSLLALTLAAA	QQCGSQAGNRRCPNNLCCSQYGYCGDTDPYCGNGCQSQCHG	GGGGGGSTGKASYYSPYIPSACYGNDESQFPPNRYMTAVSDGNPNLWQNRQGCG
A. attenuatus (cont.)			KHYRIRCQGNGCRNGDTITIKVVEFCPNGCHGRVFDLSEEAFGAIADKNVGVIT
A. attenuatus (cont.)			VQYEPVNSADAIPWAQEKLIAEVGRIGK
Aphanopetalum resinosum	MVNFSSCFLLMLCIACVARA	EECGRQAGGRTCSGNICCSQYGYCGTSDDYCSPSKNCQSNCAG	GGGSGGGGGGGQGTSNVRATYHYYNPEQHGWDLTAVSAYCSTWDAGKPLAWRKR
A. resinosum (cont.)			HGWTAFCGPAGPRGQAACGKCLRVTNSWTGAKTTVRIVDQCSNGGLDLDVGVFQ
A. resinosum (cont.)			KLDTNGRGYAQGHLIVSYEFVNCGDGIYNPL
Apios americana	MGKFEVCLAVLLCLMVTAIA	EQCGRQAGGQTCPDNLCCSQYGWCGNTEDYCSPSKGCQSNCG	GGSGGGGGGGGGESASNVRATYHYYEPEKHGWDLNAVSAYCSTWDAGKPYSWRSK
A. americana (cont.)			YGWGXCGPVGPRGQDSCGKCLQVTNTGTGAQTTVRIVDQCSNGGLDLDVGVFNR
A. americana (cont.)			LDTDGRGYQQGHLIVNYQFVDCGNQLDLTKPLFSILDAP
Apios americana	MGKFEVCLAVLLCLMVTAIA	EQCGRQAGGQTCPDNLCCSQYGWCGNTEDYCSPSKGCQSNCG	GGSGGGGGGGGGESASNVRATYHYYEPEKHGWDLNAVSAYCSTWDAGKPYSWRSK
A. americana (cont.)			YGWGXCGPVGPRGQDSCGKCLQVTNTGTGAQTTVRIVDQCSNGGLDLDVGVFNR
A. americana (cont.)			LDTDGRGYQQGHLIVNYQFVNCG
Arabidopsis thaliana	MKIRLSITIILLSYTVATVAG	QQCGRQGGGRTCPGNICCSQYGYCGTTADYCSPTNNCQSNCWG	SGPXGPGESASNVRATYHFYNPAQNNWDLRAVSAYCSTWDADKPYAWRSKYGWT
A. thaliana (cont.)			AFCGPAGPRGQASCGKCLRVKNTRTNAAVTVRIVDQCSNGGLDLDISMFNQIX
Arabis alpina	MKRLDILTIILLAYTVALVAG	QQCGSQAGGKTCSGNICCSQYGYCGTTADYCSPANKCQSNCWGSG	PPPSGGESASNVRATYHYYNPEQNNWDLRAVSAYCSTWDADKPYAWRSKYGWTA
A. alpina (cont.)			FCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCSNGGLDLDVAMFNRIDTD
A. alpina (cont.)			GNGYQQGHLIVDYQFVDCGNELTEQSKNILVSATDRV
Arachis duranensis	MARVGTCIVLSLCLIMVTMAMG	EQCGYQVGGAYCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG	GSTPTPTPPTPTPGTPVRATYHYYLPEQNGWDLMAVSAYCSTYDANKPLSWRSK
A. duranensis (cont.)			YGWTAFCGPVSPGFPAACGRCILVTNSATKASEIVRIVDQCSNGGLDLDVGVFN
A. duranensis (cont.)			RIDTDGRGYQQGHLDVTYSF
Arachis duranensis	MARVGTCIVLSLCLIMVTMAMG	EQCGYQVGGAYCANSLCCSKYGWCGTTYDYCSPDAGCHSNCWG	GSTPTPTPPTPTPGTPVRATYPL
Arachis hypogaea	MKMGNKLPLILVAMALVMIASVSA	QNCGCASGLCCSKYGYCGTGDAYCGDGCQEGPCYSS	PSTPSTPSSGSGGVNVADVVTQDFFNSSGGVNVADVVTQDFFNSIINQASANCE
A. hypogaea (cont.)			GKNFYTRDAFLSALNSYSTFGQLSSTDDSKREIAAAFAHFTHETGYFCYINEQN
A. hypogaea (cont.)			GASHNYCDSSYTQYPCNPNKQYFGRGPLQLTWNYNYGAAGNANNFDGLNAPETV
Arachis hypogaea	MGRVGTCIVVSLCLMVTMATG	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG	GPTPTPTPPTPTPGTPVRATYHYYLPEQNGWDLMAVSAYCSTYDANKPLSWRSKY
A. hypogaea (cont.)			GWTAFCGPNSPGFPAACGRCILVTNSATKASEIVRIVDQCSNGGLDLDVGVFNR
A. hypogaea (cont.)			IDTDGRG
Arachis hypogaea	MLNAKLPYLLSLTLLLLLLGAKA	EQCGRQAGGALCPNNYCCSQFGWCGETDDYCLASKGCQSQCRG	PPTPSPPRPPSPAPPGGDVANVISSSIFDQMLKYRNDPRCHANGFCTYNAFITA
A. hypogaea (cont.)			ARSFNGFGTTGDDATRKKEVAAFLGQTSHETTGGWAS

Arachis ipaensis	MSYTKIILNLMLSLLILTLFLLHGGAA	EQCGSQGGGALCPNNLCCSRWGWCGTGDDYCLVGNGCQSQCKD	SPTPTPTPPTPTPSPPSGGDISKYINSSMFDRMIKYRNDPRCHANGFYTNDAFIN
Arachis ipaensis	MGRVGTCIVVSLCLMVTMATG	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG	GPTPTPTPPTPTPGTPVRATYHYYLPEQNGWDLMAVSAYCSTYDANKPLSWRSK
A. ipaensis (cont.)			YGWTAFCGPNSPGFPAACGRCILVTNSATKASEIVRIVDQCNNGGLDLDVGVFA
A. ipaensis (cont.)			RGHLIVDYQFVNCGDGTTSENEPLFSIMDVAVD
Argemone mexicana	MATKNLSIFALVLFLTTIVSIVSA	QQCGRQAGGQTCPGNLCCSQYGYCGDTDDYCLPSNNCQSKCKG	SSGGGGGGGGGSGESASNVRATYHYYNPAQNGWDLTRVSAYCSTWDANKPLAWR
A. mexicana (cont.)			QKYGWTAFCGPVGPRGQASCGRCLRVTNRGTGAQITVRIVDQCSNGGLDLDYDG
A. mexicana (cont.)			VFKPIDTDGQGYLRGHMFVDYQFVNCGD
Asparagus densiflorus	MKEFLILTLFLTISSSILSSA	EQCGSQAGGAVCPNGLCCSKFGYCGTTSAYCCDGCQSQGTP	TPNPPSGGGGSVSSIISSSLFNQMLLH
Astilbe chinensis	MERMNLCMIIMLLPYLVSSVAA	QQCGRQAGGRTCANNLCCSQFGYCGSTAEYCSTNCQSNCNSG	GGGGGGQSASNVRATYHFYNPEQNGWNLNAVSAYCSTWDANKPLEWRKRYGWTA
A. chinensis (cont.)			FCGPVGPRGRDSCGKCLRVTNTGTGTQATVRIVDQCSNGGLDLDAGVFKKLDTN
A. chinensis (cont.)			GRGNAQGHLIVNYQFVNCGD
Astragalus membranaceus	MGNFGIGLVVVSILGCLIVTAIG	EQCGRQGGGQTCPNNQCCSQYGWCGDTDEYCSPSKNCQSNCKG	DTGGGESASNVRATYHYYNPEQNGWDLNAVSAYCSTWDAGKPYSWRSKYGWTAF
A. membranaceus (cont.)			CGPVGPRGQASCGKCLRVTNTGTGAQQTVRIVDQCSNGGLDLDVGVFNRLDTDG
A. membranaceus (cont.)			RGYQQGHLIVNYQFVDCGNELEXPPSCLRMWTTKLLIYPLCCSN
Astragalus propinquus	MGNFGIGLVVVSILGCLIVTAIG	EQCGRQGGGQKCPNNLCCSQYGWCGDTDEYCSPSKNCQSNCKG	DTGGGESASNVRATYHYYNPEQNGWDLNAVSAYCSTWDAGKPYSWRSKYGWTA
A. propinquus (cont.)			FCGPVGPRGQASCGKCLRVTNTGTGAQQTVRIVDQCSNGGLDLDVGVFNRLD
A. propinquus (cont.)			TDGRGYQQGHLIVNYQFVDCGNELDITKPLFSILHHA
Athrotaxis cupressoides	MALNKVSIVVLVAIVFLMGNVGVMA	DPTCSPSGNFWCNNGRCCSIYNWCGSGAAYCAKGNCLAQCWP	KAAALNQSSLNLSQQNLTSPSP
Athrotaxis cupressoides	MALNKVSIVVLVAIVFLMGNVGVMA	DPTCSPSGNFWCNNGRCCSIYNWCGSGAAYCAKGNCLAQCWP	GVTAVTTYSLTQRNLTSTKTNNP
Athrotaxis cupressoides	MTNRKYGGEKIKIALYMLLVLMASVMVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	KAAALNQSSLNLSQQNLTSPSP
Atropa belladonna	MEMEKLSTILLALVLFSIAAGGNA	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSRCSG	AAGGESAQNVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPFAWRSKYGWT
A. belladonna (cont.)			AFCGP
Austrocedrus chilensis	MGEYKSTVSVLMVIVLLGCNIGLPMVMG	DPTCSPSGNFFCNNGRCCSVYNWCGTGNAYCGKGNCLAQCSP	GVTAVTTYTLNGKNLTSNNRN
Austrocedrus chilensis	MADRRLRLGLHILLVMMVIVEKVNG	DPTCSPAGNFLCNTGRCCSQYNWCGSTSDYCASGNCLAQCWP	NAVALNQSTLQVAKKNLTSNSSP
Austrocedrus chilensis	MAYKSSSLALYMILGLMLIRVVKS	DPTCSPAGNFWCNTGRCCSIYNWCGSTSDYCASGNCLAQCWS	NTVALNKSCLNVSLHNRSSPSP
Austrotaxus spicata	MAKVIETMLLLLCSIFAFIVIGLPSSVSAA	QCGKDADGALCSSGLCCSEWGWCGNTPDHCGAGCQSQCG	EEPTPTPPTPPTPAGEGVASII
Austrotaxus spicata	MLITIFVATEMEMVMA	DPTCSPRGNFFCNSGRCCSKFNWCGSGAQYCAKGNCIAQCWPG	VSALTKYSYTQILAGRNLTGSATVNNP
Austrotaxus spicata	MAENRLRLAFYMLVLMAVMVEG	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP	TVAALNNSSLTLLQQRNLTSIHQNNP
Basella alba	MRFWAVFLLSWFSATLGGLAQ	EQCGSQAGGALCPGGLCCSQYGYCGTTSAYCGEGCQSQCT	SNVPPPPPFTPPPPPPP
Bauhinia tomentosa	MGKLVVSLVVLALVCIAEA	EQCGKQAGGQICPNNLCCSQYGWCGSTDDYCNTGKGCQSNCAG	GGGGSGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDAGKPYSWRSKYGWT
B.tomentosa (cont.)			AFCGPAGPTGAAACGKCLRVTNAWTGAQITVRIVDQCSNGGLDLDVGVFNRL
B.tomentosa (cont.)			DTNGRGYQQGHLIVNYQFIDCGNELIHPLLSILGH
Bazzania trilobata	METKASVAILALFLILSPTFIAG	QQCGSQAGGRTCPNGNCCSQWGWCGNTDDYCAGGCQSQCG	SNRCGRDFWGRVCSSNRCCSAWGWCGTDSLACGSGCQSQCHGLGLPTPVSQ
B. trilobata (cont.)			QLAMIMLNSTLPAHLMEVAKLI
Boehmeria nivea	MERLGLGLGLLALTMLASHFGGARA	EQCGRQAGFAICPNGLCCSQYGWCGDTYDYCSPDKNCQSNCWG	GPTPPTPPSGSETIVRATYHYYNPRENNWDLLAVSAFCSTWDASKPYSWRS
B. nivea (cont.)			KYGWTAFCGPSGPTGRDSCGRCLRL
Boswellia sacra	MSLYKCSWLLALCTLASTLGALA	EQCGSQAGGALCPGGLCCSKFGWCGSTSDYCSSGCQSQC	NGASPS
Boykinia jamesii	MEKRLNLCMIIMLLLCLVASAIA	QQCGKQAGGRTCANNLCCSQYGYCGSTAEYCFTNCQSNCNSG	GGGGESASNVRATYHYYNPEQNGWDLNAVSAYCSTWDANKPLEWRKRYGWTA
B.jamesii (cont.)			FCGPVGPRGQASCGKCLRVTNTGTGTQATVRIVDQCSNGGLDLDAGVFQQLD
B.jamesii (cont.)			TNGRGNAQGHLIVNYQFINCGD
Brassica napus	MKIRLSIAIILLSYTAAMVAG	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG	SGPPSGSGESATGVRATYHYYNPAQNNWDLRAVSAYCSTWDADKPYSWRSKY
B. napus (cont.)			GWTAFCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCSNGGLDLDVAMF
B. napus (cont.)			NQIDTDGDGYRKGHLIVDYQFVDCGNELIDQPADFKNILVSATDRV
Brassica napus	MALTKLSLVLFLCFLGLYSETVKS	QNCGCAPNLCCSQFGYCGSTDAYCGTGCRSGPCRSPG	GTPSPPGGGSVGSIVTQAFLLTVLSIKPVVVAPGRISTPVILSLTPQILS
B. napus (cont.)			QTLPIRLLDVKLQPCLLISPTRLDISAT
Brassica napus	MATHNVLLKNVLMIFLFTLTIMTETAFS	QNCGKTGCKGNMCCSRWGYCGTTNAYCGTGCQSGPCKS	KPKPTPTPSGSGGLNAGPRGTIASVVTPAFFNGIMSKVGRGCPAKGFYTRQA
B. napus (cont.)			FIVAAQSFAAYKGTVAKREIAAMLAQFSHESGSFCYKEEIAR
Brassica oleracea	MKIRLSIAIILLSYTAAMVAG	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG	GGPPSGSGESATGVRATYHYYNPAQNNWDLRAVSAYCSTWDADKPYSWRSKY
B. oleracea (cont.)			GWTAFCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCSNGGLDLDVAM
B. oleracea (cont.)			FNQIDTDGDGYRKGHLIVDYQFVDCGNELIDQPADFKNILVSATDRV
Brugmansia sanguinea	MEKLSSTILLALVLFITIGAGANA	QQCGSQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG	SGPTPGGGGASAQNVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYAWRSK
B. sanguinea (cont.)			YGWTAFCGPAGPRGRDSCGKCLRVTNTRTRAQTTVRIVDQCSNGGLDLDINV
B. sanguinea (cont.)			FRQIDTDGVGNQQGHLIVNYQFVDCGDNVNVPLLSVVDKE
Bryum argenteum	MAQVGQTQGMWIMRVASLLVVMQLVAVVVA	QQCGSQAGGQNCPGGLCCSQYGYCGTTAEYCGGGCQSQCG	SGGFSADGEKAASPTTSGKASYYTAPYVPSACNGNDQNQFPANMLFAAGGD
B. argenteum (cont.)			SSSANIWNSGKNCNKFFNIQCTGKGCRNNNVIKVKVVDRCPNGCTGGRAFD
B. argenteum (cont.)			LSSQAFAAIADPAVGVIDIKYSPA
Bursera simaruba	MGRVMNTFIVVLLCLTGVAIA	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDSG	EGVGGGSASNVLATYHLYNSQDHGWDLNAASAYCSTWDANKPYSWRXASMA
B. simaruba (cont.)			GLHSAVPSEHTANPPVESA
Caiophora chuquitensis	MGAKMGRLGVVMLVTIFIAVMIGCAAA	QQCGRQGGGRTCQNNLCCSQYGYCGSTDDYCLPSNGCQSNCRPSG	PGESANNVRATYHFYDPERVGWDLRAVSAYCSTWDADKPLAWRQKYGWTAFC
C. chuquitensis (cont.)			GPAGPRGRDSCGRCLRVTNTGTGAQITVRIVDQCSNGGLDLDFNGVFKRIDT

C. chuquitensis (cont.)			DGQGYAKGHLIVNYQFVNCNDGVNSLLSSIFDL
Cajanus cajan	MGKLGLGLVVLLCLIATAVA	EQCGSQAGGKTCPNNLCCSQYGWCGNAEDYCSPSKNCQSNCWG	GGGESASNVRATYHYYQPELHGWDLNAVSAYCSTWDAGEPYSWRSKYGWTAF
C. cajan (cont.)			CGPVGPRGQASCGKCLRVTNTGNGSPDHCEDR
Callitris gracilis-March	MAFTSSSVAMMVMILGLLLVMVKS	DPTCSPAGGFWCNIGRCCSIYNWCGSTWEYCAPGNCLAQCWP	YSRPLNHSSFHHPPHNRSTTSP
Callitris macleayana-March	MEMKKVMVMAMVMMVVGVMVGVRSVEG	EPTCSPAGRIYCSPGRCCSKFNWCGSSRAYCGPNCIAQCHSP	RTPLEYASTLLSATNFTNP
Callitris macleayana-March	MQILAERADMAPNKLRIGLYVLLVLVIVQING	DPTCSPAGNFLCNSGRCCSIYNWCGDTSDYCAAGNCLAQC	WPNQLALNQSALNLAQRNITSSSP
Callitris macleayana-March	MVMNKLCMGIIILLALVGLKVKG	DPTCSPAGSFLCNSGRCCSKYNWCGSTSDYCASGSCLAQCWP	NAVALNQSALHVAKQNLTSSSP
Calocedrus decurrens	MAYKISSLPLHMLLVLLVLMNVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NALALANHSSCLNVISHRNRTSAPAP
Calocedrus decurrens	MEAKAVAFILAVTLAAPLLTLA	ENCGSQGGGKACSGGECCSKWGWCGTTDDHCLPDRGCQSNCHG	GGGGGGGEKAYNVRSTYHEYYPERNNWNLYAVSAYCSTWDGEKPLWWRQKYGW
C. decurrens (cont.)			TAFCGPVGPRGQASCGKCLRVTNRETGAQATVRIVDQCSNGGLDLDINVFRQ
C. decurrens (cont.)			LDTNGQGNARGHLMVDYEFV
Camelina sativa	MNIRLCITIILLSYTAATVAG	QQCGRQAGGRTCPGNICCSQYGYCGTTPEYCSPANNCQSNCGG	SGPSGSGESAKNVRATYHIYNPAQNNWDLRAVSAYCSTWDADKPYAWRSKYGW
C. sativa (cont.)			TAFCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCSNGGLDLDVAMFNQI
C. sativa (cont.)			DTDRNGYQQGHLIVDYQFVNCGD
Camellia sinensis	MAKIGLSMALIMVFPFLLALATA	QQCGSQAGGQTCPNNLCCSQYGYCGTTDDYCSPSKGCQSNCQTG	GGGASGESATNVRATYHYYYPEQNGWDLNAVSAYCSTWDANKPLAWRSKYGWT
C. sinensis (cont.)			AFCGPVGAHGQASCGKCLRVTNTGTGAQQTVRIVDQCSNGGLDLDSGVFNKLD
C. sinensis (cont.)			TDGKGYAQGHLTVNYQFVDCGD
Cannabis sativa	MGKLSILVLALVLVVVAIGGARA	EQCGRQAGGVKCPNGLCCSQHGWCGNTFDYCSPQAGCQSNCWG	GPTPPSGETVVRATYHFYNPAENNWDLMAVSAFCSTWDANKPFSWRSKYGWT
C. sativa (cont.)			AFCGPSGPRGRDSCGKCLRLRNVGTGATITARIVDQCSNGGLDLDFNTIFKP
C. sativa (cont.)			LDTDGRGFQQGSMQVGYTFVDCGNELSEDEHQHPLSSIIEEKTQGV
Cannabis sativa	MGKLSILVLALVLVVVAIGGARA	EQCGRQAGGVKCPNGLCCSQFGWCGDTFDYCSPQAGCQSNCWG	GPAPPSGETVVRATYHFYNPAENNWDLMAVSAFCSTWDANKPFSWRSKYGW
Capsella rubella	MKIRLSIAIILLSYTAAMVAG	QQCGRQAGGRTCPGNICCSQYGYCGTTPEYCSPANNCQSNCG	GSAPSGSGESAKNVRATYHIYNPAQNNWDLRAVSAYCSTWDADKSYAWRSKY
C. rubella (cont.)			GWTAFCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCSNGGLDLDVAMF
C. rubella (cont.)			NQIDTDRNGYQQGHLIVDYQFVNCGD
Capsicum annuum	MEKRTSTTILLALVLFIVSGVANA	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG	SGPTPEGSAQNVRATYHLYNPQNVGWDLNAVSAYCSTWDANKPLAWRSKYGWT
C. annuum (cont.)			AFCGPVGPRGQASCGKCLRVTNRRTRAQTTVRIVDQCSNGGLDLDINVFRQI
C. annuum (cont.)			DTDGVGNQQGHLMVDYQFVNCGDNVNVPLLSVVDTQ
Capsicum annuum	MKFQVVILVLFALLLTRTSA	QNCGRQAGGRVCANRLCCSQFGFCGRTREYCGPGCQSNCRR	YATDTTGEGENVSNDEHKNNGGTSAATTRCCWGFFS
Carica papaya	MGSRAFAFTIVLVGLLLGAATA	QNCGRQAGGRTCANNLCCSQFGFCGTTDDHCSPAKNCQSNCRG	SSDTGGGTATNVRATFHFYNPEQNGWDLRAVSAFCSTWDADKPLAWRRRHGWT
C. papaya (cont.)			AFCGPVGPRGQPSCGRCLRVTNTETGAQATVRIVDQCANGGLDLDAGVFRQLD
C. papaya (cont.)			TNGQGNARGHLIVNYSFVNLW
Cassytha filiformis	MVEG	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCPPG	VSPLSLSQSLLSIANHTAQP
Castanea mollissima	MERLGMCSVVLLCLVANAIA	QQCGWQVGGKTCSDNLCCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGGGGGESASNVRATYHYYNPEQHGWDLNAVSAYCSTWDASKSYAWR
C. mollissima (cont.)			SKYGWTAFCGPAGPRGQASCGKCLKVTNAWTGAQTTVRIVDQCSNGGLDLD
C. mollissima (cont.)			VGVSNDWTQMEEDMLKATLW
Casuarina glauca	MERVSFCVLLSLCLSASVAA	QQCGRQAGGRTCANNLCCSQYGYCGSTDDYCSPSKNCQSNCKSG	GGGGESASNVRATYHFYNPEQNGWDLRAVSAYCSTWDADKPLSWRKKYGWTAF
C. glauca (cont.)			CGPAGPRGQAACGKCLRVTNSGTGAQTTVRIVDQCSNGGLDLDAGVFNKLD
C. glauca (cont.)			TDGRGYAQGHLIVSYQFVDCGNGGPVESA
Celtis occidentalis	MSRLCIWALVLVAVGRAAA	EQCGRQAGGAICPNGLCCSQFGWCGNTAEHCSPQAGCQSNCWG	GPPSGETVVRATYHFYNPQQNNWDLRAVSAFCSTWDADKSFSWRSKYGWTA
C. occidentalis (cont.)			FCGPSGPRGRDSCGRCLRLRNIRTGATITA
Celtis occidentalis	MMGRLCILALVLLAIGGAAA	EQCGRQAGGAICPNGLCCSQYGWCGNTAEYCSPQEGCQSNCWG	GPTPGPPSGETVVRATYHYYNPPENNWDLRAVSAYCATWDADKPYAWRSK
C. occidentalis (cont.)			YGWTAFCGPTGPTGPRDSCGRCLRLRNTRTGATITARIVDQCSNGGLDLDF
C. occidentalis (cont.)			NTIFRPLDTDGVGYQQGSMQVGYTFVDCGDGISDHDHDHPLSSILEEKSYGV
Cercidiphyllum japonicum	MGRRVNLCMVFLLCVVVAIATA	QQCGRQAGGQTCANNLCCSEFGYCGSTDDYCSPSKNCQSNCKSG	SSGGGGGGGGESASNVRATYHFYNPEQNGWDLNAVSAYCSTWDANKPLAWR
C. japonicum (cont.)			RKYGWTAFCGPVGPRGQASCGKCLRVTNSGTGTQATVRIVDQCSNGGLDL
C. japonicum (cont.)			DEGVFRQLDTDGKGNAQGHLTVNYQFVDCGD
Cercis canadensis	MGKLGVSLVVLALVCVIERDIAEA	EQCGRQGGGRTCPNNLCCSQYGWCGNGDDYCSPSKGCQSNCG	GGGGGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDAAKPYSWRSKYGW
C. canadensis			TAFCGPAGPRGQAACGKCLRVTNAGTGAQTTVRIVDQCSNGGLDLDVGVF
C. canadensis			NRLDTDGRGYQQGHLIVNYQFVDCGNDIINPLLSILNP
Cicer arietinum	MGKLELGSIVAILTLGCLICTAIG	EQCGRQGGGKTCPNNLCCSQYGYCGDSDDYCSPSKNCQSNCHG	SGESGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDASKPYSWRSKYG
C. arietinum (cont.)			WTAFCGPVGPRGQASCGKCLRVTNTGTGAQETVRIVDQCSNGGLDLDVG
C. arietinum (cont.)			VFNRIDTDGRGYQQGHLIVNYQFVDCGNQLDITHPLLSIFNA
Cicer arietinum	MGKLELGSIVAILTLGCLICTAIGE	QCGRQGGGKTCPNNLCCSQYGYCGDSDDYCSPSKNCQSNCHG	SGESGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDASKAYSWRSKY
C. arietinum (cont.)			GWTAFCGPVGPRGQASCGKCLRVTNTGTGAQETVRIVDQCSNGGLDLD
C. arietinum (cont.)			VGVFNRIDTDGRGYQ
Cinnamomum_camphora	MAYSSKVLIWALAVICCSWFMVEA	QQCGRQAGGRVCDGGLCCSQWGYCGSTDEYCSPSQGCQSNCG	GSSPSPPGEGTLVRSTYHFYNPEQHGWDLNAVSAFCSTWDADKPLSWRSK
C. camphora (cont.)			YPWTAFCGPQGPTGRDACGKCLRVTNQGTGAQIVVRIVDQCSNGGLDL
C. camphora (cont.)			DVAAFNQIDTDGRGYAQG
Citrus aurantium	MVLLICSMAIAAA	QQCGRQAGGRTCANNLCCSQYGYCGSTDEYCSPSKNCQSNCRPG	SSSSGDGQSASNVRATYHFYNPEQNGWDLNAVSAYCSTWDANKPLSWRR
C. aurantium (cont.)			KYGWTAFCGPFGPRGQASCGKCLRVTNRATRAQATVRIVDQCSNGGLD

C. aurantium (cont.)			LDAGVFRQLDTDRRGNAQGHLMVDYQFVNCGD
Clematis chinensis	MKHLHLGVVLLYLASLLLGASAQA	QCGRDAGGALCANNLCCSQFGFCGDTPPYCGDGCQSQCRPG	STPTPSTPTPTPTGGDLGSLIPESLFNEMLKHRSDPACQGGFYTYSAF
C. chinensis (cont.)			INAATSFAGFATTGNTDVRRREIAAFFAQTSHETTGGWATAPDGPYAW
C. chinensis (cont.)			GYCFIRERGDNPGDFCRPDPQWPCAPNR
Cleome viscosa	MRRVGVVLISVVLCIAATAIA	DKPQCGRNADGKKCPNNLCCSAWGFCGSTEDYCSPEKHCQDNCWY	SAKDVFAGYAAYNAEQNDWDLVAVNAFCAESDADKPYSWRSKYGWTA
C. viscosa (cont.)			FCGPVGPRGKDACGKCLNVKNSKSGAQITVRIVDECGNGGLNLDFDV
C. viscosa (cont.)			FKKLDTDGDGIRKGHLTVDYEFVNCGDDDDAELINSDARKALFSVIDRV
Conopholis americana	MGRLSFCLVLLLSLISLLLLAATTSAIG	EQCGRQASGKKCDNNLCCSQYGYCGSTGEFCLLSKNCQSNCQS	TTSGGGGGGGNTGGGGESASNVRATYHLYNPEQNGWNLNAVSGYCSTWD
C. americana (cont.)			ANKPLAWRSKYGWTAFCGPVGAHGQASCGKCLKVTNTGTGAQTTVRIV
C. americana (cont.)			DQCSNG
Coriaria nepalensis	MGKLISHCFVLILSLLASAATA	QQCGRQAGGKTCANNLCCSQYGFCGTTDDYCSPSKQCQSNCKSG	GVGGGGSRSDK
Crassula perforata	MWRVGLRLFYVLFLLCCATSAHVVSA	QICGKQAGGKTCPGNICCSQWGYCGTSDDYCSPSNNCQSNCKG	GGGDKASNVRATYHYYNPEQHGWDLNAXXALTVLHGMLTSLWHGARNTAGL
C. perforata (cont.)			RSVDLLVPAVKLHAASA
Crossopetalum rhacoma	MMALAMAMTMMVEA	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCP	PRVSPLPLSQITLSITNHTTKTQAHHPNP
Cucumis melo	MKTYSLIILSFAFLLGAASA	EQCGRQANGALCPNNLCCSQFGFCGDTDDYCKNGCQSQCRG	SSTPTPSGGSGVGSIISESLYNQMLKYSRDPRCPSNGFYTYNAFITAAR
C. melo (cont.)			SFPTFGTTGDATTRKREIAAFFGQTSHETTGGWSTAPDGPYAWGYCFI
C. melo (cont.)			RERNQQTYCTPSQQWPCAPGQQYYGRGPIQLTHNYNYEPAGKAIGATL
Cucumis melo	MKSQLSCLIILAFAFLLGSTSA	EQCGRQANGALCPNNLCCSQYGWCGDTDAYCKDGCQSQCRG	STSPTPPSTPSGSGVGSIITESLYKQMLKYSTDSRCPSNGFYTYNAFIT
C. melo (cont.)			AAQSFPGFGTTGDDATRKRELAAFFGQTSHETTGGWASAPDGPYAWGYC
C. melo (cont.)			FIREINQDVYCTPSSQWPCVSGQKYYGRGPIQLTHNYNYGPAGNALSLN
Cunninghamia lanceolata	MPESRLRSAVYVLLVLMAVMVKVKVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NAVALNQYTFNLTQQNLTSAP
Cupressus dupreziana	MGQYKSTVATLLVIVFLGCNIGLPMVMG	DPTCSPKGNFFCNSGRCCSIYNWCGTGAAYCAKGNCLAQCSP	GVTAVTTYTLTGKNLTSTHNN
Cupressus dupreziana	MAERRIRWALCMLFVLMAVVVING	DPTCSPAGNFWCNSGRCCSQYNWCGSTSDYCASGNCLAQCWP	NAMALNQSSLINVSMKKNITSASP
Cupressus dupreziana	MAYKISTLTLHMLLLLMINVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NAISFVNHSSCLNVIISHHNRTSPSP
Cycas micholitzii	MEIKNVMLTMLALLCVAFGMKGVRC	EQCGWQVGGAACDGDLCCSQYGFCGSTTEYCGQGCQSKCSG	GGGAGESADNIRATYHEYYPERHNWDLNAVSAYCSTWDASKPLWWRQ
C. micholitzii (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVTNRDTKSQATVRIVDQCSNGGL
C. micholitzii (cont.)			DLDINVFKQLDTNGRGVAQGHLMVNYEFVDCGDGDGVDAQLLR
Datura metel	MDKLSSTMLLALVLFITIGAGANA	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCLPSQGCQSQCSG	SGPTPGGGGASAQNVRATYHIYNPQNVGWDLNAVSAYCSTWDANKP
D. metel (cont.)			YAWRSKYGWTAFCGPAGPRGRDSCGKCLRVTNTRTKAQTIVRIVDQCS
D. metel (cont.)			NGGLDLDINVFRQIDTDGVGNQQGHLIVNYQFVNCGDNVNVPLLSVVDKE
Deutzia scabra	MTRSNLYTMFFLFLFLAAITIGTA	QQCGRQASGQTCANNLCCSQYGYCGTTDDYCLPSNNCQSNCKSG	GGGESASNVRATYHYYYPEQNGWDLNKVSAYCSTWDASKPLAWRRKYGW
D. scabra (cont.)			TAFCGPVGPRGQNSCGKCLRVTNTGTGAQETVRIVDQCSNGGLDLDWG
D. scabra (cont.)			VFQRLDTNGNGYAQGHLTVNYQFVDCGD
Dicranum scoparium	MDQQGQRILVASVFVILQLMALTAA	EQCGRQAGGALCPGGLCCSQFGFCGTTPEYCENGCQSNCG	GSGPSPGPSGTGRASFYTAPYVPSACFGFDQGQFPANMFFAAAGDSAG
D. scoparium (cont.)			ANIWNNGHNCGRNYRIRCQGNGCRGSGAITIKIVDRCPNGCSGGRTF
D. scoparium (cont.)			DLSSQAFAAIADPNVGVITVSYSLASPHDDVAVPWEKEQLIAEVGHGGQ
Dillenia indica	MGLKVYLCLMLLVSSIIGALA	QQCGSQAGGKTCPNNLCCSQYGYCGSTDPYCLPSNNCQSNCKG	GGGGSSGGGGNXGGESASNVRATYHLYNPQQHNWDLNAVSAYCSTWDAN
D. indica (cont.)			KPLSWRSKYGWTAFCGPVGPHGQAACGKCLRVTNTGTGSQAIVRIVDQC
D. indica (cont.)			SNGGLDLDINVFKQLDTNGNGYAQGHLIVNYQFVDCGD
Dioon edule	MEASNKVLTILALLCVVVVFGRREVRG	QQCGRQVSGRECDGNLCCSQFGYCGSTPEYCDPSQGCQSKCTGAG	GEKAENIRATYHEYYPERHNWDLLAVSAYCSTWDANKPLSWRQRYGWTA
D. edule (cont.)			FCGPVGPRGQAACGKCLRVTNRDTGSQATVRIVDQCSNGGLDLDINVF
D. edule (cont.)			KQLDTNGRGVAQGHLMVNYEFVNCGDGDGVADARVELLR
Diselma archeri	MAYKSSSLAVLMILGLMLIRAKS	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCAPWNCLAQCWP	YAVALNQSSLNLSQHNRSSTSP
Diselma archeri	MTDSRLRLGLHILLVLMVIVQTVNG	DPTCSPAGNFWCDSGRCCSIYNWCGDTSDYCASGNCLAQCWP	DAVALNQSTLSVSQNNLTLSS
Dombeya burgessiae	MSNLRLYLVLFSLCLIASVAAA	QQCGRQAGGRTCANNLCCSQHGYCGTTSEYCSPSQNCQSNCRG	TPTTGGGESASNVRATYHFYNPAQNGWDLNKVSAYCSTWDANKPLEWRRRYGW
D. burgessiae (cont.)			TAFCGPVGPRGQASCGRCLRVTNRATRAQVTVRIVDQCSNGGLDLDVAMFQQ
D. burgessiae (cont.)			IDTDGRGHAQGHLMVDYQFVNC
Dombeya burgessiae	MDKVKVSILFVLLVSSVGTTLG	EQCGRQAGGALCPDNLCCSQYGWCGSTDPYCLPENNCQSNCRS	PRPGPGPGPGPGPEATVRSTYHFYNPEQHGWDLMAVSAYCSTYDAGKPLS
D. burgessiae (cont.)			WRSKYGWTAFCGPVGPGFPAACGRCLRVTNTRTGAQQTVRIVDQCSNGG
D. burgessiae (cont.)			LDLDVGVFNRLDTD
Draba hispida	MKRLNFIFIILFVYTAATVIG	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPANNCQSNCWG	SGPSGPGESASNVRATYHYYNPEQNNWDLRAVSAYCSTWDADKPYAWRS
D. hispida (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVRNTRTN
Draba hispida	MKRLDIITFILLAYTAAMVAG	QQCGRQAGGKTCPGNICCSQYGYCGTGADYCSPAKNCQSNCQGTG	PSGPGESATNVRATYHYYNPQQNNWDLRAVSAYCSTWDADKSYAWRSKYG
D. hispida (cont.)			WTAFCGPAGPRGQASCGKCLRVRNTRTNASVTVRIVDQCSNGGLDLDVAM
D. hispida (cont.)			FNRIDTDGNGYQQGHLIVDYQFVDCGNELIDQSDSKNILVSALDRV
Draba ossetica	MKRLDIFTFILLAYTAAMVAG	QQCGRQAGGQTCPGNICCSQYGYCGTTEDYCSPANNCQSNCQGTG	PGDSASNVRATYHYYNPEQNNWDLRAVSAYCSTWDADKPYAWRSKYGWTA
D. ossetica (cont.)			FCGPAGPRGQASCGKCLRVRNTRTNAAVTVRIVDQCGNGGLDLDVAMFN
D. ossetica (cont.)			RIDTDGFGYQQGHLIVDYQFVDCNDSKNILVSAIDRV
Drimys winteri	MVFLLCLAAAA	EAQQCGRQAGGRTCAGNLCCSQYGYCGTADDYCLPSNNCQSNCKSSG	VGGGGGGESASNVRATYHFYNPEQNGWNLNSVSAYCSTWDANKPLAWR
D. winteri (cont.)			SKYGWTAFCGPVGPHGQASCGRCLRVTNVATSAQATVRIVDQCSNGG

D. winteri (cont.)			LDLDDGVFKRLDTDGRGNAQGHLIVNYQFINCGD
Elaeocarpus sylvestris	MGKSNLCMVLFLCVVASAIA	QQCGKQGGGKKCANKLCCSQYGYCGTTPPYCSKSQGCQSNCKG	SVGDAVGESASNVKATYHHYNPQQNDWDLKAVSASCSTSHANKPVEWR
E. sylvestris (cont.)			RKYGWTAFSGGQDACGKCLRVTNTATRAQATVRVVDQSSNGGLDMD
E. sylvestris (cont.)			AAVFEQLDTDGKGNAQGHLMVNYQFVDCAD
Ephedra sinica	MKILSDICFNRKMGKVAMMLMLVAVAAGVVVG	EQCGRQAGGKVCDGGMCCSQFGWCGTTNEYCGAGCQSNCG	GGGGGESASGVRSTYHEYYPERHNWDLNAVSAYCSTWDASKPLSWRKK
E. sinica (cont.)			YGWTAFCGPVGPRGQASCGKCLKVTNSGTGASATVRIVDQCSNGGLD
E. sinica (cont.)			LDVNVFNRIDTDGRGHAQGHLIVSYQFVNCGDNLAEAFSSALA
Eschscholzia californica	MATKNLFTIFLLFLTATASVKA	QQCGRQGGGSTCRDNLCCSQYGYCGNTDAYCLPSNNCQSNCRG	GGGGGGGGGGGGGESADNVRATYHFYRPEQNGWDLNRVSAYCSTWDAN
E. californica (cont.)			KPLAWRQKYGWTAFCGPVGPRGQASCGKCLRVTNRGTGAQTIVRIV
E. californica (cont.)			DQCSNGGLDLDEGVFKQLDTNGRGYADGHMFVNYQFVNCGD
Euptelea pleiosperma	MVLLLCLTAGATA	QNCGRQAGGRTCDGNICCSQWGFCGTTNDHCLPSNNCQSNCKSG	GGGSGGGESASNVRATYHIYNPEQNGWDLNRVSAYCSTWDANKPLE
E. pleiosperma (cont.)			WRRKYGWTAFCGPVGPRGQASCGRCLRVTNTGTGSQVTVRIVDQCS
E. pleiosperma (cont.)			NGGLDLDEGVFKQLDTNGRGNAQGHMIVNYQFVDCGD
Eutrema halophilum	MKKSRLSIAIILLSYTVATVAG	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG	SGPSGPGESASNVRATYHFYNPEQNNWDLRAVSAYCSTWDADKPY
E. halophilum (cont.)			AWRSKYGWTAFCGPAGPRGQASCGKCLRVRNTRTNAVVTVRIVD
E. halophilum (cont.)			QCSNGGLDLDVAMFNRLDTDGVGYQQGHLIVDYQFVDCGNDLIH
Eutrema salsugineum	MKKSRLSIAIILLSYTVATVAG	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG	SGPSGPGESASNVRATYHFYNPEQNNWDLRAVSAYCSTWDADK
E. salsugineum (cont.)			PYAWRSKYGWTAFCGPAGPRGQASCGKCLRVRNTRTNAVVTVR
E. salsugineum (cont.)			IVDQCSNGGLDLDVAMFNRLDTDGVGYQQGHLIVDYQFVDCGNDLIHQPADS
Fagopyrum tataricum	MKSVKAITNLLLLISLLVSFLIFSSVSEAA	QCGAQGGGATCPGGLCCSQWGWCGSTPKYCGAGCQSNC	RPNVAGAQNDVDPDAVKPTTNAHP
Falcatifolium taxoides	MTMMVEA	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCP	PRVSPLPLSQITLSITNHTTKTQAHHPNP
Falcatifolium taxoides	MAMEHKATMFVLLMGVVATFPLLGGA	ENCGRQAGGQICNGGECCSQYGWCGTTPDHCGTGCQSNCG	GGGGEKATNVRSTYHEYNPQNINWDYYKASVYCSTWDGGKPLSWRKKYGWT
F. taxoides (cont.)			AFCGPVGPRGQASCGKCLRVTNRETRSQTTVRIVDQCSNGGLDLDVNVFNQLD
F. taxoides (cont.)			TNGQGHARGHLMVDYEFVNCGDGEWVGAQSS
Ficus pumila	MGRLCFCVLVALLFLVVGSSA	EQCGKQASGKTCPNNLCCSQYGWCGSTDEYCSRSKGCQSNCG	GGGDGASANNIRATYHYYNPEQNGWDLNAVSAYCSTWDAGKPYSWRSKYGWT
F. pumila (cont.)			AFCGPAGPTGQASCGKCLRVTNTYTQAQLTVRIVDQCSNGGLDLDYNTAFRK
F. pumila (cont.)			LDTDGRGYQQGHLIVNYQFVDCGNGFDGNPLLSIVGDQ
Ficus religiosa	MGRLCFCVLVALLFLVVGSSA	EQCGRQAGGKTCPNSLCCSQYGWCGSTDEYCSPSKGCQSNCG	GSSGGASATNIRATYHYYNPEQNGWDLNAVSAYCSTWDAGKPYSWRSKYG
F. religiosa (cont.)			WTAFCGPAGPTGQAACGKCLRVTNTNTQTQLTVRIVDQCSNGGLDLDYN
F. religiosa (cont.)			TAFRKLDTNGL
Flaveria bidentis	MKTFLLSSFLLLLASVSSA	QNCGRQAGNAPCSNGNCCSQYGYCGNTPAHCLPENNCQSQCRP	SGGGGGGGGXAVAAAGVVVVLDPSLLRPCLTRCSSTETTRDVTQMDFTLT
F. bidentis (cont.)			MLSLMLQQRIMDSAPLEVSMIKRESSRLSSLKPPMKQQVDGQLHQMVS
F. bidentis (cont.)			LHGDIALLENKTRRIGTVTRLLGRVPKATLVEDLSNSLTITTMGNLESQLIRT
Fokienia hodginsii	MAESRLRLALYTLLVLMAVTVNG	DPTCSPAGNFWCDSGRCCSIYNWCGSISDYCASGNCLAQCWP	DAAALNQSSLTVFRQNLTSVSP
Fokienia hodginsii	MAESRLRWTLYTLLVLMAVMVKVNG	DPTCSPAGNFWCNSGRCCSQYNWCGSTSDYCASGNCLAQCWP	NALALNQYSSNVSHQNITSLSP
Fokienia hodginsii	METKAVAFILAVTLAAPLLCAA	QNCGSQAGGKVCSGGECCSKYGWCGTTDDHCLPDRGCQSNCAG	GGGGGGGSGEKASNVRSTYHEYYPERNNWSLYAVSAYCSTWDGEKPLWWRQKYG
F. hodginsii (cont.)			WTAFCGPVGPRGQASCGKCLRVTNRDTGAQATVRIVDQCSNGGLDLDINVFRQL
F. hodginsii (cont.)			DTNGQGNAKGHLMVDYEFVDCGDGLNLEEAQVM
Fontinalis antipyretica	MVRVATLVALLLCSLLHLASA	QQCGRAVNNTPCANLLNCCSQAGYCGTTDAYCVTGCQSGPCR	TTFSPPPPPTP
Galax urceolata	MGMVILTVALIIYLSTTAVA	QQCGQQAGGTLCANNLCCSQYGYCGTTDDYCLPSNNCQSNCRPG	GGGGGGESASNVRATYHYYNPAQVNWDLNAVSAYCSTWDAGMPLAWRQKYGWT
G. urceolata (cont.)			AFCGPVGPTGQASCGKCLQVTNVRTGAQATVRIVDQCSNGGLDLDVSVFQQL
G. urceolata (cont.)			DTDGNGYAQGHLTVNYQFVDCGDGLKRFQQLPISNN
Galax urceolata	MKTTAVALILIAFVLAKVLRGSA	EQCGSQAGGALCPGGLCCSKFGWCGSTSDYCGNGCQSQCG	GSTPTPGPSGGGVSGLISADLFNQMLLHRSDASCPAHN
Galega orientalis	MGNFGVCFLVVLTLGCLIATTIA	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCQG	GGGGGGGGGGGGESASNVRSTYHYYRPETVGGTKR
Geranium carolinianum	MECINVVLVLCVAVALAGGAVG	QQCGRQGGGRSCGGNMCCSQYGYCGTSDDYCSPSKGCQSNCAG	GGGGGGGQSASNVRATYHYYNPEQHGWDLNAVSAYCSTWDAGKPLSWRKKYGWT
G. carolinianum (cont.)			AFCGPVGPRGQASCGKCLRVTNTWTGAQATVRIVDQCSNGGLDLDVGVFQRLD
G. carolinianum (cont.)			TNGRGYAQGHMIVNYQFVNCGNDLELDLDVGVF
Geranium maculatum	MKFFWSVVAVALLVSLIGGEMAAA	QNCGRQAGGRRCANRLCCSQFGFCGSTHEYCGFGCQSQCPR	AAASVDPDAATAGDQATAAGNP
Geum quellyon	MGVVGNLVLLLVCLISGCVA	QQCGKQTGGKTCAGNLCCSQYGYCGTTDDYCSQSKGCQSNCAG	GGGGGAGGGGGGGESANNVRSTYHLYNPQQHGWDLGAVSAYCATWDASKPYS
G. quellyon (cont.)			WRSKYGWTAFCGPVGPRGQAACGKCLRVTNTATRAQTTVRIVDQCSNGGL
G. quellyon (cont.)			DLDVGVFRQLDTDGSGYARGHLIVNYQFINCG
Gleditsia sinensis	MGKVGVVVAVLVCLIGRAIG	EQCGRQAGGQQCPDNLCCSQFGYCGSTDDYCSPSKNCQSNCKG	DDSGGASASNVRATYHYYQPELHGWDLNAVSAYCSTWDASKPYSWRSKYGWTA
G. sinensis (cont.)			FCGPAGPHGQAACGKCLRVTNTGTGAQITVRIVDQCSNGGLDLDVGVFNRLD
G. sinensis (cont.)			TDGRGYQQGHLIVNYQFIDCGNELDFNPVLSILNH
Gleditsia triacanthos	MGKVGVLVAVLVCLIGRAIG	QQCGRQGGGKQCPDNLCCSQFGYCGTTDDYCSPSKNCQSNCKG	DDGGGASASNVRATYHYYQPELHGWDLNAVSAYCSTWDASKPYSWRSKYG
G. triacanthos (cont.)			WTAFCGPAGPHGQASCGKCLRVTNTGTGAQITVRIVDQCSNGGLDLDV
G. triacanthos (cont.)			GVFNRLDTDGRGYQQGHLIVNYQFIDC
Glycine max	MGKAWVGLVVLLCLIVTAIA	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG	GGGGESASNVRATYHYYEPEQHGWDLNAVSAYCSTWDASKPYSWRSK
G. max (cont.)			YGWTAFCGPVGPRGRDSCGKCLRVTNTGTGANTIVRIVDQCSNGGL
G. max (cont.)			DLDVGVFNRIDTDGRGYQQGHLIVNYQFVDCGNELDLTKPLLSILDAP

Glycine max	MGKAWVGLVVLLCLIVTAIA	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG	GGGGGGGGGGGGESASNVRATYHYYEPEQHGWDLNAVSAYCSTWDASK
G. max (cont.)			PYSWRSKYGWTAFCGPVGPRGRDSCGKCLRVTNTGTGANTIVRIVD
G. max (cont.)			QCSNGGLDLDVGVFNRIDTDGRGYQQGHLIVNYQFVDCGNELDLTKPLLSILDAP
Glycyrrhiza glabra	MEKIGVGLAVLLLACLIVTAMG	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCQG	GGGGGGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDASKPYSWRSKYGWT
G. glabra (cont.)			AFCGPVGPRGQASCGKCLRVTNTGTGAQATVRIVDQCSNGGLDLDVGVFNR
G. glabra (cont.)			IDTDGRGYQQGHLIVNYQFVNCGNELVDLMTNPLLSILDH
Glycyrrhiza uralensis	MKMRFRETVGVVLVCLMMVVGSSRG	EQCGRQAGGAVCPGGLCCSKFGWCGSSPDYCGDGCQSQCG	GGGGGGGGDLGSIISRDTFNQMLKHRNDGGCPAKGFYTYDAFIAAAKAFPNFA
G. uralensis (cont.)			KTGDTPTRKREVAAFFGQTSHETT
Gompholobium polymorphum	MGKVGVGVVVLVCLIVRAIA	EQCGQQAGGQTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCWG	GGGGGGGGESASNVXAPHIITINRSYMDGT
Gossypium hirsutum	MDKVKFVMLFVFLVSLLGATMG	EQCGRQAGGALCPNNLCCSQYGWCGNTDDYCSPEKNCQSNCRSG	PPPPTGEGATVRSTYHFYNPEQNGWDLMAVSAYCSTWGANKPYSWRSKYGWTA
G. hirsutum (cont.)			FCGPVGPSFPAACGRCLRVRNSRTGAQEIVRIVDRCSNGGLDLDVGVFNRLD
G. hirsutum (cont.)			TDGVGHAQGHLTIRYDFVDCGNGFNPLLASVIDS
Gossypium raimondii	MGNLSLSLVSLLCLLATAAS	QQCGRQAGGRTCANNLCCSQFGYCGTTDDYCSPSRGCQSNCRG	GGGSGGGGESASNVRATYHYYNPAQNGWNLNAVSAYCSTWDANKPLAWRQKY
G. raimondii (cont.)			GWTAFCGPVGPRGQASCGRCLRVTNRATRAQVTVRIVDQCSNGGLDLDVAM
G. raimondii (cont.)			FQQIDTDGRGHAQGHLMVDYQFVAC
Gyrocarpus samericanus	MTRSHSVLWLVLLVVTLLLVVTKA	QQCGRQAGGRICDNNLCCSQFGWCGSTDEYCLPTRGCQSNCRG	GGGGGGGGGGGGGGGEQATVRSTYHLYNPEQHGWSLTAVSAYCSTWDADKP
G. samericanus (cont.)			LWWRQKYGWTAFCGPVGPRGRDSCGKCLRVTNQATGAQRVVRIVDQCANGG
G. samericanus (cont.)			LDLDINVFRQLDTDGRGMAQGHLTVRYQFVNCNDGLTELDAAS
Gyrostemon ramulosus	MKTIILSLFLLCSLLAATTA	QQCGRQAGGRTCANNLCCSQYGYCGNTDDYCSPAKNCQSNCWP	DGGESASNVRATYHFYNPEQNNWDLRAVSAYCSTWDADKPYSWRSKYGW
G. ramulosus (cont.)			TAFCGPVGPRGQASCGKCLRVTNTGTGAQATVRIVDQCSNGGLDLDVG
G. ramulosus (cont.)			VFNKIDTDGRGYQQGHLIVNYQFVNCGDGLSADDHQHPLLSIINA
Halocarpus bidwillii	MKMEAVRMMALVMVLAMGVNMMMMVEG	EPTCSPAGRIYCSPGRCCSRFNWCGDTPAYCGRGNCIANCHX	QIHLSLTNLTSTATP
Halocarpus bidwillii	MKMEKMRLMALLMMLAMAMMMMMVVVEG	DPSCNPAAKIYCNPGRCCSKFNWCGTLTSYCGKDYCIAQCPPG	VSPLSHSQIHLSIANLTTTKPQAHVHNTP
Hamamelis virginiana	MGRVMNTFIVVLLCLTGVAIA	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDS	GEGVGGGSASNVLATYHLYNSQDHGWDLNAASAYCSTWDANKPYSWR
H. virginiana (cont.)			SXMAEQVWLDCILRSRRSTRPILLWKVLECDKYRDWS
Hedera helix	MKNAGLFLVLLLCLVAGGTA	QQCGSQGGGKTCANNLCCSQYGYCGTTDEYCSPSKNCQSNCKS	NGGDSSSGESASNVRATYHMYEPEKNGWNLNAVSAFCSTWDADKPLEWR
H. helix (cont.)			SKYGWTAFCGPVGPRGQASCGKCLRVTNTGTGAQATVRIVDQCSNGGL
H. helix (cont.)			DLDVGVFRKLDTDGKGNAQGHLIVNYQFVNC
Heuchera sanguinea	MEILNLCLTIMLLTCLVASATA	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTNNCQSNCNSAG	GGGESATNVRATYHYzNPAQNGWNLNAVSAYCSTWDANKPLAWRKKYG
H. sanguinea (cont.)			WTAFCGPAGPRGQASCGKCLRVTNTRTGAQATVRIVDQCSNGGLDLDY
H. sanguinea (cont.)			DGVFKPLDTDGNGYAQGHLTVNYQFVNCGD
Hevea brasiliensis	MGRVLNIFIVVLLCLTGVAIA	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKNSG	EGVGGGSASNVLATYHLYNSQDHGWDLNAASAYCSTWDANKPYSW
Hevea brasiliensis	MNIFIVVLLCLTGVAIA	EQCGRQAGGKLCPNNLCCSQYGWCGSSDDYCSPSKNCQSNCKG	GGGGGGG
Hevea brasiliensis	MTNIDMRASAFTVSCLLLASLLPGTSA	EQCGRQAGGALCPGGLCCSQFGWCGNTPDYCGATCQSQCGDG	DLTSIISRATFNQMLKHRNDGGCPAKGFYTYDAFISAAKAFP
H. brasiliensis (cont.)			TFGTTGGLLRHVKGEIAAL
Hibiscus cannabinus	MPKLSLCLVSLLCLLANAAA	QQCGRQAGGQTCANNLCCSQYGYCGTTDDYCSPSKSCQSNCRPG	GSGGESASNVRATYHFYNPAQNGWDLNAVSAYCSTWDANKPLAWRQKYGWTAFC
H. cannabinus (cont.)			GPVGPRGQASCGRCLRVTNRGTGAQVTVRIVDQCSNGGLDLDVAMFQQIDTDR
H. cannabinus (cont.)			RGYAQGHLMVDYQFVTC
Hibiscus cannabinus	MDKVKFVMLFVFLVSLVGATMG	EQCGRQAGGALCPNNLCCSQYGWCGSTDEYCLPENKCQSNCRG	PQPQPQPQPGPGGGATVRSTYHYYNPEQNNWDLNAVSAYCSTWDANKPYAW
H. cannabinus (cont.)			RSKYGWTAFCGPVGPGFPAACGRCLNVRNTRTGAQTTVRIVDQCSNGGLDL
H. cannabinus (cont.)			DWGVFQRLDTDGEGYRQGHMTVRYDFVDCGNGFNPLLASVVDN
Humulus lupulus	MKWCFSVFIIFSFLIGGALG	EQCGRQAGGASCPNGLCCSQYGWCGTTNDYCSPQAGCQSNCKG	SSTPTNPPTGGETVVRATYHYYNPQQNNWDLRAVSAYCSTWDADKPYSWRS
H. lupulus (cont.)			KYGWTAFCGPS
Humulus lupulus	MKLNTCLWTILWFGILVGLVISDDDDKP	ECGRDAFGEFCPKSQCCSQWGYCGNTSVYCGRGCQSNCWNS	PPPPTPTPPPPSPCPPSPPPKPCPPSPPPKPC
Hypecoum procumbens	MRTINLCVIFLLSLALAVNA	QQCGRQAGGATCQNNLCCSQYGYCGSTDAYCLPSNNCQSNCRG	SSPGTGGGESASNVRATYHFYNPEQNGWDLGRASTYCATWDANKPLEWRRR
H. procumbens (cont.)			YGWTAFCGPVGPRGQASCGKCLRVTNTRTGTQATVRIVDQCSNGGLDLDDG
H. procumbens (cont.)			VFKQLDSDGRGNFDGHLTVNYQFVNCGD
Impatiens balsamifera	MRGENVAMMMIVFSLISTAAVLG	QQCGRQAGGRLCDNNLCCSQWGYCGSTDEYCLTSNGCQSNCRG	GGGSEGSASNVRATYHFYNPQQVNWNLYAVSAYCSTWDGDKPLAWRKK
I. balsamifera (cont.)			YGWTAFCGPSGPTGQASCGKCLLVTNSRTGAKATVRIVDQCSNGGLDL
I. balsamifera (cont.)			DEGVFRQIDTDGSGYAQGHLIVSYQFVDCGNSLPFYSS
Ipomoea nil	MKYCTMFIVLLGLGSLLLTPTTIMA	QQCGRQASGRLCGNGLCCSQWGYCGSTAAYCGAGCQSQCKSTA	ASSTTTTTANQSTAKSDPAGGAN
Isoetes tegetiformans	MVLAMAMTMMIMVEG	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCSPG	VSPLSLSQSLLSIANHTAQP
Itea virginica	MGRDSLCILLILCVAYATNA	QQCGKQAGGKTCAGDICCSQYGYCGTTDDYCSSSKNCQSNCKG	GDGGQSASNVRATYHYYNPEQHGWDLNAVSAYCSTWDANKPLAWRQKYGW
I. virginica (cont.)			TAFCGPAGPRGQASCGKCLRVTNTYTGTQTTVRIVDQCSNGGLDLDVGVF
I. virginica (cont.)			QKLDTNGRGYAQGHLIVNYEFVNCGDGINPLFSLIDA
Juglans nigra	MERLCVWLVLSLCLFAAATTA	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS	SSSGGESASNVRATYHFYNPEQNGWNLNAVSAYCSTWDANKPLAWRSKYGWTA
J. nigra (cont.)			FCGPVGPRGQAACGKCLRVTNSGTGAQATVRIVDQCSNGGLDLDYAGVFQ
J. nigra (cont.)			KLDTDGKGYAQGHLIVSYQFVNCGD
Juniperus scopulorum	MAERRLRWALCMFFVLMAVVVING	DPTCSPAGNFLCNSGRCCSKYNWCGSTSDYCASGNCLAQCWP	NALALNQSTLINVSMERNITSASP
Larrea tridentata	MLKMENASILLVFLLFILTTASA	QQCGSQAGGRICDGNLCCSQYGWCGSTAEYCSPDQKCQSNCWG	PNPPQPPSPGGEGASNIRATYHMYNAEQNGWNLYAVSAYCSTWDGDKPL

L. tridentata (cont.)			SWRSKYGWTAFCGPVGPRGQASCGKCLRVTNRATRAQTIVRIVDQCS
L. tridentata (cont.)			NGGLDLDFSVFQRLDTDGRGYQQGHLMVDYQFVNCGDGVSSM
Lathyrus sativus	MGKFGTCLVVLTLVCLIATTIA	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCHG	SGGESASNVRATYHYYQPEQHGWDLNAVSAYCSTWDASKPYSWRS
L. sativus (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVRNSGTGAEETVRIVDQCSN
L. sativus (cont.)			GGLDLDVGVFNRIDTDGRGYQQGHLIVSYQFVDCGNELDMINPLFSIMDAKQ
Laurelia sempervirens	MGTWAFIFLFLASLLCASA	QQCGSQRGGAVCTGRLCCSQFGYCGNTDPYCGAGCQSKCSSG	GSTPTPTPSGGGGAVGSLLSLDQFENMLK
Lepidosperma gibsonii	MRKISTYIELFSLAIVCLGALCLTSQ	AQTCGSQAGGALCADNMCCSQYGYCGSTSDYCGTGCQSQCG	GDGGGG
Limnanthes douglassii	MKIPSSFTLLLFLCLIGTSLA	QQCGSQAGGRTCPNNICCSQYGYCGTTEDYCSPSKNCQSNCWG	GGGGGGGGGGGGGETATNVRATYHMYEPEKHNWDYMAVSAYCSTWDANKP
L. douglassii (cont.)			YSWRSKYGWTAFCGPVGPRGQASCGKCLRVTNTYTGAQQTVRIVDQCSN
L. douglassii (cont.)			GGLDLDVGVFNRLDTNGRGYAQGHLIVNYQFVDCGDGVNSVMSE
Limonium spectabile	MSPLALVSFLCLALAIGTVTA	QQCGSQAGGRTCANNLCCSQYGYCGSTDAYCATGCQSNCKTG	GGGGSPPSSGGGGASASNVRATYHLYNPQQNGWNLYAVSAYCSTWDG
L. spectabile (cont.)			GKSLAWRSKYGWTAFCGPVGPRGRDSCGKCLRVTNTGTGAQATVRIV
L. spectabile (cont.)			DQCSNGGLDLDVNVFRQLDTNGRGNAMGHMIVNYQFVDCRD
Linum hirsutum	MERLRVANISFTLVLMTLLATIAAA	QNCGRQAGGATCANNLCCSQFGFCGTTDDHCNPSKNCQSNCRPSG	GGGGGGGGGGGESASNVRSTYHLYNPEQNGWNLNAVSAYCSTWDANKPLAWRQ
L. hirsutum (cont.)			KYGWTAFCGPVGPRGQASCGRCLRVRNRGTGAAETVRIVDQCSNGGLDLD
L. hirsutum (cont.)			VGVFRRIDTDGRGIAQGHLMVDYQFVNCGD
Linum leonii	MRASVLSIVLLCFFTAIARA	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKNCQSNCRPG	GGSSGGGGGSGESASNVRATYHLYNPEQNGWNLNAVSAYCSTWDAGKPLA
L. leonii (cont.)			WRQKYGWTAFCGPVGPRGQASCGRCLRVTNRGTGASETVRIVDQCSNGG
L. leonii (cont.)			LDLDVGVFNRIDTDRRGNAQGHLMVDYQFVNCGD
Linum lewisii	MRASVLSIVLLCFFTAIARA	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKNCQSNCRP	GGGSSGGGGGGESASNVRATYHLYNPEQNGWNLNAVSAYCSTWDAGK
L. lewisii (cont.)			PLAWRQKYGWTAFCGPVGPRGQASCGRCLRVTNRGTGASETVRIVDQCS
L. lewisii (cont.)			NGGLDLDVGVFNRIDTDRRGNAQGHLMVDYQFVNCGD
Linum macraei	MERLNIVAMVRYILLLCFFTAIASG	QDCGRQAGGRTCANNLCCSQWGFCGTTDDHCNPSKNCQSNCRP	AGGGGGSPSGGGGESASNIRATYHLYNPEQNGWNLNAVSAYCSTWDANKPLA
L. macraei (cont.)			WRQKYGWTAFCGPVGPRGQASCGRCLRVTNRATGAQQTVRIVDQCSNGGL
L. macraei (cont.)			DLDVGVFRRLDTDGRGNAQGHLMVDYQFVNCGD
Linum strictum	MERSSTNIVAMVRYILLLLCFFTALASA	QNCGRQAGGRTCAGNLCCSQYGFCGTTDDHCNPSKNCQSNCRPG	GGGGGGGGGXVVARVLRMSGRRIICTTLSRTVGT
Linum usitatissimum	MKASMLVTTLLLCFLAAIASA	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKACQSNCKP	ASGGGGTGGGGGGGDSATNVRATYHMYEPEKIGWNLMTASAYCSTWDASK
L. usitatissimum (cont.)			SLAWRQKYGWTAFCGPVGPTGQASCGRCLRVTNKGTGAQETVRIVDQCS
L. usitatissimum (cont.)			NGGLDLDVGVFRRIDTDGRGNAQGHLMVDYQFVNCGD
Liriodendron tulipifera	MASHLGLWLLFLMVCAASIAA	EQCGRQAGGATCPGNLCCSQYGWCGSTDEYCTNNCQSNCKPSG	GPPSGGEGATVRATYHFYNPEQNGWDLNAVSAYCSTWDASKSYAWRSRY
L. tulipifera (cont.)			RWTAFCGPVGPSGQASCGKCLSVTNRATGAQRTVRIVDQCSNGGLD
L. tulipifera (cont.)			LDEGVFRQLDTDGSGYAQGHLMVDLMVDYQFVNCGD
Loropetalum chinense	MEKLSLSMVFIMIFLAGMDGAIA	QQCGKQAGGKTCPGNICCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGGQSASNVRATYHYYDPEQHGWDLNAVSAYCSTWDANKPL
L. chinense (cont.)			AWRQRYGWTAFCGPAGPHGQAACGKCLRVTNAWTGAKATVRIVDQCSN
L. chinense (cont.)			GGLDLDAGMFRQLDTDGRGNAQGHLIVNYEFVNCGDGINNPLLSIVDV
Lotus japonicus	MARLVEAVLAVLLPCLIVTAMG	EQCGSQAGGQLCPNNQCCSQYGWCGNTEDYCSSSKGCQSRCWG	PTPTPGGESASNVYASYHYYRPEQVGWDYSGTYCATWDAGKSLAWRSKYGWTAF
L. japonicus (cont.)			CGPVGPRGQASCGRCLRVTNSRTGAQQTVRIVDQCANGGLDLDWGVFSKLDTD
L. japonicus (cont.)			GVGYQQGHMTVSYQFVDCGNELELSKPLLSIIDAA
Lotus japonicus	MKMRLGIAILVSFILVGWCRG	EQCGSQAGGALCPGGICCSKYGWCGSTSEYCGEGCQSQCGG	STGGGDLGSIISRDTFNTMLKHRDDGGCPAKGFYTYDAFISAAKA
L. japonicus (cont.)			YPSFGTTGDTSTRKREIAAFF
Lotus japonicus	MKMRLGIAILVSFILVGWCRG	EQCGSQAGGAVCPGGICCSKYGWCGSTSEYCGEGCQSQCGG	STGGGDLGSIISRDTFNTMLKHRDDGGCPAKGFYTYDAFISAAKA
L. japonicus (cont.)			YPSFGTTGDTSTRKREI
Lotus japonicus	MKFQVVILVLFALLLTRTSA	QNCGRQAGRRVCANRLCCSQFGFCGTTREYCGAGCQSNCR	RYATDTTGEGENVNNDEHKNNGGPN
Lycium barbarum	MEKLTTLLLILFSIALGANA	QQCGRQKGGALCSGNLCCSQYGWCGSTPEYCSPSQGCQSRCSG	GGGGGGGGESAQNVRATYHIYNPQNVGWDLMAVSAYCSTWDANKPLAWR
L. barbarum (cont.)			KKYGWTAFCGPVGPRGRDSCGRCLRVTNTRTGAQTTVRIVDQCSNGG
L. barbarum (cont.)			LDLDINVFRQIDTDGVGNHQGHLIVNYQFVDCGDN
Lycium sp	MGKISISMLLIICVAYGASA	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGGGGGGAAASNVRATYHYYNPEQHGWDLNAVSAYCSTWDAKK
Lycium sp (cont.)			SLAWRKKYGWTAFCGPAGPRGQASCGKCLRVTNATTGAKTTVRIVDQCS
Lycium sp (cont.)			NGGLDLDVGVFNKLDTDGRGYAQGHLTVNYEFVDCGDGINNPLSSIIEA
Lycopersicon cheesmanii	MEKMNKLSSSTPILLALVLCISLTSVTNA	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG	GPTPTPTPTPGGGAQVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYSWR
L. cheesmanii (cont.)			SKYGWTAFCGPVGPRGRDSCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDL
L. cheesmanii (cont.)			DINVFRQIDTDGVGNQQGHLIVN
Maesa lanceolata	MAKTSSLFIALLLCLVASAAA	QQCGSQAGGKTCANNLCCSQYGYCGSTDDYCLPSKNCQSNCKSG	GSPSSSGQSASNVKATYHYYNPEQNGWNLNAVSAYCSTWDANKSL
M. lanceolata (cont.)			EWRSKYGWTAFCGPVGPHGQASCGKCLRVTNTGTGAQQIVRI
M. lanceolata (cont.)			VDQCSNGGLDLDYNTVFKKLDTDGKGYARGHLIVNYQFVDC
Magnolia grandiflora	MSSQLGLCLLFLMSCAASIT	AEQCGRQAGGATCPGNLCCSQYGWCGTTDEYCTNNCQSNCRPSG	GPSPSGGGATVRATYHFYNPEKNGWDLNAVSAYCSTWDAGKSYAW
M. grandiflora (cont.)			RSRYRWTAFCGPVGPRGRASCGRCLRVTNRATGAQRTVRIVDQCSN
M. grandiflora (cont.)			GGLDLDEGVFRQLDTDGRGYAQGHLMVDYVFVNCNDGLGVES
Malus domestica	MKLQTLIILSLSLLLGISA	EQCGSQAGGAVCPNGLCCSQFGWCGTTSDYCAAGCQSQCSS	TPKPTPTPTPSGGGGDVSSLVSSSVFDQMLKYRNDGRC
Manihot esculenta subsp. Peru	METVKSNICMFLLLCLTGGAIA	EQCGSQAGGQLCPNNLCCSRYGWCGTTDEYCSDGCQSNCKG	GGGGGGGGQIARNVRATYHLYNPQNHGWDLNAVSAFCSTWDANKPYSW

M. esculenta (cont.)			RSKYGWTAFCGPVGAHGRPSCGKCLRVTNTRTKAQTIARIVDQCSNG
M. esculenta (cont.)			GLDLDVNVFRKLDTDGVGYQQGHLIVDYQFVDCGNSLNPLFSVMDN
Manoao colensoi	MEKAKIVAVVMVLAMVMLMVEG	EPTCSPAGRIYCSRGRCCSKFNWCGTGPAYCGKSNCIAQCS	AGVSPLPHSQIFLAITNLTSLPNP
Marchantia paleacea	MLFYRSPMALVLLPLVVPFLLQVSSA	QECGSEAGGALCPDNYCCSKAGWCDSGDAYCGEGCQSGPCNY	DTSLNGDALSNPEAPLRPGSCGDGVEDVVSKAQFEAIF
M. paleacea (cont.)			ENHDRLYSYRAFVAAAGFFPE
Marchantia polymorpha	MYTKVIRVATILLLLFHLSSA	EQCGRQAGNAVCPNNLCCSQYGWCGSTSDYCVTGCQSGPCSG	GGTPSTPSGSKTGEVSYYTAPFVPSACFGDDAGQFPSNNFFAAGGDGAPNIW
M. polymorpha (cont.)			NNRANCGKWFRIQCTGNGCTSSATISVRIVDRCPNGCVGGRAFDLSDTA
M. polymorpha (cont.)			FRAIANTDVGHVTVNYSGPYDNA
Medicago truncatula	MGKLVVLILVCLIAATIA	EQCGRQAGGKTCPNNLCCSQYGYCGTTDEYCGPNCQSNCHG	SSGGGESASNVRATYHYYQPDQHGWDLNAVSAYCSTWDASKPYSWRSKYGWT
M. truncatula (cont.)			AFCGPVGPRGQASCGKCLRVTNSGTGAQETVRIVDQCSNGGLDLDVGVFN
M. truncatula (cont.)			RIDTDGRGYQQGHLIVSYQFVDCGNELDITNPLFSIIDAKQ
Meliosma cuneifolia	MGSFTLYVCLISVVCLLAGASA	QQCGRQAGGRTCAGNICCSQWGYCGTTDEYCLPSNNCQSNCKSG	GGGGGGGGXDCV
Metasequoia glyptostroboides	MAQSRLRLALYILLVLMAVTVKG	DPTCSPAGNFWCDSGRCCSIYNRCGSTSDYCASGNCLAQCWP	NAVALNQSSLNLSQQNLTSASP
Metasequoia glyptostroboides	MGMETKAVVFALAVAVAVPLLCAA	QNCGSQAGGNVCSGGECCSKWGWCGTTDDHCLPDRGCQSNCRG	GEGGEKAENVRSTYHEYYPERNNWDLYAVSAYCSTWDGQKPLWWRQKY
M. glyptostroboides(cont.)			GWTAFCGPVGPRGQASCGKCLRVTNRETGSQATVRIVDQCSNGGLDLDI
M. glyptostroboides(cont.)			NVFKQLDTNGQGNARGHLMVDYQFVDCGDGDNSASQFM
Metasequoia glyptostroboides	MLVCLIVALIVVGTPAFA	ENCGRQASGAVCPGGECCSEWGWCGNTVDHCRTPGCQSQCG	GGAGPTPSGQGVASIITESVFNELLKHRT
Microbiota decussata	MAEIRLRMGVDMLLVLIVAVMVING	DPTCSPAGNFWCDSGRCCSIYNWCGSTADYCASGNCLAQCWP	NALRLNQSTLIEVSQKNITSVSP
Microstrobos fitzgeraldii	MATMMMVEG	DPTCSPAGGFGCNAGRCCSRFNWCGGTAEYCRRPNCISQCWP	GVTPLPLSQSLLSITNYTTQPQAHVETHXTPSSCGDS
Microstrobos fitzgeraldii	MEKARAVALVMVLAMVLSVTMMVEG	EPTCSPAGRIYCNRGRCCSKYNWCGTGPAYCSKSNCIAQCS	AGVSPLSQSQILLAITNLTAHVDNTQDLNP
Microstrobos fitzgeraldii	MEKVRLMAVLMVLLMAMTMMVEG	DPSCSPTVTGRISCNSGRCCSKFNWCGTGVSYCGKDYCVAQC	PPRISPLPQSQSVASITNLTVHSTP
Microstrobos fitzgeraldii	MIVVAMAVAMVQG	EPTCSPAGRIYCSSGRCCSKYNWCGTGPAYCSKSNCIAQCS	AAVSPLAHSQILLSIANLTHTTP
Mitella pentandra	MEILNLCLTIMLLTCLVGSATA	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTNNCQSNCNQG	GGGGGGESATNVRATYHYYNPAQNGWNLNAVSAYCSTWDANKPLAWR
M. pentandra (cont.)			QKYGWTAFCGPAGPRGQASCGKCLRVTNTRTGAQVTARIVDQCSN
M. pentandra (cont.)			GGLDLDYGGVFKPLDTDGNGYAQGHLIVNYQFVNCGD
Monotropa uniflora	MRIVAVVSLSLLSIIGVFA	QQEQCGSQAGGKLCPGSLCCSRFGFCGNTTDYCSTGCQSNCVG	GGGSGGSPATPSPTPSGGGTDIGSLISKATFEQMLKHRNDAACQGKG
Monotropa uniflora	MGSPMFDICVALMIICLTTTAIGGAA	QQCGRQVAGRLCPGDQCCSQWGYCGTTDDYCLPSKNCQSNCRPSG	GGGGESASNVRATYHFYNPEQVGWDLNAVSAYCSTWDGNKSLAWR
M. uniflora (cont.)			KKYGWTAFCGPVGPRGQASCGKCLRVTNTWTGAQTTVRIVDQCSN
M. uniflora (cont.)			GGLDLDAGVFRQLDTDGRGNAQGHLIVNYQFVSCGDVAQPLFSILEDQP
Morinda citrifolia	MAMALELGLWLVFCGCCIASIAA	EQCGRQAGGALCPNNLCCSQYGWCGNTGEYCLPSNNCQSNCWG	SGPSGGTTVRATYHFYNPEQNGWNLYAVSAFCSTWDGDKPYSWRSKYR
M. citrifolia (cont.)			WTAFCGPSGPTGRDSCGKCLRVTNKRTGAQAIVRIVDQCSNGGLDLD
M. citrifolia (cont.)			EGVFRQLDTDGSGIAQGHLMVDYVFVNCNDGLSYAID
Moringa oleifera	MAKLSFLSLFLLCLVATATA	QNCGRQAGNRACANELCCSQYGFCGSTSEYCSRANGCQSNCRG	GGGAGGAGGGAGGGSP
Morus notabilis	MARLFYCMLVLFLWLVVGALA	EQCGRQAGGKTCPNNLCCSQYGWCGNTDEYCSSSKGCQSNCQG	GGGGGGGGGGDGASATNVRATYHYYNPEQNGWNLNAVSAYCSTWDA
M. notabilis (cont.)			GKSYAWRSKYGWTAFCGPSGPRGQASCGRCLKVTNTRTGAQLTVRIVDQ
M. notabilis (cont.)			CSNGGLDLDYNTAFRKLDTDGRGYQQGHLIVNYQFVDCGD
Morus nigra	MGRISFSLIVLGLAWLAVGSRG	EQCGRQAGGAICPNGLCCSQYGWCGSTYEYCSPEVNCQSNCWSAP	APGPEYVVRATYHYYNPAQNNWDLLAVSAYCSTYDANKPYSWRSKYG
M. nigra (cont.)			WTAFCGSGSSFPDACGKCLRLRNVRTGATITARIVDRCSNAGLDLDY
M. nigra (cont.)			STVFQPLDTDGVGWQQGSMDVGYTFVDCGDELSDHPLSSIIDEKSS
Mydocarpus sp	MKTDGLCLVFLLSLVAGAAA	QQCGRQAGGTTCANNLCCSQYGYCGTTDDYCSPSQNCQSNCKSG	GAGRGGESASNVRATYHYYEPEKNGWNLNAVSAFCSTWDADKPLAWRSKY
Mydocarpus sp (cont.)			GWTAFCGPVGPSGQAACGRCLMVTNTGTGKQATVRIVDQCSNGGLDL
Mydocarpus sp (cont.)			DVGVFQQLDTDGSGNAQGHLTVNYQFVDCG
Myriophyllum aquaticum	MKYMSSILVALALILGVASGAG	EQCGSQAGGAVCPNGLCCSQFGFCGSTADYCGTKCQSQCSG	GGSGGGSTPTPTPSGGSGDIS
Nageia nagi	MEKIRVMALVMVLAMAMTIMVEG	DPTCSPAGRQYCNAGRCCSKFNWCGTGPAYCGKGNCIAQCP	PRVSPLSLSQIIPSVSNLTAQPQAHHPNP
Neocallitropsis pancheri	MANRRLGNGFHILLVLIVIVQEVNG	DPTCSPAGNFWCNSGRCCSIYNWCGDTSDYCASGNCLAQCWQ	DTVALNHSTLSVAKQNLTSSSP
Neocallitropsis pancheri	MASKSSSLAVLVMILGLMLIRVKS	DPTCSPAGRFWCNSGRCCSIYNWCGSTWEYCAPGNCLAQCWP	YSRPLNHSSSHLSPHNRSSTSP
Nicotiana tabacum	MGKLSTLLFALVLYVIAAGANA	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG	GGGGGGGGGGGGGAAQNVRATYHIYNPQNVGWDLYAVSAYCSTWDGNK
N. tabacum (cont.)			PLAWRRKYGWTAFCGPVGPRGRDSCGKCLRVTNTGTGAQTTVRIVD
N. tabacum (cont.)			QCSNGGLDLDVNVFRQLDHRRKRESTWPPYCELRVC
Nicotiana tabacum	MRLCKFTALSSLLFSLLLLSASA	EQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGNCQSQCPG	GPTPTPPPHPVVGTSAVSSQVPCLIRCLSIATIMHAKERDSTVTMPLSMLLGL
N. tabacum (cont.)			FLALVPVAIPLPVKEKSRLSLLKPPMKLLEDGQQHQMVHMHGVIAGLENKV
N. tabacum (cont.)			APATTVPQVVSGLVLLVENISDEAPSKFHTTITTGLVEEP
Nyssa ogeche	MKFSAVITVCLLLVTLYQGGSA	QQCGIDVGGTLCSGGLCCSKWGFCGTTDAHCGDGCQSQCDG	GSGGGGGGGGGSSSKNGQSYQGL
Ochna mossambicensis	MEKSSILYNILIFVILVTTITKTAMA	QNCGRQAGGRTCANNLCCSQWGYCGSTDDHCSPSKNCQSNCRG	SGGGSGGGGGESATNVRATYHLYNPEQHNWDLNAVSAYCSTWDANKPIA
O. mossambicensis (cont.)			WRRKYGWTAFCGPAGPRGQGACGKCLRVTNTGTGAQTTVRIVDQCSN
O. mossambicensis (cont.)			GGLDLDIAAFRQIDTDGRGNAQGHLIVNYQFVNCGD
Oncotheca balansae	MAENRLRLAFSMLVLMAVMVKG	DPTCSPAGNFWCDTGRCCSIYNWCGSTADYCAQGNCLAQCWP	SVDAVDHYGLSLLQEKNLTSAHQNNP
Oncotheca balansae	METKTKAVVFLLVALLSLCAA	QNCGSQAGGRACGGGECCSKWGWCGTTDDHCLPERGCQSNCRG	GGGEKGYNVRSTYHEYYPERHNWDLYAVSAYCSTWDGGRPLWWRKKYGW
O. balansae (cont.)			TAFCGPVGPRGQASCGKCLRVKNRETGSQATVRIVDQCSNGGLDLDI
O. balansae (cont.)			NVFKQLDTNGQGVARGHLMVDYEFVDCGDGAGASQFL

Ophioglossum vulgatum	MGFLNRRRLNHLAVLLLLLQQWSRVGRA	QQCGSQAGGAVCSNNLCCSKYGYCGSTDPYCGSGCQSGPCSSG	GGGTPAPSGGGGSVSSIITSSIFDTFLNHRTDGSCPANGFYTYNAFITAAAAYN
O. vulgatum (cont.)			GFGTTGDATTQKRELAAFFAHAAHETSEGCYKEEINKQTYCQAT
O. vulgatum (cont.)			AQWSCASGKQYYGRGPLQITWN
Papaver setigerum	MATKNLSVTALLLFLTAIFSAVSA	QQCGSQAGGRTCQGNLCCSQYGYCGDTNEYCLPSNNCQSNCRG	PSVPGTGESASNVRATYHIYSPAQNGWDLYRVSAYCSTWDGNRPLAWRQKYG
P. setigerum (cont.)			WTAFCGPVGPRGQASCGRCLRVTNRGTGAQTTVRIVDQCSNGGLDLEEGV
P. setigerum (cont.)			FRQIDTNGIGINDGHMFVDYQFVNCGD
Papuacedrus papuana	MAEKRLRLALYMILGLMLIRVRT	DPTCSPAGNFWCKSGRCCSIYNWCGSTSDYCAPGNCLAQCWP	YIAGLNQSCLNVSPHNSSSASP
Papuacedrus papuana	MAENKLRMGLYILLVLLLIVKIYG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	DAVALNQSSLQVAKQNITSNSSP
Papuacedrus papuana	METKAATVTFILAATLAAPLLCGG	QNCGSQAGGKVCGGSECCSKWGWCGTTDDHCLPDRGCQSNCRG	GGGGGGGGSGEKAFNVRSTYHEYYPERHNWDLYAVSAYCSTWDGGK
P. papuana (cont.)			PLWWKQKYGWTAFCGPVGARGQPSCGKCLRVTNRDTGSQATVRIV
P. papuana (cont.)			DQCSNGGLDLDINVFRQLDTNGQGNARGHLMVDYQFVDCGDGESFEGVHEFM
Parasitaxus usta	MERAKIVAFVMVLAVAMTMMIMVEG	DPTCSPAGRVYCNPGRCCSKFNWCGTTRANCGKGNCIAQCHP	AVSPLSLYQAFLSITNHTAQP
Parasitaxus usta	MKMERAKIVAFVMVLAMAMTMMIMVEG	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCP	PGVSPLSLSQSLLSIANHTAQP
Parasitaxus usta	MANAGVVLVWAVVASSLAAALA	EDCGKQAAGKLCPGGLCCSQWGWCGKTADHCGKNCQSQCSG	GGPTPTPPTPXRSGCLLHHHSKRLQQYVEAQERGKLPRQE
P. usta (cont.)			FLLLQRLHYCCQFLPGLRNHRRSHHQEKRARRFLRPNIA
P. usta (cont.)			RNDRWMGIRPRWPVRVGVLLPQRKWRRRLLQPQLGALRCREAVL
Passiflora caerulea	MGRGMKSWLVLLLCFLGGAMA	EQCGQQAGGKTCPGNICCSQYGWCGTTDEYCSPDNNCQSNCKG	GGGGGGGGGGGQSAKNVRATFNDYNPQQHGWDLNAVSAYCATWDAGKPYSWRSKYG
P. caerulea (cont.)			WTAFCGPVGPQGRDSCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDLDVNVFNQ
P. caerulea (cont.)			LDTDGNGNFQGHLIVDYQFVDCGDSVNNPLYSIIEDRLM
Peganum harmala	MVKLSLYIVLSLSILAATTTA	QQCGRQAGGRLCPNNLCCSQFGYCGSSDDYCSPSKNCQSNCKPSG	GGGGGGGGGASASNVRATYHLYNPSQIGWDLMAASAYCATWDANKPLAWRRK
P. harmala (cont.)			YGWTAFCGPAGPRGRESCGKCLRVTNRGTGAQTIARIVDQCSNGGLDLDIAVFR
P. harmala (cont.)			QIDTDGRGHAQGHLMVDYQFVNCGNEISEIFLGEASS
Pereskia aculeata	MKTPTIPMKFSLLSLLILLMGAHEGEA	QQCGAQAGGALCRDGLCCSIWGYCGSTDPYCGPDNCQSQCPG	SSPPPPPPSPTPPPPPPLPPCPGGDLSDIFP
Peumus boldus	MASQKSLVWAMAMICCSCLATKTEA	VQCGRQAGGQLCNGNLCCSQFGWCGSTDEYCLPSSGCQSNCKTSG	GTTVRATYHLYNP
Phaseolus vulgaris	MGKARVVIQLLFCLIVTATA	QQCGRQAGGSRCSGNLCCSQFGWCGNTAEYCSPSQNCQSNCWG	SGGSASNVRATYHYYQPEQHGWDLNAVSAYCSTWDAGKSYAWRSKYGWTAFCGP
P. vulgaris (cont.)			VGPTGRDSCGRCLRVTNTGTGAQTTVRIVDQCSNGGLDLDVGVFNRLDT
P. vulgaris (cont.)			NGQGYQRGHLIVNYEFVDCGNELNLALHAPN
Phellodendron amurense	MERLCVWLVLSLCLFAAATTA	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS	SSSGGESASNVRATYHFYNPEQNGWNLNAVSAYCSTWDANKPLAWRSKY
P. amurense (cont.)			GWTAFCGPVGPRGQAACGKCLRVTNSGTGAQATVRIVDQCSNG
P. amurense (cont.)			GLDLDYAGVFQKLDTDGKGYAQGHLIVSYQFVNCGD
Phellodendron amurense	MIFSLCLLATAAA	QHCGRQAGGRTCPNNLCCSQFGYCGTTDDYCSPSKNCQSNCQSAG	GGGGGGGQ
Philonotis fontana	MARQGQSMVMLLASAFVVLSLMSLTAA	EQCGRQAGGKTCPDNLCCSQWGYCGNTDEFCKNDCQSQCG	GGGGGGGGGGQSGTGQASFYGPEYVPSACFGFDPNQFPANR
P. fontana (cont.)			HFAAGGDGSPNIWXSLRLPTPTWV
Philonotis fontana	MARQGQSMMLLASGIIMVLSLMSQLTAA	EQCGRQAGNQKCPKDQCCSYWGWCGDTPEFCEHGCQSQCSG	GGGAPGGSTGQASFYGPPYSPNACYGFDTGLFPANLYIAAAGDGNPNLWDN
P. fontana (cont.)			GRNCGRFFRIQCQGKGCRADRGTITVKIVDRC
Phyllocladus hypohyllus	MEKMRMMALVVVLAMAMTMVVEG	EPTCSPAGRIYCSQGRCCSKFNWCGTGPAYCSKSNCIAQCS	AGISPLPQSQILLSITNLTHVANTP
Phyllocladus hypohyllus	MAKKATILVVMMGVAVALPLLCVA	QNCGRQAGGVVCSGGECCSQYGWCGTTPDHCGTGCQSNCG	GGGGGEKGYNVRSTYHEYYPERINWDYYTASVYCSTWDGGKPLSWRNKYGWT
P. hypohyllus (cont.)			AFCGPVGPRGQASCGKCLRVTNRETGAQETVRIVDQCSNGGLDLDVNVFNR
P. hypohyllus (cont.)			LDTNGQGYARGHLMVDYEFVNCGDGQTVAVE
Physcomicromitrium sp	MIINMNVLFLIATIVGRDVAA	QCTAPPRSDFQCGPANGNASCPSGKCCSWWGWCGATSFYCNQRSLCTS	LTSSPTPAKTPTLAPTPVNTTFPTPVLTPAKTPSRIPPISVRGASVFNW
Physcomicromitrium sp (cont.)			CGTSKSIALTFDESAFMEPEITNLLNDLKRLQMKATFF
Pilgerodendron uviferum	MAAIRLRLGLHMLLILMAVMVKVKG	DPTCSPAGNFWCDSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NTLPLNQSTLTVAQNNLTSPSP
Pilgerodendron uviferum	MGQYKSTMTVLVVIVLLGCSIGIPMVMG	DPTCSPRGNFFCNSGRCCSIYNWCGTGPAYCGKGNCLAQCSP	GVTAVTTYSLNGKNLTSNNRN
Pinus taeda	METSVCNMKSMKFSAMAIALLTMATMNLYFVSA	EQCGQQAGGALCPGGLCCSKWGWCGNTDDHCGQDCQSQCGG	STPTPPSPTPGGQGVASIITESIFNELLKHRNDAGCKASGFYTYSA
P. taeda (cont.)			FIAADNAFPSFGTAGDVATRKRELA
Pinus parviflora	MALYKLSVTVLVMLMFTAMLRVMA	DPTCSPAGNFFCNNGRCCSRFNWCGTGPSYCGRGNCIAQCP	PAVTALTSYSFSAKNLTSANP
Pisum sativum	MRNFGVGVVVLTLAYLIATTIA	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCHG	SGGGSGESASNVRATYHYYRPEQHGWDLNAVSAYCSTWDASKPYSWRSKYGWTA
P. sativum (cont.)			FCEPVGPRGQASCGKCLRVRNSGTGAQETVRIVDQCSNGGLDLDVGVFN
P. sativum (cont.)			RIDTDGRGYQQGHLIVSYQFVDCGNEVDMINPLSPSWMLT
Platycladus orientalis	MAYKIVSLALYMLLMLMNVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NALSLLNHSSCLNVIPNHNRSSVPSP
Platycladus orientalis	MKEVAVMVVMIVVMAVSVEG	DPTCSPAGRFFCNSGRCCSKYNWCGSTSDYCAKNNCIAQCWP	GFTATSSPLTIANHTVPLHNNP
Platycladus orientalis	MRIKIRMADSRSRMGLYILLVFMTVNIHG	DPTCSPAGNFWCNSGRCCSVYNWCGSTSDYCASGNCLAQCWP	NAVALNQSTLISVSQKNITSVSP
Plumbago zeylanica	MEKNIVRRQPLLFLIIYTLASLAVAAVA	QQCGRQAGGRTCPNGLCCSQYGYCGTTDDYCLKSKNCQSNCKG	SGGSPPSTGGGGGSASNVRATYHLYNPQQNGWSLYAVSAFCSTWDGDKP
P. zeylanica (cont.)			LSWRSKYGWTAFCGPVGPRGRDSCGKCLRVTNTRTNAQA
Podocarpus coriaceus	MEKIRVMALVMVLAIAMTMMVEG	DPSCSPAGRQYCNDGRCCSKFNWCGTGPAYCSKGNCIGQCPP	GVSPLPLAQIIPSIANLTTQPEAHHPNP
Polypodium plectolens	MEGQAKLNIAFLFPLLLLSSTALLSCWA	QQCGSQAGGALCANNLCCSQYGYCGSTSAYCAVPGCQSQCG	GSSSTGGSSGVSSIITSDIFNQFFSQNSASG
Polyscias fruticosa	MGSQTHNVTKKARLLLVFLLCLVAGGTA	QQCGRQGGGKTCANNLCCSQYGYCGTTDDYCSPSKNCQSNCNSHG	GGVVGGESASNVRATYHIYEPEKKGWDLNAVSAFCSTWDADKPLEWRSKYG
P. fruticosa (cont.)			WTAFCGPVGPHGQASCGKCLRVTNTGTGAEATVRIVDQCSNGGLDLDV
P. fruticosa (cont.)			GVFRQLDTDGKGNAQGHLIVNYQFVNC
Populus deltoids	MRTVYDTFLIVLLCLIAGATG	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG	SGGESASNVRATYHLYNPQDHGWDLNAVSAYCSTWDASKPYSWRSKYGWTA

P. deltoids (cont.)			FCGPAGPRGQASCGKCLRVTNTRTGAQTTARIVDQCSNGGLDLDVNVF
P. deltoids (cont.)			RTIDTDGDGYXKGHLIVNYQFVDCXDSINSPKPLLSIIDDQ
Populus euphratica	MQPPGMERLSLLIVFLLCLAATGIA	QNCGRQAGGRTCANNLCCSEWGFCGTSDDHCSPSKSCQSNCRASG	GSGGGGSSGGGESASNVRATYHFYNPEQNGWDLNAVSAYCSTWDANKPLAW
P. euphratica (cont.)			RRQYGWTAFCGPVGPRGQASCGRCLRVTNTGTGAQATVRIVDQCSNGGL
P. euphratica (cont.)			DLDAGVFRQIDTDGRGNAQGHLIVNYQFVNC
Populus euphratica	MRRVYDTFLIVLLCLIGGATG	EQCGKQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG	GGGESASNVRATYHLYNPQEHGWDLNAVSAYCSTWDASKPYSWRSKYGW
P. euphratica (cont.)			TAFCGPAGPRGQASCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDLDVN
P. euphratica (cont.)			VFRTIDTDGDGYAKGHLMVNYQFVDCGDSINSPKPLLSIIDDQ
Populus euphratica	MKIWTFTVFSLLLSLLLGCSA	EQCGSQAGGALCPGGLCCSQFGWCGKNNDYCGTGCQSQCG	GAGGGDLGSIISSAKFDEMLKHRNDGGCPGKGFYTYNAFISAANAFPGFGTTGDA
P. euphratica (cont.)			DTRKREIAAFLGQTSHETTGGWQTAPDGPYAWGYCFVKEQNPGSYCSPSSTYP
Populus nigra	MRRVYDTFLIVLLCLIAGATG	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG	GGGESASNVRATYHLYNPQDHGWDLNAVSAYCSTWDASKPYSWRSKYGWTAFC
P. nigra (cont.)			GPAGPRGQASCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDLDVNVFRTIDTDGD
P. nigra (cont.)			GYAKGHLIVNYQFVNCGDSINSPKPLLSIIDDQ
Populus tremula	MRRVYDTFLIVLLCLIGGAAG	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG	GGGGGESASNVRATYHLYNPQDHGWDLNAVSAYCSTWDASKPYSWRSKYGWTA
P. tremula (cont.)			FCGPAGPSGQASCGKCLRVTNTRTGAQTTARIVDQCSNGGLDLDVNVFRTID
P. tremula (cont.)			TDGDGYAKGHLIVNYQFVDCGDSINSPKPLLSIIDDQ
Populus trichocarpa	MKRLSLFMVFLLCLAGTAIA	QNCGRQAGGQTCANNLCCSQWGYCGTSDDHCNPSKNCQSNCRSS	DTGGGGESASNVRATYHYYNPDQNGWDLNAVSAYCSTWDANKPLAWRRKYGW
P. trichocarpa (cont.)			TAFCGPVGPRGQASCGKCLRVTNTGTGAQVTVRIVDQCSNGGLDLDAGVFRQI
P. trichocarpa (cont.)			DTDGRGNAQGHLIVNYQFVNCGD
Populus trichocarpa	MKRLSLFMVFLLCLAATAIA	QNCGRQAGGQTCANNLCCSQWGYCGTSDDHCNPSKNCQSNCRSSG	SGGGTGGGGESASNVRATYHYYNPDQNGWDLNAVSAYCSTWDANKPLAWRR
P. trichocarpa (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVTNTGTGAQVTVRIVDQCSNGGLDLD
P. trichocarpa (cont.)			AGVFRQIDTDGRGNAQGHLIVNYQFVNC
Populus trichocarpa	MESQI	INCGRQAGGRTCANNLCCSEWGFCGTSDDHCSPSKNCQSNCRP	SGGGGGGGSSGGGESASNVRATYHFYNPEQNGWDLNAVRAYCSTWDANKPLA
Prumnopitys andina	MEKISLMVFVMVLAMMMMVEG	DPSCNPAGKIYCNPGRCCSKFNWCGTLESYCGKNYCIAQCP	PGVSPLSPSQILLSIANLTAQPQAHVDLNP
Pseudotaxiphyllum elegans	MARHGKMLMLIALAVSLLALTLADG	EQCGSQADGATCPNNLCCSKWGYCGSTDAYCNPNEGCQSNCWG	EVHALDRGLGDPRCGQAGGGDICNFGRCCSIFDFCGSTDAYCGNDCAS
P. elegans (cont.)			QCPPRNGGLRASTGKASYYTRYVPSACYGNDESQLPRNRHMAAVSDGHP
P. elegans (cont.)			NLWKNGEGCGKHYRVRCEGNGCRNADAITIKVVDRC
Pseudotaxus chienii	MKKRLTLFMALLIIIISSGSVSG	DPTCSPAGRFFCNTGRCCSKFNWCGSTAEYCASPKCIAQCWP	GVTATATATATPTATAXPTATATATPPSLLTITNYTTTPNNRHP
Pseudotaxus chienii	MAENRLRLAFYMLVLMAVMVEG	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP	TVAALNNSSLTLLQQRNLTSIHQNNP
Pyrenacantha malvifolia	MDRIAILLAWSLLVLIGTAMG	QQCGRQAGGRLCANNMCCSQWGYCGTTDDYCLPSNGCQSQCRG	GGGGTEVRATYHFYNPQDVGWDLNAVSAFCSTWDASKPLWWRSKY
P. malvifolia (cont.)			GWTAFCGPVGPQGQAACGKCLRVRNKRTGASTTVRIVDRCSNG
P. malvifolia (cont.)			GLDLDVNVFRALDTDGSGYAQGHLIVTYEFVDCGDGSNGGDDALLITSVTDRVNI
Quercus petraea	MERLGLCSVVLLCLVANAIA	QQCGWQVGGKTCSNNLCCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGGGGGGGESASNVRATYHYYNPEQHGWDLNAVSAYCST
Q. petraea (cont.)			WDAGKSYAWRSKYGWTAFCGPAGPRGQASCGKCLKVTNAWTGAQTTV
Q. petraea (cont.)			RIVDQCSNGGLDLDVGVFQRLDTDGRGYAQGHLMVNYQSVDRGDGFNPLLSII
Quillaja saponaria	MKFMYTVFVFTLILVLLKISA	EQCGKQAGGQVCPNGLCCSQFGFCGTTNEYCETGRGCQSQCKPSG	TGGGPTPKPTXAVVVAMLPLLLALIFSTNCLNIETMANARAKDFILTML
Q. saponaria (cont.)			SSLRQDLLVGLVPQVILARGRLLLTWLKHPMKPQEDGVVPQMAHMHG
Q. saponaria (cont.)			VIALSEKKTGLHIVIPAKGHVLLANYIMAEDLFN
Raphanus raphanistrum subsp.	MKIKLSITLILLSYTAAMVAG	QQCGRQASGRTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG	GGPSGSGESASNVRATYHYYNPAQNNWDLNAVSAYCSTWDANKP
R. raphanistrum (cont.)			YSWRSKYGWTAFCGPVGPRGQASCGKCLRVRNTRTNAAVTVR
R. raphanistrum (cont.)			IVDQCSNGGLDLDVAMF
Reseda odorata	MRDLSLILLIAILCLCVAAEKARG	QQCGRQAGGRTCANNLCCSQYGYCGTTDDYCSPSKGCQSNCRG	GGDSGGQSASNVRATYHLYNPERNNWDLRAVSAYCSTWDADK
R. odorata (cont.)			PYSWRSKYGWTAFCGPVGPRGRXG
Retrophyllum minus	MERAKIVAFVMVLAMAMTMMIMVEG	DPTCSPAGRVYCNPGRCCSKFNWCGTTRANCGKGNCIAQCHP	AVSPLSLYQAFLSITNHTAQ
Rhamnus japonica	MASLRLGHCLSLVLLIFLVAACATA	QQCGQQVGGKTCSGNLCCSQYGYCGNTDDYCSPSKNCQSNCWG	GGGDSGEGSASNVRATYHFYNPEQNGWDLNAVSAYCATWDA
R. japonica (cont.)			SKPYSWRSKYGWTAFCGPVGPRGQASCGKCLKVTNTRTG
R. japonica (cont.)			AQRTVRIVDQCSNGGLDLDAGVFTQLDTDGSGYAQGHLIVNYDFVNCGNDLVD
Rhodiola rosea	MEIVKVWVFISLGLLATATA	QNCGRQAGSATCANNQCCSQWGYCGTTDDHCLPSNNCQSNCRTEG	TGGGGQASNVRATYHNYNPQNIGWDLNTASVFCSTWDADKPLAWRS
R. rosea (cont.)			RHGWTAFCGPVGTGGRDSCGKCLRVTNSGTGAQAVVRIVDQCSNG
R. rosea (cont.)			GLDLDVNVFNQIDTDG
Rhodiola rosea	MRSMGNVGLFSVLFFILSAYVTS	AQICGKQAGGKTCAGNICCSQWGYCGTTDDYCLPSNNCQSNCKG	GGGGEKASNVRATYHNYNPEQHGWDLNAVSAYCSTWDANKPLAWRKK
R. rosea (cont.)			YGWTAFCGPVGPRGQASCGKCLKVTNAWTGAQTTVRIVDQCSNG
R. rosea (cont.)			GLDLDVGVFNKLDTNGQGYAKGHLMVNYEFVNCGNELDYQLLSSTTISDQ
Ricinus communis	MGSVIDVLMVVLLCSLGVAFG	EQCGRQAGGKLCPNNLCCSQYGWCGSTDDYCSPSKNCQSNCKG	GGGGESASNVRATYHFYNPQQHGWDLNAVSAYCSTWDANKPYSWRSKYG
R. communis (cont.)			WTAFCGPAGPRGQTSCGKCLRVTNTRTGAATTVRIVDQCSNGGLDL
R. communis (cont.)			DVNVFRQLDTDGIGYQQGHLIVNYQFVNCGDSFNPLFSIMDDQ
Robinia pseudoacacia	MGKVKVGLVIFACLIVSAIG	EQCGRQAGGQTCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCRG	GEGGESASNVRATYHYYQPELHGWDLNAVSAYCSTWDAGKPYSWRSKYGWTA
R. pseudoacacia (cont.)			FCGPVGPHGQASCGKCLRVTNTGTGAQETVRIVDQCSNGGLDLDVGVFNRLD
R. pseudoacacia (cont.)			TDGLGYQRGHLIVNYQFVECGNELELELNNPLLSILDAH
Roridula gorgonias	MTYPLVLLSLAAILVTCCWA	EQCGSQAGGALCPGGLCCSKFGWCGGTGDYCGDGCQSQCG	GGATPTPG

Ruellia brittoniana	MALSFSHIKPSMFFALAAILLAAIAPALG	QNCGCDARLCCSRWGYCGSTIEYCGQGCQSGPCFTSPPSNG	NSVSNIVTEAFFNGI
Salix dasyclados	MRIYDAFVIVLLCLIRGAIG	EQCGSQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG	GGGGESASNVRATYHLYNPQDNGWDLNAVSAYCSTWDASKPYSWRSKYG
S. dasyclados (cont.)			WTAFCGPAGPRGQASCGKCLTVTNTRTGARTTVRIVDQCANGGLDLD
S. dasyclados (cont.)			VNVFRTIDTDGDGYAKGHLIVNYQFVDCGDSVKNPEPLFSIIEDQ
Salix fargesii	MRIYDTFVIVLLCLIRGAIG	EQCGSQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG	GGGGESASNVRATYHLYNPQDNGWDLNAVSAYCSTWDASKPYSWRSKY
S. fargesii (cont.)			GWTAFCGPAGPRGQASCGKCLTVTNTRTGARTTVRIVDQCANGGLDL
S. fargesii (cont.)			DVNVFRTIDTDGDGYAKGHLIVNYQFVDCGDSIKIPEPLFSIIEDQ
Saxegothaea conspicua	MEVGQKATIFVFMIAVALPLFGSG	ENCGRQAGGQVCNGEECCSQWGWCGTTPDHCGTGCQSNCG	GGEKGTNVRSTYHMYNPQNINWDYYAASVYCSTWDGGNPLSWRKKYGWT
S. conspicua (cont.)			AFCGPVGPRGQASCGKCLRVTNRNTRAQTTVRIVDQCSNGGLDLDVNV
S. conspicua (cont.)			FNQLDTNGQGYAQGHMMVDYEFVNCGDGETLAVESSEPLLLRSVAHEEENLMP
Schizolaena sp.	MGRSLQLAVFLLCLLGGAV	AQQCGWQVGGQTCPDNLCCSQYGYCGTTDDYCSPSKNCQSNCKG	GGGGGGGGGGESASNVRATYHLYNPEQHGWDLNAVSAYCSTWDASKPYSWR
Schizolaena sp. (cont.)			SKYGWTAFCGPVGPHGQAACGKCLRVTNSGTGAQETVRIVDQCSNGGLD
Schizolaena sp. (cont.)			LDVGVFQRLDTDGRGYAQGHLIVNYQFVDCGDGITSSTGPSFLNKLLET
Schwetschkeopsis fabronia	MKSVLFVVIVVLSLVASIEA	ACSKTSPCANGLCCSQYGYCGTTAEYCGAGCQSQCG	GTPATPKPTTSGGGGGTGKGYITYHLYNQGDSLTTVACSDGANGLMTRWGYS
S. fabronia (cont.)			TLSPMYPNVAAFSGATWNSPNCGKCAKLTNGAKTAYVTVIDQCGGAPAGYT
S. fabronia (cont.)			AHFDVAPGAFTNLFGSTSQGVGSVDWSFVANSFCQGNRG
Sciadopitys verticillata	MAECKLRWAFILVLLGLLMVLAVA	DPTCSPAGNFWCNSGRCCSKYNWCGDTEEYCGEGNCIAQCWA	SVSAVTEYSVNLFQKQNLTTAQKNNP
Sciadopitys verticillata	MTTMKKLVVFVLMVGVAAPLLCAA	ENCGQQVGGGVCPGGECCSQYGWCGTTDAHCGAGCQSNCRTP	SPSPSGGGDKASNVRSTYHEYYPERHNWDLYAVSAYCSTWDGGKPQWWRQ
S. verticillata (cont.)			KYGWTAFCGPVGPRGQASCGKCLLVKNRETGSQATVRIVDQCSNGGLDLD
S. verticillata (cont.)			INVFKQLDTNGQGVARGHLMVDYEFVNCGDGVEDSPFLRSVTDAN
Selaginella apoda	MAASTSRVVLALALMLVVALIAVQA	DPTCSPAGRQYCNPGRCCSKFNWCGTGRAYCGRGNCIAGPCWR	SAAGVNNDQFQVSAAANFSSYPAVVHGLATEEELP
Sequoiadendron giganteum-Glau	MAESRLRLALHMLIVLMTVTVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NAAALNQSSLNVSERNLTSGSP
Serenoa repens	MDKIRVMVLMMALAMAMTMMVEA	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCPP	RVSPLPLSQITLSITNHTTKTQAHHPNP
Sideroxylon reclinatum	MGLRVGSVWTMMVVMCATIVGTAVG	QQCGRQAGGALCENNLCCSQWGYCGTTDEYCLPSNNCQSNCRG	GGTPPSGEGTTVRSTYHLYNPQNVGWDLNAVSAYCSTWDAGKPLWWRQKYG
S. reclinatum (cont.)			WTAFCGPVGPTGQAACGKCLLVTNTGTGAQATVRIVDQCSNGGLDLDVNV
S. reclinatum (cont.)			FNQIDTDGKGYQQGHLTVKYQFVDCGDGVFSII
Simmondsia chinensis	MKWAATMVLLLALPYLLVLSSA	QRCGRQGGGQRCRGRLCCSNYGYCGTGSAYCAPGSCQSQCHH	RFVDMGQDFITEATEGYA
Solanum lycopersicum	MKFHVVILMILALLLTATSA	QQCGRQARGRACANRLCCSQYGFCGSTRAYCGFGCQSNCRR	YDATGKGQKLNDDEHKNNDGHN
Solanum lycopersicum	MNKLSSSTPILLALVLCISLTSVTNA	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG	GPTPTPTPGGGAQVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYSWRSKY
S. lycopersicum (cont.)			GWTAFCGPVGPRGRDSCGKCLRVTNTRTGAQTTVRIXXQCSNGGLDLDINV
S. lycopersicum (cont.)			FRQIDTDGVG
Solanum torvum	MEKLNSSTILLGVVLFISIGAIANA	QQCGRQRGGALCDGGLCCSQFGWCGSTPEYCSPSQGCQSQCSG	SGPTPEGSAQNVRATYHLYNPQNVGWDLYAVSAYCSTWDGNKP
S. torvum (cont.)			YSWRSKYGWTAFCGPVGPRGKPPAASA
Solanum tuberosum	MKLPIFTTLSLLFSLLLLIASA	EQCGSQAGGAICASGLCCSKFGWCGNTDTYCGPGNCQSQCPS	VPSPKLPTTDPGPLGDQLGSIISNSMFDQMLKHRNDNVCKGKNNFYSYNA
S. tuberosum (cont.)			FITAARSYPGFGTTGDITTRKREIAAFFAQTSHET
Solanum tuberosum	MVKLSCAPILLVLILCISLTSVANA	QQCGSQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCSG	SGPDPGQGGSAQNVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYAWR
S. tuberosum (cont.)			SKYGWTAFCGPVGPRGRDSCGRCLRVTNTRTGAQTTVRIVDQCSNGGL
S. tuberosum (cont.)			DLDINVFRQIDTDGVGNQQGHLIVNYQFVNCGDNVNVPLLSV
Solanum tuberosum	MVKLSCGPILLALVLCISLTSVANA	QQCGRQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCTG	SGPDPGQGGSAQNVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYAWRS
S. tuberosum (cont.)			KYGWTAFCGPVGPRGRDSCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDL
S. tuberosum (cont.)			DINVFQQIDTDGVGNQQGHLIVNYQFVNCGDNVNVPLFSVI
Solanum dulcamara	SMEKLSSSTILLAVVLFISIGAVANA	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG	SGPIPTPGGSAQNVRATYHFYNPQNVGWDLNAVSAYCSTWDANKPLAWRR
S. dulcamara (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDL
S. dulcamara (cont.)			DINVFRQIDTDGLGNQQGHLIVNYEFVDCGDNVNVPLIMSVVDKE
Solanum lasiophyllum	MEKMNKLSSSTPILLALVLCISLTSVTNA	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG	GPTPTPTPTPGGGAQVRATYHIYNPQNVGWDLNAVSAYCSTWDANKPYSWR
S. lasiophyllum (cont.)			SKYGWTAFCGPVGPRGRDSCGKCLRVTNTRTGAQTTVRIVDQCSNGGLDLD
S. lasiophyllum (cont.)			INVFRQIDTDGVGNQQGHLIVNYQFVNCGDNVNVPLLSVVDKE
Solanum sisymbriifolium	MEKLNSTTSILLALILFILAAVAKA	QQCGRQKGGALCAGNLCCSQFGWCGSTPEYCSPSQGCQSRCSG	SGSTPTTPGGSAQNVRATYHLYNPQNVGWDLNAVSAYCSTWDANKPLAWRR
S. sisymbriifolium (cont.)			KYGWTAFCGPVGPRGRDSCGRCLRVTNTRTGAQTIVRIVDQCSNGGLDLDI
S. sisymbriifolium (cont.)			NVFRQIDTDGVGNQQGHLIVNYQFVDCGDN
Solanum sisymbriifolium	MSKLSSSTILLGMVLFISIGAIANA	QQCGRQMGGALCGGNLCCSQYGWCGSTPEYCSPSQGCQSQCSGSG	PIPTPGGSAQNVRATYHLYNPQNVGWDLNAVSAYCSTWDANKPYAWRS
S. sisymbriifolium (cont.)			KYGWTAFCGPVGPRGQASCGKCLRVTNRRTRAQTTVRIVDQCSNGGLD
S. sisymbriifolium (cont.)			LDINVFRQIDTDGVGYQQGHLFVDYEFVNCGDNVNVNAPLIKSVVDKE
Sphagnum lescurii	MGQYKSTVAALLVIIMFLGCNIGLPMVMG	DPTCSPKGNFFCNSGRCCSVYNWCGSGASYCAKGNCLAQCWP	GVTAVTTYTLTGKNLTSTHNN
Stangeria eriopus	MEYKNLVCVVLVLAVVIGESSVSVKG	DPTCSPAGQVYCNSGRCCSKFNWCGSTAAYCQRPNCIAQCWP	GVTALTLSESLLSLSNYTSEAATMAAENNNP
Staphylea trifolia	MNRVRICMVLLLGLVAGAIA	QQCGSQAGGRTCDNNLCCSQYGYCGTTDEYCLPSNNCQSNCKG	SSPTGGESASNVRATYHYYYPEQNGWDLNAVSAYCSTWDANKPLEWRRKY
S. trifolia (cont.)			GWTAFCGPVGPRGQASCGKCLRVTNTGTGAQATVRIVDQCSNGGLDLDA
S. trifolia (cont.)			GVFQQLDTNGNGIAQGHLIVNYQFVNCGD
Strychnos spinosa	MRNLILILLSLTTFSTILA	EQCGSQAGNALCPNGLCCSQYGWCGSTPEYCTTGCQSQCG	GSSTPIPSNPPAAPGGGGGGDIGSIIXPS
Sundacarpus amarus	MEKISMMALVMVLAXEG	DPSCNPMAKVYCNPGRCCSKFNWCGTLESYCGKNYCIAQCP	TGISPLSPSQILLSTANLTAQTQVHVDPNP

Synsepalum dulcificum	MGMRVASVWTVMMVMCATIIGTAVA	QQCGSQAGGALCANNLCCSQWGYCGDTDEYCLPSNGCQSNCRG	GGTPPGPPTGEGATVRSTYHLYNPQNVGWDLNAVSAYCSTWDASKPLSWRQ
S. dulcificum (cont.)			KYGWTAFCGPVGPTGQAACGKCLLVTNTGTGAQATVRIVDQCSNGGLDLDV
S. dulcificum (cont.)			NVFNQIDTDRRGYQQGYLTVNYQFVDCGDGVFSII
Taiwania cryptomerioides	MVESRLRLPLHILLVLMTVEVRVRG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP	NAVALNQYTFNTSQLNLTSASP
Talinum sp	MKLYVVCALITTLPCFLGTITSA	EQCGSQAGSALCPNKLCCSKFGWCGSTDQYCGPGCQSQCTPSG	PSPTPPTPTPSSGGDVGSIISPTLFDQMLKHRNENS
Tamarix hispida	MTIMSSKHAYTLLLLLSLAISGATAV	QCGTQAGGQLCPNNLCCSQYGWCGSTDDYCSQANNCQSNCNG	GGGGSPPSTGGSGESATNVRATYHLYNPQQNGWNLYAVSAYCSTWDGDK
T. hispida (cont.)			PLSWRSQYGWTAFCGPVGPRGQAACGKCLRV
Tapiscia sinensis	MEKSGLLTALLLCLLAVATA	QQCGRQAGGRTCANNLCCSQYGYCGTTAEYCSQNCQSNCSPG	GGGGGGGGGGGESASNVRATYHLYNPEQHGWDLNAVSAYCSTWDANKPLA
T. sinensis (cont.)			WRQKYGWTAFCGPVGPRGQASCGRCLRVTNRGTSAQETVRIVDQCSNGG
T. sinensis (cont.)			LDLDVGAFNQLDTDRRGYAQGHLMVDYQFVNCGD
Tarenaya hassleriana	MNKQGLSIAFVVLLLYTAAVAG	EQCGRQAGGRTCPDNLCCSQYGYCGDTEDYCSPEKNCQSNCWG	GGGGGGGGGGGESASNIRATYHYYNPEQHNWDLRAVSAYCSTWDAKKPYTW
T. hassleriana (cont.)			RSKYGWTALCGPVGPRGQASCGKCLRITNTRTRAQVTARIVDQCGNGGLDM
T. hassleriana (cont.)			DVGVFNRLDTDGVGYQQGHLIVDYQFVDCGNELLESNSQNTLPSFLDGV
Taxodium distichum	MAESRFRLALYILLVLMAVVVNVKG	DPTCSPAGNFWCNSGRCCSIYNWCGNTSDYCASGNCLAQCWP	NAVALDQSNLSVSRQNLTSAFP
Taxodium distichum	MALNKVSITVLVGLMLLMGNVGVTA	DPTCSPAGNFFCNSGRCCSIYNWCGTGPAYCGKGNCLAQCSP	GVAAVTTYSLAQKNLTSAQLINP
Taxodium distichum	MLLVLMAVMVNVKG	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCAFGNCLAQCWP	NVLALNQSSLSAVQHNLTSASP
Taxodium distichum	MGLNKVSIAVLVVVVLLMGNVGVMA	DPTCSPKGNFWCNNGRCCSIYNWCGTGPAYCSKGNCLAQCSP	GVTAVTTYSLTQKNLTSVNP
Taxodium distichum	MRMSRLTYAICMLLVLMAVMLKVKVKG	DPTCSPAGNFWCDSGRCCSIYNWCGSTSDYCASGNCLAQCWP	YAVALNESSLNLSQQNLTSALHGNP
Taxus baccata	MAENKLRLAFYMLMLMTVMVEG	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP	TVAAVDNSSFSLLQHQNLTSTHNP
Taxus baccata	MVIAVILIIISVSG	DPTCSPAGRFFCNTGRCCSKFNWCGSTAEYCASRNCIAQCWP	GVTATATPSLLTITNYTTPKNHHP
Tellima breviflora	MEILNLCLTIMLLTCLVASATA	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTTNCQSNCKPG	GGGGESAXNVRATYHYYNPAQNGWNLNAVSAYCSTWDANKPLAWRQK
T. breviflora (cont.)			YGWTAFCGPAGPRGQASCGRCLRVTNTRTGAQVTARIVDQCSNGGL
T. breviflora (cont.)			DLDYSGVFKPLDTDGNGYAQGHLIVNYQFVNCGD
Tellima breviflora	MGKISISMLLIICVAYGASA	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGGGGGGAAASNVRATYHYYNPEQHGWDLNAVSAYCSTWDAKKSL
T. breviflora (cont.)			AWRKKYGWTAFCGPAGPRGQASCGKCLRVTNATTGAKTTVRIVDQCSNGGL
T. breviflora (cont.)			DLDVGVFNKLDTDGRGYAQGHLTVNYEFVDCGDGINNPLSSIIEA
Thalictrum thalictroides	MANKIICITFLLCLIGVATA	QNCGRQAGGATCPGNICCSQWGWCGTTDDHCLPSNNCQSNCRG	GGPGPGPG
Thellungiella halophila	MKKSRLSIAIILLSYTVATVAG	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG	SGPSGPGESASNVRATYHFYNPEQNNWDLRAVSAYCSTWDADKPYAWRSKY
T. halophila (cont.)			GWTAFCGPAGPRGQASCGKCLRVRNTRTNAVVTVRIVDQCSNGGLDLDVAM
T. halophila (cont.)			FNRLDTDGVGYQQGHLIVDYQFVDCGNDLIHQPADSKNILVSAIDRV
Theobroma cacao	MKNTQKMGNLSLCLVFLVSLLASTATA	QQCGRQAAGRTCANNLCCSQFGYCGTTNEYCSPSKSCQSNCWPSG	GGGGGESASNVRATYHFYNPPQNGWDLNAVSAYCSTWDANKPLAWRQKYGW
T. cacao (cont.)			TAFCGPVGPRGQASCGRCLRVTNRGTRAEAIVRIVDQCSNGGLDLDAAVFQ
T. cacao (cont.)			QIDTDGRGYAQGHLMVDYQFVNCGD
Theobroma cacao	MDKVNTVSRLLVFLVSLVGAAVA	EQCGWQAGGTICPDNLCCSQYGWCGNTDAYCLPENNCQSNCKSSG	PGGETATVTSTYHFYNPEQHGWDLMAVSAYCSTWDASKPFSWRSKYGWT
T. cacao (cont.)			AFCGPVGPTFPAACGRCLRVTNTRTNAQEIVRIVDRCSNGGLDLDVDV
Theobroma cacao	MKNTQKMGNLSLCLVFLVSLLASTATA	QQCGRQAAGRTCANNLCCSQFGYCGTTDEYCSPSKSCQSNCWPSG	GGGGGESASNVRATYHFYNPPQNGWDLNAVSAYCSTWDANKPLAWRQKYG
T. cacao (cont.)			WTAFCGPVGPRGQASCGRCLRVTNRGTRAEAIVRIVDQCSNGGLDLDAAVF
Thuidium delicatulum	MVRVAPLVALLLCSLLHFASA	QQCGRDVNNTLCADPLNCCSQYGYCGTTDAYCVTGCQSGPCRD	TFSPPPPPAPPSPPSPPPPPRPSVPPSPSSGAGKIINRKLFEALYP
Tinospora cordifolia	MEKTNIFVVLLLCAIGVASA	QQCGRQAGGRTCPGNLCCSQWGYCGDTNEYCSNNCQSNCRPG	AGPGGGGESANNVRATYHIYNPASINWDYLRASVYCATWDANKPLEWRRR
T. cordifolia (cont.)			YGWTAFCGPVGPRGQASCGRCLRVTNTRTGAQATVRIVDQCSNGGLDLDV
T. cordifolia (cont.)			GVFRQLDTDGDGNAKGHLIVNYQFVNC
Torreya nucifera	MKTKAVVFVLVVAVAVGGPXCTA	QNCGSQVDGRVCSGGECCSKWGFCGTTDDHCLPERGCQSNCRG	GGSGEKAYNVRSTYHEYYPERHNWDLYAVSAYCSTWDGGRPLWWRQKYGWT
T. nucifera (cont.)			AFCGPVGPRGQASCGKCLRVTNRETGSQATVRIVDQCSNGGLDLDINVFN
T. nucifera (cont.)			KLDTNGQGVARGHLMVDYEFVGCGDGAHAQFL
Torreya nucifera	MAEKRLRLAFSVLVFMAVKGLA	DPTCSPAGNFWCNSGRCCSIYNWCGSTAEYCAQGNCLAQCWP	SASALDPYQYGLNLLNQQNLTTSPLNNP
Torreya nucifera	MALYKLSVTVLVMLMLTVMGMVMA	DPTCSPSGNFFCNDGRCCSRFNWCGTGPQYCGKGNCIAQCP	PSVTALTSYSLSGKNLTSANP
Torreya taxifolia	MAEKRLRLAFSVLVFMAVKGLA	DPTCSPAGNFWCNSGRCCSIYNWCGSTAEYCAQGNCLAQCWP	SASALDRYQYGLNLLNQQNLTTSPLNNP
Torreya taxifolia	MKMALYKLSVTVLIMLMFTGMAMA	DPTCSPAGNFFCNNGRCCSKFNWCGSGAQYCGKGNCIAQCP	PSVTALTRYSLSGKNLTSANP
Trianthemum portulacastrum	MKLCLVTSVLVIVILPCLLERTSA	EQCGRQAGGARCPTGLCCSQYGWCGTTNAYCGTGCQSGCPG	GGGGGGSTPTPSLP
Triticum aestivum	MAKLAMPLAATFLTLGLAVLLLSAAGPAVA	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVV	SSGAAAAETSGGKPVGSERTR
Triticum turgidum	MAKLAMPLTATFLAFGLAVLLLSAAGPAVA	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVA	SSGAAAAEASGGKPVGERAHP
Triticum turgidum subsp. Duru	MAKLAMPLTATFLAFGLAVLLLSAAGPAVA	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVA	AAAAEASGGKPVGS
Trochodendron araliodes	MKIPSSFTLLLFLCLIGTSLA	QQCGSQAGGRTCPNNICCSQYGYCGTTEDYCSPSKNCQSNCWG	GGGGGGGGGGGGGETATNVRATYHMYEPEKHNWDYMAVSAYCSTWDANKPY
T. araliodes (cont.)			SWRSKYGWTAFCGPVGPRGQASCGKCLRVTNTYTGAQQTVRIVDQCSNGGL
T. araliodes (cont.)			DLDVGVFNRLDTNGRGYAQGHLIVNYQFVDCGDGVNSVMSE
Trochodendron araliodes	MSSPNLWMVVLLCLAAGATA	QQCGRQAGGRTCDGNLCCSQYGYCGTTDDYCSPSKNCQSNCQG	GGGGGGGGESANNVRATYHYYNPEQNGWDLNAVSAYCSTWDAGKSLAWRSK
T. araliodes (cont.)			YRWTAFCGPVGPRGQASCGKCLRVTNTRTGAQRTVRIVDQCSNGGLDLDA
T. araliodes (cont.)			VAFQQLDTDGNGYAQGHLIVNYQFVDCGNLITPLPLLSILD
Tropaeolum majus	MKLLSFIAILSMLVVITTAA	PQCGRQRIGGRCPKGQCCSRFGYCGSGPAHCAPANCYSQCKL	TEPTVTPNAVGGGGGGGGEXKNTLYVLCVYV
Tropaeolum majus	MVNLFLIAIFLSLTVSGIA	QQCGRQAGGKTCRDNLCCSQYGYCGTTDDYCSPAKGCQSNCKG	GGGESASNVRATYHLYNPEQHGWDLFAVSAYCSTWDGTKPFSWRSKYGW

T. majus (cont.)			TAFCGPVGPRGQASCGKCLKVTNTRTGANEIVRIVDQCGNGGLDLDVGVF
T. majus (cont.)			NRLDTDGNGYAQGHLTVNYQFVNCGDAFNSVLRSITY
Tropaeolum peregrinum	MVKAFVIASFVIFLSLAIGGIA	QQCGRQAGGQTCRDNLCCSQYGYCGTTDDYCSPAKNCQSNCKG	SGESASNVRATYHNYNPEQHGWDLMAVSAYCSTWDATKPYSWRSKYGWT
T. peregrinum (cont.)			AFCGPVGPRGQASCGKCLKVTNTRTGANEIVRIVDQCSNGGLDLDVGVF
T. peregrinum (cont.)			NRLDTDGIGYAQGHLTVNYLFVNCGDAFNSLLRSVTY
Urginea maritima	MKAWLLACFLLASLLGSSA	QHCGRQAGGKLCPDNLCCSKWGYCGPTADHCGDECQSGPCYDKQ	PPPPPSSPPPP
Vaccaria hispanica	MVTMKSFALLTLLITLPYLMGVTTA	FQCGRQAGGARCSNGLCCSQFGYCGSTPPYCGAGQCQSQCALD	VPTEALAQAQAWAQAQALQAQAAAGAP
Vigna unguiculata	MGKAGVAIVLLLCLIVTATA	QQCGRQAGGRTCSGNLCCSQYGWCGNTEEYCSPSQNCQSNCWG	GGGSGGGSASNVRATYHYYQPELHGWDLNAVSAYCSTWDAGKSYAWRSKY
V. unguiculata (cont.)			GWTAFCGPVGPTGRDSCGKCLRVTNTGTGSQTTVRIVDQCSNGGLDLDVG
V. unguiculata (cont.)			VFNRLDTDGQGYQRGHLIVNYEFVDCGDNNLNLVLDTPK
Viola canadensis	MGRLSILVLVFFLGFVGSVLA	QDCGRQAGGRTCANNLCCSQWGFCGTTDDHCSPSKNCQSNCRG	GTPTPTPXKC
Vitex agnus castus	MGRVMNTFIVVLLCLTGVAIA	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDSG	EGVGGGSASNVLATYHLYNSQ
Vitis vinifera	MEGKIIYMVAVLLSLVACAAA	QQCGRQAGGRTCANNLCCSQYGYCGTTAEYCSPSQSCQSNCQSG	GGSSGGGGTGASASNVRATYHYYIRSNMDGT
Widdringtonia cedarbergensis	MGKAWVGLVVLLCLIVTAIA	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG	GGGGGGGGGGGGESASNVRATYHYYEPEQHGWDLNAVSAYCSTWDASKPYSWR
W. cedarbergensis (cont.)			SKYGWTAFCGPVGPRGRDSCGKCLRVTNTGTGANTIVRIVDQCSNGGLDLD
W. cedarbergensis (cont.)			VGVFNRIDTDGRGYQ
Widdringtonia cedarbergensis	MKDSKLRLRLHISIILMVVGQTVNG	DPTCSPAGNFWCDSGRCCSIYNWCGDTSDYCASENCLAQCWP	NAAALNQSTMNVAQNNLTSLSP
Xanthocercis zambesiaca	MAKLWVGFVVVLCSIATRIA	EQCGWQAGGQLCPNGLCCSQYGYCGSTADYCSPDKNCQSNCWG	GGGGGGSASNVRATYHYYQPELHGWDLNAVSAYCSTWDAGKPYSWRSKYGW
X. zambesiaca (cont.)			TAFCGPVGPRGQEACGKCLRVTNTGTQAQEIVRIVDQCSNGGLDLDVGVF
X. zambesiaca (cont.)			NRLDTDGRGYQQGHLIVNYEFVECGNELEELTTNPLLSNILE
Ziziphus jujuba	MGKLSHCVFLVLLLCLVAGAAA	QQCGRDVGGKTCDGNLCCSQYGWCGSTEEYCSPSKNCQSNCWG	GGGGGGGGDSGEGSASNVRATYHYYNPEQHGWDLNAVSAYCSTWDANKPY
Z. jujuba (cont.)			SWRSKYGWTAFCGPSGPRGQASCGKCLRVTNTRTGAQKIVRIVDQC
Z. jujuba (cont.)			SXQWRIRFGCWGIQPIGH

Appendix D. Full-length precursors of putative 10C-hevein-like peptides

10C-Hevein-like Peptides

Plant name	Signal peptide	Mature domain	C-terminal tail
Citrus aurantium	MRLIGSLLIFSLVLSFVLGGSA	GNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCGE	CDPDDGTAGDGGELGKIISRKMFEDLLEYRNDKRCPARCFYTYDAFIEAA
C. aurantium (cont.)			KAFPAFGNSGNETMRKREIAAFFAQTGHETTGGWPDAPGGEYAWGYCFNR
C. aurantium (cont.)			EVGAASSDYCDPNYPCRGKYYGRGPIQLSWNYNY
Citrus sinensis	MRLTWSLLIFSLVLSFVLGGSA	QNCGSPVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCGE	CDPDDGTAGDGGELGKIISRKMFEDLLEYRNDKRCPARCFYTYDAFIEAA
C. sinensis (cont.)			KAFPAFGNSGNETMRKREIAAFFAQTGHETTGGWPDAPGGEYAWGYCFNR
C. sinensis (cont.)			EVGAASSDYCDPNYPCRGKYYGRGPIQLTWNCNYLRCGEALGVD
Cuscuta_pentagonia	MKLTSSSSSSVLVITTMALAMMLIATTDAF	QCGGQAGGALCANRWCCSRWGYCGLSCEYCGNSCQSYCHPDCRGV	AIKFDNETDVVNTKPMDAGGLN
Elymus spicatus	MKPHMSTRGTRVAAILLAVVLAAMLATAVNG	AQKCGDQARGAKCPNCLCCGKYGFCGSTPDYCDVGCQSQCRGCRDG	VVGQALPTESDPTRAAAASSASATGLNLTASTRGP
Gnetum_montanum	MISAKKMQRPAMMRIVVAALVMSSMLVTEG	SILVTEVSSGGECNCVDKLCGEGFCCSKWGFCGTGEAWCGTGNCQCQCPS	VPPPPPSPRPPSPPRSPPPRPPSPPRSPPPRRPSLTSALL
Lolium perenne	MMRGLSVVAILAAAFAVSAHA	QQCGSQAGGATCANCLCCSQYGYCGSTSAYCGAGCQSQCNGCGG	TPTPTPSGGVSSIISQSLFDQMLLH
Oryza sativa Japonica Group	MRALAVVVVATAFAVVAVRG	EQCGSQAGGALCPNCLCCSQYGWCGSTSAYCGSGCHSQCSGSCGG	GGPTPPSGGGGSGVPSIAVALALPTDASXTERRGVP
Oryza sativa Japonica Group	MRALAVVAMVATAFLAAAVHA	EQCGSQAGGAVCPNCLCCSQFGWCGSTSDYCGAGCQSQCSAAGCG	GGGPTPPSGSGGSGVASIVSRS
Panicum virgatum	MMRGGGAVLTAAVALVLLAVAMAMSTTVRA	QQQCGSEAGGKLCPNCLCCSQWGYCGSTSDYCGDGCQSQCHGCDVAS	IISRSVFDDLLPYRDDALCPARGFYTYDAFIAAANAFPGFGTTGDAETR
P. virgatum (cont.)			KRELVAFLAQTSHLTTGGGTHSPGGPYAWGYCFKEQVEDLRADYCRPSN
P. virgatum (cont.)			RWPCVAGKQYYGRAPSSSPGTTTTPGREGIGG
Plagiochila asplenioides	MGTIHKSSASVIMRLILIGALVGVALA	VRECGPSNPCKGGECCSQFGSCGCSNAYCGKGCIDNCNG	CGGRWATSAGSNQATYYTTYIPSSCYGFDTSKFPAGNLIAAASSDVFRNR
P. asplenioides (cont.)			AACGTYYEITCKGSATGGGNPCKRNPAVKVMVVDLCPGCNQNSFDLSKEA
P. asplenioides (cont.)			FSKIADLDAGRISISAKRVSKYHRTVEETITEVV
Poncirus trifoliate	MRLIGSLLVFSLVLSFVLGGSA	QNCGSGVVCPGGECCSQYGWCGITTDHCCGGCQSDCNQVVCGE	CDPDDGTAGDGGELGKIISREMFAHLLEYKTDHRCPARCF
Solanum dulcamara	MKFSLALVFMILALLVTTTYA	EQCGRQSGRRKCPNRLCCSKFGWCGTTCEYCGSGCQSNCRGGCATT	MFSNETVNNSGKHLEDGTLN
Triticum aestivum	MRAFALFAVLAMAAAMAVAE	QCGSQAGGATCPNCLCCSRFGWCGSTSDYCGDGCQSQCSGCGSTP	VTPTPSGGGGVSSIISRALFDRMLLHRNDGACQAKGFYTYDAFVAAAGAF
T. aestivum (cont.)			RGFGTTGSTDTRKREVAAFLAQTSHETTGGWATAPDGAFAWGYCFKQERG
T. aestivum (cont.)			ATSNYCTPSAQWPCAPGKSYYGRGPIQLSHNYN
Triticum aestivum	MKPDMSATALRAPRVAAILLAVVLAAVLATAVNG	DQKCGDQARGAKCPNCLCCGKYGFCGSGDAYCGEGSCQSQCRGCRD	DVVGQALPAEPGSTRATAASSASATGLNLTATTGGP
Triticum aestivum	MSAPRAPPSLGLATAAMAILAVLAAALATTARA	QTCGSQAGGARCPNCLCCSRFGFCGSGSEWCGAGCQSQCSGCPAP	GGQGVASIVPRDLFERLLLHRNDAACLARGFYTYDAFLAAAAAFPAFAGT
T.aestivum (cont.)			SEGLSVETRKREVAAFLGQTSHETTGGVAVRARRPLLVGATASSRERDPPS
T.aestivum (cont.)			DYCQPRP
Triticum turgidum	MRGVVVVAMLAAAFAVSAHA	EQCGSQAGGATCPNCLCCSKFGFCGTTSDYCGTGCQSQCNGCSGG	TPVPVPTPSGGGVSSIISQSLFDQMLLHRNDAACLAKGFYNYGAFVAAANS
T. turgidum (cont.)			FSGFATTGSTDVKKREVAAFLAQTSHETTGGWPTAPDGPYSWGYCFNQER
T. turgidum (cont.)			GATSDYCTPSSQWPCAP

Plant name	Accession number
6C-Hevein-like Peptides	
Aerva lanata	PDQH-2053815
Aerva persic	GBCQ-2018132
Alternanthera caracasana	WGUG-2017653
Alternanthera sessilis	BWRK-2077393
Alternanthera sessilis	LUNL-2061656
Alternanthera sessilis	OYST-2062883
Amaranthus retroflexus	WMLW-2009398
Amaranthus retroflexus	WMLW-2061382
Atriplex hortensis	AKTA-2039893
Atriplex prostrata	MUCT-2036997
Atriplex rosea	PDXY-2042561
Beta maritima	FVXD-2056435
Chamaseyce mesebyranthemum	RHAU-2012379
Chamaseyce mesebyranthemum	RHAU-2051397
Flaveria bidentis	QBGG-2040842
Heliotropium calcicola	RGMN-2104521
Heliotropium tenellum	DIHD-2019944
Heliotropium texanum	MDJK-2144026
Kochia scopari	WGET-2036072
Portulaca mauii	BLWH-2018971
8C-Hevein-like Peptides	
Actinidia deliciosa	FG478012.1
Actinidia eriantha	FG507620.1
Actinidia setosa	FG481050.1
Actinidia chinensis	FG459770.1
Agathis robusta	MIXZ-2008294
Ajuga reptans	UCNM-2047888
Alangium chinense	FWBF-2012524
Allium commutatum	KBXS-2000239
Alnus glutinosa	FQ341594.1
Amentotaxus argotaenia	IAJW-2003903
Amentotaxus argotaenia	IAJW-2018454
Amentotaxus argotaenia	IAJW-2018453
Amentotaxus argotaenia	IAJW-2126957
Ammopiptanthus mongolicus	JZ389716.1
Ammopiptanthus mongolicus	JZ471525.1
Anemone hupenhensis	ZUHO-2009007
Anomodon attenuatus	QMWB-2011191
Aphanopetalum resinosum	TOKV-2053682
Apios americana	NXOH-2005138
Apios americana	NXOH-2005139
Arabidopsis thaliana	R90112.1
Arabis alpina	TZWR-2041647

Arachis duranensis	GW946290.1
Arachis duranensis	GW956564.1
Arachis hypogaea	FS977976.1
Arachis hypogaea	JK205505.1
Arachis hypogaea	JK209988.1
Arachis ipaensis	GW960675.1
Arachis ipaensis	GW991589.1
Argemone mexicana	BFMT-2017722
Asparagus densiflorus	FGRF-2048625
Astilbe chinensis	CKKR-2060572
Astragalus membranaceus	HJMP-2014725
Astragalus propinquus	MYMP-2058807
Athrotaxis cupressoides	XIRK-2000808
Athrotaxis cupressoides	XIRK-2000807
Athrotaxis cupressoides	XIRK-2000806
Atropa belladonna	BOLZ-2159794
Austrocedrus chilensis	YYPE-2002937
Austrocedrus chilensis	YYPE-2002936
Austrocedrus chilensis	YYPE-2029866
Austrotaxus spicata	BTTS-2032679
Austrotaxus spicata	BTTS-2072841
Austrotaxus spicata	BTTS-2072606
Basella alba	CTYH-2105366
Bauhinia tomentosa	JETM-2091845
Bazzania trilobata	WZYK-2087147
Boehmeria nivea	ACFP-2028246
Boswellia sacra	FCCA-2043010
Boykinia jamesii	OOVX-2042183
Brassica napus	FG577475.1
Brassica napus	DW153474.1
Brassica napus	EE445484.1
Brassica oleracea	EX114407.1
Brugmansia sanguinea	AIOU-2068307
Bryum argenteum	JMXW-2009405
Bursera simaruba	JSZD-2010254
Caiophora chuquitensis	VTLJ-2010251
Cajanus cajan	GW350060.1
Callitris gracilis-March	IFLI-2000417
Callitris macleayana-March	RMMV-2006113
Callitris macleayana-March	RMMV-2002319
Callitris macleayana-March	RMMV-2004820
Calocedrus decurrens	FRPM-2002883
Calocedrus decurrens	FRPM-2008761
Camelina sativa	XM_010419957.1
Camellia sinensis	FS957358.1
Cannabis sativa	BJSW-2013786
Cannabis sativa	PDIE-2001895
Capsella rubella	XM_006298557.1

Capsicum annuum	BM066673.1
Capsicum annuum	EL814425.1
Carica papaya	EX282008.1
Cassytha filiformis	VYLQ-2036312
Castanea mollissima	GO920272.1
Casuarina glauca	LNER-2050327
Celtis occidentalis	KYAD-2004733
Celtis occidentalis	KYAD-2050740
Cercidiphyllum japonicum	NUZN-2000485
Cercis canadensis	RKFX-2042782
Cicer arietinum	XM_004494647.2
Cicer arietinum	GR404613.1
Cinnamomum_camphora	BCGB-2043744
Citrus aurantium	FC874612.1
Clematis chinensis	JG580697.1
Cleome viscosa	UPZX-2042161
Conopholis americana	FAMO-2092069
Coriaria nepalensis	NNGU-202373
Crassula perforata	KWGC-2013004
Crossopetalum rhacoma	IHCQ-2093561
Cucumis melo	CN845588.1
Cucumis melo	JG498505.1
Cunninghamia lanceolata	OUOI-2010547
Cupressus dupreziana	QNGJ-2000610
Cupressus dupreziana	QNGJ-2067951
Cupressus dupreziana	QNGJ-2000609
Cycas micholitzii	XZUY-2048192
Datura metel	JNVS-2004019
Deutzia scabra	OTAN-2004858
Dicranum scoparium	NGTD-2096910
Dillenia indica	EHNF-2006841
Dioon edule	WLIC-2119047
Diselma archeri	GKCZ-2070649
Diselma archeri	GKCZ-2068869
Dombeya burgessiae	WYIG-2035881
Dombeya burgessiae	WYIG-2180644
Draba hispida	GTSV-2005036
Draba hispida	GTSV-2008655
Draba ossetica	LJQF-2059299
Drimys winteri	WKSU-2129629
Elaeocarpus sylvestris	THHD-2009061
Ephedra sinica	VDAO-2033117
Eschscholzia californica	RKGT-2004462
Euptelea pleiosperma	QTJY-2059221
Eutrema halophilum	BY825221.1
Eutrema salsugineum	XM_006408097.1
Fagopyrum tataricum	JK729372.1
Falcatifolium taxoides	QHBI-2007953

Falcatifolium taxoides	ROWR-2009883
Ficus pumila	HQ224868.1
Ficus religiosa	EDHN-2049024
Flaveria bidentis	ZNZC-2010710
Fokienia hodginsii	UEVI-2033737
Fokienia hodginsii	UEVI-2008402
Fokienia hodginsii	UEVI-2056865
Fontinalis antipyretica	DHWX-2066666
Galax urceolata	ADHK-2058358
Galax urceolata	UFHF-2060702
Galega orientalis	AY952142.1
Geranium carolinianum	VKGP-2118929
Geranium maculatum	YGCX-2138168
Geum quellyon	DHAW-2015118
Gleditsia sinensis	VHZV-2087875
Gleditsia triacanthos	GEHT-2054485
Glycine max	FG990284.1
Glycine max	XM_003554661.3
Glycyrrhiza glabra	PEZP-2068102
Glycyrrhiza uralensis	FS265023.1
Gompholobium polymorphum	VLNB-2018423
Gossypium hirsutum	DW516477.1
Gossypium raimondii	XM_012637384.1
Gyrocarpu samericanus	BSVG-2014470
Gyrostemon ramulosus	UAXP-2077977
Halocarpus bidwillii	OWFC-2045614
Halocarpus bidwillii	OWFC-2009897
Hamamelis virginiana	DMKC-2000892
Hedera helix	EITK-2139802
Heuchera sanguinea	ERIA-2073359
Hevea brasiliensis	HS993022.1
Hevea brasiliensis	HS994451.1
Hevea brasiliensis	JZ538450.1
Hibiscus cannabinus	OLXF-2017798
Hibiscus cannabinus	OLXF-2090625
Humulus lupulus	AQGE-2045110
Humulus lupulus	ES655104.1
Hypecoum procumbens	NMGG-2058911
Impatiens balsamifera	JEXA-2062741
Ipomoea nil	NHAG-2048417
Isoetes tegetiformans	PKOX-2006241
Itea virginica	UWFU-2035071
Juglans nigra	DXQW-2058926
Juniperus scopulorum	XMGP-2055217
Larrea tridentata	UDUT-2047060
Lathyrus sativus	KNMB-2055925
Laurelia sempervirens	WAIL-2059125
Lepidosperma gibsonii	WBIB-2057434

Limnanthes douglassii	CRNC-2000081
Limonium spectabile	WOBD-2081469
Linum hirsutum	HNCF-2000289
Linum leonii	AEPI-2048092
Linum lewisii	BHYC-2052042
Linum macraei	KCPT-2102059
Linum strictum	TXMP-2007103
Linum usitatissimum	OGSY-2094087
Liriodendron tulipifera	CK764856.1
Loropetalum chinense	TQOO-2139966
Lotus japonicus	FS322371.1
Lotus japonicus	AV421457.1
Lotus japonicus	BI419767.1
Lotus japonicus	BM060586.1
Lycium barbarum	LWCK-2001931
Lycium sp	OSMU-2011251
Lycopersicon cheesmanii	UGJI-2125875
Maesa lanceolata	DTOA-2089413
Magnolia grandiflora	WBOD-2101820
Malus domestica	EB105992.1
Manihot esculenta subsp. Peruviana	JG999405.1
Manoao colensoi	CDFR-2008765
Marchantia paleacea	IHWO-2059313
Marchantia polymorpha	JPYU-2036258
Medicago truncatula	XM_013595074.1
Meliosma cuneifolia	AALA-2011529
Metasequoia glyptostroboides	NRXL-2000516
Metasequoia glyptostroboides	NRXL-2058286
Metasequoia glyptostroboides	NRXL-2001780
Microbiota decussata	XQSG-2037376
Microstrobos fitzgeraldii	BBDD-2007819
Microstrobos fitzgeraldii	BBDD-2070246
Microstrobos fitzgeraldii	BBDD-2004959
Microstrobos fitzgeraldii	BBDD-2071547
Mitella pentandra	DAYQ-2052183
Monotropa uniflora	LRTN-2089067
Monotropa uniflora	LRTN-2016473
Morinda citrifolia	ULGV-2007220
Moringa oleifera	CZPV-2001465
Morus notabilis	XM_010103989.1
Morus nigra	XVJB-2002447
Mydocarpus sp	AJFN-2095016
Myriophyllum aquaticum	IUSR-2074233
Nageia nagi	UUJS-2000623
Neocallitropsis pancheri	JDQB-2004699
Neocallitropsis pancheri	JDQB-2064704
Nicotiana tabacum	FS393955.1
Nicotiana tabacum	FS392106.1

Nyssa ogeche	VUSY-2130965
Ochna mossambicensis	TVCU-2008600
Oncotheca balansae	PVGM-2013088
Oncotheca balansae	PVGM-2142572
Ophioglossum vulgatum	QHVS-2011139
Papaver setigerum	FNXH-2023053
Papuacedrus papuana	OVIJ-2005726
Papuacedrus papuana	OVIJ-2001691
Papuacedrus papuana	OVIJ-2082654
Parasitaxus usta	JZVE-2002559
Parasitaxus usta	JZVE-2002557
Parasitaxus usta	JZVE-2003275
Passiflora caerulea	SIZE-2011788
Peganum harmala	OBTI-2041811
Pereskia aculeata	JLOV-2001160
Peumus boldus	KRJP-2083636
Phaseolus vulgaris	FE686122.1
Phellodendron amurense	PGKL-2070733
Phellodendron amurense	PGKL-2050474
Philonotis fontana	ORKS-2002446
Philonotis fontana	ORKS-2002448
Phyllocladus hypohyllus	JRNA-2063583
Phyllocladus hypohyllus	JRNA-2063388
Physcomicromitrium sp	YEPO-2070125
Pilgerodendron uviferum	ETCJ-2001405
Pilgerodendron uviferum	ETCJ-2008954
Pinus taeda	CF396457.1
Pinus parviflora	IIOL-2071721
Pisum sativum	GH719715.1
Platycladus orientalis	BUWV-2040465
Platycladus orientalis	BUWV-2044272
Platycladus orientalis	BUWV-2045303
Plumbago zeylanica	CB818080.1
Podocarpus coriaceus	SCEB-2044852
Polypodium plectolens	ZQYU-2018462
Polyscias fruticosa	EDBB-2084228
Populus deltoids	CK318298.1
Populus euphratica	XM_011040263.1
Populus euphratica	AJ778967.1
Populus euphratica	AJ776776.1
Populus nigra	DB889512.1
Populus tremula	DN496890.1
Populus trichocarpa	DN485080.1
Populus trichocarpa	XM_006375777.1
Populus trichocarpa	DT474845.1
Populus trichocarpa	BU874345.1
Prumnopitys andina	EGLZ-2020031
Pseudotaxiphyllum elegans	QKQO-2004260

Pseudotaxus chienii	YLPM-2033428
Pseudotaxus chienii	YLPM-2037863
Pyrenacantha malvifolia	QZZU-2062507
Quercus petraea	CU656360.1
Quillaja saponaria	OQHZ-2016253
Raphanus raphanistrum subsp. Maritimus	EY897555.1
Reseda odorata	SWPE-2000451
Retrophyllum minus	VGSX-2061426
Rhamnus japonica	EILE-2037280
Rhodiola rosea	ZJUL-2011613
Rhodiola rosea	JUL-2011100
Ricinus communis	PAZJ-2060672
Robinia pseudoacacia	JK974151.1
Roridula gorgonias	HUSX-2031570
Ruellia brittoniana	AYIY-2073961
Salix dasyclados	IEPQ-2024372
Salix fargesii	RZTJ-2087618
Saxegothaea conspicua	QCGM-2071011
Schizolaena sp.	WMUK-2091805
Schwetschkeopsis fabronia	IGUH-2160246
Sciadopitys verticillata	YFZK-2040645
Sciadopitys verticillata	YFZK-2040888
Selaginella apoda	LGDQ-2003229
Sequoiadendron giganteum-Glaucum	QFAE-2000266
Serenoa repens	HXJE-2113249
Sideroxylon reclinatum	OXYP-2002914
Simmondsia chinensis	CVDF-2094468
Solanum lycopersicum	BW688395.1
Solanum lycopersicum	FS204664.1
Solanum torvum	FS109782.1
Solanum tuberosum	AJ878108.1
Solanum tuberosum	AM907316.1
Solanum tuberosum	CN216370.1
Solanum dulcamara	GHLP-2043688
Solanum lasiophyllum	DLAI-2102397
Solanum sisymbriifolium	NMDZ-2017071
Solanum sisymbriifolium	NMDZ-2017724
Sphagnum lescurii	GOWD-2075409
Stangeria eriopus	KAWQ-2005435
Staphylea trifolia	PTLU-2099986
Strychnos spinosa	GGJD-2019841
Sundacarpus amarus	KLGF-2006233
Synsepalum dulcificum	WRPP-2126204
Taiwania cryptomerioides	QSNJ-2000100
Talinum sp	LKKX-2068262
Tamarix hispida	EH051455.1
Tapiscia sinensis	WWKL-2010000
Tarenaya hassleriana	XM_010526614.1

Taxodium distichum	FHST-2007221
Taxodium distichum	FHST-2004844
Taxodium distichum	FHST-2004843
Taxodium distichum	FHST-2003453
Taxodium distichum	FHST-2061111
Taxus baccata	WWSS-2000781
Taxus baccata	WWSS-2000780
Tellima breviflora	CTSS-2001567
Tellima breviflora	CTSS-2043698
Thalictrum thalictroides	GBVZ-2064238
Thellungiella halophila	AK352622.1
Theobroma cacao	XM_007030502.1
Theobroma cacao	CU471310.1
Theobroma cacao	CU495512.1
Thuidium delicatulum	EEMJ-2032575
Tinospora cordifolia	JG647038.1
Torreya nucifera	HQOM-2008514
Torreya nucifera	HQOM-2011119
Torreya nucifera	HQOM-2011118
Torreya taxifolia	EFMS-2012045
Torreya taxifolia	EFMS-2012045
Trianthemum portulacastrum	TNVE-2099453
Triticum aestivum	BG604435.1
Triticum turgidum	BF293407.1
Triticum turgidum subsp. Durum	AJ716574.1
Trochodendron araliodes	SWOH-2054532
Trochodendron araliodes	SWOH-2054532
Tropaeolum majus	GH162958.1
Tropaeolum majus	GH164293.1
Tropaeolum peregrinum	MYZV-2053677
Urginea maritima	KOFB-2059461
Vaccaria hispanica	JZ158858.1
Vigna unguiculata	FG848837.1
Viola canadensis	NJLF-2011087
Vitex agnus castus	DMLT-2086965
Vitis vinifera	FQ438078.1
Widdringtonia cedarbergensis	AUDE-2007932
Widdringtonia cedarbergensis	AUDE-2002285
Xanthocercis zambesiaca	ZSSR-2105305
Ziziphus jujuba	ZHEE-2005858

10C-Hevein-like Peptides

Citrus aurantium	DN622484.1
Citrus reticulata	EY794233.1
Citrus sinensis	CF835995.1
Cuscuta_pentagonia	AHRN-2012568
Elymus spicatus	FF360240.1
Gnetum_montanum	GTHK-2007511

Lolium perenne	GR513336.1
Oryza sativa Japonica Group	CI631870.1
Oryza sativa Japonica Group	CI756576.1
Panicum virgatum	FL945970.1
Plagiochila asplenioides	NWQC-21549421
Poncirus trifoliate	EY833419.1
Solanum dulcamara	GHLP-2050216
Triticum aestivum	BQ805737.1
Triticum aestivum	GH724149.1
Triticum aestivum	JZ886728.1
Triticum turgidum subsp. Durum	AJ612726.1