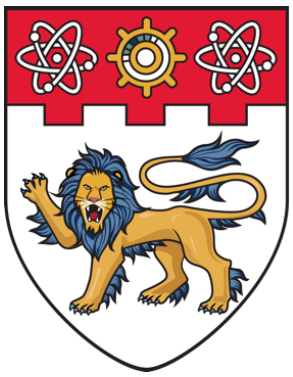


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

DISCOVERY AND CHARACTERIZATION OF HEVEIN-LIKE PEPTIDES IN FUNCTIONAL FOOD AND MEDICINAL PLANTS

STEPHANIE VICTORIA TAY CHOO GEK

SCHOOL OF BIOLOGICAL SCIENCES

2020

DISCOVERY AND CHARACTERIZATION OF HEVEIN-LIKE PEPTIDES IN FUNCTIONAL FOOD AND MEDICINAL PLANTS

STEPHANIE VICTORIA TAY CHOO GEK

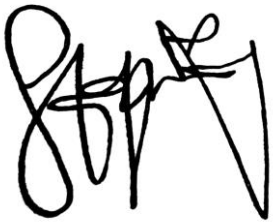
SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.



25th August 2020

. . . . . .

Date

. . . . . .

Stephanie Victoria Tay Choo Gek

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

25th August 2020

. . . . . .

Date

. . . . . .

Professor James P. Tam

Authorship Attribution Statement

*(A) This thesis does not contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author

25th August 2020

. . . . . .

Date

. . . . . .

Stephanie Victoria Tay Choo Gek

Acknowledgements

The four years of my PhD journey has truly been a fulfilling one. The completion of this major milestone is owed to all the people who have motivated, supported and challenge me throughout my journey.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor James P. Tam for providing me the opportunity to pursue my research in his laboratory. His guidance along with great patience and immense care has encouraged me at all stages of my thesis work. This thesis has been possible only because of his help and support. I will forever be grateful to him for providing me with the endless resources and constructive advices for my research and as a researcher. I would also like to thank my co-supervisor, Associate Professor Liu Chuan Fa and my thesis advisory committee members, Associate Professor Newman Sze and Associate Professor Julien Lescar for their consistent guidance and advice throughout my PhD.

This thesis would have not been possible without the encouragement and support from my awesome lab members. Special thanks to Dr. Wong Kaho for being a wonderful and insightful mentor, who introduce my first glance of research into the world of bioinformatics. I would also like to thank Dr. Nguyen Kien Truc Giang for showing me the ropes in the fundamentals of extraction and isolation of peptides. Dr. Bamaprasad Dutta for assisting me with the functional assays and his guidance. I am also blessed to form many meaningful friendships in this lab with Dr. Huang Jiayi, Dr. Tan Wei Liang, Dr. Shruthi Kini, Dr. Geeta Kumari, Tang Fan and Shuan Tan. Our lunch breaks, dinner and gatherings will always be deeply cherished. I would also like to thank our collaborators Dr. Fan Jingsong for helping me with the NMR experiments and solving the peptides' structures and, Dr. Aida Serra for teaching me and assisting me with proteomic analysis work. I am also thankful to all the URECA, FYP and BS9001 students who worked with me and help me learn along with them and development my mentorship skills.

I believe that a strong support group is essential to survive the challenging journey towards obtaining a doctorate. I am extremely blessed to have befriended wonderful people locally and internationally that I can called friends forever during this journey. A shoutout to Chinmayi Prasana and Harvey Chan for being my best mates throughout this whole journey and providing me the emotional support system to face the toughest day. I will always remember our mini celebrations we had when we hit a certain milestone during our PhD journey. I would also like to express my heartfelt appreciation to my best friend Miss Brenda Ong for providing a listening ear and undying support. I am deeply and wonderfully blessed to have such wonderful friends to share my happy moments and I look forward to more in the future.

This PhD journey would not been a fruitful and well-rounded one without the help of fellow residential mentor peers from North Hill which have greatly encouraged me to not only excel academically but also lend a helping hand to undergraduates who are pursuing their bachelors. I would like to express my heartfelt appreciation to Associate Professor Goh Wang Ling for her endless assistance and giving me the opportunity to be a residential mentor in Tanjong Hall. I've learnt so much under her mentorship and to develop myself as a better role model for the undergraduates, and to make their academic journey more enjoyable.

I would also like to take this opportunity to acknowledge the financial support rendered by Nanyang Technological University for presenting me with a research scholarship so that I'm able to pursue my PhD academic journey.

Finally, there are no words to express how truly fortunate I am to have my parents and brother that showed their understanding, love and undying selfless support. They have been my pillar or strength when times were rough and my greatest source of inspiration. With this robust support system, I was always able to overcome challenges with ease and with every step achieved during this journey, they were always there to celebrate with me. It is their constant prayers and blessing that have help me come this far and I hope that I have done them proud.

Table of Contents

List of Figures

List of Tables

Abbreviations

6C	Six cysteine
8C	Eight cysteine
10C	Ten cysteine
A. alternata	*Alternaria alternate*
A. sativa	*Avena sativa*
ACN	Acetonitrile
AMP	Antimicrobial peptide
BLAST	Basic local alignment search tool
C. lunata	*Curvularia lunata*
C. quinoa	*Chenopodium quinoa*
cDNA	Complementary deoxyribonucleic acid
COSY	Correlation spectroscopy
CRP	Cysteine-rich peptide
C-tail	C-terminal tail
D_2O	Deuterium oxide
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
ER	Endoplasmic reticulum
EST	Expressed sequenced tags
EtOH	Ethanol
FA	Formic acid

F. oxysporum	*Fusarium oxysporum*
GlcNAc	N-acetylglucosamine
HLP	Hevein-like peptide
HPLC	High performance liquid chromatography
IC_{50}	Half-maximal inhibitory concentration
IAA	Iodoacetamide
IC_{50}	Half maximal inhibitory concentraton
kDa	kilo Dalton
MALDI-TOF MS	Matrx-assisted laser desorption/ionization time of flight mass spectroscopy
MeOH	Methanol
MS	Mass spectrometry
MS/MS	Tandem mass spectroscopy
M.W	Molecular weight
NaCl	Sodium Chloride
NEM	N-ethylmaleimide
NH_4HCO_3	Ammonium bicarbonate
NaH_2PO_4	Sodium dihydrogen phosphate
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser effect spectroscopy
R. solani	*Rhizoctonia solani*
RP	Reverse-phase

SCX	Strong cation exchange
tBLASTn	Translated nucleotide BLAST
TCEP	Tris(2-carboxyethyl)phosphine
TCM	Traditional Chinese Medicine
TFA	Trifluoroacetic acid
UPLC	Ultra-performance liquid chromatography

Abstract

Hevein-like peptides (HLPs) are a family of cysteine-rich, structurally compact, plant defense peptides. However, they remain underexplored and poorly understood. The objectives of my thesis consist of two parts. The first part focuses on the discovery of novel HLPs and to examine their structure, biosynthesis, distribution, occurrence and functions in selected plants. They include two functional grains, *Chenopodium quinoa* (quinoa), *Avena sativa* (oats), and a medicinal plant, *Eleutherococcus trifoliatus* (Three-leaved *Eleutherococcus*， 苦力心). From these plants, I discovered 13 novel HLPs: three 6C-chitin-binding HLPs, nine 8C-chitin-binding HLPs and a 8C-non-chitin-binding HLP. The second part focuses on data mining using bioinformatics to identify 1204 novel HLPs from 252 plant families. Together this study suggests that HLPs are one of the largest family of CRPs and provides insights into their sequence, molecular and functional diversity, potentially leading to development of novel peptidyl drugs, antimicrobial agents and pathogen-resistant transgenic crops.

CHAPTER ONE

Introduction

1. Emerging Trends of Crop Protectants in Organic and Functional Food

The “quest for good health” is increasingly popular in this present era. As such, there has been a paradigm shift from “convention” nutrition to organic and functional food. Functional food are food that provide added health advantages beyond the provision of essential nutrients [1]. Some examples include fruits, vegetables, legumes and whole grain. According to a global functional food study, the global function food size market projected to reach an astounding USD 275.77 billion by 2025 to cater to the increasing demand [2]. Specifically, the whole grains account for approximately 80% of total food consumption. However, these crops are highly susceptible to pests and pathogens. Reports have shown that these pests and pathogens destroyed at least 33% of the global food production valued at USD 100 billion [3]. Therefore, developing strategies to counteract pests and pathogens has become the main driving force to improve agricultural products, organic and functional foods. The conventional method of pest control employs chemical insecticides and fumigation that have detrimental effects on the environment and also significant health risks to consumers [4]. Furthermore, emerging pesticidal resistance in many insects also poses a serious threat [5]. Therefore, the continual search for an alternative pesticidal compound is required.

Substantial efforts and development of plant-derived pesticides have begun to complement or to replace existing pesticides over the past decades. Naturally-occurring pesticides from plants can provide effective control of pests and are environmentally safe. Plant-derived biopesticides/ crop protectants found in families such as Annonaceae, Asteraceae, Canellaceae, Labiatae, Meliaceae and Rutaceae are uprising crop protectant. The bioactive constituents of these families act as insect repellents, feeding deterrents, insect toxicants, growth inhibitors, chemosterilants, and attractants [6]. These insecticidal phytochemicals have different molecular targets as crop protectants. They include azadirachtin, pyrethrin, nicotine, rotenone, ryania and sabadilla [7].

Azadirachtin, a secondary metabolite isolated from *Azadirachta indica* acts as a feeding deterrent against a broad spectrum of pests with a median lethal dose range LD_{50} of 13 000 mg/kg [8, 9]. Pyrethrin, an organic compound isolated from *Chrysanthemum cinerariaefolium* targets the sodium and potassium ion channels in the nerve cells of pests and cause extended nerve firing [10]. Nicotine, a potent parasympathomimetic alkaloid isolated from *Nicotiana tabacum* is by far one of the most toxic naturally-occurring botanical pesticides with an LD_{50} of 50–60mg/kg [11, 12]. In contrast, rotenone, an isoflavoid isolated from the roots of legumes, from the Fabaceae family can interfere with the electron transport chain, leading to generation of oxygen reactive species that cause mitochondrial dysfunction [13, 14]. Furthermore, a study has shown that this bioactive compound can also inhibit microtubule assembly [15]. Ryania, an alkaloid isolated from *Ryania speciose*, is an insecticide that can bind to ryanodine calcium release channel and alters the

calcium release from sarcoplasmic reticulum, which cause muscle contraction and rapid cell death [16, 17]. Finally, sabadilla, an alkaloid isolated from *Schoenocaulon officinale* can target the nerve cell and cause paralysis in many insects [18]. Collectively, these plant-derived protectants have different modes of action against pests. Despite their advantages, certain drawbacks need to be addressed [19]. The broad spectrum of botanical pesticides are non-discriminatory and can impact beneficial insects. Besides, most of these plant-derived protectants are small molecules that tend to be easily degraded under the sunlight exposure and require frequent applications.

Over the past twenty years, considerable efforts have shifted to explore protectants in the chemical spaces of peptides and proteins as alternative solutions to complement their small-molecules counterparts. The rise in proteinaceous biopesticide research brought to light a spectacular array of bioactive compounds from plants. For example, early studies by Lipke *et al* showed that soyin, a trypsin inhibitor isolated from soya beans, was toxic to the larvae of *Tribolium confusum* (flour beetle) [20]. In addition, Brunelle *et al* showed the fusion of cathepsin D inhibitor from tomato to the N-terminus of corn cystatin II produced a hybrid inhibitor against the digestive proteinase of *Leptinotarsa decemlineata* (potato beetle) [21]. Following these studies, many novel protein/peptide-based biopesticides have been identified [22-25]. Although these biopesticides demonstrated their ability, they are still not on equal competitive ground with their chemical counterpart when factors such as cost of production, ease of delivery and kill efficiency are taken in to consideration [26]. As such, biopesticides may not replace the use of chemical

pesticide in a large-scale setting but may complement it effectively. However, with the current advances in technology, it is foreseen that more protein/peptide-based biopesticide with enhanced potency against pests and pathogens will be discovered to facilitate agriculture processes and long-term storage of crops.

2. Natural Products in Drug Discovery

Natural products have revolutionized the drug discovery field and made remarkable contributions to the development of novel biotherapeutics. Over three decades, natural product-derived therapeutics encompassed more than 30% of the US Food and Drug Administered (FDA) approved drugs. Notably, more than 50% of the small molecules therapeutics originated from natural products or are natural product-inspired [27]. Plant herbals are used back to 2600 BC and introduced into healthcare, attributing to the presence of bioactive compounds in medicinal plants [28]. Furthermore, long-standing proofs of these bioactive principles were also reported in traditional folk practices such as Ayurveda and traditional Chinese medicine (TCM). As such, these medicinal plants were widely studied and proven to have a therapeutic value which is considered as an essential tool for the identification and development of novel drug leads in modern times. At present, commonly used drugs from natural products are broadly classified into two major categories, the small-molecules drugs which occupy the chemical space of <500 Da and protein biologics that are > 10 000 Da. Small-molecules therapeutics identified through high-throughput bioactivity screenings and rationale design approach, based on ligands, mechanisms, or receptors [29], displayed a myriad of pharmacological activities like anti-microbial, anti-viral and anti-cancer, which

enabled the treatment of many diseases and are endorsed by pharmaceutical companies because of their metabolic stability, ease of production and high oral bioavailability [30].

However, in the late 20th century, novel therapeutic classes that fell on the opposite end of the size spectrum began to rise which led to a paradigm shift. This shift was linked to the advancement in recombinant protein expression systems, enhanced molecular biology tools and refined protein purification methods. These factors, together with the recognition that protein therapeutics showed greater potency and higher selectivity that lead to lesser off-target effects, a major drawback of small-molecule therapeutics. As such, many pharmaceutical companies scaled back on the production of small-molecule drugs and shifted the focus to protein therapeutics. These protein drugs include growth factors, insulin, and recombinant antibodies that are termed as biologics. However, protein biologics also carry a disadvantage of having a large molecular size which is not ideal for oral administration and therefore requires alternative drug delivery methods such as intravenous, intramuscular, intranasal, or subcutaneous delivery [31].

With the progressive advancement of genome sequencing in the 21st century, it was touted as the driving force that caused the massive increase in drug development through the identification of a vast collection of new drug targets. Although genome technologies have generated a large amount of gene expression data, this did not translate into the equivalent amount of new validated drug targets [29]. It is postulated that protein-protein interactions are likely responsible for many new targets identified from sequencing that displayed shallow interaction sites

spanning across large surface areas [29]. This class of target is not very tractable in small-molecule drugs and remains inaccessible to proteins, which are membrane-impermeable but is potentially suitable for peptides.

At present, there have been over 7000 naturally occurring peptides reported. These peptides identified showcased multiple physiological roles such as hormones, anti-infectives, growth factors, ion channel ligands, and neurotransmitters [32, 33]. Peptide-derived drugs are specific and efficacious signaling molecules that can trigger a cascade of intracellular activities by binding them to cell surface receptors or ion channels. This specificity renders peptide-therapeutic drugs relatively safe, higher selectivity, and efficiency profiles compared to the small-molecule drug. Furthermore, the production of peptide-therapeutics is cheaper and less complex compared to protein biologics [31].

Over the decade, peptides have increasingly contributed to an array of applications in biotechnology and medicine. At present, there are over 60 FDA-approved peptide therapeutics that are commercially available. More than 140 peptide therapeutics are in the clinical trial phase, and nearly 500 peptide therapeutics are in preclinical development. Morever, the global peptide drug market predicted to rise from US$14.1 billion in 2011 to an estimated US$25.4 billion in 2018 [31], suggests the high commercial value of peptide therapeutics. Interestingly, metabolic diseases and oncology are the main driving fields that forwarded the usage of peptide therapeutics.

Peptides are ideal candidates for site-specific modifications that enable improvement on the potency, specificity, safety of therapeutics. Besides, with the

rapid advancement of “omics” technologies, the future development of peptide therapeutics appears promising. However, the low oral bioavailability and metabolic stability coupled with the high cost of production, remain a major concern and impedes the widespread application of peptides as drugs.

3. Cysteine-rich Peptides in plants

Plants have developed an arsenal of defense mechanisms to counter abiotic stresses such as drought, cold, heavy metals, air pollutants and pathogenic attacks from bacteria, fungi, and viruses [34]. The cell wall of the plant represents the first line of defense against infection by pathogens. Pathogens that breached the cell wall are confronted with the plant’s innate immune system. In response to the pathogenic invasion, plants up-regulate a set of genes to resist the foreign invasion. The release of secondary metabolites such as tannins, polyphenols, pathogenesis-related proteins, and enzyme inhibitors facilitate the host-defense mechanism [34-37]. They interact with the receptors on the cell surface, disrupt the cell membrane, which lead to cell death of pathogens. Current reviews showed more than 17 plant families exhibited defense-related properties such as anti-fungal, antioxidant, anti-viral and proteinase inhibitory activities, in which a fraction of it was associated with peptides found in the plants [34].

Small-molecules metabolites. Natural products have been a reliable source of inspiration in agriculture research and drug discovery. Substantial efforts to identify and characterize the myriad of active principles in medicinal plants and functional foods, responsible for their biological effects were performed. Nonetheless, these bioactive compounds studied were primarily focused on the secondary metabolite

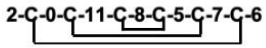
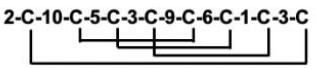
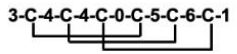
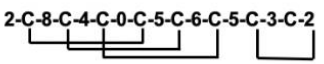
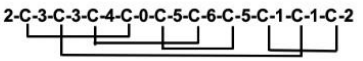
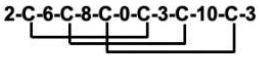
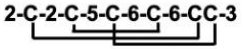
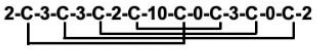
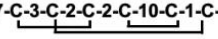
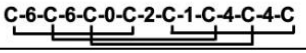
of the plants [38] These examples include, pyrethrin, an anti-feedant isolated from *Chrysanthemum cinerariaefolium* [10], nicotine, a nerve agent isolated from *Nicotiana tabacum* [11, 12] , rotenone, a mitochondria inhibitor isolated from *Lonchocarpus sp*. [13, 14], aspirin, an antipyretic agent isolated from *Salix alba* [39], quinine, an anti-malarial drug isolated from *Cichona officinalis* [40] and morphine, an alkaloid isolated from *Papaver somniferum* [41]. Pesticide manufacturing industries prefer small-molecule over protein-based biopesticides because they can be easily synthesized and have higher killing efficiency. Furthermore, pharmaceutical companies also favor the development of small-molecule drugs due to low production cost and greater oral bioavailability. Yet, the shift to macromolecular therapeutic like protein biologics suggests that it may be time for natural product research to pursue a similar path. Nonetheless, proteinaceous compounds are generally not considered in drug design due to the misconception that they are easily degradable by heat and chemical process in the gastrointestinal tract, rendering poor bioavailability [29, 42].

Cysteine-rich antimicrobial peptides (AMPs). AMPs are defense peptides widely found in different molecular forms such as linear peptides, polycyclic peptides, circular peptides and cyclotides [43, 44]. Many of these AMPs are cysteine-rich, a feature that allow the formation of highly cross-linked disulfide bridges that produce a cystine-dense structure and confer resistance to chemical and enzymatic degradation [34], debunking the misconception of proteinaceous-derived therapeutics in drug discovery and further supporting the fact that peptide/protein-based biopesticides are more stable. These cysteine-rich peptides

(CRPs) are involved in different roles in plant physiology, such as cell signaling, reproduction, and defense. However, omics studies and sequence analysis revealed that CRPs are a class of underexplored peptides [45].

Plant CRPs account for 3% of gene products. The expression of these CRPs can be constitutive or induced, but most importantly, it is tissue-specific [34]. Generally, plant CRPs are classified into different families based on the cysteine motifs, which feature a unique cysteine pattern and spacing in an evolvable scaffold. The hypervariability in their sequences produce molecular diversity and functional promiscuity within each peptide family. In my thesis, I limit my definition of a focused group of CRPs as those consisting of six to ten cysteines with a molecular range of 2–6 kDa. As such, these CRPs are structurally compact and likely to display pharmacological or therapeutic potentials. Table 1.1 shows the focused families of plant CRPs, which include thionins, defensins, hevein-like peptides, knottin-type peptides and α-hairpinin peptide. The list of CRPs isolated from plants has surpassed thousands, and it is envisioned to increase in the future.

Table 1.1: List of Plant CRP families. This table is modified from Tam *et al* [34]

	Representative member				
CRP Family	**No. of Disulfide**	**Peptide Name**	**AA No.**	**Disulfide Motif**	**Structural motif**
6C-Thionin	3	Crambin	46		Gamma (I^-) fold β1-α1-α2-β2-coil motif
8C-Thionin	4	β-Purothionin	45		
8C-Defensin	4	NaD1	47		CSαβ motif β1-coil-α-β2-β3
10C-Defensin	5	PhD1	47		
6C-HLP	3	Ac-AMP1	29		Gly- and Cys-rich Central β-strands & (short helical) coil
8C-HLP	4	Hevein	43		
10C-HLP	5	EAFP1	41		
Knottin	3	PAFP-S	38		Cystine-knot short β-strand and coil
Cyclotide	3	Kalata B1	29		
α-Hairpinin	2	Ec-AMP1	37		α1-turn-α2
Jasmintide	3	jS1	27		
β-ginkgotide	3	gB1	20		
Lybatide	4	Lyba2	33		Cystine-stabilized αT
Potentide	3	pA3	35		
Ginsentide	4	TP1	31		

Classification of CRPs. The defensin and cyclotide represent the major plant CRP families that are well-studied with thousands of articles and reviews reporting on their characterization and biological functions. A search was performed on PubMed using key words such as “cysteine-rich peptides,” “defensin,” “cyclotide” and “hevein-like peptide” to understand the extent of research done on plant major CRP families. The search revealed 5607 CRP-related articles, indicating a high number of CRPs studied. Furthermore, a total of 7755 and 460 defensin- and cyclotide-related publications, strongly suggest that these plant families are well-established (Figure 1.1). Plant defensins are cationic CRPs that are ubiquitous in plants and are expressed in various plant tissues [46-48]. The magnitude of research on plant defensins has evidently shown that they are essential in plant defense and their biological functions have been well-characterized. Table 1.2 shows the list of reported defensins as well as their biological function

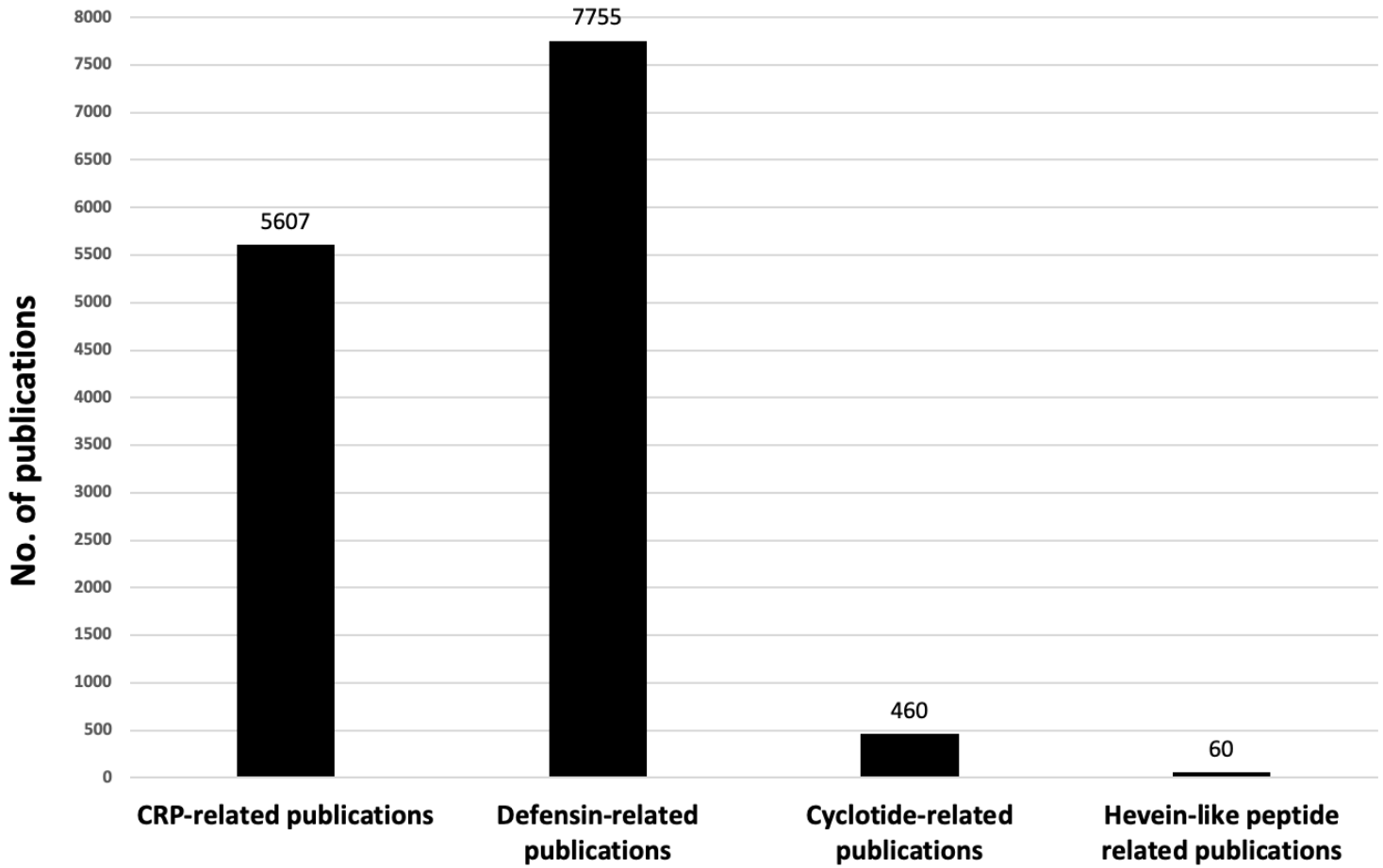


Figure 1.1: Number of publications related to CRPs and their major plant families. The search was performed on google scholar on 19th June 2020.

The cyclotides are a family of CRPs that contained a head-to-tail cyclized backbone, which is stablized by three disulfide bonds that form a cystine-knot. Such a unique circular backbone confers stability to biological, chemical and thermal degradation, which appear as promising candidates for designing of peptide biologics and therapeutic tools [49, 50]. To date, 4990 cyclotide-related articles have been reported, and their biological function has been well-characterized. Table 1.3 shows the list of cyclotides reported and their characterized function.

On the contrary, the hevein-like peptide (HLPs) family appears to be less well-established compared to defensins and cyclotides. However, in recent years, our laboratory has discovered several hevein-like peptides from different plant families [51-57]. This suggests that hevein-like peptides are an emerging group of CRPs which are highly underexplored and thus, it raises the attention to explore further and understand this family of peptides.

Table 1.2: Non-exhaustive list of defensins and their reported biological activities

Defensin Peptide	Plant Species	Biological Activity	Reference
Rs-AFP1-4	*Raphanus sativus*	Anti-fungal	Terras. *et al*. (1995)
Ah-AMP1	*Aesculus hippocatanum*	Anti-fungal	Osborn.*et al*. (1995)
AlfAFP	*Medicago sativa*	Anti-fungal	Gao. *et al*. (2000)
MsDef1	*Medicago sativa*	Anti-fungal	Abdallah. *et al*. (2010)
Wasabi defensin	*Eutrema japonicum*	Anti-fungal	Kanazaki. *et a*l (2002)
Pea defensin	*Pisum sativum*	Anti-fungal	Wang. et al. (1999)
BjD	*Brassica juncea*	Anti-fungal	Anuradha. *et al* (2008)
Pth-St1	*Solanum tuberosum*	Anti-bacterial	Moreno, M. *et al*. (1994)
Fabatin-1	*Vicia faba*	Anti-bacterial	Zhang, Y. *et al*. (1997)
Fabatin-2	*Vicia faba*	Anti-bacterial	Zhang, Y. *et al*. (1997)
SoD1-7	*Spinacia oleracea*	Anti-bacterial	Yount, N.Y. *et al*. (1995)
VrCRP	*Vigna radiata*	Insecticidal	Chen, K.C. *et al*. (2002)
γ1-H	*Hordeum vulgare*	Protein synthesis inhibitor	Colilla, F.J. *et al*. (1990)
γ1-P	*Triticum turgidium*	Protein synthesis inhibitor	Mendez, E. *et al*. (1990)
ω-H	*Hordeum vulgare*	Protein synthesis inhibitor	Mendez, E. *et al*. (1996)
HvAMP1	*Hardenbergia violacea*	Protein synthesis inhibitor	Harrison, S.J. *et al*. (1997)
SIa1-3	*Sorghum bicolor*	α-amylase inhibitor	Bloch, C. Jr. *et al*. (1991)
CfD2	*Cassia fistula*	Proteinase inhibitor	Wijaya, R. *et al*. (2000)
γ1-Z	*Zea mays*	Sodium channel inhibitor	Stiekema, W.J. *et al*. (1988)
γ2-Z	*Zea mays*	Sodium channel inhibitor	Stiekema, W.J. *et al*. (1988)
Sesquin	*Vigna sesquipedalis*	Anti-cancer	Wong J. H. *et al*. (2005)
Limenin	*Phaseolus limensis*	Anti-cancer	Wong J. H. *et al*. (2006)
Lunatusin	*Phaseolus lunatus*	Cytotoxic and anti-cancer	Wong J. H. *et al*. (2005)
Purple pole defensin	*Phaseolus vulgaris*	Anti-cancer	Lin P., *et al*.(2010)
Coccinin	*Phaseolus coccineus*	Anti-cancer	Ngai P. H. K. *et al*.(2004)
Phaseococcin	*Phaseolus coccineus*	Anti-cancer	Ngai P. H. K. *et al*.(2005)
γ-Thionin	*Capsicum chinense*	Anti-cancer	Anaya-López J. L. *et al*.(2006)
NaD1	*Nicotiana alata*	Anti-cancer	Poon I. K. H. *et al*.(2014)
Mitogenic defensin	*Phaseolus vulgaris*	Cytotoxic and anti-cancer	Wong J. H. *et al*.(2006)
Vulgarinin	*Phaseolus vulgaris*	Anti-cancer	Jack H. W. *et al*.(2005)
Cloud bean defensin	*Phaseolus vulgaris*	Anti-cancer	Wu X. *et al*.(2011)
Nepalese	*Phaseolus angularis*	Anti-cancer	Ma D. Z. *et al*.(2009)
Gymnin	*Gymnocladus chinensis Baill*	Anti-cancer	Wong J. H. *et al*.(2003)

Table 1.3: Non-exhaustive list of cyclotides and their reported biological activities

Cyclotide Peptide	Plant Species	Biological Activity	Reference
Kalata B1	*Oldenlandia affinis*	Insecticidal	Jennings. *et al*. (2001)
Kalata B1	*Oldenlandia affinis*	VEGF-A antagonist	Gunasekera. *et al* (2008)
Kalata B1	*Oldenlandia affinis*	Dengue NS2B-NS3 protease inhibitor	Gao Y. *et al* (2010)
Kalata B1	*Oldenlandia affinis*	Bradykinins B1 receptor inhibitor	Wong. *et al* (2012)
Kalata B1	*Oldenlandia affinis*	Melanocortin 4 receptor agonist	Eliasen R. *et al* (2012)
Kalata B1	*Oldenlandia affinis*	Neuropilin-1/2 antagonist	Getz JA.*et al* (2012)
Kalata B1	*Oldenlandia affinis*	Immunomodulator	
Kalata B1	*Oldenlandia affinis*	Nematocidal	Colgrave *et al*.(2008)
Kalata B2	*Oldenlandia affinis*	Insecticidal	Jennings. *et al*. (2005)
Cycloviolacin O1	*Viola odorata*	Molluscicidal	Plan *et al*. (2008)
Cycloviolacin O2	*Viola odorata*	Nematocidal	Colgrave *et al*.(2008)
Cycloviolacin O2	*Viola odorata*	Cytotoxic and anti-cancer	Wang CK. *et al* (2014)
Cliotides	*Clitoria ternatea*	Cytotoxic and anti-cancer	Nguyen G. K. T. *et al* (2011)
Viphi A	*Viola philippica*	Cytotoxic and anti-cancer	He W. *et al* (2011)
Viphi F	*Viola philippica*	Cytotoxic and anti-cancer	He W. *et al* (2011)
Viphi G	*Viola philippica*	Cytotoxic and anti-cancer	He W. *et al* (2011)
MCoTI-I	*Momordica cochinensis*	CXCR4 antagonist	Aboye T. *et al* (2015)
MCoTI-I	*Momordica cochinensis*	P53-Hdm2/HdmX antagonist	Ji Y. *et al (2013)*
MCoTI-II	*Momordica cochinensis*	FMDV 3C protease inhibitor	Thongyoo P. *et al* (2008)
MCoTI-II	*Momordica cochinensis*	β-tryptase inhibitor	Sommerhoff. *et al* (2010)
MCoTI-II	*Momordica cochinensis*	Human elastase inhibitor	Gunasekera. *et al* (2008)
MCoTI-II	*Momordica cochinensis*	CTLA-4 antagonist	Maass F. *et al* (2015)
MCoTI-II	*Momordica cochinensis*	VEGF receptor agonist	Chan LY. *et al* (2011)
MCoTI-II	*Momordica cochinensis*	BCR-ab1 kinase inhibitor	Huang YH. *et al* (2015)
EETI	*Ecballium elaterium*	Integrin-binding knottin conjugated with contrast microbubbles	Willmann JK. *et al* (2010)
Parigidin-br1	*Palicourea rigida*	Insecticidal	Pinto *et al*. (2012)
Cter M	*Clitoria ternatea*	Insecticidal	Poth. *et al*. (2011)
HB7	*Hedyotis biflora*	Anti-cancer	Ding X. *et al* (2014)

3.1 Classification of CRPs Based on the Disulfide Connectivity

CRPs has a high content of cysteine residues in their primary sequence which are responsible for forming multiple disulfide bonds. These disulfide bonds are involved in maintaining the structure and conferring metabolic stability. Besides, they facilitate the proper folding of proteins via disulfide bridges [58] The cysteine framework of CRPs encompassed two parts, the cysteine spacing and arrangement of the cysteine residues in the primary sequence, and the disulfide connectivity of the cysteine residues. Coupling these two factors could generate an extensive amount of sequence and structure diversity. Theoretically, CRPs with a molecular range between 2 to 6 kDa typically contain six to ten cysteine residues may generate 176 unique disulfide connectivity combinations. However, according to previous reports, the types of disulfide connectivity can be classified into two major classes: the type I cystine-knot disulfide connectivity and the type II thionin-like disulfide connectivity (Figure 1.2).

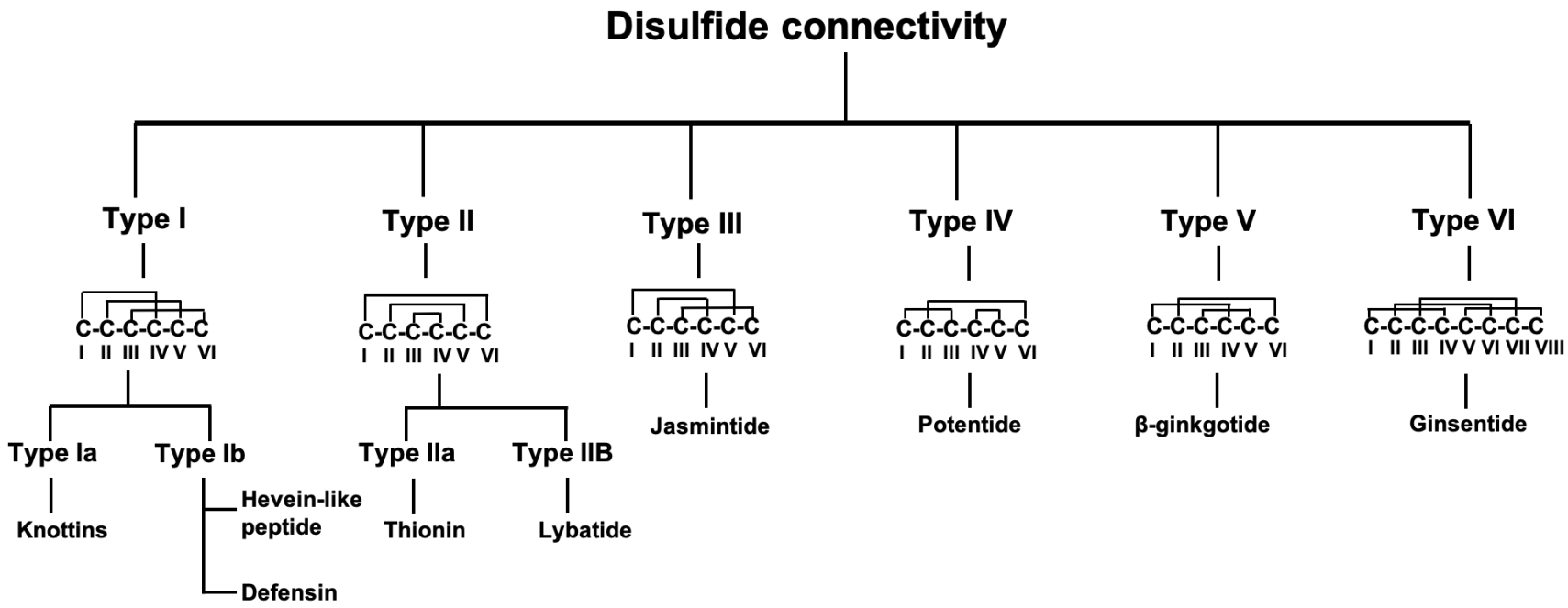


Figure 1.2: Classification of CRPs disulfide connectivity. Type I comprise of a cystine-knot and Type II comprise of a symmetric cysteine arrangement. The figure was modified from J. Huang (2019) (Doctoral dissertation, Nanyang Technological University, Singapore). Retrieved from: https://dr.ntu.edu.sg/handle/10356/137776.

The type I cystine-knot disulfide connectivity is the most common disulfide connectivity found in plant CRPs. The cystine-knot displays a CysI–CysIV, CysII–CysV and CysIII–CysVI disulfide connectivity which is normally found in knottins and 6C-hevein-like peptides (HLPs) [51, 59]. The type Ia contains six cysteine residues that are characterized by a disulfide bond that pierce through a ring formed by the other two disulfide bonds (Figure 1.3). In addition, the formation of additional disulfide bonds does not perturb the knot-topology and can expand the library of type I disulfide connectivity. This disulfide connectivity is termed as type Ib, and is typically observed in defensins and 8C-HLPs [52, 60]

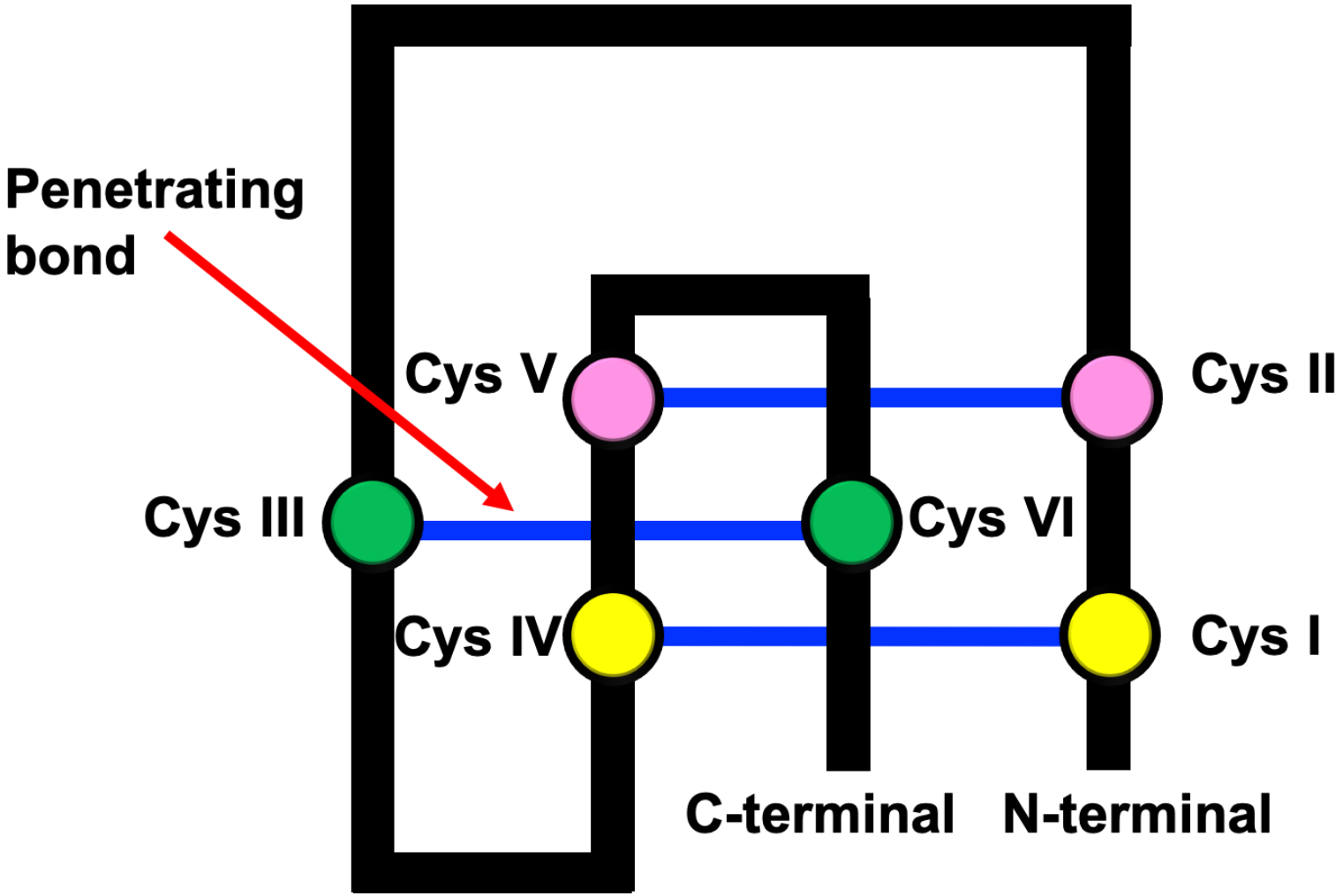


Figure 1.3: Schematic diagram of cystine-knot connectivity. The cysteine pairs are represented by circles with Cys I connected to Cys IV, Cys II to Cys V and Cys III to Cys VI. The blue line indicates the formation of a disulfide bond.

On the contrary, the type II thionin-like disulfide connectivity has a symmetric arrangement and is display as CysI–CysVI, CysII–CysV and CysIII–CysIV. This type of disulfide connectivity can be found in plant CRPs such as thionins [61] and α-hairpinins [62], and are termed as type IIa. Recently, our laboratory discovered a novel disulfide 8C-CRP connectivity pattern from lybatides. The disulfide pattern of lybatides is a variant from the type IIa that adopts a symmetric disulfide connectivity of CysII–CysVIII, CysIII–CysVII and CysIV–CysV with an additional disulfide bond CysI–CysVI [63]. Therefore, this new disulfide connectivity was termed as type IIb.

Our laboratory has also discovered four novel disulfide patterns from 6C- and 8C-CRPs, expanding the library of existing disulfide patterns. These include jasmintides isolated from *Jasminum sambac* [64], potentides from *Potentilla anserine* [65], β-ginkgotide from *Ginkgo biloba* [66] and ginsentides from the Ginseng family [57]. Jasmintides has a unique disulfide connectivity of CysI–CysV, CysII–CysIV and CysIII–CysVI and hence was termed as type III. Potentides contain a unique disulfide pattern of CysI–CysIII, CysII–CysVI and CysIV–CysV and was termed as type IV. The disulfide connectivity of β-ginkgotide is CysI–CysIV, CysII–CysVI and CysIII–CysV and was termed as type V and lastly, ginsentides have a disulfide connectivity of CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII was termed as type VI.

The cysteine spacing in CRPs is another key factor for CRP classification, which determines the number and the length of the intercysteinyl loops. These loops are formed between cysteine residues and represent the peptide backbone. The CC motif comprises of two adjacent cysteines and is well-represented in CRP families.

For example, 6C-CRPs typically have a cysteine spacing of C-C-CC-C-C in which the CC-motif is located in the center and this cysteine spacing can be observed in 6C-HLPs [51] and cystine-knot α-amylase inhibitors (CKAIs) [24]. In contrast, thionins have a cysteine spacing of CC-C-C-C-C with the CC- motif presented at the N-terminus [67]. In trans-defensins, the CC motif is located at the C-terminus as C-C-C-C-CC [68]. In some cases, the CC motif may be presented twice. This was observed in β-ginkgotide with a cysteine spacing of C-CC-C-CC. This unique cysteine spacing endows β-ginkgotide with three intercysteinyl loops. Interestingly, some 6C-CRPs have a cysteine spacing of C-C-C-C-C-C that do not contain the CC- motif. This cysteine spacing is often observed in carboxypeptidase inhibitors [69] and cyclotides [70-72].

Collectively, the different permutations of disulfide connectivity and cysteine spacing in CRPs can generate a myriad of different structural folds with 1065 possible permutations. At present, six major cysteine frameworks have been identified (Figure 1.2) in comparison to conotoxins which have a total of 27 different cysteine frameworks [73]. Therefore, this suggests that the number of cysteine frameworks found in plant CRPs is still inferior and underexplored (Table 1.4).

Table 1.4: Cysteine frameworks of conotoxins and CRPs equivalent

Conotoxin cysteine pattern	No. of Cys	Connectivity	CRPs with similar framework	Conotoxin reference	CRP reference
CC-C-C	4	I-III, II-IV		Gray,W.R. et al. (1981)	
CCC-C-C-C	6			Ramilo,C. et al. (1992)	
CC-C-C-CC	6			Sato,S. et al. (1983)	
CC-C-C-C-C	6	I-V, II-III, IV-VI	6C-Thionins	Fainzilber,M. et al. (1995)	Stec, B. (2006)
CC-CC	4	I-III, II-IV		Walker,C.S. et al. (1999)	
C-C-CC-C-C	6	I-IV, II-V, III-VI	6C-HLPs, CKAIs	Olivera,B.M. et al. (1984)	Tam. et al (2015)
C-C-C-C-C-C-C-C-C-C	10			England,L.J. et al. (1998)	
C-C-C-C-C-C	6	I-IV, II-V, III-VI	Cyclotides	Lirazan,M.B. et al. (2000)	Craik, D. J., et al (2010)
CC-C.[PO]C	4	I-IV, II-III		Balaji,R.A. et al. (2000)	
C-C-CC-CC-C-C	8	I-IV, II-VI, III-VII, V-VIII		Jimenez,E.C. et al. (2003)	
C-C-C-C-CC-C-C	8			Brown,M.A. et al. (2005)	
C-C-C-CC-C-C-C	8			Aguilar,M.B. et al. (2005)	
C-C-C-C	4	I-III, II-IV	α-Hairpinins, 8C-HLPs	Moller,C. et al. (2005)	Duvick et al. (1992)
C-C-CC-C-C-C-C	8			Peng,C. et al. (2008)	Tam. et al (2015)
C-C-CC	4			Pi,C. et al. (2006)	
C-C-CC-C-CC-C	8			Yuan,D.D. et al. (2008)	
C-C-CC-CC	6			Chen,J.S. et al. (1999)	
C-C-C-CCC-C-C-C-C	10			Chen,P. et al. (2008)	
C-CC-C-CC-C-C-C-C	10			Loughnan,M.L. et al. (2009)	
CC-C-C-C-CC-C-C-C	10			Möller,C. and Marí,F. (2011)	
C-C-C-C-C-C-C-C-C	8		8C-Defensins	Elliger,C.A. et al. (2011)	Lay, F. T., et al (2015)
C-C-C-CC-C	6			Ye,M. et al. (2012)	
C-CC-C	4			Luo,S. et al. (2013)	
C-C-C-C-CC	6		Jasmintides*	Aguilar,M.B. et al. (2013)	Kumari, G., et al (2018)
C-C-C-C-CC-CC	8		Lybatides*	Bernáldez,J. et al. (2013)	
C-CC-C-C-C	6		β-ginkgotide*	Kancherla,A.K. et al. (2015)	Wong, K. H., et al (2017)
CC-C-C-C-C-C-C#	8	I-VIII, II-VII, III-VI, IV-V	8C-Thionins		Mak AS and Jones BL (1976)
C-C-C-C-CC-C-C-C-C#	10	I-X, II-V, III-VII, IV-VIII, V-IX	10C-Defensins		Lay FT et al. (2003)
C-C-C-CC-C-C-C-C-C#	10	I-V, II-IX, III-VI, IV-VII, VIII-IX	EAFP, WAMP		Andreev Y A et al. (2012)
C-C-CC-C-CC-C-C-C#	10	I-IV, II-V, III-VI, VII-X, VIII-IX	Ee-CBP		Van den Bergh KP et al. (2004)

#represents framework not identified in conotoxin

*represents CRPs identified from our laboratory

4. Hevein-like Peptides

Hevein-like peptides (HLPs) are a family of CRPs that consist of 29 to 45 amino acid residues. In our discovery program, the cysteine spacing C-C-CC-C-C characterized by a CC motif located at the center of the cysteine spacing was commonly encountered. Functionally, this family of peptide can be classified into two subfamilies: the chitin-binding HLPs and the non-chitin-binding HLPs. These subfamilies can be further classified depending on the number of cysteines present and the peptide's bioprocessing. Figure 1.4 summarizes the overview of the classification of HLPs [34]. This family of HLPs is the focus of this thesis.

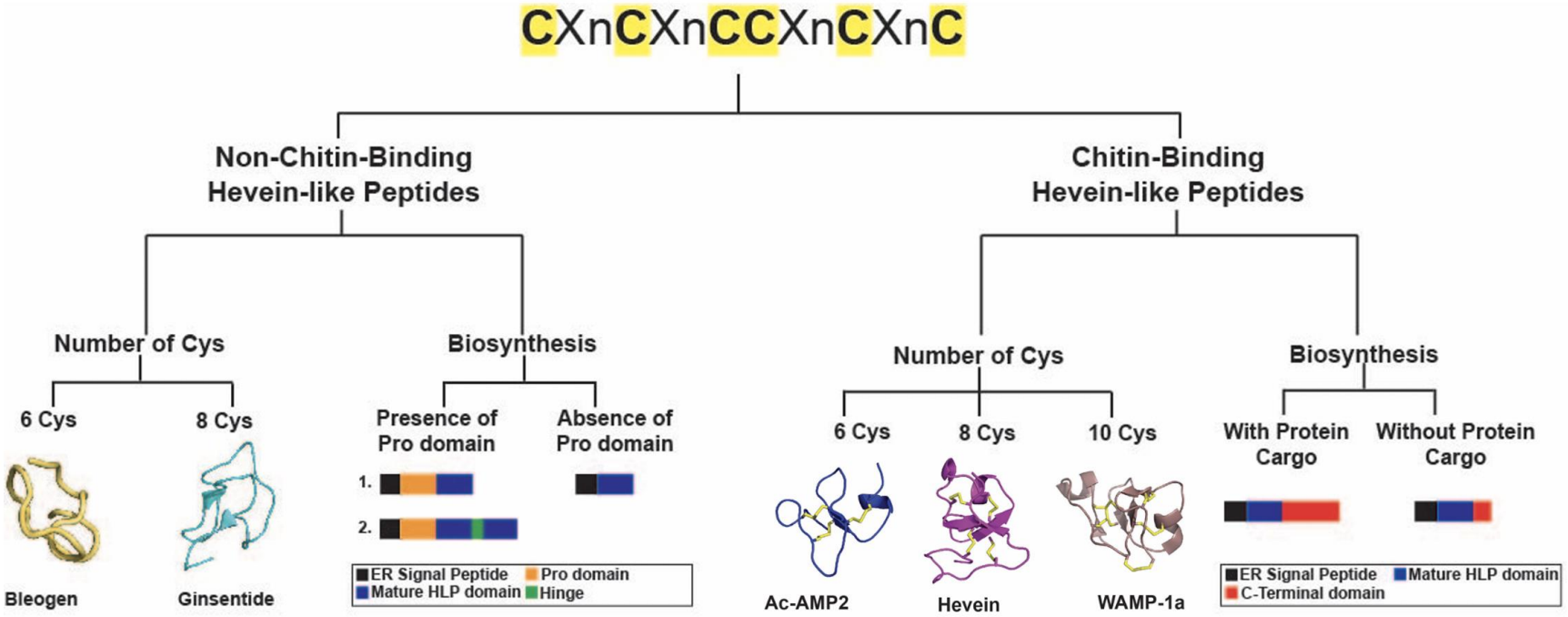


Figure 1.4: Classification of hevein-like peptides. The hevein-like peptides consist of a C-C-CC-C-C motif that is classified into chitin-binding and non-chitin-binding hevein-like peptides. Further classifications are based on the number of cysteines and biosynthesis. Examples of non-chitin-binding hevein-like peptides include bleogen pB1 (PDB:5XBD) and ginsentide TP1 (2ML7). Examples of chitin-binding hevein-like peptides include Ac-AMP2 (PDB: 1MMC), hevein (PDB:1HEV) and WAMP-1a (PDB: 2LB7).

4.1 Chitin-binding Hevein-like Peptides

4.1.1 Discovery of Hevein

Hevein was discovered from *Hevea brasiliensis* (rubber tree) and it is the most abundant constitutent from latex (Figure 1.5). In 2019, according to the Association of Natural Rubber Producing Countries (ANRPC), the rubber tree contributed to the global production of 13.804 million tons of natural rubber. Natural rubber is a polymer of isoprene obtained from the latex of *H. brasiliensis* that comprises approximately 40% rubber particles, proteins, sterol glycosides, organic solutes such as starch and alkaloids, and inorganic substances like silica and iron oxide [74]. A study has shown that the latex composition contained three different fractions: a white rubber particle top layer, a yellow serum middle layer and a lutoid basal layer [75]. Lutoids are vacuoles that constitute approximately 20% of the latex volume and consist of proteins that are involved in pathogenic defense, chitin catabolism and ion transport [76-78]. To investigate the protein components found in the basal layer of latex, Archer *et al*. performed an electrophoretic analysis. They demonstrated that the basal layer comprised of seven electrophoretically distinct protein constitutes [79, 80].

Nearly after a decade, Archer *et al.* successfully identifed a 10 000 Da protein from the basal layer of the latex which was rich in cysteine, aspartic acid, serine and tryptophan amino acid residues [81]. The group named the protein “hevein”. Fifteen years after discovering the protein hevein, the full primary sequence of hevein was successfully deduced by Walujuno *et al* in 1975. Walujuno and his team

demonstrated that hevein is a dimer that consists 43 amino acid residues and is cysteine and glycine-rich [82].

Figure 1.5: *Hevea brasiliensis* tree. The rubber tree is the primary source of the world's natural rubber with a global production of 13.9 million tons in 2018. The picture was adapted from https://www.nparks.gov.sg/gardens-parks-and-nature/heritage-trees/ht-2018-294.

Interestingly, the sequence of hevein bears homology to a family of chitin-binding proteins such as chitinases from rice [83], stinging nettle lectin [84] and chitin-binding lectins from wheat [85]. Notably, the primary sequence of hevein was similar to Urtica dioca agglutinin (UDA), a protein that binds to chitin. This protein exerts anti-fungal effect on a plethora of chitin-containing fungal strains [86]. Therefore, accounting on the sequence homology, the anti-fungal properties of hevein was examined against chitin-containing fungal strains, which revealed that fungal strains treated with hevein had growth retardation [87]. The fungi inhibitory properties of hevein was associated with its chitin-binding domain, which differentiates it from other CRPs such as knottins [34], CKAI [24], thionins [88] and cyclotides [71].

In 1990, the bioprocessing of hevein was elucidated from the cDNA clone (HEV1) of hevein [89]. The primary precursor sequence of hevein encodes for 204 amino acids. The biosynthetic arrangement of hevein comprises of a three-domain architecture that includes a signal peptide followed by a mature hevein domain and an extended C-terminal. Shortly after, in 1991, the structure of hevein was determined by X-ray crystallography with a resolution of 2.8 Å [90]. However, advances in X-ray crystallography techniques has improved the resolution of hevein to 1.5 Å [91]. In addition, NMR techniques were also employed to solve the structure of hevein [92] The structural studies revealed a knotted-topology with an additinal disulfide bond at the C-terminal. The compact structure coupled with the chitin-binding properties of hevein prompted Asensio *et al.* to investigate the interaction between hevein and carbohydrates [93]. The NMR titration study of the

interactions of hevein with carbohydrates such as GluNAc, chitobiose and chitotriose, yielded binding affinities at millimolar range. Furthermore, they reported that the amino acid residues Ser19, Trp21, Trp23 and Tyr30 were the key residues responsible for the binding interaction.

Following the discovery of hevein, it remained as the only peptide identified from this CRP family for 32 years. The advancement in technologies led to a significant increase in the discovery of peptides such as AC-AMP1, AC-AMP2 from *Amaranthus caudatus* and IWF-4 from *Beta vulgaris* [94, 95]. These peptides were homologous to hevein and chitin-binding proteins. Interestingly, they contained six cysteine residues with 29 to 30 amino acid residues. Since they contained only six cysteine residues, they were speculated as the truncated variants of hevein. Functionally, these peptides were able to bind to chitin and exhibit stronger potency against phytopathogenic fungal strains compared to hevein and the chitin-binding proteins.

In 1998, following the discovery of two eight cysteine hevein variants named as Pn-AMP 1 and Pn-AMP2 isolated from *Pharbitis nil*, the term “hevein-like peptide” was coined [96]. Subsequently, variants that contained ten cysteine residues were discovered from *Eucommia ulmoides* and these peptides were named as EAFP1 and EAFP2 [97]. Since then, the library of chitin-binding HLPs have significantly expanded. A review published by our laboratory classified these chitin-binding HLPs into three subfamilies: 6C-, 8C- and 10C-chitin-binding HLPs based on the number of cysteines [34]. At present, a total of 12 6C-chitin-binding HLPs, 21 8C-chitin-binding HLPs and five chitin-binding HLPs have been reported. Figure 1.6

shows the major milestones of the hevein-like peptide family. Interestingly, in 2016 and 2017, our laboratory had a major discovery on a group of peptides that adopt the same cysteine spacing and disulfide connectivity as hevein, named as roseltides [55] and bleogens [56]. This group of peptides do not have a chitin-binding domain. In this thesis, they are termed as non-chitin-binding HLPs which will be further discussed in the non-chitin-binding HLP sections.

Overall, the diversity and occurrence of chitin-binding HLPs suggests that they are possibly one of the bioactive components that are inherently expressed in plants.

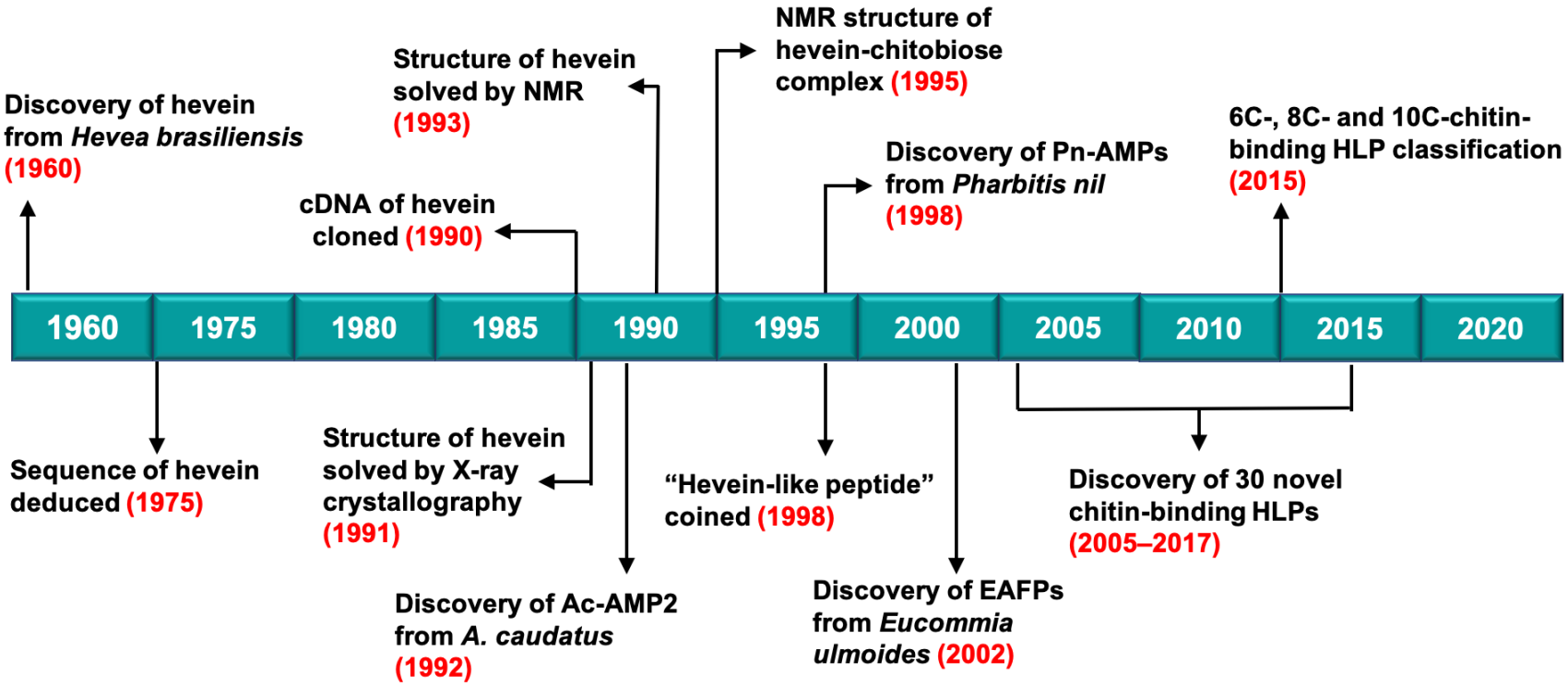


Figure 1.6: Major milestones of chitin-binding hevein-like peptides. Hevein was first discovered in 1960. The sequence of hevein was deduced in 1975. The cDNA of hevein was clone in 1990 and subsequently, the molecular structure of hevein was solved by X-ray crystallography and NMR in 1991 and 1992, respectively. Hevein remained as the only hevein-like peptide from this family up to 1992 when Ac-AMP2 was discovered. Ac-AMP2 contain six cysteines and is considered the truncated version of hevein. In 1998 the term "hevein-like peptide" was coined. Subsequently, more hevein-like peptides were discovered such as Pn-AMPs from *Pharbitis nil* and EAFPs from *Eucommia ulmoides*. In a span of 12 years, a total of 30 novel hevein-like peptides were discovered and hence classification of this family was initiated.

4.1.2 Classification of Chitin-binding Hevein-like Peptides

Chitin-binding HLPs are subclassified based on two major characteristics: the content of cysteine and their biosynthesis pathway (Figure 1.7). Chitin-binding HLPs contain cysteine residues, which are generally in even numbers and can be further divided into 6C-, 8C-, and 10C-chitin-binding HLPs. For example, the 6C-chitin-binding HLP subfamily includes peptides such as Ac-AMPs from *Amanranthus caudatus* [94], Ar-AMP from *Amaranthus retroflexus* [98], IWF-4 from *Beta vulgaris* [95], altides from *Alternanthera sessilis* [51] and SmAMPs from *Stellaria media* [99]. Additionally, the 8C-chitin-binding HLP subfamily is well represented by peptides such as hevein from *Hevea brasiliensis* [81], vaccatides from *Vaccaria hispanica* [53], ginkgotides from *Ginkgo biloba* [54], morintides from *Moringa oleifera* [52], Fa-AMPs from *Fagopyrum esculentum* [100] and Pn-AMPs from *Pharbitis nil* [96]. The 10C-chitin-binding HLPs subfamily has the highest cysteine content which include EAFPs from *Eucommia ulmoides* [97] Ee-CBPs from *Euonymus europaeus* [101], and WAMPs from *Triticum kiharae* [85]. Likewise, the biosynthesis of chitin-binding HLPs is further divided into chitin-binding HLPs with or without protein cargo. The protein cargo of chitin-binding HLPs is generally encased in the C-terminal tail. The protein cargo provides additional function to the chitin-binding HLPs. For example, the extended C-terminal tail in hevein encodes a Barley wound-inducing protein (Barwin) containing 144 amino acid residues. This protein assists hevein to bind to fungal cell wall and play an essential role in plant pathogenic defense [102].

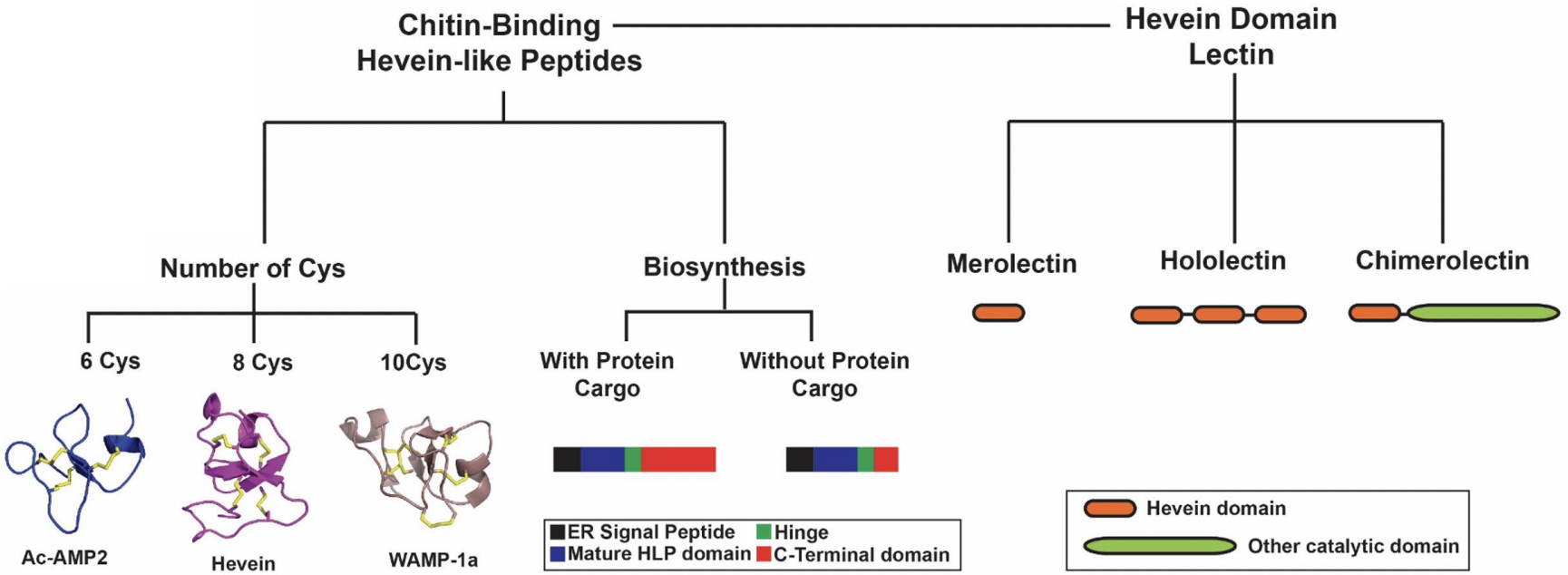


Figure 1.7: Classification of chitin-binding hevein-like peptides and proteins. The classification of chitin-binding hevein-like peptides are based on the number of cysteine present and biosynthesis. Chitin-binding hevein-like peptides are associated with a family of carbohydrate-binding protein called hevein domain lectins. In general, hevein domain lectins are classified into merolectin, hololectin and chimerolectin. PDB of Ac-AMP2, Hevein and WAMP-1a are, 1MMC, 1HEV and 2LB7, respectively.

Interestingly, the chitin-binding HLPs are associated with a family of carbohydrate-binding proteins named lectins. Plant lectins have a strong affinity towards carbohydrates, and they bind reversibily to specific carbohydrates [103]. According to the carbohydrate recognition domain (CRD), lectins can be classified into 12 families: *Agaricus bisporus* homolog, amaranthin domain, homolog of class V chitinases, cyanovirin domain, *Euonymus europaeus* lectin domain, *Galanthus nivalis* agglutinin domain, hevein domain, jacalin domain, legume domain, lys M domain, nictaba-like domain and ricin-B domain [104]. Specifically, the hevein domain lectins may consist of more than one hevein domains and can be subdivided into merolectin [87], hololectin [105] and chimerolectins [106]. In general, merolectins contain a single hevein domain that do not possess agglutination and glycoconjugate precipitation properties [103]. In contrast, hololectins contain more than one hevein domains that can be identical or homologous. This family of lectins have multiple hevein binding sites that enable them to agglutinate cells and precipitate glycoconjugates [84]. Lastly, chimerolectins are fusion proteins that contain one or more hevein domains arrayed in tandem with a catalytic domain such as chitinase domain, providing them with additional catalytic function [107, 108]. Based on this classification, the chitin-binding HLPs are considered as merolectins as they possess a single hevein domain.

Collectively, this classification provided a clearer understanding on the diversity of chitin-binding HLPs and it is expected to be more complex in the future as more novel chitin-binding HLPs are continually discovered.

4.1.3 Sequence diversity of Chitin-binding Hevein-like Peptides

Chitin-binding HLPs generally contain 29–50 amino acid residues that are rich in cysteine and glycine. Additionally, these cysteine and glycine residues are highly conserved in their primary sequences. At present 12 6C-chitin-binding HLPs have been successfully identified and characterized. These include: two Ac-AMPs from *A. caudatus* [94], one Ar-AMP from *A. retroflexus* [98], one IWF-4 from *B. vulgaris* [95], six altides from *A. sessilis* [51] and two SmAMPs from *S.media* [99]. These peptides contain six cysteines are 29–35 amino acids in length. The cysteine arrangement of 6C-chitin-binding HLP is displayed as $X_{3-6}CX_6C\text{-}X_4CC\text{-}X_5C\text{-}X_6C\text{-}X_{1-3}$, in which X represents other amino acids and the subscript stands for the number of amino acid residues. Amino acids between the cysteines are defined as "inter-cysteine loop".

Currently, within the 8C-chitin-binding HLP subfamily, 21 peptides have been isolated and characterized. These peptides include hevein from *H. brasiliensis* [81], two vaccatides from *V. hispanica* [53], 11 ginkgotides from *G. biloba* [54], two morintides from *M. oleifera* [52], two Fa-AMPs from *F. esculentum* [100]] two Pn-AMPs from *P. nil* [96] and Avesin A from *A. sativa* [109]. These peptides contain 37–45 amino acid residues and display a cysteine arrangement of $X_{2-6}C\text{-}X_{1-9}C\text{-}X_{4-5}CC\text{-}X_5C\text{-}X_6C\text{-}X_{1-5}C\text{-}X_{2-6}C\text{-}X_{2-10}$. Interestingly, in the 10C-chitin-binding HLPs, three different types of cysteine spacing patterns were generated from the shuffling of the cysteine positions. These cysteine spacing patterns were termed as type I, type II and type III. Ee-CBP isolated from *E. europaeus* displayed the type I cysteine spacing pattern as $X_2C\text{-}X_3C\text{-}X_3C\text{-}X_4CC\text{-}X_5C\text{-}X_6C\text{-}X_4C\text{-}XC\text{-}XC\text{-}X_{2.}$ [110]. Type II was

observed in WAMPs isolated from *T. kiharae,* revealed a cysteine spacing of X_2C-X_8C-X_4CCX_5C-X_6CC-X_4C-X_3C-X_2C-X [85]. In contrast, type III was observed in EAFPs isolated from *E. ulmoides* that showed a cysteine spacing of $X_{2-3}C$-$X_{4-8}C$-X_2C-XCC-X_5C-X_6C-$X_{3-4}CX_3C$-$X_{2-3}C$-X_{2-5} [97]. The flexibility of the cysteine residues may result in functional diversity within the 10C-chitin-binding HLP subfamily. Such variation was observed in the family of defensins in which the substitution of alanine in the non-cysteine residues of defensin VrD1 led to a modification in the enzyme inhibitory function [111].

The cysteine spacing of 6C-, 8C-. and 10C-chitin-binding HLPs are seperated into four, six and eight inter-cysteine loops, respectively (Figure 1.8). A prominent feature that is unique to the chitin-binding HLP family is the chitin-binding domain. This highly conserved domain is generally located between loop 3 and loop 4 [53]. The chitin-binding domain consist of a specific motif, $SX\varphi G\varphi CGX_4\varphi$, where X represents any amino acid residue and φ represents aromatic amino acid residue. This motif is highly conserved and is responsible for binding interaction with chitin [112].

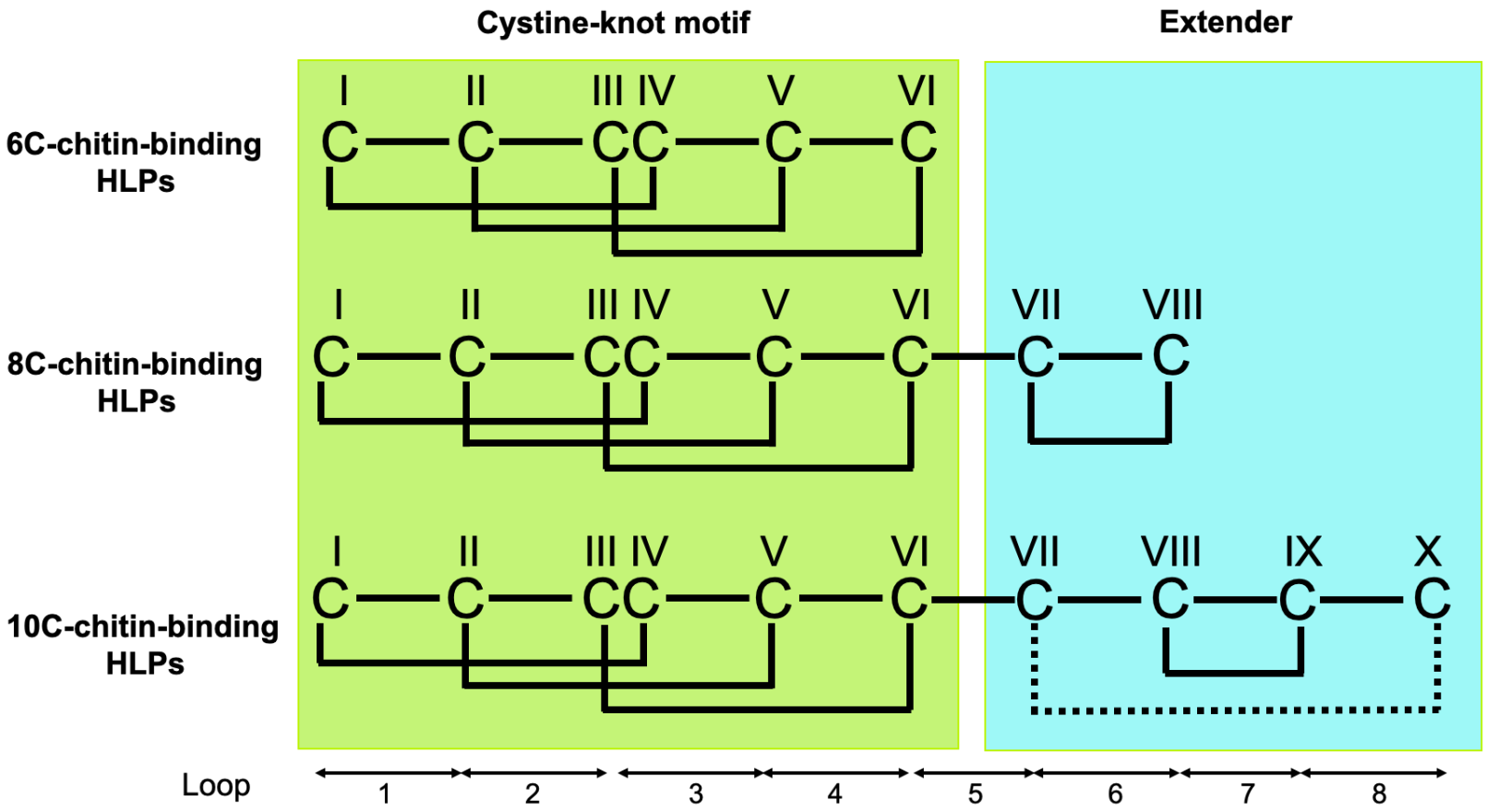


Figure 1.8: Schematic diagram of hevein-like peptides (HLPs) subfamilies. The backbone segments between cysteine residues are divided into loops and labeled as loop 1–8. 6C-, 8C-, and 10C-chitin-binding HLPs contain four, six and eight loops, respectively.

Likewise, lectin proteins have multiple units of chitin-binding domains, which was observed in *Urtica dioica* agglutinin (UDA) and wheat germ agglutinin (WGA) [113, 114]. Similarly, plant chitinases also have a chitin-binding domain in tandem to a catalytic domain. For example, chitinase A has a chitin-binding domain fused to class IIIb chitinase domain at the C-terminal [115]. Hence, it was postulated that the chitinases bind to the cell wall of chitin-containing fungal and induce anti-fungal effect.

Although chitin-binding HLPs contain one chitin-binding domain, it is not fused to a catalytic domain. Yet, they displayed fungal inhibitory effects that were more potent than chitin-binding proteins. Moreover, the chitin-binding domain is also associated to latex fruit syndrome [116]. Individuals who are sensitive to natural rubber may develop an allergic response to fruits such as papaya, carrot, tomato and avocados. It was speculated that the chitin-binding domain is responsible for the cause. Studies have reported that the IgE antibodies can target the chitin-binding domain of class I chitinases in fruits because of the high homology to hevein, resulting in the cross-reactivity of allergen [116, 117]. Therefore, structural studies were performed on the chitin-binding domain to understand the epitope involved in the allergy [118-120].

4.1.4 Biosynthesis of Chitin-binding Hevein-like Peptides

The chitin-binding HLPs are encoded by genes and synthesized by the ribosome and processed in the endoplasmic reticulum (ER) as linear precursors. These precursors are further cleaved by endopeptidases to release the mature chitin-binding HLPs. The advancement of omics technologies has greatly refined the

discovery of novel chitin-binding HLPs and provided a better understanding of the evolution, distribution, occurrence, and precursor organization of these peptides. Overall, the precursor organization of chitin-binding HLPs is conserved and constitutes a three-domain arrangement, namely a signal peptide domain, a mature domain and a C-terminal tail. The presence of an signal peptide is evidence the bioprocessing of chitin-binding HLPs require the secretory pathway [22, 121], where the synthesis and transportation of soluble protein and peptides occur (Figure 1.9). The signal peptide targets the precursor peptide to the ER and is cleaved by signal peptidase I (SPaseI). The cleavage of the C-terminal tail is performed by C-terminal processing peptidase, and the mature chitin-binding HLP is released. Notably, the precursor organization of these peptides differs from CRPs such as CKAIs and carboxypeptidase inhibitors, which comprise a pro-peptide domain prior to the mature peptide domain [59, 122-125].

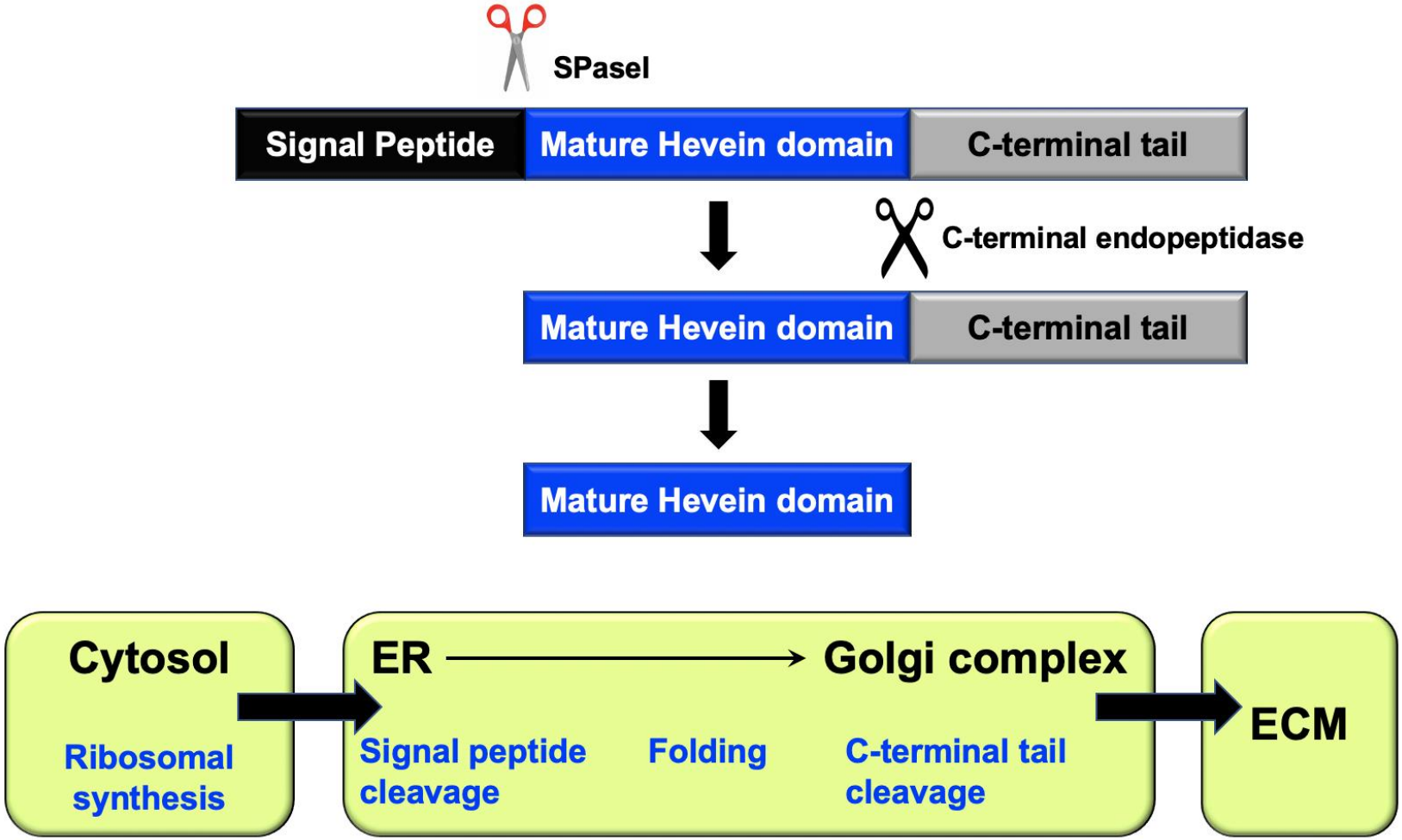


Figure 1.9: Biosynthesis of chitin-binding hevein-like peptides. The precursor organization of 6C-, 8C- and 10C-chitin-binding hevein-like peptides comprise of a signal peptide, a mature domain and a C-terminal tail. The biosynthesis pathway involves the endoplasmic reticulum, Golgi complex and the extra-cellular matrix.

The C-terminal tail of chitin-binding HLPs is another striking feature. Aforementioned, the C-terminal tail of hevein constitute 144 amino acid residues. This long C-terminal tail bears 74% to 79% homology to the Barwin gene encoded in potato [89]. However, a study on the C-terminal tail of Ac-AMP2 precursor reported that it is not homologous to any of the previously characterized carboxyl-terminal domains of chitin-binding proteins and identified protein in the Swiss-Prot database [126]. Interestingly, the C-terminal tail of AC-AMP2 displays similar characteristics with the C-terminal pro-peptide of tobacca glucanase and Graminea lectins [127, 128]. These characteristics include: (1) the absence of a C-terminal tail from the mature peptide (2) an N-glycosylation site (3) the presence of a valine residue eight amino acids before the glycosylation site and (4) is alanine- and valine-rich [126]. The C-terminal propeptide of Gramineae lectins is well known for sorting determinant that is targeted to the vacuoles [126]. Therefore, it was hypothesized that Ac-AMP2 may have the equivalent function. However, a study in transgenic tobacco revealed the localization of Ac-AMP2 in the intercellular fluid fraction of the plant refuted the hypothesis [129]. Likewise, the C-terminal tails of most reported chitin-binding HLPs were not homologous to any conserved domains [51, 52, 96, 98, 100]. Nevertheless, efforts were made to clone the cDNA of Ee-CBP isolated from *E. europaeus*. This 10C-chitin-binding HLP expressed a class I chitinase-like precursor that is cleaved during post-translation modification to produce the mature Ee-CBP peptide [101, 130].

On the whole, the versatility of chitin-binding HLP precursors proposes that they are a diverse family of CRPs and warrants investigation into their biosynthesis and precursor arrangement to understand their evolutionary relationship.

4.1.5 Molecular Structure of Chitin-binding Hevein-like Peptides

The structure of hevein provided a basis for analyzing the chitin-binding HLPs. In general, the secondary motif of chitin-binding HLPs comprised of a coil β1- β2- coil-β3 with modifications based on the β3 strand and the number of short turns [34]. This configuration is unique from other CRPs such as thionins or defensins, which have a secondary motif of γ-fold and CSαβ, respectively [131, 132]. The anti-parallel β-strands of chitin-binding HLPs form the core β-sheet along with a coil region at the N-terminal and a short helical turn at the C-terminal that is stabilized by disulfide bridges. The cysteine framework of chitin-binding HLPs followed the type I disulfide connectivity, which forms a compact knotted topology [133]. The N-terminal coil region is connected to the anti-parallel β-strands by two disulfides (CysI–CysIV and CysII–CysV) to form a ring, allowing the third disulfide bond (CysIII–CysVI) to pass through and connect the C-terminal to the first β-strand. This knotted topology is extensively distributed in different phyla such as fungi, plants and animals and is pervasive in toxins of scorpions and cone snails [134].

The 6C-chitin-binding HLPs has the most compact structure capable of inducing anti-fungal effects on a wide range of phytopathogenic fungi. The structural framework of these peptides consists of an N-terminal coil region, two anti-parallel β-strands and a C-terminal helical turn (Figure 1.10A). This type of structural scaffold exposes the residues responsible for chitin-binding on the surface [135].

In contrast, the 8C-chitin-binding HLPs comprise of three β-strand, attributing to the additional disulfide bond at the C-terminal. (Figure 1.10B).

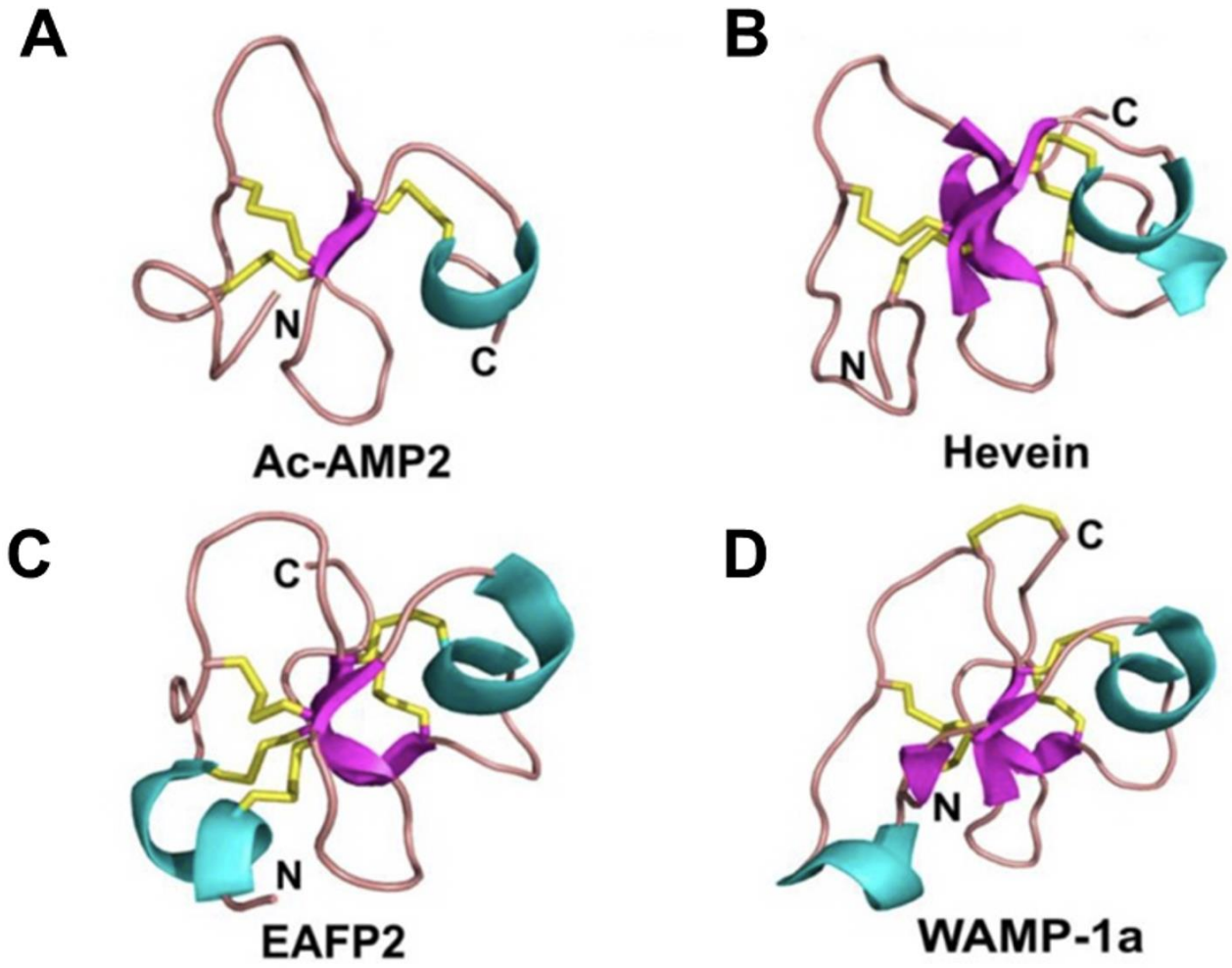


Figure 1.10: NMR structures of 6C-, 8C- and 10C chitin-binding hevein-like-peptides. (A) Ac-AMP2, **(B)** hevein, **(C)** EAFP2 and **(D)** WAMP-1a. The secondary structure is illustrated by different colors. Cyan represents the α-helix, purple represents the β-strands, pink represents the loops and yellow represents the disulfide bonds. This figure is adapted from Tam *et al* [34].

The 10C-chitin-binding HLPs contain five disulfide linkages. However, the position of the fifth disulfide bond is promiscuous. The fifth disulfide bond may be located entirely at the C-terminus or conjoined between the C-terminus and one of the intercysteine loops formed by the knotted-topology. For example, Ee-CBP isolated from *E. europaeus* has the fifth disulfide bond in the C-terminus [110]. In contrast, the fifth disulfide bond in EAFP2 isolated from *E. ulmoides*, bridges the N-terminal coil region to the third β-strand. Structural studies have shown that EAFP2 comprises of a 3_{10} helix, an α-helix and three β-strand [136, 137]. The structure of EFAP2 constitutes a chitin-binding domain and hydrophobic patches. The first four disulfide bonds of EAFP2 resemble hevein, and the last disulfide bond interlink the N-terminal Cys7 with the C-terminal Cys37 to provide a rigid structure (Figure 1.10C) [136]. EAFP2 is an amphipathic peptide that contains hydrophobic and cationic patches. The chitin-binding domain of EAFP2 involves a serine and three tyrosine residues. Structural analysis showed the exposure of aromatic residues to solvent led to a hydrophobic pocket formation that surrounds the serine residue [137].

Additionally, WAMP-1a isolated from *Triticum kiharae* has a fifth disulfide linkage that joins the C-terminus to the loop, resembling the cysteine framework of chitinases (Figure 1.10D) [138]. Considering the variations of the fifth disulfide bonds, the 10C-chitin-binding HLPs appear to be the most structurally diverse chitin-binding HLP subfamily.

4.1.6 Binding Interaction of Hevein-Chito-Oligosaccharide complex

The availability and highly compact structure of hevein and chitin-binding HLPs obtained through chemical synthesis [139], molecular cloning [120], isolation and purification [81] makes them suitable model for chitin-binding interaction study. A collection of structural and thermodynamic studies were conducted using techniques such as NMR spectroscopy [140, 141] X-ray crystallography [91] and isothermal titration calorimetry (ITC) [140, 142] to understand the interaction of chitin oligosaccharides between hevein [141], pseudohevein [143], HEV32 (truncated six cysteine hevein mutant) [144], Hev32S19D (hevein-32 mutant with a Ser19 and Asp19 mutation) [145], Ac-AMP2 [146] and wheat germ agglutinin (WGA) [147]. These studies agreed that the aromatic residues in the chitin-binding domain stabilize the chitin-binding interaction via Van der Waals contacts and CH-π stacking. A sequence logo was generated based on the alignment of the chitin-binding domains of 6C-, 8C- and 10C-chitin-binding HLPs to show the sequence conservation in the chitin-binding domain (Figure 1.11). The amino acids involved in the chitin-binding interaction were assigned from I to VI. The serine residue at the position I is highly conserved throughout the chitin-binding HLPs. In position II, an aromatic residue is favored with a higher frequency of phenylalanine in 6C-chitin-bindng HLPs, tryptophan in 8C-chitin-binding HLPs and tyrosine in 10C-chitin-binding HLPs. At position IV, aromatic amino acids are also favored. The glycine residues at position III and V, and tyrosine at position VI are conserved throughout the chitin-binding HLPs.

Figure 1.11: Sequence logo of 6C-, 8C- and 10C-hevein-like peptides chitin-binding domain. Amino acids occupying the chitin binding domain are numbered I to VI. Amino acid highlighted in purple are the key amino acid residues involved in chitin-binding. Cysteine residues are highlighted in yellow. Glycine residues are highlighted in green.

Multiple studies were performed to gain insights on how the modification of chitin-binding HLPs affect its interaction with chitin. Chemical synthesis of Hev32, a hevein mutant with truncation of ten amino acids at the C-terminal, was synthesized to understand the binding mechanism [144]. Data comparison of the thermodynamics of hevein and Hev32 with chitin revealed that the C-terminal truncation in Hev32 resulted in a non-significant loss of of ΔG^o, indicating that the truncation at the C-terminal has little effect on the binding affinity. Moreover, the larger H^o value of Hev32 was compensated for by the increase entropy ΔS suggests that the binding process is enthalpy-driven while entropy opposes binding. Therefore, the N-terminal 32 amino acid residues of hevein are the minimum residues need for efficient binding to chitin [144].

Since serine and the aromatic amino acids in the chitin-binding domain are involve in chitin-binding, mutations were made to these key residues and the effects were studied. The substitution of serine to aspartic acid in Hev32S19D resulted in a significant decreased in the binding affinity to chitin when compared to Hev32 [145]. The reduced binding affinity may be due to the absence of the hydrogen side-chain of aspartic acid, resulting in the loss of hydrogen bond to provide the driving force for the shuffling of tryptophan at position II, to facilitate CH-π stacking for the stabilization of the peptide-sugar complex [145]. Additionally, the significance of tryptophan at position II was investigated using techniques such as ITC and ^{1}H-NMR titration on pseudohevein, which has a tyrosine residue at position II instead of tryptophan [141]. The ^{1}H-NMR spectra showed numerous peaks with chemical shifts suggested the formation of complexes between pseudohevein and chitin.

Furthermore, ITC data comparisons of both hevein and pseudohevein titrated with the sugar showed a minute difference of in ΔH^{o} values, indicating that the change in binding affinity remains insignificant despite the substitution of tryptophan to tyrosine at position II [141].

Furthermore, the position II of Ac-AMP2 resides a phenylalanine residue. A study showed that substitution of phenylalanine residue to non-natural amino acids can either increase or decrease the binding efficiency [139]. Additionally, the substitution of aromatic amino acids to alanine resulted in a significant reduction of binding affinity. Therefore, this further indicates that the CH-π stacking interaction between the hydrophobic carbohydrate surface and the side-chain of aromatic amino acid residues is essential for efficient chitin-binding [139].

4.1.7 Biological activity of Chitin-binding Hevein-like Peptides

Aforementioned, hevein bears a high sequence homology with UDA therefore, its anti-fungal activity was investigated [87]. Despite the absence of a chitinase domain, hevein was able to induce anti-fungal effects on eight strains of phytopathogenic fungi. Furthermore, the anti-fungal activity of hevein was not lost under heat exposure, suggesting thermal stability. Besides, the 6C-chitin-binding HLP subfamily appear to have higher potency compared to hevein (Table 1.5). For example, the IC_{50} of AC-AMP2 and hevein against *Bortrytis cinereal* were 8 µg/mL and 500 µg/mL, respectively. *B. cinereal* is a chitin-containing fungus that infects agriculture crops causing gray mold [148].

Table 1.5: Anti-fungal activity (IC_{50}) of chitin-binding hevein-like peptides

Fungi	IC_{50} (μg/ml)								
Peptide	**6C-chitin-binding HLP**		**8C-chitin-binding HLP**				**10C-chitin-binding HLP**		
	Ac-AMP2	IWF-4	Hevein	Pn-AMP1	Fa-AMP1	mO1	Ee-CBP	EAFP1	WAMP-1a
Alternaria alternata	-	-	-	-	-	25	-	-	
Alternaria brassicola	4	-	-	-	-	60	3	-	-
Ascochyta lycopersici	-	-	-	-	-	-	-	155	-
Ascochyta pisp	8	-	-	-	-	-	-	-	-
Bipolaris sorokoniana	-	-	-	-	-	-	-	-	5
Bortrytis cinereal	8	-	500	16	-	-	1	-	20
Circospora berticola	-	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 capsicl	-	-	-	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	-	-	-	-	-	-	-	-

"-" denotes that the peptide has not been tested against the fungal strain. Ac-AMP2[94], IWF-4 [95] , Hevein [81], Pn-AMP1 [96], Fa-AMP1 [100], mO1 [52], Ee-CBP [101], EAFP1 [97]and WAMP-1a [85].

Within the 8C-chitin-binding HLP subfamily, Pn-AMPs can surpress gram-positive bacteria and fungi (chitin and chitin-free) growth [96]. Likewise, Fa-AMPs can inhibit both gram-positive and gram-negative bacteria [100]. In the 10C-chitin-binding HLP subfamily, Ee-CBP and WAMP-1a can inhibit both gram-positive bacteria and chitin-containing fungi [85, 110]. Furthermore, EAFP1 was reported to inhibit chitin-containing fungi with an IC_{50} range of 35–155 µg/mL [97].

Despite the establishment of the anti-fungal inhibitory effects in chitin-binding HLPs, their mode of action remains to be elucidated. However, three potential mechanisms have been presented. The first mechanism suggests that the size and chitin-binding properties of hevein, allow penetration to the cell wall of fungi and impede growth by binding to the developing chitin chain which results in the disparity between chitin-synthesis and hydrolysis [87]. Nevertheless, this mechanism cannot explain the anti-fungal inhibitory effects of chitin-binding HLPs such as Pn-AMPs against chitin-free fungi. Thus, a second mechanism was proposed. This mechanism suggests that the basic isoelectric point of Pn-AMPs allows it to penetrate fungal hyphae and localize at the hyphal tips of fungi resulting in the rupture of the hyphal tip and leakage of cytoplasmic material [96]. Lastly, the third mechanism suggests the inhibition of fungal metalloproteases. The chitin-binding domains of WAMPs resembled class I chitinases hence, their vulnerability to the glycine-cysteine cleavage by a fungalysin (Fv-cmp), belonging to the zinc metalloprotease (M36) family was assessed. Interestingly, WAMPs showed high tolerance towards the protease and inhibited its activity against chitinase. This inhibition was likely due to the additional serine residue found between the glycine-

cysteine cleavage site in WAMPs, conferring resistant to Fv-cmp. Hence, it was speculated that WAMPs bind to fungal metalloprotease and inhibit the enzyme activity against chitinase. As such, the intact chitinase can breakdown the fungal cell wall and prevent fungal growth [149].

While the chitin-free fungi inhibitory mode of action remains unclear, the anti-bacterial mechanism against gram-positive bacteria in chitin-binding HLPs appears to have been established. The cell wall of gram-positive bacteria contains peptidoglycan, which constitutes 90% of its dry weight. This polymeric carbohydrate consists of repeating units of *N*-acetylglucosamine (NAG), and *N*-acetylmuramic acid (NAM) similar to chitin can facilitate the binding of the anti-microbial peptide on the bacterial surface and induce membrane permeabilization [150]. Similarly, chitin-binding HLPs can bind to NAG units of the peptidoglycan and cause a “sponge-like effect” to attract more peptides on the membrane interface, which disrupt membrane integrity and eventually cause cell death [150]. Overall, these studies showed that chitin-binding HLPs have multiple mechanisms of action against fungi and bacteria based on their sequence, isoelectric point and the cell wall make-up of microorganisms. However, additional studies are still required to resolve the mode of action of chitin-binding HLPs in fungi to develop enhanced anti-fungal agents.

4.2 Non-chitin-binding Hevein-like Peptides

4.2.1 Discovery and Biological Characterization of Non-chitin-binding Hevein-like Peptides

As part of our laboratory discovery program, our laboratory has recently discovered a collection of CRPs that contained a similar cystine-knot motif as cystine-knot alpha amylase inhibitors (CKAI) [24, 59, 122] and chitin-binding HLPs [51, 53, 54]. However, these CRPs do not possess CKAI function and lack the chitin-binding domain. Hence, our laboratory defined this group of CRPs as non-chitin-binding HLPs. This subfamily of HLPs typically contain 27–40 amino acids and contain six to eight cysteine residues.

The discovery of non-chitin-binding HLPs was from *Hibiscus sabariffa*, commonly known as roselle or red sorrel [55]. This plant belongs to the Malvaceae family and is cultivated in many regions, including Africa and tropical regions such as Southeast Asia and India [151]. *H. sabariffa* displayed a myriad of pharmacological functions that include, anti-inflammatory [152], anti-hypertensive [153, 154], anti-hyperlipidemia [155, 156], antipyretic [152]and act as diuretic agents [157, 158]. In traditional Chinese medicine, it is used as a remedy for cough [159]. In 2016, Loo *et al.* isolated and characterized a group of non-chitin-binding HLPs from *H. sabariffa* that were termed as roseltides, rT1 to rT8 [55]. These peptides contained six cysteine residues, ranged from 27–39 amino acid residues, and consist of a cystine-knot motif similar to chitin-binding HLPs. However, they lack a chitin-binding domain, which is often located in loops 3 and 4 of the peptide's primary sequence. Interestingly, transcriptome analysis showed that roseltides are

ribosomally synthesized and contained a three-domain organization, including an N-terminal signal peptide, a pro-peptide domain and a C-terminal mature domain which is similar to CKAIs [55]. This precursor organization differed from chitin-binding HLP with the absence of a C-terminal tail and the presence of a pro-peptide domain. [55]. Functionally, rT1 is the smallest member of roseltide reported to be metabolically stable to acid and proteolytic treatments and is also a knottin-type neutrophil elastase inhibitor [55]. This family of elastase inhibitors was first discovered from the squash family in 1989 [160]. Yet, sequence analysis suggested that rT1 bears no homology to the squash knottin-type elastase inhibitor. The neutrophil elastase is a proteolytic enzyme essential for neutrophil function and can elicit inflammatory response[161]. The study showed that roseltide rT1 played an integral role in the interaction between the S1, S2 and S3 pockets of the human neutrophil elastase [55]. Hence, these findings indicated that roseltide rT1 is a 6C-non-chitin-binding HLP that can inhibit neutrophil elastase.

Shortly after, in 2017, the same group discovered a new suite of non-chitin-binding HLPs from *Pereskia bleo*, termed as bleogens pB1–pB15 [56]. This group of peptides contained six cysteine residues and ranged from 27–40 amino acids in length. In this period, the term “non-chitin-binding hevein” was defined. The cysteine spacing of bleogens follows the chitin-binding HLP with the absence of a chitin-binding domain. Additionally, bleogen pB1 was reported to exhibit heparin-binding activity and possessed anti-*Candida* properties with a minimal inhibitory concentration range of 5–10 μM [56]. Notably, transcriptome analysis of bleogens revealed three different precursor architectures, which were classified as type I,

type II and type IIb [56]. The type I precursor architecture reported in pB1 comprised an N-terminal signal peptide and mature peptide domain, whereas type II consists of the three-domain arrangement similar to roseltides [55] and type IIb contained a tandemly-repeating mature peptide domain. The diversity in the precursor architecture and function of bleogens further expanded our understanding of non-chitin-binding HLPs in plants.

In the following year, Tam *et al.* made a significant discovery of a suite of 8C-non-chitin-binding HLPs from the Araliaceae family: *Panax ginseng* (Asian ginseng), *Panax nontoginseng* (notoginseng) and *Panax quinquefolius* (American ginseng) [57]. These cysteine- and glycine-rich peptides were termed ginsentides, TP1–TP14 and contained 31–33 amino acids. Ginsentides contain a similar cysteine spacing as roseltides [55] and bleogen [56], and they are the first group of non-chitin-binding HLPs identified with eight cysteines. The precursor organization of ginsentide is similar to roseltide and the type II precursor architecture of bleogen, which constitutes an N-terminal signal peptide, a pro-peptide domain and a mature peptide domain. Although gisentides adopt a cystine-knot fold similar to CKAI and the CB-HLP family, it contains an unusual disulfide connectivity. This disulfide connectivity was revealed as CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII and had not been reported before [57]. Hence, this data provided new insights into the cysteine-rich peptides' cysteine framework and expanded the library of non-chitin-binding HLPs.

The progressive discovery of non-chitin-binding HLPs has raised the interest to characterized their biological function. In early 2019, Kam *et al.* discovered that

roseltide rT1 can penetrate cells and target the peptide to the mitochondria [162]. The cellular uptake of rT1 undergoes glycosaminoglycan-dependent endocytosis, and the peptide is localized in the mitochondria by binding to the mitochondrial protein import receptor TOM20. Furthermore, rT1 can enhance the production of ATP through mitochondrial membrane hyperpolarization, and target identification analysis further validated that rT1 targets the human mitochondrial membrane ATP synthase subunit O [162]. In the same year, the group also discovered that roseltide rT7 is a proteasome inhibitor [163]. The long and proline-rich nature of the intercysteine loop 4 in rT7 contained an IIML motif that catalyzed proteasomal degradation through a canonical substrate-binding mechanism. Imaging and cell-based analysis indicated that the penetration of rT7 in the cells led to the accumulation of ubiquitinated protein, inhibition of the 20S proteasomes, inhibition of tumor necrosis factor-induced $I_KB\alpha$ degradation and inhibition of intracellular adhesion molecule-1 (ICAM-1) expression [163]. Figure 1.12 shows the discovery timeline of non-chitin-binding HLPs. Collectively, these studies have highlighted the molecular and functional diversity of non-chitin-binding HLPs and showed their potential as leads for development of orally active biotherapeutics.

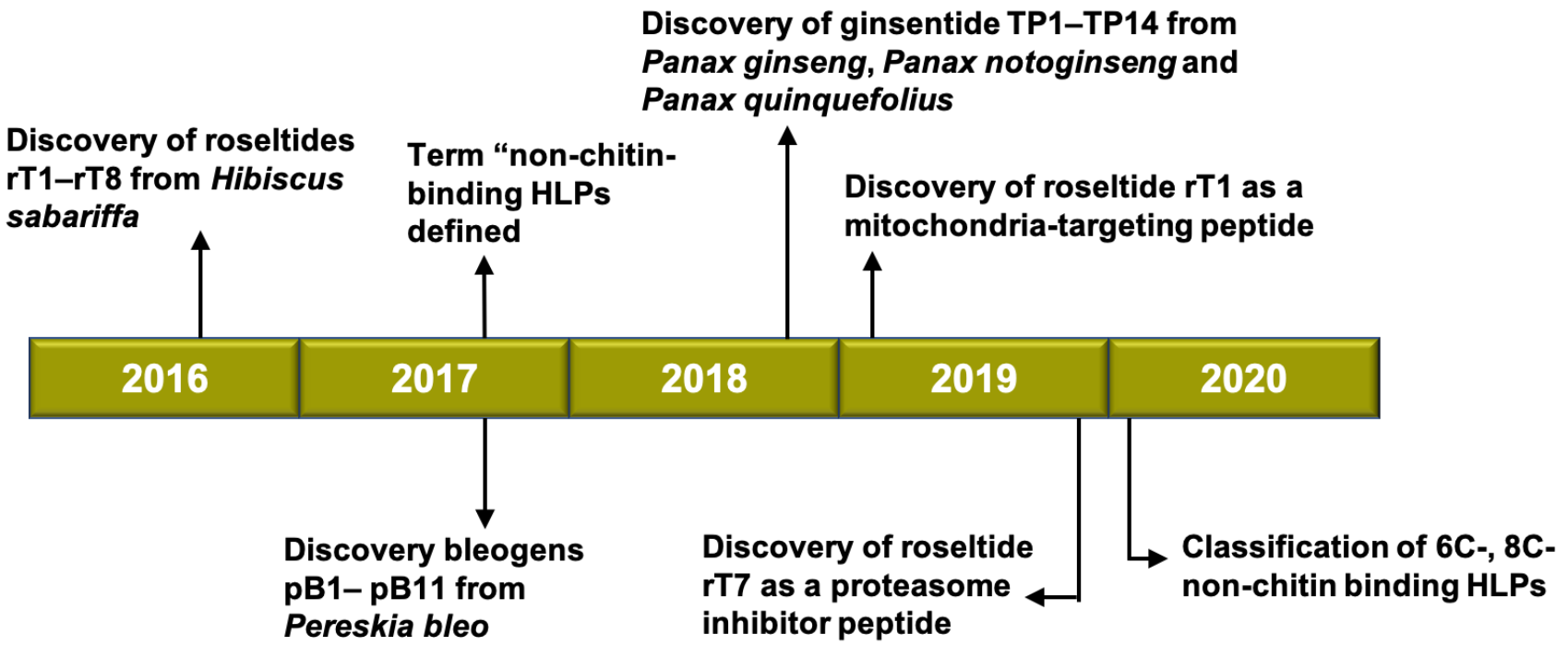


Figure 1.12: Discovery timeline of non-chitin-binding hevein-like peptides. The first collection of non-chitin-binding hevein-like peptide was discovered from *Hibiscus sabariffa.* Following that, 15 6C-non-chitin-binding hevein-like peptides were discovered from *Pereskia bleo* in 2017. In the same year the term "non-chitin-binding hevein-like peptide was defined". Subsequently 8C-non-chitin-binding hevein-like peptides were discovered from the Ginseng family with a unique disulfide connectivity. In 2019, the biological functions of non-chitin-binding hevein-like peptides were characterized. The discovery of non-chitin-binding hevein-like peptides has led to their classification based on the number of cysteine present.

4.2.2 Classification of Non-chitin-binding Hevein-like Peptides

The classification of non-chitin-binding HLPs is similar to chitin-binding HLPs. They are subclassified based on two key features: the number of cysteine residues present and their precursor architecture (Figure 1.4). Non-chitin-binding HLPs are cysteine-rich and can be further divided into 6C- and 8C-non-chitin binding HLPs. The 6C-non-chitin-binding HLP subfamily includes roseltides isolated from *H. sabariffa* [55] and bleogens isolated from *P. bleo* [56] whereas the 8C-non-chitin-binding HLP subfamily includes ginsentides isolated from Araliaceae family [57]. Similarly, the precursor architecture of non-chitin-binding HLPs can be further divided into two-domain, three-domain and tandem-repeated domains precursor.

This new subfamily of HLPs defined by our laboratory as non-chitin-binding HLPs has contributed to our knowledge of the diversity in HLPs. Furthermore, the progressive research in non-chitin binding HLPs may lead to the discovery of new peptides in other plants. Therefore, the classification of these peptides is foreseen to be more extensive in the near future.

4.2.3 Sequence Diversity of Non-chitin-binding Hevein-like Peptides

In general, non-chitin-binding HLPs contain 27–40 amino acid residues that are rich in cysteine. Thus far, 23 6C-non-chitin-binding HLPs have been successfully isolated and characterized. These include roseltides rT1–rT8 from *H. sabariffa* [55] and bleogen pB1–pB15 from *P. bleo* [56]. The 6C-non-chitin-binding HLP subfamily comprise of four intercysteinyl loops (Table 1.6). The cysteine spacing of roseltide is displayed as C-X_{6-9}C-X_{2-7}-CC-X_{3-4}-C-X_{3-13}C in which X represents other amino acid residues, and the subscript stands for the number of amino in each

intercysteine loop. The cysteine spacing of bleogen pB1–pB15 is represented as C-X_6-C-X_{6-9}CC-X_{1-4}C-X_{4-15}C. According to the cysteine spacing, the size of intercysteine loop 4 in roseltides and bleogens is the longest, comprising up to 13 and 15 amino acid residues, respectively. All the loop sizes in roseltides varied. In contrast, the size of loop 1 in bleogen is fully conserved, comprising six amino acids. Apart from the conserved cysteine positions, none of the amino acid in the 6C-non-chitin-binding HLPs were absolutely conserved, which is an indication of sequence diversity.

At present, 14 8C-non-chitin-binding HLPs have been successfully isolated and characterized. They are ginsentide TP1–TP14 isolated from the Araliaceae family [57] and display a cysteine spacing C-X_6-C-X_{6-7}CC-X_{1-4}C-X-C-X_{4-6}C-X_{1-4}C. The 8C-non-chitin-bindng HLP subfamily contained six intercysteine loops (Table 1.6). The cysteine positions in ginsentides are absolutely conserved. In addition, the size of intercysteine loop 1 and 4 is fully conserved comprising six and one amino acid, respectively, However, sequence variations in loop 1, 2, 3, 5 and 6 suggested that within the same plant family, sequence diversity can occur.

Table 1.6: Consensus sequences of selected non-chitin-binding hevein-like peptides

Peptide	Peptide Mature Sequence	Reference
6C-non-chiting binding HLP loop	I II III IV	
Roseltide		
rT1	-CIPRG--GICL-VALSGCCNSP--GCIF----------GICA-----	Loo *et al* 2016
rT2	-CLPSG--SSCS-MWKGECCNGF---CIPVS--------MVSGICP-----	Loo *et al* 2016
rT4	-CNISG--SFCYSVST--CCSGK---CTSPSIFP-----PTPAHCV-----	Loo *et al* 2016
rT7	-CVSSGIVDACS-----ECCEPD--KCIMLPTWP------PRYVCSV----	Loo *et al* 2016
rT8	-CKPVG--ASCSDPSE--CCSGI---CLFPINFPINFPIPTSGTCFRKITG	Loo *et al* 2016
Bleogen		
pB1	QCKPNG--AKCTEISIPPCCSNF---CLR---YAG----QKSGTCANR---	Loo *et al* 2017
pB2	-CVGAG--SYCNIVFGPKCCSQF--FCVP------PGIAG--GMC------	Loo *et al* 2017
pB8	-CATRY--ESCNPQEGLECCGDF--ICILPPIWP-----PVPGRC------	Loo *et al* 2017
pB14	-CKNLN--EICNISARVVCCEGL--TCVEHKHE--KGMLGVPGTC------	Loo *et al* 2017
pB15	-CATVG--ETCSPSTQHFCCGHL--SCKPNK----PGVLGGGGKC------	Loo *et al* 2017
8C-non-chiting binding HLP loop	I II III IV V VI	
Ginsentide		
TP1	-CKSGG--AWCG-FDPHGCCGN--CGCLV--GFCYGTGC-	Tam *et al* 2018
TP2	-CKSSG--AWCG-FDPHGCCGN--CGCLV--GFCYGTGC-	Tam *et al* 2018
TP7	-CKSGG--TWCG-FDPHGCCGN--CGCLV--GFCYGTGC-	Tam *et al* 2018
TP11	-CLKNG--QFCW-GNPSGCCGN--CGCLIIPGVCYGTGC-	Tam *et al* 2018
TP14	-CLKVG--KICLGRGLKECCPSATCGCLL--GFCIK--C-	Tam *et al* 2018

4.2.4 Biosynthesis of Non-chitin-binding Hevein-like Peptides

The non-chitin-binding HLPs are gene-encoded products synthesized by the ribosome and targeted to the endoplasmic reticulum (ER) for bioprocessing. Generally, the precursor gene arrangement of these peptides consists of a three-domain architecture which includes an N-terminal ER signal peptide (27–35 amino acid residues), a long pro-peptide domain (34–63 amino acid residues) and a mature domain (27–40 amino acid residues) [55-57]. Ribosomal protein synthesis occurs in the cytosol and then translocated to the ER where the signal peptide is cleaved. Following that, protein disulfide isomerase and chaperone facilitate the proper folding of the pro-peptide, misfolded pro-peptide will remain inside the ER lumen targeted for degradation [164]. The correctly folded pro-peptide will leave the ER and transported to a vacuolar compartment where it will be cleaved by pro-peptide convertase to release the mature domain (Figure 1.13A). This precursor arrangement is similar to the precursor architecture of CKAIs [59].

In contrast to chitin-binding HLPs, the precursor architecture of non-chitin-binding HLPs do not contain a C-terminal tail [52-54] but a pro-peptide at the N-terminal of the mature domain. Many studies have shown that the N-terminal pro-peptide may assist the folding of the nascent protein into their tertiary structure during protein synthesis [165, 166]. For example, the N-terminal pro-peptide of pro-aminopeptidase processing protease (PA protease) function as an intramolecular chaperone to facilitate the folding of this enzyme [165]. This suggests that the pro-peptide of non-chitin-binding HLPs may have similar function. However, the pro-peptide sequence of non-chitin-binding HLPs did no bear homology with the pro-

peptide of PA protease. Therefore, the function of the pro-peptide of non-chitin-binding HLPs remain to be elucidated.

Interestingly, Loo *et al*. showed that non-chitin-binding HLPs have other precursor gene arrangements [56]. For example, bleogen pB1 contained a two-domain architecture that constitutes a N-terminal signal peptide and a mature domain. Besides, the precursor architecture of bleogen pB12–pB15 contained tandem repeated mature domains (Figure 1.13B). The tandem repeats are speculated to facilitate plant adaptation to harsh environmental stressors and to thrive in these conditions [167]. Although, these repeats are found in a variety of plant families [167], bleogen pB12–pB15 appear to be the first suite of non-chitin-binding HLPs reported to contained tandem-repeated domains. Overall, the variations in the precursor architecture of non-chitin-binding HLPs suggest that they are a diverse subfamily of HLPs, contributing to functional diversity. Therefore, by understanding the biosynthesis of non-chitin-binding HLPs, it may provide clues with regards to their biological functions.

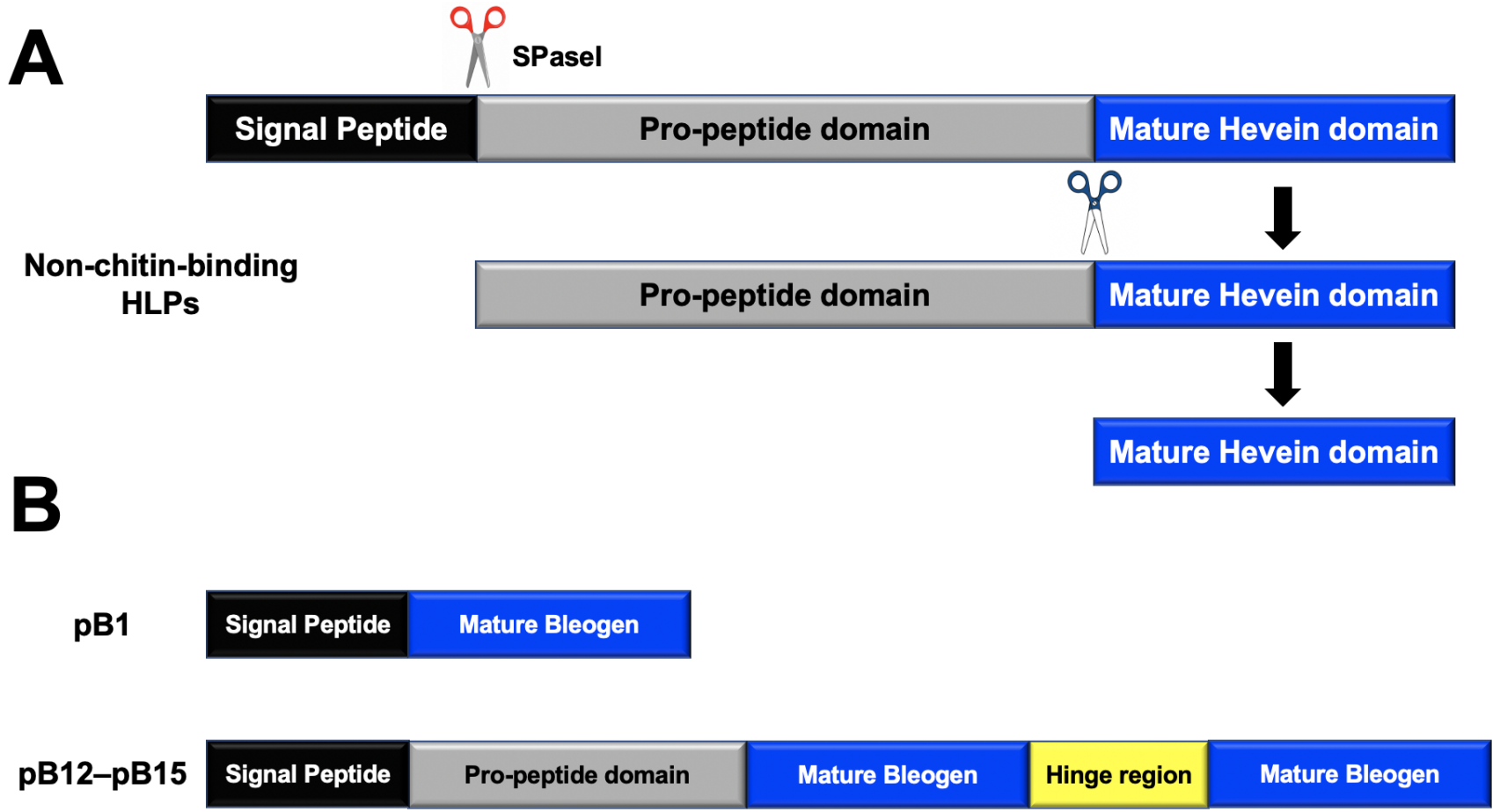


Figure 1.13: Biosynthesis of non-chitin-binding hevein-like peptides. **(A)** The precursor organization of most non-chitin-binding hevein-like peptides include a signal peptide, a pro-peptide domain and a mature domain. **(B)** Other types of precursor organization of non-chitin-binding hevein-like peptides include a two-domain precursor that comprising a signal peptide and a mature domain (pB1) or a five-domain precursor that comprise a signal peptide, a pro-peptide and a hinge region between tandem-repeated mature domains (pB12– pB15).

4.2.5 Structure of Non-chitin-binding Hevein-like Peptides

The structure of non-chitin-binding HLPs contains a cystine-knot fold similar to CKAIs [122] and chitin-binding HLPs [52], with a disulfide connectivity of CysI–CysIV, CysII–CysV and CysIII–CysVI. In the chitin-binding HLP subfamily, they contain anti-parallel β-strands and a short helical turn at the C-terminal [34]. However, within the 6C-non chitin-binding HLPs, they appear to have slight variations. For example, the secondary motif of roseltide rT1 does not contain any β-strands or α-helices (Figure 1.14A). Although bleogen pB1 structure contains anti-parallel β-strands, it has no α-helices (Figure 1.14B).

The structure of 8C-non-chitin-binding HLP subfamily comprises a central β-sheet made up of two anti-parallel β-strands. This structure configuration is different from the 8C-chitin-binding HLPs which typically consist of three β-strands [34]. In addition, 8C-non-chitin-binding HLPs contain a unique disulfide connectivity. For example, the disulfide connectivity of ginsentide TP1 is revealed as CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII. The first three disulfide linkages form the cystine-knot fold, and the last disulfide linkage is a penetrating disulfide bond (Figure 1.14C) [57]. Hence, the structural diversity in non-chitin-binding HLPs may provide new insights for target identification studies to further elucidate their potential molecular mechanism in cellular processes.

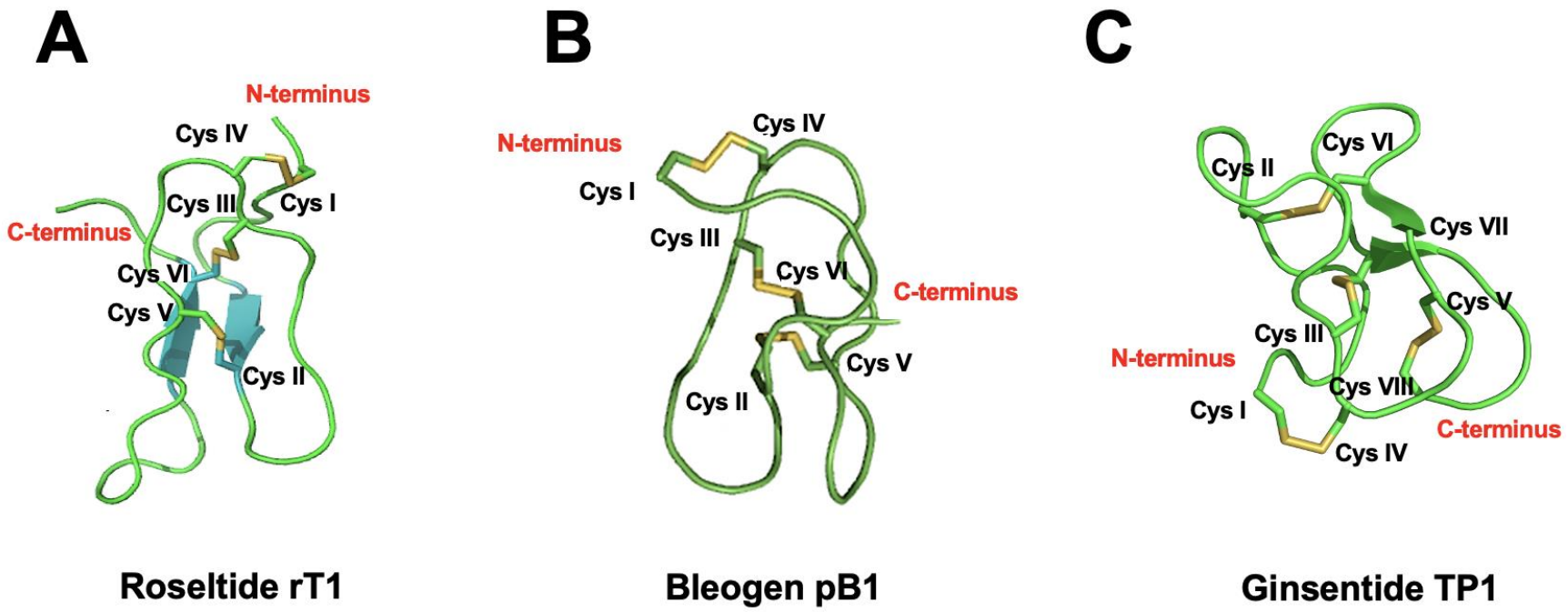


Figure 1.14: NMR structures of non-chitin-binding hevein-like peptides. **(A)** Roseltide rT1 (PDB: 5GSF), **(B)** Bleogen pB1 (PDB:5XBD) and Ginsentide TP1 **(C)** TP1 (PDB:2ML7). The 3D-structure of pB1 and TP1 contain anti-parallel β-strands. In contrast, rT1 do not contain any secondary structure.

5. Grain as Functional Food

The earliest record of grains domestication dates back to approximately 23 000 years ago in Israel [168]. Along with its long-documented history, grains have become the most important source of the world's total food and an essential staple in developing countries in the present day. According to a global statistic, 2.1 billion tons of grain are produced worldwide, suggesting its economic importance [169]. Apart from their nutritional value as functional foods, grains contained many bioactive components that assist in plant defense. Concerning HLPs, many of these grains were reported to contain high amounts of lectins [170, 171]. Lectins are chitin-binding proteins that may contain multiple chitin-binding domains or a chitin-binding domain in tandem to a catalytic domain such as chitinase [108]. As such, it is speculated that HLPs are constitutively expressed in plants. Generally, grains can be broadly classified into cereal or pseudocereal grains. Cereal grains are grasses from the Poacea family that are cultivated for its grains. Some examples include barley, maize, millet, oats, rice, rye, sorghum and wheat. In contrast, pseudocereal grains are non-grasses cultivated similarly for its grain. These include amaranth grains, buckwheat and quinoa.

5.1 Cereal Grain

5.1.1 Barley

Barley (*Hordeum vulgare*) is ranked fourth in the global production of grains. In 2018 to 2019, the worldwide production of barley was an astounding 140 million metric tons. The barley's primary usage mainly caters to daily diet consumption [172, 173] and malt production [174]. The chemical composition of barley includes

carbohydrate (74%), protein (9.3%), fat (1.9%), fiber (3.7%), vitamins, minerals and other inorganic compounds [175, 176]. Barley contains 19 amino acids but is low in essential amino acids such as lysine and methionine, which is the reason why traditional barley recipes are prepared with beans to compensate for the amino acid deficient [177]. In beverage industries, barley is an essential ingredient in the production of beer and whiskey [178]. Also, the grains are widely used to make porridges, bread, pudding, biscuits, soups and stews. Therefore, barley is considered as a valuable crop in the food industries.

Barley also contains a variety of phytochemicals such as phenolic acids, flavonoid, folate, phytosterol and tocols, which exhibit potent antioxidant, antiproliferative and anti-cholesterol activities [179]. Functionally, barley was reported to help suppress a variety of health-related diseases including, diabetes, cancer, obesity, circulatory disorders and cholesterol levels [180-185]. Interestingly, young barley leaves extract was reported to degrade organophosphate pesticides [186]. In plant defense, many studies have reported that barley contains chitinase proteins, which displayed anti-fungal activity against a plethora of fungi including *Botrytis cinereal, Pestalotia theae, Bipolaris orzya, Curvularia lunata* and *Rhizoctonia solani* [187-191]. Taken together, barley grain is a multi-purpose grain that has shown its importance as a functional food.

5.1.2 Maize

In 2018, maize (*Zea mays*), also known as corn, was ranked the highest in the worldwide production of grains, producing almost 1.1 billion ton per annum. Maize has become a staple food in many regions of the world. Most of the maize produced

are for ethanol production, animal feed and food products such as sweet corn, popcorns, corn flour, corn oil and alcoholic beverages like bourbon whiskey [192, 193]. Maize kernels are composed of mainly water (76%), carbohydrates (19%), proteins (3%) and fat (1%). Similar to barley, it contains suboptimal amounts of essential amino acid such as lysine and tryptophan. Hence, for a complete diet, legumes are included to overcome the amino acid deficient [194].

The phytochemicals found in maize include phenols, anthocyanins, flavonoids, phytosterols, carotenoids, vitamins and xanthophylls, which have health-promoting effects such as antioxidant and antiproliferative activities [195-197]. These phytochemicals have also shown their effects to reduce the risk of chronic diseases such as cancer [198], cardiovascular disease [199], obesity [200] and type 2 diabetes [201]. As such, many consume this grain not only for its nutrition but also for health benefits. Besides, maize also contains anti-fungal properties against a wide range of phytopathogenic fungi. For example, Huynh *et al.* reported two chitinases, chit A and chit B from maize seeds that displayed anti-fungi effects against *Fusarium oxysporum Alternaria solani, Sclerotinia sclerotiorum, Gaeumannimyces graminis* and *Phytophthora infestans,* with inhibition concentration as low as 0.5 μg [202]. Hence, with their wide range of biological function, maize is considered a valuable crop that can be further exploited for the potential therapeutic effects.

5.1.3 Millet

The global millet production is approximately 27.8 million ton per annum. India dominates as the global millet producer, whereas Africa is the largest global

consumer [203]. In developing countries such as China, India, Nigeria, Mali and Sudan, food and nutritional security present major challenges [204]. Millet is a promising solution as this functional crop is readily available, cheap in cost, rich in nutrients and is also drought resistant. Hence, it can be grown in countries with water scarcity to supply to the locals. There are 19 species of millet. However, only four species are cultivated for food consumption. This includes *Eleusine coracan*, *Panicum millaceum*, *Pennisetum glaucum* and *Setaria italica* . These grains contain approximately 70% carbohydrates, 11% protein, 7% fiber, 5% fat and has high amounts of vitamins and minerals [205]. For consumption, millet grains can be prepared in many ways and are generally found in a variety of traditional food such as bread, porridge, millet wine and powder formulations.

Millet grains are rich in phytochemicals such as polyphenols, lignans, β-glucan, inulin, phytates sterols, tannins and flavonoids which have antioxidant and other pharmacological properties [206] [207]. Biologically, these bioactive compounds can prevent a variety of diseases by lowering cholesterol [208], reducing cardiovascular plasma triglyceride [209], preventing the progression of cancer [210] and improving metabolic syndrome [211]. Furthermore, millet also contains anti-microbial activity that inhibits fungal growth [212-214]. A study by Radhajeyalakshmi *et al.* showed that *P. glaucum* contained a 45 kDa chitinase, which can inhibit the fungal growth of *Rhizoctonia solani* and *Trichoderma viride* [213]. Collectively, millets contain many major and micronutrients that are used as “food medicine,” representing an alternative cereal in developing countries.

5.1.4 Oat

Oat (*Avena sativa*) is ranked sixth in the global production of grains. It was estimated that 22 million tons of oats is produced annually. This grain provides a vital food staple for people in many countries and consumed as medicinal herbal supplements for its highly acclaimed health benefits. The major composition of oats includes carbohydrate (60%), protein (15%), dietary fiber (8.5%), fat (9%) and beta-glucan (4%) and other inorganic compounds [215]. This nutrient-rich grain is found in a variety of baked goods such as muesli, granola bars, cookies and cakes. It is also eaten as porridge and used as a thickener in soup. In China, oat flour is processed into noodles and consumed as a staple food by the local [216].

The phytoconstituents of oats include avenanthramides, alkaloids, saponins, tocols, sterols and flavonoids, which show biological properties such as sedative, antioxidant and anti-microbial activities [217, 218]. Traditionally, oats are used as a folk medicine to treat insomnia, nervous exhaustion and rheumatism [219]. Pharmacologically, oats were reported to lower cholesterol [220], enhance nitric oxide production [221], production of anti-inflammatory transforming growth factor beta 1 (TGFβ1) [222], stimulate wound- healing [223] and anti-microbial activity [218]. Such diverse pharmacological function is the reason why people continue to consume oats as part of their diet. Furthermore, oats also exhibit anti-fungal activity [109, 224]. For example, avesin A, an 8C-chitin-binding HLP isolated from oats, can inhibit *Fusarium culmorum*.[109]. Collectively, the use of oats appears to be helpful in health maintenance and the potential of oats as therapeutic agents and nutraceutical are promising.

5.1.5 Rice

Rice (*Oryza sativa*) is ranked as the third most grain produced globally with an estimated 495 million ton per annum. It is consumed worldwide, especially in Asia and Africa. Although maize has shown to be the highest amount of grains produced globally, rice provides one-fifth of the global dietary energy supply [225]. This indicates that the dietary consumption of rice is more preferred compared to maize. The major composition of the cooked rice grain is composed of water (68%) carbohydrate (28%) and protein (3%). Although rice may be a staple food, it does not contain essential amino acids and hence, must be paired with other sources of protein such as fish and meat for a complete diet. A variety of food products that are made from rice include milk, wine, noodles and rice cakes such as mochi, a traditional Japanese rice cake [226].

Rice is believed to contain some medicinal properties despite the lack of scientific evidence. In traditional Chinese medicine, rice is known to strengthen the spleen and cure indigestion [227]. Additionally, in Indian folklore medicine, rice water is prescribed as an ointment for surface inflammation [228]. Furthermore, some clinical studies have shown that it can alleviate acute diarrhea [229], reduced cholesterol in human [230] and contained antioxidant and anti-microbial properties [231, 232]. In plant defense, rice contains chitinase which has been widely reported for its broad spectrum of anti-fungal activity against numerous fungal species [233-235]. Therefore, the anti-fungal properties in rice chitinases have been exploited to produce transgenic crops. For example, the incorporation of rice chitinase RCG3 gene into tomato showed higher resistance to *Alternaria solani* compared to the

wild-type tomato. Therefore, it indicated that the chitinase gene enhanced tomato crop resistance towards pathogenic invasion [233]. As such, the widespread usage of rice may not be limited to dietary consumption but to non-food applications to produce an enhanced transgenic pathogen-resistant crop and anti-microbial agents.

5.1.6 Rye

Rye (*Secale cereal*) contributes approximately ten million ton of worldwide production of grains. It is primarily cultivated in Eastern, Central and Northern parts of Europe. However, due to its increasing demand, it has spread to countries like Canada, North America, South America, Australia and China [236]. Furthermore, this crop can be grown in marginal soil, demonstrating its high adaptability compared to other cereal grain. The major constituents of rye grain include carbohydrate (28%), protein (20%), dietary fiber (54%), fat (2%), vitamins and minerals. Unlike other cereal grains, rye has a high protein content. In general, rye grain can be ground into flour to produce a variety of baked goods such as bread, beer and crisp bread. Additionally, the grain is also used to brew beer and produce rye whiskey [237, 238]. Therefore, the vast food applications of rye suggest its importance in the food and beverage industry.

Although there are little reports on the clinical aspect of rye, some studies have shown that the phenols found in rye can prevent the oxidation of low-density lipoprotein (LDL), which may reduce the risk of atherosclerosis and cardiovascular-associated diseases [239, 240]. In plant defense, rye contains proteins such as chitinases that are very well characterized and have shown to induce anti-fungal effects on a plethora of fungi [241, 242]. For example, rye seed chitinase-a (RSC-

a) inhibited *Trichderma sp*, *Fusarium oxysporum* and *Rhizoctonia solani* with IC_{50} in picomole concentration suggests its potent anti-fungal activity [242]. Taken together, the protein-rich rye grain may be an alternative cereal to compensate for the amino acid deficient in other cereal. However, further studies are required to identify more bioactive compounds and elucidate their pharmacological functions.

5.1.7 Sorghum

Sorghum (*Sorghum bicolor*) is ranked as the fifth-highest grain produced globally [243]. It is estimated that approximately 58 million ton of sorghum is produced every year. This functional crop, once native to Africa, is now cultivated worldwide for food consumption [244]. Like millet, sorghum varieties are also drought and heat tolerant which allow them to grow in developing countries with arid regions. As such, this grain is considered an important staple for people in developing countries. The major components in sorghum include carbohydrate (26%), dietary fiber (26%), protein (22%) fat (4%), vitamins and minerals [245]. This grain is used to make food and alcoholic products such as flour, bread, muffins, cookies, pasta, beer and kaoliang liquor, a popular liquor in China [246]. It is also used in ethanol production and bioplastics[247, 248]. As such, the multi-purpose use of this grain has tremendously boosted its attractiveness in the food and ethanol industry.

Biologically, sorghum contains a wide range of phenolic compounds that are capable of inducing antioxidant activity [249]. For example, policosanols in sorghum have shown to reduce LDL and increase high-density lipoprotein (HDL) which is promising for the prevention and therapy of heart-related disease [250]. Furthermore, chitinases identified in sorghum have reported to exhibit anti-fungal

properties [251-254]. For example, CH1, CH2 and CH3 chitinases purified from sorghum seeds have shown affinity towards chitin and inhibited the growth of phytopathogenic fungi like *Trichoderma viride*, *Rhizoctonia solani* and *Fusarium moniliforme* [254]. Hence, apart from its nutritional value and industrial application, sorghum has a considerable potential as biotherapeutics.

5.1.8 Wheat

Wheat (*Triticum sp.*) is the second-highest grain produced globally. In 2018, the global production of wheat was approximately 734 million tons. Due to the increasing demand, this functional crop is grown on more land area than other cereal grains. The major compositions of wheat consist of carbohydrate (71%), protein (13%), water (13%), fat (1.5%) and other inorganic compounds [255]. Although it has a high protein content, it lacks essential amino acids such as lysine [256]. In food application, wheat can be found in a variety of products such as flour, bread, pasta, noodles and plant-based meat used in vegetarian Chinese cuisine [257].

Apart from its nutritional value, wheat also contained medicinal properties. Traditionally, they are used to treat a wide range of ailments such as constipation [258], skin irritation [259], and toothache [260]. Clinically, the phytochemicals in wheat like phenols, ferulic acid, flavonoids and carotenoids have shown to contain antioxidant [261], anticancer [262], anti-cholesterol [263] and antihypertensive [264] properties. Hence, it may be the underlying reason why this grain is highly sought after. Furthermore, the anti-fungal properties of wheat are also highly studied. For example, phenolic acids in wheat have shown to inhibit an array of *Fusarium*

species [265]. Additionally, proteinaceous compounds in this grain such as wheat germ agglutinin (WGA) and WAMPs also displayed anti-fungal activities [114] [85]. WGA is a lectin that contains four chitin-binding domains whereas WAMPs are 10C-chitin-binding HLPs. This further suggests that chitin-binding HLPs may be constitutively expressed in grains, which contributes to their anti-fungal activity. Collectively, wheat has shown its nutritional value and medicinal benefits. Hence, it is foreseen that the demand for wheat will continue to surge in the future.

5.2 Pseudocereal Grain

5.2.1 Amaranth grains

Amaranth grains (*Amaranthus sp*.) are grains that belong to the genus of *Amaranthus*. There are approximately 75 species of *Amaranthus*. However, only a few are cultivated for food consumption [266]. These include *Amaranthus caudatus*, *Amaranthus hypochondriacus* and *Amaranthus cruentus*. These grains have been cultivated for over 8000 years and it was a form of a staple for the Aztec people [267]. Previously, these grains were not widely studied. However, in the 1980s, research began to peak due to the recognition of its nutritional value and agronomic potential [268]. The major components of cooked amaranth grain include water (75%), carbohydrate (19%), protein (4%), fats (2%), vitamins and minerals [269]. Besides, these grains contain high amounts of lysine, which is usually deficient in cereals [270]. At present, amaranth grains are mostly cultivated in Mexico as a food crop and is processed in many ways for consumption. Some examples of food products made from amaranth grains include tamales, tortillas and atole. Culturally, it is used as food offerings for the spirits on the Day of the Dead in Mexico.

To date, many phytochemical compounds such as phenols, sterols and flavonoids from these grains have reported to contained health-promoting effects that include antioxidants [271], anti-cholesterol [272], antitumor [273], antihypertensive [274] and anti-anemic [275]. Furthermore, Ac-AMP2, a 6C-chitin-binding HLP isolated from *A. caudatus* has also showed its potential as an anti-fungal agent [94]. Although most of the research focus on the nutrition beneficial function of amaranth, there are bioactive compounds in the grains that showed potential therapeutic properties. Hence, future research is essential to elucidate amaranth grains mode of action in human.

5.2.2 Buckwheat

Buckwheat (*Fagopyrum esculentum*) is a common pseudocereal belonging to the Polygonaceae plant family. The global production of this grain is estimated at 1.5 million tons every year [276]. The crop is native to East Asia but has progressively spread to Europe and North America [277]. The major compositions of buckwheat are carbohydrate (72%), protein (13%), dietary fiber (10%), fat (3%), vitamins and minerals [278]. Therefore, its highly nutritious value has been recognized and used in many different regions to prepare traditional food cuisines. For example, buckwheat noodles are popular staples adored by people in China and Japan. In eastern Europe, buckwheat porridge is an everyday household staple whereas in Korea, a well-loved jelly called “memilmuk” is made from the starch of buckwheat [276].

Apart from its nutritional value, this grain appears to contain medicinal properties. Traditionally, buckwheat is used as a remedy to treat constipation and anemia [279].

At present, many phytochemicals identified in this grain such as fagopyrin, phenols, lignan, flavonoids and phytosterol, make it a biologically active pseudocereal [280] Functionally, phenols such as hydrobenzoic acids can stimulate antioxidant activity [281]. These antioxidant properties is commonly associated with biological functions such as antiaging [282], antimutagenic [283] and anticarcinogenic [284]. In addition, rutin, a flavanol compound rich in buckwheat, has shown medical potential against edema and hemorrhagic diseases. It plays an vital role in regulating blood circulation [285]. Furthermore, Fa-AMP1, an 8C-chitin-binding HLP isolated from buckwheat provided evidence for the grain's anti-fungal activity [100]. Taken together, buckwheat has established itself as an essential nutraceutical. By understanding the functions of these bioactive compounds, the putative therapeutic value of this grain can be used to prevent or treat various diseases.

5.2.3 Quinoa

Quinoa (*Chenopodium quinoa*), a pseudocereal belonging to the Amaranthaceae plant family, has become increasingly popular. The global production of quinoa from 2009 to 2019 showed a three-fold increase from 75 000 to 230 000 metric tons [286]. The crop which was originally cultivated in the high altitude of Andes mountain and termed as "the mother of grains" by the Incas, is now grown in arid regions such as Saudi Arabia and Uzbekistan [287]. Therefore, this showcased the resilience and high adaptability of the quinoa crop. The major components of quinoa include carbohydrate (69%), protein (16.5%), fat (6.3%), vitamins and minerals [288]. Its unique protein profile features a high quantity of lysine and

methionine, which are generally low in cereal [289]. Hence, it is favored by health enthusiasts and is considered a functional food in modern times.

The grains of quinoa are commonly used to make salad, soups and a variety of baked dishes. In traditional medicine practice, quinoa is used as a remedy against anthrax [290], sore throat [291], urinary tract infection [292], and to improve blood circulation [293]. Furthermore, clinical studies using mouse models have reported that the seed extract of quinoa reduced plasma and liver cholesterol [294], decrease lipid peroxidation [295], increase blood plasma antioxidant capacity [296] and promote wound-healing [297]. There are over 100 components identified from quinoa including saponins, carotenoids, tocopherols, betalains and phenols [289, 298]. These compounds play an essential role in plant defense and show great bioactivities such as antioxidant [299], anti-fungal [300], antiviral [301], antithrombotic [302], anti-obesity[303] and anti-inflammatory [304] activities. As such, the combination of dietary and medicinal benefits is why quinoa continues to be an important functional food from Andean times to present.

As part of our laboratory discovery program of novel HLPs, many medicinal plants have been well-studied [51-57]. However, grain-derived HLPs are poorly studied compared to medicinal plant HLPs. In contrast, grains that encode the chitinase gene such as barley, maize, millet, rice, rye, sorghum are also well-established [189, 202, 213, 233, 242, 254]. Table 1.7 summarized the sequence comparison of selected chitin-binding domains in grains. Thus far, grain-derived HLPs were reported in Ac-AMPs from amaranth grains [94], Fa-AMPs from buckwheat [100], Avesin A from oat [109] and WAMPs from wheat [85]. To the best of our knowledge,

no HLPs have been reported in quinoa and although Avesin A has been reported in oat, it is not well-represented. Therefore, the grains of quinoa (*Chenopodium quinoa*) and oat (*Avena sativa*) (Figure 1.15) are selected as representatives to characterize the HLPs in cereal and pseudocereal, respectively.

Table 1.7: Comparison of HLPs and chitin-binding domains of lectin in functional grains

Grain	Protein/Peptide Name	Chitin-domain Sequence	Reference
Barley	CHI1 HORVU	----EQCGSQAGGATCPNCLCCSRFGWCGSTP-YCGDGCQS-QCSGC--	Kirubakaran *et al* 2007
Maize	CHIA MAIZE	--------QNCG---CQPNFCCSKFGYCGTTDAYCGDGCQSGPCRS---	Huynh *et al* 1992
Millet	CHI-1 MILLET	--------QSCG---CQPNFCCSKFGYCGTTIDYCGDGCRSGPCIGS--	Radhajeyalakshmi *et al* 2003
Rice	CHI2 ORYZA	----EQCGAQAGGARCPNCLCCSRWGWCGTTSDFCGDGCQS-QCSGC--	Mizuno *et al* 2008
Rye	CHIA RYE	---AEQCGSQAGGATCPNCLCCSRFGWCGSTSDYCGDGCQS-QCAGCGG	Yamagami *et al* 1993
Sorghum	CHIB SORGHUM	-------AQNCG---CQAGYCCSKFGYCGTTDDYCGDGCQSGPCRS---	Krishnaveni *et al* 1999
Wheat	WGA A	----QRCGEQGSNMECPNNLCCSQYGYCGMGGDYCGKGCQNGACWT---	Smith *et al* 1989
Wheat	WGA B	---SKRCGSQAGGATCTNNQCCSQYGYCGFGAEYCGAGCQGGPCRA---	Smith *et al* 1989
Wheat	WGA C	-----KCGSQAGGKLCPNNLCCSQWGFCGLGSEFCGGGCQSGACST---	Smith *et al* 1989
Wheat	WGA D	-----PCGKDAGGRVCTNNYCCSKWGSCGIGPGYCGAGCQSGGCDG---	Smith *et al* 1989
Amaranth	Ac-AMP1	---VGEC--VRGR--CPSGMCCSQFGYCGKGPKYCG----------	Broekaert *et al* 1992
Amaranth	AC-AMP2	---VGEC--VRGR--CPSGMCCSQFGYCGKGPKYCGR----------	Broekaert *et al* 1992
Buckwheat	Fa-AMP1	----AQCGAQGGGATCPGGLCCSQWGWCGSTPKYCGAGCQS-NCK----	Fujimura *et al* 2003
Oat	Avesin A	---WSGCSP------CPGNECCSKYGYCGLGGDYCGAGCQSGPCYG---	Li *et al* 2006

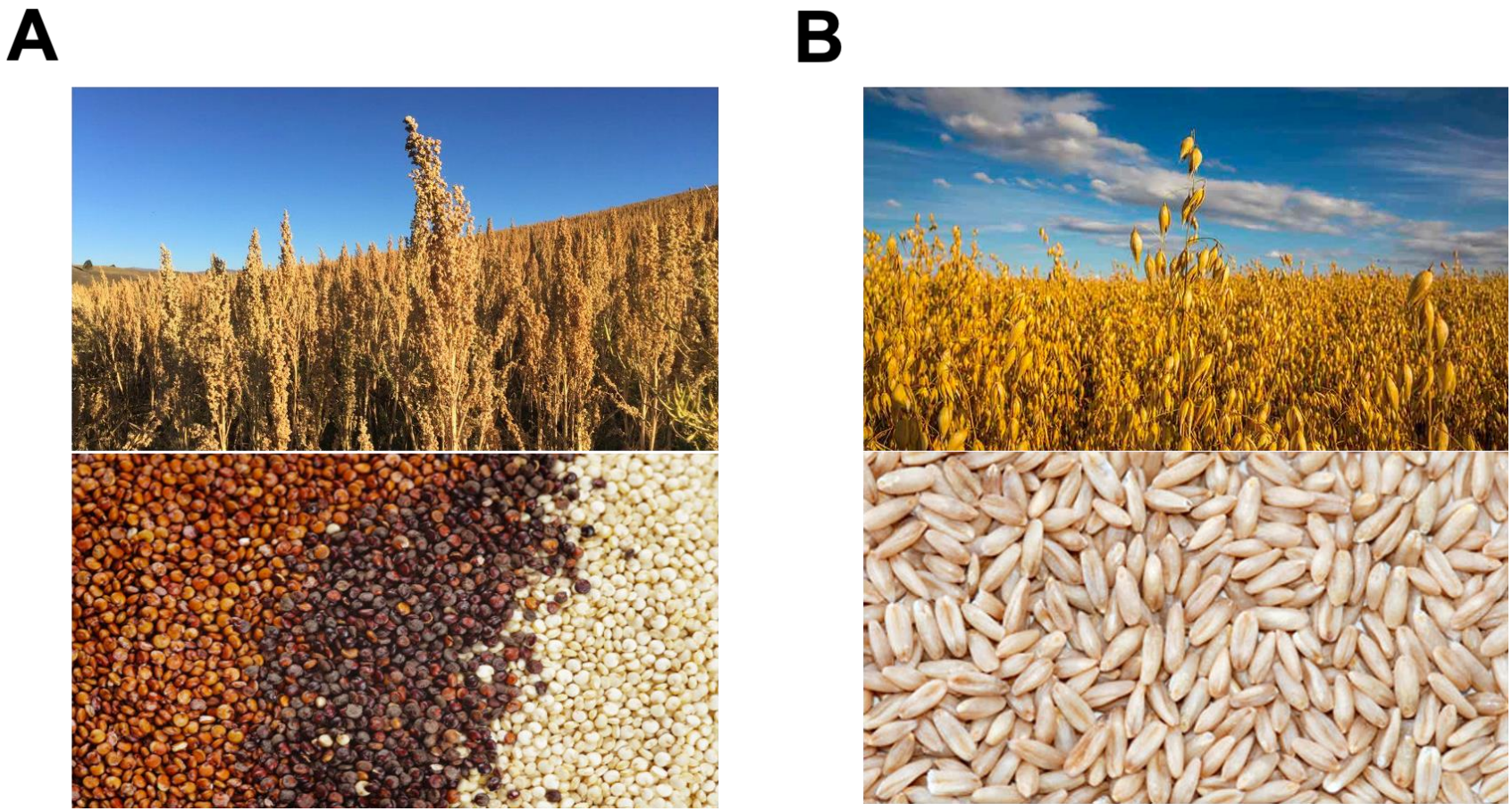


Figure 1.15: Functional grains selected for the study. (A) *Chenopodium quinoa* and **(B)** *Avena sativa*. The images were adapted from https://www.bbc.com/news/world-latin-america-45008830 and https://www.fwi.co.uk/arable/crop-management/nutrition-and-fertiliser/oat-yields-held-back-low-fertiliser-rates for *Chenopodium quinoa* and https://www.huffpost.com/entry/quinoa-facts-didnt-know_n_4494695 and https://www.fieldstoneorganics.ca/products/organic-whole-grains/oats.php for *Avena sativa.*

6. *Eleutherococcus trifoliatus*

Eleutherococcus trifoliatus (苦力心), formerly known as *Acanthopanax trifoliatus,* is a thorny shrub belonging to the Araliaceae family (Figure 1.16). This medicinal plant is commonly cultivated in Southeast Asia and is known to be incorporated into food cuisines for consumption [305]. The roots and leaves of *E. trifoliatus* are widely used in traditional Chinese medicine (TCM) for its ginseng-like activity and to treat ailments such as diarrhea, common cold, and cough [306]. In other folk-medicine practices, it provides a curative effect for the treatment of ulcers, impotence, gout, neuralgia and ameliorates memory [307, 308].

Although the literature has shown few pharmacological studies on this plant, it has recently been recognized, with an increasing number of studies performed to elucidate its bioactive compounds. Thus far, there are more than 80 components identified in *E. trifoliatus*. The major compounds comprise mainly of secondary metabolites like terpenoids phenylpropanoid, flavonoids, alkanes, lipids and steroids which show great biological effects such as antioxidative [309], anti-ulcer [310], anti-nociceptive [311], anti-tumor [312], anti-hyperalgesia [313]and anti-inflammatory [314] effects *in vivo* or *in vitro*. For example, Chien *et al.* demonstrated that the extract of *E. trifoliatus* can inhibit lipopolysaccharide-induce (LPS) inflammatory response [314]. The *in vitro* study showed that the treatment of *E. trifoliatus* to mouse macrophages cells inhibited the protein expression of nitric oxide synthase and cyclooxygenase-2 [314]. These proteins play an integral role in the pathogenesis of inflammation, and the ability to inhibit these proteins suggests the presence of anti-inflammatory activity. Furthermore, the group performed *in*

vivo studies on mice model with LPS-induced acute lung injury and showed that the lung tissue of mice treated with *E. trifoliatus* extract reduced neutrophil infiltration near the pulmonary vessel when compared to the control. Western blot further confirmed that the treatment of *E.trifoliatus* extract in mice significantly inhibits the production of $I_kB\alpha$ and NF_kB degradation in the lung tissue, suggesting inflammation inhibitory effects[314].

Despite the vast pharmacological studies, most of these bioactive compounds are attributed to small molecules. Since TCM application of *E. trifoliatus* requires decoction to exert medicinal effects, we postulate that the bioactive principles in this medicinal plant could be thermal-stable peptides such as CRPs that are abundant in plants.

Figure 1.16: Medicinal plant *Eleutherocuccus trifoliatus* selected for the study. This plant belongs to the Araliaceae plant family. Its roots and leaves are commonly used in traditional Chinese medicine to treat a variety of ailments due to its ginseng-like properties. The image was obtained from https://www.flickriver.com/photos/mingiweng/48611471767/.

CHAPTER TWO

Hypothesis and Specific Aims

Hevein-like peptides (HLPs) represent a highly underexplored family of CRPs with diverse biological functions, including chitin-binding and anti-fungal activities. They are found in all functional grains, which include rice, sorghum, barley, millet, rye, wheat and oats as well as certain medicinally-important herbs. Importantly, the significance of their conservation in grains, cereals and functional food are poorly understood. The hypothesis of my thesis is that the defense function of HLPs is an evolutionarily selected trait, and these HLPs are active constituent compounds important for the evolutionary survival of functional grains and certain medicinal plants.

The specific aims of my thesis are:

1. To discover novel HLPs from functional grains (*Chenopodium quinoa* and *Avena sativa*) and medicinal plant (*Eleutherococcus trifoliatus*)
2. To characterize their biosynthesis, disulfide connectivity, 3D-structure, metabolic stabilities and putative biological functions
3. To study the diversity and distribution of HLPs in plants using bioinformatic approach

Chapter one describes the background of CRPs and HLPs. Chapter two states my hypothesis and specific aims. Chapter three describes the materials and method. Chapter four reports the characterization of the first suite of 6C-chitin-binding HLPs from *C. quinoa* with novel tandem-repeated precursor architecture. The structure-

activity relationship, metabolic stability and potential applications in wound healing and as an anti-fungal agent will be studied. Chapter five reports the characterization of nine novel 8C-chitin-binding HLPs from *Avena sativa* with a unique four tandem repeats in their precursor sequence that has not been reported before. Their binding affinities to chitin-containing fungi, metabolic stability and putative biological activities will also be described. Chapter six reports the characterization of the first 8C-non-chitin-binding HLP isolated from *Eleutherococcus trifoliatus*. The biosynthesis, metabolic stability and potential cytoprotective function will be discussed in this chapter. Finally, chapter seven revealed a total of 1204 novel HLP sequences from 252 plant families through *in silico* identification. These peptides were expressed mostly in angiosperms followed by gymnosperms, bryophytes, euphyllophyta and lycopodiophyte, suggesting that HLP could be the largest family of plant CRPs. In summary, this thesis describes the discovery of novel 6C- and 8C-HLPs from *Chenopodium quinoa*, *Avena sativa* and *Eleutherococcus trifoliatus,* and the identification of putative HLPs in plants through bioinformatic approach.

CHAPTER THREE

Materials and Methods

1. Materials

1.1 Chemicals and Reagents

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

Acetic acid	Merck
Acetonitrile (ACN)	Fisher
Ammonium bicarbonate (NH_4HCO_3)	Sigma-Aldrich
Dichloromethane (DCM)	Merck
Dithiothreitol (DTT)	Sigma-Aldrich
Ethanol (EtOH)	Merck
Formic acid (FA)	Sigma-Aldrich
Iodoacetamide (IAA)	Sigma-Aldrich
Isopropanol	Fisher
Methanol	Merck
Sodium chloride (NaCl)	Sigma-Aldrich
Trifluoroacetic acid (TFA)	Merck

1.2 Cell Lines

All cell lines used in this thesis were purchased from the American Type Culture Collection (ATCC) and cultured according to manufacturer's instructions.

Table 3.1: Cell lines used in this study

Cell Line	Catalog Number
A431 cells	ATCC® CRL-1555™
C2C12	ATCC® CRL-1772™
HeLa cells	ATCC® CRM-CCL-2™
Human umbilical vein endothelial cells (HUVEC)	ATCC® CRL-1730™
H9c2 cells	ATCC® CRL-1446™
SH-SY5Y cells	ATCC® CRL-2266™

1.3 Fungal Strains

All fungal strains used in this thesis were obtained from China Center of Industrial Culture Collection (Beijing, China) and cultured according to manufacturer's instructions.

Table 3.2: Fungal strains used in this study

Fungal Strain	Catalog Number
Alternaria alternata	CICC 2465
Curvularia lunata	CICC 40301
Fusarium oxysporum	CICC 2532
Rhizoctonia solani	CICC 40259

1.3 Plant Material

Chenopodium quinoa (willd), *Chenopodium quinoa* (red), *Chenopodium quinoa* (black), and *Avena sativa* were purchased from the local supermarket. *Eleutherococcus trifoliatus* was obtained from School of Biological Sciences herb garden. Macroscopic and microscopic analysis were performed by Mr. Paul Leong at the Singapore Botany Center to authenticate the plant samples. A voucher of each sample (NTUH-CQ20170503-01, NTUH-CQ20170503-02, NTUH-CQ20170503-03, NTUH-CQ20170503-04 and NTUH-CQ20170503-05) was deposited at the Nanyang Technological University Herbarium, School of Biological Sciences, Singapore.

2. Instrumentation

2.1 MALDI-TOF MS and MS/MS

Mass spectrometry was performed on the ABI 4800 MALDI-TOF/TOF system (Applied Biosystems, Framingham, MA, USA). The instrument features a solid-state laser (diode pumped Nd: YAG laser) pulsing at a repetition rate of 200 Hz. A saturated solution of CHCA in 60% ACN, 0.05% TFA was used as a MALDI matrix. Samples were mixed in a 1:1 ratio (v/v) with the matrix and 0.5 µL of the mixture was spotted onto a target plate. The instrument was calibrated externally using a series of peptides standards obtained from Sigma-Aldrich (MSCal1). Both MS and MS/MS spectra were obtained using the reflectron mode. The laser intensity was set between 3000–4500. The accelerating voltage applied for MS and MS/MS were 20 kV and 8 kV respectively with an average spectrum from 1000 and 5000 shots in MS and MS/MS mode, respectively.

2.2 HPLC and UPLC Analysis

The Shimadzu system were used for both the high-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography (UPLC). Reverse-phase-HPLC (RP-HPLC) was performed using Phenomenex C18 column (particle size 5 µm; pore size, 300Å; Hesperia, CA, USA) with dimensions of 250 x 21 mm for preparative purification, and 250 x 4.6 mm for analytical purification at a flowrate of 5 mL/min and 1mL/mins respectively. A polyLC polysulfoethyl A column was used for strong cation exchange (SCX)-HPLC with dimensions of 250 x 4.6 mm at a flowrate of 1 mL/min.

2.3 LC-ESI-MS/MS

The reduced-alkylated peptide sample was desalted using Milipore Ziptips and lyophilized. The lyophilized peptide was re-dissolved in 0.1% formic acid prior to MS analysis. An Orbitrap Elite mass spectrometer (Thermo Scientific Inc., Bremen, Germany) coupled with a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific, Bremen, Germany) was used to perform LC/MS-MS analysis. Elution was performed over 1 h 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. A Michrom's Thermo CaptiveSpray nanoelectrospray ion source (Bruker-Michrom, Auburn, USA) was used to generate the spray. A Full FT-MS (350-2000 m/z, resolution 60.000, with 1 µscan per spectrum) was alternated with Full FT-MS and a FT-MS/MS scan applying 27%, 30% and 32% normalized collision energy in high-energy collisional dissociation (110-2000 m/z, resolution 30.000, with 2 µscan

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

analysis with parent error tolerance and a fragment error tolerance of 10 ppm and 0.05 Da respectively.

2.4 Spectrophotometric Analysis of Protein Concentration

The Beer-lambert law was used to calculate the concentration of the purified peptide based on the equation

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

Where,

A: The absorbance of peptide dissolved in milliQ water at 280 nm measure using the NanoDrop (Thermo Fisher Scientific. USA)

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

I: cell path length (cm)

The theoretical ε value of a protein at 280 nm was calculated based on [315]

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

Where,

*n*Trp : the number of tryptophan residues

*n*Tyr : the number of tyrosine residues

*n*SS: the number of disulfide linkages

3. Proteomics

3.1 Preliminary Screening of CRPs in grains

Grains of *Chenopodium quinoa* and *Avena sativa*, and the plant of *Eleutherococcus trifoliatus* were screened. Each of the sample (0.1 g) were homogenized with 1 mL of milliQ water and centrifuged at 1000g for 10 min at 4 °C. The supernatant was concentrated using Millipore Ziptip and spotted on the MS target plates for MS analysis.

3.2 Peptide Extraction and Purification

Seeds (5–10 kg) of *Chenopodium quinoa* and *Avena sativa* and the leaves (1kg) of *Eleutherococcus trifoliatus* were extracted separately and homogenized in water (1:1 ratio). Debris was filtered using a muslin cloth and the filtrate was centrifuged at 10 000 rpm for 20 min at 4 °C. After centrifugation, the supernatant was then subjected to 80% ammonia precipitation for 1 h and centrifuged at 10 000 rpm for 20 min at 4 °C. The pellets were collected, pooled, dissolved in 10% ACN for 1 h and centrifuged at 10 000 rpm for 10 min at 4 °C. The supernatant was collected, filtered through 1 µM and 0.45 µM pore size filter and loaded on a C18-flash column (Grace Davison, Columbia, MD, USA). Increasing concentration of ethanol (20%–80%) was used for elution from the column. Presence of the desired peptides in the eluents was confirmed by MALDI-TOF MS and these fractions were pooled and loaded on a SCX flash column containing 100 mL of SP-Sepharose Fast Flow

cation exchange resin (GE Healthcare, CA, USA). The column was washed with buffer A (20 mM NaH_2PO_4; pH 2.5) and eluted with buffer B (20 mM NaH_2PO_4 and 1 M NaCl; pH 2.5). Fractions from the SCX flash column containing the peptide of interest were pooled and purified using several rounds of RP-HPLC and SCX-HPLC. A linear gradient from buffer A (20 mM NaH_2PO_4; pH 2.5) to buffer B (20 mM NaH_2PO_4 and 1 M NaCl; pH 2.5) was performed for the removal of small molecules. Fraction from SCX-HPLC that contained the desired peptides were pooled purified by RP-HPLC using buffer A (water with 0.1% TFA) and buffer B (100% ACN with 0.1% TFA).

3.3 Reduction and Alkylation

The number of disulfide bonds present in each peptide was determined by the mass shift after reduction and alkylation of the crude extracts. Reduction and alkylation were performed as previously described [316]. Briefly, the peptide extracts were incubated at 37 ºC with 20 mM dithiothreitol (DTT) in 20 mM ammonium bicarbonate (NH_4HCO_3) buffer at pH 8 for one hr. Subsequently, peptides were alkylated with 40 mM iodoacetamide (IAA) at room temperature for 1 h.

3.4 *De novo* Sequencing

Approximately 50 µg of purified peptides was dissolved in 20 mM ammonium bicarbonate (NH_4HCO_3) buffer and was incubated with 20 mM dithiothreitol (DTT) at 37 ºC for 1 h to reduce the disulfide bonds. Digestion of the reduced samples was performed at room temperature for 15 min with trypsin and chymotrypsin respectively. *De novo* sequencing of the peptides was performed based on the presence of both b- and y-ion series in the tandem mass spectrometry (MS/MS)

profiles as previously described [316]. IIe/Leu and Lys/Gln were assigned based on the genomic or transcriptomic sequences.

3.5 Disulfide Mapping

Purified peptide was partially reduced in 50 mM tris (2-carboxyethyl) phosphine (TCEP) at 37 °C for 25 min. Next, 250 mM N-ethylmaleimide (NEM) was added and incubated at 37 °C for 30 min. The partially alkylated peptide was quenched immediately by injecting into analytical RP-HPLC. Reduced and alkylated species were separated with a linear gradient of 20%–60% buffer B (100% ACN with 0.1% TFA). Fractions were collected and analyzed by MALDI-TOF MS to verify NEM-alkylated intermediate species. The two- and four- NEM alkylated intermediated species were fully reduced and alkylated by 20 mM DTT and 40 mM iodoacetamide (IAA), respectively at 37 °C for 1 h.

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% NaN_3 in D2O or 10% D2O. Nuclear Overhauser effect spectroscopy (NOESY) experiments were acquired with 200 and 300 ms mixing times [317, 318]. Total correlation spectroscopy (TOCSY) [319] data were recorded with a mixing time of 69 or 78 ms using MLEV17 spin lock pulses [320]. Vicinal coupling constants were measured using the double-quantum-filtered (DQF)-COSY [321] and one-dimensional (1D)-1H-NMR experiments. All 2D-NMR data were recorded in the phase sensitive model using the time-

proportional phase increment method [322], 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 D2O solution. The water signal was suppressed using water-gated pulse sequences [323] or excitation sculpting [324] combined with pulsed-field gradients. All NMR data were processed using Bruker TOPSPIN 2.1 or NMRPipe [325] 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

The structure calculation was calculated by hybrid distance geometry using the software CNSsolve 1.3 [326]. Based on the intensities of the NOE peaks, the NOE distance restraints were loaded and were categorized into three groups: strong (1.8−3.0 Å), medium (1.8−3.4 Å) and weak (1.8−5.0 Å). Backbone dihedral angle restraints were determined based on the 3JHN-Hα coupling constraints in 1D 1H NMR spectrum. The φ angle was considered to be between −100° to −160° when the coupling constant was more than 8 Hz. The structure was verified by PROCHECK program [327] and displayed using PyMOL version 1.849 [328].

4.3 Ligand Peptide Docking

Prior to ligand and peptide docking, both the peptide and ligand were prepared using the Chimera version 1.11.2 for the addition of hydrogen atoms and

conversion of PDB format to MOL2 format. The GOLD 5.4.0 version [329] was utilized to perform ligand peptide docking. The GOLD score function takes into consideration of the hydrogen bond and Van de Waals energy. To define the active site pocket, one atom on the active site was chosen to define the pocket radius. The other settings in the program were set at default function.

5. Stability Assays

5.1 Heat Stability Assay

Purified peptides were incubated at 95 °C for 2 h. At each time interval (0 min, 30 min, 60 min, 90 min and 120 min), 20 µL of the treated sample was aliquoted in triplicates and quenched in an ice bath for 10 min. RP-UPLC was performed to determine the amount of peptide present before and after the heat treatment. Peaks were collected from UPLC and monitored by MALDI-TOF MS. Peak area from the UPLC were analyzed were compared to evaluate stability. Alkylated peptide was used as control for this assay.

5.2 Acid Stability Assay

Purified peptides were incubated with 1 M hydrochloric acid (HCl) for 2 h. At each time interval (0 min, 30 min, 60 min, 90 min and 120 min), 20 µL of the treated sample was aliquot in triplicates and quenched with 2 M sodium hydroxide (NaOH). RP-UPLC was performed to determine the amount of peptide present before and after the acid treatment. Peaks were collected from UPLC and monitored by MALDI-TOF MS. Peak area from the UPLC were analyzed were compared to evaluate stability. Alkylated peptide was used as control for this assay.

5.3 Endoproteolytic Enzyme Stability Assay

Purified peptides were added to 10 mM of hydrochloric acid with 4mg/mL of pepsin and incubated at 37ºC for 6 h at a final peptide to enzyme ratio of 20:1 (mol/mol). At each time interval (0 h, 2 h, 4 h and 6 h), 20 µL of the treated sample was aliquoted in triplicates and quenched with 2 M sodium hydroxide (NaOH). RP-UPLC was performed to determine the amount of peptide present before and after pepsin treatment. Peaks were collected from UPLC and monitored by MALDI-TOF MS. Peak area from the UPLC were compared to evaluate stability. Alkylated peptide served as control for this assay.

5.4 Exoproteolytic Enzyme Stability Assay

Purified peptides were added to 20 mM tricine and 0.05% BSA with 20 U/mL of aminopeptidase I enzyme and incubated at 37 °C for 6 h at a final peptide to enzyme ratio of 20:1 (mol/mol). At each time interval (0 h, 2 h, 4 h and 6 h), 20 µL of the treated sample was aliquoted in triplicates and quenched with 2 M hydrochloric acid (HCl). RP-UPLC was performed to determine the amount of peptide present before and after treatment. Peaks were collected from UPLC and monitored by MALDI-TOF MS. Peak area from the UPLC were analyzed were compared to evaluate stability. Alkylated peptide served as control for this assay.

5.5 Serum Stability Assay

Stability of purified peptide in human serum was assessed according to the protocol by Jenssen *et al*. [330]. Briefly, 25% human serum (Sigma Aldrich, Singapore) was prepared in phosphate buffer solution (PBS) and incubated at 37 °C for 20 min. Purified peptides were added to the temperature-equilibrated human serum and

incubated at 37 ºC for 36 h. At each time interval (0 h, 12 h, 24 h and 36 h), 50 µL of the treated sample was aliquot in triplicates and added to 100 µL of 95% ethanol and incubated at 4 °C for 10 min for precipitation reaction of the serum proteins. The sample was centrifuge at 13 000 rpm for 10 min to pellet down the precipitated proteins. The supernatant (50 µL) was then injected to the RP-UPLC to determine the amount of peptide present before and after serum treatment. Peaks were collected from UPLC and monitored by MALDI-TOF MS. Peak area from the UPLC were analyzed were compared to evaluate stability. Alkylated peptide served as control for this assay.

6. Bioassays

6.1 Chitin Binding Assay

The assay was performed as previously described [51]. Briefly, purified peptides were mixed with chitin beads (BioLabs, UK) in chitin binding buffer (140 mM NaCl, 10 mM Tris, 1 mM EDTA and 0.1% (v/v) Tween at pH 8 and incubated at room temperature for 30 min. After incubation, the mixture was centrifuge at 13 000 rpm for 10 min and the supernatant was collected. Chitin beads were then washed with chitin binding buffer. Elution of chitin binding peptides was done with 1 M of acetic acid at 55 °C. The supernatant and eluent were injected into RP-UPLC. Peaks were collected from UPLC and monitored by MALDI-TOF MS.

6.2 Disc Diffusion Assay

The anti-fungal activity of purified peptides were examined using a radial disk diffusion assay as previously described [331]. Four fungal strains, *Alternaria alternata*, *Curvularia lunata*, *Fusarium oxysporum* and *Rhizoctonia solani* were

selected for this study. Briefly, the fungal strains were grown in 90 mm x 15 mm Petri dishes containing 25 mL of potato dextrose agar at 25 °C. When sufficient fungal mycelial growth was observed, a hole was punched in the fungal culture plate and transferred to the center of the new agar plate. The plate was then incubated at 25 °C for 48 h until a radial mycelial colony form. Paper discs with the diameter of 0.65 cm were placed at the growing ends of the mycelial at equidistant of 1 cm. Aliquots of purified peptide of 1 mg/mL, 5 mg/mL and 10 mg/mL dissolved in 20 µL of milliQ were added to the respective paper discs and 20 µL of deionized water was used as a negative control. The plates were incubated for 48 h at 25 °C until the radial mycelia colony covers half the surface of the filtered paper disk of the negative control. Formation of a crescent-shape inhibition zone around the disk indicated susceptibility of fungi to the purified peptides.

6.3 Microbroth Dilution Assay

The half maximal inhibitory concentration levels (IC_{50}) of the purified peptides were determined by the microbroth dilution assay [332]. Fungal spores were harvested from a 4-day old actively growing fungal plate and suspended in 4 mL of half-strength potato dextrose broth. In the 96-well microplate, 80 µL of spore suspension (1 X 10^5 cells/ mL) was mixed with 20 µL of peptides at varying concentrations and incubated at 25 °C for 24 h. The cells were fixed with 100% methanol for 15 min and aspirated. Following, staining was done for 45 min with crystal violet dye and milliQ water was used to wash off the excess dye. Elution was performed using 1:1(v/v) ethanol/ 0.1N HCl. Absorbance was measured at 570 nm using Infinite@ 200 PRO Tecan microplate reader (Tecan Group Ltd, Germany).

Control wells were half-strength media containing fungal spore without peptide treatment. 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 the GraphPad Prism 6 for Windows GraphPad Software, San Diego California USA, www.graphpad.com

6.4 Cytotoxicity Assay

The 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) assay was used to evaluate the viability of cells (HUVEC, H9C2 and SH-SY5Y) treated with different concentration of the peptide of interest. Cells in a 96-well plate (1 x 10^4 cell/well) were treated with different concentration of the peptide (0, 1.25, 2.5, 5, 10, 20, 40, 80 µM) for 24 h. MTT (Biosharp, China) was added to each well, and incubated for 4 h at 37 °C. Subsequently, the medium was removed and dimethyl sulfoxide (Sigma-Aldrich, USA) was added (150 µL/well) to solubilize the formazan crystals. Optical density was measured at 570 nm using Infinite@ 200 PRO Tecan microplate reader (Tecan Group Ltd, Germany), and cell viability was calculated as a percentage of the control optical density.

6.5 Cell Migration Assay

A341 cells (1 X 10^6) were seeded in 60 mm dishes in 3 mL of Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin. The cells are incubated in 5% CO_2 incubator at 37 °C for 48–72 h to reach 100% cell confluence. For long term peptide treatment, after 24 h, the old medium was replaced with peptide (10 µM) containing medium and

incubated for another 24 h. Subsequently, the cells are washed thrice with PBS and once with serum free DMEM medium. The scratch was performed using a 10 µL pipette tip and removal of the cell debris was done by washing the cells with serum free DMEM medium and then replace with 3 mL of serum free DMEM medium. To obtain the same field during the image acquisition, reference points was made by etching the dish lightly with a razor blade on the outer bottom of the dish. After the reference points were made, the dish was placed under a phase-contrast microscope to acquire the first image of the scratch. After imaging, the cells were incubated in 5% CO_2 incubator at 37 °C for 6–24 h. After incubation, the dish was place under the phase-contrast microscope, matching the reference point and the second image was acquired. Images acquired for each sample were quantified by measuring the distance between one side of the scratch and the other using Image Pro-Plus software (Media Cybernetic, USA). The migration distance was calculated by comparing the images from time 0 to the last time point.

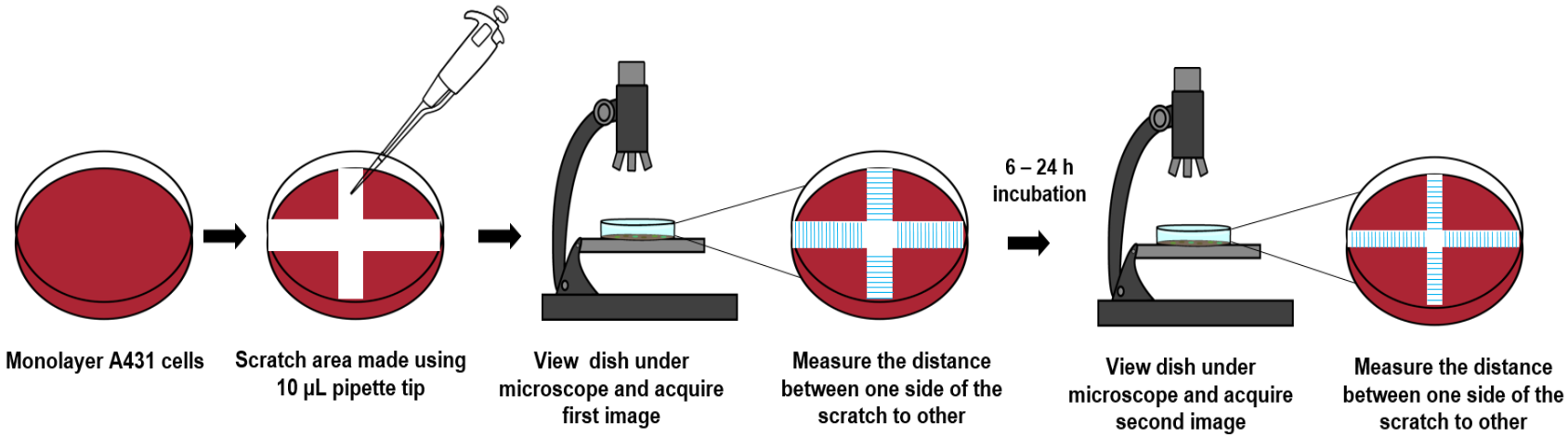


Figure 3.1: Schematic diagram of scratch assay

6.5 Cell Penetrating Assay

HeLa cells were seeded in a 12 well chamber slide (ibidi, Martinsried, Germany) with a initial cell density of 5×10^4 cell/mL and incubated for 24 h at 37 °C. After 24 h, 200 nM of MitoTrackerTM Red CMSRos (Invitrogen, CA, USA) was added to the cells and incubated for 20 min at 37 °C. PBS was used to wash the well twice before the addition of 30 μM of peptide and further incubated for 90 min. After 90 min, the medium was removed, and the well was washed with PBS twice. Subsequently, 4% polyformaldehyde was added and incubated for 1 min before washing with PBS. To permeate the cell membrane, 0.1% Triton X in PBS was added. Mounting medium FluoroshieldTM containing 4′, 6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich, Missouri, USA) was added to the chamber slide and covered with coverslips for confocal microscopy.

7. Bioinformatic Analysis

7.1 Data Mining

Database search were performed using methods modified from previous literature [333, 334]. tBLASTn was employed to search for expressed sequence tags (ESTs) or whole genome shotgun contig (WGS) encoding putative hevein-like peptide precursors in two databases, National Center for Biotechnology Information (NCBI) [335], 1000 Plants Project (OneKP) [336] and PhytAMP via query search using all reported hevein-like peptide sequences. The retrieved sequences were then used as queries in subsequent tBLASTn searches until no novel sequences were identified. The maximum target sequence and expected threshold were set at 1000 and 10, respectively. The sequences were translated by ExPaSy translation tool.

Accession numbers of all putative hevein-like peptide sequences are provided in the appendix. Selection of the remaining sequence are based on the following criteria: (1) The open reading frame must comprise of a start (ATG) and stop (TAA, TAG, and TGA) codons, following the C-terminal tail. (2) The translated amino acid sequence must contain six, eight or ten cysteine residues. The sequences were then submitted to SignalP 4.0 [337] for the identification of the signal peptide cleavage site. Replicated sequences from the same plant or different species of the same genus sharing identical full-length precursor were eliminated from the dataset. The sequences were then aligned using ClustalW [338].

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 ± standard error of the mean (SEM) in which p-values less than 0.05 were considered as statistically significant. Weblogo was used to visualize the aligned sequence [339]. Lastly, the construction and annotation of the phylogenetic tree was performed using the free online tool, Interactive tree of life (iTOL) [340].

CHAPTER FOUR

Chenotides: Tandem-repeated 6C-chitin-binding Hevein-like Peptides with Anti-fungal Activity from *Chenopodium quinoa*

1. Introduction

Chenopodium quinoa (*C. quinoa*) is a pseudocereal grown in the Andean region for thousands of years. It belongs to the Amaranthaceae family and is approximately two meters in height with deep penetrating roots. This adaptogenic plant was a form of staples for the Andean people and exhibited high tolerance towards extreme climate changes such as drought, high salinity and cold. It also demonstrates high adaptability, allowing growth in marginal land and cultivation at high altitudes of 3800 m above sea level. The grains of *C. quinoa* are highly nutritious, containing a myriad of minerals, vitamins and high protein contents. Furthermore, it contain a rich amount of amino acid such as histidine, lysine and methionine, which are generally deficient in cereal [341].

With the rising *C. quinoa* consumption, many studies have been done to characterize the bioactive compounds in this superfood.[286]. A study performed by Zhu *et al.* identified six flavanol glycosides with high antioxidant capacity in *C. quinoa* seeds and proposed it as a good source of free radical scavengers [342]. Interestingly, *C. quinoa* is one of the rare crop found to contain phytoecdysteroids that exert an array of pharmacologic and metabolic effects in mammals [343-345]. Moreover, secondary metabolites such as thiols, polyphenols and saponins

commonly found in commercially-discarded seed coats of *C. quinoa* were reported to play a role in the microbial defense of the plant [345, 346]. In addition, a study demonstrated that *C. quinoa* exerts hypocholesterolemic effects on mice models by reducing the plasma and liver total cholesterol level [294]. Therefore, these findings further support *C. quinoa* extracts as a potential nutraceutical and its medicinal importance as a functional crop.

Despite the major documentation of *C. quinoa,* its biological activities are primarily associated to the plant extract's secondary metabolite. Thus far, minimal studies with regards to peptides in *C. quinoa* have been reported. This is attributed to the misconception that peptide-derived biologics are unstable, easily degraded by heat, prone to digestion by proteolytic enzymes in the gastrointestinal tract and have low bioavailability due to the poor absorption in the guts [347, 348]. However, the progressive recognition of cysteine-rich peptides (CRPs), has refuted this misconception. CRPs are a class of underexplored biomolecules rich in cysteine residues and can be found extensively throughout the plant kingdom with molecular weight between 2–6 kDa [349]. This class of peptides forms intramolecular disulfide bonds, endowing the peptide with high stability against thermal, acidic and proteolytic degradation [71, 124, 350]. Some examples of CRPs include thionins [67], defensin [351], hevein-like peptides (HLPs) [81], knottin [122], α-hairpinin [62], Lipid Transfer Proteins (LTP) [352] and snakin [353].

This chapter reports the isolation and characterization of three new tandem-repeated 6C-chitin-binding HLPs from *Chenopodium quinoa*. These peptides are collectively termed as chenotides, cQ1–cQ3. Proteomic, genomic and disulfide

mapping revealed that chenotides contain the same cysteine motif and disulfide connectivity with previous reported 6C-chitin-binding HLPs such as Ar-AMP, Ac-AMP2, IWF4, aSG1 and aSR1 [51, 94, 95, 98]. Phylogenetic analysis showed that chenotides belong to a new class of 6C-chitin-binding HLP based on its unique feature of a tandem repeat that has not been reported. Structural studies showed that chenotides form a compact cystine-knot core, conferring them metabolic stability against acid, heat and enzymatic degradation. Additionally, biological assay showed that chenotide cQ2 is a non-cytotoxic peptide that displayed anti-fungal activity and cell migration properties. Collectively, our study provides insights into the sequence, biosynthesis, and bioactivity and expanded the existing library of HLPs. As such, the bioactive compounds discovered from *C. quinoa* may have therapeutic significance and present a lucrative idea for biotechnological applications.

2. Results

2.1 Isolation and purification of chenotides

Small-scale screenings were performed on the seeds of *Chenopodium quinoa* var. Willd, red and black using MALDI-TOF MS. The MS spectrum showed a cluster of peptides between the molecular weight of 2–4 kDa, suggesting the presence of CRPs (Figure 4.1). The MS profiles of *C. quinoa* var. Willd, red and black contained similar putative peptides, including 2893 Da, 2836 Da and 3151 Da shown in Figure 4.1A, 4.1B and 4.1C. Therefore, for further characterization of the peptides, *C. quinoa* var. Willd was used. To determine the number of cysteines present, the crude extract was S-reduced using dithiothreitol (DTT) and S-alkylated with iodoacetamide (IAA). The S-alkylated peptides of *C. quinoa var*. Willd showed a mass increase of 348 Da (Figure 4.2), indicating that the peptides are CRPs that contained six cysteine residues with three disulfide bonds.

Large scale extraction of the plant material (2 kg) was performed to obtain more peptide quantity. The seeds of *C. quinoa* were extracted in water with a weight to volume ratio of 1:10 for 2 hr. The C_{18} flash column was used to filter the extract to remove colored pigments and other impurities. Fractional elution was done using 20%–80% ethanol. Eluted fractions containing peptide of interest were purified by SCX-HLPC and RP-HPLC. The putative peptides that have been successfully isolated from *C. quinoa var*. Willd were termed chenotides and labeled as cQ1, cQ2 and cQ3 with relative monoisotopic molecular mass [M + H] of 2965.5, 2893.4 and 2836.4 Da, respectively (Figure 4.1A).

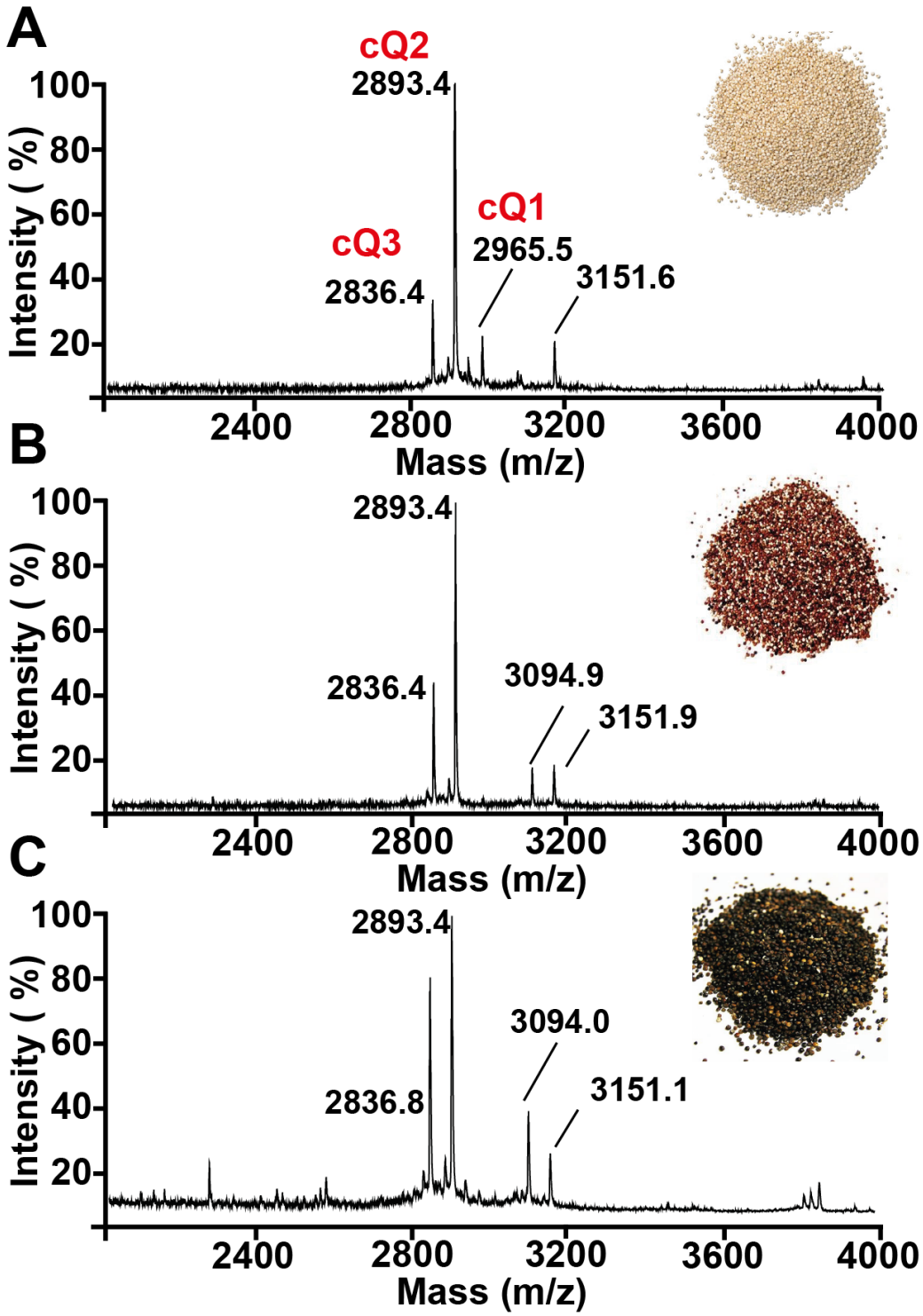


Figure 4.1: MALDI-TOF MS profile of crude extracts. (**A**) *C. quinoa* var. Willd, (**B**) *C. quinoa* var. red (**C**) *C. quinoa* var. black. Clusters of peaks in the range of 2-4 kDa indicate the presence of putative CRPs. Chenotides cQ1, cQ2 and cQ3 are labeled with its corresponding peaks.

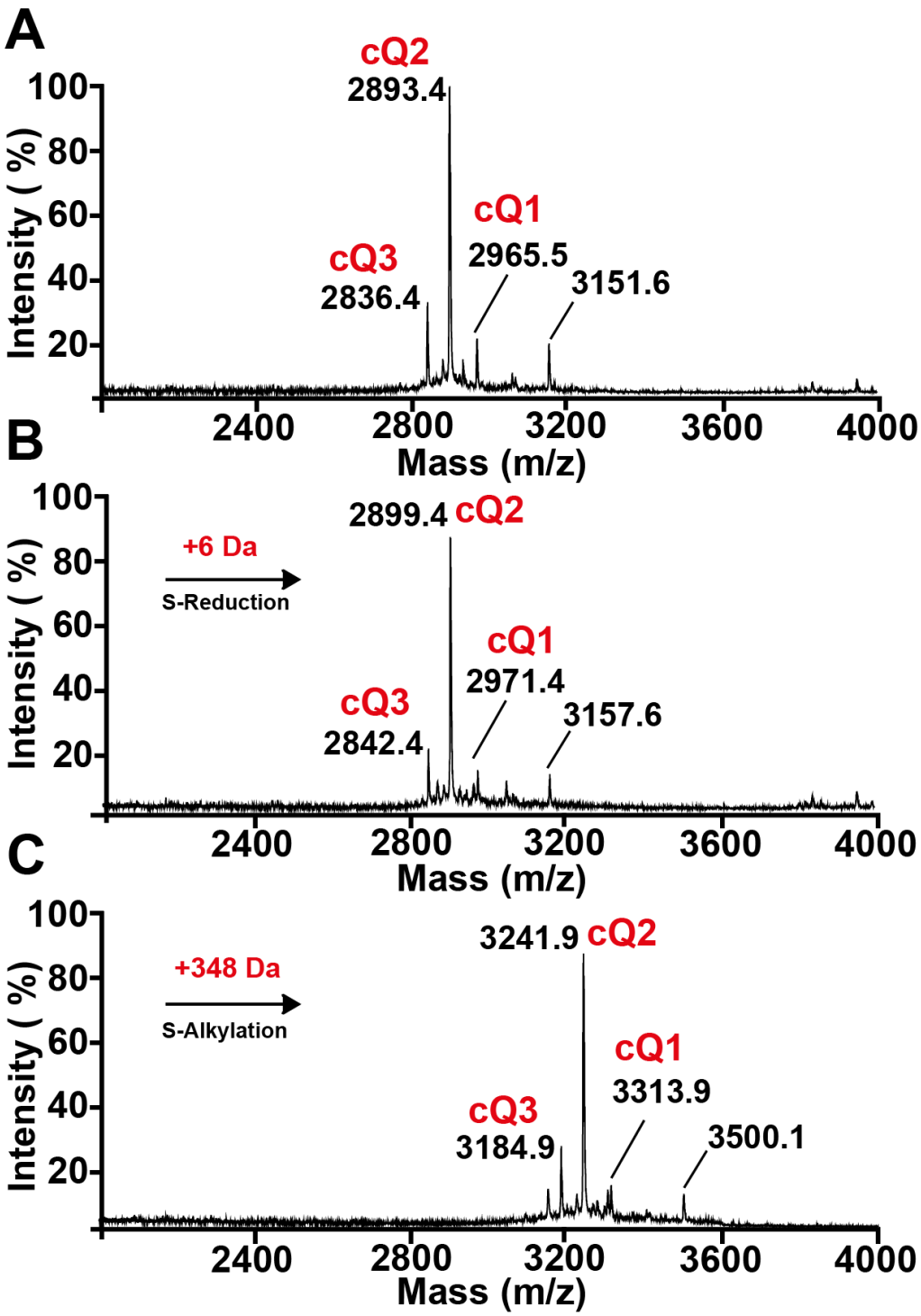


Figure 4.2: MALDI-TOF MS profile of chenotides from *C. quinoa* var. Willd. (**A**) Crude extract containing chenotides cQ1, cQ2 and cQ3 (**B**) S-reduction of crude extract by DTT at 37ºC for 1 h. (**C**) S- alkylation of crude extract by IAA at 37ºC for 1 h.

2.2 *De novo* sequencing of chenotides

De novo sequencing was employed to deduce the sequence of chenotides. The sequences of chenotide cQ1, cQ2, and cQ3 are shown in Figures 4.3, 4.4, and 4.5 respectively. Trypsin digestion of S-reduced cQ1 revealed two distinct fragments, 1695.6 Da and 1293.4 Da (Figure 4.3). The *de novo* sequencing of both fragments revealed that cQ1 contained 31 amino acid residues. Likewise, for cQ2 and cQ3, *de novo* sequencing of both chenotides appeared to be the truncated versions of cQ1. Chenotide cQ2 contained 30 amino acids with an absence of a C-terminal alanine compared to cQ1 whereas cQ3 contained 29 amino acids with the absence of both alanine and glycine at the C-terminal.

The primary challenge of *de novo* sequencing is to differentiate the isobaric residues such as leucine/isoleucine and lysine/glutamine. Therefore, the assigned sequences were validated through tBLASTn using the whole genome sequence database of *C. quinoa* [354]. The assigned chenotide sequences were input as the query sequence, and tBLASTn was performed on the whole-genome shotgun contig (WGS) database of *C. quinoa* (Genbank accession number: LPWI01001647.1). This method not only resolves the ambiguity of the isobaric residues, but it is also able to provide additional information regarding the biosynthesis pathway of chenotides.

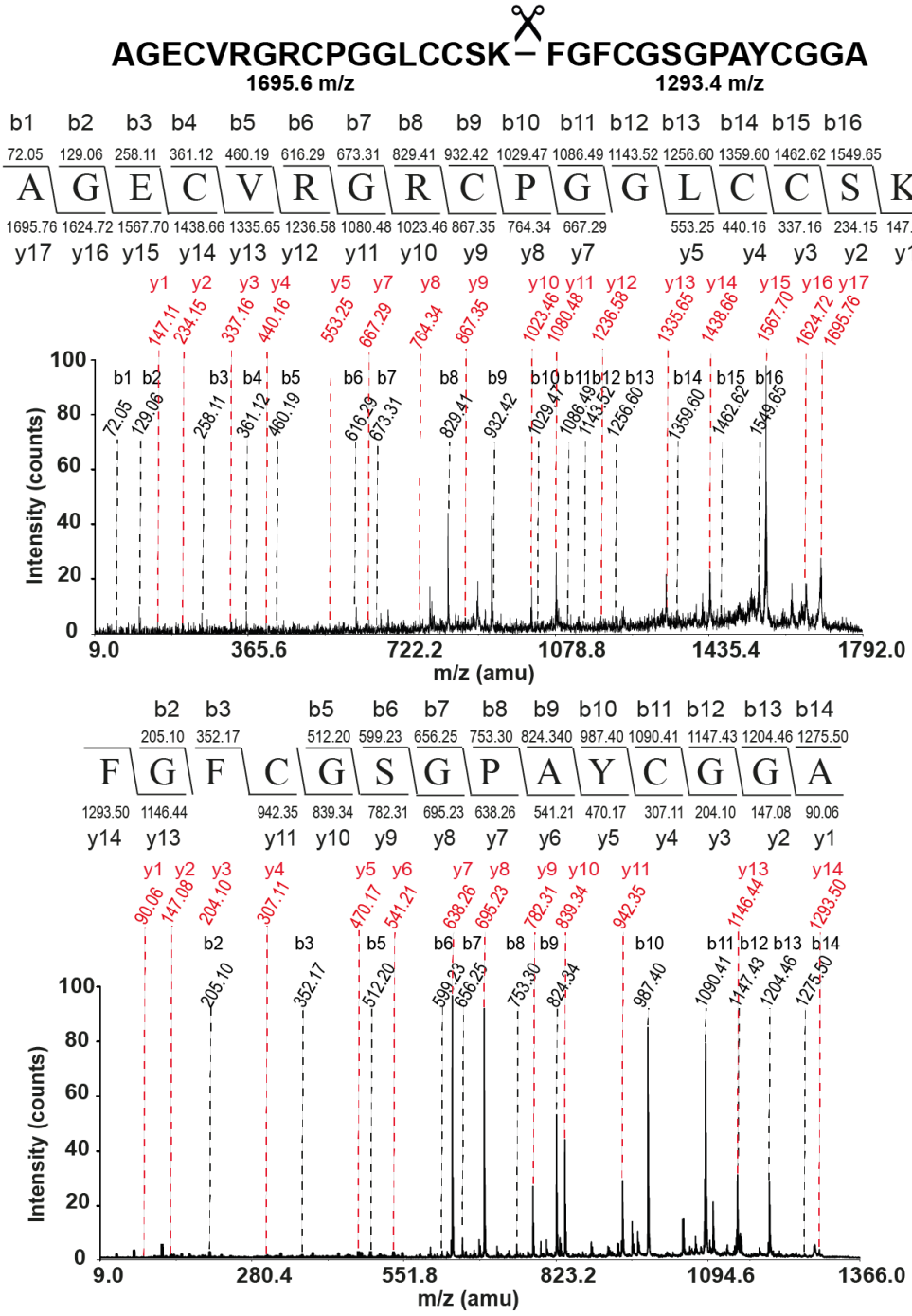


Figure 4.3: *De novo* sequencing of cQ1. Tandem MALDI-TOF TOF MS/MS profiles of two trypsinized fragments of cQ1 (1695.6 m/z and 1293.4 m/z) provided the full sequence of cQ1.

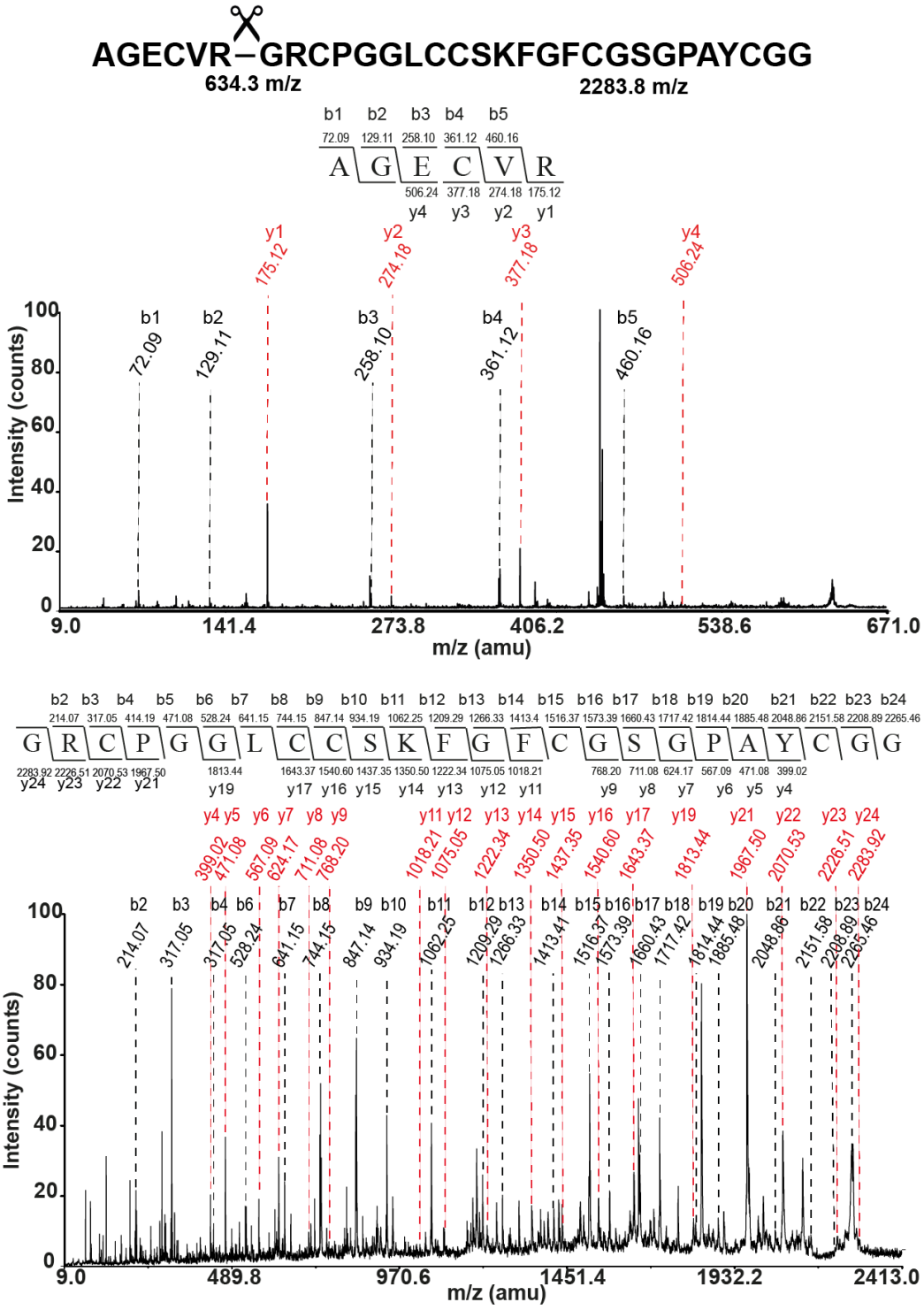


Figure 4.4: *De novo* sequencing of cQ2. Tandem MALDI-TOF TOF MS/MS profiles of two trypsinized fragments of cQ2 (634.3 m/z and 2283.8 m/z) provided the full sequence of cQ2.

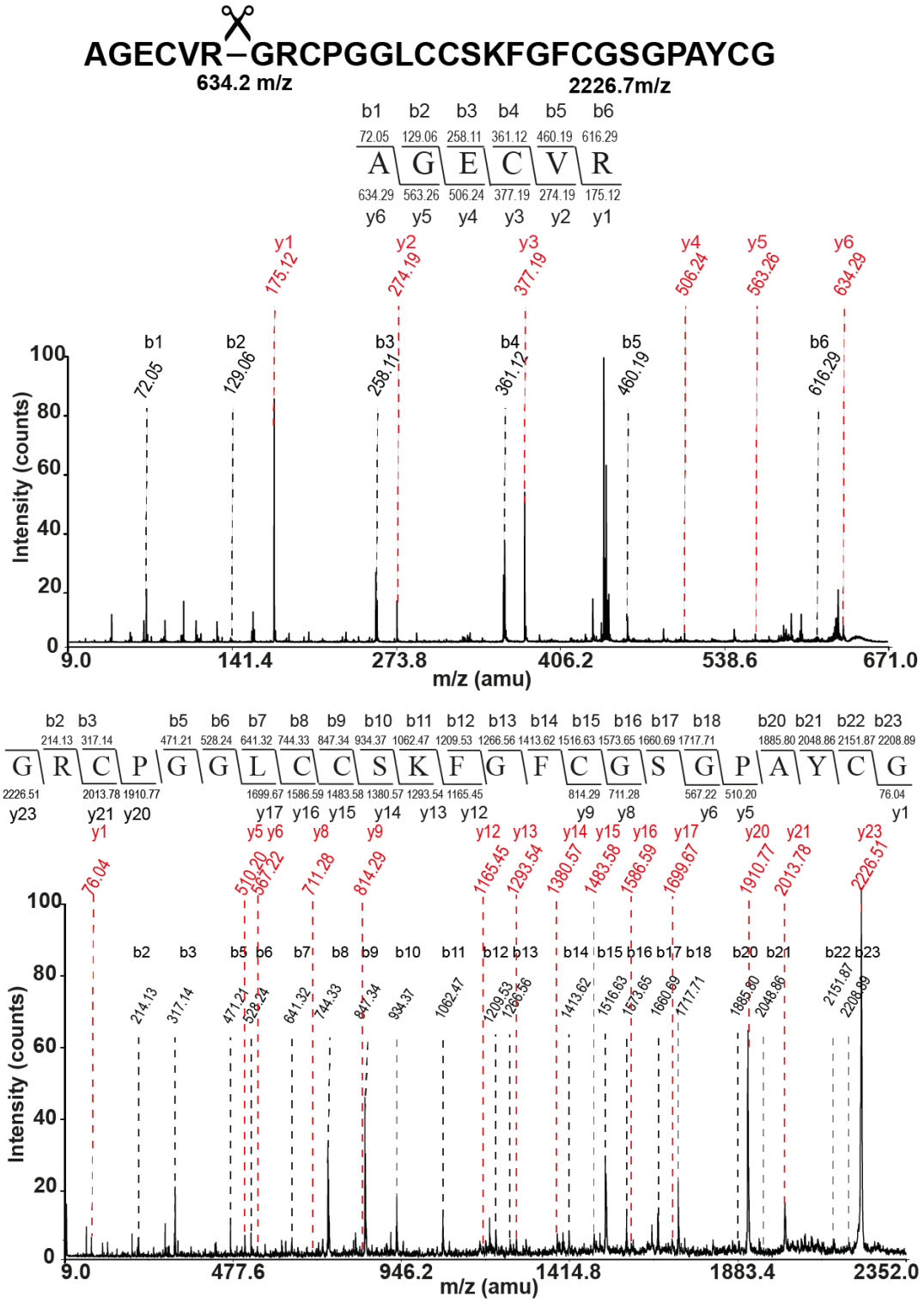


Figure 4.5: *De novo* sequencing of cQ3. Tandem MALDI-TOF TOF MS/MS profiles of two trypsinized fragments of cQ2 (634.2 m/z and 2226.7 m/z) provided the full sequence of cQ3.

2.3 Homology search and sequence alignment

The amino acid sequences of the isolated and purified chenotides were determined by tandem mass spectrometry. Their putative sequences were confirmed by the genomic data of *Chenopodium quinoa* [354]. Table 4.1 summarized the sequence alignment of chenotides cQ1–cQ3 and reported 6C-chitin-binding HLPs. Chenotides residue length ranges from 29 to 31 amino acids and are cysteine- (six residues), glycine- (7–9 residues) and proline-rich (2–3 residues), accounting for approximately 60 % of the primary sequence (Figure 4.6). BLAST analysis associated chenotides to a family of 6C-chitin-binding HLP containing a chitin-binding domain and shared an average sequence similarity of 87%. In addition, chenotide cQ1 shared the highest sequence similarity of 96.4% with aSR1, isolated from *A. sessilis* (Table 4.1).

The cysteine motif pattern $X_{3\text{-}6}$**C**X_6**C**-X_4**CC**-X_5**C**-X_6**C**-$X_{1\text{-}3}$ is highly conserved among chenotides and the reported 6C chitin-binding HLPs. Besides the conserved cysteines, Gly2, Gly7, Gly12, Gly19, Gly22, Gly24, Arg8, Ser16, Pro25 and Tyr27 (numbering based on cQ1) are conserved throughout the 6C chitin-binding HLP subfamily.

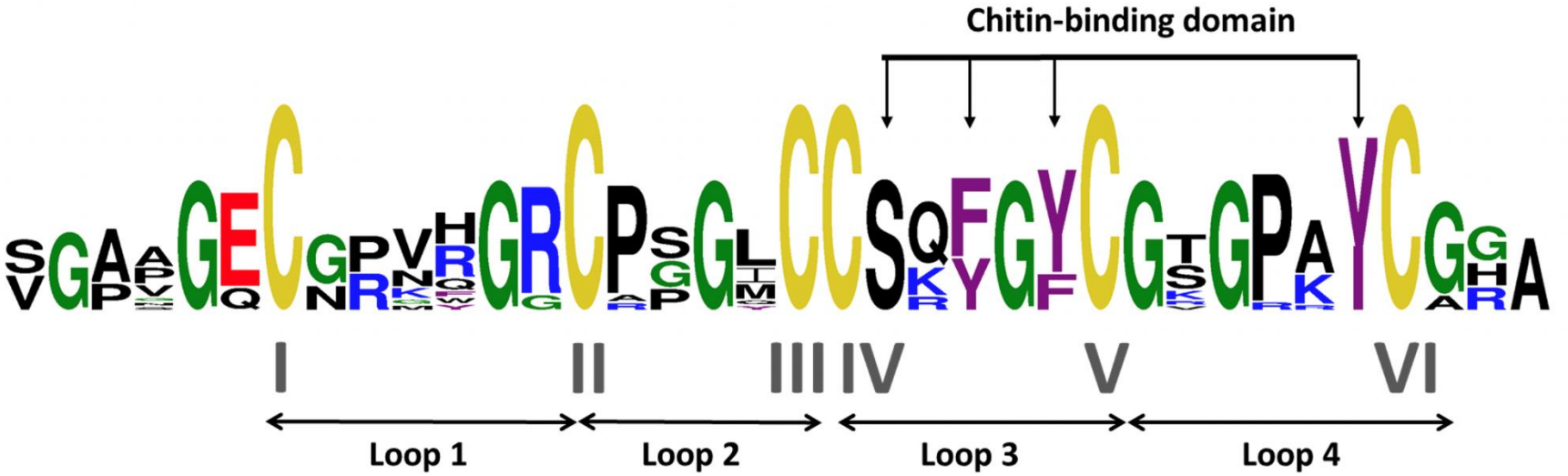


Figure 4.6: Sequence logo of aligned chenotides and reported 6C-chitin-binding-hevein-like peptides. Based on the cysteine pattern, the sequence can be divided into four loops. Loop 3 and 4 contain the conserved chitin-binding domain which is indicated by the arrows.

Table 4.1: Sequence comparison of the primary peptide sequences of chenotides and reported 6C-chitin-binding-hevein-like peptides

Peptide	Species	Amino acid sequence	Mass (Da)[1]	Charge[2]	pI	Method[3]	Similarity (%)	References
Loop		I II III IV						
		* * * *						
cQ1	*C. quinoa*	---AGEC--VRGRCPGGLCCSKFGFCGSGPAYCGGA	2963	+2	8.35	G, P	100	This work
cQ2	*C. quinoa*	---AGEC--VRGRCPGGLCCSKFGFCGSGPAYCGG	2892	+2	8.35	G, P	100	This work
cQ3	*C. quinoa*	---AGEC--VRGRCPGGLCCSKFGFCGSGPAYCG-	2835	+2	8.35	G, P	100	This work
Ac-AMP1	*A. caudatus*	---VGEC--VRGRCPSGMCCSQFGYCGKGPKYCG-	3025	+3	8.66	T, P	89.3	Broekaert et al. 1992
Ac-AMP2	*A. caudatus*	---VGEC--VRGRCPSGMCCSQFGYCGKGPKYCGR	3181	+4	8.92	T, P	89.3	Broekaert et al. 1992
Ar-AMP	*A. retroflexus L*	---AGEC--VQGRCPSGMCCSQFGYCGRGPKYCGR	3153	+3	8.68	T, P	89.7	Lipkin et al. 2005
IWF4	*B. vulgaris L.*	---SGECN-MYGRCPPGYCCSKFGYCGVGRAYCG-	3181	+2	8.27	T, P	80.0	Nielsen et al. 1997
aSG1	*A. sessilis*	--APGQC--NHGRCPSGLCCSQYGYCGTGPAYCG-	3004	+1	7.82	T, P	89.3	Kini et al. 2015
aSG2	*A. sessilis*	---AGEC--NHGRCPSGLCCSQYGYCGTGPRYCG-	2992	+1	7.82	T, P	86.2	Kini et al. 2015
aSG3	*A. sessilis*	--APGQC--NHGRCPSGICCSQYGYCGTGPAYCGG	3060	+1	7.82	T, P	89.7	Kini et al. 2015
aSR1	*A. sessilis*	---VGEC--VQGRCPPGLCCSRFGYCGTGPAYCG-	2932	+1	7.75	T, P	96.4	Kini et al. 2015
aSR2	*A. sessilis*	--APGEC--KHGRCPPGICCSQYGYCGTGPAYCG-	3029	+1	7.81	T, P	89.3	Kini et al. 2015
aSR3	*A. sessilis*	--APGEC--KHGRCPPGICCSQYGYCGTGPAYC--	2973	+1	7.81	T, P	88.9	Kini et al. 2015
SmAMP1.1a	*S. media*	VGPGGECGR-FGGCAGGQCCSRFGFCGSGPKYCAH	3305	+2	8.34	T, P	78.6	R et al. 2012
SmaAMP3	*S. media*	SGPNGQCGPGWGGCRGGLCCSQYGYCGSGPKYCAH	3460	+2	8.29	T, P	75.9	R et al. 2012

[1]Mass (Da): represents the experimentally found molecular weight. [2]Charge: represents the total charge of the molecule, and calculated by the sum of positive (lysine, arginine and histidine residues) and negative (glutamate and aspartate residues) charges. [3]Methods: represents the primary sequences obtained by genomic (G) transcriptomic (T) and/or proteomic (P) method. Cys and chitin-binding domain are highlighted in yellow and green, respectively.

In Figure 4.6, the height of the sequence logo reveals the sequence conservation, whereas the symbol height within the stack indicates the prevalence of each amino acid in that position [355]. As observed, intercysteine loop 1 is highly variable while chitin-binding domain residues reside in loop 3 and 4. These residues are highly conserved in all 6C-CB-HLPs. This data shows that despite the slight differences in the sequences of chenotides from the other 6C-CB-HLPs, they showed some degree of sequence conservation, which suggest that they may evolved from the same ancestor.

To further expand on the library of chenotides, cQ1 was input as the query sequence, and a BLAST search using the whole-genome shotgun database of *C. quinoa* was performed. An additional 12 novel chenotides were identified from the genomic sequence and collectively named cQG1–cQG12. Table 4.2 summarized the sequence alignment of cQ1–cQ3 and cQG1–cQG12. Chenotides identified from the genomic sequence had amino acid lengths that ranged from 28 to 31 amino acids and were also rich in cysteine, glycine and contain a chitin-binding domain (Figure 4.7).

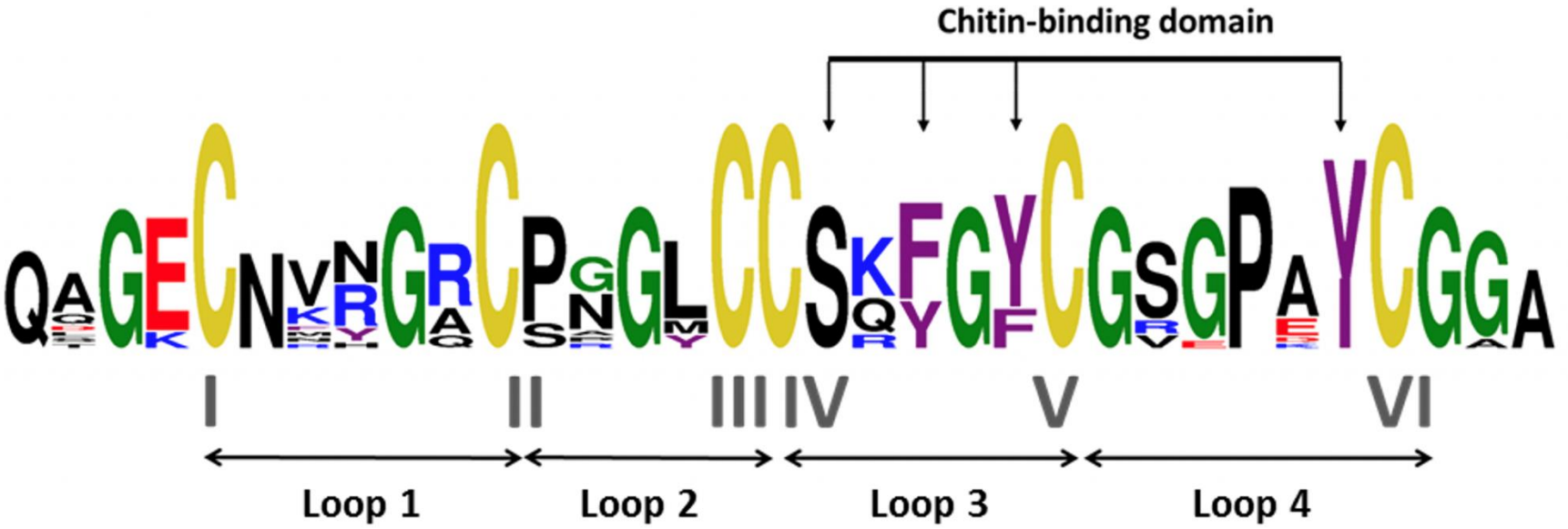


Figure 4.7: Sequence logo of aligned chenotides and genome identified chenotides. Based on the cysteine pattern, the sequence can be divided into four loops. Loop 3 and 4 contain the conserved chitin-binding domain which is indicated by the arrows.

Table 4.2: Sequence comparison of the primary peptide sequences of chenotides and genome identified chenotides

Peptide	Amino acid sequence	No. of amino acid	Mass (Da)[1]	Charge[2]	pI	Method[3]	Similarity (%)
Loop	**I II III IV**						
	* * * *						
cQ1	-AGEC-VRGRCPGGLCCSKFGFCGSGPAYCGGA	31	2963	+2	8.35	G, P	100
cQ2	-AGEC-VRGRCPGGLCCSKFGFCGSGPAYCGG-	30	2892	+2	8.35	G, P	100
cQ3	-AGEC-VRGRCPGGLCCSKFGFCGSGPAYCG--	29	2835	+2	8.35	G, P	100
cQG1	-AGEC-KRGRCSGGLCCSRFGYCGSGPEYCGG-	30	3013	+2	8.34	G	90.0
cQG2	-TGECNKYGRCPPGYCCSKYGYCGVGPKYCGA-	31	3277	+3	8.60	G	79.3
cQG3	-LGEC-KRGRCSGGLCCSKFGYCGSGPEYCGG-	30	3027	+3	8.33	G	89.7
cQG4	-QGEC-VNGACPNGLCCSQFGYCGSGPAYCGG-	30	2895	-1	4.00	G	89.7
cQG5	-DGEC-NRGRCSGGMCCSKFGFCGRGPAYCG--	29	2971	+2	8.34	G	89.3
cQG6	-QGEC-VNGACPNGLCCSRFGYCGSGPAYCGG-	30	2923	0	5.96	G	89.7
cQG7	---EC-FNGACPNGLCCSQFGYCGSEPAYCGG-	28	2830	-2	3.79	G	82.1
cQG8	-SGECNMYGRCPAGYCCSKYGYCGVGPAYCG--	30	3112	+1	7.51	G	83.3
cQG9	QAGKC-VNGQCPNGLCCSQYGYCGSGPDYCGG-	31	3082	0	5.79	G	90.0
cQG10	QAGKC-VNGQCPNGLCCSKYGYCGSGPEYCGG-	31	3096	+1	7.76	G	90.0
cQG11	--GEC-VNGACPNGLCCSQFGYCGSGPAYCGG-	29	2767	-1	4.00	G	89.7
cQG12	-AGEC-RHGRCPRGMCCSQYGYCGRGPAYCG--	29	3091	+3	8.68	G	86.2

[1]Mass (Da): represents the experimentally found molecular weight. [2]Charge: represents the total charge of the molecule, and calculated by the sum of positive (lysine, arginine and histidine residues) and negative (glutamate and aspartate residues) charges. [3]Methods: represents the primary sequences obtained by genomic (G) and/or proteomic (P) method. Cys and chitin-binding domain are highlighted in yellow and green, respectively.

2.4 Disulfide mapping of chenotides

Chenotide cQ2 was selected as a representation of chenotides for further characterization, attributing to its high peptide yield. The disulfide connectivity was deduced by partial reduction of cQ2 with tris (2-carboxyethyl) phosphine (TCEP) to produce intermediates containing either one or two disulfide bonds. These intermediates were then alkylated with N-ethylmaleimide (NEM) under acidic condition (pH 3) to prevent disulfide bonds shuffling and purified using RP-HPLC. MALDI-TOF MS was employed to identify the number of NEM-label cysteine (S-NEM) residue. For each S-NEM labeled cysteine, a mass increase of 126.15 will be observed. Therefore, partially alkylated intermediates that displayed a mass shift of 252.30 or 504.60 Da indicate that it has one or two reduced disulfide bonds, respectively. Figure 4.8B revealed four fractions that were collected from the HPLC. Fraction 1 contained the native cQ2 with three intact disulfide bonds (3S-S). Fraction 2 contained two S-NEM with two intact disulfide bonds (2S-S). Fraction 3 contained four S-NEM and one intact disulfide bonds (1S-S). Lastly, fraction 4 contained six S-NEM with no intact disulfide bonds (0S-S). Fraction 2 and 3 were subsequently fully reduced by DTT and IAA. The positions of the NEM- and IAA-labelled cysteines were identified by MS/MS fragmentation to elucidate the disulfide connectivity of chenotide cQ2. Figure 4.8 showed the results of the disulfide mapping. The 2S-S intermediates displayed the CysIII–CysVI (Figure 4.8C and 4.8D) and CysI–CysIV (Figure 4.8E and 4.8F) disulfide linkages. The 1S-S intermediate displayed both CysI–CysIV and CysIII–CysVI disulfide linkages (Figure 4.8G and 4.8H). Based on the disulfide mapping results, the third disulfide

linkage, CysII–CysV, was deduced. Overall, the disulfide connectivity of cQ2 was revealed as CysI–CysIV, CysII–CysV and CysIII–CysVI (Figure 4.8I)

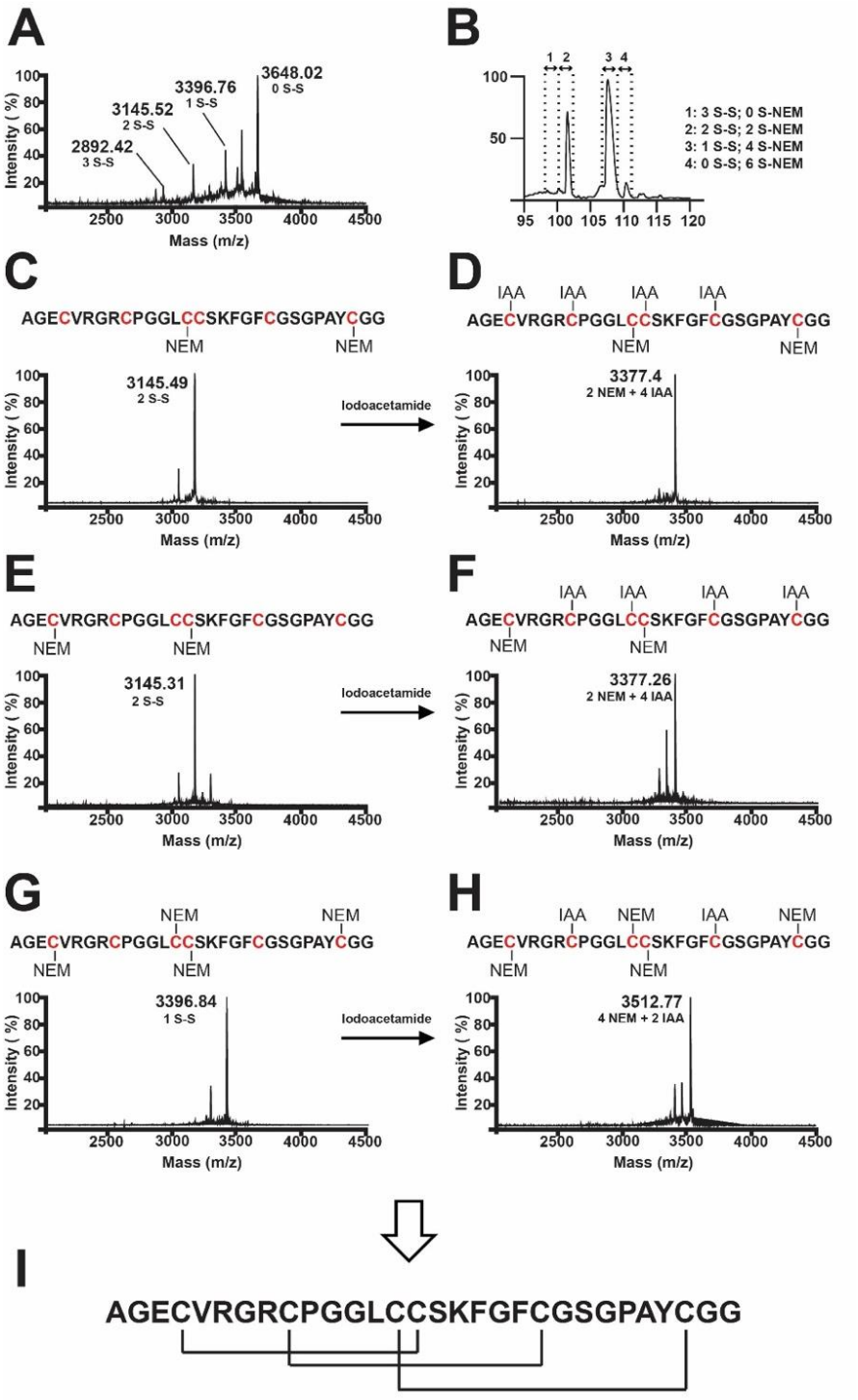


Figure 4.8: Disulfide mapping chenotide cQ2. **(A)** MALDI-TOF MS profile of partially alkylated cQ2. **(B)** HPLC chromatogram of partially alkylated cQ2. The mixture was separated by HPLC and four fractions were collected. **(C and E)** MALDI-TOF MS profile of fraction 2 with two-N-ethylmaleimide alkylated intermediate and **(D and F)** fully reduced species. **(G)** MALDI-TOF MS profile of fraction 3 with four-N-ethylmaleimide alkylated intermediate and **(H)** fully reduced species. **(I)** Overall disulfide connectivity of chenotide cQ2.

2.5 NMR structural study of chenotides

The solution state NMR successfully elucidated the structure of cQ2. (Figure 4.9A). The structure was solved using the distance restraints obtained from 2D NOESY, and the hydrogen bond restraints. All spin-spin systems of cQ2 were identified, and approximately 98% of the proton resonances were unambiguously assigned. The solution structure of cQ2 was resolved based on a total of 187 NMR derived distance restraints and eight hydrogen bonds. Figure 4.9A showed the NMR ensemble of the 20 lowest energy cQ2 structures. The root-mean-square deviation (RMSD) value of the 20 best structures for residues Glu3-Cys9 and Leu13-Try28 was 0.58±0.24 Å and that for all heavy atoms was 1.21 ±0.27 Å (Table 4.3 and 4.4). The structure of cQ2 was well-defined by several number of medium and long-range NOEs, which consisted of two anti-parallel β-strands (β 1: Cys14-Ser16 and β 2: Phe20-Gly22) and a α-helix turn (α1: Gly24-Try28). (Figure 4.9B). The N-terminus of cQ2 has no secondary structure (Ala1-Leu13). PROCHECK analysis suggested that all residues were distributed in the allowed region of the Ramachandran map (Table 4.3). The three disulfide bonds presented in cQ2 followed the classical disulfide connectivity pattern (CysI–CysIV, CysII–CysV, CysIII–CysVI), shown in Figure 4.9B. The disulfide bonds CysI–CysIV and CysII–CysV link the N-terminal loop to β1 and β2 strands respectively. The third disulfide bond CysIII–CysVI links the C-terminal to β1 strand. The acidic residue (Glu3) and basic residues (Arg6 and Arg8) are all located in the N-terminus (Figure 4.9A). The 3D structure of chenotide cQ2 was deposited to Protein Data Bank with an accession number 5ZV6. Figure 4.9C shows the electrostatic surface topology comparison of cQ2 and Ac-AMP2 (PDB: 1MMC). The key residues involved in

chitin-binding are highlighted in purple in Ac-AMP2 (Ser16, Phe18, Tyr20 and Tyr27). The equivalent is also observed in chenotide cQ2 (Ser16, Phe18, Phe20 and Tyr27), revealing that both cQ2 and Ac-AMP2 are structurally similar and contain the chitin-binding domain.

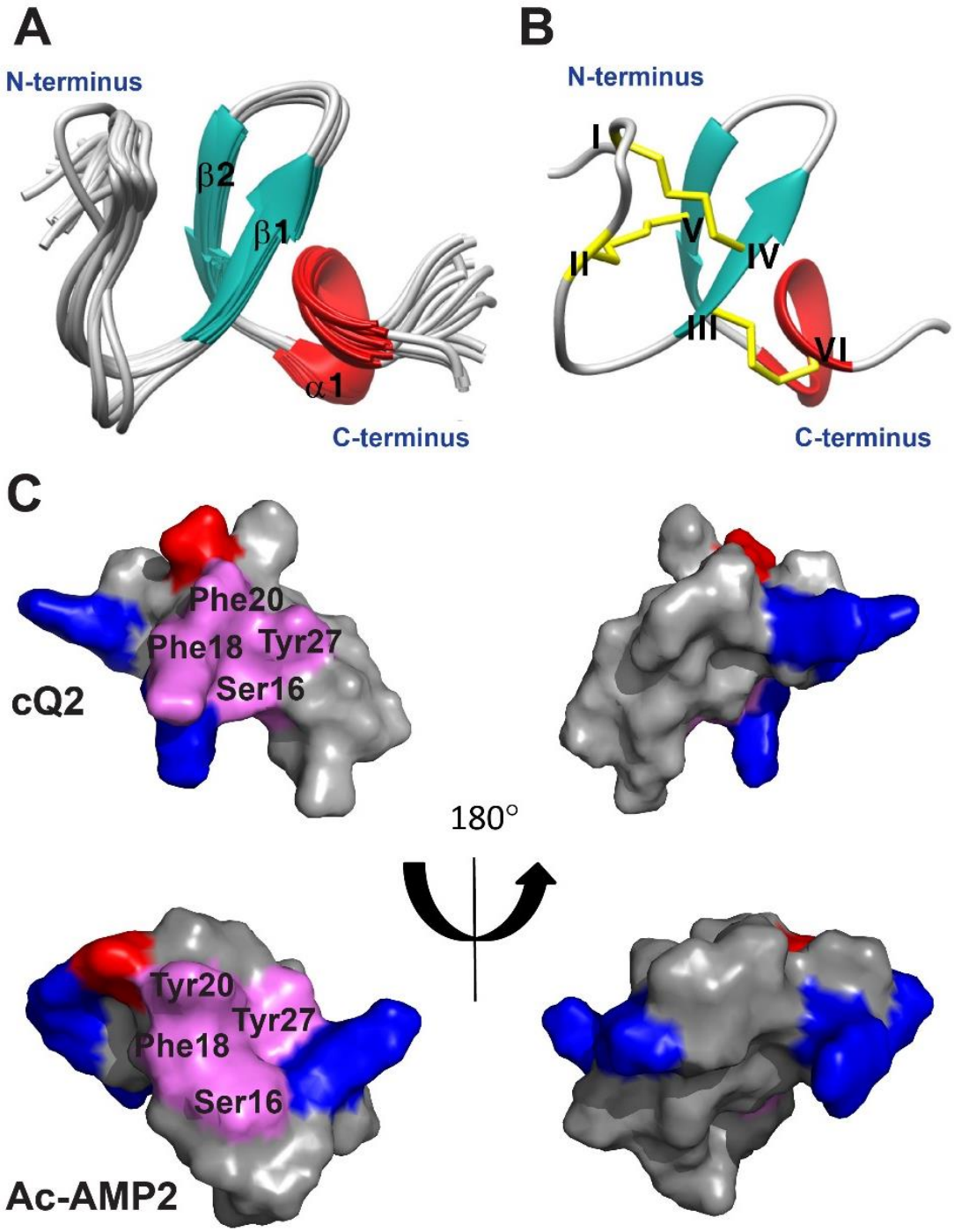


Figure 4.9: Solution NMR structure of cQ2. (A) Superposition of the cQ2 backbone traces from the final 20 ensembles solution structures and restrained energy minimized structure. **(B)** Ribbon representation of cQ2 structure **(C)** Surface topology comparison between cQ2 and Ac-AMP2 (PDB: 1MMC) in flip views showing the chitin-binding domain. The positive, negative and hydrophobic residues are displayed as blue, red and grey respectively. The residues (cQ2: Ser16, Phe18, Phe20 and Tyr27 and Ac-AMP2: Ser16, Phe18, Tyr 20 and Tyr27) were highlighted in purple represents the chitin-binding domain.

Table 4.3: Structural statistics for the final 20 conformers of cQ2[a]

Distance restraints	
Intra-residue (i-j = 0)	70
Sequential ($\|i$-$j\|$ = 1)	68
Medium range ($2 \leq \|i$-$j\| \leq 4$)	13
Long range ($\|i$-$j\| \geq 5$)	28
Hydrogen bond	8
Total	187
Average RMSD to the mean structure (Å)[b]	
Backbone atoms	0.58 ± 0.24
Heavy atoms	1.21 ± 0.27
ϕ/ψ space[c]	
Most favored region (%)	81.4
Additionally allowed region (%)	18.6
Generously allowed region (%)	0.0
Disallowed region (%)	0.0
RMSD from covalent geometry	
Bonds (Å)	0.0061 ± 0.0002
Angles (deg.)	0.540 ± 0.0319
Impropers (deg.)	0.307 ± 0.021
RMSD from experimental restraints	
NOEs (Å)	0.023 ± 0.0069

[a] Selected from 100 calculated conformers according to overall energy.

[b] Calculated with MOLMOL using range 3-9, 13-28.

[c] Calculated with PROCHECK-NMR.

Table 4.4: Proton chemical shift assignments for each amino acid residues of cQ2

	HN (ppm)	Hα (ppm)	Hβ (ppm)		Others (ppm)
A1		4.170	1.562		
G2	8.889	4.018, 4.139			
E3	8.503	4.462	1.898	2.048	Hγ, 2.288, 2.335
C4	7.714	4.373	2.792	3.143	
V5	8.620	4.117	1.892		Mγ, 0.863, 0.880
R6	9.474	3.838	1.582		Hγ, 1.811, 1.936
G7	8.349	3.732, 4.080			
R8	7.820	4.713	1.817	1.886	Hγ, 1.572, 1650
C9	8.708	5.141	2.485	2.791	
P10		4.603	2.294		Hγ, 1.875; Hδ, 3.886, 3.520
G11	8.492	3.757, 3.824			
G12	8.679	3.758, 4.043			
L13	7.433	4.351	1.316	1.577	Hγ, 1.701; Mδ, 0.602, 0.699
C14	9.074	4.480	2.559	3.767	
C15	8.751	4.808	2.809	2.926	
S16	9.848	4.903	4.293	4.348	
K17	8.981	4.043	1.494	1.572	Hγ, 1.096; Hδ, 0.741, 1.106
F18	7.434	4.716	2.890	3.603	Hδ, 7.1075; He, 7.360
G19	7.954	3.561, 3.952			
F20	7.524	5.195	3.164	2.586	Hδ, 7.075; He, 7.357
C21	8.819	5.609	2.667	2.820	
G22	8.798	3.588, 1.938			
S23	8.025	5.116	3.698	3.776	
G24	8.322	3.976, 4.560			
P25		4.258	2.099		Hδ, 3.812
A26	8.759	4.125	1.293		
Y27	7.671	4.125	2.473	2.855	Hδ, 7.215; He, 6.710
C28	8.528	4.566	2.768	3.285	
G29	8.008	3.959, 4.085			
G30	8.075	3.889			

2.6 Putative biosynthesis of chenotides

The whole-genome shotgun database of *C. quinoa* provided information about the putative biosynthesis pathway of cQ1. Figure 4.10A shows the full-length sequence of cQ1 and chitin-binding HLPs such as Ac-AMP2 [94], Ar-AMP [98], IWF4 [95], aSG1 and aSG2 [51], aSR1 [51], hevein [89] and Ee-CBP [101]. These precursors shared the same three-domain architecture which comprises an ER signal peptide, a mature domain and a C-terminal tail. This precursor arrangement suggests that cQ1 is a secretory peptide similar to other chitin-binding HLP subfamilies. The signal peptide is cleaved by a signal peptidase (SPaseI), while the C-terminal tail and hinge region is cleaved by an endopeptidase to release the mature domain. The sequences of signal peptide of cQ1 is shown in Figure 4.10A. Compared to the other chitin-binding HLPs, cQ1 has the longest signal peptide comprising of 26 amino residues. The C-terminal of all the precursor signal peptides end with either Ala or Gly. Notably, the C-terminal tail of hevein and Ee-CBP were significantly longer (142–200 amino acid residues) compared to the others because they contain a protein cargo or catalytic domain such as Barwin-like protein or chitinases.

A distinguishing feature of cQ1 was the presence of a pro-mature domain which was not observed in the rest of the 6C, 8C, and 10C-chitin-binding HLPs (Figure 3.10A). The pro-mature domain contained a tandem-repeated sequence. Hence, a comparison was made between cQ1 and chitin-binding proteins such as UDA1 [356] and WGA [357] isolated from *Urtica dioica* and *Triticum aestivum*, respectively (Figure 4.10B). Similarly, the biosynthesis of these chitin-binding proteins followed a three-domain architecture of an ER signal peptide, tandem-repeated mature

domains and a C-terminal tail. Interestingly, cQ1 comprises a tandem repeat with identical sequence whereas UDA1 and WGA comprise two and four tandem repeats, respectively, with each repeat having a different sequence. Unlike UDA1 and WGA, the hinge region in cQ1 is cleaved off during post-translational processing to give rise to two peptides of the same molecular weight (2963 Da), whereas the hinge region of both UDA1 and WGA remained intact. As a result, this differentiates cQ1 from the chitin-binding proteins. Together, this data showed that cQ1 is the first 6C-chitin-binding HLP containing a tandem repeat that has been identified thus far.

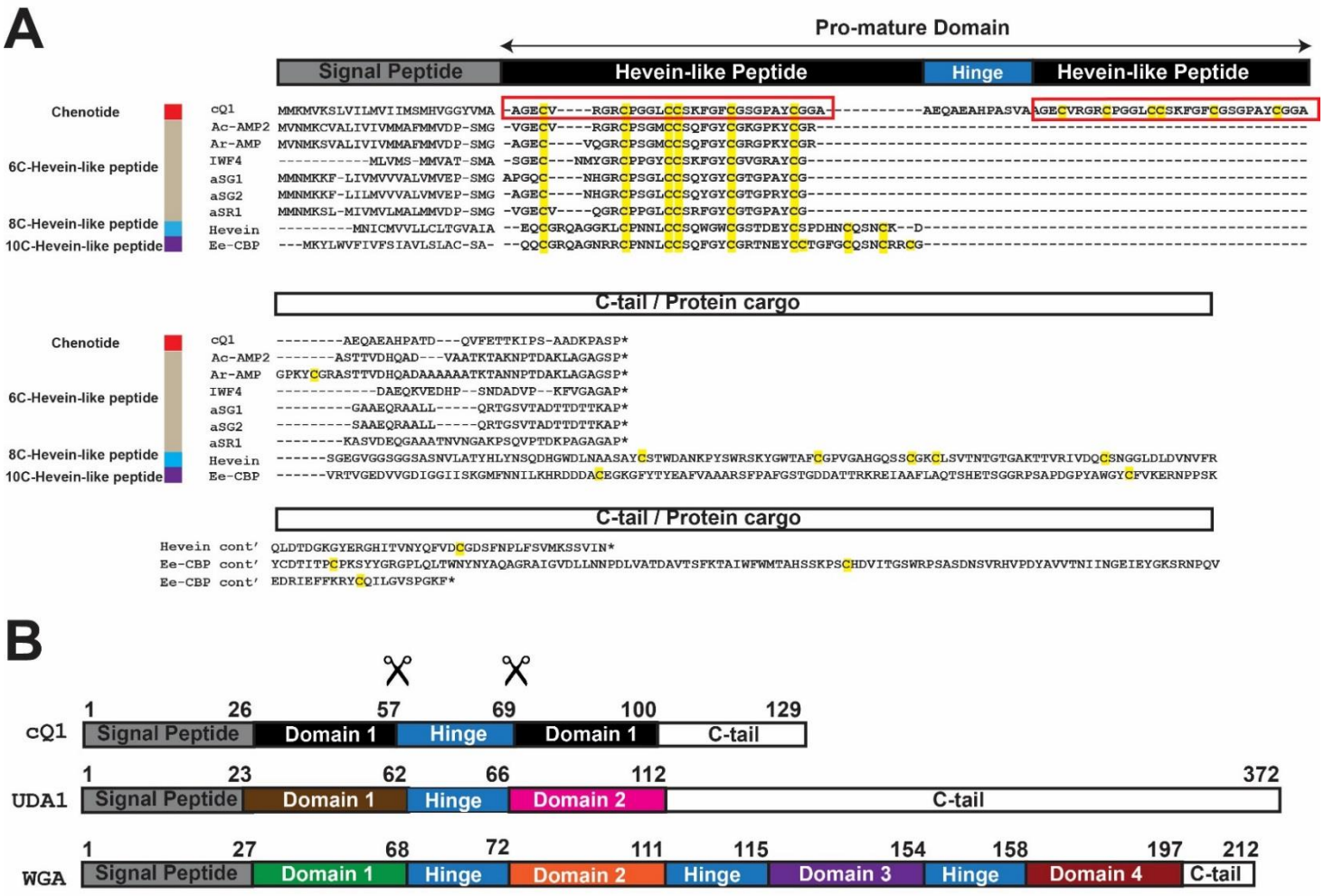


Figure 4.10: Multiple gene alignments of tentative full-length coding sequence of chenotide and chitin-binding hevein-like peptides. (**A**) Precursor sequence of chenotide aligned with reported chitin-binding hevein-like peptides. Chenotide precursor contained a three-domain architecture including a signal peptide, pro-mature domain and C-terminal tail. In 8C- and 10C-hevein-like peptide, the C-terminal tail codes for a protein cargo such as Barwin-like protein or chitinase. (**B**) Schematic comparison of chenotide with chitin-binding proteins UDA1 and WGA.

2.7 Distribution and data-mining of chenotides and 6C-hevein-like peptides

A tBLASTn search using cQ1 mature peptide as a query was performed to determine the occurrence and distribution of chenotides *in planta.* The results revealed 137 sequence homologs from a diversity of plants with an E-value of < 0.0005. Additionally, manual filtering was applied to remove sequences consisting of (1) a short signal peptide of less than ten residues, (2) an odd number of cysteine residues and (3) identical sequence from the same species. A total of 64 chenotide-like peptide homologs from 34 different plant species were identified. A circular phylogenetic tree was constructed to compare the similarities between chenotides, chenotide-like homologs and other 6C-CRPs. Their aligned sequences were calculated with the neighbor-joining clustering algorithm shown in Figure 4.11. Based on Figure 4.11, chenotide-like peptides were detected only in the plant kingdom.

For chenotide-like peptide homologs, three putative mature peptides: aL1, pM1 and aS6 were expressed in multiple plants with identical mature domain sequences but varied in the signal peptides and/or C-terminal tail. For example, aS6 was expressed in eight plants: *Spinacia oleracea*, *Alternanthera sessilis*, *Atriplex hortensis*, *Atriplex prostrata*, *Atriplex rosea*, *Beta maritima*, *Chamaseyce mesebyranthemum*, *Heliotropium calcicola*. The expression of chenotides in different families is consistent with natural product occurrence, meaning the same product can be found in different species of the plant family [358]. The data mining results also revealed two major clusters: (1) chenotides and chenotide-like homolog (highlighted in turquoise) and the 6C-chitin-binding HLPs (highlighted in salmon)

were separated from (2) the other 6C-CRPs (Figure 4.11A). Chenotides and chenotide-like homologs are found only in angiosperms, and they are distributed in ten families, including *Amaranthaceae, Caryophyllaceae, Chenopodiaceae, Dermateaceae, Euphorbiaceous, Heliotropiacea, Linaceae, Portulacaceae, Selaginellaceae* and *Taxaceae.* On the whole, data mining of the transcriptomic data and phylogenetic analysis revealed that chenotides represent a new class of 6C-chitin-binding HLPs, which are mainly distributed in the angiosperms.

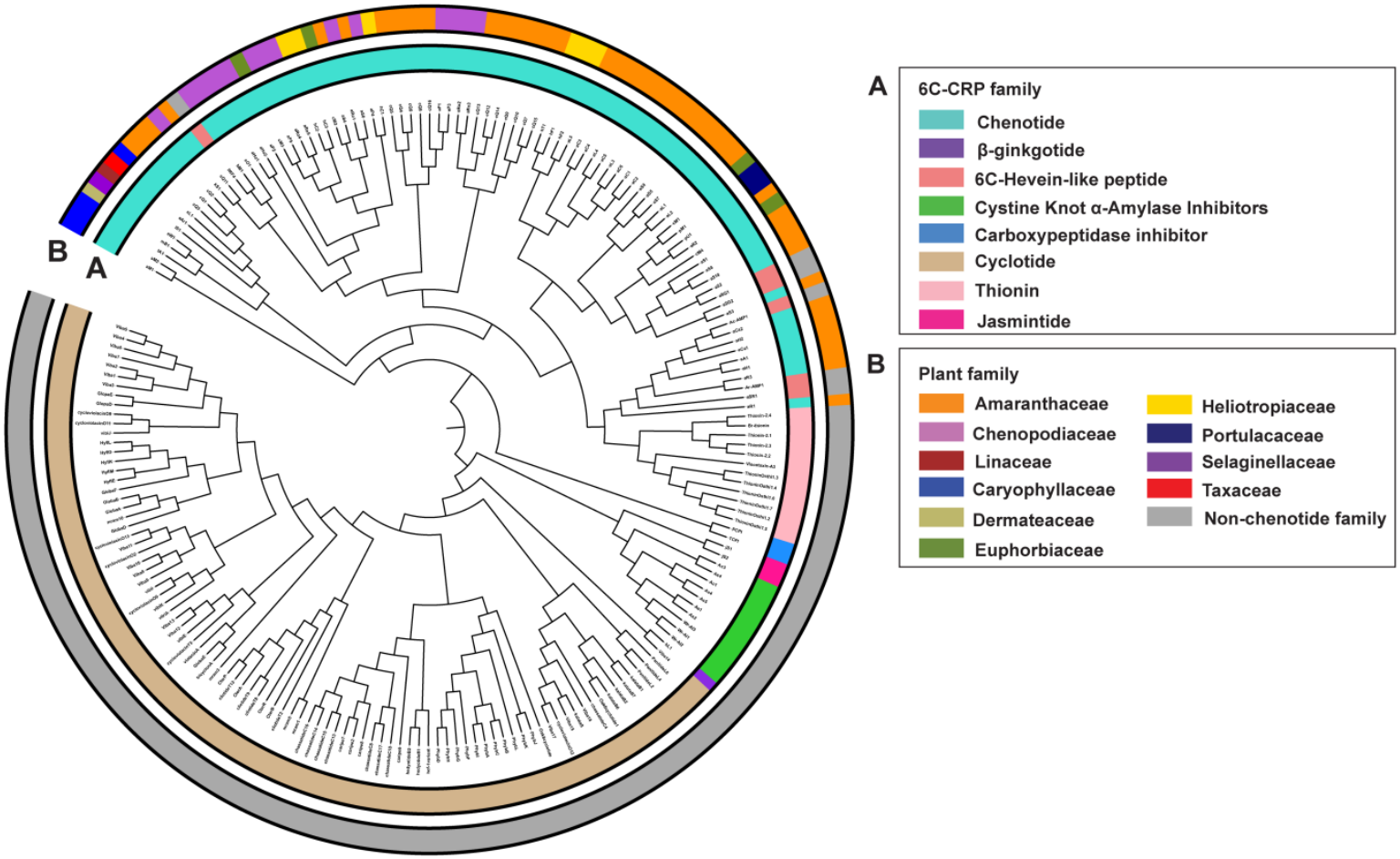


Figure 4.11: Distribution of chenotides and other 6C-cysteine-rich peptides in the plant kingdom. The precursor sequences were aligned by MUSCLE and the phylogenetic tree was generated by iTOL. **(A)** 6C-cysteine-rich peptide family and **(B)** plant's family.

2.8 Chitin-binding activity of chenotides

Both native and S-alkylated chenotide cQ2 were employed to study the chitin-binding properties of chenotides. Purified native and S-alkylated cQ2 (5 µg) were incubated with chitin beads. The UPLC chromatograph revealed that the S-alkylated cQ2 was detected in the supernatant, as shown in Figure 4.12A, suggesting that S-alkylated cQ2 did not bind to the chitin beads. This indicated that the tertiary structure is essential for chitin binding interaction. In contrast, the native cQ2 bound to the chitin beads within 15 min (Figure 4.12B). To optimize the elution of cQ2 from the chitin beads, salt, acid and heat elution at varying concentrations and temperatures were optimized. cQ2 was unable to elute from the chitin beads up to 1 M NaCl. The best elution condition observed was 1 M acetic acid at 55 °C (Figure 4.12C). This indicated that cQ2 has strong chitin binding affinity, and strong acid with heat is required to elute the peptide from the beads.

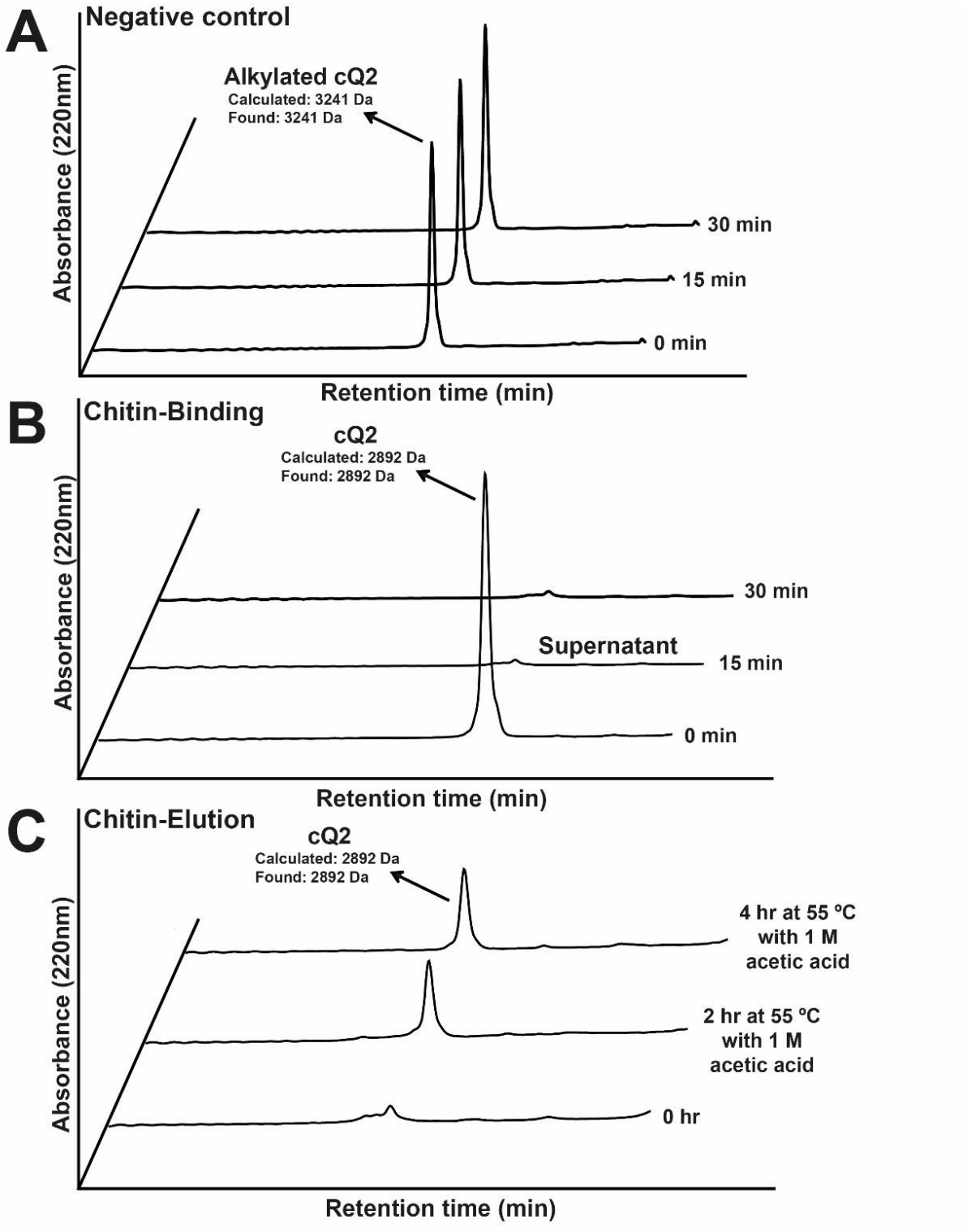


Figure 4.12: Chitin-binding activity of cQ2. (A) Negative control of S-reduced and S-alkylated cQ2 **(B)** native cQ2 chitin-binding profile **(C)** native cQ2 chitin elution profile. The absence of native cQ2 in the supernatant suggests cQ2 have chitin-binding properties.

2.9 Anti-fungal properties of chenotides

To determine the anti-fungal properties of chenotides, the disc diffusion assay was employed. Four phyto-pathogenic fungal strains, namely, *Alternaria alternata*, *Curvularia lunata*, *Fusarium oxysporum* and *Rhizoctonia solani* were utilized. The fungi were grown at 25 °C for 24–48 h until a radial colony was observed. Paper discs (6 mm) with varying concentrations of cQ2 (1, 5 and 10 mg/mL) were placed at the growing ends of the mycelial and incubated for additional 24 h at 25 °C. The presence of crescent-shaped inhibition zones indicates the fungal susceptibility towards cQ2. Based on Figure 4.13, it was observed that cQ2 at 10 mg/mL concentration has the highest inhibition efficiency in all four fungal strains, suggesting that cQ2 inhibits fungal growth by preventing hyphae budding in a dose-dependent fashion.

Since the susceptibility of fungal strains to chenotide cQ2 has been established, the micro-broth dilution assay was utilized to establish the half-maximal inhibitory concentration (IC_{50}). To evaluate the IC_{50} of chenotide against the phyto-pathogenic fungal strains, 1 x 10^5 fungal spores were incubated on the microtiter plate with cQ2 at varying concentrations at 25 °C for 24 h. The cells were subsequently fixed and stained with crystal violet and absorbance was read at 570 nm. The IC_{50} was calculated based on the dose-response curve plotted by peptide concentration against the percentage of inhibition (Figure 4.13). Chenotide cQ2 displayed strong potency against *F. oxysporum* with a remarkable IC_{50} of 0.31 µM. The IC_{50} against *Rhizoctonia solani*, *Alternaria alternata* and *Curvularia lunata* ranged from 2.98 µM to 9.51 µM.

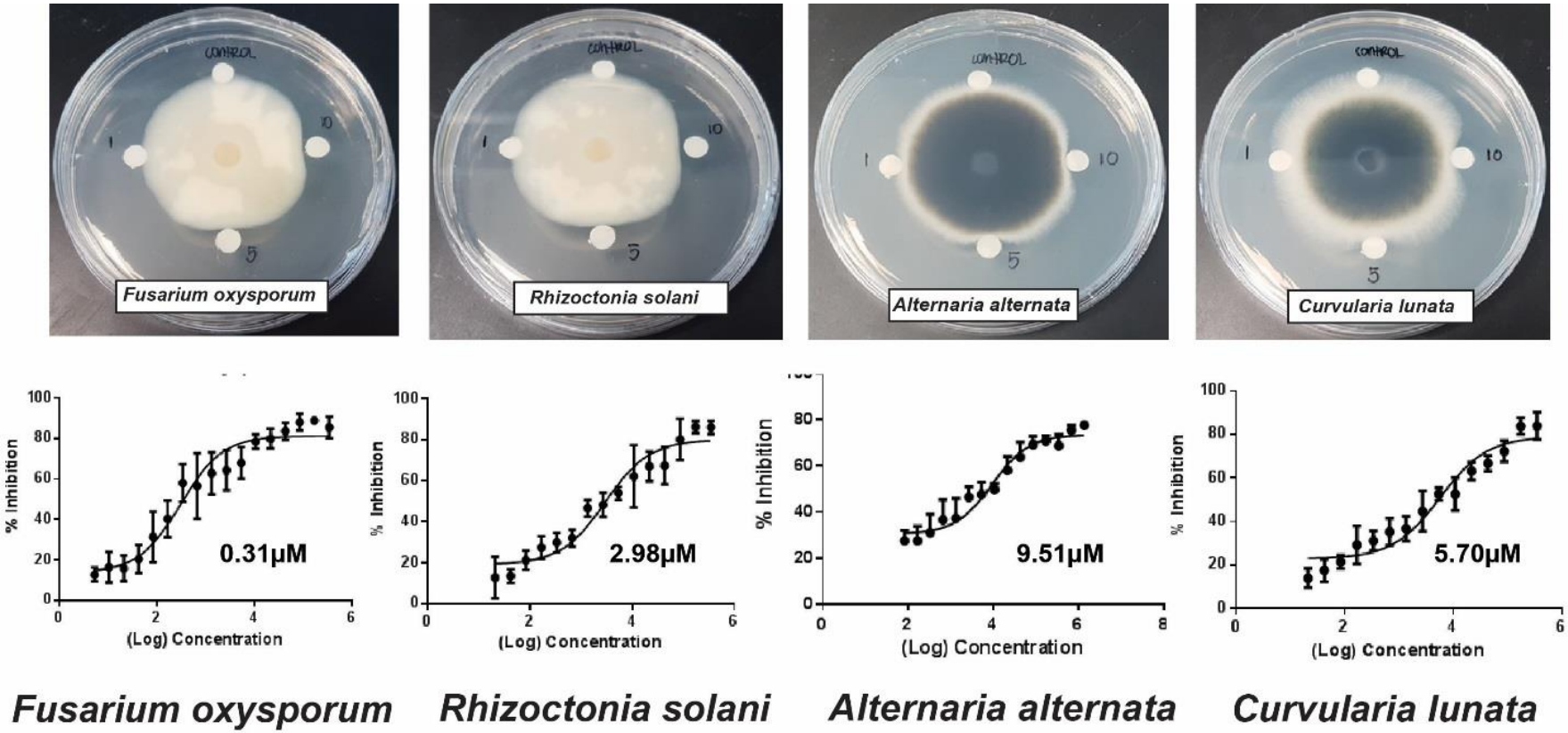


Figure 4.13: Anti-fungal activity of chenotide. Fungal inhibition of cQ2 against *Fusarium oxysporum*, *Rhizoctonia solani*, *Alternaria alternata* and *Curvularia lunata*. Formation of arc-shaped inhibition zones in the disc diffusion assay indicated susceptibility of fungi towards chenotide. The IC_{50} was calculated based on the dose-response curve obtained from the micro-broth dilution assay.

To further understand the putative mechanism of the anti-fungal activity of chenotides, the bright field microscopy was employed to detect for morphological changes in the hyphal growth. Varying concentrations of cQ2 (0.3, 0.03 and 0.003 mg/mL) were used to treat the four phyto-pathogenic fungal strains. According to Figure 4.14, there was significant retardation in the hyphal growth of all four phyto-pathogenic fungal strains treated with cQ2. Compared to the untreated control, fungal strains treated with cQ2 had swollen hyphal tips, growth retardation and short branching. Therefore, these results showed that cQ2 is an anti-fungal peptide that impedes phyto-pathogenic fungi' growth by preventing hyphal elongation at the growing tip of the mycelia.

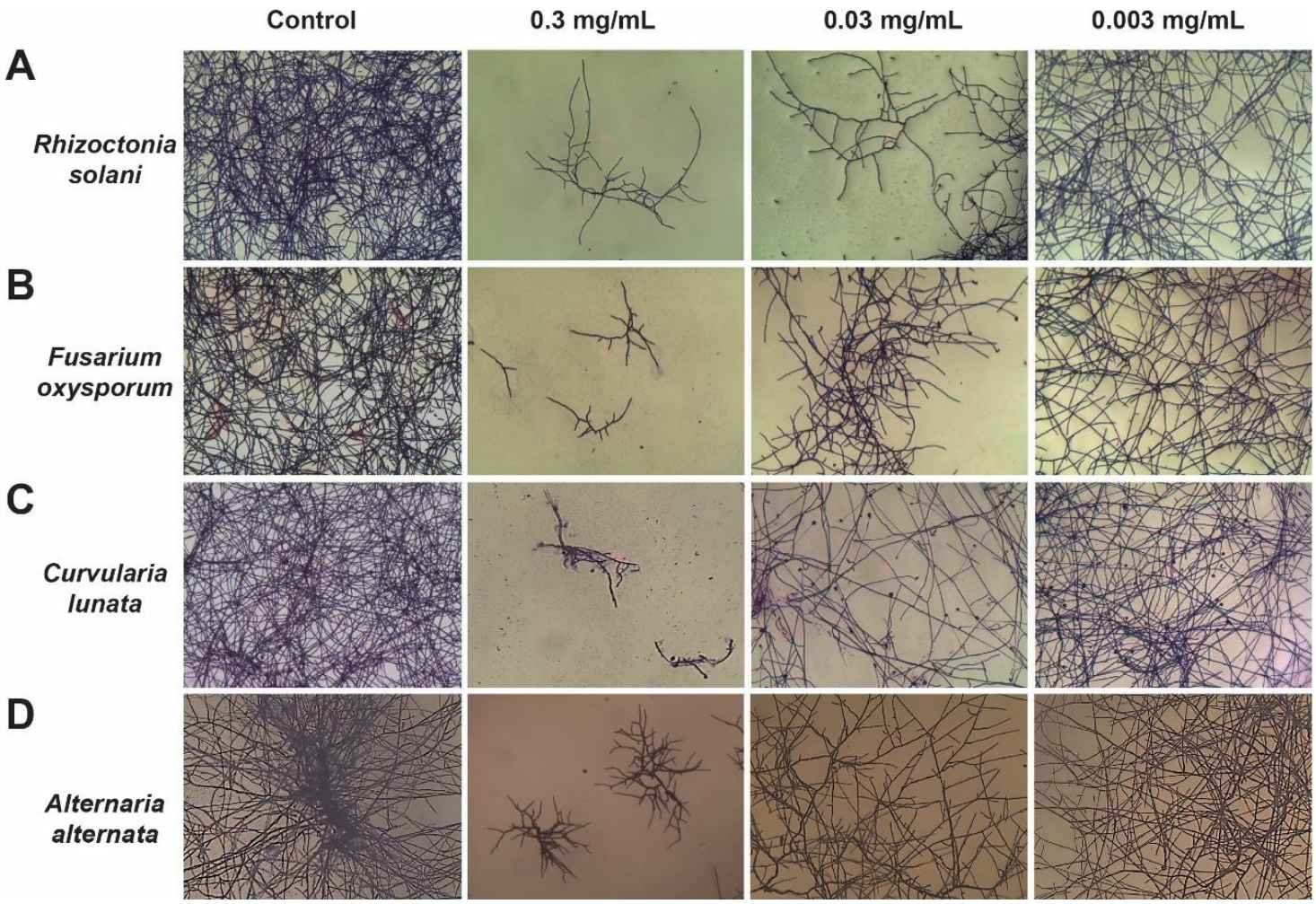


Figure 4.14: Bright field microscopy of hyphal growth inhibition with cQ2. **(A)** *Rhizoctonia solani*, **(B)** *Fusarium oxysporum*, **(C)** *Curvularia lunata* and **(D)** *Alternaria alternate* treated with 0.003–0.3 mg/mL of cQ2. Formation of stunted hyphae ends indicated that chenotide inhibits hyphal growth at the ends of the fungal mycelial.

2.10 Metabolic stability of chenotides

The hallmark of an orally active drug is the ability to withstand adverse conditions such as thermal, acid and enzymatic degradation. To assess the stability of chenotide cQ2 under heat and acidic conditions, the peptide was subjected to 95 °C or 1 M HCl for two hours. For stability towards enzymatic degradation, cQ2 was subjected to either pepsin or aminopeptidase I for 6 h. Lastly, for serum stability, cQ2 was subjected to horse or human serum for 48 h and 36 h respectively. The presence of the peptide and the extent of the peptide degradation were evaluated using MALDI-TOF MS and UPLC chromatographs. Based on the result shown in Figure 4.15, cQ2 remained relatively stable to heat, acid and proteolytic degradation, retaining approximately 80% of the peptide after each treatment. Using the S-alkylated cQ2 as a control, it was observed that the S-alkylated cQ2 had degraded over the time course of each treatment. The average half-life of S-alkylated cQ2 for all treatments is three hours. In contrast, the average percentage of native cQ2 remaining at t_{120} timepoint for all treatments is 85.3%. These results strongly suggested that the presence of the highly crossed-linked disulfide bonds in the native cQ2 prevented the hydrolysis of the peptide, conferring high metabolic stability. Together, this data provided evidence that chenotide is stable against heat, acidic and proteolytic degradation, and further suggests that the scaffold of chenotide may be a suitable candidate to engineer orally active peptidyl biotherapeutics.

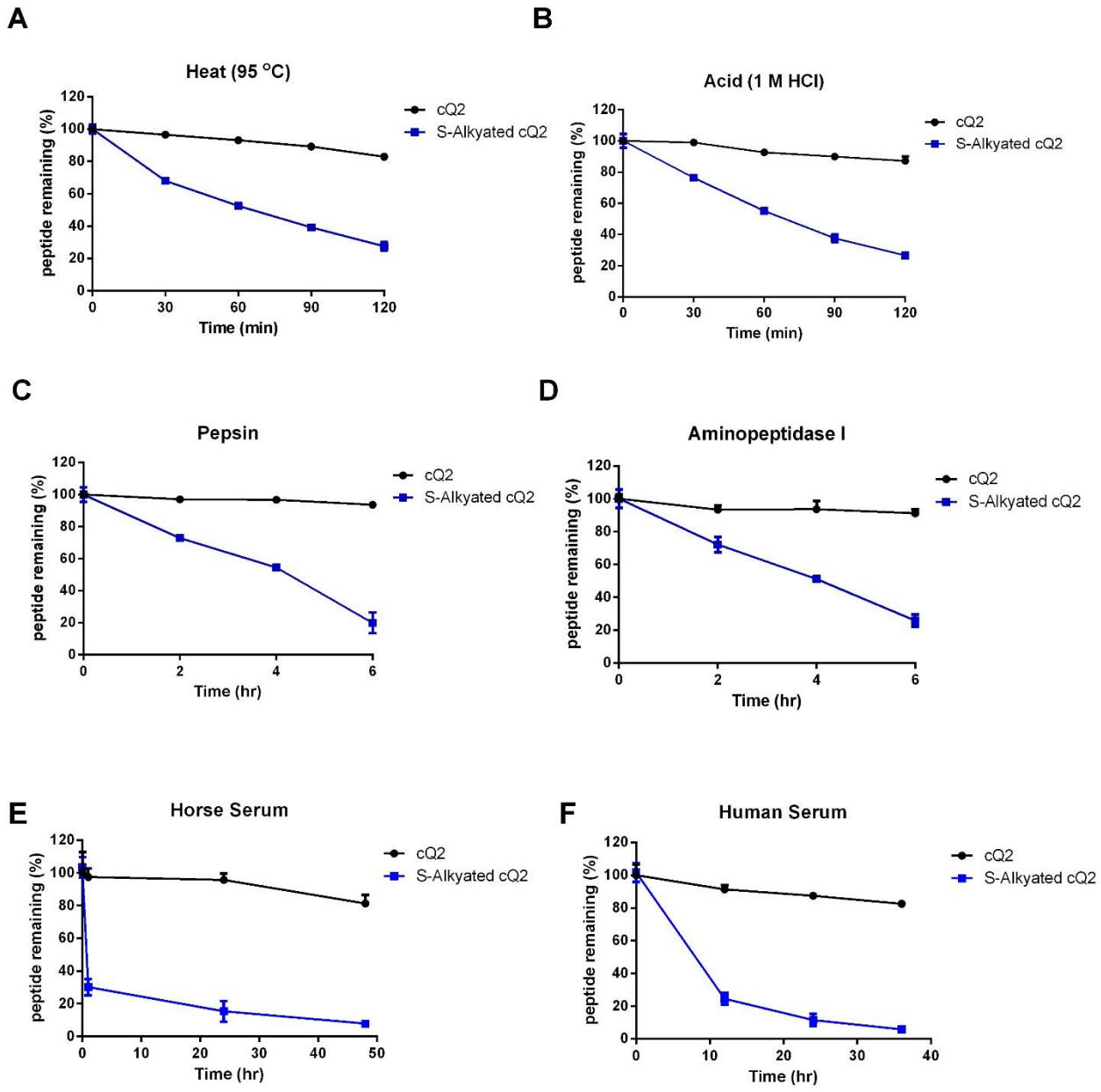


Figure 4.15: Stability test of chenotide. **(A)** Heat stability of cQ2 at 95 °C for 2 h **(B)** Acid stability of cQ2 incubated in 1 M HCl for 2 h. **(C)** Pepsin stability of cQ2 incubated for 6 h. **(D)** Aminopeptidase I stability of cQ2 incubated for 6 h **(E)** Horse serum stability of cQ2 incubated for 48 h **(F)** Human serum stability of cQ2 incubated for 36 h. The percentages of peptide remaining were calculated from the area under the graph of the UPLC chromatographs.

2.11 Cytotoxic assay of chenotide

To determine the cytotoxicity effect of chenotide, three cell lines namely, H9c2, HUVEC and SH-SY5Y were utilized for this assay. Each cell line was treated with different concentrations of cQ2 (1.25, 2.5, 5, 10, 20, 40, 80 µM) and incubated for 24 h. MTT was then added, and the mixture was further incubated 4 h at 37 °C to assess the cell viability. Based on the results shown, cQ2 did not exhibit any cytotoxicity effect on all three cell lines up to 80 µM (Figure 4.16).

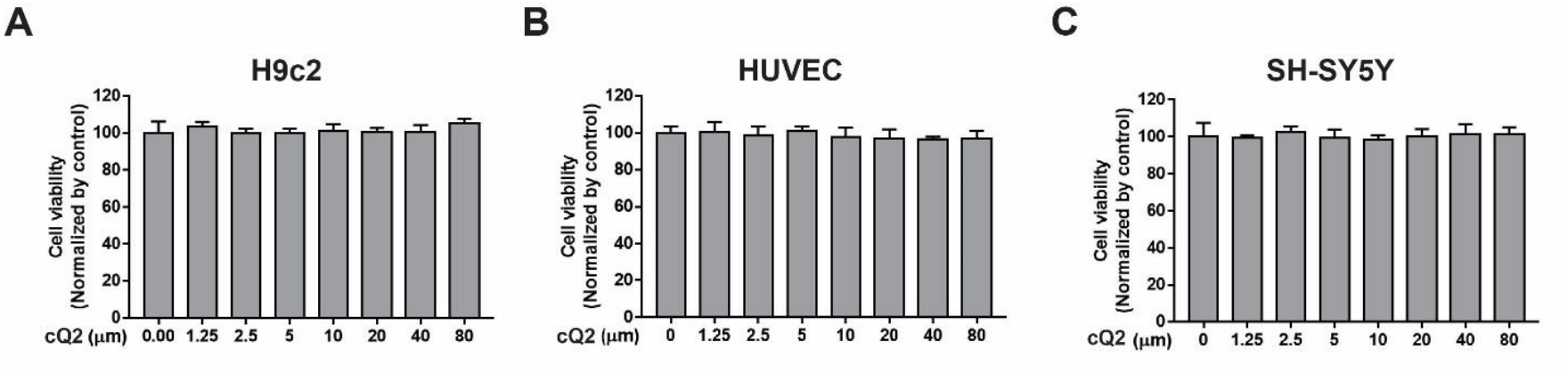


Figure 4.16: Cytotoxic activity of chenotide cQ2. **(A)** H9c2, **(B)** HUVEC and **(C)** SH-SY5Y **cells.** MTT assay was used to determine the cell viability of the cell lines co-incubated with cQ2 at different concentration up to 80 µM. 0.1% DMSO was used as a negative control. All result results were expressed as ± S.E.M (n=3). *p<0.0.5 compared to control.

2.12 Cell migration properties of chenotides

To evaluate the potential wound-healing activity of chenotides, the scratch assay was performed on A431 cells. The A431 cell line is selected for this assay because it is derived from skin carcinoma cells and is robust and easy to handle. The cells were co-incubated with cQ2 for 6 h under serum-free medium, and the extent of cell migration was measured. Figure 4.17 and 4.18 showed the results of the scratch assay. Cells treated with cQ2 showed a larger migration distance (108 ± 24 µM) compared to the untreated control (90 ± 22 µM). This suggests that cQ2 may exhibit moderate wound-healing activity. It has been established that epidermal growth factor (EGF) stimulate cell migration with concentration as low as 100 ng/mL [359, 360]. Therefore, we expect the degree of EGF effect in the same assay to stimulate cell migration greater than cQ2. However, this is a preliminary study and hence, more analysis is required to determine cQ2 role in wound healing.

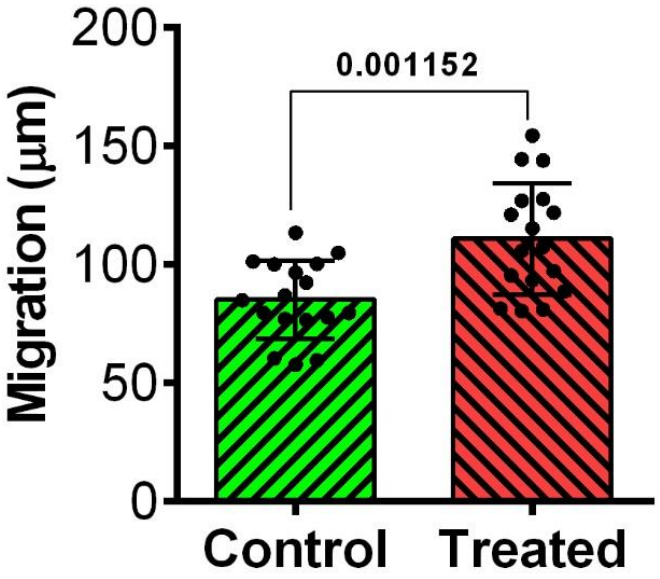


Figure 4.17: Cell migration activity of cQ2 in A431 cells. A431 cells were treated with 10 µM of cQ2 for 6 h in serum-free medium. All result results were expressed as ± S.E.M (n=3). *p<0.0.5 compared to control.

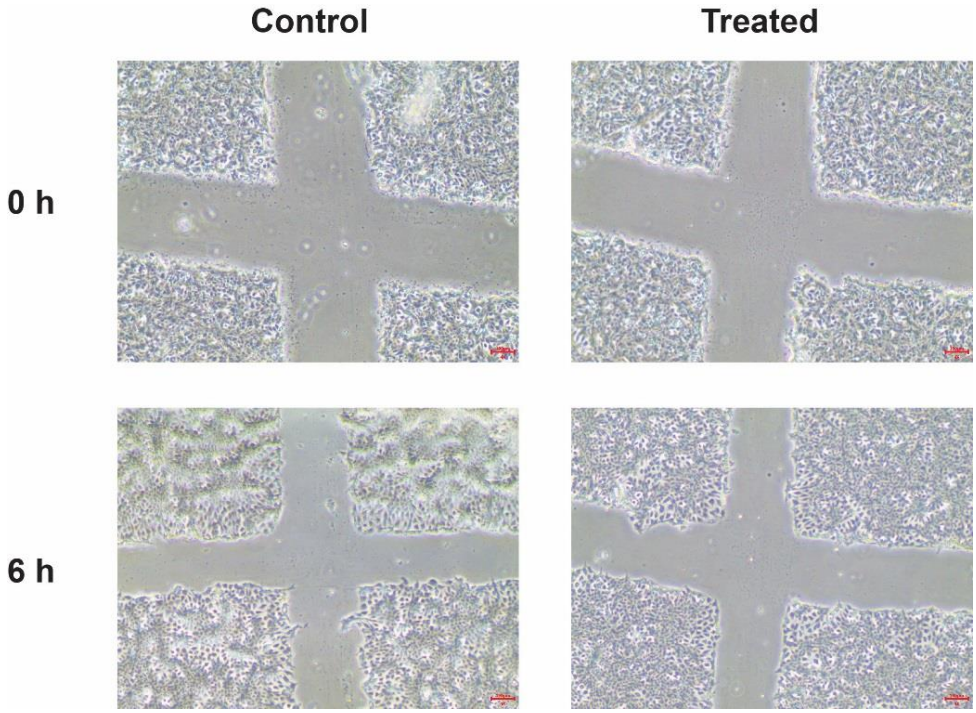


Figure 4.18: Microscopy images of the A431 cell scratch assay. At time point 0, the images were taken immediately after the scratch has been made. After 6 h of incubation, the control and cells treated with 10 µM of cQ2 were imaged again. Images were taken under 4X objective lens.

3. Discussion

Chitin-binding hevein-like peptides represent one of the most important innate immunity integrant that provides plant defense against fungi, insects and pathogens. This class of peptides can be classified into three subclass based on the number of cysteine, namely, 6C-, 8C- and 10C- [34]. At present, twelve 6C-chitin-binding HLPs have been reported, namely; six aSGs and aSRs from *A. sessilis* [51], two Ac-AMPs from *A. caudatus* [94], Ar-AMP from *A. retroflexus* [98], IWF-4 from *B. vulgaris* [95], SmAMP1.1 and SmAMP3 from *S. media* [99, 361]. Here, we report the isolation and characterization of three novel 6C-chitin-binding HLPs from the seeds of *C. quinoa var*. Willd and identified 12 additional chenotides from the whole genome sequence, expanding the current library of 6C-chitin-binding hevein like peptides. Chenotides contain 28–31 amino acid residues. Structurally, it consist of a cystine-knot conformation that has been previously reported in CRPs such as cystine knot-type α-amylase inhibitors (CKAIs) [59, 362]. The cystine-dense conformation confers chenotides high metabolic stability against heat, acid and proteolytic degradation [58]. The presence of a chitin-binding domain enabled chenotides to bind to chitin. The aromatic residues found within the chitin-binding domain mediates the interaction between the peptide and ligand [112]. Chenotides exhibit anti-fungal activity by inhibiting the growth of phyto-pathogenic fungi in a dose-dependent fashion with an IC_{50} of 0.31 µM to 9.51 µM. Chenotides are also non-cytotoxic peptides up to 80 µM that exert a mild cell migration stimulating effect. Since cellular migration plays an important role in wound healing, the results imply that chenotides may exhibit wound-healing effect.

3.1 Sequence comparison of chenotides

Sequence comparison of chenotides with previously reported 6C-chitin-binding HLPs showed >75% sequence similarity (Table 4.1). Additionally, the sequence comparison of chenotides identified from the genomic database showed >79% sequence similarity (Table 4.2). Chenotides also shared sequence similarity with other 6C-, 8C-, and 10C-chitin-binding HLPs. The sequence alignment of chenotides with previous reported 6C-chitin-binding HLPs revealed that the N-terminal of the peptides were highly variable compared to the C-terminal (Table 4.1). The C-terminal Gly was fairly conserved in the 6C-chitin-binding HLPs but replaced by Ala in cQ1, Arg in Ac-AMP and Ar-AMP, and His in SmAMP1.1a and SmaAMP3. Interestingly, the 6C-chitin-binding HLPs generally differed by one amino acid residue at the C-terminal. For instance, cQ1 and cQ2 differ by one alanine at the C-terminal, this equivalent was also observed in asG2 and asG3 in which they are differed by one glycine [51]. This occurrence could be attributed to the various post-translational modification and proteolytic degradation by exopeptidase during purification procedures [96].

The sequence diversity of chenotides is apparent. All the intercysteine loops were not conserved and differed by 2–4 amino acid residues even though they were from the same plant species (Table 4.2). Apart from the six conserved cysteine residues, the rest of the other amino acids were not conserved. Although the chitin-binding domain in loop 3 and loop 4 was fairly conserved, slight variations were observed. For example, cQ1 consists of a Phe residue instead of Tyr, which was observed in cQG2. This phenomenon could be linked to a genetic mutation in the gene

encoding for cQ1 which is improbable to alter the chitin-binding ability because the benzene ring of Phe can promote CH-π stacking interaction [144].

Comparing chenotides to the reported 6C-chitin-binding HLPs, cQ1 showed the highest similarity of 96.4% with aSR1 from *A. sessilis.* This suggests that they may be evolutionarily related. Even though chenotides, Ac-AMPs, Ar-AMPs, IWF-4, aSG and aSR belonged to the Amaranthaceae family, sequence variation was observed, yet they all have chitin-binding properties. This clearly suggests that the 6C-chitin-binding HLPs evolved from the same ancestor and this gene is likely selected and conserved throughout the Amaranthaceae family due to its role in plant defense. Therefore, to understand the evolutionary relationship within the 6C-chitin-binding-HLP subfamily, screening of more plants closely related to the Amaranthaceae family may help establish the evolutionary relationship.

3.2 Biosynthetic diversity of chenotides and 6C-cysteine-rich peptides

Genomic analysis using GenBank database revealed the full-length precursor sequence of chenotides. Chenotides comprise a three-domain architecture, including an ER signal peptide, a pro-mature domain and a C-terminal uniquely different from previous reported 6C-chitin-binding HLPs. The presence of a signal peptide indicated that the bioprocessing of chenotides undergoes the secretory pathway, which is evidently reported in many CRPs [363]. This unique feature distinguishes them from small peptide (5–10 amino acids) that are not ribosomally synthesized [22]. Sequence comparison of chenotide cQ1 precursor revealed that it contained tandem repeat, which resulted in a significantly longer mature domain than the other chitin-binding HLPs. The tandem repeat found in cQ1 contains

identical sequences. At the proteomic level, the molecular weight of 2963 Da was detected, proposing that post-translational modification may have resulted in the cleaving of the hinge region giving rise to two peptides of the same molecular weight. This phenomenon could be a consequence of divergent evolution in which gene duplication of the ancestral gene may occur to allow the plant to adapt to the environment, enhancing survivability [364]

The precursor arrangement of 6C-chitin-binding HLPs and other known 6C-CRPs revealed the biosynthetic diversity in plants (Figure 4.19). The precursors of 6C-chitin-binding HLPs share similarities with thionins [61]. Thionins are a class of well-represented 6C-CRPs known to elicit toxicity towards bacteria, fungi and animal cells by lysing the cell membrane [61, 88, 125]. Notably, the length of the C-terminal tail in thionins appeared to be longer (61 amino acid residues). A comparison of the precursor arrangement of CKAI showed that it consists of an N-terminal pro-peptide and the absence of a C-terminal tail. This similar arrangement was also reported in jasmintides [64]. CKAIs are the smallest family of proteinaceous α-amylase inhibitors and are cysteine- and proline-rich [24, 365]. In contrast to CKAIs, 6C-chitin-binding HLPS are cysteine- and glycine-rich. Another class of 6C-CRPs is the carboxypeptidase inhibitors, which contained a four-domain architecture [34, 69]. This similar architecture can also be found in β-gingkotide, which is well known for its novel hyper disulfide-constrained scaffold [66]. As a result, the diversity in the precursor organization of chenotide, 6C-chitin-binding HLPs and other 6C-CRPs may have led to the functional diversity of these peptides.

By understanding the precursor arrangement of chenotides, it can be exploited to produce transgenic crops with better resistance against invasive pathogenic attack. This method would drastically reduce the use of harmful chemical pesticides on plants that pose detrimental effects to the environment and adverse effects on humans. At present, 6C-chitin-binding HLPs have not been utilized to produce transgenic crops in contrast to α-amylase inhibitors, which have been clone in plants and are considered “Generally Recognized as Safe” for human consumption [24]. This presents a new opportunity to bioprospect chitin-binding HLPs from plants as an alternative approach to develop pathogen-resistant crops.

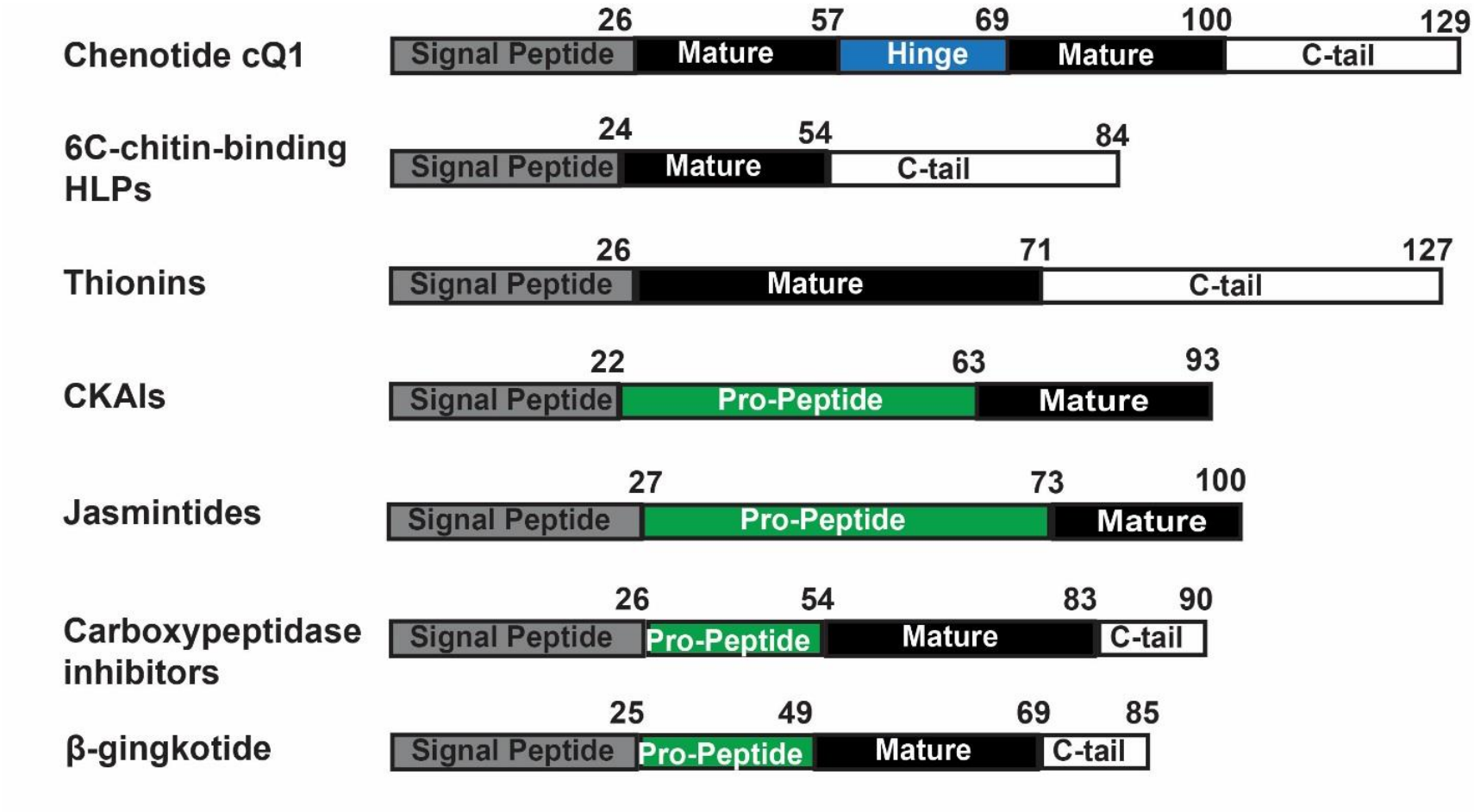


Figure 4.19: Schematic comparison of the biosynthetic architecture of chenotide cQ1 and other 6C-cysteine-rich peptides. The average amino acid length for each domain is listed above each peptide accordingly.

3.3 Chitin-binding interaction of chenotides

A peptide-ligand docking of cQ2 with N-acetylglucosamine was performed using GOLD 5.4.0 version [329]. The GOLD score function takes into consideration of the hydrogen bond and van de Waals energy. The settings in the program were set at default function. The best docking model is shown in Figure 4.20 with a GOLD score of 51.24. The hypothetical peptide-ligand interaction of chenotide cQ2 with *N*-acetylglucosamine hexamer suggests that the ligand binds to the chitin-binding pocket, and the interaction is stabilized by Van de Waals contact and CH-π interaction. Chitin is an oligosaccharide that is made up of long chains of *N*-acetylglucosamine [366, 367]. It forms the primary component in the fungi' cell wall and the exoskeleton of insects [368]. The cQ2 interaction with *N*-acetylglucosamine hexamer showed that the aromatic residues, Phe18, Phe20 and Tyr27 are responsible for the binding interaction. These aromatic residues are located on the third and fourth loop of chenotides, within the chitin-binding domain (S*X*φGφCGX_4φ). Notably, this conserved motif was also reported in Ac-AMP2 [135], Ar-AMP [98], altides [51], hevein [81] and WGA [369]. For example, the aromatic residues in Ac-AMP2 (Phe18, Tyr 20 and Tyr27) was reported to mediate the interaction between the peptide and the sugar moiety [135]. Thus, this finding suggests that the chitin-binding domain is conserved in the chitin-binding HLP family, and the peptide-ligand complex is stabilized primarily by Van de Waals contact and ch-π interaction. By understanding the binding interaction, it enables the design and engineering of putative potent anti-fungal agent that have better binding efficiency towards chitin. Overall, this data provided insights into the binding

interaction of chenotides with the sugar moiety chitin, showed that the chitin-binding motif is well-conserved in the chitin-binding HLP family and can be exploited for the use of designing effective anti-fungal agent that targets chitin.

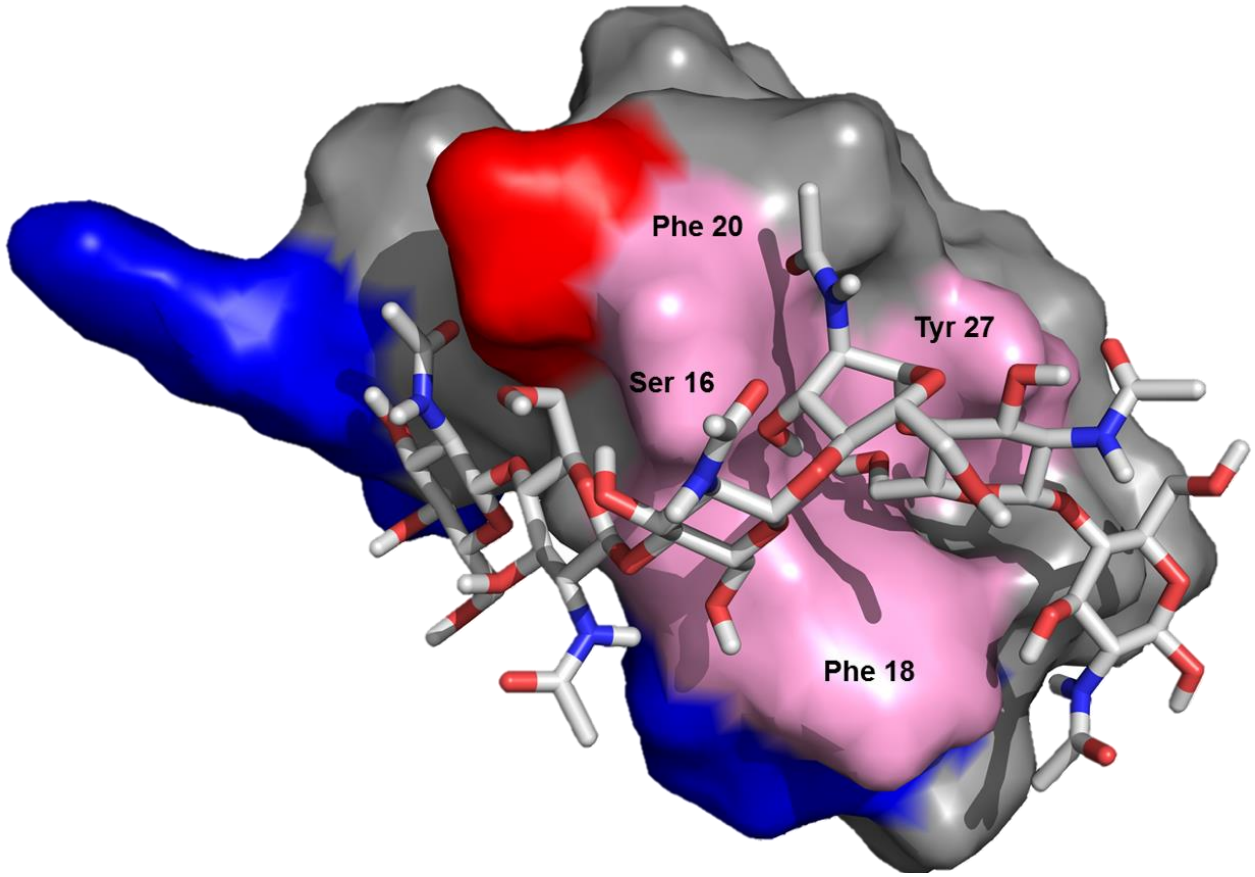


Figure 4.20: Hypothetical peptide-ligand interaction of cQ1 and N-acetylglucosamine hexamer. The figure shows N-acetylglucosamine hexamer binding to the chitin-binding pocket of cQ2. cQ2 is shown in surface representation while the ligand, N-Acetylglucosamine is shown in ball and stick. Surface highlighted in blue and red are basic (Arg, His, Lys) and acidic (Asp, Glu), respectively. The computational docking was performed using GOLD 5.4.0 version [329] with a GOLD score of 51.24.

3.4 Anti-fungal activity of chenotides

The susceptibility of four phyto-pathogenic fungi namely, *A. alternata*, *C. lunata*, *F. oxysporum*, and *R. solani* to chenotide cQ2 indicated that cQ2 exhibits anti-fungal activity. This was evident by the presence of crescent-shaped inhibition zones at the growing end of the mycelial. Furthermore, the IC_{50} of cQ2 against these fungal strains ranged from 0.31 µM to 9.51 µM, subjecting to the fungal strain tested. This data is in agreement with previously reported 6C chitin-binding HLPs such as Ar-AMP, Ac-AMP2, IWF4, aSG1 and aSR1 [51, 94, 95, 98].

Interestingly, the potency of 6C-chitin-binding HLPs appeared to be higher than 8C-chitin-binding HLP like gingkotides which have an IC_{50} ranging from 1.6 µM to 16.3 µM. Bright-field microscopy images revealed the morphological difference between the treated and untreated fungi. Fungi strains treated with cQ2 showed morphological change, such as stunted hyphal growth with shorter branching. A similar phenomenon was observed in IWF-4 isolated from *B. vulgaris*. Co-incubation of the peptide with *Cercospora beticola* also resulted in the retardation of the hyphae and swollen hyphal tips [95]. This similarity led to a postulation that both cQ2 and IWF-4 mediates anti-fungal activity through the same mechanism.

The main component of the cell wall of fungi is formed from a long carbohydrate polymer called chitin, which is composed of multiple cross-linked nascent chitin chains. These newly synthesize chitin chains provide hyphal flexibility that are subsequently rigidified through cross-linking [370]. Although the early synthesize of chitin chains enable flexibility of the hyphal, these developing chains are susceptible to anti-fungal agents. Therefore, we speculated that chenotide cQ2

inhibits the growth of the four phyto-pathogenic fungal strains by binding to the newly synthesize chitin chains. As a result, it impedes the rigidification of the chitin chains and hyphal growth.

Overall, the study focused on the sequence diversity, biosynthetic diversity, chitin-binding interaction, distribution, and the anti-fungal mechanism of a novel group of tandem-repeated 6C-chitin-binding HLPs from *C. quinoa.* This data has expanded the existing library of 6C-chitin-binding HLPs from 12 to 15. Despite the establishment of anti-fungal activity in chenotides, further study is necessary to understand their mode of action. Together, the study could be utilized to develop omnipotent anti-fungal agents for biotechnological applications.

CHAPTER FIVE

Avenatides: 8C-chitin-binding Hevein-like Peptides from the Globally Consumed Grain *Avena sativa*

1. Introduction

Avena sativa, commonly known as oats, is a type of cereal grain cultivated for its seeds. It belongs to the Poaceae family and is ranked sixth in the global cereal production. The grains are rich in protein, fiber and minerals and remained as one of the most important grain crops in developing countries. In developed countries, *A. sativa* is reported to have a dual function as a nutraceutical and therapeutic agent [217]. Traditionally, *A. sativa* is a folk medicine used to treat a wide range of ailments such as insomnia, nervous exhaustion, burns and eczema [219]. Besides, they also exhibit antispasmodic, antitumour and diuretic functions [219]. The progressive research of this globally consumed grain showed that it displayed a myriad of pharmacological activities such as anti-microbial [218], antioxidant [221], antichloesterolaemic [220], immunomodulatory [371] and wound-healing [372], which are highly sought after as active natural ingredient in Food and Drug Administration (FDA) [373].

However, researches on the active ingredient in oats primarily focus on phytoconstituents such as alkaloids [374], avenanthramides [221], tocols [375], flavonoids [376] and saponins [377]. Therefore, this raised the question of whether peptide constituent may also contribute to the therapeutic function of *A. sativa*.

This chapter reports the isolation and characterization of nine novel tandem-repeated 8C-chitin-binding hevein-like peptides (HLPs) from *Avena sativa*. These peptides are collectively termed as avenatide, av1–aV9. Proteomics and transcriptomic analysis revealed that avenatides contain the same cysteine motif and disulfide connectivity with previously reported 8C-chitin-binding HLPs such as vH1 [53], mO1 [52], Fa-AMP1 [100] and Pn-AMP1 [378]. Phylogenetic analysis showed that avenatides belong to a new class of 8C-chitin-binding HLP due to its unique feature of a tandem-repeat that has not been reported. The putative biosynthesis pathway of avenatides revealed that the sequence of the hinge regions plays a role in determining the cleaving of the peptides. Structurally, avenatides are cysteine-dense, conferring them metabolic stability against acid, heat and proteolytic degradation. In addition, biological assays showed that avenatide aV1 peptide that displayed anti-fungal and cell migration properties. Overall, this work shed light on the sequence, biosynthesis, bioactivity, and expanded the existing library of 8C-chitin-binding HLPs. As such, the peptide-derived bioactive compounds discovered from *A. sativa* may be considered for developing as anti-infectives and potential bio therapeutics.

2. Results

2.1 Screening, isolation and disulfide determination of avenatides

1 kg of *A. sativa* grain was used for a large-scale extraction using a weight to volume ratio of 1:10. C_{18} flash column was used to remove colored pigments and other impurities from the filtered extract. Fractional elution was done using 20%–80% ethanol. Eluted fractions containing peptide of interest were purified by SCX-HLPC and RP-HPLC. Based on the MS screening profile, clusters of putative CRPs within the range of 2–5 kDa were detected (Figure 5,1A). S-reduction with DTT and S-alkylation with IAA were done to determine the number of cysteines present. A mass increase of 464 Da indicated that these CRPs contained eight cysteine residues (Figure 5.1B and 5.1C). CRPs that have been successfully isolated from *A. sativa* were termed as avenatide and labeled as aV1–aV6 with relative monoisotopic molecular mass [M + H] of 3730.0, 3747.95, 3767.9, 3785.9, 3846.0 and 3858.0 Da, respectively.

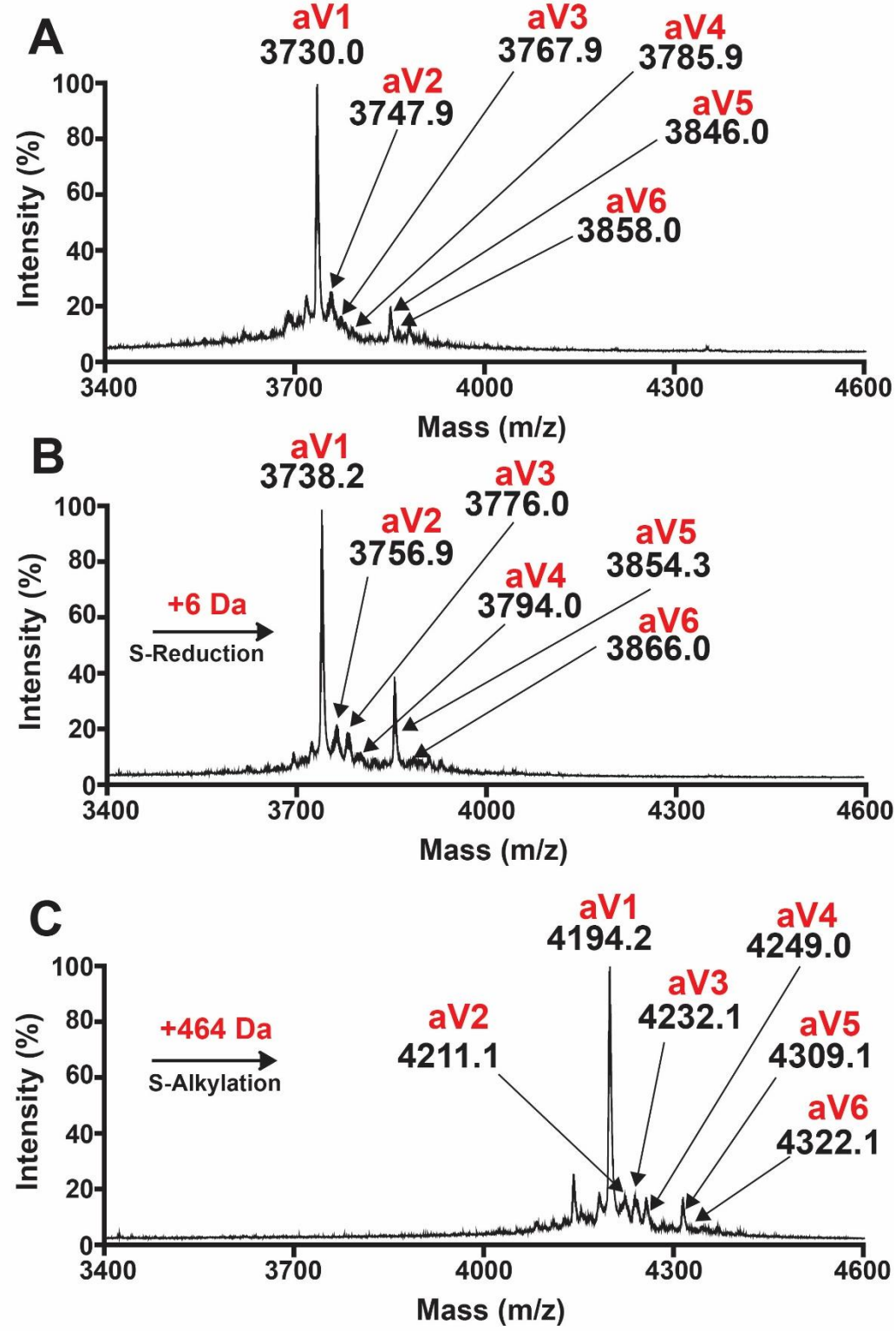


Figure 5.1: MS profile of avenatides. MS profile of **(A)** avenatides **(B)** S-reduction of avenatides and **(C)** S-reduction of avenatides. A mass increase of 464 Da after S-reduction and S-alkylation indicate the presence of eight cysteine residues.

2.2 Sequence determination of avenatides

LTQ Orbitrap Elite MS/MS was employed to determine the primary sequences of avenatides. The S-alkylated avenatides were analyzed by nanospray tandem MS for *de novo* sequencing, and their primary sequences were deduced by evaluating the mass differences between the *b*- and *y*-series ions (Figure 5.2–5.4). Results from the MS/MS sequencing revealed that S-alkylated aV1 contain a putative sequence of ACSSSSPCPGN**X**CCS**X**WGYCG**X**GGDYCGSGC**X**SGPCTGA, where **X** represents the isobaric amino acid Leu/Ile or Lys/Gln. The assignment was resolved by cDNA sequences from GenBank and the full primary sequence was deduced as ACSSSSPCPGNQCCSKWGYCGLGGDYCGSGCQSGPCTGA (Figure 5.2A). The other avenatides were similarly determined, and based on proteomics and transcriptomic analysis, nine avenatides (aV1–aV9) were successfully deduced.

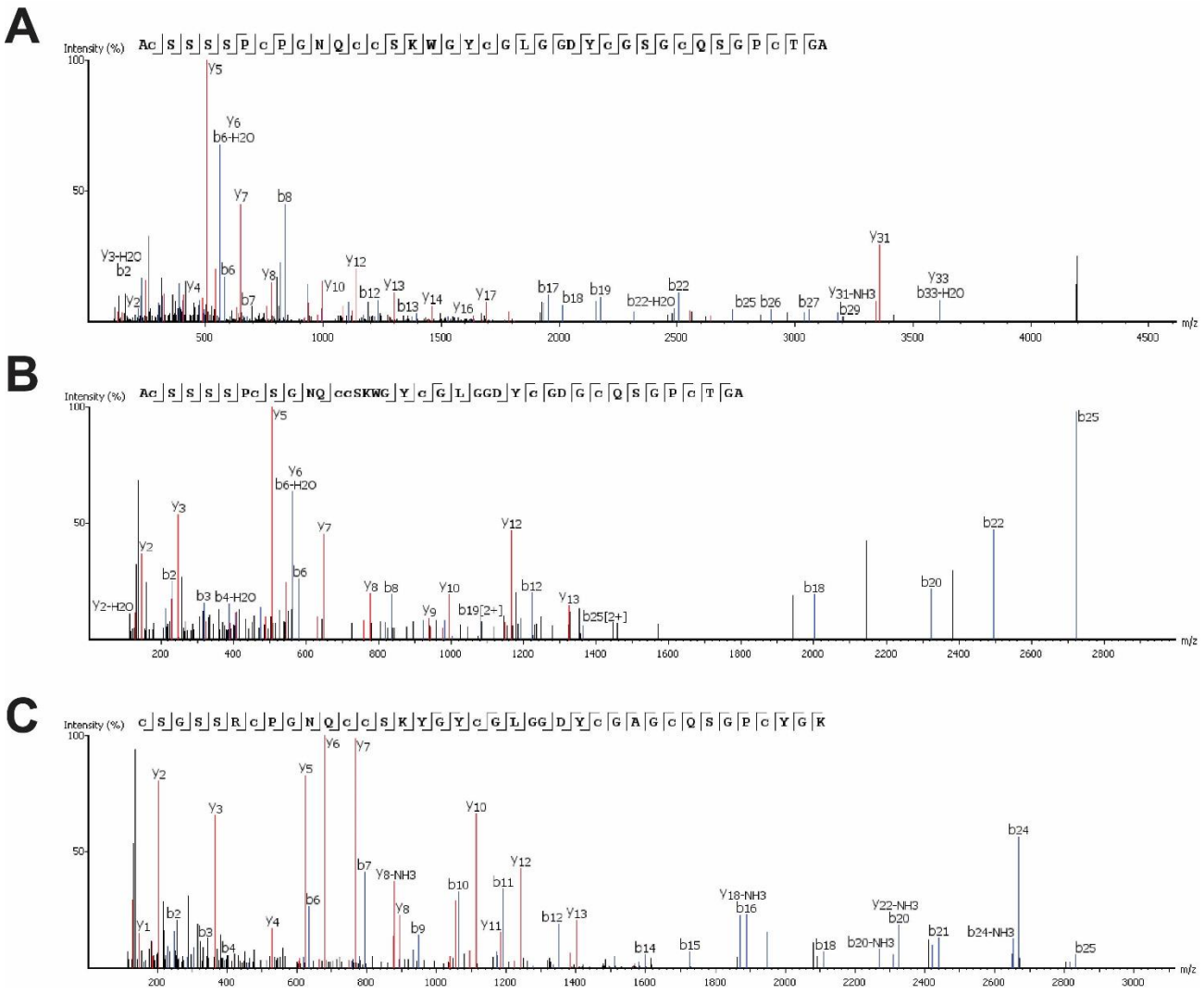


Figure 5.2: LC-ESI-LTQ-Orbitrap MS/MS positive mode mass spectra of avenatides. Mass spectra of **(A)** aV1 **(B)** aV2 and **(C)** aV3 scanned between mass ranged from 0 to 4500 m/z. Transcriptomic data was used to confirm the assignment of isobaric residues Leu/Ile or Lys/Gln.

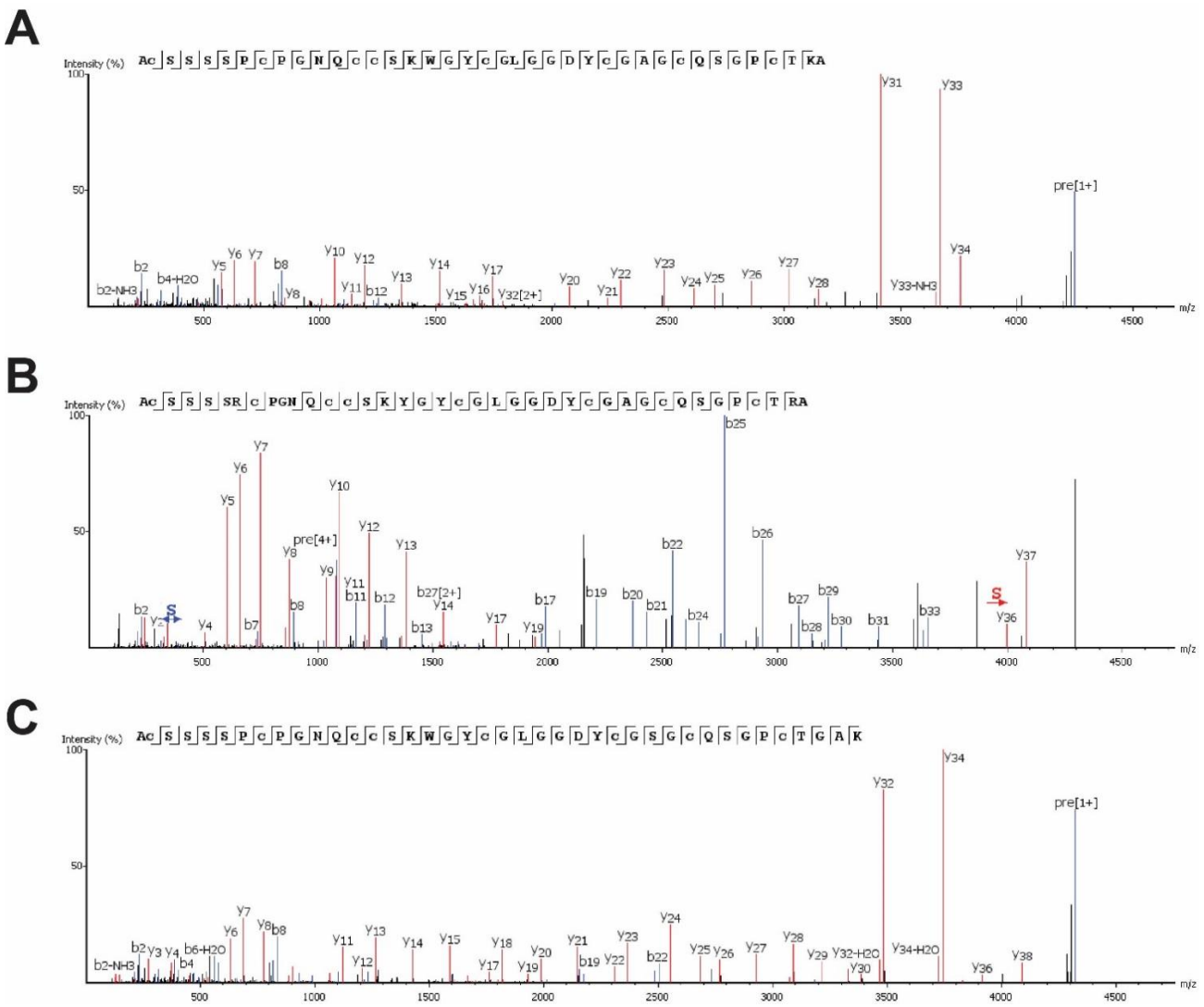


Figure 5.3: LC-ESI-LTQ-Orbitrap MS/MS positive mode mass spectra of avenatides. Mass spectra of **(A)** aV4 **(B)** aV5 and **(C)** aV6 scanned between mass ranged from 0 to 4500 m/z. Transcriptomic data was used to confirm the assignment of isobaric residues Leu/Ile or Lys/Gln.

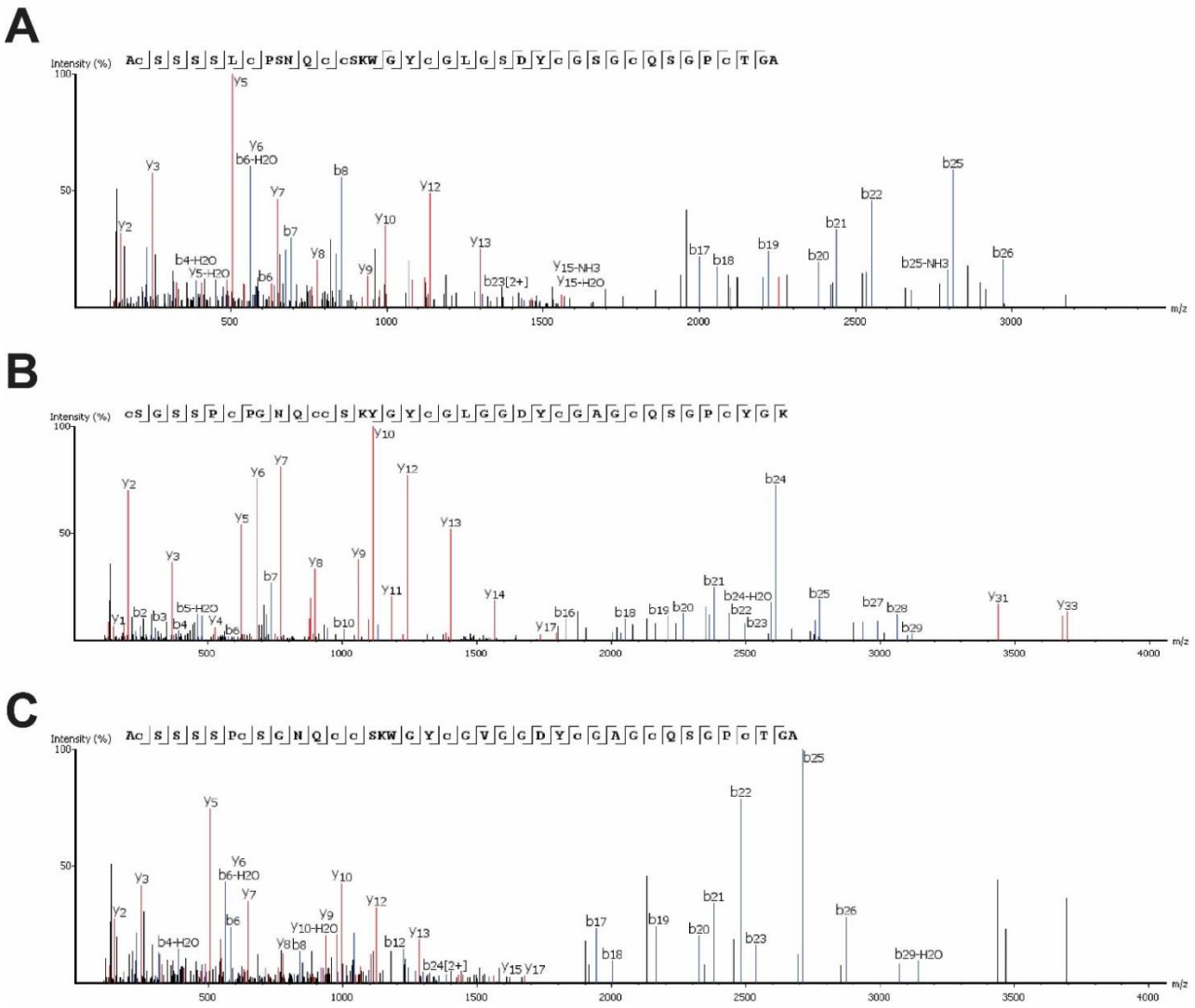


Figure 5.4: LC-ESI-LTQ-Orbitrap MS/MS positive mode mass spectra of avenatides. Mass spectra of **(A)** aV7 **(B)** aV8 and **(C)** aV9 scanned between mass ranged from 0 to 4500 m/z. Transcriptomic data was used to confirm the assignment of isobaric residues Leu/Ile or Lys/Gln.

2.3 Sequence comparison of avenatides and reported 8C-chitin-binding HLPs

Nine novel peptides collectively named as avenatide were isolated and sequenced at the proteomic and transcriptomic levels. Table 5.1 summarized the sequence alignment of avenatides aV1–aV9 and reported 8C-chitin-binding HLPs. Avenatides residue length ranges from 28 to 39 amino acids and are cysteine- (eight residues), glycine- (9–10 residues) and serine-rich (6–8 residues), accounting for approximately 64 % of the primary sequence. Sequence comparison revealed that avenatide aV1 is the truncated version of aV6, with a deletion of a lysine residue at the N-terminal. BLAST analysis associated avenatides to a family of 8C-chitin-binding HLP containing a chitin-binding domain and shared an average sequence similarity of 66%. Avenatide aV1 shared the highest sequence similarity of 91.7% with Avesin A, isolated from *Avena sativa. L*. The high sequence similarity observed is because both peptides derived from the same plant species, *A. sativa*. In contrast, aV1 shared the lowest sequence similarity of 60.6% with mO1, isolated from *M. oleifera* because they belonged to different plant species and family. Interestingly, avenatides was observed to be rich in serine (6–8 residues) compared to the other 8C-chitin-binding HLPs (3–5 residues).

The cysteine motif pattern $X_{1\text{-}3}$**C**-$X_{2\text{-}8}$**C**-X_4**CC**-X_5**C**-$X_{3\text{-}5}$**C**-$X_{3\text{-}5}$**C**-$X_{3\text{-}5}$**C**$_{\mathbf{3\text{-}5}}$ is highly conserved among avenatides and the reported 8C-chitin-binding HLPs. All of them contained six intercysteine loops and were not conserved. Apart from the conserved cysteine spacing, Ser15, Gly21 and Tyr26 (numbering based on aV1)

were absolutely conserved. Due to the peptide abundance, aV1 was selected for further characterization of avenatides.

Table 5.1: Sequence comparison of the mature peptide sequences of avenatides and reported 8C-chitin-binding-hevein-like peptides

Peptide	Species	Amino acid sequence	Mass (Da)[1]	Charge[2]	Similarity (%)	Method[3]	Reference
Loop		**I II III IV V VI**					
		* * * *					
aV1	*A. sativa*	---A--C---SSSSPCPGNQCCSKWGYCGLGGDYCGSG--CQSGPCTGA--	3729.3	0	100	T, P	This work
aV2	*A. sativa*	---A--C---SSSSPCSGNQCCSKWGYCGLGGDYCGDG--CQSGPCTGA--	3747.3	-1	94.9	T, P	This work
aV3	*A. sativa*	------C---SGSSRCPGNQCCSKYGYCGLGGDYCGAG--CQSGPCYGK--	3767.4	+2	91.9	T, P	This work
aV4	*A. sativa*	---A--C---SSSSPCPGNQCCSKWGYCGLGGDYCGAG--CQSGPCTKA--	3784.4	+1	97.4	T, P	This work
aV5	*A. sativa*	---A--C---SSSSRCPGNQCCSKYGYCGLGGDYCGAG--CQSGPCTRA--	3848.5	+2	94.9	T, P	This work
aV6	*A. sativa*	---A--C---SSSSPCPGNQCCSKWGYCGLGGDYCGSG--CQSGPCTGAK-	3857.4	+1	100	T, P	This work
aV7	*A. sativa*	---A--C---SSSSLCPSNQCCSKWGYCGLGSDYCGSG--CQSGPCTGA--	3805.4	0	94.7	T	This work
aV8	*A. sativa*	------C---SGSSPCPGNQCCSKYGYCGLGGDYCGAG--CQSGPCYGK--	3708.4	+1	91.9	T	This work
aV9	*A. sativa*	---A--C---SSSSPCSGNQCCSKWGYCGVGGDYCGAG--CQSGPCTGA--	3689.3	0	97.4	T	This work
Avesin A	*A. sativa*	WSG---C------SPCPGNECCSKYGYCGLGGDYCGAG--CQSGPCYG---	3680.3	-1	91.7	T,P	Le at al. 2003
vH1	*V. hispanica*	---FQ-CGRQAGGARCSNGLCCSQFGYCGSTPPYCGAGQ-CQSQ-CA----	4145.7	+2	66.7	T, P	Wong et al. 2017
mO1	*M.oleifera*	----QNCGRQAGNRACANQLCCSQYGFCGSTSEYCSRANGCQSN-CRGG--	4603.8	+3	60.6	T, P	Kini et al 2017
gB1	*G. biloba*	---DPTCSKLGD-FKCNPGRCCSKFNYCGSTAAYCGRGN-CIAQ-CPSNVS	4715.4	+3	60.9	T, P	Wong et al. 2016
Hevein	*H. brasiliensis*	----EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSN-CKD---	4715.8	-2	64.3	T, P	Archer et al. 1960
Fa-AMP1	*F. esculentum*	----AQCGAQGGGATCPGGLCCSQWGWCGSTPKYCGAG--CQSN-CK----	3876.5	+2	66.7	T, P	Fujimura et al. 2003
Pn-AMP1	*P.nil*	----QQCGRQASGRLCGNRLCCSQWGYCGSTASYCGAG--CQSQ-CRS---	4314.7	+4	64.5	T, P	Koo et al. 1998

[1]Mass (Da): represents the experimentally found molecular weight. [2]Charge: represents the total charge of the molecule, and calculated by the sum of positive (lysine, arginine and histidine residues) and negative (glutamate and aspartate residues) charges. [3]Methods: represents the primary sequences obtained by (T) and/or proteomic (P) method. Cys and chitin-binding domain are highlighted in yellow and green, respectively.

2.4 Structural analysis of avenatides

Solution-state NMR resolved the structure of aV1(Figure 5.5) using the distance restraints provided from 2D NOESY, and the hydrogen bond restraints. All spin-spin systems of aV1 were identified, and approximately 98% of the proton resonances were unambiguously assigned. The solution structure of aV1 was determined based on a total of 260 NMR derived distance restraints and eight hydrogen bonds. Figure 5.5A showed the NMR ensemble of the ten lowest energy aV1 structures. The root-mean-square deviation (RMSD) value of the ten best structures for residues Ser3–Gly9 and Gly18–Thr37 was 1.23±0.26 Å and for all heavy atoms was 1.68 ±0.27 Å (Table 5.2 and 5.3). The structure of aV1 was well-defined by a number of medium and long-range NOEs, which consisted of two anti-parallel β-strands (β1: Cys13–Ser15 and β2: Try19–Gly21) (Figure 5.5A). The disulfide connectivity of avenatide aV1 was deduced as CysI–CysIV, CysII–CysV, CysIII–CysVI, CysV–CysVIII. The structure contained a cystine-knot motif and a fourth disulfide bond at the C-terminal. The acidic residue (Asp25) was located closer to the C-terminal, and the basic residue (Lys16) was located in the core of the structure (Figure 5.5B). The 3D structure of avenatides aV1 was deposited to Protein Data Bank with an accession number 6M5C. Figure 5.5C showed the electrostatic surface topology comparison of aV1 and morintide mO1 (PDB: 5WUZ). The key residues involved in chitin-binding is highlighted in purple in aV1 (Ser15, Trp18, Tyr19 and Tyr26), the equivalent can be observed in mO1 (Ser19, Tyr21, Phe23 and Tyr30). The acidic residue of mO1 was located in the core of the structure and the basic residues (Arg5, Arg10, Arg33 and Arg42) were distributed

in both the N- and C-terminal of the structure. Morintide mO1 contained more charged residues and an overall charge of +3 compared to aV1 with an overall neutral charge. Although mO1 had the lowest sequence similarity with aV1, structurally, they both contained a chitin-binding active site. This suggests that the chitin-binding domain is essential for plants and may have been evolutionary conserved.

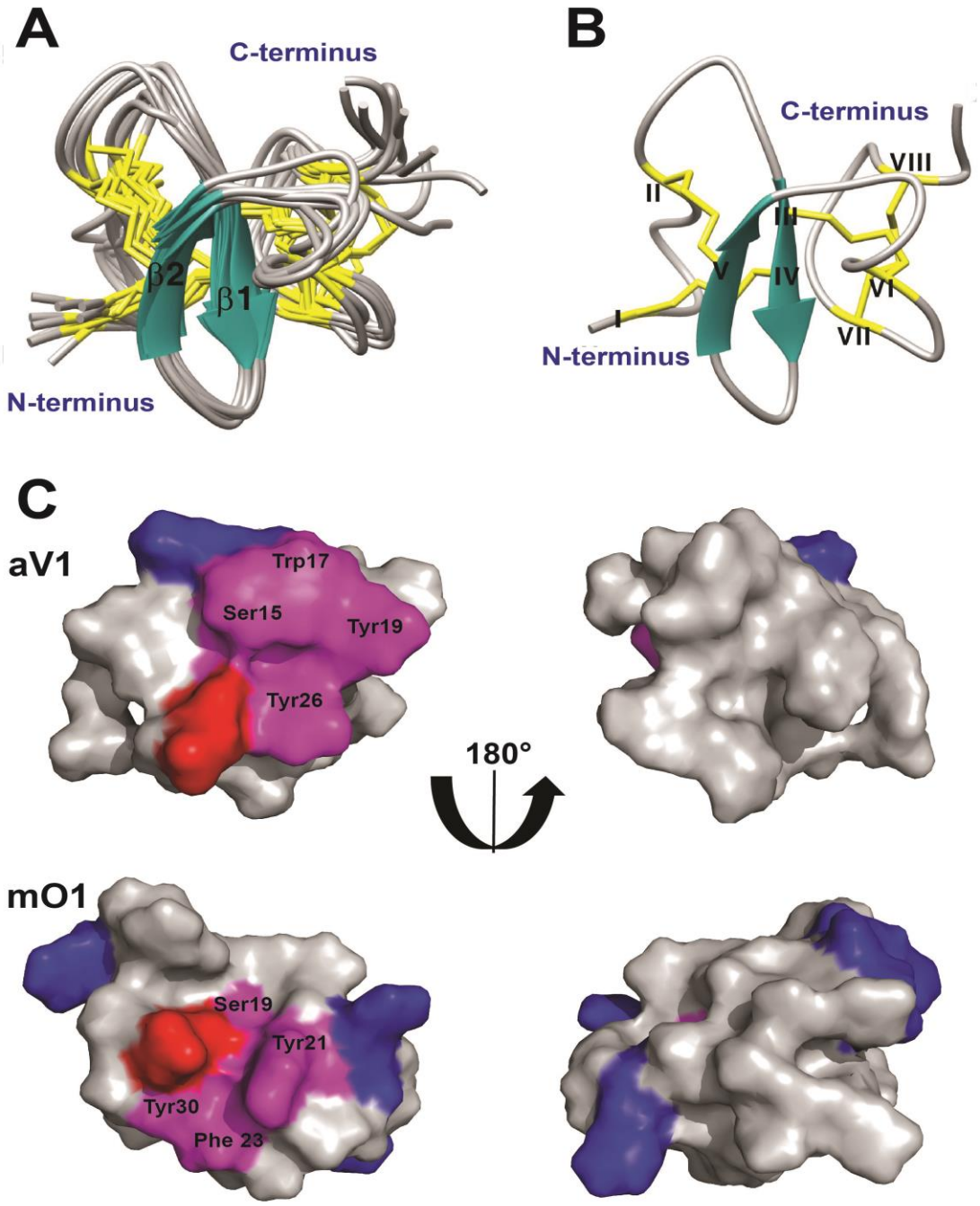


Figure 5.5: Solution NMR structure of aV1. **(A)** Superposition of the aV1 backbone traces from the final 10 ensembles solution structures and restrained energy minimized structure. **(B)** Ribbon representation of aV1 structure **(C)** Surface topology comparison between aV1 and mO1 (PDB: 5WUZ) in flip views showing the chitin-binding domain. The positive, negative and hydrophobic residues are displayed as blue, red and grey respectively. were highlighted in purple represents the chitin-binding domain.

Table 5.2: Structural statistics for the final 10 conformers of aV1[a]

Distance restraints	
Intra-residue (i-j = 0)	91
Sequential ($\|i$-$j\|$ = 1)	92
Medium range ($2 \leq \|i$-$j\| \leq 4$)	33
Long range ($\|i$-$j\| \geq 5$)	44
Hydrogen bond	8
Total	268
Average rmsd to the mean structure (Å)[b]	
Backbone atoms	1.23 ± 0.26
Heavy atoms	1.68 ± 0.27
ϕ/ψ space[c]	
Most favored region (%)	72.8
Additionally allowed region (%)	24.8
Generously allowed region (%)	2.0
Disallowed region (%)	0.4
RMSD from covalent geometry	
Bonds (Å)	0.0069 ± 0.0023
Angles (deg.)	0.027 ± 0.027
Impropers (deg.)	0.374 ± 0.038
RMSD from experimental restraints	
NOEs (Å)	0.0361 ± 0.0030

[a] Selected from 100 calculated conformers according to overall energy.

[b] Calculated with MOLMOL using range 3-8, 11-37.

[c] Calculated with PROCHECK-NMR.

Table 5.3: Proton chemical shift assignments for each amino acid residues of peptide aV1

	HN (ppm)	Hα (ppm)	Hβ (ppm)		Others (ppm)
A1		4.252	1.563		
C2	7.981	5.116	3.318	2.869	
S3	8.429	4.303	4.178	3.966	
S4	8.909	4.210	4.056		
S5	7.973	4.585	3.932	3.766	
S6	7.144	4.871	3.932	3.561	
P7		4.541	2.201		Hγ, 2.328; Hδ, 3.648, 3.401
C8	8.877	4.589	3.005	2.638	
P9		4.535	1.866		Hγ, 2.267; Hδ, 3.818,3.650
G10	8.525	3.735			
N11	8.616	4.562	2.991	2.607	
Q12	6.892	4.210	1.947	1.869	Hγ, 2.519
C13	8.725	4.723	3.840	2.784	
C14	7.925	5.198	3.036	2.744	
S15	9.874	4.904	4.363	4.318	Hγ, 6.302
K16	8.844	3.815	1.549	1.083	Hγ, 0.533, 0.263
W17	7.203	4.816	3.727	2.967	Hδ1,7.300, Hε1, 10.177
G18	8.098	4.063, 3.643			
Y19	7.458	4.977	3.454	2.956	Hδ, 7.110; H*ε*, 6,841
C20	8.824	5.710	3.085	2.838	
G21	8.814	3.445, 1.983			
L22	7.943	4.964	1.569		Hγ, 1.446; Hδ, 1.036, 0.909
G23	7.994	4.553, 3.876			
G24	8.802	3.999, 3.814			
D25	8.953	4.485	2.483		
Y26	7.669	4.063	2.837	2.563	Hδ, 7.105; H*ε*, 6.741
C27	8.269	5.139	3.576	2.766	
G28	7.518	4.310, 3.827			
S29	8.483	4.079	3.809	3.758	
G30	9.095	3.954, 3.565			
C31	7.916	4.512			
Q32	9.646	4.480	2.093	1.467	Hγ, 2.090, 2.175
S33	7.533	4.558	3.990		
G34	8.691	4.550, 3.715			
P35		4.791	1.779		Hγ, 2.244; Hδ, 3.151
C36	8.734	4.864	3.551	2.530	
T37	9.142	4.246			Hγ2, 1.203
G38	8.644	3.969, 3.877			
A39	7.806	4.134	1.299		

2.5 Putative biosynthesis of avenatides

The transcriptomic data can provide information on the putative biosynthesis pathway of avenatides. Two precursor sequences, designated as *av1* and *av2* revealed that avenatides contained a three-domain architecture comprising a signal peptide, a pro-mature domain and a C-terminal tail (Figure 5.6A). The presence of a signal peptide indicated that avenatides were ribosomally synthesized [22]. Interestingly, both *av1* and *av2* contained tandem-repeated mature domains that differentiated them from previously reported 8C-chitin-binding HLPs such as vH1 [53], mO1 [52] and gB1 [54] (Figure 4.8B). The post-translational processing of *av1* precursor in the endoplasmic reticulum produce four mature avenatides, namely, aV3, aV5, aV7 and aV4. Whereas, the post-translational processing of *av2* precursor produce the other four avenatides, namely, aV8, aV2, aV6 and aV9. Notably, av7, aV8 and aV9 were only detected at the transcriptomic level. Additionally, aV1 appears to be a truncated version of aV6. Based on observation, the C-terminal of all precursor signal peptides ended with either an Ala or Gly. However, the C-terminal tail of avenatides precursor were shorter (eight amino acid residues) compared to vH1, mO1 and gB1 (16–28 amino acid residues) and were absolutely conserved. Together, this data suggests that avenatides belong to a new 8C-chitin-binding HLPs sub-family that contained tandem-repeated domains.

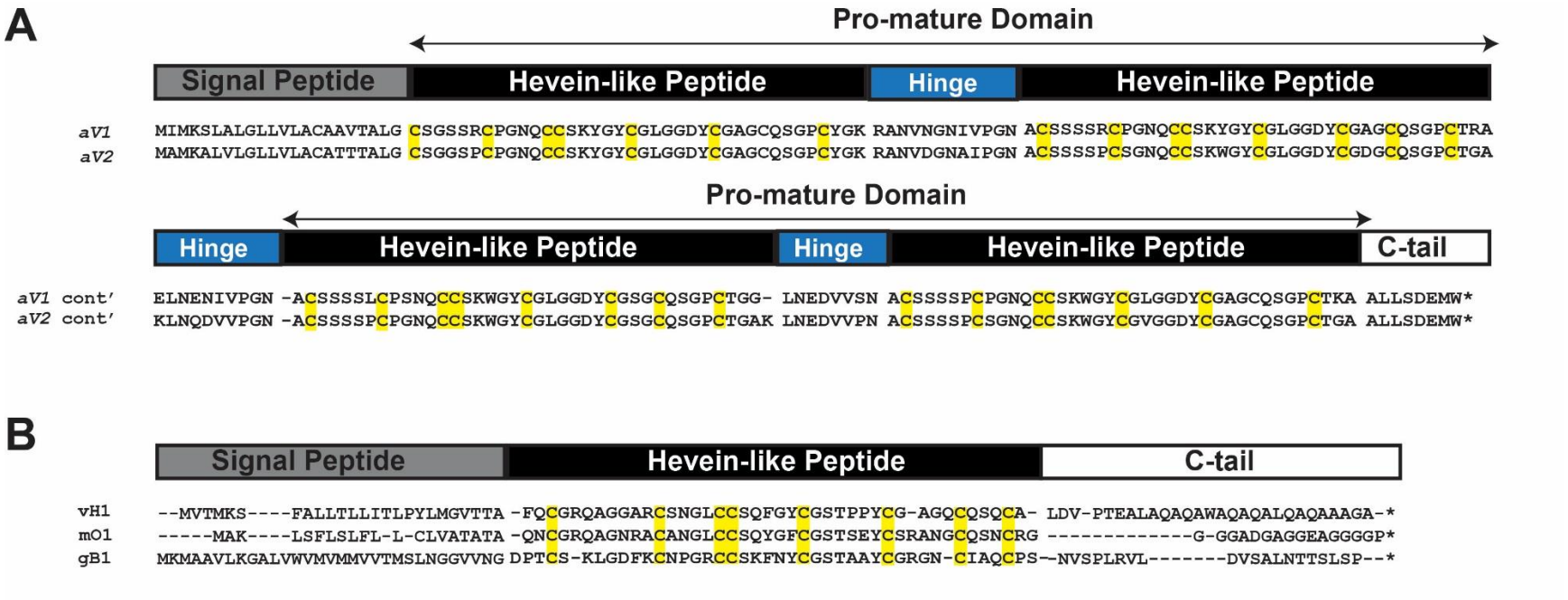


Figure 5.6: Gene alignment and putative biosynthesis pathway of 8C-chitin-binding hevein-like peptides. (**A**) Precursor sequences of avenatide. Avenatide precursors contained a three-domain architecture including a signal peptide, pro-mature domain and C-terminal. (**B**) Precursor sequences of reported 8C-chitin-binding HLPs.

2.6 Evolutionary analysis of chitin-binding HLPs and proteins

To study the evolutionary relationship of avenatides, chenotides, reported chitin-binding HLPs and chitin-binding protein such as chitinases and lectins, a phylogenetic tree was constructed using a neighbor-joining clustering algorithm (Figure 5.7). Data analysis showed four major clusters, namely, cluster 1–4. Cluster 1 consists of chitin-binding HLPs with a protein cargo. Cluster 2 was divided into two subtypes. Cluster 2A consists of 8C- and 10C-chitin-binding HLPs without protein cargo and chitinases, whereas cluster 2B consists of lectins. Cluster 3 and 4 represents the other chitin-binding HLPs that do not have a protein cargo. Although aV1 and gB1–5 are 8C-chitin-binding HLPs, they were in separate clusters. This differential clusters arise because aV1 was derived from angiosperm while gB 1–5 were derived from gymnosperm origins [54]. Notably, avenatide aV1 was clustered adjacent to WGA1 as they both contain four tandem-repeats, suggesting a close evolutionary relationship between tandem-repeated chitin-binding HLPs and lectins.

Interestingly, hevein, an 8C-chitin-binding HLP and Ee-CBP a 10C-chitin-binding HLP isolated from *Euonymus europaeus L.* were in the same cluster as both peptides contained a protein cargo. Hevein expressed a Barwin protein cargo while Ee-CBP expressed a class I chitinase-like domain [101]. These protein cargos are catalytic domains expressed in addition to the chitin-binding domain, providing a duel-catalytic function. Additionally, in cluster 2A the chitinases were observed to evolve later compared to hevein. This is confirmed by an evolutionary study that showed the N-terminal domain of class I chitinases is similar to hevein [379]. As such, these class I chitinases are speculated to be the product of gene

incorporation in the hevein gene. Horizontal gene transfer (HGT) is a process where genes are transferred between species. For example, Bergthorsson *et al.* demonstrated that the ribosomal gene *rps11* is widely distributed plant families [380], and this study was confirmed by Richardson *et al.* that showed the *rsp 11* gene was found in 280 flowering plants [381] . Such prevalence of HGT allows the introduction of genetic novelties in plants to adapt to new niches, resources and spread evolutionary success. Our data collectively showed that tandem-repeated chitin-binding HLPs are closely related to chitin-binding proteins like chitinases and lectins as a result of divergent evolution. The evolutionary conserved chitin-binding domain is important in plant and HGT is likely the mechanism responsible for the spread of chitin domain genes, introducing new functions in plants.

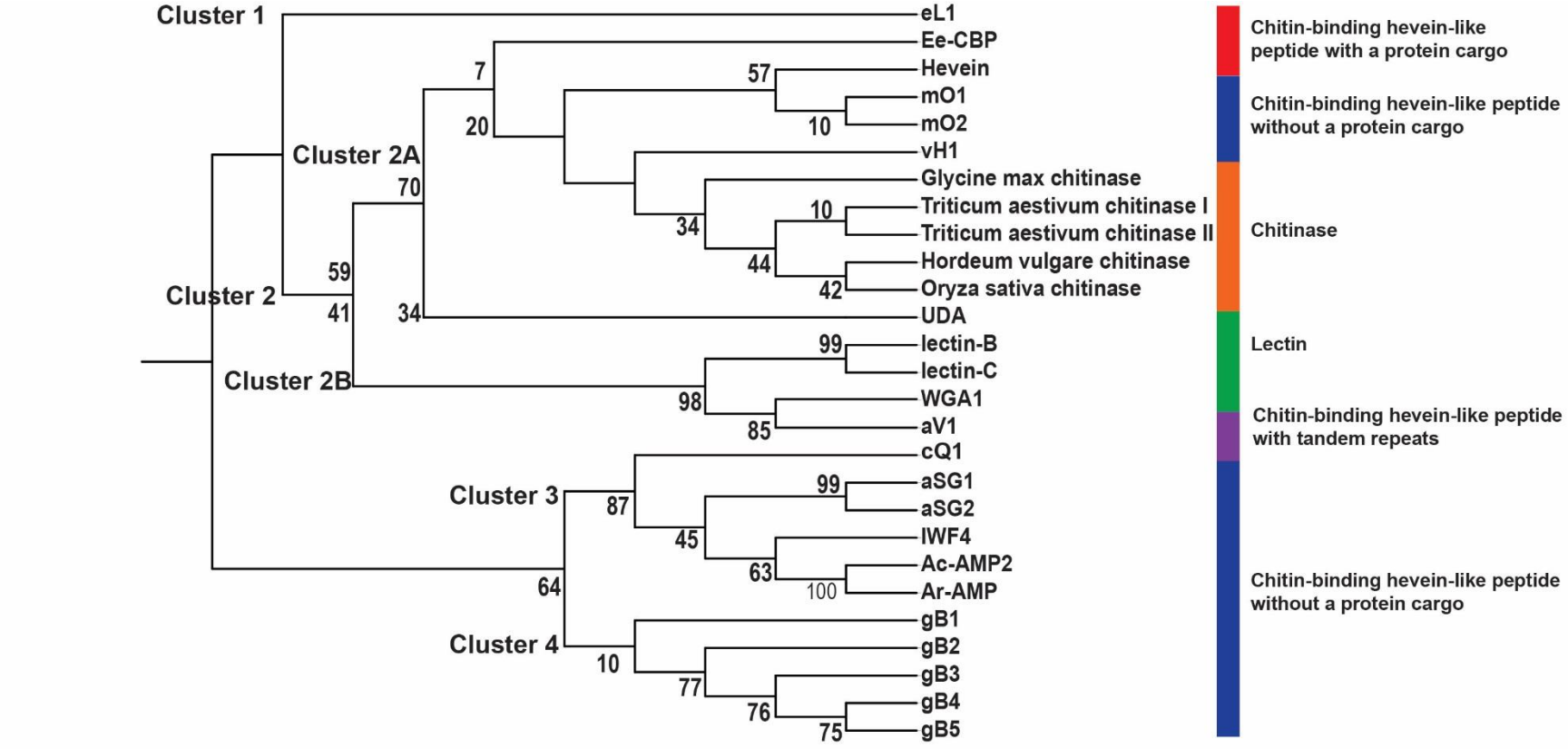


Figure 5.7 Phylogenetic tree analysis of chitin-binding-hevein-like peptides, chitinases and lectins. Neighbor-clustering algorithm was used to analyze the precursor sequences. iTOL was used to construct the phylogenetic tree.

2.7 Chitin-binding activity of avenatides

Chitin beads were used to determine the chitin-binding properties of avenatides. The native and S-alkylated aV1 (5 µg) were incubated with chitin beads up to 4 h at 25 °C. The UPLC chromatograph revealed that the S-alkylated aV1 was detected in the supernatant after 4 h (Figure 5.8A), suggesting that S-alkylated aV1 did not bind to the chitin beads. This showed that the tertiary structure is essential for chitin binding interaction. On the contrary, the native aV1 bound to the chitin beads after 1 h of incubation as shown by the decrease peak height (Figure 5.8B). High salt elution up to 1 M NaCl was not effective in eluting aV1, indicating a strong binding interaction. The successful elution of aV1 was optimized at 1 M acetic acid at 55 °C (Figure 5.8C). This confirmed that aV1 is an 8C-chitin-binding HLPs that has high affinity towards chitin beads and strong acid with heat is necessary for the elution of the peptide.

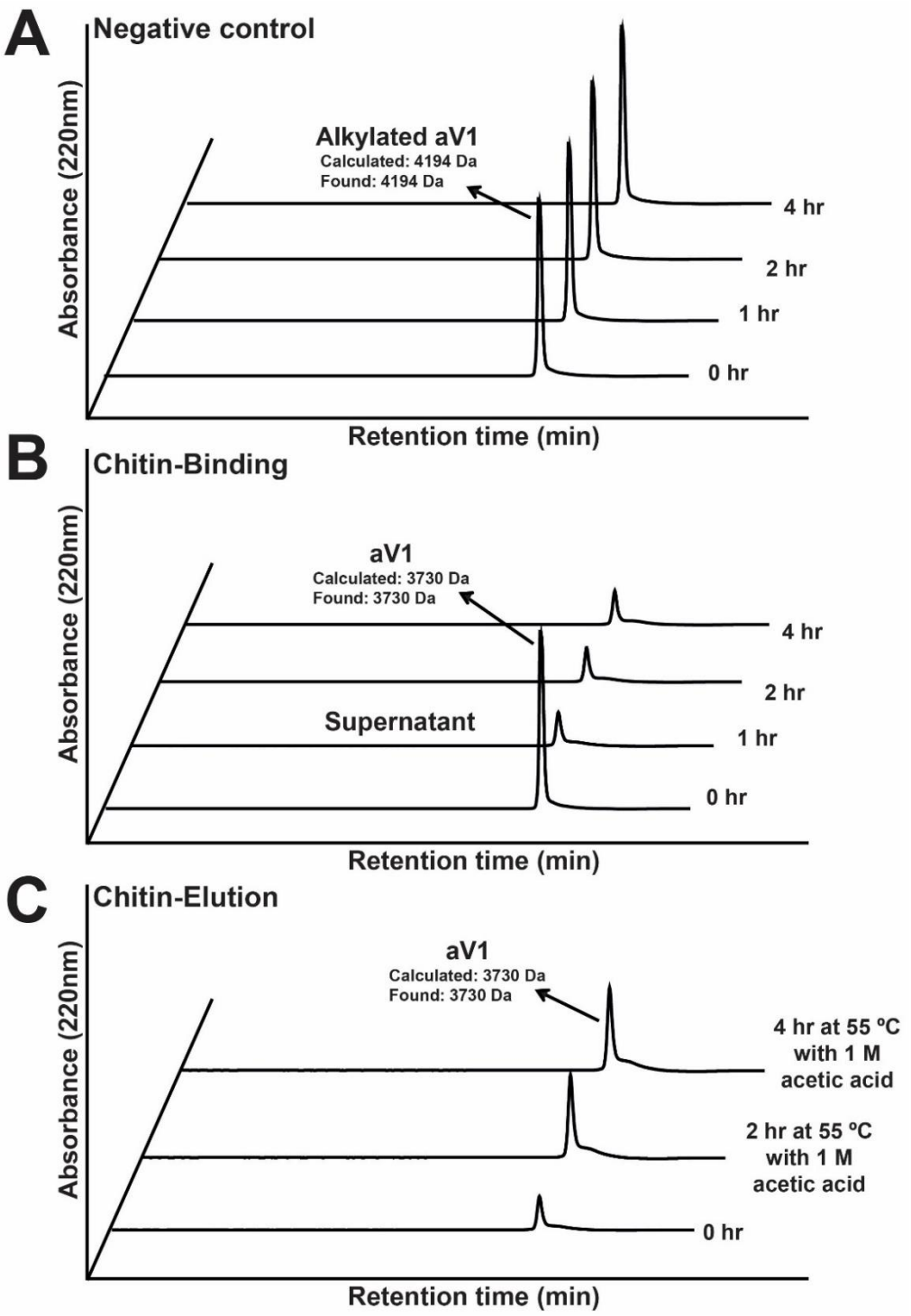


Figure 5.8: Chitin-binding activity of aV1. **(A)** Negative control of S-reduced and S-alkylated aV1 **(B)** native aV1 chitin-binding profile **(C)** native aV1 chitin elution profile. Absence of native aV1 in the supernatant suggests aV1 have chitin-binding properties.

2.8 Anti-fungal activity of avenatides

The disc diffusion assay was employed to assess the susceptibility of four phyto-pathogenic fungi strains, namely, *Alternaria alternata*, *Curvularia lunata*, *Fusarium oxysporum* and *Rhizoctonia solani* to aV1. The fungi strains were inoculated onto potato dextrose agar and incubated at 25 °C up to 72 h until a radial colony forms. Avenatides with different concentrations (2.5, 5 and 10 mg/mL) were loaded onto the paper discs and placed at the growing end of each mycelium. All the fungi strains showed the formation of an arc-shape inhibition zone, suggesting that aV1 displayed anti-fungal activity (Figure 5.9).

The micro-broth dilution assay was performed to calculate the half-maximal inhibitory concentration (IC_{50}) of avenatide aV1 against the four phyto-pathogenic fungi strains. The fungal spores (1×10^5) were seeded onto a microtiter plate and incubated with varying concentrations of aV1 for 24 h at 25 °C. The cells were fixed and stained with crystal violet after incubation. The absorbance was measure at 570 nm, and the IC_{50} was calculated based on the dose-response curve (Figure 5.9). The IC_{50} of aV1 ranged from 53.72 µM to 239 µM, depending on the strain of fungi tested.

The bright-field microscopy facilitated the detection of morphological changes to the fungi strains treated with aV1. The fungi strains were co-incubated different concentration of aV1 (0.3, 0.03, 0.003 mg/mL) for 24 h at 25 °C. Fungi treated with aV1 showed growth retardation, swollen hyphal tips and short branching compared to the control (Figure 5.10). Overall, these data indicate that av1 displayed anti-fungal activity and inhibit fungal growth in a dose-dependent fashion.

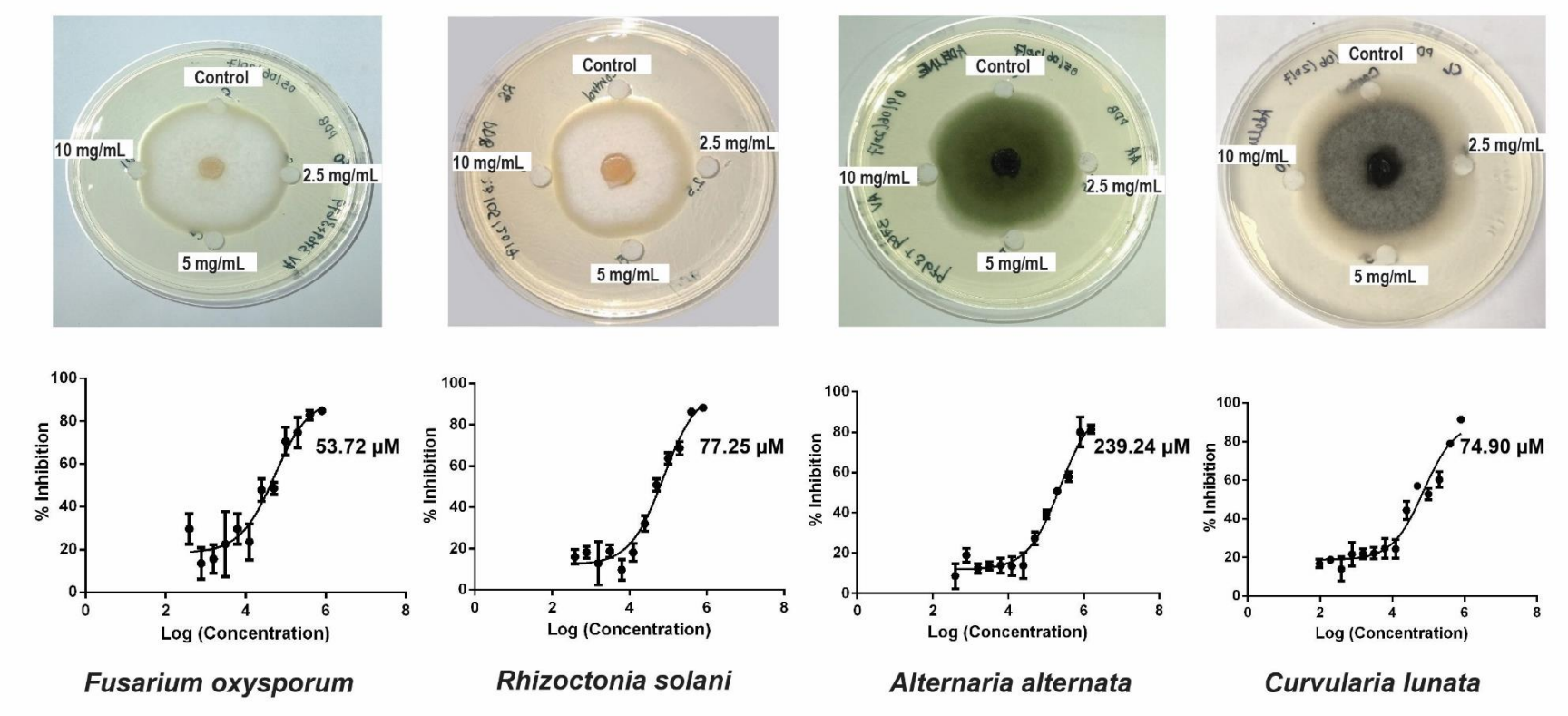


Figure 5.9: Anti-fungal activity of avenatide. Fungal inhibition of aV1 against *Fusarium oxysporum*, *Rhizoctonia solani*, *Alternaria alternata* and *Curvularia lunata*. Formation of arc-shaped inhibition zones in the disc diffusion assay indicated susceptibility of fungi towards avenatide. The IC_{50} was calculated based on the dose-response curve obtained from the micro-broth dilution assay.

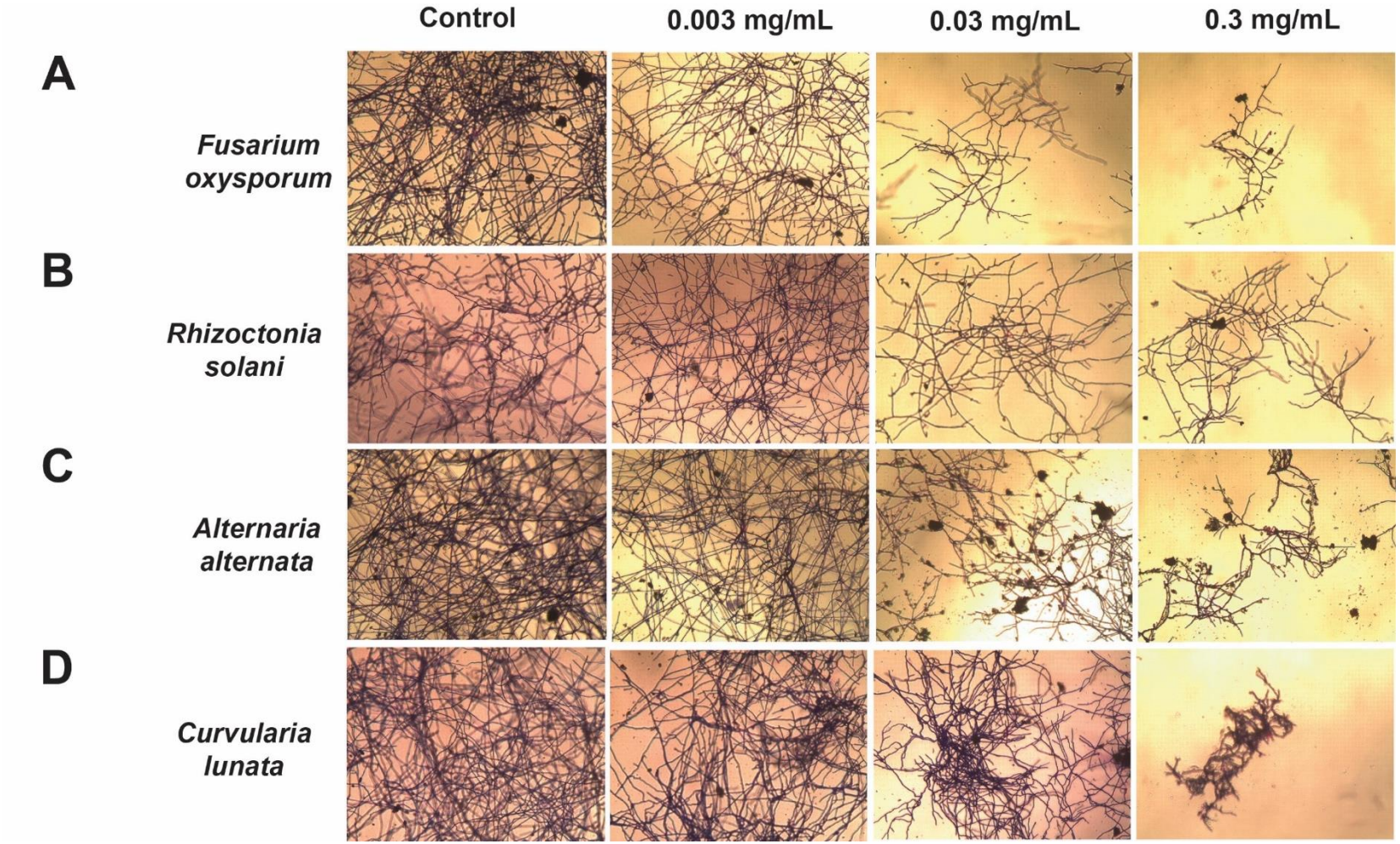


Figure 5.10: Bright field microscopy of fungal growth inhibition with aV1. **(A)** *Fusarium oxysporum*, **(B)** *Rhizoctonia solani*, **(C)** *Alternaria alternata* and **(D)** *Curvularia lunata* treated with 0.003–0.3 mg/mL of aV1. Formation of stunted hyphae ends indicated that avenatide inhibits hyphal growth at the ends of the fungal mycelial.

2.9 Metabolic stability of avenatides

Peptide-derived drugs are generally not considered the prime candidate for developing orally active drug due to its low bioavailability and susceptibility to enzymatic degradation. To determine the metabolic stability of avenatides, aV1 was subjected to heat at 100 °C, acid for 2 h, trypsin or carboxypeptidase A for 6 h. The presence and extent of degradation of the peptide were assessed using MALDI-TOF MS and the UPLC chromatographs. Based on the HPLC chromatogram, more than 85% of the native aV1 was retained for each of the treatment (Figure 5.11). In contrast, the S-alkylated aV1 gradually degraded after treatment with heat, acid or enzymes with more than 60% degradation. The average half-life of S-alkylated aV1 for all treatments is 2.5 hours. In contrast, the average percentage of native aV1 remaining at t_{120} timepoint for all treatments is 89.7%. Therefore, this data showed that the cross-linked disulfide bonds in aV1 prevented the degradation of the peptide and demonstrated that aV1 is stable against thermal, acidic and proteolytic degradation.

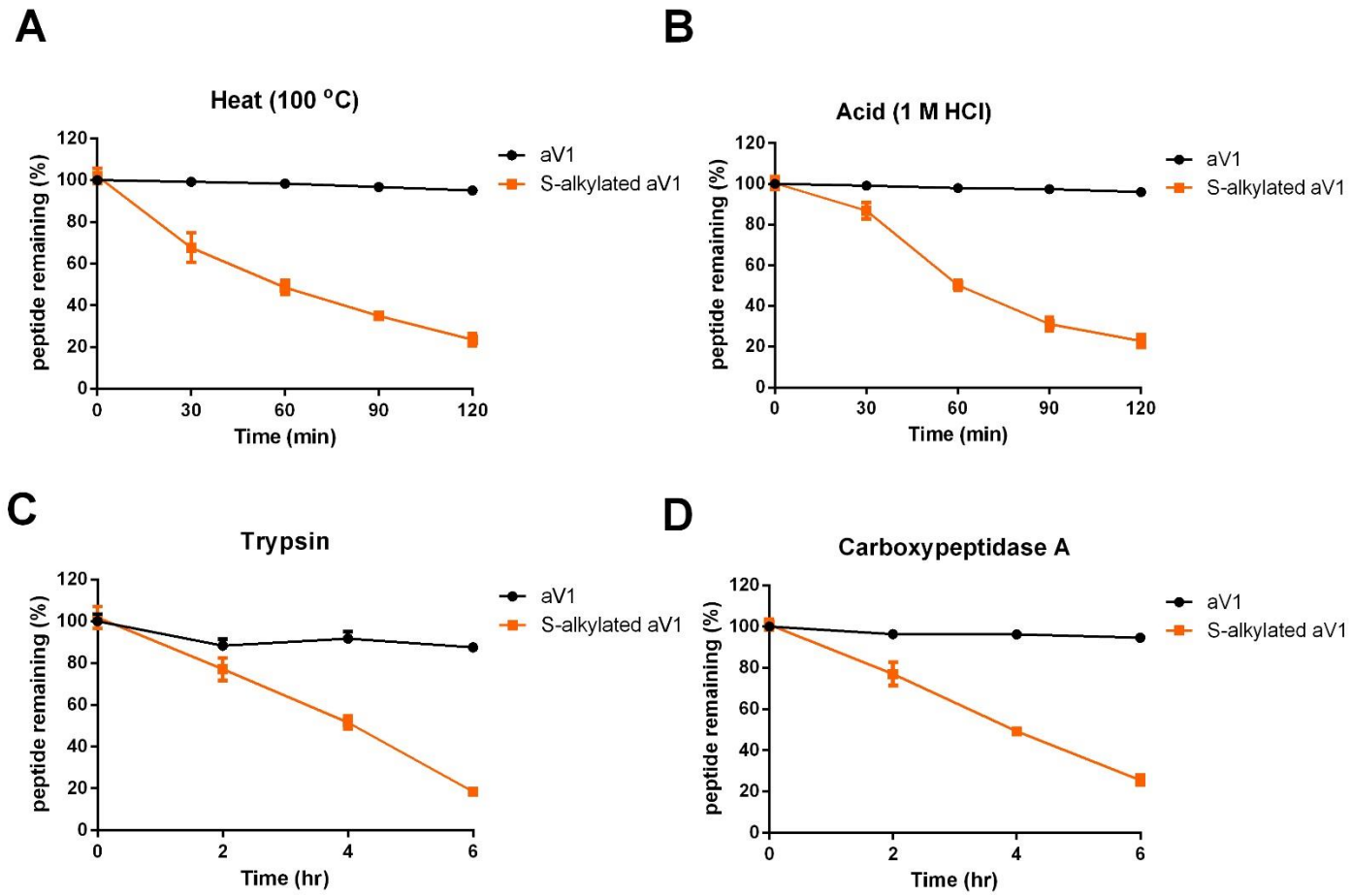


Figure 5.11: Metabolic stability test of avenatide. **(A)** Heat stability of aV1 at 100 °C for 2 h **(B)** Acid stability of aV1 incubated in 1 M HCl for 2 h **(C)** Trypsin stability of aV1 incubated for 6 h. **(D)** Carboxypeptidase A stability of aV1 incubated for 6 h. The percentages of peptide remaining were calculated from the area under the graph of the UPLC chromatographs.

2.10 Cell migration properties of avenatides

The scratch assay was used to investigate the extent of cell migration in C2C12 mouse myoblast cell line. This cell line was selected for this experiment because it is robust and easy to handle. The cells were treated with 10 µM of aV1 for 6 h in serum-free medium, and the migration distance was measured. Figure 5.12 and 5.13 showed that cells treated with aV1 had a greater migration distance (244 ± 24 µM) compared to the untreated control (208 ± 20 µM). Therefore, this data suggests that aV1 promote mild wound-healing activity in C2C12 mouse myoblast cell line.

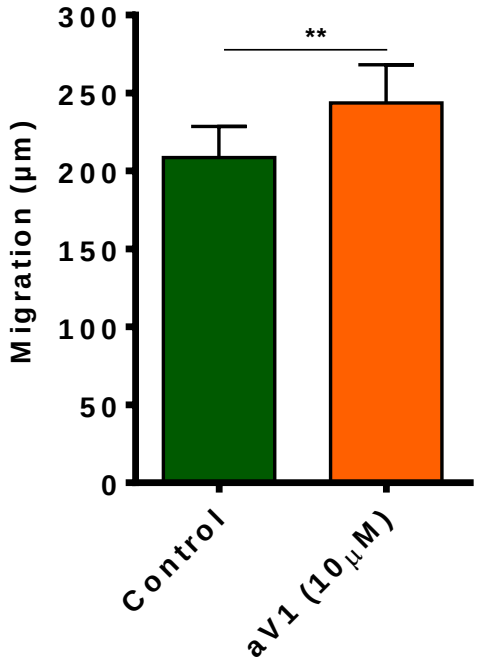


Figure 5.12: Cell migration activity of aV1 in C2C12 cells. C2C12 cells were treated with 10 µM of aV1 for 6 h in serum-free medium. All result results were expressed as ± S.E.M (n=3). *p<0.0.5 compared to control.

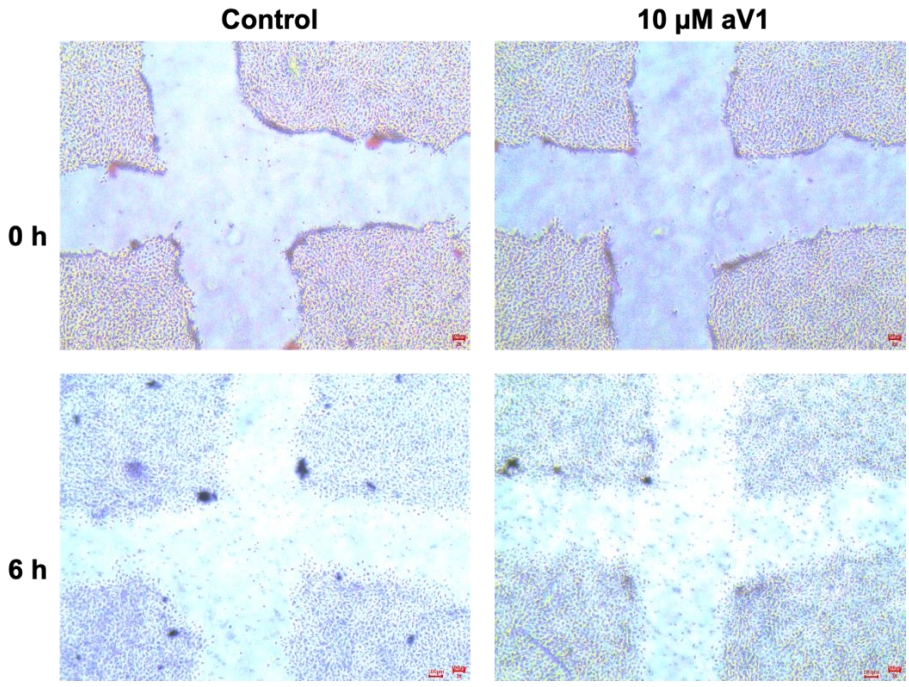


Figure 5.13: Microscopy images of the C2C12 cell scratch assay of aV1. At time point 0, the images were taken immediately after the scratch has been made. After 6 h of incubation, the control and cells treated with 10 µM of aV1 were imaged again. Images were taken under 4X objective lens.

3. Discussion

This study reports the discovery and characterization of nine novel 8C-chitin-binding hevein-like peptides termed as avenatides, isolated from *A. sativa*. Avenatides consist of 38–40 amino acids. Avenatides were detected at both proteomic and transcriptomic level. Avenatides shared sequence homology with the other 8C-chitin-binding HLPs such as Avesin A [109], vH1 [53], mO1[52], gB1 [54], hevein [81], Fa-AMP1 [100] and Pn-AMP1 [96]. Structural analysis of aV1 showed three disulfide bonds with a knotted-topology and an additional disulfide linkage located at the C-terminal. Transcriptomic analysis showed that the avenatides precursors comprise a three-domain architecture, including a signal peptide, a pro-mature domain and a short C-terminal tail that is different from the other 8C-chitin-binding HLPs. The disulfide-constrained structure of aV1 confers high stability against thermal, acid and enzymatic degradation. In addition, aV1 contained a chitin-binding domain which promotes chitin-binding interaction. Avenatide aV1 inhibited the growth of *A. alternata*, *C. lunata*, *F. oxysporum* and *R. solani* in a dose-dependent manner with an IC_{50} range of 53.72 µM to 239 µM. Avenatides exhibit mild wound-healing properties."

3.1 Sequence comparison with 8C-chitin-binding HLPs

Sequence comparison of avenatides with reported 8C-chitin-binding HLPs showed that the cysteine-spacing were absolutely conserved. The serine (Ser15), glycine (Gly21) and tyrosine (Tyr26) residue were conserved in all 8C-chitin-binding HLPs. A sequence logo of aligned avenatides and 8C-chitin-binding HLPs was generated

to show the sequence conservation and occurrence of amino acid residue at each position (Figure 5.14).

Based on the sequence logo, the 8C-chitin-binding HLPs consist of six intercysteine loops. None of the loops showed absolute sequence conservation. The C-terminal showed higher sequence conservation compared to the N-terminal. Intercysteine loop 1 has the longest loop compared to loop 6, which is the shortest loop. The size of intercysteine loop 2, 3 and 4 are fairly conserved, with four residues in loop 2, five residues in loop 3 and five residues in loop 4. In contrast, the intercysteine loop 1 varies from five to eight resides, three to five in loop 5 and three to four in loop 6, suggesting the high molecular diversity within the 8C-chitin-binding HLP subfamily. The residues involved in chitin-binding are found in loop 3 and 4 where the serine is absolutely conserved. The aromatic amino acid residues in loop 3 showed variations, whereas the tyrosine in loop 4 is conserved. Such extensive variations within the 8C-chitin-binding HLP subfamily suggest that these peptides emerged through divergent evolution, leading to functional diversity. This phenomenon was reported in WAMP, a 10C-chitin-binding HLP, isolated from *T. kiharae.* Substitution of Ser20 to Gly20 in the chitin-binding site reduced binding affinity to chitin, and the presence of an additional Ser36 led to a gain-of-function which inhibits fungal metalloproteases [149]. Similarly, genetic variants were observed in CRPs like allotides from the CKAI family [59], and cliotides from the cyclotide family [316]. Such natural variants occur as an evolutionary defense response to pathogen invasion to enhance the plant survivability and adaptation.

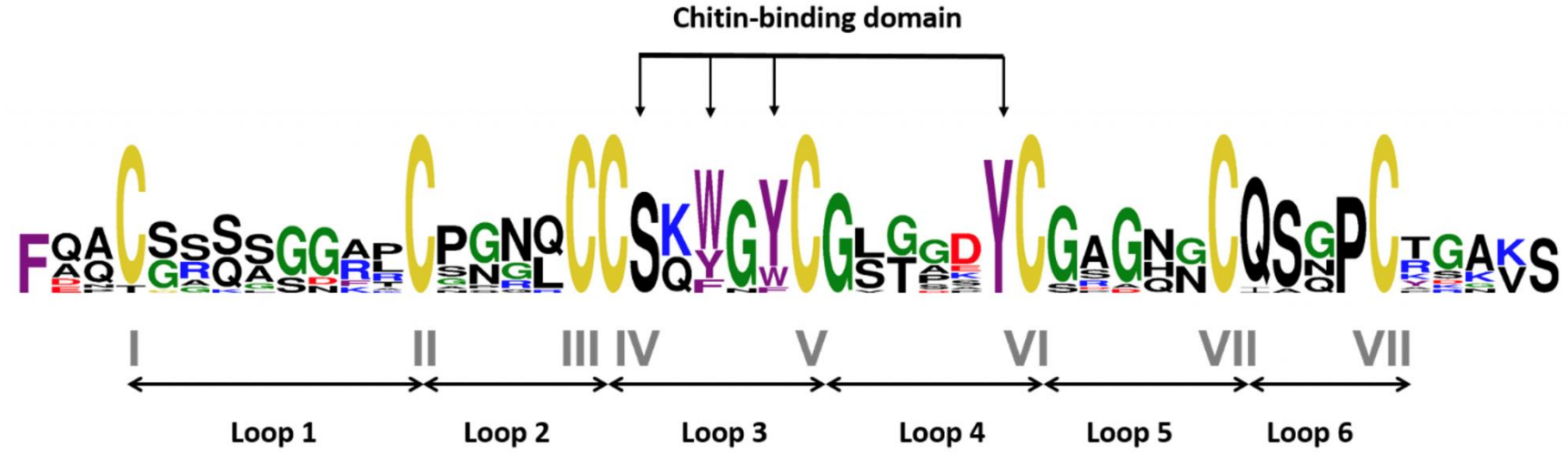


Figure 5.14: Sequence logo of aligned avenatides and reported 8C-chitin-binding-hevein-like peptides. Based on the cysteine pattern, the sequence can be divided into six loops. Loop 3 and 4 contain the conserved chitin-binding domain which is indicated by the arrows.

3.2 Bioprocessing of avenatides and 8C-CRPs

The biosynthetic pathway of avenatides revealed a three-domain precursor, comprising ER signal peptide, a pro-mature domain and a C-terminal tail. The precursor peptides are targeted to the ER, where the signal peptidase (SPaseI) cleave the signal peptide and subsequently the cleavage of the C-terminal tail to release the mature domain. The architectural comparison of avenatides and reported 8C-CRPs revealed biosynthetic diversity. Generally, all chitin-binding HLPs comprise a three-domain precursor, including; a signal peptide, a mature domain and a C-terminal tail as reported in mO1 [52]. However, within the 8C-chitin-binding HLPs subfamily, biosynthetic diversity was observed (Figure 5.15A). Avenatides comprised a pro-mature domain, which consists of four tandem repeats that are processed to individual mature domains, detected at the proteomic level. To our best knowledge, this unique feature of a pro-mature domain in 8C-chitin-binding HLPs has not been reported. In contrast, hevein, the first chitin-binding HLP discovered, also contained the classical three-domain precursor with a significantly long C-terminal tail (144 amino acid residues) that encodes for a Barwin domain [81] (Figure 5.15A). The Barwin domain is a protein domain that facilitates the chitin-binding HLPs to fungal cell wall by co-expressing the chitin-binding and Barwin domains, it increased the binding efficiency of hevein to chitooligosaccharides [81].

In the 8C-non-chitin-binding HLP subfamily, they comprised a three-domain precursor, which is different from the 8C-chitin-binding HLPs. Although they shared the same tandem cysteine-motif of C-C-CC-C-C-C-C, the precursor of ginsentide

had a pro-peptide prior the mature domain and the absence of a C-terminal tail [57] (Figure 5.15A). Interestingly, the 8C-thionins and lybatides shared the same precursor arrangement as the 8C-chitin-binding HLPs. In thionins, the C-terminal tail is significantly longer compared to lybatides [63, 67]. In contrast, the 8C-defensin comprised of a two-domain precursor consisting of a signal peptide and a mature domain without a C-terminal tail [382]. Such architectural diversity may have led to the functional difference in the 8C-CRP family.

The unique precursor arrangement of avenatides led to a schematic comparison with chitin-binding proteins UDA1 [356] and WGA [357] isolated from *Urtica dioica* and *Triticum aestivum*, respectively (Figure 5.15B). UDA1 comprised of two chitin-binding domain with a significantly long C-terminal tail that contains a catalytic class I chitinase domain [105]. Interestingly, the precursor arrangement of WGA and avenatides is similar. Both precursors contained four tandem-repeated chitin-binding domains. However, unlike WGA, the hinge region of *av1* appeared to be longer, and the mature domains were cleaved during post-translation processing (Figure 5.15B). This raised the question of whether the composition of the hinge region play a role in determining the cleavage of tandem-repeated chitin domains during processing, differentiating chitin-binding HLPs from chitin-binding proteins.

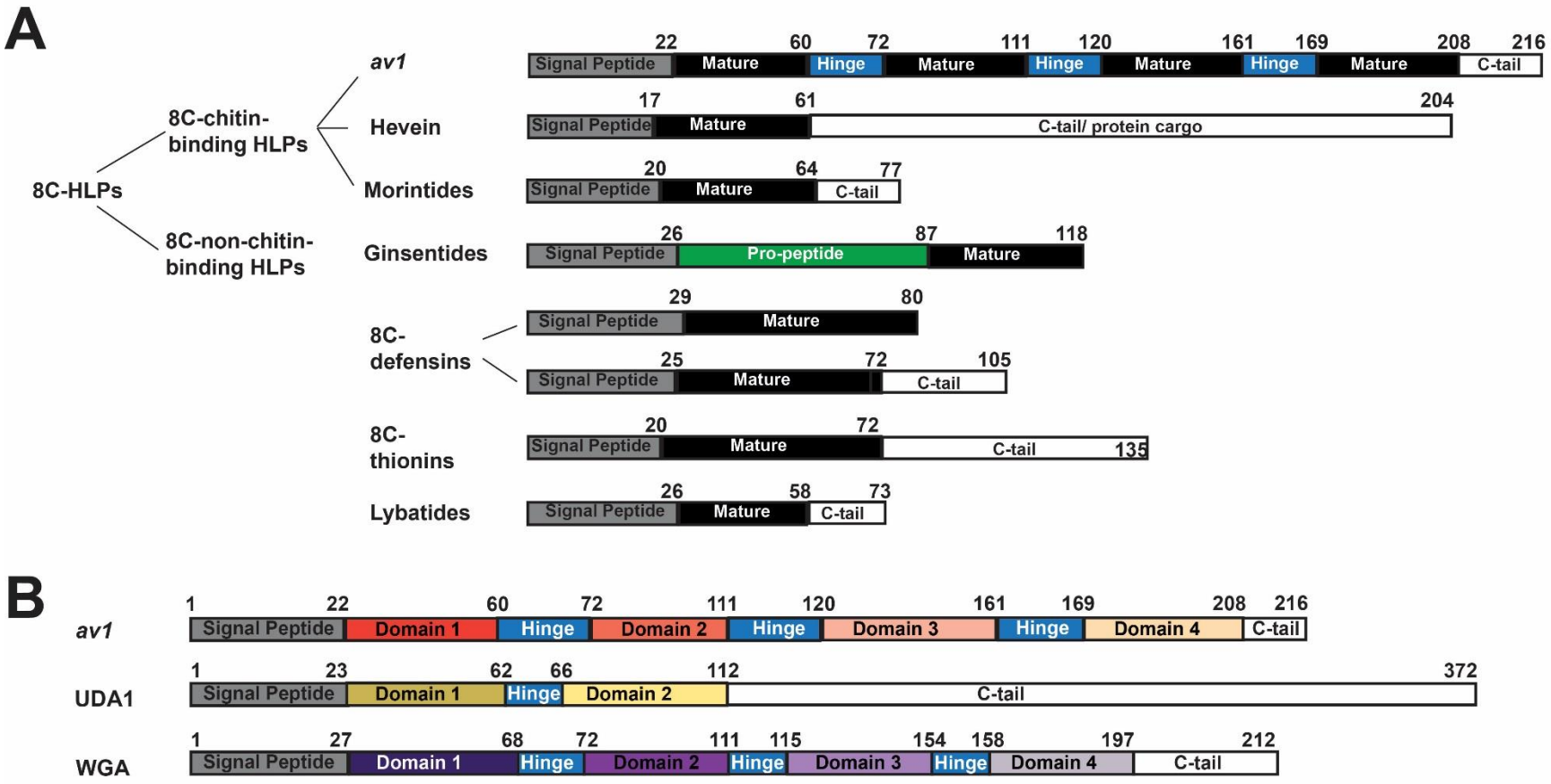


Figure 5.15: Schematic comparison of the biosynthetic architecture. **(A)** Avenatides and 8C-cysteine-rich-peptides **(B)** Schematic comparison of avenatide with chitin-binding proteins UDA1 and WGA. The average amino acid length for each domain are listed above each peptide accordingly.

3.3 Hinge region of tandem-repeated HLPs and chitin-binding proteins

The hinge region is defined as the region between two mature domains. The hinge region can be divided into two groups: cleavable and non-cleavable. To investigate the hinge region's characteristics, a BLAST search was performed on the NCBI and OneKP databases. A total of 60 chitin-binding proteins and 183 tandem repeats/chitin-binding HLPs with protein cargo were identified. The analysis of non-cleavable hinge regions from chitin-binding proteins revealed that the hinge regions are short (6 ± 4 amino acid residues) and proline-rich, which prevent endopeptidase degradation. In contrast, cleavable hinge regions from tandem repeats/chitin-binding HLPs with protein cargo are long (13 ± 3 amino acid residues) and are either Gly/Ser-rich or acidic, hence, susceptible to endopeptidase degradation. The hinge regions of avenatides are acidic and have an average of ten amino acid residues in length, which resembles the characteristics of a cleavable hinge region. Therefore, it is likely that the length and amino acid composition determine whether a hinge region is cleaved. A cleavable hinge region may be associated with evolutionary adaptation in plants. Chitin-binding proteins like WGA are also known as lectins, which exhibit cell agglutination and precipitation of glycoconjugates, causing hemagglutination [334]. The plants may have evolutionarily adapted to have a longer cleavable hinge to allow efficient production of the peptides in response to biotic stress such as pathogenic fungal invasion [383]. Taken together, this data suggests that the characteristics of the hinge sequence could be responsible in distinguishing chitin-binding HLPs from chitin-binding protein.

3.4 Distribution and occurrence of tandem-repeat in *planta*

A BLAST search was performed on the NCBI and OneKP databases to determine the distribution of tandem-repeat protein and peptides in *planta*. The search criteria include: (1) reported chitin-binding HLP or protein (2) possess at least one tandem repeat or a protein cargo (3) complete precursors sequence with a start and stop codon. Manual filtering was performed to remove: (1) signal peptide with less than ten residues (2) untranslated amino acids and (3) identical sequences from the same species. A total of 261 sequences were obtained from 211 plant species belonging to 96 different families (Figure 5.16). A total of five clusters were obtained, including: chitin-binding proteins, chitin-binding HLPs with Barwin domain, chitin-binding HLPs with chitinase domain, chitin-binding HLPs without protein cargo and others (Figure 5.16A). Interestingly, it appeared that tandem repeats in plants are more common than expected. The Poaceae and Solanaceae plant families accounted for 29% and 27% respectively, of the total sequences identified. The majority of the sequences identified from these families were chitin-binding proteins containing two to four tandem repeats with eight cysteines each (Figure 5.16B and 5.16C). Notably, most of these plant species belonged to the Poaceae family were cereal grains that include: *Oryza sativa* (rice), *Triticum aestivum* (wheat) and *Hordeum vulgare* (barley).

Genes that benefit the survival of plants are evolutionarily selected and retained [383]. Often, these genes may undergo duplication to protect the plants from environmental stressors [383]. A similar phenomenon was observed in Tiptop3 isolated from *M. cochinchinensis* that encodes for eight cyclic peptides that

possess potent trypsin inhibitors and insecticidal activity [384]. Collectively, our data showed the occurrence of tandem-repeated domains in *planta,* and since the chitin-binding domain likely plays a vital role in plant defense, it provides an evolutionary advantage and hence, is conserved.

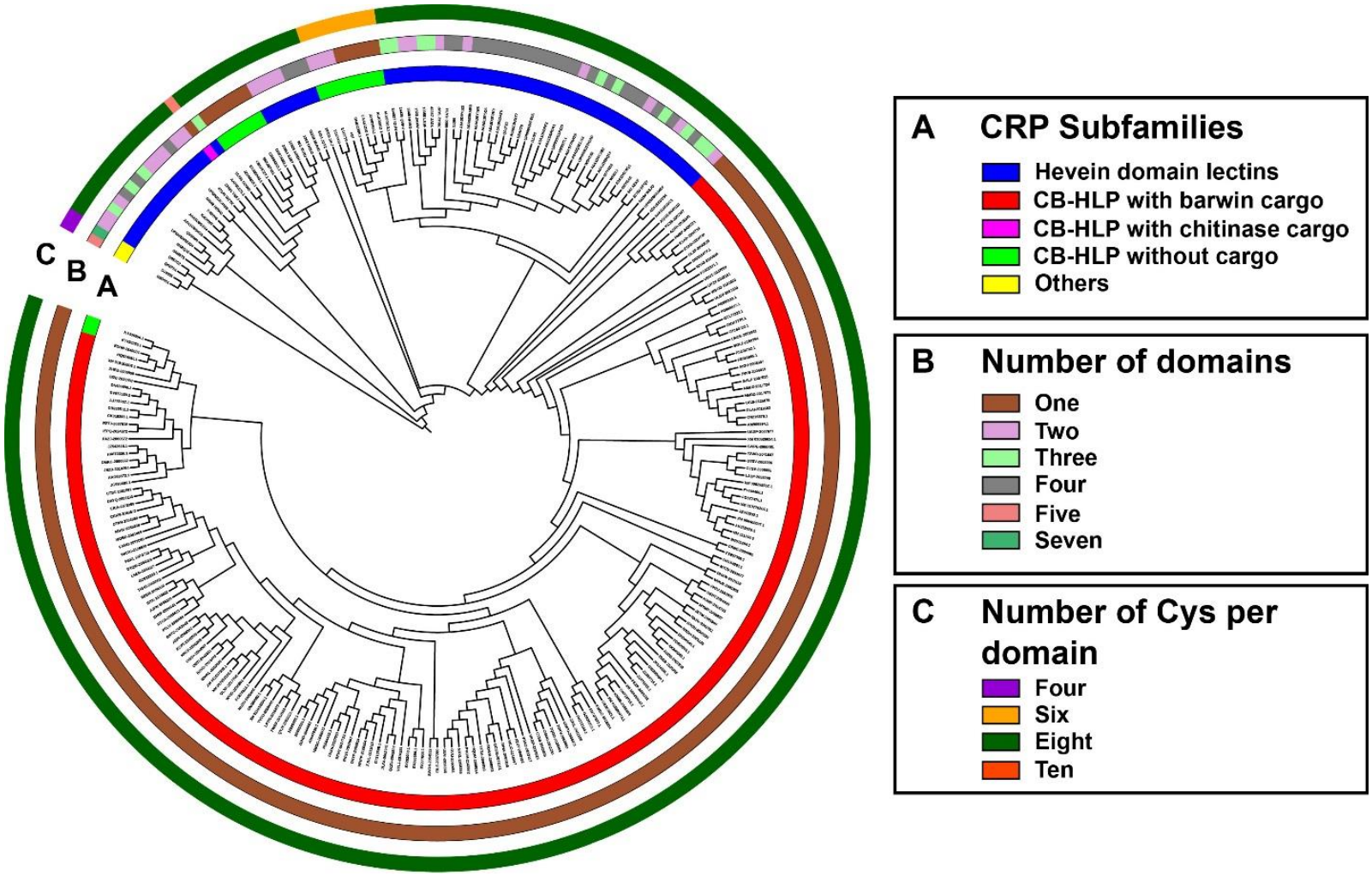


Figure 5.16: Phylogenetic tree of tandem-repeated proteins and chitin-binding hevein-like peptides. Classification based on **(A)** CRP subfamilies, **(B)** number of mature domains, and **(C)** number of cysteine residues per domain

This study has successfully isolated and characterized nine 8C-chitin-binding HLP, termed as avenatide aV1–aV9 from *A. sativa*. Sequence characterization of avenatides revealed that they contain 28–39 amino acid residues. The biosynthetic pathway of avenatide precursors adopts a three-domain architecture, which includes a signal peptide, a pro-mature domain and a C-terminal tail. Notably, the presence of four tandem repeats in the mature domain of avenatide precursors differentiate avenatides from the other reported 8C-chitin-binding HLPs. Phylogenetic tree illustrated the close evolutionary relationship between avenatides and chitin-binding proteins. Further analysis of the hinge region strongly suggests that the hinge region's composition may have a role in determining the cleavage of tandem repeats, distinguishing 8C-chitin-binding HLPs from chitin-binding proteins. NMR analysis showed that aV1 has a disulfide connectivity of CysI–CysIV, CysII–CysV, CysIII–CysVI and CysVII–CysVIII, forming a compact structure that endows aV1 with high metabolic stability against heat, acid and enzymatic degradation. In addition to the high metabolic stability, aV1 also exhibited anti-fungal activity against four phytopathogenic fungal strains with IC_{50} range of 53–239 μM and displayed mild wound healing properties by promoting cell migration in C2C12 mouse myoblast cell. Together, this data provided new insights into the sequence, structure and biosynthetic diversity of 8C-chitin-binding HLPs. Avenatides represents the first suite of 8C-chitin-binding HLPs that contained a pro-mature domain. As a group, they represent a novel subfamily of 8C-chitin-binding HLPs that contain tandem repeats. These finding suggests that avenatides may be suitable candidates to develop potential anti-microbial agents or as scaffolds for engineering metabolic-stable peptide therapeutics.

CHAPTER SIX

Eleutide: 8C-non-chitin-binding Hevein-like Peptides isolated from *Eleutherococcus trifoliatus*

1. Introduction

Eleutherococcus trifoliatus also formerly known as *Acanthopanax trifoliatus* or "Sam kar pei" is a ginseng-like thorny shrub that belongs to the Araliaceae family. There are 38 species of *Eleutherococcus* and 47% of it originated from China [385]. Attributing to its ginseng-like acitivity, the different parts of the *E. trifoliatus* have been used as a medicinal treatment in East and Southeast Asia. The leaves and roots of *E. trifoliatus* are widely used in traditional Chinese medicine (TCM) to treat inflammatory ailments like bruise, gout and other illnesses such as impotence and alleviating neuralgia [306]. The decoction of the bark is believed to help improve memory and treat common ailments such as cold, diarrhea, gastric pain and ulcer [307, 311]. Interestingly, this plant is also commonly used as an ingredient in *"lei cha"*, a popular traditional Hakka Chinese dish in Malaysia and is believed to exert tonic effects. Pharmacological studies has reported that *E. trifoliatus* exhibits anti-oxidative [309], anti-mutagenic [386], anti-ulcers [310] , anti-tumor [312] and anti-nociceptive activity [311].

Phytochemical studies on the active ingredients in *E. trifoliatus* mainly focus on small molecules such as continentalic acid [387], phenylpropanoid glycosides [388], triterpenoids [389], alkanes, lipids and steroids [390-392]. Thus far, no disulfide-constrained peptides have been reported in this plant species. Peptides and

protein-derived compounds are generally not considered as the active components in medicinal plants due to the misconception that they are unstable during the decoction process in TCM and easily degraded by endopeptidase like pepsin in the gastrointestinal tract after ingestion [34, 316, 393]. However, this misconception has been debunked with recent findings showing that a large number of cysteine-rich peptides (CRPs) that contained multiple disulfide bonds can confer metabolic resistance to chemical and enzymatic degradation [51-54, 56, 57, 63, 66, 394, 395].

CRPs are a class of peptides with a molecular range from 2–6 kDa and contained 3–5 disulfide bonds. Based on their cysteine-motifs and disulfide connectivity, CRPs can be divided into different plant families such as defensins, thionins, hevein-like peptides (HLPs) [34]. These families are further classified based on the number of cysteines. At present, there are three major classes of 8C-CRPs identified. This include 8C-defensins, 8C-thionins and 8C-HLPs [34]. The 8C-HLPs contained an evolutionarily conserved tandem cysteine motif (-CC-) and based on the presence or absence of a chitin-binding domain, they can be divided into two subfamilies namely, 8C-chitin-binding HLPs and 8C-non-chitin-binding HLPs. The 8C-chitin-binding HLPs have been widely reported in numerous plant species such as hevein from *H. brasiliensis* [81], Fa-AMP1 from *F. esculentum* [100], Pn-AMP1 from *P. nil* [96], morintides from *M. oleifera* [52], gingkotides from *G. biloba* [54], vaccatide from *V.hispanica* [53]. The key characteristic of the 8C-chitin-binding-HLPs is the presence of a chitin-binding domain, enabling the peptides to bind to chitin, the second most abundant biopolymer in the world and the major component of the fungal cell wall and exoskeleton of insects [396, 397]. In contrast, the 8C-

non-chitin-binding HLPs do not have a chitin-binding domain despite sharing the same tandem cysteine motif as the 8C-chitin-binding HLPs. Currently, the only report of 8C-non-chitin-binding HLPs are ginsentides, isolated from ginseng species [57]. Hence, this raise a research interest to identify and characterized potential 8C-non-chitin-binding HLPs from other plant species.

This study reports the isolation and characterization of a novel 8C-non-chitin-binding HLP from *Eleutherococcus trifoliatus*. The peptide was termed as eleutide eT1. Proteomics and transcriptomic analysis revealed that eleutide contained the same cysteine motif and disulfide connectivity as ginsentide. The precursor of eleutide showed a three-domain architecture comprising a signal peptide, a pro-peptide and a mature domain. The phylogenetic tree revealed that eleutide belongs to the same cluster with ginsentides, suggesting a close evolutionary relationship. The model structure of eleutide bears structural similarities with ginsentides, which confers them metabolic stability against thermal, acid and proteolytic degradation. In addition, cell-based assay revealed that eleutide eT1 is a non-cytotoxic peptide that promotes cell migration. Taken together, these bioactive compounds discovered from *E. trifoliatus* showed potential as a biotherapeutic in wound-healing and helped expand the 8C-non-chitin-binding HLP subfamily.

2. Results

2.1 Screening and isolation of eleutide

Small scale extractions were performed on the different parts (leaves, stem, roots, flower and fruits) of *E. trifoliatus* to screen for putative CRPs. The MS profiles revealed similar putative peptides within the molecular range of 3500–3800 Da (Figure 6.1). Due to the peptide abundance found in the leaves extract, it was selected for further characterization (Figure 6.1A). 1 kg of *E. trifoliatus* leaves was used for a large-scale extraction using milliQ water with a weight to volume ratio of 1:10. After multiple rounds of AEX-HLPC and RP-HPLC, the putative peptide designated as eleutide eT1 with relative monoisotopic mass $[M + H]^+$ of 3669.96 Da was successfully isolated (Figure 5.1A).

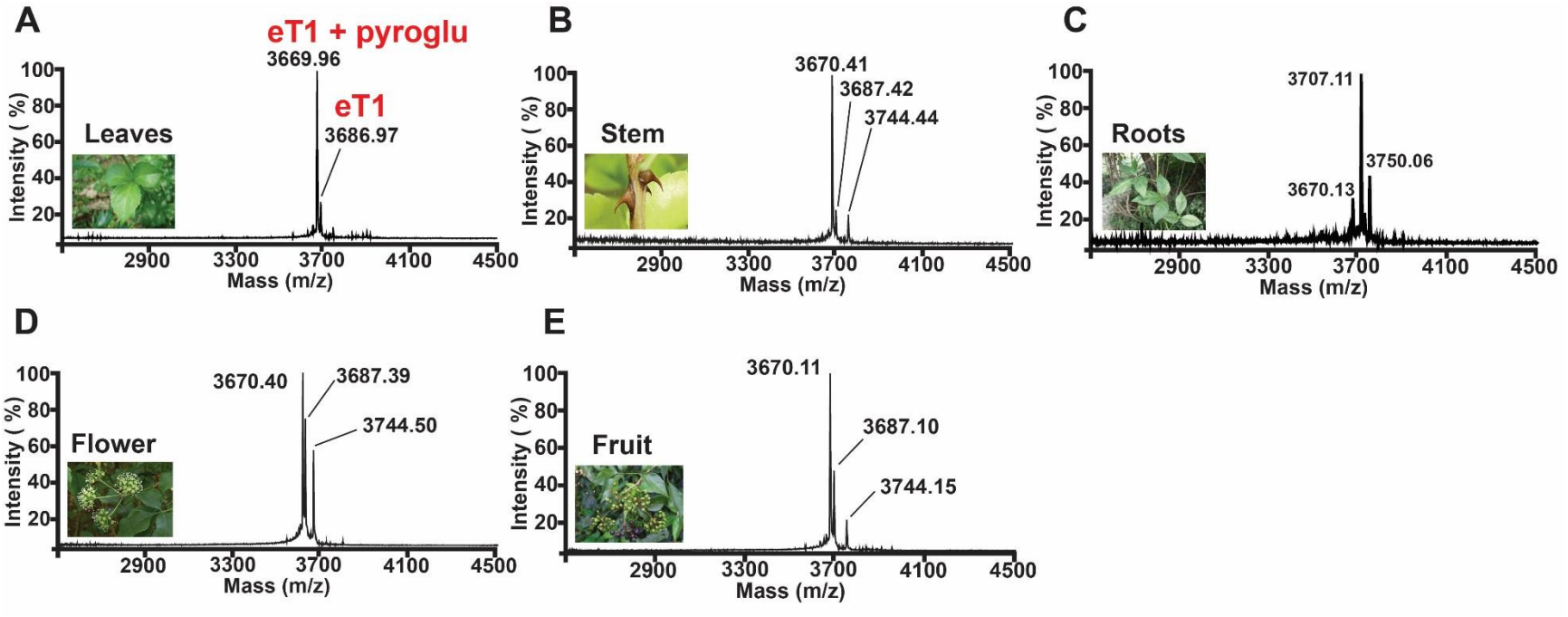


Figure 6.1: MS screening of the different parts of *E. trifoliatus*. (A) *E. trifoliatus* leaves **(B)** *E. trifoliatus* stem **(C)** *E. trifoliatus* roots **(D)** *E. trifoliatus* flower **(E)** *E. trifoliatus* fruits. Cluster of putative CRPs were observed in the molecular range of 3500–3800 Da.

2.2 S-reduction and S-alkylation of eleutide

S-reduction with DTT and S-alkylation with IAA were performed to determine the number of cysteines present in eleutide eT1. A mass increase of 464 Da indicated that eT1 is a CRP that contained eight cysteine residues (Figure 6.2).

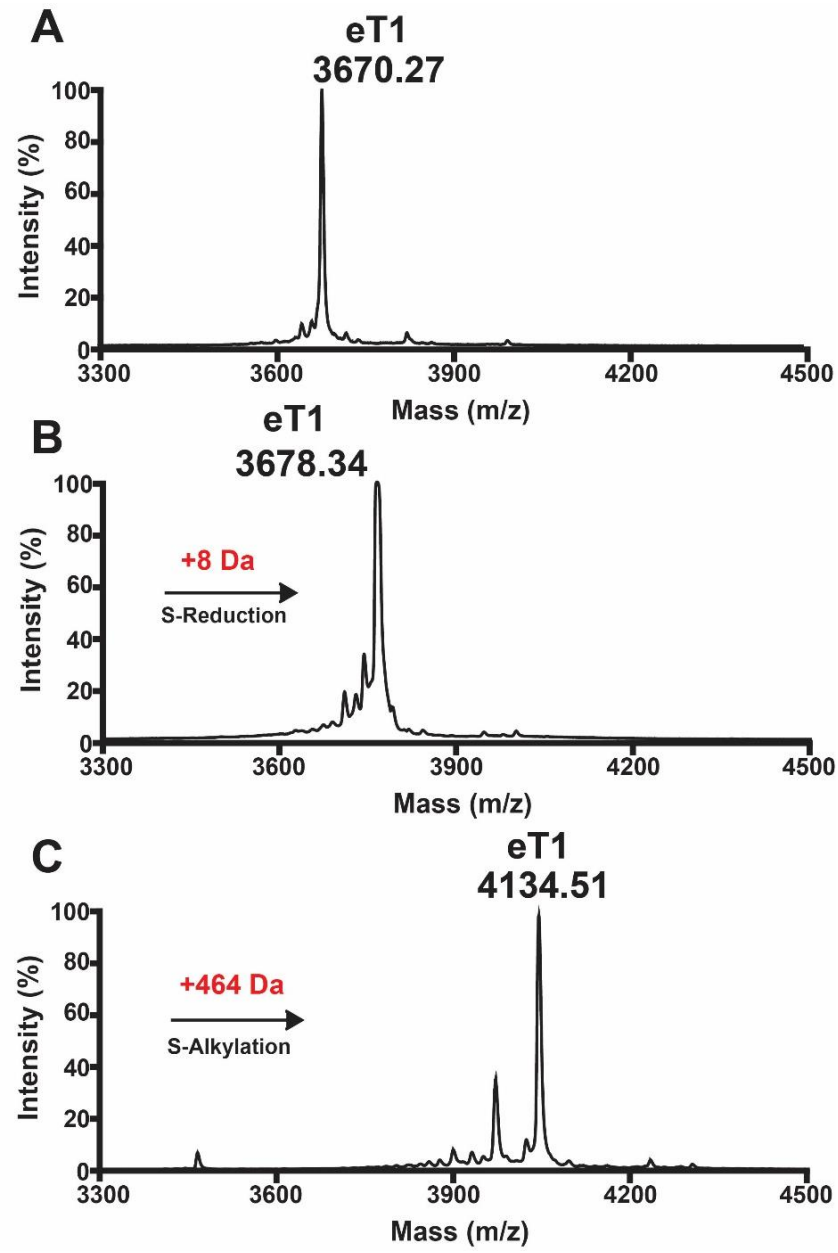


Figure 6.2: S-reduction and S-alkylation of eleutide eT1. **(A)** Native eT1, **(B)** S-reduction of eT1 and **(C)** S-alkylation of eT1. Mass increase of 464 Da indicated that eT1 contained eight cysteine residues.

2.3 MS/MS sequencing of eleutide

LTQ Orbitrap Elite MS/MS was used to elucidate the primary sequence of eleutide eT1. The S-alkylated eT1 was analyzed by nanospray MS/MS, and the sequence was deduced by analyzing the mass difference between the *b*- and *y*-series ions (Figure 6.3). The sequence of eT1 was deduced as ZVCSTAGQSCGGEQVCCDGCICNSKFIRPYCFGEC, where Z presents the N-terminal pyroglutamic acid. The primary sequence of eT1 revealed a cysteine-motif of C-C-CC-C-C-C-C that is similar to the 8C-chitin-binding HLPs. Overall, eleutide eT1 is 35 amino acid residues in length, which contained eight cysteine residue, two basic residues, three acidic residues and three aromatic residues, giving the peptide an overall net negative charge.

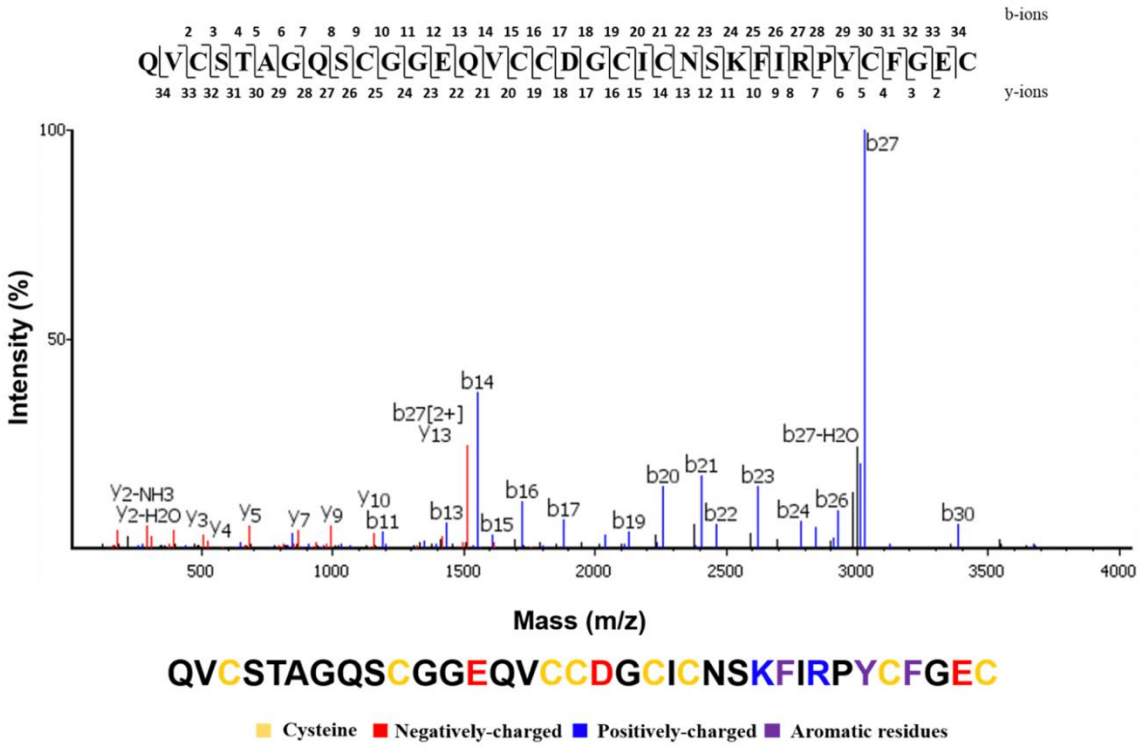


Figure 6.3: MS spectra of eleutide eT1 from LC-ESI-LTQ-Orbitrap in positive ion mode

2.4 Sequence comparison of eleutide

A tBLASTn was performed using eleutide eT1 as a query on the EST database from Genbank to identify putative eleutide-like homologs. The results revealed 123 putative eleutide-like homologs. For sequence alignment, eT1, ginsentides (TP1 and TP2) and ten homologs were selected and summarized in Table 6.1. Sequence alignment revealed a cysteine-spacing pattern of X_{1-4}-**C**-X_6-**C**-X_{5-7}-**CC**-**C**-X_{4-8}-**C**-X-**C**-X_{1-6}. Importantly, the N-terminal of eT1, cA1 and cA2 contained a pyroglutamic acid, and all sequences ended with a cysteine at the C-terminal. Such sequences found at both the N- and C-terminal render these peptides resistant to degradation by endopeptidase. A total of six intercysteine loops were observed. None of the loops were absolutely conserved. However, the loop size of loop 1 and 4 were absolutely conserved. Loop 5 has the largest loop whereas loop 4 has the smallest loop. The loop 5 of eT1 was longer than the other peptides and contained three aromatic amino acid resides (Phe26, Tyr30 and Phe32). Eleutide eT1 shared the highest sequence similarity of 54.5% with eC1 from *Eleusine coracana* and has the lowest sequence similarity of 35.3% with oS1 from *Oryza sativa*. Collectively, this data showed the sequence diversity and expanded the library of putative 8C-non-chitin-binding HLPs in different plant families.

Table 6.1: Sequence alignment of the mature peptide sequences of eleutide, eleutide-like homologs and ginsentide

Accession No.	Peptide	Species	Amino acid sequence	Mass (Da)[1]	Charge[2]	pI	Method[3]	Similarity (%)
Loop			I II III IV V VI					
This work	eT1	*E. trifoliatus*	--QVCSTAGQSC-GGEQ-VCCDGCICNSKFIRPYCFG---EC	3670.21	-1	4.68	T, P	100
WZYK-2086234	bT1	*B. trilobata*	----CLNGGGYCGSFTREACCYNCVCM----MAFCV-----C	3109.70	0	5.95	T	42.3
GT020922.1	cA1	*C. arabica*	QEGECSPLGEPC-AGNPWGCCPGCICIWQ-LTDRCIG---NC	3914.48	-3	4.00	T	50.0
GR983294.1	cA2	*C. arabica*	QEGECSPLGEAC-AGNPWGCCPGCICIWQ-LTDRCVG---NC	3874.42	-3	4.00	T	52.9
CXSJ-2102568	eC1	*E. coracana*	----CIPMGGFC-LFNLRGCCGSCGCL----AGFCWRDASSC	2439.09	+1	7.70	T	54.5
CO499109.1	gH1	*G. hirsutum*	----CIAKGGFC-LFDPTSCCRPCGCL----AGWCYNIDHDC	3573.16	-1	5.30	T	38.7
CO497769.1	gH2	*G. hirsutum*	----CIAKGRFC-LFDLTSCCRPCGCL----AGWCYNIDHNC	3687.35	-1	4.68	T	41.9
FG958636.1	oS1	*O. sativa*	----CLPAGGFC-LFRPMDCCGNCGCTYP--VGVCYR--SRC	3553.23	+2	8.22	T	35.3
CU472771.1	tC1	*T. cacao*	----CLSAGGFC-MFIPMDCCGNCGCLFP--MGFCYG--SGC	3396.09	-1	3.80	T	46.9
CU482998.1	tC2	*T. cacao*	----CPSAGGFC-MFNPMDCCGNCGCLYP--MGICYG--SGC	3362.97	-1	3.80	T	44.1
CA720771.1	tA1	*T. aestivum*	----CLPPGGFC-MFRPMDCCGNCGCLYP--AGVCY--GTRC	3496.20	+1	7.70	T	45.0
CN845637.1	TP1	*P. ginseng*	----CKSGGAWC-GFDPHGCCGNCGCLV----GFCY--GTGC	3053.10	0	6.68	T, P	46.9
DF555489.1	TP2	*P. ginseng*	----CKSSGAWC-GFDPHGCCGNCGCLV----GFCY--GTGC	3083.11	0	6.68	T, P	50.0

[1]Mass (Da): represents the experimentally found molecular weight. [2]Charge: represents the total charge of the molecule, and calculated by the sum of positive (lysine, arginine and histidine residues) and negative (glutamate and aspartate residues) charges. [3]Methods: represents the primary sequences obtained by (T) and/or proteomic (P) method. Cys are highlighted in yellow.

2.5 Putative biosynthesis pathway of eleutide

The precursor sequence of eleutide eT1, ginsentides and eleutide homologs shared the same three-domain architecture, which includes: an ER signal peptide, a pro-peptide and a mature domain, suggesting that eT1 is processed via the secretory pathway (Figure 6.4). Sequence comparison of all the precursor signal peptides showed that they are 22–24 amino acid in length. In addition, the C-terminal of the signal peptide ends with either an Ala or Gly, suggesting the cleavage site between the signal peptide and the pro-peptide is highly conserved. The pro-peptide of eT1 and ginsentides are significantly longer (61–63 amino acid residues) than the majority of the eleutide-like homologs (20–39 amino acid residues). The schematic comparison of eleutide with 8C-CRPs like 8C-thionins, 8C-defensins, 8C-chitin-binding HLPs and 8C-thionins are summarized in Figure 6.5. Eleutide contained a pro-peptide located between the signal peptide, and mature domain suggests that it does not share the same precursor architecture as the other 8C-CRPs like 8C-chitin-binding HLP, 8C-defensins and 8C-thionins. Moreover, the mature domains of the other 8C-CRP were significantly longer (44–52 amino acids in length) compared to eleutide and ginsentide (31–35 amino acid in length). Eleutide contained a pro-peptide and shared similar precursor arrangement as ginsentide. Therefore, they are classified under the 8C-non-chitin-binding HLP subfamily.

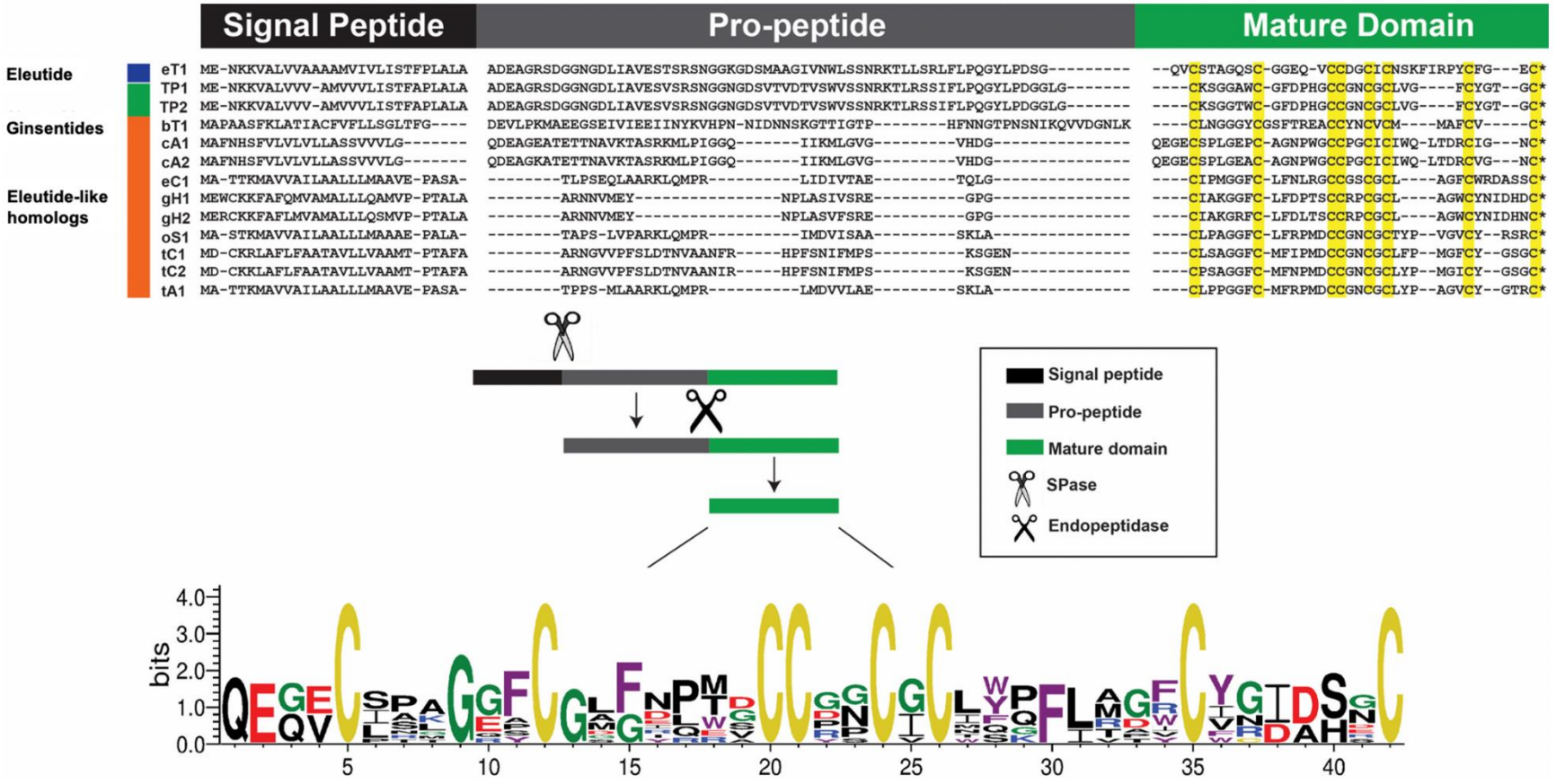


Figure 6.4: Gene alignment and putative biosynthesis pathway of eleutide, ginsentide and eleutide-like homologs. Eleutide, ginsentide and eleutide-like homolog precursors contained a three-domain architecture including a signal peptide, pro-peptide and mature domain.

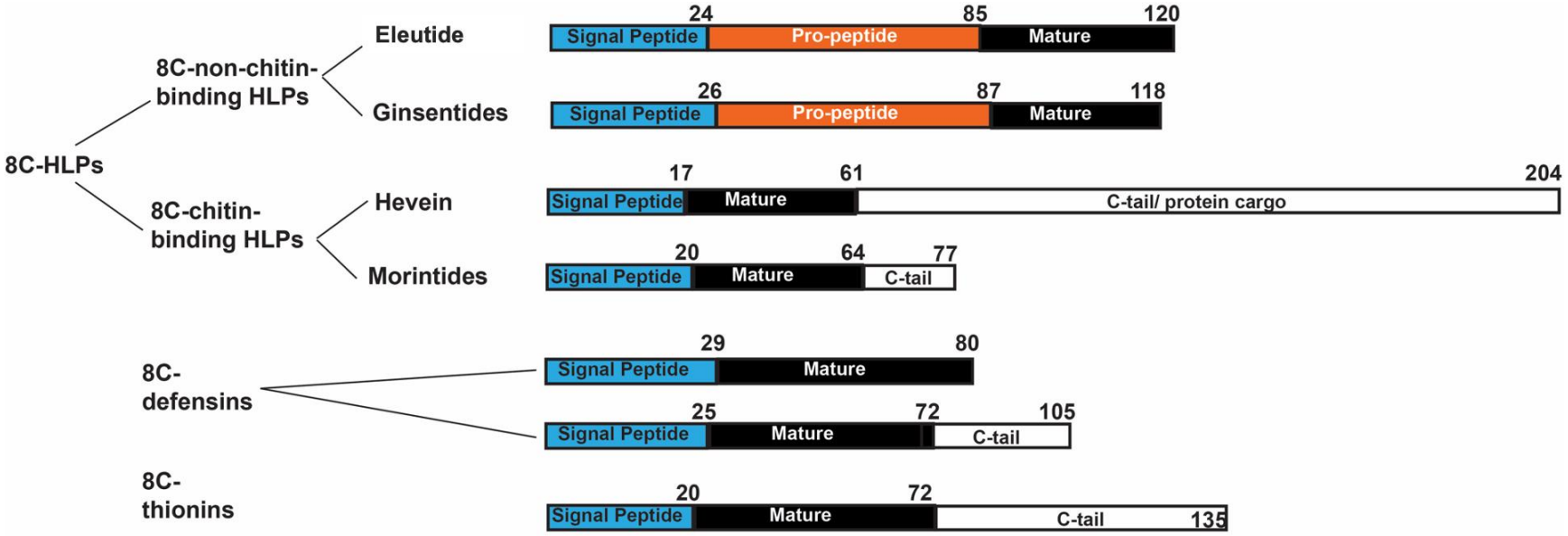


Figure 6.5: Schematic comparison of the biosynthetic architecture of eleutide with other 8C-CRPs. The average amino acid length for each domain are listed above each peptide accordingly.

2.6 Phylogenetic analysis of eleutide

A phylogenetic tree was constructed using the precursor sequences of eleutide, ginsentides, eleutide-like homologs and 8C-chitin-binding HLPs to study the evolutionary relationship of these peptides (Figure 6.6). Data analysis identified two major clusters. Cluster 1 consists of all the 8C-non-chitin-binding HLPs, whereas cluster 2 comprises of the 8C-chitin-binding HLPs. The evolutionary tree revealed that eleutide eT1 is closely related to ginsentides compared to the other 8C-non-chitin-binding HLPs. This could be attributed to the fact that eleutide and ginsentide belong to the same family of Araliaceae. The differential cluster between the 8C-chitin-binding HLPs and the 8C-non-chitin-binding HLPs is likely due to the presence of the chitin-binding domain. Although eT1 is clustered in the same cluster as ginsentides and the other 8C-non-chitin-binding HLPs, it appeared that eT1 diverged at the node independently, suggesting sequence diversity within the 8C-non-chitin-binding HLP subfamily. Taken together, the phylogeny tree provided insights into the evolutionary relationship of 8C-HLPs and revealed that eT1 belonged to the 8C-non-chitin-binding HLP subfamily.

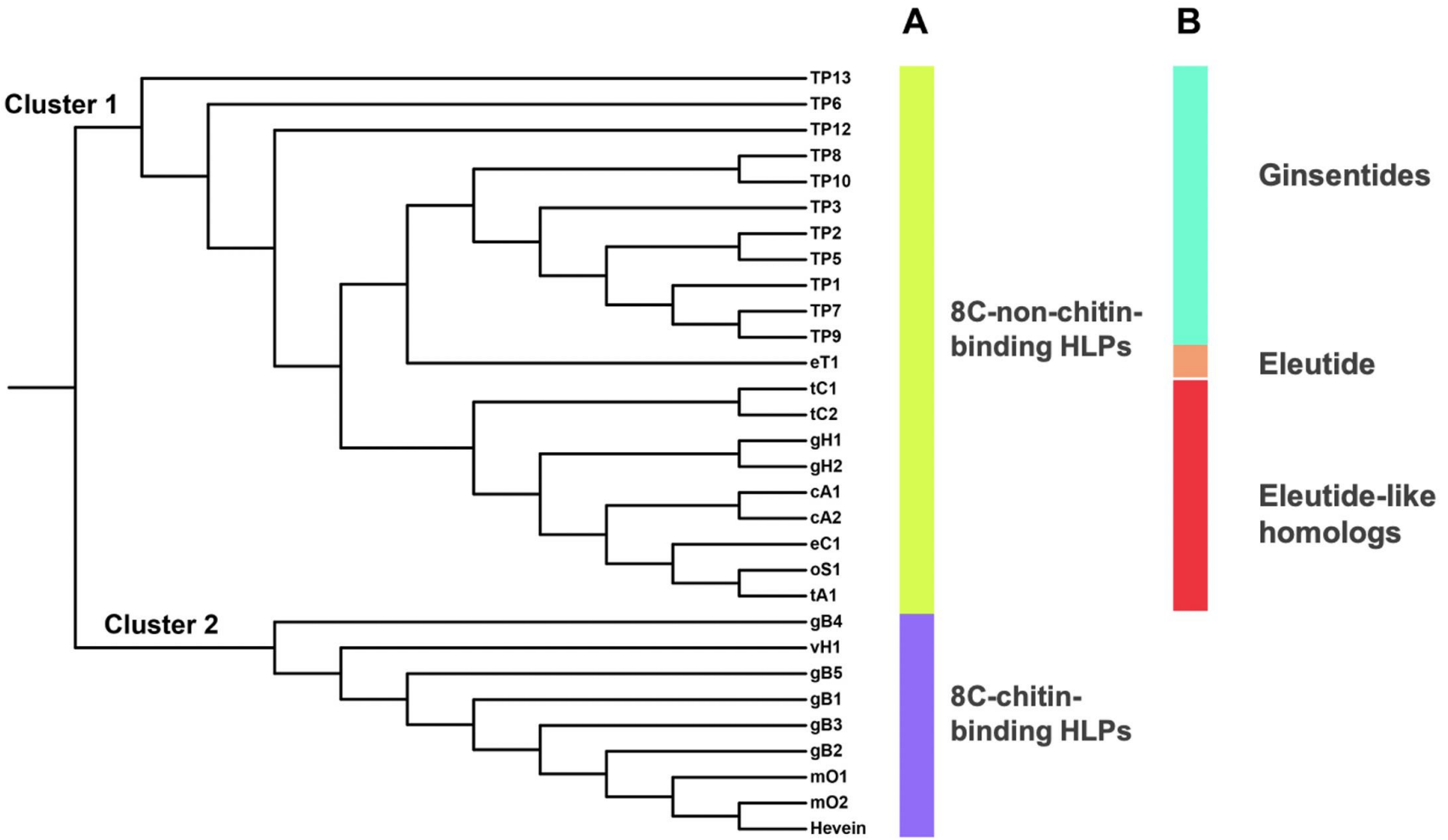


Figure 6.6: Phylogenetic tree analysis of eleutide, ginsentides, eleutide-like homologs and 8C-chitin-binding HLPs. The precursor sequences were aligned by MUSCLE. All the peptides are clustered separately based on **(A)** HLP subfamilies **(B)** Peptides within the subfamily.

2.7 Modeled structure of eleutide

Structural modeling of eT1 was performed using the online SWISS-MODEL platform (Figure 6.7A). Surface representation of eT1 revealed three acidic amino acid residues (Glu13, Asp18 and Glu34), two basic amino acid residues (Lys25 and Arg28) and three aromatic amino acid residues (Phe26, Tyr30 and Phe32) highlighted in red, blue and purple, respectively (Figure 6.7B). The best templates were based on the structures of coffeetide (Figure 6.7C) (PDB: 6JI7), Tau-theraphotoxin-Hs1a (PDB: 6CUC) and Asteropsin G (PDB: 2N3P) with a global model quality estimation score (GMQE) of 0.68, 0.66 and 0.64, respectively. The modeled structure of eT1 has two anti-parallel β-strands. β1 consist of residues Cys20–Cys22 and β2 consists of residues Cys31–Gly33.

Based on the generated model, it was evident that eT1 is an 8C-non-chitin-binding HLP with four disulfide bonds. Due to the high structural homology, the disulfide connectivity of eT1 was presumed to be the same as coffeetide (Figure 6.7D). The three disulfide bonds CysI–CysIV, CysII–CysVI and CysIII–CysVII forms a cystine-knot fold and the fourth disulfide bond CysV–CysVIII connects the β1 to the C-terminus. Similar to ginsentide (PDB: 2ML7) (Figure 6.7E), eT1 possesses a pseudocyclic structure in which both the N- and C-terminus contained disulfide bonds CysI–CysIV and CysV–CysVIII, respectively (Figure 6.7F). The structural alignment of eT1 with coffeetide and ginsentide gave an RMSB score of 0.71 and 2.55, respectively.

Coffeetide has three acidic amino acid residues (Glu2, Glu4 and Glu10), two aromatic amino acid residues (Trp17 and Trp27) and no basic amino acid residues

giving the structure an overall net negative charge. The acidic residues of coffeetides were located closer to the N-terminus, whereas the aromatic residues were located in the center of the structure. Ginsentide has one acidic residue (Asp11) and one basic residue (Lys2) located closer to the N-terminus, giving the structure an overall neutral charge. It was observed that intercysteine loop 5 of eT1 was longer compared to coffeetide and ginsentide. Additionally, the two anti-parallel β-strands in eT1, coffeetide and ginsentide comprised of only hydrophobic residues. Collectively, these findings showed that eleutide eT1 showed structural homology with coffeetide and ginsentide with a disulfide connectivity of CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII. Such a highly cross-linked disulfide structure with intramolecular hydrogen bonds can provide peptide stability against proteolytic degradation.

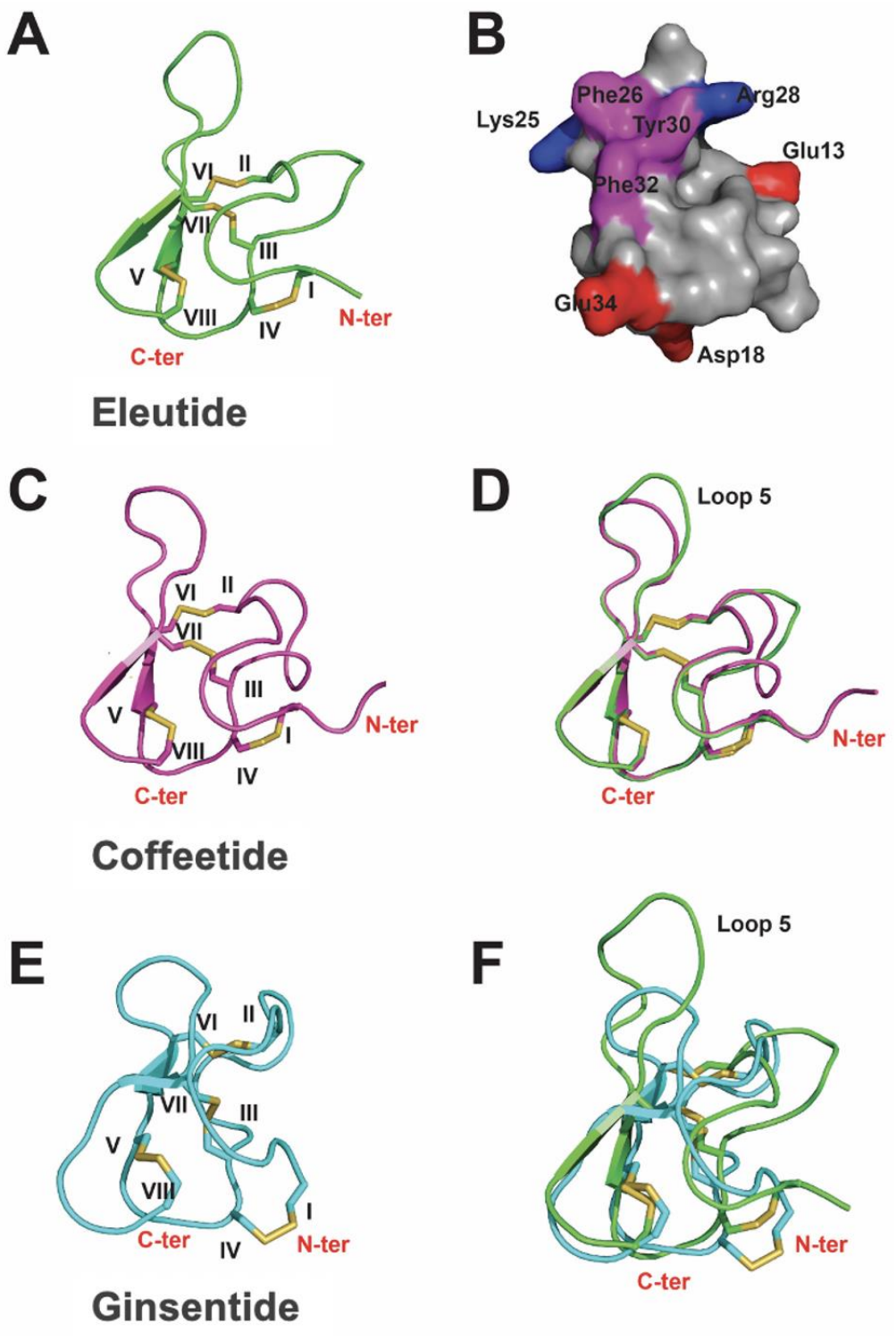


Figure 6.7: Modeled structure of eleutide eT1. **(A)** Cartoon representation eT1, **(B)** Surface representation of eT1 **(C)** Structure of coffeetide (PDB: 6JI7), **(D)** Structure alignment of coffeetide with eT1 **(E)** Structure of ginsentide (PDB: 2ML7) **(F)** Structure alignment of ginsentide with eT1. For surface representation, the acdic, basic and aromatic amino acid residues are highlighted in red, blue and purple, respectively.

2.8 Thermal, acid and proteolytic stability of eleutide

The metabolic stability of eleutide against heat, acid and proteolytic degradation was assessed by RP-UPLC. Eleutide eT1 was subjected to 95 °C or 1 M HCl for 2 hr and pepsin or aminopeptidase I for 6 h. Computation of the RP-UPLC chromatograms revealed that eT1 remain relatively stable to these treatments retaining more than 80% of the peptide after each treatment (Figure 6.8). In contrast, the S-alkylated eT1 showed more than 70% degradation for each treatment, suggesting that the linearized eT1 was susceptible to thermal, acid and enzymatic degradation. Furthermore, the average half-life of S-alkylated eT1 for all treatments is 2.5 hours. In contrast, the average percentage of native eT1 remaining at t_{120} timepoint for all treatments is 95.6%. Therefore, this data showed that the highly cross-linked disulfide bonds in eT1 coupled with the pyroglutamic acid at the N-terminal and a cysteine residue at the C-terminal prevented the degradation of the peptide.

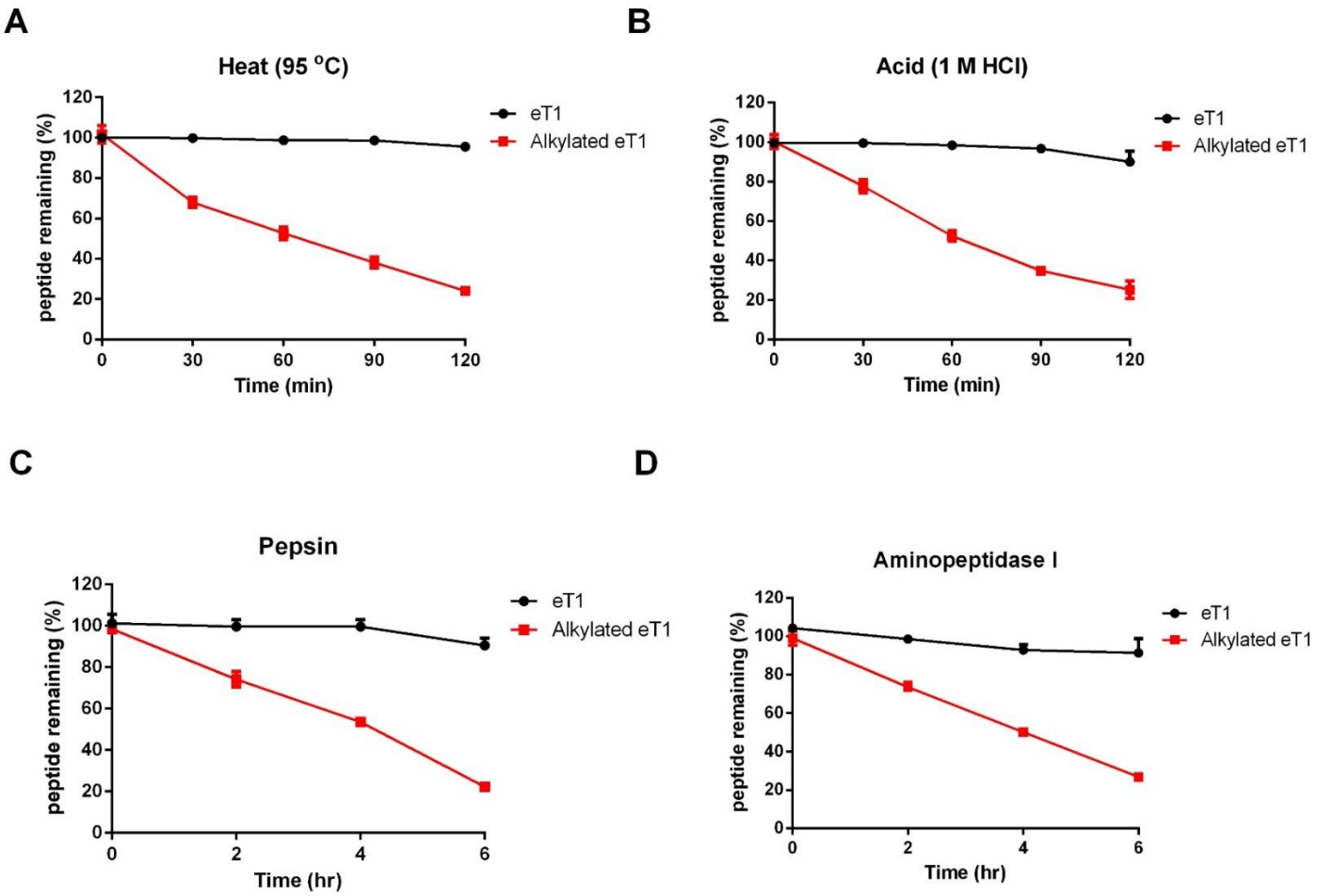


Figure 6.8: Metabolic stability test of eleutide **(A)** Heat stability of eT1 at 95 °C for 2 h **(B)** Acid stability of eT1 incubated in 1 M HCl for 2 h **(C)** Pepsin stability of eT1 incubated for 6 h. **(D)** Aminopeptidase I stability of eT1 incubated for 6 h. The percentages of peptide remaining were calculated from the area under the graph of the UPLC chromatographs.

2.9 Chitin-binding activity of eleutide

Aforementioned, the intercysteine loop 5 of eleutide contained three aromatic amino acid residues, Phe26, Tyr30, Phe32, and a serine reside (Ser24) which slightly resemble the chitin-binding motif, SXφGφCGX$_4\varphi$, where X represents any amino acids, and φ represents any aromatic residues. Therefore, to evaluate the chitin-binding properties of eleutide, eT1 was incubated with chitin beads for up to 24 h at 25 °C. Based on the RP-UPLC chromatogram, it showed that eT1 did not bind to the chitin beads and hence, confirming that it is an 8C-non-chitin-binding HLP (Figure 6.9).

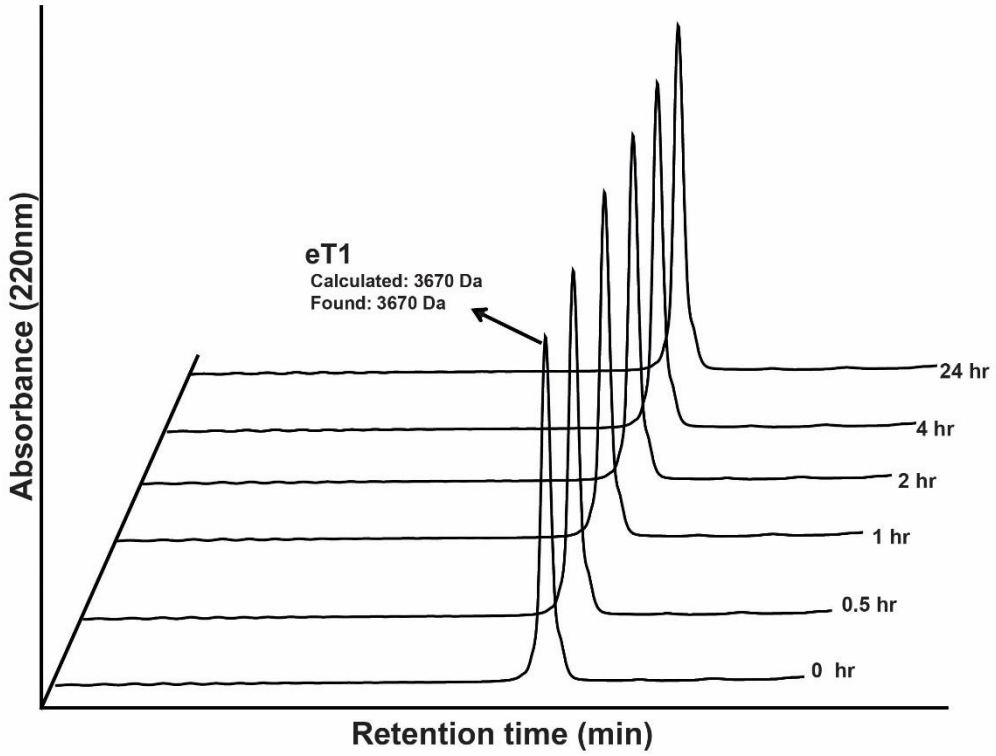


Figure 6.9: Chitin-binding activity of eT1. The presence of eT1 in the supernatant after 24 h incubation suggests that eT1 did not bind to the chitin beads.

2.10 Cell viability assay of eleutide

To examine the effects of eleutide eT1 in hypoxia stress, the rat cardio-myoblast H9c2 cells were employed and the procedure was followed according to Dutta *et al*. [398]. The cells were cultured with or without eT1 and incubated in a normoxic and hypoxic environment for 48 h before evaluating the cell viability via MTT assay. Figure 6.10A and 6.10B showed that 40 µM of eT1 increased in H9c2 cell viability by 15% and 9% in normoxic and hypoxic conditions, respectively. In the normoxic environment, cell viability increased with the increasing concentration of eT1, whereas in the hypoxic environment, cell viability showed a decrease of eT1 at concentrations up to 10 µM and increased in 20 µM of eT1. For the subsequent MTT assays, 20 µM of eT1 was used. Based on Figure 6.10C, cells treated with eT1 showed an increased cell viability of 11% and 2% after 24 h incubation in normoxic and hypoxic conditions, respectively compared to their respective control (PBS). Figure 6.10D illustrates the data normalized by the control PBS-normoxic at 24 h. Similarly, the same test was performed for 48 h (Figure 6.10E). In the normoxic condition, cell viability was increased by 5% compared to the PBS control. However, no significant increase in cell viability was observed in the hypoxic condition. This could be attributed to overcrowding of cells during prolonged incubation. Another reason could be due to the prolonged treatment, eT1 may have already been used up. Figure 6.10F shows the data normalized by PBS-normoxic at 48 h. Overall our data suggest that eT1 exhibits mild cytoprotective effect in the H9c2 cells.

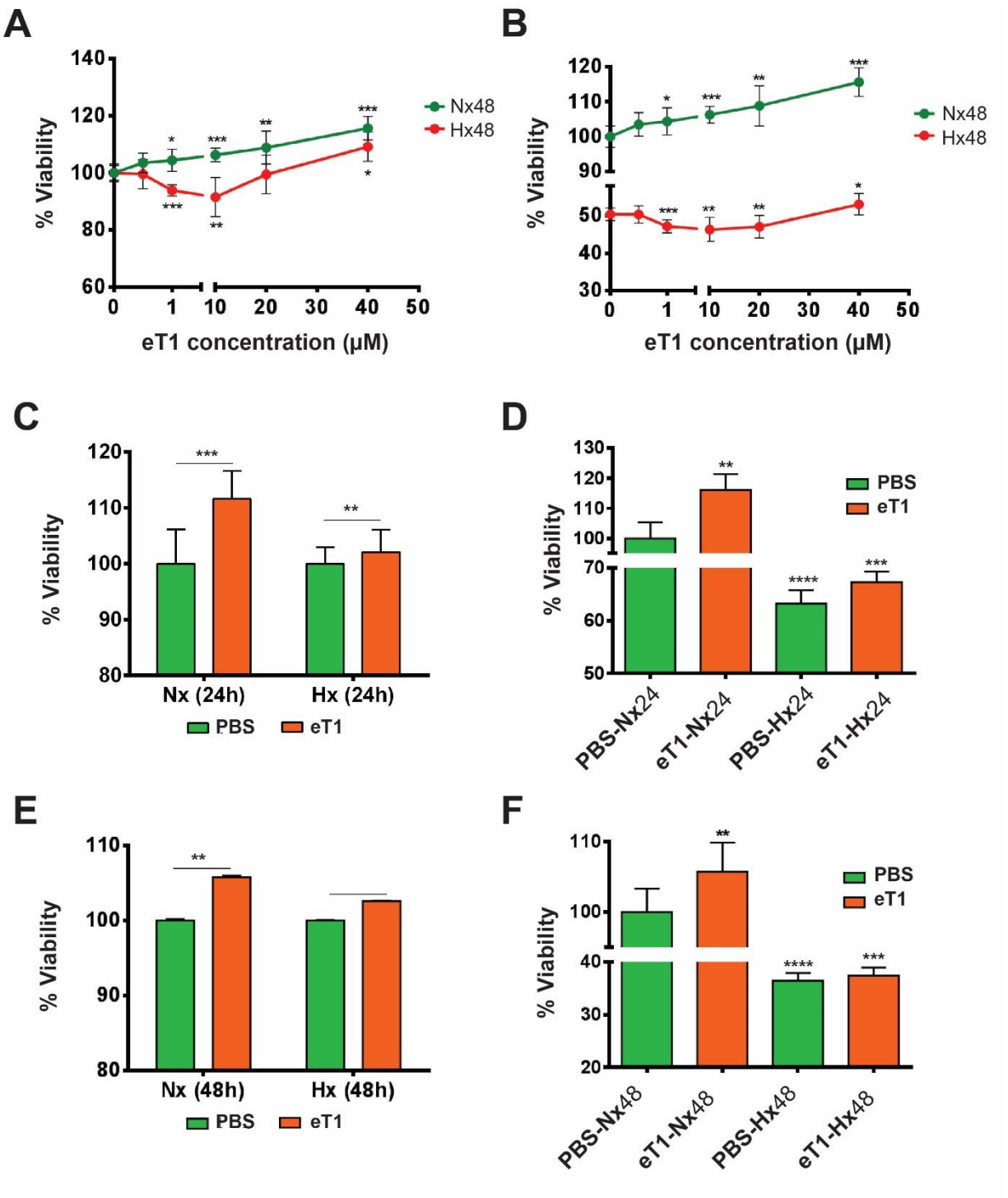


Figure 6.10: Cell viability assay of H9c2 cells with eT1. **(A)** Dose-response of eT1 on H9c2 cells **(B)** Dose response of eT1 on H9c2 cells normalized by normoxic 48 h **(C)** Relative viability of H9c2 cells upon 20 µM of eT1 treatment for 24 h **(D)** Relative viability of H9c2 cells upon 20 µM of eT1 treatment for 24 h normalized by PBS-normoxic 24 h **(E)** Relative viability of H9c2 cells upon 20 µM of eT1 treatment for 48 h (F) Relative viability of H9c2 cells upon 20 µM of eT1 treatment for 48 h normalized by PBS-normoxic 48 h.

2.11 Cell migration properties of eleutide

The cell migration properties of eleutide were assessed using the scratch assay. The C2C12 cell line was selected for this experiment because it is robust and easy to handle. The cells were co-incubated with eT1 (20 µM) for 6 h in serum-free medium, and the migration distance was measure after 6 h. Figure 6.11 and 6.12 showed the extent of cell migration. Cells treated with eT1 had a more significant migration distance (237 ± 20 µM) compared to the untreated control (197 ± 17 µM). Many studies have shown that EGF stimulates cell migration to promote wound recovery [359, 360]. For example, the findings of Lauand *et al.* show that the stimulation of EGF in A549 and HK2 cells promoted cell migration [359]. Therefore, we postulate that EGF would also promote cell migration in C2C12 cells. However further analysis is required to confirm this hypothesis. Collectively, our data suggests that eT1 promotes cell migration and hence, exhibits mild wound-healing properties.

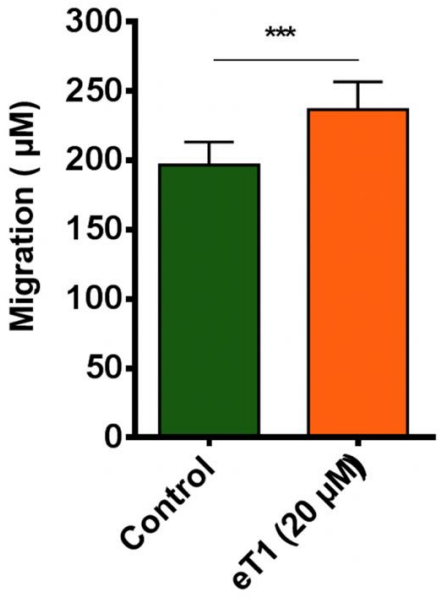


Figure 6.11: Cell migration activity of eT1 in C2C12 cells. C2C12 cells were treated with 20 µM of eT1 for 6 h in serum-free medium. All result results were expressed as ± S.E.M (n=3). *p<0.05 compared to control.

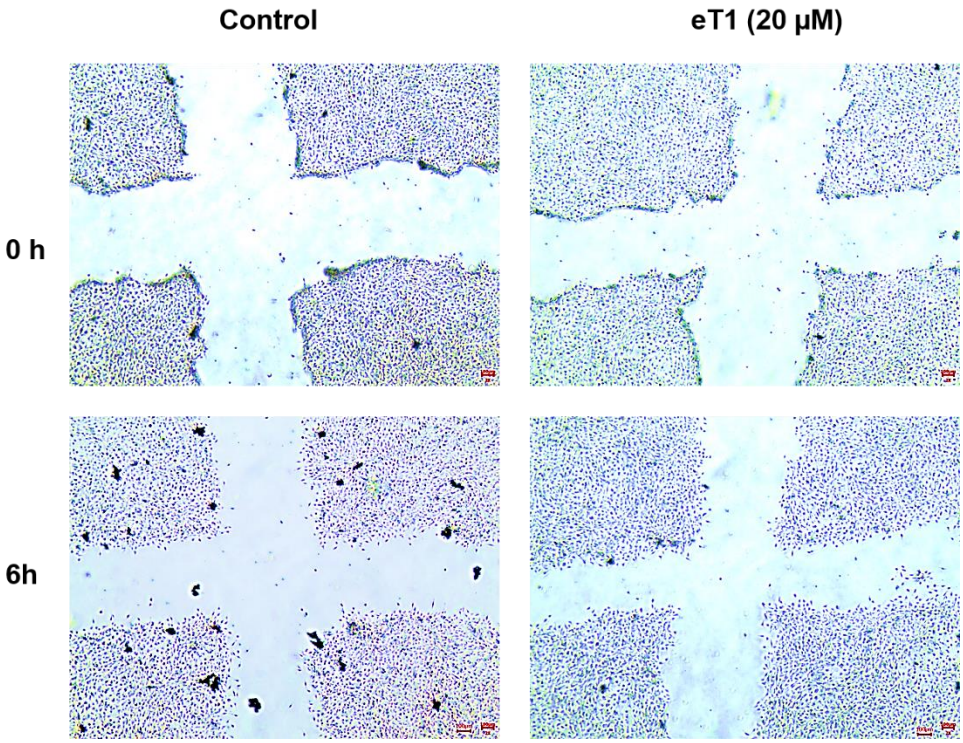


Figure 6.12: Microscopy images of the C2C12 cell scratch assay of eT1. At time point 0, the images were taken immediately after the scratch has been made. After 6 h of incubation, the control and cells treated with 20 µM of eT1 were imaged again. Images were taken under 4X objective lens.

3. Discussion

This chapter reports the discovery and characterization of an 8C-non-chitin-binding hevein-like peptide from *E. trifoliatus,* termed as eleutide. Eleutide eT1 consists of 35 amino acid residues, which are glutamine- and glycine-rich. Eleutide eT1 share sequence homology with ginsentide [57], and data mining revealed 123 putative eleutide-like homologs. Structural analysis of eT1 revealed that eT1 has a disulfide connectivity of CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII, forming a pseudocyclic peptide. Proteomic analysis revealed that the precursor of eT1 consist of a three-domain arrangement that includes an ER signal peptide, pro-peptide and a mature domain. The pseudocyclic structure of eT1 renders high metabolic stability against heat, acid and enzymatic degradation. Chitin-binding assay confirmed that eT1 is a non-chitin-binding HLP. Functionally, eT1 displayed cytoprotective effect to modulate cell viability against hypoxia-induced stress condition. It also promoted cell migration in C2C12 cells, which is a characteristic of wound-healing properties.

3.1 Sequence comparison of eT1 with ginsentide and 8C-chitin-binding HLPs

Plant CRPs can be divided into different families based on the types of cysteine-motif and disulfide connectivity [59]. HLPs are a family of CRPs that contained a conserved cysteine motif (C-C-CC-C-C) and can be further divided into two subfamilies, namely, chitin-binding HLPs and non-chitin-binding HLPs [34]. The 8C-chitin-binding HLP subfamily contain a conserved chitin-binding domain (SXφGφCGX$_4\varphi$), where X and φ represent any amino acids and aromatic amino

acid, respectively [52-54, 96, 100, 109]. On the contrary, 8C-non-chitin-binding HLPs do not have a chitin-binding domain, which was previously reported in ginsentides [57]. The primary sequence of eT1 showed that it belongs to the 8C-non-chitin-binding HLP family. Therefore, a sequence logo was generated to compare the sequences between eT1 and ginsentides (TP1–14). (Figure 6.13). Eleutide eT1 are glutamine- and glycine-rich. In contrast, ginsentides are proline- and glycine-rich. The acidic residues of eT1 are widely distributed in intercysteine loop 2, 3 and 6, whereas in gisentides, the acidic residues are located in the N-terminal. The N-terminal tail of eleutide is stabilized by a pyroglutamic acid, whereas ginsentides have a cysteine residue at the N-terminal tail. As such, this forms a pseudocyclic structure that confers the peptides' high metabolic stability [57]. Interestingly, the loops 5 and 6 of eT1 contained a serine and three aromatic amino acid residues (S$X\varphi X_3\varphi C\varphi$), which was initially presumed to resemble the chitin-binding domain in 8C-chitin-binding HLPs (Figure 6.13). However, the chitin domain is normally located in loop 3 and 4 of 8C-chitin-binding HLPs [81]. Furthermore, the chitin-binding assay showed that eT1 did not bind to chitin beads, confirming that eT1 is an 8C-non-chitin-binding HLP.

Pairwise alignment of eT1 with 8C-chitin-binding HLPs showed a low sequence identity (22.2–52.4%) and similarity (32.4–52.4%) (Table 6.2). This is because 8C-chitin-binding HLPs are basic and hydrophilic. In contrast, eT1 is acidic and hydrophobic. Interestingly, pairwise alignment of eT1 with ginsentides also showed low sequence identity (31.2–38.1%) and similarity (40.9–51.4%). Although eT1 and ginsentide belong to the Araliaceae plant family and shared a similar feature of the

cysteine motif **C**-X_6-**C**-X_{5-7}-**CC**$_2$-**C**-X_{4-8}-**C**-X-**C**-X_{1-6}**C** which comprise a -CC- and -CXC- motif, it appears that sequence variations may have contributed to the low identity and similar score. For example, the sizes of intercysteine loop 1, 3 and 4 are absolutely conserved, whereas the size of loop 2, 5 and 6 varies. The sequence variability between eT1 and ginsentides may suggests functional diversity within the 8C-non-chitin-binding HLP subfamily. However, to validate this hypothesis, further biological characterizations of the peptides are required to confirm that they indeed have different functions.

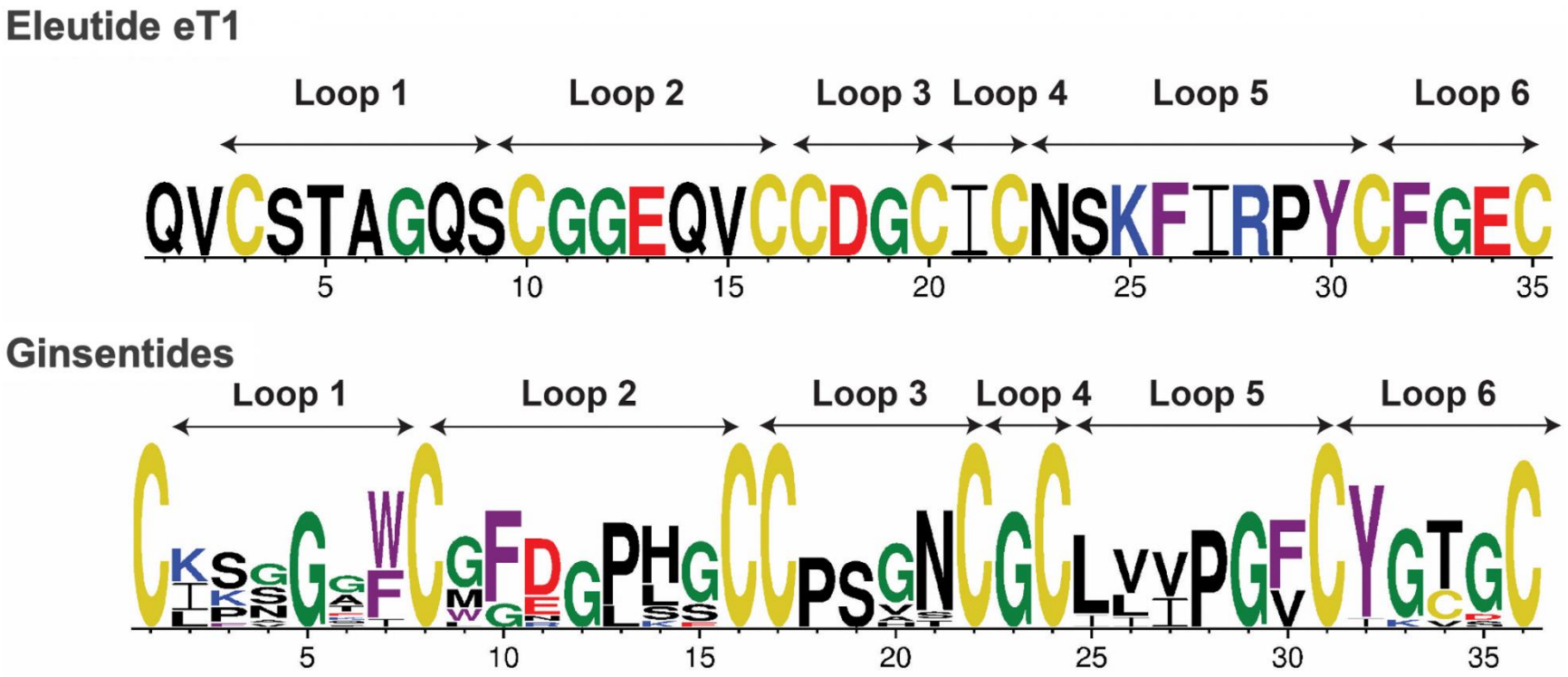


Figure 6.13: Sequence logo of eT1 and ginsentides. Based on the cysteine pattern, the sequence can be divided into six loops. The cysteine, glycine, acidic, basic and aromatic amino acid residues are highlighted in yellow, green, red, blue and purple, respectively.

Table 6.2: Pairwise alignment of eT1 with ginsentides and 8C-chitin-binding HLPs

Peptide	Identity (%)	Similarity (%)	Charge	pI
Ginsentides				
TP1	31.2	46.9	0	6.68
TP2	31.2	50.0	0	6.68
TP3	34.4	50.0	0	6.68
TP4	34.4	43.8	-1	4.37
TP5	31.4	51.4	-1	5.21
TP6	38.1	42.9	-1	4.00
TP7	31.2	46.9	0	6.68
TP8	31.2	50.0	-1	5.07
TP9	31.2	46.9	0	6.68
TP10	31.2	46.9	-1	5.07
TP11	37.5	40.6	+1	7.70
TP12	38.1	42.9	-1	5.24
TP13	38.1	42.9	-1	4.00
TP14	31.8	40.9	+4	8.74
8C-Chitin-binding HLPs				
aV1	47.6	47.6	0	5.82
aV2	52.4	52.4	-1	4.21
aV3	34.3	40.0	+2	8.18
aV4	47.6	47.6	+1	7.72
aV5	47.6	47.6	+2	8.21
aV6	47.6	47.6	+1	7.72
aV7	47.6	47.6	0	5.82
aV8	34.3	40.0	+1	7.68
aV9	30.3	36.4	0	5.82
Avesin A	32.4	38.2	-1	4.37
vH1	34.3	48.6	+2	8.22
vH2	34.3	48.6	+2	8.22
mO1	25.8	41.9	+3	8.53
mO2	29.0	45.2	+3	8.53
gB1	27.0	32.4	+3	8.50
gB2	27.0	32.4	+3	8.50
Hevein	22.2	38.9	-2	4.83
Fa-AMP1	30.3	39.4	+2	8.22
Fa-AMP2	30.3	39.4	+2	8.23
Pn-AMP1	29.7	35.1	+4	8.76
Pn-AMP2	29.7	35.1	+4	8.76

3.2 Putative biosynthesis of eleutide

Eleutide eT1 precursor contains a three-domain arrangement that includes an ER signal peptide, a pro-peptide and a mature domain (Figure 6.4). The presence of an ER signal peptide indicates that the bioprocessing of eT1 is via the secretory pathway, which is reported in many plant CRPs [53, 56, 57, 59, 63, 64, 66, 394, 395]. Peptides that contain the ER signal peptides are translocated from the cytoplasm to the ER membrane, where the signal peptide is cleaved by signal peptidase (SPaseI). Subsequently, the pro-peptide is cleaved by an endopeptidase, and the mature peptide is released for post-translational modification. The cleavage site between the pro-peptide and the mature domain is likely between the glycine at the C-terminal of the pro-peptide and glutamine at the N-terminal of the mature domain. The N-terminal glutamine in the mature domain is then converted to pyroglutamate through a spontaneous non-enzymatic reaction. Importantly, precursor architecture comparison of eT1 with the other 8C-CRPs revealed that eT1 adopts the same precursor architecture as ginsentides and is evidently different from the 8C-chitin-binding HLPs, 8C-defensins and 8C-thionins (Figure 5.5). This indicated that eT1 belongs to the 8C-non-chitin-binding HLP subfamily that is similar to ginsentide. Interestingly, the pro-peptides of eT1 and ginsentides appeared to be longer than eleutide-like homologs, suggesting that they may contain motifs that contribute to additional function. However, motif search using the pro-peptide of eT1 as a query on the online sever ScanProsite did not match any motifs from the database. Collectively, we have provided insights into the

bioprocessing of eleutide eT1 and showed that eT1 belongs to the 8C-non-chitin-binding HLP subfamily.

3.3 Distribution and occurrence of 8C-non-chitin-binding HLPs in *planta*

A tBLASTn search using eT1 and query was performed to examine the distribution and occurrence of 8C-non-chitin-binding HLPs in *planta*. Manual filtering was applied to remove sequences that have an ER signal peptide of less than ten amino acid residues, odd number cysteine residues and identical sequence from the same plant species. A total of 122 sequences homologs were identified, of which 14 of them were ginsentides. The identified sequences belonged to 16 different plant families, namely, Araliaceae, Blasiaceae, Boraginaceae, Brassicaceae, Dilleniaceae, Fabaceae, Griseliniaceae, Lepidoziaceae, Malvaceae, Marattiaceae, Microteaceae, Myodocarpaceae, Petiveriaceae, Poaceae, Rubiaceae and Salicaceae. A circular phylogenetic tree was constructed to illustrate the distribution (Figure 6.14). The full primary sequences were calculated with the neighbor-joining clustering algorithm with 1000 bootstrap. Based on the phylogeny tree, the sequence homologs were detected only in *planta*. Interestingly, the mature domains of ginsentides TP1–TP3 and TP5–TP6 were expressed in different ginseng species. For example, TP1 was expressed in *P. ginseng* and *P. notoginseng* whereas TP2 was expressed in *P. ginseng* and *P. quinquefolius*. Such expression of the same peptide in different plant species aligns with natural product occurrences in which the same product can be found in different species of the plant family [358]. The data mining revealed three major clusters, namely, eleutide eT1, ginsentides and eleutide-like homologs (Figure 6.14A). Notably most of the

sequences identified belonged to Araliaceae (16.3%), Malvaceae (20.3%), Poaceae (23.6%) and Rubiaceae family (22%), accounting for approximately 82% of the total sequences identified (Figure 6.14B). Eleutide eT1 appeared more closely related to the eleutide-like homologs from the Rubiaceae family, although it is a ginseng-like peptide. This is observed in Table 6.1 in which cA1 and cA2 from *C. arabica*, belonging to the Rubiaceae family had a higher sequence similarity (50–50.2%) compared to ginsentide TP1 and TP2 which had sequence similarity of 46.9% and 50%, respectively. Furthermore, the intercysteine loop 5 of eT1, and coffeetide, belonging to the Rubiaceae family, is much longer compared to ginsentide, which was confirmed by the structural alignment (Figure 6.7). As such, it is postulated that the biological activity of eT1 would be much similar to peptides isolated from the Rubiaceae family compared to ginsentides. Overall, the data mining results revealed that eT1 and eleutide-like homologs are found in only angiosperms and are distributed in 16 plant families. Taken together, our data expanded the existing library of 8C-non-chitin-binding HLP subfamily.

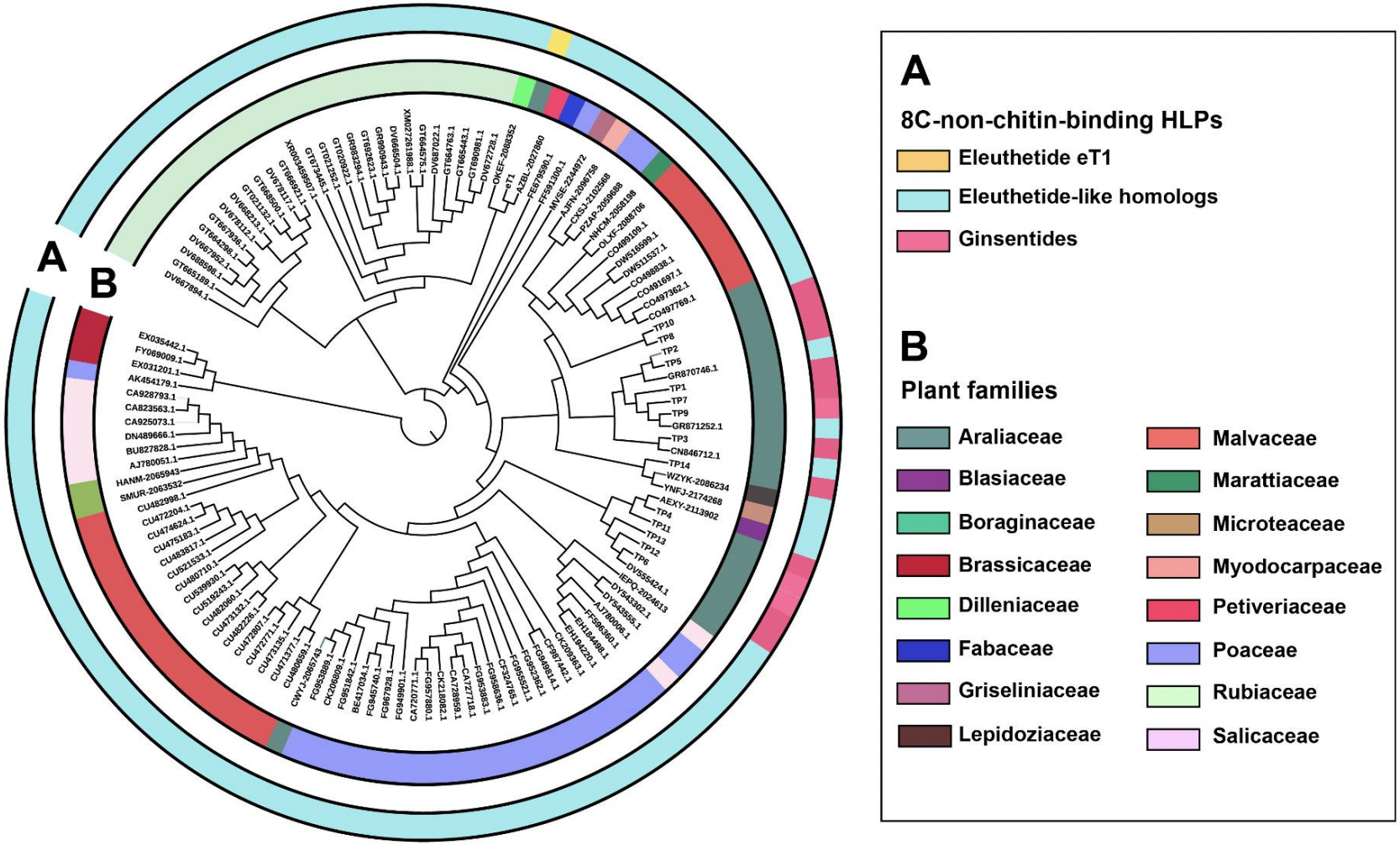


Figure 6.14: Distribution and occurrence of 8C-non-chitin-binding hevein-like peptides. Classification based on **(A)** 8C-non-chitin-binding HLPs and **(B)** plant families.

3.4 Bioactivity of eleutide

The cytoprotective and wound-healing effects of eleutide eT1 agree with the ethnomedicinal uses of *E. trifoliatus* for the treatment of inflammation and wound healing [313, 314, 399]. As part of the stress response to modulate homeostasis to increase cellular viability, autophagy facilitates the removal of cellular damages caused by environmental stress like hypoxia-induced stress. Autophagy is a cellular mechanism that degrades intracellular components and maintains cellular homeostasis. Specifically, the LC3-interacting region (LIR) motif, facilitates the targeting of autophagy receptors to LC3 or AT8 family proteins docked to the phagophore membrane, triggering an array of biochemical pathways [400]. The LIR motif contained a conserved sequence of F/W/Y-X-X-L/I/V, a sequence motif with two amino acids separated by an aromatic and hydrophobic residue [400, 401]. This motif contains typically two acidic amino acid residues to form a short recognition sequence of six amino acid residues attached into the AT8 binding pocket. A motif search using eT1 as a query was performed, and the sequence SKFIRP located in loop 5 of eT1 matched with the LIR motif with a probability score $3.37e^{-3}$, suggesting that eT1 may play a role in autophagy. Our findings showed that eleutide eT1 displayed mild cytoprotective effects that may alleviate hypoxia-induced pathogenesis and induced cell proliferation in H9c2 cardiomyocytes. Similarly, this result agrees with β-ginkgotide isolated from *G. biloba,* which contained an LIR-like motif (DEYGCI) and induced cytoprotective effects in both cardiomyocytes and neuronal cells [398]. Additionally, eT1 promotes cell migration, which is a hallmark of wound healing. Taken together, these findings showed that

eT1 could potentially reduce hypoxia-induced pathogenesis, induce cell proliferation and promote cell migration.

This study has successfully isolated and characterized an 8C-non-chitin-binding HLP, eleutide eT1, from *E. trifoliatus*. Sequence analysis revealed that eT1 contained 35 amino acid residues with eight cysteine and a conserved tandem cysteine (-CC-) motif. Biosynthesis analysis revealed that eT1 precursor sequence adopt a three-domain architecture similar to gisentides. Phylogenetic analysis showed that eT1 belongs to the 8C-non-chitin-binding HLP subfamily. Structurally, eT1 possesses a disulfide connectivity of CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII, which forms a cystine-dense pseudocyclic structure. Data mining revealed a total of 123 eleutide-like homologs that are distributed in 16 plant families. Biologically, eT1 displayed high metabolic stability against thermal, acid and proteolytic degradation. In addition, eT1 exhibits mild cytoprotective effect, increase cell proliferation and promote cell migration. Collectively, our data reports eleutide as an 8C-non-chitin-binding HLP, which expands the library of 8C-non-chitin-binding HLPs and provides insights into the sequence, structure and functional diversity in 8C-non-chitin-binding HLPs. As such, eleutide could be a potential scaffold for synthesizing metabolically stable orally active biotherapeutic.

CHAPTER SEVEN

In silico Identification of Hevein-like Peptides in *Planta*

1. Introduction

Bioactive peptides are found in many organisms from prokaryotes such as bacteria to eukaryotes like plants and mammals. These natural diversities of peptides often exhibit a myriad of functional and structural diversities that represent as interesting candidates to study [402]. The traditional method of identifying putative bioactive peptides is through a series of fractionation in which the crude plant extract undergoes fractionation by HPLC and the fractions are subjected to bioactivity screenings [403]. Even though this approach has successfully discovered novel peptide-derived biologics for many decades, it has its own set of limitations for being labor-intensive and time-consuming. Therefore, this imposes the demand to develop a high-throughput screening technique to overcome the limitations of the conventional approach and identify novel bioactive peptides in plants efficiently. Furthermore, the progressive advances in the field of omics have contributed to a magnitude of data and provided a new direction for discovering new bioactive peptides. At present, there are over 74 million expressed sequence tags (ESTs) entries in the public databases [404]. The use of bioinformatics coupled with sophisticated computational algorithms can facilitater screening for a myriad of plant genomes in a high-throughput manner compared to the traditional method [405]. This method has been used to identify miRNAs [406], transcriptional

regulators [407] and bacterial ADP-ribosyltransferases [408]. Despite its wide usage, it is not vastly used to identify putative CRPs.

HLPs represent an emerging group of CRP family that exhibit an arsenal of biological activities, which render them suitable prospects to study their distribution and occurrence in plants. The natural sequence diversity in HLPs allows them to be broadly classified into chitin-binding HLPs and non-chitin-binding HLPs. Further classification of the HLPs is based on the number of cysteines present, namely, 6C-, 8C- and 10C-HLPs. At present, 35 6C-HLPs , 35 8C-HLPs and five 10C-HLPs have been reported from Amaranthaceae (10 peptides), Araliaceae (14 peptides), Cactaceae (15 peptides), Caryophyllaceae (4 peptides), Celastraceae (1 peptide), Convolvulaceae (2 peptides), Eucommiaceae (2 peptides), Euphorbiaceae (1 peptide), Ginkgoceae (11 peptides), Malvaceae (8 peptides), Moringaceae (2 peptides), Poaceae (3 peptides) and Polygonaceae (2 peptides) plant families (Figure 7.1). Despite the wide distribution of HLPs in plants, their evolution, distribution and occurrence remain poorly understood. Therefore, this study aims to understand their evolution, distribution and occurrence in different plant families using the a bioinformatic approach.

The search was performed on NCBI, oneKP and PhytAMP databases using reported HLPs as queries to identify putative HLPs. Manual filteration was done to consolidate the hits and a total of 1204 novel HLPs were successfully identified from 252 plant families.

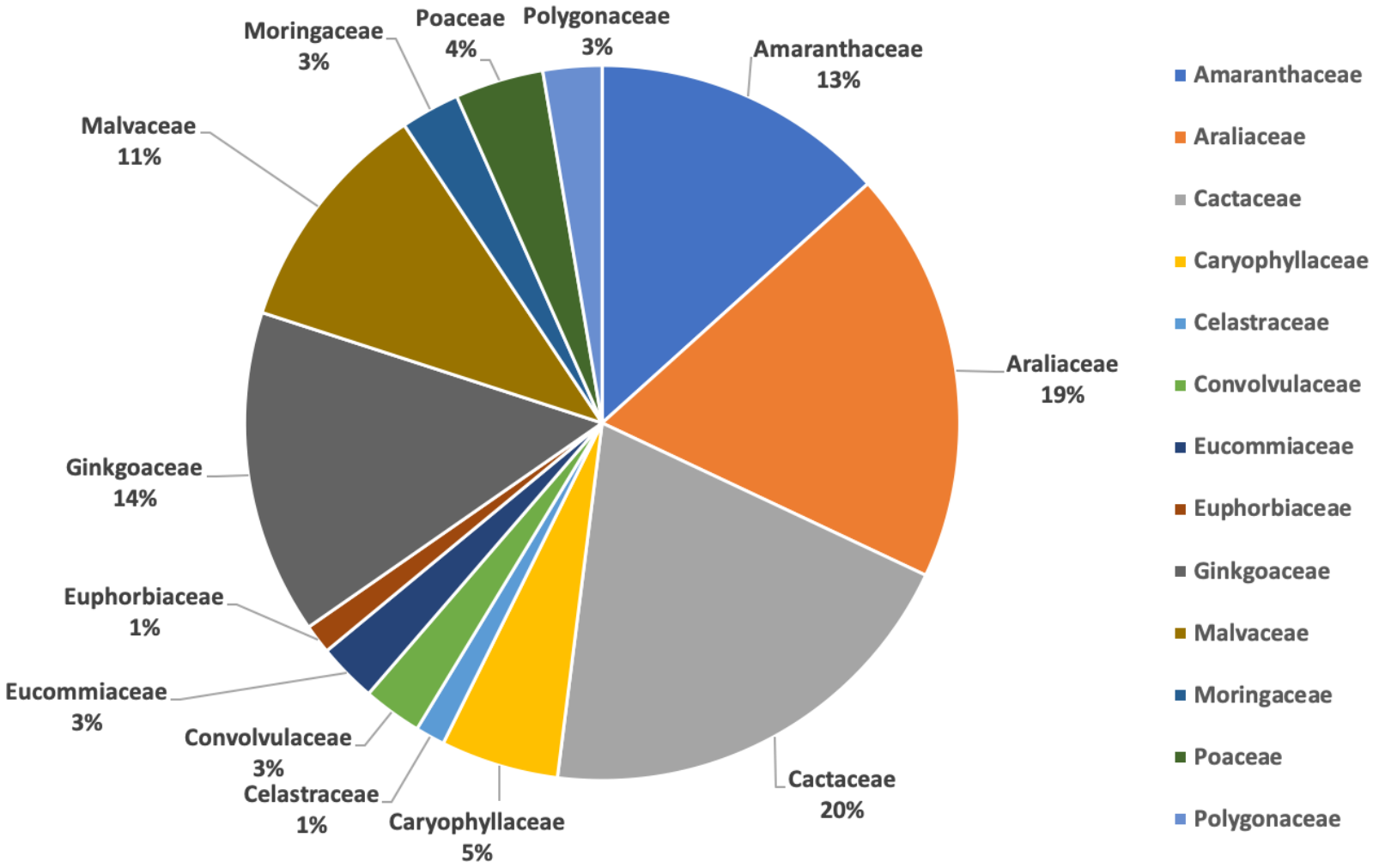


Figure 7.1: Distribution of reported hevein-like peptides plant families. A total of 75 hevein-like peptides have been reported to date and are distributed in 13 different plant families. The Amaranthaceae, Araliaceae, Cactaceae and Ginkgoceae families represents the major families and account for more than 60% of the reported peptides.

The cladograms were constructed to elucidate the relations of HLPs among plant species, and their precursor organizations were examined. The sequence alignment of HLPs revealed seven types of precursor organizations. The chitin-binding and non-chitin-binding HLPs consist of four and three types of precursor architecture, respectively. Apart from identifying HLPs from angiosperms (flowering plants), gymnosperms (non-flowering plants) and bryophyte (non-vascular plants), our data also identified HLPs in lycopodiophyta (club mosses) and euphyllophyta (ferns). At present, the 10C-HLP subfamily reports three types of disulfide connectivity. According to our search, a novel 10C-HLP disulfide connectivity was identified and termed as type IV. Taken together, our findings provided additional insights on the sequence diversity, biosynthetic diversity, distribution and occurrence of HLPs in *planta*.

2. Results

2.1 Expressed sequence tag-based discovery of hevein-like peptides in *planta*

The EST-based searches were performed on three primary databases, namely, NCBI (https://www.ncbi.nlm.nih.gov/), PhytAMP (http://phytamp.hammamilab.org/) and oneKP (https://db.cngb.org/onekp/) to identify novel putative HLP sequences in plants. The reported HLPs were used as query sequence against the EST databases, which consist of nucleotide sequences translated into open reading frames. In this study, the criteria for selection of HLPs sequences include: (1) Sequence coverage of more than 80%, (2) sequence identity of more than 60% and (3) comprise both start and stop codons. The initial search generated a total of 2475 HLPs from the three databases, of which 342 were 6C-chitin-binding HLPs, 225 were 6C-non-chitin-binding HLPs, 1047 were 8C-chitin-binding HLPs, 653 were 8C-non-chitin-binding HLPs, and 208 were 10C-chitin-binding HLPs. Due to the overlap of sequences from the databases, the replicated sequences from the same plant species were manually removed. A total of 1271 sequences were removed, of which 287 were 6C-chitin-binding HLPs, 153 were 6C-non-chitin-binding HLPs, 160 8C-chitin-binding HLPs, 531 8C-non-chitin-bionding HLPs and 140 10C-chitin-binding HLPs. After the removal, a total of 1204 putative HLP sequences remained (Figure 7.2). Collectively, the prevalence of HLPs in plants indicated that they are a diverse distributed family of CRP that are underexplored. Therefore, the study of their distribution, occurrences and the evolutionary relationship may further expand our understanding of HLPs.

Criteria for selection of sequences:

1. **Sequence coverage ≥ 80%**
2. **Sequence identity ≥ 60%**
3. **Presence of signal peptide, mature domain and C-tail**
4. **Comprise start and stop codons**

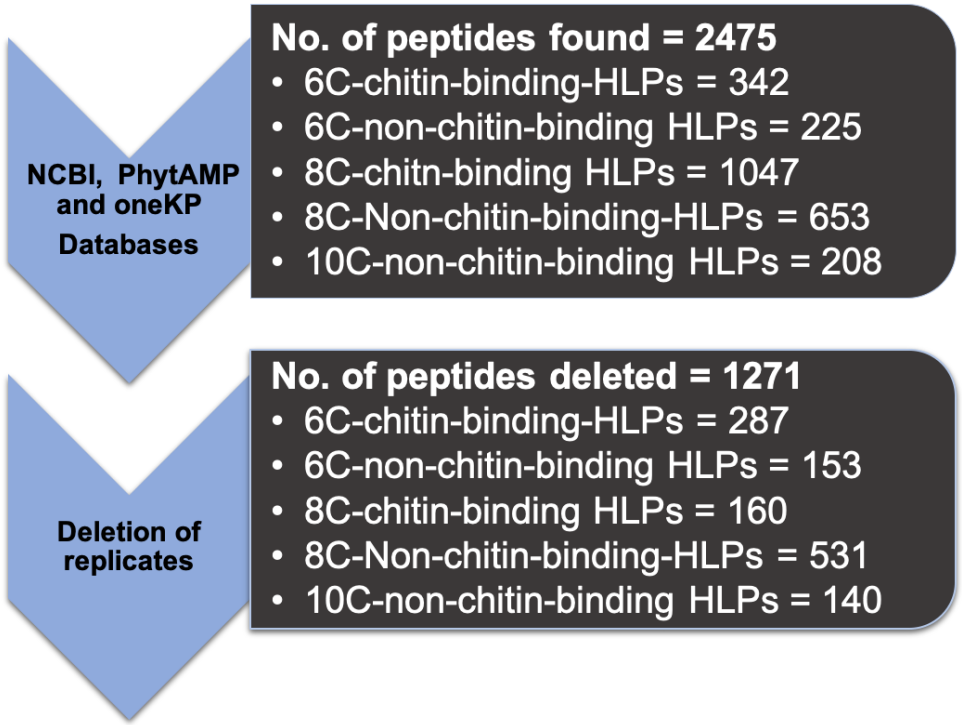


Figure 7.2: Expressed sequence tag-based search of putative hevein-like peptides. Data mining of hevein-like peptides using the above criteria from NCBI, PhytAMP and oneKP revealed a total of 2475 HLP sequences. The deletion of replicates from the same plant species resulted in the total of 1204 putative HLPs identified.

2.2 Distribution and occurrence of hevein-like peptides in *planta*

The EST data mining identified a total of 252 plant families belonging to 87 plant orders expressing HLPs. In the 6C-chitin-binding HLP subfamily, 55 novel sequences were distributed across 34 plant species from 14 plant families (Figure 7.3A). The Amaranthaceae family was the major plant family, accounting for 67% of the plant families identified. This suggests that the Amaranthaceae plant family primarily contribute to the 6C-chitin-binding HLPs in plants. Additionally, within the 6C-non-chitin-binding HLP subfamily, 72 novel sequences were identified and distributed in 42 plant species from 15 plant families (Figure 7.3B). The Amaranthaceae, Apocynaceae, Cactaceae and Caryophyllaceae represent the major plant families, accounting for 68% of the plant families identified.

The 8C-chitin-binding HLPs represent the highest amount of HLPs identified, with 887 novel sequences distributed across 610 plant species from 241 plant families (Figure 7.4A). The Fabaceae and Cupressaceae were the major plant families identified, accounting for 12% and 6%, respectively. In the 8C-non-chitin-binding HLP subfamily, 122 novel sequences were identified and distributed in 34 plant species from 13 plant families (Figure 7.4B). The Araliaceae, Malvaceae, Poaceae and Rubiaceae represent the major plant families, accounting for 86% of the total plant families identified.

Lastly, within the 10C-chitin-binding HLP subfamily, 68 novel sequences were identified and distributed in 48 plant species from 15 plant families (Figure 7.5). Such vast distribution of HLPs in plants suggests that they are essential and likely constitutively expressed in plants.

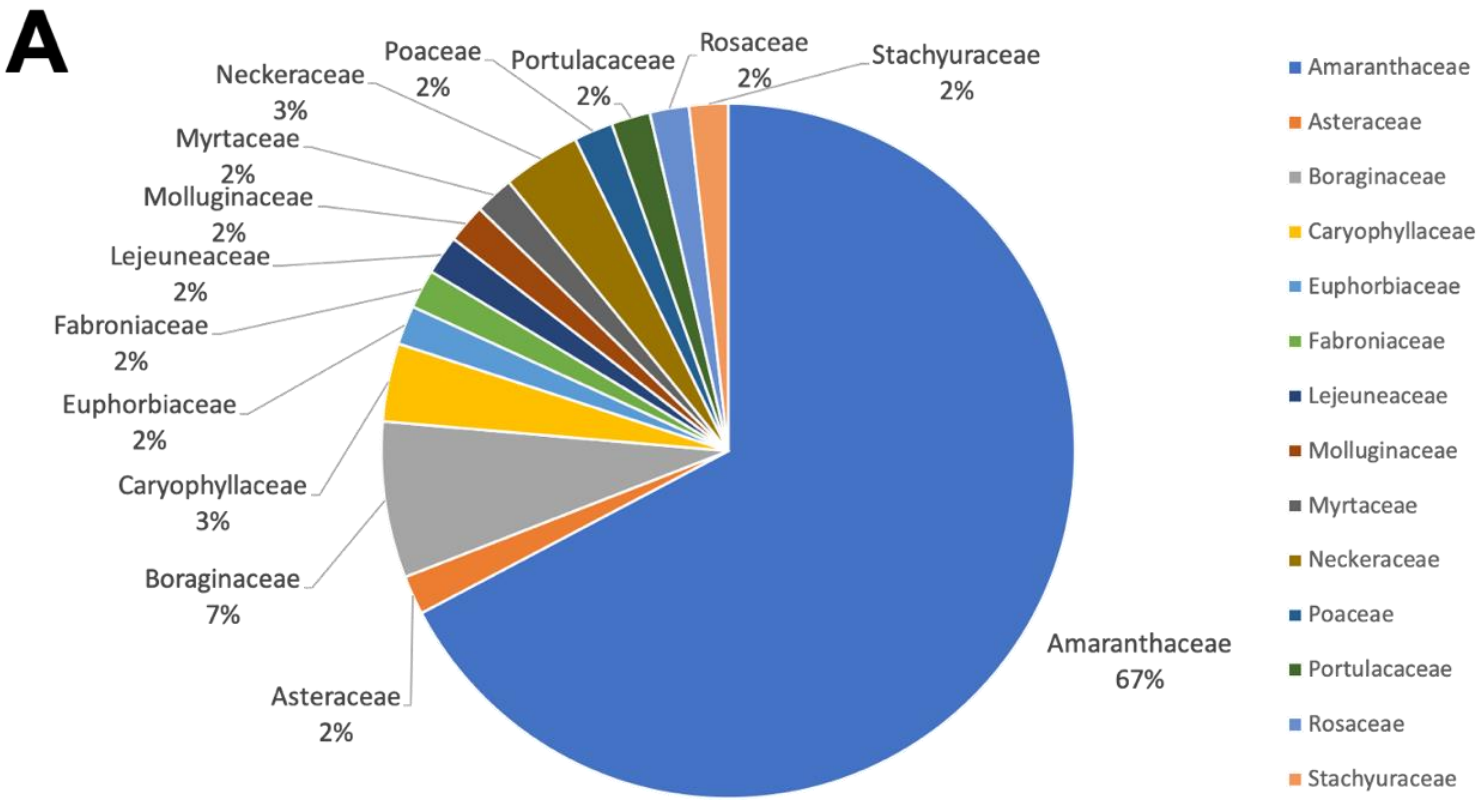


Figure 7.3A: Distribution of 6C-chitin-binding hevein-like peptide plant families. In the 6C-chitin-binding HLP subfamily, 14 plant families were identified.

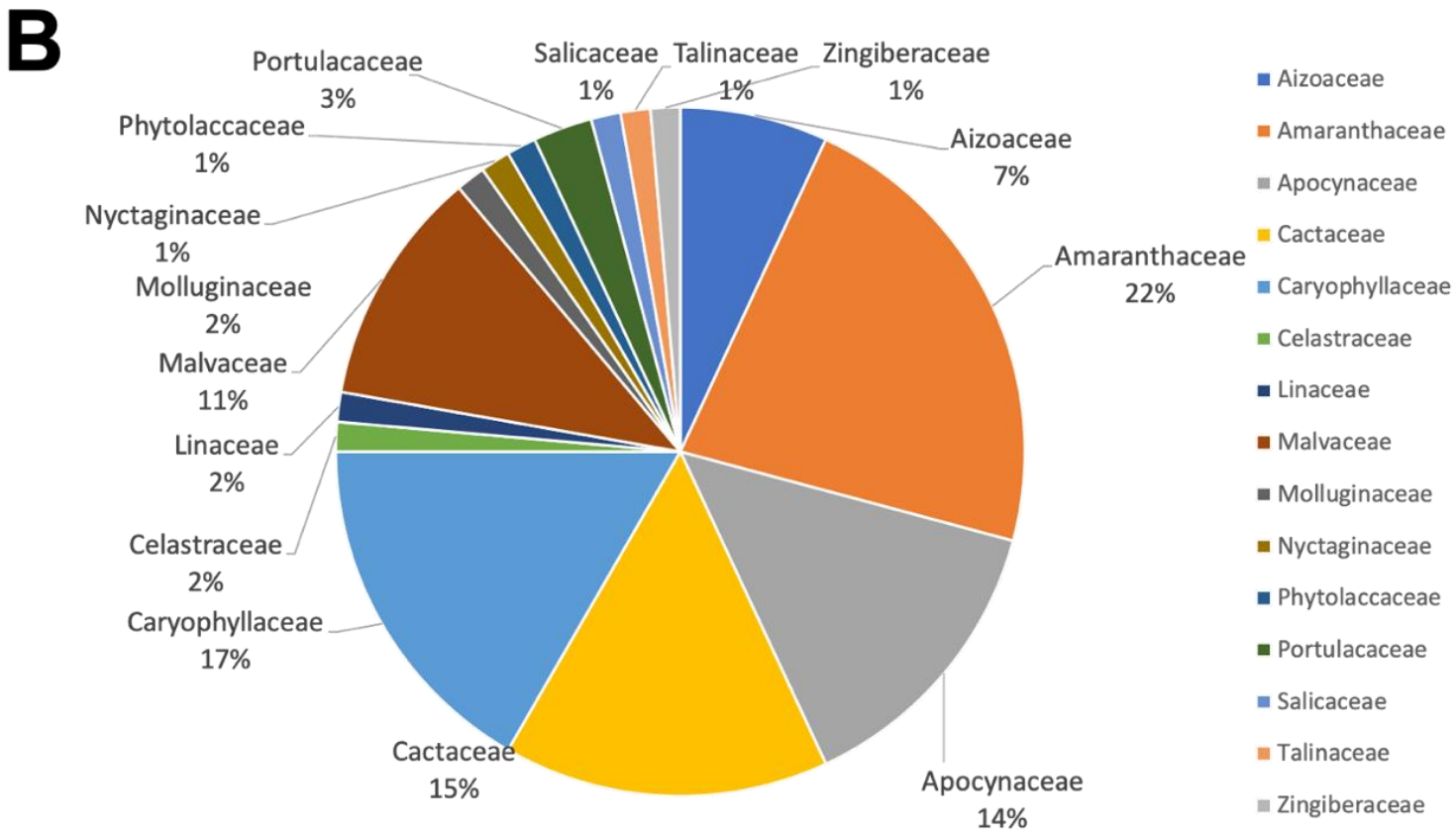


Figure 7.3B: Distribution of 6C-non-chitin-binding hevein-like peptide plant families. In the 6C non-chitin-binding HLP subfamily, 15 plant families were identified.

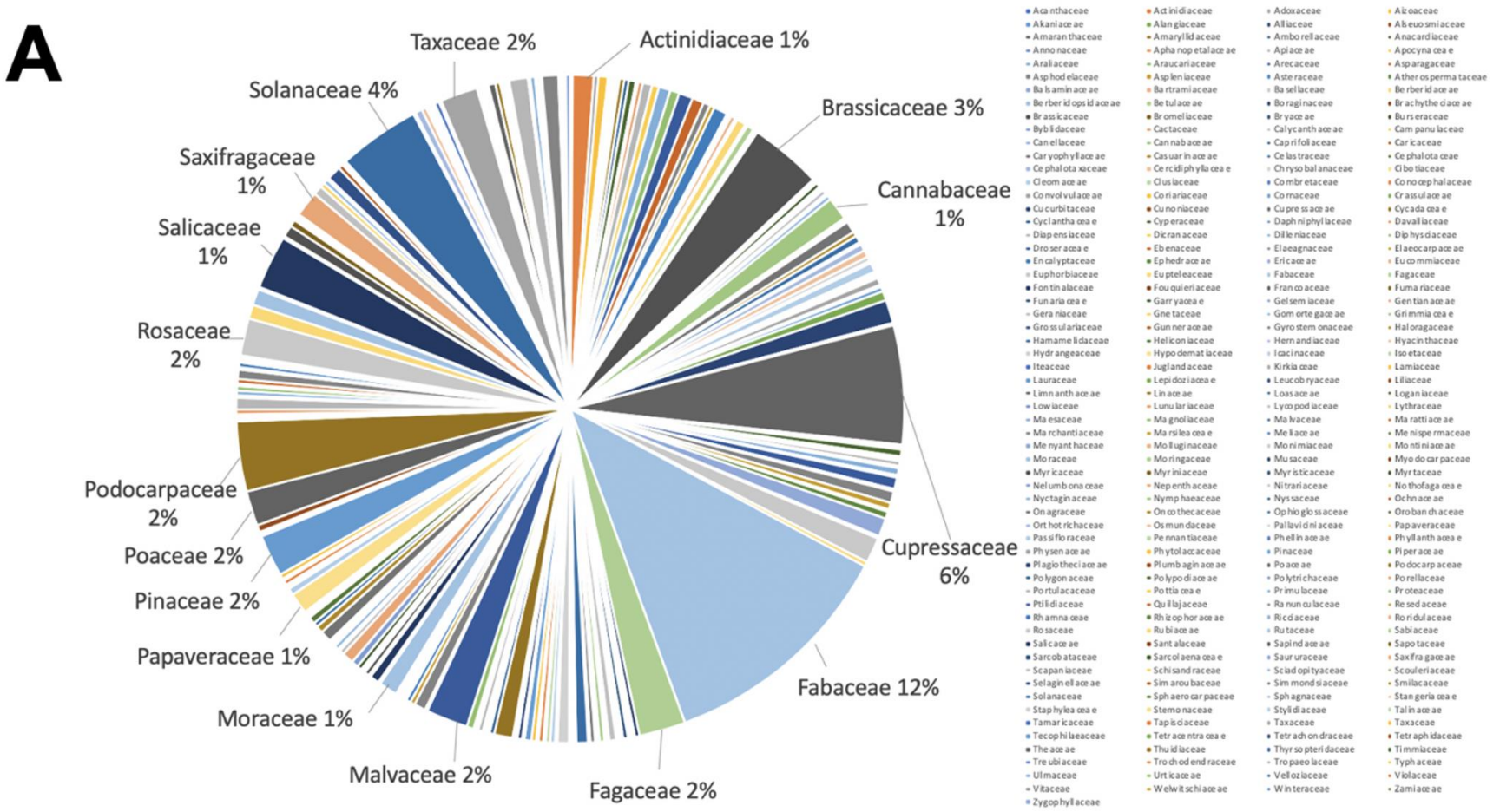


Figure 7.4A: Distribution of 8C chitin-binding-hevein-like peptide plant families. In the 8C-chitin-binding HLP subfamily, 241 plant families were identified.

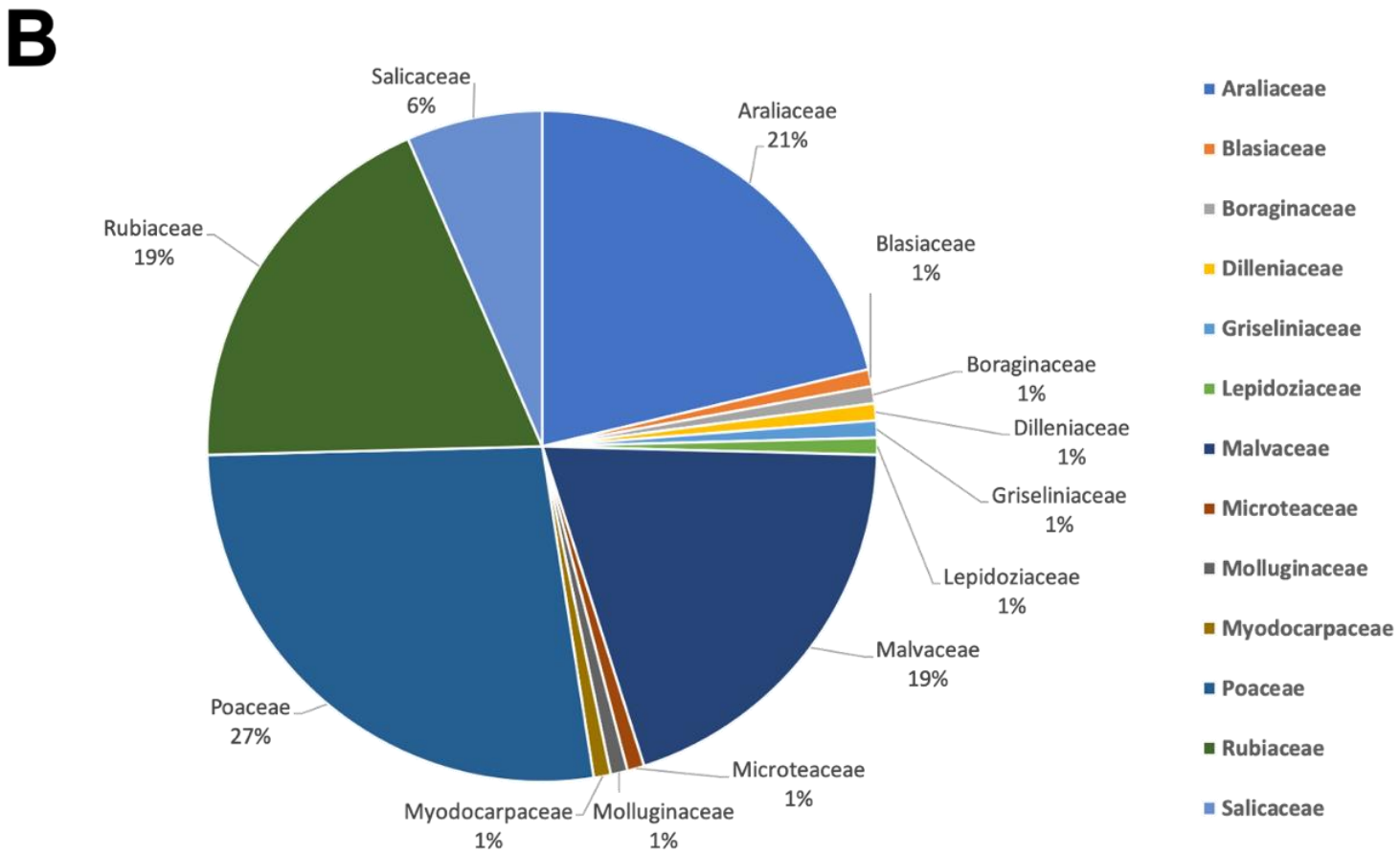


Figure 7.4B: Distribution of 8C non-chitin-binding-hevein-like peptide plant families. In the 8C-non-chitin-binding HLP subfamily, 13 plant families were identified.

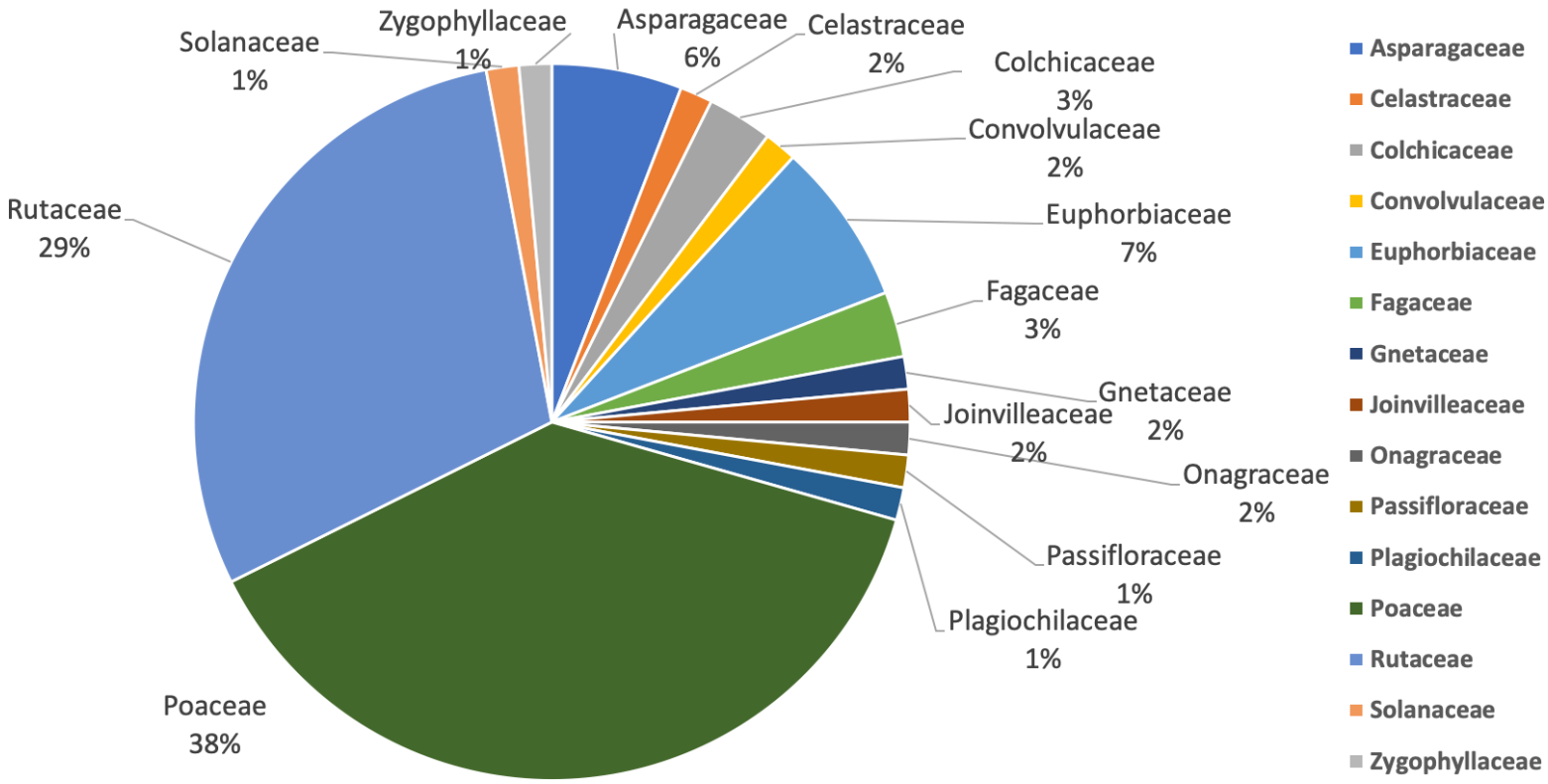


Figure 7.5: Distribution of 10C-hevein-like peptide plant families. In the 10C-chitin-binding HLP subfamily, 15 plant families were identified.

Within the 6C-HLP subfamily, a total of 127 novel HLPs were identified. These peptides comprise of mainly angiosperms (97%) and bryophyta (3%) (Figure 7.6A). In contrast, a total of 1009 novel 8C-HLPs were identified. These peptides are mainly distributed in angiosperms (80%), gymnosperms (13%), bryophyta (5%), lycopodiophyta (1%) and euphyllophyta (1%) (Figure 7.6B). The 10C-HLPs are mainly distributed in angiosperms (97%), gymnosperms (1.5%) and bryophyta (1.5%) (Figure 7.6C). Interestingly, 85 HLPs were identifed in more than one plant species. These HLPs were found in plants of the same genus but different species or different genus and species, belonging to the same plant family (Appendix A).

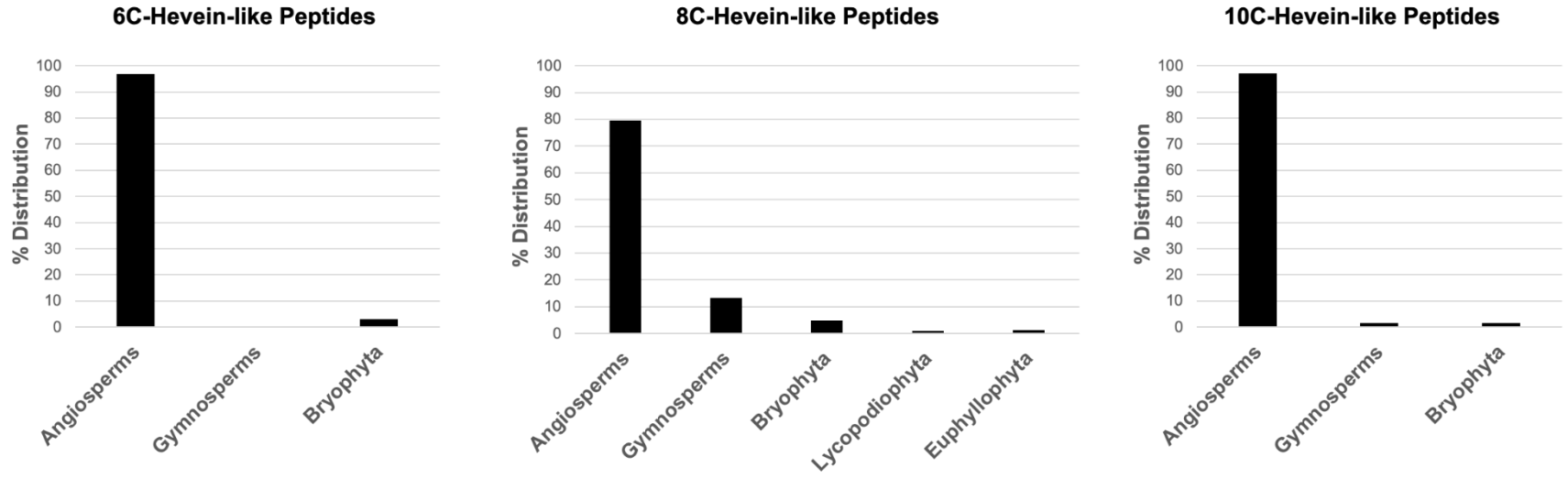


Figure 7.6: Distribution of putative hevein-like peptides in *planta*. (A) 6C-hevein-like peptides, **(B)** 8C-hevein-like peptides and **(C)** 10C-hevein-like peptides.

Cladograms were constructed to analyze the evolution and distribution of the putative HLPs. The expression of putative 6C-HLPs were distributed mainly in angiosperms and bryophyta. Four 6C-HLPs accounting for 3% of the total 6C-HLP, originated from non-vascular plants termed as bryophyta and were distributed in the order of Hypnales and Porellales. The remaining 6C-HLPs identified in angiosperms include monocots and eudicots. Plant families derived from monocots include Poaceae and Zingiberaceae, whereas plant families belonging to eudicots were derived from orders such as Asterales, Boraginales, Gentianales, which are categorized under the common clade of Asterid and Rosids (Figure 7.7). Therefore, this suggests that plants belonging to these orders may contain putative 6C-HLPs, and by examining these plant families, the library of 6C-HLP can be further expanded.

In contrast, the putative 8C-HLPs were distributed in bryophyta and tracheophyte, which include: lycopodiophyta, euphyllophyta, gymnosperms and angiosperms (Figure 7.8). Among bryophytes, 8C-HLPs were found in main clades such as Bryidae, Jungermaniidae and Marchantiidae. This suggests that the peptides originated from the early evolution of bryophytes. Within lycopodiophyta, plant families were distributed in the order of Isoetales, Lycopodiales and Selaginellales, whereas in euphyllophyta, plant families were distributed in the order of Polypodiales, Ophioglossales and Marattiales. In gymnosperms, plant families were distributed in the order of Cycadales, Ephedrales, Gnetales, Araucariales, Cupressales, Cyatheales, Osmundales, Polypodiales and Salviniales. The majority of the 8C-HLPs from gymnosperm were observed to have the highest abundance

in the order of Genetales from the sub-class Gnetidae and also in less common order such as Osmunales and Cyatheales, indicating the pervasive occurrence of these peptides among gymnosperms. Similar to 6C-HLPs, the 8C-HLPs were prevalent in angiosperm (80%). The occurrence of these peptides in monocots are in orders of Asparagales, Arecales, Liliales, Poales and Zingiberales, whereas 8C-HLPs originating from dicots were mainly from the common clade of Asterids and Rosids. The extensive distribution of 8C-HLP revealed that they are highly diverse and were likely the earliest member of the HLP family expressed in plants.

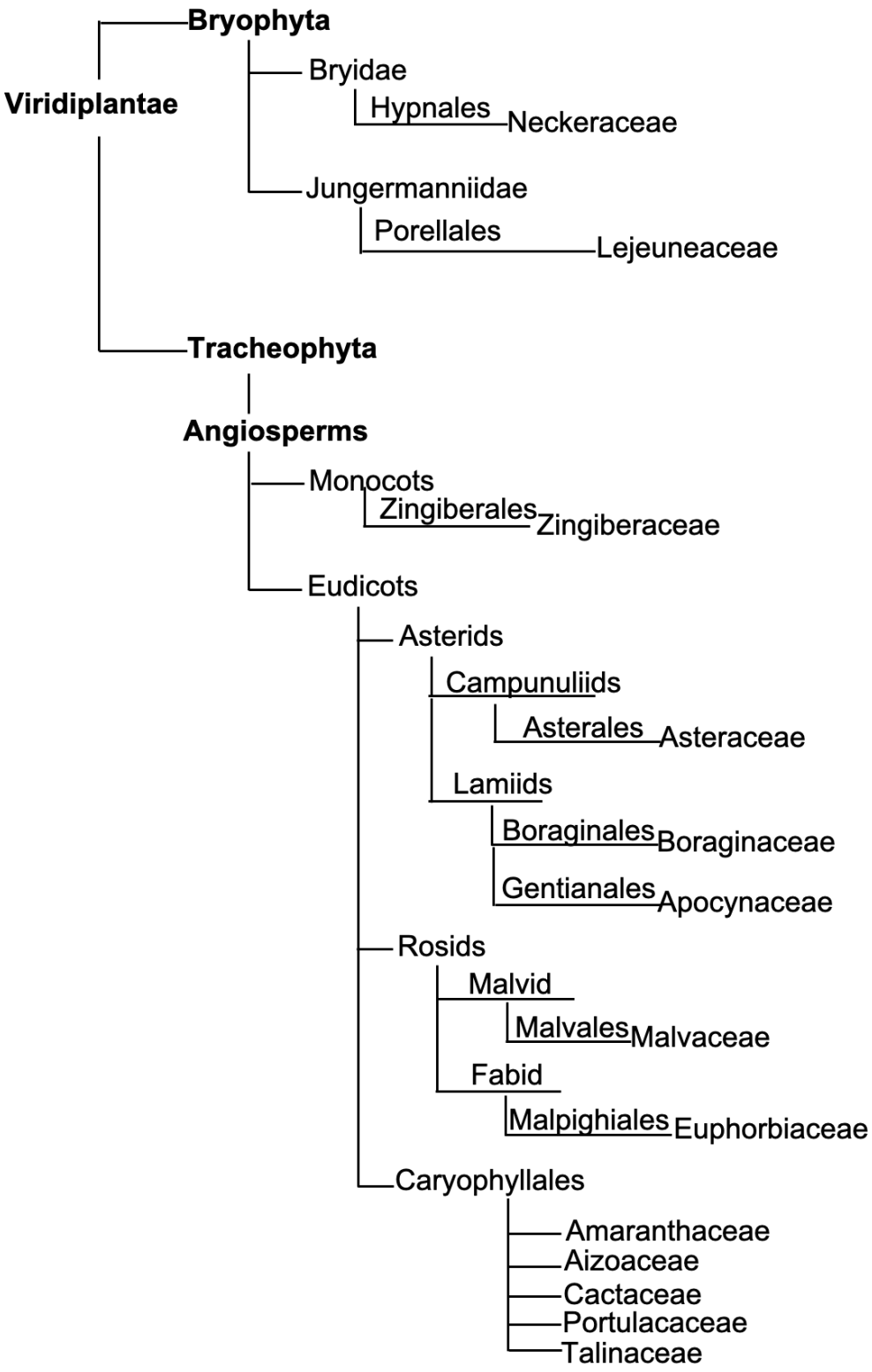


Figure 7.7: Cladogram of 6C-hevein-like peptide distribution. A total of 127 6C-HLPs are found distributed in bryophyta and angiosperms.

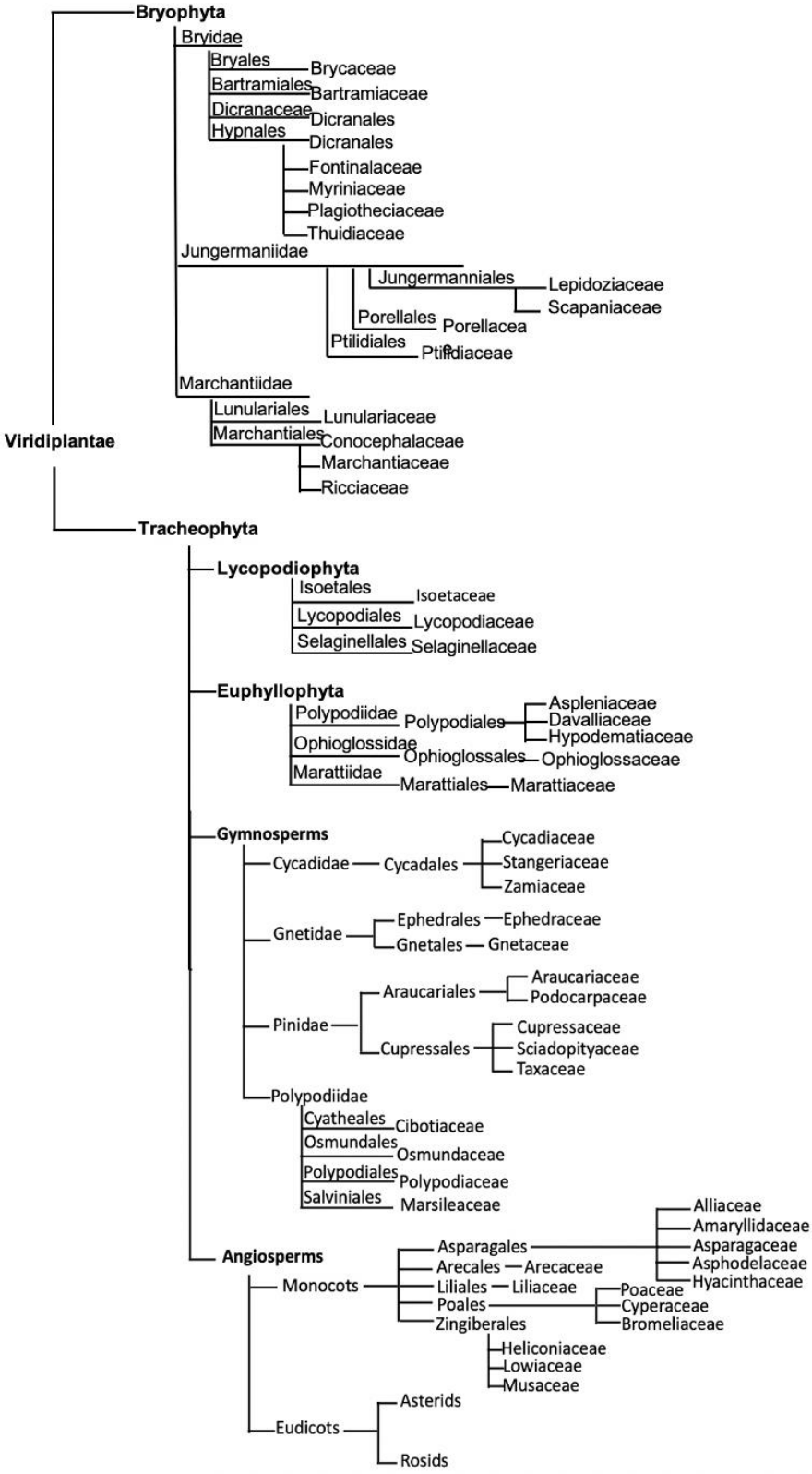


Figure 7.8: Cladogram of 8C-hevein-like peptide distribution. A total of 1009 8C-HLPs are found distributed in bryophyta, lycopodiophyta, euphyllophyta and angiosperms.

Putative 10C-HLPs were identified in bryophytes, gymnosperms and angiosperms (Figure 7.9). The 10C-HLPs found in bryophytes and gymnosperms were found in Plagiochilaceae and Gnetaceae plant families, respectively. Similar to both 6C- and 8C-HLPs, the 10C-HLPs were mostly found in angiosperms. The Poaceae family (38%) represented the major monocot plant family, whereas the Rutaceae family (29%) represented the major eudicot family in 10C-HLPs. Hence, it can be postulated that monocot and eudicot angiosperms of 10C-HLPs forms a close evolutionary relationship.

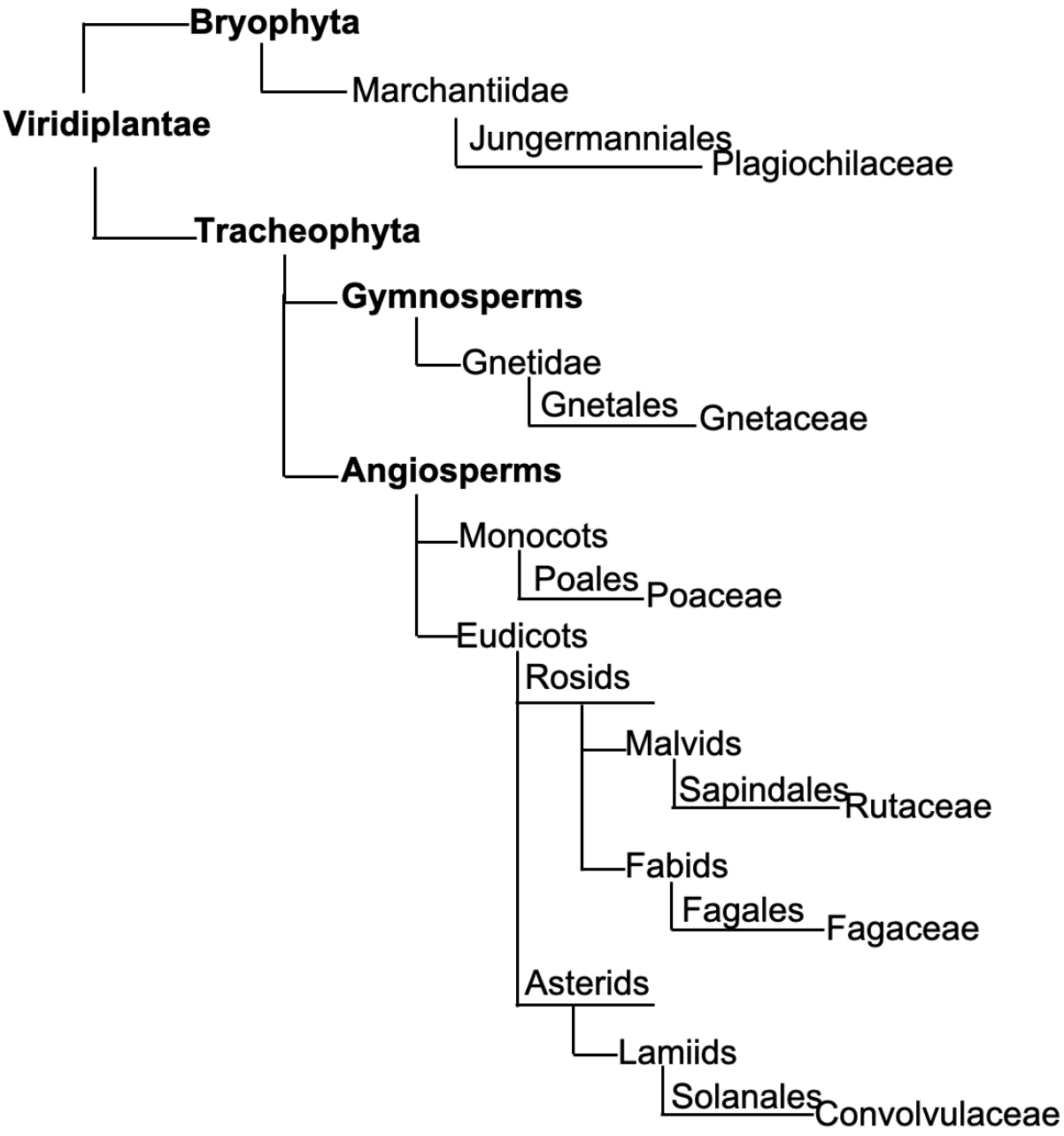


Figure 7.9: Cladogram of 10C-hevein-like peptide distribution. A total of 68 8C-HLPs are found distributed in bryophyta, lycopodiophyta, euphyllophyta and angiosperms.

Taken together, the substantial expression of HLPs in the *planta* strongly indicate that they are one of the most extensive family of CRPs and are essential to plants because of their anti-microbial function, which facilitate plant defense.

2.3 Sequence comparison of hevein-like peptides

The total amino acid content of the HLPs was determined to understand the sequence properties of the putative HLPs. Further sequence analysis was performed using ClustalW to align HLP sequences obtained from the databases and previously reported HLPs. Sequence logos were generated to show the sequence alignment and sequence conservation. Sequence comparison revealed that the cysteine spacing of 6C- and 8C-HLPs were conserved. Interestingly, within the 10C-HLP subfamily, a new cysteine spacing pattern that differed from previously reported Type I, Type II and Type III cysteine spacing was identified. Hence, the new cysteine pattern was termed as Type IV.

2.3.1 6C-chitin-binding hevein-like peptides

Sequence comparisons were performed on 12 previously reported 6C-chitin-binding HLPs with the 55 putative 6C-chitin-binding HLPs acquired from the EST databases. Figure 7.10A shows the total amino acid content of 6C-chitin-binding HLPs, which are rich in cysteine (20%), glycine (24%), proline (6%), serine (7%) and tyrosine (10%) and have an average overall net charge of +1. To further illustrate the sequence alignment of 6C-chitin-binding HLPs, a sequence logo was constructed (Figure 7.10B). The height of the sequence logo reveals the sequence conservation, whereas the symbol height within the stack indicates the prevalence of each amino acid in that position. The 6C-chitin-binding HLPs contain 25–39

amino acids with a conserved cysteine spacing. Apart from the conserve cysteine spacing, Gly20, Gly23, Gly25, Gly30, Ser16, and Tyr 29 were highly conserved (numbered according to accession number MUCT-2027217) (Appendix B). The 6C-chitin-binding HLPs can be divided into four intercysteine loops, which contained four to eight amino acid residues. Intercysteine loop 1 has the largest loop size comprising of eight amino acids and is highly variable. In contrast, loop 2 has the smallest loop size consisting of four amino acids. The conserved chitin-binding domain is located in intercysteine loop 3 and 4. Within the chitin-binding domain, the first aromatic residue favors either a phenylalanine or tyrosine, whereas the second aromatic residue favors any aromatic amino acid. The third aromatic residue highly favors tyrosine as it appeared to be absolutely conserved in all 6C-chitin-binding HLPs. The disulfide connectivity of 6C-chitin-binding HLPs has previously been established as CysI–CysIV, CysII–CysV and CysIII–CysVI, which forms a cystine-knot motif that provides a stable compact structure. Based on the highly conserved cysteine spacing among all the 6C-chitin-binding HLPs, the putative 6C-chitin-binding HLPs were assumed to also contain a cystine-knot motif similar to aSG1 which the disulfide connectivity was deduced by solution-state NMR [51].

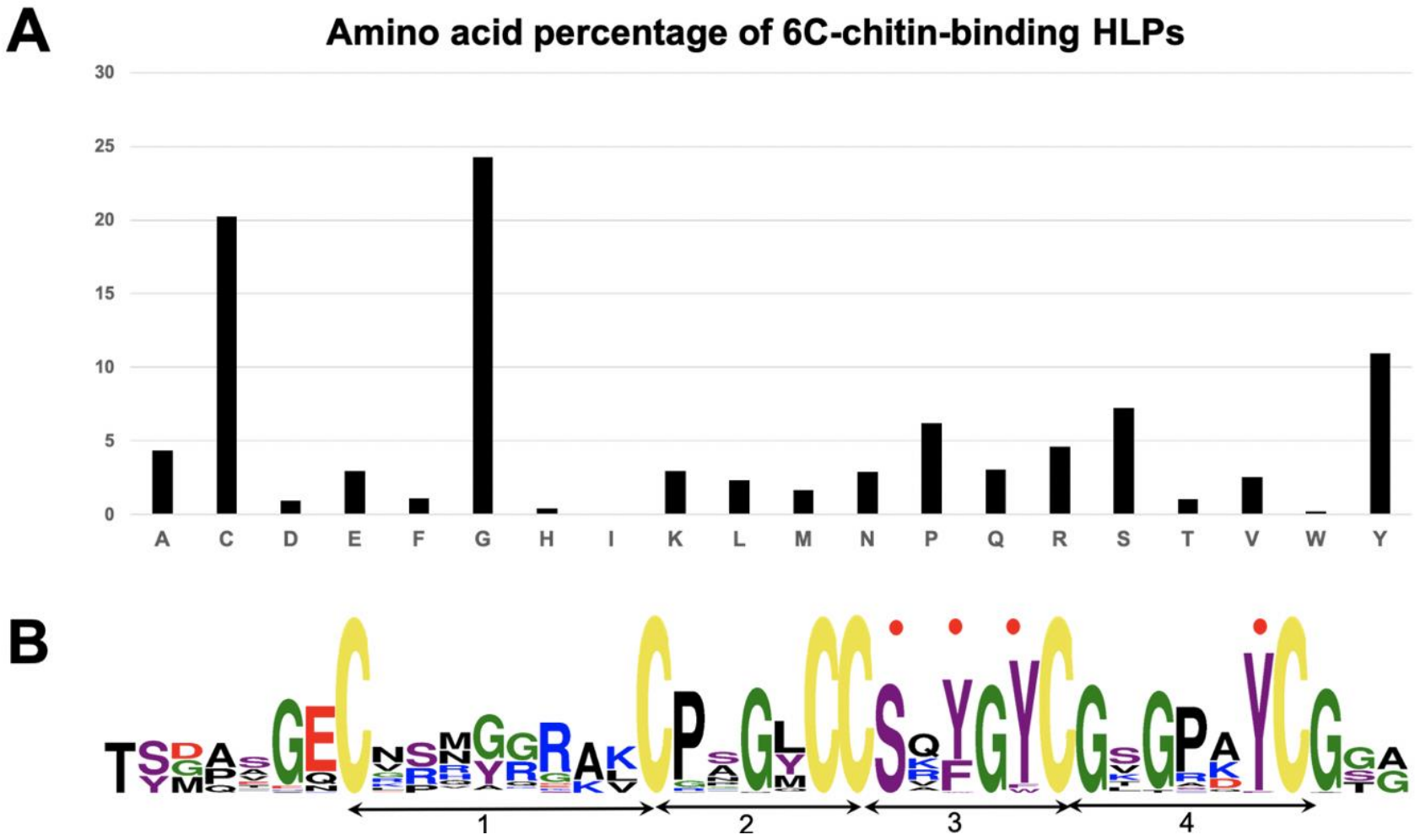


Figure 7.10: Sequence comparison of 6C-chitin-binding HLPs. (A) Total amino acid content of 6C-chitin-binding HLP in percentage **(B)** Sequence logo of 6C-chitin-binding HLP. The 6C-chitin-binding HLPs contain four loops and the residues in the chitin-binding domain are marked with a red dot above.

2.3.2 6C-non-chitin-binding hevein-like peptides

Sequence comparison of 23 previously reported 6C-non-chitin-binding HLPs with 72 putative 6C-non-chitin-binding HLPs from the EST databases showed that the total amino acid content of 6C-non-chitin-binding HLPs are rich in cysteine (17%), glycine (11%), isoleucine (5%), proline (10%), serine (6%) and tyrosine (7%) (Figure 7.11A). The average overall net charge of these peptides is neutral. A sequence logo was generated to illustrate the sequence alignment of 6C-non-chitin-binding HLPs (Figure 7.11B). Based on the sequence logo, the cysteine spacing of 6C-non-chitin-binding HLPs is absolutely conserved. Similar to the 6C-chitin-binding HLPs, the 6C-non-chitin-binding HLPs can be divided into four intercysteine loops. All four loops are highly variable and contained 7–15 amino acid residues. In contrast to 6C-chitin-binding HLPs, the 6C-non-chitin-binding HLPs do not contain a chitin-binding domain. The intercysteine loop 4 has the largest loop size comprising of 15 amino acids, whereas loop 3 has the smallest loop size, consisting of seven amino acids. On average, the loop size of intercysteine loop 1 contained six to eight amino acids. However, an exception was observed in one peptide from *Curcuma olena* where the loop 1 comprises 13 amino acid residues (Appendix C). The disulfide connectivity of 6C-non-chitin-binding HLPs was previously reported in bleogen [56] and roseltide [55] showed that they are similar to the 6C-chitin-binding HLPs (CysI–CysIV, CysII–CysV and CysIII–CysVI). Therefore, attributing to the conserved cysteine spacing of the putative and reported 6C-non-chitin-binding HLPs, their disulfide connectivity was postulated to be similar.

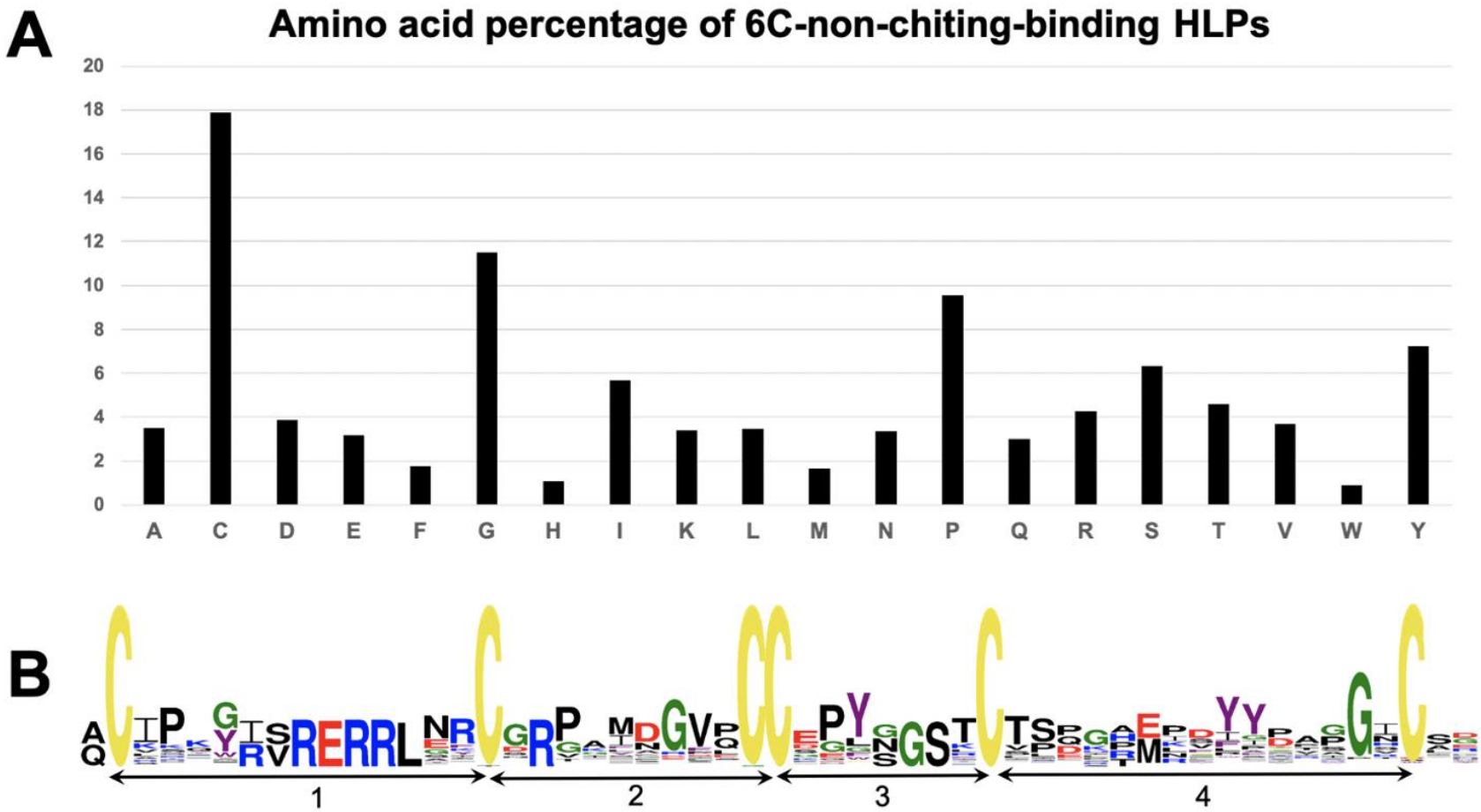


Figure 7.11: Sequence comparison of 6C-non-chitin-binding HLPs. (A) Total amino acid content of 6C-non-chitin-binding HLP in percentage **(B)** Sequence logo of 6C-non-chitin-binding HLP. The 6C-non-chitin-binding HLPs can be divided into four loops.

2.3.3 8C-chitin-binding hevein-like peptides

Sequence comparison of 887 putative 8C-chitin-binding HLPs from the EST databases with previously reported 8C-chitin-binding HLPs revealed the total amino acid content. The 8C-chitin-binding HLPs sequences are rich in cysteine (19%), glycine (18%), glutamine (10%) and serine (8%) with an average overall net charge of neutral (Figure 7.12A). A sequence logo was constructed to illustrate the sequence alignment (Figure 7.12B). The cysteine spacing of 8C-chitin-binding HLPs is highly conserved and is divided into six intercysteine loops. The sequences in all six loops varied and contained four to nine amino acid residues. Intercysteine loop 1 has the highest sequence variation, whereas loop 6 has the lowest sequence variation, suggesting the high sequence diversity within the 8C-chitin-binding HLP subfamily. Apart from the conserved cysteine spacing, the chitin-binding domain residing in intercysteine loop 3 and 4 is fairly conserved. Within the chitin-binding domain, the first aromatic residue is replaced with histidine in 27 sequences, of which one originated from bryophyta and the remaining from eudicots, belonging to the main clade of Asterids and Rosids (Appendix D). In addition, the third aromatic residue is also replaced with a histidine in 95 sequences, of which three originated from bryophyta, four from lycopodiophyta, 40 from gymnosperms and the remaining from angiosperms belonging to Asterids, Rosids, Poales, Asparagales and Ranunculales. This suggests that the peptides are closely related between all five clades. The structural study has revealed the structure of hevein [92] and based on the high conservation of the cysteine spacing, the putative 8C-chitin-binding HLPs are speculated to be similar to hevein.

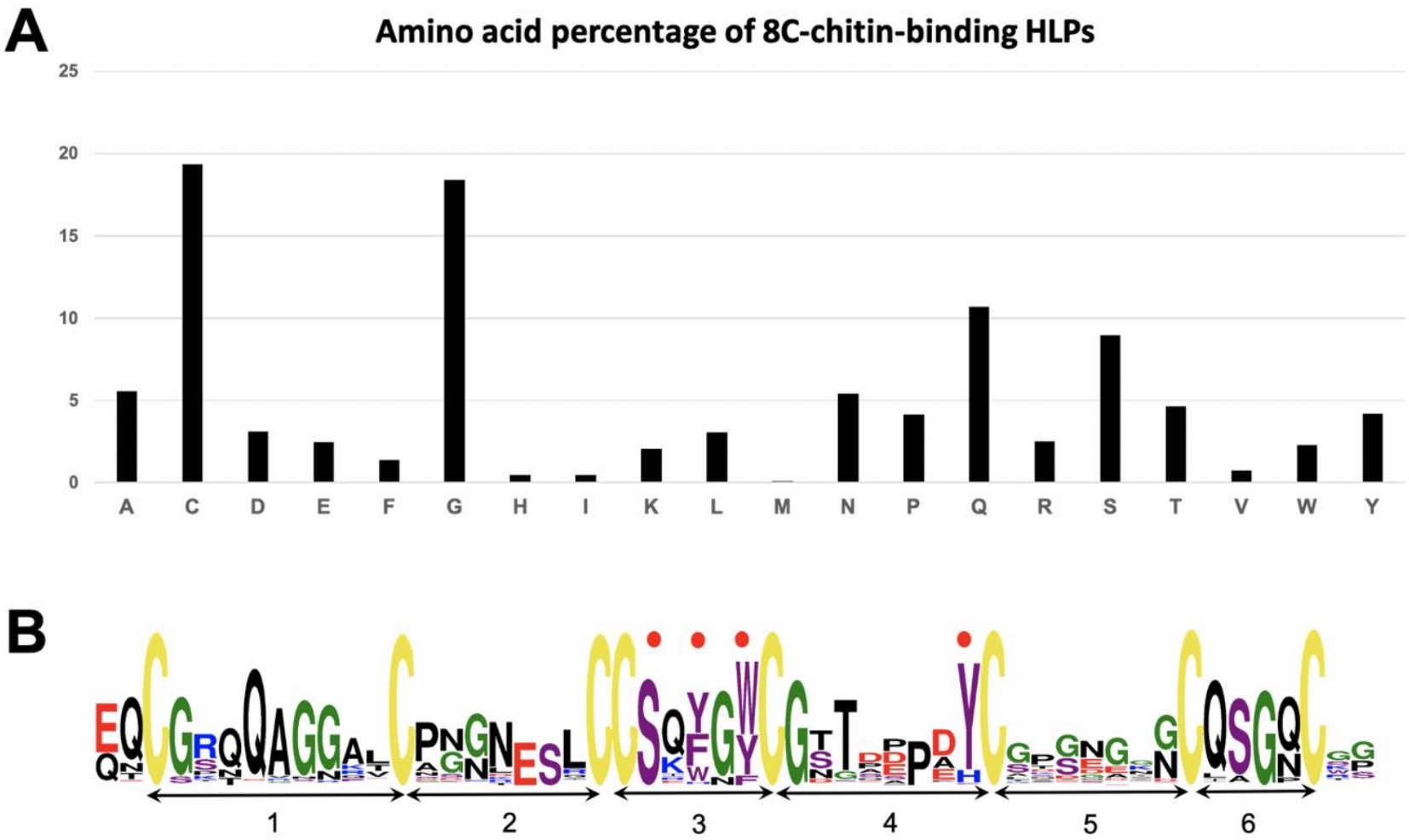


Figure 7.12: Sequence comparison of 8C-chitin-binding HLPs. (A) Total amino acid content of 8C-chitin-binding HLP in percentage **(B)** Sequence logo of 8C-chitin-binding HLP. The 8C-chitin-binding HLPs can be divided into six loops and the residues in the chitin-binding domain are marked with a red dot above.

2.3.4 8C-non-chitin-binding hevein-like peptides

Sequence comparison of 122 putative 8C-non-chitin-binding HLPs obtained from the EST databases with 14 previously reported 8C-non-chitin-binding HLPs revealed that the theses peptides were rich in cysteine (23%), glycine (17%) and proline (7%) (Figure 7.13A). In addition, these peptides have an average overall net charge of -1. Sequence alignment of the 8C-non-chitin-binding HLPs showed that their cysteine spacing is highly conserved. Like the 8C-chitin-binding HLP subfamily, the 8C-non-chitin-binding HLPs are divided into six intercysteine loops (Figure 7.13B). All the intercysteine loops were highly variable and contained 4–13 amino acid residues. None of the amino acids within the intercysteine loops were absolutely conserved. Despite the similarity, the 8C-non-chitin-binding HLPs do not contain a chitin-binding domain. Notably, the loop size of intercysteine loop 6 contained four to six amino acids. However, the loop 6 of one peptide from *Oryza sativa* comprise of 13 amino acids (Appendix E). In contrast, intercysteine loop 4 has the smallest loop size, consisting of one amino acid. The reported 8C-non-chitin-binding HLPs display a disulfide connectivity of CysI–CysIV, CysII–CysVI and CysIII–CysVII and CysV–CysVIII which differed from the 8C-chitin-binding HLPs. Due to the highly conserved cysteine spacing between the putative and reported 8C-non chitin-binding HLPs, it was assumed that their disulfide connectivity is similar. Although the 8C-non-chitin-binding HLPs contained the same cysteine motif as the 8C-chitin-binding HLPs, the cysteine spacing clearly showed a difference, suggesting diversification within the 8C-HLPs subfamily.

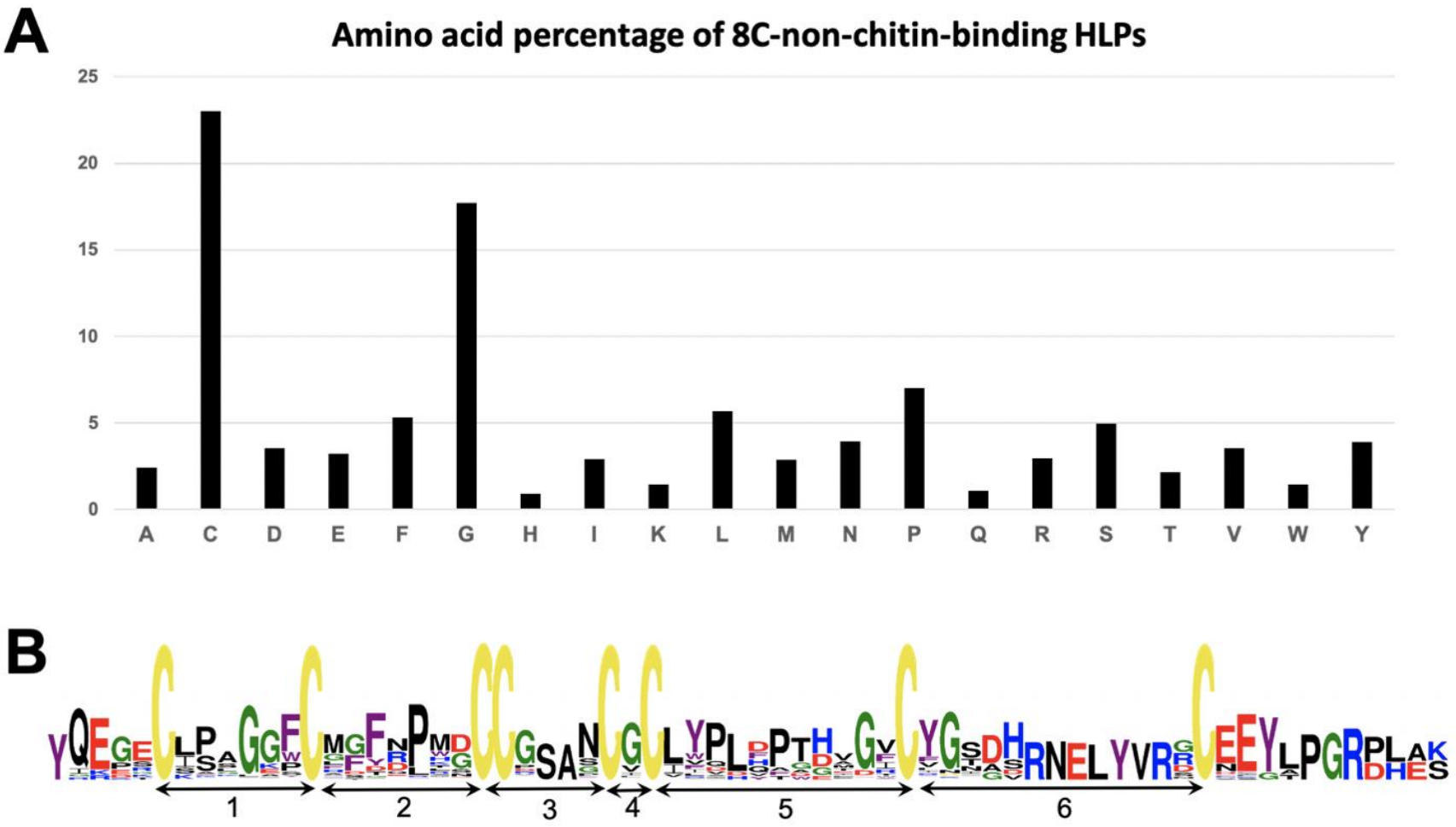


Figure 7.13: Sequence comparison of 8C-non-chitin-binding HLPs. (A) Total amino acid content of 8C-non-chitin-binding HLP in percentage **(B)** Sequence logo of 8C-non-chitin-binding HLP. The 8C-non-chitin-binding HLPs can be divided into six loops.

2.3.5 10C-chitin-binding hevein-like peptides

Sequence comparison was performed on five previously reported 10C-chitin-binding HLPs and 68 putative 10C-chitin-binding HLPs obtained from the EST databases (Appendix F). The total amino acid content of these peptides revealed a high amount of cysteine (23%), glycine (20%), glutamine (9%) and serine (10%) residues with an average overall net charge of -1 (Figure 7.14). Sequence alignment of these peptides revealed four types of cysteine spacing termed type I to type IV.

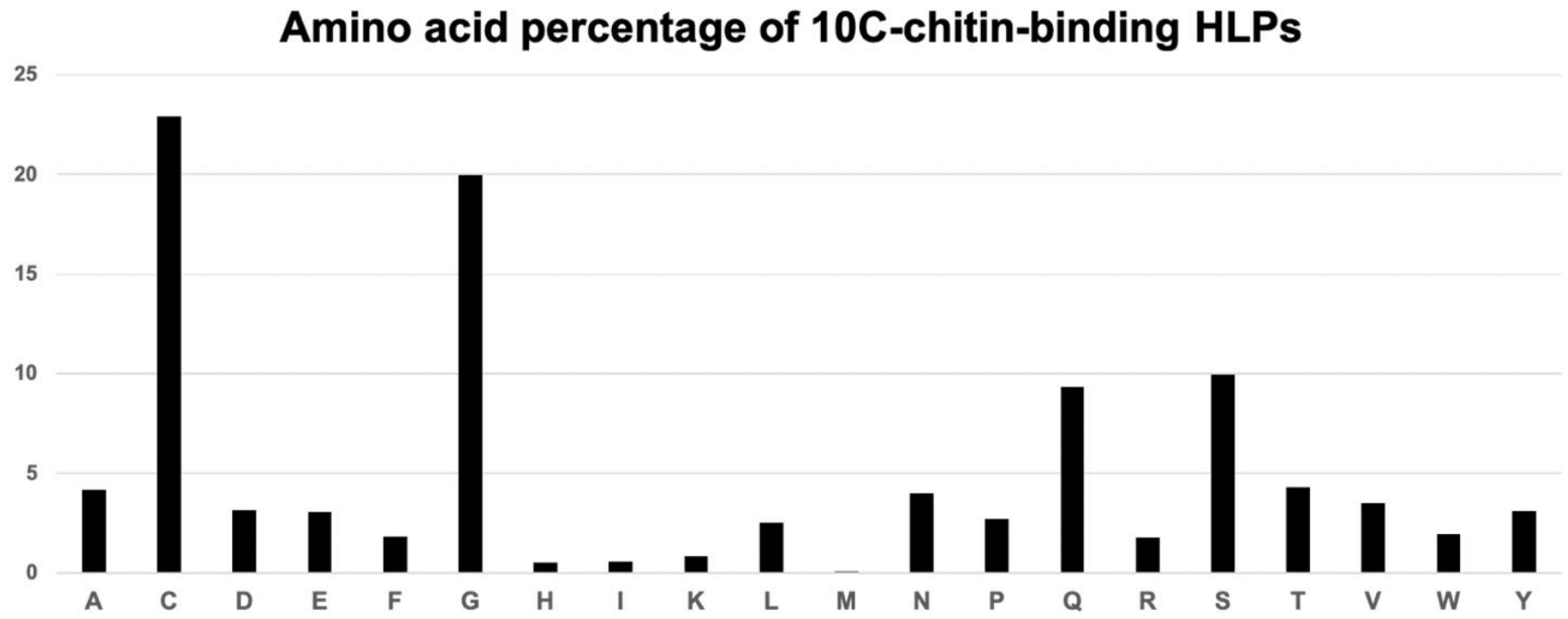


Figure 7.14: Total amino acid content of 10C-chitin-binding HLPs. The 10C-chitin-binding HLPs are rich in cysteine, glycine, glutamine and serine amino acids.

The type I cysteine spacing is similar to Ee-CBPs isolated from *E. europaeus* [110]. This cysteine spacing was observed in 39 peptides originated from monocot and eudicot angiosperms belonging to Rosids, Liliales and Asparagales. The sequence logo revealed a total of seven intercysteine loops (Figure 7.15A). Loop 1 has the largest loop size comprising eight amino acids and is highly variable, whereas loop 6 has the smallest loop size, consisting of three amino acids. The chitin-binding domain is located in loop 3 and 4 in which the first aromatic residue favors any aromatic amino acid. The second aromatic residue favors either a tryptophan or tyrosine and the third aromatic residues favor tyrosine. Interestingly, 51.3% of the peptides belong to the Rutaceae family, suggesting the abundance of 10C-chitin-binding HLPs expressed in this particular plant family. The cysteine spacing of the putative 10C-chitin-binding HLPs and Ee-CBPs was observed to be absolutely conserved. Therefore their disulfide connectivity pattern was postulated to be similar to Ee-CBP (CysI–CysIV, CysII–CysV and CysIII–CysVI and CysVII–CysX and CysVIII–CysIX) [110].

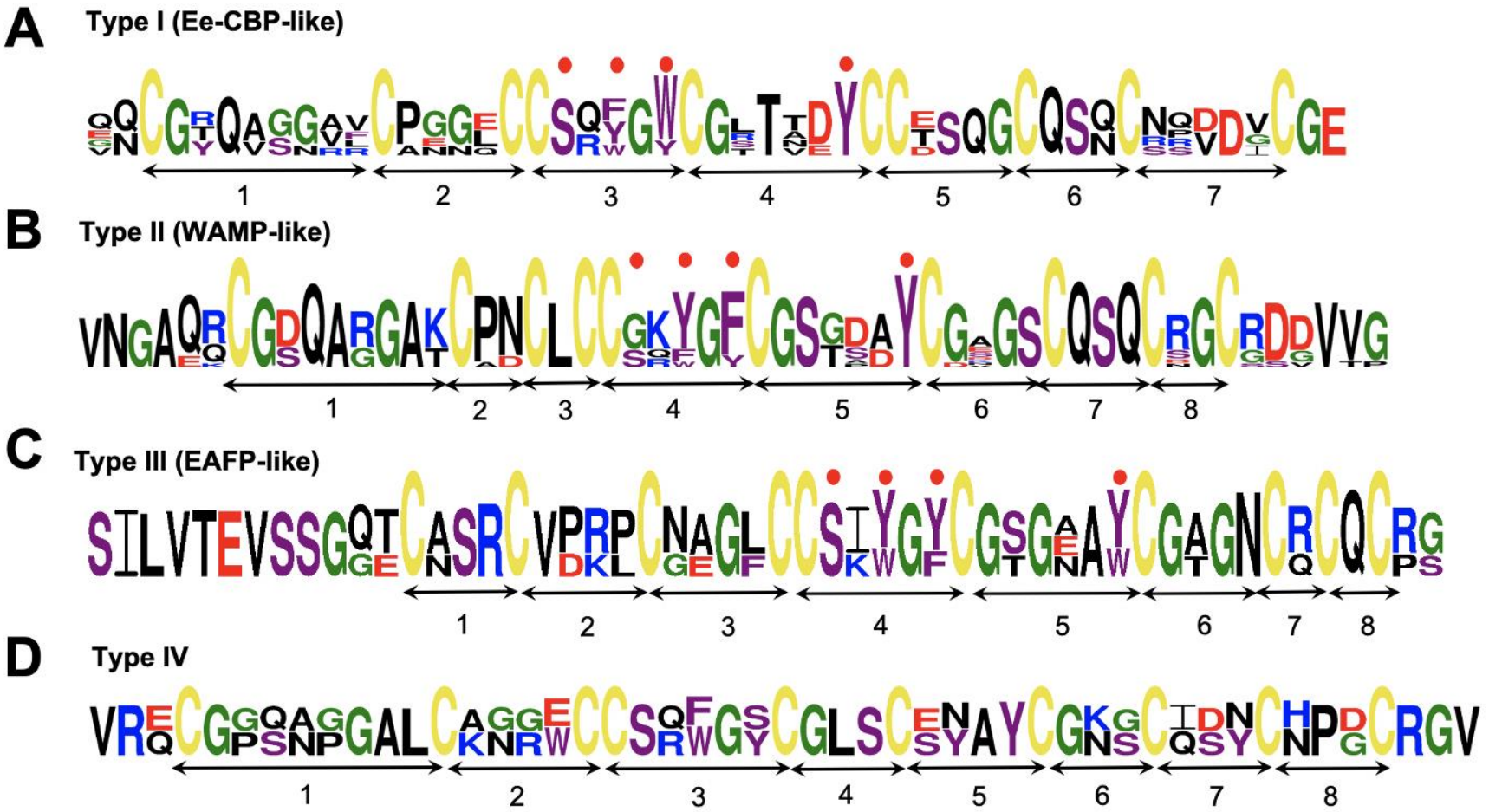


Figure 7.15: Sequence logo of 10C-chitin-binding HLPs. (A) Type I cysteine spacing pattern **(B)** Type II cysteine spacing pattern (C) Type III cysteine spacing pattern and (D) Type IV cysteine spacing pattern. Type I to III were previously reported whereas Type IV is a novel pattern found in *Cuscuta pentagonia* and *Plagiochila asplenioides* originating from angiosperms and bryophyta, respectively. The residues in the chitin-binding domain are marked with a red dot above.

The type II cysteine spacing pattern is similar to WAMPs isolated from *T. kiharae* [85]. 27 peptides derived from monocots angiosperms displayed the type II cysteine spacing. In contrast to type I, type II cysteine spacing revealed a total of eight intercysteine loops (Figure 7.15B). Sequence comparison showed that loop 1 was highly variable, whereas loops 2, 3 and 7 were highly conserved. Loop 2 comprises two residues, whereas loop 3 comprises a conserved leucine residue. Loop 7 comprises a Gln-Ser-Gln motif that is absolutely conserved. In contrast to type I, the chitin-binding domain of type II resides in loop 4 and 5. Notably, within the chitin-binding domain, the serine residue of two peptides from *Elymus spicatus* and *Triticum aestivum* was replaced with glycine (Appendix F). This similar observation was also reported in WAMPs [85]. In addition, loop 6 comprised conserved glycine and serine residues while loop 8 comprised a conserved glycine residue. Interestingly, all the putative 10C-chitin-binding HLPs containing the type II cysteine spacing belonged to the Poaceae family. Similarly, WAMPs from *T. kiharae* also belonged to the same plant family. Therefore, attributing to the conserved cysteine spacing, the putative peptides were assumed to display the same disulfide connectivity as WAMPs, following the pattern of CysI–CysV, CysII–CysVI and CysIII–CysX and CysIV–CysVII and CysVIII–CysIX [85].

The type III cysteine spacing was previously reported in EAFPs isolated from *E. ulmoides* [97]. This cysteine spacing was observed in one peptide from *Gnetum montanum*. Like type II, the cysteine spacing of type III comprises eight intercysteine loops and the chitin-binding domain is located in loop 4 and 5 (Figure 7.15C). Although the putative peptide may contain similar cysteine spacing as

EAFPS, sequence comparison revealed that the sequences were not absolutely conserved. This is likely attributed to the difference in peptide origin as the putative peptide derived from gymnosperm, whereas EAFPs derived from angiosperm. Accounting for the conservation of the cysteine spacing, the putative peptide was assumed to have similar disulfide connectivity as EAFPs that is displayed as CysI–CysV, CysII–CysIX and CysIII–CysVI and CysIV–CysVII and CysVIII–CysX [97].

The data mining of 10C-chitin-binding HLPs led to the discovery of a new cysteine spacing pattern termed as type IV (Figure 7.15D). This unique cysteine spacing was observed in *Cuscuta pentagonia* and *Plagiochila asplenioides* from angiosperm and bryophyta origin, respectively. The presence of an additional cysteine residue within the chitin-binding domain appears to be unique to the type IV cysteine spacing. Hence, additional studies are required to isolate these peptides so that their structure, disulfide connectivity and chitin-binding activity can be established.

Collectively, these data have shown the extensive molecular diversity of 6C- 8C- and 10C-HLPs that are distributed in the plant kingdom from non-vascular plants to higher plants from angiosperms. The 6C-HLPs appears to evolve later as they are less diverse compared to the 8C- and 10C-HLPs. Within the 8C-HLP subfamily, two types of disulfide connectivity, each belonging to the chitin-binding and non-chitin binding HLPs showed the diversity within its subfamily. Additionally, the disulfide pairs shuffling in 10C-HLPs resulted in four types of cysteine-spacing, expanding the subfamily diversity. Although the the primary sequences of 6C-, 8C- and 10C- HLPs varied, their cysteine spacing remains conserved within each

subfamily. Furthermore, the chitin-binding domain present in each of the chitin-binding HLP subfamily is conserved. Hence, the HLP family showed great diversity in their primary and secondary structure, yet, a degree of conservation in their cysteine spacing to maintain their structure and in the key residues of the chitin-binding HLP family to facilitate chitin-binding interaction.

2.4 Putative biosynthesis of hevein-like peptides

EST databases can provide details on the precursor sequence of a peptide. The primary sequences of the putative HLP peptides has revealed the sequence diversity. Hence, studying the bioprocessing pathways and processing sites of the precursor sequence, may provide further insights on the HLP family.

Within the chitin-binding HLP subfamily, four types of precursor organization was reported (Figure 7.16A). The type I is the most common precursor organization, consist of a signal peptide domain, a mature domain and a C-terminal tail, which has been widely reported [34, 51-53, 101]. The type II precursor organization is similar to type I, with difference in the length of the C-terminal tail. The C-terminal tail in type II consists of a protein cargo which have been reported in hevein [81]. The type III and type IV comprised of two and four tandem-repeated mature domains, and it was observed in chenotides (thesis chapter four) and avenatides (thesis chapter five), respectively. The presence of a signal peptide indicated that their bioprocessing involves the secretory pathway, and the peptides are ribosomally synthesized, which involves the cleavage of the signal peptide and the C-terminal tail to release the mature peptide.

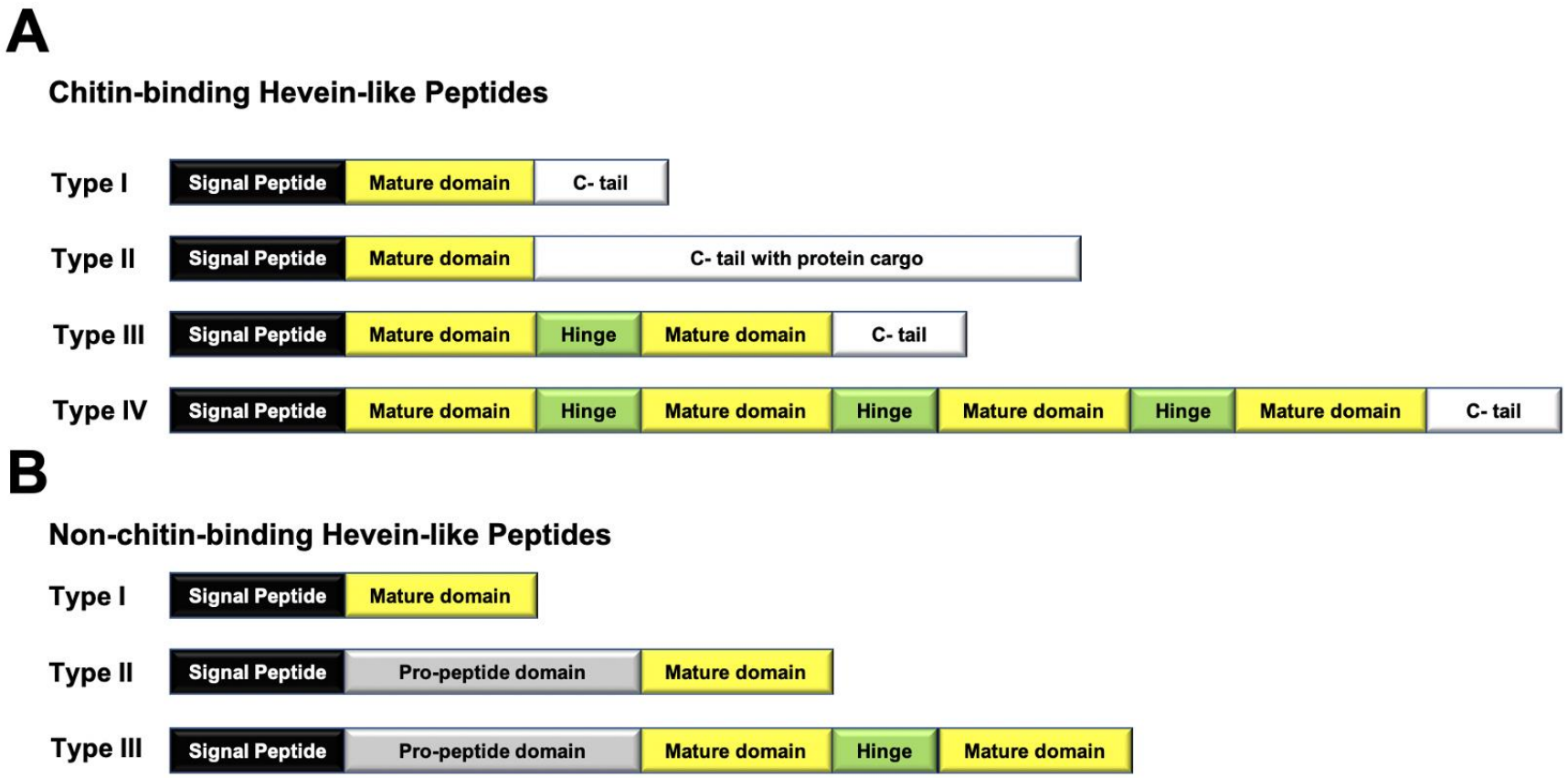


Figure 7.16: Precursor organization of hevein-like peptides. (A) Chitin-binding hevein-like peptides **(B)** Non-chitin-binding hevein-like peptides.

Precursor sequence comparison of the putatitve chitin-binding HLPs identified from the EST database revealed that they adopt either the type I or type II precursor organization. The average lengths of the signal peptide, mature domain and C-terminal tail are 23, 41 and 146 amino acids, respectively. The signal peptide is typically cleaved after an alanine or glycine residue. Interestingly, the signal peptides of two 6C-chitin-binding HLPs from *Chenopodium quinoa* were cleaved after a lysine or glutamic acid residue. Signal peptides of 8C-chitin-binding HLPs from *Akania lucens*, *Austrocedrus chilensis, Brassica napus, Callitris gracilis, Diospyros kaki, Eschscholzia californica, Papaver somniferum, Pistacia vera* and *Sphagnum palustre,* were observed to cleaved after a serine residue. In addition, the signal peptide was cleaved after an arginine or leucine residue in *Aegilops tauschii* and *Aesculus pavia*. In 10C-chitin binding HLPs, the signal peptide was cleaved after threonine and phenylalanine in *Brachypodium distachyon* and *Cuscuta pentagonia*, respectively. The C-terminal of chitin-binding HLPs are rich in glycine residues therefore, it was a challenge to determine the exact cleavage site at the C-terminal tail.

Notably, the C-terminal tail of chitin-binding HLPs varied in length with short C-tails that ranged from three to 35 amino acid residues to long C-tails with 36–631 amino acid residues. Due to the diversity in the precursor organization, the long C-terminal tail may encode other catalytic protein domains. Hence, BLAST analysis of the C-tails was performed using the NCBI conserved domain database. Of the 1010 C-terminal tails identified, 253 chitin-binding HLPs with short C-tails showed no homology to any known protein domain in the databases, whereas the remaining

peptides with long C-tail encoded for lysozyme-like domain (51.1%), PKc-like domain (0.3%), RlpA-DPBB1 domain (1.9%), Barwin domain (21.4%) GapA domain (0.1%), CE4-CICDA-like domain (0.1%) or Cerato-platanin domain (0.1%) (Figure 7.17). The lysozyme-like domain is the most common protein domain found in 6C-, 8C- and 10C-chitin-binding HLPs. Although the Barwin domain is the second most common protein domain, it is found exclusively in 8C-chitin-binding HLPs. The pKC-like domain was expressed in both 8C- and 10C- chitin-binding HLPs originating from bryophyta and angiosperm, respectively. Similarly, the RlpA-DPBB1 domain was expressed in both 8C- and 10C- chitin-binding HLP originating exclusively from bryophyta. Less common protein domains such as GapA domain, CE4-CICDA-like domain and Cerato-platanin domain were expressed exclusively in 8C-chitin-binding HLPs. Collectively, the putative chitin-binding HLPs from the EST databases adopt either the type I or type II precursor organization, which is processed via the secretory pathway. In addition, the variations in the cleavage site as well as C-terminal tail composition suggest that there could be site-specific processing enzymes involve, which warrant further studies.

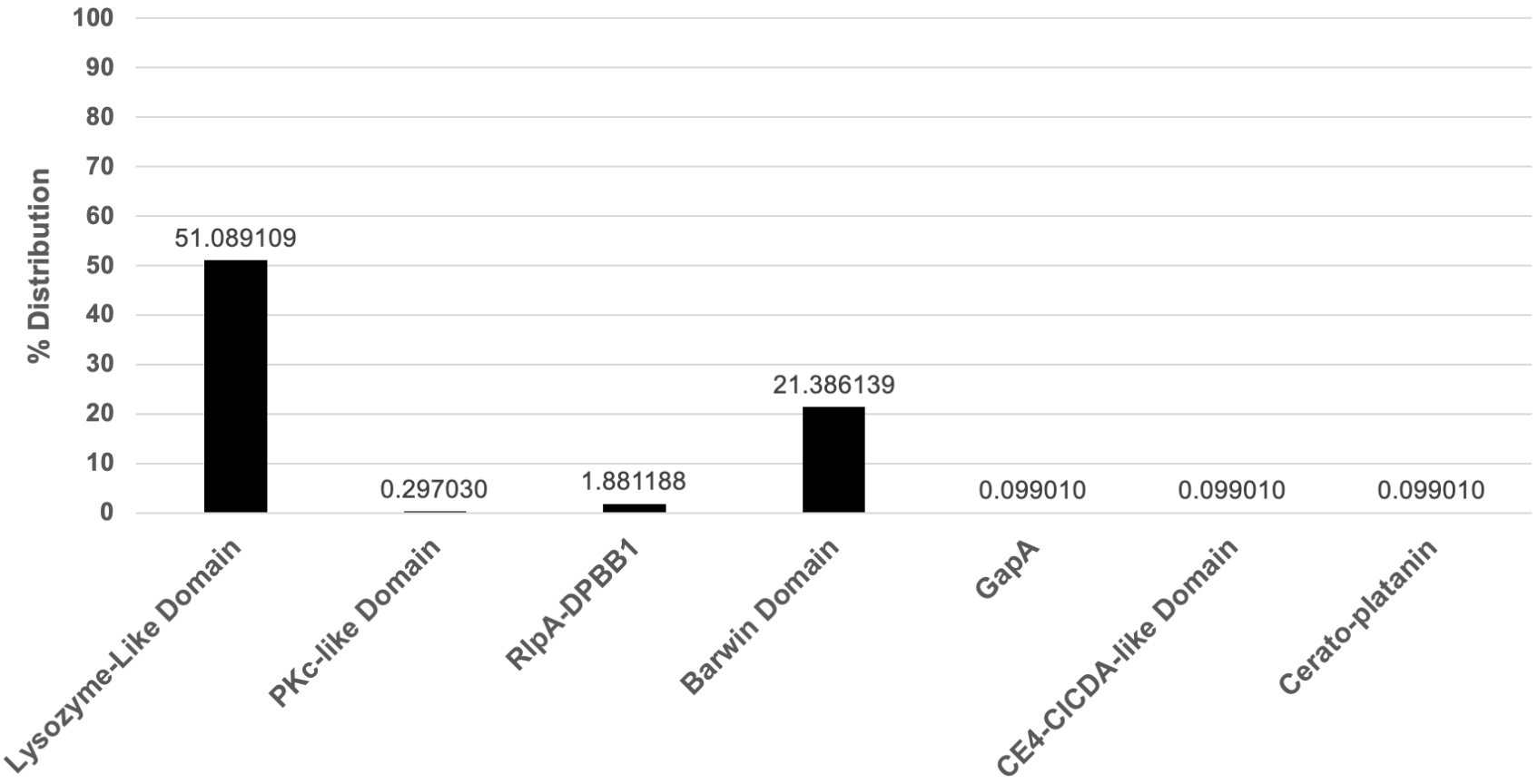


Figure 7.17: Percentage distribution of chitin-binding hevein-like peptides C-terminal tail domains. The most common protein domain expressed in these putative peptides is the lysozyme-like domain, followed by Barwin domain. Less common protein domains include pKC-like domain, RlpA-DPBB1, GapA domain, CE4-CICDA-like domain and Cerato-platanin domain.

Within the non-chitin-binding HLP subfamily, three types of precursor organization were reported [55-57] (Figure 7.16B). The type I comprise a two-domain precursor architecture consisting a signal peptide and mature domain. The type II is the most common three-domain precursor architecture consisting of a signal peptide, a pro-peptide domain and a mature domain, whereas the type III consists of a tandem-repeated mature domain. In contrast to the chitin-binding HLPs, the non-chitin-binding HLPs do not have a C-terminal tail. Similarly, the presence of a signal peptide suggests they are ribosomally-synthesized

Sequence comparison of the putative non-chitin-binding HLPs identified from the EST databases showed that they contained the type II precursor organization. The average lengths of the signal peptide, pro-peptide and mature domain are 25, 36 and 34 amino acids, respectively. Similar to the chitin-binding HLPs, the signal peptide is typically cleaved after an alanine or glycine residue. However, the signal peptide of 6C-non-chitin-binding HLPs from *Cypselea humifusum, Saponaria officianalis, Schiedea membranacea* and *Wrightia religiosa* was cleaved after a valine residue. In addition, the signal peptides of 6C-non-chitin-binding HLPs from *Alternanthera tenella* and *Atriplex prostrata* were cleaved after leucine and isoleucine residue, respectively. BLAST analysis of the pro-peptide domains of non-chitin-binding HLPs showed no homology to any known protein deposited in the NCBI conserve domain database. Collectively, the putative non-chitin-binding HLPs obtained from the EST databases adopt the type II precursor organization, and the bioprocessing of these peptides involves the secretory pathway.

3. Discussion

3.1 Overall distribution and occurrence of hevein-like peptides

The EST databases have provided great insights into the distribution, occurrence and evolution of HLPs. A total of 1204 putative HLPs were identified from 252 plant families originating from ancient bryophytes to flowering angiosperms. Furthermore, the high occurrence of HLPs in *planta* strongly indicates that they likely the largest family of CRPs. The distribution of HLPs provided information regarding their origin and evolution. Interestingly, HLPs were identified from bryophytes, which are the first terrestrial plants that diverged from green algae ancestors between 510–630 million years ago [409]. In addition, the abundance of 8C- chitin-binding HLPs in bryophytes compared to 6C-, 8C-non-chitin-binding and 10C-HLPs suggests that this subfamily could be the most ancient member of the HLP family, whereas the 6C- ,8C-non-chitin-binding and 10C-HLPs appear to be more recent members.

The variation in the number of cysteines, the position of cysteine and the length of 6C- 8C-non-chitin-binding and 10C-HLPs in the mature domain suggests that they orginated from gene mutations in 8C-chitin-binding HLPs (Figure 7.18). The 6C-HLPs appears to be the truncated variant with the deletion of two cysteine residues at the C-terminal of the 8C-chitin-binding HLPs in Asterales, Boraginales, Caryophyllales, Celastrales, Crossosomatales, Gentianales, Hypnales, Malvales, Malpighiales, Myrtales, Porellales and Rosales which resulted in divergent evolution (Figure 7.18). The 6C-HLPs could be evolutionarily selected due to their size and their biological role in plant defense. Such evolutionary selection has been reported in chitinases and snake toxin, which endows functional diversity and widen

the spectrum of targeted preys [410, 411]. In contrast, the unique disulfide connectivity of 8C-non-chitin-binding HLPs appear to diverge from 8C-chitin-binding HLPs in Apiales, Boraginales, Dilleniales, Gentianales, Jungermanniales, Malvales, Malpighiales and Poales. This divergence could be attributed to adaptive evolution, resulting in disulfide shuffling, which contributes to molecular, functional and structural diversity. Nevertheless, more studies are required to understand the underlying mechanisms responsible for the truncation and occurrence of the unique disulfide connectivity.

The 10C-HLPs are the extended variant of 8C-chitin-binding HLPs (Figure 7.18). Based on the cladogram in Figure 7.9, the 10C-HLPs evolved from 8C-chitin-binding HLPs from Fagales, Gentianales, Jungermanniales, Poales, Sapindales and Solanales. Moreover, four different types of cysteine spacing pattern were observed. It is likely after the divergence from the 8C-chitin-binding ancestors, positive Darwinian selection occurred to retain their expression, attributing to the enhanced resistance to pathogenic attacks in plants.

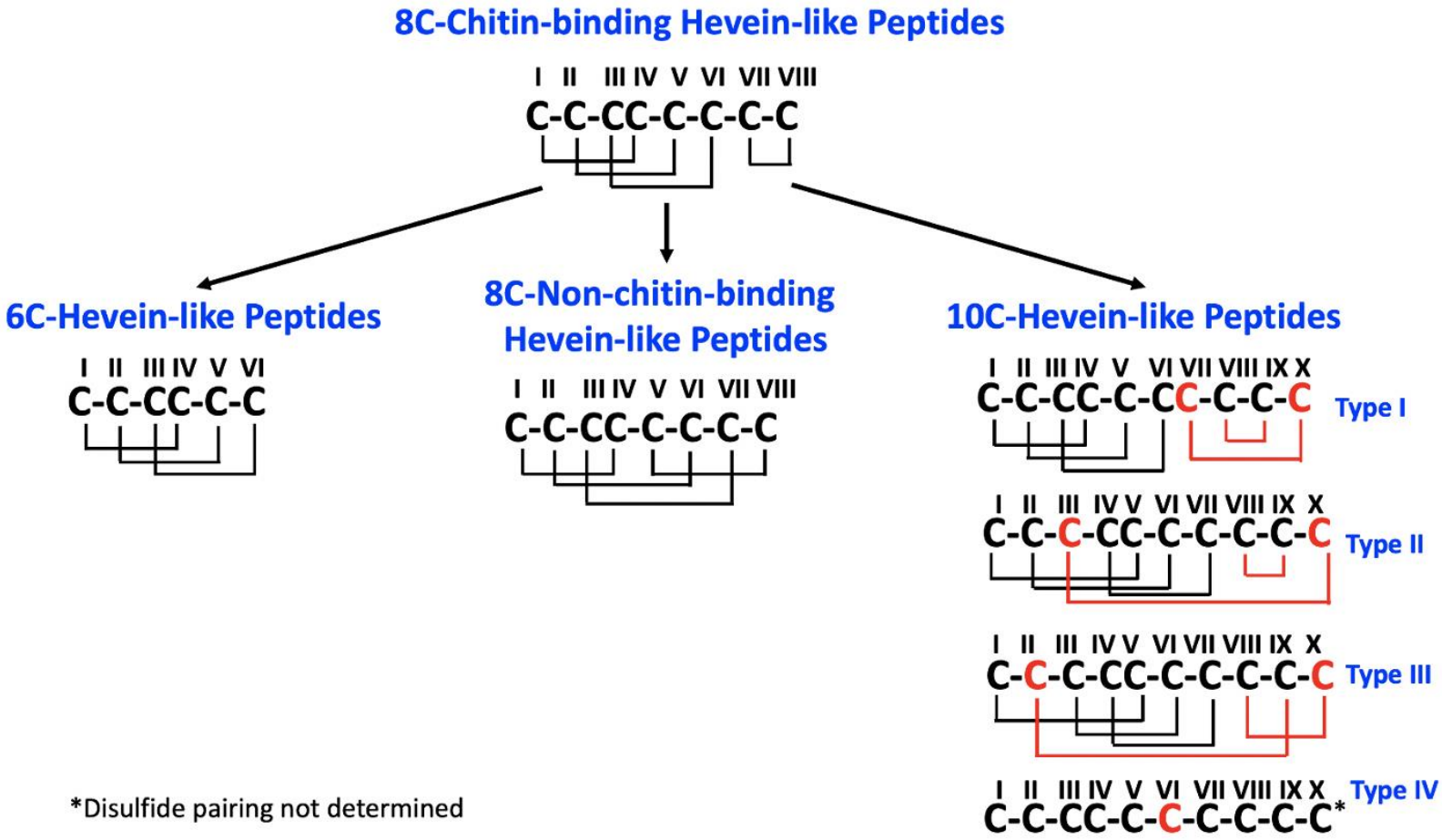


Figure 7.18: Proposed evolutionary pathways of hevein-like peptides. The 8C-chitin-binding HLPs appear to be the earliest member of the HLP family. Truncation of the 8C-chitin-binding HLPs gave rise to the 6C-HLPs. In contrast, point mutations of the 8C-chitin-binding HLPs resulted in two additional cysteine residues in 10C-HLPs which further diverged to four distinct cysteine-spacing patterns. The 8C-non-chitin-binding HLPs evolved from 8C-chitin-binding HLPs with a unique disulfide pattern.

Another proposed explanation for the evolution of 10C-HLPs suggested that WAMPs (10C-HLPs) isolated from *T. kiharae* could be the truncated variant of Class I chitinase [85]. The well-established Class I chitinases contain a chitin-binding domain in tandem to its catalytic domain [149]. A sequence analysis study revealed that the chitin-binding domains of Class I chitinase bear high sequence homology with WAMPs. Hence, it was postulated that alternative RNA splicing of Class I chitinase genes may have contributed to truncated chitinases consisting of the chitin-binding domain only [149]. Although this explanation holds true for WAMPs, it cannot account for the different cysteine spacing of the other 10C-HLPs. Therefore, this warrants further research to understand the evolution of 10C-HLPs.

The evolutionary framework of HLPs greatly differs from other family of CRPs such as defensins and cyclotides in whereby their linear precursors arose through convergent evolution [412, 413]. Collectively, the data indicate that HLPs are highly diverse ancient members of the CRP family, and the subfamilies of HLPs emerged from closely related plant species through divergent evolution.

3.2 Molecular diversity of hevein-like peptides

The extensive distribution, diversity and unique characteristics of HLPs led to the study of their molecular diversity and the consequence of functional diversity from sequence variation. Our results have revealed their molecular diversity by showing a multitude of sequence variations in their loop sizes and amino acid compositions. Within the chitin-binding HLP subfamily, the 6C chitin-binding HLPs showed fairly high conservation of sequences than the 8C- and 10C-chitin-binding HLPs. This is likely due to the relative relatedness of plant species belonging to the same plant

family and thus, leading to lower sequence variation. As for the 8C- and 10C- chitin-binding HLPs, apart from the conserved cysteine spacing, the serine and two glycine residues residing in the chitin-binding domain, none of the other residues were absolute conserved. This suggests that the residues are essential for stabilizing the configuration of the peptide during chitin interaction. It is commonly accepted that variable non-essential residues that are not associated with biological function contribution are inclined to mutations compared to conserved residues which normally reside in the inner pocket of the peptide [414]. Yet, the residues responsible for chitin-binding in these peptides were exposed on the surface and mutations of these key residues were observed. For example, the substitution of serine residue to glycine in 10C-HLPs was observed in four peptides from Poaceae and Rutaceae plant families. This similar observation was reported in WAMPs isolated from *T. kiharae* [85]. As a consequence of the point mutation, the chitin-binding affinity was reduced. Nevertheless, the reduced binding affinity was compensated by an additional serine residue in intercysteine loop six, resulting in the additional function of inhibiting fungal chitinases [149]. Besides the inhibition of fungal chitinase, the substitution of glycine and the addition of serine altered their overall net charge which led to a gain-in-function against bacteria [138].

Similarly, in 8C-chitin-bindng HLP subfamily, the tyrosine residue involved in chitin-binding interaction is replaced by histidine in 4.5% of gymnosperms and 6.2% of angiosperms. Therefore, it is postulated that horizontal gene transfer may have occurred. However, further studies are required to understand the gene transfer phenomenon to confirm the hypothesis. Although mutations were observed in the

key residues, the chitin-binding interaction might not be significantly reduced as the histidine residue consists of an imidazole ring that can facilitate CH-π stacking interaction between the peptide and chitin polymer. Therefore, it is likely that many mutations have occurred during the evolution of chitin-binding HLPs to provide the plants with a myriad of defense peptides to target a plethora of pathogens [415].

In the non-chitin-binding HLP subfamily, apart from their conserved cysteine spacing, there was no absolute conservation of any residues. They also do not contain a chitin-binding domain, and hence their biological function remains to be elucidated. Regardless, a study has shown that bleogen pB1 (6C-non-chitin-binding HLP) isolated from *P. bleo* exhibited anti-candida and heparin-binding properties [56]. In addition, roseltides (6C-non-chitin-binding HLPs) isolated from *H. sabdariffa* exhibited neutrophil inhibitory [55], proteasomal inhibitor [163] and mitochondrial-targeting properties [162]. Thus, this suggests that the molecular diversity in the non-chitin-binding HLPs confers functional diversity. Interestingly, 41.8% of the 8C-non-chitin-binding HLPs contain cysteine residues at both N- and C-terminal. In addition, 16.4% of the 8C-non-chitin-binding HLPs contain a pyroglutamic acid and cysteine at the N- and C-terminal, respectively. The presence of these amino acid residues at each terminus of these peptides confers highly stability against endopeptidases. This was observed in ginsentide TP1 (8C-non-chitin-binding HLP), which resulted in a pseudocyclic structure that is highly resistant to heat, acid and proteolytic degradation [57]. As such, further studies are necessary to elucidate the biological roles of the putative non-chitin-binding HLPs to gain insights on the effects of molecular diversity in this subfamily.

All HLPs contained a knotted-topology that is established by three disulfide linkages. This knotted-topolgy is present in many CRPs such as CKAIs and conotoxin from marine cone snails [73, 122] and additional disulfide linkages can be found in 8C- and 10C-HLPs which may contribute to functional diversity. A observation was reported in conotoxin where the additional disulfide bonds can modify the specificity of the peptides towards the receptors [416]. Collectively, the HLPs have shown their molecular diversity and its impact on functional diversity. Nevertheless, the important characteristic of the conserved cysteine spacing in the primary sequence suggests the uniformity in the diversity of this HLP family.

3.3 Biosynthetic diversity and precursor architecture of hevein-like peptides

The full-length precursor sequences provided insight on the different types of HLPs precursor organizations. The precursor sequences of all HLPs consist of a signal peptide, which indicated that these peptides are ribosomally synthesized. Within the chitin-binding HLPs, four distinct precursor organizations were identified (Figure 7.16A). The majority of the chitin-binding HLPs either adopt a type I or type II precursor architecture, which comprises a signal peptide, a mature domain and either a short or long C-terminal tail. This precursor organization is also observed in CRPs like thionins [61]. In addition, the natural diversity of the C-terminal tail of chitin-binding HLPs is evident. Short C-terminal tail does not contain a protein cargo. In contrast, long C-terminal tail encoded for protein cargo such as lysozyme-like domain, Barwin domain and less common protein domains like gapA domain and RlpA-DPBB1 domain. For example, the Barwin domain is known for binding to N-acetylglucosamine, whereas GapA domain is involved in glycolysis [89, 417].

These chimeric precursors of chitin-binding HLPs can enhance the binding efficiency of peptides to the fungal cell wall and digest the nascent chitin chains [89] and the mechanism is similar to Class I chitinases [108]. Therefore, it is postulated that Class I chitinases evolved from chitin-binding HLP precursors that contained chimeric catalytic domain. Such precursor organization comprising of a long C-terminal tail was reported in CRPs like cyclotides from *Clitoria ternatea* where the C-tail encodes for pea albumin protein [316].

The type III and IV precursor organizations containing tandem-repeated domains are less common compared to type I and type II. These tandem-repeated precursor architectures can be observed in carbohydrate-binding proteins like UDA1 [356] and WGA [357] isolated from *Urtica dioica* and *Triticum aestivum*, respectively. It is likely that gene duplication has occured during the evolution of chitin-binding HLPs. Furthermore, it has been well-established that chitin-binding HLPs play a vital role in plant defense. As such, the increased gene dosage is considered an evolutionary advantage, and duplication can confer enhanced fitness benefit for the plant [418].

The non-chitin binding HLPs consist of three distinct precursor organizations (Figure 7.16B). The majority of the non-chitin-binding HLPs adopt the type II precursor organization consist of a signal peptide, a pro-peptide and a mature domain. Similarly, the presence of a signal peptide indicated that the non-chitin-binding HLPs are ribosomally synthesized and involved the secretory pathway. A similar precursor organization was reported in CRPs like CKAIs [122]. On the contrary, the type I and type III precursor architectures are less common. The type

I consist a signal peptide and a mature domain, whereas the type III consist a signal peptide, a pro-peptide and tandem-repeated mature domains that were reported exclusively in bleogen isolated from *P. bleo* [56]. Although the precursor organization of non-chitin-binding HLPs contains a pro-peptide, the function of these pro-peptides has yet to be established as they showed no homology to any known protein domains in the NCBI conserved database. However, some studies have shown that the pro-peptide can facilitate the proper folding of peptides into their structure during peptide synthesis [165, 166]. Therefore, further studies are required to gain deeper insights into the function of the pro-peptide.

Collectively, the natural diversity of the HLPs precursor organization is the consequence of divergent evolution, which is likely responsible for their functional diversity. Therefore, HLPs may possibly be the largest family of CRPs.

3.4 Potential application of hevein-like peptides

The discovery of novel HLPs appeared to be more abundant than expected. With the additional knowledge obtained, this diverse family of CRPs showed potential value as plant-derived agrochemical and peptide biologics for biotechnological applications.

3.4.1 Development of pathogenic-resistant crops

The increasing threat of plant fungal pathogens has majorly affected a multitude of agricultural crops which led to significant losses in the many agribusinesses [3]. Hence, this urges the importance of combating this issue by developing an effective method to manage these fungal pathogens. The growth of transgenic crops with enhanced resistance to fungal invasion presents a promising solution. Defensins

are a family of CRPs that are well-studied and have shown their potential to target fungal pathogens [46, 419, 420]. For example, the expression of wasabi defensin WT1 in crops like potato and rice have shown to increase resistance against *Botrytis cinerea*, *Erwinia carotovora* and *Magnaporthe grisea* [46]. Furthermore, the overexpression of pnAMP-h2 cDNA (8C-chitin-binding HLP) in transgenic tobacco plants enhanced their resistance against *Phytophthora parasitica,* which is the source of black shank disease [421]. Therefore, with the current understanding of the sequence, structure and antimicrobial activity of HLPs, they can be employed to develop transgenic crops with better resistance against a plethora of fungal pathogens. Moreover, they present as less harmful alternatives to chemically derived anti-fungal agents and their abundance as well as the ease of expression in plants further encourage their utilization in transgenic crop development.

3.4.2 Scaffold for development of orally active peptidyl bio therapeutics

The majority of CRPs have a cystine-knot motif, which serves as a scaffold for peptide grafting to produce bioactive peptides with novel biological activities. Hence, the cystine-knotted HLPs contained several characteristics that endow them as potential candidates for engineering peptidyl biotherapeutics with enhanced bioavailability and stability *in vivo*. Firstly, the abundance of HLPs in plants makes them readily available via many techniques such as direct peptide isolation, molecular cell expression and chemical synthesis. Secondly, the intercysteine loops of HLPs are highly variable and hence show promising potential to graft bioactive peptides [422, 423]. Moreover, this method has been widely employed on cyclic scaffolds from trypsin inhibitors, kalata B1, as well as linear scaffold from

Ecballium elaterium [424-426]. For instance, the grafting of a six-residue arginine-rich sequence into kalata b1 produced an antiangiogenic agent which function as an antagonist for vascular endothelial growth factor receptor binding [424]. A similar approach was also used to graft three proangiogenic sequences into MCoTI-II trypsin inhibitor from *Momordica cochinchinensis,* which displayed promising angiogenic activity [425]. Furthermore, many of these cystine-knot peptides have been engineered to enhance their affinity for integrins that enable interaction with the target molecule to become scaffolds for radioactive imaging probes [426, 427]. For example, the engineered (99m)Tc-SAAC-S02 demonstrated its potential as a single-photon emission computed tomography (SPECT) agent to detect the overexpression of integrin αvβ6 expression in tumors, which may provide early cancer diagnosis and better prognosis [426]. Finally, the cystine-knot scaffold of HLPs endows exceptional stability to proteolytic degradation, which enhances the bioavailability of grafted peptide [34, 59]. Furthermore, the carbohydrate-binding property of chitin-binding HLPs can be used to develop peptide biologics which targets glyco-peptides on the cell membrane. The high specificity of these peptides, coupled with its high bioavailability, can overcome the limitation of off-target effects prevalent in small-molecule drugs.

Taken together, this study has shown the distribution, evolution, sequence and biosynthetic diversity of HLPs using the bioinformatic approach. The extensive distribution of HLPs in bryophytes, euphyllophyte. lycopodiophyta gymnosperms and angiosperms suggest their prevalence and importance in plant defense. The sequence diversity of 6C-, 8C- and 10C-HLPs showed that their cysteine-spacing

and key-residues involved in chitin-binding are conserved. Insights on the precursor organization of HLPs showed great diversity. While some of the C-terminal tail of chitin-binding HLPs encode functional protein cargos such as lysozyme-like and Barwin domains, the function of the pro-peptide in non-chitin-binding HLPs remains to be elucidated. As such, these findings allowed us to speculate the evolutionary pathway of HLPs in which the 6C-, 10C- and 8C-non-chitin-binding HLPs arose from 8C-chitin-binding HLP ancestors through divergent evolution of closely related plant families.

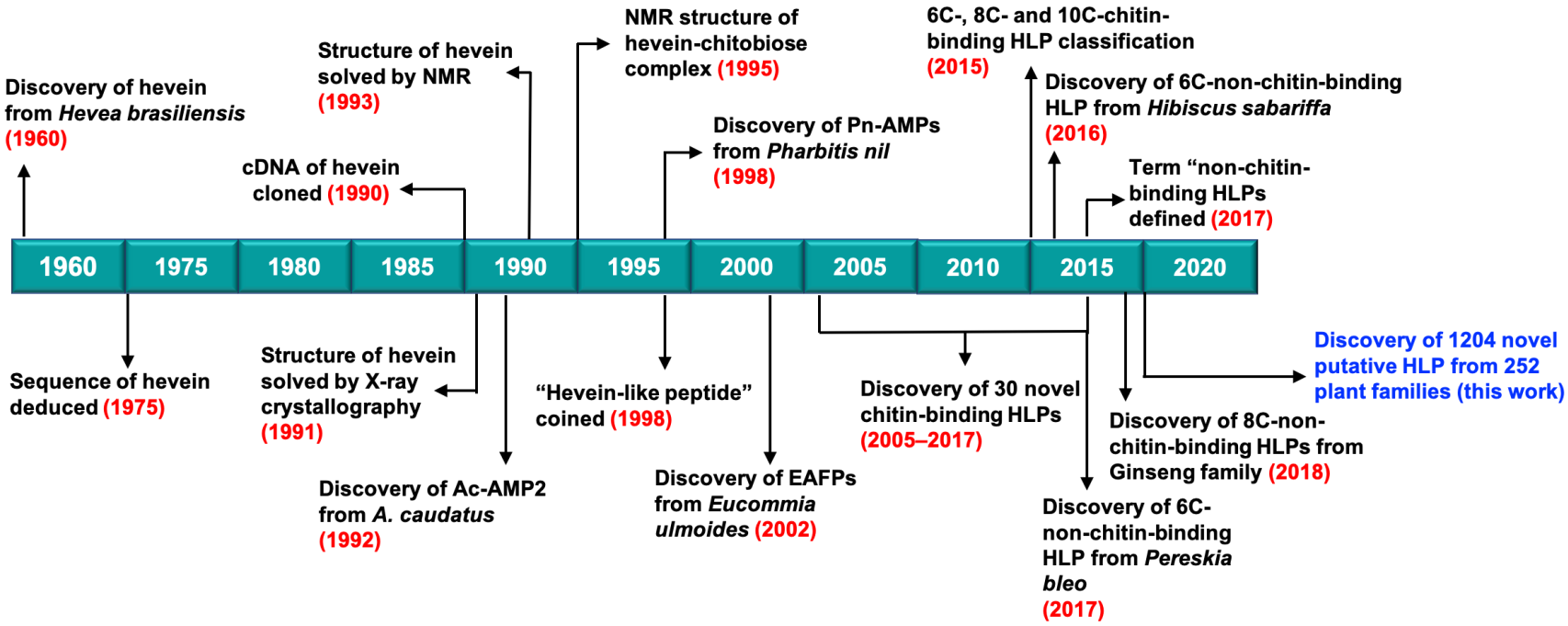


Figure 7.19: Revised timeline of hevein-like peptides. Approximately 60 years after the discovery of hevein, our study on the distribution and occurrence of HLPs in plants has expanded current knowledge on this peptide family. A total of 1204 novel putative HLPs discovered through EST-based bioinformatic approach suggest that the HLPs could possibly be the largest family of CRPs.

Significance, Conclusion and Future Outlook

Hevein-like peptides are a family of CRPs that play an important role in plant defense which include anti-fungal activity. In this thesis, I provide evidence that the presence of chitin-binding HLPs provides an evolutionary advantage for plant defense in all functional grains which include rice, barley, wheat, sorghum, rye, oat and millet. In addition, my bioinformatics study shows that HLPs are prevalent in plants and suggests that they represent a hitherto unknown large family of CRPs.

This family of CRPs can be broadly divided into two subfamilies, namely chitin-binding and non-chitin-binding HLPs. The subfamilies can be further divided based on the cysteine content and the precursor organization. The molecular diversity and prevalence of HLPs in functional foods and medicinal plants have prompted us to isolate and characterized novel HLPs from functional foods such as *Chenopodium quinoa* and *Avena sativa* as representatives from pseudocereal and cereal, respectively, and medicinal herbs like *Eleutherococcus trifoliatus*.

Chapter four discusses the isolation and characterization of three novel 6C-chitin-binding HLPs, collectively termed as chenotides, from *C. quinoa*. This finding has led to the expansion of the existing 6C-chitin-binding HLPs from 12 to 15. Our data reveal the high molecular diversity of chenotides. In addition, NMR analysis and disulfide mapping of chenotide cQ2 show the presence of three disulfide bonds with a disulfide connectivity of CysI–CysIV, CysII–CysV and CysIII–CysVI, forming a cystine-knot that endows exceptional metabolic stability against heat, acid and proteolytic degradation. Since the chitin-binding assay has established the chitin-binding property of chenotide cQ2, the anti-fungal effects of chenotides was also

studied. Chenotide cQ2 displays a broad spectrum of anti-fungal effect on four phytopathogenic fungi strains with micromolar IC_{50}. Microscopic analysis of fungal strains treated with cQ2 reveal the growth retardation of hyphae with swollen hyphal tips, and increased short branching indicated that chenotides exerts anti-fungal effect by impeding hyphal elongation at the growing tip of the mycelia in a dose-dependent fashion. Proteomic studies showed that chenotides are synthesized as a precursor protein and processed to produce the mature chenotide peptide. Importantly, the precursor organization of chenotides differed from previously reported 6C-chitin-binding HLPs. Their precursor organization includes a signal peptide, an identical tandem-repeated mature domain and a short C-terminal tail, suggest that the biosynthesis involves the secretory pathway. Thus, chenotides are the first suite of 6C-chitin-binding HLPs reported to contain an identical tandem-repeated mature domain. Collectively, these findings provide new insights into the 6C-chitin-binding HLP subfamily, and the knowledge gained can be useful for biotechnological applications such as developing potent anti-fungal agents against fungal pathogens and producing transgenic crops with enhanced resistance to fungal invasions.

Chapter five describes the isolation and characterization of nine 8C-chitin-binding HLP termed as avenatides from *A. sativa*. Avenatides contain 28–39 amino acid residues in length and sequence comparisons with previously reported 8C-chitin-binding HLPs revealed the conserved cysteine-spacing, which is divided into six intercysteine loops. Structural analysis showed avenatide aV1 contained four disulfide bonds that display a disulfide connectivity of CysI–CysIV, CysII–CysV,

CysIII–CysVI and CysVII–CysVIII, forming a compact structure, conferring aV1 with high metabolic stability against thermal, acid and enzymatic degradation. Chitin-binding studies of aV1 revealed that it contained chitin-binding property and elution of aV1 from chitin beads required heating in high acid concentration suggests the strong binding interaction between aV1 and chitin. Consequently, this enables aV1 to inhibit phytopathogenic fungal strains with IC_{50} in the micromolar range. Notably, the transcriptomic analysis revealed the precursor organization of avenatides, which differed from previously reported 8C-chitin-binding HLPs. The precursor organization includes a signal peptide, four tandem-repeated mature domains and a short C-terminal tail similar to chitin-binding proteins. Phylogenetic analysis confirmed the close evolutionary relationship between avenatides and chitin-binding proteins. However, further analysis of the hinge region strongly suggests that the hinge region's composition may have a role in determining the cleavage of the tandem repeats, distinguishing avenatides from chitin-binding proteins. Taken together, avenatides represent the first suite of 8C-chitin-binding HLPs with tandem repeats. This finding expanded the library of 8C-chitin-binding HLPs, provided additional knowledge on the sequence, structure and biosynthetic diversity, and could provide clues to their evolutionary pathway.

Chapter six reports the isolation and characterization of a novel 8C-non-chitin-binding HLP termed as eleutide from *E. trifoliatus.* Eleutide eT1 consists of 35 amino acid residues. Sequence analysis with previously reported 8C-non-chitin-binding HLPs shows a conserved cysteine-spacing that is divided into six intercysteinyl loops. Interestingly, the N- and C-terminal of eT1 contain a

pyroglutamic acid and cysteine, respectively, confering protection against endopeptidases. Structural modeling of eT1 revealed four disulfide bonds with a disulfide connectivity of CysI–CysIV, CysII–CysVI, CysIII–CysVII and CysV–CysVIII, forming a cystine-dense pseudocyclic structure. This pseudocyclic structure of eT1 displays exceptional stability against heat, acid and proteolytic degradation. The precursor organization of eleutide eT1 adopts a three-domain architecture that includes a signal peptide, a prop-peptide and a mature domain, suggesting that the peptide is ribosomally synthesized. Functional analysis of eleutide eT1 reveal that it displays cytoprotective effect, increases cell proliferation and promotes cell migration, suggesting its potential therapeutic properties. Together, these findings expanded the current library of 8C-non-chitin-binding HLP library and provided new insights into the sequence, structure and functional diversity within the subfamily. As such, eleutide eT1 shows promising value as a scaffold for peptide grafting of metabolically stable orally active biotherapeutic.

Chapter seven discusses the diversity, distribution and evolution of the HLP family using an EST-based bioinformatic approach. A total of 1204 novel HLPs were discovered from 252 plant families belonging to 89 orders. The extensive distribution and high incidence of HLPs strongly suggest that they are likely the largest family of CRPs constitutively expressed. Sequence analysis of 6C- 8C- and 10C-HLPs highlighted the molecular diversity of the HLP family. While most of the peptides had conserved cysteine spacing pattern, the 10C-HLPs had four different cysteine spacing patterns, which were designated as type I to type IV based on their disulfide connectivity. The cysteine pattern of type I to III has been reported

previously, whereas type IV is a novel cysteine pattern that was discovered through the EST-based search of this work from plant species *Cuscuta pentagonia* and *Plagiochila asplenioides.* Additional studies are to be conducted to isolate the peptides from these plants to resolve the structures and disulfide pairings. The flexibility and cysteine-shuffling in the 10C-HLPs further emphasize the molecular diversity of the HLP family. The precursor sequence of these HLPs provided insights on the biosynthetic diversity. Notably, the C-terminal tail's diversity in chitin-binding HLPs revealed that a fraction of HLPs contain chimeric precursors in which the C-tail encodes for protein cargo such as for lysozyme-like, Barwin, RlpA-DPBB, PKc-like, CE4-CICDA-like, Cerato-platanin or GapA domains. These protein cargos function as additional catalytic domains that further contribute to the functional diversity of chitin-binding HLPs. On the contrary, the pro-peptides of non-chitin-binding HLPs bear no homology to any known proteins, and hence, their function remains to be elucidated. These findings shed light on the evolution of HLPs and proposed that the 8C-chitin-binding HLP subfamily is the most ancient member of the HLP family whereas the 6C-, 8C-non-chitin-binding and 10C-HLPs evolved from the 8C-chitin-binding HLPs through divergent evolution. The 6C-HLP evolved from the C-terminal truncation of 8C-chitin-binding HLPs, whereas the 10C-HLPs evolved from point mutations in 8C-chitin-binding HLPs, resulting in two additional cysteine residues. Lastly, the 8C-non-chitin-binding HLPs evolved from the 8C-chitin-binding HLPs through the shuffling of disulfide bonds resulting in a different disulfide connectivity. Together, this study provides a comprehensive

understanding of the HLP family with regards to their distribution, evolution and molecular diversity.

Significance and conclusion: My thesis attempts to answer two major questions relating to a family of underexplored plant defense peptides namely, hevein-like peptides. The first question relates to why they are important, particularly in functional grains and medicinal herbs. The second question relates to the prevalence of hevein-like peptides in the plant kingdom. To answer the first question, we describe the discovery and characterization of novel HLPs from *C. quinoa*, *A. sativa* and *E. trifoliatus,* and determine their sequence, biosynthetic, structural and functional diversity. In addition, the non-chitin-binding HLPs represent a new front which are highly underexplored, presenting a new research angle to discover their potential biotherapeutic function in humans. To answer the second question, we perform a bioinformatic search of putative HLPs, which expand the existing HLPs from 75 to 1204. The abundance of HLPs distributed across the plant kingdom further supports that they are an evolutionary selected trait that has been retained in plants, attributing to their defense function (Figure 8). In summary, my thesis provides knowledge on the distribution and evolution of HLPs and a better understanding of this family. Importantly, my work highlights the molecular diversity of HLPs as one of the largest family of CRPs. In addition, these results can be used for the agricultural and pharmaceutical industries for the development of plant-derived anti-fungal agents and orally stable peptide biologics.

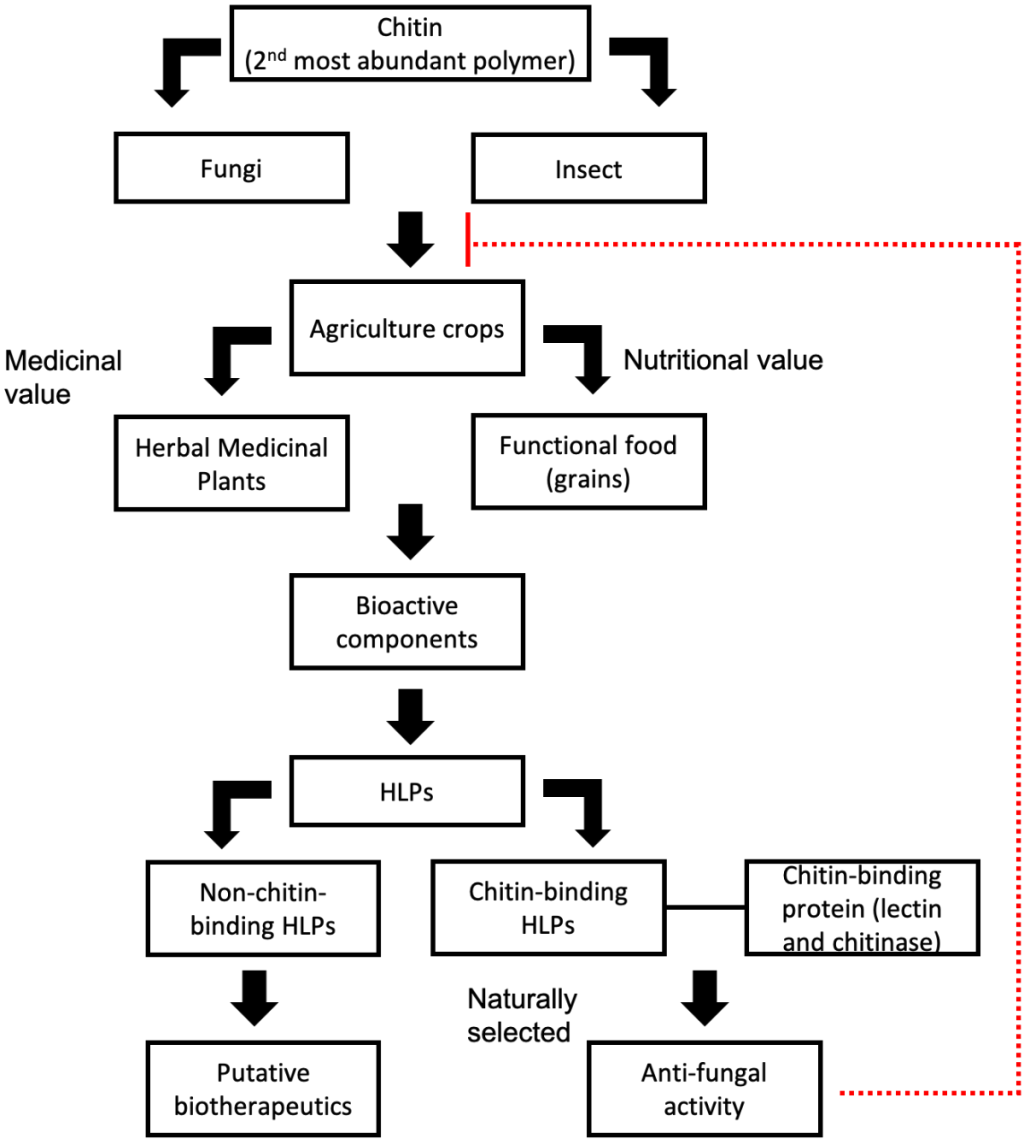


Figure 8: Summary of hevein-like peptides' impact on plant, pathogens and its potential biotechnological applications. Chitin is the 2nd most abundant polymer that can be found in fungal cell wall and the exoskeleton of insects. These fungi and insects constantly pose threats to agricultural crops which are grown commercially for their medicinal and nutritional value. In response to these pathogen invasions, these crops contained bioactive component, particular HLPs which are constitutively expressed and are broadly divided into chitin-binding and non-chitin-binding HLPs. The non-chitin-binding HLPs are speculated to contain putative biotherapeutic functions whereas the chitin-binding HLPs are evolutionary selected for their anti-fungal properties to response to pathogenic attack. Therefore, the wide distribution and abundance of HLPs in plants establish them as the largest family of CRPs. Together with their diverse functions, they showed promising value to be further develop for biotechnological applications as peptide biologics or robust anti-fungal agents.

Publications and Presentations

1. Huang, J., Wong, K. H., **Tay, S. V.**, Serra, A., Sze, S. K., & Tam, J. P. (2019). Astratides: Insulin-modulating, insecticidal, and antifungal cysteine-rich peptides from *Astragalus membranaceus*. Journal of natural products, 82(2), 194-204.

2. Huang, J., Wong, K. H., **Tay, S. V.**, How, A., & Tam, J. P. (2019). Cysteine-rich peptide fingerprinting as a general method for herbal analysis to differentiate *Radix Astragali* and *Radix Hedysarum*. Frontiers in plant science, 10, 973.

3. Huang,J., Wong, K. H., Tan, W. L., **Tay, S. V.,** & Tam, J. P (2020) Identification and characterization of a wolfberry carboxypeptidase inhibitor from *Lycium barbarum*

(submitted)

4. **Tay, S. V.,** Huang, J., Wong, K. H., & Tam, J.P. Chenotides: Anti-fungal Tandemly repeating 6C-chitin-binding Hevein-like Peptides from *Chenopodium quinoa* (manuscript in preparation)

5. **Tay, S. V.**, Wong, K. H., Huang, J., & Tam, J.P. Avenatides: Anti-fungal 8C-chitin-binding Hevein-like Peptides from a Common Cereal *Avena sativa* (manuscript in preparation)

6. **Tay, S. V.**, Huang, J., Wong, K. H., & Tam, J.P. Eleutide: Discovery and Characterization of 8C-non-chitin-binding Hevein-like Peptide from *Eleutherococcus trifoliatus* (manuscript in preparation)

References

1. Hasler, C.M., A.S. Bloch, C. Thomson, E. Enrione, and C. Manning, *Position of the American Dietetic Association: functional foods.* Journal of the American Dietetic Association, 2004. **104**(5): p. 814-826.

2. Size, E.O.M., *Share & Trends Analysis Report By Application (Cleaning & Home, Medical, Food & Beverages, Spa & Relaxation), By Product, By Sales Channel, And Segment Forecasts, 2019-2025.* Report ID, 2019: p. 978-1.

3. Jacobson, M., *Plants, insects, and man—their interrelationships.* Economic Botany, 1982. **36**(3): p. 346-354.

4. Champ, B.R. and C.E. Dyte, *Report of the FAO global survey of pesticide susceptibility of stored grain pests*. 1976: FAO.

5. Champ, B., *Occurrence of resistance to pesticides in grain storage pests. Pesticides and humid tropical grain storage systems* (Ed. BR Champ, E Highly), Australian Centre for International Agricultural Research Proceedings, 1985. **14**: p. 229-255.

6. Jacobson, M., *Botanical pesticides: past, present, and future*. 1989, American Chemistry Socitety Publications.

7. Singh, S., *Natural plant products-As protectant during grain storage: A review.* Journal of Entomology and Zoology Studies, 2017. **5**(3): p. 1873-1885.

8. Rembold, H., *Azadirachtins: their structure and mode of action*. 1989, American Chemistry Society Publications.

9. Isman, M.B., *Botanical insecticides, deterrents, and repellents in modern agriculture and an increasingly regulated world.* Annual Review Entomology, 2006. **51**: p. 45-66.

10. Casida, J.E., *Pyrethrum flowers and pyrethroid insecticides.* Environmental Health Perspectives, 1980. **34**: p. 189-202.

11. Yamomoto, I., *Nicotine to nicotinoids*. 1999, Springer, Tokyo, Japan. p. 3-27.

12. Ujváry, I., *Nicotine and other insecticidal alkaloids*, in *Nicotinoid insecticides and the nicotinic acetylcholine receptor*. 1999, Springer. p. 29-69.

13. Davidson, W., *Rotenone as a contact insecticide.* Journal of Economic Entomology, 1930. **23**(5): p. 868-874.

14. Hayes, W.J. and E.R. Laws, *Handbook of pesticide toxicology.* 1991.

15. Heinz, S., A. Freyberger, B. Lawrenz, L. Schladt, G. Schmuck, and H. Ellinger-Ziegelbauer, *Mechanistic investigations of the mitochondrial complex I inhibitor rotenone in the context of pharmacological and safety evaluation.* Scientific Reports, 2017. **7**: p. 45465.

16. Dayan, F.E., C.L. Cantrell, and S.O. Duke, *Natural products in crop protection.* Bioorganic & Medicinal Chemistry, 2009. **17**(12): p. 4022-4034.

17. Copping, L.G. and S.O. Duke, *Natural products that have been used commercially as crop protection agents.* Pest Management Science: formerly Pesticide Science, 2007. **63**(6): p. 524-554.

18. Allison, J.R., *Sabadilla seed insecticide*. 1963, Google Patents.

19. Chandler, D., A.S. Bailey, G.M. Tatchell, G. Davidson, J. Greaves, and W.P. Grant, *The development, regulation and use of biopesticides for integrated pest management.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2011. **366**(1573): p. 1987-1998.

20. Lipke, H., G. Fraenkel, and I.E. Liener, *Growth inhibitors. Effect of soybean inhibitors on growth of Tribolium confusum.* Journal of Agricultural and Food Chemistry, 1954. **2**: p. 410-414.

21. Brunelle, F., C. Girard, C. Cloutier, and D. Michaud, *A hybrid, broad‐spectrum inhibitor of Colorado potato beetle aspartate and cysteine digestive proteinases.* Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America, 2005. **60**(1): p. 20-31.

22. Jennings, C., J. West, C. Waine, D. Craik, and M. Anderson, *Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis.* Proceedings of the National Academy of Sciences, 2001. **98**(19): p. 10614-10619.

23. Chen, K.-C., C.-Y. Lin, C.-C. Kuan, H.-Y. Sung, and C.-S. Chen, *A novel defensin encoded by a mungbean cDNA exhibits insecticidal activity against bruchid.* Journal of Agricultural and Food Chemistry, 2002. **50**(25): p. 7258-7263.

24. Franco, O.L., D.J. Rigden, F.R. Melo, and M.F. Grossi‐de‐Sá, *Plant α‐amylase inhibitors and their interaction with insect α‐amylases.* The

Federartion of European Biochemical Societies Journal, 2002. **269**(2): p. 397-412.

25. Lawrence, P.K. and K.R. Koundal, *Plant protease inhibitors in control of phytophagous insects.* Electronic Journal of Biotechnology, 2002. **5**(1): p. 5-6.

26. Whetstone, P.A. and B.D. Hammock, *Delivery methods for peptide and protein toxins in insect control.* Toxicon, 2007. **49**(4): p. 576-596.

27. Li, F., Y. Wang, D. Li, Y. Chen, and Q.P. Dou, *Are we seeing a resurgence in the use of natural products for new drug discovery?* 2019, Taylor & Francis.

28. Atanasov, A.G., B. Waltenberger, E.-M. Pferschy-Wenzig, et al., *Discovery and resupply of pharmacologically active plant-derived natural products: A review.* Biotechnology Advances, 2015. **33**(8): p. 1582-1614.

29. Craik, D.J., D.P. Fairlie, S. Liras, and D. Price, *The future of peptide‐based drugs.* Chemical Biology & Drug Design, 2013. **81**(1): p. 136-147.

30. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the 30 years from 1981 to 2010.* Journal of Natural Products, 2012. **75**(3): p. 311-335.

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

32. Padhi, A., M. Sengupta, S. Sengupta, K.H. Roehm, and A. Sonawane, *Antimicrobial peptides and proteins in mycobacterial therapy: current status and future prospects.* Tuberculosis, 2014. **94**(4): p. 363-373.

33. Giordano, C., M. Marchiò, E. Timofeeva, and G. Biagini, *Neuroactive peptides as putative mediators of antiepileptic ketogenic diets.* Frontiers in Neurology, 2014. **5**: p. 63.

34. Tam, J.P., S. Wang, K.H. Wong, and W.L. Tan, *Antimicrobial Peptides from Plants.* Pharmaceuticals (Basel), 2015. **8**(4): p. 711-57.

35. Sels, J., J. Mathys, B.M. De Coninck, B.P. Cammue, and M.F. De Bolle, *Plant pathogenesis-related (PR) proteins: a focus on PR peptides.* Plant Physiology and Biochemistry, 2008. **46**(11): p. 941-950.

36. Selitrennikoff, C.P., *Antifungal proteins.* Applied Environmental Microbiology, 2001. **67**(7): p. 2883-2894.

37. Van Loon, L.C., M. Rep, and C.M. Pieterse, *Significance of inducible defense-related proteins in infected plants.* Annual Review Phytopathology, 2006. **44**: p. 135-162.

38. Rodrigues, T., D. Reker, P. Schneider, and G. Schneider, *Counting on natural products for drug design.* Nature Chemistry, 2016. **8**(6): p. 531.

39. Wick, J., *Aspirin: a history, a love story.* The Consultant Pharmacist®, 2012. **27**(5): p. 322-329.

40. Ball, P., *Quinine steps back in time: Chemists have long memories. The claim, dating back to 1918, that a crucial step in a synthesis of quinine had been carried out has been validated experimentally, closing a chapter in this fascinating story.* Nature, 2008. **451**(7182): p. 1065-1067.

41. Norn, S., P. Kruse, and E. Kruse, *History of opium poppy and morphine.* Dansk Medicinhistorisk Arbog, 2005. **33**: p. 171-184.

42. Baumann, A., *Early development of therapeutic biologics-pharmacokinetics.* Current Drug Metabolism, 2006. **7**(1): p. 15-21.

43. Rao, A.G., *Antimicrobial peptides.* Moleculart Plant-Microbe Interaction, 1995. **8**(6): p. 13.

44. Montalbán-López, M., M. Sánchez-Hidalgo, R. Cebrián, and M. Maqueda, *Discovering the bacterial circular proteins: bacteriocins, cyanobactins, and pilins.* Journal of Biological Chemistry, 2012. **287**(32): p. 27007-27013.

45. Silverstein, K.A., W.A. Moskal Jr, H.C. Wu, B.A. Underwood, M.A. Graham, C.D. Town, and K.A. VandenBosch, *Small cysteine‐rich peptides resembling antimicrobial peptides have been under‐predicted in plants.* The Plant Journal, 2007. **51**(2): p. 262-280.

46. Lay, F. and M. Anderson, *Defensins-components of the innate immune system in plants.* Current Protein and Peptide Science, 2005. **6**(1): p. 85-101.

47. Lacerda, A., É.A.R. Vasconcelos, P.B. Pelegrini, and M.F. Grossi-de-Sa, *Antifungal defensins and their role in plant defense.* Frontiers in Microbiology, 2014. **5**: p. 116.

48. Guzmán-Rodríguez, J.J., A. Ochoa-Zarzosa, R. López-Gómez, and J.E. López-Meza, *Plant antimicrobial peptides as potential anticancer agents.* BioMed Research International, 2015. **2015**.

49. Gould, A. and J.A. Camarero, *Cyclotides: Overview and biotechnological applications.* ChemBioChem, 2017. **18**(14): p. 1350-1363.

50. Weidmann, J. and D.J. Craik, *Discovery, structure, function, and applications of cyclotides: circular proteins from plants.* Journal of Experimental Botany, 2016. **67**(16): p. 4801-4812.

51. Kini, S.G., P.Q. Nguyen, S. Weissbach, A. Mallagaray, J. Shin, H.S. Yoon, and J.P. Tam, *Studies on the Chitin Binding Property of Novel Cysteine-Rich Peptides from Alternanthera sessilis.* Biochemistry, 2015. **54**(43): p. 6639-49.

52. Kini, S.G., K.H. Wong, W.L. Tan, T. Xiao, and J.P. Tam, *Morintides: cargo-free chitin-binding peptides from Moringa oleifera.* BioMedCentral Plant Biology, 2017. **17**(1): p. 68.

53. Wong, K.H., W.L. Tan, S.G. Kini, T. Xiao, A. Serra, S.K. Sze, and J.P. Tam, *Vaccatides: Antifungal Glutamine-Rich Hevein-Like Peptides from Vaccaria hispanica.* Frontiers in Plant Science, 2017. **8**.

54. Wong, K.H., W.L. Tan, A. Serra, T. Xiao, S.K. Sze, D. Yang, and J.P. Tam, *Ginkgotides: Proline-Rich Hevein-Like Peptides from Gymnosperm Ginkgo biloba.* Frontiers in Plant Science, 2016. **7**: p. 1639.

55. Loo, S., A. Kam, T. Xiao, G.K. Nguyen, C.F. Liu, and J.P. Tam, *Identification and characterization of roseltide, a knottin-type neutrophil elastase inhibitor derived from Hibiscus sabdariffa.* Scientific Reports, 2016. **6**(1): p. 1-16.

56. Loo, S., A. Kam, T. Xiao, and J.P. Tam, *Bleogens: Cactus-Derived Anti-Candida Cysteine-Rich Peptides with Three Different Precursor Arrangements.* Frontiers in Plant Science, 2017. **8**.

57. Tam, J.P., G.K. Nguyen, S. Loo, S. Wang, D. Yang, and A. Kam, *Ginsentides: cysteine and glycine-rich peptides from the ginseng family with unusual disulfide connectivity.* Scientific Reports, 2018. **8**(1): p. 1-15.

58. Betz, S.F., *Disulfide bonds and the stability of globular proteins.* Protein Science, 1993. **2**(10): p. 1551-1558.

59. Nguyen, P.Q., T.T. Luu, Y. Bai, G.K. Nguyen, K. Pervushin, and J.P. Tam, *Allotides: Proline-Rich Cystine Knot alpha-Amylase Inhibitors from Allamanda cathartica.* Journal of Natural Product, 2015. **78**(4): p. 695-704.

60. Lin, P., Y. Li, K. Dong, and Q. Li, *The Antibacterial Effects of an Antimicrobial Peptide Human beta-Defensin 3 Fused with Carbohydrate-*

Binding Domain on Pseudomonas aeruginosa PA14. Current Microbiology, 2015. **71**(2): p. 170-6.

61. Ponz, F., J. Paz-Ares, C. Hernández-Lucas, P. Carbonero, and F. García-Olmedo, *Synthesis and processing of thionin precursors in developing endosperm from barley (Hordeum vulgare L.).* The European Molecular Biology Organization Journal, 1983. **2**(7): p. 1035.

62. Duvick, J., T. Rood, A.G. Rao, and D.R. Marshak, *Purification and characterization of a novel antimicrobial peptide from maize (Zea mays L.) kernels.* Journal of Biological Chemistry, 1992. **267**(26): p. 18814-18820.

63. Tan, W.L., K.H. Wong, and J.P. Tam, *Lybatide: Naturally-Occurring Disulfide-Stapled Helical Peptides from Lycium barbarum.* The Federation of American Societies for Experimental Biology Journal, 2017. **31**(1 Supplement): p. lb115-lb115.

64. Kumari, G., A. Serra, J. Shin, P.Q. Nguyen, S.K. Sze, H.S. Yoon, and J.P. Tam, *Cysteine-rich peptide family with unusual disulfide connectivity from Jasminum sambac.* Journal of Natural Products, 2015. **78**(11): p. 2791-2799.

65. Shen, Y., L. Xu, J. Huang, A. Serra, H. Yang, and J.P. Tam, *Potentides: New Cysteine‐Rich Peptides with Unusual Disulfide Connectivity from Potentilla anserina.* ChemBioChem, 2019. **20**(15): p. 1995-2004.

66. Wong, K.H., W.L. Tan, T. Xiao, and J.P. Tam, *β-Ginkgotides: Hyperdisulfide-constrained peptides from Ginkgo biloba.* Scientific Reports, 2017. **7**.

67. Stec, B., *Plant thionins–the structural perspective.* Cellular and Molecular Life Sciences, 2006. **63**(12): p. 1370-1385.

68. Shafee, T.M., F.T. Lay, M.D. Hulett, and M.A. Anderson, *The defensins consist of two independent, convergent protein superfamilies.* Molecular Biology and Evolution, 2016. **33**(9): p. 2345-2356.

69. Park, J.D. and D.H. Kim, *Cysteine derivatives as inhibitors for carboxypeptidase A: synthesis and structure− activity relationships.* Journal of Medicinal Chemistry, 2002. **45**(4): p. 911-918.

70. Gunasekera, S., N.L. Daly, M.A. Anderson, and D.J. Craik, *Chemical synthesis and biosynthesis of the cyclotide family of circular proteins.* International Union of Biochemistry and Molecular Biology Life, 2006. **58**(9): p. 515-24.

71. Craik, D.J., N.L. Daly, T. Bond, and C. Waine, *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): p. 1327-36.

72. Qin, Q., E.J. McCallum, Q. Kaas, J. Suda, I. Saska, D.J. Craik, and J.S. Mylne, *Identification of candidates for cyclotide biosynthesis and cyclisation by expressed sequence tag analysis of Oldenlandia affinis.* BioMed Central Genomics, 2010. **11**: p. 111.

73. Kaas, Q., J.-C. Westermann, and D.J. Craik, *Conopeptide characterization and classifications: an analysis using ConoServer.* Toxicon, 2010. **55**(8): p. 1491-1509.

74. Bourgeat-Lami, E. and M. Lansalot, *Organic/inorganic composite latexes: the marriage of emulsion polymerization and inorganic chemistry*, in *Hybrid latex particles*. 2010, Springer. p. 53-123.

75. Cook, A. and B. Sekhar, *Fractions from Hevea brasiliensis latex centrifuged at 59,000 g.* Rubber Chemistry and Technology, 1954. **27**(1): p. 297-301.

76. Marin, B. and C. Lioret, *A plant vacuolar system: the lutoids from Hevea brasiliensis latex.* Plant Physiology. Vegetable, 1982. **20**(2): p. 311-331.

77. Southorn, W., *Complex particles in Hevea latex.* Nature, 1960. **188**(4745): p. 165-166.

78. Wang, X., M. Shi, D. Wang, et al., *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): p. 5146-5159.

79. Archer, B. and E. Cockbain, *The proteins of Hevea brasiliensis latex. 2. Isolation of the α-globulin of fresh latex serum.* Biochemical Journal, 1955. **61**(3): p. 508.

80. Archer, B. and B. Sekhar, *The proteins of Hevea brasiliensis latex. 1. Protein constituents of fresh latex serum.* Biochemical Journal, 1955. **61**(3): p. 503.

81. Archer, B.L., *The proteins of Hevea brasiliensis Latex. 4. Isolation and characterization of crystalline hevein.* Biochemical Journal, 1960. **75**: p. 236-40.

82. Walujono, K., *Amino acid sequence of hevein*. 1975.

83. Xu, Y., Q. Zhu, W. Panbangred, K. Shirasu, and C. Lamb, *Regulation, expression and function of a new basic chitinase gene in rice (Oryza sativa L.).* Plant Molecular Biology, 1996. **30**(3): p. 387-401.

84. Beintema, J.J. and W.J. Peumans, *The primary structure of stinging nettle (Urtica dioica) agglutinin. A two-domain member of the hevein family.* Federation of European Biochemical Societies Letters, 1992. **299**(2): p. 131-4.

85. Odintsova, T.I., A.A. Vassilevski, A.A. Slavokhotova, et al., *A novel antifungal hevein-type peptide from Triticum kiharae seeds with a unique 10-cysteine motif.* Federation of European Biochemical Societies Journal, 2009. **276**(15): p. 4266-75.

86. Peumans, W.J., M. De Ley, and W.F. Broekaert, *An unusual lectin from stinging nettle (Urtica dioica) rhizomes.* Federation of European Biochemical Societies Letters, 1984. **177**(1): p. 99-103.

87. Van Parijs, J., W.F. Broekaert, I.J. Goldstein, and W.J. Peumans, *Hevein: an antifungal protein from rubber-tree (Hevea brasiliensis) latex.* Planta, 1991. **183**(2): p. 258-64.

88. Bohlmann, H., S. Clausen, S. Behnke, et al., *Leaf-specific thionins of barley-a novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defence mechanism of plants.* European Molecular Biology Organization Journal, 1988. **7**(6): p. 1559-65.

89. Broekaert, I., H.I. Lee, A. Kush, N.H. Chua, and N. Raikhel, *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): p. 7633-7.

90. Rodriguez-Romero, A., K.G. Ravichandran, and M. Soriano-Garcia, *Crystal structure of hevein at 2.8 A resolution.* Federation of European Biochemical Societies Letters, 1991. **291**(2): p. 307-9.

91. Reyes-Lopez, C.A., A. Hernandez-Santoyo, M. Pedraza-Escalona, G. Mendoza, A. Hernandez-Arana, and A. Rodriguez-Romero, *Insights into a conformational epitope of Hev b 6.02 (hevein).* Biochemistry Biophysical Research Communications, 2004. **314**(1): p. 123-30.

92. Andersen, N.H., B. Cao, A. Rodriguez-Romero, and B. Arreguin, *Hevein: NMR assignment and assessment of solution-state folding for the agglutinin-toxin motif.* Biochemistry, 1993. **32**(6): p. 1407-22.

93. Asensio, J.L., F.J. Canada, M. Bruix, A. Rodriguez-Romero, and J. Jimenez-Barbero, *The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose.* European Journal of Biochemistry, 1995. **230**(2): p. 621-33.

94. Broekaert, W.F., W. Marien, F.R. Terras, et al., *Antimicrobial peptides from Amaranthus caudatus seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins.* Biochemistry, 1992. **31**(17): p. 4308-14.

95. Nielsen, K.K., J.E. Nielsen, S.M. Madrid, and J.D. Mikkelsen, *Characterization of a new antifungal chitin-binding peptide from sugar beet leaves.* Plant Physiolology, 1997. **113**(1): p. 83-91.

96. Koo, J.C., S.Y. Lee, H.J. Chun, et al., *Two hevein homologs isolated from the seed of Pharbitis nil L. exhibit potent antifungal activity.* Biochimica et Biophysica Acta, 1998. **1382**(1): p. 80-90.

97. Huang, R.H., Y. Xiang, X.Z. Liu, Y. Zhang, Z. Hu, and D.C. Wang, *Two novel antifungal peptides distinct with a five-disulfide motif from the bark of Eucommia ulmoides Oliv.* Federation of European Biochemical Societies Letters, 2002. **521**(1-3): p. 87-90.

98. Lipkin, A., V. Anisimova, A. Nikonorova, et al., *An antimicrobial peptide Ar-AMP from amaranth (Amaranthus retroflexus L.) seeds.* Phytochemistry, 2005. **66**(20): p. 2426-31.

99. R, R.S., D.V. V, A.K. Nikonorova, et al., *Transformation of tobacco and Arabidopsis plants with Stellaria media genes encoding novel hevein-like peptides increases their resistance to fungal pathogens.* Transgenic Research, 2012. **21**(2): p. 313-25.

100. Fujimura, M., Y. Minami, K. Watanabe, and K. Tadera, *Purification, characterization, and sequencing of a novel type of antimicrobial peptides, Fa-AMP1 and Fa-AMP2, from seeds of buckwheat (Fagopyrum esculentum Moench.).* Biosciences, Biotechnology and Biochemistry, 2003. **67**(8): p. 1636-42.

101. Van den Bergh, K.P., E.J. Van Damme, W.J. Peumans, and J. Coosemans, *Ee-CBP, a hevein-type antimicrobial peptide from bark of the spindle tree (Euonymus europaeus L.).* Meded Rijksuniv Gent Fak Landbouwkd Toegep Biol Wet, 2002. **67**(2): p. 327-31.

102. Svensson, B., I. Svendsen, P. Hoejrup, P. Roepstorff, S. Ludvigsen, and F.M. Poulsen, *Primary structure of barwin: a barley seed protein closely*

related to the C-terminal domain of proteins encoded by wound-induced plant genes. Biochemistry, 1992. **31**(37): p. 8767-8770.

103. Peumans, W.J. and E. Van Damme, *Lectins as plant defense proteins.* Plant Physiology, 1995. **109**(2): p. 347.

104. Van Damme, E.J., *History of plant lectin research.* Methods in Molecular Biology, 2014. **1200**: p. 3-13.

105. Zhao, K.J. and M.L. Chye, *Methyl jasmonate induces expression of a novel Brassica juncea chitinase with two chitin-binding domains.* Plant Molecular Biology, 1999. **40**(6): p. 1009-18.

106. Guan, Y. and M.L. Chye, *A Brassica juncea chitinase with two-chitin binding domains show anti-microbial properties against phytopathogens and Gram-negative bacteria.* Plant Signal Behaviour, 2008. **3**(12): p. 1103-5.

107. Olsnes, S., E. Saltvedt, and A. Pihl, *Isolation and comparison of galactose-binding lectins from Abrus precatorius and Ricinus communis.* Journal of Biological Chemistry, 1974. **249**(3): p. 803-810.

108. Beintema, J.J., *Structural features of plant chitinases and chitin-binding proteins.* Federation of European Biochemical Societies Letters, 1994. **350**(2-3): p. 159-63.

109. Li, S.S. and P. Claeson, *Cys/Gly-rich proteins with a putative single chitin-binding domain from oat (Avena sativa) seeds.* Phytochemistry, 2003. **63**(3): p. 249-55.

110. Van den Bergh, K.P., P. Proost, J. Van Damme, J. Coosemans, E.J. Van Damme, and W.J. Peumans, *Five disulfide bridges stabilize a hevein-type antimicrobial peptide from the bark of spindle tree (Euonymus europaeus L.).* Federation of European Biochemical Societies Letters, 2002. **530**(1-3): p. 181-5.

111. Yang, Y.F., K.C. Cheng, P.H. Tsai, C.C. Liu, and T.R. Lee, *Alanine substitutions of noncysteine residues in the cysteine‐stabilized $\alpha\beta$ motif.* Protein Science, 2009. **18**(7): p. 1498-1506.

112. Asensio, J.L., F.J. Cañada, H.-C. Siebert, et al., *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**(7): p. 529-543.

113. Van Damme, E.J., W.F. Broekaert, and W.J. Peumans, *The Urtica dioica Agglutinin Is a Complex Mixture of Isolectins.* Plant Physiology, 1988. **86**(2): p. 598-601.

114. Wright, H.T., D.M. Brooks, and C.S. Wright, *Evolution of the multidomain protein wheat germ agglutinin.* Journal of Molecular Evolution, 1985. **21**(2): p. 133-138.

115. Onaga, S. and T. Taira, *A new type of plant chitinase containing LysM domains from a fern (Pteris ryukyuensis): roles of LysM domains in chitin binding and antifungal activity.* Glycobiology, 2008. **18**(5): p. 414-23.

116. Diaz-Perales, A., C. Collada, C. Blanco, R. Sanchez-Monge, T. Carrillo, C. Aragoncillo, and G. Salcedo, *Cross-reactions in the latex-fruit syndrome: A relevant role of chitinases but not of complex asparagine-linked glycans.* Journal of Allergy and Clinical Immunology, 1999. **104**(3): p. 681-687.

117. Wagner, S. and H. Breiteneder, *The latex-fruit syndrome*. 2002, Portland Press Limited.

118. Chen, Z., V. Kampen, M. Raulf‐Heimsoth, and X. Baur, *Allergenic and antigenic determinants of latex allergen Hev b 1: peptide mapping of epitopes recognized by human, murine and rabbit antibodies.* Clinical & Experimental Allergy, 1996. **26**(4): p. 406-415.

119. Alenius, H., N. Kalkkinen, T. Reunala, K. Turjanmaa, and T. Palosuo, *The main IgE-binding epitope of a major latex allergen, prohevein, is present in its N-terminal 43-amino acid fragment, hevein.* Journal of Immunology, 1996. **156**(4): p. 1618-25.

120. Karisola, P., J. Mikkola, N. Kalkkinen, 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): p. 2621-2628.

121. Mergaert, P., K. Nikovics, Z. Kelemen, N. Maunoury, D. Vaubert, A. Kondorosi, and E. Kondorosi, *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): p. 161-173.

122. Nguyen, P.Q., S. Wang, A. Kumar, L.J. Yap, T.T. Luu, J. Lescar, and J.P. Tam, *Discovery and characterization of pseudocyclic cystine-knot alpha-amylase inhibitors with high resistance to heat and proteolytic degradation.* Federation of European Biochemical Societies Journal, 2014. **281**(19): p. 4351-66.

123. Kimura, R.H., A.T. Tran, and J.A. Camarero, *Biosynthesis of the cyclotide Kalata B1 by using protein splicing.* Angewandte Chemie, 2006. **45**(6): p. 973-6.

124. Nguyen, G.K., S. Zhang, W. Wang, C.T. Wong, N.T. Nguyen, and J.P. Tam, *Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry.* Journal of Biology and Chemistry, 2011. **286**(52): p. 44833-44.

125. Nguyen, G.K., W.H. Lim, P.Q. Nguyen, and J.P. Tam, *Novel cyclotides and uncyclotides with highly shortened precursors from Chassalia chartacea and effects of methionine oxidation on bioactivities.* Journal of Biology and Chemistry, 2012. **287**(21): p. 17598-607.

126. De Bolle, M.F., K.M. David, S.B. Rees, J. Vanderleyden, B.P. Cammue, and W.F. Broekaert, *Cloning and characterization of a cDNA encoding an antimicrobial chitin-binding protein from amaranth, Amaranthus caudatus.* Plant Molecular Biology, 1993. **22**(6): p. 1187-90.

127. Shinshi, H., H. Wenzler, J.-M. Neuhaus, G. Felix, J. Hofsteenge, and F. Meins, *Evidence for N-and C-terminal processing of a plant defense-related enzyme: primary structure of tobacco prepro-β-1, 3-glucanase.* Proceedings of the National Academy of Sciences, 1988. **85**(15): p. 5541-5545.

128. Chrispeels, M.J. and N.V. Raikhel, *Short peptide domains target proteins to plant vacuoles.* Cell, 1992. **68**(4): p. 613-616.

129. De Bolle, M.F., R.W. Osborn, I.J. Goderis, et al., *Antimicrobial peptides from Mirabilis jalapa and Amaranthus caudatus: expression, processing, localization and biological activity in transgenic tobacco.* Plant Molecular Biology, 1996. **31**(5): p. 993-1008.

130. Van den Bergh, K.P., P. Rouge, P. Proost, et al., *Synergistic antifungal activity of two chitin-binding proteins from spindle tree (Euonymus europaeus L.).* Planta, 2004. **219**(2): p. 221-32.

131. Bontems, F., C. Roumestand, P. Boyot, B. Gilquin, Y. Doljansky, A. Menez, and F. Toma, *Three‐dimensional structure of natural charybdotoxin in aqueous solution by 1H‐NMR Charybdotoxin possesses a structural motif found in other scorpion toxins.* Federation of European Biochemical Societies Journal, 1991. **196**(1): p. 19-28.

132. Milbradt, A.G., F. Kerek, L. Moroder, and C. Renner, *Structural characterization of hellethionins from Helleborus purpurascens.* Biochemistry, 2003. **42**(8): p. 2404-2411.

133. Pallaghy, P.K., R.S. Norton, K.J. Nielsen, and D.J. Craik, *A common structural motif incorporating a cystine knot and a triple ‐stranded β ‐sheet in toxic and inhibitory polypeptides.* Protein Science, 1994. **3**(10): p. 1833-1839.

134. Zhu, S., H. Darbon, K. Dyason, F. Verdonck, and J. Tytgat, *Evolutionary origin of inhibitor cystine knot peptides.* The Federation of American Societies for Experimental Journal, 2003. **17**(12): p. 1765-1767.

135. Martins, J.C., D. Maes, R. Loris, H.A. Pepermans, L. Wyns, R. Willem, and P. Verheyden, *1 H NMR Study of the Solution Structure of Ac-AMP2, a Sugar Binding Antimicrobial Protein Isolated fromAmaranthus caudatus.* Journal of Molecular Biology, 1996. **258**(2): p. 322-333.

136. Huang, R.-H., Y. Xiang, G.-Z. Tu, Y. Zhang, and D.-C. Wang, *Solution structure of Eucommia antifungal peptide: a novel structural model distinct with a five-disulfide motif.* Biochemistry, 2004. **43**(20): p. 6005-6012.

137. Xiang, Y., R.-H. Huang, X.-Z. Liu, Y. Zhang, and D.-C. Wang, *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): p. 86-97.

138. Dubovskii, P.V., A.A. Vassilevski, A.A. Slavokhotova, T.I. Odintsova, E.V. Grishin, T.A. Egorov, and A.S. Arseniev, *Solution structure of a defense peptide from wheat with a 10-cysteine motif.* Biochemical and Biophysical Research Communications, 2011. **411**(1): p. 14-18.

139. Muraki, M., H. Morii, and K. Harata, *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): p. 385-9.

140. Diaz-Perales, A., C. Collada, C. Blanco, R. Sanchez-Monge, T. Carrillo, C. Aragoncillo, and G. Salcedo, *Class I chitinases with hevein-like domain, but not class II enzymes, are relevant chestnut and avocado allergens.* Journal of Allergy and Clinical Immunology, 1998. **102**(1): p. 127-33.

141. Asensio, J.L., F.J. Canada, M. Bruix, C. Gonzalez, N. Khiar, A. Rodriguez-Romero, and J. Jimenez-Barbero, *NMR investigations of protein-carbohydrate interactions: refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside.* Glycobiology, 1998. **8**(6): p. 569-77.

142. Soedjanaatmadja, U.M., J. Hofsteenge, C.M. Jeronimus-Stratingh, A.P. Bruins, and J.J. Beintema, *Demonstration by mass spectrometry that*

pseudo-hevein and hevein have ragged C-terminal sequences. Biochimica et Biophysica Acta, 1994. **1209**(1): p. 144-8.

143. Asensio, J.L., H.C. Siebert, C.W. von Der Lieth, et al., *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): p. 218-36.

144. Aboitiz, N., M. Vila-Perello, P. Groves, J.L. Asensio, D. Andreu, F.J. Canada, and J. Jimenez-Barbero, *NMR and modeling studies of protein-carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides.* ChemBioChem, 2004. **5**(9): p. 1245-55.

145. Chavez, M.I., M. Vila-Perello, F.J. Canada, D. Andreu, and J. Jimenez-Barbero, *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): p. 1461-8.

146. Verheyden, P., J. Pletinckx, D. Maes, H.A. Pepermans, L. Wyns, R. Willem, and J. Martins, *1H NMR study of the interaction of N, N′, N″-triacetyl chitotriose with Ac-AMP2, a sugar binding antimicrobial protein isolated from Amaranthus caudatus.* Federation of European Biochemical Societies Letters, 1995. **370**(3): p. 245-249.

147. Espinosa, J.F., J.L. Asensio, J.L. Garcia, et al., *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**(13): p. 3965-78.

148. Williamson, B., B. Tudzynski, P. Tudzynski, and J.A. van Kan, *Botrytis cinerea: the cause of grey mould disease.* Molecular Plant Pathology, 2007. **8**(5): p. 561-580.

149. Slavokhotova, A.A., T.A. Naumann, N.P. Price, E.A. Rogozhin, Y.A. Andreev, A.A. Vassilevski, and T.I. Odintsova, *Novel mode of action of plant defense peptides - hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases.* Federation of European Biochemical Societies Journal, 2014. **281**(20): p. 4754-64.

150. Malanovic, N. and K. Lohner, *Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides.* Biochimica et Biophysica Acta-Biomembranes, 2016. **1858**(5): p. 936-946.

151. Ali, B.H., N.A. Wabel, and G. Blunden, *Phytochemical, pharmacological and toxicological aspects of Hibiscus sabdariffa L.: a review.* Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives, 2005. **19**(5): p. 369-375.

152. Ali, K., A. Ashraf, and N.N. Biswas, *Analgesic, anti-inflammatory and anti-diarrheal activities of ethanolic leaf extract of Typhonium trilobatum L. Schott.* Asian Pacific Journal of Tropical Biomedicine, 2012. **2**(9): p. 722-726.

153. Herrera-Arellano, A., J. Miranda-Sánchez, P. Ávila-Castro, et al., *Clinical effects produced by a standardized herbal medicinal product of Hibiscus sabdariffa on patients with hypertension. A randomized, double-blind, lisinopril-controlled clinical trial.* Planta Medica, 2007. **73**(01): p. 6-12.

154. Ajay, M., H. Chai, A. Mustafa, A.H. Gilani, and M.R. Mustafa, *Mechanisms of the anti-hypertensive effect of Hibiscus sabdariffa L. calyces.* Journal of Ethnopharmacology, 2007. **109**(3): p. 388-393.

155. Chen, C.-C., J.-D. Hsu, S.-F. Wang, et al., *Hibiscus sabdariffa extract inhibits the development of atherosclerosis in cholesterol-fed rabbits.* Journal of Agricultural and Food Chemistry, 2003. **51**(18): p. 5472-5477.

156. Ajiboye, T.O., H.O. Raji, A.O. Adeleye, N.S. Adigun, O.B. Giwa, O.B. Ojewuyi, and A.T. Oladiji, *Hibiscus sabdariffa calyx palliates insulin resistance, hyperglycemia, dyslipidemia and oxidative rout in fructose‐induced metabolic syndrome rats.* Journal of the Science of Food and Agriculture, 2016. **96**(5): p. 1522-1531.

157. Jiménez-Ferrer, E., J. Alarcón-Alonso, A. Aguilar-Rojas, A. Zamilpa, J. Tortoriello, and M. Herrera-Ruiz, *Diuretic effect of compounds from Hibiscus sabdariffa by modulation of the aldosterone activity.* Planta Medica, 2012. **78**(18): p. 1893-1898.

158. Alarcón-Alonso, J., A. Zamilpa, F.A. Aguilar, M. Herrera-Ruiz, J. Tortoriello, and E. Jimenez-Ferrer, *Pharmacological characterization of the diuretic effect of Hibiscus sabdariffa Linn (Malvaceae) extract.* Journal of Ethnopharmacology, 2012. **139**(3): p. 751-756.

159. Wilson, F. and M.Y. Menzel, *Kenaf (Hibiscus cannabinus), roselle (Hibiscus sabdariffa).* Economic Botany, 1964. **18**(1): p. 80-91.

160. Hara, S., J. Makino, and T. Ikenaka, *Amino acid sequences and disulfide bridges of serine proteinase inhibitors from bitter gourd (Momordica*

charantia LINN.) seeds. The Journal of Biochemistry, 1989. **105**(1): p. 88-92.

161. Doring, G., *The role of neutrophil elastase in chronic inflammation.* American Journal of Respiratory and Critical Care Medicine, 1994. **150**(6): p. S114.

162. Kam, A., S. Loo, B. Dutta, S.K. Sze, and J.P. Tam, *Plant-derived mitochondria-targeting cysteine-rich peptide modulates cellular bioenergetics.* Journal of Biological Chemistry, 2019. **294**(11): p. 4000-4011.

163. Kam, A., S. Loo, J.-S. Fan, S.K. Sze, D. Yang, and J.P. Tam, *Roseltide rT7 is a disulfide-rich, anionic, and cell-penetrating peptide that inhibits proteasomal degradation.* Journal of Biological Chemistry, 2019. **294**(51): p. 19604-19615.

164. Gruber, C.W., M. Čemažar, R.J. Clark, T. Horibe, R.F. Renda, M.A. Anderson, and D.J. Craik, *A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins.* Journal of Biological Chemistry, 2007. **282**(28): p. 20435-20446.

165. Tang, B., S. Nirasawa, M. Kitaoka, C. Marie-Claire, and K. Hayashi, *General function of N-terminal propeptide on assisting protein folding and inhibiting catalytic activity based on observations with a chimeric thermolysin-like protease.* Biochemical and Biophysical Research Communications, 2003. **301**(4): p. 1093-1098.

166. Andersson, E., L. Hellman, U. Gullberg, and I. Olsson, *The role of the propeptide for processing and sorting of human myeloperoxidase.* Journal of Biological Chemistry, 1998. **273**(8): p. 4747-4753.

167. Sharma, M. and G.K. Pandey, *Expansion and function of repeat domain proteins during stress and development in plants.* Frontiers in Plant Science, 2016. **6**: p. 1218.

168. Snir, A., D. Nadel, I. Groman-Yaroslavski, Y. Melamed, M. Sternberg, O. Bar-Yosef, and E. Weiss, *The origin of cultivation and proto-weeds, long before Neolithic farming.* Public Library of Sciences One, 2015. **10**(7).

169. Co-operation, O.f.E. and Development, *OECD-FAO agricultural outlook 2018-2027*. 2018: Organisation for Economic Co-operation and Development Publishing.

170. Leslie, J.F., S. Muthukrishnan, M. Swegle, A.J. Vigers, and C.P. Selitrennikoff, *Variability in antifungal proteins in the grains of maize, sorghum and wheat.* Physiologia Plantarum, 1993. **88**(2): p. 339-349.
171. Gomez, L., I. Allona, R. Casado, and C. Aragoncillo, *Seed chitinases.* Seed Science Research, 2002. **12**(4): p. 217-230.

172. Bekele, B., K. Makkouk, A. Yusuf, F. Alemayu, and A. Lencho, *Occurrence and distribution of barley yellow dwarf virus (BYDV) isolates in central Ethiopia.* International Journal of Pest Management, 2001. **47**(2): p. 115-119.

173. Ceccarelli, S., S. Grando, V. Shevstov, H. Vivar, A. Yayaoui, M. El-Bhoussini, and M. Baum, *The ICARDA strategy for global barley improvement.* Rachis, 1999. **18**(2): p. 3-12.

174. Bhatty, R., *Production of food malt from hull-less barley.* Cereal Chemistry, 1996. **73**(1): p. 75-80.

175. Bekele, B., F. Alemayehu, and B. Lakew, *Food barley in Ethiopia.* Food Barley: Importance, Uses, and Local Knowledge, 2005: p. 53-82.

176. Grando, S. and H.G. Macpherson, *Food barley: importance, uses and local knowledge.* International Center for Agricultural Research in the Dry Areas, Aleppo, Syria, 2005.

177. Matz, S.A., *Chemistry and technology of cereals as food and feed*. 1991: Springer Science & Business Media.

178. Perrocheau, L., H. Rogniaux, P. Boivin, and D. Marion, *Probing heat‐stable water‐soluble proteins from barley to malt and beer.* Proteomics, 2005. **5**(11): p. 2849-2858.

179. Idehen, E., Y. Tang, and S. Sang, *Bioactive phytochemicals in barley.* Journal of Food and Drug Analysis, 2017. **25**(1): p. 148-161.

180. Nishiyama, T., Y. Hagiwara, H. Hagiwara, and T. Shibamoto, *Inhibitory effect of 2″-O-glycosyl isovitexin and α-tocopherol on genotoxic glyoxal formation in a lipid peroxidation system.* Food and Chemical Toxicology, 1994. **32**(11): p. 1047-1051.

181. Shibamoto, T., Y. Hagiwara, H. Hagiwara, and T. Osawa, *Flavonoid with strong antioxidative activity isolated from young green barley leaves*. 1994, American Chemistry Societies Publications.

182. Behall, K.M., D.J. Scholfield, and J. Hallfrisch, *Lipids significantly reduced by diets containing barley in moderately hypercholesterolemic men.* Journal of the American College of Nutrition, 2004. **23**(1): p. 55-62.

183. Behall, K.M., D.J. Scholfield, and J. Hallfrisch, *Comparison of hormone and glucose responses of overweight women to barley and oats.* Journal of the American College of Nutrition, 2005. **24**(3): p. 182-188.

184. Behall, K.M., D.J. Scholfield, and J.G. Hallfrisch, *Barley β-glucan reduces plasma glucose and insulin responses compared with resistant starch in men.* Nutrition Research, 2006. **26**(12): p. 644-650.

185. Lahouar, L., S. El-Bok, and L. Achour, *Therapeutic potential of young green barley leaves in prevention and treatment of chronic diseases: an overview.* The American Journal of Chinese Medicine, 2015. **43**(07): p. 1311-1329.

186. Durham, J.J., J. Ogata, S. Nakajima, Y. Hagiwara, and T. Shibamoto, *Degradation of organophosphorus pesticides in aqueous extracts of young green barley leaves (Hordeum vulgare L).* Journal of the Science of Food and Agriculture, 1999. **79**(10): p. 1311-1314.

187. Kirubakaran, S.I. and N. Sakthivel, *Cloning and overexpression of antifungal barley chitinase gene in Escherichia coli.* Protein Expression and Purification, 2007. **52**(1): p. 159-166.

188. Kragh, K.M., S. Jacobsen, and J.D. Mikkelsen, *Induction, purification and characterization of barley leaf chitinase.* Plant Science, 1990. **71**(1): p. 55-68.

189. Jacobsen, S., J. Mikkelsen, and J. Hejgaard, *Characterization of two antifungal endochitinases from barley grain.* Physiologia Plantarum, 1990. **79**(3): p. 554-562.

190. Roberts, W.K. and C.P. Selitrennikoff, *Isolation and partial characterization of two antifungal proteins from barley.* Biochimica et Biophysica Acta-General Subjects, 1986. **880**(2-3): p. 161-170.

191. Swegle, M., J.-K. Huang, G. Lee, and S. Muthukrishnan, *Identification of an endochitinase cDNA clone from barley aleurone cells.* Plant Molecular Biology, 1989. **12**(4): p. 403-412.

192. Pimentel, D. and T.W. Patzek, *Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower.* Natural Resources Research, 2005. **14**(1): p. 65-76.

193. Ranum, P., J.P. Peña‐Rosas, and M.N. Garcia‐Casal, *Global maize production, utilization, and consumption.* Annals of the New York Academy of Sciences, 2014. **1312**(1): p. 105-112.

194. Park, K., S. Park, B. Choe, S. Jong, and S. Lee, *Current status of quality improvement in maize.* Korean Journal of Crop Science (Korea R.), 1988.

195. Urias-Lugo, D., J. Heredia, M. Muy-Rangel, J. Valdez-Torres, S. Serna-Saldívar, and J. Gutiérrez-Uribe, *Anthocyanins and phenolic acids of hybrid and native blue maize (Zea mays L.) extracts and their antiproliferative activity in mammary (MCF7), liver (HepG2), colon (Caco2 and HT29) and prostate (PC3) cancer cells.* Plant Foods for Human Nutrition, 2015. **70**(2): p. 193-199.

196. Adom, K.K. and R.H. Liu, *Antioxidant activity of grains.* Journal of Agricultural and Food chemistry, 2002. **50**(21): p. 6182-6187.

197. Siyuan, S., L. Tong, and R. Liu, *Corn phytochemicals and their health benefits.* Food Science and Human Wellness, 2018. **7**(3): p. 185-195.

198. Díaz-Gómez, J.L., F. Castorena-Torres, R.E. Preciado-Ortiz, and S. García-Lara, *Anti-cancer activity of maize bioactive peptides.* Frontiers in Chemistry, 2017. **5**: p. 44.

199. Anderson, J.W., T.J. Hanna, X. Peng, and R.J. Kryscio, *Whole grain foods and heart disease risk.* Journal of the American College of Nutrition, 2000. **19**(sup3): p. 291S-299S.

200. Keenan, M.J., M. Janes, J. Robert, et al., *Resistant starch from high amylose maize (HAM‐RS2) reduces body fat and increases gut bacteria in ovariectomized (OVX) rats.* Obesity, 2013. **21**(5): p. 981-984.

201. Kendall, C.W., A. Esfahani, A.J. Hoffman, et al., *Effect of novel maize-based dietary fibers on postprandial glycemia and insulinemia.* Journal of the American College of Nutrition, 2008. **27**(6): p. 711-718.

202. Huynh, Q.K., C.M. Hironaka, E.B. Levine, C. Smith, J. Borgmeyer, and D. Shah, *Antifungal proteins from plants. Purification, molecular cloning, and antifungal properties of chitinases from maize seed.* Journal of Biological Chemistry, 1992. **267**(10): p. 6635-6640.

203. Lu, H., J. Zhang, K.-b. Liu, et al., *Earliest domestication of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago.* Proceedings of the National Academy of Sciences, 2009. **106**(18): p. 7367-7372.

204. McDonough, C.M., L.W. Rooney, and S.O. Serna-Saldivar, *The millets.* Food Science and Technology-New York-Marcel Dekker-, 2000: p. 177-202.

205. Sarita, E.S. and E. Singh, *Potential of millets: nutrients composition and health benefits.* Journal of Scientific and Innovative Research, 2016. **5**(2): p. 46-50.

206. Liu, R.H., *Whole grain phytochemicals and health.* Journal of Cereal Science, 2007. **46**(3): p. 207-219.

207. Chandrasekara, A. and F. Shahidi, *Bioaccessibility and antioxidant potential of millet grain phenolics as affected by simulated in vitro digestion and microbial fermentation.* Journal of Functional Foods, 2012. **4**(1): p. 226-237.

208. Coulibaly, A., B. Kouakou, and J. Chen, *Phytic acid in cereal grains: structure, healthy or harmful ways to reduce phytic acid in cereal grains and their effects on nutritional quality.* American Journal of Plant Nutrition and Fertilization Technology, 2011. **1**(1): p. 1-22.

209. Lee, S.H., I.-M. Chung, Y.-S. Cha, and Y. Park, *Millet consumption decreased serum concentration of triglyceride and C-reactive protein but not oxidative status in hyperlipidemic rats.* Nutrition Research, 2010. **30**(4): p. 290-296.

210. Chandrasekara, A. and F. Shahidi, *Antiproliferative potential and DNA scission inhibitory activity of phenolics from whole millet grains.* Journal of Functional Foods, 2011. **3**(3): p. 159-170.

211. Hegde, P.S., G. Chandrakasan, and T. Chandra, *Inhibition of collagen glycation and crosslinking in vitro by methanolic extracts of Finger millet (Eleusine coracana) and Kodo millet (Paspalum scrobiculatum).* The Journal of Nutritional Biochemistry, 2002. **13**(9): p. 517-521.

212. Radhajeyalakshmi, R., K. Yamunarani, K. Seetharaman, and R. Velazhahan, *Existence of thaumatin-like proteins (TLPs) in seeds of cereals.* Acta Phytopathologica et Entomologica Hungarica, 2003. **38**(3-4): p. 251-257.

213. Radhajeyalakshmi, R., B. Meena, R. Thangavelu, S. Deborah, P. Vidhyasekaran, and R. Velazhahan, *A 45-kDa chitinase purified from pearl millet (Pennisetum glaucum (L.) R. Br.) shows antifungal activity/Reinigung einer 45-kDa Chitinase mit antimykotischer Aktivität aus Perlhirse (Pennisetum glaucum (L.) R. Br.).* Zeitschrift für Pflanzenkrankheiten und

Pflanzenschutz/Journal of Plant Diseases and Protection, 2000: p. 605-616.

214. Bennetzen, J.L., J. Schmutz, H. Wang, et al., *Reference genome sequence of the model plant Setaria.* Nature biotechnology, 2012. **30**(6): p. 555.

215. Rasane, P., A. Jha, L. Sabikhi, A. Kumar, and V. Unnikrishnan, *Nutritional advantages of oats and opportunities for its processing as value added foods-a review.* Journal of Food Science and Technology, 2015. **52**(2): p. 662-675.

216. Ranhotra, G.S. and J.A. Gelroth, *Food uses of oats*, in *The Oat Crop*. 1995, Springer. p. 409-432.

217. Peterson, D.M. *Oat-a multifunctional grain*. in *Proceedings 7th International Oat Conference/Pirjo Peltonen-Sainio and Mari Topi-Hulmi (eds.)*. 2004. MTT.

218. Maizel, J., H. Burkhardt, and H. Mitchell, *Avenacin, an antimicrobial substance isolated from Avena sativa. I. Isolation and antimicrobial activity.* Biochemistry, 1964. **3**(3): p. 424-426.

219. Singh, R., S. De, and A. Belkheir, *Avena sativa (Oat), a potential neutraceutical and therapeutic agent: an overview.* Critical Reviews in Food Science and Nutrition, 2013. **53**(2): p. 126-144.

220. De Groot, A., R. Luyken, and N. Pikaar, *Cholesterol-lowering effect of rolled oats.* The Lancet, 1963. **282**(7302): p. 303-304.

221. Nie, L., M.L. Wise, D.M. Peterson, and M. Meydani, *Avenanthramide, a polyphenol from oats, inhibits vascular smooth muscle cell proliferation and enhances nitric oxide production.* Atherosclerosis, 2006. **186**(2): p. 260-266.

222. Aries, M., C. Vaissiere, B. Fabre, M. Charveron, and Y. Gall. *Immunomodulatory activity of Avena rhealba: interest in skin inflammatory disorders*. Journal of Investigative Dermatology, 1999.

223. Aries, M., C. Vaissiere, I. Ceruti, M. Charveron, and Y. Gall. *Avena rhealba activity in cutaneous wound healing process*. Journal of Investigative Dermatology, 1999.

224. Sørensen, H.P., L.S. Madsen, J. Petersen, J.T. Andersen, A.M. Hansen, and H.C. Beck, *Oat (Avena sativa) seed extract as an antifungal food preservative through the catalytic activity of a highly abundant class I*

chitinase. Applied Biochemistry and Biotechnology, 2010. **160**(6): p. 1573-1584.

225. Gnanamanickam, S.S., *Rice and its importance to human life*, *biological control of rice diseases*. 2009, Springer. p. 1-11.

226. JULIANO, B.O., *Rice starch: Production, properties, and uses*, in *Starch: chemistry and technology*. 1984, Elsevier. p. 507-528.

227. Ma, J., Y. Li, Q. Ye, et al., *Constituents of red yeast rice, a traditional Chinese food and medicine.* Journal of Agricultural and Food Chemistry, 2000. **48**(11): p. 5220-5225.

]

228. Umadevi, M., R. Pushpa, K. Sampathkumar, and D. Bhowmik, *Rice-traditional medicinal plant in India.* Journal of Pharmacognosy and Phytochemistry, 2012. **1**(1): p. 6-12.

229. Zavaleta, N., D. Figueroa, J. Rivera, J. Sánchez, S. Alfaro, and B. Lönnerdal, *Efficacy of rice-based oral rehydration solution containing recombinant human lactoferrin and lysozyme in Peruvian children with acute diarrhea.* Journal of Pediatric Gastroenterology and Nutrition, 2007. **44**(2): p. 258-264.

230. Hegsted, M., M.M. Windhauser, S.K. Morris, and S.B. Lester, *Stabilized rice bran and oat bran lower cholesterol in humans.* Nutrition Research, 1993. **13**(4): p. 387-398.

231. Qiu, Y., Q. Liu, and T. Beta, *Antioxidant properties of commercial wild rice and analysis of soluble and insoluble phenolic acids.* Food Chemistry, 2010. **121**(1): p. 140-147.

232. Huang, J., S. Nandi, L. Wu, et al., *Expression of natural antimicrobial human lysozyme in rice grains.* Molecular Breeding, 2002. **10**(1-2): p. 83-94.

233. Jabeen, N., Z. Chaudhary, M. Gulfraz, H. Rashid, and B. Mirza, *Expression of rice chitinase gene in genetically engineered tomato confers enhanced resistance to Fusarium wilt and early blight.* The Plant Pathology Journal, 2015. **31**(3): p. 252.

234. Mizuno, R., Y. Itoh, Y. Nishizawa, Y. Kezuka, K. Suzuki, T. Nonaka, and T. Watanabe, *Purification and characterization of a rice class I chitinase, OsChia1b, produced in Esherichia coli.* Bioscience, Biotechnology, and Biochemistry, 2008. **72**(3): p. 893-895.

235. Nishizawa, Y., Z. Nishio, K. Nakazono, M. Soma, E. Nakajima, M. Ugaki, and T. Hibi, *Enhanced resistance to blast (Magnaporthe grisea) in transgenic Japonica rice by constitutive expression of rice chitinase.* Theoretical and Applied Genetics, 1999. **99**(3-4): p. 383-390.

236. Bushuk, W., *Rye production and uses worldwide.* Cereal Foods World, 2001. **46**(2): p. 70-73.

237. Zhuang, S., R. Shetty, M. Hansen, A. Fromberg, P.B. Hansen, and T.J. Hobley, *Brewing with 100% unmalted grains: barley, wheat, oat and rye.* European Food Research and Technology, 2017. **243**(3): p. 447-454.

238. Lahne, J., *Aroma characterization of American rye whiskey by chemical and sensory assays.* 2010.

239. Andreasen, M.F., A.-K. Landbo, L.P. Christensen, Å. Hansen, and A.S. Meyer, *Antioxidant effects of phenolic rye (Secale cereale L.) extracts, monomeric hydroxycinnamates, and ferulic acid dehydrodimers on human low-density lipoproteins.* Journal of Agricultural and Food Chemistry, 2001. **49**(8): p. 4090-4096.

240. Bondia-Pons, I., A.-M. Aura, S. Vuorela, M. Kolehmainen, H. Mykkänen, and K. Poutanen, *Rye phenolics in nutrition and health.* Journal of Cereal Science, 2009. **49**(3): p. 323-336.

241. Taira, T., T. Ohnuma, T. Yamagami, Y. Aso, M. Ishiguro, and M. Ishihara, *Antifungal activity of rye (Secale cereale) seed chitinases: the different binding manner of class I and class II chitinases to the fungal cell walls.* Bioscience, Biotechnology, and Biochemistry, 2002. **66**(5): p. 970-977.

242. Yamagami, T. and G. Funatsu, *Purification and some properties of three chitinases from the seeds of rye (Secale cereale).* Bioscience, Biotechnology, and Biochemistry, 1993. **57**(4): p. 643-647.

243. Ramatoulaye, F., C. Mady, and S. Fallou, *Production and use sorghum: a literature review.* Journal of Nutritional Health & Food Science, 2016. **4**(1): p. 1-4.

244. Clottey, V., L. Wairegi, A. Bationo, A. Mando, and R. Kanton, *Sorghum-and millet-legume cropping systems*. 2015, Africa Soil Health Consortium.

245. Billa, E., D.P. Koullas, B. Monties, and E.G. Koukios, *Structure and composition of sweet sorghum stalk components.* Industrial Crops and Products, 1997. **6**(3-4): p. 297-302.

246. Taylor, J.R., T.J. Schober, and S.R. Bean, *Novel food and non-food uses for sorghum and millets.* Journal of Cereal Science, 2006. **44**(3): p. 252-271.

247. Prasad, S., A. Singh, N. Jain, and H. Joshi, *Ethanol production from sweet sorghum syrup for utilization as automotive fuel in India.* Energy & Fuels, 2007. **21**(4): p. 2415-2420.

248. Trujillo-de Santiago, G., C. Rojas-de Gante, S. García-Lara, L. Verdolotti, E. Di Maio, and S. Iannace, *Thermoplastic processing of blue maize and white sorghum flours to produce bioplastics.* Journal of Polymers and the Environment, 2015. **23**(1): p. 72-82.

249. Dykes, L. and L.W. Rooney, *Sorghum and millet phenols and antioxidants.* Journal of Cereal Science, 2006. **44**(3): p. 236-251.

250. Varady, K.A., Y. Wang, and P.J. Jones, *Role of policosanols in the prevention and treatment of cardiovascular disease.* Nutrition Reviews, 2003. **61**(11): p. 376-383.

251. Prom, L.K., R.D. Waniska, A.I. Kollo, W.L. Rooney, and F.P. Bejosano, *Role of chitinase and sormatin accumulation in the resistance of sorghum cultivars to grain mold.* Journal of Agricultural and Food Chemistry, 2005. **53**(14): p. 5565-5570.

252. Ratnavathi, C. and R. Sashidhar, *Induction of chitinase in response to Aspergillus infection in sorghum (Sorghum bicolor (L) Moench).* Journal of the Science of Food and Agriculture, 2004. **84**(12): p. 1521-1527.

253. Waniska, R., R. Venkatesha, A. Chandrashekar, et al., *Antifungal proteins and other mechanisms in the control of sorghum stalk rot and grain mold.* Journal of Agricultural and Food Chemistry, 2001. **49**(10): p. 4732-4742.

254. Krishnaveni, S., G. Liang, S. Muthukrishnan, and A. Manickam, *Purification and partial characterization of chitinases from sorghum seeds.* Plant Science, 1999. **144**(1): p. 1-7.

255. Kahlon, T.S., *Nutritional implications and uses of wheat and oat kernel oil.* Cereal Foods World (USA), 1989.

256. Shewry, P.R. and S.J. Hey, *The contribution of wheat to human diet and health.* Food and Energy Security, 2015. **4**(3): p. 178-202.

257. Day, L., M. Augustin, I. Batey, and C. Wrigley, *Wheat-gluten uses and industry needs.* Trends in Food Science & Technology, 2006. **17**(2): p. 82-90.

258. Badiali, D., E. Corazziari, F.I. Habib, et al., *Effect of wheat bran in treatment of chronic nonorganic constipation.* Digestive Diseases and Sciences, 1995. **40**(2): p. 349-356.

259. Teglia, A. and G. Secchi, *New protein ingredients for skin detergency: Native wheat protein–surfactant complexes.* International Journal of Cosmetic Science, 1994. **16**(6): p. 235-246.

260. Mujoriya, R. and R.B. Bodla, *A study on wheat grass and its Nutritional value.* Food Science and Quality Management, 2011. **2**: p. 1-8.

261. Adom, K.K., M.E. Sorrells, and R.H. Liu, *Phytochemical profiles and antioxidant activity of wheat varieties.* Journal of Agricultural and Food Chemistry, 2003. **51**(26): p. 7825-7834.

262. Mo, Y. and L.-Y. Lim, *Paclitaxel-loaded PLGA nanoparticles: potentiation of anticancer activity by surface conjugation with wheat germ agglutinin.* Journal of Controlled Release, 2005. **108**(2-3): p. 244-262.

263. Nishina, P.M. and R.A. Freedland, *The effects of dietary fiber feeding on cholesterol metabolism in rats.* The Journal of Nutrition, 1990. **120**(7): p. 800-805.

264. Jia, J., H. Ma, W. Zhao, Z. Wang, R. He, and L. Luo, *Study on the antihypertensive and antioxidant properties of wheat germ protein fractions hydrolysates.* Acta Nutrimenta Sinica, 2010. **32**(1): p. 55-59.

265. Gauthier, L., M.-N. Bonnin-Verdal, G. Marchegay, L. Pinson-Gadais, C. Ducos, F. Richard-Forget, and V. Atanasova-Penichon, *Fungal biotransformation of chlorogenic and caffeic acids by Fusarium graminearum: new insights in the contribution of phenolic acids to resistance to deoxynivalenol accumulation in cereals.* International Journal of Food Microbiology, 2016. **221**: p. 61-68.

266. Zhu, F., *Structures, physicochemical properties, and applications of amaranth starch.* Critical Reviews in Food Science and Nutrition, 2017. **57**(2): p. 313-325.

267. O'Brien, G.K. and M. Price, *Amaranth grain and vegetable types.* ECHO Technical note, 2008.

268. Monteros, C., C. Nieto, C. Caicedo, M. Rivera, and C. Vimos, *" INIAP Alegría": Primera variedad mejorada de amaranto para la Sierra ecuatoriana.* 1994.

269. Cole, J.N., *Amaranth from the past for the future*. 1979: Rodale Press.

270. Myers, R.L. and D.H. Putnam, *Growing grain amaranth as a specialty crop.* Center for Alternative Crops & Products, 1988.

271. Paśko, P., M. Sajewicz, S. Gorinstein, and Z. Zachwieja, *Analysis of selected phenolic acids and flavonoids in Amaranthus cruentus and Chenopodium quinoa seeds and sprouts by HPLC.* Acta Chromatographica, 2008. **20**(4): p. 661-672.

272. Mendonça, S., P.H. Saldiva, R.J. Cruz, and J.A. Arêas, *Amaranth protein presents cholesterol-lowering effect.* Food Chemistry, 2009. **116**(3): p. 738-742.

273. Barrio, D.A. and M.C. Añón, *Potential antitumor properties of a protein isolate obtained from the seeds of Amaranthus mantegazzianus.* European Journal of Nutrition, 2010. **49**(2): p. 73-82.

274. Ramírez-Torres, G., N. Ontiveros, V. Lopez-Teros, J.A. Ibarra-Diarte, C. Reyes-Moreno, E.O. Cuevas-Rodríguez, and F. Cabrera-Chávez, *Amaranth protein hydrolysates efficiently reduce systolic blood pressure in spontaneously hypertensive rats.* Molecules, 2017. **22**(11): p. 1905.

275. Whittaker, P. and M. Ologunde, *Study on iron bioavailability in a native Nigerian grain amaranth cereal for young children, using a rat model.* Cereal Chemistry, 1990. **67**(5): p. 505-508.

276. Cai, Y., H. Corke, D. Wang, and W. Li, *Buckwheat: Overview.* Elsevier 2016.

277. Berglund, D.R., *Buckwheat production.* North Dakota State University and U.S. Department of Agriculture 2003.

278. Pomeranz, Y. and K. Lorenz, *Buckwheat: structure, composition, and utilization.* Critical Reviews in Food Science & Nutrition, 1983. **19**(3): p. 213-258.

279. Ratan, P. and P. Kothiyal, *Fagopyrum esculentum Moench (common buckwheat) edible plant of Himalayas: a review.* Asian Journal of Pharmacy Life Science ISSN, 2011. **2231**: p. 4423.

280. Ahmed, A., N. Khalid, A. Ahmad, N. Abbasi, M. Latif, and M. Randhawa, *Phytochemicals and biofunctional properties of buckwheat: a review.* The Journal of Agricultural Science, 2014. **152**(3): p. 349-369.

281. Przybylski, R., Y. Lee, and N. Eskin, *Antioxidant and radical‐scavenging activities of buckwheat seed components.* Journal of the American Oil Chemists' Society, 1998. **75**(11): p. 1595.

282. Christa, K. and M. Soral-Śmietana, *Buckwheat grains and buckwheat products–nutritional and prophylactic value of their components–a review.* Czech Journal of Food Sciences, 2008. **26**(3): p. 153-162.

283. Kwak, C., S. Park, S. Lim, S. Kim, and M. Lee, *Antioxidative and antimutagenic effects of Korean buckwheat, sorghum, millet and Job's tears.* Journal of The Korean Society of Food Science and Nutrition, 2004.

284. Brindzová, L., M. Mikulášová, M. Takácsová, S. Mošovská, and A. Opattová, *Evaluation of the mutagenicity and antimutagenicity of extracts from oat, buckwheat and wheat bran in the Salmonella/microsome assay.* Journal of Food Composition and Analysis, 2009. **22**(1): p. 87-90.

285. Couch, J.F., J. Naghski, and C.F. Krewson, *Buckwheat as a source of rutin.* Science, 1946. **103**(2668): p. 197-198.

286. Jacobsen, S.-E., *The worldwide potential for quinoa (Chenopodium quinoa Willd.).* Food Reviews International, 2003. **19**(1-2): p. 167-177.

287. Bhargava, A., S. Shukla, and D. Ohri, *Chenopodium quinoa—an Indian perspective.* Industrial Crops and Products, 2006. **23**(1): p. 73-87.

288. Jancurová, M., L. Minarovičová, and A. Dandar, *Quinoa–a rewiev.* Czech Journal of Food Sciences, 2009. **27**(2): p. 71-79.

289. James, L.E.A., *Quinoa (Chenopodium quinoa Willd.): composition, chemistry, nutritional, and functional properties.* Advances in Food and Nutrition Research, 2009. **58**: p. 1-31.

290. Zalles, J. and M. De Lucca, *Utasan Utjir Qollanaka.* Medicinas que hay en nuestro jardín. Descripción y uso de, 2006. **100**.

291. Horák, M., *4.1. 2 Medicinal Plants in the Tropical Andean Region. Quinoa (Chenopodium quinoa Willd.) and Coca (Erythroxylum sp.), millenary treasures for medicinal treatment Granda L.; Rosero MG; Rosero A.* Folia Primatologica: p. 58.

292. Bussmann, R.W. and D. Sharon, *Traditional medicinal plant use in Northern Peru: tracking two thousand years of healing culture.* Journal of Ethnobiology and Ethnomedicine, 2006. **2**(1): p. 47.

293. Quinoa, F., *An ancient crop to contribute to world food security.* Regional Office for Latin America and the Caribbean, 2011. **2**: p. 73-87.

294. Takao, T., N. Watanabe, K. Yuhara, et al., *Hypocholesterolemic effect of protein isolated from quinoa (Chenopodium quinoa Willd.) seeds.* Food Science and Technology Research, 2005. **11**(2): p. 161-167.

295. Pasko, P., H. Barton, P. Zagrodzki, et al., *Effect of diet supplemented with quinoa seeds on oxidative status in plasma and selected tissues of high fructose-fed rats.* Plant Foods for Human Nutrition, 2010. **65**(2): p. 146-151.

296. Paśko, P., P. Zagrodzki, H. Bartoń, J. Chłopicka, and S. Gorinstein, *Effect of quinoa seeds (Chenopodium quinoa) in diet on some biochemical parameters and essential elements in blood of high fructose-fed rats.* Plant Foods for Human Nutrition, 2010. **65**(4): p. 333-338.

297. Vilcacundo, R. and B. Hernández-Ledesma, *Nutritional and biological value of quinoa (Chenopodium quinoa Willd.).* Current Opinion in Food Science, 2017. **14**: p. 1-6.

298. Tang, Y. and R. Tsao, *Phytochemicals in quinoa and amaranth grains and their antioxidant, anti‐inflammatory, and potential health beneficial effects: a review.* Molecular Nutrition & Food Research, 2017. **61**(7): p. 1600767.

299. Tang, Y., X. Li, B. Zhang, P.X. Chen, R. Liu, and R. Tsao, *Characterisation of phenolics, betanins and antioxidant activities in seeds of three Chenopodium quinoa Willd. genotypes.* Food Chemistry, 2015. **166**: p. 380-388.

300. Stuardo, M. and R. San Martín, *Antifungal properties of quinoa (Chenopodium quinoa Willd) alkali treated saponins against Botrytis cinerea.* Industrial Crops and Products, 2008. **27**(3): p. 296-302.

301. Verma, H. and V. Baranwal, *Antiviral activity and the physical properties of the leaf extract ofChenopodium ambrosoides L.* Proceedings: Plant Sciences, 1983. **92**(6): p. 461-465.

302. Hirose, Y., T. Fujita, T. Ishii, and N. Ueno, *Antioxidative properties and flavonoid composition of Chenopodium quinoa seeds cultivated in Japan.* Food Chemistry, 2010. **119**(4): p. 1300-1306.

303. Foucault, A.-S., P. Even, R. Lafont, et al., *Quinoa extract enriched in 20-hydroxyecdysone affects energy homeostasis and intestinal fat absorption in mice fed a high-fat diet.* Physiology & Behavior, 2014. **128**: p. 226-231.

304. Yao, Y., X. Yang, Z. Shi, and G. Ren, *Anti‐inflammatory activity of saponins from quinoa (Chenopodium quinoa Willd.) seeds in*

lipopolysaccharide‐stimulated RAW 264.7 macrophages cells. Journal of Food Science, 2014. **79**(5): p. H1018-H1023.

305. Sithisarn, P., S. Jarikasem, and K. Thisayakorn, *Acanthopanax trifoliatus, a potential adaptogenic Thai vegetable for health supplement.* Phytochemicals-Bioactivities and Impact on Health. Intech, Rijeka, Croatia, 2011: p. 253-68.

306. Perry, L.M. and J. Metzger, *Medicinal plants of east and southeast Asia: attributed properties and uses*. 1980: MIT press.

307. Wiart, C., *Ethnopharmacology of medicinal plants: Asia and the Pacific*. 2007: Springer Science & Business Media.

308. Loi, D.T., *Glossary of Vietnamese Medical Plants.* Hanoi S & T Publication, 2001: p. 379-82.

309. Sithisarn, P. and S. Jarikasem, *Antioxidant activity of Acanthopanax trifoliatus.* Medical Principles and Practice, 2009. **18**(5): p. 393-398.

310. Roslida, A., H. Siti Nurusaadah, and O. Fezah, *Gastroprotective effect of Acanthopanax trifoliatus on experimentally induced acute ulcer in rats.* Pharmacologyonline, 2010. **2**: p. 828-41.

311. Sainan, S., M. Sulaiman, M. Somchit, M. Hidayat, and Z. Zakaria. *Evaluation of antinociceptive activities of Acanthopanax trifoliatus in mice*. Proceedings of the 20 th Scientific Meeting of The Malaysian Society of Pharmacology and Physiology, Penang. Malaysia, 2005.

312. Chen, C.-H., *Study of anti-tumor effect of Acanthopanax trifoliatus; understand of knowledge and attitude of Chinese herb between nurses and society in eastern Taiwan*. 2006, PhD thesis, Graduate Institude of Medical Sciences, Tzu Chi University, Taiwan.

313. Hamid, R.A., T.H. Kee, and F. Othman, *Anti-inflammatory and anti-hyperalgesic activities of Acanthopanax trifoliatus (L) Merr leaves.* Pharmacognosy Research, 2013. **5**(2): p. 129.

314. Chien, T.-M., P.-C. Hsieh, S.-S. Huang, J.-S. Deng, Y.-L. Ho, Y.-S. Chang, and G.-J. Huang, *Acanthopanax trifoliatus inhibits lipopolysaccharide-induced inflammatory response in vitro and in vivo.* The Kaohsiung journal of Medical Sciences, 2015. **31**(10): p. 499-509.

315. Pace, C.N., F. Vajdos, L. Fee, G. Grimsley, and T. Gray, *How to measure and predict the molar absorption coefficient of a protein.* Protein Science, 1995. **4**(11): p. 2411-2423.

316. Nguyen, G.K., S. Zhang, N.T. Nguyen, P.Q. Nguyen, M.S. Chiu, A. Hardjojo, and J.P. Tam, *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 Biology and Chemistry, 2011. **286**(27): p. 24275-87.

317. Jeener, J., B. Meier, P. Bachmann, and R. Ernst, *Investigation of exchange processes by two ‐dimensional NMR spectroscopy.* The Journal of Chemical Physics, 1979. **71**(11): p. 4546-4553.

318. Kumar, A., R. Ernst, and K. Wüthrich, *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules*, in *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich*. 1995, World Scientific. p. 103-108.

319. Davis, D.G. and A. Bax, *Assignment of complex proton NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy.* Journal of the American Chemical Society, 1985. **107**(9): p. 2820-2821.

320. Bax, A. and D.G. Davis, *MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy.* Journal of Magnetic Resonance (1969), 1985. **65**(2): p. 355-360.

321. Rance, M., O. Sørensen, G. Bodenhausen, G. Wagner, R. Ernst, and K. Wüthrich, *Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering.* Biochemical and Biophysical research communications, 1983. **117**(2): p. 479-485.

322. Wüthrich, K., M. Billeter, and W. Braun, *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*, in *Nmr In Structural Biology: A Collection of Papers by Kurt Wüthrich*. 1995, World Scientific. p. 292-304.

323. Liu, M., X.-a. Mao, C. Ye, H. Huang, J.K. Nicholson, and J.C. Lindon, *Improved watergate pulse sequences for solvent suppression in NMR spectroscopy.* Journal of Magnetic Resonance, 1998. **132**(1): p. 125-129.

324. Hwang, T.-L. and A. Shaka, *Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients.* Journal of Magnetic Resonance. Series A (Print), 1995. **112**(2): p. 275-279.

325. Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.

326. Brünger, A.T., P.D. Adams, G.M. Clore, et al., *Crystallography & NMR system: a new software suite for macromolecular structure determination.* Acta Crystallographica Section D: Biological Crystallography, 1998. **54**(5): p. 905-921.

327. Laskowski, R.A., J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, and J.M. Thornton, *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* Journal of Biomolecular NMR, 1996. **8**(4): p. 477-486.

328. DeLano, W.L., *Pymol: An open-source molecular graphics tool.* CCP4 Newsletter on Protein Crystallography, 2002. **40**(1): p. 82-92.

329. Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor, *Development and validation of a genetic algorithm for flexible docking.* Journal of Molecular Biology, 1997. **267**(3): p. 727-748.

330. Jenssen, H. and S.I. Aspmo, *Serum stability of peptides*, in *Peptide-based drug design*. 2008, Springer. p. 177-186.

331. Ye, X.Y. and T.B. Ng, *A new antifungal peptide from rice beans.* Journal of Peptide Research, 2002. **60**(2): p. 81-7.

332. Wiegand, I., K. Hilpert, and R.E. Hancock, *Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances.* Nature Protocols, 2008. **3**(2): p. 163.

333. Marsh, A.J., O. O'Sullivan, R.P. Ross, P.D. Cotter, and C. Hill, *In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria.* BioMed Central Genomics, 2010. **11**(1): p. 679.

334. Porto, W.F., V.A. Souza, D.O. Nolasco, and O.L. Franco, *In silico identification of novel hevein-like peptide precursors.* Peptides, 2012. **38**(1): p. 127-36.

335. Boguski, M.S., T.M. Lowe, and C.M. Tolstoshev, *dbEST—database for "expressed sequence tags".* Nature Genetics, 1993. **4**(4): p. 332.

336. Matasci, N., L.-H. Hung, Z. Yan, et al., *Data access for the 1,000 Plants (1KP) project.* Gigascience, 2014. **3**(1): p. 2047-217X-3-17.

337. Petersen, T.N., S. Brunak, G. von Heijne, and H. Nielsen, *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nature Methods, 2011. **8**(10): p. 785-786.

338. McWilliam, H., W. Li, M. Uludag, et al., *Analysis tool web services from the EMBL-EBI.* Nucleic Acids Research, 2013. **41**(W1): p. W597-W600.

339. Crooks, G.E., G. Hon, J.-M. Chandonia, and S.E. Brenner, *WebLogo: a sequence logo generator.* Genome Research, 2004. **14**(6): p. 1188-1190.

340. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* Nucleic Acids Research, 2016. **44**(W1): p. W242-W245.

341. Galwey, N., *The potential of quinoa as a multi-purpose crop for agricultural diversification: a review.* Industrial Crops and Products, 1992. **1**(2-4): p. 101-106.

342. Zhu, N., S. Sheng, D. Li, E.J. Lavoie, M.V. Karwe, R.T. Rosen, and C.T. HO, *Antioxidative flavonoid glycosides from quinoa seeds (Chenopodium quinoa Willd).* Journal of Food Lipids, 2001. **8**(1): p. 37-44.

343. Foucault, A.S., V. Mathé, R. Lafont, et al., *Quinoa extract enriched in 20‐hydroxyecdysone protects mice from diet‐induced obesity and modulates adipokines expression.* Obesity, 2012. **20**(2): p. 270-277.

344. Graf, B.L., A. Poulev, P. Kuhn, M.H. Grace, M.A. Lila, and I. Raskin, *Quinoa seeds leach phytoecdysteroids and other compounds with anti-diabetic properties.* Food Chemistry, 2014. **163**: p. 178-185.

345. Brady, K., C.-T. Ho, R.T. Rosen, S. Sang, and M.V. Karwe, *Effects of processing on the nutraceutical profile of quinoa.* Food Chemistry, 2007. **100**(3): p. 1209-1216.

346. Woldemichael, G.M. and M. Wink, *Identification and biological activities of triterpenoid saponins from Chenopodium quinoa.* Journal of Agricultural and Food Chemistry, 2001. **49**(5): p. 2327-2332.

347. Vugmeyster, Y., X. Xu, F.-P. Theil, L.A. Khawli, and M.W. Leach, *Pharmacokinetics and toxicology of therapeutic proteins: advances and challenges.* World Journal of Biological Chemistry, 2012. **3**(4): p. 73.

348. Xu, X. and Y. Vugmeyster, *Challenges and opportunities in absorption, distribution, metabolism, and excretion studies of therapeutic biologics.* The American Association of Pharmaceutical Scientists Journal, 2012. **14**(4): p. 781-791.

349. Hammami, R., J.B. Hamida, G. Vergoten, and I. Fliss, *PhytAMP: a database dedicated to antimicrobial plant peptides.* Nucleic Acids Research, 2009. **37**(suppl 1): p. D963-D968.

350. Tam, J.P., Y.A. Lu, J.L. Yang, and K.W. Chiu, *An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides.* Proceedings of the National Academy of Sciences USA, 1999. **96**(16): p. 8913-8.

351. Ai-Guo, G., S.M. Hakimi, C.A. Mittanck, et al., *Fungal pathogen protection in potato by expression of a plant defensin peptide.* Nature Biotechnology, 2000. **18**(12): p. 1307.

352. Kader, J.-C., *Lipid-transfer proteins in plants.* Annual Review of Plant Biology, 1996. **47**(1): p. 627-654.

353. Berrocal-Lobo, M., A. Segura, M. Moreno, G. López, F. Garcıa-Olmedo, and A. Molina, *Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection.* Plant Physiology, 2002. **128**(3): p. 951-961.

354. Jarvis, D.E., Y.S. Ho, D.J. Lightfoot, et al., *The genome of Chenopodium quinoa.* Nature, 2017. **542**(7641): p. 307.

355. Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner, *WebLogo: a sequence logo generator.* Genome Research, 2004. **14**(6): p. 1188-90.

356. Lerner, D.R. and N.V. Raikhel, *The gene for stinging nettle lectin (Urtica dioica agglutinin) encodes both a lectin and a chitinase.* Journal of Biological Chemistry, 1992. **267**(16): p. 11085-11091.

357. Smith, J.J. and N.V. Raikhel, *Nucleotide sequences of cDNA clones encoding wheat germ agglutinin isolectins A and D.* Plant Molecular Biology, 1989. **13**(5): p. 601-603.

358. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery.* Nature Reviews. Drug discovery, 2005. **4**(3): p. 206.

359. Lauand, C., P. Rezende-Teixeira, B.A. Cortez, E.L. de Oliveira Niero, and G.M. Machado-Santelli, *Independent of ErbB1 gene copy number, EGF stimulates migration but is not associated with cell proliferation in non-small cell lung cancer.* Cancer Cell International, 2013. **13**(1): p. 1-15.

360. Tong, J. and Z. Wang, *Analysis of epidermal growth factor receptor-induced cell motility by wound healing assay*, in *ErbB Receptor Signaling*. 2017, Springer. p. 159-163.

361. Rogozhin, E.A., M.P. Slezina, A.A. Slavokhotova, et al., *A novel antifungal peptide from leaves of the weed Stellaria media L.* Biochimie, 2015. **116**: p. 125-32.

362. Valk, V., A. Lammerts van Bueren, R.M. van der Kaaij, and L. Dijkhuizen, *Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain alpha-amylase enzymes.* Federation of European Biochemical Societies Journal, 2016. **283**(12): p. 2354-68.

363. Gillon, A.D., I. Saska, C.V. Jennings, R.F. Guarino, D.J. Craik, and M.A. Anderson, *Biosynthesis of circular proteins in plants.* The Plant Journal, 2008. **53**(3): p. 505-515.

364. Shubin, N., C. Tabin, and S. Carroll, *Deep homology and the origins of evolutionary novelty.* Nature, 2009. **457**(7231): p. 818-823.

365. Peng, H., Y. Zheng, M. Chen, Y. Wang, Y. Xiao, and Y. Gao, *A starch-binding domain identified in alpha-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch.* Federation of European Biochemical Societies Letters, 2014. **588**(7): p. 1161-7.

366. Tang, W.J., J.G. Fernandez, J.J. Sohn, and C.T. Amemiya, *Chitin is endogenously produced in vertebrates.* Current Biology, 2015. **25**(7): p. 897-900.

367. Dahlquist, F.W., L. Jao, and M. Raftery, *On the binding of chitin oligosaccharides to lysozyme.* Proceedings of the National Academy of Sciences USA, 1966. **56**(1): p. 26-30.

368. Collinge, D.B., K.M. Kragh, J.D. Mikkelsen, K.K. Nielsen, U. Rasmussen, and K. Vad, *Plant chitinases.* Plant Journal, 1993. **3**(1): p. 31-40.

369. Giovanini, M.P., K.D. Saltzmann, D.P. Puthoff, M. Gonzalo, H.W. Ohm, and C.E. Williams, *A novel wheat gene encoding a putative chitin-binding lectin is associated with resistance against Hessian fly.* Molecular Plant Pathology, 2007. **8**(1): p. 69-82.

370. Wessels, J., *A steady‐state model for apical wall growth in fungi.* Plant Biology, 1988. **37**(1): p. 3-16.

371. Estrada, A., C.-H. Yun, A.V. Kessel, B. Li, S. Hauta, and B. Laarveld, *Immunomodulatory activities of oat β-glucan in vitro and in vivo.* Microbiology and immunology, 1997. **41**(12): p. 991-998.

372. Akkol, E.K., I. Süntar, I.E. Orhan, H. Keles, A. Kan, and G. Çoksari, *Assessment of dermal wound healing and in vitro antioxidant properties of Avena sativa L.* Journal of Cereal Science, 2011. **53**(3): p. 285-290.

373. Michelle Garay, M., M. Judith Nebus, and B. Menas Kizoulis, *Anti-inflammatory activities of colloidal oatmeal (Avena sativa) contribute to the effectiveness of oats in treatment of itch associated with dry, irritated skin.* Journal of Drugs in Dermatology, 2015. **14**(1): p. 43-48.

374. Das, G. and M.M. Joseph, *Phytochemical screening of oats (avena sativa).* World Journal of Pharmaceutical Research, 2017.

375. Gutierrez-Gonzalez, J.J., M.L. Wise, and D.F. Garvin, *A developmental profile of tocol accumulation in oat seeds.* Journal of Cereal Science, 2013. **57**(1): p. 79-83.

376. Bahraminejad, S., R. Asenstorfer, I. Riley, and C. Schultz, *Analysis of the antimicrobial activity of flavonoids and saponins isolated from the shoots of oats (Avena sativa L.).* Journal of Phytopathology, 2008. **156**(1): p. 1-7.

377. Osbourn, A.E., *Saponins in cereals.* Phytochemistry, 2003. **62**(1): p. 1-4.

378. Lee, O.S., B. Lee, N. Park, et al., *Pn-AMPs, the hevein-like proteins from Pharbitis nil confers disease resistance against phytopathogenic fungi in tomato, Lycopersicum esculentum.* Phytochemistry, 2003. **62**(7): p. 1073-9.

379. Araki, T. and T. Torikata, *Structural classification of plant chitinases: two subclasses in class I and class II chitinases.* Bioscience, Biotechnology and Biochemistry, 1995. **59**(2): p. 336-8.

380. Bergthorsson, U., K.L. Adams, B. Thomason, and J.D. Palmer, *Widespread horizontal transfer of mitochondrial genes in flowering plants.* Nature, 2003. **424**(6945): p. 197-201.

381. Richardson, A.O. and J.D. Palmer, *Horizontal gene transfer in plants.* Journal of Experimental Botany, 2007. **58**(1): p. 1-9.

382. Pelegrini, P.B. and O.L. Franco, *Plant γ-thionins: novel insights on the mechanism of action of a multi-functional class of defense proteins.* The International Journal of Biochemistry & Cell Biology, 2005. **37**(11): p. 2239-2253.

383. Panchy, N., M. Lehti-Shiu, and S.H. Shiu, *Evolution of Gene Duplication in Plants.* Plant Physiology, 2016. **171**(4): p. 2294-316.

384. Craik, D.J. and U. Malik, *Cyclotide biosynthesis.* Current Opinion in Chemical Biology, 2013. **17**(4): p. 546-54.

385. Arboretum, R., G.W. w Warszawie, and A. w Rogowie, *Woody species of Araliaceae at the Rogów Arboretum.* Rocznik Dendrologiczny, 2006. **54**: p. 35-50.

386. Meng, Z., Y. Sakai, Y. Ose, et al., *Antimutagenic activity of medical plants in traditional Chinese medicines.* Mutation Research/Environmental Mutagenesis and Related Subjects, 1988. **203**(5): p. 378-379.

387. Phuong, N.T., K.A. Lee, S.J. Jeong, C.X. Fu, J.K. Choi, Y.H. Kim, and J.S. Kang, *Capillary electrophoretic method for the determination of diterpenoid isomers in Acanthopanax species.* Journal of Pharmaceutical and Biomedical Analysis, 2006. **40**(1): p. 56-61.

388. Yook, C.-S., S.-Y. Chang, J.-H. Lai, S.-K. Ko, J.-H. Jeong, and N. Toshihiro, *Lupane-glycoside of Acanthopanax trifoliatus forma tristigmatis leaves.* Archives of Pharmacal Research, 1999. **22**(6): p. 629-632.

389. Van Kiem, P., X.F. Cai, C. Van Minh, J.J. Lee, and Y.H. Kim, *Lupane-triterpene carboxylic acids from the leaves of Acanthopanax trifoliatus.* Chemical and Pharmaceutical Bulletin, 2003. **51**(12): p. 1432-1435.

390. Chen, F.-c., Y.-M. Lin, and S. Lin, *Constituents of three-leaved Acanthopanax.* Phytochemistry, 1972.

391. Du, J. and L. Gao, *Chemical constituents of the leaves of Acanthopanax trifoliatus (Linn) Merr.* Zhongguo Zhong yao za zhi= Zhongguo zhongyao zazhi= China journal of Chinese Materia Medica, 1992. **17**(6): p. 356-7, 383.

392. Muselli, A., T.M. Hoi, L.D. Cu, L.D. Moi, J.M. Bessière, A. Bighelli, and J. Casanova, *Composition of the essential oil of Acanthopanax trifoliatus (L.) Merr.(Araliacaea) from Vietnam.* Flavour and Fragrance Journal, 1999. **14**(1): p. 41-44.

393. Wong, C.T., D.K. Rowlands, C.H. Wong, T.W. Lo, G.K. Nguyen, H.Y. Li, and J.P. Tam, *Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment.* Angewandte Chemie International Edition England, 2012. **51**(23): p. 5620-4.

394. Huang, J., K.H. Wong, S.V. Tay, A. Serra, S.K. Sze, and J.P. Tam, *Astratides: Insulin-Modulating, Insecticidal, and Antifungal Cysteine-Rich*

Peptides from Astragalus membranaceus. Journal of Natural Products, 2019. **82**(2): p. 194-204.

395. Huang, J., K.H. Wong, S.V. Tay, A. How, and J.P. Tam, *Cysteine-rich peptide fingerprinting as a general method for herbal analysis to differentiate Radix Astragali and Radix Hedysarum.* Frontiers in Plant Science, 2019. **10**: p. 973.

396. Berthelot, K., F. Peruch, and S. Lecomte, *Highlights on Hevea brasiliensis (pro)hevein proteins.* Biochimie, 2016. **127**: p. 258-70.

397. Rinaudo, M., *Chitin and chitosan: Properties and applications.* Progress in Polymer Science, 2006. **31**(7): p. 603-632.

398. Dutta, B., J. Huang, J. To, and J.P. Tam, *LIR motif-containing hyperdisulfide β-ginkgotide is cytoprotective, adaptogenic, and scaffold-ready.* Molecules, 2019. **24**(13): p. 2417.

399. Wang, H., D. Li, Z. Du, et al., *Antioxidant and anti-inflammatory properties of Chinese ilicifolius vegetable (Acanthopanax trifoliatus (L) Merr) and its reference compounds.* Food Science and Biotechnology, 2015. **24**(3): p. 1131-1138.

400. Birgisdottir, Å.B., T. Lamark, and T. Johansen, *The LIR motif–crucial for selective autophagy.* Journal of Cell Science, 2013. **126**(15): p. 3237-3247.

401. Noda, N.N., Y. Ohsumi, and F. Inagaki, *Atg8‐family interacting motif crucial for selective autophagy.* Federation of European Biochemical Societies Letters, 2010. **584**(7): p. 1379-1385.

402. Koehbach, J., A.F. Attah, A. Berger, et al., *Cyclotide discovery in Gentianales revisited—identification and characterization of cyclic cystine‐knot peptides and their phylogenetic distribution in Rubiaceae plants.* Peptide Science, 2013. **100**(5): p. 438-452.

403. Kitts, D.D. and K. Weiler, *Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery.* Current Pharmaceutical Design, 2003. **9**(16): p. 1309-1323.

404. Boguski, M.S., T.M. Lowe, and C.M. Tolstoshev, *dbEST—database for "expressed sequence tags".* Nature Genetics, 1993. **4**(4): p. 332-333.

405. Kanehisa, M. and P. Bork, *Bioinformatics in the post-sequence era.* Nature Genetics, 2003. **33**(3): p. 305-310.

406. Kim, H.-J., K.-H. Baek, B.-W. Lee, D. Choi, and C.-G. Hur, *In silico identification and characterization of microRNAs and their putative target genes in Solanaceae plants.* Genome, 2011. **54**(2): p. 91-98.

407. Elkon, R., K.I. Zeller, C. Linhart, C.V. Dang, R. Shamir, and Y. Shiloh, *In silico identification of transcriptional regulators associated with c-Myc.* Nucleic Acids Research, 2004. **32**(17): p. 4955-4961.

408. Masignani, V., E. Balducci, D. Serruto, et al., *In silico identification of novel bacterial ADP-ribosyltransferases.* International Journal of Medical Microbiology, 2004. **293**(7-8): p. 471-478.

409. Von Konrat, M., A.J. Shaw, and K.S. Renzaglia, *A special issue of Phytotaxa dedicated to Bryophytes: The closest living relatives of early land plants.* Phytotaxa, 2014. **9**(1): p. 5-10.

410. Kini, R.M. and R. Doley, *Structure, function and evolution of three-finger toxins: mini proteins with multiple targets.* Toxicon, 2010. **56**(6): p. 855-867.

411. Bishop, J., A.M. Dean, and T. Mitchell-Olds, *Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution.* Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5322-5327.

412. Shafee, T.M., F.T. Lay, T.K. Phan, M.A. Anderson, and M.D. Hulett, *Convergent evolution of defensin sequence, structure and function.* Cellular and Molecular Life Sciences, 2017. **74**(4): p. 663-682.

413. Gruber, C.W., A.G. Elliott, D.C. Ireland, et al., *Distribution and evolution of circular miniproteins in flowering plants.* The Plant Cell, 2008. **20**(9): p. 2471-2483.

414. Creighton, T.E. and N.J. Darby, *Functional evolutionary divergence of proteolytic enzymes and their inhibitors.* Trends in Biochemical Sciences, 1989. **14**(8): p. 319-324.

415. Franco, O.L., *Peptide promiscuity: an evolutionary concept for plant defense.* Federation of European Biochemical Societies Letters, 2011. **585**(7): p. 995-1000.

416. Balaji, R.A., A. Ohtake, K. Sato, P. Gopalakrishnakone, R.M. Kini, K.T. Seow, and B.-H. Bay, *λ-Conotoxins, a New Family of Conotoxins with Unique Disulfide Pattern and Protein Folding Isolation and Characterization from the Venom of Conus ,Marmoreus.* Journal of Biological Chemistry, 2000. **275**(50): p. 39516-39522.

417. Petersen, J.r., R. Teich, B. Becker, R.d. Cerff, and H. Brinkmann, *The GapA/B gene duplication marks the origin of Streptophyta (charophytes and land plants).* Molecular Biology and Evolution, 2006. **23**(6): p. 1109-1118.

418. Otto, S.P. and J. Whitton, *Polyploid incidence and evolution.* Annual Review of enetics, 2000. **34**(1): p. 401-437.

419. Stotz, H.U., B. Spence, and Y. Wang, *A defensin from tomato with dual function in defense and development.* Plant Molecular Biology, 2009. **71**(1-2): p. 131-143.

420. Zhu, Y.J., R. Agbayani, and P.H. Moore, *Ectopic expression of Dahlia merckii defensin DmAMP1 improves papaya resistance to Phytophthora palmivora by reducing pathogen vigor.* Planta, 2007. **226**(1): p. 87-97.

421. Koo, J.C., H.J. Chun, H.C. Park, 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): p. 441-452.

422. Clark, R.J., N.L. Daly, and D.J. Craik, *Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design.* Biochemical Journal, 2006. **394**(1): p. 85-93.

423. Krause, S., H.U. Schmoldt, A. Wentzel, M. Ballmaier, K. Friedrich, and H. Kolmar, *Grafting of thrombopoietin‐mimetic peptides into cystine knot miniproteins yields high‐affinity thrombopoietin antagonists and agonists.* The Federation of European Biochemical Societies Journal, 2007. **274**(1): p. 86-95.

424. Gunasekera, S., F.M. Foley, R.J. Clark, L. Sando, L.J. Fabri, D.J. Craik, and N.L. Daly, *Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides.* Journal of Medicinal Chemistry, 2008. **51**(24): p. 7697-7704.

425. Chan, L.Y., S. Gunasekera, S.T. Henriques, et al., *Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds.* Blood, The Journal of the American Society of Hematology, 2011. **118**(25): p. 6709-6717.

426. Zhu, X., J. Li, Y. Hong, et al., *99mTc-labeled cystine knot peptide targeting integrin αvβ6 for tumor SPECT imaging.* Molecular Pharmaceutics, 2014. **11**(4): p. 1208-1217.

427. Jiang, L., R.H. Kimura, X. Ma, et al., *A radiofluorinated divalent cystine knot peptide for tumor PET imaging.* Molecular Pharmaceutics, 2014. **11**(11): p. 3885-3892.

Appendices

Appendix A: Replicate sequence of putative 6C-, 8C- and 10C-hevein-like peptides in plants

6C-chitin-binding hevein-like peptides

Plant Species	Plant Family	Mature sequence
Amaranthus albus	Amaranthaceae	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
Amaranthus blitum	Amaranthaceae	
Amaranthus hybridus	Amaranthaceae	
Amaranthus tricolor	Amaranthaceae	
Amaranthus hybridus	Amaranthaceae	
Amaranthus caudatus	Amaranthaceae	
Amaranthus hypochondriacus	Amaranthaceae	
Alternanthera sessilis	Amaranthaceae	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
Atriplex hortensis	Amaranthaceae	
Atriplex prostrata	Amaranthaceae	
Atriplex rosea	Amaranthaceae	
Heliotropium calcicola	Boraginaceae	
Mollugo nudicaulis	Molluginaceae	
Aerva persica	Amaranthaceae	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
Alternanthera caracasana	Amaranthaceae	
Heliotropium filiforme	Boraginaceae	
Heliotropium tenellum	Boraginaceae	
Heliotropium texanum	Boraginaceae	
Amaranthus retroflexus	Amaranthaceae	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
Euphorbia mesembryanthemifolia	*Euphorbiaceae*	
Flaveria bidentis	Asteraceae	
Portulaca umbraticola	Portulacaceae	

6C-non-chitin-binding hevein-like peptides

Plant Species	Plant Family	Mature Sequence
Achyranthes aspera	Amaranthaceae	CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS
Amaranthus cruentus	Amaranthaceae	
Amaranthus paniculatus	Amaranthaceae	
Celosia argentea	Amaranthaceae	
Gypsophila vaccaria	Caryophyllaceae	CIPKYNRCGPAMDGVQCCEPYTCTSQYYGNCS
Vaccaria hispanica	Caryophyllaceae	
Aerva lanata	Amaranthaceae	CIPKYNRCGTHYGVPCCEGTCTSEYYGICS\
Portulaca mauii	Portulacaceae	
Dianthus caryophyllus	Caryophyllaceae	CIPQYNRCGPAMDGVPCCEPYTCSSQYYGSCS
Saponaria officianalis	Caryophyllaceae	

8C-chitin-binding hevein-like peptides

Plant Species	Plant Family	Mature sequence
Portulaca molokiniensis	Portulacacea	CGRQAGGALCPNRLCCSQYGWCGNTDPYCGAGCQS QC
Portulaca oleracea	Portulacacea	
Crossopetalum rhacoma	Celastraceae	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGN CIGQCP
Falcatifolium taxoides	Podocarpaceae	
Amentotaxus argotaenia	Taxaceae	DPTCSPAGNFFCNNGRCCSRFNWCGTGPSYCGRGNC IAQCP
Pinus parviflora	Pinaceae	
Fokienia hodginsii	*Cupressaceae*	DPTCSPAGNFWCDSGRCCSIYNWCGSISDYCASGNCL AQCWP
Flaveria bidentis	Asteraceae	
Pilgerodendron uviferum	*Cupressaceae*	DPTCSPAGNFWCDSGRCCSIYNWCGSTSDYCASGNC LAQCWP
Taxodium distichum	*Cupressaceae*	

Torreya nucifera	Taxaceae	DPTCSPAGNFWCNSGRCCSIYNWCGSTAEYCAQGNC LAQCWP
Torreya taxifolia	Taxaceae	
Papuacedrus papuana	*Cupressaceae*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNC LAQCWP
Athrotaxis cupressoides	*Cupressaceae*	
Sequoiadendron giganteum Glaucum	*Cupressaceae*	
Cupressus dupreziana	*Cupressaceae*	
Calocedrus decurrens	*Cupressaceae*	
Platycladus orientalis	*Cupressaceae*	
Cunninghamia lanceolata	*Cupressaceae*	
Taiwania cryptomerioides	*Cupressaceae*	
Fokienia hodginsii	*Cupressaceae*	DPTCSPAGNFWCNSGRCCSQYNWCGSTSDYCASGN CLAQCWP
Cupressus dupreziana	*Cupressaceae*	
Austrotaxus spicata	Taxaceae	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNC LAQCWP
Pseudotaxus chienii	Taxaceae	
Taxus baccata	Taxaceae	
Retrophyllum minus	Podocarpaceae	DPTCSPAGRVYCNPGRCCSKFNWCGTTRANCGKGNC IAQCHP
Parasitaxus usta	Podocarpaceae	
Actinidia eriantha	Actinidiaceae	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDC QSQCG
Actinidia chinensis	Actinidiaceae	
Actinidia deliciosa	Actinidiaceae	
Populus deltoids	Salicaceae	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSK NCQSNCKG
Populus nigra	Salicaceae	
Populus tremula	Salicaceae	
Populus trichocarpa	Salicaceae	
Pinus banksiana	Pinaceae	
Pinus jeffreyi	Pinaceae	
Pinus ponderosa	Pinaceae	

Pinus radiata	Pinaceae	EQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQDC QSQCG
Manihot esculenta	Euphorbiaceae	EQCGRQAGGALCPGGLCCSKFGWCGNTPDYCGADC QSQCSAG
Manihot grahamii	Euphorbiaceae	
Passiflora caerulea	Passifloraceae	EQCGRQAGGALCPGGLCCSQFGWCGSTNDYCGPGC QSQCG
Prosopis alba	Fabaceae	
Cajanus cajan	Fabaceae	EQCGRQAGGALCPGGLCCSQFGWCGSTPDYCGKDC QSQCG
Cajanus platycarpus	Fabaceae	
Canella winterana	*Canellaceae*	EQCGRQAGGALCPGGLCCSQFGWCGTTNDYCGPGC QSQCG
Pueraria montana var. lobata	Fabaceae	
Salix sachalinensis	Salicaceae	EQCGRQAGGALCPGGQCCSKFGWCGTTDAYCSGDC QSQCG
Salix purpurea	Salicaceae	
Castanea crenata	Fagaceae	EQCGRQAGGALCPGGQCCSQFGWCGTTADYCNTGC QSQCG
Quercus suber	Fagaceae	
Arabidopsis lyrata	Brassicaceae	EQCGRQAGGALCPNGLCCSEFGWCGDTEPYCKQPG CQSQCTPG
Boechera fecunda	Brassicaceae	
Gleditsia triacanthos	Fabaceae	EQCGRQAGGALCPNGLCCSQYGWCGTTSDYCGTNC QSQC
Gleditsia sinensis	Fabaceae	
Leucobryum albidum	Leucobryaceae	EQCGRQAGGATCPGGLCCSQYGYCGTTPDYCGGGC QSNCG
Leucobryum glaucum	Leucobryaceae	
Fragaria vesca	Rosaceae	EQCGRQAGGATCPNGLCCSEYGWCGTTPDYCATGC QSQCTP
Fragaria vesca	Rosaceae	
Lathyrus sativus	Fabaceae	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKN CQSNCHG
Pisum sativum	Fabaceae	
Glycyrrhiza glabra	Fabaceae	EQCGRQAGGAVCPGGLCCSKFGWCGSSPDYCGDGC QSQCG
Glycyrrhiza lepidota	Fabaceae	
Glycyrrhiza uralensis	Fabaceae	

Bursera simaruba	Burseraceae	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDH NCQSNCKDSG
Vitex agnus-castus	Lamiaceae	
Galega orientalis	Fabaceae	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKN CQSNCQG
Glycyrrhiza glabra	Fabaceae	
Glycine max	Fabaceae	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSK NCQSNCWG
Widdringtonia cedarbergensis	Cupressaceae	
Prunus persica	Rosaceae	EQCGRQAGNAVCPNGLCCSQHGWCGTTADYCATGC QSQCTS
Prunus dulcis	Rosaceae	
Cucumis melo	Cucurbitaceae	EQCGRQANGALCPNNLCCSQFGFCGDTDDYCKNGCQ SQCRG
Cucumis sativus	Cucurbitaceae	
Quercus suber	Fagaceae	EQCGRQANGAVCPNGLCCSQHGWCGTTNDHCGNGC QSQCKP
Quercus lobata	Fagaceae	
Nicotiana tabacum	Solanaceae	EQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGN CQSQCPG
Nicotiana tomentosiformis	Solanaceae	
Salix dasyclados	Salicaceae	EQCGSQAGGQTCPNNLCCSQFGWCGDTDDYCSPSK NCQSNCKG
Salix sachalinensis	Salicaceae	
Salix viminalis	Salicaceae	
Salix eriocephala	Salicaceae	EQCGSQAGGQTCPNNLCCSQYGWCGDTDDYCSPSK NCQSNCKG
Salix fargesii	Salicaceae	
Isoetes sp	Isoetaceae	EQCGSQAGGVLCANNLCCSQYGWCGSTSAYCGTGC QSQCG
Treubia lacunose	Treubiaceae	
Glycine max	Fabaceae	EQCGTQAGGALCPNRLCCSKFGWCGDTDSYCGEGC QSQCKSATP
Glycine soja	Fabaceae	
Arachis hypogaea	Fabaceae	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAG CQSNCWG
Arachis ipaensis	Fabaceae	
Nepenthes khasiana	Nepenthaceae	FQCGQQAGGALCHSGLCCSQWGWCGTTSDYCGNGC QSQCG
Nepenthes sp.	Nepenthaceae	

Daphniphyllum macropodum	Daphniphyllaceae	QEQCGSQAGGALCPGGLCCSQYGYCGTTSAYCGEGCQSQC
Tamarix chinensis	Tamaricaceae	
Glyptostrobus pensilis	*Cupressaceae*	QNCGRQAGNAPCSNGNCCSQYGFCGNTPEHCSPANNCQSQCTG
Lactuca sativa	Asteraceae	
Triticum turgidum	Poaceae	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVA
Triticum turgidum subsp. Durum	Poaceae	
Aegilops tauschii subsp. tauschii	Poaceae	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVASSG
Triticum aestivum	Poaceae	
Juglans nigra	Juglandaceae	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS
Phellodendron amurense	*Rutaceae*	
Vaccinium angustifolium	Ericaceae	QQCGRQAGGKLCPGNQCCSQWGYCGTTDDYCLSSNNCQSNCKP
Vaccinium myrtilloides	Ericaceae	
Lycium sp.	Solanaceae	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG
Tellima breviflora	Saxifragaceae	
Brassica napus	Brassicaceae	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
Brassica oleracea	Brassicaceae	
Brassica oleracea var. viridis	Brassicaceae	
Brassica rapa subsp. pekinensis	Brassicaceae	
Citrus aurantium	Rutaceae	QQCGRQAGGRTCANNLCCSQYGYCGSTDEYCSPSKNCQSNCRPG
Citrus limonia	Rutaceae	
Phaseolus angustissimus	Fabaceae	QQCGRQAGGSRCSGNLCCSQFGWCGNTAEYCSPSQNCQSNCWG
Phaseolus vulgaris	Fabaceae	

Brassica napus	Brassicaceae	QQCGRQAQGALCPNGLCCSEYGWCGTTEAYCGRGC
Brassica rapa	Brassicaceae	QSQCTP
Brassica napus	Brassicaceae	QQCGRQAQGALCPNGLCCSQYGWCGTTEAYCGRGC
Brassica oleracea var. oleracea	Brassicaceae	QSQCTP
Lycopersicon cheesmanii	Solanaceae	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQ
Solanum lasiophyllum	Solanaceae	GCQSQCRG
Solanum lycopersicum	Solanaceae	
Capsicum annuum	Solanaceae	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQ
Nicotiana benthamiana	Solanaceae	GCQSQCSG
Solanum dulcamara	Solanaceae	
Beta vulgaris	Amaranthaceae	QQCGSQAGGKKCPNNLCCSPWGYCGTGPDYCGNGC
Panicum virgatum	Poaceae	QSGPCSG
Eutrema halophilum	Brassicaceae	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNN
Eutrema salsugineum	Brassicaceae	CQSNCWG
Quercus petraea	Fagaceae	QQCGWQVGGKTCSNNLCCSQYGYCGTTDDYCSPSK
Quercus robur	Fagaceae	NCQSNCQG
Drosera capensis	*Droseraceae*	VQCGSEVGGALCPNGLCCSKYGYCGTTSAYCGPGCQ
Drosera rotundifolia	*Droseraceae*	SQCG

8C-non-chitin-binding hevein-like peptides

Plant Species	Plant Family	Mature Sequence
Panax ginseng	*Araliaceae*	CIPGGGFCMFEPLSCCVNCGCILVPGVCYCG
Panax quinquefolius	*Araliaceae*	
Populus tremuloides	Salicaceae	CISSGGFCFTQPMNCCGNCGCLYPLGICYGSDC
Populus euphratica	Salicaceae	
Populus deltoides	Salicaceae	CISSGGWCFTQPKNCCGNCGCLYPIGICFGSDC
Triticum aestivum	Poaceae	
Mollugo nudicaulis	Molluginaceae	CKSAGEWCGFSWTDCCNSCGCLAGFCYGTSC
Panax notoginseng	*Araliaceae*	
Panax ginseng	*Araliaceae*	CKSAGTWCGFDPHGCCGSCGCLVGFCYGVSC
Panax quinquefolius	*Araliaceae*	
Panax ginseng	*Araliaceae*	CKSGGAWCGFDPHGCCGNCGCLVGFCYGTGC
Panax notoginseng	*Araliaceae*	
Panax quinquefolius	*Araliaceae*	CKSGGSWCGFDPHGCCGNCGCLVGFCYGTGC
Panax ginseng	*Araliaceae*	
Panax ginseng	*Araliaceae*	CKSSGAWCGFDPHGCCGNCGCLVGFCYGTDC
Panax quinquefolius	*Araliaceae*	
Blasia sp.	Blasiaceae	CLKNGEFCWGDPSGCCGNCGCLIIPGVCYGTGC
Panax ginseng	*Araliaceae*	
Bazzania trilobata	Lepidoziaceae	CLNGGGYCGSFTREACCYNCVCMMAFCVCG
Microtea debilis	Phytolaccaceae	
Triticum aestivum	Poaceae	CLPAGGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE
Oryza sativa	Poaceae	
Oryza sativa	Poaceae	CLPAGGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE
Oryza sativa	Poaceae	
Eragrostis curvula	Poaceae	
Theobroma cacao	Malvaceae	

Elymus wawawaiensis	Poaceae	CLPSGGFCMFRPKDCCGNCGCLYPIGVCYGSRCEE
Oryza sativa	Poaceae	
Panax ginseng	*Araliaceae*	GCKSSGAWCGFDPHGCCGNCGCLVGFCYGTGC
Panax quinquefolius	*Araliaceae*	
Coffea Arabica	Rubiaceae	QEGECSPLGEPCAGNPWGCCPGCICIWQLTDRCVGNC
Coffea Canephora	Rubiaceae	

10C-chitin-binding hevein-like peptides

Plant Species	Plant Family	Mature Sequence
Zea diploperennis	Poaceae	EQCGSQAGGALCPNCLCCSQFGWCGSTSDYCGSGCQSQCSGSCG
Zea mays	Poaceae	
Citrus sunki	Rutaceae	GNCGGGVVCPGGECCSRFGWCGLTTAYCCEGCQSNCNQVVCG
Citrus unshiu	Rutaceae	
Citrus clementina	Rutaceae	GNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
Citrus sinensis	Rutaceae	
Citrus reticulata	Rutaceae	QNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
Citrus sinensis	Rutaceae	
Citrus clementina	Rutaceae	GNCGSGVVCPGGECCSRFGWCGLTIDHCCDGCQSNCNQVVCG
Citrus reticulata	Rutaceae	
Citrus sunki	Rutaceae	

Appendix B: Accession numbers and mature sequences of putative 6C-chitin-binding hevein-like peptides

Accession No.	Plant Species	Mature Sequence
VKOE-2007081	*Aerva lanata*	TGECKHGKCPNGLCCSQYGYCGSGPAYCG
PDQH-2053815	*Aerva lanata*	TGECKHGKCPNGLCCSQYGYCGSGPAYCGGA
MEAH-2007358	*Aerva persica*	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
GBCQ-2018132	*Aerva persica*	AGECRNGRCPSGLCCSRYGYCGSGPAYCGX
KTQI-2011188	*Alternanthera brasileana*	TGECKHGRCPSGLCCSQYGYCGTGPAYCG
WGUG-2017653	*Alternanthera caracasana*	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
ALK00913.1	*Alternanthera sessilis*	AGECNHGRCPSGLCCSQYGYCGTGPRYCG
OYST-2062883	*Alternanthera sessilis*	APGQCKHGRCPSGLCCSQYGYCGTGPAYCG
ALK00912.1	*Alternanthera sessilis*	APGQCNHGRCPSGLCCSQYGYCGTGPAYCGG
LUNL-2060100	*Alternanthera sessilis*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
LUNL-2061656	*Alternanthera sessilis*	TGECVNGLCPRGLCCSQYGYCGKGKAYCGS
EF066731.1	*Amaranthus albus*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
EF066734.1	*Amaranthus blitum*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
P27275-1	*Amaranthus caudatus*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
EF066735.1	*Amaranthus hybridus*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
EF066733.1	*Amaranthus hybridus subsp. cruentus*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
Q71U16-1	*Amaranthus hypochondriacus*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
WMLW-2009398	*Amaranthus retroflexus*	DGECVNGSCPNGLCCSQFGYCGSGPAYCG
WMLW-2061382	*Amaranthus retroflexus*	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
EF066738.1	*Amaranthus tricolor*	VGECVRGRCPSGMCCSQFGYCGKGPKYCG
AKTA-2039893	*Atriplex hortensis*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
MUCT-2027217	*Atriplex prostrata*	AGECGRGRSCPRGMCCSRFGYCGRGPDYCG
MUCT-2036997	*Atriplex prostrata*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
PDXY-2042561	*Atriplex rosea*	QSGQCTDGQCPNGLCCSQYGYCGSGPAYCG
LBZM-2049046	*Atriplex rosea*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
FVXD-2056435	*Beta maritima*	SGECNMYGRCPPGYCCSKFGYCGVGRAYCG
DV501762.1	*Beta vulgaris*	ECNMYGRCPPGYCCSKFGYCGVGRAYCG
DN911519.1	*Beta vulgaris*	ECNMYGRCPPGYCCSRYGYCGVGRAYCG
DN911732.1	*Beta vulgaris*	SGECNMYGRCPPGYCCSKYGYCGVGRAYCG
DN911712.1	*Beta vulgaris*	SGECNMYGRCPPGYCCSRYGYCGVGRAYCG
DN911508.1	*Beta vulgaris*	TSMASGECNMYGRCPPGYCCSKFGYCGVGRAYCG

SMMC-2134691	*Chenopodium quinoa*	CKSKGDCPKGMCCSKYGYCGSGPDYCG
SMMC-2093191	*Chenopodium quinoa*	CVNGACPNGLCCSRFGYCGSGPAYCG
CN782031.1	*Chenopodium quinoa*	GGECDMYGRCPAGYCCSKYGYCGVGPAYCG
CN782027.1	*Chenopodium quinoa*	SGECDMYGRCPAGYCCSKYGYCGVGPAYCG
CN782024.1	*Chenopodium quinoa*	SGECNMYGRCPAGYCCSKYGYCGVGPAYCG
CN782004.1	*Chenopodium quinoa*	SGECNMYGRCPAGYCCSKYGYCGVXPAYCG
LSLA-2010195	*Euphorbia mesembryanthemifolia*	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
QBGG-2040842	*Flaveria bidentis*	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
VJSJ-2098871	*Heliotropium calcicola*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
IPPG-2038781	*Heliotropium filiforme*	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
DIHD-2019944	*Heliotropium tenellum*	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
MDJK-2144026	*Heliotropium texanum*	AGECRNGRCPSGLCCSRYGYCGSGPAYCG
CHJJ-2124255	*Lejeuneaceae sp*	NCLVGGRCGLGQCCSAYGFCGSGAQYCG
FJ422811.1	*Malus hupehensis*	EQCGSQAGGAVCPNGLCCSQYGWCGTTSDYCATG
UNSW-2112382	*Mollugo nudicaulis*	SGECNMYGRCPAGYCCSVYGYCGLGPKYCG
TMAJ-2141759	*Neckera douglasii*	GNCLVGTRCGVGLCCSAFGYCGGGAQYCG
TMAJ-2134096	*Neckera douglasii*	NCLIGGCGAGACCSAYGYCGVGYCG
FE621809.1	*Panicum virgatum*	HQCGSQAGGKKCPNNMCCSPWGYCGTGPDYCG
KDCH-2010472	*Portulaca umbraticola*	EGECNNGQCPAGLCCSQYGYCGSGPDYCG
IGUH-2164233	*Schwetschkeopsis fabronia*	NCLVGGRCGVGQCCSAYGFCGSGAQFCG
VYGG-2035767	*Stachyurus praecox*	EQCGRQAGGALCPGGLCCSEYGWCGTTQDYCG
FN663151.1	*Stellaria media*	SGPNGQCGPGWGGCRGGLCCSQYGYCGSGPKYC
FN663152.1	*Stellaria media*	YDPNGKCGRQYGKCRAGQCCSQYGYCGSGSKYC
XM_030584468.1	*Syzygium oleosum*	QQCGSQAGEAKCVGGLCCSKFGYCGSGPAYCG

Appendix C: Accession numbers and mature sequences of putative 6C-non-chitin-binding hevein-like peptides

Accession No.	Plant Species	Mature Sequence
KU641478.1	*Achyranthes aspera*	CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS
EMIG-2051455	*Aerva lanata*	CIPKYNRCGTHYGVPCCEGTCTSEYYGICS
MEAH-2094198	*Aerva persica*	CIPKFNRCGTHYGVPCCEGTCTSEYYGICS
JX437183	*Allamanda cathartica*	CIAHYGKCDGIINQCCDPWLCTPPIIGFCL
JX437181	*Allamanda cathartica*	CIAHYGKCDGIINQCCDPWLCTPPIIGICI
JX437182	*Allamanda cathartica*	CRPYGTRCDGVINQCCDPYWCTPPIYGWCK
JX437184	*Allamanda cathartica*	CVSHYGKCDGIINQCCDPWLCTPPIIGFCL
KP318736.1	*Alstonia scholaris*	CRPYGYRCDGVINQCCDPYHCTPPLIGICL
KP318737.1	*Alstonia scholaris*	CRPYGYRCDGVINQCCDPYRCTPPLIGICL
KP318738.1	*Alstonia scholaris*	CVPQYGVCDGIINQCCDPYYCSPPIYGHCI
KTQI-2020846	*Alternanthera brasileana*	CVNGLCPRGLCCSQYGYCGKGKAXMPSWVVLQPVWLLWQRQGLLWLC
DEMH-2067715	*Alternanthera tenella*	CIPQWNRCASAQGDQCCPPFKCTSMYYGNCS
XSSD-2000083	*Amaranthus cruentus*	CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS
KU641480.1	*Amaranthus paniculatus*	CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS
WMLW-2064164	*Amaranthus retroflexus*	CIPKWNRCGPKMDGVLCCEPYTCTSDYYGNCS
BDIW-2050673	*Atriplex hortensis*	CIPKYNRCGPAMDGVHCCPPYTCTSQYYGNCS
EPVF-2036269	*Atriplex prostrata*	CIPKSNRCGPAMDGVQCCPPYTCTSQYYGSCS
FVXD-2056545	*Beta vulgaris*	CIPQYNRCGPAMDGVPCCAPYTCTSDYFGRCA
CUTE-2055964	*Blutoparon vermiculare*	CIPRGNRCGGDGVPCCEGTCTSLFYGICS
JAFJ-2143998	*Bougainvillea spectabilis*	CIPKANRCGPAMDGVPCCEPYTCTSLHFGICS
KU641479.1	*Celosia argentea*	CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS
SMMC-2141945	*Chenopodium quinoa*	CIPKANRCGPAINKVPCCSPYTCTSEYYGICS
XR_002511380.1	*Chenopodium quinoa*	CKVAGEFCGGFAGFACCKGLTCVLEGSFPDAGGVCKRKY
OYLU-2093904	*Curcuma olena*	CSRGRRSRERRLPGCRTGRRSCCRTGSGSRCHSRRRRSRNSRIGCSR
GJNX-2003168	*Cypselea humifusum*	CIPQYNRCGPAMDGVLCCEPYTCTSQYYGSCA
FY392019.1	*Dianthus caryophyllus*	CIPQYNRCGPAMDGVPCCEPYTCSSQYYGSCS

JZ158018.1	*Gypsophila vaccaria*	CIPKYNRCGPAMDGVQCCEPYTCTSQYYGNCS
JZ156586.1	*Gypsophila vaccaria*	CKRSGELCGGIAGFRCCEGLKCVLDGDYADAGGVCKPRY
5GSF	*Hibiscus sabdariffa*	CIPRGGICLVALSGCCNSPGCIFGICA
6KLM	*Hibiscus sabdariffa*	CVSSGIVDACSECCEPDKCIIMLPTWPPRYVCSV
WGET-2037902	*Kochia scoparia*	CRRRGYRCDGIFNQCCDPYLCTPPLVGICM
5XBD	*Leuenbergeria bleo*	QCKPNGAKCTEISIPPCCSNFCLRYAGQKSGTCANR
TXMP-2033532	*Linum strictum*	CDSDGKCKKGLCCSEYNYCGSGPLYCGGNDDDENSPPPPPPGTK
HURS-2043309	*Mollugo pentaphylla*	CIPQYNRCGPKADGVPCCAPYTCTSDYYGSCR
MRKX-2087054	*Phytolacca bogotensis*	CIPKYNRCGPKMDRVPCCEPYTCSSDYYGICG
RXEN-2018078	*Polycarpaea repens*	CIPEGNRCGYDGVPCCGEYTCTSAFYGICS
CPLT-2043895	*Portulaca grandiflora*	CIPYMNRCARLIDVPCCEPYTCTSDYYGSCL
BYNZ-2074587	*Portulaca mauii*	CIPKYNRCGTHYGVPCCEGTCTSEYYGICS
INQX-2103958	*Salix acutifolia*	CLPSGGFCMFQPMNCCGDCGCLYPAG
SKNL-2009318	*Saponaria officianalis*	CIPQYNRCGPAMDGVPCCEPYTCSSQYYGSCS
OLES-2031190	*Schiedea membranacea*	CIPQYNRCGPAMDGVPCCEPYTCTSQYYGRCA
HZTS-2022862	*Sesuvium portulacastrum*	CIPQYNRCGPAMDGVLCCEPYTCTSQYYGRCA
ZJDK-2051982	*Sesuvium ventricosum*	CIPQYNRCGPAMDGVLCCEPYTCTSQYYGHCS
GH293984.1	*Silene latifolia*	CKHGGEFCGGFAGFECCQGLKCVLDGDYADAGGICYPIF
GH294699.1	*Silene latifolia*	CKHGGEFCGGFAGSECCQGLKCVLDGDYADAGGICYPIL
MFEA-2090077	*Stackhousia spathulata*	CLKNGEFCWGDPSGCCGNCGCLIIP
LKKX-2012767	*Talinum sp.*	CISYMNRCGPKMDGVQCCPPYKCTSEYYGSCIDRE
TNVE-2115229	*Trianthema portulacastrum*	CIPQYNRCGPAMDGVPCCEPYTCTSQYYGHCA
Wong et al., 2017	*Vaccaria hispanica*	CVPLSNRCGPKADGIPCCKPYTCSSDYYGHCN
KF679826.1	*Wrightia religiosa*	CAQKGEYCSVYLQCCDPYHCTQPVIGGICA
KF679827.1	*Wrightia religiosa*	CAQKGEYCSVYLQCCKPYQCTQPVIGGICA
KF679828.1	*Wrightia religiosa*	CAQKGEYCSVYLQCCKPYRCTQPVIGGICA
BERS-2062894	*Zaleya pentandra*	CIPQYNRCGPAMDGVLCCEPYTCTSQYYGHCA

Appendix D: Accession numbers and mature sequences of putative 8C-chitin-binding hevein-like peptides

Accession No.	Plant Species	Mature Sequence
VSRH-2006343	*Abies lasiocarpa*	EDCGQQAGGAVCPGGSCCSKYGWCGNTRDHCGEGCQSQCG
XM_027489992.1	*Abrus precatorius*	EQCGKQAGGAVCPNGLCCSEFGWCGNTESYCGKGCQSQCKPSG
XM_027500887.1	*Abrus precatorius*	EQCGSQAGGALCPGGLCCSKFGWCGTTPDYCGDGCQSQCSG
JQ677143.1	*Acacia koa*	EQCGRQAGGALCPGGLCCSQYGWCGQTYDYCGTGCQSQCG
AB839211.1	*Acacia mangium*	EQCGRQAGGALCPGGLCCSQFGWCGSTDDYCGRGCQSQCGG
MGBF-2001913	*Acacia pycnantha*	EQCGRQAGGALCPGGLCCSQFGWCGSTDEYCGRGCQSQCG
VFFP-2043955	*Acer negundo*	EQCGRQAGGALCPGGLCCSQFGWCGNTPDYCGTDCQSQCSG
HILW-2008099	*Acmopyle pancheri*	EQCGVQADETVCPGGLCCSQFGWCGTTNDYCGDGCQSQCSG
HILW-2099671	*Acmopyle pancheri*	EQCGRQAGGVVCPGGLCCSQWGWCGNTPDHCGKDCQSQCG
FG402028.1	*Actinidia arguta*	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGADCQSQCG
QAUE-2030996	*Actinidia chinensis*	QGEQCGSQAGGKLCPGGLCCSRFGFCGTTSDYCGTGCQSQCG
FG454416.1	*Actinidia chinensis*	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCG
FG454759.1	*Actinidia chinensis*	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCGG
FG478012.1	*Actinidia deliciosa*	QFQCGRQAGGRKCPKKICCSKWGYCGTTPDFCLPSNGCQSNCKK
FG442870.1	*Actinidia deliciosa*	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCG
FG507620.1	*Actinidia eriantha*	QQCGRQAGGRLCDNNLCCSQWGYCGNTDEYCLPSNNCQSNCKG
FG483112.1	*Actinidia eriantha*	ENCGRQAGGALCPGGQCCSKWGWCGTTPDHCGTDCQSQCG
FG481050.1	*Actinidia setosa*	QYQCGRQAGGRKCPNKVCCSKWGYCGTTPEFCLPSKGCQSNCKK
JQ029764.1	*Aegiceras corniculatum*	QQCGKDVGGKLCDGGLCCSQYGYCGSTKEYCGTNCQSQCG
XM_020330685.1	*Aegilops tauschii subsp. tauschii*	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVASSG
XM_020311190.1	*Aegilops tauschii subsp. tauschii*	CGKQADGMECPNNLCCSKDGYCGLGVDYCSTGAGCQSGAC
XM_020325269.1	*Aegilops tauschii subsp. tauschii*	QLQYCEKRCGKQADGMECPNNLCCSKDGYCGLGVDYCSAAAGCQSG
FO996961.1	*Aeschynomene indica*	EQCGKQANGATCPNGLCCSQYGWCGSTDSYCLPSQGCQSNCKG
HBHB-2038747	*Aesculus pavia*	EQCGSQAGGALCPGGQCCSQFGWCGTTPEYCSTGCQSQC
HBHB-2252018	*Aesculus pavia*	QQVLAEQCGNQAGGAVCPGGLCCSKFGFCGTTSEYCTNGCQSQCTPG
MIXZ-2008294	*Agathis robusta*	DPTCSPAGQVYCNSGRCCSKFNWCGSTAAYCQRPNCIAQCWPG
QICX-2149426	*Ailanthus altissima*	EQCGSQAGGAVCPGGLCCSKFGFCGSTTEYCTNGCQSQCTPG

UCNM-2047888	*Ajuga reptans*	QNCGCASNECCSQWGYCGTTNDYCGRGCQSGPCFAP
HYZL-2141262	*Akania lucens*	EQCGRQAGGAQCPNGLCCSQHGWCGTTPEYCGAGSCQSQCSP
FWBF-2012524	*Alangium chinense*	QQCGRQAGGRLCPGNICCSQWGYCGTTDEYCLPSQNCQSNCKPSG
DVXD-2071211	*Allionia incarnata*	EQCGRQAGGALCPNNLCCSQYGWCGNTDPYCGKGCQSRC
CF443312.1	*Allium cepa*	EQCGRQAGGALCPGGLCCSKFGYCGTTQDYCGDGCQSQCG
KBXS-2000239	*Allium commutatum*	QQCGGQAGGALCSNGLCCSQYGYCGSTSDYCGPGCQSQCG
FQ341594.1	*Alnus glutinosa*	QQCGSQAGGKTCADNLCCSQYGYCGTTDDYCLPSNNCQSNCKSSG
URDJ-2008416	*Amborella trichopoda*	QQCGRDRGGAVCAGGLCCSKWGFCGSTNDYCGEGCQSQCTG
URDJ-2001732	*Amborella trichopoda*	QNCGKERGGAVCAGGLCCSRWGFCGSTDEHCGEGCQSQCKP
IAJW-2126957	*Amentotaxus argotaenia*	QCGRQDRGRRCPRGLCCSQFGYCGSTPAYCCAGCQSQ
IAJW-2003902	*Amentotaxus argotaenia*	DPTCSPAGNFFCNNGRCCSRFNWCGTGPSYCGRGNCIAQCP
IAJW-2018453	*Amentotaxus argotaenia*	DPTCSPAGNFWCDTGRCCSIYNWCGSTAEYCAQGNCLAQCWSS
IAJW-2142165	*Amentotaxus argotaenia*	DPTCSPAGRFFCNSGRCCSKFNWCGSTSDYCARPNCVAQCWPG
JZ471525.1	*Ammopiptanthus mongolicus*	EQCGSQAGGKKCPNNLCCSQYGYCGDSEDYCSPSKNCQSNCWG
JZ389716.1	*Ammopiptanthus mongolicus*	EQCGSQAGGKKCPNNLCCSQYGYCGNSEDYCSPSKNCQSNCWG
EL680321.1	*Ammopiptanthus mongolicus*	EQCGRQAGGAVCPNGLCCSQHGWCGTTDPYCGAGCQSQCKPG
XM_020224255.1	*Ananas comosus*	EQCGSQAGGATCPNGLCCSKFGYCGSTSAYCGDGNCQSQCG
ZUHO-2009007	*Anemone hupehensis*	ENCGRQVGGATCPGNHCCSQWGWCGTTDEHCLPSKNCQSNCRG
TQKZ-2059353	*Angelica archangelica*	EQCGKQAGNALCPNGMCCSQFGWCGTTPDYCSNCQSQCG
QMWB-2011191	*Anomodon attenuatus*	QQCGSQAGNRRCPNNLCCSQYGYCGDTDPYCGNGCQSQCHG
TOKV-2053682	*Aphanopetalum resinosum*	EECGRQAGGRTCSGNICCSQYGYCGTSDDYCSPSKNCQSNCAG
TOKV-2056941	*Aphanopetalum resinosum*	EQCGNQAGGATCPGGQCCSKFGWCGTADAYCGDGCQSQCKKP
NXOH-2003366	*Apios americana*	EQCGRQAGGALCPNGLCCSKFGWCGSTDSYCGEGCQSQC
NXOH-2051850	*Apios americana*	EQCGRQAGGALCPGGLCCSQFGWCGSTNDYCGEGCQSQCPG
NXOH-2005138	*Apios americana*	EQCGRQAGGQTCPDNLCCSQYGWCGNTEDYCSPSKGCQSNCG
XM_002884861.2	*Arabidopsis lyrata subsp. lyrata*	EQCGRQAGGALCPNGLCCSEFGWCGDTEPYCKQPGCQSQCTPG
R90112.1	*Arabidopsis thaliana*	QQCGRQGGGRTCPGNICCSQYGYCGTTADYCSPTNNCQSNCWG
TZWR-2041647	*Arabis alpina*	QQCGSQAGGKTCSGNICCSQYGYCGTTADYCSPANKCQSNCWGSG

GW956564.1	*Arachis duranensis*	EQCGYQVGGAYCANSLCCSKYGWCGTTYDYCSPDAGCHSNCWG
GW946290.1	*Arachis duranensis*	EQCGYQVGGAYCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG
JK206082.1	*Arachis hypogaea*	EQCGKQANGSLCPNNLCCSKFGYCGNTNDYCGDGCQSQCKG
JK209988.1	*Arachis hypogaea*	EQCGRQAGGALCPNNYCCSQFGWCGETDDYCLASKGCQSQCRG
JK205505.1	*Arachis hypogaea*	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG
FS977976.1	*Arachis hypogaea*	QNCGCASGLCCSKYGYCGTGDAYCGDGCQEGPCYSS
JK203181.1	*Arachis hypogaea*	EQCGKQANGALCPNSLCCSQFGWCGTTDPYCLVGCQSQCKG
GW960675.1	*Arachis ipaensis*	EQCGSQGGGALCPNNLCCSRWGWCGTGDDYCLVGNGCQSQCKD
GW991589.1	*Arachis ipaensis*	EQCGYQVGGAHCANSLCCSKYGWCGTTYDYCSPDAGCQSNCWG
XTZO-2004677	*Araucaria rulei*	QNCGCASGLCCSKYGYCGTGAAYCGDGCQQGPCTS
XTZO-2062231	*Araucaria rulei*	EQCGRQAGGVVCPGGLCCSQWGWCGNTPDHCGTGCQSQCG
ACWS-2009835	*Araucaria sp*	EDCGRQAGGRVCPGGVCCSQWGWCGNTADHCGTGCQSQCG
ODDO-2102205	*Ardisia humilis*	QQCGKAVGGKVCAGGLCCSQYGYCGSTSDYCGTNCQSQCG
BFMT-2017722	*Argemone mexicana*	QQCGRQAGGQTCPGNLCCSQYGYCGDTDDYCLPSNNCQSKCKG
EY087350.1	*Artemisia annua*	QDCGTQGGNRPCTNGNCCSQWGFCGNTPDHCSPANNCQSQCSG
DSUV-2009216	*Asclepias curassavica*	KQCGDQAGGAICRGGECCSQWGWCGTTEEYCGHGCQSNC
YADI-2015264	*Asclepias syriaca*	EQCGTQAGNALCPNGLCCSRFGWCGSTDEYCKDGCQSQCTG
FGRF-2048625	*Asparagus densiflorus*	EQCGSQAGGAVCPNGLCCSKFGYCGTTSAYCCDGCQSQGTP
PSKY-2005232	*Asplenium nidus*	EQCGRQAGGALCPNRLCCSRFGYCGDTCPYCGPAC
KJZG-2002211	*Asplenium platyneuron*	EQCGRQAGGALCPNNLCCSYYGFCGNAPNYCGQNCQSQCG
CKKR-2060572	*Astilbe chinensis*	QQCGRQAGGRTCANNLCCSQFGYCGSTAEYCSTNCQSNCNSG
HJMP-2014725	*Astragalus membranaceus*	EQCGRQGGGQTCPNNQCCSQYGWCGDTDEYCSPSKNCQSNCKG
HJMP-2015212	*Astragalus membranaceus*	EQCGSQAGGALCPGGLCCSRHGWCGSTGDYCGEGCQSQCG
MYMP-2058807	*Astragalus propinquus*	EQCGRQGGGQKCPNNLCCSQYGWCGDTDEYCSPSKNCQSNCKG
MYMP-2009938	*Astragalus propinquus*	EQCGSQAGGALCPGGLCCSKHGWCGSNGDYCGEGCQSQCG
XIRK-2000806	*Athrotaxis cupressoides*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
XIRK-2000807	*Athrotaxis cupressoides*	DPTCSPSGNFWCNNGRCCSIYNWCGSGAAYCAKGNCLAQCWP
BOLZ-2159794	*Atropa belladonna*	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSRCSG
GIPR-2013846	*Aucuba japonica*	EQCGTQAGGALCPGGLCCSKFGWCGNGGDYCSTGCQSQCG
YYPE-2002936	*Austrocedrus chilensis*	DPTCSPAGNFLCNTGRCCSQYNWCGSTSDYCASGNCLAQCWP

YYPE-2029866	*Austrocedrus chilensis*	DPTCSPAGNFWCNTGRCCSIYNWCGSTSDYCASGNCLAQCWS
YYPE-2002937	*Austrocedrus chilensis*	DPTCSPSGNFFCNNGRCCSVYNWCGTGNAYCGKGNCLAQCSP
BTTS-2072841	*Austrotaxus spicata*	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP
BTTS-2007338	*Austrotaxus spicata*	DPTCSPRGNFFCNSGRCCSKFNWCGSGAQYCAKGNCIAQCWPG
BTTS-2072606	*Austrotaxus spicata*	ERCGLQAGGAVCHGGLCCSRSGWCGNTPDYCNGCQSQCREEP
BTTS-2032679	*Austrotaxus spicata*	QCGKDADGALCSSGLCCSEWGWCGNTPDHCGAGCQSQCG
UVDC-2040487	*Azadirachta indica*	EQCGSQAGGALCPGGLCCSKFGYCGNTPAYCTNGCQSQCTSG
OFTV-2021147	*Barbilophozia barbata*	QGQCQGNNLCTDRTLCCSQYNFCGSNSSYCGSGCKGGPC
CTYH-2105366	*Basella alba*	EQCGSQAGGALCPGGLCCSQYGYCGTTSAYCGEGCQSQCT
JETM-2091845	*Bauhinia tomentosa*	EQCGKQAGGQICPNNLCCSQYGWCGSTDDYCNTGKGCQSNCAG
JETM-2093297	*Bauhinia tomentosa*	EQCGSQAGGALCPGGSCCSKYGWCGGTSEYCGDGCQSQCG
WZYK-2087147	*Bazzania trilobata*	QQCGSQAGGRTCPNGNCCSQWGWCGNTDDYCAGGCQSQCG
HAEU-2010316	*Berberidopsis beckleri*	EQCGRQAGGAVCPGGQCCSQFGWCGTTPEYCTDGCQSQCSSTG
JG622500.1	*Berberis nervosa*	QGGCGSNAGGATCAGGLCCSTSGFCGSTEAYCGPGNCQSQCP
FE620753.1	*Beta vulgaris*	QQCGSQAGGKKCPNNLCCSPWGYCGTGPDYCGNGCQSGPCSG
CWZU-2006082	*Betula pendula*	EQCGKQAGGALCPGGQCCSQFGWCGTTADYCTNGCQSQCTG
VNMY-2014788	*Bischofia javanica*	EQCGRQAGGALCPGGQCCSKFGWCGKTNDYCGSGCQSQCSG
AF135137.1	*Boechera fecunda*	EQCGRQAGGALCPNGLCCSEFGWCGDTEPYCKQPGCQSQCTPG
ACFP-2028246	*Boehmeria nivea*	EQCGRQAGFAICPNGLCCSQYGWCGDTYDYCSPDKNCQSNCWG
FCCA-2043010	*Boswellia sacra*	EQCGSQAGGALCPGGLCCSKFGWCGSTSDYCSSGCQSQC
BEGM-2017838	*Botrypus virginianus*	EQCGKQAGGKVCPNNLCCSQYGYCGSTTAYCGKGCQSQCSG
OOVX-2042183	*Boykinia jamesii*	QQCGKQAGGRTCANNLCCSQYGYCGSTAEYCFTNCQSNCNSG
DW153474.1	*Brassica napus*	QNCGCAPNLCCSQFGYCGSTDAYCGTGCRSGPCRSPG
XM_022700474.1	*Brassica napus*	QNCGCTSGLCCSQYGYCGTGPDYCGSGCRSGPCTG
EE445484.1	*Brassica napus*	QNCGKTGCKGNMCCSRWGYCGTTNAYCGTGCQSGPCKS
FG577475.1	*Brassica napus*	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
XM_013792769.2	*Brassica napus*	QQCGRQAQGALCPNGLCCSEYGWCGTTEAYCGRGCQSQCTP
XM_013838100.2	*Brassica napus*	QQCGRQAQGALCPNGLCCSQYGWCGTTEAYCGRGCQSQCTP
EV219966.1	*Brassica napus*	EQCGRQAGGALCPNGLCCSEFGWCGDTESYCKQPGCQSQCG
GR437477.1	*Brassica napus*	QNCGTYGCAGNLCCSRYGYCGTTLDYCGTGCRSGPCSSG

EX114407.1	*Brassica oleracea*	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
AM388980.1	*Brassica oleracea var. alboglabra*	EQCGRQAGGALCPNGLCCSEFGWCGDTEPYCKQPGCQSQCG
XM_013770397.1	*Brassica oleracea var. oleracea*	QQCGRQAQGALCPNGLCCSQYGWCGTTEAYCGRGCQSQCTP
DK569514.1	*Brassica oleracea var. viridis*	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
XM_009136986.2	*Brassica rapa*	QQCGRQAQGALCPNGLCCSEYGWCGTTEAYCGRGCQSQCTP
EX075813.1	*Brassica rapa subsp. pekinensis*	QQCGRQAGGQTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
EX030278.1	*Brassica rapa subsp. pekinensis*	QQCGRQAGGQTCAGNXCCSQYGYCGTTADYCSPDNNCQSNCWG
AIOU-2068307	*Brugmansia sanguinea*	QQCGSQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
JMXW-2009405	*Bryum argenteum*	QQCGSQAGGQNCPGGLCCSQYGYCGTTAEYCGGGCQSQCG
JSZD-2010254	*Bursera simaruba*	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDSG
GDZS-2115154	*Byblis gigantea*	QNCNCPPGWCCSQWGYCGTTEPYCGGGCQEGPCYDKQQP
VTLJ-2000125	*Caiophora chuquitensis*	EQCGRQAGGAVCPGGLCCSQFGWCGSTTPDYCGAGCQSQCG
VTLJ-2010251	*Caiophora chuquitensis*	QQCGRQGGGRTCQNNLCCSQYGYCGSTDDYCLPSNGCQSNCRPSG
XM_020382083.2	*Cajanus cajan*	EQCGRQAGGALCPGGLCCSQFGWCGSTPDYCGKDCQSQCG
GW350060.1	*Cajanus cajan*	EQCGSQAGGKTCPNNLCCSQYGWCGNAEDYCSPSKNCQSNCWG
XM_020368743.2	*Cajanus cajan*	QNCGCAEGLCCSQYGYCGTGDAYCGTGCQQGPCYNP
MK606416.1	*Cajanus platycarpus*	EQCGRQAGGALCPGGLCCSQFGWCGSTPDYCGKDCQSQCG
IFLI-2000417	*Callitris gracilis*	DPTCSPAGGFWCNIGRCCSIYNWCGSTWEYCAPGNCLAQCWP
RMMV-2002319	*Callitris macleayana*	DPTCSPAGNFLCNSGRCCSIYNWCGDTSDYCAAGNCLAQC
RMMV-2004820	*Callitris macleayana*	DPTCSPAGSFLCNSGRCCSKYNWCGSTSDYCASGSCLAQCWP
RMMV-2006113	*Callitris macleayana*	EPTCSPAGRIYCSPGRCCSKFNWCGSSRAYCGPNCIAQCHSP
FRPM-2002883	*Calocedrus decurrens*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
FRPM-2008761	*Calocedrus decurrens*	ENCGSQGGGKACSGGECCSKWGWCGTTDDHCLPDRGCQSNCHG
XM_010419957.1	*Camelina sativa*	QQCGRQAGGRTCPGNICCSQYGYCGTTPEYCSPANNCQSNCGG
MG720758.1	*Camellia fraterna*	QQCGSQAGGALCPGGLCCSEHGWCGTTDAYCGPGCQSQCSP
XM_028253825.1	*Camellia sinensis*	QNCGCAAGLCCSQYGYCGTGNAYCGAGCQSGPCTG
FS957358.1	*Camellia sinensis*	QQCGSQAGGQTCPNNLCCSQYGYCGTTDDYCSPSKGCQSNCQTG
IFCJ-2113840	*Canella winterana*	EQCGRQAGGALCPGGLCCSQFGWCGTTNDYCGPGCQSQCG

DDEV-2107832	*Canella winterana*	EQCGSQAGGALCPNGLCCSKFGWCGSTDDYCMNGCQSQCKSG
PMRP-2007200	*Cannabis sativa*	EQCGRQAGGALCPGGLCCSEFGWCGDTSEYCNDKCQSQCG
PDIE-2001895	*Cannabis sativa*	EQCGRQAGGVKCPNGLCCSQFGWCGDTFDYCSPQAGCQSNCWG
BJSW-2013786	*Cannabis sativa*	EQCGRQAGGVKCPNGLCCSQHGWCGNTFDYCSPQAGCQSNCWG
GR221643.1	*Cannabis sativa*	DHQCGPNLGNPPCGDGRCCSIHNFCGSVSSYCRGGNCRYQCWFA
XM_006298557.1	*Capsella rubella*	QQCGRQAGGRTCPGNICCSQYGYCGTTPEYCSPANNCQSNCG
EL814425.1	*Capsicum annuum*	QNCGRQAGGRVCANRLCCSQFGFCGRTREYCGPGCQSNCRR
BM066673.1	*Capsicum annuum*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
GD054660.1	*Capsicum annuum*	QNCGRQAGGRFCANRLCCSQFGFCGTTREYCGPGCQSNCR
GD054975.1	*Capsicum annuum*	QNCGRQAGGRVCANRLCCSQFGFCGTTREYCGAGCQSNCR
EX282008.1	*Carica papaya*	QNCGRQAGGRTCANNLCCSQFGFCGTTDDHCSPAKNCQSNCRG
VYLQ-2036312	*Cassytha filiformis*	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCPPG
NHUA-2075831	*Castanea crenata*	EQCGRQAGGAVCPNGLCCSQYGWCGNTNDYCGNGCQSQCSSG
NHUA-2065072	*Castanea crenata*	EQCGRQAGGALCPGGQCCSQFGWCGTTADYCNTGCQSQCG
GO920272.1	*Castanea mollissima*	QQCGWQVGGKTCSDNLCCSQYGYCGTTDDYCSPSKNCQSNCQG
X95610.1	*Castanea sativa*	EQCGRQAGGAACANNLCCSQFGWCGNTAEYCGAGCQSQCS
HQ414236.1	*Casuarina equisetifolia*	EQCGRQAGGALCPGGQCCSQYGWCGSTSDYCSTGCQSQCG
LNER-2050327	*Casuarina glauca*	QQCGRQAGGRTCANNLCCSQYGYCGSTDDYCSPSKNCQSNCKSG
AVJK-2023427	*Cavendishia cuatrecasasii*	EQCGSQAGGKVCPGGLCCSKFGFCGSTAAYCTGGCQSNCG
GGEA-2062097	*Cedrus libani*	EDCGQQAGGALCPGGLCCSKWGWCGNTQAHCGQDCQSQCG
KYAD-2004733	*Celtis occidentalis*	EQCGRQAGGAICPNGLCCSQFGWCGNTAEHCSPQAGCQSNCWG
KYAD-2050740	*Celtis occidentalis*	EQCGRQAGGAICPNGLCCSQYGWCGNTAEYCSPQEGCQSNCWG
KYAD-2003083	*Celtis occidentalis*	EQCGRQAGGALCPGGLCCSQFGFCGTTADYCGGGCQSQCDRG
WEQK-2065627	*Centella asiatica*	EQCGRQAGNAICPNGLCCSQFGFCGSTPDYCTKNCQSQCSTG
YZVJ-2050978	*Cephalotus follicularis*	EQCGRQAGGALCPGGQCCSKYGWCGTTSDYCNSECQSQCG
NUZN-2058310	*Cercidiphyllum japonicum*	EQCGSQAGGAKCPGGQCCSKYGWCGTTSDYCGDGCQSQCTPTG
NUZN-2000485	*Cercidiphyllum japonicum*	QQCGRQAGGQTCANNLCCSEFGYCGSTDDYCSPSKNCQSNCKSG
NUZN-2052281	*Cercidiphyllum japonicum*	QEQCGSQAGGALCAGGLCCSQHGYCGSTADYCSTGCQSQC
RKFX-2045308	*Cercis canadensis*	EQCGRQAGGAGCPGGLCCSKFGWCGSAADYCGDGCQSQCG
RKFX-2042782	*Cercis canadensis*	EQCGRQGGGRTCPNNLCCSQYGWCGNGDDYCSPSKGCQSNCG

XFFT-2009888	*Cercocarpus ledifolius*	EQCGRQAGGAVCPGGQCCSQYGWCGSGGDYCSTGCQSQCG
YEPO-2078519	*cf. Physcomicromitrium sp*	EQCGRQAGGAICPGGLCCSRFGWCGSTEEYCGDGCQSQCG
AIGO-2010558	*Chamaecyparis lawsoniana*	QNCGSQVGGKVCSGGECCSKFGWCGTTDDHCLPDRGCQSNCG
FJ749130.1	*Chimonanthus praecox*	EDCGSQAGGALCPGGLCCSKFGFCGTTADYCGTGCQSQCSG
ZBVT-2060803	*Chrysobalanus icaco*	EQCGTQAGGALCPNNLCCSQWGWCGTTSDYCGTGCQSQCG
ORJE-2088434	*Cibotium glaucum*	EQCGSQAGFALCGGDYCCSKWGWCGSGDDYCGDGCQSGPCG
XM_004494647.2	*Cicer arietinum*	EQCGRQGGGKTCPNNLCCSQYGYCGDSDDYCSPSKNCQSNCHG
GR404613.1	*Cicer arietinum*	QCGRQGGGKTCPNNLCCSQYGYCGDSDDYCSPSKNCQSNCHG
BCGB-2043744	*Cinnamomum camphora*	QQCGRQAGGRVCDGGLCCSQWGYCGSTDEYCSPSQGCQSNCG
FC874612.1	*Citrus aurantium*	QQCGRQAGGRTCANNLCCSQYGYCGSTDEYCSPSKNCQSNCRPG
BB999737.1	*Citrus hassaku*	ENEQCGWQAGGALCPNENCCSKGGFCGITDAYCGEGCQSQCHHLSSFLD
BB999801.1	*Citrus hassaku*	ENEQCGWQAGGALCPNENCCSKGGFCGITDAYCGEGCQSQCHHP
EY871065.1	*Citrus limonia*	QQCGRQAGGRTCANNLCCSQYGYCGSTDEYCSPSKNCQSNCRPG
JG580697.1	*Clematis chinensis*	QCGRDAGGALCANNLCCSQFGFCGDTPPYCGDGCQSQCRPG
MBQU-2012677	*Cleome gynandra*	YQCGMQNGGALCPNGLCCSKWGWCGSTEAYCGEGCQSQC
UPZX-2042161	*Cleome viscosa*	DKPQCGRNADGKKCPNNLCCSAWGFCGSTEDYCSPEKHCQDNCWY
CSUV-2018711	*Cochlearea officinalis*	EQCGRQAGGALCPNGLCCSEFGWCGNTEPYCKQPGCQSQCTPG
SUAK-2039541	*Codariocalyx motorius*	EQCGRQAGGAVCPGGLCCSQFGWCGAFPDPAYCGDGCQSQCG
XR_003450397.1	*Coffea arabica*	EQCGKQAGNALCPNRLCCSQYGWCGSTADYCTNGCQSQCKNG
XM_027223444.1	*Coffea arabica*	EQCGRQAGNALCPNGLCCSQYGWCGSTPDYCSPSNNCQSQCSG
ILBQ-2047344	*Conocephalum conicum*	QQCGTQADNALCLNSLCCSKFGFCGVGDEYCGSDCQSNCNSTATTTG
FAMO-2092069	*Conopholis americana*	EQCGRQASGKKCDNNLCCSQYGYCGSTGEFCLLSKNCQSNCQS
CPOC-2001595	*Convolvulus arvensis*	QVRCGSQAGGALCPGGLCCSQYGFCGNKPEYCLPSSGCQSQCS
FY987602.1	*Coptis japonica var. dissecta*	QNCGRQAGGRTCPGNICCSQWGWCGTSNDHCTNNCQSNCRG
NNGU-202373	*Coriaria nepalensis*	QQCGRQAGGKTCANNLCCSQYGFCGTTDDYCSPSKQCQSNCKSG
BFJL-2003077	*Cornus florida*	QQCGTQAGGALCPNGLCCSQWGYCGTTLPYCGDGCQSQC
BFJL-2192413	*Cornus florida*	EQCGSQAGGAKCPGNLCCSKFGWCGNKPEYCGDGCQSQCG
ZGQD-2001293	*Corydalis linstowiana*	EQCGRQAGGAVCGGGLCCSQYGYCGSTLPYCGDGCQSQCTPG

KT943844.1	*Corylus avellana*	QQCGRQAGGKTCANNLCCSQYGYCGTTDDYCSPSKNCQSNCKTSG
KWGC-2013004	*Crassula perforata*	QICGKQAGGKTCPGNICCSQWGYCGTSDDYCSPSNNCQSNCKG
IHCQ-2017247	*Crossopetalum rhacoma*	EQCGTQAGGAVCPENLCCSKWGWCGNTIEYCGEGCQSQCG
IHCQ-2093561	*Crossopetalum rhacoma*	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCP
VVPY-2063263	*Croton tiglium*	EQCGRQAGGAKCPGGQCCSKFGWCGKAAAYCGDGCQSQCG
VVPY-2063371	*Croton tiglium*	EQCGRQANNARCPNGLCCSQFGWCGNTAEYCGAGNCQSQCPG
AB196451.1	*Cryptomeria japonica*	QNCGCNGLCCSQYGYCGSGEAYCGAGCKEGPCSSSSP
AK412731.1	*Cryptomeria japonica*	QNCGCNGLCCSQYGYCGSGEAYCGTGCKEGPCSSSSSP
AB096360.1	*Cryptomeria japonica*	ENCGSQAGGAVCPGGLCCSQYGWCGNTPDHCRVPGCQSQCG
CN845588.1	*Cucumis melo*	EQCGRQANGALCPNNLCCSQFGFCGDTDDYCKNGCQSQCRG
XM_008442677.2	*Cucumis melo*	EQCGRQANGALCPNNLCCSQYGWCGDTDAYCKDGCQSQCG
JG498505.1	*Cucumis melo*	EQCGRQANGALCPNNLCCSQYGWCGDTDAYCKDGCQSQCRG
XM_008442681.2	*Cucumis melo*	EQCGWQAGGAVCPNGLCCSQYGWCGTVNAYCGAGCQSQCRRR
XM_004134785.3	*Cucumis sativus*	EQCGRQANGALCPNNLCCSQFGFCGDTDDYCKNGCQSQCRG
XM_031887505.1	*Cucumis sativus*	EQCGRQTNGALCPNNLCCSQYGWCGNTDPYCKDGCQSQCG
XM_023121536.1	*Cucurbita maxima*	EQCGRQASGAVCPNGLCCSQFGWCGTTRAYCGHGCQSRCR
XM_023094025.1	*Cucurbita moschata*	EQCGRQASGAVCPNGLCCSQFGWCGTTRAYCGHGCQSRCRR
XM_023689213.1	*Cucurbita pepo subsp. pepo*	EQCGRQAGGVVCPNGLCCSQFGWCGTTRAYCGHGCQSRCRR
OUOI-2010547	*Cunninghamia lanceolata*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
TIUZ-2008371	*Cunonia capensis*	EQCGSQAGGALCPGGQCCSKYGWCGATGEYCGTGCQSQCG
QNGJ-2000609	*Cupressus dupreziana*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
QNGJ-2067951	*Cupressus dupreziana*	DPTCSPAGNFWCNSGRCCSQYNWCGSTSDYCASGNCLAQCWP
QNGJ-2000610	*Cupressus dupreziana*	DPTCSPKGNFFCNSGRCCSIYNWCGTGAAYCAKGNCLAQCSP
RDYY-2007556	*Cyanastrum cordifolium*	QQCGSQAGGALCPNGLCCSKYGYCGSTDPYCGAGNCQSQCG
XZUY-2048192	*Cycas micholitzii*	EQCGWQVGGAACDGDLCCSQYGFCGSTTEYCGQGCQSKCSG
XM_025140254.1	*Cynara cardunculus var. scolymus*	QNCKCARNLCCSKYGFCGTGPAYCGKDCRGGPCSQP
XM_025140291.1	*Cynara cardunculus var. scolymus*	QNCKCARNLCCSQYGYCGTGAAYCGKGCQSGPCSLR
PWSG-2008860	*Cyperus papyrus*	QQCGSQAGGATCANNLCCSKYGYCGSTSDYCGDGCQSQCG

FMWZ-2006066	*Dacrycarpus compactus*	ENCGRQAGGQICNGGECCSQYGWCGVTPDHCGTGCQSNCG
IZGN-2071461	*Dacrydium balansae*	EQCGRQAGGVVCPGGLCCSQWGWCGNTPDHCGTGCQSQCSG
FYTP-2003750	*Daphniphyllum macropodum*	QEQCGSQAGGALCPGGLCCSQYGYCGTTSAYCGEGCQSQC
JNVS-2004019	*Datura metel*	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCLPSQGCQSQCSG
XM_017380025.1	*Daucus carota subsp. sativus*	EQCGKQAGNALCPNGMCCSEFGWCGTTPDYCSDTCQSQCG
XM_017380027.1	*Daucus carota subsp. sativus*	EQCGKQAGNALCPNGMCCSQFGWCGTTAEYCTKCQSQCG
OQWW-2038289	*Davallia fejeensis*	QQCGSQAGGALCANNLCCSQFGYCGSTSDYCGTGCQSQC
BJKT-2003149	*Delosperma echinatum*	EQCGTQAGNALCPNGLCCSQYGWCGSSTLYCGTGCQSQC
OTAN-2099584	*Deutzia scabra*	EQCGRQAGGALCPGGACCSQFGWCGTTADYCTNGCQSQCTP
OTAN-2004858	*Deutzia scabra*	QQCGRQASGQTCANNLCCSQYGYCGTTDDYCLPSNNCQSNCKSG
OTAN-2096703	*Deutzia scabra*	QQQCGTQAGNQLCAGGLCCSQYGYCGNTANYCGSGCQSQCTAG
OTAN-2005526	*Deutzia scabra*	EQCGSQAGGKLCAGGLCCSKFGYCGNTAPYCTTGCQSQCSPG
QURC-2010804	*Dichroa febrifuga*	EQCGKQAGGALCPGGLCCSQFGWCGKTQEYCGPGCQSQCTP
NGTD-2096910	*Dicranum scoparium*	EQCGRQAGGALCPGGLCCSQFGFCGTTPEYCENGCQSNCG
EHNF-2086784	*Dillenia indica*	EQCGRQAGGALCPGGLCCSQYGWCGSASEYCATGCQSQCG
EHNF-2086972	*Dillenia indica*	QEQCGSQAGGNKCPGGLCCSKYGYCGSTADYCAQGNCQSQCG
EHNF-2006841	*Dillenia indica*	QQCGSQAGGKTCPNNLCCSQYGYCGSTDPYCLPSNNCQSNCKG
FJ040804.2	*Dimocarpus longan*	EQCGRQAGGAVCPNGLCCSQHGWCGSTTEYCGTGCQSQCR
KT223143.1	*Dionaea muscipula*	EQCGSQAGGALCPNGLCCSQYGWCGTTSAYCGAGCQSQCG
WLIC-2119047	*Dioon edule*	QQCGRQVSGRECDGNLCCSQFGYCGSTPEYCDPSQGCQSKCTGAG
KX871200.1	*Diospyros kaki*	AEQCGKQAGGALCPNGLCCSQYGYCGDTNDYCLKGCQSQCKSG
AWOI-2070580	*Diphyscium foliosum*	ENCGKTKCSDNLCCSQYGYCGSSSAYCGTGCQNGPC
GKCZ-2068869	*Diselma archeri*	DPTCSPAGNFWCDSGRCCSIYNWCGDTSDYCASGNCLAQCWP
GKCZ-2070649	*Diselma archeri*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCAPWNCLAQCWP
WYIG-2180644	*Dombeya burgessiae*	EQCGRQAGGALCPDNLCCSQYGWCGSTDPYCLPENNCQSNCRS
WYIG-2035881	*Dombeya burgessiae*	QQCGRQAGGRTCANNLCCSQHGYCGTTSEYCSPSQNCQSNCRG
GTSV-2008655	*Draba hispida*	QQCGRQAGGKTCPGNICCSQYGYCGTGADYCSPAKNCQSNCQGTG
GTSV-2005036	*Draba hispida*	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPANNCQSNCWG
LJQF-2059299	*Draba ossetica*	QQCGRQAGGQTCPGNICCSQYGYCGTTEDYCSPANNCQSNCQGTG
WKSU-2129629	*Drimys winteri*	EAQQCGRQAGGRTCAGNLCCSQYGYCGTADDYCLPSNNCQSNCKSSG
LC037409.1	*Drosera adelae*	QQCGYQAGGALCPNGLCCSKYGYCGTTSAYCGDGCQSQCG
MN481117.1	*Drosera binata*	EQCGYQAGGAVCPNGLCCSQYGYCGTTSAYCGSGCQSQCG

MK093978.1	*Drosera capensis*	VQCGSEVGGALCPNGLCCSKYGYCGTTSAYCGPGCQSQCG
KU516826.1	*Drosera rotundifolia*	VQCGSEVGGALCPNGLCCSKYGYCGTTSAYCGPGCQSQCG
XM_022907240.1	*Durio zibethinus*	EQCGRQAGGALCPGGLCCSQYGWCGNTDAYCKKENGCQSQCSG
XM_022907239.1	*Durio zibethinus*	EQCGSQAGGALCPGGLCCSQWGWCGSTDEYCKNGCQSQCNG
XM_022884198.1	*Durio zibethinus*	QQCGRQAGGALCANGLCCSQFGWCGNTPEYCGTGCQSQCSG
RBYC-2021348	*Elaeagnus pungens*	EQCGKQAGGKVCPGGQCCSQYGWCGTTDQYCKKGCQSQCG
RBYC-2077387	*Elaeagnus pungens*	ENCGKQAGGAVCPNGLCCSQYGWCGSTNEHCTTGCQSQCRKTASG
RBYC-2003243	*Elaeagnus pungens*	EQCGRQAGGALCPGGLCCSKHGWCGNTDAYCKDGCQSQCSG
AF061806.1	*Elaeagnus umbellata*	EQCGNQAGGKVCPGGQCCSQYGWCGTTDEYCKKGCQSKCG
XM_010943102.3	*Elaeis guineensis*	QQCGSQAGGKTCPNGLCCSQFGYCGSTSAYCGNGCQSQCSG
XM_010943097.3	*Elaeis guineensis*	QQCGSQAGGQTCPNGLCCSQYGYCGSTSAYCGDGCQSQCSG
EY411996.1	*Elaeis guineensis*	QCGSQAGGKTCPNGLCCSQFGYCGSTSAYCGNGCQSQCSG
THHD-2013091	*Elaeocarpus sylvestris*	EQCGRQAGGALCPGGQCCSKYGWCGTTSDYCNSDCQSQCG
THHD-2056890	*Elaeocarpus sylvestris*	EQCGSQAGGAVCPGGLCCSKFGYCGSTPDYCTNGCQSQCTP
THHD-2009061	*Elaeocarpus sylvestris*	QQCGKQGGGKKCANKLCCSQYGYCGTTPPYCSKSQGCQSNCKG
TIJL-2030394	*Eleusine coracana*	CPGCMCCSRFGWCGTTSDYCGQGCQSNC
PZAP-2011360	*Eleusine coracana*	CPNCLCCSKWGWCGNTADHCGAGCQSQC
KEFD-2060274	*Encalypta streptocarpa*	EQCGSQAGGKTCPGGLCCSQYGWCGIGNEYCGQGCQSNCG
JG720657.1	*Ephedra distachya*	EQCGRQAGGAVCPGGACCSQFGWCGTTQDYCGAGCQSQCSG
JG720009.1	*Ephedra distachya*	EQCGRQAGGAVCPGGACCSQFGWCGTTQNYCGTGCQSQCSG
VDAO-2033117	*Ephedra sinica*	EQCGRQAGGKVCDGGMCCSQFGWCGTTNEYCGAGCQSNCG
URZI-2013338	*Epifagus virginiana*	EQCGSQAGGAVCPNGLCCSQYGWCGNTDVYCGAGNCQSQC
GH617811.1	*Epimedium sagittatum*	CGRQVGGRTCGSNICCSQWGFCGTTDDHCLPSNNCQSNCR
GH615104.1	*Epimedium sagittatum*	VGGCTCDGNICCSQWGFCGTTDDHCLPSNNCQSNCRG
RKGT-2057060	*Eschscholzia californica*	IPFNEWTSDRCGTFGGSVCPEGSCCSIWGYCGNTNDYCLYNCYSQCTKLIP
ERXG-2060416	*Eschscholzia californica*	IPFNEWTSDRCGTFGGSVCPEGSCCSIWGYCGNTNDYCVYNCYSQCTKLIP
RKGT-2004462	*Eschscholzia californica*	QQCGRQGGGSTCRDNLCCSQYGYCGNTDAYCLPSNNCQSNCRG
KJ413009.1	*Eucommia ulmoides*	EQCGRQAGGASCPSGLCCSNFGWCGNTPEYCGSGNCQSQCG
HS256865.1	*Euonymus alatus*	EQCGTQANGALCPGGLCCSQFGWCGNTNAYCGDGCQSQCG

QTJY-2061015	*Euptelea pleiosperma*	QEQCGSQAGGKVCPGGLCCSKFGYCGNTPDYCTNGCQSQCTTP
QTJY-2059221	*Euptelea pleiosperma*	QNCGRQAGGRTCDGNICCSQWGFCGTTNDHCLPSNNCQSNCKSG
AK352622.1	*Eutrema halophilum*	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG
XM_006408097.1	*Eutrema salsugineum*	QQCGSQAGGQTCPGNICCSQYGYCGTTADYCSPDNNCQSNCWG
KPUM-2100737	*Exacum affine*	EQCGAQAGNAVCPNGLCCSQFGWCGSTDDYCKNGCQSQC
XGFU-2129416	*Exocarpos cupressiformis*	EQCGSQAGGVVCPGGACCSKFGWCGSGPDYCSAGCQSNCG
JK729372.1	*Fagopyrum tataricum*	QCGAQGGGATCPGGLCCSQWGWCGSTPKYCGAGCQSNC
FR617289.1	*Fagus sylvatica*	EQCGRQAGGALCPGGQCCSKFGWCGTTSDYCTNECQSQCG
QHBI-2007953	*Falcatifolium taxoides*	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCP
ROWR-2009883	*Falcatifolium taxoides*	ENCGRQAGGQICNGGECCSQYGWCGTTPDHCGTGCQSNCG
LC222264.1	*Ficus carica*	EQCGRQASGKTCPNNLCCSQYGWCGSTDEFCSRSKGCQSNCG
HQ224868.1	*Ficus pumila var. awkeotsang*	EQCGKQASGKTCPNNLCCSQYGWCGSTDEYCSRSKGCQSNCG
EDHN-2009181	*Ficus religiosa*	EQCGKQAGGALCPGGMCCSQYGWCGTTSEYCGEGCQSQCG
EDHN-2049024	*Ficus religiosa*	EQCGRQAGGKTCPNSLCCSQYGWCGSTDEYCSPSKGCQSNCG
EDHN-2009182	*Ficus religiosa*	EQCGSQAGGALCPGGLCCSKYGWCGNTPDYCNDGCQSQCG
ZNZC-2010710	*Flaveria bidentis*	DPTCSPAGNFWCDSGRCCSIYNWCGSISDYCASGNCLAQCWP
QBGG-2007208	*Flaveria bidentis*	QNCGRQAGNAPCSNGNCCSQYGYCGNTPAHCLPENNCQSQCRP
UEVI-2033737	*Fokienia hodginsii*	DPTCSPAGNFWCDSGRCCSIYNWCGSISDYCASGNCLAQCWP
UEVI-2008402	*Fokienia hodginsii*	DPTCSPAGNFWCNSGRCCSQYNWCGSTSDYCASGNCLAQCWP
UEVI-2056865	*Fokienia hodginsii*	QNCGSQAGGKVCSGGECCSKYGWCGTTDDHCLPDRGCQSNCAG
DHWX-2074382	*Fontinalis antipyretica*	EQCGSQAGNQKCPNNLCCSKFGYCGDTDEYCGNGCQSQCHG
DHWX-2066666	*Fontinalis antipyretica*	QQCGRAVNNTPCANLLNCCSQAGYCGTTDAYCVTGCQSGPCR
YSRZ-2089351	*Fouquieria macdougalii*	EQCGRQAGGALCPGGLCCSKFGWCGDTGEYCGEGCQSQCG
CX662023.1	*Fragaria vesca*	EQCGRQAGGATCPNGLCCSEYGWCGTTPDYCATGCQSQCTP
XM_004287696.2	*Fragaria vesca subsp. vesca*	EQCGRQAGGATCPNGLCCSEYGWCGTTPDYCATGCQSQCTP
HDWF-2070528	*Francoa appendiculata*	VQCGRDADGKLCPNGLCCSQYGYCGTTIPYCGPGCQSQC
UFHF-2060702	*Galax urceolata*	EQCGSQAGGALCPGGLCCSKFGWCGSTSDYCGNGCQSQCG
ADHK-2058358	*Galax urceolata*	QQCGQQAGGTLCANNLCCSQYGYCGTTDDYCLPSNNCQSNCRPG
AY952142.1	*Galega orientalis*	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCQG

AY253986.1	*Galega orientalis*	EQCGSQANGAVCPNGLCCSQFGYCGNTDQYCGAGCQSQCK
FWCQ-2032626	*Garcinia oblongifolia*	QNCGCASNLCCSQYGYCGTGDAYCGTGCKEGPC
HGSM-2003788	*Gelsemium sempervirens*	EQCGSQAGNALCPNGLCCSKFGWCGSTDDYCRDGCQSQC
VKGP-2120376	*Geranium carolinianum*	QQCGRQAGGAVCAGGLCCSQYGYCGSGPDYCGTGCQSQCG
VKGP-2118929	*Geranium carolinianum*	QQCGRQGGGRSCGGNMCCSQYGYCGTSDDYCSPSKGCQSNCAG
YGCX-2138168	*Geranium maculatum*	QNCGRQAGGRRCANRLCCSQFGFCGSTHEYCGFGCQSQCPR
DHAW-2015118	*Geum quellyon*	QQCGKQTGGKTCAGNLCCSQYGYCGTTDDYCSQSKGCQSNCAG
DHAW-2010903	*Geum quellyon*	ENCGRQAGGAICPNGLCCSQFGWCGTTPQHCTQNCQSQCTP
VHZV-2083997	*Gleditsia sinensis*	EQCGRQAGGALCPNGLCCSQYGWCGTTSDYCGTNCQSQC
VHZV-2089852	*Gleditsia sinensis*	EQCGRQAGGAVCPGGLCCSQFGWCGSTDEYCGAGCQSQCSG
VHZV-2087875	*Gleditsia sinensis*	EQCGRQAGGQQCPDNLCCSQFGYCGSTDDYCSPSKNCQSNCKG
GEHT-2004072	*Gleditsia triacanthos*	EQCGRQAGGALCPNGLCCSQYGWCGTTSDYCGTNCQSQC
GEHT-2054485	*Gleditsia triacanthos*	QQCGRQGGGKQCPDNLCCSQFGYCGTTDDYCSPSKNCQSNCKG
GR842224.1	*Glycine max*	CGRQAGGQTCGNNLCCSQYGWCGNSEDHCSPSKNCQSTCWG
FG990284.1	*Glycine max*	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG
EH039090.1	*Glycine max*	QNCGRQAGGQTCGNNLCCSQYGWCGNSEDHCSPSKNCQSTCWG
GR828879.1	*Glycine max*	EQCGTQAGGALCPNRLCCSKFGWCGDTDSYCGEGCQSQCK
CA853535.1	*Glycine max*	EQCGTQAGGALCPNRLCCSKFGWCGDTDSYCGEGCQSQCKSATP
ZUQW-2125577	*Glycine soja*	EQCGTQAGGALCPNRLCCSKFGWCGDTDSYCGEGCQSQCKSATP
PEZP-2011582	*Glycyrrhiza glabra*	EQCGRQAGGAVCPGGLCCSKFGWCGSSPDYCGDGCQSQCG
PEZP-2068102	*Glycyrrhiza glabra*	EQCGRQAGGKTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCQG
PEZP-2012969	*Glycyrrhiza glabra*	EQCGRQAGGAVCPNGLCCSQHGWCGTTDAYCGEGCQSQC
JTQQ-2003664	*Glycyrrhiza lepidota*	EQCGRQAGGAVCPGGLCCSKFGWCGSSPDYCGDGCQSQCG
FS265023.1	*Glycyrrhiza uralensis*	EQCGRQAGGAVCPGGLCCSKFGWCGSSPDYCGDGCQSQCG
FS275297.1	*Glycyrrhiza uralensis*	EQCGRQAGGAVCPNGLCCSQHGWCGTTDAYCGEGCQSQCKG
AB075372.1	*Glyptostrobus pensilis*	QNCGRQAGNAPCSNGNCCSQYGFCGNTPEHCSPANNCQSQCTG
DN954328.1	*Gnetum gnemon*	CGQQAGGALCPGGLCCSKYGWCGSTADYCGDGCQSQCX
MAQO-2107189	*Gomortega keule*	EQCGSQAGGALCPGGLCCSRYGWCGSTNDYCGEGCQSQCG
VLNB-2018423	*Gompholobium polymorphum*	EQCGQQAGGQTCPNNLCCSQYGYCGNTDDYCSPSKNCQSNCWG
JG842791.1	*Gossypium arboreum*	EQCGRQAGGALCPGGLCCSQFGWCGSTADYCTVPGCQSQCSG

DW516477.1	*Gossypium hirsutum*	EQCGRQAGGALCPNNLCCSQYGWCGNTDDYCSPEKNCQSNCRSG
XM_012637384.1	*Gossypium raimondii*	QQCGRQAGGRTCANNLCCSQFGYCGTTDDYCSPSRGCQSNCRG
GRRW-2003693	*Grevillea robusta*	EQCGSQAGGAVCPGGACCSQYGWCGTTNDYCGTGCQSQCSG
XMQO-2010683	*Gunnera manicata*	EQCGRQAGGALCPGGACCSQFGWCGTTADYCTNGCQSQCTPS
QZXQ-2006988	*Gymnocladus dioicus*	EQCGRQAGGALCPNGLCCSQYGWCGTTNDYCATNCQSQC
QZXQ-2018079	*Gymnocladus dioicus*	EQCGRQAGGAVCPGGLCCSQFGWCGSTNEYCGAGCQSQCG
BSVG-2014470	*Gyrocarpus samericanus*	QQCGRQAGGRICDNNLCCSQFGWCGSTDEYCLPTRGCQSNCRG
UAXP-2079326	*Gyrostemon ramulosus*	EQCGRQAGGALCPNGLCCSEFGWCGNTDPYCGKGCQSQCTG
UAXP-2077977	*Gyrostemon ramulosus*	QQCGRQAGGRTCANNLCCSQYGYCGNTDDYCSPAKNCQSNCWP
SIIK-2012655	*Hakea drupaceae*	EQCGSQAGGALCPGGACCSKYGWCGTTNDYCGSGCQSQCSSG
OWFC-2009897	*Halocarpus bidwillii*	DPSCNPAAKIYCNPGRCCSKFNWCGTLTSYCGKDYCIAQCPPG
OWFC-2045614	*Halocarpus bidwillii*	EPTCSPAGRIYCSPGRCCSRFNWCGDTPAYCGRGNCIANCHX
YHXT-2011001	*Hamamelis virginiana*	EQCGKQAGGANCPGGQCCSQYGWCGTASDYCGSGCQSQCAP
DMKC-2000892	*Hamamelis virginiana*	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDS
YHXT-2056758	*Hamamelis virginiana*	QEQCGSQAGGALCAGGLCCSKFGFCGSTPAYCTNGCQSQCTSG
EITK-2031556	*Hedera helix*	QQCGKQAGNALCPNGLCCSQFGYCGNTPDYCTNCQSQCG
EITK-2139802	*Hedera helix*	QQCGSQGGGKTCANNLCCSQYGYCGTTDEYCSPSKNCQSNCKS
TNWF-2128136	*Heliconia sp*	EQCGSQAGGALCPGGLCCSQYGWCGNTAAYCGNGCQSQCSG
XM_021418087.1	*Herrania umbratica*	QQCGSQAGGALCANGLCCSQFGWCGNTPEYCGTGCQSQCG
CMCY-2028506	*Hesperaloe parviflora*	QQCGSQAGGAVCPNGLCCSQFGYCGSTDPYCGKGCQSQCG
ERIA-2072338	*Heuchera sanguinea*	EQCGSQAGNAKCPGGQCCSKFGWCGTTSDYCGSGCQSQC
ERIA-2073359	*Heuchera sanguinea*	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTNNCQSNCNSAG
JZ538450.1	*Hevea brasiliensis*	EQCGRQAGGALCPGGLCCSQFGWCGNTPDYCGATCQSQCGDG
HS993022.1	*Hevea brasiliensis*	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKNSG
HS994451.1	*Hevea brasiliensis*	EQCGRQAGGKLCPNNLCCSQYGWCGSSDDYCSPSKNCQSNCKG
OLXF-2025952	*Hibiscus cannabinus*	EQCGRQAGGALCPGGLCCSQFGWCGNTDAYCKNGCQSQCTP
OLXF-2090625	*Hibiscus cannabinus*	EQCGRQAGGALCPNNLCCSQYGWCGSTDEYCLPENKCQSNCRG
OLXF-2017798	*Hibiscus cannabinus*	QQCGRQAGGQTCANNLCCSQYGYCGTTDDYCSPSKSCQSNCRPG

JQ289153.1	*Hippophae rhamnoides*	EQCGKQAGGKVCPGGQCCSQYGWCGTTDQYCKNGCQSQCG
CSSK-2009761	*Houttuynia cordata*	QQCGTQAGGALCAGGMCCSQYGYCGTDAAYCGEGCQSQCNG
ES655104.1	*Humulus lupulus*	ECGRDAFGEFCPKSQCCSQWGYCGNTSVYCGRGCQSNCWNS
AQGE-2052547	*Humulus lupulus*	EQCGRQAGGALCPGGLCCSQFGWCGNTDEYCKDGCQSQCKP
AQGE-2045110	*Humulus lupulus*	EQCGRQAGGASCPNGLCCSQYGWCGTTNDYCSPQAGCQSNCKG
NMGG-2058911	*Hypecoum procumbens*	QQCGRQAGGATCQNNLCCSQYGYCGSTDAYCLPSNNCQSNCRG
WPHN-2013558	*Idiospermum australiense*	EQCGRQAGGALCPGGLCCSQYGWCGTTADYCGPTCQSQCG
VZCI-2001282	*Illicium floridanum*	QQQCGEDVGGTLCPNGLCCSRWGYCGTTPPYCGDGCQSQC
JEXA-2062741	*Impatiens balsamifera*	QQCGRQAGGRLCDNNLCCSQWGYCGSTDEYCLTSNGCQSNCRG
NHAG-2048417	*Ipomoea nil*	QQCGRQASGRLCGNGLCCSQWGYCGSTAAYCGAGCQSQCKSTA
XM_019318656.1	*Ipomoea nil*	QQCGRQASGRLCGNRLCCSQWGYCGSTASYCGAGCQSQCRS
PYHZ-2072627	*Isoetes sp*	EQCGSQAGGVLCANNLCCSQYGWCGSTSAYCGTGCQSQCG
PKOX-2006241	*Isoetes tegetiformans*	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCSPG
UWFU-2035071	*Itea virginica*	QQCGKQAGGKTCAGDICCSQYGYCGTTDDYCSSSKNCQSNCKG
XM_012209964.2	*Jatropha curcas*	EQCGRQANNAQCPNGLCCSQFGWCGNTPEYCGAGNCQSQCPG
WTDE-2126717	*Johnsonia pubescens*	EQCGKQAGGAVCPGGMCCSQYGFCGTTSQYCGTNCQSQCG
WTDE-2126750	*Johnsonia pubescens*	EQCGRQANGALCPNALCCSQFGWCGKTDEYCKKGCQSQCG
WTDE-2023969	*Johnsonia pubescens*	FQCGLLAGDKECPDNLCCSRYSWCGSSAEYCGQGCQSGPCYS
DXQW-2058926	*Juglans nigra*	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS
XM_018979984.1	*Juglans regia*	EQCGMQAGGALCPNGLCCSQWGWCGNTAAYCGTGCQSQCP
XMGP-2055217	*Juniperus scopulorum*	DPTCSPAGNFLCNSGRCCSKYNWCGSTSDYCASGNCLAQCWP
NWMY-2061736	*Kadsura heteroclite*	EQCGSQVGGATCDGGLCCSRWGWCGSTDEYCGEGCQSNC
KTWL-2001822	*Kaliphora madagascariensis*	EQCGRQAGSAVCPNGLCCSQYGWCGSTSDYCGDGCQSQCTG
BCAA-2071511	*Kirkia wilmsii*	EQCGWQAGGALCPNNLCCSQYGWCGSTDPYCGNGCQSQC
XM_023878188.1	*Lactuca sativa*	QNCGRQAGNAPCSNGNCCSQYGFCGNTPEHCSPANNCQSQCTG
ZQWM-2006577	*Lagarostrobos franklinii*	EQCGRQAGGAVCPGGLCCSQWGWCGNTKDHCGTGCQSQCG
RJNQ-2117967	*Lagerstroemia indica*	EQCGSQAGGAVCPGGLCCSQWGWCGSTADYCSTGCQSQCG
UDUT-2047060	*Larrea tridentata*	QQCGSQAGGRICDGNLCCSQYGWCGSTAEYCSPDQKCQSNCWG

GO319239.1	*Lathyrus odoratus*	EQCGRQSNGAFCPNGLCCSQFGWCGDTDSYCGTGCQSRCKS
KNMB-2055925	*Lathyrus sativus*	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCHG
WAIL-2059125	*Laurelia sempervirens*	QQCGSQRGGAVCTGRLCCSQFGYCGNTDPYCGAGCQSKCSSG
WXVX-2017468	*Ledum palustre*	EQCGTQAGGALCPGGLCCSKFGWCGNGGDYCSNGCQSQCG
WBIB-2057434	*Lepidosperma gibsonii*	AQTCGSQAGGALCADNMCCSQYGYCGSTSDYCGTGCQSQCG
HS977103.1	*Leucaena leucocephala*	EQCGRQAGGALCPGRLCCSQFGWCGSTNDYCGPGCQSQCG
VMXJ-2009688	*Leucobryum albidum*	EQCGRQAGGATCPGGLCCSQYGYCGTTPDYCGGGCQSNCG
RGKI-2000950	*Leucobryum glaucum*	EQCGRQAGGATCPGGLCCSQYGYCGTTPDYCGGGCQSNCG
WGTU-2071701	*Leucostegia immersa*	QQCGSQAGGAVCANNLCCSQYGYCGSTADYCGTGCQSQC
HBUQ-2036853	*Licania michauxii*	EQCGSQAGGALCPNNLCCSQFGWCGNTNDYCGPGCQSQCG
MH806400.1	*Lilium longiflorum*	QQCGTQRGGAVCPGGLCCSQFGYCGSTIAYCGDGCQSQCIG
CRNC-2000081	*Limnanthes douglassii*	QQCGSQAGGRTCPNNICCSQYGYCGTTEDYCSPSKNCQSNCWG
CX263350.1	*Limonium bicolor*	EQCGSQAGGAVCPNGLCCSKYGWCGSTDTYCKDGCQSQCG
WOBD-2081469	*Limonium spectabile*	QQCGSQAGGRTCANNLCCSQYGYCGSTDAYCATGCQSNCKTG
MYVH-2061367	*Linum grandiflorum*	QNCGSQAGNAVCPGGLCCSQFGWCGNTPEHCTTGCQSQC
HNCF-2000289	*Linum hirsutum*	QNCGRQAGGATCANNLCCSQFGFCGTTDDHCNPSKNCQSNCRPSG
AEPI-2048092	*Linum leonii*	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKNCQSNCRPG
BHYC-2052042	*Linum lewisii*	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKNCQSNCRP
KCPT-2102059	*Linum macraei*	QDCGRQAGGRTCANNLCCSQWGFCGTTDDHCNPSKNCQSNCRP
TXMP-2007103	*Linum strictum*	QNCGRQAGGRTCAGNLCCSQYGFCGTTDDHCNPSKNCQSNCRPG
OGSY-2094087	*Linum usitatissimum*	QNCGRQAGGQTCANNLCCSQYGYCGTTDDHCSPSKACQSNCKP
OGSY-2028337	*Linum usitatissimum*	QNCGSQAGNAVCPGGLCCSRFGWCGNTPDHCTNGCQSQC
CK764856.1	*Liriodendron tulipifera*	EQCGRQAGGATCPGNLCCSQYGWCGSTDEYCTNNCQSNCKPSG
WAXR-2009837	*Litchi chinensis*	EQCGRQAGSALCPGGQCCSQFGWCGTTPAYCGDGCQSQCG
IZLO-2060473	*Lobelia siphilitica*	QPQCGRDAGNAVCPNGLCCSQFGYCGNTPDYCNNCQSQCG
MUMD-2011940	*Lomandra longifolia*	EQCGSQAGGALCPNNLCCSKFGWCGSTDDYCKDGCQSQCQQG
OMDH-2065588	*Loropetalum chinense*	EQCGKQAGGAKCPGGQCCSQFGWCGTISDYCGSGCQSQCTP
HQRJ-2007580	*Loropetalum chinense*	QQCGKQAGGKTCPGNICCSQYGYCGTTDDYCSPSKNCQSNCQG
AV410121.1	*Lotus japonicus*	EQCGSQAGGALCPGGICCSKYGWCGSTSEYCGEGCQSQCG

AV421457.1	*Lotus japonicus*	EQCGSQAGGALCPGGICCSKYGWCGSTSEYCGEGCQSQCGG
BI419506.1	*Lotus japonicus*	EQCGSQAGGAVCPGGICCSKYGWCGSTSEYCGEGCQSQCG
BI419767.1	*Lotus japonicus*	EQCGSQAGGAVCPGGICCSKYGWCGSTSEYCGEGCQSQCGG
FS322371.1	*Lotus japonicus*	EQCGSQAGGQLCPNNQCCSQYGWCGNTEDYCSSSKGCQSRCWG
BT133701.1	*Lotus japonicus*	EQCGTQAGGAVCPNGLCCSKYGYCGNTDSYCGADCQSQCKKSG
BM060586.1	*Lotus japonicus*	QNCGRQAGRRVCANRLCCSQFGFCGTTREYCGAGCQSNCR
VVVV-2006290	*Ludovia sp*	QQCGSQAGGALCPGGLCCSQYGYCGTTTPYCGKGCQSQCG
TXVB-2072372	*Lunularia cruciata*	NNTIQRCGKDYGGAKCRNSLCCSEFGFCGTNASYCGPGCQSQCSLHATDP
XM_019558790.1	*Lupinus angustifolius*	QSCGCTEGLCCSQYGYCGTGDAYCGTGCKEGPCYASP
LWCK-2001931	*Lycium barbarum*	QQCGRQKGGALCSGNLCCSQYGWCGSTPEYCSPSQGCQSRCSG
OSMU-2078285	*Lycium sp*	EQCGRQAGGARCPSGLCCSNFGWCGDTQAYCGPGNCQSQCSSG
OSMU-2011251	*Lycium sp.*	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG
UGJI-2125875	*Lycopersicon cheesmanii*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG
ULKT-2009410	*Lycopodiella apressa*	QQCGSQASGAKCANNLCCSQYGYCGSTDAYCGQGCQSGPC
ENQF-2077819	*Lycopodium annotinum*	QLQCGSAAGNKVCSGGLCCSKFGYCGTTDAYCGGGCQSGRCIG
DTOA-2089413	*Maesa lanceolata*	QQCGSQAGGKTCANNLCCSQYGYCGSTDDYCLPSKNCQSNCKSG
WBOD-2101820	*Magnolia grandiflora*	AEQCGRQAGGATCPGNLCCSQYGWCGTTDEYCTNNCQSNCRPSG
EB105992.1	*Malus domestica*	EQCGSQAGGAVCPNGLCCSQFGWCGTTSDYCAAGCQSQCSS
CN446645.1	*Malva pusilla*	EQCGRQAGGALCPDRLCCSQFGWCGDTDPYCKDGCQSHCSG
XM_021764317.1	*Manihot esculenta*	EQCGSQAGGQLCPNNLCCSRYGWCGTTDEFCSDGCQSNCKG
DB921422.1	*Manihot esculenta*	EQCGRQAGGALCPGGLCCSKFGWCGNTPDYCGADCQSQCSAG
JG999405.1	*Manihot esculenta subsp. Peruviana*	EQCGSQAGGQLCPNNLCCSRYGWCGTTDEYCSDGCQSNCKG
XNLP-2089391	*Manihot grahamii*	EQCGRQAGGALCPGGLCCSKFGWCGNTPDYCGADCQSQCSAG
CDFR-2008765	*Manoao colensoi*	EPTCSPAGRIYCSRGRCCSKFNWCGTGPAYCGKSNCIAQCS
XPAF-2007050	*Mapania palustris*	QQCGSQAGGATCPDNLCCSQYGYCGSTSAYCGDGCQSQCSG
UXCS-2010144	*Marattia sp*	EQCGSQAAGAVCPNGLCCSQWGWCGSTDAYCGSGCQSQC
TFYI-2003205	*Marchantia emarginata*	QVCGSQAGGALCQGNQCCSQYGYCGTTPAYCGTGCQSQC
HMHL-2050343	*Marchantia paleacea*	EQCGQQAGNALCPDNLCCSQFGWCGNTIDYCGNGCQSGPC
IHWO-2059313	*Marchantia paleacea*	QECGSEAGGALCPDNYCCSKAGWCDSGDAYCGEGCQSGPCNY
JPYU-2036258	*Marchantia polymorpha*	EQCGRQAGNAVCPNNLCCSQYGWCGSTSDYCVTGCQSGPCSG

HM224449.1	*Medicago sativa*	EQCGSQANRAVCPNGLCCSKFGWCGTTDQYCGAGCQSQCR
XM_013595074.2	*Medicago truncatula*	EQCGRQAGGKTCPNNLCCSQYGYCGTTDEYCGPNCQSNCHG
ES611787.1	*Medicago truncatula*	QQCGRQANGAVCANRLCCSQFGYCGNTADYCGAGCQSQCTS
AJ503831.1	*Medicago truncatula*	EQCGSQANRAVCPNGLCCSKFGWCGTTDQYCGAGCQSQCRS
VCCF-2053609	*Melia azedarach*	EQCGSQAGGALCPGGLCCSKFGYCGNTPAYCTDGCQSQCNSG
AALA-2011529	*Meliosma cuneifolia*	QQCGRQAGGRTCAGNICCSQWGYCGTTDEYCLPSNNCQSNCKSG
IXVJ-2005589	*Menyanthes trifoliata*	EQCGRQAGNAVCPNGLCCSQFGYCGNKPEYCNTGCQSQC
DKFZ-2059098	*Mertensia paniculata*	QQCGSQAGGALCPNGLCCSQHGYCGSTDDYCKNNCQSQCG
NRXL-2000516	*Metasequoia glyptostroboides*	DPTCSPAGNFWCDSGRCCSIYNRCGSTSDYCASGNCLAQCWP
NRXL-2001780	*Metasequoia glyptostroboides*	ENCGRQASGAVCPGGECCSEWGWCGNTVDHCRTPGCQSQCG
NRXL-2058286	*Metasequoia glyptostroboides*	QNCGSQAGGNVCSGGECCSKWGWCGTTDDHCLPDRGCQSNCRG
XQWC-2158723	*Michelia maudiae*	EQCGRQAGGALCPGGLCCSQFGWCGNTLPYCGEGCQSQCG
XQSG-2037376	*Microbiota decussata*	DPTCSPAGNFWCDSGRCCSIYNWCGSTADYCASGNCLAQCWP
MHGD-2009333	*Microcachrys tetragona*	EQCGKQASGAVCPGGLCCSQWGWCGNTADHCGNGCQSQCG
YPIC-2013971	*Microstegium vimineum*	GGALCPNCLCCSKFGWCGTTSDYCGNGCQSQC
BBDD-2004959	*Microstrobos fitzgeraldii*	DPSCSPTVTGRISCNSGRCCSKFNWCGTGVSYCGKDYCVAQC
BBDD-2007819	*Microstrobos fitzgeraldii*	DPTCSPAGGFGCNAGRCCSRFNWCGGTAEYCRRPNCISQCWP
BBDD-2070246	*Microstrobos fitzgeraldii*	EPTCSPAGRIYCNRGRCCSKYNWCGTGPAYCSKSNCIAQCS
BBDD-2071547	*Microstrobos fitzgeraldii*	EPTCSPAGRIYCSSGRCCSKYNWCGTGPAYCSKSNCIAQCS
GQAV-2094745	*Mirabilis jalapa*	EQCGRQASGALCPNNLCCSQYGWCGNTEPYCGTGCQSQC
DAYQ-2052183	*Mitella pentandra*	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTNNCQSNCNQG
RNBN-2042717	*Mollugo cerviana*	EQCGKQAGGALCPNGLCCSQFGWCGTTTDYCGTGCQSQC
LRTN-2097367	*Monotropa uniflora*	EQCGRQAGGALCPGGLCCSQFGWCGNGGDYCGTGCQSQCG
LRTN-2016473	*Monotropa uniflora*	QQCGRQVAGRLCPGDQCCSQWGYCGTTDDYCLPSKNCQSNCRPSG
LRTN-2089067	*Monotropa uniflora*	QQEQCGSQAGGKLCPGSLCCSRFGFCGNTTDYCSTGCQSNCVG
ULGV-2007220	*Morinda citrifolia*	EQCGRQAGGALCPNNLCCSQYGWCGNTGEYCLPSNNCQSNCWG
CZPV-2001465	*Moringa oleifera*	QNCGRQAGNRACANELCCSQYGFCGSTSEYCSRANGCQSNCRG

JX432966.1	*Morus alba var. multicaulis*	SEQQCGRDVGGALCHGNLCCSHWGFCGTTAIFCDVDQGCQSQCWSSP
XVJB-2002447	*Morus nigra*	EQCGRQAGGAICPNGLCCSQYGWCGSTYEYCSPEVNCQSNCWSAP
XM_010103989.1	*Morus notabilis*	EQCGRQAGGKTCPNNLCCSQYGWCGNTDEYCSSSKGCQSNCQG
XM_009394595.2	*Musa acuminata subsp. Malaccensis*	EQCGRQAGGALCPGGLCCSQYGWCGNTDPYCGQGCQSQCTG
XM_009395786.2	*Musa acuminata subsp. Malaccensis*	EQCGTQAGGALCPGGLCCSQYGFCGSTEAYCGSGCQSQCG
AY532922.1	*Musa balbisiana*	EQCGRQAGGALCASGKESLCCSQFGWCGNTDDYCGSQEGCQSQCPG
GU391234.2	*Musa sp. AB group*	EQCGRQAGGALCPGGLCCSQYGWCGNTDPYCGQGCQSQCTGS
AJFN-2095016	*MyOdocarpus sp.*	QQCGRQAGGTTCANNLCCSQYGYCGTTDDYCSPSQNCQSNCKSG
INSP-2052424	*Myrica cerifera*	EQCGKQAGGAVCPNGLCCSGFGWCGSTTEYCGANCQSQCG
INSP-2052907	*Myrica cerifera*	EQCGRQAGAAVCPNGLCCSQHGWCGTTPDYCGTGCQSQC
IUSR-2074233	*Myriophyllum aquaticum*	EQCGSQAGGAVCPNGLCCSQFGFCGSTADYCGTKCQSQCSG
OBPL-2033644	*Myristica fragrans*	EQCGRQAGGALCPGGLCCSEFGWCGSGPAYCGHGCQSQCG
UUJS-2000623	*Nageia nagi*	DPTCSPAGRQYCNAGRCCSKFNWCGTGPAYCGKGNCIAQCP
YHFG-2075257	*Nandina domestica*	QCGEQASGKLCPGGLCCSQWGYCGSTEAYCGDGCQSQC
IQYY-2022155	*Narcissus viridiflorus*	EQCGSQAGGAVCPNGLCCSQYGFCGSTDPYCGKGCQSNCG
XM_010244546.2	*Nelumbo nucifera*	EQCGRQAGGALCPGGLCCSQYGWCGNTGDYCGEGCQSQCSG
XM_010244547.2	*Nelumbo nucifera*	EQCGWQAGGALCPGGLCCSQYGWCGNTDTYCGEGCQSQCR
PZRT-2001562	*Nelumbo nucifera*	QQCGRQAGGALCAGGLCCSQWGYCGSTTNYCGTGCQSQCSG
JDQB-2004699	*Neocallitropsis pancheri*	DPTCSPAGNFWCNSGRCCSIYNWCGDTSDYCASGNCLAQCWQ
JDQB-2064704	*Neocallitropsis pancheri*	DPTCSPAGRFWCNSGRCCSIYNWCGSTWEYCAPGNCLAQCWP
AY618887.1	*Nepenthes khasiana*	EQCGSQAGGAACPGGLCCSQFGWCGTTDDYCEAGCQSQCSSSG
AY618885.1	*Nepenthes khasiana*	EQCGSQAGGAVCPGGLCCSQYGWCGTTDDYCGAGCQSQCSFSG
AY618883.1	*Nepenthes khasiana*	FQCGQQAGGALCHSGLCCSQWGWCGTTSDYCGNGCQSQCG
MH430924.1	*Nepenthes sp. MF-2019*	EQCGSQAGGAACPGGLCCSQYGWCGTTDDYCEAGCQSQCSSSDG
MH430923.1	*Nepenthes sp. MF-2019*	FQCGQQAGGALCHSGLCCSQWGWCGTTSDYCGNGCQSQCG
HM588678.1	*Nicotiana attenuata*	EQCGSQAGGALCPSGLCCSKFGWCGDTNDYCAPGNCQSQCRG
CK285068.1	*Nicotiana benthamiana*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
FS392106.1	*Nicotiana tabacum*	EQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGNCQSQCPG
FS393955.1	*Nicotiana tabacum*	QQCGRQRGGALCSGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG

XM_009619007.2	*Nicotiana tomentosiformis*	EQCGSQAGGARCPSGLCCSKFGWCGNTNDYCGPGNCQSQCPG
ABCD-2074237	*Niphotrichum elongatum*	EQCGKQAGGASCPGGLCCSQYGWCGTTTPYCEDGCQSNCG
HOKG-2093902	*Nolina atopocarpa*	QQQCGSQAGGAVCPNGLCCSSYGYCGSTDRYCGSGCQSQCG
TJLC-2009721	*Nothofagus obliqua*	FQCGTQAGGAVCPNGLCCSQWGWCGTTNDFCGSGCQSNC
AREG-2046875	*Nothotsuga longibracteata*	ENCGKQAGGAVCPGGLCCSEYGWCGNTQGYCGQGCQSQCG
XM_031637854.1	*Nymphaea colorata*	QQCGTQAGGALCAGGLCCSQWGYCGSTSAYCGTGCQSQC
VUSY-2130965	*Nyssa ogeche*	QQCGIDVGGTLCSGGLCCSKWGFCGTTDAHCGDGCQSQCDG
TVCU-2008600	*Ochna mossambicensis*	QNCGRQAGGRTCANNLCCSQWGYCGSTDDHCSPSKNCQSNCRG
ROLB-2078944	*Oenothera elata*	EQCGRQAGGALCPGGQCCSQYGWCGTTSDYCSTGCQSQCSG
ECSL-2001992	*Oenothera filiformis*	QNCGCASNLCCSQYGYCGSGNAYCGTGCQSGPC
TGOP-2046141	*Oenothera gaura*	QNCGCASNLCCSQYGYCGSGDAYCGTGCQSGPC
HPNZ-2034913	*Oenothera longituba*	QNCGCSSNLCCSQYGYCGSGTAYCGPGCKSGPC
YIVZ-2020758	*Oenothera rosea*	EQCGRQAGGALCPDGQCCSKFGWCGTTSDYCSTGCQSQCSG
JZ943226.1	*Olimarabidopsis pumila*	QQCGRQAGGQTCPGNICCSQYGYCGTTADYCSPANNCQSNCWG
PVGM-2148710	*Oncotheca balansae*	EQCGSQAGGAVCGGGQCCSRYGWCGITADYCGDGCQSQCSSP
PVGM-2013088	*Oncotheca balansae*	DPTCSPAGNFWCDTGRCCSIYNWCGSTADYCAQGNCLAQCWP
PVGM-2142572	*Oncotheca balansae*	QNCGSQAGGRACGGGECCSKWGWCGTTDDHCLPERGCQSNCRG
QHVS-2011139	*Ophioglossum vulgatum*	QQCGSQAGGAVCSNNLCCSKYGYCGSTDPYCGSGCQSGPCSSG
LSKK-2019644	*Orchidantha maxillaroides*	EQCGSQAGGAVCPGGLCCSKYGWCGSTAAYCGNGCQSQCG
UHBY-2021469	*Oresitrophe rupifraga*	EQCGKQAGGALCPNGNCCSQFGYCGNTPDYCTNGCQSQCG
CMEQ-2010424	*Orthotrichum lyellii*	ECGTQANNKKCPNNLCCSQYGYCGDTSLYCGTPGCQSQC
VIBO-2004398	*Osmunda javanica*	EQCGRQAGGKLCPNSLCCSQYGYCGSDPAYCGTGCQSQCG
YFGP-2104103	*Pallavicinia lyellii*	QAACGSQVGGALCPNGLCCSQFGFCGSTADYCGAGCQSQC
FJ790420.1	*Panax ginseng*	EQCGKQAGMALCPNGLCCSQFGWCGSTPEYCTNCQSQCG
MH782149.1	*Panax notoginseng*	EQCGKQAGMALCPNGLCCSQFGWCGSTPEYCTNCQSQCSEP
DN141224.1	*Panicum virgatum*	QQCGSQAGGKKCPNNLCCSPWGYCGSGPDYCGNGCQSGPCSG
FE621365.1	*Panicum virgatum*	QQCGSQAGGKKCPNNLCCSPWGYCGTGPDYCGNGCQSGPCSG
FNXH-2023053	*Papaver setigerum*	QQCGSQAGGRTCQGNLCCSQYGYCGDTNEYCLPSNNCQSNCRG
XM_026539927.1	*Papaver somniferum*	QQCGRQAGGRTCQGNLCCSQYGYCGDTDEYCLPSNSCQSNCR

XM_026581480.1	*Papaver somniferum*	QQCGRQAGGRTCQGNLCCSQYGYCGDTNEYCLPSNNCQSNCR
XM_026590851.1	*Papaver somniferum*	QQCGSQAGGRTCQGNLCCSQYGYCGDTNEYCLPSNNCQSNCKG
OVIJ-2005726	*Papuacedrus papuana*	DPTCSPAGNFWCKSGRCCSIYNWCGSTSDYCAPGNCLAQCWP
OVIJ-2001691	*Papuacedrus papuana*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
OVIJ-2082654	*Papuacedrus papuana*	QNCGSQAGGKVCGGSECCSKWGWCGTTDDHCLPDRGCQSNCRG
JZVE-2002557	*Parasitaxus usta*	DPTCSPAGRQYCNDGRCCSKFNWCGTGAAYCSKGNCIAQCP
JZVE-2002559	*Parasitaxus usta*	DPTCSPAGRVYCNPGRCCSKFNWCGTTRANCGKGNCIAQCHP
JZVE-2003275	*Parasitaxus usta*	EDCGKQAAGKLCPGGLCCSQWGWCGKTADHCGKNCQSQCSG
SIZE-2004173	*Passiflora caerulea*	EECGRQAGGALCPGGLCCSQYGWCGNTDAYCGAGCQSQCG
SIZE-2011788	*Passiflora caerulea*	EQCGQQAGGKTCPGNICCSQYGWCGTTDEYCSPDNNCQSNCKG
SIZE-2004177	*Passiflora caerulea*	EQCGRQAGGALCPGGLCCSQFGWCGSTNDYCGPGCQSQCG
OBTI-2043901	*Peganum harmala*	EQCGRQAGGKLCPGGQCCSQFGWCGTDSQYCGTGCQSQCSG
OBTI-2041811	*Peganum harmala*	QQCGRQAGGRLCPNNLCCSQFGYCGSSDDYCSPSKNCQSNCKPSG
YKFU-2013674	*Peltoboykinia watanabei*	EQCGSQAGGAKCQGGQCCSKFGWCGTTSDYCSTGCQSQCG
SBZH-2131130	*Pennantia corymbosa*	EQCGRQAGGALCPGGDCCSQFGWCGTTAEYCGDGCQSQCTP
JLOV-2001160	*Pereskia aculeata*	QQCGAQAGGALCRDGLCCSIWGYCGSTDPYCGPDNCQSQCPG
WIGA-2099273	*Persea borbonia*	EQCGRQAGGALCPGGLCCSQFGWCGSTSDYCGSTCQSQCG
KRJP-2083636	*Peumus boldus*	VQCGRQAGGQLCNGNLCCSQFGWCGSTDEYCLPSSGCQSNCKTSG
HO777299.1	*Phaseolus acutifolius*	EQCGRQAGGALCPGGLCCSNWGFCGFTIDYCGQGCQSQCG
HO777873.1	*Phaseolus acutifolius*	EQCGRQAGGALCPGGLCCSQFGWCGSTTDYCGKGCQSQCG
FD781929.1	*Phaseolus angustissimus*	QQCGRQAGGSRCSGNLCCSQFGWCGNTAEYCSPSQNCQSNCWG
FE679590.1	*Phaseolus vulgaris*	AQQCGRQAGGALCPGGNCCTQFRWCGSTTDYCGPGCQSQCG
FE679592.1	*Phaseolus vulgaris*	AEQCGRQAGGALCPGGNCCSQFGWCGSTTDYCGPGCQSQCG
FE679587.1	*Phaseolus vulgaris*	EQCGRQAGGALCPGGNCCSQFGWCGSTTDYCGKDCQSQCG
JK039277.1	*Phaseolus vulgaris*	EQCGRQAGGALCPGGNCCSQFGWCGSTTDYCGPGCQSQCG
JK038827.1	*Phaseolus vulgaris*	EQCVRQAGGALCPRGNCCSQFGWCGSTTDYCGPGCQSQCG
FE898289.1	*Phaseolus vulgaris*	QQCGRQAGGSRCSGNLCCSQFGWCGNTAEYCSPSQNCQSNCWG
AUIP-2030349	*Phelline lucida*	EQCGRQAGNALCPNGLCCSQYGYCGSTPEYCTQNCQSQCSG
PGKL-2073482	*Phellodendron amurense*	EQCGRQAGGALCPGGQCCSQFGWCGTAPDYCTNGCQSQCSG

PGKL-2050474	*Phellodendron amurense*	QHCGRQAGGRTCPNNLCCSQFGYCGTTDDYCSPSKNCQSNCQSAG
PGKL-2070733	*Phellodendron amurense*	QQCGKQAGGKTCSNNLCCSQYGYCGSTDDYCSPSKGCQSNCQS
ORKS-2002446	*Philonotis fontana*	EQCGRQAGGKTCPDNLCCSQWGYCGNTDEFCKNDCQSQCG
ORKS-2002448	*Philonotis fontana*	EQCGRQAGNQKCPKDQCCSYWGWCGDTPEFCEHGCQSQCSG
XM_008813877.3	*Phoenix dactylifera*	QQCGSQAGGTTCPDDLCCSQYGYCGSTSAYCGAGCQSQCSG
XM_026800205.1	*Phoenix dactylifera*	QQCGSQAGGTTCPNGLCCSQYGYCGSTSAYCSTGCQSQCSG
YGAT-2044856	*Phyllanthus sp*	EQCGRQAGGAVCPGGLCCSQFGWCGNTDTYCGNGCQSQCG
JRNA-2063583	*Phyllocladus hypophyllus*	EPTCSPAGRIYCSQGRCCSKFNWCGTGPAYCSKSNCIAQCS
JRNA-2063388	*Phyllocladus hypophyllus*	QNCGRQAGGVVCSGGECCSQYGWCGTTPDHCGTGCQSNCG
JRNA-2003842	*Phyllocladus hypophyllus*	EQCGRQAGGALCPGGLCCSKWGWCGNTADHCGTDCQSQCG
XM_024524440.1	*Physcomitrella patens*	EQCGRQAGGATCPGGLCCSQFGWCGSNAEYCGDGCQSQCG
YEPO-2070125	*Physcomitrium sp.*	QCTAPPRSDFQCGPANGNASCPSGKCCSWWGWCGATSFYCNQRSLCTS
RUUB-2010722	*Physena madagascariensis*	EQCGAQAGGALCPGGACCSQWGWCGTTTEYCGPGCQSQCKP
SXCE-2000039	*Physocarpus opulifolius*	EQCGRQANGAVCPNGLCCSQFGWCGSTTDYCATGCQSQC
BKQU-2099109	*Phytolacca americana*	EQCGRQAGGALCPNSLCCSQFGWCGNTNPYCGPGCQSQC
MRKX-2002612	*Phytolacca bogotensis*	QGHEGHGVGEILLMGKLGEPVCGVSASGRVCPNGHCCSEWGYCGTTNEYC GKGCQSQC
DV972520.2	*Picea glauca*	CGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
ETCJ-2001405	*Pilgerodendron uviferum*	DPTCSPAGNFWCDSGRCCSIYNWCGSTSDYCASGNCLAQCWP
ETCJ-2008954	*Pilgerodendron uviferum*	DPTCSPRGNFFCNSGRCCSIYNWCGTGPAYCGKGNCLAQCSP
KIIX-2004332	*Pilularia globulifera*	QTCGSQAGGAICPNNACCSQWGYCGTGSAYCGTGCQSGPCP
KIIX-2074234	*Pilularia globulifera*	SCATCADNLCCSQYGYCGSTSDYCGTGCQNGPCTSG
MK521170.1	*Pinus banksiana*	EQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCG
MK521203.1	*Pinus contorta*	QNCGCASGLCCSKYGYCGSTPAYCGRGCRSGPCSSSG
AY705804.1	*Pinus halepensis*	EQCGRQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCG
MFTM-2080210	*Pinus jeffreyi*	EQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCG
IIOL-2071721	*Pinus parviflora*	DPTCSPAGNFFCNNGRCCSRFNWCGTGPSYCGRGNCIAQCP
CT582939.1	*Pinus pinaster*	EQCGRQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCGG
JBND-2013698	*Pinus ponderosa*	EQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCG

DZQM-2014352	*Pinus radiata*	EQCGQQAGGALCPGGLCCSKWGWCGNTDAHCGQDCQSQCG
CF396457.1	*Pinus taeda*	EQCGQQAGGALCPGGLCCSKWGWCGNTDDHCGQDCQSQCGG
DR012435.1	*Pinus taeda*	EQCGRQASGALCPGGLCCSQWGWCGNTPAHCGQQCQSQCGG
MUNP-2000009	*Piper auritum*	QMCGWQAGGRLCPDNLCCSQWGYCGSTPPYCGEGCQSQC
XM_031411934.1	*Pistacia vera*	QYCGCQAHLCCSEFGYCGTGSAYCGNGCQEGPCY
GH719715.1	*Pisum sativum*	EQCGRQAGGATCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCHG
BUWV-2040465	*Platycladus orientalis*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
BUWV-2045303	*Platycladus orientalis*	DPTCSPAGNFWCNSGRCCSVYNWCGSTSDYCASGNCLAQCWP
BUWV-2044272	*Platycladus orientalis*	DPTCSPAGRFFCNSGRCCSKYNWCGSTSDYCAKNNCIAQCWP
TPEM-2001428	*Platyspermation crassifolium*	EQCGRQAGGAICPNGLCCSQYGFCGSTPEYCTDGCQSQCSG
CB818080.1	*Plumbago zeylanica*	QQCGRQAGGRTCPNGLCCSQYGYCGTTDDYCLKSKNCQSNCKG
SCEB-2006039	*Podocarpus coriaceus*	EQCGTQAGGALCPGGLCCSQWGWCGTTPDYCGNGCQSQCG
SCEB-2044852	*Podocarpus coriaceus*	DPSCSPAGRQYCNDGRCCSKFNWCGTGPAYCSKGNCIGQCPP
WFBF-2042355	*Podophyllum peltatum*	QEQCGKQAGEAVCPGGLCCSQWGYCGSSNDHCGTGCQSQCG
ZQYU-2018462	*Polypodium plectolens*	QQCGSQAGGALCANNLCCSQYGYCGSTSAYCAVPGCQSQCG
COBX-2007364	*Polypremum procumbens*	AEQCGRQASGALCPNGLCCSQYGWCGSTPAYCTNGCQSQCG
EDBB-2084228	*Polyscias fruticosa*	QQCGRQGGGKTCANNLCCSQYGYCGTTDDYCSPSKNCQSNCNSHG
SZYG-2006556	*Polytrichum commune*	EQCGTQAGGALCPNNLCCSQYGWCGSTSDYCGNGCQSQCG
CB240244.1	*Populus alba x Populus glandulosa*	EQCGSQAGSALCPGGLCCSQFGWCGNTNDYCGTGCQSQCG
CK318298.1	*Populus deltoids*	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
AJ778967.1	*Populus euphratica*	EQCGKQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG
AJ776776.1	*Populus euphratica*	EQCGSQAGGALCPGGLCCSQFGWCGKNNDYCGTGCQSQCG
XM_011040263.1	*Populus euphratica*	QNCGRQAGGRTCANNLCCSEWGFCGTSDDHCSPSKSCQSNCRASG
DV464270.2	*Populus fremontii x Populus angustifolia*	EQCGSQAGGAVCPGGLCCSQFGWCGSTNDYCGTGCQSQCG
DB889512.1	*Populus nigra*	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
DN496890.1	*Populus tremula*	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
BU874345.1	*Populus trichocarpa*	EQCGKQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
DT474845.1	*Populus trichocarpa*	INCGRQAGGRTCANNLCCSEWGFCGTSDDHCSPSKNCQSNCRP

DN485080.1	*Populus trichocarpa*	QNCGRQAGGQTCANNLCCSQWGYCGTSDDHCNPSKNCQSNCRSS
XM_006375777.1	*Populus trichocarpa*	QNCGRQAGGQTCANNLCCSQWGYCGTSDDHCNPSKNCQSNCRSSG
CN522571.1	*Populus trichocarpa x Populus deltoides*	EQCGRQAGDALCPGGLCCSSYGWCGTTVDYCGIGCQSQCDG
DT503232.1	*Populus trichocarpa x Populus deltoides*	EQCGSQAGGAVCPGGLCCSQFGWCGSTNDYCGNGCQSQCG
KRUQ-2107849	*Porella navicularis*	NCLVGGCGAGLCCSAYGYCGTGYCGAAAGTCRNYGCP
UUHD-2017041	*Porella pinnata*	EQCGRQAGGAICPNQLCCSQFGYCGTTSDYCGTGCQSNC
LFQH-2049300	*Portulaca amilis*	EQCGKQAGGALCPNRLCCSQYGWCGNTDPYCGAGCQSQC
BYNZ-2017723	*Portulaca mauii*	EQCGRQAGGALCPNHLCCSQFGWCGNTDQYCGAGCQSQC
UQCB-2020805	*Portulaca molokiniensis*	CGRQAGGALCPNRLCCSQYGWCGNTDPYCGAGCQSQC
TRLB-2022680	*Portulaca oleracea*	CGRQAGGALCPNRLCCSQYGWCGNTDPYCGAGCQSQC
GCYL-2045988	*Portulaca suffruticosa*	KNCNCPSNLCCSQYGYCGTGDAYCGAGCRAGPC
XM_028941752.1	*Prosopis alba*	EECGRQAGGALCPGGLCCSQYGWCGSTNEYCGTGCQSQCG
XM_028934376.1	*Prosopis alba*	EQCGRQAGGALCPGDLCCSQYGWCGSTDEYCGENCQSKCKG
XM_028940321.1	*Prosopis alba*	EQCGRQAGGALCPGGLCCSQFGWCGSTNDYCGPGCQSQCG
XM_028941724.1	*Prosopis alba*	EQCGRQAGGALCPGGLCCSQYGWCGSTNEYCGTGCQSQCG
XM_028918746.1	*Prosopis alba*	EQCGRQAGGALCPNGLCCSQYGWCGTTNDYCSNNCQSQCKPSG
EGLZ-2005642	*Prumnopitys andina*	QNCGCSSELCCSQYGYCGNGDDYCGSGCQEGPCYSNSVP
EGLZ-2020031	*Prumnopitys andina*	DPSCNPAGKIYCNPGRCCSKFNWCGTLESYCGKNYCIAQCP
EGLZ-2031851	*Prumnopitys andina*	ENCGRQAGGALCPGGACCSQWGWCGNTPDHCGTGCQSQCG
XM_021975002.1	*Prunus avium*	EQCGRQAGNAVCPNGLCCSQFGWCGTTADYCATGCQSQCASTP
XM_021952175.1	*Prunus avium*	GTAEQCGRQAGGALCPGGQCCSKYGWCGTAPDYCSTGCQSQCG
XM_021974997.1	*Prunus avium*	EQCGRQAGNAVCPNGLCCSQHGWCGTTADYCSTGCQSQCASTP
BQ641037.1	*Prunus dulcis*	EQCGRQAGNAVCPNGLCCSQHGWCGTTADYCATGCQSQCTS
XM_008243757.2	*Prunus mume*	EQCGRQAGNAVCPNGLCCSQYGWCGSTAAYCATGCQSQCASTP
FC864188.1	*Prunus persica*	EQCGRQAGNAVCPNGLCCSQHGWCGTTADYCATGCQSQCTS
AQFM-2001500	*Pseudolarix amabilis*	EDCGQQAGGALCPGGLCCSKWGWCGNTADHCGDGCQSQCG
UPMJ-2002634	*Pseudolycopodiella caroliniana*	AQQCGSQASGATCSNNLCCSKYGYCGSTAAYCGQGCQSGPCT
QKQO-2004260	*Pseudotaxiphyllum elegans*	EQCGSQADGATCPNNLCCSKWGYCGSTDAYCNPNEGCQSNCWG
YLPM-2037863	*Pseudotaxus chienii*	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP

YLPM-2033428	*Pseudotaxus chienii*	DPTCSPAGRFFCNTGRCCSKFNWCGSTAEYCASPKCIAQCWP
IOVS-2012852	*Pseudotsuga menziesii*	EQCGRQAGGALCPGGLCCSKWGWCGNTQTHCGQDCQSQCG
AB048531.1	*Psophocarpus tetragonolobus*	EQCGRQAGGGVCPGGLCCSKFGWCGSTAEYCGEGCQSQCWG
DNQA-2001518	*Psychotria douarrei*	EQCGRQAGNALCPNGLCCSQYGWCGSTDDYCKKGCQSQCRPG
BQEQ-2014152	*Psychotria ipecacuanha*	EQCGKQAGFALCPNGLCCSQYGWCGSTSEYCGDGCQSQC
PCNH-2069255	*Psychotria marginata*	EQCGRQAGYAVCPNGQCCSQYGWCGSTSDYCGDGCQSQC
HPXA-2008397	*Ptilidium pulcherrimum*	NCLISGCGGGGCCSAYGYCGTGYCGAAAGTCLNYGCP
HO708597.1	*Pueraria montana var. lobata*	EQCGRQAGGALCPGGLCCSQFGWCGTTNDYCGPGCQSQCG
QZZU-2064820	*Pyrenacantha malvifolia*	EQCGRQAGGALCPGGLCCSQFGWCGNTAEYCDKDKGCQSNCWG
QZZU-2062507	*Pyrenacantha malvifolia*	QQCGRQAGGRLCANNMCCSQWGYCGTTDDYCLPSNGCQSQCRG
IKFD-2051736	*Quassia amara*	EQCGRQAGGALCPDGLCCSEHGWCGDKPEYCGTGCQSQC
XM_031099390.1	*Quercus lobata*	EQCGSQAGGAVCPSGLCCSRFGWCGSTSDYCGNGCQSQCTSG
XM_031099388.1	*Quercus lobata*	EQCGSQAGGVVCPSGLCCSRFGWCGSTSDYCGNGCQSQCTSG
XM_031099393.1	*Quercus lobata*	EQCGRQANGAICPNGLCCSQHGWCGTTTEYCGNGCQSQCKP
XM_031099392.1	*Quercus lobata*	EQCGRQANGAVCPNGLCCSQHGWCGTTNDHCGNGCQSQCKP
CU656360.1	*Quercus petraea*	QQCGWQVGGKTCSNNLCCSQYGYCGTTDDYCSPSKNCQSNCQG
FN714255.1	*Quercus petraea*	EQCGRQANGAICPNGLCCSQHGWCGTTTEYCGNGCQSQCKPA
FN699861.1	*Quercus robur*	QQCGWQVGGKTCSNNLCCSQYGYCGTTDDYCSPSKNCQSNCQG
HENI-2049377	*Quercus shumardii*	QQCGWQVGGKTCSNNLCCSQHGYCGTTDDYCSPSKNCQSNCQG
XM_024026236.1	*Quercus suber*	EQCGRQAGGALCPGGQCCSQFGWCGTTADYCNTGCQSQCG
XM_024062919.1	*Quercus suber*	EQCGRQARGAVCANNLCCSQFGWCGNTPEYCGNTCQSQCTSG
XM_024030363.1	*Quercus suber*	EQCGRQVGGAVCVNNLCCSQFGWCGNTPEYCGNTCQSQCTG
XM_024035062.1	*Quercus suber*	EQCGSQAGGALCRNGLCCSKFGWCGSTSQYCGSGCQSQCSSG
XM_024026235.1	*Quercus suber*	EQCGTQAGGALCPDGLCCSQYGWCGTTADYCDNGCQSQCSG
XM_024072440.1	*Quercus suber*	EQCGTQAGGALCPNGLCCSQWGWCGTTAAYCDDGCQSQCSG
XM_024029648.1	*Quercus suber*	EQCGRQANGAVCPNGLCCSQHGWCGTTNDHCGNGCQSQCKP
OQHZ-2016253	*Quillaja saponaria*	EQCGKQAGGQVCPNGLCCSQFGFCGTTNEYCETGRGCQSQCKPSG
OQHZ-2019573	*Quillaja saponaria*	EQCGSQAGDALCPGGLCCSRFGWCGTTNDYCGDGCQSQCTG
RDOO-2115922	*Racomitrium varium*	EQCGKQAGGASCPGGLCCSQYGWCGTTPPYCEDGCQSNCG
BNCU-2002975	*Radula lindenbergia*	QNCGTEGGGALCQSGMCCSNFGYCGSGDAYCASGNCQSQCP

EY897555.1	*Raphanus raphanistrum subsp. Maritimus*	QQCGRQASGRTCAGNICCSQYGYCGTTADYCSPDNNCQSNCWG
XM_018635095.1	*Raphanus sativus*	EQCGRQAGGALCPNGLCCSEFGWCGNTEPYCKQPGCQSQCG
SWPE-2000451	*Reseda odorata*	QQCGRQAGGRTCANNLCCSQYGYCGTTDDYCSPSKGCQSNCRG
VGSX-2061426	*Retrophyllum minus*	DPTCSPAGRVYCNPGRCCSKFNWCGTTRANCGKGNCIAQCHP
EILE-2037280	*Rhamnus japonica*	QQCGQQVGGKTCSGNLCCSQYGYCGNTDDYCSPSKNCQSNCWG
ZTLR-2046308	*Rhizophora mangle*	EQCGKQAGGALCPGGLCCSQFGWCGDADAYCGAGCQSQCG
XM_030679960.1	*Rhodamnia argentea*	QQCGSQAGGAKCAGGLCCSKFGYCGSTPAYCADGCQSQCTP
JUL-2011100	*Rhodiola rosea*	AQICGKQAGGKTCAGNICCSQWGYCGTTDDYCLPSNNCQSNCKG
ZJUL-2011613	*Rhodiola rosea*	QNCGRQAGSATCANNQCCSQWGYCGTTDDHCLPSNNCQSNCRTEG
YUOM-2002381	*Rhus radicans*	EQCGRQAGGALCPGGQCCSQFGWCGTTSDYCTTGCQSQCG
YUOM-2007942	*Rhus radicans*	QHCGCGAHLCCSEFGYCGTGSAYCGSGCQGGPCYNERTP
JADL-2045621	*Rhynchostegium serrulatum*	EQCGTQAGGQLCPNNLCCSQYGYCGDTPPYCGENCQSQC
SYHW-2007441	*Ribes aff. giraldii*	EQCGKQAGGAVCPNGLCCSQFGFCGNTAEYCGTGCQSQC
WJLO-2036720	*Riccia berychiana*	VDCGSQAGGALCPNNQCCSQYGFCGTTPAYCGTGCQSQC
PAZJ-2060672	*Ricinus communis*	EQCGRQAGGKLCPNNLCCSQYGWCGSTDDYCSPSKNCQSNCKG
JK974151.1	*Robinia pseudoacacia*	EQCGRQAGGQTCPNNLCCSQYGYCGDTDDYCSPSKNCQSNCRG
HUSX-2031570	*Roridula gorgonias*	EQCGSQAGGALCPGGLCCSKFGWCGGTGDYCGDGCQSQCG
XM_024310382.1	*Rosa chinensis*	EQCGRQAGGASCPGGQCCSKFGWCGTTPDYCGTGCQSQCG
XM_024329626.1	*Rosa chinensis*	EQCGRQAGGATCPNGLCCSEYGWCGSTAEYCTKNCQSQCSSG
AYIY-2073961	*Ruellia brittoniana*	QNCGCDARLCCSRWGYCGSTIEYCGQGCQSGPCFTSPPSNG
IEPQ-2008285	*Salix dasyclados*	EQCGSQAGGALCPGGQCCSKFGWCGTTDAYCSGDNCQSQCG
IEPQ-2024372	*Salix dasyclados*	EQCGSQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG
GLVK-2018625	*Salix eriocephala*	EQCGSQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
RZTJ-2014561	*Salix fargesii*	EQCGSQAGGALCPGGLCCSKYGWCGNTDDYCGNGCQSQCG
RZTJ-208761	*Salix fargesii*	EQCGSQAGGQTCPNNLCCSQYGWCGDTDDYCSPSKNCQSNCKG
LFOG-2016379	*Salix purpurea*	EQCGRQAGGALCPGGQCCSKFGWCGTTDAYCSGDCQSQCG
TDTF-2019612	*Salix sachalinensis*	EQCGRQAGGALCPGGQCCSKFGWCGTTDAYCSGDCQSQCG
TDTF-2076763	*Salix sachalinensis*	EQCGSQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG
KKDQ-2079291	*Salix viminalis*	EQCGSQAGGQTCPNNLCCSQFGWCGDTDDYCSPSKNCQSNCKG
QRNU-2153740	*Sambucus canadensis*	QCGKQANDTLCPNNLCCSKWGFCGITDYYCGHGCQSQCG

GIWN-2008267	*Sarcobatus vermiculatus*	EQCGRQAGGAVCPNGLCCSQYGWCGNTNAYCGTGCQSNC
SERM-2050332	*Sarcodes sanguinea*	EQCGRQSGGALCPGGLCCSQFGWCGNGGDYCGTGCQSQCG
QCGM-2071011	*Saxegothaea conspicua*	ENCGRQAGGQVCNGEECCSQWGWCGTTPDHCGTGCQSNCG
IRBN-2159154	*Scapania nemorosa*	TEEGECDENNVCALGLCCSRFGYCGPGEEYCGSGCQGGSC
IRBN-2014956	*Scapania nemorosa*	DQGQCVNKTCPDALCCSKEGYCDTGDAYCGFGCQGGPC
IRBN-2159389	*Scapania nemorosa*	QGQCQNNICAAGLCCSQYGYCGNDSAHCGPGCLSGPC
OLES-2000665	*Schiedea membranacea*	EQCGRQAGGALCPGGLCCSQFGWCGSDNTYCGPGCQSQCG
OLES-2138639	*Schiedea membranacea*	EQCGTQANNALCPNGLCCSQFGFCGTTPAYCGTGCQSQCSP
WMUK-2091805	*Schizolaena sp.*	AQQCGWQVGGQTCPDNLCCSQYGYCGTTDDYCSPSKNCQSNCKG
IGUH-2160246	*Schwetschkeopsis fabronia*	ACSKTSPCANGLCCSQYGYCGTTAEYCGAGCQSQCG
YFZK-2040645	*Sciadopitys verticillata*	DPTCSPAGNFWCNSGRCCSKYNWCGDTEEYCGEGNCIAQCWA
YFZK-2040888	*Sciadopitys verticillata*	ENCGQQVGGGVCPGGECCSQYGWCGTTDAHCGAGCQSNCRTP
BPSG-2016915	*Scouleria aquatica*	ASCGSTAGGVLCPNNLCCSQYGYCGQTSAYCGTGCQSQC
MK044883.1	*Sedum alfredii*	QQCGRQAGGALCANGLCCSEHGYCGSTDPWCGAGCQSQCRG
LGDQ-2003229	*Selaginella apoda*	DPTCSPAGRQYCNPGRCCSKFNWCGTGRAYCGRGNCIAGPCWR
ZFGK-2039791	*Selaginella kraussiana*	EDCGRQAGGRTCPGGNCCSKWGWCGVTPDHCGDDCQSNCG
ABIJ-2008792	*Selaginella lepidophylla*	QDCGSQASFASCPPAQCCSQYGYCGTTAAYCGSGCQSQC
XM_002976162.2	*Selaginella moellendorffii*	EDCGRQAGGRTCPPGICCSKWGWCGVTTDHCGDGCQSQCG
XM_002987070.2	*Selaginella moellendorffii*	EDCGRQAGGRTCPPGNCCSKWGWCGVTPDHCGDGCQSQCG
ZZOL-2007633	*Selaginella stauntoniana*	EDCGRQAGGRTCPPGICCSKWGWCGVTPDHCGEDCQSQCG
QFAE-2004303	*Sequoiadendron giganteum*	ENCGRQAGGAVCPGGQCCSEWGWCGNTPDHCRVPGCQSQCG
QFAE-2000266	*Sequoiadendron giganteum Glaucum*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
HXJE-2113249	*Serenoa repens*	DPSCSPAGRQYCNDGRCCSKFNWCGTGAAYCGKGNCIGQCPP
HZTS-2101312	*Sessuvium portulacastrum*	QGDRVEQPMQLMPLWQKLGGPECGRQGGGKVCPNNRCCSRWGYCGDTDAF CGNGCQSQCDSG
OPZX-2050165	*Sessuvium ventricosum*	QGEVAEQVQVKPLSLWHKLTSFPTCGWQVLGRVCPDGRCCSKWGYCGDTE DYCGSGCQSNC
OXYP-2002914	*Sideroxylon reclinatum*	QQCGRQAGGALCENNLCCSQWGYCGTTDEYCLPSNNCQSNCRG
FZQN-2003040	*Silene latifolia*	EQCGRQAGGALCPGGLCCSKFGFCGSNDPAYCGEGCQSQCG

CVDF-2094468	*Simmondsia chinensis*	QRCGRQGGGQRCRGRLCCSNYGYCGTGSAYCAPGSCQSQCHH
VMNH-2021807	*Sinapis alba*	EQCGSQAGGALCPNGLCCSSVGFCGTTEPYCGPGCQSQC
MWYQ-2009109	*Smilax bona-nox*	EQCGSQAGGVTCPGGLCCSKFGYCGTTATYCGAGCQSQC
GHLP-2043688	*Solanum dulcamara*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
DLAI-2102397	*Solanum lasiophyllum*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG
BW688395.1	*Solanum lycopersicum*	QQCGRQARGRACANRLCCSQYGFCGSTRAYCGFGCQSNCRR
FS204664.1	*Solanum lycopersicum*	QQCGRQRGGALCGGNLCCSQFGWCGSTPEYCSPSQGCQSQCRG
BM413094.1	*Solanum lycopersicum*	QQCGSQAGGALCANGLCCSQYGYCGTTPDYCGQGCQSQC
FS021711.1	*Solanum melongena*	EQCGRQAGGARCAPGLCCSNFGWCGNTNDYCGAGKCQSQCPSG
FG646507.1	*Solanum phureja*	QQCGSQASGALCANGLCCSRYGYCGTTPAYCGPGCQSQC
NMDZ-2017071	*Solanum sisymbriifolium*	QQCGRQKGGALCAGNLCCSQFGWCGSTPEYCSPSQGCQSRCSG
NMDZ-2017724	*Solanum sisymbriifolium*	QQCGRQMGGALCGGNLCCSQYGWCGSTPEYCSPSQGCQSQCSGSG
FS109782.1	*Solanum torvum*	QQCGRQRGGALCDGGLCCSQFGWCGSTPEYCSPSQGCQSQCSG
FS110407.1	*Solanum torvum*	EQCGSQAGGALCPSGLCCSRFGWCGDTDAYCAPGNCQSQCRG
AJ878108.1	*Solanum tuberosum*	EQCGSQAGGAICASGLCCSKFGWCGNTDTYCGPGNCQSQCPS
CN216370.1	*Solanum tuberosum*	QQCGRQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCTG
JX683414.1	*Solanum tuberosum*	QQCGSQAGGALCANGLCCSEYGYCGTTTAYCGSGCQSQC
AM907316.1	*Solanum tuberosum*	QQCGSQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
DR751744.1	*Solanum tuberosum*	QQCGRQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCSG
AM909101.1	*Solanum tuberosum*	QQCGSQRGGALCGNNLCCSQFGWCGSTPEYCSPSQGCQSQCNG
LQJY-2003271	*Solanum xanthocarpum*	EQYGSHSPEGSCGSLADGAPCPSGLCCSKWGWCGDTDAYCAPENCASQCP
BLVL-2042202	*Sorbus koehneana*	EQCGSQAGGAVCPNGLCCSQYGWCGTTSDYCATDCQSQCG
HERT-2005714	*Sphaerocarpos texanus*	QDCGRQKNDEICRNNLCCSEHGFCGTGPEWCGAGCQSNCNTTSDDG
GOWD-2015396	*Sphagnum lescurii*	QQCGSQAGNALCANGLCCSQYGYCGSTAAYCGTGCQSQC
GOWD-2075409	*Sphagnum lescurii*	DPTCSPKGNFFCNSGRCCSVYNWCGSGASYCAKGNCLAQCWP
RCBT-2010499	*Sphagnum palustre*	QQCGEQADGASCENNLCCSEWGFCGTGDNYCGHGCQNGPCTV
KAWQ-2054190	*Stangeria eriopus*	QQCGSQAGGALCEGGRLCCSQWGFCGLGDPWCGQGCQSGPCSG
KAWQ-2005435	*Stangeria eriopus*	DPTCSPAGQVYCNSGRCCSKFNWCGSTAAYCQRPNCIAQCWP
PTLU-2099986	*Staphylea trifolia*	QQCGSQAGGRTCDNNLCCSQYGYCGTTDEYCLPSNNCQSNCKG
VBHQ-2003597	*Stemona tuberosa*	QNCGCAAGLCCSQYGYCGQGPAYCGTGCKGGPC

GGJD-2019841	*Strychnos spinosa*	EQCGSQAGNALCPNGLCCSQYGWCGSTPEYCTTGCQSQCG
FXGI-2053483	*Stylidium adnatum*	EQCGRQAGNVVCPNGLCCSQYGFCGSTPAYCGTNCQSQCG
MN064559.1	*Stylosanthes guianensis*	EQCGKQANGALCPNNLCCSQYGFCGNTNDYCGKGCQSQCKG
KLGF-2006233	*Sundacarpus amarus*	DPSCNPMAKVYCNPGRCCSKFNWCGTLESYCGKNYCIAQCP
WRPP-2126553	*Synsepalum dulcificum*	EQCGRQAQGALCPGGLCCSKFGWCGDTSDYCGDGCQSQCRP
WRPP-2126204	*Synsepalum dulcificum*	QQCGSQAGGALCANNLCCSQWGYCGDTDEYCLPSNGCQSNCRG
GRKU-2016385	*Syntrichia princeps*	QVDCGREANYALCANNQCCSQYGYCGTTDAYCGIGCQSQCP
NEBM-2079661	*Syzygium macranthum*	EQCGSQAGGKLCPGGQCCSKYGWCGTTPEYCTNGCQSQCG
QSNJ-2000100	*Taiwania cryptomerioides*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCASGNCLAQCWP
SILJ-2019357	*Talbotia elegans*	KEQCGRNASNAQCPNMMCCSGWGFCGSESAWCGEGCQSQC
LKKX-2068262	*Talinum sp.*	EQCGSQAGSALCPNKLCCSKFGWCGSTDQYCGPGCQSQCTPSG
HTDC-2009019	*Tamarix chinensis*	QEQCGSQAGGALCPGGLCCSQYGYCGTTSAYCGEGCQSQC
EH051455.1	*Tamarix hispida*	QCGTQAGGQLCPNNLCCSQYGWCGSTDDYCSQANNCQSNCNG
WWKL-2010000	*Tapiscia sinensis*	QQCGRQAGGRTCANNLCCSQYGYCGTTAEYCSQNCQSNCSPG
XM_010526614.1	*Tarenaya hassleriana*	EQCGRQAGGRTCPDNLCCSQYGYCGDTEDYCSPEKNCQSNCWG
GR933848.1	*Tarenaya spinosa*	EQCGRQAGGALCPNGLCCSQYGWCGDTEPYCGQGCQSQCG
FHST-2004844	*Taxodium distichum*	DPTCSPAGNFFCNSGRCCSIYNWCGTGPAYCGKGNCLAQCSP
FHST-2061111	*Taxodium distichum*	DPTCSPAGNFWCDSGRCCSIYNWCGSTSDYCASGNCLAQCWP
FHST-2007221	*Taxodium distichum*	DPTCSPAGNFWCNSGRCCSIYNWCGNTSDYCASGNCLAQCWP
FHST-2004843	*Taxodium distichum*	DPTCSPAGNFWCNSGRCCSIYNWCGSTSDYCAFGNCLAQCWP
FHST-2003453	*Taxodium distichum*	DPTCSPKGNFWCNNGRCCSIYNWCGTGPAYCSKGNCLAQCSP
AB096607.1	*Taxodium distichum var. distichum*	ENCGRQAGGAVCPGGLCCSQYGWCGNTPAHCQVPGCQSQCG
WWSS-2000781	*Taxus baccata*	DPTCSPAGNFWCNTGRCCSIYNWCGSTAAYCAQGNCLAQCWP
WWSS-2000780	*Taxus baccata*	DPTCSPAGRFFCNTGRCCSKFNWCGSTAEYCASRNCIAQCWP
ZYAX-2004409	*Taxus cuspidata*	EQCGRQASGVTCPGGLCCSKWGWCGNTPDHCGDGCQSQCG
CTSS-2001440	*Tellima breviflora*	QQCGRQAGGALCPNGLCCSQFGYCGSTPEYCTNGCQSQCG
CTSS-2043698	*Tellima breviflora*	QQCGRQAGGKTCQGNVCCSQYGYCGTTDDYCSPSKNCQSNCQG
CTSS-2001567	*Tellima breviflora*	QQCGRQAGGRTCANNLCCSQYGYCGTTAEFCTTNCQSNCKPG
AJJE-2095822	*Terminalia neotaliala*	EQCGRQAGGALCPGGQCCSTYGWCGTTPEYCSNGCQSQCTG
HVBQ-2128592	*Tetraphis pellucida*	EQCGSQAGGALCPNNFCCSQFGFCGTNPAYCGNGCQSGPC

BBBA-2081782	*Tetrastigma obtectum*	EQCGRQAGGKVCPGGACCSKYGWCGNTADYCGDGCQSQCSSTG
GBVZ-2064238	*Thalictrum thalictroides*	QNCGRQAGGATCPGNICCSQWGWCGTTDDHCLPSNNCQSNCRG
CU471310.1	*Theobroma cacao*	EQCGWQAGGTICPDNLCCSQYGWCGNTDAYCLPENNCQSNCKSSG
CU495512.1	*Theobroma cacao*	QQCGRQAAGRTCANNLCCSQFGYCGTTDEYCSPSKSCQSNCWPSG
XM_007030502.1	*Theobroma cacao*	QQCGRQAAGRTCANNLCCSQFGYCGTTNEYCSPSKSCQSNCWPSG
CU520991.1	*Theobroma cacao*	EQCGRQAGGALCPGGLCCSQFGWCGNTDDYCKKENGCQSQCSG
CU478322.1	*Theobroma cacao*	QNCGCAEGLCCSRWGFCGTGDDFCGTGCQEGPCNP
AXAF-2002983	*Thladiantha villosula*	EQCGRQANNAVCPNGLCCSQFGWCGNTAEYCTTGCQSQC
EEMJ-2032575	*Thuidium delicatulum*	QQCGRDVNNTLCADPLNCCSQYGYCGTTDAYCVTGCQSGPCRD
EWXK-2121685	*Thyrsopteris elegans*	EQCGSQAGGALCPDTLCCSEWGWCGATDNYCGEGCQSQC
SLOI-2002722	*Tiarella polyphylla*	QQCGRQAGGALCANGLCCSQHGYCGTTSDYCGTGCQSQCG
ZQRI-2065945	*Timmia austriaca*	SNCGSQAGGVICPNNLCCSSIGWCGQTSAHCGTGCQSQC
JG647038.1	*Tinospora cordifolia*	QQCGRQAGGRTCPGNLCCSQWGYCGDTNEYCSNNCQSNCRPG
HQOM-2016415	*Torreya nucifera*	QNCGCPTGLCCSQYGYCGTGDAYCGQGCREGPCYLSSP
HQOM-2011119	*Torreya nucifera*	DPTCSPAGNFWCNSGRCCSIYNWCGSTAEYCAQGNCLAQCWP
HQOM-2011118	*Torreya nucifera*	DPTCSPSGNFFCNDGRCCSRFNWCGTGPQYCGKGNCIAQCP
HQOM-2008514	*Torreya nucifera*	QNCGSQVDGRVCSGGECCSKWGFCGTTDDHCLPERGCQSNCRG
EFMS-2012045	*Torreya taxifolia*	DPTCSPAGNFFCNNGRCCSKFNWCGSGAQYCGKGNCIAQCP
EFMS-2069914	*Torreya taxifolia*	DPTCSPAGNFWCNSGRCCSIYNWCGSTAEYCAQGNCLAQCWP
EFMS-2004797	*Torreya taxifolia*	EQCGKQAGGALCPGGLCCSQWGWCGNTLIYCGNGCQSQCG
FITN-2088818	*Treubia lacunose*	EQCGSQAGGVLCANNLCCSQYGWCGSTSAYCGTGCQSQCG
TNVE-2099453	*Trianthemum portulacastrum*	EQCGRQAGGARCPTGLCCSQYGWCGTTNAYCGTGCQSGCPG
KVAY-2042228	*Tribulus eichlerianum*	EQCGRQAGGALCPGGLCCSEFGWCGDTDPYCGKGCQSQCG
BB912457.1	*Trifolium pratense*	EQCGSQANGAVCPNGLCCSKFGYCGNTDQYCGAGCQSQCK
AK455605.1	*Triticum aestivum*	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVASSG
BG604435.1	*Triticum aestivum*	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVV
M25536.1	*Triticum aestivum*	QRCGEQGSNMECPNNLCCSQYGYCGMGGDYCGKGCQNGAC
AK454179.1	*Triticum aestivum*	KCPNCLCCGKYGFCGSGDAYCGEGSCQSQCRG
DQ462308.2	*Triticum aestivum*	QLQSCPTRCGKQADGMECPNNLCCSKDGYCGLGVDYCSAGAGCQSG
BF293407.1	*Triticum turgidum*	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVA

AJ716574.1	*Triticum turgidum subsp. Durum*	QNCNCPAGMCCSQWGYCGTGPDYCGAGCQSGPCTVA
SWOH-2012447	*Trochodendron araliodes*	EQCGSQAGGALCPGGLCCSQFGWCGSTADYCNGGCQSQCSQTG
SWOH-2054532	*Trochodendron aralioides*	QQCGRQAGGRTCDGNLCCSQYGYCGTTDDYCSPSKNCQSNCQG
GH162518.1	*Tropaeolum majus*	APQCARQRIGGRCPKGQCCSRFGYCGSGPAHCAPANCYSQCKLTEP
GH167980.1	*Tropaeolum majus*	APQCGRQGIGGRCPKGQCCSRFGYCGSGPAHCAPANCYSQCKLTEP
GH166799.1	*Tropaeolum majus*	APQCGRQRIGGRCPKGLCCSRFGYCGSGPAHCAPANCYSQCKLTEP
GH162958.1	*Tropaeolum majus*	APQCGRQRIGGRCPKGQCCSRFGYCGSGPAHCAPANCYSQCKLTEP
GH170148.1	*Tropaeolum majus*	APQCGRQRISGRCPKGQCCSRFGYCGSGPAHCAPANCYSQCKLTEP
GH161739.1	*Tropaeolum majus*	APQCRRQRIGGRCPKGHCCSRFGYCGSGPAHCAPANCYSQCTLTEP
GH164293.1	*Tropaeolum majus*	QQCGRQAGGKTCRDNLCCSQYGYCGTTDDYCSPAKGCQSNCKG
MYZV-2053677	*Tropaeolum peregrinum*	QQCGRQAGGQTCRDNLCCSQYGYCGTTDDYCSPAKNCQSNCKG
GAMH-2003755	*Tsuga heterophylla*	EDCGQQAGGAVCPGGSCCSKWGWCGITQDHCGDGCQSQCG
GAMH-2003756	*Tsuga heterophylla*	QNCGEQAGGVLCPGDFCCSQWGWCGNTPDHCGDGCQSQCG
BRUD-2052794	*Typha latifolia*	EQCGRQAGGALCPNGLCCSQFGFCGNTDPYCGKGCQSQCG
DQ078281.1	*Ulmus americana*	EQCGSQAGGAVCPVGLCCSKFGWCGSTNEYCGDGCQSQCG
DQ078282.1	*Ulmus pumila*	DQCGSQAGGAVCPGGLCCSKFGWCGNTNEYCGDGCQSQCG
KOFB-2059461	*Urginea maritima*	QHCGRQAGGKLCPDNLCCSKWGYCGPTADHCGDECQSGPCYDKQ
PSJT-2036056	*Uvaria microcarpa*	EQCGRQAGGALCPGGLCCSQFGWCGNTPPYCDQGCQSQCSG
JZ158858.1	*Vaccaria hispanica*	FQCGRQAGGARCSNGLCCSQFGYCGSTPPYCGAGQCQSQCALD
JZ156691.1	*Vaccaria hispanica*	QNCGCDPDLCCSKFGYCGSSAAYCGQGCQSGPCDTPSG
MK292723.1	*Vaccinium angustifolium*	QQCGRQAGGKLCPGNQCCSQWGYCGTTDDYCLSSNNCQSNCKP
MK292724.1	*Vaccinium myrtilloides*	QQCGRQAGGKLCPGNQCCSQWGYCGTTDDYCLSSNNCQSNCKP
FFFY-2059513	*Valeriana officianalis*	EQCGTQAGGALCPGGHCCSQFGWCGTTVDYCTVPGCQSQC
HLJG-2152978	*Viburnum odoratissimum*	EQCGRQAGGALCPNGLCCSQYGYCGNTDDYCKNGCQSQC
FL503412.1	*Vicia faba*	EQCGSQAGGAVCPNGLCCSKFGFCGGTDAYCKDGCQSQCK
HX949455.1	*Vigna angularis*	EQCGSQAGGALCPGGLCCSQFGWCGSTDDYCGQGCQSQCG
XM_014653809.2	*Vigna radiata var. radiata*	EQCGKQAGGALCPNGLCCSKFGWCGDTDSYCGEGCQSQCKSG
XM_014646008.2	*Vigna radiata var. radiata*	EQCGRQAGGALCPGGLCCSQFGWCGSTDDYCGRGCQSQCG
XM_014653811.2	*Vigna radiata var. radiata*	QRCGSQGGGALCANGLCCSQHGWCGSTNEYCGTGCQSQCQP
FG848837.1	*Vigna unguiculata*	QQCGRQAGGRTCSGNLCCSQYGWCGNTEEYCSPSQNCQSNCWG

NJLF-2057700	*Viola canadensis*	EQCGSQAGFALCPNGLCCSQFGWCGSTSEYCGTGCQSQC
DMLT-2086965	*Vitex agnus-castus*	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKDSG
CF205030.1	*Vitis hybrid cultivar*	EQCGRQAGGKVCPGGACCSKFGWCGNTADYCGSGCQSQCSSTG
KP274873.1	*Vitis pseudoreticulata*	QQCGRQAGGRTCANNLCCSQYGYCGTTAAYCSPSQSCQSNCQSG
FQ438078.1	*Vitis vinifera*	QQCGRQAGGRTCANNLCCSQYGYCGTTAEYCSPSQSCQSNCQSG
CB920456.1	*Vitis vinifera*	QQCGRQASGKRCAGGLCCSQYGYCGSTRPYCGVGCQSQCRG
CB349756.1	*Vitis vinifera*	QQCGRQAGGKRCAGGLCCSQYGYCGSTRPYCGVGCQSQCRG
CB978544.1	*Vitis vinifera*	QQCGRQARGKRCAGGLCCSQYGYCGSTRPYCGAGCQSQCRG
TOXE-2009668	*Welwitschia mirabilis*	QNCGCPSDLCCSQYGYCGTGEAYCGTGCKEGPC
AUDE-2007932	*Widdringtonia cedarbergensis*	EQCGRQAGGQTCPNNLCCSQYGWCGNTEEYCSPSKNCQSNCWG
AUDE-2002285	*Widdringtonia cedarbergensis*	DPTCSPAGNFWCDSGRCCSIYNWCGDTSDYCASENCLAQCWP
RMWJ-2055404	*Wisteria floribunda*	EQCGSQAGGAVCPGGLCCSKFGWCGSTNEYCGKGCQSQCSG
EDEQ-2067968	*Wrightia natalensis*	EQCGRQAGNALCPNGLCCSQFGWCGDTDPYCKDGCQSQCNG
ZSSR-2108833	*Xanthicercis zambesiaca*	EQCGTQAGGALCPGGLCCSQFGWCGSTNDYCGPGCQSQCG
ZSSR-2105305	*Xanthocercis zambesiaca*	EQCGWQAGGQLCPNGLCCSQYGYCGSTADYCSPDKNCQSNCWG
YBML-2122140	*Yucca brevifolia*	QQCGSQAGGAVCPNGLCCSQYGYCGSTDPYCGNGCQSQCG
ZHEE-2005858	*Ziziphus jujuba*	QQCGRDVGGKTCDGNLCCSQYGWCGSTEEYCSPSKNCQSNCWG

Appendix E: Accession numbers and mature sequences of putative 8C-non-chitin-binding hevein-like peptides

Accession No.	Plant Species	Mature Sequence
DY543555.1	*Agrostis stolonifera*	CLPSGGFCMFRPTDCCGNCGCLYPMGVCYGSRCEE
DY543302.1	*Agrostis stolonifera*	CLPSGGFCMFRPTDCCGNCGCLYPVGVCYGSRCEE
WZYK-2086234	*Bazzania trilobata*	CLNGGGYCGSFTREACCYNCVCMMAFCVCG
AEXY-2113902	*Blasia sp.*	CLKNGEFCWGDPSGCCGNCGCLIIPGVCYGTGC
XR_003459507.1	*Coffea Arabica*	QEGECSPAGKPCRPVRCCDSCLCIVDYPTHVGTCRGNC
GR983294.1	*Coffea Arabica*	QEGECSPLGEACAGNPWGCCPGCICIWQLTDRCVGNC
GT020922.1	*Coffea Arabica*	QEGECSPLGEPCAGNPWGCCPGCICIWQLTDRCIGNC
GR990943.1	*Coffea Arabica*	QEGECSPLGEPCAGNPWGCCPGCICIWQLTDRCVGNC
GT692623.1	*Coffea Arabica*	QEGECSPLGKPCAGNPWGCCPGCICIWQLTDRCIGNC
GT021132.1	*Coffea Arabica*	QEGECSPLGKPCRYNPRGCCDFCVCVVADVTDEEGSCRGNC
GT021252.1	*Coffea Arabica*	QEPSCIPVGEPCAGNPGGCCDGCICIWQLTDRCAGSC
GT673445.1	*Coffea Arabica*	QEPSCLPAGESCTGNPWGCCPGCICIWQLTERCVGNC
XM_027261988.1	*Coffea Arabica*	RKDTCIGLLESCKDDPYGCCPGCVCLWPGDLCRGDC
DV668213.1	*Coffea Canephora*	QEGECSALGKPCRYNPSGCCGLCVCVIPDATDEGSCIGIC
DV678112.1	*Coffea Canephora*	QEGECSALGKPCRYNPSGCCGLCVCVIPDPTDEGSCIGIC
DV678117.1	*Coffea Canephora*	QEGECSPFGKPCRYNPWGCCDSCVCVATPADEGRCLGNC
DV666504.1	*Coffea Canephora*	QEGECSPLGEPCAGNPWGCCPGCICIWQLTDRCVGNC
DV688598.1	*Coffea Canephora*	QEGECSPLGKPCKYNPWGCCGSCLCIVDQPTHEGTCVGNC
DV667952.1	*Coffea Canephora*	QEGECSPLGKPCRYNPWGCCGSCLCIVDQPTHEGTCVGNC
DV687022.1	*Coffea Canephora*	QERRCIPALGSCVGNPRGCCFGCTCIRQLTDRCLGYCRI
GT664575.1	*Coffea racemosa*	GKDTCIGLLESCKDDPWGCCPGCVCLWPGDLCRGSC
GT668500.1	*Coffea racemosa*	QEEKCSPAGKPCRYNPRGCCDFCVCVVDFPGGEGSCLGNC
GT664298.1	*Coffea racemosa*	QEGECSPFGKPCRYNPWGCCGSCLCVVDHPTHEGTCVGNC
GT665189.1	*Coffea racemosa*	QEGECSPFGKPCRYNPWGCCGSCVCVVDHPTHEGTCLGNC
GT666921.1	*Coffea racemosa*	QEGKCSPAGKPCDPWGCCDFCVCVVDFPGGEGRCAGNC
GT664763.1	*Coffea racemosa*	QEPRCIPALGSCVGNPWGCCFGCMCIRQLTDRCLGYC
GT665740.1	*Coffea Rosmosa*	QEPRCIPVFGSCVGNPWGCCFGCMCIRQHTNRCLGYC
PZAP-2059688	*Eleusine coracana*	CIPMGGFCLFNLRGCCGSCGCLAGFCWRDASSCDL
CXSJ-2102568	*Eleusine coracana*	CIPMGGFCLGNLRGCCGSCGCLAGFCWRPASSCDS
FF591300.1	*Elymus wawawaiensis*	CISAGGSCFIHPKSCCGSCGCLYPIGICVGSSC
FF596360.1	*Elymus wawawaiensis*	CLPSGGFCMFRPKDCCGNCGCLYPIGVCYGSRCEE

44667515	*Eragrostis curvula*	CLPAGGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE
EH194220.1	*Eragrostis curvula*	CLPSGGFCMFHPKDCCGSCGCLYPIGVCFGSSC
EH184498.1	*Eragrostis curvula*	CLPSGGFCMFRPKDCCGSCGCLYPIGVCFGSSC
CO497362.1	*Gossypium hirsutum*	CIAKGGFCLFDLTSCCRPCGCLAGWCYNIDHDCNEYLPGR PLES
CO491697.1	*Gossypium hirsutum*	CIAKGGFCLFDLTSCCRPCGCLVGWCYNIDHNCNEYLPGR PL
CO499109.1	*Gossypium hirsutum*	CIAKGGFCLFDPTSCCRPCGCLAGWCYNIDHDCNEYLGRD HAK
CO497769.1	*Gossypium hirsutum*	CIAKGRFCLFDLTSCCRPCGCLAGWCYNIDHNCNEYLPGR PLES
CO498838.1	*Gossypium hirsutum*	CITKGGFCLFDLTSCCRPCGCLAGWCYNIDHDCSEYLGRD HAK
DW516599.1	*Gossypium hirsutum*	CKPKGSFCLFDLQSCCRPCGCLAGWCYNIDHDCNEYT
MVSE-2244972	*Griselinia littoralis*	CISSGGFCMFNPRDCCGSCGCLYPMGICYGSSC
JBYE-2151639	*Hedera helix*	CQPFDAPCDTFYGFYCCGSCTCTYVDFWHTSRCTGSC
CWYJ-2065743	*Heracleum lanatum*	CLPAGGFCMFRPMDCCGTCGCLYPVGVCFGNDC
OKEF-2088352	*Hibbertia grossulariifolia*	CSPLGGKCGDLVECCSGCVCIWPTYTCVGHC
OLXF-2088706	*Hibiscus cannabinus*	CIPKGGWCLFDIMGCCKPCGCLAGFCWVVGDDCN
YNFJ-2174268	*Microtea debilis*	CLNGGGYCGSFTREACCYNCVCMMAFCVCG
CUVY-2127511	*Mollugo nudicaulis*	CKSAGEWCGFSWTDCCNSCGCLAGFCYGTSC
AJFN-2096758	*Mydocarpus sp*	CIPKGGFCLFDLRGCCGMCGCLAGVCFNYDHPCEE
2.08E+08	*Oryza sativa*	CLPSGGFCMFRPKDCCGNCGCLYPIGVCYGSRCEE
2.18E+08	*Oryza sativa(Indica group)*	CGFCMYRPMDCCGNCGCLYPVGVCYGSRCEE
39571644	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE
2.19E+08	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCVCLYPVGVCYGSRCEE
2.18E+08	*Oryza sativa(Indica group)*	CLPAGGFCMFRRMDCCGNCGCLYPVGVCYGSRCEE
FG958636.1	*Oryza sativa(Indica group)*	CLPAGGFCLFRPMDCCGNCGCTYPVGVCYRSRCEE
FG955521.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCGCIYPVGVCYGSRCEE
FG951842.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCGCLYPVGVCTAHGVRNELYVR TCPFY
FG967928.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE
FG952362.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCGCPYPVGVCYGSRCEE
FG945740.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMDCCGNCSCLYPVGVCYGSRCEE
CF987442.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMGCCGNCGCLYPVGVCYGSRCEE
FG949814.1	*Oryza sativa(Indica group)*	CLPAGGFCMFRPMHCCGNCGCLYPVGVCYGSRCEE
FG953889.1	*Oryza sativa(Indica group)*	CLPAGGFCMYRPMDCCGNCGCLYPVGVCYGSRCEE
FG949901.1	*Oryza sativa(Indica group)*	CLPAGGYCMYRPMDCCGNCGCLYPVGVCYGSRCEE
FG953883.1	*Oryza sativa(Indica group)*	CLPARGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE

FG957880.1	*Oryza sativa(Indica group)*	CLPTGGFCMFRPMDCCGNCGCIYPVGVCYGSRCEE
CF324765.1	*Oryza sativa(japonica group)*	CLPAGGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE
2.84E+08	*Panax ginseng*	CKSGGSWCGFDPHGCCGNCGCLVGFCYGTGC
DV555424.1	*Panax ginseng*	CIPGGGFCMFEPLSCCVNCGCILVPGVCYCD
UNSW-2022849	*Panax notoginseng*	CKSAGEWCGFSWTDCCNSCGCLAGFCYGTSC
HANM-2065943	*Pholisma arenarium*	CISAGGFCFFDPMNCCGNCGCLYPVGICVGTNC
EDBB-2001794	*Polyscias fruticosa*	YIRESCIPLGGDCTDLFDCCPGCVCIITDLTCDGNCFRGA
EDBB-2001791	*Polyscias fruticosa*	YIRESCSPLGGKCGDLVECCSGCVCIWPTYTCVGHC
CA823563.1	*Populus deltoides x Populus trichocarpa*	CISSGGWCFTQPKNCCGNCGCLYPIGICFGSDC
BU827525.1	*Populus euphratica*	CISSGGFCFTQPMNCCGNCGCLYPLGICYGSDC
AJ780051.1	*Populus euphratica*	CISSGGWCFTQPMNCCGNCGCLYPIGICFGSDC
AJ780006.1	*Populus euphratica*	CLPSGGFCFFQPKNCCGNCGCLYPMGICFGSNC
DN489666.1	*Populus tremula x Populus tremuloides*	CISSGGFCFTQPMNCCGNCGCLYPLGICYGPDC
CA925073.1	*Populus tremuloides*	CISSGGFCFTQPMNCCGNCGCLYPLGICYGSDC
CA928793.1	*Populus tremuloides*	CISSGGFCFTQPVNCCGNCGCLYPLGICYGSDC
IEPQ-2024613	*Salix dasyclados*	CLPSGGFCMFQPMNCCGNCGCLYPIGVCYGSNC
CI251708.1	*Theobroma cacao*	CLPAGGFCMFRPMDCCGNCGCLYPVGVCYGSRCEE
CU482060.1	*Theobroma cacao*	CLFAGGFCMFNPLDCCGNCGCLFPLGICFGSGC
CU473132.1	*Theobroma cacao*	CLFAGGFCMFNPMDCCGNCGCLYPMGICFGSGC
CU482226.1	*Theobroma cacao*	CLFAGGFCMFNPMDCCGNCGCLYPMGICYGSGC
CU472771.1	*Theobroma cacao*	CLSAGGFCMFIPMDCCGNCGCLFLMGFCYGSGC
CU472807.1	*Theobroma cacao*	CLSAGGFCMFIPMDCCGNCGCLFPMGFCYGSGC
CU539930.1	*Theobroma cacao*	CLSAGGFCMFNPLDCCGNCGCLFPMGICFGSGC
CU519243.1	*Theobroma cacao*	CLSAGGFCMFNPLDCCGNCGCLFPMGICYGSGC
CU480659.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLFPMGFCYGSGC
CU483817.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLFPMGICYGSGC
CU473135.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLYPLGFCYGSGC
CU480710.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLYPLGICYGSGC
CU471377.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLYPMGFCYGSGC
CU521533.1	*Theobroma cacao*	CLSAGGFCMFNPMDCCGNCGCLYPMGICYGSGC
CU475183.1	*Theobroma cacao*	CLSAGGFCVFNPMDCCGNCGCLYPMGICYGSGC
CU474624.1	*Theobroma cacao*	CLSAGGSCMFNPMDCCGNCGCLYPMGICYGSGC
CU482998.1	*Theobroma cacao*	CPSAGGFCMFNPMDCCGNCGCLYPMGICYGSGC
CA821362.1	*Triticum aestivum*	CISSGGWCFTQPKNCCGNCGCLYPIGICFGSDC

CK206809.1	*Triticum aestivum*	CLPAGGFCMFRPMDCCGNCGCGYPAGVCYGTRCEE
BE417034.1	*Triticum aestivum*	CLPAGGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE
CA727718.1	*Triticum aestivum*	CLPASGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE
CA728959.1	*Triticum aestivum*	CLPSSGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE
CK218082.1	*Triticum aestivum*	CLPVGGFCMFRPMDCCGNCGCLYPAGVCYGTRCEE

Appendix F: Accession numbers and mature sequences of putative 10C-chitin-binding hevein-like peptides

Accession No.	Plant Species	Mature Sequence
AY453406.1	*Bambusa oldhamii*	QQCGSQAGGATCPDCLCCSQWGYCGSTADYCGDGCQSQCDGCG
XM_010233871.3	*Brachypodium distachyon*	EQCGSQAGGATCPDCLCCSRFGFCGSTSDYCGSGCQSQCSGCSSVVTP
XM_003569556.3	*Brachypodium distachyon*	EQCGSQAGGATCPNCLCCSRFGFCGSTSDYCGSGCQSQCSGCG
DN622484.1	*Citrus aurantium*	GNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCGE
FC920927.1	*Citrus clementina*	GNCGSGVVCPGGECCSRFGWCGLTIDHCCDGCQSNCNQVVCG
FC931884.1	*Citrus clementina*	GNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
AB081944.1	*Citrus jambhiri*	QNCGSGVVCPGGECCSQYGWCGLTTDHCCEGCQSNCDQVVCG
EY866470.1	*Citrus limettioides*	QNCGSGVVCPGGECCSQYGWCGLTTDHCCEGCQSNCNQVVCG
DC891630.1	*Citrus limon*	GNCGSGVVCPGGECCSQYGWCGLTTDHCCEGCQSNCNQVVCG
EY786360.1	*Citrus reticulata*	GNCGSGVVCPGGECCSRFGWCGLTIDHCCDGCQSNCNQVVCG
EY784150.1	*Citrus reticulata*	QNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
EY784928.1	*Citrus reticulata*	GNCGSGVVCPGGECCSRFGWCGLPIDHCCDGCQSNCNQVVCG
EY790130.1	*Citrus reticulata*	GNCGSGVVCPGGECCTLFGWCGLTIYHCCDGCQSNCSQVVCG
EY794233.1	*Citrus reticulata*	QNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCGE
CF506844.1	*Citrus sinensis*	GNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
CX044825.1	*Citrus sinensis*	QNCGSGVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCG
CF835995.1	*Citrus sinensis*	QNCGSPVVCPGGECCSRFGWCGLTTDYCCEGCQSNCNQVVCGE
EY887500.1	*Citrus sunki*	GNCGGGVVCPGGECCSRFGWCGLTTAYCCEGCQSNCNQVVCG
EY889333.1	*Citrus sunki*	GNCGSGVVCPGGECCSRFGWCGLTIDHCCDGCQSNCNQVVCG
CV710690.1	*Citrus trifoliata*	QNCGSGVVCPGGECCSQYGWCGITTDHCCGGCQSDCNQVVCG
EY829795.1	*Citrus trifoliata*	QNCGSGVVCPGGECCXQYGWCGITTDHCCGGCQSDCNQVVCG
DC899586.1	*Citrus unshiu*	GNCGGGVVCPGGECCSRFGWCGLTTAYCCEGCQSNCNQVVCG
NHIX-2010488	*Colchicum autumnale*	EQCGSQAGGATCPNGLCCSQFGFCGSTSDYCCSGCQSQCSCG
AHRN-2012568	*Cuscuta pentagonia*	QCGGQAGGALCANRWCCSRWGYCGLSCEYCGNSCQSYCHPDCRGV

FF360240.1	*Elymus spicatus*	AQKCGDQARGAKCPNCLCCGKYGFCGSTPDYCDVGCQSQCRGCRDG
AY277395.1	*Euonymus europaeus*	QQCGRQAGNRRCANNLCCSQYGYCGRTNEYCCTSQGCQSQCRRCG
EU837265.1	*Festuca arundinacea*	QQCGSQAGGATCANCLCCSQYGYCGSTSAYCGAGCQSQCNGCG
GDKK-2059433	*Gloriosa superba*	EQCGSQAGGATCPNGLCCSKYGFCGTTSDYCCSGCQSQCSCG
GTHK-2007511	*Gnetum montanum*	SILVTEVSSGGECNCVDKLCGEGFCCSKWGFCGTGEAWCGTGNCQC QCPS
AJ469417.1	*Hordeum vulgare*	EQCGSQAGGATCPNCLCCSRFGFCGSTSDYCGTGCQSQCNGCG
GH220479.1	*Hordeum vulgare subsp. vulgare*	EQCGSQAGGATCPNCLCCSRFGWCGSTSDYCGDGCQSQCSGCG
WXNT-2053439	*Joinvillea ascendens*	QQCGKQAGGKVCPNGLCCSQYGYCGSTSDYCCAGCQSQCNCG
UDUT-2001059	*Larrea tridentata*	EQCGRQAGGALCPGGLCCSQFGWCGDTPDYCCNGCQSQCGSGCG
GR513336.1	*Lolium perenne*	QQCGSQAGGATCANCLCCSQYGYCGSTSAYCGAGCQSQCNGCGG
XFJG-2002855	*Maianthemum canadense*	QQCGSQGGGAVCPNGLCCSQFGYCGTTSAYCCSGCQSQCSCG
XM_021755214.1	*Manihot esculenta*	EQCGDQAGGALCPEGQCCSQWGWCGSTADFCCEGCQSQCNPDNICFG
XM_021753285.1	*Manihot esculenta*	EQCGSQAGGAVCPGGLCCSKYGWCGSTTEYCCDGCQSQCRPNICG
XM_021755215.1	*Manihot esculenta*	EQCGYQAGGALCPEGQCCSQWGWCGSTADFCCEGCQSQCNPDNICFG
XM_021755212.1	*Manihot esculenta*	EQCGYQAGGALCPEGQCCSQWGWCGTTVEYCCDGCQSQCNPDDICG
XM_021755213.1	*Manihot esculenta*	EQCGYQVGGALCPEGQCCSQWGWCGSTVDYCCDGCQSQCNPDDICG
UASK-2013329	*Oenothera speciosa*	SQCGSQAGGAVCPDGLCCSKWGYCGTTDDYCCTGCQSQCTGCG
XM_006657298.2	*Oryza brachyantha*	EQCGSQAGGALCPNCLCCSKWGWCGSTSDYCGAGCQSQCSGCG
CI631870.1	*Oryza sativa (Japonica Group)*	EQCGSQAGGALCPNCLCCSQYGWCGSTSAYCGSGCHSQCSGSCGG
CI756576.1	*Oryza sativa (Japonica Group)*	EQCGSQAGGAVCPNCLCCSQFGWCGSTSDYCGAGCQSQCSAAGCG
FL945970.1	*Panicum virgatum*	QQQCGSEAGGKLCPNCLCCSQWGYCGSTSDYCGDGCQSQCHGCDVAS
XUAB-2018020	*Paraneurachne muelleri*	EQCGEQAGGALCPNCLCCSKWGWCGSTQDYCGEGCQSQCDGCG
MF581109.1	*Passiflora edulis*	AECGRQAAGVRCPGSLCCSQRGWCGTADEYCCISRGCQSQCHGCAAVG
TCYS-2014660	*Peliosanthese minor*	QQCGSQGGGAVCPNGLCCSQFGYCGTMPAYCCSGCQSQCSCG
FP101062.1	*Phyllostachys edulis*	EQCGSQAGGAACPNCLCCSRFGWCGSTDAYCGAGCQSQCSGCSG
NWQC-21549421	*Plagiochila asplenioides*	VRECGPSNPCKGGECCSQFGSCGCSNAYCGKGCIDNCNGCG
AF000964.1	*Poa pratensis*	EQCGSQAGGATCPNCLCCSKFGFCGNTSDYCGTGCQSQCNGCSG
AF000966.1	*Poa pratensis*	EQCGSQAGGATCPNCLCCSKFGFCGTTSDYCGTGCQSQCNGCSG
EY833419.1	*Poncirus trifoliate*	QNCGSGVVCPGGECCSQYGWCGITTDHCCGGCQSDCNQVVCGE

XM_024041040.1	*Quercus suber*	EQCGTQAEGALCPDGLCCSQYGYCGTTAEYCCTGCQSQCCG
XM_024041039.1	*Quercus suber*	VQCGTQAGGAFCPGGLCCSQYGWCGTTADYCCTGCQSQCSSDGCG
AB051578.1	*Secale cereale*	EQCGSQAGGATCPNCLCCSRFGWCGSTSDYCGDGCQSQCAGCG
XM_004966450.3	*Setaria italica*	EQCGSQAGGAVCPNCMCCSKWGWCGTTSDYCGTGCQSQCSGCG
GHLP-2050216	*Solanum dulcamara*	EQCGRQSGRRKCPNRLCCSKFGWCGTTCEYCGSGCQSNCRGGCATT
BE594394.1	*Sorghum bicolor*	EQCGTQAGGALCPNCLCCSKFGWCGSTSDYCGSGCQSQCTGSCG
GH724149.1	*Triticum aestivum*	DQKCGDQARGAKCPNCLCCGKYGFCGSGDAYCGEGSCQSQCRGCRD
BQ805737.1	*Triticum aestivum*	QCGSQAGGATCPNCLCCSRFGWCGSTSDYCGDGCQSQCSGCGSTP
JZ886728.1	*Triticum aestivum*	QTCGSQAGGARCPNCLCCSRFGFCGSGSEWCGAGCQSQCSGCPAP
AJ612726.1	*Triticum turgidum*	EQCGSQAGGATCPNCLCCSKFGFCGTTSDYCGTGCQSQCNGCSGG
YBML-2014359	*Yucca brevifolia*	QQCGSQAGGRVCPNAYCCSQYGYCGTTSDYCCSGCQSQCNCG
ICNN-2031036	*Yucca filamentosa*	QQCGREAGGRVCPNGLCCSQYGYCGTTSAYCCSGCQSQCNCG
AY532759.1	*Zea diploperennis*	EQCGSQAGGALCPNCLCCSQFGWCGSTSDYCGSGCQSQCSGSCG
AY532758.1	*Zea diploperennis*	EQCGSQAGGALCPNCLCCSQYGWCGSTSDYCGSGCQSQCSGSCG
FL135058.1	*Zea mays*	EQCGSQAGGALCPNCLCCSQFGWCGSTSDYCGSGCQSQCSGSCG