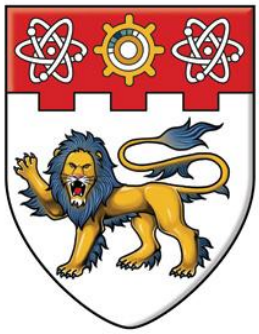


NANYANG TECHNOLOGICAL UNIVERSITY

STUDY OF MOLECULAR AND FUNCTIONAL DIVERSITY OF PLANT CYSTINE KNOT MINIPROTEINS

NGUYEN QUOC THUC PHUONG

SCHOOL OF BIOLOGICAL SCIENCES

2013

STUDY OF MOLECULAR AND FUNCTIONAL DIVERSITY OF PLANT CYSTINE KNOT MINIPROTEINS

NGUYEN QUOC THUC PHUONG

SCHOOL OF BIOLOGICAL SCIENCES

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

2013

Acknowledgments

The very first person who I am deeply indebted to is my eminent supervisor, Professor James P. Tam. First, he offered me the valuable PhD scholarship. The scholarship opened a new stage in my life and gave me the chance to explore a dynamic, advanced world outside my hometown during my study in Nanyang Technological University. My thirst for learning was assuaged and I did learn a lot during the past four years in Singapore under his guidance. More importantly, the scholarship fulfilled the dream of my late father. My father always wanted to see his own daughter go beyond the border of Vietnam and explore the world. The pride and happiness in my father's eyes four years ago always remind me of how meaningful this scholarship is to me. Thus, I sincerely thank Professor Tam for giving me a chance to make my father proud and happy. Second, my supervisor is my professional model. I have learnt from him not only professional knowledge but also the high ethics standard of a scientist. I have seen myself improve day by day through the hard course of doing research and writing papers under constant guidance and thoughtful encouragement of Professor Tam. Without him, I could never be what I am today.

I also wish to express my heartfelt thanks to Professor Dingxiang Liu, Dr. Huang Mei, and Mr. Fung To Sing. They have been giving me kind support and helpful instruction for all the molecular biology work, particularly the antiviral activity study. I felt in love with molecular biology thanks to their guidance and have learnt a lot from their expertise.

I am grateful to all the people who in many ways have made their contribution to enable this thesis at various stages: Dr. Nguyen Kien Truc Giang for initial experimental ideas and continuous help to get me familiar with the laboratory equipments, Mr. Zhang Sen for his useful guidance in cloning techniques at the beginning of my project, Mr. Antony Hardjojo, Mr. Cheong Yue Sheng, Mr.

Justin Ooi Seng Geap, Ms. Jolene Koh Hong Kiew, Mr. Chuan Huat Teo, Ms. Yunshan Lam, and Ms. Pee Yun Ting Anita for their kind assistance during their final year projects and attachment periods, and all other members of Professor Tam's group for their daily help and support. There are many more professors, researchers, staff and students at School of Biological Sciences whose names are not listed here yet I would like to extend my thanks to. They have in one way or another taught me how to do biological research and assisted me during my PhD life.

I sincerely thank all my friends in Vietnam and abroad, particularly Ms. Pham Tu Anh, Mr. Luu Minh Quan, and Mr. Pham Hung Thinh for their continuous care and encouragement whenever I am in need. With love, I thank my lovely little sister, Ms. Nguyen Quoc Mai Phuong, for her dear presence in Singapore, which has given me the motivation to stay strong and optimistic. She has always been my dearest little sister since I was born and till I am dead. I also thank Ms. Luu Thanh Thuy for being my best colleague ever and also a caring sister during the past two years. She had helped me a lot in the project of α-amylase inhibitors.

With all my heart, I dedicate my great appreciation to my boyfriend, Mr. Vo Minh Ngoc, who is always by my side and provides me moral support when I need it the most. He has been sharing with me joy and sadness, success and difficulty throughout the time with a sincere, deep love for me. He has showed me the wound-healing power of true love despite physical and geographical distance. Thanks to him, I regained my passion to live and to give.

Last but not least, I owe my deepest gratitude to my wonderful parents and other beloved family members for their love, patience and support during this long, challenging project. To me, the day I was born is my luckiest day because that is the day I started to have my Mom and Dad as my parents and brother and sisters as my siblings. I have done my best on the thesis because the success I have upon thesis submission is the gift I long want to give my Mom and Dad for their unconditional love and care. More importantly, I hope

they will smile on my graduation day because their smiles have always been the greatest achievement in my life.

Table of Contents

CHAPTER 2

MATERIALS AND METHODS

List of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

List of Tables

Abbreviations

AcN	acetonitrile
AEP	asparaginyl endopeptidase
AIDS	acquired immunodeficiency syndrome
AZP	artificial zinc-finger protein
CCK	cyclic cystine knot
CFU	colony-forming unit
CHCA	alpha-cyano-4-hydroxycinnamic acid
CK	cystine knot
CPP	cell-penetrating peptide
CRP	cysteine-rich peptide
CTAB	cetyl trimethylammonium bromide
CTPP	C-terminal propeptide
DAPI	4′,6′-diamidino-2-phenylindole
DCM	dichloromethane
DENV2	dengue virus type 2
DMEM	Dulbecco’s modified Eagle’s medium
dNTP	deoxyribonucleotide triphosphate
DTT	dithiolthreitol
EndoGlu-C	endoproteinase Glu-C

ER	endoplasmic reticulum (to indicate the signal peptide marking the proteins destined to endoplasmic reticulum)
GFCK	growth factor cystine knot
HD_{50}	hemolytic dose which hemolyzes 50% of a blood cell population
hpi	hour post infection
HPLC	high-performance liquid chromatography
HRP	horseradish peroxidase
HIV	human immunodeficiency virus (infection)
IAA	iodoacetamide
ICK	inhibitor cystine knot
IBV	infectious bronchitis virus
IC_{50}	half-maximal inhibitory concentration
EC_{50}	half-maximal effective concentration
kb	kilo base pairs
kDa	kilo Dalton
LB	Luria-Bertani broth
MALDI	matrix-assisted laser desorption ionization
MIC	minimum inhibitory concentration
MOI	multiplicity of infection
MS	mass spectrometry
MS/MS	tandem mass spectrometry

MW	molecular weight
NEM	N-ethylmaleimide
NTPP	N-terminal propeptide
PA1a	pea albumin chain a
PA1b	pea albumin chain b
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBST	phosphate buffered saline with 0.1% Tween 20
PCR	polymerase chain reaction
PFA	paraformaldehyde
PVDF	polyvinylidene fluoride
PCR	polymerase chain reaction
RP	reverse phase
RSV	respiratory syncytial virus
RT	room temperature
RT-PCR	real-time polymerase chain reaction
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
TCEP	tris[2-carboxyethyl]phosphine
TEMED	tetramethylethylenediamine

TFA	trifluoroacetic acid
T_m	melting temperature
UPLC	ultra-performance liquid chromatography
UTR	untranslated region

Abstract

Cystine knot (CK) peptides are structurally diverse and well represented in plant defense system. They are exceptionally stable to thermal denaturation and resistant to proteolytic degradation. As such, they can be used as scaffolds for engineering peptidyl drugs. The focus of my thesis is to isolate and characterize two classes of CK peptides by proteomic and genomic methods. These include α-amylase inhibitors from Apocynaceae and cyclotides from Rubiaceae, Fabaceae, and Poaceae families. From three Apocynaceae species, a family of 30-amino acid-residues α-amylase inhibitors was found, and they are the smallest CK α-amylase inhibitors known to date. In addition, we showed that they are entry inhibitor of a Class I virus by binding to the S and M proteins of infectious bronchitis disease.

For cyclotides which are end-to-end cyclic peptides, we studied their biosynthesis and showed that the processing enzymes may be responsible to generate both sequence and form diversity (both linear and cyclic versions). Our work suggests that CK peptides serve as a scaffold to generate molecular diversity in forms, sequences and functions as an evolution advantage for plant defense system.

Chapter 1

Introduction

Cysteine-rich peptide is a part of plant defense system

In response to challenges by pathogens and pests, plants have evolved many ways to defense themselves. The first line of self-defense is usually attributed to the thick, complex cell wall which might undergo dynamic changes under attacks [1]. Next, plants utilize their cellular innate immune system to ward off such assaults. The innate immune system in plants employs a wide range of chemical and biological compounds such as proteins, secondary metabolites, and reactive oxygen species [2]. These defense compounds could have either constitutive or induced expressions.

The inherent functional activities of plant defense compounds also benefit human. The use of medicinal plants for treatment of a variety of ailments has dated back to perhaps prehistoric times. It is estimated that about 25% of the prescribed drugs worldwide are obtained or derived from plants [3]. Their established usages in folklore remedy have guided ethnopharmacological studies in identifying biologically active principles, as recognized by the Traditional Medicine Division of the WHO (as cited by Rates, 2001 [3]). Small molecules (<500 Da), such as alkaloids, flavonoids, tannins, and polyphenols, have so far been the subjects of the vast majority of pharmacological, physiological and biochemical studies [4]. Some small-molecule drugs of plant origin have long been considered basic and essential in medicinal industry by World Health Organization (WHO), including quinine from *Cinchona* spp. as antimalarial drug, digoxin from *Digitalis* spp. for heart medication, and morphine from *Papaver somniferum* as potent analgesic compound. The favor

of small drugs by most pharmaceutical companies is attributed to their ease and low cost of production as well as good oral availability.

Recently, there has been growing interest towards alternative medication using macromolecules of natural sources, for example nucleic acids and proteins [5]. The shift of interest away from phytochemicals, either natural or synthetic, is explained by their limitation of small footprints. In many cases off-target effects, long-term toxicity, and thus ineffective therapies have been observed. Among the newly emerged macromolecular therapeutics, peptides and proteins, collectively known as biologics, have been receiving increasing attention owing to their high specificity, diversity and widespread occurrence in nature.

Biologics are peptides and proteins with therapeutic values. They exist widely in nature and function effectively along with small secondary metabolites. However, their activities have not been recognized until recently for two main reasons. First, most canonical proteins, such as enzymes and antibodies, have poor survival upon heat preparation and harsh digestion in the stomach. Second, the majority of known protein therapeutics have poor oral bioavailability, which has become a persistent challenge for drug administration [6]. However, cumulative evidence on cysteine-rich peptides (CRPs) suggests otherwise.

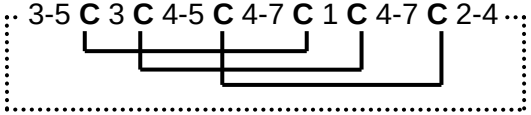
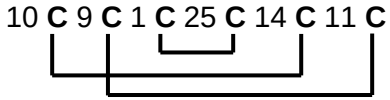
Cysteine-rich peptides are peptides with a high content of cysteine residues (typically 4, 6, or 8) which are crosslinked by multiple disulfide bonds into different cysteine motifs [7]. Thanks to the rigidified structures with a compact cystine core, most CRPs are structurally and thermodynamically stable [7, 8]. Many known CRPs have the amphipathic structruces, which likely promote their membrane permeability and improve their bioavailability. Furthermore, they are highly diverse in sequence and structure and have been found widespread in the plant kingdom. A sequence motif search through the genomes of the model dicot *Arabidopsis thaliana* and monocot *Oriza sativa* (rice) revealed that CRPs surprisingly take up 2-3% of the gene repertoire of

each species [9]. Taken together, CRPs likely contribute to the library of active components in phytomedicines.

Hammami and his coworkers [10] have documented sequence, structure, and bioactivity of most published CRPs in PhytAMP database, together with a minority of glycine – rich (sherpherin), glutamic acid – rich (vicillin-like family), arginine – rich (MBP-1 family) antimicrobial peptides. Based on their identities in primary sequence, cysteine pattern and spacing, nine CRP families have been clearly defined (Table 1.1). However, mounting incidences of CRPs identified in recent years suggest that many more are to be discovered by both proteomic and genomic approaches [11].

Table 1.1. Classification of plant cysteine – rich peptides. The size and consensus sequence of each family are summarized from available information in PhytAMP database after selected elimination of unrelated members. Numbers represent the range of intercysteinyl spacing in the backbone. The disulfide connectivities are shown by solid lines whereas the cyclic backbone of cyclotide is indicated by the dotted line. The disulfide connectivity of snakin remains unresolved.

Peptide family	Representative member	Plant species	No. of residues	CRP sequence	Reference
Plant defensin	Rs-AFP2	*Raphanus sativus* (radish)	45 – 57	1-9 **C** 8-11 **C** 3-5 **C** 3 **C** 9-11 **C** 4-8 **C** 1 **C** 3 **C** 0-7	[12]
Thionin	α - purothionin	*Hordeum vulgare* (barley)	42-47	2 **CC** 7 **C** 3 **C** 8-10 **C** 3 **C** 1 **C** 7 **C** 3-6	[13]
Lipid transfer protein	Ace-AMP1	*Allium cepa* (onion)	90 - 94	2-3 **C** 9 **C** 12-15 **CC** 17-18 **C** 1 **C** 20-23 **C** 13-15 **C** 4-5	[14]
Hevein (Chitin-binding peptide)	Ac-AMP2	*Amaranthus caudatus*	30-45	2-3 **C** 4 **C** 4 **CC** 5 **C** 6 **C** 1-2	[15]
Knottin – type	Mj-AMP1	*Mirabilis jalapa*	36-38	0-2 **C** 6 **C** 8 **CC** 3 **C** 10 **C** 3	[16]
Impatiens	Ib-AMP1	*Impatiens balsamina*	20	5 **CC** 8 **C** 3 **C**	[17, 18]

Peptide family	Representative member	Plant species	No. of residues	CRP sequence	Reference
Cyclotide	Kalata B1	*Oldenlandia affinis*	28-37		[19]
β-Barrelin (Macadamia)	MiAMP1	*Macadamia integrifolia*	76		[20]
Snakin	StSN1	*Solanum tuberosum*	63-67	4-7 **C** 3 **C** 3 **C** 8-9 **C** 3 **C** 2 **CC** 2 **C** 1 **C** 11 **C** 1-2 **C** 12-15 **C** 1	[21]

Inspired by the open venue for plant CRP applications in pharmaceutical industry, our group investigates CRPs from medicinal plants indigenous to Singapore and Southeast Asia using both proteomic and genomic approaches. Particularly, we have focused on the cystine knot family by virtue of its special features, which will be discussed in the following section.

Cystine knot miniprotein

Cystine knot miniprotein is a class of CRPs that possess a unique structural feature defined as a “cystine knot” (CK) [22, 23]. The cystine knot is a motif characterized by three intramolecular disulfide bonds, where two disulfide bonds and connecting backbones form an embedded ring penetrated by the third one [24, 25]. This fascinating arrangement of disulfide bonds provides a defined three-dimensional structure, which is typical for proteins, to these peptide-sized macromolecules. The cystine knot motif is prominently found in disulfide-rich domains of toxins, hormones, growth factors, and inhibitors [26].

Discovery and classification of the cystine knot

The term “cystine knot” was originally coined by McDonald and Hendrickson (1993) to describe the protomeric structures of several growth factors previously determined [27]. In their review, they noted that nerve growth factor (NGF), transforming growth factor β (TGFβ2), and platelet-derived growth factor BB (PDGF-BB) shared an unusual motif of knotted disulfide bonds associated with two pairs of antiparallel β strands. They believed that there were evolutionary and functional implications in the conserved structural motif among sequence-diverse proteins.

With the discovery of many cystine knot-containing peptides (CK peptides), the cystine knot was recognized as a versatile motif for proteins with different functional properties [25]. For examples, trypsin inhibitor EETI II [28], cone snail conotoxin ω-CgTx [29], and uterotonic agent kalata B1, though having

little sequence and function identity, were all found to contain the knotted disulfide arrangement. However, careful study of different cystine knot miniproteins revealed that they were not superimposable despite sharing the same connectivity of CysI-IV, CysII-V, and Cys III-VI (The cysteine residues involved in the knot are numbered from the N- to C-terminus) [30]. The discovery led to the introduction of “inhibitor cystine knot” (ICK) family by Pallaghy *et al.* (1994) and later formalized by Isaacs (1995) to indicate the CK peptides adopting some inhibitory activities, distinguishing them from cystine knot growth factors [25, 30]. ICK family included the “knottin” family earlier introduced by Le Nguyen *et al*. (1990) to group trypsin inhibitor EETI II, carboxypeptidase inhibitor CPI, and ω-conotoxin [31]. Later, with an increasing number of kalata B1 homologs identified, all having cyclic backbone in addition to embedded disulfide ring, Craik *et al.* (1999) suggested the classification for “cyclic cystine knot” (CCK) family [32]. Thus, cystine knot miniproteins currently comprise of three families (Figure 1.1):

- Growth factor cystine knots (GFCKs)
- Inhibitor cystine knots (ICKs)
- Cyclic cystine knots (CCKs)

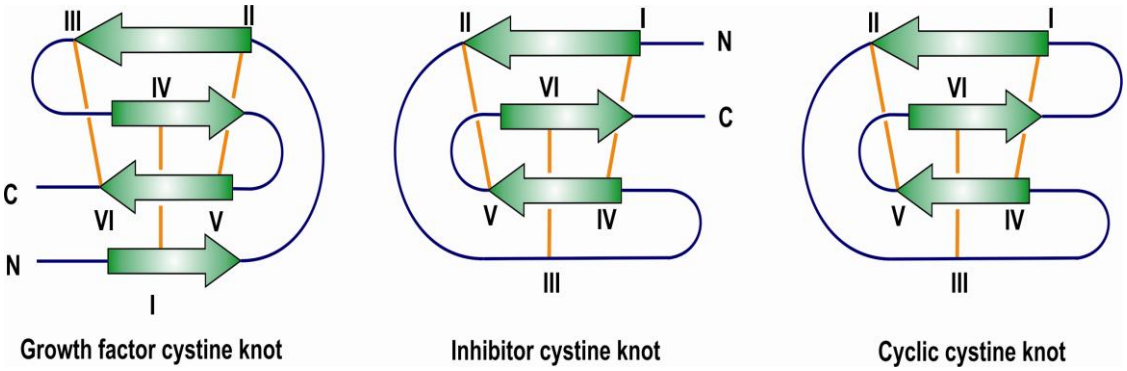


Figure 1.1. Schematic illustration of the structures of three cystine knot classes. The arrows represent the β-strands. The half cystines are numbered from the N- to C-terminus. The straight lines illustrate disulfide linkages. Comprehensive data on known members of inhibitor cystine knot and cyclic cystine knot families are documented at the KNOTTIN database [33] (http://knottin.cbs.cnrs.fr/) and the Cyclic Protein Database (Cybase at http://cybase.org.au), respectively. Overlapping profiles of many cyclotides can be found in both databases.

Basic structural features of cystine knot families

As seen in Figure 1.1, the three families differ from each other by the penetrating disulfide bond. This bond is the linkage between CysI-IV for GFCK family and between CysIII-VI for ICK and CCK families. CCK peptides have additional head-to-tail cyclized backbone as compared to ICK peptides. It is noted that only the connectivity of CysI-IV, CysII-V, and CysIII-VI is not a sufficient condition for the knot formation. For instance, the conotoxin GIIIB shares identical disulfide connectivity and performs similar function of a sodium channel blocker as the ICK conotoxin GS, yet does not contain a CK motif [34, 35]. It is believed that the spacing between half cystines involved in the knotted ring must also be considered because it tightens and stabilizes the structure. The cystine knot proteins are therefore sometimes classified based on the number of the residues in the ring, which ranges from 8 to 10 members for GFCK family [36] and 8 to 15 for ICK and CCK families [22]. To date, the ICK proteins have been found in animal and plants and the CCK proteins exclusively in plants.

An outstanding feature of cystine knot miniproteins is their high stability upon treatment in harsh conditions, such as boiling, intestinal or serum proteolytic digestion, and extreme acidic or basic conditions [37]. The GFCK miniproteins, however, are more susceptible to degradation perhaps because of their long, flexible protruding loops from the cystine knot core. With our program interest in heat-stable biologics which are potentially active compounds in herbal decoctions, the remainder of this thesis will focus on two stable CK families: CK α-amylase inhibitors and cyclotides, representing ICK and CCK proteins, respectively.

Cystine knot α-amylase inhibitor

Overview of α-amylase inhibitor

The family of α-amylases (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) is ubiquitous and widely studied. They digest starch, glycogen, and other α-linked polysaccharides to oligosaccharides, which are further hydrolyzed to monosaccharides by glucosidases to provide the basic energy supply of many organisms including human. The α-amylase inhibitors are thus important for regulating starch metabolism. They block carbohydrate digestion by inhibiting α-amylases and α-glucosidases.

In plants, they play an important role in the innate defense against predatory insects. In fact, α-amylase inhibitors have been shown to be crucial to many plant-microbe interactions as well as to the proliferation of pest and pathogens that feed on starch [38, 39]. The loss of crops and the growth of some pathogens were shown to correlate with the α-amylase production of the intruders [40].

Naturally occurring α-amylase inhibitors are present in plants and microorganisms. They vary greatly in both structure and size, ranging from small molecules (<1 kDa) such as carbohydrates, organic acids [41] and polyphenolics [42] to large proteins with the mass >56 kDa [43]. Both classes of α-amylase inhibitors have attracted attentions as tools in agriculture and antidiabetic management [44, 45]. Genes encoding amylase inhibitors have been successfully introduced into crop plants, rendering them pathogen-resistant. Transgenic pea plants expressing α-AI1, an amylase inhibitor from kidney bean (*Phaseolus vulgaris*), exhibited complete resistance to *Z. subfasciatus* and did not show any reduction in nutritional value [46]. In diabetes care, many small molecules of synthetic sugar-analogs inhibitors including acarbose, miglitol, and voglibose have been developed [47]. The synthetic acarbose has been commercialized for treating newly diagnosed type

2 diabetes as a first-line medicine [48]. The Phase 2®, a water-soluble extract of kidney bean containing the proteinaceous α-amylase inhibitor α-AI1 [49], is currently produced as weight-loss products of about 200 brands worldwide [50].

Classification of plant proteinaceous α-amylase inhibitor

One of the earliest studies on α-amylase inhibitors was the study on malted wheat α-amylase inhibitor by Chrzaszcz and Janicki in 1933 [51]. Since then, much has been known about various families of α-amylase inhibitors. Many reviews have been published, covering their occurrence, activity specificities, chemical and structural properties, inhibitory mechanisms, and applications of transgenic plants in crop protection and nutrition remedy [45, 52-54]. Dendrograms for some groups of inhibitors have been calculated to illustrate their structural and evolutional relationship with related proteins [53].

Proteinaceous α-amylase inhibitors are known for their structural diversity and are classified into six groups based on their similarities in sequences and three-dimensional structures by Franco *et al.* and Svensson *et al.* [52, 55]. They include knottin-type, γ-purothionin-like, CM-proteins, Kunitz-type, thaumatin-like, and legume lectin-like inhibitors. Some basic features of these inhibitor families are summarized in Table 1.2.

Plant α-amylase inhibitors, regardless of their structural classes, have different specificity and active forms. Some exhibit broad-spectrum activities (for example wheat inhibitor 0.19 [52, 56], common bean α-AI1 [52]), whereas some are specific (for example wheat inhibitor 0.53 [52, 56] and zeamatin [57] being specific to insect α-amylases). Most inhibitors act as monomers. Several inhibitors are active in homodimers (for example wheat inhibitor 0.19 and 0.53 [56, 58], $\alpha_2\beta_2$ form of *P. vulgaris* α-AI4 [43]) or heterooligomers (for example common bean α-AI1 [43]). Post-tranlational processing, such as proteolytic trimming or glycosylation, may be involved to produce the active forms of the CM- and lectin-like inhibitors [43, 59-61].

Given the discussed advantages of ICK pharmaceutically, our work focuses on the knottin-type α-amylase inhibitors as a representative of ICK family. From this point onwards, we use the term “cystine knot inhibitors” to refer to knottin-type inhibitors for a consistent nomenclature in this thesis.

Table 1.2. Classification of α-amylase inhibitors

Inhibitor type	Source	α-Amylase target	Size (aa)	No. of disulfide bonds	Representatives	References
Knottin-type (CK)[a]	Amaranth	Insect	32	3	AAI	[62, 63]
γ-Purothionin-like	Sorghum	Insect	47-48	4	SIα1, SIα2, SIα3	[64]
CM-proteins (Including LTP)[b]	Barley, wheat, ragi, rye, chili pepper	Mammalian, insect, bacterial	124-160	5	*Ca*-LTP$_1$	[65]
Kunitz-type	Barley, wheat, rice	Cereal, insect	176-181	1-2	BASI, WASI, RASI	[66-68]
Thaumatin-like	Maize	Insect	173-235	5-8	Zeamatin	[57, 69]
Legume-lectin-like	Common bean	Mammalian, insect	240-250	4 (intra) + 2 (inter)	α-AI1, α-AI2	[43]

a. Knottin-type inhibitors will be referred to as cystine knot inhibitors in this thesis.

b. The sequence and number of disulfide bonds of lipid transfer protein (LTP) from chili pepper Ca-LTP$_1$ remain unknown. The size of Ca-LTP$_1$ inhibitor was estimated about 9 kDa based on reducing Tricine-SDS-PAGE.

Characteristics of cystine knot α-amylase inhibitor

The first cystine knot α-amylase inhibitor was discovered in 1994 by Carugo *et al.* [70]. The compound identified in his work is Amaranth α-amylase inhibitor (AAI) from the crop plant *Amaranthus hypochondriacus* (Apocynaceae family), which remains the only member of CK α-amylase inhibitor family reported to date. AAI has 32 amino acids, currently being the smallest known α-amylase inhibitor, with six cysteines arranged in a typical cystine knot *abcabc* motif [70] (Figure 1. 2). AAI exhibited specific inhibition against insect enzyme, *Tenebrio molitor* α-amylase (TMA), while remaining inactive towards human α-amylases [62]. Apart from α-amylase inhibition activity, little is known about other biophysical and pharmaceutical properties of cystine knot α-amylase inhibitor.

The NMR structure of the free AAI and its X-ray crystal structure in complex with TMA have been determined previously [63, 71]. The structure in complex revealed a substrate mimicry action of AAI through the blockade of four out of six subsites in the central active crevice of TMA. Arg7 of AAI forms a strong salt bridge with the catalytic acid Asp287 of TMA. Its sidechain, together with the hydroxyl group of Tyr28, is connected to the other two catalytic residues, Glu222 and Asp185, via intricate water-mediated hydrogen-bonding network. A remarkable difference between the free and enzyme-bound structures of AAI is the conformation of Pro20. Upon binding to enzyme, the trans Pro in the free form undergoes a trans-cis isomerization, as also observed in hirustasin, a CK protease inhibitor [72, 73]. The isomerization increases the rigidity of the molecule due to the cis-Pro conformational constraint. This leads to concurrent conformational changes in two of three disulfide bonds and a large number of side-chain reorientations towards the amylase binding sites for the optimal enzyme-inhibitor interaction.

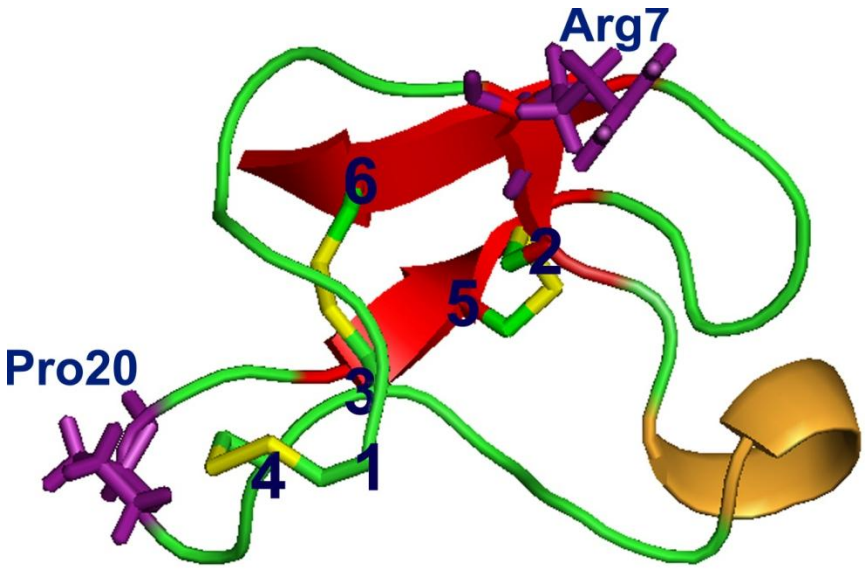


Figure 1.2. Three-dimentional structure of Amaranth α-amylase inhibitor.The disfulfide bonds are highlighted in yellow, β-sheet in red, and α-helix in dark orange. The Cys numbering is indicated. Two residues shown to be crucial for inhibitory activities, Arg7 and Pro20, are shown in sticks. The structure was obtained from Protein Data Bank (PDB number: pdb1qfd) and prepared by PyMOL program.

Bifunctional α-amylase inhibitor

Many α-amylase inhibitors have been reported to be bifunctional [74]. The first function is their activities against α-amylases from various sources. The second function in most cases is inhibitory activity against digestive proteases. For instances, RBI from finger millet [75] and zeamatin from maize [76] are α-amylase/trypsin inhibitors; RASI from rice [77-79] is α-amylase/subtilisin inhibitor. Chitinase activity has been occasionally found as the second function, such as for LCAI from Jobs tear's seed [80] and PvCAI from wild bean seeds [81]. A combination of α-amylase and protease/chitinase inhibition is believed to be useful for plant protection against pathogenic intruders and microorganisms [52].

To date, the reported bifunctional inhibitors are large molecules (>10 kDa). The inhibition sites of the two functions are independent for the double-headed inhibitor RBI because stable complex of α-amylase – RBI – trypsin has been observed [82]. Similarly, inhibitory mechanism at two different sites has been suggested for WASI [52]. Its α-amylase inhibitory activity was retained after incubation with proteinase K.

There has been no report on the second function for α-amylase inhibitors of <10 kDa, particularly the CK family. In this study, we discovered the antiviral effect of alstotides S1 and S3, CK α-amylase inhibitors from *A. scholaris*, against avian infectious bronchitis virus (IBV) through its binding to viral S and M proteins. The novel function, together with its mechanism studied in this work, provides additional evidence of the wide-spectrum defense function of CK inhibitors.

Cyclotide

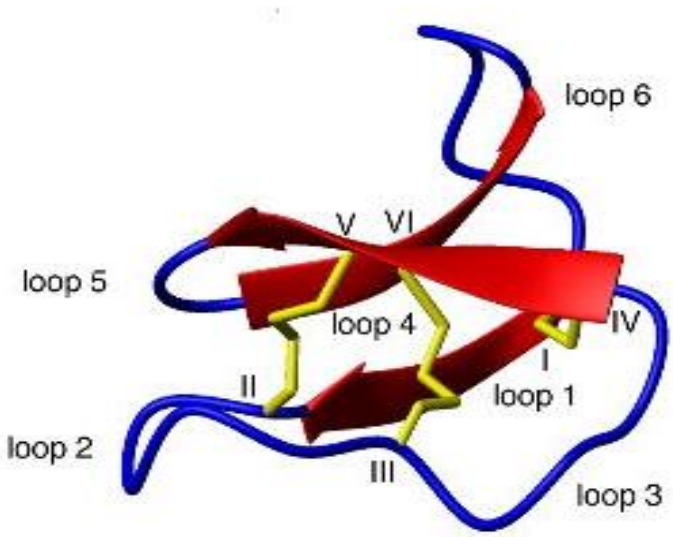
Discovery of cyclotide

Cyclotide is a family of cysteine-rich peptides recently receiving considerable attention from scientist community due to their exceptionally stable structure. Cyclotides have an end-to-end cyclized backbone of 28 to 37 amino acids with six strictly conserved cysteine residues engaged in a cystine knot [32]. The history of cyclotide dated back to the early 1970s when the Norwegian doctor Lorents Gran noticed the traditional use of *Oldenlandia affinis* decoction among tribal women in Congo to accelerate birth [83, 84]. The decoction stimulated uterus contraction after ingestion or topical application. Gran later determined the active principles to be a small peptide, kalata B1, with prolonged uterotonic activity and named it after the local name “kalata kalata” of the herb. The discovery of cyclotides has recently been reviewed extensively by Gran [85] and others [86, 87] in response to a frenzy of research on cyclotides.

The unusual cyclized backbone of kalata B1 had hindered any sequencing attempts using chemical techniques for more than 20 years since its discovery. The cyclic backbone and the cystine knot motif remained unknown till 1995, when Saether *et al*. used a combination of NMR and enzymatic cleavages to study this molecule [88]. Prior to that report, four of kalata B1 homologs were independently reported to possess diverse activities in 1993-1994 during drug screening programs [86]. Together, these molecules suggested the emergence of a family of cyclic peptides as promising therapeutic leads and stable drug scaffolds. It thus boosted the interest of scientists in exploring this class of macromolecules and soon led to the report on their novel chemical synthesis by Tam *et al.* in 1998 via thia zip reaction [89].

Structural characteristics and classification of cyclotide

A combination of head-to-tail cyclized backbone and a knotted disulfide bond arrangement defines a cyclic cystine knot (CCK). This unique structure was found to be conserved throughout the cyclotide family. It is the head-to-tail cyclization, the state of two termini ligated seamlessly, that inspired Craik *et. al* to coin the term cyclotide, aka “cyclopeptide”, in 1999 [32]. The backbone is cross-braced by three disulfide bonds into a cystine knot topology. Given their endless backbone, cyclotides have six intercysteinyl segments, normally referred to as “loops 1-6” and numbered from cysteine I towards cysteine VI (Figure 1.3). Thus, the knot ring is comprised of loop 1, loop 4, and two disulfide bonds (CysI-IV, CysII-V). Loop 1 and loop 4 are highly conserved among cyclotide members and typically contain three and one amino acids, respectively. It is believed that the eight-membered knot ring of cyclotides (4 cysteine and 4 intercysteinyl residues) represents the smallest ring that can be penetrated by a disulfide bond [86]. This topology provides a compact cyclotide core as compared to a flexible structure formed by a laddered arrangement of three disulfides such as that in heterodimeric RTD-1 [90]. It also gives rise to exceptional chemical, endoproteolytic and thermal stabilities while the cyclic backbone makes cyclotides resistant to exoproteases [91, 92].



Loop	**6**		**1**		**2**		**3**		**4**		**5**		**6**
Cys		**I**		**II**		**III**		**IV**		**V**		**VI**	
Möbius													
Kalata B1	GLPV	**C**	GET	**C**	VGGT	**C**	NTPG	**C**	T	**C**	SWPV	**C**	TRN
Bracelet													
Cycloviolacin O1	GIP	**C**	AES	**C**	VYIP	**C**	TVTALLG	**C**	S	**C**	SNRV	**C**	YN
Trypsin Inhibitor													
McoTI-I	GGV	**C**	PKILQR	**C**	RRDSD	**C**	PGA	**C**	I	**C**	RGNGY	**C**	GSGSD

Figure 1.3. Structure and classification of cyclotides. The structure of the prototypical kalata B1 illustrates the basic structure of cyclotides (from www.cyclotide.com). The Cys numbering scheme and connecting loop nomenclature are indicated. The disfulfide bonds are highlighted in yellow to depict the knotted structure formed by three interlocking disulfide bonds and the distorted triple - stranded β-sheet. Sequences of one representative from each of the three cyclotide subfamilies were given for comparison.

Cyclotides are classified into two main subfamilies, Möbius and bracelet, based on the presence or absence of a *cis*-Pro residue in loop 5, respectively. The peptide bond preceeding this *cis*-Pro in Möbius family causes a conceptually 180° structural twist in the circular backbone, resembling a Möbius chain [32, 93]. In contrast, the bracelet members have a *trans*-peptide bond at the corresponding position due to the lack of the *cis*-Pro. The third subfamily of cyclotides, the trypsin inhibitor cyclotides, was later added when two sequences *Momordica cochinchinensis* trypsin inhibitor I (MCoTI-I) and II (MCoTI-II) were discovered [94]. Despite their low sequence homology with known cyclotides, these two novel peptides possess a CCK motif and are thus classified into cyclotide family.

Occurrence of cyclotide

Since their discovery in 1970s to the 2000s, cyclotides were found in three families: Rubiaceae, Violaceae, and Cucurbitaceae. Rubiaceae is the fourth largest family with approximately 11,150 species in about 660 genera [95]. Meanwhile, the Violaceae family is much smaller, comprising about 900 flowering species in 23 genera [96]. A screening study of approximately 20 Violaceae species and >200 Rubiaceae species found a hit rate of 100% for cyclotides in every violaceous plant and of 5-10% in rubiaceous plants [97]. It is estimated that about 10,000 – 50,000 novel cyclotides in Rubiaceae family are to be discovered [97] though the real quantity may far outnumber the initial estimation. To date, there has been only one species from the Cucurbitaceae family, *Momordica cochinchinensis*, that was investigated for the presence of cyclotides. A total of four cyclic and three linear knottins have been identified from this cucurbitaceous species [98-100].

The expression level of cyclotide varies within each plant family. *O. affinis*, belonging to Rubiaceae family, has been thus far reported to have the highest

yield of 1-2 gram per kilogram of wet weight [101]. An expressed sequenced tag study on *O. affinis* plant revealed a distinct commitment of 2.8% of the plant transcriptome to encode cyclotide precursors [102].

Field collection and screening of more than 140 Rubiaceae– and Violaceae –related species suggest the potential existence of cyclotides in 12 Apocynaceae plants [103]. In fact, in this thesis we prove that the cyclotide-like hits for one of these 12 Apocynaceae plants were triggered by a family of cystine knot α-amylase inhibitors, not cyclotides. This α-amylase inhibitor family has also been found to distribute rather commonly in the Apocynaceae family. Thus, Apocynaceae family may not be a good candidate for cyclotide-producing family as suggested.

The search for cyclotides has also been performed by bioinformatics approach. Data mining on the genome of plants from families other than the three known cyclotide-producing families revealed a potentially wider distribution of cyclotides throughout the plant kingdom, for example in Poaceae family [104] (Figure 1.4). In 2011, independent works from our laboratory and Craik group provided evidence on the existence of cyclotides in the butterfly pea *Clitoria ternatea* species, making Fabaceae officially the fourth cyclotide-producing family [105, 106]. Very recently, we have reported cyclotide sequences from the grass *Panicum laxum (*Poaceae family), confirming cyclotide presence in the fifth plant family. As compared to other cyclic peptides reported to date for microorganisms [107, 108], plants [109, 110], and animals [111], cyclotides seem to constitute thus far the largest family with over 190 sequences, 1 X-ray and 30 NMR structures documented in KNOTTIN database [28, 29] and many more awaiting discovery. It has also been suggested that cyclotides may surpass the defensin family in number and diversity [112, 113].

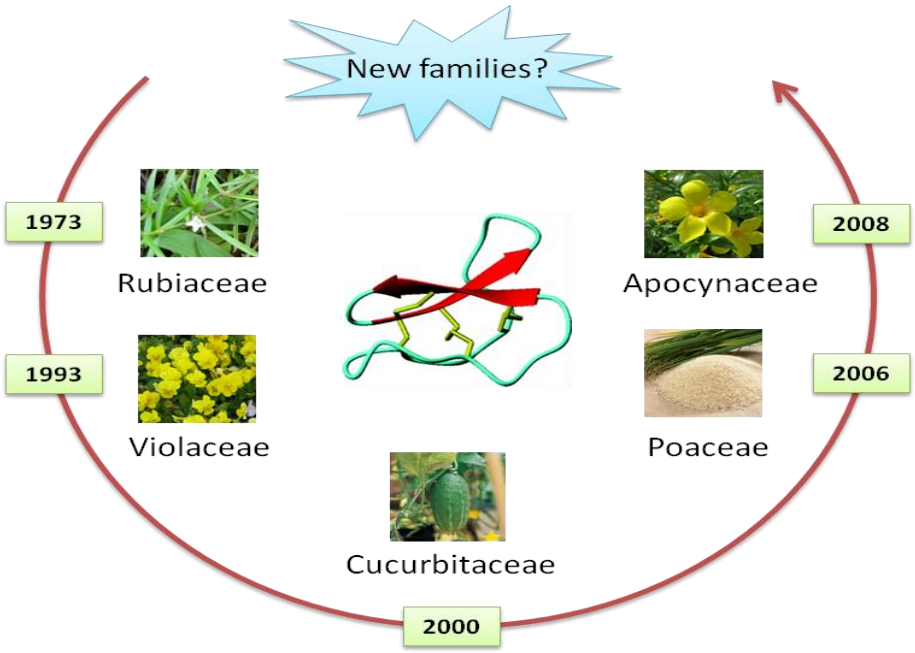


Figure 1.4. Summary of the discovery of cyclotides in various families by different methods. The prototypical member, kalata B1, was discovered after an observation of the birth-accelerating effect of the O. affinis decoction. Violapeptide I from Viola arvensis, the first violaceous cyclotide, was found in a hemolytic bioassay-guided screening in a search for saponins 20 years later. It took another 7 years for the identification of MCoTI-I and MCoTI-II from Momordica cochinchinensis, marking the introduction of the third subgroup of cyclotides with trypsin inhibition effect. Recently two novel families are reported to contain the cyclotide-like sequences in their genome (Poaceae) or to show mass spectra with peaks resembling cyclotides (Apocynaceae). The graph was built on the information available prior to this project.

Biological activity of cyclotide

A wide range of biological activities have been reported for cyclotides so far. They include cytotoxic [114-117], hemolytic [118, 119], antifouling [120], anti-HIV [121-127], antimicrobial [128], insecticidal [93, 129], uterotonic [85], anthelmintic [130, 131], and neurotensin antagonism activities [119]. The typical active concentrations for these functions are: antibacterial and antifungal minimum inhibitory concentrations (MICs) 0.2-50 µM, anti-HIV IC_{50} 100 nM, hemolytic HD_{50} >10 µM. The insecticidal properties of cyclotides suggest their main function in plants as host-defense agents [132].

Although further studies are still required to give insights into the detailed mechanism of each function, previous results have strongly suggested that cyclotides may exert their effects via membrane interactions [132]. NMR experiments determined that the membrane binding surface to the dodecylphosphocholine (DPC) micelle corresponds to the hydrophobic patch of kalata B1 and kalata B7 [133, 134]. The hydrophobic patch of the bracelet cycloviolacin O2 involves residues in loop 2 and 3, whereas the hydrophobic patch in the Möbius kalata B1 stretches over loop 2 and 5. Thus, despite sharing similar cystine knot motif, bracelet and Möbius cyclotides orient themselves in different membrane binding modes and have different activities [135].

Although the hydrophobic patch is critical for the biological functions, Ala and Lys scan on kalata B1 model revealed a “bioactive face” to be different from the hydrophobic patch [136]. The “bioactive face” for insecticidal activity comprises the Glu residue in loop 1 and the surrounding hydrophilic patch. Dye leakage and electrophysiological studies provided evidence on the pore formation of kalata B1 in the vesicles membranes [137]. Therefore, it is suggested that multimeric cyclotides, particularly in quartet, due to self-association in solution form pores at

membranes to exert biological effects following the channel-like mechanism (Figure 1.5) [137, 138].

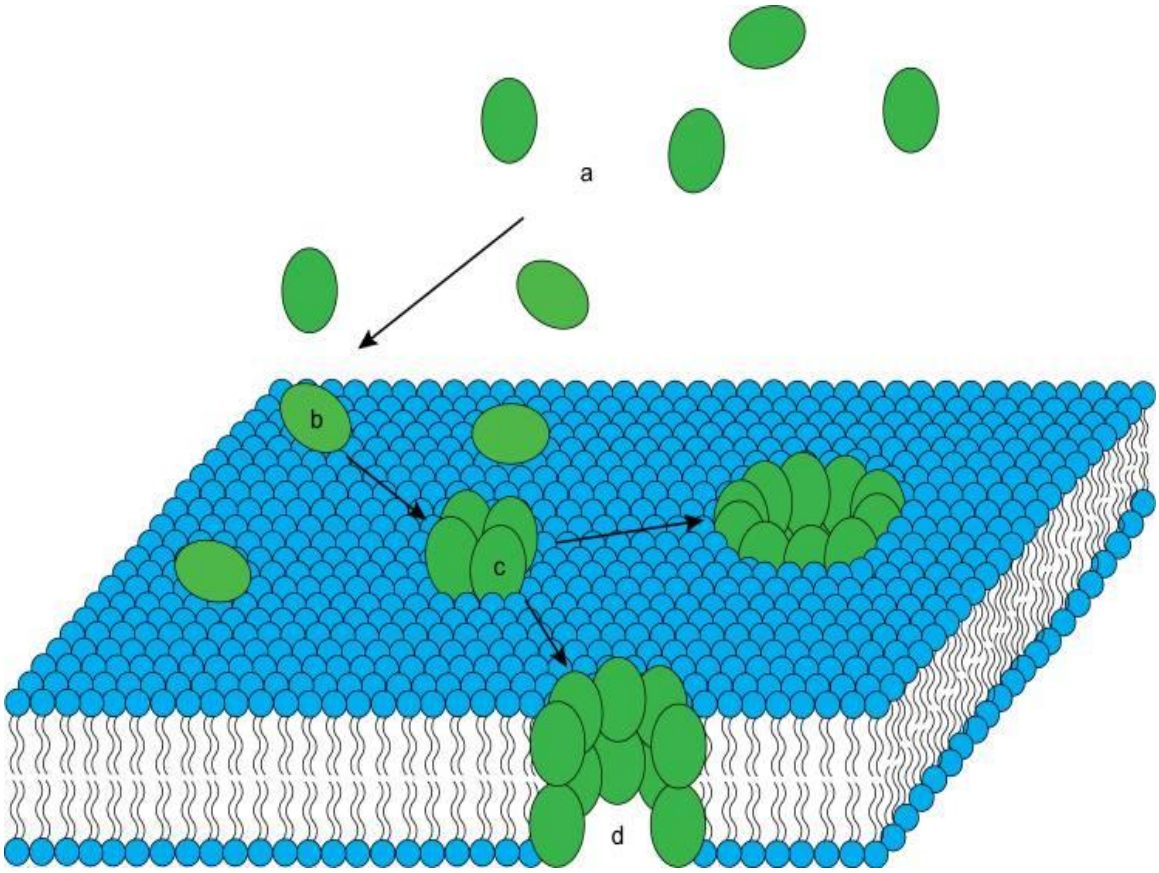


Figure 1.5. Proposed membrane binding mechanism of kalata B1. Kalata B1 exists as monomers in solution (a). Upon contact with the membrane surface (b), kalata B1 molecules diffuse laterally and aggregate to form tetramers or octamers (c), which is followed by pore-formation (d) to disrupt the membrane. (Figure taken from Huang et al. [137])

Biosynthesis of cyclotide

Cyclotides are gene-encoded miniproteins. They are ribosomally synthesized as large linear precursors and post-translationally modified to join the N- and C-termini. Prior to this work, cyclotide genes were reported for the Violaceae and Rubiaceae family to be remarkably similar. Cyclotide precursors generally comprise a signal peptide destined to endoplasmic reticulum (ER), a pro-peptide followed by single or multiple cyclotide domains and a short hydrophobic tail (Figure 1.6). Each cyclotide domain may contain an N-terminal repeat (NTR) sequence followed by a mature cyclotide sequence. In this project, we discovered the precursor structure of the Fabaceae and Poaceae cyclotides. To our surprise, the gene architecture of Fabaceae cyclotide precursors is completely different from those of Violaceae, Rubiaceae, and Poaceae families. The new gene structure gives hints about the evolution of cyclotides in the plant kingdom and will be discussed further in Chapter 5.

The cellular processing to produce mature cyclic cyclotides follows a multiple-step mechanism, as shown in the study of kalata B1 precursor by Gillon *et al.* [139] (Figure 1.5). The study suggested ER signal sequence to be cleaved first. ER signal sequence functions to direct cyclotide precursors to ER lumen co- or post-translationally (Figure 1.5). The cleavage of the signal peptide, possibly by SPase I [140], releases a stable 11-kDa product into the ER lumen. In this highly oxidative environment, the folding and disulfide bond formation occur with the assistance of residential reductases and chaperones to produce correctly folded prepropeptides. Particularly, protein disulfide isomerase (PDI), an ER chaperone, was shown to assist kalata B1 to obtain the native folding (CysI-IV, CysII-V, and CysIII-VI) out of 15 possibilities to arrange three disulfide bonds among six cysteines [141]. Misfolded peptides are retained inside ER lumen for subsequent ER-associated degradation [142, 143].

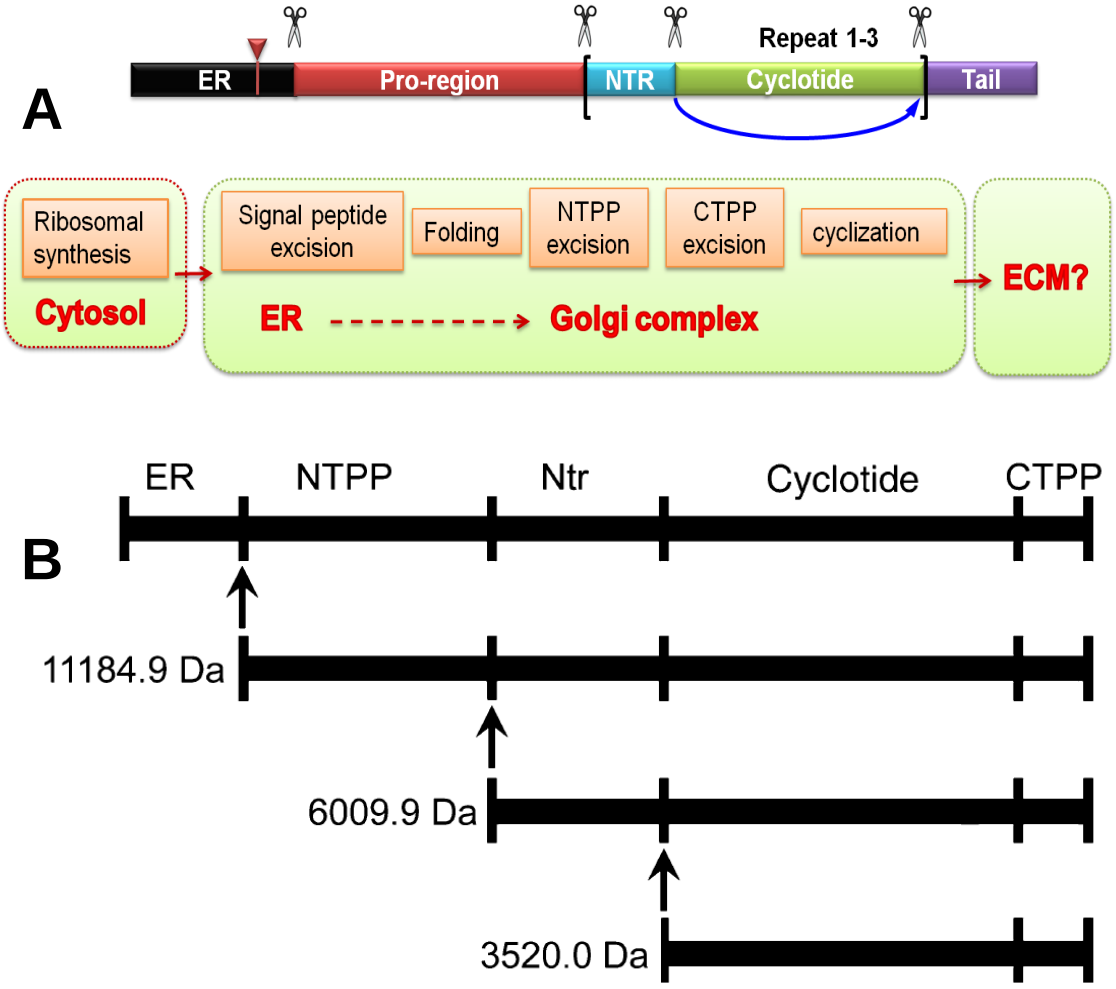


Figure 1.6. Proposed biosynthesis pathway of cyclotide. ***A****, Proposed biosynthesis of cyclotide. NTPP, N-terminal propeptide; NTR, N-terminal repeat; CTPP, C-terminal propeptide; ER, endoplasmic reticulum; ECM, extracellular matrix;* ***B****, Sequential processing sites, indicated by arrows, during biosynthesis of kalata B1. Figure 1.6B taken from Gillon et al. [139]*

The correctly folded prepropeptides are probably transported to the Golgi apparatus. The NTPP and NTR are sequentially cleaved from the precursor. In Gillon's study, both corresponding 6-kDa and 3.5-kDa products were detected (Figure 1.5). The NTPP domain is not quite conserved among reported cyclotide genes. Besides, there is only one NTPP in multi-domain cyclotide precursors. It is thus suggested that NTPP has no direct role on the processing of mature cyclotides. By contrast, there are several functions hypothesized for the NTR preceeding each cyclotide domain, which will be discussed in Chapter 5.

The synthesis of the mature cyclotides requires at least two more events: proteolytic cleavage at C-terminus and transpeptidation reaction to cyclize the backbone. Originally, many researchers have examined the possibility that cyclotides are cyclized via a nonenzymatic process, for example a self-excision via an intein-like mechanism. However, there has been no successful report so far to support this hypothesis. Instead, evidence from several works well established that these two events occur concurrently and are catalyzed by asparaginyl endopeptidase (AEP) [139, 144]. AEP is an endopeptidase that specifically cuts after Asn and, to a lesser extent, Asp to produce an acyl-enzyme intermediate. The ultimate Asn and Asp have been found almost completely conserved at the C-terminus of all cyclotides reported.

Linear cyclotide

In 2006, the discovery of the first linear analogue, violacin A from *Viola odorata* L. [145], surprised the research community. Its gene sequence revealed a premature stop codon at the position of the ultimate Asn/Asp residue, which is essential for backbone cyclization. Similar mechanism may also explain for the linearity of psyle C from *Psychotria leptothyrsa* [146] and kalata B20-lin from *O. affinis* [112] found in 2010. Both molecules lack this ligation residue at their C-terminus (Table 1.2). Two linear variants of kalata B9 and kalata B10 from *O.*

affinis were identified by MS/MS in 2007 [147]. However, Plan *et al*. suggested that these acyclic molecules result from the chemical cleavage of a labile Asn/Asp-Gly bond, which connects the two termini.

Although violacin A has more general flexibility in its backbone than the cyclic kalata B1 due to the distorted termini, it still maintains similar overall fold to its cyclic counterpart [145]. Violacin A remains resistant to endoproteases such as trypsin and thermolysin, suggesting that the cystine knot core, not the cyclic backbone, is responsible for this stability. In contrast, violacin A is susceptible to proteolytic cleavage by exoprotease aminopeptidase M, which unambiguously results from the lack of cyclic backbone.

Table 1.3. Linear cyclotides discovered from Violaceae, Rubiaceae and Solanaceae families

Uncyclotides/ Acyclotides	Species	Sequences	Ref.
Violacin A	*V. odorata*	SAIS**C**GET**C**FKFK**C**YT---PR**C**S**C**SYPV**C**K	[145]
Psyle C	*P. leptothyrsa*	KL**C**GET**C**FKFK**C**YT---PG**C**S**C**SYPF**C**K	[146]
Kalata B20-lin	*O. affinis*	GSIP**C**GES**C**VWIP**C**-ITAIAG**C**S**C**SNKV**C**YR	[112]
Phyb K	*P. x hybrida*	STD**C**GEP**C**VYIP**C**TITALLG**C**S**C**LNKV**C**VRP	[148]
Kalata B9-lin	*O. affinis*	GSVFN**C**GET**C**VLGT**C**YT---PG**C**T**C**NTYRV**C**TKD	[147]
Kalata B10-lin	*O. affinis*	GLPT**C**GET**C**FGGT**C**NT---PG**C**S**C**SSWPI**C**TR	[147]

Aims of the study

Cystine knot miniproteins are highly diverse and well represented in plant defense system. Insights into many aspects of these molecules greatly benefit their applications in drug development. This Ph.D. project was thus designed to study two classes of plant cystine knot miniproteins: CK α-amylase inhibitors and cyclotides. The specific aims are as follows:

1. To characterize cystine knot α-amylase inhibitors from three medicinal Apocynaceae species by proteomic and genomic approaches

The sequence and structure of the only known member of CK α-amylase inhibitors, AAI, have been reported. The occurrence of other members as well as their biosynthesis, unfolding pathway, and bioactivity profiles still remain unknown. This project aims to

(i) develop a routine screening and purification platform for CK α-amylase inhibitors using state-of-the-art instrumentation
(ii) characterize the CK inhibitors from different Apocynaceae species and their gene organization
(iii) study their biophysical and pharmaceutical properties for applications in medicine and agriculture.

2. To study antiviral mechanism of alstotide S1 using molecular biology techniques

In the study, we found that alstotide S1, a CK α-amylase inhibitor from *A. scholaris,* inhibited IBV proliferation in the cell-based system. To determine the detailed mechanism of this antiviral effect is thus appealing. It is of the interest of this project to

(i) determine the molecular mechanism of the anti-IBV activity of alstotide S1

(ii) study the specificity or broad-spectrum activity of alstotide S1 against other viruses.

3. To understand the molecular mechanism of how plants generate diversity within cyclotide family by genomic approach

Plants generate families of homologous peptides diverse in both sequence and form to serve a wide range of functions. This phenomenon is known as protein promiscuity and is utilized widely by plants for better survival and adaptation. In many cases, the post-translational modifications that determine the diversity are pre-encoded in the precursor sequences. It is thus essential and powerful to study the protein diversity in plants via genomic approach.

Our laboratory has identified 300 cysteine-rich peptides and the profiles of >300 medicinal plants using MS and MS/MS analysis (Wang S., Nguyen K.T.G, How J.G.A., J.P. Tam, unpublished data). Particularly in cyclotide-producing plants, a cluster of up to 50 MS peaks can be detected in one plant profile, implying that cyclotide is one of the most diverse groups known to date. Using cyclotide as the model, my thesis aims to

(i) develop a genomic approach to study the precursor sequences of cyclotides in plants

(ii) investigate the bioprocessing mechanism leading to the hyperdiversity in form and sequence of cyclotides

Chapter 2

Materials and Methods

Materials

Chemicals and reagents

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

Buffers, acids, bases, and salts	Sigma Aldrich (USA), Merck (USA), 1st Base (Singapore)
Organic solvent	Fisher Scientific (USA), Merck (USA)
DTT, TCEP, IAA, CHCA	Sigma Aldrich (USA)
Triton-X, agarose	Bio-Rad (USA)
LB media	BD (USA)
Cell culture media	PAA Laboratories (UK)
X-gal, dNTP mix	Fermentas (Canada, USA)
NeutrAvidin UltraLink bead	Thermo Scientific (USA)

Enzymes

The enzymes used for MS sequencing (trypsin, chymotrypsin, and EndoGlu-C) were of sequencing grade and purchased from Roche (Switzerland). The

enzymes for proteolytic stability tests were purchased from Sigma Aldrich (USA). The enzymes used in molecular cloning were of molecular biology grade and were bought from Fermentas (Canada, USA), Promega (USA), and NEB (UK).

Kits

The kits for molecular cloning were purchased from Invitrogen (Life Technologies, USA), Qiagen (USA), Clontech (Takara Bio, Japan), and Promega (USA). ACCQ-tag Ultra kit for amino acid analysis was from Waters (USA)

Cell lines, virus and antibodies

All the cell lines used in this thesis were received as kind gifts from the laboratory of Dr. Ding Xiang Liu, Nanyang Technological University, Singapore. These cell lines included African green monkey kidney (Vero) cells, p53-deficient human non-small lung cancer (H1299) cells, human hepatoma (Huh7) cells, baby hamster kidney (BHK) cells, and Madin-Darby canine kidney (MDCK) cells. Except for H1299 cells, all the cells were cultured in complete Dulbecco's modified Eagle's medium (DMEM, Hyclone) supplemented with 10% fetal bovine serum (FBS, Hyclone) and 1% penicillin/streptomycin (PAA Laboratories). H1299 cells were maintained in Roswell Park Memorial Institute medium (RPMI-1640) supplemented with 10% FBS and 1% penicillin/streptomycin. Cultured cells were incubated in 37°C, 5% CO_2 humidified incubator.

The virus stocks used in this study, including infectious bronchitis virus (IBV) strain Beaudette, human respiratory syncytial virus type A (RSV A), and dengue virus type 2 (DENV2), together with primary antibodies targeting IBV proteins were donated by the laboratory of Dr. Ding Xiang Liu. Primary antibody against RSV and DENV were the commercial products of Abcam (ab43812 and

ab90602, respectively). Secondary antibodies conjugated with horseradish peroxidase (HRP) were bought from Santa Cruz (USA).

Plant materials

All the plants included in this study were collected from throughout Singapore, Vietnam, or China and then cultivated in Nanyang Technological University Herbatorium for a timely collection. Different plant parts, such as leaves, flowers, seeds, buds, and roots, were harvested and used immediately or stored in ice water for 1-3 h prior to RNA extraction. For DNA extraction, freezed samples kept at -80°C for 1-3 months can be used instead. Voucher specimens were verified by Herbarium Department (Singapore Botanical Garden).

Plants used for cloning project include *Hedyotis biflora*, *Hedyotis diffusa, Hedyotis chrysotricha*, *Hedyotis uncinella, Hedyotis affinis (Oldenlandia affinis), Panicum laxum, Chassalia curviflora, Morinda citrifolia, Clitoria ternatea, Allamanda oenotheraefolia, Alstonia scholaris, and Wrightia religiosa.* The last three species were also used for large-scaled extraction of cystine knot α-amylase inhibitors.

Genomics

RNA extraction

The samples were subjected to total RNA extraction using Purelink^{TM} Micro-to-Midi Kit (Invitrogen) at a mini scale for almost all samples but *Chassalia curviflora* and *Morinda citrifolia.* Cysteine-rich peptides are only highly expressed in the fruits of *C. curviflora* and *M. citrifolia*, making RNA extraction using the kit unsuccessful. A rapid CTAB-based method developed by Gambino and his

colleagues [149] was employed instead. The qualified concentration and integrity of the isolated total RNA were tested by 1.0% agarose gel electrophoresis.

DNA extraction

The chromosome DNA extraction was performed using the PureLink™ Genomic Plant DNA Purification Kit (Invitrogen). For some plants such as *Hedyotis diffusa* and *Clitoria ternatea* which are rich in secondary metabolites, particularly polyphenolic terpenoids and tannins [150], the yield was low. A modified version of Kim's protocol (1997) [151] was thus followed to improve the yield and the quality of DNA extract. About 100 mg of leaf material was ground with the aid of liquid nitrogen. After grinding, 300 µl of extraction buffer (250 mM NaCl, 25 mM EDTA, 0.5% SDS, 200 mM Tris–HCl pH 8.0) with 1% (v/v) 2-mercaptoethanol was added to the homogenate which was then incubated at room temperature for 1 h. Freshly prepared 20% stock of polyvinylpyrrolidone (PVP, Sigma, MW 10,000) to make 6% of final volume and 150 µl of 7.5 M ammonium acetate were added separately. The mixture was left sitting on ice for 30 min to facilitate the hydrogen bonding and precipitation and then centrifuged at 10 000 x *g* for 10 min at 4°C. The supernatant was transferred to a fresh tube containing 1 volume of isopropanol (~ 450 µl). The tube was inverted for a few times, left at –20°C for 30 min, and centrifuged at 10,000 x *g* for 10 min. The supernatant was discarded, leaving the DNA pellet to be air-dried and subsequently resuspended in 500 µl TE buffer (10 mM Tris–HCl, 0.1 mM EDTA pH 8.0). Two µl of RNase (1 mg/ml) was added to the solution and incubated at 37°C for 15 min. Proteolysis was performed by proteinase K (20 mg/ml) at a final concentration of 100 µg/ml for 1 h in 50°C water bath with gentle shaking. One volume of chloroform – isoamyl alcohol (24:1) was added and completely emulsified before centrifugation at 10 000 x *g* at 4°C for 5 min. The supernatant was transferred to a fresh tube and the chloroform – isoamyl alcohol extraction was repeated once. The clear supernatant was finally transferred to a fresh tube to which 1 volume of ice-cold

isopropanol was added. The whole mixture was left at –20°C for 10 min and centrifuged at 10 000 x *g* for 10 min. After discarding the supernatant, the pellet was washed with 1 ml of 80% ethanol, vacuum-dried and redissolved in 30 µl of TE buffer or distilled water. DNA concentration and integrity were determined using NanoPhotometer™ (Implen, Northstar Scientific, Germany) and 1.0% agarose gel electrophoresis.

Rapid amplification of cDNA ends (RACE) and PCR analysis

The RNA extract served as the template to generate 3′ RACE (Rapid Amplification of cDNA Ends) cDNA population with GeneRacerTM Kit (Invitrogen, following the manufacturer's instruction. Degenerate forward primers were designed via a combination of CODEHOP [152, 153] or iCODEHOP [154] program and manual modification to achieve optimum degeneracy. The obtained PCR products amplified from degenerate primers and universal primer UAP/AUAP on a PTC-100 Peltier Thermal Cycler were run on 1.0% agarose gel electrophoresis. The target fragments were excised, purified with Wizard® SV Gel and PCR Clean-up System (Promega) and cloned with pGEM®-T Easy Vector System (Promega) into chemically competent DH5α cells, following the manufacturer's instruction. The insert-containing plasmids were sequenced via 1st Base Company's service.

Specific primers to walk upstream were derived from the newly found partial sequences and primed to 5′ RACE cDNA prepared with SMARTer™ RACE cDNA Amplification Kit (Clontech) (Figure 2.1). The whole procedure from the PCR reaction to the final sequencing step was repeated to complete the transcripts of interest.

In order to determine intron position and sequence on DNA template, specific primers were designed from obtained cDNA sequences. Forward primers were

derived from the 5′ UTR or ER signal peptide, and reverse primers were obtained from 3′ UTR or peptide domain (Figure 2.2).

A typical PCR setting on a PTC-100 Peltier Thermal Cycler was as follows: initial denaturation at 94°C for 5 min, main amplification for 35 cycles (denaturation at 94°C for 30 sec, annealing at T_m - 3°C for 30 sec, elongation at 72°C for 45 – 60 sec), final elongation at 72°C for 10 min, and cooling to 4°C. PCR products were tested on a 1.5% agarose electrophoresis, purified, cloned into pGEM-T Easy Vector, transformed into DH5α cells, and sequenced as previously described.

Sequence analysis

Sequencing results were analyzed using the BioEdit software. ExPasy translate tool (http://www.expasy.ch/tools/dna.html) was used to predict the amino acid sequences of the clones. Prediction of endoplasmic reticulum (ER) signal sequences of the precursor sequences was performed using SignalP 3.0 prediction server (http://www.cbs.dtu.dk/services/SignalP/) [155, 156]. Logo sequences were built with WebLogo application, version 3 (http://weblogo.berkeley.edu/) [157].

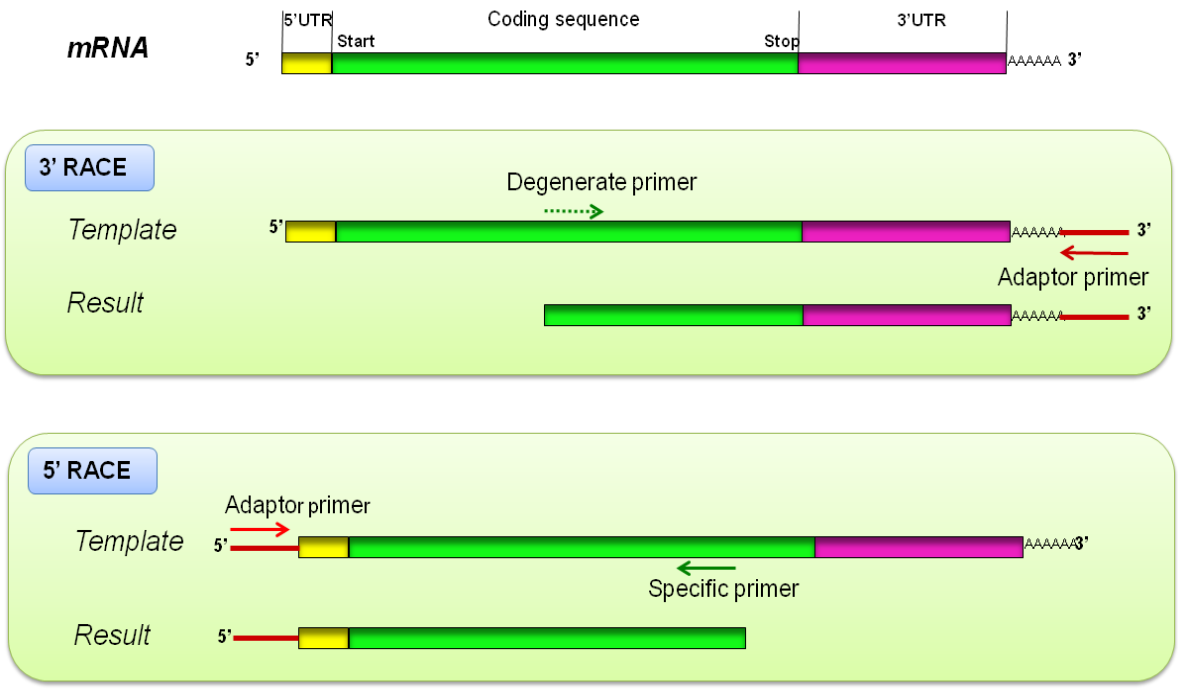


Figure 2.1. Strategy for genetic cloning at mRNA level. The general organization of mRNA is depicted, starting with 5′ untranslated region (5′ UTR), continuing with the coding sequence, and ending with 3′ untranslated region (3′ UTR) to which a poly A tail is added at the end of transcription process. The 3′ RACE cDNA is the template on which a combination of a degenerate primer and an adaptor were employed to first get the 3′ end partial sequence. The newly found segment became the base to design a specific primer to walk upstream 5′ RACE cDNA in order to achieve a complete transcript.

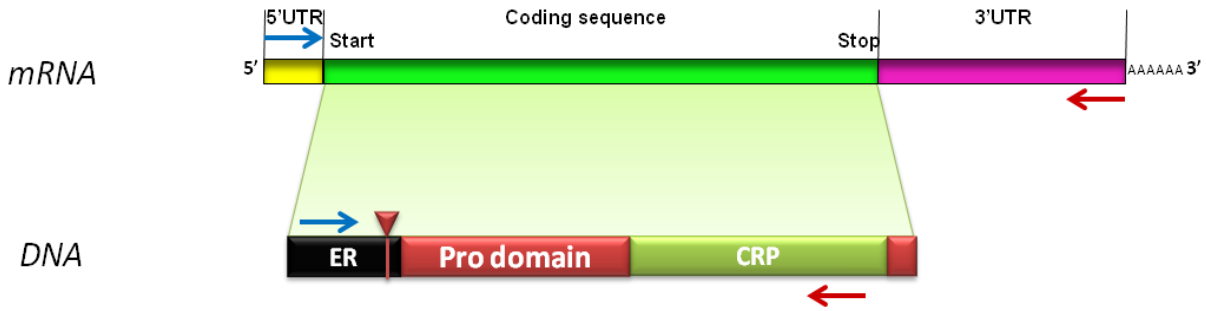


Figure 2.2. Strategy for genetic cloning at DNA level. To determine intron locations and sequences in the open reading frame (ORF), a pair of specific primers, including a forward primer designed either in 5′UTR or ER region and a reverse primer in 3′UTR or CRP domain, are utilized in a polymerase chain reaction. Intron information is later deduced from the difference between DNA and mRNA levels.

Proteomics

HPLC and UPLC

High performance liquid chromatography (HPLC) in this study was perform on Shimadzu system. The elution of peptides were monitored by UV detector at 220, 254 and 280 nm. For reverse-phase HPLC, Grace Vydac C18 columns (particle size 5 μm, pore size 300 Å) with dimensions of 250 x 22 mm, 250 x 10 mm, and 250 x 4.6 mm were used for preparative, semi-preparative, and analytical purifications at a flow rate of 6, 3, and 1 mL/min, respectively. For strong cation exchange (SCX), PolyLC polysulfoethyl A columm with dimensions of 250 x 9.4 mm and 250 x 4.6 mm were used at 3 and 1 mL/min flow rates, respectively.

MALDI-TOF MS and MS/MS

Mass spectrometry analysis for the crude extract and HPLC fractions were performed on the ABI 4800 MALDI-TOF/TOF system (Applied Biosystems, Framingham, MA, USA). The matrix used for matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) and MS/MS was comprised of saturated CHCA (~10 mg/mL) in 60% acetonitrile, 0.05% TFA. The samples were mixed with the matrix (ratio 1:1 v/v) on the spot of the target plate.

Both MS and MS/MS spectra were scanned in the reflection mode with laser intensity between 3000-4000. Accelerating voltages for MS and MS/MS were 20 and 8 kV, respectively. The mass spectra were averaged from 1000 and 5000 shots in MS and MS/MS, respectively.

Protein extraction and purification

800 g of leaves (*A. oenotheraefolia, A. scholaris* or *W. religiosa*) was homogenized and extracted twice in 50% ethanol (v/v). After centrifugation (7000 rpm, 10 min) to remove the plant debris, the supernatant was partitioned with dichloromethane (DCM) in the ratio of 2:1 to 3:1. The aqueous upper layer was concentrated, filtered, and loaded into C18 flash column (Grace Davison, US). Elution was done with increasing concentrations of EtOH (20%, 40% to 70% with 10% interval). The presence of 3-4-kDa cysteine-rich peptides in all fractions was monitored by MS. We confirmed the α-amylase inhibitory activity of these fractions by amylase assay.

To separate and to purify individual α-amylase inhibitors, several dimensions of SCX and RP-HPLC were employed. SCX utilized a linear gradient from buffer A (20% acetonitrile, 10 mM KH_2PO_4, pH 2.8) to buffer B (25% acetonitrile, 0.5 M KCl, 10 mM KH_2PO_4, pH 2.8) over 70 min. Inhibitors-containing fractions were subsequently purified using a Vydac C18 column with varied gradients. Mobile phases for RP-HPLC were: buffer A (0.1 % trifluoroacetic acid in water) and buffer B (100% acetonitrile, 0.1% trifluoroacetic acid).

De novo sequencing with MALDI-TOF MS/MS

About 40 µg of each purified peptide was dissolved in 50 mM ammonium bicarbonate (NH_4HCO_3) buffer (pH 7.8) containing 50 mM dithiothreitol (DTT) and incubated at 37°C for 2 h. Digestion with trypsin or chymotrypsin at a final peptide/enzyme ratio of 80:1 (w/w) was done at room temperature (RT) for 5 min. *De novo* discovery of CRP sequences was based on both b- and y- ion series in the tandem mass spectrometry (MS/MS) profiles.

Quantitative amino acid analysis

20 µg of peptide was fully reduced by incubation with 50 mM DTT in 50 mM NH_4HCO_3 buffer (pH 7.8) for 1 h at 37°C. Double amount of iodoacetamide (IAA) as compared to DTT was added to the reduction mixture. The S-tagging was allowed to proceed at 37°C for another hour. The alkylated peptide was subsequently purified using UPLC. After removing the solvent by lyophilization, peptide was digested by chymotrypsin in 50 mM NH_4HCO_3 buffer. Subsequently, lyophilized peptide fragments after UPLC separation were hydrolyzed by 6 N HCl, 1 % phenol, in vacuum at 110°C for 24 h. Derivatization and quantification of amino acids were done following AccQ·Tag_{ultra} UPLC method [158].

Small-scale extraction to screen for the presence of cystine knot inhibitors in the Apocynaceae family

To study the potential expression of knottin inhibitors in the Apocynaceae family, different plant parts of nine Apocynaceae species available in Nanyang Technological Univeristy campus were collected. The tested plants included *Allamanda oenotherafolia (golden trumpet), Cerbera odollam* (suicide tree)*, Plumeria rubra* (frangipani), *Allamanda cathartica* (golden trumpet), *Nerium oleander* (oleander), *Wrightia religiosa* (water jasmine), *Wrightia antidysenterica* (arctic snow), *Alstonia scholaris* (blackboard tree), and *Plumeria obtusa* (frangipani). Small-scale extractions were performed on RP-HPLC column.

Briefly, about 1 g of plant sample was homogenized in 500 µl of 50% ethanol (v/v). The extracts were diluted three times in water to increase solvent polarity and loaded to C18 solid phase extraction column. The column was washed with 20% acetonitrile to desorb any hydrophilic substances. The putative CK-peptide-containing fractions were eluted with 80% acetonitrile. Their occurrence in the eluate was profiled using MALDI-TOF MS.

Disulfide mapping

In this study we determined the disulfide connectivity of two knottin α-amylase inhibitors: allatide O2 from *A. oenotheraefolia* and alstotide S3 from *A. scholaris*. The routine protocol used for allatide O2 as described below was unsuccessful to determine the disulfide connectivity of alstotide S3. Thus, a more elaborate protocol with remarkably shallow HPLC gradient was applied to separate unstable intermediates.

Disulfide mapping for allatide O2

Allatide O2 (0.2mg) was partially reduced in 100 mM citrate buffer (pH 3.0) with 20 mM tris(2-carboxyethyl) phosphine (TCEP) at 37°C for 30 min. N-ethylmaleimide (NEM) was added to a final concentration of 40 mM and incubated at 37°C for 40 min. The reaction was quenched by loading the sample into a Vydac analytical C18 column (250 x 4.6 mm) and NEM-alkylated intermediate species separated with a linear gradient of 10-60% buffer B. Intermediate species were fully reduced with 20 mM DTT (37°C, 40 min) and *S*-alkylated with 40 mM IAA (37°C, 30 min) before purification by HPLC. *S*-alkylated peptides were then subjected to trypsin digestion and MS/MS sequencing.

Disulfide mapping for alstotide S3

About 0.3 mg of alstotide S3 dissolved in 0.5 ml of 30% acetonitrile/H_2O and two-fold volume of 40 mM tris(2-carboxyethyl)phosphine (TCEP) in 200 mM citrate buffer, pH 3 were incubated at 75°C for 3.5 min. All reagents were equilibrated to the reaction temperature before mixing. The partial reduction was quenched by immediate injection onto analytical RP-HPLC column and eluted with shallow gradient focusing the 38-40% fraction in 30 min.

Alkylation of reduced cysteine residues was carried out by adding an excess amount of NEM directly to the HPLC fraction and incubating the mixture at 45°C for 45 min. Different NEM-alkylated species were then separated by RP-HPLC focusing the 38.5 – 40% fraction in 1 h. MS scanning of the eluted fractions showed that we obtained alstotide S3 alkylated with two to six NEM groups.

For complete reduction and alkylation, fractions that contain alstotide S1 alkylated with two/three and three/four/five NEM groups respectively were added with equal volume of 40 mM dithiothreitol (DTT) in 25 mM NH_4HCO_3 buffer, pH 7.8. After a 15 min incubation at 70°C, 1/5 volume of 0.5M iodoacetamide (IAA) in water was added, and the reaction was allowed to happen at 37°C for 2 h before injection to RP-HPLC. Collected fractions were lyophilized and subjected to MS/MS sequencing described earlier.

Spectrophotometric determination of protein concentration

The concentrations of purified peptides were calculated from the Beer-Lambert law:

$$A = \varepsilon \,.\, l \,.\, c$$

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

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

l : cell path length (cm)

The thereotical ε value of a protein at 280 nm was calculated as follows [159]:

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

where n_{Trp} : the number of Trp residues

n_{Tyr} : the number of Tyr residues

n_{SS} : the number of disulfide bonds in the protein

Stability assay

Heat stability assay

Purified peptides were heated to 100°C for 1 h and then subjected to ultra performance liquid chromatography (UPLC). Corresponding peptides without the heat treatment served as controls. Collected peaks from UPLC were monitored by MALDI-TOF MS.

Enzymatic stability assay

Purified peptides were incubated with or without chymotrypsin (at a final peptide/enzyme ratio of 10:1 w/w) in 20 mM NH_4HCO_3 (pH = 8.0) at 37°C for 4 h. Corresponding peptides completely reduced with 50 mM DTT (2 h, 37°C) was treated in the same way and used as controls. Treated samples or controls were subjected to UPLC and the collected peaks were monitored by MALDI-TOF MS.

Bioassays

Alpha-amylase inhibition assay

Amylase purification

Alpha-amylase was isolated from yellow mealworm, larvae of *Tenebrio molitor,* following procedure described previously [160] with some modifications. Yellow mealworm extraction was done in standard buffer (20 mM acetic acid/NaOH (pH

5.4), 1 mM $CaCl_2$). Proteins were precipitated using 90% saturated ammonium sulfate. Protein precipitate was dissolved in standard buffer and dialyzed against the same buffer (overnight, 4°C). Subsequently, the solution was subjected to an anion exchange chromatography (buffer A: 0.02 M KH_2PO_4/K_2HPO_4 (pH 5.4), buffer B: KH_2PO_4/K_2HPO_4, 500 mM KCl (pH 5.4), gradient of 1.5% buffer B per minute, flow rate of 3 ml/min). α-Amylase assay was done to screen for fractions containing TMA. These fractions were subsequently dialyzed against distilled water (overnight, 4°C) and enzyme aliquots were stored at – 20°C.

Amylase inhibition assay

Assays for α-amylase were carried out in 96-well microtitre plates (Nunc) following the method developed previously [161]. In principle, enzyme TMA with or without being treated with peptides (20 min, 37°C) was incubated with 1% starch (in 20 mM sodium phosphate buffer, pH 6.7) for five minutes. The same amount of color reagent (3,5-dinitrosalicilic acid, Sigma, and sodium potassium tartrate, Sigma) [162] was dispensed into each well and color developed for 20 min at 100°C. Absorbance at 540 nm of water-diluted samples was read to determine the α-amylase activity. To build the standard curve, different maltose concentrations were incubated with color reagent (20 min, 100°C). Similar inhibition experiments were done for human salivary, porcine pancreatic, and *Aspergillus oryzae* α-amylase (Sigma). One enzyme unit was defined as the required amount of enzyme that liberated one mg maltose under the assay conditions. The remaining enzyme activity (%) was derived by using the following formula:

$$Enzyme\ activity\ (\%) = \frac{A_{540\ Sample}\ x\ 100}{A_{540\ Negative\ control}}$$

*: Absorbance values had been subtracted by the corresponding blanks. The blanks were obtained similarly with the enzyme being added after the color

reagent.

Hemolysis assay

Fresh type AB blood was donated by a volunteer student. The erythrocytes from 1 mL of blood were collected by centrifugation (700 rpm, 15 min) and washed thrice with 10 mL of PBS buffer (pH 7.2), each followed by 1000 rpm, 5 min centrifugation. 500 µL of centrifuged erythrocytes was diluted 100x in PBS buffer to prepare the stock dispersion. Stock peptides were diluted in PBS to final 2x concentrations of 200, 150, 100, and 50 µM. The stock dispersion was added to the wells in 96-well plates together with diluted peptide at the ratio of 1:1 (v:v). The final concentrations of peptides tested were thus 100, 75, 50, and 25 µM. The mixtures were incubated at 37°C for 4 h with gentle shaking. After centrifugation at 1000 rpm for 5 min, the supernatants were transferred to new wells and the absorbance at 415 nm measured in a Tecan plate reader. Equal volume of PBS or 8% Triton X-100/PBS was used as negative and positive control, respectively. The experiments were done in triplicate. Hemolysis data were calculated as the percentage of the complete hemolysis.

Cytotoxicity assay

Cytotoxicity of the purified allatides was tested using PrestoBlue™ Cell Viability Reagent (Invitrogen). Briefly, African green monkey kidney (Vero) cells were seeded into a 96-well plate and incubated with allatides at five different concentrations (100 µM, 75 µM, 50 µM, 25 µM, and 1 µM) for 24 h at 37°C. After incubation, the wells were dispensed with the reagent and left at 37°C for 2 h. The fluorescence was subsequently read as instructed by the manufacturer. 1% Triton X-100 solution was used as positive control.

Radial diffusion assay

Radial diffusion assay was performed with Gram negative *Escherichia coli* and Gram positive *Staphylococcus aureus* bacteria. The nutrient-poor underlay gel (0.03% w/v Tryptic Soy Broth, 0.25% w/v agarose in 10 mM sodium phosphate buffer) prewarmed at 42°C was inoculated with $4x10^6$ CFUs and poured into 100 mm Petri dishes. 3 µl of peptide samples were added into 2-mm wells punched on the solidified underlay gel. The dishes were incubated at 37°C for 3 h to allow for the diffusion of peptides into the gel. Nutrient-rich overlay gel (6% w/v Tryptic Soy Broth, 1% w/v agarose in Mili-Q water) prewarmed at 42°C was then poured onto the underlay gel with equal volume. The petri dishes were incubated at 37°C overnight. The antibacterial activity was calculated from the the diameter of the clear zones surrounding the wells. D4R, an in-house peptide dendrimer with potent antibacterial activity, was used as positive control. The experiments were done in duplicate.

Antiviral activity study

Antiviral assay

Vero cells (4×10^4 cells) were seeded to 48-well plates and incubated for 24 h. Various concentrations of alstotides S1 and S3 (25 µM, 50 µM, 75 µM, 100 µM) diluted in DMEM were mixed with IBV (equivalent amount to MOI=1) and incubated at room temperature for 30 min. The mixtures were then inoculated into the cell monolayers for 1 h at 37°C with occasional shaking. The supernatants were removed and the media were replaced with DMEM. The cells treated with virus only served as the negative control. After 24-hour incubation, cytopathic effect was observed under light microscope to ensure complete infection of the negative control. The virus supernatants were collected for plaque

assay and the monolayers lysed with Laemmli sample buffer with 20 µM DTT for Western blotting. Each concentration of alstotides S1 and S3 was assayed in duplicate.

Western Blotting

The lysate of aS1-treated/nontreated infected cell samples, non infected cells, and the viral pellets in 2x Laemmli loading buffer (4% SDS, 10% 2-mercaptoehtanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris HCl) were separated on 10% SDS-PAGE gels and transferred to the polyvinylidene fluoride (PVDF) membrane (Millipore, Singapore) by wet transfer method. The membranes were blocked with 5% low-fat milk in PBST at room temperature for 1 h with constant shaking. The membranes were then incubated with diluted antibody targeting specific viral proteins (dilution 1:4000-1:8000) for 1 h at room temperature or 4°C overnight, which was followed by thee washes with PBST, 20 min each. The blots were subsequently probed with HP-conjugated secondary antibodies (dilution 1:1000-1:2000) for 1 h at RT or 4°C overnight, followed by thee washes with PBST, 20 min each. The Western Lightning Plus ECL (PerkinElmer) was used as detection reagent and the exposed Kodak X-ray medical films were developed by Kodak X-OMAT 2000.

Plaque assay

Vero cells grown to confluency in 6-well plates were washed once with PBS and infected with 200 µl of serial dilutions of the virus stocks in DMEM at 37°C for 2 h with occasional shaking. The inoculum was aspirated, the cells washed with PBS and overlayed with 2 mL of 0.3% agarose in water. The plates were incubated in humidified 5% CO_2 incubator at 37°C for 3 days. The cells were subsequently fixed with 4% paraformaldehyde at RT for 0.5 h and virus plaques were visualized and counted by staining with 0.2% crystal violet in 20% EtOH.

Determination of $TCID_{50}$

Vero cells ($1.5x10^4$ cells/well) were plated onto 96-well plate 1 day before infection. Cell cultures were washed once with PBS and infected with serial dilutions of virus samples in DMEM. 10 replicates were prepared for each dilution. Non-infected cells dispensed with DMEM served as controls. The cultures were incubated for 3-5 days at 37°C till the cytopathic effects appeared. TCID50 titers were calculated using the Excel spreadsheet from Yale School of Medicine available at www.med.yale.edu/micropath/pdf/Infectivity%20calculator.xls.

Reverse transcription and real-time PCR analysis

At the time of collection, the supernatants were removed. In order to determine the expression level of viral and host proteins, total RNA of the monolayers was extracted using Trizol reagents following the manufacturer's instruction. Different primers were used in the reverse transcription from total RNA to cDNA (Table 2.1) using Transcriptor First Strand cDNA Synthesis kit (Applied Biosystems) according to the manufacturer's protocol. 1 µg of total RNA was used in each reaction to prepare 20 µL of cDNA. The reactions were performed at 50°C for 1 h before the Transcriptor Reverse Transcriptase was inactivated at 85°C for 5 min.

Table 2.1. Primers used to study viral replication

Source	Target	Primers for reverse-transcription	Primers for realtime PCR
Host	GAPDH	Oligo $(dT)_{18}$	GAPDH-F: 5′ GGAGCGAGATCCCTCCAAAAT 3′ GAPDH-R: 5′ GGCTGTTGTCATACTTCTCATGG 3′
Virus	+gRNA	IBV3835-F	IBV3835-F: 5′ GGAGACAACAACTTGCATG 3′ IBV4321_R: 5′ TCGCGCCACTCTAGTATAAG 3′

+gRNA: positive-stranded genomic RNA

Quantitative real-time PCRs were performed using 1 µL of cDNA and Power SYBR Green PCR 2× Master Mix on the 7500 Real-time PCR system (Applied Biosystems). Cycling conditions were as follows: 1 cycle of 95°C for 10 min and 40 cycles of 95°C for 15 s and 60°C for 1 min. Each reaction was done in triplicate. The data were analyzed with the 7500 software (Applied Biosystems).

Time-of-drug-addition assay

Vero cells seeded in 48-well plates were incubated with IBV solution (MOI=1) at 4°C for 2 h to facilitate viral attachment. Alstotide S1 (50 and 100 µM) was added together with the virus at the beginning of the incubation for sample A (Figure 2.3). After the low-temperature incubation, the cells were washed once with PBS to remove unbound virus particles and the medium replaced with new DMEM before the cells were transferred to 37°C. At the transfer or 3, 6, and 8 h post transfection, alstotide S1 was added to parallel cultures at the final concentration of 50 and 100 µM. Equal amount of peptide was also dispensed to sample A to replace the washed-out peptide. At 24 h post infection, the cytopathic effects of infection were monitored under the light microscopy. The supernatant were collected for plaque assay and the cell monolayers for Western blotting against IBV N and S proteins. IBV-infected Vero cells without peptide treatment served as negative control.

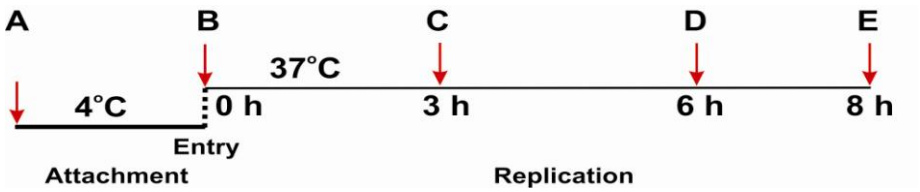


Figure 2.3. Experimental design of time-of-drug-addition assay

Transfection

Subconfluent Vero cells grown on 100 mm disks were trypsinized, washed twice with cold DMEM and resuspended in 1.2 mL of ice-cold PBS. The cells must be kept cold during the procedure. Vero suspension (400 µL) was transferred to a pre-chilled electroporation cuvette together with 2 µg of viral RNA (in up to 5 µL). Electroporation was performed with one pulse at 450 V, 50 µF, and infinitive resistance using Gene Pulser II Electroporation System (Bio-Rad, USA). After electroporation, the cells were left on ice for 1 min, resuspended in prewarmed complete media, and plated in 24-well plates for Western blotting (500 µL/well) and 12-well plates for quantitative PCR (1,000 µL/well). The cells were incubated at 37°C, 5% CO_2 and harvested after 3, 18, and 36 h. The samples collected at 3 h after seeding served as the starting control to calculate the percentage replication of IBV virus.

Preparation of biotinylated peptide

The biotinylation of alstotide S1 was performed based on the efficient reaction of NHS-activated biotin with primary amine at pH 7-9 to form stable amide bonds. Biotin reagent was stored at 4°C and equilibrated to room temperature prior to the reaction to avoid moisture condensation. 1 mL of 15 mM biotin in DMSO (6.8 mg/mL) was mixed with 1 mL of alstotide S1 in PBS, pH 7.4 (1 mg/mL) and incubated at RT for 30 min or 4°C for 24 h. The products were monitored by MS and purified by analytical RP-HPLC. Eluted fractions containing biotinylated peptide was collected, lyophilized and dissolved in water prior to storage at -20°C.

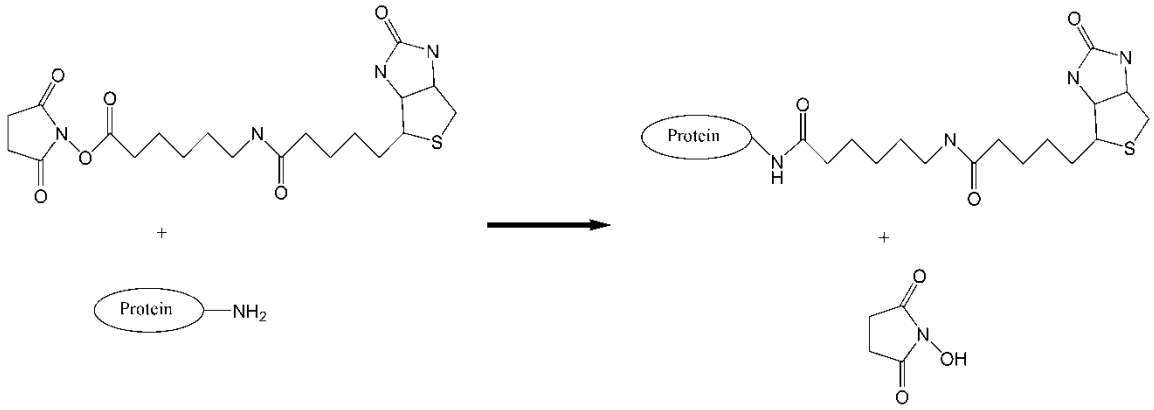


Figure 2.4. Reaction of NHS-LC-Biotin with primary amino group at pH 7-9. N-hydroxysuccinimide (NHS) is a good leaving group, which will be replaced by the protein through the formation of a new amide bond. NHS and excessive NHS-LC-biotin are removed during the final desalting step.

In vitro pull-down assay

Lysates containing IBV proteins were prepared from IBV-infected Vero cells seeded onto 100 mm disks. At about 24-28 hpi when CPE was observed for 80-90% cell population, the adherent cell monolayer was washed twice with ice-cold PBS. 2 mL of cold PBS was added to each disk to facilitate cell scraping using a plastic policeman. The cell suspension was transferred to a 15 mL tube and centrifuged at 3000 x *g* at 4°C for 5 min. After the removal of PBS, 1.5 mL of cold lysis buffer was added to every 10^7 cells, vortexed to resuspend the cells and left on ice for 20-30 min with occasional mixing. Lysis buffer includes 50 mM Tris-HCl, pH 8.0, 5mM ethylenediamine tetraacetic acid (EDTA), 150 mM NaCl, and 0.5% NP40 with 1mM DTT and protease inhibitors added before use. The lysate was ready for pull-down assay after centrifugation at >12,000 x *g* at 4°C for 20 min. The negative control lysate was prepared in similar manner from non-infected Vero cell population.

Biotinylated alstotide S1 (20 µg) was incubated with 30 µL of NeutrAvidin UltraLink beads (Thermo Scientific) for 30 min at RT with rocking. The beads were washed twice with 500 µL PBS before incubation with the lysates for 2 h at 4°C with rocking. Unbound proteins and unspecific binding proteins were removed later by two washes with lysis buffer and two washes with SNNTE buffer (50 mM Tris-HCl, pH 7.4, 5mM EDTA, 5% sucrose, 1% NP40, and 0.5M NaCl; 1mM DTT and protease inhibitors added before use). Protein complexes were eluted with Laemmli sample buffer and heated to 100°C for 5 min. IBV proteins that interact with alstotide S1 were detected by SDS-PAGE electrophoresis and Western blotting with corresponding antibodies.

Co-immunoprecipitation assay

The experiment was performed in a similar manner to co-immunoprecipitation assay except that the protein complex was pull-downed directly by NeutrAvidin beads. Lysate of infected, aS1-treated cells was incubated with NeutraAvidin

beads for 30 min at RT with rocking. The beads were subsequently washed thrice with lysis buffer and the protein complexes eluted with Laemmli sample buffer. Western blotting was performed to analyze protein-alstotide S1 interaction.

Immunofluorescence assay

Vero cells seeded at a density of 6 x 10^4 in 4-well chamber slide were inoculated with DMEM alone, IBV (MOI = 1), biotinylated-AS1-treated (100 µM) IBV, or biotinylated aS1 (100 µM) for 16 h at 37°C until the cytopathic effect was observed for 70-80% of the cell population. The supernatant was removed and the cells washed thrice with ice-cold PBS, 5 min each, followed by fixation with 4% paraformaldehyde for 20 min at RT. After being rinsed with cold PBS, the cell monolayers were permeabilized with 0.2% Triton X-100 in PBS for 10 min at 4°C and washed thrice with cold PBS, 5 min each. The cells were subsequently blocked with blocking buffer (5% FBS in PBS) for 20 min and incubated with rabbit anti-IBV-N protein antibody in blocking buffer at RT for 2 h, followed by three washes with cold PBS. Secondary antibody was goat anti-rabbit immunoglobulin conjugated with Alexa Fluor 594 (dilution 1:200, Life Technologies) in blocking buffer at 4°C for 2 h. After washes with PBS, the cells were stained with Neutravidin Dylight 488 conjugate (dilution 1:200, Thermo Scientific) at 4°C for 2 h. The slide was mounted in Vectashield with counterstain 4′,6′-diamidino-2-phenylindole (DAPI, Vecto Labs). The fluorescence was examined on a Zeiss LSM 710 META confocal microscope at the magnification of ×63.

Antiviral assay on other viruses

Vero cells (4×10^4 cells) were seeded to 48-well plates and incubated for 24 h. Various concentrations of alstotides S1 (25 µM, 50 µM, 75 µM, 100 µM) diluted in DMEM were added to the cells together with DENV2 or RSV A. The cultures were incubated for 2 (RSV A) to 3 (DENV2) days until >50% CPE appeared.

The cells treated with virus only served as the negative control. The monolayers were lysed with Laemmli sample buffer for Western blotting to visualize the antiviral activity of alstotide S1 against RSV and DENV.

Chapter 3

Characterization of Cystine Knot α-Amylase Inhibitors in the Apocynaceae Family

Introduction

The Apocynaceae, commonly known as the dogbane family, is a flowering plant family that includes trees, shrubs, and vines. This family comprises up to 395 genera and over 5100 species [163] primarily distributed in the tropics and subtropics [164]. Many Apocynaceae species exude milky latex and some are poisonous. Plants of this family are often rich in secondary metabolites such as alkaloids and glycosides, particularly in seeds and latex [164]. Some Apocynaceae plants were of economic uses such as rubber and fiber sources while others for ornamentation or cooking. Several toxic genera, such as *Nerium, Acokanthera, Cerbera, Thevetia, Carissa* and *Plumeria*, provide sources of cardiac glycosides and alkaloids [165-167] for heart, high blood pressure and cancer treatment.

Apocynaceae species have been long used in folklore medicines for treating fever, diarrhea, skin diseases, tumors, cardiopathy, and other diseases [168-170]. Pharmaceutically, extracts from Apocynaceae species were shown to display antiplasmodial [171, 172], cytotoxic [171, 173, 174], antidermatophytic [175], and antimalarial properties [176, 177]. Here we studied three Apocynaceae species largely available in Singapore and Southeast Asia, representing three different genera with known medicinal values: *Allamanda oenotheraefolia, Alstonia scholaris,* and *Wrightia religiosa*.

Allamanda oenotheraefolia

Allamanda oenotheraefolia is an ornamental small woody, evergreen shrub, also known as Bush Allamanda, Golden Trumpet Bush, Yellow Bell, or Buttercup Flower (Figure 3.1). *A. oenotheraefolia* is native to Brazil and now adapted to many sunny, well-watered areas around the world. Bush Allamanda is highly similar in appearance to *A. cathartica*, the vine Golden Trumpet. However, as their name suggested, Bush Allamanda only grow into bushes and their flowers are smaller than those of *A. cathartica* [178].

There are not many reports on the biological activity of *A. oenotheraefolia* extracts but its closely related species *A. cathartica* has been known for many medicinal values. *A. cathartica* roots have applications in treating malaria and jaundice in Suriname's traditional medicine while flowers are used as laxative [169]. The aqueous extract of *A. cathartica* leaves exhibited wound healing effects which reduced macrophages and inflammation on rat model [179]. *A. cathartica* also has antibiotic activity against Gram-positive bacteria *Staphylococcus* [179]. We suspect that *A. oenotheraefolia* may possess similar bioactivity profiles.

Alstonia scholaris

A. scholaris, known as devil tree, is an evergreen tree found in the tropics with the height up to 100m. *A. scholaris* is one of the most studied plants of *Alstonia* genus [180]. Different plant parts have been used as a folklore remedy for a wide spectrum of ailments. The bark is used in treating fever, diarrhea and tumors, the milky juice for ulcer, rheumatic pains and earache, and the leaf juice for snakebites [181]. In pharmaceutical study, crude extracts of various plant parts display antimicrobial, antidiarrhoeal, antiplasmodial, teratogenic, immunomodulatory, wound healing, analgesic, and antioxidant activities [182].

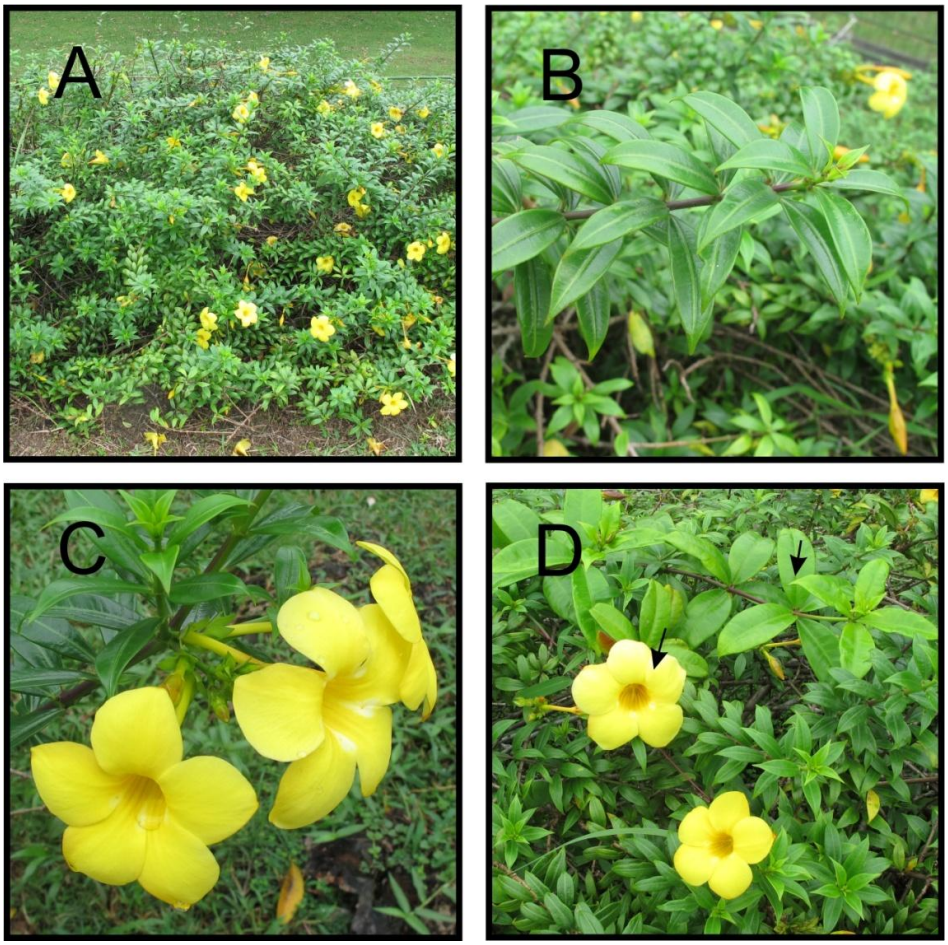


*Figure 3.1. A. oenotheraefolia in the campus of Nanyang Technological University, Singapore. A. oenotheraefolia grows into an evergreen small shrub (**A**). The leaves grow in whorls of four, about 6-10 cm (**B**). A. oenotheraefolia has bright yellow flowers (**C**), which are surprisingly similar to yet smaller than those of its closely related A. cathartica. The arrows in Figure 3.1D point to the A. cathartica flower and leave.*

Wrightia religiosa

Wrightia religiosa has the common names as Sacred Buddhist, Wondrous Wrightia, Wild Water Plum, or Water Jasmine. It is a native plant of Southeast Asia mainly used for bonsai purpose. *W. religiosa* grows into tropical shrubs with thin, slightly hairy foliage, simple oval-shaped leaves and white, pendulous fragrant flowers on twiggy branches [116, 183]. The plant has been traditionally used as medicinal herb in Ayurvedic medicine and the root for treatment of skin diseases [184].

Despite the long uses of Apocynacae plants in folklore medicine, most of the studies have so far focused on small secondary metabolites, such as alkaloids and glycosides. A larger chemical space which includes therapeutic miniproteins of <5 kDa, particularly cysteine-rich peptides, has not been explored although its presence in the Apocynaceae species was shown in a field trip screening for cyclotides in 2008 [97]. This chapter therefore reports on the isolation and characterization of 3-4 kDa cysteine-rich peptides from *A. oenotheraefolia*, *A. scholaris*, and *W. religiosa* by proteomic and genomic methods. The work led to the discovery of a family of 30-amino-acid CK α-amylase inhibitors in these three plants and potentially in many other Apocynaceae species. Novel CK α-amylase inhibitors display tissue- and species-specific expression profiles together with cytotoxicity, hemolysis, and antiviral effect. These findings provide better understanding of the distribution of CK α-amylase inhibitors in the Apocynaceae family and their potential roles as active components in traditional medicine.

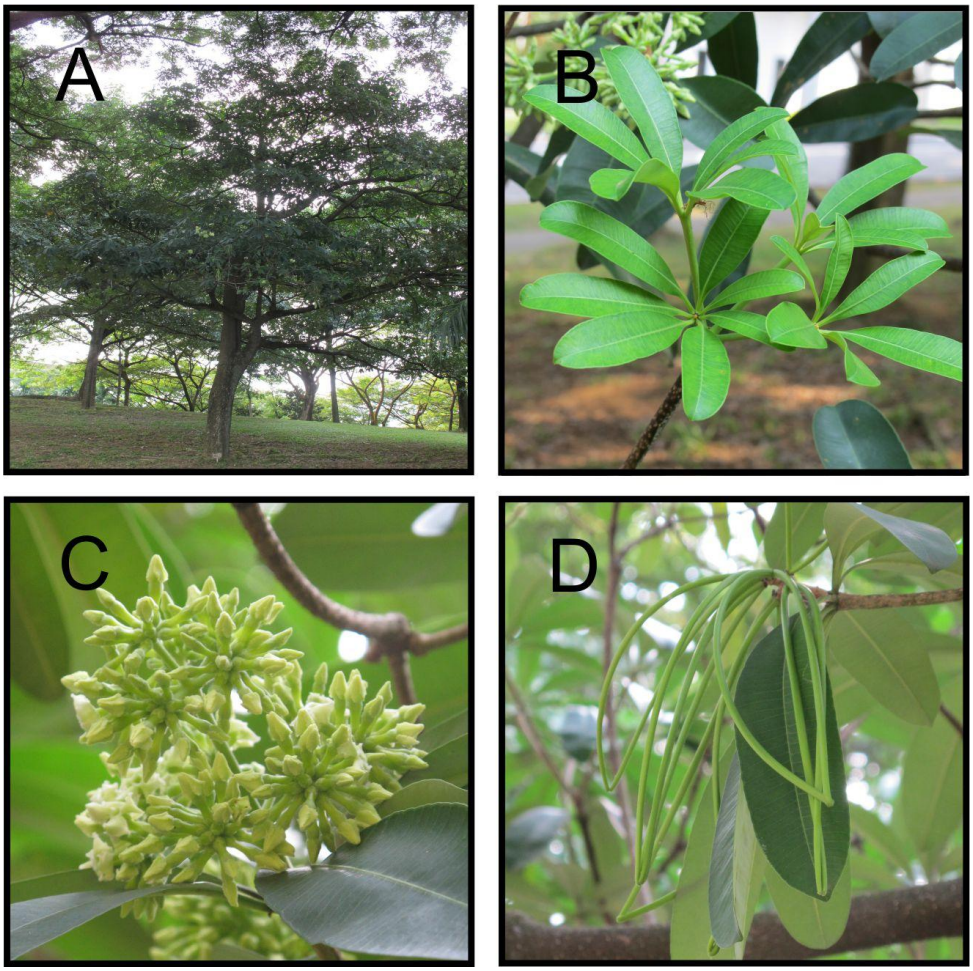


*Figure 3.2. A. scholaris in the campus of Nanyang Technological University, Singapore. A. scholaris plant is a 10-meter evergreen tree (**A**). The leaves grow in whorls (**B**), about 10-15 cm. A. scholaris has strongly fragrant, slightly greenish flowers (**C**) and paired fruit pods, 30-60 cm long (**D**).*

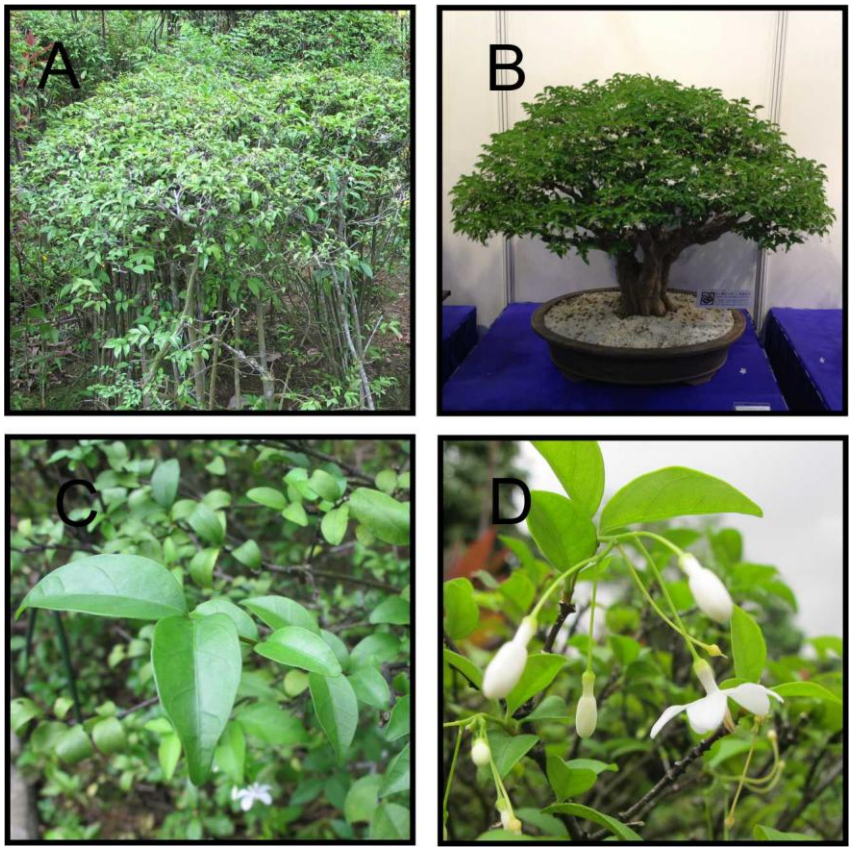


*Figure 3.3. W. religiosa in the campus of Nanyang Technological University, Singapore. W. religiosa is an ornamental plant (**A** and **B**) with small oval-shaped leaves (**C**) and white, pendulous fragrant flowers (**D**).*

Results

Screening for cystine knot peptides in medicinal Apocynaceae species

We investigated the presence of cystine knot peptides in the Apocynaceae plants by small-scale extractions on RP-HPLC column. Briefly, about one g of plant sample was homogenized in 50% ethanol (v/v) and loaded to C18 solid phase extraction column. The column was washed with 20% acetonitrile and eluted with 80% acetonitrile. The occurrence of putative cystine knot peptides in the eluate was profiled using MALDI-TOF MS. The screening was based on CK peptide-defining characteristics of high hydrophobicity (eluted at 25-55% acetonitrile from C18 column), mass range from 3 to 5 kDa, and the presence of three disulfide bridges [185].

Nine species available within Nanyang Technological University campus area were screened, including *Nerium oleander* (oleander), *Alstonia scholaris* (blackboard tree), *Wrightia antidysenterica* (arctic snow), *Wrightia religiosa* (water jasmine), *Cerbera odollam* (suicide tree), *Plumeria rubra* (frangipani), *Plumeria obtusa* (frangipani), *Allamanda oenotheraefolia* (Bush golden trumpet), and *Allamanda cathartica* (golden trumpet). Except for *Cerbera odollam*, the other eight plants exhibited strong positive signals in the mass range of 3-5 kDa (Figure 3.4). There were two major patterns of peptide expression observed in the MS profiles of these eight plants. *W. religiosa*, *P. rubra*, *P. obtusa*, *A. oenotheraefolia,* and *A. cathartica* expressed the two clusters of peptides, one centered around 3.0-3.5 kDa and the other around 3.7-4.1 kDa. In contrast, *N. oleander*, *A. scholaris*, and *W. antidysenterica* showed only strong signals for the lower-mass cluster of 3.0-3.5 kDa.

These plant extracts were subsequently subjected to S-reduction and S-alkylation with iodoacetamide (IAA). Mass-shifts of 348 Da were observed for almost all MS peaks in the range of 3-5 kDa, indicative of the presence of cysteine-rich peptides having six cysteine residues in their primary sequences

(Each half-cystine S-alkylation causes a 58-Da mass increase). Three plants, with high-intensity MS profiles of CRPs and of abundant materials, including *A. oenotheraefolia, A. scholaris,* and *W. religiosa*, were investigated further in large-scale extraction and characterization. Peptides with the higher mass range (>3.8 kDa) were not characterized in this study.

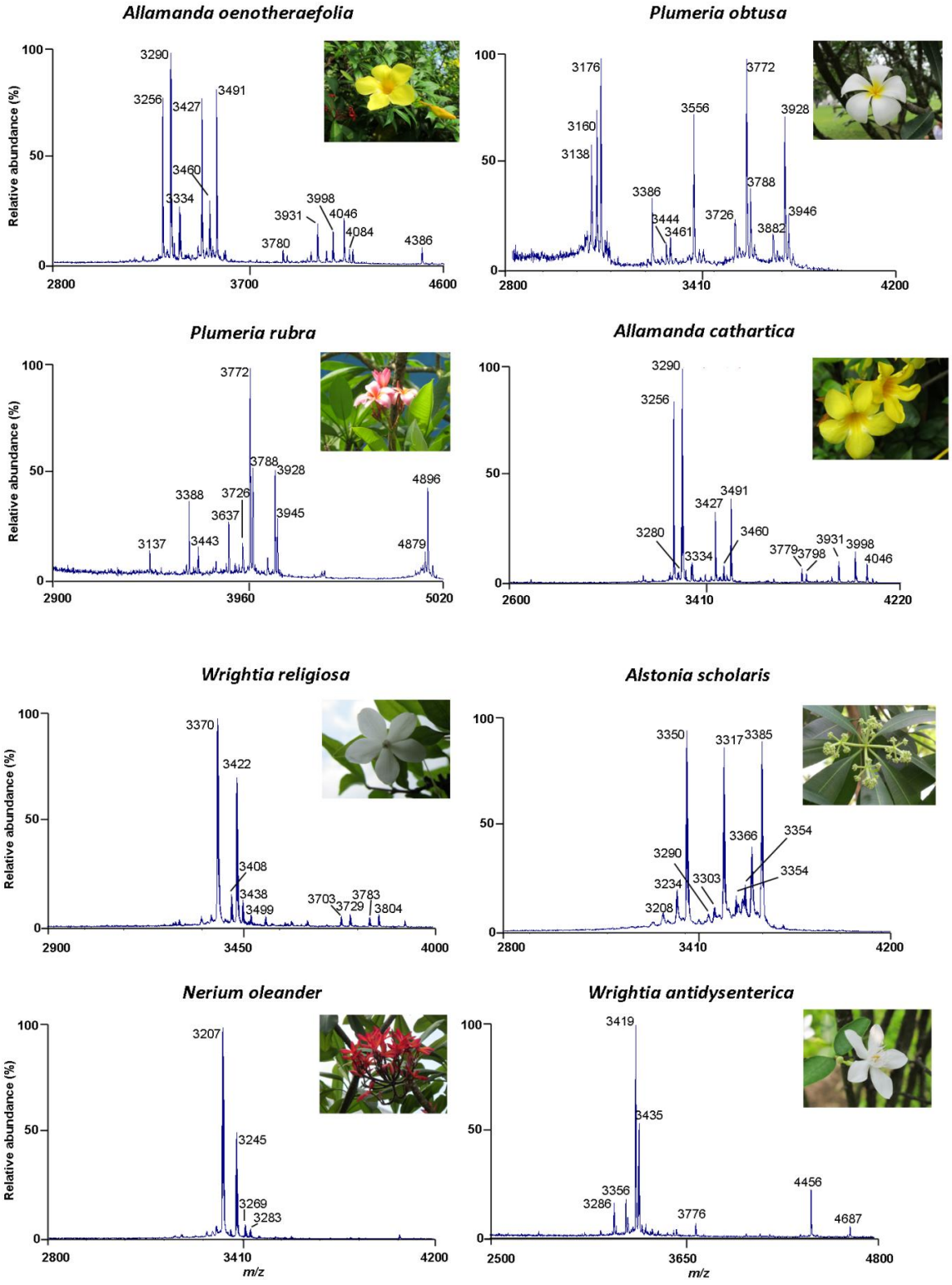


Figure 3.4. Species-specific expression profiles of cystine knot peptides from Apocynaceae species

Purification and de novo sequencing of putative CK α-amylase inhibitors

Allamanda oenotheraefolia

To isolate α-amylase inhibitors from *A. oenotheraefolia* for characterization, 800 g of fresh leaf material was homogenized and extracted twice in 50% ethanol (v/v) in a large-scale extraction. Chlorophylls and lipids were removed by partition with dichloromethane. The aqueous layer was loaded onto a C18 column, which was washed with ten volumes of 20% ethanol to partially remove hydrophilic compounds and color-bearing substances. The α-amylase inhibitor-enriched fractions eluted with 40% to 70% EtOH were collected and passed through several rounds of reversed-phase and ion-exchange HPLC.

Seven peptides, named allatides O1-O7 (aO1-aO7), were purified with allatide O4 giving the highest yield of 5 mg. Each purified allatide was disulfide-reduced (*S*-reduced) to free Cys by dithiothreitol and digested with trypsin and chymotrypsin. The resulted fragments were sequenced by tandem mass spectrometry MALDI-CID MS/MS and the sequences deduced by the analysis of b- and y-ions in their profiles. As an example (Figure 3.5), the *S*-reduced allatide O2 had the *m/z* value of 3433, indicating a mass increase of 6 Da as compared to the native peptide. Enzymatic digestion of the *S*-reduced allatide O2 gave two fragments with *m/z* values of 845 and 2949, respectively. The total *m/z* value of these two digested fragments was 19 units larger than the *m/z* of the whole *S*-reduced peptide, suggesting a single tryptic cleavage of the backbone. The assembly of the two fragment sequences yielded the full sequence of allatide O2.

A limitation of *de novo* sequencing by MS/MS is the difficulty in assigning isobaric residues, such as Leu/Ile and Lys/Gln. To clarify the sequence ambiguity in such cases, three main methods are available:

- Gene cloning: The genetic sequences of the precursors provide unambiguous information about the peptide domain primary sequences.
- Amino acid analysis: The peptides are reduced, enzymatically cleaved into smaller fragments and the amino acid composition of each fragment is determined by acid hydrolysis and UPLC.
- Enzyme digestion: The peptides are cut after specific residues, which is followed by MS confirmation of the digested fragments. The common enzymes used in this method include trypsin and chymotrypsin. Trypsin cleaves peptide bonds after basic residue Lys and Arg but not Gln. Chymotrypsin cuts after Phe, Tyr, Trp, and to a lesser extent, Met and Leu, but not Ile.

Among the three methods, gene cloning has several advantages which will be discussed in Chapter 5. Therefore, most of the sequences identified in this work were confirmed by the genetic information of corresponding peptides. Isobaric residues in the peptides without gene sequences were clarified by amino acid analysis. Other assignments were performed based on sequence homology with the confirmed peptides thanks to their high homology within each species.

Table 3.1 summarizes the primary sequences of allatides O1 to O7. Isobaric residues of allatide O2 was determined by amino acid analysis, allatides O1, O3-O5 by genomic analysis, and allatides O6-O7 by sequence homology with allatides O4 and O1, respectively. All allatides were found to contain 30 amino acid residues with a high occurrence of proline (3 to 4 residues), glycine (3 residues) and cysteine (6 residues). Together, they account for ~40% of their sequences. Sequence analysis based on the Cys motif showed that allatides may contain a CK connectivity, a CC-type of CK peptides with adjacent Cys residues which is found in AAI and ω-conotoxin. Because they are Cys-, Pro- and Gly- rich, BLAST search showed that allatides share moderately high homology of about 50% and 23% sequence identity to AAI and ω-conotoxin, respectively. However, allatides share high sequence homology among their

members, with 20 out of 30 residues conserved. In many cases, the allatides differ from each other only by one or two residues (allatides O1 and O7, O2 and O3, O4, O5, and O6).

Table 3.1. Putative cystine-knot α-amylase inhibitors from A. oenotheraefolia

Peptide		Sequence [a]	Approach [b]	MW [c]	No of residues	Charge
Intercysteine segment		**1 2 3 4**				
Allatide O1	aO1	CIAHYG-KCDGIIN--QCCDPWLCTPPIIGICI	P+G	3255	30	0
Allatide O2	aO2	CRP-VGTRCDGVIN--QCCDPYWCTPPIYGWCK	P+A	3426	30	+1
Allatide O3	aO3	CRP-YGTRCDGVIN--QCCDPYWCTPPIYGWCK	P+G	3490	30	+1
Allatide O4	aO4	CIAHYG-KCDGIIN--QCCDPWLCTPPIIGFCL	P+G	3289	30	0
Allatide O5	aO5	CVSHYG-KCDGIIN--QCCDPWLCTPPIIGFCL	P+G	3291	30	0
Allatide O6	aO6	CHAHYG-KCDGIIN--QCCDPWLCTPPIIGFCL	P	3313	30	+1
Allatide O7	aO7	CHAHYG-KCDGIIN--QCCDPWLCTPPIIGICI	P	3279	30	+1
AAI[d]		CIPKWN-RCGPKMDGVPCCEPYTCTSDYYGNCS		3538	32	0
ω-conotoxin GVIA		CKSO-GSSCSOTSY--NCCR--SCNOYTKR-CY*		3408	27	+5

a. Disulfide connectivity is shown by the solid lines.

b. Approaches to obtain α-amylase inhibitor sequences are proteomic (P), genomic (G), and amino acid analysis (A).

c. Monoisotopic masses are reported in this table.

*d. AAI, the α-amylase inhibitor from A. hypochondriacus, and ω-conotoxin GVIA are included for sequence comparison. The * indicates amidated C-terminus*

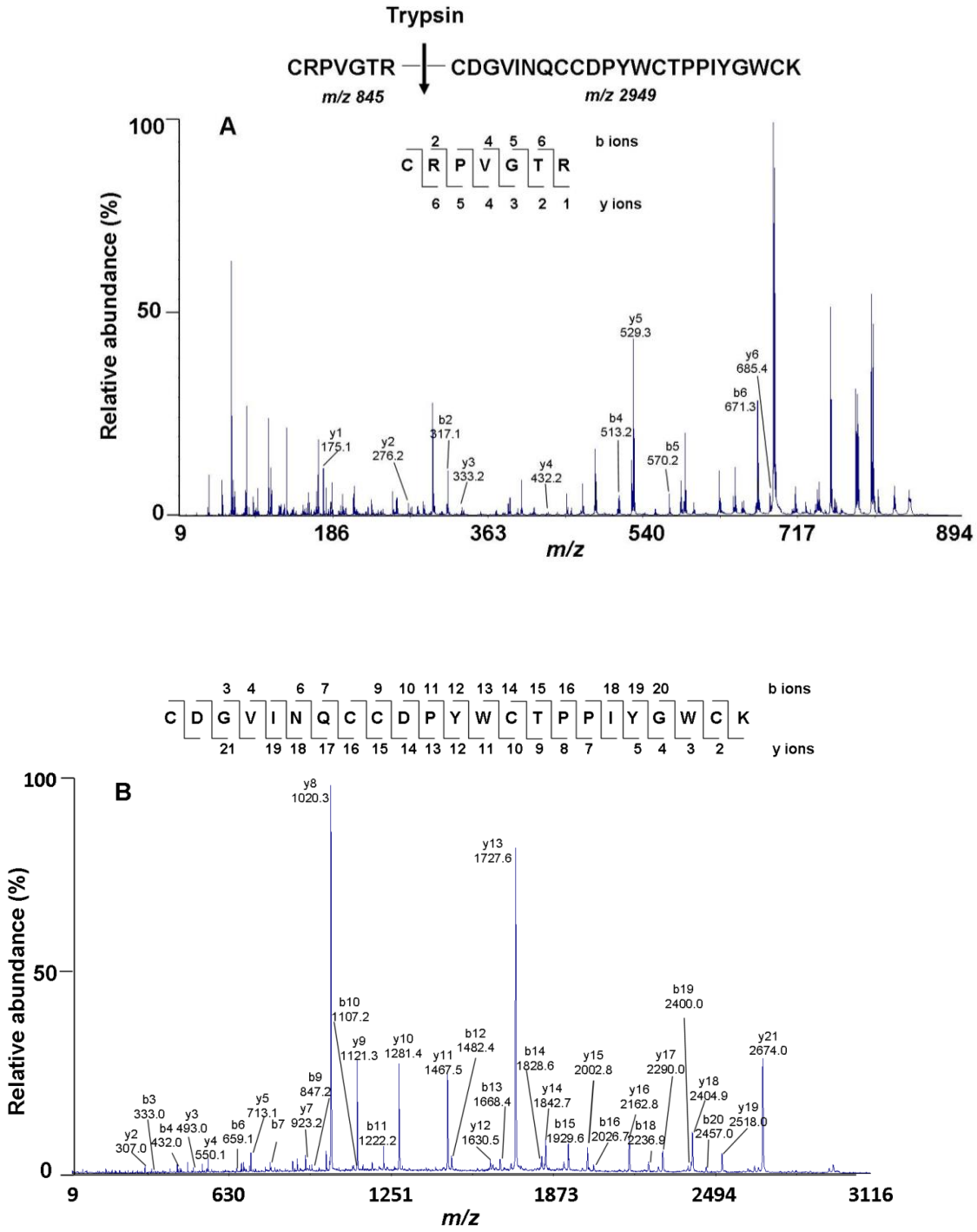


*Figure 3.5. De novo sequencing of allatide O2 using MALDI-TOF/TOF MS/MS. Tandem MS/MS profiles of two tryptic fragments (m/z 845 - **A** and 2949 - **B**) provide the full allatide O2 sequence. Ile/Leu assignment was determined by amino acid analysis.*

Table 3.2. Amino acid analysis result for allatide O2. Four fragments obtained after chymotrypsin digestion were subjected to hydrolysis, derivatization, and separation with UPLC. Two Ile residues were confirmed. Numbers on the left column are calculated from amino acid analysis results (Cald.); those on the right are theoretical values predicted from MS/MS data (Theo.).

Amino acid	**Fragments**							
	1814 Da		**936 Da**		**1467 Da**		**1179 Da**	
	Cald.	*Theo.*	*Cald.*	*Theo.*	*Cald.*	*Theo.*	*Cald.*	*Theo.*
Arg	1.1	1						
Asp/Asn	2.9	3						
Thr	1.0	1	1.0	1	0.5	1	1.0	1
Glu/Gln	0.2	1						
Pro	1.0	1	2.4	2	1.8	2	2.3	2
Gly	2.2	2			1.4	1	1.1	1
Cys		3		1		2		1
Val	0.5	1						
Ile	0.6	1	0.9	1	0.8	1	1.1	1
Tyr	1.2	1	1.1	1	1.0	1	1.2	1
Lys					1.4	1		
Trp				1		2		2
Total		**15**		**7**		**11**		**9**
Sequences	**GTRCDGVINQCCDPY**		**WCTPPIY**		**WCTPPIYGWCK**		**WCTPPIYGW**	

Alstonia scholaris

MS profile of *A. scholaris* demonstrates a cluster of CRPs around 3.3 kDa. The expression level of CRPs in *A. scholaris* is almost 8-fold higher than that in *A. oenotheraefolia.* Large-scale aqueous ethanol extraction was performed to purify four predominant peptides, alstotides S1-S4 (aS1-aS4). The yield ranged from 3 mg (alstotide S2) to 40 mg (alstotide S1) from 800 g fresh leave material.

Each purified alstotide was disulfide-reduced and sequenced by tandem MS/MS in a manner described previously for allatides. Their sequences are summarized in Table 3.3 with residue assignment of alstotide S3 provided in Figure 3.6. Sequence ambiguities by isobaric residues Ile and Leu were resolved by either gene cloning (alstotides S1, S2, and S4) or amino acid analysis (alstotide S3) (Table 3.4). All four alstotides are highly homologous to allatides: backbone containing 30 amino acids of which 19 are absolutely conserved among alstotides and allatides, rich in cysteine (6 residues), proline (4 residues) and glycine (3 residues). Particularly, the cysteine pattern and spacing of alstotides resemble the CC-type CK motif found in AAI and allatides. Alstotides S1 and S2 are nearly identical, differing by only one residue at position 20 where His in alstotide S1 is replaced by Arg in alstotide S2**.** The remarkably high sequence homology between alstotides from a woody tree and allatides from a shrubby plant provides interesting evidence of the distribution of CK peptides in the Apocynaceae family.

Table 3.3. Putative cystine-knot α-amylase inhibitors from A. scholaris

Peptide		Sequence [a]	Approach [b]	MW [c]	No of residues	Charge
Astotide S1	aS1	CRP-YGYRCDGVIN--QCCDPYHCTPPLIGICL	P+G	3365	30	+1
Astotide S2	aS2	CRP-YGYRCDGVIN--QCCDPYRCTPPLIGICL	P+G	3384	30	+1
Astotide S3	aS3	CVPRFG-RCDGIIN--QCCDPYICTPPLVGICT	P	3249	30	0
Astotide S4	aS4	CVPQYGV-CDGIIN--QCCDPYYCSPPIYGHCI	G	3316	30	-1
AAI[d]		CIPKWN-RCGPKMDGVPCCEPYTCTSDYYGNCS	-	3538	32	0

a Disulfide connectivity for CRPs is shown by the solid lines

b Approaches to obtain CRP sequences are either proteomic (P) or genomic (G)

c Monoisotopic masses are reported in this table. d AAI, included for sequence comparison, is the α-amylase inhibitor from A. hypochondriacus

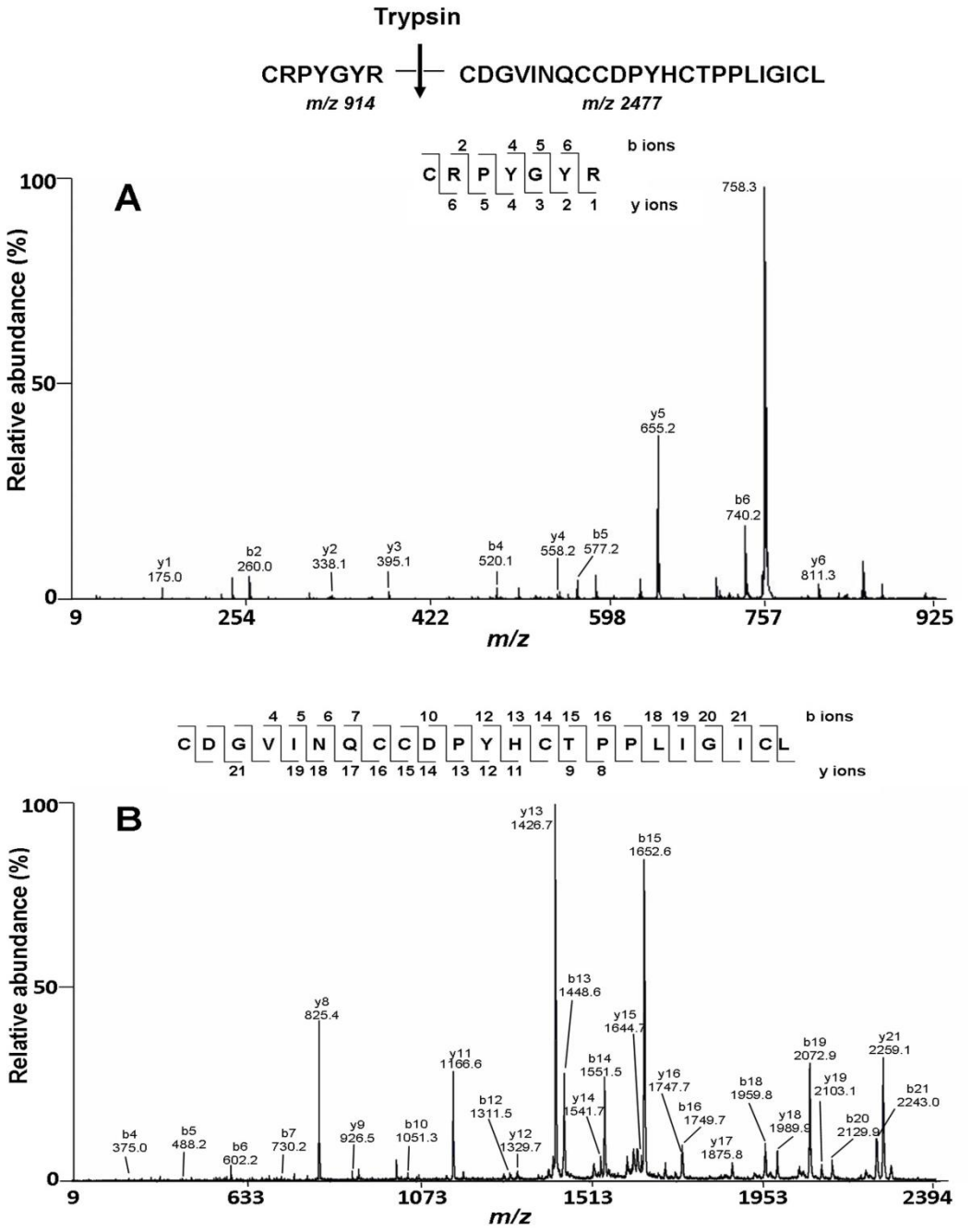


Figure 3.6. De novo sequencing of alstotide S1 using MALDI-TOF/TOF MS/MS. Alstotide S1 peptide was reduced by DTT and directly digested with trypsin. Two tryptic fragments with m/z 914 (CRPYGYR) and 2477 (CDGVINQCCDPYHCTPPLIGICL) were analyzed through their b- and y-ion series, which collectively provided full alstotide S1 sequence.

Table 3.4. Amino acid analysis result for alstotide S3. Chymotrypsin cleaved the IAA-alkylated alstotide S3 predominantly at two sites, giving us two main fragments. Cleavage sites are numbered as follows: CVPRF [1] GRCDGIINQCCDPY – 1727 Da [2] LCTPPLVGICT – 1230 Da. Amino acid analysis of these two fragments and the whole alstotide S3 peptide confirmed that among five isobaric Ile/Leu amino acid residues, the first two and the last one were Ile, and the rest were Leu. Numbers on the left column are calculated from amino acid analysis results (Cald.); those on the right are theoretical values predicted from MS/MS data (Theo.).

	Fragments					
Amino acid	**1230 Da**		**1727 Da**		**3250 Da**	
	Cald.	*Theo.*	*Cald.*	*Theo.*	*Cald.*	*Theo.*
Arg			1.00	1	1.79	2
Gly	0.94	1	2.04	2	2.81	3
Asp/Asn			3.39	3	2.89	3
Thr	1.77	2			1.88	2
Glu/Gln			1.59	1	1.00	1
Pro	2.07	2	1.42	1	4.22	4
Cys		2		3		6
Tyr			1.08	1	1.49	1
Val	1.00	1			2.21	2
Ile	0.99	1	1.60	2	2.78	3
Leu	2.06	2			2.45	2
Phe					1.81	1
Total	**11**		**14**		**30**	
Sequence	**LCTPPLVGICT**		**GRCDGIINQCCDPY**		**CVPRFGRCDGIINQCCDPYLCTPPLVGICT**	

Wrightia religiosa

Examination of the *W. religiosa* extract revealed two groups of CRPs around 3.2 kDa and 3.8 kDa, respectively. Large-scale purification of the leave extract and *de novo* MS/MS sequencing yielded three peptides of the former group, designated as wrightides R1 to R3 (wR1-wR3) and six of the latter, wrightides R4 to R9 (wR4-wR9). Wrightides R1 to R3 have 30 residues and wrightides R4 to R9 have 30-35 residues. All wrightides contain six cysteine residues arranged in spacings similar to that observed for the CC type of CK motif.

Despite sharing a backbone length of 30 residues, wrightides R1-R3 do not possess the strictly conserved intercysteinyl spacings as do allatides and alstotides. Loops 2 and 4 of wrightides contain 5 and 8 amino acids, respectively, whereas corresponding loops in allatides and alstotides contain 6 and 7 amino acids, respectively. Loop 3 of wrightides R1 and R3 is identical to that of alstotide S1. Loop 1 is the most diverse loop among wrightides, allatides, and alstotides. Except for the six cysteine residues, the absolutely conserved residues among peptides from the three plants include Gly5/6, Gln14, Glu17, Pro18 (corresponding to Pro20 in AAI), Tyr/Trp19, Thr22, Pro24, and Glu27 (alstotide numbering). The critical Arg7, which is believed to be important for the binding of AAI to α-amylase, is missing in wrightides R1-R3. This residue is conserved or replaced by the basic Lys7 in allatides and alstotides. Taken together, wrightides R1-R3 likely belong to the CC type of CK α-amylase inhibitors found in the third Apocynaceae species.

Wrightides R4-R9 represent a separate group of negatively charged CRPs in the water jasmine plant. Wrightides R4-R8 coeluted in the preparative RP-HPLC (Figure 3.8) and share identical sequences except some residues at the two termini. Among them, wrightide R4 is the largest peptide (35 residues) and wrightide R8 is the smallest (30 residues). Each peptide differing from the preceded peptide by 1 amino acid. In analytical RP-HPLC, these peptides were closely eluted in the range of 35-38% acetonitrile although the hydrophobicity increases slightly from wrightide R4 to R8 according to their

retention time. Wrightide R9 was cloned by genomic method and has not been detected at protein level.

As only the two most abundant wrightides R4 and R5 were detected in the crude extract, it is unclear whether wrightides R6-R9 are naturally present in the plant or they are the degradation products occurring during sample work-up. In any cases, the presence of two groups of CRPs in *W. religiosa* plant implies that plants may produce different CRPs in a single species to diversify their defense mechanism or to serve some endogenous functions.

Table 3.5. Putative cystine-knot α-amylase inhibitors from A. scholaris

Peptide		Sequence [a]	Approach [b]	MW [c]	No of residues	Charge
Intercysteine segment		**1 2 3 4**				
Wrightide R1	wR1	---CAQKGEYCSVYLQ-CCDPYHCTQPVIGGICA	P	3245	30	0
Wrightide R2	wR2	---CAQKGEYCSVYLQ-CCKPYQCTQPVIGGICA	G+P	3249	30	+1
Wrightide R3	wR3	---CAGINEFCNGWTIQCCDPYRCTGPVLGGLCA	P	3257	30	
Wrightide R4	wR4	EDECIPAGEPCSAENYNCCKGSICIWPLYR--CVGST	G+P	3803	35	-3
Wrightide R5	wR5	EDECIPAGEPCSAENYNCCKGSICIWPLYR--CVGS-	P	3702	34	-3
Wrightide R6	wR6	EDECIPAGEPCSAENYNCCKGSICIWPLYR--CVG--	P	3614	33	-3
Wrightide R7	wR7	-DECIPAGEPCSAENYNCCKGSICIWPLYR--CVGS-	P	3573	33	-2
Wrightide R8	wR8	--ECIPAGEPCSAENYNCCKGSICIWPLYR--CVGS-	P	3458	32	-1
Wrightide R9	wR9	EDECIPAGEPCSAESYNCCKGSICIWPLYR--CVGST	G	3776	35	-3

a Disulfide connectivity for CRPs is shown by the solid lines

b Approaches to obtain CRP sequences are either proteomic (P) or genomic (G)

c Monoisotopic masses are reported in this table.

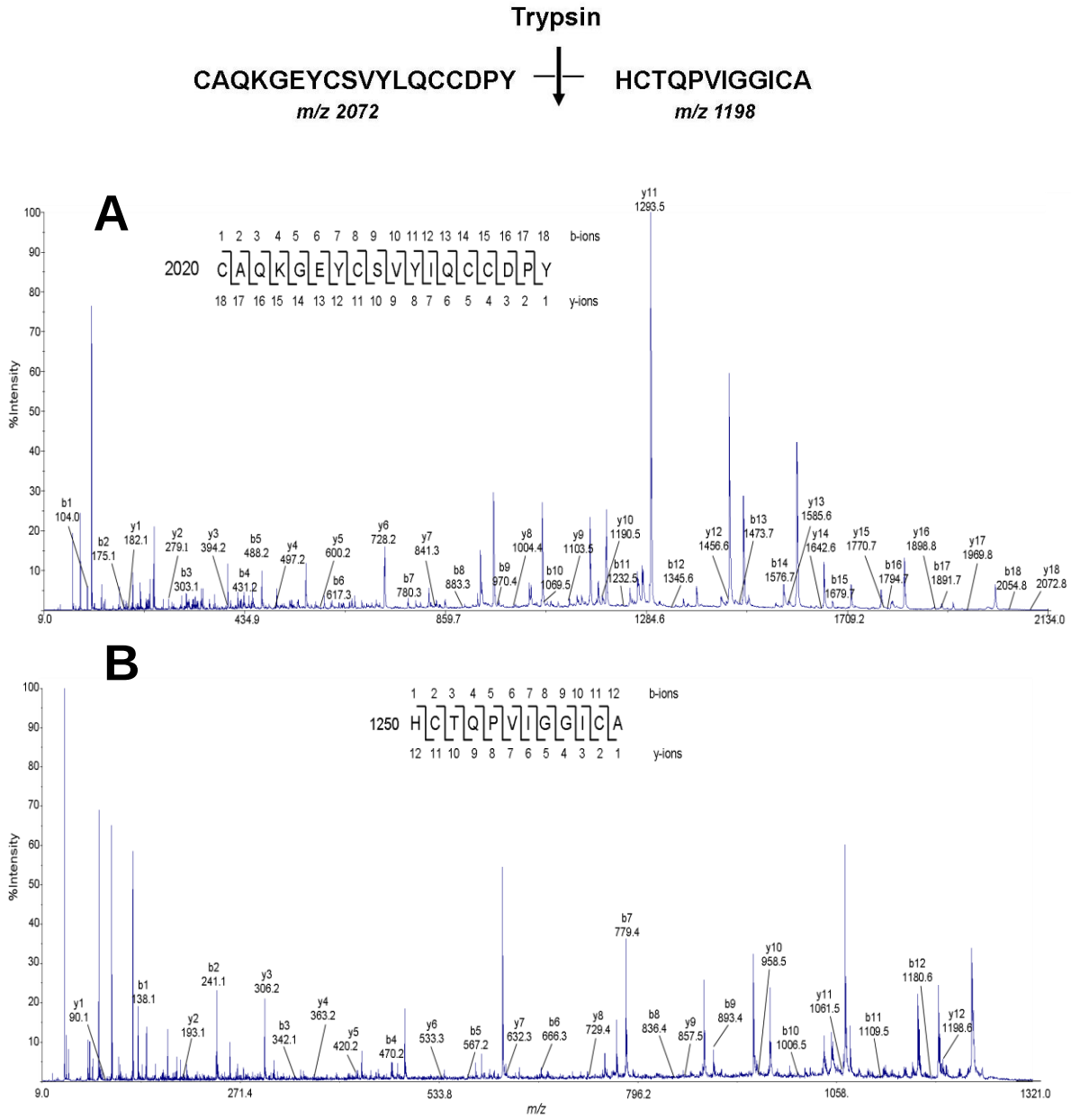


Figure 3.7. De novo sequencing of wrightide R1 using MALDI-TOF/TOF MS/MS. MS/MS spectra of 2072 Da and 1198 Da fragments obtained by trypsin digestion of wrightide R1 provide full wrightide R1 sequence.

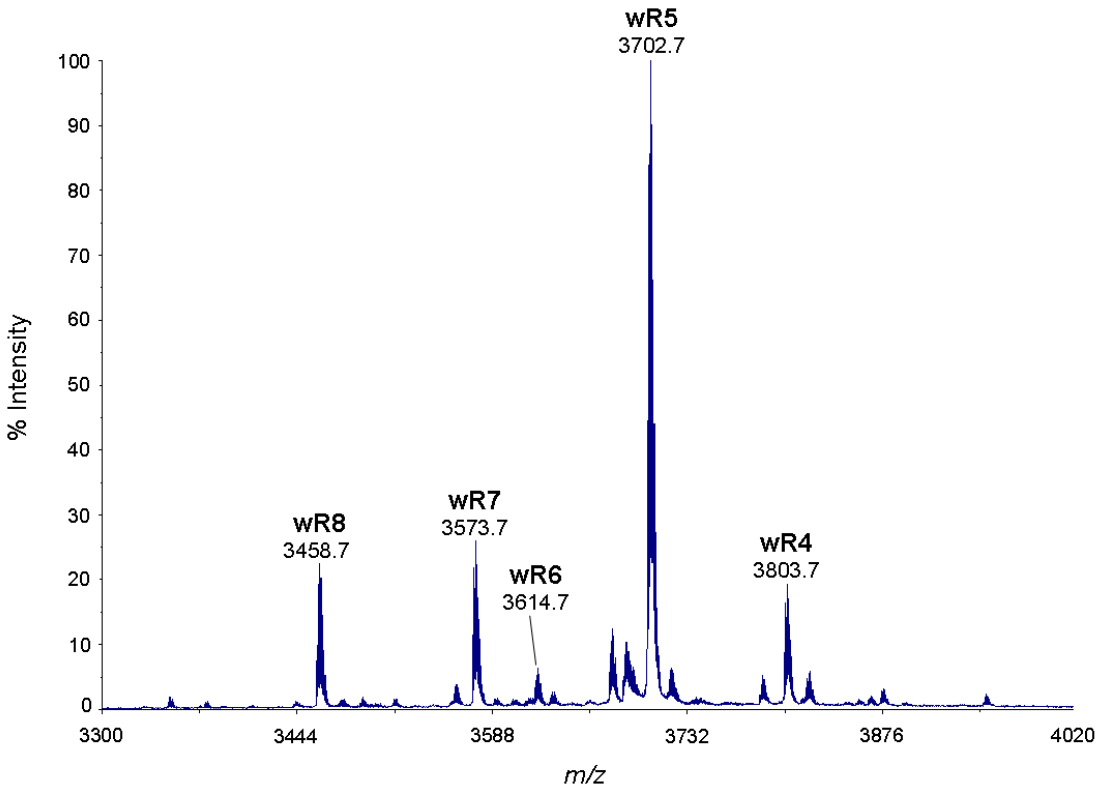


Figure 3.8. Coelution of wrightides R4-R8 monitored by MS. Wrightides R4-R8 differ from each other by one residue at either N- or C-terminus. During preparative RP-HPLC, wrightides R4-R8 were coeluted at ~36-38% acetonitrile. The spectrum shows the co-presence of wrightides R4-R8 in a single preparative RP-HPLC fraction. The peptides were separated closely in analytical RP-HPLC with the retention time increasing slightly from wR4 to wR8.

Tissue- and region-specific expression of CK α-amylase inhibitors

To get a better understanding of the tissue specificity of CK peptides in the Apocynaceae family, different plant parts from each species were collected and profiled in mass spectrometry. *W. religiosa* samples from Vietnam and Singapore were also compared to study the regional expression of CK α-amylase inhibitors.

Allamanda oenotheraefolia

Figure 3.9 presents allatide expression profiles in four different parts of *A. oenotheraefolia,* including flower, leaf, stem, and root. This figure highlights the abundance of 3-4 kDa peptides in *A. oenotheraefolia* with at least 20 peptides present in each plant part. Flower, leaf, and stem display high similarity in the expression profiles while the root expresses some distinct peptides (*m/z* of 3774, 3322, 3072, and 3037). In general, two main clusters, one around 3.0-3.5 kDa and the other around 3.7-4.1 kDa, were found in all tissues tested.

Alstonia scholaris

MS profiles of flower, leave, stem, and fruit of *A. scholaris* are shown in Figure 3.10. The peptide expression profiles in different tissues varied relative abundance, but the major peaks (*m/z* 3250, 3317, 3366, and 3385 corresponding to alstotides S1-S4) were found in all tested plant parts. Peptides with the mass higher than 3.5 kDa were not detected in all tested tissues in the experimental condition.

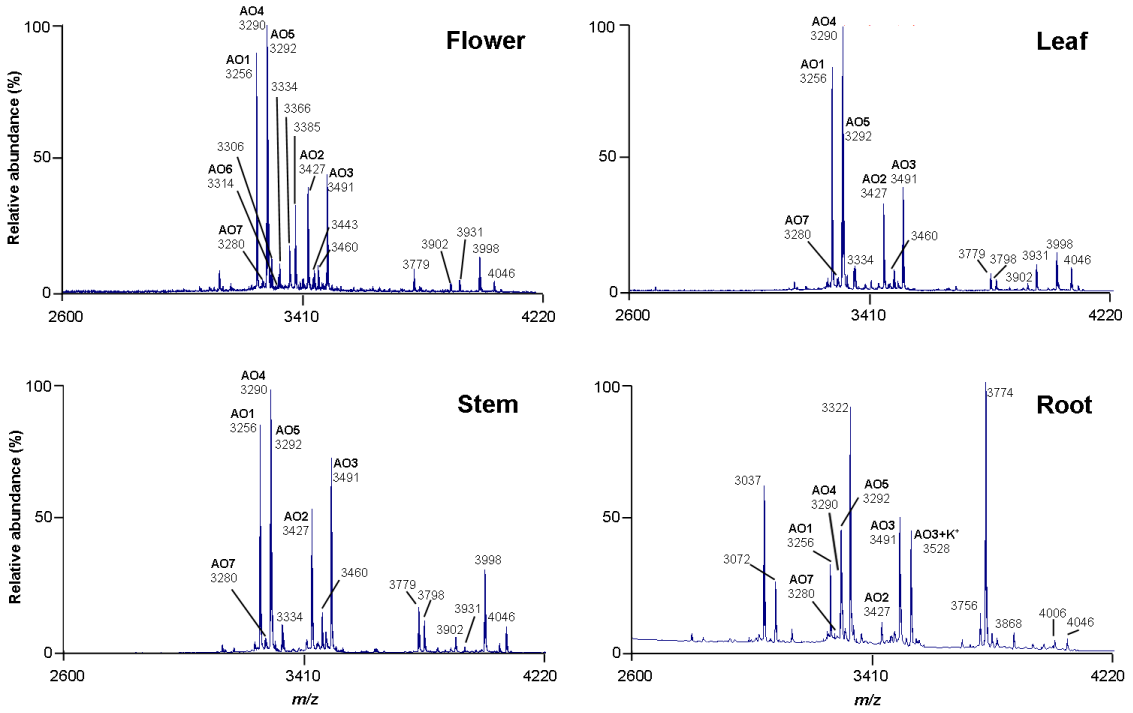


Figure 3.9. Tissue-specific expression of allatides from A. oenotheraefolia plant

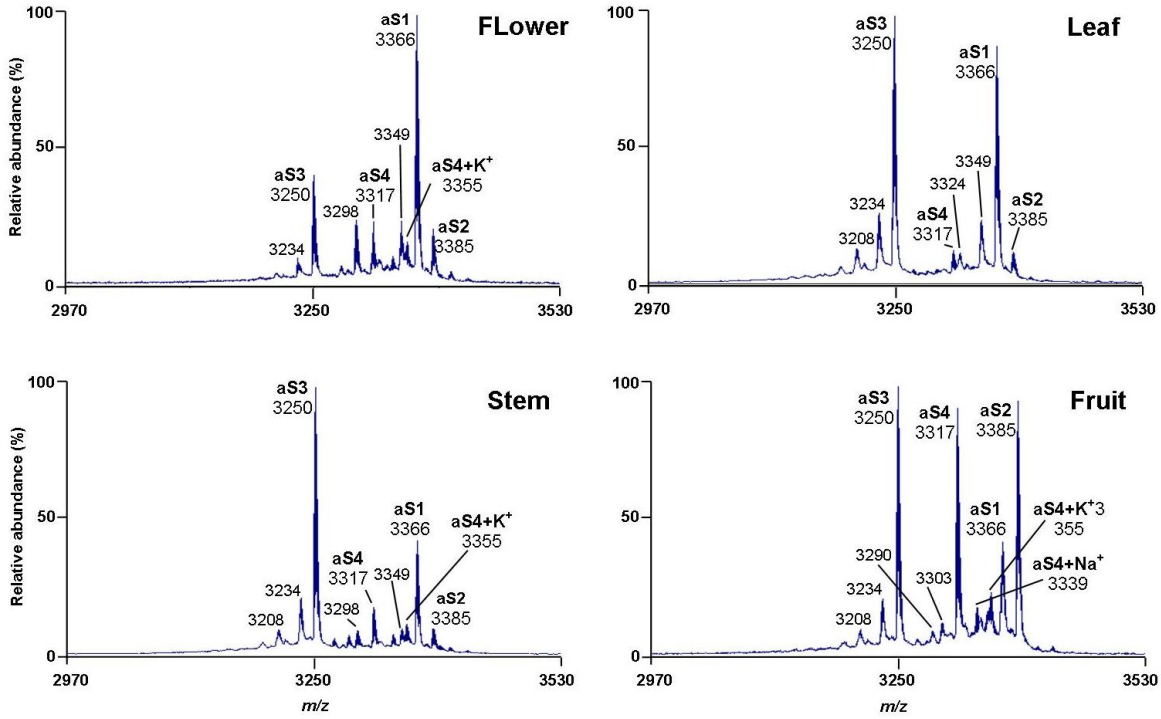


Figure 3.10. Tissue-specific expression of alstotides from A. scholaris plant

Wrightia religiosa

The expression profiles for flower and leaf of *W. religiosa* from Vietnam and Singapore were compared in Figure 3.11. Tissue and regional specificity can be observed in these spectra. Prominent expression of the peptides with *m/z* 3370 and 3407 was found in flowers from both countries, whereas individual wrightides R1 and R2 are the most abundant in Vietnam and Singapore leaves, respectively. Wrightides R4 and R5 were found ubiquitously in every sample. The cluster of higher mass range (~4.5 kDa) was visible in the spectra of the leaves but not of the flowers from both countries.

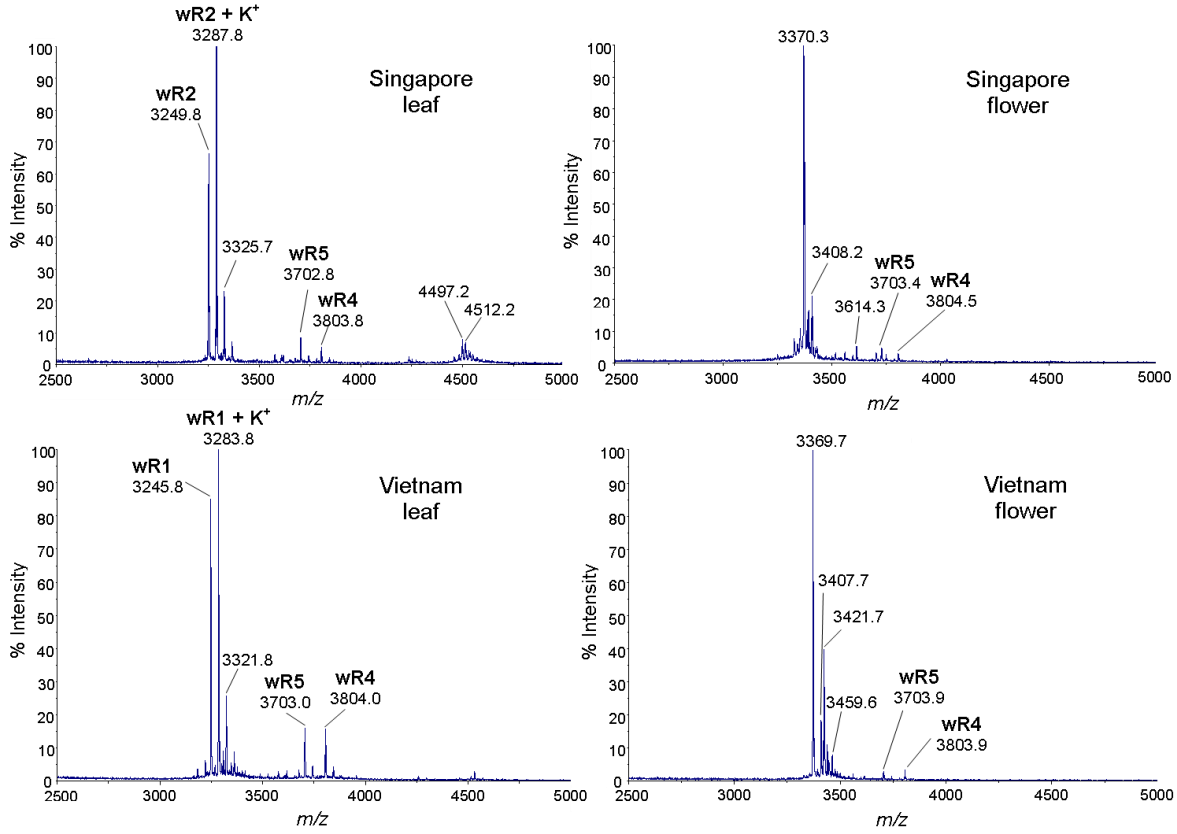


Figure 3.11. Tissue- and region-specific expression of wrightides from W. religiosa plant

Gene cloning of CK α-amylase inhibitors

Allamanda oenotheraefolia

To identify the genetic structures of allatide precursors, 3′ RACE PCR was performed using a degenerate primer targeting the sequence INQCCDPW. Partial 3′ end gene sequence of allatides O4/O5 was obtained. One specific primer targeting 3′ untranslated region (UTR) of the newly discovered gene was designed for 5′ RACE PCR. Three full-length genes encoding allatides O1, O4, and O5 were identified by assembling 3′ RACE and 5′ RACE clones and named *aoc1*, *aoc4*, and *aoc5*, respectively.

Figure 3.12 shows the deduced amino acid sequences of three allatide precursors. They contain an ER signal sequence followed by a pro-domain and a CRP-domain at the C-termini. The absolute identity in ER signal domain and pro-domain was observed for three allatide precursors (allatides O1, O4, and O5). To gain insights into the genetic structure of novel allatides at DNA level, a pair of gene-specific primers targeting 3′ and 5′ UTR were used. An intron located in the middle of ER signal sequence was revealed.

Alstonia scholaris

Allatides are gene-encoded CK peptides with three-domain precursors. To determine if alstotide gene transcripts are composed of similar domains, we studied alstotide precursor sequences in similar manner. RNA extract of fresh leaves was reverse-transcribed to 3' and 5' RACE cDNA libraries. A degenerate primer targeting the conserved INQCCDPY motif was used in 3' RACE PCR and the partial sequence of alstotide S1 was obtained. Next, we performed 5' RACE study to obtain the complete transcript of alstotide S1, named *asc1*. A specific primer against the signal peptide of *asc1* clone was subsequently used to amplify *asc2* and *asc4* clones of alstotides S2 and S4 transcripts, respectively. The three precursors encoded by *asc1*, *asc2*, and *asc4* were highly homologous to each other and to allatide clones (Figure

3.12). We also found an intron of 135 and 192 base pairs in the middle of the signal peptide of genomic sequences encoding alstotides S2 and S4, respectively.

Wrightia religiosa

Using the same genomic method, we obtained three full clones for wrightide R2, R4 and R9, designated as *wrc2, wrc4*, and *wrc9*, respectively. The clone *wrc2* appeared to be highly homologous with the allatide and alstotide precursors. In these precursors, the mature domain is followed immediately by a stop codon, which is one residue away from the last cysteine (CysVI). In contrast, *wrc4* and *wrc9* have a rather long tail of 9 residues after the CysVI, which corresponds to a 5-residue addition to the C-terminus of mature wrightide R4. At this point, it remains unclear whether the last 5-residue fragment is removed in a single cleavage by a specific endogenous enzyme to produce mature wrightide R4. Or alternatively, it could be gradually trimmed to produce a series of products differing by one residue, among which wrightide R4 and R5 are the most abundant.

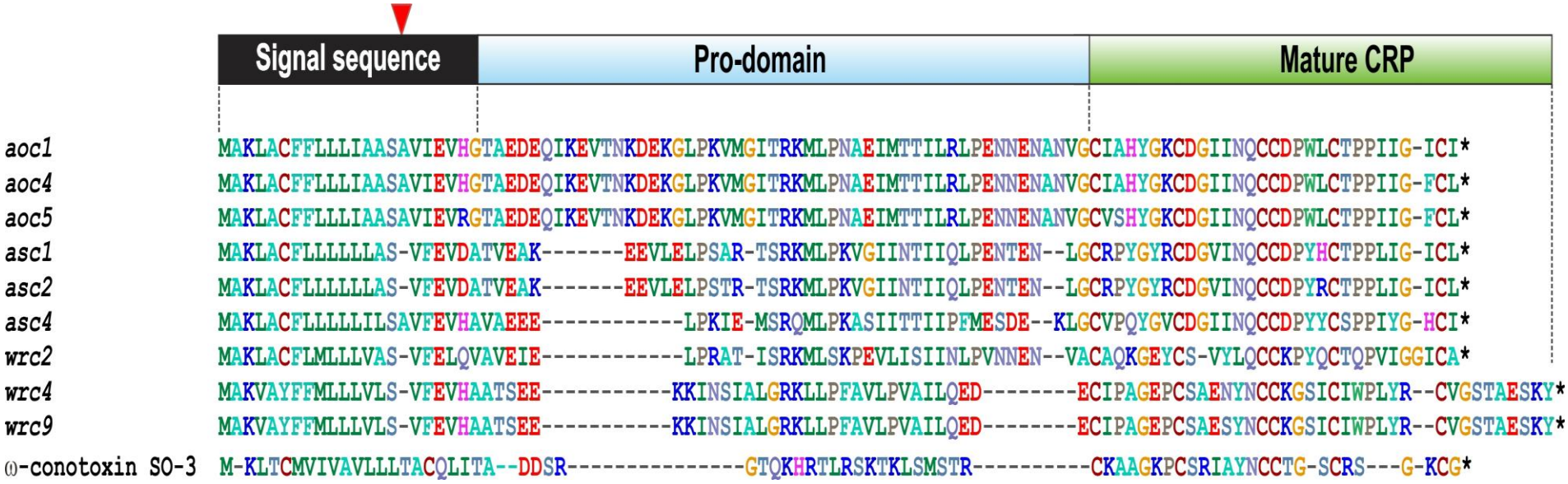


Figure 3.12. Sequence alignment of CK α-amylase precursors. Three sequences were obtained by clones from 3′ RACE PCR, 5′ RACE PCR, and DNA PCR. The precursors generally comprise three domains: ER signal domain, pro-domain, and mature CRP domain. It remains unclear whether the last 5 residues of wrc4 and wrc5 belong to another domain, the C-terminal tail. One single intron is located in the middle of ER signal sequence, indicated by red inverted triangle. The clones wrc4 and wrc5 do not contain intron whereas the intron information has not been confirmed for wrc2. Precursor sequence from ω conotoxin SO-3 [186] was included for comparison.

Top-down disulfide mapping

Allatide O2

To confirm the putative CK fold of allatides, we used a topdown approach to map the disulfide connectivity of allatide O2 by a differential *S*-reduction and *S*-alkylation method to label cysteine residues [187]. Allatide O2 was partially reduced with tris(2-carboxyethyl) phosphine (TCEP) by cleaving one or two disulfide bonds to generate 1*SS* and 2*SS* intermediates. The *S*-reduced species were immediately *S*-alkylated with an excess of NEM. Both steps were performed under acidic condition (pH 3.5) to prevent the scrambling of disulfide bonds. The partially NEM-tagged species were purified by RP-HPLC.

Figure 3.13 shows the HPLC profile with four separated peaks, which were collected and analyzed by MALDI-TOF MS. The mass-shifts of *S*-alkylated species, as compared to the native allatide O2, could be used to predict the number of reduced disulfide bonds. Each NEM-tagged cysteine increased the *m/z* value by 126; hence, the mass gain of 252 Da, 504 Da, and 756 Da would indicate peptides with 2-, 4-, and 6-tagged cysteines, respectively. Peptides eluted in peak 1 (*m/z* 3679) were assigned as a 1*SS* intermediate having four *S*-alkylated cysteines and one intact *SS* bond. Peak 3 was the 2*SS* intermediate having two *S*-alkylated cysteines and two intact *SS* bonds. Peak 4 was completely reduced allatide O2 with all six cysteines being NEM-tagged. To determine their disulfide connectivity, 1*SS* and 2*SS* species were subsequently fully reduced with DTT and *S*-alkylated with another alkylation reagent (IAA). The *S*-tagged peptides were digested with trypsin, and tryptic fragments were analyzed by tandem mass spectrometry. The trypsin cleaved allatide O2 after the Arg7, and two digested fragments were obtained. Figure 3.14 shows the b- and y-ion assignments of the C-terminal fragments of the 1*SS* and 2*SS* species, containing CysII-VI. The alkylated groups on CysI were deduced to obtain even numbers of both *S*-NEM and *S*-IAA tags. The connectivity of the 2*SS* species was determined to be CysI-IV with two NEM-

tagged groups on CysI and CysIV and four IAA-tagged on other cysteine residues. Likewise, another disulfide bridge was identified as CysII-V because the 1*SS* species had two IAA-tagged groups on CysII and CysV and four NEM-tagged groups on remaining cysteine residues. The third *SS* bond was deduced to be CysIII-VI. Taken together, our results unequivocally confirmed that the disulfide connectivity of the allatide O2 has a CK motif of CysI-VI, CysII-V, and CysIII-VI.

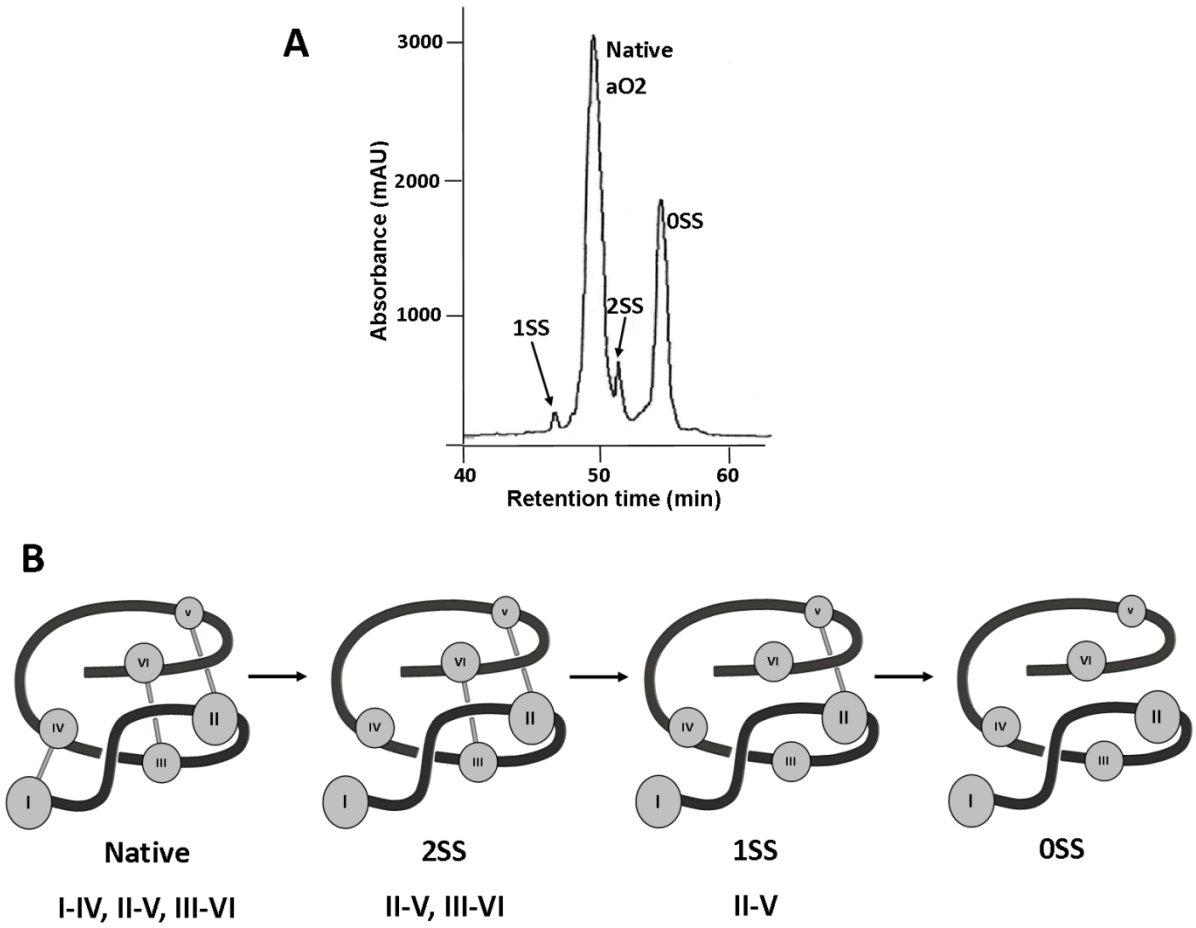


Figure 3.13. Top-down disulfide mapping of allatide O2. ***A****, RP-HPLC profiles of partially S-reduced and NEM-alkylated allatide O2. 1SS-intermediates contained one intact disulfide bond, 2SS-intermediates two intact disulfide bonds, and 0SS contained fully S-NEM alkylated allatide O2;* ***B****, Schematic representation of unfolding pathway of allatide O2 showing CysI-IV bond is likely the first to be broken following by CysIII-VI bond and finally by CysII-V bond.*

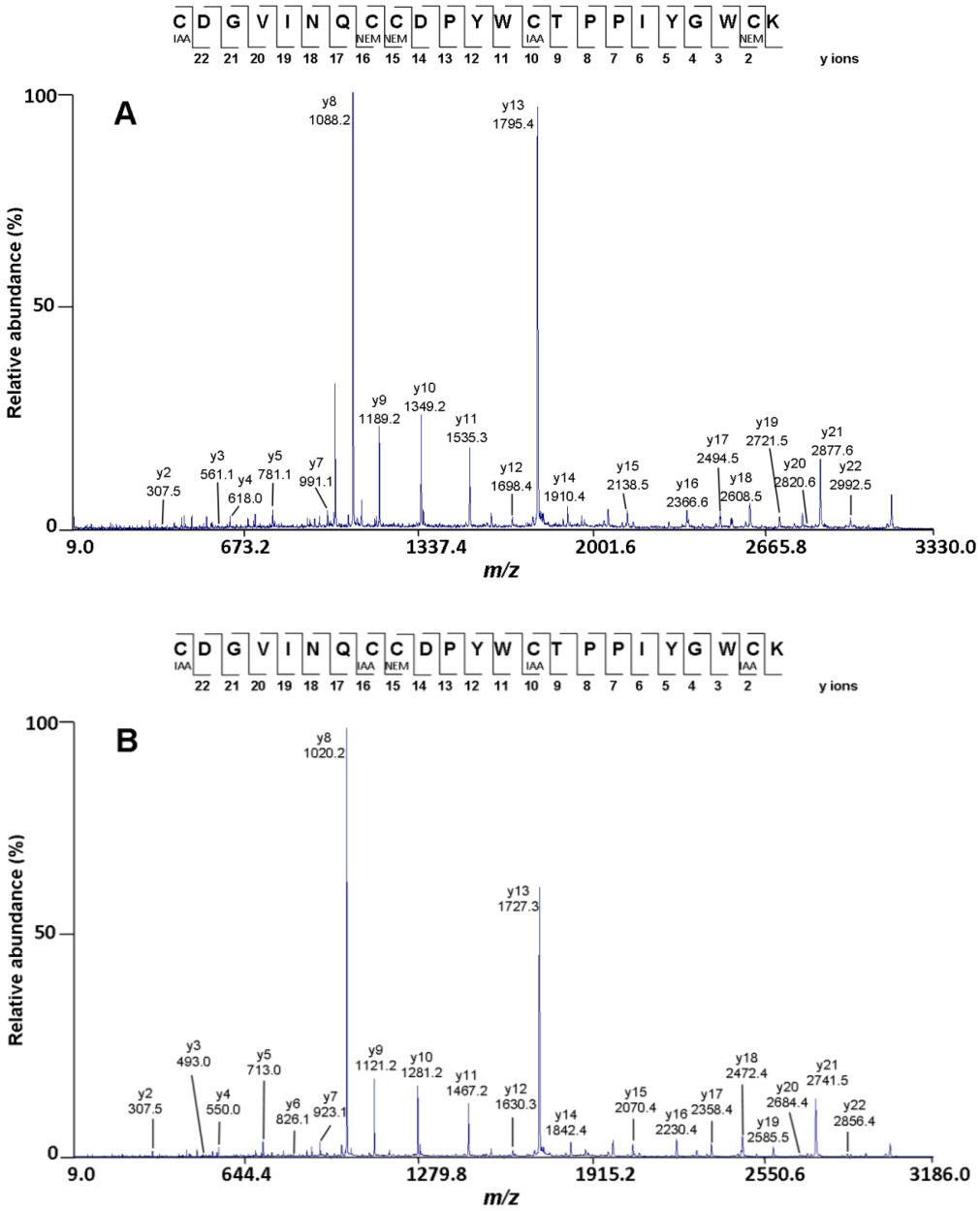


Figure 3.14. Tandem MS/MS profiles for disulfide connectivity determination of allatide O2. The figure shows MS/MS profiles of CDGVINQCCDPYWCTPPIYGWCK fragments obtained by tryptic digestion of fully NEM- and IAA-alkylated allatide O2. 1SS- and 2SS-intermediate species collected after RP-HPLC in Figure 3.13A were S-tagged with IAA. MS/MS profiles of C-terminal fragments of the fully alkylated allatide O2 derived from 1SS and 2SS species were showed in Figure 3.14A and 3.14B, respectively. The disulfide connectivity was confirmed to be CysI-IV, CysII-V, and CysIII-VI.

Alstotide S3

Though highly homologous to allatide O2, partial reduction of alstotide S3 at 37°C in various reducing conditions did not yield significant amounts of intermediates as obtained for allatide O2 (Table 3.6). TCEP reacts kinetically, not thermodynamically. Thus, we employed a fast, high-temperature partial reduction at 75°C for 3.5 min by TCEP followed by immediate HPLC purification (Figure 3.15A and 3.15B). Intermediates with one or two remaining disulfide bonds were labeled 1*SS* and 2*SS*, respectively. Intermediate species were then *S*-alkylated at reduced half-cystine with excess NEM (Figure 3.15C) and monitored by MALDI-TOF MS. In the experimental condition, intermediates with incomplete NEM-tagging (labeled as 3-NEM, 5-NEM) were coeluted with 1*SS* and 2*SS* species (4-NEM and 2-NEM, respectively) in partially resolved peaks from shallow-gradient RP-HPLC. Extended incubation with NEM led to a reduced amount of 2*SS* species in the mixture which was inadequate for subsequent analysis.

Mixtures of 1*SS* or 2*SS* with incompletely NEM-tagged species were fully reduced with dithiothreitol. The products were subsequently alkylated with IAA. Each differentially labeled species was digested by chymotrypsin and sequenced by tandem MS/MS through b- and y- ion assignments of each chymotryptic fragment (Figure 3.16). The analysis of 2*SS* species with two NEM tags on CysI and CysIV revealed a disulfide bond between these two cysteine residues. Likewise, a disulfide linkage between CysIII and CysVI was implicated by the positions of two IAA groups in 1*SS* species. Derived from these results, the last disulfide bond was between CysII and CysV. Thus, alstotide S3 possesses a CK motif.

Table 3.6. Different conditions for partial reduction of alstotide S3. About 0.3 mg of alstotide S3 dissolved in 0.5 ml of 30% acetonitrile /H_2O and two-fold volume of TCEP in 200 mM citrate buffer, pH 3 were incubated at different temperatures for varying time periods. Condition 9 provided satisfactory amount of intermediates for subsequent S-alkylation step.

Condition	TCEP concentration	Temperature	Time	Reference
1	100mM	64°C	3 min	[188]
2	50mM	37°C	30 min	Used for aO2
3	200mM	37°C	60 min	This work
4	200mM	64°C	3 min	This work
5	200mM	16°C	16-24hr	This work
6	200mM	75°C	5 min	This work
7	40mM	75°C	5 min	This work
8	40mM	80°C	5 min	This work
9	**40mM**	**75°C**	**3.5 min**	**This work**

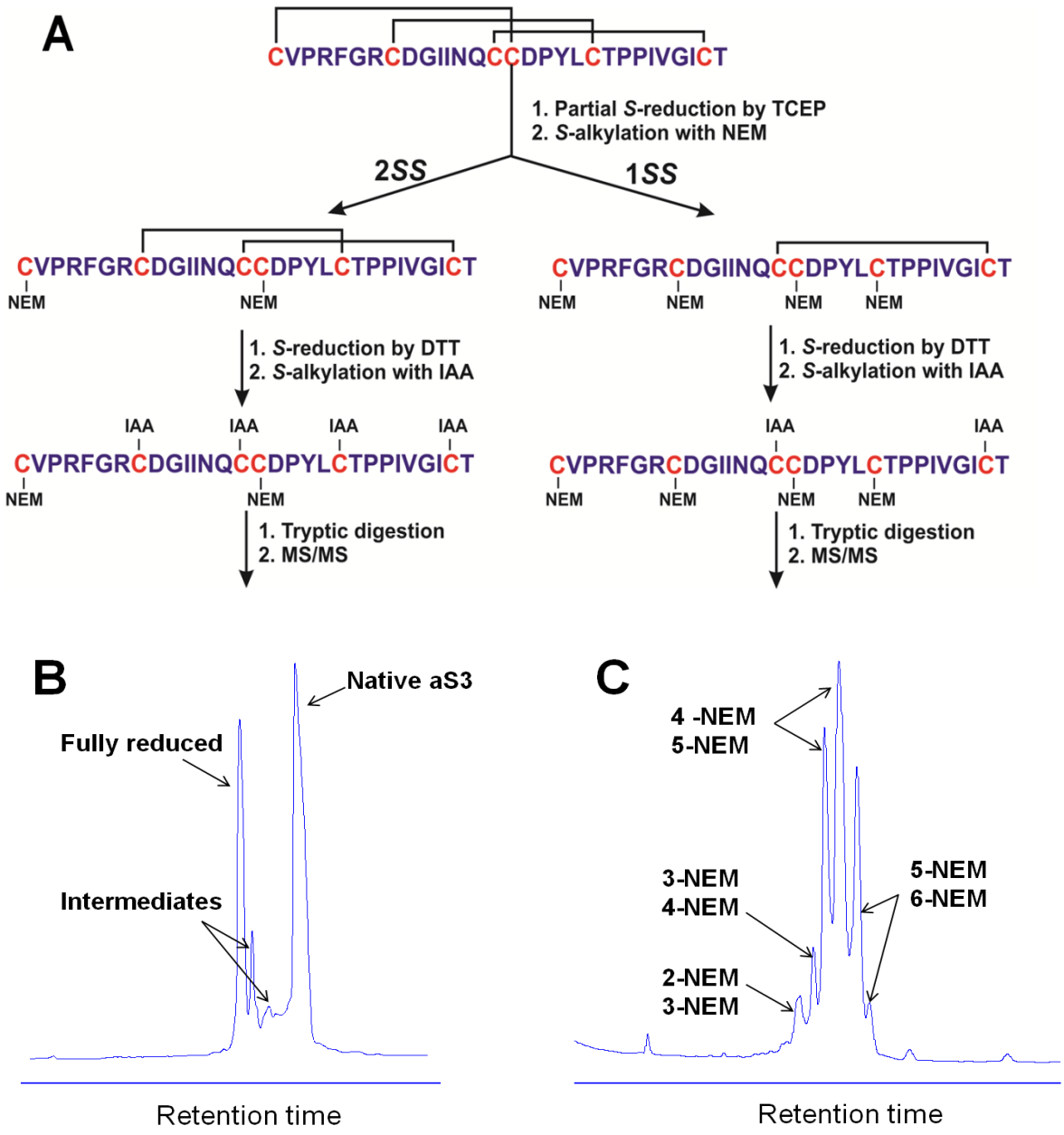


Figure 3.15. Top-down disulfide mapping of alstotide S3. ***A****, Schematic diagram of partial reduction and S-alkylation for disulfide mapping of alstotide S3;* ***B****, HPLC profile of partially reduced alstotide S3 after treatment with 40 mM of TCEP at 75℃ for 3.5 min. Partially resolved peaks were collected from a RP-HPLC run with 38-40% acetonitrile, 30 min gradient;* ***C****, HPLC profile of NEM- alkylated alstotide S3. The profile was run on a gradient of 38.5-39.8% acetonitrile in 60 min.*

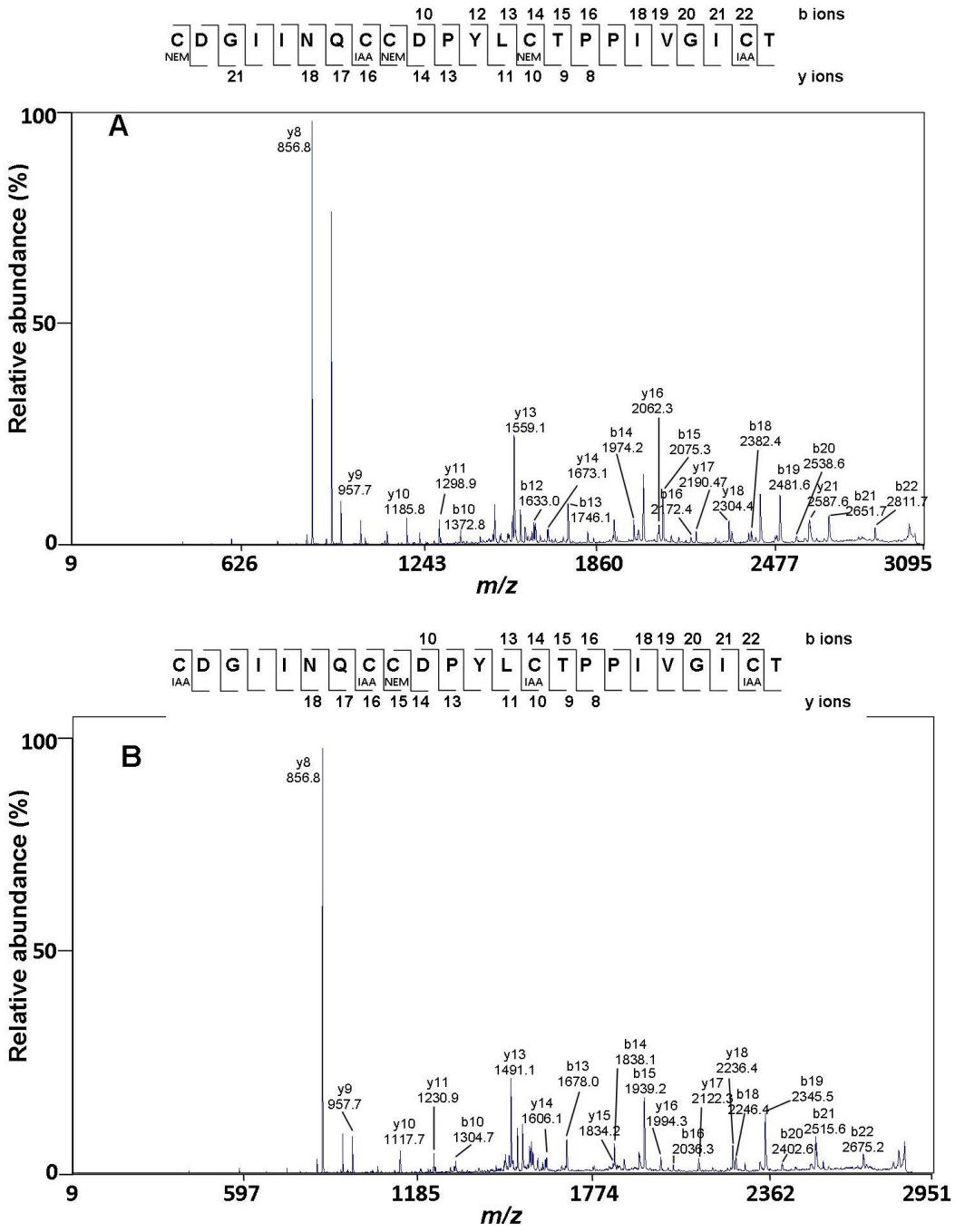


*Figure 3.16. Tandem MS/MS profiles for disulfide connectivity determination of alstotide S3. The figure shows MS/MS profiles of tryptic fragments of alkylated 1SS- (**A**) and 2SS- (**B**) intermediates. The disulfide connectivity was confirmed to be CysI-IV, CysII-V, and CysIII-VI.*

Heat and proteolytic stability

Heat treatment and protease degradation generally lead to significant loss of bio-active macromolecules in herbal remedies. To determine whether allatides would survive boiling and protease degradation, allatide O4 was heated at 100°C for 1 h or incubated with chymotrypsin for 4 h. The intact peptide remaining after heat and protease treatment was monitored by UPLC and mass spectrometry**.** More than 90% of allatide O4 was observed in the UPLC profiles after heat treatment (Figure 3.17A). The MS profiles of corresponding peaks showed that both peaks contained mainly native allatide O4 (*m/z* 3290) with a small amount of degraded products.

In enzymatic stability test, allatide O4 fully reduced with dithiothreitol served as the control, which was completely hydrolyzed after 4-hour incubation with chymotrypsin at 37°C (Figure 3.17C and 3.17D). Native allatide O4, however, was relatively resistant to protease degradation with >95% of peptides remaining intact (Figure 3.17B).

In similar manner, we investigated thermal and enzymatic stability of alstotides and wrightides (Figure 3.18 to 3.21). The results showed that >90% of alstotides and wrightides tested survived the heat and proteolytic treatment. Our results provided a strong evidence of the stability of CK α-amylase inhibitors against thermal and proteolytic treatment.

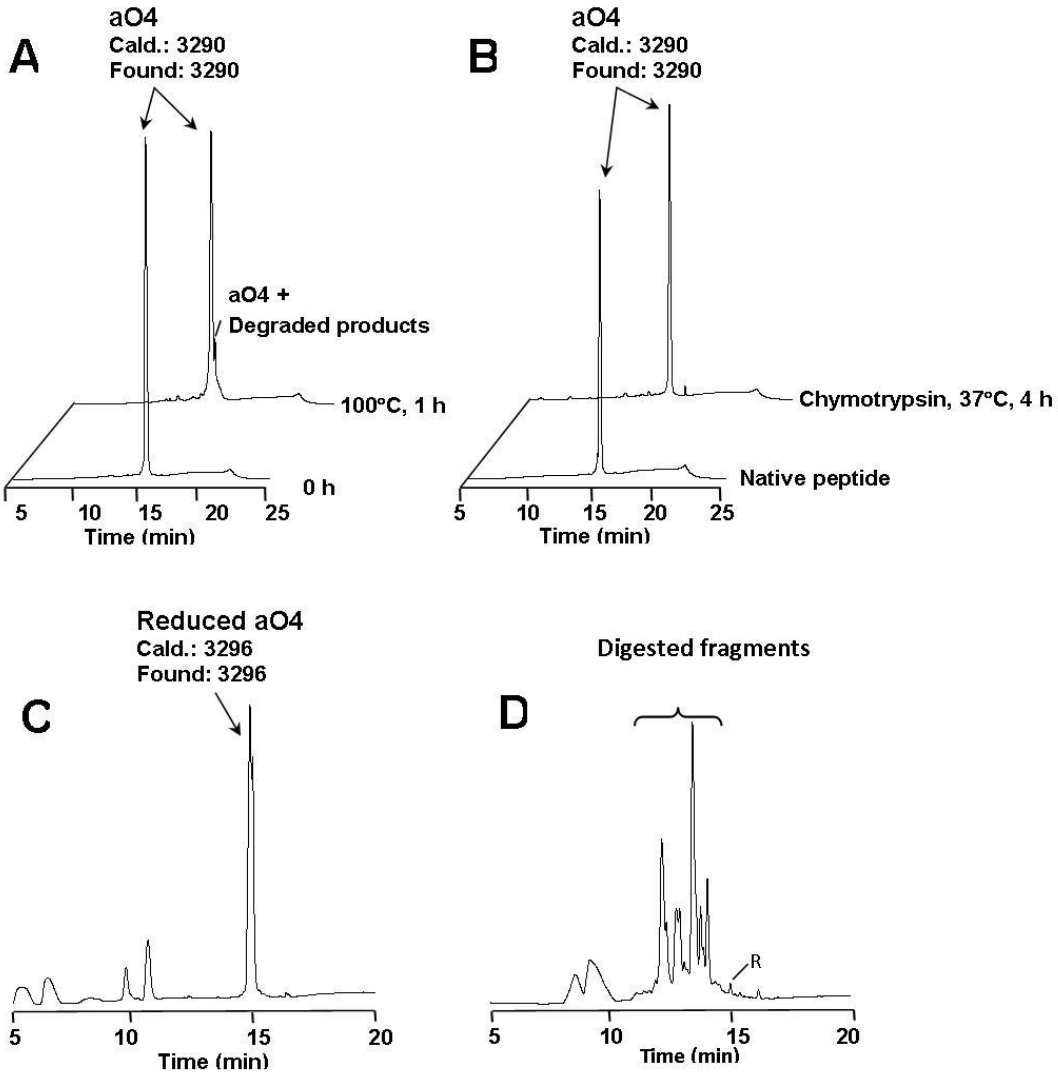


*Figure 3.17. Thermal (**A**) and enzymatic (**B**) stability of allatide O4. The peptide was heated at 100℃ for 1 h or incubated with chymotrypsin at 37℃ for 4 h. Allatide O4 was resistant to both treatments. The minor peak in the profile (**A**) after heat treatment contained mainly allatide O4 with a small amount of degraded products as examined by MS profile. Enzymatic treatment of fully reduced allatide O4 served as a control. Allatide O4 fully reduced with dithiothreitol (**C**) was completely digested with chymotrypsin at 37℃ for 4 h (**D**).*

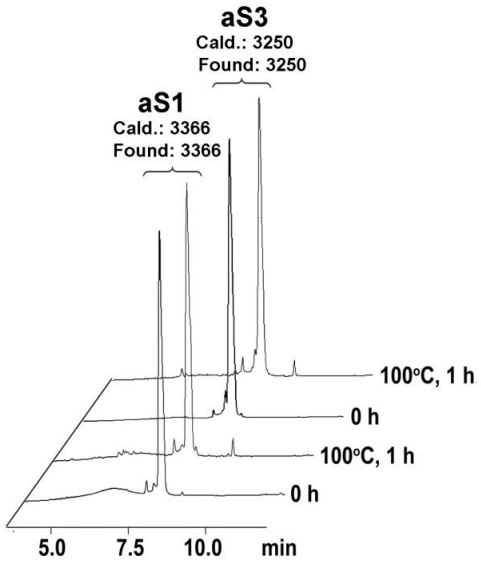


Figure 3.18. Heat stability of alstotides S1 and S3. The peptides were dissolved in water and heated in boiling water for 1 h. The treated samples and the untreated controls were subjected to UPLC (Vydac C18-TP) using TFA (0.1%) as an ion-pairing agent.

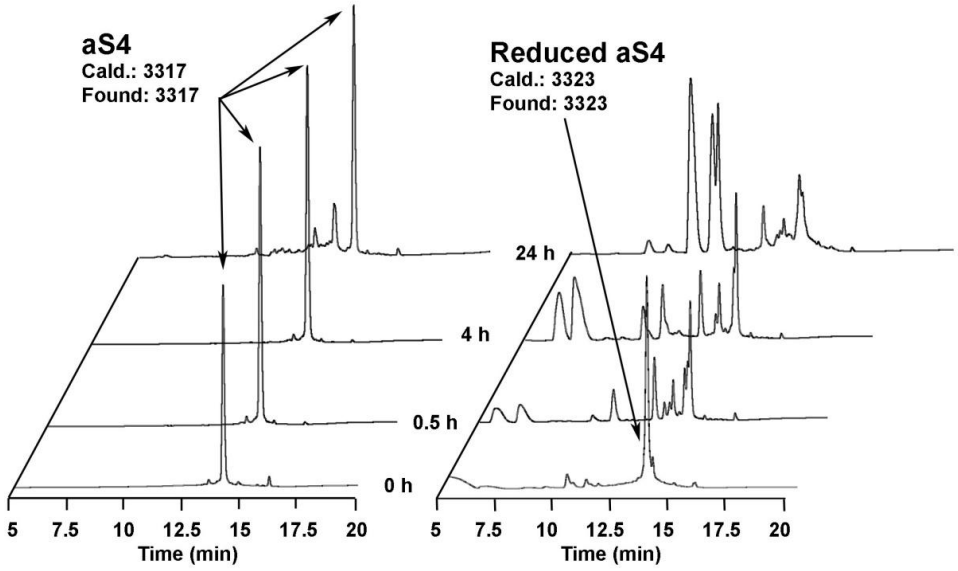


Figure 3.19. Proteolytic stability of alstotide S4. Alstotide S4 was digested with chymotrypsin for 0.5, 4, and 24 h and the products monitored by UPLC and MS (left panel). Reduced aS4 was treated in similar conditions to serve as a control (right panel)

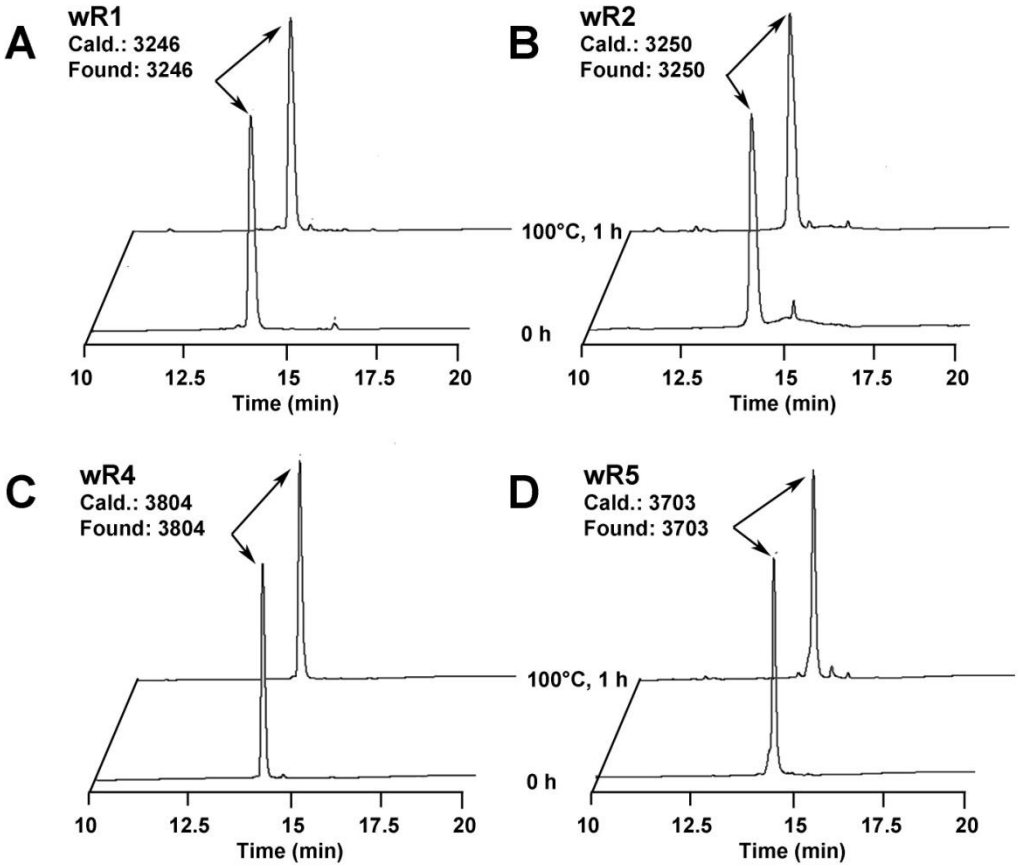


Figure 3.20. Heat stability of wrightides

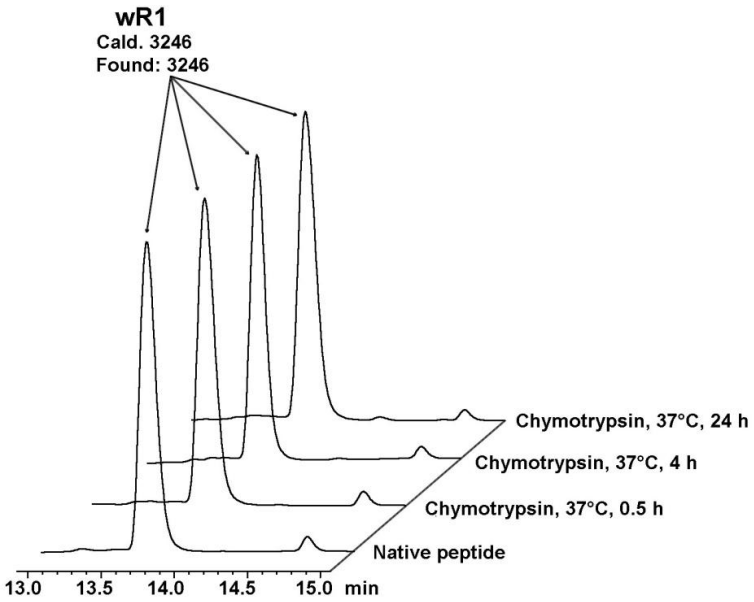


Figure 3.21. Proteolytic stability of wrightides

Amylase inhibitory activity

To determine the α-amylase inhibitory activities of allatides, we carried out the inhibition assays with α-amylases from yellow mealworm (*Tenebrio molitor*), human saliva, porcine pancreas, and fungus (*Aspergillus oryzae*). *Tenebrio molitor* α-amylase (TMA) was known to be inhibited by AAI [62]. Since TMA is not commercially available, we purified the enzyme using ammonium sulfate precipitation followed by an anion exchange chromatography at pH 5.4 and dialysis. Amylase inhibitory activities of allatides O4, O2/3, and O1/5 (allatides O2/3 and O1/5 were tested as mixtures due to coelution) were examined using Bernfield method [189]. Allatides exhibited inhibitory activities against TMA in a dose-dependent manner with IC_{50} ranging from 2 to 3 μM (Figure 3.22A). Allatides O1 to O5 at up to 100 μM did not inhibit α-amylases from fungus or mammals.

Similarly, activity of TMA before and after incubation with alstotides (aS1, aS3 or aS4) or wrightides (wR1, wR2, wR4, and wR5) were determined. The results showed that all alstotides and wrightides tested had strong inhibitory effects towards TMA (Figure 3.22B and 3.22C). IC_{50} of alstotides and wrightides varies in a narrow range from 1.9 to 5.2 μM. Similar to AAI and allatides, alstotides and wrightides did not inhibit human and fungal α-amylases at up to 100 μM.

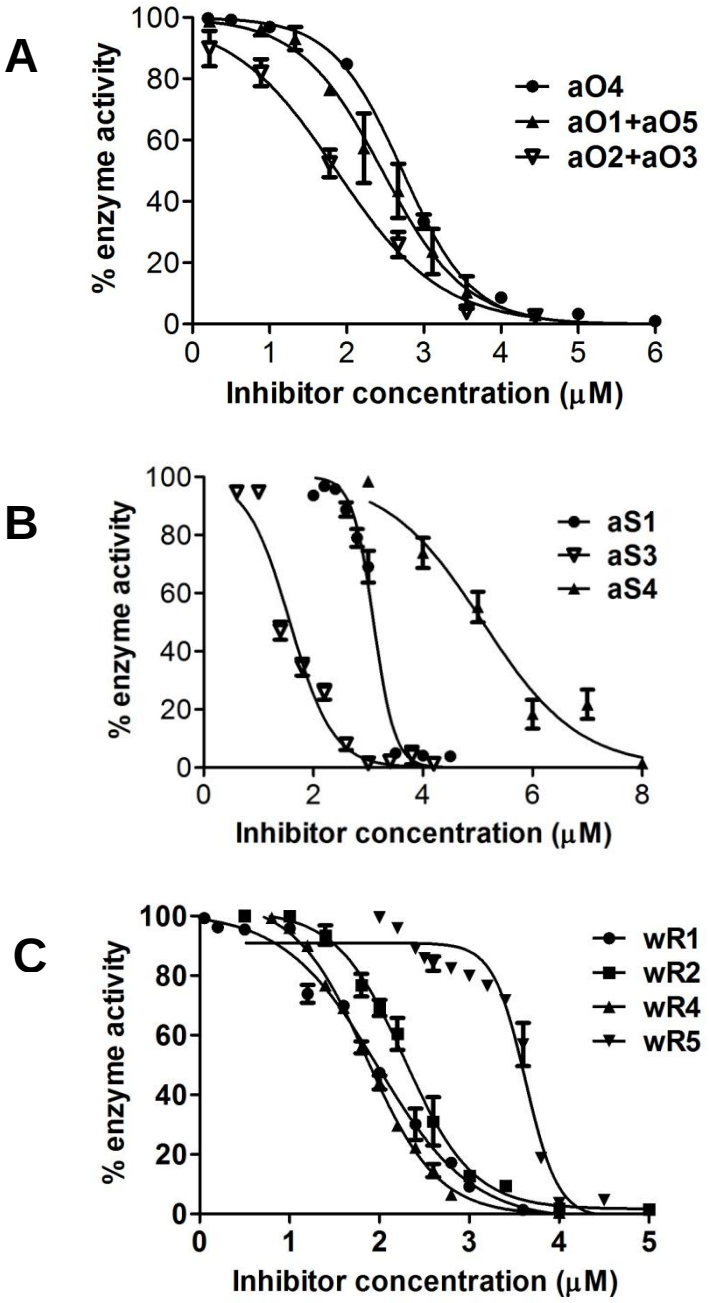


Figure 3.22. Inhibition of Tenebrio molitor α-amylase by CK α-amylase inhibitors. Peptides were preincubated with TMA for 20 min at 37˚C. Hydrolysis was started by adding 1% starch, proceeded for 5 minutes, and stopped by adding a color reagent containing DNS.

Bioactivity assays

Purified CK α-amylase inhibitors, including allatides O4, O2/3, O1/5, alstotides S1, S3, and wrightides R2, R4, R5, were examined for their cytotoxic, hemolytic, and antibacterial activities. In cytotoxicity assay, the viability of BHK, Huh-7 and Vero cells after treatment with peptides was measured using Presto blue reagent as the indicator. Allatides (aO4, aO2/3, aO1/5) and wrightides (wR2, wR4, and wR5) were noncytotxic to three cell lines tested at concentrations up to 100 μM. Alstotides S1 and S3 did not affect the proliferation of Vero and Huh-7 cells at up to 100 μM. However, they inhibited 60% and 70% of BHK cell viability at 100 μM, respectively.

The hemolytic effects were investigated on human type O erythrocytes. Cliotide T4, a cyclic peptide with strong hemolytic activity [190], was utilized as positive control. Allatides O4, O2/3, O1/5, alstotide S1, and wrightides R2, R4, R5 did not exhibit any hemolytic effects at concentrations up to 100 μM. In contrast, alstotide S3 lysed erythrocytes with HD_{50} of 47 μM, which was significantly weaker than that of the positive control melittin (HD_{50} 3.5 μM).

We examined antibacterial activities of CK α-amylase inhibitors against *E. coli* using radial diffusion assay. D4R, a peptide dendrimer with a broad-spectrum antimicrobial activity and prepared by in-house peptide synthesis [191], was used as positive control. All tested peptides did not exhibit bactericidal effects against *E. coli* at concentrations up to 100 μM.

CK α-amylase inhibitors from *A. oenotheraefolia*, *A. scholaris*, and *W. religiosa* were also screened for antiviral activity against infectious bronchitis virus (IBV) in the concentration range of 0-100 μM. Among all the tested CK peptides (aS1, aS3, aO4, wR2, and wR4), only aS1 and aS3 exhibited antiviral properties in the experimental condition. Further study was performed to gain insights into the antiviral mechanism of alstotide S1 and will be presented in Chapter 4.

Computer modeling of 3D structure

With the knowledge of the knotted structure of allatides, we used the module Knoter1D3D [192] to construct three-dimensional structure of alstotide S3 based on the crystal structure of AAI (PDB number 1qfd [71]). Alstotide S3 had an egg-like shape, compact structure, and similar fold as AAI (Figure 3.23) except for the lack of the α-helix in loop 2 and one β-strand at the end of loop 3 (containing CysV) in alstotide S3 structure. The motif of α-helix and antiparallel β-sheet connected by disulfide bonds, known as cysteine-stabilized alpha beta (CSαβ) motif [193, 194], has been found in AAI and many cysteine-rich bioactive peptides [195-197].

Among the residues involved in AAI and TMA interactions [198], the Pro20 and Arg7 (AAI numbering) residues were believed to play an important role in α-amylase inhibitory activity of AAI. The Pro20 experienced a trans-cis conformational change upon being contact with TMA to stabilize the AAI-TMA complex. This residue is conserved in all CK α-amylase inhibitors found in this study. In alstotide S3, the according Pro18 is placed in a similar position as compared to AAI. The Pro18 residue could therefore undergo similar structural changes after the binding of alstotide S3 to TMA enzyme. In contrast, the Arg7, which was previously shown to directly interact with the catalytic Asp287 of TMA via a salt bridge, may not be necessary for CK inhibitors' activity. This residue is missing in alstotide S4. However, alstotide S4 exhibited comparable α-amylase inhibition as compared to alstotides S1 and S3 in our *in vitro* amylase assay (Figure 3.22B).

Although the overall shape of alstotide S3 showed a close resemblance to that of AAI, a significant difference in the surface hydrophobicity between the two peptides was observed (Figure 3.23C and 3.23D). In AAI, the hydrophilic hemisphere comprised of residues Lys4-Arg7 and Tyr21-Ser32 was made up of mostly polar and charged residues [199]. This hemisphere was shown to be responsible for 27 out of the total 40 contacts (distance < 4.0 Å) between AAI

and TMA. On the contrary, the corresponding intercysteine segment 4 of alstotide S3 including residues Tyr19-Thr30 consisted of eight nonpolar and only three polar, uncharged residues, which potentially changed the type of interaction between inhibitors and enzyme TMA.

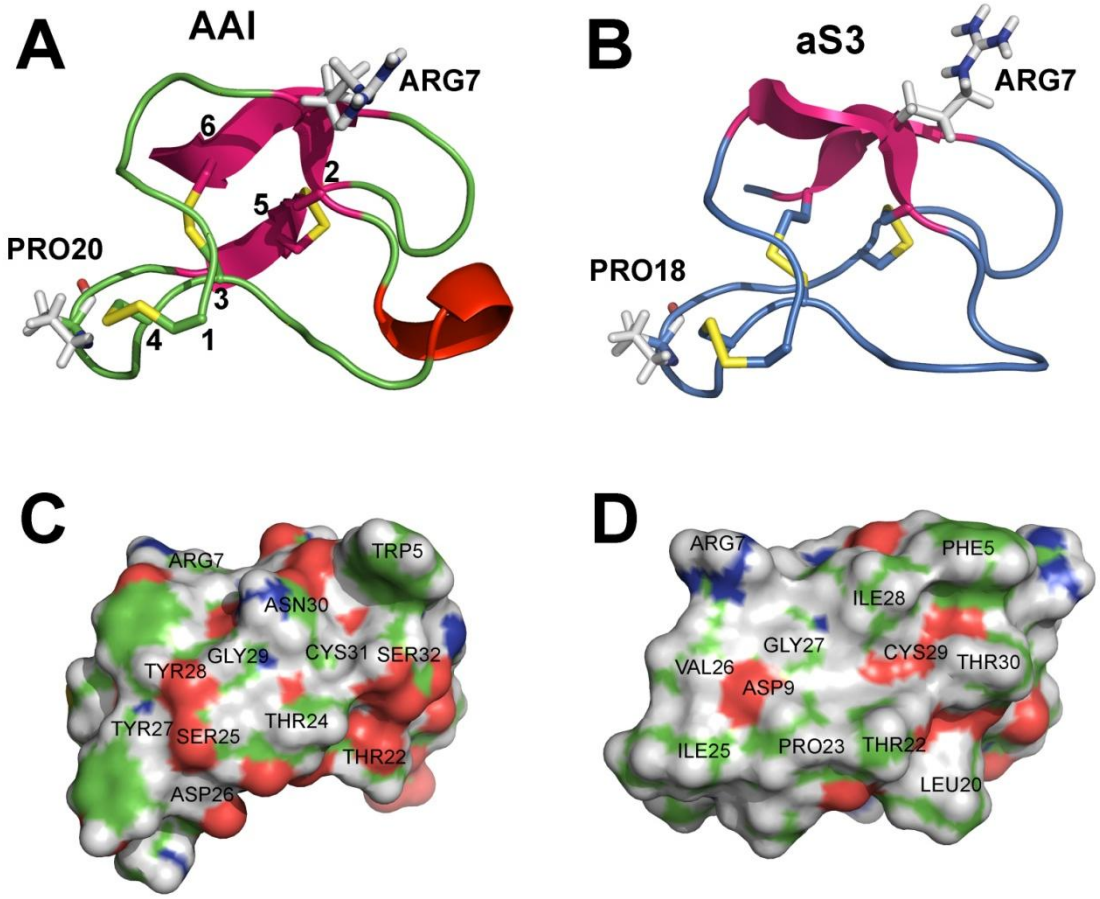


Figure 3.23. Three-dimensional modeling structures of AAI and alstotide S3. The structure of alstotide S3 (B and D) was constructed using the module Knoter1D3D, available at http://pat.cbs.cnrs.fr [200], based on AAI template (A and C, PDB number: 1qfd). The images were generated using Pymol program. There is substantial similarity in the overall conformations of alstotide S3 and AAI. The Arg7 and Pro18/Pro20 are highlighted because of their putative importance for inhibitory activity. Figures 3.23B and 3.23D compare the active hydrophilic hemisphere in AAI and the corresponding yet more hydrophobic hemisphere in alstotide S3.

Discussion

Proteinaceous α-amylase inhibitors are widespread in nature, and medicinal plants potentially represent a rich source of diverse biologics [85, 190, 201]. In this work, we report on the discovery and characterization of 20 novel CRPs from the Apocynaceae family. They include seven allatides (aO1-aO7) from *A. oenotheraefolia,* four alstotides (aS1-aS4) from *A. scholaris*, and nine wrightides (wR1-wR9) from *W. religiosa* species. These peptides contain 30 amino acid residues and adopt the same cystine-knot disulfide arrangement found in ω-conotoxins and AAI. They inhibit starch hydrolysis by the yellow mealworm α-amylase. Our work provides the first evidence of the existence of a CK-type of α-amylase inhibitors from Apocynaceae plants.

It has been 18 years since the AAI discovery in 1994, and AAI has remained the only representative of small cysteine-rich peptides with 32 amino acid residues and a cystine-knot structure that exhibits α-amylase inhibitory activity [202]. The discovery of allatides, alstotides, and wrightides extends the list of CK peptides in plants that display α-amylase inhibitory activities. With two amino acids shorter than AAI, the CK family in Apocynaceae plants now represent the smallest proteineaceous α-amylase inhibitors reported.

Sequence conservation of CK α-amylase inhibitors

CK α-amylase inhibitors are highly conserved within species and moderately conserved between species. For example, allatides share high sequence identity with each other, with 63% of amino acid conservation. The consensus sequence for loop 2 to loop 4 (based on the intercysteinyl residues from the N-terminus) throughout allatides O1-O7 is CDGXINQCCDPXXCTPPI-XGXCX. The sequence identity within alstotides or within wrightides is 66% and 53%, respectively. The consensus sequences of loops 2-4 among CK inhibitors from the three plant species is CXXXXQCCXPY(W)XCT(S)PP(V)XXGXCX. The crystal complex of AAI and TMA showed that loop 2 and loop 4 of AAI were involved in an extended hydrogen bond network with TMA [198]. Except for

the conserved cysteines, Pro20, Thr24 (not conserved in aS4), and Gly29 (AAI numbering), CK α-amylase inhibitors found in this study show no other sequence identity with AAI regarding the residues involved in the enzyme-inhibitor interaction network. Thus, Apocynaceae CK α-amylase inhibitors might interact with TMA through a different network of amino acid residues despite the similar size and cysteine pattern with AAI. The question to be addressed is whether loop 2 and loop 4 in these peptides also play a role in the interface surface between AAI and TMA. This will necessitate a structural study of Apocynaceae CK α-amylase inhibitor(s) in its complex with TMA using crystallography technique.

In contrast to the three conserved loops 2-4, loop 1 is the most variable loop with only the glycine residue in the middle of the loop being conserved. Thus, the high sequence variability of loop 1 between Apocynaceae CK inhibitors and AAI suggests that this loop may not be important for binding to the TMA α-amylase. The question remains whether this glycine residue plays any role in the cellular function of the CK peptides. Glycine residue in similar position is the only residue, apart from the six cysteines, conserved among reported ω-conotoxins that block N-type calcium channel [203]. The replacement of this glycine with alanine results in difficulty oxidizing and folding the reduced ω-conotoxins properly [204].

A. oenotheraefolia, A. scholaris, W. religiosa, and *A. hypochondriacus* are distantly related within the core eudicot clade of the Angiosperm. The first three species belong to the Apocynaceae family of the Gentianales order while *A. hypochondriacus* belongs to the Amaranthaceae family of the Caryophyllales order. CK α-amylase inhibitors from Apocynacae plants and AAI from *A. hypochondriacus* very likely have evolved independently from different ancestral genes in these species. With distinct sequences but similar structural fold, Apocynacae α-amylase inhibitors and AAI display the same biological function of inhibiting insect α-amylase. There is clear evidence of convergent evolution, through which proteins with diverse sequences from

different species or different lineages independently evolve to adopt similar structural folds to perform the same functions.

Expression of α-amylase inhibitors in the Apocynaceae family

Eight out of nine species of the Apocynaceae family investigated in this study were found to express the putative CK α-amylase inhibitors, based on their MS profiles with mass range of 3-5 kDa and a mass shift of 348 Da after *S*-alkylation with IAA indicating the presence of 6 Cys residues. Eight plants belong to two subfamilies of the Apocynaceae family including the Apocynoideae (*Alstonia scholaris*, *Nerium oleander*, *Wrightia religiosa*, and *Wrightia antidysenterica*) and the Rauvolfioideae (*Allamanda oenotheraefolia*, *Allamanda cathartica*, *Plumeria rubra*, and *Plumeria obtuse*). Three of such plants have been studied in this project and shown to contain CK α-amylase inhibitor sequences. All the identified 30-residue peptides contain same cysteine pattern and spacing as AAI. Taken together, our study shows that the prevalent CK peptides with MW 3-5 kDa in many Apocynaceae species are likely plant defense peptides and could act as α-amylase inhibitors.

Biosynthesis of CK α-amylase inhibitors

The genetic structure of CK α-amylase inhibitor precursors revealed a three-domain precursor commonly found in many CRPs: an ER signal domain, a pro-domain, and a single inhibitor-encoding domain. This type of precursor structure has been reported for many other CRPs, including ω-conotoxin SO3 [205], hedyotide B2 [201], and cycloviolacin-O8 [206]. A single intron located in the middle of the ER signal domain was also found in CK inhibitor genes, which is a common feature identified in the genes encoding CK types of CRPs. The gene organization starting with a signal peptide provides hints about the biosynthesis pathway of CK α-amylase inhibitors. Their ribosomal synthesis is ER-targeted following the conventional pathway for secretory proteins as suggested for many cysteine-rich peptides [139, 207, 208]. These characters distinguish them from smaller peptides of 5-12 amino acid residues which are

synthesized by multi-enzyme complexes instead of ribosomes [209]. Furthermore, an outstanding feature of the characterized precursors is their prodomains being significantly longer than the functional domains (~40-50 residues over 30 residues). It is interesting to study if there is any regulation of the prodomain over the expression and processing of allatides during cellular synthesis as in the case of transforming growth factor β1 [210].

Within a single species, remarkable identity was found among allatides O1, O4, and O5 clones and between alstotides S1 and S2 clones. It is thus likely that alstotides, and perhaps other CK inhibitors, evolve from recent gene duplication and mutation events within plant species through divergent process. The divergent process is a mechanism adopted by many plants to better adapt to environmental challenges and enhance their survival. Two well-known 43-kDa inhibitors αAI-1 and αAI-2 from the common bean (*Phaseolus vulgaris* L.) were previously reported to resemble each other with 78% sequence identity [211].

Recently, the transgenic plant approach has gained favorable attention over the use of chemical insecticides due to its beneficial safety to the environment [212]. There have been concerns over health risk to consumers when transgenic plants become popular. Current transgenic plants expressing human α-amylase inhibitors α-AI1 and α-AI2 are evaluated to be safe for human consumption. These large proteins are susceptible to heating during cooking, which leads to the loss of their activity in the end products. The heat- and enzyme-resistant property of allatides may have an advantage over the large proteinaceous α-amylases in transgenic plants.

Top-down mapping of disulfide linkage in CK α-amylase inhibitors

The disulfide connectivity of allatides and alstotides was confirmed to be CysI-IV, CysII-V, and CysIII-VI. A previous study on oxidative folding and reductive unfolding of AAI failed to detect any intermediates during reduction of AAI with different concentration of DTT, only the native or fully reduced species found in the reactions, which was proposed as the all-or-none mechanism [213]. The all-or-none model is in contrast to recent findings showing that the unfolding of CK-peptides follow sequential reductive pathways with discrete disulfide intermediates.

In disulfide mapping experiment of allatide O2, we obtained discrete intermediates in the reductive unfolding of allatide O2 with TCEP at 37°C for 20 min. Figure 3.13 and 3.14 summarize the unfolding pathway for allatide O2 by analyzing its disulfide intermediate species. In our experiment, only one 1*SS* species having the intact CysII-V linkage was isolated, suggesting this bond to be the most stable among the three disulfide bonds. Likewise, we only detected one 2*SS* species with breakage of the CysI-IV linkage, suggesting that this *S-S* bridge was the first to be broken. The computer-generated 3D model of allatide O2 reveals that the CysI-IV linkage is the most solvent-exposed bond.

This result is consistent with the reductive unfolding pathway of alstotide S3 (derived from our work in similar manner) and the linear CK peptide of the –CXC- type, hedyotide B2, in which the CysI-IV linkage is the first to be broken among the three disulfide bridges [201]. However, the reduction order of the other two bonds in allatide O2 was different from that of alstotide S3 and hedyotide B2. Generally, allatide O2 is unfolded in the order of CysI-IV, CysIII-VI, and CysII-V whereas alstotide S3 and hedyotide B2 in the order of CysI-IV, CysII-V, and CysIII-VI. Thus, our study provides insights on the sequence-dependent reductive unfolding of CK α-amylase inhibitors.

The importance of high proline content and a cis-proline bond

In a dendrogram calculated for cysteine-rich peptides and proteins that share the *abcabc* disulfide topology of AAI, Carugo *et al.* [70] classified AAI into cluster 5. Cluster 5 comprises 16 structural homologues that have diverse origins and biological roles. Most appear to be inhibitors of other proteins. They include venom of snails and spiders as well as gurmarin, the plant-derived sweet-taste suppressant used to treat diabetes. All members of this cluster have a CK fold with a triple-stranded beta-sheet stabilized by three disulfide bonds. Although members of this cluster display low sequence homology, they share the common features of having high proline content and high frequency of cis prolines that often appear in similar positions as the Pro20 of AAI in the enzyme-complexed form.

Previous studies show that in the AAI-TMA complex, there is a trans-cis isomerization of a peptide bond at Pro20 which exists as a trans Pro bond in the free form. [71, 199]. The trans-cis Pro isomerization upon binding to an enzyme has also been observed in hirustasin, a knottin-like protease inhibitor ([72, 73]. The consequence of this trans-cis isomerization is the more compact AAI structure due to conformational constraint of cis-Pro residue and concurrent changes in two disulfide bonds, which is accompanied by a large number of side-chain reorientations towards the amylase binding sites for optimal enzyme-inhibitor interaction.

Each allatide or alstotide contains four Pro residues which represent >13% of the residues and are substantially higher than about 5% of nonhomologous protein structures in the protein database [214]. Sequence comparison of allatides, alstotides and AAI reveals two of four Pro residues (Pro3 and Pro20, AAI numbering) are conserved in all sequences. The high proline content in Apocynaceae CK inhibitors and the conservation of these two proline residues with AAI suggest that the trans-cis proline isomerization leading to side-chain reorientation to the amylase-binding site is a common mechanism.

Potential applications of stable cystine-knot inhibitors

The CK structure is prevalent in a strikingly vast number of unrelated protein families from plants, animals, and microbes [215]. They are appealing for three features: small size and extraordinarily stable to heat and enzymatic degradation, and tolerance to sequence variation [216-218]. The discovery of α-amylase inhibitors adopting CK fold, therefore, is noteworthy. Firstly, the small size renders CK inhibitors more accessible to chemical synthesis [215]. Secondly, the CK peptides in general are highly tolerant to sequence variations and the spacing of the half cystines [219] to allow α-amylase inhibitors as a scaffold for protein engineering to attain new functions, like the successful example of grafting bradykinin-antagonist peptides DALK or DAK onto kalata B1 scaffold, a cyclic CK peptide [220].

One potential application of considerable interest is to engineer Apocynaceae CK inhibitors to be active towards mammalian α-amylases since the prevalent health problems relate metabolic diseases such as obesity and diabetes. TMA and porcine pancreatic a-amylase (PPA) share about 54% sequence identity [160]. The TMA-AAI interaction interface includes mostly polar interaction with ten direct and seventeen water-mediated hydrogen bonds and one salt bridge [198]. Many other enzyme-inhibitor complexes, including α-amylase inhibitor α-AI1 with PPA and TMA [221], the microbial α-amylase inhibitor tendamistat from *Strepmyces tendae* with PPA [222], and the *Ragi* bifunctional α-amylase/trypsin inhibitor with TMA [47], were reported to depend mainly on hydrophobic interactions, particularly those between aromatic residues in PPA and α-AI1 complex. The lack of extended hydrophobic interactions and the small overall surface interface may account for the lack of AAI inhibition on mammalian α-amylases [198]. Likewise, these could also account for Apocynaceae CK inhibitors being inactive against human α-amylases. In one study, a small peptidic recognition motif with high content of aromatic and positively charged residues was designed *de novo* and showed to be active against TMA, PPA, and several other α-amylases of class 13 [223, 224].

Therefore, it is possible that careful incorporation of aromatic and positively charged residues into CK inhibitors templates may render them active against mammalian enzymes.

There are advantages of choosing natural proteinaceous inhibitors over acarbose, organic acids, or polyphenics for health management applications. First, side effects, such as diarrhea and flatulence, were frequently observed in clinical subjects after ingesting acarbose [225]. With generally more specific activity, proteinaceous inhibitors may not induce such undesirable effects. Second, production of transgenic plants expressing nonproteinaceous inhibitors encounters difficulty due to their complex bioprocessing pathways [212]. Gene-encoded modified CK inhibitors may provide a simple approach in their expression in transgenic plants. Third, if successfully engineered to obtain activities against human α-amylases, CK inhibitors will be much beneficial in diabetes and obesity therapy as these heat- and enzyme-stable inhibitors likely survive the harsh conditions during herb preparation and inside our digestive tract.

Chapter 4

Antiviral Activity: Novel Function of Cystine Knot α-Amylase Inhibitor

Introduction

Challenges in antiviral drug development

Viral diseases have devastating effects, sometimes lethal, on hundreds of millions of people worldwide. Public health all over the world is constantly threatened by a wide range of viral infections, ranging from familiar ailments like common cold and flu to severe illness like HIV/AIDS and hemorrhagic fever. Many newly emergent human pathogens including SARS coronavirus, influenza H5N1, and H1N1 lead to recent outbreaks that exact enormous toll on human lives and huge-economic loss.

Worldwide medical challenges of viral infection have fostered efforts in developing antiviral drugs as an alternative treatment to vaccines and antibodies. Earliest versions of antiviral drugs were reported in the 1950s-1960s at the 2nd Conference on Antiviral Substances. Since then, many more drugs have been developed. Most commercial antiviral therapeutics are nucleoside analogues (for example idoxuridine and adenine arabinoside) or small molecules (for example 2-(α-hydroxybenzyl)benzimidazole, β-thiosemicarbazone, amantadine and rimantadine). However, despite a number of novel antiviral substances reported in the past decades, most drugs have been plagued by many unsatisfactory therapeutic characteristics. They generally have poor oral bioavailability, high cytotoxicity, and off-target effects. Besides, re-emerging and resistant viruses are also a challenge. In this context, antiviral peptides are potentially a promising alternative choice [226].

Many scientists have reported efficient antiviral peptides against a wide range of viruses [227-232]. Synthetic analogs have been developed, sometimes with the aid of bioinformatics, from the naturally occurring antiviral molecules to obtain desired pharmaceutical properties. At least 15 antimicrobial peptides have entered or completed clinical trials [233].

Infectious bronchitis virus (IBV)- structure and replication cycle

Infectious bronchitis virus (IBV) was chosen as the model to study the antiviral activity of alstotides. IBV belongs to the third genera of Coronaviridae family. It is the causative agent of highly infectious respiratory disease in fowl, particularly young chickens. Infected species may experience severe respiratory distress, retarded growth, reduced egg production and eggshell quality. Mortality sometimes goes up to 100%, depending on the secondary infections by other pathogenic agents that follow IBV infection. Current major treatment for IBV relies on vaccines, but the high rate of genetic makeup mutations of IBV greatly complicates the disease control [234, 235]. Thus, IBV is still one of the major causes of economic loss in poultry industry worldwide.

IBV virions are enveloped, spherical, and 100 to 120 nm in diameter. The genome inside the envelope is a single-stranded, positive-sense RNA about 27.5-27.7 kb [236]. IBV particles have four main structural proteins, namely the spike (S) proteins, the membrane (M) proteins, the envelope (E) proteins, and nucleocapsid (N) proteins (Figure 4.1). S and M proteins are glycosylated whereas N protein is phosphorylated. Both the native and modified forms could be observed on SDS-PAGE gels.

- S protein (~130 kDa – averaged from GenBank data) is a club-shaped projection out of the surface, about 20 nm in length. They attached to the envelope in trimers, each post-translationally cleaved into two subunits, N-terminal S1 (92 kDa) and C-terminal S2 (84 kDa). S proteins have two known functions: S1 attaches to host cell receptors

and S2 triggers membrane fusion between IBV virion and the host cell to release the genome into the cell [237, 238].

- M protein (~25.5 kDa) is a triple-spanning membrane protein existing in the most abundance in the IBV virion [239]. The ectodomain constitutes only about 10% of M protein whereas triple-membrane-spanning domain comprises one-third of the molecule, leaving a large C-terminal α-helix endodomain inside the envelope [240, 241]. Interactions between M proteins is the driving force for virion budding and assembly [242]. Their endodomain associated with the virion core functions in the core structure maintenance [239].
- E protein is small (~12.4 kDa) and exists in low amount [238]. They are highly hydrophobic and topologically buried in the membrane. E protein is essential for the assembly of a stable, rounded infectious particle, particularly in interaction with M protein [239, 243].
- N protein (~45 kDa) is rich in basic residues. They associate with RNA genome to form stable ribonucleoprotein complexes in a beads-on-a-string fashion [244]. Binding to the RNA is the primary function of N protein. Besides, N protein also interacts with M protein and facilitates the virus particle assembly [239].

The replication cycle of IBV includes the attachment and entry into the infected cells, the replication of genomic RNAs and transcription of subgenomic RNAs, the translation of their mRNAs, and the assembly and release of progeny virions (Figure 4.2). All steps of the IBV replication cycle are suggested to occur outside the nucleus [238].

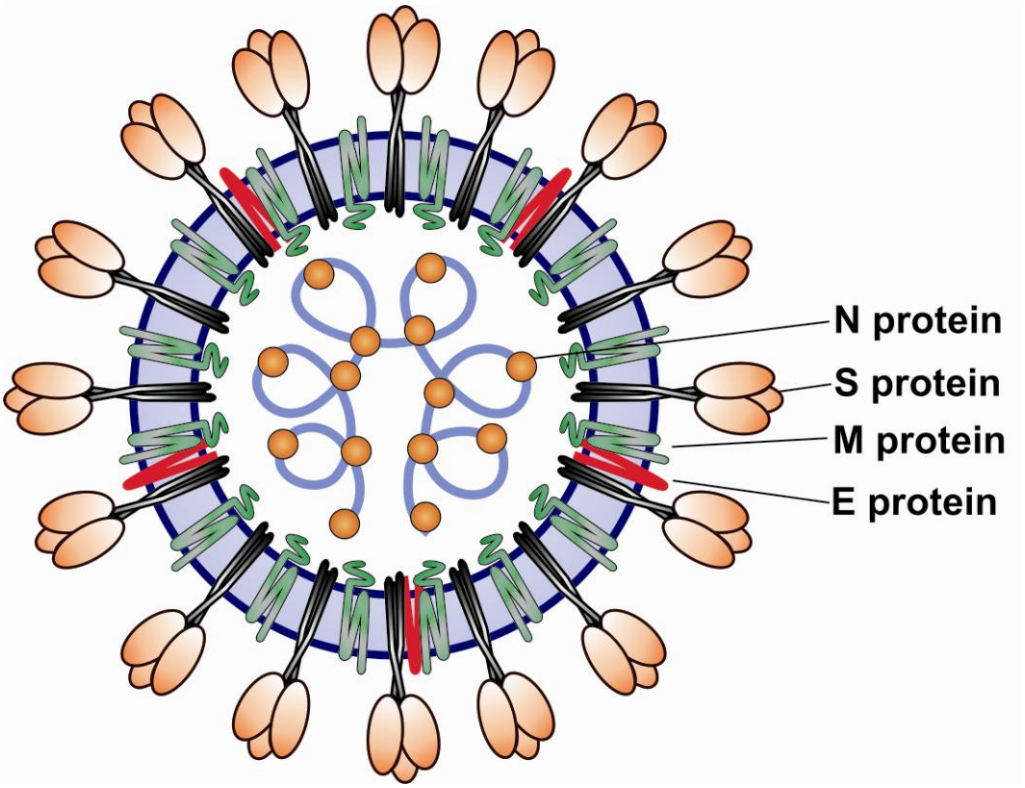


Figure 4.1. Schematic diagram of IBV virion structure. IBV virions are enveloped particles about 100-120 nm in diameter. The nucleocapsid in the virion core includes nucleocapsid (N) protein associated with the 27.6 kb RNA genome [236]. IBV genome is a single-stranded, positive-sense RNA. The envelope is a lipid bilayer containing the spike (S) protein, the membrane (M) protein, and the envelope (E) protein.

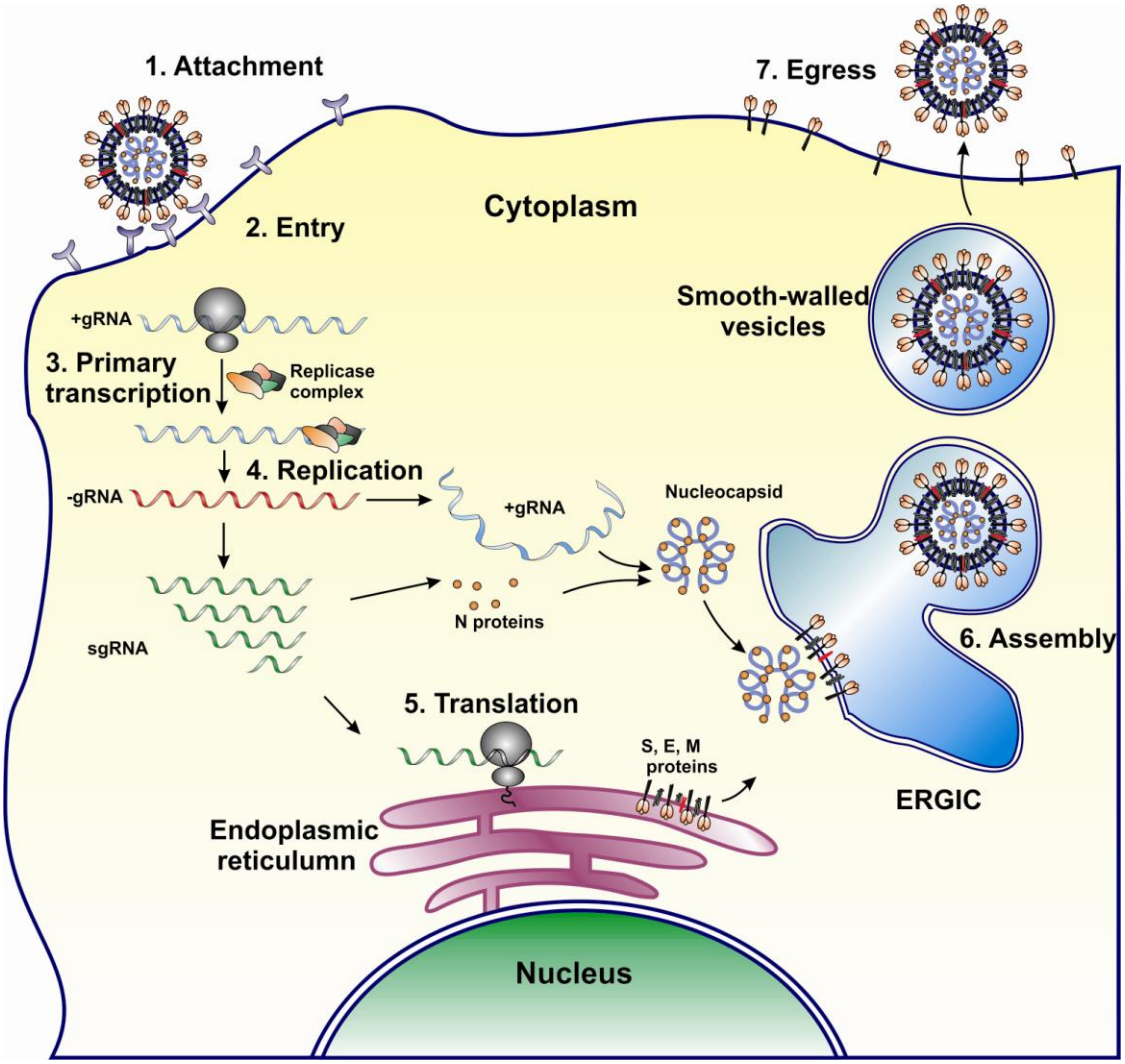


Figure 4.2 The IBV replication cycle. The S proteins bind to host cell receptors, mediate membrane fusion to release genome RNA (+gRNA) into the cytoplasm. The replicase complex is generated using cellular machinery and processed into 15 non-structural proteins. The non-structural proteins assist the synthesis of negative-stranded genomic RNA (-gRNA) and a nested set of subgenomic RNAs (sgRNA) in a discontinuous transcription process. Subgenomic RNAs are translated into structural proteins (S, E, M, N and accessory proteins) at the endoplasmic reticulumn. S, E, and M proteins associated with the double-layer membrane assemble with the nucleocapsid in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Smooth-walled vehicles bud from the ERGIC before their transport to the plasma membrane and egress from the host cell.

Result

Antiviral activities of alstotides

Cytotoxicity assay showed that alstotides S1 and S3 did not significantly affect the Vero cell proliferation at concentrations up to 100 µM. Thus, subsequent assays were conducted with peptide concentrations in the range of 0-100 µM.

To evaluate the ability of alstotides to inhibit IBV infectivity, antiviral assay was performed using virus pre-treatment method. Diluted virus in DMEM (MOI = 1) was incubated with increasing concentrations of alstotides S1 and S3 (25, 50, 75, and 100 µM) for 30 min at room temperature before infecting the cells for 2 h at 37°C (Figure 4.3A and 4.3C). The results showed that alstotides S1 and S3 could inhibit plaque formation in a dose-dependent manner with an estimated EC_{50} of 35 µM and 55 µM, respectively. Particularly, at the concentration of 100 µM alstotide S1 exhibited a 98.1 ± 0.3% inhibition at 24 hpi whereas alstotide S3 88.5 ± 1.3%. The inhibitory effects were further visualized in the time-course assay (Figure 4.3B and 4.3D). In subsequent experiments, alstotide S1 was chosen as the model to study the antiviral mechanism due to its stronger antiviral activity and abundant amount as compared to alstotide S3.

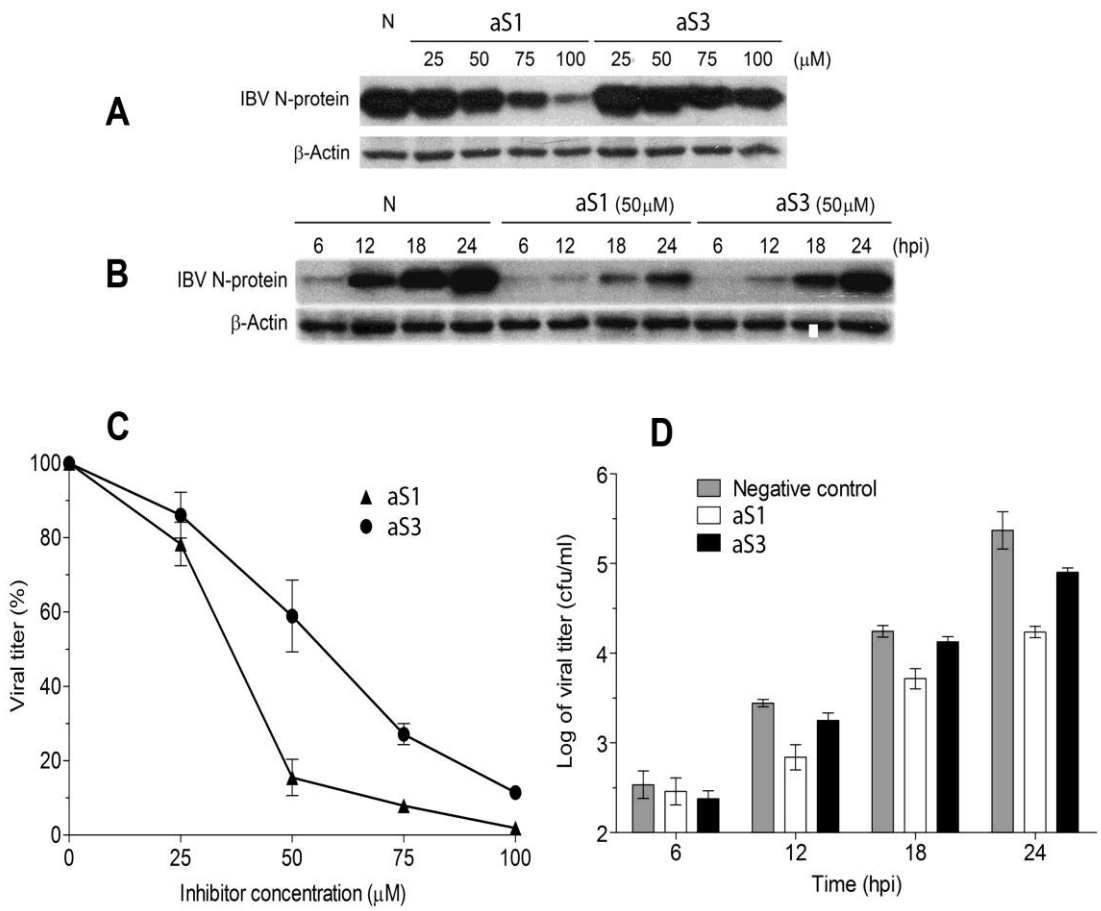


Figure 4.3. Evaluation of antiviral effect of alstotides S1 and S3 against IBV infection. IBV N protein levels in host cells were assessed by Western blot. IBV titers in culture medium were measured by plaque assay. ***A*** *and* ***C****, Dose-dependent inhibition of altotides S1 and S3 were determined at serially diluted concentrations (25, 50, 75, and 100 μM) at 24 hpi. Both alstotides S1 and S3 exhibited antiviral effect against IBV infection of Vero cells with estimated* IC_{50} *of 35 and 55 μM, respectively;* ***B*** *and* ***D****, Time-course experiments were conducted at the concentration of 50 μM to demonstrate the activity of alstotides S1 and S3. The experiments were repeated at least twice. All the measurements were done in triplicate.*

Time-of-drug-addition assay

To determine the stage(s) in viral proliferation cycles upon which alstotides act, we performed a time-of-drug-addition assay. Alstotide S1 at 50 and 100 μM was introduced to synchronized infected cells at different time points during the infection.

Figure 4.4 showed the expression levels of IBV proteins for the aS1-treated samples and negative control. Incubation at 4°C facilitates viral attachment to but not entry into the cells which occurs after the sample is transferred to 37°C. Late addition of peptide at 1 hpi usually allows for only the action of inhibitors targeting replication of the nucleic acid materials and protein translation. In this assay, the samples added early at the beginning of attachment or entry processes exhibited comparable positive effects (Lane A and B). Meanwhile, alstotide S1 significantly lost its activity when added during later stages from 3 hpi onwards (Lane C, D, and E). The inhibitory effects of alstotide S1 to different extents in these samples were visualized by microscopic images in Figure 4.5. The cytopathic effects shown by the formation of multinucleated cells (syncytium formation) in samples A-E agreed with the Western blotting result in Figure 4.4. Comparable amounts of the endogenous control actin in all the samples excluded the effect of alstotide S1 on general cellular protein synthesis. Taken together, our results provided evidence of the entry step as the main mechanism of alstotide S1 against IBV.

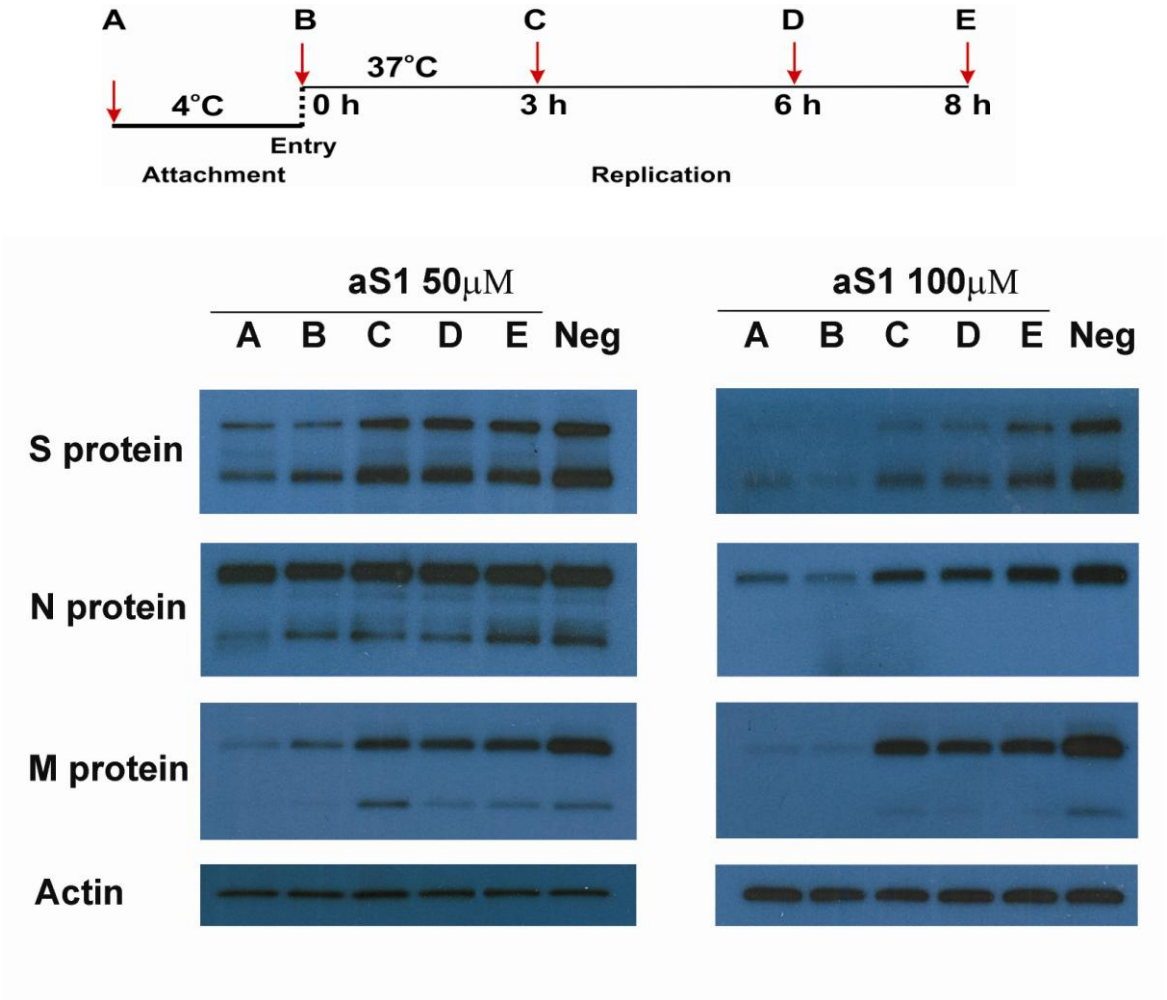


Figure 4.4. Time-of-drug-addition study. Parallel cultures of Vero cells with synchronized IBV infections were treated with alstotide S1 (50 and 100 μM) at different time points: before and after 2-hour incubation at 4℃ for samples A and B, respectively, at 3, 6, or 8 hpi for samples C, D, and E, respectively. Alstotide S1 was not added to the negative (Neg) sample. The cell monolayers were collected 24 hpi.

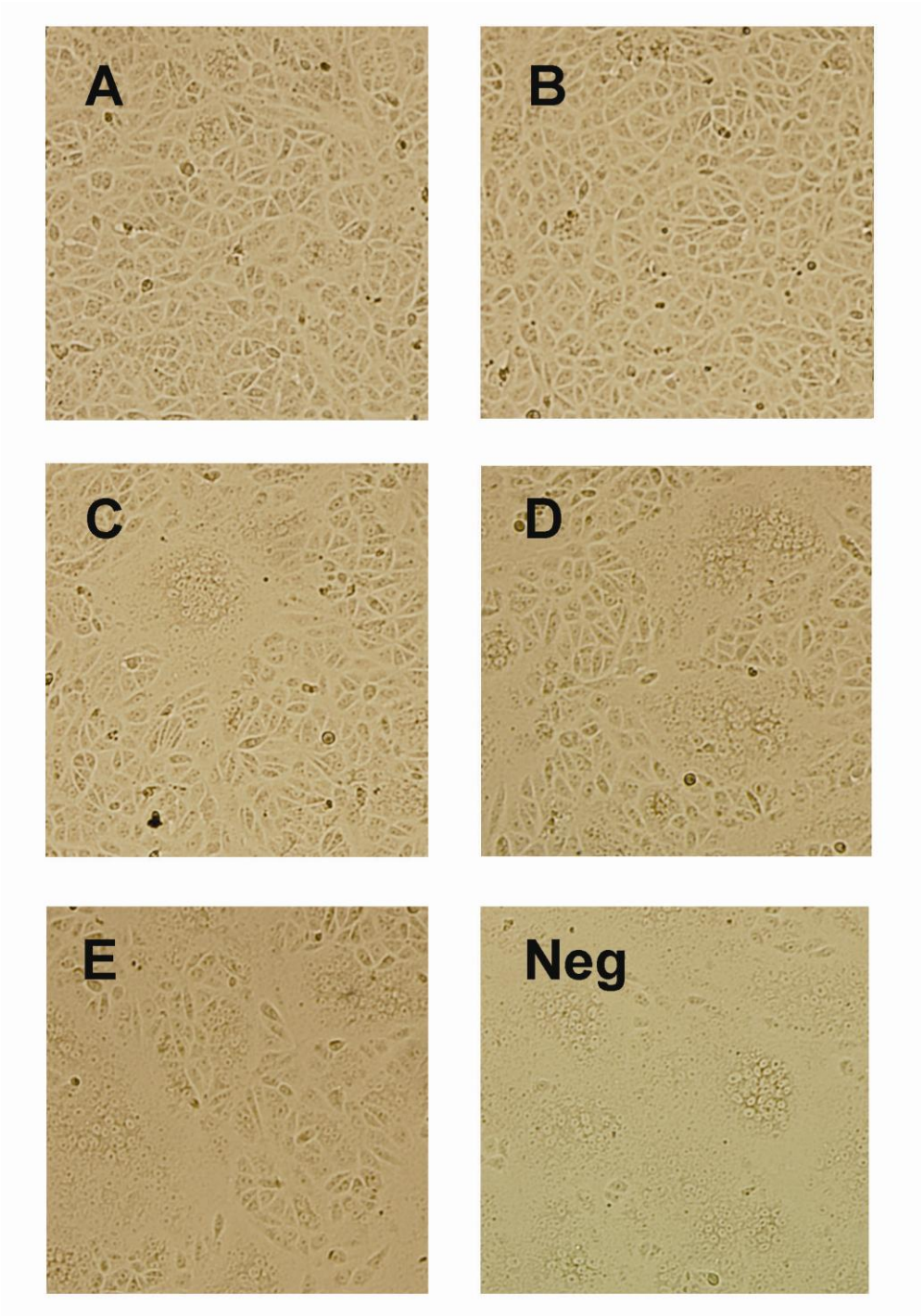


Figure 4.5. Microscopic images of time-of-drug-addition study. Images A to E correspond to the samples A to E in Figure 4.4. Alstotide S1 was added to the infected cell population at the beginning of 4℃ incubation (A), after transfer to 37℃ (B) or 3, 5, and 8 h later on (C, D, and E, respectively). The images were taken at 24 hpi.

Alstotide S1 targets the early stage of viral replication cycle

To unequivocally eliminate the possibility of alstotide S1 affecting viral RNA replication and virion assembly, we performed a transfection experiment. In this assay, the attachment, fusion and entry steps were excluded (Figure 4.6). The cells were transfected with purified viral RNA by electroporation followed by treatment with alstotide S1 (100 µM). The replication of IBV during primary and secondary infections was investigated after 18 and 40 h post-transfection, respectively (Figure 4.7). Parallel aS1-non-treated samples were included as controls.

The results (Figure 4.7) showed that alstotide S1 has minimal effect on viral RNA replication and assembly. During the primary infection (18 h), quantitative RT-PCR for positive-stranded genomic RNA (panel B) did not show any significant difference between aS1-treated sample and the control. This means that alstotide S1 exhibited no significant inhibitory effect when the entry step was removed. During secondary infection (40 h), alstotide S1 impeded the IBV replication as the virions entered new host cells. A clear inhibition of N protein (panel A) and (+) genomic RNA synthesis (panel B) was observed 40 h post transfection in both Western blotting and quantitative RT-PCR.

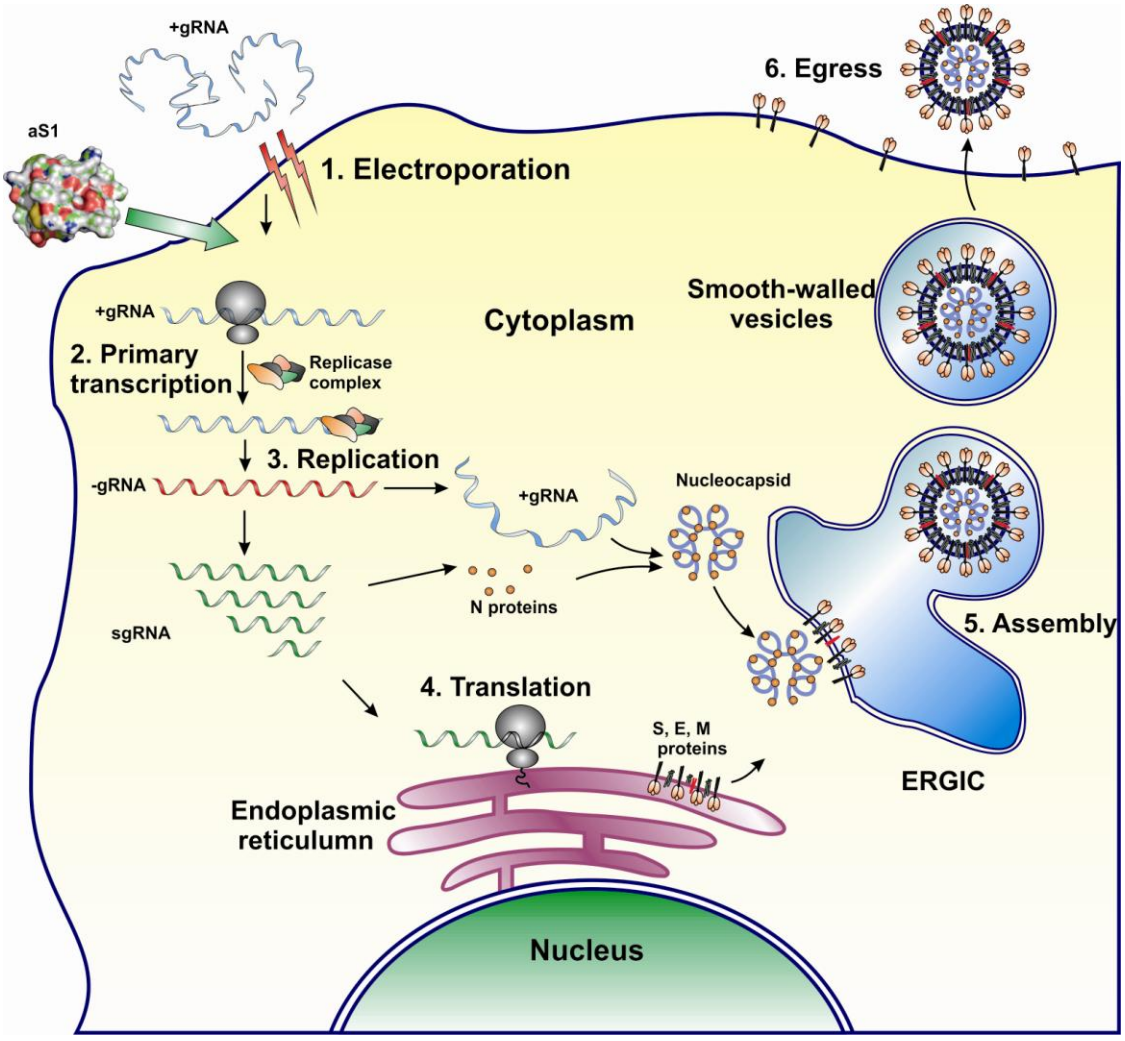


Figure 4.6. Viral replication cycle without entry step. The purified viral RNA is electroporated into Vero cells, followed by treatment with alstotide S1 (100 μM). The replication cycle of IBV occurring intracellularly includes primary transcription of replicase complex, the generation of (-) genomic RNA and subgenomic RNAs, the translation of structural protein and the assembly of progeny virions before egress from the host cell.

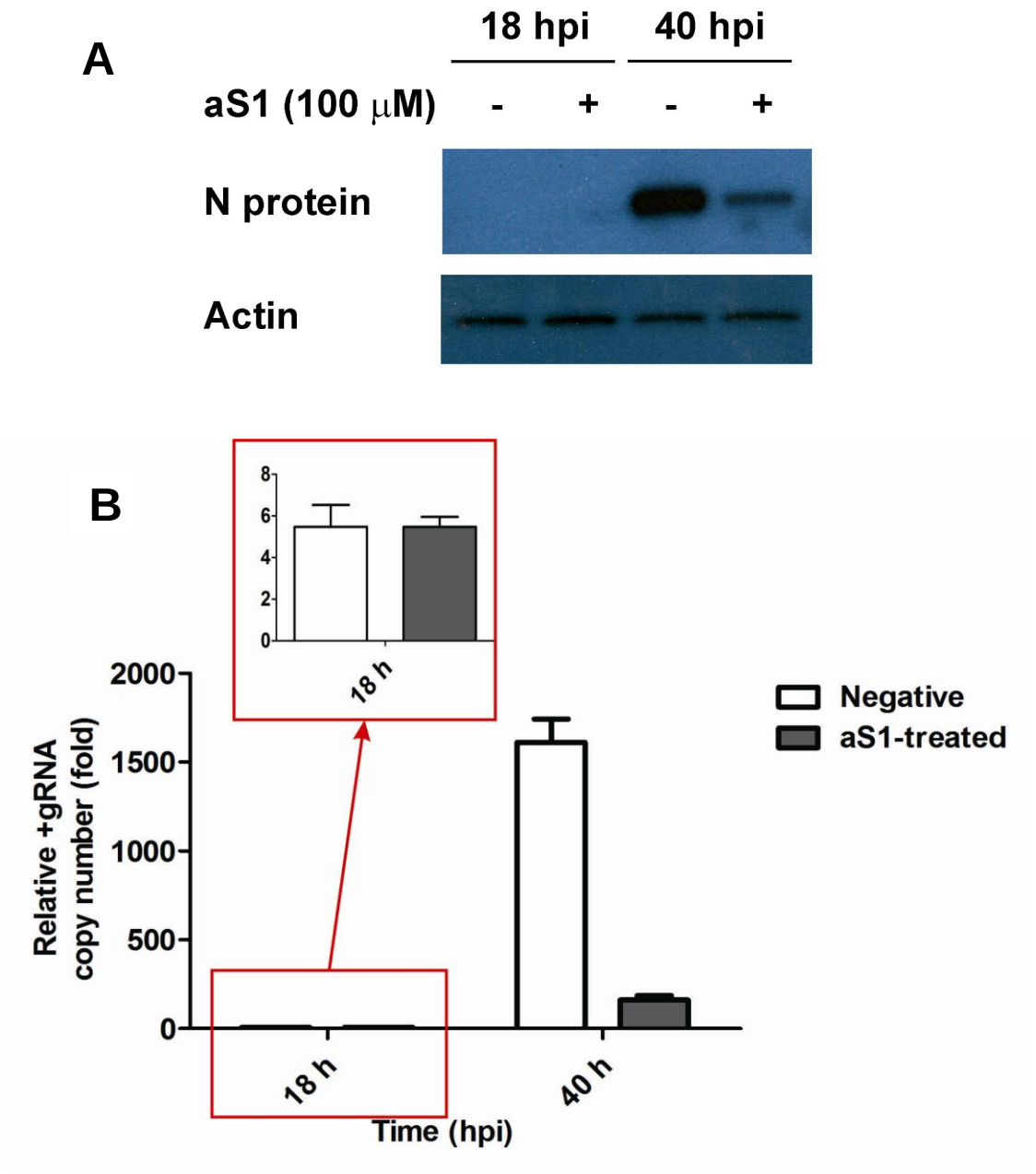


Figure 4.7. Quantitative results of virus proliferation after IBV RNA electroporation and peptide treatment. ***A****, Western blotting of N proteins visualizes minimal antiviral effect of alstotide S1 during primary infection (18 h) and strong effect during secondary infection (40 h). The amount of N proteins in the treated and control samples 18 h post transfection was low and undetectable in the experiment condition;* ***B****, Realtime PCR quantification of positive-stranded genomic RNA after 18 and 40 h post transfection was normalized against that of the 3-h-post-transfection sample.*

Biotinylation of alstotide S1

Labeling the peptide to facilitate its tracking and detection in studying molecular mechanism is essential. With the advantage of having no lysine residue, alstotide S1 has only one biotinylation site using NHS-LC-biotin (Thermo Scientific). Biotin was ligated covalently to the primary amino group at the N-terminus of the alstotide S1 molecule via a 6-aminocaproic acid (LC) spacer. As the LC spacer only functions to reduce steric hindrance of biotin binding, we use the term biotinyl-aS1 to refer to the covalently bonded aS1-LC-biotin chain in this thesis. The reaction was quenched after 30 min at RT following the manufacturer's instruction with the major product having a single biotin group (Figure 4.8). The MS spectra showed a mass increase of 339 Da from *m/z* 3366 to 3705, corresponding to the mass of one LC-biotin group. The HPLC profile also showed an increase in hydrophilicity of alstotide S1 molecule upon the addition of the biotin group. The biotinyl-aS1 was eluted at about 26% acetonitrile while the native alstotide S1 at about 37% acetonitrile in the analytical RP-HPLC purification.

Extended incubation at low temperature (24 h, 4°C) led to the production of double-biotinylated products as examined in mass spectrometry. A peak of *m/z* 4045 was observed in the MS spectrum, indicating the addition of two LC-biotin molecules. The secondary biotinylation site, to a much lesser extent than the primary site, is likely the Thr20. Its hydroxyl group may act as a weak nucleophile, attacking the carboxyl group in NHS-LC-biotin molecule. The reaction covalently ligated the second biotin to the singly biotinylated alstotide S1. To minimize the effect of biotin groups to the intrinsic activity of alstotide S1, only singly biotinylated alstotide S1 was used in subsequent experiments.

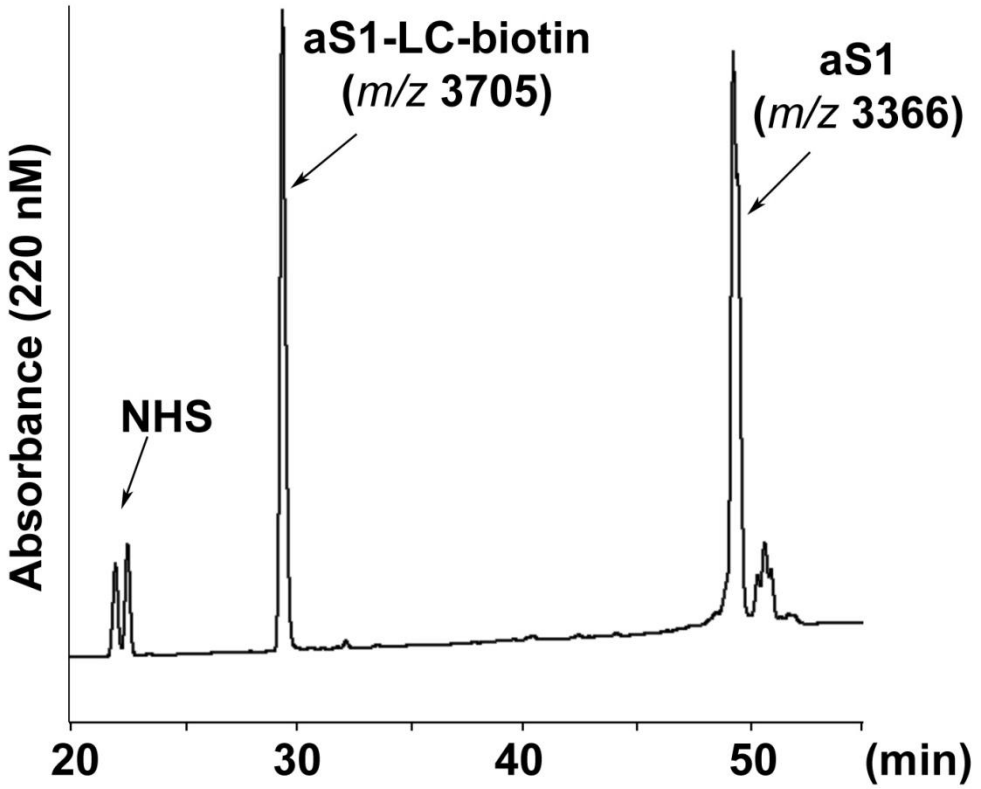


Figure 4.8. Biotinylation of alstotide S1 using NHS-LC-Biotin. Figure 4.8 shows the HPLC purification profile of alstotide S1 biotinylation. The biotinylation reaction was performed using NHS ester reaction chemistry between NHS ester-activated crosslinker and the target protein in physiologic condition. The reaction was quenched after 30 min at RT. The addition of one biotin group rendered the alstotide S1 molecule more hydrophilic and eluted earlier in RP-HPLC.

Interaction of IBV proteins with alstotide S1

To gain insights into the interaction between alstotide S1 and viral proteins, we performed a pull-down assay using Neutravidin beads. Neutravidin recognizes the biotin group in the biotinyl-aS1 molecules and immobilizes the molecules onto agarose beads. This complexed particle was used to pull-down viral proteins that interact with the peptides. As shown in Figure 4.9A, biotinyl-aS1 was associated *in vitro* with IBV spike (S) and membrane (M) proteins but not with nucleocapsid (N) proteins.

The interaction of alstotide S1 and M protein was also proved to be an *in vivo* association in a co-immunoprecipitation experiment (Figure 4.9C). Biotinyl-aS1 was allowed to form complexes with M protein in IBV-infected Vero cell system. The alstotide S1-M protein complex from the cell lysate was pull-downed by Neutravidin beads and M protein subsequently analyzed in Western blotting. As expected, a clear band was observed in the last lane of Figure 4.9C, confirming the intracellular interaction between alstotide S1 and viral M protein.

Antiviral activity of alstotide S1 against IBV mutant

Alstotide S1 was shown previously in this study to be an α-amylase inhibitor. It likely exerts substrate mimicry as does AAI. M protein is glycosylated at two Asn residues. It is thus possible that alstotide S1 has carbohydrate-like properties and binds to oligosaccharide moieties of S and M proteins.

To verify this hypothesis, we tested the antiviral effect of alstotide S1 against an IBV mutant, F5L5, having non-glycosylated M protein. The N-linked glycosylation site (MPNETNC) at the N-terminal ectodomain of wild-type IBV M protein was changed to MPSKGEE in F5L5 virus. There is only one major band at ~22 kDa for the non-glycosylated M protein of F5L5 virions in Western blot, as opposed to the major band at ~28 kDa for wild-type IBV M protein (Figure 4.9A and 4.9B).

In the antiviral assay, Vero cells were incubated with F5L5 virus (MOI 1) and alstotide S1 (100 μM) in a co-treatment manner. Western blotting result showed a complete inhibition of alstotide S1 (100μM) on F5L5 mutant replication (Figure 4.9B). This result indicated that alstotide S1's antiviral mechanism is independent from the mutated ectodomain, which was confirmed by the pull-down assay in Figure 4.9A. In the pull-down experiment, stronger bands of S and M proteins were observed when IBV- or F5L5-infected cell lysates were incubated with alstotide S1 immobilized on Neutravidin beads. This result means that alstotide S1 interacted with both the glycosylated and non-glycosylated M proteins from wildtype IBV and mutant F5L5, respectively (Figure 4.9A). Interaction between alstotide S1 and M protein was further supported in the co-immunoprecipitation experiment (Figure 4.9C). Biotinylated – alstotide S1 (75 μM) were allowed to bind to M protein in the IBV-infected Vero cells during 24-hour incubation at 37°C. The complex was later pull-downed by Neutravidin beads and quantified in Western blotting. The band detected for M protein in Western blot showed that alstotide S1 interacted with M protein *in vivo*.

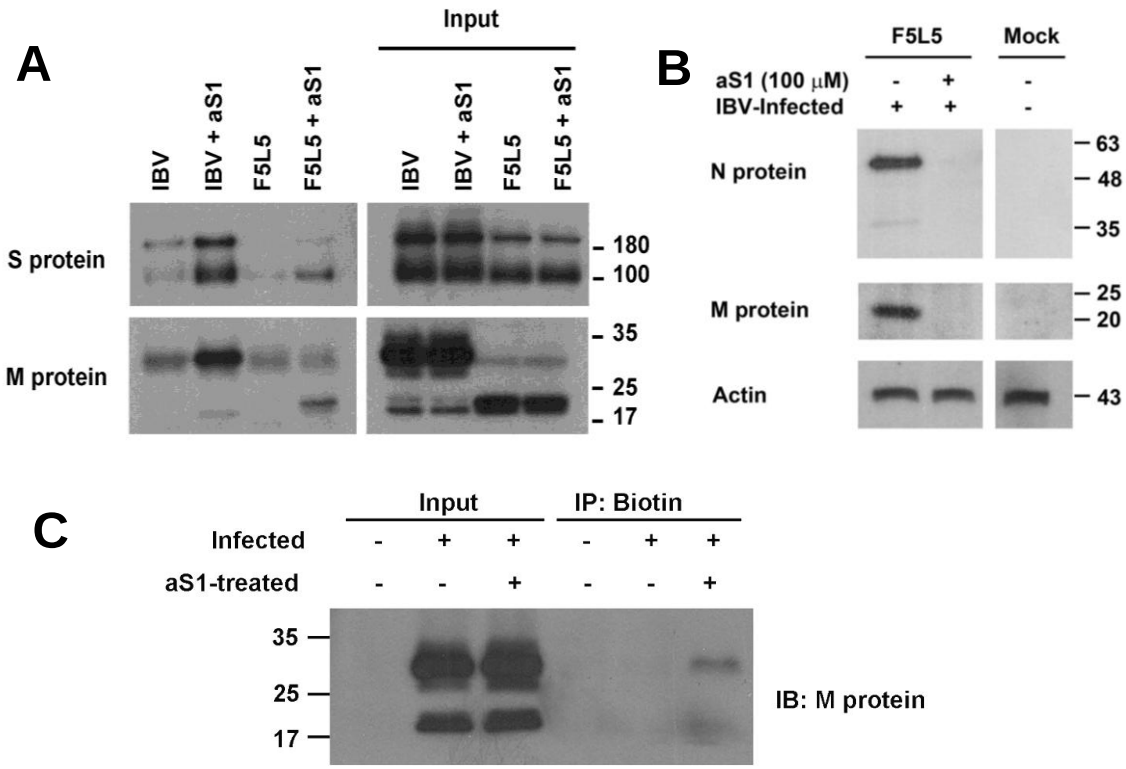


Figure 4.9. Interaction between alstotide S1 and IBV proteins. ***A****, In vitro interaction between alstotide S1 and S or M protein was identified by pull-down assay. Cell lysates from IBV-infected cells were incubated with biotinyl-aS1 immobilized on Neutravidin beads. Viral proteins associated with biotinyl-aS1 were analyzed by blotting with corresponding antibodies as indicated. Alstotide S1 interacted with both glycosylated (~28 kDa) and non-glycosylated (~22 kDa) IBV M protein. Non-infected cell lysates were included as control. A bead with biotinyl-aS1 on its surface is called bait. Input is 1/100 of total cell lysate applied onto the beads or baits. Protein markers are shown in kDa;* ***B****, F5L5 virus is an IBV mutant with non-glycosylated M protein. Vero cells were infected with IBV and treated with alstotide S1. The cell monolayer was collected at 80-90% CPE and subjected to Western blotting. Protein markers are shown in kDa;* ***C****, In vivo co-immunoprecipitation of alstotide S1 and M protein was performed in Vero cells. IBV-infected cells were treated with alstotide S1 (75 μM) and incubated for 24 h at 37℃. Cell lysates were precipitated with Neutravidin beads, which interact with biotin, (IP:biotin) and blotted with anti-M protein antibody (IB:M protein). Protein markers are shown in kDa.*

Cellular uptake of alstotide S1 into Vero cell system

One of the major disadvantages of current therapy using large molecules is the low bioavailability. It is always of high concern to determine whether a new drug can penetrate the cell membrane. On the other hand, the time-of-drug-addition assay showed the prominent effect of alstotide S1 when added before the entry stage of the virus. To act early in the viral replication cycle, a common mechanism for the inhibitors is to bind to either virus or cell membrane, blocking the virion-cell receptor interaction or inhibiting the membrane fusion. Thus, it is important to study if alstotide S1 is retained extracellularly via interaction with host receptors or penetrates into the cells to exert its effect. In the latter case, it is of our interest to determine the intracellular distribution of alstotide S1.

Through immunofluorescence staining on a Zeiss LSM 510 confocal microscope, we performed the co-localization of biotinyl-aS1 and IBV N protein in Vero cell system. Vero cells were mock-infected or infected with IBV (MOI = 1), treated or non-treated with biotinyl-aS1 and incubated for 16 h. The cells were fixed, permeabilized, stained with antibodies and visualized under the confocal microscope. Figure 4.10 showed that alstotide S1 accumulated intracellularly in the perinuclear region in Vero cells. The perinuclear distribution of alstotide S1 agrees with the cellular location of M protein reported previously [245]. Z-stack analysis further confirmed that alstotide S1 is able to traverse the cell membrane and accumulated intracellularly.

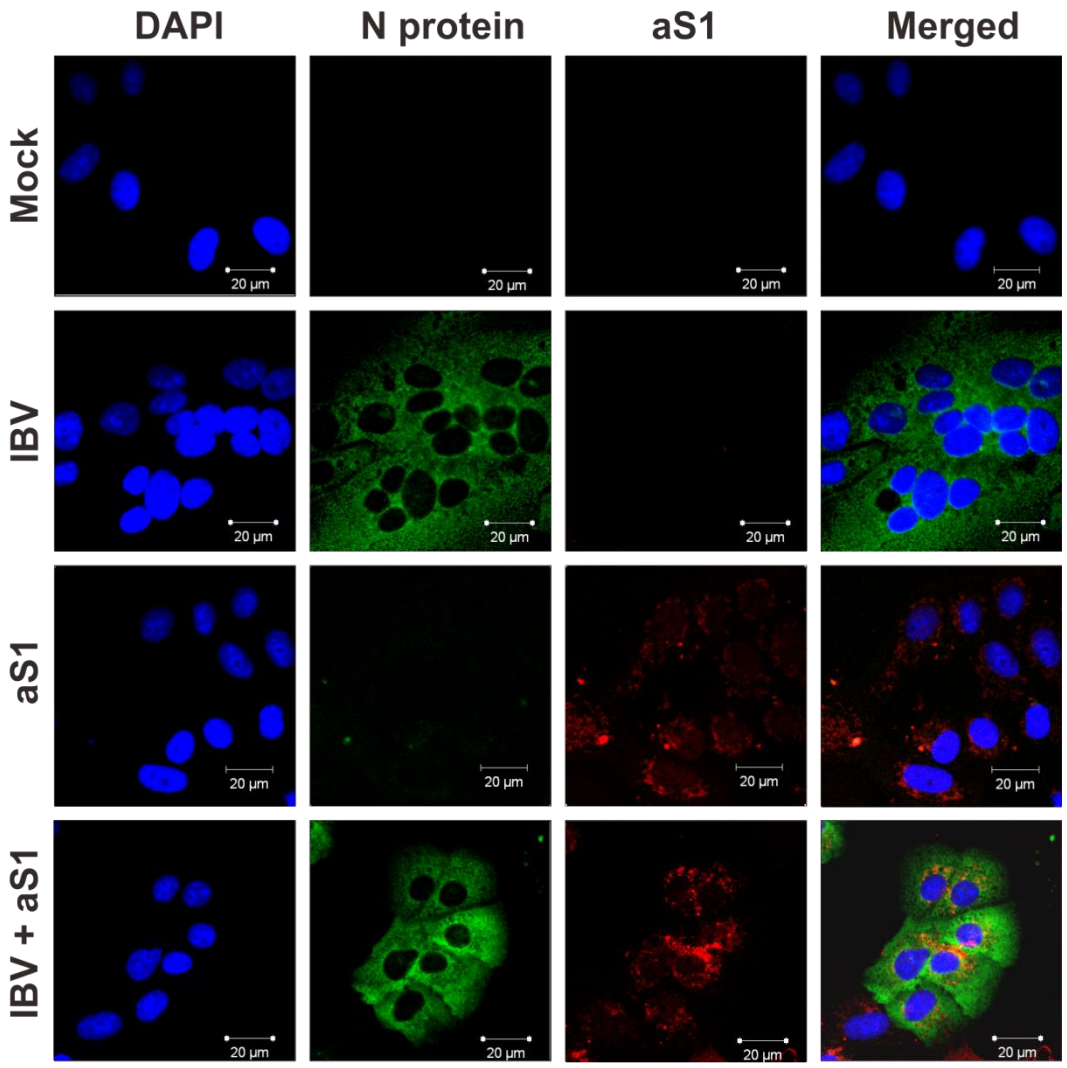


Figure 4.10. Colocalization of alstotide S1 and IBV N protein by indirect immunofluorescence. Vero cells at 60% confluency were treated with biotinyl-alstotide S1 (10 µM) at the point of infection by IBV. The cells were stained with rabbit anti-IBV N protein sera, followed by FITC-labeled anti-rabbit antibody (green) and then Neutravidine conjugated to Dylight 594 (red). The nuclear DNA was visualized with DAPI in the mounting medium (blue). Parallel samples without alstotide S1 treatment (labeled IBV) or IBV infection (labeled aS1) or both (mock) were prepared as controls. Magnification: x63.

Selective antiviral activity of alstotide S1

This study has so far identified the antiviral activity of alstotide S1 over the replication of IBV, a contagious virus on domestic fowl. With global constant need to find new effective drugs targeting human viral diseases, alstotide S1 was tested against two human viruses: dengue virus type 2 (DENV2) and respiratory syncytial virus type A (RSV A). The effect of alstotide S1 on these two viruses was evaluated in a co-treatment antiviral assay. The peptide at different final concentrations (25, 50, 75, and 100 µM) was added to the cell culture at the point of viral infection and the cell lysate harvested after 2 (RSV A) to 3 days (DENV2).

The assessment of DENV2 NS3 protein expression in the infected cultures by Western blotting revealed a moderate inhibitory effect of alstotide S1 against this virus (Figure 4.11A). The effect was concentration-dependent, with EC_{50} about 90 µM based on semi-quantitative analysis by densitometry. In contrast, alstotide S1 did not inhibit the propagation of RSV A at up to 100 µM (Figure 4.11B), as observed from comparable HA and NA proteins in all infected cultures. The differential effects of alstotide S1 on IBV but not RSV, though tested on the same Vero cell system, indicated that the antiviral activity of alstotide S1 is virus-specific or infected cell-specific and not through interference to the healthy host cell metabolism.

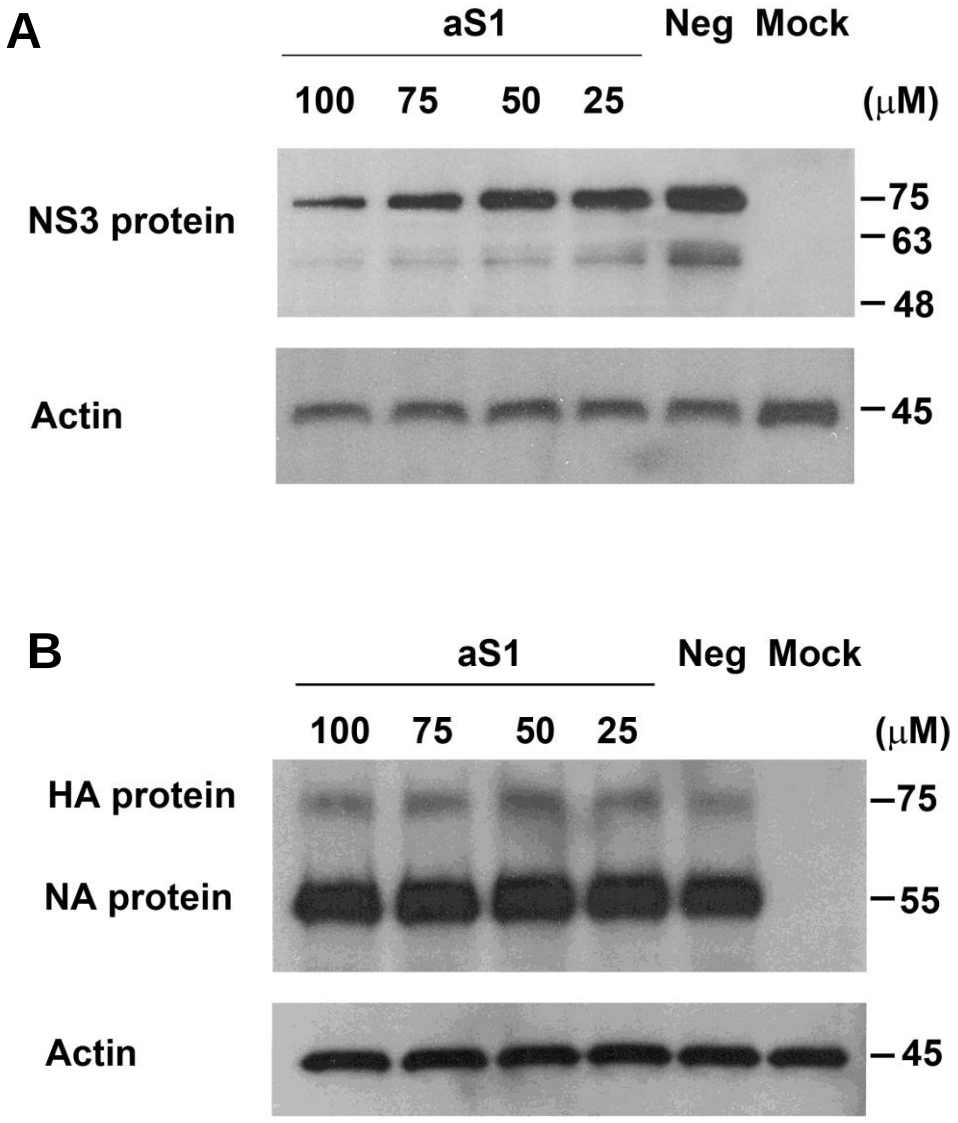


*Figure 4.11. Evaluation of antiviral effect of alstotide S1 against Dengue type 2 and RSV type A infection. The antiviral assay was performed in a co-treatment manner. Vero cells seeded in 48-well plates were infected with the virus and treated with serially diluted concentration of alstotide S1 at the same time. We tested Dengue type 2 virus (**A**) and RSV type A (**B**). The cultures were incubated for 2 (RSV A) to 3 (DENV2) days until >50% CPE appeared. The cell monolayers were collected and analyzed with corresponding antibodies. Protein markers are shown in kDa.*

Discussion

Alstotides are novel plant-derived antiviral drugs

Medicinal plants have been considered a rich, promising source of alternative antiviral chemotherapeutics [246-248]. Many groups have reported on a variety of plant extracts that inhibit different viruses. However, in most cases only the activities of total plant extracts are documented. Limited works reported on the active compounds responsible for the antiviral effects [248] and even fewer for the mechanism study. This is due to the lack of appropriate isolation and characterization instrumentation in many molecular biology laboratories, particularly those in tropical developing countries where medicinal plants are abundantly available. With the well-established routine purification platform in our laboratory, we have succeeded in characterizing the purified bioactive components from *A. scholaris* and studying their antiviral activity against IBV and DENV.

The discovery of alstotides' inhibition against IBV in this study marks the first report of drugs specifically targeting infectious bronchitis disease. To date, there has been no published work on an anti-IBV drug for the treatment of infectious bronchitis among domestic fowl. When the disease occurs, the morbidity and mortality largely depend on the secondary infections (for example Mycoplasma, *E. coli*, and Newcastle disease). The current therapy for infectious bronchitis is the administration of small-molecule drugs, for example sodium salicylate, to control secondary infections, not the acute bronchitis disease itself. Prior vaccination for 2-week-old chickens is an expensive and labor-intensive yet moderately effective prophylaxis. New strains of IBV emerging through genetic point mutation and recombination have been occasionally identified from vaccinated chicken flocks [249, 250]. The challenges in the control of infectious bronchitis disease make it appealing to study alstotide S1 as an alternative drug which can be applied during the endemics.

Despite tremendous efforts of scientists to seek an effective therapy in many past decades, dengue has been one of the major threats to global public health with reemerging epidemics in over 60 countries worldwide during 2000-2007 [251]. At least 2.5 billion people living in the tropical and contiguous temperate areas with dengue-affected history are at risk to develop dengue fever and its severe complications, including dengue hemorrhagic fever and dengue shock syndrome [251]. Therefore, the discovery of any new drug gives chances to rescue a remarkable population from the threat of dengue, which follows by the save of huge economic loss.

Antiviral mechanism of alstotide S1

Early acting drug

At a noncytotoxic concentration of 100 μM, purified alstotide S1 blocked 98.1% of IBV proliferation in Vero cells. EC_{50} of alstotide S1 against IBV is 35 μM. The inhibition effect of alstotide S1 was exerted mainly in the early viral replication cycle prior to the early replicase synthesis. The claim is supported by two lines of evidence. First, time-of-drug-addition assay revealed a significant drop of activity when alstotide S1 was added 3 hpi onwards, which is after viral entry into the cells and during early viral synthesis and replication. Second, when the viral RNA was electroporated into the cells, skipping the attachment and entry steps, alstotide S1 did not show any effect upon viral gene replication during primary infection (18 h post-transfection) yet exhibited strong anti-IBV activity during secondary infection (40 h post transfection). Secondary infection entails normal viral attachment and entry into the cells and therefore allows for the anti-IBV effect by alstotide S1. Together, the results showed that alstotide S1 is an early acting antiviral agent.

The range of active concentrations, the action during the early phases and intracellular location of alstotide S1 are reminiscent of the features previously described for the EB (entry blocker) peptide. EB is a 20-residue broad-spectrum antiviral peptide derived from human fibroblast growth factor 4

protein [252]. It inhibits unrelated viruses, including herpes simplex virus type 1 [253], avian influenza [254], vaccinia [255], cowpox and monkey pox [256], in a concentration- and MOI-dependent manner. The EC_{50} of EB varied from 15 μM to >150 μM against vaccinia when MOI changed from 0.05 to 50 [255]. EB may target either attachment or entry step in different viral infections [254, 256] although it was found located in the nucleus. Similarly, alstotide S1 may act in an MOI-dependent manner, which partially accounts for its moderately high EC_{50} at MOI=1 in this study. Alstotide S1 also demonstrated its effect in early viral replication cycle, which takes place at the lipid bilayer membrane, while the immunofluorescence experiments revealed its location in the perinuclear region. Whether the intracellular location of alstotide S1 plays any role in its antiviral properties will be discussed later in the context of its interaction with virion proteins. As for its activity at the surface membrane, alstotide S1 likely acts during the fusion/entry step of viral invasion rather than during the attachment. This hypothesis stems from the observation that alstotide S1 added at the beginning of either attachment (4°C) or entry (37°C) showed comparable inhibition on IBV protein synthesis (Figure 4.4). However, the solid elimination of attachment as the major step for the action of alstotide S1 requires more work to be done.

Entry inhibitor via S protein interaction

IBV, RSV, and DENV are members of three enveloped virus families: Coronaviridae, Paramyxoviridae, and Flaviviridae, respectively. In enveloped virus families, virus-mediated membrane fusion during entry is an essential step to trigger the infection and is increasingly viewed as an attractive target for antiviral intervention. One or more viral surface glycoproteins, called fusion proteins, are responsible for the virus-cell fusion. They are spike (S) protein for IBV, fusion (F) protein for RSV, and envelope (E) protein for DENV [257, 258].

The discovery of alstotide S1 acting during early stages of IBV infection led us to the hypothesis that alstotide S1 interferes with viral entry through its binding to fusion glycoprotein in a “carbohydrate-love” manner, given its carbohydrate

substrate mimicry mechanism as an α-amylase inhibitor. The pull-down assay showing that alstotide S1 binds to IBV glycosylated S and M proteins favors this hypothesis. However, alstotide S1 also interacts with the non-glycosylated M protein, implicating that alstotide S1 does not bind to the carbohydrate moiety of M protein. Whether the interacting surface between alstotide S1 and S protein lies on the glycosylation site still remains unknown.

Antiviral tests on DENV and RSV suggest the mechanism via endocytic interference of alstotide S1. For the three viruses included in this study, their fusion proteins belong to two structural classes: class I (mainly α-helix, perpendicular to virion surface in IBV and RSV) and class II (mainly β-sheet, parallel to virion surface in DENV). The fact that alstotide S1 is active against IBV and DENV, but not RSV, eliminates any correlation between the activity of alstotide S1 and general secondary structure of the affected fusion proteins. Interestingly, the entry pathway of these viruses could provide a common mechanism for the action of alstotide S1. While the entry of RSV and DENV has been well-established to be triggered by host-receptor binding and endosomal low pH environment, respectively, the mechanism for IBV has been suggested by conflicting evidence. For many years, large syncytia formed by cell-cell fusion at neutral pH [259] was used to generalize for IBV entry as a non-endocytic direct pH-independent pathway. However, it was demonstrated that IBV fusion is pH-dependent [260] and IBV enters cells via clathrin-dependent endocytosis [261]. Thus, it is possible that alstotide S1 acts through interaction with fusion protein during the common endocytic route of IBV and DENV entry.

Alstotide S1 and M protein interaction - Single or mixed mechanism?

In our attempt to determine the viral factors involved in the activity of alstotide S1, we identified a specific interaction between alstotide S1 and IBV M protein. The interaction was found to occur both *in vitro* and *in vivo.* M protein is comprised of triple membrane-spanning domain flanked by a small N-terminal glycosylated ectodomain and a large C-terminal endodomain. A preliminary

effort to map the site of interaction between the two molecules excludes the possibility of alstotide S1 binding to the glycosylated ectodomain of M protein.

The finding of a drug targeting the early entry step and interacting with both S and M proteins is rare. Most of the entry inhibitors with their mechanism revealed so far bind to the fusion proteins at conserved sequence. Examples include C34 [262] and the approved drug T20 for HIV-1 treatment [263-265], a number of T-20 – like peptides designed on the conserved heptad repeat region for the inhibition of RSV [266], human parainfluenza virus type 3 (HPIV-3) [266], measles virus (MV) [266, 267], Newcastle disease virus (NDV) [268], Hendra virus (HeV) [269], and Sendai virus [270].

M protein in coronavirus virions is long believed to play major role in the assembly, budding, and maturation processes [271], which are late stages in viral replication cycle. The fact that alstotide S1 interacts with M protein and acts mainly in the early stages of the infection could be a coincidence. Or it could also be possible that M protein is involved in the early fusion/uncoating step in an unknown manner through which alstotide S1 elicits its antiviral effect. A recent study has shown that in flavivirus, M protein may modulate the normal pH-dependent function of E fusion protein during fusion reaction [272]. In vesicular stomatitis virus (VSV), M protein binds to host dynamin and this interaction affects clathrin-dependent endocytosis [203]. If M protein plays role in endocytosis, it would be interesting to utilize M protein to study the entry mechanism of IBV infection, which still remains unclear.

Alternatively, alstotide S1 may adopt an M protein-independent mechanism to inhibit IBV entry into the host cells. The alstotide S1-M protein association may otherwise produce inhibition on IBV assembly, budding or maturation processes, which involve M protein. A slight positive effect observed in the time-of-drug-addition assay when alstotide S1 was added at late stages could be accounted by this hypothesis. The localization of alstotide S1 in the perinuclear region by immunofluorescence resembles the distribution of M protein in the ER-Golgi intermediate compartment (ERGIC) and Golgi area

[245]. Thus, alstotide S1 may affect the normal function of its binding partner, M protein, intracellularly and thus inhibit the production of infectious virions.

Applications of alstotide S1 as an antiviral drug or cell-penetrating carrier

In recent years, the limited success of vaccines and antibodies to control re-emerging and resistant viruses calls for the need of new therapeutic candidates [273]. In this aspect, there has been a shift in the interest of antiviral drug development towards bioactive peptides. Generally, antiviral peptides have the advantages of convenient oral administration, small size, low cytotoxicity, mild side effects, good solubility, and effective production cost [274]. There are more than 500 antiviral peptides documented in the AVPpred web server [226], which facilitates the study and design of peptides as alternative antiviral therapeutics.

Given many valued features of cystine knot scaffold and their distinct properties identified in this study, alstotides are attractive candidates for antiviral drug development in two aspects: broad-spectrum, stable therapeutics against virus and cell-penetrating carriers.

Broad-spectrum, stable therapeutics

In the context of growing concern over human health and huge loss due to viral infections annually, the identification of alstotide S1 as a broad-spectrum antiviral drug is economically advantageous. Alstotide S1 demonstrated inhibition against the unrelated viruses IBV and DENV, two important threats to poultry and human, respectively. It also has the potential to inhibit other viruses. The conventional inhibitors targeting conserved regions in fusion proteins are usually selective for the virus of origin [266]. Perhaps with a hypothesized mechanism targeting endocytosis, a broader spectrum could be covered by alstotide S1.

The stable alstotide S1 molecule is potentially able to overcome the two main obstacles of oral drugs: low oral bioavailability and susceptibility to enzyme degradation. Enfuvirtide, widely known as T20, was introduced in 2003 as the first effective inhibitor of HIV and is currently in clinical use. A major disadvantage of T20 is that it has a low oral bioavailability, which entails subcutaneous administration of T20 by injection twice a day. Moreover, if orally administered, T20 and other drugs must survive the harsh acidic and proteolytic environment in the digestive system before they reach the circulation system. Linear peptides, like T20, are generally highly susceptible to proteolytic degradation. Alstotide S1 has superior properties in these two aspects. Being able to traverse the cell lipid bilayer, alstotide S1 most likely can pass the gut membrane and become orally available. With its high tolerance to proteases, alstotide S1 survives degradation by digestive enzymes in human body. Therefore, alstotide S1 can be directly used as a broad-spectrum, stable antiviral drug, or it can undergo protein engineering to enhance desired properties for more effective applications.

Cell-penetrating carrier

In the past decades, there has been increasing interest for the use of cell-penetrating peptides (CPPs) as the carrier of “antiviral epitope cargo” across the cell membrane [275]. CPPs are generally water-soluble, and/or polybasic with a net positive charge at physiological pH [276]. Their size ranges from a few (<10) to 30-35 amino acids. CPPs include Tat-derived peptides, homeodomain-derived peptides, signal-sequence-based peptides, and chimeric peptides. CPPs carry biologically active cargos into the cells with high efficiency and low cytotoxicity [277]. A wide range of cargos have been reported to be transported into the cells by CPPs, such as peptides, proteins, oligonucleotides, DNA, nanoparticles and liposomes [278, 279].

Many successful examples of grafting antiviral therapeutics onto CPP scaffolds have been reported. An example is AZP-R9, the conjugation of the 9-mer-of-arginine carrier (R9) with artificial zinc-finger protein (AZP). AZP is a protein

generated to inhibit DNA replication of human papillomavirus type 18 (HPV-18) [280]. The conjugated peptide maintains anti-HPV property with EC_{50} <31 nM and IC_{50} for cytotoxicity >10 µM. Their selective index (>300) is therefore better than the commercialized cidofovir drug for HPV.

With its location intracellularly in the perinuclear region, alstotide S1 is apparently a cell-penetrating peptide. Its open end can be ligated covalently to the biologically active cargoes. The potential applications of alstotide S1 as CPP carrier is aided by its high stability and non-cytotoxicity to many human cell lines at micromolar range. Furthermore, its inherent antiviral activity may provide additional beneficial effects to the main therapeutics in a synergistic manner.

Chapter 5

Study of the Hyperdiversity of Cystine Knot Peptides by Genomic Approach

Introduction

In recent years, with the aid of advanced technology, many proteins and peptides have been discovered and studied. Cumulative evidence of their structure and function has challenged the conventional concept in biology: “one protein produced from RNA is associated with one dedicated function”. A growing number of single proteins or a group of homologous proteins possessing related structure display a wide range of functions. This gives rise to the idea of protein promiscuity, which has been reviewed recently from different perspectives [281-284]. In general, promiscuity is believed to be utilized by plants as an essential strategy to adapt and evolve in various environmental conditions [283].

Protein promiscuity could be divided into two classes: pure and family promiscuities. Pure promiscuity indicates different and perhaps unusual functions of a single protein under different conditions such as pH and protein concentration. In contrast, family promiscuity is related to a group of homologous proteins which share an identical structural scaffold and display multiple functions. Family promiscuity is common in plant defense peptides [283], perhaps due to their inherent function in warding off different kinds of pests and pathogens.

Together with the cysteine-stabilized αβ (CSαβ) motif of defensin, the cyclic cystine knot of cyclotide (CCK) is found as a typical, well- studied scaffold for family promiscuity [284]. Over 150 sequences of cyclotides and linear

cyclotides have been documented in the Cyclic Protein Database (Cybase) and over 200 more identified in our program. A wide range of biological functions have been described for this cystine knot family, as reviewed in Chapter 1. On the CCK scaffold, cyclotide members share different levels of sequence identity from almost identical to completely different except six conserved cysteines. In our study, examples of cyclic and linear counterparts sharing exactly primary sequences have been found. Together, cyclotide members display diversity in both sequence and form (cyclic and linear) on the same CCK scaffold.

In examining ribosomal defense peptides, Arnison *et al.* observed that the hints for diversity, for example post-translational modification, are pre-determined and encoded in the gene precursors [285]. The genomic approach therefore could reveal the origin of diversity beyond the structure and sequence of mature peptides. Therefore, as part of this project, we used genomic approach to study the diversity that leads to protein promiscuity of a cystine knot family, the cyclotide family. Genomic approach was also used here as a complementary tool to proteomic methods for profiling cyclotides in medicinal plants localizing to Southeast Asia.

Results

A routine genomic method has been established in our laboratory to study various medicinal plants endogenous to Southeast Asia, particularly Singapore. The strategy at both RNA and DNA level is similar to that applied for CK α-amylase inhibitors presented in Chapter 3. Briefly, 3′ RACE PCR was performed using a degenerate primer targeting a conserved sequence of cyclotides. One or several partial 3′ end genes were normally obtained. Specific primers targeting 3′ untranslated region (UTR) of the newly discovered genes were designed for 5′ RACE PCR. Assembling 3′ RACE and 5′ RACE clones yielded full-length cyclotide transcripts, from which specific primers were designed to clone DNA sequences.

A total of 101 genes were identified from eight species belonging to three families (Rubiaceae, Fabaceae, and Poaceae) as listed in Table 5.1. Among them, over 60 genes were derived from this project. Most of the genes share common features known for precursor elements found in CRPs, such as

(i) Signal peptide predicted by SignalP 3.0 [155, 156] starts with Met and could be divided into three regions, namely n – region (positively charged, 1 – 5 residues), h – region (hydrophobic, 7 – 15 residues), and c – region (polar, 3-7 residues). They are generally 20 – 25 residues long.

(ii) Intron is a DNA AT-rich fragment with conserved GT and AG sequences at two terminal processing sites.

(iii) Most cyclotides have a terminal Asn/Asp residue which marks the processing site of asparaginyl endopeptidase (AEP).

In this chapter, genomic work on cyclotides from seven plants, namely *Hedyotis biflora, Clitoria ternatea, Hedyotis uncinella, Hedyotis chrysotricha, Hedyotis diffusa, Chassalia chartaceae*, and *Panicum laxum* will be presented in chronological order of the experiments.

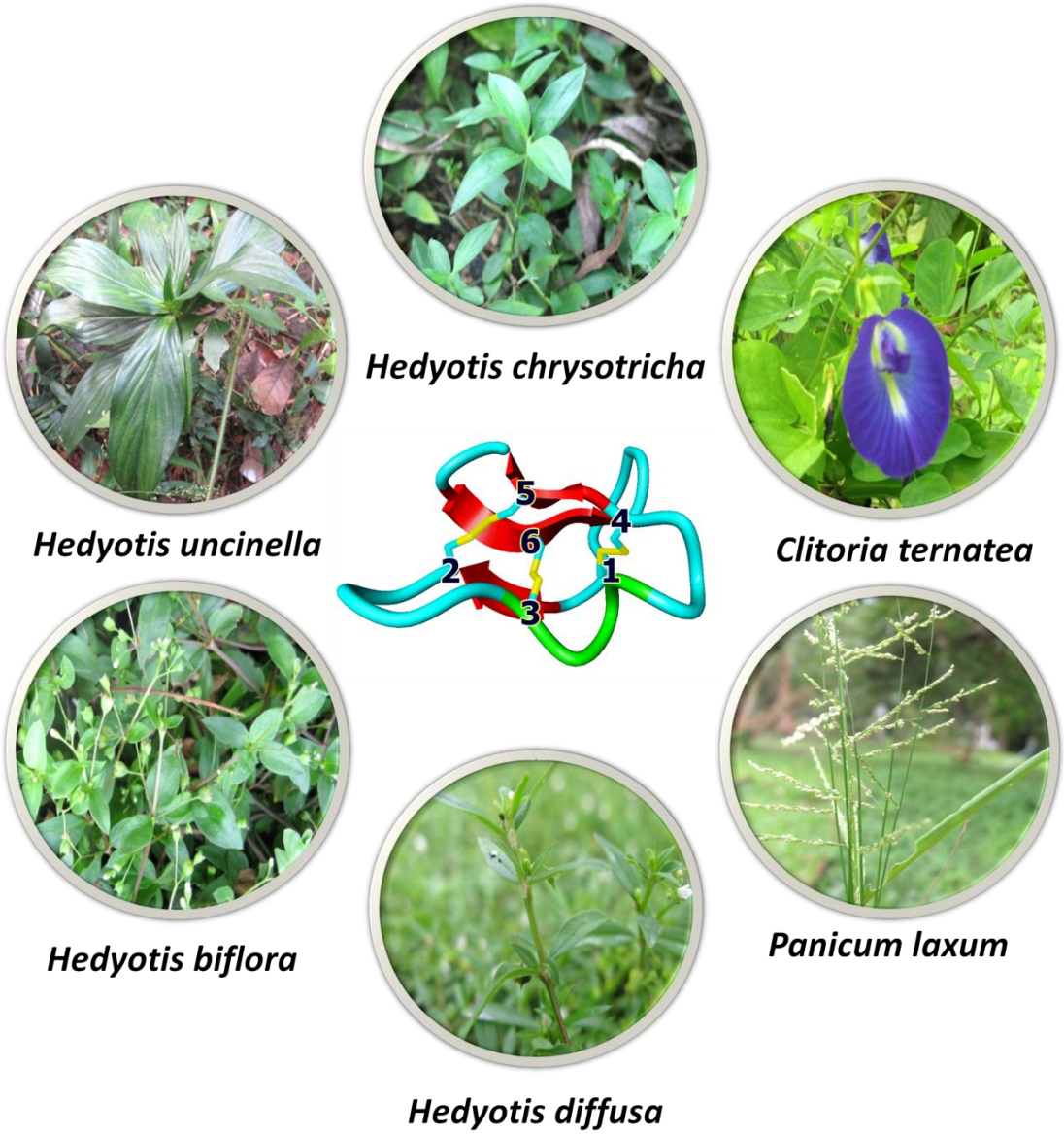


Figure 5.1. Cyclotide-producing plants from Rubiaceae, Fabaceae, and Poaceae families. *Hedyotis chrysotricha, Hedyotis uncinella, Hedyotis biflora,* and *Hedyotis diffusa* belong to the Rubiaceae family. *Clitoria ternatea* and *Panicum laxum* belong to the Fabaceae and Poaceae families, respectively.

Table 5.1. The number of cyclotides obtained in our project. The total number of sequences also includes linear counterparts of cyclotides.

Species	Family	No. of protein sequences	No. of gene sequences
Hedyotis biflora	Rubiaceae	38	17
Hedyotis diffusa	Rubiaceae	6	3
Hedyotis chrysotricha	Rubiaceae	39	6
Hedyotis uncinella	Rubiaceae	58	31
Chassalia chartaceae	Rubiaceae	19	10
Morinda citrifolia	Rubiaceae	12	2
Clitoria ternatea	Fabaceae	33	25
Panicum laxum	Poaceae	9	7
Total		**213**	**101**

Hedyotis biflora

The cloning project of cyclotides from *Hedyotis biflora* (Rubiaceae family) was initiated by Dr. Wang W. and Dr. Nguyen K.T.G. and assisted by Mr. Zhang S. before I joined the program. *Hedyotis biflora* is a small herb native to tropical and subtropical areas. It is traditionally used to treat body pain in fever and malaria by Indian tribes [286].

Only bracelet cyclotides, hedyotides B1-B31, were identified from *H. biflora,* of which 18 were confirmed by genetic sequences. Hedyotide precursor has a conventional structure found in many Rubiaceae cyclotides, comprised of an ER signal peptide, a propeptide, a cyclotide domain, and a C-terminal hydrophobic tail. The hydrophobic tail in hedyotide precursor contains 13 amino acids, which is longer than the normal range of 3-11 residues found in other cyclotide precursors.

We found several clusters of related cyclotides which differ from each other only at the two termini in primary sequences. They include hedyotides B4/B13/B15, B9/B16/B29, B10/B11, B12/B14/B30, and B17/B31. Two new features were observed for these clusters, which had not been reported before for cyclotides from the same species. First, there are two pairs of cyclotides, hedyotides B9/B29 and B12/B14, in which one cyclotide is different from the other by a single Gly at the N-terminus. The GGAVP motif at the N-terminus of hedyotides B9 and B12 is shortened by one Gly to produce GAVP motif in hedyotides B29 and B14, respectively. Both the long and the shortened variants are cyclic. Second, these clusters of almost identical sequences comprise both cyclic and linear analogs. It is common in these clusters that when all the Gly at the N-terminus is removed, only linear counterparts could be found, as in the case of hedyotides B16 and B30. Intriguingly, we also found linear cyclotides that share complete identity to “cyclic” cyclotides, for example hedyotides B1a and B6a. In other words, the linear and cyclic variants are completely identical except for an open or cyclized backbone.

Similar examples have only been identified when cyclotides were expressed in non-cyclotide-producing *A. thaliana* plant with inefficient processing machinery [139]. Such pairs of identical linear and cyclic variants have not been observed in naturally cyclotide-producing plants.

To address the question that whether members within each cluster arise from a single gene or from different genes, we particularly targeted those clusters in our cloning experiments. However, despite extensive efforts, we were only able to find one gene, and thus one precursor, for each cluster. As an example, the gene and deduced amino acid sequence for cluster hedyotides B4/B13/B15 are presented in Figure 5.2. It is likely that each cluster comprises differentially processed products of a single precursor.

The cloning of the linear hedyotide B2 precursor (Figure 5.3) revealed a premature stop codon in place of the conserved Asn at the C-terminus, the target of AEP for cyclization. Previously, the clone *Oak9* encoding kalata B20-lin was also found to have a stop codon after Arg30, producing a peptide lacking the processing site Asn [112]. The sequence that follows the stop codon in *Oak9* is highly homologous to that of *Oak8*. The stop codon in *Oak9* is therefore the result of a single nucleotide mutation from a conventional cyclotide precursor.

Hedyotide B2 precursor has the conventional motif -LKGIQ- at N-terminal cleavage site. Alignment with other cyclotide precursors suggested a cleavage site between residue K and G which should give rise to a protein starting with GIQCGES. However, hedyotide B2 identified in our program lacks the first G residue. Despite our exhaustive efforts, the species starting with GIQCGES has not been found.

Four DNA sequences were obtained using signal peptide primer derived from hedyotide B4 transcript and a reverse primer against the conserved 3′ UTR. A conserved position of one intron, 104-110 bases, was found in the middle of residue 17 of the signal peptide. Intriguingly, introns of hedyotides B1 and B2

are almost identical with only one base different despite their highly different mature domain. On the other hand, hedyotides B1 and B2 introns showed moderate homology to introns from hedyotides B4 and B12. The finding reflects the long-held view that intron is largely conserved due to the lack of pressure for adaption as compared to exon.

Table 5.2. Cyclotides found in Hedyotis biflora

Peptides	Amino acid sequences							Möbius /		Cyclic /	MW	
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)	Sources
		I	II	III	IV	V	VI					
hB1	GTR	C GET C	FVLP	C WSAKFG	C	Y C	QKGF C	YRN	Bracelet	P,cDNA,DNA	Cyclic	3403	Aerial
hB1a	GTR	C GET C	FVLP	C WSAKFG	C	Y C	QKGF C	YRN		P,cDNA,DNA	*Linear*	3421	Aerial
hB1b	TR	C GET C	FVLP	C WSAKFG	C	Y C	QKGF C	YRNE		P,cDNA,DNA	*Linear*	3364	Root
hB1c	GTR	C GET C	FVLP	C WSAKFG	C	Y C	QKGF C	YRNE		P,cDNA,DNA	*Linear*	3550	Root
hB1d	TR	C GET C	FVLP	C WSAKFG	C	Y C	QKGF C	YRN		P,cDNA,DNA	*Linear*	3493	Root
hB2	IQ	C GES C	VWIP	C ISSAWG	C	S C	KNKI C	SS		P,cDNA,DNA	*Linear*	2986	Root
hB3	GIS	C AET C	VLIP	C ISSVIG	C	T C	QNKR C	YKN	Bracelet	P,cDNA	Cyclic	3182	Root
hB5	GTR	C GET C	FVLP	C QSAKFG	C	Y C	QKGF C	YRD	Bracelet	P	Cyclic	3346	Root
hB6	GTP	C AET C	IFIP	C TVTALIG	C	S C	QNQI C	YKN	Bracelet	P	Cyclic	3270	Root
hB6a	GTP	C AET C	IFIP	C TVTALIG	C	S C	QNQI C	YKN		P	*Linear*	3288	Root
hB7	GTP	C AES C	VYLP	C VTAVIG	C	T C	QNRV C	YLN	Bracelet	P,cDNA	Cyclic	3156	Root
hB8	GTP	C AES C	VYIP	C TVTALIG	C	S C	QNQI C	YKN	Bracelet	P	Cyclic	3258	Root
hB4	IP	C GES C	AFIP	C LTSLLG	C	T C	QNKV C	YRDE		P,cDNA,DNA	*Linear*	3260	Root
hB13	GIP	C GES C	AFIP	C LTSLLG	C	T C	QNKV C	YRD	Bracelet	P,cDNA,	Cyclic	3170	Root
hB15	GIP	C GES C	AFIP	C LTSLLG	C	T C	QNKV C	YRDE		P,cDNA,	*Linear*	3317	Root
hB9	GGAVP	C GET C	VYIP	C ITAAIG	C	S C	QNNV C	YHN	Bracelet	P,cDNA,DNA	Cyclic	3236	Root
hB29	GAVP	C GET C	VYIP	C ITAAIG	C	S C	QNNV C	YHN	Bracelet	P,cDNA,DNA	Cyclic	3179	Root
hB16	AVP	C GET C	VYIP	C ITAAIG	C	S C	QNNV C	YHN		P,cDNA	*Linear*	3140	Root
hB10	REE	C GET C	YILP	C VTPD	C	I C	SGGQ C	YKIE		P,cDNA	*Linear*	3206	Root
hB11	EE	C GET C	YILP	C VTPD	C	I C	SGGQ C	YKIE		P,cDNA	*Linear*	3050	Root
hB11a	pEE	C GET C	YILP	C VTPD	C	I C	SGGQ C	YKIE		P,cDNA	*Linear*	3032	Root

Peptides	Amino acid sequences							Möbius /		Cylic /	MW	
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)	Sources
hB11b	E C GET C YILP C VTPD				C I C SGGQ		C YKIE		P,cDNA	*Linear*	2921	Root
hB12	GGVP C GES C VWIP C ISSVFG				C T C QNSDKA		C YHN	Bracelet	P	Cyclic	3440	Root
hB14	GVP C GES C VWIP C ISSVFG				C T C QNSDKA		C YHN	Bracelet	P	Cyclic	3383	Root
hB30	VP C GES C VWIP C ISSVFG				C T C QNSDKA		C YHN		P	*Linear*	3344	Root
hB17	GNP C GES C VYIP C ITTVVG				C S C QNSV		C YHN	Bracelet	P	Cyclic	3126	Root
hB31	NP C GES C VYIP C ITTVVG				C S C QNSV		C YHN		P	*Linear*	3087	Root
hB18	GGAVP C GET C VYIP C ITAAIG				C S C QNKV		C YRD	Bracelet	P	Cyclic	3270	Root
hB19	GIP C GES C VIIP C ISTVIG				C T C ENKV		C LRN	Bracelet	P	Cyclic	3100	Root
hB20	GGAVP C GES C VYIP C ITTVVG				C S C QNSV		C YHN	Bracelet	P	Cyclic	3239	Root
hB21	GTP C GES C VFIP C VTTVIG				C T C QNRV		C YLN	Bracelet	P	Cyclic	3156	Root
hB22	GIP C GES C IFIP C FTAALG				C S C QNKV		C YRN	Bracelet	P	Cyclic	3173	Root
hB23	GIS C GES C VFIP C ISTVLG				C S C QNKV		C YKN	Bracelet	P	Cyclic	3131	Root
hB24	GTIP C GES C IFIP C ISTVIG				C S C KNNV		C YRN	Bracelet	P	Cyclic	3270	Root
hB25	GTS C GEG C VFIP C ITAVIG				C S C QNKV		C YLN	Bracelet	P	Cyclic	3058	Root
hB26	Q C GES C VWIP C ISSAWG				C S C KNKI		C SS	Bracelet	P	*Linear*	2872	Root
hB27	pQ C GES C VWIP C ISSAWG				C S C KNKI		C SS	Bracelet	P	*Linear*	2856	Root

Abbreviation: P protein, pE or pQ: pyroglutamate

(Courtesy of Tam J.P., Wang W., Nguyen K.T.G., Zhang S.)

Figure 5.2. The nucleotide and deduced amino acid sequence of hedyotide B4/B13/B15 from H. biflora. Intron is presented in low case whereas the ORF in upper case, signal peptide in bold and cyclotide domain in blue with highlighted cysteine residues. The orange triangle marks the signal peptide cleavage site predicted with SignalP 3.0 using hidden Markov models (HMM). Neural network from the same website, however, predicted the processing site between residue 19 and 20 (CVG – VL) but careful comparison with available data almost excluded this possibility. The asterisk is put beneath the stop codon.

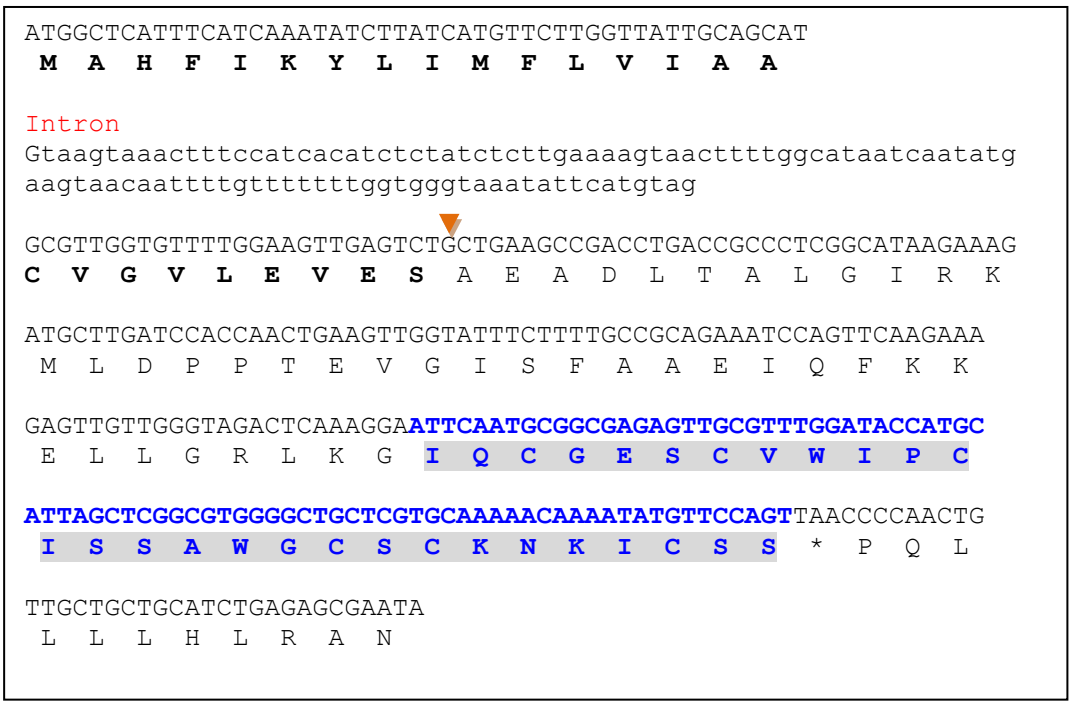


Figure 5.3. The nucleotide and deduced amino acid sequence of hedyotide B2 from H. biflora. Intron is presented in low case whereas the ORF in upper case, signal peptide in bold and cyclotide domain in blue with highlighted cysteine residues. The orange triangle marks the signal peptide cleavage site predicted with SignalP 3.0 using hidden Markov models (HMM). The asterisk is put beneath the stop codon. Data taken from [201]

Clitoria ternatea

Clitoria ternatea is a perennial herbaceous Fabaceae plant commonly known as “butterfly pea”. *Clitoria ternate* species has been long used in Ayurvedic medicine with a broad medicinal spectrum, such as memory enhancing, antistress, anticonvulsant, and sedative activites [287]. The decoction from roots and flowers is also considered as emmenagogue agent in Cuba [91]. *C. ternatea* plant has strikingly high expression of cyclotides which yielded a 1000-fold stronger signal in MS spectrum as compared to other *Hedyotis* plants when purified under the same condition. A total of 33 cyclotides and linear cyclotides were identified (Table 5.3), among which 22 were confirmed by both proteomic and genomic methods, 8 only by proteomic and 3 only by genomic method. The cyclotides were designated as cliotides T1-T25. Together, these novel sequences make Fabaceae the fourth family reported to produce cyclotides with proteomic evidence in addition to Violaceae, Rubiaceae, and Cucurbitaceae families [105].

To obtain the genomic profile for *C. ternatea*, we chose cliotide T2 as our starting target. Cliotide T2 was one of the most abundant cyclotides in all plant parts. Owing to the exceptionally high expression of cyotides, the PCR using degenerate primer targeting ‘GEFIKCGE’ segment (5′ GGnGArTTyAThAArTGyGGnGA 3′) successfully amplified cliotide T2 clone, *ctc2* (courtesy of Zhang Sen). Analysis of the encoding behavior of *H. biflora* cyclotides showed that some multiple codon-encoded residues, particularly cysteine, at certain positions have their favorable codon(s) over other codons. Therefore, we employed the specific codons encoding Cys, Gly, and Glu in loop 1 from cliotide T2 gene to design primer cT3_B1 (5′ ACnTGT GGG GAA ACn TGT AC 3′). A single round of 3′ RACE PCR using this primer produced a sharp band in agarose gel electrophoresis from which partial precursor of cliotide T3 was obtained.

To rapidly and efficiently get the clones of many cliotides, we designed a low-degeneracy primer against the conserved loop 1 ‘CGESC’ on the knowledge of cliotides T2 and T3 genes. This strategy proved efficient with four partial cliotide genes found in a single PCR, namely *ctc4, ctc5, ctc8,* and *ctc9* encoding cliotides T4, T5, T8, and T9, respectively.

The alignment of all known cliotide sequences by ClustalW revealed a highly, though not absolutely, conserved fragment ‘AESEGFLLKD’ in the tail downstream of the cyclotide domain. Using a specific reverse primer against this conserved region in 5′ RACE PCR, we successfully cloned the full genes of several sequences, including cliotides T7 and T3. To our surprise, we identified *ctc20* clone encoding cliotide T20, which escaped our conventional extraction protocol owing to its unusually high net charge +4.

To massively clone full-length sequences of cliotides, we used the short stretch at the beginning of all known signal peptides as PCR primers, namely ‘MAYVRL’ and ‘MASLRIA’, in 3′ RACE PCRs. These reactions amplified a novel cyclotide, cliotide T16, whose expression fell below the detection level in our experimental setup. Similarly, applying primers targeting signal peptide and conserved ‘AESEGFLLKD’ 3′ tail fragment in PCRs on DNA template, we identified the genes of cliotides T17 and T18 which have not been found in MS spectra.

Analysis of the deduced protein precursors revealed a novel arrangement of cliotides which was strikingly different from that of Rubiaceae and Violaceae families (Figure 5.4). Cliotides in Fabaceae family apparently lacked the pro-region connecting the signal peptide and the NTR domain. The SignalP 3.0 program predicted the processing site of signal peptide immediately located before the cyclotide domain. In addition, the unusually long tail at C-terminus in cliotide precursors harbors a CRP. BLAST search showed that this C-terminal CRP was highly homologous to the pea albumin chain a (PA1a), a storage protein typical for legume species. As such, our discovery unveiled an unusual gene arrangement in which cyclotide was linked to a storage protein,

not with another cyclotide domain. The storage protein PA1a is normally found in the same transcript with a cystine knot peptide, PA1b, in legume species. To our surprise, the cliotide domain is located precisely in place of the CK peptide PA1b. Both cliotide and other pea albumin precursors contains an ER signal, PA1b or cyclotide domain, a short linker, and PA1a peptide followed by a tail (Figure 5.5).

The question about the biosynthetic origin of cyclotide gene in legume species *C. ternatea* has not been answered. We tried to seek the presence of cyclotides in related species to determine cyclotide distribution in the Fabaceae family. Due to the limited members of *Clitoria* genus and sparse availability of those plants in local areas, we were only able to screen two species, namely *Clitoria fairchildiana* and *Clitoria laurifolia*. None of them were found to express cyclotides in our experimental condition. Some distantly-related Fabaceae species, such as *Pisum sativum*, *Vignaunguiculata*, and *Vigna radiate*, have also been screened but no positive signal of cyclotides were detected. However, using MS profiling the peaks in the mass range of pea albumin PA1a and PA1b were found in all Fabaceae plants tested, except *C. ternatea*.

To provide insights into the gene structure as well as the relationship between cyclotides in Fabaceae family and other families, introns from cliotide and hedyotide clones were compared. The results showed that intron is located at its conserved position in the middle of the signal peptide in both cliotide and hedyotide genes. However, they significantly varied in length from 107 to 196 residues. Surprisingly, cliotide T7 was found to lack intron, as shown by two independent experiments (Figure 5.5).

Table 5.3. Cyclotides found in Clitoria ternatea

Peptides	Amino acid sequences							Mobius /	Method	Cylic /	MW	Sources
Loop	6	1	2	3	4	5	6	Bracelet		Linear	(Da)	
		I	II	III	IV	V	VI					
cT1	GIP	C GES	C VFIP	C ITGAIG	C	S C	RSRV C	YRN	Bracelet	P,cDNA	Cyclic	3084	F,Sh,Po,R,St
cT2	GEFLR	C GES	C VQGE	C YTPG	C	S C	DWPI C	RRN	**Möbius**	P,cDNA,DNA	Cyclic	3260	F,Sh,Po
cT2a	GEFLR	C GES	C VQGE	C YTPG	C	S C	DWPI C	RRN		P,cDNA,DNA	*Linear*	3278	F,Sh,Po
cT2b	EFLR	C GES	C VQGE	C YTPG	C	S C	DWPI C	RRN		P,cDNA,DNA	*Linear*	3221	F,Sh,Po
cT2c	FLR	C GES	C VQGE	C YTPG	C	S C	DWPI C	RRN		P,cDNA,DNA	*Linear*	3092	F,Sh,Po
cT2d	LR	C GES	C VQGE	C YTPG	C	S C	DWPI C	RRN		P,cDNA,DNA	*Linear*	2945	F,Sh,Po
cT3	GLPT	C GET	C TLGT	C YVPD	C	S C	SWPI C	MRN	**Möbius**	P,cDNA,DNA	Cylic	3058	F,L,St
cT4	GIP	C GES	C VFIP	C ITAAIG	C	S C	RSRV C	YRN	Bracelet	P,cDNA	Cyclic	3098	F,Sh,Po,R,St
cT5	GIP	C GES	C VFIP	C ISTVIG	C	S C	RNRV C	YRN	Bracelet	P,cDNA	Cyclic	3169	F,Po,R
cT6	SIP	C GES	C VYIP	C ITTIVG	C	S C	RNSV C	YSN	Bracelet	P	Cyclic	3119	F
cT7	GIP	C GES	C VFIP	C TVTALLG	C	S C	RDRV C	YRN	Bracelet	P,cDNA,DNA	Cyclic	3227	F,Se,R
cT8	GIP	C GES	C VFIP	C ISSVVG	C	S C	RSRV C	YNN	Bracelet	P,cDNA,DNA	Cyclic	3072	Se
cT9	GIP	C GES	C VFIP	C LTTVVG	C	S C	RNRV C	YNN	Bracelet	P,cDNA	Cyclic	3127	Se
cT10	GVP	C AES	C VWIP	C TVTALLG	C	S C	RDRV C	YLN	Bracelet	P,cDNA,DNA	Cyclic	3251	Se
cT11	GVIP	C GES	C VFIP	C ISTVIG	C	S C	RNRV C	YRN	Bracelet	P,cDNA	Cyclic	3268	F,R
cT12	GIP	C GES	C VYIP	C TVTALLG	C	S C	RDRV C	YRN	Bracelet	P	Cyclic	3243	R,St
cT13	GIP	C AES	C VWIP	C TVTALLG	C	S C	RDRV C	YLN	Bracelet	P	Cyclic	3265	R,St
cT14	GIP	C GES	C VFIP	C TITALLG	C	S C	RDRV C	YRN	Bracelet	P	Cyclic	3241	R,St
cT15	GLPI	C GET	C FRTR	C YTRG	C	S C	SYPV C	RRN	**Möbius**	P,cDNA	Cyclic	3165	R
cT16	ARIP	C GES	C VWIP	C TITALVG	C	A C	HERV C	YRSS	Bracelet	cDNA		-	-
cT17	GTVP	C GES	C VFIP	C ITGIAG	C	S C	RNRV C	YLN	Bracelet	DNA		3155	-
cT18	RIP	C GES	C VWIP	C LTGYFG	C	Y C	QSRV C	YHN	Bracelet	DNA		3377	-
cT19	GSVIR	C GES	C LLGR	C YTPG	C	T C	SRPI C	RRN	**Möbius**	P,cDNA	Cyclic	3124	R
cT19a	GSVIR	C GES	C LLGR	C YTPG	C	T C	SRPI C	RRN		P,cDNA	*Linear*	3125	R

Peptides	Amino acid sequences							Mobius /	Method	Cylic /	MW	Sources
Loop	6	1	2	3	4	5	6	Bracelet		Linear	(Da)	
cT20	GSAIRC	GES	C LLGR	C YTPG	C T	C DRPI	C RRN	**Möbius**	P,cDNA	Cyclic	3152	R
cT20a	AIR	C GES	C LLGR	C YTPG	C T	C DRPI	C RRN		P,cDNA	*Linear*	3026	R
cT20b	IR	C GES	C LLGR	C YTPG	C T	C DRPI	C RRN		P,cDNA	*Linear*	2955	R
cT21	DIQ	C AET	C VHSP	C IGP	C Y	C RHGLI	C YRN	Bracelet	P	Cyclic	3102	R
cT22	GSVIG	C GET	C LRGR	C YTPG	C T	C DHGI	C RRN	Bracelet	P,cDNA	Cyclic	3107	R
cT22a	VIG	C GET	C LRGR	C YTPG	C T	C DHGI	C RRN		P,cDNA	*Linear*	2981	R
cT23	GLPI	C GET	C FTGT	C YTPG	C T	C SYPV	C RRN	**Möbius**	P	Cyclic	3022	R
cT24	DTTP	C GES	C VWIP	C VSSIVG	C S	C QNRV	C YQN	Bracelet	P	Cyclic	3326	SC
cT25	DTIP	C GES	C VWIP	C ISSIIG	C S	C RDRV	C YHN	Bracelet	P	Cyclic	3404	SC

Abbreviation: P protein, F flower, Sh shoot, St stem, Se seed, R root, Po pod, SC Seed Coat

(Courtesy of Tam J.P., Nguyen K.T.G.,Nguyen T.K.N., Chiu S., Zhang S., Hardjojo A.)

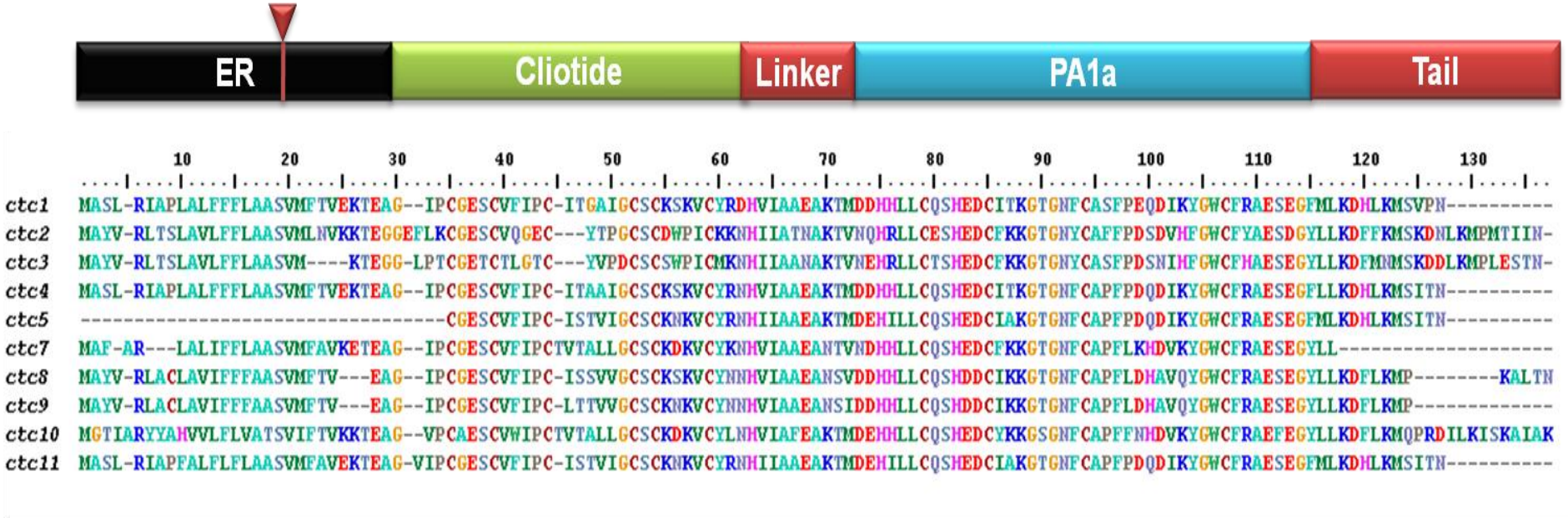


Figure 5.4. Schematic presentation and deduced amino acid sequences of cliotide precursors. The novel gene organization of cliotides is entirely different from that of previously identified cyclotides from Rubiaceae and Violaceae families. The precursors comprise an ER signal peptide, cliotide domain, a short linker, PA1a, and a C-terminal tail. The intron is here depicted by a triangle in the ER signal peptide. The alignment highlights the homology of signal peptide among cliotide precursors. PA1a has not been found in C. ternatea extract. The organization of the later part is predicted based on the alignment with known pea albumins. (Data published in [190])

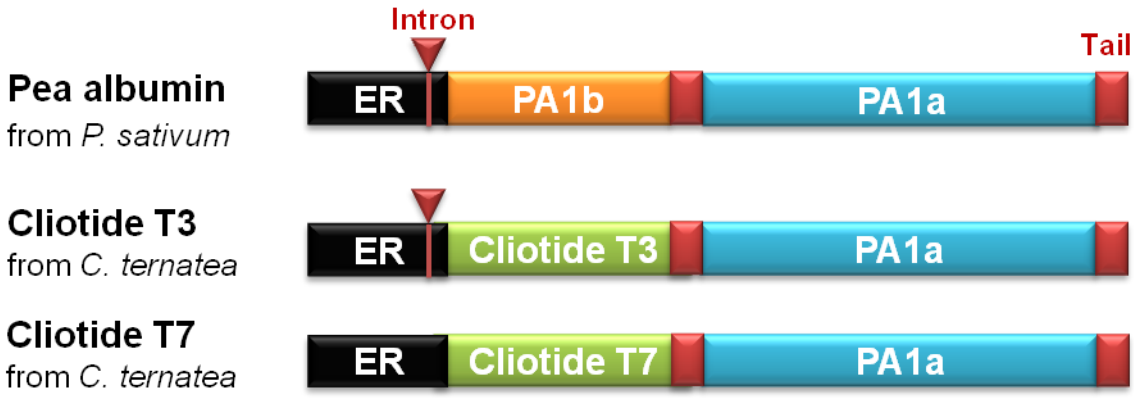


Figure 5.5. Schematic gene organization of cliotides and pea albumin. Each precursor comprises an ER signal peptide (black), mature cyclotide domains (orange) or PA1b (green), linker peptide, pea albumin domain (blue), and predicted short hydrophobic tails (red). Intron locations are marked with a triangle. The gene structure of cT3 precursor also represents those of cliotides T2 and T8.

Hedyotis uncinella

Hedyotis uncinella is a perennial, erect herb that can grow up to 70 cm high. *H. uncinella* has long been accepted in Traditional Chinese Medicine as a drug to treat ulceratic gingivitis, enteritis, diarrhea, and dysentery. It has a wide distribution over China yet appears sparsely in highlands of Vietnam. Because of the lack of a well-equipped greenhouse, we were unable to nurture the plant in the Nanyang herbatorium. The whole-plant sample for this study was directly collected from Lam Dong province, Vietnam and preserved in humid soil during its transport to Singapore.

More than 40 cyclotides and linear variants were identified from *H. uncinella* by proteomic methods (courtesy of Dr. Wang S.) (Table 5.4). The cyclotides were designated as hedyotides U1-U33. The expression levels were generally comparable to cyclotides from *H. biflora*. By far *H. uncinella* contains the highest number of cyclotides and linear cyclotides identified within one plant species. *Viola hederacea* (Violaceae family) was previously reported to contain 66 cyclotides in different tissues, but only ten sequences were reported at the protein level [288]. In addition to the clones encoding cyclotides identified by MS/MS, 14 out of 31 clones obtained from *H. uncinella* encode novel sequences which have not been found (Figure 5.6). These genes could be silent (only found at DNA level) or expressed at an undetectable level (found at cDNA level but not by MS analysis), or detected yet impossible to sequence by MS/MS (appearing as tiny and poorly resolved peaks). They accounted for nearly 30% of cyclotides identified from *H. uncinella*, reflecting the advantage of genomic approach as a complementary tool to proteomic method.

To our surprise, the cyclotide sequences from *H. uncinella* displayed remarkable variation. In some cases, residues considered essential for cyclotides to maintain their unique topology and functional activities are missing. Twelve cyclotides/linear cyclotides have Asp residue in place of almost strictly conserved Glu in the middle of loop 1. Hedyotides U35 and U37

lack both negatively charged Asp and Glu at similar position. 21 cyclotides/linear cyclotides have a basic amino acid, either Lys or His, at position 1 or 3 of loop 1, which is rare among cyclotides. Small residues like Gly, Ala, Ser or Thr are normally found at such positions in known cyclotides. Surprisingly, hedyotide U38 and U40 with the sequences ‘GDEYCNNETCVMFPCRSPGCRCDGMYCVKYD’ and ‘GDEFCNNETCVMFP-CRS-PGCRCNGMYCVKYD’, respectively, harbor four residues rather than three in loop 1.

Table 5.4. Selective cyclotides found in Hedyotis uncinella

Peptides	Amino acid sequences							Möbius /		Cylic /	MW	
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)	Sources
		I	II	III	IV	V	VI					
hU1	GGDT	C HET	C FFLP	C FTPG	C S	C SGTQ	C VTN	Bracelet	P,A,cDNA,DNA	Cyclic	2991	L
hU1a	GGDT	C HET	C FFLP	C FTPG	C S	C SGTQ	C VTN		P,E,A,cDNA	*Linear*	3008	L
hU1b	GDT	C HET	C FFLP	C FTPG	C S	C SGTQ	C VTN		P,cDNA	*Linear*	2951	L
hU1c	DT	C HET	C FFLP	C FTPG	C S	C SGTQ	C VTN		P,cDNA	*Linear*	2894	L
hU2	GVVP	C GDT	C FLGP	C NDSS	C S	C IAGI	C LRN	Bracelet	P,DNA	Cylic	2880	L
hU4	GGDT	C HET	C FFLP	C FTSG	C S	C SGTQ	C VTN	Bracelet	P,cDNA	Cyclic	2981	L
hU13	GIP	C GES	C VFIP	C ATAAIG	C S	C QNKV	C MRN	Bracelet	P,A,cDNA	Cyclic	3051	R
hU19	GGNY	C GET	C RVLG	C ITAIAG	C H	C VNKN	C EHN	Bracelet	P,E,cDNA	Cyclic	3215	R
hU21	GNY	C GET	C RVLG	C ITAIAG	C H	C VNKN	C EHN	Bracelet	P,E,cDNA	Cyclic	2951	L
hU26	GGV	C KET	C IFLP	C FSTG	C S	C SGRI	C YYN	Bracelet	P,A,cDNA	Cyclic	2987	L
hU27	GGI	C KET	C IFLP	C FSTG	C S	C SGRI	C YYN	Bracelet	P,cDNA	Cyclic	3001	L
hU29	GGM	C KET	C IFLP	C FSTG	C S	C SGNI	C YYN	Bracelet	cDNA,DNA	Cyclic	2978	St
hU30	GGI	C KET	C IFLP	C FSTG	C S	C SGRI	C YYD	Bracelet	P,cDNA	Cyclic	3002	L
hU31	GHF	C SDS	C IIFD	C DDKS	C S	C INHF	C YKN		cDNA	pCyclic	3181	R
hU32	GLF	C GDP	C FFGP	C NDPS	C S	C IARF	C YKN		P,cDNA	Cyclic	3040	St,R
hU33												
hU33a	SYP	C GDPC	FFGP	C NDPS	C S	C IARF	C YKNE		P,cDNA	*Linear*	3215	St
hU33b	YP	C GDP	C FFGP	C NDPS	C S	C IARF	C YKNE		P,cDNA	*Linear*	3130	St
hU33c	YP	C GDP	C FFGP	C NDPS	C S	C IARF	C YKN		P,cDNA	*Linear*	3001	St
hU34	GQP	C GES	C VWIP	C ISSVVG	C S	C QNKI	C YLN	Bracelet	DNA	pCyclic	3166	
hU35	IAYQ	C GGS	C VYVP	C ISSVAYP	C T	C IEKK	C ILLE		cDNA	*pLinear*	3552	R
hU36	GVY	C NET	C IIFP	C FTPG	C S	C SPFASL	C YNN	Bracelet	DNA	pCyclic	3229	

Peptides	Amino acid sequences							Möbius /		Cylic /	MW	
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)	Sources
		I	II	III	IV	V	VI					
hU37	GRF	C GQA	C SIFP	C VTPG	C S	C SRGA	C FRN	Bracelet	DNA	pCyclic	2903	
hU38	GDEY	C NNET	C VMFP	C RSPG	C R	C DGMY	C VKYD	Bracelet	cDNA	pCyclic	3534	R
hU39	GDMF	C IDT	C MIFN	C YSSR	C E	C TRGL	C YTK		DNA	*pLinear*	3356	
hU40	GDEF	C NNET	C VMFP	C RSPG	C R	C NGMY	C VKYD	Bracelet	cDNA	pCyclic	3517	St
hU41	GGDT	C HET	C FVLP	C FSSG	C S	C SGYQ	C VKN	Bracelet	DNA	pCyclic	3008	
hU42	GGI	C KET	C IFLP	C FSTG	C S	C SGNI	C YYN	Bracelet	cDNA	pCyclic	2959	R
hU43	GLF	C GDP	C FFGP	C NDPS	C S	C IARF	C YKN	Bracelet	cDNA	pCyclic	3040	R
hU44	GLTP	C GES	C IWIP	C ISAAIG	C S	C QNKV	C YRN	Bracelet	cDNA	pCyclic	3066	R
hU45	GES	C FET	C FILP	C FNPD	C S	C SGSY	C YTN	Bracelet	cDNA	pCyclic	3265	St

Abbreviation: P protein, St stem, R root, L leaf

(Courtesy of Tam J.P., Wang S., unpublished data)

** pCyclic: presumably cyclic based on theory of processing pathway for known cyclotides*

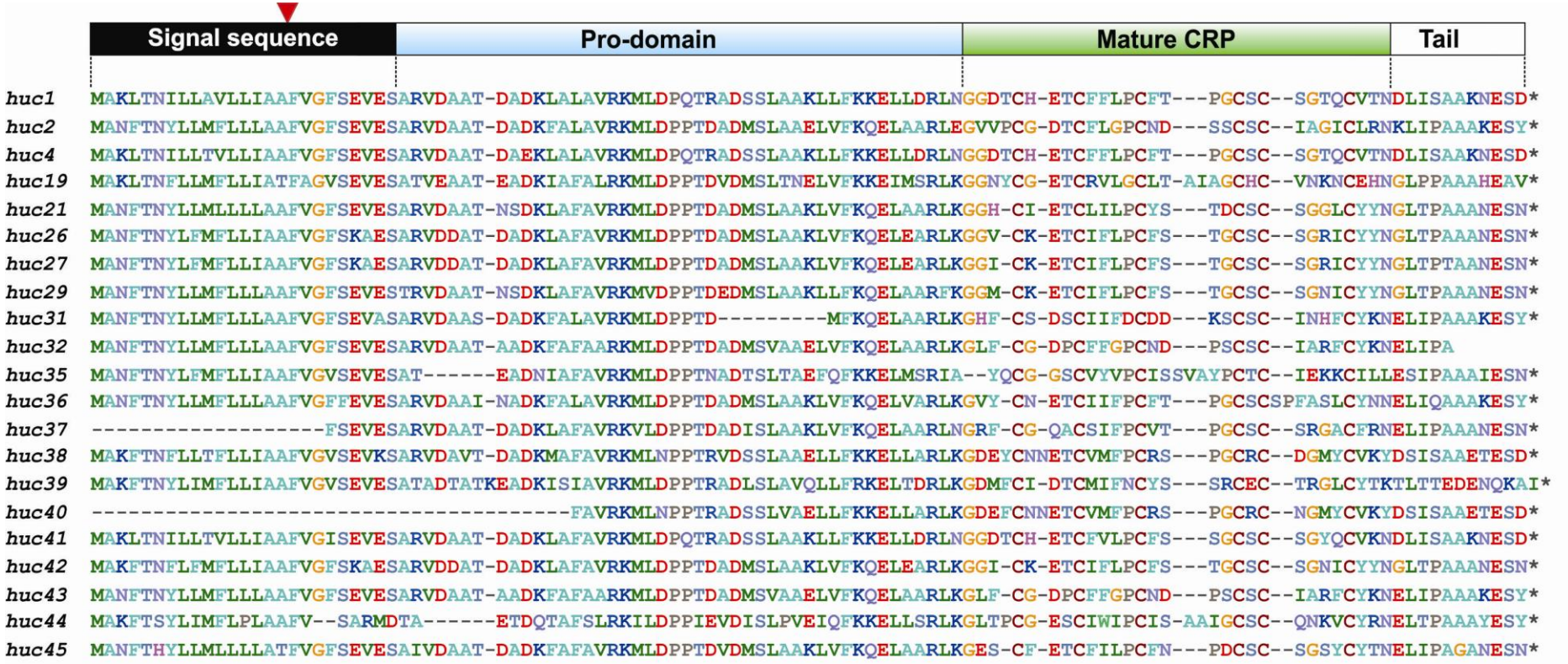


Figure 5.6. Schematic presentation and deduced amino acid sequences of cyclotide precursors from Hedyotis uncinella. The precursor comprises an ER signal peptide, a prodomain, and a mature cyclotide domain followed by a tail. The intron is here depicted by a triangle in the ER signal peptide.

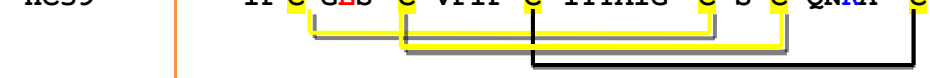
Hedyotis chrysotricha

Hedyotis chrysotricha is a medicinal herb commonly used in enteritis and hepatitis treatment in Southeast Asia and China. The plants were collected from Liuzhou, China and recultivated in Nanyang herbatorium. MS/MS analysis has identified 14 uncyclotides and 25 cyclotides, named hedyotide C1-C39 (courtesy of Wang S., unpublished data). Two cDNA sequences were obtained for two clusters: hedyotides C25/C36/C37 and C27/C38/C39 (Figure 5.7). Like H.*biflora* cyclotides, only one clone was found for each cluster, suggesting that members of the same cluster are different protein products from a single precursor. One intron was identified at its conserved position in the middle of the 20-amino-acid signal peptide.

Sequence analysis revealed unusual motif at the two termini of linear cyclotides from *H. chrysotricha*. Hedyotides C1-C10 start with LFC/FC at the N-terminus and end with YTQD/YTQ at the C-terminus. In these cases, the N-terminal processing site is Asp instead of Asn. It is possible that the inefficient enzymatic cleavage by AEP after Asp residue produces linear species with two open ends which are later trimmed off by a cellular exopeptidase. Interestingly, in cyclotide/linear cyclotide pairs of hedyotides C25/C36 and C27/38, the linear variant has an overhanged Glu at the N-terminus as compared to its cyclic counterpart. This could result from an unspecific activity of AEP despite the presence of a favorable Asn residue at the conserved C-terminal position.

Table 5.5. Selective cyclotides found in Hedyotis chrysotricha

Peptides	Amino acid sequences							Mobius /		Cylic /	MW
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)
		I	II	III	IV	V	VI				
hC25	GVP	C	GES	C	VFIP	C	ISAAIG	C S C	QNKA	C YRN	Bracelet	P,cDNA	Cyclic	3069
hC36	GVP	C	GES	C	VFIP	C	ISAAIG	C S C	QNKA	C YRNE	Bracelet	P,cDNA	*Linear*	3216
hC37	VP	C	GES	C	VFIP	C	ISAAIG	C S C	QNKA	C YRNE	Bracelet	P,cDNA	*Linear*	3159
hC27	GIP	C	GES	C	VFIP	C	ITTAIG	C S C	QNKA	C YRN	Bracelet	P,cDNA	Cyclic	3127
hC38	GIP	C	GES	C	VFIP	C	ITTAIG	C S C	QNKA	C YRNE	Bracelet	P,cDNA	*Linear*	3274
hC39	IP	C	GES	C	VFIP	C	ITTAIG	C S C	QNKA	C YRNE	Bracelet	P,cDNA	*Linear*	3217



Abbreviation: P protein, St stem, R root, L leaf

(Courtesy of Tam J.P., Wang S., Nguyen K.T.G., Hardjojo A.)

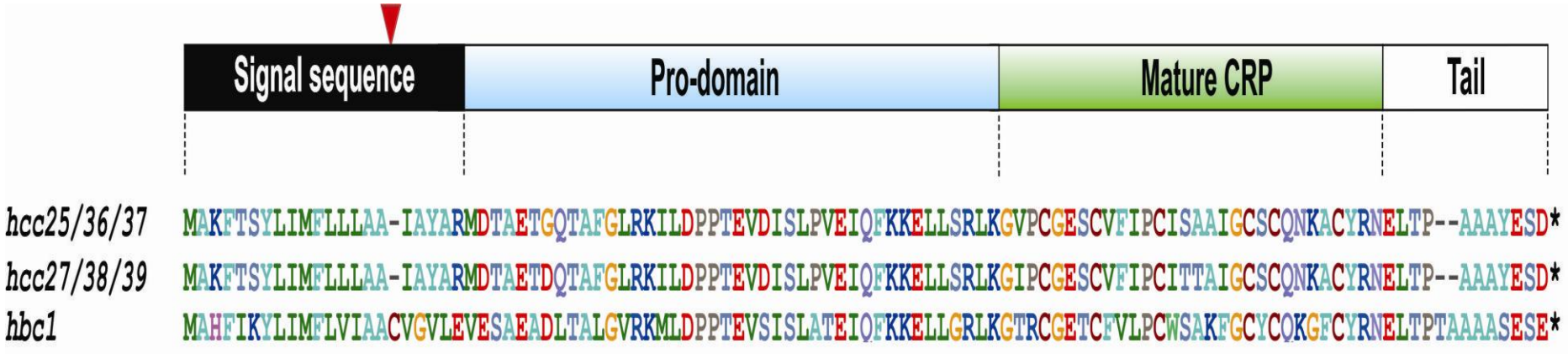


Figure 5.7. Schematic presentation and deduced amino acid sequences of cyclotide precursors from Hedyotis chrysotricha. The precursor comprises an ER signal peptide, a prodomain, and a mature cyclotide domain followed by a tail. The intron is here depicted by a triangle in the ER signal peptide. The precursor of hedyotide B1 (hbc1 clone) was included for comparison.

Hedyotis diffusa

The herbal plant *Hedyotis diffusa*, also known as Bai Hua She She Cao or snake needle grass, was readily collected in Nanyang herbatorium or along the sidewalk in Nanyang Technological University campus (Singapore). The proteomic work identified 6 linear cyclotides from H. *diffusa*, named hedyotides D1-D6 (Table 5.6). Their expression was low, about ten-fold lower than that of *H. biflora* (0.5 mg/kg for *H. biflora*), and varied seasonally.

Linear hedyotides D1-D6 lack a small residue, e.g. Gly, at the N-terminus and the indispensible Asn/Asp at C-terminus. Sequence alignment showed that hedyotides D1-D6 belonged to two clusters, hedyotides D1/D2/D3 and D4/D5/D6. Members in each cluster differ from each other by one residue at the C-terminus, as observed commonly for cyclotides/linear cyclotides from *H. biflora, H. uncinella,* and *H. chrysotricha*. Particularly, the C-terminus, ending with VKSVS or VKIVS, was perhaps gradually trimmed off to produce many linear variants from a single precursor.

The full sequences of hedyotide D4/D5/D6 were obtained at both RNA and DNA levels (Figure 5.8), revealing a gene organization similar to that of cyclotides from other Rubiaceae species. The cyclotide domain was separated from the signal peptide by a pro-region and followed by a 10-amino-acid hydrophobic tail. An intron of 202 bp was located in the middle of the signal peptide in DNA sequence.

Table 5.6. Cyclotides found in Hedyotis diffusa

Peptides Loop	Amino acid sequences 6	1	2	3	4	5	6	Method	Cylic / Linear	MW (Da)
	I	II	III		IV	V	VI			
hD1	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKS	P	*Linear*	3369
hD2	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKSV	P	*Linear*	3468
hD3	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKSVS	P	*Linear*	3555
hD4	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKI	P,cDNA,DNA	*Linear*	3581
hD5	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKIV	P,cDNA,DNA	*Linear*	3494
hD6	VG	C YER C	VWGP C	ISKIVG	C	S C	DTTTYD C	VKIVS	P,cDNA,DNA	*Linear*	3395

Abbreviation: P protein

(Courtesy of Tam J.P., Nguyen K.T.G.)

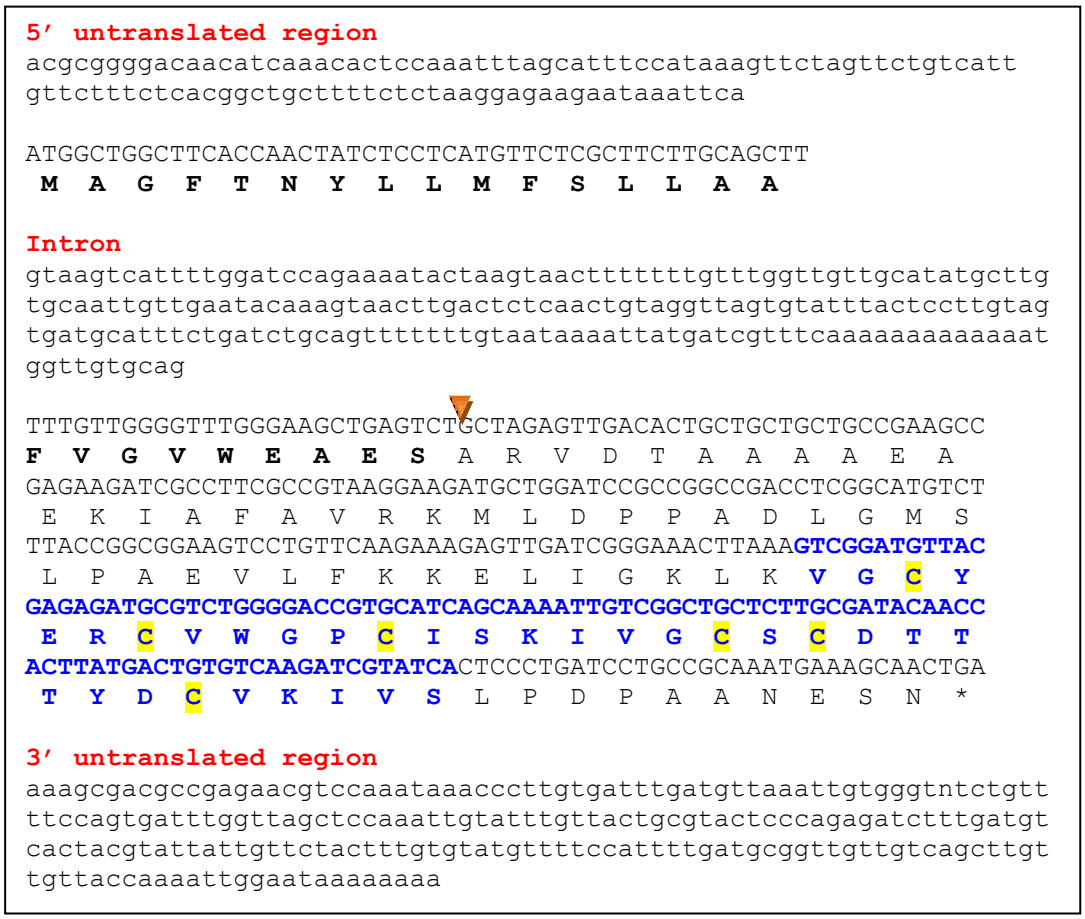


Figure 5.8. The nucleotide and deduced amino acid sequences of hedyotides D4/D5/D6 from H. diffusa. Untranslated regions and intron are presented in low case whereas the ORF in upper case, signal peptide in bold and cyclotide domain in blue with highlighted Cysteine residues. The orange triangle marks the signal peptide cleavage site predicted with SignalP 3.0 using hidden Markov models (HMM). The asterisk is put beneath the stop codon.

Chassalia chartaceae

Chassalia chartaceae plant was obtained by Lim W. in a field trip collection in MacRitchie reservoir park, Singapore. It is a medium-sized tree with white, five-petal flowers and dark blue berries typical for Rubiaceae family. A total of 15 cyclotides and linear cyclotides were sequenced by proteomic method and named chassatides C1-C15 (Table 5.6). The cloning work yielded 17 precursor sequences, including two partial clones encoding chassatides C7 and C11 and nine full-length clones encoding chassatides C2, C4, C8 and, in addition, C13-C18 (Figure 5.9). Chassatides C13-C18 have not been found at the protein levels despite our exhaustive analysis of *C. chartaceae* MS profiles.

Chassatide precursors share high sequence homology to each other yet significantly different from other rubiaceaous cyclotide, such as kalatas B1 and B7 (*O. affinis*), hedyotides B1 and B2 (*H. biflora*) (Figure 5.10). Interestingly, chassatide precursors are the shortest among all known cyclotides. Particularly, chassatide precursors contain 75-78 residues whereas kalata B1 124 residues and hedyotide B2 107 residues. Other known cyclotide precursors, on average, contain 100 to 200 amino acids [284]. The chassatide precursors are shortened with the NTR absent and the hydrophobic tail comprised of only two residues. The normal range reported for the C-terminal hydrophobic tail of cyclotides is 3-11 amino acids.

Similar to hedyotide B2, the premature stop codon instead of the ultimate Asn/Asp was identified in linear chassatides C7, C8, and C11. Their sequences explain biosynthetic origin of their linearity. They also provide evidence for the essential role of the ultimate Asn/Asp for backbone cyclization.

Table 5.7. Cyclotides found in Chassalia chartaceae

Peptides Loop	Amino acid sequences 6	1	2	3	4	5	6	Mobius / Bracelet	Method	Cylic / Linear	MW (Da)	Source
		I	II	III	IV	V	VI					
cC1	GDA	C GET C	FTGI C	FTAG	C S C	NPWPT	C TRN	**Möbius**	P	Cyclic	2991	Pe
cC2	GIP	C AES C	VWIP C	TITALMG	C S C	KNNV	C YNN	Bracelet	P,cDNA	Cyclic	3283	Pe
cC2a	GIP	C AES C	VWIP C	TITALM*G	C S C	KNNV	C YNN	Bracelet	P,cDNA	Cyclic	3299	Pe
cC3	GIP	C GES C	VWIP C	ISSALG	C S C	KNKV	C YRN	Bracelet	P	Cyclic	3166	Pe
cC4	GAS	C GET C	FTGI C	FTAG	C S C	NPWPT	C TRN	**Möbius**	P,cDNA	Cyclic	2963	Pe
cC5	GVIP	C GES C	VFIP C	ISSVVG	C S C	KNKV	C YRN	Bracelet	P	Cyclic	3240	Pe
cC6	GVIP	C GES C	VFIP C	ISSVIG	C S C	KNKV	C YRN	Bracelet	P,cDNA	Cyclic	3254	Pe
cC7	IP	C GES C	VWIP C	LTAIAG	C S C	KNKV	C YT		P,cDNA	*Linear*	2956	Pe
cC8	AIP	C GES C	VWIP C	ISTVIG	C S C	SNKV	C YR		P,cDNA	*Linear*	3085	Pe
cC9	GIP	C GEA C	DFIP C	VTTVIG	C S C	KDKV	C YNN	Bracelet	P	Cyclic	3128	R
cC10	GEY	C GES C	YLIP C	FTPG	C Y C	VSRQ	C VNKN	Hybrid	P	Cyclic	3212	R
cC11	IP	C GES C	VWIP C	ISGMFG	C S C	KDKV	C YS		P,cDNA	*Linear*	3009	R
cC11a	IP	C GES C	VWIP C	ISGM*FG	C S C	KDKV	C YS		P,cDNA	*Linear*	3025	Pe
cC12	EY	C GES C	YLIP C	FTPG	C Y C	VSRQ	C VNKN	Hybrid	P	Cyclic	3155	Pe
cC13	GFP	C AES C	VYIP C	TVTALLG	C S C	RNRV	C YRN	Bracelet	P,cDNA,DNA	Cyclic	3374	Pe
cC14	GIP	C AES C	VYIP C	TITALFG	C S C	KDKV	C YNN	Bracelet	cDNA	Cyclic	3291	Pe
cC15	GIP	C AES C	VYIP C	TITALLG	C S C	KDKV	C YKN	Bracelet	cDNA,DNA	pCyclic	3271	Pe
cC16	GVP	C AES C	VYIP C	TITALFG	C S C	KDKV	C YNN	Bracelet	cDNA,DNA	pCyclic	3277	Pe
cC17	IP	C GES C	VYIP C	ISAVLG	C S C	QNKV	C YR	Bracelet	cDNA	*pLinear*	3002	Pe
cC18	GIP	C GES C	VFIP C	ISALLG	C S C	SNKV	C YNN	Bracelet	cDNA,DNA	pCyclic	3070	Pe
												Pe

Abbreviation: P protein, Pe pedicel, R root, M oxidized methionine* *(Data taken from [289])*

** pCyclic presumably cyclic, pLinear presumably linear (based on theory of processing pathway for known cyclotides)*

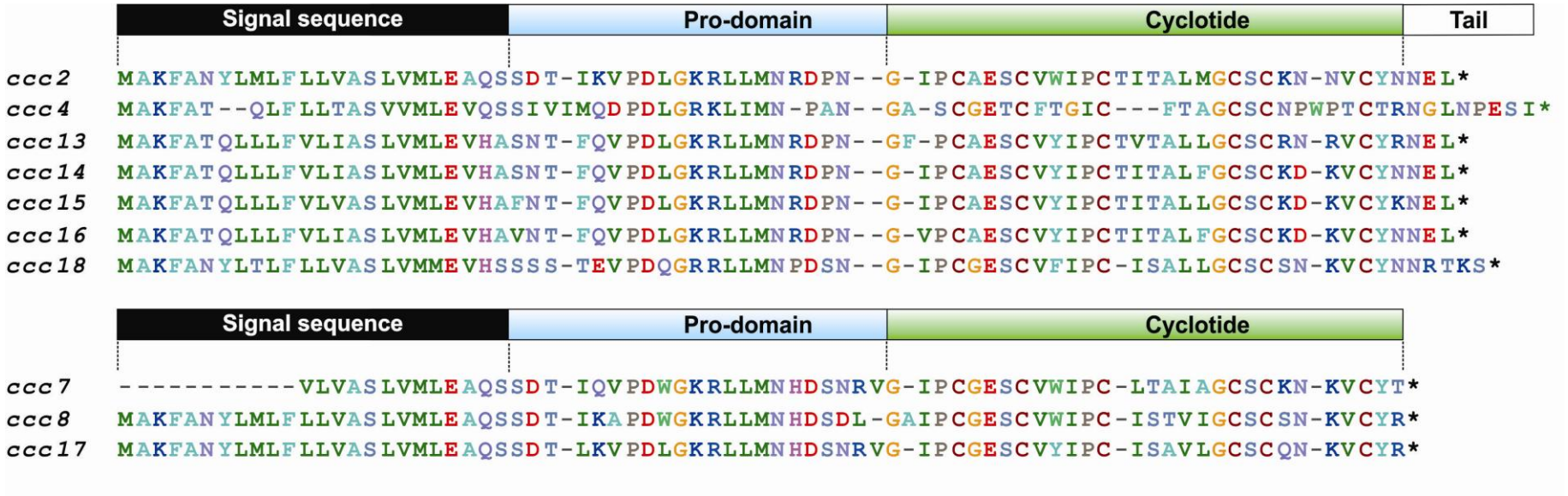


Figure 5.9. Schematic presentation and deduced amino acid sequences of cyclotide precursors from Chassalia chartaceae. The precursor basically comprises an ER signal peptide, a prodomain, and a mature cyclotide domain followed by a tail. The tail is missing in linear chassatide precursors.

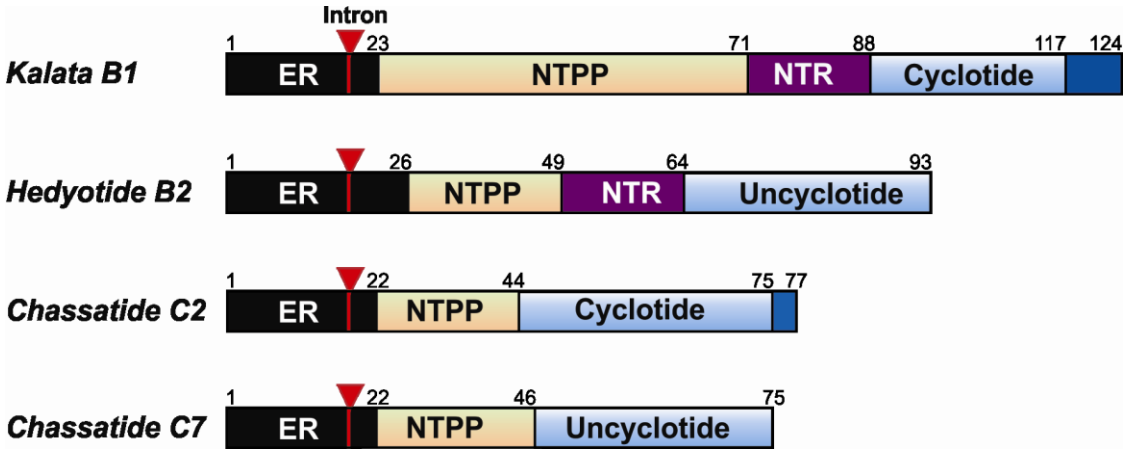


Figure 5.10. Gene structure comparison of chassatides with other rubiaceous cyclotides. Each precursor comprises an ER signal peptide (black), N-terminal propeptide (NTPP) (light orange), N-terminal repeat (NTR) (purple), cyclotide/uncyclotide domain (light blue), and a short hydrophobic tails (dark blue). Intron locations are marked with a triangle. The number above the diagrams indicates the length of each domain. Data published in [289].

Panicum laxum

Panicum laxum is perennial, medium-sized grass in the genus *Panicum* (panic *grass*) and belongs to the Poaceae family (*Grass* Family). It grows to about 30 to 60 cm high and is commonly found in all kinds of pastures with certain humidity, usually due to imperfect drainage. *P. laxum* is a source of cattle fodder and grass for sheep grazing in tropics and subtropics.

The discovery of cyclotides from *P. laxum* was inspired by the data mining into the nucleotide database of several Poaceae plants, in which cyclotide-like genes were found abundantly [104]. Screening of several Poaceae grasses readily available in Nanyang Technological University campus (Singapore) revealed a strikingly high expression of cyclotides in *P. laxum* (about 0.25 mg/g of wet material). A total of eight linear cyclotides, designated as panitides L1-L8, were identified (Table 5.8), marking the first discovery of cyclotides in a monocot plant. Previously, cyclotides have been found only in dicot plant families, including Rubiaceae, Violaceae, Cucurbitaceae, Fabaceae, and Solanaceae families. Four full-length clones (*plc1, plc2, plc4, plc6*) and three partial clones (*plc7, plc8, plc9*) were obtained with the novel panitide L9 having incomplete sequence and as yet identified at protein level (Figure 5.11). DNA sequence of panitide L4 showed that panitide precursor has no intron, similar to violaceous cyclotide precursors. All panitides lack CTPP domain and the Asn/Asp residue essential for cyclization by AEP.

Table 5.8. Cyclotides found in Panicum laxum

Peptides	Amino acid sequences							Mobius /		Cylic /	MW
Loop	6	1	2	3	4	5	6	Bracelet	Method	Linear	(Da)
		I	II	III	IV	V	VI				
pL1	pQLPI	C GET	C VLGT	C YTPG	C R	C QYPI	C VR	**Möbius**	P,cDNA	*Linear*	3052
pL2	pQLPI	C GET	C VLGR	C YTPN	C R	C QYPI	C VR	**Möbius**	P,cDNA	*Linear*	3164
pL3	pQAF	C GET	C LLGK	C YTPG	C S	C HTGI	C LK	Hybrid	P	*Linear*	2812
pL4	pQAF	C GET	C LLGK	C YTPG	C S	C HTGI	C LK	Hybrid	P,cDNA,DNA	*Linear*	2788
pL5	pQLPI	C GET	C VLGT	C YTPG	C S	C AYPI	C AR	**Möbius**	P	*Linear*	2898
pL6	pQLPI	C GET	C VLGT	C YTPG	C S	C AYPI	C VR	**Möbius**	P,cDNA	*Linear*	2926
pL7	pQAF	C GET	C VLGT	C YTPG	C S	C NFGI	C LK	Hybrid	P,cDNA	*Linear*	2794
pL8	pQD	C GET	C VLGT	C YTPG	C S	C SAYPL	C V	**Möbius**	P,cDNA	*Linear*	2648
pL9		C GET	C VLGP	C STPG	C S	C NFGI	C LK	Hybrid	cDNA	-	-

Abbreviation: P protein, pQ pyroglutamic acid *(Data taken from [290])*

Panitide L9 (abbreviated as pL9) is predicted from the gene sequence.

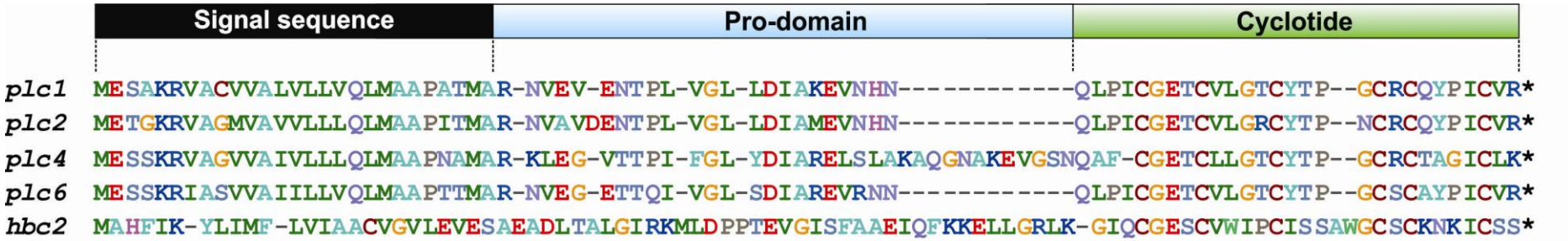


Figure 5.11. Schematic presentation and deduced amino acid sequences of cyclotide precursor from Panicum laxum. The truncated precursor comprises an ER signal peptide, a prodomain, and a mature cyclotide domain. The precursor of hedyotide B2 (hbc2 clone) was included for comparison. Data published in [290].

Discussion

Genomic approach as a validating tool to confirm proteomic sequence

In contrast to many small cyclic peptides, typically 8-10 residues and synthesized by non-ribosomal peptide synthetases [110], cyclotides are gene-encoded and ribosomally synthesized *in planta*. The genetic nature of cyclotides renders genomic approach suitable as a validating tool in support of proteomic method for peptide characterization (Figure 5.12).

With the genes obtained, many ambiguities in MS/MS sequence reading could be readily clarified, such as

(a) similar mass between isoleucine and leucine, glutamine and lysine, phenylalanine and oxidized methionine

(b) poor signals at two termini

(c) high signal – to – noise ratio due to the low abundance of the peptide or the presence of impurity.

Genomic approach is also advantageous in that some peptides undetectable in MS spectra could be revealed in cloning work, as observed for *H. uncinella* and *C. chartaceae*. The failure to detect such peptides in MS could result from the expression variation, e.g. hedyotides U34-45, or a bias in purification protocol targeting hydrophobic CRPs, e.g. the overlooked cliotide T20 due to its unusual net charge +4. Cliotide T20 is normally eluted in early fractions of our routine reverse phase HPLC and thus mistaken as impurities or small molecules that are out of our interest.

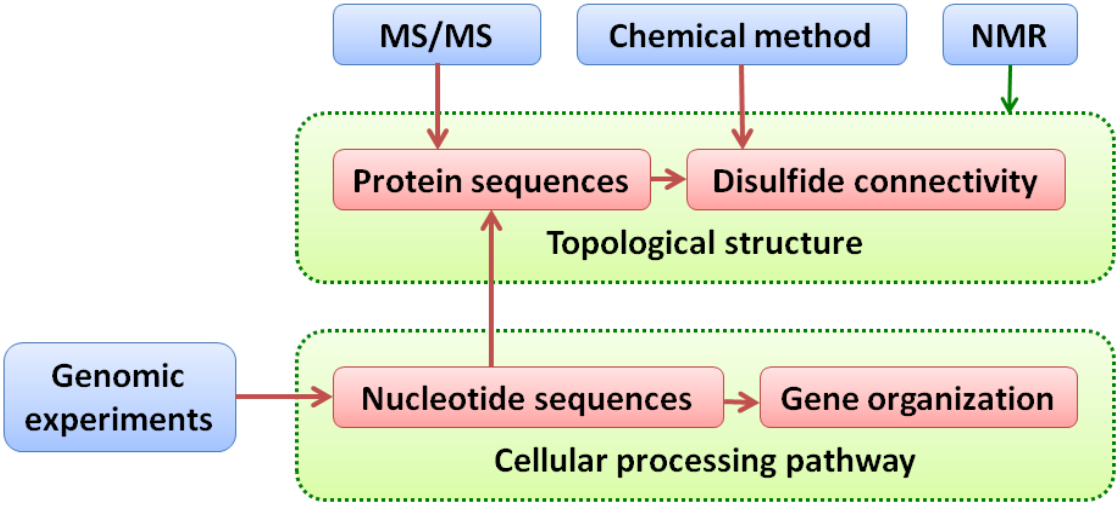


Figure 5.12. Genomic approach in relationship with other methods for peptide study. The primary sequences of proteins are determined by either MS/MS reading or NMR analysis. It becomes the base for disulfide connectivity mapping using chemical method. In this context, genomic approach aids in clarifying the ambiguity in the peptide chain determination. It also gives clues for the cellular processing pathway which may help enhance their potential applications in medicinal industry.

Diversity in cyclotide gene organization

Various gene arrangements of cyclotides from five cyclotide-producing families Fabaceae, Rubiaceae, Violaceae, Poaecae, and Solanaceae are depicted in Figure 5.13. In general, cyclotide precursors share the same basic organization, comprising an ER signal peptide, an N-terminal pro-peptide (NTPP or pro-region) followed by up to 3 cyclotide domains and a short hydrophobic tail. Each cyclotide domain can be further divided into N-terminal repeat and mature cyclotide domains. One intron at conserved position was found in all rubiaceous cyclotide precursors but not in violaceous precursors. Intron at similar postion was present in all cliotides (Fabaceae family), except cliotide T7, and missing in panitides (Poaceae family).

It has been known that different precursor structures in terms of the number of domains, domain type, and intron were employed by plants for different CRP classes (Figure 5.14). However, such a high precursor diversity in one particular CRP family, as shown for cyclotide in Figure 5.13, is rare. This diversity suggests that plants evolve different gene architectures, likely together with different mechanisms of gene expression modulation, to facilitate better adaptation and survival.

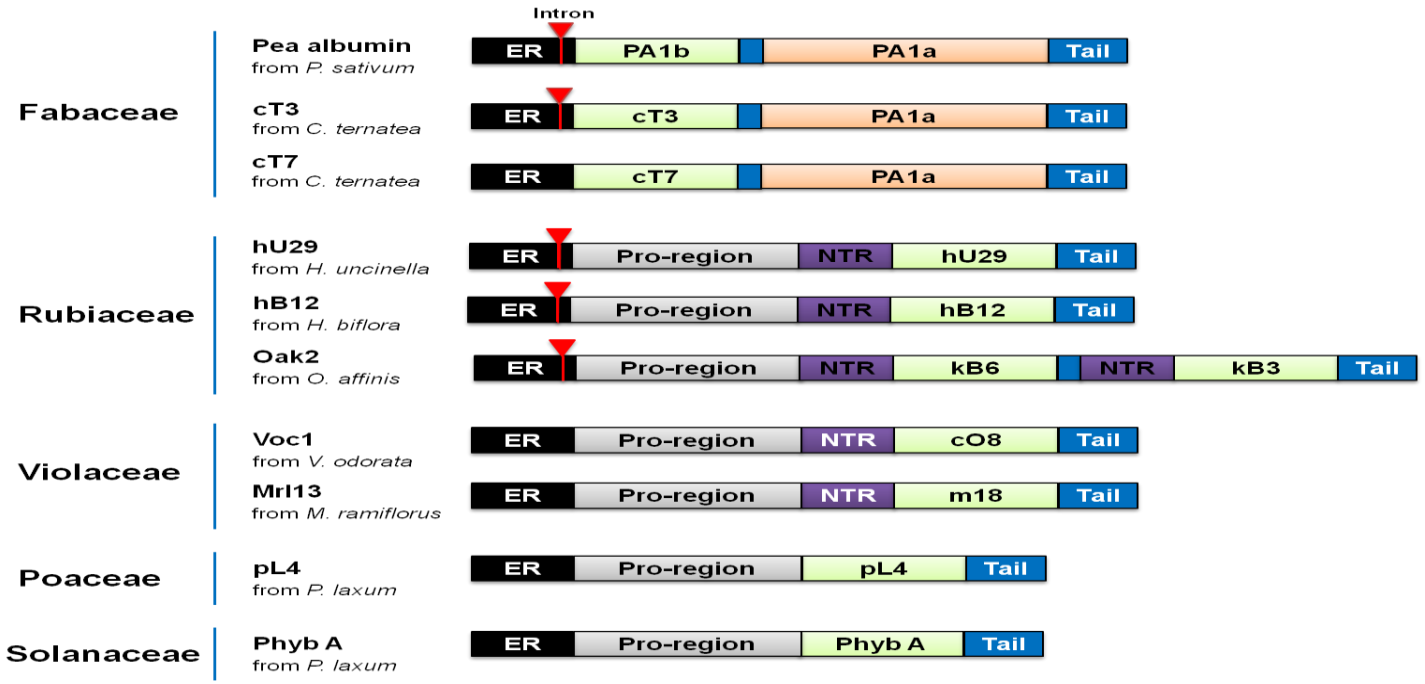


Figure 5.13. Comparison of the gene organization of cyclotides from different families. Pea albumin gene is included to depict the similar arrangement to cliotide genes. Cyclotide precursor generally comprises a signal peptide, N-terminal pro-peptide, mature cyclotide domains or pea albumin chain b domain, N-terminal repeat (NTR) and predicted short hydrophobic tails. Cliotide precursor contains an ER signal peptide, a cyclotide domain, a shortlinker, and pea albumin chain a domain followed by a C-terminal tail. Intron locations are marked with a triangle and remain unknown for solanaceous cyclotides. The gene arrangement of Oak2, Voc1, Mrl13, and Phyb A were adapted from Gruber et al. and Aaron et al. [97, 148].

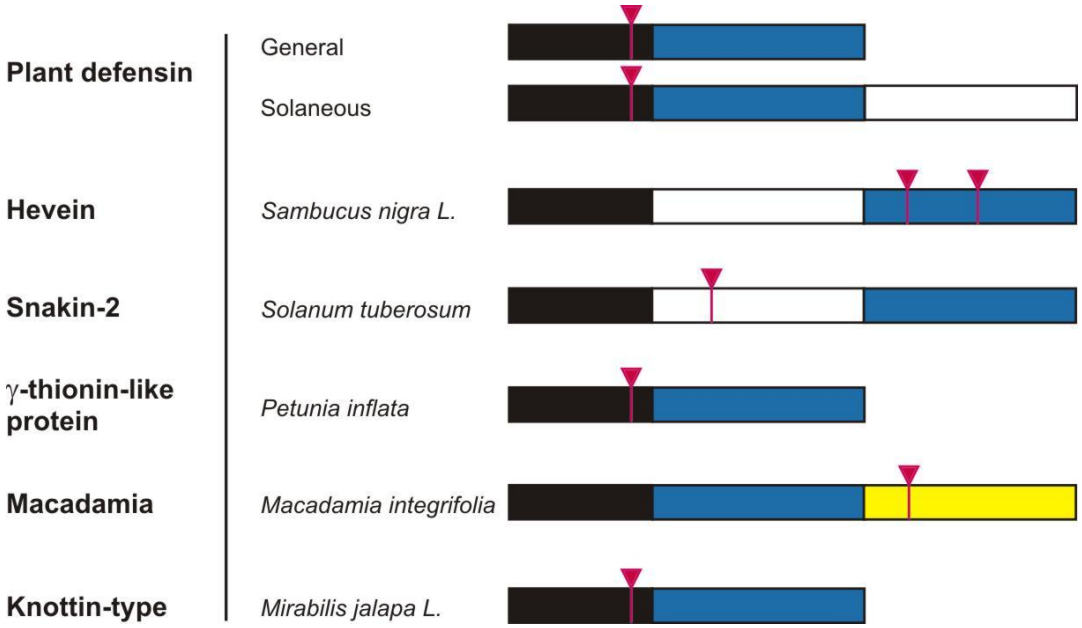


Figure 5.14. Gene organization of plant CRPs. Gene elements are depicted in boxes coloured as follows: ER signal peptide in black, mature peptide in blue, pro-peptide or 3' tail in white, and vicilin, a 7S storage protein in pea seed, in yellow. The red triangles indicate intron location. The relative sizes of gene fragments are only indicative. Adapted from [7, 21, 163-165, 291-293]

Multidomain precursor structure

The structure of multi-domains within a single transcript has been reported for many peptide precursors from both plants and animals (Table 5.9). It is believed that this organization could be a mechanism to boost the simultaneous expression of related compounds in response to stimulating factors, such as pathogenic invasion. This enables a broader spectrum of activities in host defense or, in the case of ubiquitin, fulfills heavy transcription demands for the normal operation of the cells. To date the longest precursor, in terms of the number of repeats in a single transcript, has been the antibacterial peptide in *Riptortus clavatus*, a hemipteran insect, with 14 domains [294]. In the plant kingdom, the 8-repeat precursor TIPTOP3 from *M. cochinchinensis* encoding trypsin inhibitor-type cyclotides [100] is currently the longest. Interestingly, this precursor, together with TIPTOP1 and TIPTOP2 from the same species, encodes for both cyclic and linear cylotides in a single transcript, highlighting the diversity in form (cyclic/linear) determined at the gene level.

In spite of being encoded in a single transcript, the expressions of different domains are not necessarily comparable, for example apidaecins, the bee antibacterial peptides [166]. Unlike the first three isoforms with relative ratio corresponding to those predicted from cDNA clones, the fourth isoform has never been detected in the bee lymph. Similarly, we were unable to find the storage protein PA1a in all parts of *Clitoria ternatea* despite the high expression level of cliotides from the same gene (Figure 5.13). The fusion of cliotide domain into an albumin precursor in *C. ternatea* may present the complexity of cyclotide production by multiple biosynthetic pathways. Therefore, multi-domain structure could be a mechanism employed by plants to modulate the expression of certain genes in response to different environmental conditions.

Table 5.9. Multidomain precursors reported for peptides from animals and plants

Peptide	Species	Characteristic	Size	Number of repeats/units	Reference
Apidaecin	Bee	Inducible, proline-arginine-rich, antimicrobial	18	12	[295]
Magainins	*Xenopusgenus* frog	Antimicrobial	23	6	[296]
Acaloleptin A	*Acalolepta luxuriosa* beetle	Domains separated by diarginyl linker, antibacterial	69-71	5	[297]
Λdiff02 repeat	*Riptortus clavatus* hemipteran	Rich in charged residue, proline-rich, without linker domain,	47	14	[294]
Polyubiquitin	Plant	Without linker domains, linking with other ubiquitins or ribosome-related protein	76	variable	[298]
Na PI	*Nicotiana alata* plant	Homologous to proteinase inhibitors	58-60	6	[299]
Ib-AMPs	*Impatiens balsamina* plant	Cysteine-rich peptides, basic, domains linked by variable acidic linkers, expressed in the seed	20	6	[18]

N-terminal repeat

N-terminal repeat (NTR) is a short repeated portion of the pro-region, about 16-20 residues [129, 284]. It preceeds each mature cyclotide domain in multidomain precursors. NTR is found quite conserved within a single plant species yet more diverse between species. NTR normally possesses a structurally conserved α-helix motif [300]. Many intrinsic functions have been suggested for NTR. They include aiding the folding process of cyclotides via interaction with their surface-exposed hydrophobic patch, detoxifying the cyclotide during their intracellular biosynthesis, playing roles in protein transport or host defense, and most commonly, flanking and signaling the cleavage site for the excision and cyclization of mature cyclotides [139, 300].

However, our discovery of cliotide genes challenges the necessity of NTR in the biosynthesis of cyclotides. The cyclotide domain in cliotide genes immediately follows the signal peptide, indicating the apparent lack of pro-region and NTR in cliotide precursors (Figure 5.13). Without the pro-domain and NTR in precursors, cliotides still maintain the canonical cyclic cystine knot topology. This clearly challenges the proposed essential functions of NTR for cyclotide biosynthesis. Furthermore, an attempt of our laboratory to clarify the function of NTR in folding process revealed that the presence of a separate NTR peptide did not aid the proper oxidative folding of a reduced cyclic analog (Wong C., unpublished data). Therefore, we hypothesize that the most likely function of NTR lies in the host defense system in a synergistic or independent action with mature cyclotides.

Hyperdiversity of cyclotide sequences

Together with >150 cyclotide precursor sequences deposited in Cyclic Protein Database (Cybase) [301], the results in this study present the hyperdiversity of cyclotide sequences using the same CCK fold. The long-standing view on sequence conservation of cyclotides has been challenged. Some critical residues thought to be strictly conserved are no longer true, except the six definitive Cys residues. Here I discuss our findings relevant to sequence conservation and diversity relating to promiscuous functions of cyclotides.

Loop 1

Loop 1 with three intercystenyl amino acids is previously believed to be too short and too conserved to be engineered in cyclotide-based drug development. Most cyclotides have the common motifs of GET, GES, AET, and AES in this loop (Table 5.10). However, a careful analysis of cyclotide sequences revealed a greater diversity in loop 1, particularly in *H. uncinella* (Rubiaceae) (Table 5.4). In contrast to the conventional three-residue length observed for loop 1 of all violaceous cyclotides, hedyotides U38 and U40, though not confirmed at protein level, have loop 1 harboring four amino acids. This result suggests that loop 1 may tolerate amino acid modification.

Glutamic acid residue in loop 1 is highly conserved among cyclotides and occasionally replaced by Asp residue in kalata B2 (*H. affinis*) and htf-1 (*H. terminalis*) [147, 301]. This glutamic acid is reported to be responsible for structure stabilization by its hydrogen bonding network to residues in loop 3 [302]. Methylation of the side chain leads to a dramatic decrease in cytotoxicity of cycloviolacin O2 [173]. The Glu residue is also believed to be involved in metal binding in an experiment with Mn^{2+} [303]. However, an inactive mutant was shown to have high response to Mn^{2+}, suggesting that metal binding may not be a crucial feature for cyclotide activity [303]. In our study, it becomes clear that plant genome employs some residues other than only Glu and Asp at the center of loop 1, for example Gly and Gln in hedyotides U35 and U37, respectively. Alanine substitution at any positions other than 6 Cys residues in

kalata B1 causes no structural change. It does only affect functional properties [303]. Particularly, Ala substitution at Gly and Glu in loop 1 and a few residues in loop 3 almost abolished the hemolytic and insecticidal activity of kalata B1. Accordingly, the unusual sequences in *H. uncinella*, if expressed, may perform different biological functions to provide broader host defense. Furthermore, they reinforce the potential role of the cyclotide scaffold as a natural combinatorial template in cyclotide-based drug engineering.

Table 5.10. Diverse sequences in loop 1 of cyclotide. GET, GES, AES, and AET are the most common sequences found in almost all species. Data for Violaceae family and H. terminalis, Rubiaceae, were obtained from CyBase [301]. The asterisk marks the cyclotides found by genomic but not proteomic method and thus their expressions have not been confirmed.

Violaceae family		**Rubiaceae family**	
Loop 1	***Species***	***Loop 1***	***Species***
GET **GES** **AES** **AET**	**common**	**GET** **GES** **AES** **AET**	**common**
GER	*V. baoshanensis*	GDS	*H. terminalis*
GEG	*V. baoshanensis, M. ramiflorus*	YER	*H. diffusa*
IET	*V. baoshanensis, G. pauciflorum*	GDT	*H. affinis*
		GEG	*H. biflora*
		IET HET KET GDT GDP SDS* GGS* GQA* NET* NNET* IDT* FET	*H. uncinella*

Loop 5

Of particular interest in loop 5 of cyclotide is the presence or absence of a cis-proline residue in Möbius and bracelet subfamilies, respectively. Möbius cyclotides accounted for about half of the cyclotides reported up to 2001 [304]. There is less sequence variation within Möbius than within the bracelet subfamily. Given the existence of Möbius analogs in most of cyclotide-producing plants reported so far, it is surprising that no Möbius molecules were as yet discovered in *H. biflora* and *H. uncinella* despite the high number of cyclotides identified from each plant. This raises questions about the co-existence of the two subfamilies in many cyclotide-producing plants. Is it a deliberate design of mother Nature to include a conceptually twist-causing Pro residue in loop 5 of Möbius cyclotides? Does the backbone twist introduced by this Pro cause any differential effect on the peptide's activity? Are Möbius cyclotides destined to function distinctively from or synergistically with their bracelet analogs? The answers for these questions perhaps call for exhaustive and extensive structure-function studies of cyclotides from various plant species.

Loop 6

Since the first cyclotide gene discovery, loop 6 has widely attracted the scientist community for two main reasons. First, it is the site for terminal cleavage and subsequently end-to-end cyclization in the biosynthesis. In other words, it is involved in the enzymatic processing that determines the form diversity of cyclotide. Studies in this loop revealed much of our current knowledge about cyclotide biosynthesis. Second, loop 6 has the largest size and most variable sequence as compared to other loops. These features render loop 6 highly suitable for grafting diverse bioactive epitopes in protein engineering.

Prior to this project, only a few examples of linear cyclotides have been reported. The discovery of 57 linear cyclotide sequences in this project from Rubiaceae, Fabaceae, and Poaceae families, together with the recent report

of cyclotides in a Solanaceae plant [148], implicates that they are not as rare as previously thought. Further studies of the bioactivities of linear cyclotides suggest that cyclic backbone is unlikely the essential feature for biological activities of cyclotides. Previously, it was believed that the disruption of the end-closed backbone caused dramatic effects on hemolytic activity of cyclotides. The linear violacin A displayed a marked reduction in hemolytic activity relative to kalata B1 [145]. Different acyclic permutants of kalata B1, either with a charged or an amidated C-terminus, completely lost the hemolytic property [305, 306]. It is explained that the linearization perhaps causes some changes in dynamic stabilities and structural features of the linear variants that are required for hemolysis [305]. Structurally, linear analogs were found to have weaker internal hydrogen bonds though they still adopt a native cyclotide-like framework [306]. The free ends were highly flexible according to solution state NMR studies [304]. Interestingly, the linear cyclotide psyle C still maintains cytotoxicity comparable to known cytotoxic cyclotides with IC_{50} at 3.5 µM [146]. Very recently, we published our study on cyclotides from *C. chartacea* in which linear cyclotides displayed strong antimicrobial, cytotoxic, and hemolytic activities [289]. Together, these results suggest that linear cyclotides possess certain biological activities and play a role in plant defense.

Thus, it is intriguing to speculate why plants produce linear cyclotide as an open structure together with their stable, cyclic counterparts. Our laboratory has found some linear cyclotides from various plants with comparable expression to their cyclic variants. Perhaps, it is an evolutionary advantage to produce different forms of homologous defense peptides which, together, may cover broader activity spectrum through a promiscuous multifunctionality strategy.

Biosynthesis of cyclotides

Alignment of cyclotide sequences obtained by proteomic methods reveals many conserved structural features important for cellular processing pathway of cyclotides [139, 307]. However, the enzymes involved in the biosynthesis as well as structural and sequence elements of the precursor required for an

efficient processing *in planta* remain unknown. In this respect, genomic work to clone cyclotide precursors is advantageous and provides insights into the entities involved in cyclotide biosynthesis.

N-terminal processing

Different enzymes have been known to be involved in the removal of flanking regions in CRP precursors. For example, the flanking region in apidaecins is gradually trimmed by dipeptidyl aminopeptidase [308] whereas that in magainin and xenopsin is cleaved by an endopeptidase to release a short peptide [166]. As for cyclotide, although enzyme(s) involved in N-terminal cleavage has not been identified, sequence alignment of the precursor near the processing site suggests the possible involvement of several proteolytic enzymes (Figure 5.15). In *Clitoria ternatea* the signal peptidase I (SPase I) is the single enzyme responsible for N-terminal processing. SPase I, when removing the signal peptide, also cleaves the N-terminus of cyclotide domain because cliotide precursors lack the pro-region. In contrast, the pro-region in Rubiaceae and Violaceae precursor apparently requires a second enzyme for its excision. Although no study has been performed to confirm the N-terminal cleavage mechanism, alignment of precursor sequences around cyclotide N-terminus suggests two possible pathways for two groups of cyclotides.

The first group with L or V at position P2 (the numbering system starting from the residue preceding the excision site in backward direction) might be processed by papain-like cysteine protease (PLCP). PLCP is the largest family among cysteine proteases found in viruses, plants, primitive parasites, invertebrates and vertebrates [309, 310]. The most extensively studied enzymes in this group are mammalian PLCPs known as cathepsins, which are highly homologous to papain. Ten out of 11 cathepsins have glycosylation sites for proteins targeted into lysosomes, leaving the last member retained in the endoplasmic reticulum. They were shown to have high preferences toward substrates with F, L, and, to some extent, V at position P2 [309]. Therefore, we hypothesize that a similar enzyme may exist in the plant ER to function in the propeptide excision.

The second group comprises members with an exclusive Asn at position P1 and variable residue at position P_2' (numbering system starting from the residue after the cleavage sit in forward direction). It is possible that AEP, which is known to process the C-terminus, is also involved in N-terminus processing. AEP, an enzyme prevalent in all plant parts, is named legumains after its activity characterization in the seed of *Vigna aconitifolica* (moth bean) [311]. Besides, AEP is also known as vacuolar processing enzyme (VPE) due to its cellular location in the vacuole and cell wall. In an *in vitro* test, this enzyme exhibited hydrolysis activity toward the scissile peptide bond at the C-terminal side of Asn, and to a much lesser extent, Asp residue (~100 folds weaker) [312]. AEP is synthesized as proprotein precursor in the ER, after signal peptide excision, and then transported to Golgi complex, targeting final location in protein storage vacuole (PSVs) or cell wall [313]. In the acidic environment of PSV or cell wall, the autocatalytic cleavage is initiated to produce an active form of AEP which later performs in the maturation [314, 315] and degradation [312, 313] of storage proteins or in the lysis of waste materials [313]. As such, our precursors can be processed by AEP to excise the pro-region, a mechanism that utilizes the host resource to produce new peptides at minimum cost.

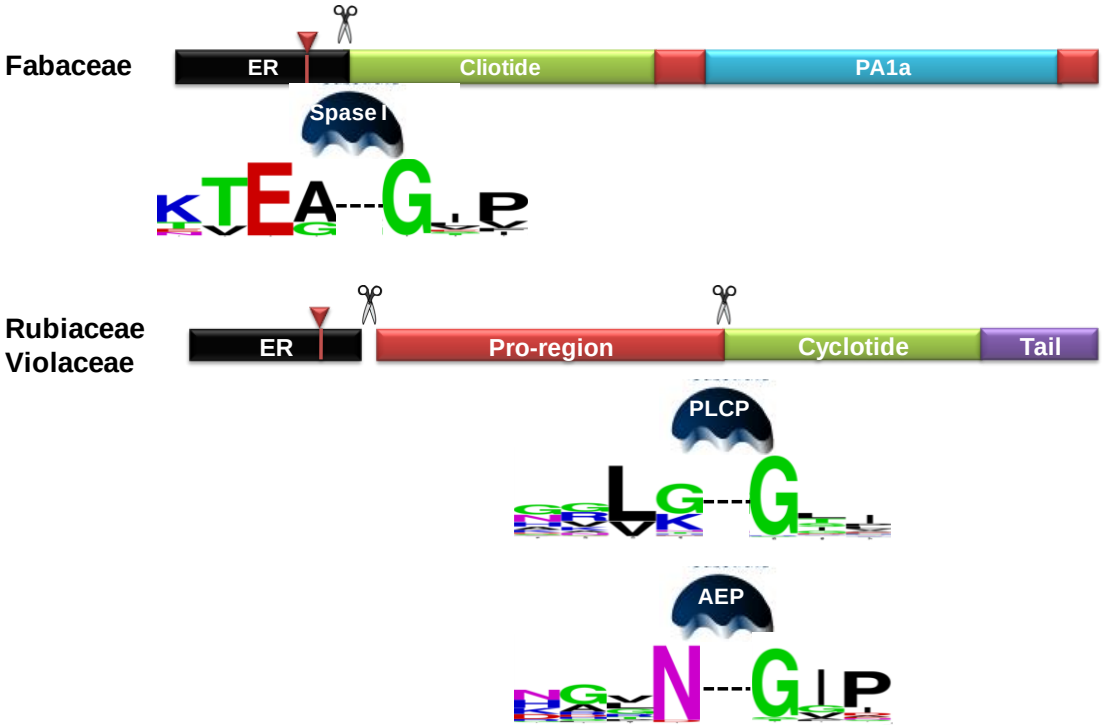


Figure 5.15. Proposed enzymes involved in N-terminal cleavage. PLCP, papain-like cysteine protease; AEP, Asparaginyl Endopeptidase. The peptides were first aligned using ClustalW program with all available sequences from CyBase and our work. Logo sequences were built with WebLogo application, version 3 (http://weblogo.berkeley.edu/) [157]. The total height of each column reflects the conservation whereas the size of each letter demonstrates the relative occurrence of that residue at a given position.

C-terminal processing

The cleavage at the C-terminus of cyclotide domain and subsequent backbone cyclization to produce mature cyclotide molecule involve the presence of AEP [139, 144]. The favorable processing site of AEP, C-terminal to Asn (occasionally Asp) [312], have been found invariably at the C-terminus of cyclotides and absent in linear variants (Figure 5.16). The C-terminal cleavage releases the hydrophobic tail and produces an acyl-enzyme intermediate, which is followed by end-to-end backbone cyclization. The detailed mechanism of cyclization still requires futher investigation but similar case reported for the mature Con A [316] suggests a combined action of AEP for peptide bond cleavage and cyclization. The energy released during C-tail excision is transiently stored in the acyl-enzyme complex and necessarily transferred to the new termini-connecting bond. It was found that truncated kalata B1 precursor without C-tail failed to produce any cyclic species *in vivo* despite the presence of the ultimate Asn and an active AEP [139]. Thus, this mechanism highlights the indispensible presence of the C-tail for successful cyclization.

A matter for consideration is to what extent the downstream hydrophobic tail is important for the C – terminal cleavage. In other words, how many residues are necessary for the C-tail excision? Gilson and his coworkers [139] proposed a model for the formation of cyclotides catalyzed by AEP [139]. In this model, they highlighted the importance of a conserved motif XXN-GLP as the recognition site of AEP during C-terminal cleavage. They also emphasized the role of the repeated GLP motif at the N-terminus for pocket binding during cyclization. The proposed mechanism implicates the requirement of at least three residues at the hydrophobic tail. However, our group has recently discovered six cyclic chassatides from *Chassalia chartaceae* with the tail of only two residues (Figure 5.16). This is the shortest hydrophobic tail flanking cyclotides reported to date. A shorter tail with only one residue apparently does not provide a qualified tail for substrate recognition of AEP as we find heydotides B1c, B15, C36, C38, and U11 in the linear form starting with G and ending with XXNE. Therefore, it is likely that a fragment of two residues is the shortest tail required for a proper cleavage and cyclization catalyzed by AEP.

In addition, both two residues were shown to contribute to the efficiency of AEP. Gly rather than Ile or Glu at P_1' position is more susceptible to *in vitro* cleavage by AEP [317]. Leu, and to a lesser extent Val/Ile at P_2' position is found conserved throughout cyclotide precursors (Figure 5.16).

In the model of Gillon *et al.*, the repeated motif GLP at the N-terminus of cyclotide swings around and fits into the binding subsites previously occupied by GLP motif of the hydrophobic tail. However, this dual function of the flanking motif is likely not the general mechanism, because

(a) N-terminal region of cyclotides and the C-terminal flanking region are rarely homologous

(b) N-terminal region of cyclotides is hyper-variable (Figure 5.16). A wide range of amino acid, either charged or uncharged, were found for cyclotides. Most noteworthy are two negatively-charged D and E residues, starting cycloviolacin 25 [318], cliotide T21, T24, T25 (Nguyen K.T.G., unpublished data) and chassatide C13 [289] in place of conventional small residue G (occasionally S)

These observations suggest a high substrate tolerance or different specificity of AEP in various cyclotide-producing species in order to generate cyclotides with diverse sequences.

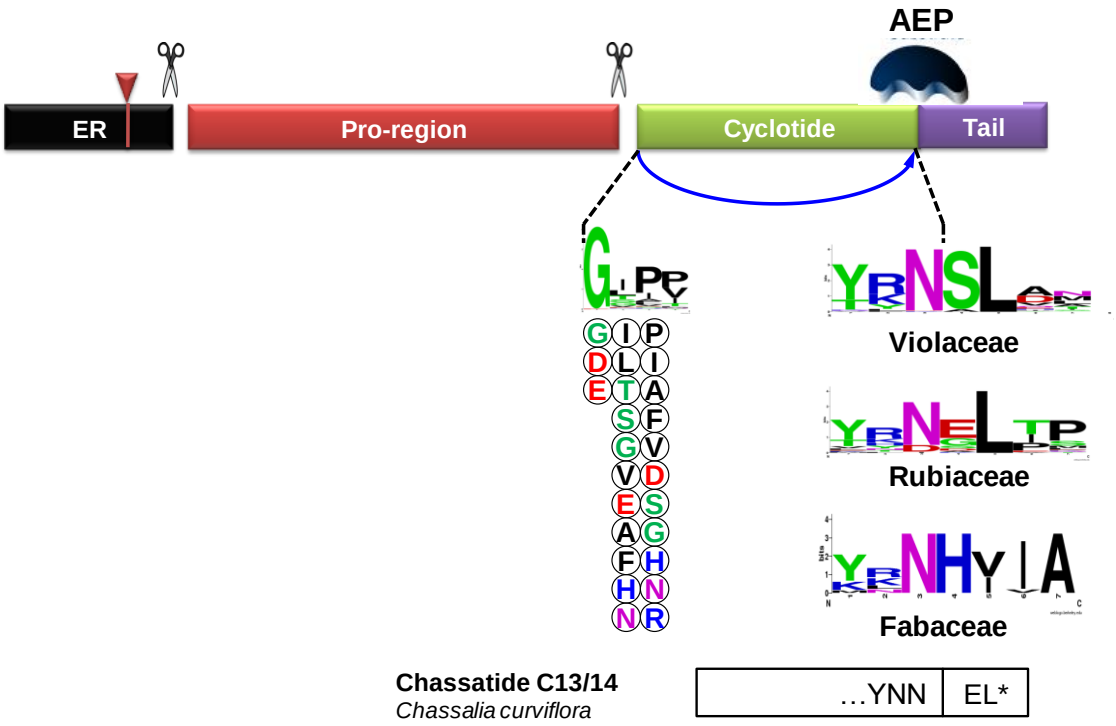


Figure 5.16. Asparaginyl endopeptidase (AEP) involved in N-terminal cleavage and cyclization

The involvement of AEP in cyclization altered the previous notion that organisms have to withstand the metabolic burden in exchange for the exceptional stability of head-to-tail cyclized backbone. For instance, bacteria employ the complex enzyme machinery for the synthesis of non-gene-encoded cyclic peptides. Examples include non-ribosomal peptide synthetases (NRPSs) for cyclosporine, mixed modular polyketide synthases (PKSs) and NRPSs for rapamycin [319, 320]. The cyclization of the MSDIN fungal toxin family is also believed to be catalyzed by a separate enzyme other than the prolyl oligopeptidase (POP) which processes their cleavage from the precursor [320]. In cyclotide-producing plants, connecting two termini could be achieved by the incorporated activity of the enzymes existing inherently for other purposes, in this case, AEP for the processing of storage protein or waste materials [139, 144]. Therefore, such utilization of the intrinsic metabolic machinery to produce molecules with superior features brings a lot of benefits, though at little cost, to the organisms along their evolution.

Fuzzy processing for promiscuous multifunctionality

The cyclotide family was originally classified comprising cyclic cystine knot peptides with high diversity in sequence. The discovery of many linear analogs highly homologous to “cyclic” cyclotides suggests a high diversity also in form of cyclotides. In this program, we observed the intriguing behavior of plants to produce cyclotide clusters diverse in both sequence and form. Within each cluster, the terminal residues, either at N- or C-terminus, are gradually trimmed in an unknown mechanism to produce a series of truncated variants. More interestingly, these variants could be both linear and/or cyclic seemingly depending on the terminal residues at both ends. Two examples of cyclotide clusters are hedyotides B4/B13/B15 and B12/B14/B30, from *H. biflora*. In the former cluster, cyclotides differ by either G or E at the termini and only one cyclic variant, which starts with G and ends with D, is found. In contrast, gradual trimming from the N-terminus produces one linear and two cyclic peptides in the latter cluster. The likelihood of these clusters being byproducts of an erroneous extraction was excluded owing to repetitive detections of such

compounds with consistent relative ratio despite multiple efforts to avoid oxidation during the extraction. Only one precursor has been found for each cluster. Together, they suggested that the peptides in each cluster are the products of intracellular processing from a single precursor.

We coined the term “fuzzy processing” for the phenomenon of plants producing a cluster of homologous cyclotides/linear cyclotides from a single precursor in a “fuzzy” manner. Such phenomenon has not been reported to date, especially the presence of two cyclic cyclotides derived from a single putative precursor or pairs of identical cyclic/linear cyclotides. The discovery of similar clusters in many species in our program raises questions toward their bioprocessing which could be far more complicated than a single “cut-and-join” event. Given that each cluster is encoded by only one gene, possible pathways for the cluster production could be at either RNA or protein level.

The possibility of producing “fuzzy products” through RNA modifications could be excluded. Two well-known mechanisms of RNA modification to produce different proteins from the same gene, namely RNA editing and alternative splicing, are apparently not involved. Uracil insertion/deletion or base deamination (C to U or A to I editing) in RNA editing causes dramatic reading frame shift, leading to different translated proteins. The members of a single-gene-derived cluster are, however, identical except at the two termini. In alternative splicing, exons are reconnected in multiple ways during pre-mRNA splicing to produce various proteins from a single gene. This mechanism requires multiple introns in a single gene. The complete absence or the presence of only one intron in cyclotide precursors eliminates the occurrence of alternative splicing in cyclotide biosynthesis.

Hence, post-translational processing is likely the mechanism for the production of cyclotide clusters differing in forms and sequences. Figure 5.17 depicts the sequences at the two termini of every cluster in this project with highlighted residues found in cyclic variants. There are two groups of linear cyclotides based on the existence of their putative cyclic analogues:

(a) Group 1 having the necessary features for cyclization. This group is characterized by the ultimate residue being N or D and the first residue being G, S, D or rarely E. So far, only one cyclotide is found to have E at terminus–connecting bond; other sequences are linear cyclotides.

(b) Group 2 with the apparent lack of necessary motif for an efficient catalysis by AEPs. They include all the members that fail to possess the above critical residues at two ends.

To explain the cellular origin of these two groups, we hypothesize the fuzzy processing to comprise two steps: fuzzy cleavage at N- and C-termini. The proposed N-terminal fuzzy cleavage occurs during or immediately after the breaking of scissile bonds at N-terminus. Along the secretory pathway, the precursors undergo the signal peptide excision, folding process, and pro-region cleavage before their interactions with AEP. During the interval between the N-terminal and C-terminal cleavages, an opportunistic enzyme gradually trims pro-cyclotide from both ends. If the truncated products still retain the required features for cyclization including the critical N-terminal residues and a conserved hydrophobic tail of minimum two residues, they are cyclized by AEP activity. Otherwise, when the fuzzy trimming cleaves beyond the critical residues, the cyclotide domain loses the essential features for cyclization. They remain in the linear form or continue to be further processed.

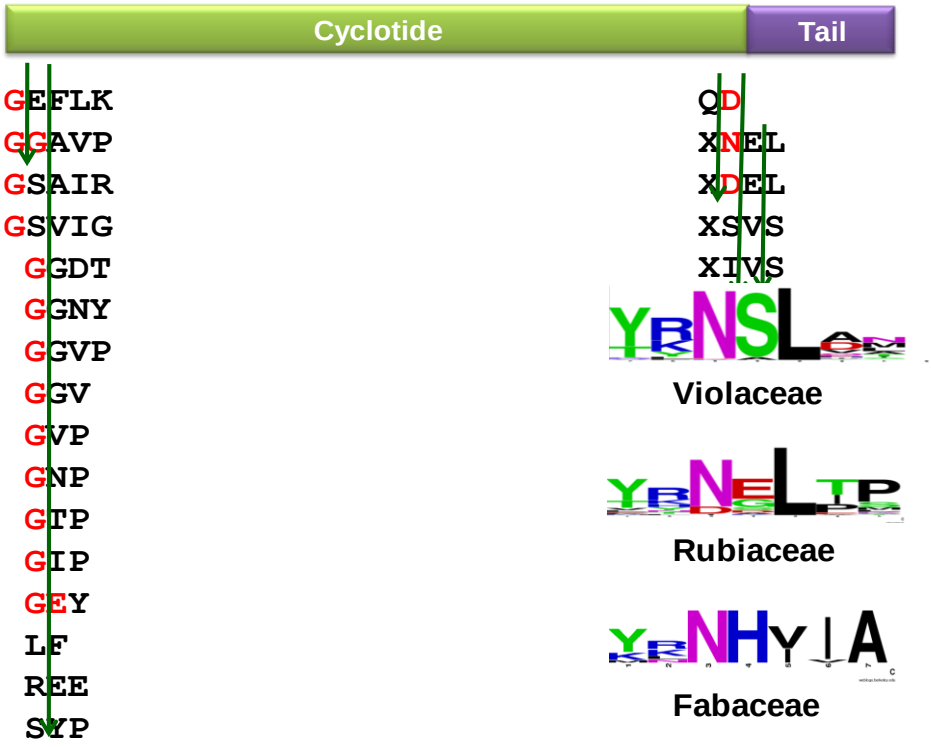


Figure 5.17. The sequences at two extremes of uncyclotides. The arrow represents the hydrolysis activity of unknown enzymes to release a series of cyclotides/linear cyclotides differing by one or two residues at two termini. The letters in red mark the sequences where we can find corresponding cyclic variants. They highlight the requirements of cyclotide primary sequences to produce a cyclic structure.

Evolutionary mechanism of cyclotides

The evolutionary history of cyclotides in the plant kingdom has not been fully understood. However, sequence analysis of available precursors from different species and families suggests that the cyclotide family may evolve through a series of duplication and mutation events. The whole gene is first multiplicated by chance to form paralogs. The newly formed copies then accumulate mutations at a higher rate owing to the lack of selective pressure - that means they have no or little deleterious effect to the organisms. Every element in those copies mutates at its own rate. This, after many generations, produces a family with diverse members different from each other by one or many gene elements. The evolution through duplication and mutation explains several conservations observed among cyclotide precursors: first, the existence of many signal peptides in a particular species, each shared by a few cyclotides; second, the expression of several common cyclotides in distantly related species, for example kalata B1 in *Oldenlandia affinis* (Rubiaceae), *Viola odorata* and *Viola yedoensis* (Violaceae) [101, 124, 318]; and third, the strikingly almost complete homology between hedyotides B1 and B2 introns or hedyotides U26 and U27 precursors. The earliest duplication and mutation events of the ancestral cyclotide genes perhaps occurred before the divergence of monocot and dicot groups, because cyclotides have been found in both monocot (Poaceae) and dicot families (Rubiaceae, Violaceae, Fabaceae, and Solanaceae).

There are three main mechanisms for gene duplication described so far, namely chromosomal duplication, unequal crossing-over, and retroposition [321]. Making a new copy of the whole genome in chromosomal duplication is obviously not suitable for hundreds of cyclotide-producing plants. Between unequal crossing-over and retroposition pathways, our discovery on cyclotides suggests the former. While unequal crossing-over is a recombination DNA event in which two parental DNA molecules exchange nonidentical fragments, retroposition is the insertion of a complementary DNA (cDNA), after being retrotranscribed from messenger RNA, into the genome in a random manner.

The sharp contrast in the outcome for retroposition is the loss of introns and of regulatory sequences as well as the retention of poly A tracts in the duplicated gene. Accordingly, given the presence of introns in Rubiaceae and Fabaceae precursors, it is likely that cyclotide gene duplication occurs via unequal crossing-over.

A notable finding in this project is the discovery of cyclotide, in place of the defense pea albumin chain b (PA1b), fused with the storage protein pea albumin chain a (PA1a) in a single precursor. There are two main mechanisms that could explain for the appearance of this intriguing gene architecture.

The first hypothesis supports a recent horizontal gene transfer (HGT) event, also known as lateral gene transfer. In contrast to vertical gene transmission from parents to offsprings, HGT refers to an exchange of genetic information between distantly related species by asexual means [322, 323], in this case from a cyclotide-producing plant to *Clitoria ternatea*. The theory is supported by several points emerged from the studies of mitochondria HGT [322].

(a) Most HGT cases puplished so far are recent events, being limited to a small subset of plant genus. The expression of cyclotides, though strikingly high in *C. ternatea*, was undetectable in *Clitoria fairchildiana* and *Clitoria laurifolia* (Nguyen K.T.G., unpublished data).

(b) Some HGTs appeared to be functional homolog replacement. In fact PA1b possesses the same cystine knot motif with comparable loop sizes and also exhibits insecticidal activity [324].

(c) A few transferred genes are chimeric combination of the foreign and native copies. Cyclotide precursors are recombined from pea albumin chain a and cyclotide in the same reading frame.

(d) The study of a *Drosophila* chimeric gene, jingwei, suggests that retrotransposition could be a mechanism to generate a new intron-containing functional gene [325]. The processed messenger RNA can capture upstream

regions, including exons and introns, of an unrelated gene. This could be the mechanism underlying HGT event in *C. ternatea* if it did occur.

However, one may argue that the chance for an in-frame and precise insertion of cyclotide in place of PA1b is low. Such an occurence seems unreasonable when considering the multi-member nature of both pea albumin and cyclotide families which may call for massive HGTs. Besides, in the plant kingdom the event is reported to occur with highest frequency for mitochondrial genes, possibly because of their high propensity to uptake exogenous double stranded DNA [322]. With small coding region, mitochondria hardly accommodate of >40 cyclotides as in *H. chrysotricha* and *H. uncinella*. Meanwhile, the exchange of nuclear genome information was only found to be mediated by phagocytosis of plastid or via transposable element activity [326]. However, the insertion of solely cyclotide domain into pea albumin precursor almost excludes the involvement of a transposable element.

The second theory focuses on the hypothesis of convergent evolution. An interesting example has recently been reported for the platypus. 83 novel putative venom genes identified from platypus showed striking homology to those found in a wide variety of vertebrates (fish, reptiles) and invertebrates (spider, starfish) [327]. In the constraints of the peptide size, the collection of primary sequences and structural topologies are limited. Species evolving from distant branches therefore end up in developing similar families of molecules for the same functions. The process probably occurs with the aid of natural selection to optimize the protection barriers. PA1b and cyclotides have backbones of similar size and constrained by the cystine knot motif. Both of them exhibit many biological functions. Therefore, PA1b and cyclotides may be the products of convergent evolution in *C. ternatea* and rubiaceous and violaceaous species.

Summary, Conclusion and Future Outlook

Cystine knot miniproteins exist in abundance among the plant kingdom. They are highly diverse and display multiple biological functions, particularly defense-related activities. In this thesis, the molecular and functional diversity of two classes of plant CK miniproteins was studied by proteomic and genomic methods. They include α-amylase inhibitors from the Apocynaceae and cyclotides from the Rubiaceae, Fabaceae, and Poaceae families.

Chapter 3 describes the isolation and characterization of a family of 30-amino acid-residues α-amylase inhibitors from three Apocynaceae species, namely *Allamanda oenotheraefolia, Alstonia scholaris*, and *Wrightia religiosa*. A total of 20 peptides were purified and sequenced from these three medicinal plants. Their knotted topology of disulfide bonds was confirmed by top-down disulfide mapping of two different peptides, namely allatide O2 and alstotide S3 from *A. oenotheraefolia* and *A. scholaris*, respectively. The 20 identified peptides share high sequence identity to each other and to AAI, the prototype Amaranth α-amylase inhibitor. They exhibited exceptional thermal and proteolytic stability and possess inhibitory property against insect α-amylase. Together, our results unequivocally prove that these peptides are CK α-amylase inhibitors. This finding is significant because there is only one member of the CK α-amylase inhibitor family known so far, which is AAI. Furthermore, by small-scale extraction and MS, we also showed the potential existence of CK α-amylase inhibitors in another five Apocynaceae species out of six tested, implicating the wider presence of this family in the Apocynaceae family. Careful protein engineering on these peptide backbones may render them active against human α-amylases and open the venue of their applications in diabetes and obesity treatment.

Among the identified CK α-amylase inhibitors, 14 sequences have been confirmed by genomic method. Their transcripts reveals a three-domain precursor of Apocynaceae CK α-amylase inhibitors, comprising an ER signal peptide, a pro-domain followed by the mature peptide domain. Their

biosynthesis is ribosomally processed, ER-targeted, and follows the secretory protein pathway. This understanding of the biosynthesis is useful for future application of CK α-amylase inhibitors in crop protection using transgenic plants.

One of the major findings in this thesis is the dual function of alstotides S1 and S3 as α-amylase inhibitor and antiviral reagent. Chapter 4 presents the study of antiviral mechanism of alstotide S1 against IBV, a worldwide threat of domestic fowl. Alstotide S1 acts mainly during the early stage of viral replication cycle and interacts with S fusion protein and M protein of IBV. The proposed mechanism is that alstotide S1 binds to S protein and blocks the S protein-mediated fusion between virus and cell membranes during viral entry. Interaction of alstotide S1 and M protein may elicit further inhibition effect during assembly, budding, and maturation processes. Alstotide S1 also inhibits the replication of DENV but not RSV. Although the antiviral mechanism against DENV remains unresolved, alstotide S1 may act through the endocytosis of DENV entry. The endocytosis during viral entry is shared between DENV and IBV but not RSV, which is possibly the main cause of the differential effect of alstotide S1 on these viruses. Interestingly, alstotide S1 was shown to penetrate the cell membrane and accumulate in the perinuclear region, potentiating its application as a cell-penetrating carrier.

Chapter 5 reports on the study of the hyperdiversity the cyclotide class of CK miniproteins through genomic method. We studied their biosynthesis and identified 101 genes of cyclotides from eight species belonging to three plant families, >60 of which were derived from this project. Cyclotide precursors may have different structures among plant families but their precursors are quite conserved within family. Sequence analysis at the processing sites of the precursors reveal potential promiscuous activity of processing enzymes as a mechanism to produce multiple protein products from a single transcript. This phenomenon is named “fuzzy processing”. Together with the multi-gene nature and multi-domain precursor structure of cyclotide family, fuzzy processing provides the hyperdiversity of cyclotides, significantly broadening the defense

spectrum and conferring the host plants better adaptation at minimum genetic cost.

Two ongoing projects have been developed following the work in this thesis: transcriptome sequencing in search of CRPs from plants and rational design of human α-amylase inhibitors. Pyrosequencing service recently available at low cost allows for high-throughput and efficient profiling of CRPs in the first project. At the same cost, transcriptome analysis by pyrosequencing reveals full profiles of CRPs after one run as compared to the laborious, multiple-round cloning using degenerate primers and RACE PCR. In the second project, structural studies by NMR and X-ray crystallography have been performed to determine 3D structures of CK α-amylase inhibitors in solution and in its complex with TMA. Comparison between these structures and known proteinaceous human α-amylase inhibitors at the active site of α-amylases suggests critical residues important for their differential inhibitory activities. Based on this knowledge, computer-aided molecular docking of inhibitors to human α-amylases may propose rationally designed, potential candidates of CK human α-amylase inhibitors. These candidates will be chemically synthesized by solid phase peptide synthesis and tested in *in vitro* and *in vivo* assays.

In conclusion, this thesis provides insights into the molecular and functional hyperdiversity of CK miniproteins through two family models: CK α-amylase inhibitor and cyclotide. CK miniproteins exist widely in nature and are a rich goldmine of pharmacological compounds for medicinal applications and drug development. However, only a small fraction of natural CK miniproteins has been studied so far. For example, about 0.1% of conotoxins, a large and highly diverse peptide family, has been pharmaceutically characterized [328]. This study therefore broadens our knowledge about the biosynthesis and functions of this family and facilitates their future applications in many aspects.

Publications

1. Nguyen, G. K. T., Zhang, S., Nguyen, N. T. K., **Nguyen, P. Q. T.**, Chiu, M. S., Hardjojo, A. and Tam, J. P. (2011) Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family. J. Biol. Chem. 286, 24275-24287

2. Nguyen, G. K., Lim, W. H., **Nguyen, P.Q.T.** and Tam, J. P. (2012) Novel cyclotides and uncyclotides with highly shortened precursors from *Chassalia chartacea* and effects of methionine oxidation on bioactivities. J Biol Chem. 287, 17598-17607

3. Nguyen, G. K., Lian, Y., Pang, E. W., **Nguyen, P.Q.T.**, Tran, T. D. and Tam, J. P. (2012) Discovery of Linear Cyclotides in Monocot Plant *Panicum laxum* of Poaceae Family Provides New Insights into Evolution and Distribution of Cyclotides in Plants. The Journal of biological chemistry

4. Luu, T.T.*, **Nguyen, P.Q.T.*,** Nguyen, G. K. T. and Tam, J. P. (2012) Allatides: smallest proteinaceous α-amylase inhibitors from *Allamanda oenotheraefolia* and their expression in Apocynaceae plants (In manuscript)

5. **Nguyen, P.Q.T.*,** Luu, T.T.* and Tam, J. P. (2012) Discovery of bifunctional antiviral/α-amylase inhibitors from medicinal plant *Alstonia scholaris* (In manuscript)

6. **Nguyen, P.Q.T.,** Teo, C.H., Lam, Y. and Tam, J. P. (2012) Discovery and characterization of thermostable cystine knot α-amylase inhibitors from *Wrightia religiosa* (In manuscript)

Poster presentation

The 7th Annual Peptide Therapeutic Symposium, organized by The Peptide Therapeutics Foundation, at La Jolla, San Diego, USA on 25-26 Oct, 2012. Title: Discovery and Characterization of Small Proteinaceous α-Amylase Inhibitors from *Allamanda oenotheraefolia*

References

1. Showalter, A.M., *Structure and Function of Plant Cell Wall Proteins.* Plant Cell, 1993. **5**(1): p. 9-23.
2. Manners, J., *Hidden weapons of microbial destruction in plant genomes.* Genome Biol., 2007. **8**(9): p. 225.
3. Rates, S.M.K., *Plants as source of drugs.* Toxicon, 2001. **39**(5): p. 603-613.
4. Gurib-Fakim, A., *Medicinal plants: Traditions of yesterday and drugs of tomorrow.* Mol. Aspects Med., 2006. **27**(1): p. 1-93.
5. Belting, M. and A. Wittrup, *Developments in Macromolecular Drug Delivery*, in *Macromolecular Drug Delivery*, M. Belting, Editor 2009, Humana Press. p. 1-10.
6. Shaji, J. and V. Patole, *Protein and peptide drug delivery: Oral approaches*. Vol. 70. 2008. 269-277.
7. Lay, F.T. and M.A. Anderson, *Defensins - components of the innate immune system in plants.* Curr. Prot. Pept. Sci., 2005. **6**: p. 85 - 101.
8. Colgrave, M.L. and D.J. Craik, *Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot†.* Biochem., 2004. **43**(20): p. 5965-5975.
9. Silverstein, K.A.T., et al., *Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants.* Plant J., 2007. **51**(2): p. 262-280.
10. Hammami, R., et al., *PhytAMP: a database dedicated to antimicrobial plant peptides.* Nucleic Acids Res., 2009. **37**(suppl 1): p. D963-D968.
11. Silverstein, K.A.T., et al., *Genome Organization of More Than 300 Defensin-Like Genes in Arabidopsis.* Plant Physiol., 2005. **138**(2): p. 600-610.
12. Terras, F.R., et al., *Analysis of two novel classes of plant antifungal proteins from radish (Raphanus sativus L.) seeds.* J. Biol. Chem., 1992. **267**(22): p. 15301-15309.
13. Ponz, F., et al., *Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin α-hordothionin.* Eur. J. Biochem., 1986. **156**(1): p. 131-135.
14. Cammue, B.P.A., et al., *A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins.* Plant Physiol., 1995. **109**: p. 445 - 455.
15. Broekaert, W.F., et al., *Antimicrobial Peptides from Amaranthus caudatus Seeds with Sequence Homology to the Cysteine/Glycine-Rich Domain of Chitin-Binding Proteins.* Biochem. J., 1992. **31**: p. 4308-4314.
16. Cammue, B.P., et al., *Isolation and characterization of a novel class of plant antimicrobial peptides from Mirabilis jalapa L. seeds.* J. Biol. Chem., 1992. **267**: p. 2228 - 2233.
17. Patel, S.U., et al., *Structural Studies of Impatiens balsamina Antimicrobial Protein (Ib-AMP1).* Biochem., 1998. **37**(4): p. 983-990.

18. Tailor, R.H., et al., *A novel family of small cysteine-rich antimicrobial peptides from seed of Impatiens balsamina is derived from a single precursor protein.* J. Biol. Chem., 1997. **272**: p. 24480 - 24487.
19. Simonsen, S.M., et al., *A continent of plant defense peptide diversity: Cyclotides in Australian Hybanthus (Violaceae).* Plant Cell, 2005. **17**(11): p. 3176-3189.
20. Marcus, J.P., et al., *Purification, Characterisation and cDNA Cloning of an Antimicrobial Peptide from Macadamia Integrifolia.* Eur. J. Biochem., 1997. **244**(3): p. 743-749.
21. Berrocal-Lobo, M., et al., *Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection.* Plant Physiol., 2002. **128**: p. 951 - 961.
22. Craik, D.J., N.L. Daly, and C. Waine, *The cystine knot motif in toxins and implications for drug design.* Toxicon, 2001. **39**(1): p. 43-60.
23. Kolmar, H., *Alternative binding proteins: Biological activity and therapeutic potential of cystine-knot miniproteins.* FEBS J., 2008. **275**(11): p. 2684-2690.
24. Pallaghy, P., et al., *A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides.* Protein Sci., 1994. **3**(10): p. 1833-1839.
25. Isaacs, N.W., *Cystine knots.* Curr. Opin. Struct. Biol., 1995. **5**(3): p. 391-5.
26. Cheek, S., S.S. Krishna, and N.V. Grishin, *Structural classification of small, disulfide-rich protein domains.* J. Mol. Biol., 2006. **359**(1): p. 215-37.
27. McDonald, N.Q. and W.A. Hendrickson, *A structural superfamily of growth factors containing a cystine knot motif.* Cell, 1993. **73**(3): p. 421-424.
28. Chiche, L., et al., *Use of restrained molecular dynamics in water to determine three-dimensional protein structure: Prediction of the three-dimensional structure of Ecballium elaterium trypsin inhibitor II.* Proteins: Structure, Function, and Bioinformatics, 1989. **6**(4): p. 405-417.
29. Davis, J.H., et al., *Solution structure of .omega.-conotoxin GVIA using 2-D NMR spectroscopy and relaxation matrix analysis.* Biochem., 1993. **32**(29): p. 7396-7405.
30. Pallaghy, P.K., et al., *A common structural motif incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides.* Protein Sci., 1994. **3**(10): p. 1833-1839.
31. Le Nguyen, D., et al., *Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure.* Biochimie, 1990. **72**(6–7): p. 431-435.
32. Craik, D.J., et al., *Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif.* J. Mol. Biol., 1999. **294**(5): p. 1327-1336.
33. Gracy, J., et al., *KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007.* Nucleic Acids Res., 2008. **36**(suppl 1): p. D314-D319.

34. Hill, J.M., P.F. Alewood, and D.J. Craik, *Solution structure of the sodium channel antagonist conotoxin GS: a new molecular caliper for probing sodium channel geometry.* Structure, 1997. **5**(4): p. 571-83.
35. Hill, J.M., P.F. Alewood, and D.J. Craik, *Three-Dimensional Solution Structure of μ-Conotoxin GIIIB, a Specific Blocker of Skeletal Muscle Sodium Channels.* Biochem., 1996. **35**(27): p. 8824-8835.
36. Avsian-Kretchmer, O. and A.J.W. Hsueh, *Comparative Genomic Analysis of the Eight-Membered Ring Cystine Knot-Containing Bone Morphogenetic Protein Antagonists.* Mol. Endo., 2004. **18**(1): p. 1-12.
37. Kolmar, H., *Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins.* Curr. Opin. Pharmacol., 2009. **9**(5): p. 608-14.
38. Bluhm, B.H. and C.P. Woloshuk, *Amylopectin Induces Fumonisin B1 Production by Fusarium verticillioides During Colonization of Maize Kernels.* Mol. Plant-Microbe Interact., 2005. **18**(12): p. 1333-1339.
39. Fakhoury, A.M. and C.P. Woloshuk, *Inhibition of Growth of Aspergillus flavus and Fungal α-Amylases by a Lectin-Like Protein from Lablab purpureus.* Mol. Plant-Microbe Interact., 2001. **14**(8): p. 955-961.
40. Kim, Y.-H., et al., *Cloning and functional expression of the gene encoding an inhibitor against Aspergillus flavus α-amylase, a novel seed lectin from Lablab purpureus.* Plant Cell Reports, 2007. **26**(4): p. 395-405.
41. Kim, M.J., et al., *Comparative study of the inhibition of alpha-glucosidase, alpha-amylase, and cyclomaltodextrin glucanosyltransferase by acarbose, isoacarbose, and acarviosine-glucose.* Arch. Biochem. Biophys., 1999. **371**(2): p. 277-83.
42. Gruber, C.W., et al., *Insecticidal plant cyclotides and related cystine knot toxins.* Toxicon, 2007. **49**(4): p. 561-575.
43. Lee, S.C., P.L. Gepts, and J.R. Whitaker, *Protein structuRes. of common bean (Phaseolus vulgaris) alpha-amylase inhibitors.* J. Agric. Food Chem., 2002. **50**(22): p. 6618-27.
44. Barrett, M. and J. Udani, *A proprietary alpha-amylase inhibitor from white bean (Phaseolus vulgaris): A review of clinical studies on weight loss and glycemic control.* Nutrition J., 2011. **10**(1): p. 24.
45. Prakash, O. and N. Jaiswal, *-Amylase Inhibition: Some Practical Considerations.* Curr. Enz. Inh., 2010. **6**(3): p. 158-163.
46. Morton, R.L., et al., *Bean α-amylase inhibitor 1 in transgenic peas (Pisum sativum) provides complete protection from pea weevil (Bruchus pisorum) under field conditions.* Proc. Natl. Acad. Sci. U.S.A., 2000 **97**(8).
47. Hanefeld, M. and F. Schaper, *The Role of Alpha-Glucosidase Inhibitors (Acarbose) Pharmacotherapy of Diabetes: New Developments*, C.E. Mogensen, Editor 2007, Springer US. p. 143-152.
48. Hanefeld, M., et al., *Therapeutic potentials of acarbose as first-line drug in NIDDM insufficiently treated with diet alone.* Diabetes Care, 1991. **14**(8): p. 732-7.
49. Le Berre-Anton, V., et al., *Characterization and functional properties of the alpha-amylase inhibitor (alpha-AI) from kidney bean (Phaseolus vulgaris) seeds.* Biochim. Biophys. Acta, 1997. **14**(1): p. 31-40.

50. Barrett, M. and J. Udani, *A proprietary alpha-amylase inhibitor from white bean (Phaseolus vulgaris): A review of clinical studies on weight loss and glycemic control.* Nutrition J., 2011. **10**(1): p. 24.
51. Chrzaszcz, T. and J. Janicki, *Sisto-amylase", ein naturlicher paralysator der amylase.* Biochem. Zeitschr., 1933. **260**: p. 354-369.
52. Franco, O., et al., *Plant alpha-amylase inhibitors and their interaction with insect alpha-amylases Structure, function and potential for crop protection.* Eur. J.Biochem. / FEBS, 2002. **269**: p. 397 - 412.
53. Svensson, B., et al., *Proteinaceous α-amylase inhibitors.* Biochim. Biophys. Acta (BBA) - Proteins & Proteomics, 2004. **1696**(2): p. 145-156.
54. Whitaker John, R., *Protease and α-Amylase Inhibitors of Higher Plants*, in *Antinutrients and Phytochemicals in Food* 1997, American Chemical Society. p. 10-30.
55. Svensson, B., et al., *Proteinaceous alpha-amylase inhibitors.* Biochim. Biophys. Acta (BBA) - Proteins & Proteomics, 2004. **1696**(2): p. 145-156.
56. Franco, O.L., et al., *Activity of wheat α-amylase inhibitors towards bruchid α-amylases and structural explanation of observed specificities.* Eur. J. Biochem., 2000. **267**(8): p. 2166-2173.
57. Schimoler-O'Rourke, R., M. Richardson, and C.P. Selitrennikoff, *Zeamatin inhibits trypsin and alpha-amylase activities.* Appl. Environ. Microbiol., 2001. **67**(5): p. 2365-6.
58. Oda, Y., et al., *Tertiary and Quaternary StructuRes. of 0.19 α-Amylase Inhibitor from Wheat Kernel Determined by X-ray Analysis at 2.06 Å Resolution.* Biochem., 1997. **36**(44): p. 13503-13511.
59. Young, N.M., et al., *Post-translational processing of two alpha-amylase inhibitors and an arcelin from the common bean, Phaseolus vulgaris.* FEBS Lett., 1999. **446**(1): p. 203-6.
60. Sanchez-Monge, R., et al., *Wheat and barley allergens associated with baker's asthma. Glycosylated subunits of the alpha-amylase-inhibitor family have enhanced IgE-binding capacity.* Biochem. J., 1992. **281 (Pt 2)**: p. 401-5.
61. Garcia-Casado, G., et al., *Role of complex asparagine-linked glycans in the allergenicity of plant glycoproteins.* GlycoBiol., 1996. **6**(4): p. 471-7.
62. Chagolla-Lopez, A., et al., *A novel alpha-amylase inhibitor from amaranth (Amaranthus hypocondriacus) seeds.* J. Biol. Chem., 1994. **269**(38): p. 23675-23680.
63. Pereira, P.J.B., et al., *Specific inhibition of insect α-amylases: yellow meal worm α-amylase in complex with the Amaranth α-amylase inhibitor at 2.0 Å resolution.* Structure, 1999. **7**(9): p. 1079-1088.
64. Nitti, G., et al., *Amino Acid Sequence and Disulphide-bridge Pattern of three γ-Thionins from Sorghum bicolor.* Eur. J. Biochem., 1995. **228**(2): p. 250-256.
65. Diz, M.S., et al., *Characterisation, immunolocalisation and antifungal activity of a lipid transfer protein from chili pepper (Capsicum annuum) seeds with novel α-amylase inhibitory properties.* Physiol. Plant., 2011. **142**(3): p. 233-246.

66. Vallee, F., et al., *Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 A resolution.* Structure, 1998. **6**(5): p. 649-59.
67. Gvozdeva, E.L., T.A. Valueva, and V.V. Mosolov, *Enzymatic oxidation of the bifunctional wheat inhibitor of subtilisin and endogenous alpha-amylase.* FEBS Lett., 1993. **334**(1): p. 72-4.
68. Ohtsubo, K. and M. Richardson, *The amino acid sequence of a 20 kDa bifunctional subtilisin/alpha-amylase inhibitor from bran [correction of brain] of rice (Oryza sativa L.) seeds.* FEBS Lett., 1992. **309**(1): p. 68-72.
69. Blanco-Labra, A. and F.A. Iturbe-ChiÑAs, *Purification and Characterization of an α-Amylase Inhibitor from Maize (Zea Maize).* J. Food Biochem., 1981. **5**(1): p. 1-17.
70. Carugo, O., et al., *Structural analysis of free and enzyme-bound amaranth α-amylase inhibitor: classification within the knottin fold superfamily and analysis of its functional flexibility.* Protein Eng., 2001. **14**(9): p. 639-646.
71. Lu, S., et al., *Solution Structure of the Major α-Amylase Inhibitor of the Crop Plant Amaranth.* J. Biol. Chem., 1999. **274**(29): p. 20473-20478.
72. Mittl, P.R., et al., *A new structural class of serine protease inhibitors revealed by the structure of the hirustasin-kallikrein complex.* Structure, 1997. **5**(2): p. 253-264.
73. Usón, I., et al., *The 1.2 Å crystal structure of hirustasin reveals the intrinsic flexibility of a family of highly disulphide-bridged inhibitors.* Structure, 1999. **7**(1): p. 55-63.
74. Prakash, O. and N. Jaiswal, *α-Amylase Inhibition: Some Practical Considerations.* Curr. Enz. Inhib., 2010. **6**(3): p. 158-163.
75. Maskos, K., M. Huber-Wunderlich, and R. Glockshuber, *RBI, a one-domain alpha-amylase/trypsin inhibitor with completely independent binding sites.* FEBS Lett., 1996. **397**(1): p. 11-6.
76. Schimoler-O'Rourke, R., M. Richardson, and C.P. Selitrennikoff, *Zeamatin inhibits trypsin and alpha-amylase activities.* Appl. Environ. Microbiol., 2001. **67**(5): p. 2365-6.
77. Lin, Y.H., et al., *Purification, crystallization and preliminary X-ray crystallographic analysis of rice bifunctional alpha-amylase/subtilisin inhibitor from Oryza sativa.* Acta crystallographica. Section F, Structural Biol. and crystallization communications, 2006. **62**(Pt 8): p. 743-5.
78. Yamasaki, T., et al., *Rice bifunctional alpha-amylase/subtilisin inhibitor: cloning and characterization of the recombinant inhibitor expressed in Escherichia coli.* Biosci., Biotech., Biochem., 2006. **70**(5): p. 1200-9.
79. Yamagata, H., et al., *Rice bifunctional alpha-amylase/subtilisin inhibitor: characterization, localization, and changes in developing and germinating seeds.* Biosci., Biotech., Biochem., 1998. **62**(5): p. 978-85.
80. Ary, M.B., M. Richardson, and P.R. Shewry, *Purification and characterization of an insect alpha-amylase inhibitor/endochitinase from seeds of Job's Tears (Coix lachryma-jobi).* Biochim. Biophys. Acta, 1989. **999**(3): p. 260-6.
81. Dayler, C.S., et al., *Identification of a novel bean alpha-amylase inhibitor with chitinolytic activity.* FEBS Lett., 2005. **579**(25): p. 5616-20.

82. Maskos, K., M. Huber-Wunderlich, and R. Glockshuber, *RBI, a one-domain alpha-amylase/trypsin inhibitor with completely independent binding sites.* FEBS Lett., 1996. **397**(1): p. 11-6.
83. Gran, L., *Isolation of oxytocic peptides from Oldenlandia affinis by solvent extraction of tetraphenylborate complexes and chromatography on sephadex LH-20.* Lloydia, 1973. **36**(2): p. 207-8.
84. Gran, L., *On the effect of a polypeptide isolated from "Kalata-Kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus.* Acta Pharmacol Toxicol (Copenh), 1973. **33**(5): p. 400-8.
85. Gran, L., F. Sandberg, and K. Sletten, *Oldenlandia affinis (R&S) DC. A plant containing uteroactive peptides used in African traditional medicine.* J. Ethnopharmacol., 2000. **70**(3): p. 197-203.
86. Daly, N.L., K.J. Rosengren, and D.J. Craik, *Discovery, structure and biological activities of cyclotides.* Advanced Drug Delivery Reviews, 2009. **61**(11): p. 918-930.
87. Craik, D.J., *Discovery and applications of the plant cyclotides.* Toxicon, 2010. **56**(7): p. 1092-1102.
88. Saether, O., et al., *Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1.* Biochem., 1995. **34**(13): p. 4147-58.
89. Tam, J.P. and Y.A. Lu, *A biomimetic strategy in the synthesis and fragmentation of cyclic protein.* Protein science: A publication of the Protein Society, 1998. **7**(7): p. 1583-92.
90. Trabi, M., H.J. Schirra, and D.J. Craik, *Three-dimensional structure of RTD-1, a cyclic antimicrobial defensin from Rhesus macaque leukocytes.* Biochem., 2001. **40**(14): p. 4211-21.
91. Pelegrini, P.B., B.F. Quirino, and O.L. Franco, *Plant cyclotides: an unusual class of defense compounds.* Peptides, 2007. **28**(7): p. 1475-81.
92. Saska, I., et al., *Quantitative analysis of backbone-cyclised peptides in plants.* J. chromatography. B, Analytical technologies in the biomedical and life sciences, 2008. **872**(1-2): p. 107-14.
93. Jennings, C.V., et al., *Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Mobius strips exist in nature?* Biochem., 2005. **44**(3): p. 851-60.
94. Felizmenio-Quimio, M.E., N.L. Daly, and D.J. Craik, *Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis.* J. Biol. Chem., 2001. **276**(25): p. 22875-82.
95. Robbrecht, E., J.-F. Manen, and ois, *The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of Coptosapelta and Luculia, and supertree construction based on rbcL, rps16, trnL-trnF and atpB-rbcL data. A new classification in two subfamilies, Cinchonoideae and Rubioideae.* Systematics and Geography of Plants, 2006. **76**(1): p. 85-146.
96. Tokuoka, T., *Molecular phylogenetic analysis of Violaceae (Malpighiales) based on plastid and nuclear DNA sequences.* J. Plant Res., 2008. **121**(3): p. 253-260.

97. Gruber, C.W., et al., *Distribution and Evolution of Circular Miniproteins in Flowering Plants.* Plant Cell Onl., 2008. **20**(9): p. 2471-2483.
98. Hernandez, J.F., et al., *Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure.* Biochem., 2000. **39**(19): p. 5722-30.
99. Heitz, A., et al., *Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins.* Biochem., 2001. **40**(27): p. 7973-83.
100. Mylne, J.S., et al., *Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis.* Plant Cell, 2012. **24**(7): p. 2765-78.
101. Gran, L., *Oxytocic principles of Oldenlandia affinis.* Lloydia, 1973. **36**(2): p. 174-8.
102. Qin, Q., et al., *Identification of candidates for cyclotide biosynthesis and cyclisation by expressed sequence tag analysis of Oldenlandia affinis.* BMC Genomics, 2010. **11**: p. 111.
103. Gruber, C.W., *Global cyclotide adventure: a journey dedicated to the discovery of circular peptides from flowering plants.* Biopol., 2010. **94**(5): p. 565-72.
104. Mulvenna, J.P., et al., *Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins?* Plant Cell, 2006. **18**(9): p. 2134-44.
105. Nguyen, G.K., et al., *Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family.* J. Biol. Chem., 2011. **286**(27): p. 24275-87.
106. Poth, A.G., et al., *Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins.* ACS Chem. Biol., 2011. **6**(4): p. 345-55.
107. Blond, A., et al., *The cyclic structure of microcin J25, a 21-residue peptide antibiotic from Escherichia coli.* Eur. J. Biochem. / FEBS, 1999. **259**(3): p. 747-55.
108. Martinez-Bueno, M., et al., *Determination of the gene sequence and the molecular structure of the enterococcal peptide antibiotic AS-48.* J. Bacteriol., 1994. **176**(20): p. 6334-9.
109. Luckett, S., et al., *High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds.* J. Mol. Biol., 1999. **290**(2): p. 525-33.
110. Tan, N.H. and J. Zhou, *Plant cyclopeptides.* Chemical reviews, 2006. **106**(3): p. 840-95.
111. Tang, Y.Q., et al., *A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins.* Science, 1999. **286**(5439): p. 498-502.
112. Mylne, J.S., et al., *Cyclotides are a component of the innate defense of Oldenlandia affinis.* Biopol., 2010. **94**(5): p. 635-46.
113. Trabi, M., et al., *Variations in cyclotide expression in viola species.* J. Natural Prod., 2004. **67**(5): p. 806-10.
114. Lindholm, P., et al., *Cyclotides: a novel type of cytotoxic agents.* Mol. Cancer Ther., 2002. **1**(6): p. 365-9.

115. Svangard, E., et al., *Mechanism of action of cytotoxic cyclotides: cycloviolacin O2 disrupts lipid membranes.* J. Nat. Prod., 2007. **70**(4): p. 643-7.
116. Svangard, E., et al., *Cytotoxic cyclotides from Viola tricolor.* J. Nat. Prod., 2004. **67**(2): p. 144-7.
117. Herrmann, A., et al., *The alpine violet, Viola biflora, is a rich source of cyclotides with potent cytotoxicity.* Phytochem., 2008. **69**(4): p. 939-52.
118. Chen, B., et al., *Cycloviolacin H4, a hydrophobic cyclotide from Viola hederaceae.* J. Nat. Prod., 2006. **69**(1): p. 23-8.
119. Witherup, K.M., et al., *Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from Psychotria longipes.* J. Nat. Prod., 1994. **57**(12): p. 1619-25.
120. Goransson, U., et al., *Reversible antifouling effect of the cyclotide cycloviolacin O2 against barnacles.* J. Nat. Prod., 2004. **67**(8): p. 1287-90.
121. Bokesch, H.R., et al., *A novel anti-HIV macrocyclic peptide from Palicourea condensata.* J. Nat. Prod., 2001. **64**(2): p. 249-50.
122. Chen, B., et al., *Isolation and characterization of novel cyclotides from Viola hederaceae: solution structure and anti-HIV activity of vhl-1, a leaf-specific expressed cyclotide.* J. Biol. Chem., 2005. **280**(23): p. 22395-405.
123. Daly, N.L., et al., *Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity.* J. Mol. Biol., 1999. **285**(1): p. 333-45.
124. Wang, C.K., et al., *Anti-HIV cyclotides from the Chinese medicinal herb Viola yedoensis.* J. Nat. Prod., 2008. **71**(1): p. 47-52.
125. Ireland, D.C., et al., *Cyclotides as natural anti-HIV agents.* Peptide Sci., 2008. **90**(1): p. 51-60.
126. Derua, R., K.R. Gustafson, and L.K. Pannell, *Analysis of the disulfide linkage pattern in circulin A and B, HIV-inhibitory macrocyclic peptides.* Biochem. Biophys. Res. Commun., 1996. **228**(2): p. 632-8.
127. Hallock, Y.F., et al., *Cycloviolins A-D, anti-HIV macrocyclic peptides from Leonia cymosa.* J. Org. Chem., 2000. **65**(1): p. 124-8.
128. Tam, J.P., et al., *An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides.* Proc. Natl. Acad. Sci. USA, 1999. **96**(16): p. 8913-8.
129. Jennings, C., et al., *Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis.* Proc. Natl. Acad. Sci. USA, 2001. **98**(19): p. 10614-9.
130. Colgrave, M.L., et al., *Anthelmintic activity of cyclotides: In vitro studies with canine and human hookworms.* Acta Trop, 2009. **109**(2): p. 163-6.
131. Colgrave, M.L., et al., *The anthelmintic activity of the cyclotides: natural variants with enhanced activity.* Chembiochem., 2008. **9**(12): p. 1939-45.
132. Craik, D.J., *Host-defense activities of cyclotides.* Toxins (Basel), 2012. **4**(2): p. 139-56.
133. Shenkarev, Z.O., et al., *Conformation and mode of membrane interaction in cyclotides. Spatial structure of kalata B1 bound to a dodecylphosphocholine micelle.* FEBS J, 2006. **273**(12): p. 2658-72.

134. Shenkarev, Z.O., et al., *Divalent cation coordination and mode of membrane interaction in cyclotides: NMR spatial structure of ternary complex Kalata B7/Mn2+/DPC micelle.* J. Inorg. Biochem., 2008. **102**(5-6): p. 1246-56.
135. Wang, C.K., et al., *Despite a conserved cystine knot motif, different cyclotides have different membrane binding modes.* Biophys. J., 2009. **97**(5): p. 1471-81.
136. Huang, Y.H., et al., *Lysine-scanning mutagenesis reveals an amendable face of the cyclotide kalata B1 for the optimization of nematocidal activity.* J. Biol. Chem., 2010. **285**(14): p. 10797-805.
137. Huang, Y.H., et al., *The biological activity of the prototypic cyclotide kalata b1 is modulated by the formation of multimeric pores.* J. Biol. Chem., 2009. **284**(31): p. 20699-707.
138. Nourse, A., et al., *A comparison of the self-association behavior of the plant cyclotides kalata B1 and kalata B2 via analytical ultracentrifugation.* J. Biol. Chem., 2004. **279**(1): p. 562-70.
139. Gillon, A.D., et al., *Biosynthesis of circular proteins in plants.* Plant J., 2008. **53**(3): p. 505-515.
140. von Heijne, G., *The signal peptide.* J. Mem. Biol., 1990. **115**(3): p. 195-201.
141. Gruber, C.W., et al., *A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins.* J. Biol. Chem., 2007. **282**(28): p. 20435-46.
142. Hebert, D.N. and M. Molinari, *In and out of the ER: protein folding, quality control, degradation, and related human diseases.* Physiol. Rev., 2007. **87**(4): p. 1377-408.
143. Stevens, F.J. and Y. Argon, *Protein folding in the ER.* Semin Cell Dev. Biol. 1999. **10**(5): p. 443-54.
144. Saska, I., et al., *An Asparaginyl Endopeptidase Mediates in Vivo Protein Backbone Cyclization.* J. Biol. Chem., 2007. **282**(40): p. 29721-29728.
145. Ireland, D.C., et al., *Discovery and characterization of a linear cyclotide from Viola odorata: implications for the processing of circular proteins.* J. Mol. Biol., 2006. **357**(5): p. 1522-35.
146. Gerlach, S.L., et al., *Isolation, Characterization, and Bioactivity of Cyclotides from the Micronesian Plant Psychotria leptothyrsa.* J. Nat. Prod., 2010. **73**(7): p. 1207-1213.
147. Plan, M.R., et al., *The cyclotide fingerprint in Oldenlandia affinis: elucidation of chemically modified, linear and novel macrocyclic peptides.* Chembiochem., 2007. **8**(9): p. 1001-11.
148. Poth, A.G., et al., *Cyclotides Associate with Leaf Vasculature and Are the Products of a Novel Precursor in Petunia (Solanaceae).* J. Biol. Chem., 2012. **287**(32): p. 27033-27046.
149. Gambino, G., I. Perrone, and I. Gribaudo, *A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants.* Phytochem. Anal., 2008. **19**(6): p. 520-5.
150. John, M.E., *An efficient method for isolation of RNA and DNA from plants containing polyphenolics.* Nucleic Acids Res., 1992. **20**(9): p. 2381.

151. Kim, C.S., et al., *A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP.* Nucleic Acids Res., 1997. **25**(5): p. 1085-6.
152. Rose, T.M., et al., *Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences.* Nucleic Acids Res., 1998. **26**(7): p. 1628-35.
153. Rose, T.M., J.G. Henikoff, and S. Henikoff, *CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design.* Nucleic Acids Res., 2003. **31**(13): p. 3763-6.
154. Boyce, R., P. Chilana, and T.M. Rose, *iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences.* Nucleic Acids Res., 2009. **37**(Web Server issue): p. W222-8.
155. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* J. Mol. Biol., 2004. **340**(4): p. 783-95.
156. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.* Protein Eng., 1997. **10**(1): p. 1-6.
157. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res., 2004. **14**(6): p. 1188-90.
158. Boogers, I., et al., *Ultra-performance liquid chromatographic analysis of amino acids in protein hydrolysates using an automated pre-column derivatisation method.* J.Chromatogr. A., 2008. **1189**(1-2): p. 406-9.
159. Pace, C.N., et al., *How to Measure and Predict the Molar Absorption-Coefficient of a Protein.* Protein Sci., 1995. **4**(11): p. 2411-2423.
160. Strobl, S., et al., *The α-amylase from the yellow meal worm: complete primary structure, crystallization and preliminary X-ray analysis.* FEBS Lett., 1997. **409**(1): p. 109-114.
161. Okolo, B.N., L.I. Ezeogu, and C.N. Mba, *Production of raw starch digesting amylase by Aspergillus niger grown on native starch sources.* J. Sci. Food Agricul., 1995. **69**(1): p. 109-115.
162. Miller, G.L., *Use of dinitrosalicylic acid reagent for determination of reducing sugar.* Anal. Chem., 1959. **31**: p. 426-429.
163. Endress, M., *Apocynaceae: Brown and now.* Telopea, 2004. **10**(2): p. 525-541.
164. Li, P.-t., A.J.M. Leeuwenberg, and D.J. Middleton, *Apocynaceae.* Flora of China, 1995. **16**: p. 143-188.
165. Kohls, S., et al., *Cardiac glycosides from Yellow Oleander (Thevetia peruviana) seeds.* Phytochem., 2012. **75**(0): p. 114-127.
166. Bandara, V., et al., *A review of the natural history, toxinology, diagnosis and clinical management of Nerium oleander (common oleander) and Thevetia peruviana (yellow oleander) poisoning.* Toxicon, 2010. **56**(3): p. 273-281.
167. Radford, D.J., et al., *Naturally occurring cardiac glycosides.* Med. J. Aust., 1986. **144**(10): p. 540-4.
168. Arulmozhi, S., et al., *Pharmacological activities of Alstonia scholaris linn. (Apocynaceae) - A review.* Phcog. Rev., 2007. **1**(1): p. 163-170.

169. Nayak, S., P. Nalabothu, and S. Sandiford, *Evaluation of wound healing activity of Allamanda cathartica. L. and Laurus nobilis. L. extracts on rats.* BMC Complement Altern. Med., 2006. **6**(12).
170. Perez, R.M., et al., *Anti-diabetic effects of compounds isolated from plants.* Phytomed., 1998. **5**: p. 55-75.
171. Wong, S., et al., *Assessment of antiproliferative and antiplasmodial activities of five selected Apocynaceae species.* BMC Complement. Altern. Med., 2011. **11**(1): p. 3.
172. Keawpradub, N., et al., *Antiplasmodial activity of extracts and alkaloids of three Alstonia species from Thailand.* Planta Med., 1999. **65**(8): p. 690-4.
173. Chang, L.C., et al., *Activity-guided isolation of constituents of Cerbera manghas with antiproliferative and antiestrogenic activities.* Bioorganic & Med. Chem. Lett., 2000. **10**(21): p. 2431-2434.
174. Siddiqui, B.S., et al., *Two cytotoxic pentacyclic triterpenoids from Nerium oleander.* Phytochem., 1995. **39**(1): p. 171-4.
175. Tiwari, T.N., V.B. Pandey, and N.K. Dubey, *Plumieride from Allamanda cathartica as an antidermatophytic agent.* Phytotherapy Res., 2002. **16**(4): p. 393-394.
176. Schwikkard, S. and F.R. van Heerden, *Antimalarial activity of plant metabolites.* Nat. Prod. Rep., 2002. **19**(6): p. 675-92.
177. Wright, C.W., et al., *Alstonia species: are they effective in malaria treatment?* J. Ethnoph., 1993. **40**(1): p. 41-5.
178. Clay, H.F., J.C. Hubbard, and R. Golt, *Allamanda oenotheraefolia - Bush Allamanda*, in *Tropical Shrubs*1987, University of Hawaii Press. p. 172.
179. Nayak, S., P. Nalabothu, and S. Sandiford, *Evaluation of wound healing activity of Allamanda cathartica. L. and Laurus nobilis. L. extracts on rats.* BMC Complement. Altern. Med., 2006. **6**(12).
180. Kumar Pratyush, et al., *Ethnobotanical and Pharmacological Study of Alstonia (Apocynaceae) - A Review.* J. Pharm. Sci. Res., 2011. **3**(8): p. 1394-1403.
181. Mazumder, P., et al., *Pharmacological activities of Alstonia scholaris linn. (Apocynaceae) - A review*. Vol. 1. 2007. 163-170.
182. Arulmozhi, S., et al., *Pharmacological activities of Alstonia scholaris linn. (Apocynaceae) - A review*. Vol. 1. 2007. 163-170.
183. Garden, C.B. *Water Jasmine Bonsai Tree*. 2008-2012 [cited 2012; Available from: http://www.chinesebonsaigarden.com/water-jasmine-bonsai-tree/.
184. Arya, P. *Medicinal Herbs - Names and Uses (Part-2)*. 2012 [cited 2012; Available from: http://www.globalhinduism.com/article/2012/01/medicinal-herbs-names-and-uses-part-2/.
185. Gruber, C.W., et al., *Distribution and Evolution of Circular Miniproteins in Flowering Plants.* Plant Cell Onl., 2008. **20**(9): p. 2471-2483.
186. Lu, B.S., et al., *Conopeptides from Conus striatus and Conus textile by cDNA cloning.* Peptides, 1999. **20**(10): p. 1139-44.
187. Gray, W.R., *Disulfide structuRes. of highly bridged peptides: a new strategy for analysis.* Protein Sci., 1993. **2**(10): p. 1732-48.

188. Goransson, U. and D.J. Craik, *Disulfide mapping of the cyclotide kalata B1. Chemical proof of the cystic cystine knot motif.* J.Biol. Chem., 2003. **278**(48): p. 48188-96.
189. Bernfeld, P., *[17] Amylases, [alpha] and [beta]*, in *Methods in Enzymology*1955, Academic Press. p. 149-158.
190. Nguyen, G.K.T., et al., *Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family.* J. Biol. Chem., 2011. **286**(27): p. 24275-24287.
191. Tam, J.P., Y.-A. Lu, and J.-L. Yang, *Antimicrobial dendrimeric peptides.* Eur. J. Biochem., 2002. **269**(3): p. 923-932.
192. Gracy, J. and L. Chiche, *Optimizing structural modeling for a specific protein scaffold: knottins or inhibitor cystine knots.* BMC Bioinform., 2010. **11**(1): p. 535.
193. Bontems, F., et al., *Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins.* Sci., 1991. **254**(5037): p. 1521-1523.
194. Cornet, B., et al., *Refined three-dimensional solution structure of insect defensin A.* Structure, 1995. **3**(5): p. 435-448.
195. Hoffmann, J.A., *Innate immunity of insects.* Curr. Opin. Immunol., 1995. **7**(1): p. 4-10.
196. Dimarcq, J.-L., et al., *Cysteine-rich antimicrobial peptides in invertebrates.* Peptide Sci., 1998. **47**(6): p. 465-477.
197. Zhu, S., B. Gao, and J. Tytgat, *Phylogenetic distribution, functional epitopes and evolution of the CSαβ superfamily.* Cell. Mol.r Life Sci., 2005. **62**(19): p. 2257-2269.
198. Pereira, P.J.B., et al., *Specific inhibition of insect alpha-amylases: yellow meal worm alpha-amylase in complex with the Amaranth alpha-amylase inhibitor at 2.0 Å resolution.* Structure (London, England : 1993), 1999. **7**(9): p. 1079-1088.
199. Pereira, P.J.B., et al., *Specific inhibition of insect α-amylases: yellow meal worm α-amylase in complex with the Amaranth α-amylase inhibitor at 2.0 Å resolution.* Structure, 1999. **7**(9): p. 1079-1088.
200. Gracy, J. and L. Chiche, *Optimizing structural modeling for a specific protein scaffold: knottins or inhibitor cystine knots.* BMC Bioinform., 2010. **11**(1): p. 535.
201. Nguyen, G.K.T., et al., *Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry.* J. Biol. Chem., 2011.
202. Chagolla-Lopez, A., A. Blanco-Labran, and A. Patthy, *A Novel a-Amylase Inhibitor from Amaranth (Amaranthus hypocondriacus) Seeds.* J. Biol. Chem., 1994. **269**(38): p. 23675-23680.
203. Raux, H., et al., *The matrix protein of vesicular stomatitis virus binds dynamin for efficient viral assembly.* J. Virol., 2010. **84**(24): p. 12609-18.
204. Zamponi, G.W., *Block of Voltage-Gated Calcium Channels by Peptide Toxins*, in *Voltage-Gated Calcium Channels* 2005, Kluwer Academic/Plenum Publishers. p. 301.
205. Olivera, B.M., et al., *Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using omega-*

conotoxin from Conus magus venom. Biochem. J., 1987. **26**(8): p. 2086-90.
206. Ireland, D.C., M.L. Colgrave, and D.J. Craik, *A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability.* Biochem. J., 2006 July 27. **400**(Pt 1): p. 1–12.
207. Mergaert, P., et al., *A Novel Family in Medicago truncatula Consisting of More Than 300 Nodule-Specific Genes Coding for Small, Secreted Polypeptides with Conserved Cysteine Motifs.* Plant Phys., 2003. **132**(1): p. 161-173.
208. Davids, B.J., et al., *A New Family of Giardial Cysteine-Rich Non-VSP Protein Genes and a Novel Cyst Protein.* PLoS One, 2006. **1**(1): p. e44.
209. Jennings, et al., *Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis.* Proc. Natl. Acad. Sci. U.S.A., 2001 Sep 11. **98**(19): p. 10614-9.
210. Sha, X., L. Yang, and L.E. Gentry, *Identification and analysis of discrete functional domains in the pro region of pre-pro-transforming growth factor beta 1.* J. Cell Biol., 1991. **114**(4): p. 827-839.
211. de Sa, M.F.G., et al., *Molecular characterization of a bean α-amylase inhibitor that inhibits the α-amylase of the Mexican bean weevil Zabrotes subfasciatus.* Planta, 1997. **203**(3): p. 295-303.
212. Franco, O.L., et al., *Plant α-amylase inhibitors and their interaction with insect α-amylases.* Eur. J. Biochem., 2002. **269**(2): p. 397-412.
213. Cemazar, M., et al., *Oxidative folding of Amaranthus alpha-amylase inhibitor: disulfide bond formation and conformational folding.* J. Biol. Chem., 2004. **279**(16): p. 16697-705.
214. Hobohm, U. and C. Sander, *Enlarged representative set of protein structures.* Protein Sci., 1994. **3**(3): p. 522-524.
215. Gracy, J., et al., *KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007.* Nucleic Acids Res., 2008. **36**(Database issue): p. D314 - 319.
216. Norton, R.S. and P.K. Pallaghy, *The cystine knot structure of ion channel toxins and related polypeptides.* Toxicon, 1998. **36**(11): p. 1573-83.
217. 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.* Biochem. J., 2006. **394**(Pt 1): p. 85-93.
218. Krause, S., et al., *Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists.* Febs. J., 2007. **274**(1): p. 86-95.
219. Nuttall, S.D., et al., *Design and expression of soluble CTLA-4 variable domain as a scaffold for the display of functional polypeptides.* Proteins, 1999. **36**(2): p. 217-27.
220. Wong, C.T.T., et al., *Orally Active Peptidic Bradykinin B1 Receptor Antagonists Engineered from a Cyclotide Scaffold for Inflammatory Pain Treatment.* Angewandte Chemie International Edition, 2012. **51**(23): p. 5620-5624.
221. Nahoum, V., et al., *A plant-seed inhibitor of two classes of alpha-amylases: X-ray analysis of Tenebrio molitor larvae alpha-amylase in*

complex with the bean Phaseolus vulgaris inhibitor. Acta Crystallogr. D Biol. Crystallogr., 1999. **55**(Pt 1): p. 360-2.
222. Wiegand, G., O. Epp, and R. Huber, *The crystal structure of porcine pancreatic alpha-amylase in complex with the microbial inhibitor Tendamistat.* J. Mol. Biol., 1995. **247**(1): p. 99-110.
223. Doleckova-Maresova, L., et al., *De novo design of alpha-amylase inhibitor: a small linear mimetic of macromolecular proteinaceous ligands.* Chem. Biol., 2005. **12**(12): p. 1349-57.
224. Ochiai, T., et al., *Screening of an alpha-amylase inhibitor peptide by photolinker-peptide array.* Biosci. Biotechnol. Biochem., 2012. **76**(4): p. 819-24.
225. Brogard, J.M., et al., *Alpha-glucosidase inhibitors: a new therapeutic approach in diabetes and functional hypoglycemia.* Rev. Med. Interne., 1989. **10**(4): p. 365-74.
226. Thakur, N., A. Qureshi, and M. Kumar, *AVPpred: collection and prediction of highly effective antiviral peptides.* Nucleic Acids Res., 2012.
227. Jones, J.C., et al., *Identification of the Minimal Active Sequence of an Anti-Influenza Virus Peptide.* Antimicrobial Agents and Chemotherapy, 2011. **55**(4): p. 1810-1813.
228. Drummond, J.E., et al., *Design and optimization of a multiplex anti-influenza peptide immunoassay.* J. Immunol. Methods, 2008. **334**(1-2): p. 11-20.
229. Bai, F., et al., *Antiviral Peptides Targeting the West Nile Virus Envelope Protein.* J. Virology, 2007. **81**(4): p. 2047-2055.
230. Lambert, D.M., et al., *Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion.* Proc. Natl. Acad. Sci. U.S.A., 1996. **93**(5): p. 2186-91.
231. Akkarawongsa, R., et al., *Multiple peptides homologous to herpes simplex virus type 1 glycoprotein B inhibit viral infection.* Antimicrob. Agents Chemother., 2009. **53**(3): p. 987-96.
232. Lalonde, M.S., et al., *Inhibition of Both HIV-1 Reverse Transcription and Gene Expression by a Cyclic Peptide that Binds the Tat-Transactivating Response Element (TAR) RNA.* PLoS Pathogens, 2011. **7**(5): p. e1002038.
233. Fjell, C.D., et al., *Designing antimicrobial peptides: form follows function.* Nat. Rev. Drug Discov., 2012. **11**(1): p. 37-51.
234. Yang, Z., et al., *Anti-Infectious Bronchitis Virus (IBV) Activity of 1,8-cineole: Effect on Nucleocapsid (N) Protein.* J. Biomol. Structure Dynamics, 2010. **28**(3): p. 323-330.
235. Mondal, S.P. and C.J. Cardona, *Comparison of four regions in the replicase gene of heterologous infectious bronchitis virus strains.* Virol., 2004. **324**(1): p. 238-248.
236. Cavanagh, D., *Coronavirus avian infectious bronchitis virus.* Vet. Res., 2007. **38**(2): p. 281-97.
237. Lai, M.M., *Molecular Biol. of coronavirus 1986.* Adv. Exp. Med. Biol., 1987. **218**: p. 7-13.

238. Masters, P.S., *The Molecular Biol. of Coronaviruses*, in *Advances in Virus Res.*, M. Karl and J.S. Aaron, Editors. 2006, Academic Press. p. 193-292.
239. Baker, S.C., *Coronaviruses: Molecular Biol.*, in *Encyclopedia of Virology (Third Edition)*, B.W.J.M. Editors-in-Chief: and M.H.V.v. Regenmortel, Editors. 2008, Academic Press: Oxford. p. 554-562.
240. Rottier, P.J.M., *The Coronaviridae*, S.G. Siddell, Editor 1995, Plenum: New York. p. 115-139.
241. Spaan, W., D. Cavanagh, and M.C. Horzinek, *Coronaviruses: structure and genome expression.* J.Gen. Virol., 1988. **69 (Pt 12)**: p. 2939-52.
242. de Haan, C.A.M. and P.J.M. Rottier, *Molecular Interactions in the Assembly of Coronaviruses*, in *Advances in Virus Res.*, R. Polly, Editor 2005, Academic Press. p. 165-230.
243. Lim, K.P. and D.X. Liu, *The Missing Link in Coronavirus Assembly: retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-golgi compartments and physical interaction between the envelope and membrane proteins.* J. Biol. Chem., 2001. **276**(20): p. 17515-17523.
244. Laude, H. and P.S. Masters, *The Coronaviridae*, S.G. Siddell, Editor 1995, Plenum: New York. p. 141-163.
245. Klumperman, J., et al., *Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding.* J.Virol., 1994. **68**(10): p. 6523-34.
246. Kitazato, K., Y. Wang, and N. Kobayashi, *Viral infectious disease and natural products with antiviral activity.* Drug Discov. Ther., 2007. **1**(1): p. 14-22.
247. Jassim, S.A.A. and M.A. Naji, *Novel antiviral agents: a medicinal plant perspective.* J. Applied MicroBiol., 2003. **95**(3): p. 412-427.
248. Mukhtar, M., et al., *Antiviral potentials of medicinal plants.* Virus Res., 2008. **131**(2): p. 111-120.
249. Yan, F., et al., *Phylogenetic analysis of S1 gene of infectious bronchitis virus isolates from China.* Avian Dis., 2011. **55**(3): p. 451-8.
250. Liu, S., et al., *Molecular characterization and pathogenicity of infectious bronchitis coronaviruses: complicated evolution and epidemiology in china caused by cocirculation of multiple types of infectious bronchitis coronaviruses.* Intervirol., 2009. **52**(4): p. 223-34.
251. Laughlin, C.A., et al., *Dengue Res. Opportunities in the Americas.* J. Infectious Diseases, 2012. **206**(7): p. 1121-1127.
252. Lin, Y.Z., et al., *Inhibition of nuclear translocation of transcription factor NF-kappa B by a synthetic peptide containing a cell membrane-permeable motif and nuclear localization sequence.* J.Biol. Chem., 1995. **270**(24): p. 14255-8.
253. Bultmann, H., J.S. Busse, and C.R. Brandt, *Modified FGF4 signal peptide inhibits entry of herpes simplex virus type 1.* J.Virol., 2001. **75**(6): p. 2634-45.
254. Jones, J.C., et al., *Inhibition of influenza virus infection by a novel antiviral peptide that targets viral attachment to cells.* J.Virol., 2006. **80**(24): p. 11960-7.

255. Altmann, S.E., et al., *Inhibition of Vaccinia virus entry by a broad spectrum antiviral peptide.* Virol., 2009. **388**(2): p. 248-59.
256. Altmann, S.E., et al., *Antiviral activity of the EB peptide against zoonotic poxviruses.* Virol. J., 2012. **9**: p. 6.
257. White, J.M., et al., *StructuRes. and mechanisms of viral membrane fusion proteins: multiple variations on a common theme.* Crit. Rev. Biochem. Mol. Biol., 2008. **43**(3): p. 189-219.
258. Huang, C.Y.H., et al., *The dengue virus type 2 envelope protein fusion peptide is essential for membrane fusion.* Virol., 2010. **396**(2): p. 305-315.
259. Li, D. and D. Cavanagh, *Coronavirus IBV-induced membrane fusion occurs at near-neutral pH.* Arch. Virol., 1992. **122**(3-4): p. 307-16.
260. Chu, V.C., et al., *The avian coronavirus infectious bronchitis virus undergoes direct low-pH-dependent fusion activation during entry into host cells.* J.Virol., 2006. **80**(7): p. 3180-8.
261. Belouzard, S., et al., *Mechanisms of coronavirus cell entry mediated by the viral spike protein.* Viruses, 2012. **4**(6): p. 1011-33.
262. Lu, M. and P.S. Kim, *A trimeric structural subdomain of the HIV-1 transmembrane glycoprotein.* J. Biomol. Struct. Dyn., 1997. **15**(3): p. 465-71.
263. LaBonte, J., J. Lebbos, and P. Kirkpatrick, *Enfuvirtide.* Nat. Rev. Drug Discov., 2003. **2**(5): p. 345-346.
264. Wild, C.T., et al., *Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection.* Prod. Nat. Acad. Sci., 1994. **91**(21): p. 9770-9774.
265. Liu, S., et al., *Different from the HIV Fusion Inhibitor C34, the Anti-HIV Drug Fuzeon (T-20) Inhibits HIV-1 Entry by Targeting Multiple Sites in gp41 and gp120.* J. Biol. Chem., 2005. **280**(12): p. 11259-11273.
266. Lambert, D.M., et al., *Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion.* Prod. Nat. Acad. Sci., 1996. **93**(5): p. 2186-2191.
267. Wild, T.F. and R. Buckland, *Inhibition of measles virus infection and fusion with peptides corresponding to the leucine zipper region of the fusion protein.* J. Gen. Virol., 1997. **78**(1): p. 107-11.
268. Young, J.K., et al., *Interaction of Peptides with Sequences from the Newcastle Disease Virus Fusion Protein Heptad Repeat Regions.* J. Virol., 1999. **73**(7): p. 5945-5956.
269. Porotto, M., et al., *Inhibition of Hendra Virus Fusion.* J. Virol., 2006. **80**(19): p. 9837-9849.
270. Rapaport, D., M. Ovadia, and Y. Shai, *A synthetic peptide corresponding to a conserved heptad repeat domain is a potent inhibitor of Sendai virus-cell fusion: an emerging similarity with functional domains of other viruses.* EMBO J., 1995. **14**(22): p. 5524-31.
271. Wang, J., et al., *Interaction of the coronavirus infectious bronchitis virus membrane protein with beta-actin and its implication in virion assembly and budding.* PLoS One, 2009. **4**(3): p. e4908.

272. Maier, C.C., et al., *A single M protein mutation affects the acid inactivation threshold and growth kinetics of a chimeric flavivirus.* Virol., 2007. **362**(2): p. 468-474.
273. Sadanand, S., *Vaccination: the present and the future.* Yale J.Biol. Med., 2011. **84**(4): p. 353-9.
274. Castel, G., et al., *Phage Display of Combinatorial Peptide Libraries: Application to Antiviral Res..* Molecules, 2011. **16**(5): p. 3499-3518.
275. Lindgren, M., et al., *Cell-penetrating peptides.* Trends in Pharm. Sci., 2000. **21**(3): p. 99-103.
276. Järver, P. and Ü. Langel, *Cell-penetrating peptides—A brief introduction.* Biochim. Biophys. Acta (BBA) - Biomembranes, 2006. **1758**(3): p. 260-263.
277. El-Andaloussi, S., T. Holm, and U. Langel, *Cell-penetrating peptides: mechanisms and applications.* Curr. Pharm. Des., 2005. **11**(28): p. 3597-611.
278. Richard, J.P., et al., *Cell-penetrating Peptides: a reevaluation of the mechanism of cellular uptake.* J. Biol. Chem., 2003. **278**(1): p. 585-590.
279. Madani, F., et al., *Mechanisms of cellular uptake of cell-penetrating peptides.* J. Biophys., 2011. **2011**: p. 414729.
280. Mino, T., et al., *Cell-permeable artificial zinc-finger proteins as potent antiviral drugs for human papillomaviruses.* Archives of Virol., 2008. **153**(7): p. 1291-1298.
281. Nobeli, I., A.D. Favia, and J.M. Thornton, *Protein promiscuity and its implications for biotechnology.* Nat. Biotech., 2009. **27**(2): p. 157-167.
282. Babtie, A., N. Tokuriki, and F. Hollfelder, *What makes an enzyme promiscuous?* Curr. Opin. Chem. Biol., 2010. **14**(2): p. 200-207.
283. Franco, O.L., *Peptide promiscuity: An evolutionary concept for plant defense.* FEBS Lett., 2011. **585**(7): p. 995-1000.
284. Pinto, M.F.S., et al., *Cyclotides: From Gene Structure to Promiscuous Multifunctionality.* J. Evidence-Based Complementary & Alternative Medicine, 2011.
285. Arnison, P.G., et al., *Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature.* Nat. Prod. Reports, 2013. **30**(1): p. 108-160.
286. Dagar, H.S. and J.C. Dagar, *Plant folk medicines among the nicobarese of Katchal Island, India.* Economic Botany, 1991. **45**(1): p. 114-119.
287. Mukherjee, P.K., et al., *The Ayurvedic medicine Clitoria ternatea--from traditional use to scientific assessment.* J. Ethnopharmacol., 2008. **120**(3): p. 291-301.
288. Trabi, M. and D.J. Craik, *Tissue-specific expression of head-to-tail cyclized miniproteins in Violaceae and structure determination of the root cyclotide Viola hederacea root cyclotide1.* Plant Cell, 2004. **16**(8): p. 2204-16.
289. Nguyen, G.K., et al., *Novel cyclotides and uncyclotides with highly shortened precursors from Chassalia chartacea and effects of methionine oxidation on bioactivities.* J. Biol. Chem., 2012. **287**(21): p. 17598-607.

290. Nguyen, G.K., et al., *Discovery of Linear Cyclotides in Monocot Plant Panicum laxum of Poaceae Family Provides New Insights into Evolution and Distribution of Cyclotides in Plants.* J. Biol. Chem., 2012.
291. Froy, O. and M. Gurevitz, *Membrane potential modulators: a thread of scarlet from plants to humans.* FASEB J., 1998. **12**(15): p. 1793-6.
292. Van Damme, E.J., et al., *A gene encoding a hevein-like protein from elderberry fruits is homologous to PR-4 and class V chitinase genes.* Plant Physiol., 1999. **119**(4): p. 1547-56.
293. Castagnaro, A., et al., *Extreme divergence of a novel wheat thionin generated by a mutational burst specifically affecting the mature protein domain of the precursor.* J. Mol. Biol., 1992. **224**(4): p. 1003-9.
294. Miura, K., et al., *Cloning of mRNA sequences for two antibacterial peptides in a hemipteran insect, Riptortus clavatus.* Zoolog. Sci., 1996. **13**(1): p. 111-7.
295. Casteels-Josson, K., et al., *Apidaecin multipeptide precursor structure: a putative mechanism for amplification of the insect antibacterial response.* EMBO J., 1993. **12**(4): p. 1569-78.
296. Terry, A.S., et al., *The cDNA sequence coding for prepro-PGS (prepro-magainins) and aspects of the processing of this prepro-polypeptide.* J. Biol. Chem., 1988. **263**(12): p. 5745-51.
297. Imamura, M., et al., *Multipeptide precursor structure of acaloleptin A isoforms, antibacterial peptides from the Udo longicorn beetle, Acalolepta luxuriosa.* Dev. Comp. Immunol., 2009. **33**(10): p. 1120-7.
298. Belknap, W.R. and J.E. Garbarino, *The role of ubiquitin in plant senescence and stress responses.* Trends Plant Sci., 1996. **1**(10): p. 331-335.
299. Atkinson, A.H., et al., *Proteinase inhibitors in Nicotiana alata stigmas are derived from a precursor protein which is processed into five homologous inhibitors.* Plant Cell, 1993. **5**(2): p. 203-13.
300. Dutton, J.L., et al., *Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins.* J. Biol Chem., 2004. **279**(45): p. 46858-67.
301. Wang, C.K., et al., *CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering.* Nucleic Acids Res., 2008. **36**(Database issue): p. D206-10.
302. Rosengren, K.J., et al., *Twists, knots, and rings in proteins. Structural definition of the cyclotide framework.* J. Biol. Chem., 2003. **278**(10): p. 8606-16.
303. Simonsen, S.M., et al., *Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity.* J. Biol. Chem., 2008. **283**(15): p. 9805-13.
304. Craik, D.J., *Plant cyclotides: circular, knotted peptide toxins.* Toxicon, 2001. **39**(12): p. 1809-13.
305. Barry, D.G., et al., *Linearization of a Naturally Occurring Circular Protein Maintains Structure but Eliminates Hemolytic Activity†,‡.* Biochem., 2003. **42**(22): p. 6688-6695.
306. Daly, N.L. and D.J. Craik, *Acyclic Permutants of Naturally Occurring Cyclic Proteins: characterization of cystine knot and β-sheet formation*

in the macrocyclic polypeptide kalata b1. J. Biol. Chem., 2000. **275**(25): p. 19068-19075.
307. Craik, D.J., *Circling the enemy: cyclic proteins in plant defence.* Trends Plant Sci., 2009. **14**(6): p. 328-35.
308. Misumi, S., et al., *Zn2+ binding to cysteine-rich domain of extracellular human immunodeficiency virus type 1 Tat protein is associated with Tat protein-induced apoptosis.* AIDS Res. Hum Retroviruses, 2004. **20**(3): p. 297-304.
309. Bromme, D., *Papain-like cysteine proteases.* Curr. Protoc. Protein Sci., 2001. **Chapter 21**: p. Unit 21 2.
310. dos Reis, F.C., et al., *The substrate specificity of cruzipain 2, a cysteine protease isoform from Trypanosoma cruzi.* FEMS Microbiol. Lett., 2006. **259**(2): p. 215-20.
311. Kembhavi, A.A., et al., *The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays.* Arch. Biochem. Biophys., 1993. **303**(2): p. 208-13.
312. Müntz, K., F.R. Blattner, and A.D. Shutov, *Legumains - a family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants.* J. Plant Physiol., 2002. **159**(12): p. 1281-1293.
313. Müntz, K. and A.D. Shutov, *Legumains and their functions in plants.* Trends Plant Sci., 2002. **7**(8): p. 340-344.
314. Shimada, T., et al., *Vacuolar processing enzymes are essential for proper processing of seed storage proteins in Arabidopsis thaliana.* J. Biol. Chem., 2003. **278**(34): p. 32292-9.
315. Hara-Nishimura, I., K. Inoue, and M. Nishimura, *A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms.* FEBS Lett., 1991. **294**(1-2): p. 89-93.
316. Sheldon, P.S., J.N. Keen, and D.J. Bowles, *Post-translational peptide bond formation during concanavalin A processing in vitro.* Biochem. J, 1996. **320 (Pt 3)**: p. 865-70.
317. Becker, C., et al., *Purification, cDNA cloning and characterization of proteinase B, an asparagine-specific endopeptidase from germinating vetch (Vicia sativa L.) seeds.* Eur. J. Biochem. / FEBS, 1995. **228**(2): p. 456-62.
318. Ireland, D.C., M.L. Colgrave, and D.J. Craik, *A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability.* Biochem. J., 2006. **400**(1): p. 1-12.
319. Bordusa, F., *Enzymes for peptide cyclization.* Chembiochem., 2001. **2**(6): p. 405-9.
320. Luo, H., H.E. Hallen-Adams, and J.D. Walton, *Processing of the phalloidin proprotein by prolyl oligopeptidase from the mushroom Conocybe albipes.* J. Biol. Chem., 2009. **284**(27): p. 18070-7.
321. Zhang, J., *Evolution by gene duplication: an update.* Trends Ecol. Evo. (Personal edition), 2003. **18**(6): p. 292-298.
322. Keeling, P.J. and J.D. Palmer, *Horizontal gene transfer in eukaryotic evolution.* Nat. Rev. Genet., 2008. **9**(8): p. 605-18.
323. Andersson, J.O., *Lateral gene transfer in eukaryotes.* Cell Mol. Life Sci., 2005. **62**(11): p. 1182-97.

324. Yamazaki, T., et al., *A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes.* Eur. J. Biochem. / FEBS, 2003. **270**(6): p. 1269-76.
325. Long, M. and C.H. Langley, *Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila.* Sci., 1993. **260**(5104): p. 91-5.
326. Bock, R., *The give-and-take of DNA: horizontal gene transfer in plants.* Trends Plant Sci., 2010. **15**(1): p. 11-22.
327. Whittington, C.M., et al., *Novel venom gene discovery in the platypus.* Genome Biol., 2010. **11**(9): p. R95.
328. Lewis, R., *Conotoxins: Molecular and Therapeutic Targets*, in *Marine Toxins as Res. Tools*, N. Fusetani and W. Kem, Editors. 2009, Springer Berlin Heidelberg. p. 45-65.

Appendix A

Primer used in the cloning project

Primer	Forward	Sequence	Encoded aa	Remarks
Allamanda oenotheraefolia				
AO1 B1	x	ATTGCTCACTAyGGnAArTGyGA	IAHYGKCD	
AO2-3 B1	x	UGTGArGGCGTnAthAAyCA	CDGVINQ	
AO1-2-3 B1	x	ATTAATCArUGyUGyGArCCnTGG	INQCCDPW	
AO4 E1		AGGCAAACATGAAAGAAATAGT	3′ UTR of aO4	
AO4 E2		CCTTCTGGCCGATCTGCTTCGA	3′ UTR of aO4	
AO4 E2′		GCCTTCTGGCCGATCTGCTTCGA	3′ UTR of aO4	
AO4 E3		GAGGCAGAAGCCGATAATCGGC	3′ UTR of aO4	
AO_SFSMAKLA	x	GCAAGAGTTTCTCCATGGCGAAGC	SFSMAKLA	Signal sequence from aO5
AO5speB	x	GCCCTCTTACATATTCATTGCAAG	5′ UTR of aO5	
AO1E1		ATAAAGCCTTCTGGTCGATCTGC	3′ UTR of aO1	
AO-F1	x	AAYAARTGGTAYTGYGAYGG	NKWYCDG	
AO-FN1	x	ATATAYAARTGYTGYGAYCC	IYKCCDP	
AO10speE		GCTTCATTAAACCCACAGAACA	3′ UTR of aO10	
AO10speEn		ACTGCTTAGCCGGAGATCTTT	3′ UTR of aO10	
AO11speE		TTGGACAGAAGAACAACATTGC	3′ UTR of aO10	
AO10speBn	x	TACAGACATGGCAAAGCTAGC	5′ UTR of aO10	
AO10speB	x	GCTTTAAACATATATACAGAC	5′ UTR of aO10	
Alstonia scholaris				
AS3_B1	x	TGCGATGGTAThAThAAyCArTG	CDGIINQC	
AS123_B1	x	ATTAATCArTGyTGyGArCCnTA	INQCCDPY	
AS12_B1n2	x	TGCACTCCGCCGATTATCGGCAT	CTPPIIG	
AS1_MAKLA	x	TGGCTAAGCTTGCTTGCTTCCTTC	MAKLACFL	Signal sequence from aS1
AS4 E1		CTGGACAGGCACCTTTATTTTAGG	3′ UTR of aS4	

AS5 E1		CAGGATTATGCTTCCTCCAGGGCA	3′ UTR of aS5	
Wrightia religiosa				
WR1B	x	AATAATTGyTGyCArGGnTCnAT	NNCCQGSI	
WR1Bn	x	GCCArGGTTCnAThTGyAThTGGC	CCQGSICIWP	
WR2B	x	AAGTGyTGyAAGAThTAyTGGTG	KCCKIYWC	
WR4E		GAGCAACGGTAATGGGAAGAGAG	3′ UTR of wR4	
WR4En		AGAAGCCAGCGAACGAAACAGT	3′ UTR of wR4	
WR1Bspe	x	AGTGGTATCAACGCAGAGTA	5′ UTR of wR1	
WR2Bspe	x	AAGCAGTGGTATCAACGCAGA	5′ UTR of wR2	
WR1aBspe	x	AAGCAGTGGTATCAACGCAGAGTA	5′ UTR of wR1	
WR3B	x	CArAArGGnGArTAyTG	QKGEYC	
WR3Bn	x	TGTGCTCArAArGGnGA	CAQKGE	
WR4B	x	AATGArTTyTGyAAyGGnTGG	NEFCNGW	
WR4Bn	x	TTATGyAAyGGnTGGACnAT	FCNGWTI	
WR5B	x	ATTCArTGyTGyGAyCCnTA	IQCCDPY	
WR5Bn	x	TGTGAyCCnTAyCGnTG	CDPYRC	
WR2Espe		CATTCACCACATAGCCCGTA	3′ UTR of wR2	
WR2speEn		TCACCACATAGCCCGTAGAACTG	3′ UTR of wR2	
Hedyotis biflora				
B12 B1	x	CTyAAAGGwGGrGTkCCATG	LKGGVPC	
B12 M1	x	ATwCCATGCAThAGCTCsGTGTTyG	IPCISSVFG	
B12 E1		CGTTATGrTAACAwGCyTTrTC	DKACYHNE	
H28 E1		GTAACATAATTTGTTTTGGCTGCAC	CTCSQNKLCY	
HBComE		CAGCAGATGGTGTTAATTCGT	3′ UTR of hB2	
HB_MAHFIKY	x	TGGCTCATTTCATCAAATATCTC	MAHFIKY	Signal peptide of hB2
MASFTKY_HB	x	ATGGCTAGTTTCACTAAATATCT	MASFTKY	Signal peptide of hB4
Clitoria ternatea				
CT3 B1	x	ACnTGTGGGGAAACnTGTAC	TCGETCT	
CT3 M1-1	x	CCnGAyTGTTCATGTTCnTGGC	PDCSC**S**WP	

CT3 M1-2	x	CCnGAyTGTTCATGTAGyTGGC	PDCSC**S**WP	
CT10 B1	x	TGTGCnGAAAGTTGTGTnTGGAT	CAESCVWI	
CT3 E		AACAAATATCCTTCAGATTCAGCA	HAESEGYLL	
MAYVRL	x	ATGGCATACGTTAGGCTT	MAYVRL	Conserved sequence at 5□end
GTGNFCA		AGCACAAAAGTTTCCAGTTCC	GTGNFCA	Conserved sequence at 3□ UTR
5end_CT3	x	AAGCAGTGGTATAACGCAGAGTACGC	5′ UTR of cT3	
CT3speB	x	GGCACAGCAAAAGAGTAATTCC	5′ UTR of cT3	
CT7speB	x	GGGTAACAAATACTTCTAATTCC	5′ UTR of cT7	
CT3speE		GGTACTTTCCAAAGGCATCTTTAAG TCA	3′ UTR of cT3	
CT8speE		CCTTTCAGTTGGTAAGTGCTTTAGGC	3′ UTR of cT8	
CT10speE		AGTAGCAGAGAATGCATCATACATTT CAAC	3′ UTR of cT10	
CT10speE1		ATTCACTACAGAGGATCAGAGGGTCT	3′ UTR of cT10	
CT10speB	x	GAACCTCTAACGAATATCAGTAATCA	5′ UTR of cT10	
CT6 B1	x	CCATGTGGrGGAAGTTGTGThTAyA	PCGESCVYI	
CT6 B1n	x	TACATTCCATGyATAACwACwATC	YIPCITTI	
MASLRIA	x	ATGGCTTCCCTTCGCATTGCT	MASLRIA	Signal peptide of cT4
CT12_3RACE	x	CCTTGTGCATCTTGTGTTTA	ACAESCVY	
CT13_3RACE	x	CCTTGTGCATCTTGTGTTTGG	ACAESCVW	
CT14_3RACE	x	CCTTGTGCATCTTGTGTTTT	ACAESCVF	
Hedyotis uncinella				
AAFALPA	x	GGGCHGCHTTYGCHCTTCCHGC	AAFALPA	
GVLEVES	x	TGTTTTGGAAGTTGAGTCTGCT	GVLEVES	
Hu2990 B1	x	TGCCATGAGACrTGyTTyTTyAT	CHETCFFI	
Hu2990 B1n	x	TGCTTCTTCAThCCTTGyTTyAC	CFFIPCFT	
Hu3268 B1n	x	TGCCArAAyAArGTTTGyAC	CQNKVCT	
Hu2879 B1	x	CCTTGCGGTGAyACTTGyTTyAT	PCGDTCFI	

Hu2879 B1n	x	TGCTTCAThGGTCCnTGyAAyGAy	CFIGPCND	
MANFTNY	x	ATGGCTAATTTTACAAACTATCT	MANFTNY	Signal sequence of hU1
HU8B	x	GGTGTCGTTCCyTGyGGnGAyAC	GVVPCGDT	
HU8Bn	x	GTTCCyTGTGGTGAyACnTGCTT	VPCGDTCF	
HU1B	x	GGAGGTGAyACnTGTCAyGAGAC	GGDTCHET	
HU1Bn	x	GAyACTTGTCAyGAGACnTGyTT	DTCHETCF	
HU18B	x	TGTGGTGArACnTGCCGnGT	CGETCRV	
HU18Bn	x	GCTAThGCTGGnTGCCAyTGCGV	AIAGCHCV	
HU27B	x	TGTGGTGAyCCnTGCTTyTTyGG	CGDPCFFG	
HU27Bn	x	TTyTTyGGTCCnTGCAAyGAyCC	FFGPCNDP	
HU1E		AGCTGAGATCAGGTCGTTTGTCA	VTNDLISA	
HU18E		CAGAATCCCAATTCAAACAGCAT	MLFELGFCFP	
HUComE		GGAGCCAACCAAATTACTGGGAA		Conserved region among HU1,18,20,25, 26,27,29,30,31
HU27E		GCTGGGATCAATTCGTTCTTGTA	YKNELIPA	
HU11B	x	GGTAThCCTTGyGGnGA	GIPCGE	
HU11B2	x	TGCGTATTTAThCCATGyGC	CVFIPCA	
HU11B2n	x	TTTAThCCATGyGCtACnGC	FIPCATA	
HU_MAKFTN	x	TTCGATGGCTAAATTCACCAACT	(DS)MAKFTNF	Signal peptide of HU13
HU19speEn		CTCATCTTCAGACGGCTTCATGAG	AHEAV-	
HU_MAKLTNI	x	GCTAAACTCACCAACATTCTCCT	AKLTNILL	Signal peptide of HU4
HU_MAFAVR	x	TTGCCGTAAGAAAGATGCTGAAT	MAFAVR	Signal peptide of HU40
HU33NTR		ACAGAAGAATAACAAGTAGAGAT	3′ UTR of hU33	
HU33Tail		TGTGGGTCTCGTCTTCAGTAGC	3′ UTR of hU33	
Hedyotis chrysotricha				
ChUncy B1	x	GCGTGGACACCTGCTTCATnTTy AAyTGyTA	VDTCFIFNCY	

ChUncy B2	x	CAACTGCTACTCCTCCTGCTGyGAr TGyAC	NCYSSSCECT	
ChCy B1	x	GGCTTCCCCGTCTGyGGnGArAC	GLPVCGET	
HC_MAKFTSY	x	ATGGCTAAGTTTACCAGCTATC	MAKFTSY	
HC_NTR2_E		TGTTTGGTCTCGTCTTCAGTC		Conserved region at 3′ end
HC11_B	x	GGTTTCCCTGTnTGTGGnGArAC	GFPVCGET	
HC11_Bn	x	TGTGGTGArACnTGCGTnGG	CGETCVG	
HC19_B	x	GGTwsnATTCCTTGTGGnGA	GSIPCGE	
HB19_Bn	x	TGCGTTTTCyTnCCnTGCAT	CVFLPCI	
HC25_B	x	TGTGGTGATwsnTGCGTnTTyAT	CGESCVFI	
Hedyotis diffusa				
D6 M1	x	GAyACTACTACTTAyGAyTGyGC	DTTTYDCA	
D6 M2	x	GArCGTTGyGTTTGGGGTCCnTG	ERCVWGPC	
D6 M3	x	GTnGGCTGCTCATGCGAyAC	VGCSCDT	
D6 M4	x	AArAThGTnGGCTGCTCATGCGA	KIVGCSCD	
D6p UTR		GGTTTATTTGGACGTTCTCGG ReverseCCGAGAACGTCCAAATAAACC	PRTSK-T	
D6p E		GAGTGATACGATCTTGACACAGTC	DCVKIVSL	
5end-D9	x	ATAAATTCAATGGCTGGCTTCACC	INSMAGFT	
5nest-D9	x	ATGGCTGGCTTCACCAACTATCTCC	INSMAGFTNYL	
5nest2-D9	x	CGCTTCTTGCAGCTTTTGTTGGG	SLLAAFVG	
D1 B1	x	TGTGGTGArACTTGCGGTmTnGGTAC	CGETCVIGT	.
D1n B1	x	GGTACTTGCTAyACTCCnGGnTGC	CVIGTCYTPGC	
Chassalia chartaceae				
CC7811ComB	x	TGCGTTTGGATHCCNTGCAT	CVWIPCI	
CC7SpeB	x	ATTCCTTGCATHACNGCNATHGC	IPCITAIA	
CC8SpeB	x	ATTCCTTGCATTWSNACNGTNAT	IPCISTVI	
CC11SpeB	x	CCTTGCATTWSNGGNATGTTYGG	PCISGMFG	
CC7SpeE		TCCACAACTTTAAATCCTCTCGC	3′ UTR of cC7	

CC8SpeE		GATTACGGCTGCACGGTCCAC	3′ UTR of cC8	
CC11SpeE		GGTCCACCACTTTAAATACTCA	3′ UTR of cC11	
CC_ComE		AAGGATAATTCACAAAAGGCAA		Common region for cC7,8,11
Panicum laxum				
HD1-F1	x	ATATGTGGNGARACNTGYGT	ICGETCV	
PL-F2	x	GCATTATGTGGNGARACNTG	AF CGETC	
PL-FN1	x	GGTACTTGTTAYACNCCNGG	GT CYTPG	
PL-FN2	x	GGTAAATGTTAYACNCCNGG	GK CYTPG	
PL5speR		ACACATCCAACAGTCCCACA	3′ UTR of pL5	
PL4speF	x	TAGAGGCCTAGTTCCCAACAGA	5′ UTR of pL4	
PL4speMAR	x	AACGCCATGGCTCGCAAATTGG	5′ UTR+signal peptide of pL4	
PL7,9speR		CACAAGCCATCTTAGCCTGCCA	3′ UTR of pL7,9	
PL1,2speR		ACACCACTCTTCTTATCCTCC	3′ UTR of pL1,2,6,8	
PL_QLAAPA	x	TGCAGCTNATGGCMGCTCCGA	QLAAPA	Conserved sequence in pro-region