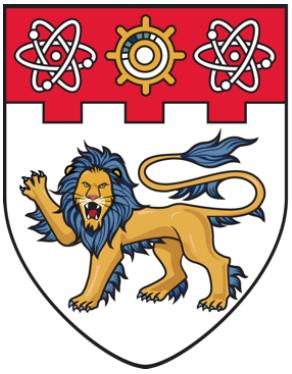


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

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

FAN TANG

SCHOOL OF BIOLOGICAL SCIENCES

2022

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

FAN TANG

SCHOOL OF BIOLOGICAL SCIENCES

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

2022

Statement of Originality

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

1st August 2022

. . . . . .

Date

. . . . . .

Fan Tang

Supervisor Declaration Statement

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

1st August 2022

Date

Professor James P. Tam

Authorship Attribution Statement

This thesis contains material from an abstract accepted at a conference in which I am listed as an author.

Chapter 5 is published as an abstract: Tang, F., Loo, S., Kam, A., & Tam, J. P. (2022). Cystine-stapled Helical Peptide from *Withania somnifera* Is Highly Stable and Cell-permeable. The FASEB Journal, 36.

The contributions of the co-authors are as follows:

- Prof James P. Tam initialized the project and edited the abstract.
- I drafted the abstract and did most of the experiments and analysis.
- The abstract was revised by Dr. Antony Kam and Dr. Shining Loo.
- Dr. Kaho Wong sequenced wS1 by mass spectrometry.
- Dr. Jingsong Fan determined the NMR structure of wS1.

1st August 2022

. . . . . .

Date

Fan Tang

. . . . . .

Fan Tang

Acknowledgments

I would like to express my sincere gratitude to my advisor Professor James P. Tam. Four years ago, he provided me the opportunity to start my Ph.D. journey. Afterward, he has been continuously supporting, patiently helping, and wisely suggesting me. He helped me gradually overcome mental barriers of doing research and started to realize the beauty of biological science. Also, I have to thank Professor Maggie Hoi and Professor Simon Lee who co-supervised me in my master's training and helped me join my current lab. They are widely respected and loved by their students. I often miss them and sincerely wish them all the best!

I would like to thank all the lab members including lab managers, graduate students, and Postdocs, for their contributions directly or indirectly to my research. My special thanks go to my lab mentors, Dr. Antony Kam and Dr. Shining Loo, as they helped me to get familiar with the new environment and to settle down. They are brilliant researchers who inspire me to think differently and make me dream of becoming a good researcher like them. My special thanks also go to my friends: Dr. Kaho Wang, Dr. Xinya Hemu, Miss Stephanie Victoria Tay, Miss Xia Yiyin, Miss WeiQin, Mr Benjamin Liew, Mr Shaun Tan, Miss Zhen Nee Fung, and Miss Huang Jiayi. Although some of them have left the lab already, their accompany made my Ph.D. studying journey more colorful.

Lastly, I have to thank my family for their contributions during these years. Without their sacrifices, I would not be able to healthily grow up, bravely face challenges, and confidently live my life. I wish my family happiness and health!

Table of Contents

List of Figures

List of Tables

Abbreviations

(±)-trans-1,2-bis(2-mercaptoacetamido)cyclohexane	BMC
1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine	MPTP
3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide	MTT
8C-cysteine-rich peptide	8C-CRP
Amino acid residues	Aa
Basic local alignment search tool	BLAST
Beijing Genomics Institute	BGI
Bioinformatic molecules	BMs
Chitin-binding	CB
Cysteine-rich peptides	CRPs
Different expression genes	DEGs
Dimethyl sulfoxide	DMSO
Dimethylformamide	DMF
Dithiothreitol	DTT
Dulbecco's modified eagle medium	DMEM
Endoplasmic reticulum	ER
Endothelial nitric oxide synthase	eNOS
Epidermal growth factor	EGF
Epidermal growth factor receptor	EGFR
Epidermal patterning factors	EPFs
ER oxidoreductase 1	ERO1
Ethylisopropylamiloride	EIPA
Expressed sequence tags	EST
Fetal bovine serum	FBS
Five flavor berries	FFBs
Fluorescein isothiocyanate	FITC
Gain-of-function	GOF
Gene Ontology	GO
Hepatocyte nuclear factor-1α	HNF-1 α
Hepatoma G2	HepG2
Heterozygous familial hypercholesterolemia	HeFH
Hevein-like peptides	HLP
High-performance liquid chromatography	HPLC
Homozygous FH	HoFH
Horseradish peroxidase	HRP
Human immunodeficiency virus	HIV
Human umbilical vein endothelial cells	HUVEC
Hydrogen chloride	HCl
Hydrophilic polyvinylidene fluoride	PVDF
Indian ginseng	IG
Interactive tree of life	iTOL

Isoelectric point	PI
Kyoto encyclopedia of genes and genomes	KEGG
Lactate dehydrogenase	LDH
Laser Scanning Microscopy	LSM
Lipid transfer proteins	LTP
Lipid transfer proteins	LTPs
Loss-of function	LOF
Low-density lipoprotein receptor	LDLR
Low-density lipoprotein	LDL
Low-density lipoprotein cholesterol	LDL-C
Maneb-Paraquat	MB-PQ
Mass spectrometry	MS
Matrix-assisted laser desorption/ionization-time of flight	MALDI-TOF
Mitogen-activated protein kinase	MAPK
Molecular weight	M.W.
Multiple sequence comparison by log- expectation	MUSCLE
N-Hydroxysuccinimide	NHS
National center for biotechnology information	NCBI
Natural products	NPs
Non-bioinformatic molecules	non-BMs
Non-chitin-binding	non-CB
Nuclear magnetic resonance	NMR
Optical density	OD
Oxidized glutathione	GSSG
Phosphate buffered saline	PBS
Proprotein convertase subtilisin/kexin type 9	PCSK9
Protein disulfide isomerase	PDI
Rapid alkalinization factor	RALF
Reduced glutathione	GSH
Resin	Rs
Side-chain protective group	SP
Signaling peptidase	SPase
Solid-phase peptide synthesis	SPPS
Stand deviation	SD
Sterol regulatory element	SRE
Sterol regulatory element-binding protein-2	SREBP-2
Tris-buffered saline	TBS
Union of Soviet Socialist Republics	USSR
US Food and Drug Administration	FDA
Western blotting	WB
Wisotide	wS
B-hydroxy β-methylglutaryl-CoA	HMG-CoA

Abstract

Plants produce an enormously diverse range of medicinally useful molecules, ranging from small-molecule metabolites to peptides, microproteins, proteins, and biopolymers. Among them, plant-derived peptides are not only under-explored, but also marginalized for their unstable stereotypes. This thesis describes my discovery and characterization of two highly stable and novel peptides, wisotide (wS1) from *Withania somnifera* and wuweizitide (wZ1) from *Schisandra chinensis*. Both wZ1 and wS1 belong to the superfamily of cysteine-rich peptides (CRPs) whose structures are stabilized by multiple cross braces of disulfides. Indeed, the structures of the two CRPs were elucidated and confirmed using NMR spectroscopy. In addition, they are highly resistant to proteolytic degradation. These desirable properties share similarities to small-molecule metabolites as drug-like molecules and dispel the misconception that peptides are inherently unstable.

Sequencing by mass spectrometry showed that wS1, with eight cysteines, contains 31 amino acids. Structural determination by NMR showed that wS1 displays the lybatide-like disulfide connectivities to support helical conformations. The helical conformation is a common structure motif found in proteins but rare in short peptides. Peptide stapling is a synthetic strategy to constrain the helical conformations of a short peptide, generally through covalent linkage of their side chains. They are also known to be cell-membrane permeable and capable of inhibiting intracellular protein-protein interactions. However, naturally-occurring stapled peptides are rare. Previously, we discovered lybatides which contain helices stapled by cystine. Moreover, lybatides form a new family of cysteine-rich peptides with novel disulfide connectivity. Here, my thesis shows that wS1 is the second member of cystine-stapled helical peptides. Importantly, the anionic cystine-stapled wS1 is cell-membrane permeable. Database

mining further revealed that >50 lybatide/wS1-like sequences are found in plants. Taken together, wS1 and the cystine-stapled helical peptides could represent useful leads of peptide biologics to probe intracellular protein-protein interactions.

Wuweizitide (wZ1) was found to contain 18 residues with six cysteines, giving wZ1 a 33% cysteine content. This hyperdisulfide peptide is also hyperstable and cell-membrane permeable. This thesis shows that wZ1 induces LDL uptake by an atypical mechanism of LDLR accumulation at lipid rafts, suggesting that it might lower circulating LDL-cholesterol by a novel mechanism. These findings not only expand our knowledge and understanding of CRP families, but also open new avenues for discovering hyperstable bioactive peptides as leads for orally-active therapeutics from medicinal plants.

CHAPTER ONE

Introduction

1.1 Natural products (NPs)-based drug discovery

NP is a revolutionary driving force in drug discovery and has greatly promoted therapeutic developments on human diseases. Up to 2021, over 328,000 NPs have been recorded into the Dictionary of Natural Products database, with ~10,000 new additions per year [1]. From 1981 to 2014, NPs or their analogs, like statins and their derivatives, account for ~25% of all drugs approved by the US Food and Drug Administration (FDA; **Figure 1.1**) [2-4]. In the small molecule category, NPs or their related compounds have over 33% occupation [2]. These nature-derived drugs are widely and intensively applied to a broad spectrum of human illnesses, such as microbial infection, malignant cancers, malaria, neurodegenerative diseases, and cardiovascular diseases. According to biological origins, NPs can be classified into three categories: 1) microorganism-derived NPs; 2) animal-derived NPs; 3) plant-derived NPs.

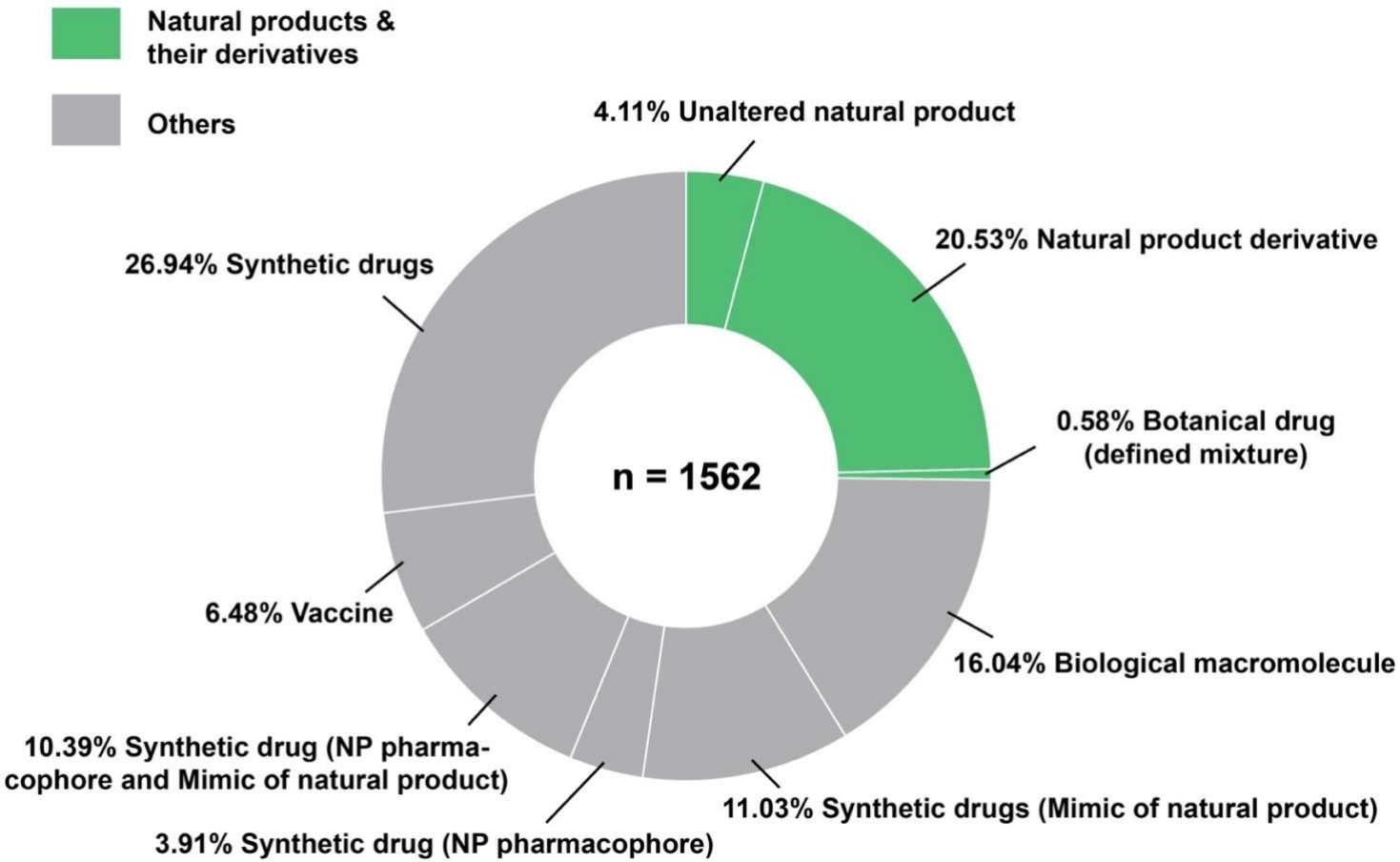


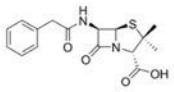
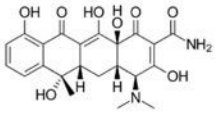
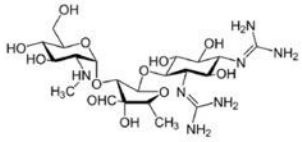
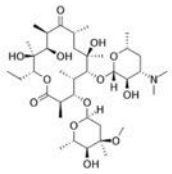
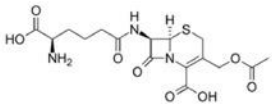
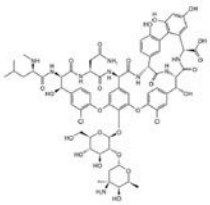
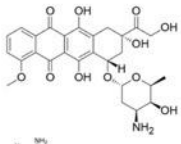
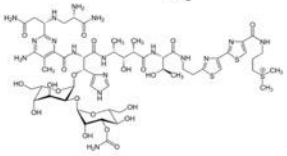
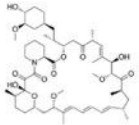
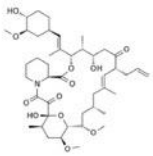
Figure 1.1 FDA-approved drugs between 1981 and 2014 in the US.

This figure was adapted from the work of David J. Newman et al. [2].

1.1.1 Microorganism-derived NPs

In 1928, Alexander Fleming, a Scottish microbiologist, discovered the penicillin from *Penicillium notatum* with potent anti-bacteria effects and then shifted scientific focus to NPs of microorganisms [5]. After the penicillin discovery, many molecules from microorganisms have been discovered and characterized around the world. In 1940s, two drugs with new scaffolds were discovered. One is chlortetracycline, a tetracycline isolated from *Streptomyces aureofaciens* [6], and another is streptomycin, an aminoglycoside obtained from *Streptomyces griseus* [7]. In 1950s, more antibiotics were discovered, including cephalosporin C (a cephalosporin antibiotic), erythromycin (a cyclic ketone antibiotic), and vancomycin (a glycopeptide antibiotic) [8, 9]. Nowadays, 69% of antibiotics are derived from NPs and ~60% of small molecules directly or indirectly originate from NPs [10, 11]. The microorganism-derived NPs not only have antibiotic activities, but also exhibit diverse other biological effects, such as anti-cancer and anti-inflammation (**Table 1.1**). Due to these functions, they are broadly used in fields, such as medicinal therapies, food preservations, agriculture developments, and scientific investigations.

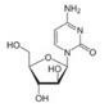
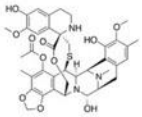
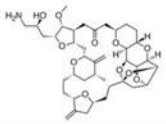
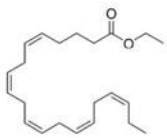
Table 1.1 Examples of natural products (NPs) from microorganisms.

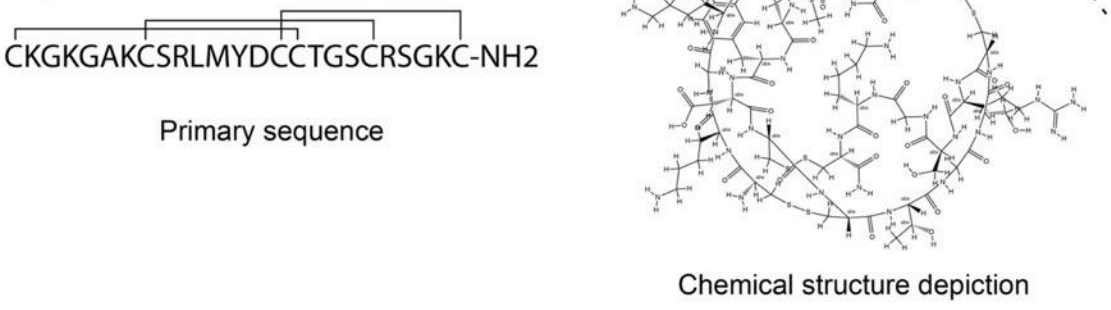
Diseases	Compounds	Origins	Structures	Year of dicovery
Microbal infection	Penicillins	*Penicillium moulds*		1928
	Tetracycline	*Streptomyces rimosus*		1948
	Streptomycin	*Streptomyces griseus*		1943
	Erythromycin	*Saccharopolyspora erythraea*		1952
	Cephalosporin C	*Acremonium*		1953
	Vancomycin	*Amycolatopsis orientalis*		1955
Cancer	Doxorubicin	*Streptomyces peucetius*		1950s-1960s
	Bleomycin	*Streptomyces verticillus*		1962
Inflammation	Sirolimus	*Streptomyces hygroscopicus*		1972
	Tacrolimus	*Streptomyces tsukubaensis*		1978

1.1.2 Animal-derived NPs

Animal NPs can be derived from the ocean, a place where life originates. The ocean covers over 70% of the earth surface and accommodates ~2.2 million eukaryotic species, of which 91% are unknown [12]. Out of 31 animal phyla available, 12 are exclusively found in the ocean [13]. These characteristics indicate the invaluable potential of marine animal sources for novel drug discovery. However, marine NPs are far from being adequately explored, and history of using marine medicines is not as ancient as terrestrial one, which is dating back to 2600 BCE [14]. To promote marine investigation, development of diving technology was started several decades ago, and this helped human activities to extend deep down to ~40 meters. In addition, other marine technologies, like submersible vehicles with remote control, were developed. These things make human exploration reach to an unprecedented level, and thus greatly promoted the NPs development from the ocean. In the past decades, different marine animal NPs from different structural classes were approved by FDA to treat diseases, including cancer, Alzheimer disease, hypertriglyceridemia, microbial infection, and neural pain. For example, cytarabine from *Cryptotheca crypta* was approved by FDA in 1969 for acute myeloid leukemia. Its mode of action is to competitively inhibit DNA polymerase in the S-phase of the cell cycle [15, 16]. In 2004, ziconotide from Conus magus was approved for ameliorating chronic pain. This ziconotide is a cysteine-rich peptide with three disulfide bonds and with high potency on blocking calcium ion channels [17]. These agents with other FDA-approved ones are listed (Table 1.2).

Table 1.2 Examples of FDA-approved drugs derived from marine animals.

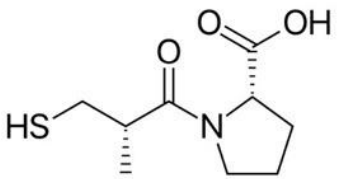
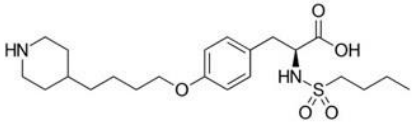
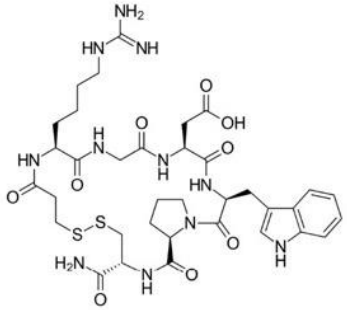
Diseases	Compounds	Marine origins	Structures	Year of approval
Cancer	Cytarabine	Cryptotheca crypta		1969
	Trabectedin	Caribbean sea squirt		2007
	Eribulin	*Halichondria okadai*		2010
Hyper-triglyceridemia	Omega-3-acid ethyl esters	Fish oil	Mixture	2004
	Ethyl eicos-apentaenoic acid	Fish oil		2012
	Epanova	Fish oil	Mixture	2014
Pain	Ziconotide	*Conus magus*	Cysteine-rich peptides	2004



Animal NPs can also be derived from land, a place with a long history of medicinal practices and experiences. On land, prevalent animal sources are often referred as venomous species, like snakes, spiders, scorpions, and frogs. In the world, it is estimated that venomous animals have over a million different species, and each species might produce a venom mixture containing up to 100 different compounds [18]. These compounds contain proteins, peptides, and other non-protein molecules, showing a broad spectrum of specific biological functions with high potency. The mixture of compounds not only helps venomous species, like snakes, to effectively capture their prey by envenomation, but also represents a rich library of bioactive compounds with tremendous potential for drug development.

For example, in the snake venom-based field, the first success came from the marketed Captopril®, for its approval by FDA in 1981 [19]. The captopril is a peptide obtained from *Bothrops jararaca* to treat hypertension via inactivating an angiotensin-converting enzyme [19]. Another two examples are tirofiban and eptifibatide. For tirofiban, it is an antiplatelet drug approved by FDA in 1998 for heart attacks [20]. This drug was developed from an Arg-Gly-Asp motif of a venous ingredient, disintegrins, from *Echis carinatus* [20, 21]. Similarly, eptifibatide is also an antiplatelet drug with its approval from FDA in 1998, and it has been granted to treat acute coronary syndrome [22]. The integrin was based on the Lys-Gly-Asp motif of disintegrins from *Sistrurus miliarus barbourin* [18, 23]. These examples together (**Table 1.3**) with those currently under clinical trials demonstrate that animal venom is a rich source of compounds with markable pharmacological potential.

Table 1.3 Examples of FDA-approved drugs derived from land animals.

Diseases	Compounds	Land origins	Structures	Year of approval
Hypertension	Captopril	*Bothrops jararaca*		1980
Platelet agglutination	Tirofiban	*Echis carinatus*		1999
	Eptifibatide	*Sistrurus miliarius barbouri*		2001

1.1.3 Plant-derived NPs

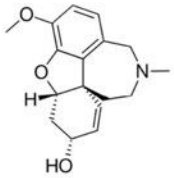
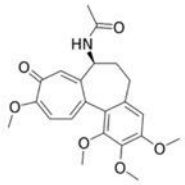
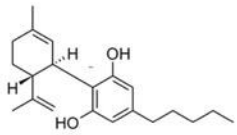
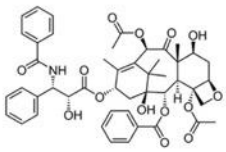
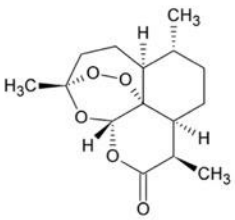
Medicinal uses of plants may start in eastern ancient civilizations. Dating back to 2600 BC, Mesopotamia included ~1000 plant-derived ingredients in their medicinal documents [24]. Plant-derived oils in the document are still in use today for diseases like cold, infections, and inflammations. Egyptian medicine can be traced back to 1500 BC and mainly recorded in the Egyptian “Ebers Papyrus” with over 700 medicines [25]. India has its medicine system, called “Ayurveda”, which might be started in 6000 BC [26]. In the system, *Astanga Hridaya*, *Charaka Samhita*, and *Susruta Samhita* books recorded hundreds of drugs [27, 28]. China has many ancient scriptures, and the oldest medicinal one is “52 Medicinal Prescriptions” which can be traced back to 1000 BC [29]. By trial and error, these ancient cultures developed their sophisticated medicine system and laid a solid basis of modern drug discovery.

Western uses of medicinal plants might begin in Greek and Roman civilizations. In the first century CE, Greek and Roman physicians (like Dioscorides and Pliny) wrote compendia on early Greco-Roman knowledge [30]. During the fifth to twelfth centuries, Arabs compiled the initial Greco-Roman knowledge, their own medicinal experience, and Chinese/Indian traditional medicine [31]. In the fifteenth and sixteenth centuries, the letterpress promoted resurrecting of Greco-Roman knowledge by putting together information from other books, such as *Herbarius Moguntinus* (1484), *The German Herbal* (1485), and *Herbarium Vivae Eicones* (1530) [30]. Up until the eighteenth century, the studies of aconite, colchicum, and foxglove switched previous empirical applications to modern rational investigation [30]. At the beginning of the nineteenth century, morphine became the first medicine with rational and systematic investigations, involving isolation, structural determination, and pharmacological

characterization [32]. Following the morphine discovery, the pharmacology field started to enter the era of NPs.

The twentieth century is an NPs century. Many revolutionary drugs were discovered and characterized by using modern biotechnologies. For example, between the 1950s and 1970s, Galantamine from *Galanthus* was discovered and used for Alzheimer's disease; vincristine from Catharanthus roseus was discovered for leukemia therapy; paclitaxel from *Pacific yew* and camptothecin from *Camptotheca acuminata* were applied for cancer chemotherapy; artemisinin from *Artemisia annua L* was shown effects on anti-malaria. These examples of NPs discovered are shown below (**Table 1.4**).

Table 1.4 Examples of approved drugs from plants.

Diseases	Compounds	Land origins	Structures	Year of approval
Alzheimer	Galantamine	Many plants, e.g. *Galanthus nivalis*		2004
Gout	Colchicine	*Colchicum autumnale*		1961
Epilepsy	Cannabidiol	*Cannabis*		2018
Cancer	Paclitaxel	*Pacific yew*		1993
Malaria	Artemisinin	*Artemisia annua*		-

1.1.4 Advantages and challenges of plant-derived NPs

Although NPs from plants historically revolutionized drug discovery and therapeutic development, pharmacology industry, in twenty-first century, starts losing enthusiasm on drug discovery in plants. Many reasons are there for this trend, such as technical obstacles of characterization and conventions on biological diversity [33], but the main reason might be the overwhelming exploration of phytochemicals, mainly represented by secondary metabolites or small molecules. In our research, we shifted the investigating focus from the overly-explored small molecules to under-explored peptidyl phytochemicals, cysteine-rich peptides.

1.2 Cysteine-rich peptides (CRPs) emerging as bioactive compounds

1.2.1 Roles of CRPs in plants

CRPs are unusual small peptides (usually less than 100 amino acid residues) with stable disulfide bonds and diverse sequences. The disulfide bonds render their scaffolds highly stable, and amino acid sequences with scaffold support in CRPs can exhibit multiple biological functions. In plants, CRP is mainly known for its functions on defending against external pathogens and mediating internal signaling, which is crucial for plant development. These biological functions or roles shall be discussed (**Table 5**).

Plants are constantly under stress from environmental pathogens. To confront the harmful attacks from pathogens, plants have developed their physical and chemical defense systems. Physical system can be built up by the strengthened cuticle and the hardened cell wall, which are natural barriers for pathogen spreading. Unlike physical barriers, chemical system does not resist pathogens by building the “physical wall”. This system is mainly responsible for supplying the front line with “chemical warriors”, like metabolites and peptides. For example, antimicrobial peptides are abundant in plants and encoded by 2-3% of genes in the genome of some plants [34]. These peptides are not only resistant to a broad spectrum of pathogens, but also often characterized by their cysteine-rich property, so-called antimicrobial CRPs. These CRPs are represented by thionins, defensins, HLP (hevein-like peptides), knottin, LTP (lipid transfer proteins), hairpinin, and snakins [35].

CRPs in plants are not always harmful to microbes. Instead, they are sometimes beneficial to microbes by establishing a symbiotic relationship with microbial plant hosts, like legumes. In this relationship, microbes provide plants with ammonium

(source of nitrogen); leguminous plants supply microbes with carboxylic acid (source of carbon) and accommodate them in nodules. CRPs bridge communications between microbes and legumes, at least in nodule development. Studies show that some CRPs, like ENOD3 and ENOD4, are encoded by nodule-specific genes in peas. In the nodule-forming process, these two CRPs begin to express at the early stage and their amount decreases after 16 days with occurrence of *nifH* (a bacteria nitrogenase) [36, 37]. In addition, over nodule-related 300 CPRs with highly conserved cysteine residues are identified from a nodule-forming legume (*Medicago truncatula*). As microbial invasion does not always activate the defense system in legumes [36], nodule-specific CRPs might be coordinators in a symbiosis between plants and microbes, instead of being anti-microbial peptides.

CPRs, like RALF (rapid alkalinization factor), actively take part in root growth. The RALF peptides were widely discovered in species like tobacco, sugarcane, and *Arabidopsis* [38-40]. RALF might bind to cell membrane receptors (ATPase) for alkalinizing the extracellular environment, because disrupting association between ligands and receptors (using suramin) can inhibit such alkalinization [41, 42]. This view was reinforced by other evidence: 1) RALF activates the MAPK signaling pathway [43]; 2) tomato RALF analog has a typical peptide-receptor dissociation (0.8×10^{-9} M) [43]; 3) RALF interacts with two proteins (25 kDa and 120 kDa) that might be receptors on the cell membrane [43].

In stomatal patterning, EPFs (epidermal patterning factors) are CRPs that control stomata formation in *Arabidopsis*. For instance, EPF1 is a negative regulator in stomatal development by acting on TMM and ER signaling pathways [44], while EPF1 mutation leads to a disorder of stomata distribution [45]. EPF1 is likely an excreted signaling CRP for neighboring cells of stomata precursors. The excreted EPF1 might

act on TMM and ER receptors of the neighboring cells for regulating stomatal patterns. More examples of stomatal regulating CRPs are STOMAGEN and CHALLAH. The former positively promotes stomatal formation, while the latter represses stomatal formation by directly interacting with stomata-related receptors [46].

Apart from playing a role in pathogen defense, microbe symbiosis, root growth, and stomatal patterning, CRPs also participate in plant reproduction, involving pollen tube formation, self-fertilizing prevention, and seed development [47]. These CRPs functions are based on regulating capacities of CRPs in cellular signaling communication, suggesting CRPs have high therapeutic potential in human diseases. Although possessing diverse biological functions, CRPs share high conservation in terms of disulfide connectivities, cysteine motif (arrangement and spacing), and cysteine numbers. Based on these conserved characteristics, we can classify CRPs into different families to seek their biosynthetic, structural, and functional similarities.

Table 1.5 Cysteine-rich peptides and their roles in plants

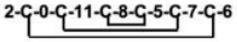
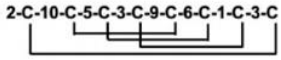
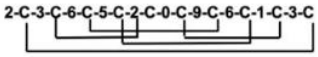
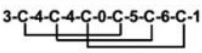
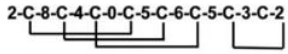
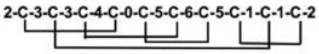
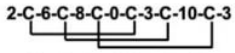
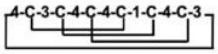
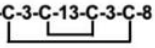
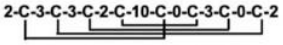
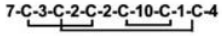
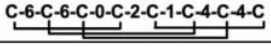
Name	Plant	Expression	Cysteine	Role	Function	Reference
RsAFP2	Radish	Pathogen induced	8	Defence	Pathogen defence	Terras *et al.*, 1992*a*
ENOD3/14	Pea	Infected nodules	4	Nodulation	Metal binding?	Scheres *et al.*, 1990*a*
RALFs	Tobacco	Roots	4	Development	Root elongation	Pearce *et al.*, 2001*b*
EPF1	*Arabidopsis*	Epidermal cells	8	Development	Stomatal positioning	Hara *et al.*, 2007, Hunt and Gray, 2009
EPF2	*Arabidopsis*	Epidermal cells	8	Development	Negative regulator of stomatal density	Hara *et al.*, 2009
STOMAGEN	*Arabidopsis*	Mesophyll cells	6	Development	Positive regulator of stomatal density	Sugano *et al.*, 2010, Kondo *et al*, 2010
CHALLAH	*Arabidopsis*	Epidermal cells	8	Development	Negative regulator of stomatal density	Abrash and Bergmann, 2010
LAT52	Tomato	Pollen	6	Reproduction	Pollen tube germination	Muschietti *et al.*, 1994
STIG1	Tomato	Pistil	16	Development	Positive regulator of pollen tube growth	Tang *et al.*, 2004
SCA/LTP5	Lily/*Arabidopsis*	Stigma, style	8	Reproduction	Pollen adhesion to pistil tube track	Mollet *et al.*, 2000; Chae *et al*, 2009
SCR/SP11	*Brassica*	Pollen	8	Reproduction	Male determinant of SI	Schopfer *et al.*, 1999
PrsS	Poppy	Stigma	4	Reproduction	Female determinant of SI	Wheeler *et al.*, 2009
LUREs	*Torenia*	Synergid cell	6	Reproduction	Pollen tube attractants	Okuda *et al.*, 2009
ES4	Maize	Synergid cell	6	Reproduction	Pollen tube rupture	Amien *et al.*, 2010
BAPs	Maize	Endosperm	4	Reproduction	Antifungal?	Serna *et al.*, 2001
AE1/3	Maize	Endosperm	8	Reproduction	Embryo/endosperm crosstalk?	Margnard *et al.*, 2000
MEG1	Maize	Endosperm	8	Reproduction	Transfer cell development	Costa *et al*, unpublished data

Note: This table is taken from the work of E Marshall et al. [29].

1.2.2 Families of plant CRPs

Among diverse CRPs in plants, we are interested in CRPs of 2 to 6 kDa, with six to eight cysteine residues. In this range, CRPs can be classified into 15 different families that are unique in their cysteine motifs with evolutionary conservation (**Table 1.6**). In my thesis, the following part is limited to plant CRPs from these families.

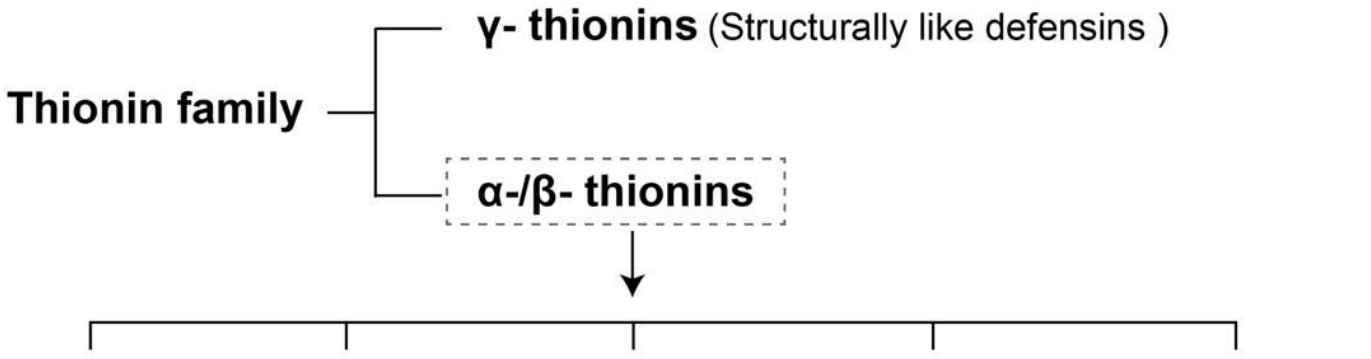
Table 1.6 Plant CRP families and their representative members.

	Representative member			
CRP Family	No. of Disulfide	Peptide Name	AA No.	Disulfide Motif
6C-Thionin	3	Crambin	46	
8C-Thionin	4	β-Purothionin	45	
8C-Defensin	4	NaD1	47	
10C-Defensin	5	PhD1	47	
6C-HLP	3	Ac-AMP1	29	
8C-HLP	4	Hevein	43	
10C-HLP	5	EAFP1	41	
Knottin	3	PAFP-S	38	
Cyclotide	3	Kalata B1	29	
α-Hairpinin	2	Ec-AMP1	37	
Jasmintide	3	jS1	27	
β-ginkgotide	3	gB1	20	
Lybatide	4	Lyba2	33	
Potentide	3	pA3	35	
Ginsentide	4	TP1	31	

Note: This table is modified from the work of James P. Tam et al. [48].

1.2.2.1 Thionin family

Thionins are antimicrobial cysteine-rich peptides with six to eight cysteine residues, 45-48 amino acid residues, and symmetric disulfide bonds. In 1977, purothionin as a prototypic thionin was discovered from *Triticurm vulgare* [49]. Following this discovery, other thionins (α-/β-/γ- thionins) were identified from different tissue of monocots and dicots, including cereal (endosperm), *Pyrularia pubera* (leaves and nuts), *Hordeum vulgare* (leaves), mistletoe (leaves and stems), and *Crambe abyssinica* (seeds) [50]. α-/β- thionins are highly homologous at their primary, secondary, and tertiary structures. These two thionins are classified into five types (I, II, III, IV, and V**; Figure 1.2**), according to their charges, cysteine motifs, residue numbers, and origins [51, 52]. Type I is basic with 45 residues from cereals versus 46-47 residues of basic type II from *Pyrularia pubera* [53]. Unlike type I and II have four disulfide bonds with similar charges, type III and IV possess three disulfide bonds, with different net charges (i.e. basic type III versus neutral type IV). For origins, type III and IV are from *Viscum album* (European mistletoe) and *Crambe abyssinica*, respectively. Lastly, type V is a truncated thionins with unknown functions from cereals, like wheat. Unlike α-/β-thionins, γ- thionins have different structures that are similar to defensins. For this reason, they were suggested to be placed in the defensin family [54]. Initially, thionins are recognized as toxins, due to their toxicities on microbes (e.g. bacteria and fungi) and some animals (e.g. insects) [55-58]. Although the mode of action is still unclear, positive/hydrophobic thionins might disrupt the membrane integrity by interacting with negative phospholipids [54, 59-61].



Type	I	II	III	IV	V
Origin	Cereals	*Pyrularia pubera*	Viscum album (European mistletoe)	*Crambe abyssinica, Phoradendron tomentosum*	*Triticum aestivum* (Wheat)
Residues	45	46-47	46	46	37
Disulfide bonds	4	4	3	3	3
Net charges	Basic	Basic	Basic	Neutral	Neutral
Examples	Purothionin, hordothionin, avenothionin	Pp-TH	Viscotoxin	Crambin, Phoratoxin	TTHV

Figure 1.2 Different types of thionins and their sequence properties.

1.2.2.2 Defensin family

“Plant defensins” term was named after animal defensins [62], which contain three types (α, β, and θ) with tissue distributing specificities (**Figure 1.3**) [63]. Like defensins from animals (e.g. invertebrates), plant defensins are generally 2-5 kDa in size with potent antimicrobial properties. As antimicrobial CRPs in plant innate immunity, plant defensins might exhibit their microbe-killing functions by at least two folds: 1) the cationic property or basic isoelectric point; 2) sequence variation. Cationic defensin can electrostatically interact with anionic lipid components in the microbial membrane, leading to membrane pore formation or its biogenesis inhibition, which causes death of microbes during their invasion. This speculated mode of action is based on studies of MtDef5 from *Medicago truncatula*. MtDef5 is an antibacterial defensin with γ-core motif, and mutation of cationic residues in this motif leads to loss of antibacterial activities [64]. Sequence variation not only can provide defensin with a broad spectrum of anti-microbial capacities, but also render defensin promiscuous, meaning possessing diverse biological functions [65]. For example, other than anti-microbe, some defensins also show activities on promoting pollen tube formation, exhibiting amylase-like properties, and increasing mental tolerance [66, 67].

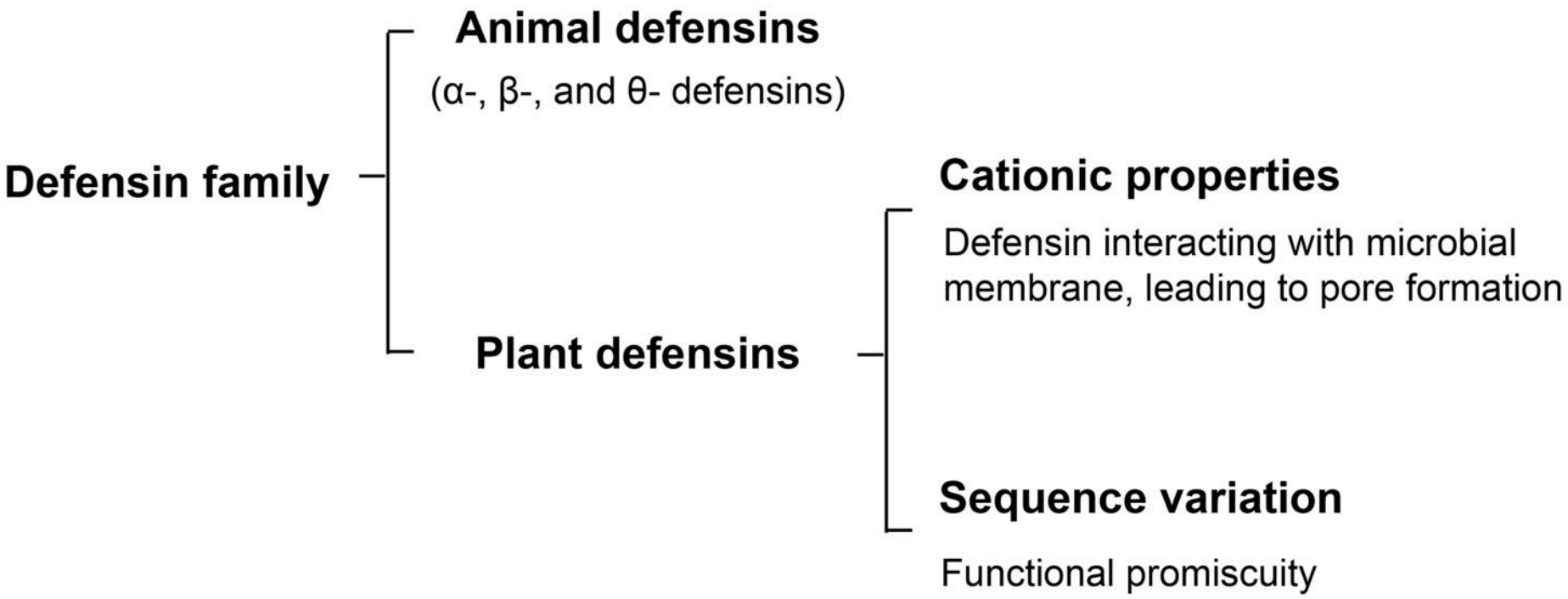


Figure 1.3 Classification of the defensin family and functional properties of plant defensins.

1.2.2.3 Hevein-like peptide family

In 1960, hevein as a novel CRP was identified from latex of *Hevea brasiliensis* [68]. For its chitin-binding domains, the hevein CRPs can inhibit many chitin-containing fungal strains, such as *Botrytis cinerea, Fusarium oxysporum,* and *Pyricularia oryzae* [69]. In 1998, hevein-like peptides (HLPs) were coined after discovering two hevein CRP homologs (Pn- AMP1 and Pn-AMP2) from *Pharbitis nil L.* [70]. Based on their functional properties, HLPs can be classified into chitin-binding and non-chitin-binding HLPs , which can be further divided into 6C-, 8C-, and 10C- subfamilies (**Figure 1.4**). Chitin-binding HLPs are known for their chitin-binding domain at the third and fourth inter-cysteine loops (a loop refers to the sequence between cysteine residues. For example, the third loop is between the third and fourth cysteine residues, which are in ascending order from the N to C terminals). In the third and fourth loops, chitin-binding HLPs have a highly conserved motif, SXφGφCGX4φ, in which X means any amino acid residues and φ represents aromatic residues [71]. Due to this motif, chitin-binding HLPs can target a cell wall component, chitin, of many fungi and exhibit a wide spectrum of anti-fungal activities [35]. In chitin-binding HLPs subfamilies, 6C is thought to be a truncated variant of 8C, while 10C is regarded as an adding variant of 8C and this addition shuffles cysteine pattern of 10C. To date, 6C-chitin-binding HLPs contain over 10 characterized members, such as Ac-AMPs (*Amanranthus caudatus*) [72], lWF-4 (*Beta vulgaris*) [73], and altides (*Alternanthera sessilis*) [74]. 8C-chitin-binding HLPs have over 20 characterized members, such as hevein (*Hevea brasiliensis*) [68], ginkgotides (*Ginkgo biloba*) [75], and Fa-AMPs (*Fagopyrum esculentum*) [76]. 10C-chitin-binding HLPs have five characterized members, involving EAFPs (*Eucommia ulmoides*) [77], WAMPs (*Triticum kiharae*) [78], and Ee-CBPs (*Euonymus europaeus*)

[79]. These few characterized chitin-binding HLPs indicate they are not only structurally, but also functionally underexplored.

Non-chitin-binding HLPs have the same cysteine pattern and disulfide connectivities as the ones in chitin-binding HLPs, but lack the chitin-binding domains in loop 3 and 4. So far, non-chitin-binding HLPs have 23 members isolated from five species (*Hibiscus sabariffa*, *Pereskia bleo*, *Panax ginseng*, *Panax nontoginseng*, and *Panax quinquefolius*). In *Hibiscus sabariffa*, our lab firstly discovered a group of non-chitin-binding HLP, termed as roseltides (rT1-rT8) [80]. The roseltides are 27 to 39 residues in length and identified as non-chitin-binding HLPs (a term created in 2017) with six cysteine residues [80]. In other words, roseltides belong to 6C-non-chitin-binding HLPs. Following the discovery of roseltides, our lab reported another group of 6C-non-chitin-binding HLPs from *Pereskia bleo* in 2017 and then coined “non-chitin-binding HLPs” [81]. This group of HLPs is termed as bleogens (pB1-pB5) with 27–40 residues in length. Unlike roseltides and bleogens have six cysteine residues, ginsentides (TP1-TP14) were shown to possess eight cysteine residues, and they are a suite of 8C-non-chitin-binding HLPs from three ginseng species (i.e. *Panax ginseng*, *Panax nontoginseng*, and *Panax quinquefolius*) [48]. However, 10C-non-chitin-binding HLPs are still unknown, suggesting our knowledge of non-chitin-binding HLP family should be expanded in the future.

Collectively, HLPs can be divided into two groups: chitin-/non-chitin-binding HLPs. To date, chitin-binding HLPs have 38 characterized members and non-chitin-binding HLPs have 23 characterized members. These reported members are relatively low compared with other CRP families, indicating that they might exhibit unique functional properties.

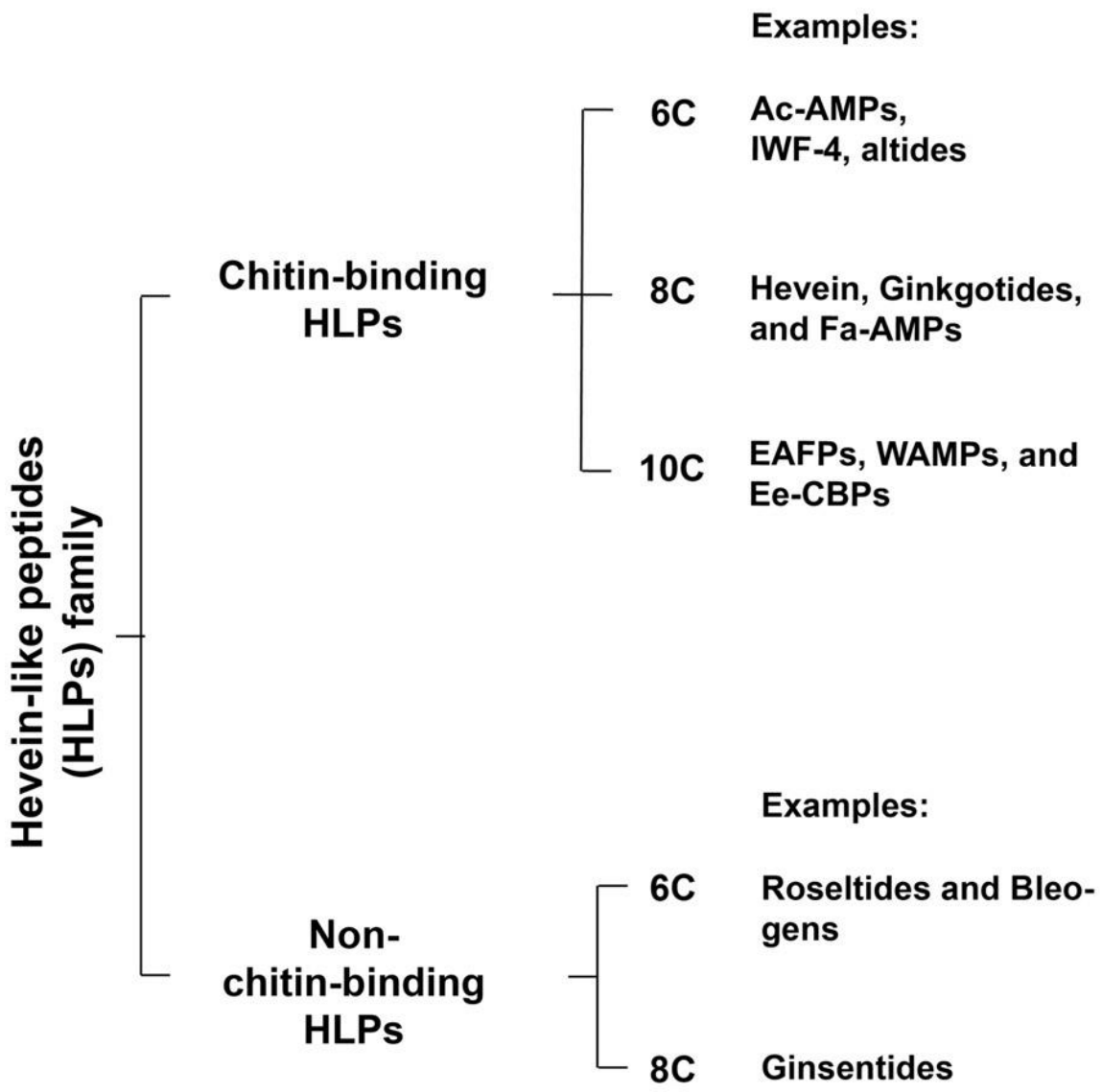


Figure 1.4 Classification of the hevein-like peptides (HLPs).

1.2.2.4 Knottin family

The first knottin was discovered as an inhibitor of potato carboxypeptidase in 1982. In following decades, many knottin-type peptides have been identified, characterized, and synthesized from different organisms. According to the knottin database (https://www.dsimb.inserm.fr/KNOTTIN/), 3320 knottin sequences from 654 organisms are identified (until 29 December, 2021). These organisms cover microbes (bacteria and fungi), animals (e.g. agouti, spiders, and insects), and plants (e.g. maize). In plants, knottins are typically ~30 amino acid residues in length with three disulfide bonds. Plant knottins can be generally classified into two groups, which are cyclic and linear knottins (**Figure 1.5**), based on whether their backbones at N and C terminals are cyclized or ligated by ligases (e.g. butelase 1 [82]). Cyclic knottins subfamily is represented by cyclotides, which have 35 determined NMR structures in the KNOTTIN database, such as structures of Circulin a, Kalata b1, and Palicourein. These cyclotides are derived from many plant species, such as *Chassalia parviflora*, *Viola odorata*, and *Oldenlandia affinis*. Functionally, these cyclotides possess diverse biological activities, involving anti-HIV, cytotoxic, and uterotonic. In contrast to cyclic knottins, linear knottins have no cyclized backbones, and they are either acyclic (meaning with sequence homology to cyclic knottins, but without head-to-tail backbone cyclization) or native linear (meaning unlike cyclotide in terms of sequences and cyclization). Acyclic knottins are exemplified by acyclotides, such as violacin A from *Viola odorata* and hedyotide B2 from *Hedyotis biflora* [83, 84]. Native linear knottins, for example, have alstotides from *Alstonia scholaris* and wrightides from *Wrightia religiosa* [85, 86]. Thus, knottins present a special CRP family with both cyclic and linear subfamily, and possess multiple biological functions, indicating their potential as agricultural and medicinal uses in the future

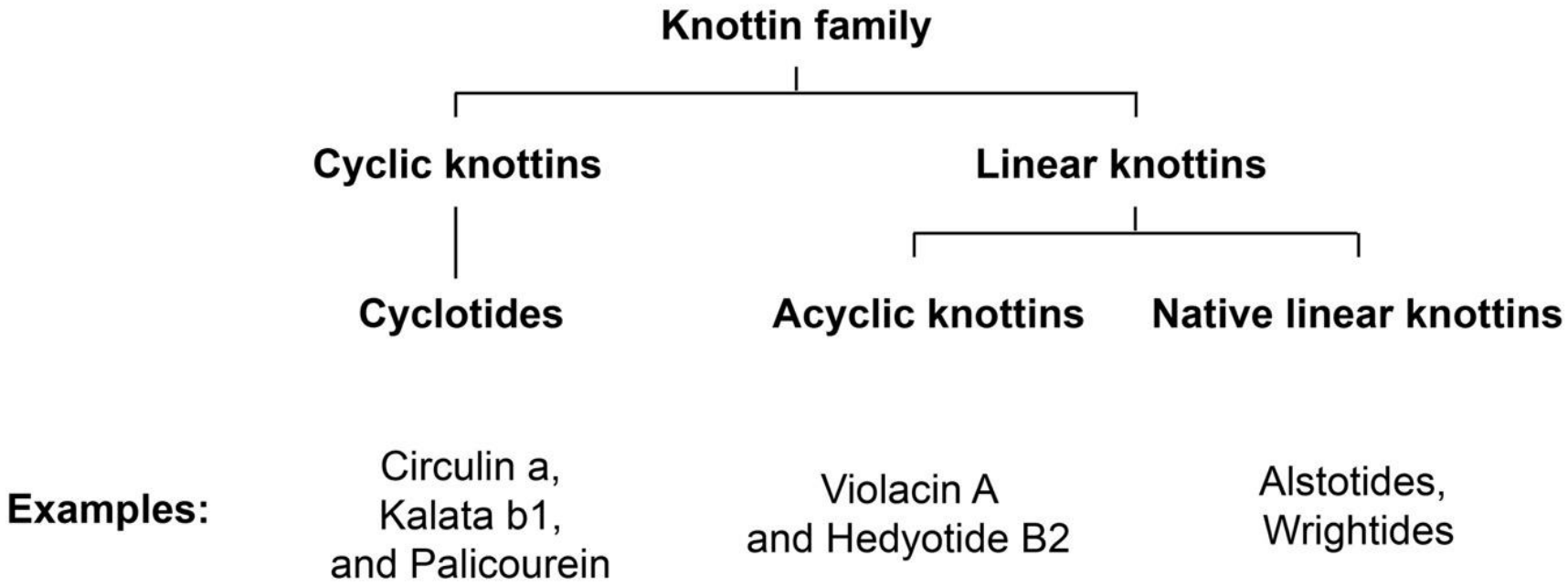


Figure 1.5 Different members in the knottin family.

1.2.2.5 Other plant CRP families

Classified plant CRP families also have α-hairpinin, lipid transfer proteins (LTPs), and snakins (**Figure 1.6**). Although being unique in their structures at primary to tertiary levels, these CRP families, so far, have few members reported. α-hairpinin is a family of Lys/Arg-rich CRP with cystine-stabilized antiparallel α-helices. This family has limited members (~10 members) reported, such as MBP-1 from *Zea mays L.* [87], MiAMP2s from *Macadamia integrifolia* [88], Ec-AMP1 from *Echinochloa crus-gali* [89]. LTPs (e.g. LTP1 and LTP2) are another CRP family with unusual molecular weight (>7 kDa) [90], compared with usual CRPs (< 6 kDa) from other families like thionin, defensin, knottin, and hevein-like peptides. LTPs exist in various plant species, like radish, maize, and wheat [90-92]. They are cationic 8C-CRPs that were thought to play a lipid-transferring role between intracellular organelles, but inconsistency was observed in newly discovered LTPs [93]. Another CRP family with a molecular weight of over 7 kDa is snakin. The snakin is a class of AMPs with 12 disulfide bonds in the plant defense system. For example, snakin-1 and -2 from *Solanum tuberosum*, as defense barriers, protect plants from pathogen invasion [94, 95]. In addition, novel CRP families have been discovered in our laboratory, including jasmintide, ginkgotide, lybatide, potentide, and ginsentide (**Figure 1.6**). These work enriched our knowledge of CRP families and triggered our interest in function investigation for future medicinal use.

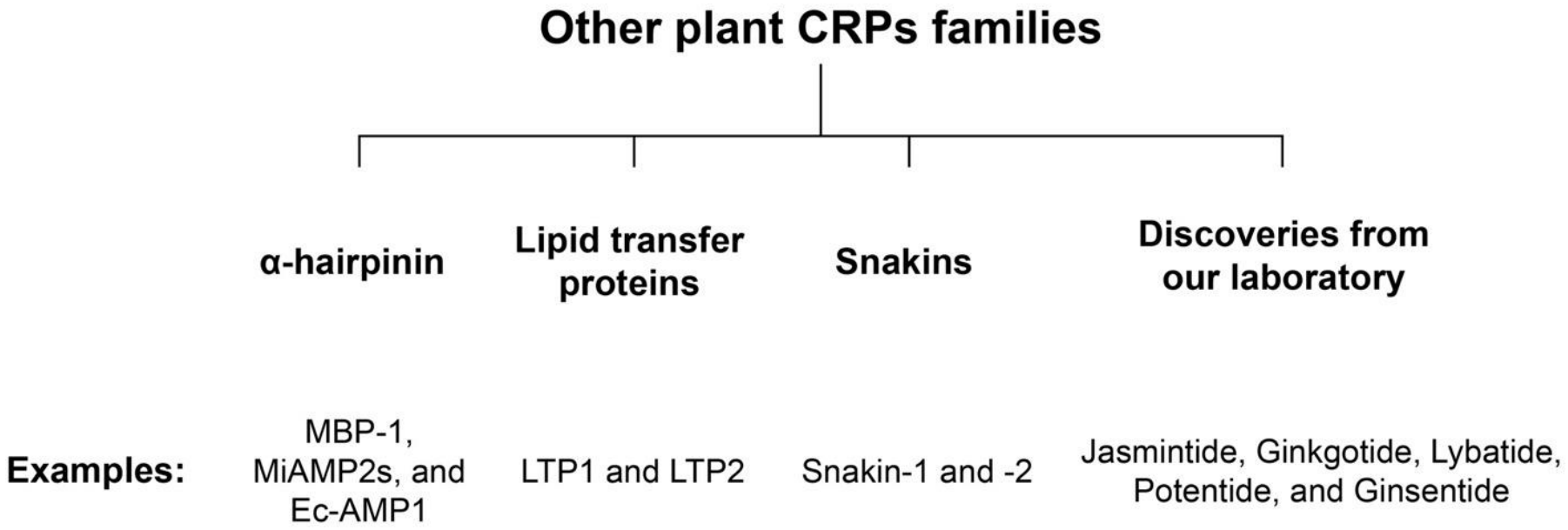


Figure 1.6 Other plant CRPs families with relatively few members.

1.2.2.6 Statistics of major CRPs in plants

In plant CRP families, defensin and cyclotide are the major CRP members. By searching PubMed database with “defensin” or “cyclotide”, we found that thousands of academic papers have been published and this number is still increasing. Specifically, until 16th December 2022, “defensin” has 8460 related publications since 1985. “Cyclotide” has 510 related publications since 1948. In contrast, “lybatide” family has only two papers published (in 2017 and 2022), and “wuweizitide” family has no publications so far (**Figure 1.7**), suggesting these two CRP families are rare *in planta* and their unusual scaffold might provide new engineering possibilities.

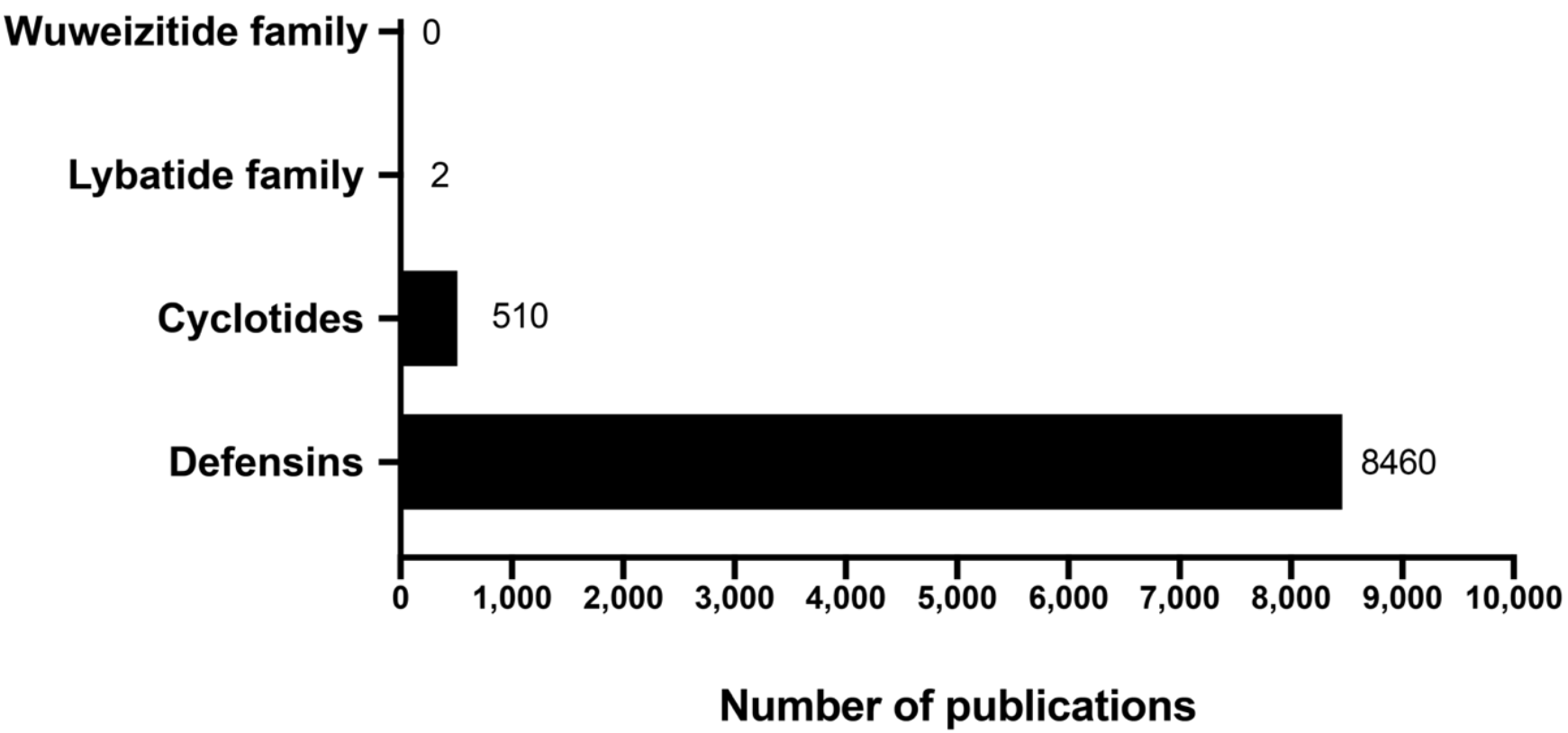


Figure 1.7 Publications of our current studies (wuweizitide and lybatide families) and the major plant CRPs.

1.2.3 General classifications of plant CRPs

For plant CRP general classification, other than the number of cysteine residues (i.e. 6C, 8C, and 10C), disulfide connectivities and cystine pattern are other two key factors. Based on these two factors, these CRPs, in theory, have hundreds of unique forms at the primary sequence level. For example, taking only disulfide connectivity into account, we found that a CRP containing 6-10 cysteine residues can exhibit 176 different connectivities. Since these diverse cysteine motifs directly determine three-dimensional structures and functional engineering possibilities, it is worthwhile to have a clear understanding of them.

1.2.3.1 Disulfide connectivity-based classification

Disulfide connectivities of plant CRPs have at least six different forms (type I-V; **Figure 1.8**). Unlike other forms that are rarely characterized, type I (cystine-knot like connectivity) and type II (thionin-like connectivity) are major classes. Type I that can be further divided into type I a and type I b is the commonest in plant CRPs, as demonstrated by a large amount of studies on defensins and cyclotides (a type of knottins) in the NCBI database (**Figure 1.7**). At the primary level, type I a has six cysteine residues and specific disulfide linkages (Cys I-IV, Cys II-V, and Cys III-VI). At the tertiary level, this knottin-like linkage, however, may display differently. For example, knottins and hevein-like peptides are with similar in terms of connectivities at the primary level, but they may differ at the tertiary level (their structural characteristics will be discussed later in this chapter). Unlike type I a with six cysteine residues, type I b has eight or ten that can form four or five disulfide bonds, respectively. These additional disulfide bonds in type I b do not intervene formation of Cys I-IV, Cys II-V, and Cys III-VI. Thus, we could consider 8C-type I b as a derivative CRP from 6C-type I a by having one more disulfide bond.

In contrast to asymmetric type I, type II is symmetric and can be further divided into type II a (6 C), type II b (8 C), and type II c (4 C), according to their number of cysteine residues contained. In type II, the simplest connectivity is type II c, represented by α-hairpinin with two symmetric disulfide bonds [96]. Type II a, like 6C-thionin [97], has a disulfide linkage of Cys I-VI, Cys II-V, and Cys III-IV. Type II b, like 8C-thionin [98], possesses one more symmetric disulfide bond formed between first and last cysteine residues (Cys I and Cys VIII), compared with 6C-thionin. The additional disulfide bond does not have to be symmetric, like in lybatides [99], a CRP with one more asymmetric disulfide bond between Cys I and Cys VI based on symmetric three disulfide bonds. These 8C CRPs with additional asymmetrical/symmetrical disulfide bonds, in a sense, could be treated as structural derivatives of 6C thionin or 4C α-hairpinin. The relationship between these 4C, 6C, and 8C type II presents an interesting evolutionary mystery.

Other CRPs with unique disulfide connectivities were type III jasmintides, type IV potentides, and type V β-ginkgotide (**Figure 1.8**). A limited number of plant CRPs members in these novel types (III-V) indicates that plant CRPs connectivities are unusual and could provide novel scaffolds for peptidyl drug carriers. Also, six types of discovered connectivities, in contrast to theoretical 176 different connectivities, suggest one possibility that these types are unctionally selected by nature in the evolution of plants.

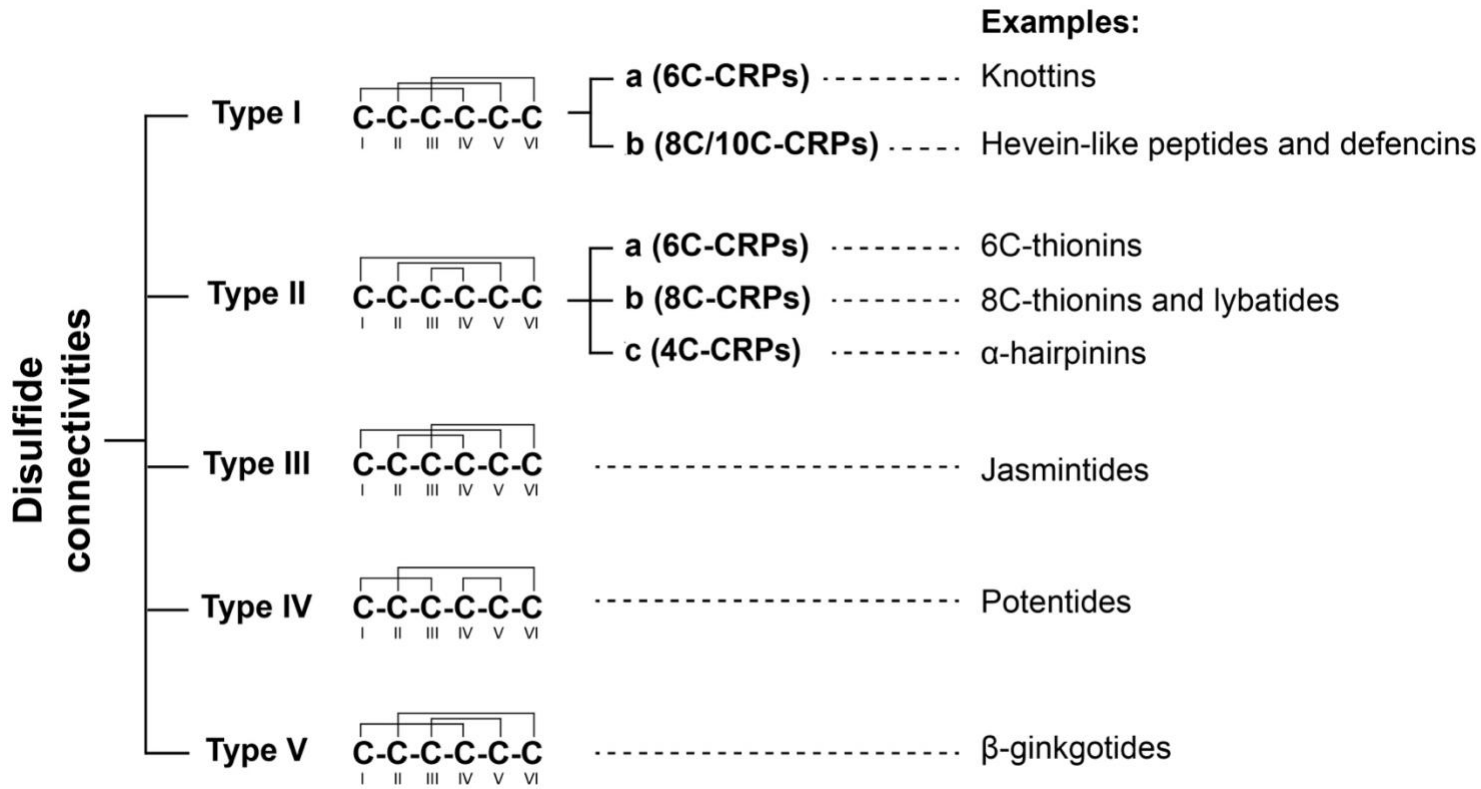


Figure 1.8 Disulfide connectivity-based classification of plant CRPs.

This figure is modified from the work of J. Huang (Ph.D. thesis, Nanyang Technological University, 2019).

1.2.3.2 Cysteine pattern-based classification

Cysteine pattern is formed by cysteine residues and intercysteinyl loops that might or might not carry functional peptidyl segments. As the length of these segments various, we decide to classify these cysteine patterns, instead of using the number of amino acid residues in each loop, based on the number of successive cysteine residues (or number of "-CC-"). From currently known plant CRP families, they can be classified into three classes: 1) non-successive Cys residues; 2) one pair of successive Cys residues; 3) two pairs of successive Cys residues (**Figure 1.9**). The first class has non-successive Cys residues in their cysteine pattern and contains four to eight cysteine residues, such as α-hairpinin with four, potentide with six, and defensin NaD1 with eight. The second class can be seen in 6C-, 8C-, and 10C- families, and in this class, one Cys pair can be formed in different places of cysteine pattern. For example, 6C-thionin crambin has the Cys pair at the N terminal; 6C-hevein like peptides have the Cys pair in the middle; jasmintides possess the Cys pair at the C terminal. The third class of CRP has two pairs of Cys residues and only two members have been isolated and characterized so far. These two members are the 6C β-ginkgotide and the 8C lybatide. These cysteine patterns determine the intercysteinyl loop length or size that is one of the key determinants for peptide functions, as the length might affect the holding capacity of sequence information and peptidyl structural conformations.

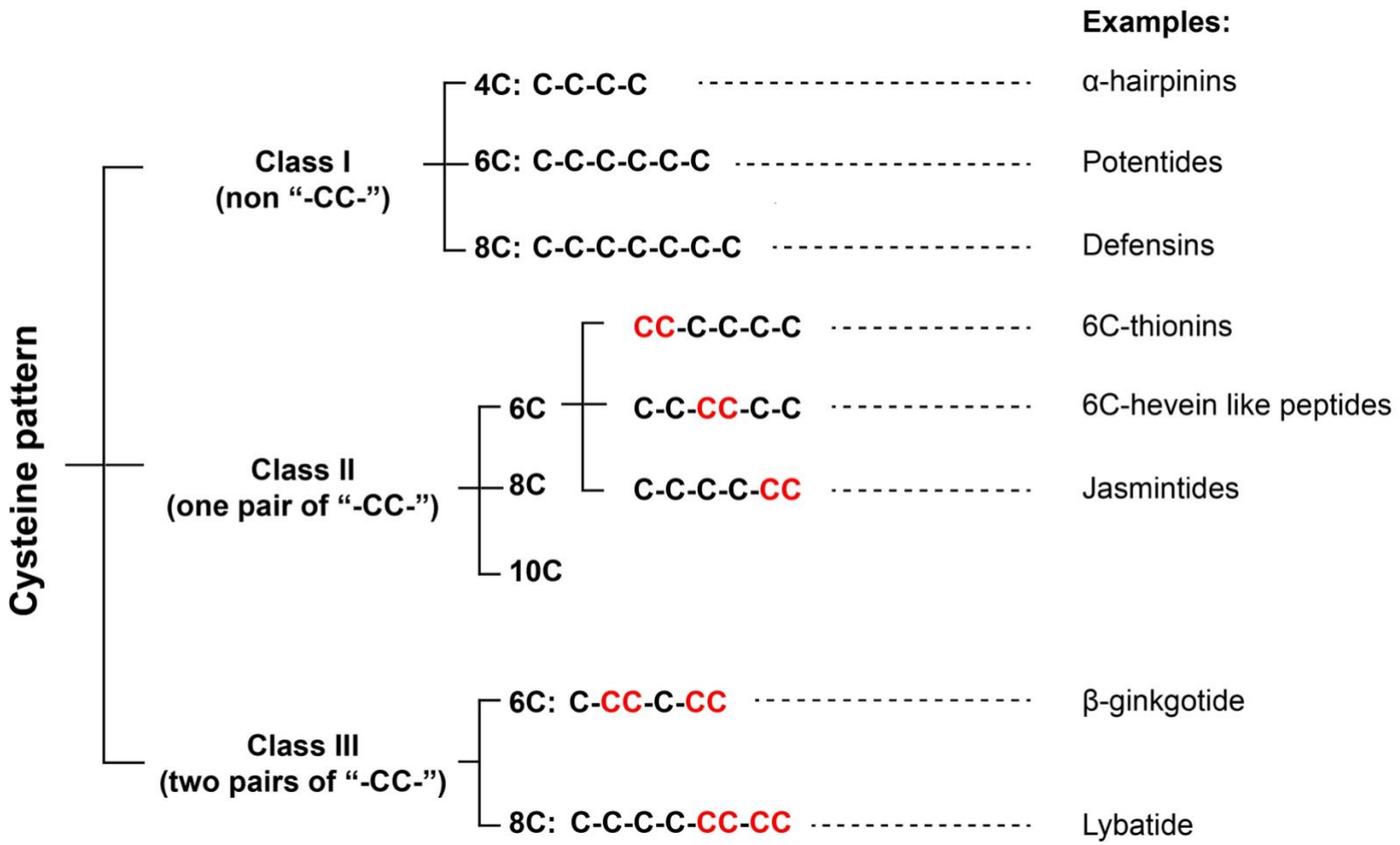


Figure 1.9 Cysteine pattern-based classification of plant CRPs.

1.2.4 Biosynthesis of CRPs

1.2.4.1 Translocating pathways to the endoplasmic reticulum (ER)

CRP precursors are biosynthesized by the cytoplasmic ribosome and then translocated into ER. In this translocating process, different excretory pathways could be involved, such as the Sec pathway, the SRP pathway, and the Tat pathway [100]. The Sec pathway can translocate unfolded proteins post-translationally. Accompanied by Sec machinery (e.g. Sec A, E, G, Y, and DF) [101], the unfolded proteins were guided to ER for excretion. The SRP pathway occurs co-translationally and involves machinery like SRP complex, SRP receptors, and Sec 61 [102, 103], to ensure an efficient translocation of proteins into ER. Lastly, the Tat pathway is responsible for translocating mature proteins. In this pathway, although participating machinery varies in different species, Tat A and C are treated as essential [104]. After ER localization via different pathways, the proteins start to fold or refold into a mature form, to be post-modified, and to be excreted into different intracellular locations or extracellular places.

1.2.4.2 CRP precursor domains

CRP precursors can in maximum exhibit four domains: signal domains, pro domains, mature domains, and c tails (**Figure 1.10 A**). Being translocated into ER, the signal domain (or signal peptide) provides protein couriers (i.e. machinery in secretory pathways) with a precise delivery address. In other words, only with signal peptides in precursors can proteins enter excretory pathways [105]. The signal peptides are generally ~25 amino acid residues in length with three regions: N-region, H-region, and C-region (**Figure 1.10 B**) [105]. N-region at the amino terminus is 1-5 residue long. By being positively charged, this N-region has at least three functions: 1) binding to the phosphate group on the ER membrane, an essential step for protein translation [106]; 2) together with C-region, conferring basic-to-neutral orientation of signal sequence [107] and making recognition of cutting-site effortless for SPase I; 3) preventing signal disorders caused by other signal peptides, as others (e.g. mitochondria-/chloroplast-organelle-targeting) could be more basic [108]. For the H-region, it is hydrophobic with 7-15 residues in length [105]. This region tends to form α-helical conformation, together with its hydrophobic property, contributing to its membrane-spanning capability. Also, H-region affects the translocation efficiency of proteins and determines ER translocating pathways [100, 109, 110]. The C-region normally consists of 3-7 hydrophilic and no charge residues [105]. As the cleavage site should not be far from the membrane binding domain (e.g. H-region), the C-region is not long [100]. Although the sequence of signal peptides varies, the cleavage site of C-region is highly conserved among different species. Specifically, the presence of Ala at the C terminal of C-region is crucial for the recognition by SPase [100]. The well-conserved cleavage site and highly-diverse signal sequences not only can ease the enzyme supply for cutting, but also can specifically translocate characteristic

proteins in cells. In addition, other domains such as pro domain and c tails also take part in the protein translocation, besides their roles in organelle targeting and structural folding [111].

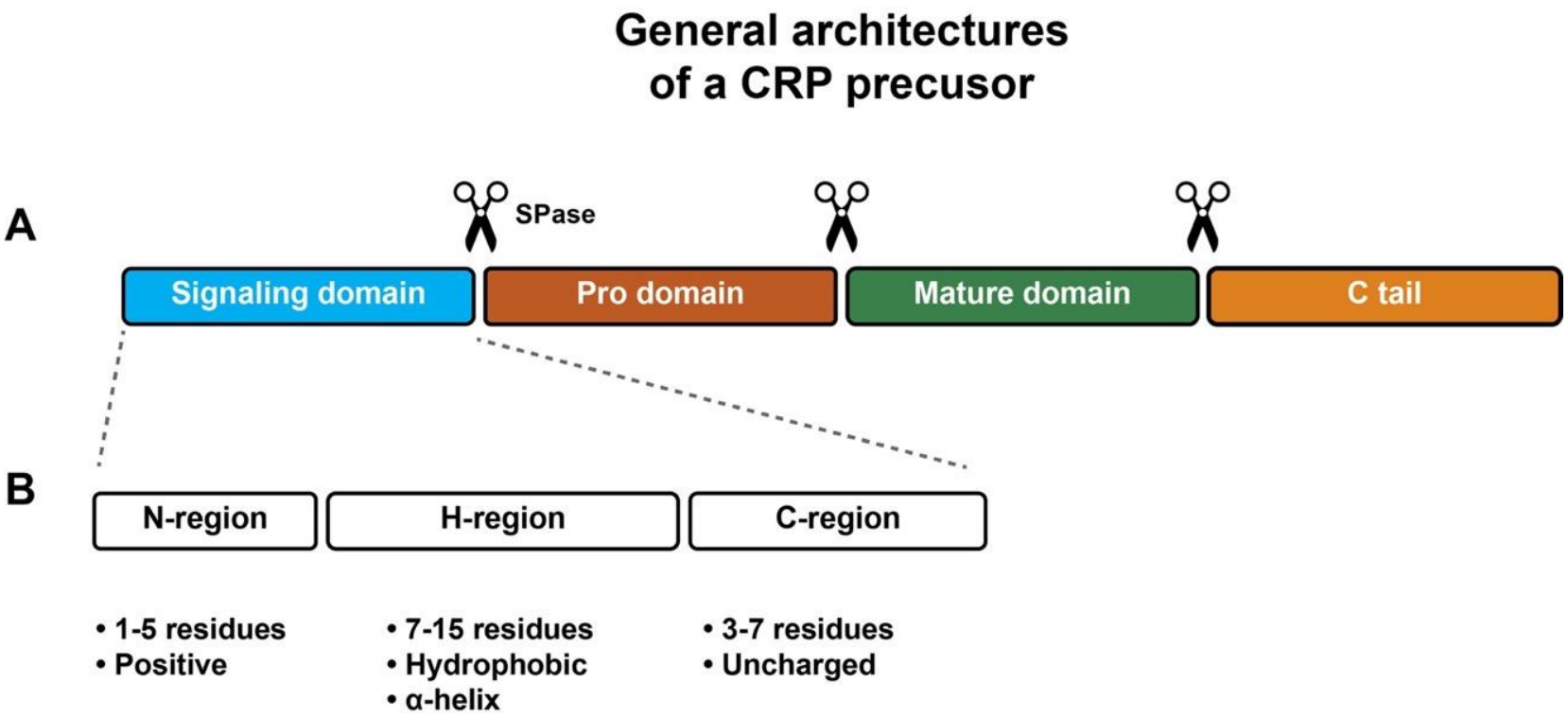


Figure 1.10 General structural architectures of a CRP precursor.

The number of domains might varies in different CRP precursors, but signaling and mature domains are conserved. For example, cyclotide precursors have four domains, while thionin precursors have only three domains.

1.2.4.3 Folding and disulfide formation in the ER

For the newly synthesized CRPs, ER is the first location in their excretory pathway. After reaching ER lumens, linear forms of CRPs start to be folded into three-dimensional conformations, which are stabilized by disulfide bonds. To ensure a proper disulfide formation, ER not only provides an oxidating environment, but also two key players: protein disulfide isomerase (PDI) and ER oxidoreductase 1 (ERO1). ER environment differs from the cytosolic one in two aspects: GSH:GSSG ratio and enzymes with thioredoxin domains. In the cytosol, GSH:GSSG ratio is higher than the ratio in ER, as glutathione exists mainly in a reduced form (or the GSH form) [112]. Other groups of enzymes (e.g. glutathione reductase, thioredoxin, and thioredoxin reductase) can form a reducing system to prevent disulfide formation in the cytosol [113]. In contrast, ER environment has a lower GSH:GSSG ratio. Proteins with thioredoxin domains in ER have higher reduction potential in the active site, suggesting a higher tendency to form disulfide bonds [114, 115]. In this case, ER can have a robust redox buffer for the disulfide formation.

Disulfide formation requires two crucial enzymes: PDI and ERO1 (**Figure 1.11**). On one hand, the PDI is ubiquitous in all eukaryotic cells and it functions as relaying disulfide to substrates, like CRPs. PDI achieves this relaying function by thioredoxin-like domains, which can accept electrons from thiol groups in substrates and lead to disulfide bond formation [116, 117]. On the other hand, oxidation form of PDI is contributed by ERO1 that functions as a vehicle to transfer oxidating power from oxygen to PDI [116, 117]. The oxygen then loses two electrons to form hydroxy peroxide, which can further be cleaned by other enzymes, such as glutathione peroxidase and peroxiredoxin IV, to relieve intracellular oxidative stress [116].

Taken together, ER as a place for CRPs maturation has a robust buffer system, containing oxidating components (e.g. low reduced glutathione) and crucial proteins (e.g. PDI and ERO1). This system can fold linear CRPs into a three-dimensional molecule with super stable structures, the premise for CRP biological functions.

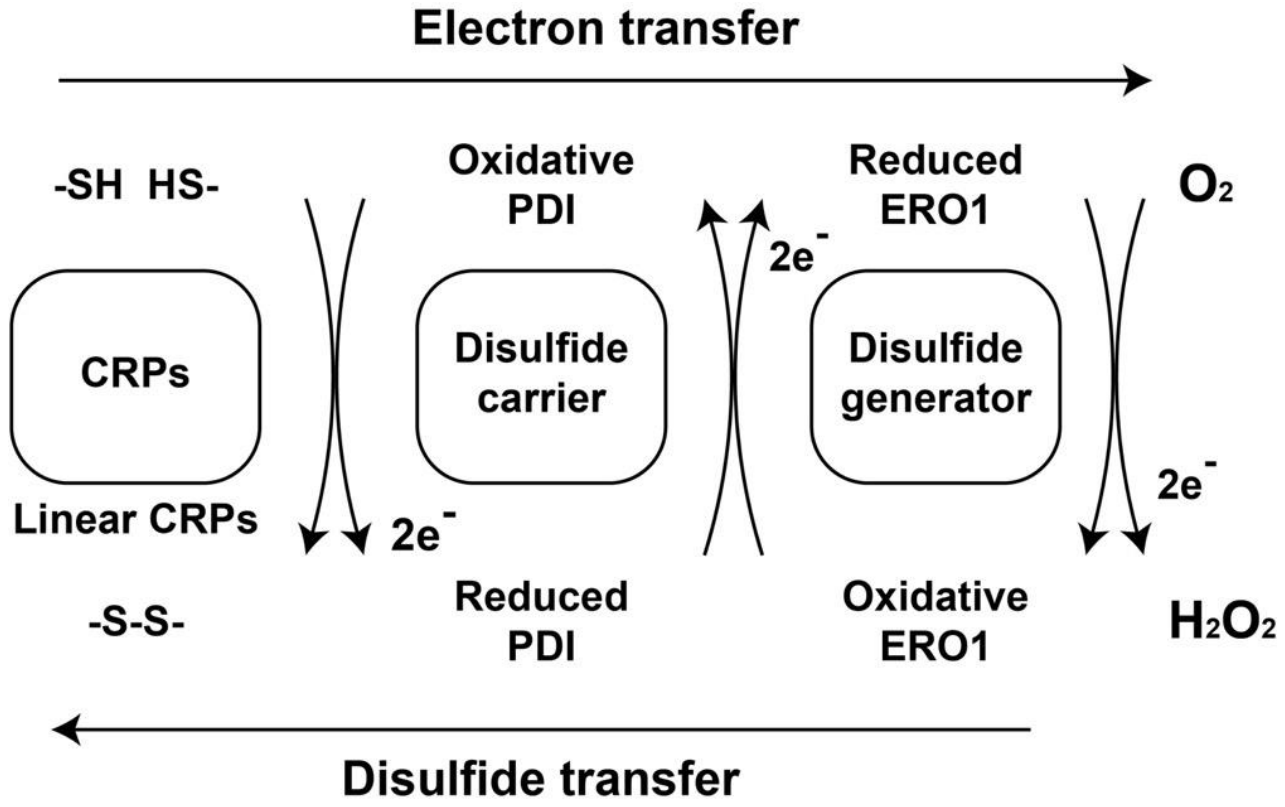


Figure 1.11 Putative disulfide formation of plant CRPs in the ER.

This figure is adapted from the work of Yayoi Onda [117].

1.2.5 Molecular structures of CRPs

1.2.5.1 Thionin structures

Thionins are unique in primary, secondary, and tertiary structures (**Figure 1.12**). They are cationic with three or four symmetric disulfide connectivity [54]. Although varying in the number of disulfide bonds, in secondary structures, 6C-/8C-thionins commonly possess a β1-α1-α2-β2-coil motif [35]. In this motif, two α-helices constitute a long arm, while two wwz2rands constitute a short arm, and these two arms then form a gamma Γ-like fold in their tertiary structures [118]. To maintain this unique fold, thionins employ three to four disulfide linkages. In 8C-thionins, there are four disulfide bonds: 1) Cys I of β1 to Cys VIII of a coil at the C-terminal; 2) Cys II of β1 to Cys VII of β2; 3) Cys III of α1 to Cys VI of loop 4; 4) Cys IV of α1 to Cys V of α2. In 6C-thionins, they also have similar disulfide connectivities, with one missing connectivity (i.e. disulfide bond between Cys II and Cys VII). In addition, thionins have a disulfide bond linking between C- and N-terminal sequences, loosely forming a pseudocyclic peptide and further stabilizing the characteristic thionins.

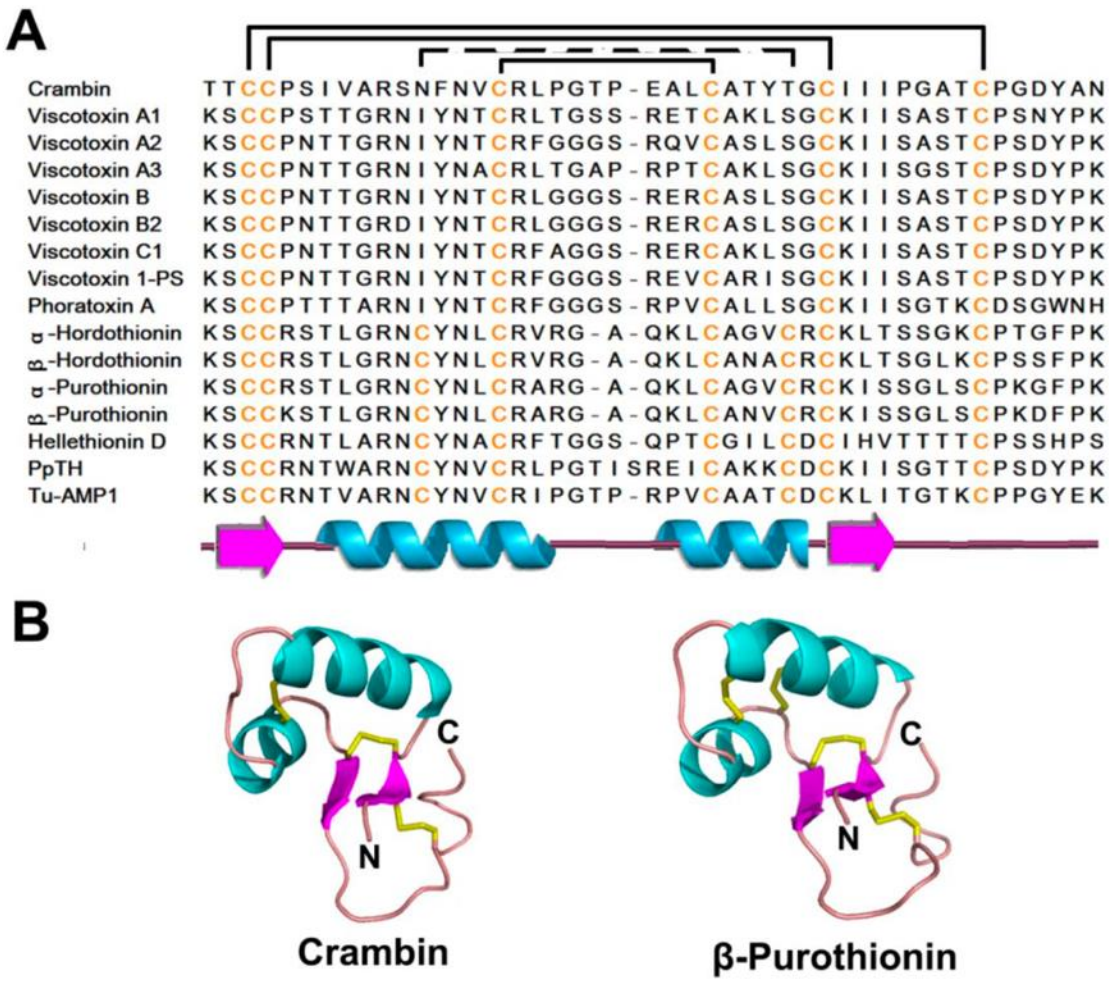


Figure 1.12 Primary, secondary, and tertiary structures of plant thionins.

(A) Thionins members with their disulfide connectivities and β1-α1-α2-β2-coil motif. **(B)** Tertiary structures of thionins with a gamma Γ-like fold. This figure is obtained from James P. Tam et al. [48]

1.2.5.2 Defensin structures

In primary structure (**Figure 1.13 A**), plant defensins typically have eight cysteine residues, forming four disulfide bonds, except for ten disulfide bonds in PhDs [119]. In these four bonds, three of them form a knottin-type connectivity at the primary level, but not at a tertiary level, and one bond at N- and C-termini forms a pseudocyclic structure in plant defensins [65]. In the secondary structure (**Figure 1.13 B**), plant defensins contain one α helix and three anti-parallel β strands, which are stabilized by three disulfide bonds [65, 120]. These secondary structures of plant defensins can also be observed in mammalian defensins [121]. In tertiary structures, plant defensins and mammalian are different from each other (**Figure 1.13 B and 1.14**). Plant defensins have a cystine-stabilized αβ fold (or Csαβ motif) that belongs to the *cis*-defensins superfamily. In this superfamily, two disulfide bonds from the CXC motif of the third β strand both links to the same α helix, and this type of topology is called *cis* (**Figure 1.14 A and C**). Mammalian or invertebrate defensins have -CC- motif in the third β strand, projecting two disulfide bonds to different secondary elements, and thus these invertebrate defensins are from *trans*-defensin superfamilies (**Figure 1.14 B and D**). Other than the well conserved Csαβ motif, amino acid and loop length of plant defensins vary, contributing to diverse functions, such as antimicrobial and neurotoxic activities.

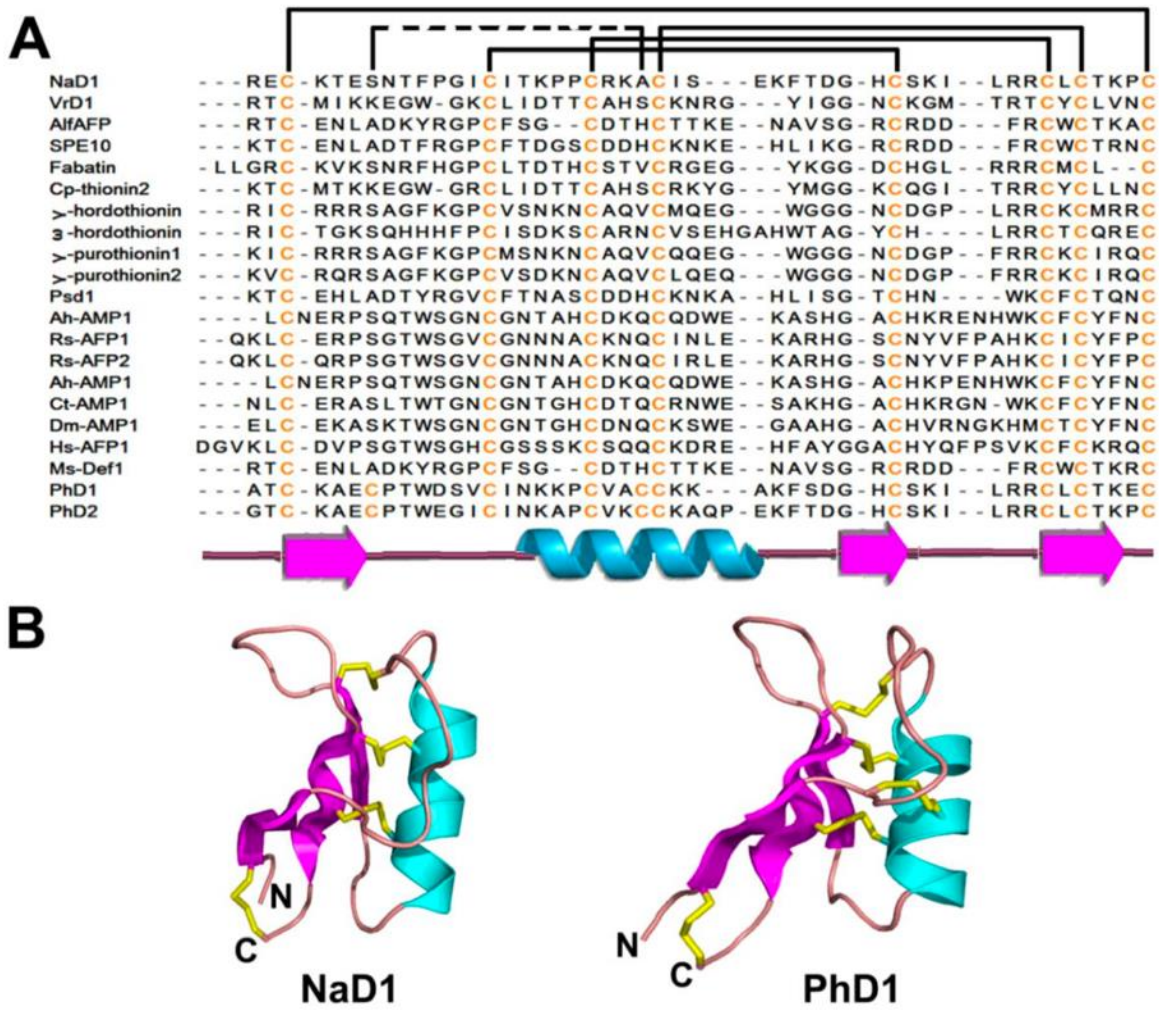


Figure 1.13 Primary, secondary, and tertiary structures of plant defensins.

(A) Defensins members with their knottin-type disulfide connectivities and secondary structures. **(B)** Tertiary structures of defensins with one α helix and three anti-parallel β strands, which are stabilized by disulfide bonds. This figure is obtained from James P. Tam et al. [48]

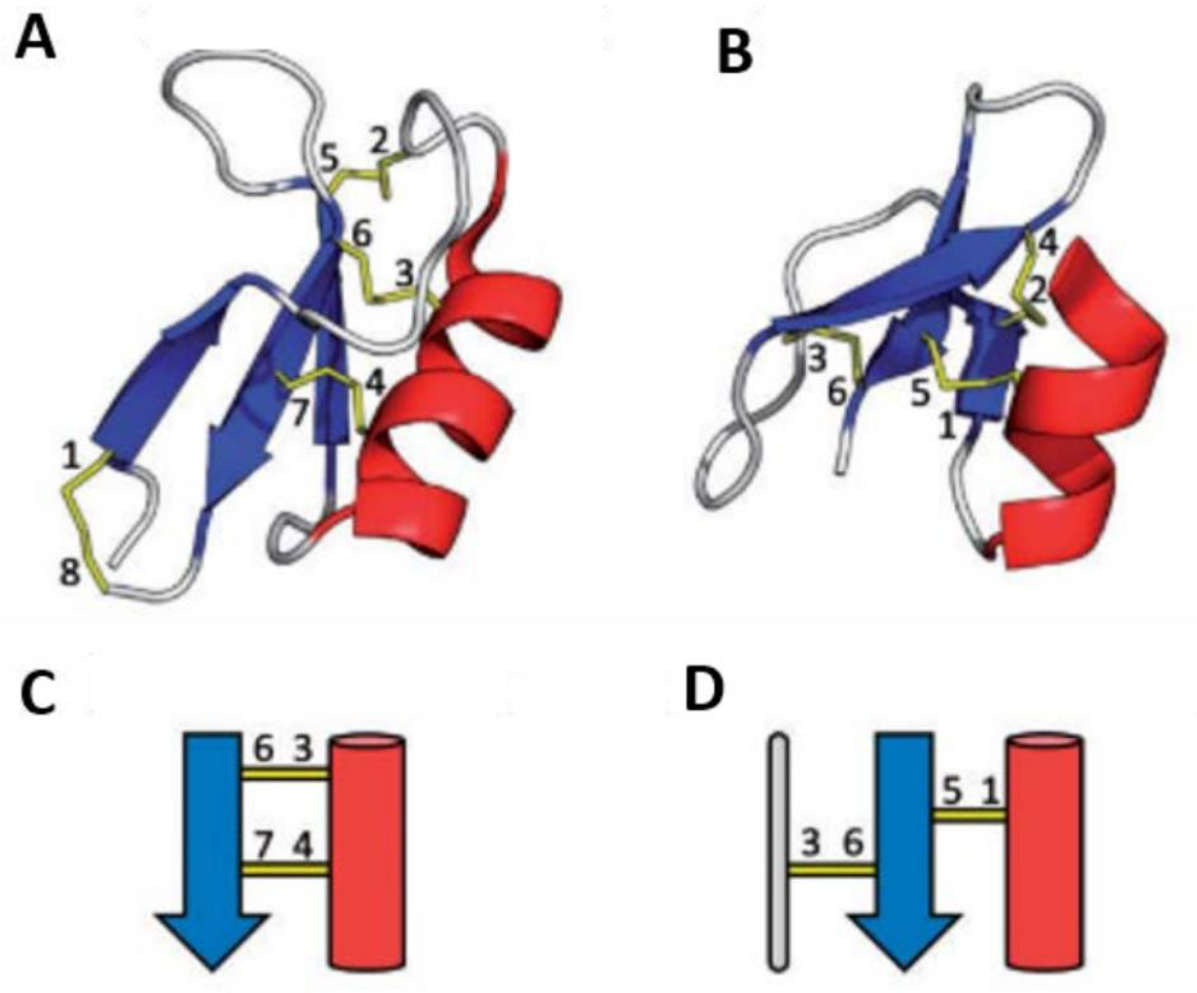


Figure 1.14 Tertiary structures of plant and mammalian defensins.

(A) Plant *cis*-defensins (NaD1, PDB: 1MR4). **(B)** Mammalian *trans*-defensins (HBD1, PDB: IJV). **(C)** *cis* topology is two disulfide bonds (yellow) from the β strand link to the same α helix. **(D)** *trans* topology is one disulfide bond (yellow, 5 to 1) from the β strand links to the α helix, while another (yellow, 6 to3) is projected in the opposite direction. This figure is from J. Huang (Ph.D. thesis, Nanyang Technological University, 2019).

1.2.5.3 Hevein-like peptide (HLP) structures

HLPs contain chitin-binding (CB) and non-chitin-binding (non-CB) HLPs subfamilies. Since the difference between these two subfamilies is the CB domain (see details below), we only used CB-HLPs as examples for structural discussion. CB-HLPs, based on cystine number can be further classified into 6C-, 8C-, and 10C- HLPs (**Figure 1.15**). The 6C-CB-HLPs contain a knottin-type disulfide connectivity at the primary level (not at the tertiary level) [35], this connectivity locks two anti-parallel β strands and two α helices [35]. Also, the 6C-CB-HLP has four intercysteinyl loops. Among these loops, the loop 3 and loop 4 contain a chitin-binding domain: SXφGφCGX4φ, in which X stands for any residues and φ represents aromatic residues (**Figure 1.15**) [35]. 8C-CB-HLPs, compared with 6C-CB-HLPs, have an extender at its C terminal, containing the fourth disulfide bond and one more β strand (like the hevein, **Figure 1.16**). Although 10C-CB-HLPs possess one more disulfide bond compared with 8C-CB-HLPs (**Figure 1.15**), the location of this additional disulfide bond varies. For example, EAFPs from *Eucommia ulmoides Oliv* have the fifth disulfide bridging between an N-terminal coil and a β strand 3 [122], while in Ee-CBP from *Euonymus europaeus*, the fifth disulfide bond locates at in the C-terminal [123]. In a word, 6C-/8C-/10C-HLPs are similar to each other in structure, but might be different in functions, as the same structural scaffold can hold different sequences.

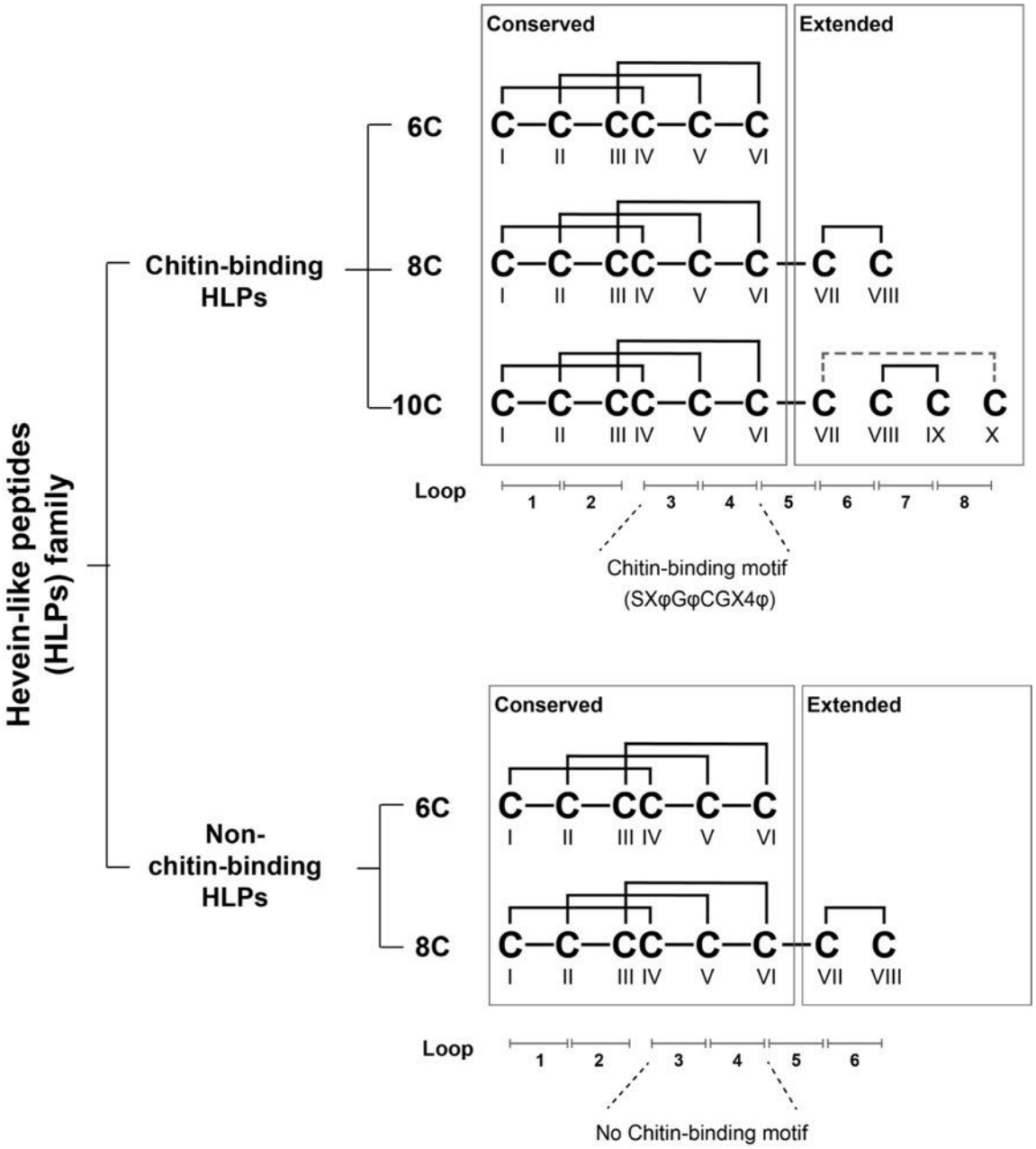


Figure 1.15 Cysteine motifs of chitin-binding and non-chitin-binding HLPs.

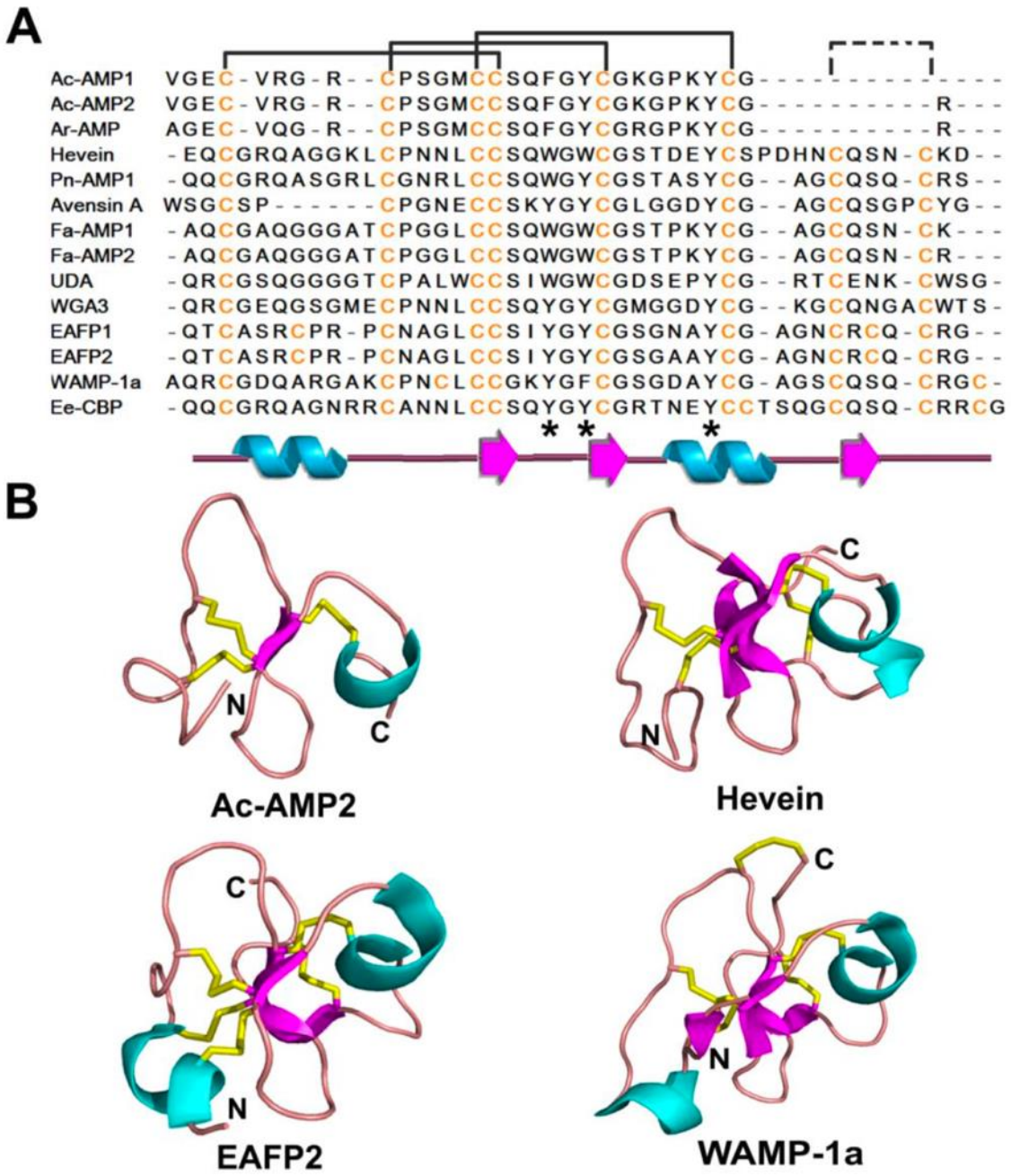


Figure 1.16 Primary, secondary, and tertiary structures of HLPs.

(A) Primary sequences of hevein-like peptides (HLPs), with aromatic residues (marked by asterisks) in the chitin-binding motif. **(B)** Tertiary structures of 6C-HLPs (Ac-AMPs), 8C-HLPs (hevein and WAMP-1a), and 10C-HLPs (EAFP2). This figure is obtained from James P. Tam et al. [48]

1.2.5.4 Knottin structures

Knottins have three disulfide bonds, forming connectivity of Cys I-IV, Cys II-V, and Cys III-VI (**Figure 1.17**). Although some defensins and HLPs share this disulfide connectivity in primary sequences, they do not have the knottin topology at the tertiary level. The core structural topology of knottins is that two disulfide bonds (Cys I-IV and Cys II-V), together with peptide backbones in between, constitute a macrocyclic that is threaded by the third disulfide bond: Cys III-VI (**Figure 1.17**). Due to this knot-like topology, knottins are treated to have “peptide promiscuity”, a term to describe their multiple biological functions discovered thus far [67, 124]. Other than the core topology, knottins also contain some characteristic secondary elements, like cystine-stabled triple β strands [124]. The linear knottins can be cyclized from head-to-tail of the peptide backbone to form cyclotide [125], or by using a disulfide bond to link the head and tail of the backbone to form pseudocyclic peptides [86] (**Figure 1.18**).

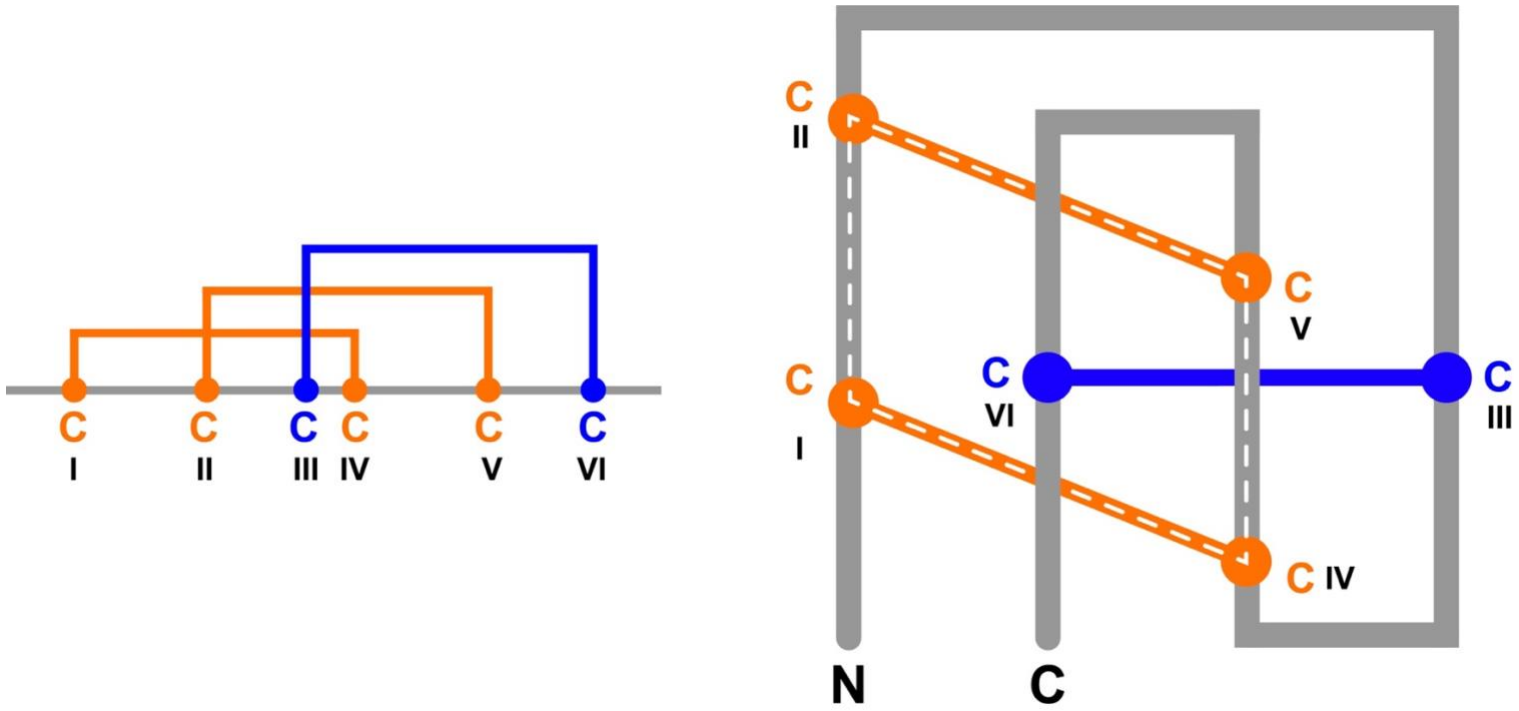


Figure 1.17 Structural topology of knottins.

The macrocyclic (the dashed line), formed by two disulfide bonds (orange) and the peptide backbone (grey), is penetrated the third disulfide bond (blue) to form a knot. This figure is adapted from a figure in Knottin database in Jan, 2022 (website link: https://www.dsimb.inserm.fr/KNOTTIN/)

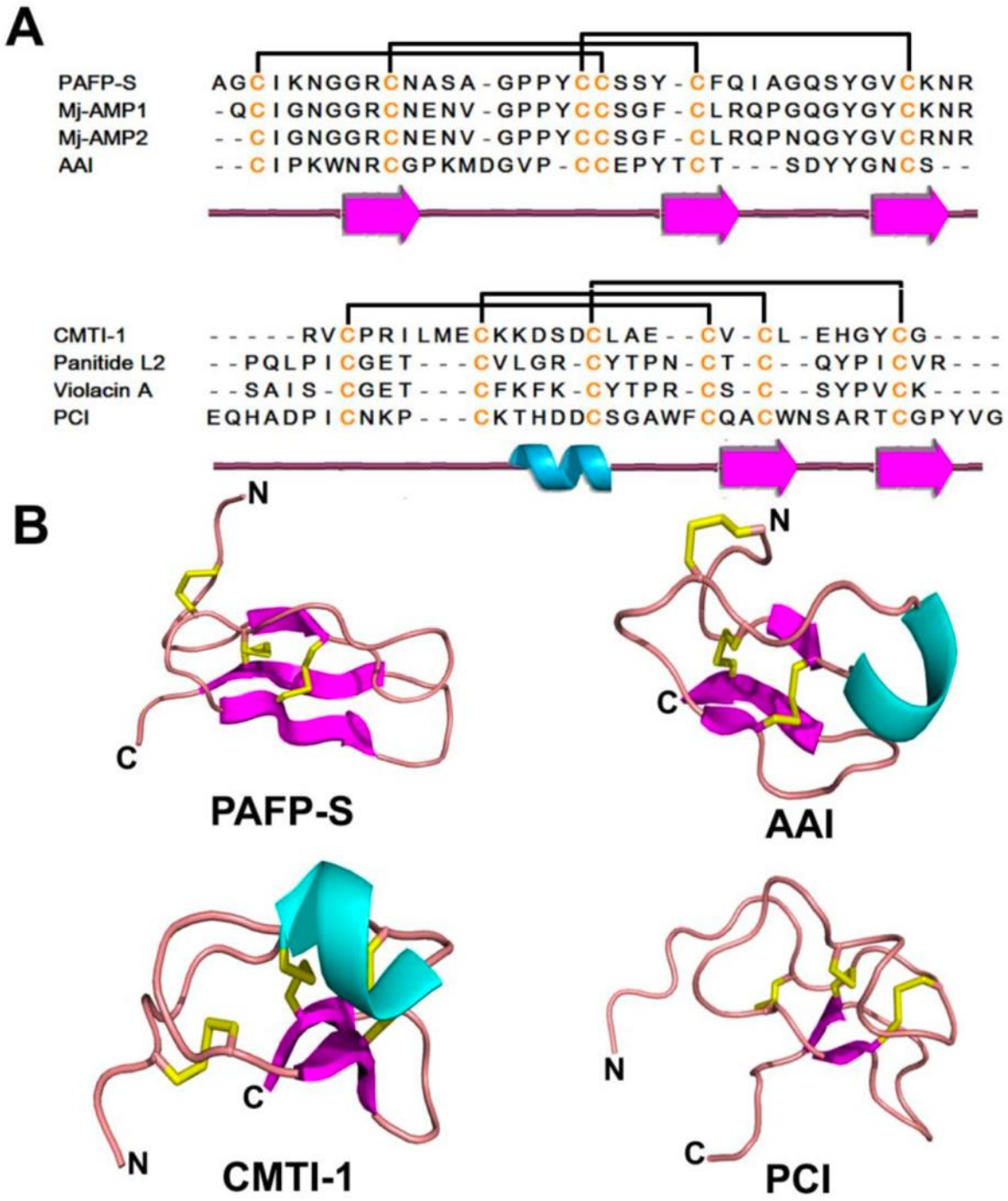


Figure 1.18 Primary, secondary, and tertiary structures of knottins.

(A) Knottins sequences and their secondary structures. **(B)** Examples of tertiary structures of knottins. This figure is obtained from James P. Tam et al. [48].

1.2.5.5 Other CRP structures

CRP families also include α-hairpinin, lipid transfer proteins (LTPs), and snakins. The α-hairpinin has two symmetric disulfide bonds, stabilizing two helices in its structures (**Figure 1.19 and 1.20**). These helices dominating properties distinguish α-hairpinin from other CRP with wwz2rands. In structures of LTPs, α-helices are stabilized by four disulfide bonds [126, 127], and the inner cavity of these α-helices is hydrophobic, responsible for binding to lipids [128-131]. The structures of snakins are not well characterized thus far, but they are predicted to show α-helices and six disulfide bonds. In addition, other CRPs with newly discovered structures include jasmintides, ginkgotides, lybatides, potentides, and ginsentides. These new CRP families enriched our knowledge on CRP structural diversity.

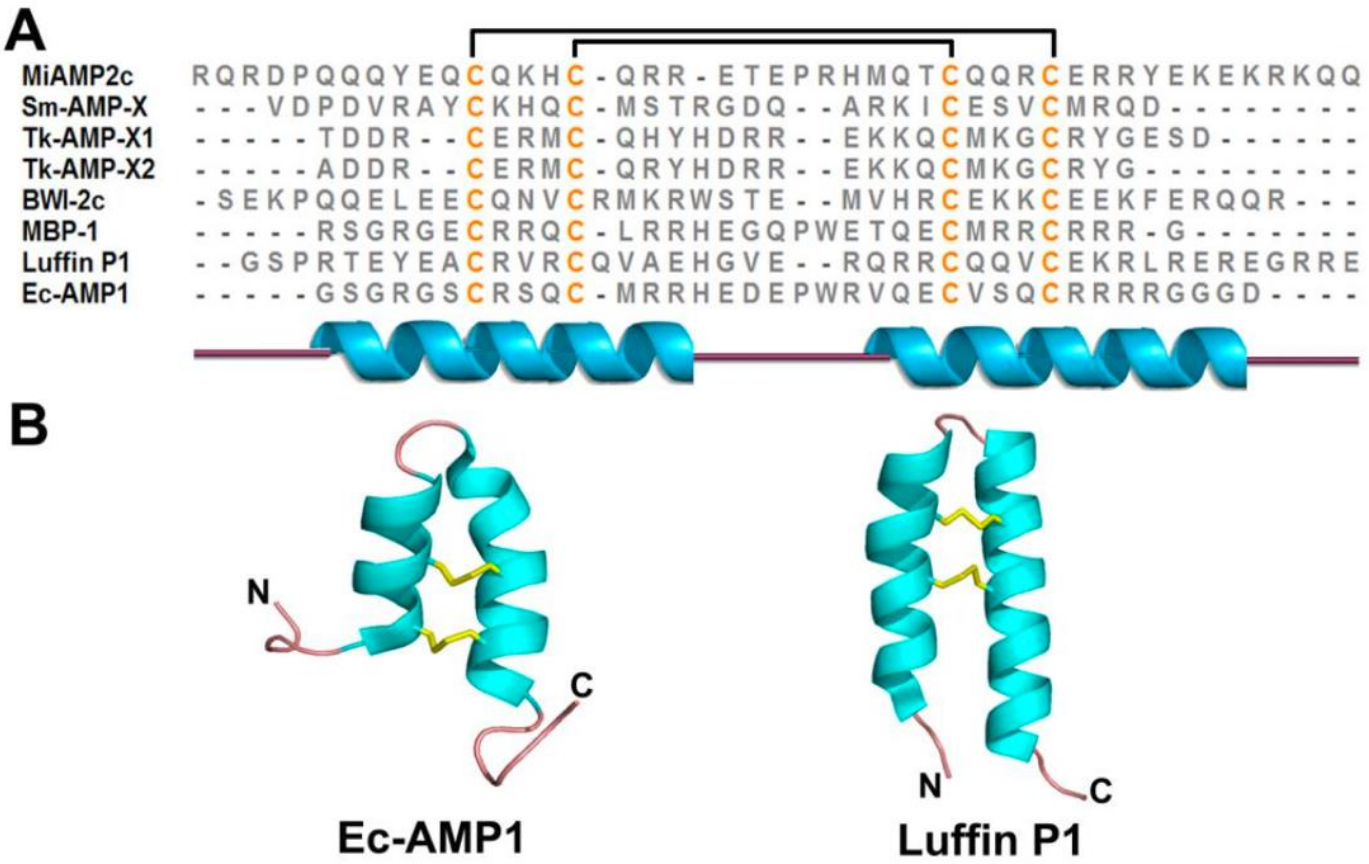


Figure 1.19 Primary, secondary, and tertiary structures of α-hairpinin.

(A) α-hairpinin sequences and their secondary structures. **(B)** Examples of tertiary structures of α-hairpinin. This figure is obtained from James P. Tam et al. [48]

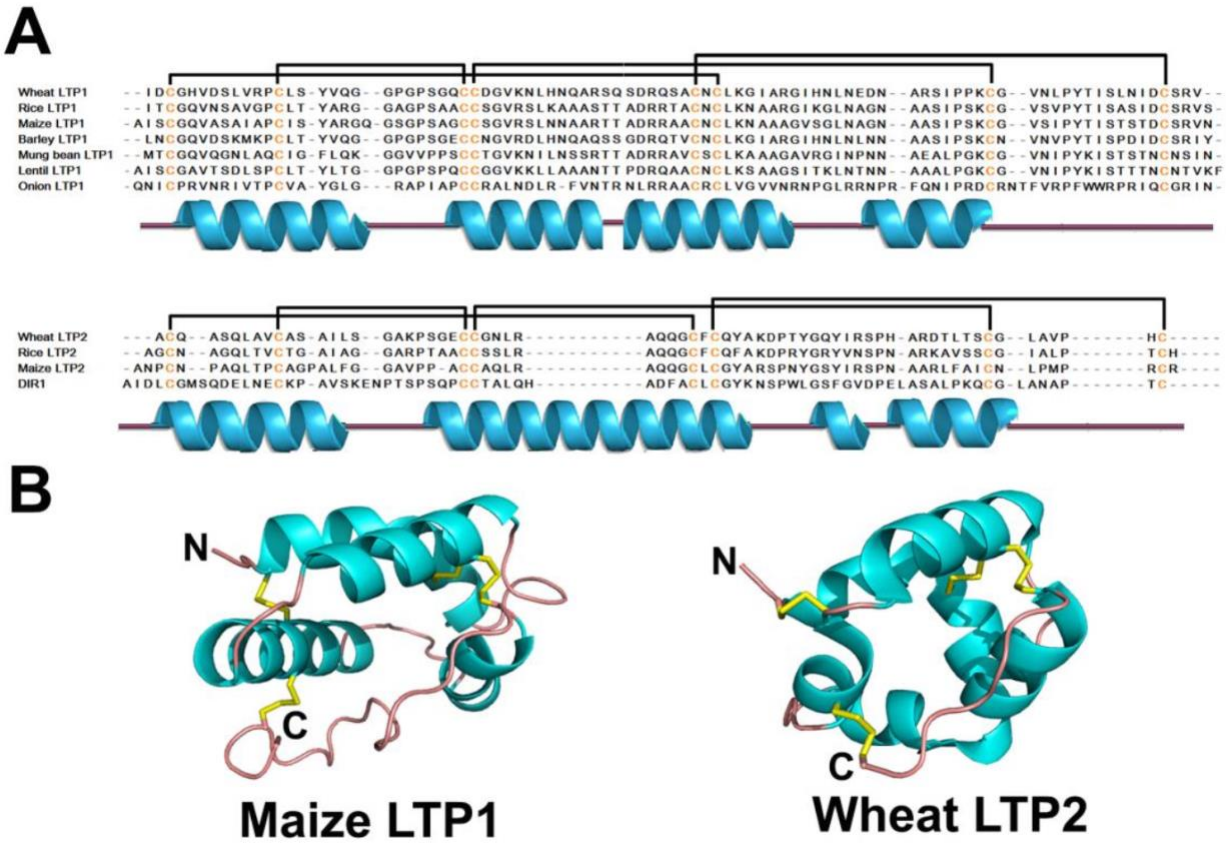


Figure 1.20 Primary, secondary, and tertiary structures of lipid transfer proteins (LTPs).

(A) Lipid transfer proteins (LTPs) sequences and their secondary structures. **(B)** Examples of tertiary structures of lipid transfer proteins (LTPs). This figure is obtained from James P. Tam et al. [48]

1.2.6 Chemical synthesis of CRPs

Compared with synthetic compounds, natural CRPs normally have more complex molecular structures, partially due to the long-time biological evolution and multiple functions. Although this complexity may render them functionally beneficial, producing CRPs in tedious conventional ways, like isolation from plants, is not desirable anymore. Solid-phase peptide synthesis (SPPS) and oxidative folding are rapid and robust chemical ways of producing CRPs. This method not only accelerates the functional investigation of CRPs, but also paves a road for structurally optimizing CRPs with better efficacies and specificities.

SPPS method was first reported by Bruce Merrifield in 1963 [132]. In following several decades, this method has been developed into a robust way for synthesizing various peptides, even peptides with post-translational modifications [133]. The SPPS generally has three key steps (**Figure 1.21**). First, the C-terminal of the first amino acid residue, with its side chain and alpha-amino groups protected, is directly or indirectly (via a linker) attached to the insoluble resin, like crosslinked polystyrene. After that, resin, with an attached residue, is washed and filtered for clearing unwanted by-products and reagent waste. Second, the alpha-amino group of the first residue on the resin is deprotected, and a new residue, with protection of its side chain and alpha-amino groups, is then coupled to the alpha-amino group of the first residue. The first and second steps are repeated until the synthesis of a full sequence is complete. Third, all protective groups on the newly synthetic peptides are removed, and then (or at the same time) the sequence is released from the resin by using cleavage reagents. In my projects of this thesis, we synthesized linear CRPs by SPPS, and then folded them into three-dimensional CRPs by oxidative folding.

Oxidative folding of CRPs is referred to the formation of correct disulfide bonds intramolecularly. In vitro, many variables should be taken into consideration, such as redox buffer, pH value, peptide concentrations, and additives. Typically, the redox buffer mimics the environment of ER lumen where CRP disulfide bonds are formed. The redox buffer contains oxidants, like glutathione, dimethyl sulfoxide, and (±)-trans-1,2-bis(2-mercaptoacetamido)cyclohexane (BMC) [134, 135]. pH value can influence deprotonation (or activation) of thiol groups that is a premise for disulfide formation [136, 137]. Peptide concertation, in aqueous folding, could not be too high (normally 20 – 50 μM), because the high concentration increases the ratio of intermolecular disulfide formation [138]. The additives might also improve the yield of CRPs folding. For example, adding salt can inhibit the electrostatic interaction [139], and detergent might improve the folding yield of hydrophobic peptides [140]. In addition, folding temperature and time also are variables that could influence oxidating folding of CRPs. Therefore, the development of SPPs and oxidative folding might be effective solutions for peptide production, which is a bottleneck of CRP scaled-up applications.

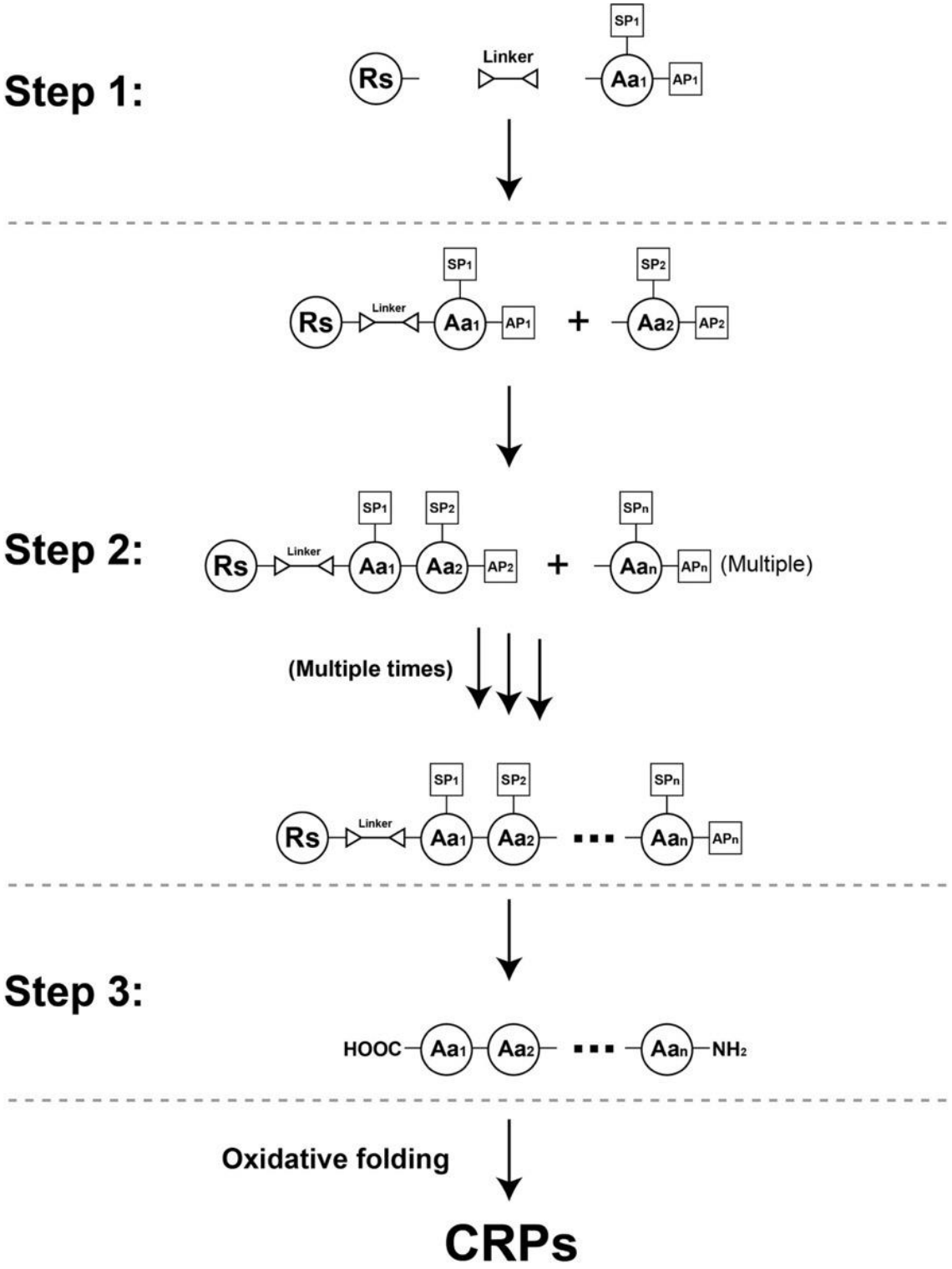


Figure 1.21 Solid-phase peptide synthesis and oxidative folding of CRPs. Abbreviations are as follows: Rs (resin), SP (side-chain protective group), and Aa (amino acid residues).

1.3. Underexplored CRPs in medicinal plants

1.3.1 Underexplored medicinal CRPs

At the primary structural level, CRPs can be classified by cysteine motifs that include disulfide connectivity, cysteine numbers, and the cysteine pattern (also known as framework or spacing). The cysteine pattern directly influences the length and distributions of intercysteinyl sequences that are crucial for biological functions of CRPs. For these reasons, we often use the cysteine pattern to structurally classify CRPs, aiming to determine their functional and evolutional similarities. Compared with CRPs patterns in animals like cone snails, plant CRPs frameworks are markedly underexplored. In cone snails, conotoxins are specific CRPs and used for immobilizing prey or predators, because they can bind to ion channels, receptors, and transporters on neuronal cells with high affinity [141]. So far, the conotoxins have 31 different patterns with cysteine numbers of 4-12 (**Table 1.7**). In contrast to conotoxins with 31 patterns, plants have relatively fewer cysteine patterns discovered. From the transcriptomic analysis of 1267, it is estimated that these plants have 46,112 CRPs in theory [142]. Because cysteine patterns are structural skeletons for functional peptide sequences, structurally under-exploration also indicates functionally under-exploration of CRPs in plants.

Table 1.7 Cystine pattern-based classification of conotoxin.

This table was taken from the ConoServer database in Jan. 2022 (website link: http://www.conoserver.org/?page=classification&type=cysteineframeworks).

Framework	Cystine pattern	# cysteines	Connectivity	Reference
I	CC-C-C	4	I-III, II-IV	Gray,W.R. *et al.* (1981) *J. Biol. Chem.* **256**:4734-4740
II	CCC-C-C-C	6		Ramilo,C. *et al.* (1992) *Biochemistry* **31**:9919-9926
III	CC-C-C-CC	6		Sato,S. *et al.* (1983) *FEBS Lett.* **155**:277-280
IV	CC-C-C-C-C	6	I-V, II-III, IV-VI	Fainzilber,M. *et al.* (1995) *Biochemistry* **34**:8649-8656
V	CC-CC	4	I-III, II-IV	Walker,C.S. *et al.* (1999) *J. Biol. Chem.* **274**:30664-30671
VI/VII	C-C-CC-C-C	6	I-IV, II-V, III-VI	Olivera,B.M. *et al.* (1984) *Biochemistry* **23**:5087-5090
VIII	C-C-C-C-C-C-C-C-C-C	10		England,L.J. *et al.* (1998) *Science* **281**:575-578
IX	C-C-C-C-C-C	6	I-IV, II-V, III-VI	Lirazan,M.B. *et al.* (2000) *Biochemistry* **39**:1583-1588
X	CC-C.[PO]C	4	I-IV, II-III	Balaji,R.A. *et al.* (2000) *J. Biol. Chem.* **275**:39516-39522
XI	C-C-CC-CC-C-C	8	I-IV, II-VI, III-VII, V-VIII	Jimenez,E.C. *et al.* (2003) *J. Neurochem.* **85**:610-621
XII	C-C-C-C-CC-C-C	8		Brown,M.A. *et al.* (2005) *Biochemistry* **44**:9150-9159
XIII	C-C-C-CC-C-C-C	8		Aguilar,M.B. *et al.* (2005) *Biochemistry* **44**:11130-11136
XIV	C-C-C-C	4	I-III, II-IV	Moller,C. *et al.* (2005) *Biochemistry* **44**:15986-15996
XV	C-C-CC-C-C-C-C	8		Peng,C. *et al.* (2008) *Peptides* **29**:985-991
XVI	C-C-CC	4		Pi,C. *et al.* (2006) *Genomics* **88**:809-819
XVII	C-C-CC-C-CC-C	8		Yuan,D.D. *et al.* (2008) *Peptides* **29**:1521-1525
XVIII	C-C-CC-CC	6		Chen,J.S. *et al.* (1999) *J Nat Toxins* **8**:341-349
XIX	C-C-C-CCC-C-C-C-C	10		Chen,P. *et al.* (2008) *Toxicon* **52**:139-145
XX	C-CC-C-CC-C-C-C-C	10		Loughnan,M.L. *et al.* (2009) *Biochemistry* **48**:3717-3729
XXI	CC-C-C-C-CC-C-C-C	10		Möller,C. and Marí,F. (2011) *Biopolymers* **96**:158-165
XXII	C-C-C-C-C-C-C-C	8		Elliger,C.A. *et al.* (2011) *Toxicon* **57**:311-322
XXIII	C-C-C-CC-C	6		Ye,M. *et al.* (2012) *J Biol Chem* **287**:14973-14983
XXIV	C-CC-C	4		Luo,S. *et al.* (2013) *PLoS ONE* **8**
XXV	C-C-C-C-CC	6		Aguilar,M.B. *et al.* (2013) *Peptides* **[ahead of print]**
XXVI	C-C-C-C-CC-CC	8		Bernáldez,J. *et al.* (2013) *Mar Drugs* **11**:1188-1202
XXVII	C-C-C-CCC-C-C	8		Jin,A.H. *et al.* (2015) *Proc. Biol. Sci.* **282**
XXVIII	C-C-C-CC-C-C-C-C-C	10		Lu *et al.* (2017) *Peptides, 94, pp.64-70* **94**:64-70
XXIX	CCC-C-CC-C-C	8		Bernáldez-Sarabia *et al.* (2019) *Toxins* **11**:128
XXX	C-C-CCC-C-C-C-CC	10		Bernáldez-Sarabia *et al.* (2019) *Toxins* **11**:128
XXXII	C-CC-C-C-C	6		Kancherla,A.K. *et al.* (2015) *ACS Chem. Biol.* **10**:1847-1860
XXXIII	C-C-C-C-C-C-C-C-C-C-C-C	12		Pardos-Blas,J.R. *et al.* (2019) *Marine drugs* **17**:453

As mentioned above, the plant CRPs have at least seven different families, including thionins, defensins, knottins, α-hairpinin, lipid transfer proteins (LTPs), and snakins. These plants CRPs are well-known for their anti-microbial activities [35], but CRPs from medicinal plants are rarely structurally and functionally investigated. To broaden plant CRPs from medicinal plants, our laboratory first characterized jasmintide from *Jasminum sambac* [143], ginkgotide from *Ginkgo biloba* [144], lybatide from *Lycium barbarum* [99], and ginsentide from ginseng [48]. Because these CRPs are derived from medicinal plants with a long history of medicinal uses and are highly stable in structures, they might have great potential of being developed into orally active drugs or leads for therapeutics. However, these CRPs from medicinal plants are rarely reported in terms of their biological functions.

1.3.2 Neglected chemical space in medicinal plants

In medicinal plants, the phytochemicals can be generally divided into three groups with different chemical spaces (or molecular weight, M.W.). These three groups are: 1) small molecules with M.W. less than 1 kDa; 2) macromolecules with M.W. more than 10 kDa; 3) peptidyl molecules with M.W. between 1 and 10 kDa (**Figure 1.23**). In modern studies of medicinal plants, small molecules are intensively studied, in contrast to rare investigations of macromolecules (represented by proteins) and peptidyl molecules. This is a stereotype that proteins and peptides are unstable under extreme conditions, like high temperature during decoction and low pH/enzymatic conditions in gastrointestinal tracts. In our research, we are interested in unusual peptides, called cysteine-rich peptides (CRPs) that are 2-6 kDa and constrained by 2-5 disulfide bonds. The bonds make them highly stable under high temperature, acidic, and enzymatic conditions.

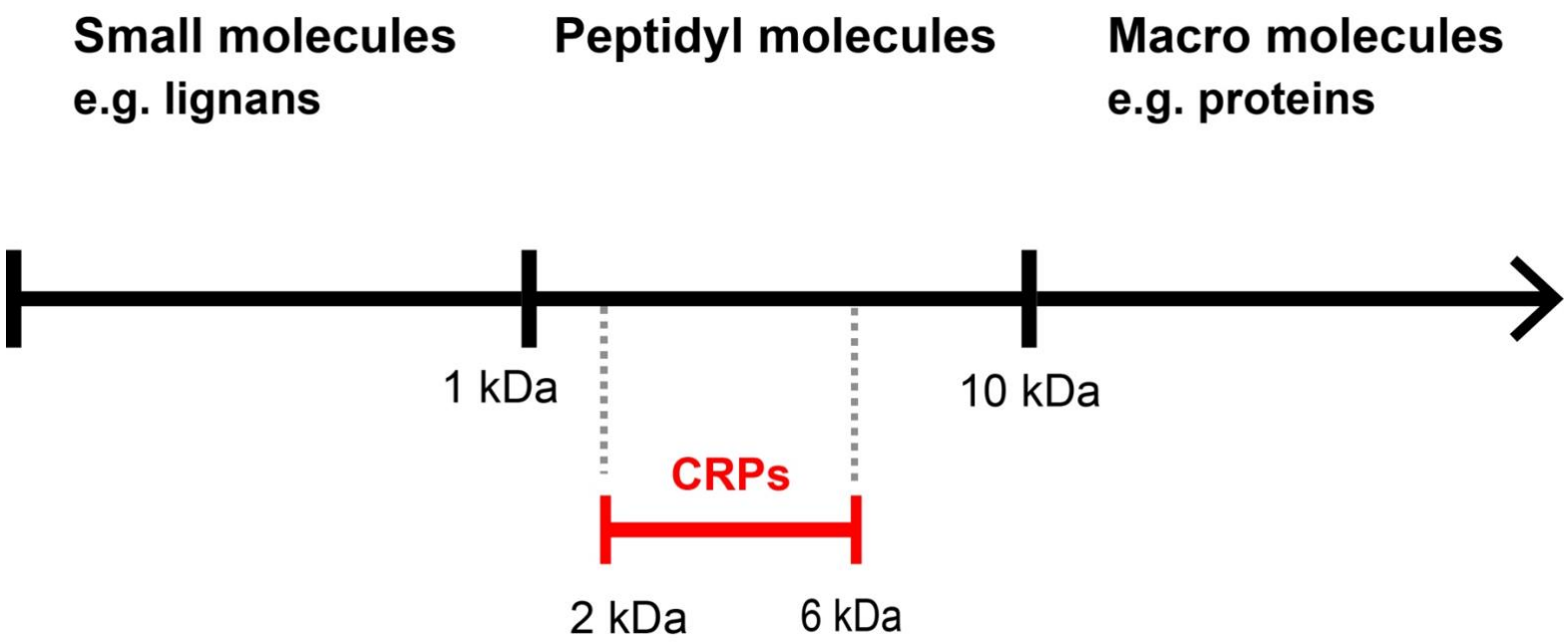


Figure 1.22 The position of CRPs in the chemical space of phytochemicals.

This figure is adapted from the work of Loo S. N. (Ph.D. thesis, Nanyang Technological University, 2017).

1.4. *Withania* somnifera

1.4.1 Indian traditional medicine *W. somnifera*

Indian ginseng (IG) refers to the root of the plant *W. somnifera* in the Solanaceae family (**Figure 1.22**). Although this plant can be found in other Asian countries, India is the most well-known producing one, especially in the Indian Northwest and middle areas [145]. The geographical preference might be that IG requires relatively restricted growing conditions. For example, cultivating conditions of IG would be best, if the altitude is 1500 m above sea level, regions have tropical or semitropical climates with ~650 mm rainfall every year, and the temperature is in the range of 20 to 38 Celsius [145]. Other factors, like soil types, can also influence geographical distributions of IG [146-148].

In ancient times, for its important pharmacological value, IG had been included in Indian, Unani, and Tibetan traditional medicine systems. In the Indian system (Ayurveda), IG is called ashwagandha (a term describing a strong sweaty-horse-like smell), and thus the IG is often administrated with milk and honey [145] to exert tonic effects like slowing aging [149, 150]. In the Unaini system, IG is called asgand with varieties and can be administrated alone or with other medicines [151, 152]. Also, the asgand is included in many formulations and used for diseases like arthritis, anxiety, and ulcers [152, 153]. The Tibetan system is one of the medicine systems with the longest history in the world. In this system, IG is called Asgandnagori or Ba-dzigandha [154, 155], and mainly used for respiratory, cardiovascular, and hepatic diseases. Though recorded in different works of literature, IG is still widely treated as Indian traditional medicine. This might partially be because of IG geographical distributions

Figure 1.23 *Withania somnifera* plant.

Flower and fruit are shown on the left, and root (Indian ginseng) is shown on the right. These pictures were purchased from Shutterstock (website link: https://www.shutterstock.com/).

1.4.2 Phytochemicals of the *W. somnifera*

Many phytochemicals of IG have been structurally characterized, including steroids, flavonoids, sitoindosides, steroidal lactones, and alkaloids [148]. Unlike others, the latter two are generally regarded as the main components, responsible for the pharmacological activities of the IG [145]. The steroidal lactones contain withanone, withaferin-A, and withanolides (A, E, F, G, H, I, J, K, L, and M). Interestingly, the steroidal lactones from the IG are structurally similar to ginsenosides from *Panax ginseng* [156, 157], suggesting they could have remarkable pharmacological potential like panacea ginseng. In the steroidal lactones, the most intensively studied one is withaferin-A, which shows a broad spectrum of pharmacological effects, including neuroprotection [158], anti-cancer [159], and anti-virus effects [160]. In the IG, alkaloids have over 12 members discovered. Among them, the withanine is the major one [145]. The structures of their representatives are shown (**Figure 1.24**).

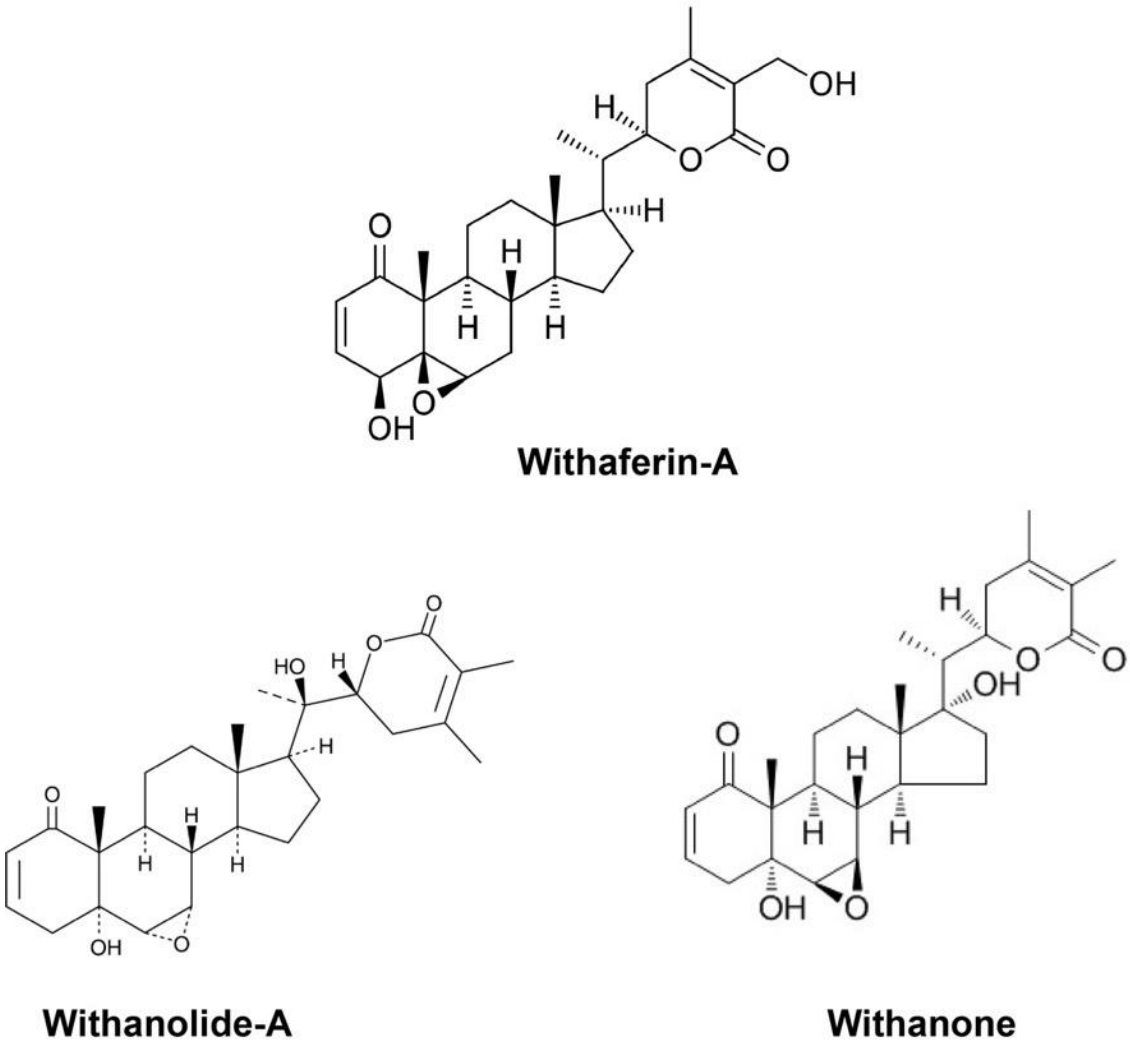


Figure 1.24 Representative phytochemicals from the Indian ginseng.

1.4.3 Pharmacological activities of the *W. somnifera*

Like Panax ginseng, Indian ginseng (IG) is also an ancient panacea with a broad range of pharmacological activities. In modern studies, these activities of IG at least cover neuroprotection, anti-infection, anti-diabetes, anti-cancer, and cardioprotection [148].

In the *W. somnifera*, the "*somnifera*" stands for "leading to sleep" in Latin, intuitively indicating its potential applications in neurological diseases. In elderly people, neurodegenerative diseases are a big health concern, and they are represented by Parkinson's disease and Huntington's disease. In Parkinson's mouse models, IG extract has neuroprotective activities, demonstrated by reversing MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) or MB-PQ (Maneb-Paraquat) induced alternations of disease-related markers and physical symptoms [161-164]. These neuroprotective effects of IG in the mouse model are further supported by using a *Drosophila melanogaster* model, in which IG extract decreases rotenone-induced locomotor disorders and neurological toxicities [165]. In Huntington's model, nitropropionic acid was used to induce Huntington-like abnormities, and IG extract can protect against these abnormities via anti-oxidant effects [166]. These modern studies suggest that IG might contain phytochemicals with neuroprotective activities.

IG can potently be against infections from bacteria, parasites, and fungi. The anti-bacteria properties of IG can be both seen in Gram-positive and Gram-negative species [167-169]. The mechanism might be partially related to gene regulation and immunology [170]. Other than having anti-bacteria activities from IG itself, IG can even synergize anti-microbial agents, rifampicin, and isoniazid [171]. Similarly, IG has anti-parasite activities itself, and it can synergize with cisplatin on *Leishmania donovani* [172]. The IG also shows anti-parasite activities in malarial mice [173]. For anti-fungi effects, components from IG, such as glycoprotein and flavonoids, show fungistatic

properties [168, 174]. These studies not only explain the native immune defense system in plants, but also suggest the drug-like potential of IG for future applications.

For anti-cancer, IG extraction and phytochemicals show activities in cells *in vitro* or in mice *in vivo*. For instance, IG leaf extract can kill cancer cells [175], and root extraction inhibits the proliferation of Hep2 cells (the human laryngeal carcinoma) [176]. Withaferin A, as a representative phytochemical of IG, induces cell death of human melanoma cells [177], human myeloid leukemia HL-60 cells [178], human breast cancer cells [179], and renal carcinoma cells [180]. The cell death of these cancer cells is via apoptosis or apoptosis-related pathways. Since the anti-cancer activities of IG can be unspecific, more work on IG specifically killing cancers is still required.

Other activities of IG are related to anti-diabetes and cardioprotection. In diabetes, IG shows pharmacological effects on rat models and humans, as demonstrated in humans by stabilization of blood glucose after administration of root powder. The anti-diabetic effects are further demonstrated in rats by IG normalizing diabetic parameters after treatments of root/leaf extracts [181-183]. In cardiovascular diseases, IG can exert protective effects on doxorubicin-induced cardiotoxicities [184], isoprenaline-induced cardio necrosis [185], and ischemia-reperfusion-caused heart injuries [186]. These studies show that IG can be a promising phytochemical pool for drug development, providing alternative solutions for diabetic and cardiac diseases.

1.5. *Schisandra chinensis*

1.5.1 Chinese traditional medicine *S. chinensis*

Schisandra chinensis is commonly known as five flavor berries (FFBs) that were included in an ancient book “Shen-nong’s Herbal Classics” that has a history of around 2000 years in China, and their medical value is described in the upper-grade category. FFBs are a type of medicine with sour, sweet, bitter, spicy, and salty tastes. Nowadays, although FFBs might be obtained from the genus Schisandra with 25 different species, medicinal FFBs are only derived from two plant species: *schisandra chinensis* and *schisandra sphenanthera* [187] (**Figure 1.25**). These two species differ from each other by growing regions, because *S. chinensis* is from north-eastern China [188], while *S. sphenanthera* is from the south of China [187] (**Figure 1.26**). For this geographic reason, FFBs can be classified into north- and south- FFBs. Another difference is that the north-FFBs possess higher medicinal value than the south counterpart, based on Chinese pharmacopeia. In the rest part of my thesis, we shall use FFBs to stand for *S. chinensis*-derived berries, if there are no specific indications.

Figure 1.25 Five-flavor berries from *Schisandra chinensis* and *Schisandra sphenanthera*.

(A) Fruit of S. chinensis. **(B)** *S. sphenanthera*. This figure is from the work of Zijian Li et al. [187].

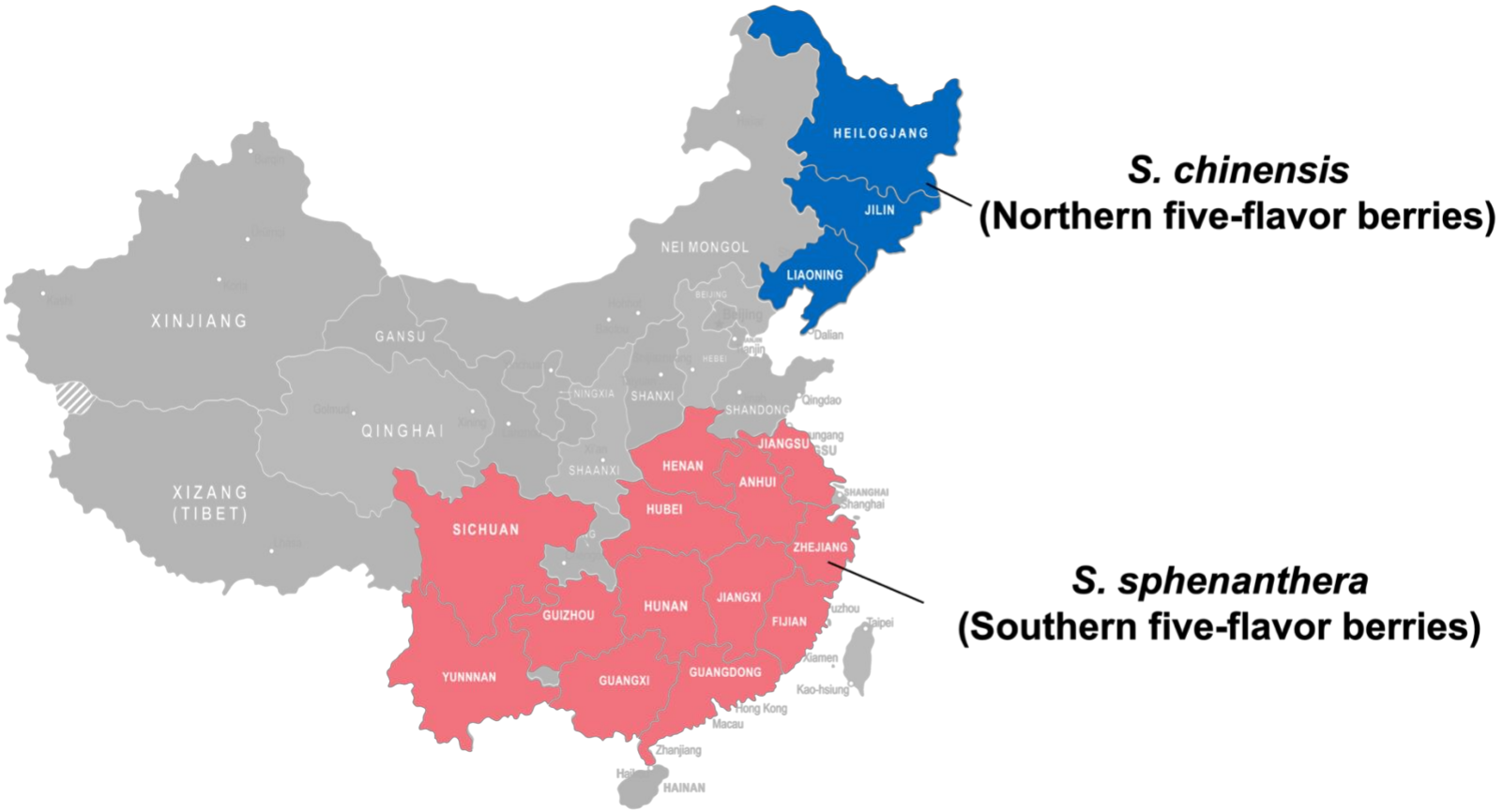


Figure 1.26 Geographical distributions of northern and southern five-flavor berries.

This picture was purchased from Shutterstock (website link: https://www.shutterstock.com/) and then modified by me.

1.5.2 Preparations of the *S. chinensis*

FFBs are often considered as a traditional Chinese medicine with thousands of years of clinic history, but their modern research started in USSR (Union of Soviet Socialist Republics, or former Russia) during 1940s-1960s. In 1943, USSR leader, Joseph Stalin, signed an order (No. 4654-P of People's Commissars Council) to study pharmacological functions and active components of FFBs. He aimed to improve physiological and mental conditions of USSR people in the Second World War. For decades of hard work, Rusia officially finalized two groups of FFBs preparations: the fruit preparation and the seed preparation. The fruit preparation can be air-dried fruit, a tincture of fruit with 95% ethanol at the ratio of 1/6 (w/v), or a mixture of fruit and water at the ratio of 1/20 (w/v) [189-191]. These fruit preparations are suggested to be administrated twice a day at ~1 g for fruit, 25 drops for the ethanol tincture, and 150 ml for water solution [189-191]. Like the fruit preparations, the seed preparations also have three different types: 1) seed powder [192]; 2) seed tincture prepared by adding five volumes of 95% ethanol into seeds [192]; 3) seed extract solution made by using one volume of 95% ethanol [193]. The seed preparations can be administrated every two days at ~1 g for the seed powder, ~25 drops for the seed tincture, or ~1.2 ml/kg for the seed extract [192, 193].

1.5.3 Phytochemicals of the *S. chinensis*

FFBs, as ancient medicines, contain remarkable molecules with functional potential. These molecules might be polysaccharides, vitamins, essential oils, and lignans [194]. Nowadays, although components like polysaccharides also are reported to exhibit biological functions [187], the main active components in *S. chinensis* are treated as a specific type of lignans: dibenzocyclooctadiene lignans. This type of lignans is characterized by its common skeletal structure, consisting of two aromatic rings attaching to an eight-carbon ring (**Figure 1.27**). In 1951, the first dibenzocyclooctadiene lignan, schizandrin in a crystalline form, was obtained from *S. chinensis* fruits [195]. Since then, around 30 different dibenzocyclooctadiene lignans were further isolated and characterized. Examples of these lignans are schisandrin (A, B, and C), schisantherin A, and Gomisin A (**Figure 1.27**). The compositions of lignans in FFBs might highly vary from batch to batch, and this variation could be caused by factors like geographically growing conditions of FFBs [196]. Since lignan compositions cannot be used as a quality of standard for FFBs, we speculate that there are other unknown compounds with pharmacological activities in FFBs remaining to be discovered and characterized.

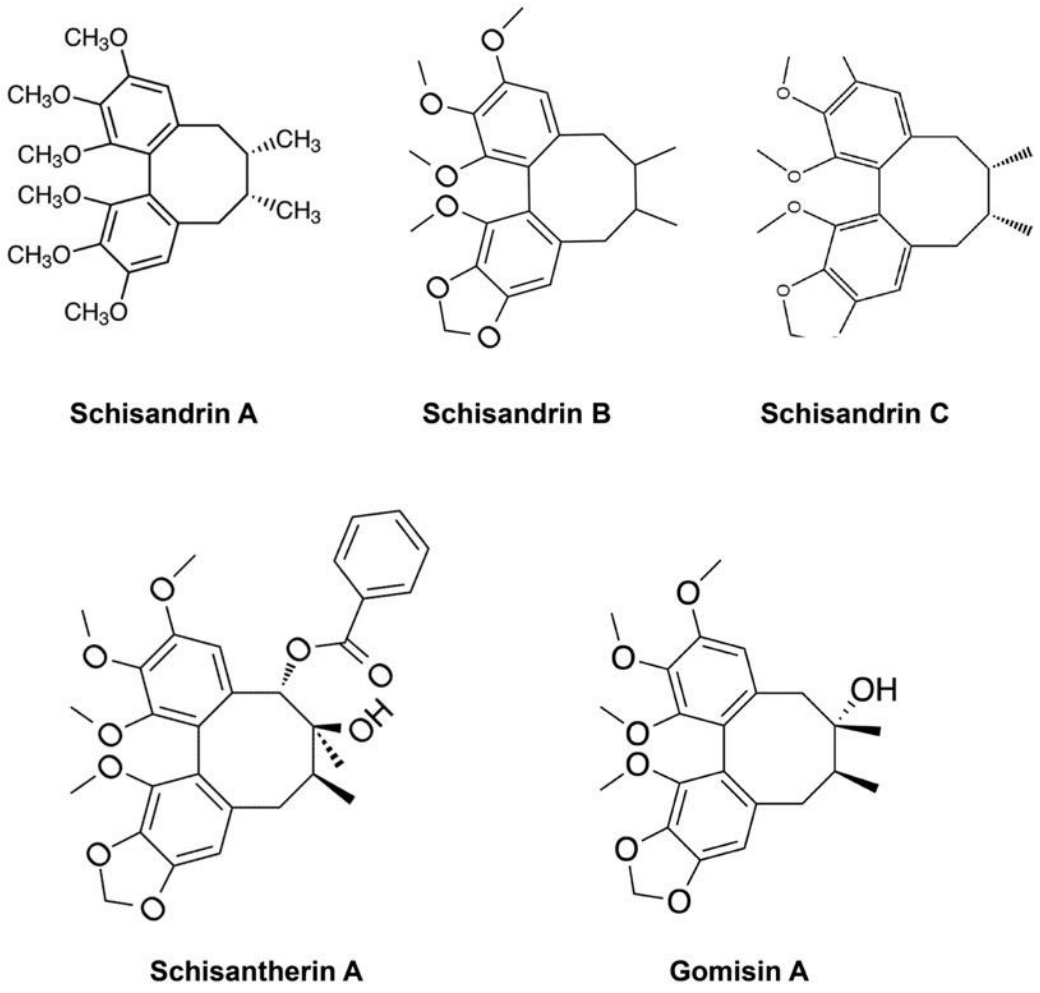


Figure 1.27 Representative phytochemicals from the *S. chinensis*.

1.5.4 Pharmacological activities of the *S. chinensis*

After 1960, based on pharmacological studies, FFBs gradually established their role in Russian pharmacopeia, and they are treated as the “adaptogen”, a term coined for substances with non-specific anti-stress activities [197, 198]. In other words, FFBs can possess various biological functions, by adapting human beings to different environmental stressors, physiologically and mentally.

Being an adaptogen, FFBs preparation regulates multiple human systems, involving gastrointestinal, cardiovascular, nervous, and inflammatory systems, especially when the human body is exposed to stressors [188]. Then, these multi-systematic regulations might synergistically lead to final pharmacological effects, such as anti-fatigue. For example, when two groups of Red Army soldiers were running 20 km, the soldiers (n=23) with the administration of FFBs seeds showed less exhaustion, thirst, and muscular pain, compared with the soldiers in the placebo group (n=22) [199]. FFBs preparations can increase not only physical capacities of individuals, but also mental performances. A group of ~22 year-old telegraph-operators (n=20), with a five min task of transmitting Morse code, had higher accurate performance after administration of FFBs preparations, compared with operators without such administration (n=23) [200, 201]. Other than showing anti-fatigue physically and mentally, FFBs preparations are also involved in balancing reflexes in post-traumatic encephalopathy [202], improving the coagulation system of pregnant women [203], and inhibiting local inflammatory activities of the skin [188].

Apart from preparations of FFBs, dibenzocyclooctadiene lignans from *S. chinensis* were also intensively studied. These lignans show pharmacological activities in neuroprotection and cardioprotection at least *in vitro*. For instance, schisandrin A show neuroprotective effects in different cell-line based models, such as LPS-induce

neuroinflammation on the microglia BV-2 cell model [204], oxygen deprivation/reperfusion on the rat cortical cell model [205], deprivation of nutrition on the neuroblastoma SH-SY5Y model [206]. Other lignans like schisandrin B can improve cognitive function *in vivo*, as demonstrated by reducing mice memory loss that was impaired by β-Amyloid [207]. Cardiovascular diseases can be caused by disturbance in many biological processes, including vascular relation [208], fibrosis [209, 210], inflammation [211, 212], and apoptosis [213]. In these processes, Schisandrin B from FFBs can exert positive pharmacological activities and might decrease the risk of suffering cardiovascular diseases. For instance, Schisandrin B can increase NO production via promoting phosphorylation of eNOS, leading to vasodilation in rat thoracic aorta [214]. Also, the schisandrin B can not only downregulate cytokines TNF-α and IL-1β to inhibit inflammation for cardioprotection [215], but also can inhibit apoptotic pathways in cardio myoblast [216]. Taken together, due to these pharmacological effects mentioned above, FFBs are widely recognized as adaptogens in modern pharmacology.

1.6. Familial hypercholesterolemia

1.6.1 Familial hypercholesterolemia

Patients with allele defections in LDLR and PCSK9 genes normally suffer familial hypercholesterolemia (FH). Heterozygous familial hypercholesterolaemia (HeFH) is the most common disorder in FH with one allele defective and another allele normal, and its incidence ratio is 1 out of 500 patients [4]. In contrast, homozygous FH (HoFH) is relatively rare and its incidence ratio is one out of a million lives. Patients with the HoFH often develop cardiovascular diseases in childhood, for their high low-density lipoprotein cholesterol (LDL-C) in circulation [4]. In other words, LDLR and PCSK9 play key roles in FH development. These two proteins, thus, shall be discussed in the following sections.

1.6.2 Low-density lipoproteins (LDL) receptors (LDLR)

1.6.2.1 LDLR entering cell by clathrin-mediated endocytosis

Unlike EGFR moving to coated pits after binding to EGF ligands [217, 218], LDLR can continually move to coated pits and then to intracellular coated vehicles with/without binding of LDL ligands. After binding to LDL, LDLR/LDL complex is internalized and delivered to endosomes that have a low pH environment inside. This acidic condition disassociates LDL from LDLR [219, 220]. The LDL is then transported into lysosomes for degradation, while LDLR in vehicles budded from endosomes recycles back to the cell membrane. Since LDLRs might remain clustered, they are easily trapped again by coated pits for another round of internalization and recycling [221]. One round of recycling takes 10-20 mins and the lifespan of LDLR is 10-30 h. These characteristics allow such ligand internalization to be highly efficient. Since LDL goes through the acidic endosome during the process, the binding domain of LDLR must be highly stable and the structural domains of LDLR are discussed in the following section.

1.6.2.2 Structural domains of LDLR

LDLR protein sequence with 839 residues can be divided into five domains (**Figure 1.28**), after SPase removes the signal sequence of 21 residues in the LDLR precursor [222]. The first domain at the N terminal contains 292 residues and it is a cysteine-rich domain. This domain is external to the cell membrane and contains seven repeats, each being ~40 residue long and possessing the C-C-C-C-C-C motif. Importantly, these repeats have significantly clustered negative residues, making them negatively binding to positive residues of ligands, like apolipoprotein E [223]. The second domain contains ~400 residues and shares the homology of the EGF percussor [224-226]. This domain contains three sequence repeats (A, B, and C), and each repeat contains six cysteine residues. The repeats are homologous only to the EGF precursor [227-229], but also to blood clotting proteins [228, 229]. Although the functional relationship between the second domain, the EGF precursor, and blood clotting proteins is unclear, these three were suggested to be likely caused by gene duplications [224]. The third domain has 58 residues, 18 of which are serine or threonine residues [225]. The domain contains O-linked carbohydrate chains with a core of N-acetylgalactosamine and might function as a strut for LDLR standing on the membrane. The fourth domain is cell-spanning with hydrophobic residues, and it is not well conserved, when human and bovine LDLRs are compared with each other. Unlike the fourth domain, the fifth domain is well-conserved between species, and it is located on the cytoplasmic side of the cell membrane. In the fifth domain, the sequence near the cell membrane is positively charged, while the C terminal far from the membrane has a negatively charged cluster. These properties in the fifth domain are important for clathrin-related binding, which is a crucial step of LDLR internalization.

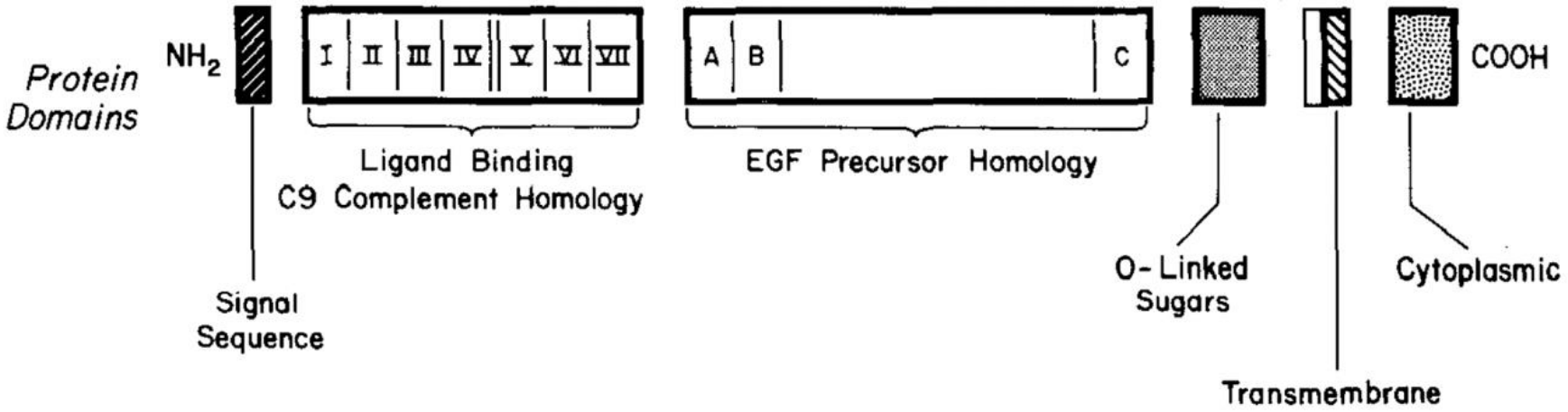


Figure 1.28 LDLR structural domains

This figure is the work of Joseph L. Goldstein [222].

1.6.3 PCSK9

1.6.3.1 PCSK9 genetics

Mutation of *PCSK9* can be gain-of-function (GOF) and loss-of-function (LOF), based on mouse and human studies. In mice, the GOF that leads to upregulation of PCSK9 can cause hypercholesterolemia [230], while the LOF that leads to downregulation of PCSK9 can cause hypocholesterolaemia [231]. In a human population study, DNA sequencing reveals *PCSK9* LOF variants (e.g. pY142X and p.C679X) are related to a 40% reduction of LDL-cholesterol (LDL-C) level [232]. The existence of these variants might differ in different populations, as Africans in the US appear to have some LOF variants more commonly than Europeans in the US [232]. Another 15-year study on LOF variants of *PCSK9* suggested that US Africans (around ~3%) had low LDL-C and a risk of suffering cardiovascular diseases. These mutation sites in *PCSK9* gene together with others are shown (**Figure 1.29 a**). Thus, these genetic studies suggest that PCSK9 could be a drug target for lowering LCL-C and decreasing the risk of suffering cardiovascular diseases.

1.6.3.2 Structural domains of PCSK9 precursor

PCSK9 precursor contains 692 amino acid residues by the *PCSK9* gene on chromosome 1P32 [233]. It has four domains, including a signaling domain, a pro domain, a catalytic domain, and a cysteine/histidine-rich domain (**Figure 1.29 b**). The signaling domain guides the precursor to enter the endoplasmic reticulum (ER) and then is cut by signal proteinases. In the ER, the pro domain is cleaved at Gln_{152} and then blockades the catalytic triad of the catalytic domain [234]. In this case, the mature PCSK9 without exposing its catalytic triad is excreted for executing its functions [235]. In contrast to its normal condition, PCSK9 excretion can be reduced, if LOF mutations happen in *PCSK9*. These mutations can be found in all of the domains, but mainly in the catalytic domain of the PCSK9 precursor [236], leading to RNA decay, autocatalytic inhibition, and protein misfolding [237, 238]. Unlike LOF, GOF is relatively rare to be studied in a population, but it is suggested to increase the risk of getting cardiovascular diseases [239-241]. These findings of structural domains and point mutations laid a solid foundation for later drug discovery of PCSK9 inhibitors.

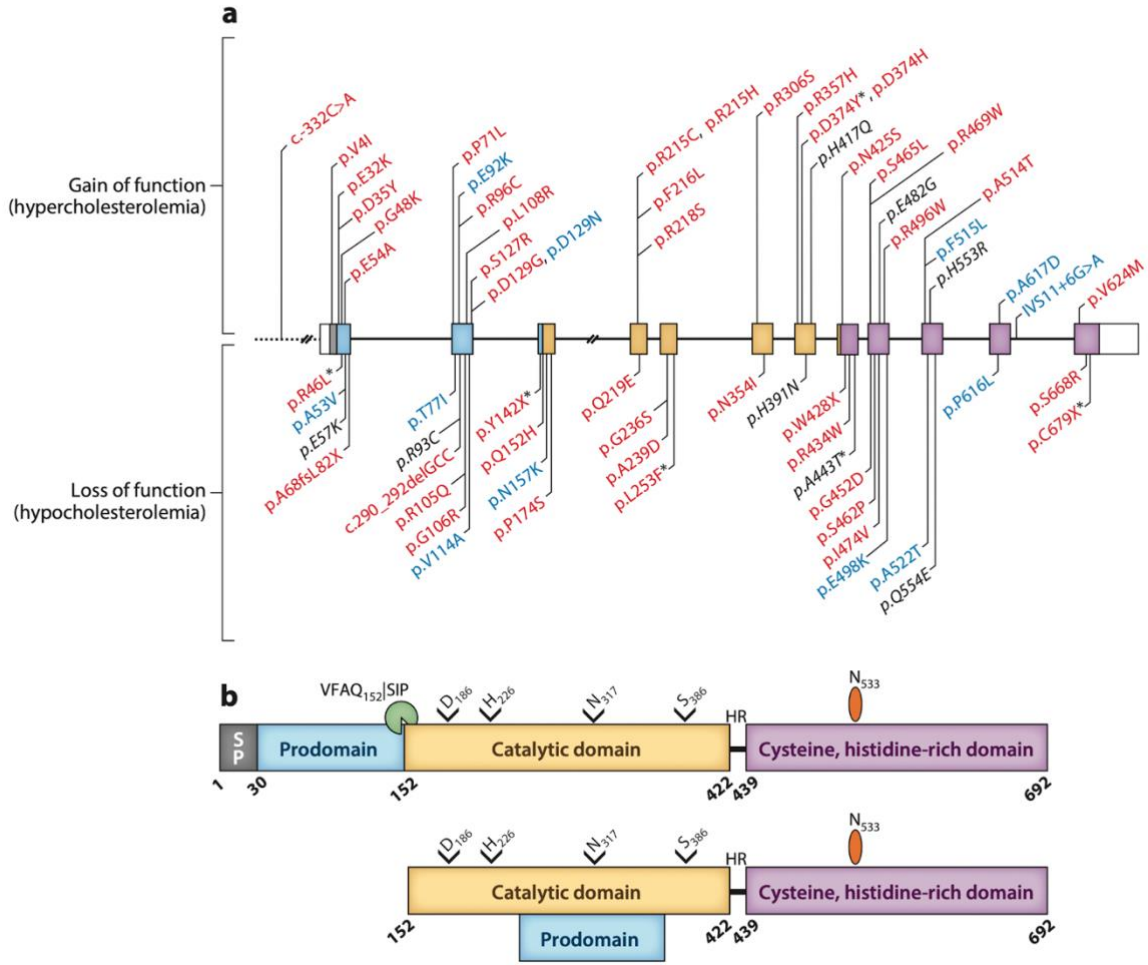


Figure 1.29 Mutation sites and structural domains of PCSK9 or its precursors.

This figure is the work of Amy C. Burke et. al [242].

1.6.3.3 Transcriptional regulation of PCSK9

PCSK9 is regulated by at least two transcription factors: sterol regulatory element-binding protein-2 (SREBP-2) and hepatocyte nuclear factor-1α (HNF-1 α). The activated SREBP binds to the sterol regulatory element (SRE) in *PCSK9* proximal promoter, leading to PCSK9 gene transcription [243]. The SREBP has two isoforms: 1) SREBP-2 that regulates cholesterol biosynthetic/uptake genes; 2) SCREBP-1c that mainly regulates genes related to fatty acid synthesis [244]. In response to low cholesterol in ER, SREBP is proteolytically processed into the active form, which then enters into the nucleus and binds to SRE. Apart from *PCSK9*, the gene of LDLR is also transcriptionally regulated by SREBP-2. For instance, statins can activate SREBP-2 by inhibiting the HMG-CoA reductase in the cholesterol biosynthetic pathway. The active form of SREBP-2 increases LDLR gene transcription and thus corresponsive protein expression, leading to LDL-C uptake in hepatocytes and LDL-C decrease in blood circulation. Since statins activate both PCSK9 and LDLR genes, the circular LDL-C lowering effect suggested that statin-induced LDLR upregulation dominates the effect, while statin-induced PCSK9 upregulation can neutralize overexpression of LDLR. Other than SREBP, the HNF-1 α also can bind to a site near SRE in *PCSK9* for upregulation [245]. The activity of HNF-1 α can be altered by other external factors. For example, insulin can induce exclusion of HNF-1 α from the nucleus, limit PCSK9 gene expression, and upregulate LDLR protein [246]. Thus, HNF-1 α and SREBP cooperate together to regulate the transcription of *PCSK9*.

1.6.3.4 Mechanism of PCSK9-induced LDLR degradation

PCSK9-induced LDLR degradation can happen intracellularly and extracellularly. Inside of cell, PCSK9 binds to LDLR in the Golgi network and increases transportation of LDLR/PCSK9 complex from *trans* Golgi to lysosomes for proteolytic degradation. Also, the LDLR degradation can be activated extracellularly, when excreted PCSK9 in the active form binds to the EGF-A domain of LDLR, which extends out of the hepatic cell membrane. The binding site of PCSK9 is located in the catalytic domain [230, 247, 248]. After binding, LDLR is endocytosed in a clathrin-dependent manner and enclosed in the endosome that has an acidic environment. This acidic condition normally helps ligand of LDLR, such as LDL, disassociate from LDLR, which is then recycled to the cell surface. Unlike LDL, PCSK9/LDLR complex did not disassociate from each other in the acidic endosome. This disassociation inhibits the recycling of LDLR and promotes LDLR trafficking to the lysosome for degradation [234, 247, 249].

1.6.4 Drugs for familial hypercholesterolemia

1.6.4.1 Statins

In 1978, lovastatin is the first statin discovered from *Aspergillus terreus* in Merck Research Laboratories. It showed potent inhibiting activities on HMG-CoA reductase, which is a key rate-limiting enzyme in the cholesterol biosynthetic pathway. Although the lovastatin later has revolutionized therapeutics of hypercholesterolemia, at the early stage, the future of lovastatin was doubtful. In 1980, the Merk company ceased clinic trials of the lovastatin and started new safety evaluations in animals, because the compactin that is structurally similar to the lovastatin showed serious toxicity in the clinic. In 1983, after confirming lovastatin safety in animals, the Merck company re-initiated clinic trails. In 1987, the US FDA panel finally approved lovastatin for clinic uses. In addition to lovastatin, other statin drugs 1980s to 2000s, including simvastatin (1988), pravastatin (1991), fluvastatin (1994), atorvastatin (1997), cerivastatin (1998), and rosuvastatin (2003). These statin structures are shown (**Figure 1.30**).

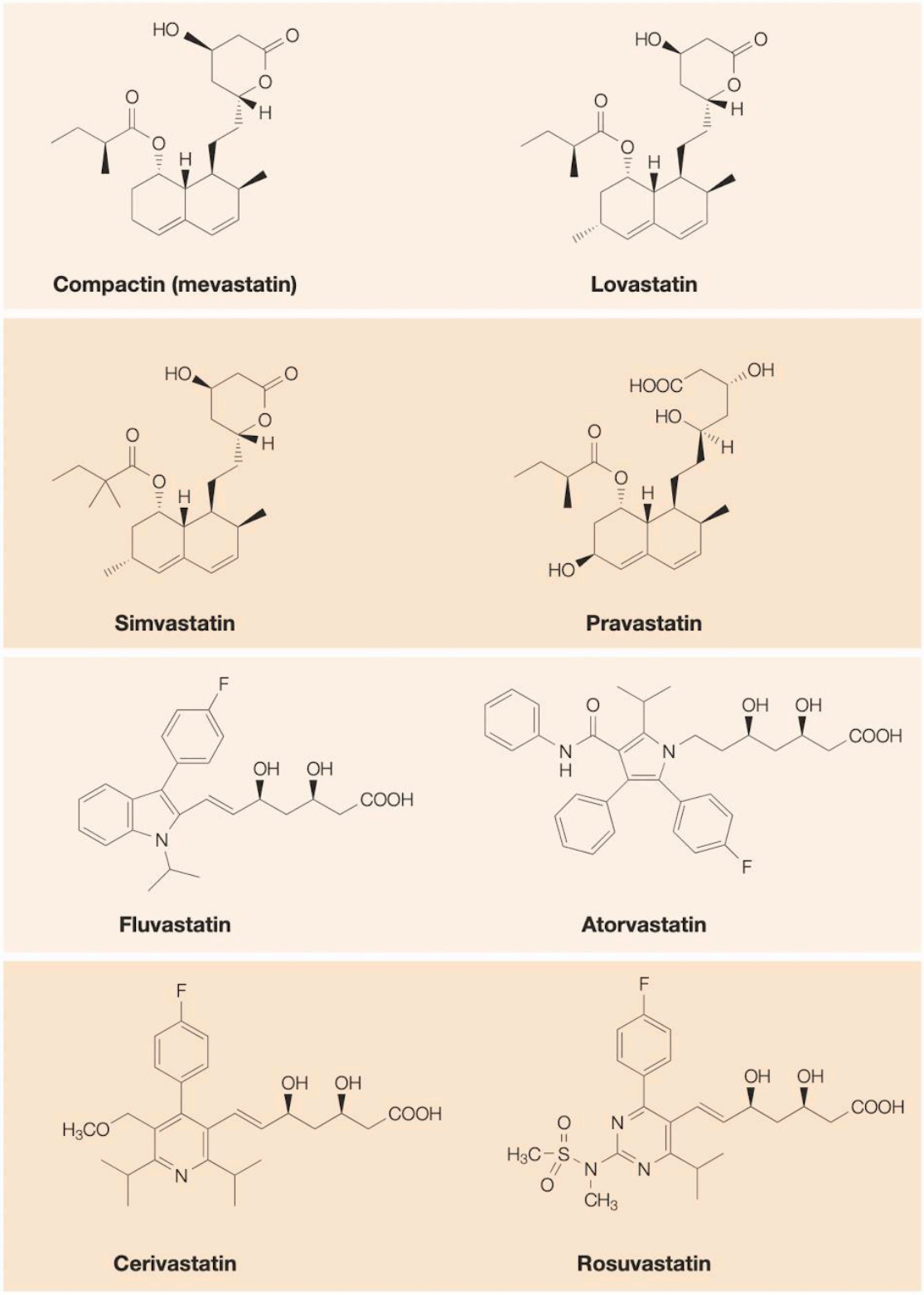


Figure 1.30 Structure of statins

This figure is the work of Jonathan A. Tobert [4].

1.6.4.2 PCSK9 inhibitors

As mentioned above, statins inhibit HMG-CoA reductase and increase LDLR expression at a genetic level. However, statins do not work well on patients that are intolerant to the reverse effects of statins [250]. Patients with HeFH often possess one normal allele of the LDLR gene. In HeFH cases, the PCSK9 inhibitors can inhibit LDLR degradation for increasing the LDLR uptake. To date, PCSK9 inhibitors approved by FDA are two monoclonal antibodies: alirocumab (approved in July, 2015) and evolocumab (approved in August, 2015). Thus, the development of PCSK9 inhibitors provides statin-intolerant patients with a novel therapeutic way of reducing LDL-C in the blood circulation.

1.7 Overview of this thesis

In my thesis, chapter one gives an overview of natural products, plant CRPs, two medicinal plants (*W. somnifera* and *S. chinensis*), areas of under-exploration, the role of LDLR (low-density lipoprotein receptors) in the familial cholesterolemia, and current therapeutics for lowering LDL-cholesterol in blood circulation. Chapter two is the hypothesis and specific aims. Chapter three is the methods and materials. The following two chapters shall discuss structural characterization and functional investigation of two novel CRPs. Specifically, chapter four discusses wisotide (wS1) from *W. somnifera* with characterization of its primary sequences, tertiary structures, homologs, biosynthesis, and phylogenetic trees. Besides, chapter four will cover chemical synthesis, orally active potential, and cell-membrane permeability of wS1, which are crucial for their future functional investigation. Chapter five has three subsections: 1) the first subsection focused on wuweizitide (wZ1) from *S. chinensis* with analysis of its primary sequences, tertiary structures, homologs, biosynthesis, and phylogenetic trees; 2) the second subsection related to cellular uptake pathways, intracellular localization, putative target identification, and transcriptomic regulations; 3) the third subsection covering upregulation of LDLR and investigation of the underlying mechanism: lipid raft accommodating LDLR. Finally, chapter six covers the overall discussion, general conclusions, and preliminary results for future directions.

CHAPTER TWO

Hypothesis and aims

Cysteine-rich peptides (CRPs) from medicinal plants possess high structural stability and great pharmacological potential. They belong to a special chemical space with the molecular weight (M.W.) of 1-10 kDa, making them not only differ from small molecules (M.W. less than 1kDa) and large molecules (M.W. more than 10 kDa) which include biologics such as proteins and nucleic acids, but also likely inherit advantages of both small and large compounds in terms of structural and functional properties. However, the CRPs from the medicinal plants are structurally and functionally underexplored. Thus, we hypothesize that hyperstable CRPs in medicinal plants can represent a novel class of orally active compounds with broad therapeutic applications.

My specific aims are:

1. To discover novel CRPs from two traditional medicinal plants: *Withania somnifera* and *Schisandra chinensis*
2. To characterize these CRPs in terms of primary sequences, three-dimensional structures, biosynthesis, and bioinformatic analysis
3. To chemically synthesize the CRPs for functional investigation
4. To determine the orally-bioactive potential of the CRPs by exposing them to enzymatic conditions
5. To investigate the molecular targets of CRPs. This aim includes testing their cell-membrane permeability, intracellular localization, transcriptomic regulations, potential target identification, and pharmacological effects.

CHAPTER THREE

Methods and materials

3.1 Materials

3.1.1 Chemicals and reagents

Names	Companies
Acetic acid	Merck, USA
Acetonitrile	Merck, USA
AF488 NHS ester	Lumiprobe, USA
Ammonium bicarbonate	Signa-Aldrich, USA
Dimethyl sulfoxide	Signa-Aldrich, USA
Dimethylformamide	Merck, USA
Dithiothreitol	Signa-Aldrich, USA
Dynasore	Signa-Aldrich, USA
Ethanol	Merck, USA
Ethylisopropylamiloride (EIPA)	Signa-Aldrich, USA
Hoechst 33342	Signa-Aldrich, USA
Iodoacetamide	Signa-Aldrich, USA
Lactate dehydrogenase (LDH)	Cell Biolabs, USA
Nystatin	Signa-Aldrich, USA
PKH26 red-fluorescent dye	Signa-Aldrich, USA
PureLinkTM RNA Mini kit	InvitrogenTM, USA
Sodium chloride	Signa-Aldrich, USA
Sodium phosphate	Signa-Aldrich, USA

Trifluoroacetic acid	Tokyo Chemical Industry, Japan
Triisopropylsilane	Signa-Aldrich, USA
1,2-ethanedithiol	Signa-Aldrich, USA
3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT)	Signa-Aldrich, USA

3.1.2 Medicinal plant sources

Root powder of *Withania somnifera* was purchased in an Indian pharmacy. Five flavor berries of *Schisandra chinensis* were purchased from the HUNG SOON Medicinal Trading Pte. Ltd. in Singapore.

3.1.3 Chromatographic materials

Names	Companies	Cat. No.
Filter papers	Whatman, UK	1004150
Funnel	VWR, USA	5110066
Preparative column	Phenomenex, USA	00G-4605-P0-AX
Analytical column	Phenomenex, USA	00G-4632-E0
SPE C18 column	Phenomenex, USA	8B-S001-EAK
SP Sepharose beads	GE healthcare, USA	17072901

3.1.4 Enzymes and biotin probes

Trypsin protease (MS grade) for proteomics and Ultra Streptavidin-HRP for probing the biotin were purchased from Thermo Scientific, USA. Histon-H4 primary antibody was from Cell Signaling Technology, USA.

3.1.5 Cell lines

HepG2 cells (human hepatoma)were purchased from ATCC, USA and HUVEC-CS (a subline of human umbilical vein endothelial cells that are immortalized) were obtained from Professor Kathy Qian Luo.

3.1.6 Culture medium

Names	Companies	Cat. No.
DMEM	Gibco, USA	11995-065
FBS	HyClone, USA	SV30160.03
Penicillin/streptomycin	HyClone, USA	SV30010
Trypsin-EDTA	Gibco, USA	15400-054
Phosphate buffered saline	HyClone, USA	SH30256.01

3.2 Instruments

Names	Companies
LSR Fortessa™ X-20	BD, USA
LSM710 and LSM 980 confocal microscopes	Zeiss, Germany
Cytation1 reader	BioTek, USA
HPLC	Shimadzu, Japan
UHPLC	Shimadzu, Japan
LC-MS/MS (Orbit trap)	Thermo Fisher Scientific, USA
Nanodrop™ 2000/2000c	Thermo Fisher Scientific, USA
NMR spectrometer (800 MHz)	Bruker, USA
5800 MALDI-TOF/TOF	Applied Biosystems, USA

3.3 Proteomic and transcriptomic analysis

3.3.1 Identification

The cysteine-rich peptide screening was conducted by using the 5800 MALDI TOF mass spectrometer (Applied Biosystems, USA). Two medicinal plants (*Withania somnifera* and *Schisandra chinensis*) were screened and show occurrence of CRPs. This identification was confirmed by *S*-reduction with dithiothreitol (DTT) and *S*-alkylation with iodoacetamide (IAA). Specifically, the root powder of *W. somnifera* was mixed with 30% ethanol at room temperature for 1h. The mixed sample was centrifuged by the 5418 microcentrifuge (Eppendorf, Germany) at 12,500 ×g for 30 s. The supernatant was collected and scanned for its mass spectrum. After that, the supernatant was lyophilized and then reduced with dithiothreitol (20 mM) in ammonium bicarbonate (50 mM, pH 8.4) at 60°C for 1 h. The reduced sample was further alkylated with iodoacetamide (40 mM) in a dark box at room temperature for 30 min. Similarly, we identified a new CRP (wZ1) from seeds of *S. chinensis*, except for some details. These are: 1) extraction with 50% ethanol; 2) reduction with 100 mM IAA; 3) conduction with ammonium bicarbonate (25 mM, pH 7.8).

3.3.2 Isolation and purification

For wS1, the root powder of *W. somnifera* (100 g) was stirred with 30% ethanol (500 ml) for 4 h at room temperature. The stirred sample was centrifuged by a J-25 centrifuge (Beckman Coulter, USA) at 8000 rpm for 30 min. The supernatant was collected and filtered by filter papers (Cat. No. 1004150, Whatman, USA). After filtering, the clean solution was diluted to <10% ethanol by using distilled water for C18-beads flush chromatography. The C18 beads were prepared in three steps: 1) loading the beads (~60% of the funnel volume) into the funnel (Cat. No. 511-0066, VWR, USA); 2) washing beads with 100% ethanol (0.5 L); 3) balancing beads with distilled water

(2.5 L). The diluted sample was loaded into the C18 beads, and the flow-through solution was collected. The loaded sample was then eluted out with 0%, 20%, 40%, and 100% ethanol solutions (200 ml each). These eluted solutions and the flow-through solution were subjected to 5800 MALDI-TOF/TOF mass spectrometry for determining existence of wS1. The solutions with wS1 were diluted and then loaded into the Shimadzu HPLC system (HPLC: high-performance liquid chromatography, LC-20 series, Japan). Two columns were used in this HPLC purification: 1) preparative column (Cat. No. 00G-4605-P0-AX, Phenomenex, USA); 2) analytical column (Cat. No. 00G-4632-E0, Phenomenex, USA). Specifically, the wS1 in 5% ethanol solution was loaded into the preparative column (250 × 21.2 mm, 5 μm, 100 Å). The wS1 in the column was eluted at room temperature by using the following HPLC programme: 15% acetonitrile (ACN) in 10 min; 15 to 30% ACN-water gradient in 30 min; 100 % ACN in 10 min; 15% ACN in 10 min. The eluted solution containing the wS1 was determined by MALDI-TOF/TOF and diluted to the solution with <15% ACN for further purification using the analytical column (250 × 4.6 mm, 5 μm, 100 Å). The purification settings of HPLC using analytical column are similar to the one using the preparative column, except for that the flow rate was changed from 5 ml/min to 1 ml/min. Finally, the purified wS1 was lyophilized to remove the water and ACN.

For wZ1, the dried berries (2 kg) were blended with a fruit blender (Toyomi, Singapore) with 50% ethanol. The blended mixture was stirred for 4 h at room temperature and spun down by a centrifuge (Model: J-25, Beckman coulter, USA) at 9000 rpm for 20 min. The supernatant was filtered with the 150mm filter paper (Cat. No. 1004150, Whatman, UK). The filtrate was adjusted to pH 2 and loaded to SP Sepharose beads (Cat. No. 17072901, GE healthcare, USA). The peptide was eluted out with 20 mM sodium phosphate, 1M sodium chloride, and 5% acetonitrile. The eluted sample with

<20% ACN was then purified by RP-HPLC using an analytical column (Cat. No. 00G-4632-E0, 250 × 21.2 mm, 5 µm, 100 Å, Phenomenex, USA). The linear gradient from 20%-50% buffer B/A was applied with a flow rate of 1 ml/min at room temperature using an HPLC system (Shimadzu, LC-20 series, Japan). Buffer A contains 0.1% trifluoracetic acid and 99.9% Milli-Q® water, and buffer B contains 0.1% trifluoracetic acid and 99.9% acetonitrile.

3.3.3 Primary sequence determination

For wS1, the purified peptides were reduced with dithiothreitol (20 mM) in ammonium bicarbonate (50 mM, pH 8.4) at 60°C for 1 h. The reduced wS1 (3394.8 *m/z*) was confirmed by using the MALDI-TOF/TOF mass spectrometer (Applied Biosystems, USA). After confirmation, the reduced peptide was digested by the trypsin (Cat. No. 90057, Pierce™, USA). The digestion was performed at 37°C with the trypsin-to-peptide ratio of 1:50 (w/w) for 24 hours. Finally, the digested sample was cleaned up by C18 column (Cat. No. 8B-S001-EAK, Phenomenex, USA) and subjected to *de novo* sequencing by Orbitrap LC-MS/MS (Thermo Fisher Scientific, USA). The wS1 sequence was confirmed by transcriptomic analysis of *W. somnifera*. Similarly, wZ1 was *de novo* sequenced and its sequence confirmed transcriptomic analysis. There are two main differences: 1) the sequencing was conducted using the 5800 MALDI TOF/TOF mass spectrometer (Applied Biosystems, USA). 2) the transcriptome of S. chinensis is not obtained from online databases, but from our experiments.

To obtain its transcriptome, fresh *S. chinensis* plants were sent to Beijing Genomics Institute (BGI) for sequencing. Briefly, total RNA from *S. chinensis* tissues was extracted using the RNApre Pure kit (Tiangen, China). The RNA quantity and purity were determined using a NanoDrop™ (Thermo Scientific, USA). Next, mRNA was purified by magnetic beads using poly-T oligos. The purified mRNA was cut into small

fragments by divalent cations at an optimized temperature. These fragments were reverse-transcribed into a single cDNA strand using random primers. Using single cDNA strands as a template, double-stranded cDNA was synthesized. It was added with an “A” base and ligated with adaptors for PCR amplification. The double-strand cDNA was used to generate ssDNA circles (single-strand DNA circles) with DNBs (DNA nanoballs). Sequencing was performed using the BGISEQ-500 platform. The raw reads were filtered, and the clean reads were de novo assembled by Trinity. Finally, wuweizitide precursors were determined by blasting the transcriptome library, using *de novo* sequenced wZ1 as a query sequence in Bioedit software.

3.4 Structural analysis

3.4.1 NMR structures

For wS1, the peptide after purification was dissolved in H_2O (containing 5% D_2O) or in pure D_2O at pH 3.5. The Bruker Avance 800 spectrometer equipped with a cryogenic probe at 30 ºC was used to require: 1) ^1H,^1H-2D TOCSY with a mixing time of 80 ms [251]; 2) ^1H,^1H-2D NOESY with a mixing time of 150 ms [252]. W5 pulse sequences were used for the water suppression [253]. By the ^1H-1D spectra using pure D2O, the amide that forms hydrogen bonds was determined in the exchange of hydrogen and deuterium. NMRPipe software was used for visualizing NMR spectra [254], and NMRFAM-Sparky was used for analysis [255]. The solution structures of wS1 were calculated with Xplor-NIH software [256]. The proton-proton distance restraints and four disulfide bond (Cys3-CYs24, Cys7-Cys-29, Cys11-Cys28, and Cys14-Cys23) restraints were employed in standard simulated annealing protocol. Based on the intensities of NOE cross-peaks, the distance restraints were divided into three classes: 1) strong, 1.8 < d < 2.9 Å, 2) medium, 1.8 < d < 3.5 Å and 3) weak, 1.8 A < d < 5 Å. A total of 100 structures were calculated and 20 lowest energy structures were chosen

for data statistics and presentation. The structure was verified using the PROCHECK program [257], and it is presented by using Chimera version 1.15 [258].

For wZ1, the purified peptide was dissolved in a water mixture (95% H2O + 5% D2O) or pure D2O at pH 3.5. The dissolved wZ1 (~1 mM) was then analyzed via NMR using a Bruker Avance 800 spectrometer equipped with a cryogenic probe at 30°C. Two data sets were generated by two experiments: 1) 1H,1H-2D TOCSY with a mixing time of 80 ms [9] and 2) 1H,1H-2D NOESY with a mixing time of 200 ms [10]. W5 pulse sequences were used for water suppression [11]. By the 1H-1D spectra in pure D2O, the amide that forms hydrogen bonds was determined by the exchange of hydrogen and deuterium. NMRPipe software was used to process the NMR spectra [12], and NMRFAM-Sparky was used for data analysis [13].The solution structures of wZ1 were calculated using Xplor-NIH software [14]. The proton-proton distance restraints and four disulfide bond restraints were used for the standard simulated annealing protocol. The distance restraints were divided into three classes based on the intensities of the NOE cross-peaks: strong (1.8 < d < 2.9 Å), medium (1.8 < d < 3.5 Å), and weak (1.8 A < d < 5 Å) intensities. A total of 100 structures were calculated, and the 20 lowest energy structures were selected for statistics and presentation. The structure was verified using PROCHEC program and presented using Chimera version 1.15 [15, 16].

3.4.2 Structural stability

The linear wS1 and wZ1 were prepared by reduction with DTT and alkylated with IAA. Synthetic peptides and S-reduced peptides were subjected to proteolytic degradation. For wS1, the synthetic wS1 and the *S*-alkylated wS1 (2 µl, 0.3 mM) were subjected to degradation of protease (Sigma, USA), pepsin (Roche, USA), and α-chymotrypsin (Sigma, USA). For wZ1, the synthetic wZ1 and S-alkylated wZ1 were incubated with the same three proteases plus the trypsin (Thermo Scientific, USA). The pepsin

digestion was performed at pH 2.4 in 10 mM phosphate buffer, whereas the other protease (trypsin, pronase, and α-chymotrypsin) reactions were performed at pH 7.4 in PBS at 37°C. RP-HPLC was used to determine the quantity of peptide remaining by calculating the area under the curve.

3.5 Bioinformatic analysis

The primary wS1 sequence was used as a query sequence to search for its precursors and homologs *in planta*. The searching database is EST (expressed sequence tags) in NCBI (National Center for Biotechnology Information). The tBLASTn was used as the searching algorithm (with other parameters in default settings). Two different wS1 precursors from *W. somnifera* (with a percent identity of 100%) were found. They have three structural domains: the signaling domain, the mature domain, and the C tail. The signaling domain was predicted by using the SignalP 4.0 Server[259]. The mature domain was the primary sequence of wS1. The C tail domain was deduced from other two domains. For the phylogenetic analysis, the top 100 sequences with significance from the tBLASTn searching were selected. These sequences with repetition were filtered out. The remaining unique sequences were aligned by the MUSCLE (Multiple Sequence Comparison by Log- Expectation) [260]. The phylogenetic analysis after the sequence alignment was displayed and annotated by the iTOL (Interactive Tree Of Life) [261]. After the sequence alignment, wS1 precursors and their representative homologs were selected from different plants, including the goji berry, the eggplant, the tobacco, and the tomato. The conservation in the mature domain of these precursors was presented by WebLogo [262, 263].

The percussor and mature sequences of wZ1 were used for searching homologs in NCBI databases, the OneKP database [264], and the ConoServer database [265, 266]. The signal domain, pro domain, and mature domain of wZ1 were determined by the

Expasy Translate and the SignalP 4.1 [267]. The isoelectric point (PI) of wZ1 was computed by the Expasy ProtParam tool [268]. From the ConoServer database [265, 266], three domains, structure, and physicochemical properties of conotoxin were obtained. Finally, sequence alignment, conservation, and phylogenetic tree of wZ1 were conducted using the methods in wS1 bioinformatic analysis.

3.6 Chemical synthesis

The chemical synthesis of wS1 and wZ1 follows procedures of roseltide rT1[269]. Briefly, linear wS1 and wZ1 on the resin were prepared by Fmoc-SPPS (fmoc-solid-phase peptide synthesis). Then, the peptides were cleaved by trifluoroacetic acid (92.5%) with an equal amount of 1,2-ethanedithiol, triisopropylsilane, and distilled water (each 2.5%). This cleavage was performed at room temperature for 2 h. After that, the peptides were folded in ammonium bicarbonate (100mM) buffer with dimethyl sulfoxide (10%) and cystamine/cysteamine (ratio of amount: 1 to 10) at 4°C for 30 min. The folded wS1 and wZ1 were purified by an analytical column (Cat. No. 00G-4632-E0, Phenomenex, USA) in the HPLC system (Shimadzu, Japan). The synthetic peptides (wS1 and wZ1) and native ones were structurally compared using HPLC co-elution and 1^{H}2D NMR spectrum overlapping.

3.8 Investigation of biological functions

3.8.1 Cell culture

HepG2 cells (HB-8065™, ATCC, USA) and HUVEC-CS (courtesy of Prof. Kathy Qian Luo) were cultured in DMEM (Cat. No. 11995-065, Gibco, USA) with 10% FBS (Cat. No. SV30160.03, HyClone, USA) and 100 units/mL penicillin/streptomycin (Cat. No. SV30010, HyClone, USA). The cell culturing conditions were at 37°C in a humidified incubator supplied with 5% CO_2. Medium for cell culture was replaced with the new one every two days. The cells were sub-cultured when cell confluency reached 80%

of a petri dish, by using the Trypsin-EDTA (Cat. No. 15400-054, Gibco, USA) to trypsinized for 10 min.

3.8.2 MTT and LDH release assays

For the MTT assay of wS1, HepG2 cells and HUVEC-CS were seeded on a 96-well plate with 3000 cells/well for 24 h. The cells were treated with wS1 in different concentrations (0 μM, 0.1 μM, 1 μM, and 10 μM) for 24h, and then incubated with 0.5 mg/mL MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) for 4 h. The medium was removed, and the purple formazan was then dissolved in 100 μl dimethyl sulfoxide (DMSO) with shaking for 5 min. The absorbance was measured at 580 nm by the cytation1 cell imaging multi-mode reader (BioTek, USA).

For cell membrane integrity of wZ1, it was measured by LDH Cytotoxicity Assay Kit (Cat. No. CBA-241, Cell Biolabs, USA). The HUVEC-CS and HepG2 cells were seeded in 96 well plates (3000 cells/well) and then treated with 1-20 μM wZ1 for 48 h. The 45 μl culture medium was collected and incubated with the 5 μl LDH reagent in a cell-culture incubator for 1 h. The OD value was read at 450 nm.

3.8.3 Fluorescent labeling

To label the N terminus of wS1 and wZ1 with AF488 fluorophore, AF488 NHS ester (Cat. No. 21820, Lumiprobe, USA) was used. The 0.86 mg AF488 HNS ester and 0.5 mg peptides were dissolved in 100 μl dimethylformamide (DMF). Then, the solution was mixed with 900 μl sodium bicarbonate buffer (100 mM, pH 8.4). The reaction was conducted at 4°C in the dark overnight. Finally, the labeled peptides were determined by the MALDI-TOF mass spectrometry and purified by the RP-HPLC.

3.8.4 Cell-membrane permeability

HepG2 cells or HUVEC-CS were seeded on a 24-well plate (30,000 cells/well) for 24 h. For the cellular uptake analysis, the HepG2 cells were treated with AF488-wS1 (1

μM) at 37°C for 0 h, 0.5 h, 1 h, 2 h, and 4 h. The treated cells were washed with phosphate-buffered saline (PBS; HyClone, USA) three times. Then, they were detached from the plate by using the 0.05% Trypsin-EDTA (Gibco, USA). The detached cells were mixed with 10% FBS and spun down at 500 ×g for 5 min. After replacing the supernatant with PBS, the cells were vortexed and analyzed by the LSR Fortessa™ X-20 flow cytometer (BD, USA). For temperature-dependent analyses, two groups of cells were pre-incubated at 4°C and 37°C separately for 10 min, and then they were treated with AF488-wS1 (1 μM) at their respective temperature for 1 h. After that, these cells were trypsinized for flow cytometry analyses. For the endocytic analysis, HepG2 cells and HUVEC-CS were pre-incubated with nystatin (50 μg/ml), dynasore (50 μM), or ethylisopropylamiloride (EIPA, 50 μM) for 30 min. These cells were then co-incubated with AF488-wS1 (1 μM) and respective inhibitors for 1 h before flow cytometry analyses. The key cytometer settings are as follows: 1) 10,000 events per sample; 2) excitation laser: 488 nm; 3) emission filter range: 530/30 nm. The generated data were analyzed by Flowjo™10 (BD, USA). Similarly, we analyzed the cellular uptake of AF488-wZ1 in HepG2 cells.

To visualize the cellular uptake of peptides, live HepG2 cells and HUVEC-CS were treated with AF488-wS1 (1 μM) or AF488-wZ1 (1 μM) for 1 h. The cells were then stained with Hoechst 33342 (1 μg/mL) and PKH26 red-fluorescent dye (2 μM) for 10 min. The stained cells were incubated at 37°C and 5% CO_2 and imaged by the LSM confocal microscope (Zeiss, Germany) after 1h and 4 h treatments.

3.8.5 Affinity-enrichment mass spectrometry

A pull-down assay was conducted using NeutrAvidin™ resin (Thermo Scientific, USA). Briefly, 20 μL of resin was incubated with biotin-LC-wZ1 (10 μg/mL) or biotin (10 μg/mL) at room temperature with rotation for 1 h and then blocked with BSA for another hour.

The resins were incubated with HepG2 cell lysate (500 µg) at 4°C with rotation overnight. After incubation, the resins were washed thoroughly 10 times with PBS. The resins were then incubated at 95°C for 5 min, reduced by DTT (20 mM) in NH_4HCO_3 (50 mM, pH 8) at 60°C for 1 h, and alkylated using IAM (50 mM) in the dark for 1 h at room temperature. The alkylated samples were digested with 1 µg/ml trypsin (Pierce, USA) and subjected to high pH (0.1% TEA) reversed-phase fractionation. The fractions were evaporated using a vacuum centrifuge.

LC-MS/MS analysis was conducted using an online Dionex UltiMate 3000 UHPLC system coupled with an Orbitrap Elite mass spectrometer (Thermo Scientific Inc., Germany). Tryptic peptides were dissolved in 0.1% formic acid solution and separated on an Acclaim PepMap RSL column (75 µm inner diameter, 15 cm, 2 µm particle size) using a 60 min gradient of mobile phase A and mobile phase B (0.1% formic acid in 90% can). Samples were sprayed through a Michrom Thermo Captive Spray nanoelectrospray ion source (Bruker-Michrom Inc., USA) with a source voltage of 1.5 kV. A full MS scan range was set to 350–1600 m/z with a resolution of 60,000 at 400 m/z. The Fourier transform-MS/MS scan range was set to 150–2000 m/z with a resolution of 15,000 at 400 m/z. The 10 most intense ions with a threshold of 500 counts were selected for high-energy collisional dissociation fragmentation and fragmented using 32% normalized collision energy with a maximum ion accumulation time of 120 ms. Parameters included an automatic gain control of 1E + 06 for the full MS scan and 2E + 05 for the MS/MS scan, active precursor ion charge state screening, and a capillary temperature of 250°C. Data acquisition was conducted in positive mode using LTQ Tune Plus software.

MetaMorpheus is a software tool used to identify peptides and proteins based on mass spectrometry fragmentation spectra profiles [270]. The raw data obtained from LC-

MS/MS was entered into MetaMorpheus. The human proteome (Proteome ID: UP000005640) was obtained from UniProt.org and used as the reference proteome in MetaMorpheus. Carbamidomethylation of cysteine was set as a fixed modification and methionine oxidation was set as a variable modification. Digestion by trypsin and a maximum of two missed cleavages were defined as the search parameters. Quantification was incorporated using label-free quantification by FlashLFQ. The quantitation was set to be normalized and matched between runs. After the protein groups were quantified, the corresponding data were entered into Perseus for statistical analysis. A two-sample t-test was conducted between the experimental and control samples. Subsequently, a volcano plot was generated to identify the proteins that were enriched at least twofold in the experimental group. These proteins were then filtered against CRAPome (Contaminant Repository for Affinity Purification) [271], a database consisting of commonly occurring background contaminants. The filtered enriched proteins were entered into STRINGdb. STRINGdb is a biological database that compiles known and predicted protein-protein interactions [272]. A protein-protein interaction network was generated to illustrate the relationships between the identified proteins.

3.8.6 Nuclear localization

HUVEC-CS and HepG2 cells were seeded on a chambered slide with a glass bottom (Ibidi, Germany), treated with 1 μM AF488-wZ1 for 1 h, and washed with PBS three times. After washing, the cells were incubated with Hoechst 333242 (1 μg/ml) for 10 min and PKH26 red-fluorescent dye for 5 min. The stained live cells were imaged in real time at 37ºC with 5% carbon dioxide for 4 hours by the LSM 710 confocal microscope (Zeiss, Germany).

For Western blotting analysis, wZ1 was labeled with LC-biotin at its N terminal by using EZ-Link NHS-LC-Biotin (Thermo Scientific, USA). WZ1-LC-biotin was confirmed by MALDI-TOF mass spectrometry, purified by HPLC, and then lyophilized. The purified wZ1-LC-biotin (20 μM) was used to treat HepG2 cells for 24 h. By using nuclear/cytosol fractionation kit (Cat. No. K266-25, Biovision, UK), the nucleus and the cytosol are separated from each other for Western blotting (WB) analysis. The fractionated nucleus and cytosol samples were denatured with WB loading buffer at 95 ºC for 5 min. These two samples were then separated by SDS-PAGE and transferred to a PVDF membrane by a semi-dry electrophoretic cell (Cat. No. 1703940, Bio-Rad, USA). At room temperature, the membrane was blocked with 5% BSA (Sigma-Aldrich, USA) for 1 h, before incubating with the Ultra Streptavidin-HRP (Thermo Scientific, USA) overnight. The next day, the membrane was washed with TBST three times for 10 min each. Finally, the membrane was shortly incubated with Clarity™ Western ECL Substrate (Bio-Rad, USA) and imaged by the ChemiDoc™ imaging system (Bio-Rad, USA).

3.8.7 RNA-seq analysis of wZ1-treated HepG2 cells

HepG2 cells were treated with or without 10 μM wZ1 for 24 h. Total RNA was then extracted and purified using the PureLink™ RNA Mini kit (Invitrogen, USA). For RNA-seq, the RNA samples (five replicates for the control group and five replicates for the treated group) were submitted to BGI for constructing the transcriptome library and sequenced using the DNBSEQ platform. The data analysis was conducted using a Dr. Tom server (BGI, China).

3.8.8 Real-time polymerase chain reaction (qPCR)

HepG2 cells were treated with or without 10 μM wZ1 for 6 and 24 h. Total RNA was then extracted and purified using the PureLink™ RNA Mini kit (Invitrogen, USA). Total

RNA (4 μg) was reverse-transcribed into cDNA using reverse transcriptase (Thermo Scientific, USA), oligo(dT)^{18} (Thermo Scientific, USA), dNTP mix (Thermo Scientific, USA), and an RNase inhibitor (Thermo Scientific, USA). Then, 100 ng of cDNA from each sample (three biological replicates and two technical replicates) were used for qPCR analysis. The mRNA expression level was measured using an SYBR™ Green kit (Bio-Rad, USA) on a CFX96™ machine (Bio-Rad, USA). The primers used for qPCR are provided in the supplementary information.

3.8.9 LDL receptor-mediated LDL uptake

For confocal imaging, HepG2 cells were seeded on an 8-well coverslip (Cat. No. 80826, Ibidi, Germany) overnight and then treated with wZ1 (10 μM) for 24 h. Then, confocal cell samples were prepared using an LDL uptake cell-based assay kit (Cat. No. 10011125, Cayman Chemical Company, USA). To avoid unspecific binding, LDL concentration in the original protocol was further diluted 125 times in our experiments. Briefly, the preparing process is as follows. First, the HepG2 cells were washed with TBS (tris buffered saline) three times and fixed with a fixative solution in the kit. After fixing, the cells were washed with TBS-Triton buffer three times. The washed cells were incubated with a blocking solution and anti-LDL receptor primary antibody. Finally, the cells were washed again with TBS-Triton three times and incubated with DyLight™ 488-conjugated secondary antibody before the confocal imaging.

For flow cytometry analysis, HepG2 cells were seeded on a 6 well-plate overnight and treated with wZ1 (10 μM) for 24 h. The next day, the cells were treated with diluted LDL. On the third day, the cells were trypsinized from the culture plates and pelleted by 500 × g for 5 min using a microcentrifuge. The cells were then washed with TBS once and gently shaken with the TBS-Triton buffer three times for 5 min each. Finally, the cells were blocked with 5% BSA for 30 min and incubated with the FITC-

conjugated LDL receptor primary antibody (Santa Cruz, USA) on ice for 1 h, before the flow cytometry analysis. The results were presented by FlowJo™ software (BD, USA).

3.8.10 Lipid raft analysis

HepG2 cells were seeded into a six-well plate and incubated with 10 μM wZ1 for 24 h. After incubation, the cells were trypsinized from the culture plates and lipid rafts were stained using the Alexa Fluor® 594 Lipid Raft Labeling Kit (Invitrogen, USA). The cells were analyzed using an LSRFortessa™ X-20 cell analyzer (BD, USA). To visualize, hepG2 cells were seeded onto an eight-well chamber slide (Ibidi, Germany) and treated with 10 μM wZ1 for 24 h. After incubation, lipid rafts were stained using the Alexa Fluor® 594 Lipid Raft Labeling Kit (Invitrogen, USA). Then, the cells were fixed and permeabilized. LDLRs were stained using a FITC-conjugated LDLR primary antibody (Santa Cruz, USA) for 1 h, and the nuclei were stained with 1 μg/ml Hoechst 333242 at room temperature for 10 min. The cells were imaged via LSM 980 confocal microscopy (Zeiss, Germany) in a Z-stack Airyscan mode.

3.8.11 Statistical analysis

The statistical analyses were conducted by the Prism 8 software (USA)), if there are no specific indications. The data were analyzed by using the ordinary one-way ANOVA and the groups were compared by Tukey's multiple comparison test. The data were presented as the mean ± SD and P-value < 0.05 was regarded as statistical significant.

CHAPTER FOUR

Wisotide: An anionic cystine-stapled helical peptide from *Withania somnifera* with cell-membrane permeability

4.1 Introduction

Withania somnifera from Solanaceae family is an evergreen shrub growing in India, Mediterranean sea, Cape of Good Hope, and high-altitude places of Himalayas [273]. The root of *W. somnifera* is known as Indian ginseng [274], and it plays crucial roles in traditional Indian medicine (called Ayurveda) [273]. Formulations of Indian ginseng are used for providing physical strength, treating neurological disorders, and improving sleep quality. In this medicine, current bioactive compounds include alkaloids, sitoindosides, and withanolides [275]. The withanolides, like withaferin A and withanolide D, are treated as main active constituents [276]. However, cysteine-rich peptides (CRPs) with bioactive potential from Indian ginseng are unknown.

CRPs lies in a chemical space of 2–6 kDa and are constrained by 2–5 intramolecular disulfide bonds [35]. These bonds render CRPs structurally resistant to thermal, chemical, and proteolytic degradations [99, 143, 277-279]. Although CRP families differ from one another in cysteine numbers, cysteine motifs, and disulfide connectivity [280], these peptides commonly exhibit a large footprint for protein interaction, showing less off-target or side effects [281]. So, these features make CRPs ideal leads for therapeutic development or excellent templates for protein design. Recently, we discovered an 8-cysteine CRP, called lybatide 2, from the cortex of the Lycium barbarum root [99]. The lybatide 2 has an unusual cystine-stapled helical scaffold [99], which enriched our current template pool of plant CRPs.

Structurally, CRP topology can be arbitrarily divided into two major groups: cystine-stabilized (CS) β-peptides and CS α-helical peptides [35]. The CS β-peptides in plants are exemplified by knottins and hevein-like peptides, whereas CS α-helical peptides are represented by defensins and thionins. Unlike CS β-peptides, which are often dominated by short β-strands without α-helices involved, CS α-helical peptides are

characterized by a cystine-stapled helix stabilized by other secondary structural elements, such as an α-helix in thionins and β-strands in defensins.

Here, we report the discovery and characterization of a novel cystine-stapled CRP, designated as wisotide (wS1), from the root of Withania somnifera. The wS1 has a short 3_{10} helix to stabilize its π-helix, a feature similar to lybatides but different from defensins containing a helix stabilized by three β-strands. Although wS1 is anionic and hydrophilic, it can cross cell membranes by micropinocytosis, likely because of its hydrophobic patches that are displayed on the disulfide-stapled helices. Furthermore, over 57 putative wS1-like peptides were found in the transcriptome of over five different plants. Thus, our findings indicate an excellent class of scaffolds for membrane-permeable peptide design.

4.2 Results

4.2.1 Discovery and structural characterization of wisotide (wS1)

4.2.1.1 Identification of wisotide (wS1)

To discover a novel cysteine-rich peptide (CRP) from *Withania somnifera*, we applied MALDI-TOF mass spectrometry (MS). By the MS scanning, we detected a 3386.8 *m/z* signaling in the crude extraction of *W. somnifera* (**Figure 4.1 A**). After *S*-reduction with dithiothreitol, the mass signaling shifted from 3386.8 to 3394.8 *m/z* (**Figure 4.1 B**). This 8 *m/z* increase might be because of four disulfide bonds formed by eight cysteine residues in the peptide. Furthermore, we alkylated the reduced sample with iodoacetamide and found that the 3394.8 *m/z* became 3850.4 *m/z*, with an increase of 463.6 *m/z* (**Figure 4.1 C**). Since iodoacetamide alkylation leads to 57 *m/z* for one cysteine residue, this 463.6 *m/z* increase suggested the existence of an eight cysteine CRP in *W. somnifera*. The wS1 was then extracted and purified in a large scale for the following studies. The yield of wS1 is ~4 mg out of 100 g power of roots.

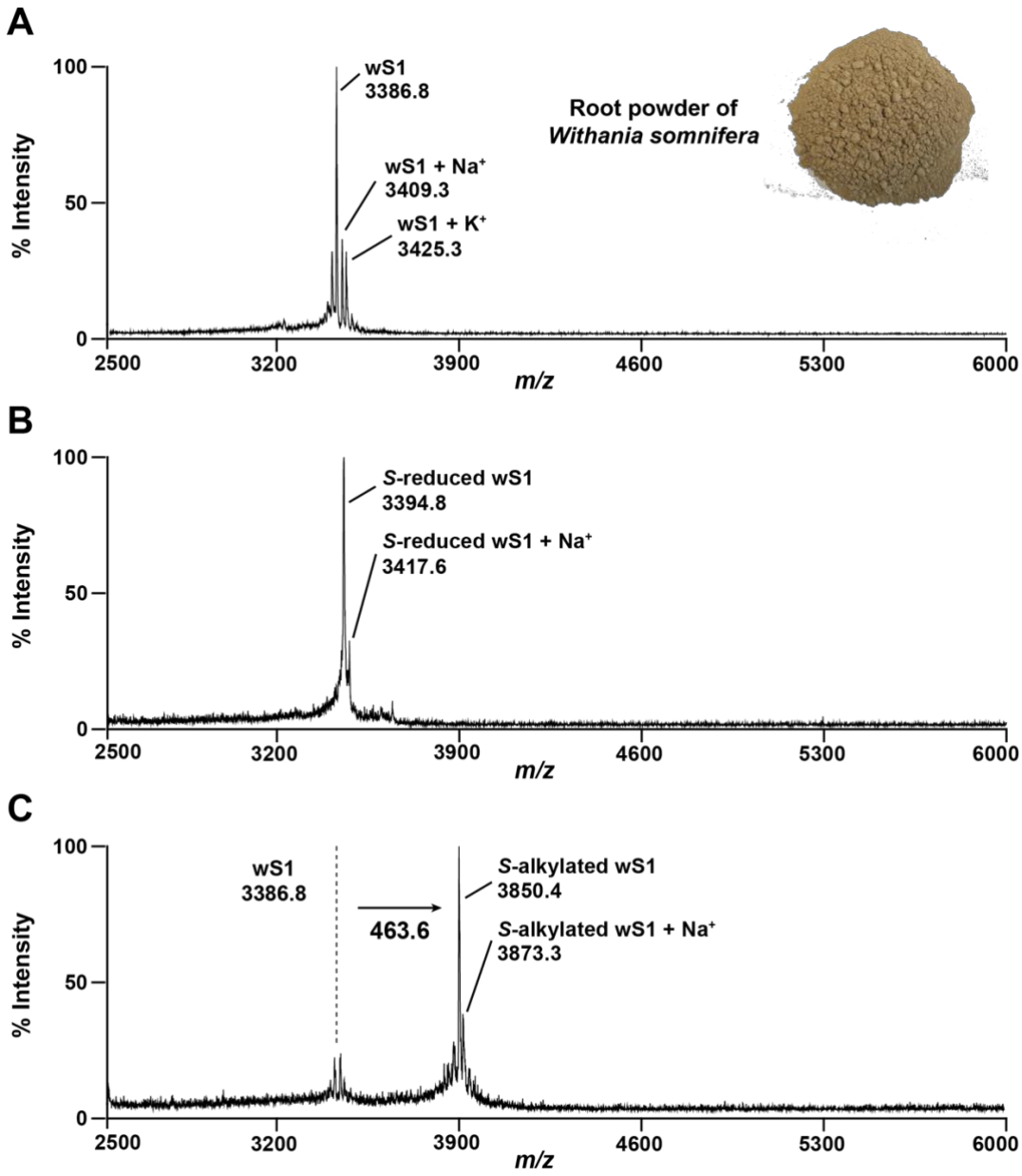


Figure 4.1 Identification of wisotide wS1 from Indian ginseng (*Withania somnifera*).

(A) *W. somnifera* root was extracted with 30% ethanol and then profiled by a MALDI-TOF mass spectrometer. **(B)** The extraction sample was lyophilized and then reduced with dithiothreitol (20 mM) in ammonium bicarbonate (50 mM, pH 8.4). The mass peak shifted forward 8 *m/z*, indicating four disulfide bonds in wS1. **(C)** The reduced sample was alkylated with iodoacetamide (40 mM). The mass peak shifted forward 463.6 *m/z*, confirming the existence of four disulfide bonds. The wS1 was identified by Dr. Wong K. H. initially.

4.2.1.2 Primary sequence determination

To sequence this CRP, we digested the linear peptide (reduced) with the trypsin, and then performed *de novo* sequencing by LC-MS/MS. The digested peptide fragments were sequenced and then mapped together to obtain a full sequence (**Figure 4.2 A**). These fragment sequences were deduced from y ions and b ions generated by LC-MS/MS. Two fragments (red lines in **Figure 4.2 A**) were used as representatives to show deducing details (**Figure 4.2 B**). After transcriptomic confirmation, the full sequence of wS1 is: ADCTEYCSNSCPFCNGQPLYQLCCINNCCPS. Thus, wS1 is a hydrophilic and negatively charged cysteine-rich peptide.

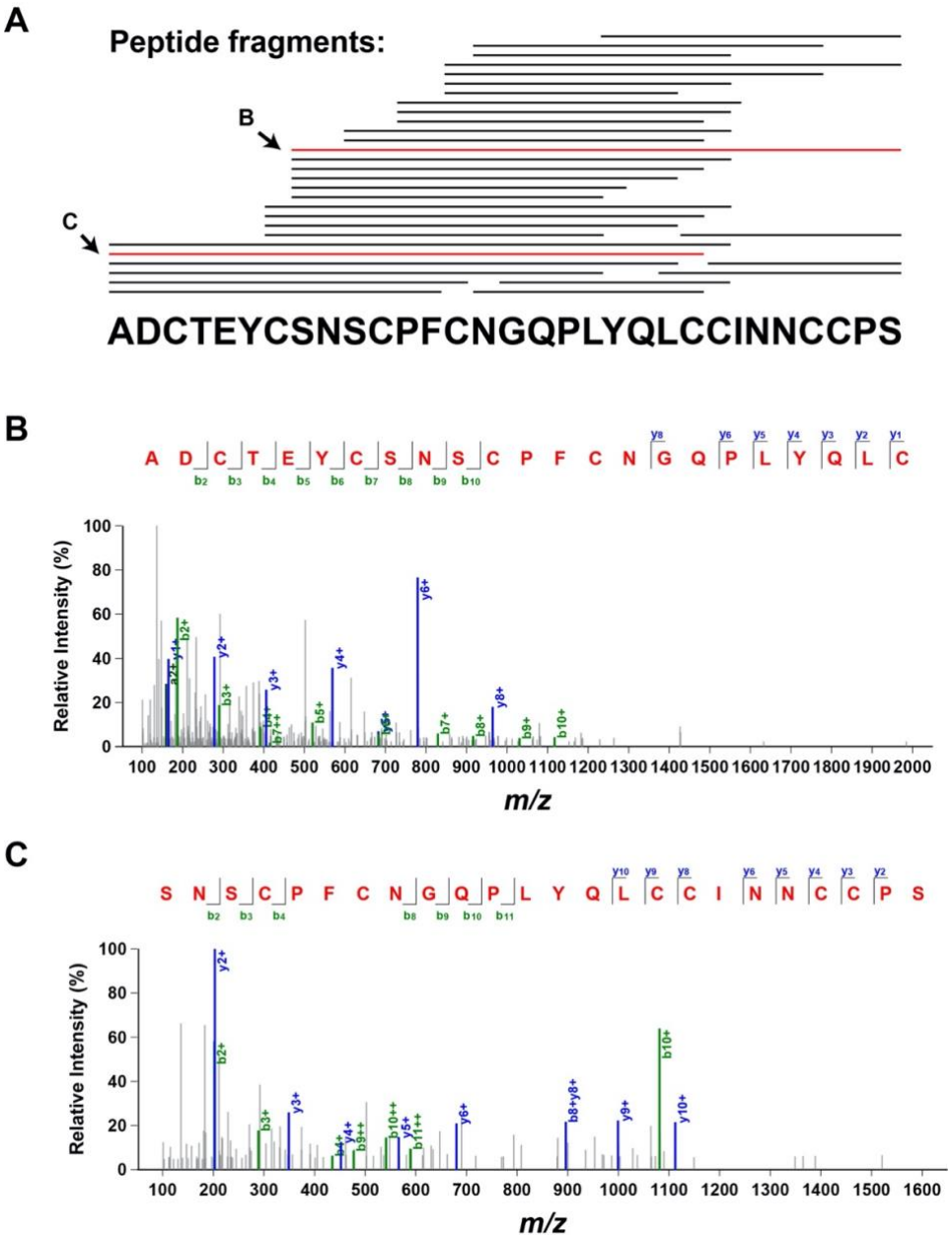


Figure 4.2 Primary sequence determination of wS1.

(A) Purified wS1 was reduced by dithiothreitol (20 mM) and digested into peptide fragments using trypsin at a trypsin/peptide ratio of 1:50 (w/w). The fragments were sequenced using LC-MS/MS and then mapped to deduce the complete wS1 sequence. **(B and C)** The representative *de novo* sequencing of two peptide fragments (red) from the (A). The sequence was confirmed with wS1 precursors using the *W. somnifera* transcriptome. The wS1 sequencing experiment was performed by Dr. Kam A. and Dr. Loo S.N., and then analyzed by us.

4.2.1.3 Disulfide connectivity and cystine-stapled helical scaffold

To characterize the disulfide connectivity and the three-dimensional structure of wS1, we used nuclear magnetic resonance (NMR) spectroscopy. Our results showed that four disulfide bonds were connected as follows: Cys I–VI, Cys II–VIII, Cys III–VII, and Cys IV–V (**Figure 4.3 A**). In the backbone of wS1, we observed two helical structures: a π-helix and a 3_{10} helix (**Figure 4.3 B**). The π helix is stabilized by four disulfide bonds, and the 310 helix has one disulfide bond (Cys IV–Cys V) connected to the π helix (Figure XX). These results that wS1 has a cystine-stapled helical scaffold.

By analyzing surface properties, we found that wS1 has two negative charge residues (D and E) that are close to each other and overall hydrophilic molecular surfaces. Although wS1 shows these unfavorable properties of cell-membrane permeability, the peptide displays hydrophobic surface patches, which are partially supported by the helical scaffold (**Figure 4.3 C**). Thus, we speculate that wS1 can cross cell membranes with these hydrophobic patches.

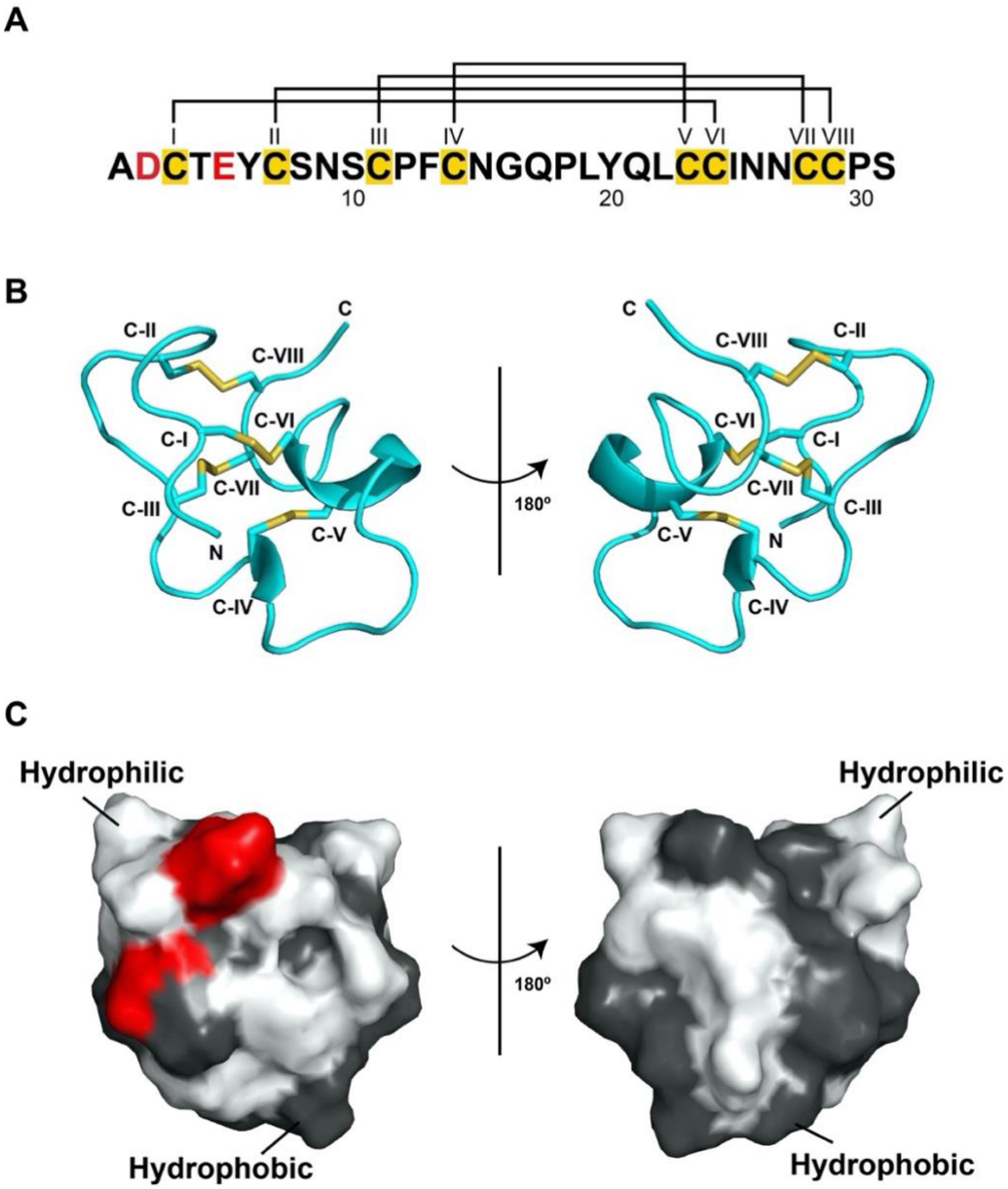


Figure 4.3 Cysteine motif, peptide backbone, and surface properties of wS1.

(A) Primary sequence contains eight cysteine residues (yellow), forming four disulfide bonds (blank lines). Negatively charged residues (D and E) are indicated in red. **(B)** The peptide backbone is locked by four disulfide bonds (yellow) and shows two common secondary structures: the 3_{10} helix and the π helix. **(C)** Structural properties of wS1 are indicated in different colors: negative properties in red, hydrophilicity in white, and hydrophobicity in dark grey. The NMR structure of wS1 was determined by Dr. Fan. J. S. and then analyzed by me.

4.2.1.4 Biosynthesis

For determining the biosynthesis of wS1, we searched the NCBI database with tBLASTn, and then determined structural domains with the SignalP 4.0 Server. We found that two wS1 precursors have the same three domains (signaling, mature, and C tail domains) and the same length (72 residues) (**Figure 4.4**). Although these precursor sequences are similar, they are still slightly different from each other in their signaling domains by one residue (at the 11^{th} position, V in wS1 1.1 versus A in wS1 1.2, Figure 4.4). These findings showed that wS1 has two slightly different precursors, but these two share the same structural domains.

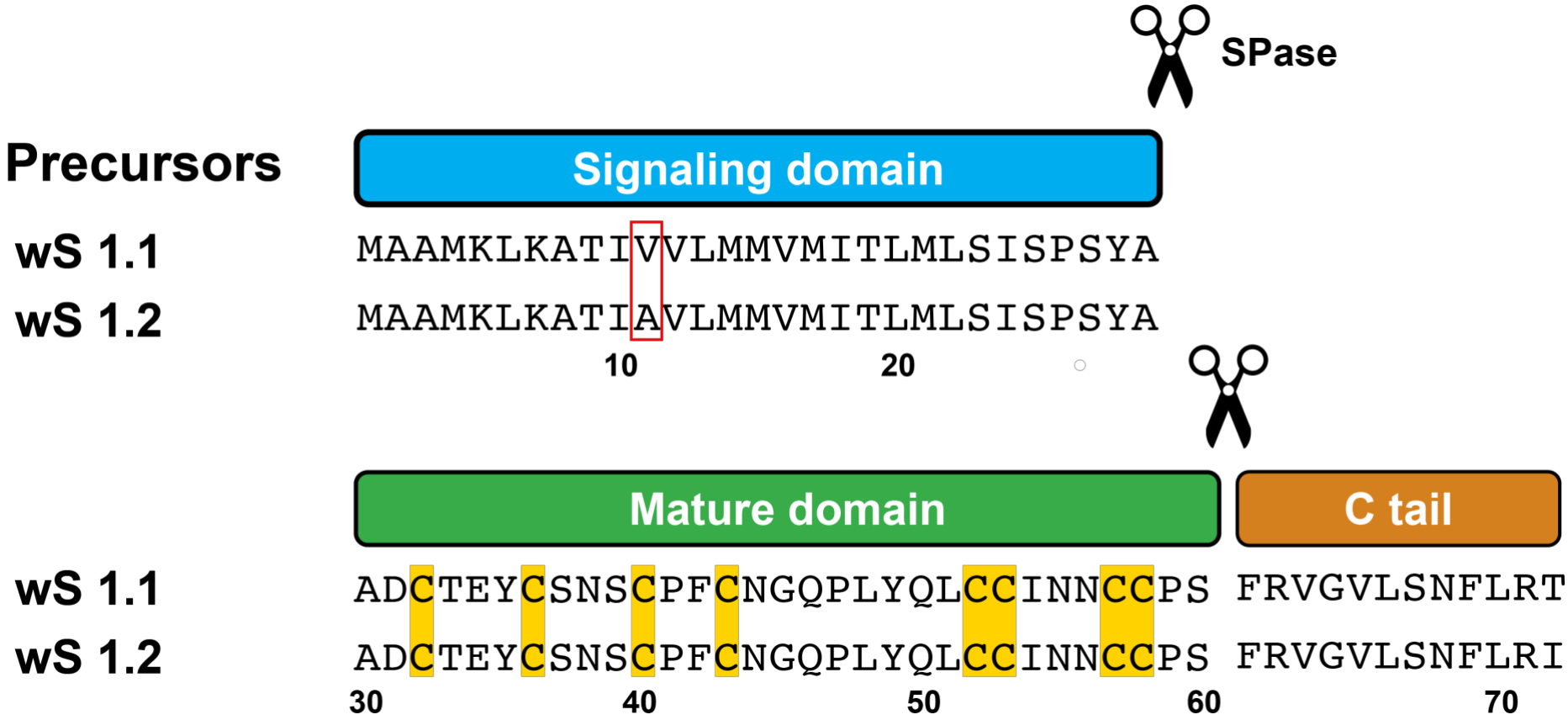


Figure 4.4 Structural domains of wS1 precursors

Two precursors are the same in three structural domains, but slightly different from each other (V in wS 1.1 versus A in wS 1.2, in the red rectangle) in the signaling domain. SPase denotes a signaling peptidase.

4.2.1.5 Phylogenetic tree

To investigate wS1 origins and homolog distribution *in planta*, we drew a phylogenetic tree (**Figure 4.5**). In the phylogenetic tree, we found that the wS1 homologs, *in planta*, are widely distributed in three different families and over different five plants (Indian ginseng, goji berry, eggplant, tomato, tobacco, and others). Although the wide distribution of wS1 homologs, most of the sequences including the wS1 sequence were found in the Solanaceae family. Very few sequences, in contrast, were found in the other two families (Caprifoliaceae and Sargassaceae). Interestingly, goji berry, apart from Indian ginseng, is the only medicinal plant among the five plants. Besides, lybatides from goji berry were the phylogenetically close to wS1, compared with other homologs. Thus, wS1 and lybatides may share the therapeutic potential.

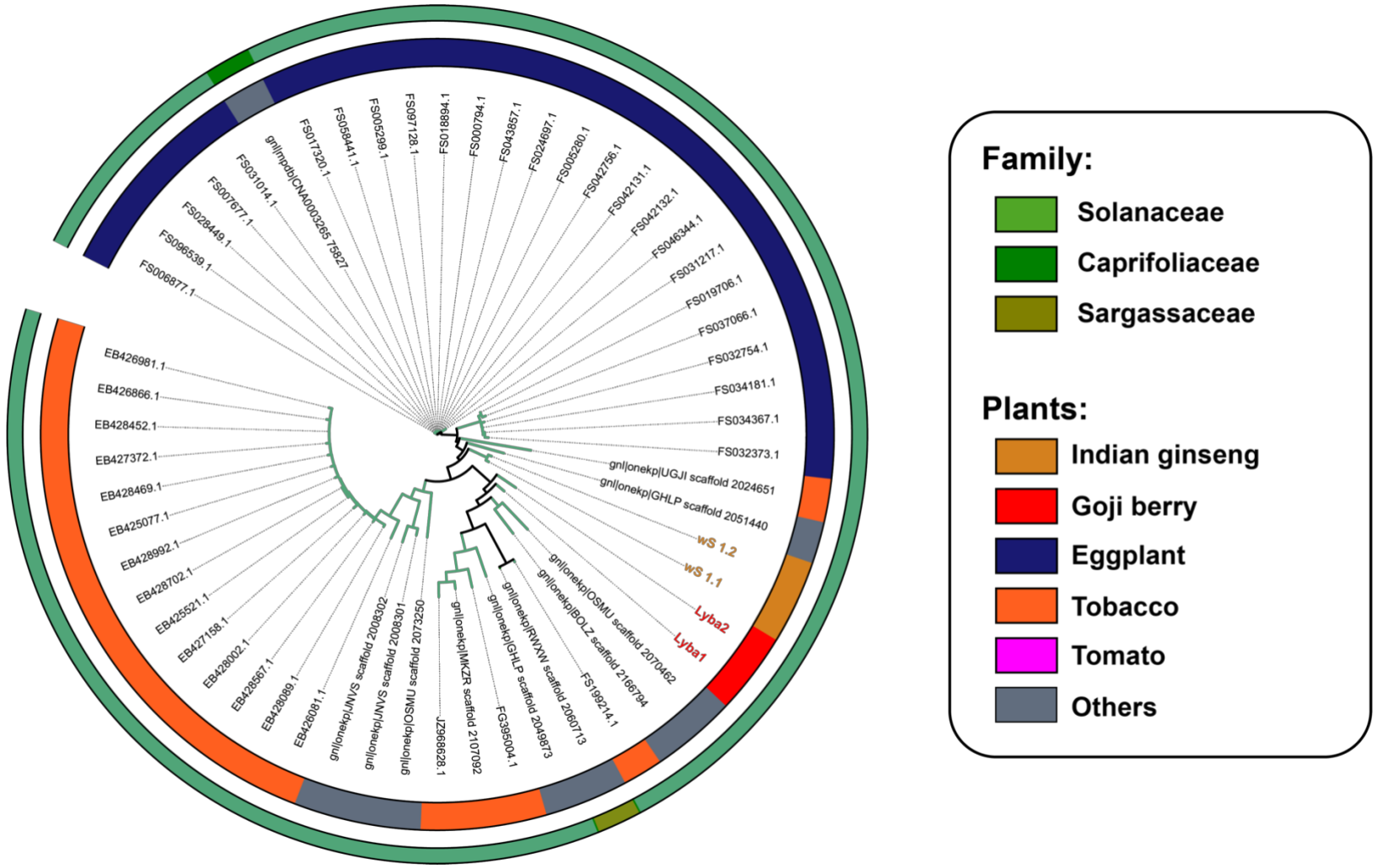


Figure 4.5 Phylogenetic tree of wS1 precursors and their homologs.

Phylogenetic analysis of wS1 precursors (yellow), lybatides (red), and other homologs with the same cysteine motif (dark gray).

4.2.1.6 wS1-like CRPs

To deduce the biosynthetic and structural similarities of wS1-like CRPs, we aligned wS1 precursors with their homologs *in planta* (**Figure 4.6**), and then compared the sequences in Weblogo to show residue conservation (**Figure 4.7**). In sequence alignment, we found that all sequences, except for lyba 1 and lyba 2, contain three structural domains. In these domains, signaling domains are cut by two signaling peptidases (SPases): one cuts at the carboxyl end of alanine residues, and another cut at the carboxyl end of glycine residues. The C tail of these sequences is mainly cleaved by serine endopeptidases, except for FS1992.14.1 from tomatoes. By analyzing residue conservation (the mature domain), we observed that these sequences contain five loops. Charged residues (D, E, R, and K), in these loops, mainly are present mainly in loop 1 and loop 4, whereas few of these are conserved in other loops. These charged residues are also well conserved at the amino-terminus of these sequences (before loop 1). Apart from charged residues, some residues are also conserved, including F in loop 3, P and Y in loop 4, and I and N in loop 5. This analysis suggested that these homologs are not only biosynthetically similar, but also likely to exhibit similar structural and physicochemical properties.

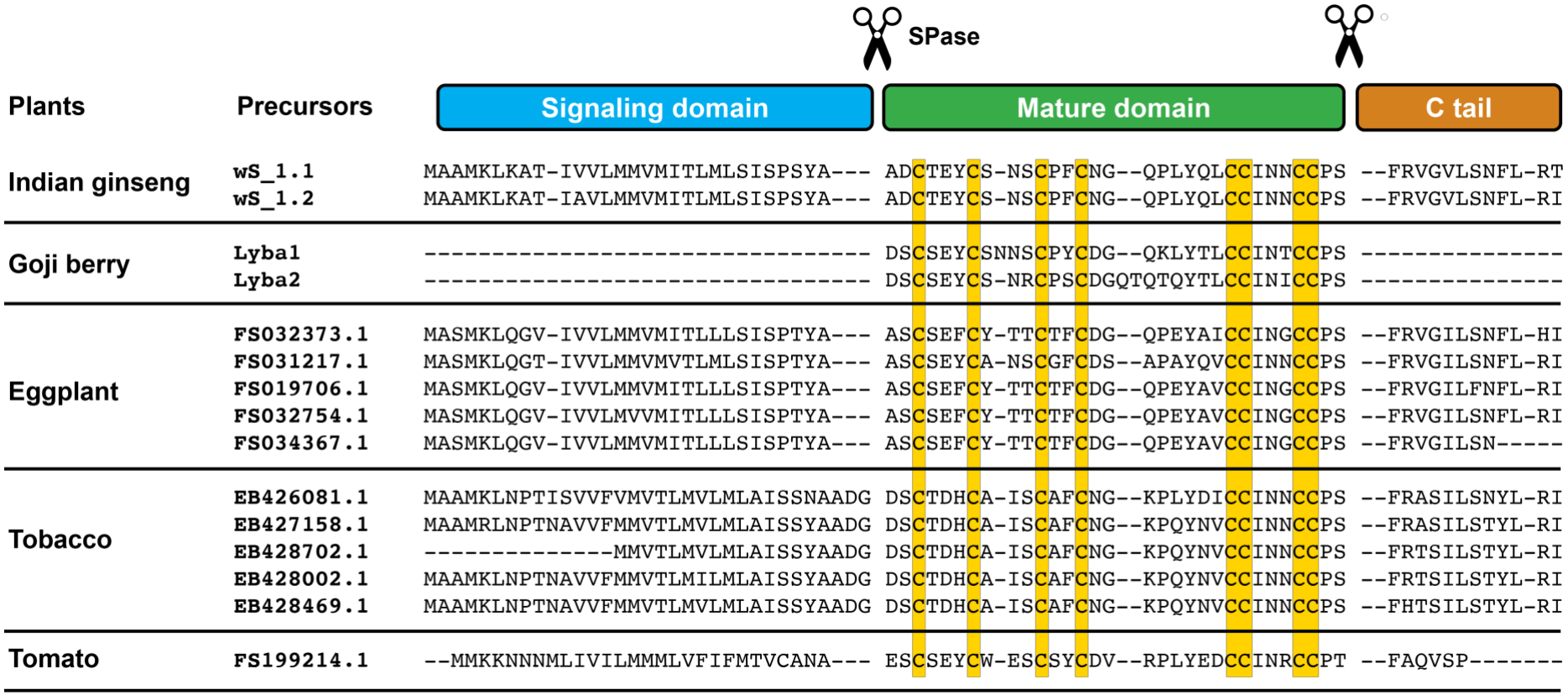


Figure 4.6 Precursor alignment of wS1 and its homologs.

Representative precursors, except for lyba 1 and lyba 2 with unknown precursors, from different plants are aligned. Cysteine residues are highlighted in yellow

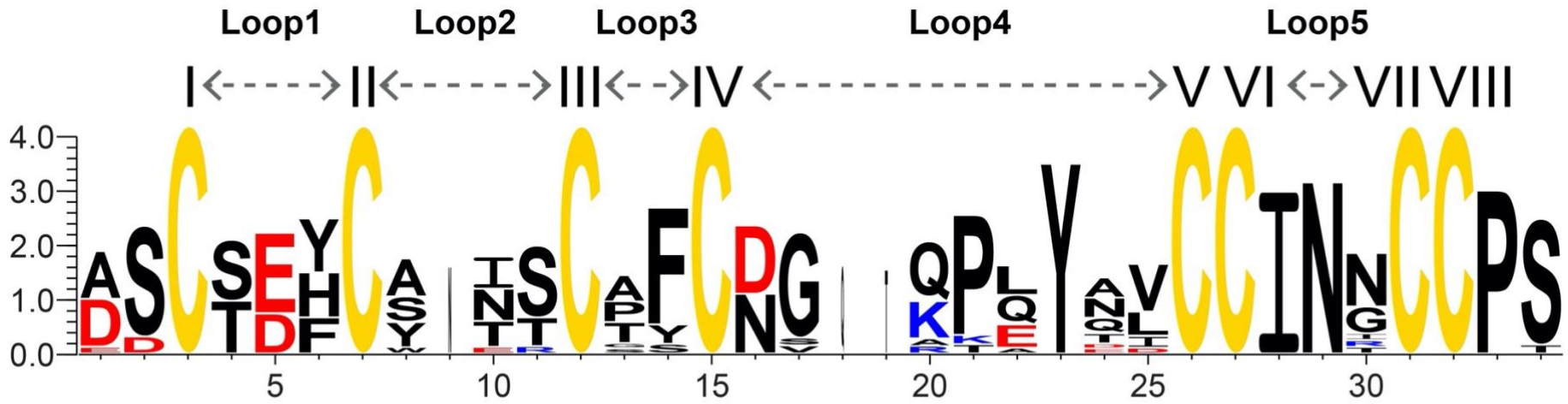


Figure 4.7 Mature domain conservation of wS1 and wS1-like peptides

Conservation of residues in the mature domain is shown. The height of each residue letter presents the residue conservation at that position. Residues are in different colors: negatively charged residues in red, positively charged residues in blue, cysteine residues in yellow, and others in black.

4.2.1.7 Comparison between wS1 and lybatides

To further compare wS1 with wS1-like peptides, lybatide, the only peptide characterized in this CRP family so far, was used as an example. By physiochemical analyses, we found that lybatides have ~30 residues in length, -2 for their overall charges, and 4.03 for their isoelectric point (pI). These parameters are almost the same as the parameters of wS1, which are 30 residues in length, -2 for its charges, and 3.67 for its pI (**Figure 4.8 A**). Besides, peptide backbone analysis showed that lyba 2 has the two common secondary structures (an α + π helix and a 3_{10} helix) and a specific C-C-C-C-CC-CC cysteine motif. Similarly, wS1 has two common secondary structures (an α helix and a β strand) with the C-C-C-C-CC-CC cysteine motif (**Figure 4.8 B**). Furthermore, we found that lyba 2 has negative charge residues, close to each other, on its structural surface (**Figure 4.8 C**). Two of these negative charges (D and E) are at the amino-terminal of lybatides (**Figure 4.8 A**). These properties of lyba 2 were also observed on wS1. Thus, wS1 and lybatides have high physiochemical and possible biological similarities.

A

Peptides	Characterized mature sequences	Amino acid residues	Charge	pI	Species/ Organisms	References
wS1	ADCTEYCS-NSCPFCNG--QPLYQLCCINNCCPS	31	-2	3.67	*Withania somnifera*	*This work*
Lyba1	DSCSEYCSNNSCPYCDG--QKLYTLCCINTCCPS	32	-2	4.03	*Lycium barbarum*	*Wei Liang Tan et al, 2017*
Lyba2	DSCSEYCS-NRCPSCDGQTQTQYTLCCINICCPS	33	-2	4.03	*Lycium barbarum*	*Wei Liang Tan et al, 2017*

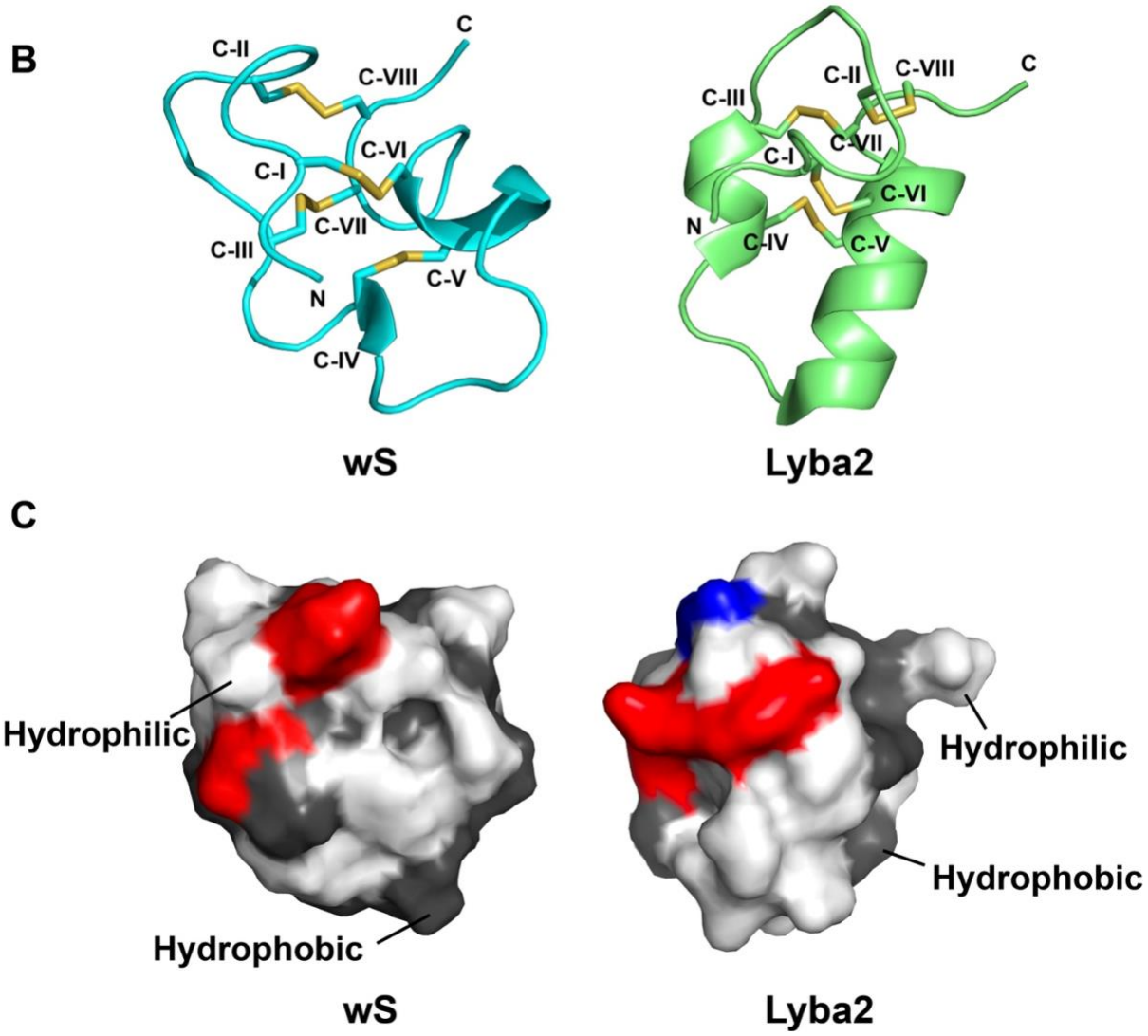


Figure 4.8 Comparison between wS1 and lybatides.

(A) Physiochemical properties of wisotide (wS1) and lybatides (lyba1 and lyba2). **(B-C)** Peptidyl backbones and surface properties of wS1 and lyba2. In the surface properties, negative charge residues are in red and positive residues are in blue.

Table 4.1 Structural statistics for the final 20 conformers of wS1.

Distance restraints	
Intra-residue (i-j = 0)	81
Sequential ($\|i$-$j\|$ = 1)	78
Medium range ($2 \leq \|i$-$j\| \leq 4$)	15
Long range ($\|i$-$j\| \geq 5$)	22
Hydrogen bond	12
Total	208
Average rmsd to the mean structure (Å)[b]	
Backbone atoms	0.78 ± 0.28
Heavy atoms	1.39 ± 0.35
ϕ/ψ space[c]	
Most favored region (%)	73.6
Additionally allowed region (%)	24.4
Generously allowed region (%)	1.6
Disallowed region (%)	0.4
rmsd from covalent geometry	
Bonds (Å)	0.003 ± 0.000
Angles (deg.)	0.485 ± 0.023
Impropers (deg.)	0.381 ± 0.027
rmsd from experimental restraints	
NOEs (Å)	0.036 ± 0.008

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

[b] Calculated with MOLMOL using range 3-16, 20-30.

[c] Calculated with PROCHECK-NMR.

Table 4.2 Proton chemical shift assignments for each residue of wS1.

	HN (ppm)	Hα (ppm)	Hβ (ppm)		Others (ppm)
A1		4.223	1.441		
D2	10.40	4.020	3.214		
C3	9.443	4.361	3.218	3.809	
T4	8.358	4.331	4.398		Mδ2, 1.319
E5	9.579	3.833	1.099	1.352	Hγ, 2.088, 2.320
Y6	7.288	4.690	2.409	3.242	Hδ, 6.988; Hε, 6.785
C7	8.358	5.244	2.825	3.510	
S8	8.958	4.389	3.914		
N9	8.281	4.700	2.605	2.682	
S10	8.040	4.709	3.736		
C11	8.319	4.612	2.130	2.416	
P12		4.467	1.994	2.368	Hγ, 1.841, 2.081; Hδ, 3.528, 3.708
F13	6.894	4.544	3.010	3.358	Hδ, 7.110; Hε, 7.263; Hζ, 7.344
C14	6.993	4.917	2.619	3.108	
N15	7.877	4.592	2.933	3.021	
G16	8.716	3.892, 4.034			Hγ,2.441, 2.655
Q17	7.609	4.538	1.978	2.215	Hγ, 2.392, 2.441; Hδ, 3.852, 3.886
P18		4.329	2.124	1.926	Mδ, 0.883, 0.922
L19	7.965	4.779	1.549	1.856	
Y20	7.461	3.940	2.873	3.445	Hγ, 1.423, 1.685
Q21	8.692	3.053	1.936		Mδ, 0.792, 0.863
L22	7.611	3.833	1.526	1.563	
C23	7.994	4.222	3.045	3.409	
C24	8.414	4.489	2.504	2.721	Hγ1, 1.189, 1.233; Hδ, 0.654; Mg2, 0.759
I25	8.368	3.948	1.784		
N26	8.181	4.841	2.575	2.838	
N27	8.298	5.088	3.00	3.272	
C28	8.044	5.143	2.814	4.344	
C29	8.594	5.389	3.138	3.354	Hγ, 1.907; Hδ, 2.957, 3.447
P30		4.821	2.017	2.352	
S31	8.387	4.482	3.902	3.964	

4.2.1.8 Structural stability

Medicinal plants usually involve decoction before oral administration. To test the stability of wS1, we used digestive enzymes to simulate oral administration. Our results showed that synthetic wS1 (with disulfide bonds) was highly stable after two hours in the enzymatic condition (**Figure 4.9 A-C**). In contrast, liner wS1 (*S*-alkylated wS1) was degraded under the same conditions. Therefore, disulfide bonds can render the structure of wS1 highly stable.

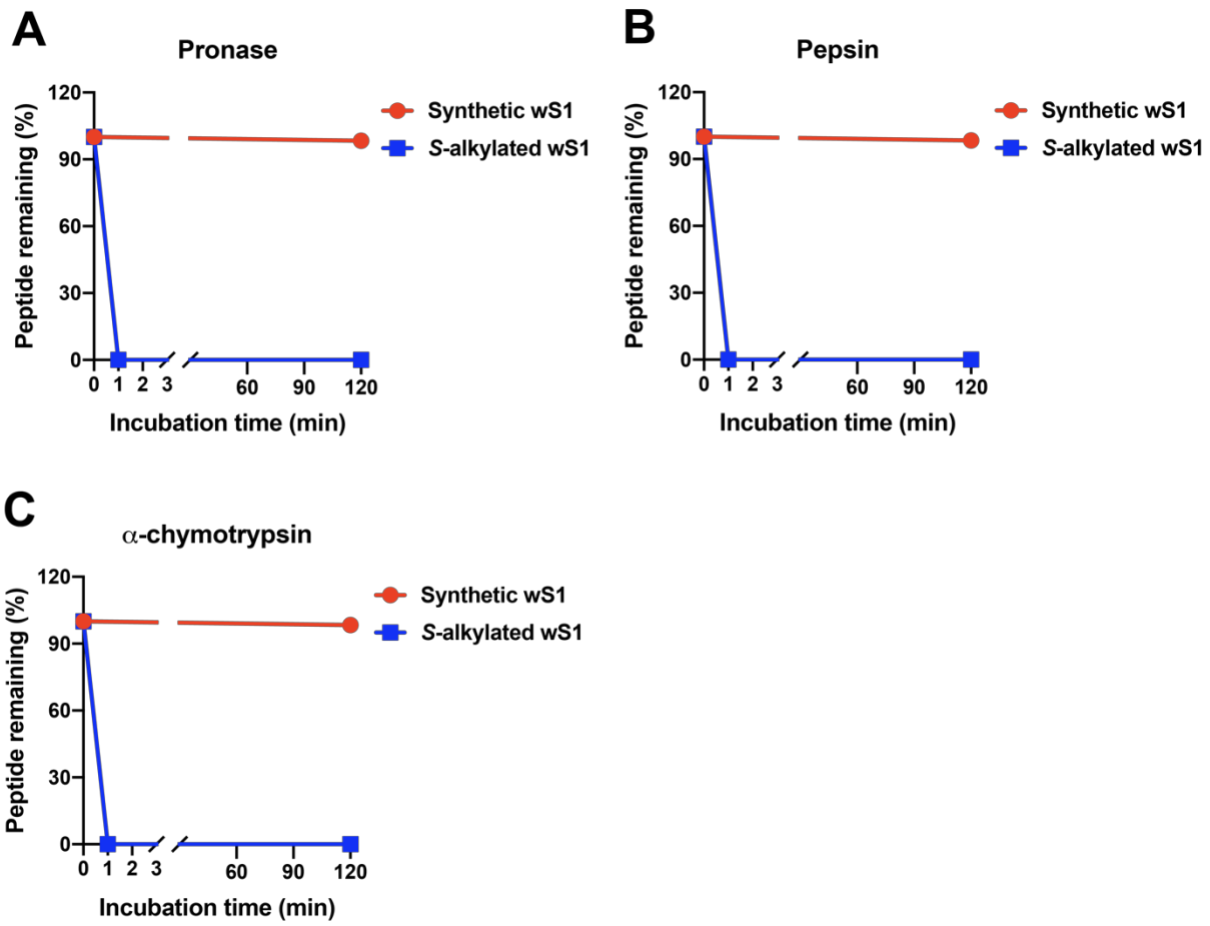


Figure 4.9 Structural stabilities of wS1.

Synthetic wS1 (with intact disulfide bonds) is 100-fold more stable than its S-alkylated form (without disulfide bonds) against proteolytic degradation. The pronase, chymotrypsin, and peptides were used at 20 μg/ml for these experiments.

4.2.2 Cell-membrane permeability of wS1

4.2.2.1 Non-cytotoxicity

To test the safety of wS1, we measure cell viability after treating cells with wS1. We did not detect any cytotoxicity after wS1 treatments (1-10 μM) for 24 h in HUVEC-CS and HepG2 cells, suggesting wS1 is non-cytotoxic up to 10 μM (**Figure 4.10**).

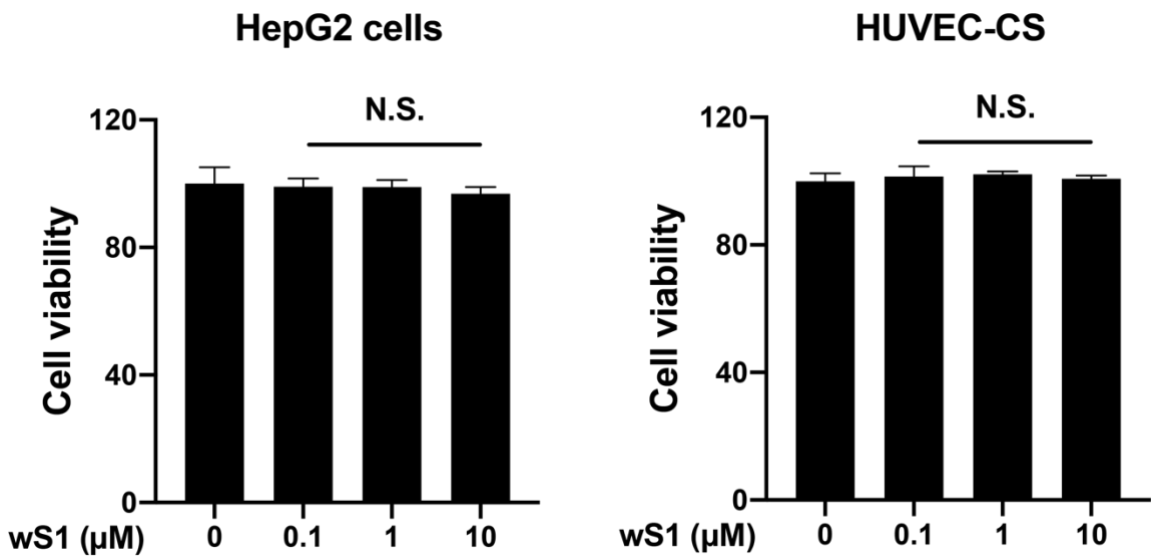


Figure 4.10 Non-cytotoxicity of wS1.

The cells were treated with wisotide (wS1) in different concentrations for 24 h. After wS1 treatment, the cell viability was measured by an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. The analysis was based on five replicates in each group. N.S. denotes no significance.

4.2.2.2 Chemical synthesis and fluorophore labeling

For biological investigation, we chemically synthesized, oxidatively folded, and fluorescently labeled wS1 (**Figure 4.11**). The linear wS1 synthesis was conducted by using the Fmoc SPPS (solid-phase peptide synthesis). Then, the peptides were cleaved by trifluoroacetic acid (92.5%) with 1,2-ethanedithiol (2.5%), triisopropylsilane (2.5%), and distilled water (2.5%). The synthetic linear wS1 was precipitated by diethyl ether and oxidatively folded in an aqueous solution at 4°C for 30 min. The solution contains cystamine/cysteamine (ratio: 1/10), DMSO (10%), and NH_4HCO_3 (100mM). In RP-HPLC co-elution, synthetic wS1 shares the same retention time with the native one (**Figure 4.12 A**). Also, they share high similarities in their ^1H2D NMR spectra (**Figure 4.12 B**), indicating their identity in three-dimensional structures. Furthermore, the synthetic wS1 (folded) was labeled with the AF488 at its amino-terminus and purified by RP-HPLC. The mass shift of 516.3 Da indicates the success of AF488 labeling on wS1 (**Figure 4.11 and Figure 4.12 C**). Thus, we show the synthetic wS1 (folded) is identical to the native one and that wS1 can be labeled with the AF488.

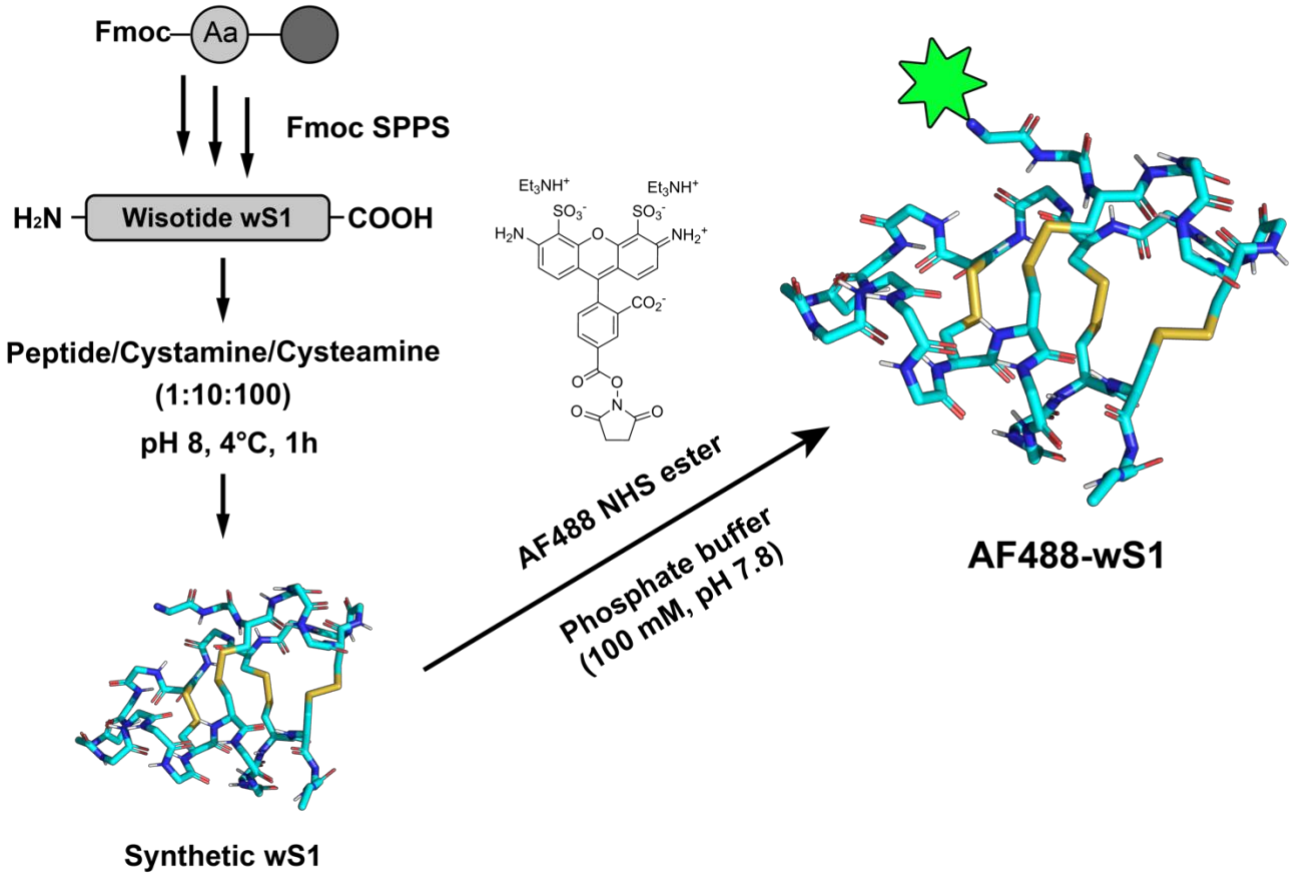


Figure 4.11 Schematic illustration for chemical synthesis, oxidative folding, and fluorescent labeling of wS1.

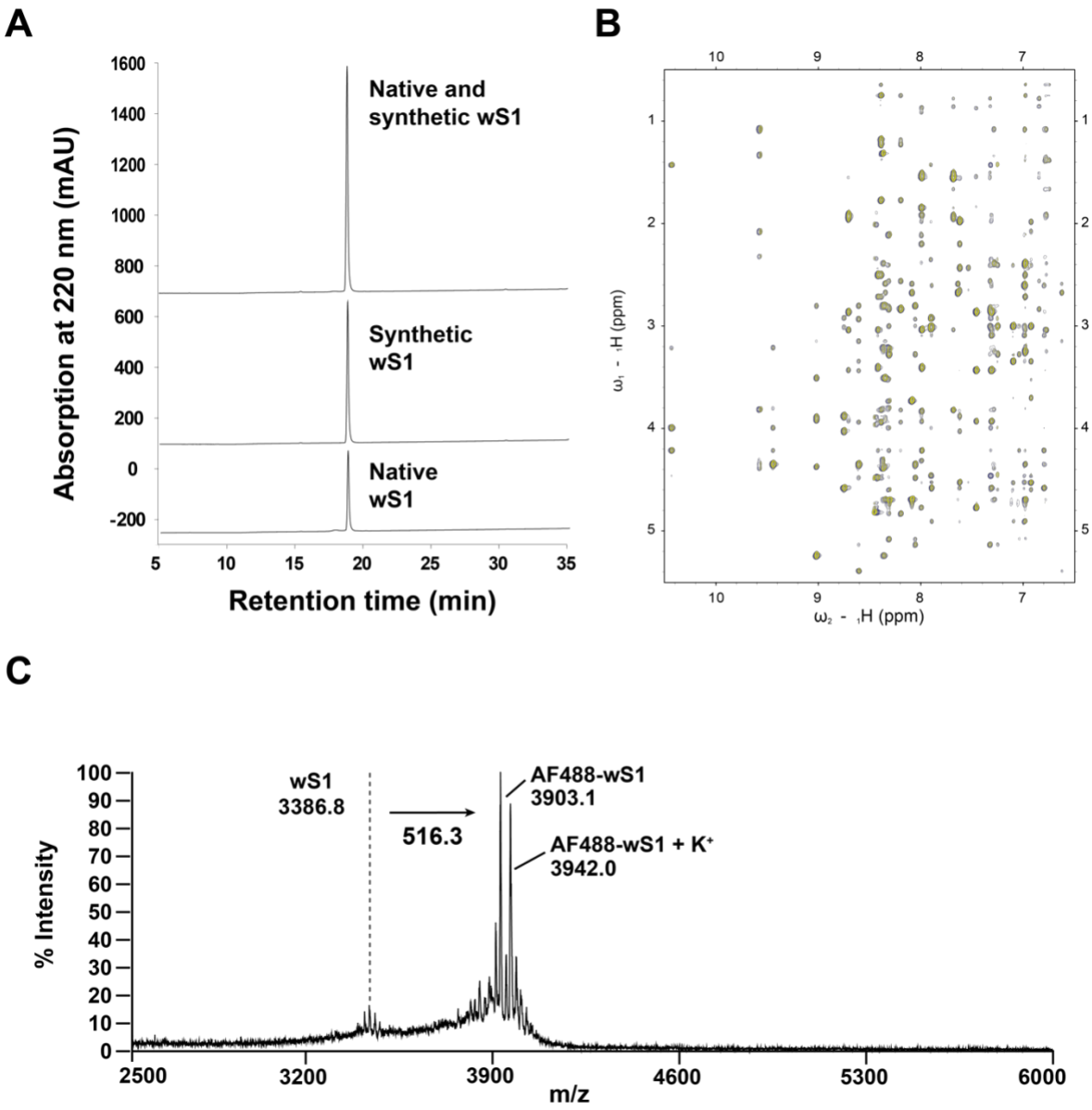


Figure 4.12 Quality control of synthetic wS1 and AF488-wS1.

(A) Co-elution of synthetic and native wS1 by using HPLC (high-performance liquid chromatography) is shown. **(B)** Overlapping 2D NMR spectra of the synthetic wS1 (yellow) and the native one (blue) are shown. **(C)** MALDI-TOF mass shift after labeling wS1 with the AF488 fluorophore. The NMR experiment for the spectra overlapping was conducted by Dr. Fan. J. S. and then analyzed by me.

4.2.2.3 Endocytosis-mediated cell entering

To determine cell entering capacities of wS1, we performed flow cytometric analysis and confocal imaging. After treating cells with AF488-wS1, we observed that cellular uptake of AF488-wS1 (at 37°C) increased in a time-dependent manner (**Figure 4.13 and Figure 4.17**). At low temperature (4°C), the cellular uptake decreased 38% compared with the uptake at 37°C (**Figure 4.14**), indicating the AF488-wS1 might cross the cell membrane in an energy-dependent manner. For endocytic pathway analysis, we found that EIPA (ethylisopropylamiloride) inhibited the cellular uptake in two different cell lines (**Figure 4.15 and Figure 4.16**), showing that AF488-wS1 can cross via macropinocytosis. We also found that dynasore (an inhibitor of clathrin-dependent endocytosis) exhibited inhibition in HUVEC-CS rather than in HepG2 cells, which might be due to the cell line variation. Unexpectedly, nystatin (an inhibitor of caveolae/lipid-mediated endocytosis) promoted the cellular uptake, suggesting that these endocytic pathways are not independent from each other, but might share common regulators.

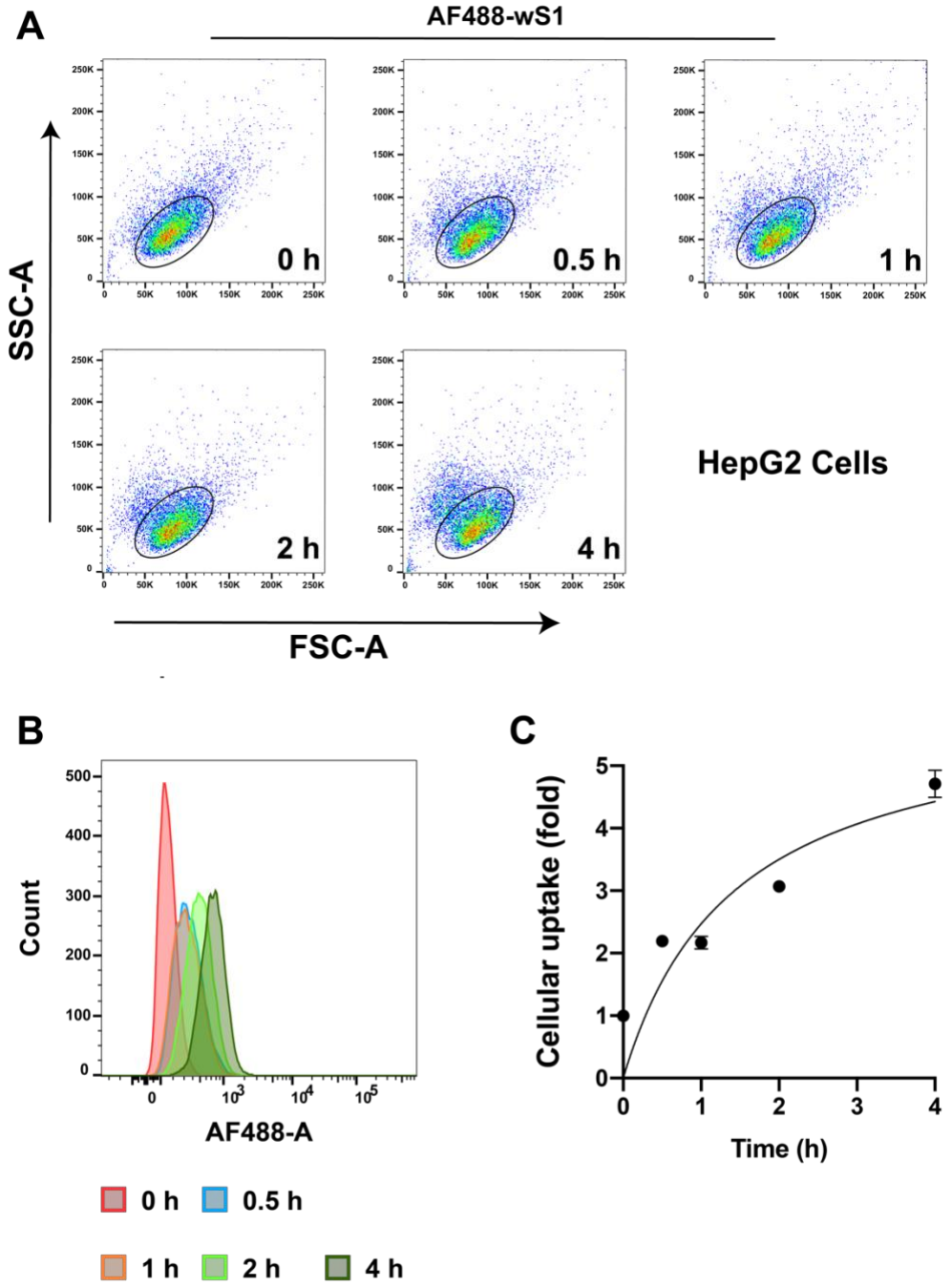


Figure 4.13 Flow cytometry analysis of AF488-wS1 cellular uptake with time.

HepG2 cells were treated with AF488 labeled wS1 (AF488-wS1) for different periods. **(A)** The cells were selected in elliptical circles for the fluorescent analysis. SSC-A indicates the side scatter area, and FSC-A stands for the forward scatter area. **(B)** The fluorescent peaks shifted forward after AF488-wS1 treatment. **(C)** Plotting cure of AF488-wS1 cellular uptake (n=3).

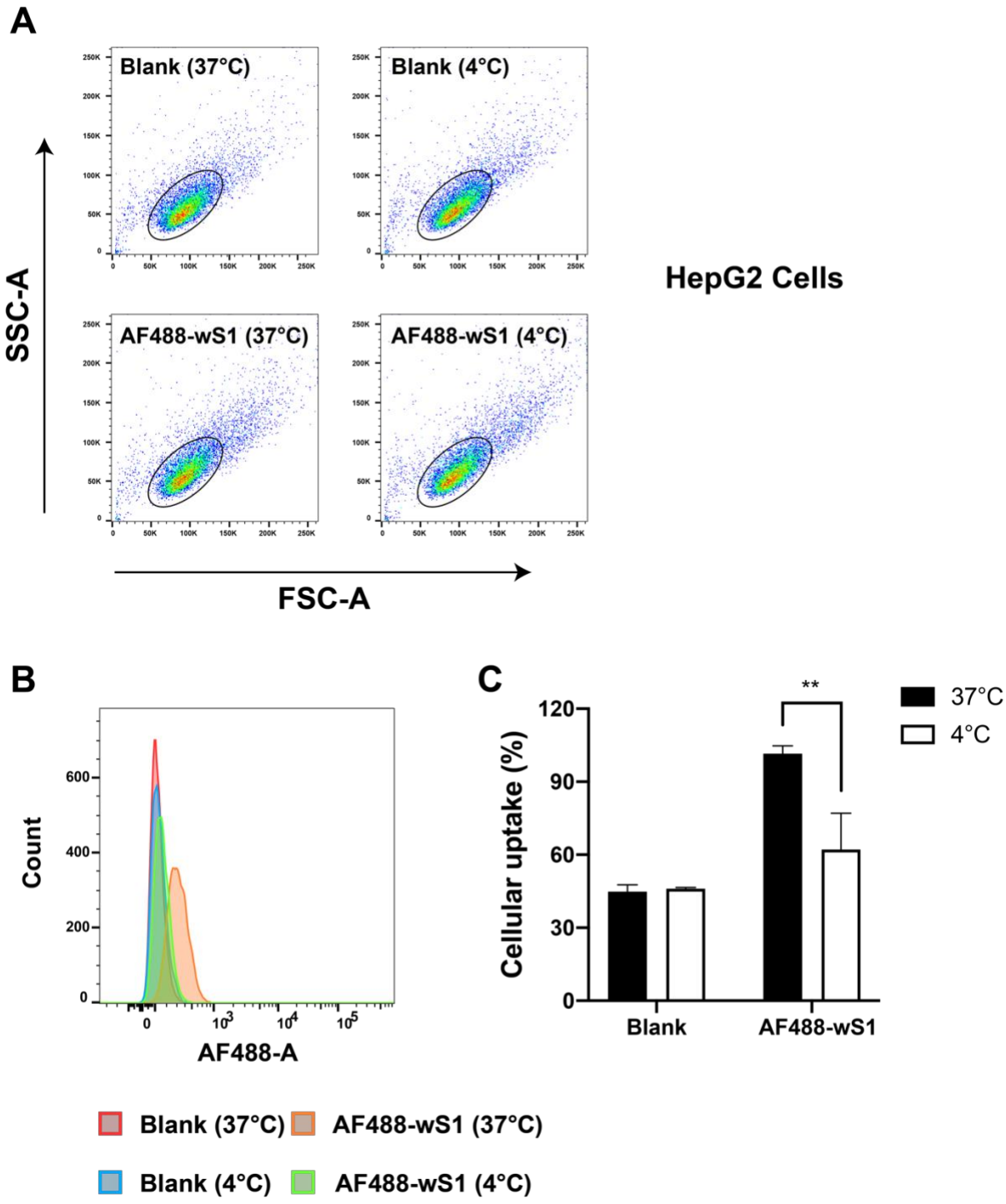


Figure 4.14 Temperature-dependent cellular uptake of AF488-wS1.

(A) HepG2 cells under different temperatures were treated with the AF488-wS for 1 h. By using SSC-A and FSC-A, the cells in elliptical circles were selected for the fluorescent analysis. **(B)** The fluorescent peaks shifted more at 37°C than at 4°C. **(C)** Quantification of (A) and (B) with statistical analysis (n=3).

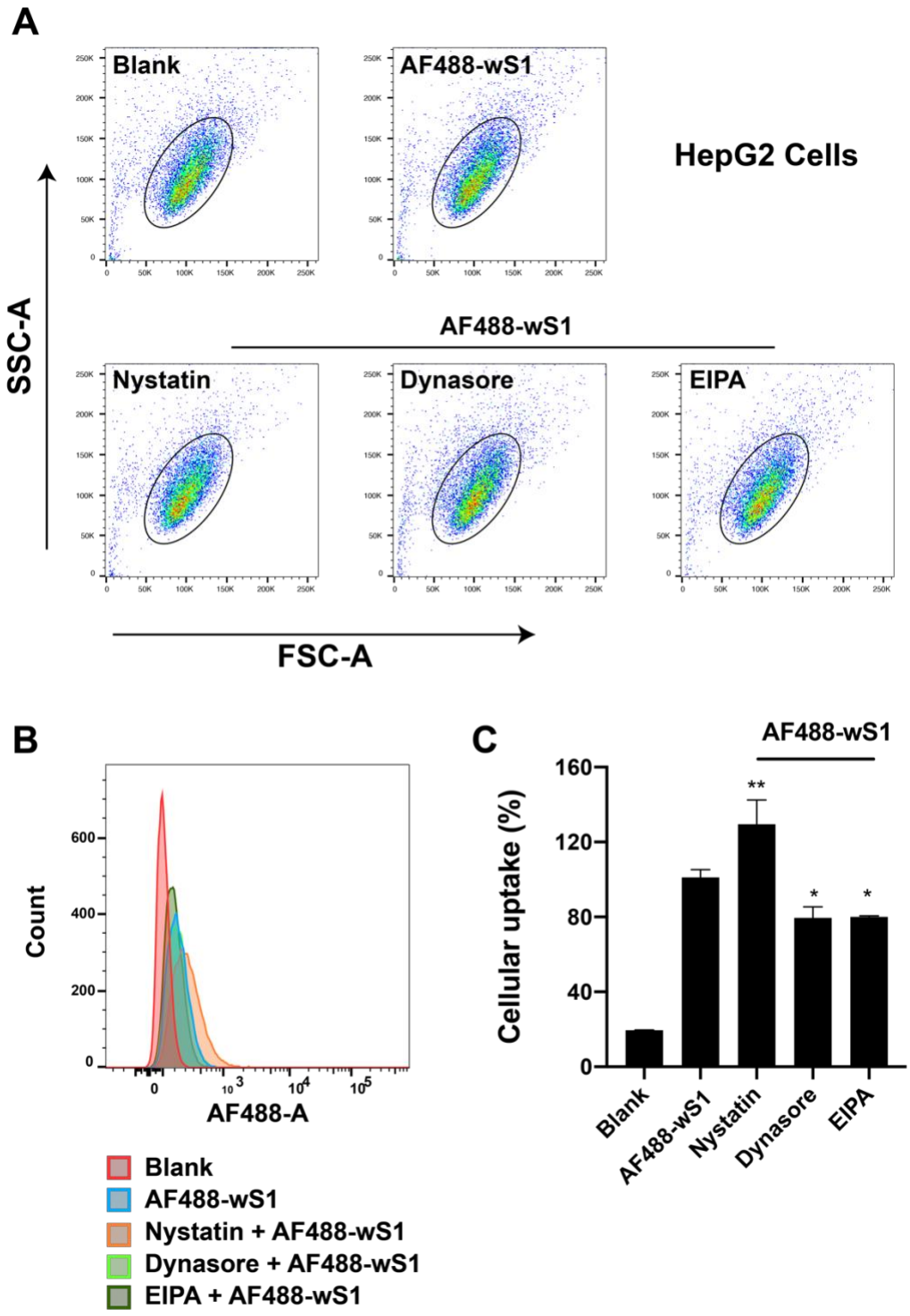


Figure 4.15 Cell entering by endocytic pathways in HepG2 cells.

(A) The HepG2 cells were treated with exocytotic inhibitors (30 min) and co-treated with AF488-wS1 and inhibitors (1 h) before being selected by ellipses. **(B)** The cells in ellipses were analyzed by peak shifting extend. **(C)** Quantification of (A) and (B) with statistical analyses (n=3).

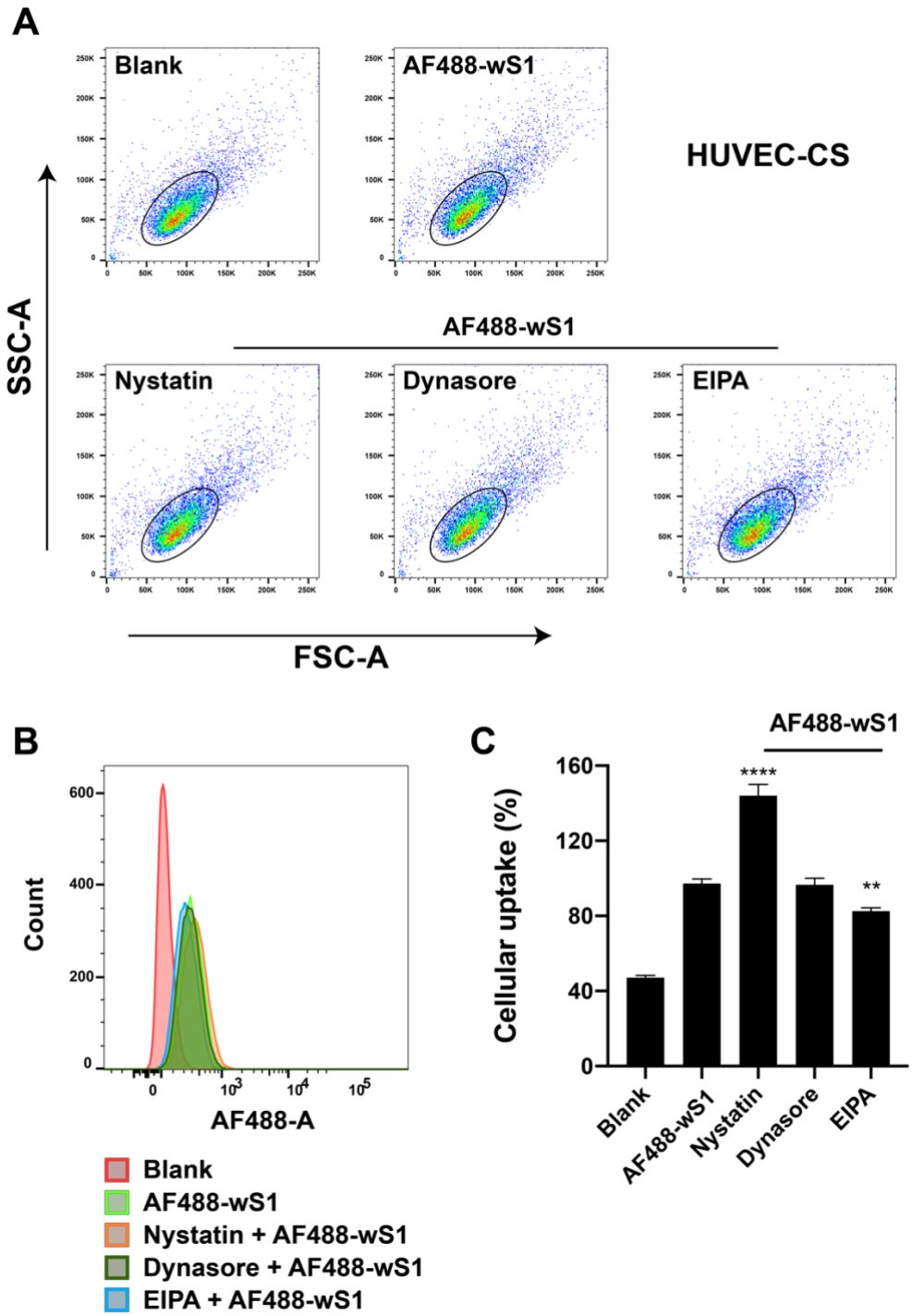


Figure 4.16 Cell entering by endocytic pathways in HUVEC-CS.

(A) The HUVEC-CS were pre-treated with endocytic inhibitors (30 min pre-treatment and 1 h co-treatment). The cells were selected by elliptical circles. **(B)** Fluorescent intensity difference between cell groups was presented by fluorescent shifting extent between peaks. **(C)** Quantification of (A) and (B) with statistical analyses (n=3).

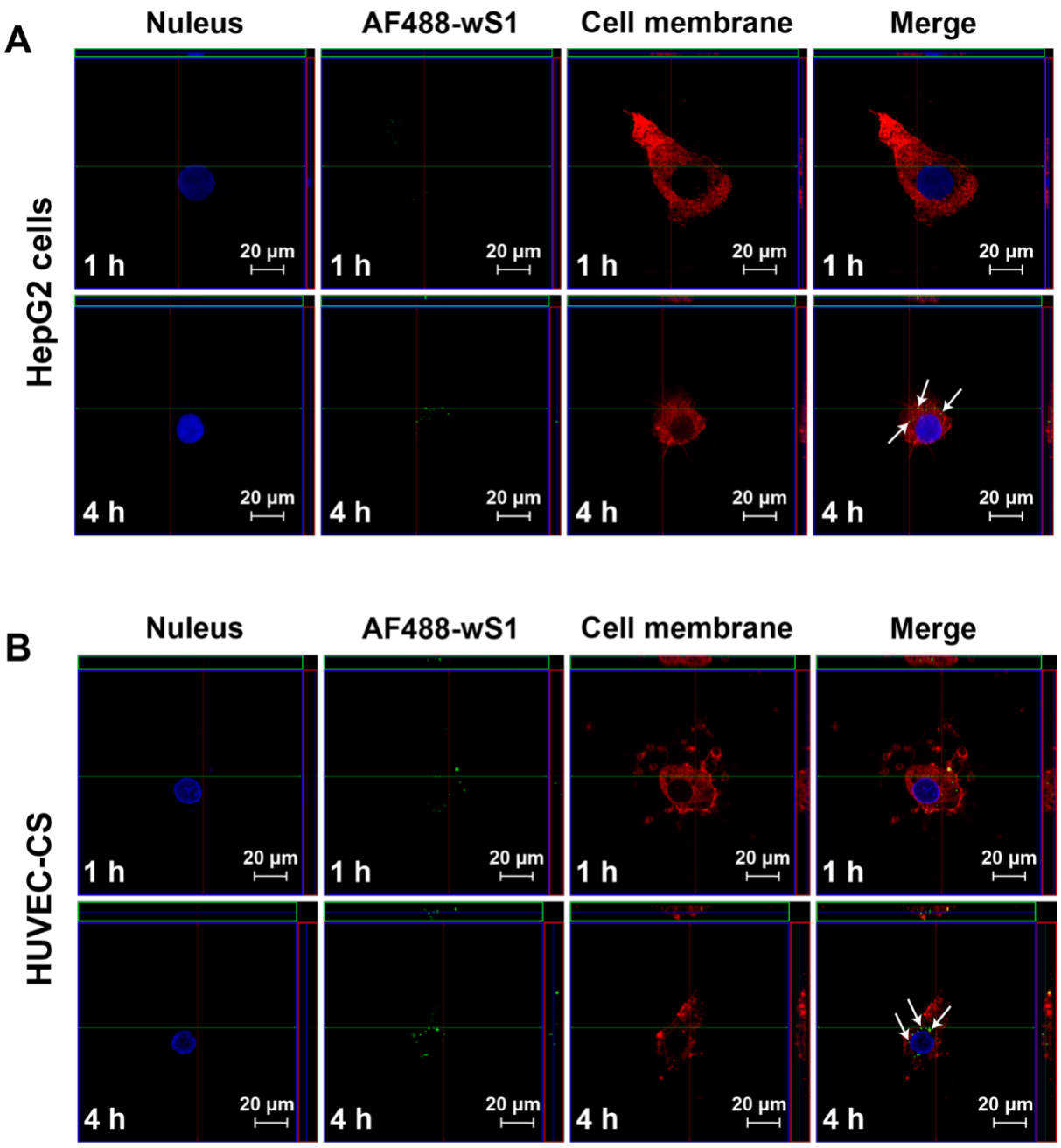


Figure 4.17 Confocal micrographs of AF488-wS1 cell entering.

(A and B) Live cells were treated with the AF488-wS1 (1μM) for 1 h and 4 h before confocal microscopic analysis. The nuclei were stained with Hoechst 33342 (1 μg/ml) for 15 min, and the plasma membrane was stained with PKH26 red-fluorescent dye (2 μM) for 5 min. White arrows indicate the internalized AF488-wS1.

4.3 Discussion

In this study, we discovered and characterized an unusual cystine-stapled helical peptide, wisotide (wS1), from the root of *Withania somnifera* or Indian ginseng. The wS1 has 31 residues in length and a cystine-stapled helical scaffold, rendering it hyperstable against proteolytic degradations. Also, the wS1 with hydrophobic patches clustered by its helical scaffold exhibited cell-membrane permeability, an important property for peptides targeting the intracellular environment.

Sequence analysis showed that wS1 has a cysteine pattern of C-C-C-C-CC-CC, with two consecutive cysteine residues (-CC-) at its C terminus. In plant CRPs, based on -CC- pairs in their cysteine motif, there are at least three known classes: class I with no -CC- pairs, class II with one -CC- pair, and class III with two -CC- pairs. Class I without the -CC- pair can be found in 4C/6C/8C CRPs, such as 4C α-hairpinin, 6C potentides, and 8C defensin (e.g., NaD1). Class II has -CC- pair located in different sites in the cysteine pattern, and this class can be found in 6C/8C/10C CRP families, such as 6C thionin (e.g., crambin) with the -CC- pair at the N terminus, 6C hevein like peptides with a pair in the middle, and jasmintides with a pair at the C terminus. Class III contains two pairs of -CC- and only two characterized members so far, including the 6C β-ginkgotide and the 8C lybatide. In this study, we present wS1 as a new member of the lybatide family, for it shares same cysteine pattern with lybatides. This pattern determines the loop length, a key determinant of intercysteinyl sequence variation and peptidyl structural conformations.

NMR spectrometry indicated that 8C wS1 has disulfide connectivity of Cys I-Cys VI, Cys II-Cys VIII, Cys III-Cys VII and Cys IV-Cys V. The disulfide connectivity in 8C CRPs can be treated as an extension of the one in the 6C CRPs, which contain at least five types. Type I is the well-known cystine knot (Cys I-Cys IV, Cys II-Cys V, and

Cys III-Cys VI), like the one in knottins, defensins, and hevein-like peptides. Type II is the symmetric connectivity (Cys I-Cys VI, Cys II-Cys V, and Cys III-Cys IV), like the one in thionins and α-hairpinins. Other types are hybrid and represented by the connectivity in jasmintides (type III), potentides (type IV), and β-ginkgotides (type V). Like lybatides, wS1 may be treated as an 8C CRP derived the 6C CRP in type II, because after removing Cys I-Cys VI, the connectivity of wS1 is symmetric. Disulfide connectivity together with the cysteine pattern result in excellent scaffold diversity of CRPs.

Structural analysis revealed that wS1 has a cystine-stapled helical scaffold, which in CRPs can be seen in two structural motifs: CSαα and CSαβ. The CSαα motif is characterized by two α helices bridged by disulfide bonds in between them. This motif can be found in α-hairpinins and thionins, although these two do not share the exact same motif: α-hairpinins have only two α-helices stabilizing one another, whereas thionins not only have two stabilized α-helices but also two anti-parallel β-strands. The CSαβ is characterized by an α-helix linked to the central β-strand of an anti-parallel β-sheet by disulfide bonds, such as in defensins. wS1 has two helices bridged by one disulfide bond and thus exhibits a CSαα motif. The inside disulfide bond of the wS1 motif might force hydrophobic side chains to be exposed outside, partially leading to cell-membrane permeability of hydrophilic and anionic wS1.

Biosynthetic analysis showed that wS1 is derived from two precursors with one residue variation, and that these precursors share three structural domains: a signal domain, a mature domain, and a C tail. The signal domain is cleaved by the signal peptidase at the ER, and then the mature domain is subsequently cleaved by the serine endopeptidase. Bioinformatic analysis revealed that 57 wisotide-like peptides are distributed in three plant families, and their precursors all have the three architectural

domains. Weblogo analysis indicated that these wisotide-like peptides have the serine highly conserved at the C terminus, indicating the importance of the serine endopeptidase in peptide maturation.

In conclusion, wS1 is an anionic cystine-stapled peptide from the root of *Withania somnifera* with cell-membrane permeability activity, which is partially attributed to hydrophobic patches on the peptide molecular surface. Moreover, 57 putative wS1-like peptides were identified from five different plants. Thus, wS1 represents an excellent class of cystine-stapled helical templates for designing membrane-permeable peptides.

CHAPTER FIVE

Discovery of a plant-derived hyperdisulfide peptide that induces LDLR-mediated LDL uptake at lipid rafts

5.1 Introduction

Among natural products, peptides are distinguished in size from small molecules and large proteins, but often possess their inherent therapeutic advantages, representing a novel class of compounds for drug discovery [281]. Natural occurring peptides can be found in fungi, animal venoms, human hormones, and plant tissues. These peptides, like macrocycles [282], conotoxins [283], chlorotoxin [284], and insulin [285], exhibit exceptionally high structural diversity and have evolutionally fine-tuned biological functions, critical for organisms to defend, prey, and signal. Over 80 peptides from nature so far have succussed as clinical drugs, like captopril for hypertension and exenatides for Type 2 diabetes [286]. Nowadays, natural occurring peptides, cysteine-rich peptides (CRPs) in particular, are a highly underexplored frontier of drug discovery. CRPs in nature are 15-50 residue miniproteins, braced by 2-5 disulfide bonds, displaying different cysteine frameworks [35, 75, 99, 143, 277-280]. CRPs occurring in plants can function as antimicrobial peptides or protease inhibitors [35, 287]; in animal, like marine snails, spiders, and scorpions, they can be a venom component [286]. Unlike small molecules, these natural miniproteins are evolutionally optimized to display a large molecular interface for protein-protein interactions, representing CRPs an ideal class of therapeutic leads with less off-target possibilities (or side effects). For example, ziconotide from the venom of *Conus magus* has been developed into a potent neuropathic painkiller in clinic [288]. Although promising, no plant-derived CRPs are marketed so far.

Hyperdisulfide CRPs are a subfamily of CRPs containing at least 30% cysteine content [289]. As cystine percentage increases and peptide size decreases, structural scaffold becomes increasingly constrained to render peptides hyperstable against thermal, chemical, and proteolytic degradation. Hyperdisulfide CRPs derived from nature can

be exemplified at least by three peptides: 1) the linaclotide with ~42% cystine in 14 residues [290]; 2) the conotoxin SIIIA with 30% cystine in 20 residues [291]; 3) the β-ginkgotides with at least 30% cystine in 18-20 residues [289]. The high cystine percentage in small peptide size indicates that hyperdisulfide CRPs likely exhibit small-molecule stability and are good starting points for orally bioactive development or drug design.

Schisandra chinensis, from Schisandraceae family, is a woody vine plant growing at high-latitude areas in Asia. Fresh fruits of *S. chinensis* ripen once a year, and they are often consumed directly with hot water or dried for long-term storage like tea. In modern pharmacopeias, these fruits are positioned as “adaptogens”, a term describing substances with none-specific anti-stress effects or multiple therapeutic functions [197, 198]. One of their pharmacological applications is to treat hyperlipidemia in blood circulation [292]. Currently, bioactive constituents in these fruits are represented by dibenzocyclooctadiene lignans, like schisandrin B, schisantherin A, and Gomisin A [196]. However, bioactive peptides in this medicine are unknown.

Here, we report the identification and characterization of a novel hyperdisulfide CRP, designated wuweizitide 1 (wZ1), from *S. chinensis*. We showed that wZ1 contains six cystine residues out of 18 total residues, processed from an 86-residue precursor with three structural domains. We suggested that wZ1 with cysteine-knot like scaffold is hyperstable against proteolytic degradation. We revealed that wZ1 crosses the cell membrane by endocytosis and increases LDL uptake by promoting LDLR distribution into lipid rafts. Thus, this study enriches our knowledge of hyperdisulfide CRPs and indicates their underexplored potential as therapeutic leads.

5.2 Results

5.2.1 Discovery and structural characterization of wuweizitide (wZ1)

5.2.1.1 Identification of wuweizitide (wZ1)

To discover novel cysteine-rich peptides, we profiled berries of *Schisandra chinensis* by MALDI-TOF mass spectrometry (MS), reduced berry extraction with dithiothreitol, and then alkylated with an iodoacetamide. By the MS profiling, we found two mass peaks at 2154.1 *m/z* and 2178.4 *m/z*, designated as wuweizitide (wZ1) and its sodium-ion binding form (wZ1+Na^+), respectively (**Figure 5.1 A**). By the reduction, we observed a mass shift of 6 *m/z* between 2154.4 *m/z* and 2160.9 *m/z*, indicating the possibility of occurring a novel peptide with three disulfide bonds, which are formed by six cysteine residues (**Figure 5.1 B**). By the alkylation, we detected a peak at 2502.7 *m/z*, with a mass increase of 348.3 *m/z*, compared with the wZ1 peak at 2154.4 *m/z*. This mass increase confirmed the occurrence of six cysteine residues, as one cysteinyl alkylation leads to a mass increase of 58 *m/z* (**Figure 5.1 C**). These results suggest that a novel cysteine-rich peptide with three disulfide bonds exists in the *S. schisandra* berries. The wZ1 was then extracted and purified in a large scale from *S. chinensis* seeds, and the yield of wZ1 is ~2 mg out of 100 g seeds.

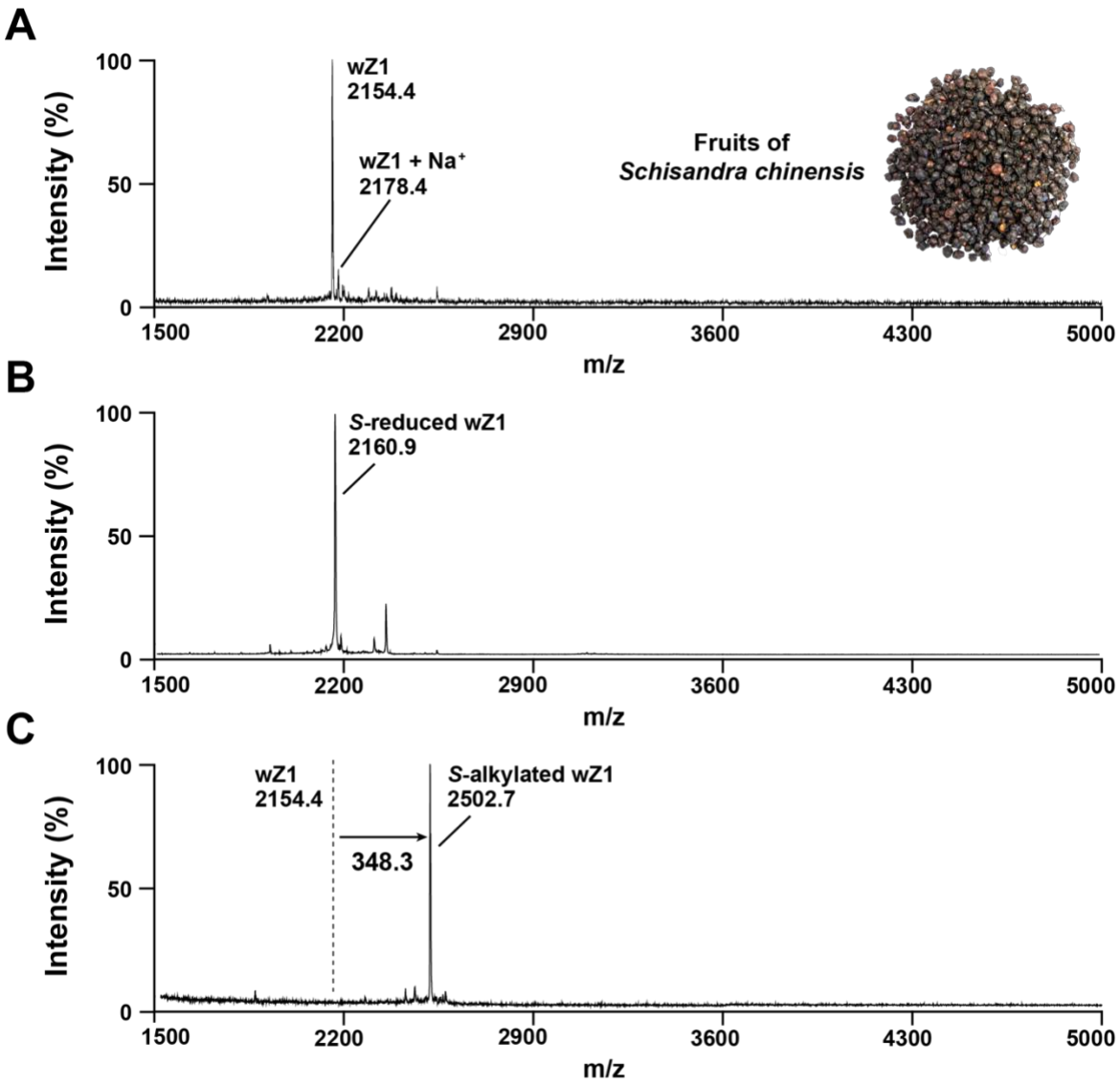


Figure 5.1 Mass spectrometry profile of wuweizitide (wZ1).

(A) MALDI-TOF MS profile of wZ1 with *m/z* value 2154.4 isolated from the fruit of *Schisandra chinensis*. To confirm that wZ1 is a six-cysteine peptide, we performed *S*-reduction and *S*-alkylation. **(B)** MS profile of the *S*-reduced wZ1 with *m/z* value 2160.9 after reduction with 25 mM dithiothreitol. **(C)** MS profile of the *S*-alkylated wZ1 with *m/z* value 2502.7 with 100 mM iodoacetamide in 25 mM ammonium bicarbonate buffer (pH 7.8). The increase in *m/z* value 348.3 confirms that wZ1 has six cysteine residues. The wZ1 was identified by Dr. Wong K. H. initially.

5.2.1.2 Primary amino acid sequence determination of wZ1

To determine the primary sequence of wZ1, we *de novo* sequenced wZ1 by using MALDI-TOF/TOF in an MS/MS mode and further confirmed its sequence by transcriptomic analysis. *De novo* sequencing showed a putative wZ1 sequence: XXCCRCXSWPYMCSVFCC (**Figure 5.2**). In this sequence, X represents isobaric residues (I/L and K/Q), which can have four possible combinations (IK, IQ, LK, and LQ). To confirm these isobaric residues, we extracted the total RNA of the live *S. chinensis* berries, sequenced the transcriptome of the berries by using the next-generation sequencing, and then conducted a *de novo* assembly of the transcriptome. By transcriptomic analysis, we found that wZ1 precursor has three architectural domains, in which the mature domain shows the primary sequence of wZ1 as IQCCRCQSWPYMCSVFCC (**Figure 5.3**). However, when aligning wZ1 sequence to the transcriptome of *S. chinensis* by using the blast (basic local alignment search tool), we did not find its homologous mRNA sequences with the CC-C-C-CC motif in *S. chinensis*. Interestingly, we found another cysteine-rich peptide (β-wuweizitide, wwz2) showing a C-CC-C-CC motif from the transcriptome of *S. chinensis* (**Figure 5.3**). These results show that two different wuweizitides exist in *S. chinensis*. WZ1 in high yield was used for structural characterization and functional investigation.

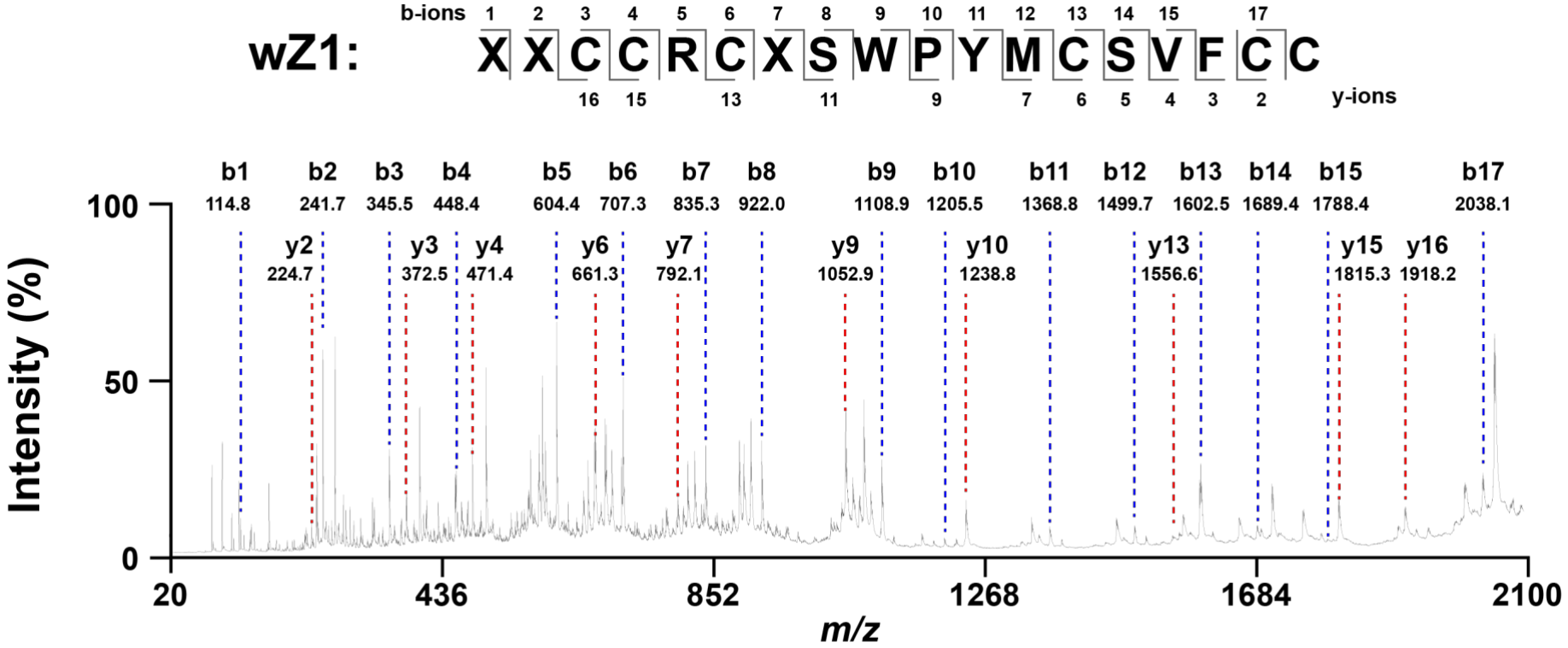


Figure 5.2 De novo sequencing of wZ1.

The primary amino acid sequence was deduced by using b-/y- ions, generated by MALDI-TOF/TOF in an MS/MS mode.

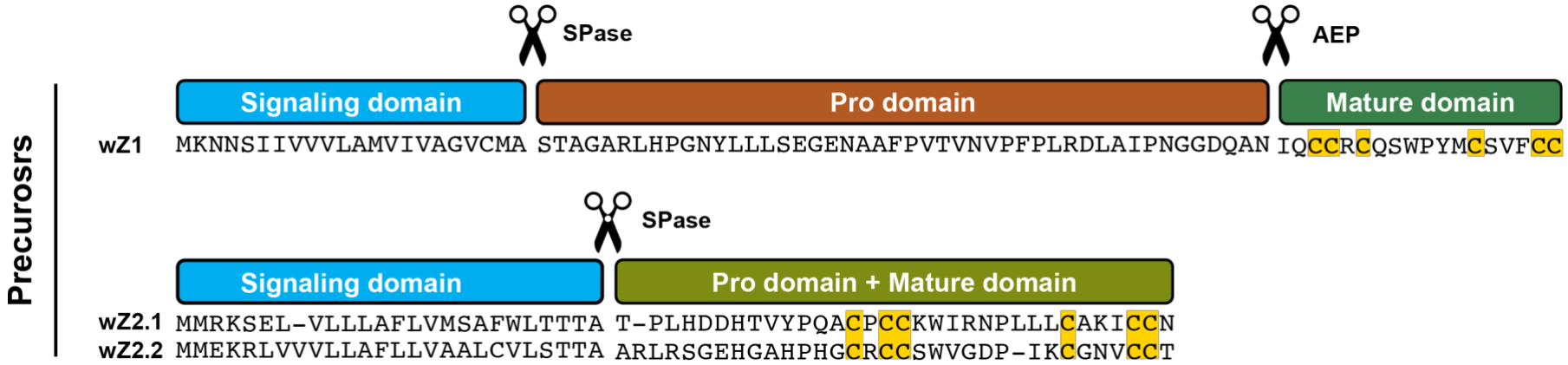


Figure 5.3 Biosynthesis of wZ1 and wuweizitide2 (wZ2).

Transcriptomic analysis confirmed the sequence of wZ1 and indicated the occurrence of wuweizitide2 (wZ2.1 and wZ2.2). The precursor sequence of wuweizitides consists of three architectural domains, including signaling, pro, and mature domains.

5.2.1.3 Chemical synthesis of wZ1

To prepare the sample for later structural and functional investigation, we chemically synthesized wZ1 and compared its structure with the native one. In detail, linear Fmoc-wZ1 with protection was obtained by solid-phase peptide synthesis (**Figure 5.4**). The Fmoc-wZ1 was then deprotected and cleaved by trifluoroacetic acid to release the linear wZ1, which was immediately folded oxidatively under redox conditions (cystamine/cysteamine) with 10% DMSO in the 100 mM NH_4HCO_3 buffer (pH 8) at 4ºC for 30 min. The folded wZ1 was purified by RP-HPLC to reach >98% purity, and the purified synthetic wZ1 was then co-eluted with native wZ1 (**Figure 5.5 A**). The co-elution result shows the same retention time for synthetic and native forms of wZ1, indicating that they share an identical hydrophobic structure. The structural identity of synthetic/native wZ1 was further confirmed by 2D NOESY spectra (**Figure 5.5 B**), in which most of the cross-peaks are overlayed with each other, while some of the peaks are not, for the existence of a few metabolites in the native wZ1 sample. These results show that the synthetic wZ1 is identical to the native one in structure.

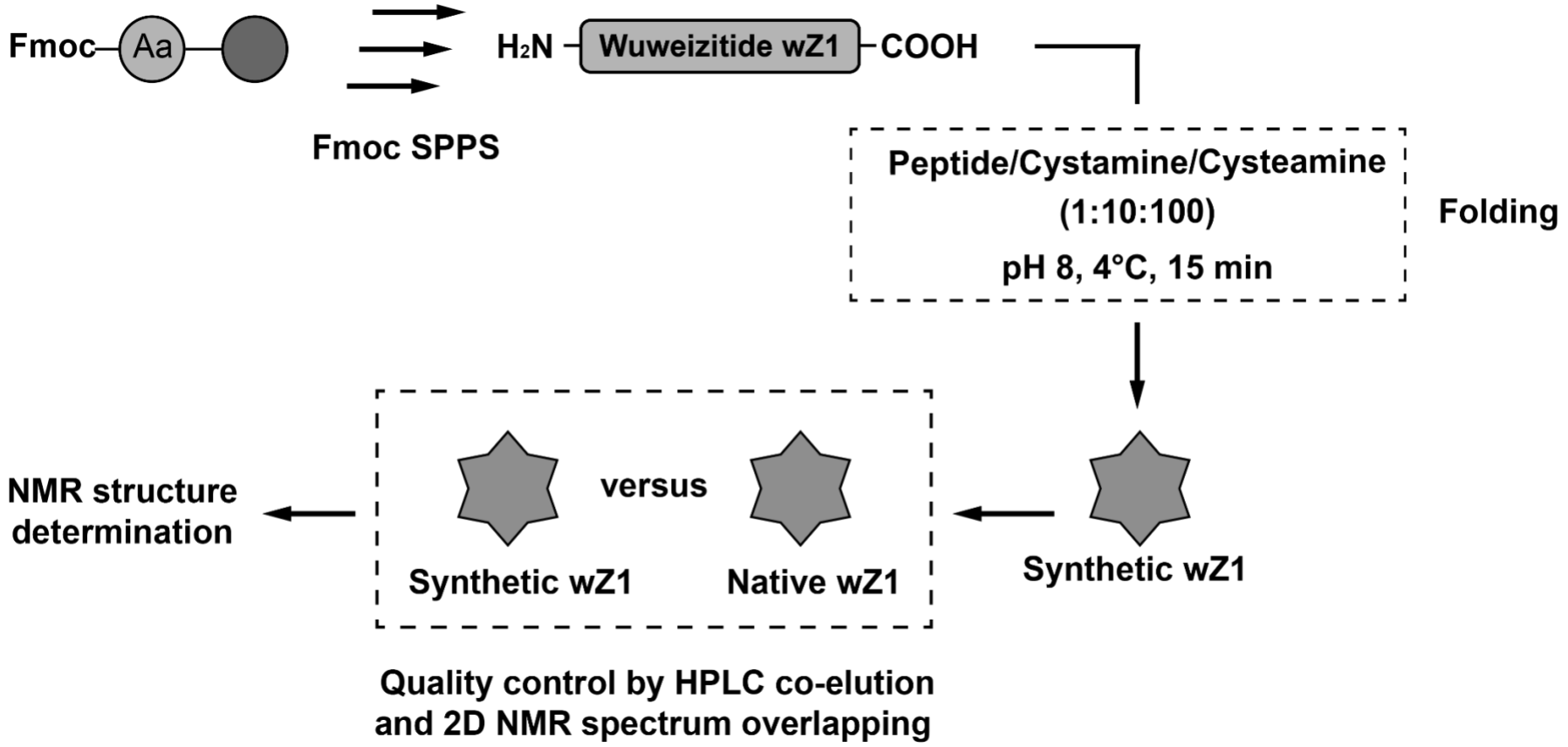


Figure 5.4 Schematic illustration of wZ1 chemical synthesis for NMR structure determination.

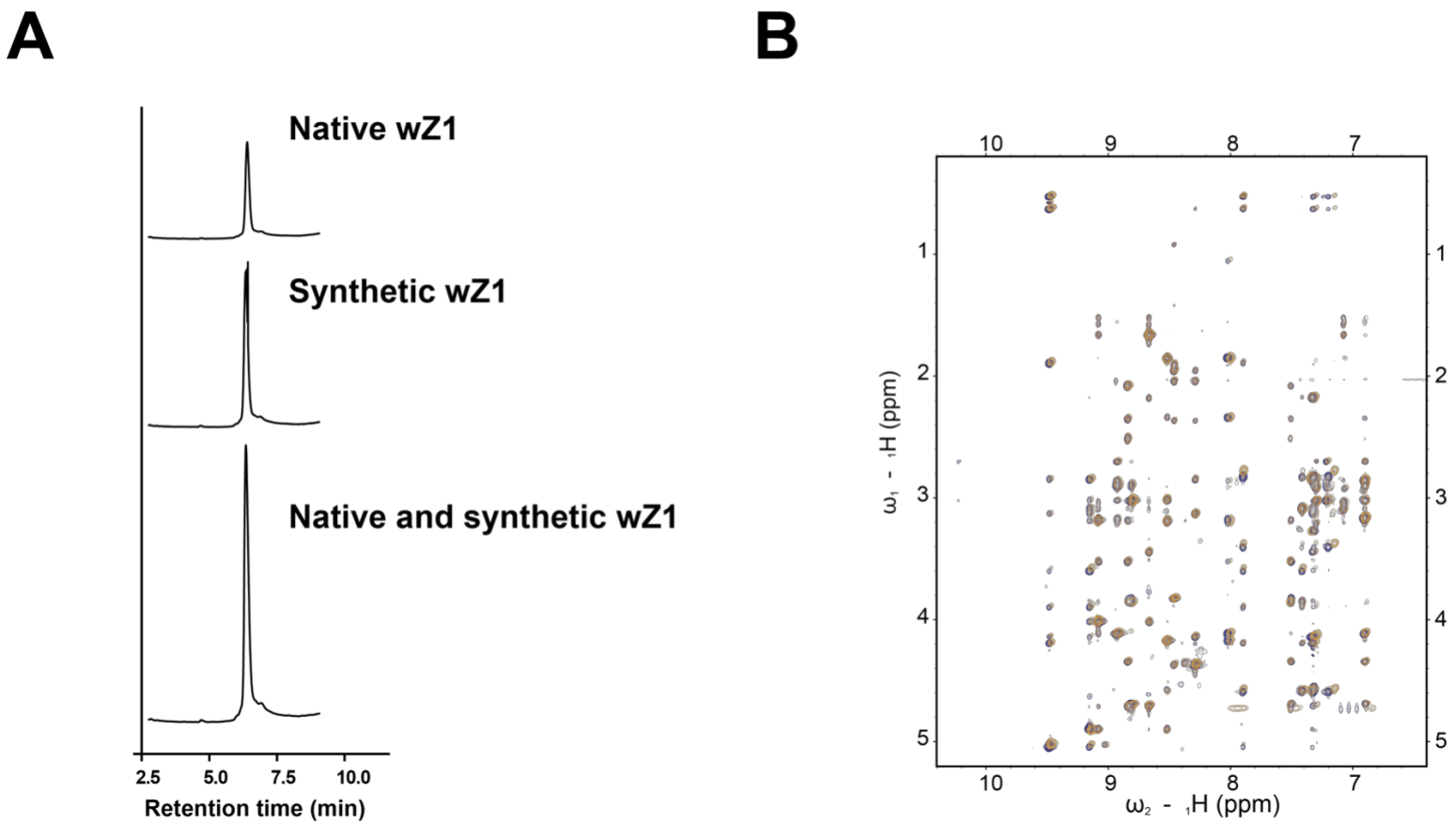


Figure 5.5 Quality control of synthetic wZ1.

(A) RP-HPLC co-elution of the native and the synthetic wZ1. **(B)** NOESY 2D-NMR overlay spectra of the native (purple) and the synthetic (yellow) wZ1. The NMR experiment was conducted by Dr. Fan J. S. and then analyzed by me.

5.2.1.4 NMR (nuclear magnetic resonance) structure of wZ1

To characterize the structural properties of wZ1, we performed NMR structural determination of synthetic wZ1. Our result showed that wZ1 has knottin-type disulfide connectivity, C-I to C-IV, C-II to C-V, and C-III to C-VI (**Figure 5.6 A and B**). Besides, wZ1 has one positive charge (blue) and overall hydrophobic property (dark grey) on its structure surface (**Figure 5.6 C**). The NMR result suggests that wZ1 might be a hyperstable cysteine-rich peptide with cell-permeable capacities.

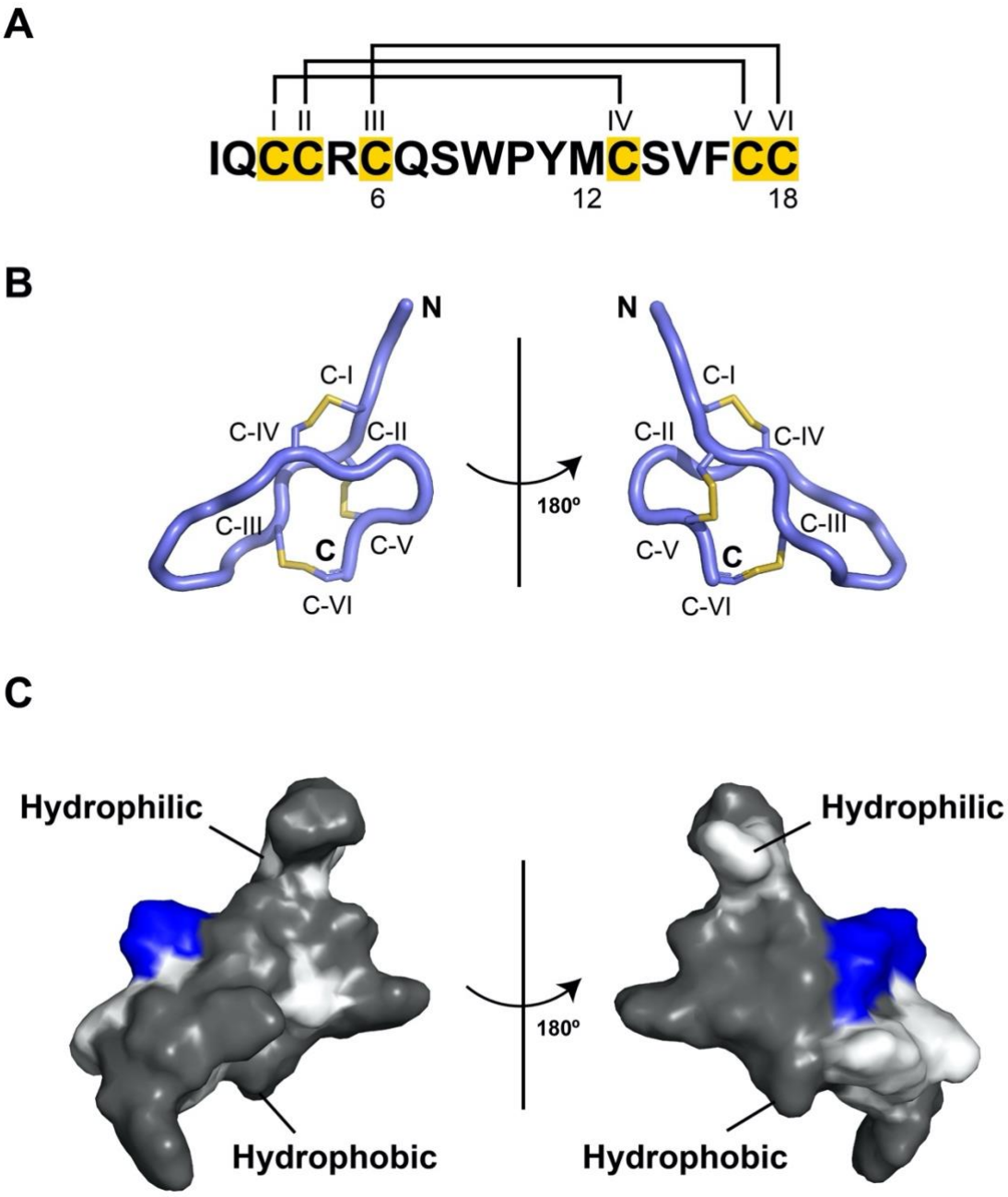


Figure 5.6 Disulfide connectivity and NMR structure of the synthetic wZ1.

(A) Primary sequence and disulfide connectivity of wZ1. WZ1 has 18 amino acids in length with disulfide connectivity of Cys I-IV, Cys II-V, and Cys III-VI. **(B)** The backbone of wZ1 with three disulfide bonds (yellow). The structural surface of wZ1 with positive residue (R) in blue, hydrophobic residues (I, W, P, M, V, and F) in dark grey, and hydrophilic residues (Q, S, and Y) in light grey. This NMR experiment was conducted by Dr. Fan J. S. and then analyzed by me.

Table 5.1 Structural statistics for the final 20 conformers of wZ1.

Distance restraints	
Intra-residue (i-j = 0)	55
Sequential ($\|i$-$j\|$ = 1)	58
Medium range ($2 \leq \|i$-$j\| \leq 4$)	12
Long range ($\|i$-$j\| \geq 5$)	18
Hydrogen bond	0
Total	143
Average rmsd to the mean structure (Å)[b]	
Backbone atoms	0.83 ± 0.26
Heavy atoms	1.64 ± 0.39
ϕ/ψ space[c]	
Most favored region (%)	78.7
Additionally allowed region (%)	17.7
Generously allowed region (%)	3.0
Disallowed region (%)	0.7
rmsd from covalent geometry	
Bonds (Å)	0.003 ± 0.000
Angles (deg.)	0.577 ± 0.047
Impropers (deg.)	0.367 ± 0.046
rmsd from experimental restraints	
NOEs (Å)	0.076 ± 0.011

[a] Selected from 100 calculated conformers according to overall energy.
[b] Calculated with MOLMOL using range 3-7, 12-18.
[c] Calculated with PROCHECK-NMR.

Table 5.2 Proton chemical shift assignments for each amino acid residues of peptide wZ1.

	HN (ppm)	Hα (ppm)	Hβ (ppm)		Others (ppm)
I1		3.820	1.899		Mδ1,0.846; Mγ2, 0.916; Hγ1, 1.127,1.419;
Q2	8.480	4.363	1.951	2.038	Hγ, 2.364
C3	8.309	4.128	2.840	3.127	
C4	7.329	4.700	2.172	3.438	
R5	8.690	4.009	1.660		Hγ,1.520,1.571; Hδ, 3.035,3.089; Hε, 7.096
C6	9.106	4.711	3.187	3.517	
Q7	8.859	4.338	2.080	2.511	Hγ, 2.348
S8	7.535	4.691	3.851		
W9	8.820		2.893	3.019	Hδ1, 7.225; Hε1, 10.230; Hε3, 7.316 Hη2, 7.199; Hζ2, 7.467; Hζ3, 7.079
P10		3.210	1.047	1.285	Hγ, 1.530, -0.178; Hδ, 2.697, 2.920
Y11	8.942	4.108	2.850	3.163	
M12	8.023	4.170	1.854		Hγ, 2.337; Mδ, 1.042
C13	8.542	4.906	3.010	3.186	
S14	9.161	5.023	3.575	3.883	
V15	9.490	4.184	1.883		Mγ, 0.518, 0.618
F16	7.917	4.563	2.779	3.374	
C17	7.431	4.550	3.087	3.848	
C18	7.362	4.571	2.834	3.273	

5.2.1.5 Stability of wZ1 with knottin-type disulfide connectivity.

To determine the stability of three-disulfide constrained wZ1, we built *in vitro* models for gastrointestinal environments using different digestive enzymes. We found that wZ1 is hyper-stable against proteolytic degradation (**Figure 5.7 A-D**), suggesting its possibility of being developed as an orally active compound.

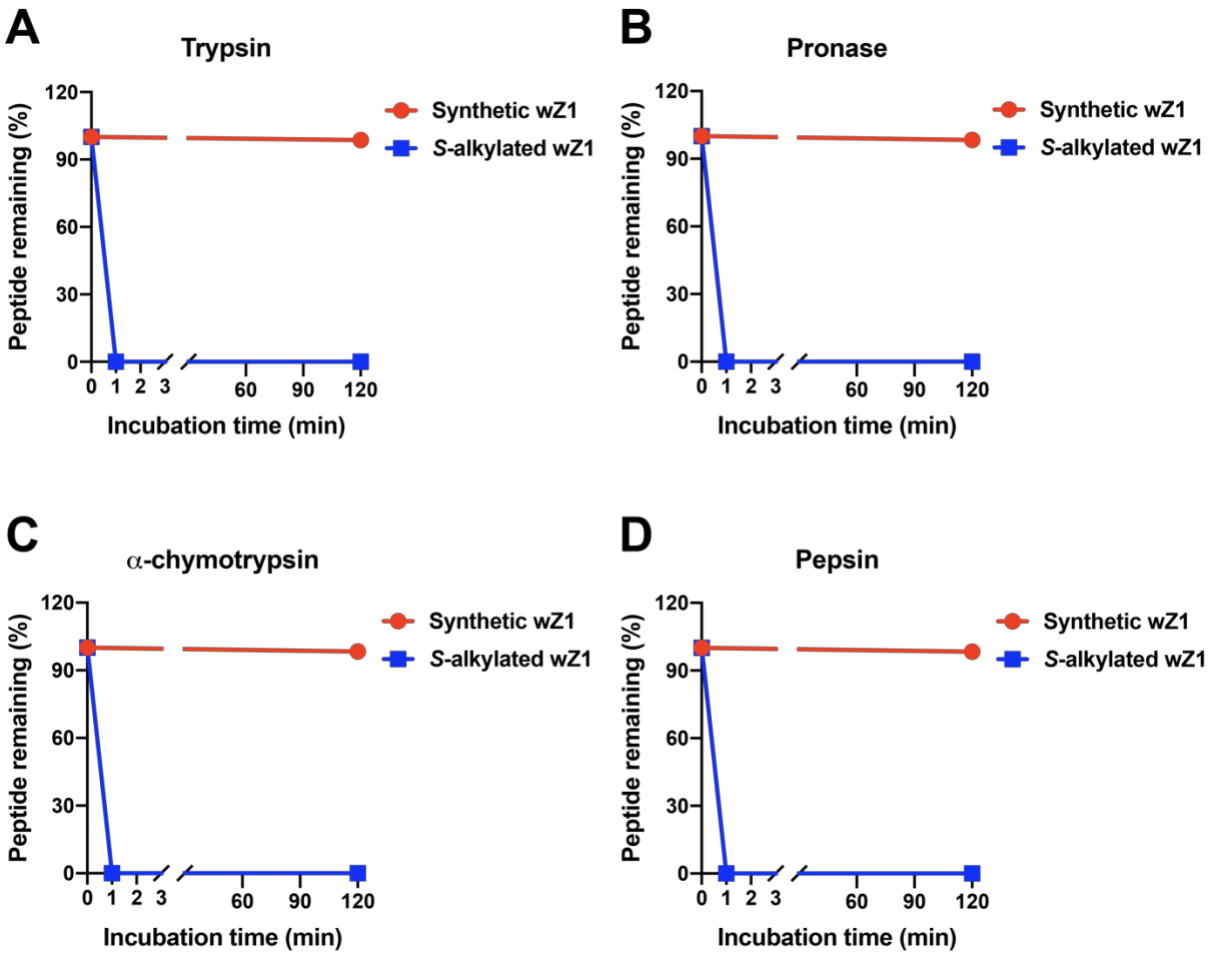


Figure 5.7 wZ1 is highly stable against proteolytic degradation.

Synthetic wZ1 and *S*-alkylated wZ1 were incubated with **(A)** trypsin (20 µg/ml), **(B)** pronase (20 µg/ml), **(C)** α-chymotrypsin (20 µg/ml), and **(D)** pepsin (20 µg/ml) at 37°C. RP-HPLC was used to quantitate the percentage of peptides remaining using an area under the curve analysis.

5.2.1.6 Phylogenetic tree of wZ1 and its conotoxin homologs

To characterize the phylogenetic relationship, regional distribution, and biosynthesis of wZ1 and its homologs, we searched the 1 KP database and NCBI databases with the primary amino acid sequence of wZ1. For the phylogenetic relationship and the regional distribution, we tried to find out its homologs *in planta* by using BLAST (Basic Local Alignment Search Tool). The results showed that no homologs were found in 1 KP database and NCBI databases, as the cysteine motif, CC-C-C-CC, is unusual *in planta*. Unlike *in planta*, we successfully found over 300 mRNA precursors in conotoxins with the same cysteine motif; the top 100 precursors with the highest similarity were then used for phylogenetic analysis. The results showed that wZ1 is distantly related to its conotoxin homologs, as wZ1 does not form any cluster with other homologs in the phylogenetic tree, and that wZ1 is geographically far from its conotoxin homologs (**Figure 5.8**).

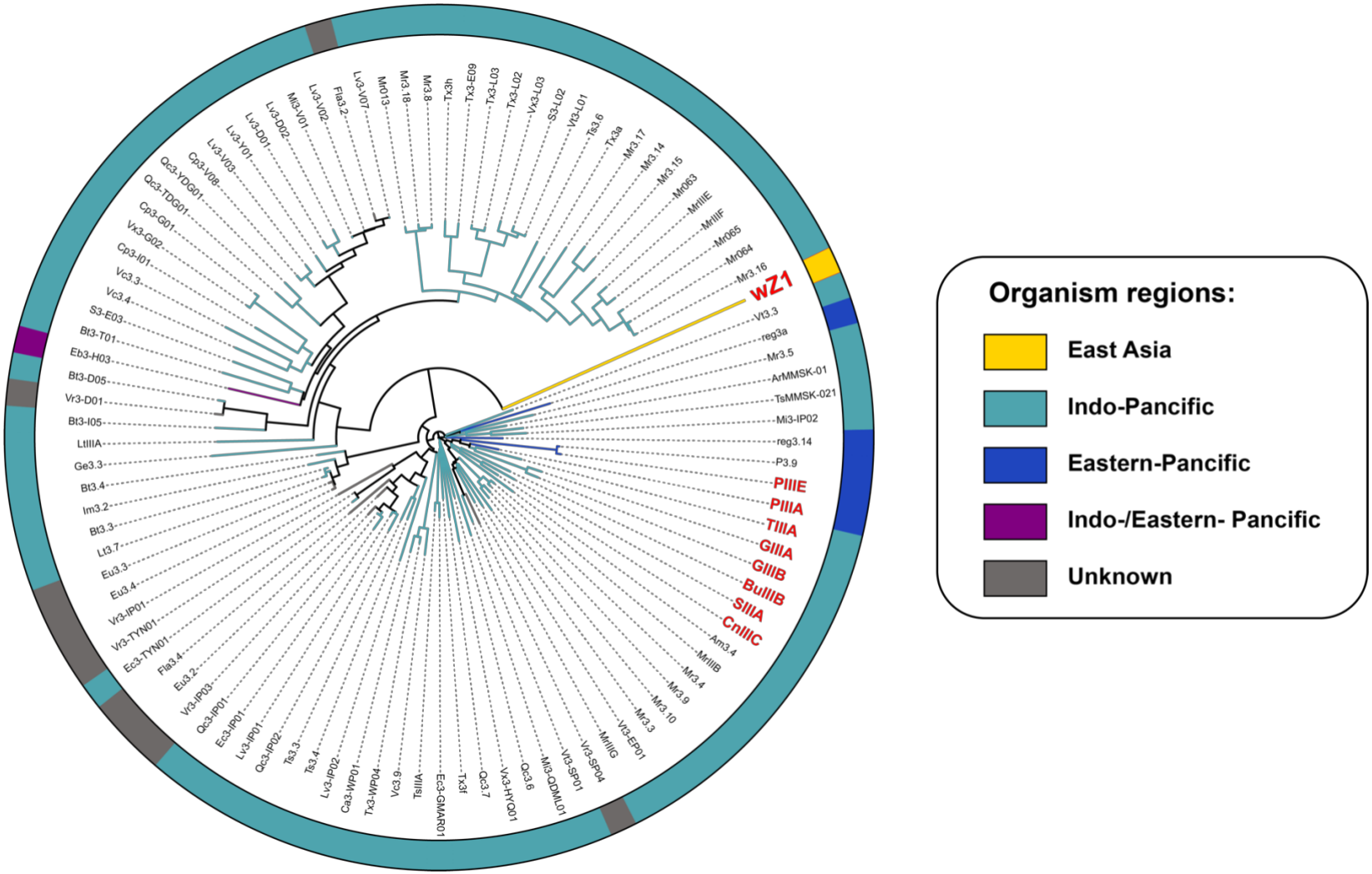


Figure 5.8 Phylogenetic tree of wZ1 and its homologs.

5.2.1.7 Precursors of wZ1 and ion-channel targeting homologs

For comparing the biosynthesis of wZ1 and its homologs, we aligned the precursor sequences of wZ1 and the ion-channel targeting peptides (highlighted in red in the phylogenetic tree). The result showed that wZ1 precursor has similar architecture to these peptide homologs. They contain three same domains: 1) signaling domain; 2) pro domain; 3) mature domain. In contrast to their similarities, wZ1 does not have a small C tail as its homologs (**Figure 5.9**). These results suggest that wZ1 is an unusual cysteine-rich peptide *in planta*, and that wZ1 is geographically and evolutionarily distant to wZ1 conotoxin homologs, but its homologs share a similar biosynthetic architecture with it.

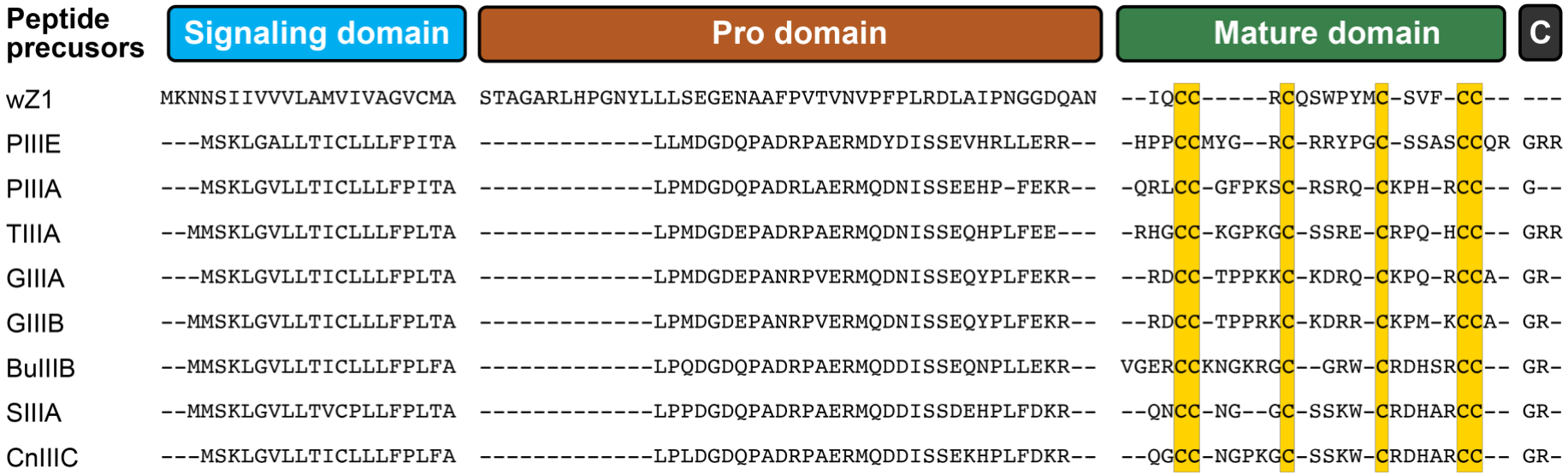


Figure 5.9 Sequence alignment of wZ1 precursor and its ion-channel targeting homologs.

The structurally and functionally characterized homologs (highlighted in red in Figure 5.8) were selected for comparison.

5.2.1.8 Physiochemical comparison of wZ1 and its ion-channel targeting homologs

Conotoxins are well known for their ability to target neuronal ion channels, which are used by snails to immobilize and catch their prey in the ocean [293]. To see if wZ1 has functional similarity to its ion-channel targeting homologs, we compared the physiochemical properties of their mature sequences (**Table 5.3**). Unlike wZ1, its ion-channel targeting homologs have post modifications on their mature sequences during the biosynthetic process. These post modifications on sequences are threefold: 1) transforming proline residues to hydroxyproline residues; 2) cyclizing glutamic acid residues to be pyroglutamic acid residues; 3) adding NH_2 to the C terminus. The modifications could not only increase the peptidyl solubility in the ocean, but also enhance its ion-channel targeting specificity, matching their roles in the hunting process of snails. To show the sequence conservation, we used the mature sequences and created a WebLog (**Figure 5.10**), in which the cysteine motif and the positive residues (H, K, and R) are high conserved, suggesting their hyper stabilities and ion-channel targeting ability.

Table 5.3 Physiochemical property of wZ1 and its ion-channel targeting homologs.

Peptides	Mature sequences	Amino acid residues	Charge	PI	Species/ Organisms	References
wZ1	--IQCC-----RCQSWPYMC-SVF-CC--	18	+1	7.78	*Schisandra chinensis*	*This work*
PIIIE	-HPPCCMYG--RC-RRYPGC-SSASCCQR-NH_2	24	+4	12.81	*Conus purpurascens*	*Ryan M Van Wagone et al, 2003*
PIIIA	-QRLCC-GFPKSC-RSRQ-CKPH-RCC---NH_2	22	+6	12.99	*Conus purpurascens*	*Alesia A Tietze et al, 2012*
TIIIA	-RHGCC-KGPKGC-SSRE-CRPQ-HCC---NH_2	22	+4	12.28	*Conus tulipa*	*Richard J Lewis et al, 2007*
GIIIA	--RDCC-TPPKKC-KDRQ-CKPQ-RCCA--NH_2	22	+5	11.87	*Conus geographus*	*K Wakamatsu et al,1992*
GIIIB	--RDCC-TPPRKC-KDRR-CKPM-KCCA--NH_2	22	+6	12.18	*Conus geographus*	*J M Hill et al, 1996*
BuIIIB	VGERCCKNGKRGC--GRW-CRDHSRCC---NH_2	24	+5	12.36	*Conus bullatus*	*Zhihe Kuang et al, 2013*
SIIIA	--QNCC-NG--GC-SSKW-CRDHARCC---NH_2	20	+2	11.62	*Conus striatus*	*Shenggen Yao et al, 2008*
CnIIIC	--QGCC-NGPKGC-SSKW-CRDHARCC---NH_2	22	+3	11.80	*Conus consors*	*Philippe Favreau et al, 2012*

Note: Sequence comparison of wZ1 and its ion-channel targeting homologs [294-299]. Hydroxyproline and pyroglutamic acid residues were labeled with * and #, respectively. The charge of the peptide was calculated by positive residues (R, K, and H) and negative residues (D and E).

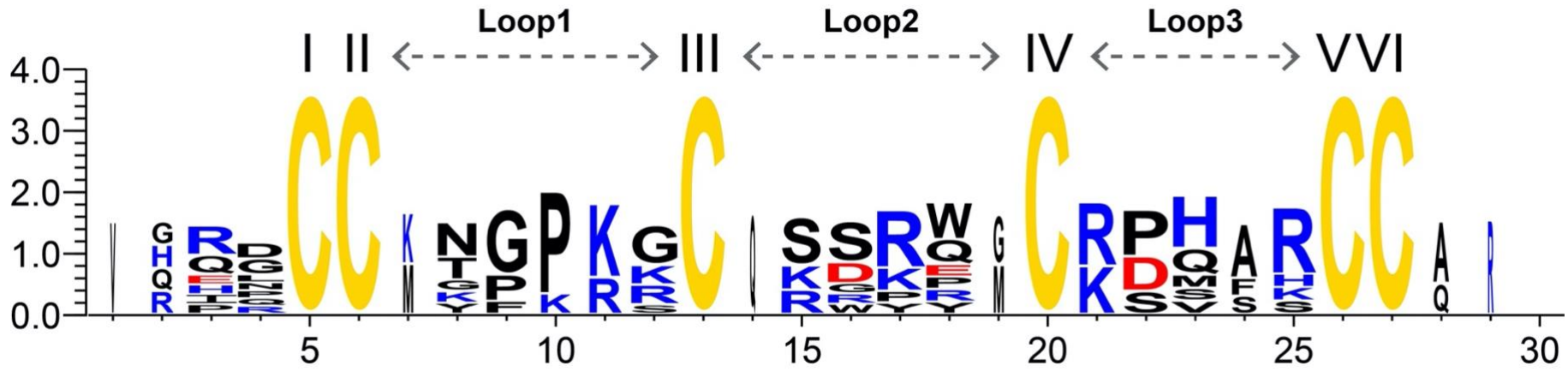


Figure 5.10 Sequence conservation of wZ1 and its ion-channel targeting homologs.

The overall height shows the amino acid conservation at a position. The height of each amino acid indicates the individual frequency at that position. The residues in this graph are labeled in orange (C), blue (H, R, and K), red (D and E), or black (others).

5.2.1.9 Structural comparison of wZ1 and its ion-channel targeting homologs

To compare their tertiary structures, we aligned the backbones of wZ1 and GIIIA, a potent ion-channel blocker (**Figure 5.11 A**). Although they share the same knottin-type disulfide connectivity, the two backbones are different from each other. Furthermore, we showed the surficial properties of GIIIA (**Figure 5.11 B**), showing that GIIIA is a highly positively charged (blue) and hydrophilic (cyan) cysteine-rich peptide. These properties are in contrast to the properties of one positive charge and hydrophobicity on the wZ1 surface. These results indicate that wZ1 might not target the ion channel due to its obvious different properties, when compared with its ion channel targeting homolog.

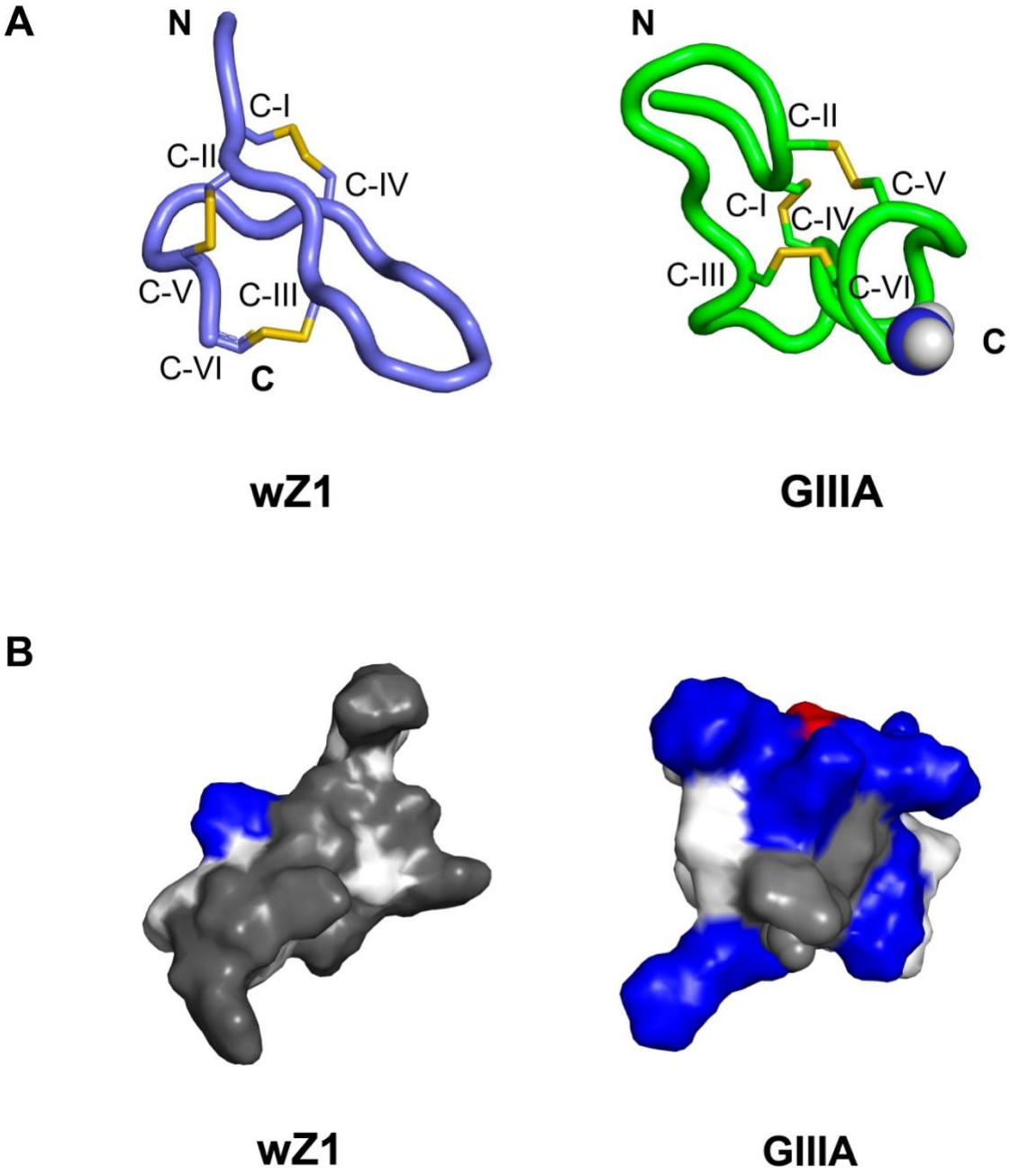


Figure 5.11 Structural comparison of wZ1 and GIIIA.

(A) The structural backbone of wZ1 and GIIIA (PDB: 1TCJ), a sodium ion-channel blocker. **(B)** Surface properties are indicated in different colors: positive charges in blue (R and K), negative charges in red (D), the hydrophobicity in dark grey (A, I, W, P, M, V, and F), and the hydrophilicity in light grey (T, *P, Q, S, and Y).

5.2.2 Cell-membrane permeability and membrane interaction of wZ1

5.2.2.1 Cell membrane integrity after wZ1 treatment

To see if wZ1 can damage the cell membrane integrity, we treated HepG2 cells and HUVEC-CS (human umbilical vein endothelial cells) with wZ1 in different concentrations, 1-20 μM, for 48 h. We measured the cell membrane integrity by using the lactate dehydrogenase (LDH) release assay **(Figure 5.12)**. We found that wZ1 did not damage the cell membrane and lead to LDH release from intracellular space to extracellular space in different cell lines.

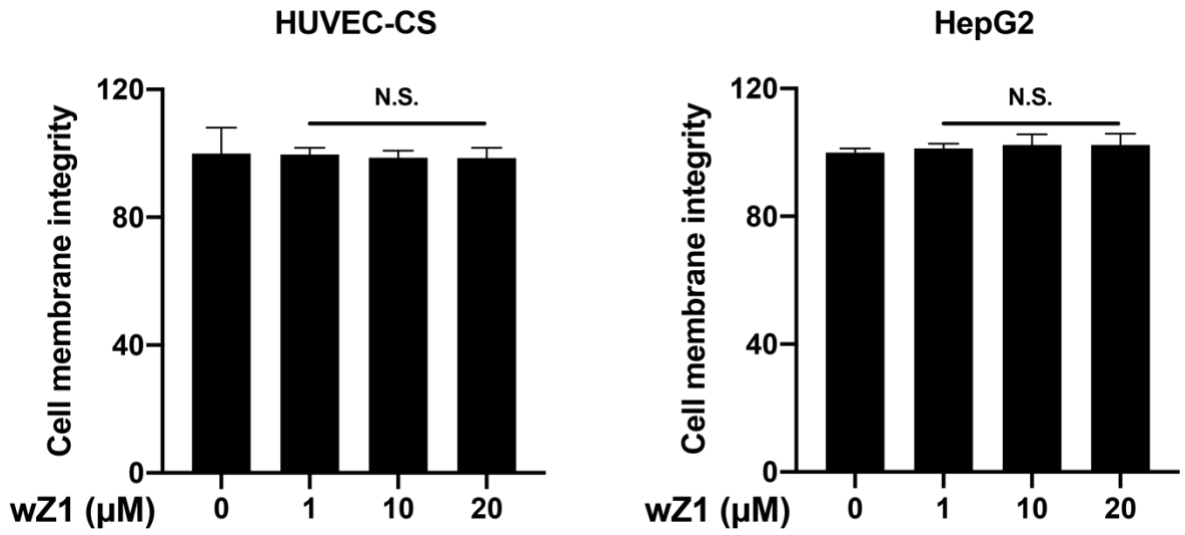


Figure 5.12 Cytotoxicity of wZ1.

WZ1 is not cytotoxic to HUVEC-CS and HepG2 cells, when tested in concentrations of up to 20 μM for 48h. The membrane integrity was analyzed by lactate dehydrogenase (LDH) release assay.

5.2.2.2 Affinity-enrichment proteomic profiling of wZ1

To explore the biological function of wZ1, we performed affinity-enrichment proteomic profiling to identify putative binding partners of wZ1 **(Figure 5.13)**. Using biotin-LC-wZ1, pull-down assays with HepG2 cell lysates were performed on NeutrAvidin-agarose beads followed by nano-liquid chromatography-tandem MS/MS analysis. Label-free quantitation was subsequently conducted along with a search against the human reference proteome (UP000005640) using MetaMorpheus version 0.0.320 software [270]. The identified protein groups with a fold-change of >2 compared with the biotin control group were identified. After subtracting the nonspecific binding proteins based on the CRAPome database [271], we identified 104 proteins in the biotin-wZ1 group **(Figure 5.13)**. Bioinformatic analysis using STRINGdb revealed that these binding proteins formed one cluster consisting of mitochondrial proteins. However, only eight of 104 pull-down proteins were associated with oxidative phosphorylation. Interestingly, 68 out of the 104 proteins were membrane-associated proteins. Based on this information, we speculated that wZ1 may interact with the lipid membrane to exert its biological effects.

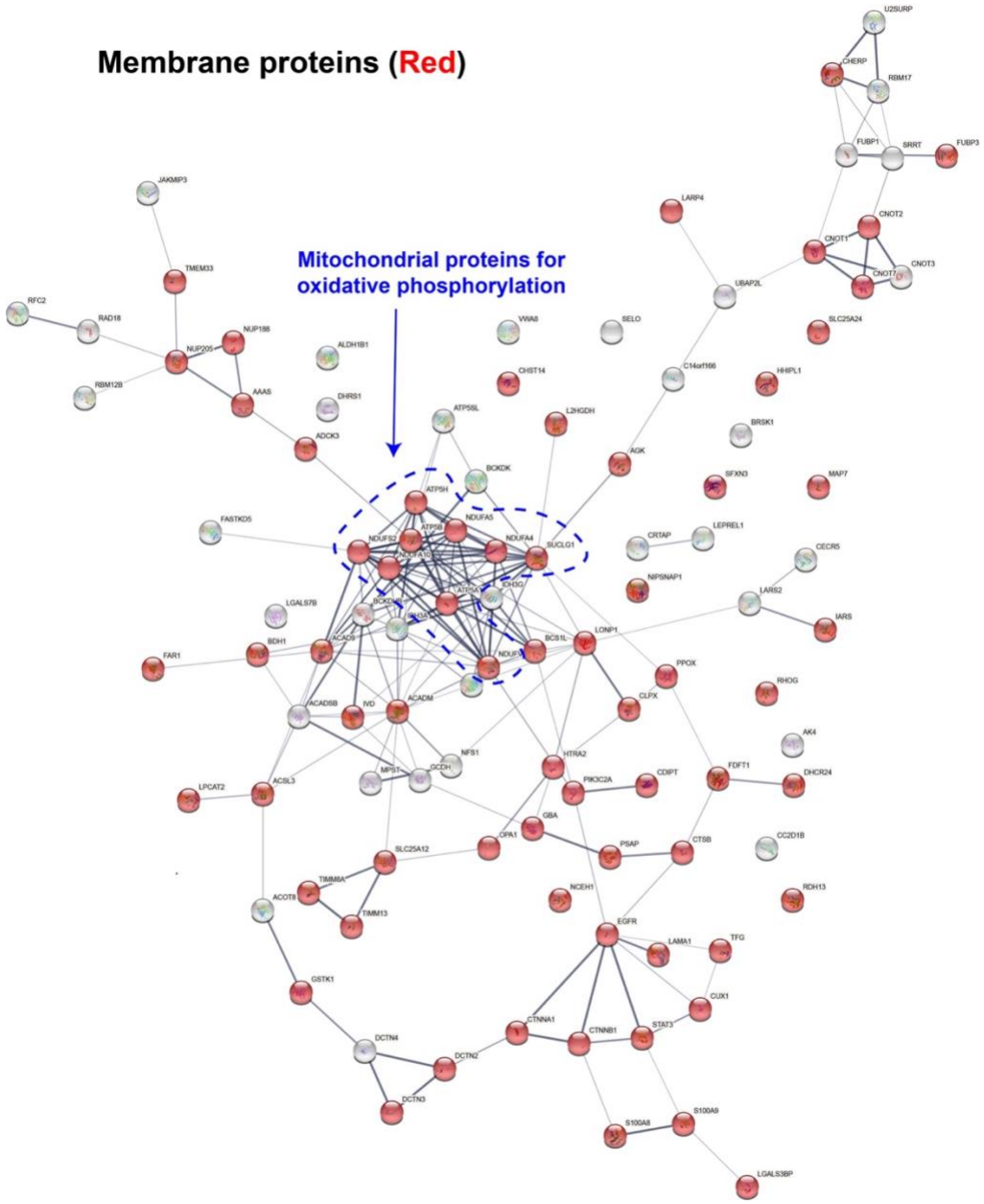


Figure 5.13 Affinity-enrichment mass spectrometry analysis of wZ1 using HepG2 cell lysate.

Biotin-LC-wZ1 was mounted on streptavidin beads and then incubated with HepG2 cell lysate at 4ºC overnight. The pulled proteins were analyzed by mass spectrometry. The membrane proteins binding to wZ1 were in red. Pull-down experiment was conducted by Dr. Kam A. and Dr. Loo S. N., and then analyzed by three of us.

5.2.2.3 AF488-wZ1 actively entering cells by the endocytosis

To determine the cell-entering capacity of wZ1 and intracellular localization, we labeled wZ1 with AF488 fluorophore or Biotin-LC at its N terminus (**Figure 5.14**). After labeling, we analyzed AF488-wZ1 treated HepG2 cells by flow cytometry. First, we selected single cells for the cellular uptake analysis (**Figure 5.15 A**). The cells started to increase the fluorescent intensity in a time-dependent manner and reached the plateau in 4 h (**Figure 5.15 B and C**). Second, we pre-exposed the cells to 37ºC and 4ºC for 30 min, treated them with AF488-wZ1 for 10 min, and then selected single cells for energy-dependent analysis (**Figure 5.16 A**). Compared with physiological temperature (37ºC), low temperature (4ºC) significantly decreased the cellular fluorescent intensity of AF488-wZ1 **(Figure 5.16 B and C)**. Third, we pre-treated cells with endocytosis inhibitors (nystatin, dynasore, and EIPA) for 30 min, and then selected single cells for endocytosis analysis (**Figure 5.17 A**). Nystatin and dynasore treatments significantly decreased the normal uptake of AF488-wZ1 **(Figure 5.17 B and C)**. These results show that AF488-wZ1 crosses the cells by energy-dependent endocytosis.

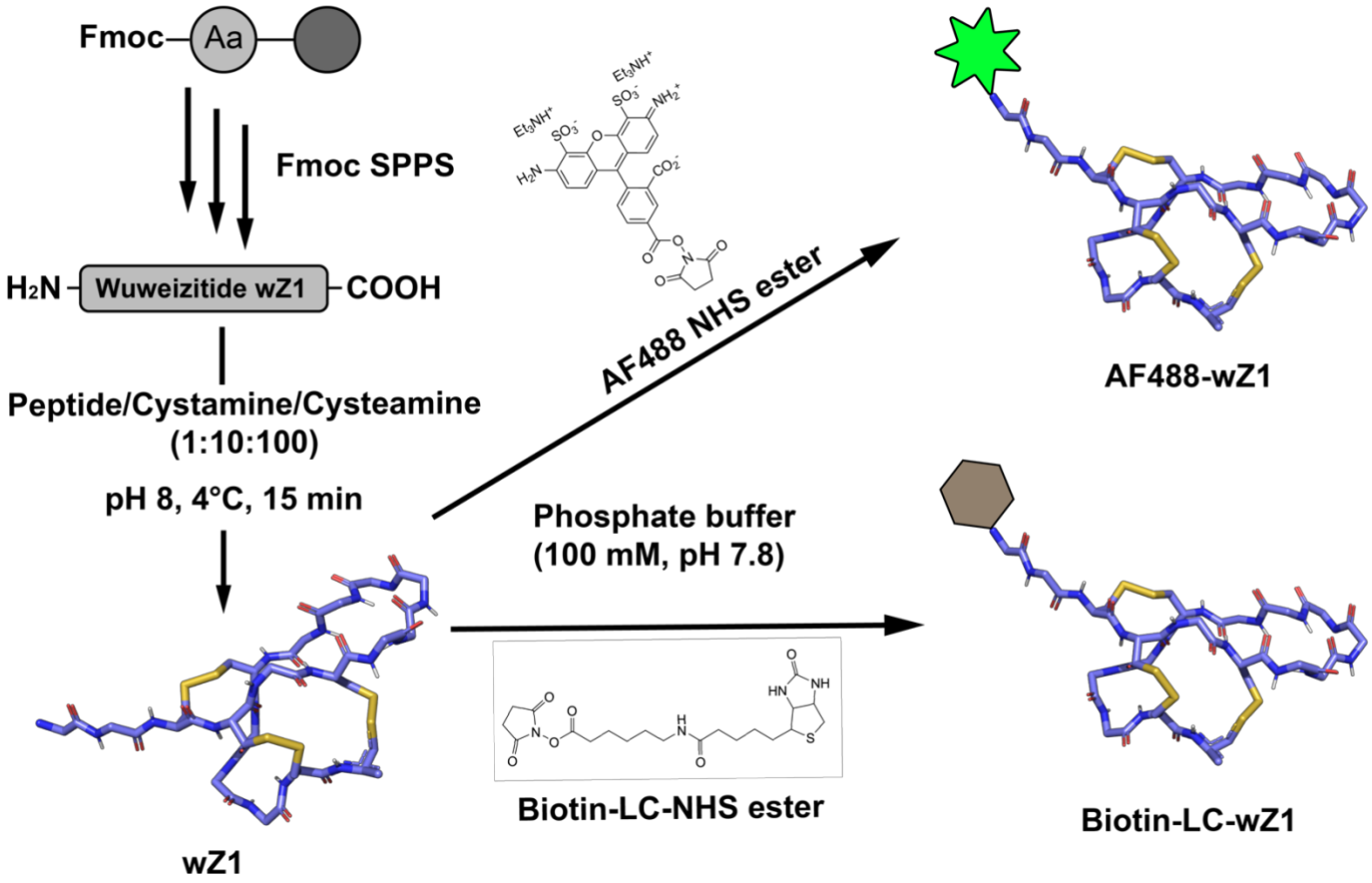


Figure 5.14 Schematic illustration for chemical synthesis and labeling of wZ1.

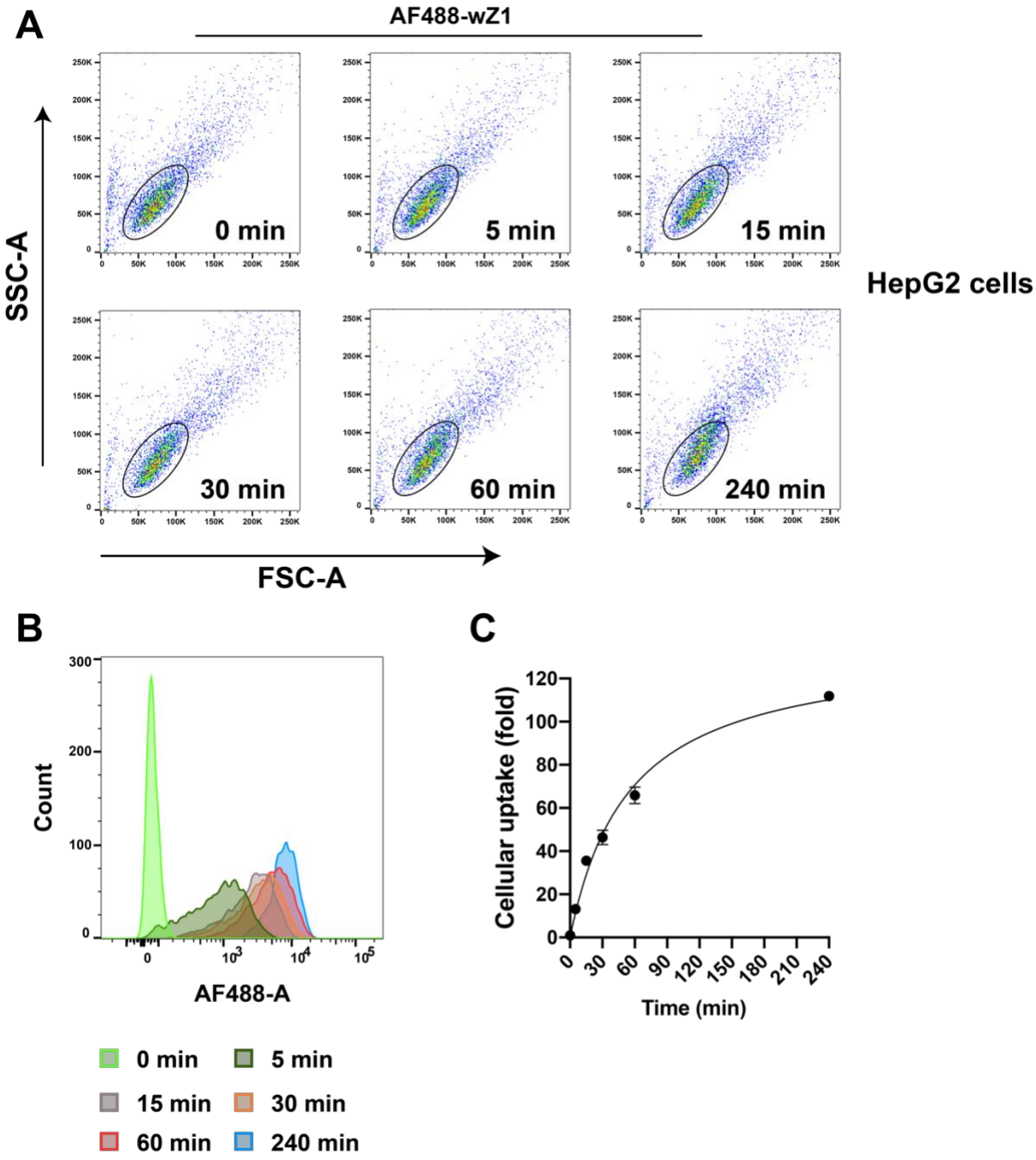


Figure 5.15 Cellular uptake of AF488- wZ1 in HepG2 cells.

(A) The HepG2 cells were treated with AF488-wZ1 for different periods and cells in elliptical circles were used for analysis. **(B)** Fluorescent peaks of the AF488--wZ1 treated cells. As the time increases, the peak shifts forward, meaning cellular uptake increases. **(C)** Quantification for cellular uptake of the AF488-wZ1 versus time.

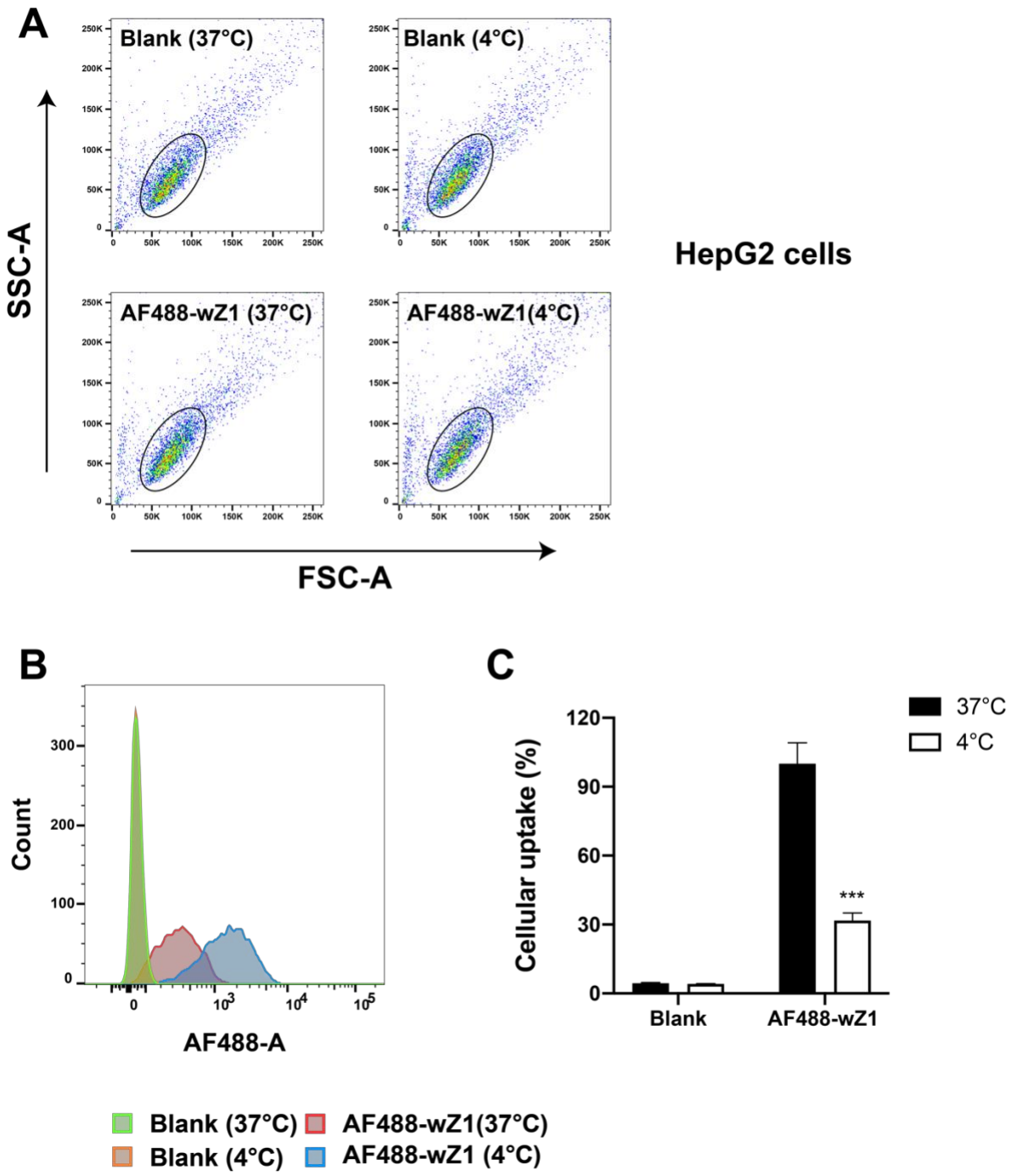


Figure 5.16 Temperature-dependent uptake of the AF488- wZ1 in HepG2 cells.

(A) The HepG2 cells were pre-incubated under 37ºC and 4ºC for 30 min and then treated with AF488- wZ1 for 10 min. The cells in elliptical circles were selected for analysis. **(B)** Fluorescent intensity for each group of cells. **(C)** Quantification of cellular fluorescent intensity in each group. Values are in mean ± S.D (n=3) with $P^{***} < 0.001$.

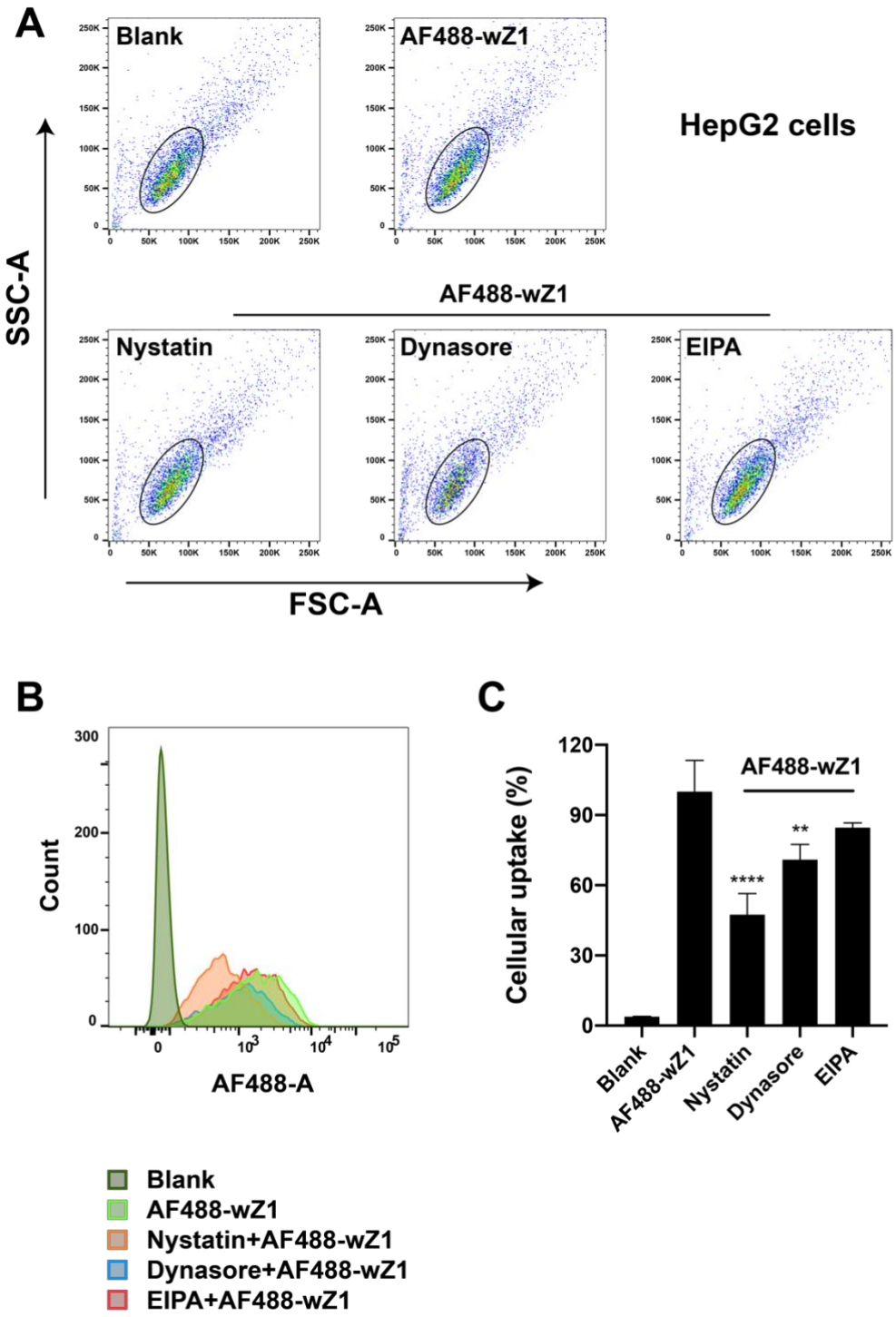


Figure 5.17 AF488- wZ1 entering HepG2 cells via endocytosis.

(A) The cells were pre-treated with different endocytosis inhibitors. The cells for analysis were selected in elliptical circles. **(B)** The cellular fluorescent intensity in each cell group. **(C)** Quantification of the cellular fluorescent intensity. Values are presented in mean ± S.D (n=3) with $^{**}P < 0.01$ and $^{****}P < 0.0001$.

5.2.3 Nuclear targeting of wZ1

5.2.3.1 Nuclear localization of wZ1

To further confirm the cell entering of AF488-wZ1 and investigate its intracellular localization, we imaged the cells after the treatment for 1h and 4h by confocal microscopy. In HepG2 cells, after 1 h treatment, we observed that most of the AF488-wZ1 molecules accumulated on the cell membrane, and that few of them were in the cytoplasm (**Figure 5.18 A, 1h**). After 4 h treatment, we found that AF488-wZ1 not only was in the cytoplasm, but also entered into the nucleus (**Figure 5.18 A, 4h**). Similar results were observed in HUVEC-CS (**Figure 5.18 B**). In addition, we used another cysteine-rich peptide, wS1, which was labeled with the same AF488 fluorophore, serving as a control. After 1h and 4 h treatments, the cellular uptake of AF488 labeled wS1 is less than the cellular uptake of AF488-wZ1, and AF488 labeled wS1 also did not accumulate in the nucleus (**Figure 5.19**). These results show that AF488-wZ1 can enter the cells and accumulate in the nucleus.

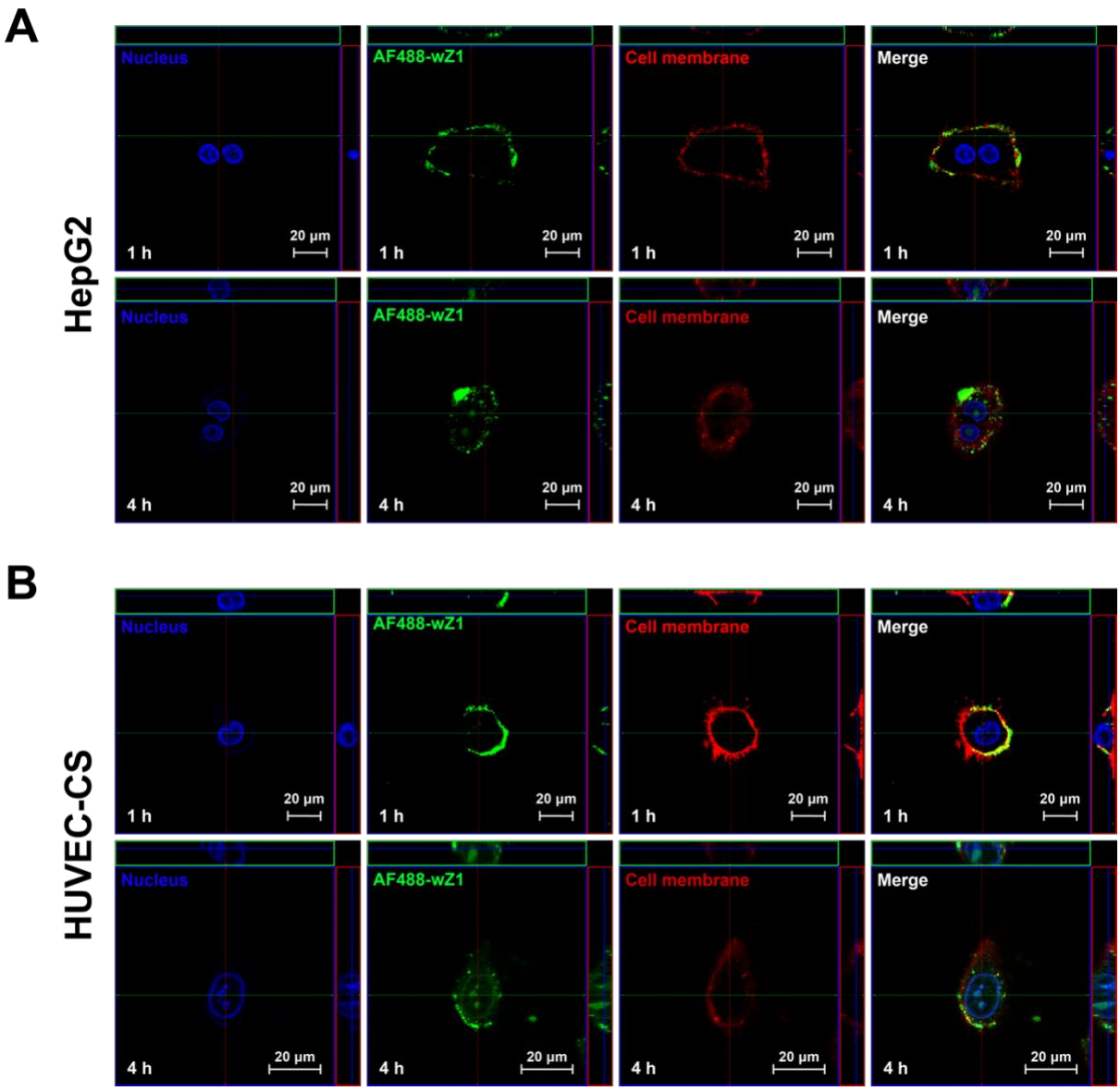


Figure 5.18 Nuclear localization of AF488-wZ1 in different cell lines.

(**A and B**) HepG2 cells and HUVEC-CS were treated with AF488-wZ1 for 1 h, washed with phosphate-buffered saline, and then imaged with a confocal microscope. (**C**) wS1, another cysteine-rich peptide, was labeled with the AF488, serving as a control.

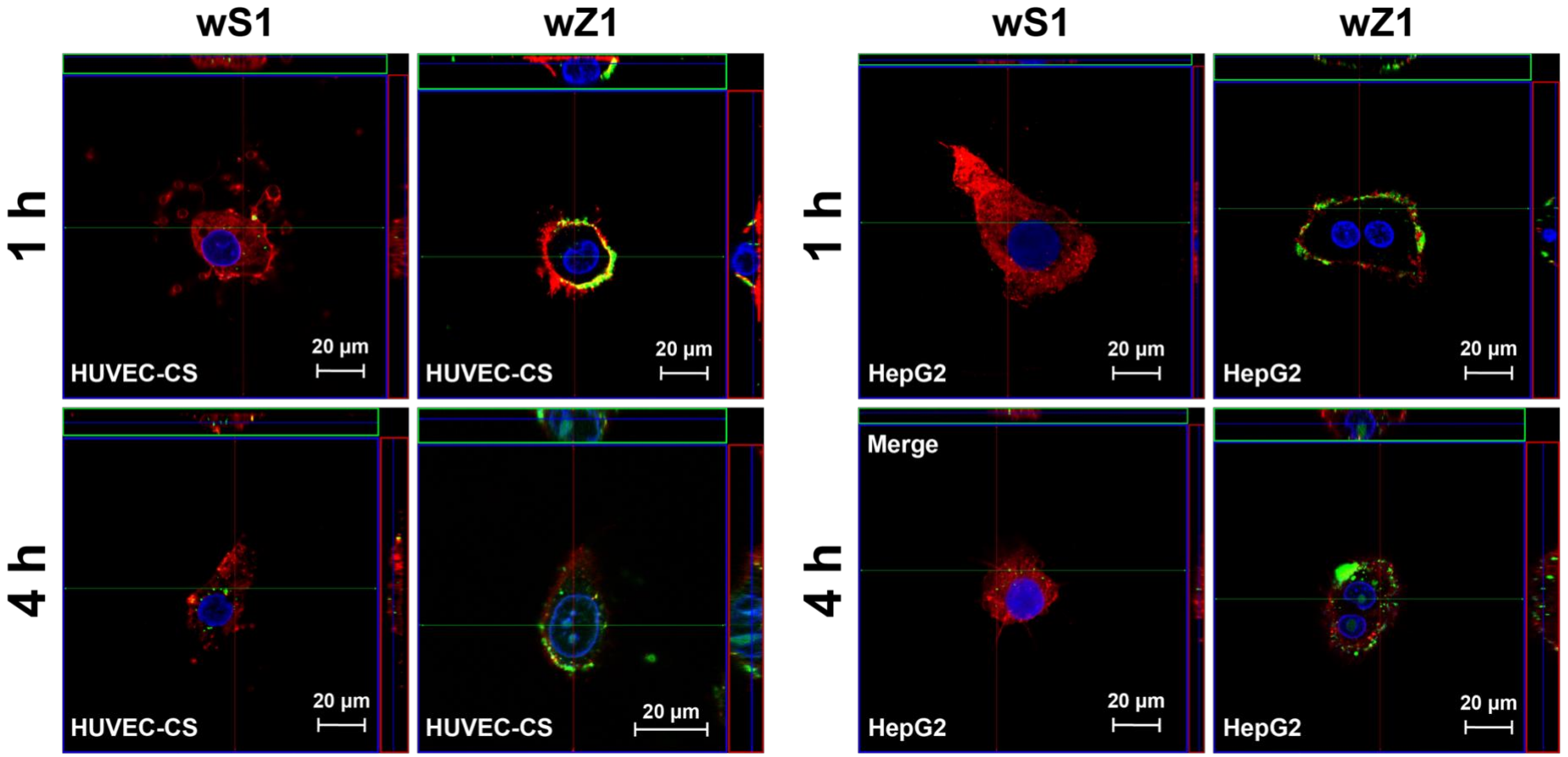


Figure 5.19 Comparison between AF488-wZ1 cell entering and AF488-wS1 cell entering.

AF488-wS1 was used as a negative control to show that wZ1 can have nuclear localization.

5.2.3.2 Verification of wZ1 nuclear localization

To exclude the influence of AF488 on the nuclear localization of wZ1, we labeled the amino terminus of wZ1 with biotin-LC and treated HUVEC-CS with biotin-LC-wZ1 for 24 h before nuclear/cytoplasmic fractionation. Our results showed that in both the blank group and biotin-LC-wZ1 treated group the nuclei were separated from the cytoplasm without cross-contamination. This is supported by two folds: 1) in the fractionated nucleus, we detected Histone H4 that is a nuclear marker, and did not detect the GAPDH that is a cytoplasmic maker; 2) in contrast, in separated cytoplasm, we detected GAPDH without Histone H4 (**Figure 5.20**). Based on a solid nuclear/cytoplasmic fractionation, we found that biotin-LC-wZ1 (M.W. = 2.5 kDa) could be detected at around 2 kDa on a Western blotting membrane, without peptidyl degradation after 24 h. These results further supported that wZ1 can target the nucleus with high structural stability inside of the cell.

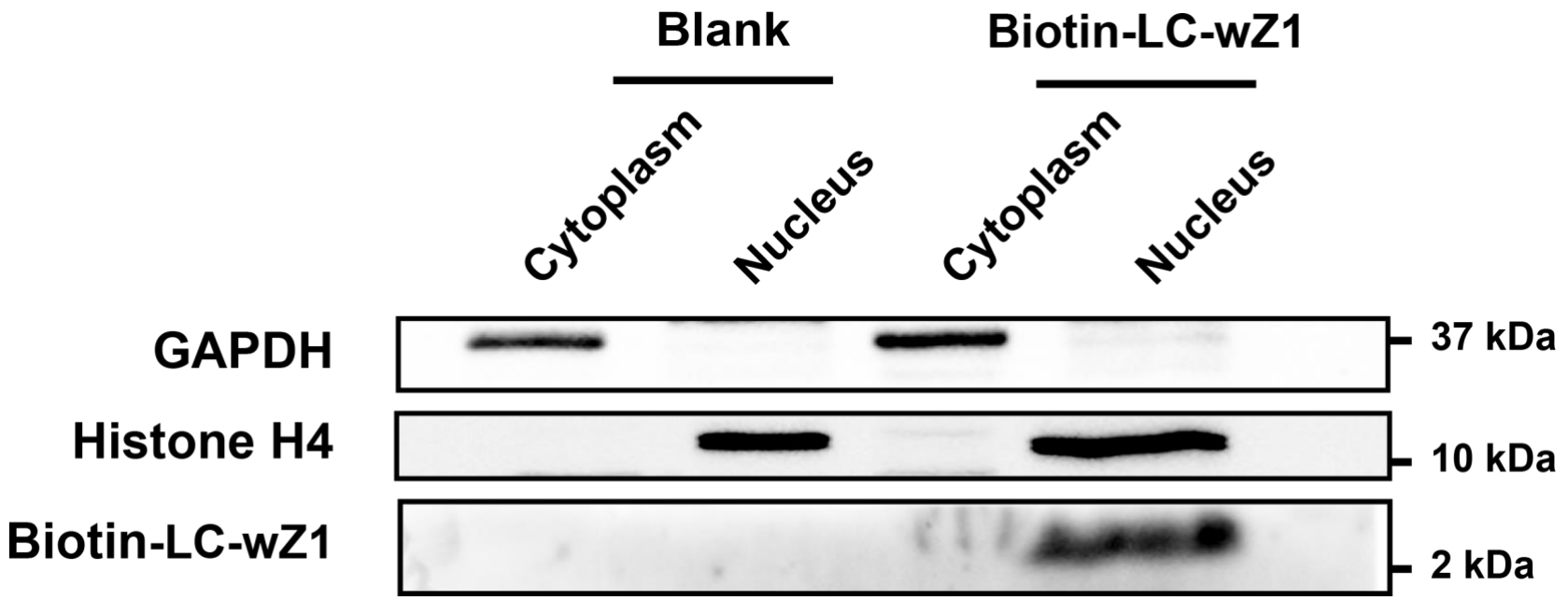


Figure 5.20 Western blotting analysis of biotin-LC- wZ1 nuclear localization.

HUVEC-CS were treated with the biotin-LC-wZ1 (20 μM) for 24 h and then the cellular nuclei were separated from the cytoplasm. GAPDH, existing in the cytoplasm rather than inside of the nucleus, was used as a cytoplasmic marker. In contrast, Histone H4, existing in the nucleus, was used as a nuclear marker. Thus, these makers were then used for assuring no cross-contamination between separated nuclei and the cytoplasm.

5.2.4 Down-regulation of cholesterol biosynthetic genes by wZ1 in HepG2 cells

To determine the influence of wZ1 on gene expression, HepG2 cells were treated with wZ1 for 24 h, followed by transcriptomic sequencing. Compared with the control group, the treatment group has 162 different expression genes (DEGs), among which 36 genes are downregulated and 126 genes are upregulated by wZ1 **(Figure 5.21).** For these DEGs, we conducted cellular process enrichment by using Gene Ontology (GO), and also performed pathway and disease enrichment by using Kyoto Encyclopaedia of Genes and Genomes (KEGG). In the GO process enrichment, the top seven hits (highlighted in red in **Figure 5.22**) were shown to be related to cholesterol biosynthesis, which plays a key role in regulating low-density-lipoprotein receptors (LDLR) [4]. Also, KEGG pathway enrichment suggested that the top one hit is steroid biosynthesis (**Figure 5.23**). Disease enrichment showed that the first hit is familial hypercholesterolemia (**Figure 5.24**). In detail, those DEGs influenced by wZ1 and involved in cholesterol biosynthesis are shown (**Figure 5.25**). Furthermore, these cholesterol biosynthetic genes are verified by another independent qPCR experiment (**Figure 5.26**). These results show that wZ1 can regulate cholesterol biosynthetic genes, which might play a role in LDLR regulation and have therapeutic potential for diseases like familial hypercholesterolemia.

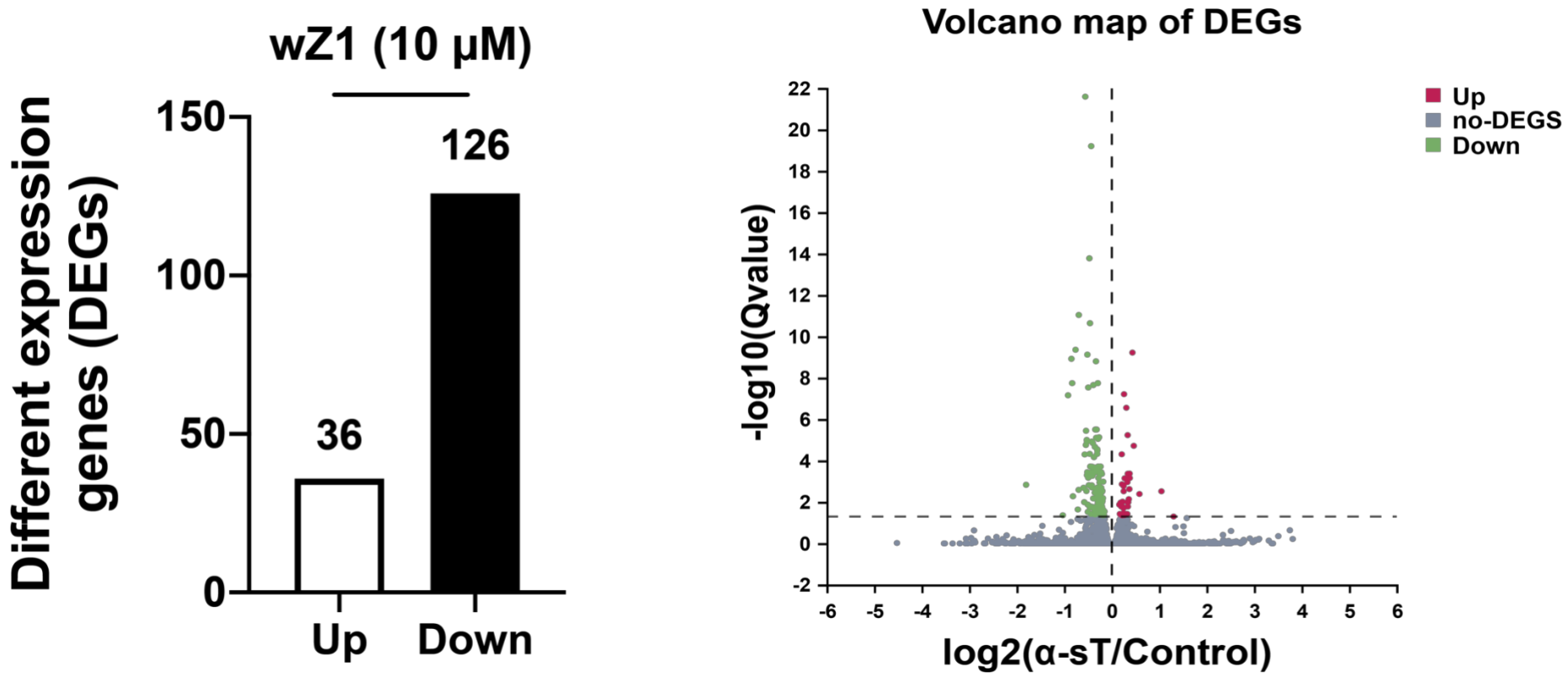


Figure 5.21 Number of different expression genes (DEGs) after wZ1 treatment in HepG2 cells.

The HepG2 cells were treated with wZ1 (10 μM) for 24 h and then the transcriptome of the cells was sequenced by the BGI company.

The DEGs with statistical significance were presented in this figure.

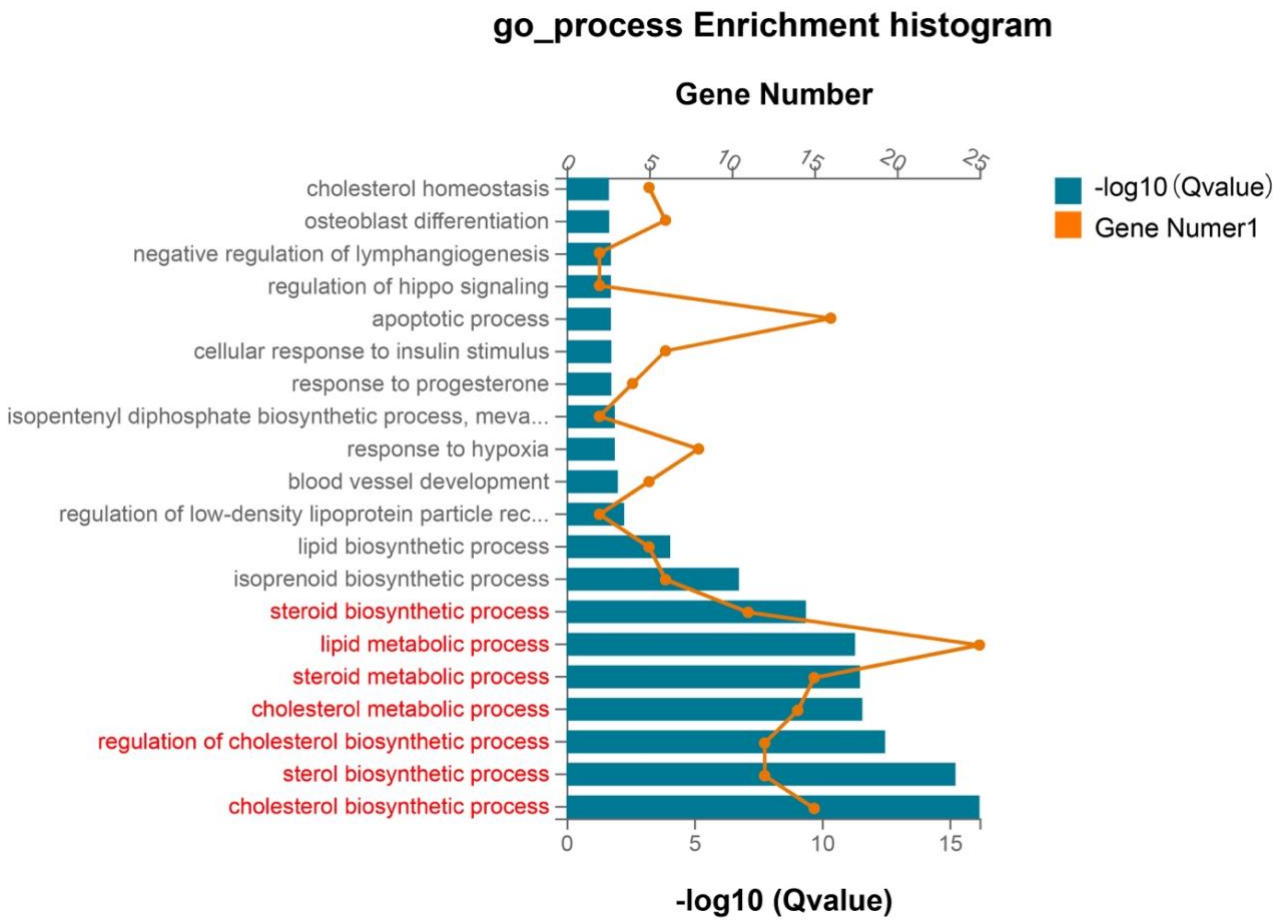


Figure 5.22 GO process enrichment of DEGs showing cholesterol biosynthesis.

The DEGs with statistical significance were enriched using the GO process in Dr. Tom server of BGI.

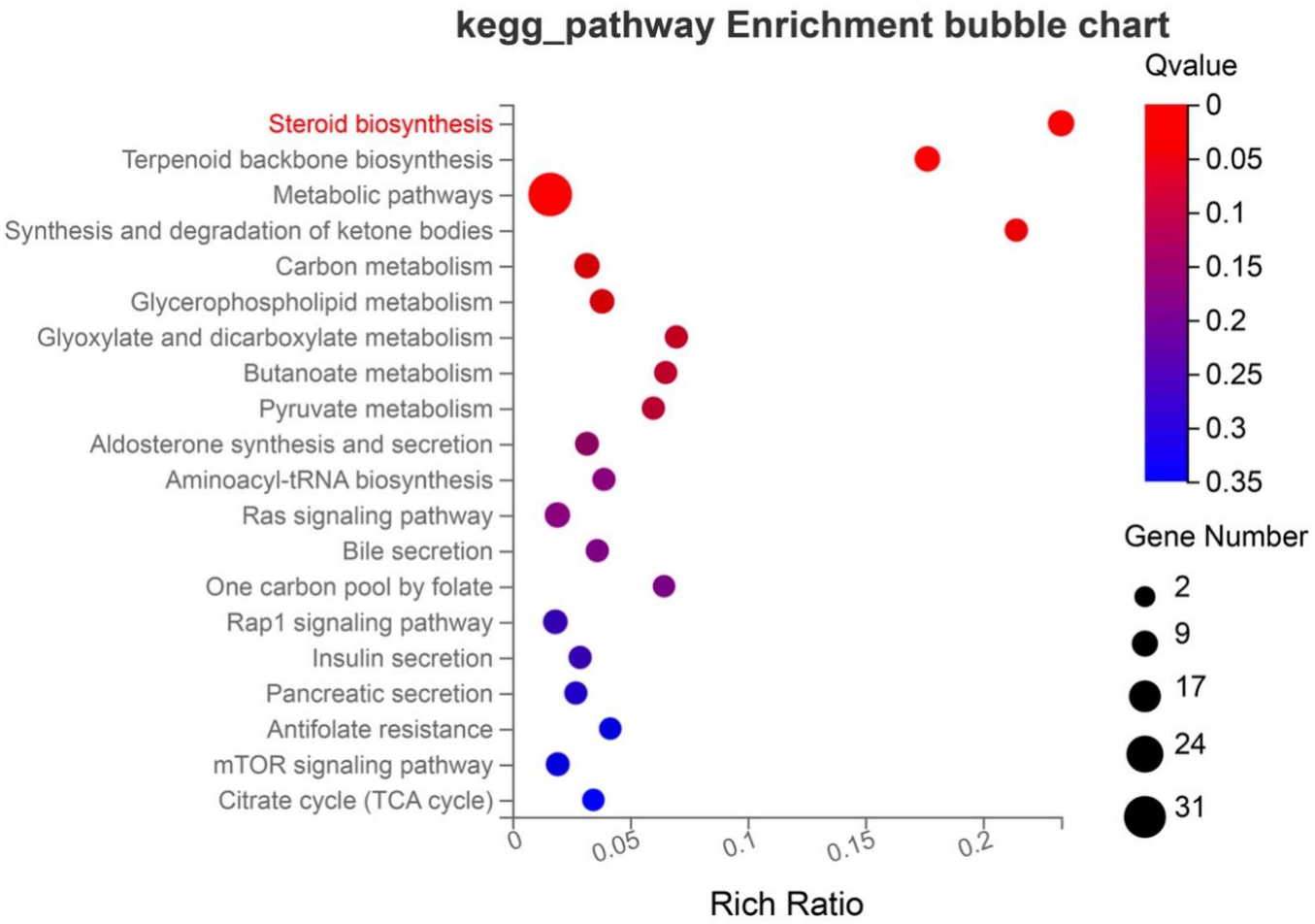


Figure 5.23 KEGG pathway enrichment of DEGs showing steroid biosynthesis.

The DEGs with statistical significance were enriched using the KEGG pathway in the Dr. Tom server of BGI.

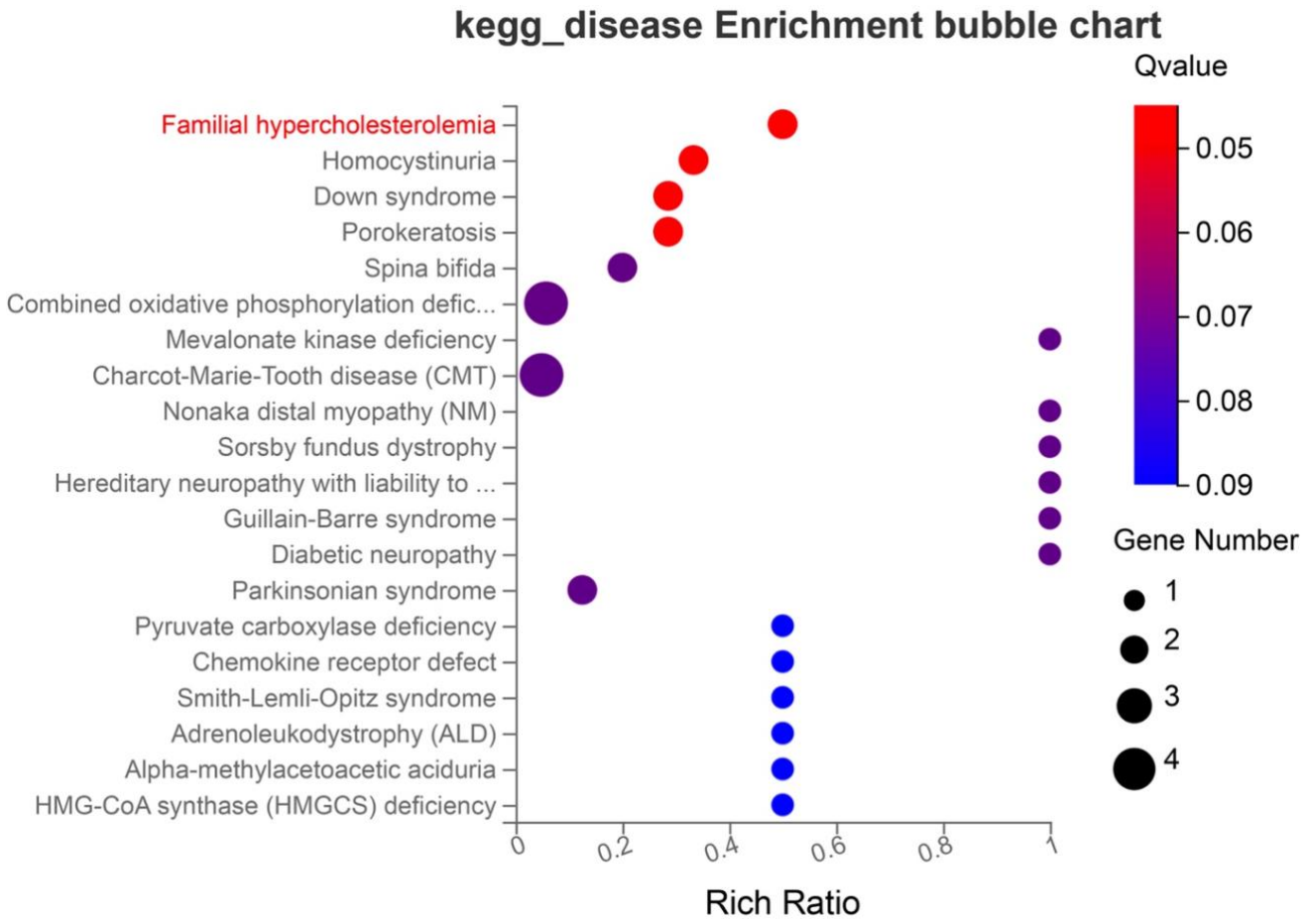


Figure 5.24 KEGG disease enrichment of DEGs showing familial hypercholesterolemia.

The DEGs with statistical significance were enriched using the KEGG pathway in the Dr. Tom server of BGI.

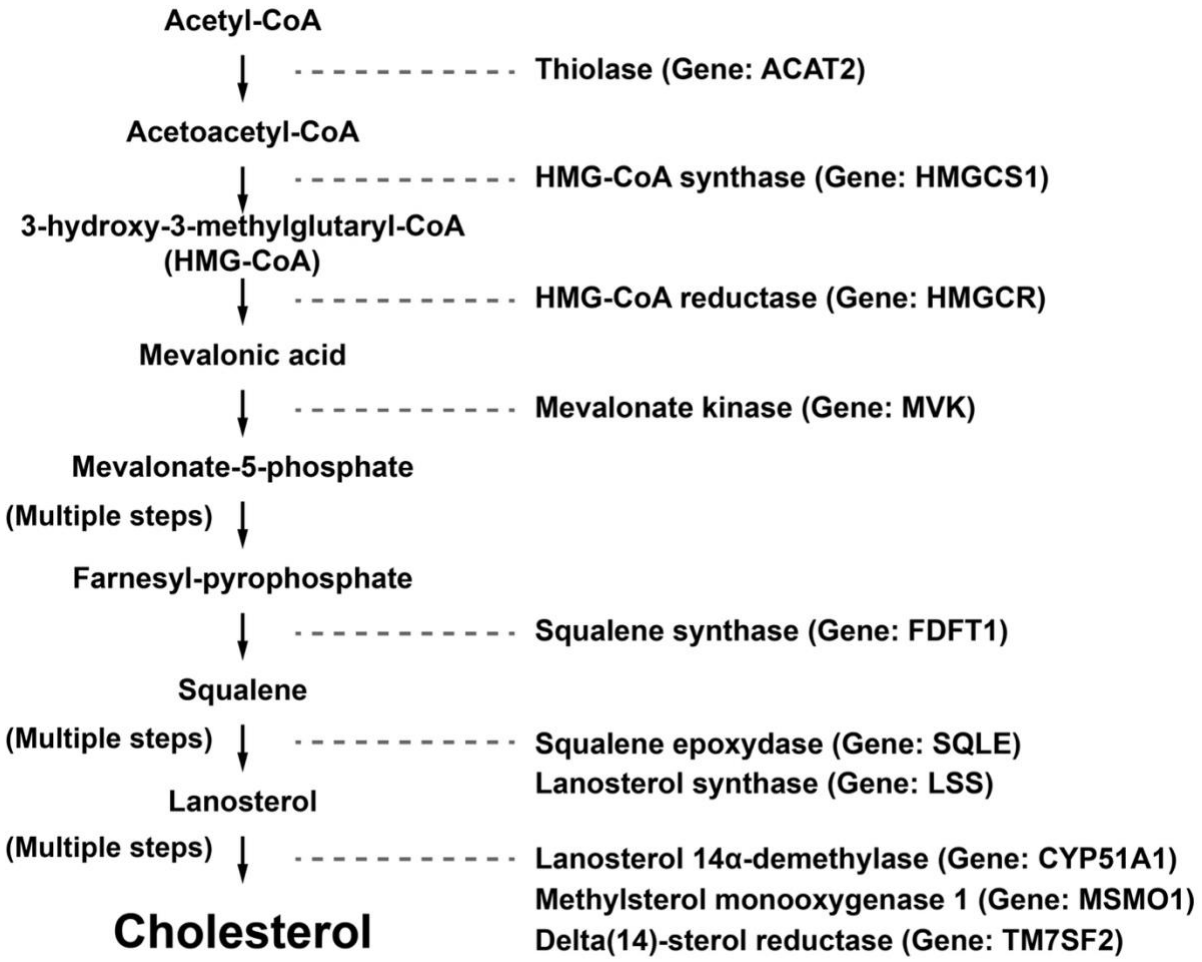


Figure 5.25 DEGs are involved in the cholesterol biosynthetic pathway.

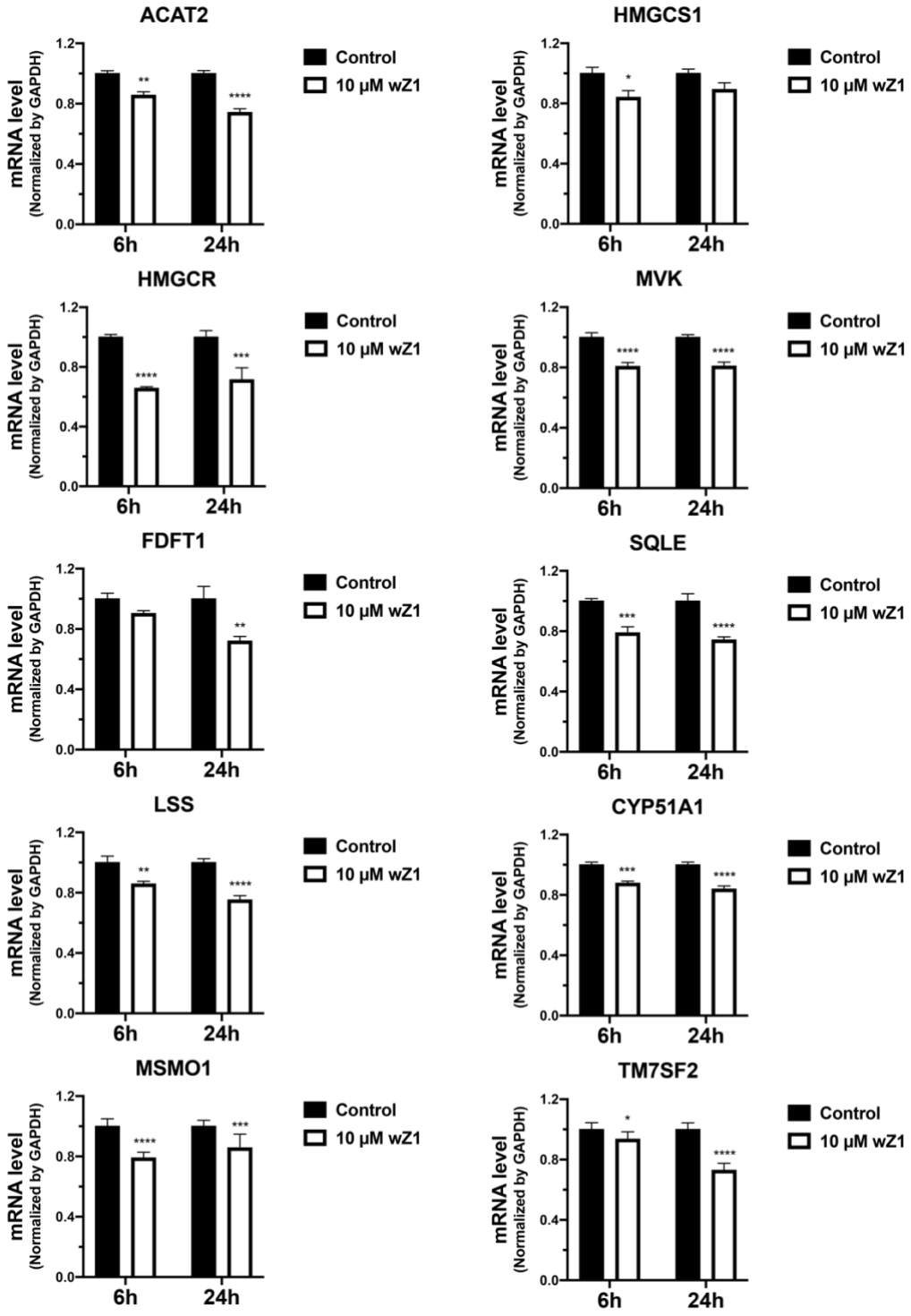


Figure 5.26 Verification of cholesterol biosynthetic DEGs by qPCR experiments. HepG2 cells were treated with wZ1 (10 μM) for 24 h before being analyzed by qPCR. The cell line and treatment conditions are identical to the ones of transcriptomic analysis.

5.2.5 wZ1 increases LDLR-mediated LDL uptake

Transcriptome profiling suggested that a 24-hour incubation with wZ1 downregulates cholesterol biosynthetic gene expression. We hypothesized that this effect may represent negative feedback caused by increased cellular cholesterol levels resulting from increased LDL uptake. Thus, we used flow cytometry to examine the effects of wZ1 on LDL uptake and LDLR expression **(Figure 5.27 and 5.28)**. The results indicated that wZ1 increases the cellular uptake of fluorescent-labeled LDL (LDL-Dylight™ 550) and the surface expression of LDLR in a dose-dependent manner using HepG2 cells. To support these results, confocal microscopic images revealed that LDL-Dylight™ 550 colocalized with LDLRs, which accumulated on the cell membrane of wZ1-treated HepG2 cells **(Figure 5.29)**. By contrast, minimal amounts of LDL were bound to LDLRs, which were evenly distributed in the cytoplasm and the membrane of the untreated cells. In addition, we found that wZ1 at 1 μM is better than lovastatin at 40 μM, in terms of increasing LDLR-mediated LDL uptake **(Figure 5.30)**. These results suggest that wZ1 is more potent than lovastatin in promoting LDL uptake in HepG2 cells.

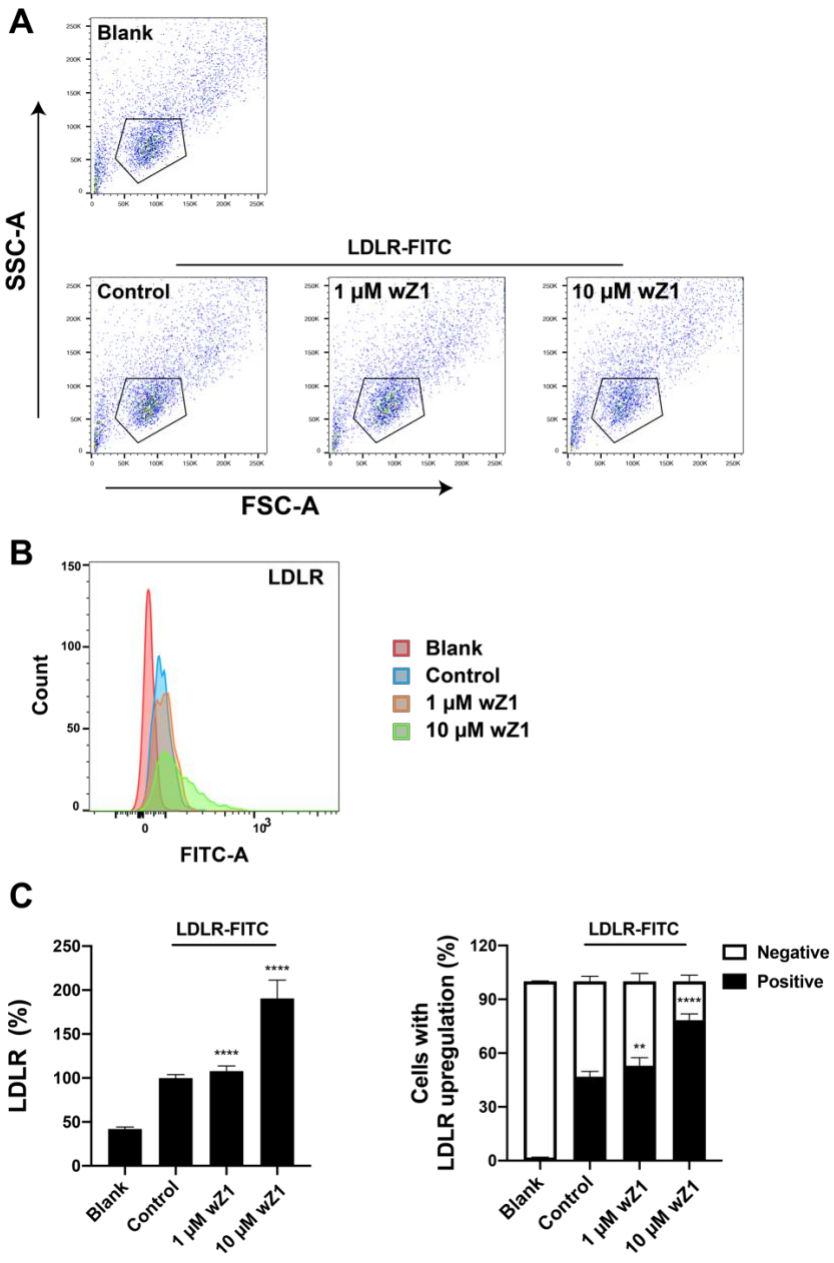


Figure 5.27 wZ1-induced upregulation of LDLR in HepG2 cells.

(A) The cells were treated with different doses of wZ1 for 24 h and then LDL receptor (LDLR) was probed by LDLR-FITC. The cells in the pentagon were selected for analysis. **(B)** Fluorescent intensity for each group of cells. **(C)** Quantification of mean fluorescent intensity of LDLR-FITC in each group and cells with LDLR upregulations. Values are presented by mean ± S.D (n=6) with $^{**}P < 0.01$ and $^{****}P < 0.0001$.

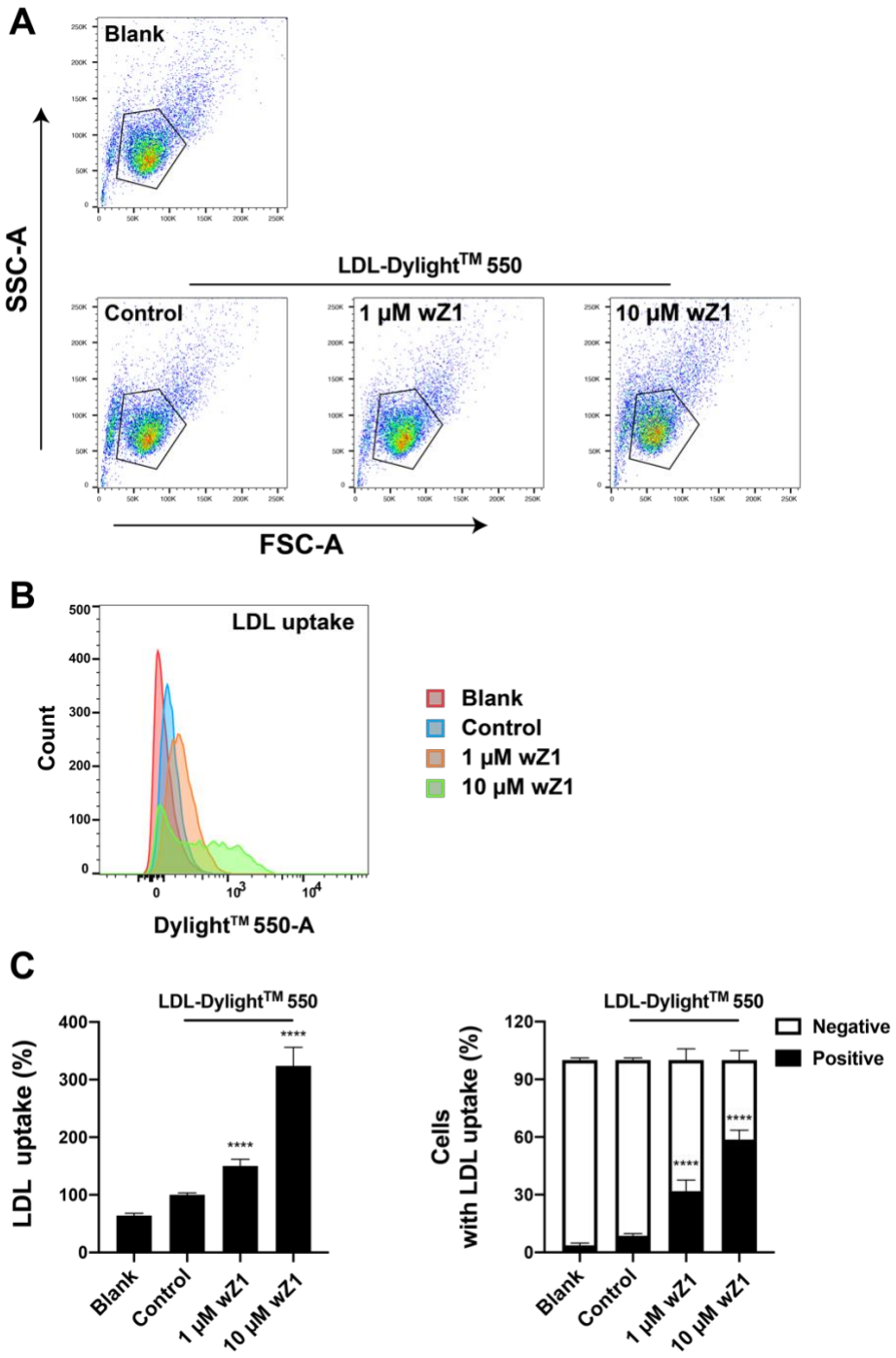


Figure 5.28 wZ1 increasing LDL uptake in HepG2 cells.

(A) The cells were treated with wZ1 in different concentrations for 24 h and then incubated with LDL-Dylight™550 for 24 h. The cells in the pentagon were selected for analysis. **(B)** Fluorescent intensity for each group of cells. **(C)** Quantification of mean fluorescent intensity of LDL-Dylight™550 in each group and cells with LDL uptake. Values are presented by mean ± S.D (n=10) with ****$P < 0.0001$.

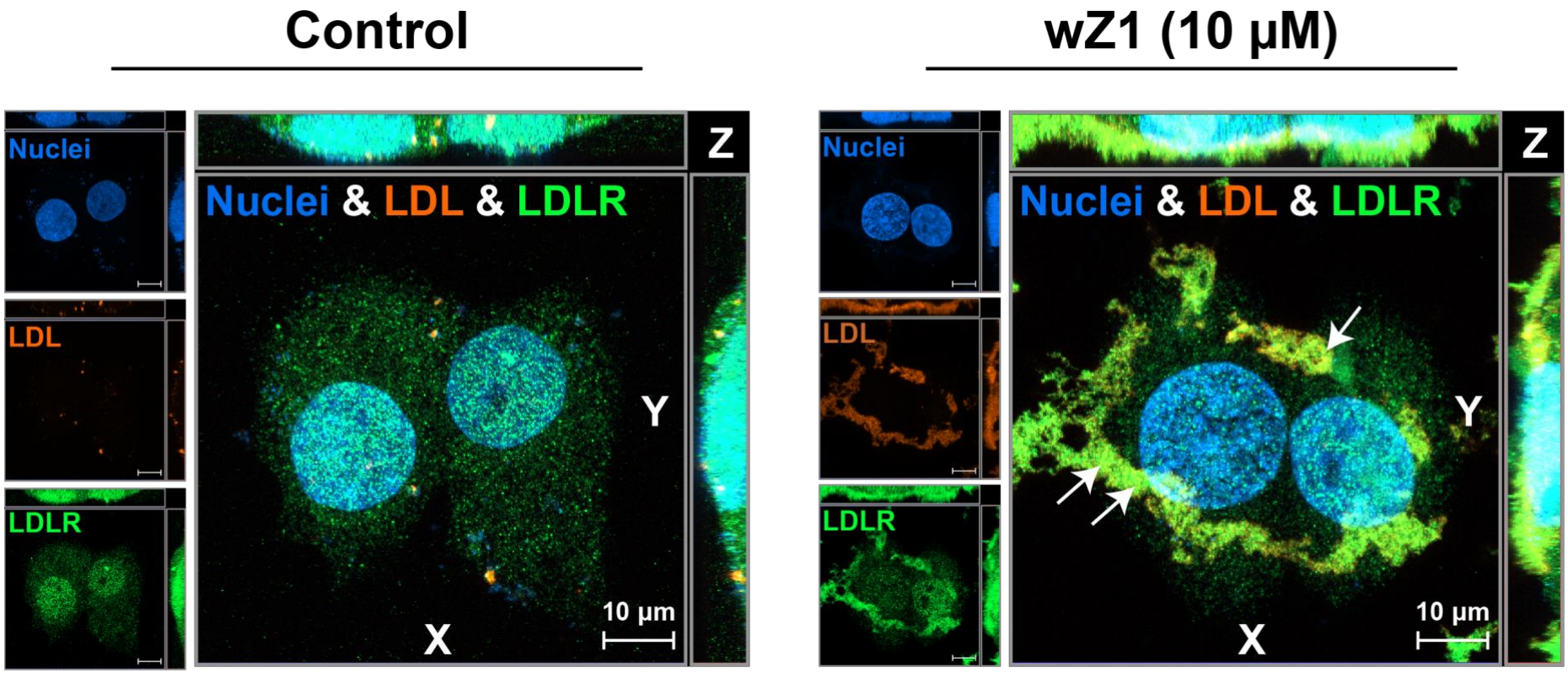


Figure 5.29 Confocal micrographs of LDLR-mediated LDL uptake after wZ1 treatment in HepG2 cells.

Confocal microscopic images of HepG2 cells treated with wZ1 (10 μM) for 24 h showed that LDL-Dylight^{TM}550 colocalizes with LDLRs at the cell membrane (white arrows). The nuclei were stained by Hoechst 333242 (1 μg/ml), and LDLRs were stained using the LDLR-FITC primary antibody conjugate. The cells were imaged in Z-stack or 3D (X, Y, and Z), and the images were shown in maximum intensity projection.

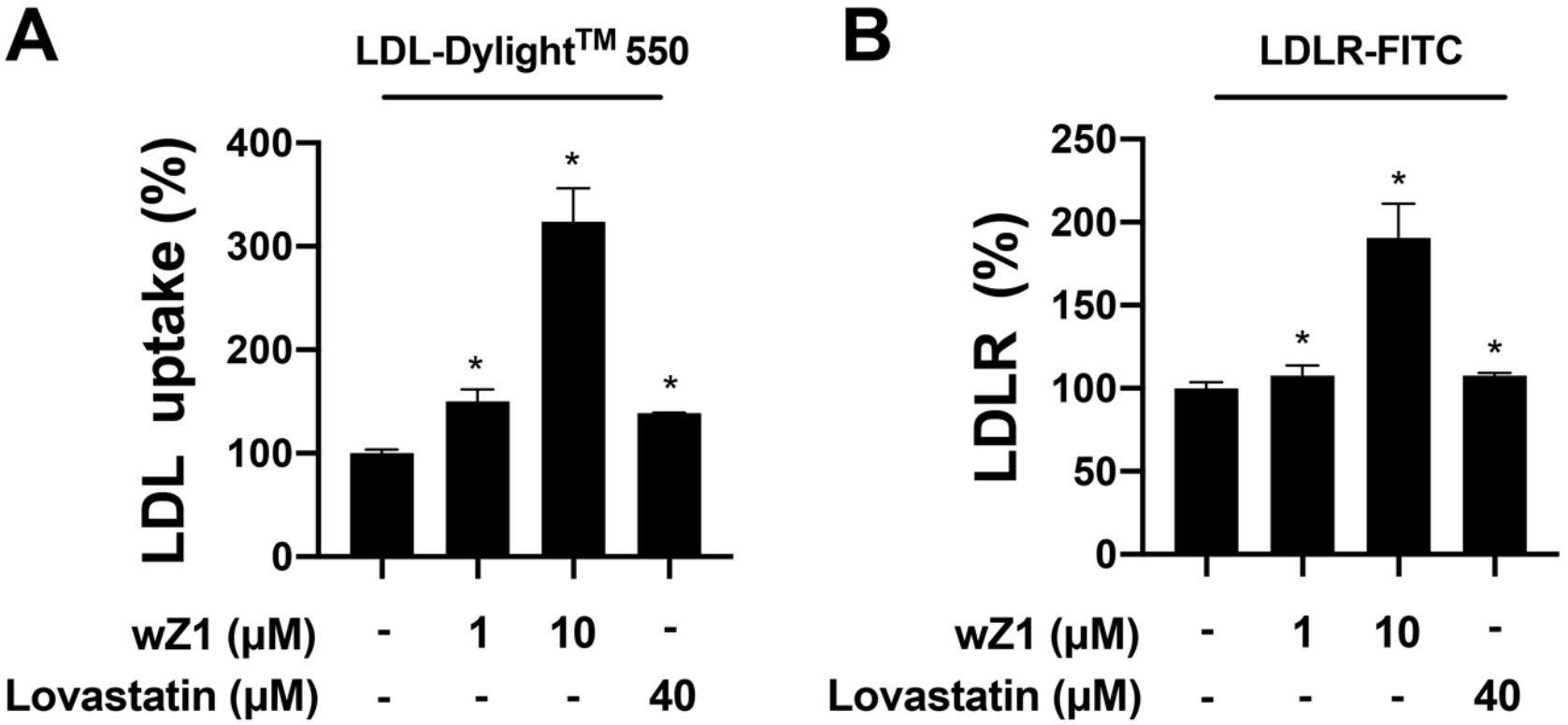


Figure 5.30 wZ1 is more potent than lovastatin on LDLR-mediated LDL uptake in HepG2 cells.

5.2.6 Increase in LDLR-mediated LDL uptake via lipid rafts

Based on our affinity-enrichment mass spectrometry experiments, we found that wZ1 does not interact directly with HMG-CoA reductase, PCSK-9 or LDLR, suggesting that wZ1 may utilize atypical mechanisms to reduce LDL uptake. Moreover, our results showed that the cellular uptake of wZ1 involves lipid rafts, and therefore, we examined the involvement of lipid rafts in wZ1-mediated LDL uptake. We found that wZ1 increased the formation of lipid rafts using flow cytometry **(Figure 5.31A)**. To determine the relationship between lipid raft formation and LDL uptake, we found that by inhibiting lipid raft formation with nystatin, wZ1-mediated LDL uptake was significantly reduced (**Figure 5.31B**). Indeed, confocal microscopic images revealed that lipid rafts (probed using CT-B-Alexa Fluor™ 594) colocalize with LDLR at the cell membrane of wZ1-treated HepG2 cells **(Figure 5.32)**. Overall, these results suggest that wZ1 increases lipid raft formation to promote LDL uptake via LDLR in HepG2 cells.

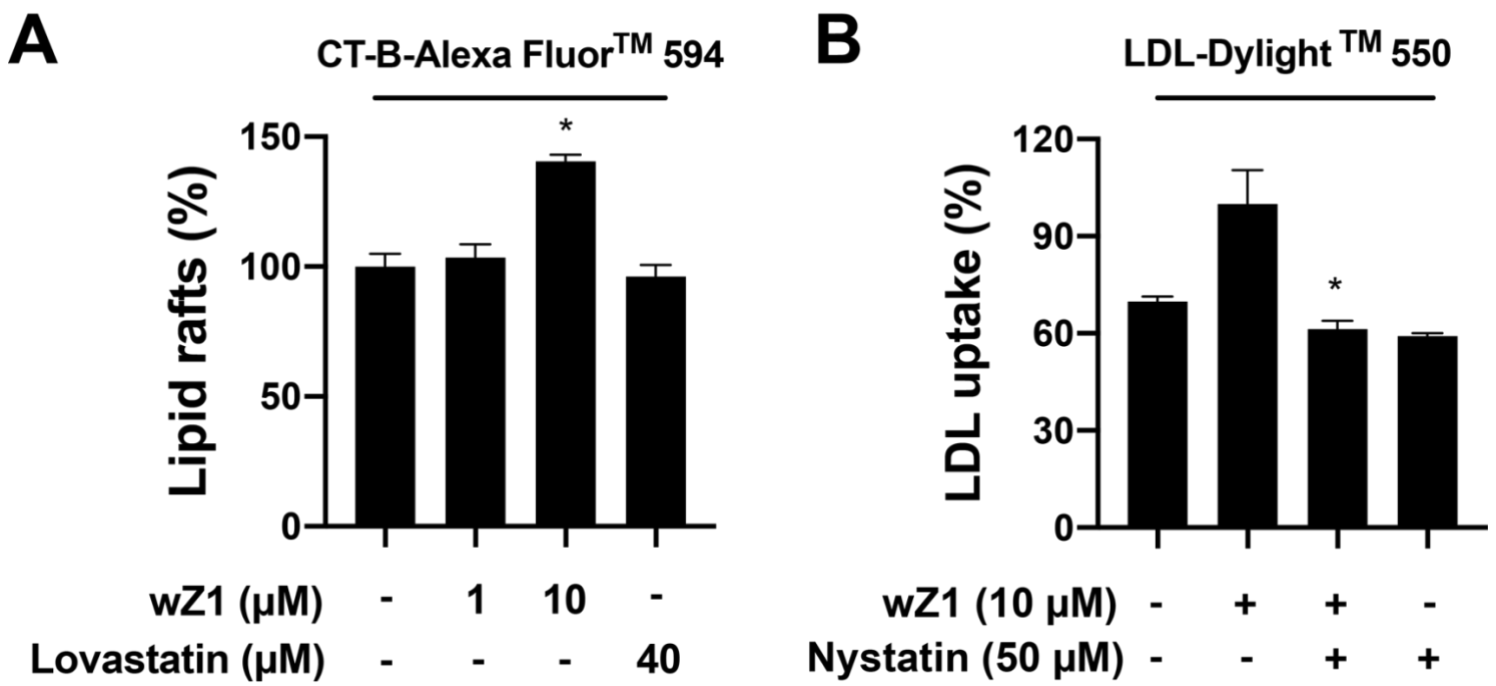


Figure 5.31 wZ1 induces LDL uptake via lipid raft formation.

(A) wZ1 promotes lipid raft formation in HepG2 cells as determined by flow cytometry using the lipid raft stain, CT-B-Alexa Fluor™ 594. *$p < 0.05$ was considered statistically significant, and values are presented as the mean ± SD (n = 4). (B) Nystatin, a cholesterol sequestering agent that disrupts lipid raft formation, inhibited wZ1-induced LDL uptake as analyzed by flow cytometry using LDL-Dylight™ 550. *$p < 0.05$ was considered statistically significant. Results are presented as the mean ± SD (n = 4).

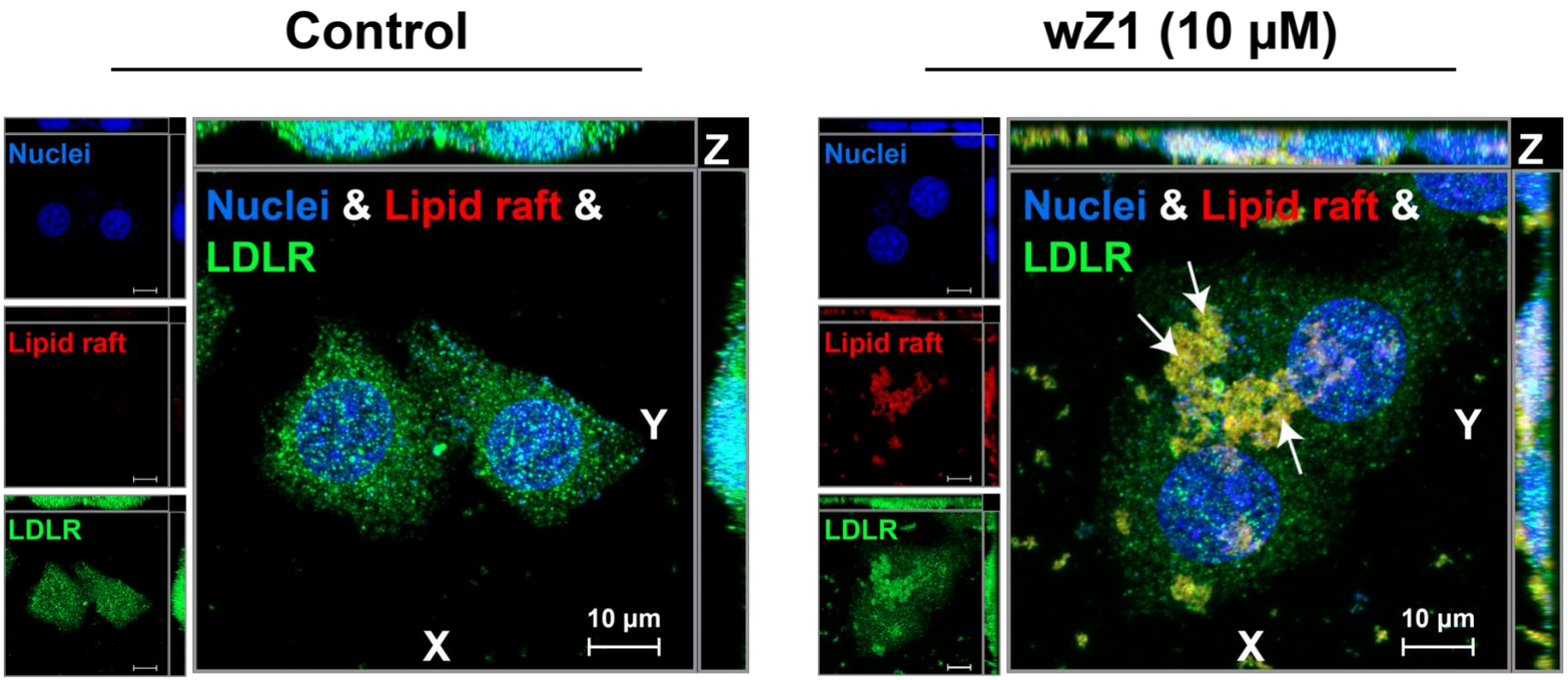


Figure 5.32 wZ1 increases redistribution of LDLR to lipid rafts.

Confocal microscopic images of HepG2 cells treated with or without wZ1 (10 µM) for 24 h show that LDLRs are clustered in lipid rafts (white arrows) after treatment with wZ1. The nuclei in blue were stained with Hoechst 333242 (1 µg/ml). LDLR in green was probed using an LDLR-FITC primary antibody conjugate. Lipid rafts in red were probed with CT-B-Alexa Fluor™ 594.

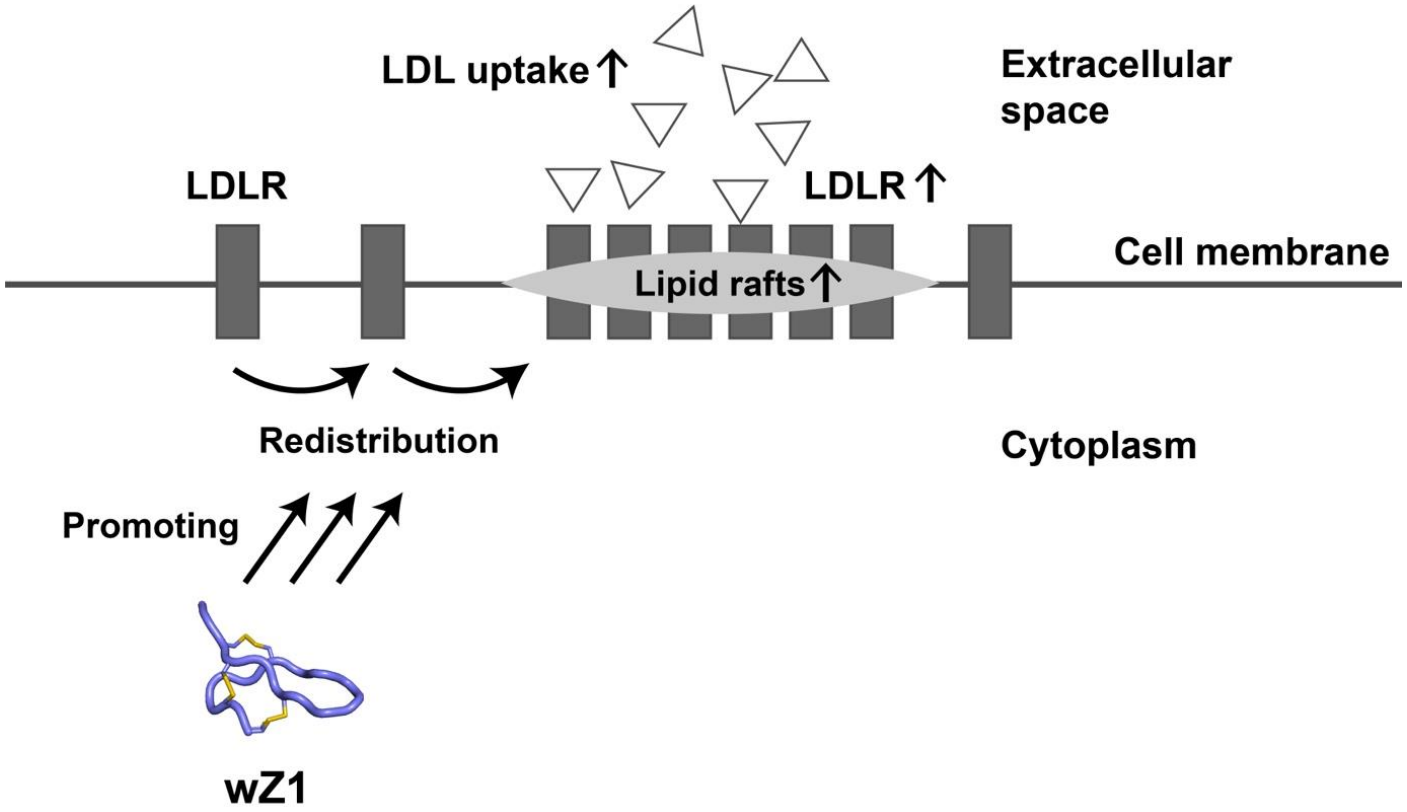


Figure 5.33 Schematic diagram for wZ1 promoting LDLR redistribution into lipid rafts to increase LDL uptake.

5.3 Discussion

S. chinensis is a Chinese traditional medicine with a long history, but its modern studies were primarily initiated in 1940s by the Soviet Union [300-304]. After 1960s, *S. chinensis* was gradually positioned as an adaptogen in medicinal applications. Based on whether the molecules have genetic information, the *S. chinensis* contains two types of biomolecules: 1) non-bioinformatic molecules (non-BMs); 2) bioinformatic molecules (BMs). As the first non-BMs, schizandrin with activities on mental performance was identified from the seeds of *S. chinensis* in 1951 [195]. Following the schizandrin discovery, other non-BMs, like deoxyschisandrin [305] and gomisin A-F [306, 307], were gradually reported. However, to our best knowledge, no bioactive BMs are reported so far. Here, we report a novel cysteine-rich peptide (CRP) from *S. chinensis* with its chemical structure and biological function.

Structurally, wuweizitide (wZ1) is a cysteine-rich peptide with six cysteine (6C) residues. *In planta*, 6C-CRPs have at least four major families, including gingotide, knottin, thionin, and jasmitide families, each of them with different disulfide connectivities and/or cysteine motifs [289]. Among these 6C-CRPs, the knottin family shows the same disulfide connectivity as wZ1, sharing knottin-type connectivity (CysI-CysIV, CysII-CysV, and CysIII-CysVI). However, no plant-derived 6C-CRP families were shown to exhibit a wZ1 cysteine motif (CC-C-C-CC). To determine the occurrence of other CRPs with such cysteine motif *in planta*, we performed data mining in NCBI and OneKP databases. Our results showed that no 6C-CRPs with CC-C-C-CC motif could be found. These results suggest that wZ1 is a novel and unusual 6C-CRPs *in planta*.

Using mass spectrometry to profile CRPs from medicinal plants, we identified wZ1 from *S. chinensis* fruit to be a positively charged CRP with six cysteine residues, which

account for 33.3% of the total amino acid content (18 residues). Thus, CRP wZ1 may be further classified as a hyperdisulfide-constrained peptide, a subfamily of CRPs characterized by their unusual cysteine-rich content (>30%). Additionally, we found that wZ1 is proteolytically processed from an unannotated 256-nucleotide smORF-encoded precursor. Because the proteolytic cleavage site occurs between Asn-Ile at wZ1 N-terminus, we speculate that asparaginyl endopeptidase is involved in its hydrolytic cleavage or maturation, like that occurring in roseltides from *Hibiscus sabdariffa* [80] and jasmintides from *Jasminum sambac* [143].

Although wZ1 homologs *in planta* are rare, we found that wZ1 possesses >300 conotoxin homologs with CC-C-C-CC motif in cone snails [265, 266]. Conotoxins are small cysteine-rich peptides secreted by the *Conus* to capture the prey, due to their high affinity and potency to the ion channels and receptors [141]. To investigate if wZ1 can target ion channels or receptors, bioinformatic analyses were conducted. The phylogenetic tree shows the predicted origination of wZ1 and its 100 conotoxin homologs, suggesting that wZ1 is distant from its conotoxin homologs phylogenetically and geographically. WZ1 was then compared with some conotoxins with known tertiary structure and ion-channel targeting functions. Although sharing similar biosynthetic architecture with conotoxins, wZ1 is not like them to have post modifications: hydroxylation of proline residue, cyclization of glutamic acid residues, and addition of NH_2 at the C terminus. Besides, we found that ion-channel targeting peptides are highly positive in charge and hydrophilic, while wZ1 is less positive in charge and hydrophobic. We also showed that the backbone and surface properties of GIIIA are significantly different from those of wZ1. Based on the points above, we speculate that wZ1 might not target the ion channel or receptors specifically, but be likely to cross the cell membrane for its biological functions.

Cellular uptake of peptides can be spontaneous entering and/or endocytosis [308]. The cellular uptake of AF488-wZ1 is in a temperature-dependent manner without damage of the cell membrane, indicating its cell entering might be through endocytosis. The endocytosis mechanisms can be classified into two types: 1) the clathrin-dependent mechanism; 2) clathrin-independent mechanisms, such as caveolae/lipid-mediated endocytosis and macropinocytosis [309]. To confirm the entering mechanisms of AF488-wZ1, we used dynasore to inhibit clathrin-dependent endocytosis and applied nystatin and EIPA to inhibit the caveolae/lipid-mediated endocytosis and the macropinocytosis, respectively. Our result shows that the nystatin and dynasore treatments both decreased the cellular uptake of AF488-wZ1, suggesting its cell entering is through multiple endocytosis mechanisms.

Under confocal microscopy, we observed that AF488-wZ1 entered the nucleus of human HepG2 cells. This observation was confirmed by biotin-wZ1 localization in nucleus fractionation. By transcriptomic analysis, we found that wZ1 downregulated cholesterol biosynthetic genes that can lead to low cholesterol level in hepatocytes. Decreasing the cholesterol level in the liver is an important target for upregulating LDLR. For example, statins can inhibit the HMG-CoA reductase, a rate-limiting enzyme in the cholesterol biosynthetic pathway. This inhibiting finally upregulates LDLR expression on the hepatic cell membrane for uptake of LDL-C in circulation, leading to low LDL-C and low risks of cardiovascular diseases.

In 1974, the existence of LDLR was confirmed by Joseph L. Goldstein and Michael S. Brown using radiolabelled LDL [310]. In 1985, these two scientists are awarded The Nobel Prize in Physiology or Medicine, for their contribution to explaining the underlying mechanism of familial hypercholesterolemia. In 1987, the Merck company introduced lovastatin to upregulate LDLR for treating familial hypercholesterolemia.

The mechanism of lovastatin is to inhibit HMG-CoA reductase, a rate-limiting enzyme in the cholesterol biosynthetic pathway, and lower cholesterol in the liver then leads to upregulating gene expression of LDLR. Another class of medicine used to increase LDLR is PCSK9 inhibitors (e.g., alirocumab and evolocumab). They reduce degradation of and promote recycling of LDLR to increase LDL uptake [242, 311]. Unlike typical inducers of LDL uptake, wZ1 increases LDL uptake by spatially distributing LDLR into lipid rafts. Apart from having a different mechanism, wZ1 as a peptide locates in a chemical space between statins (small molecules) and PCSK9 inhibitors (monoclonal antibodies). WZ1 likely inherits both stable properties of small molecules and less off-target properties of monoclonal antibodies. In other words, wZ1 may represent a novel class of leads for drug research and development.

Since LDLR is traditionally known as a non-raft transmembrane protein [312], we hypothesize that wZ1 can alter lipid raft composition, leading to the formation of more rafts that can accommodate LDLR. The redistribution of LDLR into lipid rafts increases LDLR membrane density and enhances the efficiency of LDL uptake. Lipid rafts are plasma membrane microdomains enriched in sphingolipid and cholesterol that are involved in the lateral compartmentalization of molecules at the cell surface [311, 313]. One plausible explanation as to why wZ1 promotes lipid raft formation may reside in its primary sequence. wZ1 has two features that could interact with lipids, including the presence of a reversed cholesterol recognition motif, $(L/V(X)_{1-5}Y(X)_{1-5}R/K)$, and a tryptophan residue that may be inserted into the bilayer to increase the hydrophobicity of the local environment [314]. A combination of these features may provide wZ1 with the ability to interact with lipid components in the cell membrane, alter membrane lipid fluidity, and trigger lipid raft formation. Thus, our results open a new avenue for lipid

research to further understand this putative atypical mechanism: raft-enhanced LDL uptake after peptide-inducing raft formation.

In conclusion, we discovered wuweizitide wZ1, a hyperdisulfide-constrained peptide from *S. chinensis*, with a function of inducing LDLR-mediated LDL uptake at lipid rafts. wZ1 is encoded by a precursor with a three-domain architecture. The mature wZ1 is a cell-membrane permeable peptide that can promote lipid raft formation. By redistributing LDLR to lipid rafts, wZ1 increases LDL uptake in Hepatoma G2 cells. Our findings indicate that hyperdisulfide peptides are highly underexplored natural products and represent a novel class of plant-derived peptide leads for orally-bioactive therapeutic development.

CHAPTER SIX

6.1 General discussion

To structurally and functionally investigate peptidyl phytochemicals from well-known medicinal plants, in this thesis, we report two cysteine-rich peptides (CRPs): wisotide (wS1) from *Withania somnifera* and wuweizitide (wZ1) from *Schisandra chinensis*. These two CRPs are identified, isolated, and structurally determined. Structurally, on one hand, wS1 belongs to a rarely reported lybatide CRP family and adopts a cystine-stable helical structure. On the other hand, wZ1 represents a novel class of CRPs with a unique cysteine pattern. For the orally-active potential, wS1 and wZ1 were tested to be both cell-membrane penetrable and structurally stable against enzymatic degradation. Interestingly, unlike wS1 with low cell-membrane permeable capacities, wZ1 can easily cross the cell membrane, enter the nucleus, and regulate gene expression, leading to upregulation of low-density-lipoprotein (LDL) receptors and then to increase LDL uptake at lipid rafts.

wS1 contains eight cysteine residues with two pairs of consecutive cystines at the C terminal of the primary sequence. According to our classification, wS1 belongs to lybatide family, in which only one member has been discovered and characterized so far [99]. In other words, wS1 is a second member of lybatide family. The structural characteristics of lybatides are their unusual cystine-stabled helices. Similarly, wS1 has this helical scaffold, indicating that wS1 might be evolutionarily and functionally close to lybatides. Other than its helicity, wS1 also shares physiochemical properties with lybatides, such as hydrophobicity and pI (isoelectric point), further supporting our speculation that these two CRPs are relevant in evolution and functions. Unlike wS1, wZ1 is structurally unique and rare *in planta*. Based on our search in databases (e.g.

NCBI, UniProt, and ONEKP), no plant homologs of wZ1 were found by us, suggesting that wZ1 is a novel CRP family and rarely found *in planta*.

Structural stability and cell-membrane permeable capacity of peptides are crucial for orally-active drug development, because they influence storage conditions, drug functional integrity, and administration ways. Due to three or four disulfide bonds, wS1 and wZ1 are super stable against proteolytic degradation, conditions simulating enzyme digestion in our stomach and intestines. Besides, cell-membrane permeable capacities of wS1 and wZ1 further demonstrate the drug-like potential of these two peptides. wS1 differs from wZ1 by hydrophobic properties, as wS1 is hydrophilic, while wZ1 is hydrophobic. This difference could explain why the cellular uptake of wS1 is not as good as the one of wZ1. Thus, we speculate that the main targets of wS1 are on the cell membrane, while the main targets of wZ1 are inside the cell.

After crossing the cell membrane, wZ1 was observed to be enriched in the nucleus under the confocal microscope. This nucleus enrichment was confirmed by Western blotting analysis after tagging wZ1 with the LC-biotin. As the nucleus is a place for storing our generic information, transcriptomic analyses after treating the HUVEC cells and HepG2 cells were performed. In HUVEC cells, wZ1 regulates genes related to inflammation, anti-oxidative stress, and apoptosis. These regulations are consistent with the adaptogenic role of *S. chinensis* in modern research **(Supplementary information)**. In HepG2 cells (the liver carcinoma), genes in cholesterol biosynthetic pathways are downregulated. Since inhibition of cholesterol biosynthesis is the key mechanism for lovastatin to upregulate LDLR, we measured the level of LDLR after wZ1 treatment and found that wZ1 upregulates LDLR and then increases LDL uptake of HepG2 cells. To the best of our knowledge, wZ1 is the first peptidyl phytochemical found in five-flavor berries possessing this regulation of LDLR so far.

In Indian Ayurvedic medicine, *W. somnifera* or Indian ginseng is a multi-functional medicinal plant and shows diverse pharmacological effects, like neuron protection, immunomodulation, anti-cancer, and cardiovascular protection [315]. In this plant, we identified wS1 that displays a well-defined (α + π) helical conformation, stabilized by four disulfides that are arranged in an integrated cystine-knot and symmetric motif. This stabilization distorts the α helix towards a π helix and introduces torsional strain, likely as a result of evolutional selection to serve critical biological functions, such as in protein-protein interactions. Inspired by nature, many chemically covalent strategies are also used to reinforce peptidyl helicity that in turn reduces exposure of polar amide groups and enhances membrane permeability of peptides [316, 317]. Currently, these stabilized helical peptides are a highly unmet need in biotechnological applications. Future work can be protein engineering or functional investigation using wS1.

In traditional Chinese medicine, *S. chinensis*, commonly known as five-flavor berries, possesses therapeutic-associated tastes: sour, sweet, bitter, astringent, and salty [188]. These tastes are believed to account for systematic regulation of *S. chinensis* on testicles, lung, stomach, liver, and heart [188]. Thus, this plant is used for treating sexual disorders, asthenia, diarrhoea, body weakness, diabetes, stress, and hyperlipidaemia [188, 196, 318]. One possible explanation for its multi-functional property is that different components of this plant (with unique tastes) exert different therapeutic functions. Currently known bioactive components are mainly dibenzocyclooctadiene lignans, a class of small molecules with effects like anti-stress [196]. Unlike these molecules, wZ1 is a peptide with a larger target-binding interface and functions as a LDLR upregulator in human HepG2 cells. This finding can provide molecular basics for the previous report that *S. chinensis* extraction lowers LDL-cholesterol in the blood of hyperlipidemic rats [319].

In summary, wS1 and wZ1 are novel CRPs from a highly underexplored chemical space of phytochemicals in medicinal plants, and these CRPs have unusual structures and huge pharmacological potential, because of their structural stabilities, sequence diversities, and naturally-occurring in well-known medical plants. The finding that wZ1 can cross the cell membrane and upregulate LDLR further enforced our belief in the remarkable therapeutic potential of CRPs from medicinal plants.

6.2 Overall conclusions

In this thesis, we hypothesized that hyperstable CRPs in medicinal plants represent a novel class of orally-bioactive compounds for therapeutic applications. Based on this hypothesis, we aimed to discover novel CRPs from medicinal plants and then to explore their biological functions.

In chapter four, we reported wisotide (wS1), a new CRP from the medicinal plant *W. somnifera* or Indian ginseng. wS1 was sequenced to be a CRP with 31 amino acid residues, including eight cysteine residues and two negative residues (D and E). This sequence result was confirmed by its mRNA precursors, which show three architectural domains: the signal domain, the mature domain, and the C tail. From NMR structural analysis, wS1 adopts a cystine-stabled helical scaffold with structurally super stability. In addition, phylogenetic trees showed homologs of wS1 *in planta*. These homologs are distributed in different plants where only goji berry, similar to the Indian ginseng, belongs to medicinal plants that are rich in potential bioactive molecules. wS1 was then compared with lybatides, CRPs from goji berry in terms of sequences, properties, and structures. We found that wS1 is the new member to the lybatide family. Lastly, wS1 was chemically synthesized and found to be non-toxic and cell-membrane permeable, indicating its high potential for drug development.

In chapter five, we introduced *S. chinensis*-derived wZ1 that forms a novel CRP family with rare known members *in planta*. Structurally, wZ1 has 18 amino acid residues with eight cysteine residues, and it has knottin-type disulfide connectivity, leading to its high structural stability. By transcriptomic analysis, the precursor of wZ1 has three domains: the signal domain, the mature domain, and the C tail. Although having rare homologs *in planta*, wZ1 shares a cysteine pattern with some conotoxins that are known for their capacity of targeting ion-channel on the cell membrane. wZ1 was then compared with

its highly homologous conotoxins that are functionally characterized to be ion channel blocking. wZ1 differs from these conotoxins in geographical distribution and structural properties (e.g., hydrophobicity and net charges). These comparisons indicated that wZ1 might not target the proteins on the cell membrane, but could have intracellular localization. Our further investigation, with the chemical synthesis of wZ1, finally unveiled one biological function of wZ1, which is to upregulate LDL receptors and promote LDL uptake in HepG2 cells. This biological function was contributed by its capacities of entering the cell membrane, being enriched in the nucleus, and down-regulating cholesterol biosynthetic genes.

Taken together, we reported the discovery, synthesis, and characterization of two hyperstable CRPs: wS1 from *W. somnifera* and wZ1 from *S. chinensis*. Importantly, we found that wZ1 can cross the cell membrane, enter the nucleus, and upregulate LDLR in HepG2 cells. These findings show that CRPs as a novel class of phytochemicals, different from small molecules, could be used as orally-bioactive leads for future therapeutic development.

6.3 Future directions

In addition to the above-mentioned findings, we have some preliminary results:

1. After transcriptomic analysis of wS1-treated stem cells, Dr. Antony Kam and Dr. Shining Loo speculate that wS1 might be related to the treatment of atherosclerosis (see **SI 1**). Our preliminary results show that wS1 can decrease LDL uptake in RAW 264.7 cells (macrophage). This uptake in macrophages is a crucial step for atherosclerosis development, indicating that wS1 might have therapeutic potential in atherosclerosis.
2. In HUVEC-CS cells, wZ1 can upregulate 158 genes and downregulate 142 genes (see **SI 2**). These genes are enriched to take part in multiple cell signaling (e.g. autophagy, metabolic pathways, and cytokine-cytokine interaction; see **SI 2**) and to different disease-related hallmarks (e.g. inflammation, oxidative stress, and apoptosis; see **SI 2**). Interestingly, these analyses are consistent with the adaptogenic role of *S. chinensis*, the medicinal plant where wZ1 is derived.
3. By pull-down assay, we found that wZ1 can bind to 16 ribosomal proteins in HUVEC-CS (see **SI 3**), and that wZ1 might upregulate the expression of ribosomal RNA (see **SI 3**). In addition, under the confocal microscope, we observed that wZ1 can be enriched in the nucleolus in HUVEC-CS, which is the crucial place for ribosome biogenesis. Thus, we speculate that wZ1 can regulate ribosome biogenesis in HUVEC-CS.
4. wZ1 can increase MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) absorbance after treating HUVEC-CS and HepG2 cells. This increase can be due to either increase in cell proliferation or metabolic activities,

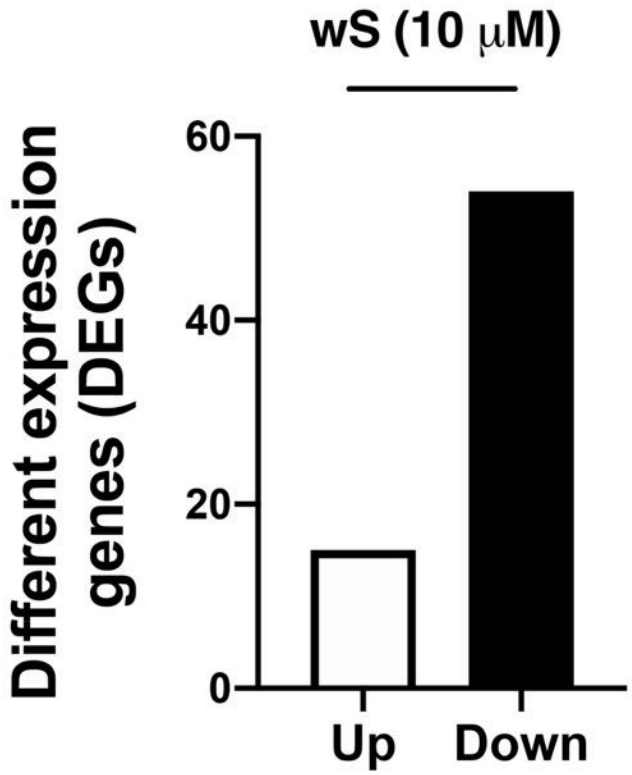
suggesting wZ1 might be used for angiogenesis, liver regeneration, and metabolic-related diseases.

Based on these preliminary results, my future work is as follows:

1. To investigate effects of wS1 on atherosclerosis by using a macrophage model.
2. To explore adaptogenic (or anti-stress) effects of wZ1 by using HUVEC cells.
3. To confirm effects of wZ1 on ribosomal biogenesis.
4. To test pharmacological effects of wZ1 has angiogenesis and liver protections.

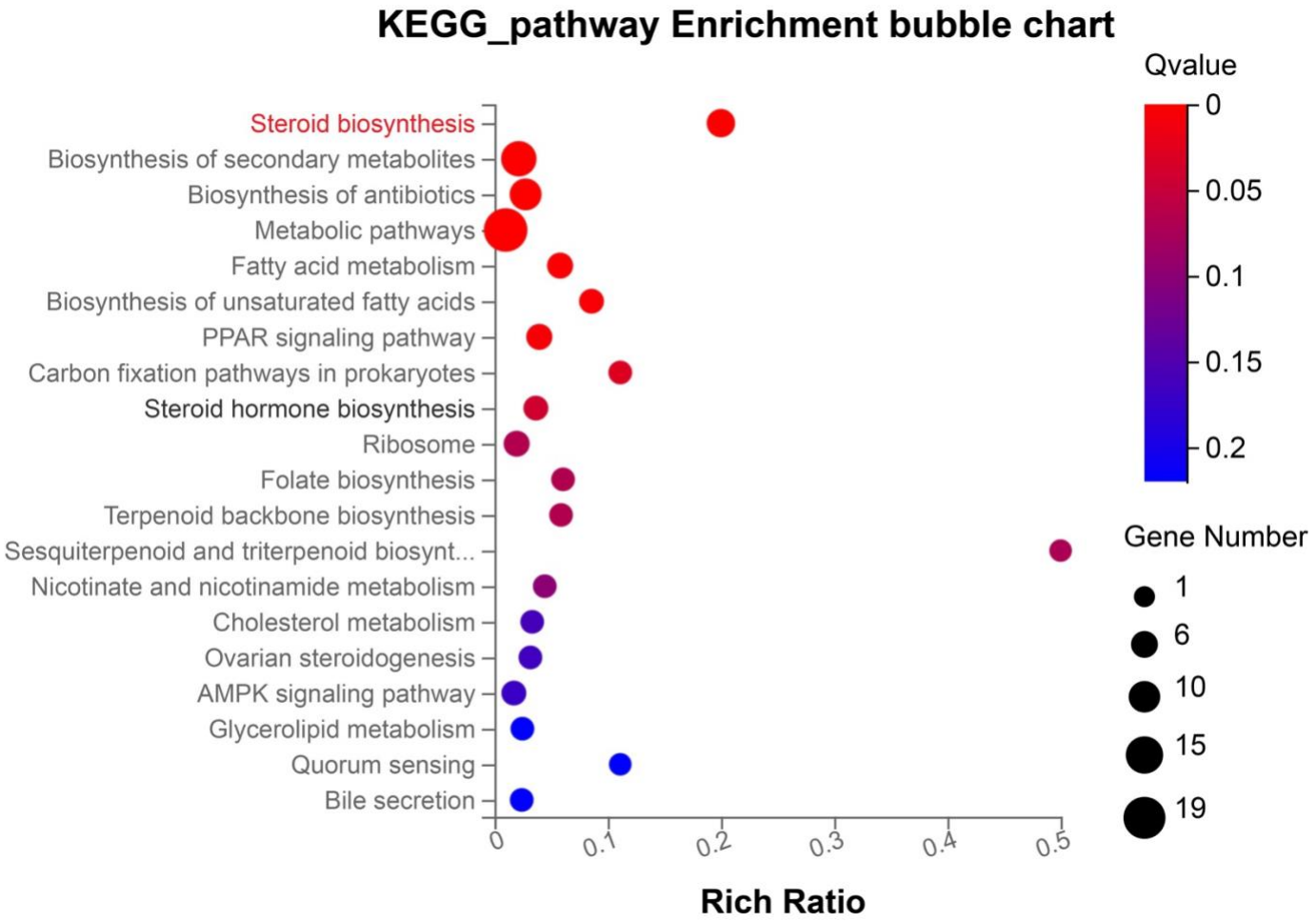
Supplementary information (SI)

SI 1. Transcriptomic analysis of wS1-treated stem cells



Supplementary figure 1. Gene expression is regulated by wS1 on mesenchymal stem cells.

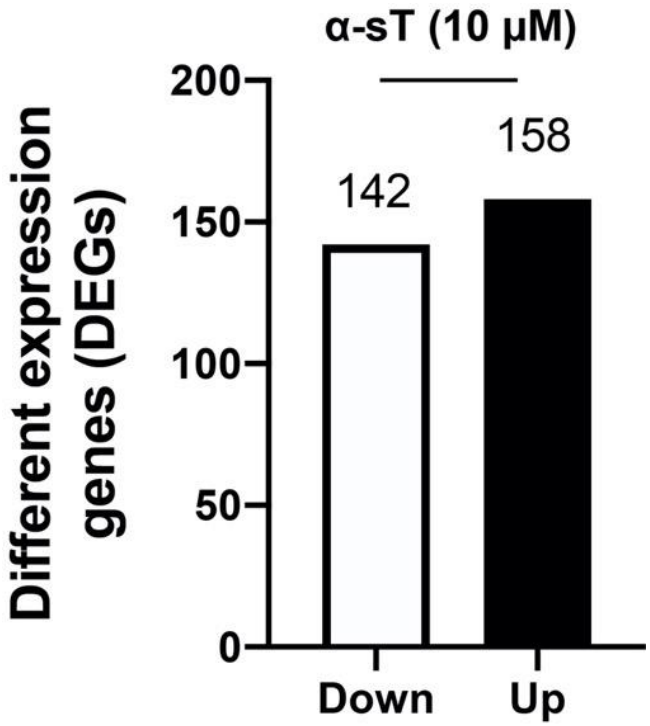
This experiment was performed by Dr. Antony Kam and Dr. Shining Loo.



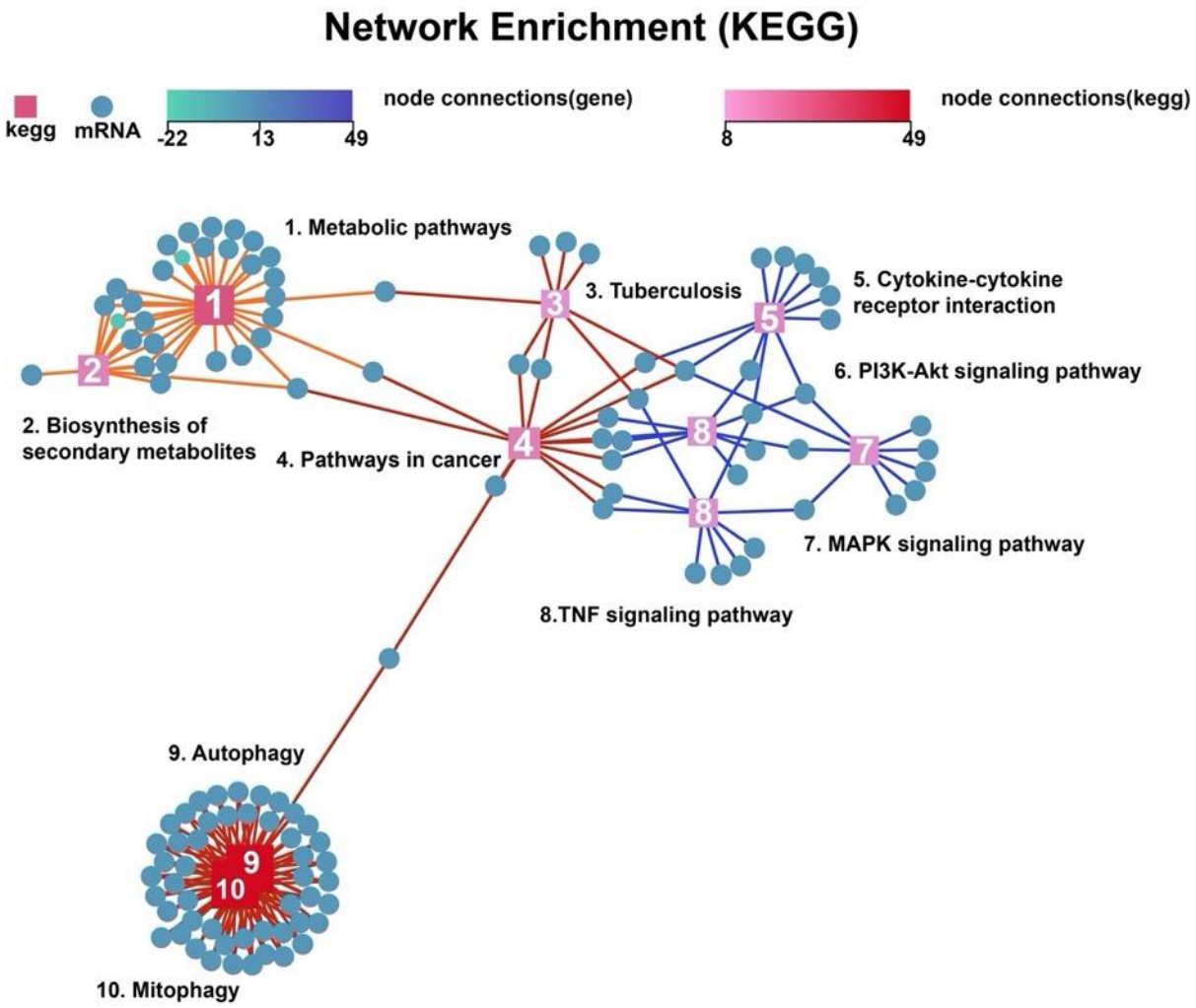
Supplementary figure 2. KEGG pathway enrichment of wS1 regulated DEGs.

This experiment was performed by Dr. Antony Kam and Dr. Shining Loo.

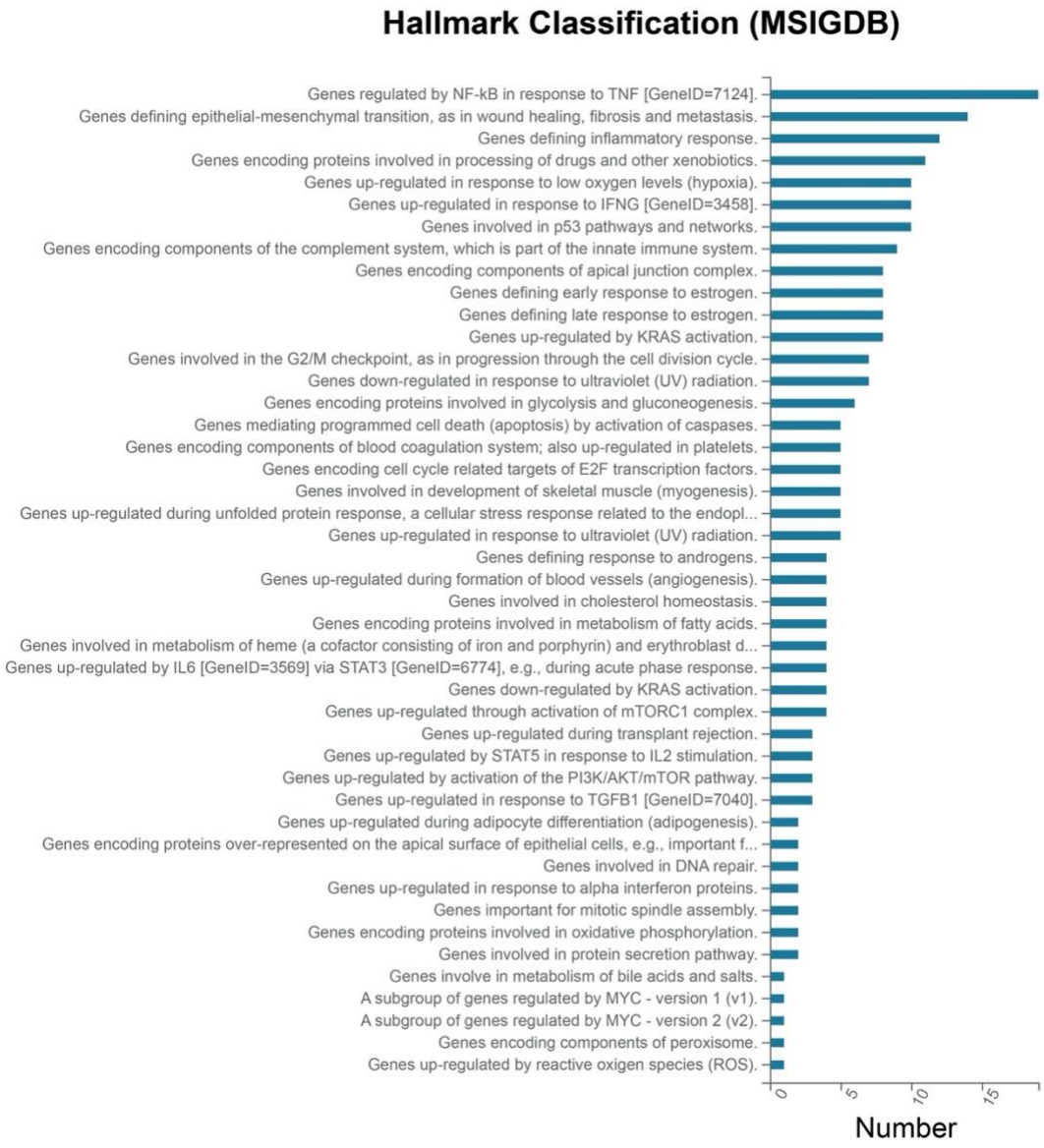
SI 2. Transcriptomic analysis of wZ1 on HUVEC-CS



Supplementary figure 3. Gene expression is regulated by wZ1.

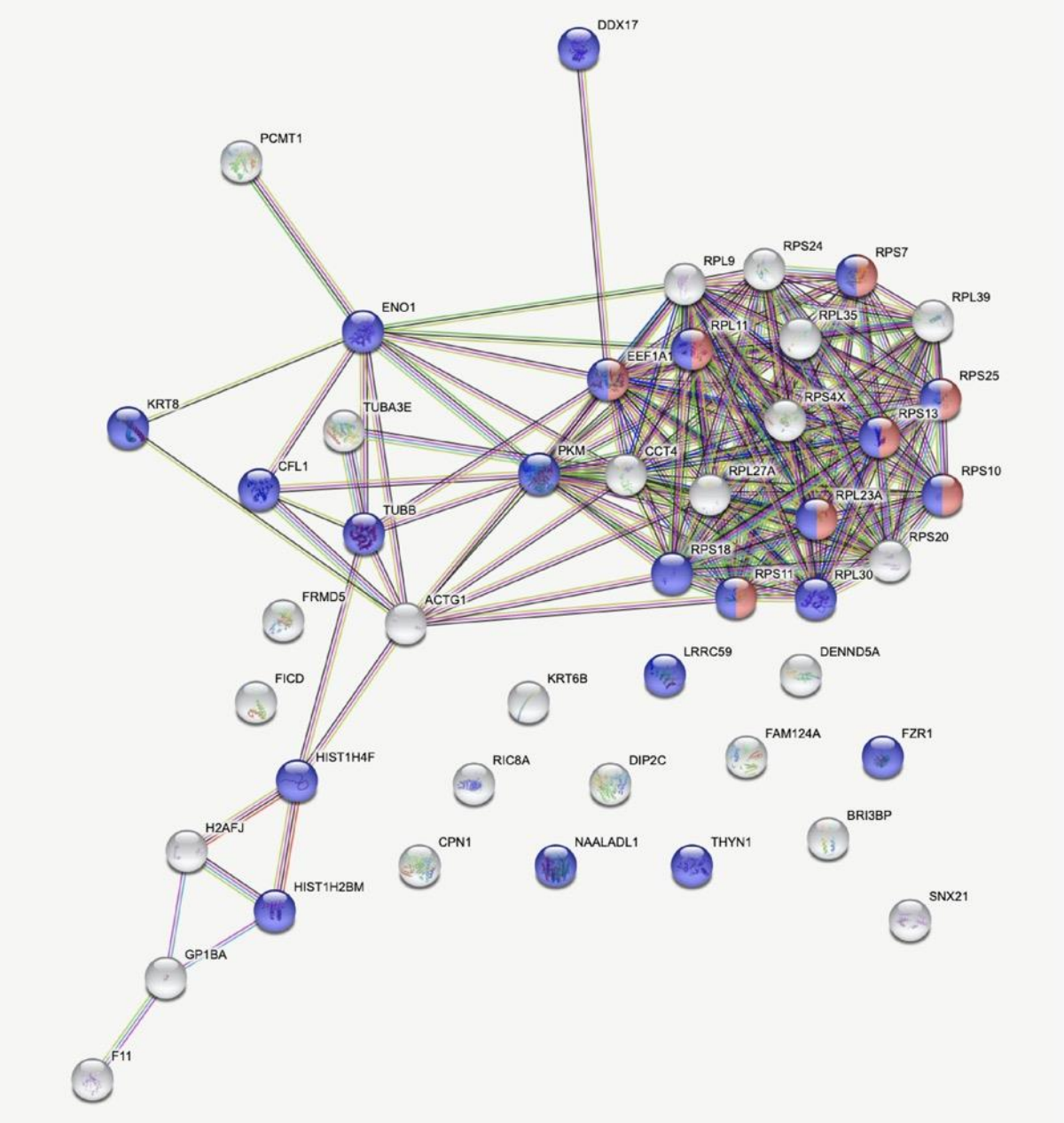


Supplementary figure 4. Cellular network enrichment of DEGs on wZ1 treated HUVEC-CS.



Supplementary figure 5. Hallmark enrichment of DEGs on wZ1 treated HUVEC-CS.

SI 3. Pull down results of wZ1 in HUVEC-CS



Supplementary figure 6. Pull down proteins clustering analysis.

Proteins in the nucleus are in blue and proteins in the nucleolus are in red. This experiment was performed with assistance from Dr. Antony Kam and Dr. Shining Loo.

Histone H2A.J
Putative trypsin-6
Histone H2B type 1-M
Disco-interacting protein 2 homolog C
Synembryn-A
Sorting nexin-21
60S ribosomal protein L39
Protein-L-isoaspartate(D-aspartate) O-methyltransferase
Coagulation factor XI
Thymocyte nuclear protein 1
T-complex protein 1 subunit delta
Aminopeptidase NAALADL1
Protein adenylyltransferase FICD
DENN domain-containing protein 5A
60S ribosomal protein L27a
Fizzy-related protein homolog
40S ribosomal protein S10
Carboxypeptidase N catalytic chain
Hemoglobin subunit gamma-1
Protein FAM124A
40S ribosomal protein S24
Tubulin alpha-3E chain
60S ribosomal protein L11
40S ribosomal protein S25
Leucine-rich repeat-containing protein 59
40S ribosomal protein S7
Cofilin-1
40S ribosomal protein S11
60S ribosomal protein L23a
60S ribosomal protein L30
60S ribosomal protein L9
Keratin, type II cytoskeletal 8
Keratin, type II cytoskeletal 6B
40S ribosomal protein S18
40S ribosomal protein S4, X isoform
Alpha-enolase
Pyruvate kinase PKM
60S ribosomal protein L35
Platelet glycoprotein Ib alpha chain
FERM domain-containing protein 5
Histone H4
Tubulin beta chain
40S ribosomal protein S13
40S ribosomal protein S20
Probable ATP-dependent RNA helicase DDX17
BRI3-binding protein
Actin, cytoplasmic 2
Elongation factor 1-alpha 1

Supplementary figure 7. Pull down protein list.

Ribosomal proteins are in yellow and histone subunits are in red. This experiment was performed with assistance from Dr. Antony Kam and Dr. Shining Loo.

SI 4. qPCR primers

Supplementary table 1. Primers of cholesterol biosynthetic gens in the qPCR experiment.

Genes	Forward primers	Reverse primers
ACAT2	TGGTGCCTTAGCTGCTGTTCCT	GGCTTGTCTAACAGGATTCTGCC
HMGCS1	AAGTCACACAAGATGCTACACCG	TCAGCGAAGACATCTGGTGCCA
HMGCR	GACGTGAACCTATGCTGGTCAG	GGTATCTGTTTCAGCCACTAAGG
MVK	GGAAAGTGGACCTCAGCTTACC	GCTTCTCCACTTGCTCTGAGGT
FDFT1	TGTGACCTCTGAACAGGAGTGG	GCCCATAGAGTTGGCACGTTCT
SQLE	CTCCAAGTTCAGGAAAAGCCTGG	GAGAACTGGACTCGGGTTAGCT
LSS	GACGACCGATTCACCAAGAGCA	AGACATGCTCCTGGAAGGCAGT
CYP51A1	CTCTTACCAGGTTGGCTGCCTT	CTTGAGACTGTCTGCGTTTCTGG
MSMO1	GCTGCCTTTGATTTGTGGAACCT	CTGCACAACCAAAGCATCTTGCC
TM7SF2	GGTCAATGGCTTCCAGTTGCTC	AACGCCAGCATGAAGCCAAACC

SI 5. Data mining of peptide homologs

Supplementary table2. Peptide precursors of wS1 homologs in the phylogenetic tree.

Peptide ID	Precursors	Organisms	Family
gnl|mpdb|CNA0003265_75827	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAYQVCCINNCCPSF	Lonicera japonica	Caprifoliaceae
gnl|onekp|RWXW_scaffold_2060713	MLIVILMMMLVFIFMTVCANAESCSEYCSESCSYCDVRPLYEDCCINRCCPTFAQVSP	Sargassum horneri	Sargassaceae
gnl|onekp|BOLZ_scaffold_2166794	MMKKNTNAVVLVLLVIIMGVCCYTGNAASCSEYCSNSCSFCNGQPLYEVCCINNCCPSLGRSIFSVLRLNFV	Atropa belladonna	Solanaceae
gnl|onekp|JNVS_scaffold_2008302	MAAMKLQATLIVVLMMITLMFAISPSCYAADGDSCTDHCSKSCDFCNGQPQYDVCCINNCCPSFRRAGRHWLN	Datura metel	Solanaceae
gnl|onekp|JNVS_scaffold_2008301	MAAMKLQATLIVVLMMITLMFAISPSCYAADGDSCTDHCSKSCDFCNGQPQYDVCCINNCCPSFRLSIIPNFLRI	Datura metel	Solanaceae
gnl|onekp|OSMU_scaffold_2070462	MMKKNSNALVVVLLLVLIMGVCCYTGNAESCSEYCSNSCPFCNGQPQYEVCCINNCCPSMALGQSLLNALRRHV	Lycium_sp	Solanaceae

gnl|onekp|OSMU_scaffold_2073250	MGAMKLKGTIAVAFIIMMVTLMLAISRCNAEDGDSCTDHCGISCDFCNGQPQYDVCCINNCCPSFRTSLRSNFLIRI	Lycium_sp	Solanaceae
gnl|onekp|UGJI_scaffold_2024651	MASMKLRGAIVVFMIVMISLLLSISASHAASCTEFCPKSCRYCDKYPVYNDCCFKSCCPKRFLSNFLSV	Lycopersicon_cheesmanii	Solanaceae
EB426081.1	MAAMKLNPTISVVFVMVTLMVLMLAISSNAADGDSCTDHCAISCAFCNGKPLYDICCINNCCPSFRASILSNYLRI	Nicotiana tabacum	Solanaceae
EB428452.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB428089.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGQPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB425521.1	MAAMKLNLTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB427158.1	MAAMRLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRASILSTYLRI	Nicotiana tabacum	Solanaceae
EB425077.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRV	Nicotiana tabacum	Solanaceae

EB428702.1	MMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB428992.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSVLSTYLRI	Nicotiana tabacum	Solanaceae
EB428567.1	MAAMKLNPTNAVVFMMVTLMVLVLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB428469.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFHTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB428002.1	MAAMKLNPTNAVVFMMVTLMILMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB426981.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCTNNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB427372.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCALCNGKPQYNVCCINNCCPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae
EB426866.1	MAAMKLNPTNAVVFMMVTLMVLMLAISSYAADGDSCTDHCAISCAFCNGKPQYNVCCINNCYPSFRTSILSTYLRI	Nicotiana tabacum	Solanaceae

JZ968628.1	MVPLNNHVLVIIWMIIVLMVVCFYTGNAPSRSEFCSNSCSYCNVRPLYAD CCINRCCPSLGESLLNILRPN	Nicotiana tomentosiformis	Solanaceae
gnl|onekp|MKZR_scaffold_2107092	MVKKNNNVLVIIWMIIVLMVVCFYTGNAESCSEFCSNSCSYCNVRPLYAD CCINRCCPSLGESLLNILRPN	Nicotiana_sylvestris	Solanaceae
FG395004.1	MVKKNNNVLVVIWMIIVLMAVYFYTGNAESCSEFCFNSCSFCNVRPLYE DCCIKRCCPSLAESLLNILRPN	Phyllostachys edulis	Solanaceae
FS199214.1	MMKKNNNMLIVILMMMLVFIFMTVCANAESCSEYCWESCSYCDVRPLY EDCCINRCCPTFAQVSP	Solanum lycopersicum	Solanaceae
FS028449.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS007677.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFL	Solanum melongena	Solanaceae
FS031217.1	MASMKLQGTIVVLMMVMVTLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS046344.1	MASMKLQGAIVVLMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae

FS005280.1	MASMKSQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS017320.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILFNFLRIQLNGSRGR	Solanum melongena	Solanaceae
FS031014.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRIQLNGSRGR	Solanum melongena	Solanaceae
FS058441.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILFNFLRI	Solanum melongena	Solanaceae
FS042132.1	MMVMITLVLSISPTYAASCSEYCANSCGFCDSAPAYQVCCINNCCPSFR VGILSNFLRI	Solanum melongena	Solanaceae
FS096539.1	MASMKLQGAIVVFMMVMITLMSSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS043857.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRIKLNGSRGR	Solanum melongena	Solanaceae
FS042131.1	MASMKLQGAIVVFMMVMITLVLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae

FS006877.1	MASMKLQGAIVVFMMVMIALMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS024697.1	MASMKLQGAVVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPA YQVCCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS042756.1	MITLMLSISPTYAASCSEYCANSCGFCDSAPAYQVCCINNCCPSFRVGIL SNFLRI	Solanum melongena	Solanaceae
FS000794.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINDCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS018894.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINSCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS097128.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QACCINNCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS005299.1	MASMKLQGAIVVFMMVMITLMLSISPTYAASCSEYCANSCGFCDSAPAY QVCCINNCCPPFRVGILSNFLRI	Solanum melongena	Solanaceae
FS034367.1	MASMKLQGVIVVLMMVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA VCCINGCCPSFRVGILSN	Solanum melongena	Solanaceae

FS034181.1	MASMKLQGVIVVLMMVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA VCCINGCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS019706.1	MASMKLQGVIVVLMMVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA VCCINGCCPSFRVGILFNFLRI	Solanum melongena	Solanaceae
FS032373.1	MASMKLQGVIVVLMMVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA ICCINGCCPSFRVGILSNFLHI	Solanum melongena	Solanaceae
FS032754.1	MASMKLQGVIVVLMVVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA VCCINGCCPSFRVGILSNFLRI	Solanum melongena	Solanaceae
FS037066.1	MASMKLQGVIVVLMMVMITLLLSISPTYAASCSEFCYTTCTFCDGQPEYA VWCINGCCPSFRVGILFNFLRI	Solanum melongena	Solanaceae
gnl|onekp|GHLP_scaffold_2049873	MMKKNNNVLVVILMMLVFMTVCFYTANGESCSEYCSNSCSYCDVRPLY EDCCINRCCPTLSQSLLNVLHPN	Solanum_dulcamara	Solanaceae
gnl|onekp|GHLP_scaffold_2051440	MGSMKLQATIVVLMMTLMLAISPTYAVNGDSCTDYCGVSCSFCNGKPLY QICCLNNCCPTFRVGILSNFLNI	Solanum_dulcamara	Solanaceae
wS1_1.1	MAAMKLKATIVVLMMVMITLMLSISPSYAADCTEYCSNSCPFCNGQPLY QLCCINNCCPSFRVGVLSNFLRT	Withania somnifera	Solanaceae

wS1_1.2	MAAMKLKATIAVLMMVMITLMLSISPSYAADCTEYCSNSCPFCNGQPLY QLCCINNCCPSFRVGVLSNFLRI	Withania somnifera	Solanaceae
Lyba1	DSCSEYCSNNSCPYCDGQKLYTLCCINTCCPS	Lycium barbarum	Solanaceae
Lyba2	DSCSEYCSNRCPSCDGQTQTQYTLCCINICCPS	Lycium barbarum	Solanaceae

Supplementary table 3. Conotoxin homologous precursors of wZ1 in the phylogenetic tree.

Peptide ID	Precursors	Organism region
P3.9	MMSKLGALLTICLLLFPITALLMDGDQPADRPAERMDYDISSEVHRLLERRHPPCCMYGRCRRYPGCSSASCCQGG	Eastern Pacific
reg3.14	MMSKLGVLLTICLLLFPLSVLPLDGDQPADQPAERMQDISAEQNPWFDPVKRCCNWPRCNVYLCGPCC	Eastern Pacific
reg3a	MMSKLRVLLTICLLLFPLSALPLDGDQPADQPAKRMWNGKLAARKPRFDKYDLVRGCCPPQWCGPDCTSPCCG	Eastern Pacific
PIIIE	MSKLGALLTICLLLFPITALLMDGDQPADRPAERMDYDISSEVHRLLERRHPPCCMYGRCRRYPGCSSASCCQRG	Eastern Pacific
PIIIA	MSKLGVLLTICLLLFPITALPMDGDQPADRLAERMQDNISSEEHPFEKR	Eastern Pacific
Bt3-I05	MLKLGMVLFIFLVLLPLATLHLDADQPVERNVNKQGLKPDERRRFILHALGQRQCCDWPWCDDCICCD	Indo-Pacific

Cp3-I01	MLKMGVMLFTFLVLFPLATLQLDADQPVERYAEDKQDLNRDERMGFILHALGQRQCCDWPW CDDCICCD	Indo-Pacific
Tx3f	MSKLGVLLTICLLLFPLTALPLDGDQPADQAAERMQAEQHPLFDQKRRCCKFPCPDSCRYLCC G	Indo-Pacific
Cp3-V08	MLKMGVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSALGQRRCCIWPEC GSCVCCL	Indo-Pacific
S3-L02	MLKMGVMLFIFLVLFPLATLHLDADQPVERYAENKQLLNTDERREIILSALRRQCCDSNSCEYP KCLCCNG	Indo-Pacific
Tx3-L02	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLSPDERREIILSALRRQCCDSNSCEYP KCLCCNG	Indo-Pacific
Vt3-L01	MLKMGVMLFIFLVLFPLATLQLDADQPVERYAENKQLLNTDERREIILSALRRQCCDSNSCEYP KCLCCNG	Indo-Pacific
Lt3.7	MMSKLGVLLTICLLLFPLTALPMDGDQPQERKEDGKSAALQPWFDPVKRCCQAACSPWPCLP CCR	Indo-Pacific
Im3.2	MMSKLGVLLAICLLMLPLTALPLDGDQPQERKEDGKSAALQPWFDPVKRCCQAACSPWLCLP CCG	Indo-Pacific

Bt3.3	MMSKLGVLLTICLLLFPLTALPMDGDQPQERKEDGKSAALQPWFDPVKR	Indo-Pacific
Ec3-GMAR01	MMSKLGVLLTICVLLFPLTAVPLDGDQPADQPAERTQNEQHPLFDQKRGCCRWPCPSICGMARCCSS	Indo-Pacific
S3-E03	MLKMGVVLFTFLVLFPLATLQLDADQPVERYVENKQDLNPDERSNFRLPLVRRCCSVSICQSPPVCECCA	Indo-Pacific
Vx3-HYQ01	MMSKLGCLLTICLVLFPLTALPLDGDQPAERPAKRTQDDIPNGQDPLIDRQINCCPWPCPDSCHYQCCH	Indo-Pacific
Vx3-L03	MLKMGVMLFIFLVLFPLATLHLDADQPVERYAENKQLLNTDERREVILSALRRQCCDRNSCEYPKCLCCNG	Indo-Pacific
Qc3.7	MSTLGVLLTICLLLFPLTALPLDGDQPADQSAERPAERTQDDIQQHPLYDPKRRCCRYPCPDSCHGSCCYK	Indo-Pacific
Tx3-L03	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLSPDERREIILSALRRQCCDRNSCEYPKCLCCNG	Indo-Pacific
Bt3-T01	MGVVLFTFLVLFPLATLQLDADQPVERYAENKQDHYPGESRGILKSALRKCCTMSVCQPPPVCTCCA	Indo-Pacific

Bt3.4	MMSKLGVLLTICLLLFPLTALPMDGDQPQERKEDEKSAALQPWFDPVKRCCQAGCSRYMCLP CCQ	Indo-Pacific
Qc3.6	MMSKLGVLLTICLVLFPLTALQLDGDQPADRPAERTQDISSEQYRKFDQRQRCCRWPCPGSC RCC	Indo-Pacific
MrIIIG	MMSKLGVLLTICLLLFALTAVPLDGDQPADRPAERMQDDISVPLDGDQPADRPAERMQDDISS ERHPMFDAVRDCCPLPACPFGCNPCCG	Indo-Pacific
Mr3.10	MMSKLGVLLTICLLLFPLTALPMDGDQPADRPAERMQNDISSERHPFFDRSKQCCHLAACRFR CTPCCW	Indo-Pacific
Mr3.18	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAKNKQLFNPHKRRGIILRAPGKRCCHRNWCDH LCSCCGS	Indo-Pacific
Mr3.5	MSKLGVLLTICLLLFPLTALPLDGDQPADQRAERTQAEKHSLPDPRMGCCPFPCKTSCTTLCC G	Indo-Pacific
Qc3-IP01	MMSKLGVLLTICLLLFPLTAVQLDGDQPVDLPALRTQDFAPEHSPWFNPVKRCCSRHCWVCIP CCPNGS	Indo-Pacific
Lv3-IP02	MMSKLGVLLTICLLLFPLTAVQLDGDQPVDLPALRTQDFAPDGDQPVDLPALRTQDFAP	Indo-Pacific

Mr3.9	MMSKLGVLLTICLLLFPLTALPMDGDQPADRPAERMQDDISSERHPFFDRSEQCCHLAACRFG CTPCCW	Indo-Pacific
MrIIIB	MMSKLGVLLTICLLLFPLTAVPLDGDQPADRPAERMQDDISSERHPFFDRSKQCCHLAACRFG CTPCCW	Indo-Pacific
Ec3-lP01	MMYKFGVLLTICLLLFPLTAVQLDGDQPVDLPALRTQDFAPEHSPWFNPVKRCCSRHCWVCIP CCPNGS	Indo-Pacific
Mr3.4	MSKLGVLLTICLLLFPLTAVPLDGDQPADRPAERMQDDISSERHPFFDRSKQCCHLPACRFGC TPCCW	Indo-Pacific
Qc3-TDG01	MLKMGVVLFTFLVLFPLATLQLDADQPVARYAENKQDFNPNERMKIMLSALRQRECCEPSWC DAGCTDGCC	Indo-Pacific
TsIIIA	MMSKLGVLLTICLLLFPLTAVPLDGDQPADQPAERKQNEQHPLFDQKRGCCRWPCPSRCGM ARCCSS	Indo-Pacific
Qc3-lP02	MMYKLGVLLTTCLLLFPLTAVQLDGDQPVDLPALRTQDFAPERSPWFDPVKRCCSRDCWVCI PCCPNGS	Indo-Pacific
Tx3h	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLNPDERREILLPALRKFCCDSNWCHIS DCECCYG	Indo-Pacific

Lv3-IP01	MMSKLGVLLTICLLLFPLTAVQLDGDQPVDLPALRTQDFAPEHSPWFDPVKRCCSRDCWVCIP CCPNGSA	Indo-Pacific
Mr3.8	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAKNKQLFNPHKRRGIILRAPGKRCCHWNWCDH LCSCCGS	Indo-Pacific
Mr3.16	MLKMGVVLFIVLVLFPLATLQLDADKPVERYAENKQLLNPDERRGIILHALGQRVCCSFGSCDS LCQCCDG	Indo-Pacific
Lv3-V03	MLKMGVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCCI WPECGSCVCCL	Indo-Pacific
Vt3.3	MMSKLGVLLTICLLLFPLTALPMDGDQPADQPAERLQDISPKETPGSDPFKRCCHWPYCAPPP LGCRCCGK	Indo-Pacific
TsMMSK -021	MMSKLGVLLTICLLLFPLTAVRLDGDQHTDRPADRMQDIATEQHPLFDPVKRCCDWPCTIGCV PCCLP	Indo-Pacific
Mr3.17	MLKMGVVLFIFLVLFTLPTLQLDADQPVERYAENKQLLNPDERRTIILHALTQPVCCPFGECKSL CYCCGA	Indo-Pacific
Mr064	MLKMGVVLFIVLVLFPLATLQLDADKPVERYAENKQLLNPDERRGIILHALGQRVCCSFGSCDN LCQCCDG	Indo-Pacific

Lv3-Y01	MLKMGVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRHRRCC FWPDCRGCYCCL	Indo-Pacific
Vc3.3	MLKMGVVLFTFLVLFPLATLQLNADRPVERYAANKQDLNPDETREMILHVLGQRLCCWSEMC HARCKCCG	Indo-Pacific
Tx3-E09	MMLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLNPDEKRGILLPALRRFCCDSNWCNI SDCECCYG	Indo-Pacific
Vc3.4	MLKMGVLLFTFLVLFPLATLQLDADRPVERYAANKQDLNPEERRKFILHALGQWQCCTMQWC DKACYCCE	Indo-Pacific
Mr013	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAKNKQLFNPHKRRGIILRAPGKRCCHWNWCDH LCLCCGS	Indo-Pacific
Mi3-IP02	MMSKLGVLLTICLLLFPITADPLVEDQPADRPADRMQDIATEQHPLFDPVKRCCDWPCGIGCIP CCLP	Indo-Pacific
Mr063	MLKMGVVLFIVLVLFPLATLQLDADKPVERYAENKQLLNPDERRGIILHALGQRVCCRKEWCH ARCTCCG	Indo-Pacific
Lv3-V02	MLKMGVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCC LWPECGGCVCCYL	Indo-Pacific

MrIIIE	MLKMGVVLFIVLVLFPLATLQLDADQPVERYAENKRLLNPDERRGIILHALGQRVCCPFGGCHE LCYCCDG	Indo-Pacific
Vt3-EP01	MMSKLGVLLTICLLLFPLTAVPLDGDQPADRPAERMQDGISSEHHPFFDSVKKKQQCCPPVAC NMGCEPCCG	Indo-Pacific
Mr065	MLKMGVVLFIVLVLFPLATLQLDADKPVERYAENKQLLNPDERRGIILHALGQRVCCSFGSCDS LCQCCDC	Indo-Pacific
Ec3-TYN01	MMSKLGALLTICLLLFPITGLPLDEDQPADLPALRMEDFAPEHSPLFDPVKRCCILCWKCTYNC CRAW	Indo-Pacific
Lv3-D01	MLKMRVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMDRQRRCC FWPMCGGCDCCYL	Indo-Pacific
Ts3.3	MMSKLGVLLTICLLLFPLTAVQLDGDQPADLPALRTQDIATDHSPWFDPVKRCCSRYCYICIPC CPN	Indo-Pacific
Lv3-D02	MLKMRVVLFTFLVLFPLATLQQDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCC FWPMCRGCDCCYL	Indo-Pacific
Tx3a	MLKMGVVLFIFLVLFPLATLQLDADQPVERYAENKQLLSPDERREIILHALGTRCCSWDVCDHP SCTCCG	Indo-Pacific

Tx3-WP04	MMSKLGALLTICLLLFPLTAVPMDGDQPADQPAQRLQDRLPTEDHPLYDPVKRCCDDSECDYSCWPCCIFS	Indo-Pacific
Mr3.14	MLKMGVVLFIFLVLFTLPTLQLDADQLVERHVENKQLLNPDERRRIISDALGQRICCPQGGCHQLCQCCGC	Indo-Pacific
Ca3-WP01	MSKLGALLTICLLLFSLTAVPLDGDQHADQPAQRLQDRLPTEDHPLYDPVKRCCDDSECDYSCWPCCIFS	Indo-Pacific
Mr3.15	MLKMGVVLFIFLVLFTLPTLQLDADQLVERHAENKQLLNPDERRRIISVALGQRVCCPHGGCHQICQCCGC	Indo-Pacific
Ge3.3	MSKLGVVLFTILVLLPLATLLLEADQPVERQQDLNPQRGTRGIMKHVMSKGMSRRGCCTGQGCWNVPICECCV	Indo-Pacific
Vc3.9	MMSKLGALLTICLLLFSLTAVPLDGDQHADQPAERLHDRLPTENHPLYDPVKRCCDDSECDYNCWPCCIFG	Indo-Pacific
LtIIIA	MLKMGVLLFTFLVLFPLTTLELDTDRPVERHAAIKQDLKPQERRGIRLHAPRDECCEPQWCDGACDCCS	Indo-Pacific
Mi3-V01	MLKMGVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCCLWPECGGCVCCY	Indo-Pacific

ArMMSK-01	MMSKLGVLLTICMLLFPLTALPLDGDQPADRPAERMQDDFISEQHPLFNPIKRCCDWPCTIGC VPCCK	Indo-Pacific
Lv3-V07	MLKMRVVLFTFLVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCC LWPECGGCVCCY	Indo-Pacific
Mr3.3	MSRLGVLLTICLLLFPLTAVPLDGDQPADRPAERLQDDISSEHHPHFDSGRECCGSFACRFGC VPCCV	Indo-Pacific
Vx3-G02	MLKMGVVLFVFLVLFPLATFQLDADQPVKRYAKYKQDLNTDKRMRLILPALRQRQCCGWEWC DDICGCCE	Indo-Pacific
Ts3.6	MLKMGVVLFVFLVLFPLATLQLDADQPVERYAENKQLVSPYERRQIILHALGQRQCCDWQWC DGACDCCA	Indo-Pacific
Bt3-D05	MLKLGMVLFIFLVLFPLATLHLDADQPVERNVNKQGLKPDERRRFGFSAPRKRECCEWEWCD GACDCCN	Indo-Pacific
Cp3-G01	MLKMGVVLFVFLVLFPLATFQLDADQPVKRYAKYKQDLNADKRMRLILPALRQRQCCGWEWC DDICGCCE	Indo-Pacific
Mi3-QDML01	MMFKLGVLLTICLVLFPLTALPLDGDQPADRPAERMQDDTSAAQNPRVAFVERCCEADCPMC QDMLCCRK	Indo-Pacific

Ts3.4	MMSKLGVLLTICLLLFPLTVLPMDGDQPADLPALRTQDIATDQSPWFDPVKRCCSRYCWKCIP CCPY	Indo-Pacific
MrIIIF	MLKMGVVLFIVLVLFPLATLQLDADKPVERYAENKQLLNPDERRGIILHALGQRVCCPFGGCHE LCLCCDG	Indo-Pacific
Qc3-YDG01	MLKMGVVLFVFLVLFPLATLQLDADQPVARYAENKQDFNPNERMKTMLSALRQRGCCDPQW CDAGCYDGCC	Indo-Pacific
Am3.4	MMYKLGVLLIICLLLFPLTAVPQDGDQPADRPAERMQDDISFEHDRFFDPVKRCCKYGWTCW LGCSPCCG	Indo-Pacific
Vt3-SP01	MMSKLGVLLIICLLLFSLNAVPLDGDQPADRYAERMQDDISSEHPLFDAVRGCCHLLACRMGC SPCCW	Indo-Pacific
GIIIA	MSKLGVLLTICLLLFPLTALPMDGDEPANRPVERMQDNISSEQYPLFEKRRDCCTPPKKCKDR QCKPQRCCAGR	Indo-Pacific
SIIIA	MMSKLGVLLTVCPLLFPLTALPPDGDQPADRPAERMQDDISSDEHPLFDKRQNCCNGGCSSK WCRDHARCCGR	Indo-Pacific
GIIIB	MMSKLGVLLTICLLLFPLTALPMDGDEPANRPVERMQDNISSEQYPLFEKRRDCCTPPRKCKD RRCKPMKCCAGR	Indo-Pacific

CnIIIC	MSKLGVLLTICLLLFPLFALPLDGDQPADRPAERMQDDISSEKHPLFDKRQGCCNGPKGCSSK WCRDHARCCGRR	Indo-Pacific
BuIIIB	MMSKLGVLLTICLLLFPLFALPQDGDQPADRPAERMQDDISSEQNPLLEKRVGERCCKNGKR GCGRWCRDHSRCCGRR	Indo-Pacific
TIIIA	MMSKLGVLLTICLLLFPLTALPMDGDEPADRPAERMQDNISSEQHPLFEERHGCCKGPKGCS SRECRPQHCCGRR	Indo-Pacific
Eb3-H03	MMLKMGVVLFVFLVLFPLATLQLDADQPVERDAENKQDLHPDERTGFILPAMRRRCCRLRRC KTHCHCCVYRP	Indo-Pacific, Eastern Pacific
Eu3.4	MMSKLGVLLTICLLLFPLTALPMDGDQPQERKEDGKSAALQPWFDPVKRCCRAACSPWLCLP CCG	Unknown
Eu3.3	MMSKLGVLLTICLLLFPLTALPMDGDQPQERKEDGKSAALQPWFDPVKRCCQAACSPWLCLP CCG	Unknown
Vr3-IP03	MMSKLGVLLTICLLLFSLTAVQLDGDQPVDLPALRTQDFAPEHSPWFNPVKRCCSRRCWVCIP CCPNGS	Unknown
Fla3.4	MMSKLGVLLTISLLLFPLTAVQLDGEQPVDLLALRTQDFAPEQSPWFDPVKRCCSKYCWECTP CCPYSSG	Unknown

Eu3.2	MMSKLGVLLTICLLLFPLTAVQLDGDQPADLPALRAQDFAPERSPWFDPVKRCCSQDCWVCIP CCPN	Unknown
Vr3-TYN01	MMSKLGVLLTICLLLFPITGLPLDEDQPADLPALRMEDFAPEHSPLFDPVKRCCILCWKCTYNC CRAW	Unknown
Vr3-IP01	MMSKLGVLLTICLLLFPITGLPLDEDQHADLPALRAQAFEPEHSPWFDPVRRCCSQDCWECIP CCPN	Unknown
Fla3.2	MLKMRVVLFTFRVLFPLATLQLDADQPRARYAENKQDFNRNERTKMILSAVSASMGRQRRCC LWPECGGCVCCY	Unknown
Vr3-D01	MLKLGMVLFIFLVLFPLATLHLDADQPVERNVNKQGLKPNERRFRFSAPRKRECCEWEWCDG ACDCCN	Unknown
Vr3-SP04	MSKLGALFVICLLLFPLTAVPLDGDQPADQPAQRMQDDISSEHPLFDAVRGCCHLLACRMGCS PCCW	Unknown

References

1. Group, T.F., *Dictionary of Natural Products*. 2018.
2. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs from 1981 to 2014.* Journal of natural products, 2016. **79**(3): p. 629-661.
3. Mishra, B.B. and V.K. Tiwari, *Natural products: an evolving role in future drug discovery.* European journal of medicinal chemistry, 2011. **46**(10): p. 4769-4807.
4. Tobert, J.A., *Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors.* Nature reviews Drug discovery, 2003. **2**(7): p. 517-526.
5. FLEMING, A., *THE DISCOVERY OF PENICILLIN.* British Medical Bulletin, 1944. **2**(1): p. 4-5.
6. Ehrlich, J., et al., *Chloromycetin, a new antibiotic from a soil actinomycete.* Science, 1947. **106**(2757): p. 417-417.
7. Waksman, S.A., H.C. Reilly, and D.B. Johnstone, *Isolation of streptomycin-producing strains of Streptomyces griseus.* Journal of bacteriology, 1946. **52**(3): p. 393-397.
8. Newton, G. and E. Abraham, *Cephalosporin C, a new antibiotic containing sulphur and D-α-aminoadipic acid.* Nature, 1955. **175**(4456): p. 548-548.
9. Geraci, J., et al. *Some laboratory and clinical experiences with a new antibiotic, vancomycin*. in *Proc. Staff, Meetings Mayo Clinic*. 1956.
10. Patridge, E., et al., *An analysis of FDA-approved drugs: natural products and their derivatives.* Drug discovery today, 2016. **21**(2): p. 204-207.
11. Matsumura, E., et al., *Microbial production of novel sulphated alkaloids for drug discovery.* Scientific reports, 2018. **8**(1): p. 1-10.
12. Mora, C., et al., *How many species are there on Earth and in the ocean?* PLoS biology, 2011. **9**(8): p. e1001127.
13. Boeuf, G., *Marine biodiversity characteristics.* Comptes rendus biologies, 2011. **334**(5-6): p. 435-440.
14. Borchardt, J.K., *The Beginnings of Drug Therapy: Ancient Mesopotamian Medicine.* Drug News Perspect, 2002. **15**(3): p. 187-192.
15. Löwenberg, B., et al., *Cytarabine dose for acute myeloid leukemia.* New England Journal of Medicine, 2011. **364**(11): p. 1027-1036.
16. El-Subbagh, H.I. and A.A. Al-Badr, *Cytarabine*, in *Profiles of Drug Substances, Excipients and Related Methodology*. 2009, Elsevier. p. 37-113.
17. Schmidtko, A., et al., *Ziconotide for treatment of severe chronic pain.* The Lancet, 2010. **375**(9725): p. 1569-1577.
18. Mohamed Abd El-Aziz, T., A.G. Soares, and J.D. Stockand, *Snake venoms in drug discovery: valuable therapeutic tools for life saving.* Toxins, 2019. **11**(10): p. 564.
19. Vidt, D.G., E.L. Bravo, and F.M. Fouad, *Captopril.* New England Journal of Medicine, 1982. **306**(4): p. 214-219.
20. Cook, J.J., et al., *Tirofiban (Aggrastat®).* Cardiovascular Drug Reviews, 1999. **17**(3): p. 199-224.
21. Rasmussen, S. and S. Husted, *Tirofiban (Aggrastat). A non-peptide glycoprotein IIb/IIIa receptor inhibitor.* Ugeskrift for laeger, 2001. **163**(4): p. 461-465.
22. Brener, S.J., et al., *Eptifibatide and low-dose tissue plasminogen activator in acute myocardial infarction: the integrilin and low-dose thrombolysis in acute*

myocardial infarction (INTRO AMI) trial. Journal of the American College of Cardiology, 2002. **39**(3): p. 377-386.
23. Phillips, D.R., et al., *Effect of Ca2+ on GP IIb-IIIa interactions with integrilin: enhanced GP IIb-IIIa binding and inhibition of platelet aggregation by reductions in the concentration of ionized calcium in plasma anticoagulated with citrate.* Circulation, 1997. **96**(5): p. 1488-1494.
24. Borchardt, J.K., *The Beginnings of drug therapy: ancient mesopotamian medicine.* Drug news & perspectives, 2002. **15**(3): p. 187-192.
25. Nunn, J.F., *Ancient egyptian medicine*. 2002: University of Oklahoma Press.
26. Society, S.G.A., *Charaka Samhita (English translation)*. 1949.
27. Kapoor, L., *CRC handbook of Ayurvedic medicinal plants*. 2018: CRC press.
28. Dev, S., *Ancient-modern concordance in Ayurvedic plants: some examples.* Environmental Health Perspectives, 1999. **107**(10): p. 783-789.
29. Pan, S.-Y., et al., *Historical perspective of traditional indigenous medical practices: the current renaissance and conservation of herbal resources.* Evidence-Based Complementary and Alternative Medicine, 2014. **2014**.
30. Sneader, W., *Drug discovery: a history*. 2005: John Wiley & Sons.
31. Cragg, G.M. and D.J. Newman, *Natural products: a continuing source of novel drug leads.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2013. **1830**(6): p. 3670-3695.
32. Serturner, F., *Ueber das Morphium, eine neue salifahige Grundlage, und die Mekonsaure als Hauptbestandtheile des Opiums.* Annalen der Physik, 1817. **25**: p. 56-89.
33. Cragg, G.M., et al., *The impact of the United Nations Convention on Biological Diversity on natural products research.* Natural product reports, 2012. **29**(12): p. 1407-1423.
34. Silverstein, K.A., et al., *Small cysteine‐rich peptides resembling antimicrobial peptides have been under‐predicted in plants.* The Plant Journal, 2007. **51**(2): p. 262-280.
35. Tam, J.P., et al., *Antimicrobial peptides from plants.* Pharmaceuticals, 2015. **8**(4): p. 711-757.
36. Scheres, B., et al., *The ENOD12 gene product is involved in the infection process during the pea-Rhizobium interaction.* Cell, 1990. **60**(2): p. 281-294.
37. Scheres, B., et al., *Sequential induction of nodulin gene expression in the developing pea nodule.* The Plant Cell, 1990. **2**(8): p. 687-700.
38. Pearce, G., et al., *Production of multiple plant hormones from a single polyprotein precursor.* Nature, 2001. **411**(6839): p. 817-820.
39. Mingossi, F.B., et al., *SacRALF1, a peptide signal from the grass sugarcane (Saccharum spp.), is potentially involved in the regulation of tissue expansion.* Plant molecular biology, 2010. **73**(3): p. 271-281.
40. Olsen, A.N., J. Mundy, and K. Skriver, *Peptomics, identification of novel cationic Arabidopsis peptides with conserved sequence motifs.* In silico biology, 2002. **2**(4): p. 441-451.
41. Pearce, G., et al., *RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development.* Proceedings of the National Academy of Sciences, 2001. **98**(22): p. 12843-12847.
42. Scheer, J.M., G. Pearce, and C.A. Ryan, *Le RALF, a plant peptide that regulates root growth and development, specifically binds to 25 and 120 kDa cell surface membrane proteins of Lycopersicon peruvianum.* Planta, 2005. **221**(5): p. 667-674.

43. Scheer, J.M., G. Pearce, and C.A. Ryan, *LeRALF, a plant peptide that regulates root growth and development, specifically binds to 25 and 120 kDa cell surface membrane proteins of Lycopersicon peruvianum.* Planta, 2005. **221**(5): p. 667-674.
44. Shpak, E.D., et al., *Stomatal patterning and differentiation by synergistic interactions of receptor kinases.* Science, 2005. **309**(5732): p. 290-293.
45. Hara, K., et al., *The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule.* Genes & development, 2007. **21**(14): p. 1720-1725.
46. Abrash, E.B. and D.C. Bergmann, *Regional specification of stomatal production by the putative ligand CHALLAH.* Development, 2010. **137**(3): p. 447-455.
47. Marshall, E., L.M. Costa, and J. Gutierrez-Marcos, *Cysteine-rich peptides (CRPs) mediate diverse aspects of cell–cell communication in plant reproduction and development.* Journal of experimental botany, 2011. **62**(5): p. 1677-1686.
48. Tam, J.P., et al., *Ginsentides: cysteine and glycine-rich peptides from the ginseng family with unusual disulfide connectivity.* Scientific reports, 2018. **8**(1): p. 1-15.
49. OHTANI, K., et al., *Complete primary structures of two subunits of purothionin A, a lethal protein for brewer's yeast from wheat flour.* The Journal of Biochemistry, 1977. **82**(3): p. 753-767.
50. Bohlmann, H. and K. Apel, *Thionins.* Annual review of plant biology, 1991. **42**(1): p. 227-240.
51. Dang, L. and E.J. Van Damme, *Toxic proteins in plants.* Phytochemistry, 2015. **117**: p. 51-64.
52. Epand, R.M., *Host defense peptides and their potential as therapeutic agents*. 2016: Springer.
53. Padovan, L., M. Scocchi, and A. Tossi, *Structural aspects of plant antimicrobial peptides.* Current Protein and Peptide Science, 2010. **11**(3): p. 210-219.
54. Stec, B., *Plant thionins–the structural perspective.* Cellular and Molecular Life Sciences CMLS, 2006. **63**(12): p. 1370-1385.
55. De Caleya, R.F., et al., *Susceptibility of phytopathogenic bacteria to wheat purothionins in vitro.* Applied microbiology, 1972. **23**(5): p. 998-1000.
56. Ebrahim-Nesbat, F., et al., *Cultivar-related differences in the distribution of cell-wall-bound thionins in compatible and incompatible interactions between barley and powdery mildew.* Planta, 1989. **179**(2): p. 203-210.
57. Evans, J., et al., *Cellular responses to Pyrularia thionin are mediated by Ca2+ influx and phospholipase A2 activation and are inhibited by thionin tyrosine iodination.* Proceedings of the National Academy of Sciences, 1989. **86**(15): p. 5849-5853.
58. Kramer, K.J., et al., *Toxicity of purothionin and its homologues to the tobacco hornworm, Manduca sexta (L.)(Lepidoptera: Sphingidae).* Toxicology and applied pharmacology, 1979. **48**(1): p. 179-183.
59. García-Olmedo, F., et al., *Plant defense peptides.* Peptide Science, 1998. **47**(6): p. 479-491.
60. Florack, D. and W. Stiekema, *Thionins: properties, possible biological roles and mechanisms of action.* Plant molecular biology, 1994. **26**(1): p. 25-37.

61. Carrasco, L., et al., *Thionins: plant peptides that modify membrane permeability in cultured mammalian cells.* European Journal of Biochemistry, 1981. **116**(1): p. 185-189.
62. Broekaert, W.F., et al., *Plant defensins: novel antimicrobial peptides as components of the host defense system.* Plant physiology, 1995. **108**(4): p. 1353.
63. Ganz, T., *Defensins: antimicrobial peptides of vertebrates.* Comptes rendus biologies, 2004. **327**(6): p. 539-549.
64. Islam, K.T., et al., *A novel bi-domain plant defensin MtDef5 with potent broad-spectrum antifungal activity binds to multiple phospholipids and forms oligomers.* Scientific reports, 2017. **7**(1): p. 1-13.
65. Van der Weerden, N.L. and M.A. Anderson, *Plant defensins: common fold, multiple functions.* Fungal Biology Reviews, 2013. **26**(4): p. 121-131.
66. de Oliveira Carvalho, A. and V.M. Gomes, *Plant defensins—prospects for the biological functions and biotechnological properties.* Peptides, 2009. **30**(5): p. 1007-1020.
67. Franco, O.L., *Peptide promiscuity: an evolutionary concept for plant defense.* FEBS letters, 2011. **585**(7): p. 995-1000.
68. Archer, B., *The proteins of Hevea brasiliensis latex. 4. Isolation and characterization of crystalline hevein.* Biochemical Journal, 1960. **75**(2): p. 236-240.
69. Van Parijs, J., et al., *Hevein: an antifungal protein from rubber-tree (Hevea brasiliensis) latex.* Planta, 1991. **183**(2): p. 258-264.
70. Koo, J.C., et al., *Two hevein homologs isolated from the seed of Pharbitis nil L. exhibit potent antifungal activity.* Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 1998. **1382**(1): p. 80-90.
71. Loo, S., et al., *Anti-Fungal Hevein-like Peptides Biosynthesized from Quinoa Cleavable Hololectins.* Molecules, 2021. **26**(19): p. 5909.
72. Broekaert, W.F., et al., *Antimicrobial peptides from Amaranthus caudatus seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins.* Biochemistry, 1992. **31**(17): p. 4308-4314.
73. Nielsen, K.K., et al., *Characterization of a new antifungal chitin-binding peptide from sugar beet leaves.* Plant Physiology, 1997. **113**(1): p. 83-91.
74. Kini, S.G., et al., *Studies on the chitin binding property of novel cysteine-rich peptides from Alternanthera sessilis.* Biochemistry, 2015. **54**(43): p. 6639-6649.
75. Wong, K.H., et al., *Ginkgotides: proline-rich hevein-like peptides from gymnosperm Ginkgo biloba.* Frontiers in plant science, 2016. **7**: p. 1639.
76. Fujimura, M., et al., *Purification, characterization, and sequencing of a novel type of antimicrobial peptides, Fa-AMP1 and Fa-AMP2, from seeds of buckwheat (Fagopyrum esculentum Moench.).* Bioscience, biotechnology, and biochemistry, 2003. **67**(8): p. 1636-1642.
77. Huang, R.-H., et al., *Two novel antifungal peptides distinct with a five-disulfide motif from the bark of Eucommia ulmoides Oliv.* FEBS letters, 2002. **521**(1-3): p. 87-90.
78. Odintsova, T.I., et al., *A novel antifungal hevein-type peptide from Triticum kiharae seeds with a unique 10-cysteine motif.* The FEBS journal, 2009. **276**(15): p. 4266-4275.
79. Van den Bergh, K., et al., *Ee-CBP, a hevein-type antimicrobial peptide from bark of the spindle tree (Euonymus europaeus L.).* Mededelingen

(Rijksuniversiteit te Gent. Fakulteit van de Landbouwkundige en Toegepaste Biologische Wetenschappen), 2002. **67**(2): p. 327-331.
80. Loo, S., et al., *Identification and characterization of roseltide, a knottin-type neutrophil elastase inhibitor derived from Hibiscus sabdariffa.* Scientific reports, 2016. **6**(1): p. 1-16.
81. Loo, S., et al., *Bleogens: Cactus-Derived Anti-Candida Cysteine-Rich Peptides with Three Different Precursor Arrangements.* Frontiers in Plant Science, 2017. **8**(2162).
82. Nguyen, G.K., et al., *Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis.* Nature chemical biology, 2014. **10**(9): p. 732-738.
83. Ireland, D.C., et al., *Discovery and characterization of a linear cyclotide from Viola odorata: implications for the processing of circular proteins.* Journal of molecular biology, 2006. **357**(5): p. 1522-1535.
84. Nguyen, G.K.T., et al., *Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry.* Journal of Biological Chemistry, 2011. **286**(52): p. 44833-44844.
85. Nguyen, P.Q.T., et al., *Antiviral cystine knot α-amylase inhibitors from Alstonia scholaris.* Journal of Biological Chemistry, 2015. **290**(52): p. 31138-31150.
86. Nguyen, P.Q., et al., *Discovery and characterization of pseudocyclic cystine-knot α-amylase inhibitors with high resistance to heat and proteolytic degradation.* The FEBS journal, 2014. **281**(19): p. 4351-4366.
87. Duvick, J., et al., *Purification and characterization of a novel antimicrobial peptide from maize (Zea mays L.) kernels.* Journal of Biological Chemistry, 1992. **267**(26): p. 18814-18820.
88. Marcus, J.P., et al., *A family of antimicrobial peptides is produced by processing of a 7S globulin protein in Macadamia integrifolia kernels.* The Plant Journal, 1999. **19**(6): p. 699-710.
89. Nolde, S.B., et al., *Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (Echinochloa crus-galli).* Journal of Biological Chemistry, 2011. **286**(28): p. 25145-25153.
90. de Oliveira Carvalho, A. and V.M. Gomes, *Role of plant lipid transfer proteins in plant cell physiology—a concise review.* Peptides, 2007. **28**(5): p. 1144-1153.
91. Sels, J., et al., *Plant pathogenesis-related (PR) proteins: a focus on PR peptides.* Plant Physiology and Biochemistry, 2008. **46**(11): p. 941-950.
92. Kader, J.-C., *Lipid-transfer proteins in plants.* Annual review of plant biology, 1996. **47**(1): p. 627-654.
93. Cammue, B.P.A., et al., *A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins.* Plant Physiology, 1995. **109**(2): p. 445-455.
94. Segura, A., et al., *Snakin-1, a peptide from potato that is active against plant pathogens.* Molecular Plant-Microbe Interactions, 1999. **12**(1): p. 16-23.
95. 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 physiology, 2002. **128**(3): p. 951-961.
96. Rogozhin, E., et al., *Characterization of hydroxyproline-containing hairpin-like antimicrobial peptide ecamp1-hyp from barnyard grass (Echinochloa crusgalli*

L.) seeds: structural identification and comparative analysis of antifungal activity. International journal of molecular sciences, 2018. **19**(11): p. 3449.
97. Hendrickson, W.A. and M.M. Teeter, *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur.* Nature, 1981. **290**(5802): p. 107-113.
98. Mak, A.S. and B.L. Jones, *The amino acid sequence of wheat β-purothionin.* Canadian Journal of Biochemistry, 1976. **54**(10): p. 835-842.
99. Tan, W.L., et al., *Lybatides from Lycium barbarum contain an unusual Cystine-stapled helical peptide scaffold.* Scientific reports, 2017. **7**(1): p. 1-11.
100. Owji, H., et al., *A comprehensive review of signal peptides: Structure, roles, and applications.* European journal of cell biology, 2018. **97**(6): p. 422-441.
101. Rusch, S.L. and D.A. Kendall, *Interactions that drive Sec-dependent bacterial protein transport.* Biochemistry, 2007. **46**(34): p. 9665-9673.
102. Luirink, J. and B. Dobberstein, *Mammalian and Escherichia coli signal recognition particles.* Molecular microbiology, 1994. **11**(1): p. 9-13.
103. Deshaies, R.J., et al., *Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex.* Nature, 1991. **349**(6312): p. 806-808.
104. Yuan, J., et al., *Protein transport across and into cell membranes in bacteria and archaea.* Cellular and Molecular Life Sciences, 2010. **67**(2): p. 179-199.
105. von Heijne, G., *The signal peptide.* The Journal of membrane biology, 1990. **115**(3): p. 195-201.
106. VAN VOORST, F. and B. De Kruijff, *Role of lipids in the translocation of proteins across membranes.* Biochemical Journal, 2000. **347**(3): p. 601-612.
107. Nilsson, I., et al., *The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence.* Journal of molecular biology, 2015. **427**(6): p. 1191-1201.
108. Berks, B.C., *The twin-arginine protein translocation pathway.* Annual review of biochemistry, 2015. **84**: p. 843-864.
109. Ng, D., J.D. Brown, and P. Walter, *Signal sequences specify the targeting route to the endoplasmic reticulum membrane.* The Journal of cell biology, 1996. **134**(2): p. 269-278.
110. De Gier, J.-W.L., et al., *Differential use of the signal recognition particle translocase targeting pathway for inner membrane protein assembly in Escherichia coli.* Proceedings of the National Academy of Sciences, 1998. **95**(25): p. 14646-14651.
111. Eder, J. and A.R. Fersht, *Pro-sequence-assisted protein folding.* Molecular microbiology, 1995. **16**(4): p. 609-614.
112. Hwang, C., A.J. Sinskey, and H.F. Lodish, *Oxidized redox state of glutathione in the endoplasmic reticulum.* Science, 1992. **257**(5076): p. 1496-1502.
113. Arnér, E.S., *Focus on mammalian thioredoxin reductases—important selenoproteins with versatile functions.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2009. **1790**(6): p. 495-526.
114. Krause, G. and A. Holmgren, *Substitution of the conserved tryptophan 31 in Escherichia coli thioredoxin by site-directed mutagenesis and structure-function analysis.* Journal of Biological Chemistry, 1991. **266**(7): p. 4056-4066.
115. Chambers, J.E., et al., *The reduction potential of the active site disulfides of human protein disulfide isomerase limits oxidation of the enzyme by Ero1α.* Journal of Biological Chemistry, 2010. **285**(38): p. 29200-29207.

116. Oka, O.B. and N.J. Bulleid, *Forming disulfides in the endoplasmic reticulum.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2013. **1833**(11): p. 2425-2429.
117. Onda, Y., *Oxidative protein-folding systems in plant cells.* International journal of cell biology, 2013. **2013**.
118. Stec, B., *Plant thionins--the structural perspective.* Cell Mol Life Sci, 2006. **63**(12): p. 1370-85.
119. Pelegrini, P.B. and O.L. Franco, *Plant γ-thionins: novel insights on the mechanism of action of a multi-functional class of defense proteins.* The international journal of biochemistry & cell biology, 2005. **37**(11): p. 2239-2253.
120. de Oliveira Dias, R. and O.L. Franco, *Cysteine-stabilized αβ defensins: from a common fold to antibacterial activity.* Peptides, 2015. **72**: p. 64-72.
121. Hollox, E.J. and R. Abujaber, *Evolution and diversity of defensins in vertebrates*, in *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*. 2017, Springer. p. 27-50.
122. Huang, R.-H., et al., *Solution structure of Eucommia antifungal peptide: a novel structural model distinct with a five-disulfide motif.* Biochemistry, 2004. **43**(20): p. 6005-6012.
123. Van den Bergh, K.P., et al., *Five disulfide bridges stabilize a hevein-type antimicrobial peptide from the bark of spindle tree (Euonymus europaeus L.).* FEBS letters, 2002. **530**(1-3): p. 181-185.
124. 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 Science, 1994. **3**(10): p. 1833-1839.
125. Craik, D.J., et al., *Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif.* Journal of molecular biology, 1999. **294**(5): p. 1327-1336.
126. Pons, J.-L., et al., *Refined solution structure of a liganded type 2 wheat nonspecific lipid transfer protein.* Journal of Biological Chemistry, 2003. **278**(16): p. 14249-14256.
127. Samuel, D., et al., *Solution structure of plant nonspecific lipid transfer protein-2 from rice (Oryza sativa).* Journal of Biological Chemistry, 2002. **277**(38): p. 35267-35273.
128. Gomar, J., et al., *Comparison of solution and crystal structures of maize nonspecific lipid transfer protein: a model for a potential in vivo lipid carrier protein.* Proteins: Structure, Function, and Bioinformatics, 1998. **31**(2): p. 160-171.
129. Tassin, S., et al., *Solution structure of Ace-AMP1, a potent antimicrobial protein extracted from onion seeds. Structural analogies with plant nonspecific lipid transfer proteins.* Biochemistry, 1998. **37**(11): p. 3623-3637.
130. Lee, J.Y., et al., *Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity.* Journal of molecular biology, 1998. **276**(2): p. 437-448.
131. Han, G.W., et al., *Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography.* Journal of molecular biology, 2001. **308**(2): p. 263-278.
132. Merrifield, R.B., *Solid phase peptide synthesis. I. The synthesis of a tetrapeptide.* Journal of the American Chemical Society, 1963. **85**(14): p. 2149-2154.

133. Palomo, J.M., *Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides.* Rsc Advances, 2014. **4**(62): p. 32658-32672.
134. Woycechowsky, K.J., K.D. Wittrup, and R.T. Raines, *A small-molecule catalyst of protein folding in vitro and in vivo.* Chemistry & biology, 1999. **6**(12): p. 871-879.
135. Steiner, A.M. and G. Bulaj, *Optimization of oxidative folding methods for cysteine-rich peptides: a study of conotoxins containing three disulfide bridges.* Journal of Peptide Science, 2011. **17**(1): p. 1-7.
136. Bach, R.D., O. Dmitrenko, and C. Thorpe, *Mechanism of thiolate− disulfide interchange reactions in biochemistry.* The Journal of organic chemistry, 2008. **73**(1): p. 12-21.
137. Snow, J.T., J.W. Finley, and M. Friedman, *Oxidation of sulfhydryl groups to disulfides by sulfoxides.* Biochemical and biophysical research communications, 1975. **64**(1): p. 441-447.
138. Annis, I., B. Hargittai, and G. Barany, *[10] Disulfide bond formation in peptides.* Methods in enzymology, 1997. **289**: p. 198-221.
139. Kubo, S., et al., *Oxidative folding of ω-conotoxin MVIIC: Effects of temperature and salt.* Biopolymers, 1996. **38**(6): p. 733-744.
140. DeLa Cruz, R., et al., *Detergent-assisted oxidative folding of δ-conotoxins.* The Journal of peptide research, 2003. **61**(4): p. 202-212.
141. Craik, D.J. and D.J. Adams, *Chemical modification of conotoxins to improve stability and activity.* ACS chemical biology, 2007. **2**(7): p. 457-468.
142. Shelenkov, A., A. Slavokhotova, and T. Odintsova, *Predicting antimicrobial and other cysteine-rich peptides in 1267 plant transcriptomes.* Antibiotics, 2020. **9**(2): p. 60.
143. Kumari, G., et al., *Molecular diversity and function of jasmintides from Jasminum sambac.* BMC plant biology, 2018. **18**(1): p. 1-13.
144. Wong, K.H., et al., *Ginkgotides: Proline-Rich Hevein-Like Peptides from Gymnosperm Ginkgo biloba.* Frontiers in Plant Science, 2016. **7**.
145. Zahiruddin, S., et al., *Ashwagandha in brain disorders: A review of recent developments.* Journal of ethnopharmacology, 2020. **257**: p. 112876.
146. Kulkarni, S. and A. Dhir, *Withania somnifera: an Indian ginseng.* Progress in neuro-psychopharmacology and biological psychiatry, 2008. **32**(5): p. 1093-1105.
147. Mirjalili, M.H., et al., *Genetic and withaferin A analysis of Iranian natural populations of Withania somnifera and W. coagulans by RAPD and HPTLC.* Natural product communications, 2009. **4**(3): p. 1934578X0900400307.
148. Dar, N.J., A. Hamid, and M. Ahmad, *Pharmacologic overview of Withania somnifera, the Indian Ginseng.* Cellular and molecular life sciences, 2015. **72**(23): p. 4445-4460.
149. Samadi, A.K., *Ashwagandha: ancient medicine for modern Times.* Journal of Ancient Diseases & Preventive Remedies, 2013.
150. Forman, M. and N. Kerna, *Merging ayurvedic ashwagandha with traditional Chinese medicine Part 1. Foundation in ashwagandha: physiological effects, clinical efficacy, and properties.* Current Research in Complementary and Alternative Medicines, 2018.
151. Uddin, Q., et al., *Phytochemical and pharmacological profile of Withania somnifera Dunal: a review.* J Appl Pharm Sci, 2012. **2**(01): p. 170-175.

152. Speziale, F., *Il Central Council for research in Unani medicine, Delhi (CCRUM, Ministry of health and family welfare, Government of India).* AM. Rivista della Società Italiana di Antropologia Medica, 1999. **4**(7-8).
153. Khare, C.P., *Indian medicinal plants: an illustrated dictionary*. 2008: Springer Science & Business Media.
154. Sharma, A., V.K. Gothecha, and N.K. Ojha, *Dyslexia: A solution through Ayurveda evidences from Ayurveda for the management of dyslexia in children: A review.* Ayu, 2012. **33**(4): p. 486.
155. Scartezzini, P. and E. Speroni, *Review on some plants of Indian traditional medicine with antioxidant activity.* Journal of ethnopharmacology, 2000. **71**(1-2): p. 23-43.
156. Grandhi, A., A. Mujumdar, and B. Patwardhan, *A comparative pharmacological investigation of Ashwagandha and Ginseng.* Journal of ethnopharmacology, 1994. **44**(3): p. 131-135.
157. Singh, B., et al., *Adaptogenic activity of a novel, withanolide-free aqueous fraction from the roots of Withania somnifera Dun.* Phytotherapy Research, 2001. **15**(4): p. 311-318.
158. Das, R., et al., *Role of Withaferin A and Its Derivatives in the Management of Alzheimer's Disease: Recent Trends and Future Perspectives.* Molecules, 2021. **26**(12): p. 3696.
159. Dutta, R., et al., *Withania somnifera (Ashwagandha) and withaferin A: Potential in integrative oncology.* International journal of molecular sciences, 2019. **20**(21): p. 5310.
160. Straughn, A.R. and S.S. Kakar, *Withaferin A: a potential therapeutic agent against COVID-19 infection.* Journal of Ovarian Research, 2020. **13**(1): p. 1-5.
161. RajaSankar, S., T. Manivasagam, and S. Surendran, *Ashwagandha leaf extract: a potential agent in treating oxidative damage and physiological abnormalities seen in a mouse model of Parkinson's disease.* Neuroscience letters, 2009. **454**(1): p. 11-15.
162. Sankar, S.R., et al., *The neuroprotective effect of Withania somnifera root extract in MPTP-intoxicated mice: An analysis of behavioral and biochemical varibles.* Cellular & molecular biology letters, 2007. **12**(4): p. 473-481.
163. Prakash, J., et al., *Neuroprotective role of Withania somnifera root extract in Maneb–Paraquat induced mouse model of parkinsonism.* Neurochemical research, 2013. **38**(5): p. 972-980.
164. Prakash, J., et al., *Withania somnifera alleviates parkinsonian phenotypes by inhibiting apoptotic pathways in dopaminergic neurons.* Neurochemical research, 2014. **39**(12): p. 2527-2536.
165. Manjunath, M., *Standardized extract of Withania somnifera (Ashwagandha) markedly offsets rotenone-induced locomotor deficits, oxidative impairments and neurotoxicity in Drosophila melanogaster.* Journal of food science and technology, 2015. **52**(4): p. 1971-1981.
166. Kumar, P. and A. Kumar, *Possible neuroprotective effect of Withania somnifera root extract against 3-nitropropionic acid-induced behavioral, biochemical, and mitochondrial dysfunction in an animal model of Huntington's disease.* Journal of medicinal food, 2009. **12**(3): p. 591-600.
167. Bisht, P. and V. Rawat, *Antibacterial activity of Withania somnifera against Gram-positive isolates from pus samples.* Ayu, 2014. **35**(3): p. 330.

168. Singh, G. and P. Kumar, *Evaluation of antimicrobial efficacy of flavonoids of Withania somnifera L.* Indian journal of pharmaceutical sciences, 2011. **73**(4): p. 473.
169. Alam, N., et al., *Methanolic extracts of Withania somnifera leaves, fruits and roots possess antioxidant properties and antibacterial activities.* BMC complementary and alternative medicine, 2012. **12**(1): p. 1-8.
170. Mwitari, P.G., et al., *Antimicrobial activity and probable mechanisms of action of medicinal plants of Kenya: Withania somnifera, Warbugia ugandensis, Prunus africana and Plectrunthus barbatus.* PloS one, 2013. **8**(6): p. e65619.
171. Arora, S., et al., *The in vitro antibacterial/synergistic activities of Withania somnifera extracts.* Fitoterapia, 2004. **75**(3-4): p. 385-388.
172. Chandrasekaran, S., et al., *An in vitro study of apoptotic like death in Leishmania donovani promastigotes by withanolides.* Parasitology international, 2013. **62**(3): p. 253-261.
173. Dikasso, D., et al., *Anti-malarial activity of withania somnifera L. Dunal extracts in mice.* Ethiopian Medical Journal, 2006. **44**(3): p. 279-285.
174. Girish, K., et al., *Antimicrobial properties of a non-toxic glycoprotein (WSG) from Withania somnifera (Ashwagandha).* Journal of Basic Microbiology, 2006. **46**(5): p. 365-374.
175. Widodo, N., et al., *Selective killing of cancer cells by leaf extract of Ashwagandha: Components, activity and pathway analyses.* Cancer letters, 2008. **262**(1): p. 37-47.
176. Mathur, R., et al., *Evaluation of the effect of Withania somnifera root extracts on cell cycle and angiogenesis.* Journal of ethnopharmacology, 2006. **105**(3): p. 336-341.
177. Mayola, E., et al., *Withaferin A induces apoptosis in human melanoma cells through generation of reactive oxygen species and down-regulation of Bcl-2.* Apoptosis, 2011. **16**(10): p. 1014-1027.
178. Malik, F., et al., *Reactive oxygen species generation and mitochondrial dysfunction in the apoptotic cell death of human myeloid leukemia HL-60 cells by a dietary compound withaferin A with concomitant protection by N-acetyl cysteine.* Apoptosis, 2007. **12**(11): p. 2115-2133.
179. Hahm, E.R., J. Lee, and S.V. Singh, *Role of mitogen-activated protein kinases and Mcl-1 in apoptosis induction by withaferin A in human breast cancer cells.* Molecular carcinogenesis, 2014. **53**(11): p. 907-916.
180. Choi, M.J., et al., *Endoplasmic reticulum stress mediates withaferin A-induced apoptosis in human renal carcinoma cells.* Toxicology in Vitro, 2011. **25**(3): p. 692-698.
181. Andallu, B. and B. Radhika, *Hypoglycemic, diuretic and hypocholesterolemic effect of winter cherry (Withania somnifera, Dunal) root.* 2000.
182. Anwer, T., et al., *Effect of Withania somnifera on insulin sensitivity in non-insulin-dependent diabetes mellitus rats.* Basic & clinical pharmacology & toxicology, 2008. **102**(6): p. 498-503.
183. Udayakumar, R., et al., *Hypoglycaemic and hypolipidaemic effects of Withania somnifera root and leaf extracts on alloxan-induced diabetic rats.* International journal of molecular sciences, 2009. **10**(5): p. 2367-2382.
184. Ashour, O.M., et al., *Evaluation of the potential cardioprotective activity of some Saudi plants against doxorubicin toxicity.* Zeitschrift für Naturforschung C, 2012. **67**(5-6): p. 297-307.

185. Mohanty, I., et al., *Mechanisms of cardioprotective effect of Withania somnifera in experimentally induced myocardial infarction.* Basic & clinical pharmacology & toxicology, 2004. **94**(4): p. 184-190.
186. Mohanty, I., et al., *Cardioprotection from ischemia and reperfusion injury by Withania somnifera: a hemodynamic, biochemical and histopathological assessment.* Molecular and cellular biochemistry, 2004. **260**(1): p. 39-47.
187. Li, Z., et al., *A review of polysaccharides from Schisandra chinensis and Schisandra sphenanthera: Properties, functions and applications.* Carbohydrate polymers, 2018. **184**: p. 178-190.
188. Panossian, A. and G. Wikman, *Pharmacology of Schisandra chinensis Bail.: an overview of Russian research and uses in medicine.* Journal of ethnopharmacology, 2008. **118**(2): p. 183-212.
189. Turova, A.D., *Medicinal Plants of the USSR and their Use.* 1974: p. 29–34.
190. Turova, A.D., Sapozhnikova, E.N, *Medicinal Plants of the USSR and their Use.* 1982.
191. USSR, *National Pharmacopoeia of the USSR.* 1968.
192. USSR, *National Pharmacopoeia of the USSR.* 1990.
193. Lupandin, A., *Schizandra chinensis effect on some parameters of the organism reactivity.* Materials on the 23rd Scientific Session of Khabarovsk Medical Institute, 1965. **151**: p. 153.
194. Wang, Z., et al., *Comparative studies on the chemical composition and antioxidant activities of Schisandra chinensis and Schisandra sphenanthera fruits.* Journal of Medicinal Plants Research, 2011. **5**(7): p. 1207-1216.
195. Balandin, D., *Schizandrin—a new stimulant from Schizandra fruits.* Materials for the Study of Stimulants and Tonics from Ginseng and Schizandra Roots, 1951: p. 45-50.
196. Szopa, A., R. Ekiert, and H. Ekiert, *Current knowledge of Schisandra chinensis (Turcz.) Baill.(Chinese magnolia vine) as a medicinal plant species: a review on the bioactive components, pharmacological properties, analytical and biotechnological studies.* Phytochemistry Reviews, 2017. **16**(2): p. 195-218.
197. Brekhman, I. and I. Dardymov, *New substances of plant origin which increase nonspecific resistance.* Annual review of pharmacology, 1969. **9**(1): p. 419-430.
198. Panossian, A., G. Wikman, and H. Wagner, *Plant adaptogens III. Earlier and more recent aspects and concepts on their mode of action.* Phytomedicine, 1999. **6**(4): p. 287-300.
199. Murtazin, I., *Schizandra chinensis as a stimulant under long physical loads.* Farmakol Toxikol, 1946. **9**: p. 12-13.
200. Galant, I., *Materials for the Study of Ginseng and Schizandra.* 1958.
201. Lebedev, A., *Schizandrin—a new stimulant from Schizandra chinensis fruits.* Summary of Thesis for a Candidate's Degree in Medicine, Tashkent, 1967.
202. Markova, L. and R. Samoilova, *The study of Schizandra effect on the higher nervous activity in cases of traumatic encephalopathy.* Transactions of the Khabarovsk Medical Institute, 1954. **13**: p. 72-77.
203. Gaistruk, A. and K. Taranovskij, *The treatment of arterial hypotension in pregnant women using Schizandra chinensis*, in *Urgent Problems of Obstetics and Gynecology*. 1968, Ministry of Health of Ukrainian SSR L'vov.

204. Song, F., et al., *Schizandrin A inhibits microglia-mediated neuroninflammation through inhibiting TRAF6-NF-κB and Jak2-Stat3 signaling pathways.* PLoS One, 2016. **11**(2): p. e0149991.
205. Wang, C.-P., et al., *Neuroprotective effect of schizandrin A on oxygen and glucose deprivation/reperfusion-induced cell injury in primary culture of rat cortical neurons.* Journal of physiology and biochemistry, 2014. **70**(3): p. 735-747.
206. Tang, M., et al., *Protection of seven dibenzocyclooctadiene lignans from Schisandra chinensis against serum and glucose deprivation injury in SH-SY5Y cells.* Cell biology international, 2015. **39**(12): p. 1418-1424.
207. Hu, D., et al., *Deoxyschizandrin isolated from the fruits of Schisandra chinensis ameliorates Aβ1–42-induced memory impairment in mice.* Planta medica, 2012. **78**(12): p. 1332-1336.
208. Lacolley, P., et al., *The vascular smooth muscle cell in arterial pathology: a cell that can take on multiple roles.* Cardiovascular research, 2012. **95**(2): p. 194-204.
209. Kong, P., P. Christia, and N.G. Frangogiannis, *The pathogenesis of cardiac fibrosis.* Cellular and molecular life sciences, 2014. **71**(4): p. 549-574.
210. Lan, T.-H., X.-Q. Huang, and H.-M. Tan, *Vascular fibrosis in atherosclerosis.* Cardiovascular Pathology, 2013. **22**(5): p. 401-407.
211. Libby, P., *Inflammation and cardiovascular disease mechanisms.* The American journal of clinical nutrition, 2006. **83**(2): p. 456S-460S.
212. Maskrey, B.H., et al., *Mechanisms of resolution of inflammation: a focus on cardiovascular disease.* Arteriosclerosis, thrombosis, and vascular biology, 2011. **31**(5): p. 1001-1006.
213. Lee, Y. and Å.B. Gustafsson, *Role of apoptosis in cardiovascular disease.* Apoptosis, 2009. **14**(4): p. 536-548.
214. Park, J.Y., et al., *Gomisin A from Schisandra chinensis induces endothelium-dependent and direct relaxation in rat thoracic aorta.* Planta medica, 2007. **73**(15): p. 1537-1542.
215. Chen, P., et al., *Beneficial effects of schisandrin B on the cardiac function in mice model of myocardial infarction.* PloS one, 2013. **8**(11): p. e79418.
216. Chiu, P.Y., et al., *Schisandrin B stereoisomers protect against hypoxia/reoxygenation-induced apoptosis and inhibit associated changes in Ca2+-induced mitochondrial permeability transition and mitochondrial membrane potential in H9c2 cardiomyocytes.* Life Sciences, 2008. **82**(21-22): p. 1092-1101.
217. Schlessinger, J., *The mechanism and role of hormone-induced clustering of membrane receptors.* Trends in Biochemical Sciences, 1980. **5**(8): p. 210-214.
218. Dunn, W.A. and A.L. Hubbard, *Receptor-mediated endocytosis of epidermal growth factor by hepatocytes in the perfused rat liver: ligand and receptor dynamics.* The Journal of Cell Biology, 1984. **98**(6): p. 2148-2159.
219. Brown, M.S., R.G. Anderson, and J.L. Goldstein, *Recycling receptors: the round-trip itinerary of migrant membrane proteins.* Cell, 1983. **32**(3): p. 663-667.
220. Helenius, A., et al., *Endosomes.* Trends in Biochemical Sciences, 1983. **8**(7): p. 245-250.
221. Robenek, H. and A. Hesz, *Dynamics of low-density lipoprotein receptors in the plasma membrane of cultured human skin fibroblasts as visualized by*

colloidal gold in conjunction with surface replicas. European journal of cell biology, 1983. **31**(2): p. 275-282.
222. Goldstein, J.L., et al., *Receptor-mediated endocytosis: concepts emerging from the LDL receptor system.* Annual review of cell biology, 1985. **1**(1): p. 1-39.
223. Innerarity, T., et al., *Normalization of receptor binding of apolipoprotein E2. Evidence for modulation of the binding site conformation.* Journal of Biological Chemistry, 1984. **259**(11): p. 7261-7267.
224. Russell, D.W., et al., *Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor.* Cell, 1984. **37**(2): p. 577-585.
225. Yamamoto, T., et al., *The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA.* Cell, 1984. **39**(1): p. 27-38.
226. Südhof, T.C., et al., *Cassette of eight exons shared by genes for LDL receptor and EGF precursor.* Science, 1985. **228**(4701): p. 893-895.
227. Scott, J., et al., *Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.* Science, 1983. **221**(4607): p. 236-240.
228. Doolittle, R.F., F. Da Feng, and M.S. Johnson, *Computer-based characterization of epidermal growth factor precursor.* Nature, 1984. **307**(5951): p. 558-560.
229. Südhof, T.C., et al., *The LDL receptor gene: a mosaic of exons shared with different proteins.* Science, 1985. **228**(4701): p. 815-822.
230. Maxwell, K.N. and J.L. Breslow, *Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype.* Proceedings of the National Academy of Sciences, 2004. **101**(18): p. 7100-7105.
231. Rashid, S., et al., *Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9.* Proceedings of the National Academy of Sciences, 2005. **102**(15): p. 5374-5379.
232. Cohen, J., et al., *Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9.* Nature genetics, 2005. **37**(2): p. 161-165.
233. Seidah, N.G. and A. Prat, *The biology and therapeutic targeting of the proprotein convertases.* Nature reviews Drug discovery, 2012. **11**(5): p. 367-383.
234. Cunningham, D., et al., *Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia.* Nature structural & molecular biology, 2007. **14**(5): p. 413-419.
235. Seidah, N.G., et al., *The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation.* Proceedings of the National Academy of Sciences, 2003. **100**(3): p. 928-933.
236. Abifadel, M., et al., *Living the PCSK9 adventure: from the identification of a new gene in familial hypercholesterolemia towards a potential new class of anticholesterol drugs.* Current atherosclerosis reports, 2014. **16**(9): p. 1-23.
237. Zhao, Z., et al., *Molecular characterization of loss-of-function mutations in PCSK9 and identification of a compound heterozygote.* The American Journal of Human Genetics, 2006. **79**(3): p. 514-523.
238. Mayne, J., et al., *Novel loss-of-function PCSK9 variant is associated with low plasma LDL cholesterol in a French-Canadian family and with impaired*

processing and secretion in cell culture. Clinical chemistry, 2011. **57**(10): p. 1415-1423.
239. Seidah, N.G., et al., *PCSK9: a key modulator of cardiovascular health.* Circulation research, 2014. **114**(6): p. 1022-1036.
240. Naoumova, R.P., et al., *Severe hypercholesterolemia in four British families with the D374Y mutation in the PCSK9 gene: long-term follow-up and treatment response.* Arteriosclerosis, thrombosis, and vascular biology, 2005. **25**(12): p. 2654-2660.
241. Hopkins, P.N., et al., *Characterization of autosomal dominant hypercholesterolemia caused by PCSK9 gain of function mutations and its specific treatment with alirocumab, a PCSK9 monoclonal antibody.* Circulation: Cardiovascular Genetics, 2015. **8**(6): p. 823-831.
242. Burke, A.C., et al., *PCSK9: regulation and target for drug development for dyslipidemia.* Annual review of pharmacology and toxicology, 2017. **57**: p. 223-244.
243. Jeong, H.J., et al., *Sterol-dependent regulation of proprotein convertase subtilisin/kexin type 9 expression by sterol-regulatory element binding protein-2s.* Journal of lipid research, 2008. **49**(2): p. 399-409.
244. Horton, J.D., J.L. Goldstein, and M.S. Brown, *SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver.* The Journal of clinical investigation, 2002. **109**(9): p. 1125-1131.
245. Li, H., et al., *Hepatocyte nuclear factor 1α plays a critical role in PCSK9 gene transcription and regulation by the natural hypocholesterolemic compound berberine.* Journal of Biological Chemistry, 2009. **284**(42): p. 28885-28895.
246. Ai, D., et al., *Regulation of hepatic LDL receptors by mTORC1 and PCSK9 in mice.* The Journal of clinical investigation, 2012. **122**(4).
247. Benjannet, S., et al., *NARC-1/PCSK9 and its natural mutants: zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol.* Journal of Biological Chemistry, 2004. **279**(47): p. 48865-48875.
248. Park, S.W., Y.-A. Moon, and J.D. Horton, *Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver.* Journal of Biological Chemistry, 2004. **279**(48): p. 50630-50638.
249. Holla, Ø.L., et al., *Degradation of the LDL receptors by PCSK9 is not mediated by a secreted protein acted upon by PCSK9 extracellularly.* BMC cell biology, 2007. **8**(1): p. 1-12.
250. Needham, M. and F. Mastaglia, *Statin myotoxicity: a review of genetic susceptibility factors.* Neuromuscular Disorders, 2014. **24**(1): p. 4-15.
251. Jeener, J., et al., *Investigation of exchange processes by two-dimensional NMR spectroscopy.* The Journal of chemical physics, 1979. **71**(11): p. 4546-4553.
252. Anil-Kumar, E.R., R. Ernst, and K. Wüthrich, *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules.* Biochem. Biophys. Res. Commun, 1980. **64**: p. 2229-2246.
253. Liu, M., et al., *Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy.* Journal of Magnetic Resonance, 1998. **132**(1): p. 125-129.
254. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-293.

255. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2015. **31**(8): p. 1325-1327.
256. Schwieters, C.D., J.J. Kuszewski, and G.M. Clore, *Using Xplor–NIH for NMR molecular structure determination.* Progress in nuclear magnetic resonance spectroscopy, 2006. **48**(1): p. 47-62.
257. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* Journal of biomolecular NMR, 1996. **8**(4): p. 477-486.
258. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.
259. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nature methods, 2011. **8**(10): p. 785-786.
260. Madeira, F., et al., *The EMBL-EBI search and sequence analysis tools APIs in 2019.* Nucleic acids research, 2019. **47**(W1): p. W636-W641.
261. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic acids research, 2021. **49**(W1): p. W293-W296.
262. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome research, 2004. **14**(6): p. 1188-1190.
263. Schneider, T.D. and R.M. Stephens, *Sequence logos: a new way to display consensus sequences.* Nucleic acids research, 1990. **18**(20): p. 6097-6100.
264. Leebens-Mack, J.H., et al., *One thousand plant transcriptomes and the phylogenomics of green plants.* Nature, 2019. **574**(7780): p. 679-685.
265. Kaas, Q., et al., *ConoServer: updated content, knowledge, and discovery tools in the conopeptide database.* Nucleic acids research, 2012. **40**(D1): p. D325-D330.
266. Kaas, Q., et al., *ConoServer, a database for conopeptide sequences and structures.* Bioinformatics, 2008. **24**(3): p. 445-446.
267. Sakurai, T. and K. Kataoka, *Basic and applied features of multicopper oxidases, CueO, bilirubin oxidase, and laccase.* The Chemical Record, 2007. **7**(4): p. 220-229.
268. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server.* The proteomics protocols handbook, 2005: p. 571-607.
269. Kam, A., et al., *Plant-derived mitochondria-targeting cysteine-rich peptide modulates cellular bioenergetics.* Journal of Biological Chemistry, 2019. **294**(11): p. 4000-4011.
270. Solntsev, S.K., et al., *Enhanced global post-translational modification discovery with MetaMorpheus.* Journal of proteome research, 2018. **17**(5): p. 1844-1851.
271. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification–mass spectrometry data.* Nature methods, 2013. **10**(8): p. 730-736.
272. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic acids research, 2021. **49**(D1): p. D605-D612.

273. Mukherjee, P.K., et al., *Withania somnifera (L.) Dunal-Modern perspectives of an ancient Rasayana from Ayurveda.* Journal of Ethnopharmacology, 2021. **264**: p. 113157.
274. India, M.o.H. and W. Family, *Pharmacopoeia of India (the Indian pharmacopoeia)*. 1985, Delhi: Controller of Publications.
275. Saleem, S., et al., *Withania somnifera L.: Insights into the phytochemical profile, therapeutic potential, clinical trials, and future prospective.* Iranian Journal of Basic Medical Sciences, 2020. **23**(12): p. 1501.
276. Chen, L.-X., H. He, and F. Qiu, *Natural withanolides: an overview.* Natural product reports, 2011. **28**(4): p. 705-740.
277. Kumari, G., et al., *Cysteine-rich peptide family with unusual disulfide connectivity from Jasminum sambac.* Journal of natural products, 2015. **78**(11): p. 2791-2799.
278. Shen, Y., et al., *Potentides: New Cysteine-Rich Peptides with Unusual Disulfide Connectivity from Potentilla anserina.* ChemBioChem, 2019. **20**(15): p. 1995-2004.
279. Wong, K.H., et al., *Vaccatides: antifungal glutamine-rich hevein-like peptides from Vaccaria hispanica.* Frontiers in plant science, 2017. **8**: p. 1100.
280. Kaas, Q., J.-C. Westermann, and D.J. Craik, *Conopeptide characterization and classifications: an analysis using ConoServer.* Toxicon, 2010. **55**(8): p. 1491-1509.
281. Craik, D.J., et al., *The future of peptide-based drugs.* Chemical biology & drug design, 2013. **81**(1): p. 136-147.
282. Driggers, E.M., et al., *The exploration of macrocycles for drug discovery—an underexploited structural class.* Nature Reviews Drug Discovery, 2008. **7**(7): p. 608-624.
283. Jin, A.-H., et al., *Conotoxins: Chemistry and biology.* Chemical reviews, 2019. **119**(21): p. 11510-11549.
284. Dardevet, L., et al., *Chlorotoxin: a helpful natural scorpion peptide to diagnose glioma and fight tumor invasion.* Toxins, 2015. **7**(4): p. 1079-1101.
285. Giovannucci, E., *Insulin, insulin-like growth factors and colon cancer: a review of the evidence.* The Journal of nutrition, 2001. **131**(11): p. 3109S-3120S.
286. Muttenthaler, M., et al., *Trends in peptide drug discovery.* Nature reviews Drug discovery, 2021. **20**(4): p. 309-325.
287. Huang, J., et al., *Identification and characterization of a wolfberry carboxypeptidase inhibitor from Lycium barbarum.* Food Chemistry, 2021. **351**: p. 129338.
288. Miljanich, G., *Ziconotide: neuronal calcium channel blocker for treating severe chronic pain.* Current medicinal chemistry, 2004. **11**(23): p. 3029-3040.
289. Wong, K.H., et al., *β-Ginkgotides: Hyperdisulfide-constrained peptides from Ginkgo biloba.* Scientific reports, 2017. **7**(1): p. 1-13.
290. Ray, K., *Linaclotide approved for constipation-predominant IBS.* Nature reviews Gastroenterology & hepatology, 2012. **9**(11): p. 616-616.
291. Wang, C.-Z., et al., *A novel conotoxin from Conus striatus, μ-SIIIA, selectively blocking rat tetrodotoxin-resistant sodium channels.* Toxicon, 2006. **47**(1): p. 122-132.
292. Sun, J.-H., et al., *Metabolomics study of the therapeutic mechanism of Schisandra Chinensis lignans in diet-induced hyperlipidemia mice.* Lipids in Health and Disease, 2017. **16**(1): p. 1-14.

293. Jin, A.-H., et al., *Conotoxins: Chemistry and Biology.* Chemical Reviews, 2019. **119**(21): p. 11510-11549.
294. Mitchell, S.S., et al., *Three-dimensional solution structure of conotoxin ψ-PIIIE, an acetylcholine gated ion channel antagonist.* Biochemistry, 1998. **37**(5): p. 1215-1220.
295. Tietze, A.A., et al., *Structurally Diverse μ-Conotoxin PIIIA Isomers Block Sodium Channel NaV1. 4.* Angewandte Chemie International Edition, 2012. **51**(17): p. 4058-4061.
296. Lewis, R.J., et al., *Isolation and structure-activity of μ-conotoxin TIIIA, a potent inhibitor of tetrodotoxin-sensitive voltage-gated sodium channels.* Molecular pharmacology, 2007. **71**(3): p. 676-685.
297. Wakamatsu, K., et al., *Structure-activity relationships of. mu.-conotoxin GIIIA: Structure determination of active and inactive sodium channel blocker peptides by NMR and simulated annealing calculations.* Biochemistry, 1992. **31**(50): p. 12577-12584.
298. 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.* Biochemistry, 1996. **35**(27): p. 8824-8835.
299. Kuang, Z., et al., *Mammalian neuronal sodium channel blocker μ-conotoxin BuIIIB has a structured N-terminus that influences potency.* ACS chemical biology, 2013. **8**(6): p. 1344-1351.
300. Lazarev, N., *Experimental data on evaluation of Far east Shizandra as a stimulator.* Proceedings Scientific Medical Council, 1946. **5**(17): p. 62-68.
301. Lebedev, A., *Comparative evaluation of the stimulating effect of various Schizandra products.* Materials for the study of stimulants and tonics from Ginseng and Schizandra roots, 1951(1): p. 103-8.
302. Lebedev, A., *On the pharmacology of Schizandra.* Materials for the Study of Ginseng and Schizandra. Far East Branch of USSR Academy of Science, Moscow, 1955: p. 178-188.
303. Kochetkov, N., A.Y. Khorlin, and O. Chizhov, *Chemical study of Schizandra chinensis. I. Schizandrin and related compounds.* Zh. Obshch. Khim, 1961. **31**: p. 3454-60.
304. Balandin, D., *Fatty oil from Schizandra seeds.* Doclady Academii Nauk SSSR, 1940. **26**: p. 592-594.
305. Kochetkov, N., A. Khorlin, and O. Chizhov, *Deoxyschizandrin-structure and total synthesis.* Tetrahedron Letters, 1962. **3**(9): p. 361-363.
306. Ikeya, Y., et al., *The constituents of Schizandra chinensis Baill. I. Isolation and structure determination of five new lignans, gomisin A, B, C, F and G, and the absolute structure of schizandrin.* Chemical and Pharmaceutical Bulletin, 1979. **27**(6): p. 1383-1394.
307. IKEYA, Y., et al., *The constituents of Schizandra chinensis Baill. II. The structure of a new lignan, gomisin D.* Chemical and Pharmaceutical Bulletin, 1979. **27**(6): p. 1395-1401.
308. Madani, F., et al., *Mechanisms of cellular uptake of cell-penetrating peptides.* Journal of biophysics, 2011. **2011**.
309. Doherty, G.J. and H.T. McMahon, *Mechanisms of endocytosis.* Annual review of biochemistry, 2009. **78**: p. 857-902.
310. Goldstein, J.L. and M.S. Brown, *The LDL receptor.* Arteriosclerosis, thrombosis, and vascular biology, 2009. **29**(4): p. 431-438.

311. Simons, K. and D. Toomre, *Lipid rafts and signal transduction.* Nature reviews Molecular cell biology, 2000. **1**(1): p. 31-39.
312. Lorent, J.H., et al., *Structural determinants and functional consequences of protein affinity for membrane rafts.* Nature communications, 2017. **8**(1): p. 1-10.
313. Sezgin, E., et al., *The mystery of membrane organization: composition, regulation and roles of lipid rafts.* Nature reviews Molecular cell biology, 2017. **18**(6): p. 361-374.
314. Vitiello, G., et al., *Fusion of raft-like lipid bilayers operated by a membranotropic domain of the HSV-type I glycoprotein gH occurs through a cholesterol-dependent mechanism.* Soft matter, 2015. **11**(15): p. 3003-3016.
315. Gupta, G.L. and A. Rana, *Withania somnifera (Ashwagandha): a review.* Pharmacognosy Reviews, 2007. **1**(1).
316. Schafmeister, C.E., J. Po, and G.L. Verdine, *An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides.* Journal of the American Chemical Society, 2000. **122**(24): p. 5891-5892.
317. Klein, M.A., *Stabilized helical peptides: A strategy to target protein–protein interactions*. 2014, ACS Publications. p. 838-839.
318. Hancke, J., R. Burgos, and F. Ahumada, *Schisandra chinensis (Turcz.) baill.* Fitoterapia, 1999. **70**(5): p. 451-471.
319. Ock, E.-S., *Effect of Schizandra chinensis extract in hyperlipidemic rats.* JOURNAL-KOREAN SOCIETY OF FOOD AND NUTRITION, 1995. **24**: p. 658-662.

Publications

1. TANG, F., KAM, A., LOO, S., WONG, K. H., TAY, S. V., HUANG, J., TAM, J. P. Discovery of a plant-derived hyperdisulfide peptide that induces LDLR-mediated LDL uptake at lipid rafts. (Manuscript in preparation). 2022.

2. TANG, F., LOO, S., KAM, A., WONG, K. H., TAY, S. V., HUANG, J., TAM, J. P. Wisotide: An anionic cystine-stapled helical peptide from *Withania somnifera* with cell-membrane permeability. (Manuscript in preparation). 2022.

3. Tang, F., Loo, S., Kam, A., & Tam, J. P. Cystine-stapled Helical Peptide from Withania somnifera Is Highly Stable and Cell-permeable. The FASEB Journal. 2022

4. LOO, S., TAY, S. V., KAM, A., TANG, F., FAN, J.-S., YANG, D. & TAM, J. P. Anti-fungal hevein-like peptides biosynthesized from quinoa cleavable hololectins. Molecules. 2021.